

Ancient DNA

DTU Next Generation Sequencing Analysis



Martin Sikora, PhD

Lundbeck Foundation Geogenetics Centre, University of Copenhagen

martin.sikora@bio.ku.dk

2 weeks ago

The New York Times

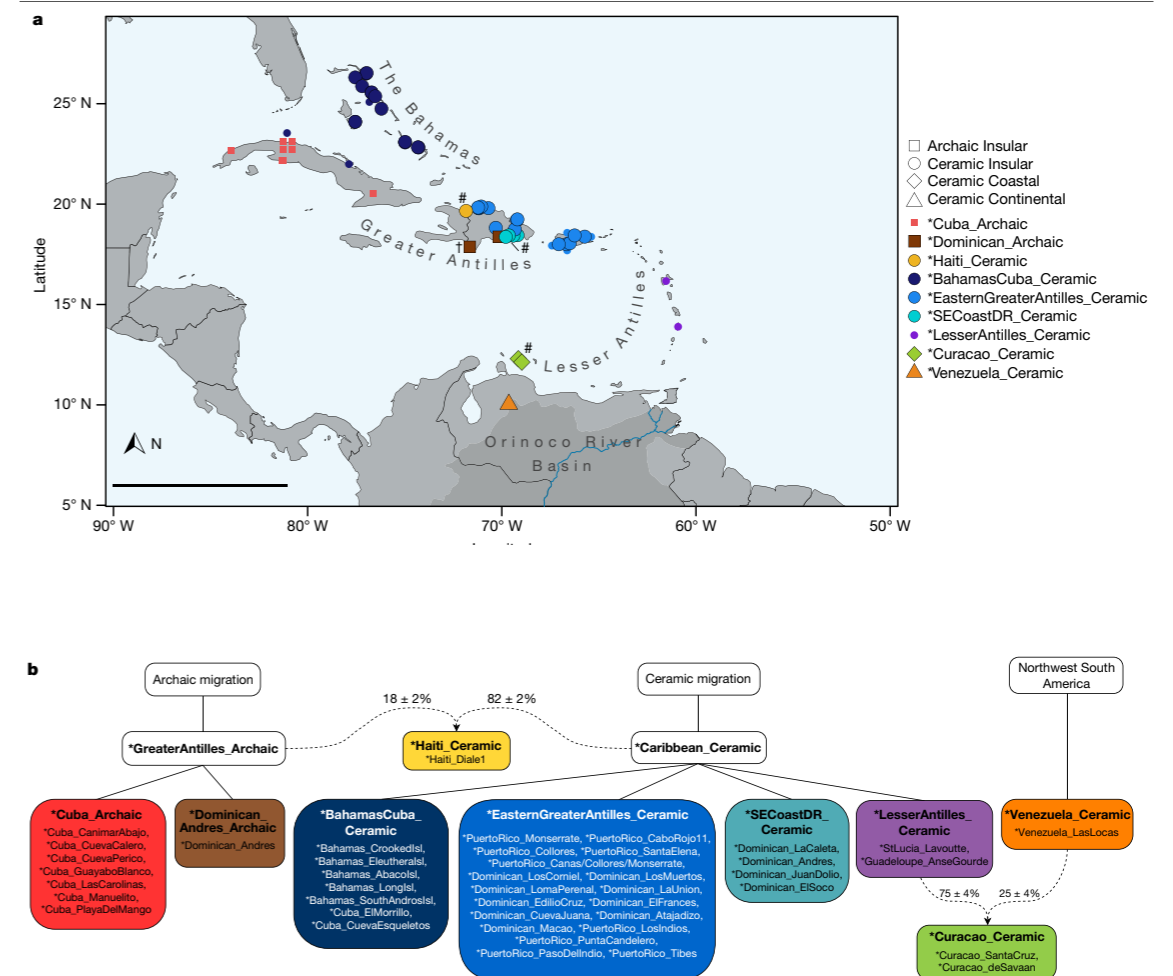
MATTER

Ancient DNA Shows Humans Settled Caribbean in 2 Distinct Waves

Millions of people living on the islands today inherited genes from the people who made them home before Europeans arrived.



Taíno ceramic vessels from eastern Dominican Republic, circa A.D. 1400. Menno Hoogland/Leiden University



1984 - the first 2 ancient DNA sequences

NATURE VOL. 312 15 NOVEMBER 1984

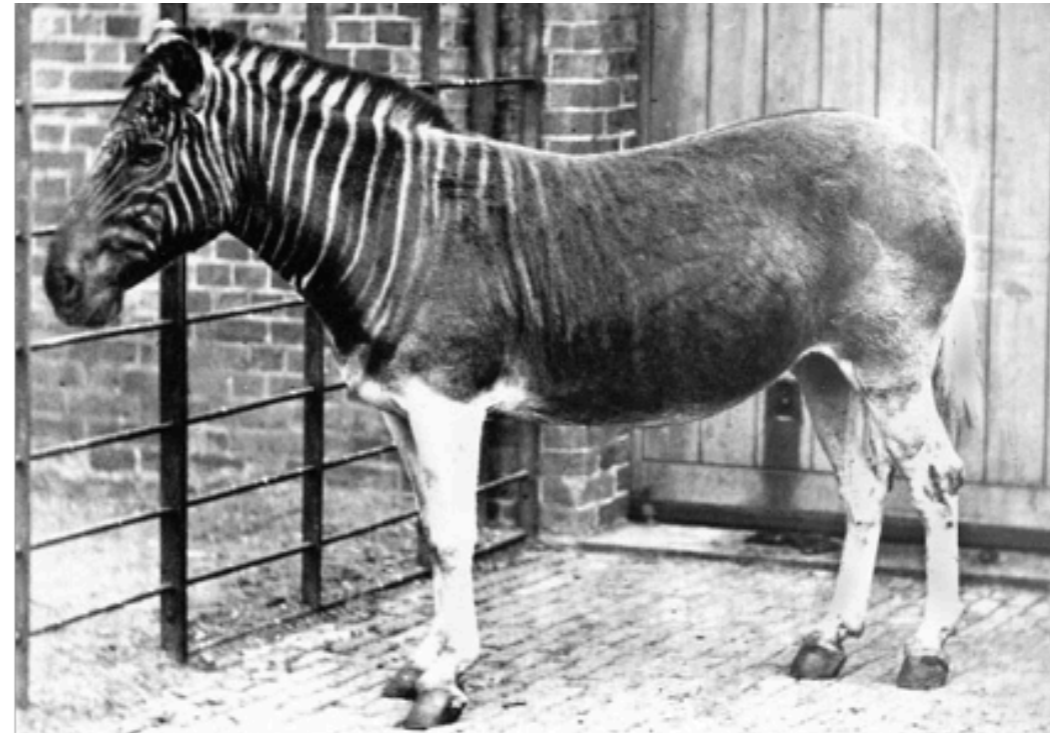
DNA sequences from the quagga, an extinct member of the horse family

Russell Higuchi*, Barbara Bowman*, Mary Freiberger*,
Oliver A. Ryder† & Allan C. Wilson*

* Department of Biochemistry, University of California, Berkeley,
California 94720, USA

† Research Department, San Diego Zoo, San Diego,
California 92103, USA

To determine whether DNA survives and can be recovered from the remains of extinct creatures, we have examined dried muscle from a museum specimen of the quagga, a zebra-like species (*Equus quagga*) that became extinct in 1883 (ref. 1). We report that DNA



Unidentified reading frame 1

Quagga	C CCA ATC CTG CTC GCC GTA GCA TTC CTC ACA CTA GTT GAA CGA AAA GTC TTA GGC TAC ATA CAA CTT CGT AAA GGA CCC AAC ATC GTA GGC CCC TAT GGC CTA CTA CAA CCC ATT AC
Zebra T G T C G*

Cytochrome oxidase I

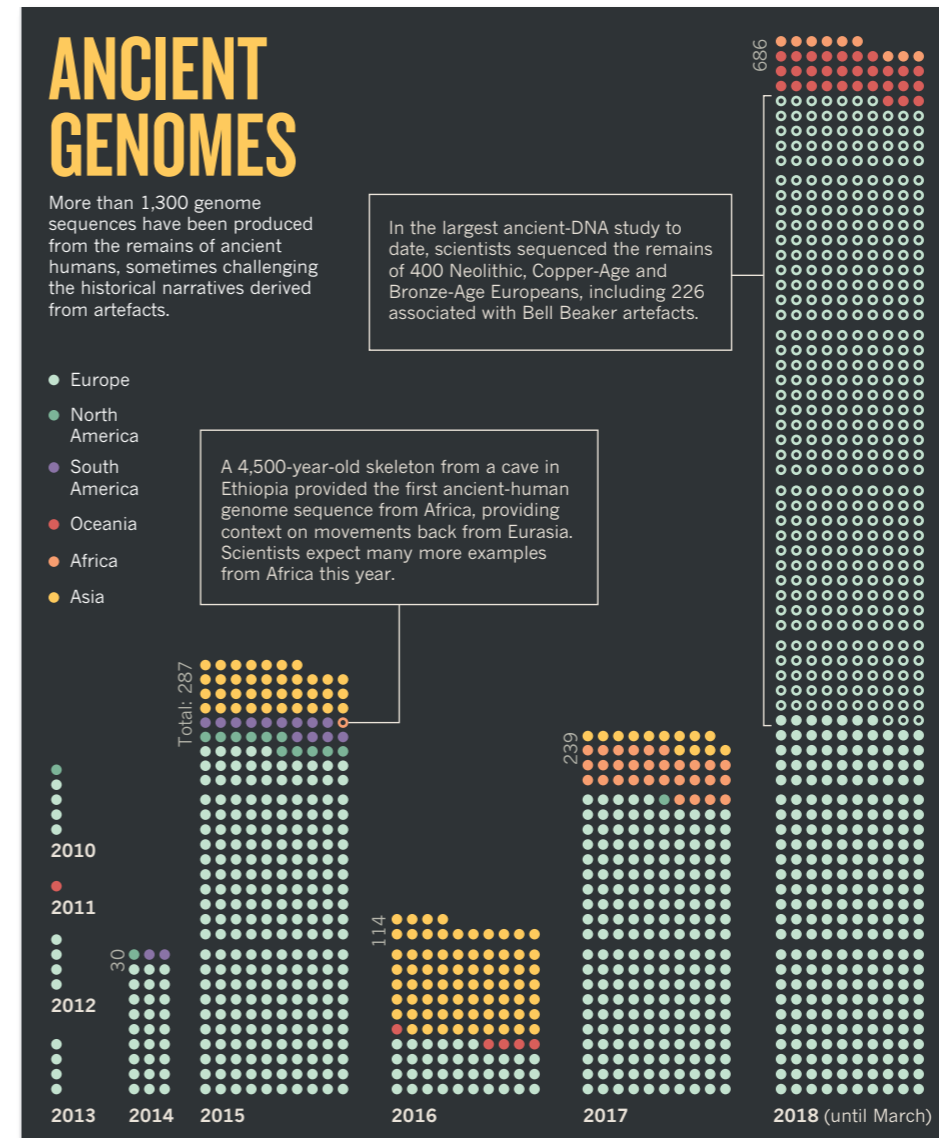
Quagga	A GGA GGA TTC GTT CAC TGA TTC CCT CTA TTC TCA GGA TAC ACA CTC AAC CAA ACC TGA GCA AAA ATT CAG TTT ACA ATT ATA TTC GTA GGG GTC AAC ATA ATT TTC TTC CCA
Zebra	G T G C A T C*

Fig. 1 Sequences of the coding strands determined for two pieces of quagga mtDNA. The sequences are arranged in triplets corresponding to the amino acids that they encode. At 12 positions, the quagga sequences differ from those of mtDNA from a mountain zebra; only for these positions is the nature of the base specified for the zebra. The two asterisks identify triplets at which the zebra and quagga differ by an amino acid replacement.

