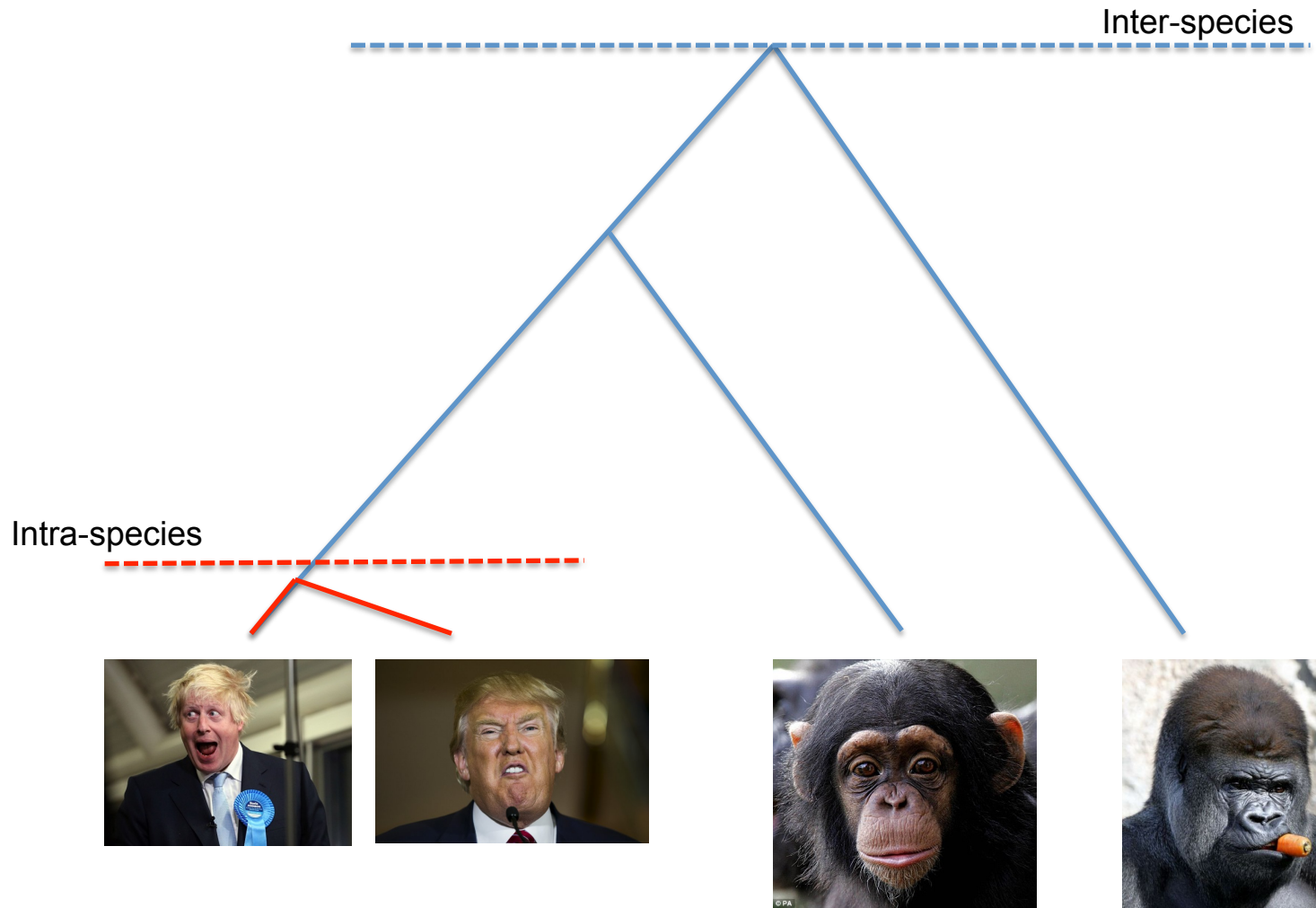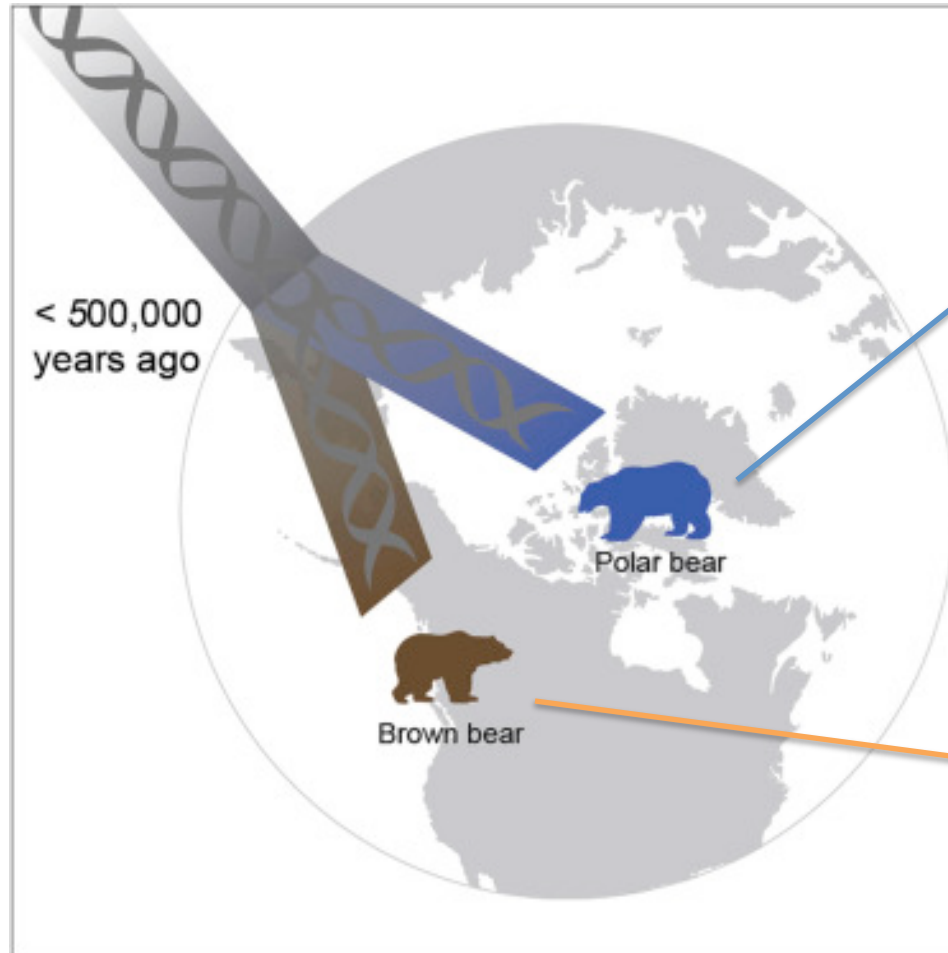# Detecting selection:
methods (II)
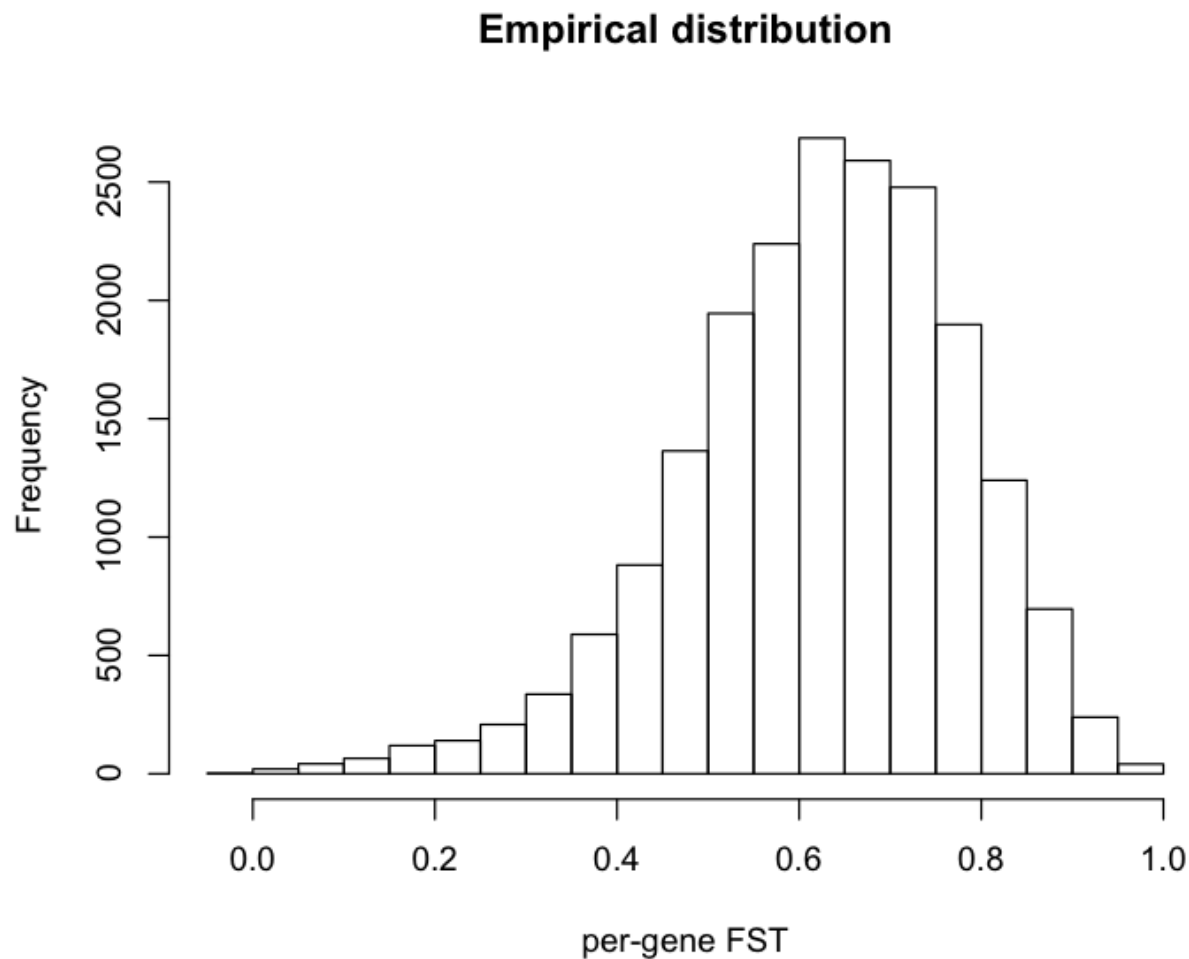
Matteo Fumagalli

