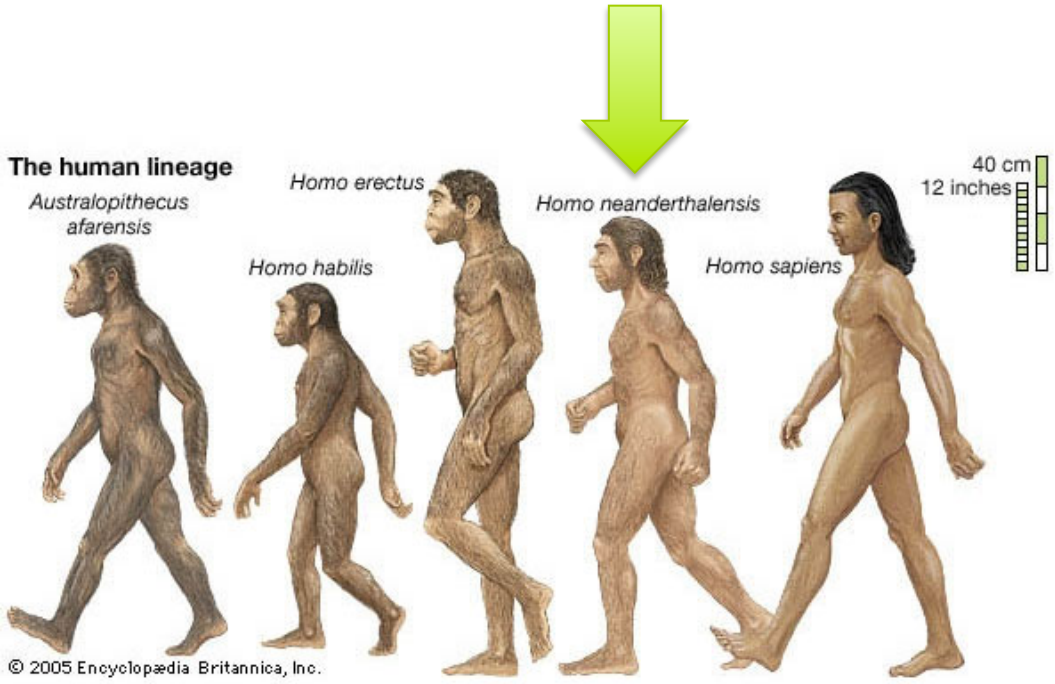
The complete genome sequence of a Neanderthal from the Altai Mountains

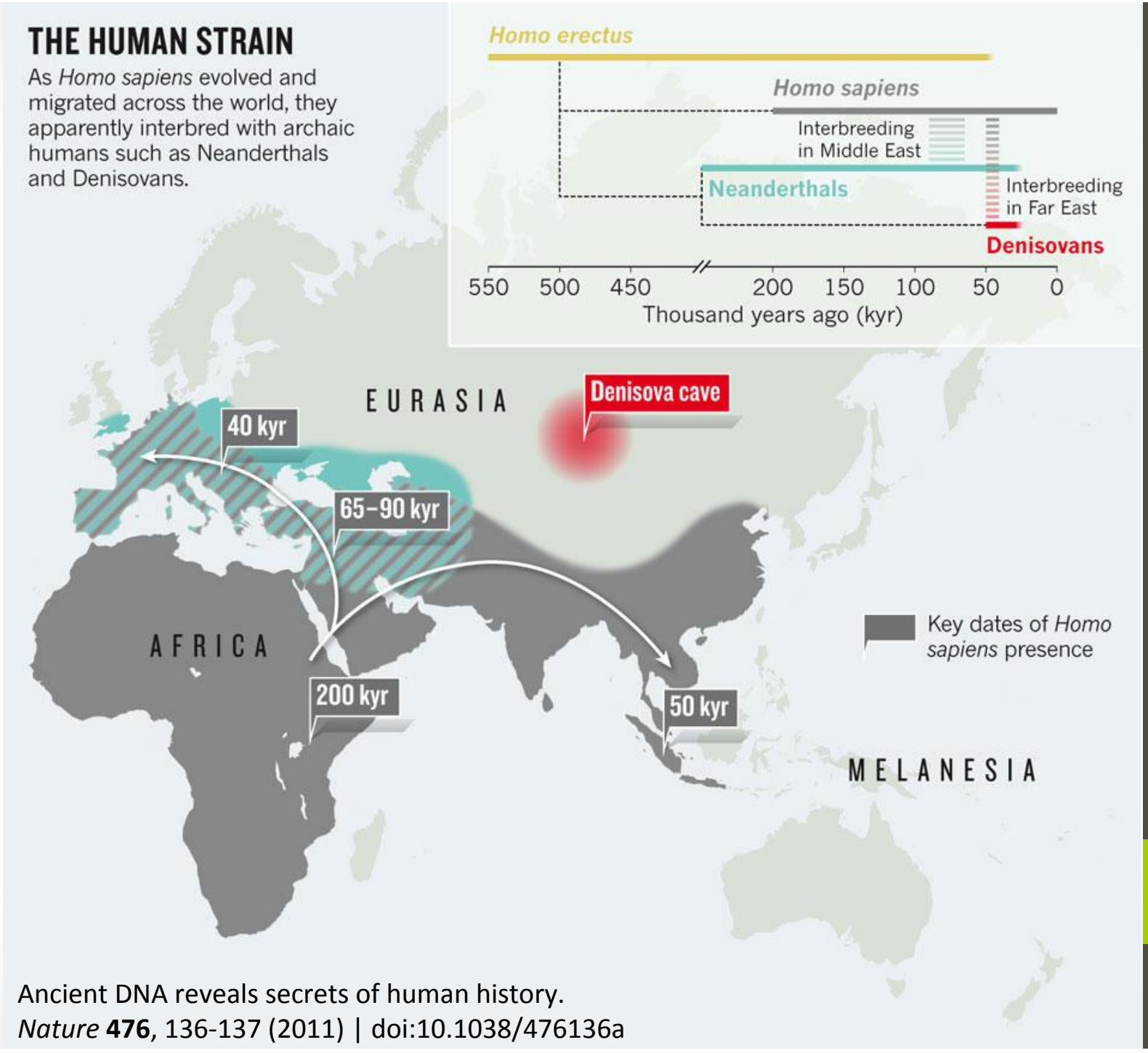
Prüfer et al., Nature, 2 January 2014. 10.1038/nature12886

Critical paper review by Jasleen Grewal (PhD Student, UBC)

Neanderthals…huh?

