3.4 Ancient DNA: a genetic time capsule

In which we discuss ancient DNA, and how the sequencing of ancient genomes has reshaped our understanding of human prehistory [517] [a].

We'll describe how aDNA has driven a paradigm shift in how we study human prehistory. This includes human relations with Neanderthals, the discovery of a new archaic species, the Denisovans, and studying human migrations in all parts of the world. While there's already far too much work on human aDNA for us to cover everything, we'll close with a case-study of the fascinating complexity of European prehistory.

[a] ***Ancient DNA (aDNA)*** *refers to DNA sequences retrieved from post-mortem biological samples, usually skeletal remains, from humans or other species.*

A short history of human ancient DNA. Recall that in 1856 workers in a German quarry in the Neander Valley happened across a remarkable collection of bones. These Neanderthal remains, dating to 40 KYA, were the first recognized example of an archaic hominin. 135 years later, in 1991, a research team in Germany led by the geneticist Svante Pääbo started testing whether one of these bone fragments might contain intact biological molecules including DNA [518].

At this time there had already been some early successes in sequencing DNA from ancient materials. In 1984, Allan Wilson's lab had obtained DNA from a museum specimen of a quagga (a zebra-like species that went extinct in 1883). There had also been successful DNA extractions from other relatively recent museum specimens such as the marsupial wolf and New Zealand moa [519].

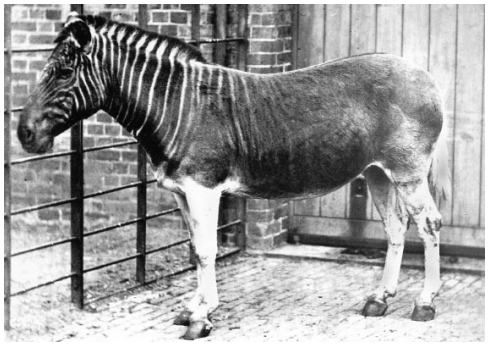


Figure 3.72: **The extinct quagga** *was the first application of aDNA sequencing, in 1984. This rather poignant photo from 1870 shows one of the last quaggas.* *Credit: Frederick York [Link]. Public Domain.*

However, it was also becoming clear that ancient DNA studies require great caution. DNA degrades over time, and ancient samples often have little or no remaining ancient DNA. But unless extreme care is taken, it is easy to contaminate the samples with modern DNA. In fact, the developing field took a bit of a black eye in the early 1990s from high-profile – but wrong – reports that DNA had been extracted from insects and plants trapped in amber for tens of millions of years, and even from Cretaceous dinosaur bones [520]. We now know that DNA very rarely survives beyond a few tens of thousands of years.

But in this case, initial testing of the Neanderthal specimen was promising, so the group next attempted PCR-amplification of short fragments of mitochondrial DNA. They focused on mtDNA for the same reasons that the out-of-Africa studies a few years earlier had used mtDNA: it is present at very high copy numbers per cell, and it is much more variable than most nuclear regions of the genome.

After six years of work, in a paper that was widely covered by the international media, Pääbo's team reported success in sequencing mtDNA from this bone [521]. Their experimental strategy used PCR amplification of 13 short overlapping DNA segments that spanned a total of 379 bp; amplicons were cloned into plasmid vectors and sequenced by traditional Sanger sequencing.

Given that these samples have been handled by researchers and museum staff for more than 130 years, there was a concern that PCR might amplify modern DNA contaminating the bone. To help address this, the team sent a separate bone sample to collaborators at Penn State University in the US, who were able to confirm the sequences of two amplicons. Here's a look at their first results:

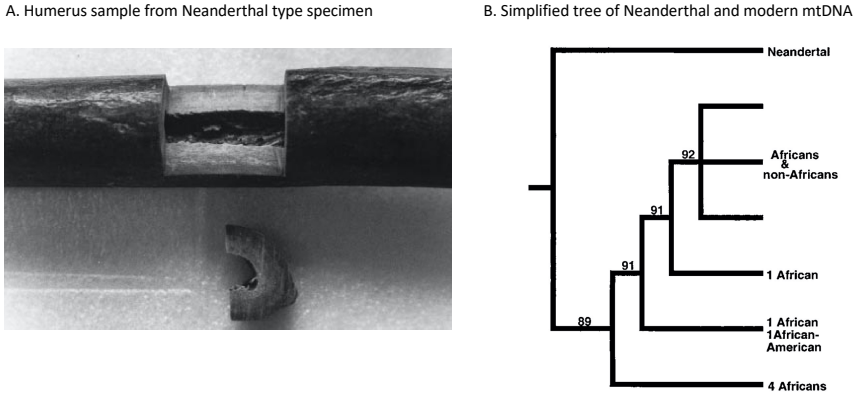


Figure 3.73: **The first Neanderthal sequencing (1997). A.** *DNA was obtained from a small fragment of the humerus.* **B.** *A simplified phylogenetic tree comparing the Neanderthal mtDNA sequence to 986 modern human sequences shows that* ***Neanderthal mtDNA lies outside the tree of all human sequences****.*

The numbers represent statistical support of the branching order based on bootstrapping. Credit: Modified from Figs 1, 7a of Matthias Krings et al (1997) [Link]. Used with permission.

One question of popular interest was whether Neanderthals might have interbred with modern humans, as they likely came into contact with one another in Europe and in the Middle East. If this had occurred, then we could detect it using haplotypes in modern humans that are closely related to those in Neanderthals [b] [522].

Notably, comparison of the Neanderthal sequence to modern mtDNA showed that it lay completely outside the distribution of modern human variation [c]. This indicates that this Neanderthal individual's mtDNA was not closely related to any lineages that contributed to modern humans. It also fit with the observations from a few years earlier showing a recent MRCA for human mtDNA [523], and the emerging consensus of a Recent African Origin for modern humans.

During the following decade, mitochondrial sequences were obtained from about a dozen Neanderthals. These results supported the main results from the first paper: Neanderthal mitochondrial sequences form their own clade, completely outside the human mitochondrial tree. Based on this work, many people in the field suspected that early humans had completely replaced Neanderthals (by around 35 KYA) with no gene flow at all from Neanderthals into humans. And yet it was also clear that introgression could not be definitively ruled out from just a single region of the genome: what if variation of mtDNA was atypical for some reason [524]? And lastly, there were examples of deeply divergent sequences in modern humans that hinted at Neanderthal admixture [525].

The next phase in the story was enabled by the huge technical advances in massively parallel sequencing that were emerging in the early 2000s. The first generation of Neanderthal sequencing relied on PCR amplification of targeted regions and Sanger sequencing. But with the new sequencing methods it was practical to start from a random soup of indi-

[b] *A note on terminology: Gene flow between species is referred to* ***introgression****, while gene flow between populations is referred to as* ***admixture****. Humans and Neanderthals can be viewed as either diverged populations or closely related species, and so both terms are used somewhat interchangeably, depending on context.*

[c] *The finding that this sequence was distinct from all modern humans was also critical evidence that it was not simply a contamination artifact.*

vidual DNA molecules, where each molecule is a starting template for sequencing – without any need to amplify specific sequences. This opened up the possibility of genome sequencing.

Meanwhile, the Pääbo lab had been screening Neanderthal bones for one that might be suitable. They identified a bone from a cave in Vindija, Croatia with unusually good DNA preservation. In a first study in 2006, they used an emerging sequencing technology called 454 Sequencing (now defunct) to shotgun sequence a total of 1 MB of DNA scattered across the nuclear genome [526]. They also shared a small fragment with the lab of sequencing expert Eddy Rubin, at the Joint Genome Institute in California. Rubin's lab collected 65 Kb of Neanderthal sequence data (my lab analyzed data for the Rubin paper) [527].

Figure 3.74: **Vindija Cave in Croatia.** *Source of the 2010 Neanderthal draft genome.* *Credit: Tomislav Kranjcic. [Link] CC BY-SA 2.0.*

At first this seemed like a huge step forward for the field, but it soon became apparent that the Pääbo data set was heavily contaminated with modern DNA. Compared to the Rubin data, the Pääbo data had a higher fraction of long sequence reads, and the long reads showed very low divergence from modern humans [528]. This is exactly the pattern that would be expected for DNA contamination: modern contaminating DNA is less degraded and thus tends to be enriched in longer DNA fragments. Following this major stumble, there was renewed emphasis on lab techniques to minimize contamination, and this laid the groundwork for the first successful genome sequencing of archaic hominins.

Ultimately, this work was rewarded by a pair of landmark studies: a 2010 paper that reported a low coverage Neanderthal "draft" genome from Vindija, and a 2014 paper that reported a high coverage Neanderthal genome from the Altai Mountains in Siberia [529]. The main bombshell result from the 2010 paper was that it provided the first direct evidence for gene flow from Neanderthals into modern humans. We now know that all non-Africans carry around 1.5%–2.0% Neanderthal ancestry.

Figure 3.75: **Vindija Neanderthal.** *Bone fragments from different individuals. DNA was recovered from all three.* *Credit: Figure 1a from Richard Green et al (2010) [Link]. Used with permission.*

In parallel, this work led to the extraordinary discovery of a new hominin, the **Denisovans**, who were a sister group to the Neanderthals. We'll cover all of these topics in more detail below.

aDNA as a tool for studying human prehistory. Starting in the early 2010s, there were also parallel efforts to study aDNA of anatomically modern humans – ranging in age from ∼40 KY for the current oldest human data [530], to many thousands of ancient genomes from the last few thousand years. Ancient DNA sequencing, especially for high-quality samples, is now routine and high-throughput [d].

This work has led to major insights into human prehistory. One consistent theme of aDNA is that population movements, admixtures, and replacements are ubiquitous. Before the rise of aDNA data, population geneticists assumed relatively simple models of recent population history: for example, distinct populations with stable levels of migration. aDNA tells us that reality is almost always far more complicated. At the end of this chapter we'll discuss the aDNA of early Europeans as an example.

[d] *A database curated by David Reich's lab reported genome data for* $>10,000$ *ancient humans as of March 2023 [Link].*

After the next section on aDNA technology, we'll return to consider the population genetics of ancient DNA in more detail.

Figure 3.76: **Woolly mammoth teeth.** *Genomic DNA was extracted from these 1.3 MY-old teeth, collected in northeastern Siberia.* *Credit: From Extended Figure 1 of Tom van der Valk et al (2021) [Link]. Used with permission.*

Ancient DNA technology. The fundamental technical challenge of ancient DNA work is that DNA degrades over time, so that there is less and less of the target DNA of interest, and contaminating DNA from other sources becomes increasingly problematic.

DNA degradation. Over time, DNA undergoes **fragmentation**, in which the DNA breaks into shorter and shorter fragments until it is too short to be sequenced or to be mapped uniquely to its correct genome location. DNA molecules also accumulate **single nucleotide errors**, although we can usually account for these in data analysis [531].

The rate of DNA decay is such that *it's often straightforward to obtain DNA from bones up to a few thousand years old, and becomes increasingly difficult for samples up to tens of thousands of years*. Successful preservation past 100 KYA is extremely rare.

The rate of DNA decay also depends greatly on climate: frozen samples are ideal (though frozen human remains are rare); otherwise *cold and dry conditions are best*. Unfortunately, this means that DNA retrieval in some of the most interesting regions of the globe – notably in Africa, and in southeast Asia – is extremely difficult past about 10,000 years.

At the time of writing, the oldest hominin data come from a limited recovery of DNA from 430,000 year-old early-stage Neanderthals at Sima de los Huesos in Spain [532]. Looking beyond hominins, the oldest credible DNA sequence data from bones are from a variety of samples preserved in permafrost, with the current record coming from mammoth DNA dating to ∼1.3 MYA and soil extracts in Greenland dating to ∼2 MYA [533].

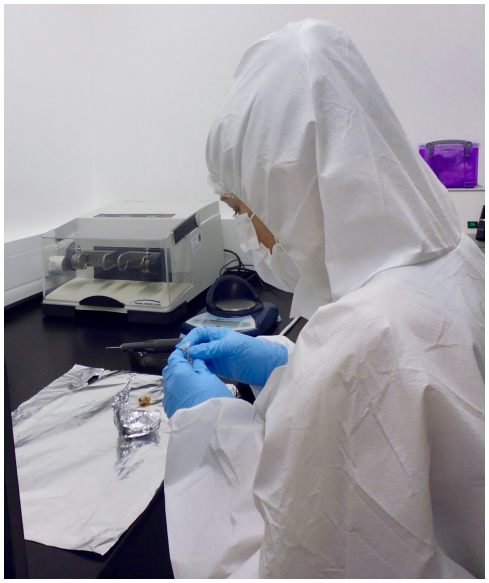


Figure 3.77: **Work in an aDNA clean room.** *Credit: Image kindly provided by Ron Pinhasi. CC BY 4*

Microbial DNA. The next major challenge is that **most ancient samples are heavily infiltrated with microbial DNA** – especially bacteria and fungi living in the bone or leaching into the bone from adjacent organic matter. A well-preserved sample may have above 50% endogenous (i.e., human) DNA, but poorly preserved samples may have less than 1% surviving hominin DNA.

The microbial DNA is unlikely to confuse the analysis, as the sequence reads should not map to the human genome, but the sequencing can become very expensive if 99% of the reads are worthless [534]. (Of course, this may still be worth doing for rare and valuable samples.)

DNA contamination from modern humans. Another serious challenge in aDNA research is the possibility of contamination with human DNA from other sources – for example anyone who has handled the bones, or cross-contamination from other samples processed in the same lab. While it's easy to exclude sequence reads from microbial DNA as they do not match the human genome, it is much harder to exclude human contaminants.

Consequently the ancient DNA field has developed strict protocols for minimizing contamination. Researchers wear gloves and masks whenever they are handling ancient material, and sample processing is conducted in dedicated clean rooms that are specialized for ancient DNA work. The bone itself is treated with UV radiation or chemical treatment to remove surface DNA prior to drilling into the bone to extract DNA from the bone's interior. As part of the protocol for extracting aDNA and preparing it for sequencing, a sample-specific DNA barcode is attached to each DNA fragment. This means that any subsequent contamination – for example at the sequencing center – can be excluded from the data analysis.

The aDNA workflow. These challenges notwithstanding, lab protocols have improved to the point that aDNA work is increasingly routine, except for the most degraded samples. Contamination is controlled by a combination of clean techniques and data analysis checks as described above. As such, **contamination is now a manageable pitfall, rather than an existential threat to the field as it was in the early days**.

