

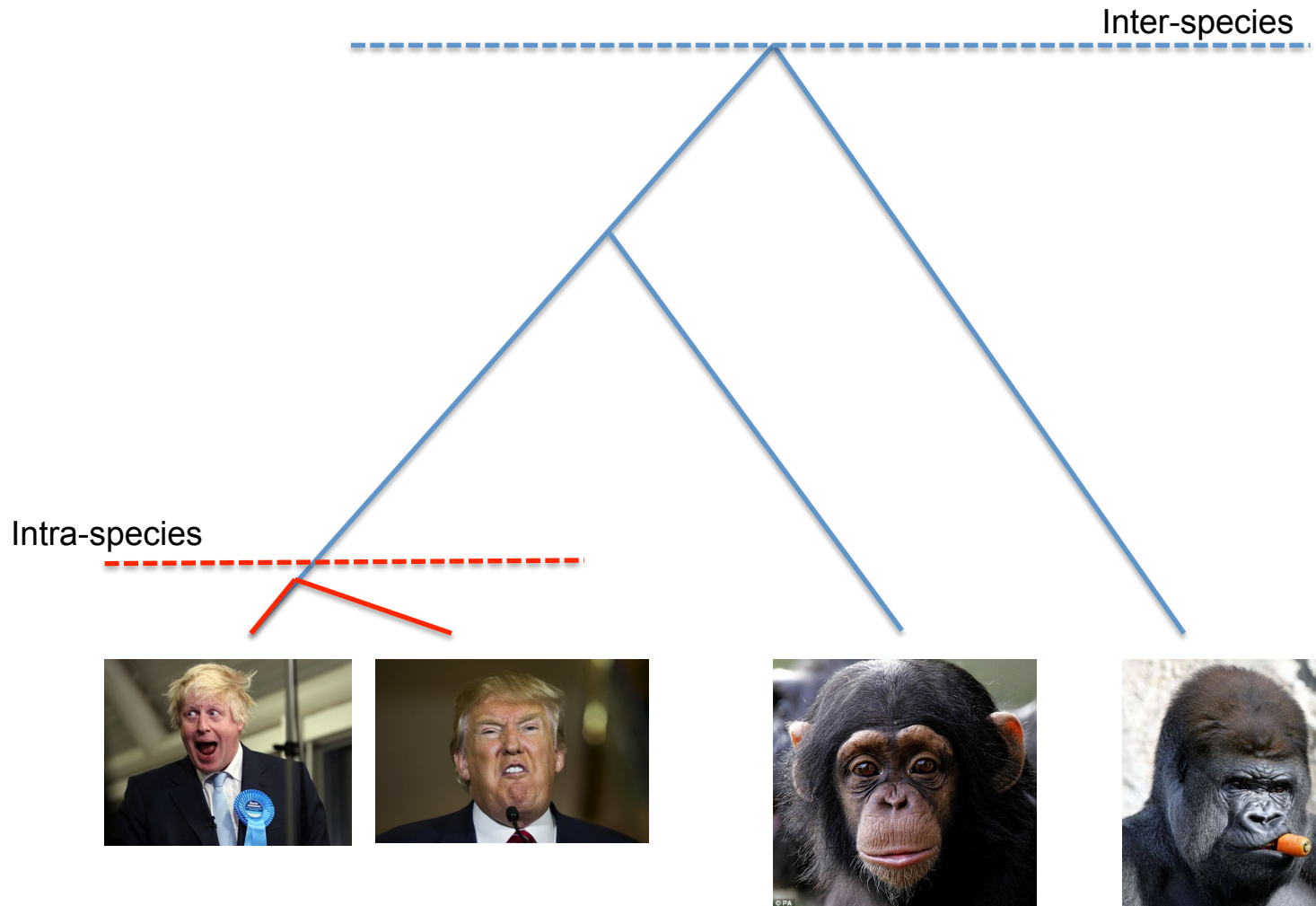
Detecting selection: methods (II)

Matteo Fumagalli

Outline

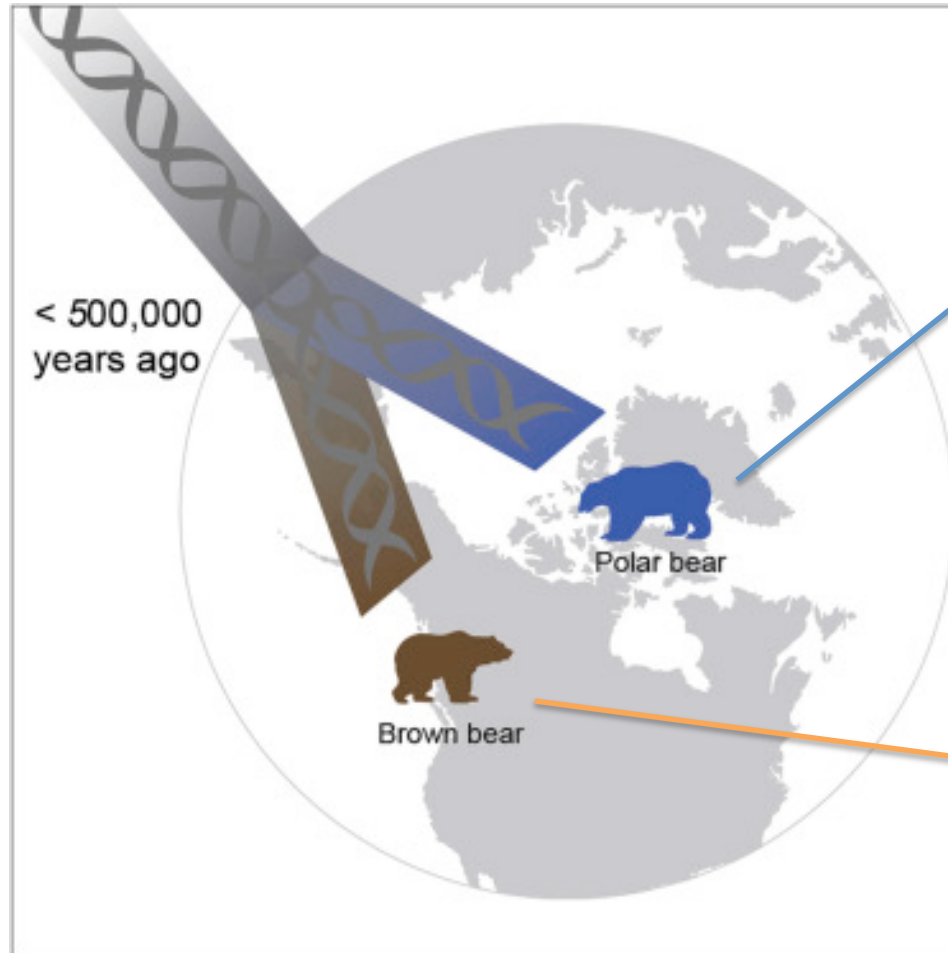
- Brief introduction to natural selection
- Modes of selection
- Inferring selection at the intra-species level
 - Genetic differentiation
 - Haplotype variation
 - Model-based approaches
 - Testing for significance
- Inferring selection at the inter-species level
- Detecting selection from low-depth sequencing data

