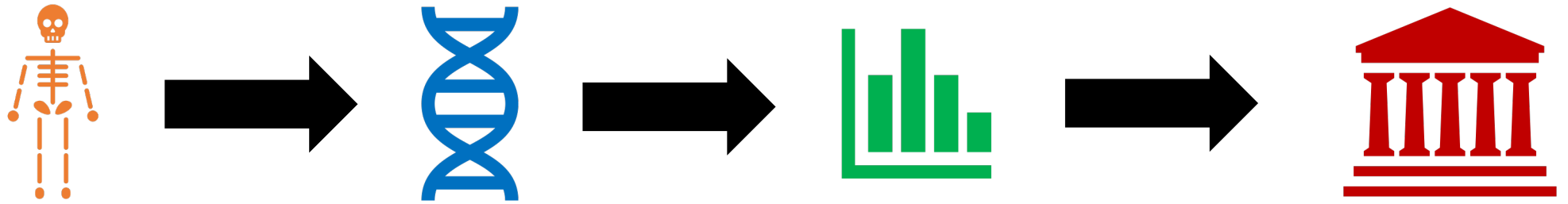
aDNA: Methods and Applications

aDNA analysis

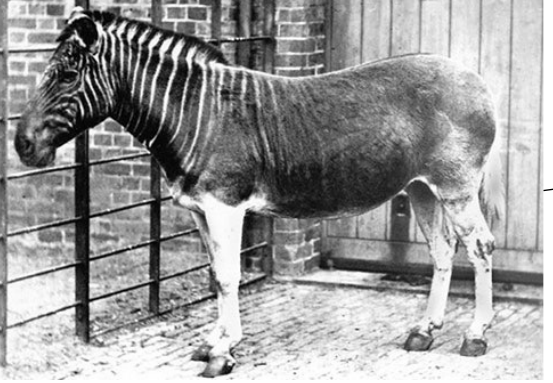


Samples

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

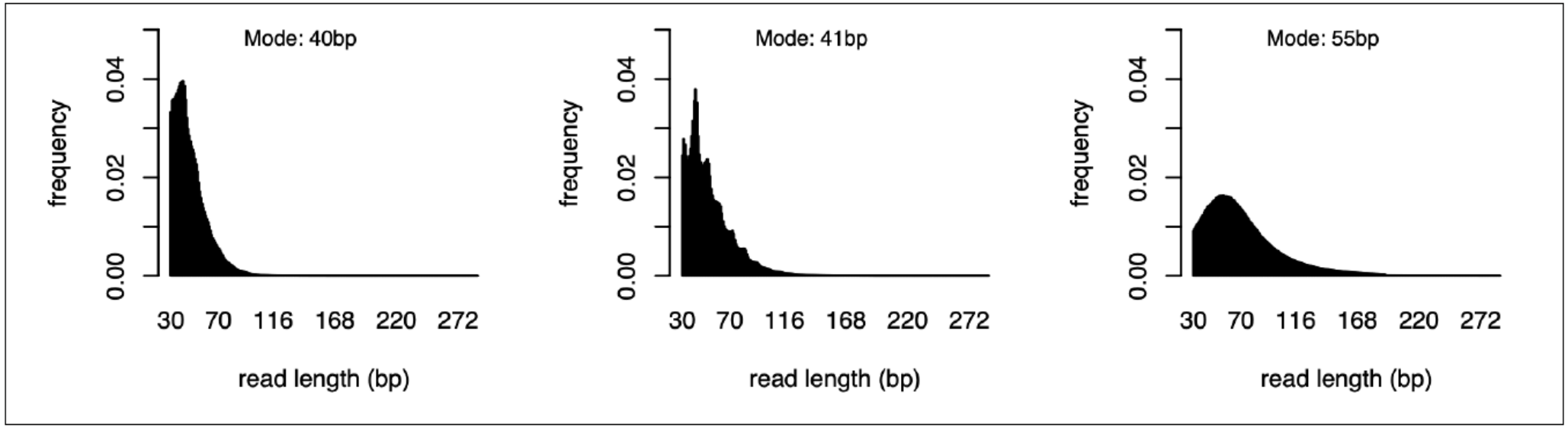
Analyse a precise moment

Characteristics of aDNA

Degradation: Fragmentation and post-mortem damage

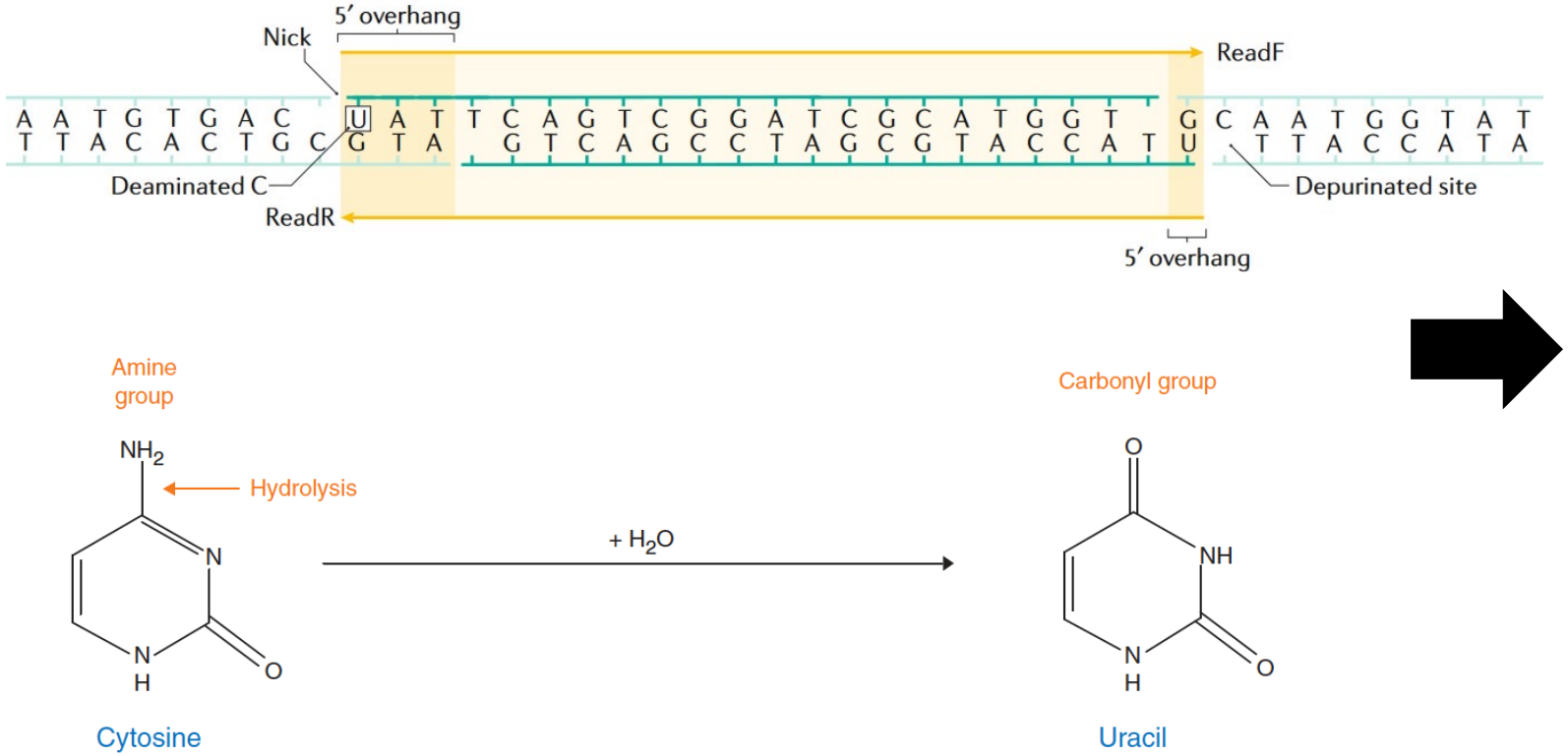
Characteristics of aDNA

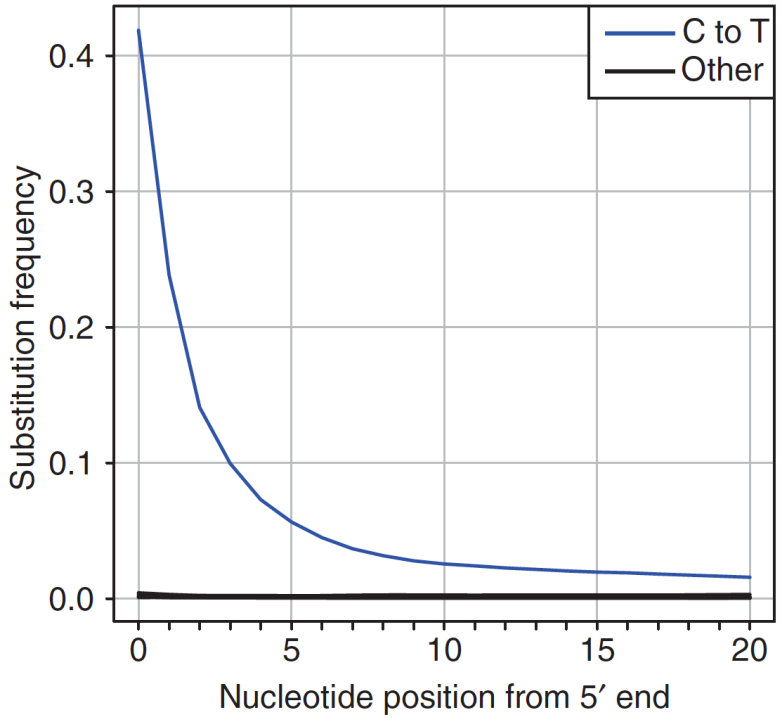
Degradation: Fragmentation and post-mortem damage



Characteristics of aDNA

Degradation: Fragmentation and post-mortem damage

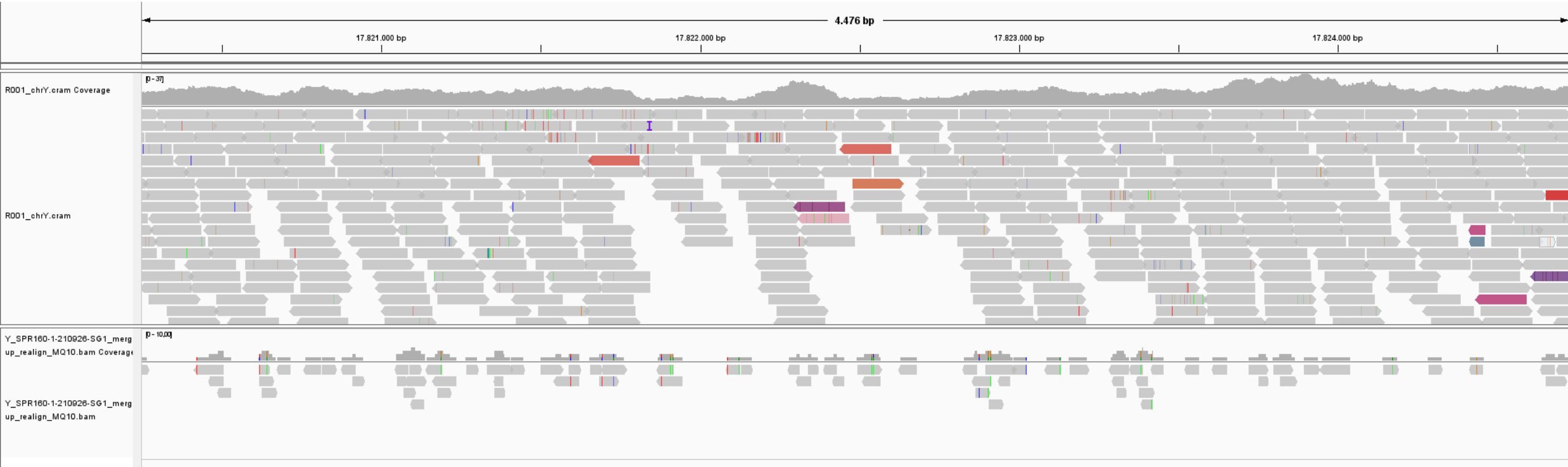




Characteristics of aDNA

Usually found in low quantities

↳ Resulting in low coverage sequences



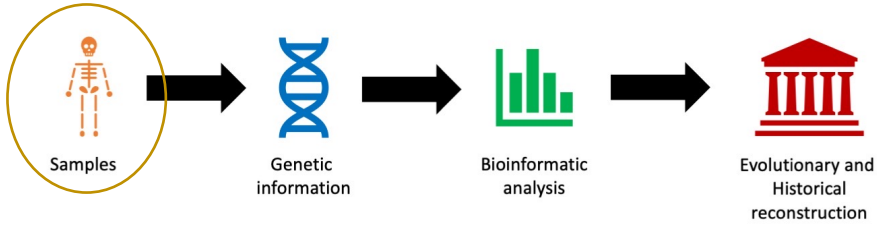
Characteristics of aDNA

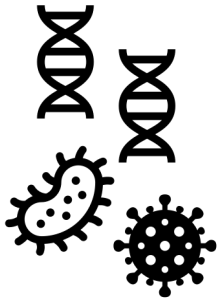
Potentially contaminated

Not mapped to the human reference sequence

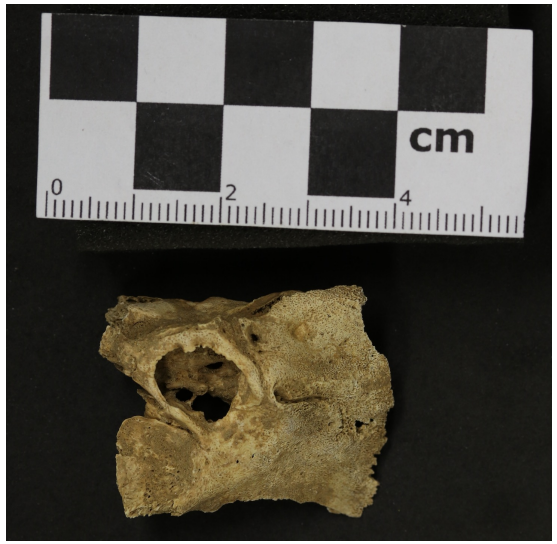
Controlled environment to prevent contamination

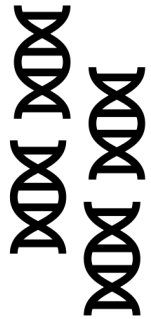
Sample collection





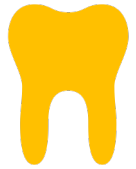
Tooth: relatively less DNA molecules, but greater chance to find ancient pathogens.





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

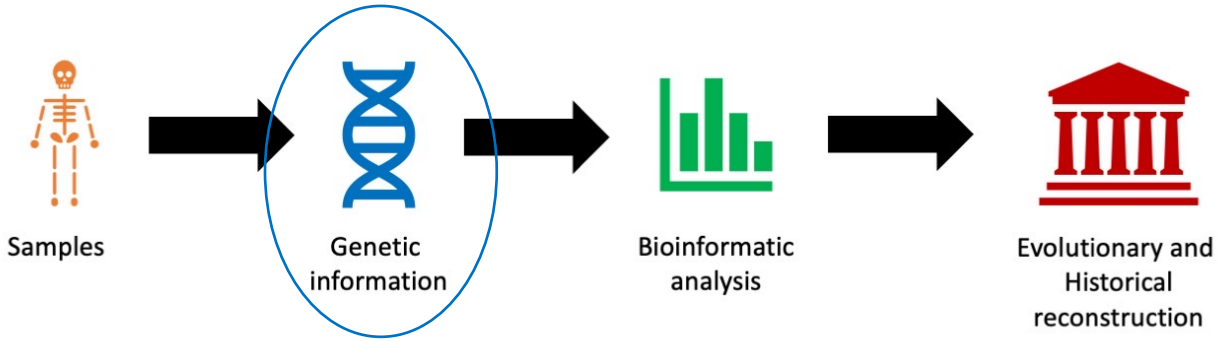
Sample collection

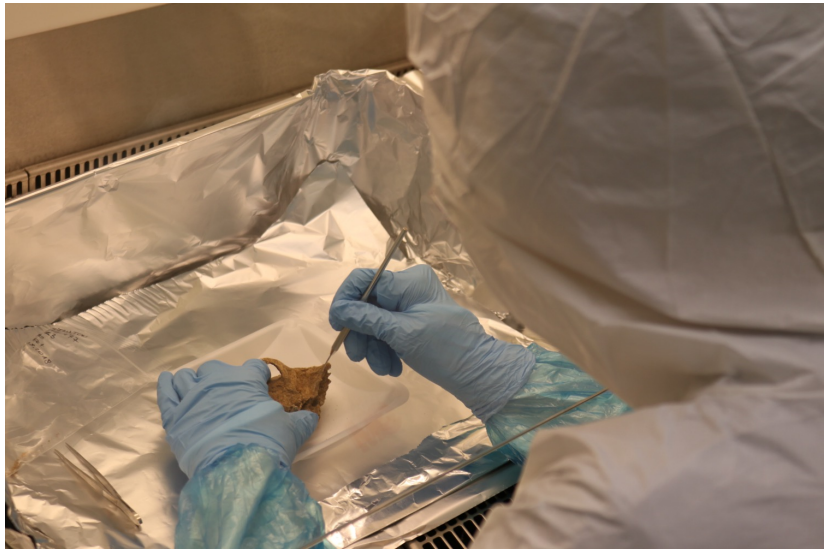
- More endogenous DNA in the petrous bone 
- Possibility to recover ancient pathogens from teeth 
- How destructive is the method? 
- Samples may be used for other analysis 

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

Drilling

Root tip for the teeth

Slices for the petrous bones

Decontamination

Washing samples with bleach to remove impurities from the sample surface

Extraction

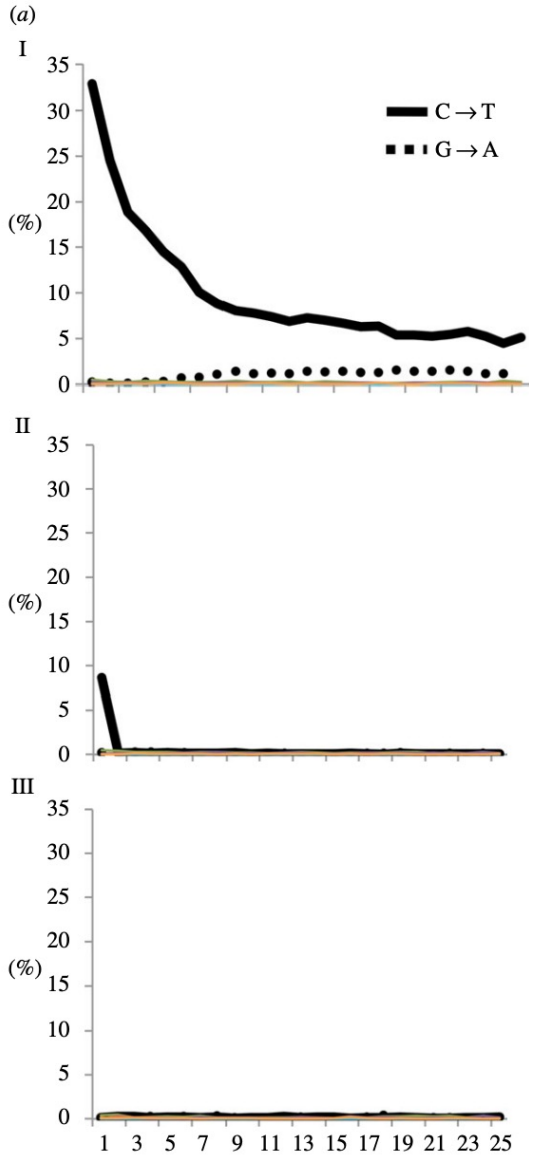
With EDTA (to remove calcium minerals) and proteinase K (to remove collagen fibers)

Purification

With silica columns

(or other methods)

UDG treatment (optional)