- *Homo neanderthalensis*
- Our early ancestors in Europe and Western Asia
- Closest evolutionary relatives of modern humans
- Went extinct 25-30,000 years ago





Ancient DNA reveals secrets of human history.
Nature **476**, 136-137 (2011) | doi:10.1038/476136a

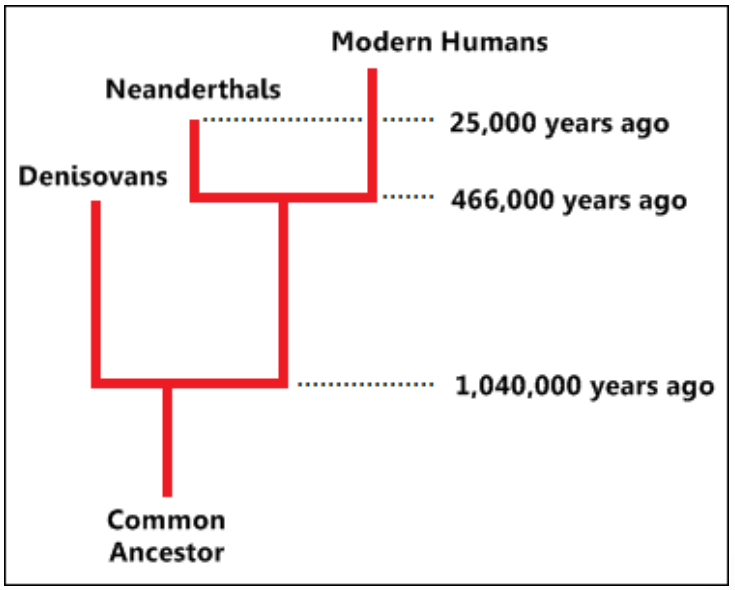
2008: Vindija Cave (Croatia)

- 21 bones found, 3 had Neanderthal mtDNA (screened by PCR)
- 3 individuals - 454 sequencing on extracts
  - 95-99% of DNA sequenced in the libraries was derived from non primate organisms!
  - mtDNA coverage 29-, 35-, and 72-fold respectively.
- 2010: Neanderthal DNA extracted, sequenced, assembled into the **Neanderthal genome sequence**[1]
- Result: Draft Neanderthal genome (1.3x coverage)
  - 1-4% of the genes that modern humans carry, come from Neanderthals

1. Green et al. A draft sequence of the Neandertal Genome. *Science.* 328, 5979 (2010). doi:10.1126/science.1188021. pmid:20448178

2008: Denisova Cave (Altai Mountains)

- Siberia
- Hominin Finger Phalanx, layer age 50,000 years
- Genome sequence generated at ~1.9x coverage in 2010[1]
- Denisovan origins (i.e. distinct from both Neanderthals and modern humans)
- 'X woman'

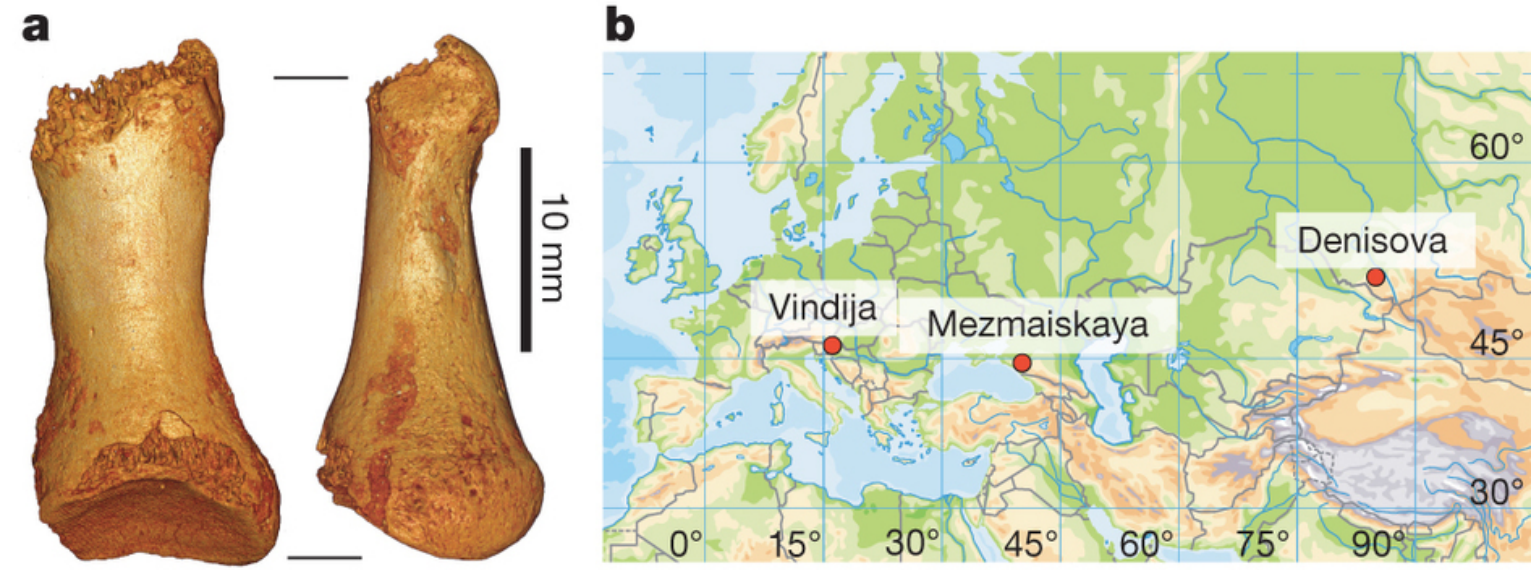


1. Reich D, et al. Genetic history of an archaic hominin group from Denisova Cave in Siberia. Nature. 2010;468:1053.

2010: Denisova Cave (Altai Mountains)

- Siberia
- Hominin toe phalanx, layer age 50,000 years + a little older
- Morphology similar to modern humans and Neanderthals
- Genomic analysis of DNA from this toe phalanx = this paper!

Neanderthal Fossils and Genomics



Location	Age of sample(s)	Year of Discovery
Vindija (Croatia)	38-44,000 years old	2008 (draft genome, mtDNA)
Mezmaiskaya (Caucasus)	29,000 years old	2000 (infant sample, mtDNA)
Denisova (Siberia)	**48-52,000 years old**	**2010 (genome, mtDNA)**

Research Questions at hand:

We have reference Neanderthal and Denisovan sequences, so:

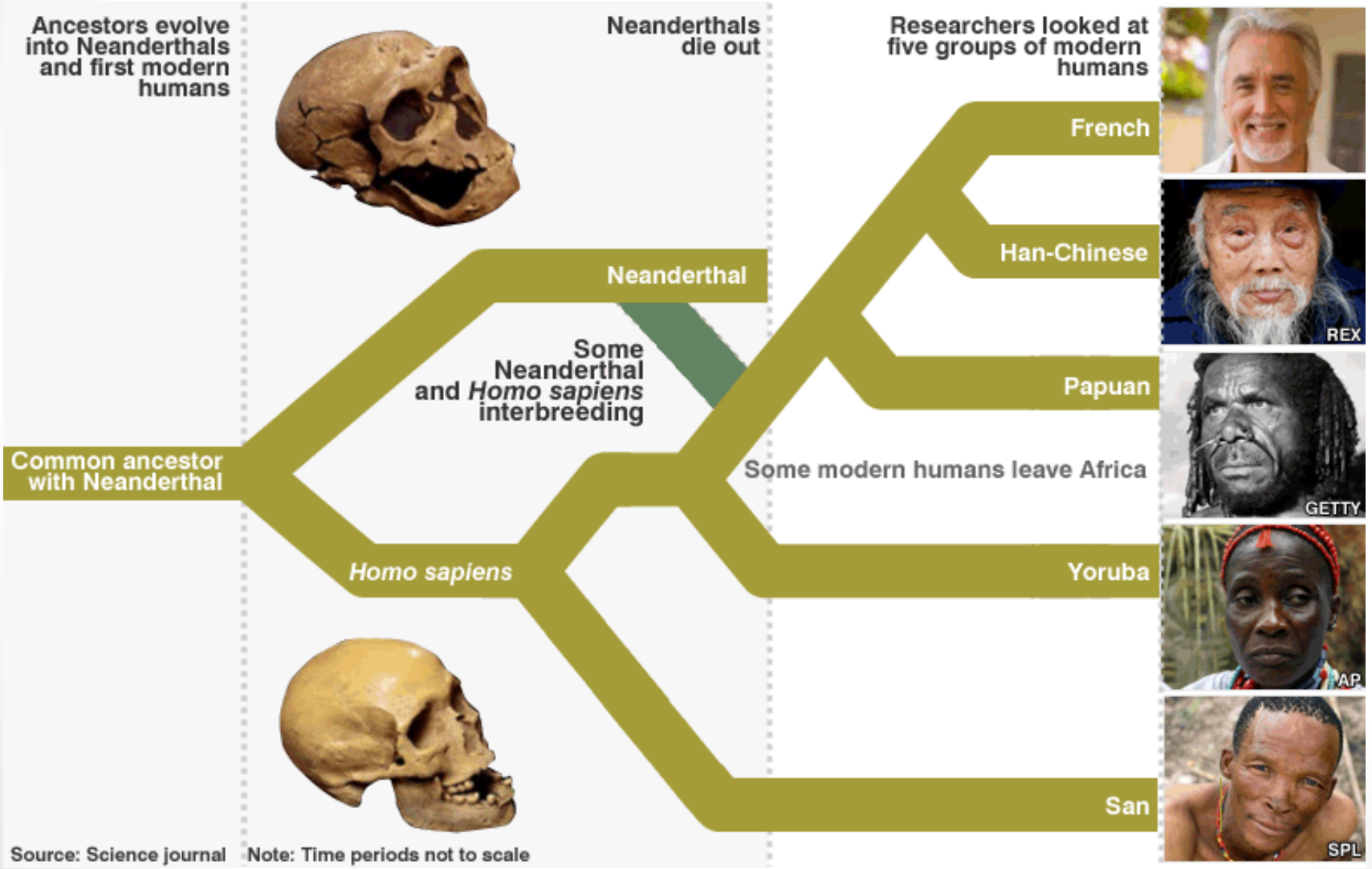
- Can we sequence this Altai Mountains sample and tell which type of hominin (human/human ancestor) it is?
- What new things about the hominins does this new genome sequence tell us?
- How does the modern human genome compare with the Neanderthals or Denisovans?

Methodology

1. Get high quality, high coverage of the genome (and mtDNA) from the toe phalanx sample in Siberia
    1. Illumina HiSeq, alignment to Human + Chimp + hybrid ancestor genomes
    2. Ibis (ML algorithm) for base calling
2. Compare with the genome (and mtDNA) of
    1. other Neanderthals (Vindija cave[1])
    2. Denisovans (Altai Mountains)
    3. Modern humans (representative set of 5 different populations)
3. How do these genomes cluster?
    1. Is the 2010 Siberian sample a Neanderthal?
4. Then set about interpreting the genome – personalized ancestral genomics!

1. Green et al. A draft sequence of the Neandertal Genome. *Science.* 328, 5979 (2010). doi:10.1126/science.1188021. pmid:20448178

What makes a modern human?



Sequencing DNA from fossils

- Fragmented DNA (random pieces)
  - <200 bp fragments of nuclear DNA
  - Chemical effects (deamination)
  - Bacterial contamination, modern human contamination
- mtDNA is well suited for genome-scale ancient DNA sequencing projects[1]
  - Smaller size relative to nuclear genome
  - Higher relative abundance (hundreds of mtDNAs per nuclear genome)
  - Ancestry tracing (track the 'mothers' in a lineage) and discrimination between different populations

1. A. W. Briggs et al., Targeted retrieval and analysis of five Neandertal mtDNA genomes. *Science* 325, 318 (2009). doi:10.1126/science.1174462 pmid:19608918

Altai Mountain Sample

- Initial sequencing of random DNA fragments
  - mtDNA seemed closely related to the Neanderthal mtDNAs
    - Aligned to the Neanderthal mitochondrial genome
  - 70% aligned to the human genome!
  - X chromosome enriched : Female
  - 52 x coverage nuclear genome sequencing

  - Female Neanderthal sample
  - How does it cluster (mtDNA)?

Phylogenetics analysis

- Bayesian clustering results:
  - Build a prior model for substituions
  - Calculate pairwise mtDNA nucleotide differences between all 13 samples + Chimpanzee outlier

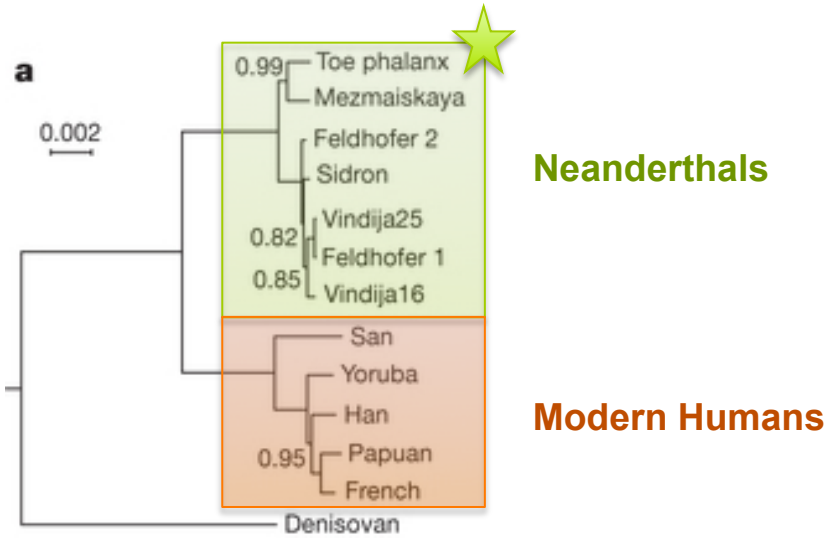


Figure 2 | Phylogenetic relationships of the Altai Neanderthal.
Bayesian clustering tree based on mitochondrial DNA

Studying the Altai Genome

- Know that it's a Neanderthal Female
- We essentially have 2 'haploid genomes' partly representing the 2 parents of this Neanderthal
- How do we trace back the evolutionary history from these two haploid genomes?
  - Enter Pairwise Sequentially Markovian Coalescent Model (PSMC)

