Selection II

Drive

Rapid spread of gene variants independent of selection.

- **Meiotic drive**
- **Postmeiotic drive**
- **Zygotic drive**
- **Molecular drive** (e.g. through gene conversion)
- **Mutation/reparation drive**

Genetic draft
(genetic hitchhiking, linked selection)

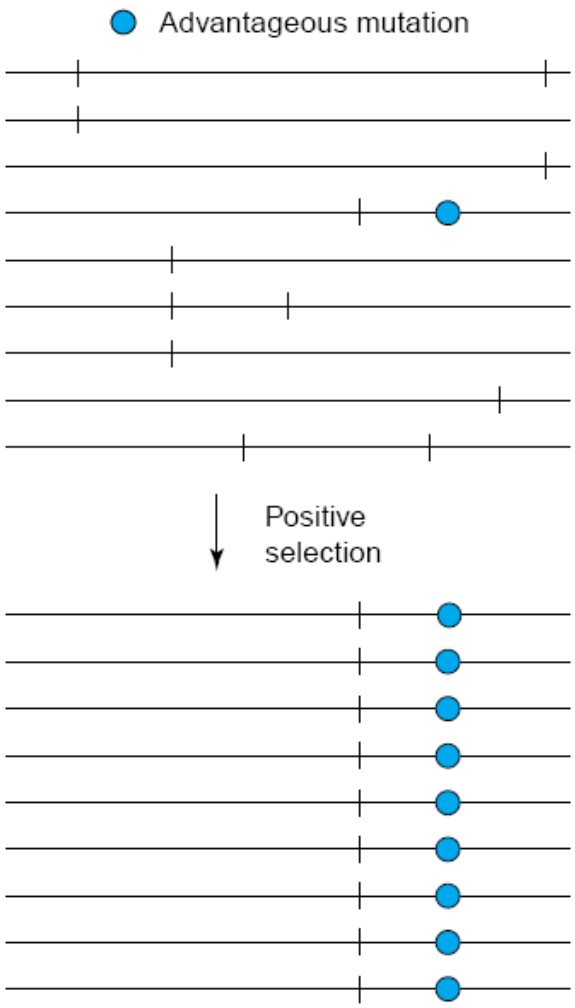
- Change in the frequency of an allele because of linkage with beneficial or detrimental allele.

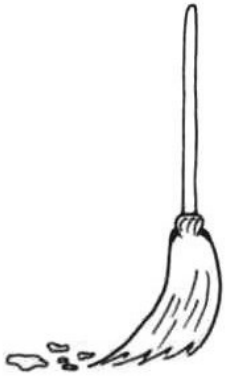
Detection of selection from molecular data

Recent positive selection

Selective sweep

- Reduction of genetic variability around positive mutation.
- Increase of linkage disequilibrium around positive mutation.
- Increase of genetic differentiation (Fst) between populations.

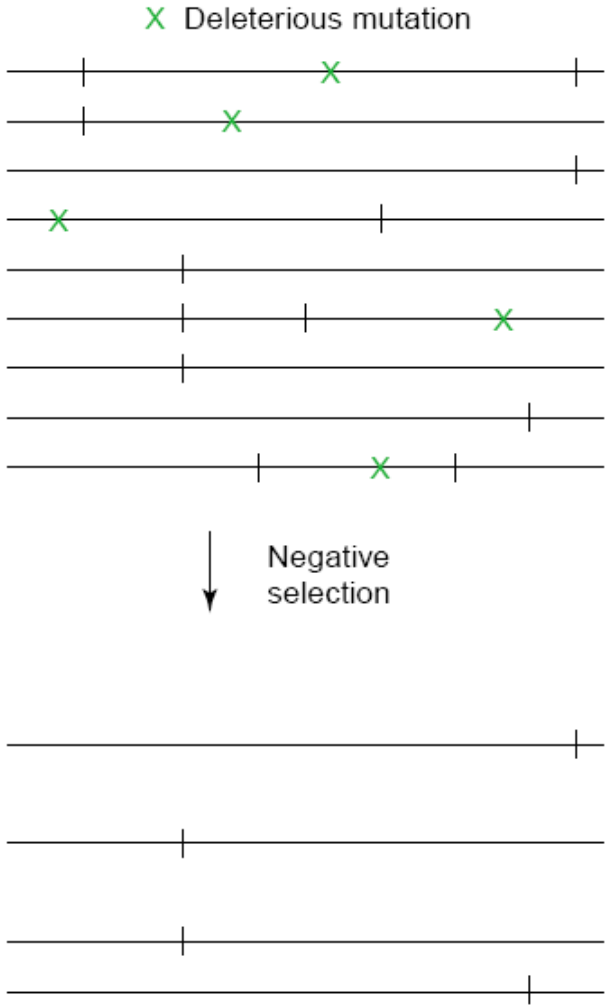




Recent negative selection

Background selection

- Reduction of genetic variation around the negative mutation, but not so marked as in case of positive selection.
- Decrease of genetic differentiation (Fst) between populations.

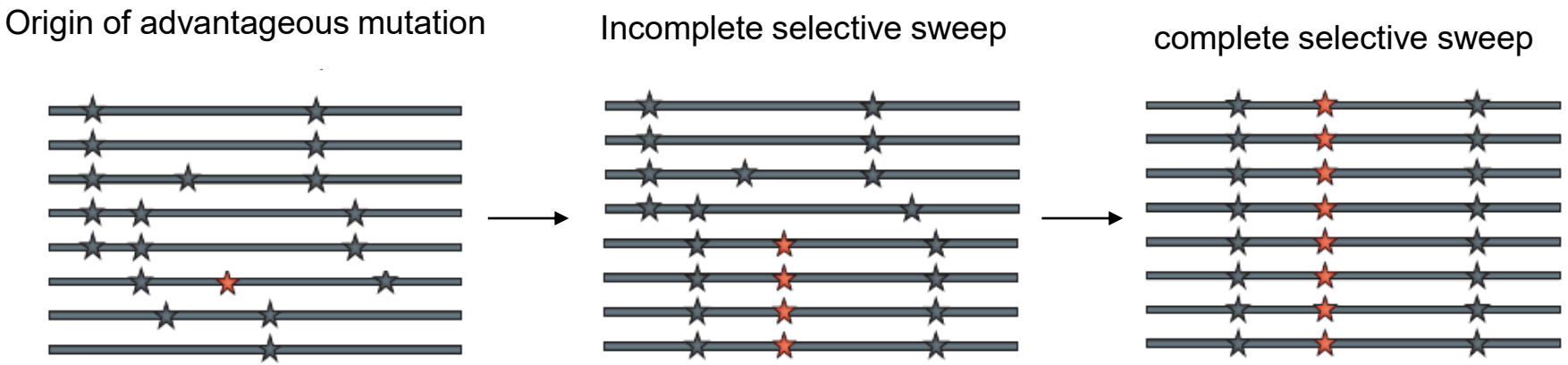


Balancing selection

- Increases levels of genetic variation
- Maintain relatively high frequecies of alternative alleles.

Methods of detection of recent positive selection (selective sweep)

- Selection not much longer than ~ N_e generations ago (in humans ~ 250 000 years).
  - Reduction of genetic variability
  - Increase of linkage disequilibrium
  - Changes in allele frequency spektra
  - Increase in genetic differentiation (Fst) between populations



Hudson-Kreitman-Aguadé (HKA) test

$\theta = 4N_e\mu$

$D = 2\mu t$

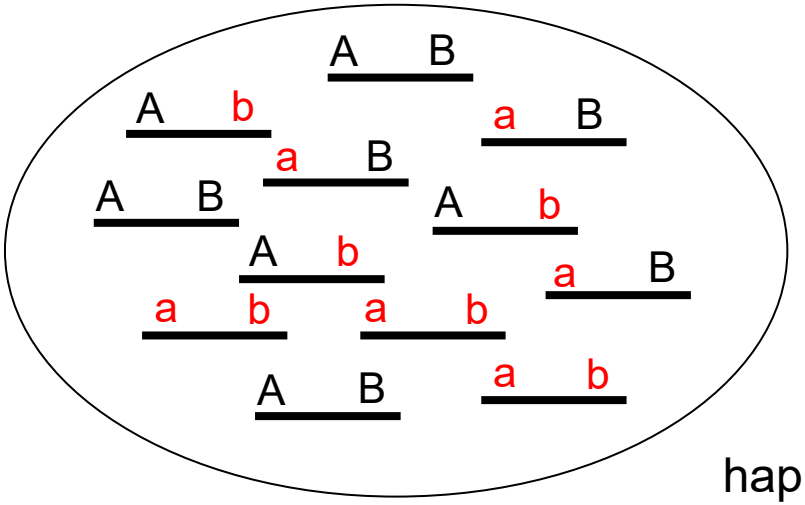
- Compares levels of within species polymorphism (θ) and between species divergence (D) in multiple loci.
- For neutral sequences θ/D ratio should be constant.
- Recent positive selection reduces θ, but does not affect D.
- Positive HKA test could be caused by selection for linked gene with the studied gene.