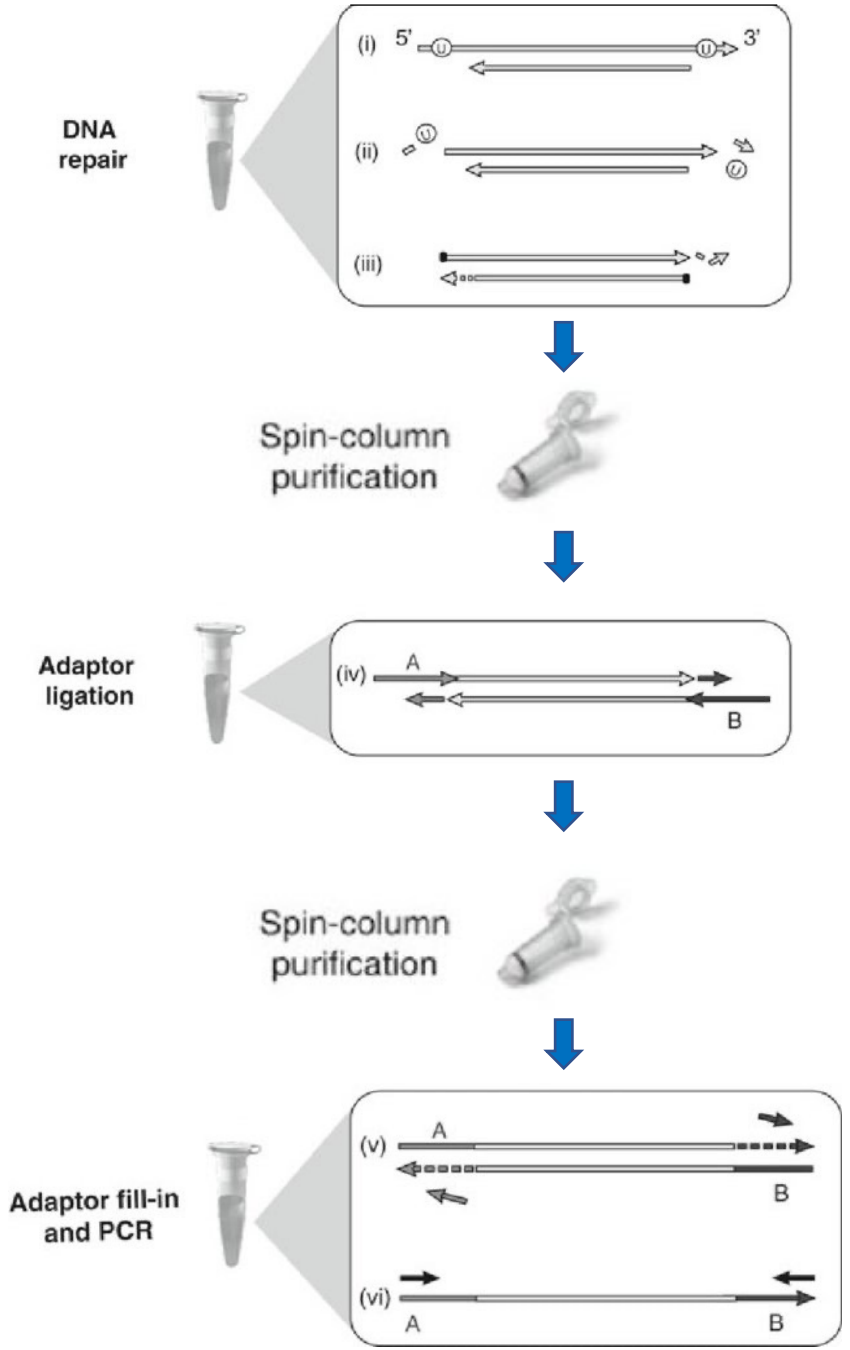
➡ No UDG treatment

➡ Partial UDG treatment

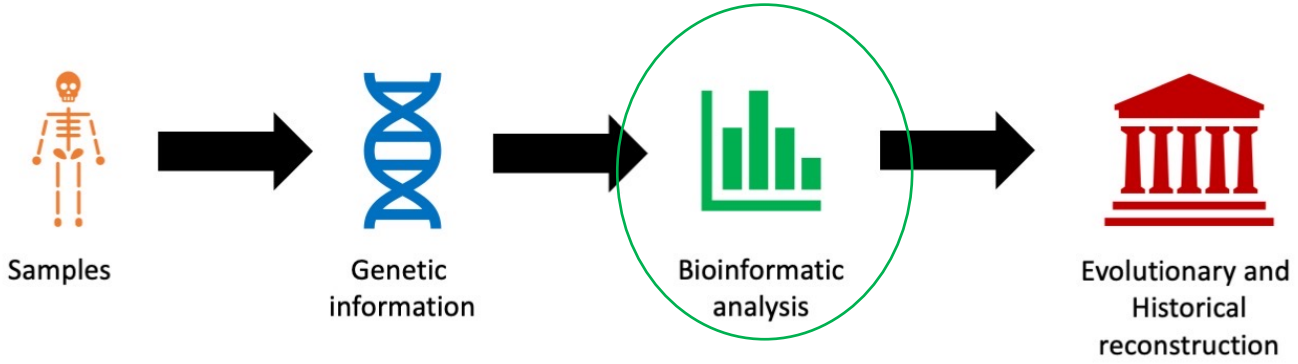
➡ Full UDG treatment

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
- Output: fastq files

A sequence identifier with information about the read

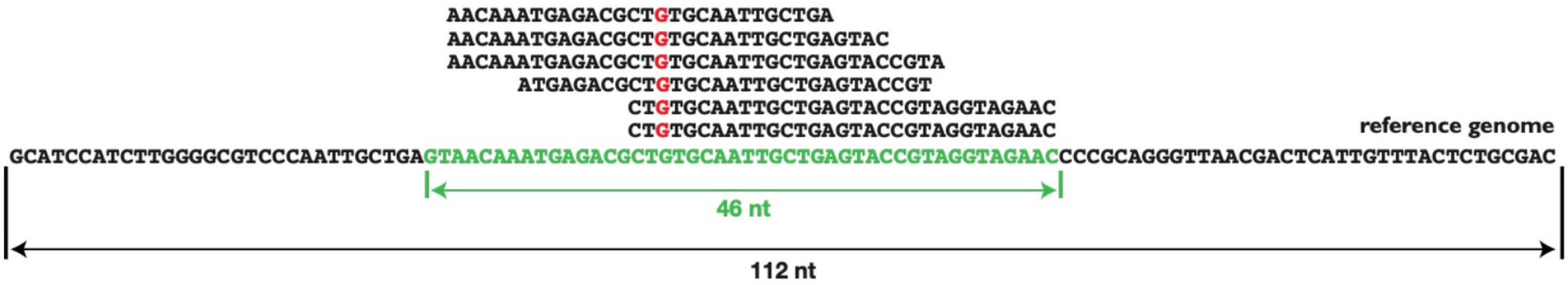
The sequence

```
@NB501163:37:HGK7WBGX7:1:11101:20397:1053 1:N:0:CGTACTAG+TACTCCTT
AGATCNGAAGAGCACACGTCTGAACTCCAGTCACCGTACTAGATCTCGTATGCCGTCTTCTGCTTGAAAAAAAAAA
+
AAAAA#/EEE/6EEEEEEAEE6AEEAEE/EEA/EEEEEEAEEEA//EAEEEEEEEAEEEA/EEEA/EAEEEE/E//
```

Base call quality scores

Mapping

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

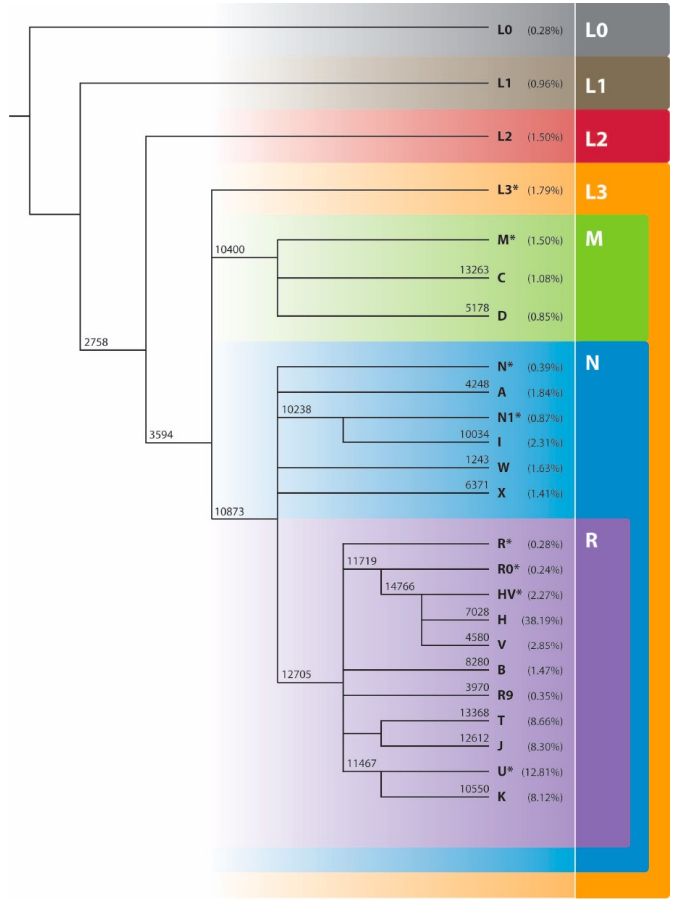


Authentication

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

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

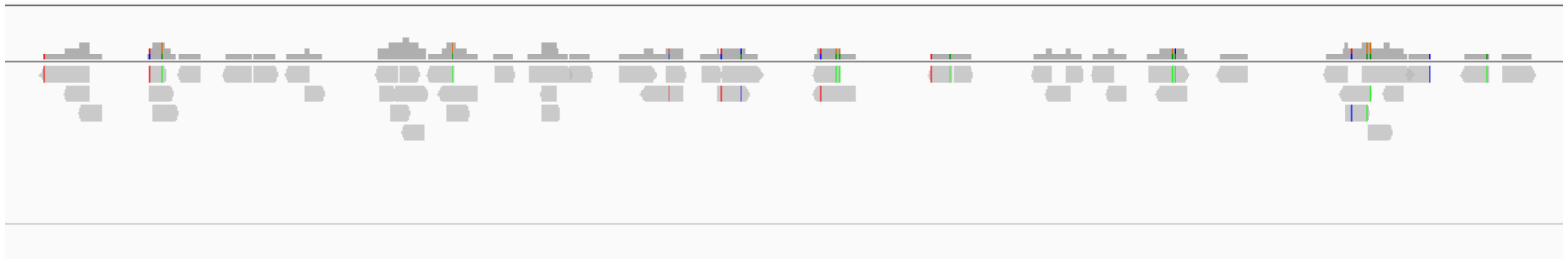
- **Contamination**
  - X-based method (only for male samples)
  - mtDNA method (Calculating the percentage of non-consensus bases at haplogroup-defining positions)



Variant calling

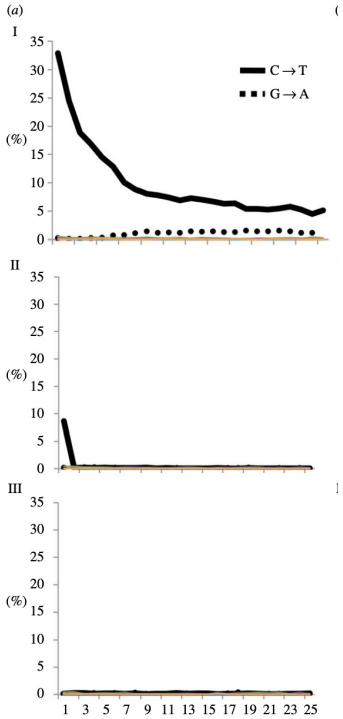
Variant type:

- 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

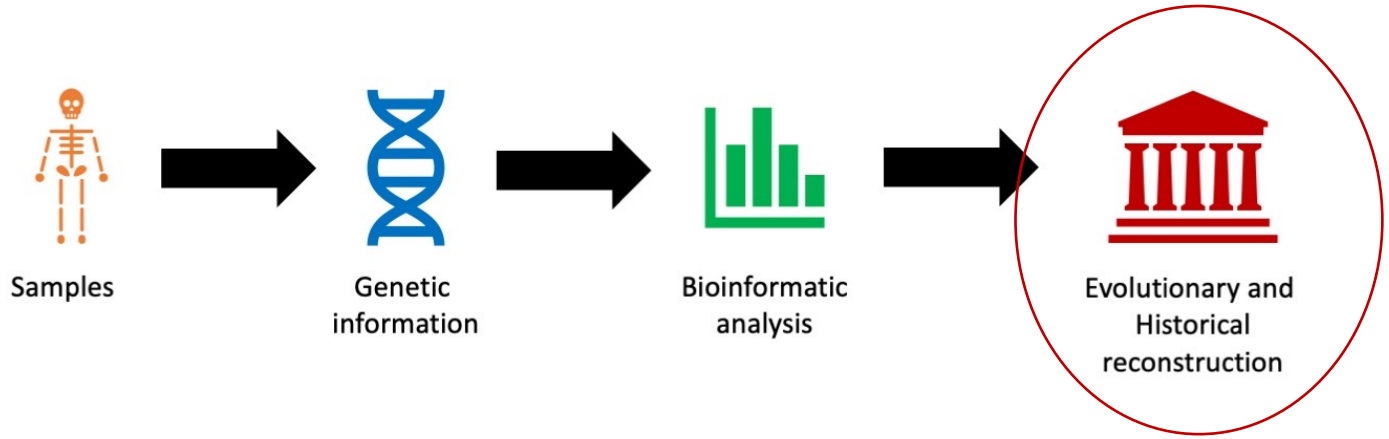


No UDG treatment

Partial UDG treatment

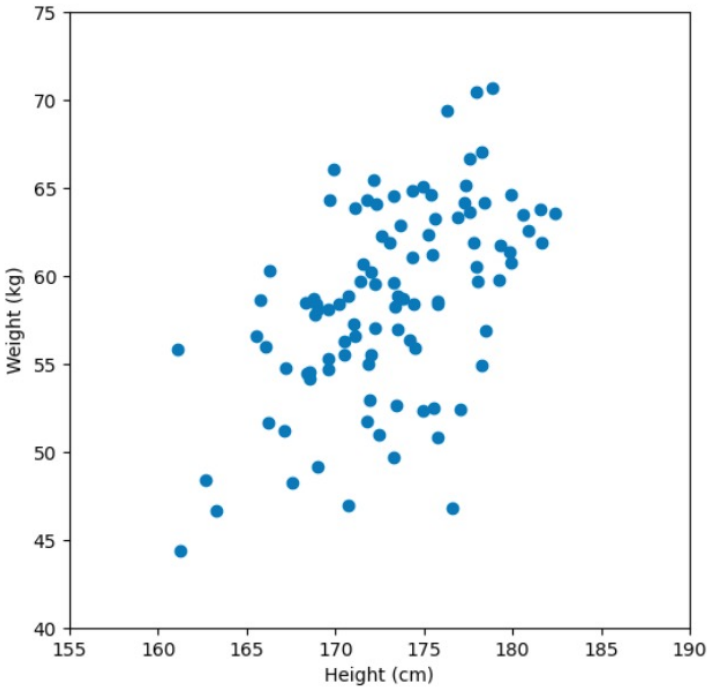
Full UDG treatment

Population genetics analysis for aDNA data

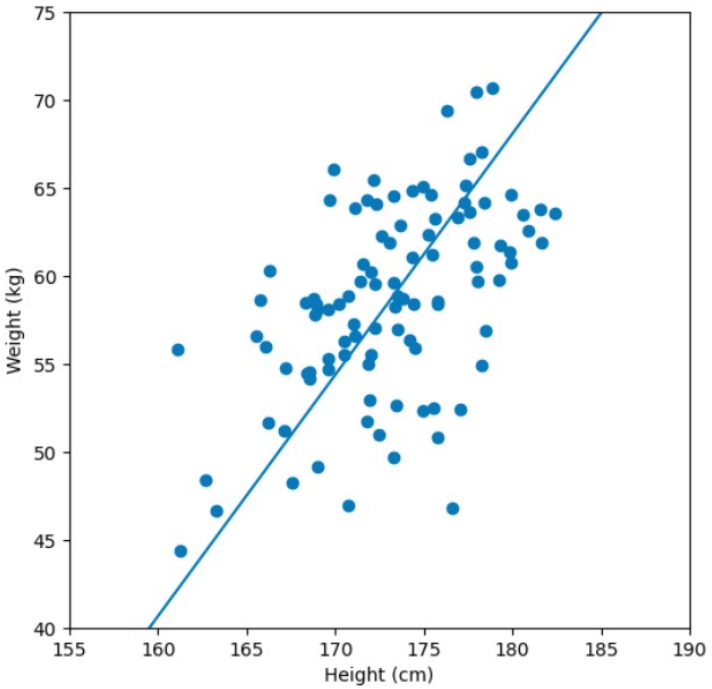


Principal Component Analysis (PCA)

- 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



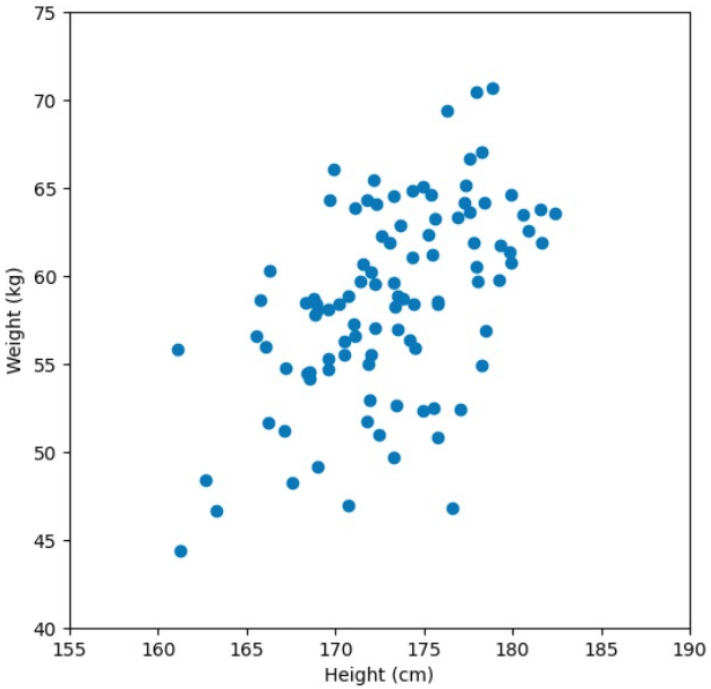
Each individual (point) is represented by two variables.



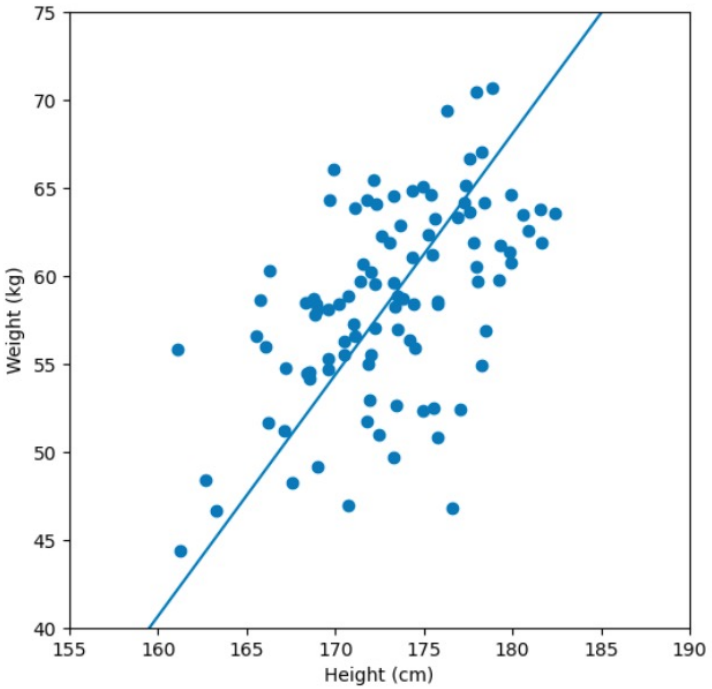
Find the axis of greatest variation (fit line) —> The principal component.

Principal Component Analysis (PCA)

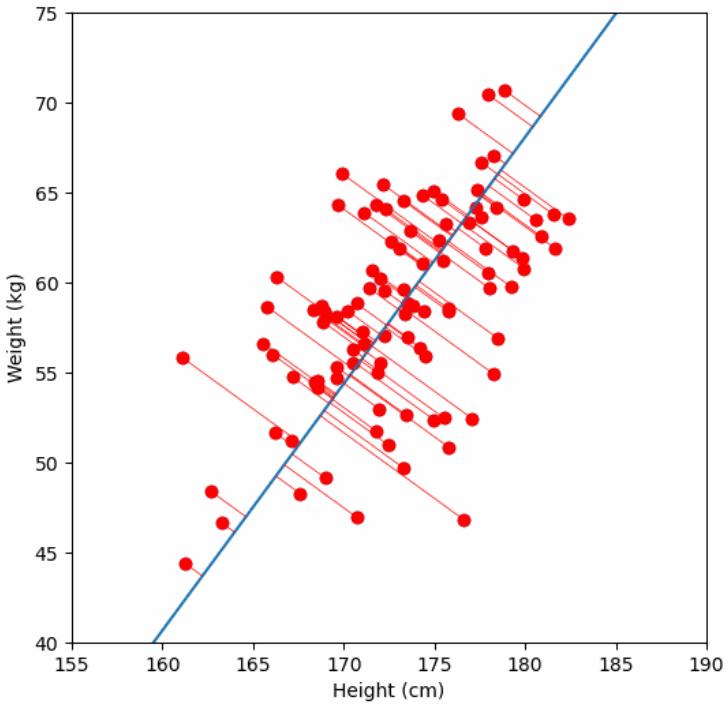
- PCA is a linear transformation to a new coordinate system
- **Reduction of dimensions**: the genetic information contained in 1M SNPs can be summarized by a few new variables



Each individual (point) is represented by two variables.



