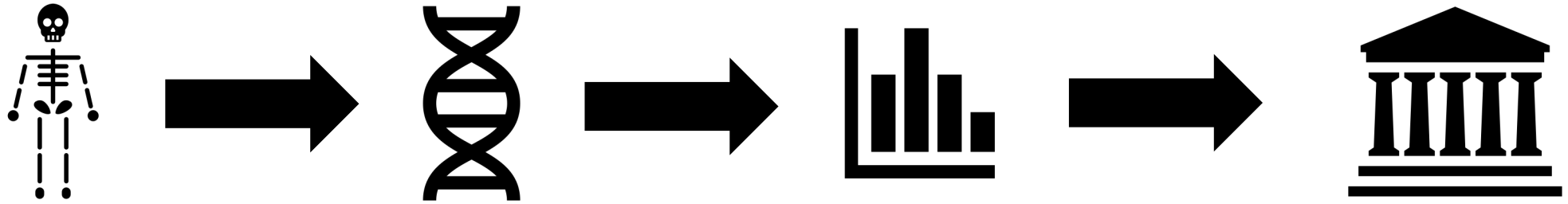
aDNA: Methods and Applications

aDNA analysis

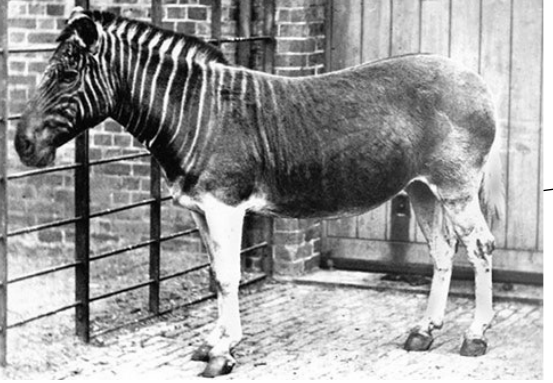


Samples

Genetic information

Bioinformatic analysis

Evolutionary and Historical reconstruction



Quagga

Neanderthal

Pleistocene horse

More than 5,000 ancient humans analysed

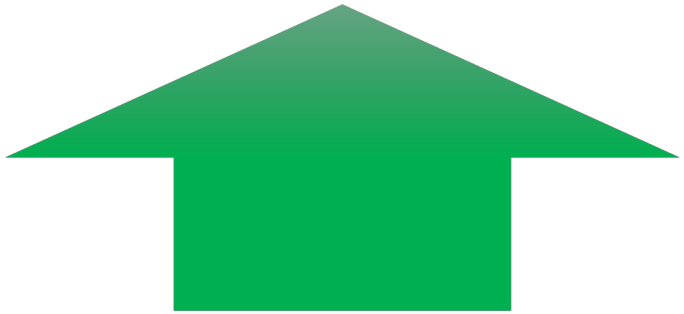
https://umap.openstreetmap.fr/en/map/ancient-human-dna_41837#5/45.106/17.534

- **1984** — First aDNA sequence
- **1988** — First aDNA PCR amplification
- **1989** — First nuclear aDNA fragment
- **1990** — First bone aDNA extraction
- **2001** — First complete ancient mitogenome
- **2003** — First environmental aDNA
- **2005** — First reverse engineering of an ancient virus; First Bayesian Skyline plot using deep aDNA time series
- **2006** — First application of NGS to ancient remains
- **2007** — First characterization of post-mortem DNA damage sequence footprints
- **2008** — First metagenomic analysis of paleofaeces
- **2010** — • USER removal of post-mortem DNA damage • First ancient hominin genomes; • Authentication criteria for NGS sequence data • *D* statistics applied to aDNA data
- **2011** — • First expression and functional testing of ancient gene and protein • First ancient bacterial genome; Post-mortem DNA damage, mapDamage
- **2012** — • *D* and *f* statistics, admixtools • Procrustes PCA projection
- **2013** — • First metataxonomic analysis of ancient dental calculus • First extraction method for ultrashort DNA fragments • Single-stranded DNA libraries • Oldest genome (560,000–780,000 years ago); • mtDNA contamination estimates, contamMix • Recalibration and base-calling, freeIBIS • Analysis of NGS data, ANGSD
- **2014** — • First contamination removal through post-mortem DNA damage • First metagenome and metaproteome of ancient dental calculus • Discovery of high endogenous content in petrous • First ancient epigenome; • First PSMC demography based on complete ancient genomes • Read processing, mapping and genotyping, PALEOMIX • Population ancestry modelling, qpAdm and qpWave • Damage selective filtering of contamination, PMD
- **2015** — • Nuclear DNA contamination estimates, DICE • mtDNA contamination estimates, Schmutzi • Kinship inference, lcMLkin
- **2016** — • Read processing, mapping and genotyping, EAGER • Ancient metagenomic profiling, metaBIT
- **2017** — First automated capture of aDNA; • Ancient epigenetic inference • epiPALEOMIX • aDNA read simulator, Gargammel
- **2018** — • Tree-based selection scans, LSD • Kinship inference, READ
- **2019** — • Damage-aware heterozygosity and ROH estimates, ROHan • Graph-aware detection of selective sweeps, GRoSS • Ancient metagenomic profiling, HOPS
- **2020** — • High-accuracy phylogenetic assignation of metagenomic data • Paleofaeces host species identifier, coproID • Admixture date estimates, DATE • LIMS for aDNA, CASCADE

Pros and Cons of aDNA analysis

Hard to analyse

Analyse a precise moment

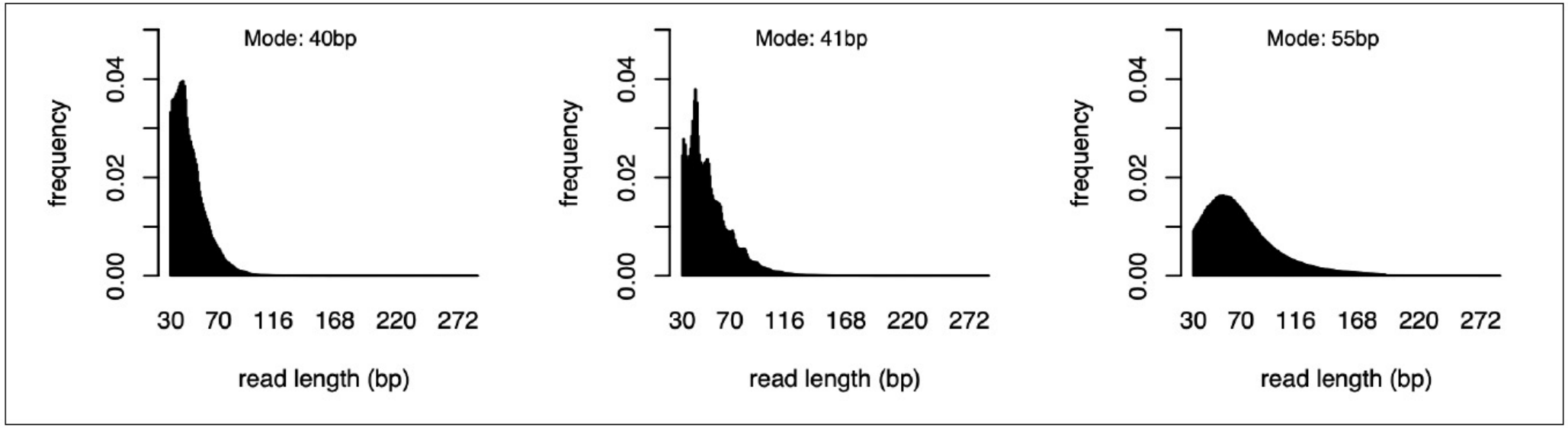


Characteristics of aDNA

Degradation: Fragmentation and post-mortem damage

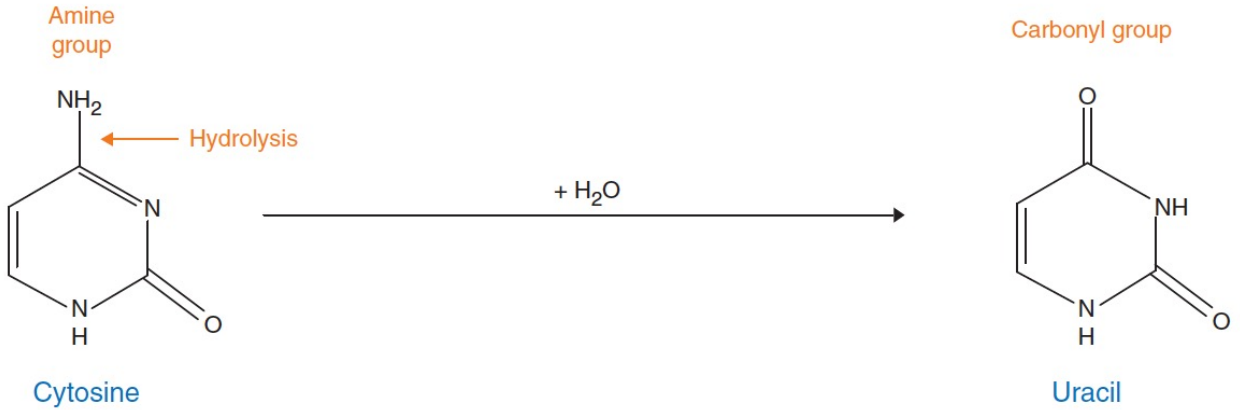
Characteristics of aDNA

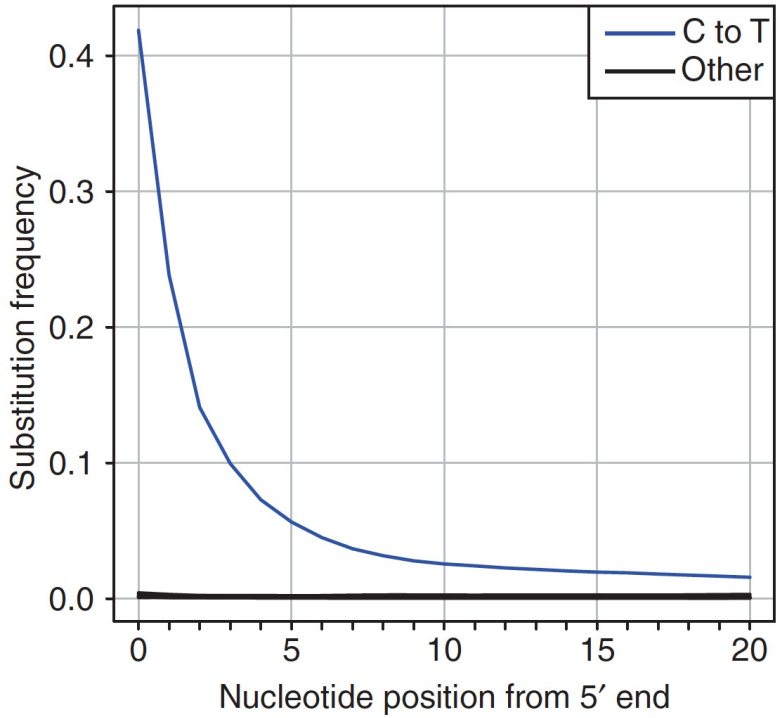
Degradation: Fragmentation and post-mortem damage



Characteristics of aDNA

Degradation: Fragmentation and post-mortem damage

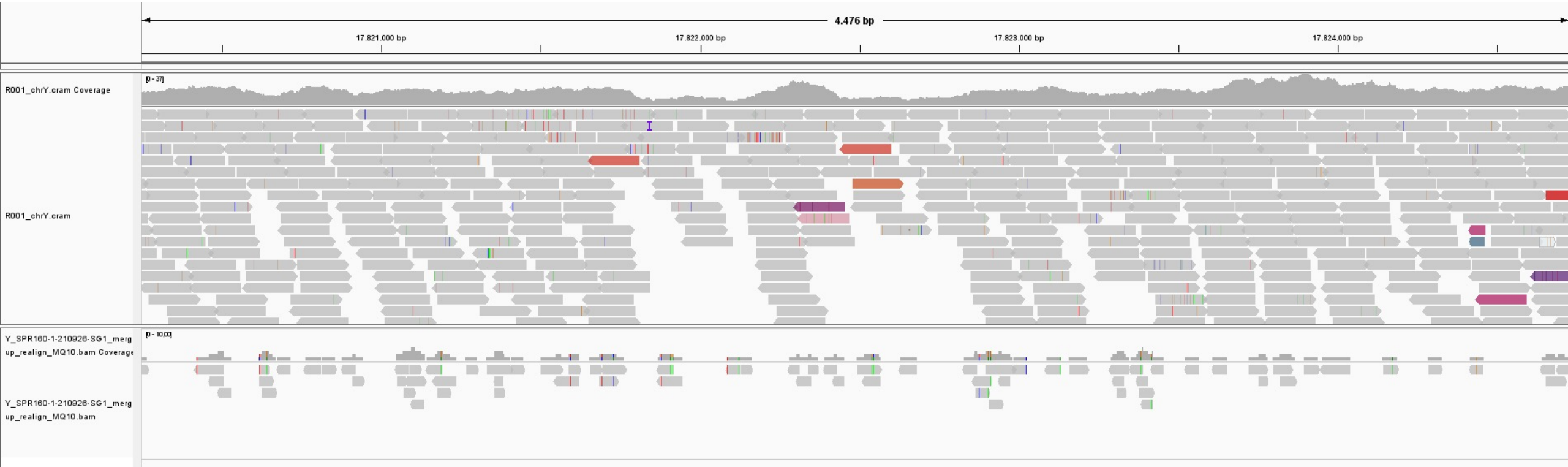




Characteristics of aDNA

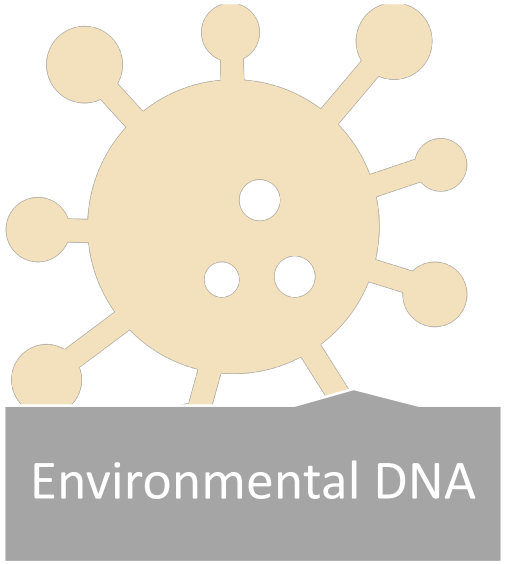
Usually found in low quantities

↳ Resulting in low coverage sequences (<1)



Characteristics of aDNA

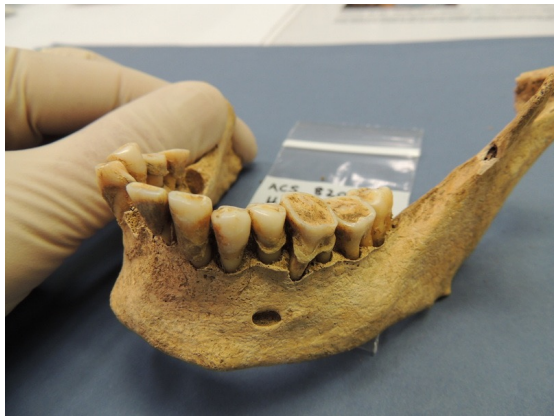
Potentially contaminated

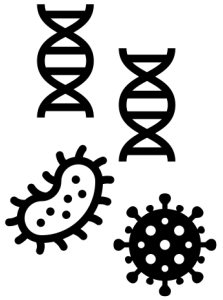


Not mapped to the human reference sequence

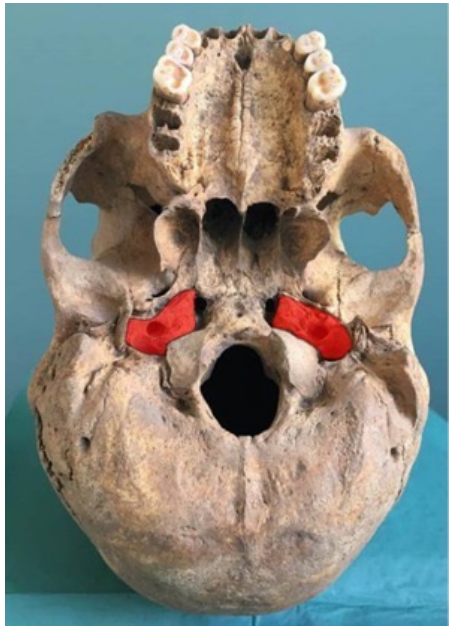
Controlled environment to prevent contamination

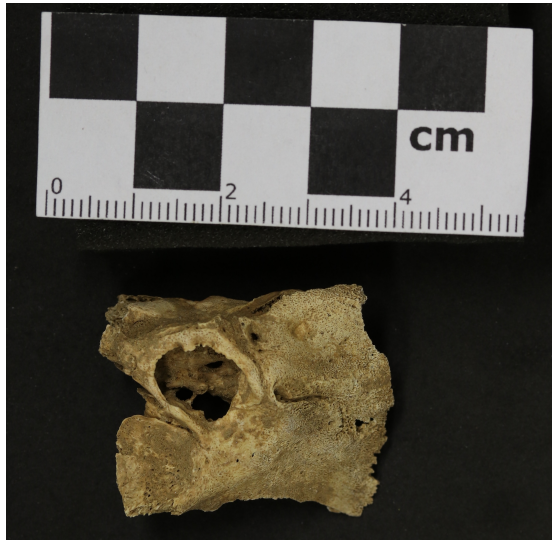
Sample collection

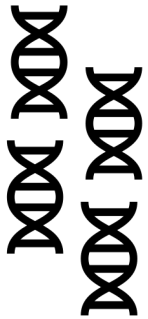




Tooth: relatively less DNA molecules, but greater chance to find ancient pathogens.

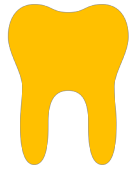






Petrous bones: relatively more DNA molecules. Not optimal for ancient pathogen search.

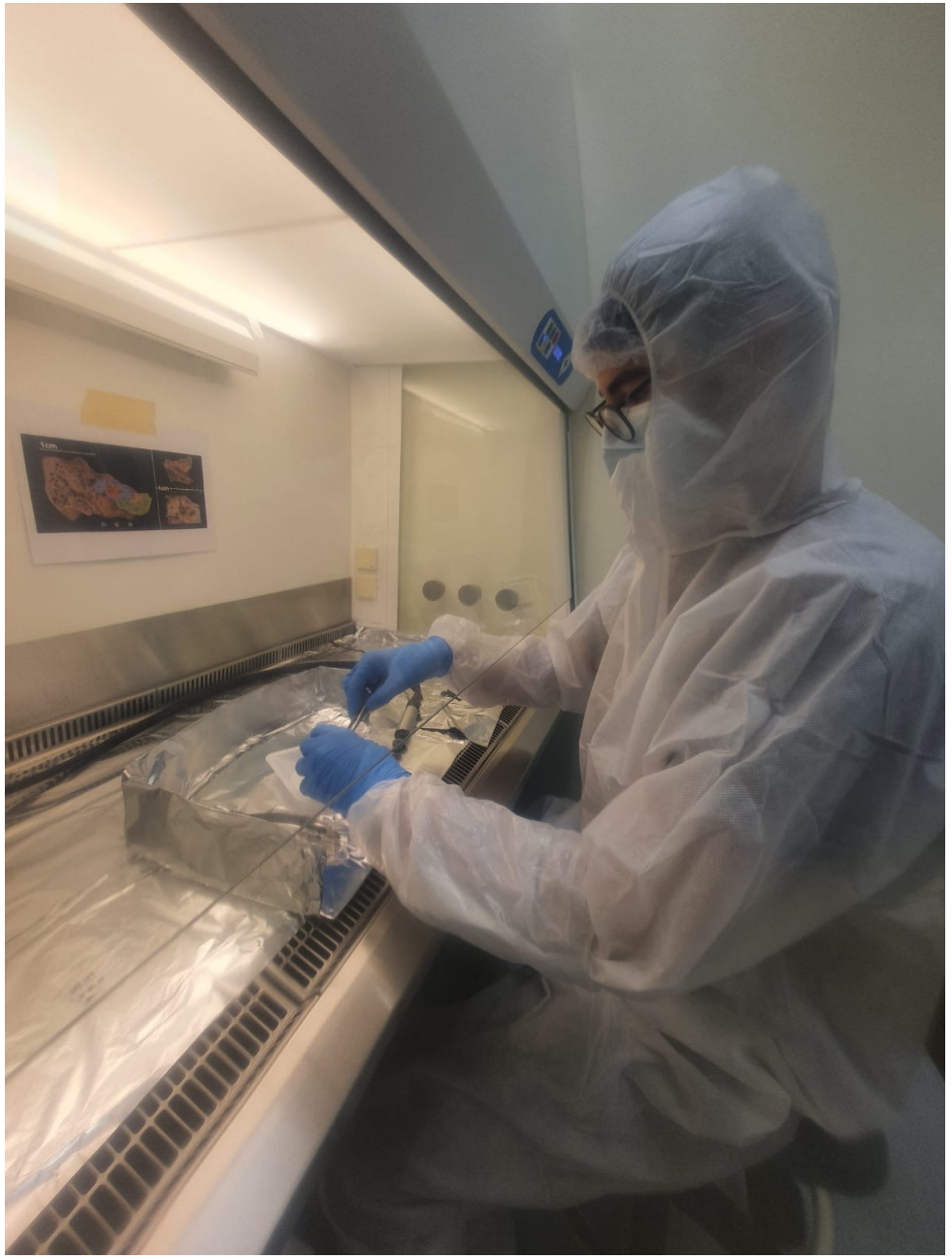
Sample collection

- More endogenous DNA in the petrous bone 
- Possibility to recover ancient pathogens from teeth 
- How destructive is the method? 
- Samples may be used for other analysis 

aDNA clean lab

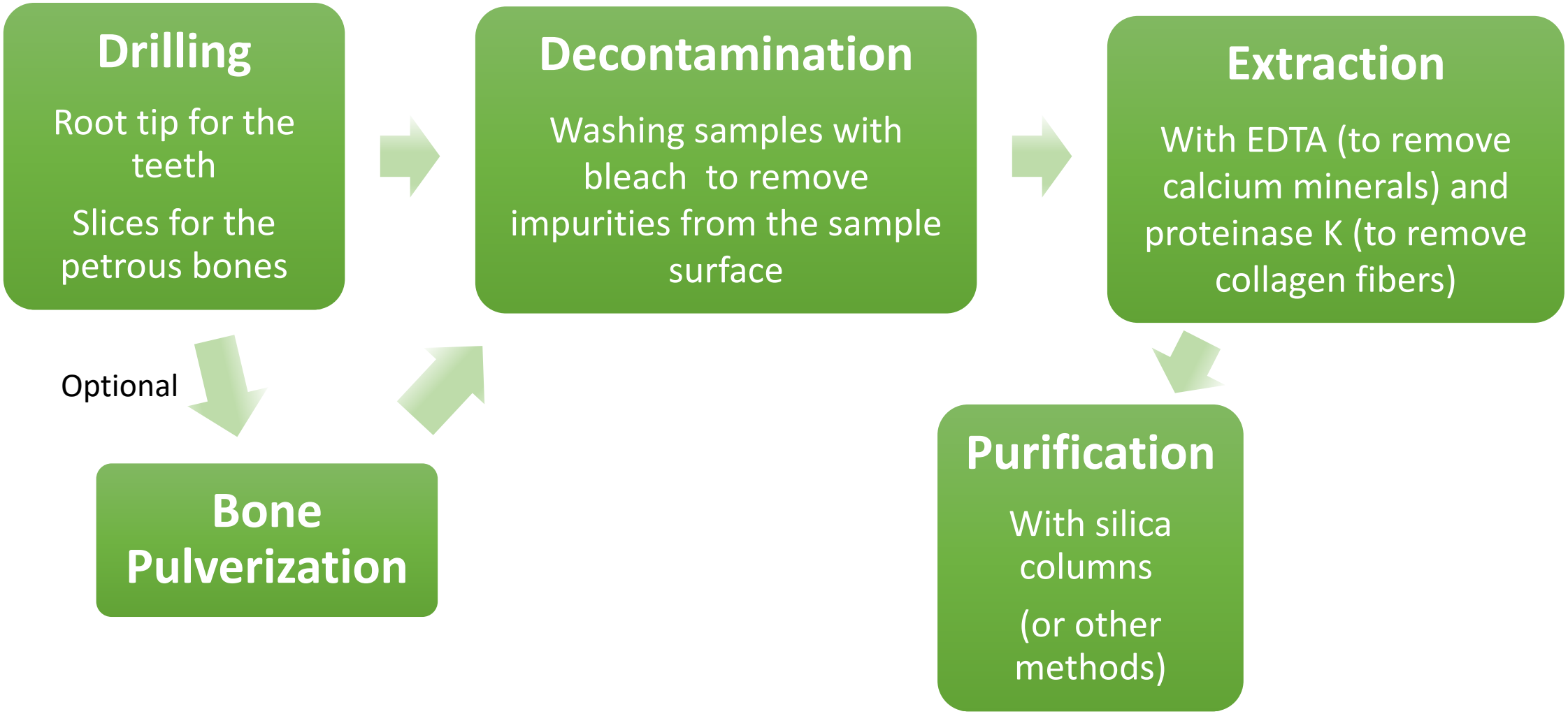
The laboratory is designed to prevent contamination:

- Controlled environment
- Positive pressure
- Filtered air
- UV light (optional)
- No entry without security devices (suits, masks, gloves etc.)
- Compartmentalized laboratory (one room for each operation)
- All objects brought from outside must be cleaned with appropriate products (or bleached)
- Daily cleaning

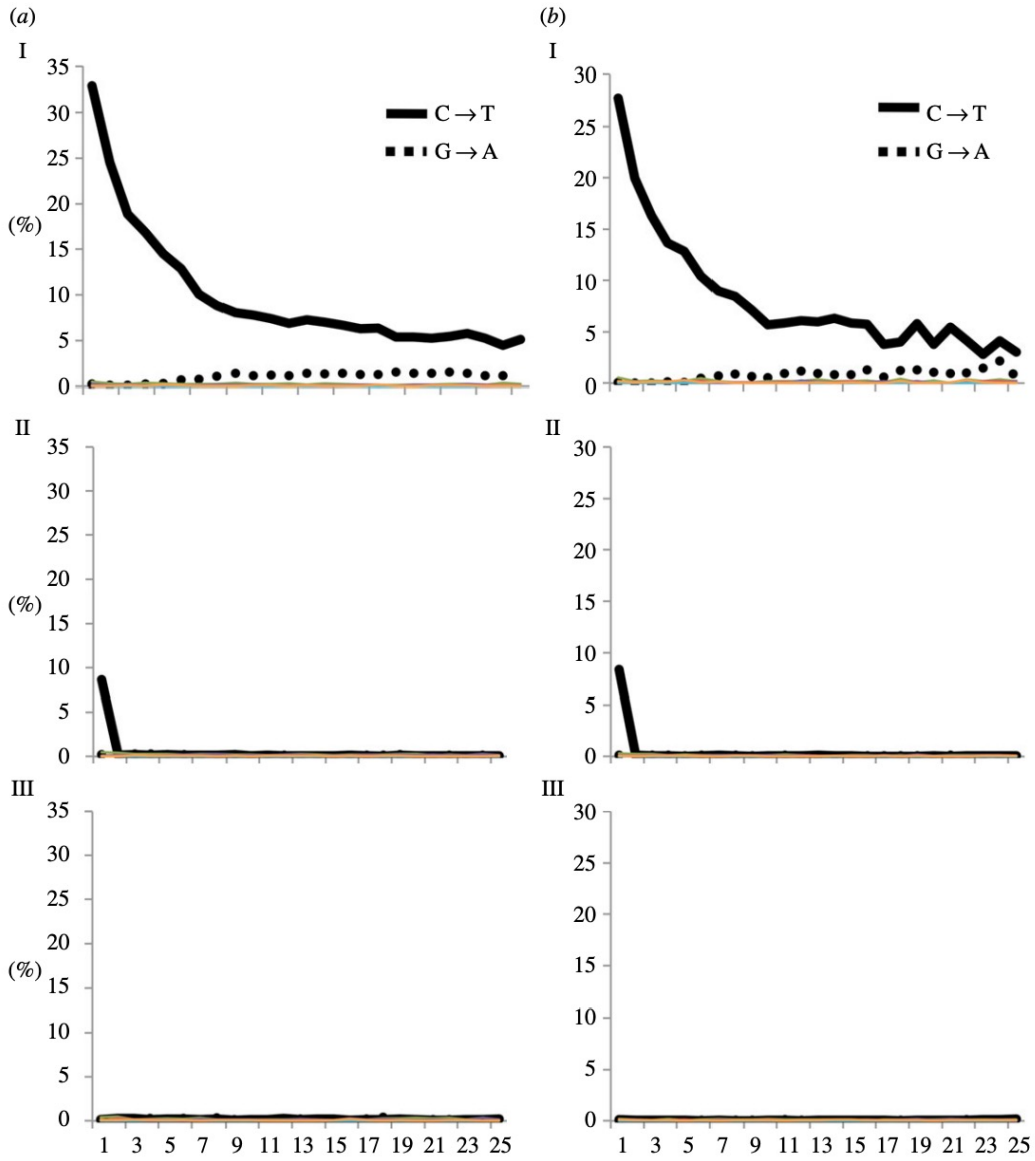


DNA extraction

There are several protocols that can be used to extract DNA from bones and teeth



UDG treatment (optional)



