Neutrality tests: Detecting selection based on patterns of polymorphism

Hancock

March 26, 2024

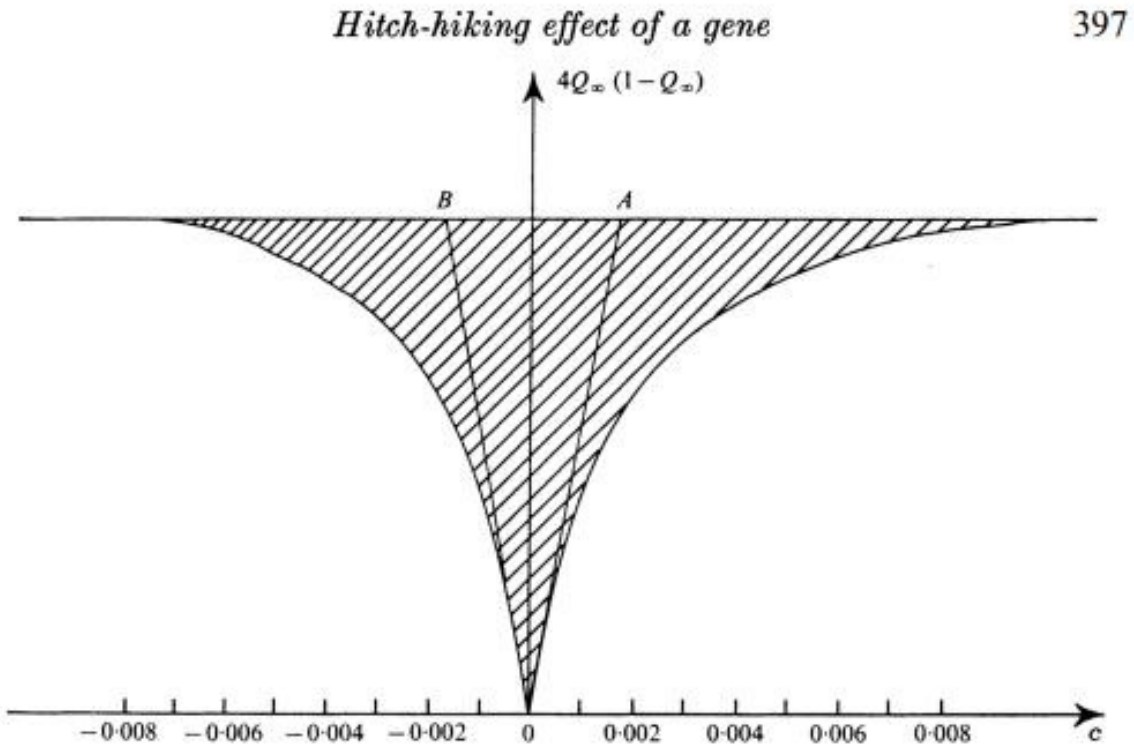


Fig. 2. $4Q_\infty(1-Q_\infty)$ is the final amount of heterozygosity at a locus, when initial frequencies of a, A are 0·5. The graph here, with $N = 10^6$ and $s = 0{\cdot}01$, is calculated from (8).

Evolutionary processes

The factors that alter allele frequencies and affect patterns of polymorphism from one generation to the next

- Genetic drift
- Gene flow

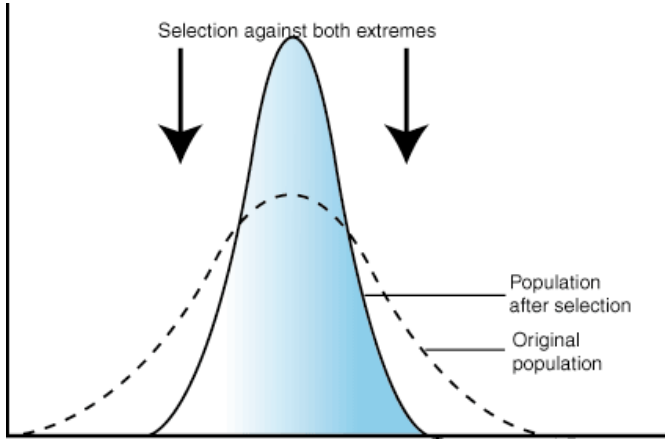
← - - - - Neutral processes

- Selection
  - Positive selection ← Adaptive evolution
  - Negative selection ← Stabilizing selection

evolution.berkeley.edu/evolibrary/article/evo_24

Selection can affect traits in different ways

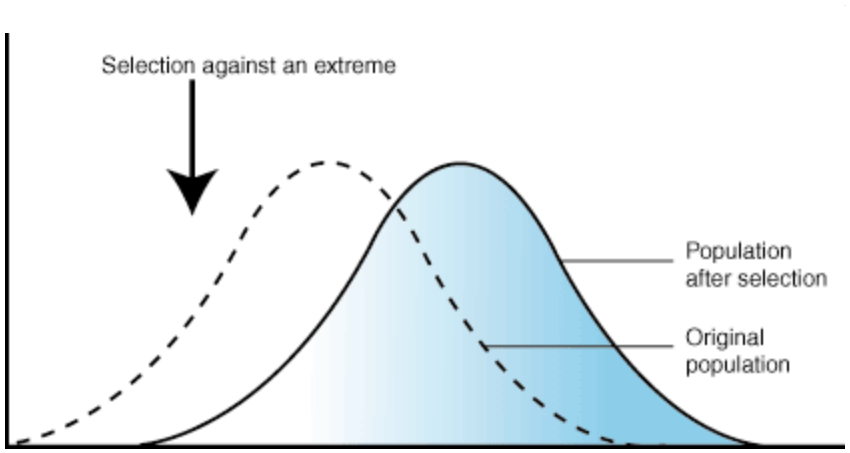
Stabilizing selection



Selection that removes variation from the population

The most common form of selection

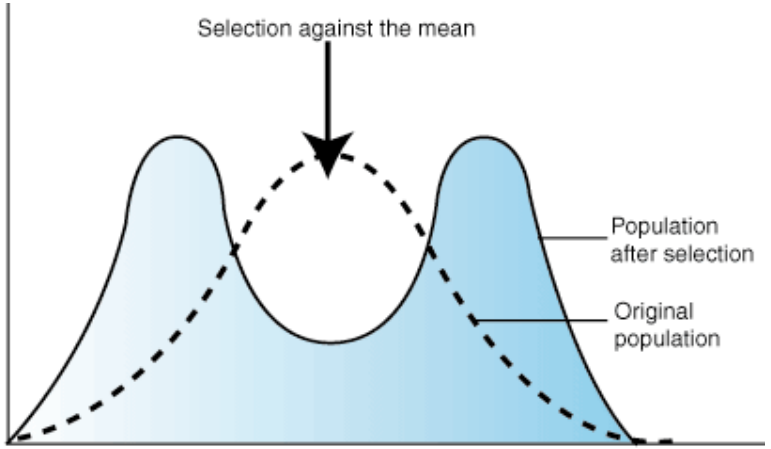
Directional selection



Selection that changes the mean trait value in the population

Type of selection responsible for adaptation to a novel environment

Disruptive selection/ Balancing selection



Selection that maintains variation in the population

Selection that occurs when there are differences in pressures across time or space

Reconstructing evolutionary history from DNA sequence data

process → pattern

Selection and demographic events

pattern in DNA sequence variation

reconstruction ←

Population genetics uses these patterns to reconstruct the evolutionary history

What processes are involved in adaptation?

- Adaptive divergence on the lineage leading to a species
- Adaptation through subtle allele frequency changes
- Selective sweeps

What patterns do these processes leave in data?

How do we assay the genome for signatures of selection?

Summary statistics can be used to quantify the pattern at a locus

- Functional genetic divergence from relatives
- Allele frequencies differ among populations (local adaptation with population-specific sweep)
- Changes in the frequency spectrum
- Variation is reduced across a haplotype

Adaptive footprints

Tests based on divergence

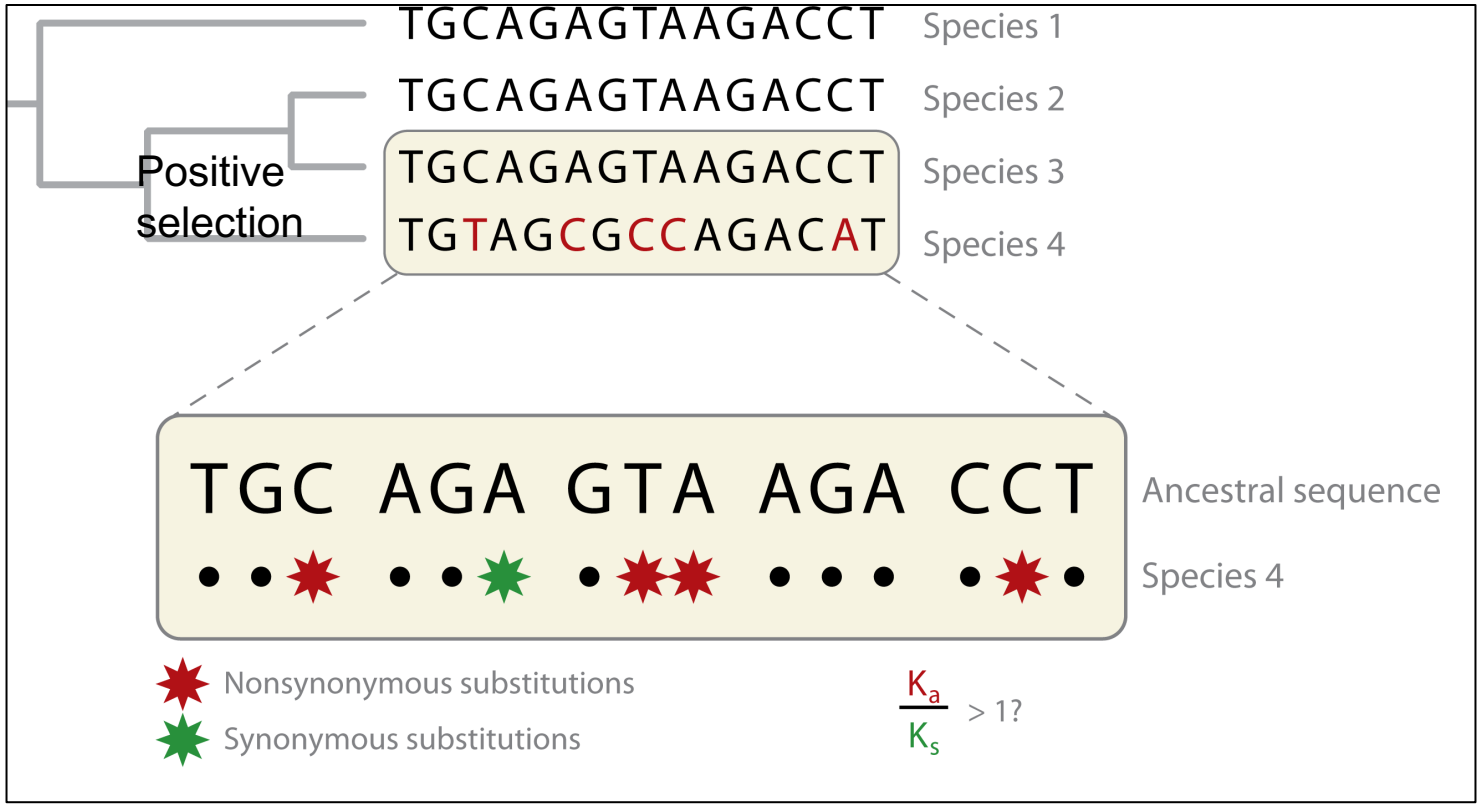
process → pattern

Adaptation (positive selection)

increased divergence of selected sites

Tests to identify evidence for positive selection on the lineage leading to a species

Identify loci with a high rate of functional evolution in a species



Use the ratio of NS to S variants to detect rapidly evolving loci across species

dN/dS ratio

Differences in fixation probabilities of selected and neutral alleles:

$$\frac{d(non-neutral\ class)}{d(neutral\ class)} \rightarrow \frac{d(non-synonymous)}{d(sysnonymous)} = \frac{dN}{dS}$$

- Positive selection leads to more non-neutral substitutions: $\frac{dN}{dS} > 1$
- Negative selection leads to fewer non-neutral substitutions: $\frac{dN}{dS} < 1$

dN/dS ratio

Ask whether: $$\frac{dN}{dS} > 1$$

- Robust test, but tends to be conservative:
  - Requires multiple adaptive fixations in a gene
  - Adaptive change must outweigh purifying selection
- Method detects rapidly evolving genes. Mostly genes involved in arms races, e.g., immunity genes, reproductive competition
- Caveat: over long time scales, repeat mutations can occur at some synonymous sites (saturation)

McDonald-Kreitman test
Divergence vs polymorphism

Improvement over dN/dS:
Normalize divergence by polymorphism to control for different rates of evolution among sites

<u>Logic</u>:
If all segregating or fixed mutation are neutral, then the proportion of fixed differences that are nonsynonymous should be the same as the proportion of segregating mutations that are nonsynonymous

McDonald-Kreitman test

Divergence vs polymorphism

	Between species	Within species
NS	dN	pN
S	dS	pS

Accounts for purifying selection using polymorphism relative to divergence

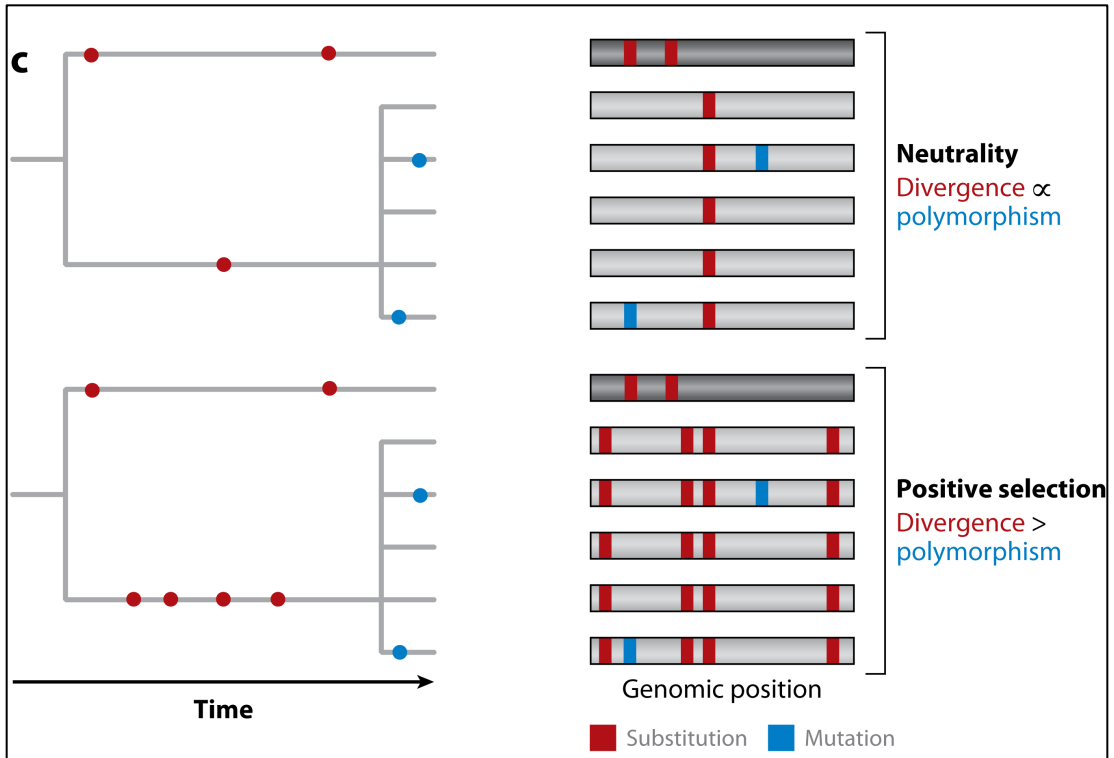
$$\frac{\#\ non-synonymous\ polymorphisms}{\#\ sysnonymous\ polymorphisms} = \frac{pN}{pS}$$

Weak negative selection: $\frac{dN}{dS} < \frac{pN}{pS}$

Positive selection: $\boxed{\frac{dN}{dS} > \frac{pN}{pS}}$

Test for significance using χ^2 test

MK uses information about diversity and divergence within and between species



Vitti JJ, et al. 2013.
Annu. Rev. Genet. 47:97–120

McDonald-Kreitman test
Divergence + Polymorphism

$$\boxed{\frac{dN}{dS} > \frac{pN}{pS}}$$

- powerful test framework
- $\alpha = 1 - \frac{dS}{dN}\frac{pN}{pS}$ *proportion of adaptive AA substitutions* (estimates: 10-20% in humans, 50% in fruitflies)
- Still requires multiple adaptive fixations
- Can also use this family of to look for signals for other putative functional sites
- False positives possible for growing populations

McDonald-Kreitman test: genome-wide example in Drosophila

Assess adaptive significance of non-coding DNA changes in *D. melanogaster*

Table 2 | Functionally relevant nucleotides in non-coding DNA

Class	C (%)*	α(%)†	p ($\alpha \leq 0$)‡	FRN (%)§
UTRs	60.4	57.5	$<10^{-3}$	83.2
5′UTRs	52.9	60.8	$<10^{-3}$	80.9
3′UTRs	70.7	52.9	$<10^{-3}$	86.2
Introns	39.5	19.3	0.007	51.2
IGRs	49.3	15.3	0.036	57.1
pIGRs	40.6	11.4	0.165	47.4
dIGRs	54.6	18.5	0.019	63.0
Introns + IGR	44.2	17.6	0.013	54.0

*Constraint (C) is estimated relative to fourfold degenerate synonymous sites.
†α is the estimated fraction of divergence driven by positive selection.
‡Probabilities ($\alpha \leq 0$) have been adjusted for effects of linkage within loci (see Supplementary Materials 2.5).
§FRN is the inferred fraction of functionally relevant nucleotides given levels of constraint and α (that is, FRN $\approx C + (1 - C)\alpha$).

Takehome:
Non-coding DNA
is not junk!

Andolfatto, 2005

Reconstructing adaptive history *within species*

process → pattern

Adaptation (positive selection)

among populations

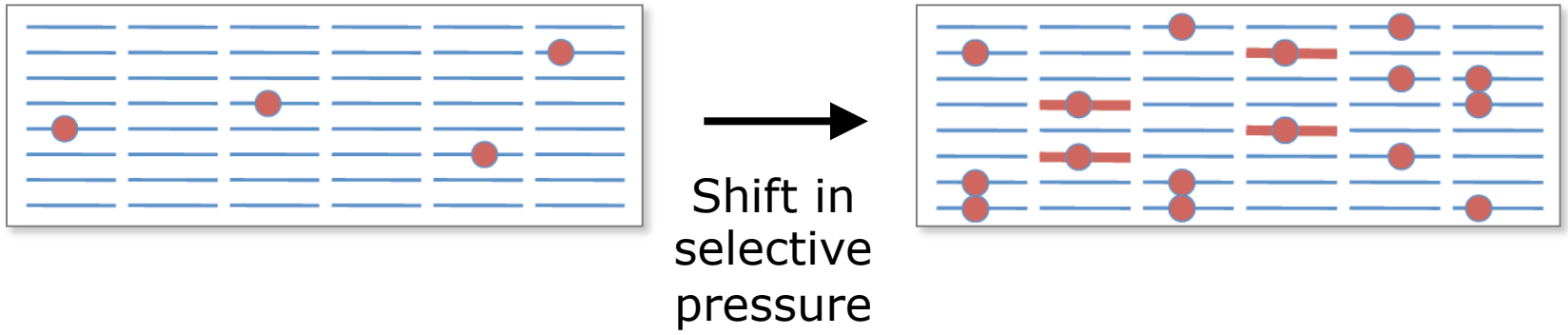
within populations

increased divergence and differentiation

- reduced polymorphism
- changes in the SFS
- increased LD

Models of adaptation: polygenic selection model

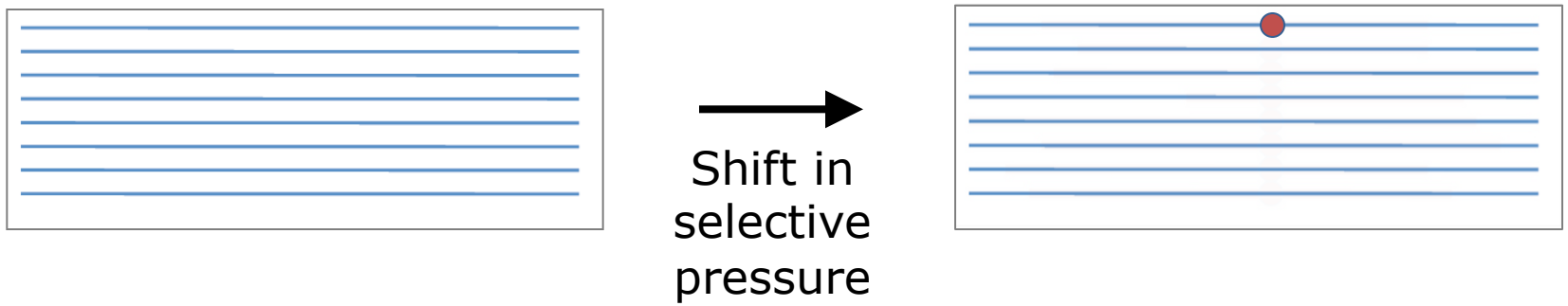
A polygenic model of selection:



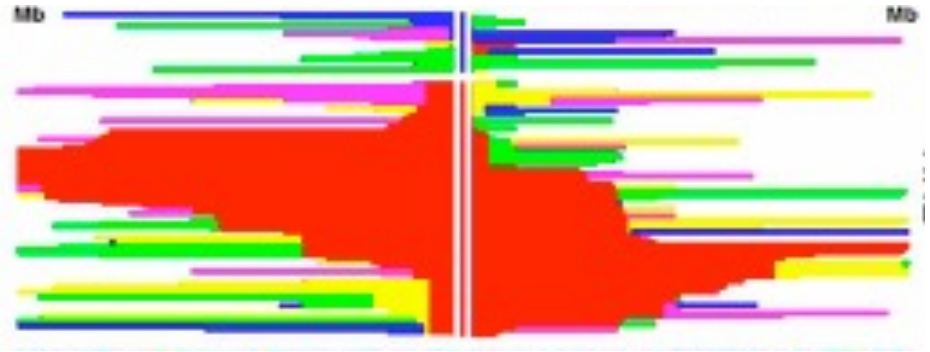
Polygenic selection may result in subtle shifts in frequencies at many loci, most of which were present in the population when the selection pressure arose

Models of adaptation: Hard sweep model (hitch-hiking)

Hard sweep model of adaptation:



Haplotype structure can be used to identify regions implicated in *hard sweeps*



When selection acts differentially across populations, identifying regions of increased differentiation can be a powerful approach

Reconstructing adaptive history *among populations within species*

process → pattern

Adaptation (positive selection)

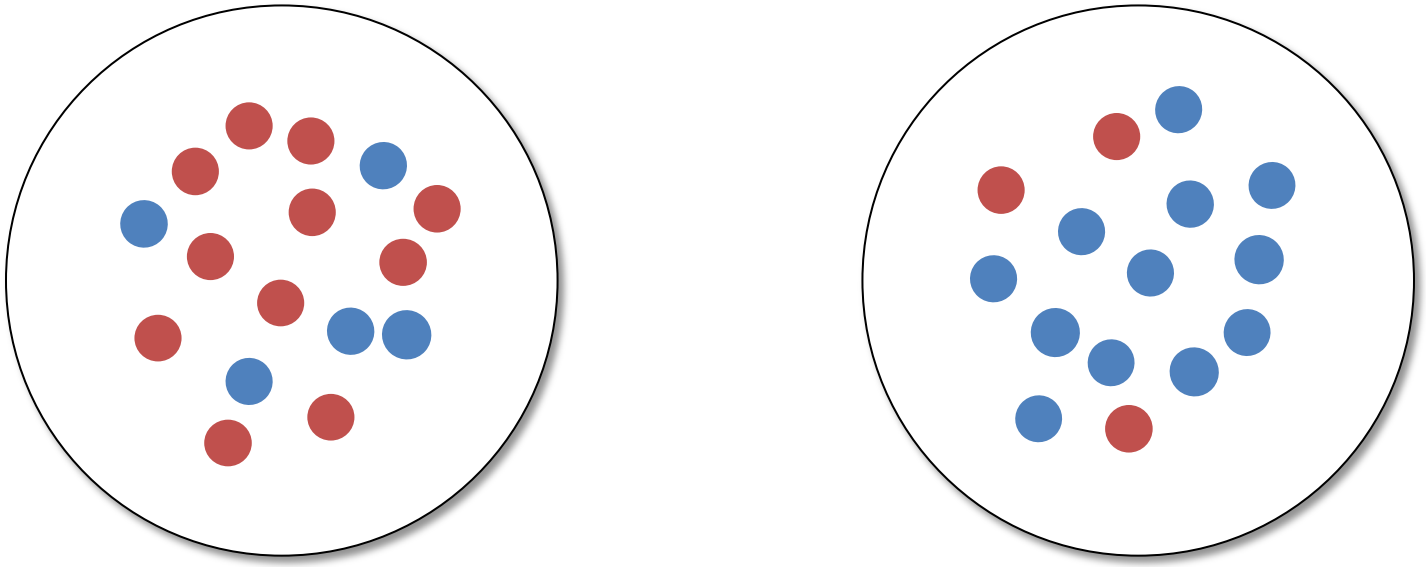
among populations

within populations

population differentiation

- reduced polymorphism
- changes in the SFS
- increased LD

Population differentiation

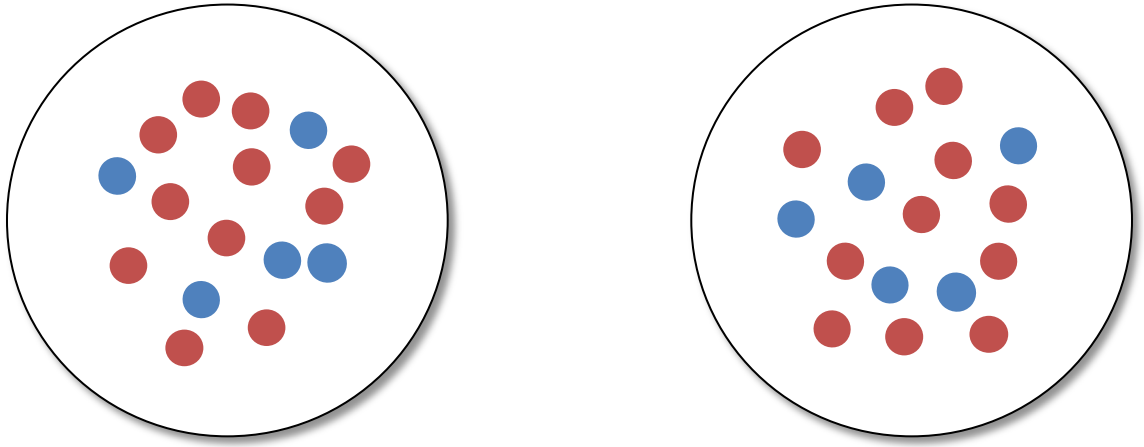


At the simplest level, population-differentiation based approaches rely on the simple assumption that the populations differ with respect to some (not necessarily defined) selection pressure

F_{ST}: Wright's fixation index

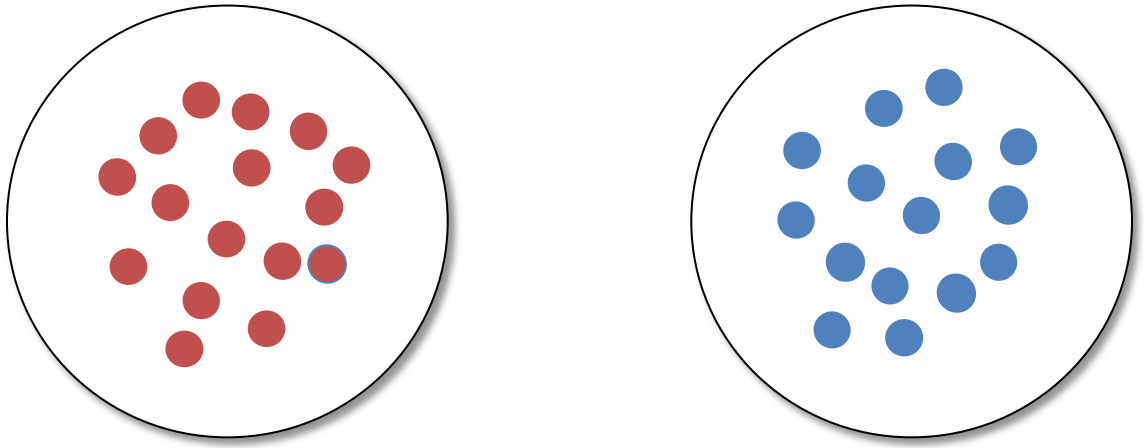
F_{ST} measures the amount of genetic variance that can be explained by population structure

