

Ancient DNA and the origin of horse domestication

Popgen KU 2020

Genetic drift based analyses

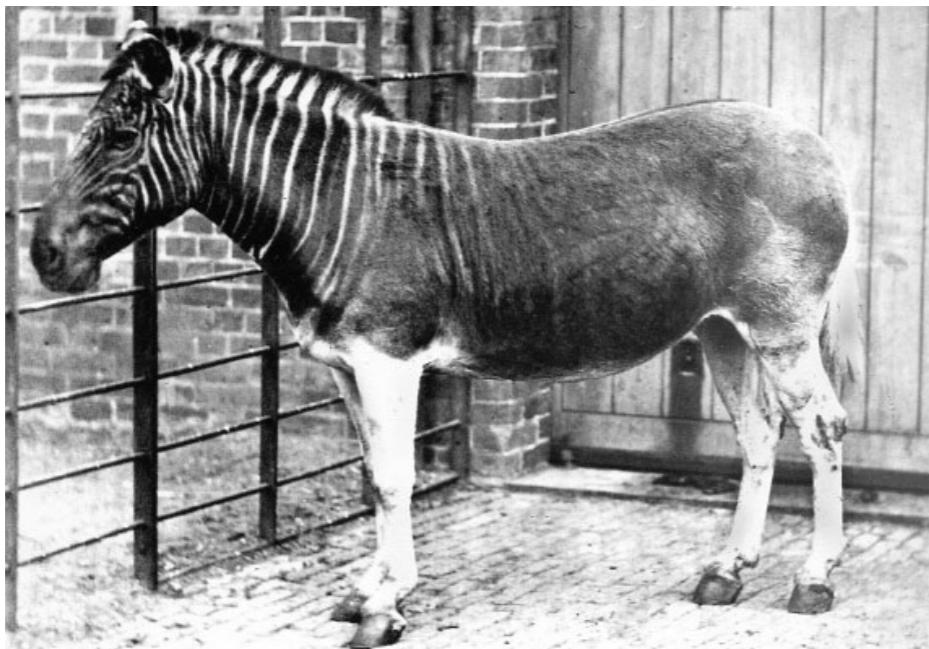
Program

- Introduction to ancient DNA (~30 min)
- Genetic drift (~15 min)
- Exercise 1 (~20 min)
- Ancient DNA and horse domestication (~20 min)
- Exercise 2 (~40 min)
- **Before we start:**

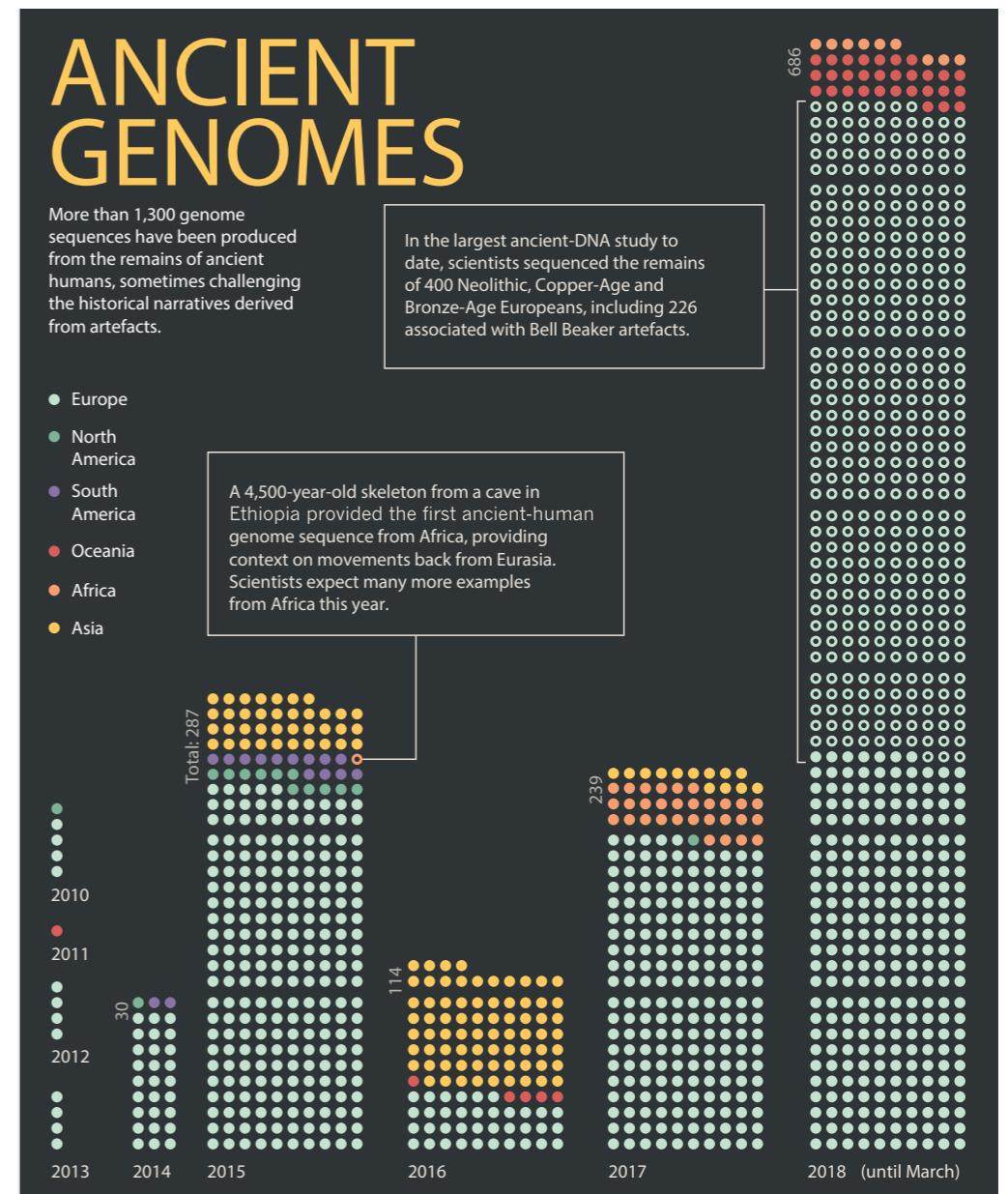
<https://github.com/KHangoj/popgenteach/blob/master/installdriftr.md>

Today Thousands of ancient samples characterized

229 bp mtDNA
in 1984



Whole Genome (7.9X)
in 2014



Ancient DNA in the 90s



65 million years old DNA

20 million years old DNA



Ancient DNA in the 90s



Contamination from human and/or food sources s old DNA

65 million years old DNA

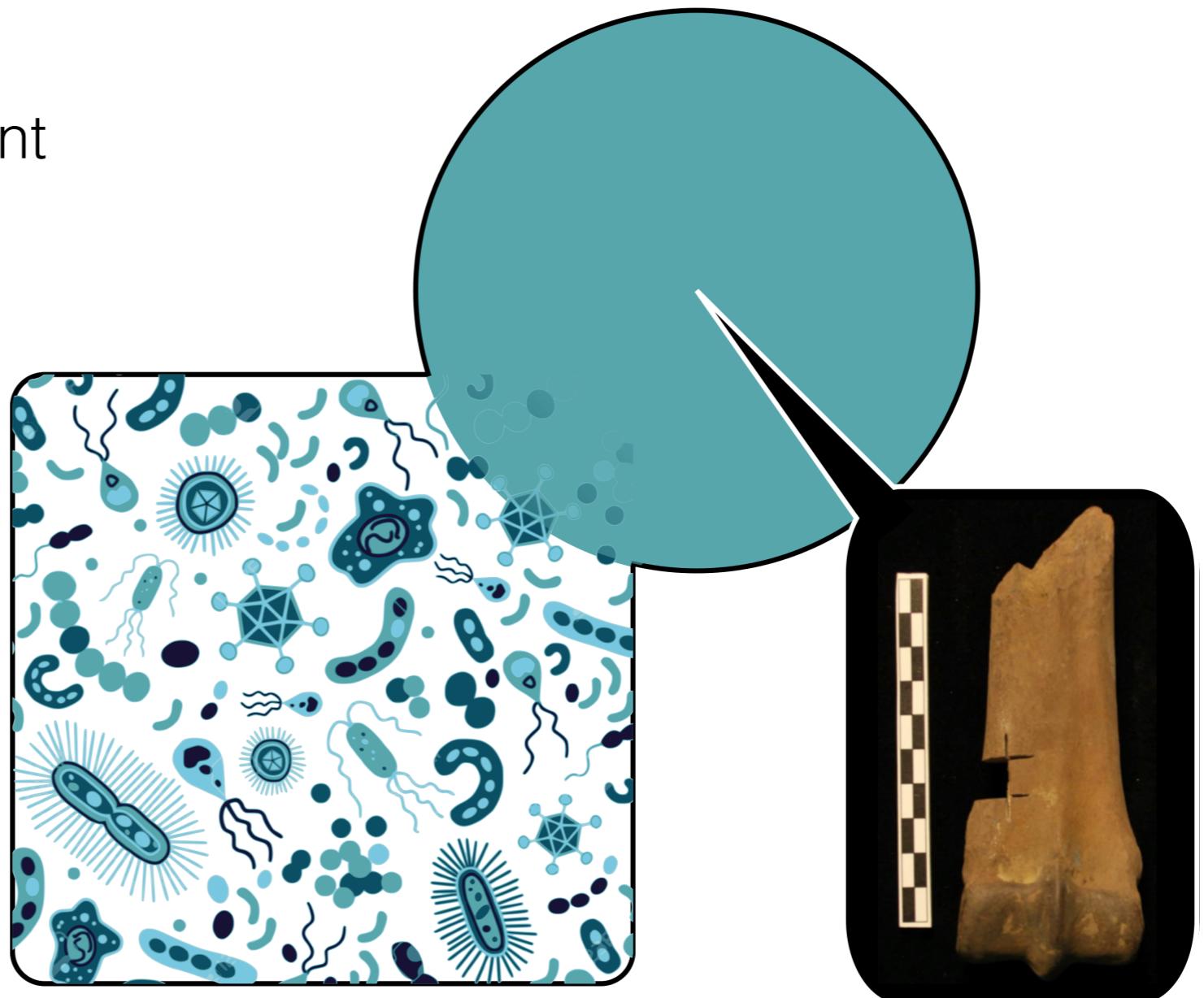


What does ancient DNA look like

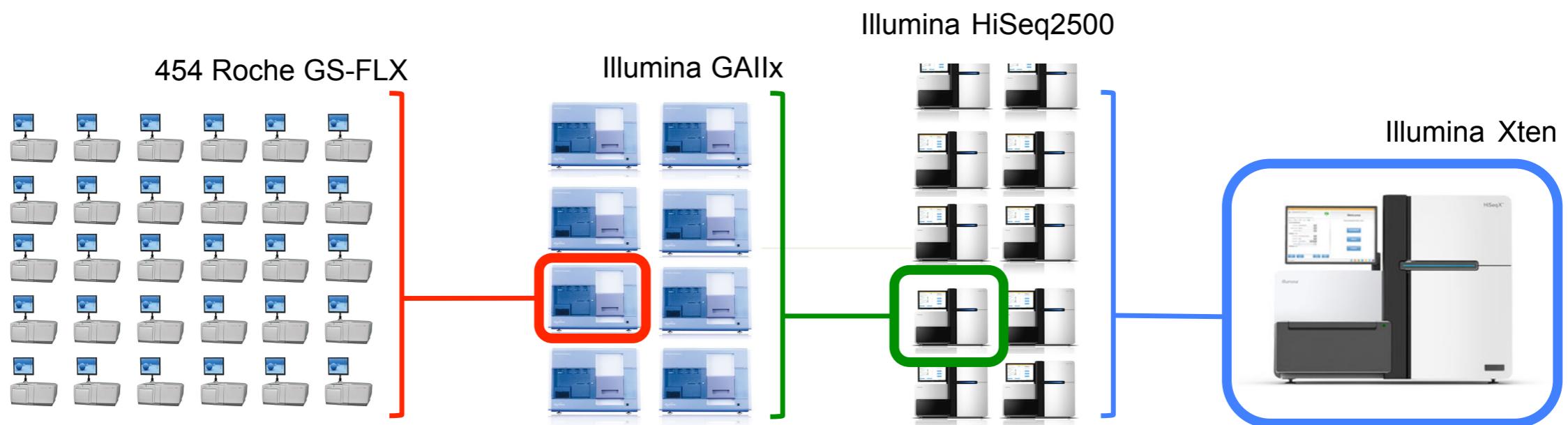
**DNA contents of
ancient samples**

Issues:

1. Low Endogenous Content
2. Short DNA molecules
3. DNA damage



How is it still possible?



Output

0.5M

250M

4G

30G

Runtime

0.5d

5d

5d

3d

Application of ancient DNA

Denisovan Cave



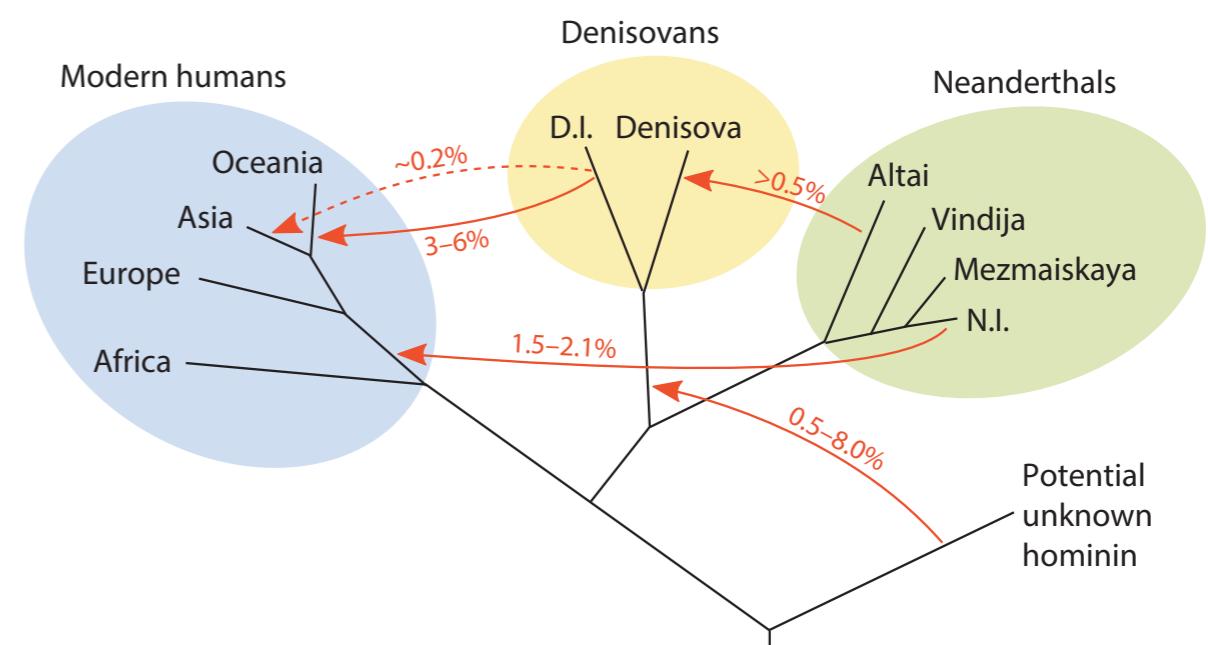
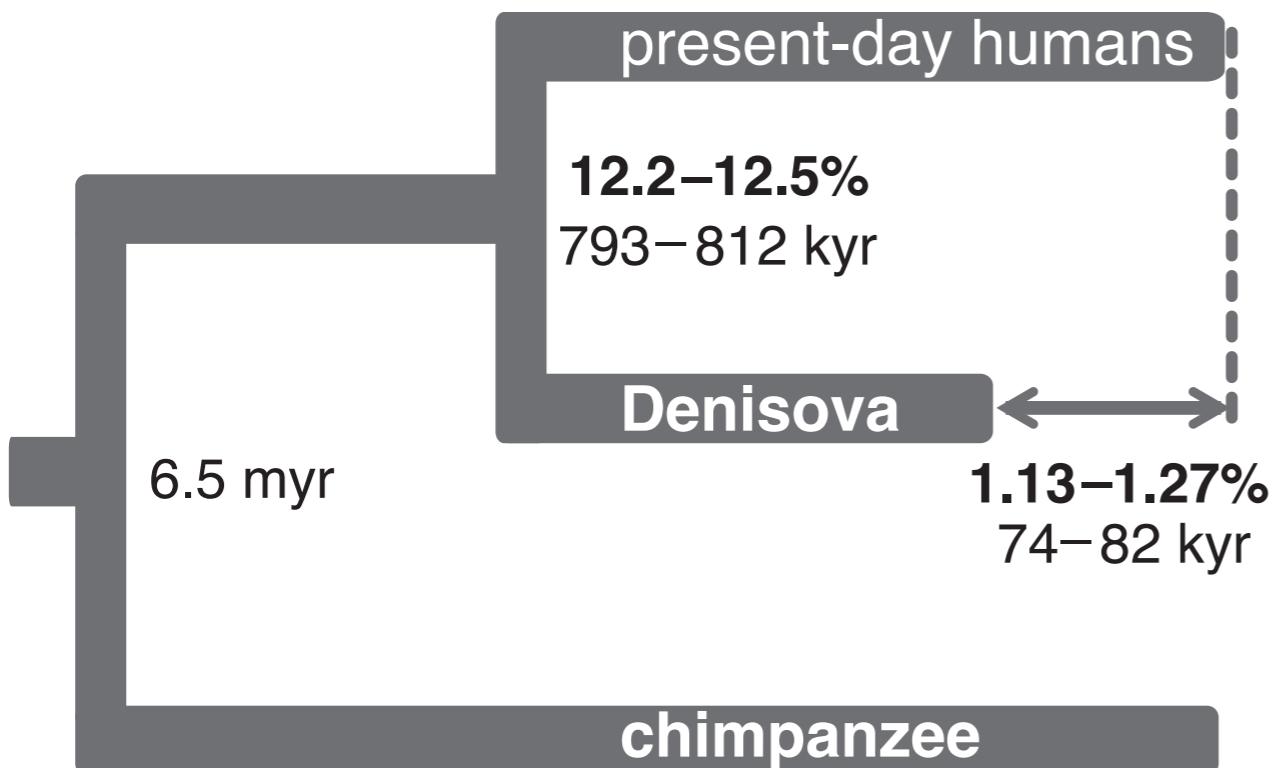
Finger bone