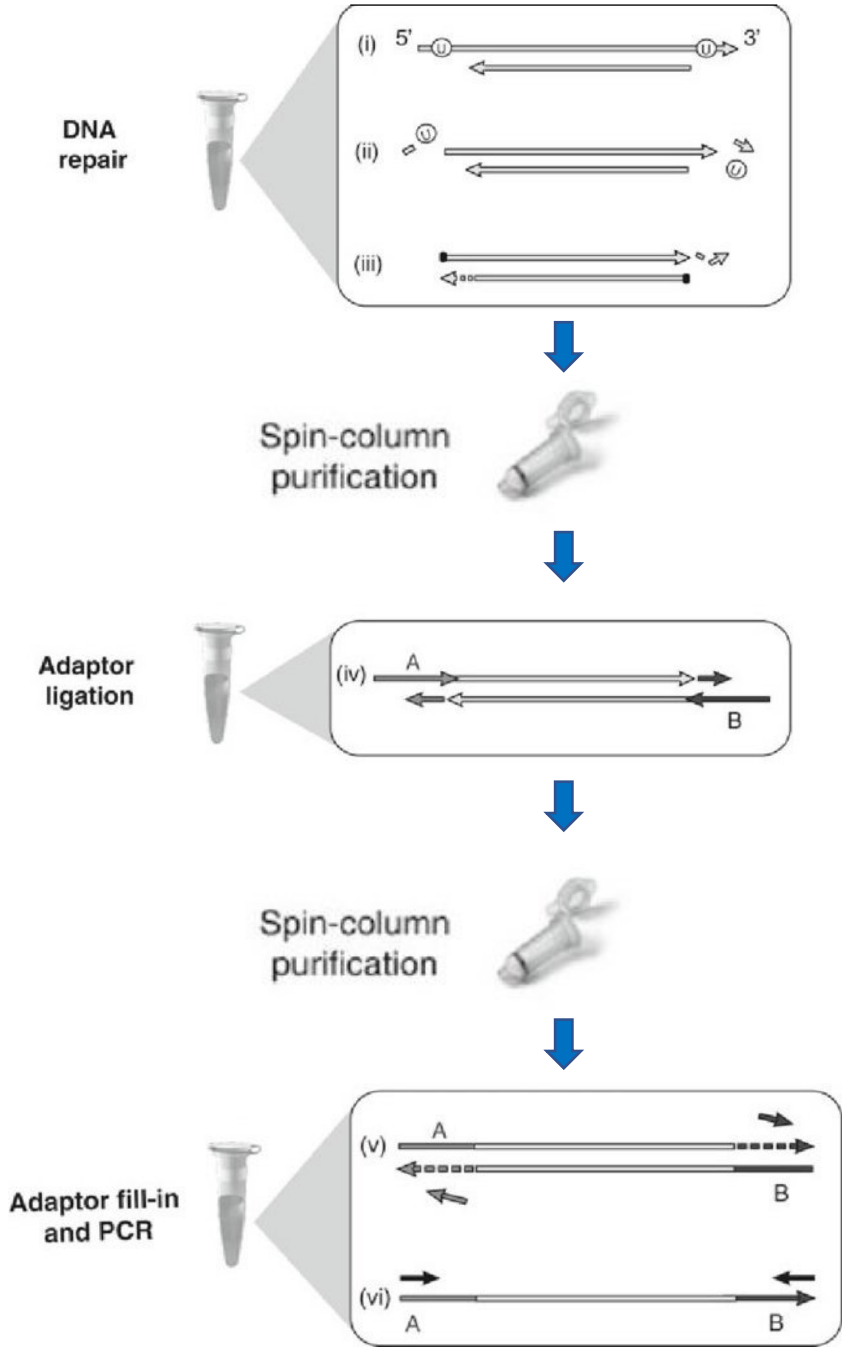
➡ No UDG treatment

➡ Partial UDG treatment

➡ Full UDG treatment

Library preparation

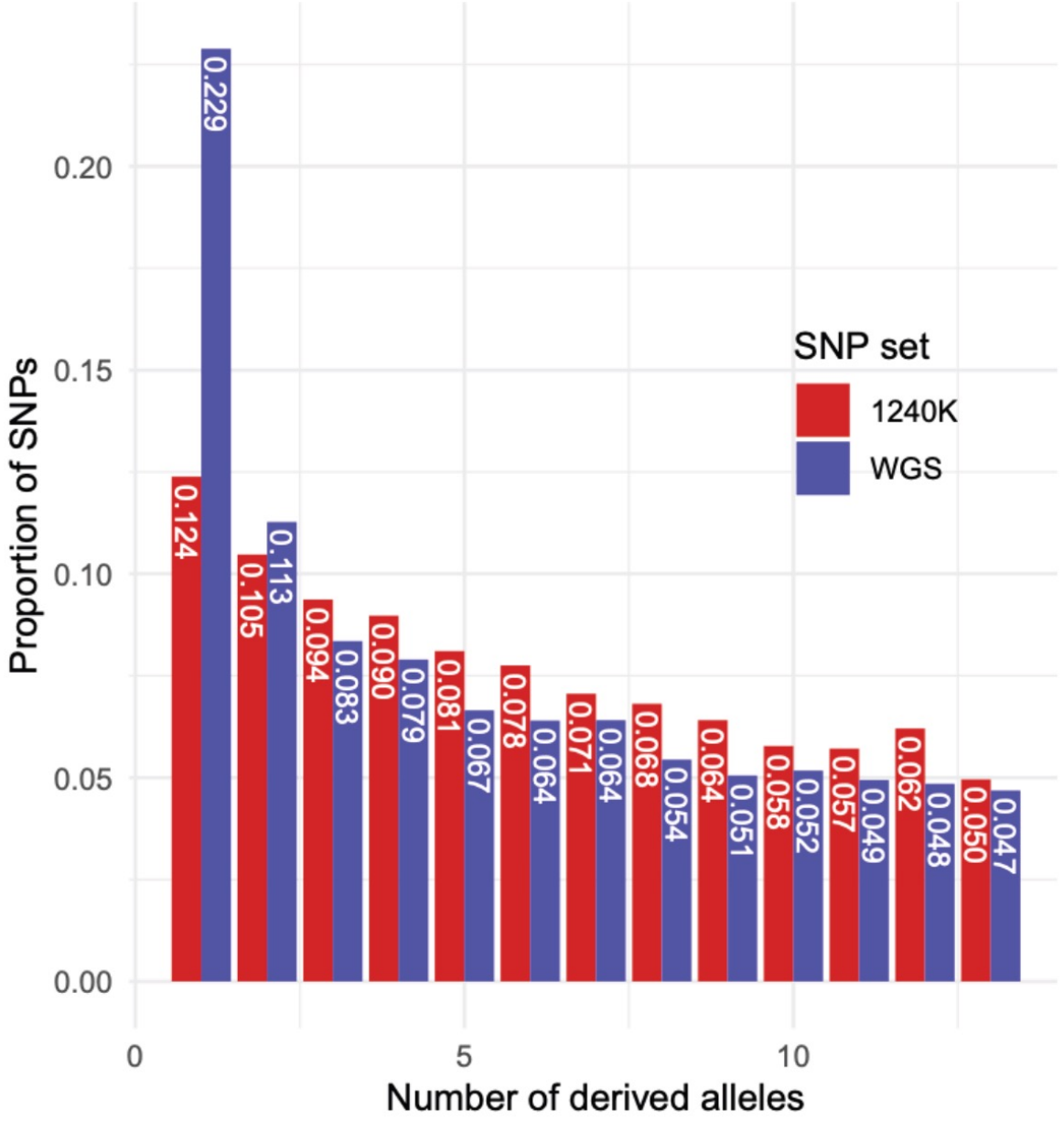
- Fragmentation (not needed in aDNA)
- DNA molecule end repair
- Adaptor ligation (with indexes for the sequencing)
- Adaptor fill-in
- PCR (outside the clean lab)



Sequencing

NGS Sequencing:

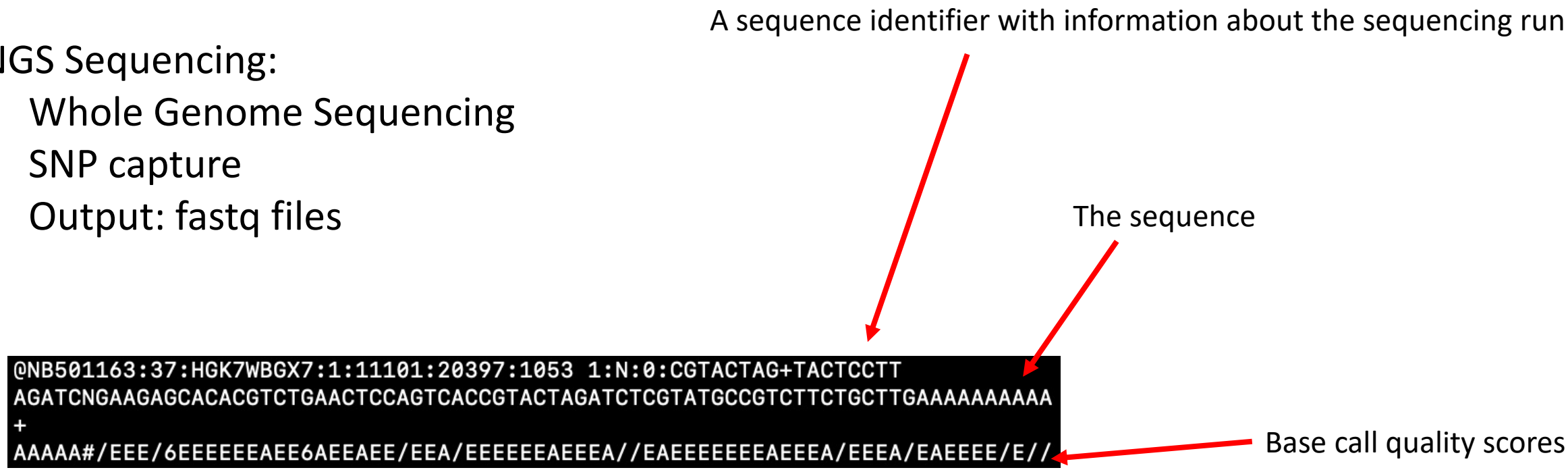
- Whole Genome Sequencing
- SNP capture



Sequencing

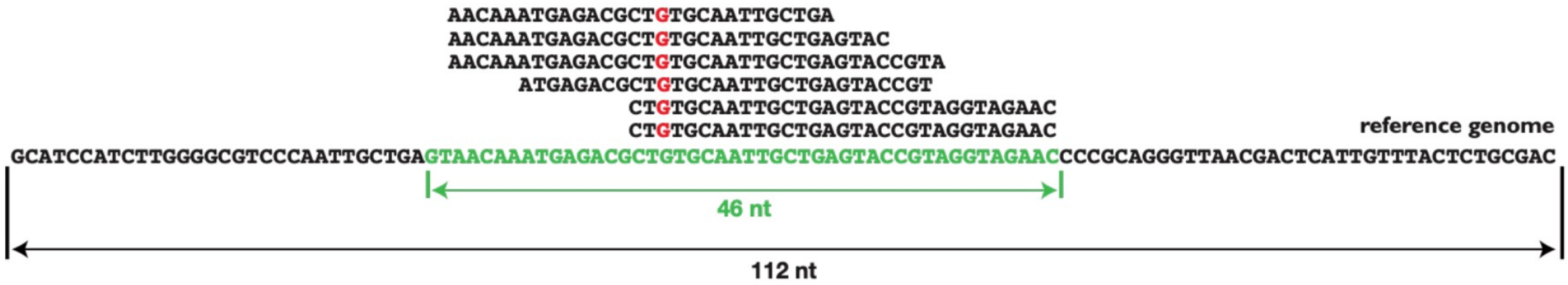
NGS Sequencing:

- Whole Genome Sequencing
- SNP capture
- Output: fastq files



Mapping

Mapping sequencing reads (from fastq files) to the reference genome

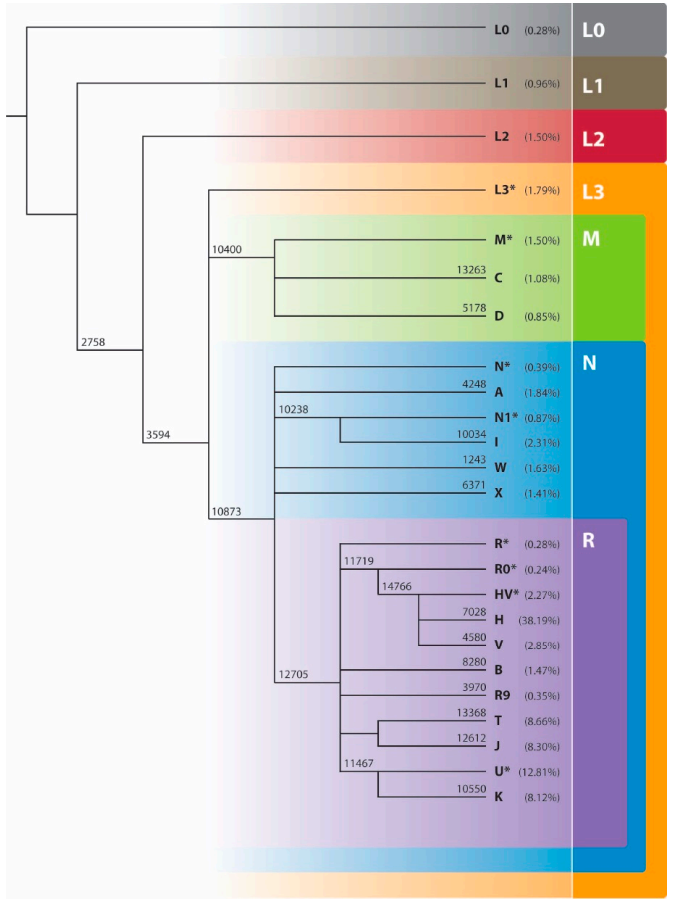


Authentication

- **Amount of endogenous DNA** (mapped/unmapped reads ratio)

- **Ancient or modern DNA**
  - Read length
  - aDNA damage

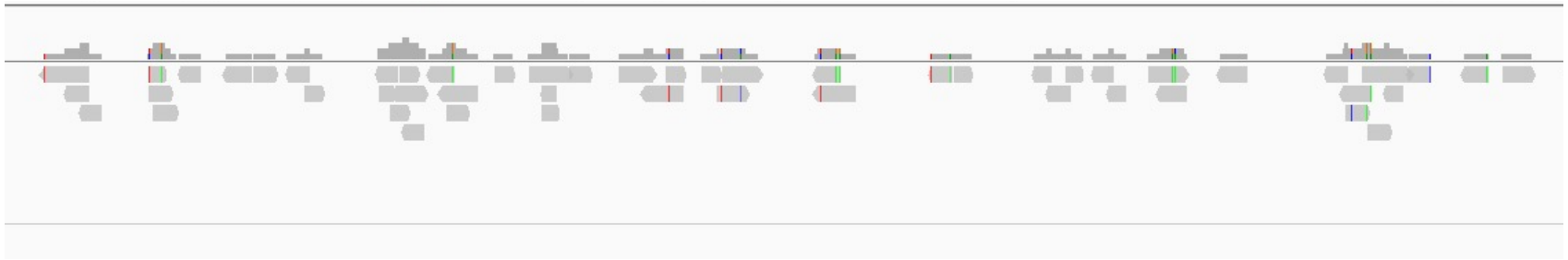
- **Contamination**
  - X-based method (only for male samples)
  - mtDNA method (Calculating the percentage of non-consensus bases at haplogroup-defining positions)



Variant calling

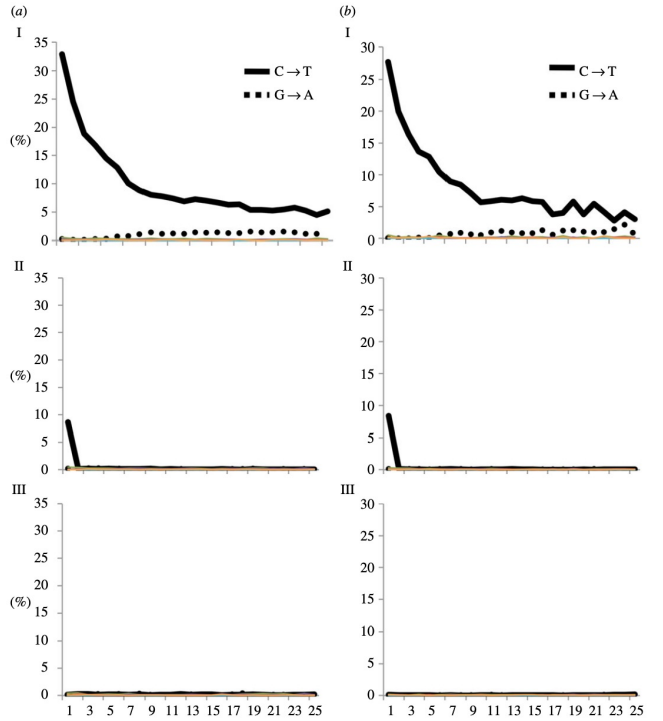
Variant type:

- Genotypes
- Pseudo-haploid genotype
- Genotype likelihoods



Deal with post-mortem damage:

- Trim reads for partially UDG-treated samples
- Remove transitions (C <-> T, G <-> A)
- Likelihood methods



No UDG treatment

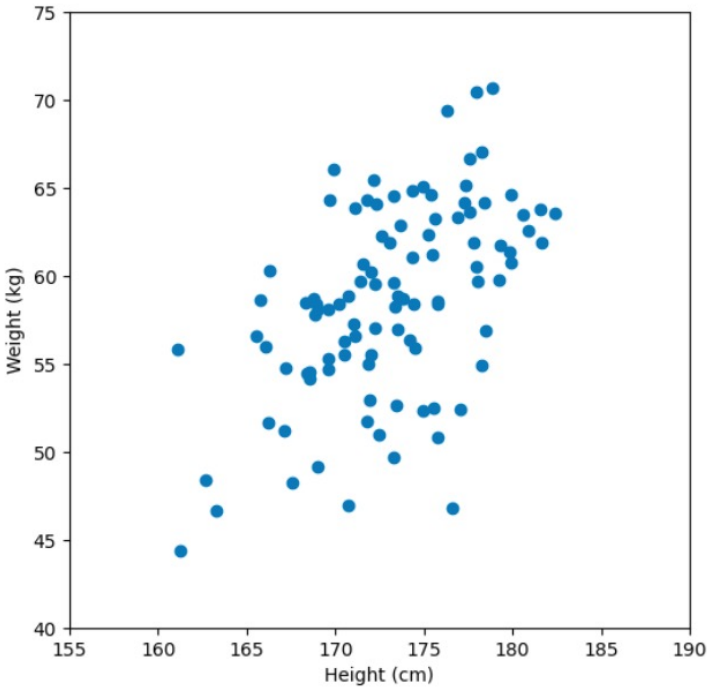
Partial UDG treatment

Full UDG treatment

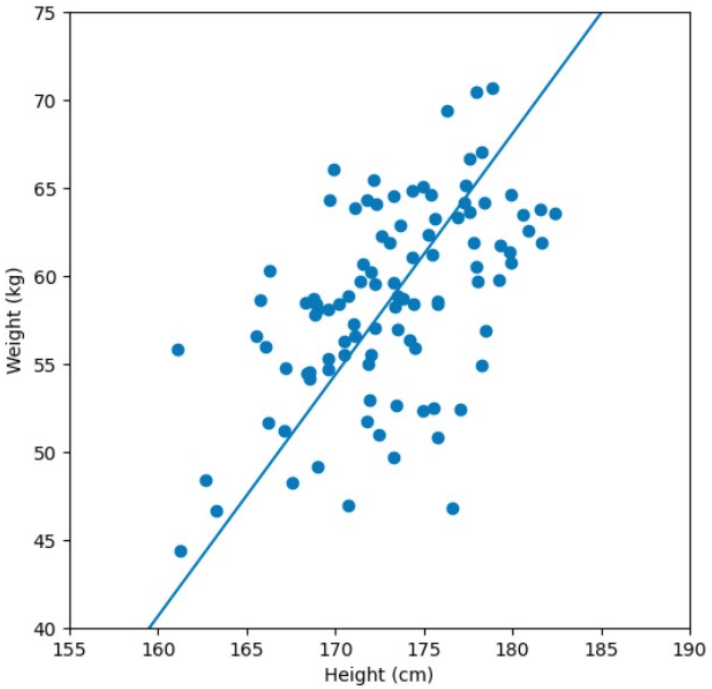
Population genetics analysis for aDNA data

Principal Component Analysis (PCA)

- PCA is a linear transformation to a new coordinate system
- **Reduction of dimensions**: the genetic information contained in 1M SNPs can be summarized by a few new variables



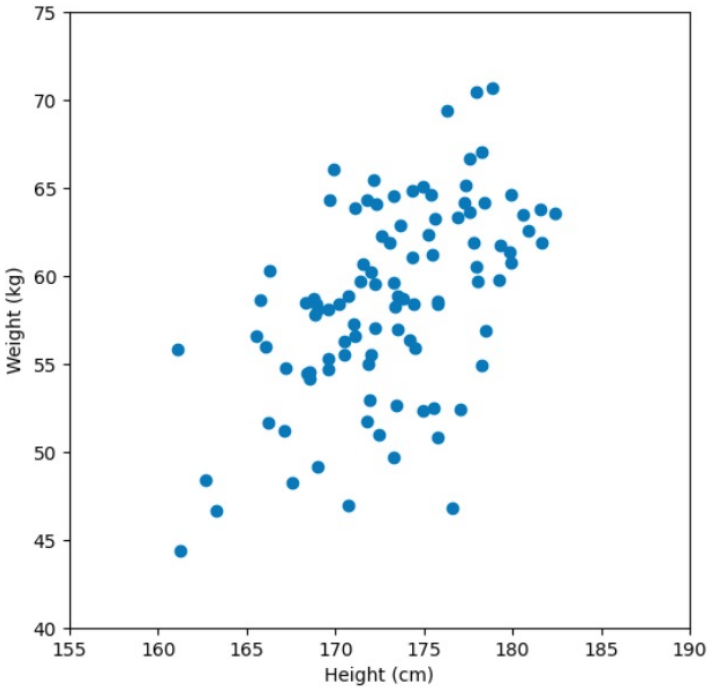
Each individual (point) is represented by two variables.



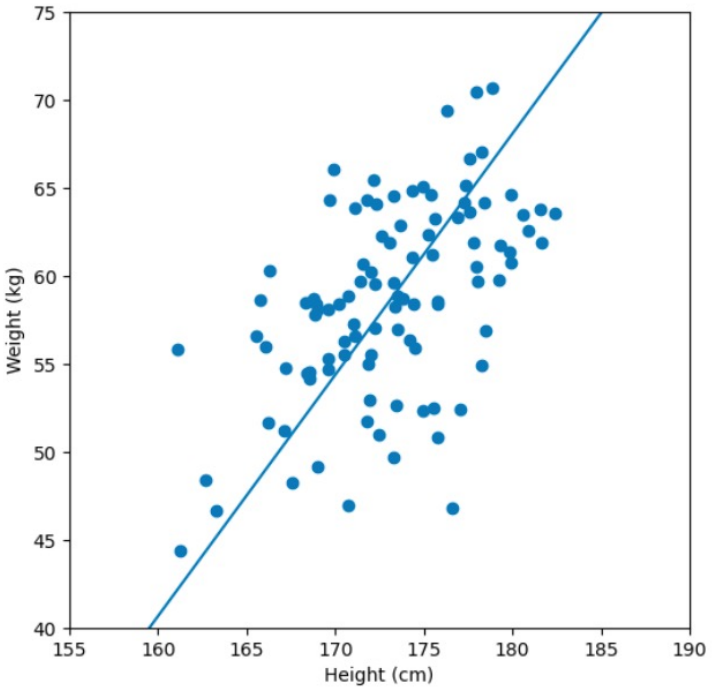
Find the axis of greatest variation (fit line) —> The principal component.

Principal Component Analysis (PCA)

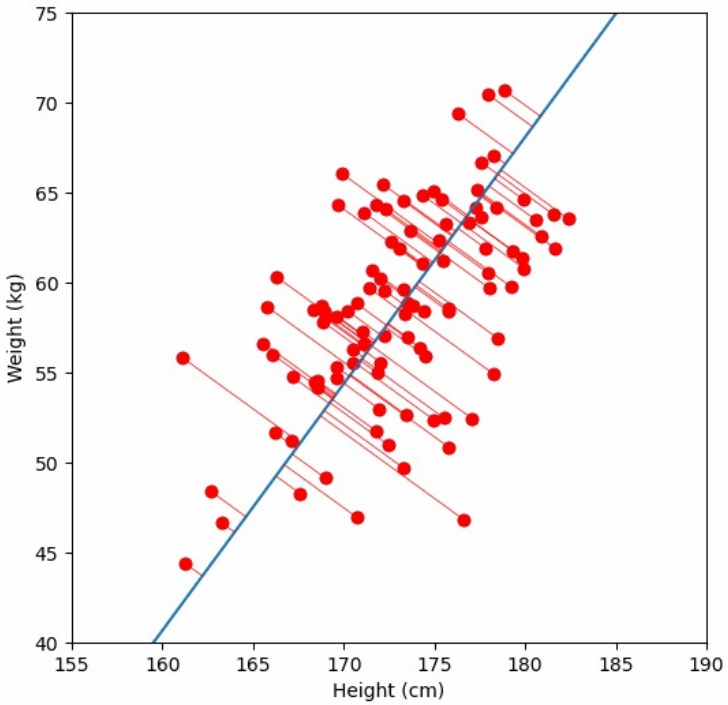
- PCA is a linear transformation to a new coordinate system
- **Reduction of dimensions**: the genetic information contained in 1M SNPs can be summarized by a few new variables



Each individual (point) is represented by two variables.