Inter-species variation
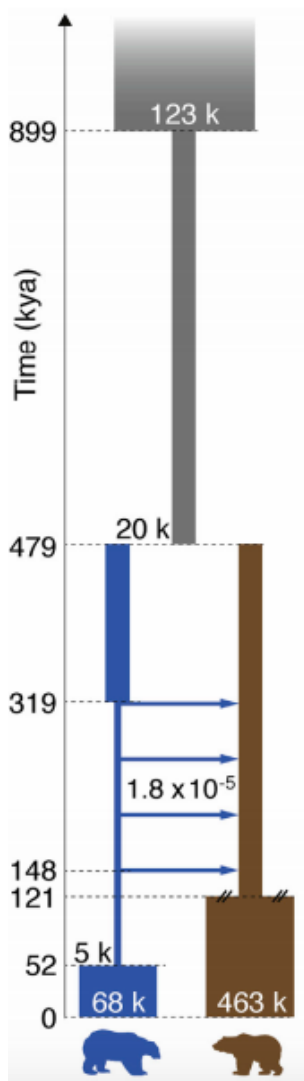

events in the deep past, macro-evolutionary trends, selection between species

Polar bears vs. Brown bears
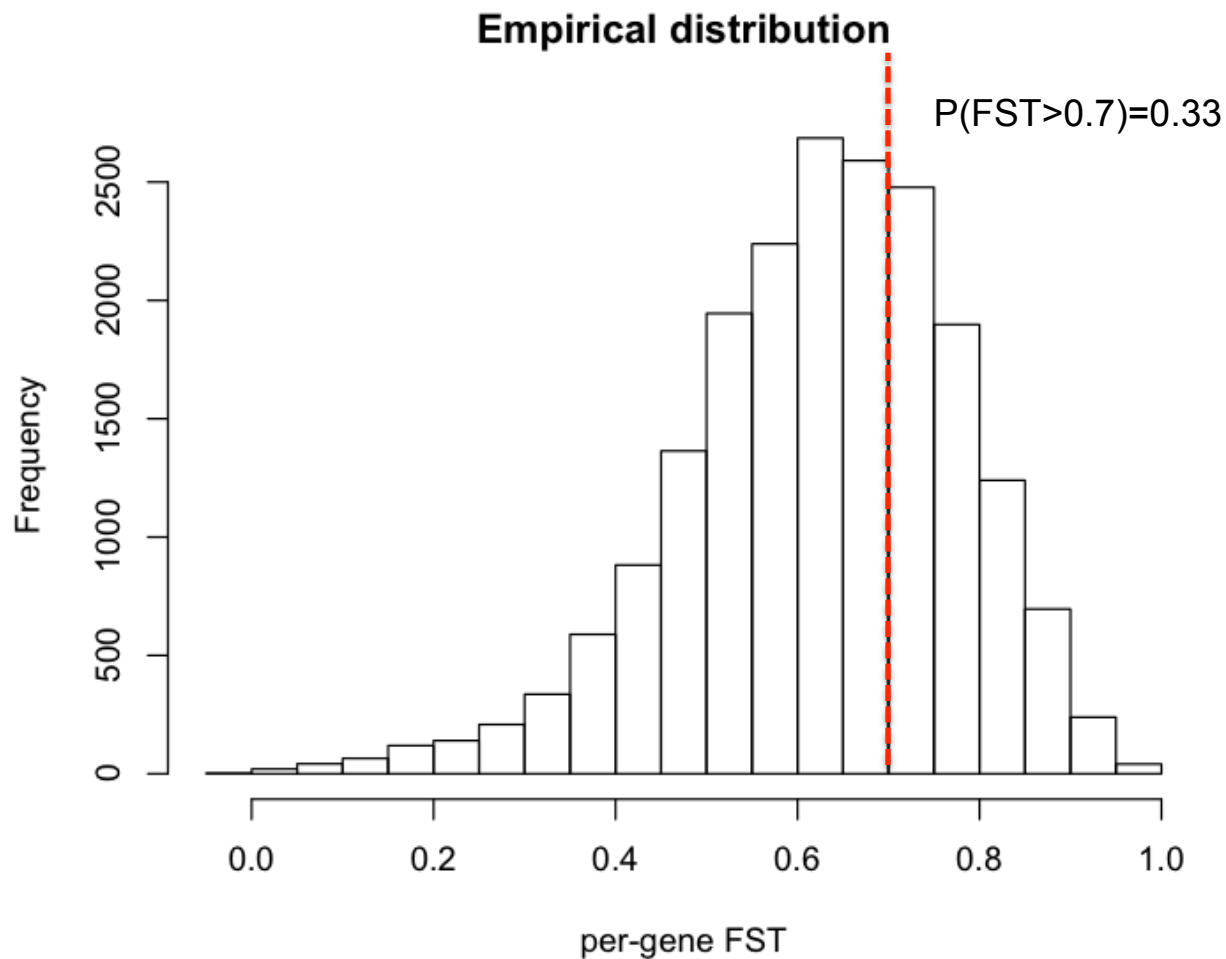

< 500,000 years ago
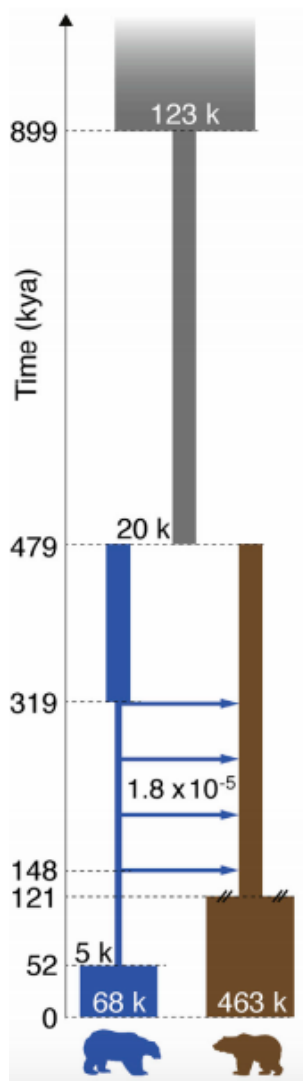