$F_{ST} = 0$



No differentiation

$F_{ST} = 1$



Complete differentiation

F_{ST}: Wright's fixation index

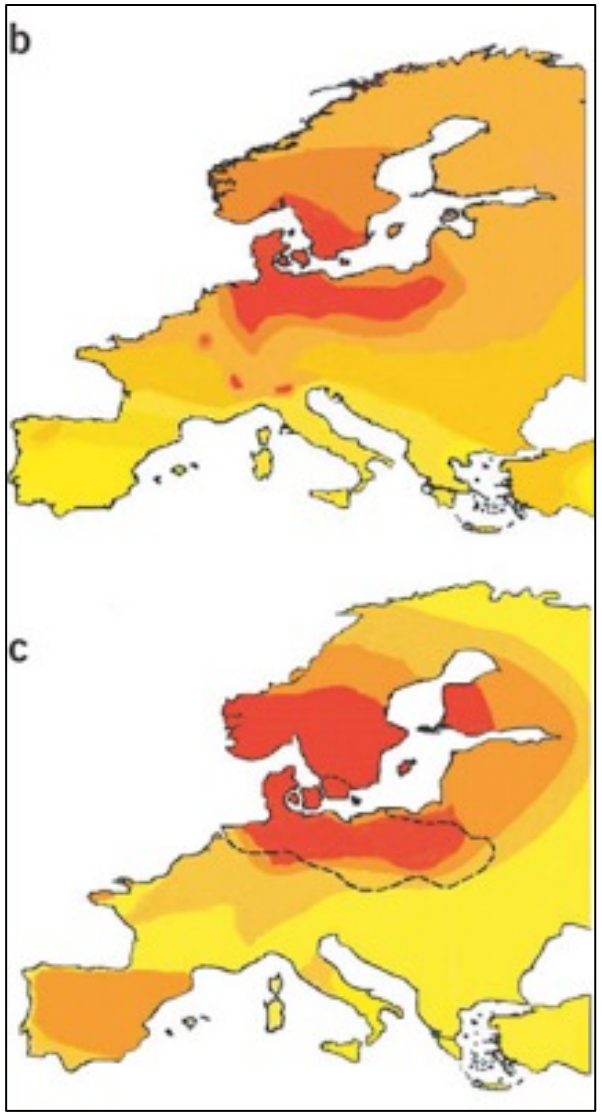
Recall:

- F_{ST} measures the amount of genetic variance that can be explained by population structure
- This is the fraction of diversity that is not due to the mean of the within population diversity

$$F_{ST} = \frac{\sigma_S^2}{\sigma_T^2} = \frac{\sigma_S^2}{\bar{p}(1-\bar{p})}$$

Where $\bar{p}$ is the average frequency of an allele in the total population, σ_S^2 is the variance in the frequency between subpopulations, weighted by the sizes of the populations and σ_T^2 is the variance of the allelic state in the total population

An example: lactase persistence in humans

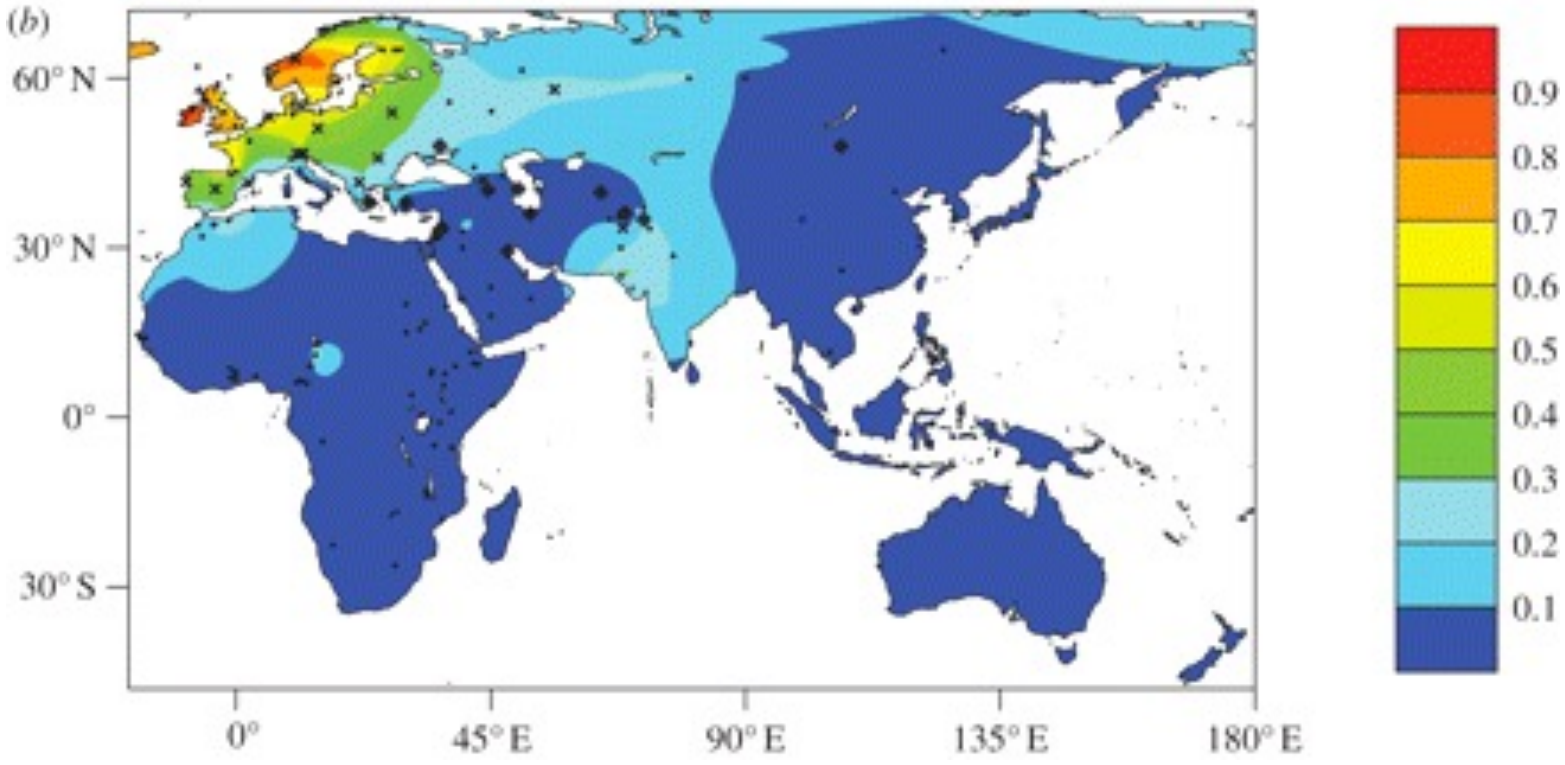


Cow milk protein diversity (proxy of length of time milk has been an important part of the diet)

Frequency of lactase persistence in humans

Geographic distribution of allele responsible for lactase persistence in Europe

Distribution of LCT -13910*T



Adaptation to dietary shift: lactase persistence in Europeans

Simoons hypothesized that the distribution of pastoralism could explain the striking differences in lactase persistence among populations

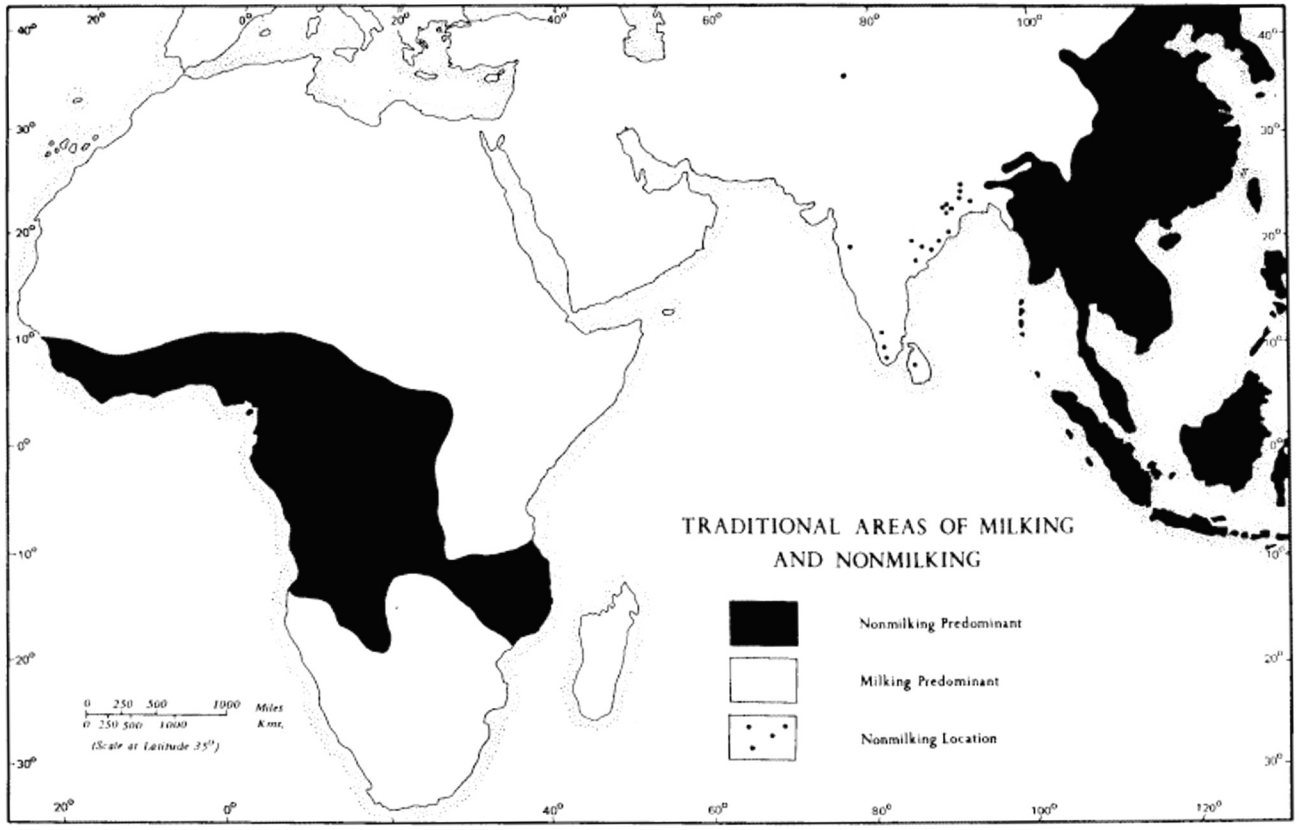
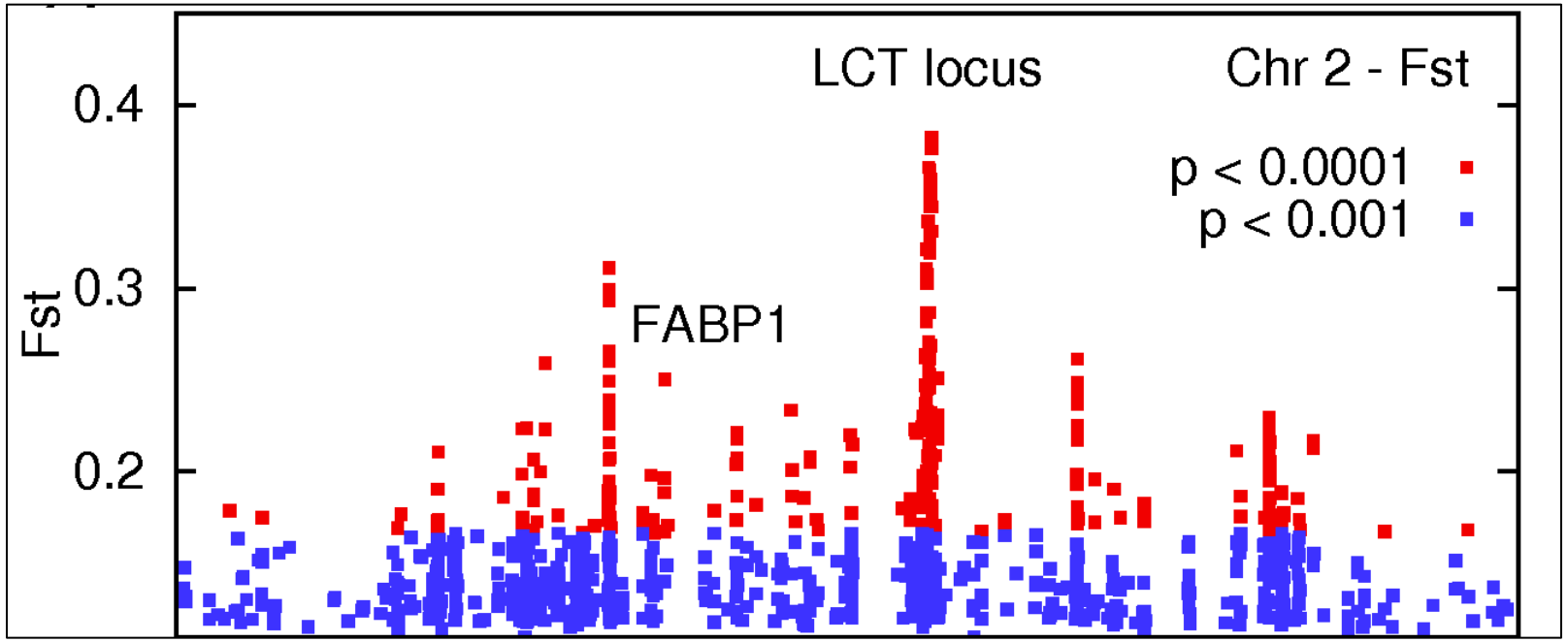


Fig 1. Traditional areas of milking and nonmilking.

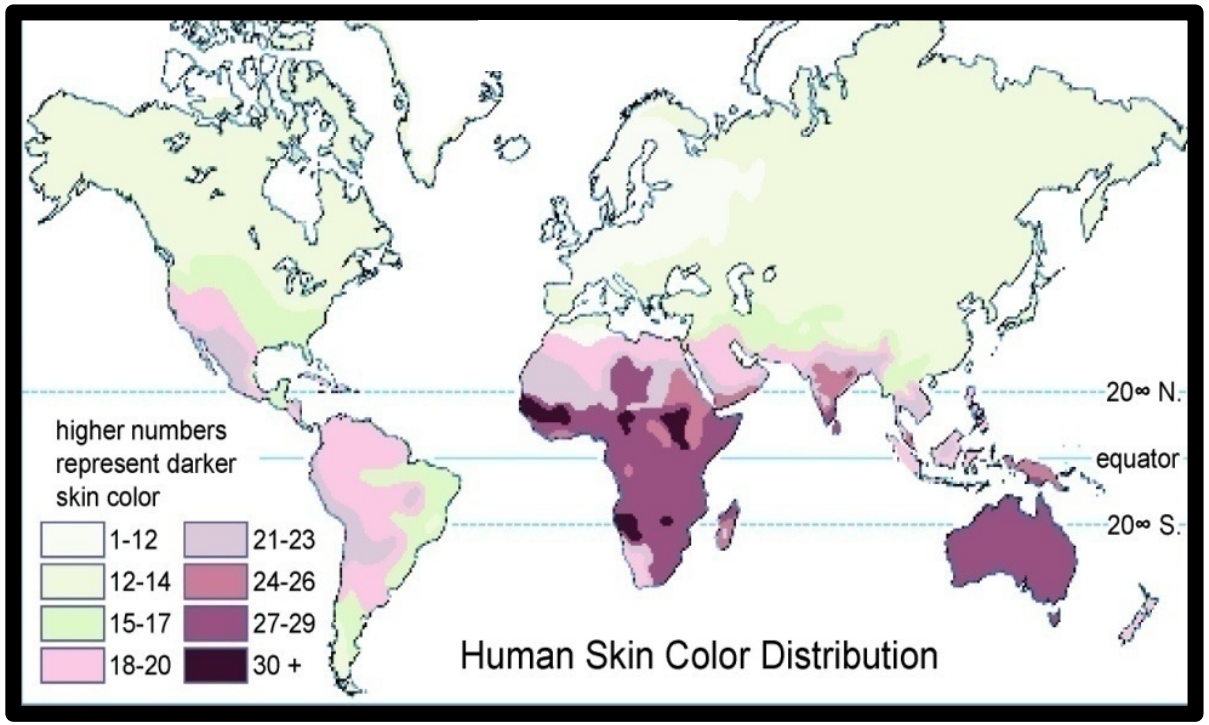
Simoons, 1970

The LCT locus is differentiated between European and African populations

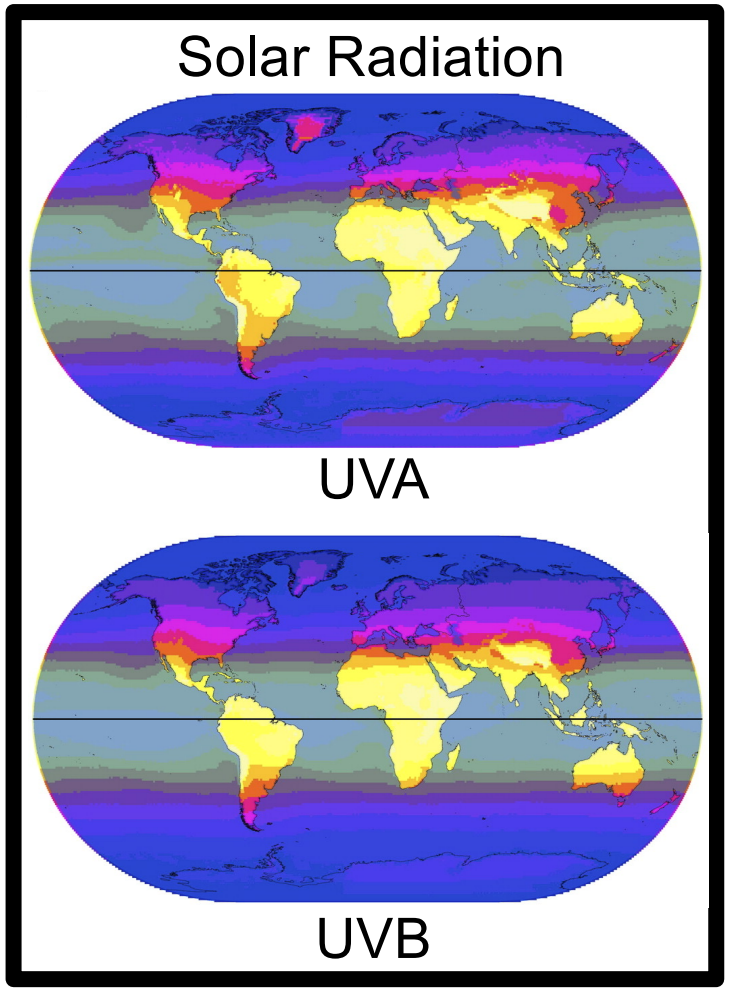


In humans pigmentation is correlated with solar radiation

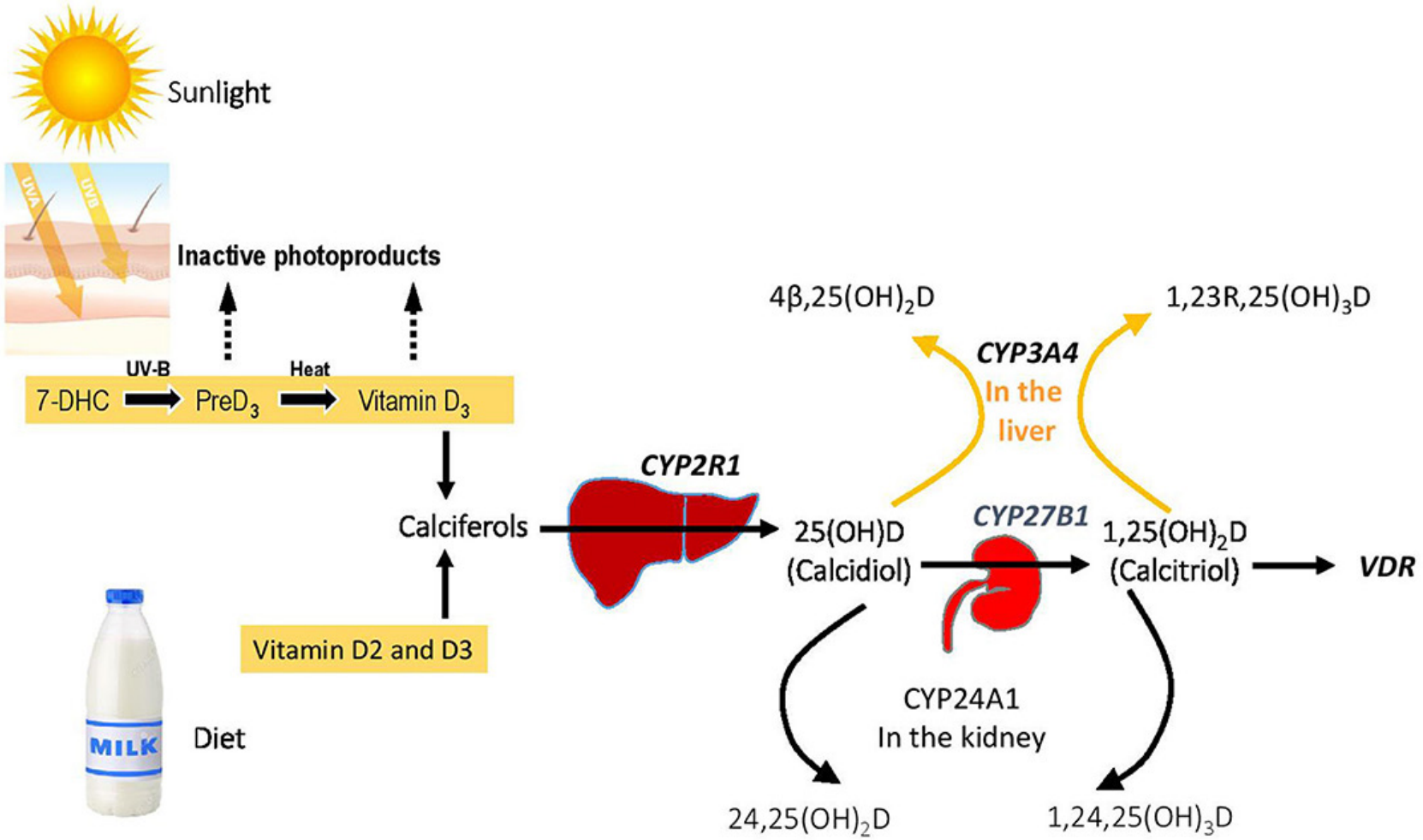
Worldwide variation in pigmentation



from Barsh PLOS Biology 2003, adapted from Biasutti 1953



Sunlight is needed for vitamin D production, so high levels of pigmentation can be detrimental at high latitude



Inadequate vitamin D levels can result in many physiological ailments

Possible symptoms include:

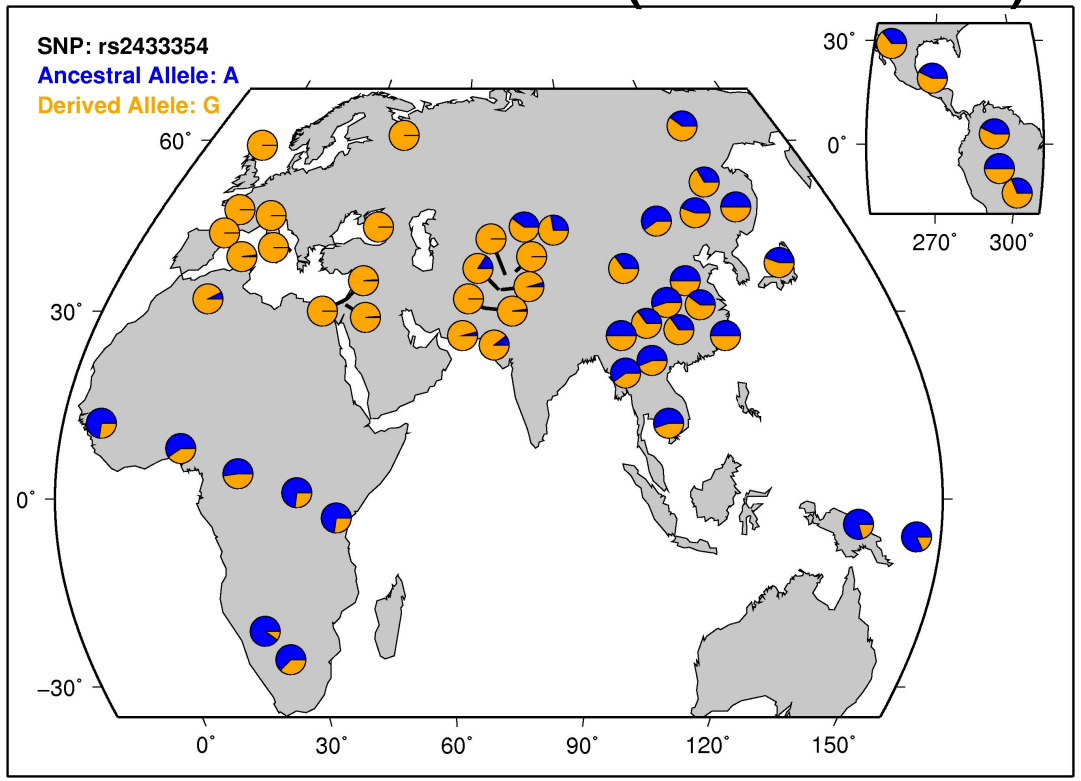
- Muscle and bone pain and increased sensitivity to pain
- Muscle weakness in body parts near the trunk of the body, such as the upper arms or thighs
- Increased risk of broken bones
- Muscle spasms, twitches or tremors
- Bowed legs (when the deficiency is severe)
- Increased risk of chronic heart failure

Tradeoff between protective effect of pigmentation (against UV damage and cancer) and deleterious effect at high latitude

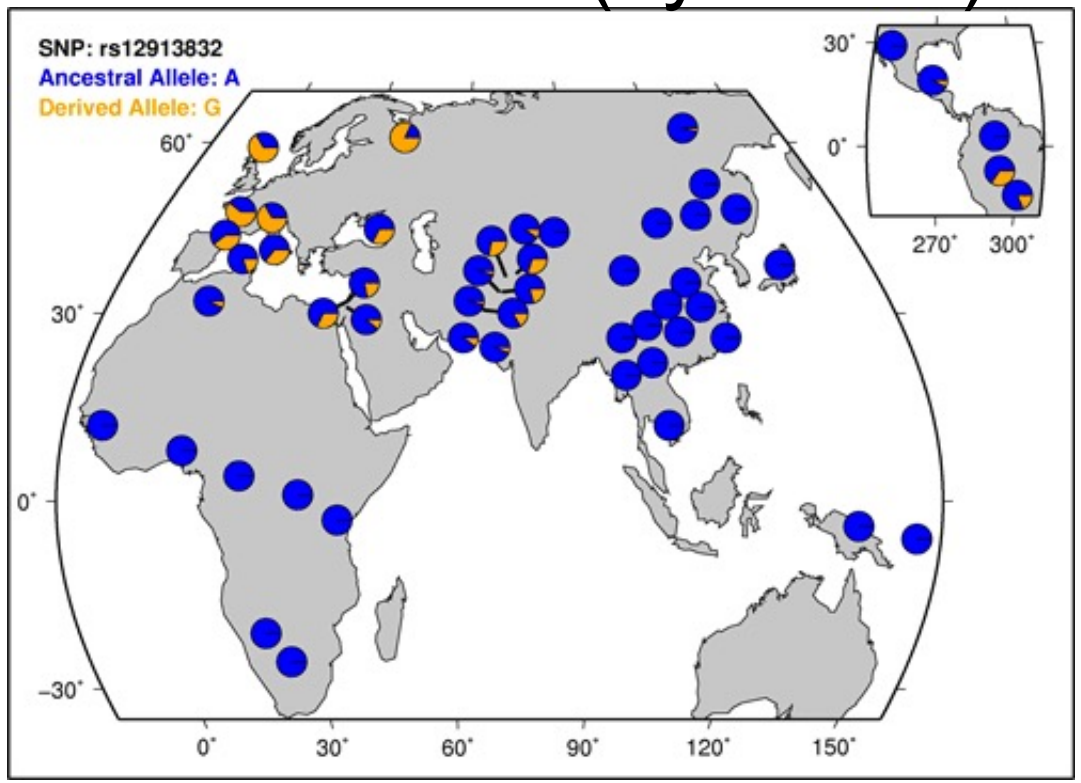
https://www.yalemedicine.org

Multiple variants involved in loss of pigmentation are differentiated among human populations

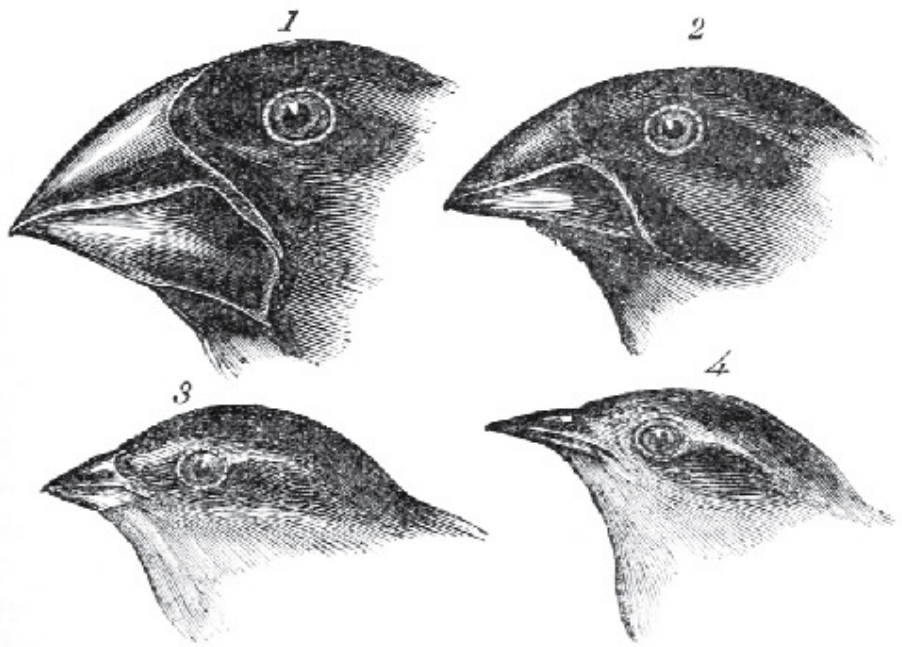
SLC24A5 variant (skin color)