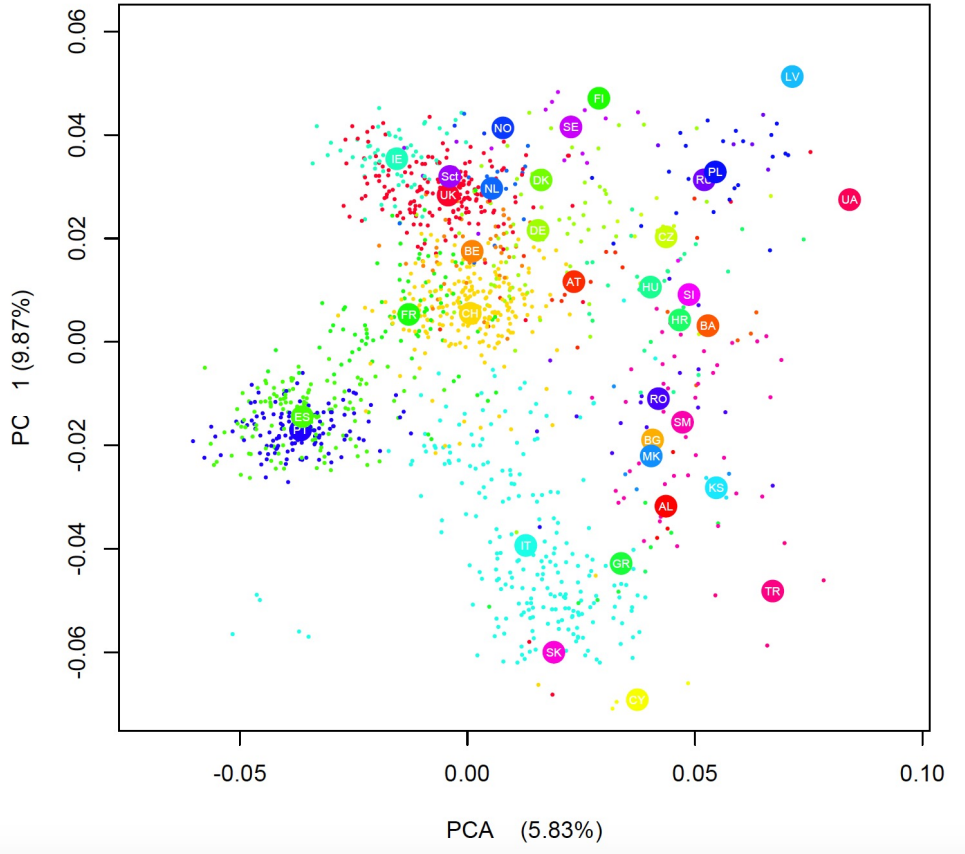
Find the axis of greatest variation (fit line) —> The principal component.



“Project” each point onto the line. Now each individual is represented by one variable.

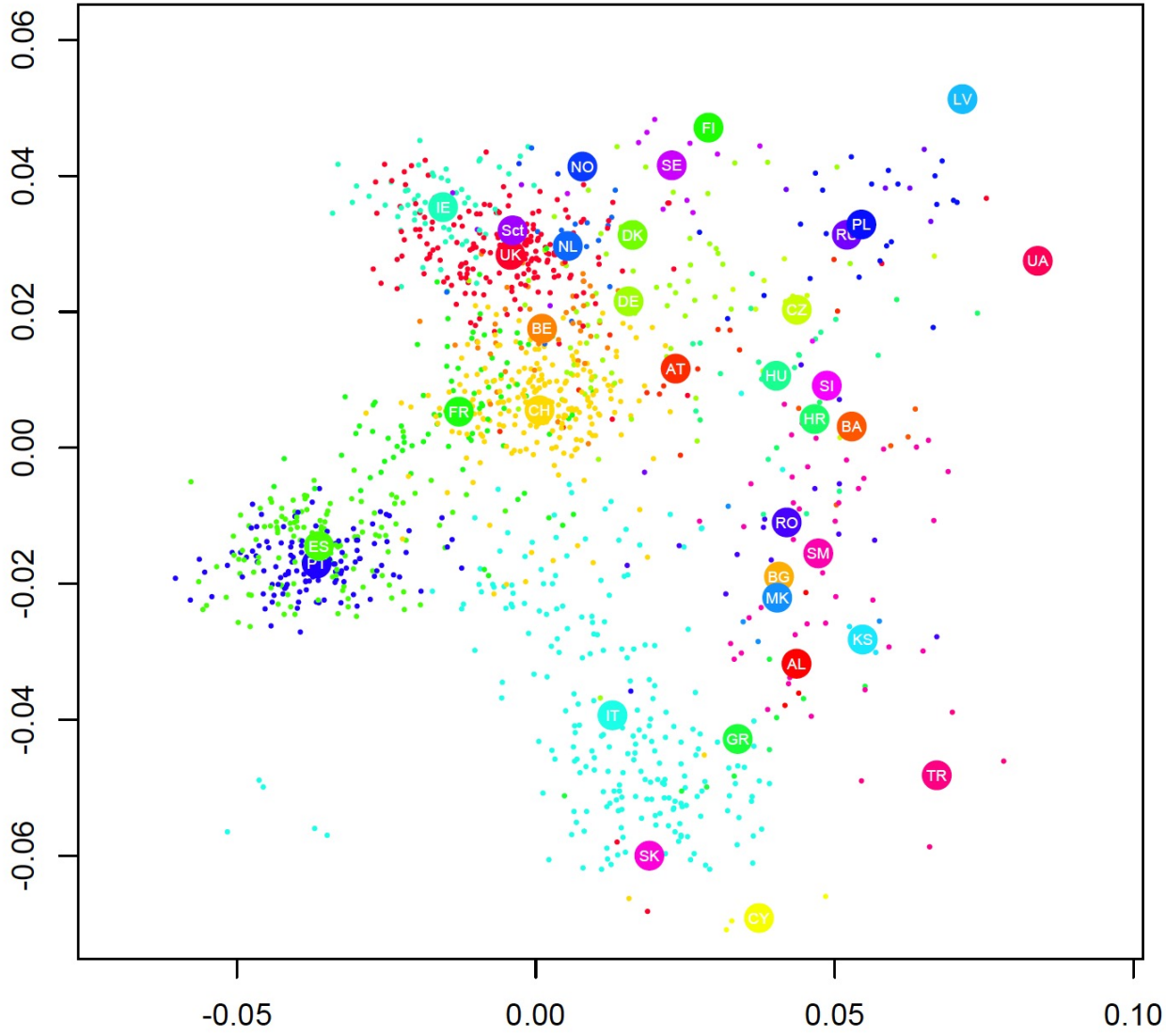
Principal Component Analysis (PCA)

	PC1	PC2
Ind(1): 0101110110101110	0.01	-0.02
Ind(2): 0111110110101111	0.50	0.03
Ind(3): 0100110110101011	0.07	-0.13
Ind(4): 0111111111101111	0.02	-0.04
Ind(5): 0101110110100001	0.01	-0.05
.	.	
.	.	
.	.	
Ind(n): 0101110110101111	-0.03	0.03



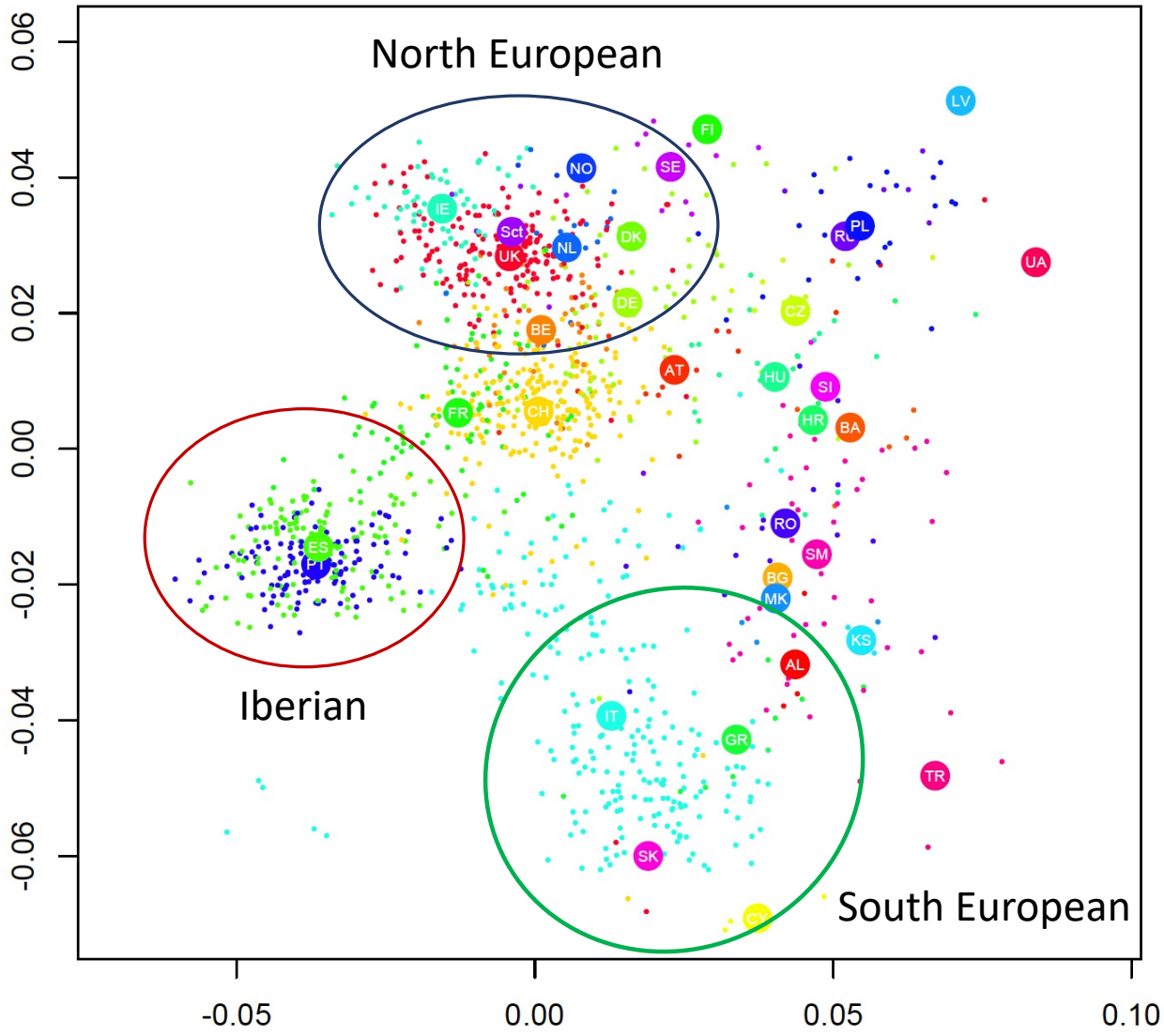
Principal Component Analysis (PCA)

- PCA reveal population structure
- Genetic Distance ≈ Physical distance
- Easily identify genetic outliers and isolated populations



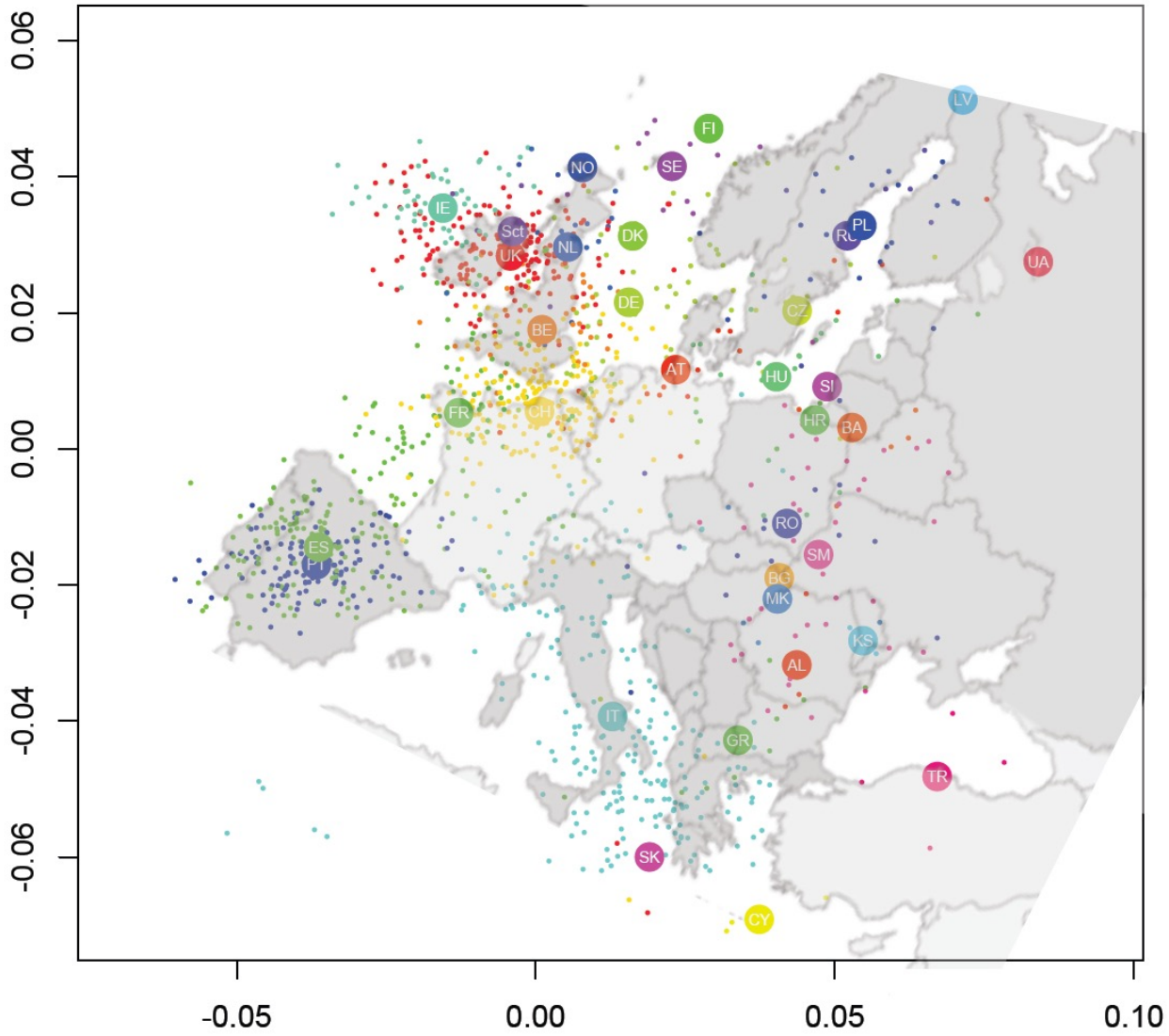
Principal Component Analysis (PCA)

- PCA reveal population structure
- Genetic Distance ≈ Physical distance
- Easily identify genetic outliers and isolated populations



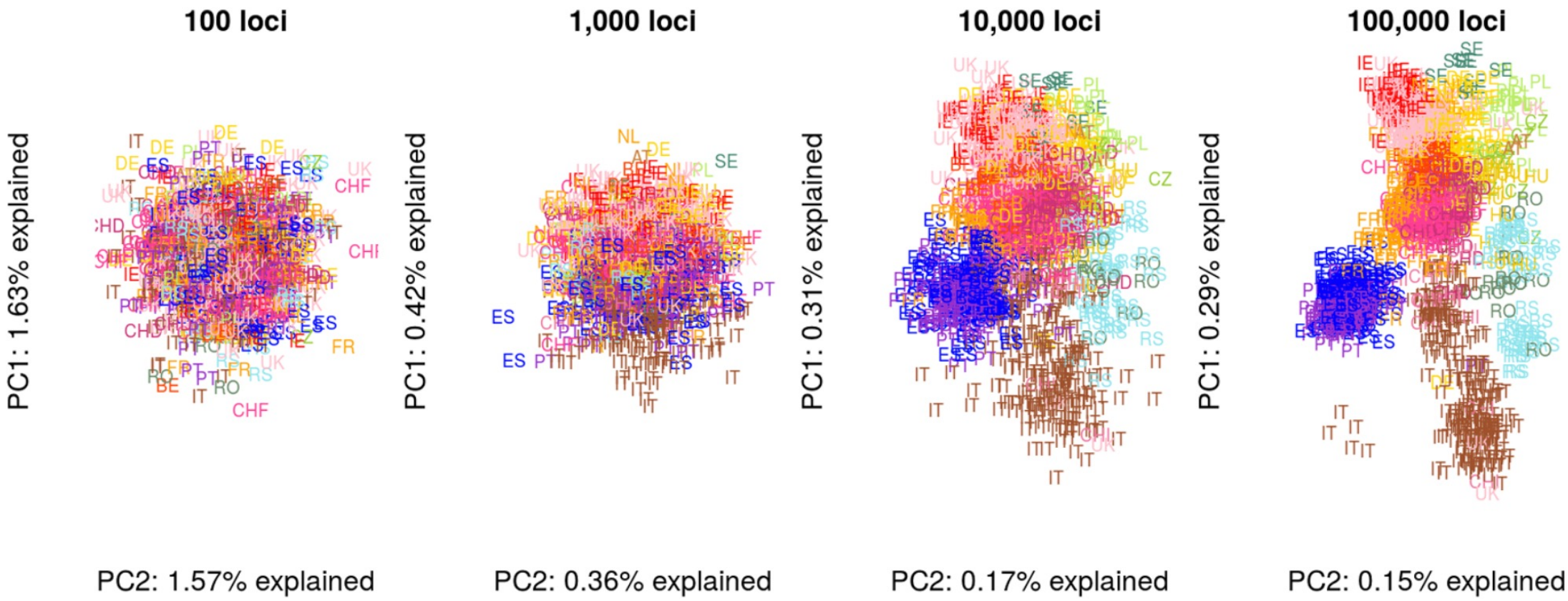
Principal Component Analysis (PCA)

- PCA reveal population structure
- Genetic Distance ≈ Physical distance
- Easily identify genetic outliers and isolated populations



Principal Component Analysis (PCA)

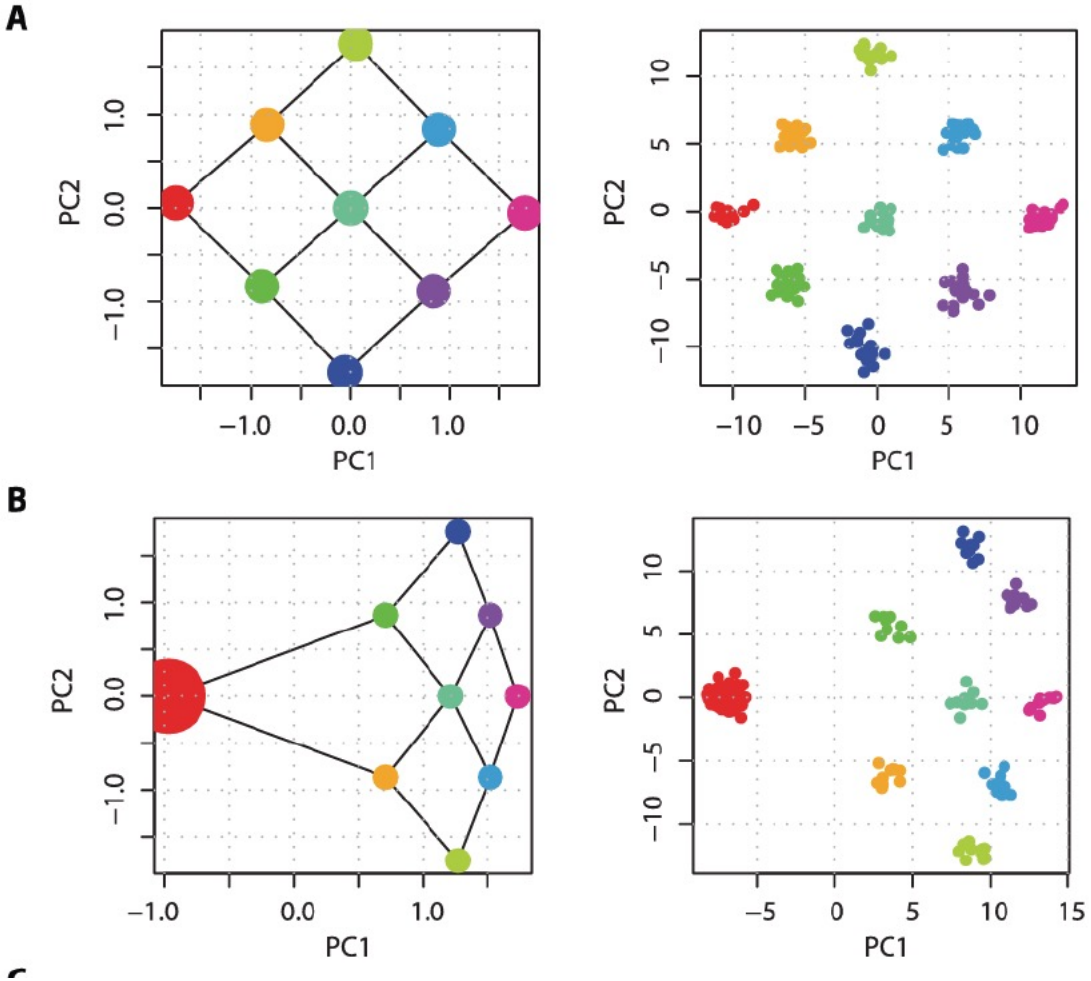
Produce good results even when the information is low



Principal Component Analysis (PCA)

Factors that influence PCA:

- Migration
- Genetic drift
- Admixture
- Population size
- SNP selection



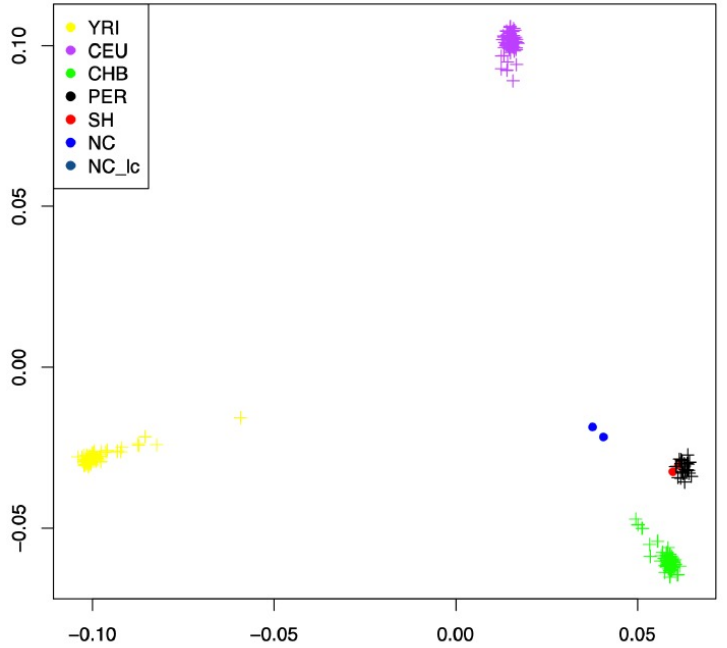
PCA with ancient samples

Low coverage individuals result in many SNPs with missing data

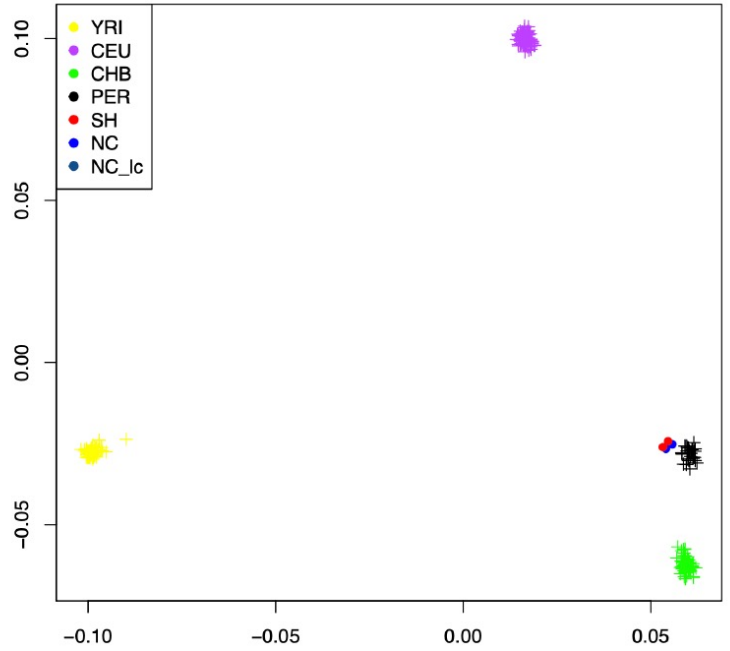
Usually, PCA methods will fill in all missing data. This results in PCA plots that have ancient individuals near/at the origin (0,0 coordinate).

Solution: Projection of ancient individuals.
We can infer eigenvectors using the reference set and then project ancient individuals onto those eigenvectors.

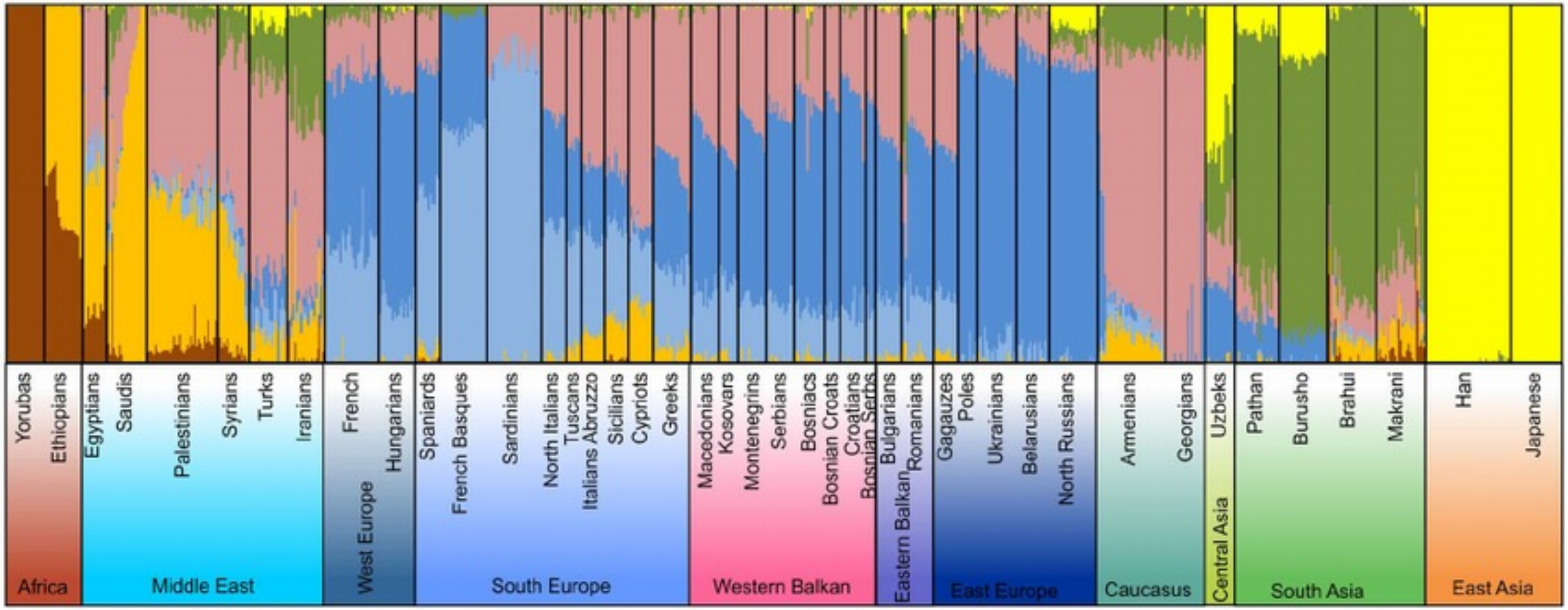
Not projected



Projected



Ancestry proportion inference (ADMIXTURE)



Ancestry proportion inference (ADMIXTURE)

Assumes that there are *k* ancestral populations and that the individuals included in the analysis have ancestry from those ancestral populations.