Polar bear

Brown bear

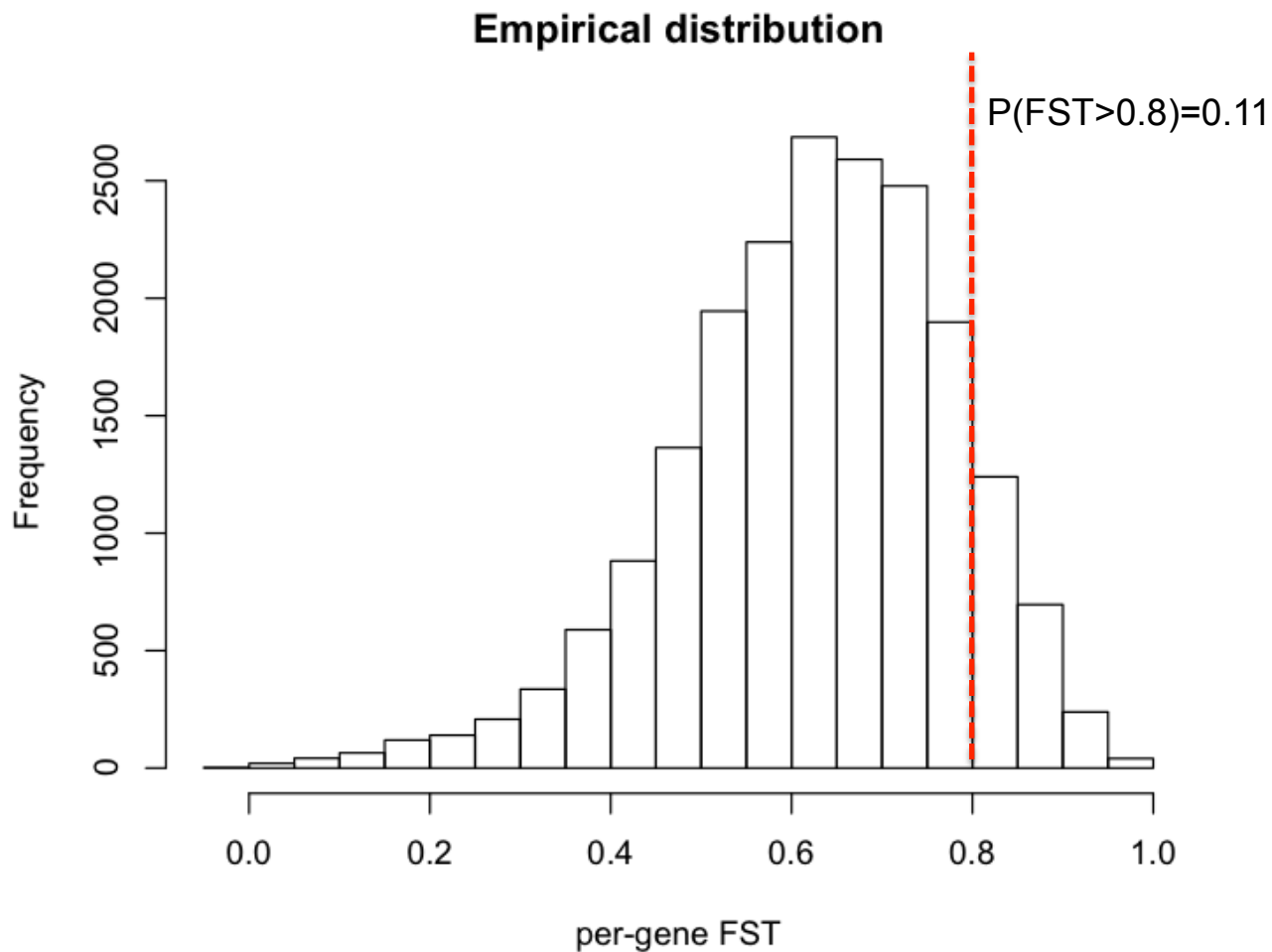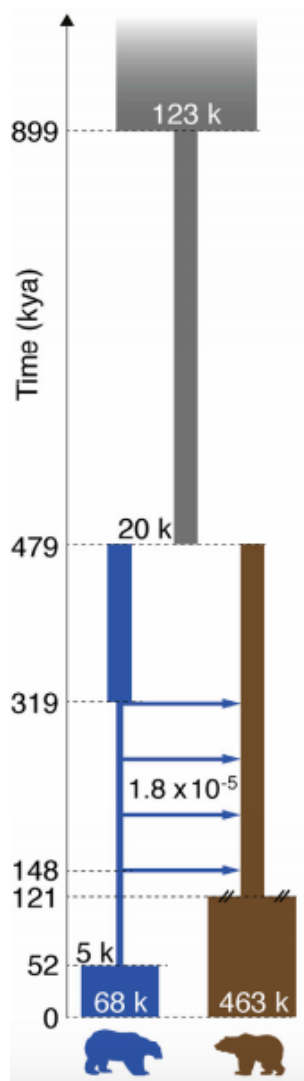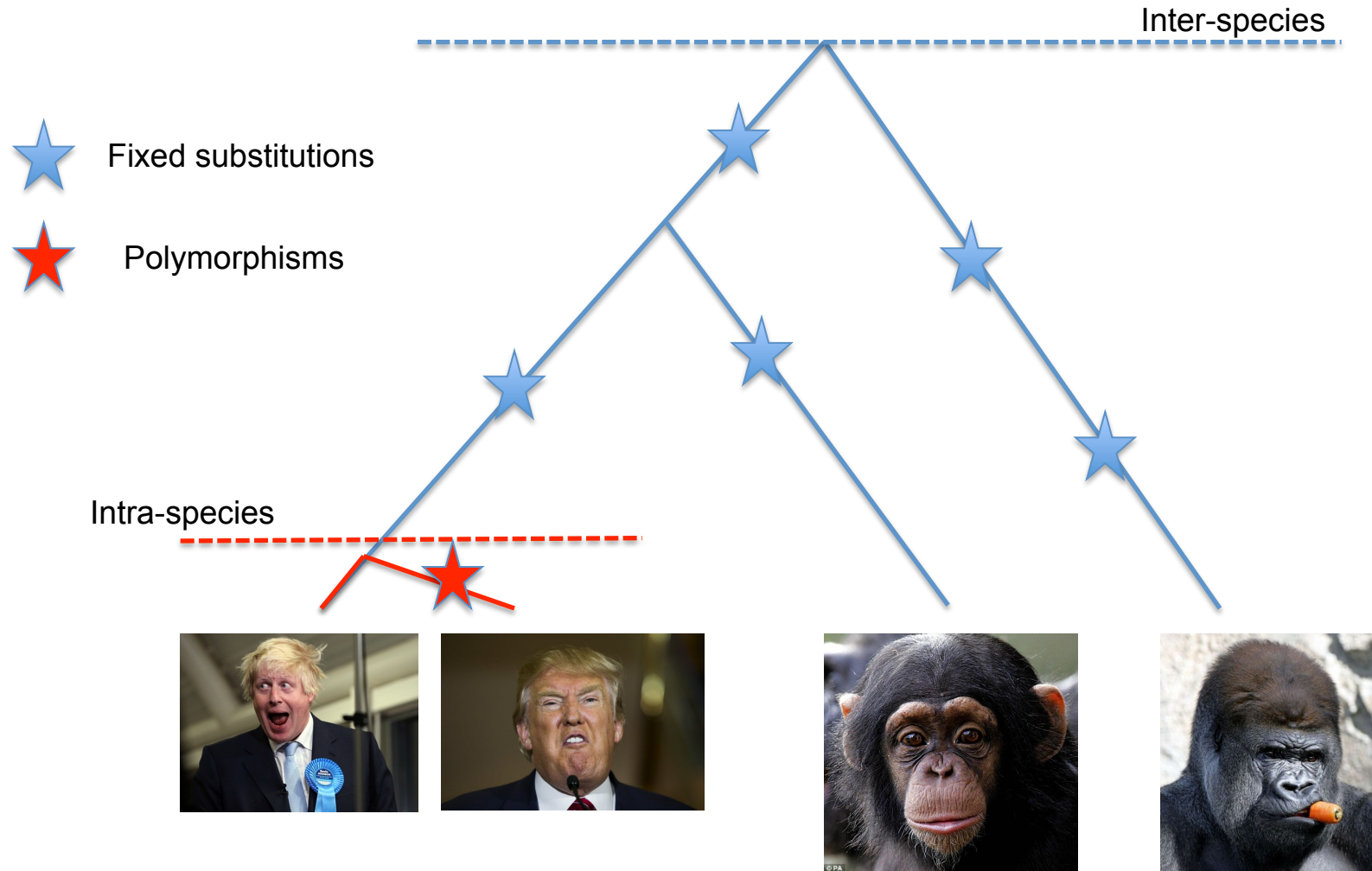Question: what are the genetic signatures of polar bears' adaptation to the Arctic environment?