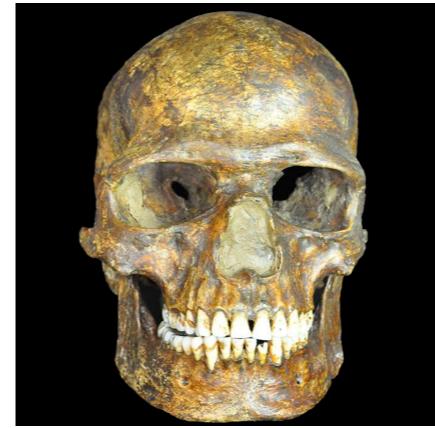
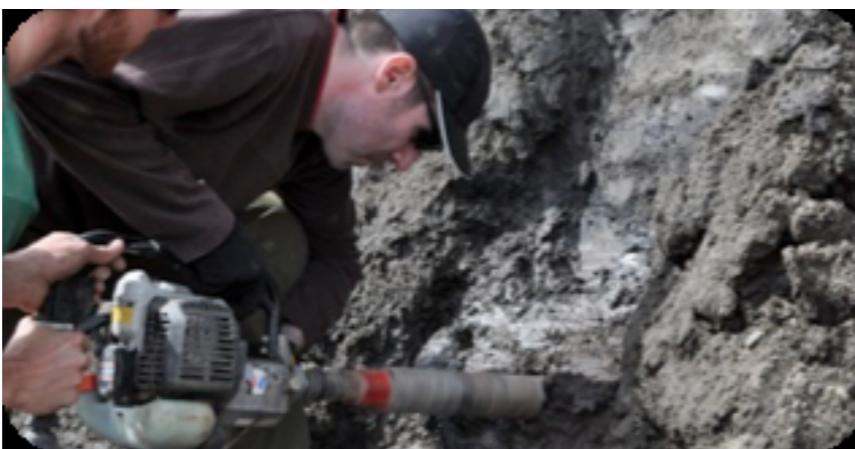
Altai mountains - Siberia

Discovery of a new lineage

Unknown extinct relative to Neanderthals sequenced to 30X



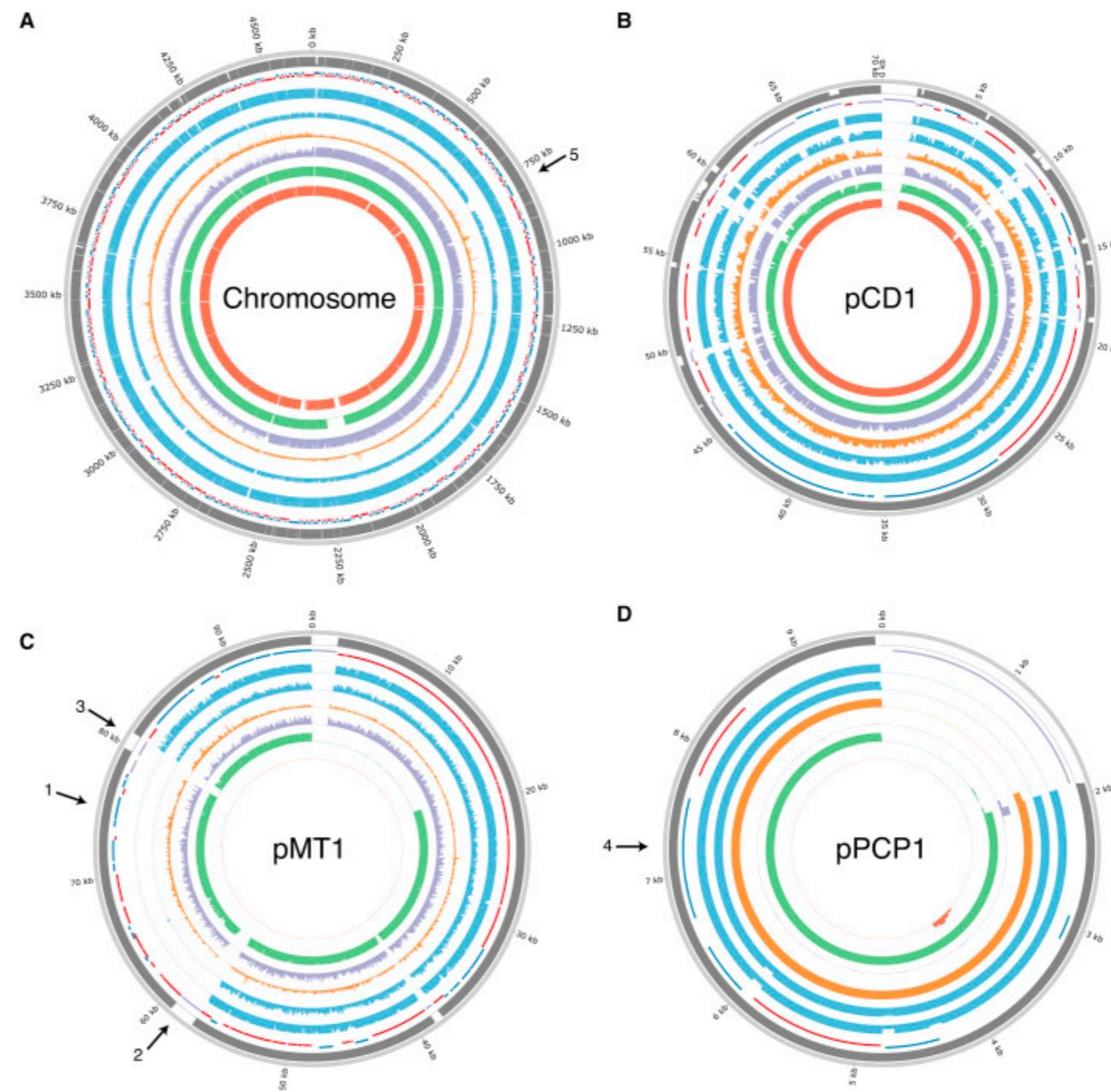
What else is ancient DNA?



Ancient DNA

Ancient pathogens: 5,000 years old *Yersinia pestis*

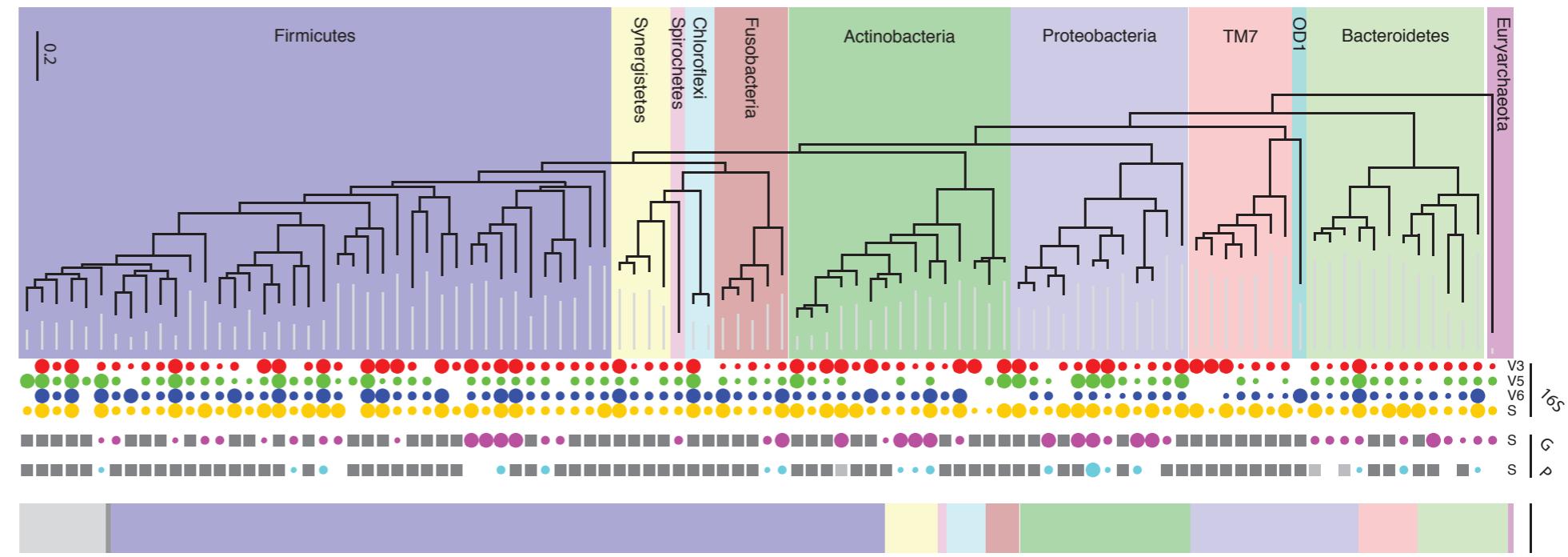
Human remains



Ancient DNA

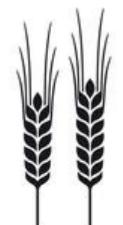
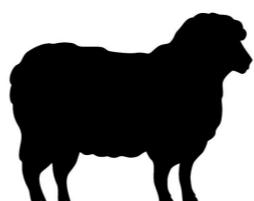
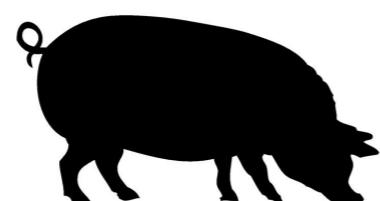
Oral microbial diversity

Dental plaque



Warinner 2014

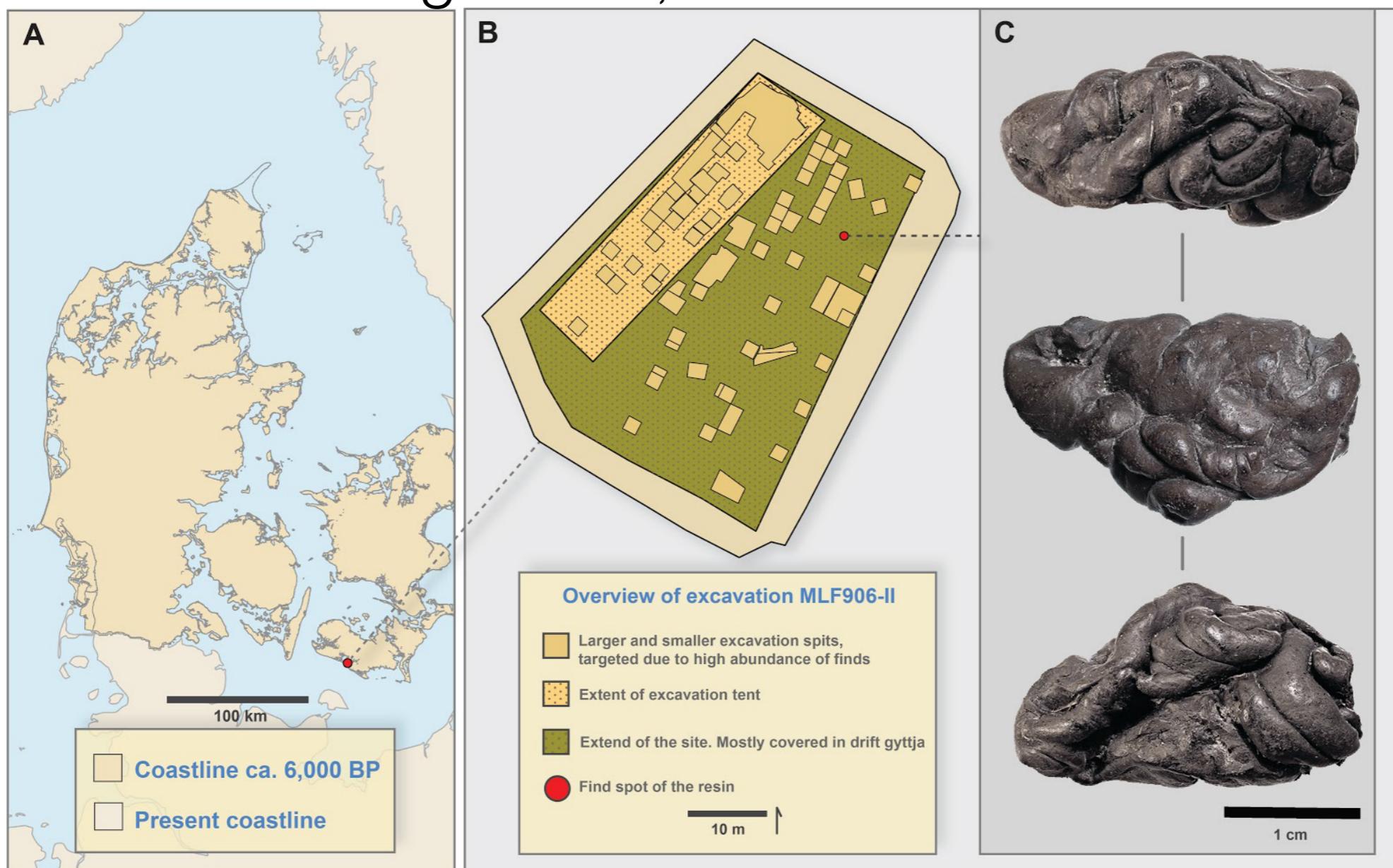
Diet



Ancient DNA

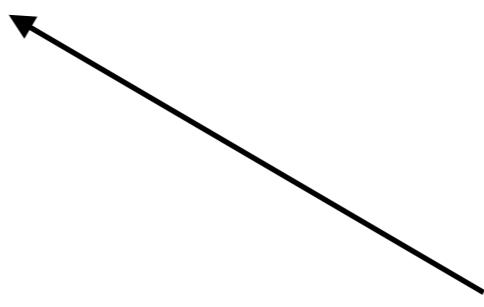
5.700 years old “chewing gum”