Inter-species variation



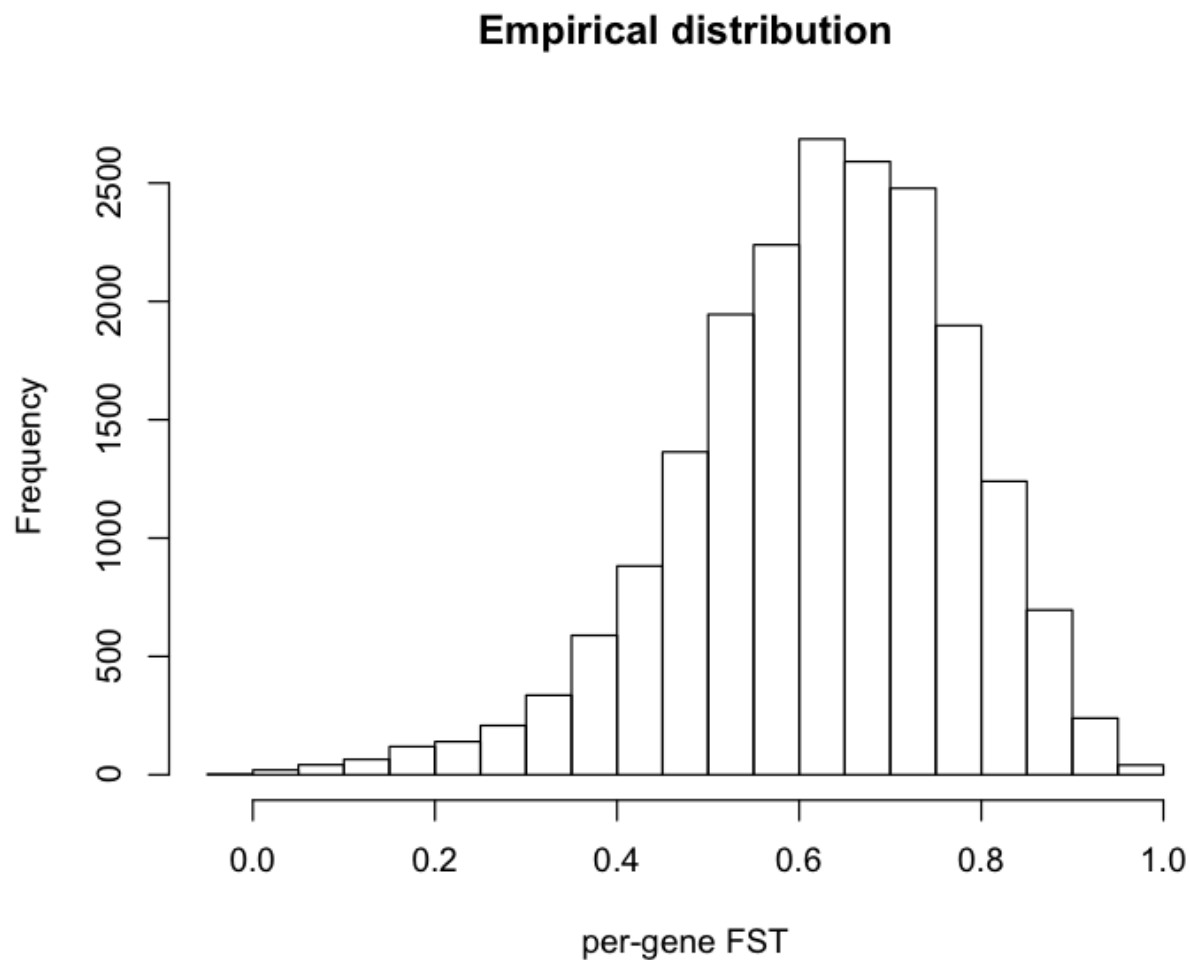
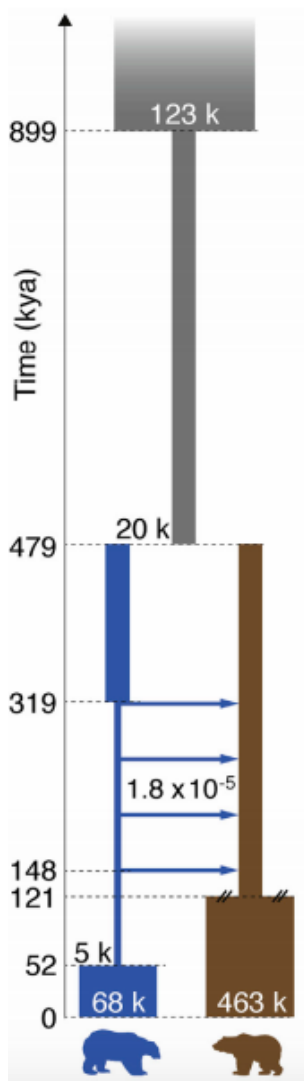
events in the deep past, macro-evolutionary trends, selection between species

Polar bears vs. Brown bears

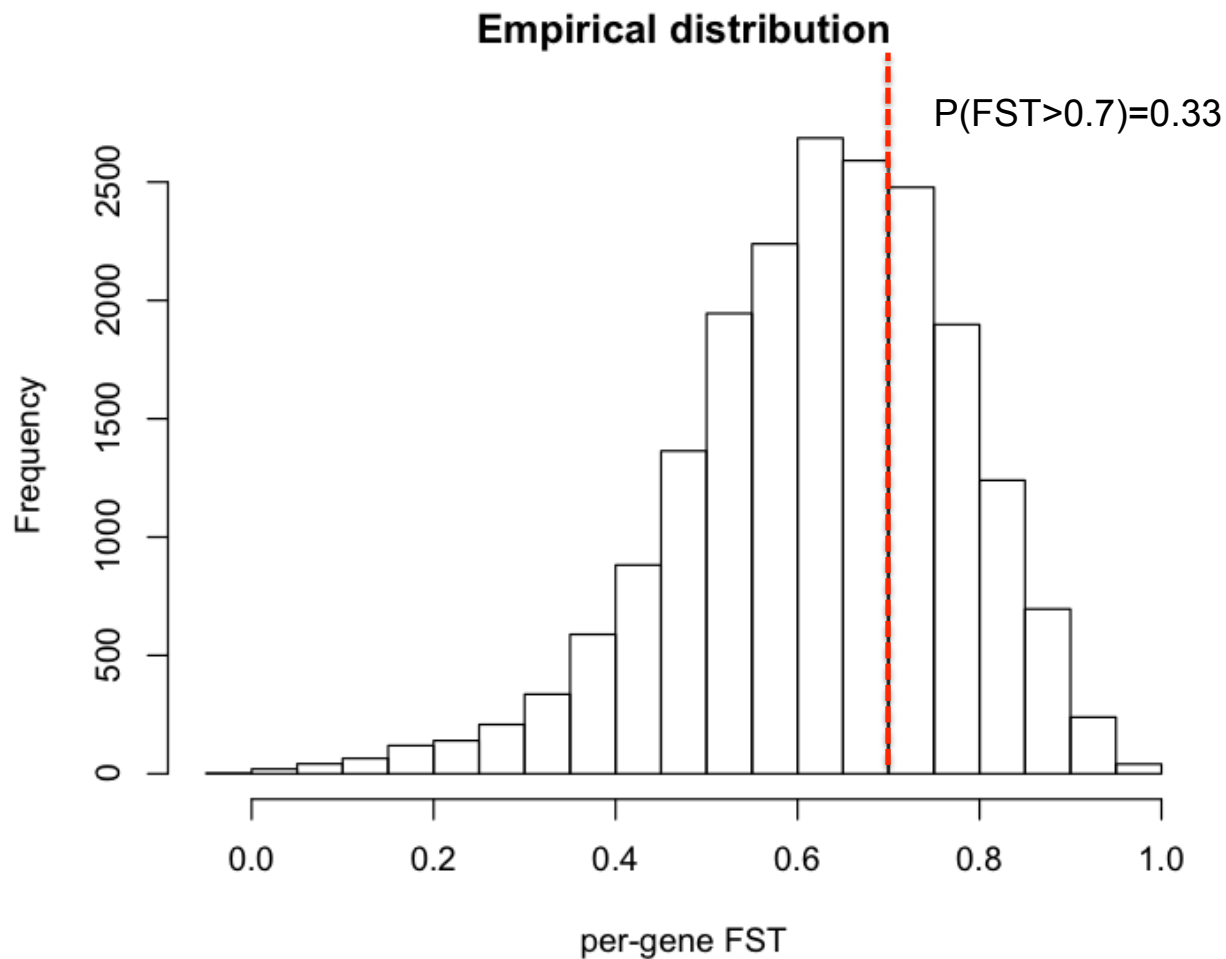
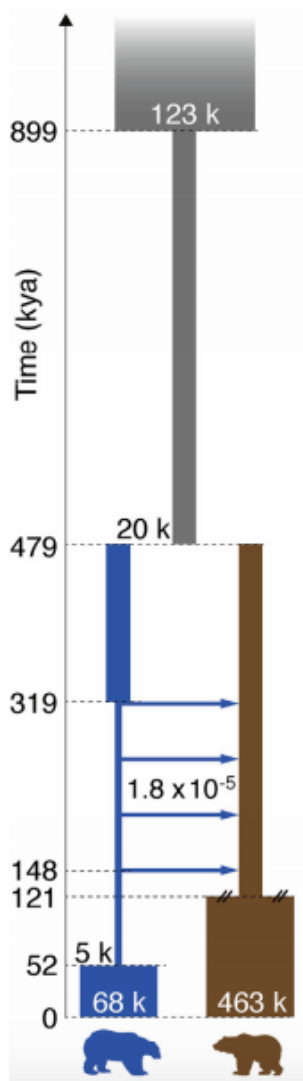


Question: what are the genetic signatures of polar bears' adaptation to the Arctic environment?