2018 - Thousands of ancient genomes



2010



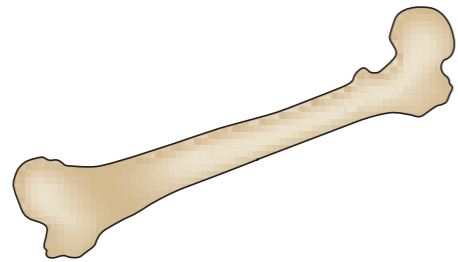
2018

Generating and authenticating aDNA data

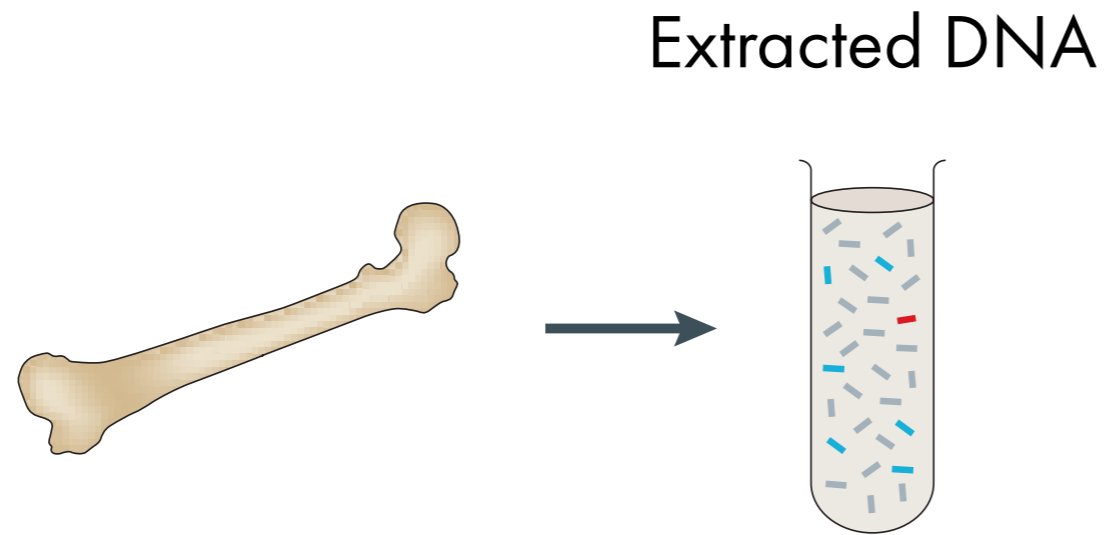


Ancient DNA sequencing

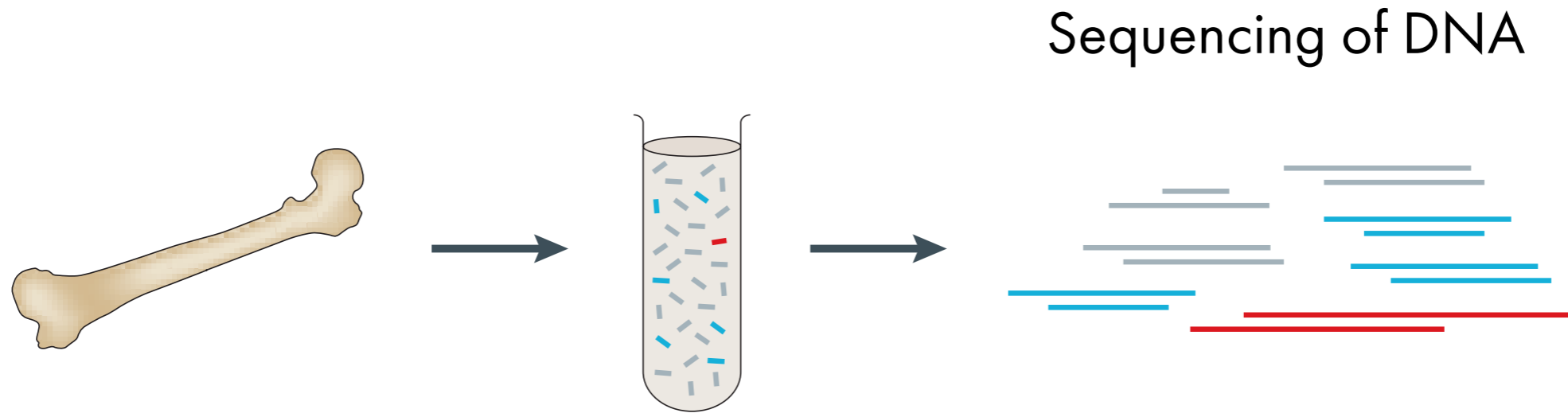
Bone, tooth, tissue



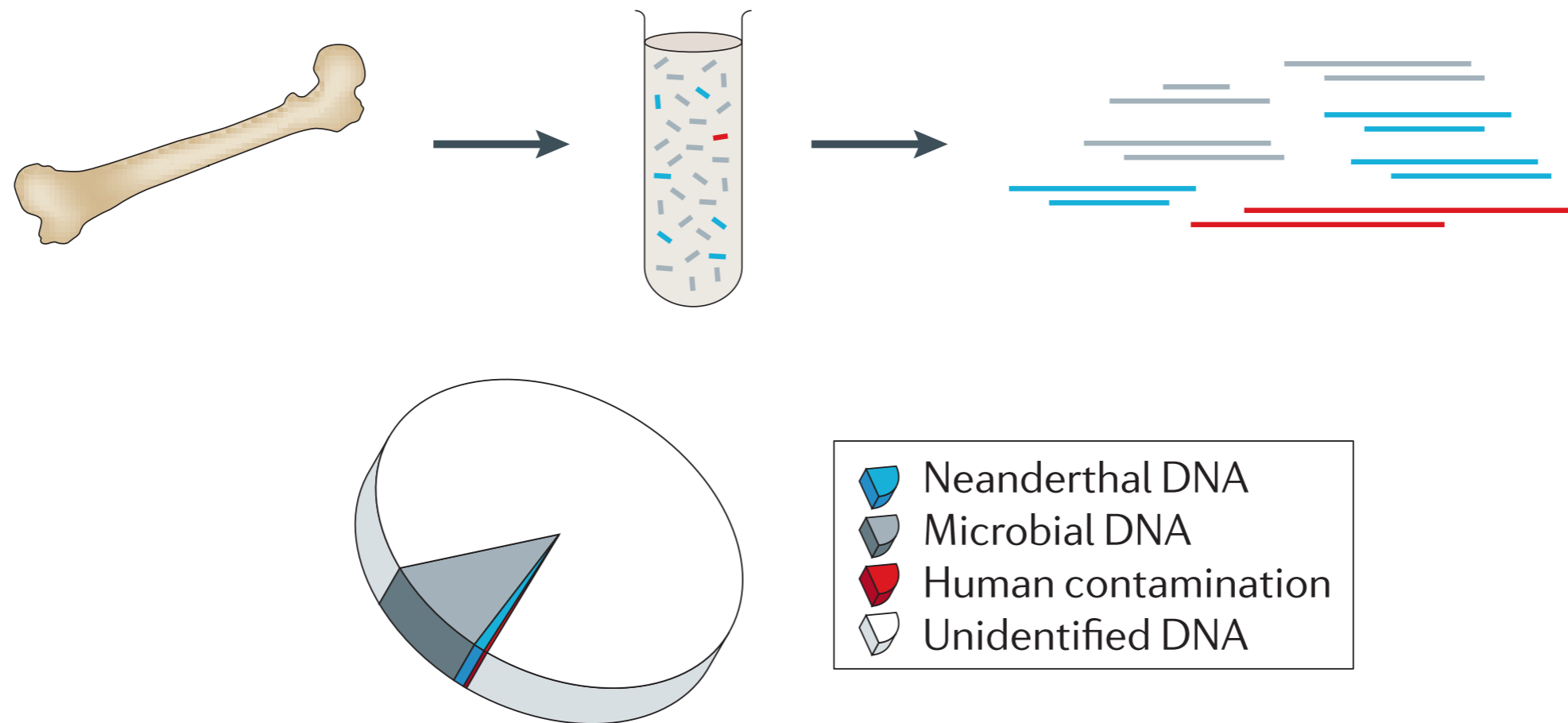
Ancient DNA sequencing



Ancient DNA sequencing



Ancient DNA sequencing

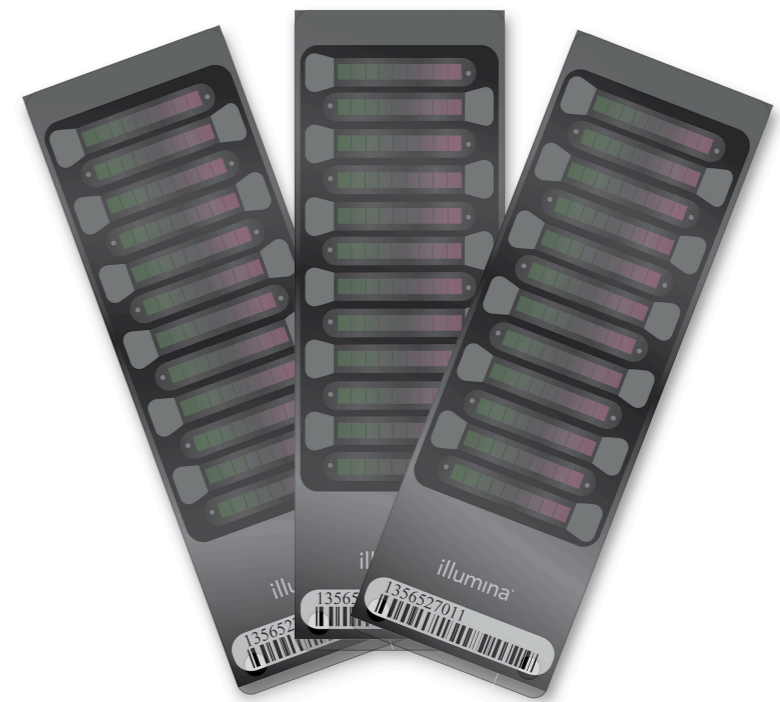


Ancient DNA studies are metagenomic studies

Sequencing approaches



Shotgun sequencing

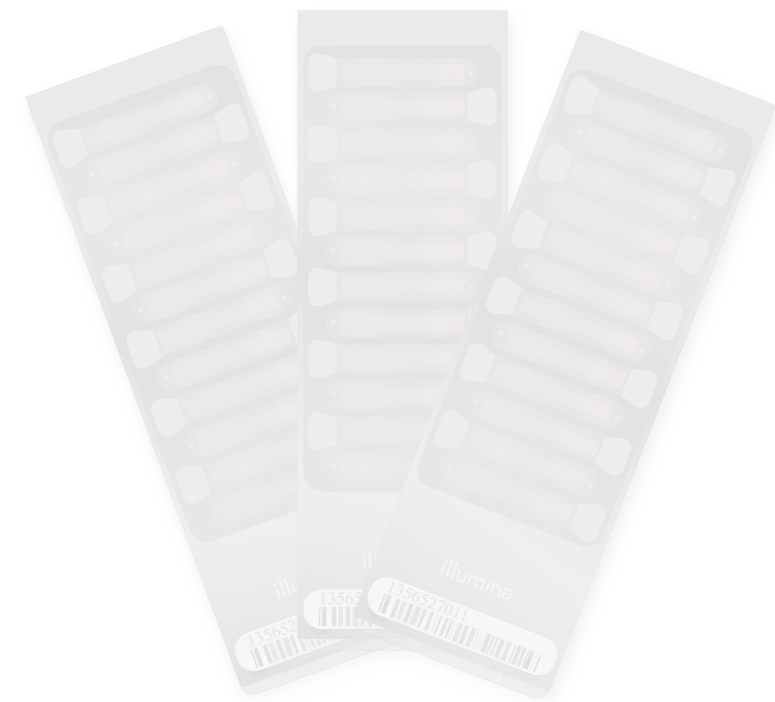


DNA capture

Sequencing approaches



Shotgun sequencing

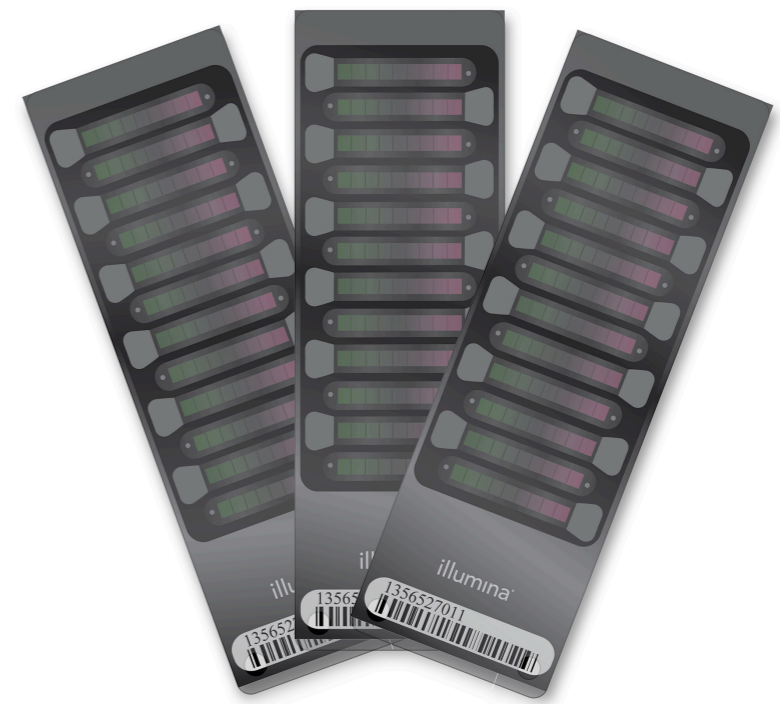


DNA capture

Sequencing approaches

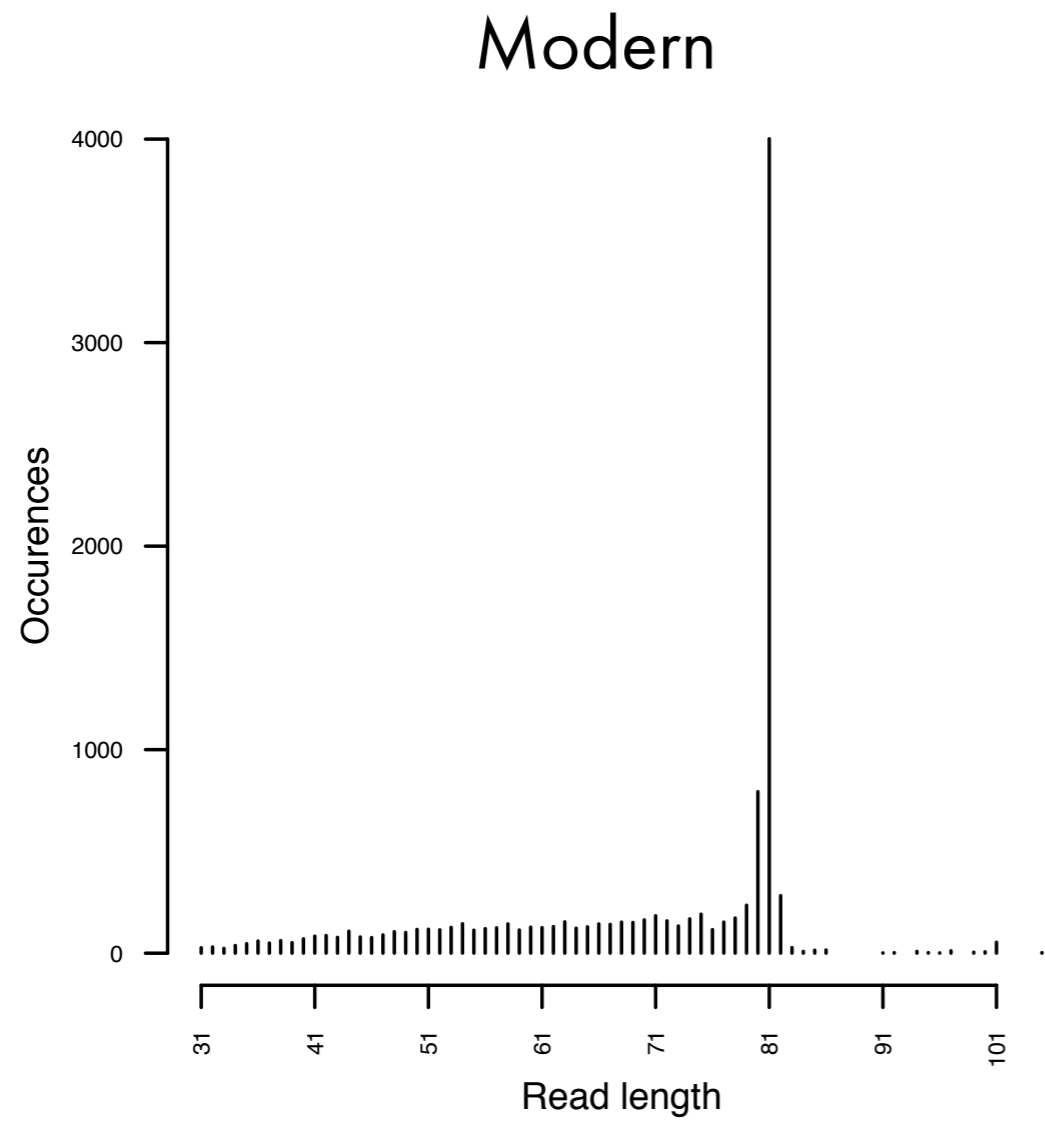


Shotgun sequencing



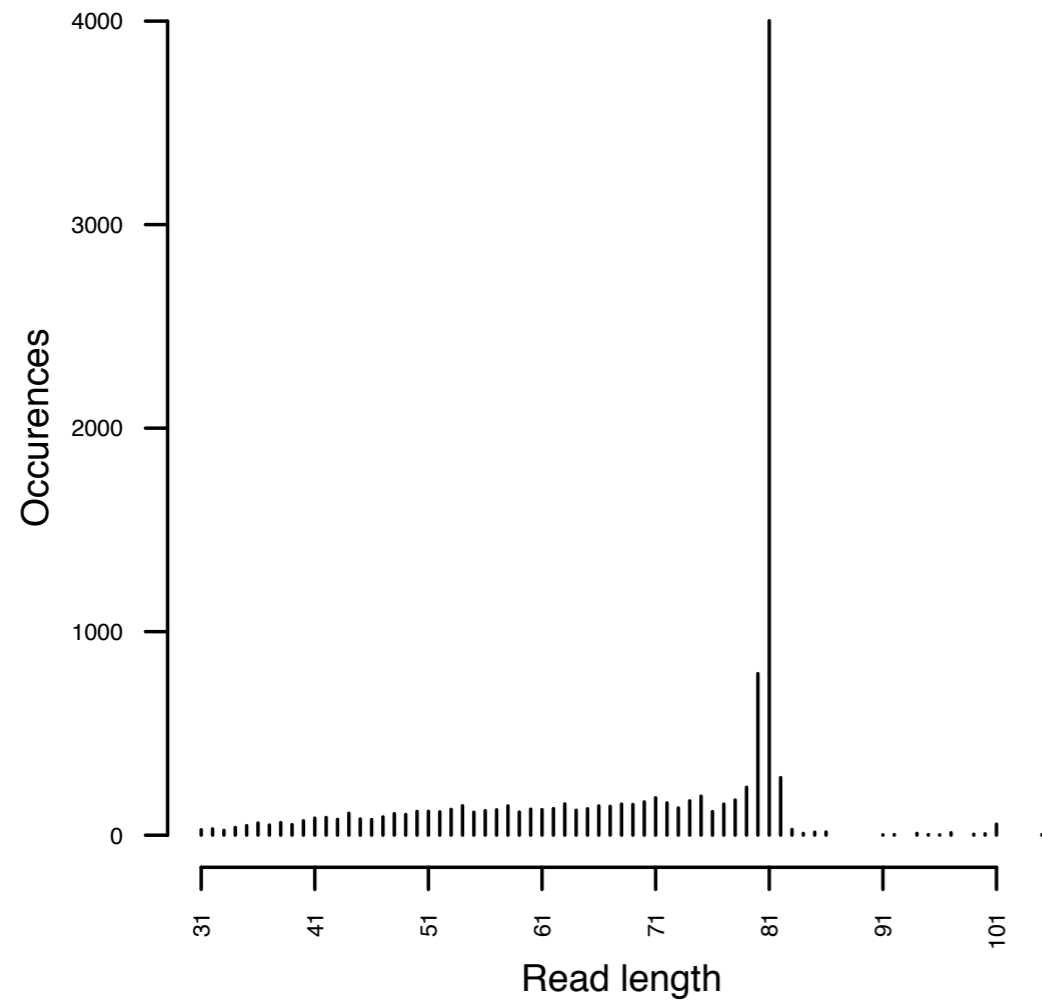
DNA capture

Characteristics of ancient DNA

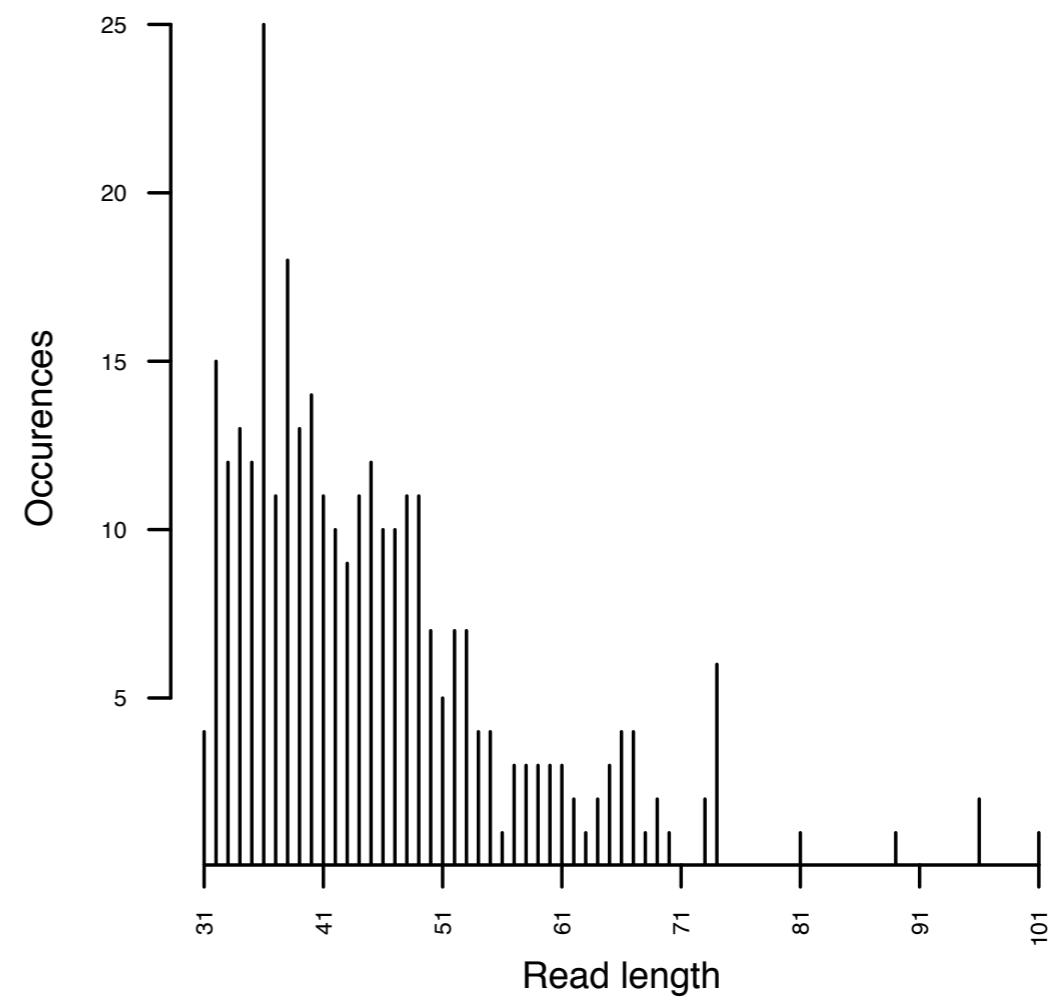


Characteristics of ancient DNA

Modern

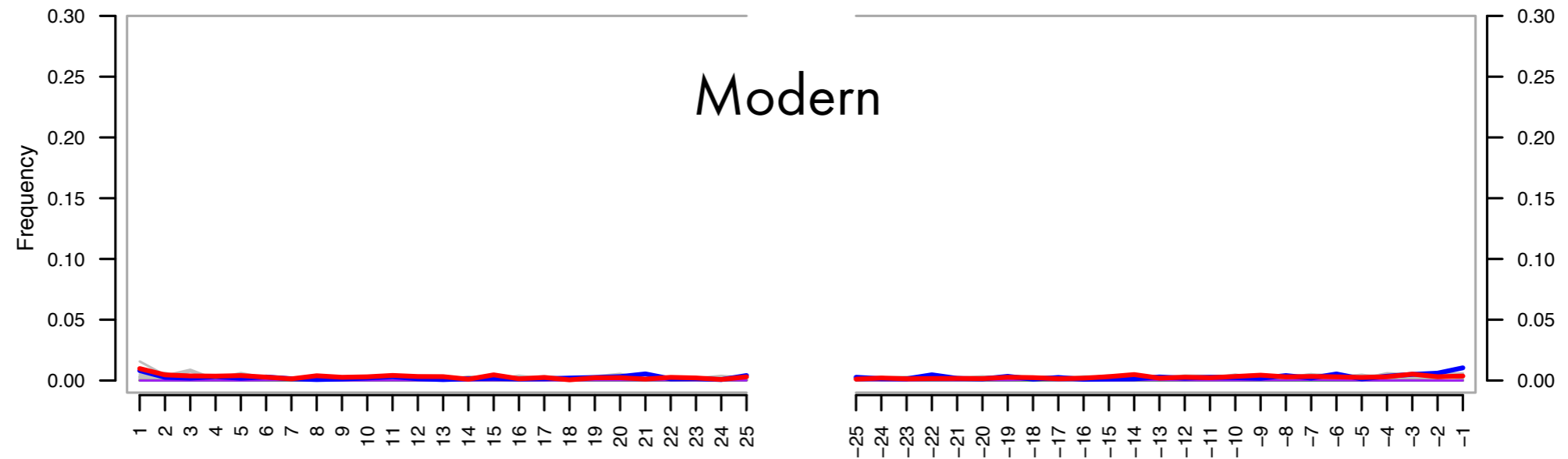


Ancient

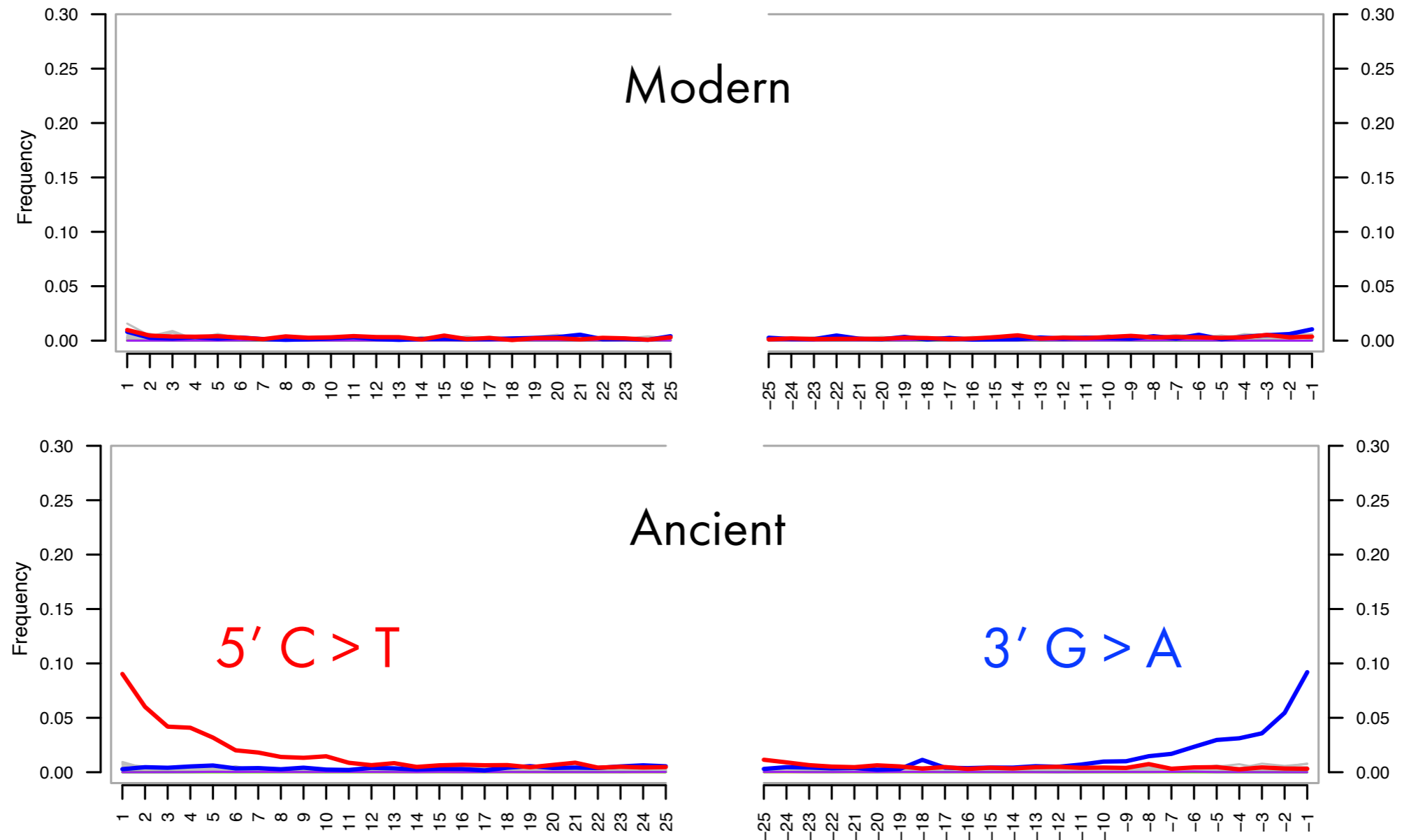


Highly fragmented - short molecules (< 100bp)