# Expected genetic differentiation

Expected genetic differentiation

P(FST>0.7)=0.33

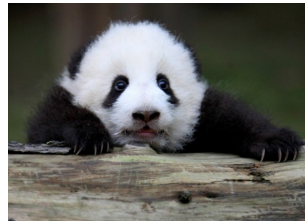# Expected genetic differentiation

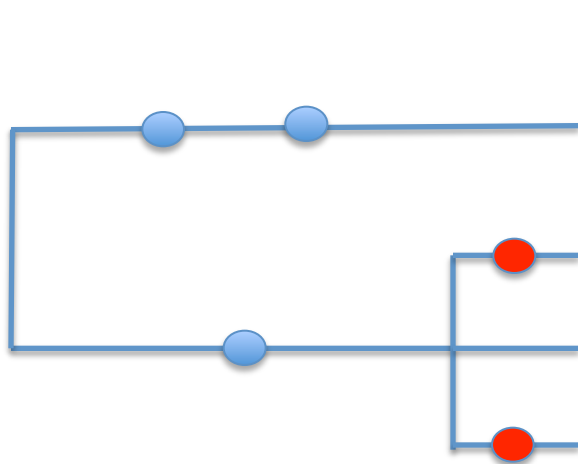Empirical distribution

P(FST>0.8)=0.11
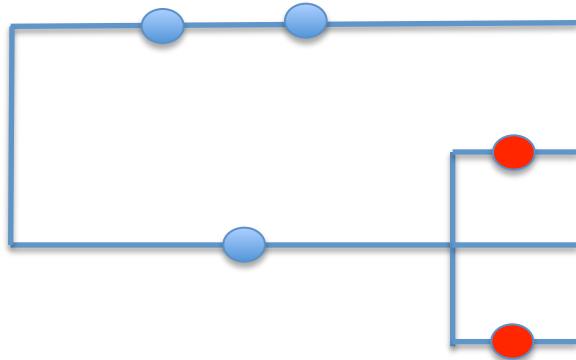
Inferring inter-species selection
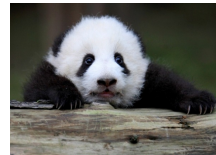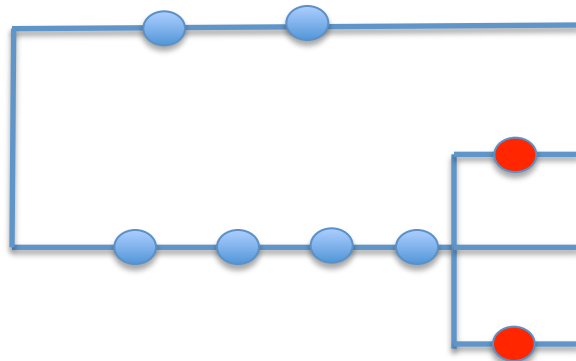
Polymorphisms and divergence

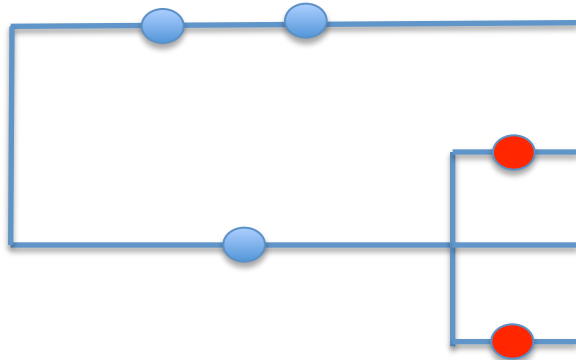As both depend on mutation rates, and levels are expected to be proportional.
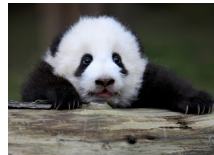
Polymorphisms and divergence

Under neutrality
(control locus or genome-wide)

Polymorphisms and divergence

Under neutrality
(control locus or genome-wide)

≈

??? selection

>

Polymorphisms and divergence

Under neutrality
(control locus or genome-wide)

$$\bigcirc \approx \bullet$$

Positive selection
(or negative selection)

$$\bigcirc > \bullet$$
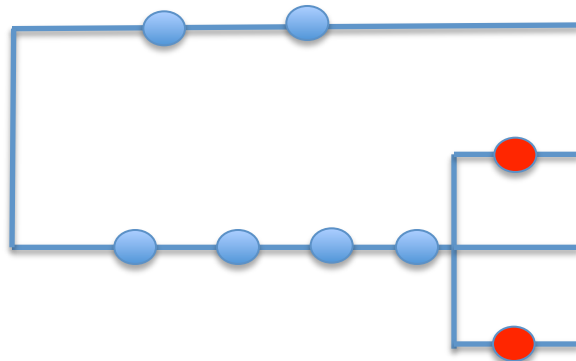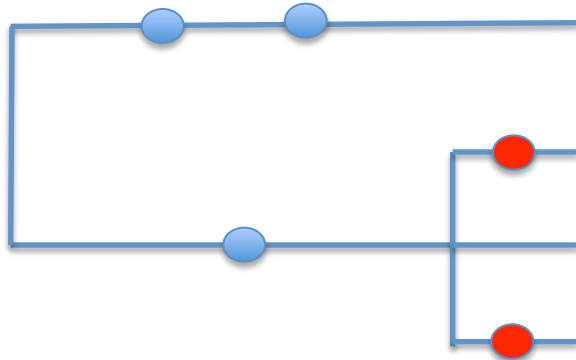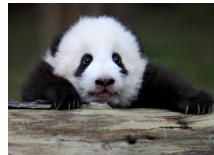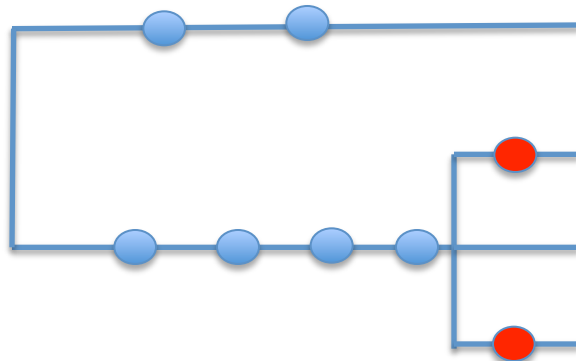
Polymorphisms and divergence