A typical workflow for shotgun sequencing is shown below:

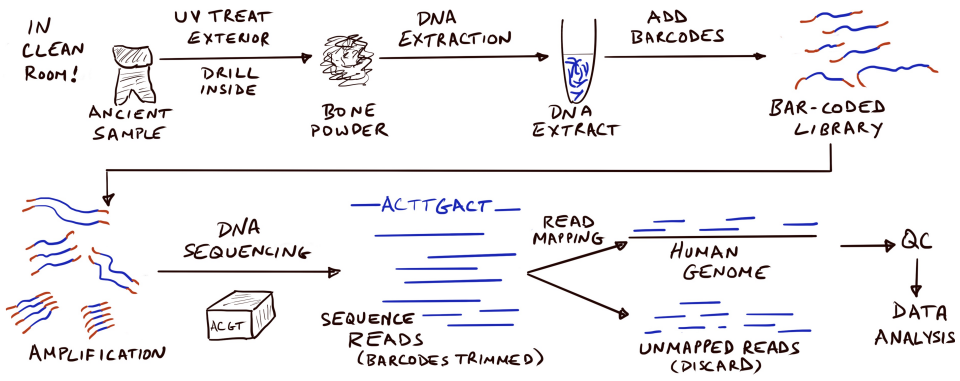


Figure 3.78: **Simplified aDNA workflow.** *The top row shows DNA extraction in a clean room to create a library of DNA fragments. The red lines indicate a short DNA barcode that identifies each sample that is added to each DNA fragment. Once this is in place the workflow can move out of the clean room. The barcodes are then computationally removed from the reads prior to read mapping. Note that it's common to perform an initial screening step using sequencing at very low read-depth, followed by deeper sequencing of the samples that pass QC.*

DNA sources. While the first Neanderthal sequencing sampled from the humerus (arm bone), it turns out that other bones, including teeth, often have better DNA preservation. But an important paper in 2014 showed that a dense part of the **petrous bone** in the inner ear has extremely high human DNA content compared to all other tested bone sites [535]. This may reflect both better preservation of human DNA in the petrous, as well as better exclusion of microbes. Since then the field has shifted entirely to sequencing from petrous bone when that is available.

Sediment DNA. Skeletal remains of ancient hominins are extremely rare and hard to find. But an exciting alternative is the possibility of retrieving DNA from the places where they lived. When early peoples were living, eating, defecating, and dying in caves, they often left traces of their DNA in the sediment of the cave. While bones are probably necessary for obtaining high quality genomes, mitochondrial genomes and low-coverage nuclear data can often be recovered directly from sediment. Recent work

with **sediment sequencing** has provided astonishing time series of archaic habitation of caves, including Denisova Cave and elsewhere, at a resolution that has not been possible using aDNA from bones [536] [e].

Lastly, there are intriguing advances in **ancient protein sequencing.** Many important skeletal remains no longer have surviving DNA. However, proteins can persist much longer, and there is recent work using mass spectrometry to obtain protein sequences from proteins present in teeth or bones. While this approach is limited to the most abundant structural proteins, and therefore carries much less information than would be available for genome sequencing, it has the potential to offer exciting insights, as in protein sequencing from a Denisovan jaw (see below) and protein sequencing from 2 MY-old Paranthropus remains [537].

[e] *For an excellent short video on sediment sequencing in Denisova Cave see Link.*

Figure 3.79: **Artifacts from Denisova Cave:** *pendants and notched bones. Denisova Cave has become the most important site in aDNA studies of archaic hominins, with DNA and physical evidence for both Neanderthal and Denisovan occupation at multiple time points from* 50–200 *KYA. An individual dated to* 90 *KYA is a Neanderthal-Denisovan hybrid, indicating that the populations overlapped, at least at that time.* *Credit: Thilo Parg [Link]. CC BY-SA 4.0.*

Population genetics of archaic hominins. We can now pick up the main story again. As you recall, the mitochondrial DNA studies of Neanderthals had provided a first look at genetic variation in archaic hominins, but it was clear that nuclear data would provide a much more comprehensive picture.

After their false start with contamination four years earlier, in 2010 the Pääbo team succeeded in generating an initial 1x Neanderthal "draft" genome, using the samples from Vindija, Croatia [538]. This was followed in 2014 by a high-coverage (30x) Neanderthal genome from the Altai Mountains of southern Siberia (referred to as the Altai Neanderthal), and in 2017 by a high-coverage genome from Vindija [539]. Additional low coverage Neanderthal data are available from various other sites.

The Denisovans. Meanwhile, the Pääbo lab continued to screen bones for suitable Neanderthal material for sequencing. One important source of ancient bone fragments was a cave in the Altai Mountains in Siberia, named Denisova Cave after a 17th Century local hermit called Denis. This cave was the source for the first high coverage Neanderthal sequence. But one tiny fragment turned out to be something truly remarkable and unexpected...

A tiny piece of bone from the fifth (pinky) finger of a hominin showed indications of good DNA preservation. Upon mitochondrial sequencing it turned out that the mtDNA was unlike anything that had ever been seen before: it was neither human nor Neanderthal [540].

The figure below shows that, as expected, all humans, and all Neanderthals sequenced up to then fall into distinct clades, with a common ancestor between the two groups at around 400 KYA. But mtDNA from the new fingerbone, shown at the bottom in red, was completely different from both, branching off at around 1.0 MYA [541]:

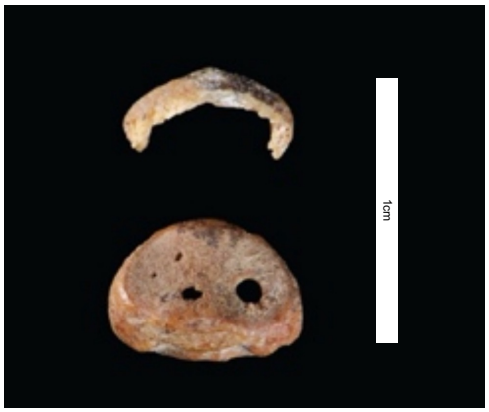


Figure 3.80: **Punching above its weight.** *The Denisovans were discovered from this tiny fragment of pinky bone, shown here from two angles. The bone dates to ∼70 KYA.* *Credit: Modified from Fig S12a of Katerina Douka et al (2019) [Link]. Used with permission.*

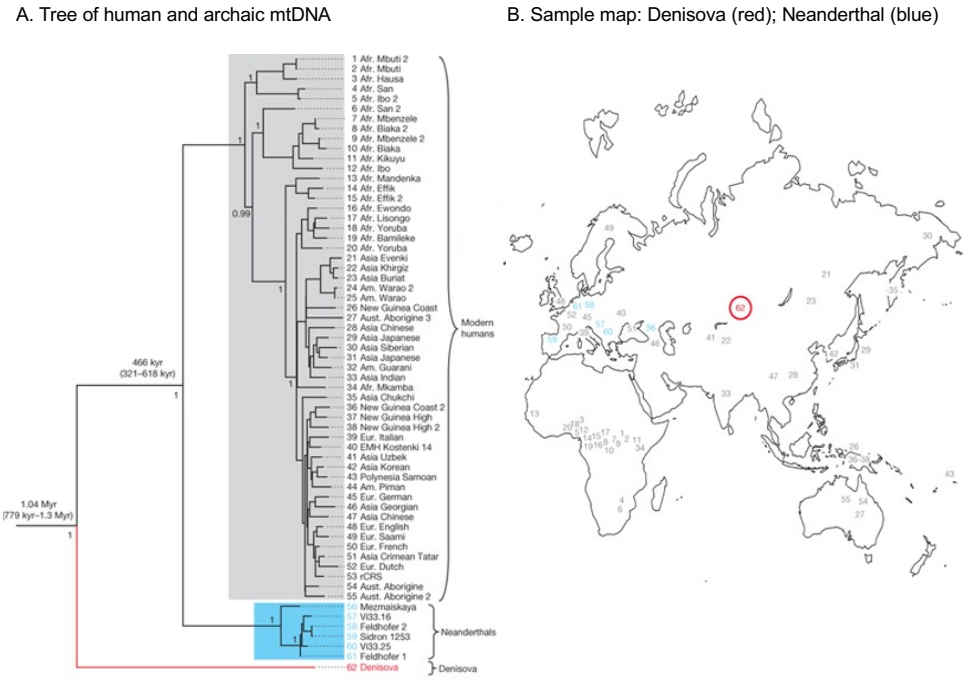


Figure 3.81: **Divergent mtDNA from Denisova. A.** *MtDNA from the new finger bone lay completely outside the tree of known variation from Neanderthals and diverse humans (lineage in red at bottom).* **B.** *Map of samples. Denisova cave in the Altai Mountains of Siberia is in red. Since then, Neanderthal sampling has extended as far east as Denisova.* *Credit: Modified from Fig 3 of Johannes Krause et al (2010) [Link]. Creative Commons.*

At first this new hominin was completely enigmatic, known entirely from this one mtDNA sequence from one bone fragment. During the next few years, the Pääbo group obtained first a low-coverage (2x) genome, in 2010, and then a high-quality (30x) genome, in 2014 [542]. These sequences showed that the new bone fragments came from a new, unknown hominin, related to Neanderthals.

This group is now known as the **Denisovans**, and it has become clear that they actually had a broad distribution across much of Asia. The team that reported on the Denisovan genome decided not to designate it as a separate species given the uncertainty about its relation to hominin fossils from elsewhere, and evidence that it had admixed with modern humans [543]. This is the first and, so far, only new hominin identified entirely through genomic approaches.

Denisovans are a sister group to Neanderthals. At first, in view of the mtDNA results, it seemed possible that the Denisovan pinky might represent something even more ancient than Neanderthals (*Homo erectus*?). But whole genome data show that the Denisovans and Neanderthals are sister groups.

Specifically, these data suggest that the ancestors of Neanderthals and Denisovans emerged from Africa ~600 KYA, before diverging within Eurasia ~400 KYA to form the distinct populations of Neanderthals, in western Eurasia, and Denisovan lineages in eastern Eurasia, respectively [544] [545] [546].

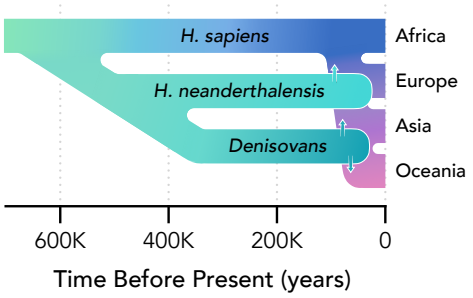


Figure 3.82: **Simplified model of Human-Neanderthal-Denisovan origins.** *See Figure 3.48 for a more detailed overview.* *Credit: Kindly contributed by Alyssa Lyn Fortier, unpublished CC BY 4*

Archaic introgression. Prior to the first aDNA studies, it had been known for decades that Neanderthals likely overlapped with anatomically modern humans in Europe and the Middle East. **This raised the question of whether they interbred, and if so, do we still carry Neanderthal DNA in our genomes today?**

As we discussed above, Neanderthal mtDNA lies entirely outside the distribution of human mitochondrial variation; this seemed to argue against the possibility of introgression. But there was no way to know if that would be true for the rest of the genome. So in 2010, having collected the first draft version of a Neanderthal genome, the Pääbo team revisited the question of Neanderthal-human introgression.

Before I get to the results, I'll describe how they tackled the question.

Optional Box: D statistics for testing introgression. *This box is more specialized and not required– but you'll find it's a nice review of coalescent techniques for structured populations that you learned in Chapter 2.4), as well as the admixture methods from Chapter 3.2.*

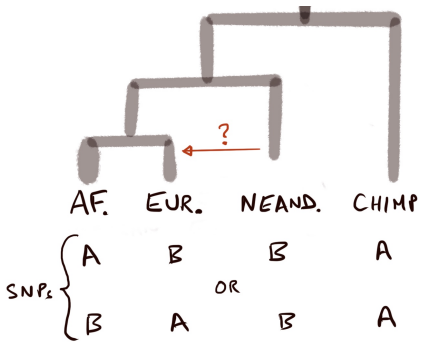
The first Neanderthal draft genome presented the following analysis challenge: Suppose that we have a single genome sequence from a particular population, how can we tell if that population contributed genetic material to other related populations? One particular challenge is that we would want any test to be robust to complexities of demographic history – for example the relationships among populations as well as population bottlenecks, expansions, etc.

In 2010, Nick Patterson, David Reich, Monty Slatkin, and colleagues developed an elegant test for this question that is robust to most complications of demography [547]. This test produces a measure known as a **D statistic** or **ABBA-BABA statistic**; it can also be viewed as a type of **F test** from Chapter 3.2, but using a single haploid sample from each population and motivated differently.

The setup is as follows. We start by choosing one genome from a modern African and one from a modern European. (In practice, since most genome data are diploid, at heterozygous sites we simply pick one of the alleles at random [548].) We also include a Neanderthal and a chimpanzee genome. We then choose sites that meet the following criteria:

• The allele in the African genome differs from the corresponding site in the European genome.