Characteristics of ancient DNA



Characteristics of ancient DNA



Post-mortem DNA damage

Increased rates of C>T and G>A substitution towards read ends

Challenges in ancient genomics

Accessibility



Suitable sample material

DNA preservation

Challenges in ancient genomics

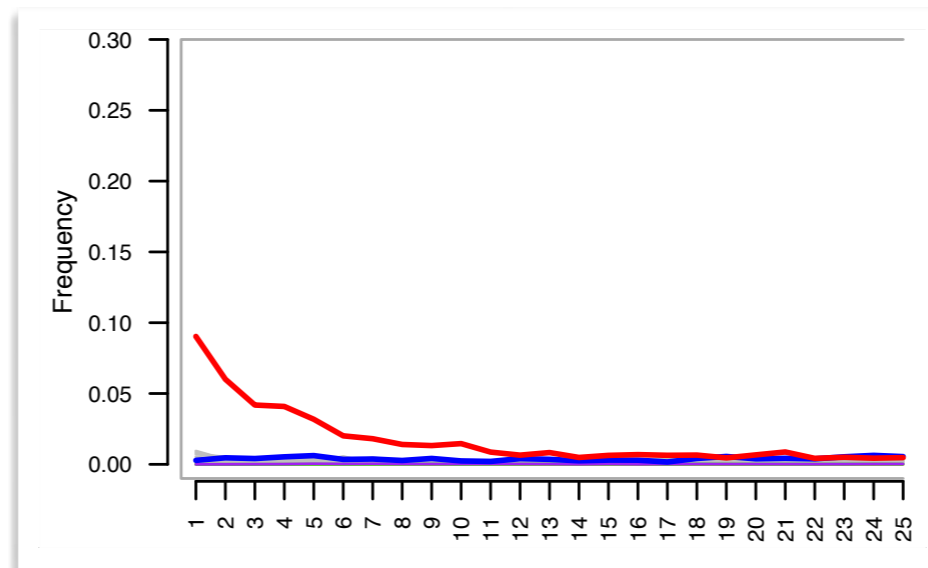
Accessibility



Suitable sample material

DNA preservation

Authentication



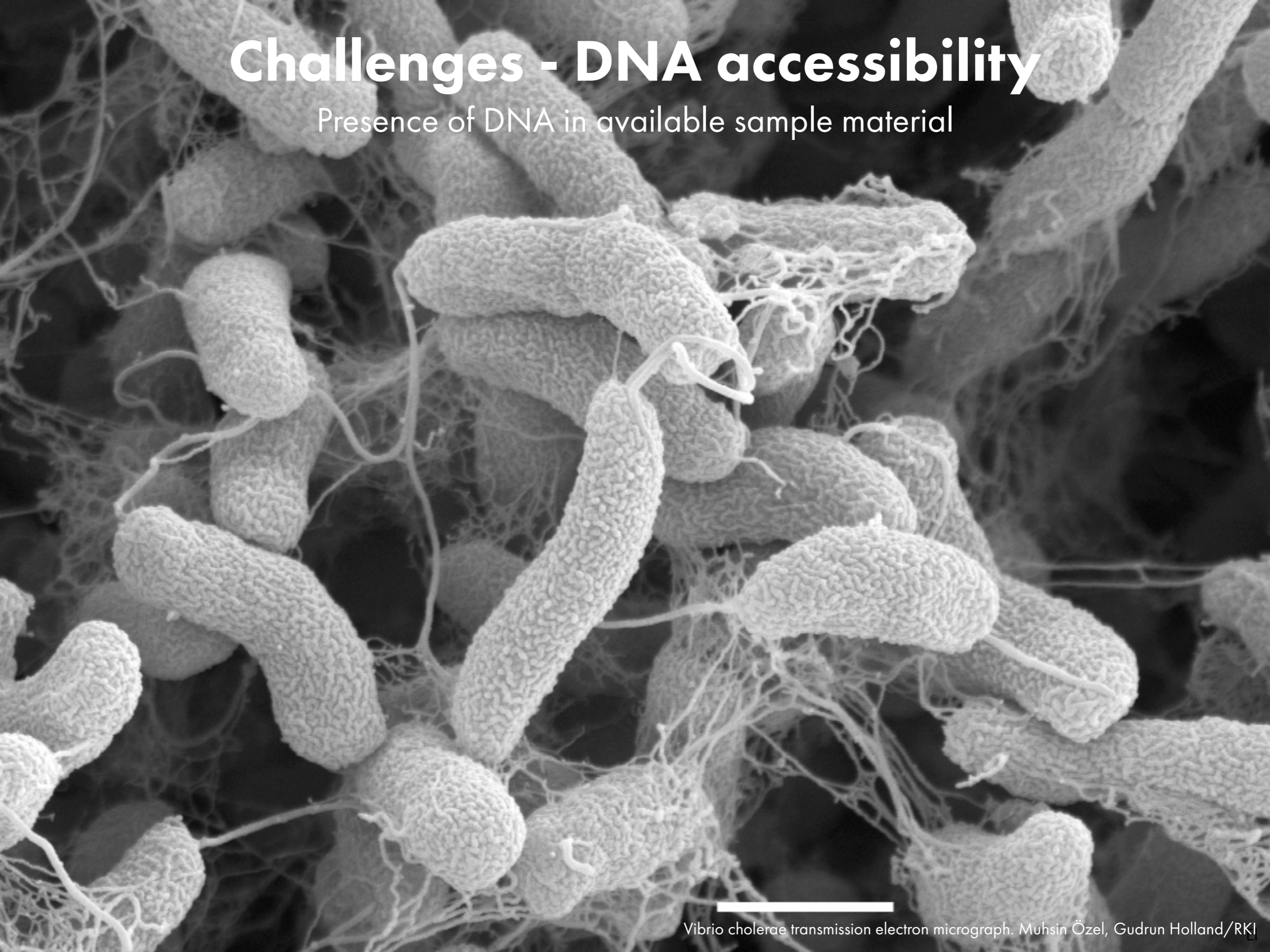
Lab contamination

Environmental contamination

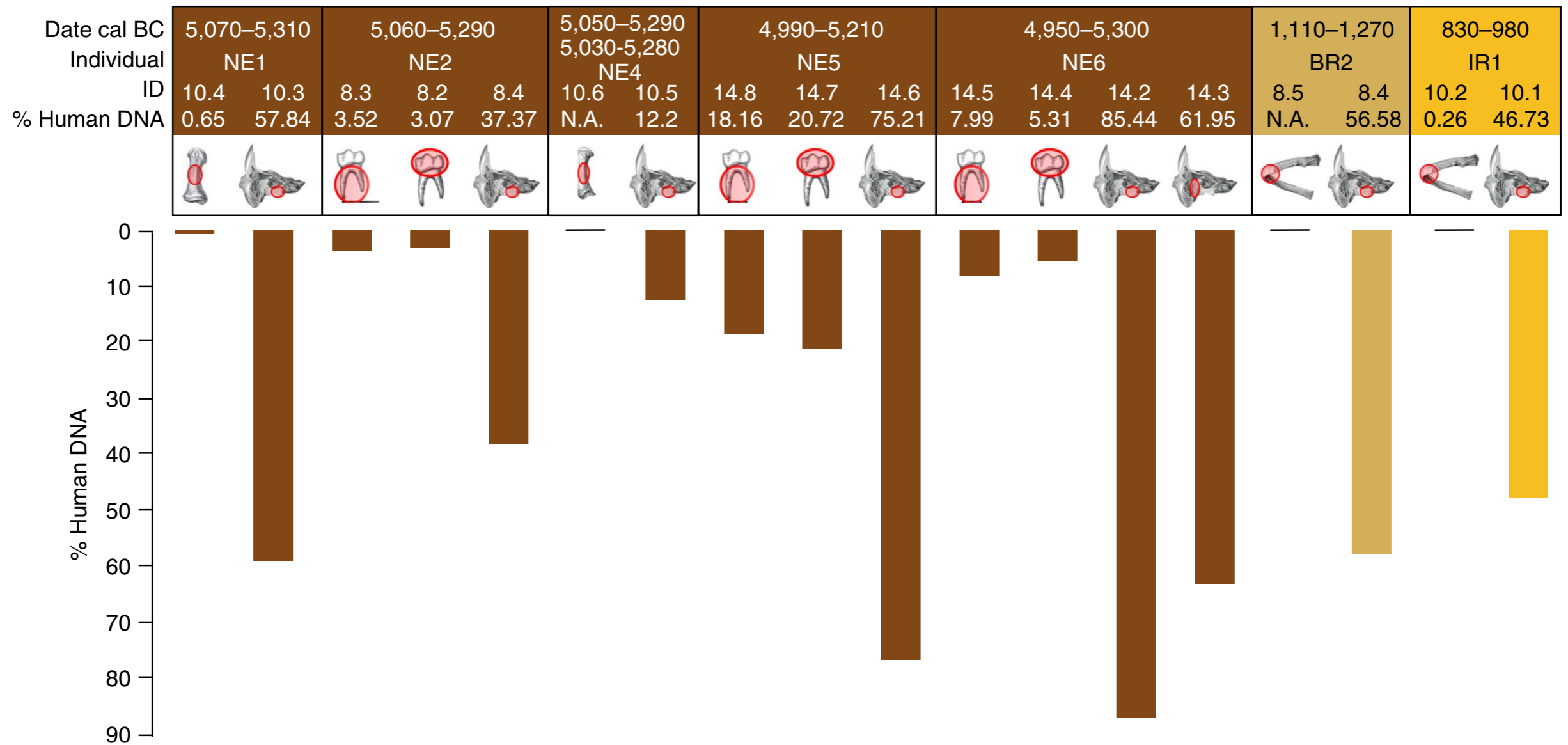
In-silico contamination (databases)

Challenges - DNA accessibility

Presence of DNA in available sample material

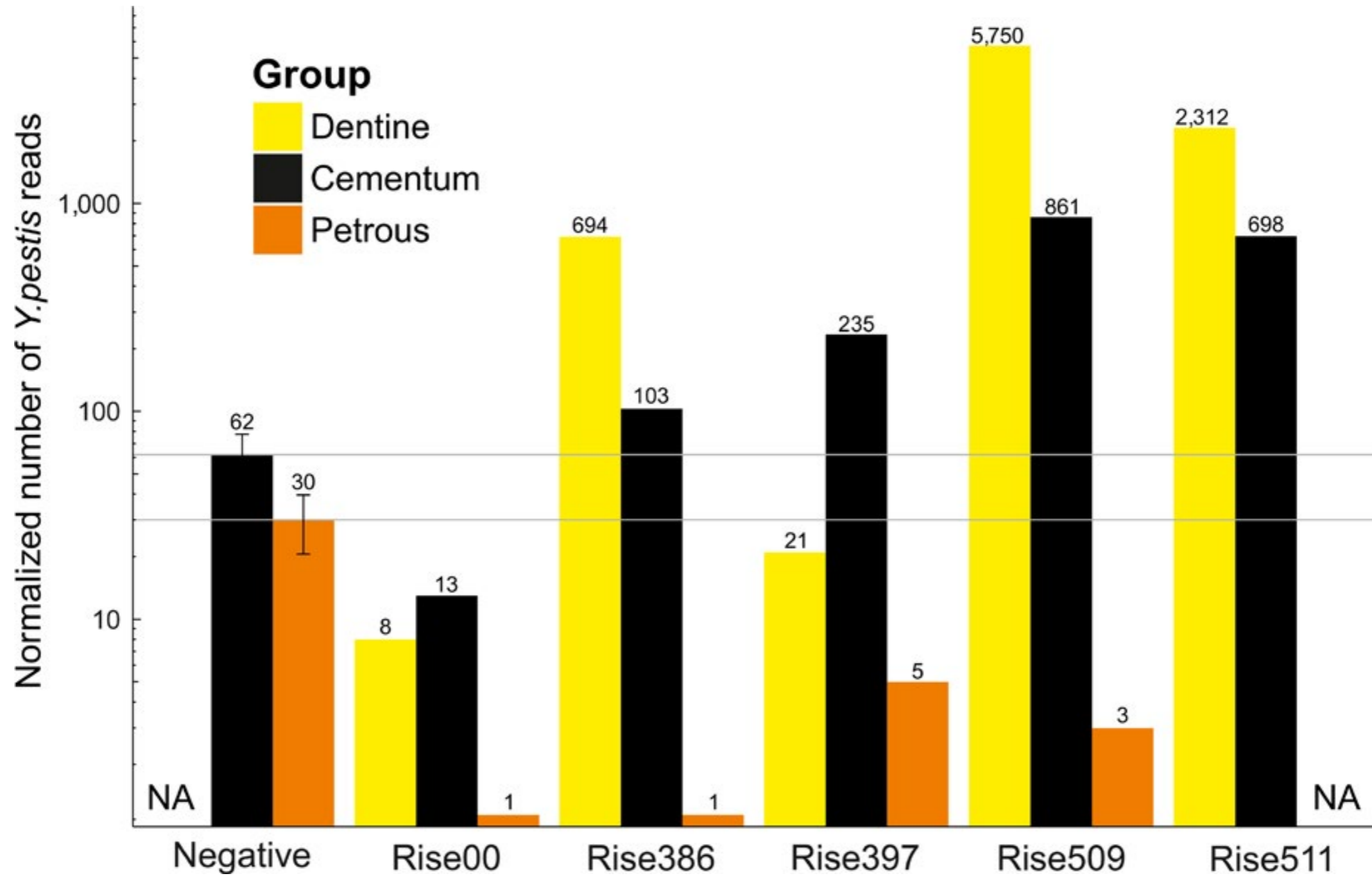


Challenges - DNA accessibility



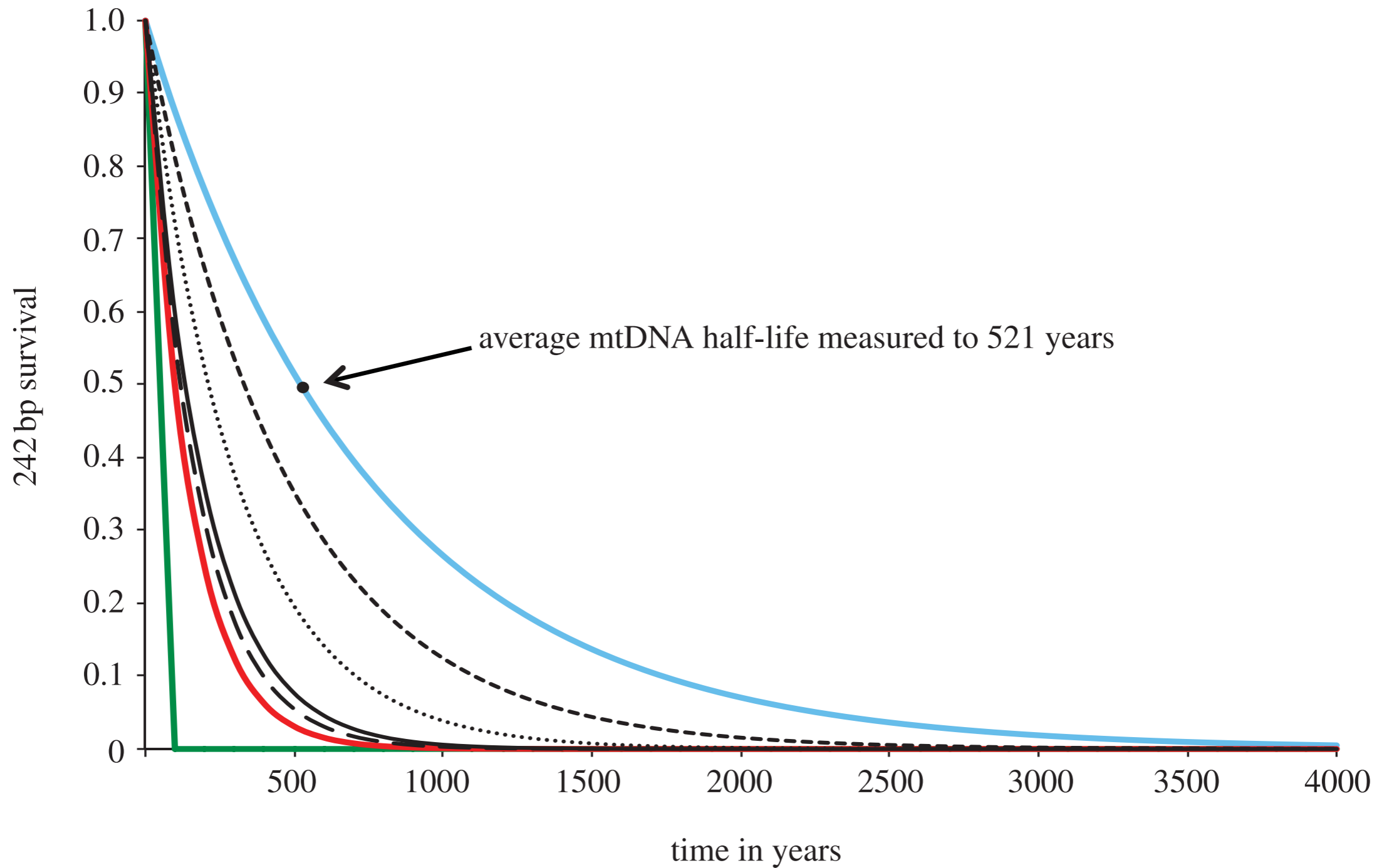
High endogenous DNA yield from petrous portion of temporal bone

Challenges - DNA accessibility



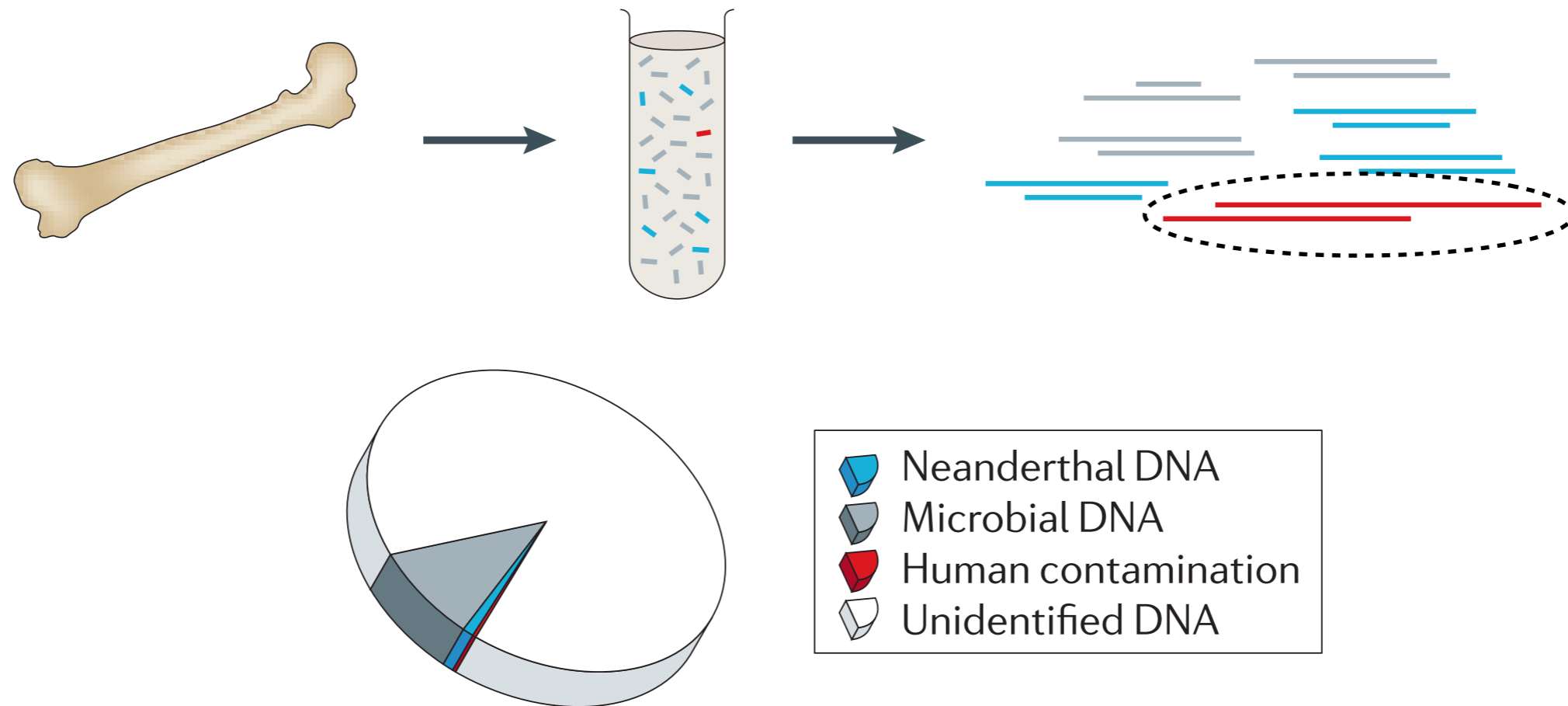
Presence of pathogen DNA in different tissues

Challenges - DNA accessibility



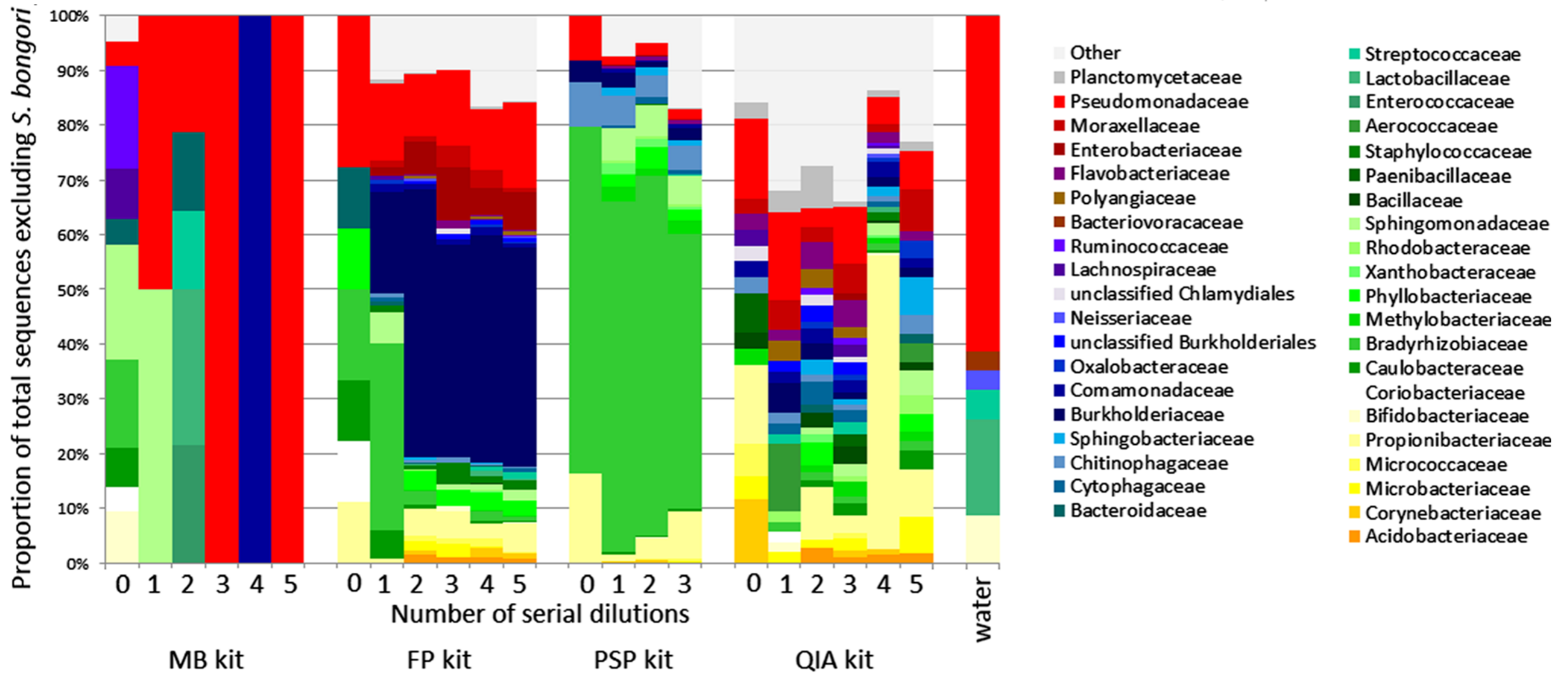
Preservation of ancient DNA affected by environmental conditions