Whole genome; mouth microbiome



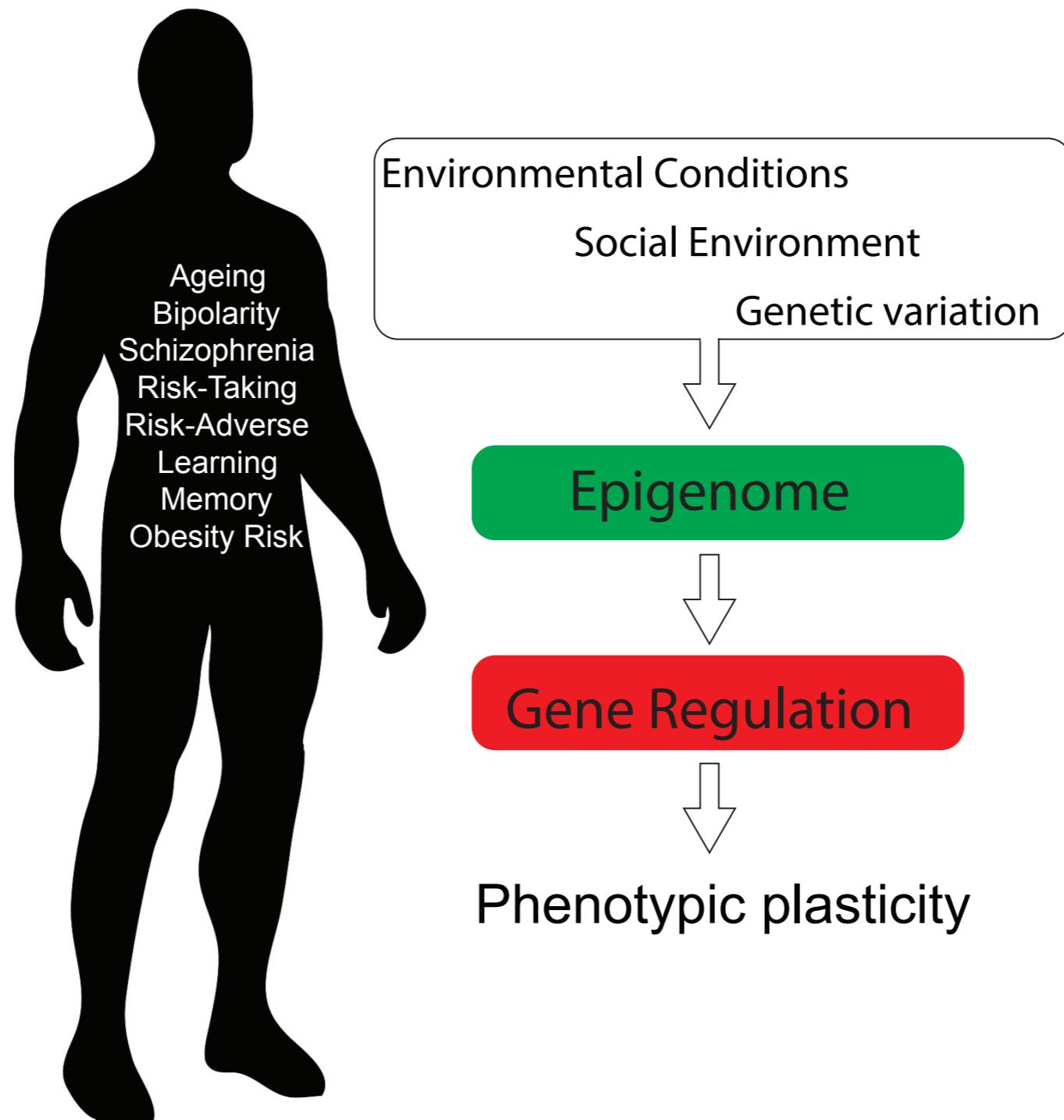
My background in ancient DNA

1. Methylation signatures from ancient DNA sequencing data
2. **Early horse domestication** and signatures of artificial selection

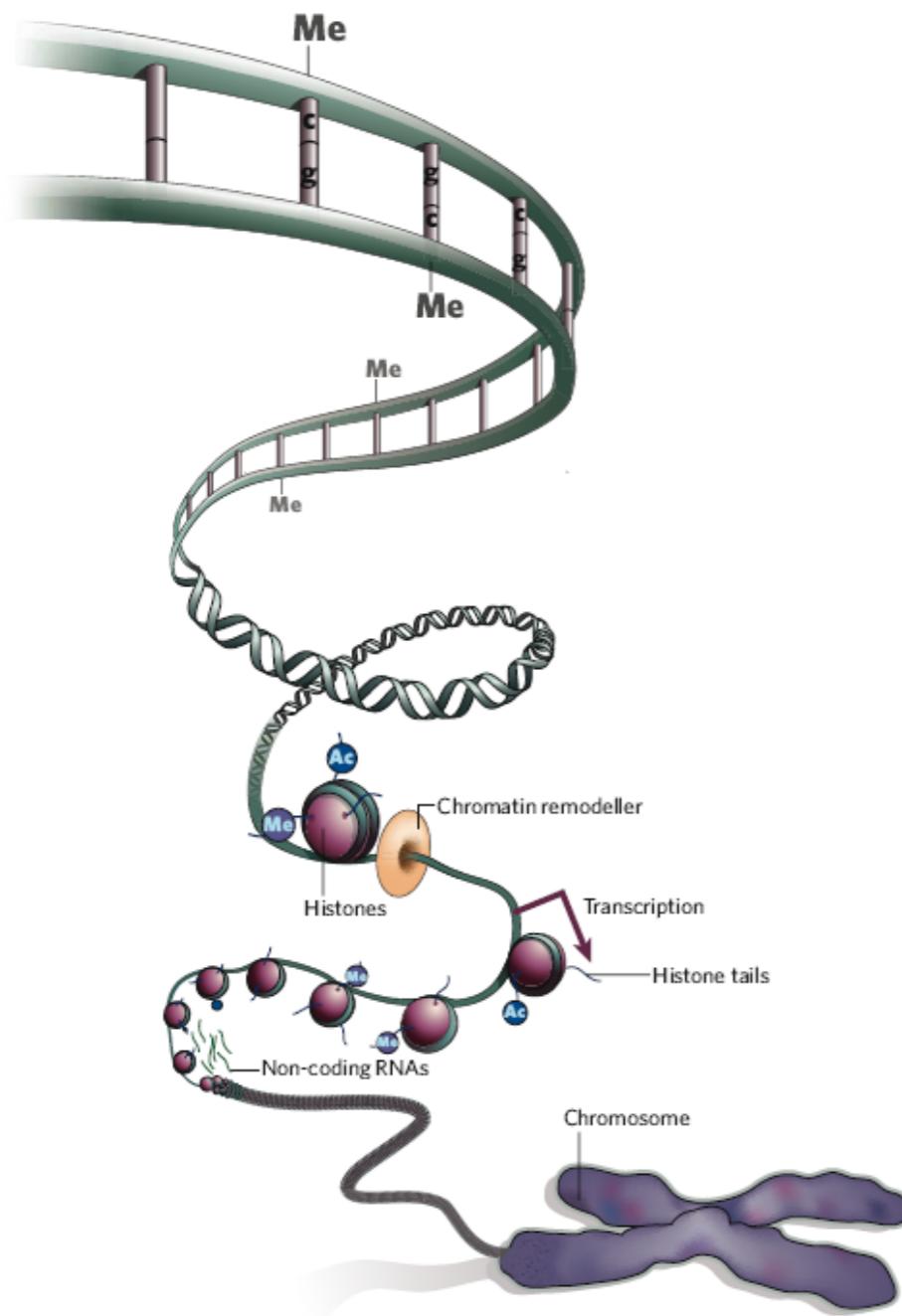


Exercise 2

Humans are much more than genomes



Methylation in mammals



Bisulphite sequencing



Bisulfite conversion

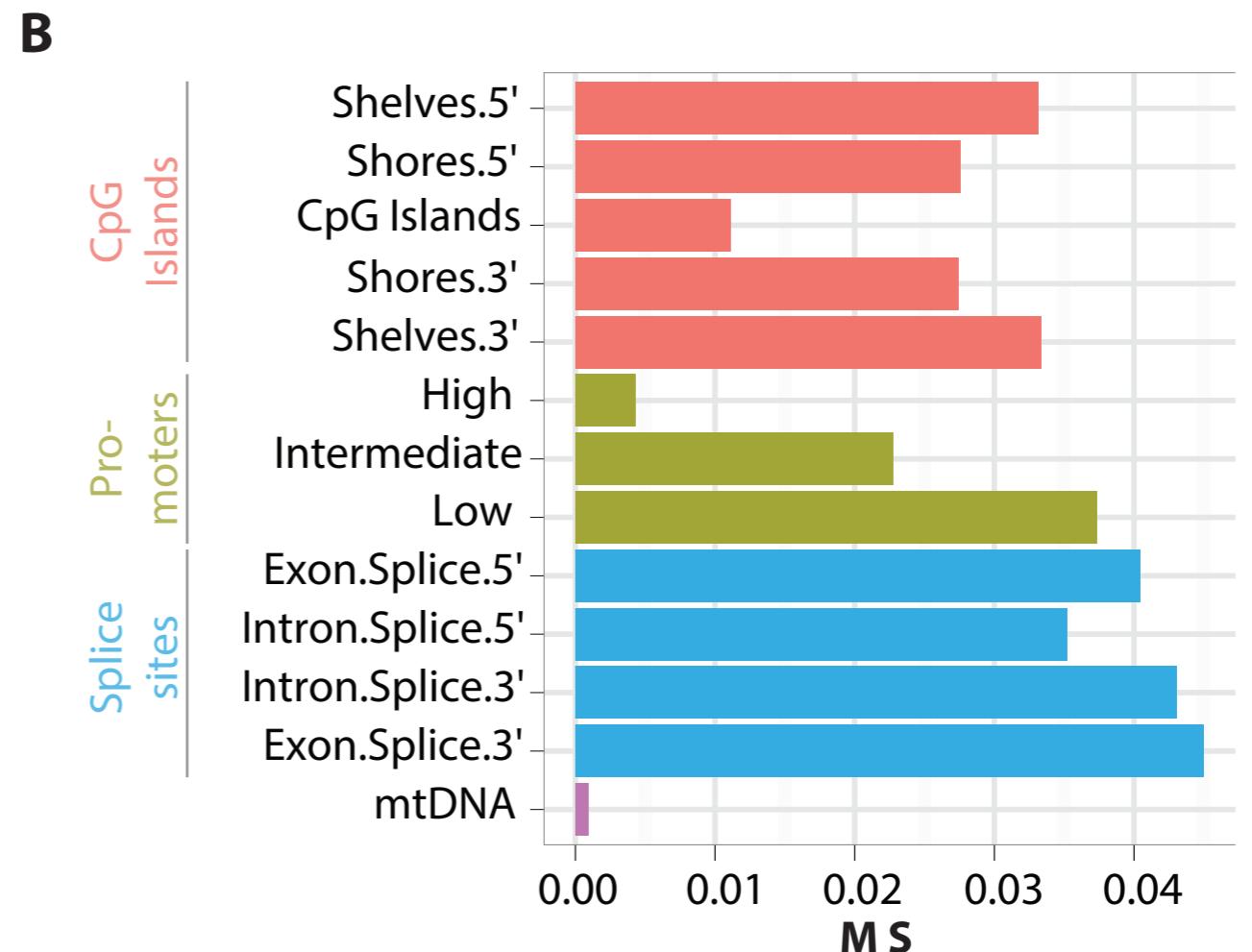
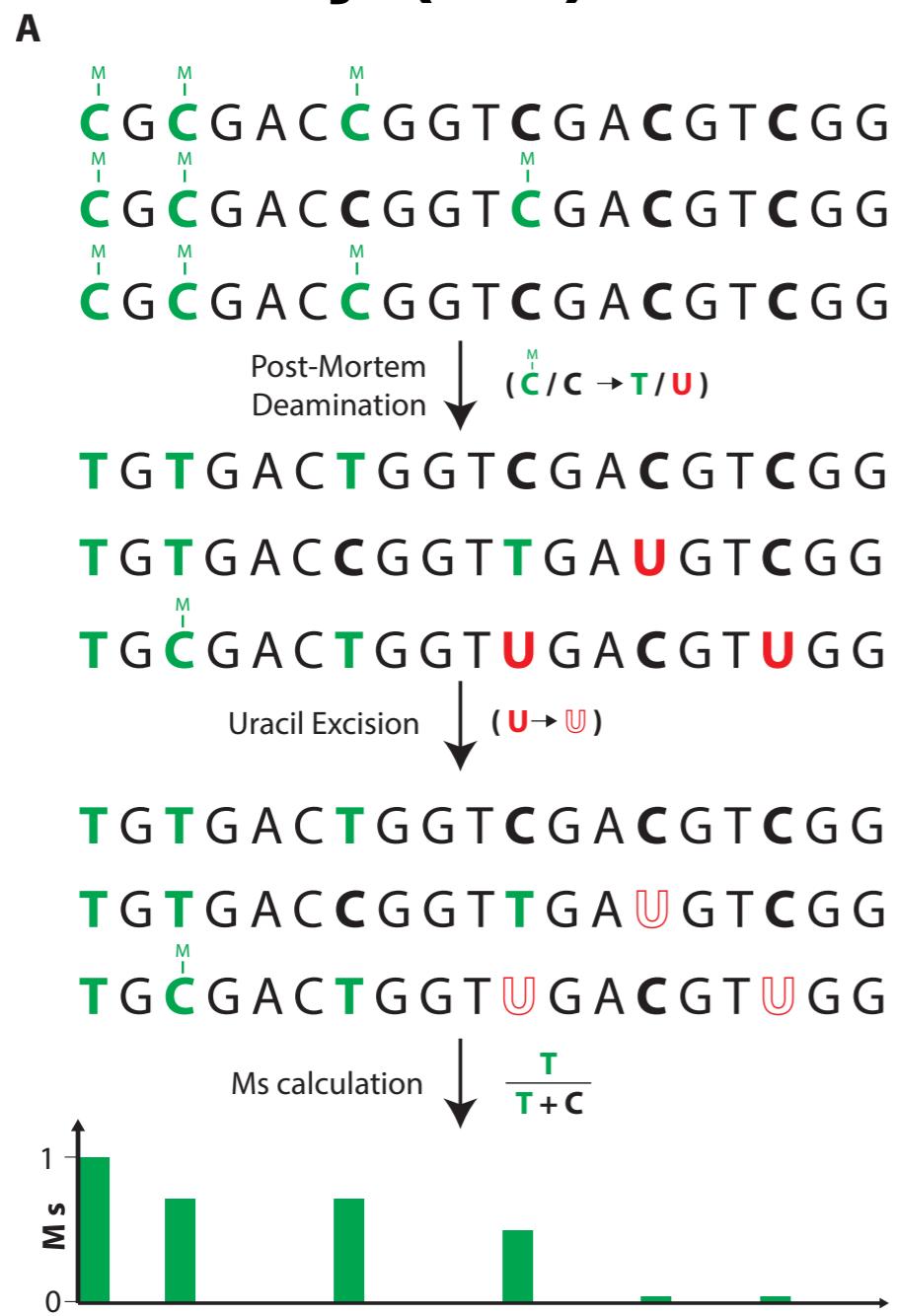