• Of those two alleles, one matches the chimpanzee (we label this allele *A*) and the other matches Neanderthal (we label this *B*). This gives us two possible configurations of alleles:



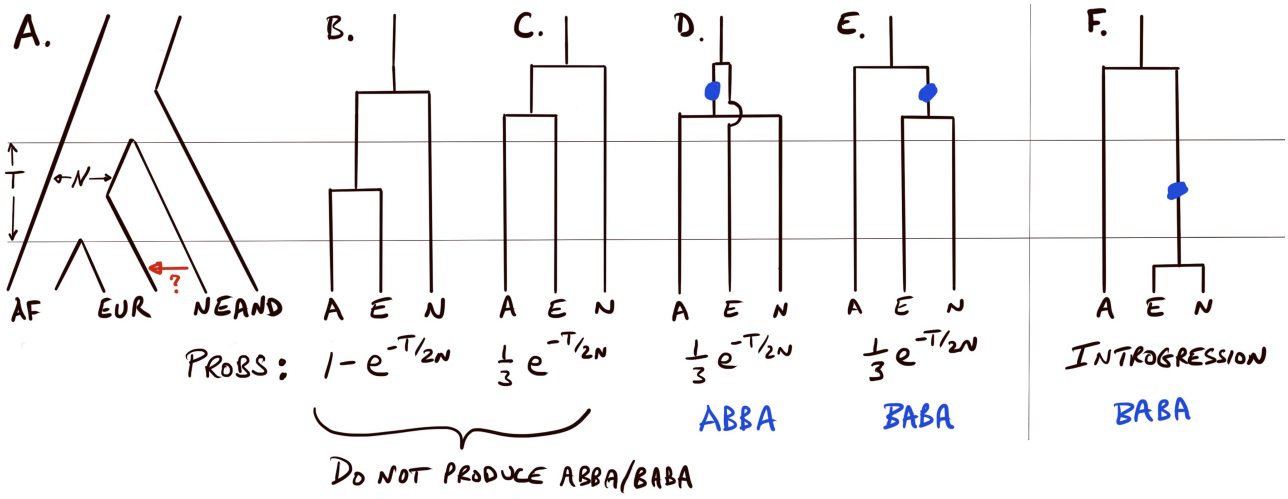
Box Figure: The ABBA-BABA Test. *This test considers SNPs where a single African and a single European genome differ, and where one matches Neanderthal while the other matches an outgroup (chimpanzee). The chimp allele is labeled 'A'. This gives two possible configurations of alleles, labeled ABBA and BABA, depending on who matches Neanderthal.*

Next, we count the number of sites, genome-wide, that fall into either ABBA or BABA configurations. Denoting these numbers as $\#ABBA$ and $\#BABA$, respectively, we define the D statistic as

$$D = \frac{\#ABBA - \#BABA}{\#ABBA + \#BABA} \tag{3.1}$$

Under the null hypothesis of no introgression, there is a symmetric relationship between Africans and Europeans with respect to Neanderthals, so the African and the European genomes should match Neanderthal at equal rates ($D \approx 0$). This prediction does not depend on the details of demographic history within each population [549]. But if there *has* been introgression into Europeans, then Europeans should match Neanderthals at a higher rate ($D > 0$).

To see why this works, and why it doesn't depend on demography, it's helpful to draw some pictures of the coalescent genealogies that produce the data. We saw this kind of approach previously when we discussed the human-chimpanzee-gorilla split in Chapter 2.4:



Box Figure: Intuition for the D statistic. A. *Population model for Africans, Europeans, and Neanderthals.* **B.–E.** *Four possible coalescent topologies, assuming no introgression, with the corresponding probabilities shown below. Only D and E produce ABBA or BABA configurations of SNPs (here the blue dot indicates the position of the mutation that produces allele B).* **F.** *shows how an introgression event from Neanderthals into Europeans can produce additional BABA sites.*

The figure above shows the probabilities of different topologies, assuming an effective population size N_e in the human ancestral population. But the key point here is that only (D) and (E) can produce ABBA or BABA configurations; by symmetry these occur at equal rates in the absence of introgression, regardless of the demographic details [550].

To summarize: Under a null hypothesis of no gene flow between Neanderthals and Europeans, we expect D = 0; if there *has* been gene flow we expect D > 0.

One last technical question is how to test for significance. We can't treat all SNPs as independent, be-

cause SNPs that are close in the genome are correlated due to LD. A standard solution is to compute standard errors using a technique known as the **block jackknife**, which treats megabase blocks of the genome as independent estimators of D [551].

In 2010, this test produced a startling result: that Neanderthals had contributed genetic material to modern humans! Here's a table of the D statistics from that first paper:

Population 1	Population 2	D×100
Yoruba	**French**	**4.6**
Yoruba	**Han**	**5.3**
Yoruba	**Papuan**	**4.2**
Yoruba	San	−0.1
French	Han	0.9
French	Papuan	0.0

Table 3.1: D statistics for different population pairs with Neanderthal. *The comparisons of Africans and non-Africans are highly significant (the standard errors are ≈0.5 for all tests). The other comparisons (between Africans, or between non-Africans) are not statistically significant.* *Data from Table S48 of Green et al (2010).*

As you can see, the test shows positive values of D for several non-African populations (these are all highly significantly >0) [f].

[f] *Now that the fact of archaic introgression is well-known, it's perhaps difficult to appreciate how extraordinarily impactful this result was in 2010 – truly reshaping people's understanding of human genetic variation.*

Follow-up work from many groups estimates that **all non-African populations have around 1.5–2% Neanderthal ancestry** [552]. In contrast, Neanderthal ancestry in sub-Saharan African populations is very low, and is likely due to gene flow from Eurasia back into Africa in the last few thousand years [553].

Where did the main introgression occur? One surprising feature immediately jumps out from these analyses: Prior to the 2010 paper, we would have expected that if Neanderthal introgression had occurred, it would have been mainly into Europeans and other western Eurasians. *If Neanderthals were mainly in northwest Eurasia, then why would they have contributed to all non-Africans in very similar amounts?*

We can explain this if *Neanderthals contributed genetic material to the founder population that gave rise to all non-Africans*, before it started to spread globally. At the time of writing, we can't tell from the currently available genetic data where this happened: Genetics can be very informative about the relationships between populations, but given the mobility of early humans, the genetic data are not very useful at telling us where those ancestors lived.

However, *the most natural hypothesis is that the admixture occurred in the Middle East*. We know that both Neanderthals and early modern humans were present in this region in about the right time-frame [554], and it makes geographic sense as a source location for all non-Africans [g]. We can hope that, in future, ancient DNA from the right time and place may resolve this question.

[g] *For a map showing where this might have occurred see Figure 3.59.*

The next question to ask is: *When did admixture occur?*

Dating Neanderthals, and human admixture. Recall from Chapter 3.2 that

after two populations mix, this creates a type of long range LD, known as **admixture LD**. Over time, recombination breaks the admixed chromosomes into smaller and smaller pieces. We can predict that the average size of an introgressed block, t generations after admixture, is $100/t$ cM [h].

[h] *Remember that introgressed blocks get whittled down over time by recombination. We can use this logic to date admixture events: see Figures 3.25 and 3.27.*

In a 2016 paper, Priya Moorjani and colleagues implemented a similar method (modified for use with archaic genomes) to estimate the timing of human-Neanderthal admixture [555]. They measured the decay of admixture LD both in present-day humans, as well as in ancient DNA from human specimens at a range of different dates.

One example from their paper is shown below. The decay curve for modern humans (CEU) indicates an admixture time of about 47 KYA. The decay curve for a 24,000-year old specimen from Siberia (Mal'ta) is much flatter – expected as it is much closer to the original admixture event. Adding together the age of Mal'ta, with the inferred time between Mal'ta and Neanderthal admixture (22,000 years) gives an estimate very similar to the date from CEU.

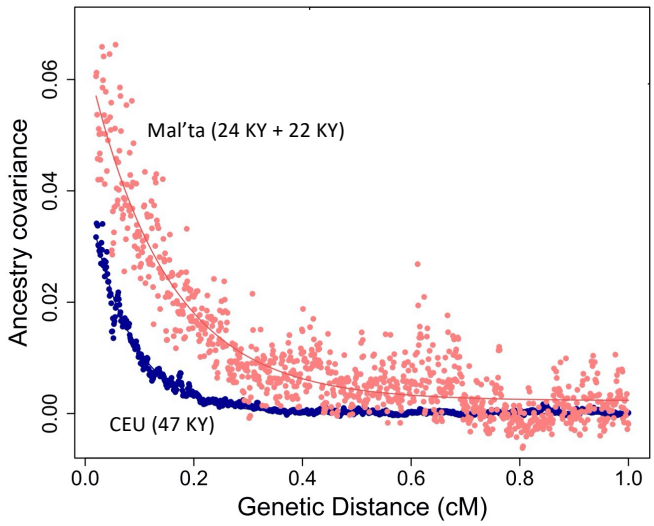


Figure 3.83: **Decay of Neanderthal admixture LD.** *The plot shows decay of admixture LD in modern Europeans (***CEU***) and in an early human specimen named* **Mal'ta***, from south-central Siberia, carbon dated to* 24 *KYA. The authors estimated an admixture time of* 47 *KYA using a present-day CEU genome, and* 22 *KYA for the Mal'ta genome (consistent with the CEU result, given Mal'ta's age).* *Credit: Modified Figure 1b of Priya Moorjani et al (2016) [Link]. Reuse permitted per PNAS license.*

Together the analyses in this paper indicate that Neanderthal introgression into anatomically modern humans occurred around 45–50 KYA.

This date has a very important implication. Given that all non-African populations alive today inherited Neanderthal ancestry from this introgression event, it implies that the main out-of-Africa diaspora cannot have started spreading globally until after this date. We know from fossil evidence that humans had reached Eurasia long before this [556], but the admixture data imply that the earlier human expansions must have been largely or completely replaced by this expansion after 50 KYA.

Direct observations of Neanderthal hybrids! Most of what we know about ancient admixture comes from statistical signals of events tens of thousands of years ago. But two remarkable fossils provide direct evidence of admixture.

The first is a human mandible from Romania known as Oase 1 from **Romania, dated to around 40 KYA**. This individual has around 6–9%

Neanderthal ancestry – far more than modern humans. Most striking is that much of the Neanderthal ancestry is contained within 3 blocks of more than 50 centiMorgans each. This implies **a recent Neanderthal ancestor, within the previous 4–6 generations** [557].

Even more astonishing is a remarkable bone fragment from Denisova Cave. Denisova Cave was at the eastern edge of the Neanderthal range, and western edge of the Denisovan range, and remains of both species have been found in the cave, ranging over tens of thousands of years. But for a long time it wasn't known if the two hominins were actually there at the same time.

This question was resolved in 2018 by the lucky find of a female bone fragment in **Denisova Cave, from around 90 KYA** [558]. Remarkably, this girl, nicknamed **"Denny", had a Neanderthal mum and a Denisovan dad**. We usually expect F1 hybrids between divergent populations to be rare, and yet this hybrid was found within the first few Denisovan genomes ever sequenced. It's unclear whether this was just an extraordinarily lucky find, or hybrids were perhaps more common than expected. Species hybrids often have reduced fertility, but we have no way of knowing in this particular case if she would have been fully fertile.

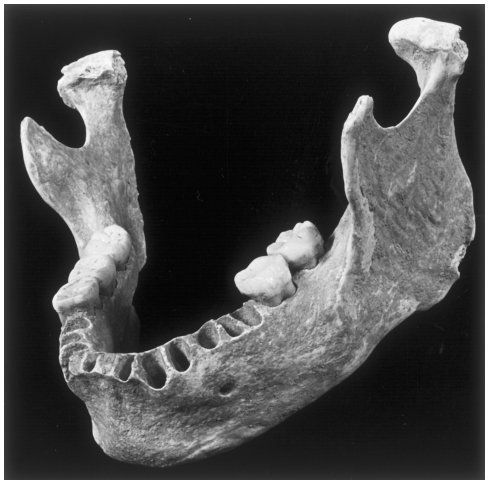


Figure 3.84: **The Oase 1 human mandible** *from Romania, dated to ∼40 KYA, had a recent Neanderthal ancestor.* *Credit: Eric Trinkaus et al (2003) [Link]. Reuse permitted.*

Denisovan introgression. Given the evidence for gene flow from Neanderthals into modern humans, it was immediately interesting to ask whether Denisovans might also have contributed genes to humans.

For Neanderthal introgression, the major surprise had been that it was fairly uniform across all non-Africans [559]. In contrast, Denisovan introgression was geographically concentrated, but in a part of the world that seemed at first to make no sense [560]:

	Neanderthal (%)	Denisovan (%)
Papuan	2.4	**5.0**
Bougainvillean	2.5	**4.3**
Tibetan	1.9	**0.5**
Mongol	2.4	**0.4**
Han	2.5	**0.4**
Karitiana	2.0	**0.4**
French	1.9	0.1
Palestinian	1.7	0.1
Tajikistan	1.9	0.1

Table 3.2: Archaic ancestry proportions, sorted by Denisovan, *for representative non-African populations. While Neanderthal introgression is relatively uniform, Denisovan introgression is heavily concentrated in Oceania and to a lesser extent in East Asia.* *Papuans and Bougainvilleans are both from islands in Papua New Guinea. Tibetan, Mongol and Han are east Asians; Karitiana are from the Amazon but carry east Asian ancestry from the original Beringian migration. The Denisovan estimates from French/Palestinian/Tajikistan are not significant. Estimates based on ratios of F statistics. Data from Supp. Table 1 of Mallick et al (2016) [Link].*

As illustrated in the table above, **Denisovan ancestry is found at high levels in Oceania (Papua New Guinea and nearby islands)** and, **to a lesser extent, in east Asia** [561]. Denisovan ancestry has also been carried by these populations into Pacific Islanders and native Americans, respectively.

So why would a hominin known only from the Altai Mountains of southern Siberia have introgressed mainly into Oceanian populations? First it

suggests that Denisovans must have had a broad geographic range despite going undetected until 2010.

But where did the ancestors of Oceanians encounter Denisovans? Did they meet Denisovans in central or southern Asia during the original eastward expansion? Or did the Denisovans themselves reach Oceania?

To tackle this question, several groups have developed methods to identify Denisovan ancestry blocks in modern populations [562].

In the plot below, Guy Jacobs and colleagues identified Denisovan-like blocks in Papuans, East Asians, and Siberians. They characterized each block in terms of its length (which tells us how long ago it passed into humans) and its divergence from the Denisovan genome sequence from Siberia.

This work reveals a fascinating complexity of Denisovan admixture: there were at least three major admixture events from three deeply diverged Denisovan lineages. One admixture event is found in east Asians, and is closely related to the Siberian Denisovan. Two other admixture events are only found in Papuans, and indicate the existence of two additional unknown Denisovan lineages. The authors estimate that these were deeply diverged from the Siberian Denisovans, by around 300 KY!