Challenges - DNA authentication



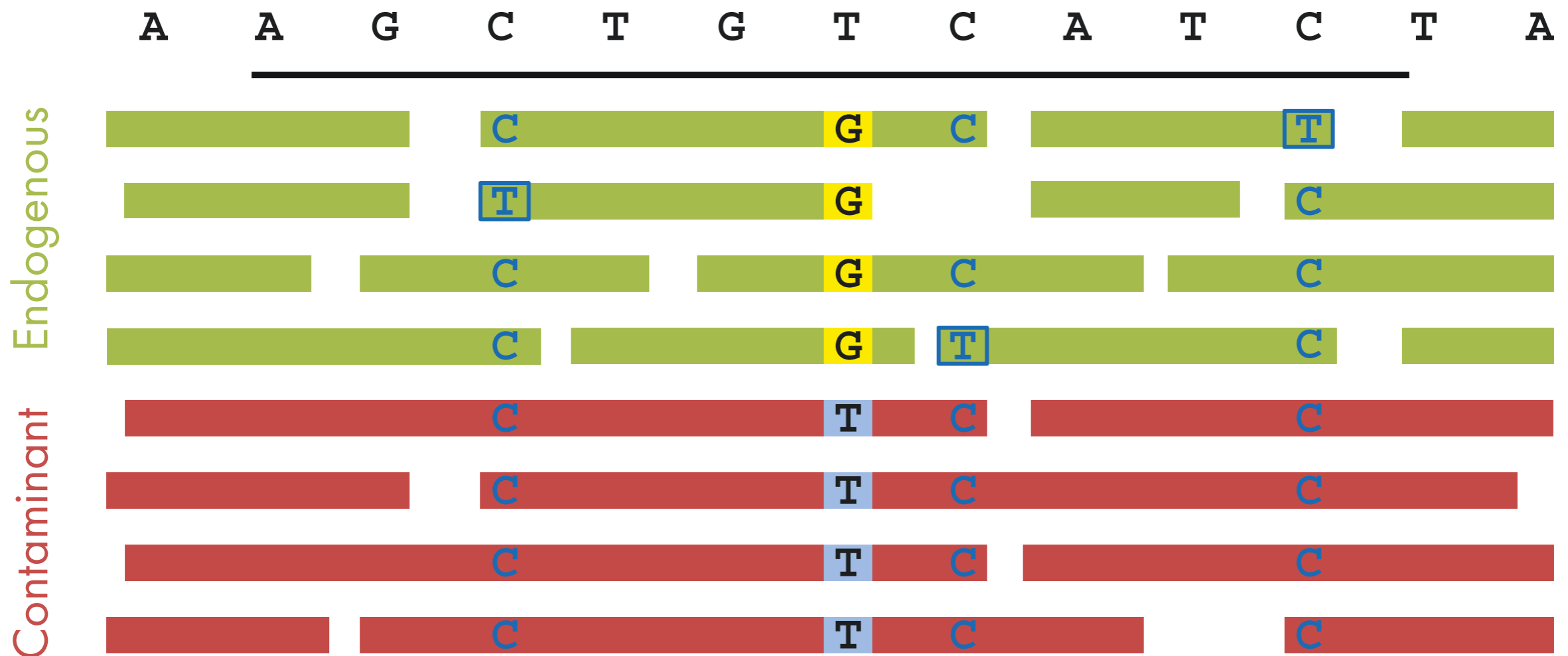
Contamination possible at all stages of processing
Low endogenous content

Challenges - DNA authentication



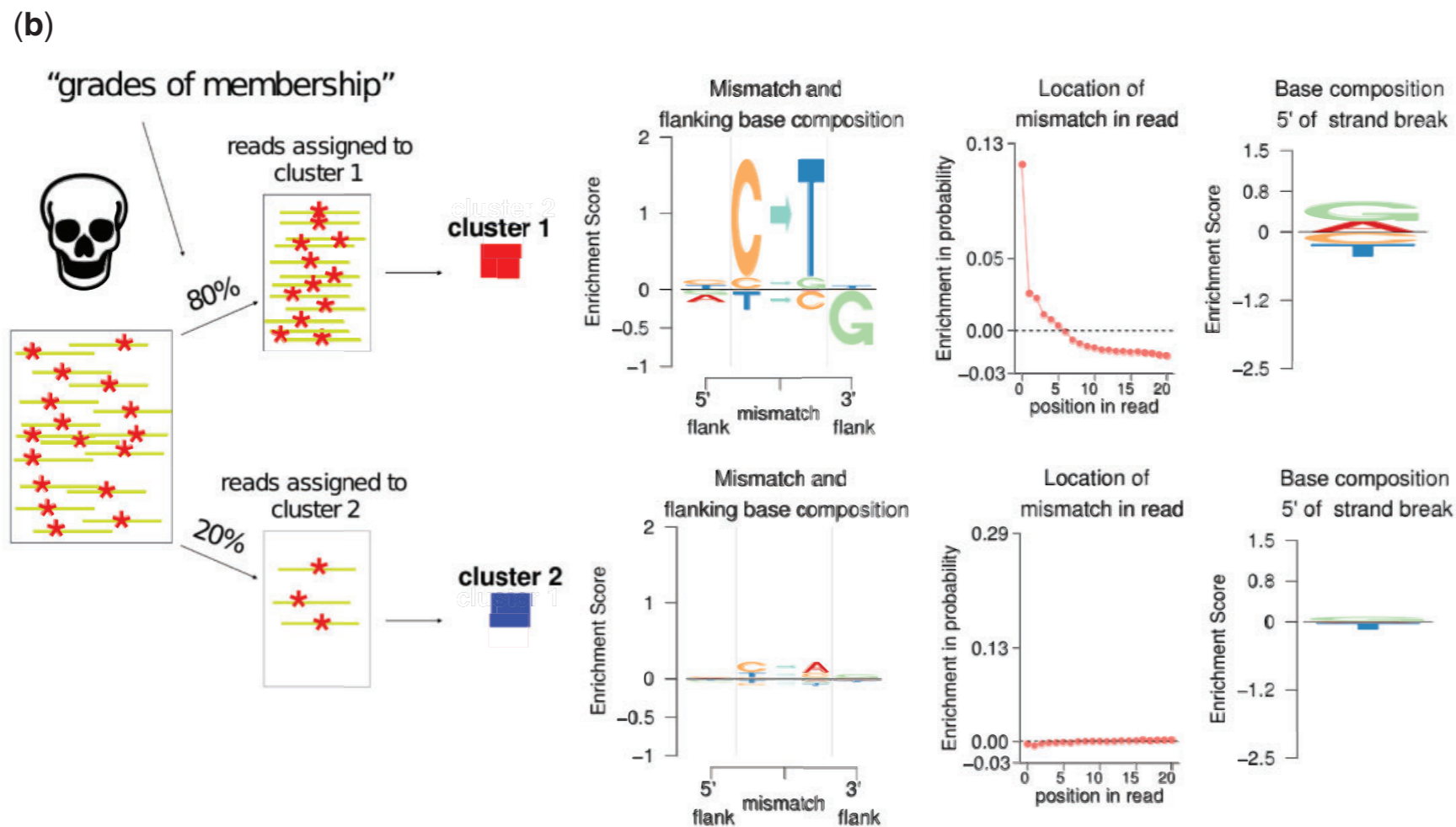
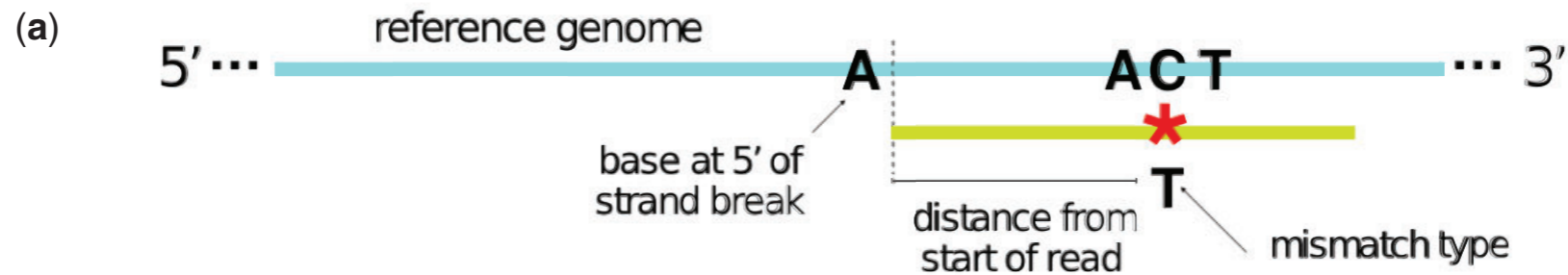
Commonly used laboratory reagents have microbial contamination

Challenges - DNA authentication



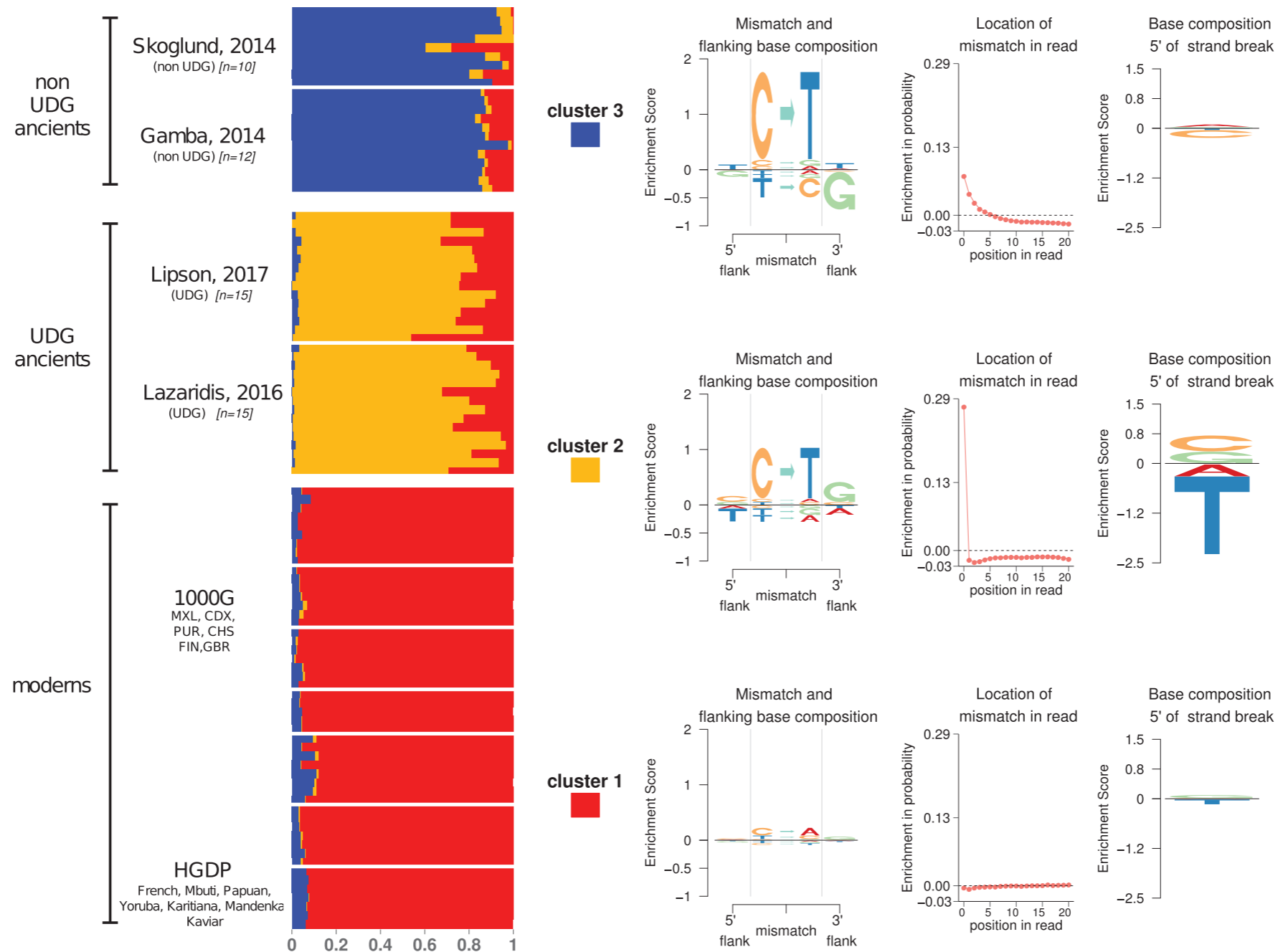
Use characteristics of ancient DNA to estimate fraction of contaminating reads

Challenges - DNA authentication



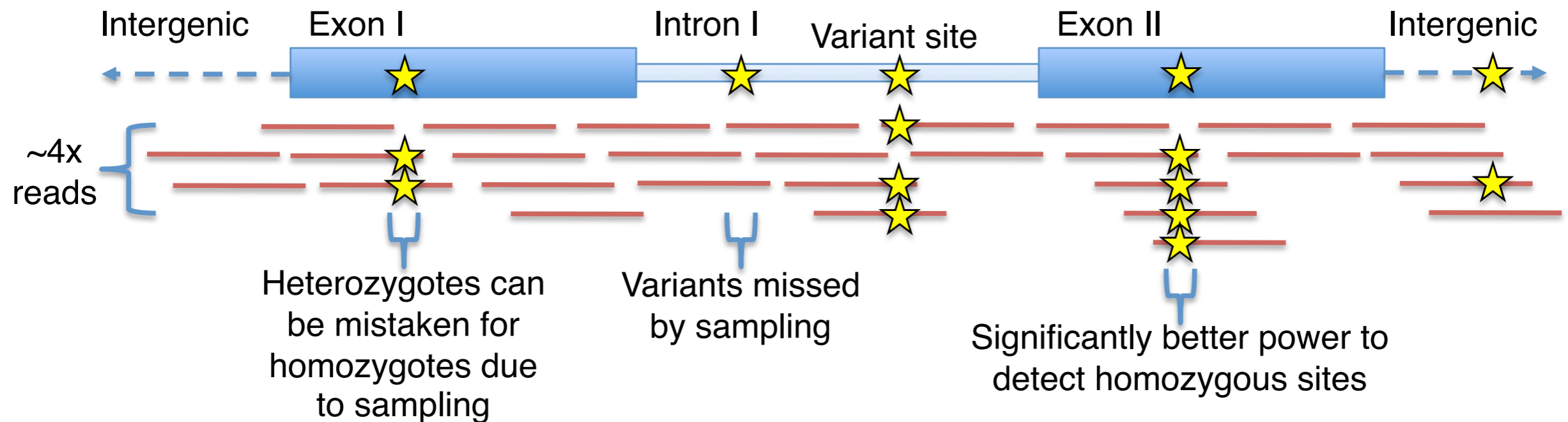
Damage profiles can be used to infer data composition

Challenges - DNA authentication



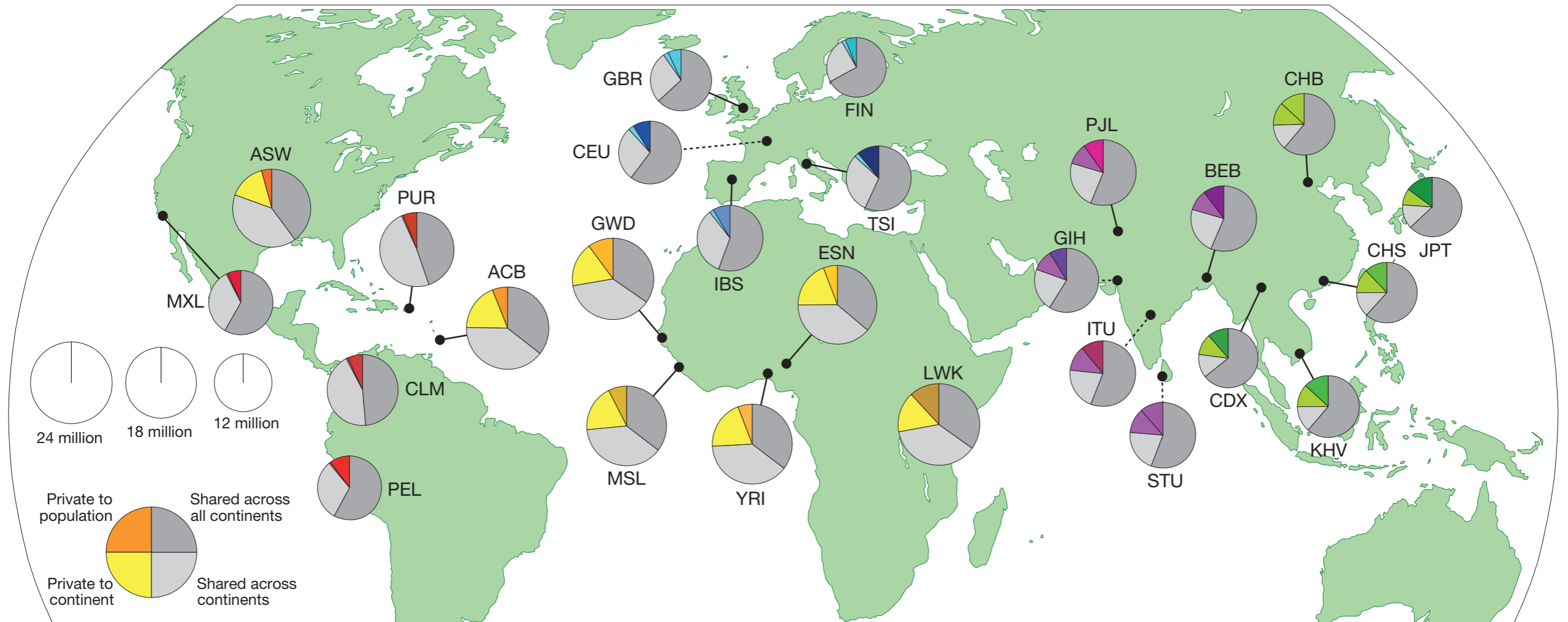
Damage profiles can be used to infer data composition

Challenges - data analysis



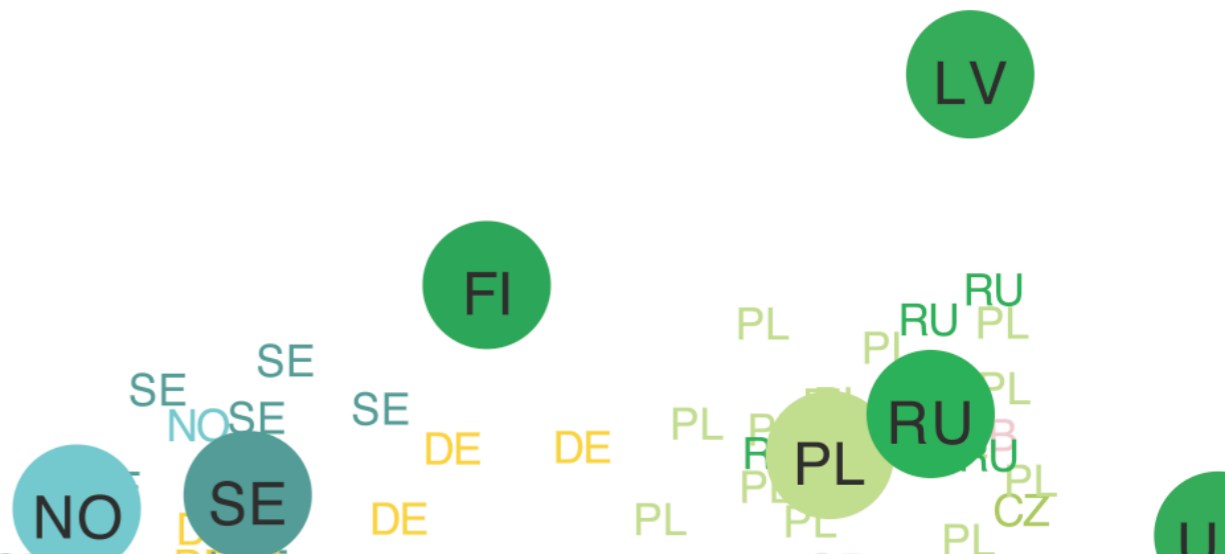
Use pre-ascertained variants to reduce false-positive variants
Use single read as “pseudo-haploid” genotype for each sample