	θ	D
Lokus 1	θ_1	D_1
Lokus 2	θ_2	D_2
Lokus 3	θ_3	D_3
Lokus 4	θ_4	D_4

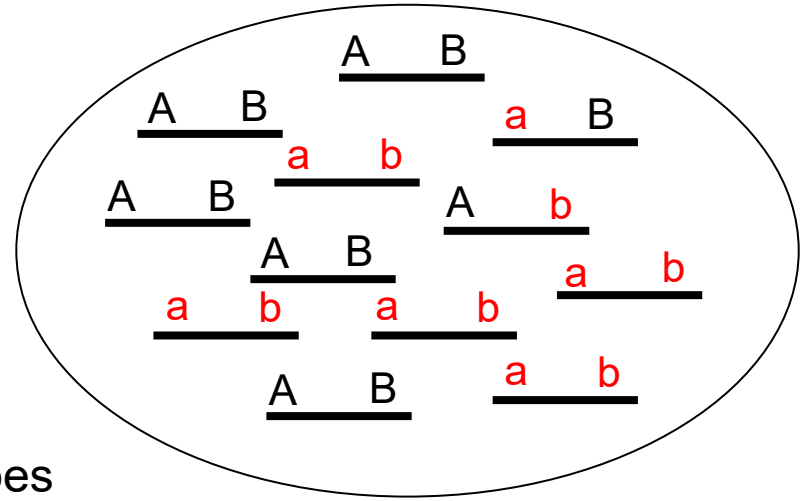
Software: http://genfaculty.rutgers.edu/hey/software#HKA

Linkage disequilibrium

- Certain combinations of allels in two or more loci occur in higher frequencies than we would expect based on their random combination.



haplotypes



AB…. 25%
ab…. 25%
aB…. 25%
Ab…. 25%

LINKAGE EQUILIBRIUM

AB…. 40%
ab…. 40%
aB…. 10%
Ab…. 10%

LINKAGE DISEQUILIBRIUM

Linkage disequilibrium (D)

D = pozorované – očekávané frekvence haplotypů

Expected frequencies of haplotypes (i.e. random combinations) are given only by allele frequencies.

haplotype	**expected frequency**	
AB	p_1q_1	8%
ab	p_2q_2	48%
Ab	p_1q_2	32%
aB	p_2q_1	12%

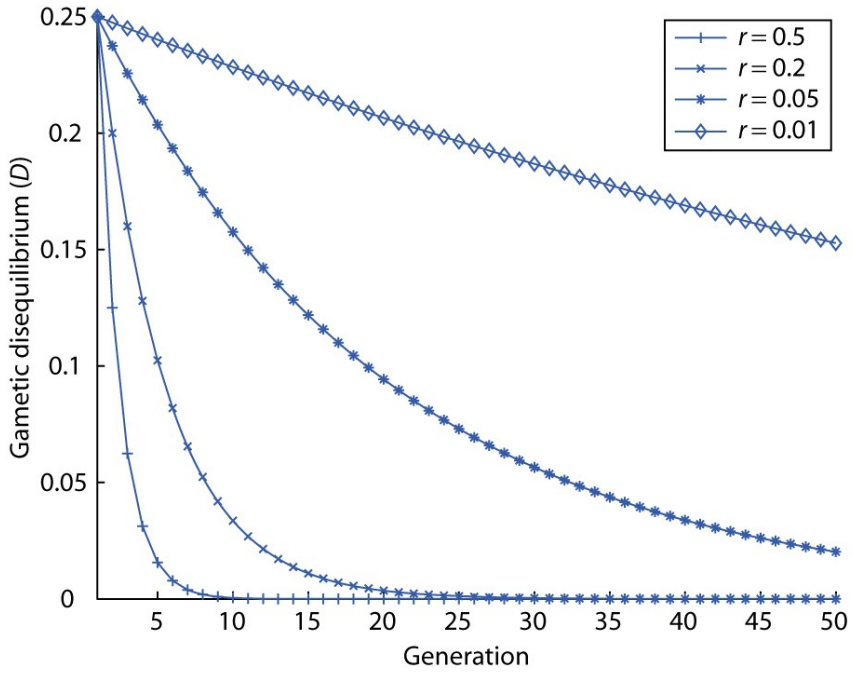
p_1 … frequency of A 40%
p_2 … frequency of a 60%
q_1 … frequency of B 20%
q_2 … frequency of b 80%

D = 0	linkage equilibrium
D > 0 či D < 0	linkage disequilibrium

D’ = D / Dmax mezi 0 a 1

Linkage disequilibrium (D)

- The level of linkage disequilibrium depends inversely on recombination rate (r) and effective population size (N_e).
- Lower levels of linkage disequilibrium in larger populations.
- Linkage disequilibrium between loci tend to decay with time (as recombination events accumulate between loci).

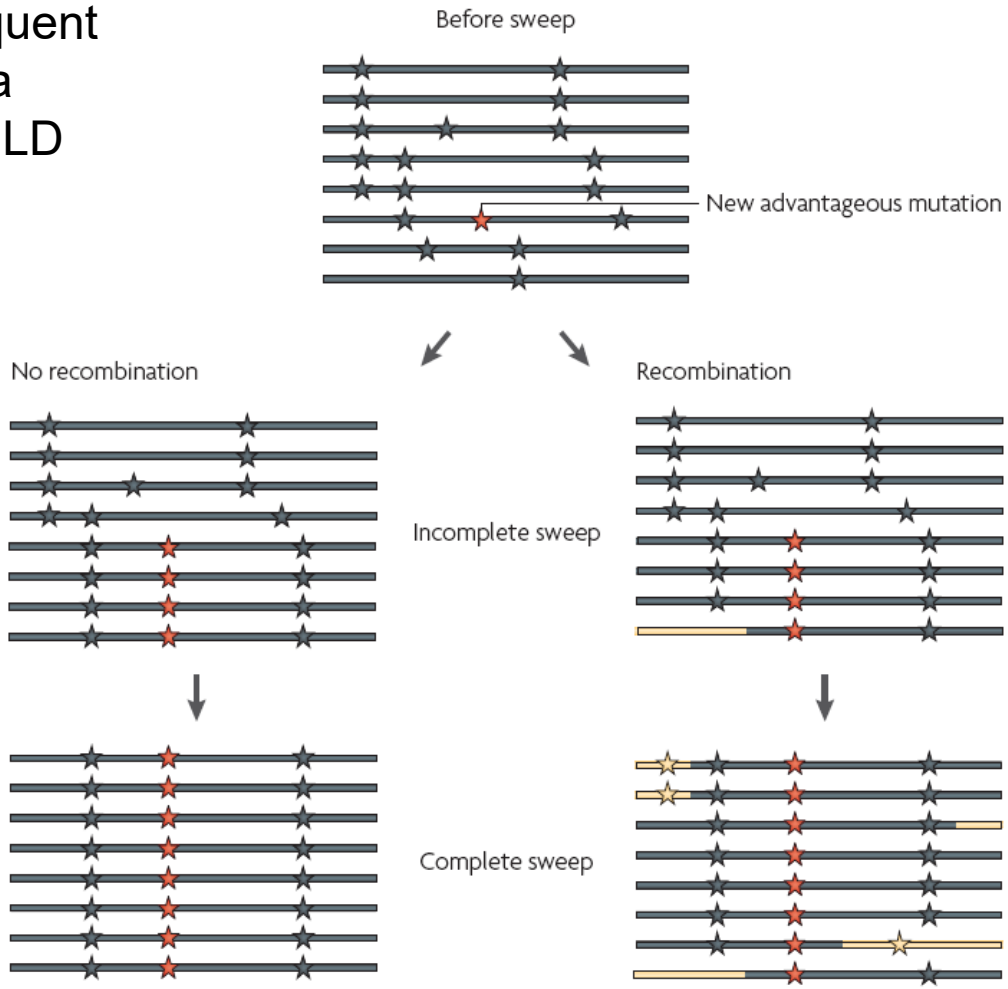


$$D = \frac{1}{4N_e r}$$

Population rekombination rate

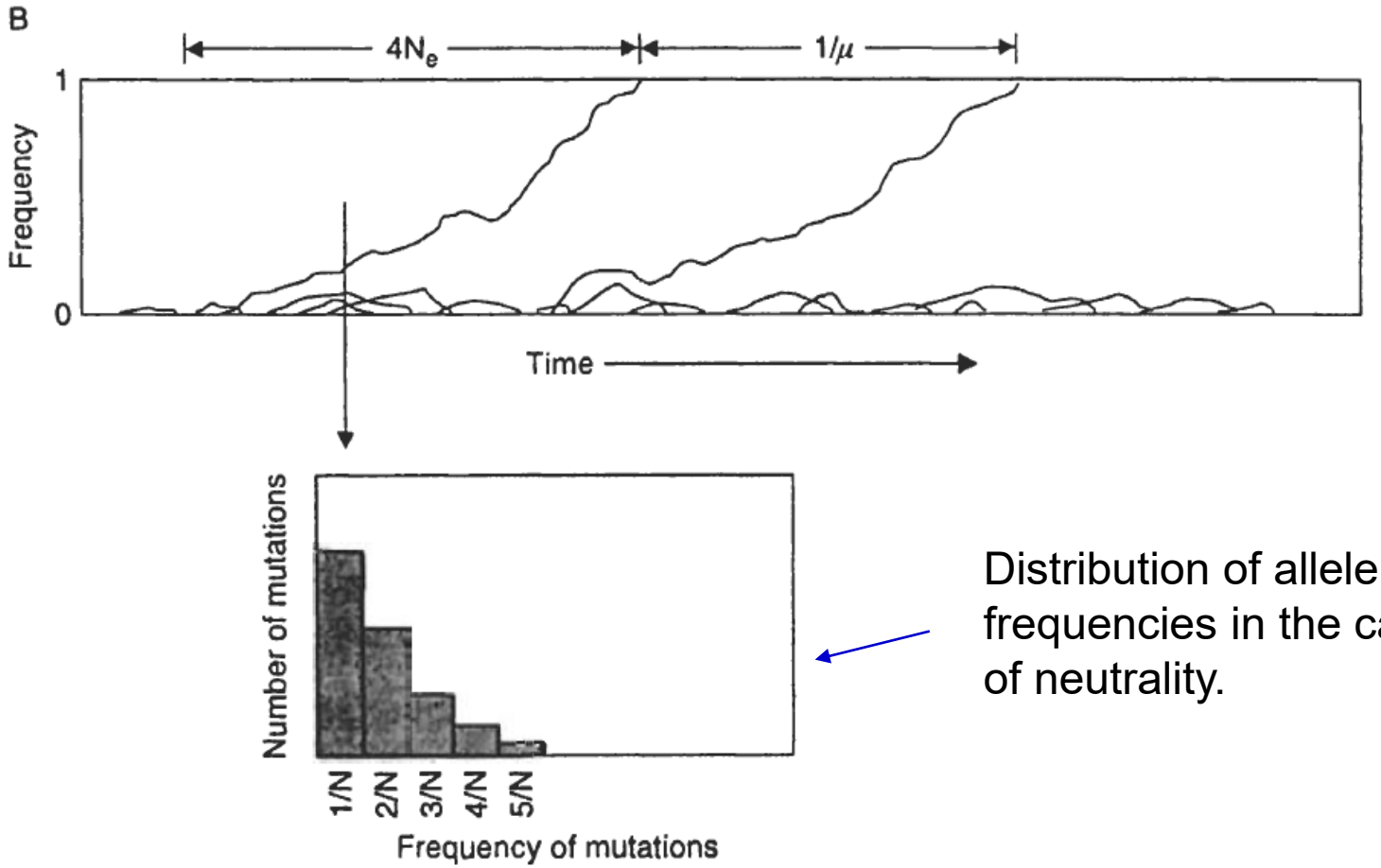
Detection of recent positive selection based on levels of linkage disequilibrium (LD)