Under neutrality
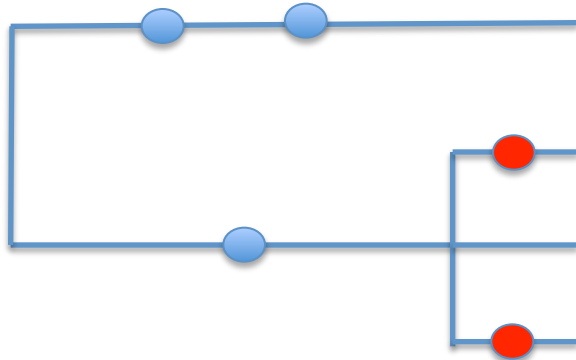(control locus or genome-wide)

Polymorphisms and divergence
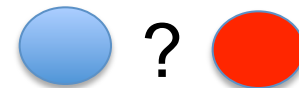

Under neutrality
(control locus or genome-wide)

🔵 ≈ 🔴

🔵 < 🔴

Polymorphisms and divergence

Under neutrality
(control locus or genome-wide)

🔵 ≈ 🔴

??? selection

🔵 < 🔴

Polymorphisms and divergence
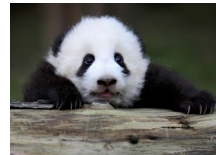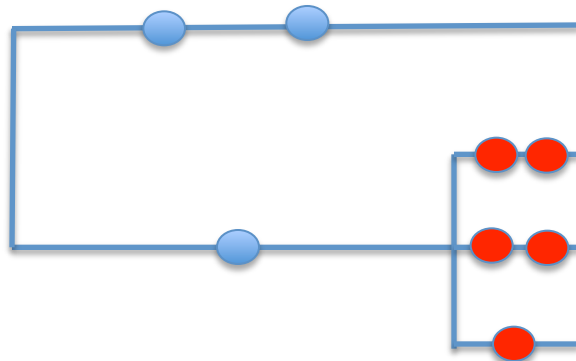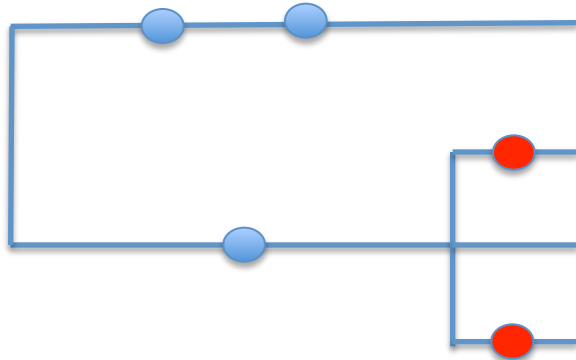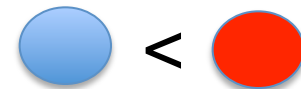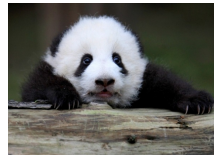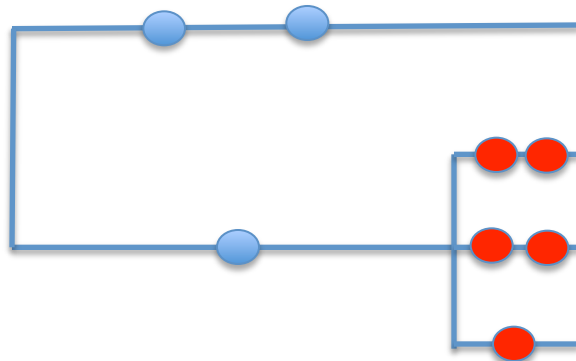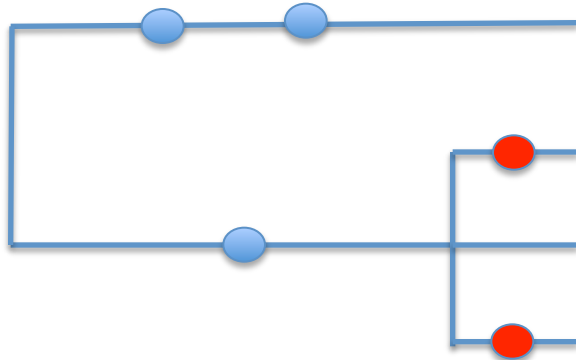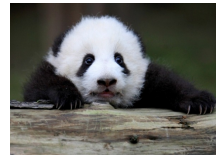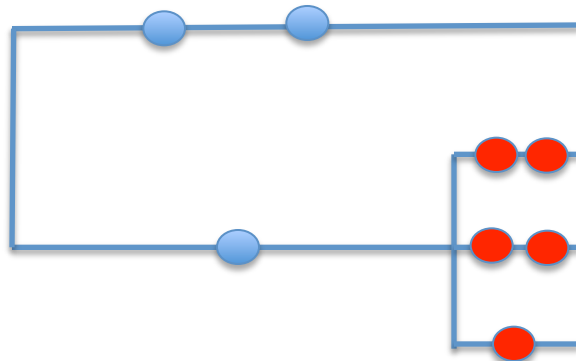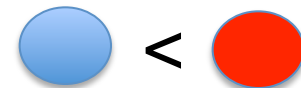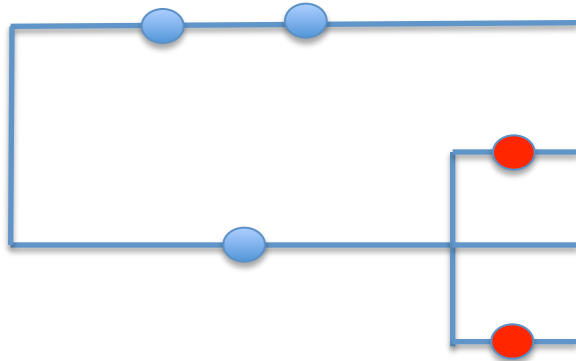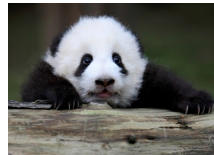