Challenges - data analysis

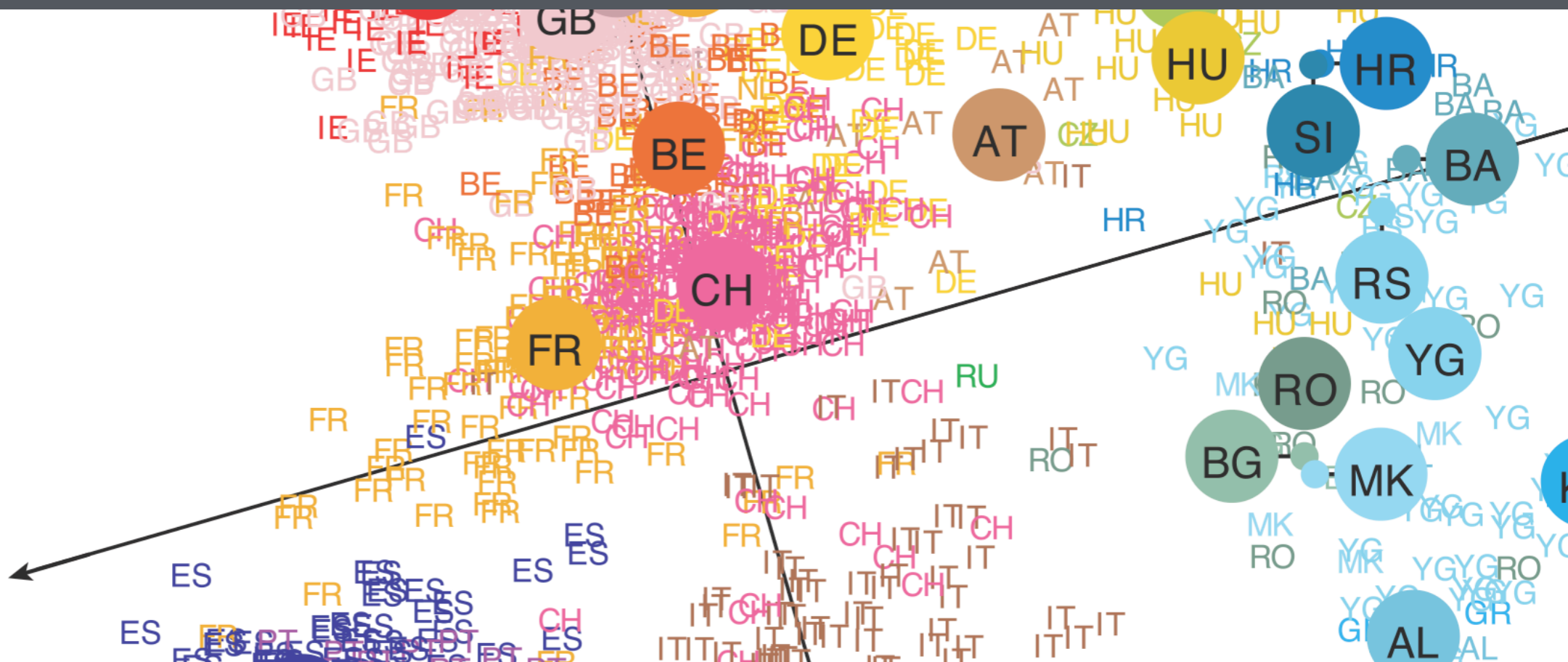


Large panel of individuals covering the genetic diversity relevant for my scientific question

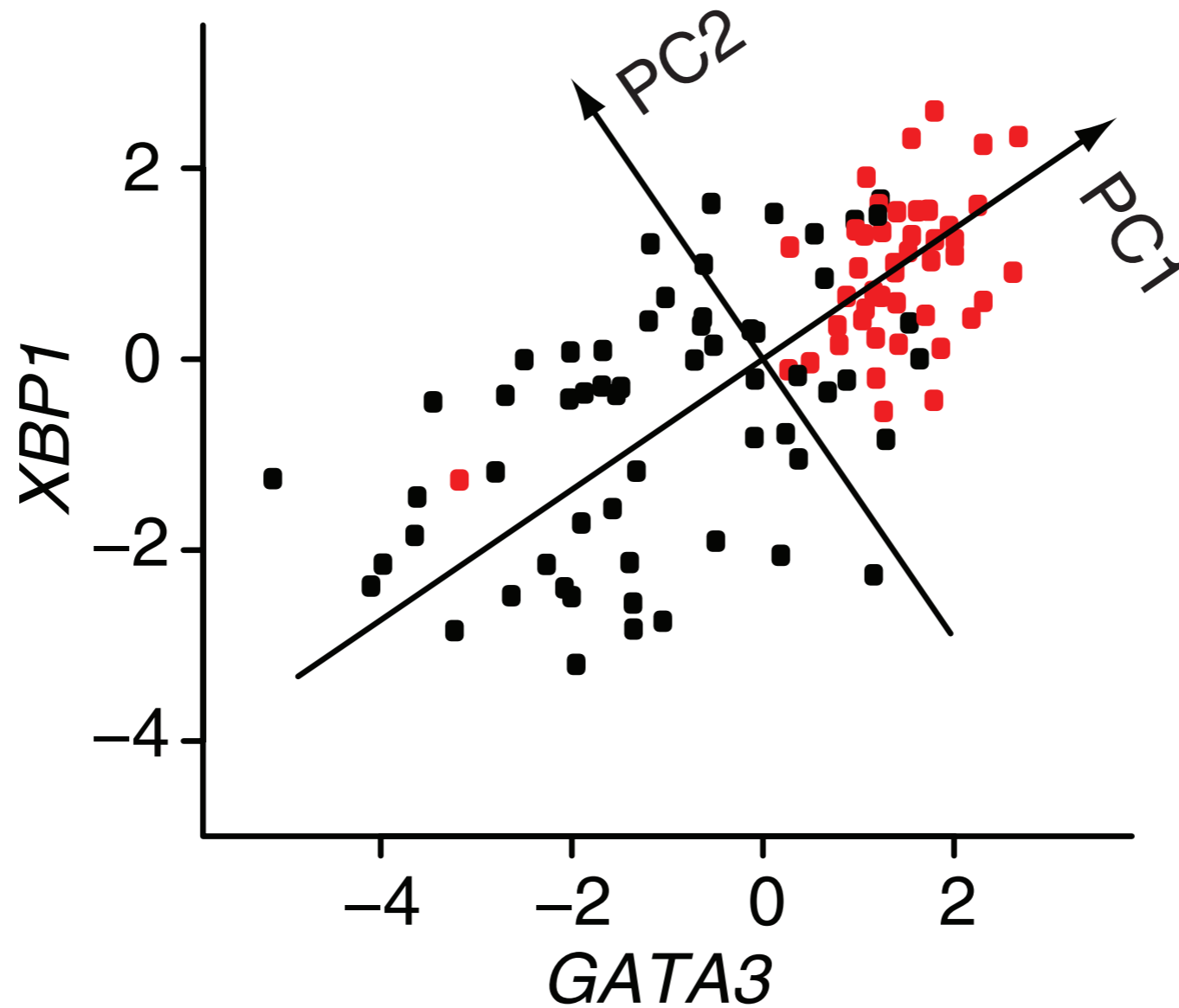
PC1



Analysing ancient DNA

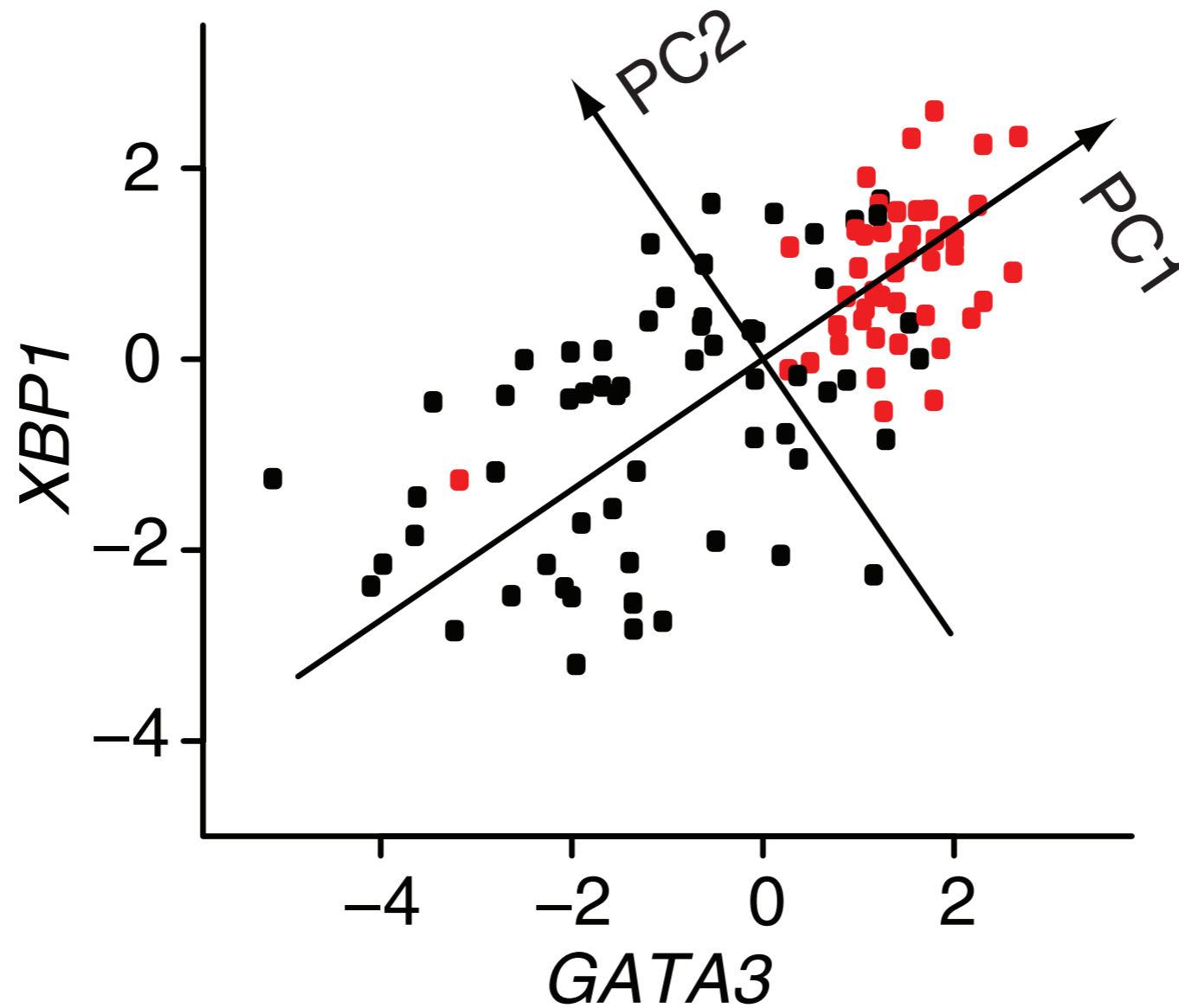


What is principal component analysis?



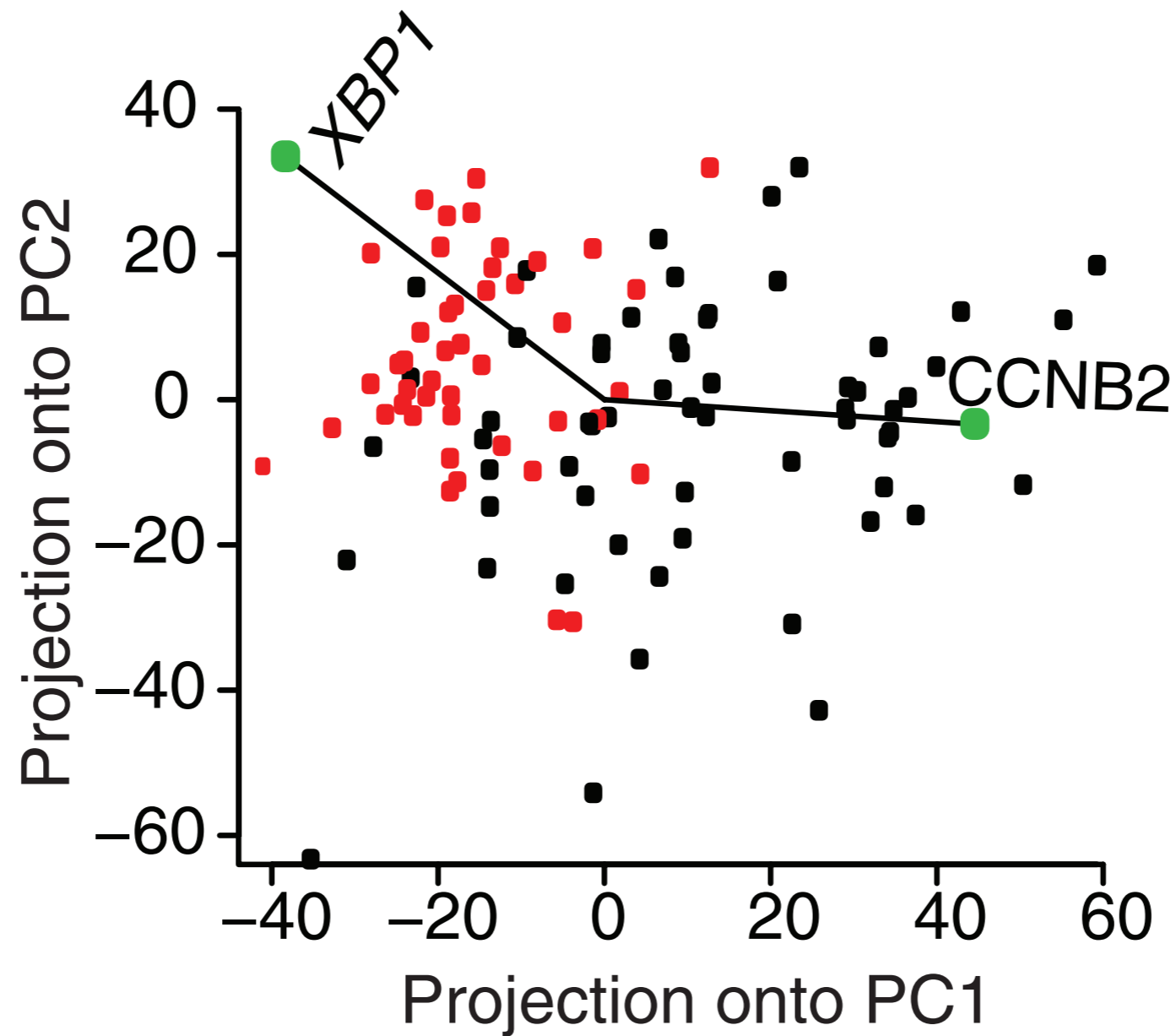
N Samples measured at K variables

What is principal component analysis?



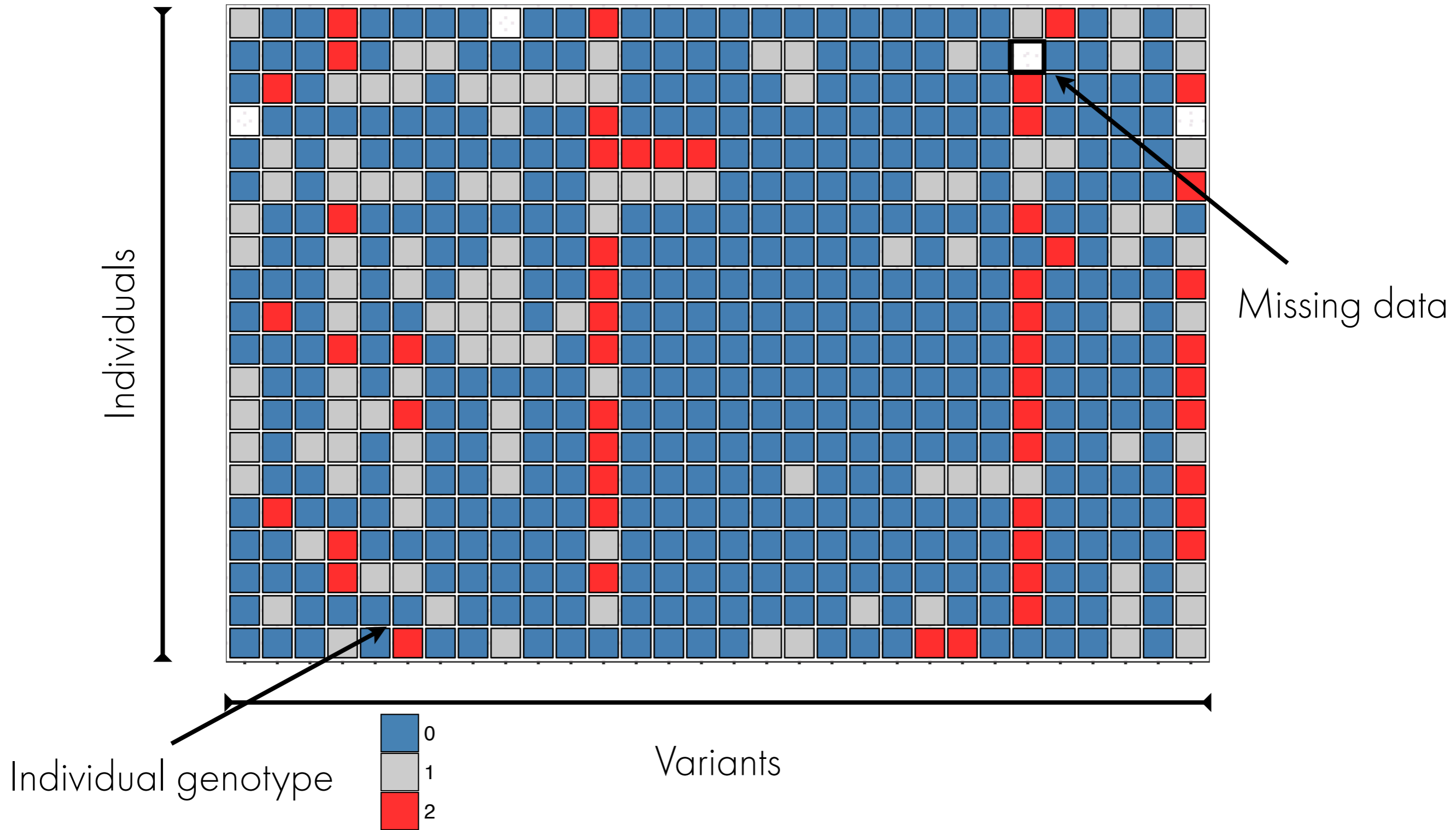
Identify new axes (linear combinations of original variables) that maximise variation in the data

What is principal component analysis?