OCA2 variant (eye color)



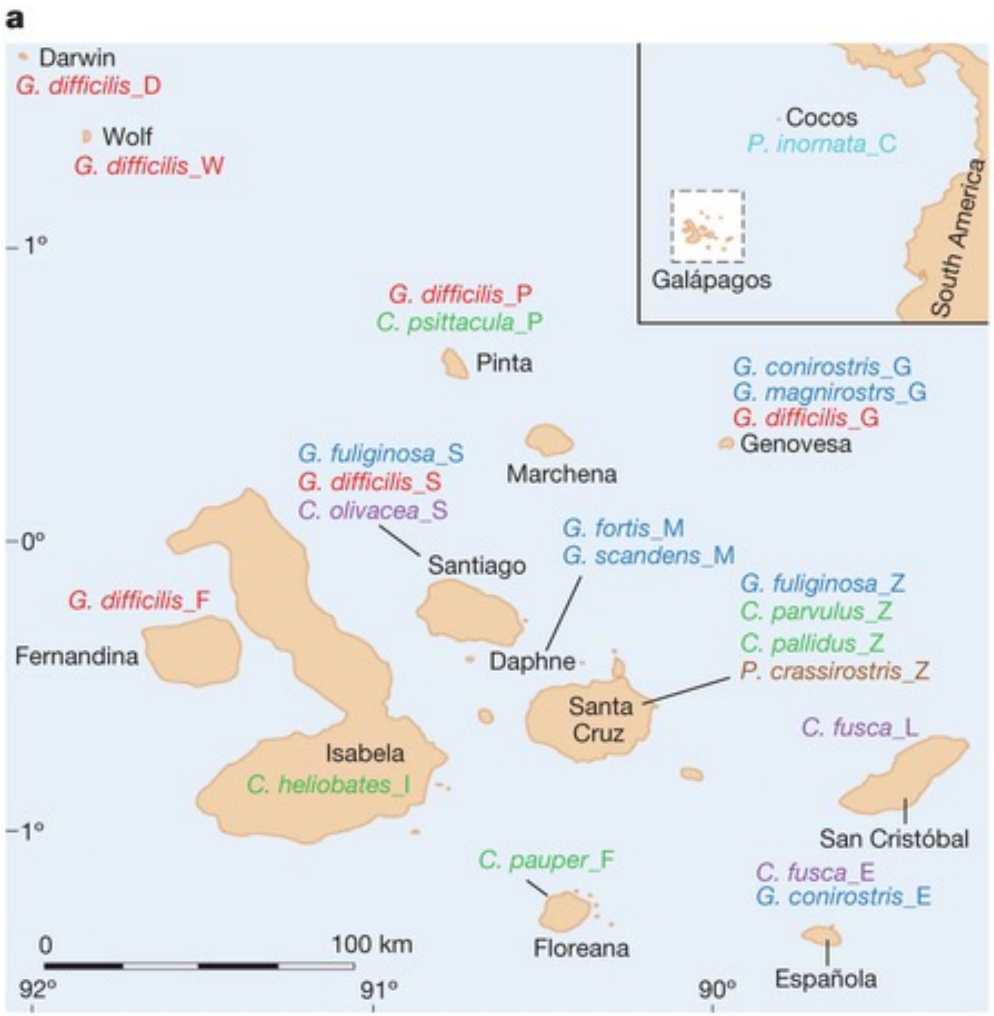
Beak morphology in Darwin’s finches: a classic example of adaptive radiation

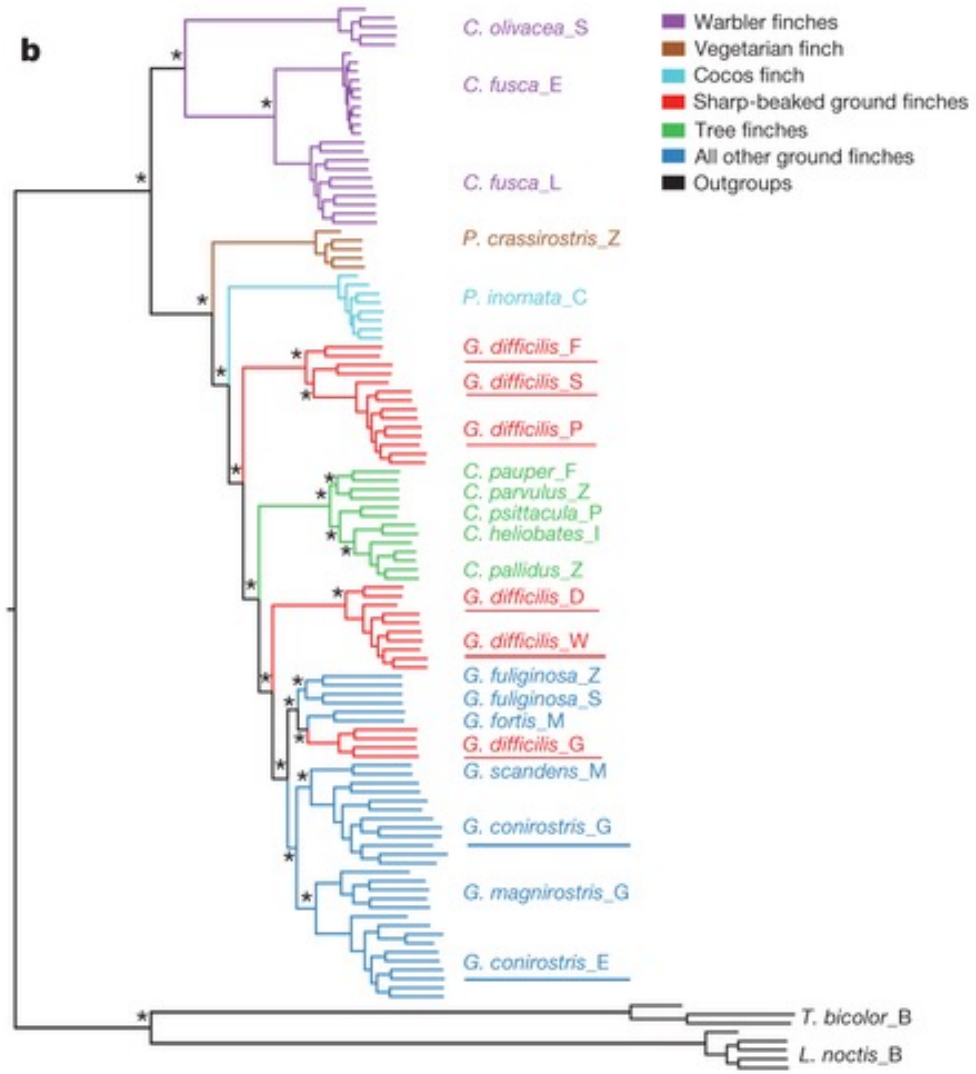


1. Geospiza magnirostris
2. Geospiza fortis
3. Geospiza parvula
4. Certhidea olivacea

Finches from Galápagos Archipelago

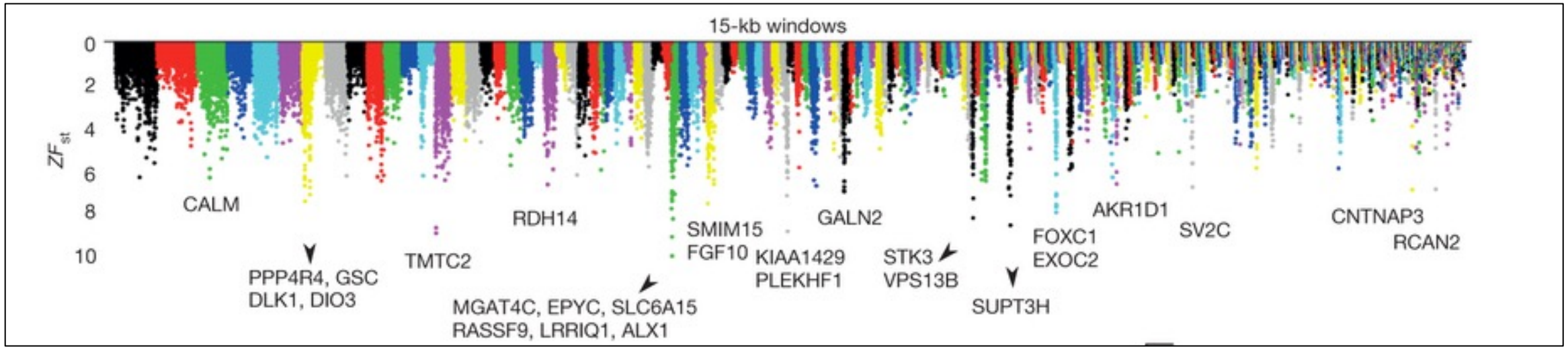
Sequencing the genomes of Darwin's finches





Lamichhaney et al., 2015

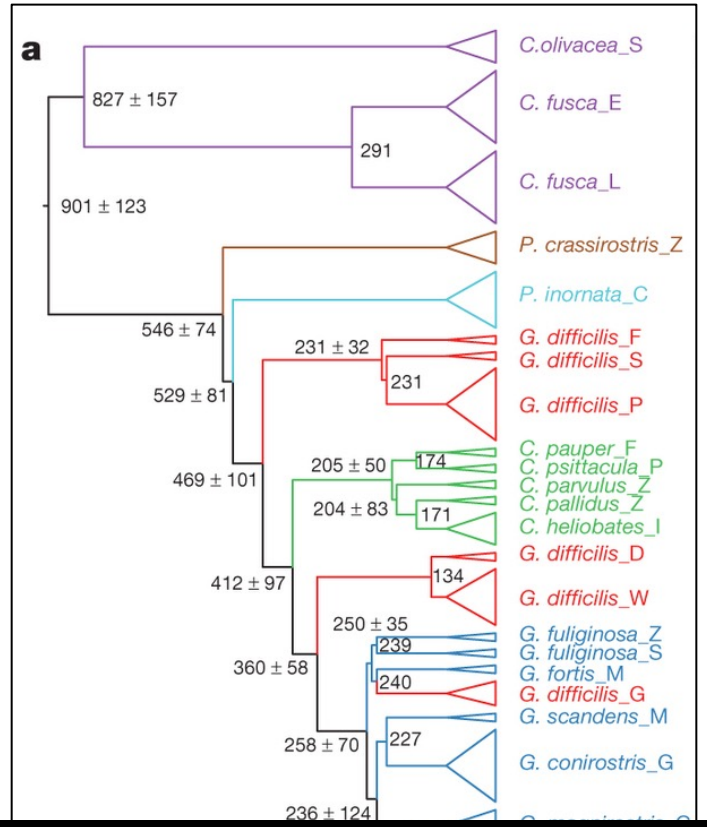
Population differentiation across the genome of Darwin's finches

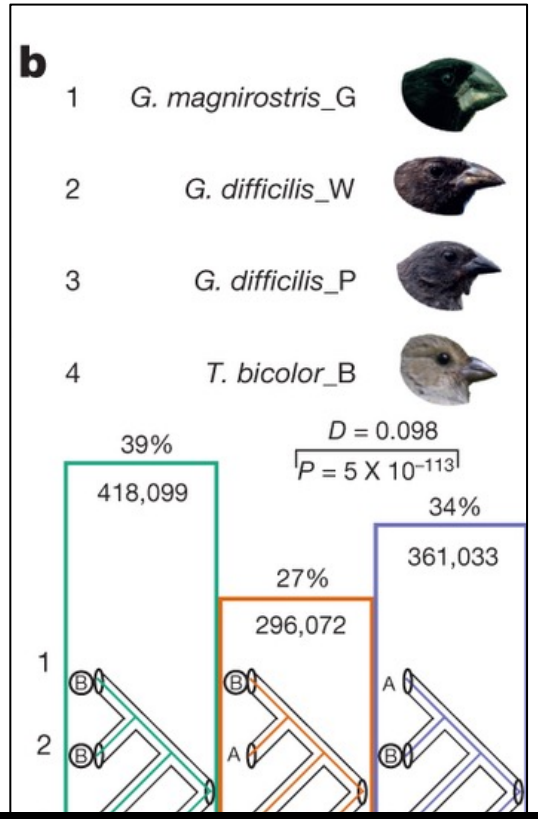


Differentiation at many loci across the genome

Lamichhaney et al., 2015

Haplotype sharing due to incomplete lineage sorting and introgression across the genome

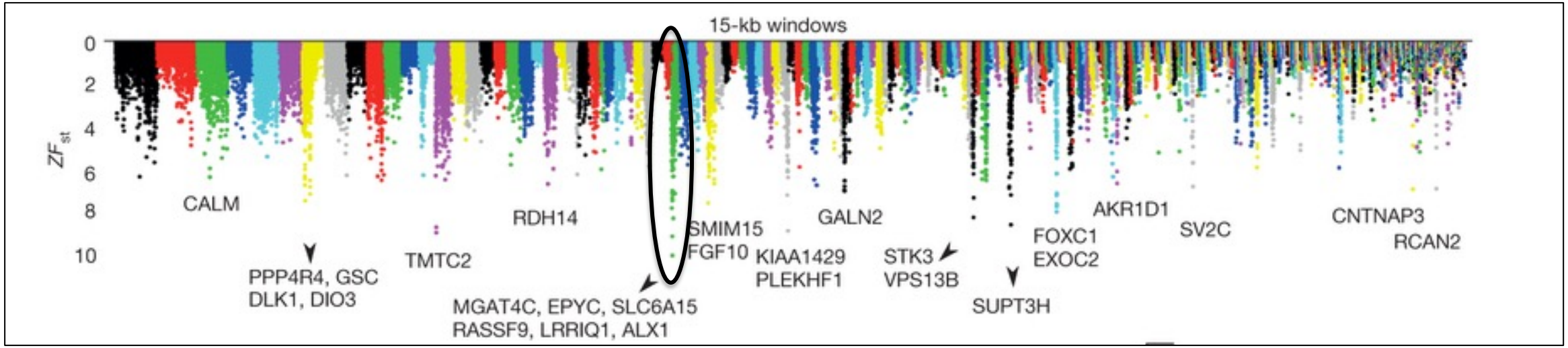




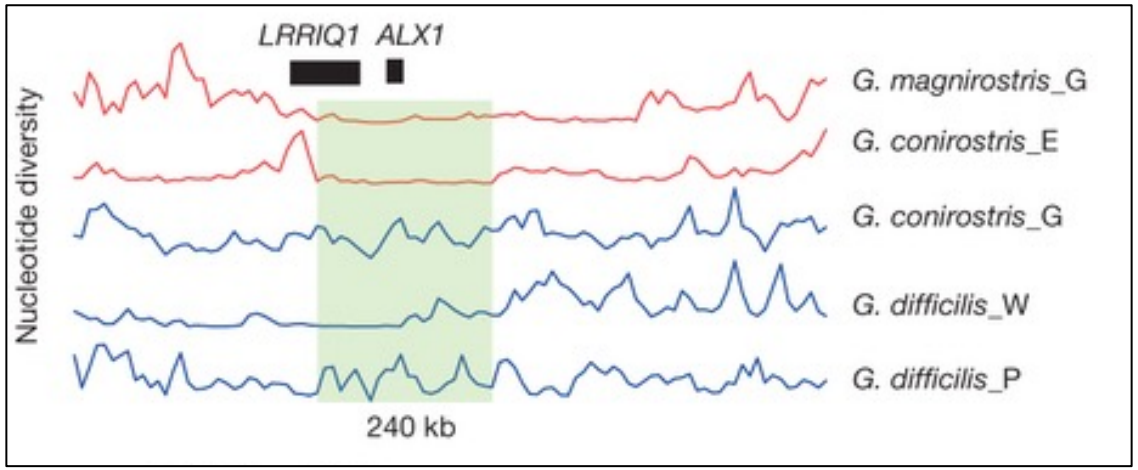
This means that there is shared variation across species, so that segregation patterns across species can be used to identify adaptive loci

Lamichhaney et al., 2015

Population differentiation across the genomes of Darwin's finches (across a combination of species)



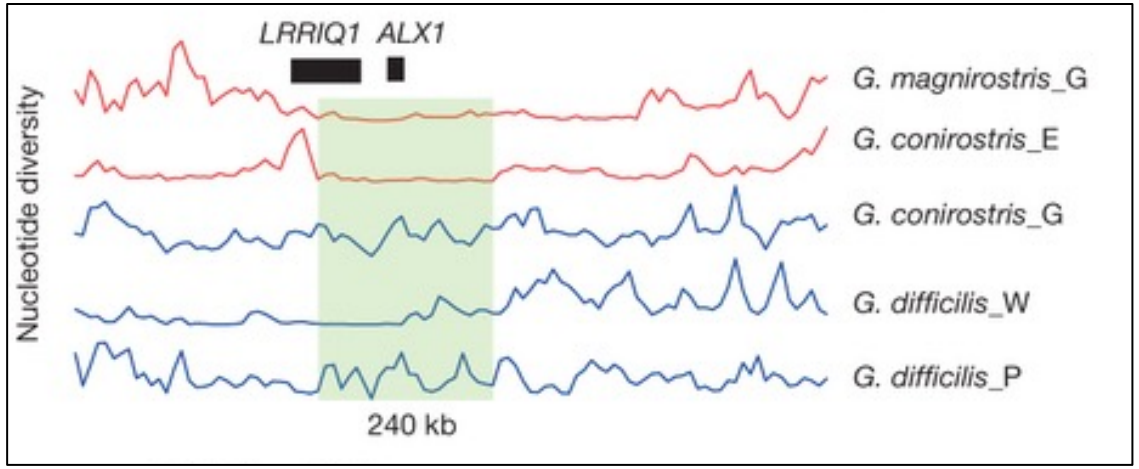
ALX1 is involved in beak development



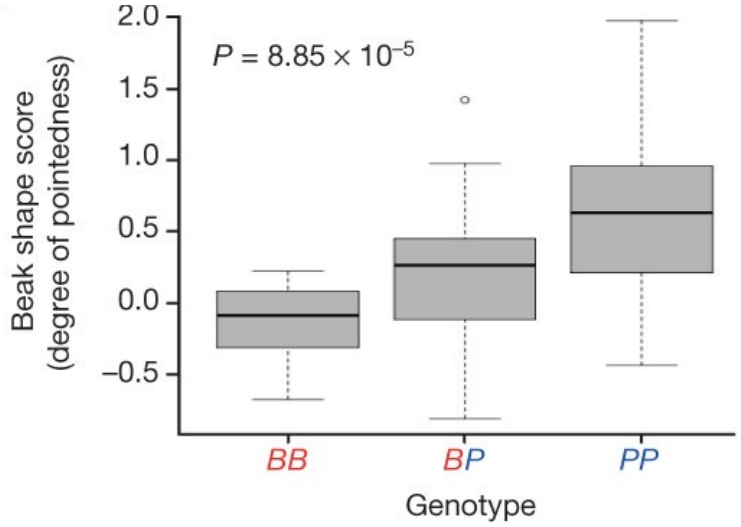
Lamichhaney et al., 2015

ALX1 haplotype is strongly differentiated among populations and associated with beak shape

Reduced diversity at ALX1, a gene involved in beak development

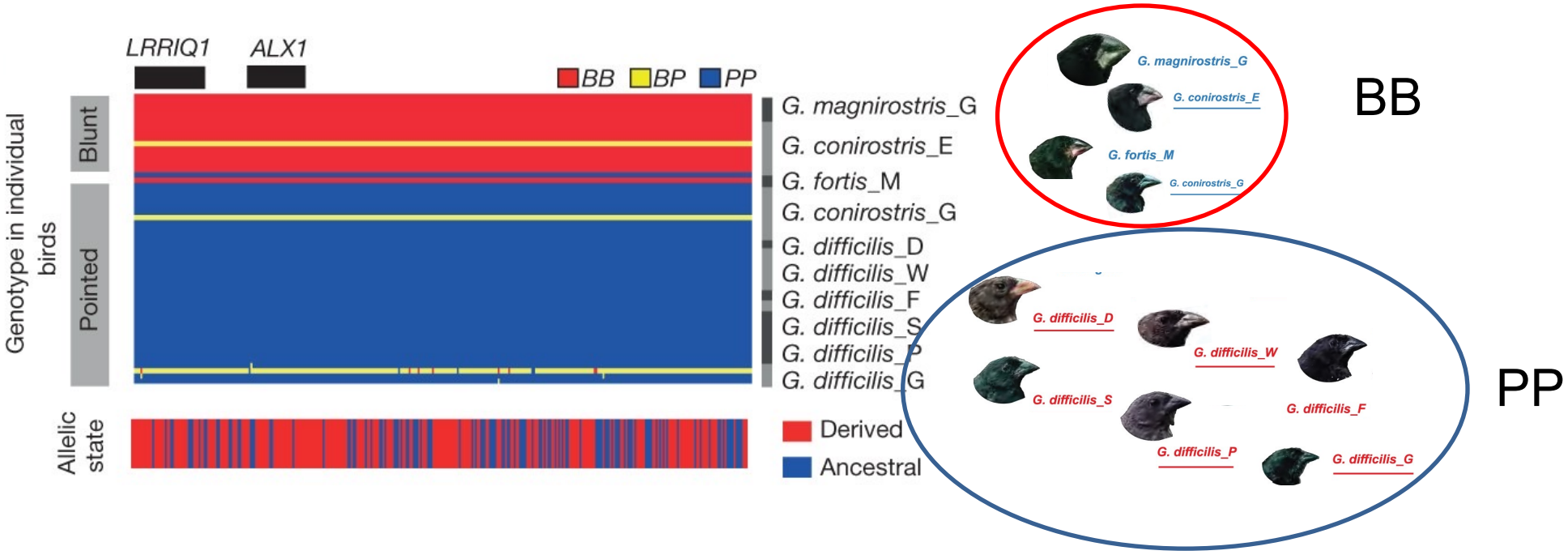


B haplotype is associated with blunt (versus pointed) beaks



Lamichhaney et al., 2015

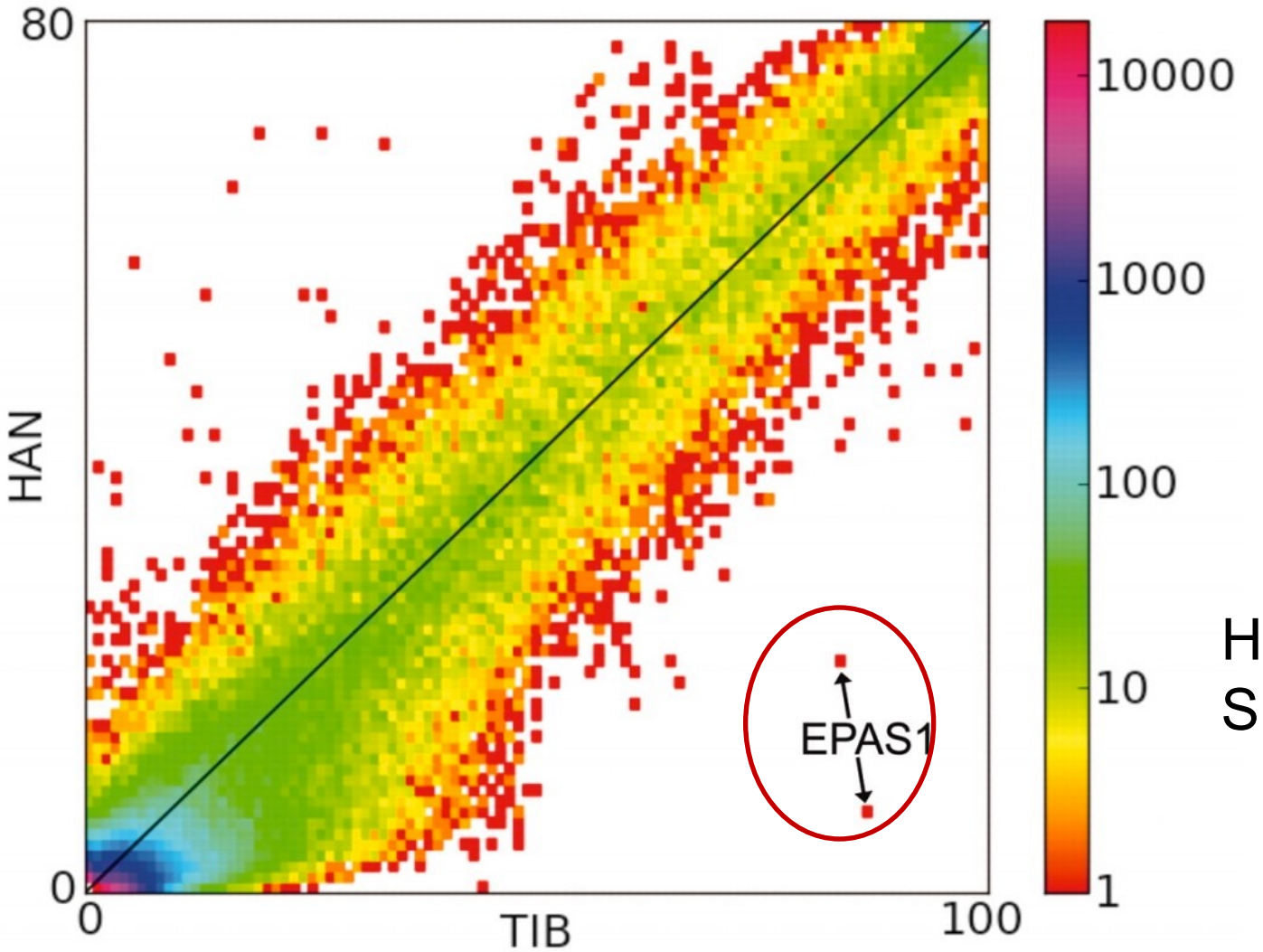
ALX1 haplotype is strongly differentiated among populations and associated with beak shape



Lamichhaney et al., 2015

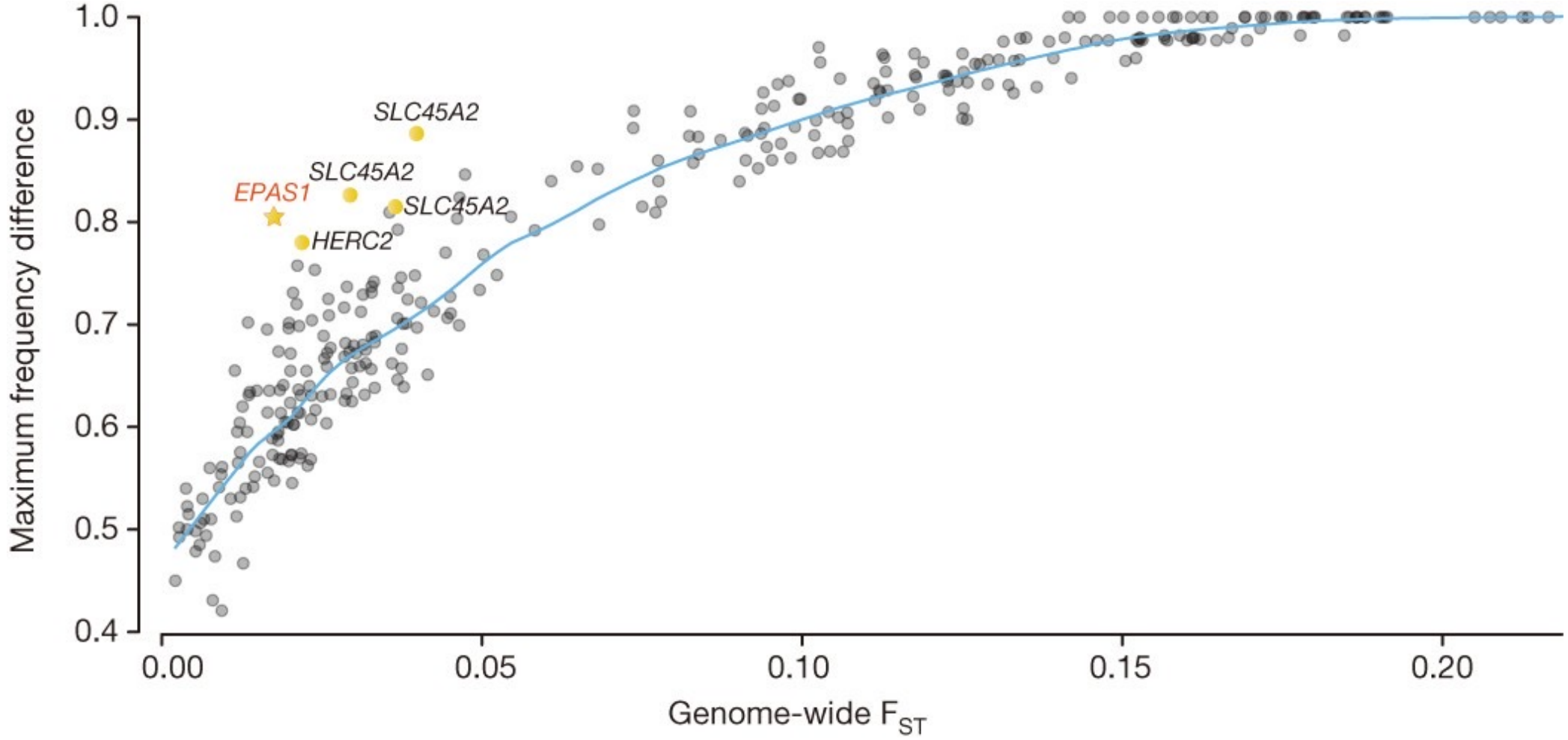
High altitude adaptation in humans

High altitude adaptation in humans identified based on frequency differences



Highly differentiated SNPs at *EPAS1*

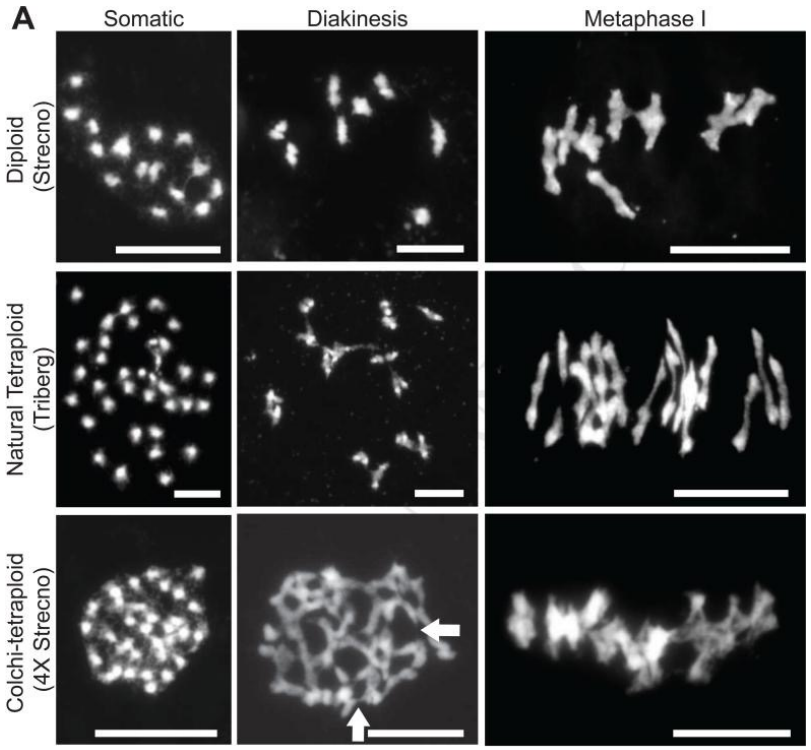
Maximal frequency differences relative to pairwise F_{ST}