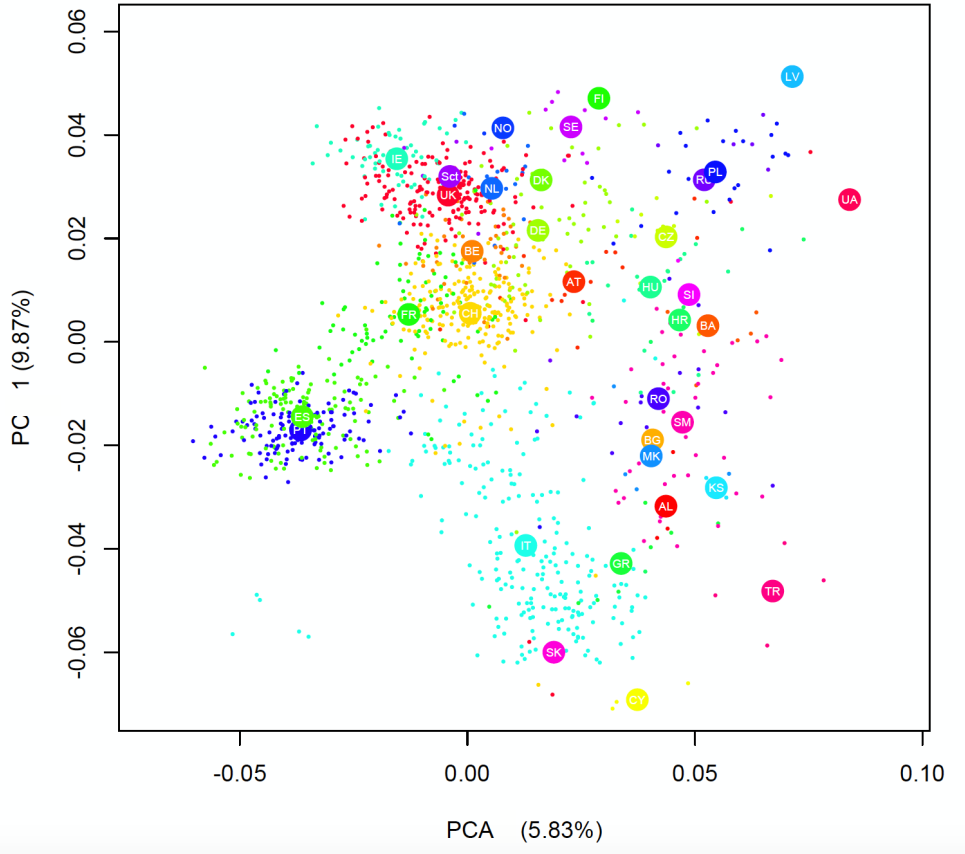
Find the axis of greatest variation (fit line) —> The principal component.



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

Principal Component Analysis (PCA)

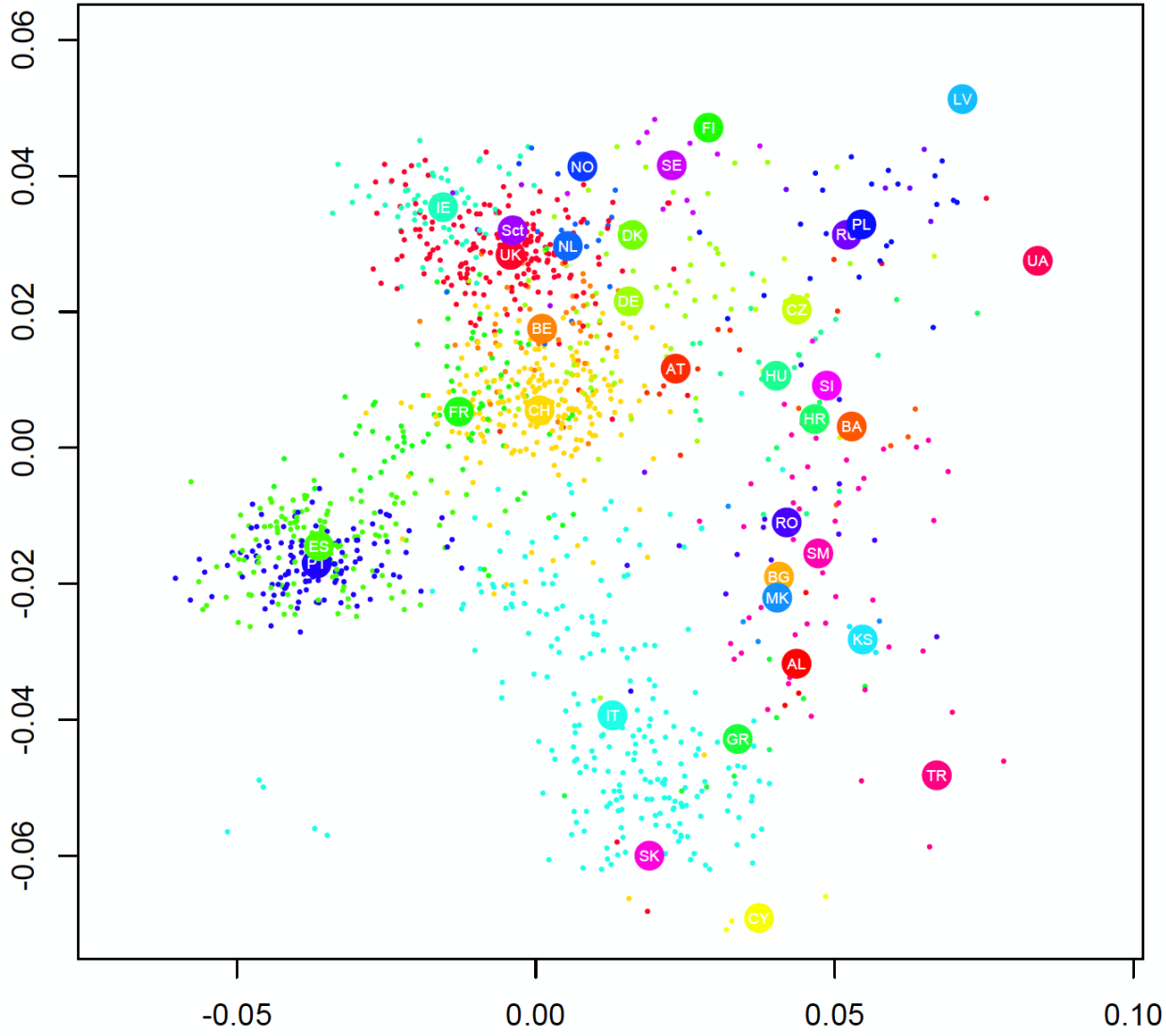
	PC1	PC2
Ind(1): 0101110110101110	0.01	-0.02
Ind(2): 0111110110101111	0.50	0.03
Ind(3): 0100110110101011	0.07	-0.13
Ind(4): 0111111111101111	0.02	-0.04
Ind(5): 0101110110100001	0.01	-0.05
.	.	
.	.	
.	.	
Ind(n): 0101110110101111	-0.03	0.03



Principal Component Analysis (PCA)

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

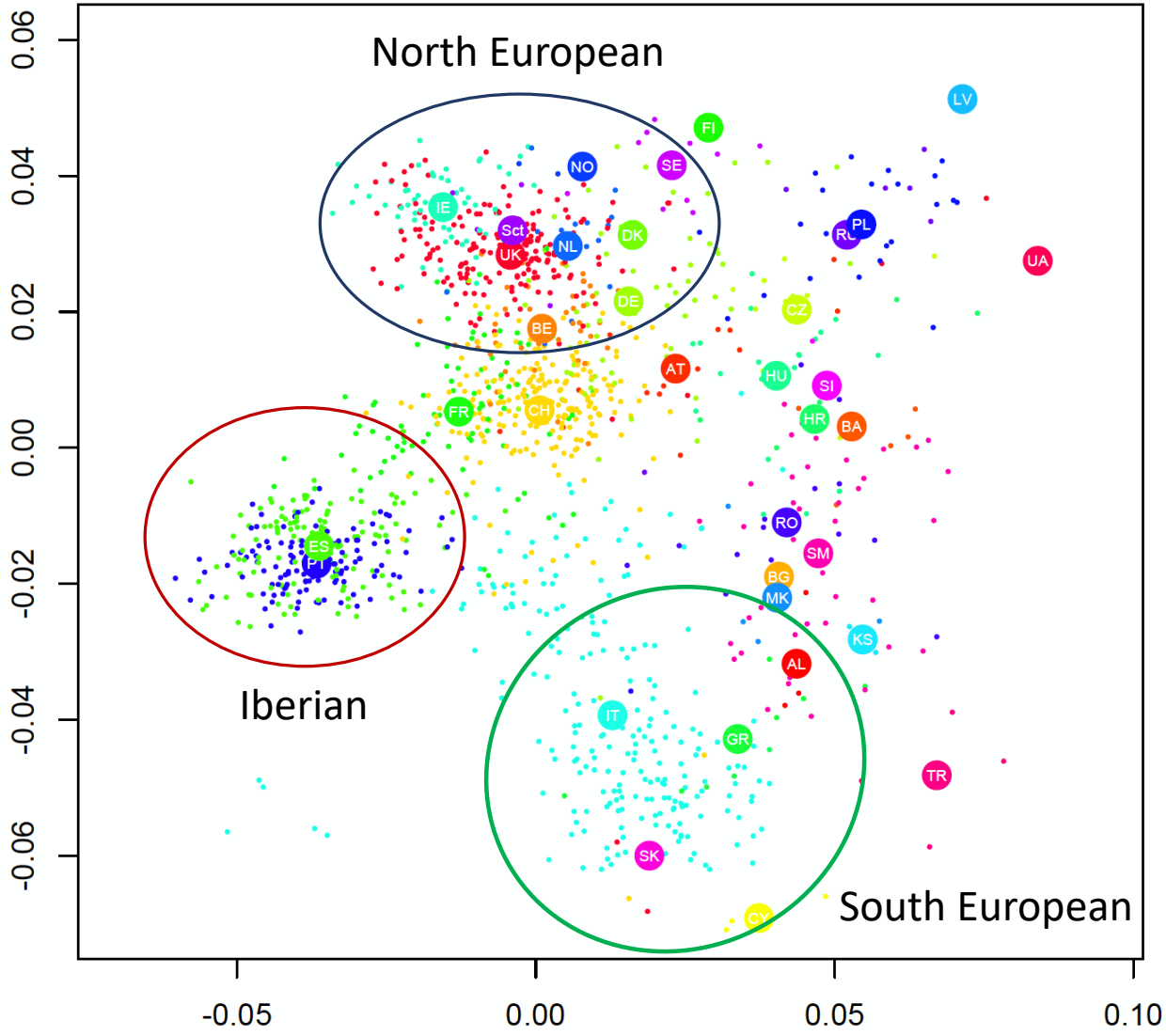
Novembre et al. 2009



Principal Component Analysis (PCA)

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

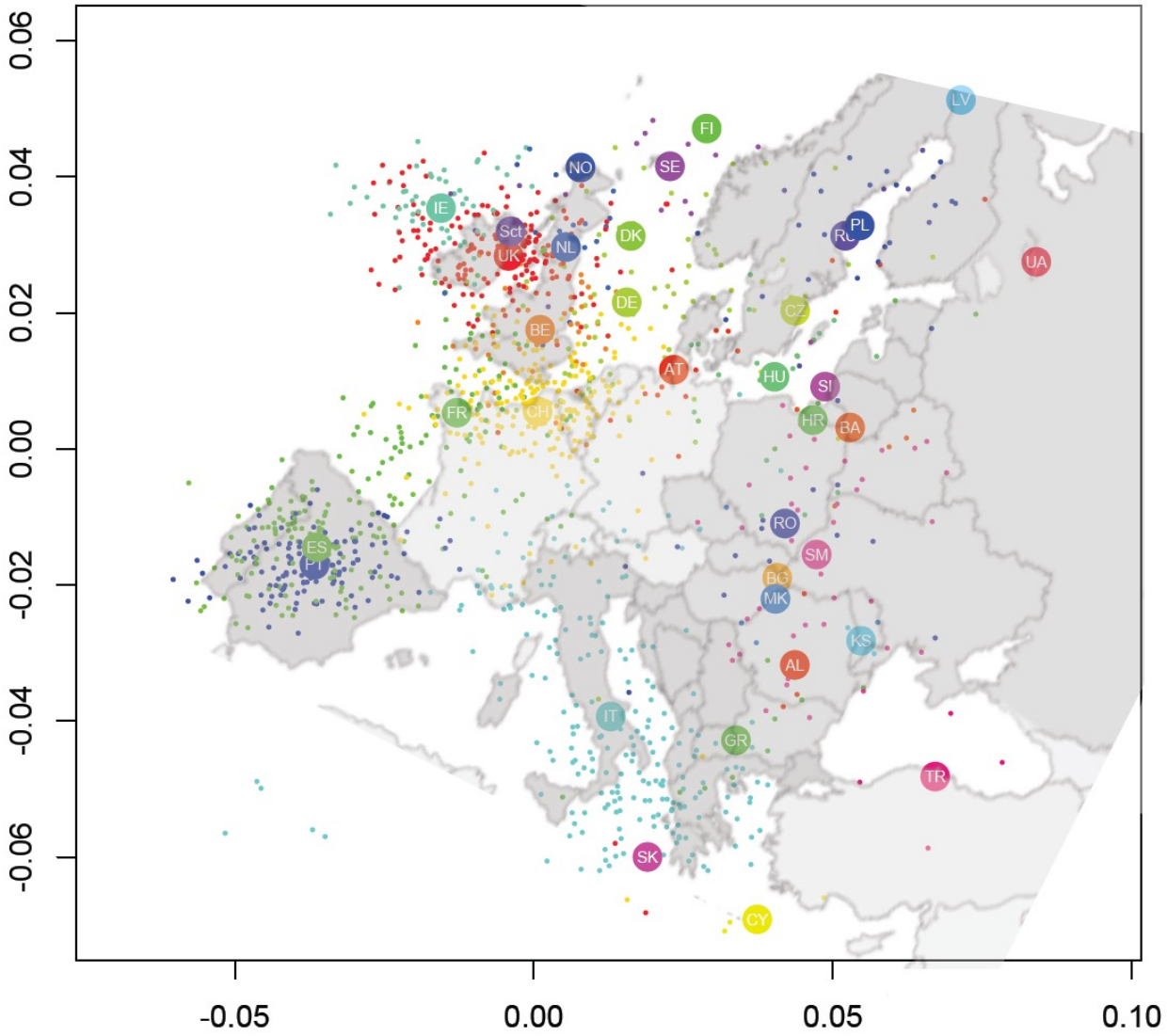
Novembre et al. 2009



Principal Component Analysis (PCA)

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

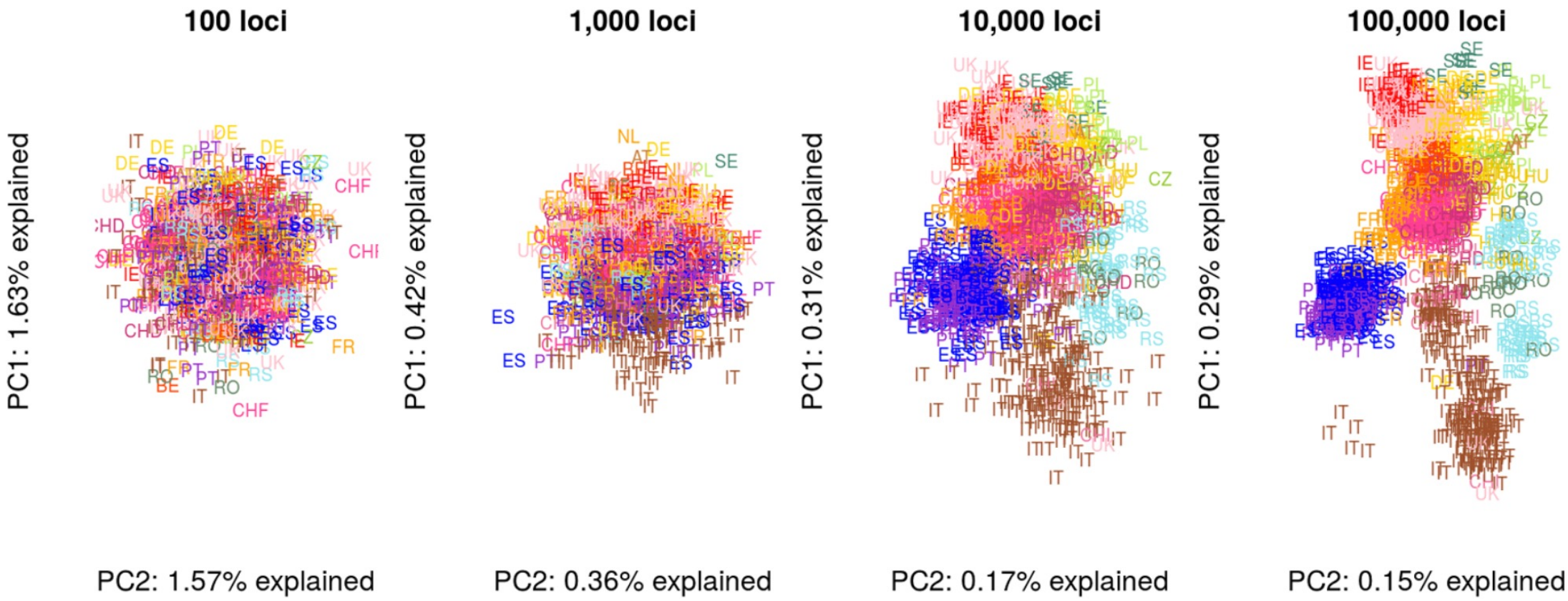
Novembre et al. 2009



Principal Component Analysis (PCA)

Produce good results even when the information is low

Novembre et al. 2009

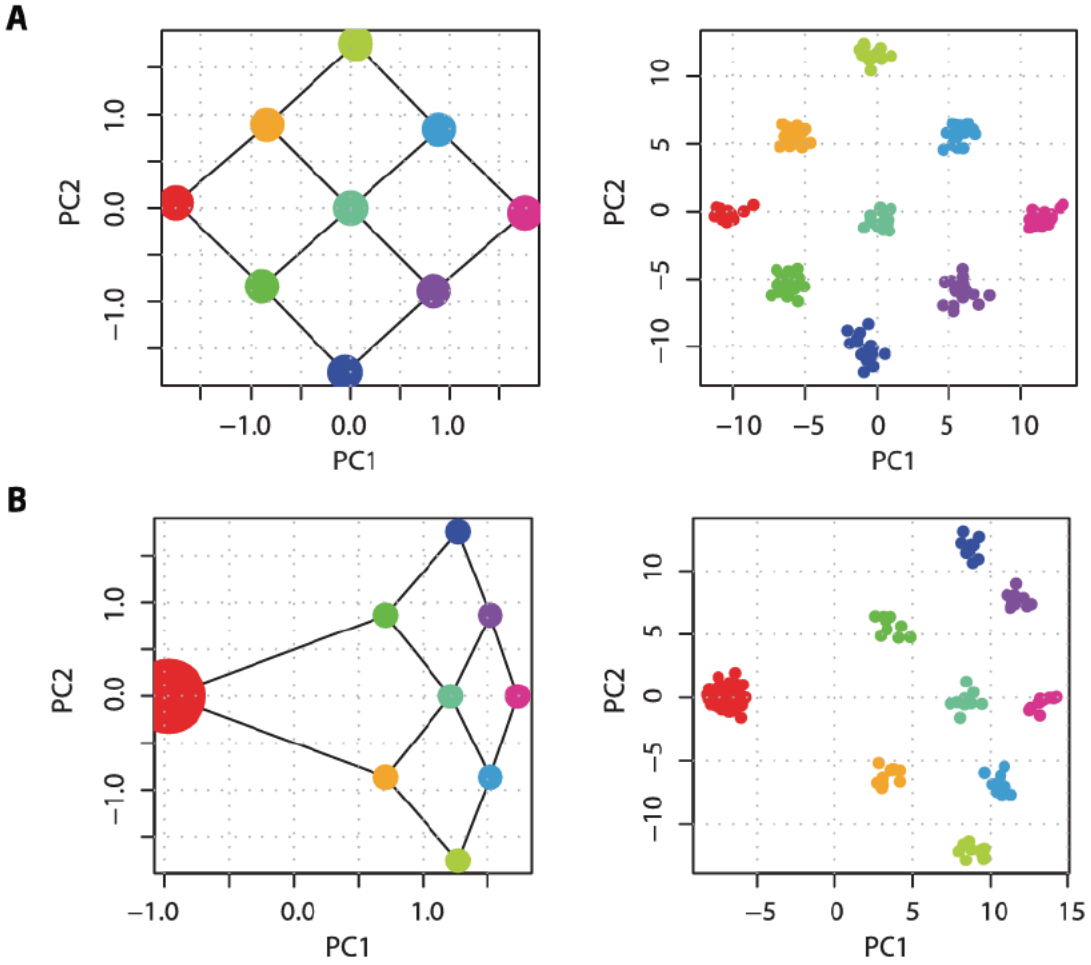


Principal Component Analysis (PCA)

Factors that influence PCA:

- Migration
- Genetic drift
- Admixture

- Population size
- SNP selection



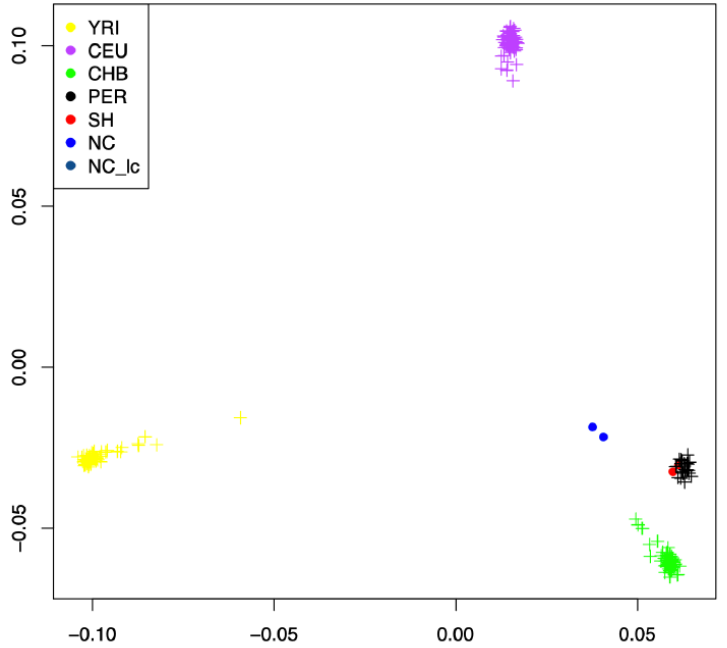
PCA with ancient samples

Low coverage individuals result in many SNPs with missing data

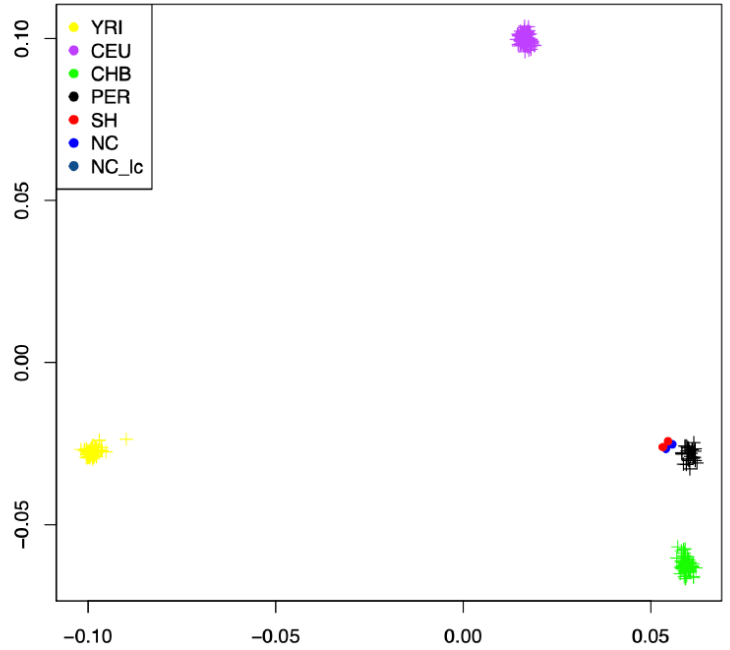
Usually, PCA methods will fill in all missing data. This results in PCA plots that have ancient individuals near/at the origin (0,0 coordinate).