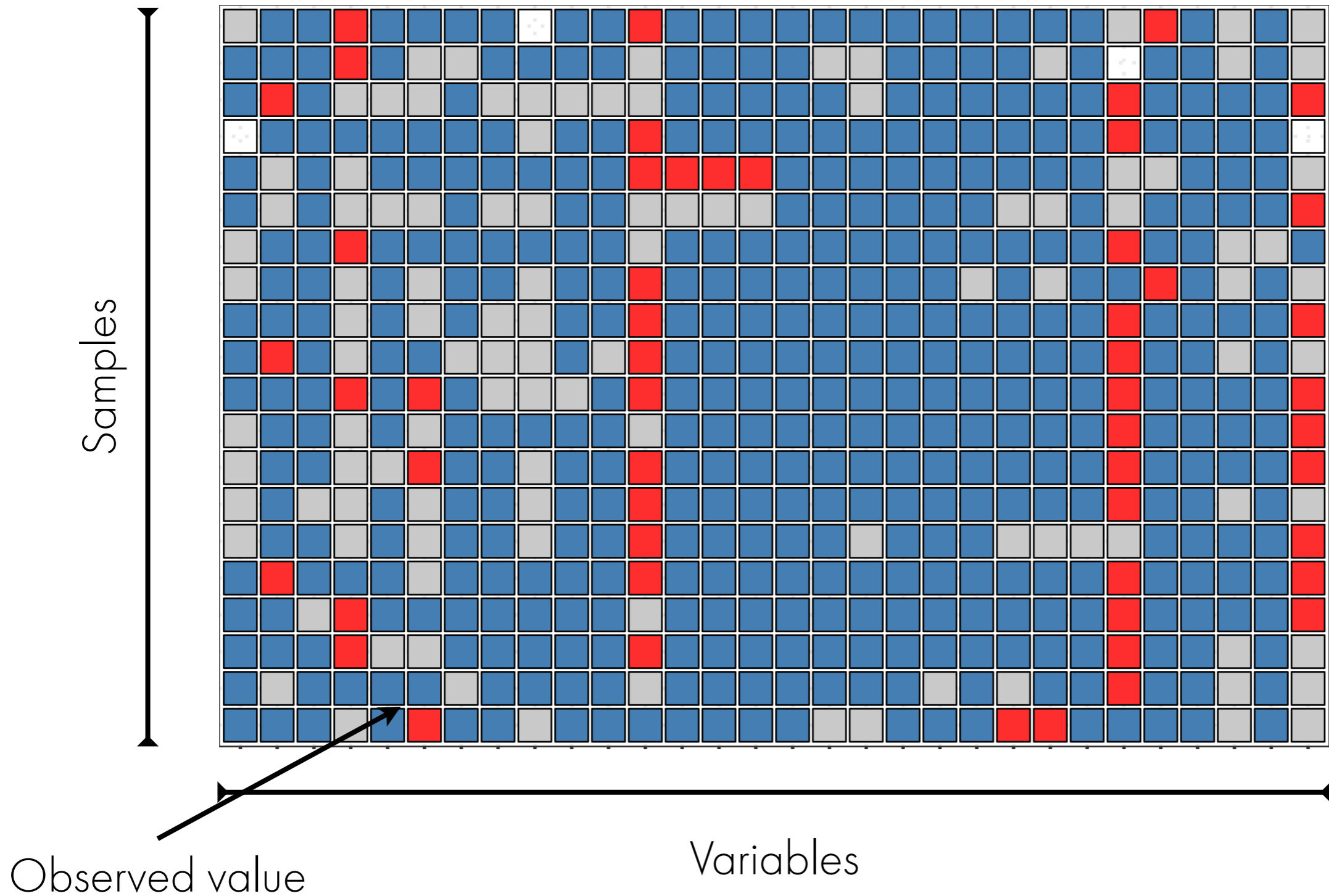


Project samples onto the principal components

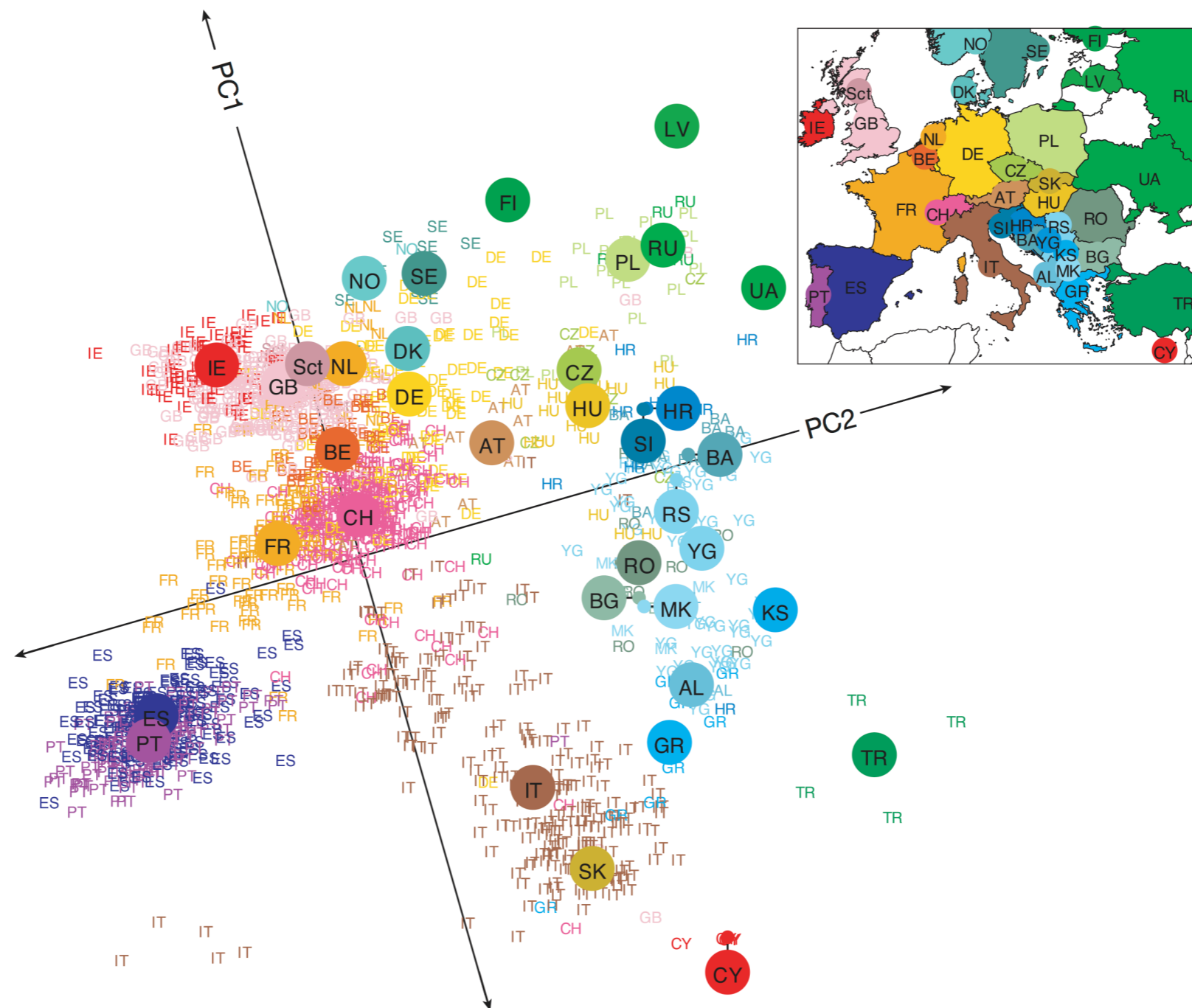
A genotype matrix



PCA on a genotype matrix

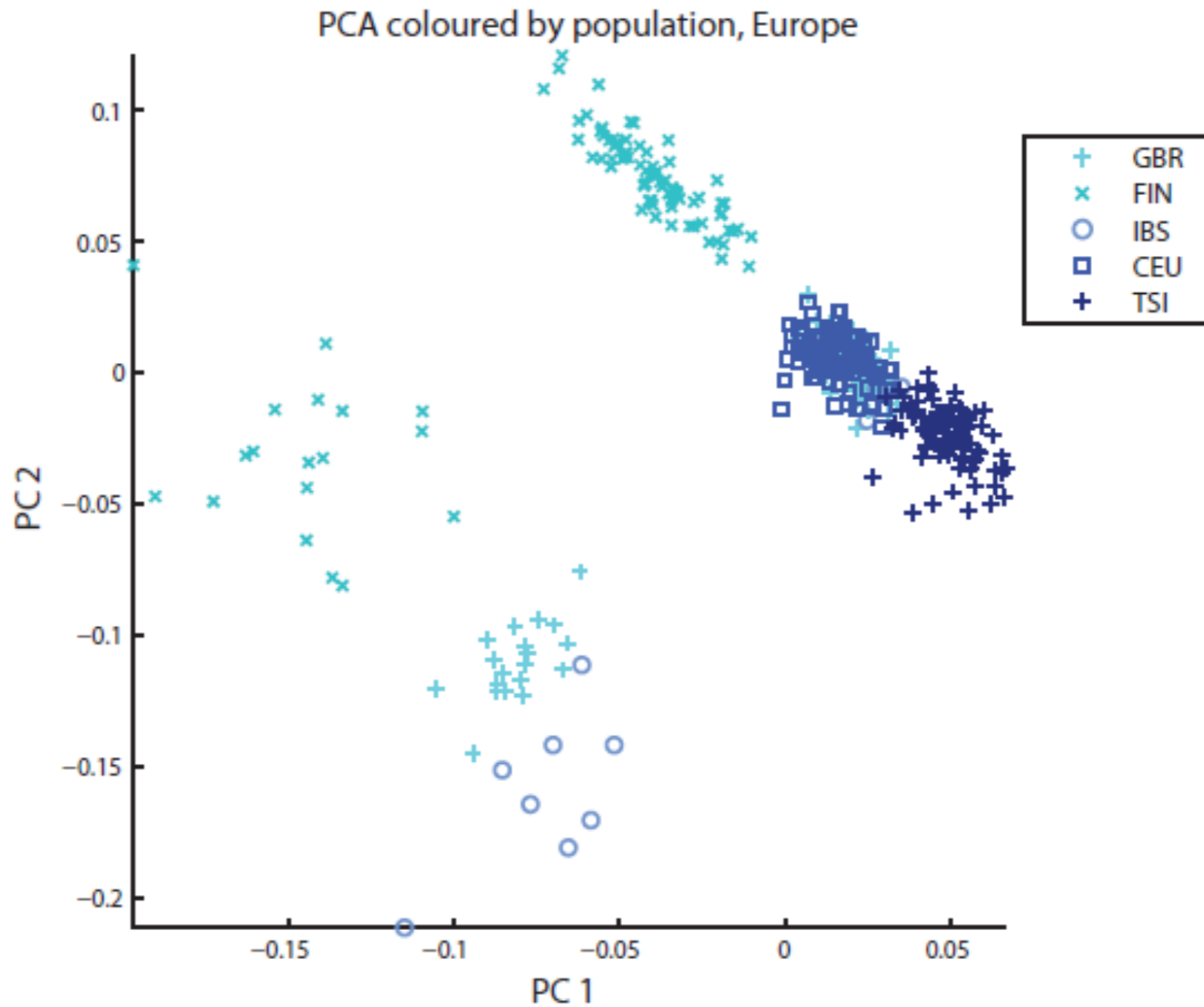


Usage case: Population structure inference

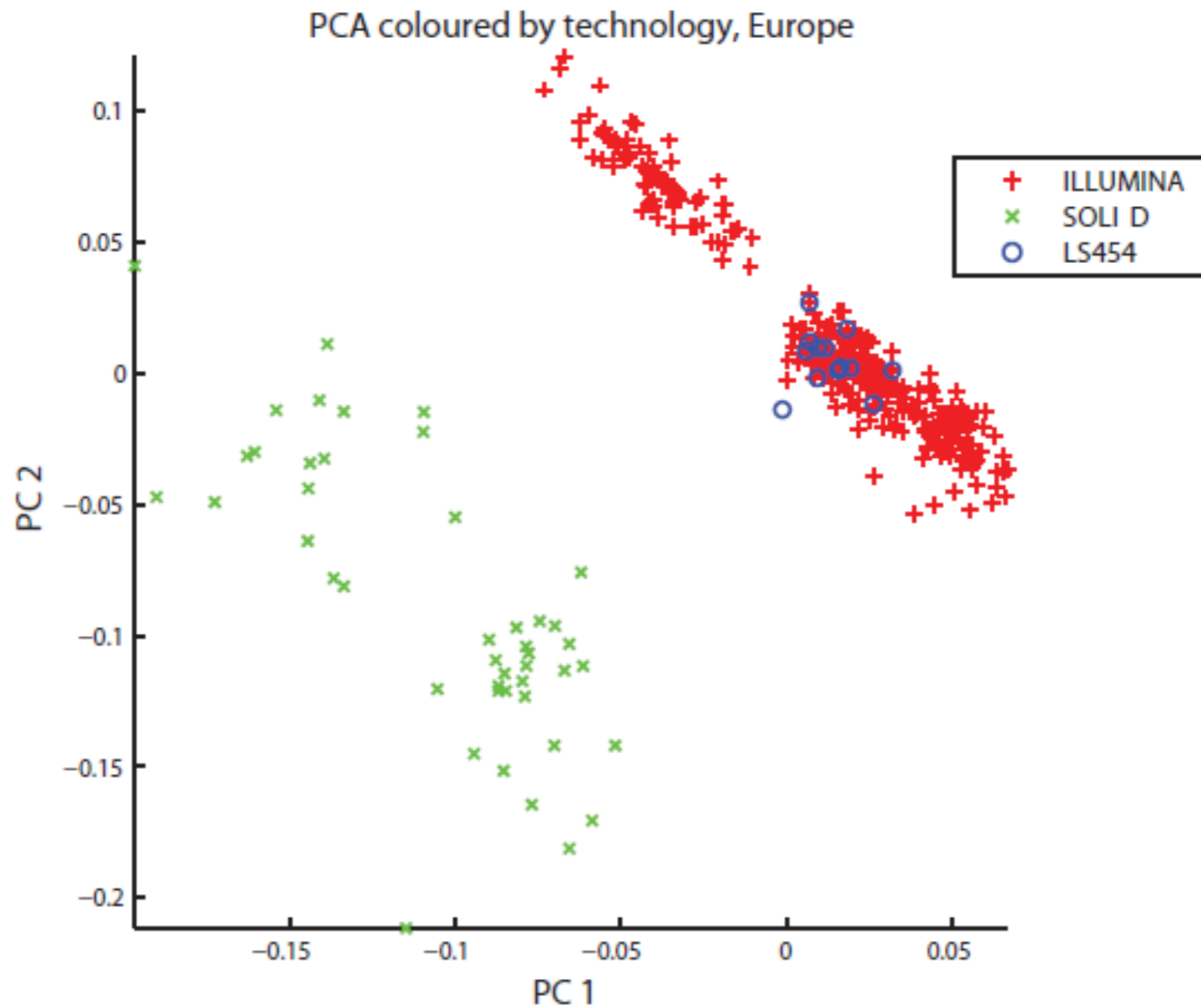


Genes mirror geography in Europe

Usage case: QC

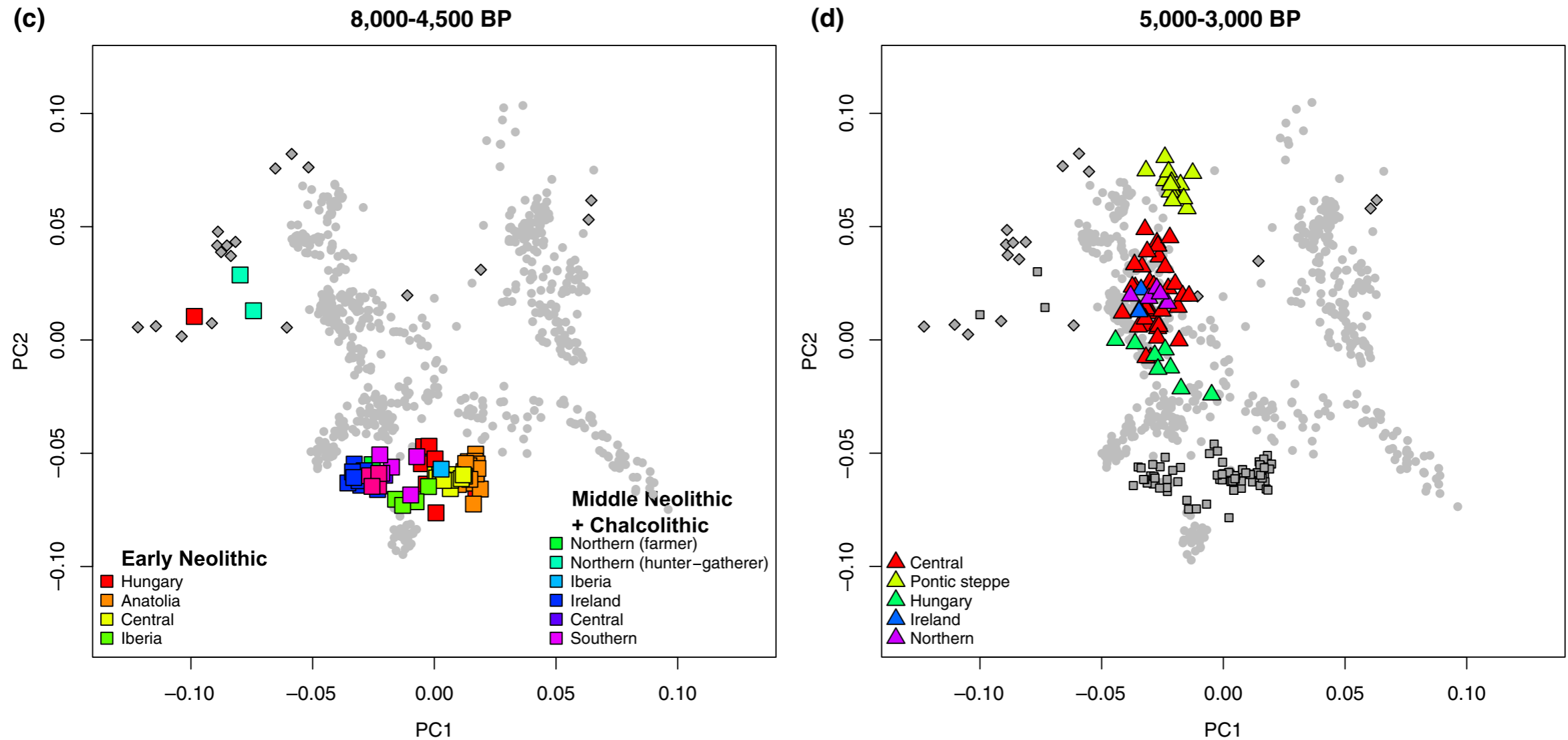


Usage case: QC



PCA can reveal batch effects in datasets

Usage case: Ancient DNA

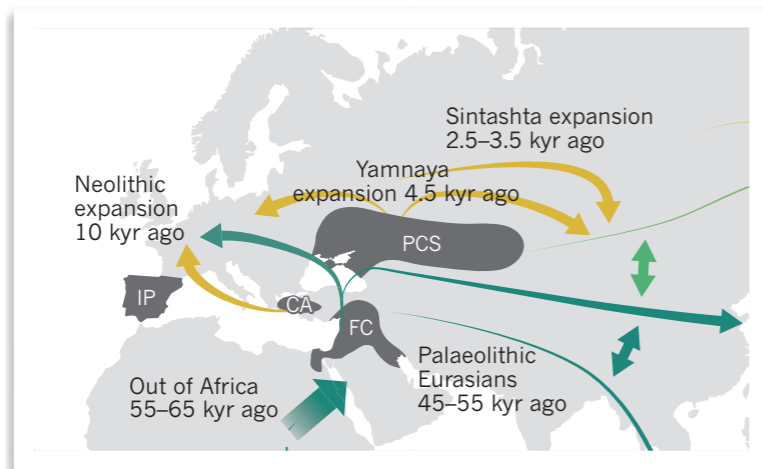


Infer principal components using high quality modern data
Project low coverage ancient samples onto inferred components



What do we study with aDNA?





Migrations

How did modern humans disperse throughout history?



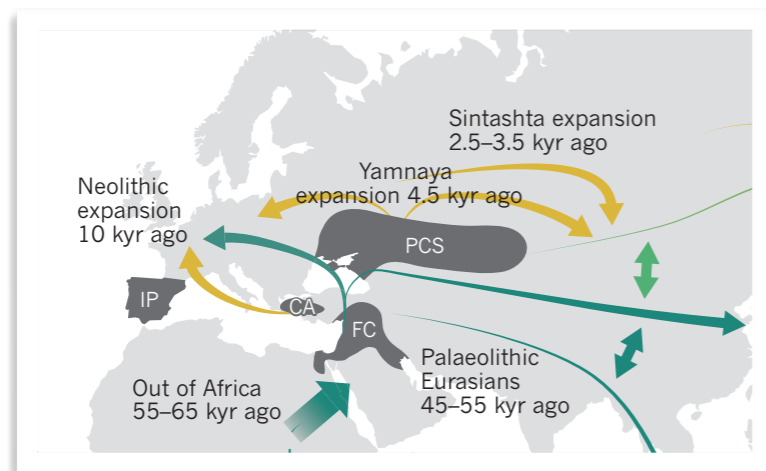
Networks

How were pre-historic human societies structured?



Microbes

What pathogens were affecting humans throughout history?



Migrations

How did modern humans disperse throughout history?



Networks

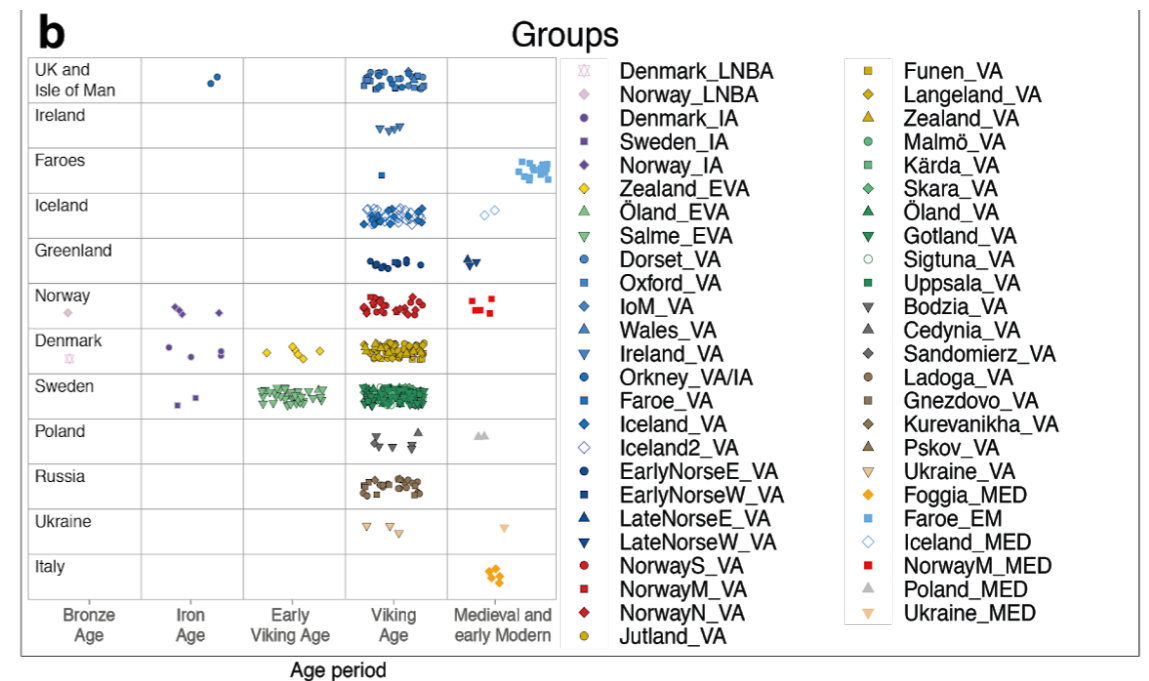
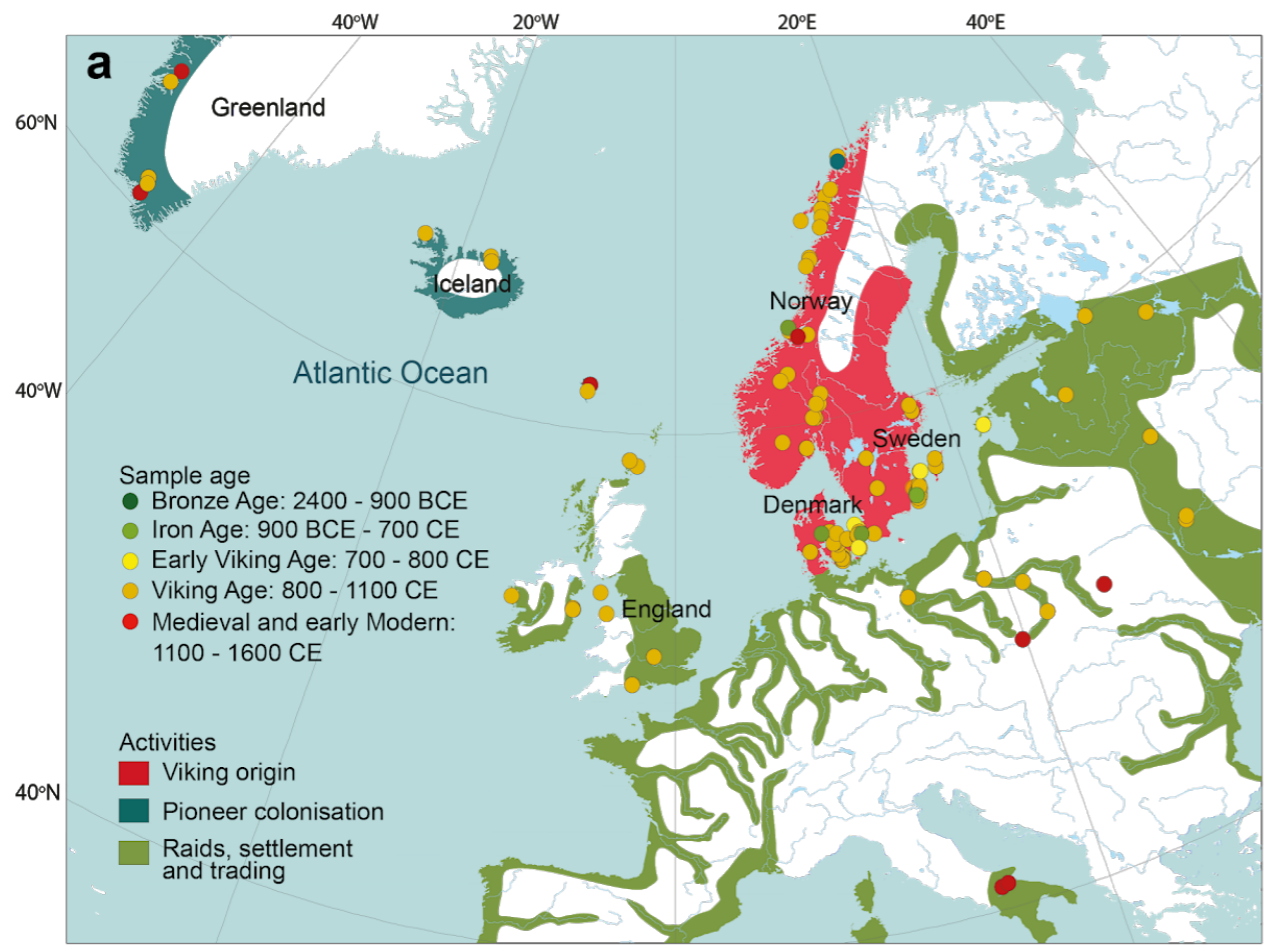
How were pre-historic human societies structured?