K = 3 ancestral population

Red

Blue

Yellow

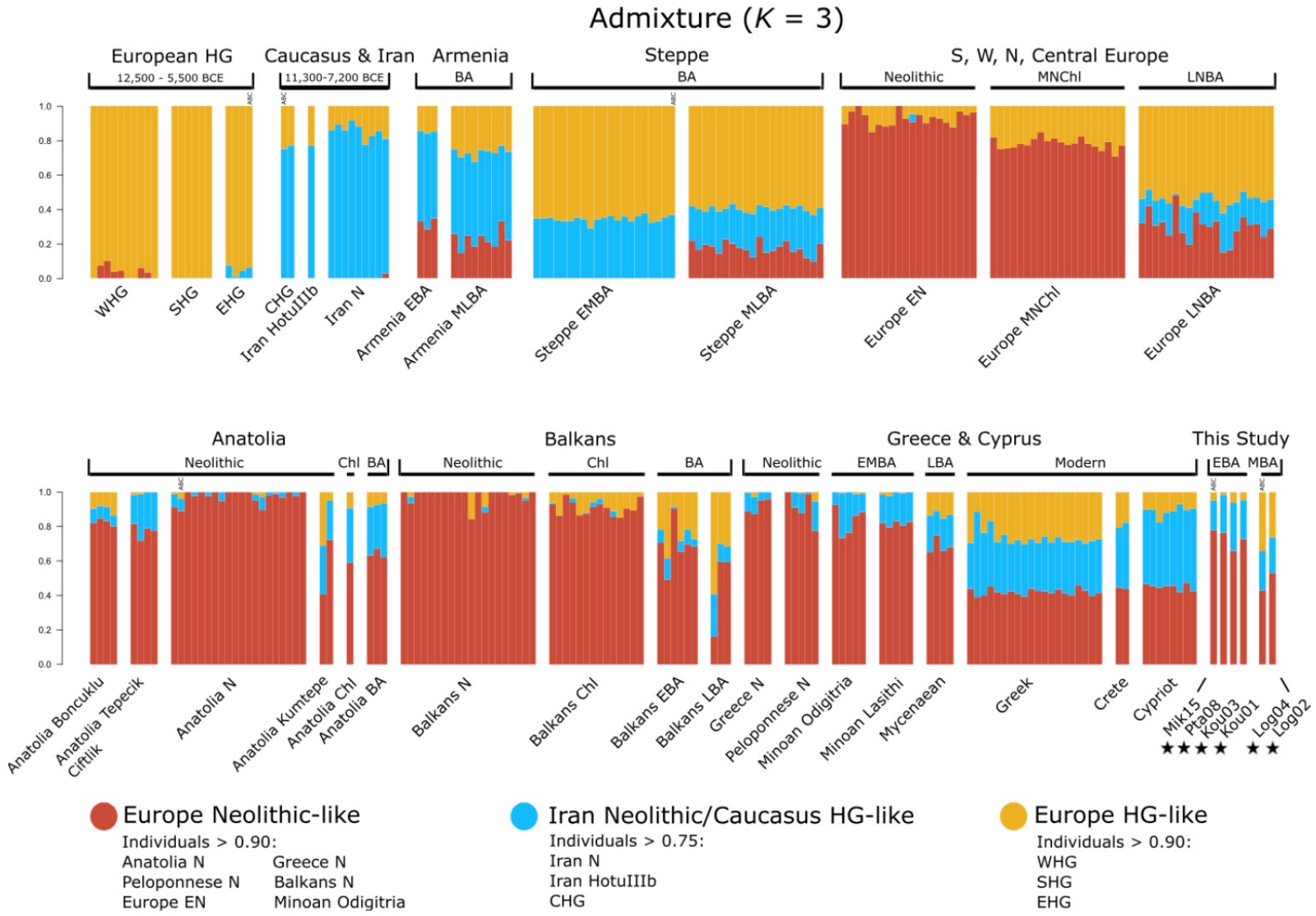
Allele frequences

	Red	Blue	Yellow
SNP1	0.1	0.5	0
SNP2	0.3	0.3	0.2
SNP3	0.6	0.1	0.1
SNP4	0.1	0.2	0.5
SNP5	0.7	0.3	0.9

Admixed individual

= 0.5 + 0.4 + 0.1

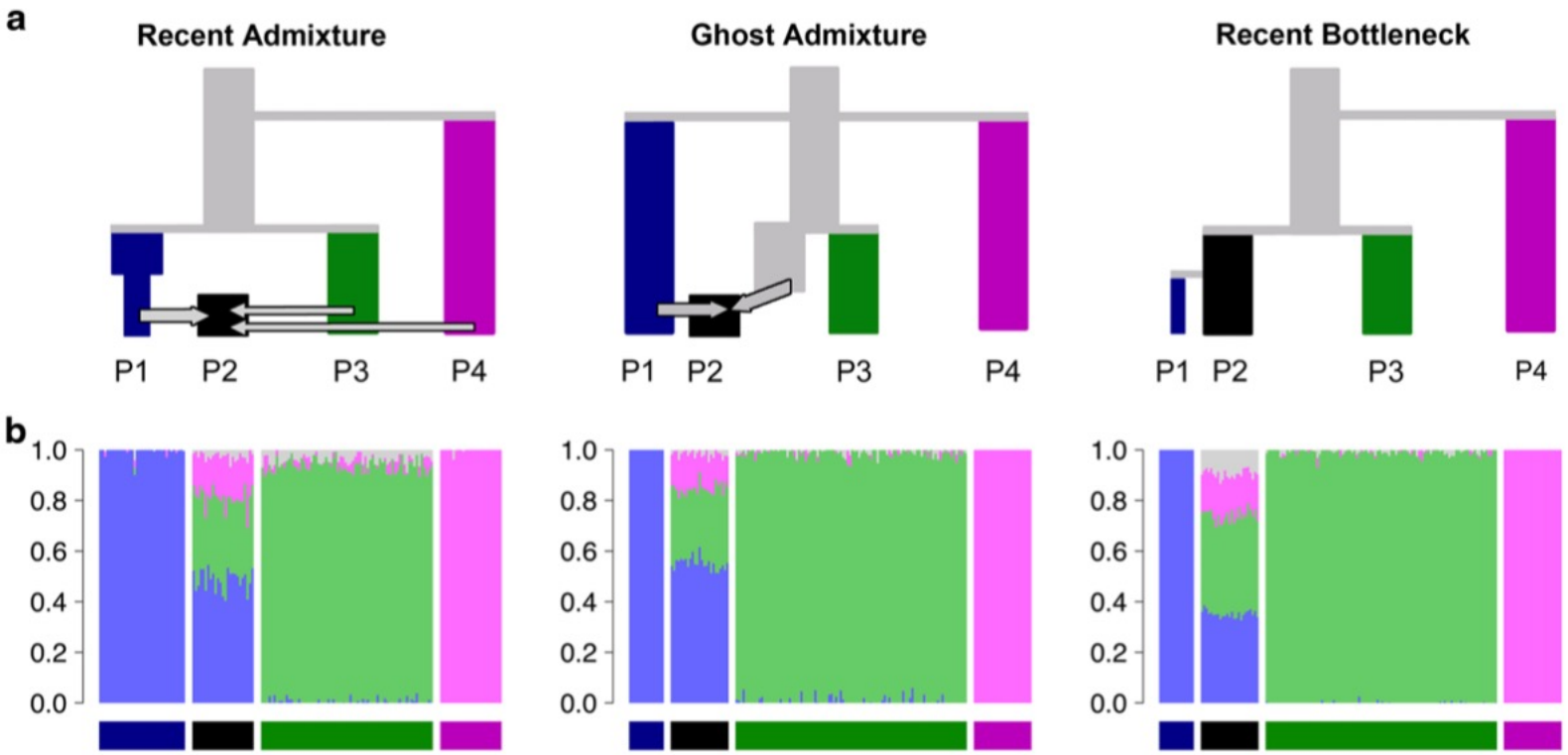
Ancestry proportion inference (ADMIXTURE)



Clemente et al. 2021 *Cell*

Ancestry proportion inference (ADMIXTURE)

Be careful when interpreting ADMIXTURE results!



Ancestry proportion inference (ADMIXTURE)

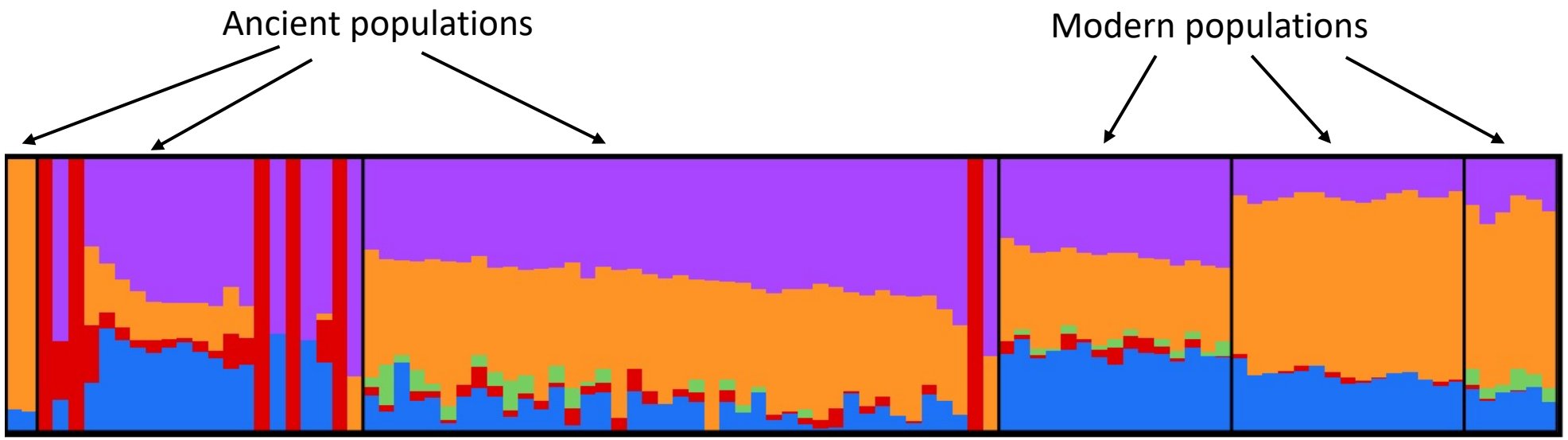
Be careful when interpreting ADMIXTURE results!

In this case, clustering will be the same as that for discrete populations

Ancestry proportion inference (ADMIXTURE)

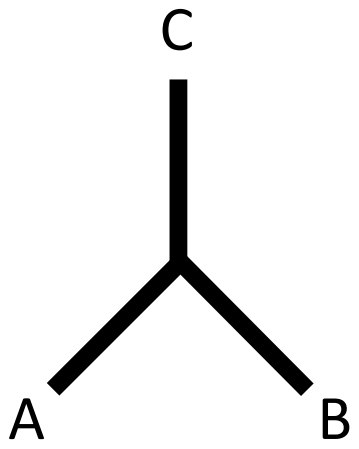
Be careful when interpreting ADMIXTURE results!

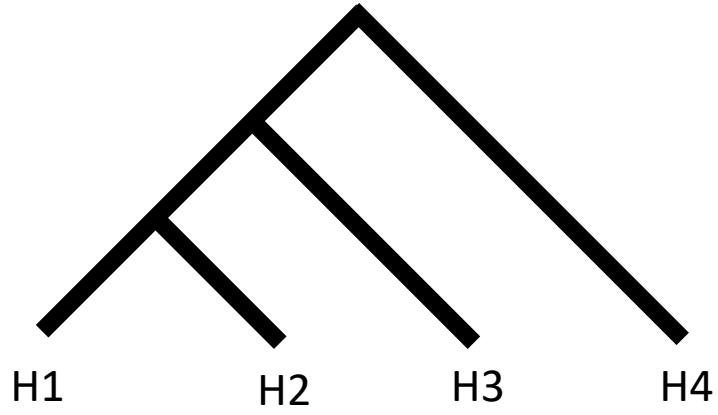
Possible problem with low coverage samples



Tests of "treeness" – *f* and Patterson's *D* statistics

- Testing if a tree of population is correct
- Identify admixture and gene flow
- Simple to analyse
- Results easy to interpret
- Statistically robust even with a small number of loci
- Ideal for aDNA data





Tests of "treeness" – *f* and Patterson's *D* statistics

Method	Applications	Test of Significance	Limitations	References
f_3	Test of whether a target population is admixed; measurement of shared ancestry in two populations; allele frequencies or sequence data	Weighted-block jackknife	Large drift in the admixed population may mask the signal of its admixture; putative genetic donor population may be incorrectly identified if it is closely related to the true donor	Reich et al. (2009) applied f_3 to characterize admixture in Indian populations; Raghavan et al. (2014) used outgroup f_3 to quantify the Western Eurasian-Siberian ancestry of Native Americans; Peter (2016) redefined f_3 in terms of coalescence times
f_4	Test of treeness of four species; quantification of admixture proportion; inferring the number of admixture events; allele frequencies or sequence data	Weighted-block jackknife	f_4 can be zero, suggesting no admixture, if the target admixed population descends equally from two donors	Reich et al. (2009) used f_4 to identify and quantify admixture proportions in Indian populations; Reich et al. (2012) demonstrated that Native American population history is consistent with at least three migrations from East Asia using f_4 (as qpWave)
Patterson's *D*	Model-based test for introgression between candidate populations; sequence data or allele frequencies	Weighted-block jackknife	Results do not imply a direction of gene flow; method cannot distinguish between ancestral population structure and introgression; ability to infer significance depends on the number of informative sites available; can be misled by contamination (also applies to and other *D*-statistics)	Used by Green et al. (2010) to support the hypothesis that Neanderthals interbred with non-African humans

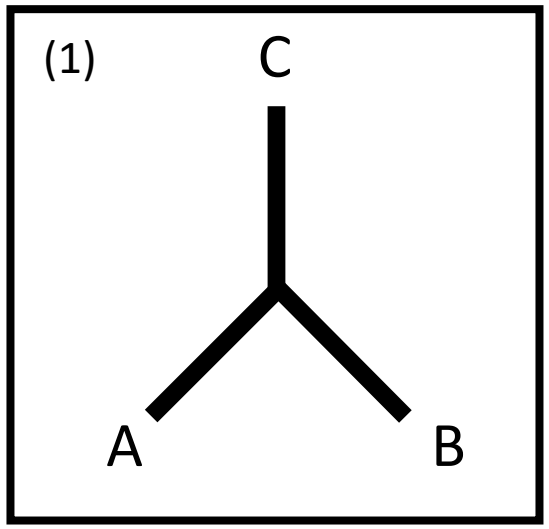
f3 statistic

$$f3(C;A,B) = \frac{1}{J}\sum_{j=1}^{J}(c_j - a_j)(c_j - b_j)$$

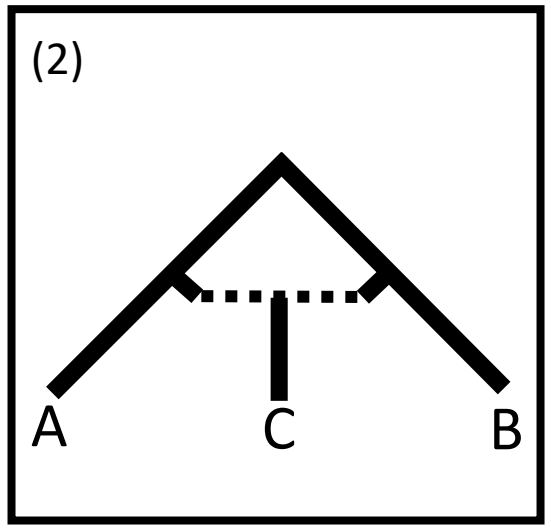
Two main purposes:

- Measuring how much two populations are similar with respect to an outgroup (1)
- Testing if a population is the result of an admixture between the other two populations (2)

f3 > 0



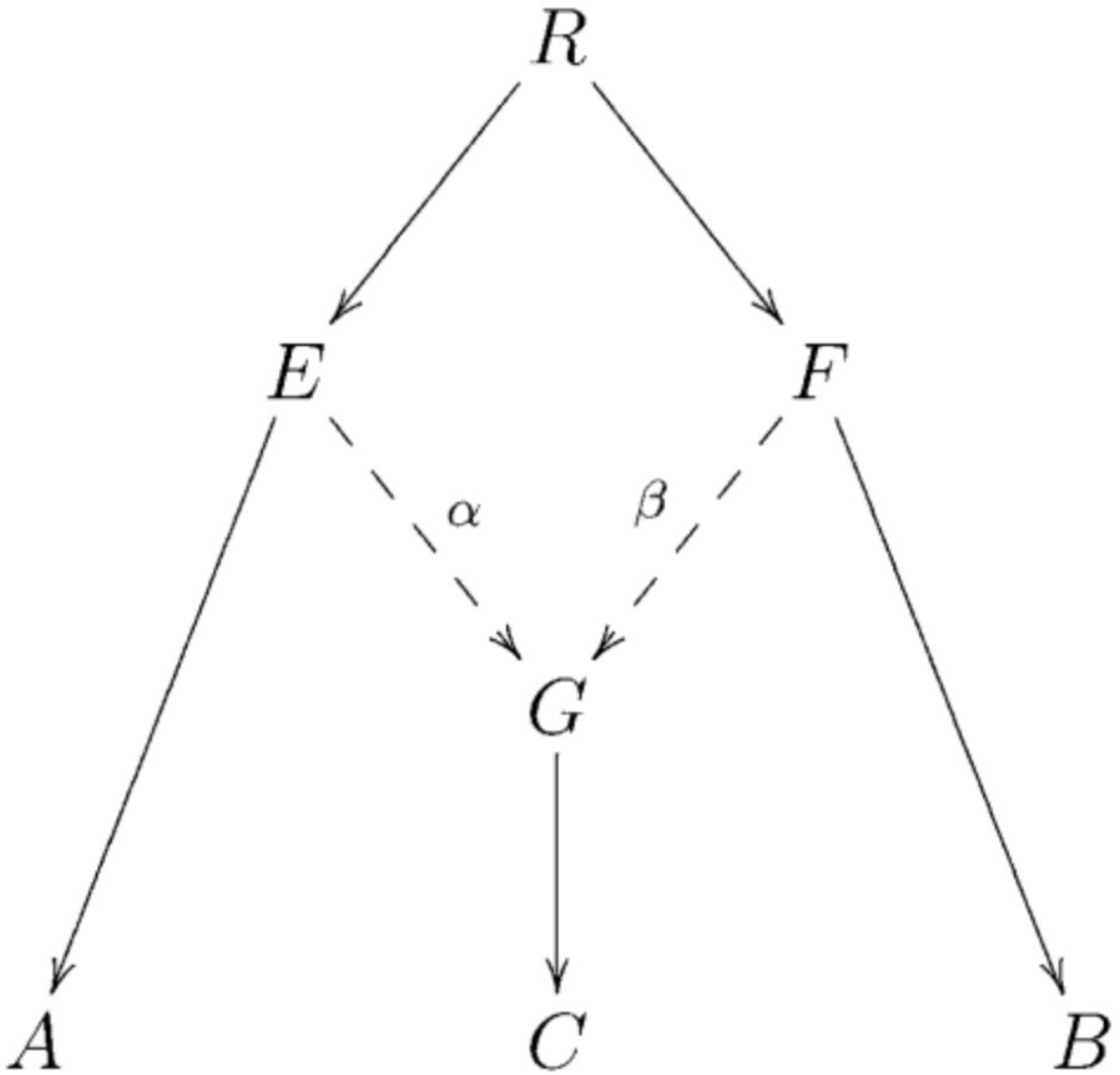
f3 < 0



f3 statistic

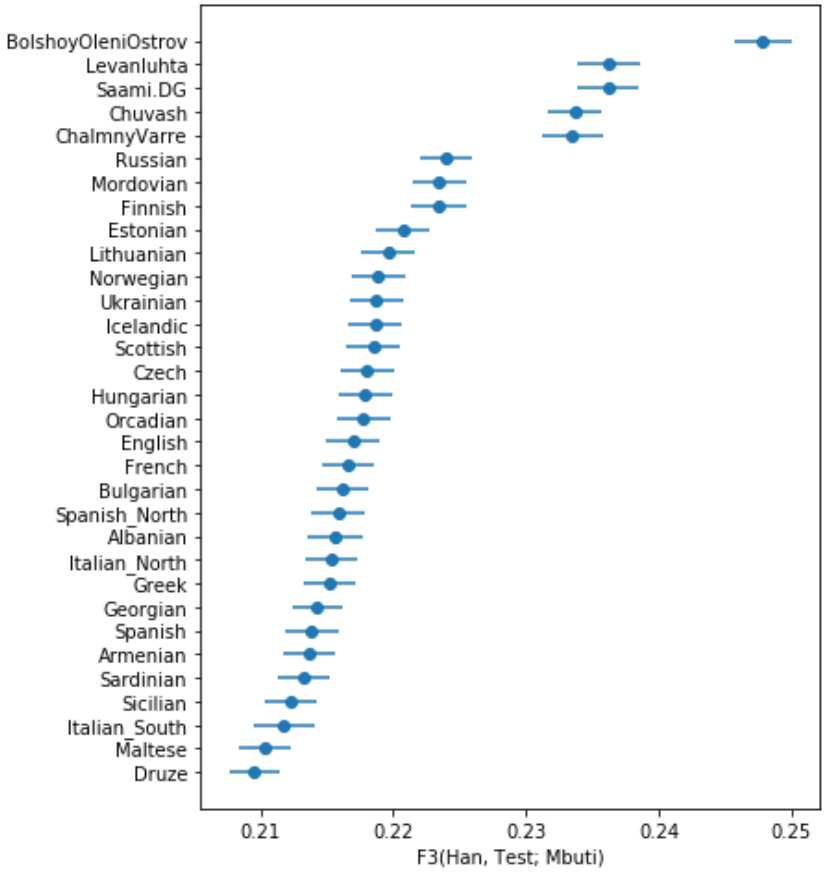
$$f3(C;A,B) = \frac{1}{J}\sum_{j=1}^{J}(c_j - a_j)(c_j - b_j)$$

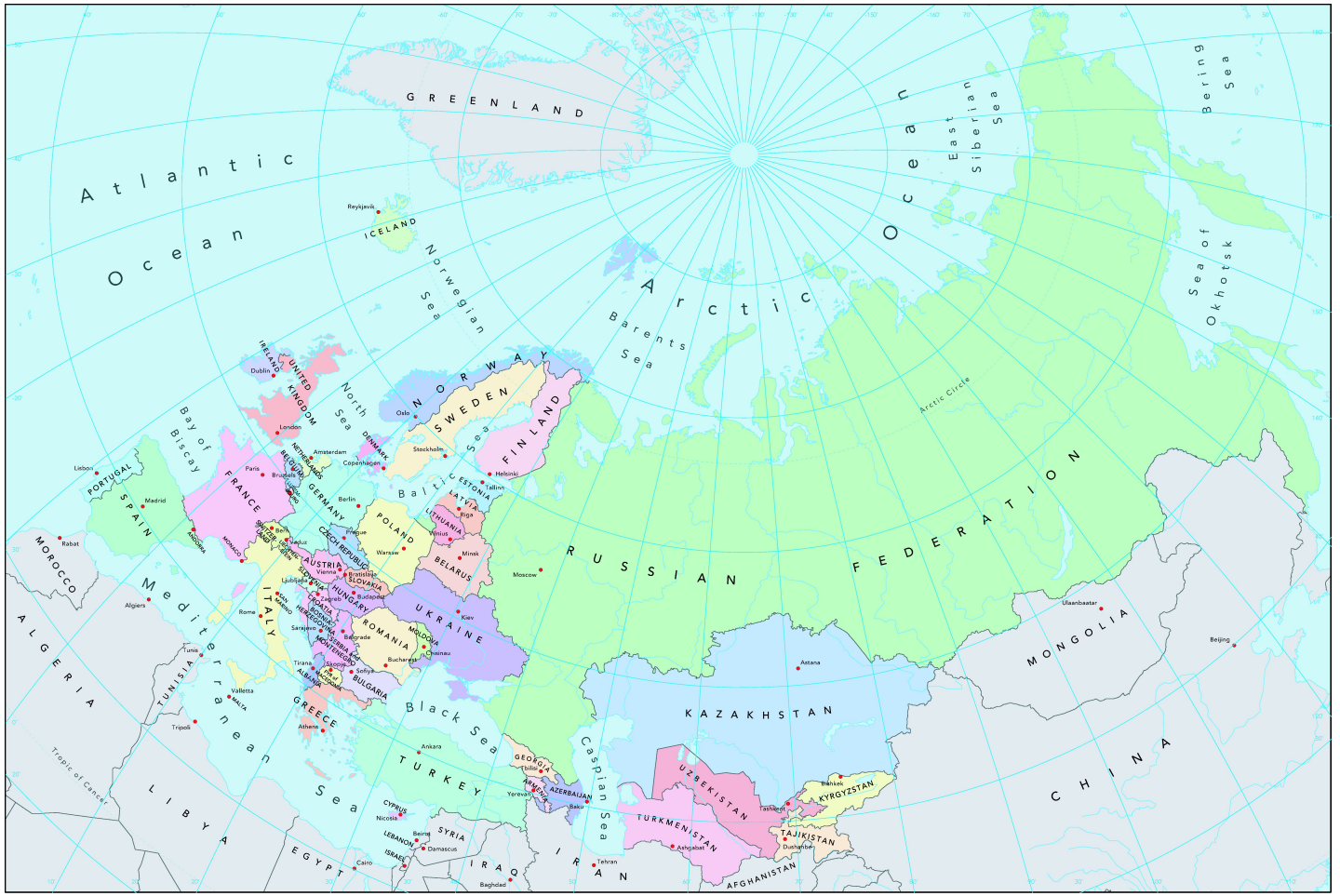
f3 < 0



Outgroup *f3* statistic – Example

Goal: We want to test the genetic affinity of European populations to East Asia, by performing the statistic ***f3*(Han, X; Mbuti),** where Mbuti is a distant African population and acts as outgroup here, Han denote Han Chinese, and X denotes various European populations





Target *f3* statistic – Example

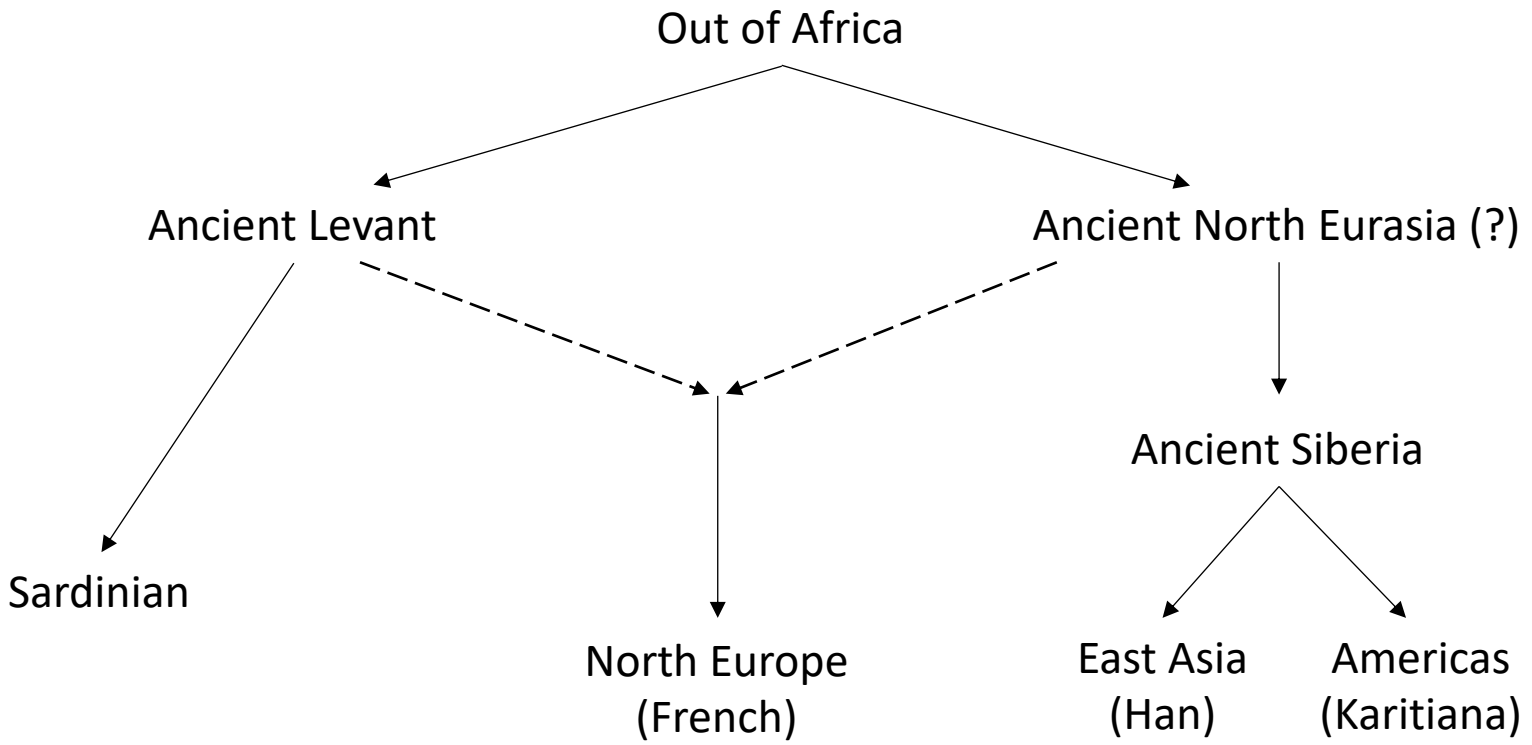
We can use target *f3* to better understand what is the genetic relationship between East Asia and Europe