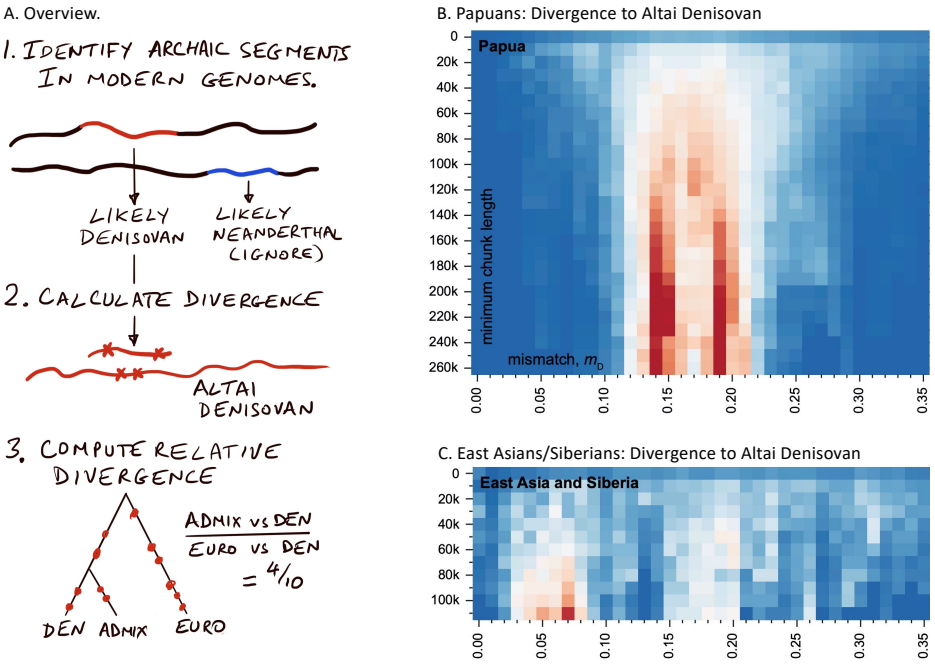


Figure 3.85: **Divergence of introgressed chunks from Altai Denisovan.** *In each plot the x-axis shows the divergence from the Altai Denisovan for Denisovan introgressed chunks (the x-axis is scaled relative to human-Denisovan divergence). The y-axis corresponds to admixed segments of at least the indicated size (there are fewer large chunks in East Asians). The color scale indicates the most frequent divergence levels within each row: red is highest; white is intermediate. The divergence estimates are more noisy for small chunks, which is why the signals are clearest at large sizes.* *Credit: From Figure 3 of Guy Jacobs et al (2019) [Link]. Used with permission.*

This discovery of deeply diverged Denisovan lineages who interbred with the ancestors of Oceanians – but with no other known human populations – strongly suggests that these Denisovan lineages had, themselves, reached Oceania.

In summary, the results suggest that **there were at least three divergent lineages of Denisovans: one in Central/East Asia**, represented by the Siberian genome sequence, and **two in Island Southeast Asia** that ad-

mixed with the ancestors of modern Papuans and Melanesians.

We do not yet have physical evidence of Denisovans in Southeast Asia, but we do know that other archaic humans, *floresiensis* and *luzonensis*, had successfully reached the islands of Indonesia and the Philippines.

Figure 3.86: **Baishiya Karst Cave**. *This Tibetan Buddhist sanctuary, at 3200m elevation, was home to Denisovans until ~60,000 years ago.* *Credit: Dongju Zhang. Linkhttps://doi.org/10.1126/science.abb6320 CC BY-SA 4.0*

The hunt for physical remains of Denisovans. At this point in the story (around the year 2020), the data presented a bit of a conundrum. Ancient Denisovans were known from just a single cave in Siberia, but the analysis of Denisovan introgression in modern humans shows they must have been widespread.

Yet the physical remains in Denisova Cave are very limited – a finger, some bone fragments, some teeth, and so we had no sense what Denisovans looked like. This left key questions open: Do we have direct physical evidence for Denisovans anywhere else outside of Denisova Cave? And without knowing what they look like: How would we know?

The next chapter in this story comes from a cave at the edge of the Tibetan Plateau, in Gansu, China. In 1980, a Tibetan Buddhist monk, using the cave as a santuary, came across a jawbone with very unusual morphology. The bone was passed to Lanzhou University where it sat in a box for many years. Nearly thirty years later it caught the attention of two scientists working at Lanzhou, Fahu Chen and Dongju Zhang [563].

Radiometric dating indicated that the jaw, now known as the "Xiahe Mandible", was around 160,000 years old – implying that it must come from an archaic hominin. The authors attempted DNA sequencing, but that failed.

Figure 3.87: **The Xiahe Mandible from the Tibetan Plateau**. *Remarkably, protein sequencing suggests a Denisovan origin – providing the first physical remains of Denisovans outside of Denisova Cave in Siberia.* *Credit: Dongju Zhang. [Link] CC BY-SA 4.0*

But they did find evidence for degraded ancient proteins. Applying **mass spectrometry**, they were able to obtain peptide sequences from six collagen genes (collagen is the main structural protein in bone). This technique is extremely limited compared to DNA sequencing, but they identified a single amino variant at which the Xiahe Mandible carries a derived allele matching the Denisovan genome, while the available Neanderthal sequences and modern genomes all carry the ancestral allele [564]. Given the likelihood of shared polymorphisms between Neanderthals and Denisovans, this is not definitive proof that the mandible is Denisovan – but is at least suggestive [565].

While the mandible did not yield any surviving DNA, a follow-up study searched for DNA using **sediment sequencing** of layers within the cave [566]. A team of Chinese archaeologists was granted access to work in the cave, but only at night and in the winter to avoid disturbing the monks. They collected sediment samples from eight archaeological layers of different depths and ages. DNA sequencing revealed extensive mammalian diversity in the cave, including extinct rhinoceros and hyena. They also identified hominin DNA in the cave.

After excluding likely contamination from modern humans, they found a clear signal of Denisovan mitochondrial DNA in two different layers

dating to 100 KYA and 60 KYA, respectively. In the tree below, you can see that the mtDNA from these two Baishiya layers cluster with previous Denisovan sequences:

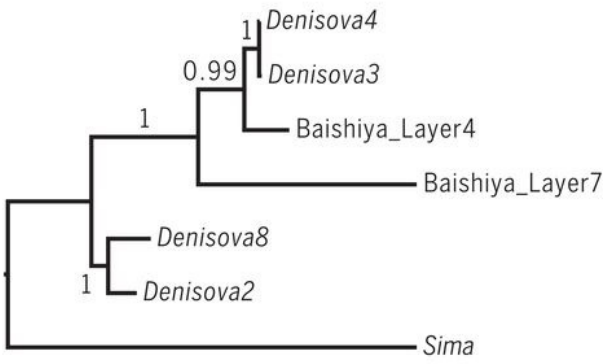


Figure 3.88: **Baishiya mtDNA clusters with Denisovans**. *Phylogenetic tree showing the close relationship between mtDNA sequences retrieved from sediments in the cave with Denisovan mtDNA from Denisova Cave. Sima (an early Neanderthal ancestor with Denisova-like mtDNA) is used as an outgroup.* *Credit: From Figure 4c of Dongju Zhang et al (2020). [Link]. Used with permission.*

The protein data, combined with definitive evidence for Denisovan presence in the cave, strongly suggest that the Xiahe mandible is the first identified Denisovan fossil outside of Denisova Cave in Siberia.

Next, in 2021, a massive ancient skull with an estimated age of at least 150 KY was reported from Harbin in northeast China. Known as the **Harbin Cranium** and nicknamed **Dragon Man**, it has a large cranial capacity and a mixture of archaic and modern features. Importantly, it shares features with other Chinese fossils from around the same period and, most notably, it potentially matches aspects of the Xiahe mandible [567]. Thus *it is tempting to suggest that all these east Asian remains are physical manifestations of Denisovans* [568]. You can see a map showing the locations of putative Denisovan remains in Figure 3.54.

In summary, although Denisovans were originally detected from a finger bone fragment in Siberia, this finger represents a diverse clade of hominins with a huge range stretching from Siberia to Tibet, to northeast China, and south into Island Southeast Asia.

Figure 3.89: **The Harbin Cranium**, *around 150 KY old from northeast China, is one of several Chinese fossils tentatively assigned to the Denisovan lineage.* *Credit: From Figure 2a of Xijun Ni et al (2020). [Link] CC BY-NC-ND 4.0*

Population genetics of Neanderthals and Denisovans. Alongside these discoveries about the Neanderthal and Denisovan populations and human admixture, the archaic genomes revealed another surprise.

With the first archaic genomes in hand, it was natural to look at basic properties of genetic variation, including heterozygosity, i.e., the fraction of sites that are heterozygous in a single genome.

Recall that in your genome, you have a heterozygous site roughly once every 1–2 Kb, depending on your ancestry – or equivalently, about 5–10 heterozygous sites per 10kb. Analysis of Neanderthals and Denisovans showed that they are much less variable, with only 2 heterozygous sites per 10 Kb:

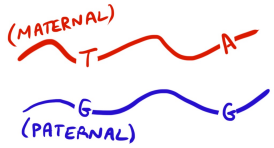


Figure 3.90: *Recall that* ***heterozygosity*** *measures the fraction of sites that differ between the corresponding positions in two haploid genomes. Under an ideal model the expected heterozygosity is* $H = 4N_e\mu$. *See Chapters 1.3 and 2.1.*

Population	Heterozygous SNPs per 10 Kb
San	10.5
Yoruba	10.1
French	7.7
Han	7.4
Papuan	6.3
Karitiana	5.8
Neanderthal	**2.1**
Denisovan	**1.9**

Table 3.3: Low heterozygosity in archaic humans. *The table shows the average density of heterozygous sites in representative genomes.* *Estimates exclude long homozygous runs due to inbreeding. Karitiana are a small indigenous population in the western Amazon. Data from Table S9.1 of Kay Prüfer et al (2014). [Link].*

In other words, genetic diversity of archaic humans was only about one third as much as in the least variable modern humans! In fact, despite their huge geographic ranges, the genetic diversity of the Neanderthals and Denisovans was lower than almost any other known species, and is similar to very scarce mammals such as wolverine and lynx [569].

Moreover, PSMC analysis [i] shows that Neanderthals and Denisovans experienced low effective population sizes for hundreds of thousands of years prior to their extinction:

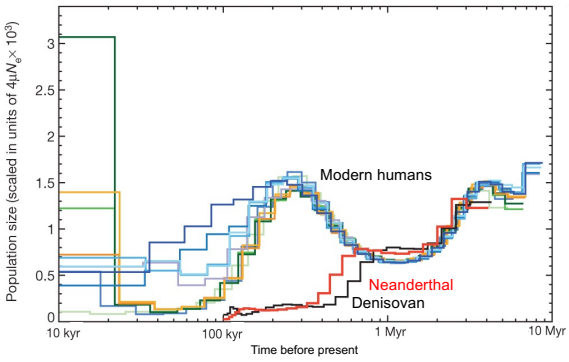


Figure 3.91: **PSMC for archaic and modern populations** *indicates long-term reduced population size for Neanderthals and Denisovans compared to 11 diverse modern populations. The Neanderthal and Denisovan lines start at the estimated sample ages.* *The timescale here shows years before present, and was modified to reflect modern mutation rate estimates. Credit: Modified Fig. 4 from Kay Prüfer et al (2014). [Link]. Used with permission.*

[i] *Recall that PSMC estimates population sizes in the past using distributions of heterozygous sites in a single genome (Chapter 3.3).*

In summary, Neanderthals and Denisovans had tiny effective populations sizes for extremely long time periods. Give their huge geographic ranges, this implies that they must have had very low population densities – perhaps living in small, scattered groups. These low population sizes may help to explain why these groups ultimately went extinct in the face of ecological competition or conflict with modern humans.

Natural selection for and against archaic DNA. So far we've focused on genome-wide average patterns of Neanderthal/Denisovan admixture into modern humans. This is appropriate for asking questions about demography and population history.

But do we see functional differences between human versus archaic admixture blocks? Does selection act on archaic alleles [j]? The answer is a resounding yes!

The first key result on this came in work from 2014 and 2016 by Sriram

[j] *As we shall see, the major signal of selection is **against** archaic DNA, albeit with limited examples where particular archaic alleles have been positively selected.*

Sankararaman and colleagues, showing a striking result: namely that **Neanderthal and Denisovan ancestry is systematically depleted near genes** and other functional elements [570] [571].

To show this, they used the B statistic, which measures the local density of functional DNA elements around any given position in the genome [572] [k]. Regions with more functional DNA are defined as having low B.

[k] *We encountered B previously when discussing background selection in Chapter 2.7; see Figure 2.123.*

In the plot below, the genome has been divided into 5 bins of B. As you can see, the average fraction of Neanderthal ancestry is *much* lower in Bin 1 (highest density of functional sequences) than in Bin 5 (lowest density):

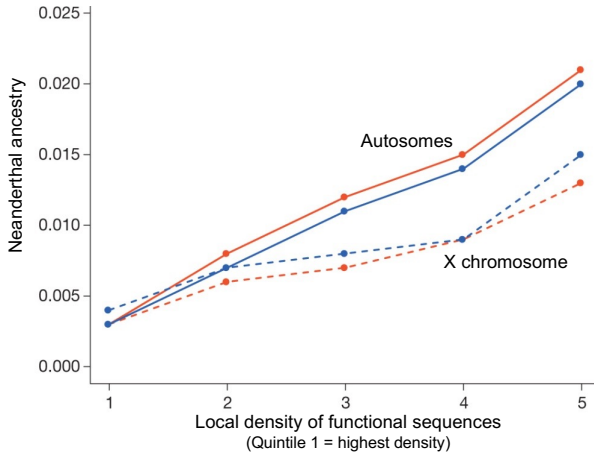


Figure 3.92: **Neanderthal ancestry is depleted from functional regions.** *The plot shows average Neanderthal ancestry of Europeans (red) and East Asians (blue) as a function of McVicker's B.* *Credit: Modified from Fig 2 of Sriram Sankararaman et al (2014) [Link]. Used with permission.*

This is in sharp contrast to predictions under a neutral model, since we would expect Neanderthal ancestry to be uniformly distributed across the genome, without regard to functional elements. Instead, if we think of regions with low functional density (Bin 5) as approximating a neutral baseline, this implies that **selection must have purged archaic ancestry from functional regions.**

Why did this purging occur? There are two main models [l]:

[l] *It's possible/likely that both models contribute, but we don't have a good handle yet on the relative importance of each.*

Model 1: Neanderthals and Denisovans were starting to speciate from humans. When populations are isolated from each other for long time periods they start to accumulate genetic incompatibilities. Over time, the accumulation of incompatibilities steadily reduces the fitness of hybrids. Eventually, there are so many incompatibilities that hybrids are either inviable or infertile, at which point we consider the populations to have formed distinct species. It's likely that humans and Neanderthals/Denisovans were in the early stages of this process [573].

During these early stages of speciation, hybrids are still viable and fertile, but usually with reduced fitness. This in turn drives selection against the specific alleles that cause incompatibility. While we do not know yet which genes or regulatory elements evolved incompatibilities, we can predict that on average they are more likely to be in genomic regions with higher density of functional elements. Hence, this model predicts that in a population that has majority human ancestry, there would tend

to be selection against Neanderthal DNA, and that this selection would be strongest on average in regions with low B [574].