PSMC

- Call SNPs on input diploid genome using SAMtools
- Infer the distribution of 'merge time' between the 2 copies of the genome, across all chromosomes
  - Estimate when the 2 haploid genomes shared a genetic ancestor
  - Works off of an average mutation rate expectation (0.5×10^{-9} bp per year)
- The faster the merge time, the smaller the effective population size at that point (**inverse relationship**)
- Can also compare two different populations and estimate when they diverged
- Results as expected (Denisovian and Neanderthal populations were smaller than *Homo sapiens…* – can discuss after presentation)

PSMC also revealed this:

- Runs of homozygosity in the Altai Neanderthal genome
  - >50 kb
  - no heterozygous site
  - <50% missing data
  - <70% missing + filtered data

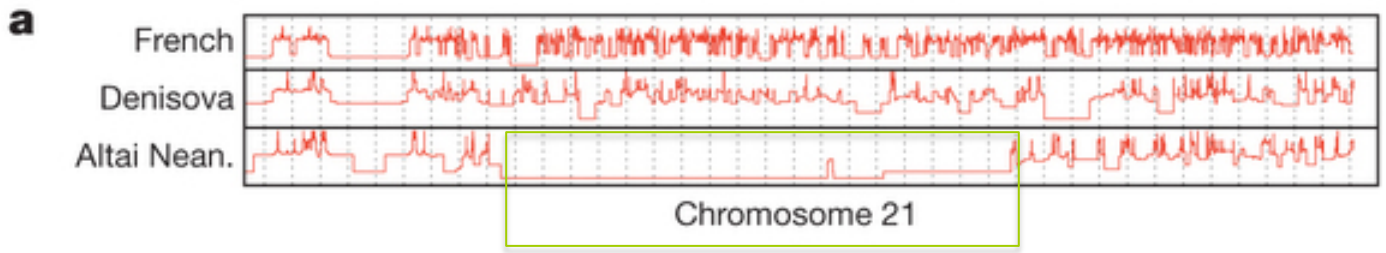


Figure 3a. Time since the most recent common ancestor in log-scale for the two alleles of a French, the Denisovan and the Altai Neanderthal individual along 40 Mb of Chromosome 21

1 cM ~= 1 million bp in the human genome

Inbreeding!

- Had 20 homozygous regions of length >10 cM → inbreeding coefficient of 1/8
- X chromosome also had long runs of homozygosity (XX sample)
  - 1 X chromosome came from the mother (M)
  - 1 X chromosome came from the father (F)
  - Father's mother, or daughter, shared maternal ancestry with M
  - Scenarios?

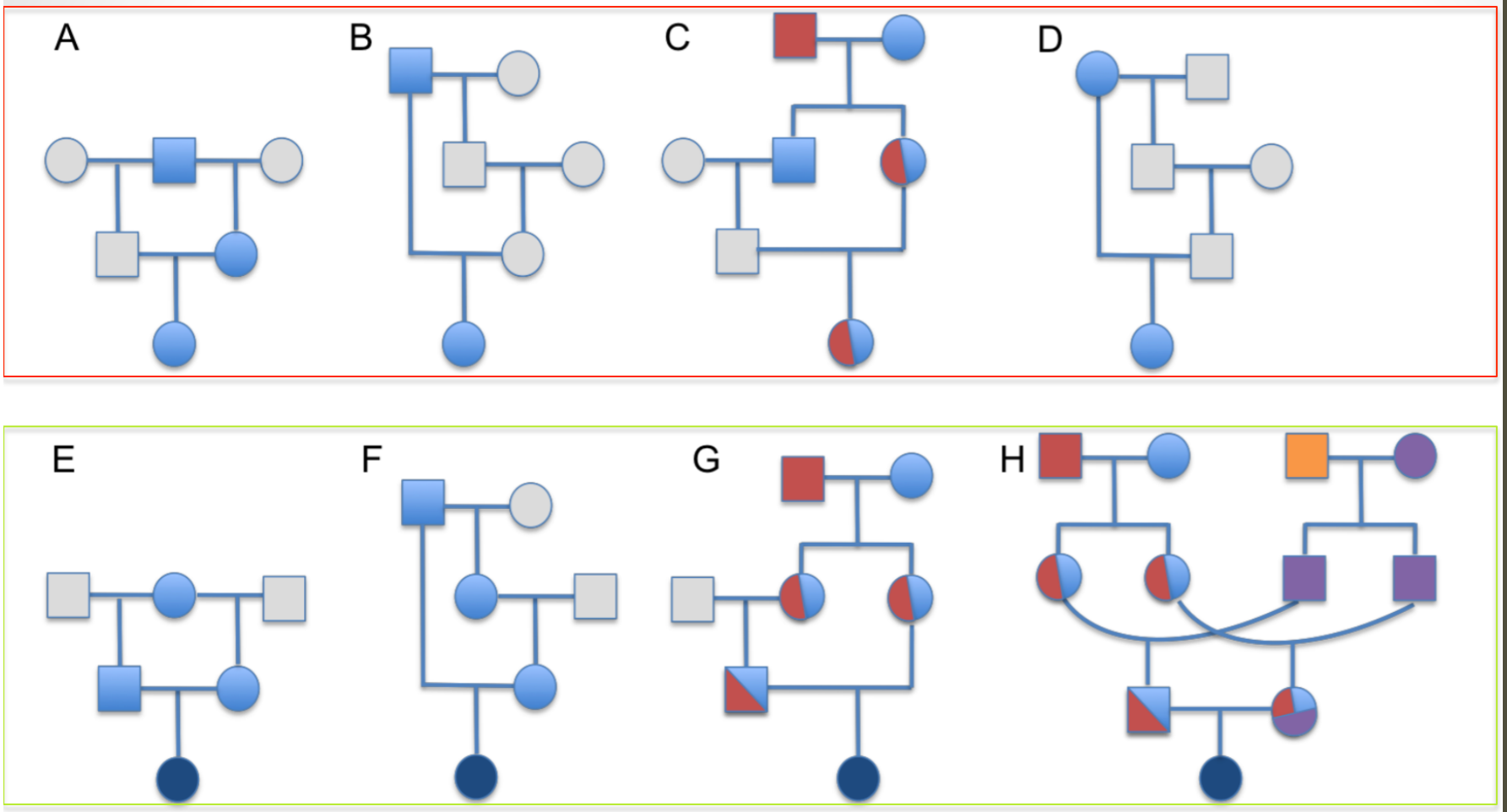


Figure S10.9 Non exhaustive illustration of pedigrees that can be excluded (A-D), or not excluded (bottom), using X chromosome information.
Gray: absence of X sequence coming from common ancestors
Dark blue: Both parents may carry X chunks inherited from same recent common ancestor
Other colours: potential presence of X sequence coming from common ancestor(s)

Perhaps only a 'recent' case of inbreeding?