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

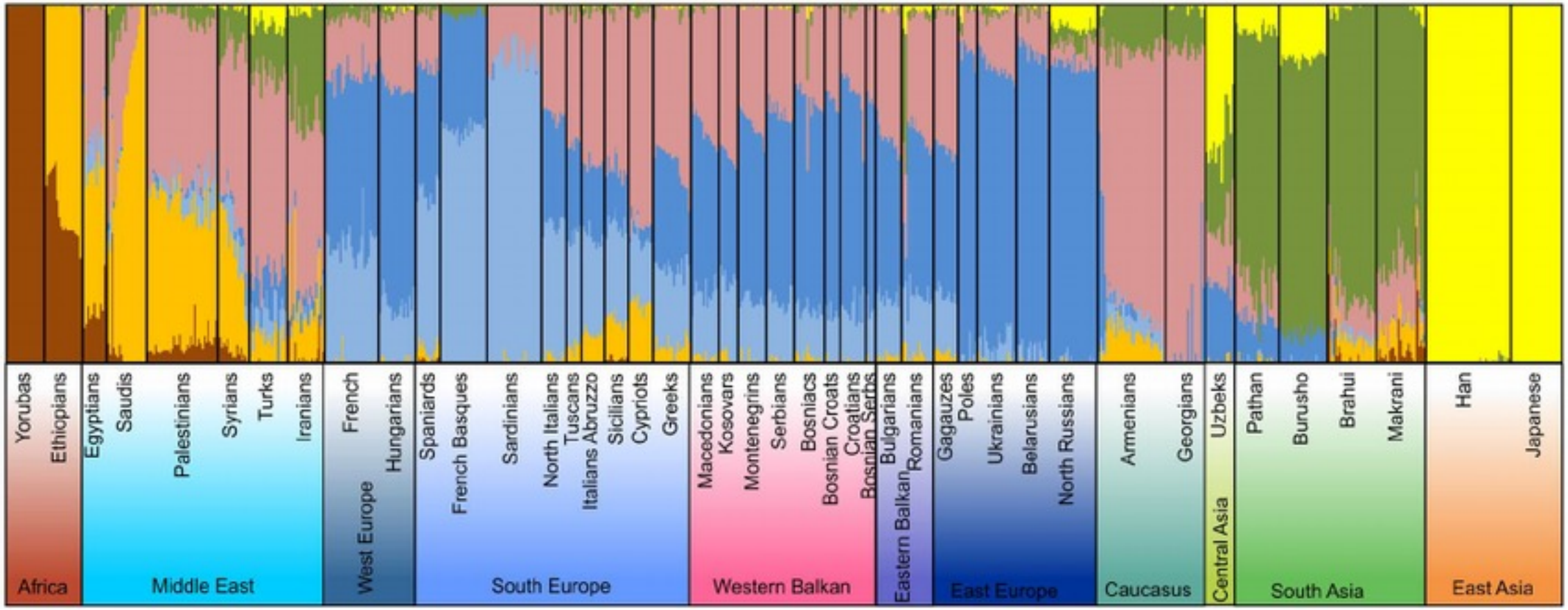
Not projected



Projected



Ancestry proportion inference (ADMIXTURE)



Allele frequency-based clustering

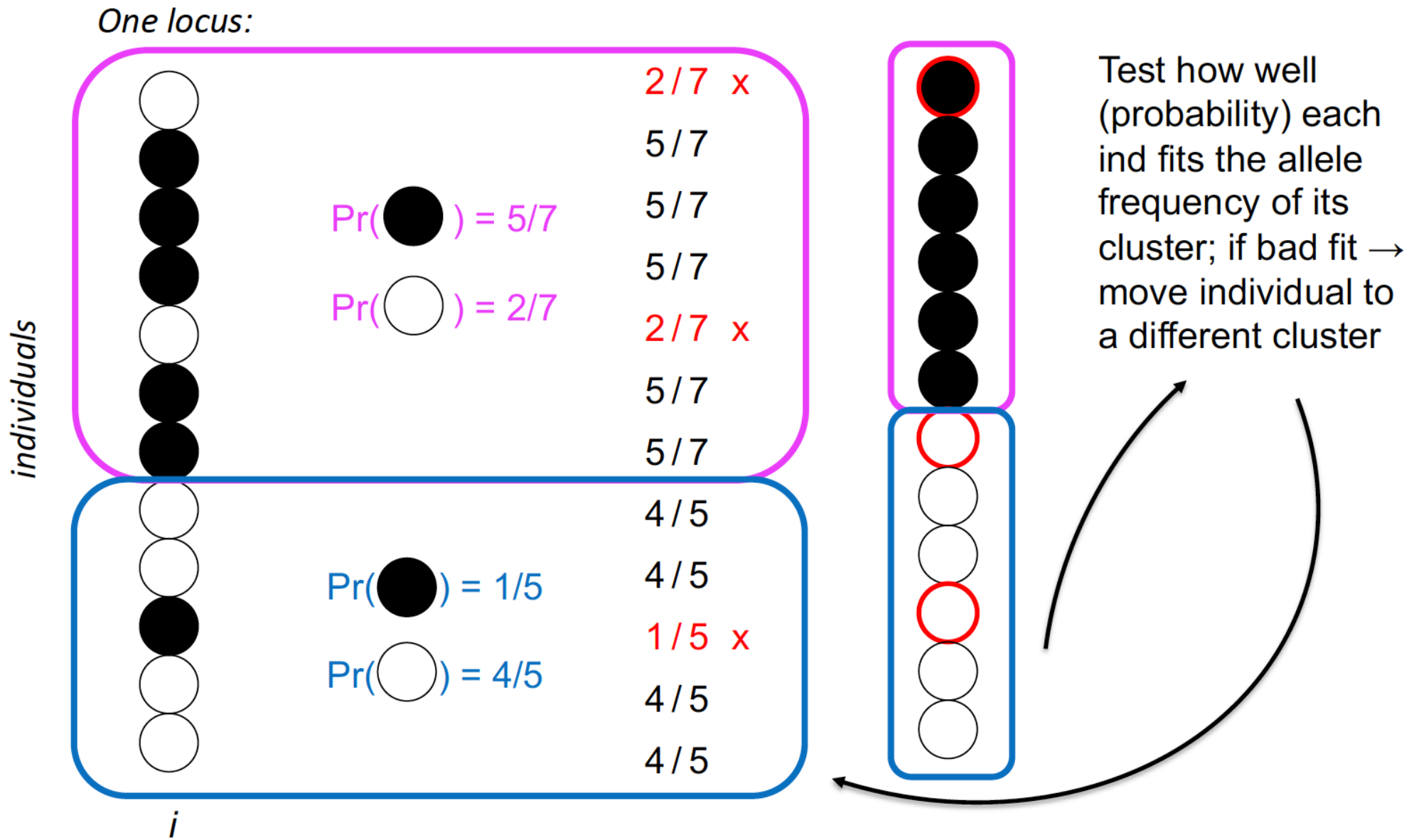
Thought experiment:

- Assume we sequenced individuals we knew are from Pop 1 and Pop 2
- We now sequence another individual, where we are unsure whether they are from Pop1 or Pop 2
- How could we try to assign this individual to Pop1 or Pop 2?

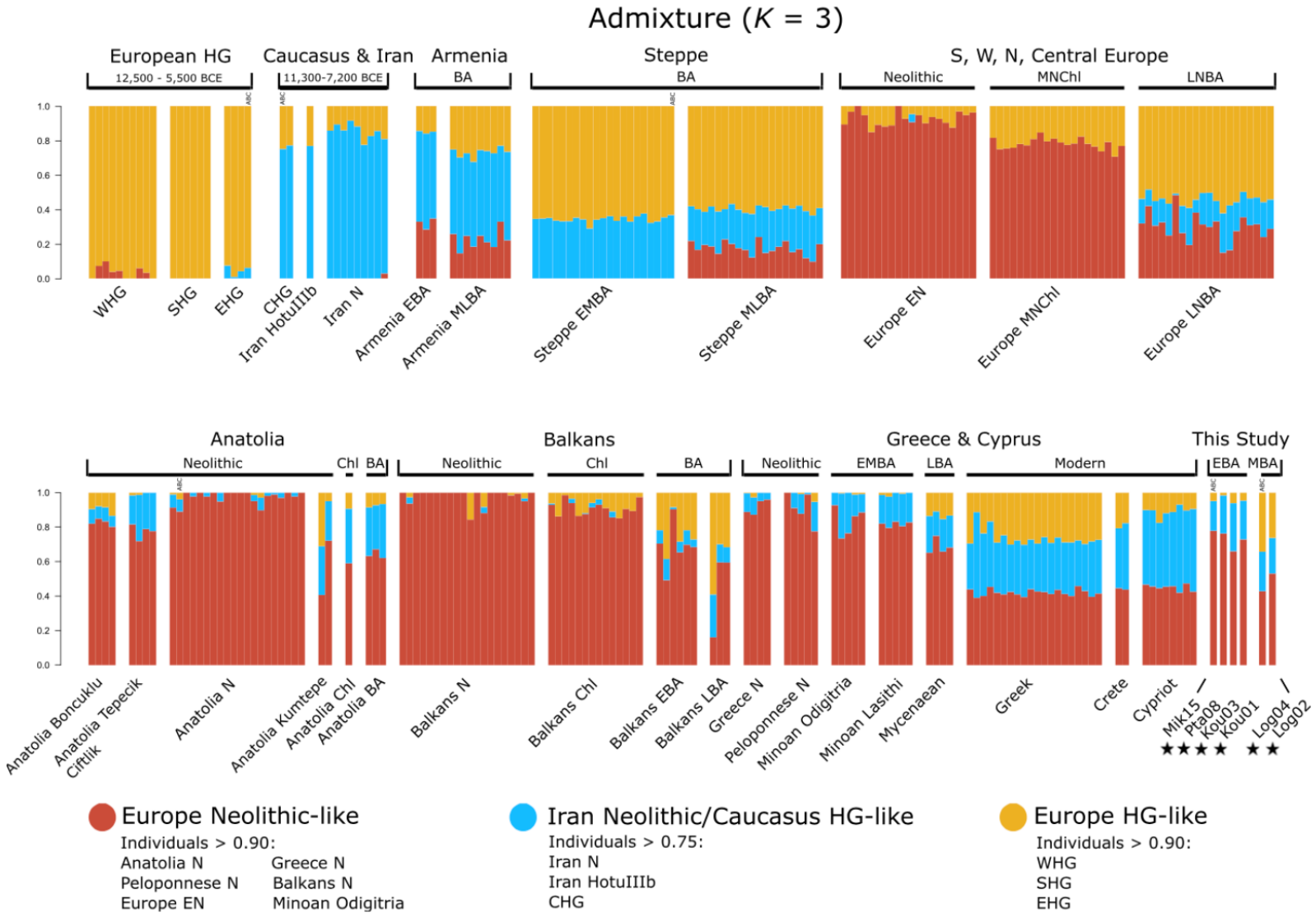
	Allele frequency in Pop1	Pop2	Genotype of individual	
SNP1	0.8	0.4	1	→ Pop1?
SNP2	0.3	0.7	0	→ Pop1?
SNP3	0.4	0.6	1	→ Pop2? (Pop1?)
SNP4	0.9	0.1	1	→ Pop1?

Allele frequency-based clustering

We can do the same **without** knowing allele frequencies in Pop1 and Pop2 by clustering



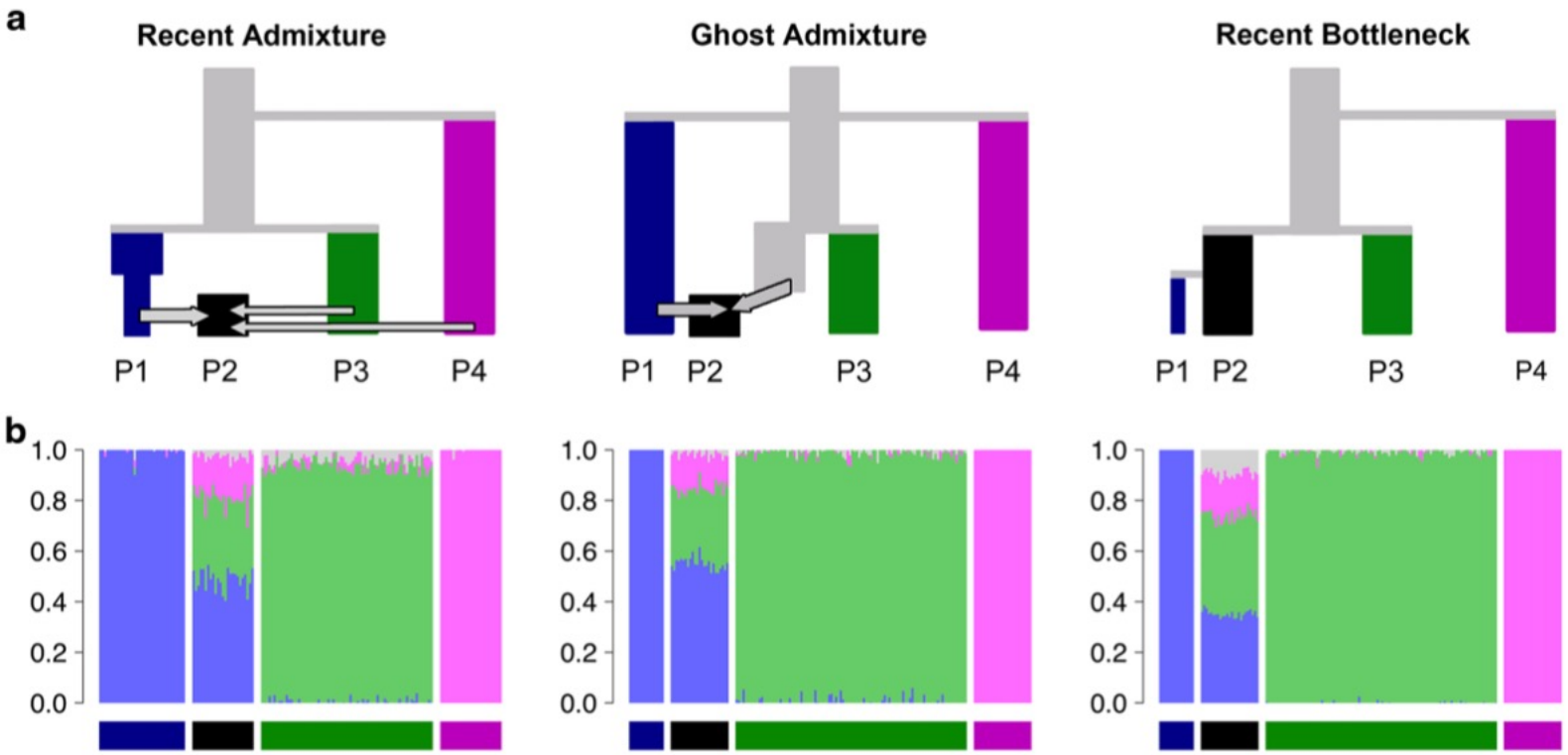
Ancestry proportion inference (ADMIXTURE)



Clemente et al. 2021 *Cell*

Ancestry proportion inference (ADMIXTURE)

Be careful when interpreting ADMIXTURE results!



Ancestry proportion inference (ADMIXTURE)

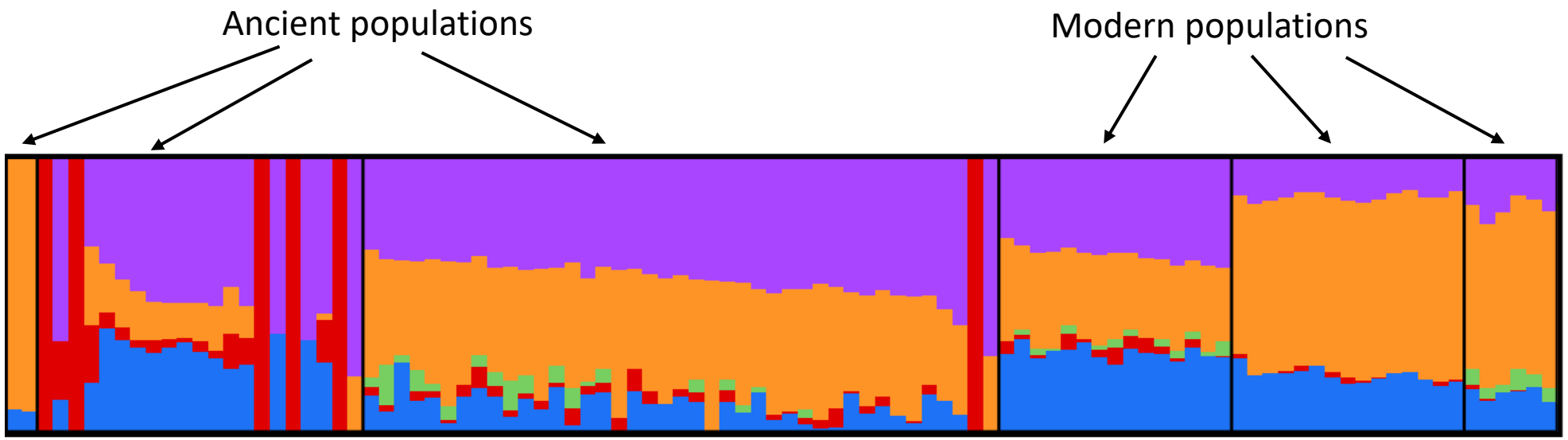
Be careful when interpreting ADMIXTURE results!

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

Ancestry proportion inference (ADMIXTURE)

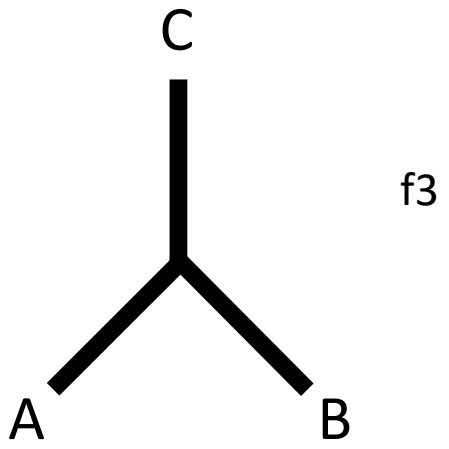
Be careful when interpreting ADMIXTURE results!