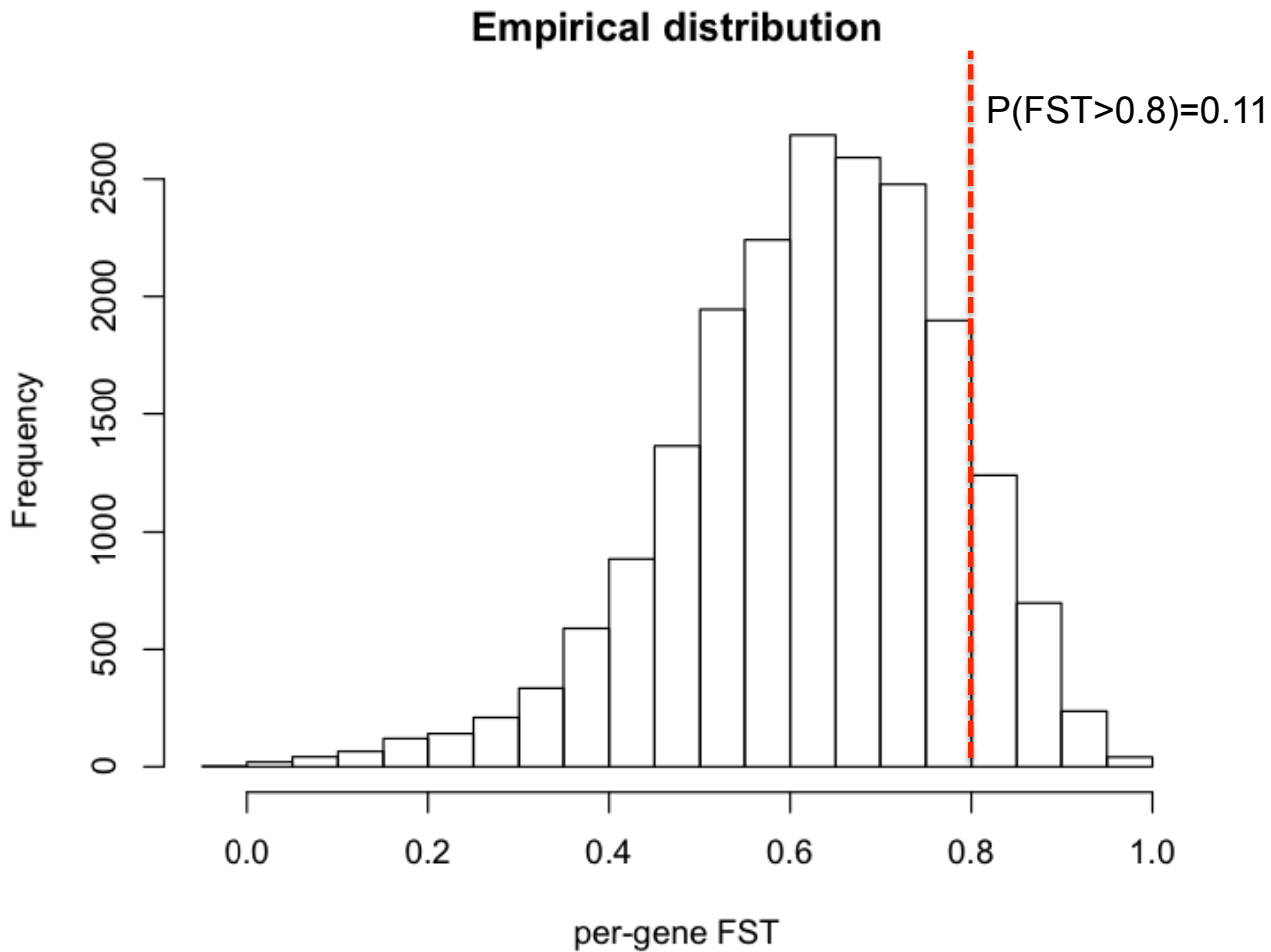
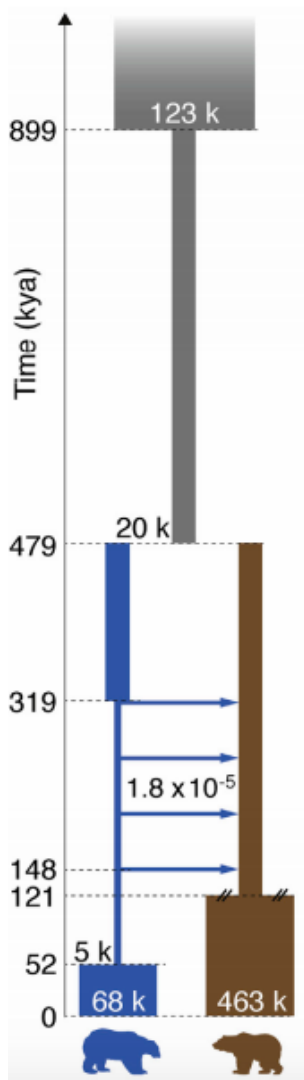
Expected genetic differentiation