Microbes

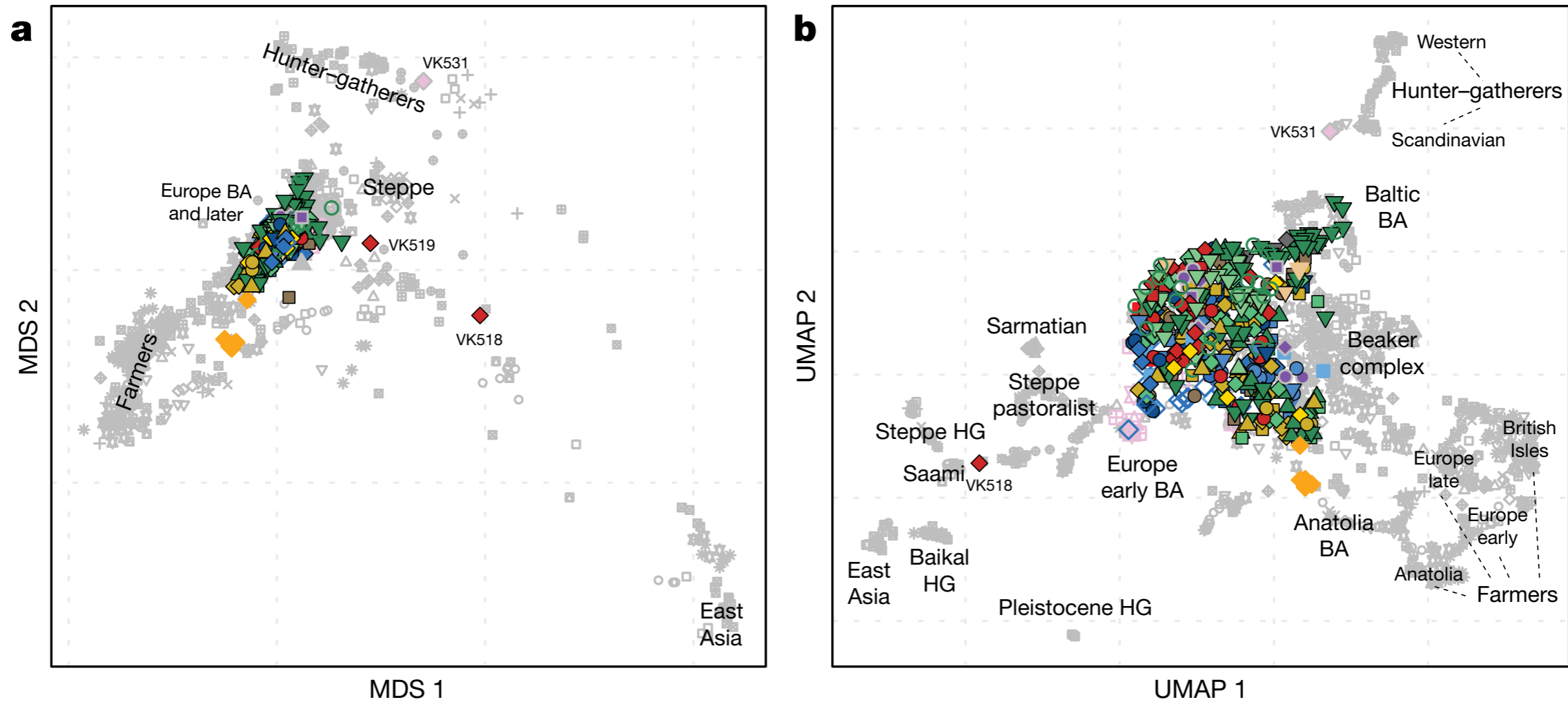
What pathogens were affecting humans throughout history?

Ancient genomics of the Viking Age



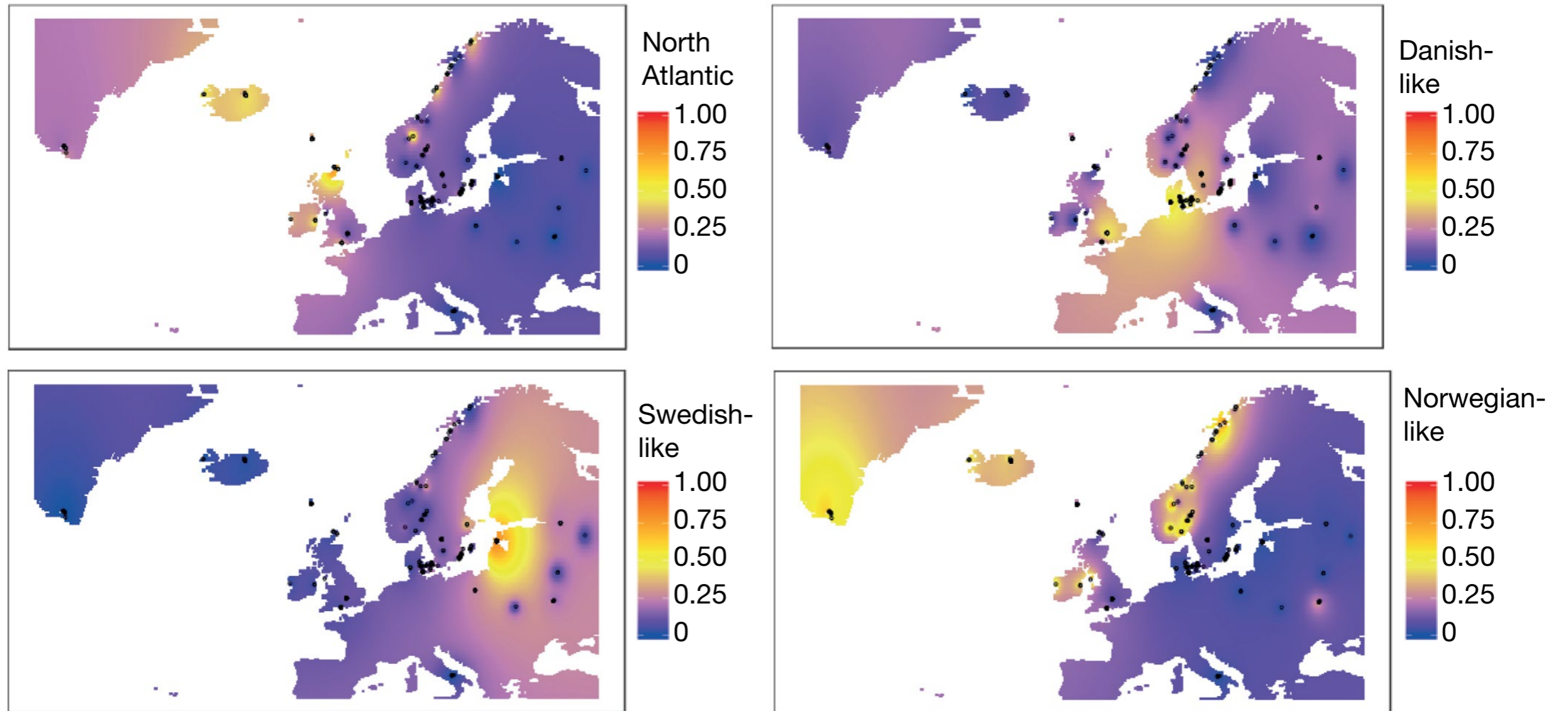
Population genomics using over 400 Viking Age genomes

Ancient genomics of the Viking Age



Population genomics using over 400 Viking Age genomes

Ancient genomics of the Viking Age



Viking Age migrations



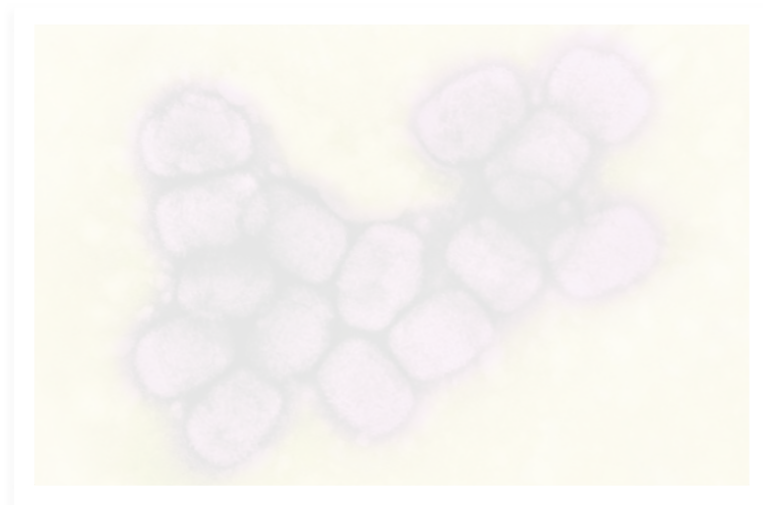
Migrations

How did modern humans disperse throughout history?



Networks

How were pre-historic human societies structured?



Microbes

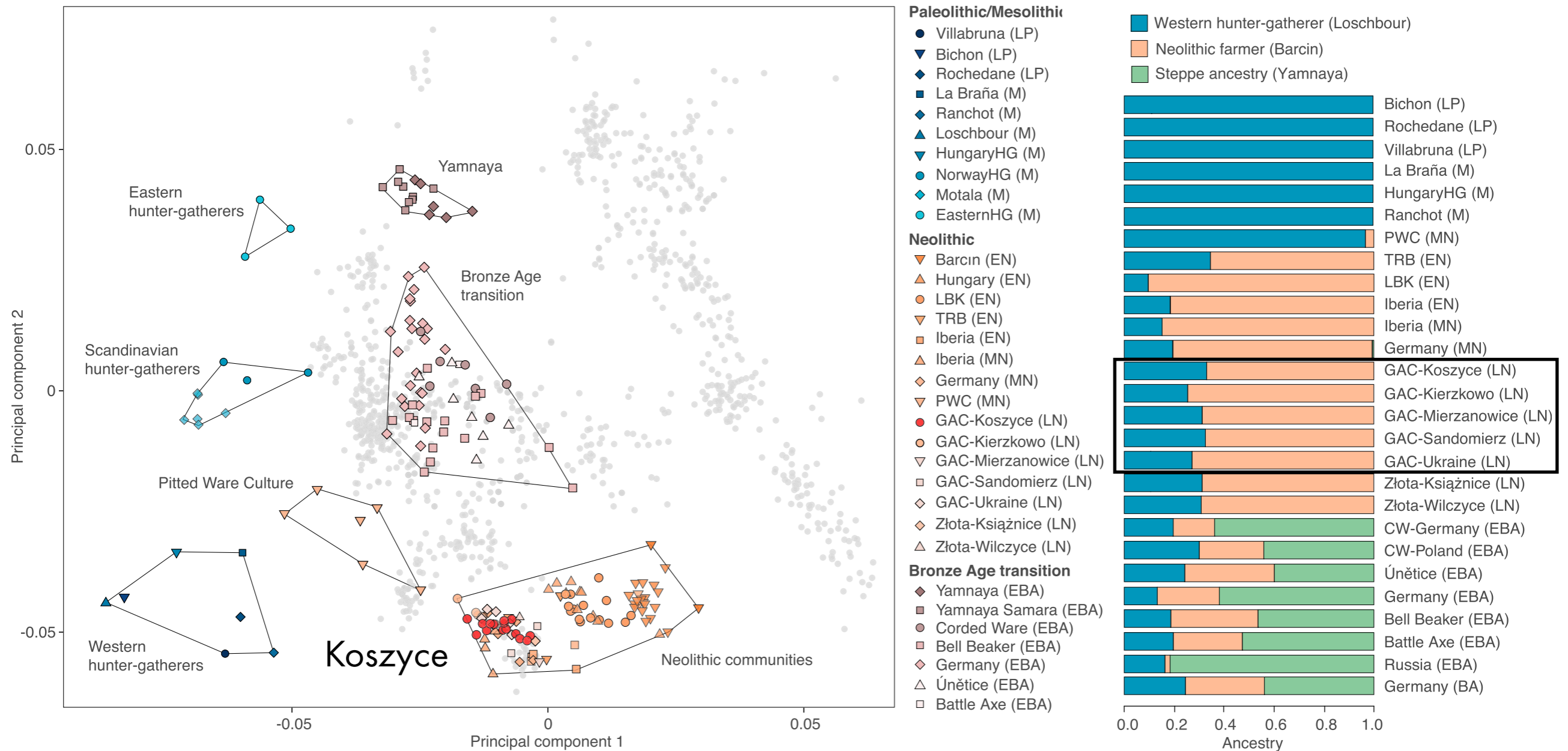
What pathogens were affecting humans throughout history?

Kinship and ancestry in a late Neolithic mass grave



15 individuals killed with blows to their heads
Associated with late Neolithic Globular Amphora culture

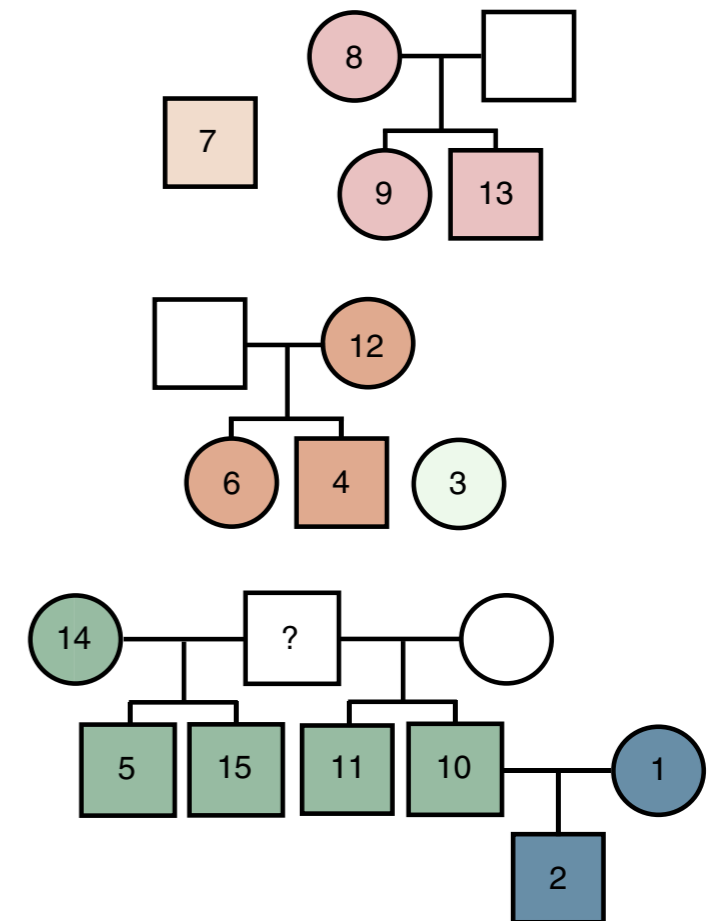
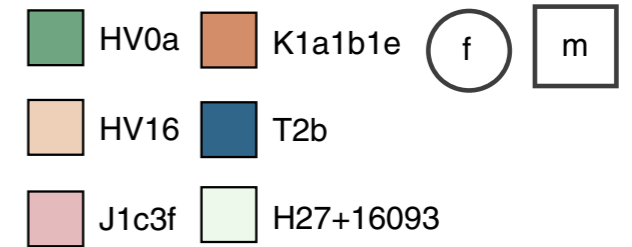
Late Neolithic ancestry peoples of Europe



Social structure of a late Neolithic community

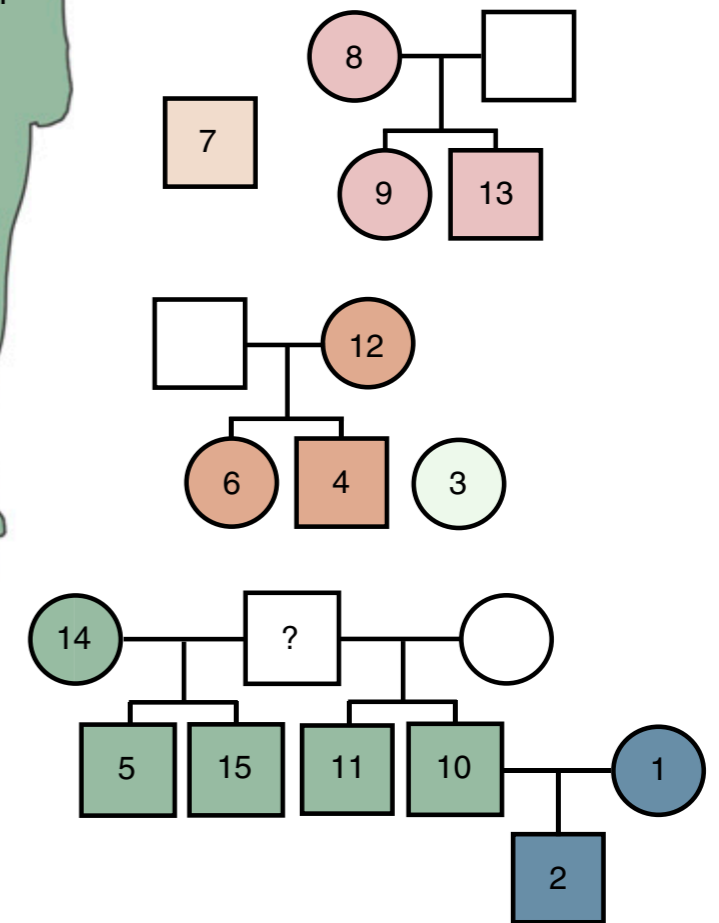
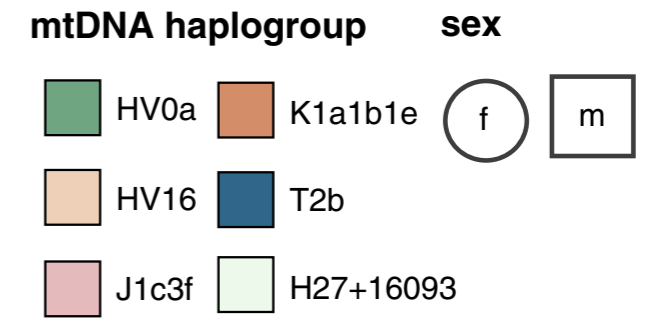
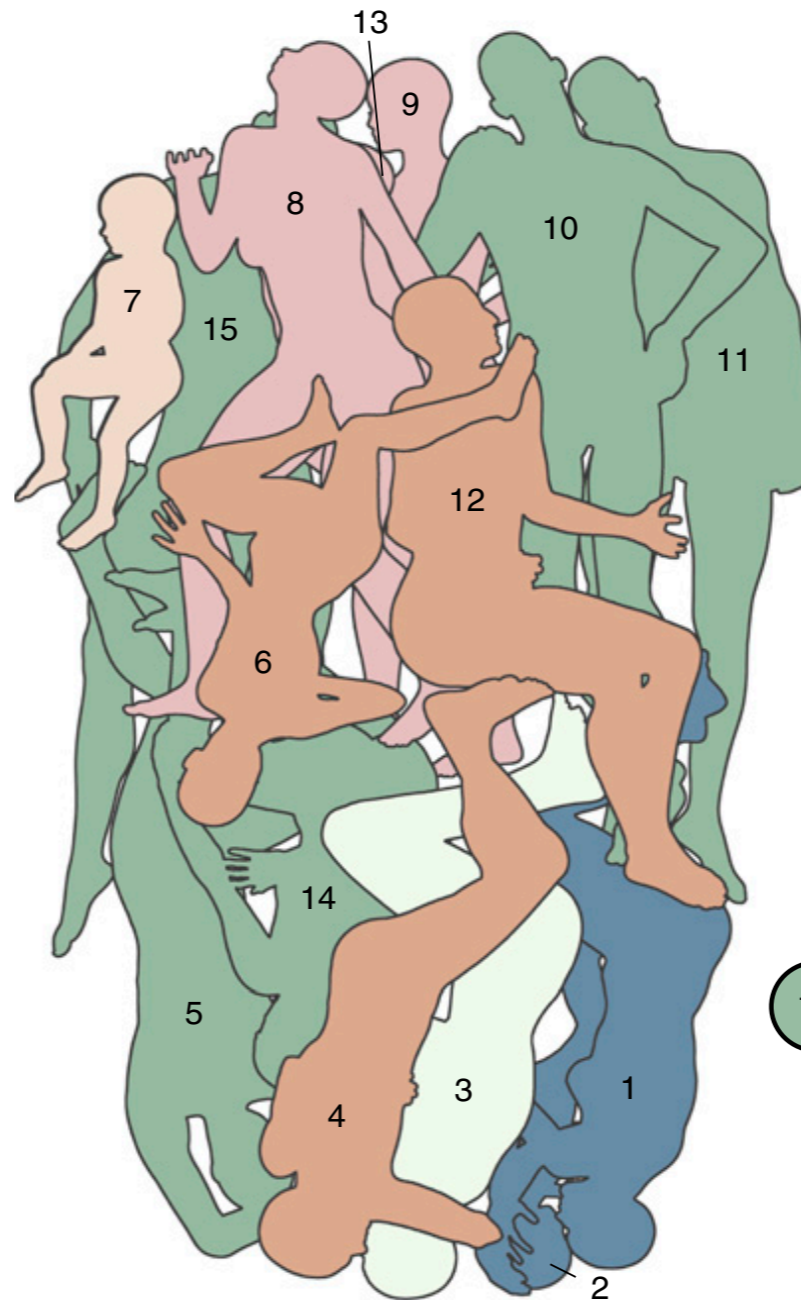


