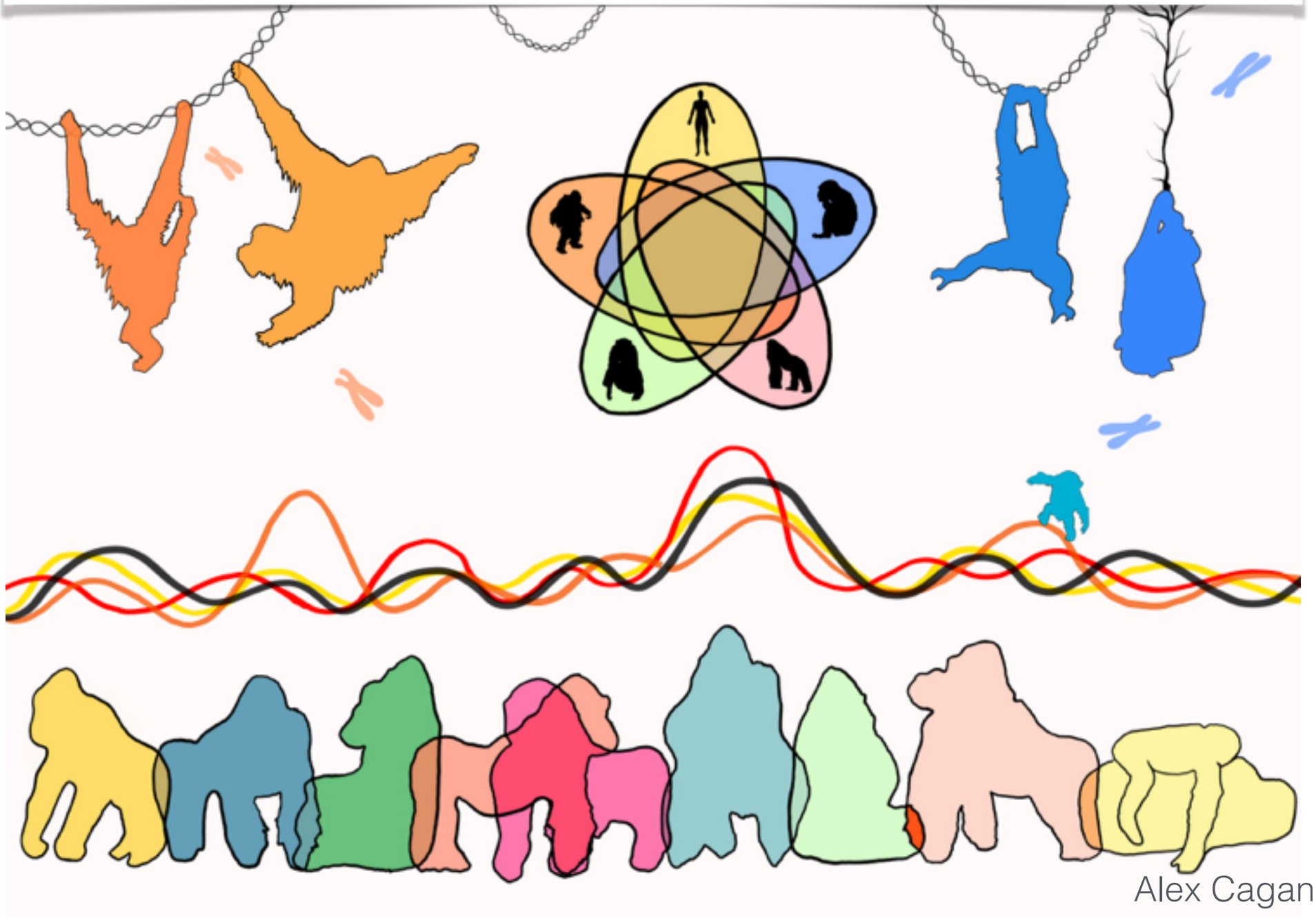
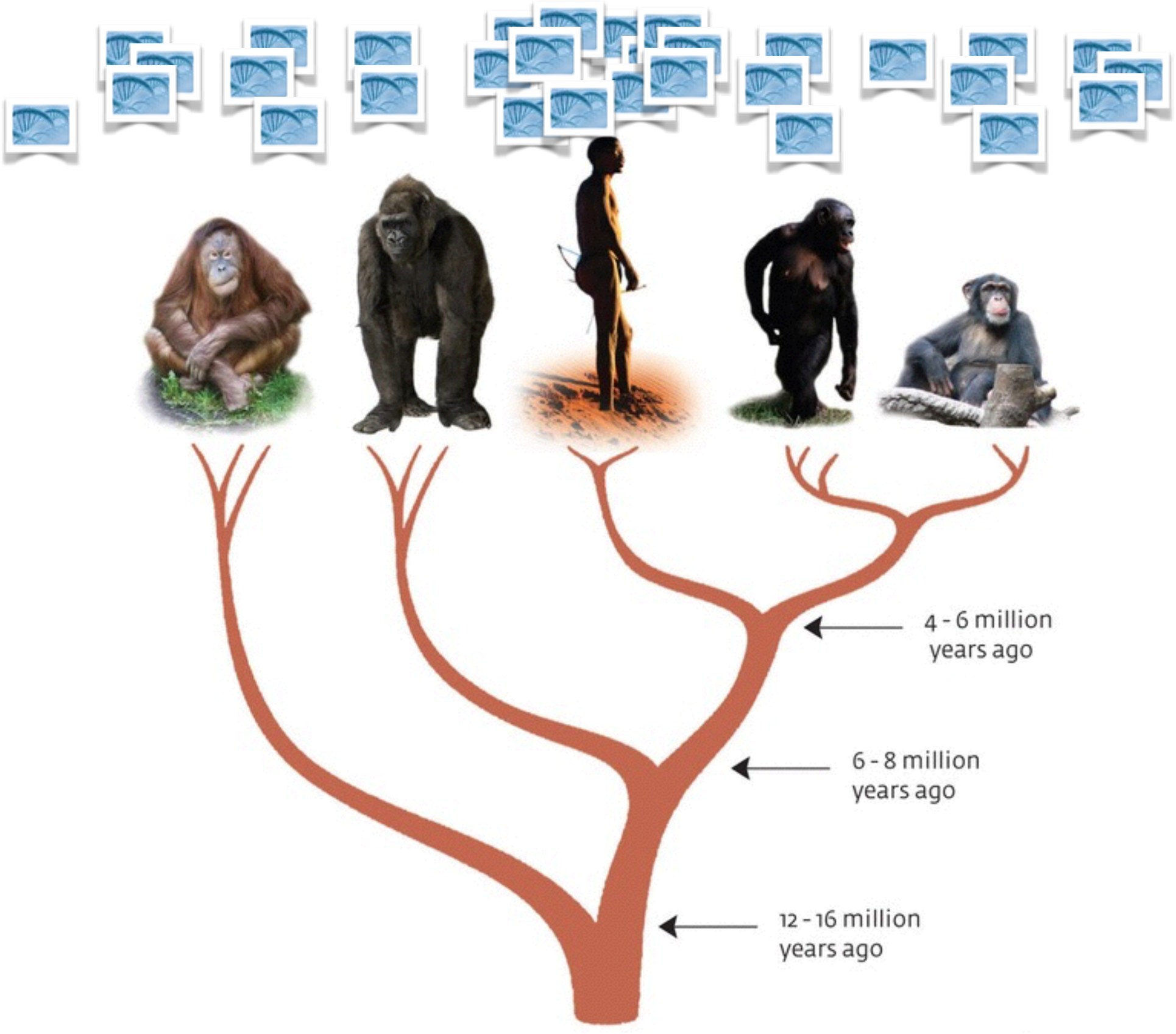
Species divergence
(our closest living relatives)



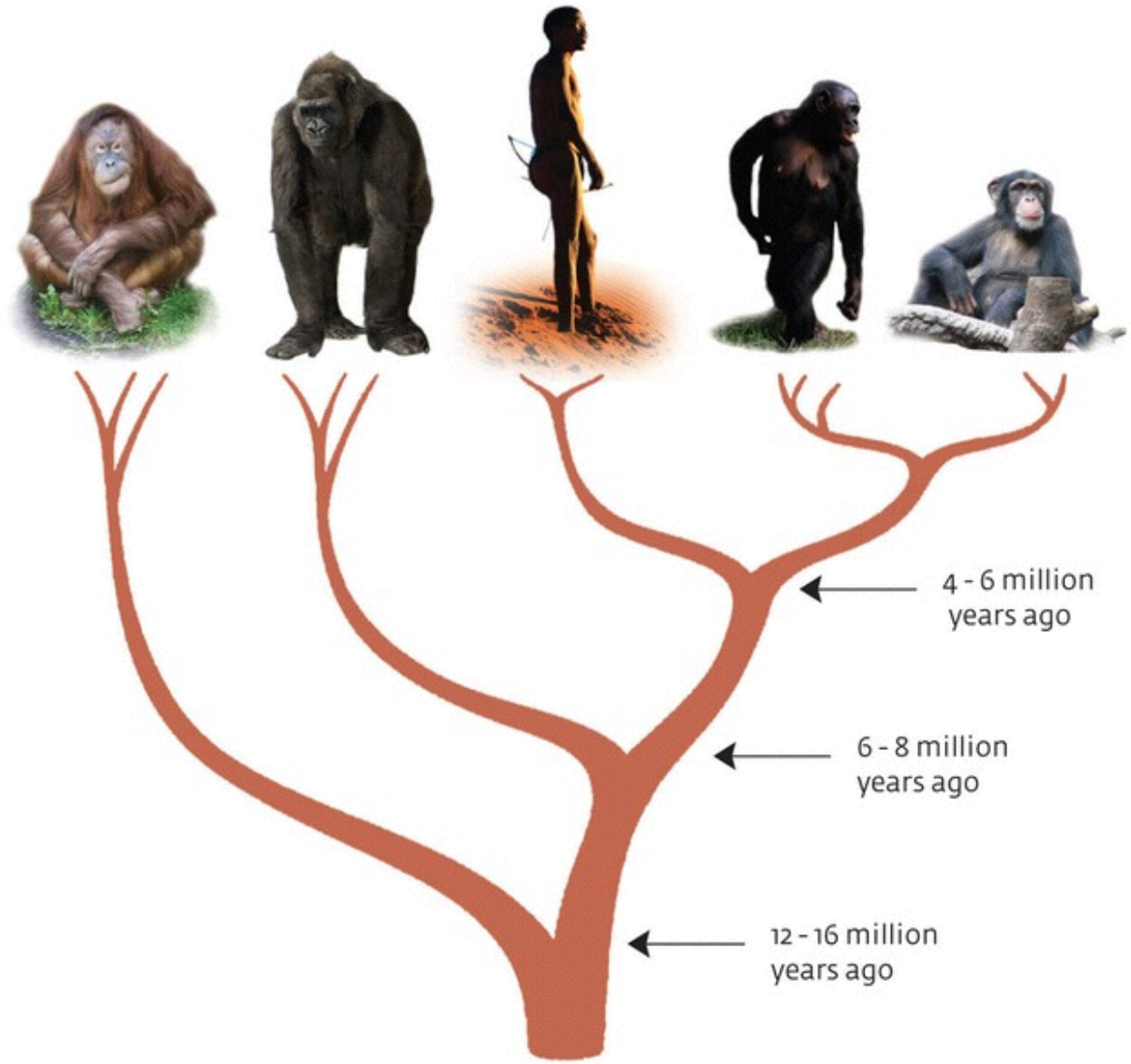
Aida Andrés



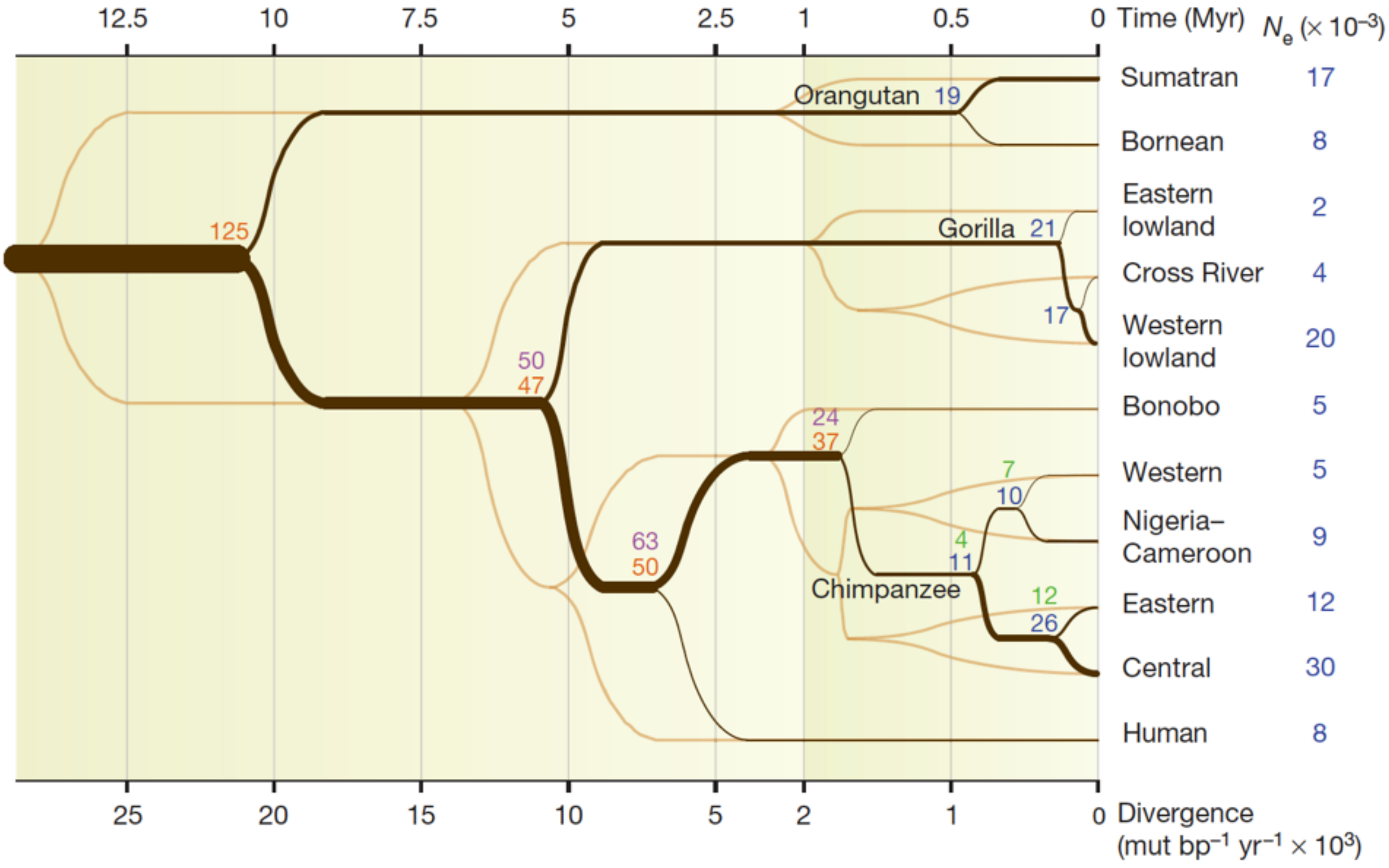
Prado-Martinez *et al.*, Science 2013; De Manuel *et al.*, Science 2016; *e.g.* 1000 Genomes dataset