- Ideally detection of relatively frequent haplotype (or haplotype fixed in a subpopulation) with high level of LD compared to other haplotypes.



Tajima's D test

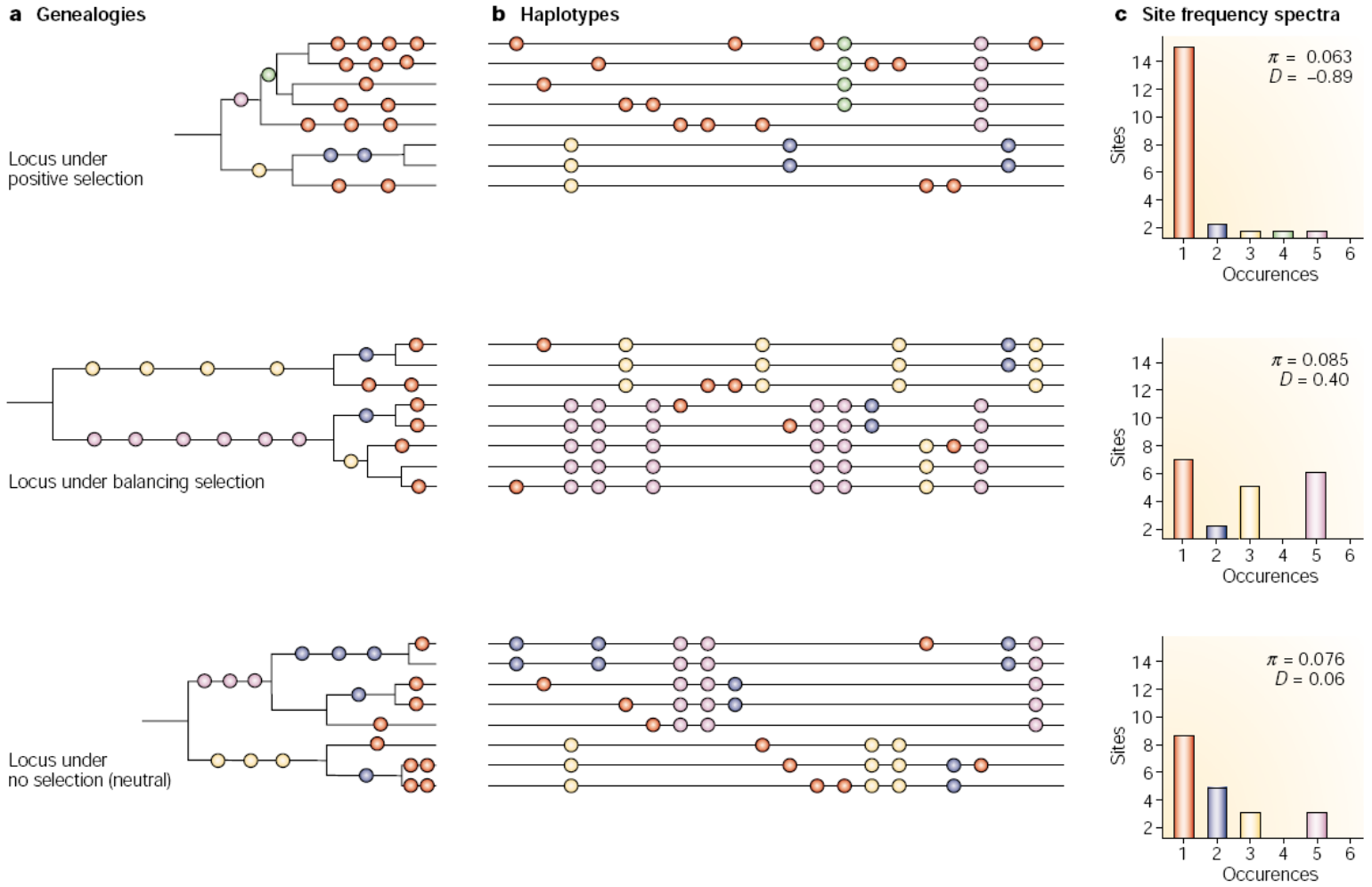
- Based on distribution of allele frequencies



Distribution of allele frequencies in the case of neutrality.

Tajima's D test

- Positive and negative selection leads to increased proportion of rare alleles.
  Balancing selection leads to the increased proportion of relatively frequent alleles.



Tajima's D test

- **Tajima's D = (π - θ)/SD (π - θ)**
- For neutral sequences: θ = π. D = 0.
- **Relatively more rare allels θ > π. D < 0. (positive or negative selection)**

  Relatively more frequent allels π > θ. D > 0. (balancing selection)

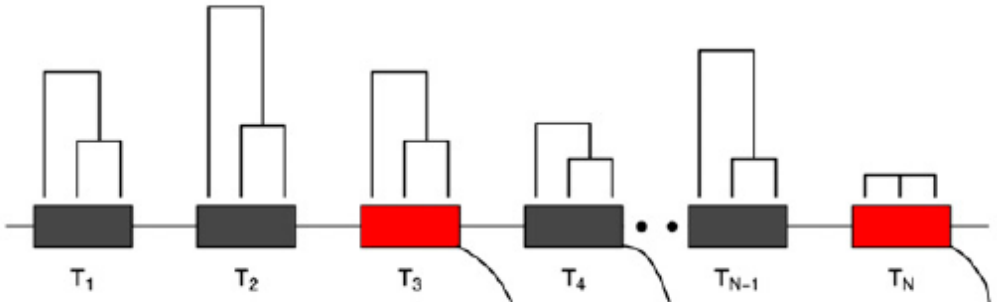
Population expansion has similar effect on allele frequency spectrum as positive and negative selection.

Bottle-neck has similar effect as balancing selection.

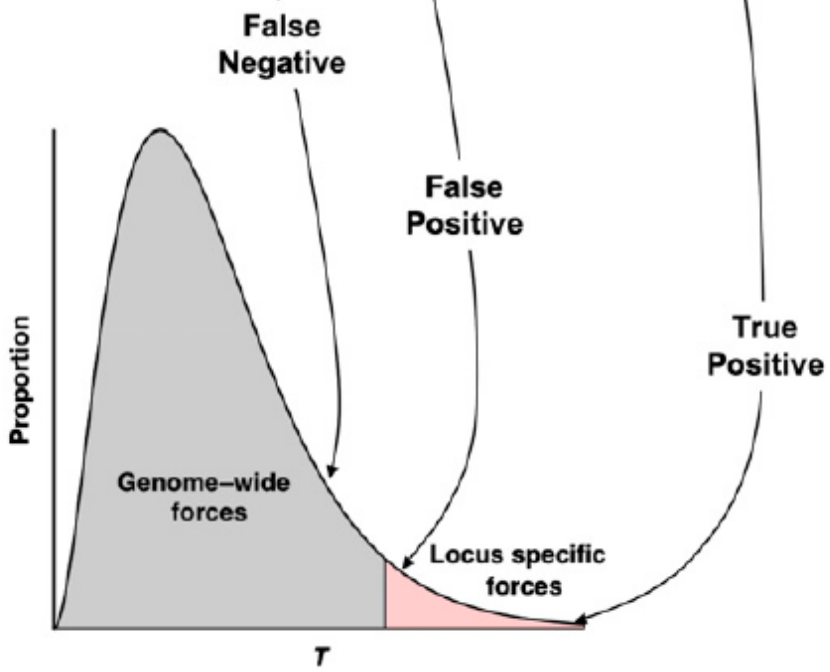
How to differentiate the effect of selection and demographic factors?

Tajima's D test for many loci.