Under neutrality
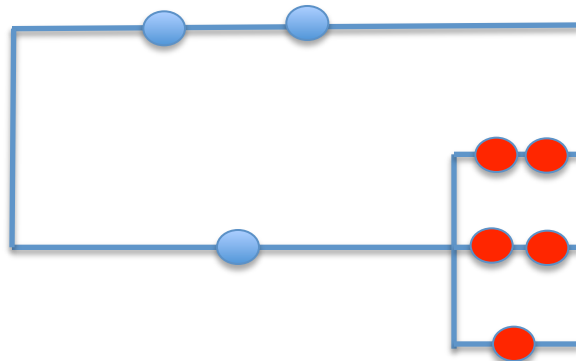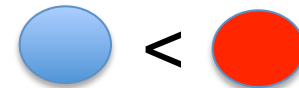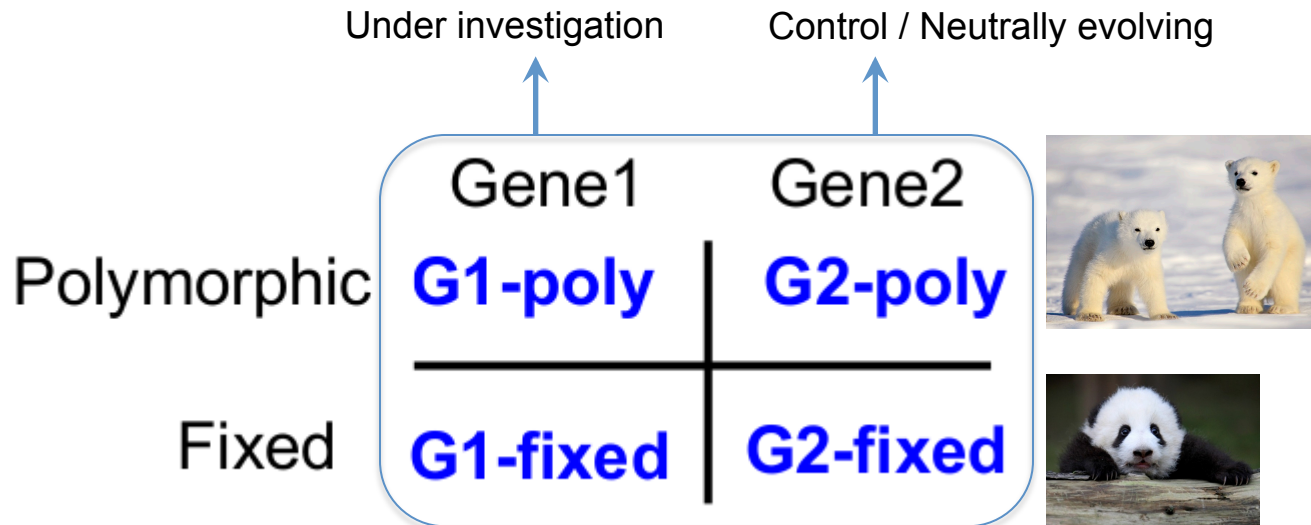(control locus or genome-wide)

Balancing selection
(or relaxation of functional constraints)

HKA test

Hudson-Kreitman-Aguadè (HKA, Hudson et al. 1987) test
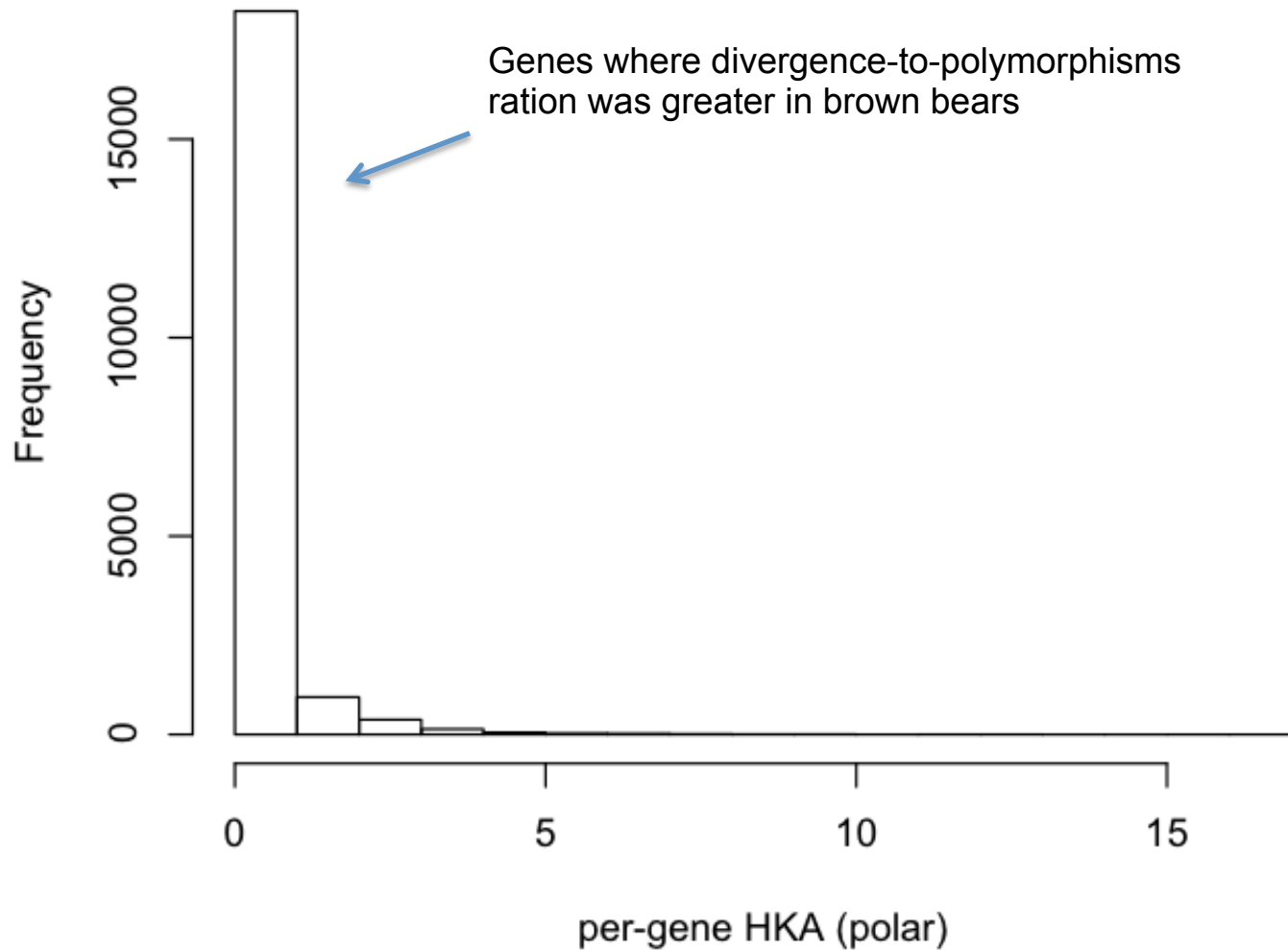