Differ in:

- Demographic history
- Social patterns
- Mating behaviour
- Environment
- Diet
- Size
- Locomotion



Extremely closely related

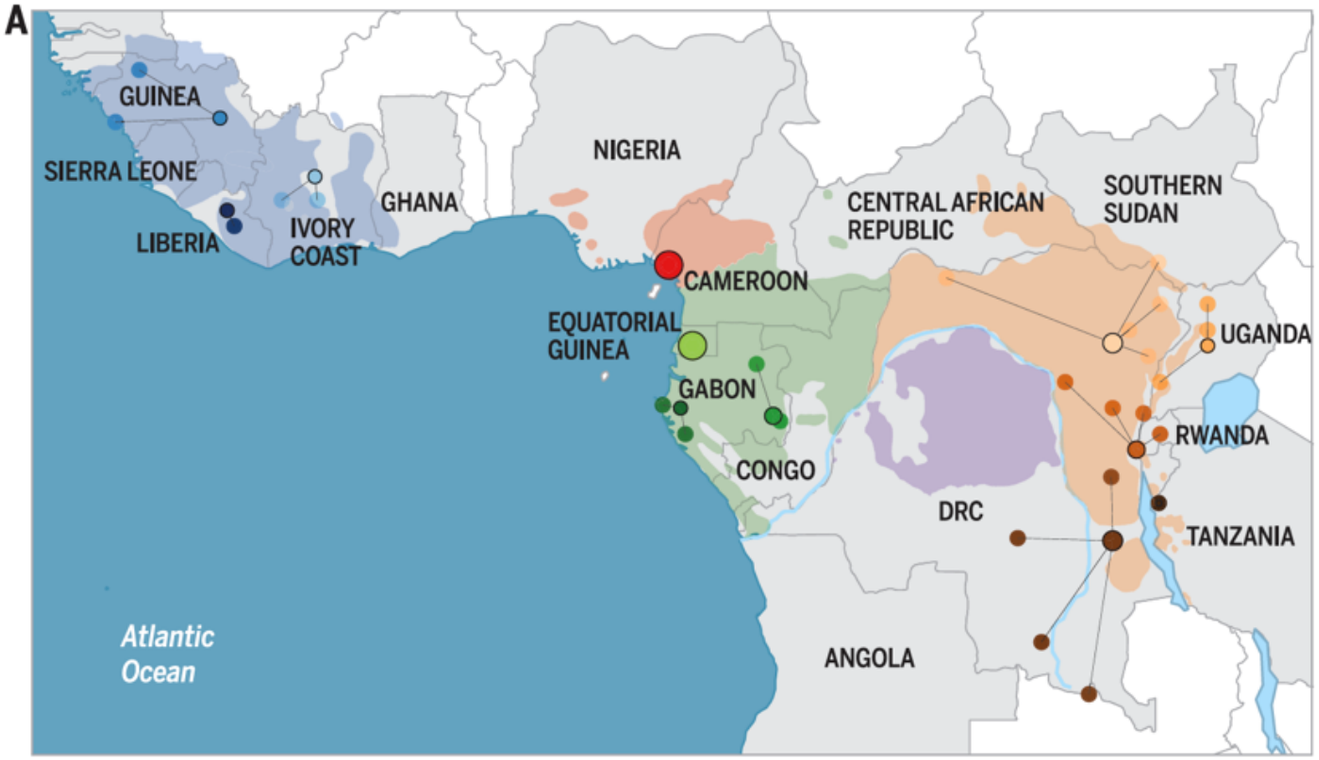


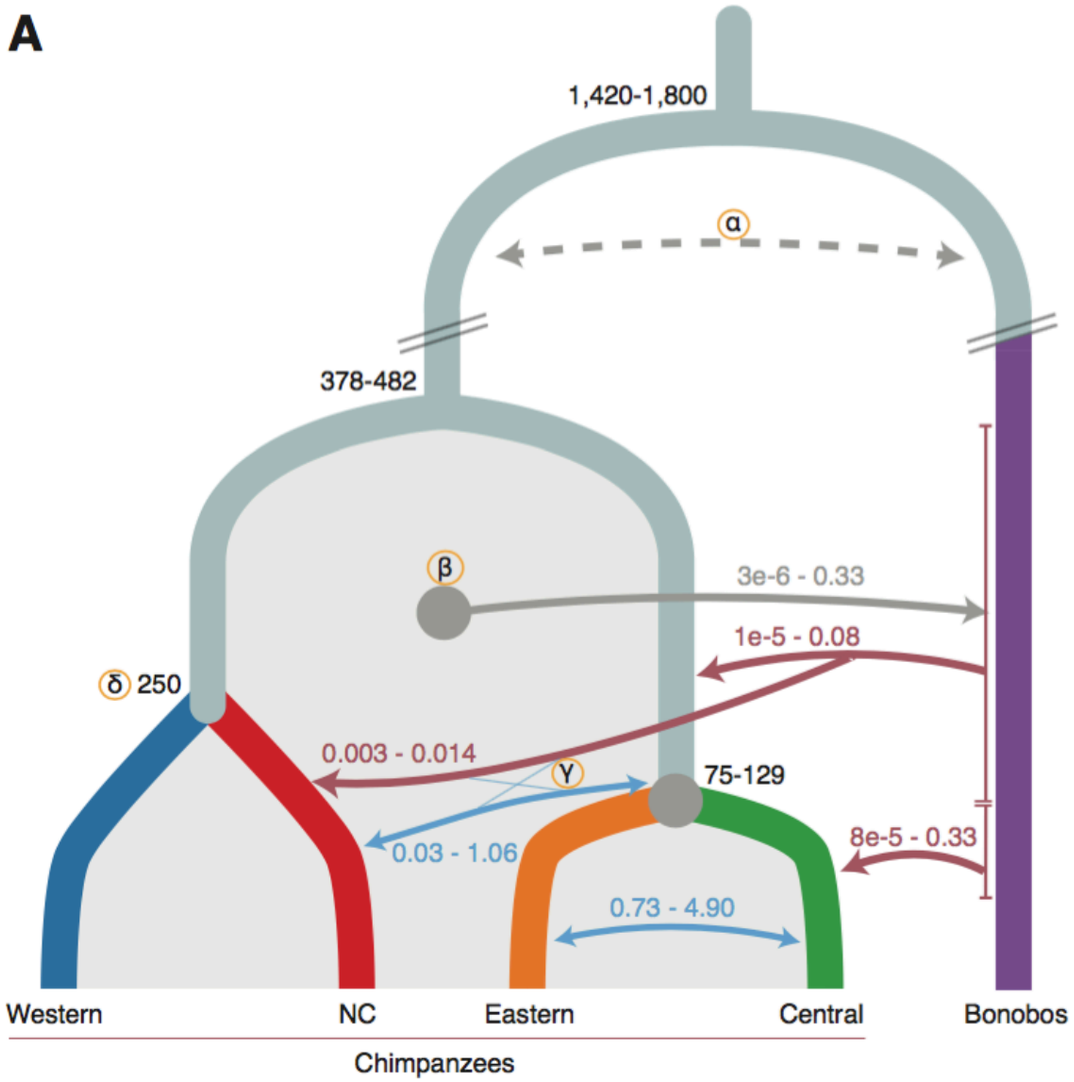
Prado-Martinez *et al.*, Science 2013

75 chimpanzee genomes (4 subspecies)

de Manuel et al., Nature 2016

Chimpanzee subspecies



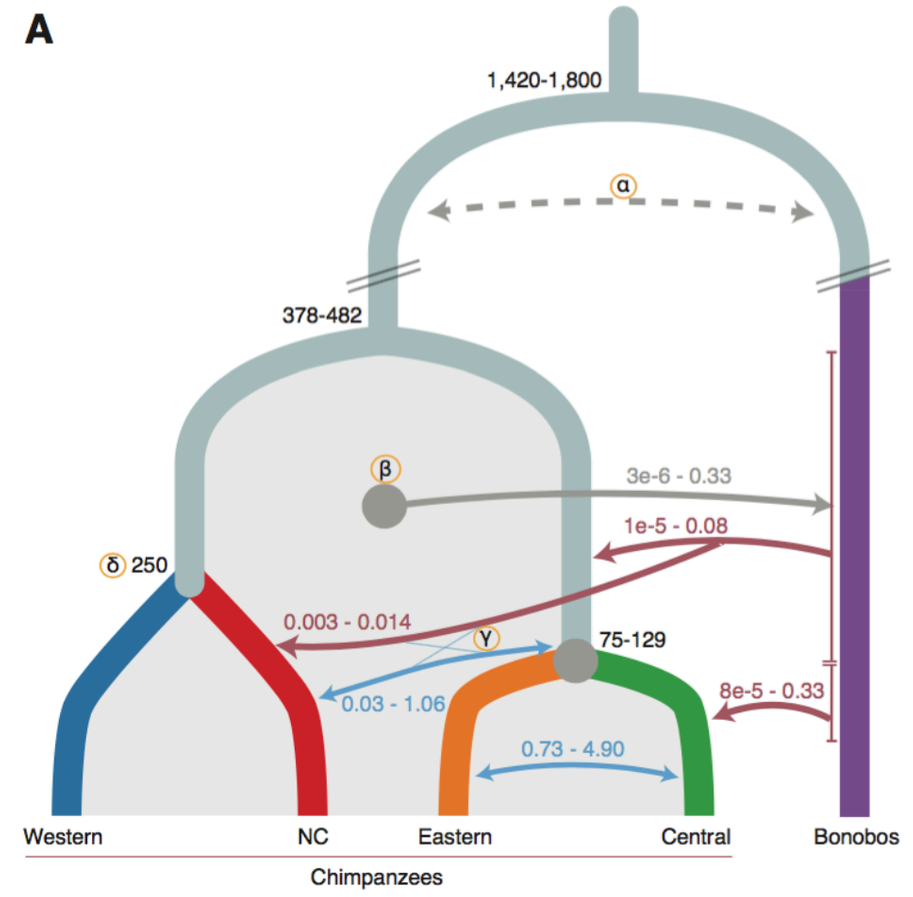


De Manuel *et al.*, Science 2016; Gonder *et al.*, PNAS 2011

Chimpanzee subspecies

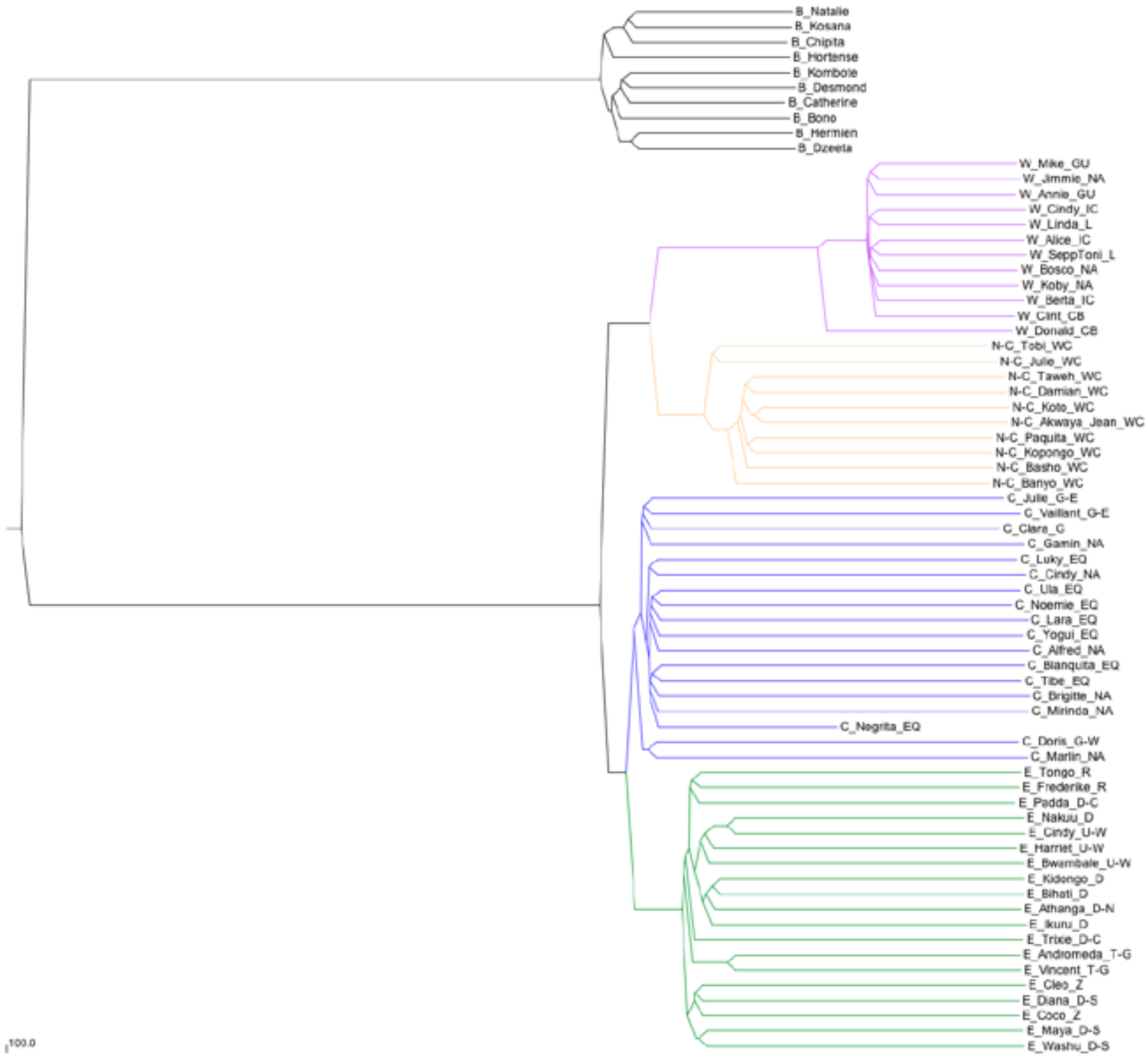
Fst	Central	Eastern	Nigeria-Cameroon	Western
Central		0.10	0.20	0.35
Eastern			0.24	0.39
Nigeria-Cameroon				0.39
Western				