Model 2: Neanderthals and Denisovans had a high load of deleterious alleles due to their small N_e. This second model proposes that, instead, **Neanderthals and Denisovans may have had an excess of deleterious alleles and that, after admixture, selection acted to remove these** [575].

Recall that Neanderthals and Denisovans both had extraordinarily low effective population sizes for an extended period of time. And remember from Chapter 2.5 that efficacy of selection is greatly reduced with small N_e. Hence, we can expect that the archaic genomes would have accumulated an excess of deleterious mutations (also referred to as **genetic load**) relative to human genomes [576]. For an alternative version of this model based on *stabilizing selection*, see: [577].

Following admixture with modern humans, blocks of archaic DNA would be more likely to carry deleterious variants compared to human haplotypes in the same region. Hence, the archaic blocks would tend to be purged by natural selection. As in the genetic incompatibilities model, this effect would be strongest in regions with high density of functional elements and/or low recombination rates.

At this point we don't yet have a good way of distinguishing between these two models in data. Based on modeling I think that Model 2 (genetic load) almost certainly contributes a good fraction of the effect; I'd bet that Model 1 (incompatibilities) probably contributes too, but at the time of writing the magnitude of this effect remains unclear.

Lastly, you may wonder if the same processes are relevant in mixture events between modern human populations. The short answer seems to be *No*. There is no evidence for DMIs between any human populations, likely due to the much shorter separation times and continual gene flow between human populations [578]. As for the load model: while non-African populations are significantly bottlenecked, the bottlenecks were short enough that there are not meaningful differences in load between human populations [579].

So, did we get anything good from the archaics? Was archaic admixture just a genetic burden – or did we get anything useful? Intuitively, it makes sense to expect that the archaics might have had useful adaptations to the climate and pathogens of Eurasia – after all, their ancestors had spent several hundred thousand years there when the first modern humans arrived.

In other species there are many examples of a process called **adaptive introgression**, in which favorable variants or haplotypes pass between closely-related species in rare hybridization events. These alleles can then spread to high frequency in the recipient population, often enabling adaptation to some new environment [580]

As for whether this happens in humans, I first need to remind you about **altitude adaptation in Tibetans** (Figure 2.67 in Chapter 2.4). In 2010

Figure 3.93: **Hybrid incompatibilities drive selection in contemporary swordtail fish**, *in a way that is very similar to Model 1. In hybrid populations of these two species, selection acts against genetic material from whichever species contributed less to the initial mixed population.*
Credit: Fig. 1A from Molly Shumer et al (2014). [Link]. CC BY 3.0.

Figure 3.94: **Adaptive introgression in snowshoe hares.** *This picture shows the white winter coat that is typical of snowshoe hares. In regions with less winter snow cover, some snowshoes have brown winter coats instead – this is due to an allele that introgressed from a related species, jackrabbits.*
Credit: Jacob W Frank [Link]. Public Domain.

three groups independently identified a region of the genome near **EPAS1** (also known as HIF2α). SNPs in this region have huge allele frequency differences (F_{ST}) between Tibetans and their close relatives, Han Chinese [581]. The EPAS1 gene encodes a transcription factor that stimulates red blood cell production and regulates hemoglobin levels in the blood, and the Tibetan haplotype is strongly associated with changes in hemoglobin levels and better performance at high altitude [582].

At first it was natural to assume that EPAS1 was undergoing a conventional selective sweep.

But a 2014 paper by Emilia Huerta-Sánchez and colleagues pointed out that something even more interesting and unusual is happening there. They pointed out that **the selected haplotype is unusually diverged from other haplotypes observed in Tibetans or Han**, with 32 highly differentiated SNPs within a 32.7 kb region [583]. This would be highly unlikely under a standard sweep model, because the sweeping haplotype is so highly diverged from other haplotypes in the population. So where did the selected haplotype come from?

The key to answering this came from a comparison to archaic genomes. While the Tibetan haplotype is distantly related to other human haplotypes, it is nearly identical to the published Denisovan genome:

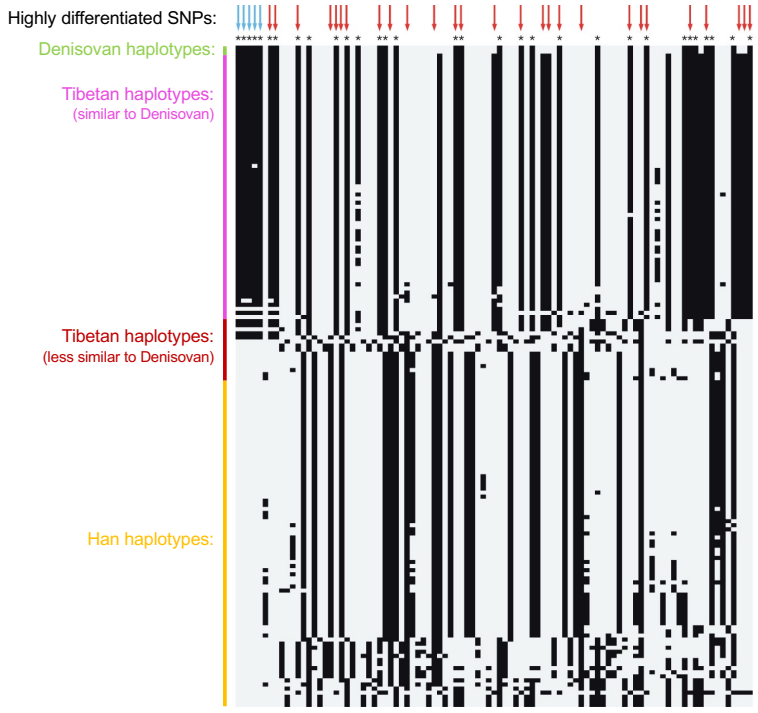


Figure 3.95: **The Tibetan EPAS1 haplotype matches the Denisovan genome.** *The plot shows haplotype variation in a 32.7 Kb region at EPAS1. Haplotypes are shown as rows and SNPs as columns with derived alleles in black. The blue arrows at the top indicate a block of five SNPs that are identical between Denisovans and nearly all copies of the advantageous haplotype in Tibetans.* *Credit: Modified from Fig 2 of Emilia Huerta-Sánchez et al (2014). [Link]. Used with permission.*

So with these observations, the EPAS1 story became even more interesting. One of the key adaptations of high-altitude Tibetans had actually introgressed from Denisovans. Back in 2014 this was particularly enigmatic because Denisovan remains had only been detected in Denisova Cave in Siberia, though of course we do now know that Denisovans were actually in Tibet. So we can speculate that the altitude-adapted EPAS1 al-

lele originated in Tibetan Denisovans. That said, there are still some gaps in this story: Why did the Siberian Denisovans carry the high-altitude allele? And how did this allele pass into modern Tibetans, as this population has probably not been in Tibet long enough to have received the allele directly from Denisovans [584].

EPAS1 remains our strongest example of adaptive introgression from archaic hominins. Other archaic haplotypes that have increased in frequencies may play roles in processes such as immunity, pathogen interactions, pigmentation and metabolism [585]. However, genome-wide, the overall importance of archaic introgression in driving human phenotypic variation was probably fairly modest, especially given that most archaic functional variation has been selected against [586].

So far, we've been focusing on the use of aDNA to study archaic hominins and their relations to Homo sapiens. In the last section of this chapter we turn our attention to the use of ancient DNA to study our own prehistory.

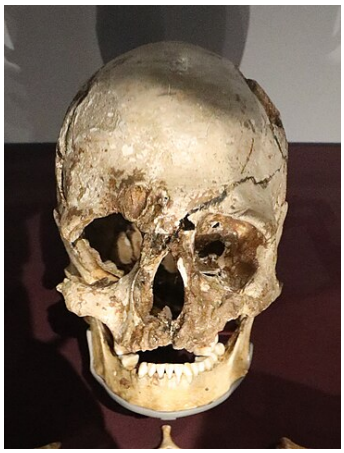


Figure 3.96: **Cheddar Man**, *10.3 KYA from Cheddar, England. In 1996, a TV program claimed to have found a local school teacher who was a descendant of Cheddar Man(!) – while such a direct and specific link was always implausible, we now know that in fact the British population of that period was almost completely replaced by subsequent waves of immigration.* *Cheddar Man was part of the Western Hunter Gatherer population that we'll meet shortly. Credit: user:geni [Link] CC BY-NC-SA 4.0*

aDNA of anatomically modern humans. Before the widespread use of aDNA, it was tempting to imagine that population histories were simple. For example, modern humans (*Homo sapiens*) first migrated into Europe around 50,000 years ago; can we assume those people were ancestors of modern Europeans? More generally, it was easy to suppose that the people living in any given place today would be mainly descended from people who lived there, or lived nearby, in the past.

But as tempting as it may be to assume such simple models, we now know from aDNA that simple models assuming continuous populations from past to present are almost invariably wrong.

Since around 2012, ancient DNA has now completely rewritten our understanding of human prehistory [587]. Thousands of ancient human genomes have now been sequenced, from all the inhabited continents.

While the field is far too large to cover systematically, we'll illustrate the power of aDNA with a look at what we've learned so far in Europe. Europe has a rich trove of archaeological remains, and much of the continent has a temperate climate, which is good for DNA preservation. We now have a dense sampling of European aDNA, and a good understanding of the main storylines [588].

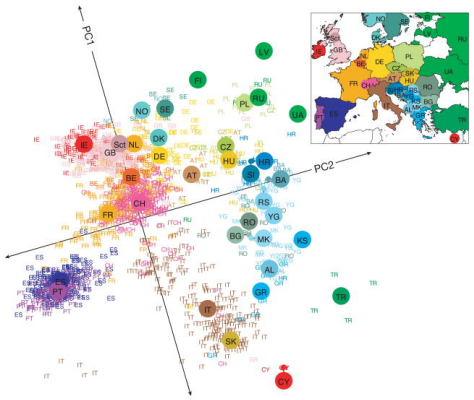


Figure 3.97: **Genes mirror geography within Europe (2008).** *PCA projection of present-day Europeans. After slightly rotating the PCA axes, the PCA projection reflects the geographic positions of countries to a remarkable degree, with only minor distortions.* *Credit: Figure 1A from John Novembre et al (2008). [Link] Used with permission.*

Case study: the genetic history of Europe. In Chapter 3.1 we discussed the population structure of present-day Europe. As illustrated in a classic plot (Figure 3.17; reproduced here in the margin), there is a striking correspondence between genetic variation and geography. What processes generated this pattern?

Two models and a clue. Prior to the aDNA era, there were two main hypotheses to explain the European PCA.

The first model, developed by the 20th Century geneticist Luca Cavalli-Sforza, proposed that the key event was **the Neolithic transition from hunter-gathering to farming in Europe**. Archaeological studies had shown that the invention of farming spread slowly across Europe, starting around 9,000 years ago in the Middle East, and reaching Northern Europe around 5,500 years ago.

This raised the question: *How* did farming spread across Europe? One possibility is that populations *learned how to farm from their neighbors*, in which case the culture could have spread without the movement of people. But instead Cavalli-Sforza proposed a model called *demic diffusion, in which the farmers themselves expanded slowly across Europe*. As they expanded (in Cavalli's model) they mated with local hunter gatherer populations, thereby diluting the farmer ancestry as they spread across Europe. Cavalli argued that the southeast-to-northwest direction of Europe's leading PC reflects this ancient spread of farmer ancestry [m].

[m] *Debate about the spread of farming is a classic example of a **pots versus people** debate in archaeology. When a new technology (often a new style of pottery) is found to have spread over time, did it do so through spread of culture or through spread of the people who used that technology? aDNA can answer such questions.*

The second model was that the European PCA might instead reflect **a steady-state population with genetic drift and local migration**. In this case, local migration would result in continuous spatial structure that could look very much like the patterns seen in Europe [589].

One early observation turned out to be an important clue:

The clue: Ötzi the Iceman. In September 1991, two hikers in the Ötzal Alps on the border of Italy and Austria came across the body of a man, partially covered by snow and ice. At first they assumed that he was a modern hiker who had died in the mountains, but it quickly became clear that he was actually an ancient frozen mummy – now dated to around 5,300 years ago (after the arrival of farming).

Figure 3.98: **Ötzi the Iceman.** *This 5,300 year-old body was found partially buried in snow and ice at the border of Italy and Austria.*
Credit: Figure 1 of Andreas Keller et al, 2012 [Link] CC BY-NC-SA 3.0

Examination of the remains showed that Ötzi (as he was nicknamed) had been walking in the mountains with a copper axe, a bow and arrows, berries, and other possessions. He had been struck by an arrow in one shoulder, and this was the likely cause of death. His last meals included meat from ibex and deer, as well as einkorn wheat, possibly in bread.

In 2012 his genome was sequenced, and compared to modern Europeans [590]. At this point, the researchers found something completely unexpected.

One might naively have expected him to cluster near other central Europeans. But as you can see below, he actually clusters away from most modern Europeans: in fact, **he is most similar to modern Sardinians! Why?** [n]

[n] *The island of **Sardinia** is located in the central Mediterranean Sea, about 1000 km south of where Ötzi was found.*

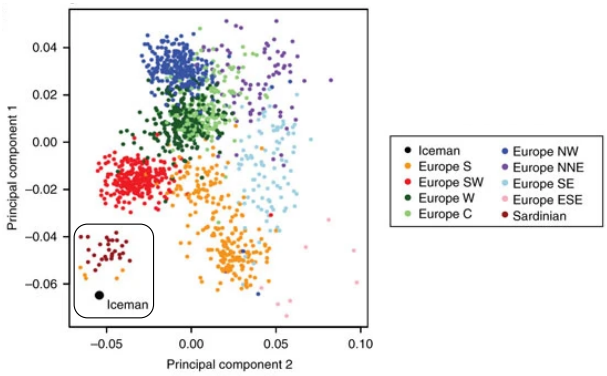


Figure 3.99: **Ötzi's genome is most similar to modern Sardinians.** *The plot shows a PCA of modern Europeans and Ötzi (each dot is one individual). The region in the lower-left of the plot contains individuals from Sardinia from two different data sets.* *Credit: Modified Figure 3f of Andreas Keller et al, 2012 [Link] CC BY-NC-SA 3.0*

The answer to this puzzle points to one of the most important events shaping the modern genetic structure of Europe. We'll get to this shortly.