Possible problem with low coverage samples

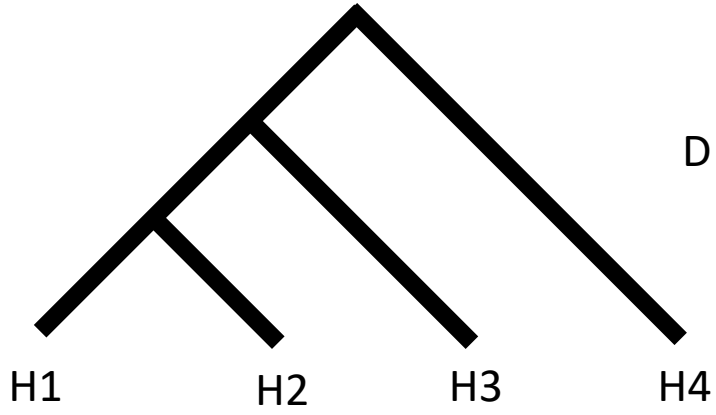


Tests of "treeness" – *f* and Patterson's *D* statistics

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



f3



D

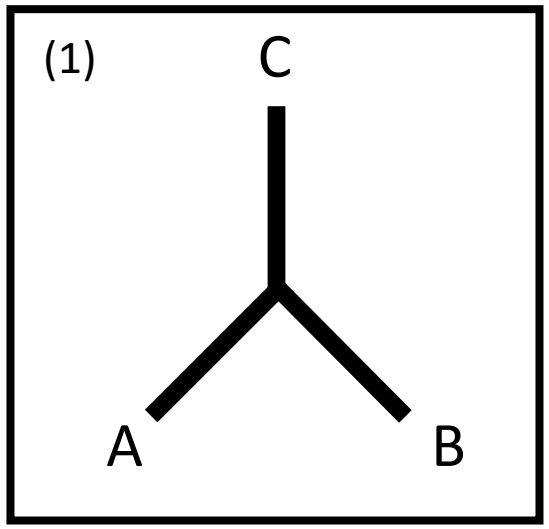
f3 statistic

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

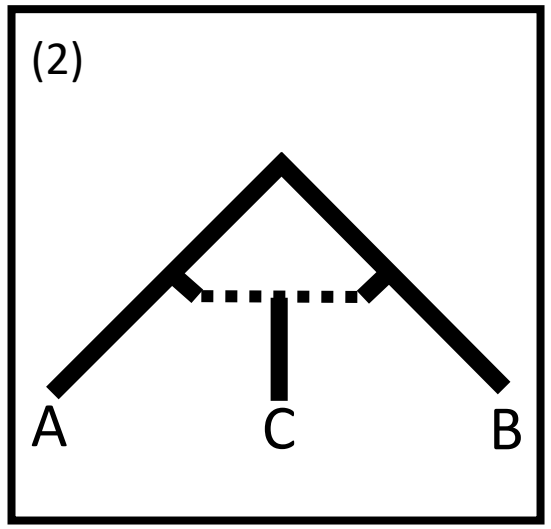
Two main purposes:

- 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



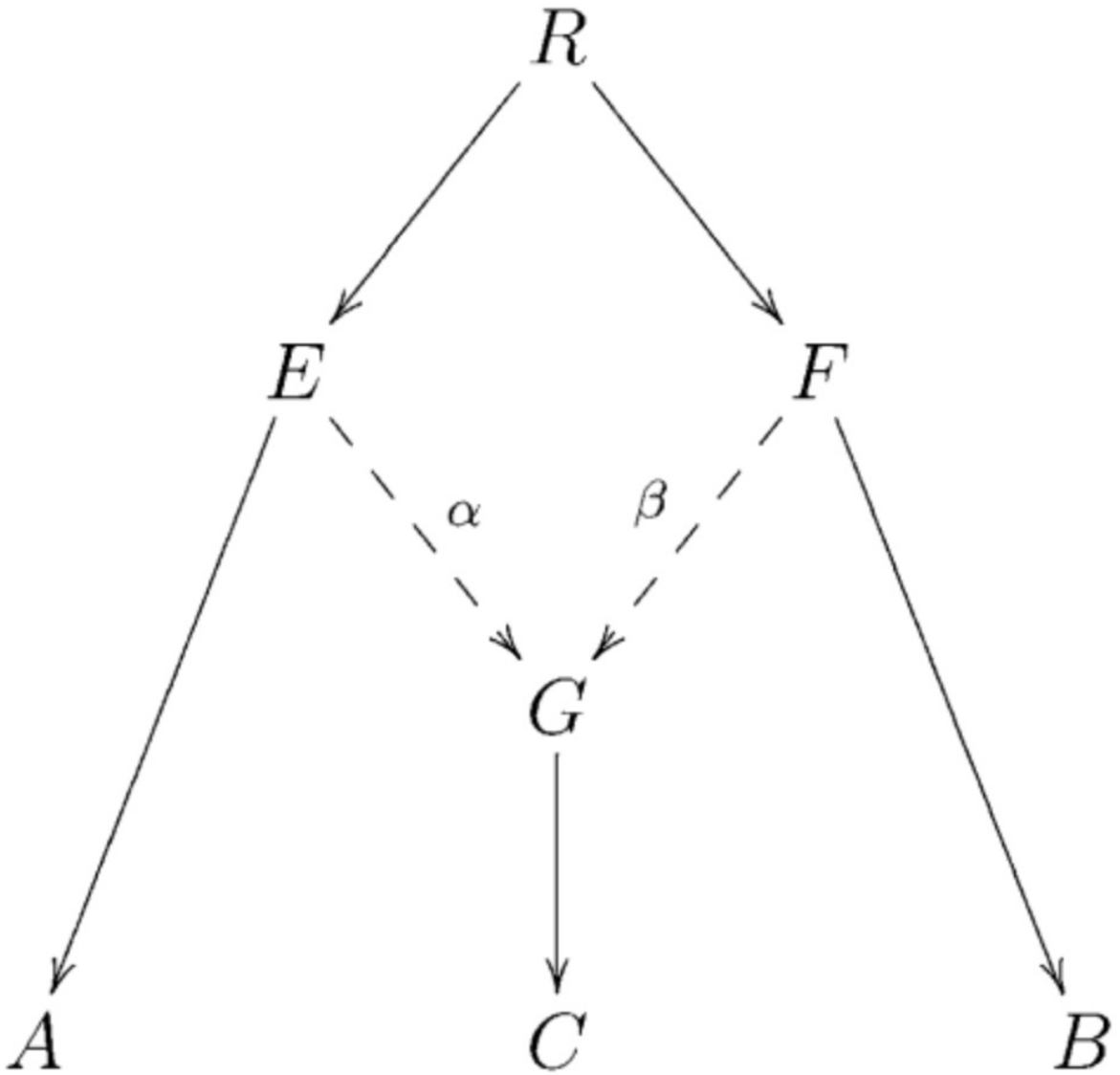
f3 < 0



f3 statistic

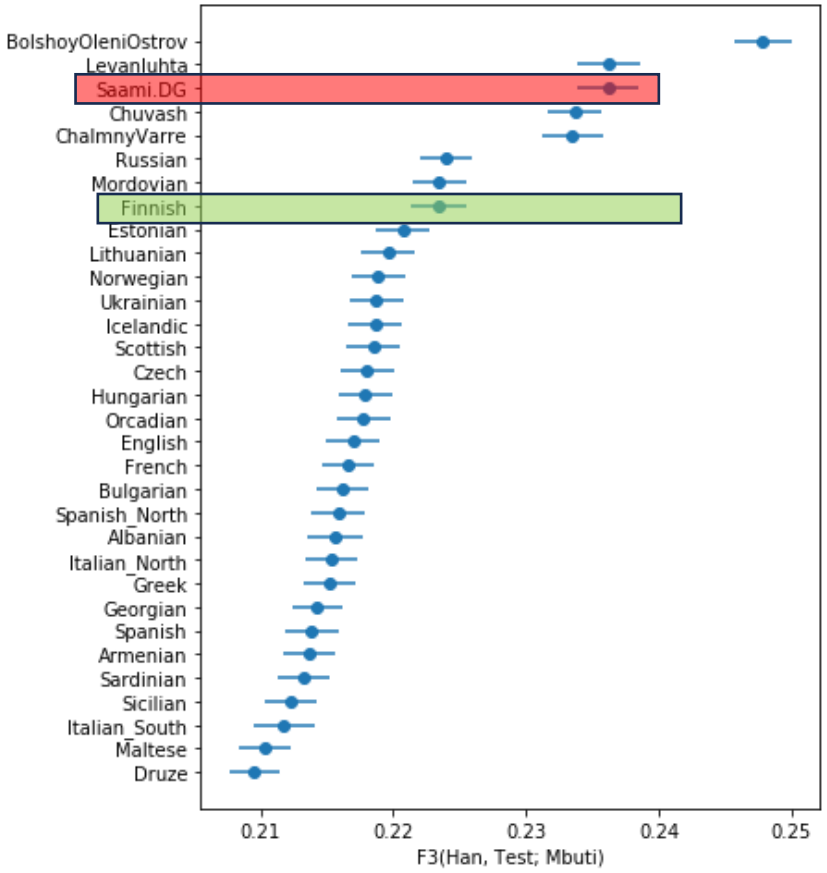
$$f3(C;A,B) = \frac{1}{J}\sum_{j=1}^{J}(c_j - a_j)(c_j - b_j)$$

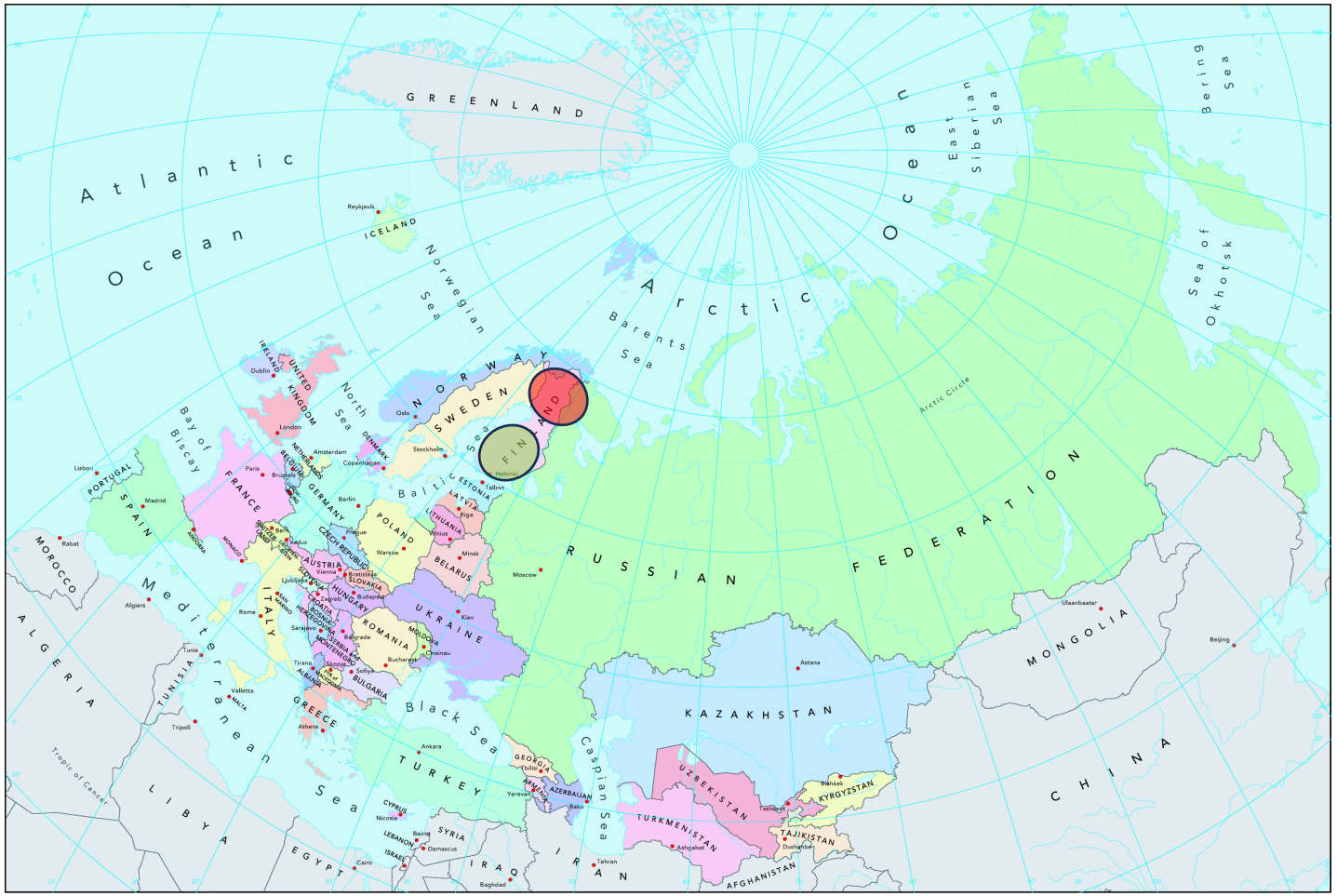
f3 < 0



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 denotes Han Chinese, and X denotes various European populations





Target *f3* statistic – Example

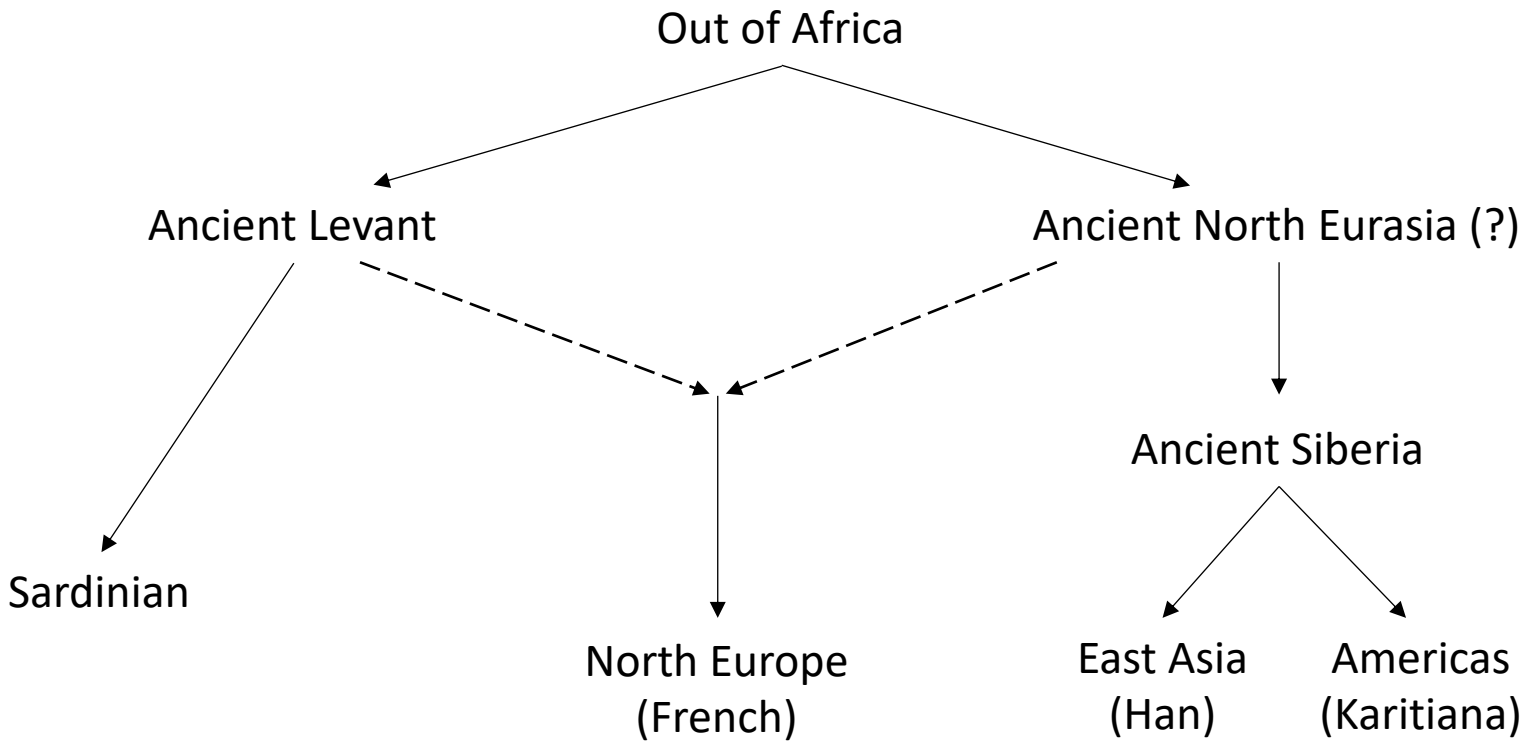
We can use target *f3* to better understand what is the genetic relationship between East Asia and Europe (and the Americas)

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

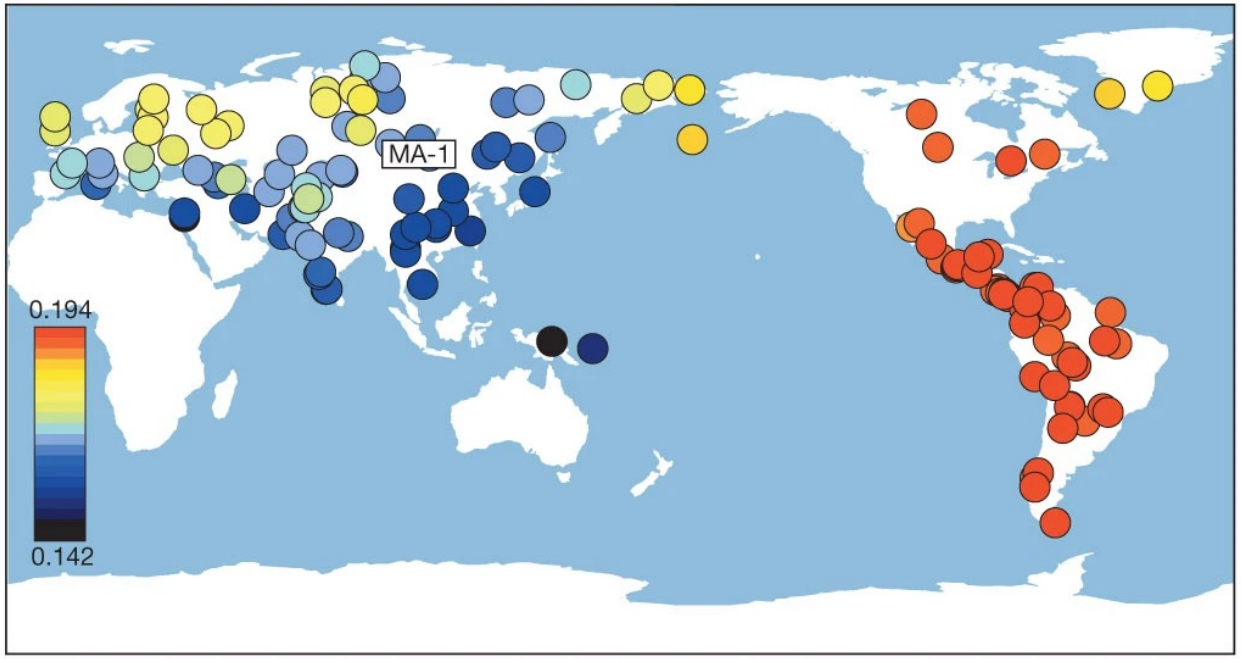
We can use target *f3* to better understand what is the genetic relationship between East Asia and Europe (and the Americas)

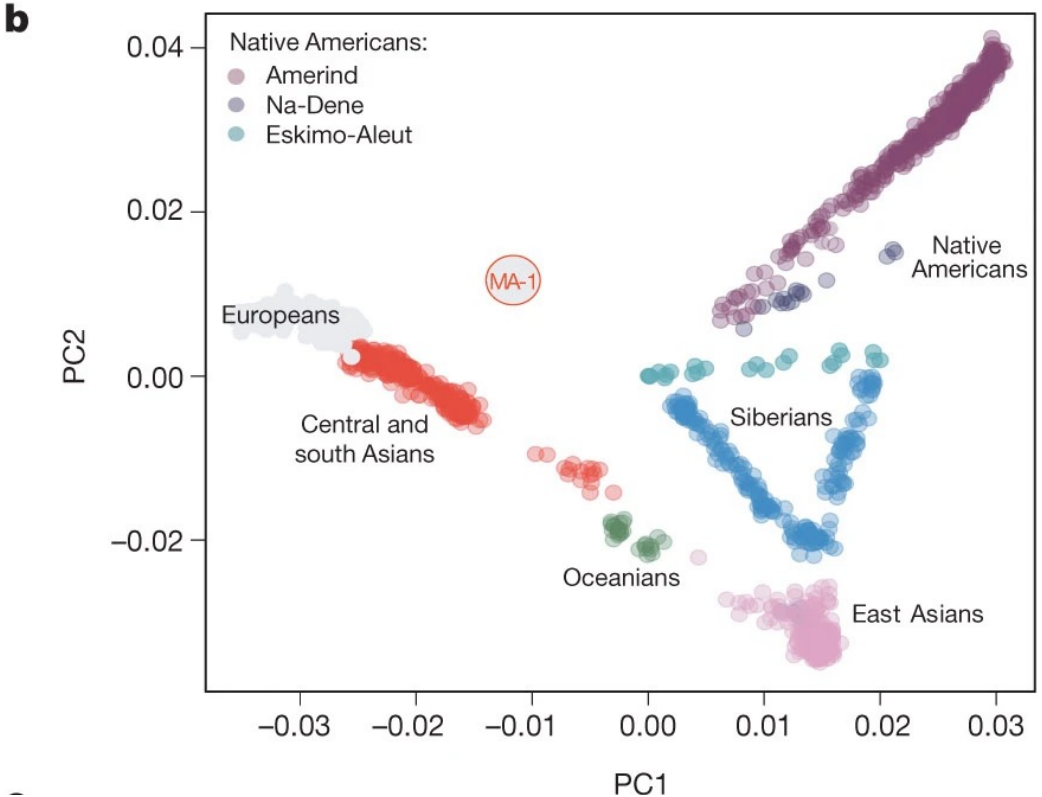
Source1	Source2	Target	f_3	*Z*-score
Japanese	Italian	Uygur	−0.0259	−74.79
Japanese	Italian	Hazara	−0.0230	−74.05
Yoruba	Sardinian	Mozabite	−0.0211	−56.95
Mozabite	Surui	Maya	−0.0149	−19.67
Yoruba	San	Bantu-SA	−0.0107	−31.39
Yoruba	Sardinian	Palestinian	−0.0107	−36.70
Yoruba	Sardinian	Bedouin	−0.0104	−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