Fst	Yoruba	Great Britain	Japan
Yoruba		0.14	0.13
Great Britain			0.11
Japan			



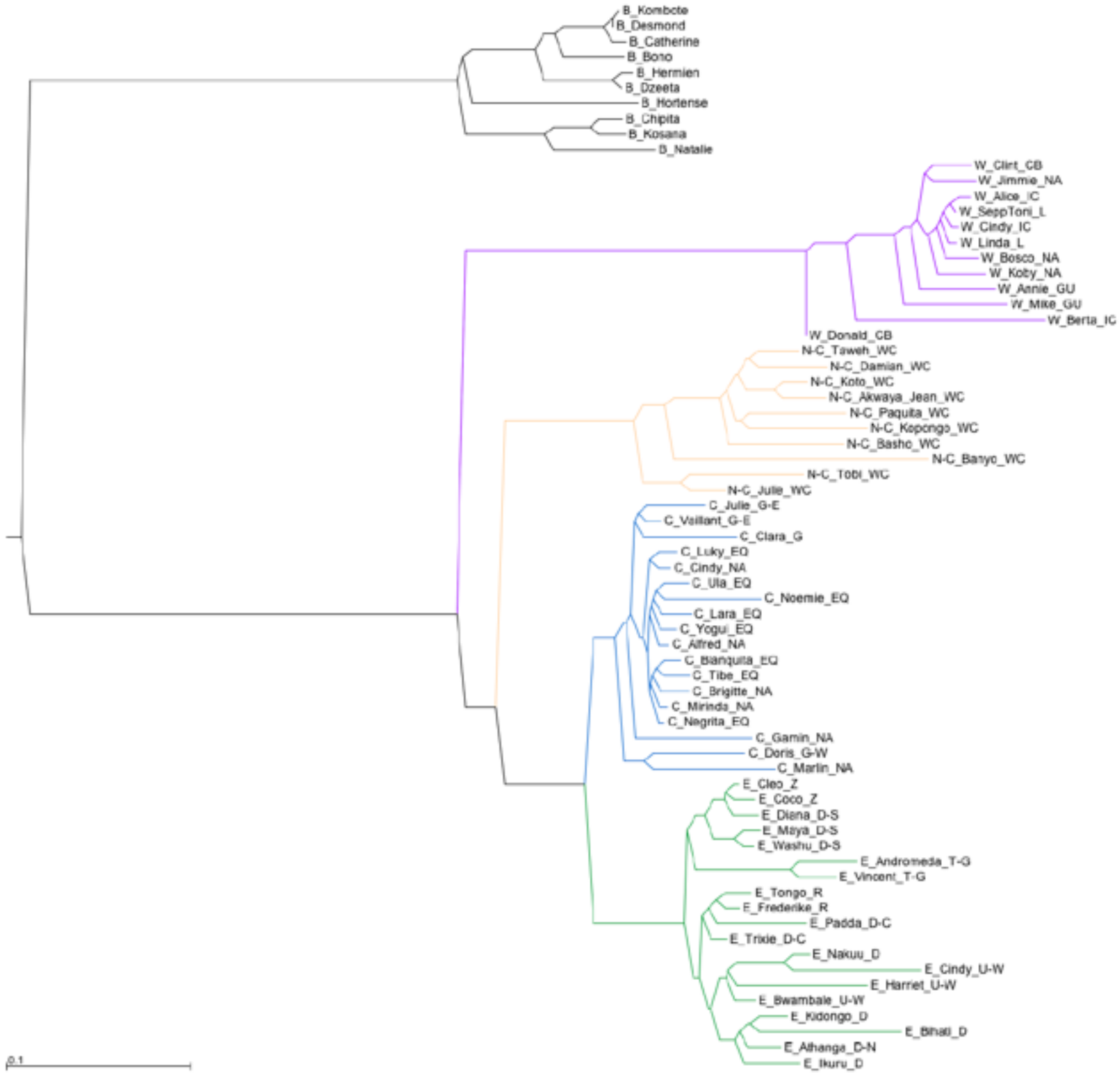
De Manuel *et al.*, Science 2016;

Chimpanzee subspecies



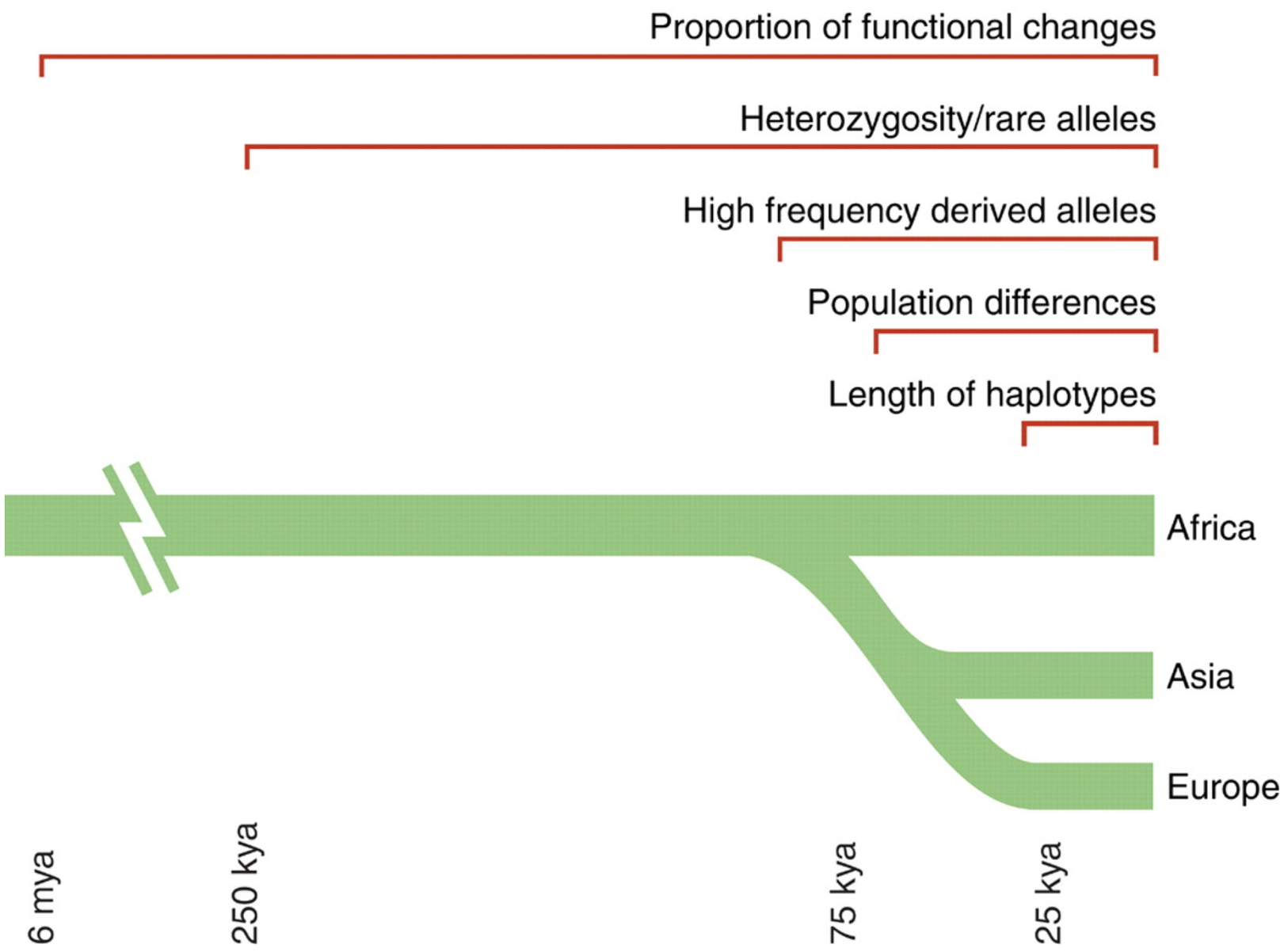
De Manuel *et al.*, Science 2016;

Chimpanzee subspecies



De Manuel *et al.*, Science 2016;

Time Scales for the Signatures of Selection



Sabeti et al., Science, 2006

Time Scales for the Signatures of Selection

Proportion of functional changes

dN/dS

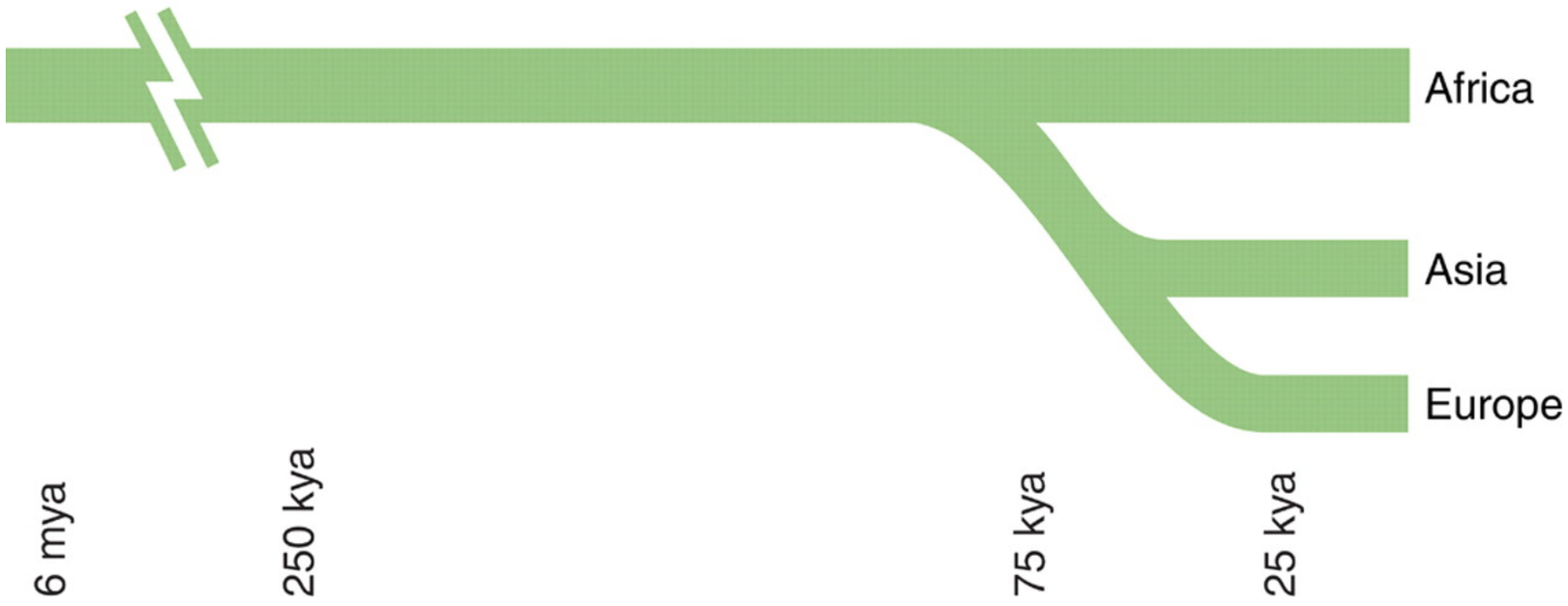
McDonald&Kreitman

Accumulation of changes (substitutions)

Lineage sorting

HKA

Extended Lineage Sorting



Sabeti et al., Science, 2006

Protein-coding evolution: Ka/Ks

non-synonymous

synonymous

	AGA									UUA					AGC					
	AGG									UUG					AGU					
GCA	CGA						GGA			CUA				CCA	UCA	ACA			GUA	
GCC	CGC						GGC		AUA	CUC				CCC	UCC	ACC			GUC	UAA
GCG	CGG	GAC	AAC	UGC	GAA	CAA	GGG	CAC	AUC	CUG	AAA		UUC	CCG	UCG	ACG		UAC	GUG	UAG
GCU	CGU	GAU	AAU	UGU	GAG	CAG	GGU	CAU	AUU	CUU	AAG	AUG	UUU	CCU	UCU	ACU	UGG	UAU	GUU	UGA
Ala	Arg	Asp	Asn	Cys	Glu	Gln	Gly	His	Ile	Leu	Lys	Met	Phe	Pro	Ser	Thr	Trp	Tyr	Val	stop
A	R	D	N	C	E	Q	G	H	I	L	K	M	F	P	S	T	W	Y	V	

$$\frac{Ka}{Ks} = \frac{\text{proportion of NS changes}}{\text{proportion of S changes}}$$