- If inbreeding happened over more than 3 generations
  - Shorter homozygous stretches (< 10 cM, more time after inbreeding event, to acquire het SNPs in regions of homozygosity)
  - **But** not so short as to originate simply by accrual of mutations in two distinct individuals (False Positives, < 2.5 cM)
- Considered homozygosity stretches of length 2.5 - 10 cM

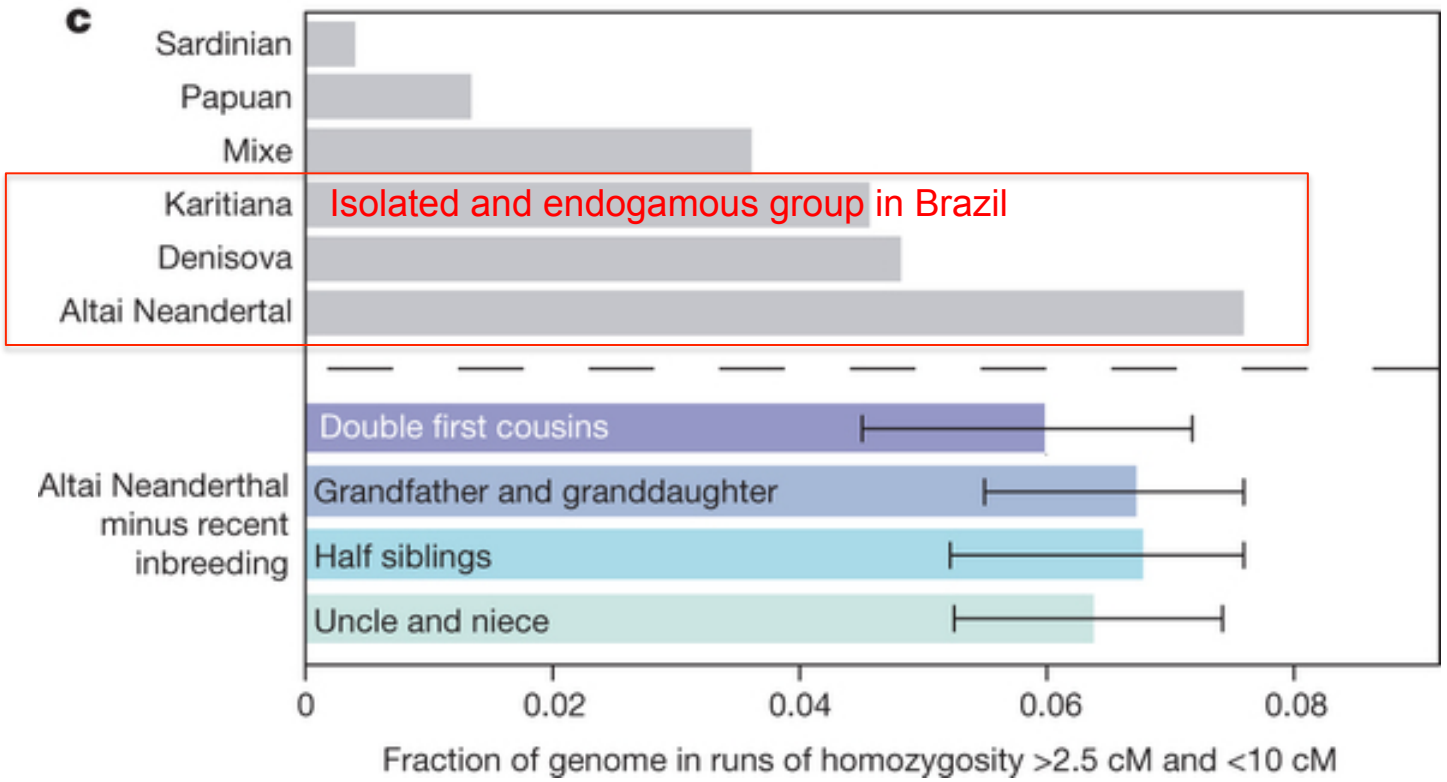


Figure 3c. Fraction of genome in runs of homozygosity between 2.5-10 cM in length for Altai Neanderthal, Denisovan, and the three present-day human individuals with the largest fractions (grey bars)

Perhaps only a 'recent' case of inbreeding?

Results:

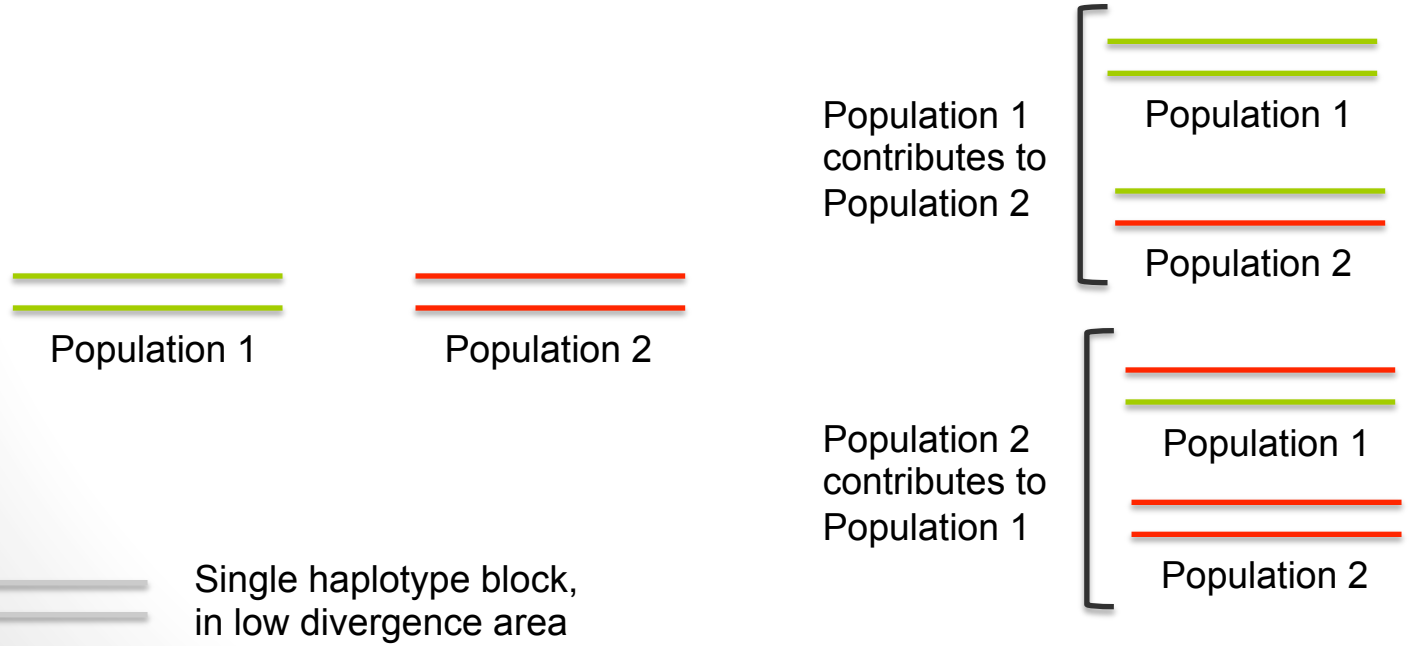
- Altai Neanderthal still had more runs than Denisovan genome ($P < 2.2 \times 10^{-16}$)
- Altai Neanderthal and Denisovan both had more runs than the Karitiana tribe
  - Karitiana tribe has long homozygous stretches
  - Isolated and endogamous group in Brazil → consanguineous individuals
  - 320 people

Other side of the coin

- More homozygous regions = more shared ancestry
- What about heterozygous regions?
- The more similar the heterozygous regions of 2 populations are, perhaps that means the more they mixed with each other?