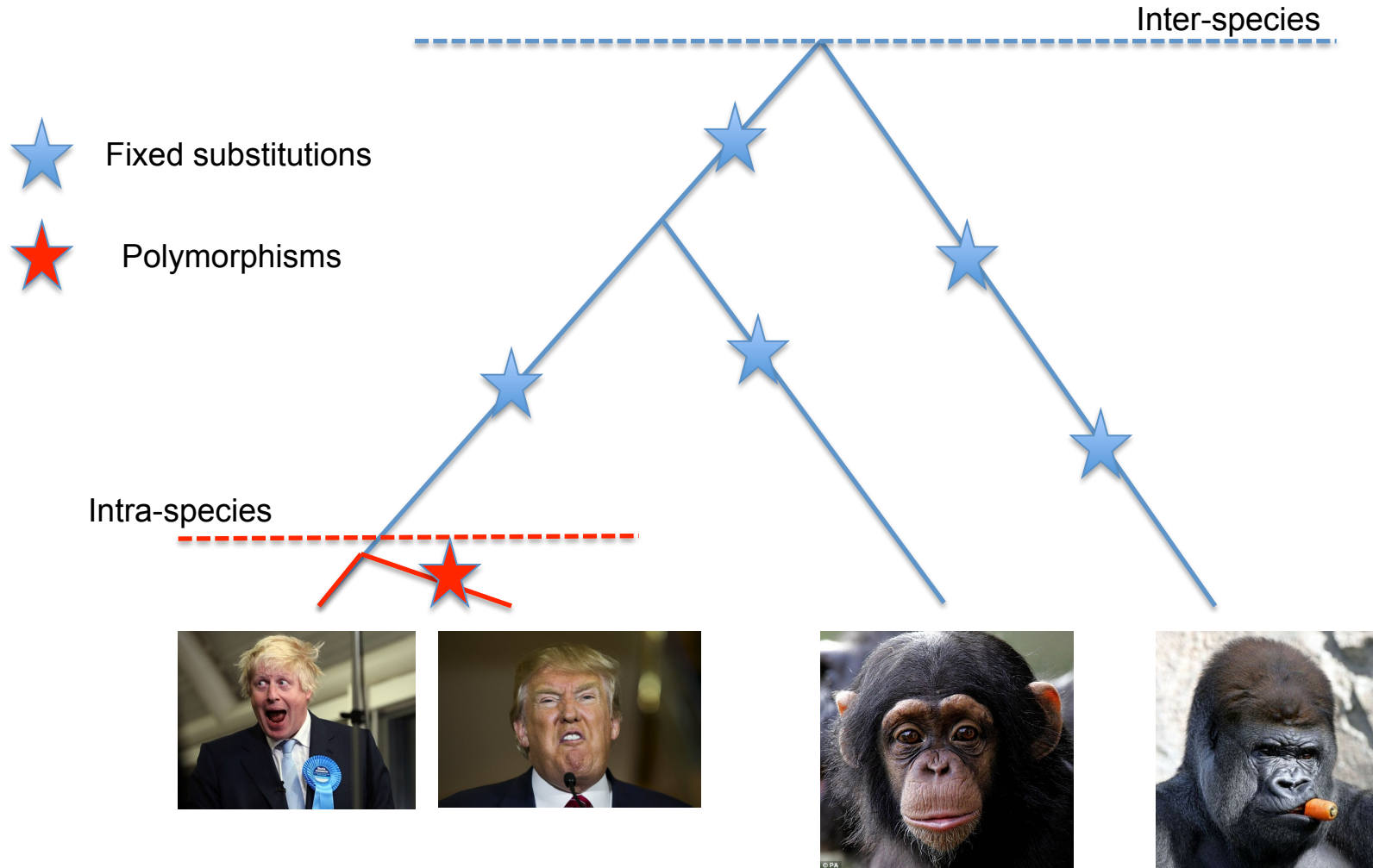
Expected genetic differentiation





Expected genetic differentiation

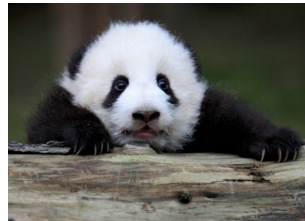
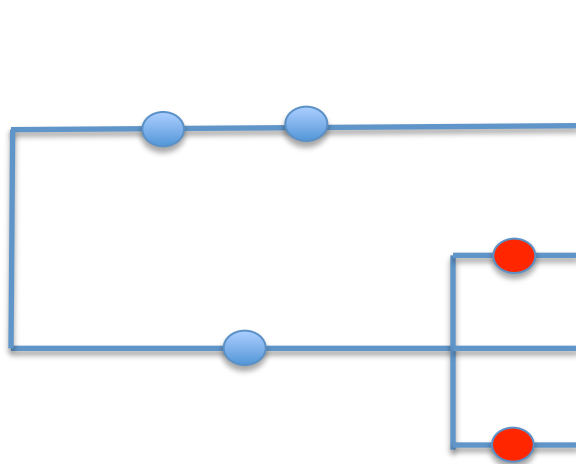


Inferring inter-species selection

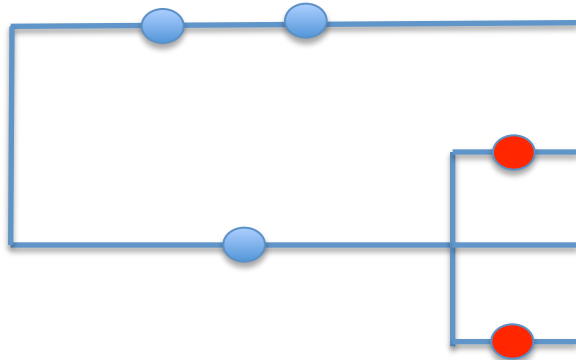


Polymorphisms and divergence

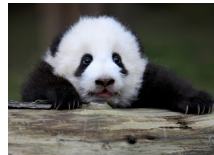
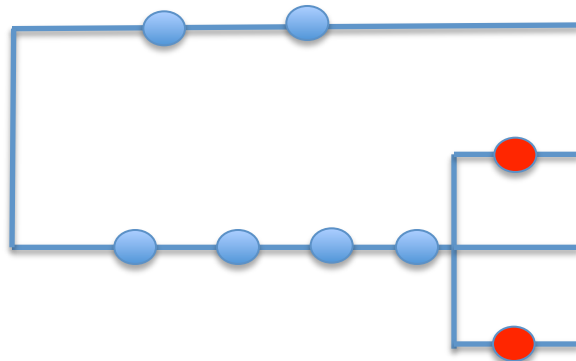
As both depend on mutation rates,  and  levels are expected to be proportional.



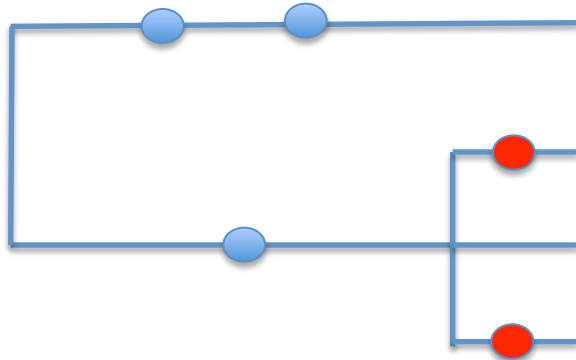
Polymorphisms and divergence



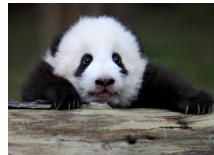
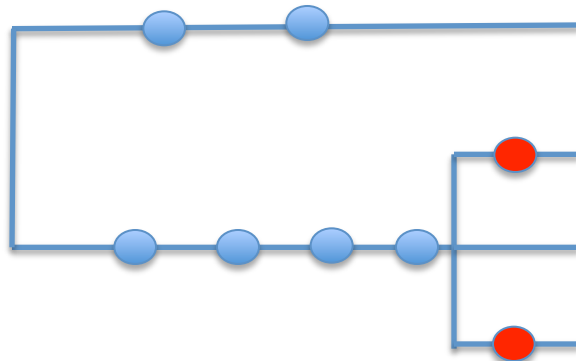
Under neutrality
(control locus or genome-wide)



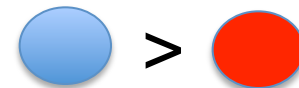
Polymorphisms and divergence



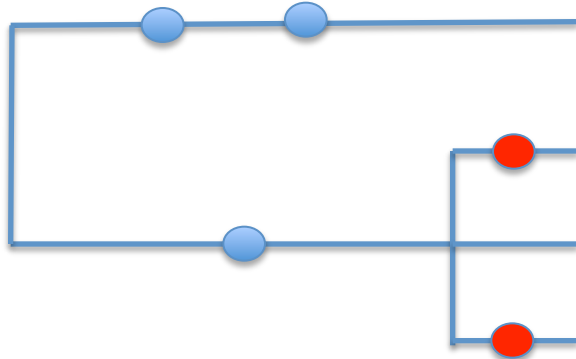
Under neutrality
(control locus or genome-wide)



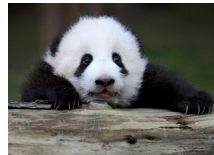
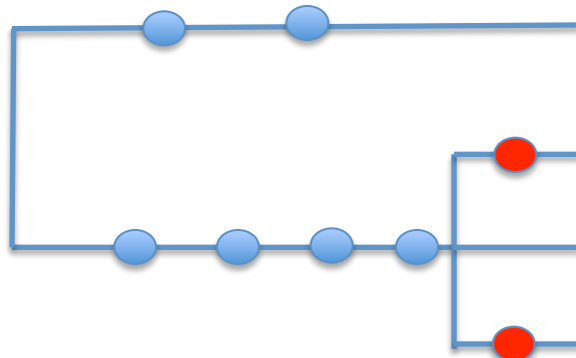
??? selection



Polymorphisms and divergence



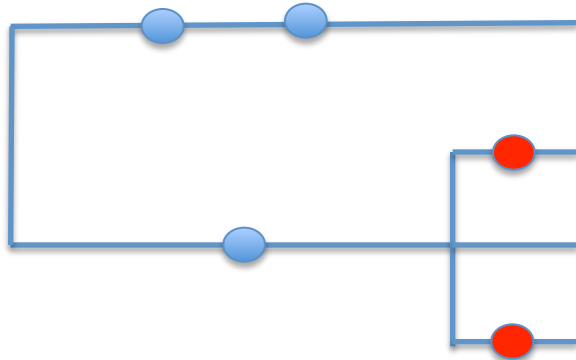
Under neutrality
(control locus or genome-wide)



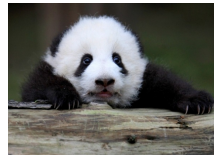
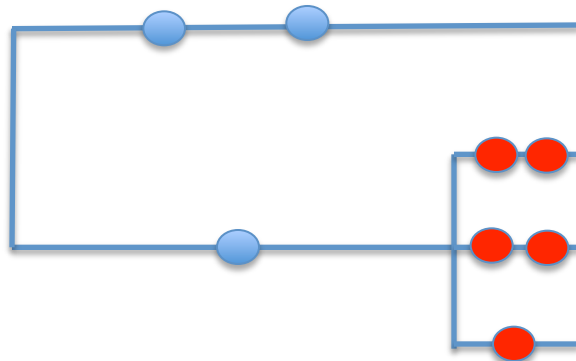
Positive selection
(or negative selection)