1. Sample loci and calculate statistic (T_i)



2. Construct empirical distribution



3. Identify "outlier" loci

SweepFinder

SweepFinder

SweepFinder is a program implementing the method described in Nielsen et al. 2005. Genomic scans for selective sweeps using SNP data. Genome Research 1566-1575. It can be used to detect the location of a selective sweep based on SNP data. It will also estimate the frequency spectrum of observed SNP data in the presence of missing data.

MOLECULAR ECOLOGY

Molecular Ecology (2016) **25**, 142–156 doi: 10.1111/mec.13351

DETECTING SELECTION IN NATURAL POPULATIONS: MAKING SENSE OF GENOME SCANS AND TOWARDS ALTERNATIVE SOLUTIONS

Detecting recent selective sweeps while controlling for mutation rate and background selection

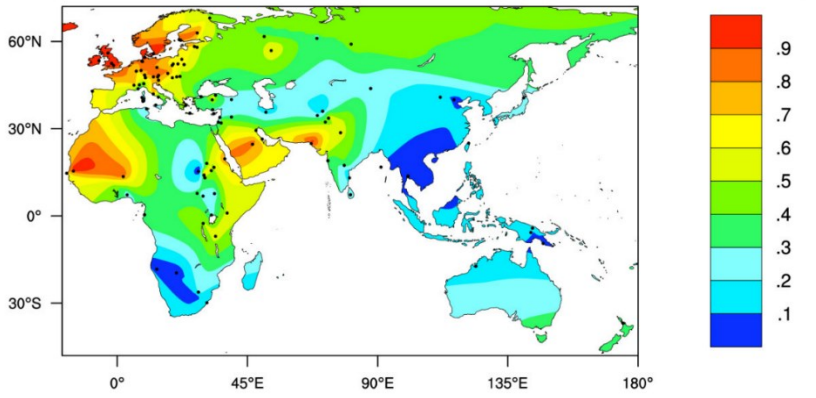
CHRISTIAN D. HUBER,*†‡ MICHAEL DEGIORGIO,§¶ INES HELLMANN** and RASMUS NIELSEN††

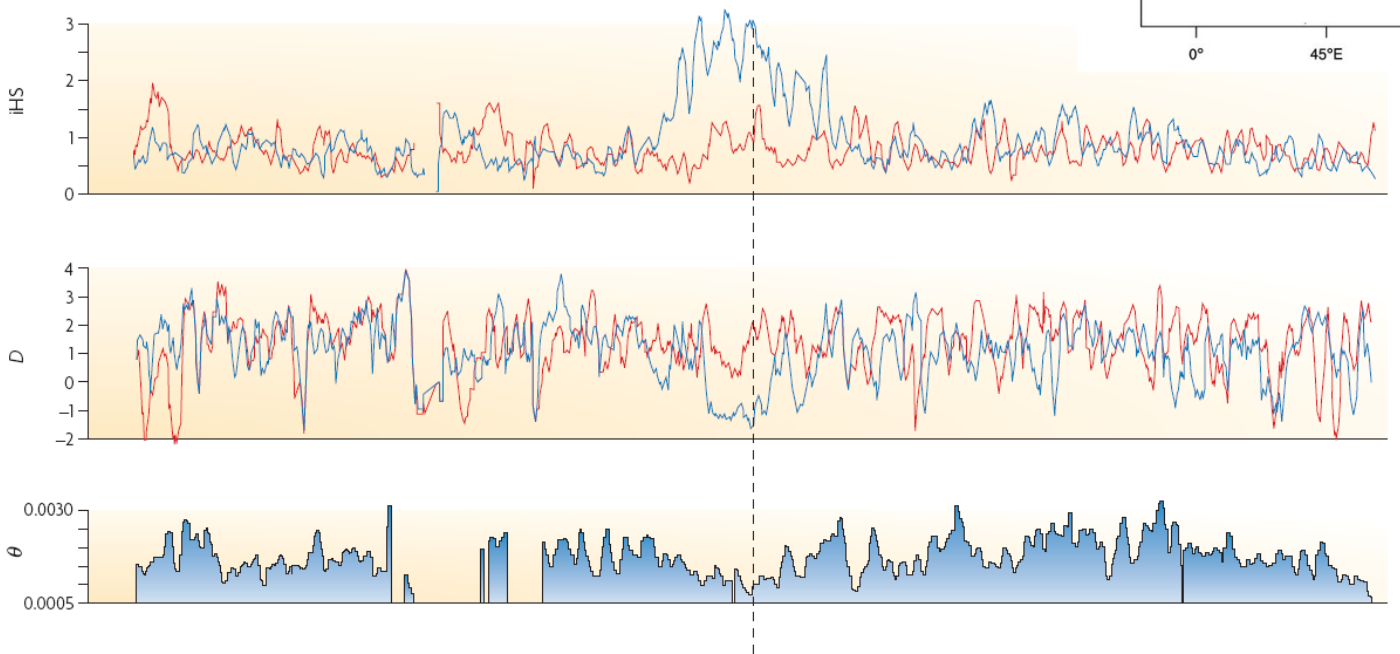
Selective sweeps in human genome

Laktase persistence

- Several independent mutations in LCT gene, allows digestion of milk in adulthood.
- Connected with the spread of parstoral farming.
- s ~1,4-15%

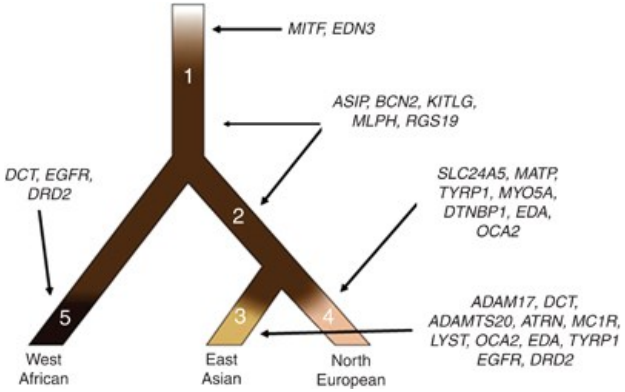
Occurrence of lactase persistence

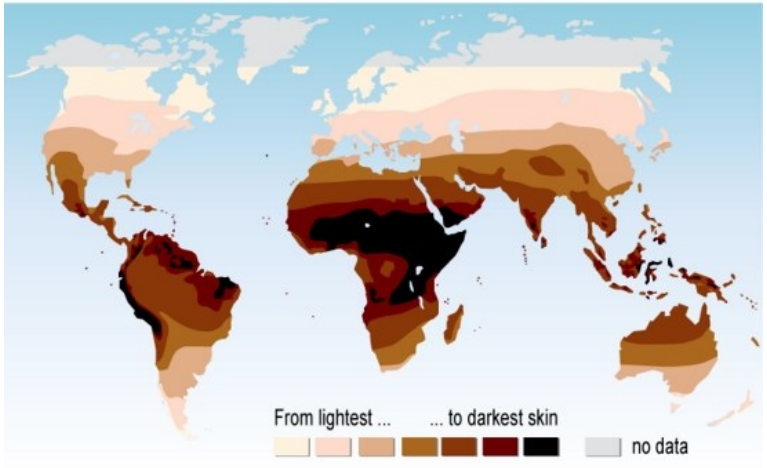




Skin color

- Genes affecting skin color in humans: HERC2, SLC45A2, TYR
- Selection for lighter skin in Europe.
- s ~ 2–10%

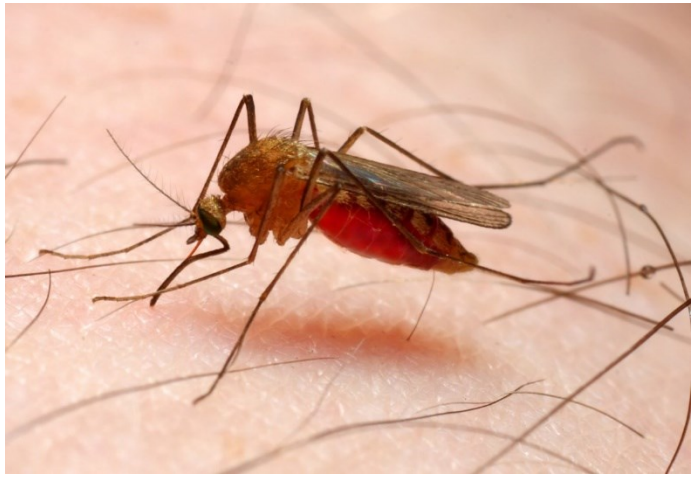


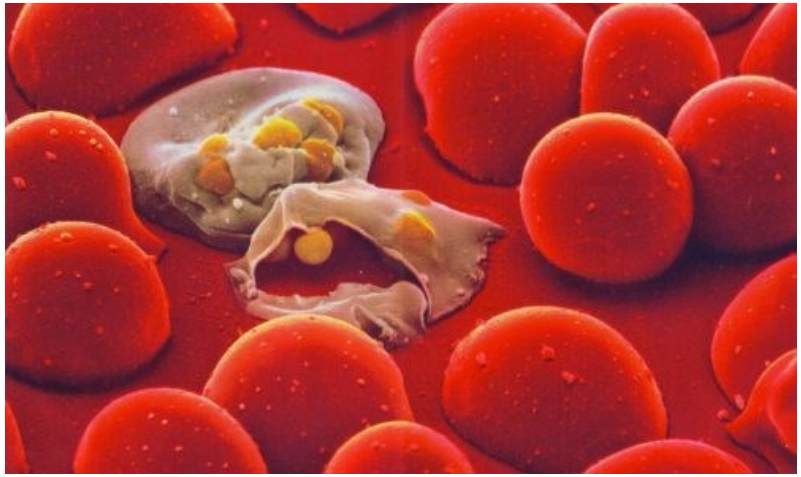


Source: Chaplin G.©, *Geographic Distribution of Environmental Factors Influencing Human Skin Coloration*, American Journal of Physical Anthropology 125:292–302, 2004; map updated in 2007.