Source1	Source2	Target	f_3	*Z*-score
Japanese	Italian	Uygur	−0.0259	−74.79
Japanese	Italian	Hazara	−0.0230	−74.05
Yoruba	Sardinian	Mozabite	−0.0211	−56.95
Mozabite	Surui	Maya	−0.0149	−19.67
Yoruba	San	Bantu-SA	−0.0107	−31.39
Yoruba	Sardinian	Palestinian	−0.0107	−36.70
Yoruba	Sardinian	Bedouin	−0.0104	−33.73
Druze	Yi	Burusho	−0.0090	−27.62
Sardinian	Karitiana	Russian	−0.0086	−20.68
Druze	Karitiana	Pathan	−0.0084	−22.25
Han	Orcadian	Tu	−0.0076	−20.64
Mbuti	Orcadian	Makrani	−0.0076	−19.56
Han	Orcadian	Mongola	−0.0075	−19.21
Han	French	Xibo	−0.0069	−16.92
Druze	Dai	Sindhi	−0.0067	−21.99
Sardinian	Karitiana	French	−0.0060	−18.36
Dai	Italian	Cambodian	−0.0060	−13.16
Sardinian	Karitiana	Adygei	−0.0057	−13.03
Biaka	Sardinian	Bantu-Kenya	−0.0054	−13.42
Sardinian	Karitiana	Tuscan	−0.0052	−11.26
Sardinian	Pima	Italian	−0.0045	−12.48
Druze	Karitiana	Balochi	−0.0044	−11.58
Daur	Dai	Han	−0.0026	−13.20
Han	Orcadian	Han-NChina	−0.0025	−7.09
Han	Yakut	Daur	−0.0025	−9.05
Druze	Karitiana	Brahui	−0.0025	−6.43
Hezhen	Dai	Tujia	−0.0021	−6.97
Sardinian	Karitiana	Orcadian	−0.0019	−4.31
She	Yakut	Oroqen	−0.0017	−5.13

Target *f3* statistic – Example

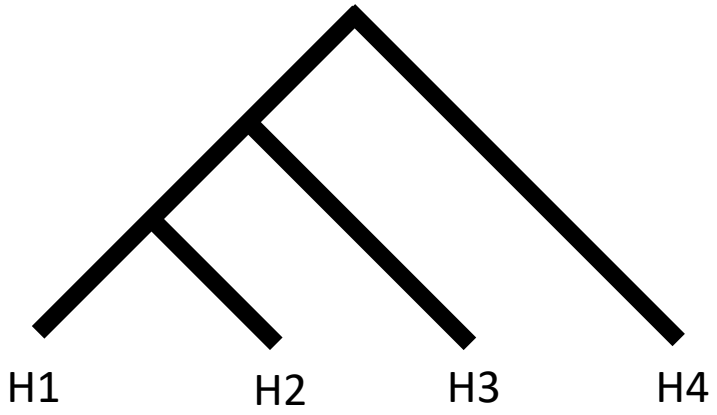
We can use target *f3* to better understand what is the genetic relationship between East Asia and Europe

Source1	Source2	Target	f_3	*Z*-score
Japanese	Italian	Uygur	−0.0259	−74.79
Japanese	Italian	Hazara	−0.0230	−74.05
Yoruba	Sardinian	Mozabite	−0.0211	−56.95
Mozabite	Surui	Maya	−0.0149	−19.67
Yoruba	San	Bantu-SA	−0.0107	−31.39
Yoruba	Sardinian	Palestinian	−0.0107	−36.70
Yoruba	Sardinian	Bedouin	−0.0104	−33.73
Druze	Yi	Burusho	−0.0090	−27.62
Sardinian	Karitiana	Russian	−0.0086	−20.68
Druze	Karitiana	Pathan	−0.0084	−22.25
Han	Orcadian	Tu	−0.0076	−20.64
Mbuti	Orcadian	Makrani	−0.0076	−19.56
Han	Orcadian	Mongola	−0.0075	−19.21
Han	French	Xibo	−0.0069	−16.92
Druze	Dai	Sindhi	−0.0067	−21.99
Sardinian	Karitiana	French	−0.0060	−18.36
Dai	Italian	Cambodian	−0.0060	−13.16
Sardinian	Karitiana	Adygei	−0.0057	−13.03
Biaka	Sardinian	Bantu-Kenya	−0.0054	−13.42
Sardinian	Karitiana	Tuscan	−0.0052	−11.26
Sardinian	Pima	Italian	−0.0045	−12.48
Druze	Karitiana	Balochi	−0.0044	−11.58
Daur	Dai	Han	−0.0026	−13.20
Han	Orcadian	Han-NChina	−0.0025	−7.09
Han	Yakut	Daur	−0.0025	−9.05
Druze	Karitiana	Brahui	−0.0025	−6.43
Hezhen	Dai	Tujia	−0.0021	−6.97
Sardinian	Karitiana	Orcadian	−0.0019	−4.31
She	Yakut	Oroqen	−0.0017	−5.13



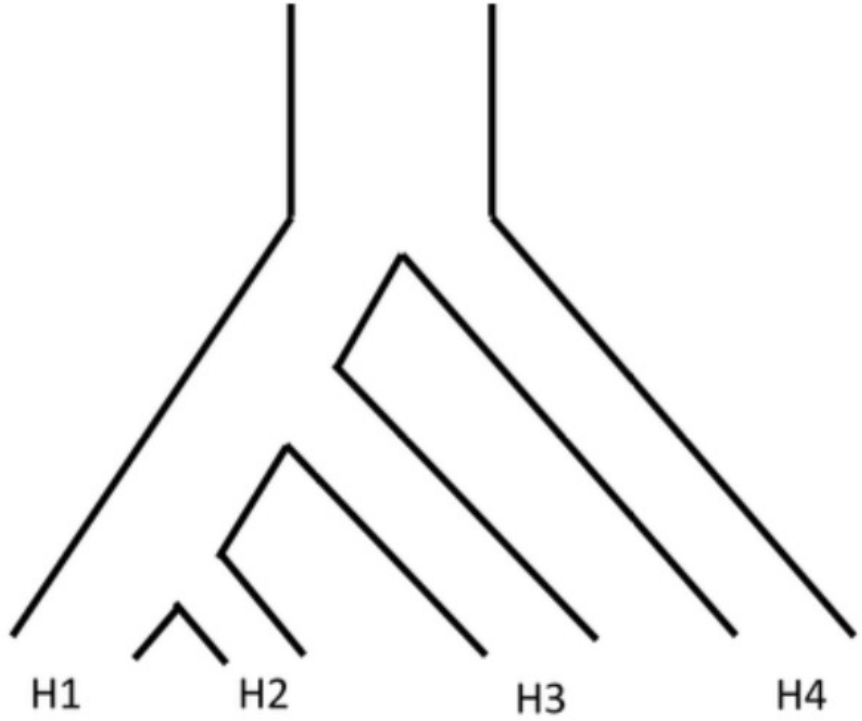
D statistic

- Detect signature of admixture between populations
- Identify the correct tree for a set of population



D statistic

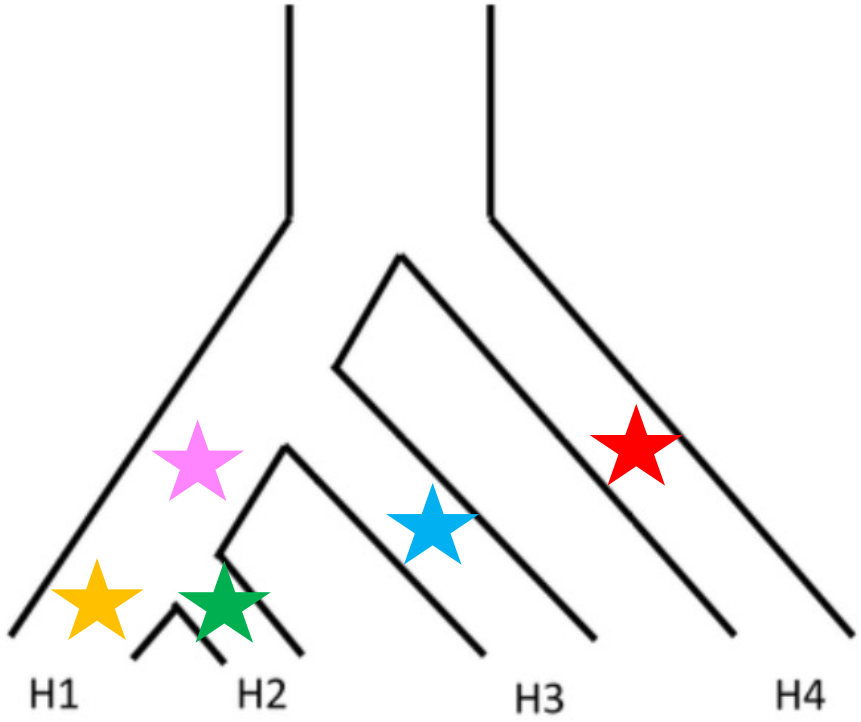
- Analyse a tree with four population
- Pick one individual for each population (it can be performed also with the whole population)
- Look at a polymorphic site – “A” is the ancestral state and “B” is the derived one
- Possible observable pattern of allele sharing



B	A	A	A
A	B	A	A
A	A	B	A
A	A	A	B
A	B	B	A
B	A	B	A
B	B	A	A

D statistic

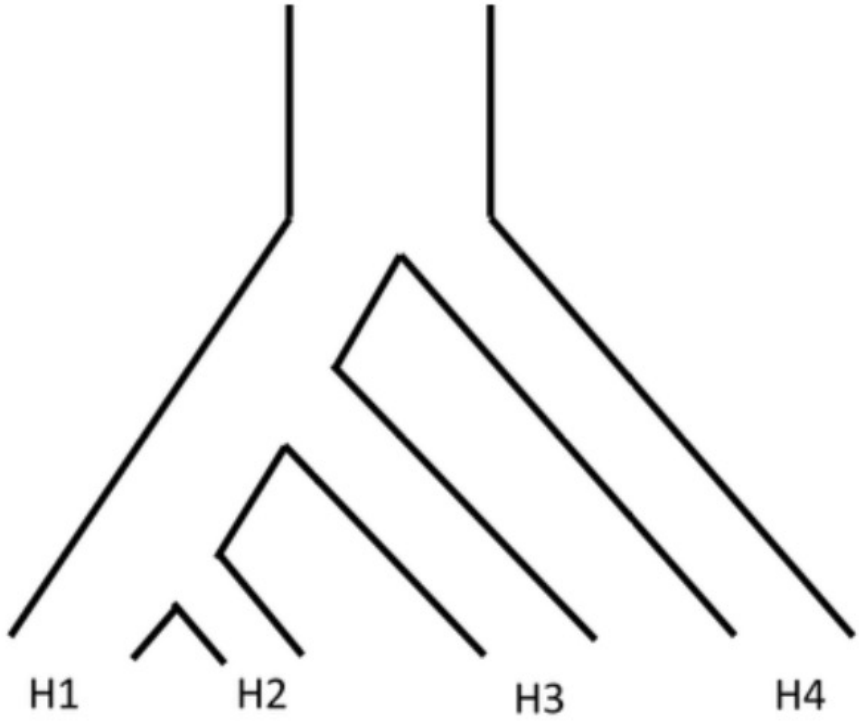
How to explain the patterns?



B	A	A	A
A	B	A	A
A	A	B	A
A	A	A	B
A	B	B	A
B	A	B	A
B	B	A	A

D statistic

How to explain the patterns?

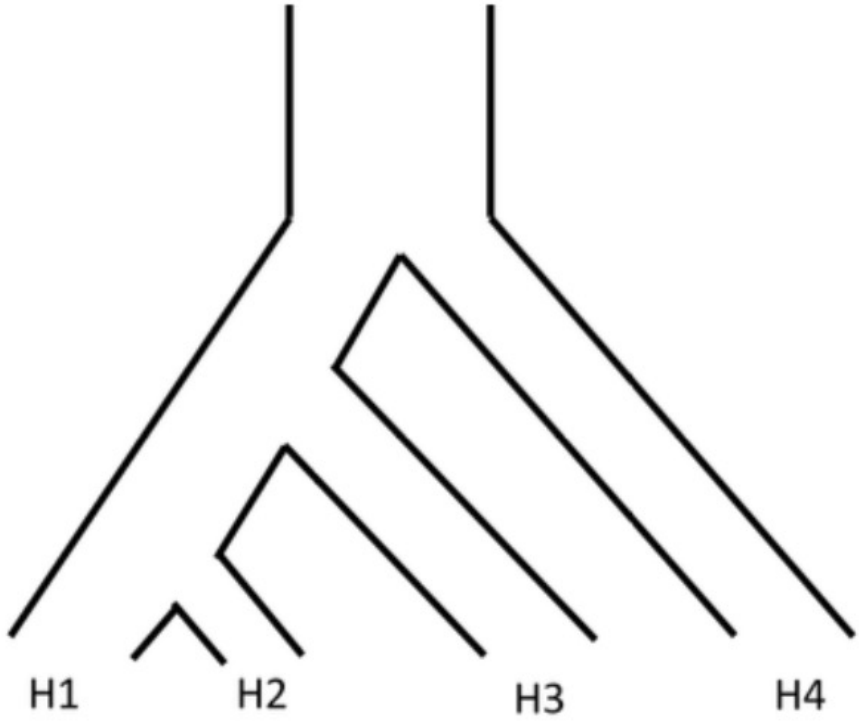


B	A	A	A
A	B	A	A
A	A	B	A
A	A	A	B
A	B	B	A
B	A	B	A
B	B	A	A

D statistic

How to explain the patterns?

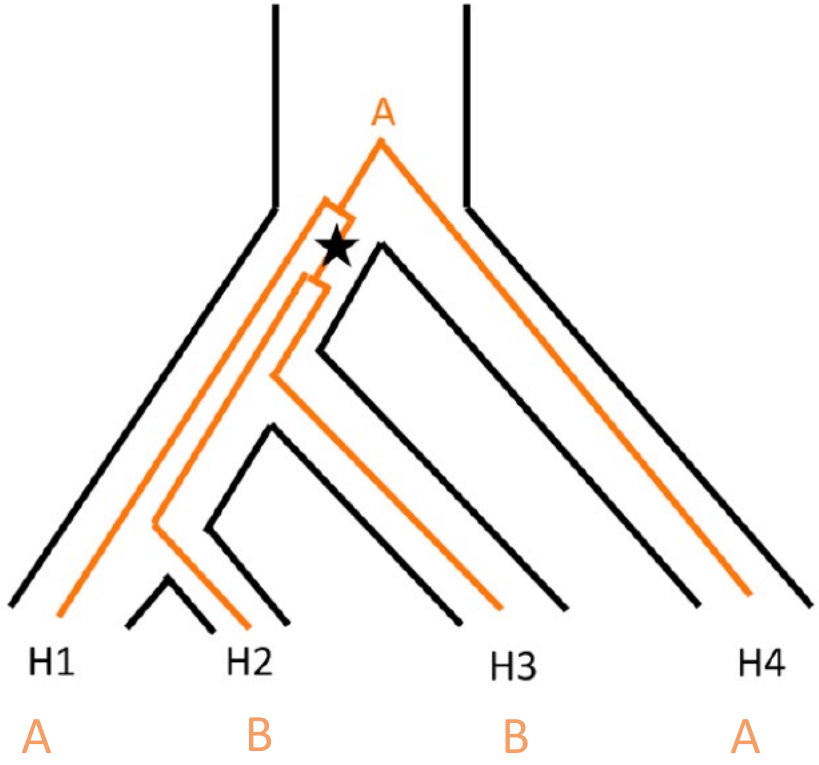
It is important to note that **gene genealogies** necessarily follow the **population tree**

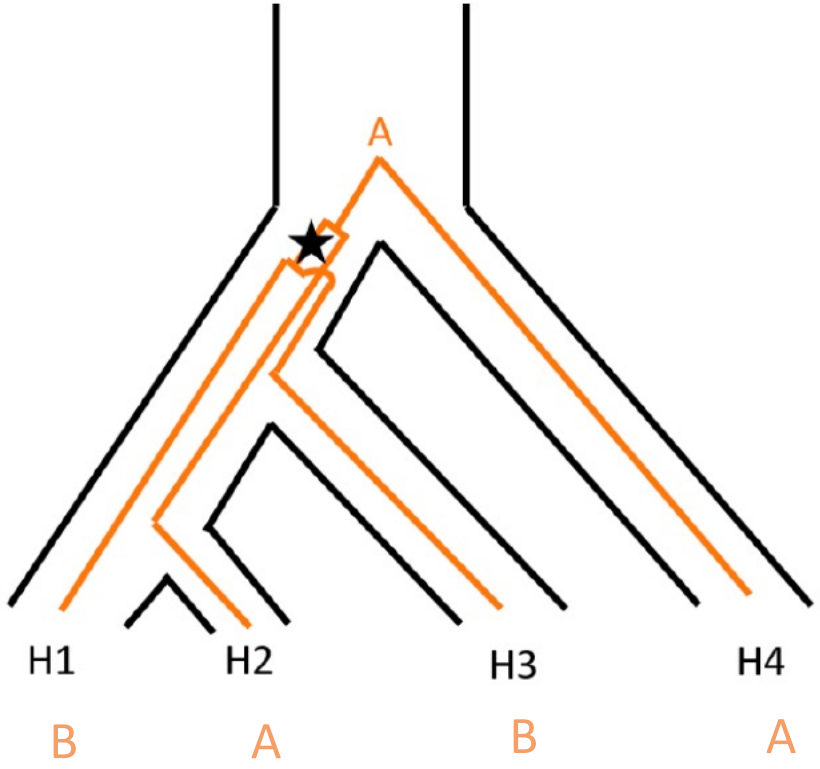


B	A	A	A
A	B	A	A
A	A	B	A
A	A	A	B
A	B	B	A
B	A	B	A
B	B	A	A

D statistic

ABBA and BABA sites

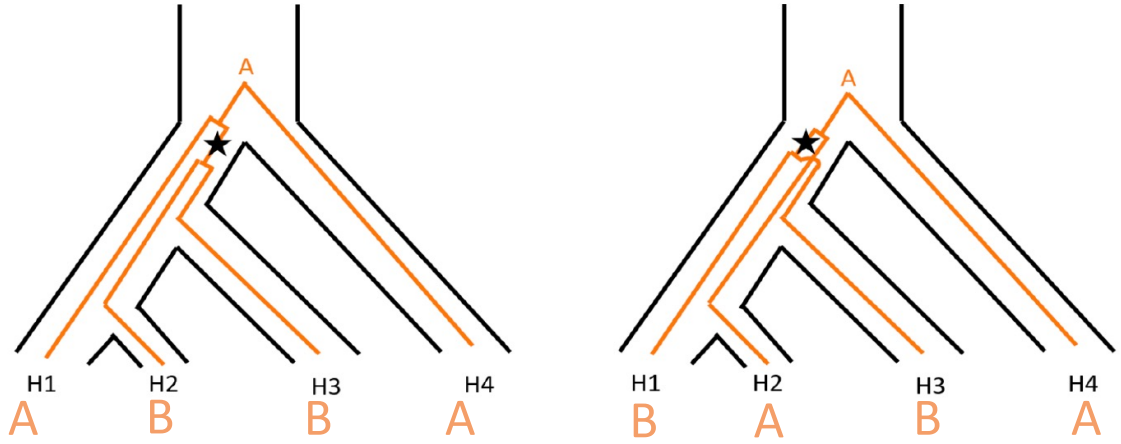




D statistic

D statistic is calculated in this way:

$$D(H_1, H_2; H_3, H_4) = \frac{(n_{ABBA} - n_{BABA})}{(n_{ABBA} + n_{BABA})}$$

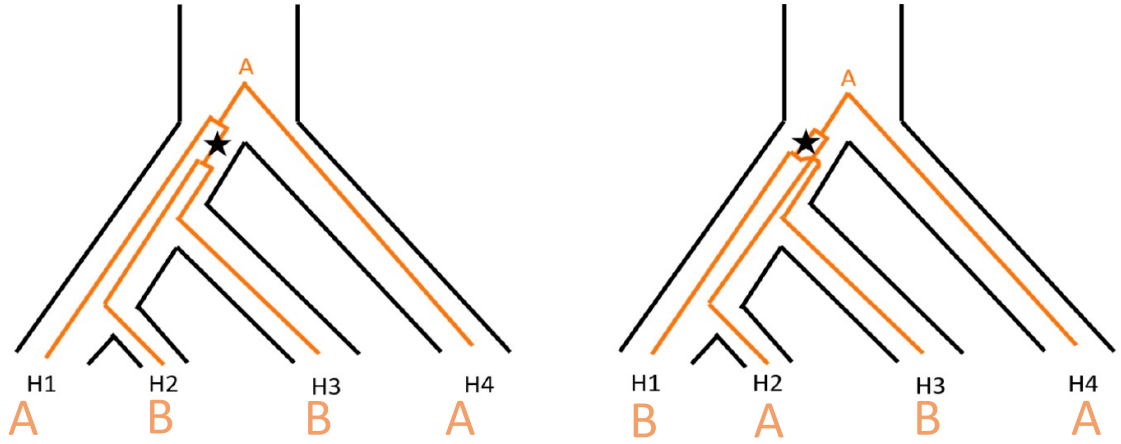


Using several (all) the loci in the genome

We are observing which pattern is the most frequent, ABBA or BABA

D statistic

$$D(H_1, H_2; H_3, H_4) = \frac{(n_{ABBA} - n_{BABA})}{(n_{ABBA} + n_{BABA})}$$

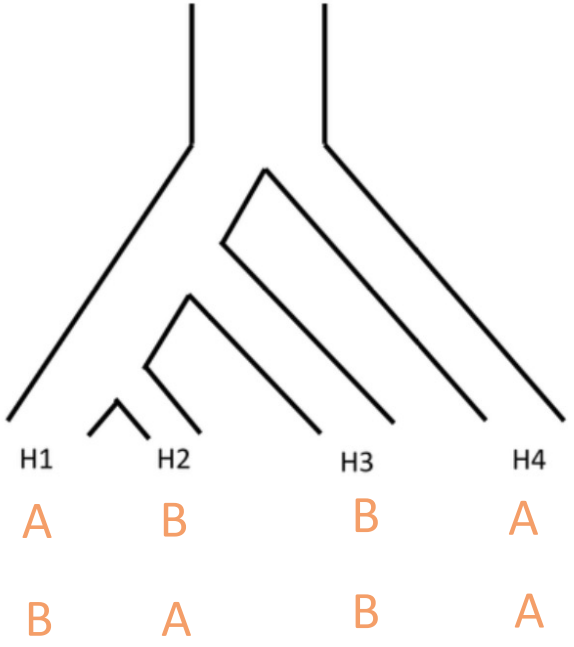


D = (1000-500)/(1000+500) = 0.33 D > 0 if ABBA is more common

D = (500-1000)/(500+1000) = -0.33 D < 0 if BABA is more common

Interpreting *D* statistic

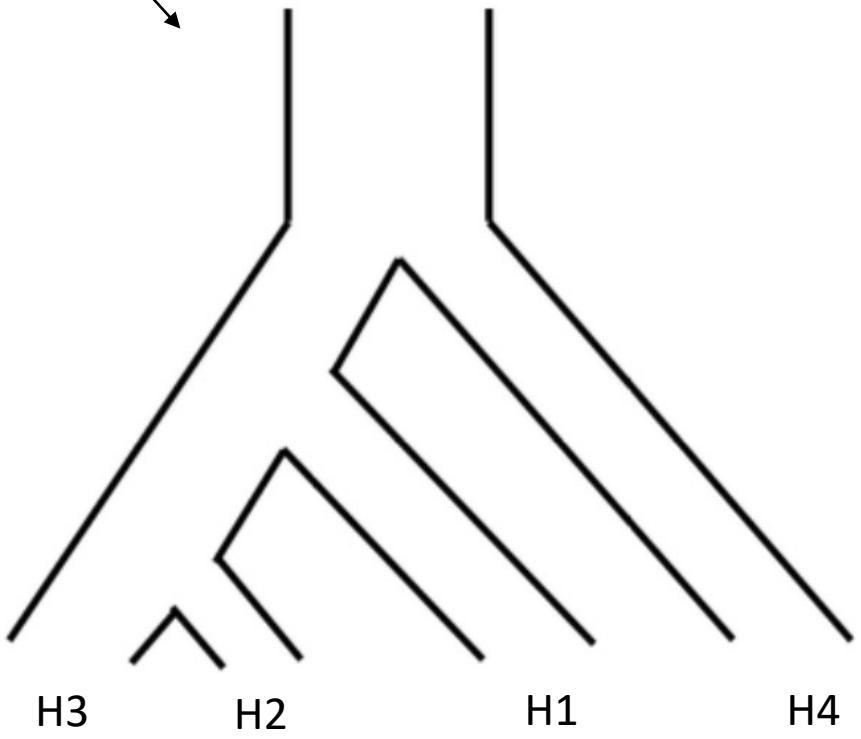
- If the tree is correct and there is no gene flow, we observe that most of the sites have a mutation pattern that is compatible with the tree (AAAA, BBAA, AABA, etc.)
- We can still observe some sites with the incompatible ABBA and BABA patterns, but there should be an approximately equal amount of each type of site.
- In this way, we should observe that *D* = 0 (or not significantly different from 0)



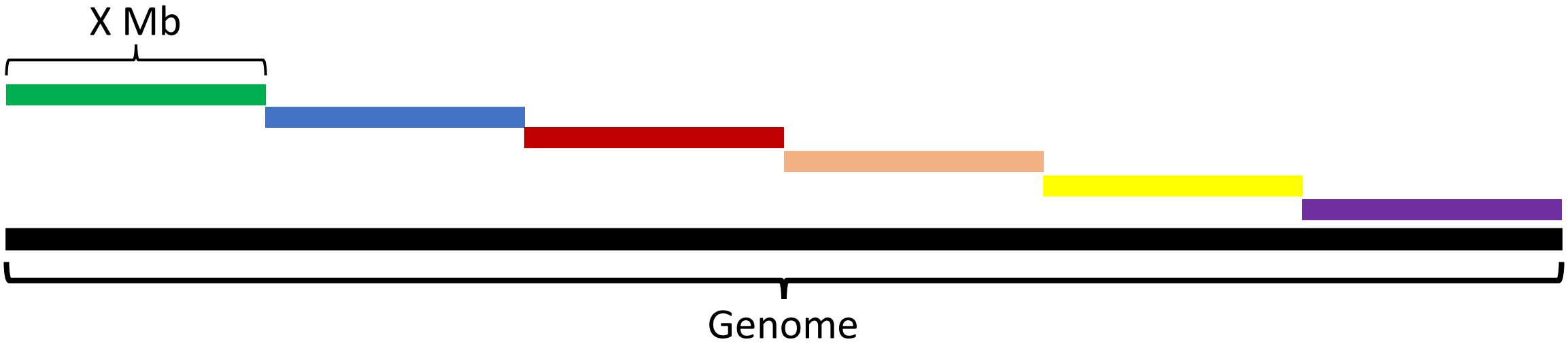
Interpreting *D* statistic

What if D ≠ 0?