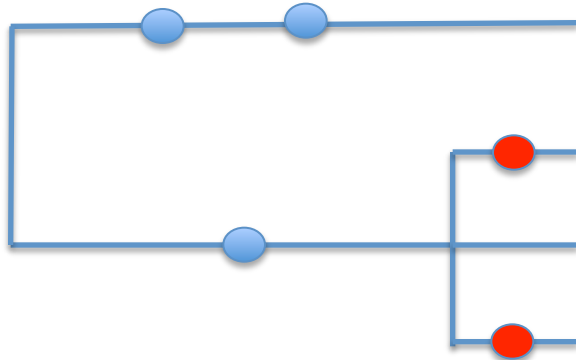
Polymorphisms and divergence



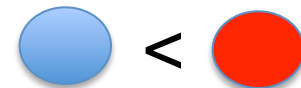
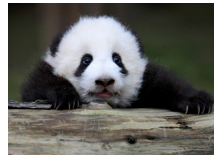
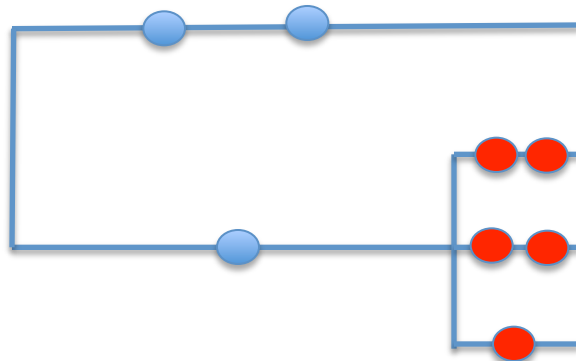
Under neutrality
(control locus or genome-wide)



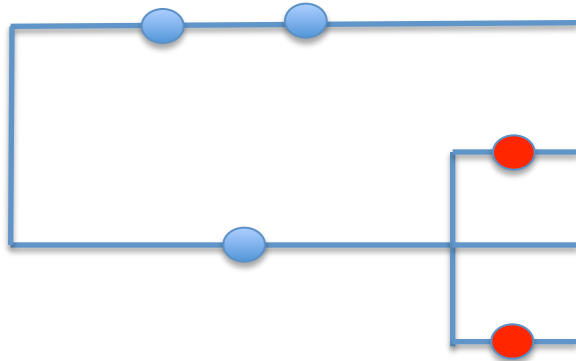
Polymorphisms and divergence



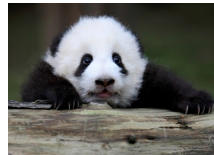
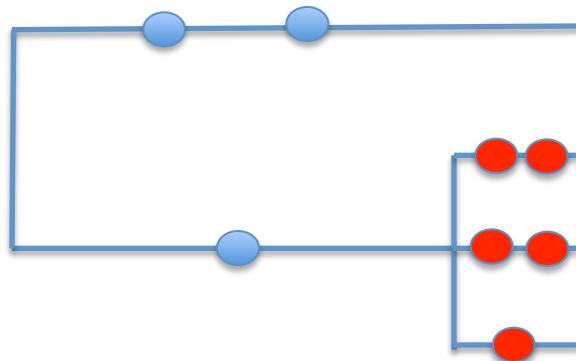
Under neutrality
(control locus or genome-wide)



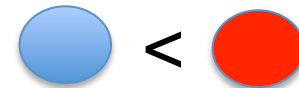
Polymorphisms and divergence



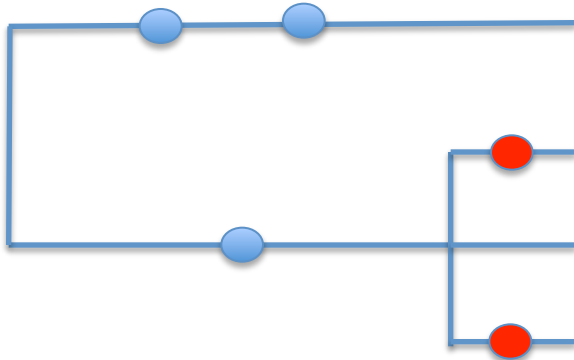
Under neutrality
(control locus or genome-wide)



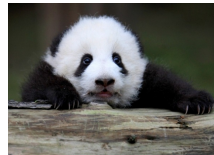
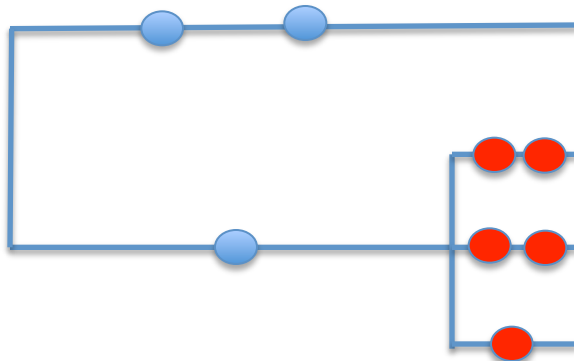
??? selection



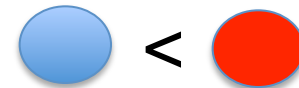
Polymorphisms and divergence



Under neutrality
(control locus or genome-wide)

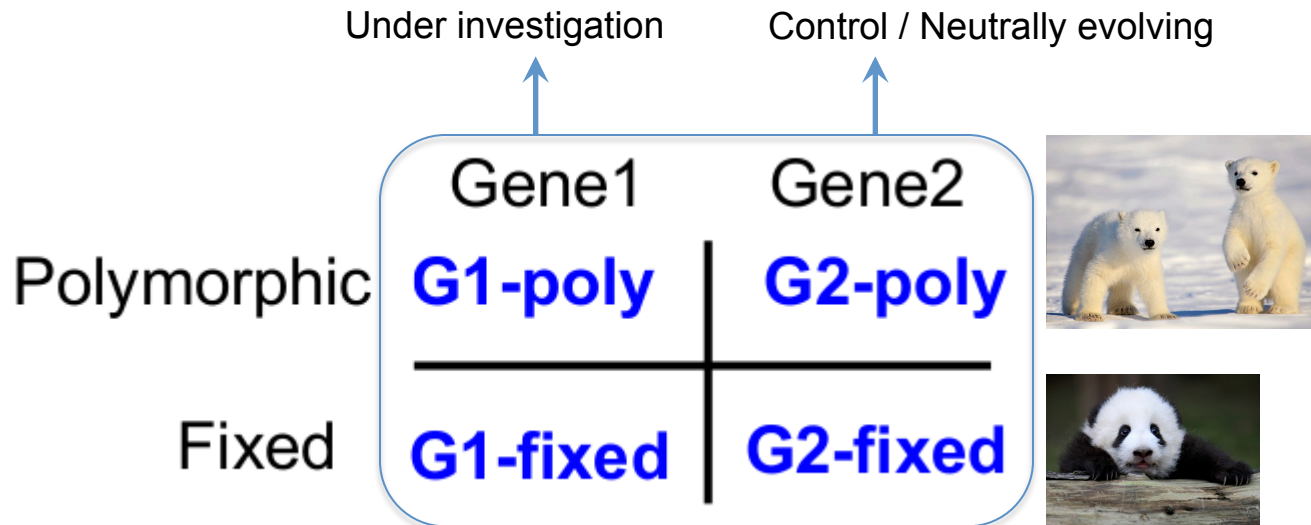


Balancing selection
(or relaxation of functional constraints)