Huerta-Sánchez et al., 2014; https://www.nature.com/articles/nature13408

Polyploids can be stable in nature, but their formation in the lab is associated with meiotic disfunction

DAPI-stained meiotic chromosome spreads

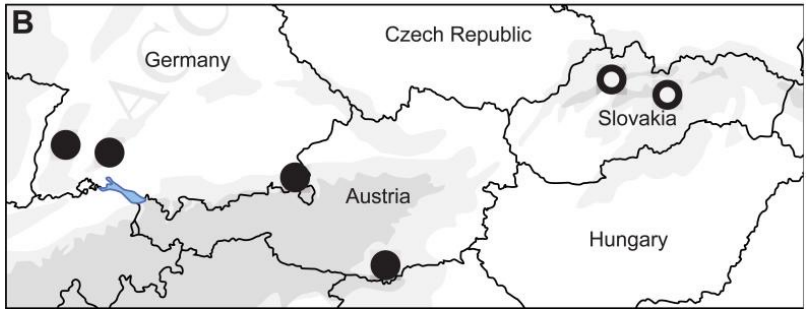


Diploid

Natural polyploid (collected from nature)

Synthetic polyploids created using colchicine, a mutagen

Map of populations sequenced. Tetraploids are indicated with closed circles; diploids with open circles

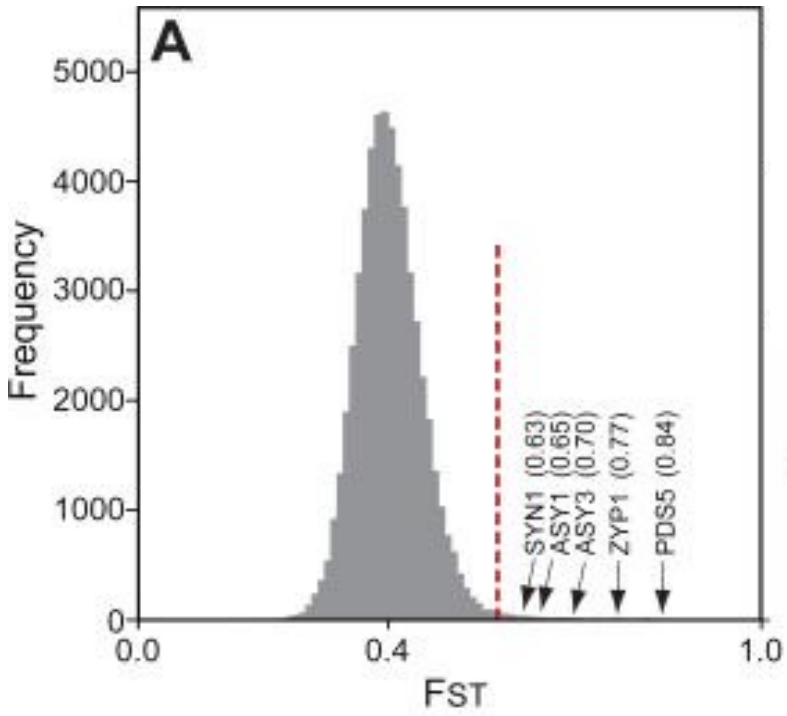


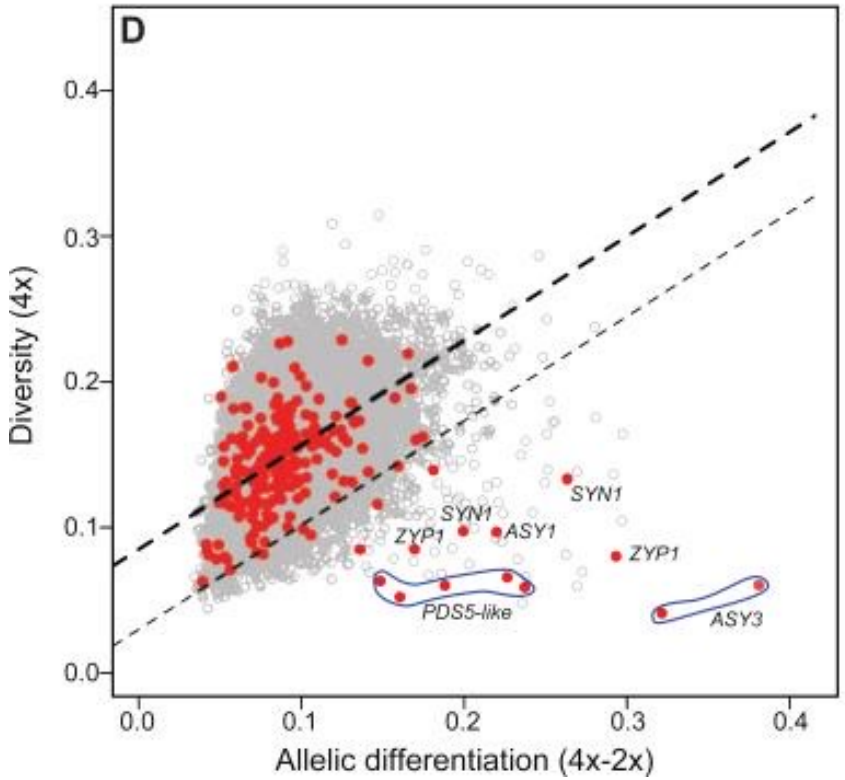
Yant et al., 2014, https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3859316/

Polyploids can be stable in nature, but their formation in the lab is associated with meiotic disfunction

What makes them viable in nature?

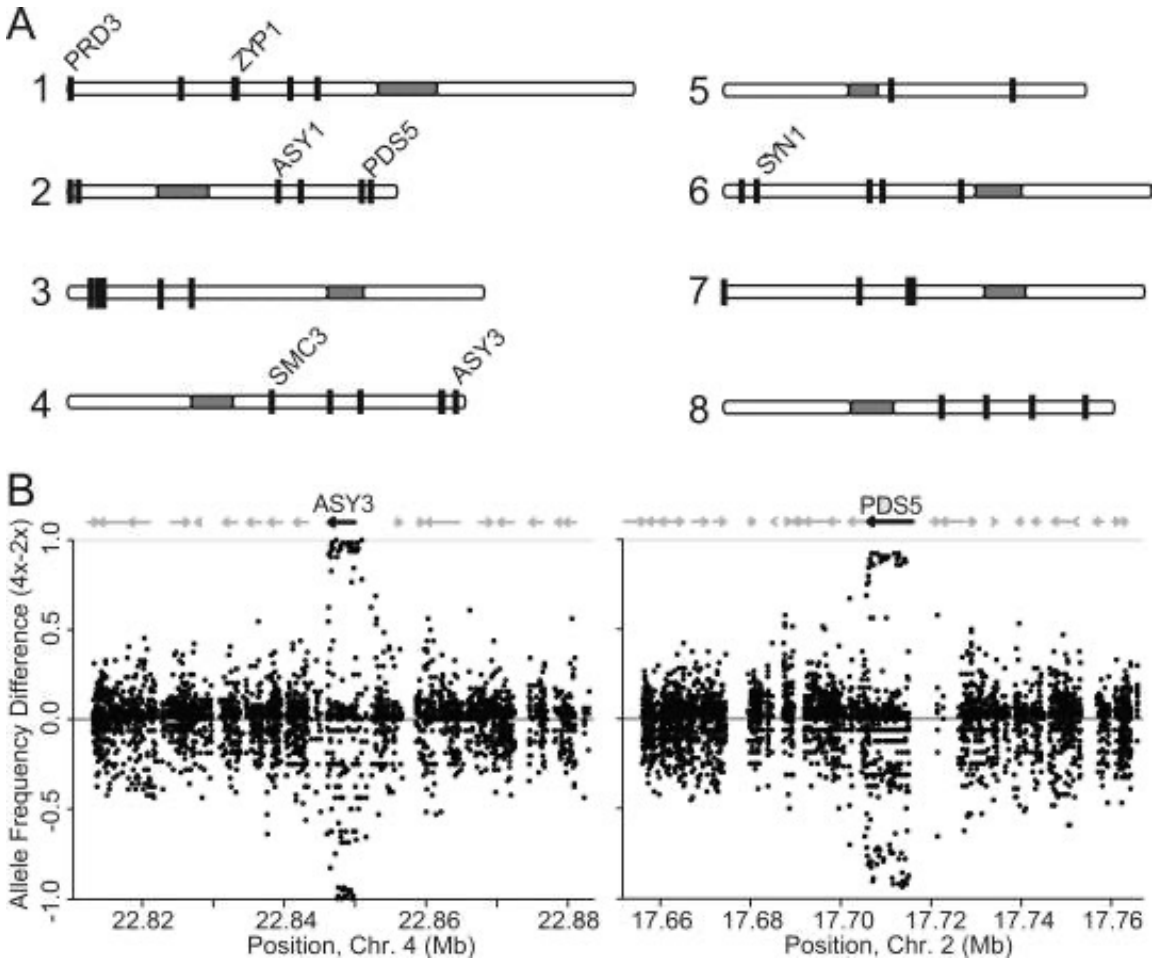
Adaptation through changes in core meiosis genes!





Yant et al., 2014, https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3859316/

Examples of meiosis genes, *ASY3* and *POS5,* with differentiation signals



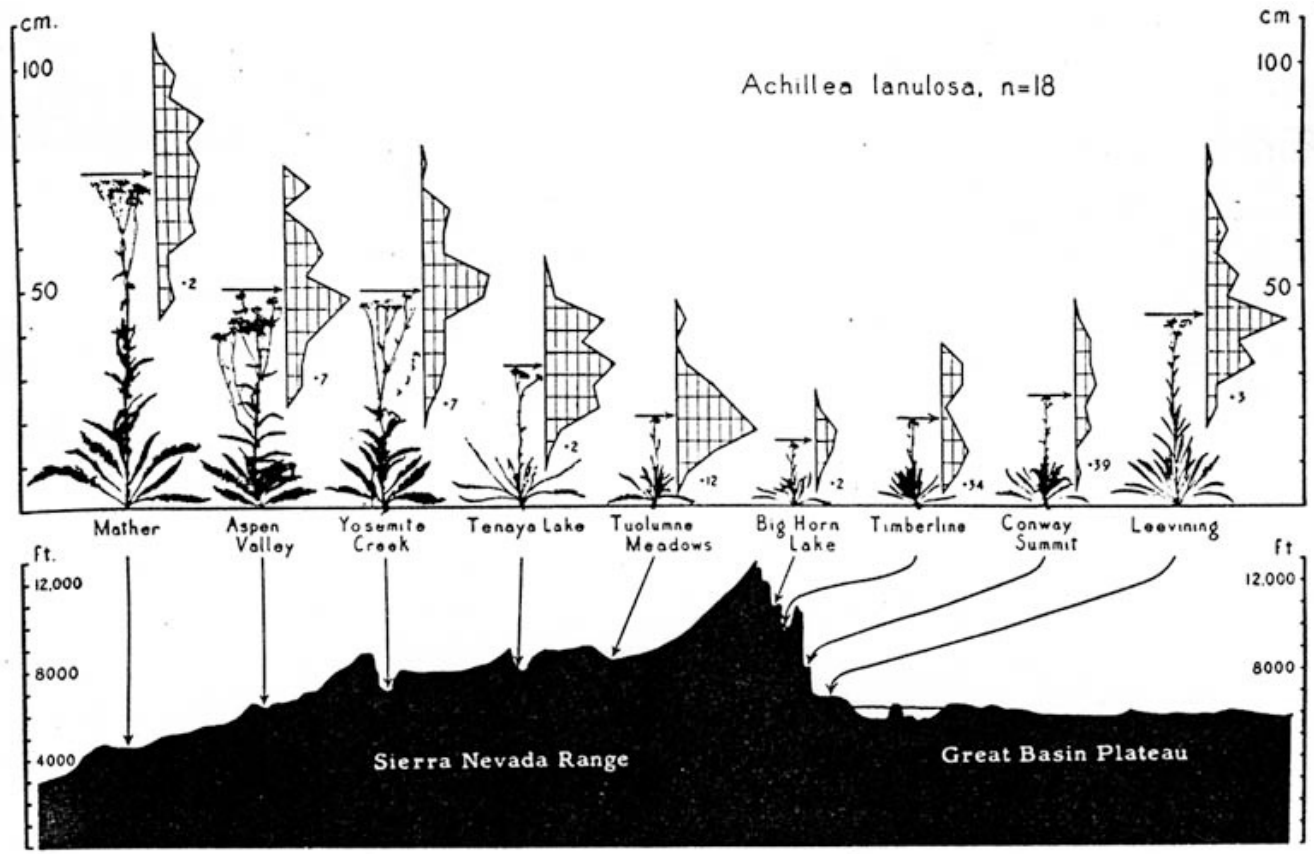
Strong differentiation between SNPs in the ASY3 and POS5 regions

Yant et al., 2014, https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3859316/

Identifying regions that are simply differentiated between two populations can be powerful in simple cases

… but can we do better?

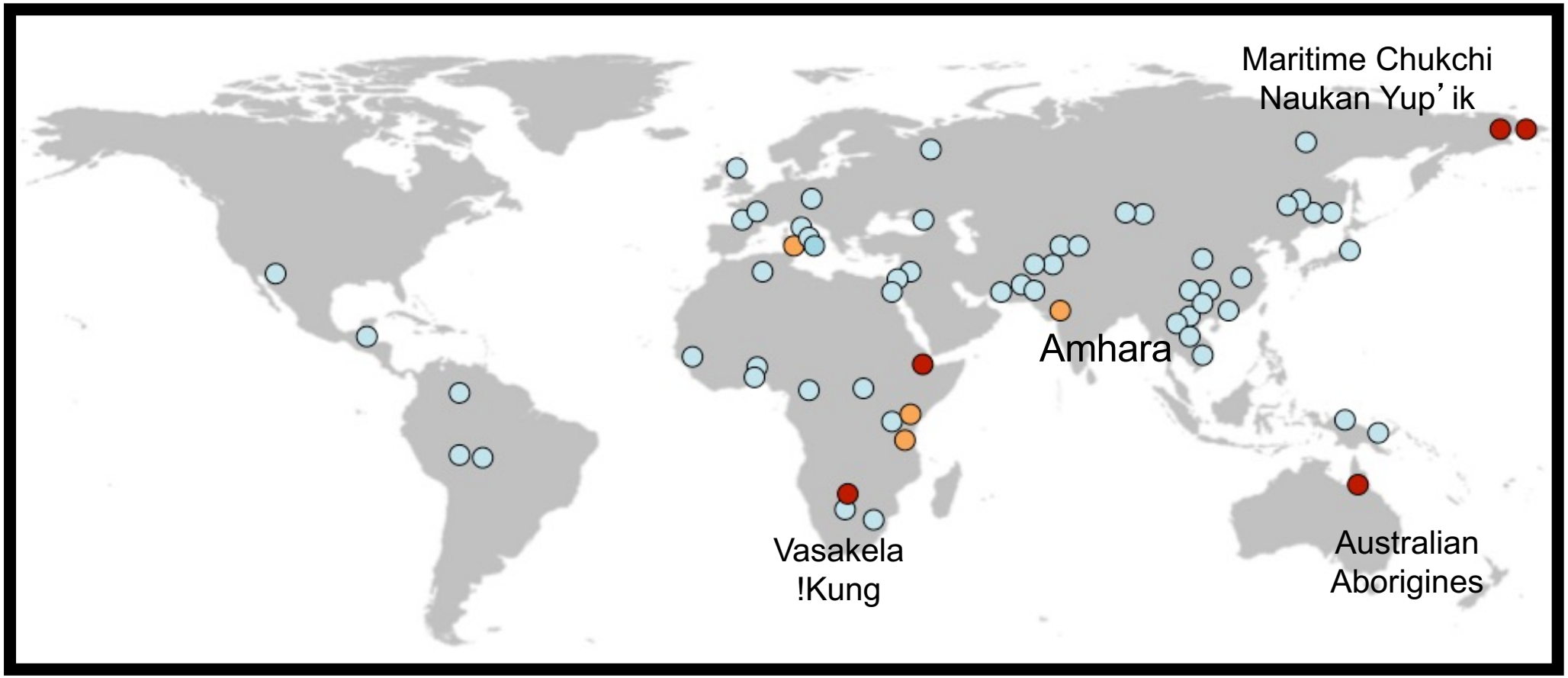
Clinal patterns represent a classic signature of adaptation



from Clausen, Keck and Heisey, 1948

Heights of yarrow plants vary with altitude

Can we use clinal patterns to identify adaptive *loci*?

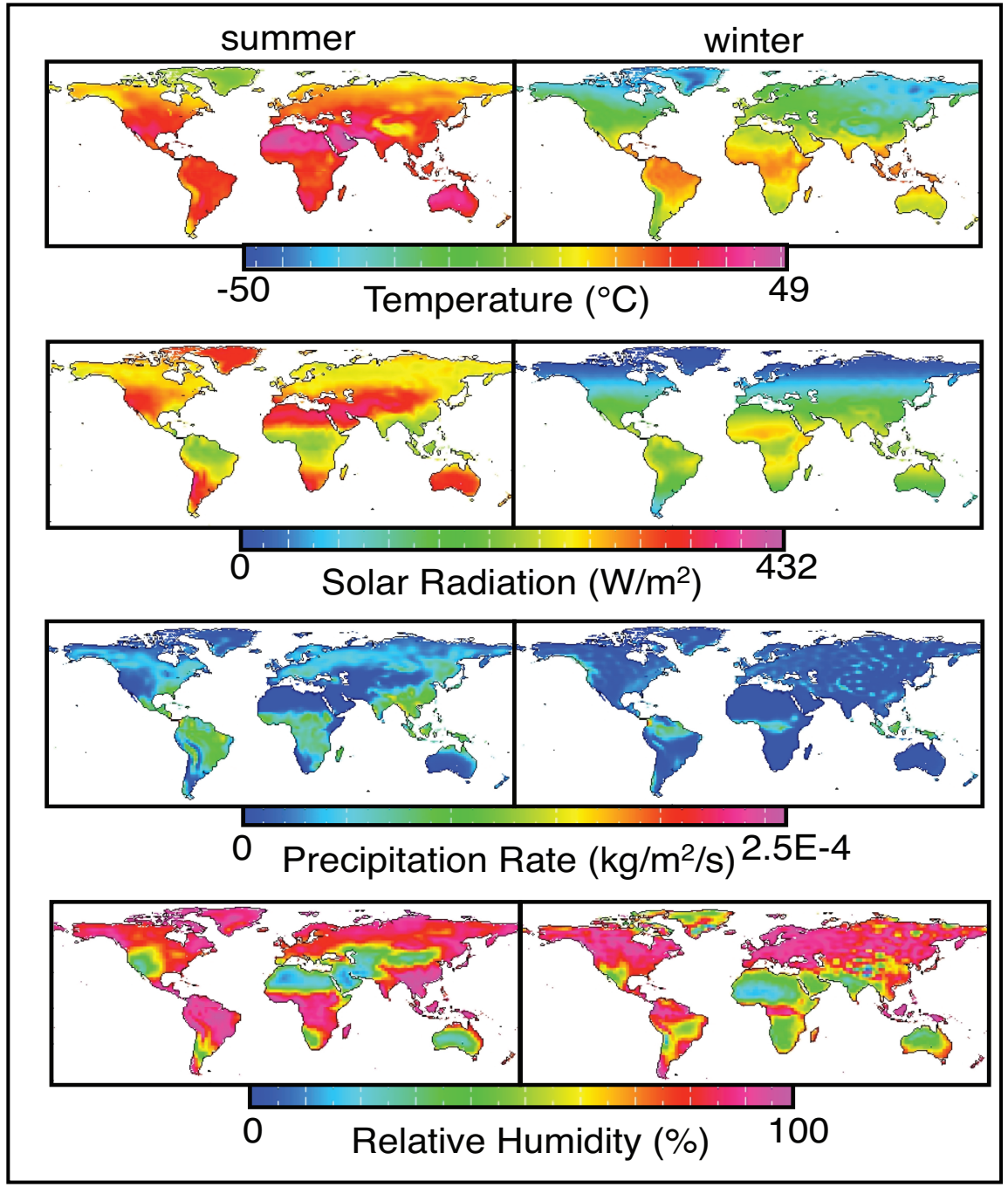


Genome-wide data from 1344 individuals from 61 populations:

- 938 individuals from the Human Genome Diversity Panel
- 288 individuals from HapMap Phase 3
- 118 individuals genotyped for these projects

Hancock et al., 2011

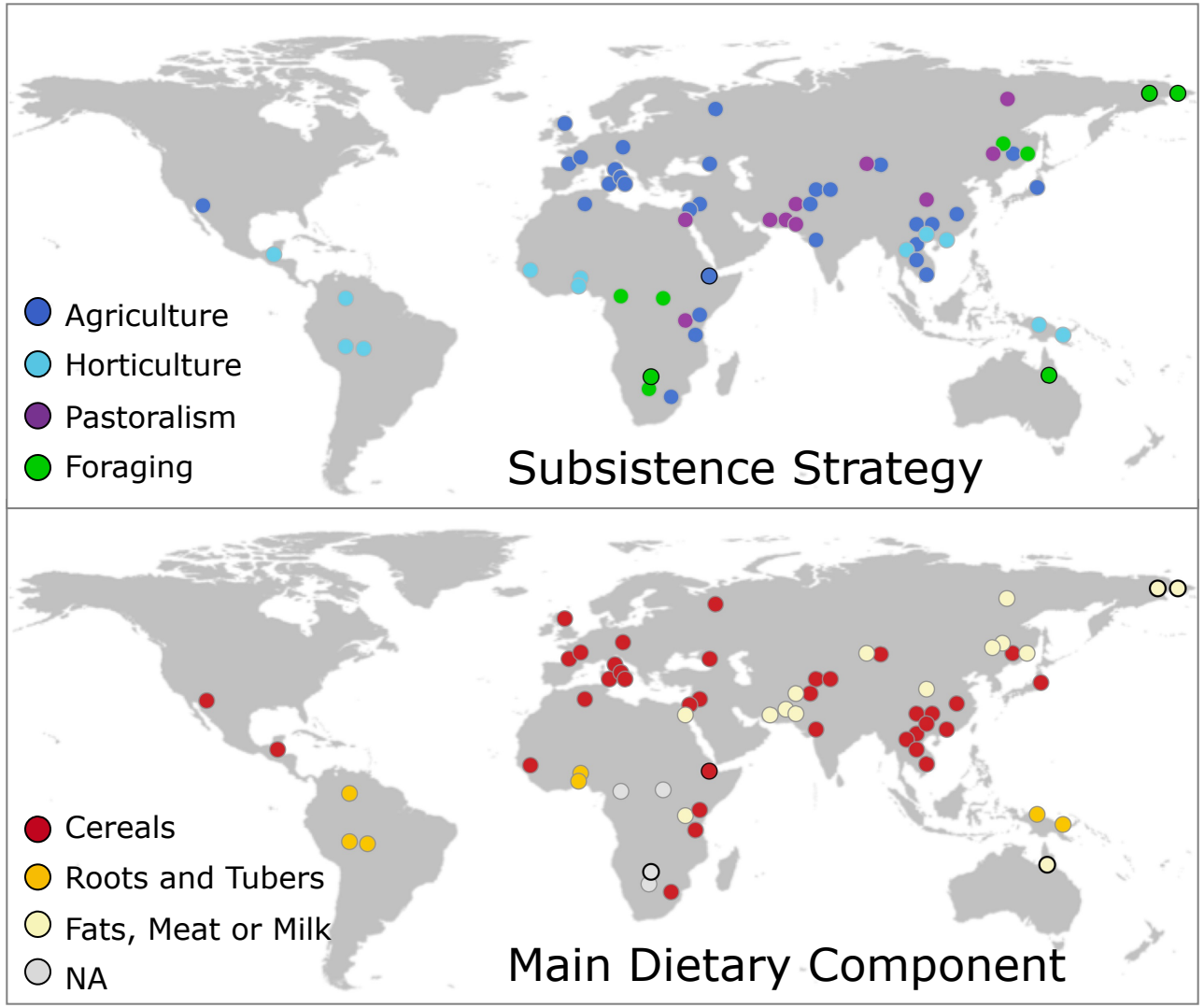
Climate Variables



Climate data source:

NCEP/NCAR Reanalysis Project (Kistler et al., 2001)

Diet and Subsistence Variables

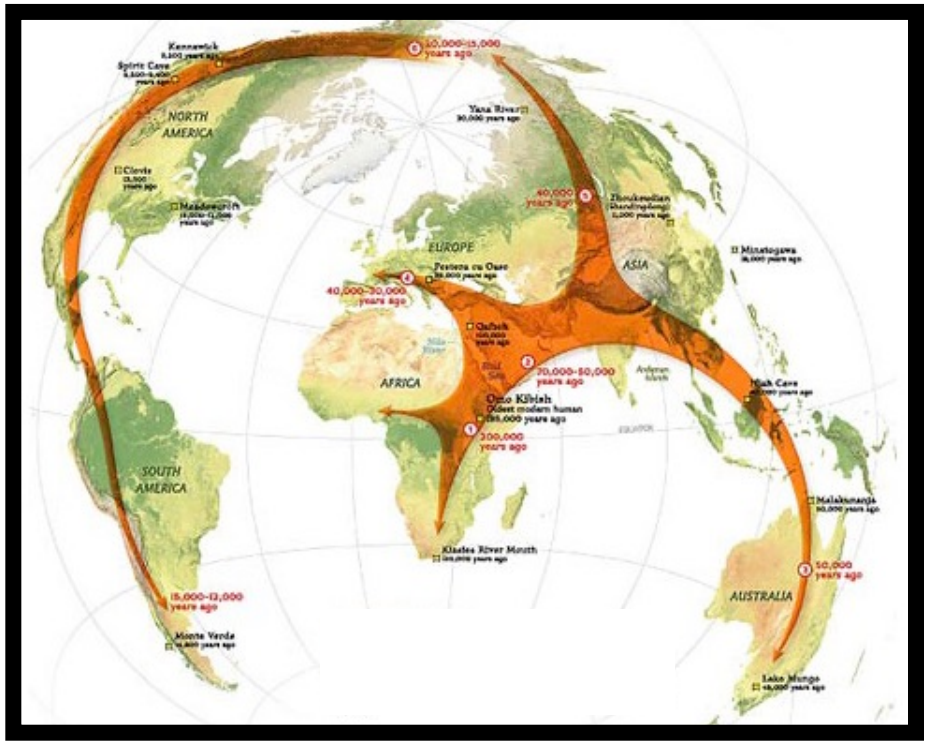


Sources:

- *Ethnographic Atlas (Murdock 1967)*
- *Encyclopedia of World Cultures (Levinson 1991-97)*

Hancock et al. 2010

Population history confounds efforts to identify adaptive loci



Significance of correlations may be over-estimated if:

- Population history is correlated with the environment

… And under-estimated if:

- The effects of selection are subtle relative to the effects of population structure on allele frequencies

Solution: Model population structure when assessing evidence for correlation with the environment