mtDNA haplogroup sex

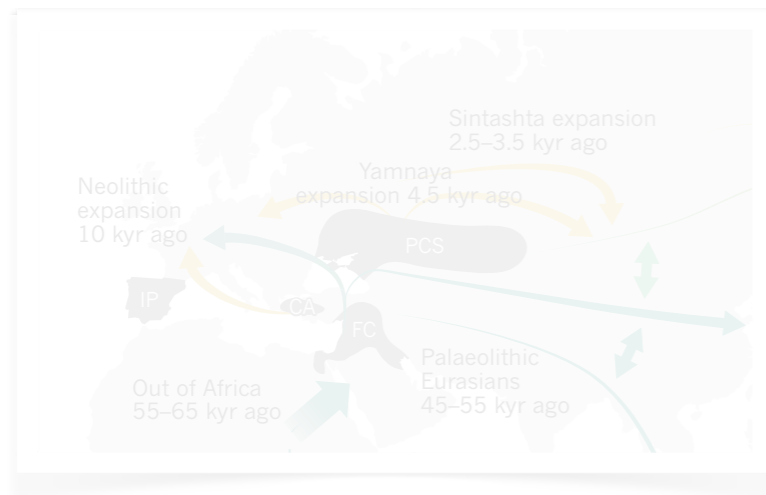


Unrelated females with children related through male lineage

Social structure of a late Neolithic community

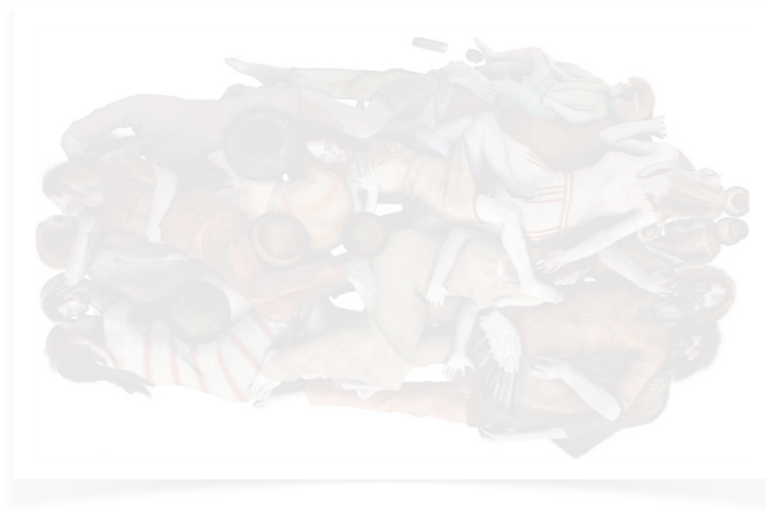


Burial positions shows individuals were buried by their kin



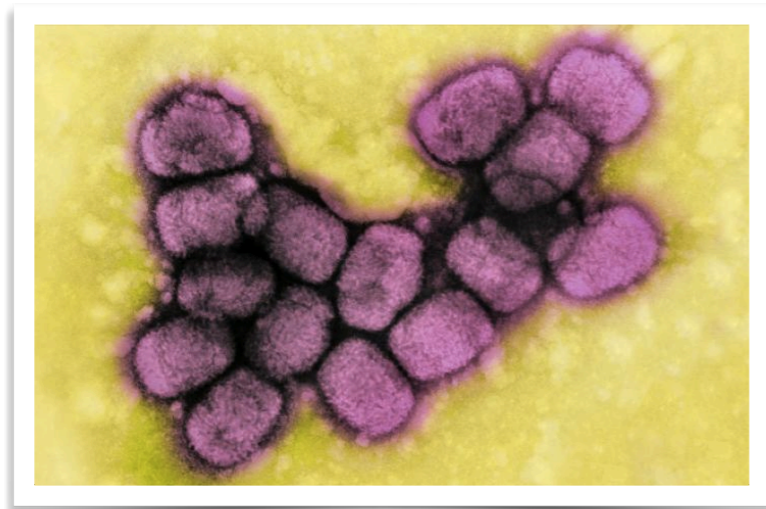
Migrations

How did modern humans disperse throughout history?



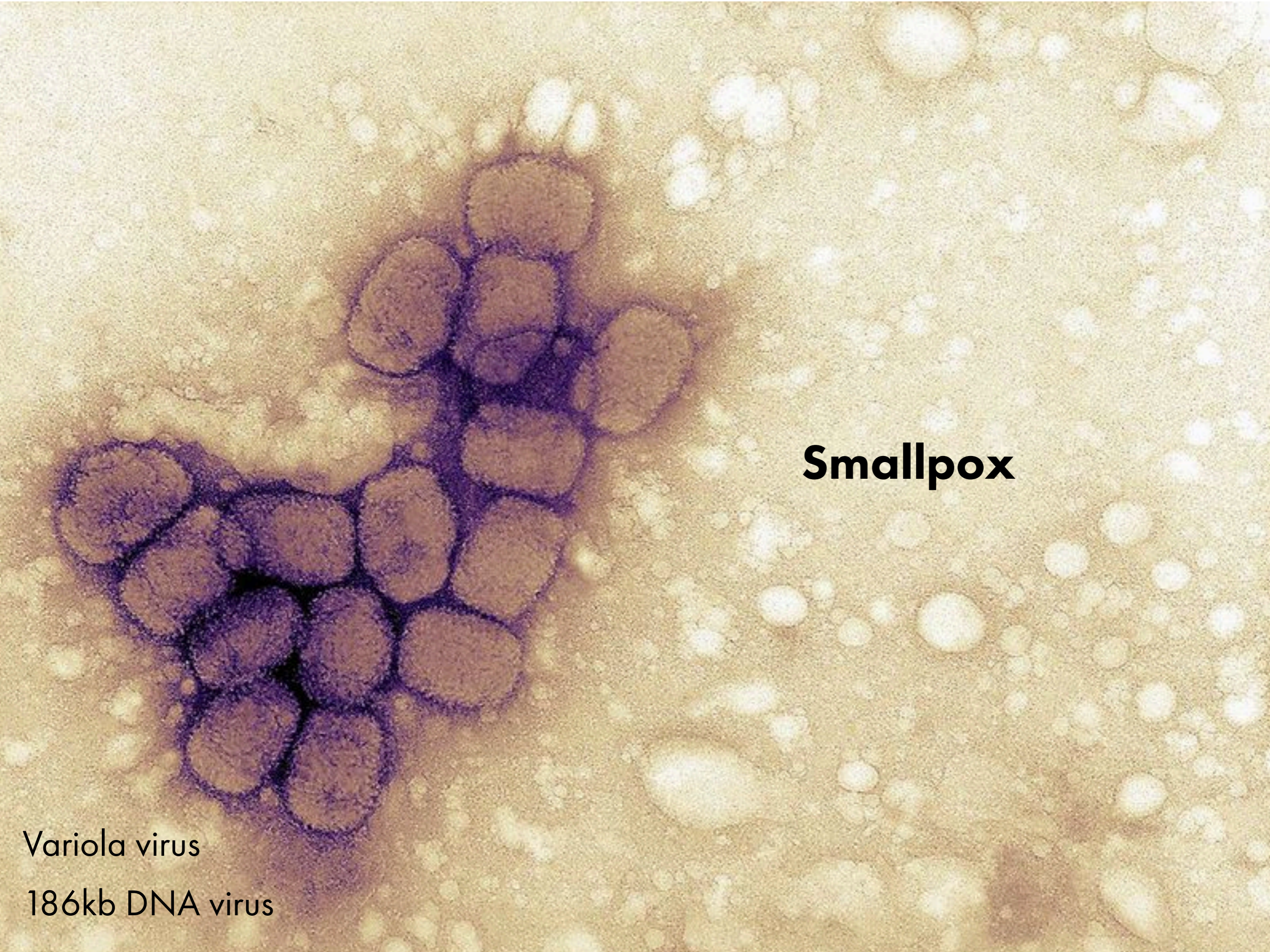
Networks

How were pre-historic human societies structured?



Microbes

What pathogens were affecting humans throughout history?



Smallpox

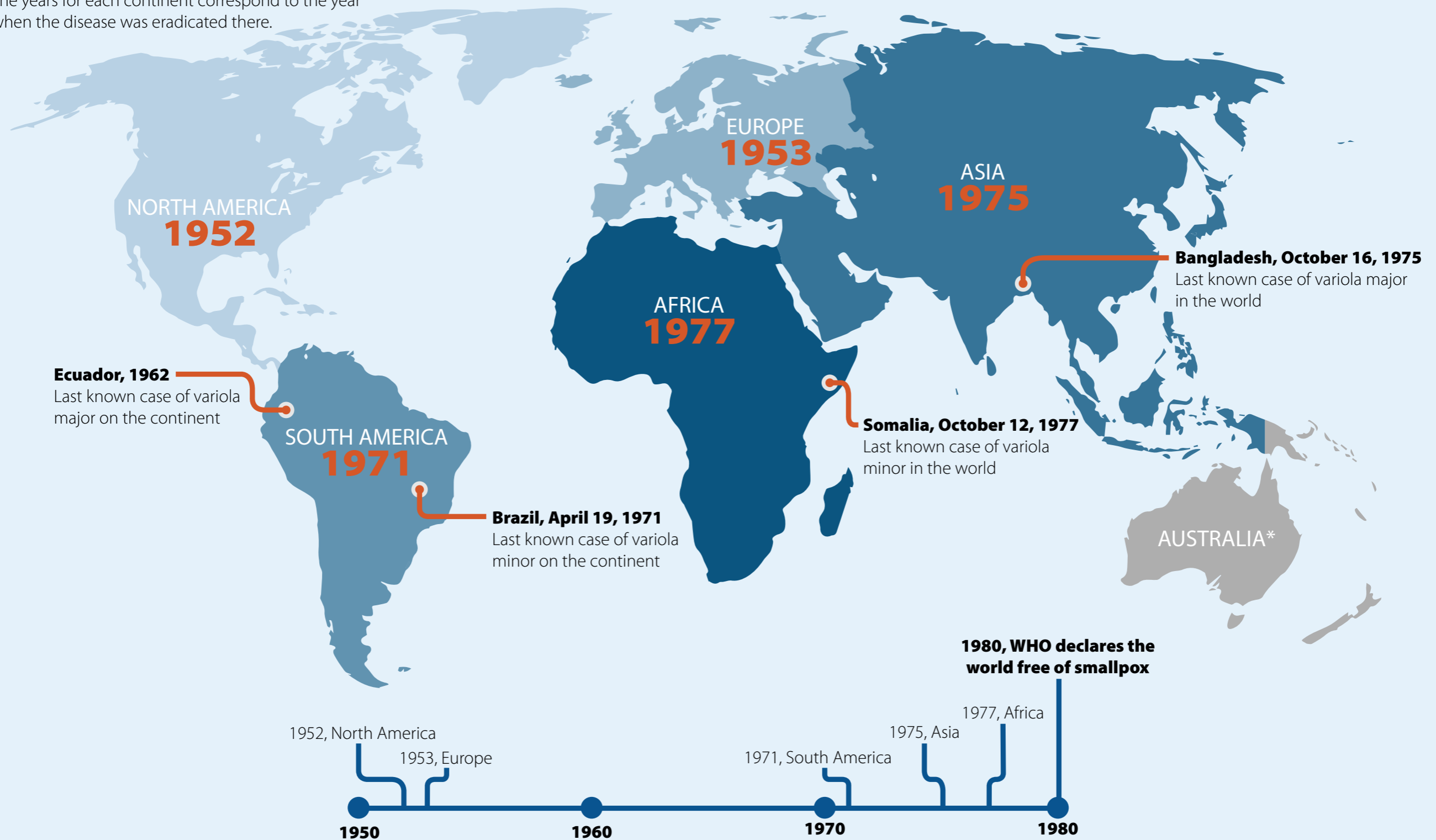
Variola virus

186kb DNA virus

GLOBAL SMALLPOX ERADICATION

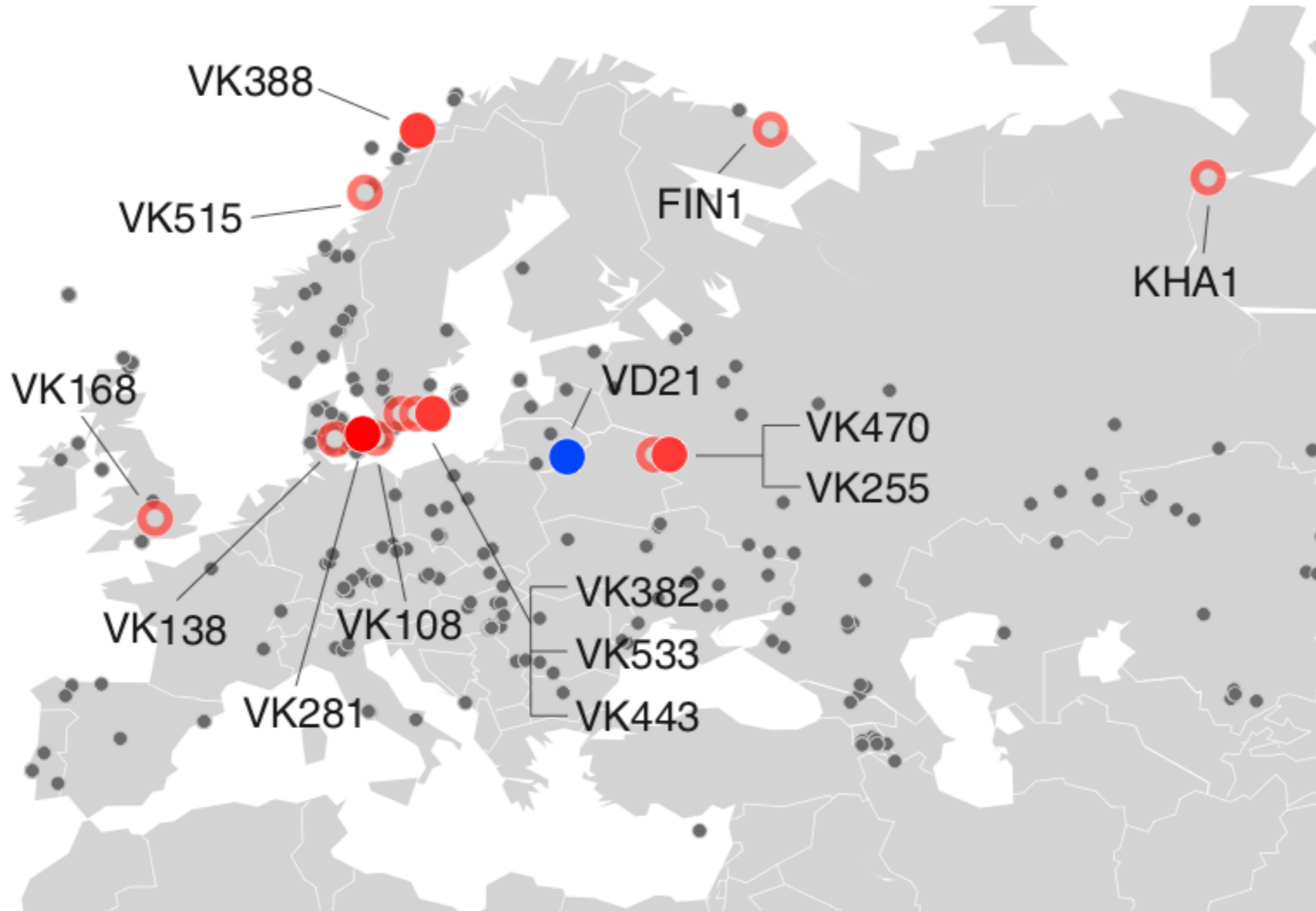
The historically important dates highlighted in the map show countries in which the last naturally acquired cases of smallpox occurred.

The years for each continent correspond to the year when the disease was eradicated there.



* Smallpox was never endemic (widespread) in Australia

The Viking Age smallpox



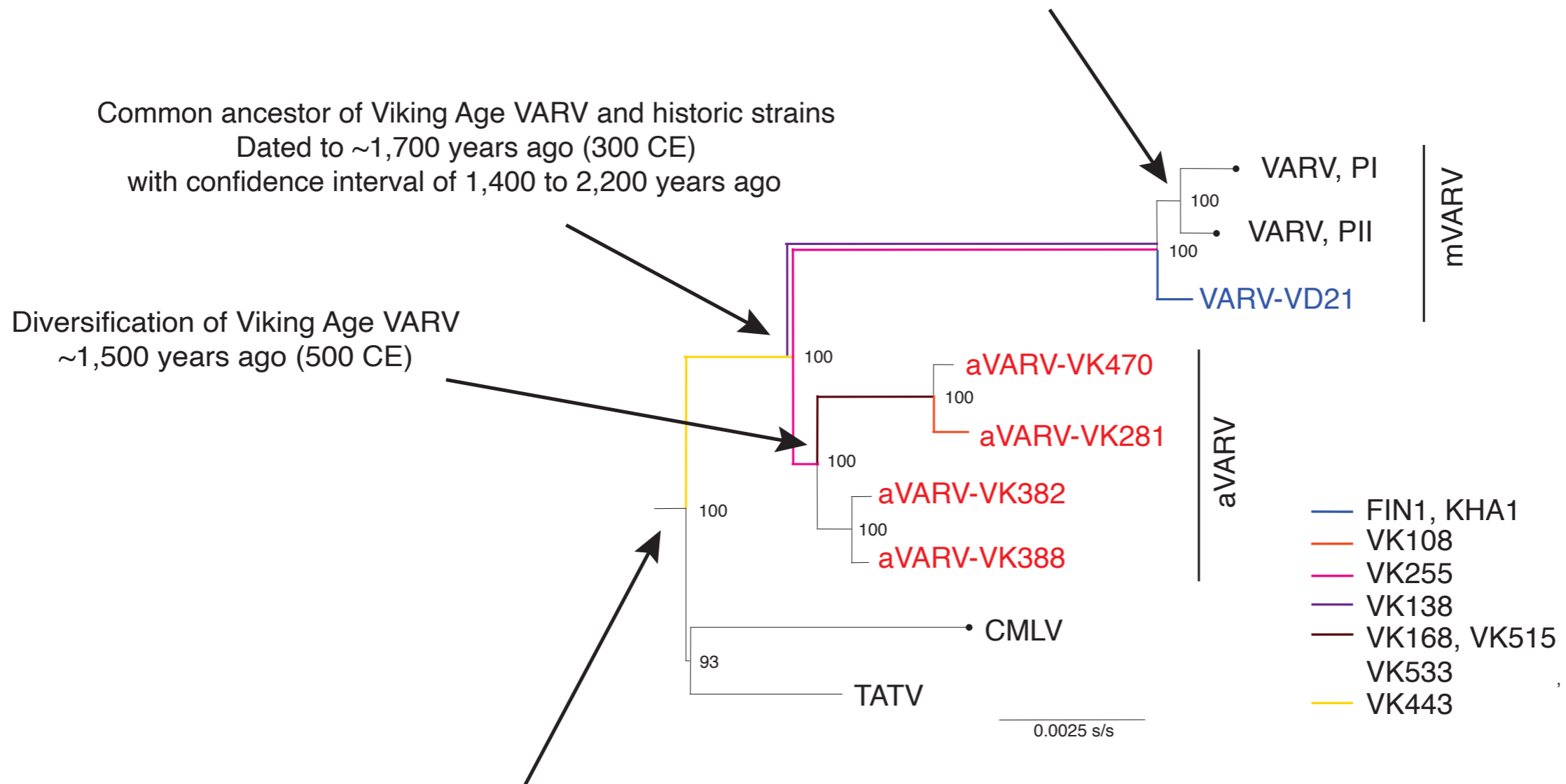
Recovery of 13 ancient variola virus genomes (0.01X - 45X)
11 from Viking Age (500 CE - 1100 CE), 2 from 19th century

A recent origin for smallpox

Recent diversification of historical VARV primary clades
Dated to ~300 years ago (1700 CE)

Common ancestor of Viking Age VARV and historic strains
Dated to ~1,700 years ago (300 CE)
with confidence interval of 1,400 to 2,200 years ago

Diversification of Viking Age VARV
~1,500 years ago (500 CE)



Common ancestor of human VARV and animal pox viruses (Camelpox, Taterapox)
No molecular dating in our analysis, but likely not older than 3,000 - 4,000 years



Interested in a project?
martin.sikora@sund.ku.dk