HKA test

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

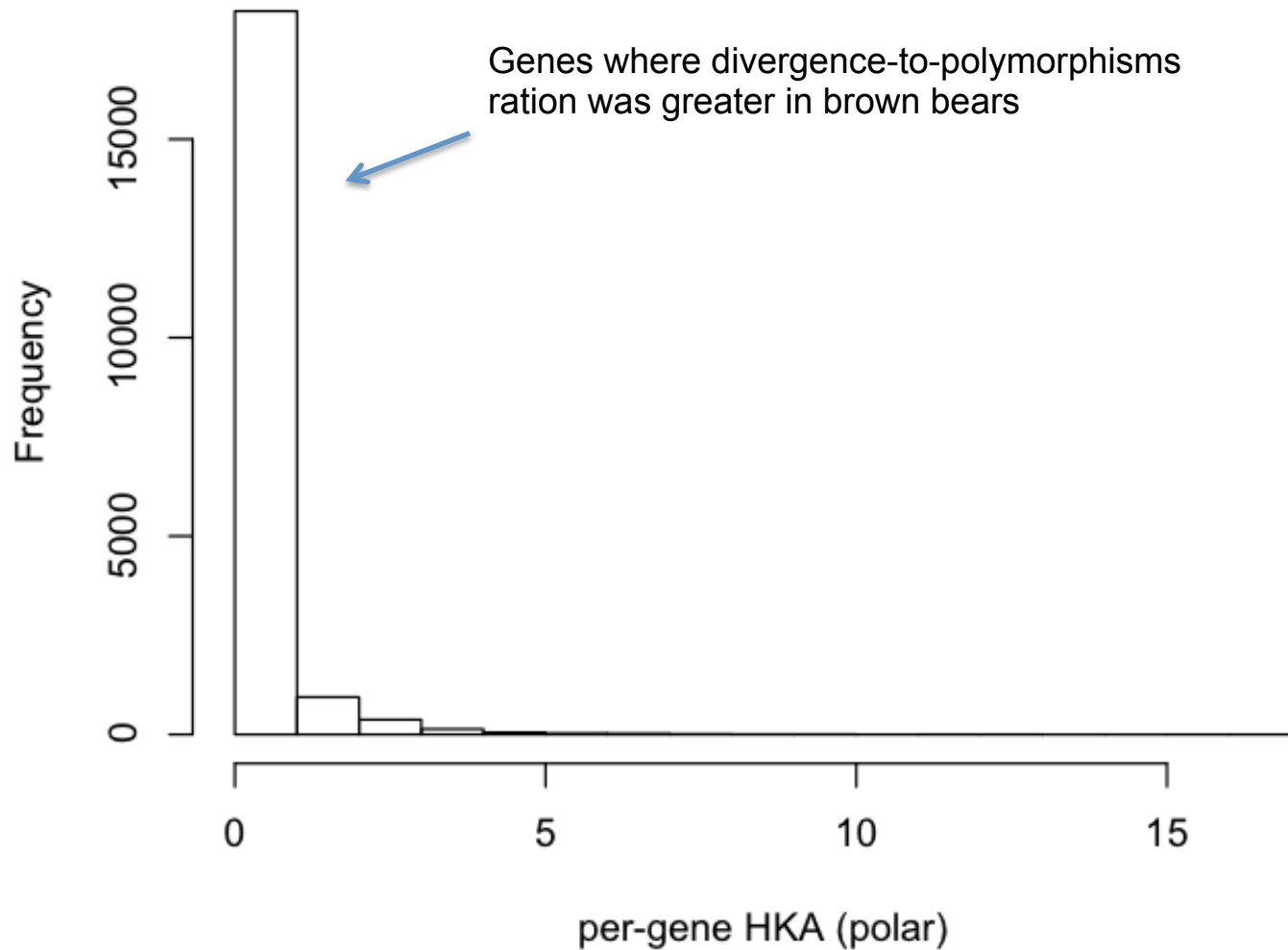


Contingency table 2x2: chi-square test

HKA test



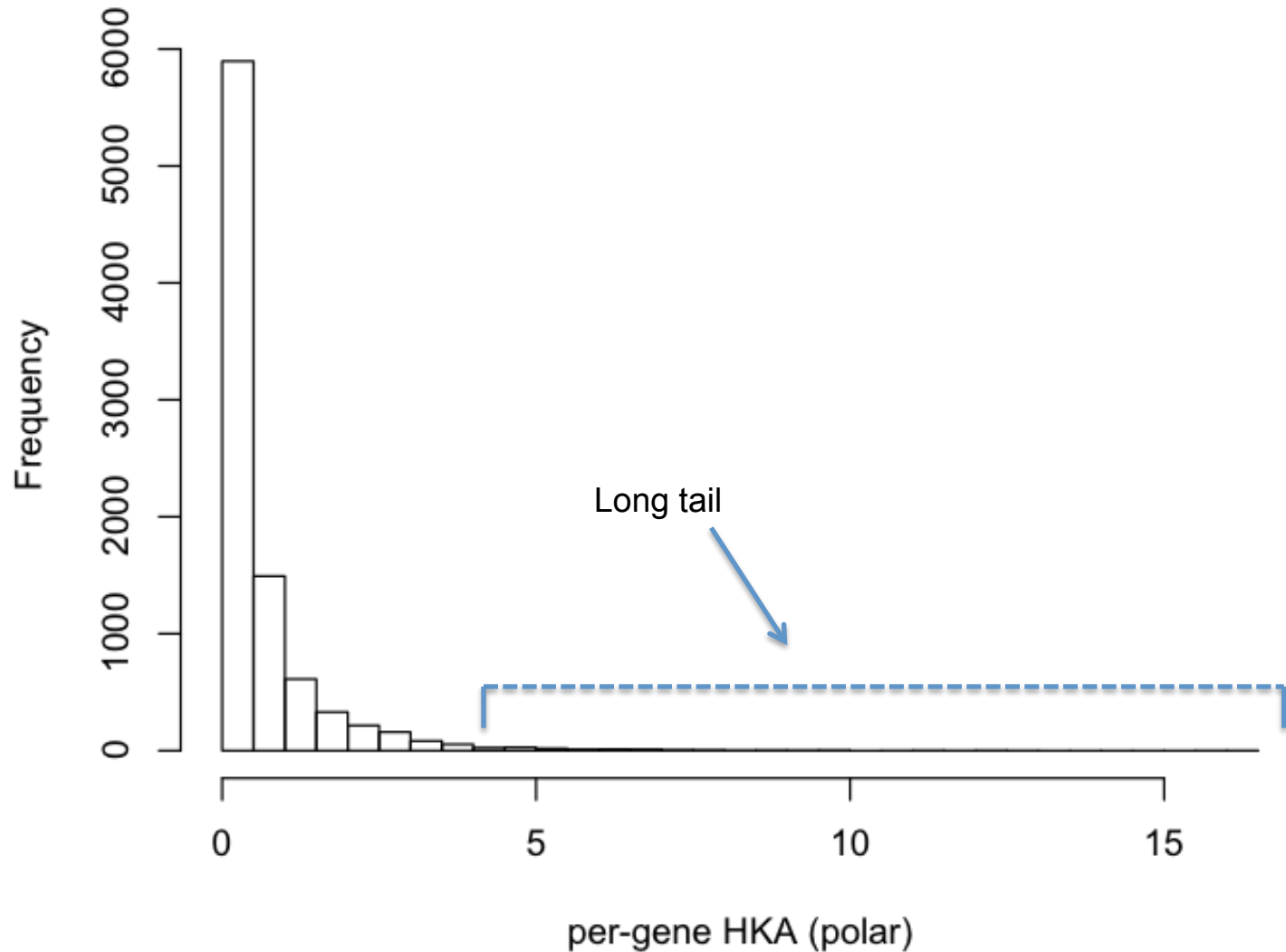
Empirical distribution ($-\log_{10}$)



HKA test



Empirical distribution ($-\log_{10}$)

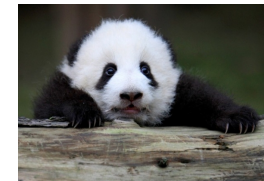


Coding polymorphisms

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

ATC CAA ...
I Q

Ancestral sequence



ATA GAA ...
I K

Species 1



ATG CAG ...
M Q

Species 2



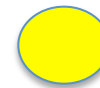
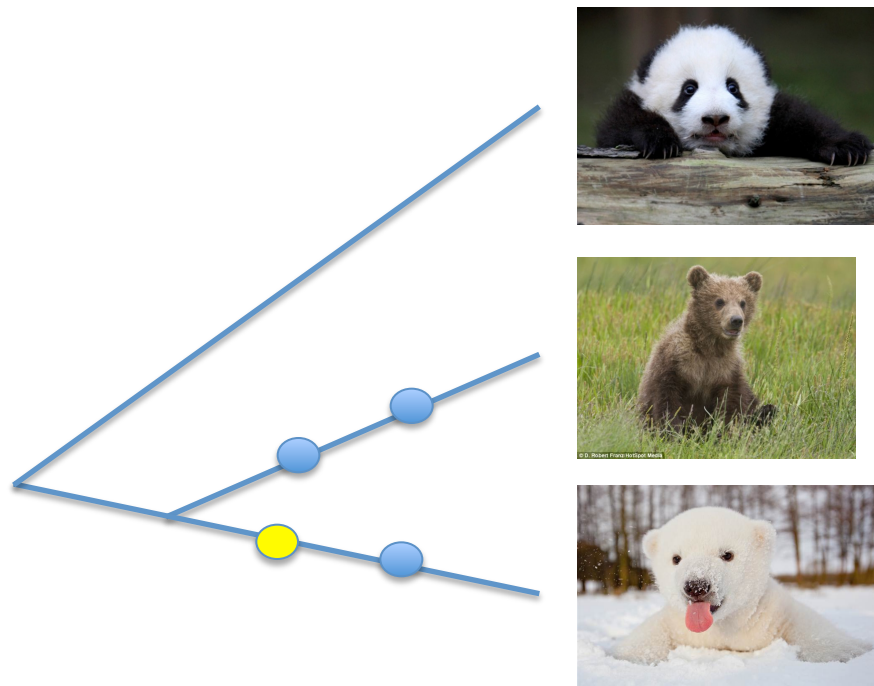
Non-synonymous