PCR Amplification

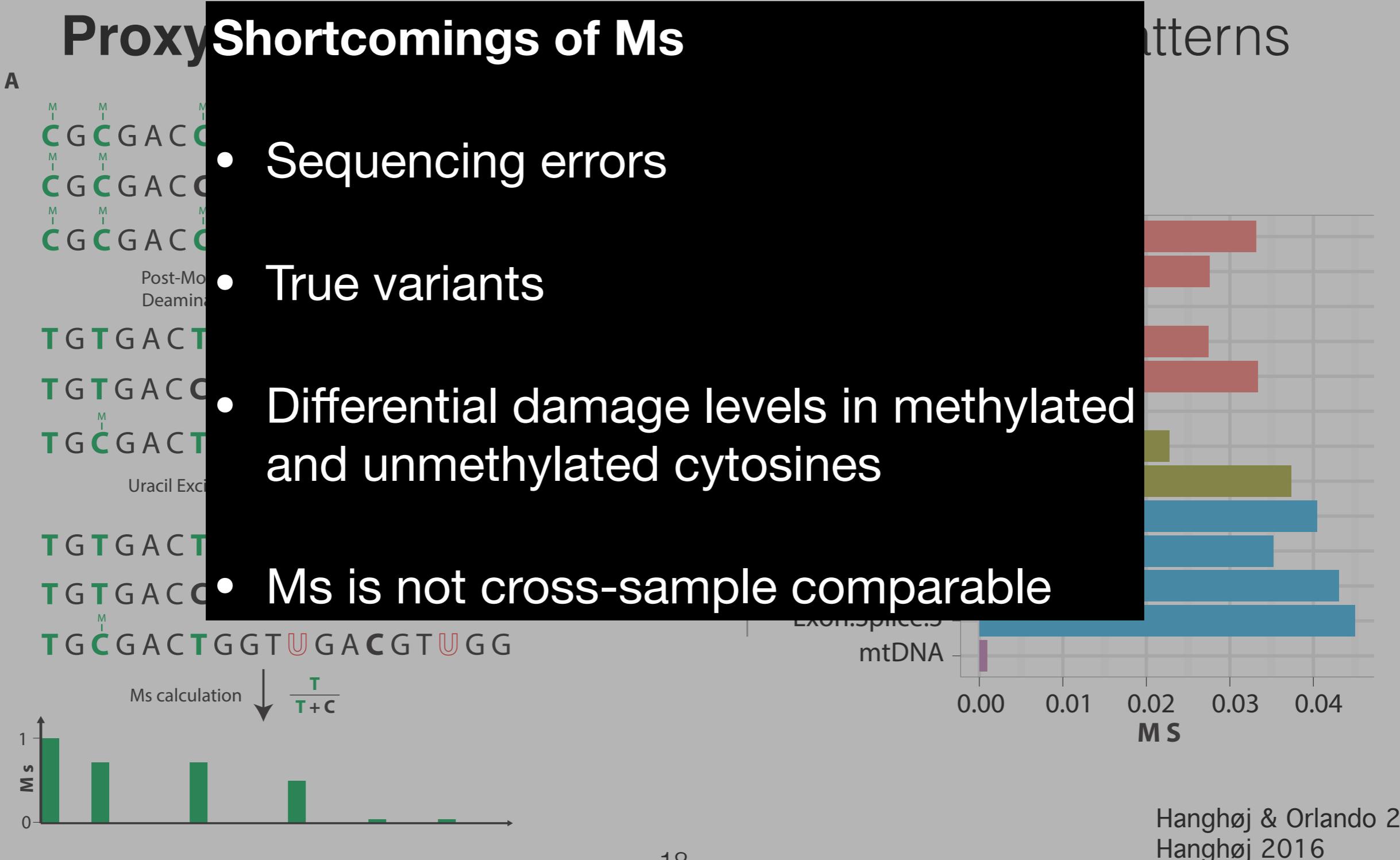


Methylation maps

Proxy (Ms) follows expected methylation patterns



Methylation maps



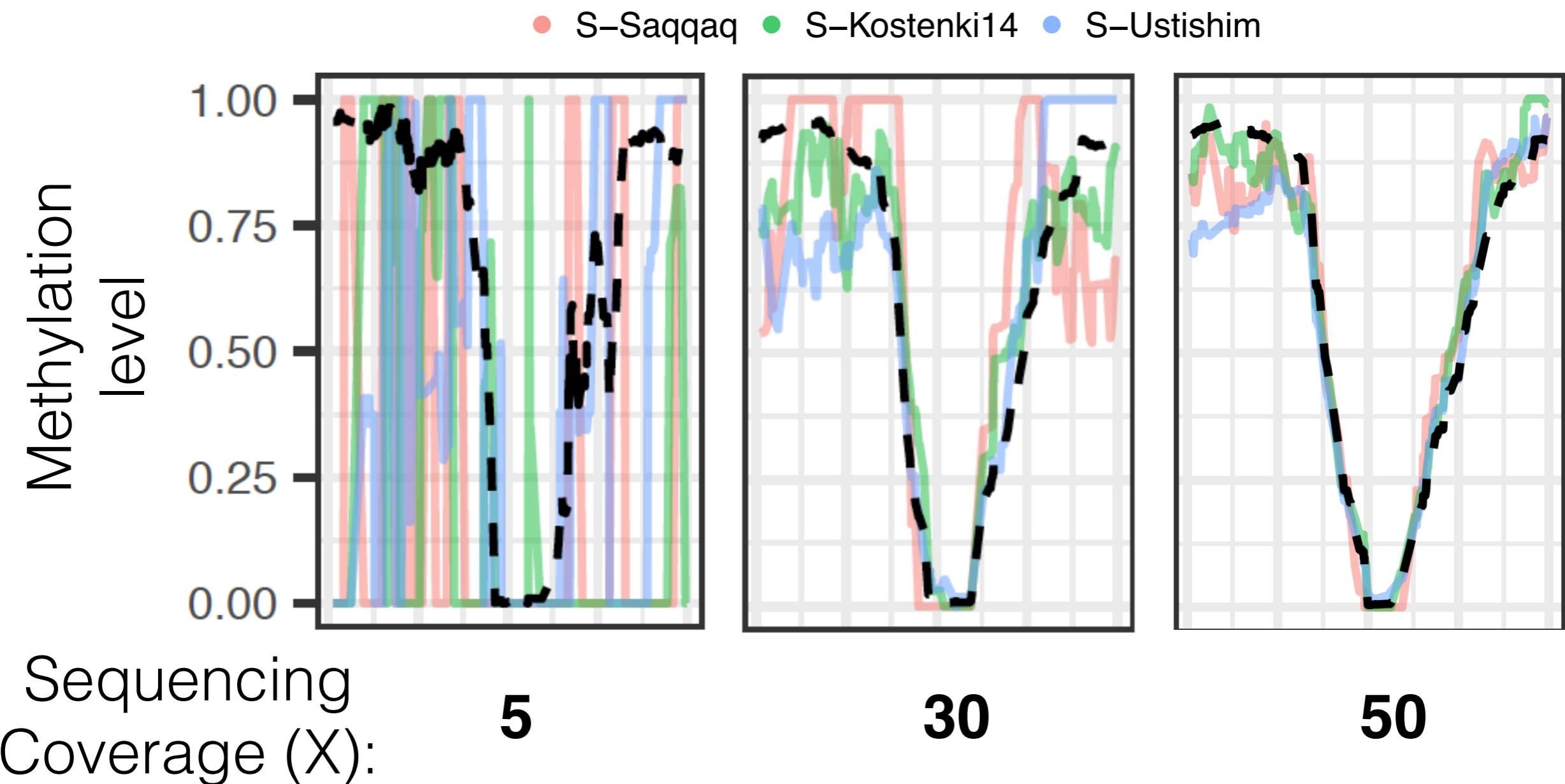
DamMet

Full probabilistic model to estimate methylation levels from ancient sequencing data

Step 1: $L(D|\mathbb{D}) = \prod_{j=1}^{\text{sites}} \prod_{i=1}^{\text{depth}} p(X_{j,i,k,v} | D_{M,k,v}, Q_{j,i}, \epsilon_{j,i}, F_{global})$

Step 2: $L(f|\mathbb{D}) = \prod_s \sum_{G \in (0\dots6)} p(G=g) p(X|s, f, G=g, D, \theta)$

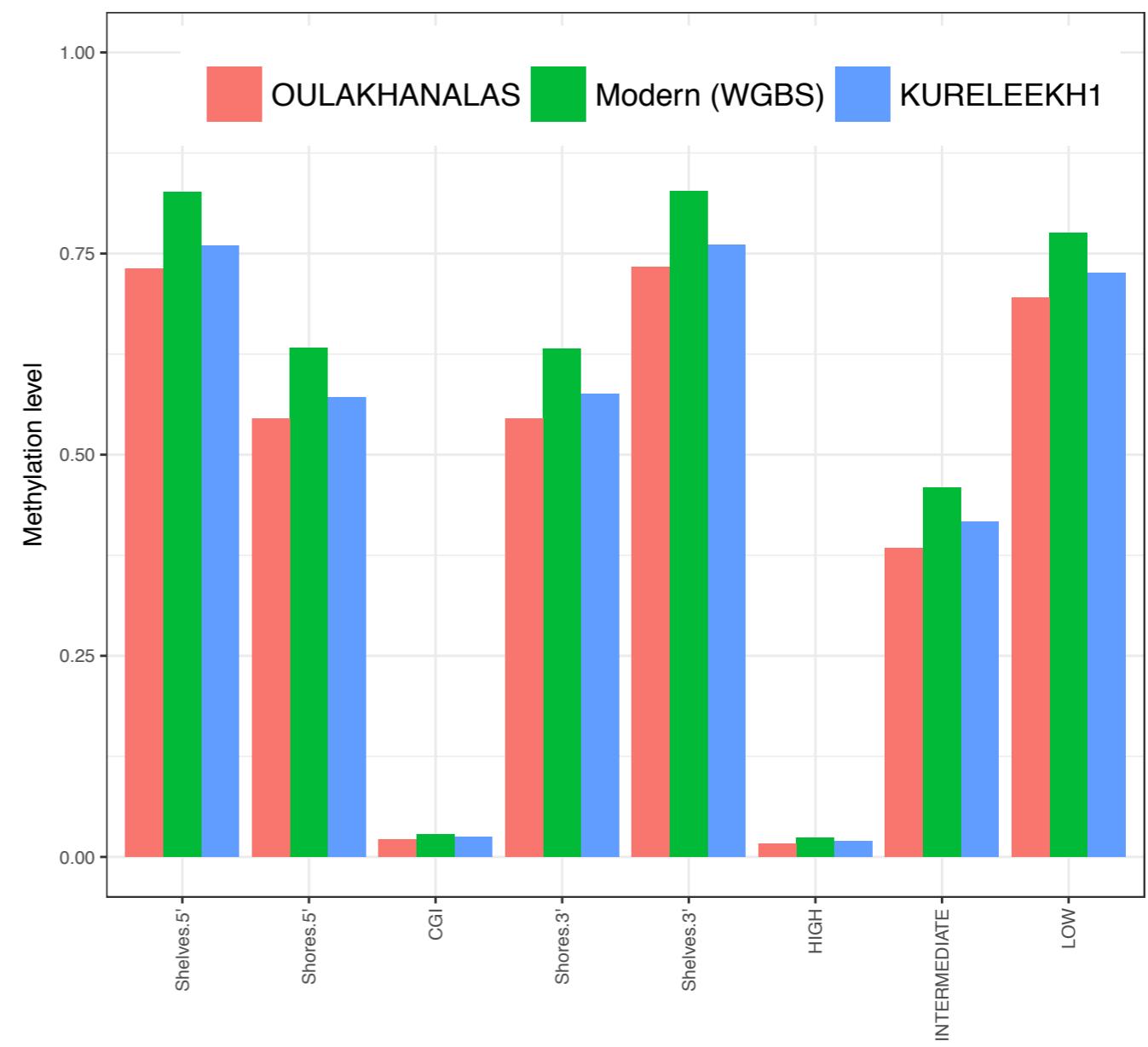
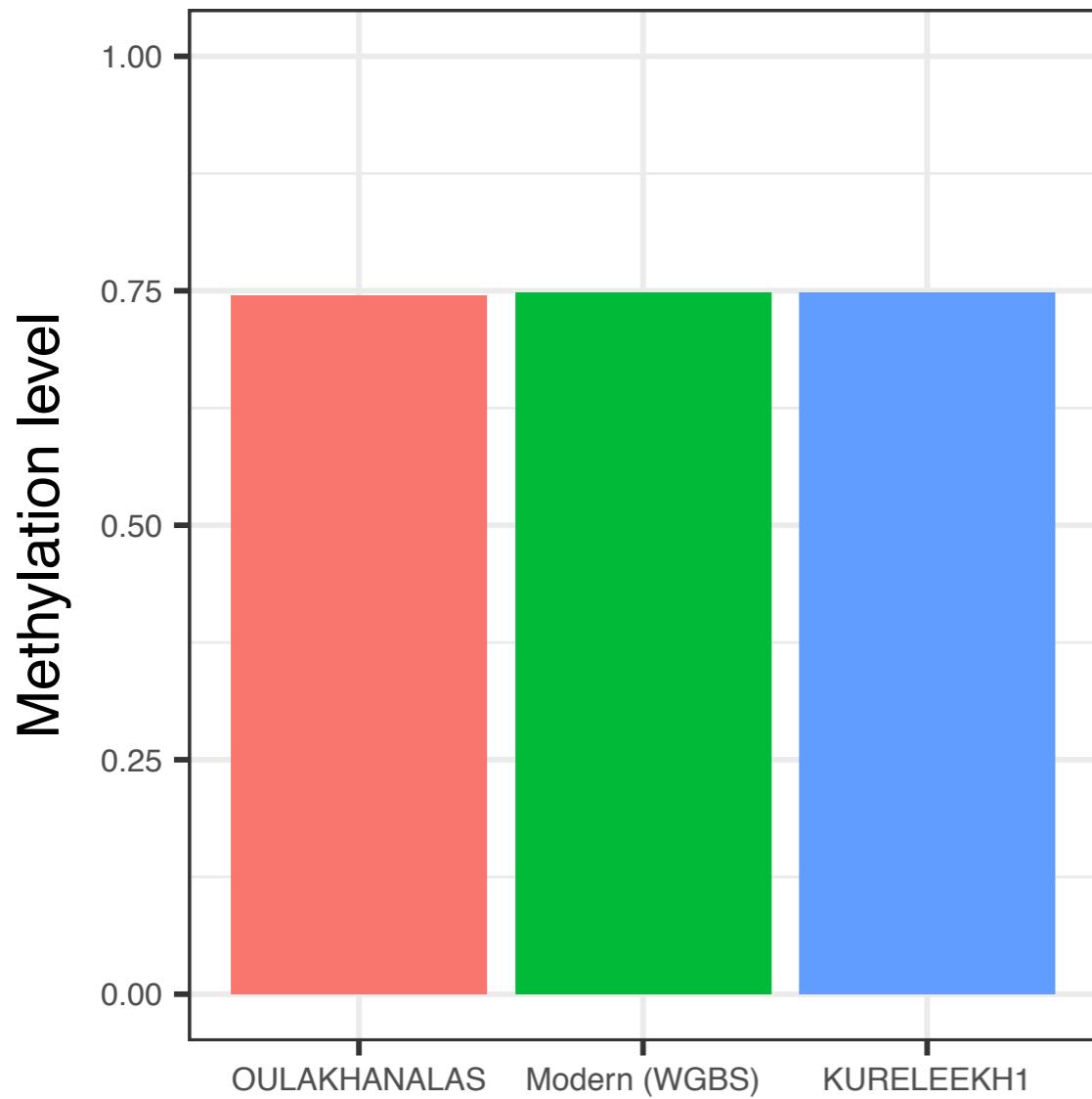
Estimate of methylation on simulated data



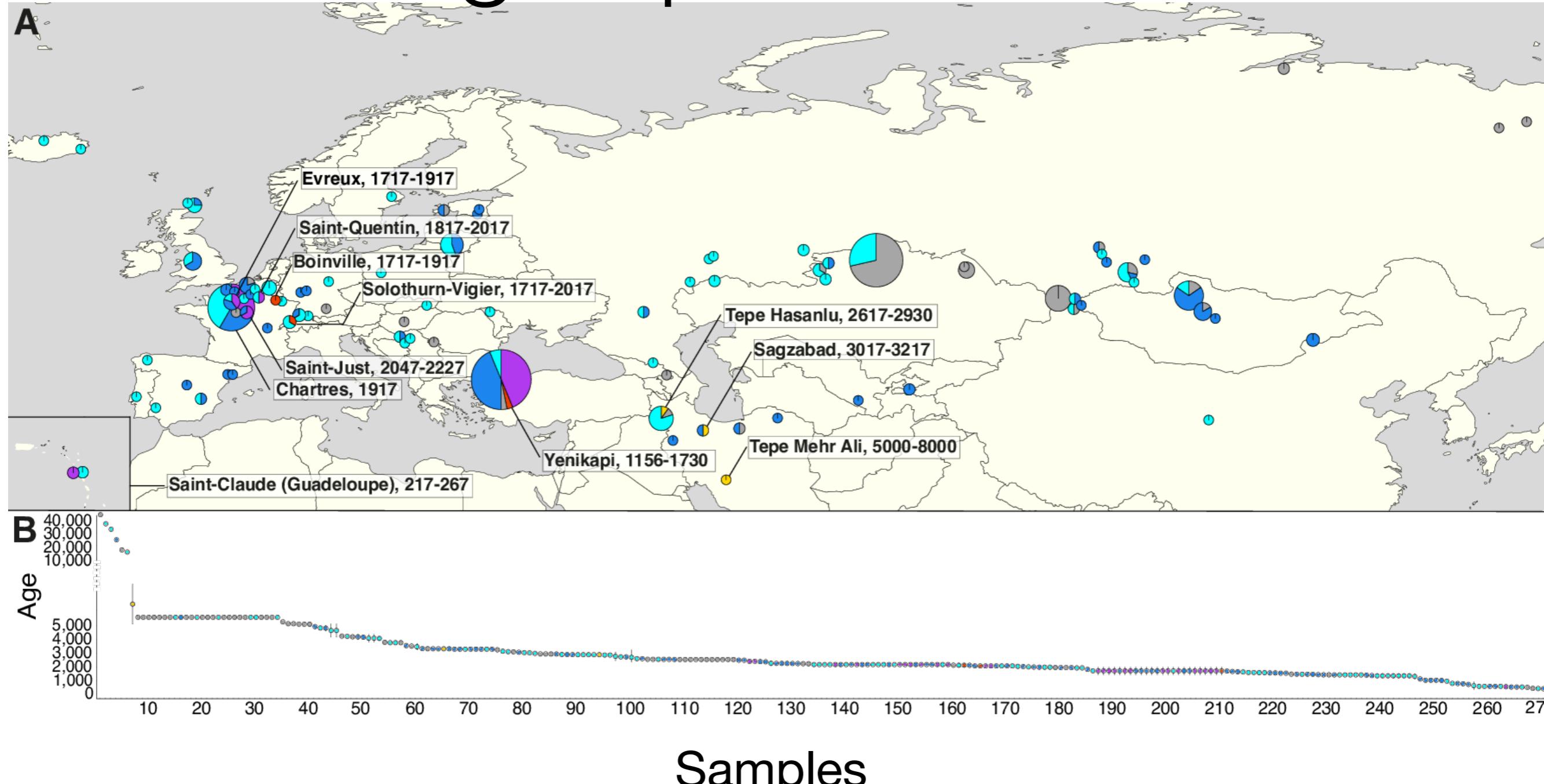
Oulalakhanalas (25X) & Kureleekh1 (44X)