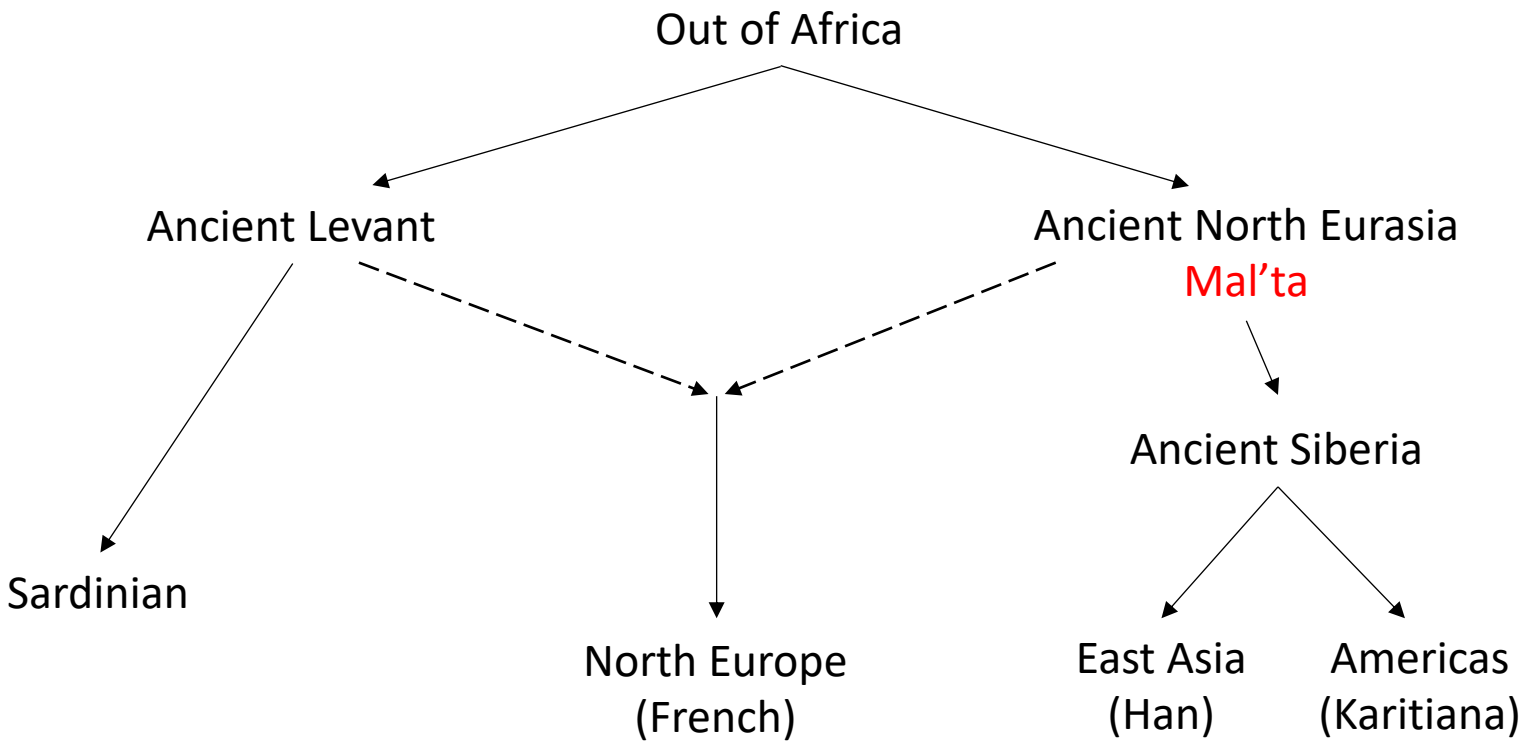


Ancient North Eurasia

24,000-year-old individual (MA-1) from Mal’ta

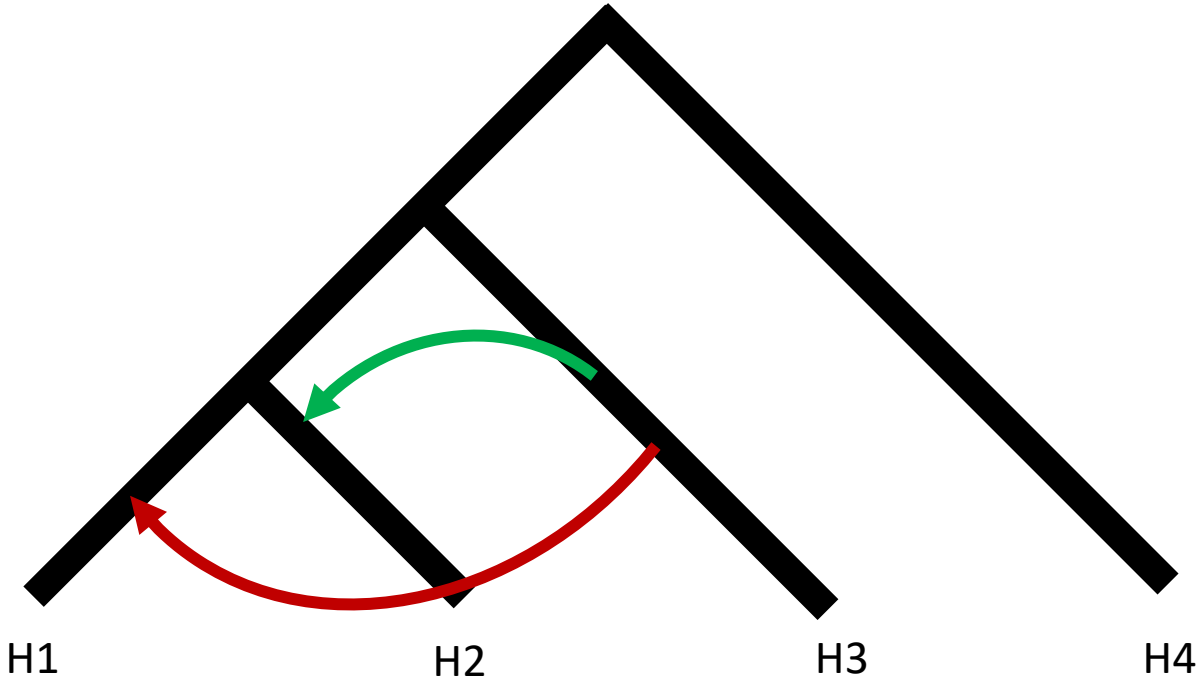





Out of Africa
Ancient Levant
Ancient North Eurasia
Mal'ta
Ancient Siberia
Sardinian
North Europe
(French)
East Asia
(Han)
Americas
(Karitiana)

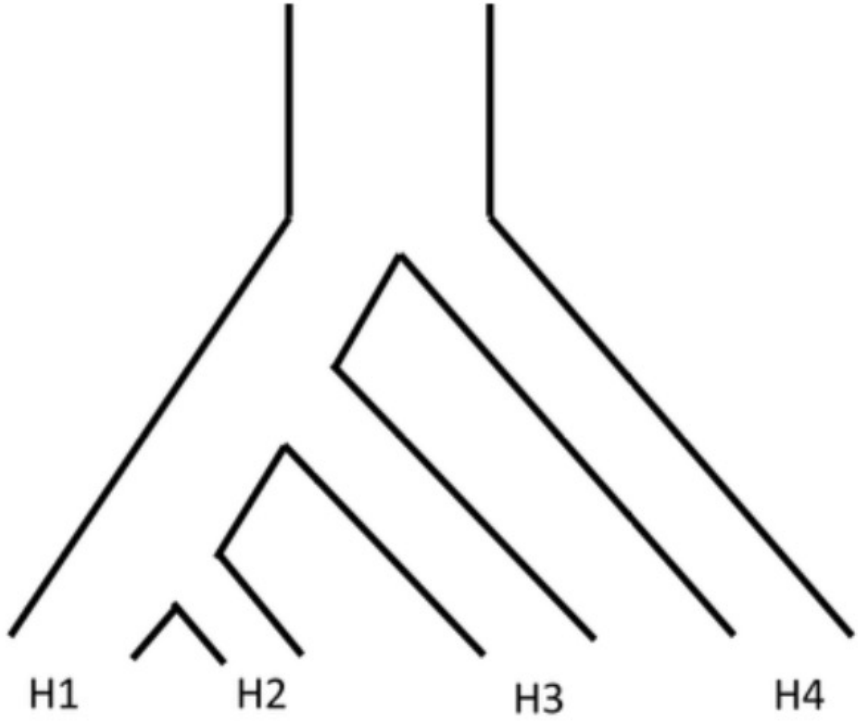
D statistic

Detect signature of admixture between populations



D statistic

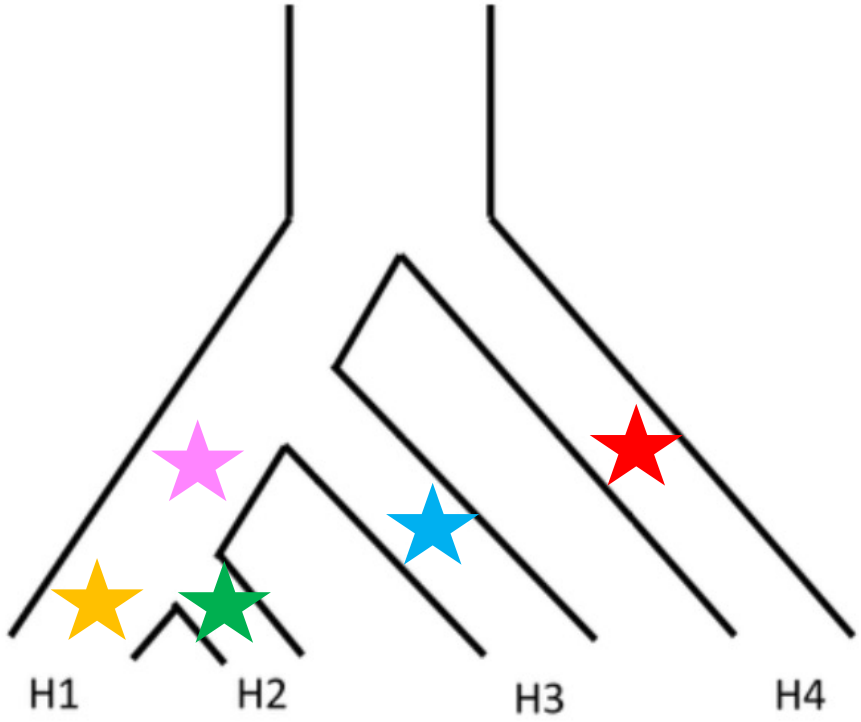
- Analyse a tree with four population
- Pick one individual for each population (it can be performed also with the whole population)
- Look at a polymorphic site – “A” is the ancestral state and “B” is the derived one
- Possible observable pattern of allele sharing



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

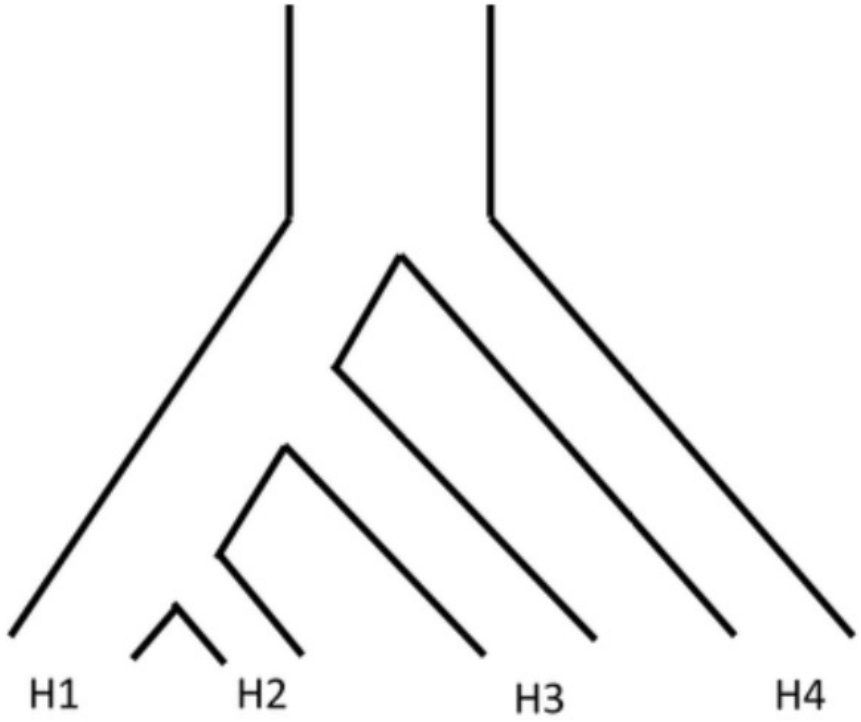
How to explain the patterns?



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?

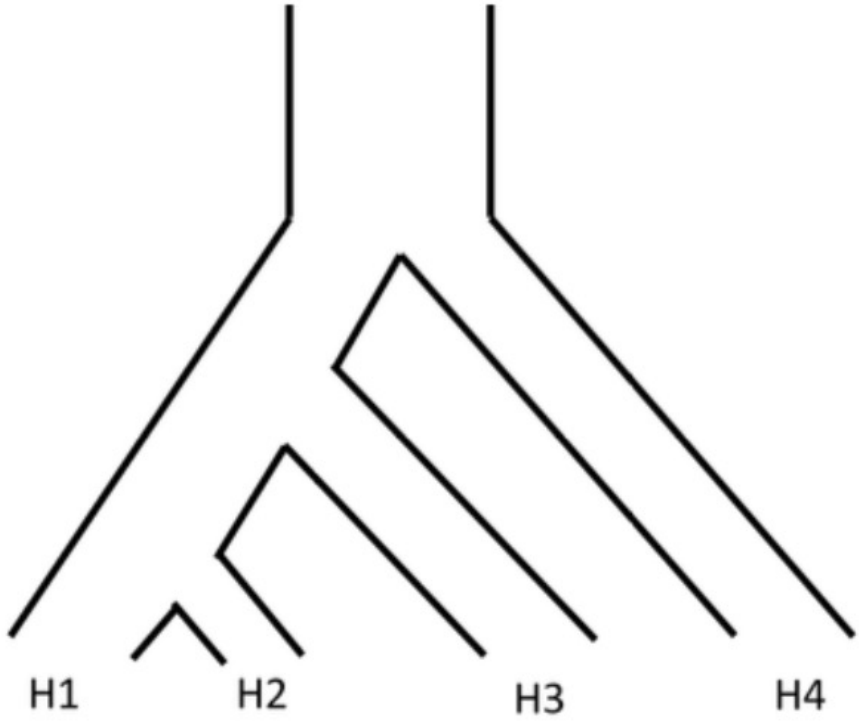


B	A	A	A
A	B	A	A
A	A	B	A
A	A	A	B
A	B	B	A
B	A	B	A
B	B	A	A

D statistic

How to explain the patterns?

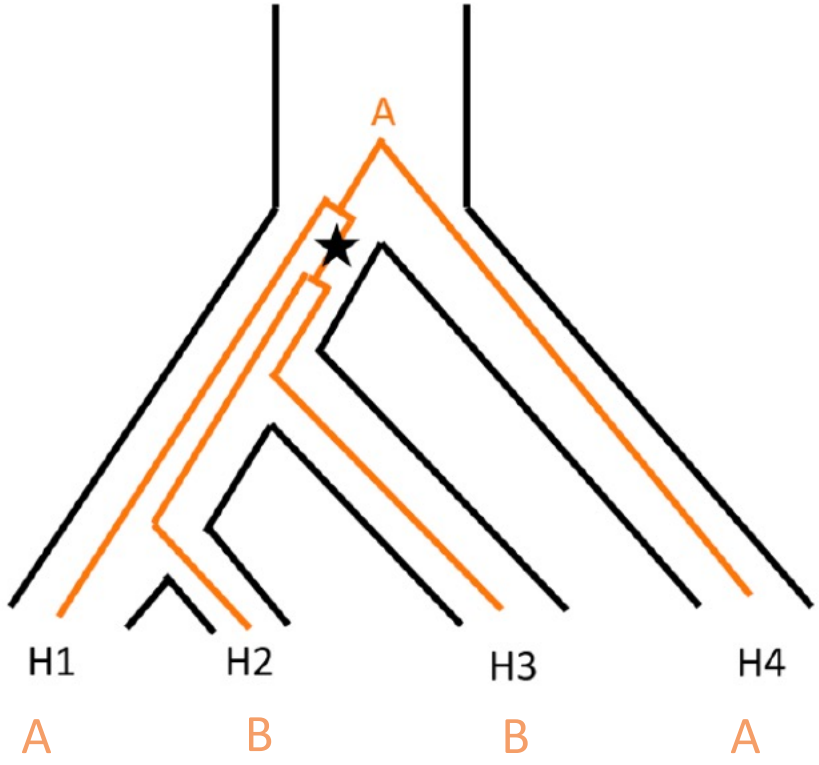
gene genealogies not necessarily follow the **population tree**

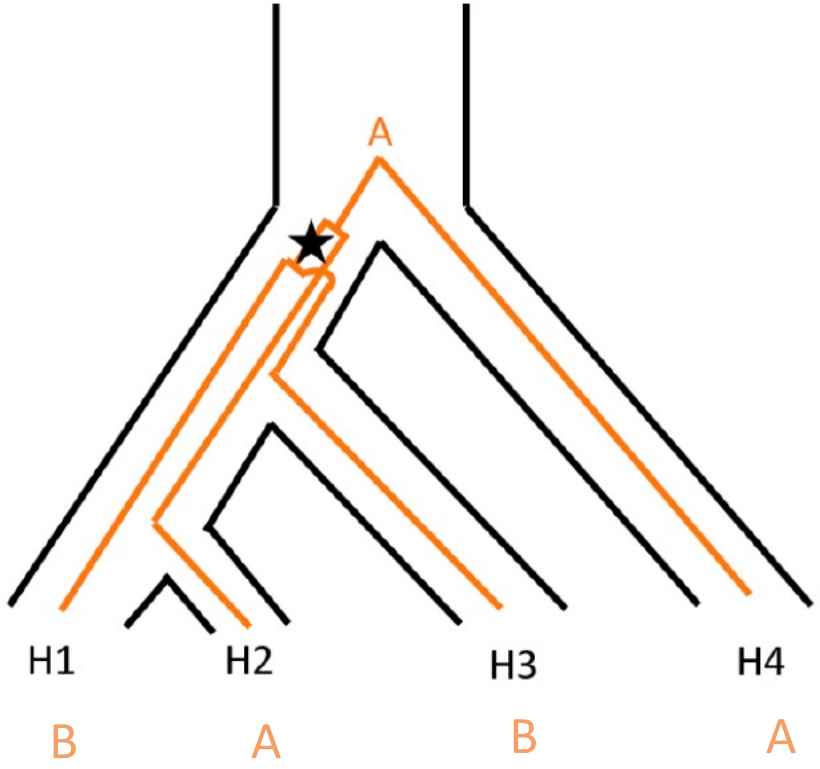


B	A	A	A
A	B	A	A
A	A	B	A
A	A	A	B
A	B	B	A
B	A	B	A
B	B	A	A

D statistic

ABBA and BABA sites

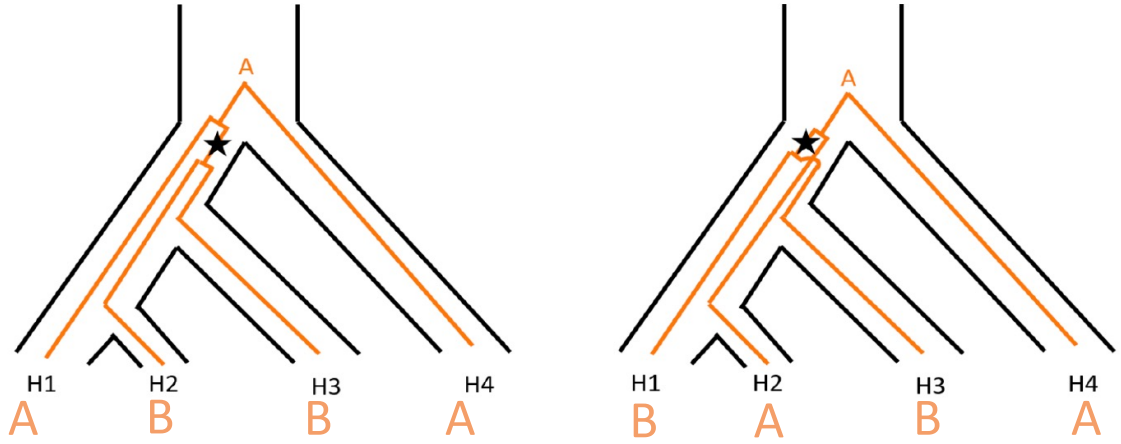




D statistic

D statistic is calculated in this way:

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

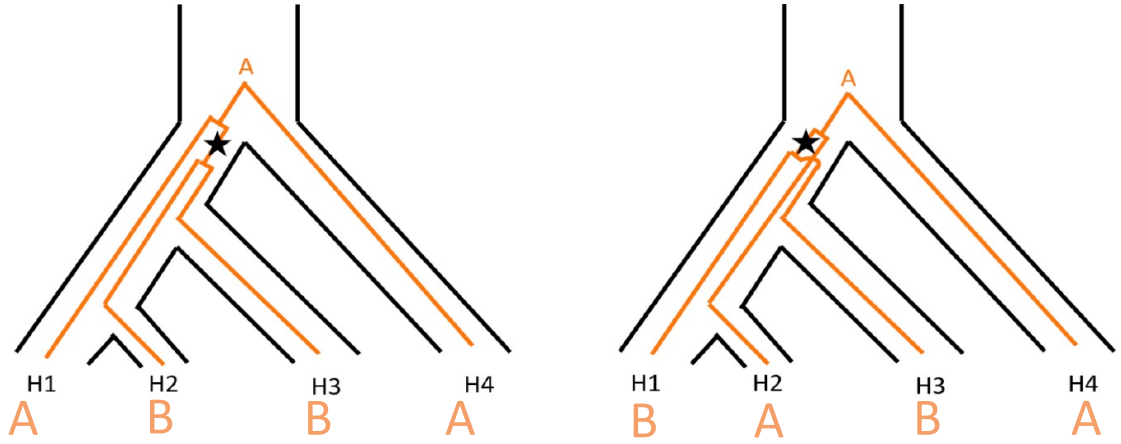


Using several (all) the loci in the genome

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

D statistic

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



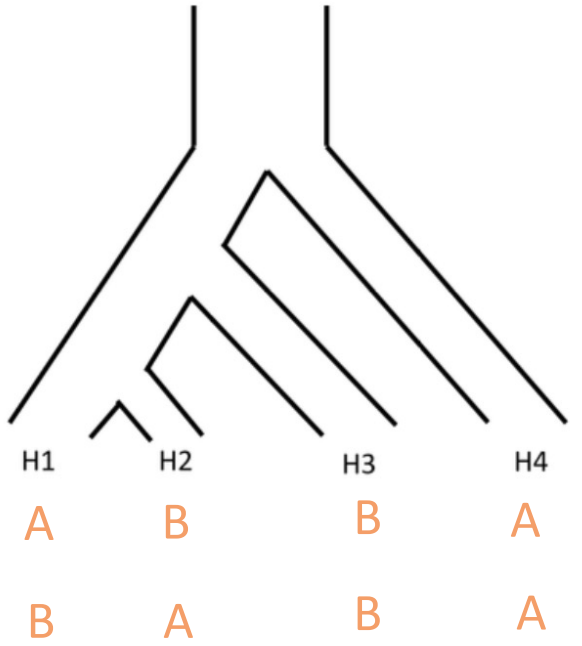
D = (1000-500)/(1000+500) = 0.33 D > 0 if ABBA is more common

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

Interpreting *D* statistic

ABBA and BABA sites should be equally represented

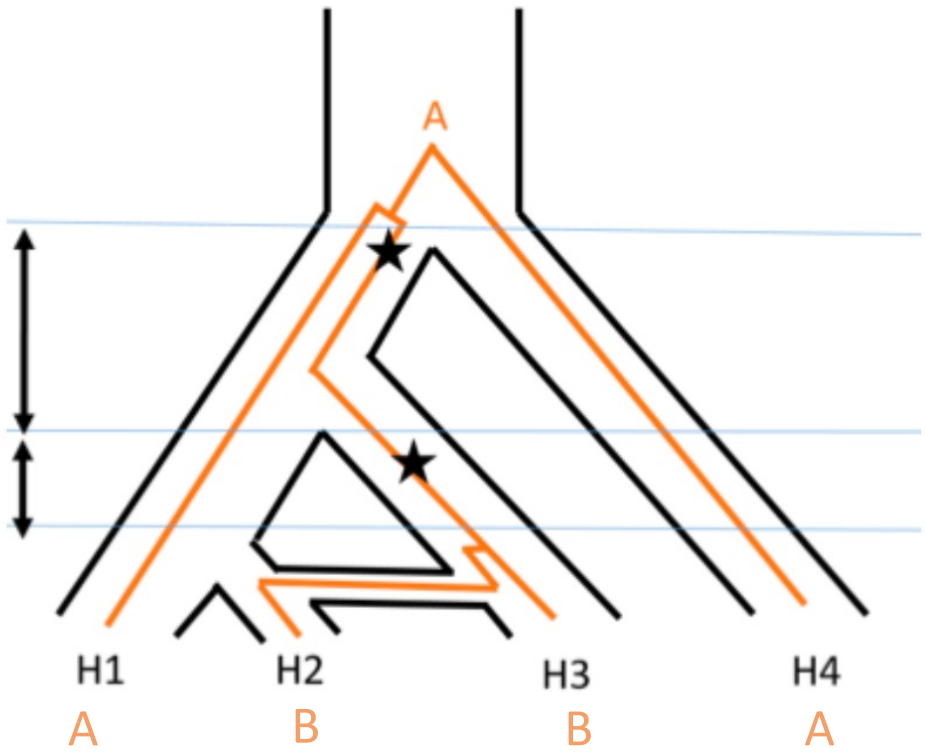
What is happening if they are not?

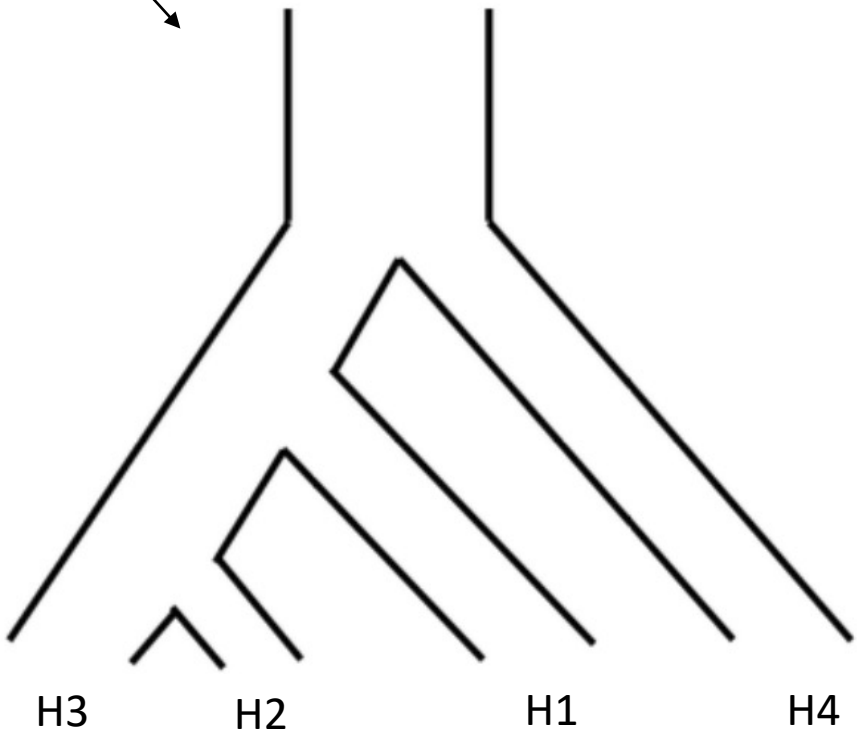


Interpreting *D* statistic

What if D ≠ 0?

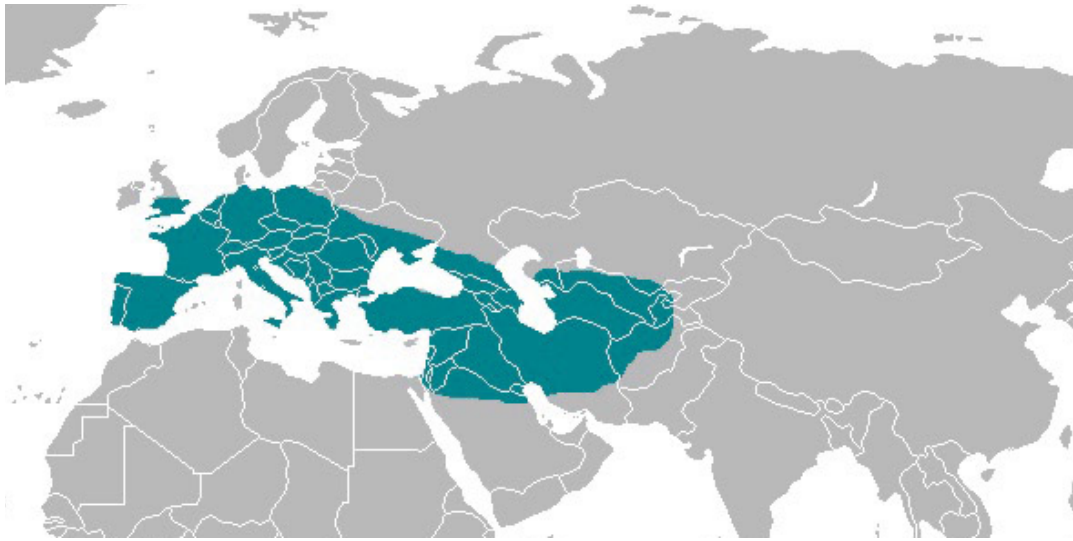
- Gene flow
- The tree is not correct

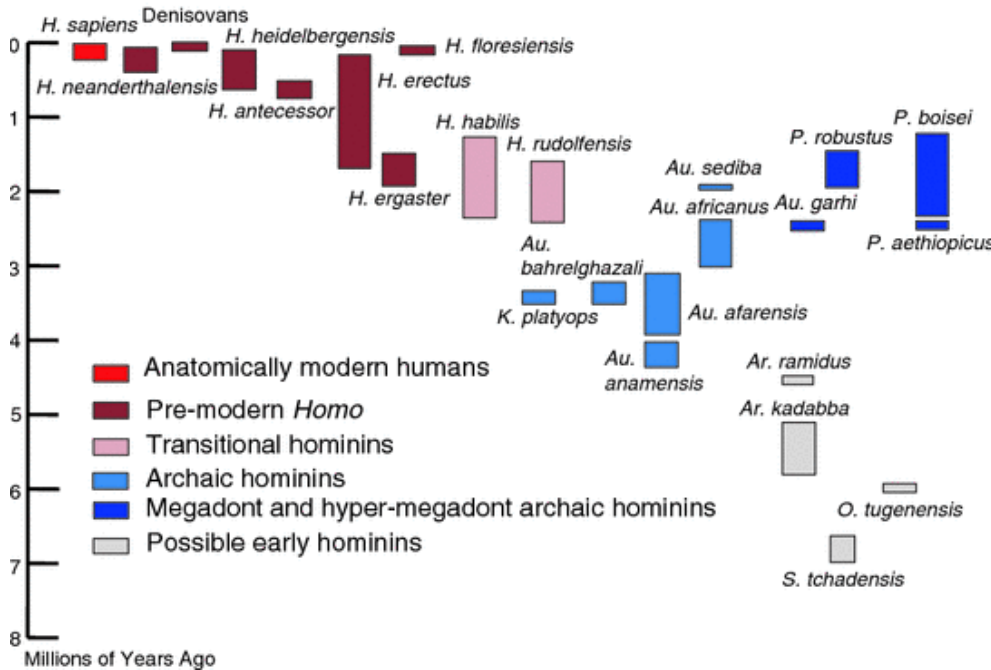




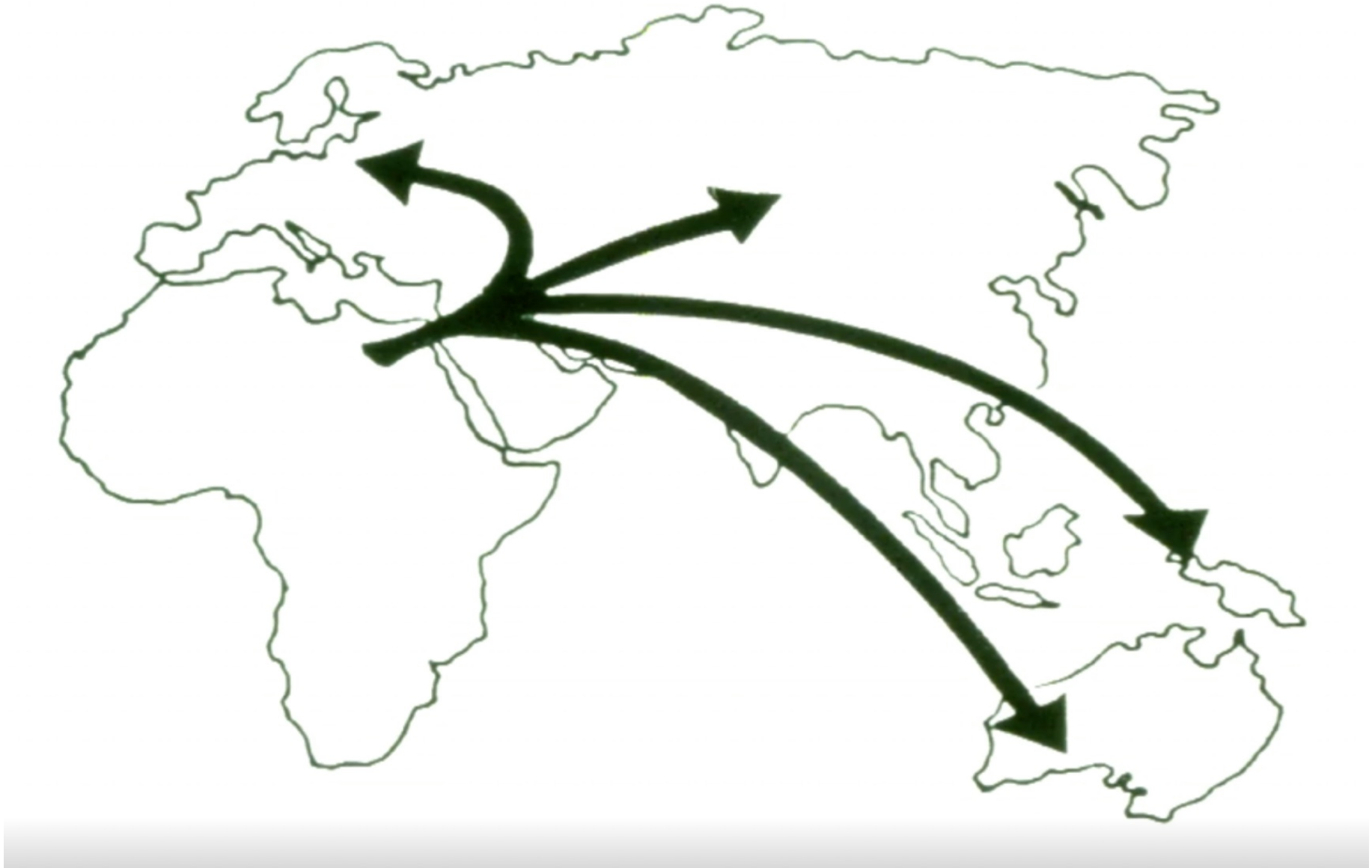
Neanderthal

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





Out of Africa



D statistic – Human/Neanderthal admixture

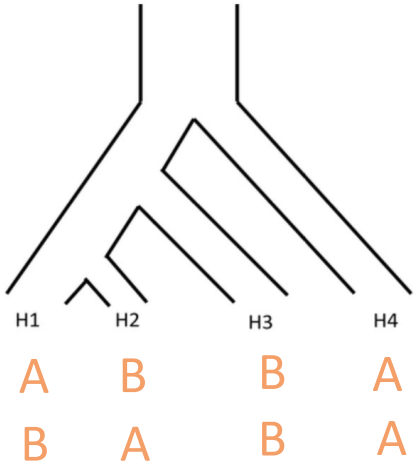
Whole genome sequences for one individual (or more) from each of the six following populations:

- 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

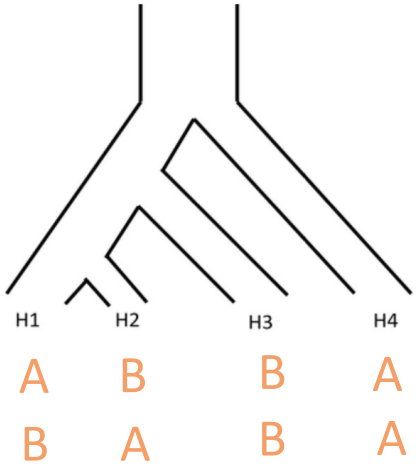
H1	H2	H3	H4	N˚ ABBA	N˚ BABA
Yoruba	Dinka	Neanderthal	Chimpanzee	44,161	44,221
Yoruba	French	Neanderthal	Chimpanzee	46,449	44,347
Yoruba	Han	Neanderthal	Chimpanzee	48,227	43,863



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

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

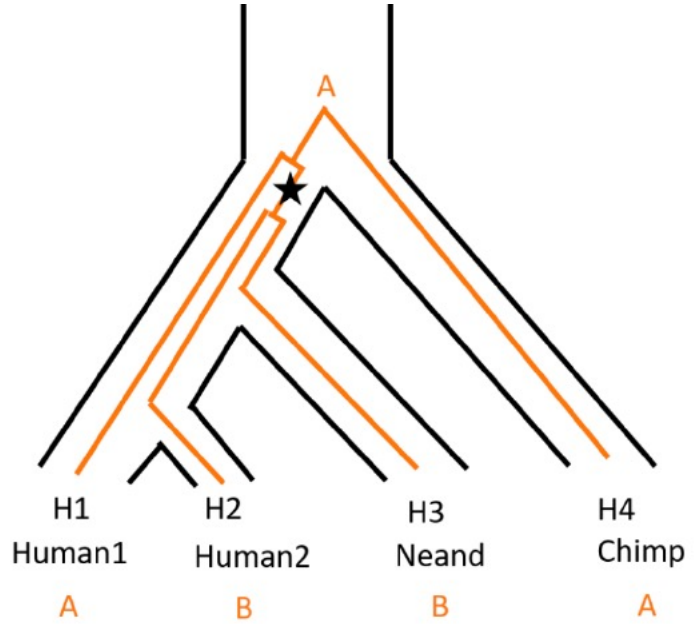
This result suggest that the pair of African genomes are symmetrically related to the Neanderthal and the chimp. Therefore, we infer that these two Africans form a clade to the exclusion of the Neanderthal and the chimp.

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

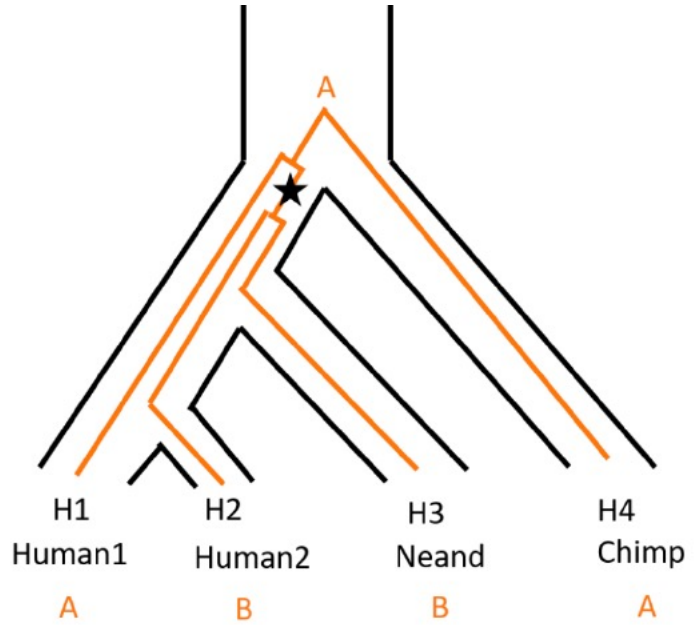
This result suggests that the **French genome shares a statistically significant larger proportion of derived alleles with the Neanderthal genome** (excess of ABBA sites), than the Yoruba does.