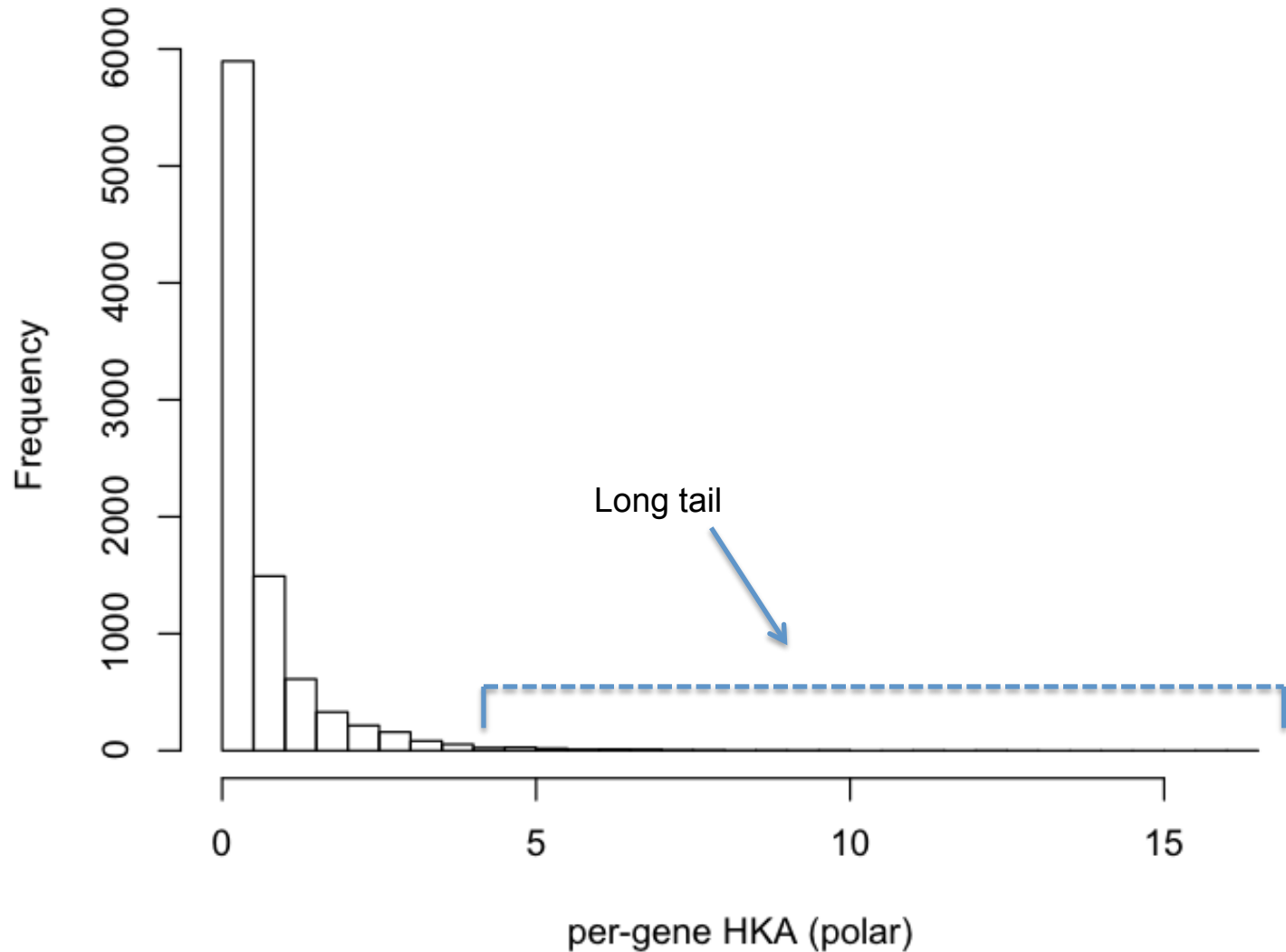Contingency table 2x2: chi-square test

HKA test

Empirical distribution (-log10)

Genes where divergence-to-polymorphisms ration was greater in brown bears
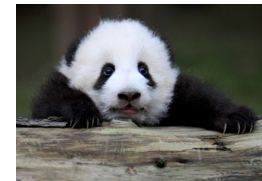
HKA test

Empirical distribution (-log10)

Long tail

Coding polymorphisms

Substitutions in the coding regions may alter (non-synonymous) or not (synonymous) the protein sequence:

ATC	CAA	...	Ancestral sequence
I	Q		

ATA	GAA	...	Species 1
I	K		

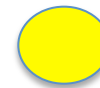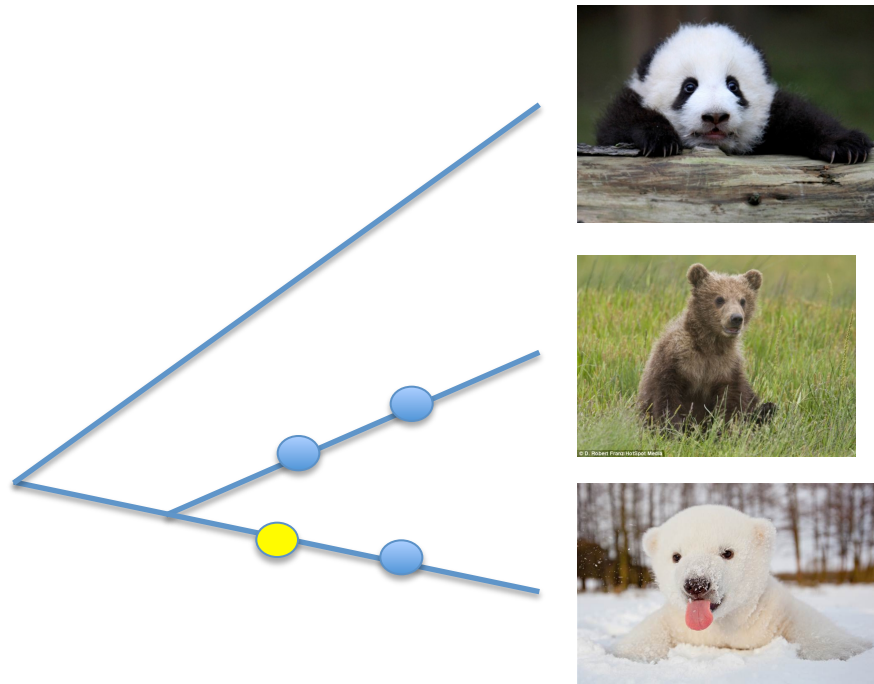|			
ATG	CAG	...	Species 2
M	Q		

Non-synonymous   Synonymous

dN/ds

Comparison between rate of **nonsynonymous** substitutions and rate of **synonymous** substitutions (also called K_a/K_s or ω)