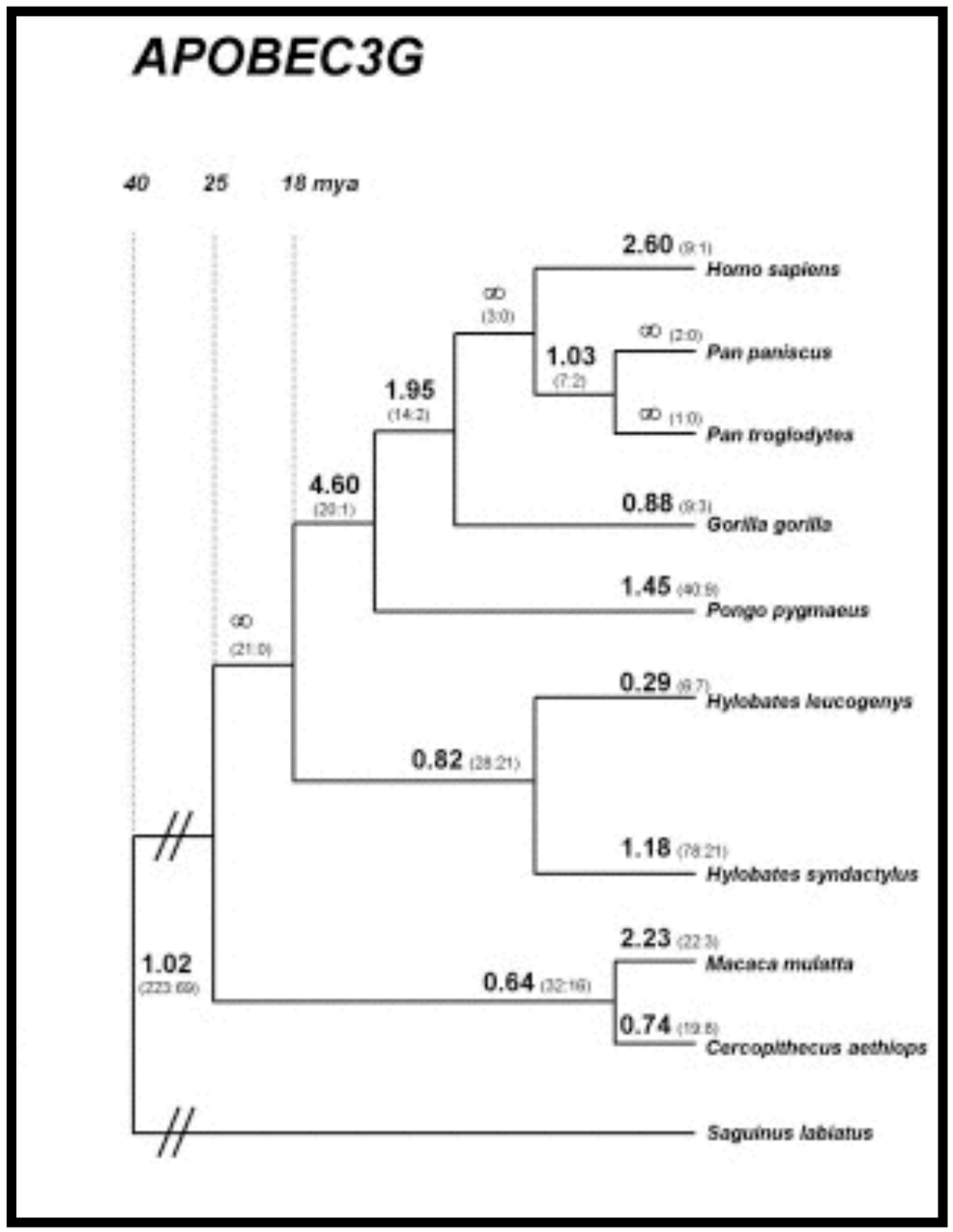
dN/dS (PAML)

Maximum likelihood approach

Along the full sequence

Per lineage

Across all lineages



Ortiz *et al.,* Retrovirology 2006

dN/dS (PAML)

Maximum likelihood approach

Along the full sequence

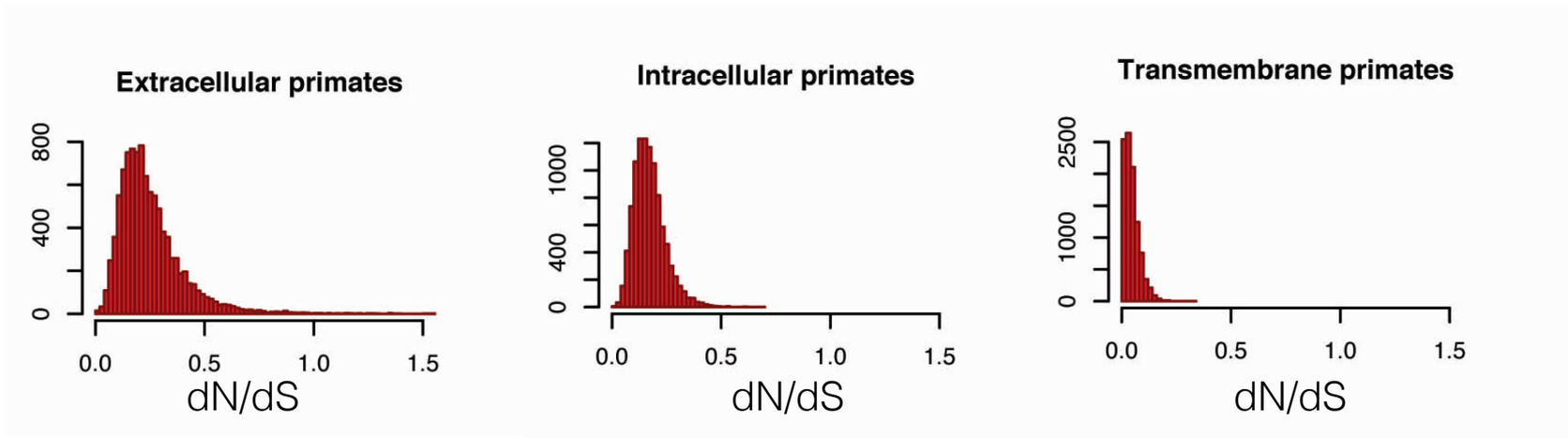
Per lineage

Across all lineages

Per codon

In particular protein sections (domains)

Freq



Andrés et al., MBE 2007

dN/dS (PAML)

Maximum likelihood approach

Along the full sequence

Per lineage

Across all lineages

Per codon

In particular protein sections (domains)

Per codon

Per codon and lineage

Calculate the likelihood of models with positive selection in particular lineages/codons, and identify putatively selected codons

McDonald & Kreitman (MK)

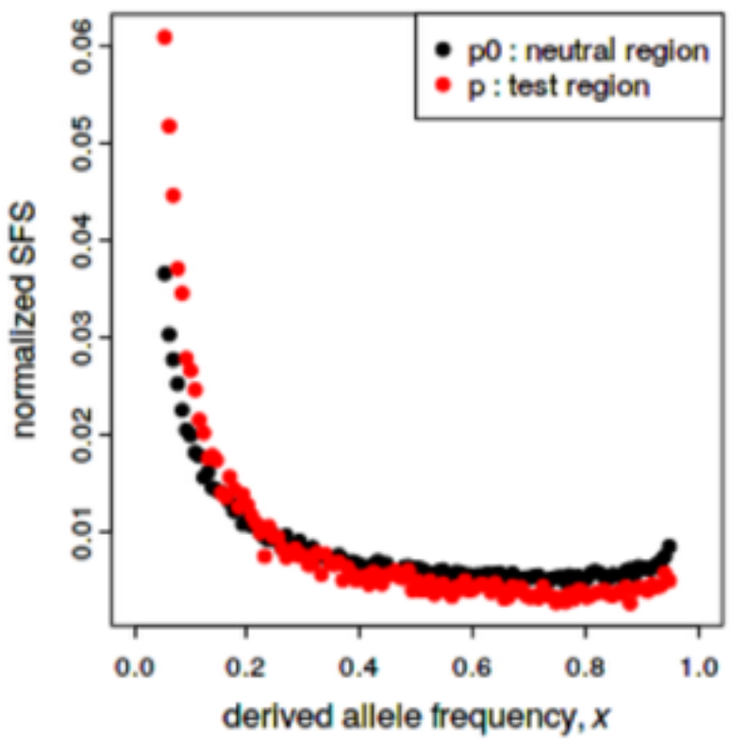
	Fixed	Polymorphic
Synonymous	D_s	P_s
Nonsynonymous	D_n	P_n

McDonald and Kreitman, 1991

McDonald & Kreitman (MK)

Estimate the proportion of non-synonymous substitutions driven by positive selection (alpha).

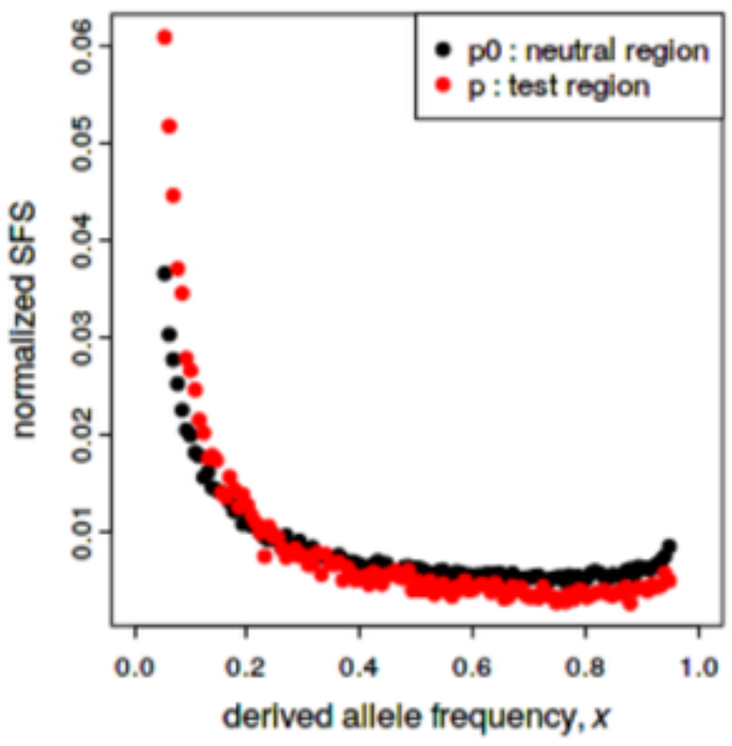
Remove rare alleles (slightly deleterious alleles)



McDonald & Kreitman (MK)

Estimate the proportion of non-synonymous substitutions driven by positive selection (alpha).

Use the full site frequency spectrum



McDonald & Kreitman (MK)

Estimate the proportion of non-synonymous substitutions driven by positive selection (alpha).

Use the full site frequency spectrum

DFE-alpha (simultaneous estimate of the DFE of new mutations at selected sites based on the SFS, and alpha)

Eyre-Walker and Keighley MBE 2009

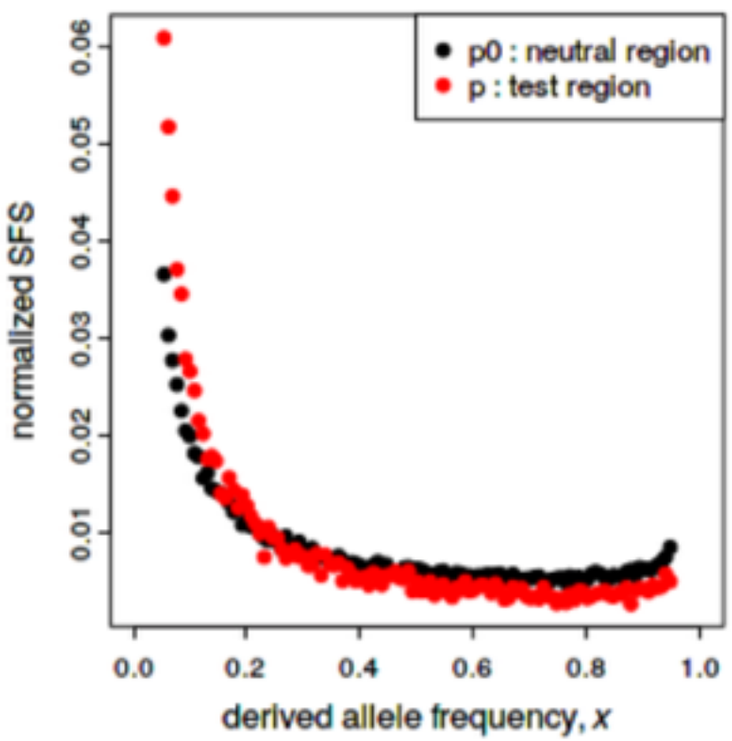
McDonald & Kreitman (MK)

Estimate the proportion of non-synonymous substitutions driven by positive selection (alpha).

Use the full site frequency spectrum

asymptoticMK

$$\alpha(x) = 1 - \frac{d_0}{d}\frac{p(x)}{p_0(x)}.$$



Messer and Petrov, PNAS 2013

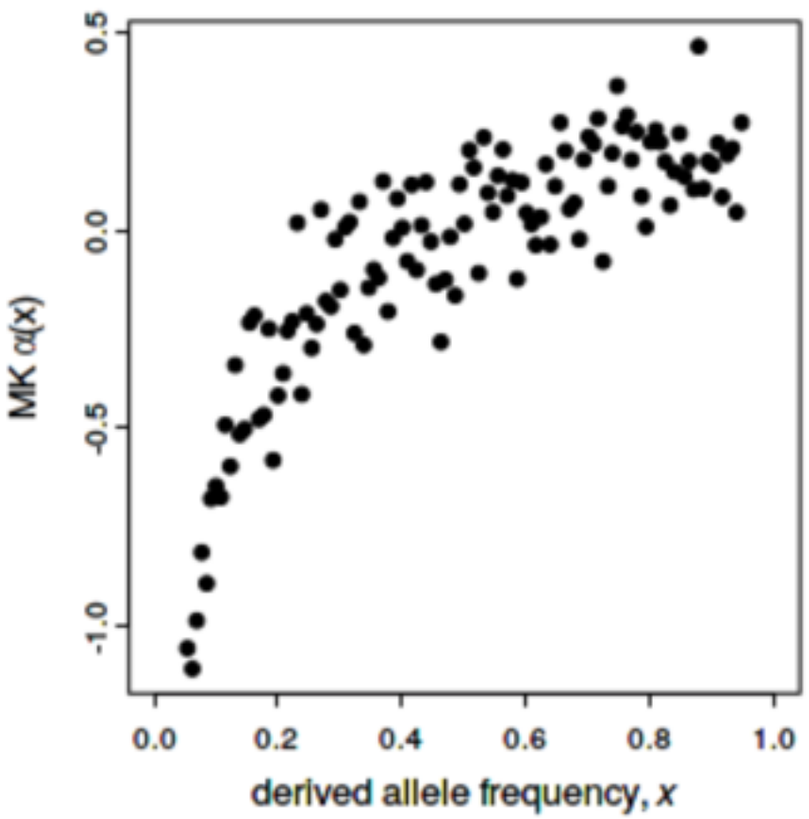
McDonald & Kreitman (MK)

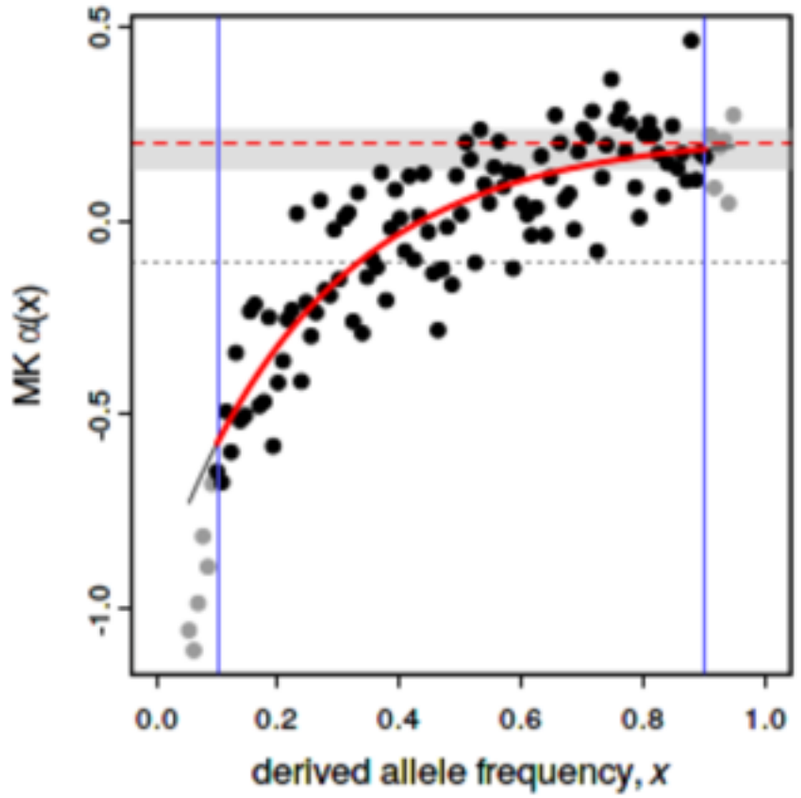
Estimate the proportion of non-synonymous substitutions driven by positive selection (alpha).