Many of the strongest correlations are with amino acid-changing variants

Climate

- *TLR6* P249S & solar radiation

  associated with malaria resistance and prostate cancer
- *TRIP6* V858I & minimum temperature

  implicated in energy metabolism and basal metabolic rate

Diet/Subsistence

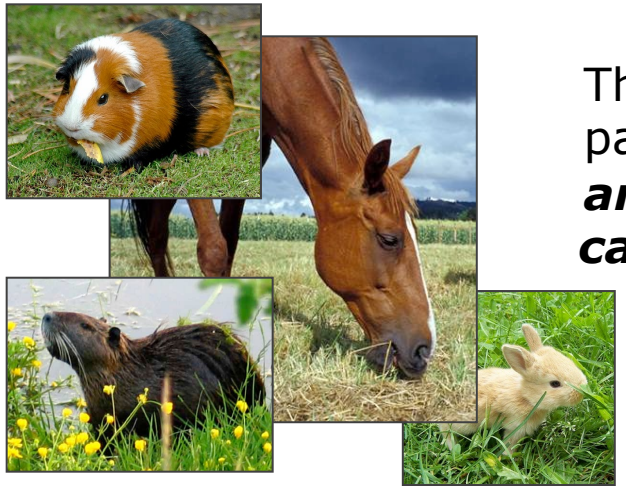
- *MTRR* K350R & roots and tubers

  involved in folic acid metabolism; associated with spina bifida
- *CCL22* D2A & pastoral subsistence

  associated with multiple sclerosis and H. pylori-related carcinoma

The strongest correlation with cereal use is a truncating amino acid change

PLRP2 hydrolyzes ***galactolipids***, the main triglyceride component in plants



This protein is found in pancreases of ***herbivores and omnivores*** but not ***carnivores or ruminants***

Structure of *PLRP2*

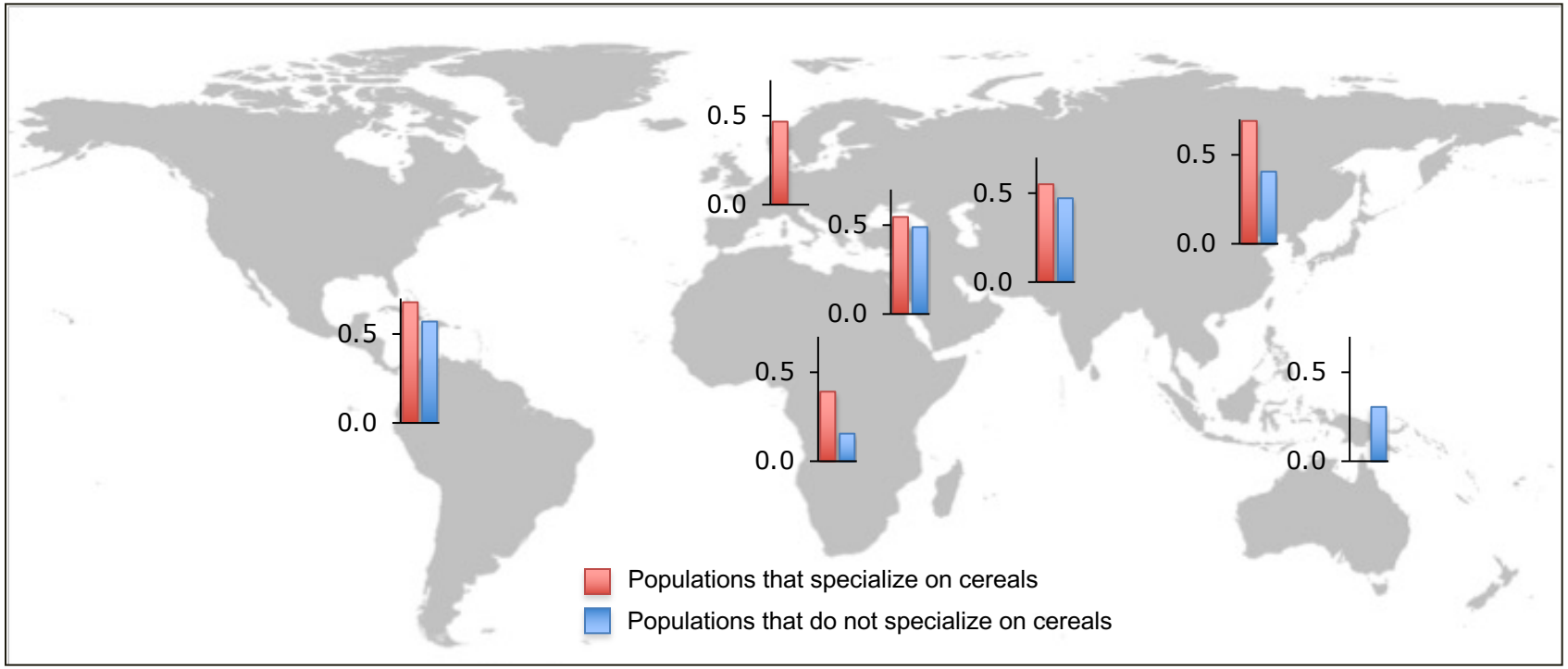
The full length protein in humans has low activity relative to other herbivores

This low activity may be due to binding to glycan chains, which interfere with binding to colipase

In the truncated protein, the residues that allow binding to the glycan chains are missing, likely increasing binding efficiency with colipase

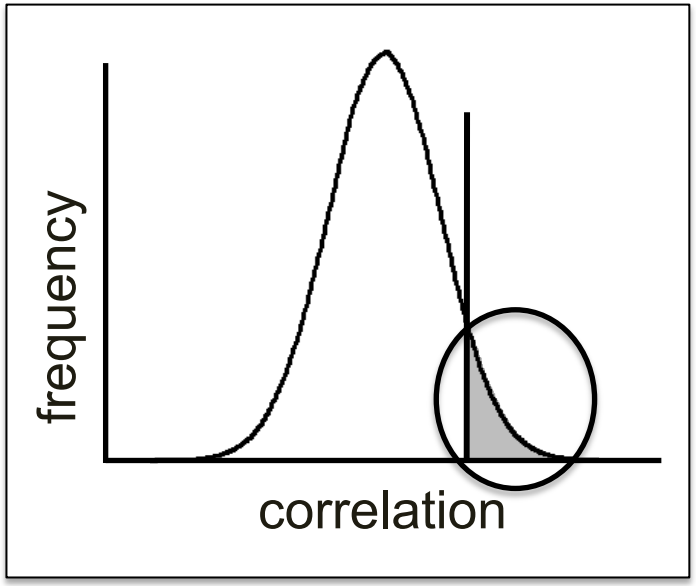
Biochemical evidence suggests that the truncated protein that increases in frequency with cereal use should have higher activity

The frequency of the truncated protein is higher in populations that use cereals



Subtle, but concordant, shifts in allele frequencies across regions suggest polygenic adaptation

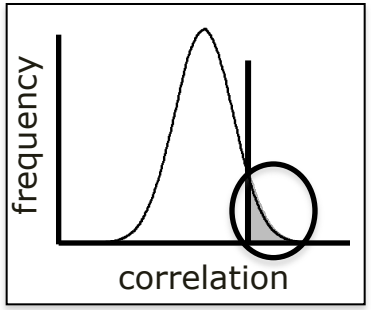
Is there evidence for adaptation to climate and subsistence *overall*?



Is the proportion of ***genic SNPs*** > the proportion ***nongenic SNPs***?

Is the proportion of ***NS SNPs*** > the proportion ***nongenic SNPs***?

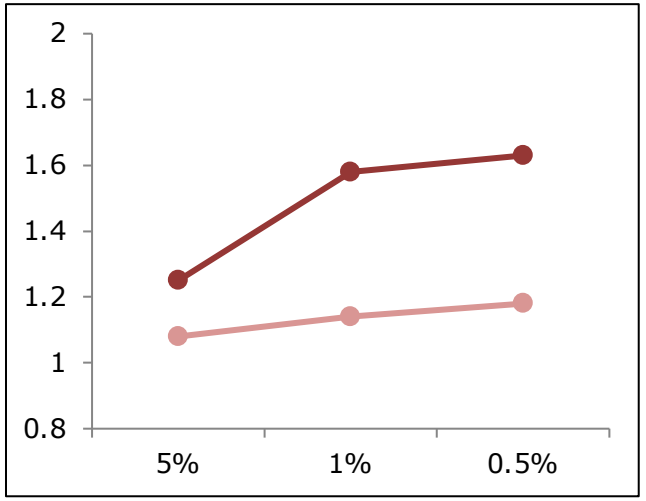
Is there evidence for adaptation to climate and subsistence *overall*?



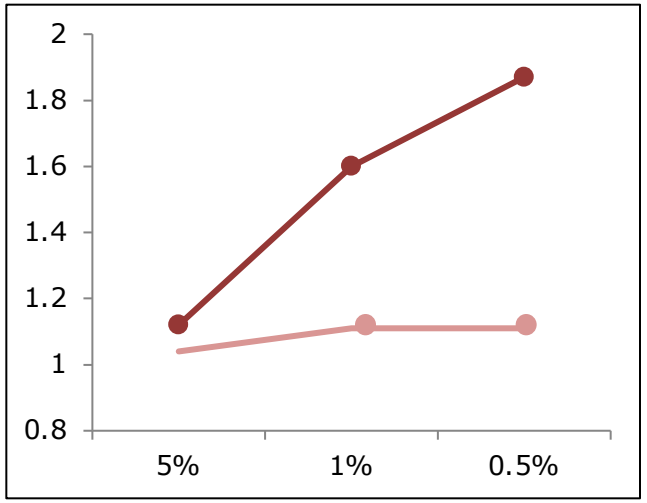
Is the proportion of ***genic SNPs*** > the proportion ***nongenic SNPs***?

Is the proportion of ***NS SNPs*** > the proportion ***nongenic SNPs***?

Climate



Subsistence



● $p < 0.05$

Hancock et al. 2010; Hancock et al. 2011

Reconstructing adaptive history *within populations*

process → pattern

Adaptation (positive selection)

among populations

within populations

increased divergence and differentiation

- **reduced polymorphism**
- **changed SFS**
- **increased LD**

Selective sweep model (Hard sweep)

- Introduced by Maynard Smith and Haigh (1974)
- Selection acts on a single copy of the beneficial allele that enters the population as new mutation after the onset of the selection pressure
- The variant rises in frequency very quickly so that there is not time for other adaptive alleles to arise in the region
- This produces a long tightly-linked haplotype, with an age that is young relative to the genomic background

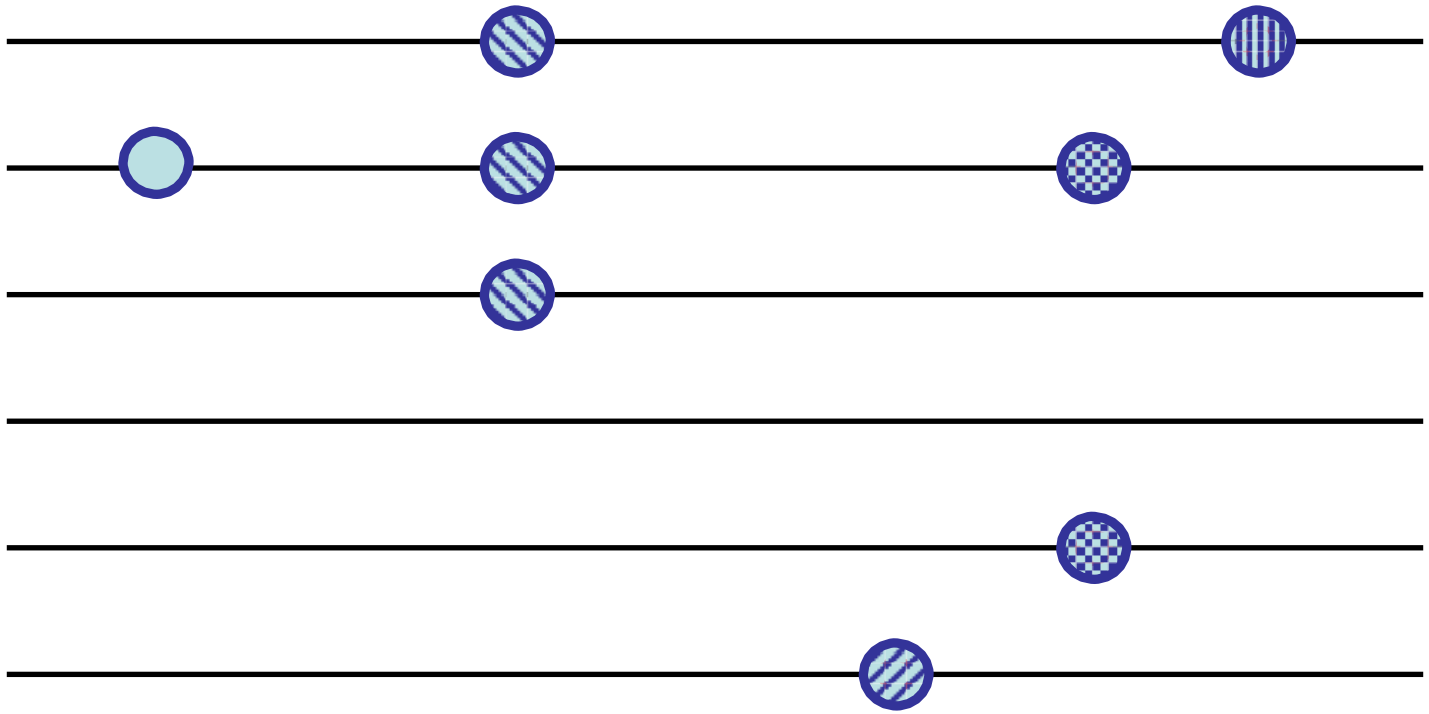
Sweep pattern

A completed sweep leaves a pattern of a haplotype driven quickly to high frequency so that the swept haplotype is younger than average relative to other loci

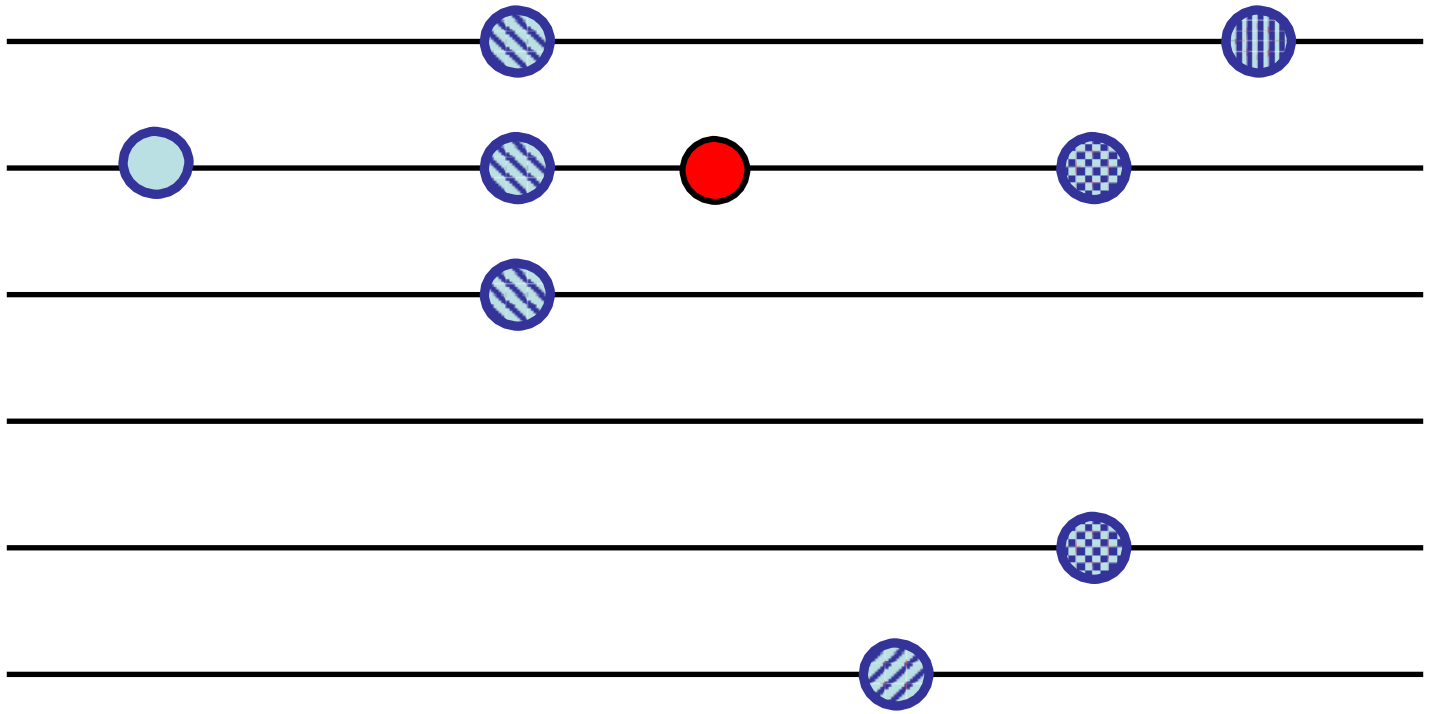
Polymorphism patterns immediately after the sweep:

- Region of reduced heterozygosity
- Excess of high frequency derived variants at the border of the region some time later
- Low frequency variants begin to accumulate

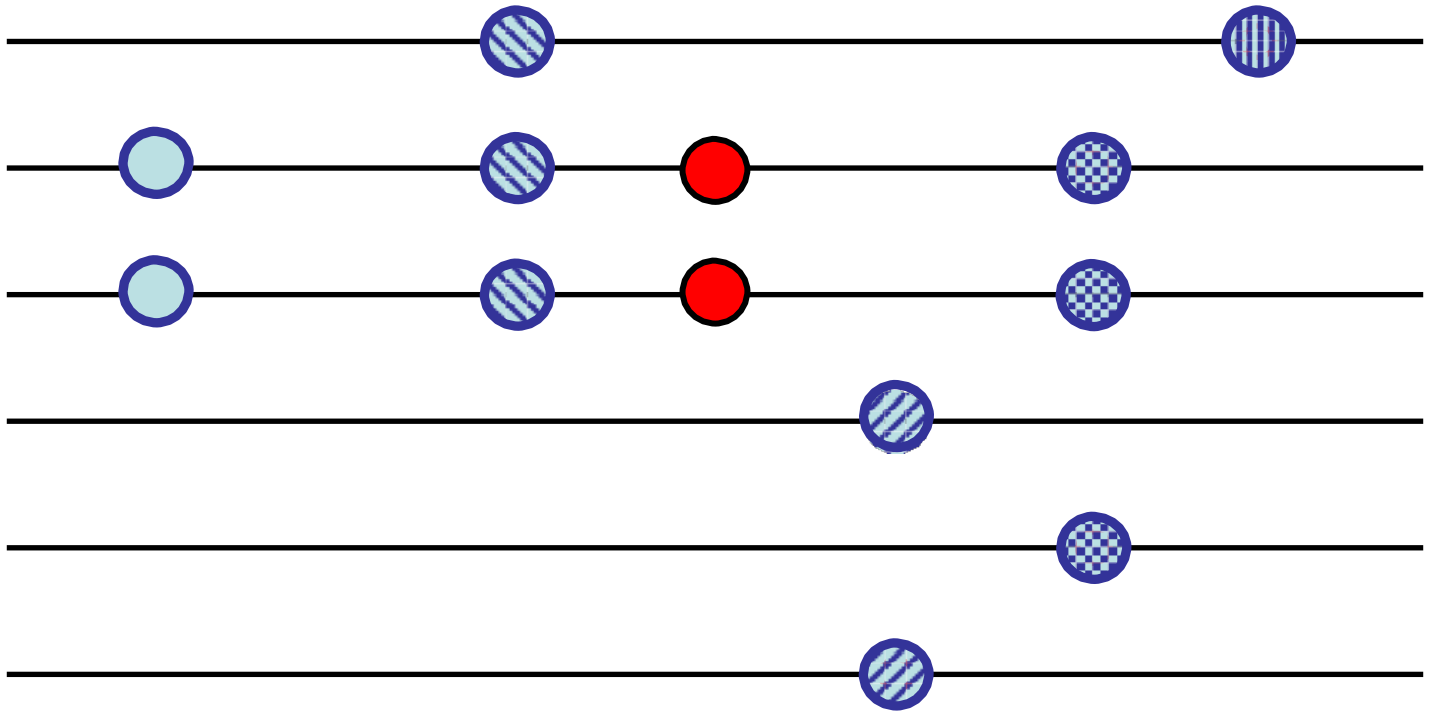
Selective sweep with recombination



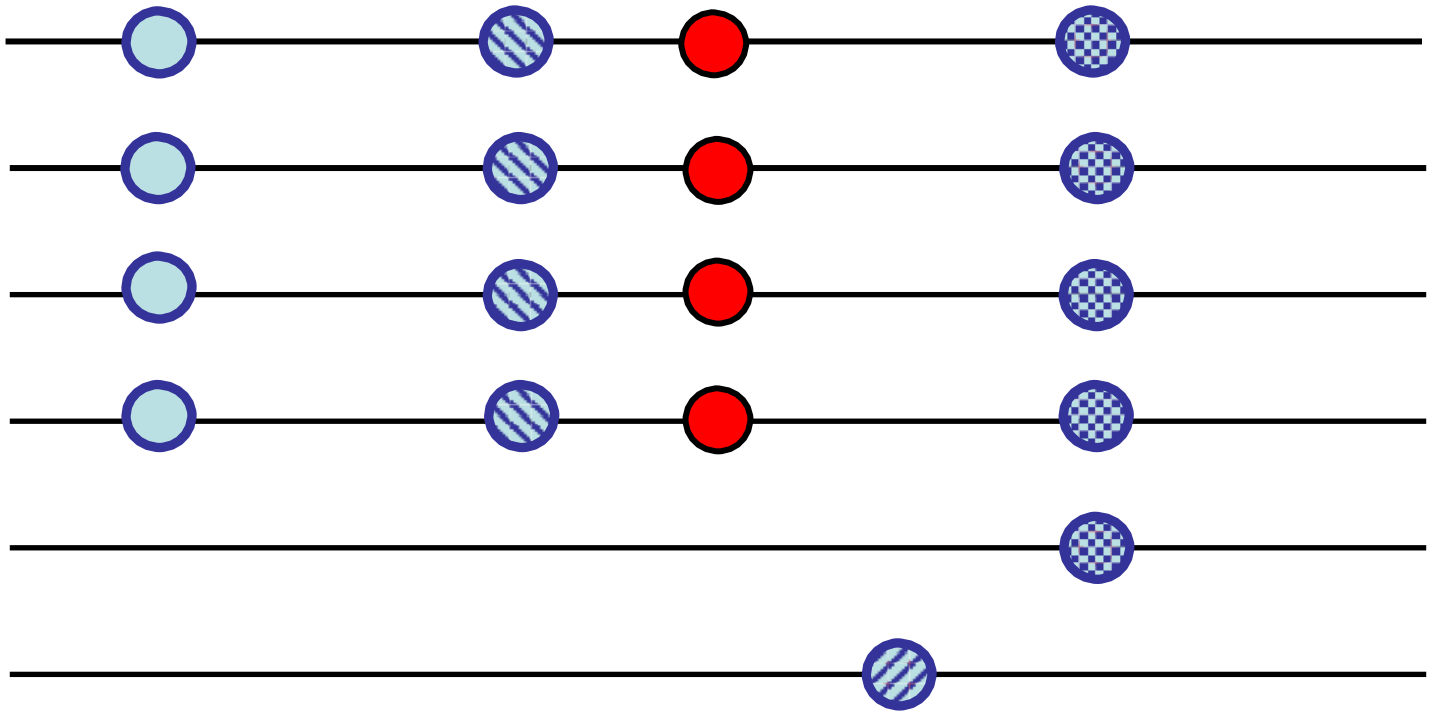
Selective sweep with recombination



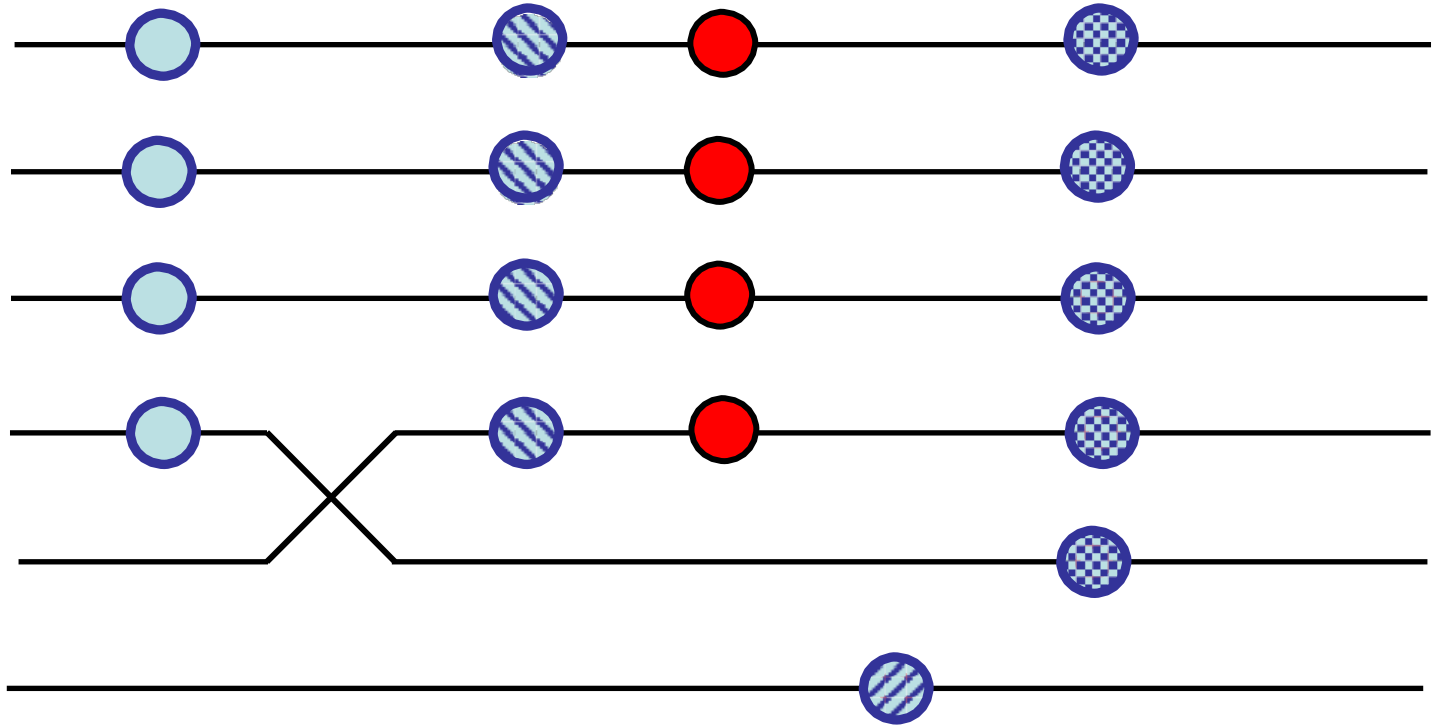
Selective sweep with recombination



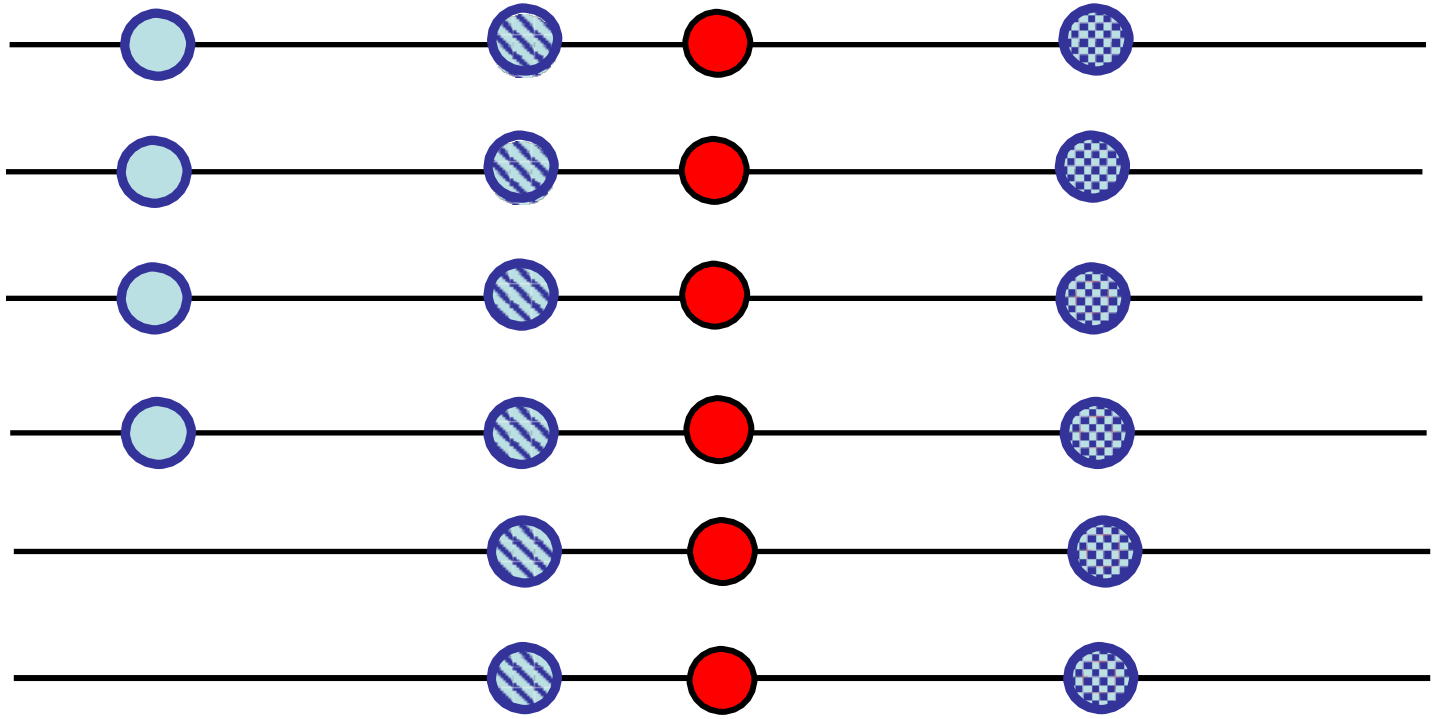
Selective sweep with recombination



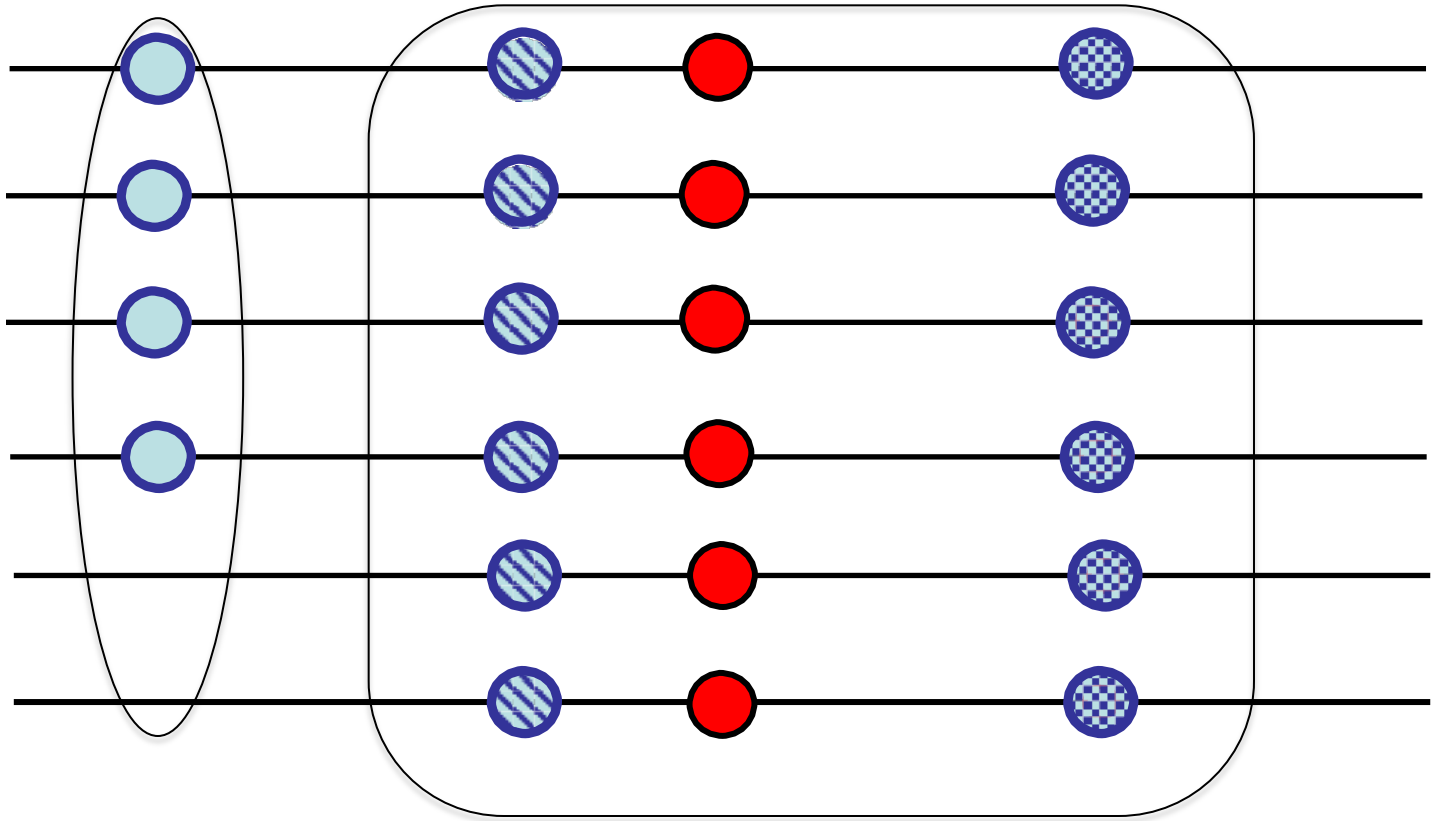
Selective sweep with recombination



Selective sweep with recombination



Selective sweep with recombination



High frequency derived variants at the edge

Region of reduced variation

How can we identify the loci responsible for adaptation?