Resistance against malaria

- Mutations in genes GYPA and GYPB coding for receptors on the red blood cells, *Plasmodium falciparum* uses to get into the cells.





Leffler et al. 2017

Local adaptations

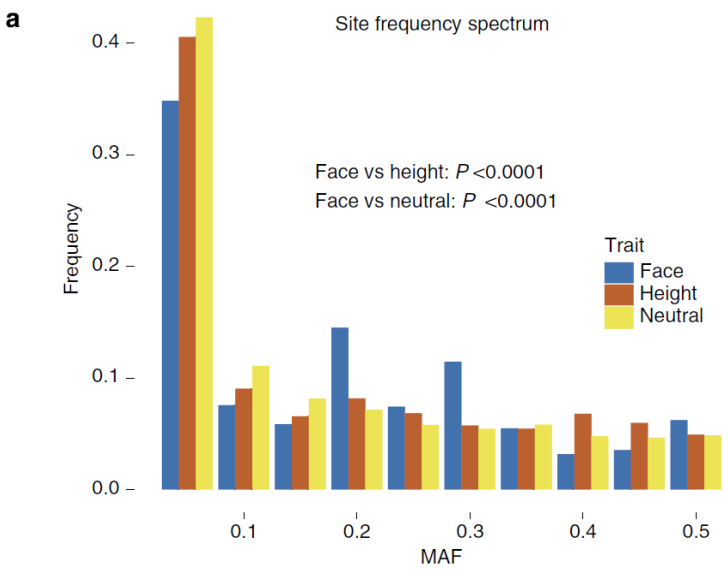
Inuits in Greenland

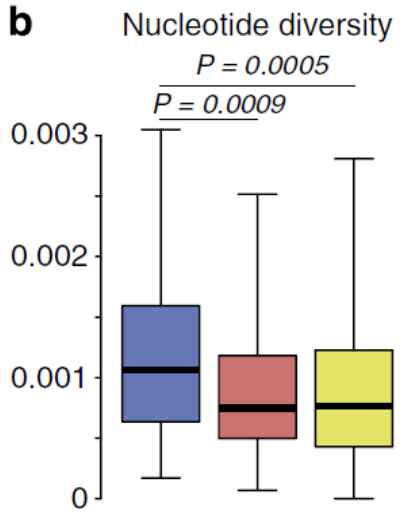
- Selective sweeps around genes FADS1, FADS2. Desaturases, metabolism of fatty acids.

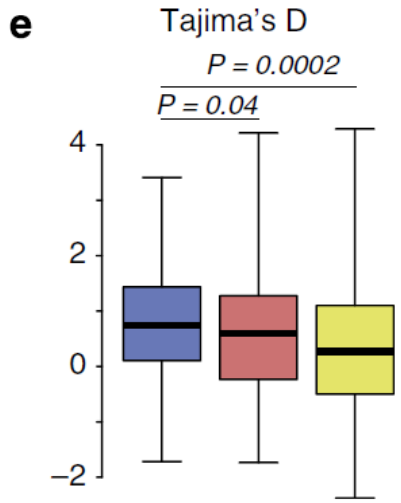
Fumagalli et al. 2015

Balancing selection

Maintains variability in human faces







Sheeman and Nachman. 2014. Nature.

Among positively selected genes there are also genes causing diseasses

- *CFTR* gene (cystic fibrósis)
- ALMS1 gene (Alstrom syndrom)
- *GBA* gene *(*Gaucher diseases*)*
- PCDH15 (Uscher syndrom)

etc.

How frequency of allele causing disease can be affected by positive selection?

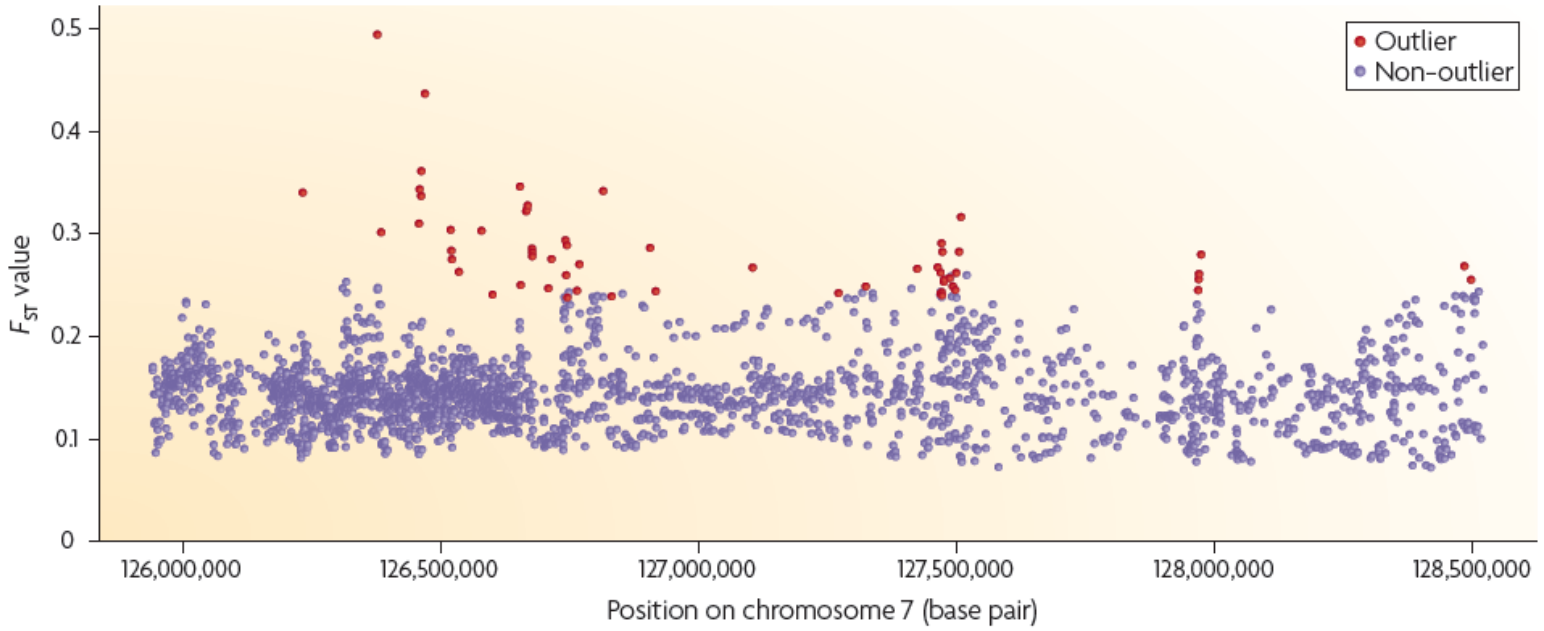
1. Mutations causing the disease are recessive and in heterozygote state they can bring some advantage (balancing selection).
2. Change of selection pressure during time (cyclical selection)

Detection of genes for local adaptations based on levels of genetic differentiation

- Genes for local adaptations have increased levels of genetic differentiation (F_{ST}) between populations.