(And for now, I'll just give you the hint that it's NOT because Ötzi was actually a Sardinian on an Alpine ski holiday gone horribly wrong!)

The first humans in Europe: complex structure, mixtures, and extinctions. The first main period of human occupation in Europe is known as the **Upper Paleolithic**, ranging from ∼54 KYA to ∼14 KYA. The later part of this time range, starting ∼33 KYA, is known as the Last Glacial Period, when much of northern Europe was covered in ice.

The people who lived in Europe at this time were a complicated – and genetically heterogeneous – set of groups known collectively as the **Cro-Magnon** people, including the cave artists featured in Figure 3.100) [591].

The first physical remains of Cro-Magnon, based on stones and teeth, date to around 54,000 years ago [592]. But remember that this is before the critical date of the Neanderthal introgression event that all non-Africans share (at around 47 KYA). How should we understand this?

We still have quite limited aDNA sampling in this time period, but recall the 40,000 year-old **Oase 1** mandible from Romania, with a recent Neanderthal ancestor (Figure 3.84). Very surprisingly, analysis of the Oase 1 genome showed that *it has no particular genetic affinity with modern Europeans* [593]. In fact, Oase 1 is slightly more closely related to East Asians and Native Americans than to Europeans!

Together, the analysis of early Cro-Magnon Europeans shows that they have little or no genetic connection to modern populations. Apparently these early *Homo sapiens* forays into Europe, spread over more than 15 KY, were ultimately unsuccessful – likely failing to survive or being actively replaced by groups that arrived later. This lack of genetic continuity also helps to explain why Europeans don't have *more* Neanderthal ancestry than other non-Africans, even though we know there were human–Neanderthal matings in Europe.

It's not until about 37 KYA that European samples first start to show significant genetic links to modern Europeans. However, these genetic links

Figure 3.100: **The Chauvet horses, ∼36 KYA.** *The walls of the Chauvet-Pont d'Arc Cave in France show paintings of large mammals including horses, lions, hyenas, bears, deer and woolly rhinos, as well as human handprints, rendered in charcoal and ochre.* *Credit: Thomas T [Link] CC BY-SA 3.0.*

This remarkable art dates to around the time of the earliest Europeans with a detectable genetic link to present-day Europeans.

to modern populations are still highly indirect, because even these later Paleolithic populations would ultimately be replaced.

A collision of 3 diverged groups forms the genetics of modern Europeans. We pick up our story about 20 KYA, during the last Ice Age. As the great European ice sheet of the Ice Age recedes northward, a group of Mesolithic hunter gatherers, known as **Western Hunter Gatherers (WHG)**, starts to spread outward from southeast Europe.

By 14 KYA the WHG have spread across most of Europe, ranging from modern-day Ireland in the west, to the edges of Central Asia in the East, and essentially replacing the earlier Cro-Magnon populations. *The WHG is the first major ancestry group that contributes directly to modern Europeans.*

But the WHG were soon to be outcompeted by a population that brought with them one of humankind's most important inventions: farming [594]. Around 9 KYA, a group known as the **Neolithic Anatolian Farmers** started to spread west out of modern-day Turkey into Europe and north Africa. They brought animals including pigs, sheep, goats, and cattle, and food plants including lentils, barley and einkorn wheat. *During the next 2000 years, the Anatolian Farmers swept across Europe.* Instead of mixing with the local WHG population, as Cavalli-Sforza had hypothesized, this was a complete replacement across most of the continent.

The last major reshaping of Europe came from the east. The **Yamnaya** population were sheep and cattle herders living in the **Russian Steppes**. Starting around 5 KYA, they acquired horses from the east, and learned to make wagons. This suddenly gave them new mobility for herding, and the ability to carry supplies across long distances. Now, they too began to sweep into Europe. Over the next thousand years, they spread west and south to contribute genes and culture throughout the region. Unlike the earlier Farmer expansion, the Steppe expansion did not replace the existing population, but often mixed with them as they spread. *By 4 KYA, the Steppe expansion had completely reshaped the genetics of Europe.*

Figure 3.101: **Corded Ware pottery.** *This style was developed by the Yamnaya in the Russian Steppes and brought to Europe in the Steppe Expansion. Dated* 4.8–4.3 *KYA, Germany.*
Credit: Einsamer Schütze [Link] CC BY-SA 3.0

These replacements are illustrated below. Clemens Weiß used the clustering program ADMIXTURE to model the three main ancestry components in around 3,500 ancient and modern samples: WHG, Anatolian Farmers, Steppe/Yamnaya, plus a fourth Iranian component that is relevant in southeast Europe. The maps below show dramatically how the average ancestry components changed over time and space [595]:

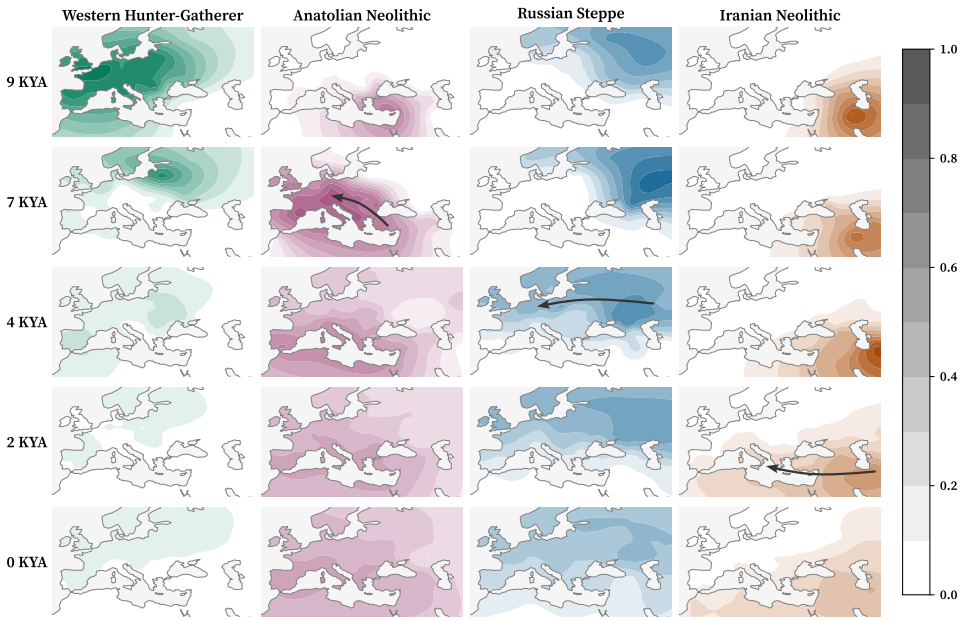


Figure 3.102: **Ancestry changes in Europe across the past 10 KY.** *Each map shows contours of ancestry fractions for a particular ancestral component (columns) in a particular time-slice (rows). The numerical values of the contours are as indicated in the grey-scale key at right. The arrows indicate major population expansions.* *These maps were estimated by running ADMIXTURE on aDNA data from the indicated region, and fitting the proportions in time and space using a 3-D Gaussian Process model. Credit: GP development, data analysis, and unpublished figure kindly contributed by Clemens Weiß; CC BY 4.*

The plots show the key population movements clearly:
∼**9 KYA:** WHG ancestry dominates most of Europe;
∼**7 KYA:** The Anatolian Farmers sweep in from modern-day Turkey, largely replacing WHG;
∼**4 KYA:** The Russian Steppe expansion spreads into Europe, *mixing* with the existing populations;
∼**4–0 KYA:** European population structure generally stable [596]; minor influx of Middle Eastern ancestry into Southern Europe.

We can now understand why Ötzi is most like modern Sardinians: Living in central Europe before the arrival of Steppe ancestry, he had mainly Anatolian Farmer ancestry. As you can see in the bottom row of maps, Sardinia is the one of the only parts of present-day Europe that was not reached by the Steppe influx. For this reason, the central Europeans of 5 KYA are most similar to modern Sardinians, of all modern-day populations [o].

The making of European population structure. We're now ready to answer the question of what produced the classic PCA plot of European genetics.

A version of the PCA is reproduced below, using an expanded set of populations. The European populations are contained in the upper cluster, forming a somewhat distorted and rotated map of Europe; the lower cluster includes adjacent populations from North Africa, the Middle East, and West Asia.

Five sets of ancient samples are projected in color into the same PCA space. As you can see, all five of these ancient populations lie outside the scale of modern population variation. We can understand why this is – the colored plots at the right show how modern Eurasians are genetic mixtures of these earlier groups:

[o] *Here's another interesting consequence of the Steppe influx: I noted in Chapter 3.2 that there is* ***a genetic link between modern Europeans and Native Americans****. This is because both derive ancestry from the Ancient North Eurasians. The ANE ancestry was first carried into Europe by the Steppe incursion.*

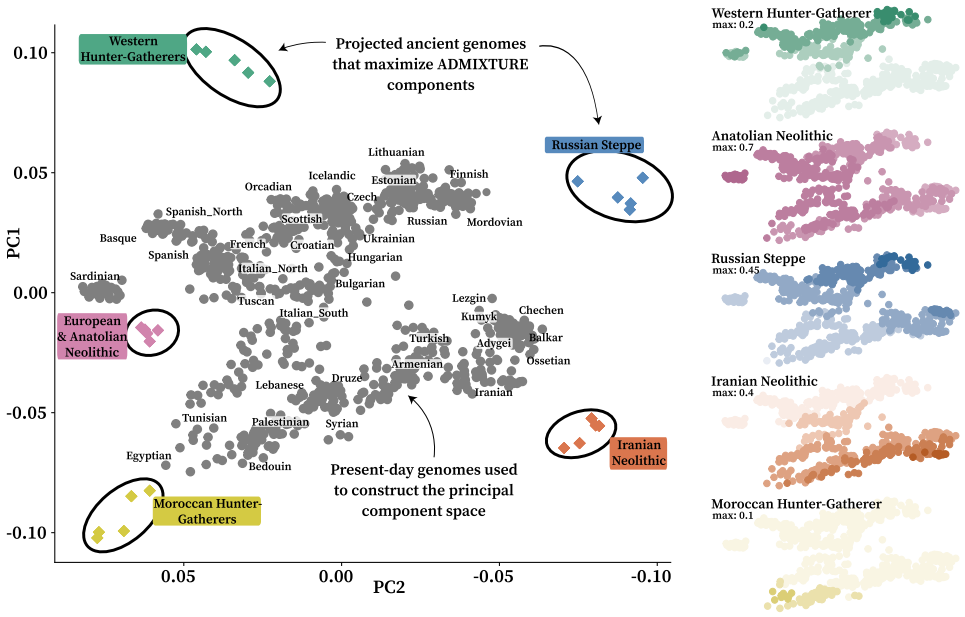


Figure 3.103: **Modern Eurasians are mixtures of ancestral populations.**
Left. *PCA of modern Eurasian individuals (in grey, with population labels). A handful of ancient samples, chosen to represent the main ancestry components, are projected into the same space (in color).*
Right. *The modern samples in the same projection were fit with 5 ancient ancestry components, as listed. The colors here show relative ancestry fractions for each individual: e.g., individuals in dark green (Finnish) have the highest WHG ancestry (max 0.2).*
Credit: Credit: Unpublished data analysis and figure kindly contributed by Clemens Weiß and Margaret Antonio; CC BY 4.

Thus, the classic PCA of Europe results from multiple admixture events, mixing together three main ancestral groups – WHG, Anatolian Farmer, and Steppe – plus other smaller components. The reality is much more complicated than anticipated by either of the two main models that I described at the start of this section [597].

Perhaps most striking, we can now understand **modern Eurasian populations as relatively recent genetic mixtures of highly divergent ancestral groups that no longer exist in their original forms.**

Summary. As we have discussed across the last four chapters, methods from population genetics have dramatically rewritten our understanding of human populations and human history. Notably, in the last chapter, I described how the tools of population genetics allow us to dig into our deep history, hundreds of thousands of years ago, complementing the classical tools of archaeology and paleontology. In this chapter I discussed how ancient DNA has further shaken up our understanding of human origins, especially within the past 50 KYA.

If there is an over-arching lesson from aDNA it is that simple models of human history are often wrong, and in entirely unexpected ways. The populations that live in a particular place now very likely have no genetic connection to the people who lived in that place in the past; and their ancestors often lived far away in unexpected locations.

Open questions: While we have already learned much about human history from genetics, I think we can expect progress on a number of fronts in the upcoming years:
• Greater resolution of our deep origins in the last 1–2 MY. Is the standard model (Figure 3.48) largely correct, or will this undergo revisions? For example, did Eurasian *Homo erectus* contribute in any meaningful way to the main stem of human evolution?

• How far can we push the boundaries on sampling very old ancient DNA, and aDNA from challenging environments with heat and humidity. Nontraditional approaches including sediment sequencing and perhaps protein sequencing may be very valuable here. It would be of immense value to achieve deeper sampling of African *Homo sapiens*, as well as additional late-stage archaics such as *Homo floresiensis* from outside the Neanderthal/Denisovan branch.
• How exactly are modern non-Africans related to each other, to earlier *Homo sapiens* outside Africa; when did their ancestors separate from the complex population structure apparent inside Africa?
• aDNA at more recent timescales – especially within the past 10 KY – will continue to merge with archaeology and history, and continue to rewrite how we understand our recent history
• What can aDNA teach us about human phenotypic evolution? What traits, and what genes, were the main drivers of adaptive change in different *Homo* lineages?

Well done! You've now finished Part 3 of the book. In the next section we turn our attention to understanding the genetic basis of human trait variation.

Notes and References.

[517]For reviews of ancient DNA see e.g.,

Nielsen R, Akey JM, Jakobsson M, Pritchard JK, Tishkoff S, Willerslev E. Tracing the peopling of the world through genomics. Nature. 2017;541(7637):302-10

Reich D. Who we are and how we got here: Ancient DNA and the new science of the human past. Oxford University Press; 2018

Skoglund P, Mathieson I. Ancient genomics of modern humans: the first decade. Annual review of genomics and human genetics. 2018;19(1):381-404

Bergström A, Stringer C, Hajdinjak M, Scerri EM, Skoglund P. Origins of modern human ancestry. Nature. 2021;590(7845):229-37

[518]Svante Pääbo was awarded the 2022 Nobel Prize for the pioneering work of his team on ancient DNA of Neanderthals and Denisovans.