- Gene flow
- The tree is not correct



Assessing the significance of the *D* statistic

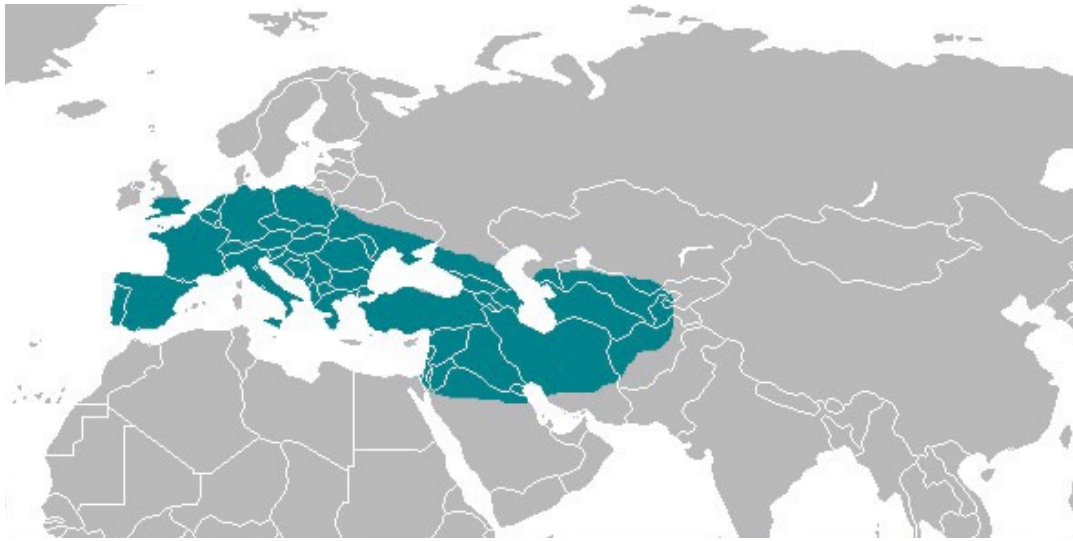


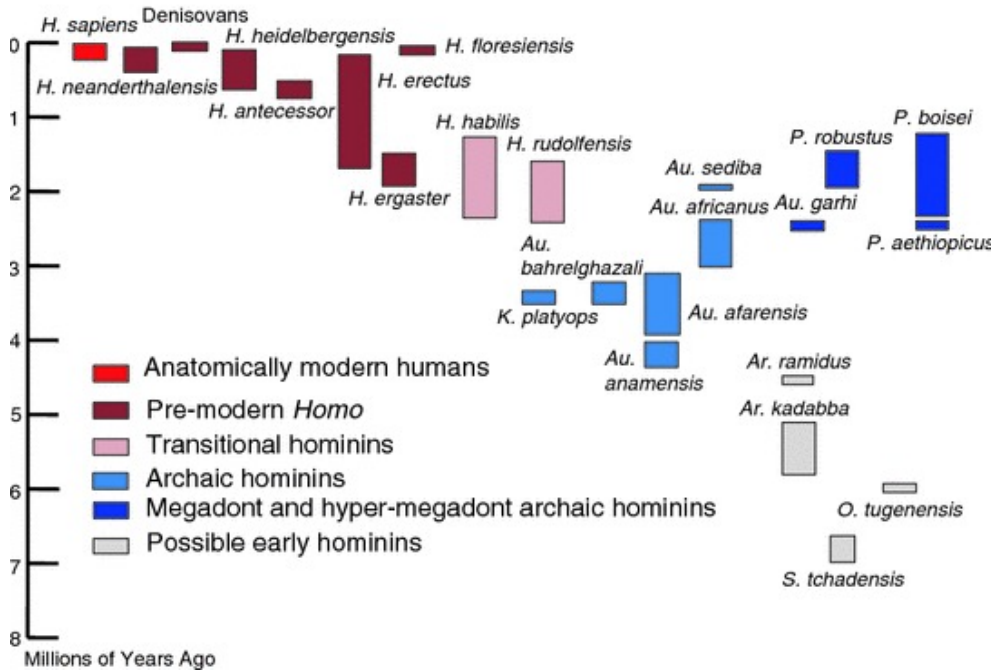
- Compute standard error by performing many *D* statistic discarding every time a different portion of the genome
- We obtain a Z-score by dividing *D* (with the whole genome) by its standard error
- Z-score > |3| indicates significant *D* value

Deep human evolutionary history revealed by aDNA

Neanderthal

- First ancient hominin discovered
- Modern humans closest relative
- Lived between ≈ 400,000 and 40,000 years ago
- Language?
- Abstract thinking?





D statistic – Human/Neanderthal admixture

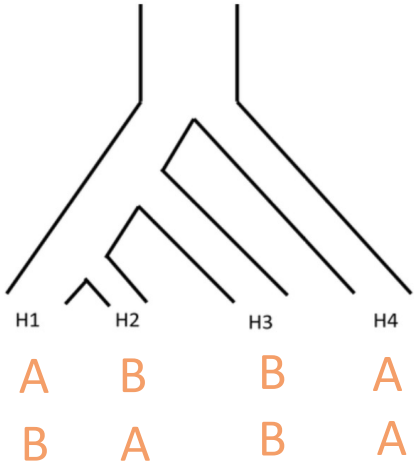
Whole genome sequences for one individual (or more) from each of the six following populations:

- Neanderthal
- Yoruba (Africa)
- Dinka (Africa)
- French (Europe)
- Han Chinese (East Asia)
- Chimpanzee (Outgroup)

We can compare their genomes and calculate the number of ABBA and BABA sites.

D statistic – Human/Neanderthal admixture

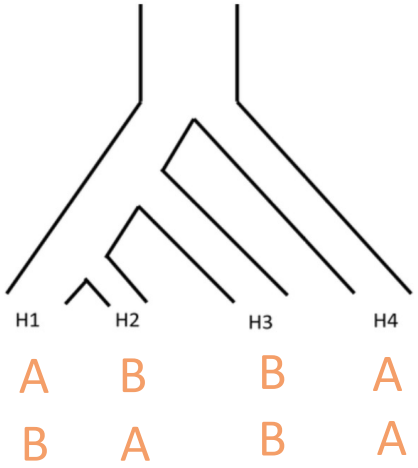
H1	H2	H3	H4	N˚ ABBA	N˚ BABA
Yoruba	Dinka	Neanderthal	Chimpanzee	44,161	44,221
Yoruba	French	Neanderthal	Chimpanzee	46,449	44,347
Yoruba	Han	Neanderthal	Chimpanzee	48,227	43,863



$$D(H_1, H_2; H_3, H_4) = \frac{(n_{ABBA} - n_{BABA})}{(n_{ABBA} + n_{BABA})}$$

D statistic – Human/Neanderthal admixture

H1	H2	H3	H4	N˚ ABBA	N˚ BABA
Yoruba	Dinka	Neanderthal	Chimpanzee	44,161	44,221
Yoruba	French	Neanderthal	Chimpanzee	46,449	44,347
Yoruba	Han	Neanderthal	Chimpanzee	48,227	43,863



$$D(H_1, H_2; H_3, H_4) = \frac{(n_{ABBA} - n_{BABA})}{(n_{ABBA} + n_{BABA})}$$

	Test	D-stat	Standard error	Z-score
Scenario 1	(Yoruba, Dinka; Neanderthal, Chimp)	-0.000678	0.00336	-0.201
Scenario 2	(Yoruba, French; Neanderthal, Chimp)	0.02315	0.00473	4.894
Scenario 3	(Yoruba, Han; Neanderthal, Chimp)	0.04738	0.00543	8.725

D statistic – Human/Neanderthal admixture

	Test	D-stat	Standard error	Z-score
Scenario 1	(Yoruba, Dinka; Neanderthal, Chimp)	-0.000678	0.00336	-0.201

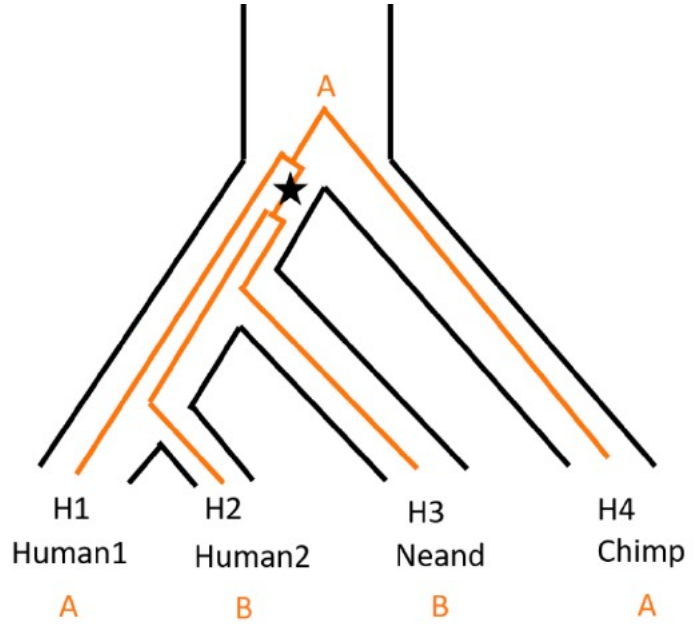
This result suggest that the pair of African genomes are symmetrically related to the Neanderthal and the chimp. Therefore, we infer that these two Africans form a clade to the exclusion of the Neanderthal and the chimp.

Moreover, we observe **no statistically significant evidence of gene flow between the African individuals and the Neanderthal**.

D statistic – Human/Neanderthal admixture

	Test	D-stat	Standard error	Z-score
Scenario 2	(Yoruba, French; Neanderthal, Chimp)	0.02315	0.00473	4.894

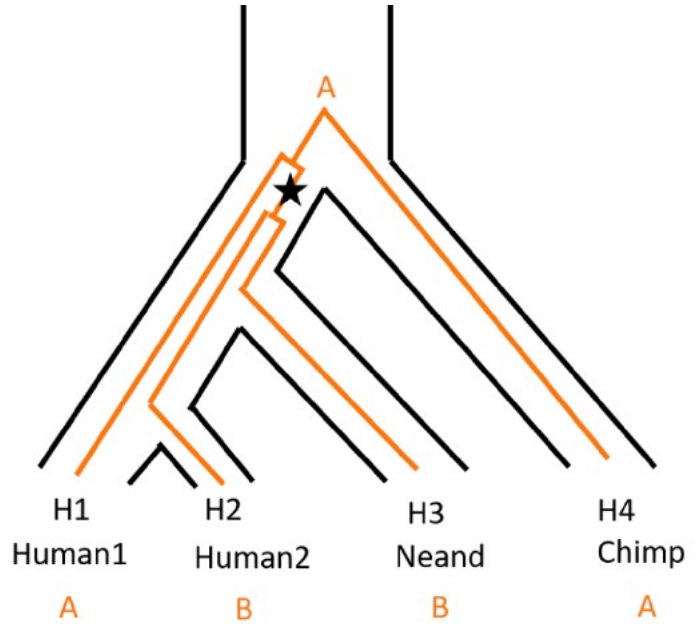
This result suggests that the **French genome shares a statistically significant larger proportion of derived alleles with the Neanderthal genome** (excess of ABBA sites), than the Yoruba does.



D statistic – Human/Neanderthal admixture

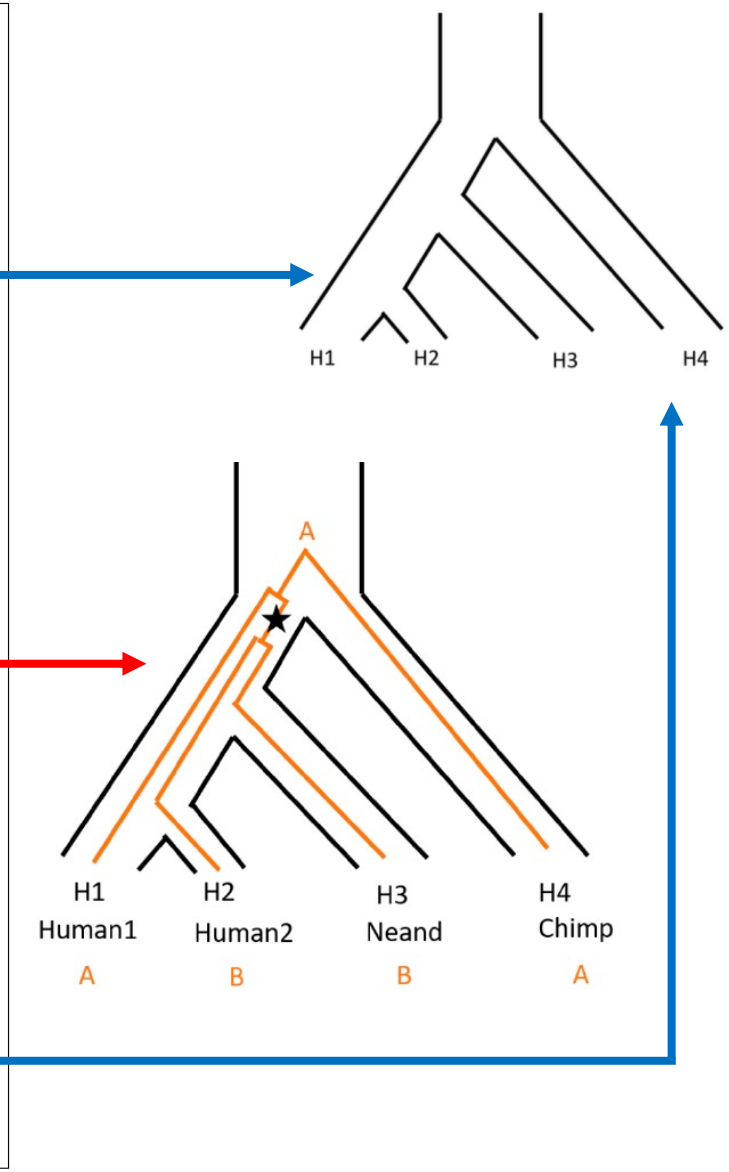
	Test	D-stat	Standard error	Z-score
Scenario 3	(Yoruba, Han; Neanderthal, Chimp)	0.04738	0.00543	8.725

Similar to what we observed for Scenario 2, this suggests that the **Han genome shares a statistically significant larger proportion of derived alleles with the Neanderthal genome** (excess of ABBA sites), than the Yoruba does.



D statistic – Human/Neanderthal admixture

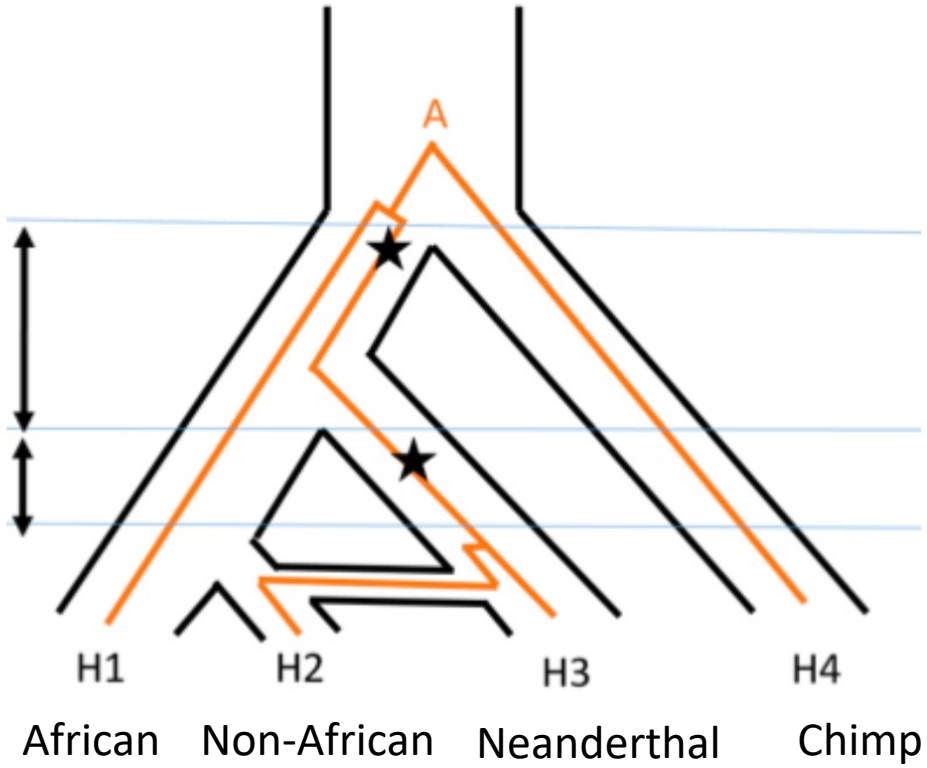
Population comparison	H_1	H_2	% Neandertal matching to H_2 – % Neandertal matching to H_1 (±1 standard error)
ABI3730 sequencing (~750 bp reads) used to discover H_1-H_2 differences			
African to African	NA18517 (Yoruba)	NA18507 (Yoruba)	-0.1 ± 0.6
	NA18517 (Yoruba)	NA19240 (Yoruba)	1.5 ± 0.7
	NA18517 (Yoruba)	NA19129 (Yoruba)	-0.1 ± 0.7
	NA18507 (Yoruba)	NA19240 (Yoruba)	-0.5 ± 0.6
	NA18507 (Yoruba)	NA19129 (Yoruba)	0.0 ± 0.5
	NA19240 (Yoruba)	NA19129 (Yoruba)	-0.6 ± 0.7
African to Non-African	NA18517 (Yoruba)	NA12878 (European)	**4.1 ± 0.8**
	NA18517 (Yoruba)	NA12156 (European)	**5.1 ± 0.7**
	NA18517 (Yoruba)	NA18956 (Japanese)	**2.9 ± 0.8**
	NA18517 (Yoruba)	NA18555 (Chinese)	**3.9 ± 0.7**
	NA18507 (Yoruba)	NA12878 (European)	**4.2 ± 0.6**
	NA18507 (Yoruba)	NA12156 (European)	**5.5 ± 0.6**
	NA18507 (Yoruba)	NA18956 (Japanese)	**5.0 ± 0.7**
	NA18507 (Yoruba)	NA18555 (Chinese)	**5.8 ± 0.6**
	NA19240 (Yoruba)	NA12878 (European)	**3.5 ± 0.7**
	NA19240 (Yoruba)	NA12156 (European)	**3.1 ± 0.7**
	NA19240 (Yoruba)	NA18956 (Japanese)	**2.7 ± 0.7**
	NA19240 (Yoruba)	NA18555 (Chinese)	**5.4 ± 0.9**
	NA19129 (Yoruba)	NA12878 (European)	**3.9 ± 0.7**
	NA19129 (Yoruba)	NA12156 (European)	**4.9 ± 0.7**
	NA19129 (Yoruba)	NA18956 (Japanese)	**5.1 ± 0.8**
	NA19129 (Yoruba)	NA18555 (Chinese)	**4.7 ± 0.8**
Non-African to Non-African	NA12878 (European)	NA12156 (European)	-0.5 ± 0.8
	NA12878 (European)	NA18956 (Japanese)	0.4 ± 0.8
	NA12878 (European)	NA18555 (Chinese)	0.3 ± 0.8
	NA12156 (European)	NA18956 (Japanese)	-0.3 ± 0.8
	NA12156 (European)	NA18555 (Chinese)	1.3 ± 0.7
	NA18956 (Japanese)	NA18555 (Chinese)	2.5 ± 0.9

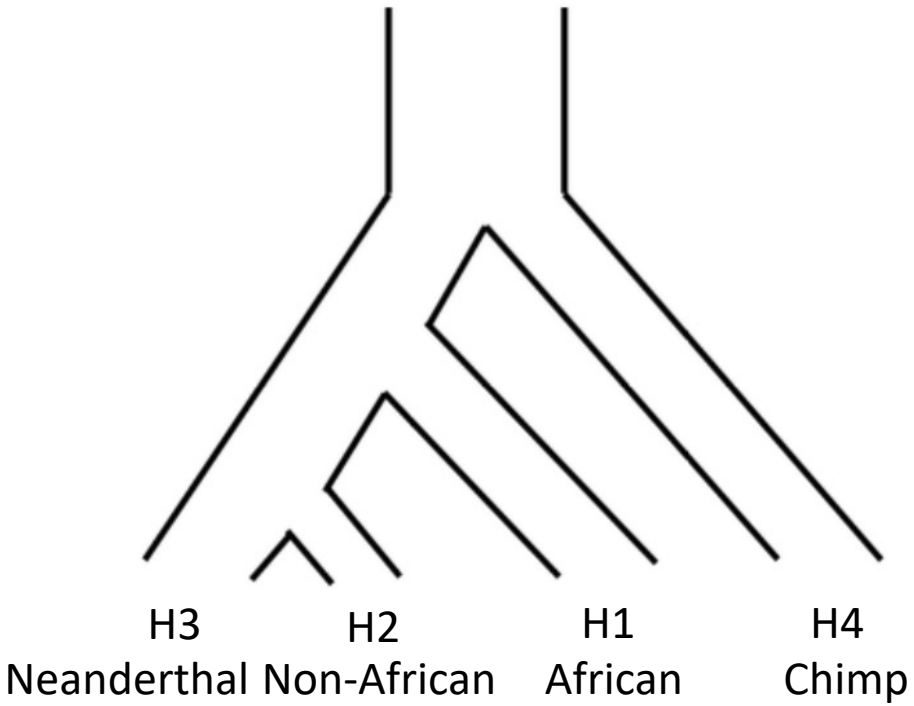


D statistic – Human/Neanderthal admixture

What if D ≠ 0?

- Gene flow
- The tree is not correct



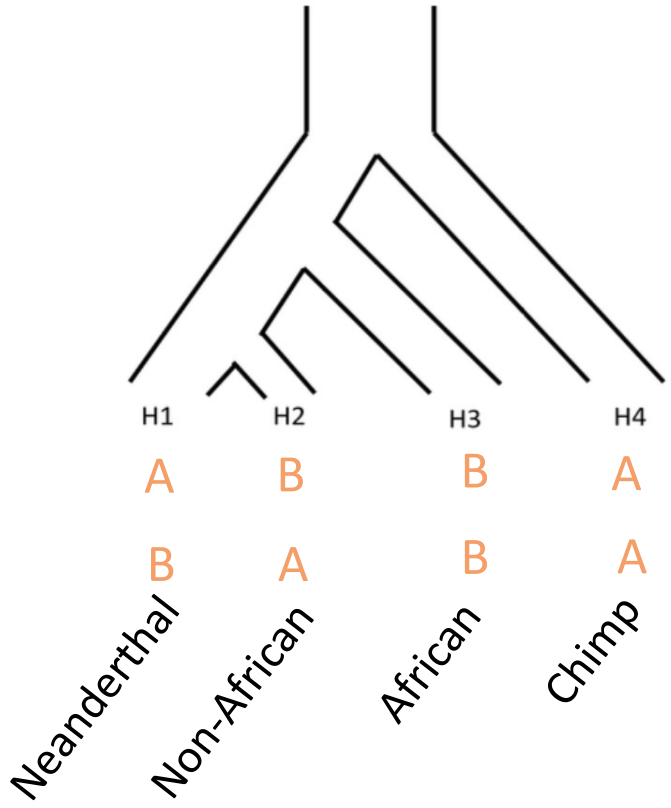


What is the right model?