Chromosome-wide

Regional



Domesticated Horses through space and time

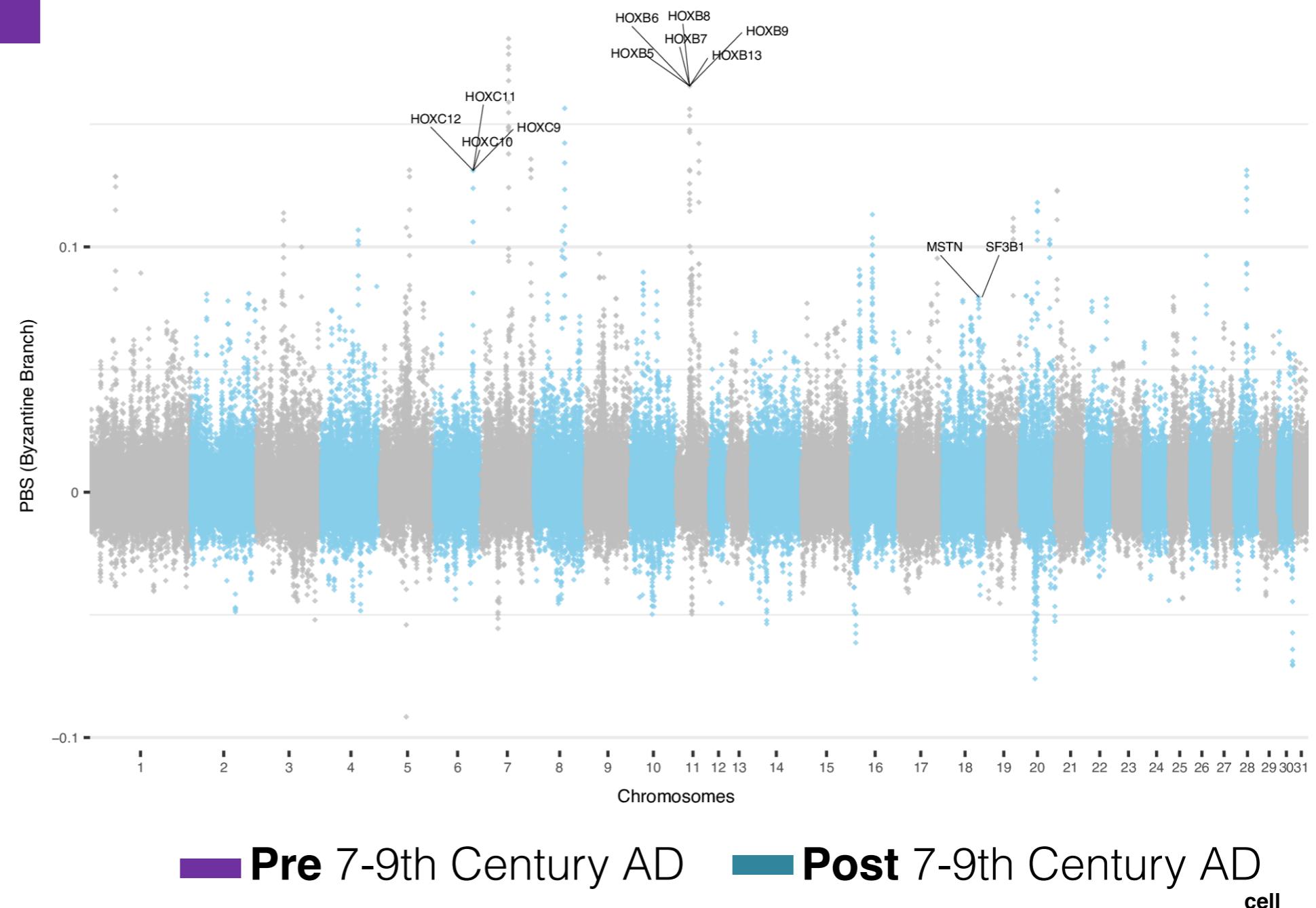
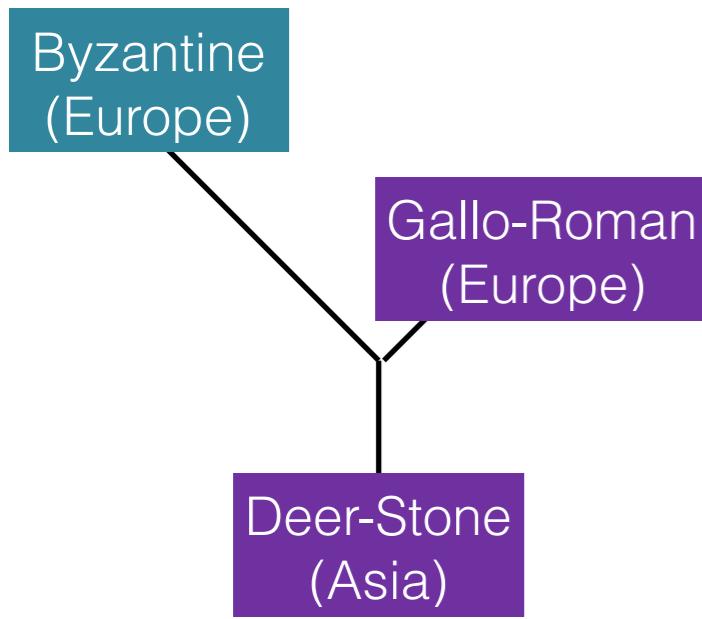
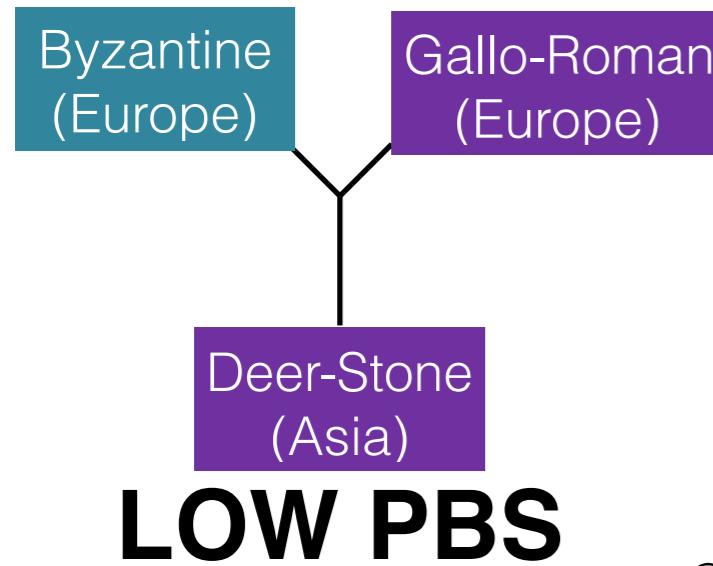


Samples

cell

Culture-Specific Selection

Population Branch Statistics (PBS)



Genetic drift - Wright Fisher

- Assuming the Wright-Fisher model (WFM), **genetic drift** solely changes the allele frequency in a population.
 - The WFM assumes absence of selection, migration, mutation.

- Genetic drift is the **change** in allele frequencies over time due to **random sampling** of alleles from the previous generation:

$$P_{ij} = \binom{2N}{j} \left(\frac{i}{2N}\right)^j \left(1 - \frac{i}{2N}\right)^{2N-j}$$

- All alleles are equally likely to be sampled.
- These changes are **population specific**.

Mini exercise

Probability of going from ten to eight alternative alleles in a population of 10 diploid individuals in a single generation?

$$\binom{2N}{j} p^j (1-p)^{2N-j} \quad p = \frac{i}{2N}$$

$$N = 10 \quad i = 10 \quad j = 8$$

Hint: use dbinom in R

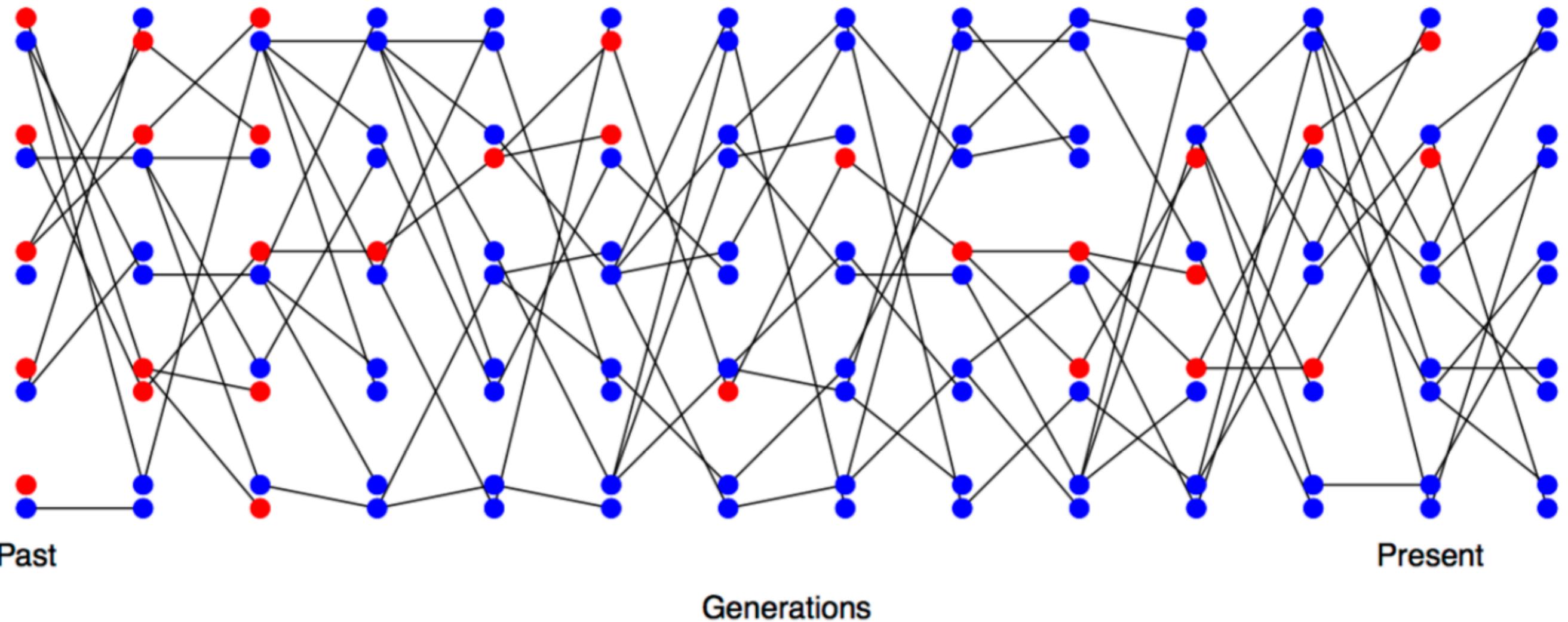
Genetic drift - reality

Many sites in the genome evolve in a (somewhat) neutral manner

- Changes in non-coding and non-functional DNA (~90% of genome; Rands 2014)
- Synonymous change in coding DNA
- Non-synonymous that don't have strong effect on functional properties
- Non-synonymous changes with phenotypic consequences but no effect on fitness

Example of genetic drift

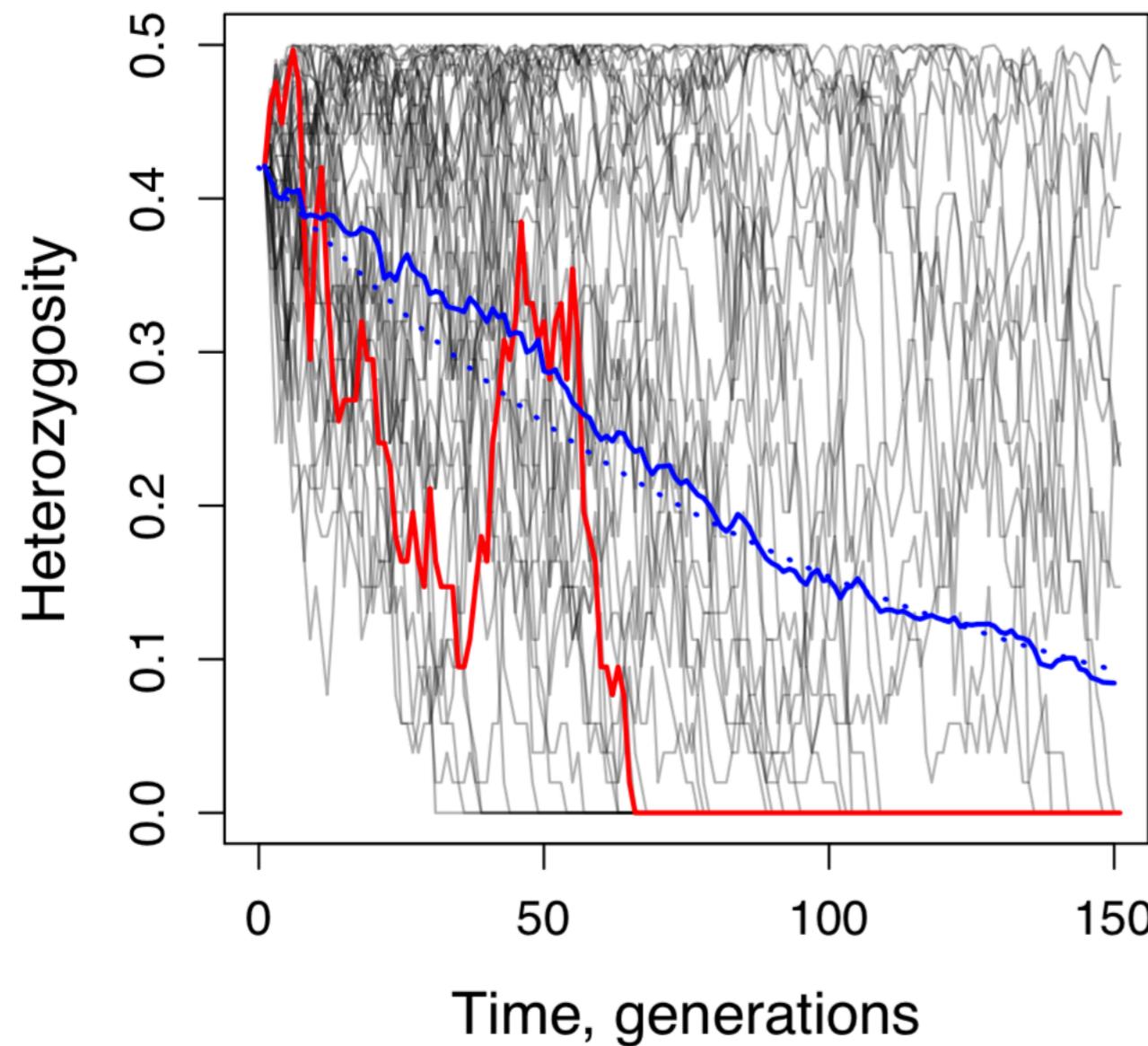
Loss of Heterozygosity



Assuming absence of mutation and migration

Example of genetic drift

Loss of Heterozygosity



$$H_t = \left(1 - \frac{1}{2N}\right)^t H_0$$

$$H_0 = 0.42 \quad N = 50$$

$$H_{150} = ?$$

$$H_{300} = ?$$

Assuming absence of mutation and migration

Exercise 1 - driftR

driftR: Population Genetic Simulations in R

Run Simulation

Starting allele frequency A
0.5

Mutation Rate
0 0.1

Fitness of genotype AA
0 1

Fitness of genotype AB
0 1

Fitness of genotype BB
0 1

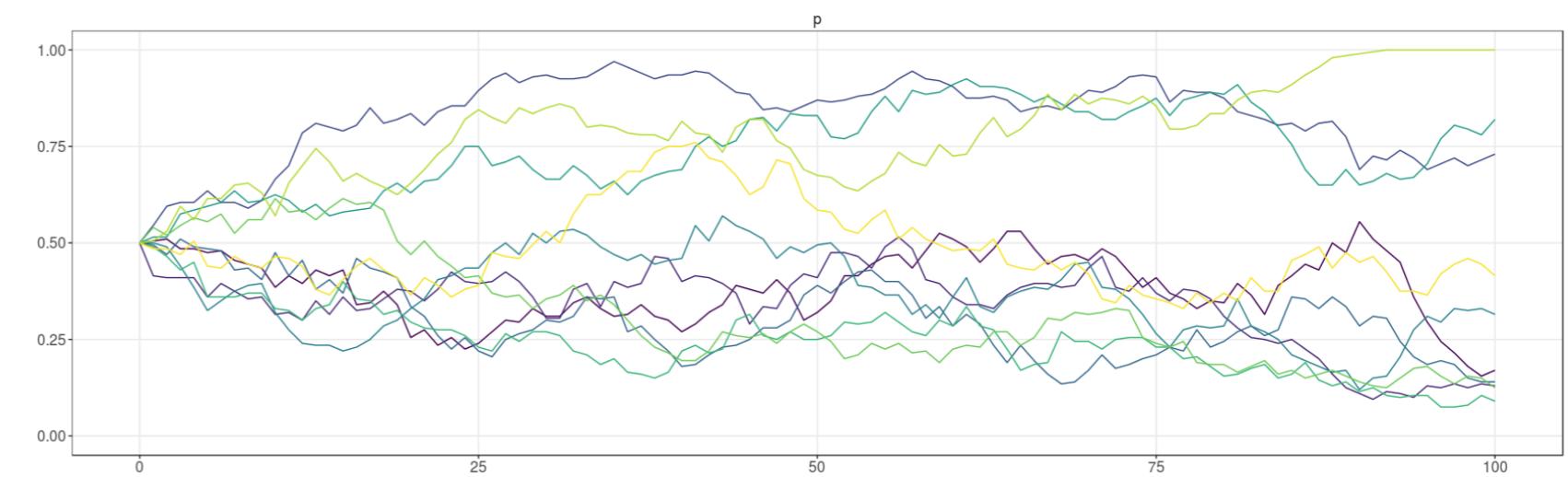
Migration Rate
0 0.2

Number of Populations
10

Population Size
100

Number of Generations
100

Infinite Population (no drift)



Final Generation State:

Fis	Hs	Ht	Fst
0.03	0.27	0.48	0.43

Replicate Runs

Click the button below to run repeated simulations of 2 populations using the current parameters (useful for estimating variance).