Did the Denisovans and Neanderthals ever cross paths?

- How different are the Neanderthal and Denisovan genomes?
  - Compared regions around shared alleles between the two genomes
  - If population 1 contributed to population 2, you'd expect more heterozygosity around the shared alleles in population 2 than if population 2 was isolated from population 1



Did the Denisovans and Neanderthals ever cross paths?

- D-statistic as a measure of gene flow
  - Measures the distance between two means, as $(u_1-u_2)/sd\_pooled$

Statistic	D	Z	Interpretation
D(Altai, Mezmaiskaya; Denisova, Chimp)	13.2%	5.9	SI 15: Gene flow between Altai related Neanderthals and Denisovans (Denisovans share more derived alleles with Altai than with Mezmaiskaya)
D(Altai, Vindija; Denisova, Chimp)	7.9%	5.6	

Extended Table 2. Selected D-statistics supporting inferences of gene flow

Did the Denisovans and Neanderthals ever cross paths?

- Found no increases in Neanderthal heterozygosity
- Found more heterozygosity in Denisovan genome
  - Shares more derived alleles with the Altai Neanderthal, than with the Vindija or Mezmaiskaya ones

Oh, the places you'll go!

- Given the two reference Denisovan and Neanderthal genomes, can we predict what parts of a present-day human genome came from which ancestor (if either)?
  - Hidden Markov Model
  - Applied to 13 experimentally phased modern human genomes

Oh, the places you'll go!

- Sardinian and French genomes: genomic regions of Neanderthal origins
- Han & Dai (China), Karitiana & Mixe (Americas): genomic regions of Neanderthal and Denisovan origin

Oh, the places you'll go!

- Did certain alleles from Neanderthals and Denisovans have functional relevance in modern humans?

Functionally important genes (relevant today?)

- Comparison with 1000 Genomes Project samples
  - 96 AA coding changes shared in 1094 modern humans
  - These are absent in Denisovan and Neanderthal genomes
  - ~3000 mutations fixed in modern humans, that influence gene expression
    - Found stretches of DNA in non-African genomes that were totally devoid of Neanderthal DNA (genes involved in motor coordination, language, speech)
  - Several variants shared with the Neanderthal genome
    - Roles published separately (2 months after this paper[1])
    - Neanderthal variants in genes that affect the risk of several diseases (lupus, biliary cirrhosis, Crohn's, type 2 diabetes)

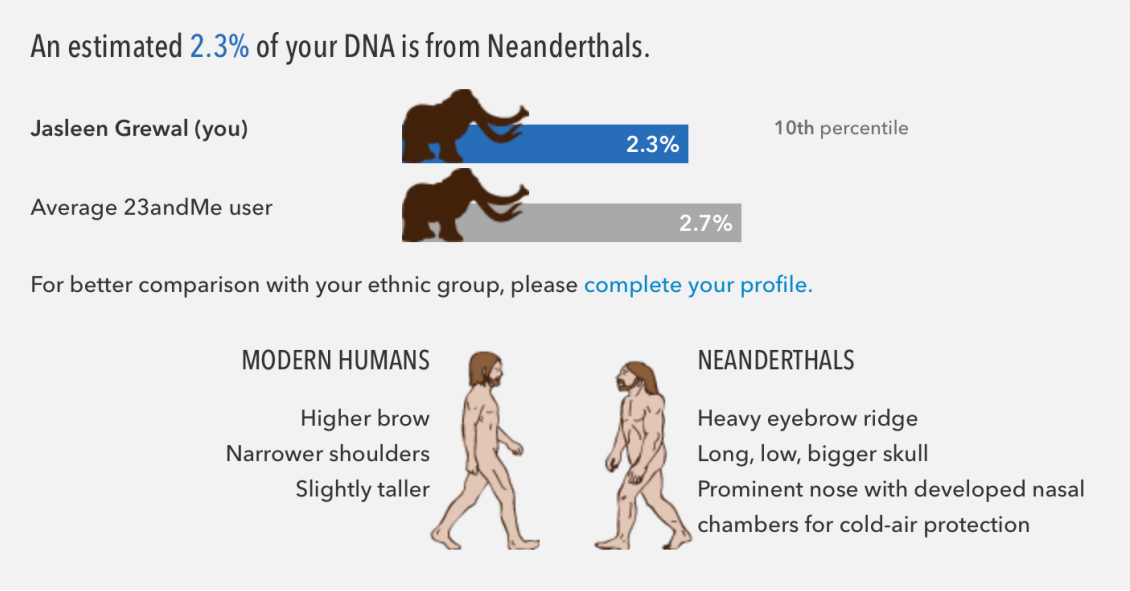
1. The genomic landscape of Neanderthal ancestry in present-day humans. Sankararaman et al. *Nature*. 2014 Mar 20;507(7492):354-7. doi: 10.1038/nature12961.

Conclusions

- The Altai Neanderthal was inbred, and its ancestral population size was small
- Neanderthals contributed to our modern day genomes, and our immune system
  - Strong gene flow from Neanderthals into Denisovans for the Human Leucocyte antigen (HLA) and CRISP cluster regions (immunity and sperm function)!
  - Caucasus Neanderthal shares more derived alleles with the modern non-Africans than do other Neanderthals
- Other mystery humans? (indication of unknown archaic genes in the Denisovans, with gene flow analysis)

From here, whither?

- Qualitative value to knowing how much of your genome is Neanderthal?
  - Comparative genomics
- Neanderthals contributed to our immune systems
- A lesson in the value of evolution (presumably, the bad bits of the *Homo erectus* genome were weeded out with the *H. Neanderthalinsis*)?



Things to think about - methodology

- Why use multiple reference genomes of modern humans, instead of the one established reference genome?
- The original Neanderthal genome was aligned partially to the reference human genome, the reference chimpanzee genome, and an inferred common ancestral sequence of humans and chimps. Why would a de novo assembly be restrictive?
- Analysis was restricted to certain regions of the genome
  - non-repetitive (so no tandem repeats)
  - Would inclusion of tandem repeats be expected to change anything in the results?
- (Im)morally speaking, incest is not a good thing
  - Did the long homozygous stretches kill off the Neanderthals? Can we test out this hypothesis in any way?
  - Does this further our understanding of modern humans?