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)

$$K_a/K_s > 1$$

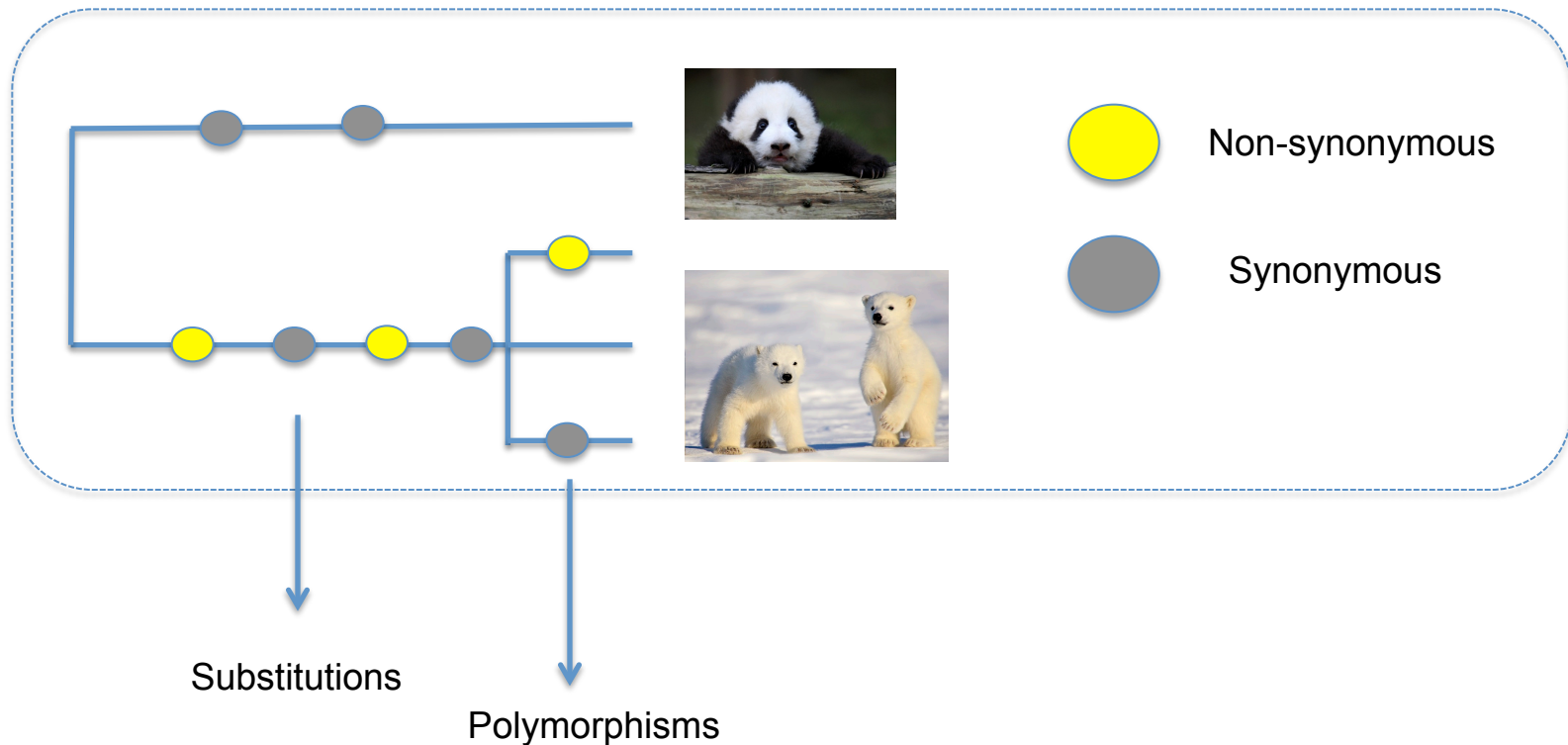
Negative selection

(against deleterious mutations)