Run Replicate Simulations

number of replicates

10

Mean state at final generation:

p	Fis	Hs	Ht	Fst
0.4455	-0.0049	0.3078	0.3780	0.1480

<https://github.com/KHangoj/popgenteach/blob/master/exercise1.md>

Horse Domestication

The impact of horses on human history
Transportation, warfare, agriculture



Introduction to exercise 2

Present-day domesticated horses

Huge interbreed phenotypic diversity

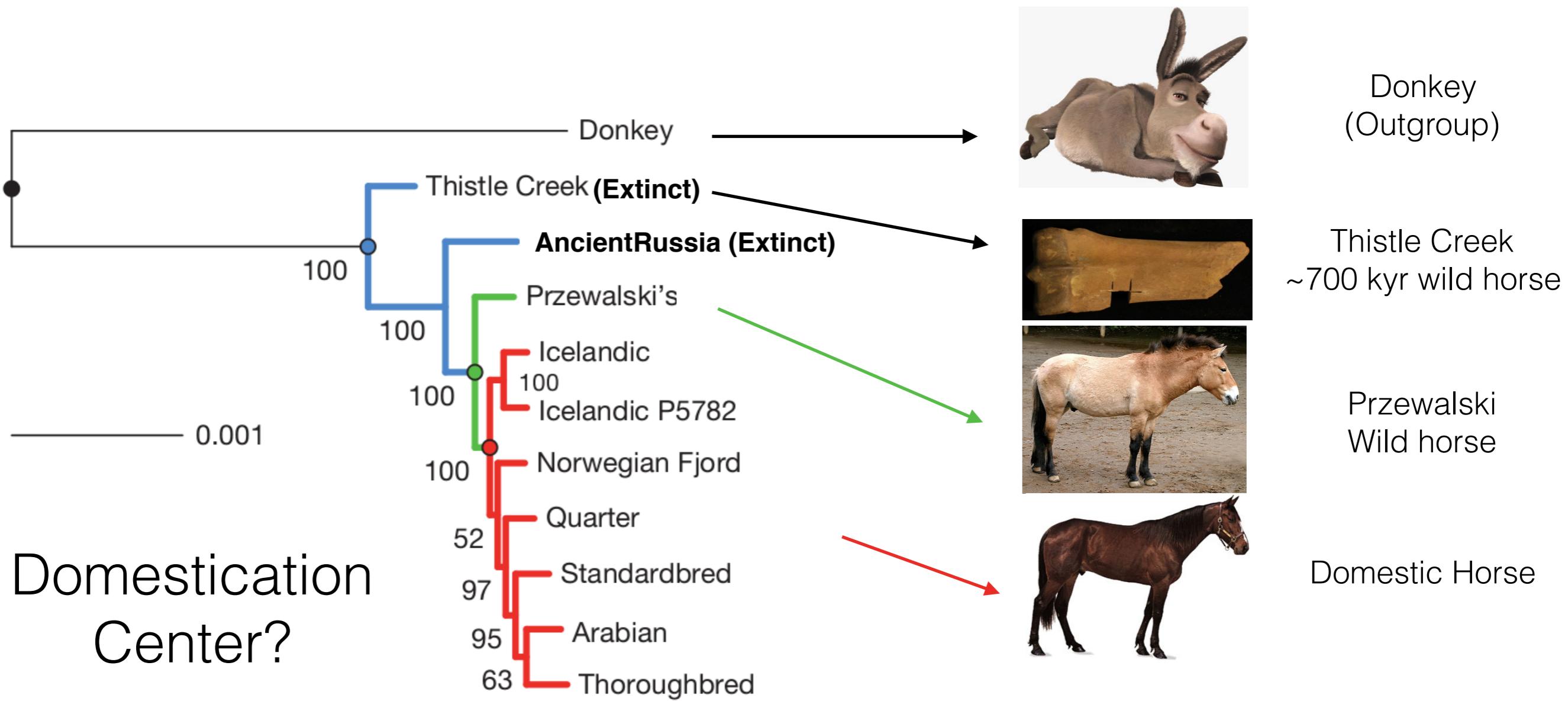


Przewalski: Last Extant Wild Horse

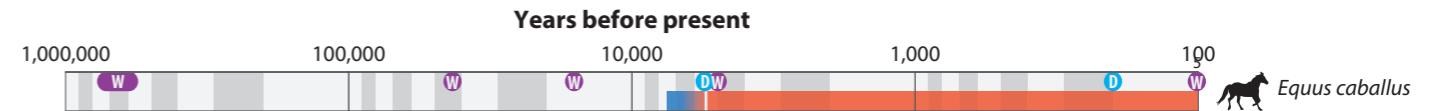
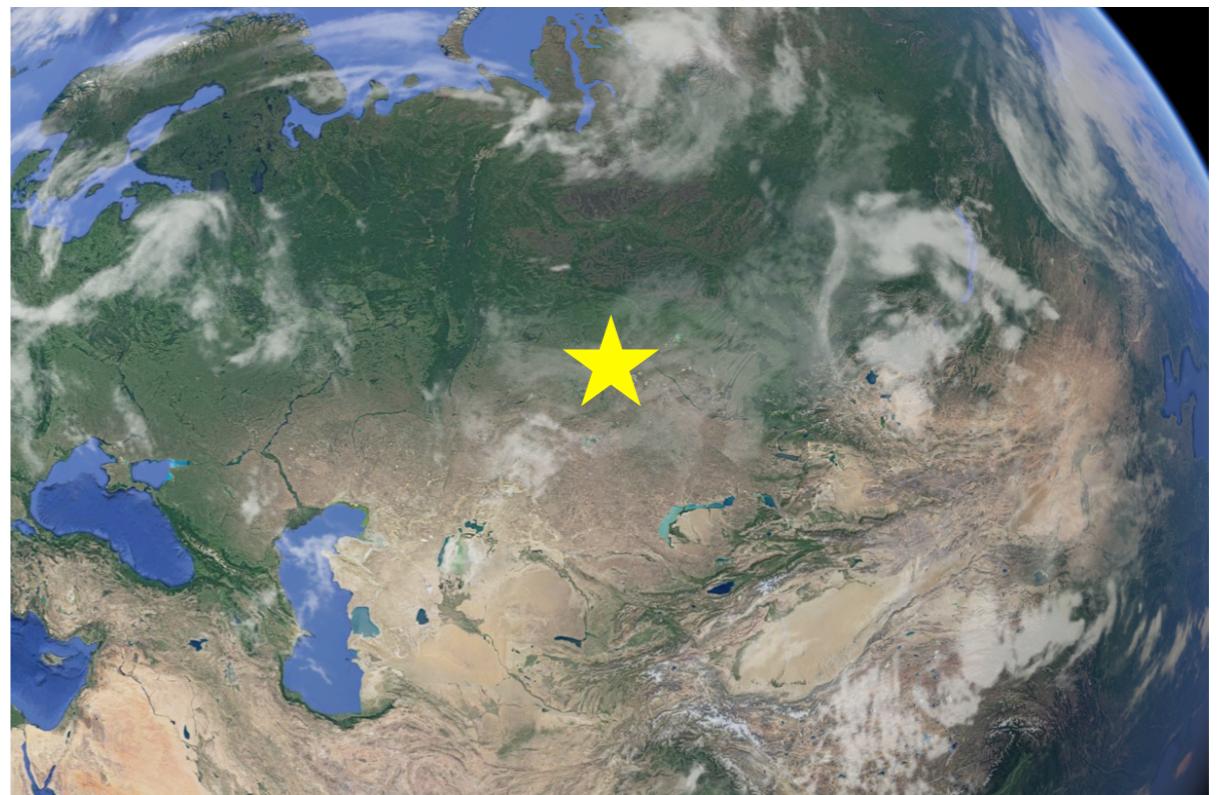
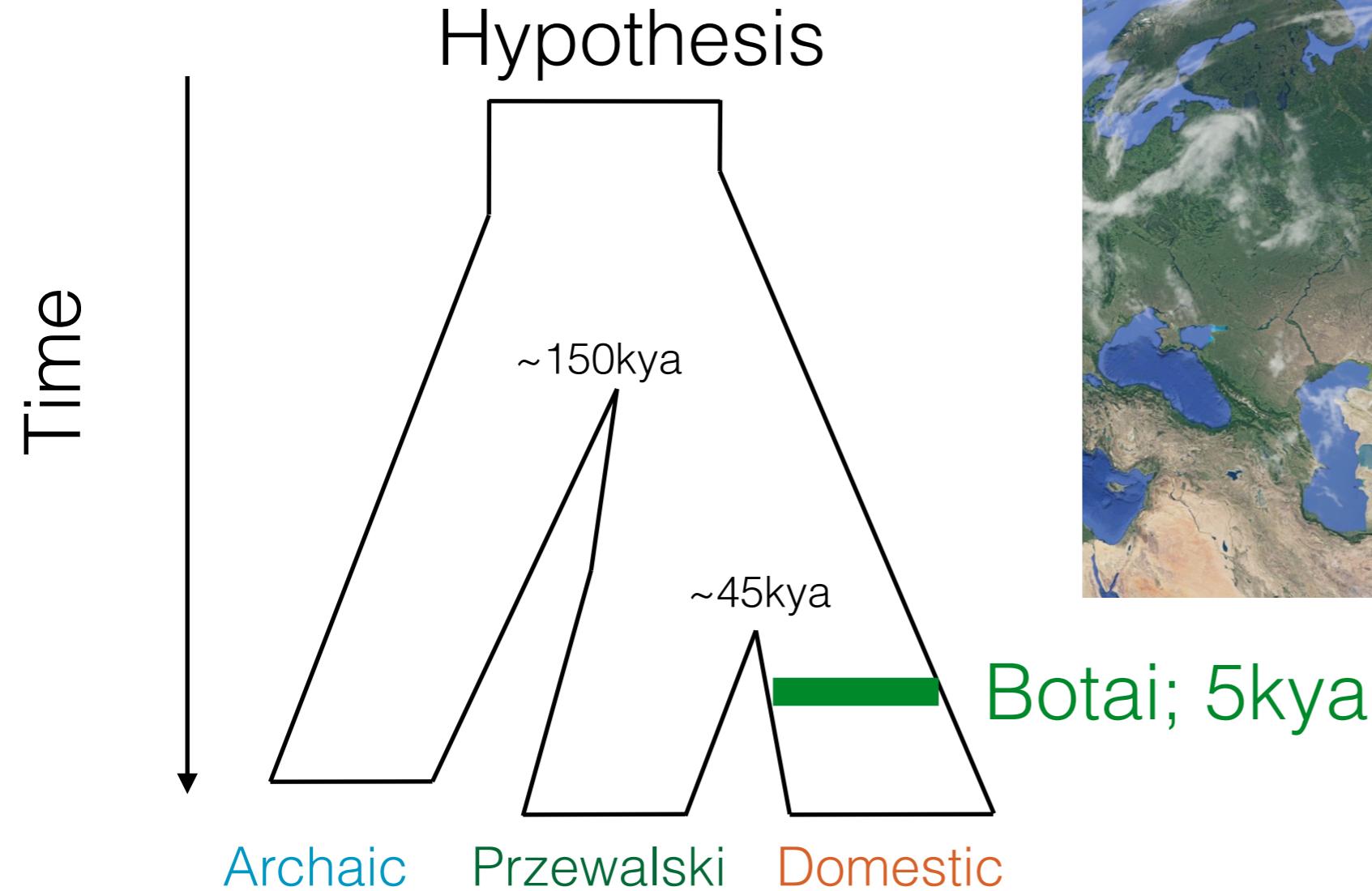


12 founders
Today: 2000 individuals

Genomic relationship of Equus (horse)

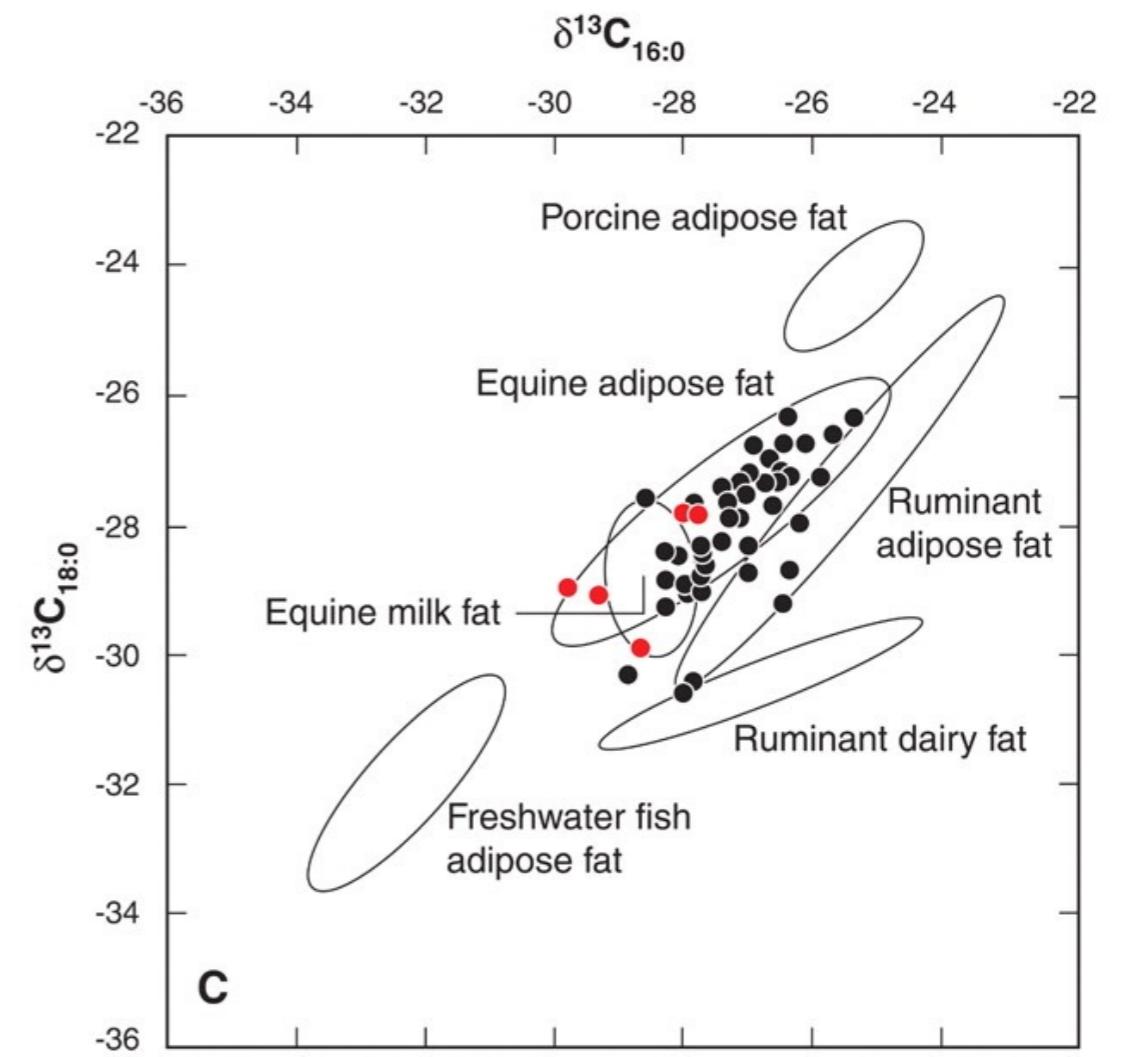
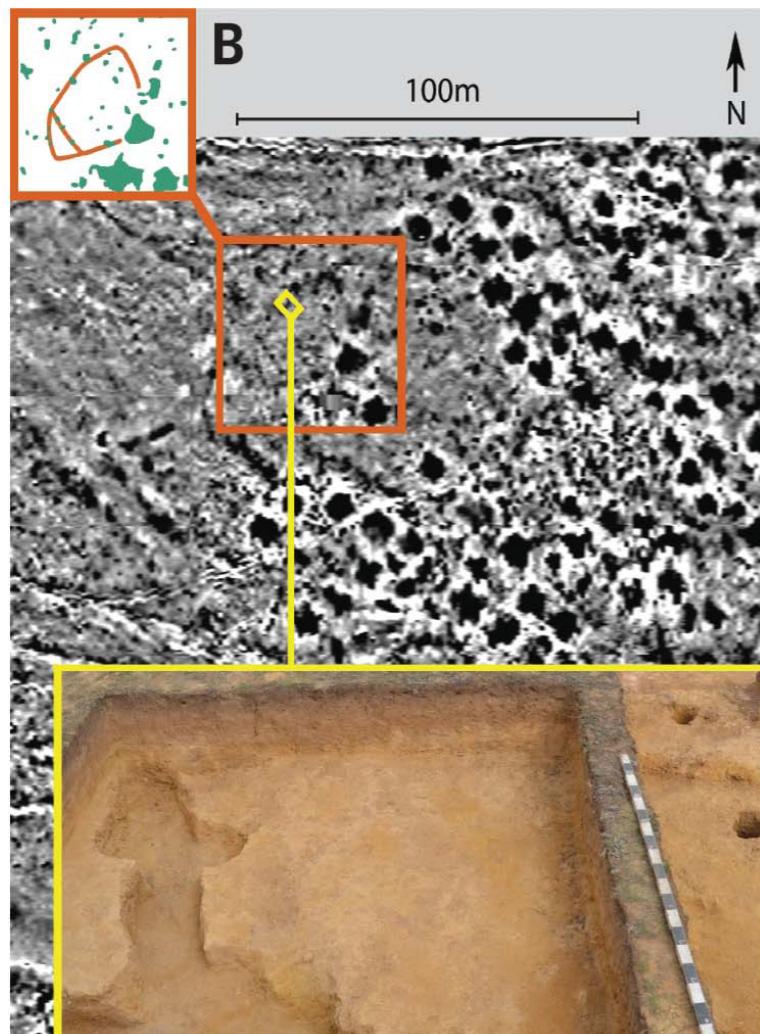