Neanderthal Fossils and Genomics

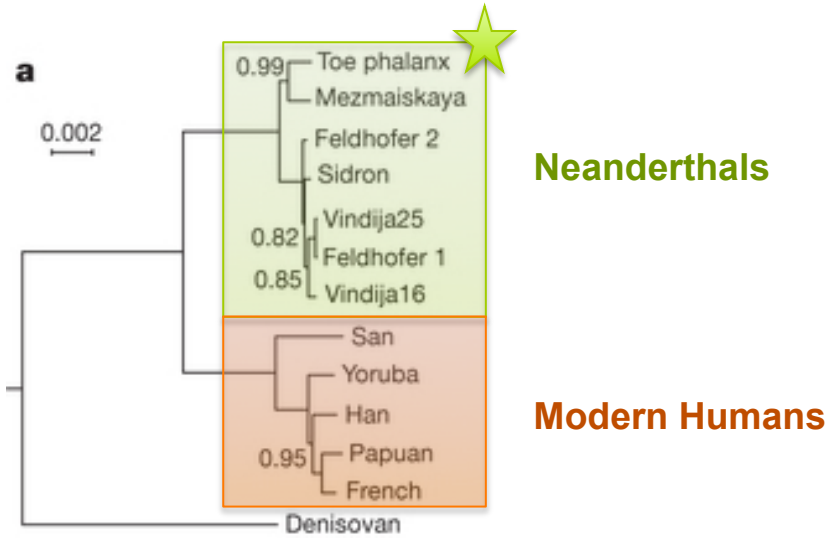


Figure 2 | Phylogenetic relationships of the Altai Neanderthal.
Bayesian clustering tree based on mitochondrial DNA

Location	Age of sample(s)	Year of Discovery
Vindija (Croatia)	38-44,000 years old	2008 (draft genome, mtDNA)
Mezmaiskaya (Caucasus)	**29,000 years old**	**2000 (infant sample, mtDNA)**
Denisova (Siberia)	**48-52,000 years old**	**2010 (genome, mtDNA)**

Trivia (if you read the supplementary)

Is it Neanderthal or Neandertal?

"all the cool kids write it with a 'T'" - John Hawks
(see http://johnhawks.net/weblog/hawks.html)

Did the Denisovans and Neanderthals ever cross paths?

- D-statistic as a measure of gene flow
  - Measures the distance between two means, as $(u_1-u_2)/sd\_pooled$

Statistic	D	Z	Interpretation
D(Altai, Mezmaiskaya; Denisova, Chimp)	13.2%	5.9	SI 15: Gene flow between Altai related Neanderthals and Denisovans (Denisovans share more derived alleles with Altai than with Mezmaiskaya)
D(Altai, Vindija; Denisova, Chimp)	7.9%	5.6	
D(Altai, Denisova; 12 Africans, Chimp)	7.0%	11.6	SI 16: Unknown archaic gene flow into Denisova: Africans share more derived alleles with Altai than with Denisova, a signal that strengthens for fixed derived alleles
D(Altai, Denisova; 12 Africans Fixed, Chimp)	13.4%	10.0	

Extended Table 2. Selected D-statistics supporting inferences of gene flow

We kept the genes for smartness?

- Copy number changes were studies in the two archaic genomes and 25 present day human genomes
- Found 3 regions duplicated only on the modern human lineage
  - BOLA2 had a copy number gain in humans
  - Is near a microdeletion associated with developmental delay, intellectual disability and autism
- Keep in mind that the reference archaic genomes had 1.3-1.8x coverage
  - Is that sufficient to elicit copy number changes in the archaic genome?
  - Can this information be interpreted as selective pressure against developmental disorders? Or do we need more data?

Did the Denisovans and Neanderthals ever cross paths?

- Found no increases in Neanderthal heterozygosity
- Found more heterozygosity in Denisovan genome
  - Shares more derived alleles with the Altai Neanderthal, than with the Vindija or Mezmaiskaya ones
- The modern Africans share more 'ancestrally fixed' alleles with the Neanderthals than with the Denisovans, even though both the ancestral species migrated out of Africa around the same time!

Statistic	D	Z	Interpretation
D(Altai, Mezmaiskaya; French, Dinka)	-16.4%	-5.8	SI 14: The archaic material in non-Africans falls within late Neanderthal variation: Non-Africans share more alleles with some Neanderthals (Mezmaiskaya/Vindija) than others (Altai).
D(Altai, Vindija; French, Dinka)	-7.0%	-4.3	

cM distance (centimorgans)

- Distance between two loci, for which the expected frequency of crossovers in a single generation is 1%
- Two loci are 1 centimorgan apart if recombination between them is observed in 1% of meioses.
- Higher cM value = ‘closer together’
- 1 cM ~= 1 million bp in the human genome

Time of population split?

- For two chromosomes from 2 different populations,
  - How long before these two 'merge'?
  - Use average mutation rate (0.5×10^{-9} bp per year)
  - Apply PSMC
- If we randomly choose alleles from an individual,
  - How often are these different from the apes (derived alleles)
  - The older the population, the fewer the derived alleles

Time of population split?

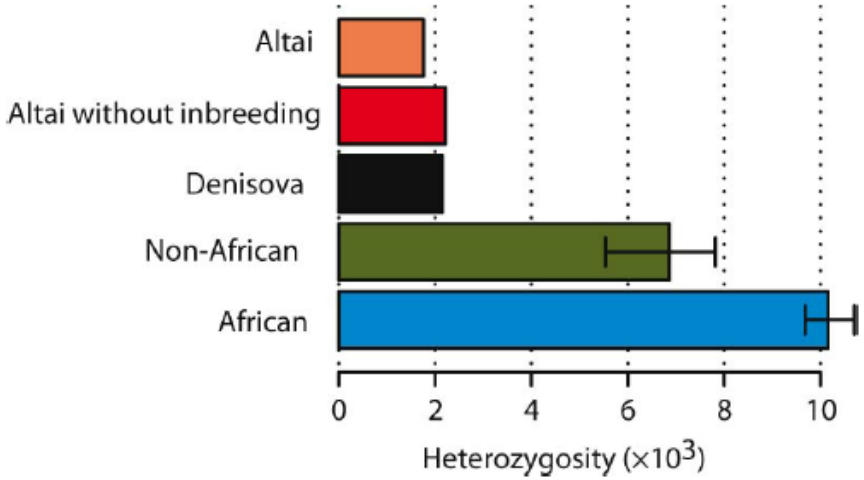
- Pooled estimates from both these methods
  - Neanderthals & Denisovans split 381-473 kya
  - Archaic and Modern Humans split 550-765 kya

Other side of the coin

- More homozygous regions = more shared ancestry
- What about heterozygous regions?
- More variability in these regions means more people were around to carry additional mutations in their genome

Heterozygosity as a proxy for population size

- Remove homozygous stretches > 2.5 cM
- Resulting regions have 2.1-2.2 heterozygous sites per 10,000
- 16-32% of present-day humans
- 81% of Denisova



Extended Data Figure 1 | Heterozygosity estimates for the Altai Neanderthal individual, the Denisovan individual, non-Africans and Africans. The bars for the latter two give the range of heterozygosity observed among 15 non-African and 10 African individuals, respectively (Supplementary Information section 9).