[519]Higuchi R, Bowman B, Freiberger M, Ryder OA, Wilson AC. DNA sequences from the quagga, an extinct member of the horse family. Nature. 1984;312(5991):282-4

Thomas RH, Schaffner W, Wilson AC, Pääbo S. DNA phylogeny of the extinct marsupial wolf. Nature. 1989;340(6233):465-7

Cooper A, Mourer-Chauviré C, Chambers GK, von Haeseler A, Wilson AC, Pääbo S. Independent origins of New Zealand moas and kiwis. Proceedings of the National Academy of Sciences. 1992;89(18):8741-4

[520]Cano RJ, Poinar HN, Pieniazek NJ, Acra A, Poinar Jr GO. Amplification and sequencing of DNA from a 120–135-million-year-old weevil. Nature. 1993;363(6429):536-8

Woodward SR, Weyand NJ, Bunnell M. DNA sequence from Cretaceous period bone fragments. Science. 1994;266(5188):1229-32

[521]Krings M, Stone A, Schmitz RW, Krainitzki H, Stoneking M, Pääbo S. Neandertal DNA sequences and the origin of modern humans. cell. 1997;90(1):19-30

[522]The **biological species concept** defines populations as distinct species if they cannot produce fertile offspring. However, this definition is not always easily applied in practice as reproductive barriers evolve gradually, and these barriers are often not absolute even between groups that are considered well-defined, distinct species (hence the concept of introgression). As we shall see, humans and Neanderthal interbred many times, in multiple locations, but the hybrids likely had reduced fertility. Thus we can view these two groups as being in the earliest stages of speciation.

[523]See discussion of the human mitochondrial MRCA in the previous chapter.

[524]Magnus Nordborg (1998) made the point that mtDNA is not sufficient to resolve the question even under a simple neutral model [REF]. Moreover we now know that mtDNA variation is often atypical of the rest of the genome: for example with both a more-recent human MRCA, and human-Neanderthal MRCA, and a more ancient human-Denisovan MRCA than most of the genome.

Nordborg M. On the probability of Neanderthal ancestry. The American Journal of Human Genetics. 1998;63(4):1237-40

[525]Plagnol and Wall provided a careful analysis arguing for an archaic contribution in both Europeans and Africans; this work was somewhat overlooked at the time but the overall conclusions were probably correct.

Plagnol V, Wall JD. Possible ancestral structure in human populations. PLoS genetics. 2006;2(7):e105

[526]Green RE, Krause J, Ptak SE, Briggs AW, Ronan MT, Simons JF, et al. Analysis of one million base pairs of Neanderthal DNA. Nature. 2006;444(7117):330-6

[527]Noonan JP, Coop G, Kudaravalli S, Smith D, Krause J, Alessi J, et al. Sequencing and analysis of Neanderthal genomic DNA. science. 2006;314(5802):1113-8

[528]These observations were first made by Sridhar Kudaravalli from my lab and used for internal discussions among the groups. A careful analysis of this was published by Jeffrey Wall and Sung Kim (2007). It's unclear why the two studies differed so dramatically: this may have been due to technical differences in protocols used by the two studies, or perhaps to chance differences between the bone fragments used.

Wall JD, Kim SK. Inconsistencies in Neanderthal genomic DNA sequences. PLoS Genetics. 2007;3(10):e175

[529]Green RE, Krause J, Briggs AW, Maricic T, Stenzel U, Kircher M, et al. A draft sequence of the Neandertal genome. Science. 2010;328(5979):710-22

Prüfer K, Racimo F, Patterson N, Jay F, Sankararaman S, Sawyer S, et al. The complete genome sequence of a Neanderthal from the Altai Mountains. Nature. 2014;505(7481):43-9

[530]Fu Q, Hajdinjak M, Moldovan OT, Constantin S, Mallick S, Skoglund P, et al. An early modern human from Ro-

mania with a recent Neanderthal ancestor. Nature. 2015;524(7564):216-9

[531]In addition to fragmentation, ancient DNA also accumulates lesions that (depending on the precise lesion) can cause errors during sequencing, or block the DNA polymerases used for sequencing. The most frequent change is **deamination of cytosines leading to C→T errors**. Cytosine deamination changes the nucleotide to uracil; during sequencing uracil is interpreted as thymine. To address this issue, many labs now use a technique known as UDG treatment which performs enzymatic repair of most uracil sites. As for conventional genome sequencing, apparent mutations that are only supported by one or very few reads are considered unreliable. For analysis of low-coverage sequencing, the methods focus entirely on variation at known, common SNPs.

Dabney J, Meyer M, Pääbo S. Ancient DNA damage. Cold Spring Harbor perspectives in biology. 2013;5(7):a012567

Briggs AW, Stenzel U, Meyer M, Krause J, Kircher M, Pääbo S. Removal of deaminated cytosines and detection of in vivo methylation in ancient DNA. Nucleic acids research. 2010;38(6):e87-7

Rohland N, Harney E, Mallick S, Nordenfelt S, Reich D. Partial uracil–DNA–glycosylase treatment for screening of ancient DNA. Philosophical Transactions of the Royal Society B: Biological Sciences. 2015;370(1660):20130624

[532]Meyer M, Arsuaga JL, De Filippo C, Nagel S, Aximu-Petri A, Nickel B, et al. Nuclear DNA sequences from the Middle Pleistocene Sima de los Huesos hominins. Nature. 2016;531(7595):504-7

[533]Van der Valk T, Pečnerová P, Díez-del Molino D, Bergström A, Oppenheimer J, Hartmann S, et al. Million-year-old DNA sheds light on the genomic history of mammoths. Nature. 2021;591(7849):265-9

[534]To tackle this problem, some aDNA labs use capture methods with DNA or RNA "baits" designed to enrich for target DNA – for example to capture fragments that overlap common SNPs in the human genome. However, the capture kits themselves can be expensive, and most labs simply apply the brute-force approach of random shotgun sequencing, followed by computational identification of human DNA reads

[535]Gamba C, Jones ER, Teasdale MD, McLaughlin RL, Gonzalez-Fortes G, Mattiangeli V, et al. Genome flux and stasis in a five millennium transect of European prehistory. Nature communications. 2014;5(1):5257

Pinhasi R, Fernandes D, Sirak K, Novak M, Connell S, Alpaslan-Roodenberg S, et al. Optimal ancient DNA yields from the inner ear part of the human petrous bone. PloS one. 2015;10(6):e0129102

[536]Vernot B, Zavala EI, Gómez-Olivencia A, Jacobs Z, Slon V, Mafessoni F, et al. Unearthing Neanderthal population history using nuclear and mitochondrial DNA from cave sediments. Science. 2021;372(6542):eabf1667

Zavala EI, Jacobs Z, Vernot B, Shunkov MV, Kozlikin MB, Derevianko AP, et al. Pleistocene sediment DNA reveals hominin and faunal turnovers at Denisova Cave. Nature. 2021;595(7867):399-403

[537]Madupe PP, Koenig C, Patramanis I, Rüther PL, Hlazo N, Mackie M, et al. Enamel proteins reveal biological sex and genetic variability within southern African Paranthropus. bioRxiv. 2023:2023-07

[538]Green et al (2010), cited above.

[539]Prufer et al (2014), cited above; and also:

Prüfer K, De Filippo C, Grote S, Mafessoni F, Korlević P, Hajdinjak M, et al. A high-coverage Neandertal genome from Vindija Cave in Croatia. Science. 2017;358(6363):655-8

[540]Krause J, Fu Q, Good JM, Viola B, Shunkov MV, Derevianko AP, et al. The complete mitochondrial DNA genome of an unknown hominin from southern Siberia. Nature. 2010;464(7290):894-7

[541]Note that the mtDNA time estimates were unaffected by the 2012 revision of the mutation rate in the nuclear genome. A recent paper estimated the divergence between human and Neanderthal mtDNA at 360–468KYA, and the divergence Denisova and human/Neanderthal mtDNA at 720–1410KYA, roughly similar to the estimates in the Krause 2010 paper.

Posth C, Wißing C, Kitagawa K, Pagani L, Van Holstein L, Racimo F, et al. Deeply divergent archaic mitochondrial genome provides lower time boundary for African gene flow into Neanderthals. Nature communications. 2017;8(1):16046

[542]Reich D, Green RE, Kircher M, Krause J, Patterson N, Durand EY, et al. Genetic history of an archaic hominin group from Denisova Cave in Siberia. Nature. 2010;468(7327):1053-60

Meyer M, Kircher M, Gansauge MT, Li H, Racimo F, Mallick S, et al. A high-coverage genome sequence from an archaic Denisovan individual. Science. 2012;338(6104):222-6

[543]By tradition, Neanderthals have their own species name, *Homo neanderthalensis*, but Denisovans do not. As I noted above, it's debatable whether these groups should be considered distinct species as there was extensive gene flow between humans and both Neanderthals and Denisovans.

[544]It is estimated that about 1% of the Denisovan genome descends from an unknown "ghost" population that has a deeper split from the human branch, probably at about 1 MYA or more

Hubisz MJ, Williams AL, Siepel A. Mapping gene flow between ancient hominins through demography-aware in-

ference of the ancestral recombination graph. PLoS genetics. 2020;16(8):e1008895.

[545]It's interesting that the mitochondrial tree is markedly different from the overall population tree estimated from nuclear genomes. The Neanderthal mtDNA shares a common ancestor with humans at about 400 KYA – more recent than the main population split. In contrast the Denisovan mtDNA is more ancient, at around 1MYA, perhaps deriving from the unknown ghost population. It's not entirely clear why the mitochondrial tree should be so unusual, although in other systems mtDNA variants frequently play key roles in speciation, which may lead to different patterns of sharing and divergence than for random parts of the genome.

[546]Estimates of key timepoints can be found int the following sources. Prüfer et al (2017): The two highest-quality Neanderthal genomes, Altai and Vindija, split at 130–145 KYA; Vindija Neanderthal and Denisovan split at 390–440 KYA; Vindija Neanderthal and Humans split at 520–630 KYA. Posth et al (2017) estimate human-Neanderthal mtDNA divergence at 360–468 KYA and human-Denisovan mtDNA divergence at 720–1,410 KYA. Note that dating in the Meyer et al (2016) paper about the Sima de los Huesos samples implies an older timeline than the Prüfer et al estimates: they date the Sima samples at 430 KYA, and place them on the Neanderthal lineage. The Denisovan ghost lineage is dated to $>$ 1MY by Hubisz et al (2020). Jacobs et al (2019) estimate two splits between Altai and Southern Denisovans at 283 and 363 KYA (cited below).

[547]Green et al (2010), cited above;
Durand EY, Patterson N, Reich D, Slatkin M. Testing for ancient admixture between closely related populations. Molecular biology and evolution. 2011;28(8):2239-52

[548]This kind of procedure works in analyses like this where SNPs are treated independently. It wouldn't be appropriate for any analysis related to LD.

[549]This test does rely on the assumption that African and European lineages coalesce with Neanderthal lineages at equal rates when the human-Neanderthal populations merge together at $\sim$ 600 KYA. This is a good assumption given that modern structure only goes back at most around $\sim$ 300 KY. For a rigorous treatment see Durand et al (2011).

[550]As discussed in more detail by Durand et al, the main thing that can blow up this assumption is if there is structure in the ancestral population in such a way that ancestors of one of the modern populations (e.g., the ancestors of Europeans) coalesce more quickly with Neanderthals than ancestors of the other population (e.g., Africans). This is not a concern in humans as the human-Neanderthal split is more ancient than the deep structure of extant humans, but it can certainly be a caveat in other analyses.

[551]In genomics the main issue is that many features of the genome, such as LD, are correlated at length-scales of 10s to 100s of KB or more. A common approach to dealing with this is to divide the genome into megabase-scale blocks (i.e., larger than the extent of LD) and then treat the estimates from different blocks as independent observations. It's then possible to compute standard errors using the jacknife. You can read a basic description of the jackknife in the Wikipedia page: [Link]. The jackknife is conceptually related to another resampling technique, the bootstrap; the bootstrap is more widely used in many statistical applications, but use of the jackknife has drifted to high frequency in recent population genetics applications.

[552]Prufer et al (2017).

[553]Chen L, Wolf AB, Fu W, Li L, Akey JM. Identifying and interpreting apparent Neanderthal ancestry in African individuals. Cell. 2020;180(4):677-87

[554]Hershkovitz I, Marder O, Ayalon A, Bar-Matthews M, Yasur G, Boaretto E, et al. Levantine cranium from Manot Cave (Israel) foreshadows the first European modern humans. Nature. 2015;520(7546):216-9

[555]Moorjani P, Sankararaman S, Fu Q, Przeworski M, Patterson N, Reich D. A genetic method for dating ancient genomes provides a direct estimate of human generation interval in the last 45,000 years. Proceedings of the National Academy of Sciences. 2016;113(20):5652-7
Iasi LN, Chintalapati M, Skov L, Mesa AB, Hajdinjak M, Peter BM, et al. Neandertal ancestry through time: Insights from genomes of ancient and present-day humans. bioRxiv. 2024

[556]See Chapter 3.3; broadly consistent with the fossils there's also genetic evidence for earlier human to Neanderthal gene flow
Li L, Comi TJ, Bierman RF, Akey JM. Recurrent gene flow between Neanderthals and modern humans over the past 200,000 years. Science. 2024;385(6705):eadi1768.

[557]Paper about the fossil: Trinkaus E, Moldovan O, Milota Ş, Bîlgăr A, Sarcina L, Athreya S, et al. An early modern human from the Peştera cu Oase, Romania. Proceedings of the National Academy of Sciences. 2003;100(20):11231-6
aDNA analysis: Fu et al (2015), cited above.

[558]Slon V, Mafessoni F, Vernot B, De Filippo C, Grote S, Viola B, et al. The genome of the offspring of a Neanderthal mother and a Denisovan father. Nature. 2018;561(7721):113-6

[559]There's some debate about whether the admixture proportions are precisely the same in all non-Africans, or merely very similar. For example, most estimates of Neanderthal introgression show Europeans with about 20% less Neanderthal ancestry than some other groups such as Han. This may be due to European admixture with a group that lacked Neanderthal introgression (perhaps from north Africa, or an early out-of-Africa population), or potentially estimation bias caused by back-migration from Europe into Africa.

[560]This was first shown in the 2010 Denisovan genome paper by Reich et al.