Earliest evidence of Domestication



Earliest evidence of Domestication

Botai Site



99% of 300k specimen originated from horses

Exercise 2

- **Hypothesis:**

- Botai, the earliest known domesticated horses, are the ancestors to present-day domesticated horses

- **20 genomes**

- 7 present-day domesticated horses
- 7 Przewalski horses
- 4 **Botai** horses (5kya)
- 2 Russian horses (5 and 42kya)

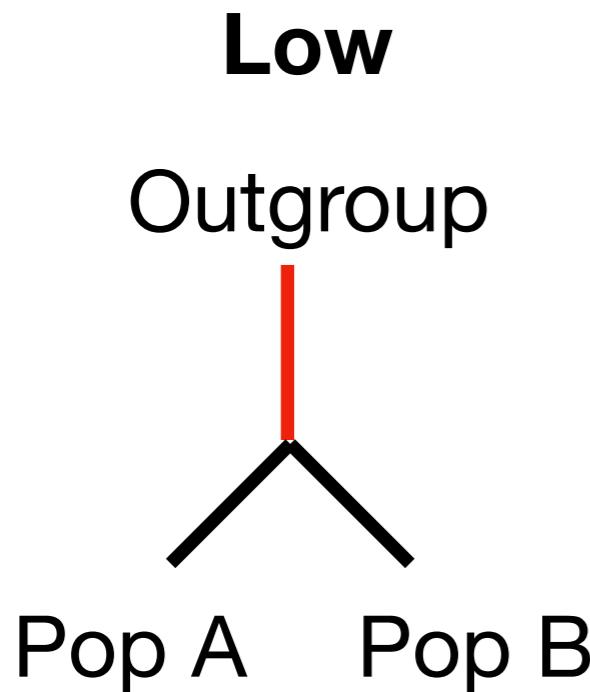
- **Two methods will be used:**

- f3-outgroup statistic
- Treemix



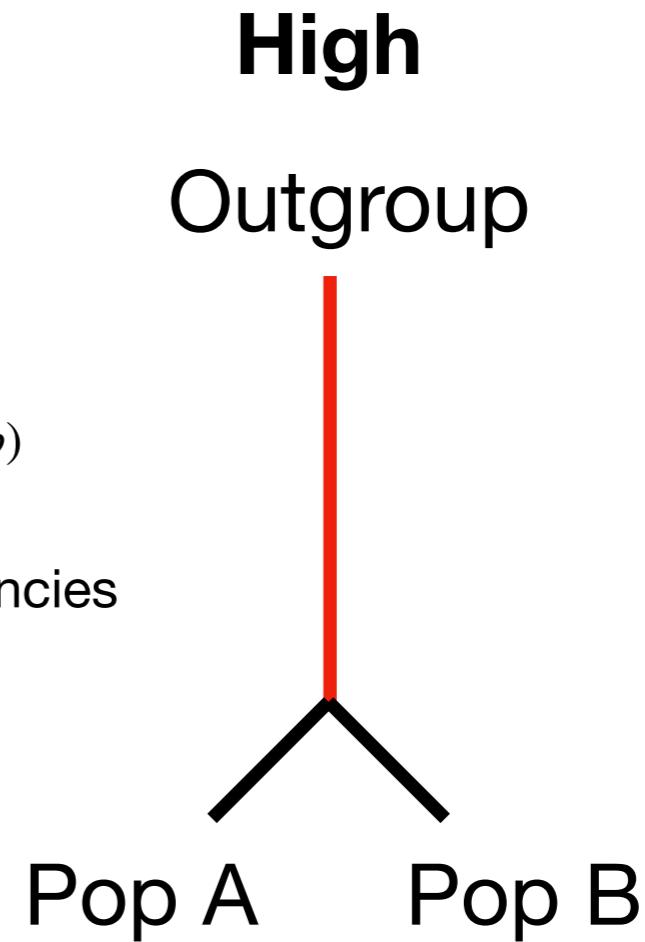
Exercise 2 - f3-outgroup

Shared genetic drift between A and B



$$\text{f3-outgroup} = \frac{1}{N} \sum^N (o - a)(o - b)$$

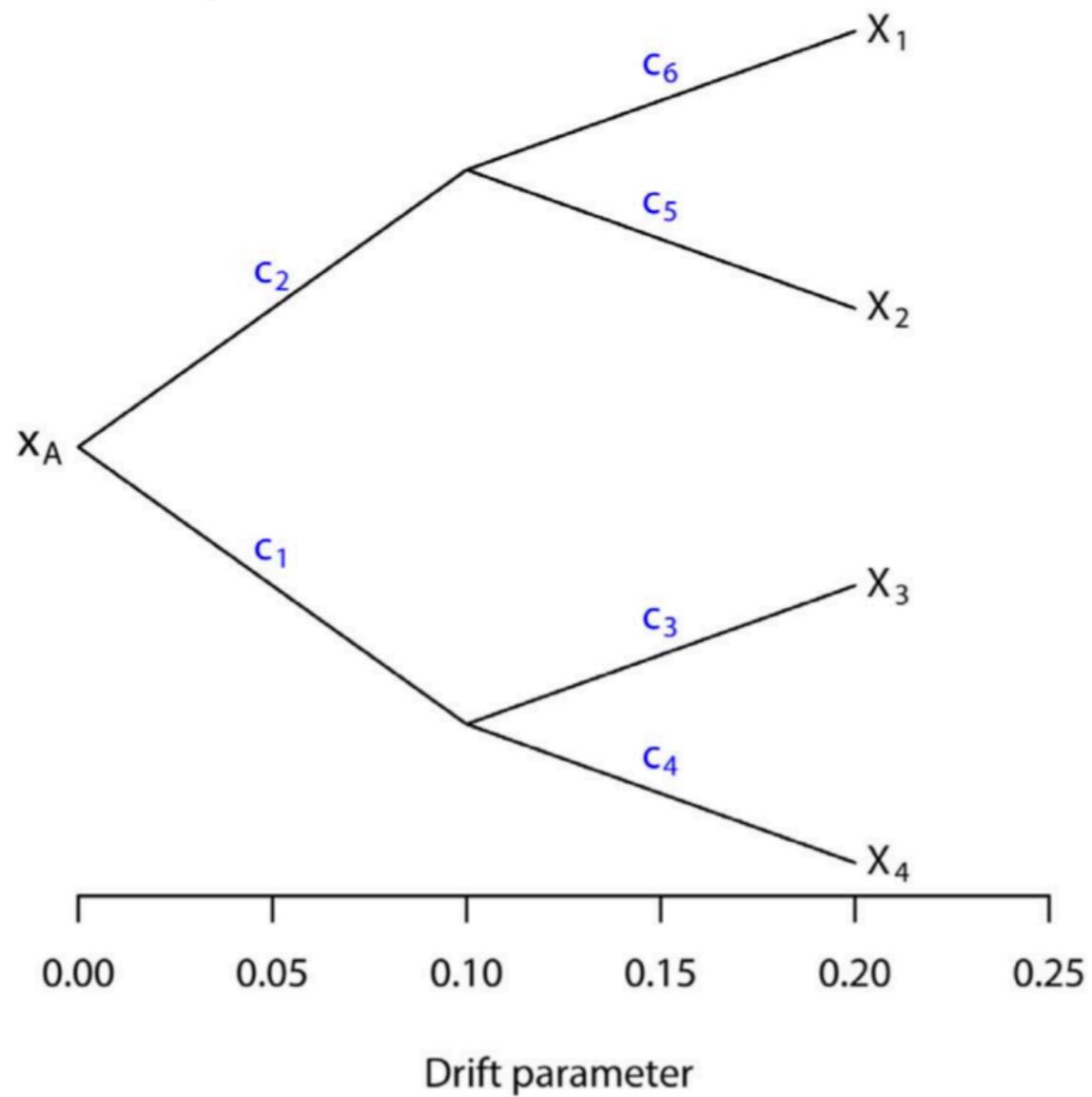
o , a , and b are allele frequencies



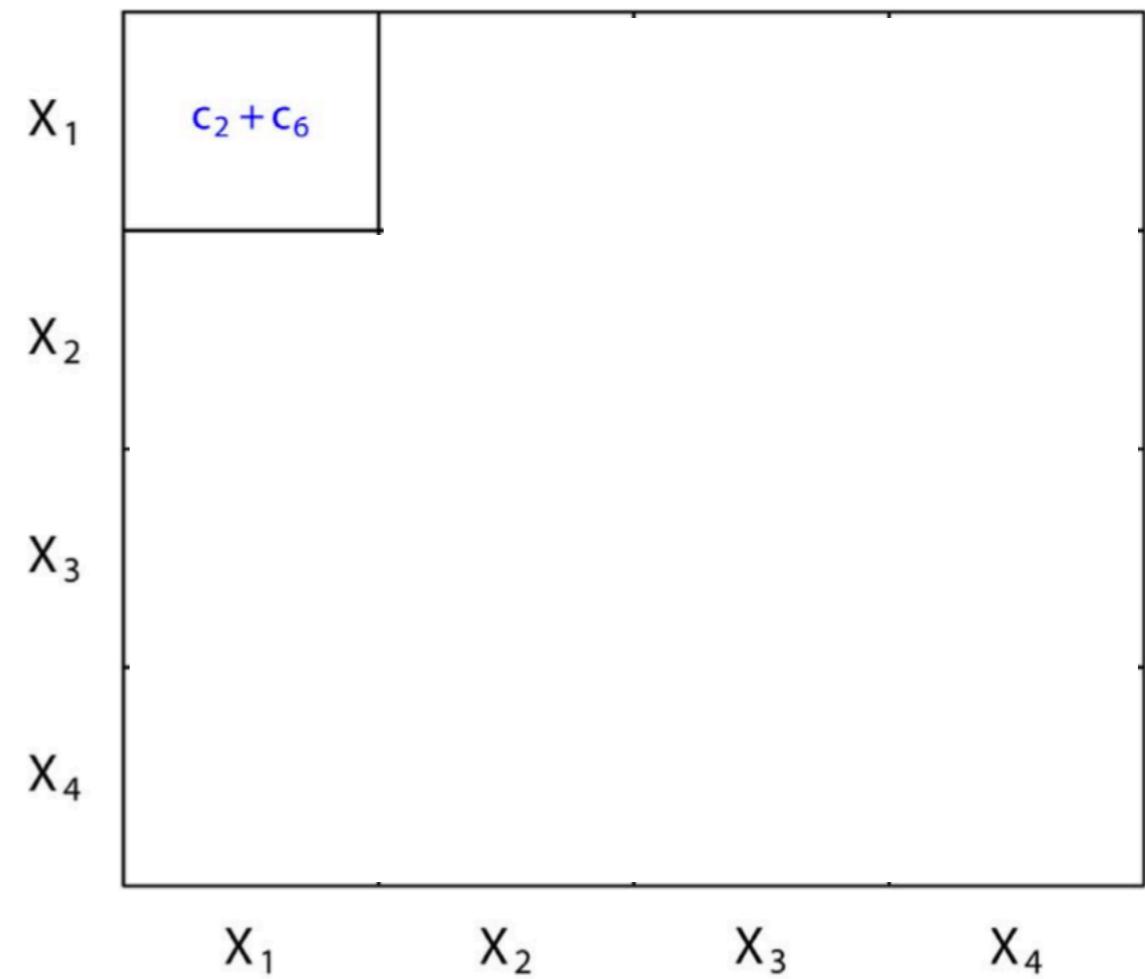
Exercise 2 - Treemix

Population graphs based on genetic drift

A. Example tree



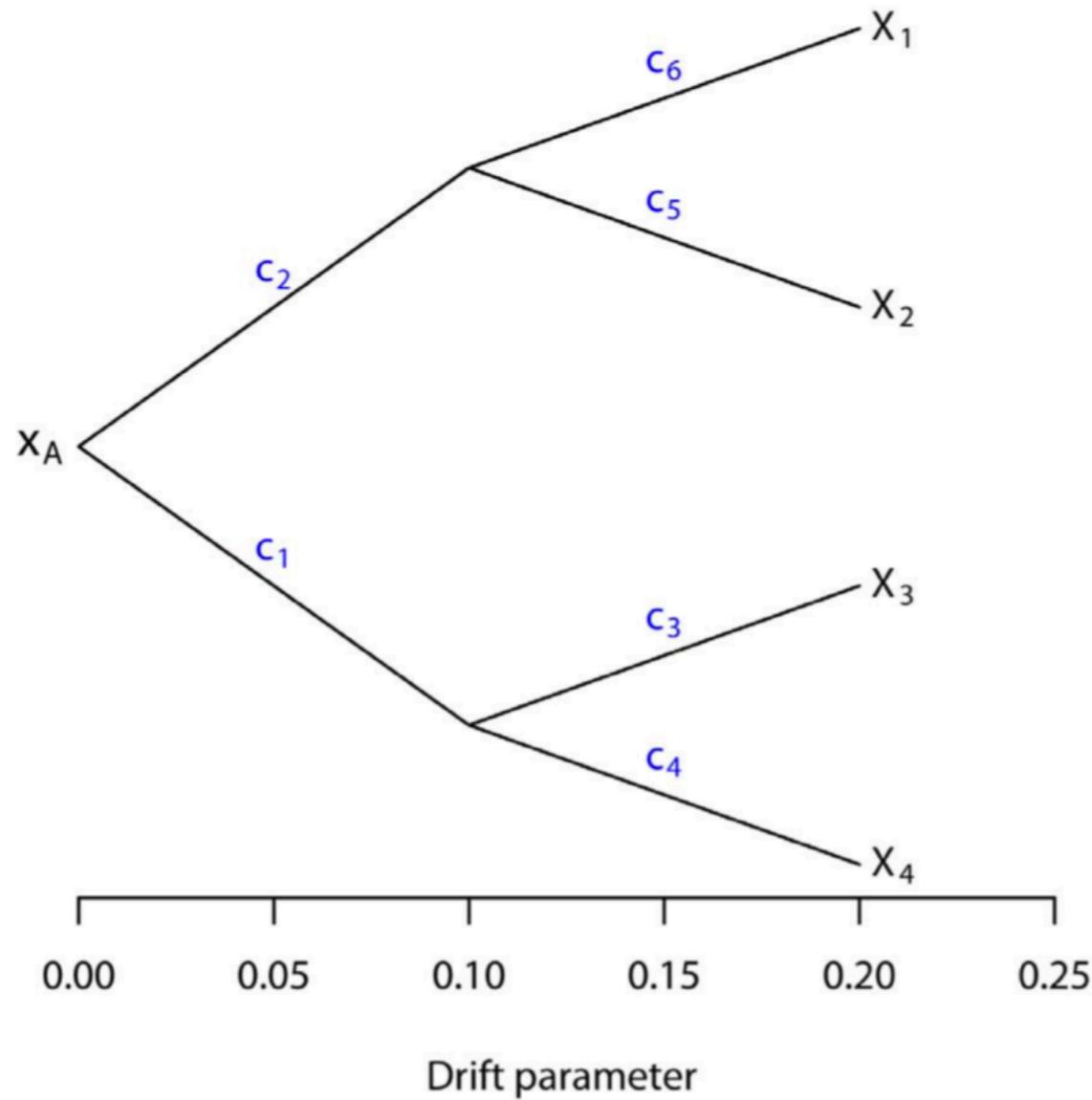
B. Covariance matrix for tree in A.