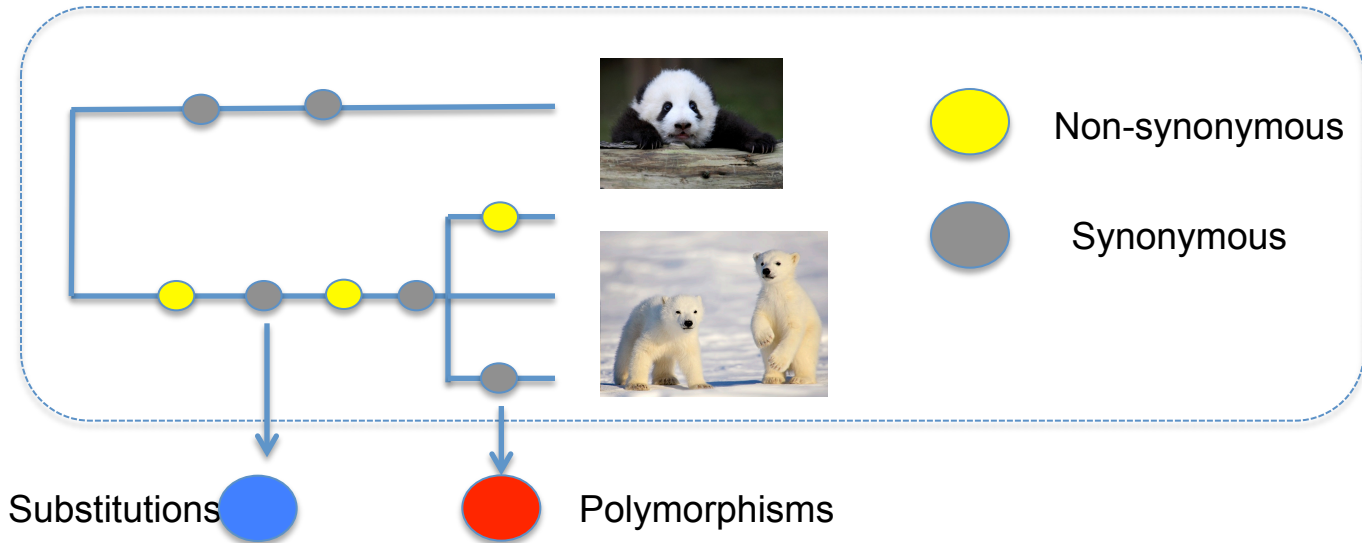
$$K_a/K_s < 1$$

Coding polymorphisms and divergence

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

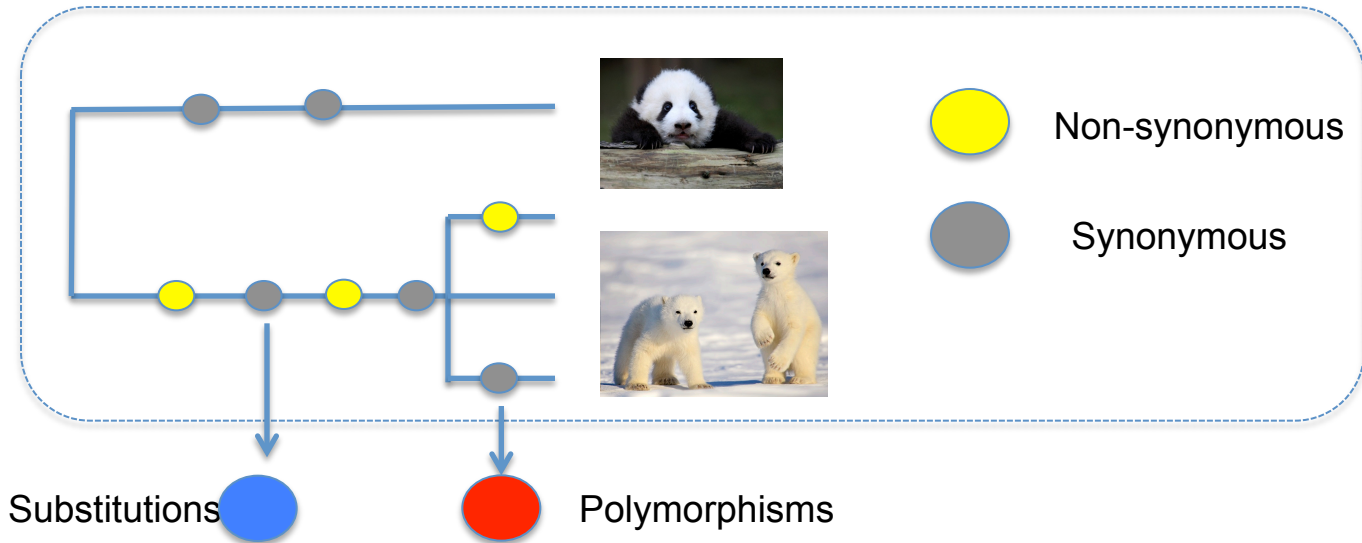






MK (McDonald–Kreitman) test



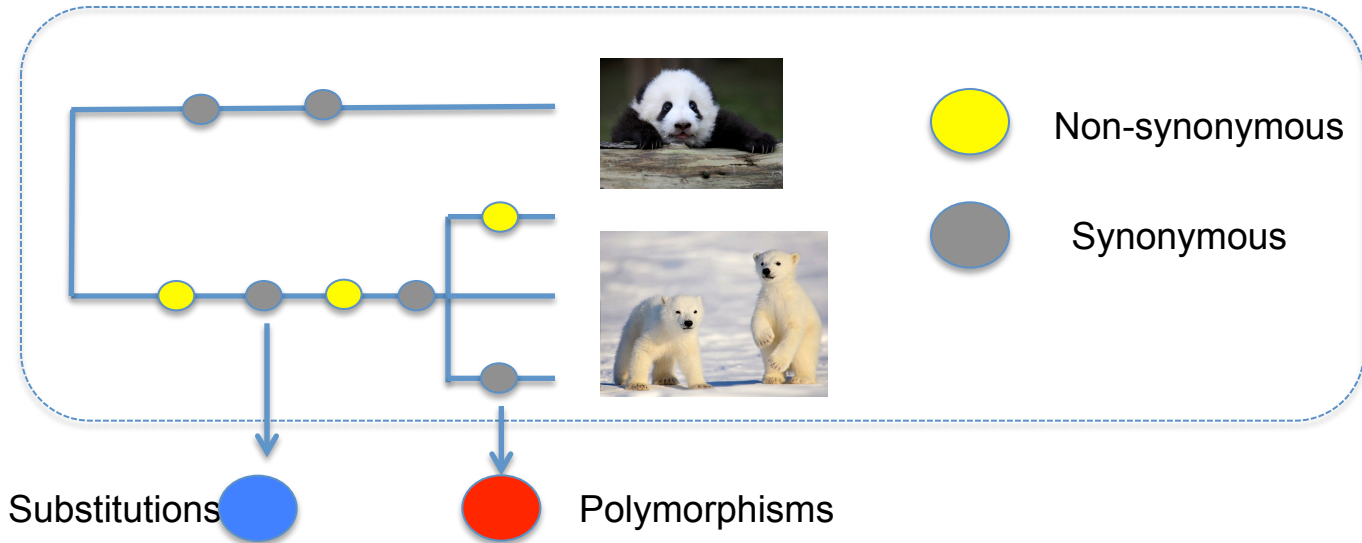
	Non-synonymous	Synonymous
Divergence		
Polymorphisms		









MK (McDonald–Kreitman) test



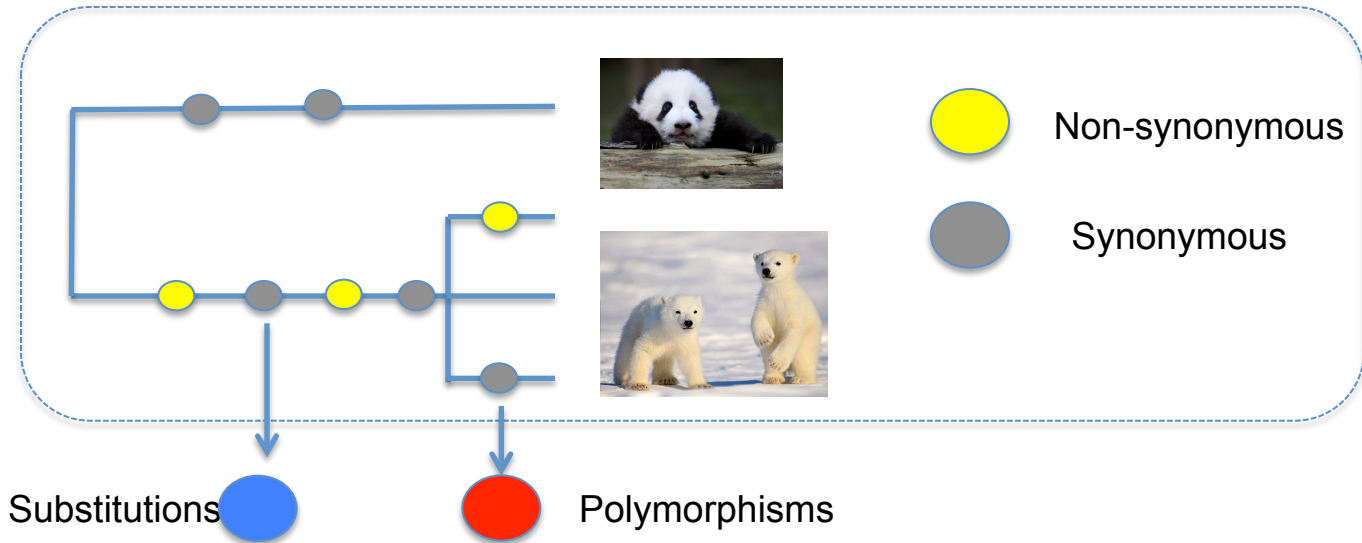
	Non-synonymous	Synonymous	
Divergence	 	 	dN/dS between species
Polymorphisms			









MK (McDonald–Kreitman) test



	Non-synonymous	Synonymous	
Divergence	 	 	dN/dS within species
Polymorphisms	 	 	

MK (McDonald–Kreitman) test

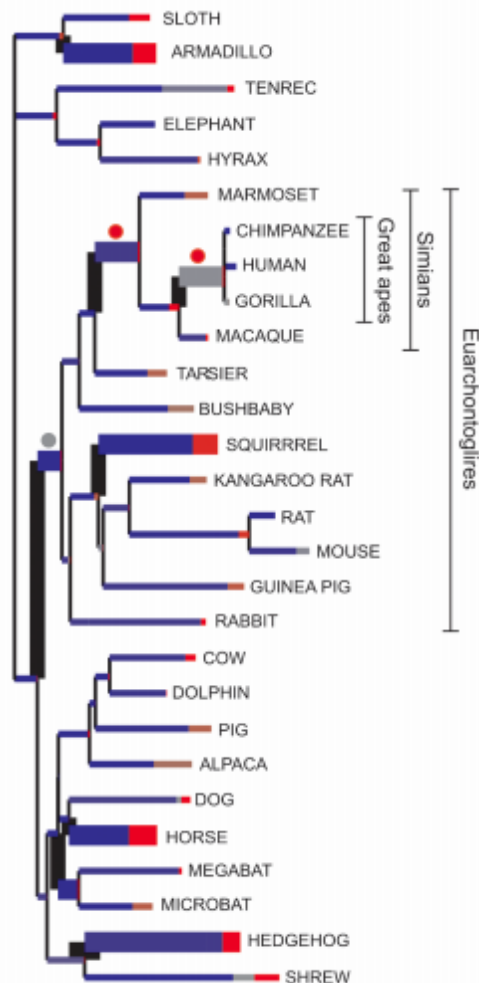


	Non-synonymous	Synonymous
Divergence	 	 
Polymorphisms	 	 

2x2 contingency table

Comparative genomics

Identify multiple-species conserved genomic elements or accelerated rates of substitution in specific lineages (e.g. PAML software).



	159		191
Sloth	L	R	E
Tenrec	L	Q	D
Elephant	L	Q	M
Hyrax	L	Q	D
Marmoset	L	R	E
Chimpanzee	L	R	E
Human	L	R	E
Gorilla	L	R	E
Macaque	L	R	E
Bushbaby	L	R	E
Kangaroo rat	L	R	E
Rat	L	R	E
Mouse	L	R	E
Guinea pig	L	R	E
Rabbit	L	R	E
Cow	L	R	E
Dolphin	L	R	E
Dog	L	R	E
Horse	L	R	E
Hedgehog	L	R	E
Shrew	L	R	E

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

...

Summary

- Methods to detect signatures of selection are grouped based on:
 - time of selection
 - summary statistics used
- Assessing statistical significance through empirical or expected distributions
- Investigating complex models of selection and adaptation