Use summary statistics based on patterns of polymorphism to identify loci that show departures from neutrality

H_0: neutral evolution

H_1: adaptation

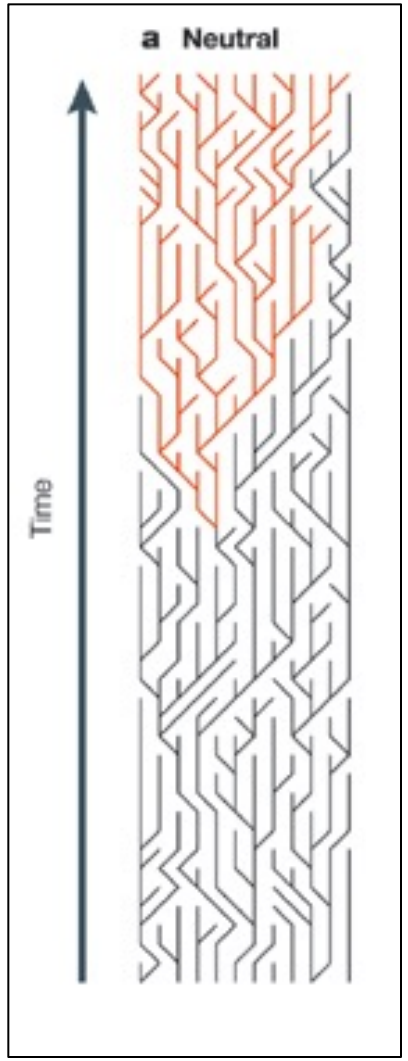
Seems simple enough…

But it can get complicated in many (*most?*) realistic scenarios

- Population structure due to historical demographic events can confound our ability to detect adaptive loci
- Some specific demographic factors that affect results are:
  - population growth
  - bottlenecks
  - hierarchical structure

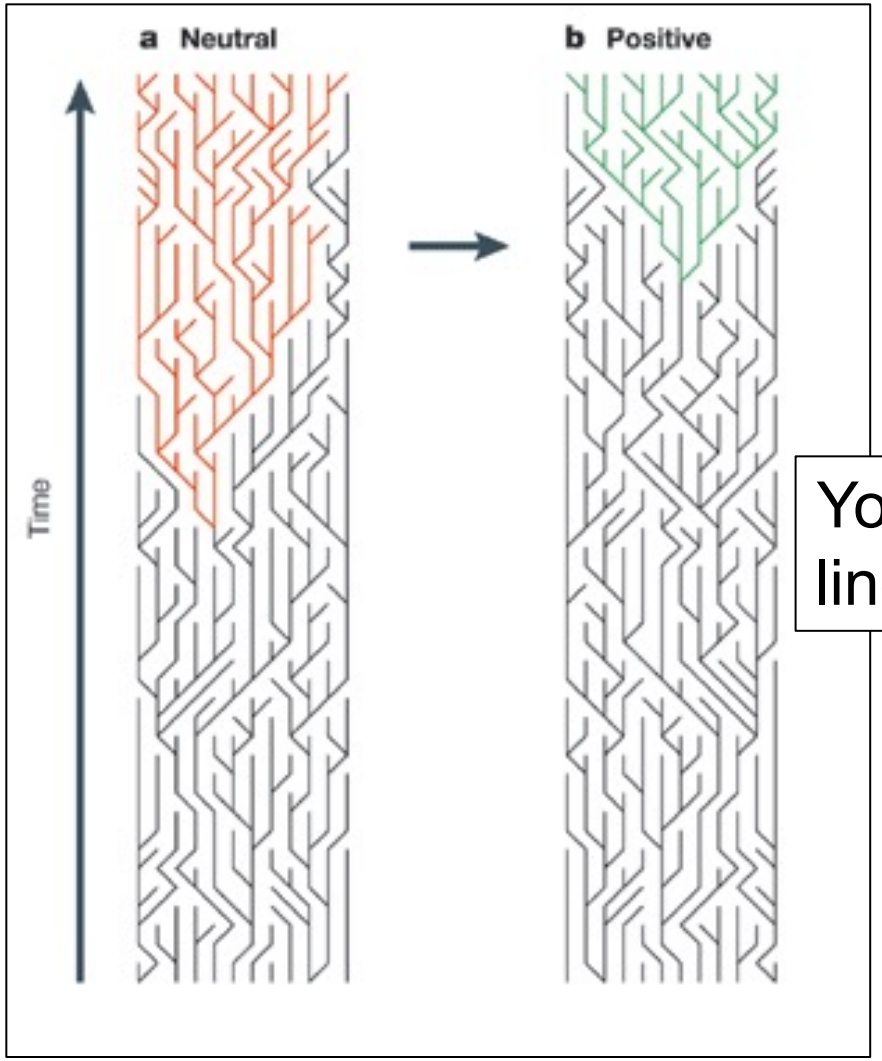
Solution:
Compare patterns at individual loci to entire genome

Effects of natural selection on gene genealogies and allele frequencies



Bamshad and Wooding NRG, 2003
https://doi.org/10.1038/nrg999

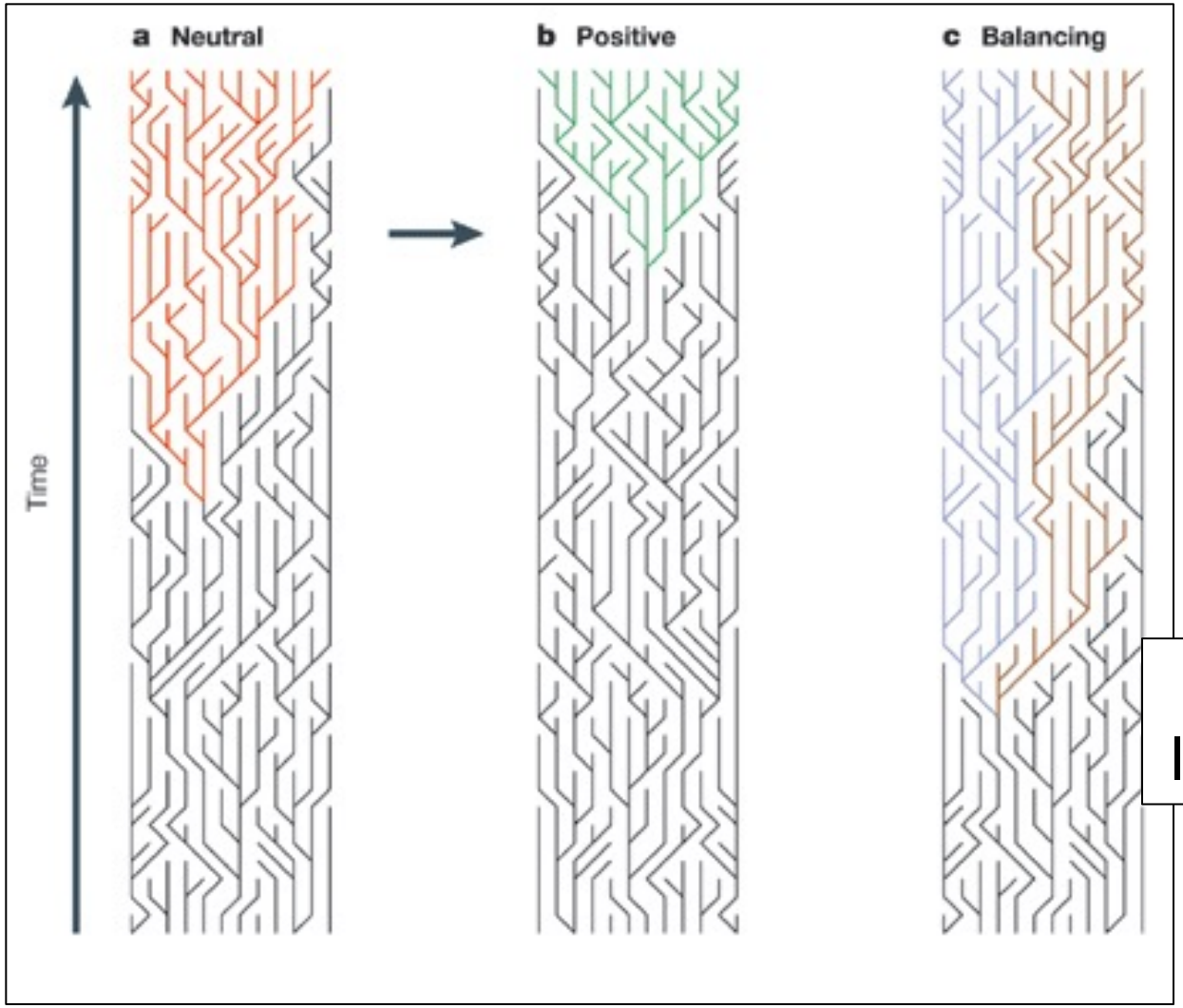
Effects of natural selection on gene genealogies and allele frequencies



Younger lineages

Bamshad and Wooding NRG, 2003
https://doi.org/10.1038/nrg999

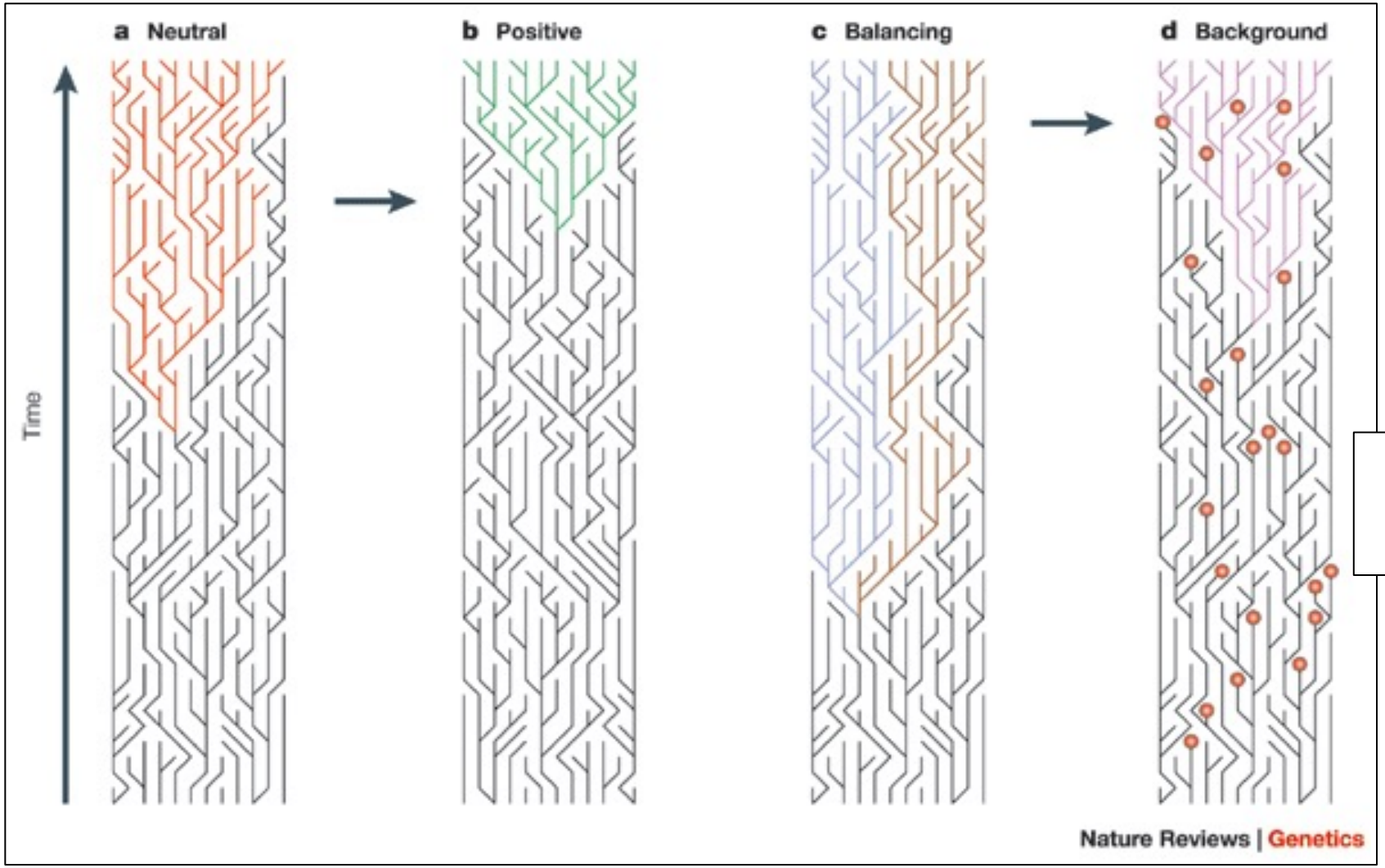
Effects of natural selection on gene genealogies and allele frequencies



Older lineages

Bamshad and Wooding NRG, 2003
https://doi.org/10.1038/nrg999

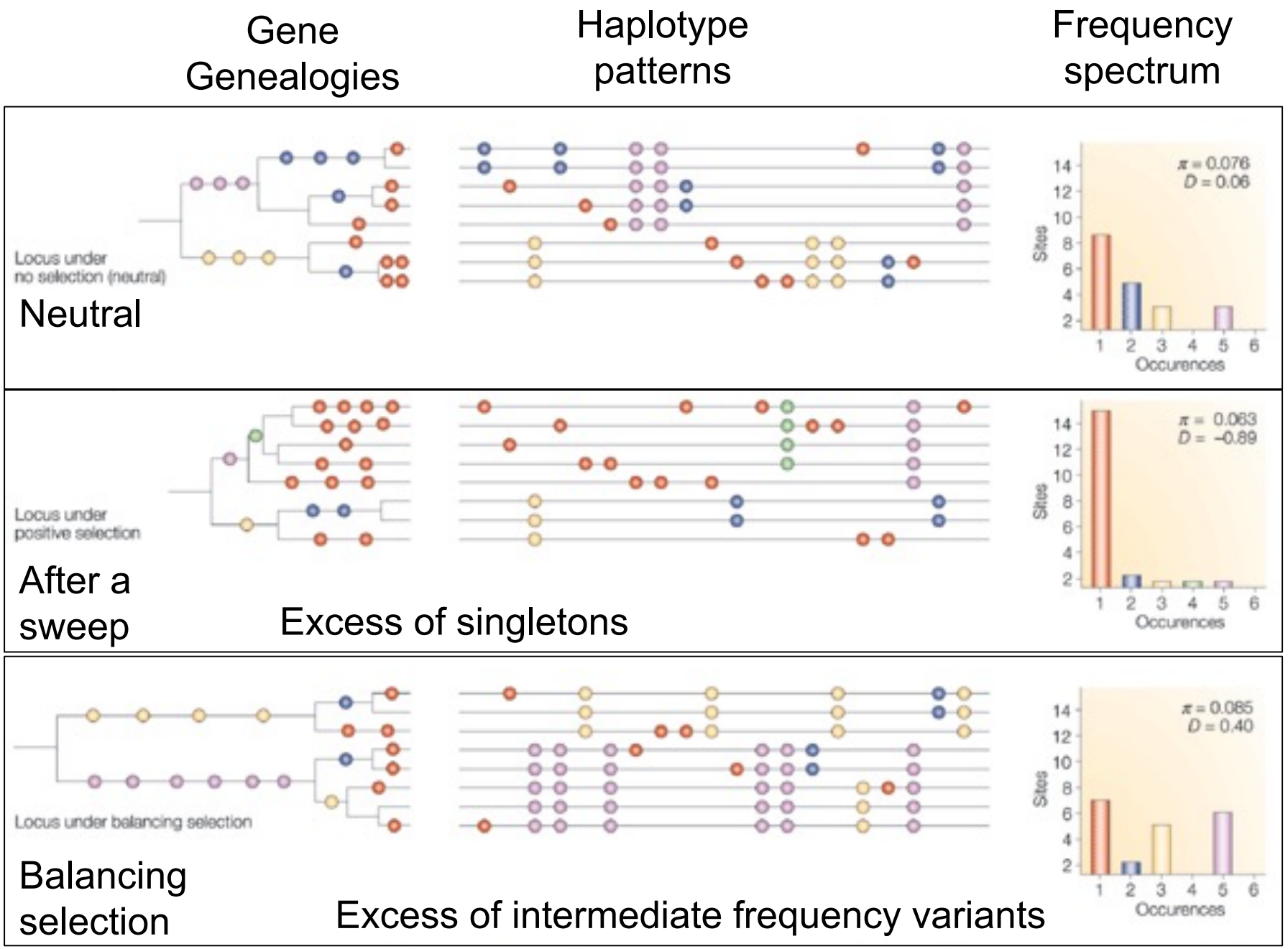
Effects of natural selection on gene genealogies and allele frequencies



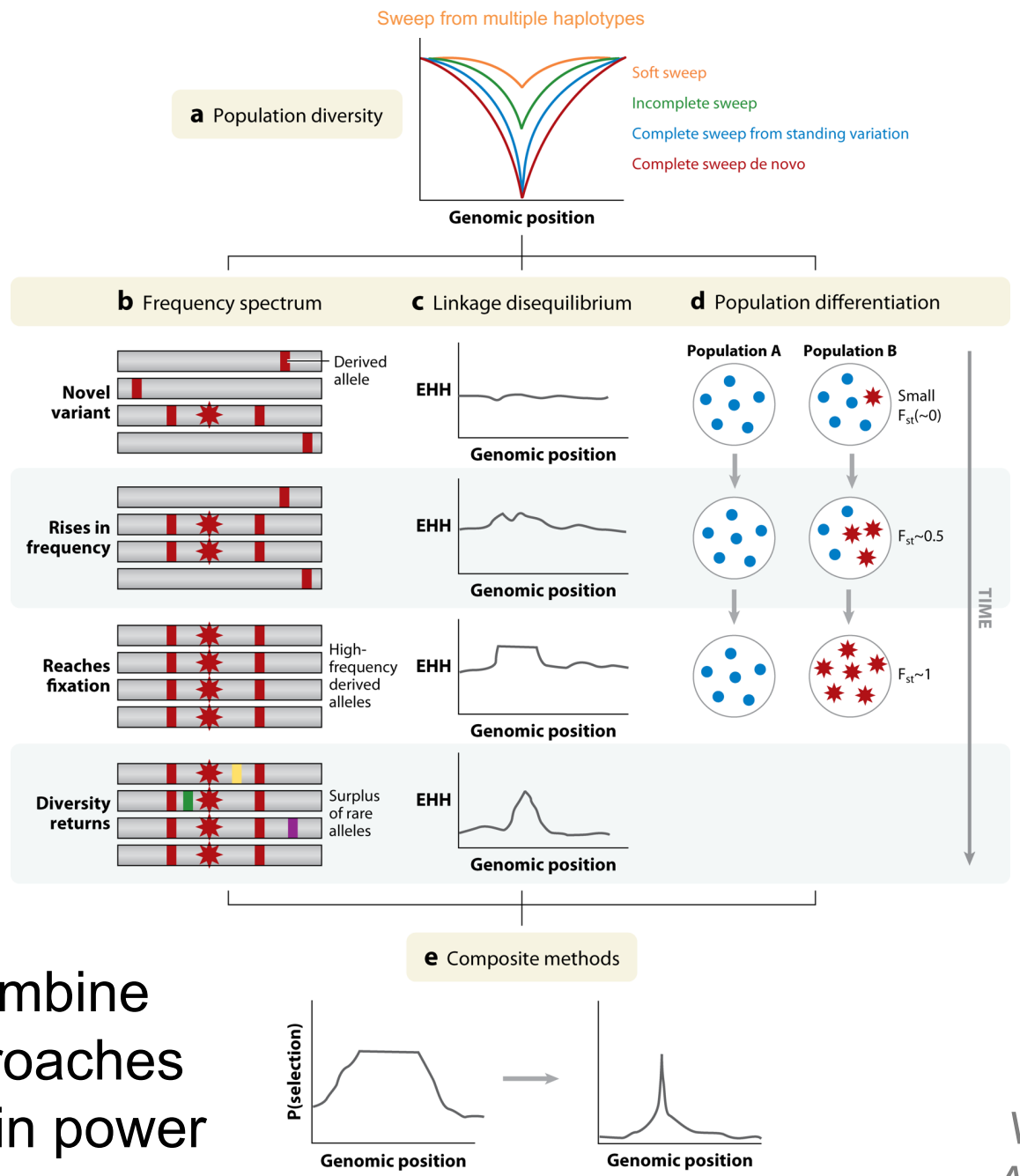
Lost lineages

Bamshad and Wooding NRG, 2003
https://doi.org/10.1038/nrg999

Signatures of selection



Signatures of positive selection



Combine approaches to gain power

Vitti et al., 2013 Ann. Rev Genet

Sweep signatures:
Tests based on polymorphism

process → pattern

Adaptation (positive selection)

reduced polymorphism, changes in the SFS, increased LD

Reduction of polymorphism

Compare local and global patterns of variation

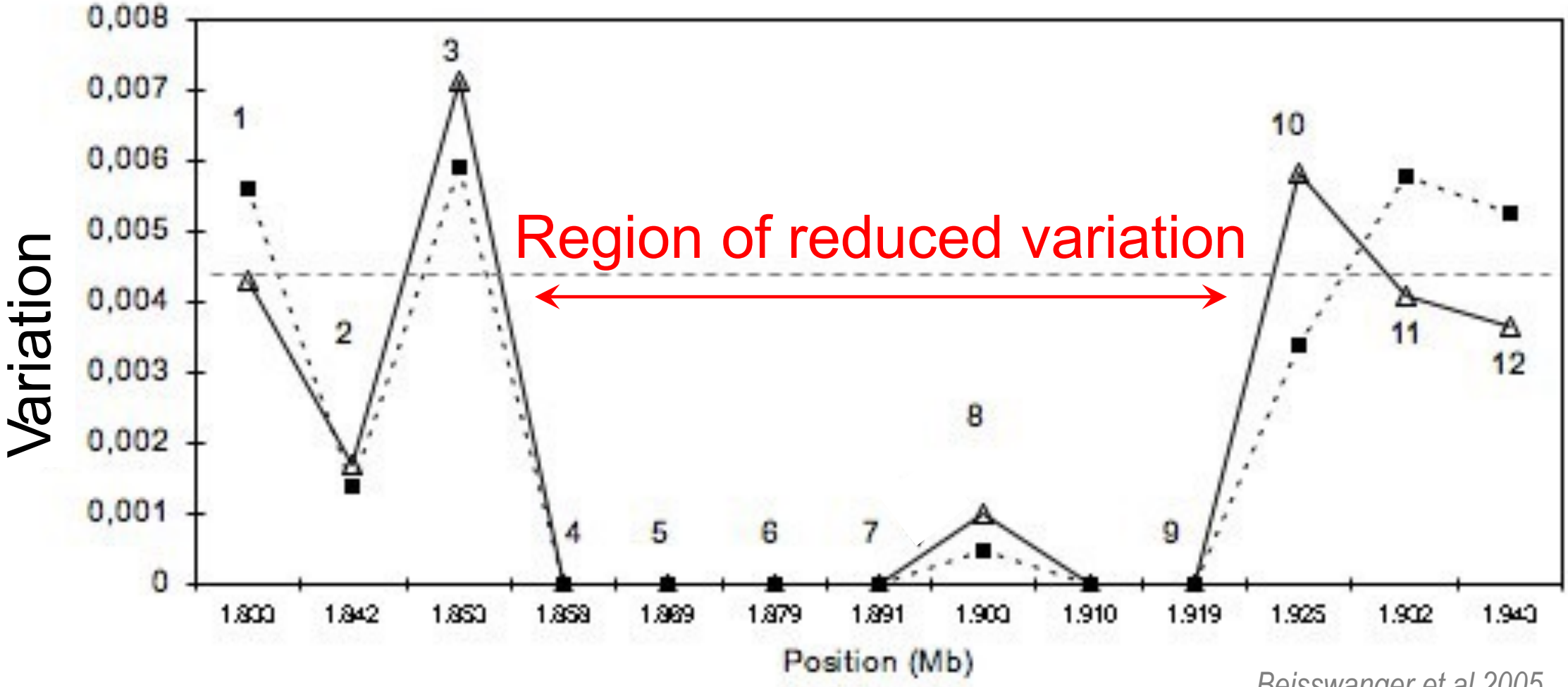
Caveats:

- Neutral coalescent is very variable
  - Even more so in the presence of population bottlenecks
- There are alternative causes for valleys of low variation
  - Selective constraints (purifying selection in coding regions)
  - Locally reduced mutation rate