Use the full site frequency spectrum

asymptoticMK

$$\alpha(x) = 1 - \frac{d_0}{d}\frac{p(x)}{p_0(x)}.$$





Messer and Petrov, PNAS 2013

asymptoticMK

asymptoticMK: Asymptotic McDonald–Kreitman Test

See below for background and usage information. If you use this service, please cite our paper:

[not yet published, please check back for a citation...]

Submit your data:

d :

d_0 :

Input file : Choose File SFSasympMK_...an_no1.txt

(Tab-delimited with named columns for x, p, and p_0) *[sample]*

x interval to fit : [0.10 , 0.90]

Submit

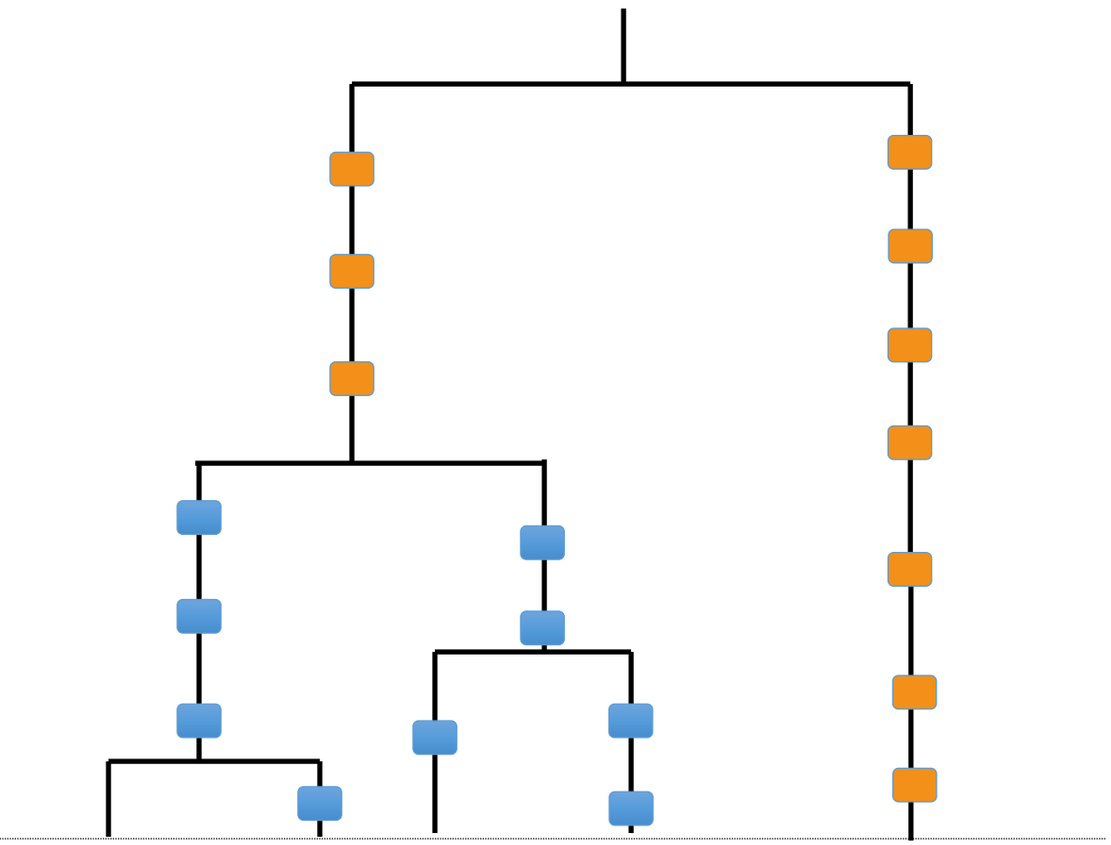
HKA (Hudson, Kreitman and Aguadé)

Locus A	#Polym	#Subst
Locus B	#Polym	#Subst

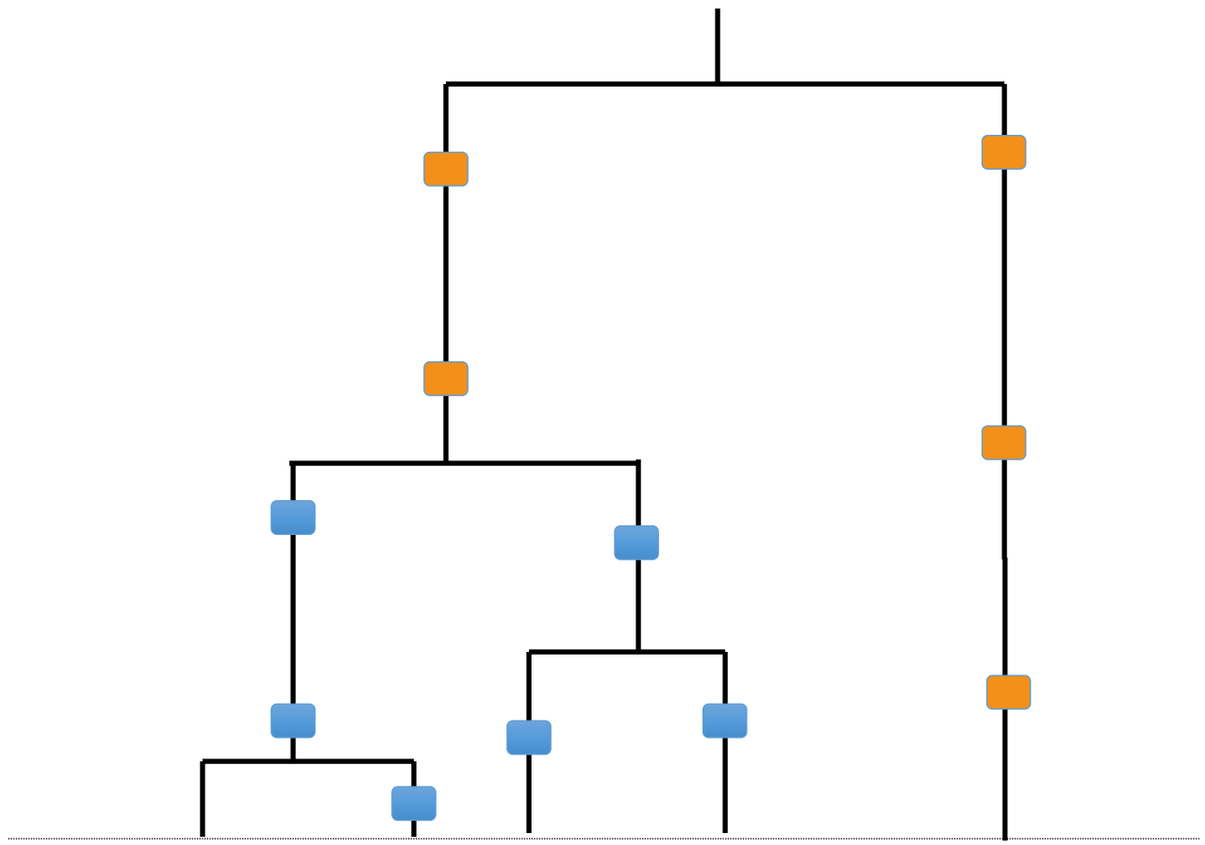
Hudson, Kreitman and Aguadé, Genetics 1987

HKA (Hudson, Kreitman and Aguadé)

Neutral



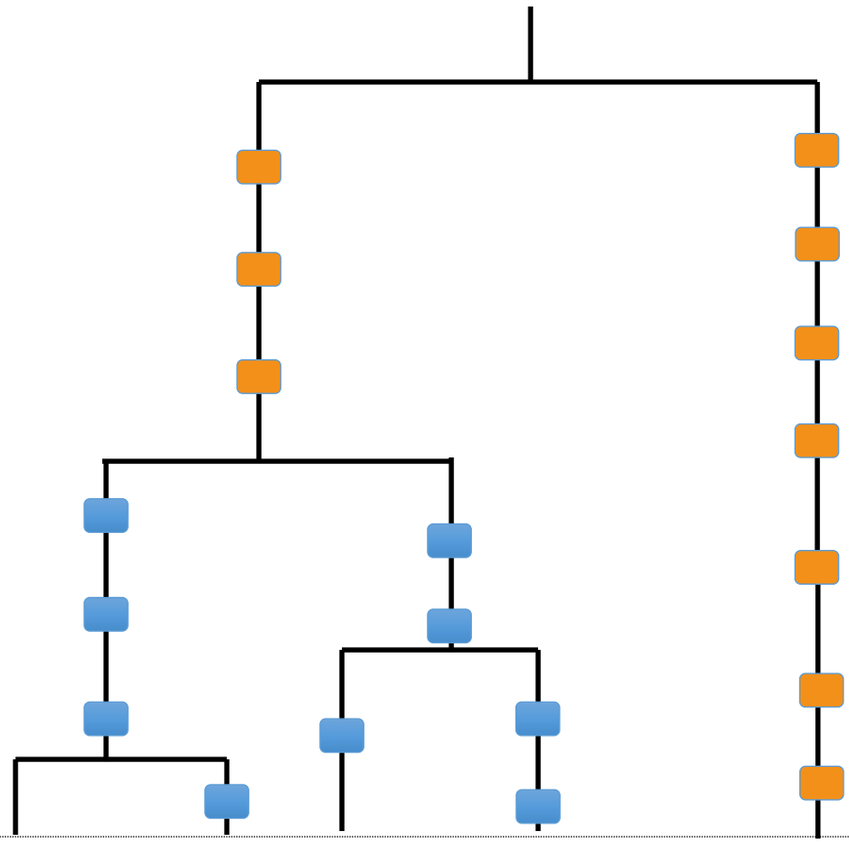
SNPs/FDs



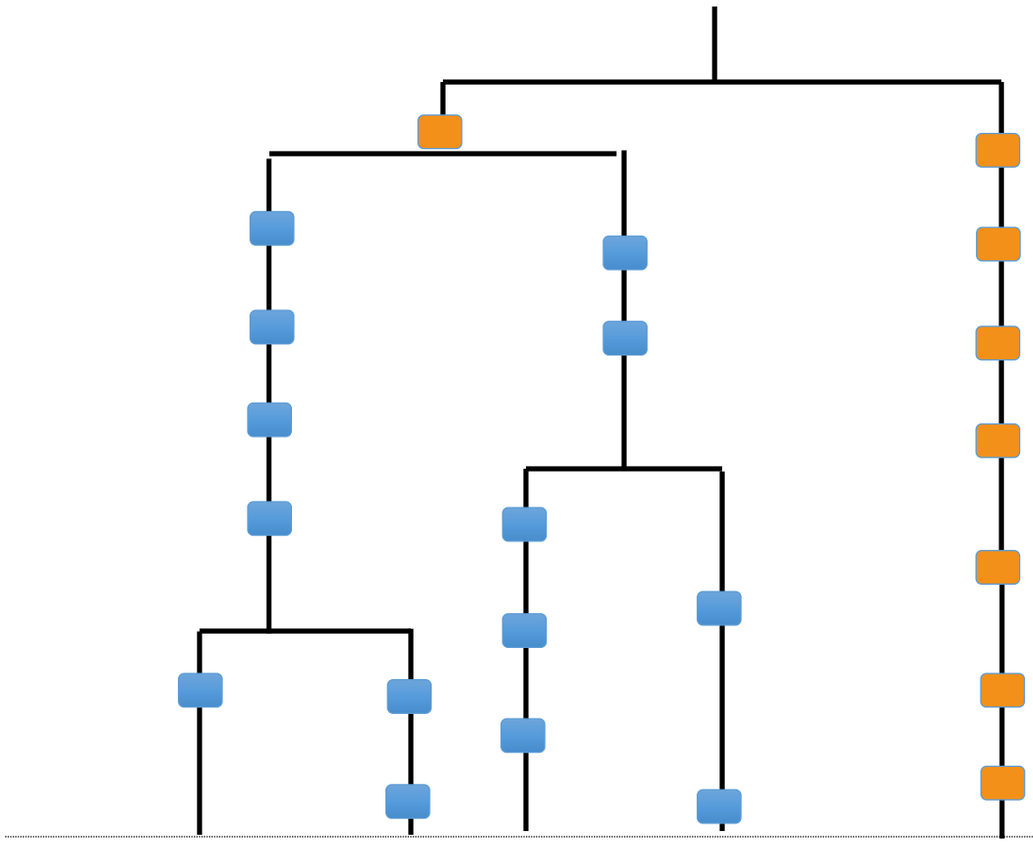
SNPs/FDs

HKA (Hudson, Kreitman and Aguadé)