Exercise 2 - Treemix

Population graphs based on genetic drift

A. Example tree



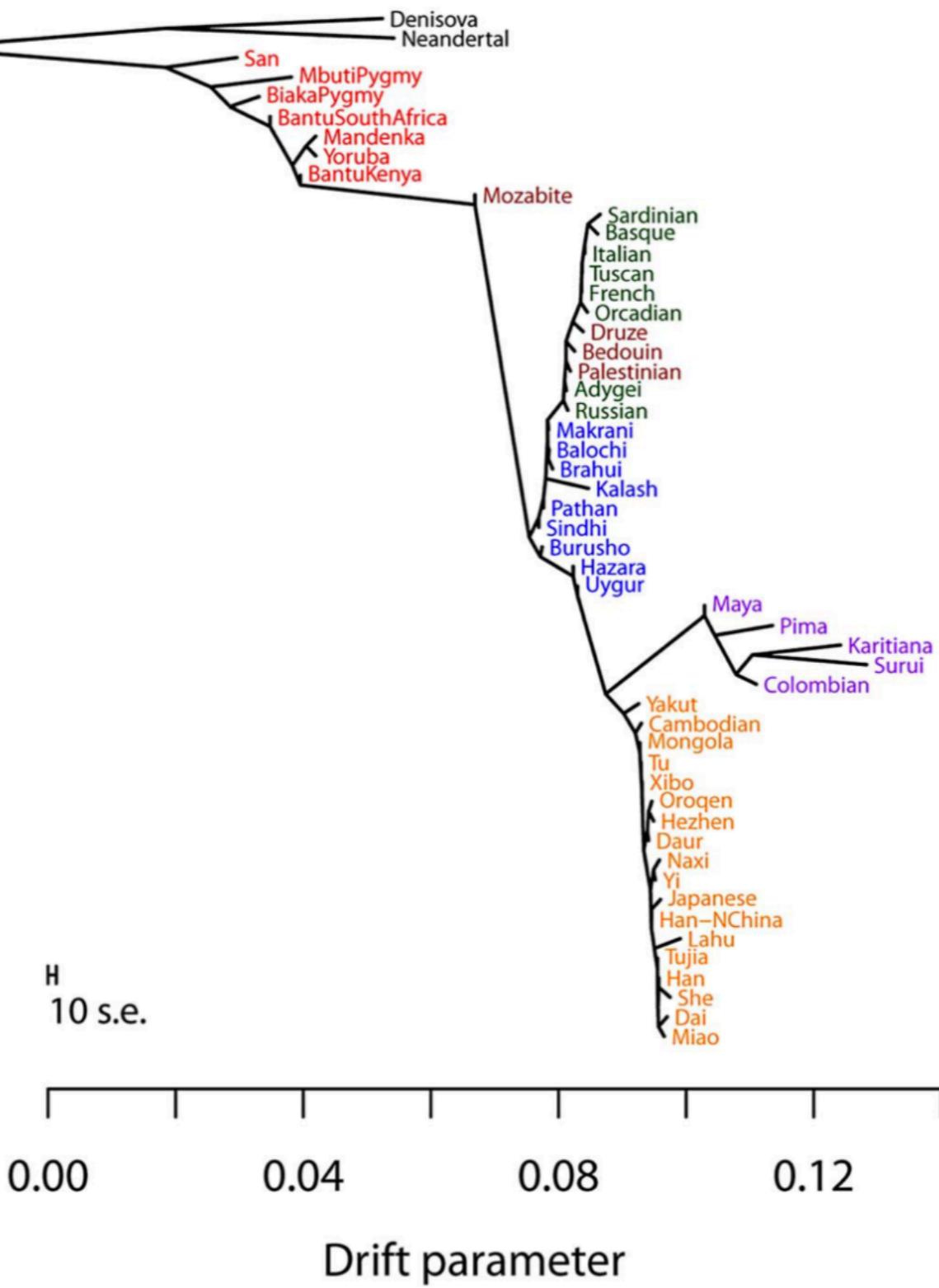
B. Covariance matrix for tree in A.

X_1	X_2	X_3	X_4	
X_1	$c_2 + c_6$	c_2	0	0
X_2	c_2	$c_2 + c_5$	0	0
X_3	0	0	$c_1 + c_3$	c_1
X_4	0	0	c_1	$c_1 + c_4$

Exercise 2 - Treemix

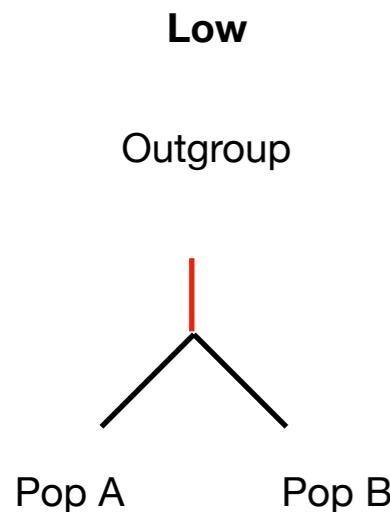
A

Population graphs
based on genetic drift

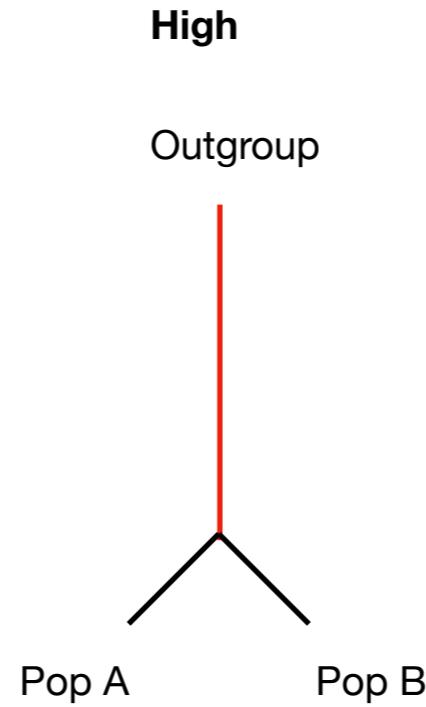


Exercise 2

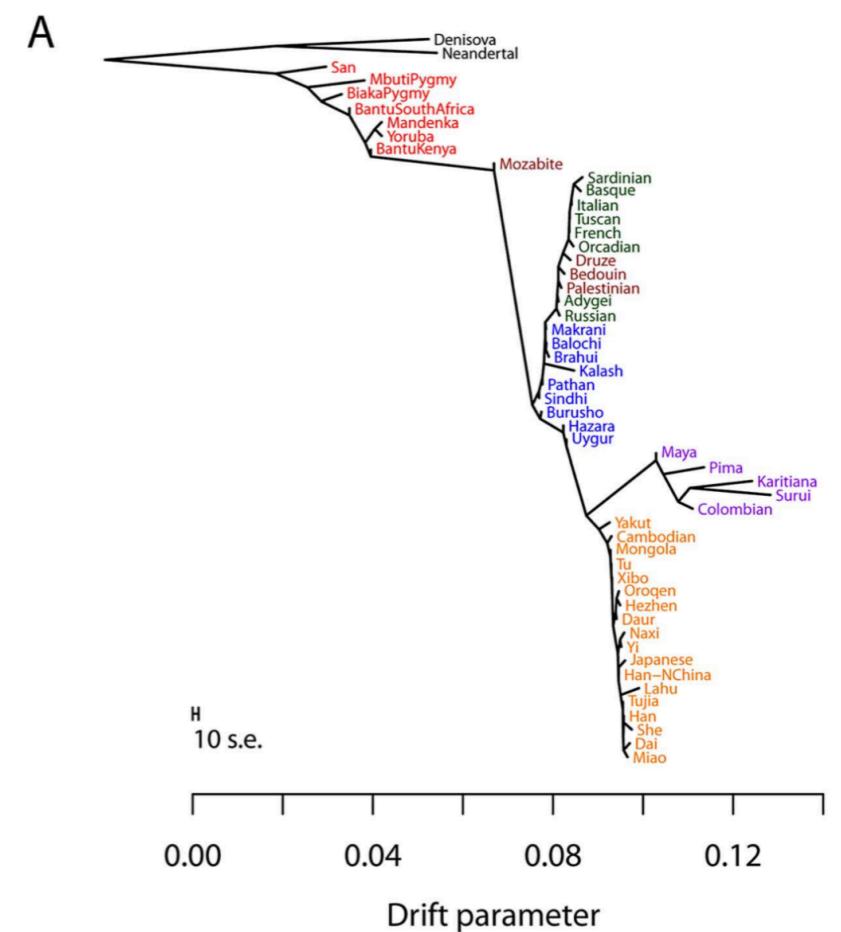
<https://github.com/KHanghoj/popgenteach/blob/master/exercise2.md>



$$f3\text{-outgroup} = \frac{1}{N} \sum^N (o - a)(o - b)$$

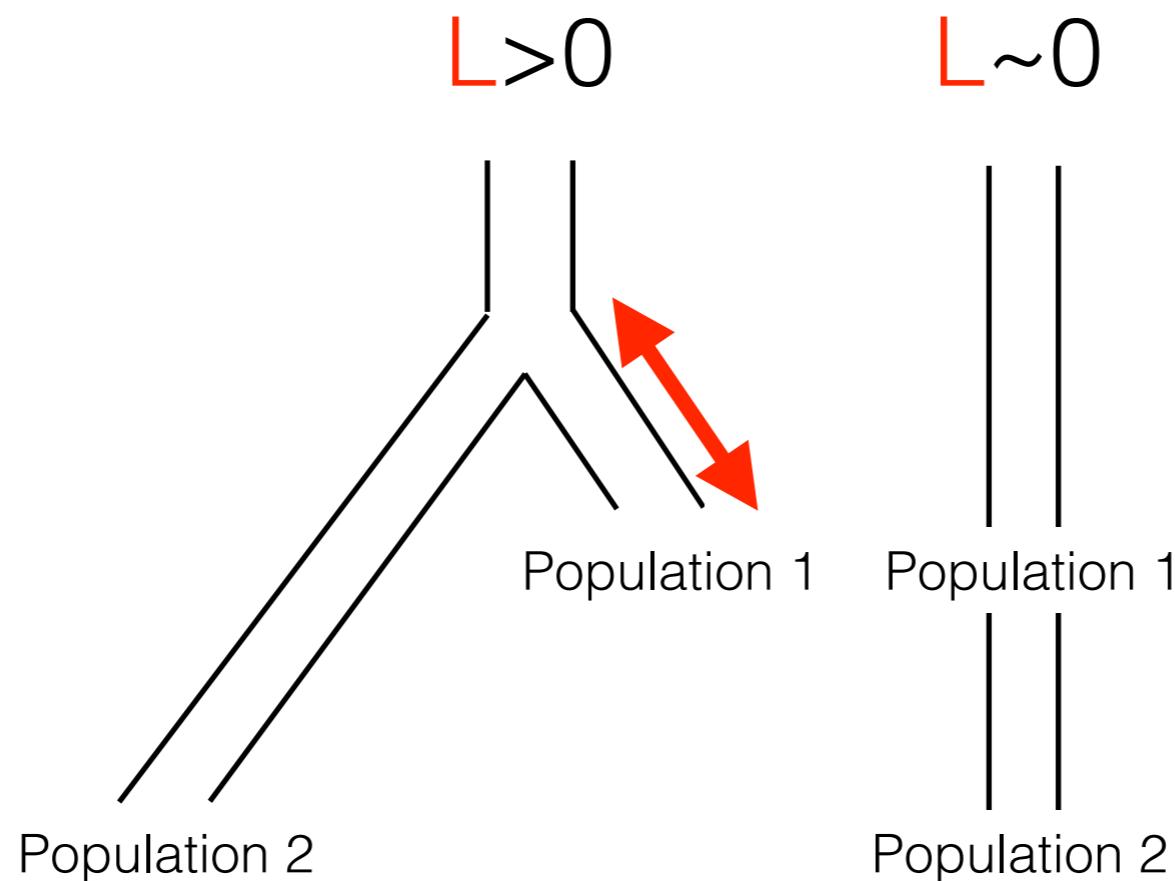


f3-outgroup

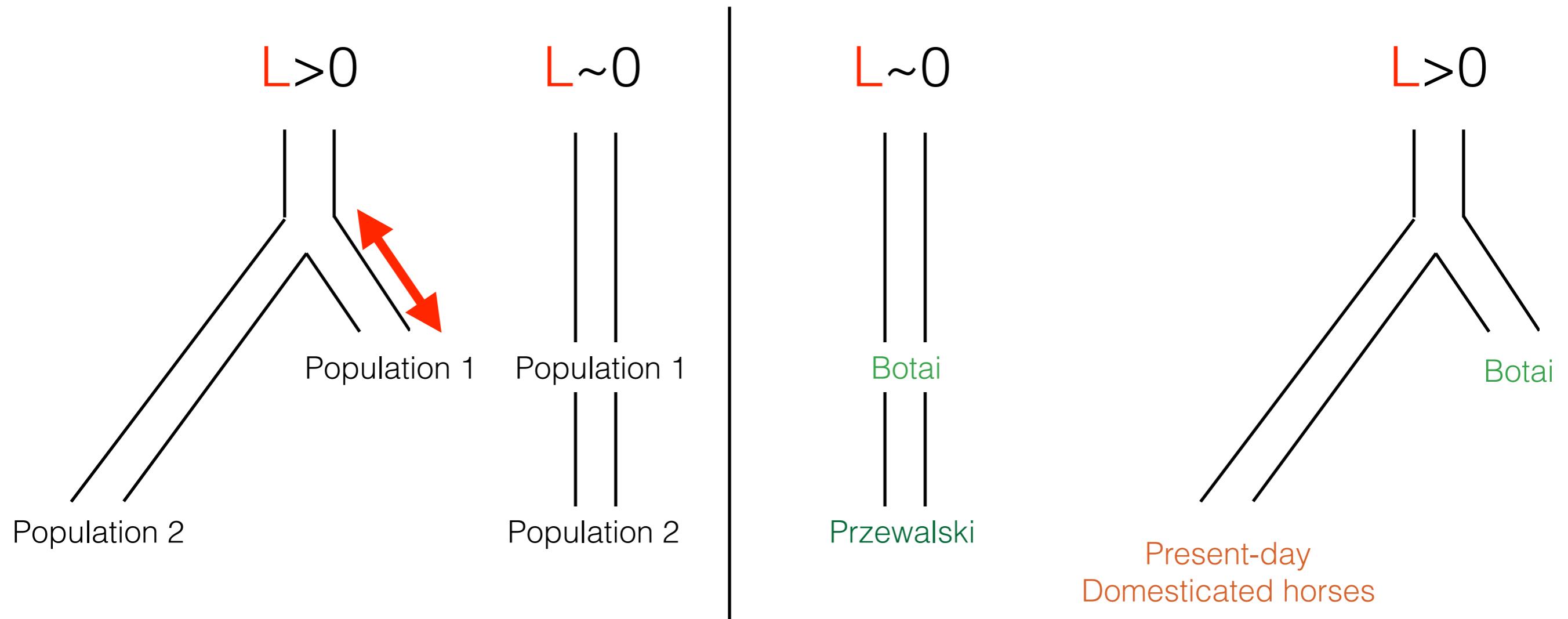


Treemix

Formal Test for Direct Ancestry



Formal Test for Direct Ancestry



Botai was the **direct** ancestor of **feral** Przewalski horses