Non-synonymous

Synonymous
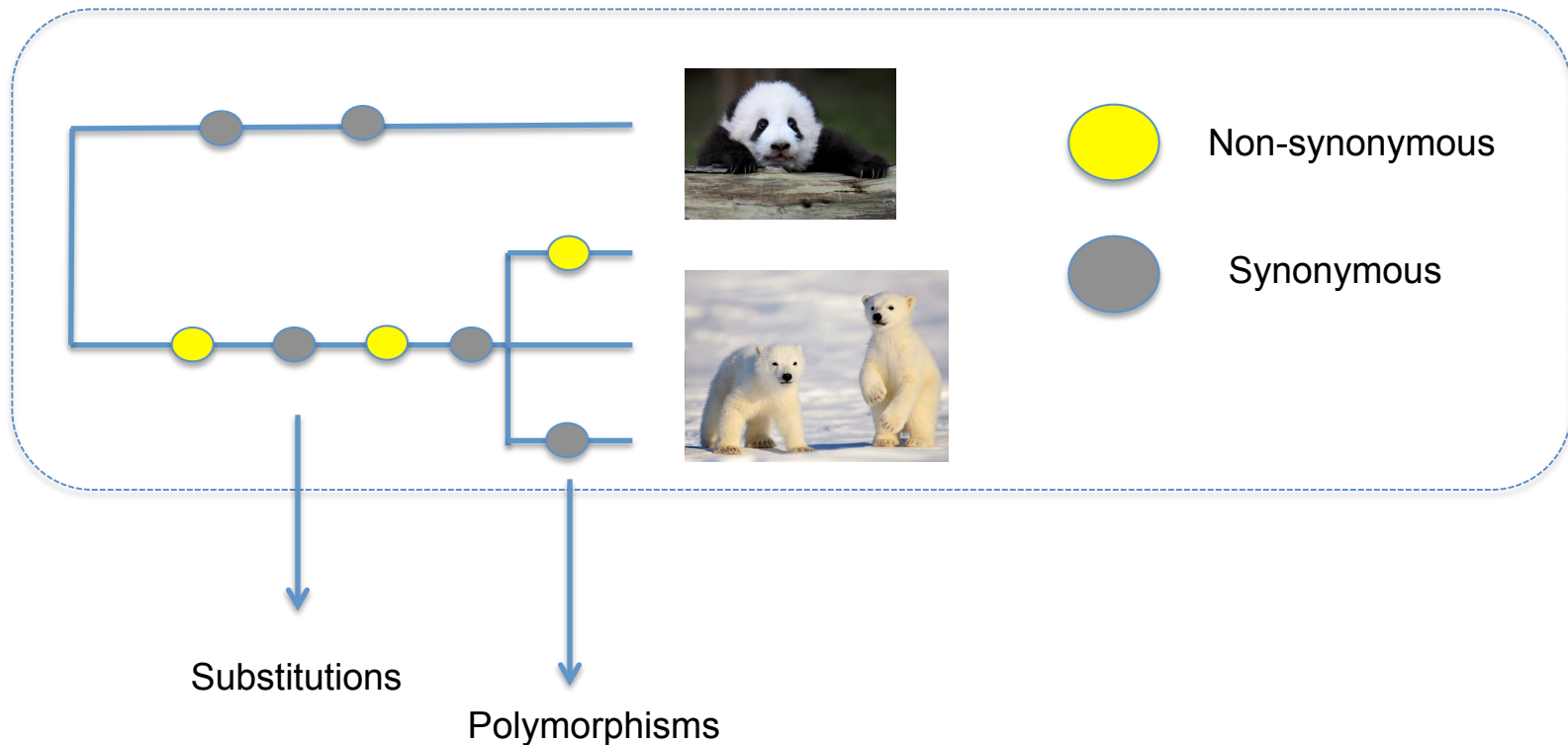
Positive selection
(novel proteins)
$K_a/K > 1$
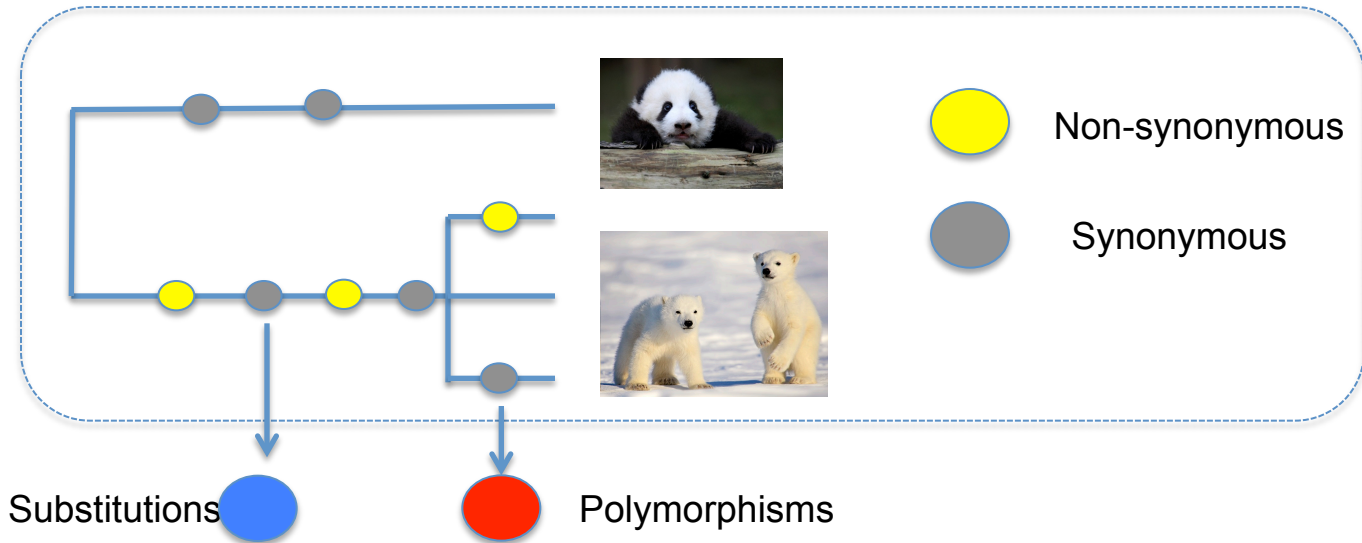
Negative selection
(against deleterious mutations
$K_a/K < 1$

Coding polymorphisms and divergence

Compare the amount of variation within a species to the divergence between species at non-synonymous and synonymous variants.
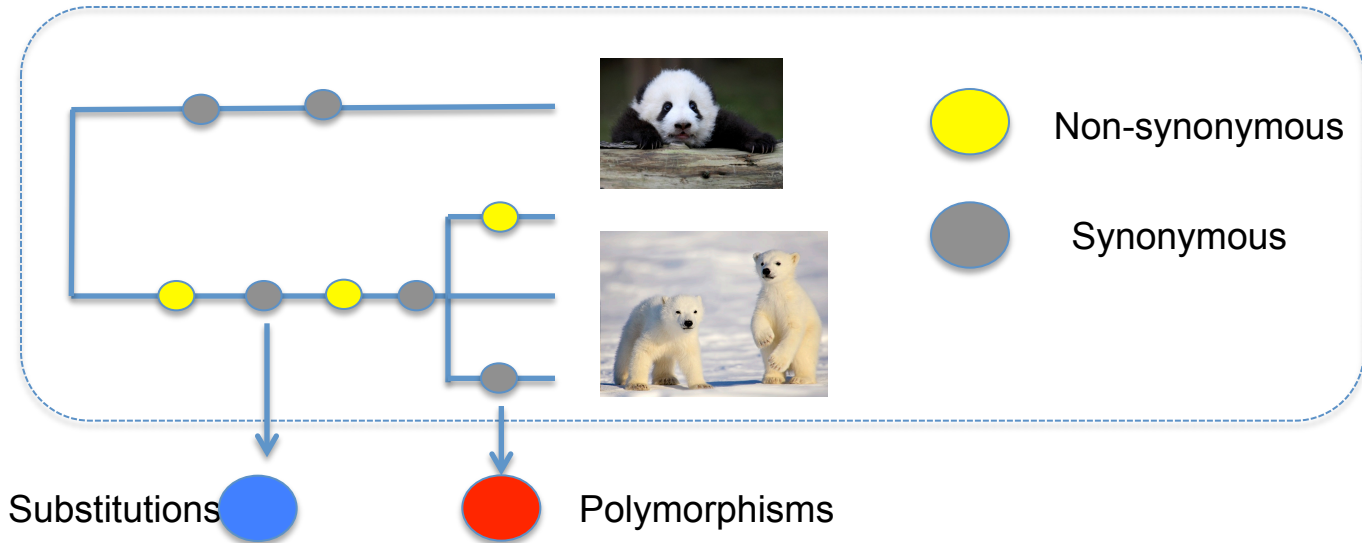

Non-synonymous

Synonymous

Substitutions

Polymorphisms

MK (McDonald–Kreitman) test

	Non-synonymous	Synonymous
Divergence		
Polymorphisms		

MK (McDonald–Kreitman) test

MK (McDonald–Kreitman) test

MK (McDonald–Kreitman) test

Non-synonymous

Synonymous

Substitutions

Polymorphisms

	Non-synonymous	Synonymous
Divergence		
Polymorphisms		

2x2 contingency table

Software available

DnaSP (http://www.ub.edu/dnasp/)

Arlequin (http://cmpg.unibe.ch/software/arlequin35/)

BayeScan (http://cmpg.unibe.ch/software/BayeScan/)

libsequence (http://www.molpopgen.org/software.html)

sweep (http://www.broadinstitute.org/mpg/sweep/)

iHS (http://coruscant.itmat.upenn.edu/software.html)

nSL (http://cteg.berkeley.edu/~nielsen/resources/software/)

Pre-computed values (USCS genome browser tables)

Homemade scripts

 …