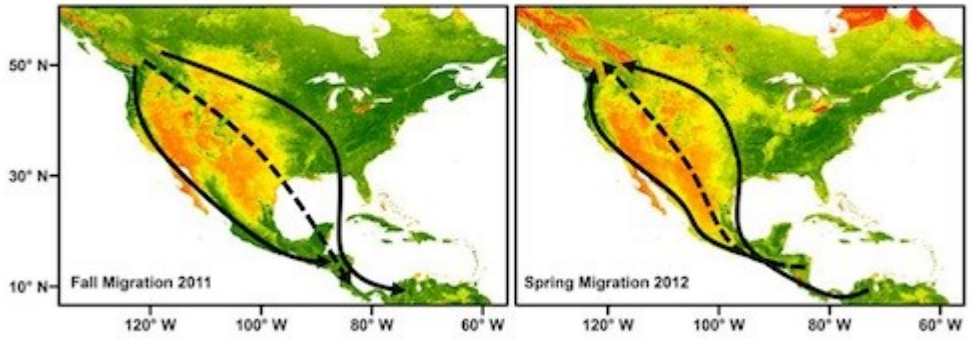
Genomic islands of increased F_{ST}

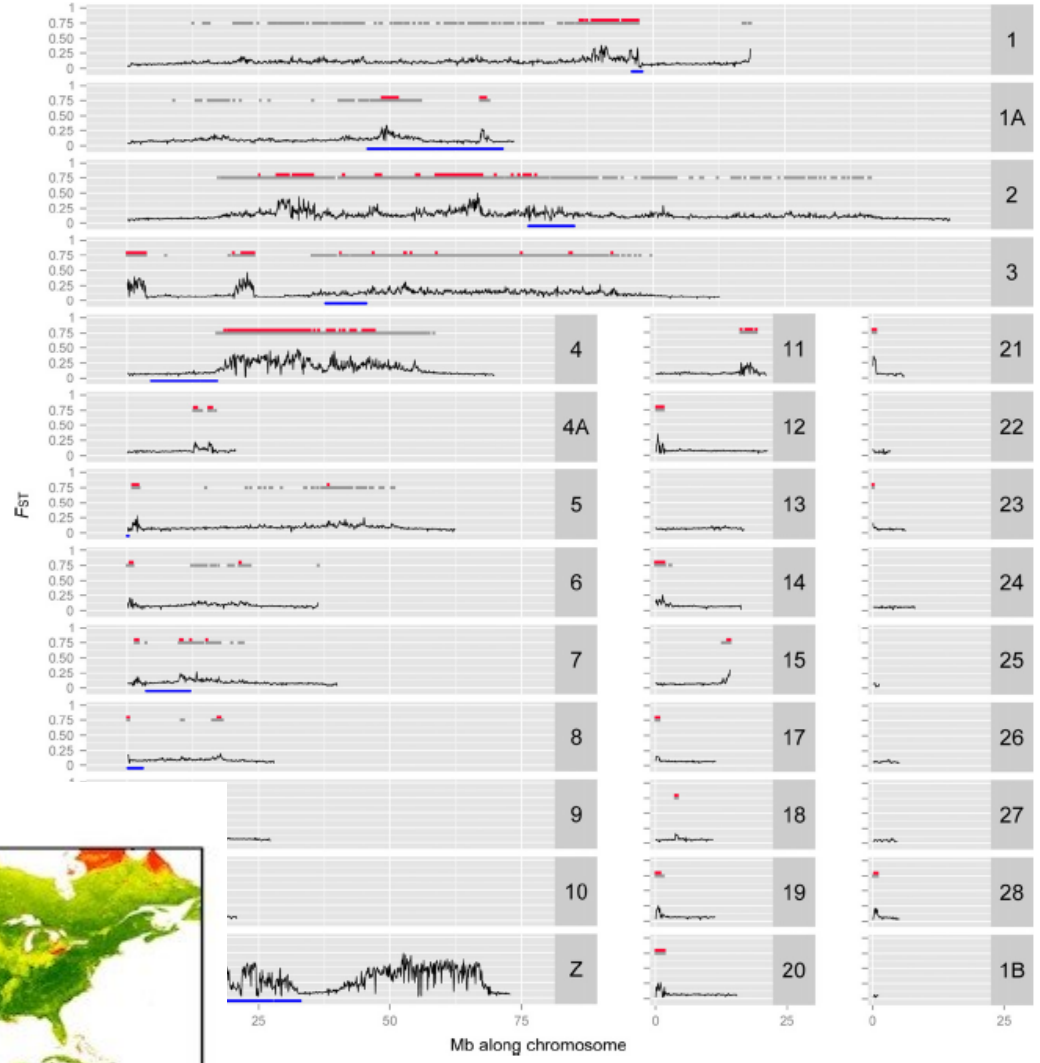
- Positive selection leads to incresed differences in allele frequencies between populations.



Islands of differentiation in Swainson's thrush (*Catharus ustulatus*)

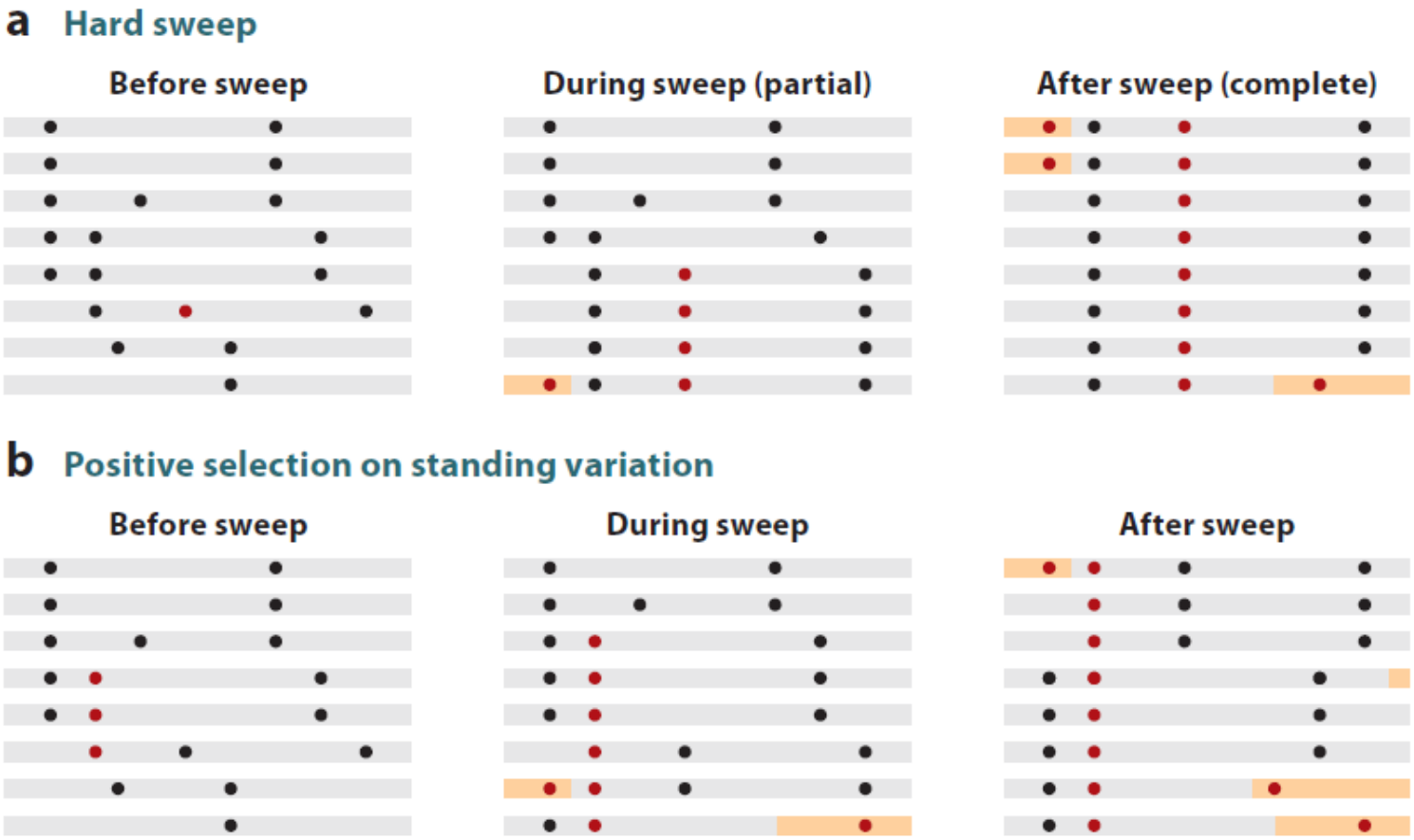
- Different migration routes





Delmore et al. 2015

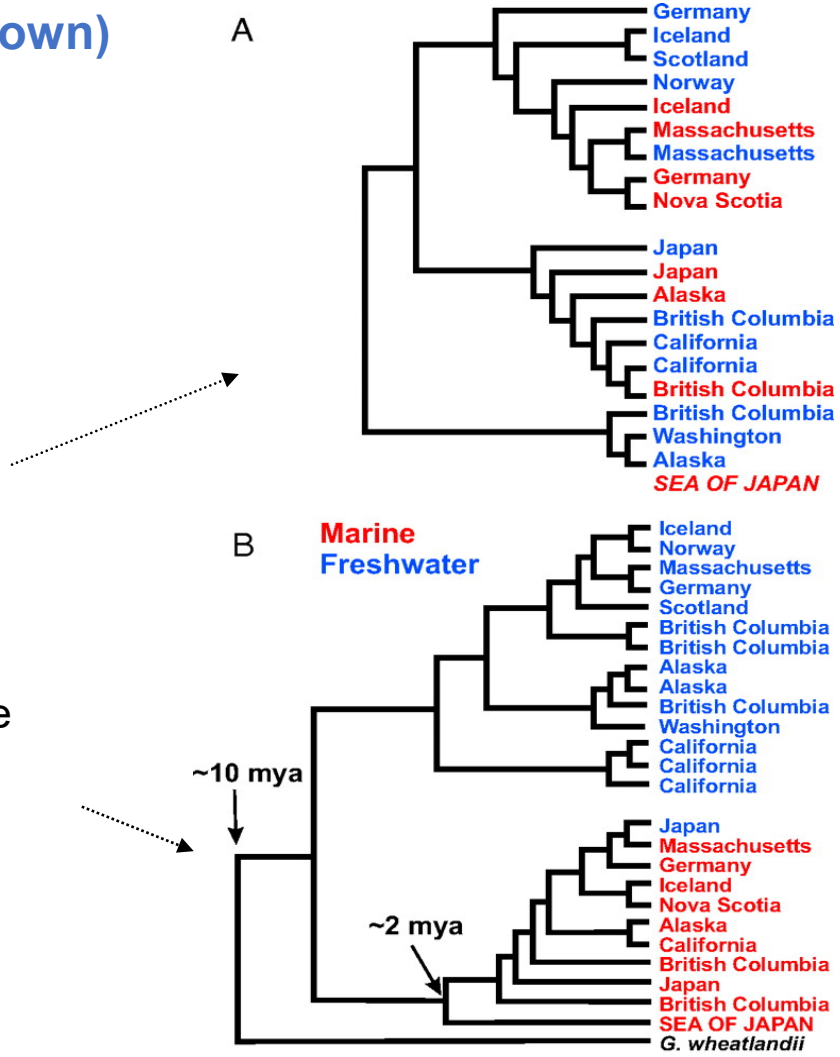
Adaptation from standing genetic variation (soft selective sweep)



Fu and Akey, 2013

Adaptation from standing genetic variation and paralel evolution

Three spine stickleback
marine form (up) a **freshwater form (down)**



- Freshwater form arose multiple times independely by colonization of new rivers.