Neutral



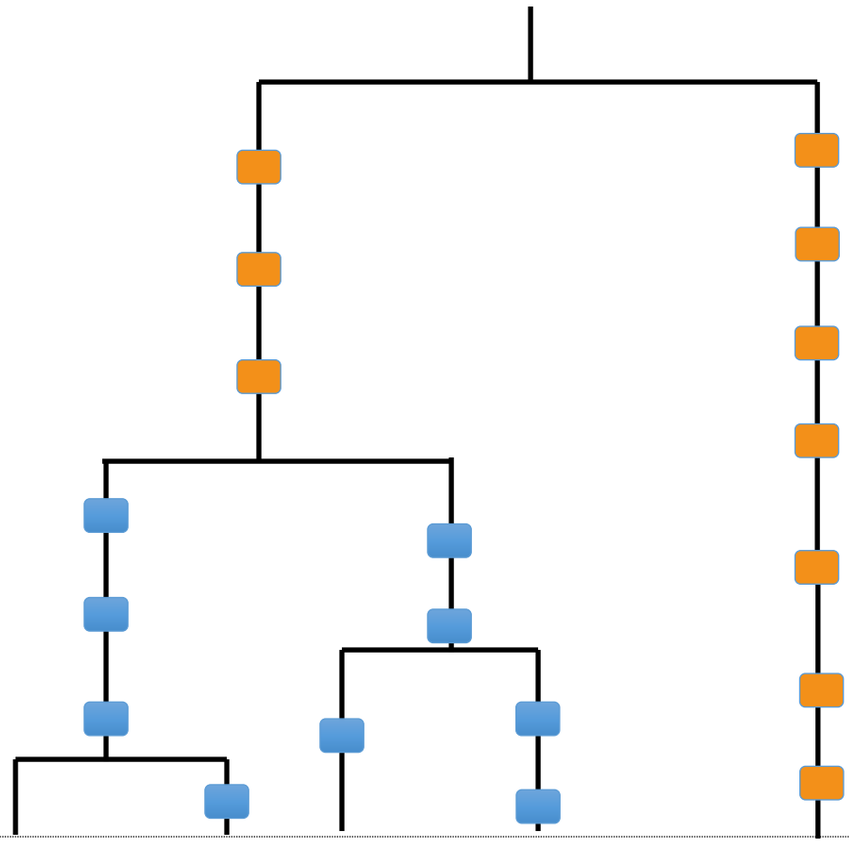
Balancing selection



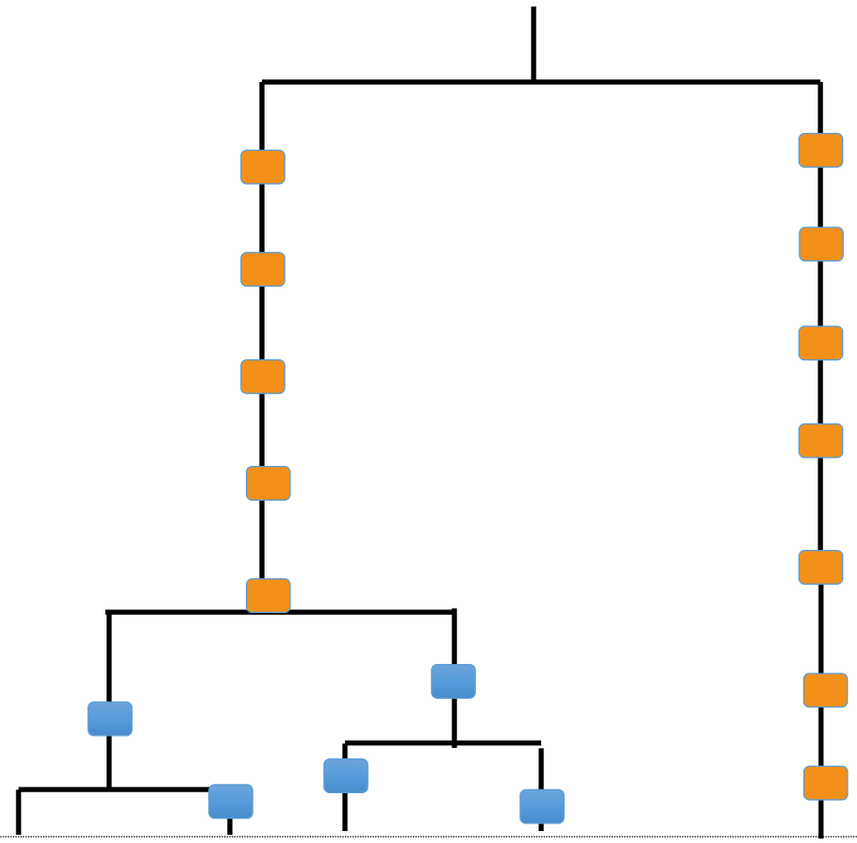
SNPs/FDs < SNPs/FDs

HKA (Hudson, Kreitman and Aguadé)

Neutral



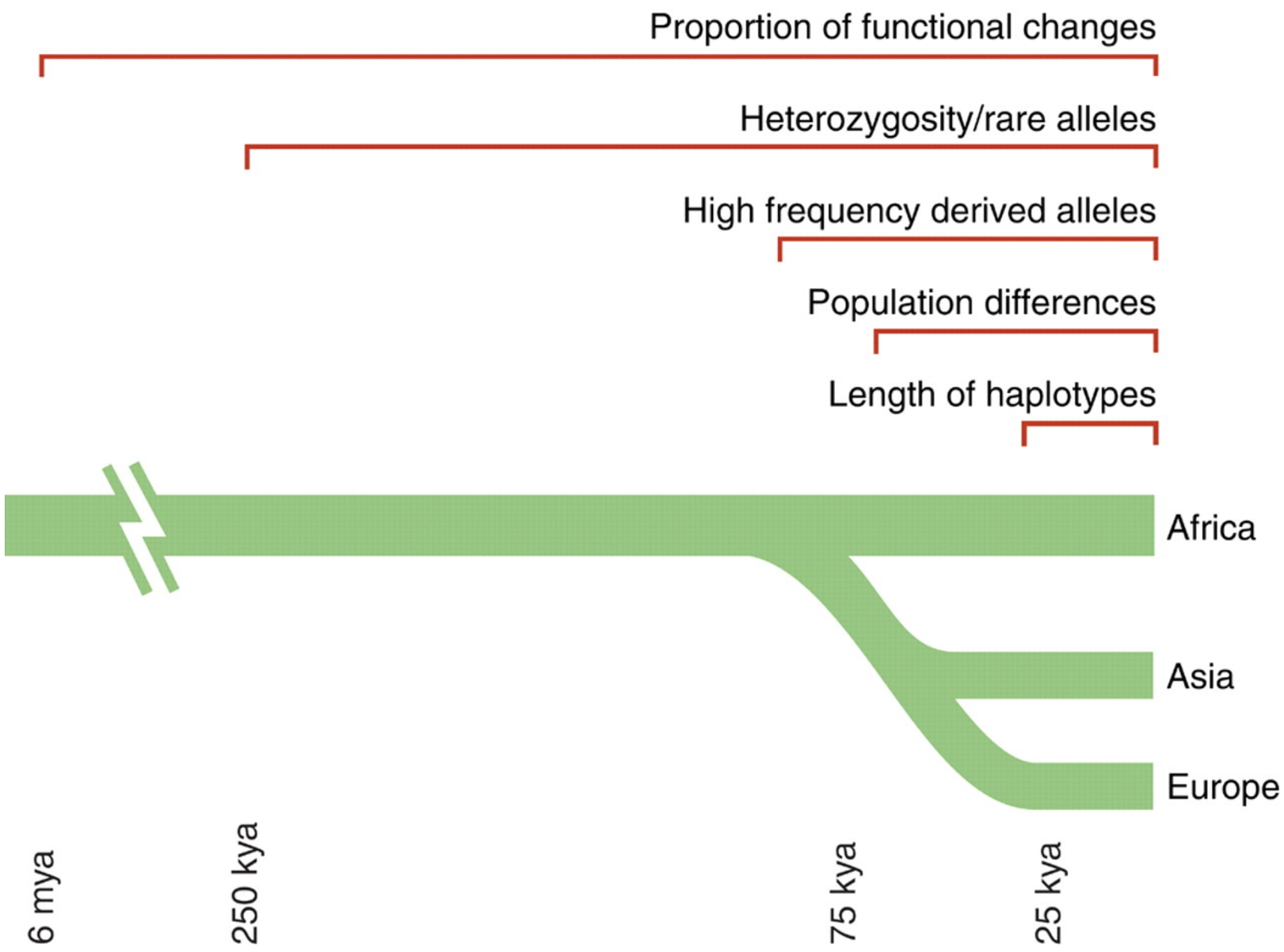
Positive selection



SNPs/FDs > SNPs/FDs

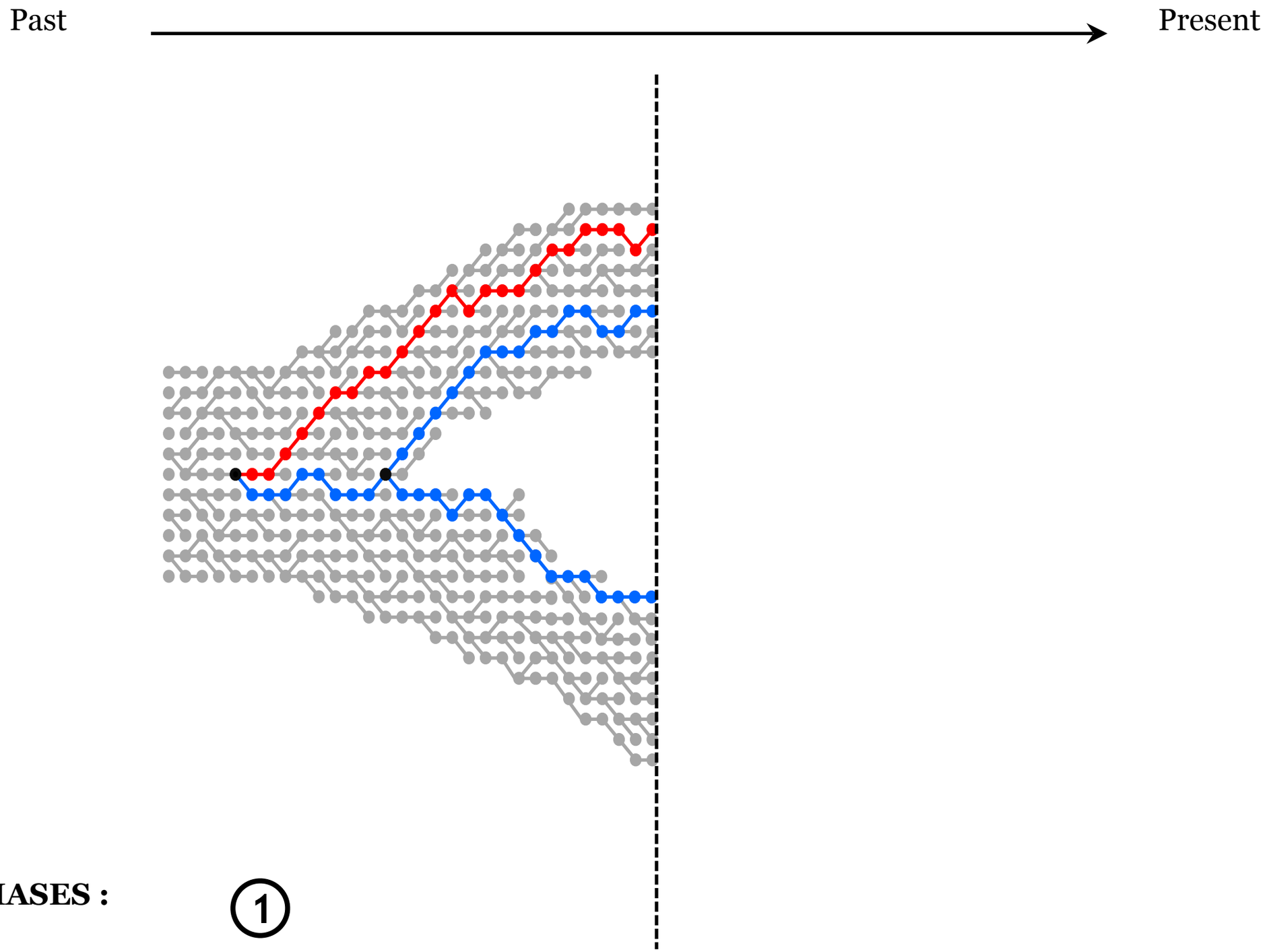
Extended lineage sorting (ELS)