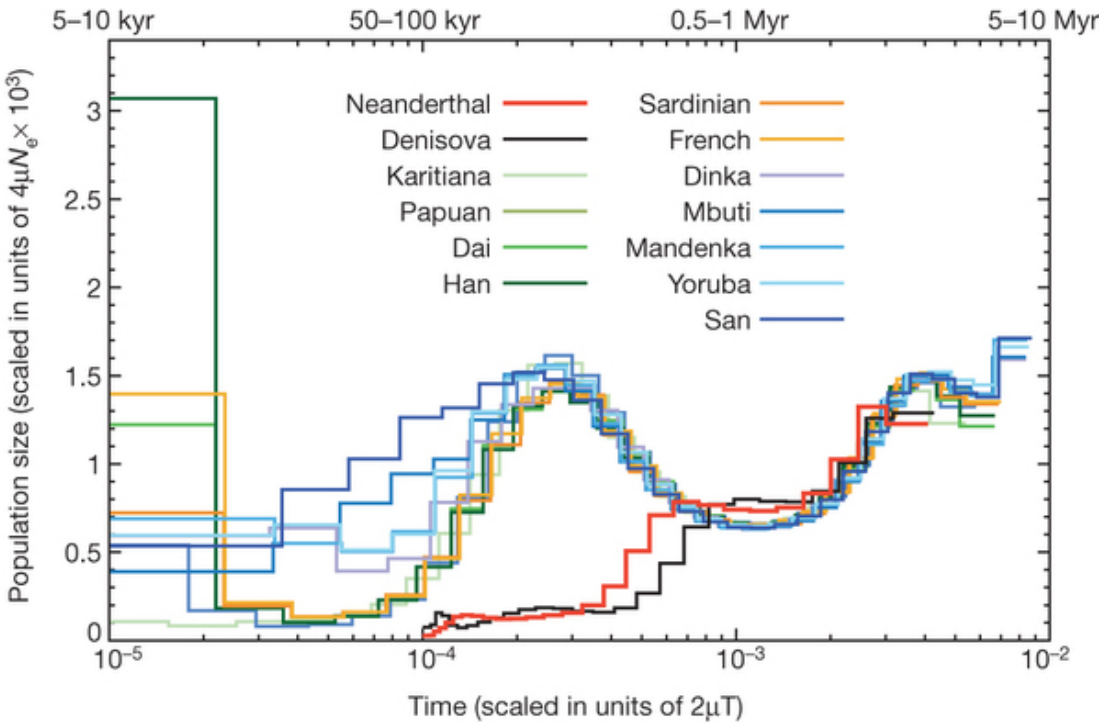


Figure 4. Inference of population size change over time

The tale of 2 chromosomes

- Chromosome 2 fusion makes us different from the great apes
  - Short stretch of telomeric repeat sequence in both directions in the interior of Chr 2, as an artefact of this fusion
- Fusion exists in Denisovan and Neanderthal sequence

Problems with DNA from fossils

- The DNA is degraded to a small average size < 200 bp
- Chemical modification (oxidative processes)
  - High average coverage of random sequence reads
- Small amounts of endogenous DNA, large amounts of DNA from colonizing microbes
  - comparing the genome sequences of closely related organisms
- Modern human contamination (initial Neanderthal fossil extracts had 11-40% DNA contamination)
  - Measurement of modern human mtDNA fragments pegs contamination in samples at <1%
  - Amplification and sequencing of low coverage positions

1. S. Pääbo, Ancient DNA: extraction, characterization, molecular cloning, and enzymatic amplification. *Proc. Natl. Acad. Sci. U.S.A.* 86, 1939 (1989).doi:10.1073/pnas.86.6.1939 pmid:2928314
2. A. W. Briggs et al., Patterns of damage in genomic DNA sequences from a Neandertal. *Proc. Natl. Acad. Sci. U.S.A.* 104, 14616 (2007). doi:10.1073/pnas.0704665104 pmid:17715061

Sequencing DNA from fossils

- 'Filter out' microbial origin fragments by comparing the genome sequences of closely related organisms
  - restriction enzyme digest to preferentially remove bacterial DNA sequences
- Use the finished human genome sequence and the chimpanzee genome to identify Neanderthal DNA sequences.

1. Green et al. A draft sequence of the Neandertal Genome. *Science.* 328, 5979 (2010). doi:10.1126/science.1188021. pmid: 20448178

From supplementary (Figure S6a.3)

To test the effect of contamination on the signal, we recalculate African-Altai and African-Denisova divergence using reads that show deamination patterns. Deamination has been found to increase with time, so reads containing deamination are more likely endogenous[6].

We selected reads with a cytosine to thymine change at the positions where residual deamination is found after uracil removal (last base 5' end + last two bases 3' end; see SI1 and SI5). Bases at these positions were required to have a minimum quality score of 30 (corresponding less than 1 error in a 1000 basepairs). The cytosine residue at the corresponding genomic position was required to be present in the consensus sequence (genotype in the VCF). In order to avoid contaminating human molecules to be falsely classified as deaminated read, we further required that no human from the 1000 Genomes Project phase1 data shows the thymine at the position under consideration. With this methodology, new BAM files containing putatively deaminated reads for the Altai Neandertal and the Denisova were obtained and genotyping was performed according to the method of SI 3. The deaminated reads from Altai Neandertal provided an average coverage of 2.4 while those from Denisova gave an average coverage of 2.0.

Using only the deaminated reads, we recalculate divergence (see Figure S6A.3). We find that Denisova continues to show deeper divergence to African individuals compared to the Altai Neandertal. We conclude that the difference is not caused by a difference in contamination.

Phylogenetics analysis

- Rate of Transversions (Purine-Pyrimidine differences)
  - Deamination increases with time (more Transitions {A->G, C->T})
  - Transversions less likely to be a chemical artifact
- Neighbour Joining results:

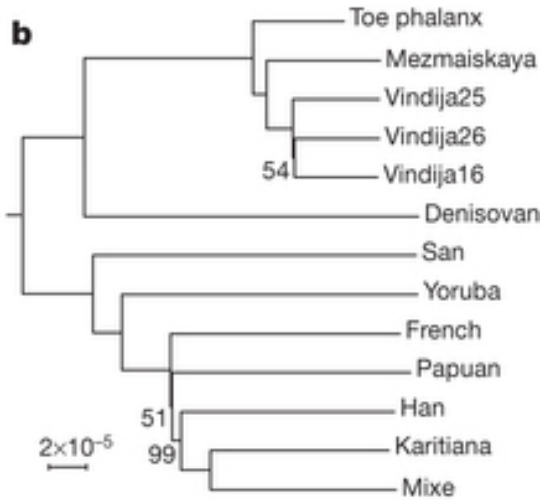


Figure 2 | Phylogenetic relationships of the Altai Neanderthal.
Neighbour-joining tree based on transversions (Purine-pyrimidine differences)

Testing the effects of filtering

Figure S6a.5
Also tested lack of filters in the estimates; Relative het estimates remained stable with a max difference of 3.5%

http://johnhawks.net/weblog/reviews/neandertals/neandertal_dna/altai-neandertal-genome-2013.html
http://johnhawks.net/weblog/reviews/neandertals/neandertal_dna/sima-de-los-huesos-dna-meyer-2013.html
http://siberiantimes.com/science/casestudy/news/n0356-fresh-discoveries-of-ancient-mans-bone-in-altai-mountains-cave/
http://science.sciencemag.org/content/328/5979/710.full

Supplementary:
http://www.nature.com/nature/journal/v505/n7481/extref/nature12886-s1.pdf