- Phylogeny based on *Ectodisplasin* (*Eda*) gene underlying some differences between forms.
- The same pattern holds for other genes underlying phenotypic differences between forms. These genes often in inversions.

Jones et al. Nature. 2012

Detection of reccurent positive selection

Repeated fixation of advantageous mutations in the same locus increases the ratio between non-synonymous substitutions (K_A) and synonymous substitutions (K_S).

$K_A/K_S = 1$ neutral evolution

$K_A/K_S > 1$ positive selection

"The Red Queen has to run faster and faster in order to keep still where she is. That is exactly what you all are doing!"

Detection of reccurent negative selection

Longterm negative selection decreases the ratio between non-synonymous substitutions (K_A) and synonymous substitutions (K_S).

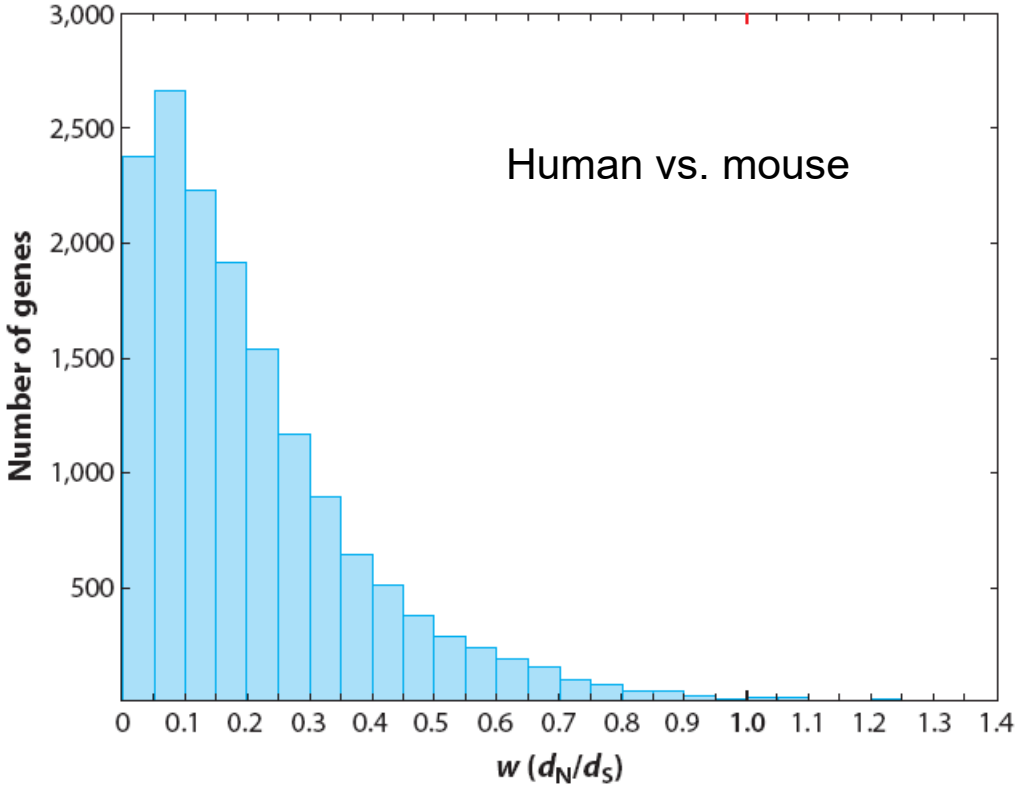
$K_A/K_S = 1$ neutral evolution
$K_A/K_S < 1$ negative selection

Ultra conserved genetic elements:
extremely low substitution rate for non-synonymous substitutions.

K_A/K_S test

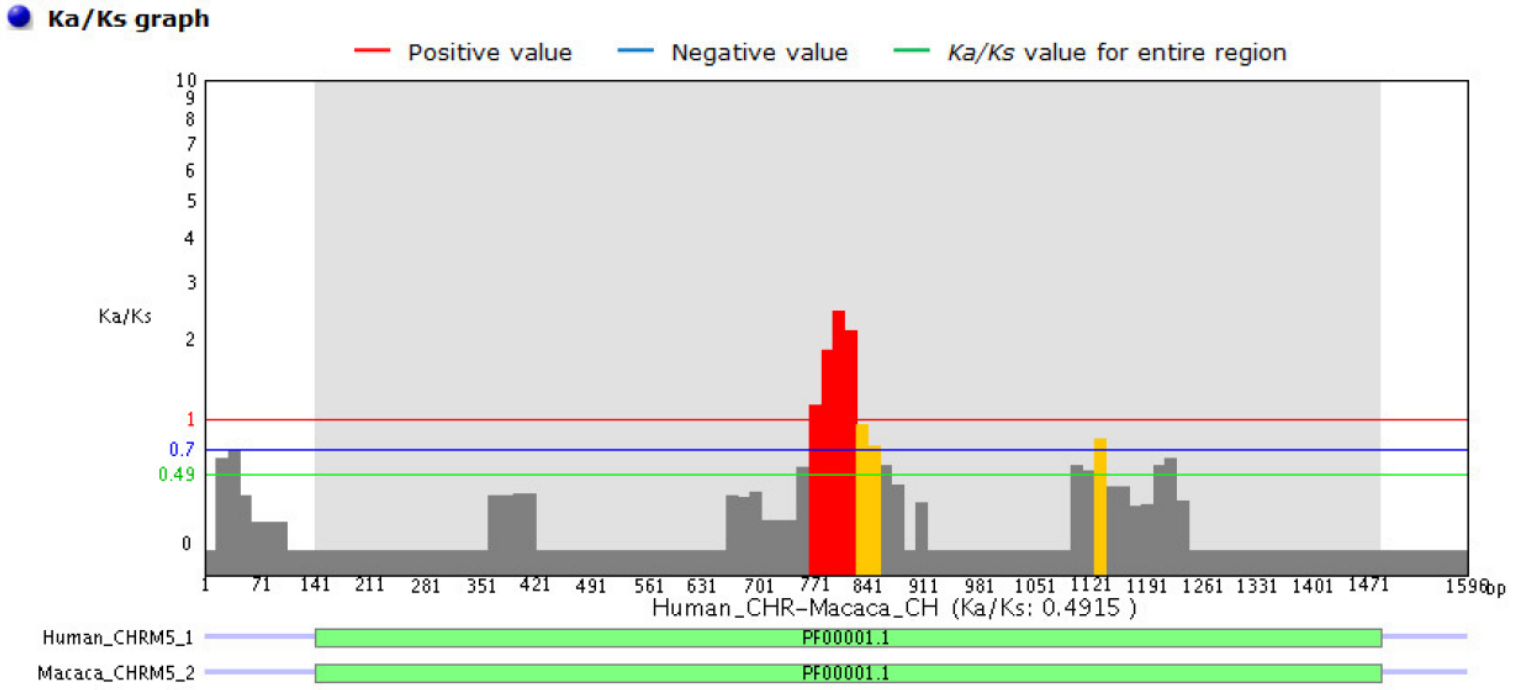
- Most genes have K_A/K_S around 0,1 – 0,2.
  Under negative selection.

- Genes with high K_A/K_S often associated with **reproduction**, **imunity**, in mammals in **olfactory sense** (e.g. OBP genes).



K_A/K_S test – site specific analysis

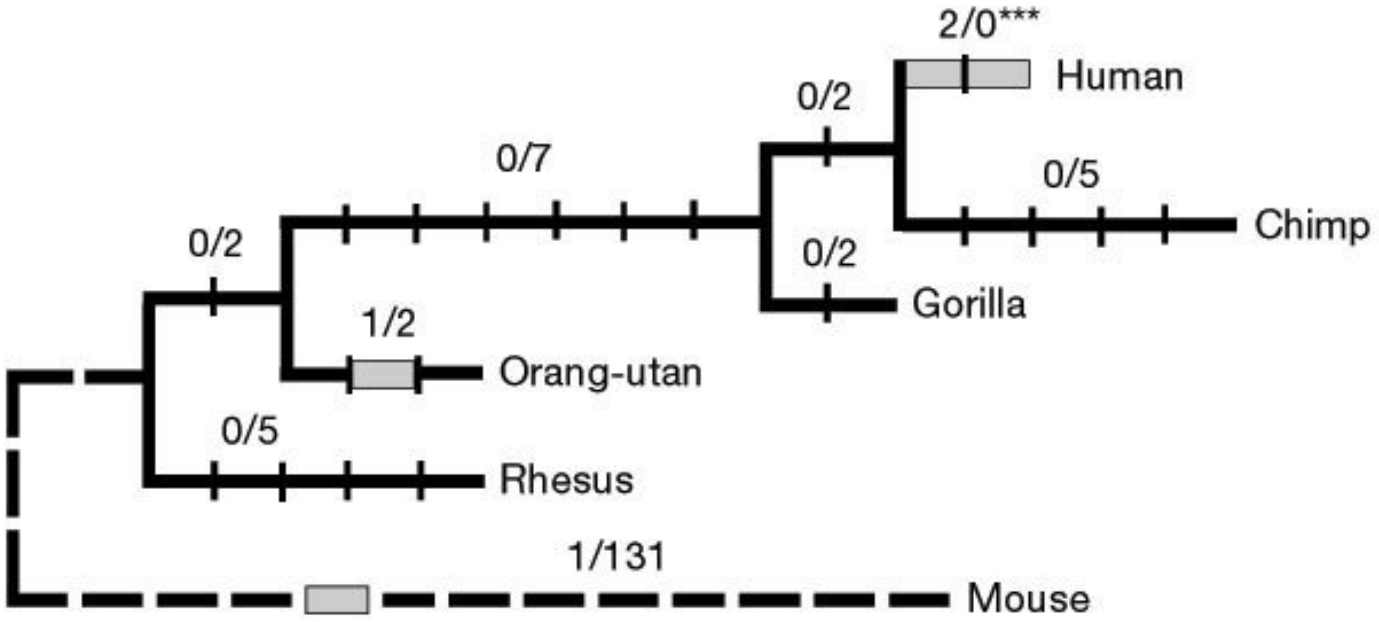
- Sliding window analysis – allows to calculate K_A/K_S for different parts of the genes.
- K_A/K_S can be calculated for individual codons if we compare multiple species.