D statistic – Human/Neanderthal admixture

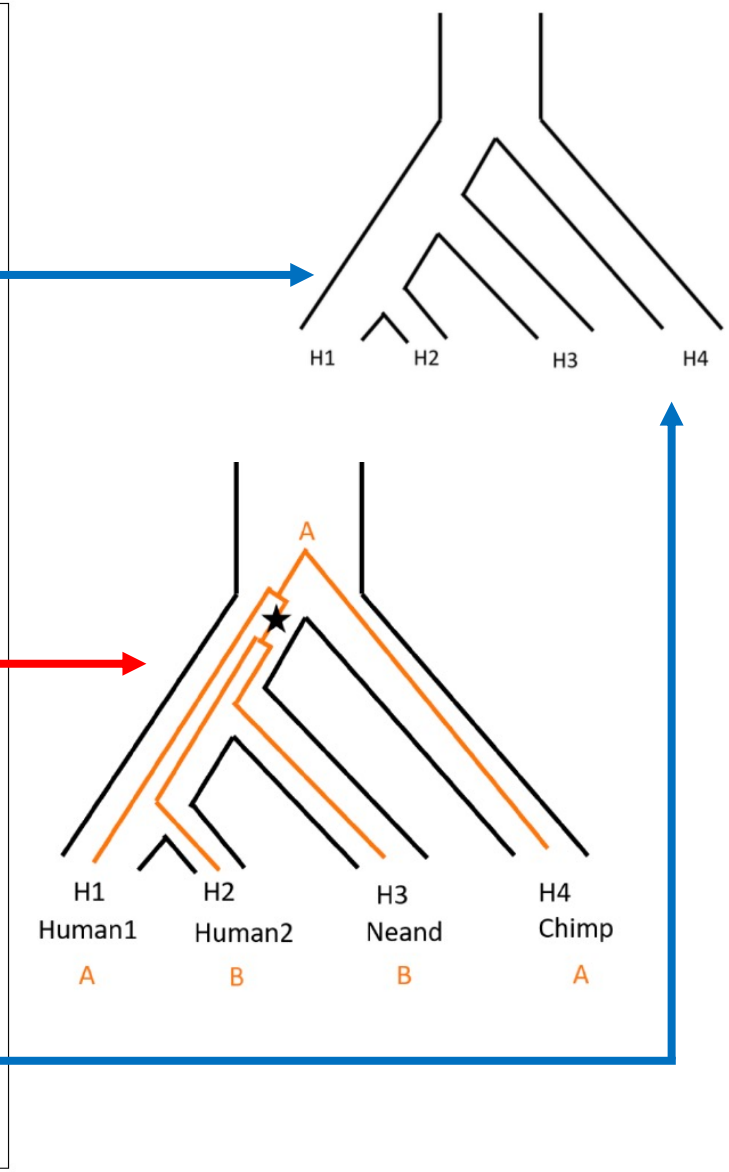
	Test	D-stat	Standard error	Z-score
Scenario 3	(Yoruba, Han; Neanderthal, Chimp)	0.04738	0.00543	8.725

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

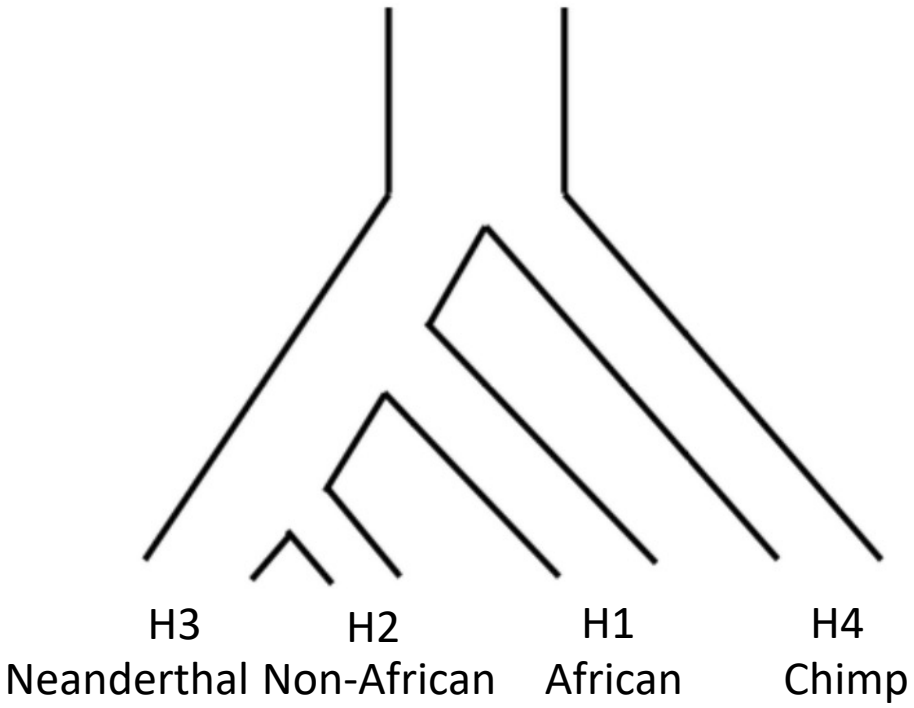
Population comparison	H_1	H_2	% Neandertal matching to H_2 – % Neandertal matching to H_1 (±1 standard error)
ABI3730 sequencing (~750 bp reads) used to discover H_1-H_2 differences			
African to African	NA18517 (Yoruba)	NA18507 (Yoruba)	-0.1 ± 0.6
	NA18517 (Yoruba)	NA19240 (Yoruba)	1.5 ± 0.7
	NA18517 (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

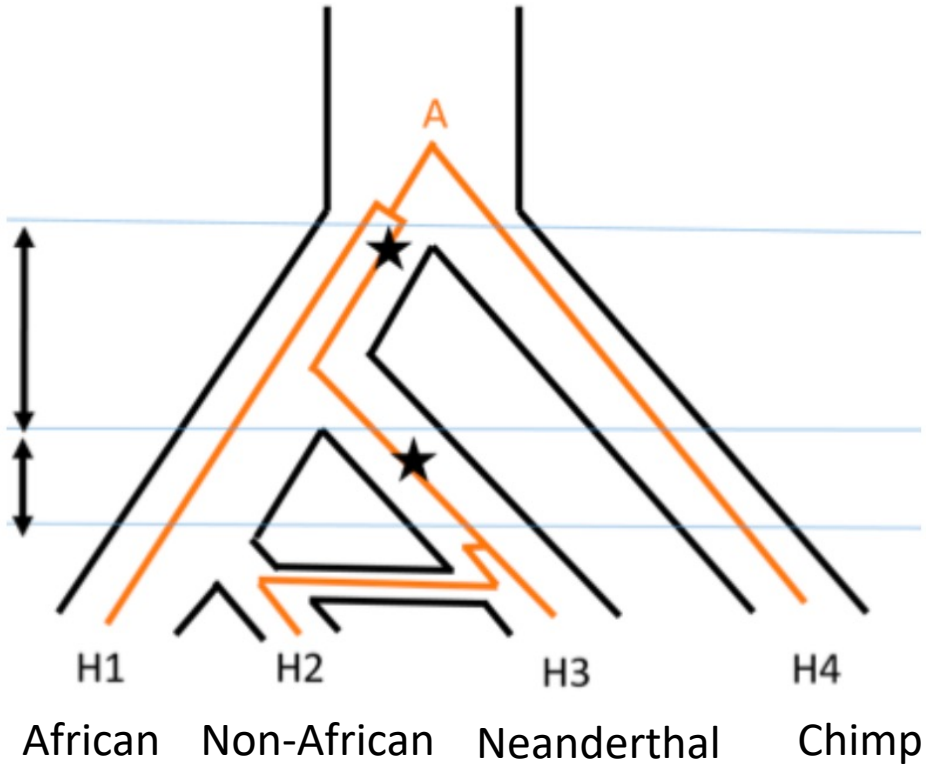


D statistic – Human/Neanderthal admixture

What if D ≠ 0?

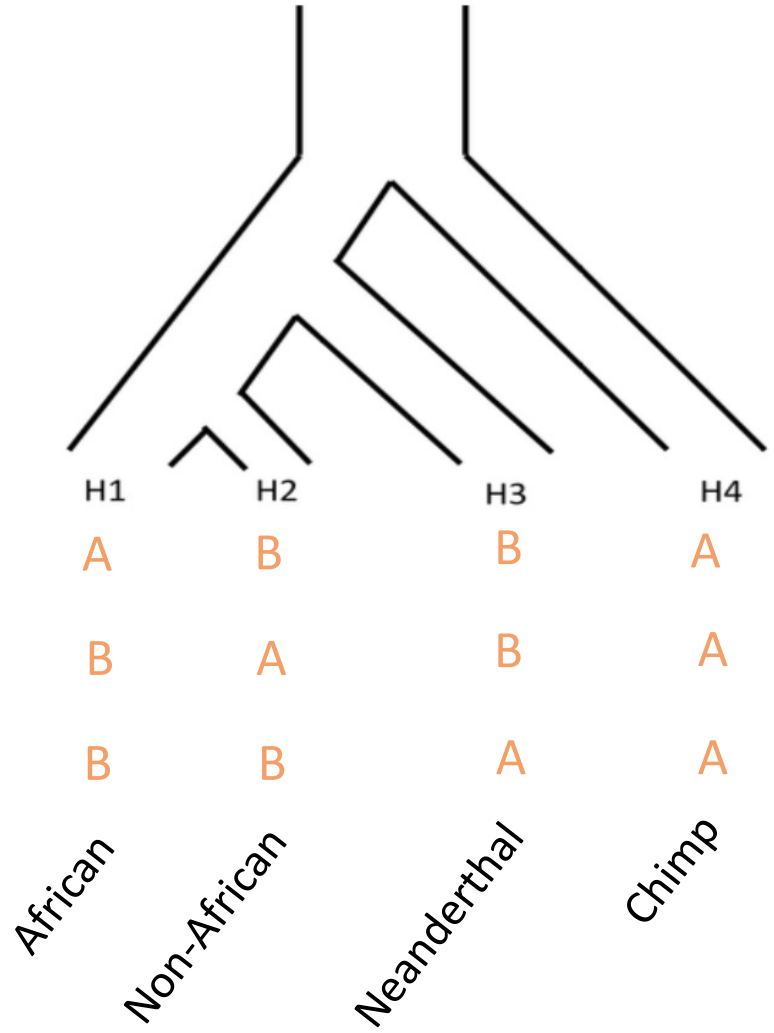
- Gene flow
- The tree is not correct





What is the right model?

D statistic – Human/Neanderthal admixture



How we can discriminate between the two model:

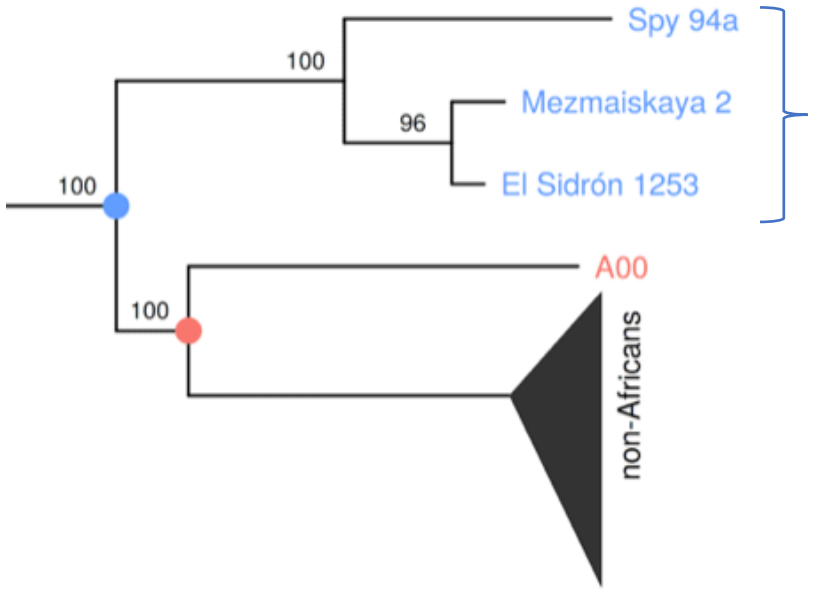
- Look for BBAA sites

D statistic – Human/Neanderthal admixture

How we can discriminate between the two model:

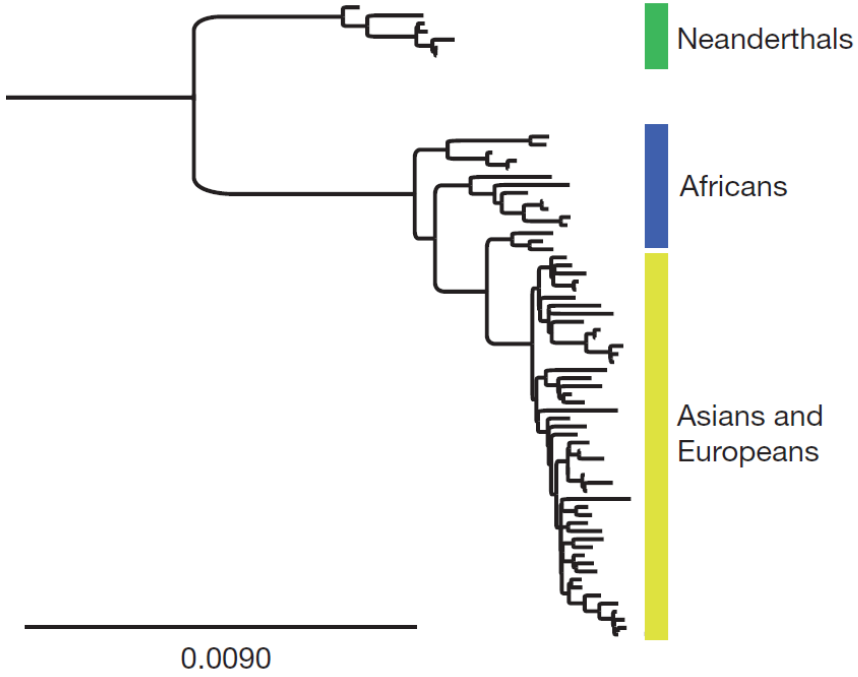
- Compare the results with different analysis

Y chromosome phylogeny



Neanderthal

mtDNA phylogeny

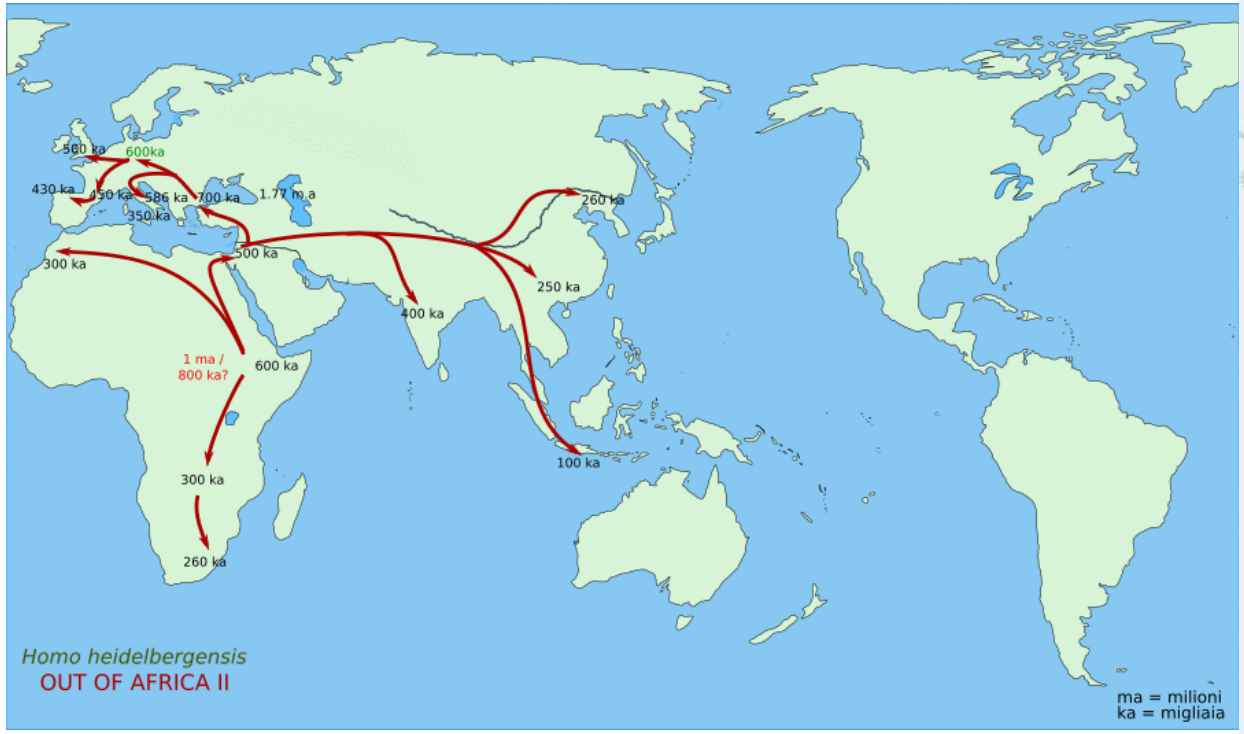


D statistic – Human/Neanderthal admixture

How we can discriminate between the two model:

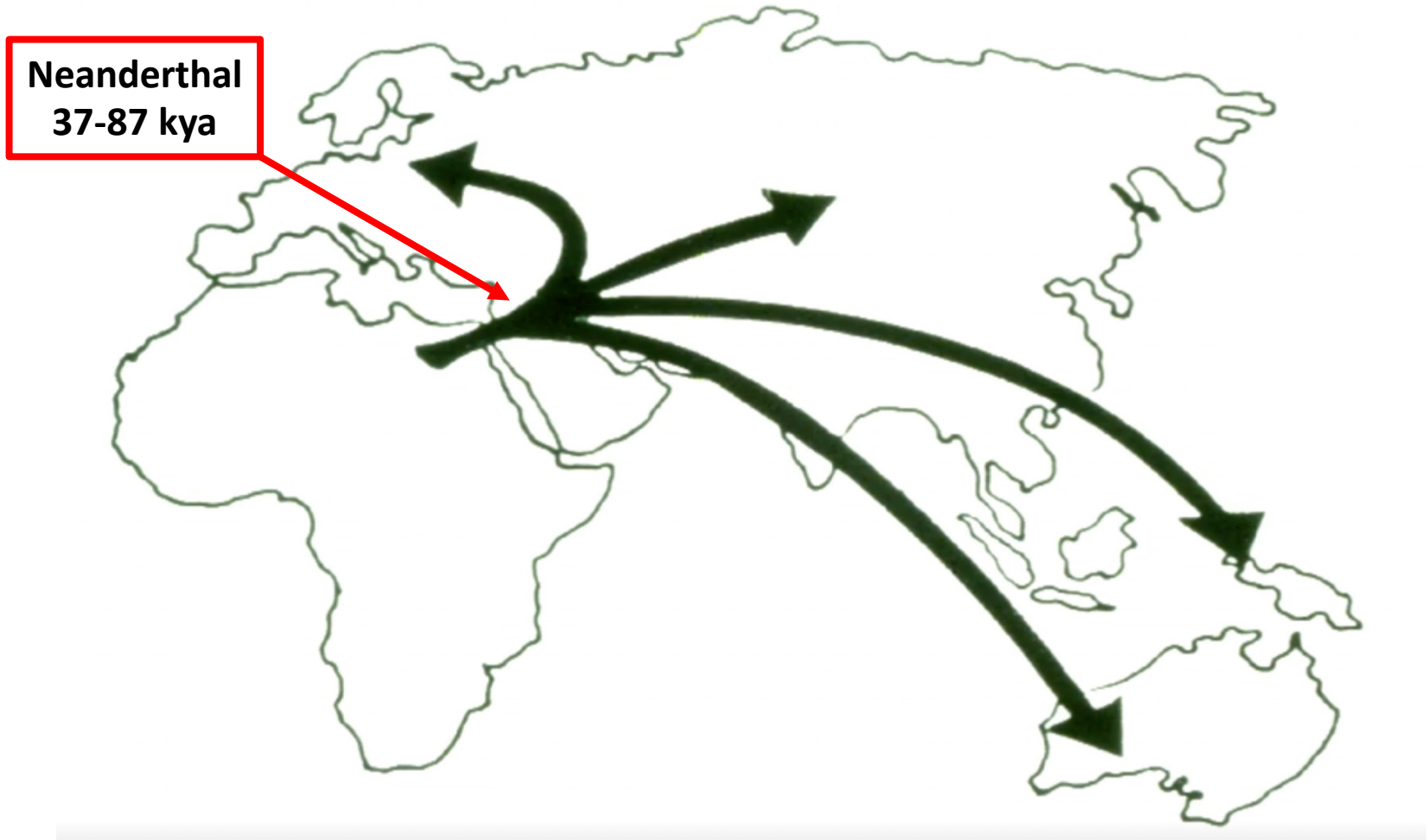
- Compare the results with different analysis

Neanderthal ancestors out of Africa ≈ 500 kya



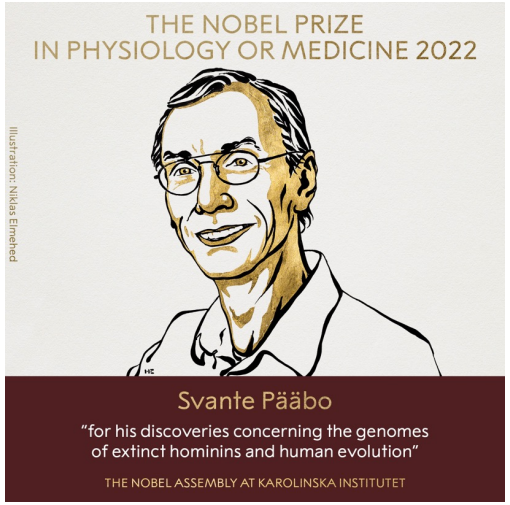
Modern humans out of Africa ≈ 100 kya

D statistic – Human/Neanderthal admixture



aDNA

- Hard to analyse (dergradation, contamination…)
- Incredibly powerful tool for evolutionary and historical reconstructions
- Insights into onset and evolution of diseases



Report

A 5,000-year-old hunter-gatherer already plagued by *Yersinia pestis*

Julian Susat [1 11], Harald Lübke [2 11], Alexander Immel [1], Ute Brinker [2], Aija Macāne [3], John Meadows [2 4], Britta Steer [5], Andreas Tholey [5], Ilga Zagorska [6], Guntis Gerhards [6], Ulrich Schmölcke [2], Mārcis Kalniņš [6], Andre Franke [1], Elīna Pētersone-Gordina [6], Barbara Teßman [7], Mari Tõrv [8], Stefan Schreiber [1 9], Christian Andree [10], Valdis Bērziņš [6], Almut Nebel [1]...Ben Krause-Kyora [1 12]

Genotype of a historic strain of *Mycobacterium tuberculosis*

Abigail S. Bouwman[a,1,2], Sandra L. Kennedy[a,2], Romy Müller[a,2], Richard H. Stephens[a], Malin Holst[b], Anwen C. Caffell[c], Charlotte A. Roberts[c], and Terence A. Brown[a,3]

Article

The major genetic risk factor for severe COVID-19 is inherited from Neanderthals

https://doi.org/10.1038/s41586-020-2818-3
Hugo Zeberg[1,2] & Svante Pääbo[1,3]
Received: 3 July 2020

nature communications

Article

https://doi.org/10.1038/s41467-024-45438-1

Cases of trisomy 21 and trisomy 18 among historic and prehistoric individuals discovered from ancient DNA