Time Scales for the Signatures of Selection



Sabeti et al., Science, 2006

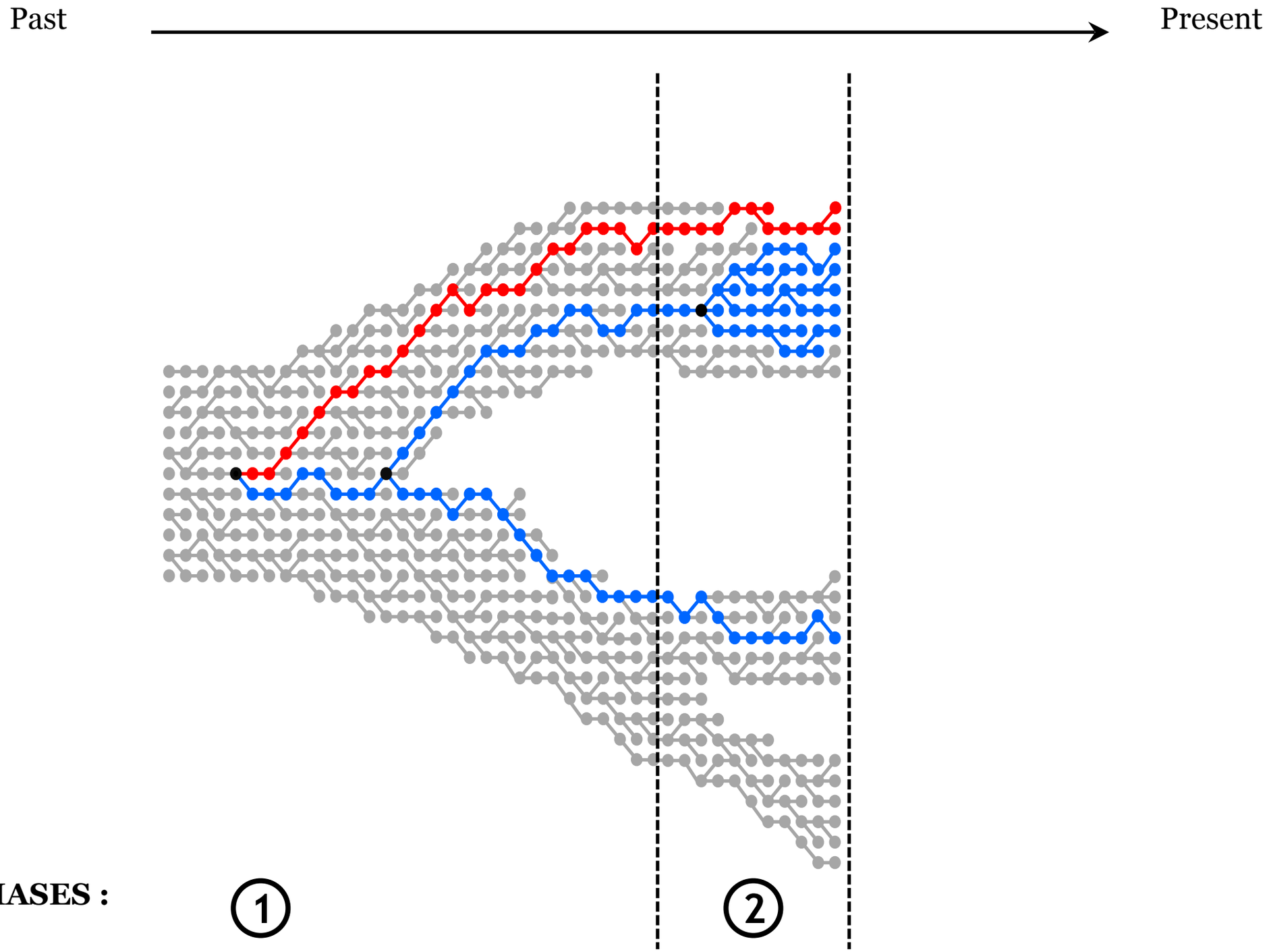
Definition: Lineage Sorting



PHASES : ①

- Random Distribution of Lineages

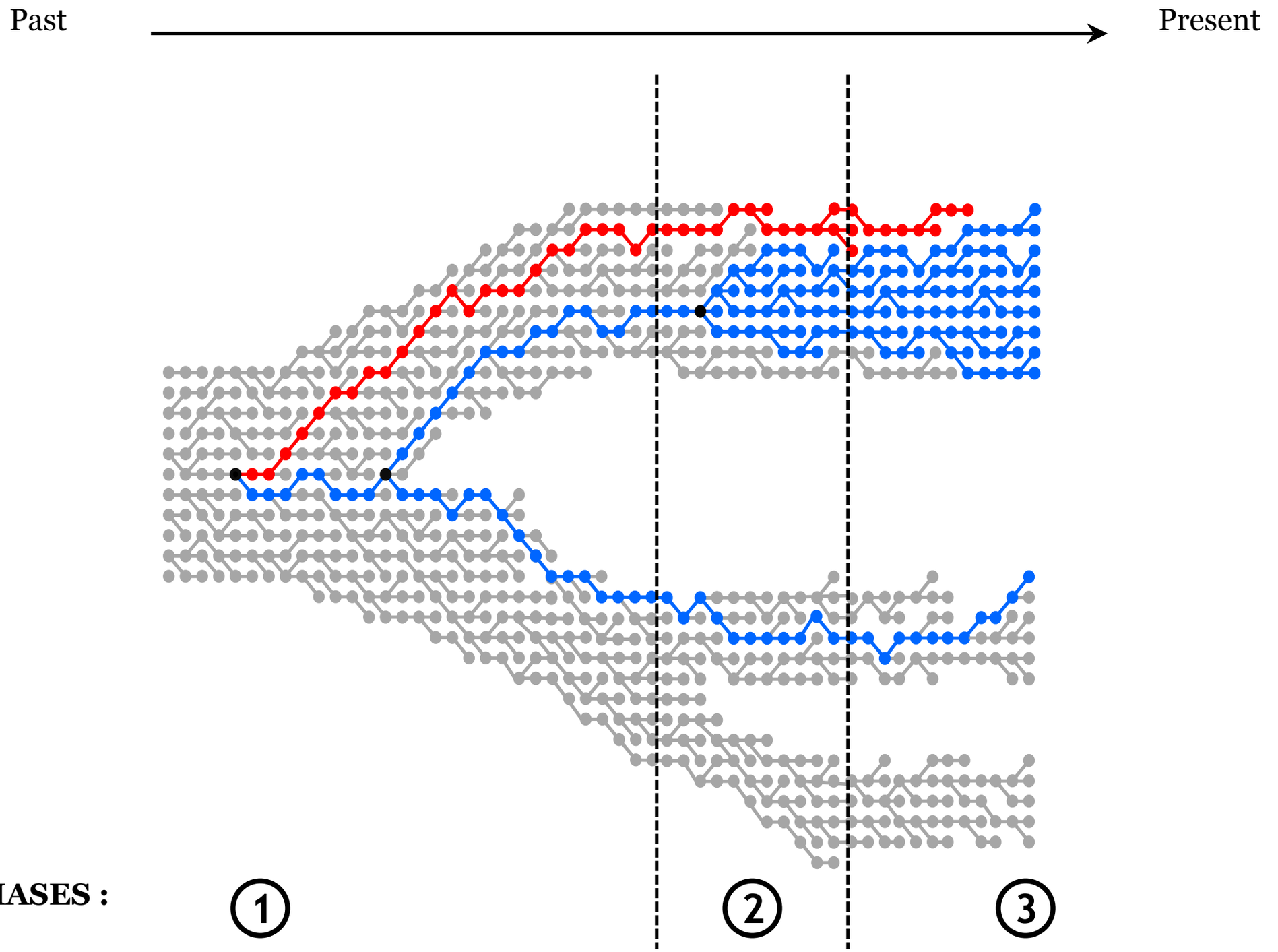
Definition: Lineage Sorting



PHASES :

- Random Distribution of Lineages
- Incomplete Lineage Sorting

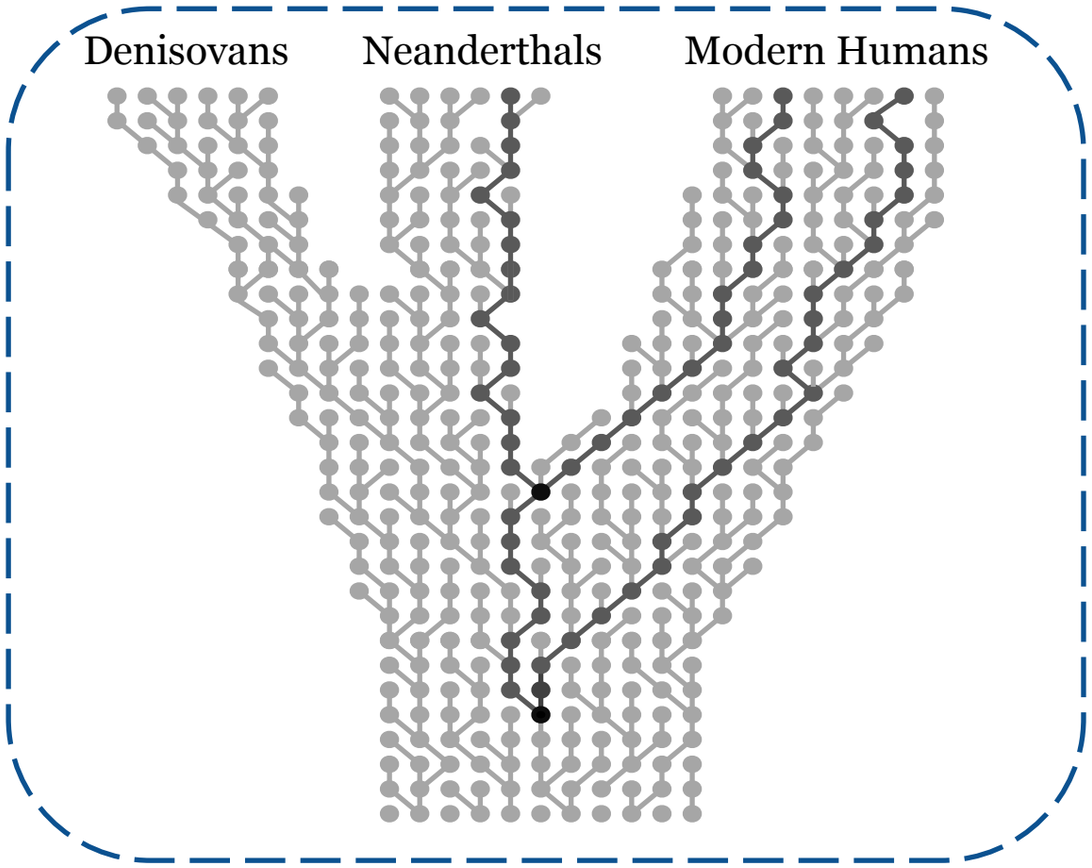
Definition: Lineage Sorting



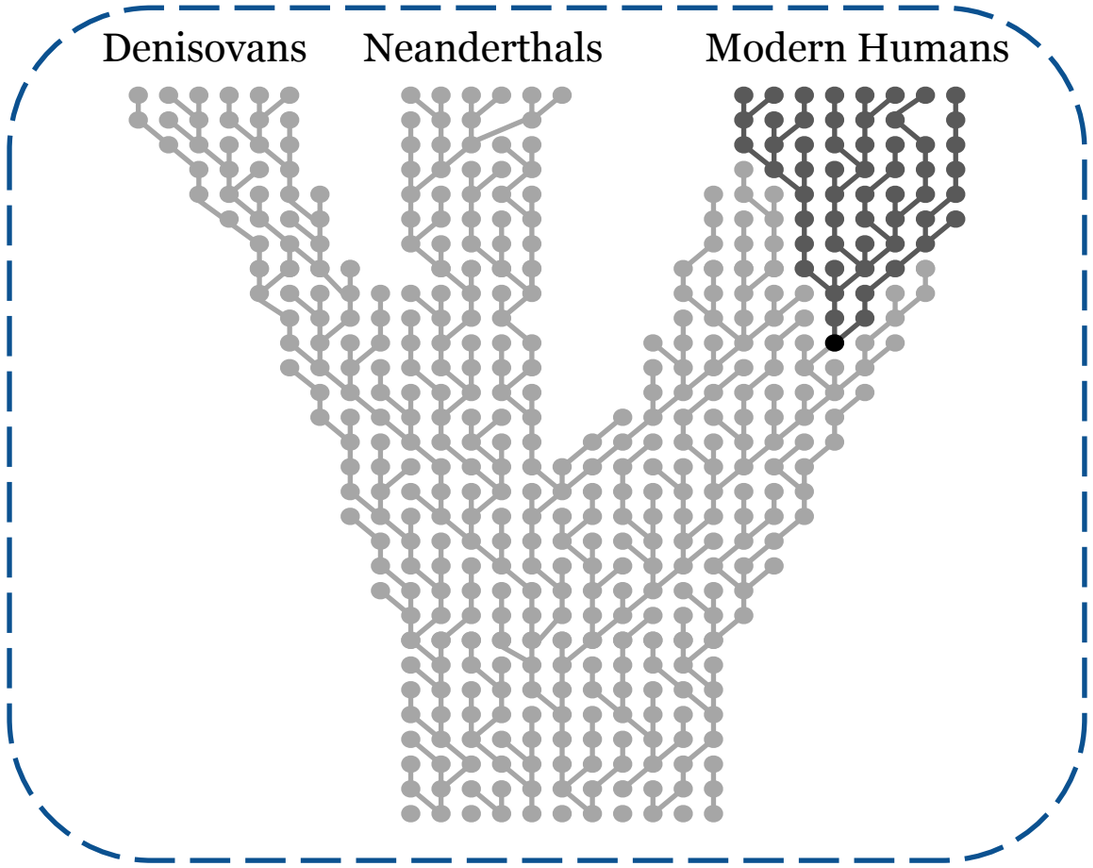
PHASES :

- Random Distribution of Lineages
- Incomplete Lineage Sorting
- Lineage Sorting

Changes in Local Genealogies along the Genome

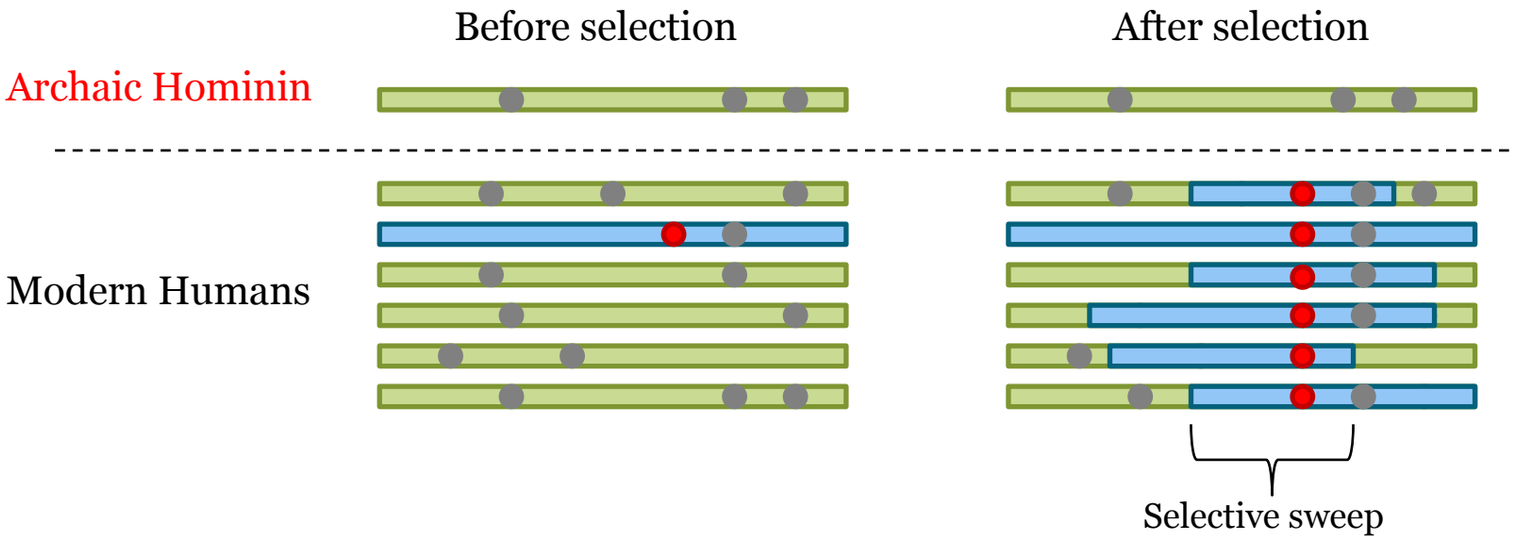


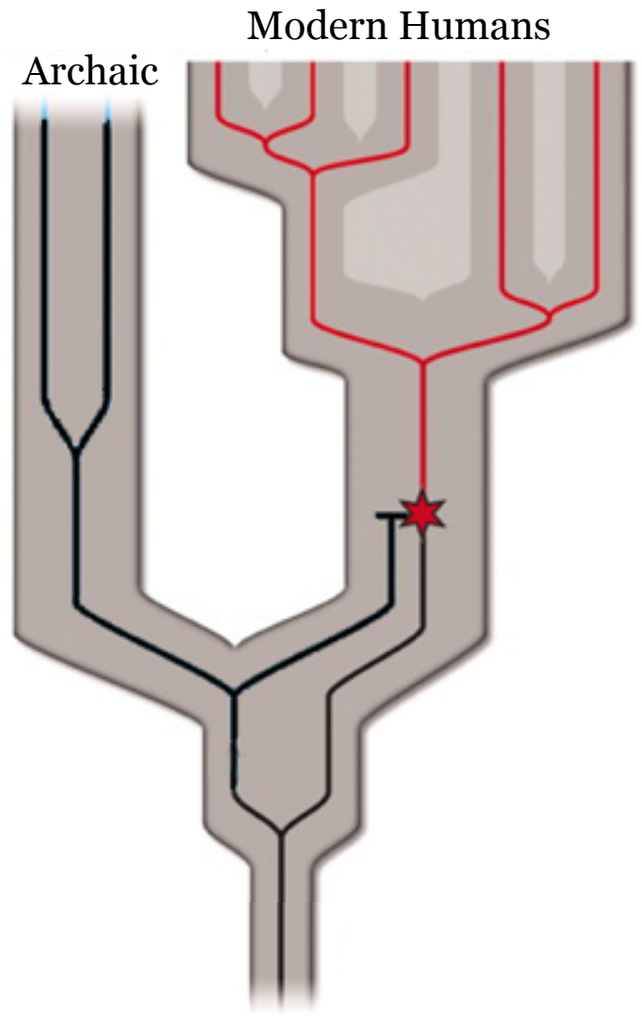
Internal regions (~90%)