[561]Jacobs GS, Hudjashov G, Saag L, Kusuma P, Darusallam CC, Lawson DJ, et al. Multiple deeply divergent Denisovan ancestries in Papuans. Cell. 2019;177(4):1010-21

[562]Prufer et al (2014); Jacobs et al (2019);
Vernot B, Akey JM. Resurrecting surviving Neandertal lineages from modern human genomes. Science. 2014;343(6174):1017-21
Skov L, Hui R, Shchur V, Hobolth A, Scally A, Schierup MH, et al. Detecting archaic introgression using an unadmixed outgroup. PLoS genetics. 2018;14(9):e1007641;
Browning SR, Browning BL, Zhou Y, Tucci S, Akey JM. Analysis of human sequence data reveals two pulses of archaic Denisovan admixture. Cell. 2018;173(1):53-61

[563]Chen and Zhang also enlisted a German team led by Jean-Jacques Hublin to help characterize the fossil. The work is described by
Chen F, Welker F, Shen CC, Bailey SE, Bergmann I, Davis S, et al. A late middle Pleistocene Denisovan mandible from the Tibetan Plateau. nature. 2019;569(7756):409-12

[564]The Xiahe Mandible carries allele R996K in the gene COL2α1. According to the Chen et al analysis, this matches the Denisovan reference but not three high-quality Neanderthal sequences nor any modern humans. The mandible also carries a unique variant in a different collagen gene. See Supplementary Information 4 of Chen et al for details.

[565]Xia H, Zhang D, Wang J, Fagernäs Z, Li T, Li Y, et al. Middle and Late Pleistocene Denisovan subsistence at Baishiya Karst Cave. Nature. 2024:1-6

[566]Zhang D, Xia H, Chen F, Li B, Slon V, Cheng T, et al. Denisovan DNA in late pleistocene sediments from Baishiya Karst Cave on the Tibetan Plateau. Science. 2020;370(6516):584-7
Gibbons A. Denisovan DNA found in cave on Tibetan Plateau. American Association for the Advancement of Science; 2020

[567]The fossils have been assigned to species known as *Homo longi* and *Homo daliensis*.
Ni X, Ji Q, Wu W, Shao Q, Ji Y, Zhang C, et al. Massive cranium from Harbin in northeastern China establishes a new Middle Pleistocene human lineage. The Innovation. 2021;2(3)

[568]For a deeper consideration of these questions (that argues that the reality is more complicated) see this blogpost by John Hawks: [Link]

[569]Leffler EM, Bullaughey K, Matute DR, Meyer WK, Segurel L, Venkat A, et al. Revisiting an old riddle: what determines genetic diversity levels within species? PLOS Biology. 2012;10(9):e1001388

[570]Sankararaman S, Mallick S, Dannemann M, Prüfer K, Kelso J, Pääbo S, et al. The genomic landscape of Neanderthal ancestry in present-day humans. Nature. 2014;507(7492):354-7
Sankararaman S, Mallick S, Patterson N, Reich D. The combined landscape of Denisovan and Neanderthal ancestry in present-day humans. Current Biology. 2016;26(9):1241-7

[571]Indeed, selection against hybrid ancestry seems to be a common process in many species, as similar observations have also been made in swordtail fish (see below), and in baboons
Vilgalys TP, Fogel AS, Anderson JA, Mututua RS, Warutere JK, Siodi IL, et al. Selection against admixture and gene regulatory divergence in a long-term primate field study. Science. 2022;377(6606):635-41

[572]To calculate local "density", distances are measured in terms of recombination units to reflect the natural length scales of selective processes including hitchhiking and background selection
McVicker G, Gordon D, Davis C, Green P. Widespread genomic signatures of natural selection in hominid evolution. PLoS genetics. 2009;5(5):e1000471

[573] To understand the evolution of incompatibilities, consider a pair of proteins A and B, that interact in some way – perhaps as part of a protein complex. Suppose that in Population 1, a new allele A' arises. If it interacts correctly with protein B it can spread neutrally. Meanwhile in Population 2, a new allele B' arises; again, if it interacts correctly with

A it can also spread neutrally. But if A' does not play nicely with B' then hybrids will have reduced fitness due to the incompatibility of A' with B'. The gene pair A' and B' is referred to as a **Dobzhansky-Muller Incompatibility (DMI)** (or sometimes Bateson-Dobzhansky-Muller Incompatibility). For a very readable history of DMIs see Orr (1996). The accumulation of DMIs is a hallmark of the process by which populations evolve gradually into distinct species that cannot interbreed.

Orr HA. Dobzhansky, Bateson, and the Genetics of Speciation. Genetics. 1996;144(4):1331-6

[574] Suppose that humans and Neanderthals had started to accumulate DMIs by the time that they started interbreeding around 50 KYA. Then hybrid individuals would have had reduced fitness (though evidently not so low as to completely block genetic material flowing from Neanderthals into modern humans). Moreover, in subsequent generations after admixture, chromosomal segments of Neanderthal descent would tend to experience reduced fitness due to the negative interactions with a genome that is mainly of modern human ancestry. Under this scenario, there would be selection to reduce the frequency of introgressed Neanderthal segments. This mode of selection against the minority ancestry in a hybrid population has been shown to occur in swordtail fish (Schumer et al (2018)). One additional piece of evidence in favor of the DMI model is that there is a modest signal that genes expressed in testis have enhanced selection against Neanderthal ancestry (Sankaraman et al (2014), (2016)). Since in other species DMIs often act through hybrid sterility, this provides further indirect evidence in favor of this model.

Schumer M, Xu C, Powell DL, Durvasula A, Skov L, Holland C, et al. Natural selection interacts with recombination to shape the evolution of hybrid genomes. Science. 2018;360(6389):656-60

[575] In this model, the focus is on alleles that are *unconditionally deleterious*, and have drifted up in archaics due to small N_e. Thus, the emphasis is different from Model 1, in which the DMIs can spread neutrally (or under positive selection) within populations, and only become deleterious when divergent populations mix. For more on this see:

Harris K, Nielsen R. The genetic cost of Neanderthal introgression. Genetics. 2016;203(2):881-91

Juric I, Aeschbacher S, Coop G. The strength of selection against Neanderthal introgression. PLoS genetics. 2016;12(11):e1006340

[576] An empirical analysis of missense variants found direct evidence for extra load in Denisovans; curiously, their analysis was not significant in Neanderthals:

Do R, Balick D, Li H, Adzhubei I, Sunyaev S, Reich D. No evidence that selection has been less effective at removing deleterious mutations in Europeans than in Africans. Nature genetics. 2015;47(2):126-31.

[577] An alternative version of the load model appeals to the concept that many traits are under stabilizing selection–a model in which extreme phenotypes (in any direction) are deleterious. Under stabilizing selection, mutation tends to increase phenotypic variance and selection acts to reduce variance. The key point of this model is that, under plausible assumptions, admixture increases phenotypic variance, and so the countervailing effect of selection will be to remove minority genotypes. Since it is the admixture process itself that is responsible for increased variance, this model does not rely on the assumption that Neanderthals/Denisovans had lower N_e and higher load.

Veller C, Simons Y. Stabilizing selection generates selection against introgressed DNA. bioRxiv. 2024:2024-08

Ragsdale AP. Archaic introgression and the distribution of shared variation under stabilizing selection. bioRxiv. 2024:2024-08.

[578] Modeling suggests that DMIs between two isolated populations accumulate at an increasing rate over time: this is the so-called snowball model. Hence a lack of DMIs between modern human populations does not necessarily rule out the existence of many DMIs between humans and Neanderthals/Denisovans given the ∼5-fold longer separation time.

Orr HA. The population genetics of speciation: the evolution of hybrid incompatibilities. Genetics. 1995;139(4):1805-13

[579] A large study of African Americans found no evidence for adaptive shifts in European ancestry either up or down (Bhatia et al, 2014). Moreover, there is no significant difference in genetic load between African and non-African populations (Simons et al, 2015; Do et al 2015, cited above.

Bhatia G, Tandon A, Patterson N, Aldrich MC, Ambrosone CB, Amos C, et al. Genome-wide scan of 29,141 African Americans finds no evidence of directional selection since admixture. The American Journal of Human Genetics. 2014;95(4):437-44

Simons YB, Turchin MC, Pritchard JK, Sella G. The deleterious mutation load is insensitive to recent population history. Nature Genetics. 2014;46(3):220-4

[580] North American snowshoe hares have a white color phase in winter to provide camouflage in snowy conditions. But in the Pacific Northwest there are populations that stay brown in winter, which is better in locations with limited snow. Recent work shows that the brown winter coat is encoded by a highly divergent haplotype of the Agouti locus. This haplotype passed into snowshoe hares through a rare hybridization event with jackrabbits, a related species that stays brown all year. Another nice example comes from Killifish.

Jones MR, Mills LS, Alves PC, Callahan CM, Alves JM, Lafferty DJ, et al. Adaptive introgression underlies polymorphic seasonal camouflage in snowshoe hares. Science. 2018;360(6395):1355-8

Oziolor EM, Reid NM, Yair S, Lee KM, Guberman VerPloeg S, Bruns PC, et al. Adaptive introgression enables evolutionary rescue from extreme environmental pollution. Science. 2019;364(6439):455-7

[581]SNPs in this region not only have the largest F_{ST} between Tibetans and Han but, relative to genome-wide F_{ST}, are among the biggest outliers between *any pair* of human populations (Huerta-Sanchez 2014)

[582]Slightly counterintuitively, high hemoglobin is associated with mountain sickness, and the Tibetan haplotype actually lowers hemoglobin.

[583]Huerta-Sánchez E, Jin X, Asan, Bianba Z, Peter BM, Vinckenbosch N, et al. Altitude adaptation in Tibetans caused by introgression of Denisovan-like DNA. Nature. 2014;512(7513):194-7

[584]It would be tempting to guess that the EPAS1 variant passed directly from Tibetan Denisovans into early humans within Tibet, however the genetic data argue against this model. Instead, recent work suggests that the Denisovan haplotype was present in east Asians at low frequencies for thousands of years, and only became advantageous much later, when the ancestors of modern Tibetans started spending more time at high elevations. Zhang et al (2021) estimated the onset of the sweep at 9,000 years ago, but with a wide credible region from 2,500 to 42,000 years ago. That paper also reviews archaeological evidence regarding the occupation of the Tibetan Plateau.

Zhang X, Witt KE, Bañuelos MM, Ko A, Yuan K, Xu S, et al. The history and evolution of the Denisovan-EPAS1 haplotype in Tibetans. Proceedings of the National Academy of Sciences. 2021;118(22):e2020803118

[585]Racimo F, Gokhman D, Fumagalli M, Ko A, Hansen T, Moltke I, et al. Archaic adaptive introgression in TBX15/WARS2. Molecular biology and evolution. 2017;34(3):509-24

Abi-Rached L, Jobin MJ, Kulkarni S, McWhinnie A, Dalva K, Gragert L, et al. The shaping of modern human immune systems by multiregional admixture with archaic humans. Science. 2011;334(6052):89-94

[586]For alternate viewpoints on this see:

Simonti CN, Vernot B, Bastarache L, Bottinger E, Carrell DS, Chisholm RL, et al. The phenotypic legacy of admixture between modern humans and Neandertals. Science. 2016;351(6274):737-41

Skov L, Coll Macià M, Sveinbjörnsson G, Mafessoni F, Lucotte EA, Einarsdóttir MS, et al. The nature of Neanderthal introgression revealed by 27,566 Icelandic genomes. Nature. 2020;582(7810):78-83.

[587]David Reich's 2018 book is a fantastic overview of the first few years of this field. For a shorter, but also excellent, review see Skoglund and Mathieson (2018, cited above), and the short Lazaridis (2018) review on Europe:

Lazaridis I. The evolutionary history of human populations in Europe. Current opinion in genetics & development. 2018;53:21-7

[588]Additionally, several leading aDNA labs are based in Europe, meaning that it has been practical for them to focus first in this region. Another key factor is that in some parts of the world there are strong cultural sensitivities (for example in north America or Australia) or legal restrictions (for example, India and China restrict or prohibit export of biological samples) that have considerably slowed aDNA research.

[589]This model was investigated in an elegant paper by John Novembre and Matthew Stephens. While they showed that equilibrium spatial structure can produce patterns similar to those observed in Europe, they were careful not to claim that this model is actually the correct explanation in Europe specifically.

Novembre J, Stephens M. Interpreting principal component analyses of spatial population genetic variation. Nature genetics. 2008;40(5):646-9

[590]Keller A, Graefen A, Ball M, Matzas M, Boisguerin V, Maixner F, et al. New insights into the Tyrolean Iceman's origin and phenotype as inferred by whole-genome sequencing. Nature communications. 2012;3(1):698

[591]Fu Q, Posth C, Hajdinjak M, Petr M, Mallick S, Fernandes D, et al. The genetic history of ice age Europe. Nature. 2016;534(7606):200-5

This newspaper article also provides an excellent overview of the early European hunter-gatherers [Link].

[592]Physical evidence for early Homo sapiens, including stones and teeth, found in a cave in France place modern humans in Europe around 54,000 years ago:

Slimak L, Zanolli C, Higham T, Frouin M, Schwenninger JL, Arnold LJ, et al. Modern human incursion into Neanderthal territories 54,000 years ago at Mandrin, France. Science advances. 2022;8(6):eabj9496

[593]Fu et al (2015), cited above

[594]Forms of agriculture were invented independently at around 10 different ancient locations across the globe [Link]; one major epicenter of agriculture was in the eastern Mediterranean, which is relevant to our story here.

[595]For a dramatic illustration of how ancestry changed over time at a single location (Rome), see Figure 2 of Antonio et al (2019):

Antonio ML, Gao Z, Moots HM, Lucci M, Candilio F, Sawyer S, et al. Ancient Rome: A genetic crossroads of Europe and the Mediterranean. Science. 2019;366(6466):708-14

[596] Antonio ML, Weiß CL, Gao Z, Sawyer S, Oberreiter V, Moots HM, et al. Stable population structure in Europe since the Iron Age, despite high mobility. Elife. 2024;13:e79714

[597] It's interesting to compare these results to the two main models that existed before ancient DNA. Certainly, the modern PCA is *not* the result of a long-term model of local drift and migration. But the Cavalli-Sforza model isn't quite correct either. He *was* correct that the Neolithic farming revolution swept through Europe as the result of people moving, not culture. However, this proceeded largely as replacement event, and not by the progressive mixing process that he imagined. However, the Steppe migration (which he didn't know about) did proceed through a progressive mixing process and, by doing so, played an essential role in shaping the genetic structure of Europe.