K_A/K_S test – lineage specific analysis

FOXP2 gene for speach, positive selection in human lineage

No. of nonsynonymous/ synonymous substitutions.



Enard W. 2002. Nature

McDonald-Kreitman (MK) test

- Compares numbers of nonsynonymous and synonymous substitutions within species (P) and between species (D) for individual genes.
- Neutral genes Ds/Ps = Dn/Pn.
- Positively selected genes Dn/Pn >> Ds/Ps.

	D	P
Synonymous	Ds	Ps
Nonsynonymous	Dn ↑	Pn

Positive selection

MK test allows to estimate proportion of aminoacid substitutions driven by positive selection (α).

Organisms with higher Ne have higher α.
Selection in larger populations is more efficient.

- Drosophila (Ne ~ 10^6) α > 50%
- Mus musculus castaneus (Ne ~ 5 x 10^5) α ~ 40-60%
- Topol osika (Ne ~ 10^5) α ~ 30-40%
- Mus musculus domesticus (Ne ~ 10^5) α ~ 13%
- Human (Ne ~ 10^4) α ~ 10 – 20%

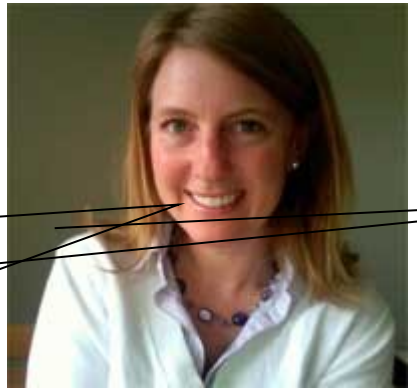
Genetic basis of adaptations

Importance of coding vs. regulatory changes

Sean B. Carroll

Mutations in regulatory regions of genes are more important. Have lower pleiotropic effects.

Most so far detected mutations underlying adaptations are in coding regions of the gene.



Hopi E. Hoekstra

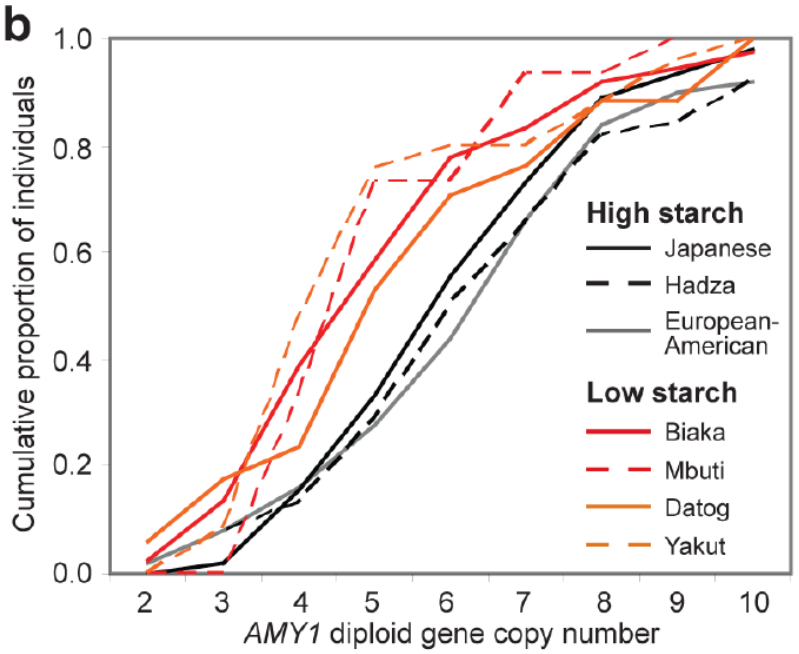
Jerry A. Coyne

Hoekstra HE and Coyne JA (2007).
The locus of evolution:evo devo and the genetics of adaptation. *Evolution.*

Adaptive evolution by gene duplications

Salivary amylase gene (*AMY1*)

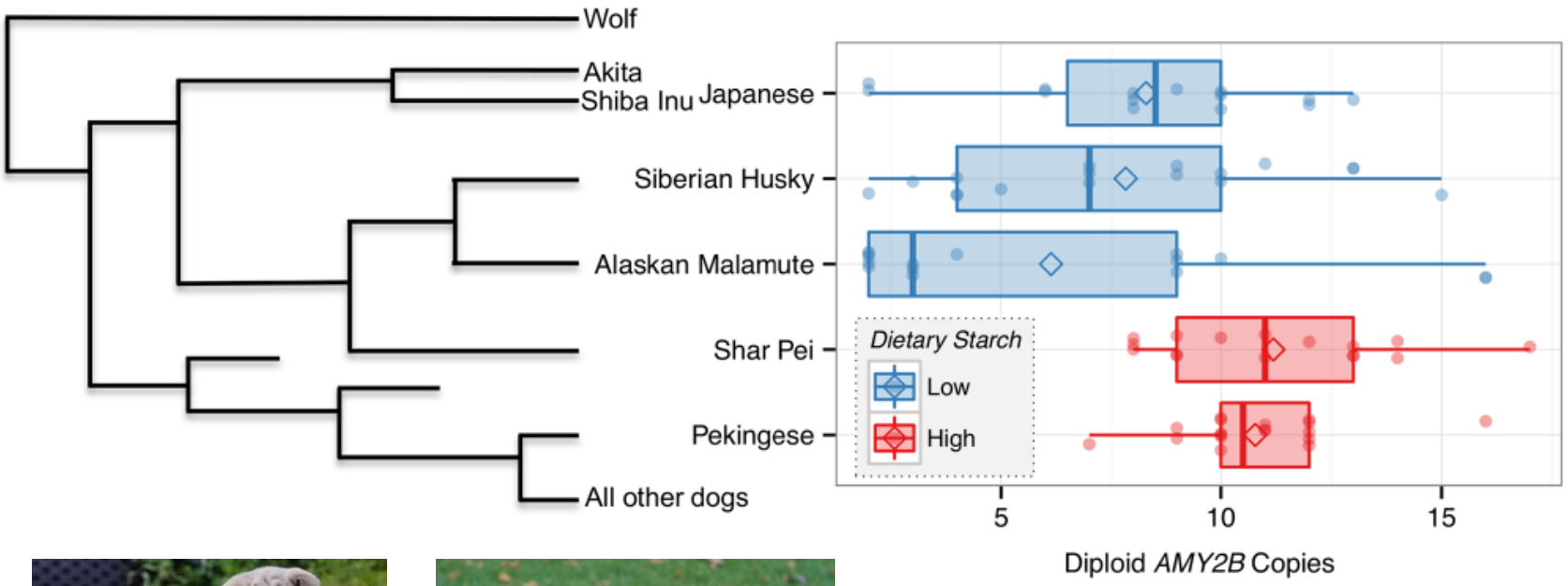
- Digest starch.
- One of the most variable gene in terms of gene copies in the genome.
- More copies of AMY1 gene more amylase enzyme in saliva.
- Populations with starch rich diet have higher number of AMY1 gen copies.



Perry et al. 2007, Nature Genetics

Number of AMY2B amylase gene copies is related to the diet of domestic dogs




Reiter et al. 2016

Adaptations underlied by epigenetic changes

nature ecology & evolution

ARTICLES

https://doi.org/10.1038/s41559-018-0569-4

An epigenetic mechanism for cavefish eye degeneration

Aniket V. Gore[1*], Kelly A. Tomins[1], James Iben[2], Li Ma[3], Daniel Castranova[1], Andrew E. Davis[1], Amy Parkhurst[1], William R. Jeffery[3] and Brant M. Weinstein[1*]

Coding and non-coding mutations in DNA contribute significantly to phenotypic variability during evolution. However, less is known about the role of epigenetics in this process. Although previous studies have identified eye development genes associated with the loss-of-eyes phenotype in the Pachón blind cave morph of the Mexican tetra *Astyanax mexicanus*, no inactivating mutations have been found in any of these genes. Here, we show that excess DNA methylation-based epigenetic silencing promotes eye degeneration in blind cave *A. mexicanus*. By performing parallel analyses in *A. mexicanus* cave and surface morphs, and in the zebrafish *Danio rerio*, we have discovered that DNA methylation mediates eye-specific gene repression and globally regulates early eye development. The most significantly hypermethylated and downregulated genes in the cave morph are also linked to human eye disorders, suggesting that the function of these genes is conserved across vertebrates. Our results show that changes in DNA methylation-based gene repression can serve as an important molecular mechanism generating phenotypic diversity during development and evolution.