D statistic – Human/Neanderthal admixture

How we can discriminate between the two model:

- Perform the complementary D-test



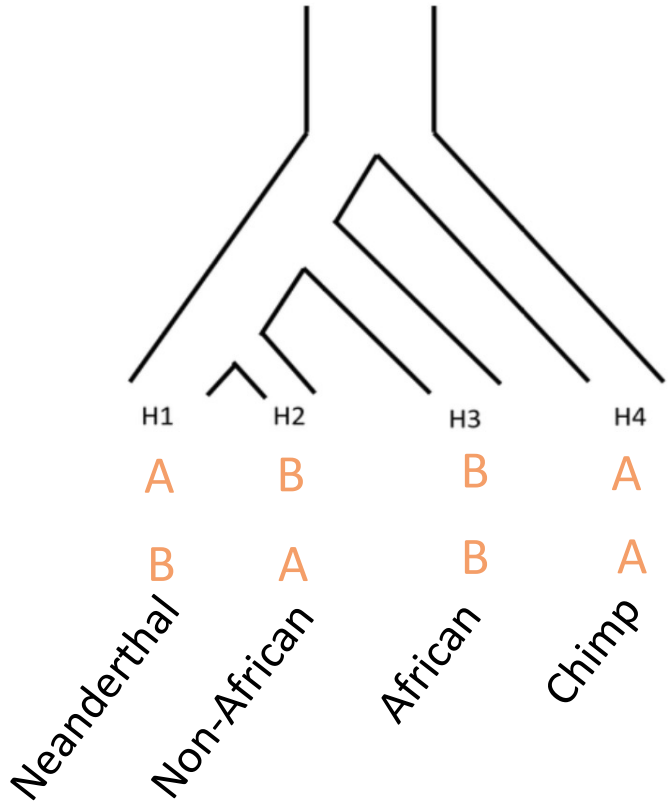
If this tree is correct, D = 0 (different from zero in a non-statistically significant way)

But, if African and non-African form a clade, we observe D > 0 (more ABBA sites)

D statistic – Human/Neanderthal admixture

How we can discriminate between the two model:

- Perform the complementary D-test



If this tree is correct, D = 0 (different from zero in a non-statistically significant way)

But, if African and non-African form a clade, we observe D > 0 (more ABBA sites)

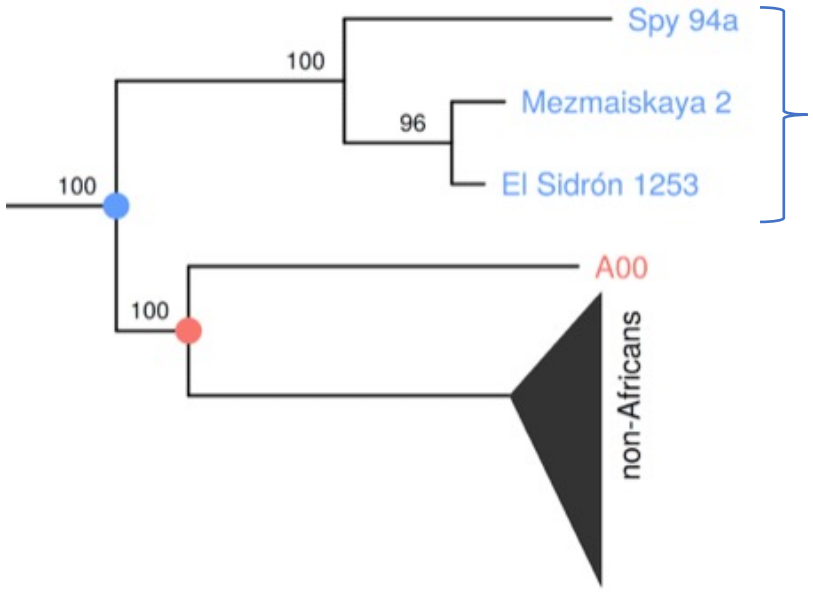
→ **D > 0**

D statistic – Human/Neanderthal admixture

How we can discriminate between the two model:

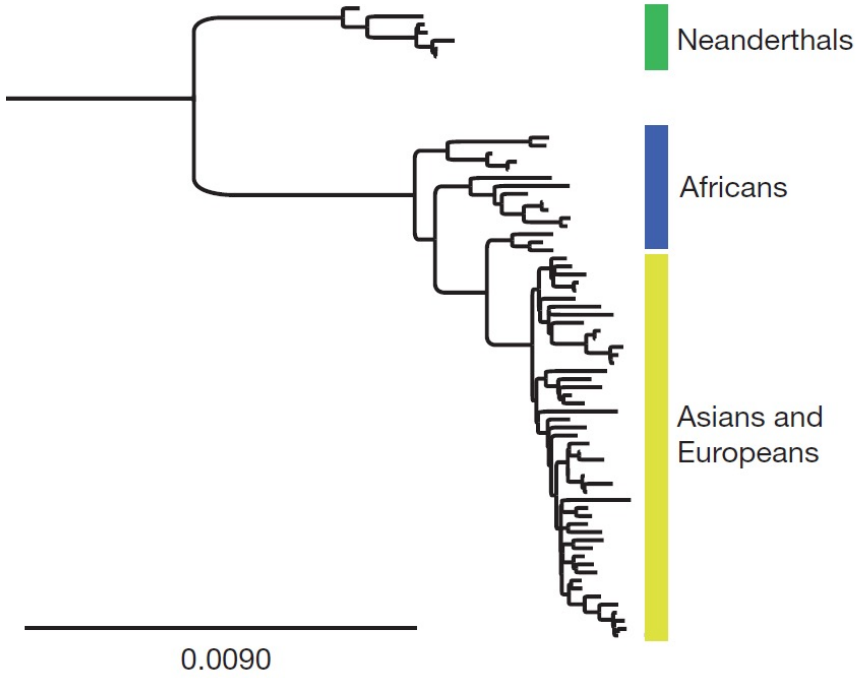
- Compare the results with different analysis

Y chromosome phylogeny



Neanderthal

mtDNA phylogeny

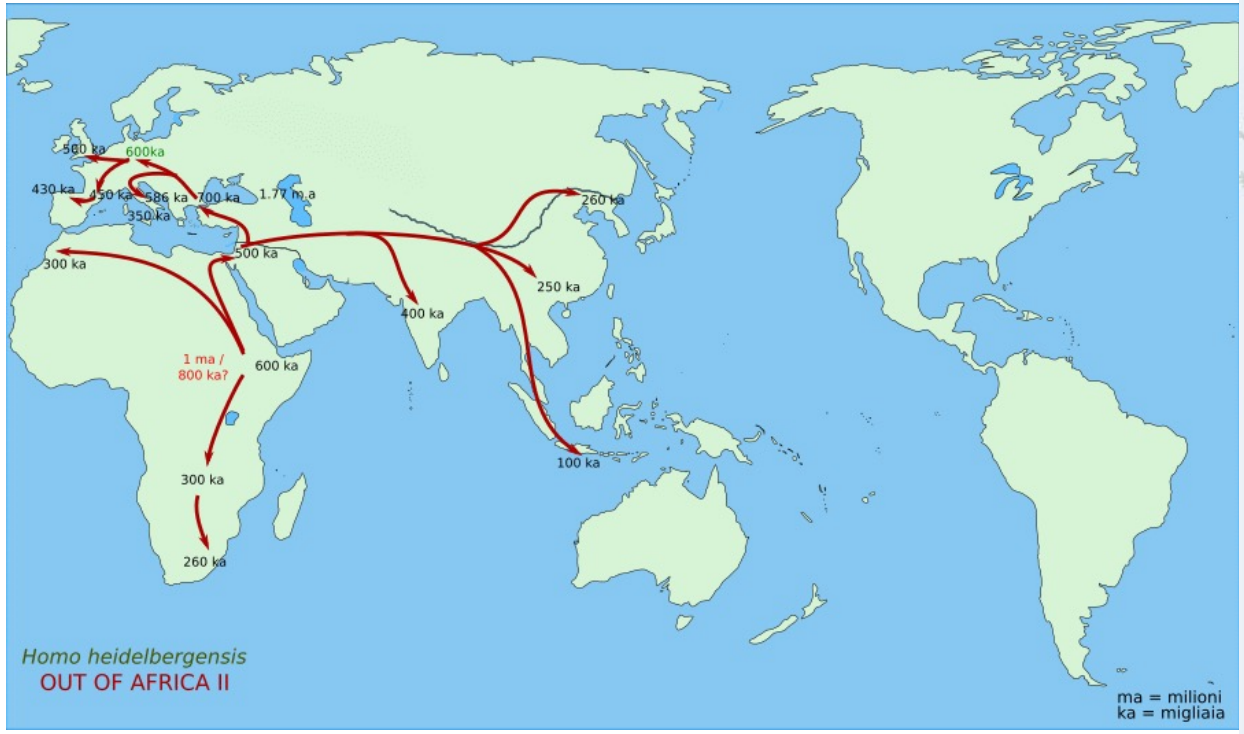


D statistic – Human/Neanderthal admixture

How we can discriminate between the two model:

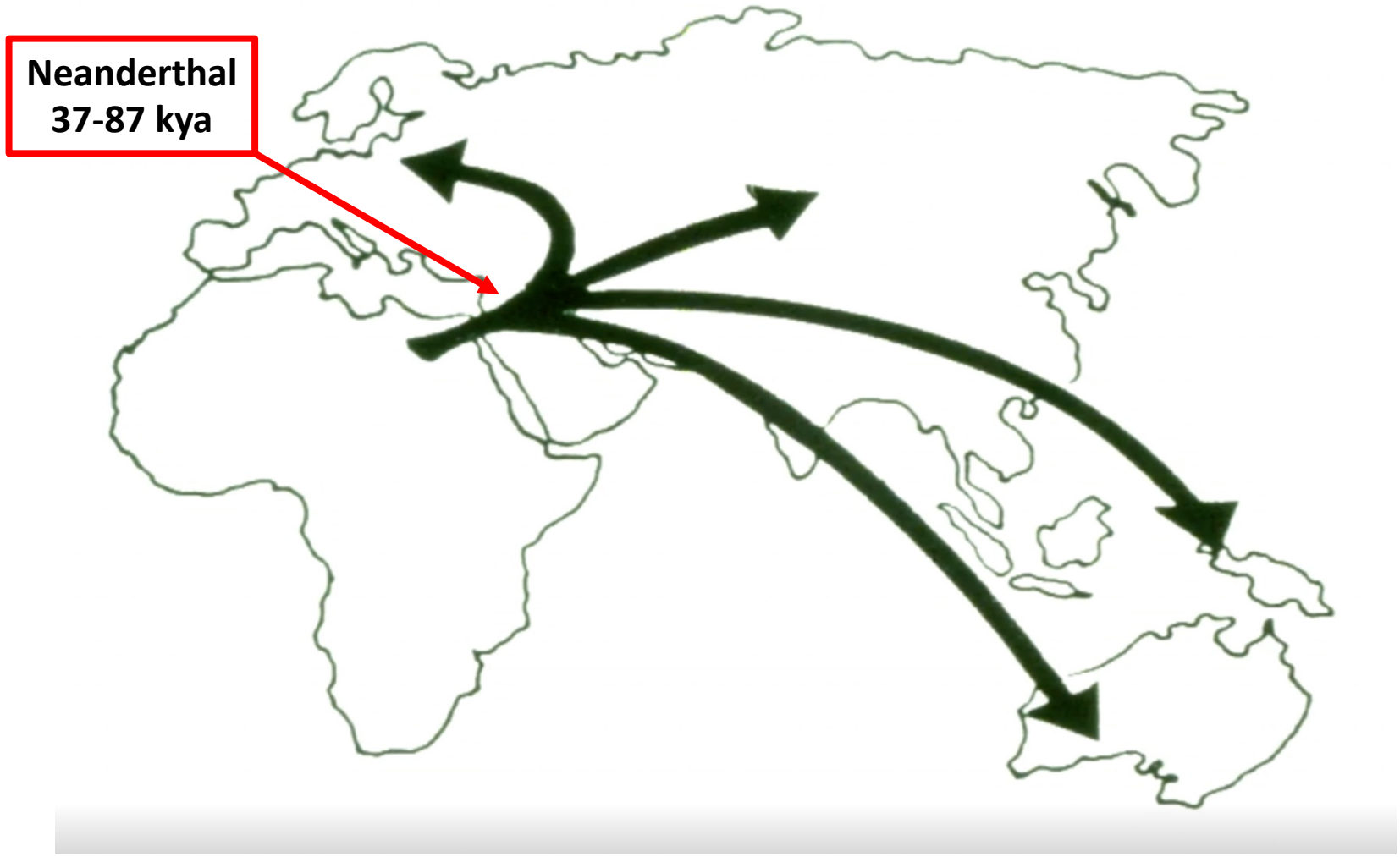
- Compare the results with different analysis

Neanderthal ancestors out of Africa ≈ 500 kya



Modern humans out of Africa ≈ 100 kya

D statistic – Human/Neanderthal admixture



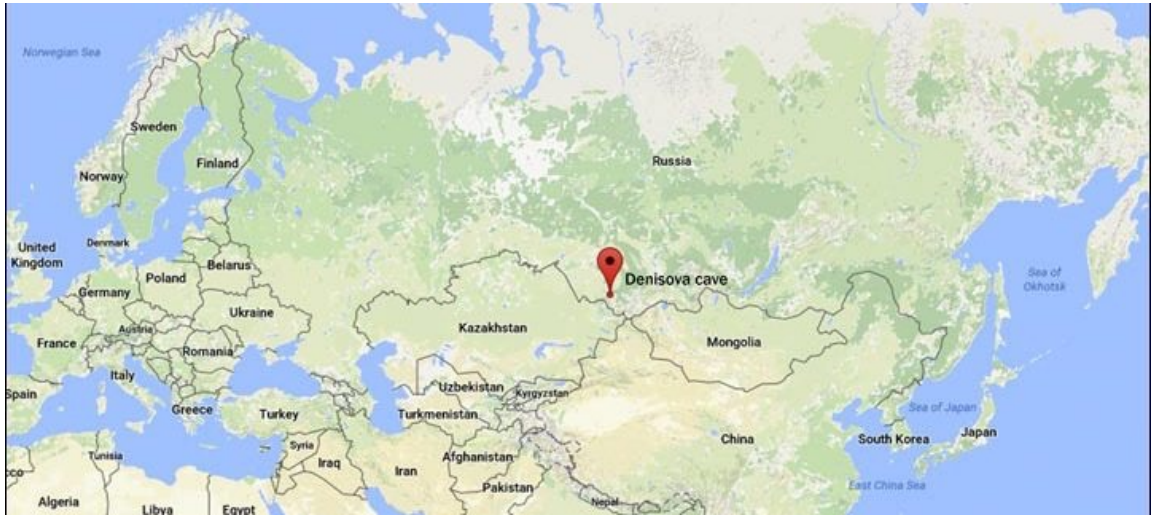
Selection of Neanderthal-specific genes

Locus	Type of selection	Associated phenotype
AMY1	Negative	Starch digestion
FOXP2	Negative	Speech
ASB1	Positive	Night-time activity
HLA	Positive	Immune response

Introgression may have facilitated the adaptation of Eurasian populations to their new environment

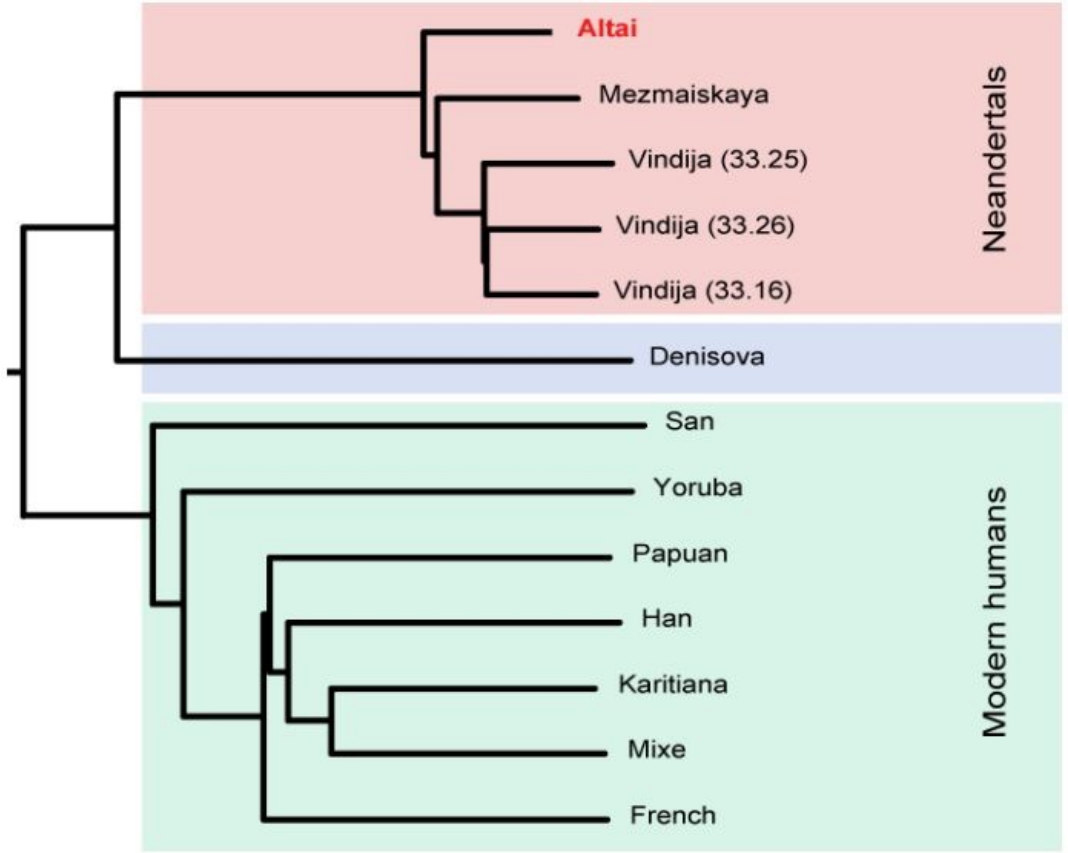
DNA from an ancient hominin in Denisova

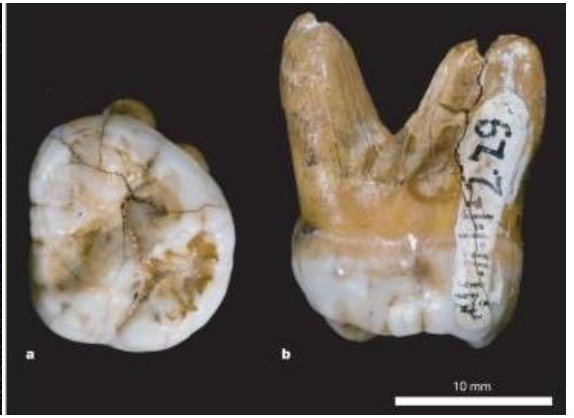
Whole genome analysis of the DNA obtained from the small phalanx

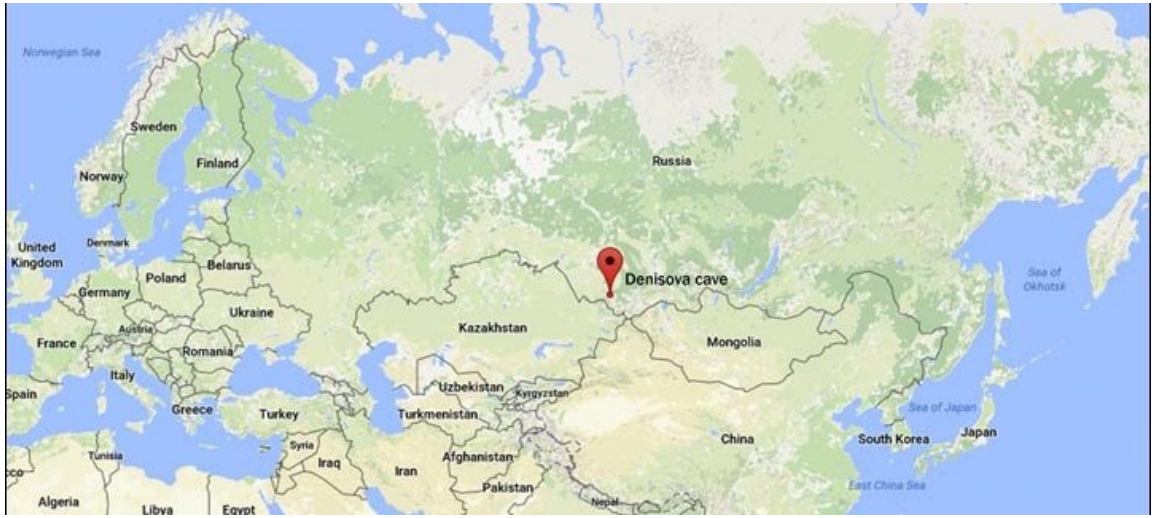


DNA from an ancient hominin in Denisova

Whole genome analysis of the DNA obtained from the small phalanx

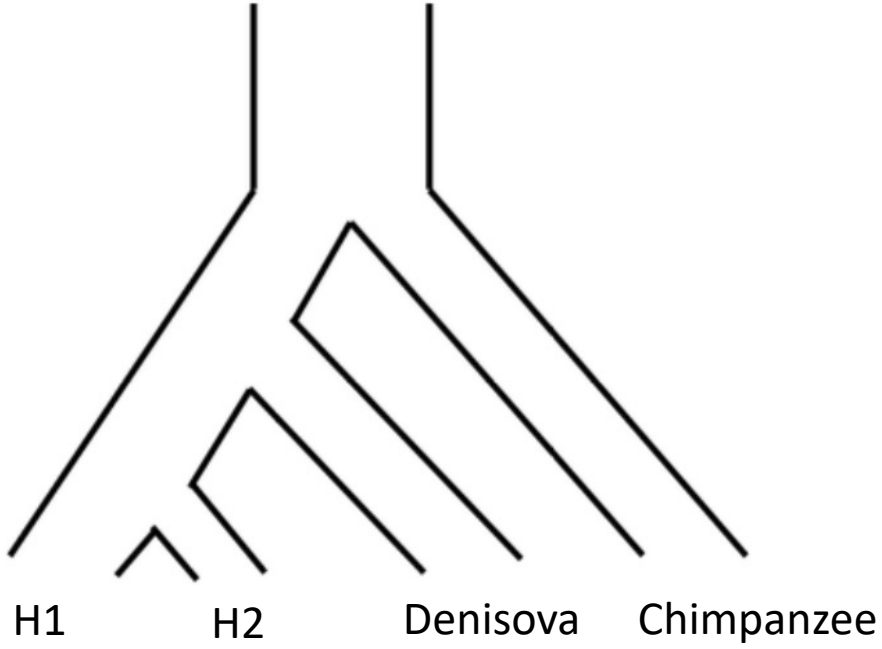




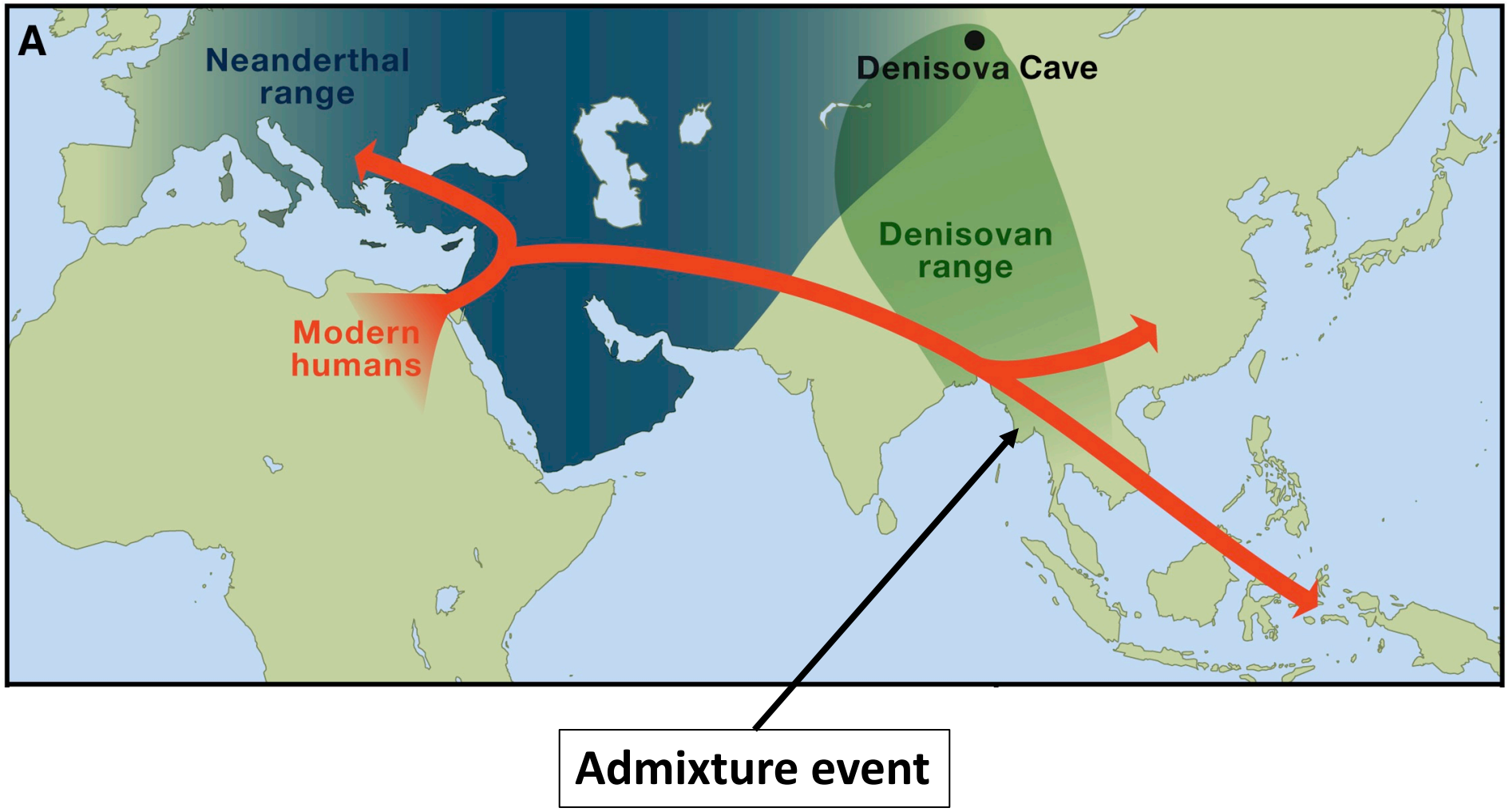


Human/Denisova Admixture

Sample H_1	Sample H_2	$D(H_1, H_2, \text{Denisova, chimpanzee})$				
		n_{BABA}	n_{ABBA}	D (%)	s.e. (%)	Z-score
Eurasian/Eurasian*						
French	Han	27,250	27,265	0.0	0.6	0.0
Karitiana	Sardinian	1,559	1,627	−2.1	1.8	−1.2
Karitiana	Cambodian	2,371	2,460	−1.8	1.5	−1.2
Karitiana	Mongolian	1,765	1,742	0.7	1.8	0.4
Sardinian	Cambodian	3,935	3,925	0.1	1.2	0.1
Sardinian	Mongolian	3,036	3,057	−0.3	1.3	−0.3
Cambodian	Mongolian	4,442	4,342	1.1	1.2	1.0
African/African*						
San	Yoruba	39,042	39,019	0.0	0.5	0.1
Melanesian/Melanesian*						
Papuan2	Bougainville	5,319	5,140	1.7	1.1	1.5
Eurasian/African*						
French	San	39,838	38,495	1.7	0.5	3.4†
French	Yoruba	34,262	33,078	1.8	0.5	3.6†
Han	San	38,815	37,439	1.8	0.5	3.4†
Han	Yoruba	33,182	32,184	1.5	0.5	2.8
Karitiana	Mbuti	2,368	2,360	0.2	1.5	0.1
Sardinian	Mbuti	4,028	3,784	3.1	1.2	2.6
Cambodian	Mbuti	6,329	5,850	3.9	1.0	4.0†
Mongolian	Mbuti	4,514	4,505	0.1	1.1	0.1
Eurasian/Melanesian*						
French	Papuan1	23,509	25,470	−4.0	0.7	−5.7†
Han	Papuan1	22,262	24,198	−4.2	0.7	−5.8†
Karitiana	Papuan2	2,201	2,641	−9.1	1.6	−5.8†
Karitiana	Bougainville	2,229	2,671	−9.0	1.5	−5.9†
Sardinian	Papuan2	3,714	4,150	−5.5	1.2	−4.5†
Sardinian	Bougainville	3,877	4,336	−5.6	1.1	−4.9†
Cambodian	Papuan2	5,457	6,272	−6.9	1.1	−6.5†
Cambodian	Bougainville	5,751	6,333	−4.8	1.0	−4.7†
Mongolian	Papuan2	4,192	4,758	−6.3	1.2	−5.3†
Mongolian	Bougainville	4,234	4,847	−6.8	1.1	−6.0†
Melanesian/African*						
Papuan1	San	35,923	32,841	4.5	0.6	7.2†
Papuan1	Yoruba	30,995	28,186	4.7	0.6	7.4†
Papuan2	Mbuti	6,124	5,233	7.8	1.1	7.2†
Bougainville	Mbuti	6,498	5,633	7.1	1.1	6.7†

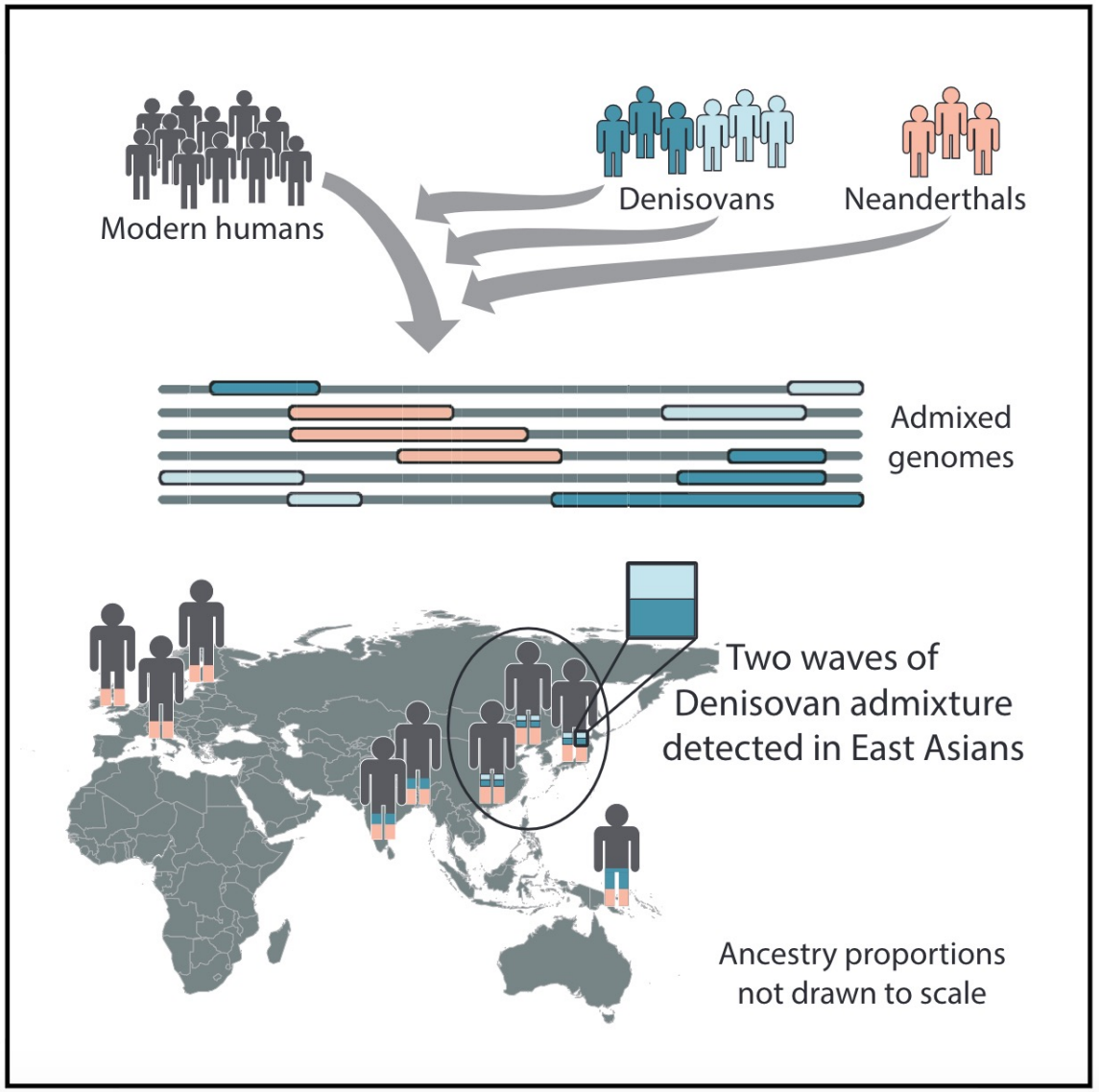


Human/Denisova Admixture



Two pulses of Denisova admixture

- Different Denisovans contributed to Asian and Melanesian modern genomes
- East Asians carry signature of both the admixture events
- Melanesians admixed only with “Oceanian” Denisovans