External regions (~10%)

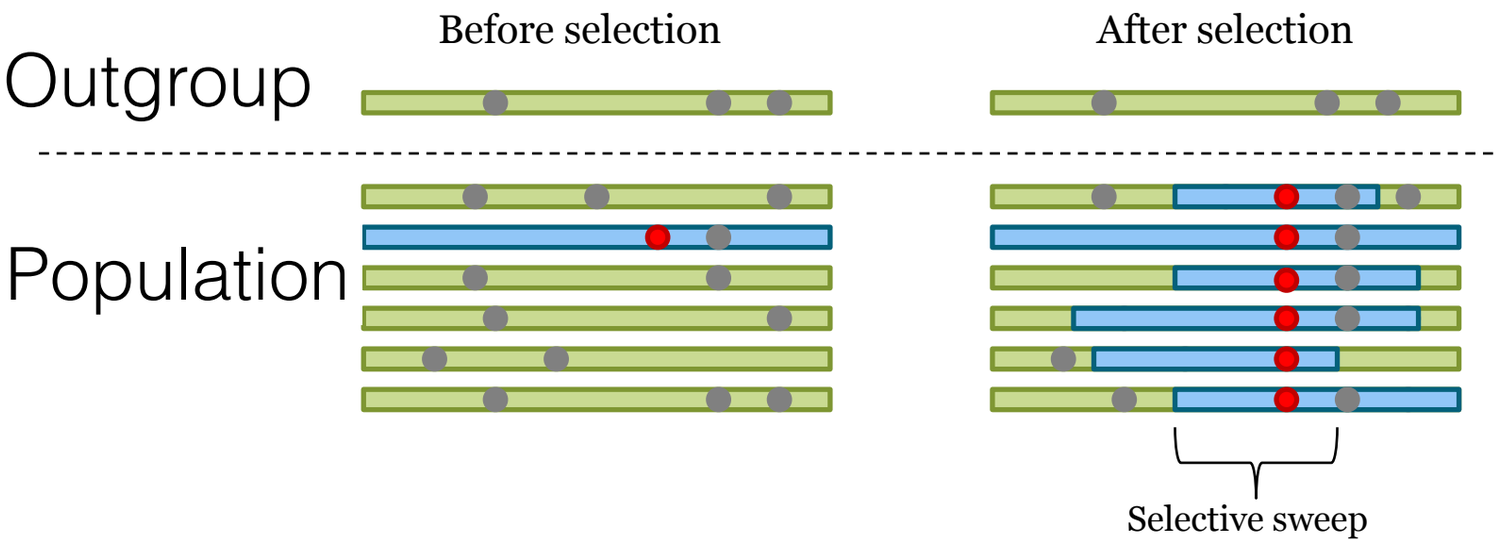
Signal of Positive Selection

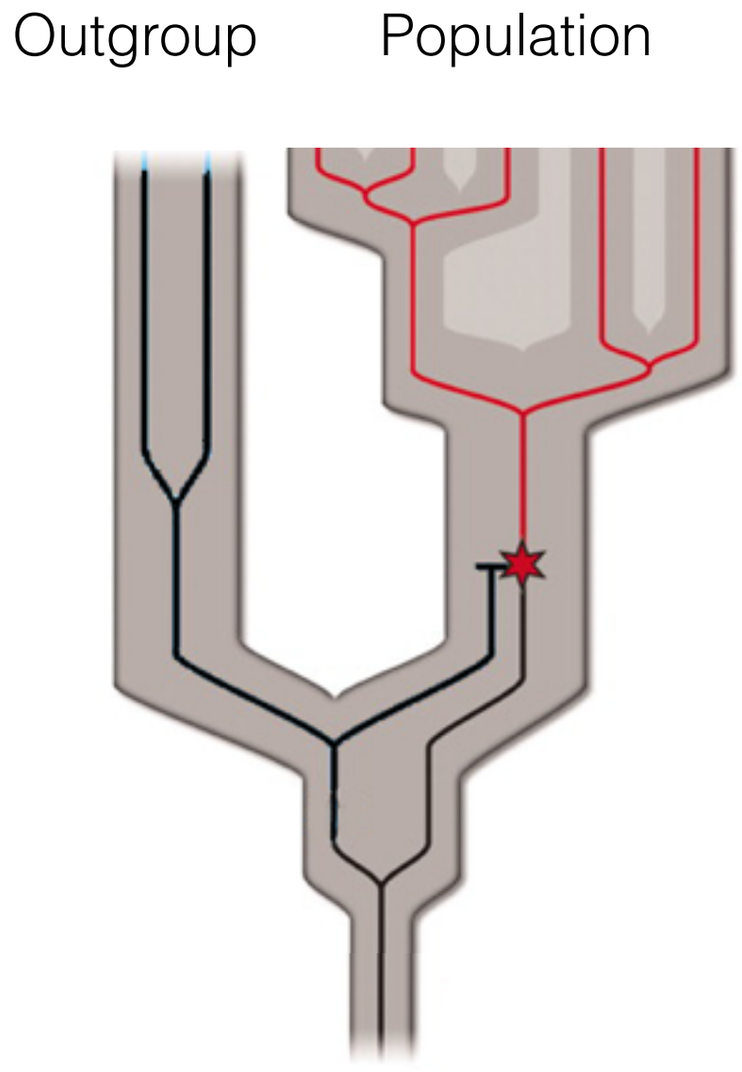




Adapted from Green et al., Science, 2010

Signal of Positive Selection



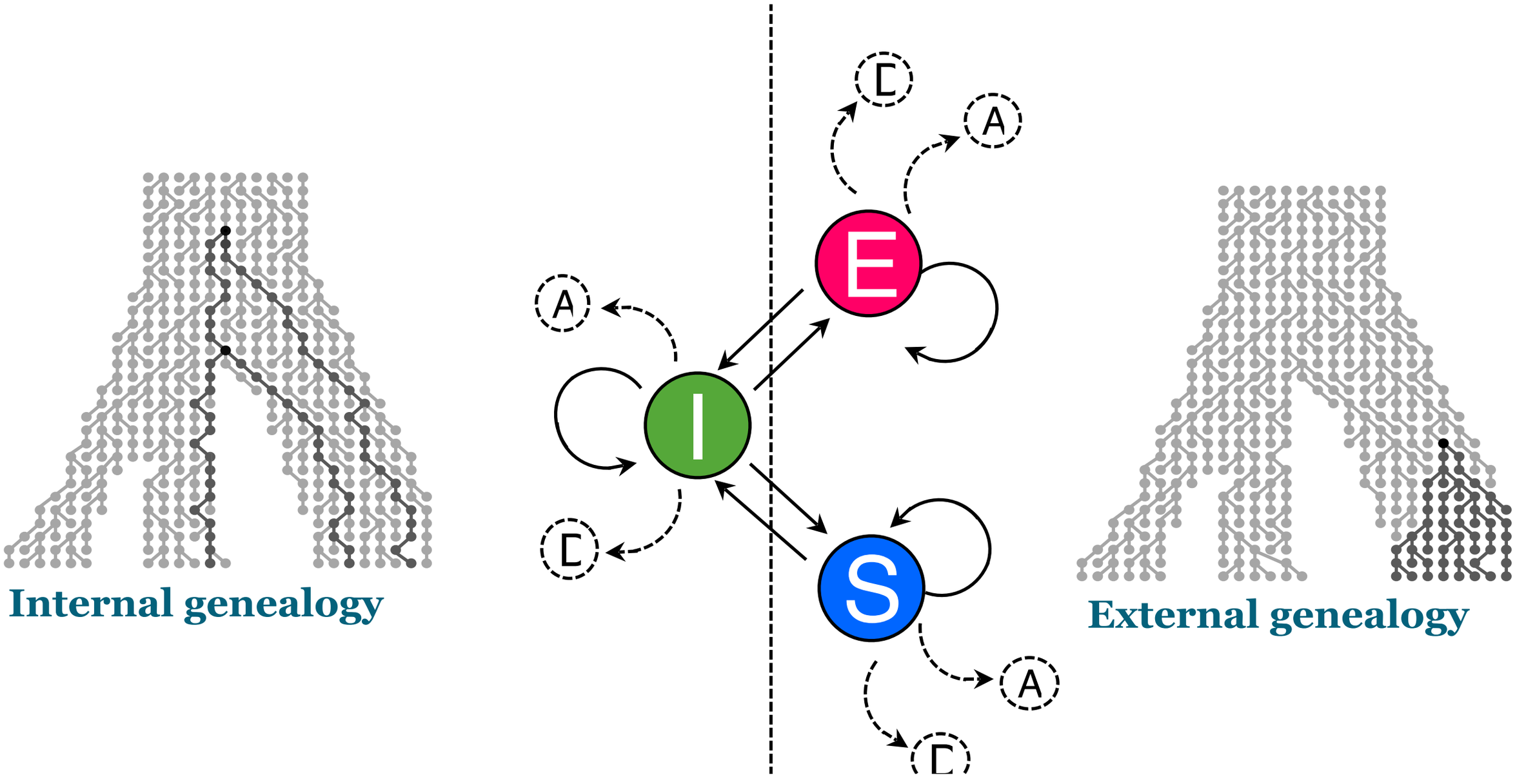


Can use several outgroups

Adapted from Green et al., Science, 2010

Cagan et al., MBE 2016

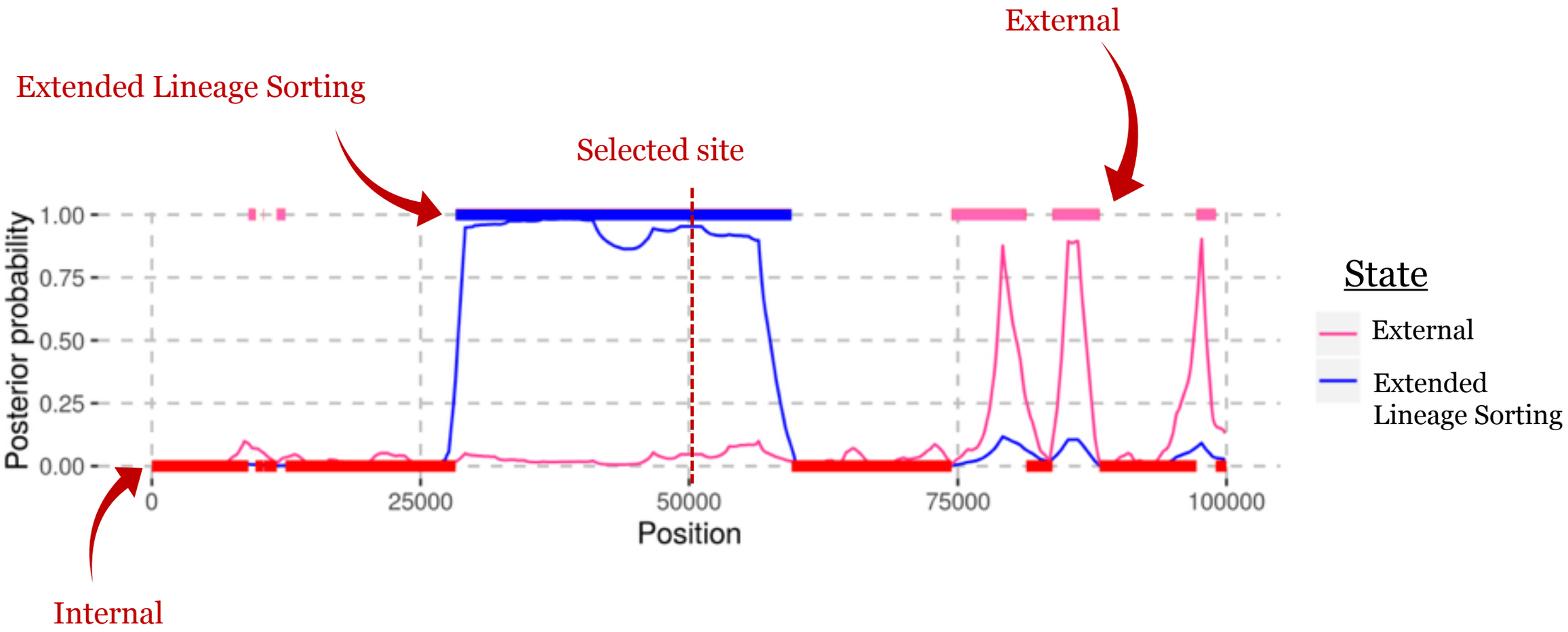
A Hidden Markov Model to Detect Extended Lineage Sorting



Estimate for each position the probability of each state

Peyregne et al., bioRxiv

Detection of Extended Lineage Sorting



Peyregne et al., bioRxiv

Extended lineage sorting (ELS)

Relevant parameters

Probability of the outgroup to share a derived allele in the population.

If external region, the probability is 0 (but error)

If internal region, the probability is 1 if the site if fixed in the population (but error)

If internal region, the probability for polymorphic sites depends on the age of the allele (frequency as a proxy)

Extended lineage sorting (ELS)

Relevant parameters

- Length of internal and external regions
- Length of ELS regions

Extended lineage sorting (ELS)

Estimate for each position the probability of each state (internal, external, ELS) and relevant parameters.

Can consider two states only (internal, external) in the absence of positive selection that creates ELS regions. This allows comparisons of the likelihood of the two models (neutral and with positive selection).

Thanks to

Stephane Peyregne

Gabriel Santpere

Joshua Schmidt

Philip Messer