➢ There is no test exclusively based on reduced variation

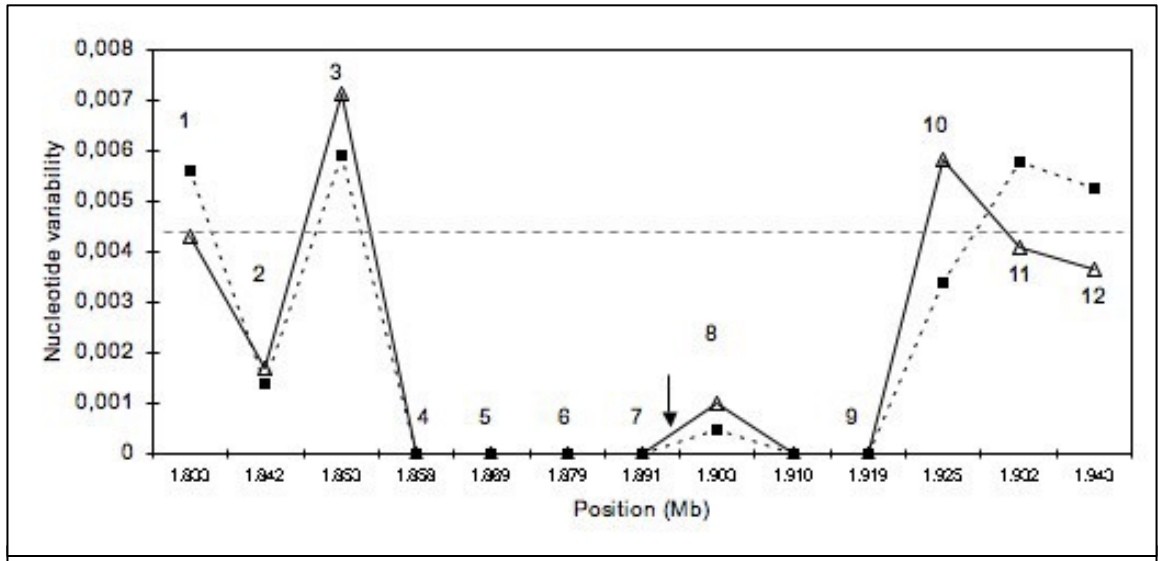
Selective sweep

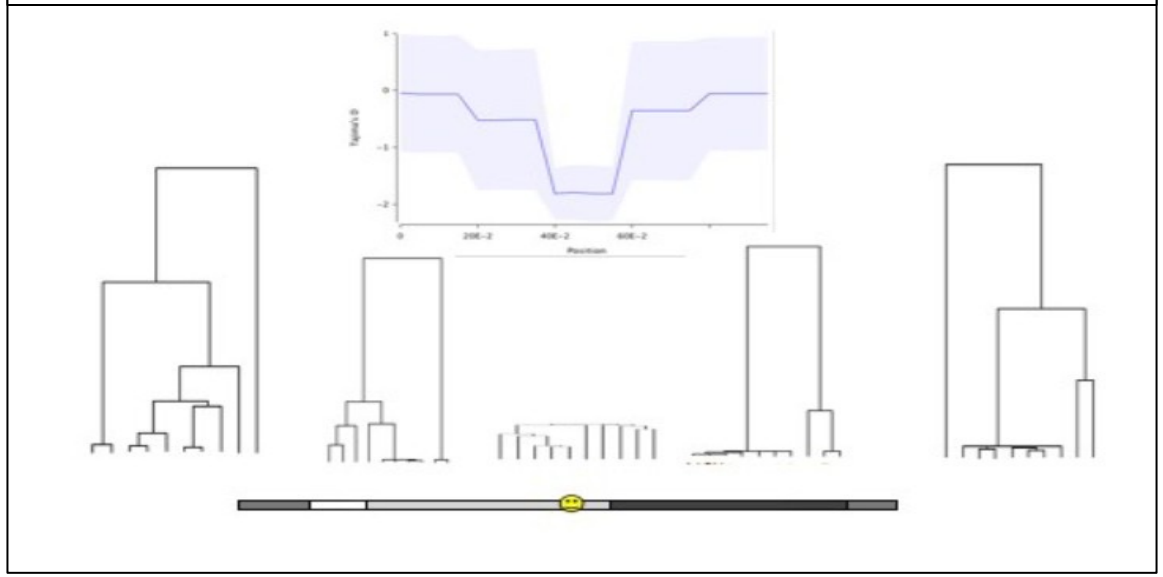
(in European *Drosophila melanogaster*)



Beisswanger et al 2005

Sweeps locally skew the frequency spectrum

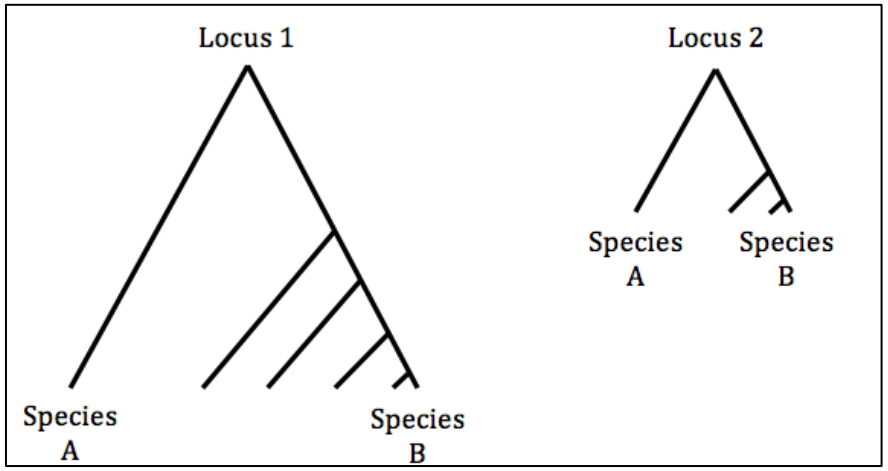




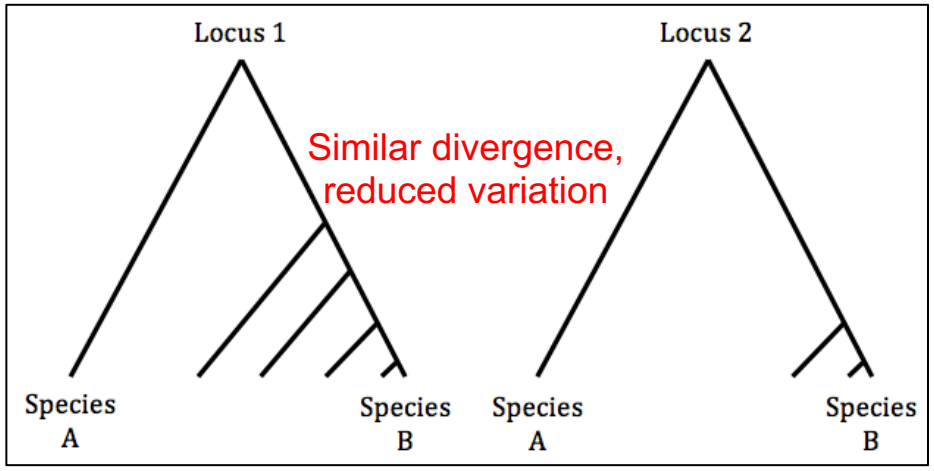
HKA Test (Hudson, Kreitman, Aguadé)

Compares ***polymorphism within species*** to ***divergence between populations***

Neutral Locus



Positively selected locus



Uses comparison between divergence and diversity to normalize for rate differences (i.e., variation in purifying selection) across loci

HKA test

Compares divergence relative to polymorphism

Inter-locus test of reduced polymorphism relative to divergence

Focal locus: $\theta_1 = 4N_e\mu_1$, $d_1 = 2t\mu_1$

locus 2 (3,4,…): $\theta_2 = 4N_e\mu_2$, $d_2 = 2t\mu_2$

- Positive selection for: $\boxed{\frac{\theta_1}{d_1} < \frac{\theta_2}{d_2}}$

- Similar to McDonald-Kreitman but looking for the opposite signal

Hudson, Kreitman, Aguade 1987

Reconstructing adaptive history *within populations*

process → pattern

Adaptation (positive selection)

among populations

within populations

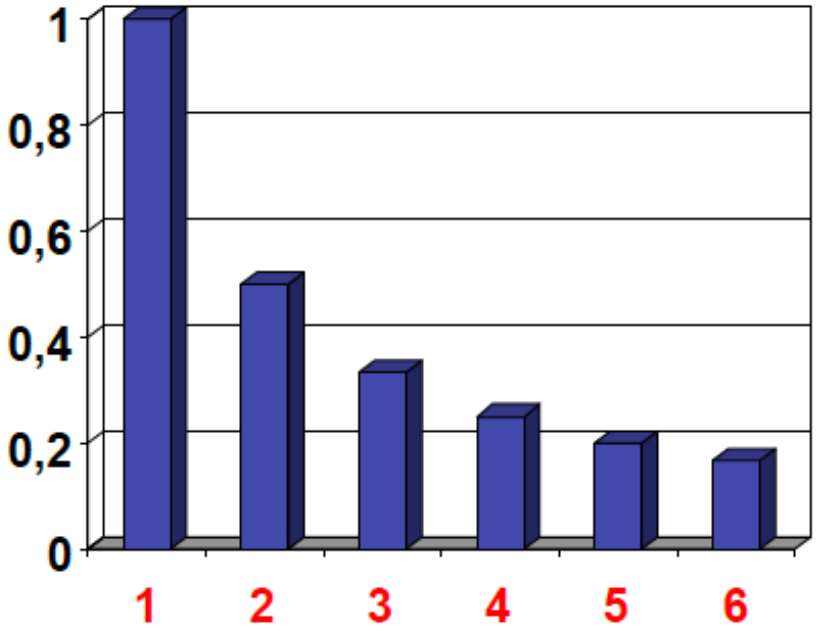
increased divergence and differentiation

- reduced polymorphism
- **changes in the SFS**
- **increased LD**

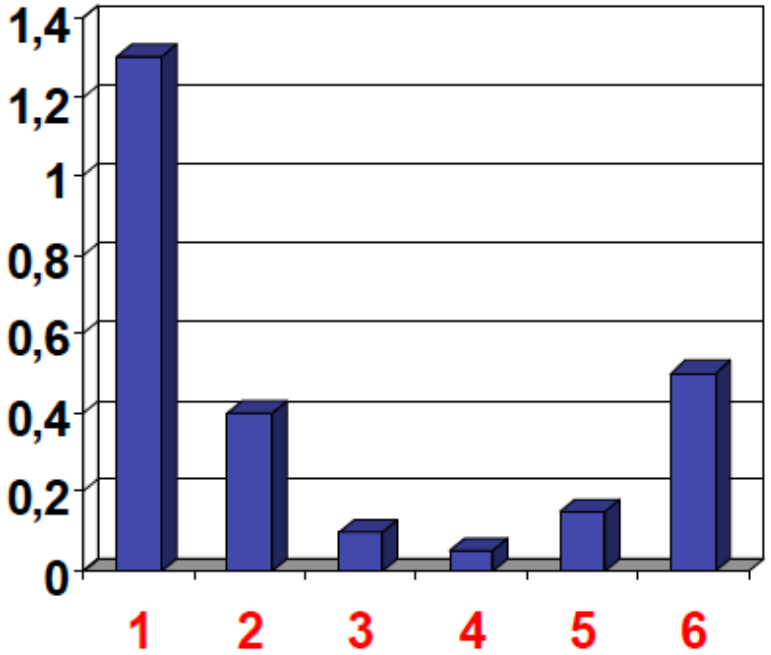
Selective sweeps: effects on the frequency spectrum

How is polymorphism distributed across frequency bins?

Selection footprint on the site frequency spectrum:



standard neutral model

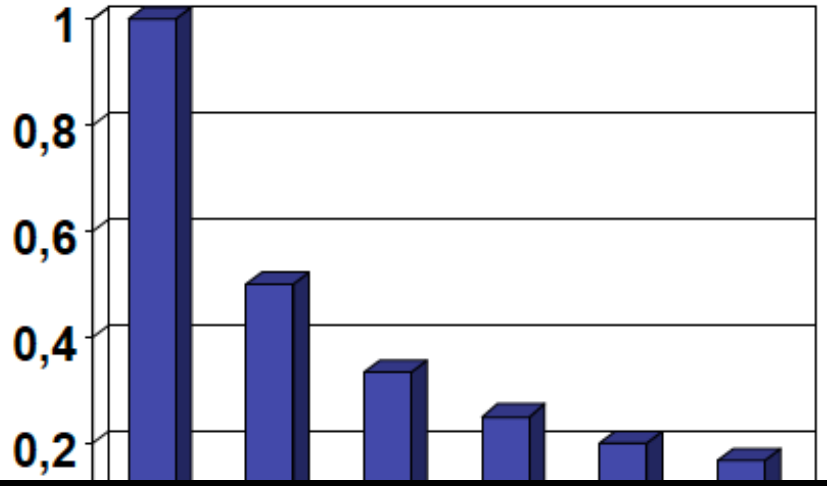


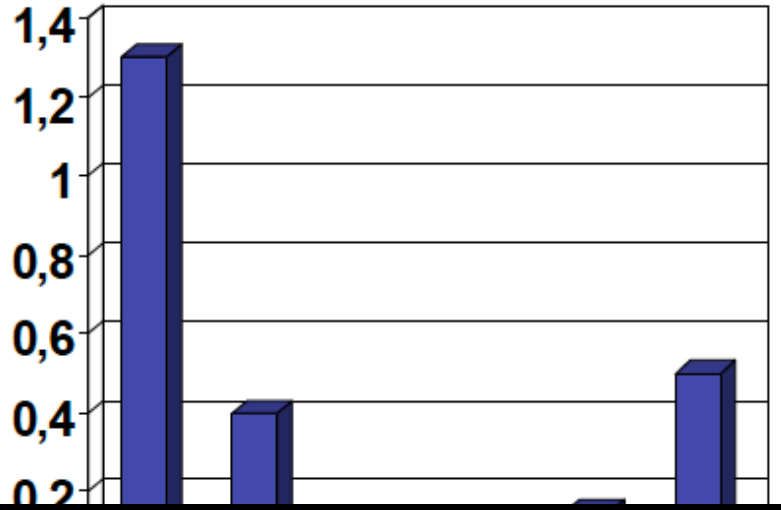
selective sweep

Selective sweeps: effects on the frequency spectrum

How is polymorphism distributed across frequency bins?

Selection footprint on the site frequency spectrum:





Increase in low frequency variants and high frequency derived variants after a selective sweep

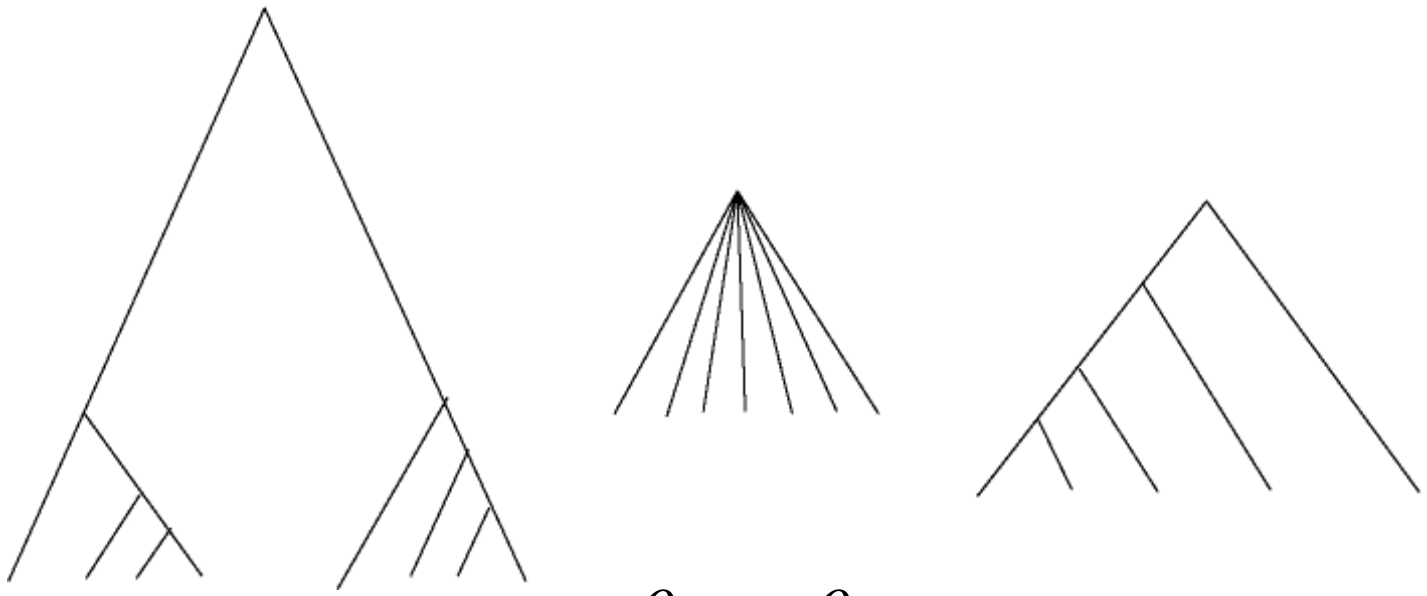
Site frequency spectrum based tests

- Define different estimators of θ $(4Nu)$ from the site frequency spectrum.
- Compare these estimators to detect deviations from neutrality.

Tajima's D

Compares estimates of θ based on the number of segregating sites (S -> θ_S) and π (the number of pairwise differences) in the sample

$$D = \frac{\widehat{\theta}_\pi - \widehat{\theta}_S}{\sqrt{\widehat{\mathbf{Var}}[\widehat{\theta}_\pi - \widehat{\theta}_S]}}$$



$\theta_\pi > \theta_S$

D positive

$\theta_\pi < \theta_S$

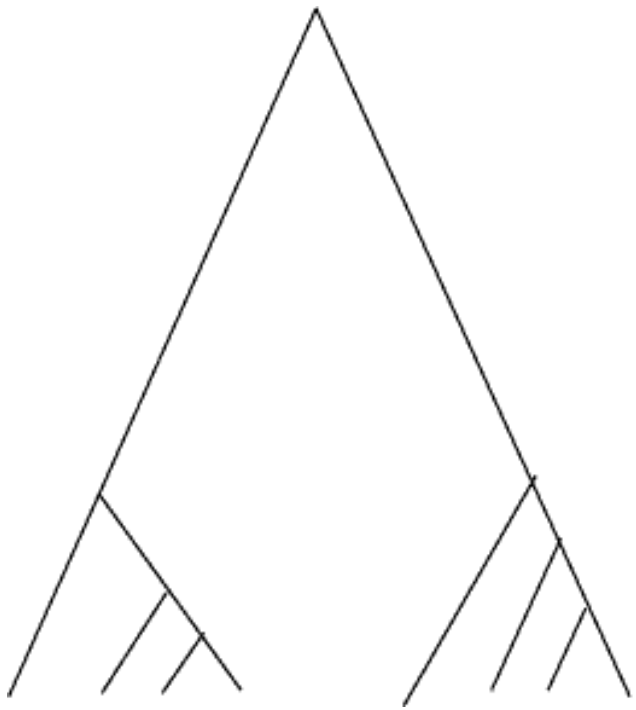
D negative

$\theta_\pi = \theta_S$

neutral

Tajima's D

Can get this from population bottleneck
Or balancing selection



$$\theta_\pi > \theta_S$$

D positive

Can get this from population growth
Or a sweep

$$\theta_\pi < \theta_S$$

D negative

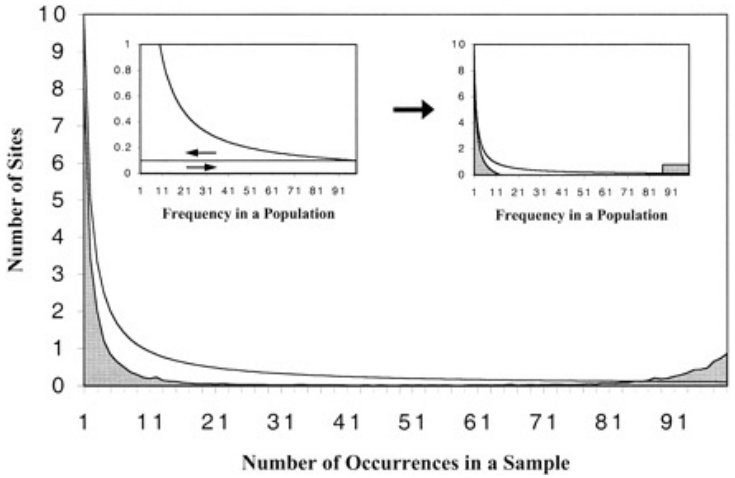
Fay and Wu's H Test

Compares estimates of θ based on k (which captures information about high frequency derived variants and π (the number of pairwise differences)

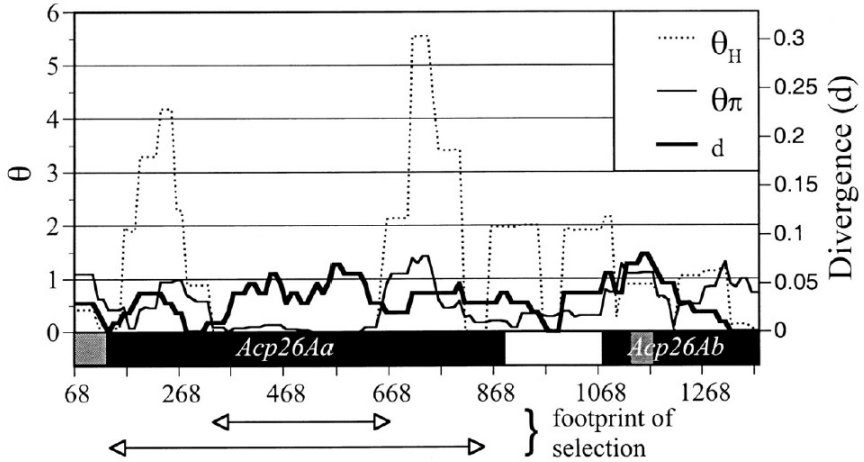
$$\widehat{\theta}_H = \sum_{i=1}^{n-1} \frac{2S_i i^2}{n(n-1)}$$