Selection of Denisova-specific genes

LETTER

doi:10.1038/nature13408

Altitude adaptation in Tibetans caused by introgression of Denisovan-like DNA

Emilia Huerta-Sánchez[1,2,3]*, Xin Jin[1,4]*, Asan[1,5,6]*, Zhuoma Bianba[7]*, Benjamin M. Peter[2], Nicolas Vinckenbosch[2], Yu Liang[1,5,6], Xin Yi[1,5,6], Mingze He[1,8], Mehmet Somel[9], Peixiang Ni[1], Bo Wang[1], Xiaohua Ou[1], Huasang[1], Jiangbai Luosang[1], Zha Xi Ping Cuo[10], Kui Li[11], Guoyi Gao[12], Ye Yin[1], Wei Wang[1], Xiuqing Zhang[1,13,14], Xun Xu[1], Huanming Yang[1,15,16], Yingrui Li[1], Jian Wang[1,16], Jun Wang[1,15,17,18,19] & Rasmus Nielsen[1,2,20,21]

- The physiological response to low oxygen differs between Tibetans and individuals of low-altitude origin
- Adaptations that confer lower infant mortality and higher fertility
- *EPAS1* (a transcription factor induced under hypoxic conditions) is the gene with the strongest signal of Tibetan specific selection
- Denisovans have the same *EPAS1* as the Tibetans
- Adaptive introgression from Denisovans into the Tibetans genome

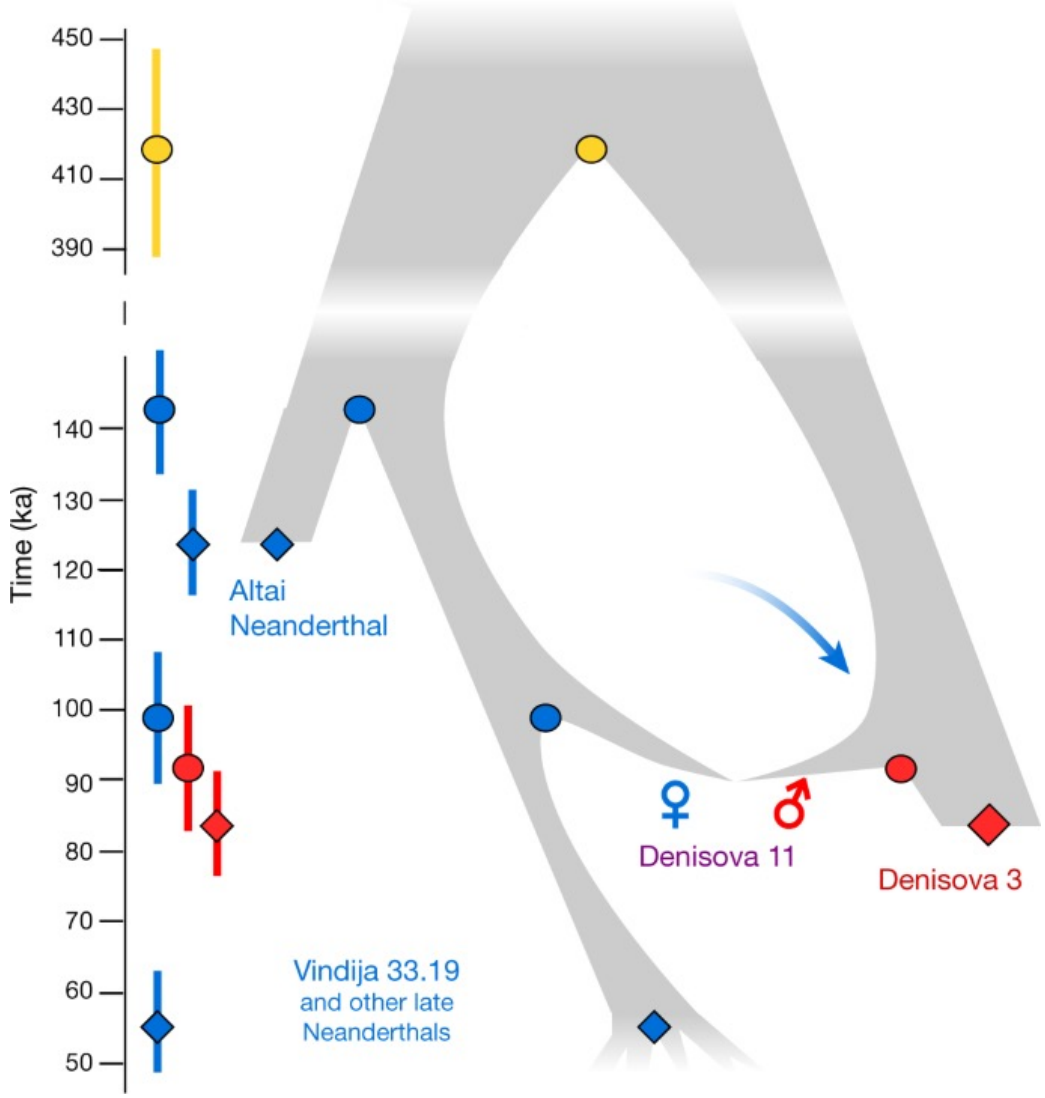
The story gets complicated

LETTER

https://doi.org/10.1038/s41586-018-0455-x

The genome of the offspring of a Neanderthal mother and a Denisovan father

Viviane Slon[1,7]*, Fabrizio Mafessoni[1,7], Benjamin Vernot[1,7], Cesare de Filippo[1], Steffi Grote[1], Bence Viola[2,3], Mateja Hajdinjak[1], Stéphane Peyrégne[1], Sarah Nagel[1], Samantha Brown[4], Katerina Douka[4,5], Tom Higham[5], Maxim B. Kozlikin[3], Michael V. Shunkov[3,6], Anatoly P. Derevianko[3], Janet Kelso[1], Matthias Meyer[1], Kay Prüfer[1] & Svante Pääbo[1]*



The story gets complicated

LETTER

doi:10.1038/nature14558

An early modern human from Romania with a recent Neanderthal ancestor

Qiaomei Fu[1,2,3]*, Mateja Hajdinjak[3]*, Oana Teodora Moldovan[4], Silviu Constantin[5], Swapan Mallick[2,6,7], Pontus Skoglund[2], Nick Patterson[6], Nadin Rohland[2], Iosif Lazaridis[2], Birgit Nickel[3], Bence Viola[3,7,8], Kay Prüfer[3], Matthias Meyer[3], Janet Kelso[3], David Reich[2,6,9] & Svante Pääbo[3]

Human/Neanderthal admixture only 4 generation before this individual

ARTICLE

doi:10.1038/nature16544

Ancient gene flow from early modern humans into Eastern Neanderthals

Martin Kuhlwilm[1]*, Ilan Gronau[2]*, Melissa J. Hubisz[3], Cesare de Filippo[1], Javier Prado-Martinez[4], Martin Kircher[1,5], Qiaomei Fu[1,6,7], Hernán A. Burbano[1,8], Carles Lalueza-Fox[4], Marco de la Rasilla[9], Antonio Rosas[10], Pavao Rudan[11], Dejana Brajkovic[12], Željko Kucan[11], Ivan Gušic[11], Tomas Marques-Bonet[4,13,14], Aida M. Andrés[1], Bence Viola[15,16], Svante Pääbo[1], Matthias Meyer[1], Adam Siepel[3,17] & Sergi Castellano[1]

From a human population more ancient than the Out Of Africa

ARTICLE

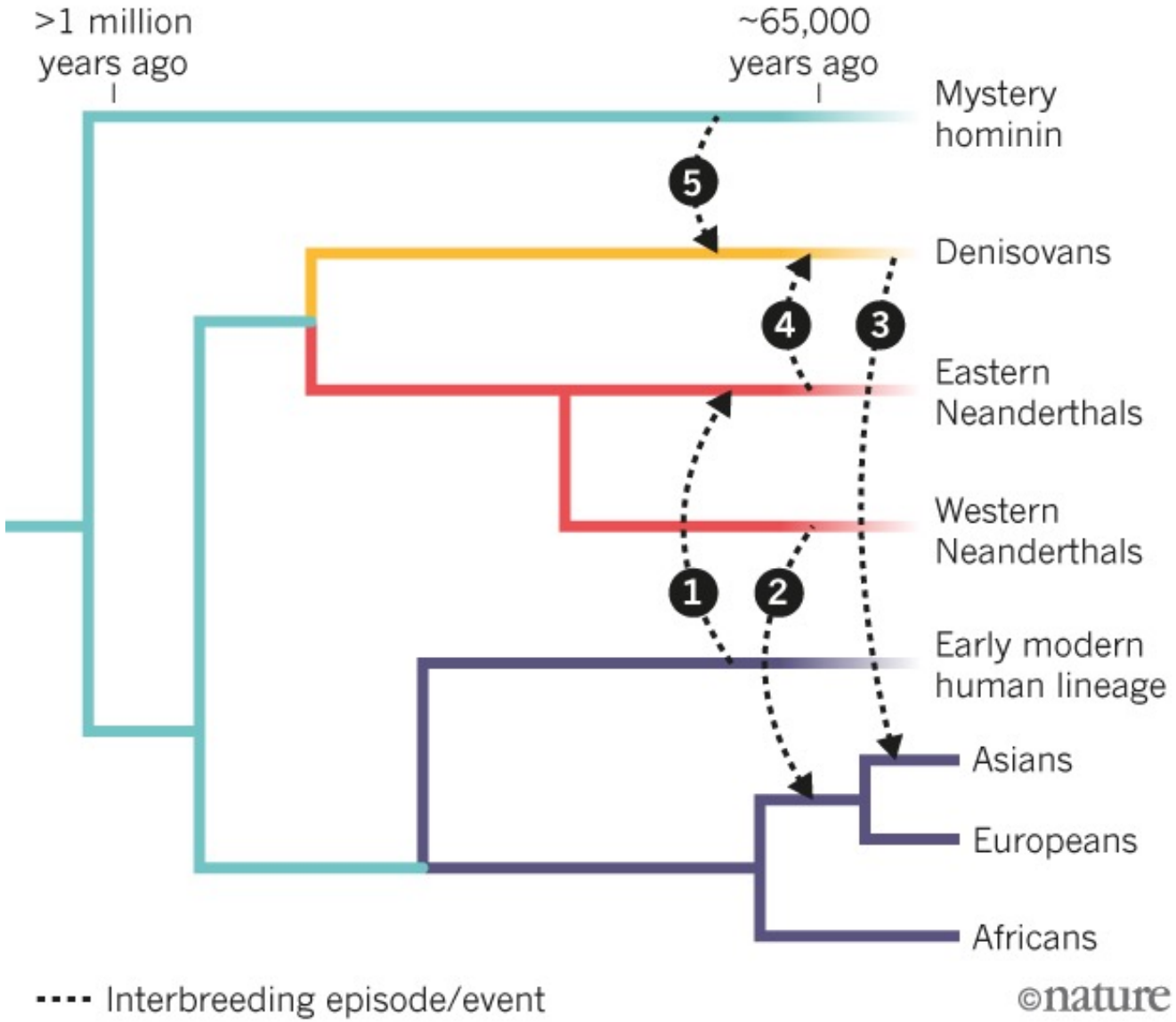
doi:10.1038/nature12886

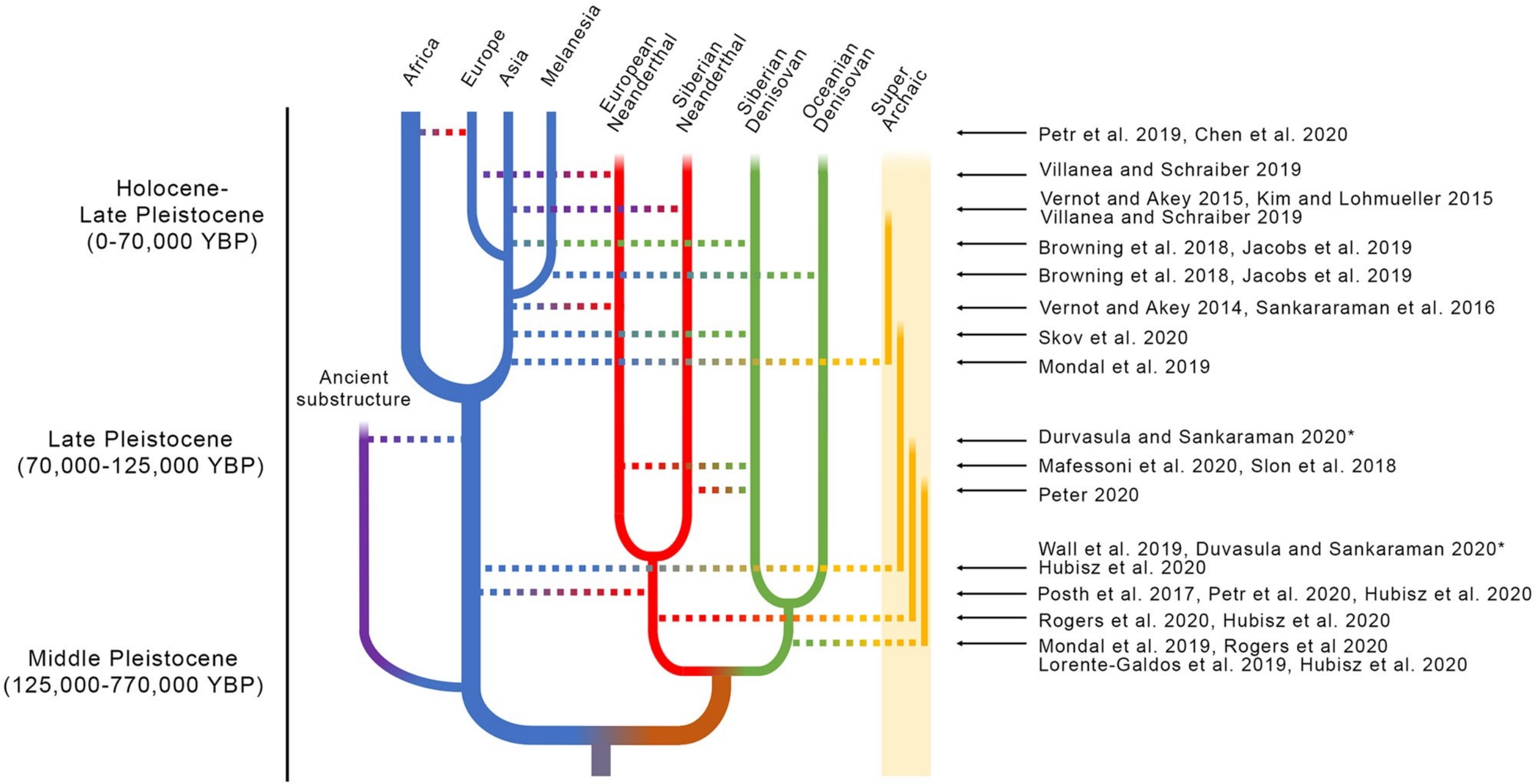
The complete genome sequence of a Neanderthal from the Altai Mountains

Kay Prüfer[1], Fernando Racimo[2], Nick Patterson[3], Flora Jay[2], Sriram Sankararaman[3,4], Susanna Sawyer[1], Anja Heinze[1], Gabriel Renaud[1], Peter H. Sudmant[5], Cesare de Filippo[1], Heng Li[3], Swapan Mallick[3,4], Michael Dannemann[1], Qiaomei Fu[1,6], Martin Kircher[1,5], Martin Kuhlwilm[1], Michael Lachmann[1], Matthias Meyer[1], Matthias Ongyerth[1], Michael Siebauer[1], Christoph Theunert[1], Arti Tandon[3,4], Priya Moorjani[4], Joseph Pickrell[4], James C. Mullikin[7], Samuel H. Vohr[8], Richard E. Green[8], Ines Hellmann[9]†, Philip L. F. Johnson[10], Hélène Blanche[11], Howard Cann[11], Jacob O. Kitzman[5], Jay Shendure[5], Evan E. Eichler[5,12], Ed S. Lein[13], Trygve E. Bakken[13], Liubov V. Golovanova[14], Vladimir B. Doronichev[14], Michael V. Shunkov[15], Anatoli P. Derevianko[15], Bence Viola[16], Montgomery Slatkin[2], David Reich[3,4,17], Janet Kelso[1] & Svante Pääbo[1]

Introgression from an unknown “super archaic” hominin into the Denisovans

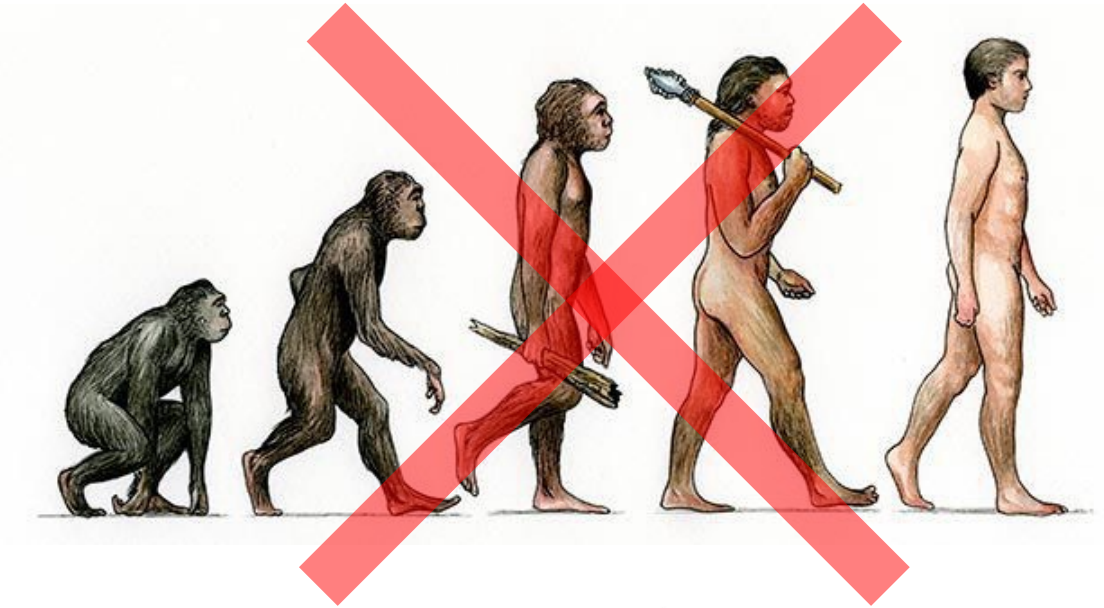
Early modern humans, Neanderthal and Denisovans all interbred with each other on multiple occasions in the past 100,000 years

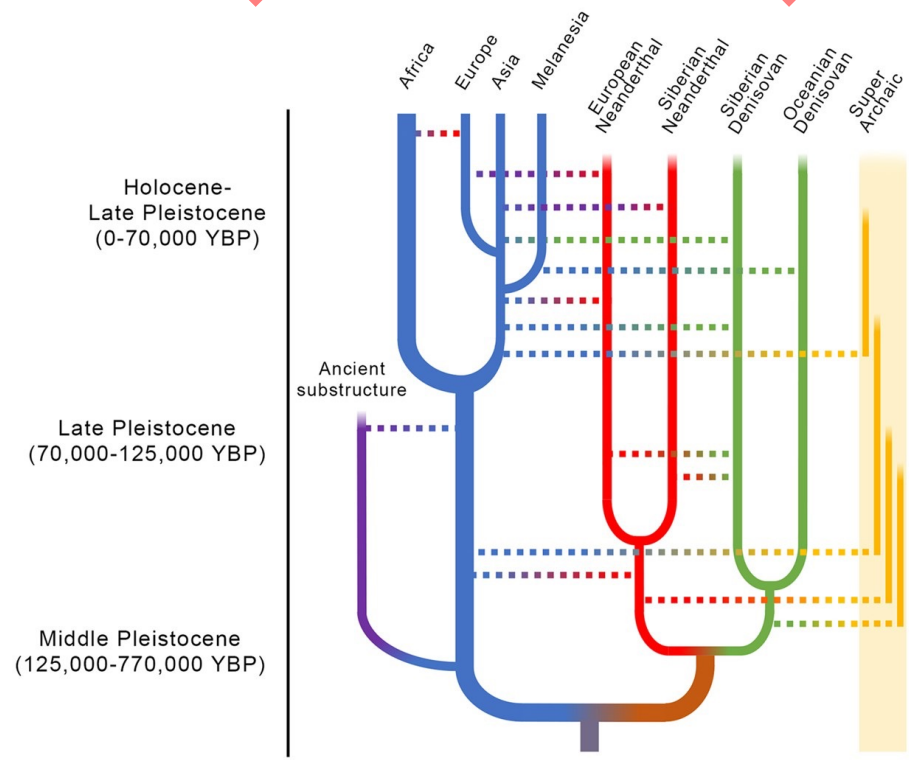



Africa
Europe
Asia
Melanesia
European Neanderthal
Siberian Neanderthal
Siberian Denisovan
Oceanian Denisovan
Super Archaic
Ancient substructure
Holocene-
Late Pleistocene
(0-70,000 YBP)
Late Pleistocene
(70,000-125,000 YBP)
Middle Pleistocene
(125,000-770,000 YBP)
Petr et al. 2019, Chen et al. 2020
Villanea and Schraiber 2019
Vernot and Akey 2015, Kim and Lohmueller 2015
Villanea and Schraiber 2019
Browning et al. 2018, Jacobs et al. 2019
Browning et al. 2018, Jacobs et al. 2019
Vernot and Akey 2014, Sankararaman et al. 2016
Skov et al. 2020
Mondal et al. 2019
Durvasula and Sankaraman 2020*
Mafessoni et al. 2020, Slon et al. 2018
Peter 2020
Wall et al. 2019, Duvasula and Sankaraman 2020*
Hubisz et al. 2020
Posth et al. 2017, Petr et al. 2020, Hubisz et al. 2020
Rogers et al. 2020, Hubisz et al. 2020
Mondal et al. 2019, Rogers et al 2020
Lorente-Galdos et al. 2019, Hubisz et al. 2020

Human evolution

- Not a linear process
- It is a bush in which the branches cross each other
- Some of these crosses have turned out to be beneficial for some modern populations

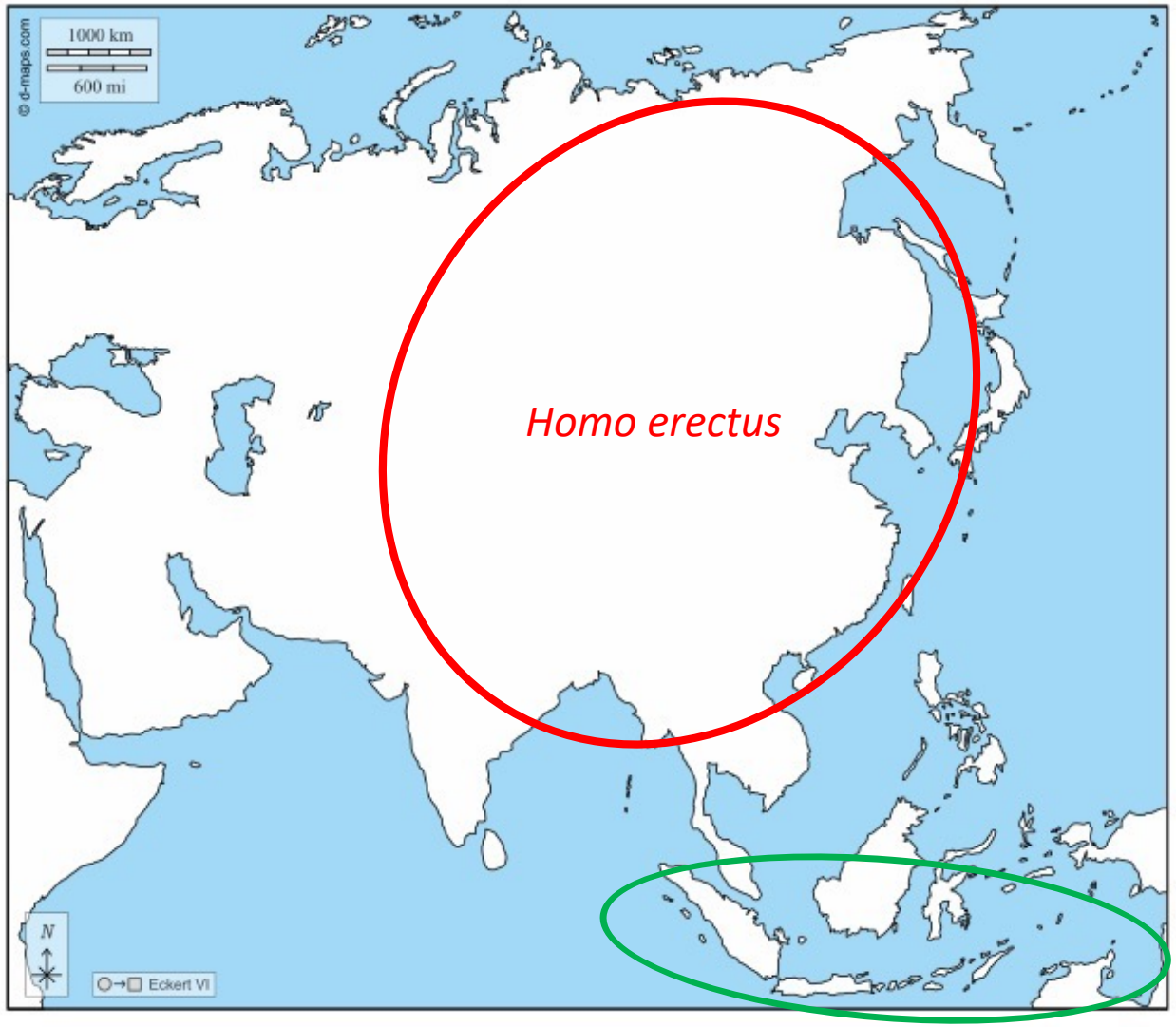




Human evolution

- Not a linear process
- It is a bush in which the branches cross each other
- Some of these crosses have turned out to be beneficial for some modern populations

- Super-archaic?
- *Homo erectus*?
- *Homo floresiensis*?



francesco.ravasini@uniroma1.it