$$H = \hat{\theta}_\pi - \hat{\theta}_H$$

Pattern in simulations

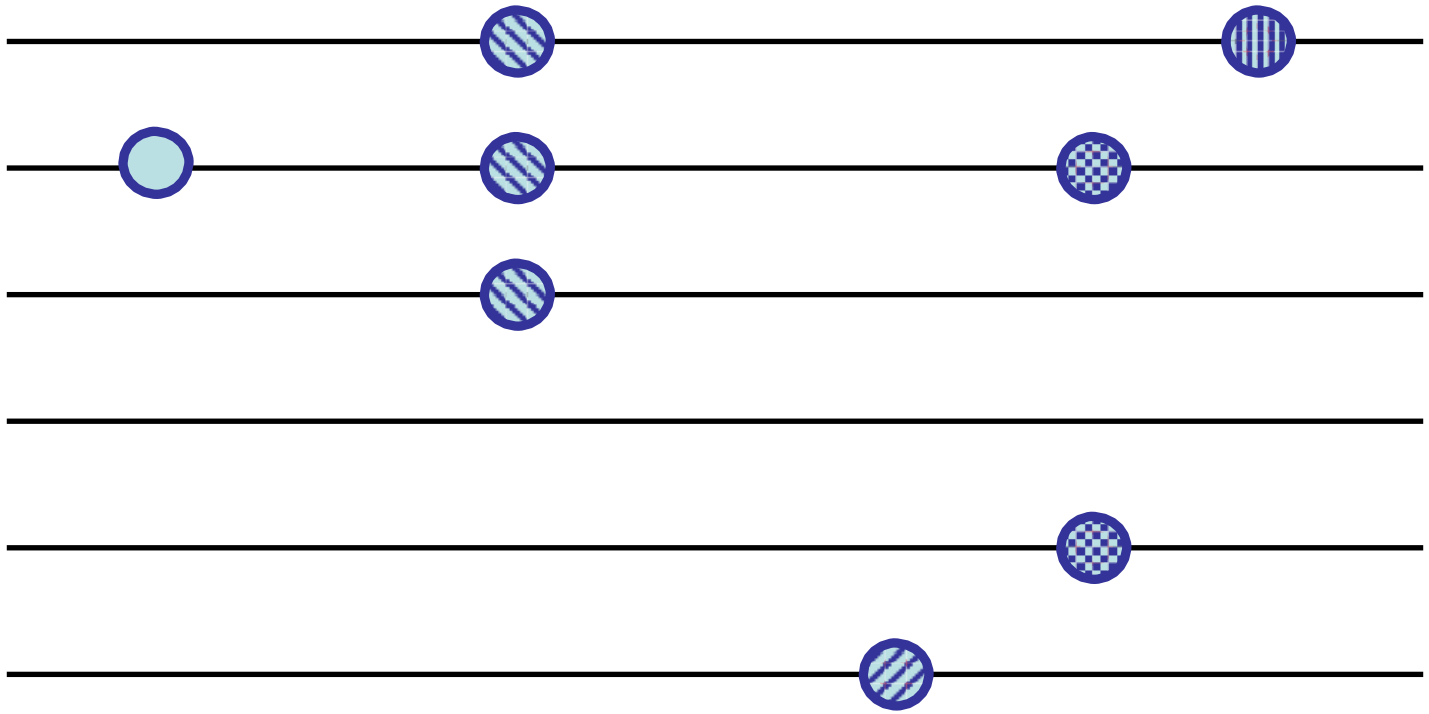


Acp26Aa

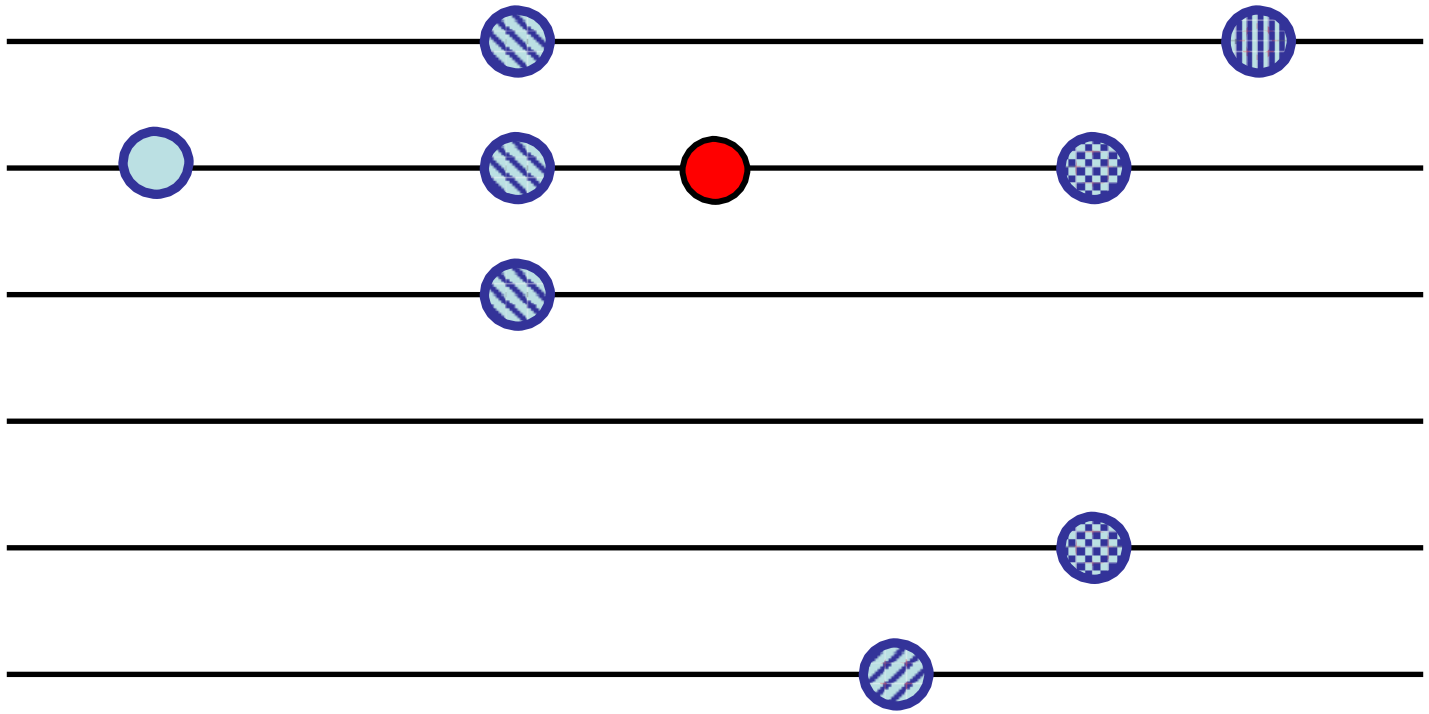


D. melanogaster accessory gland gene

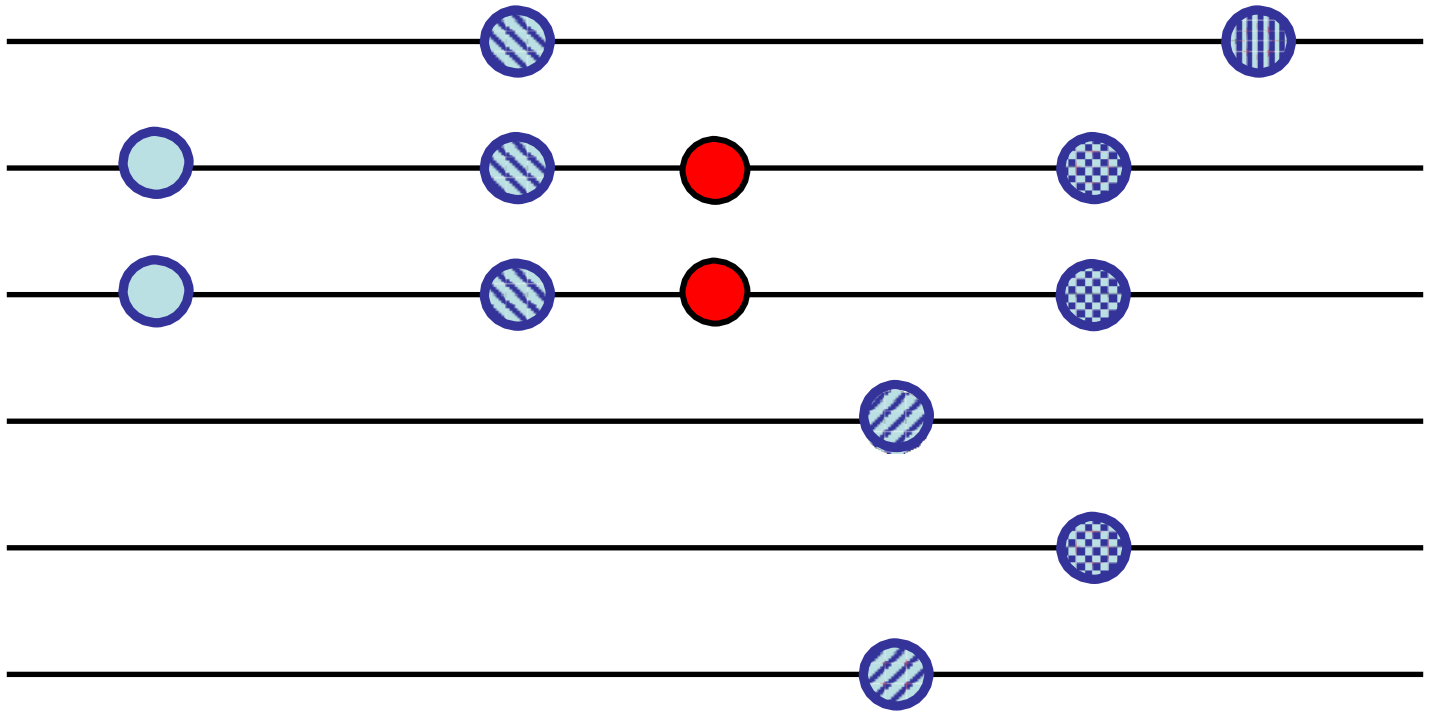
Selective sweep with recombination



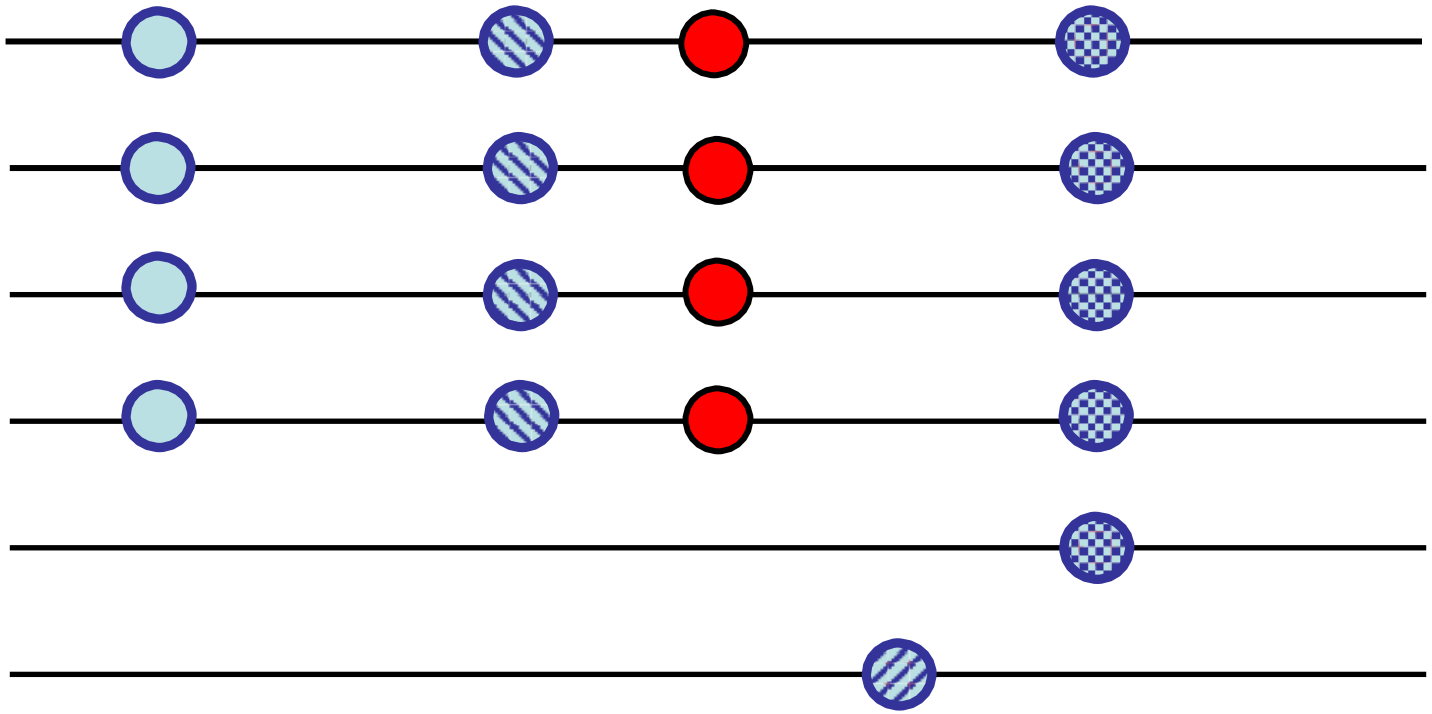
Selective sweep with recombination



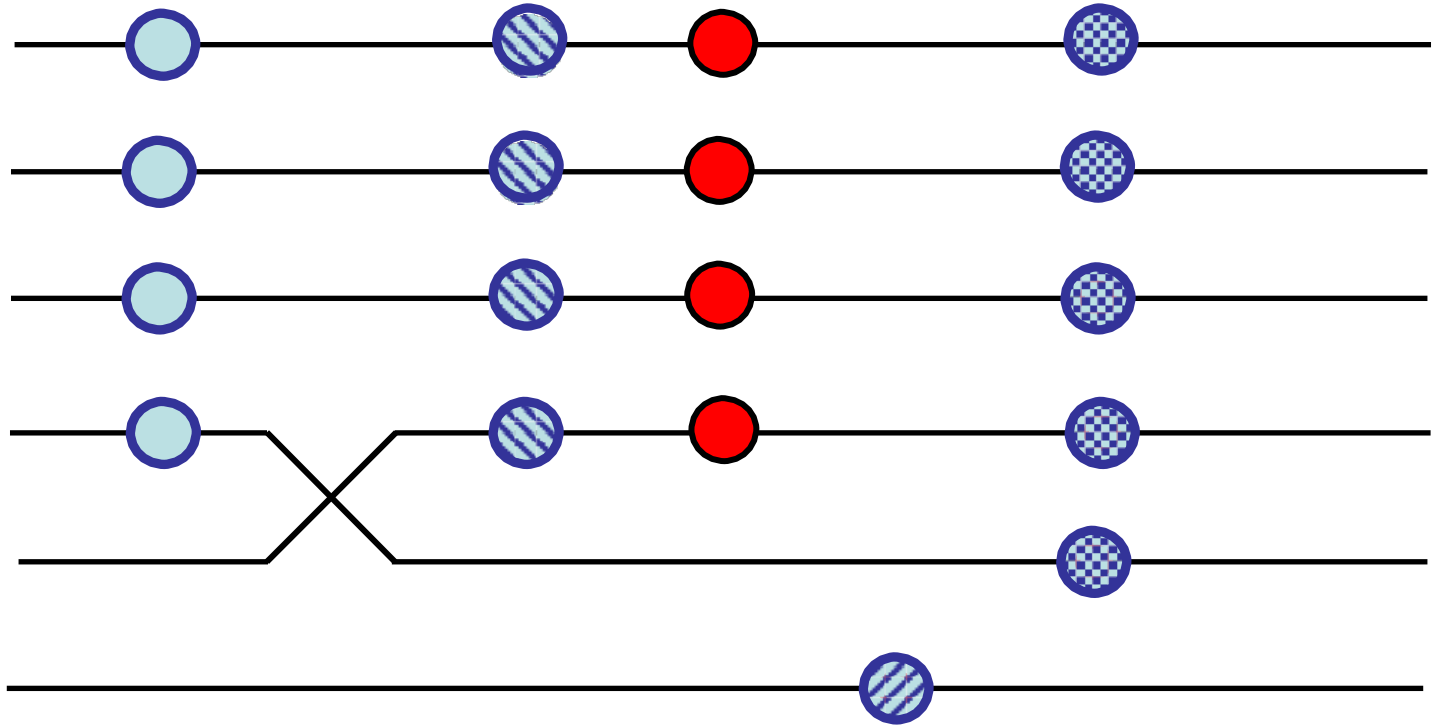
Selective sweep with recombination



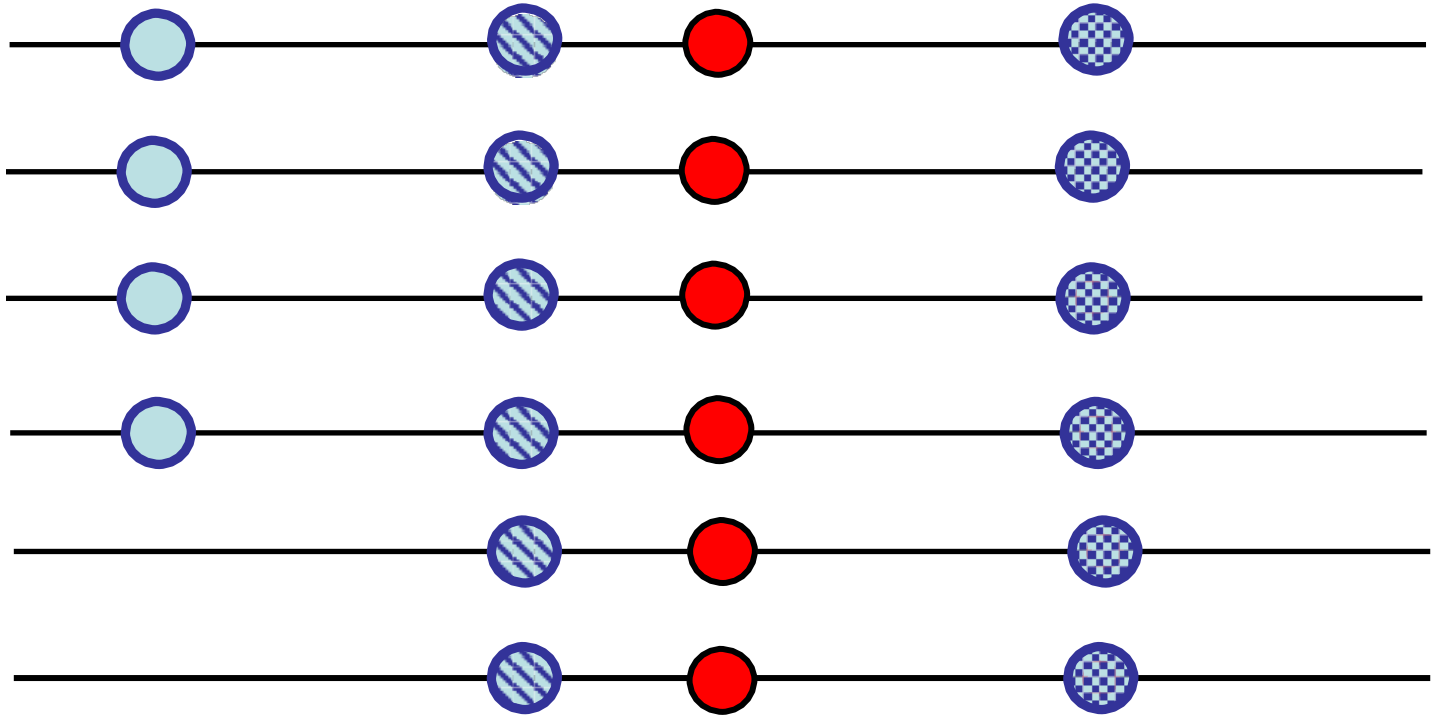
Selective sweep with recombination



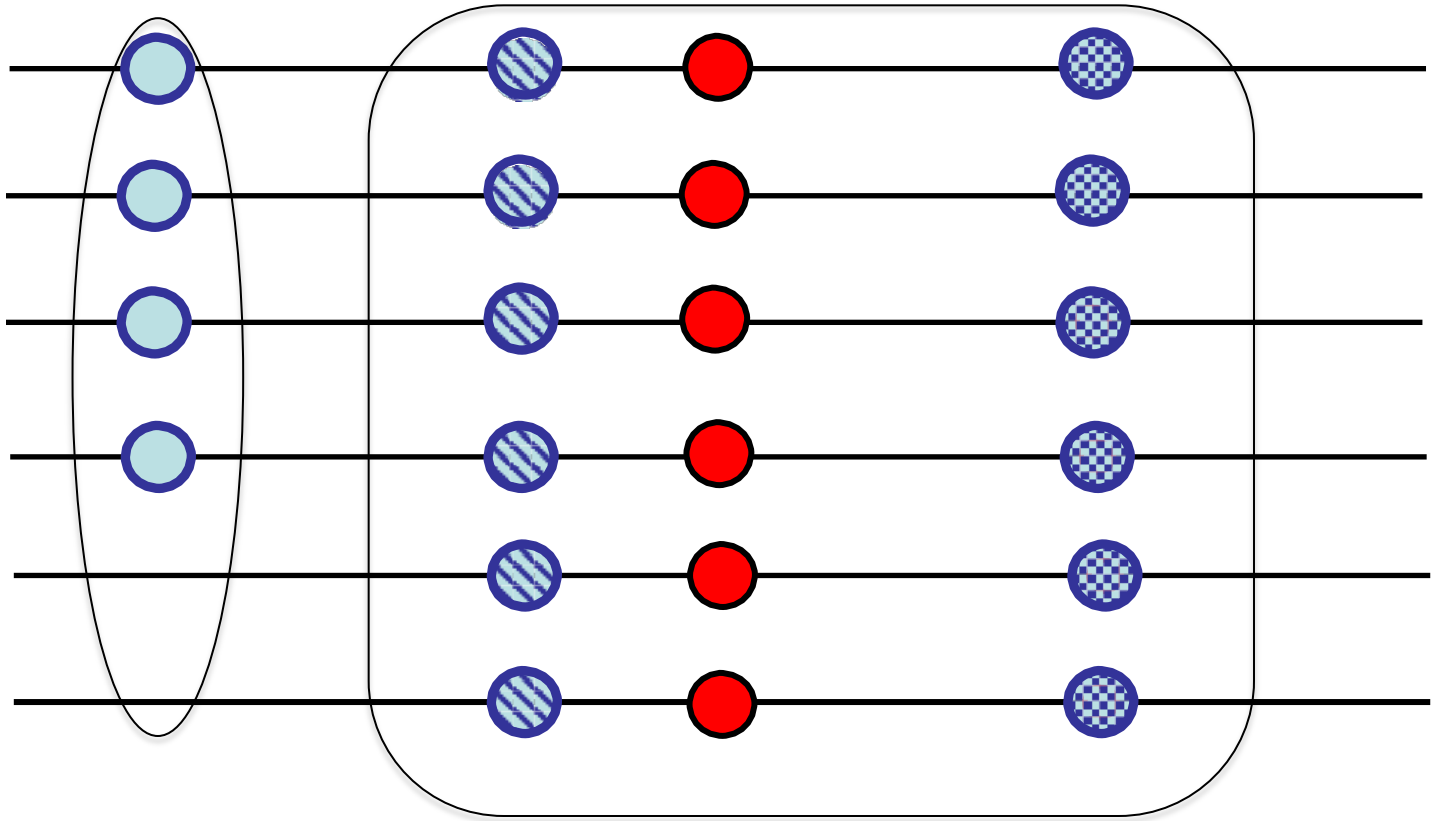
Selective sweep with recombination



Selective sweep with recombination



Selective sweep with recombination



High frequency derived variants at the edge

Region of reduced variation

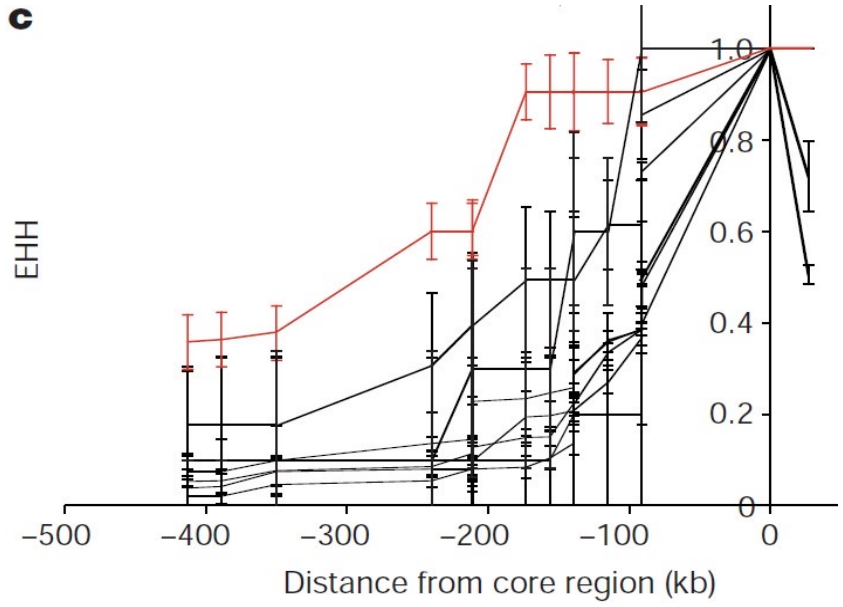
Linkage disequilibrium: haplotype tests

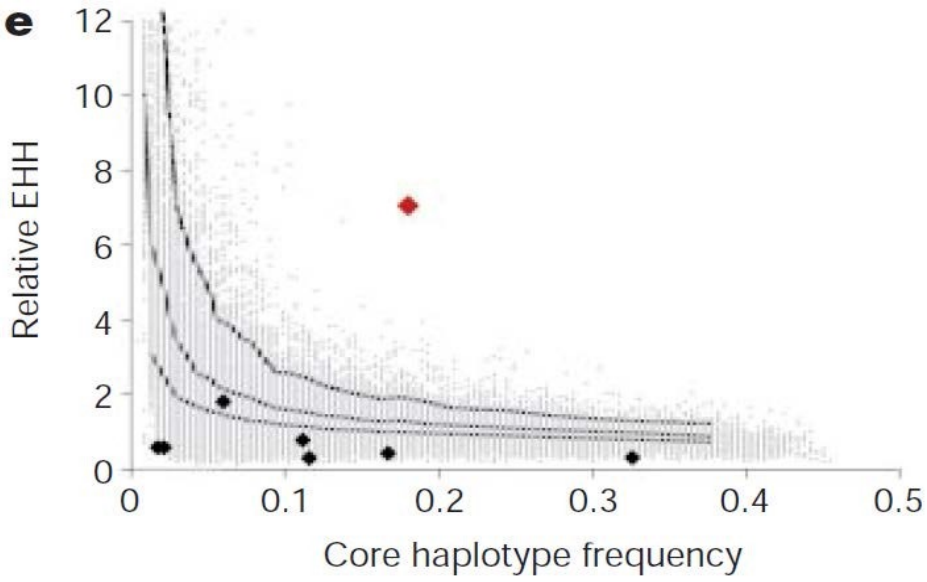
Many different approaches:

- Number of different haplotyes
  - Low number for given polymorphism indicates selection
- Frequency of the major haplotype
  - Unusually high frequency indicates selection
- Length and frequency of core haplotype
  - EHH: extended haplotype homozygosity measures the reduction in frequency of a core haplotype. Slow reduction indicates selection

Linkage disequilibrium: EHH test

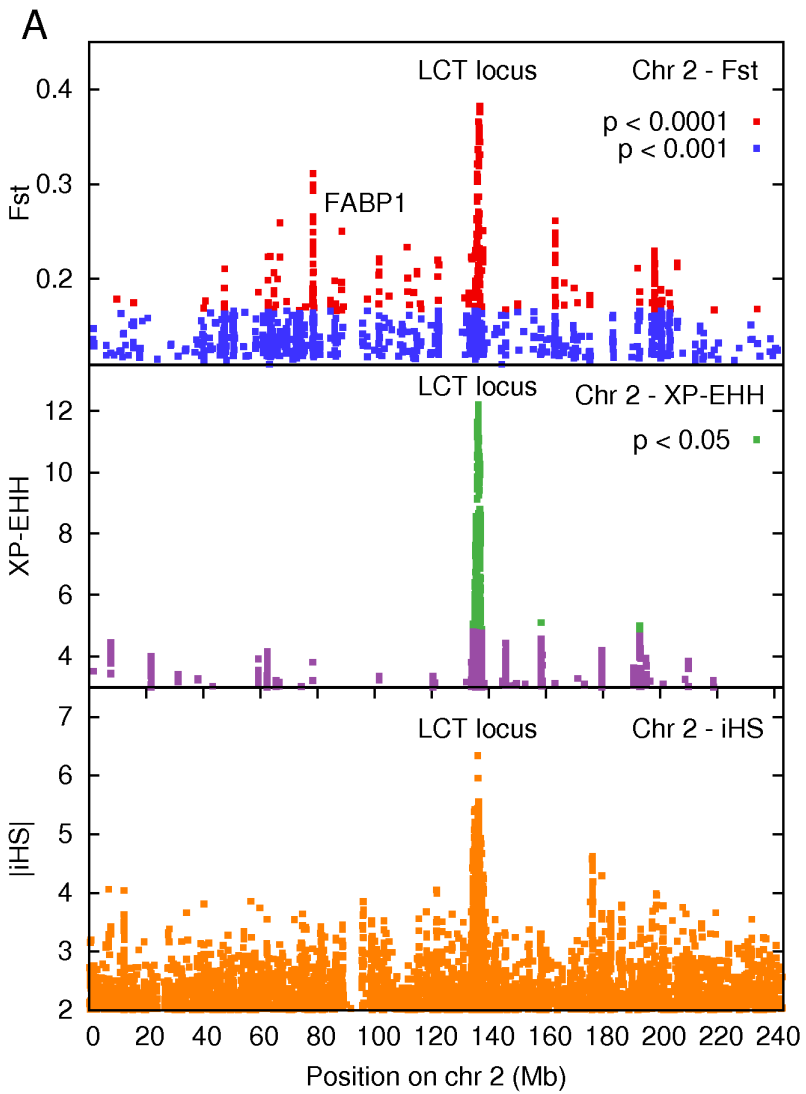
- Logic: High frequency haplotypes typically do not extend over a long region. With positive selection, one long major haplotype is created.
- EHH score: homozygosity of core haplotype up to a given distance relative to other haplotypes (identifies incomplete sweeps!)





Sabeti et al 2002

Lactase persistence region in Europeans has reduced haplotype heterozygosity



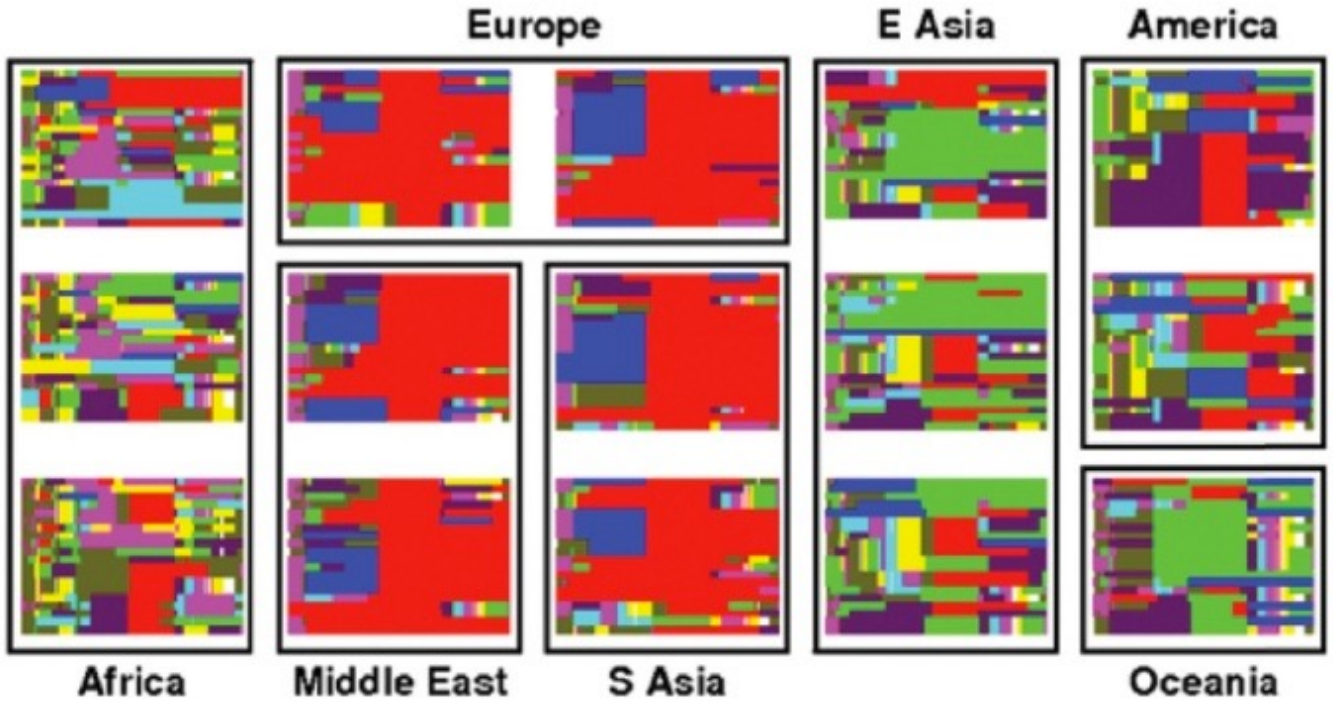
Differentiation (F_{ST})

Differentiation + haplotype-based score

Integrated haplotype score (iHS) in the lactase region

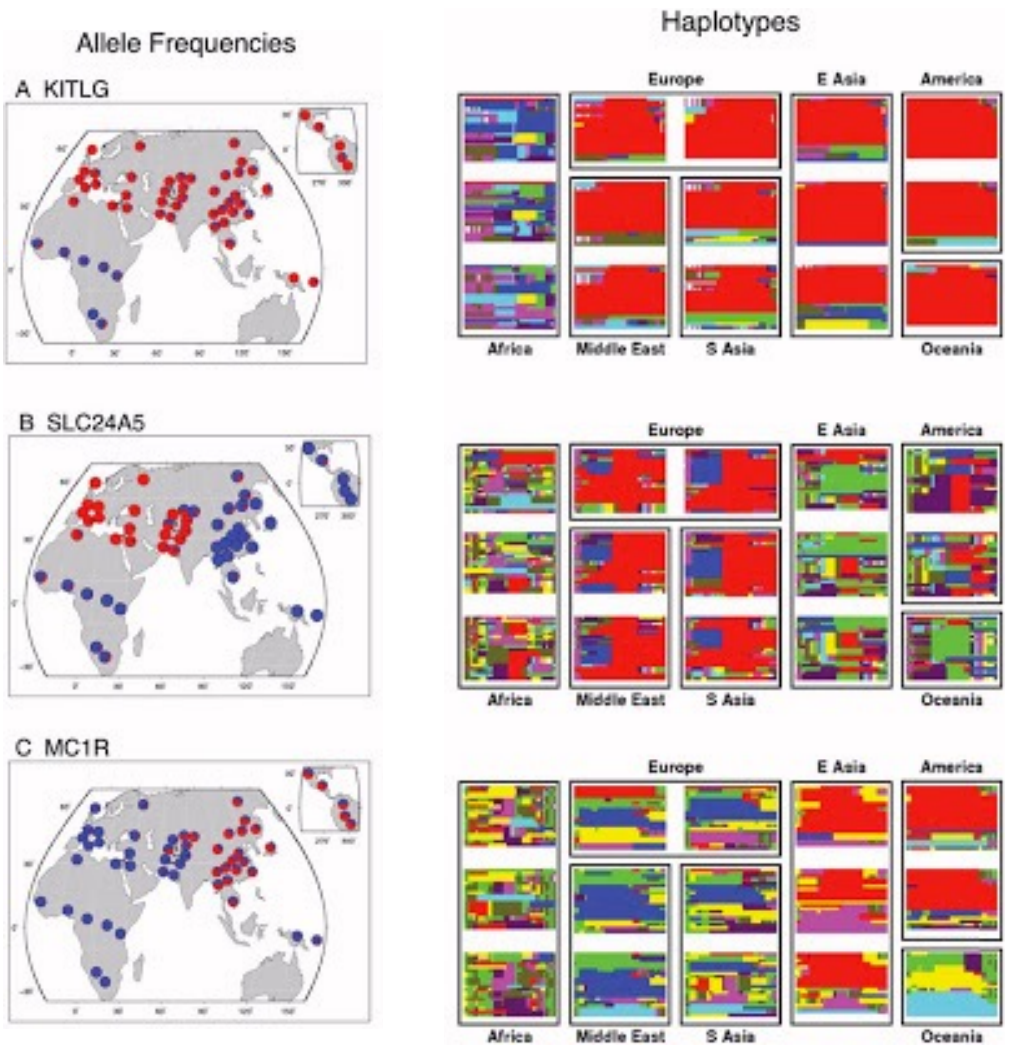
Voight et al., 2006

Haplotype-based signature of positive selection in Europeans at *SLC24A5*



Coop et al. 2009

Haplotype patters across pigmentation loci differ across populations



Simple hard sweep model is appealing

<u>Guiding assumption under a simple model of hard sweeps versus neutrality</u>: Most of the genome is assumed to be evolving neutrally, while only a subset of loci/variants are subject to positive selection

But reality is often more complex. Detecting sweeps can be difficult because:

- Confounding effects of background selection (negative selection)
- Selection from standing genetic variation
- Selection from multiple variants

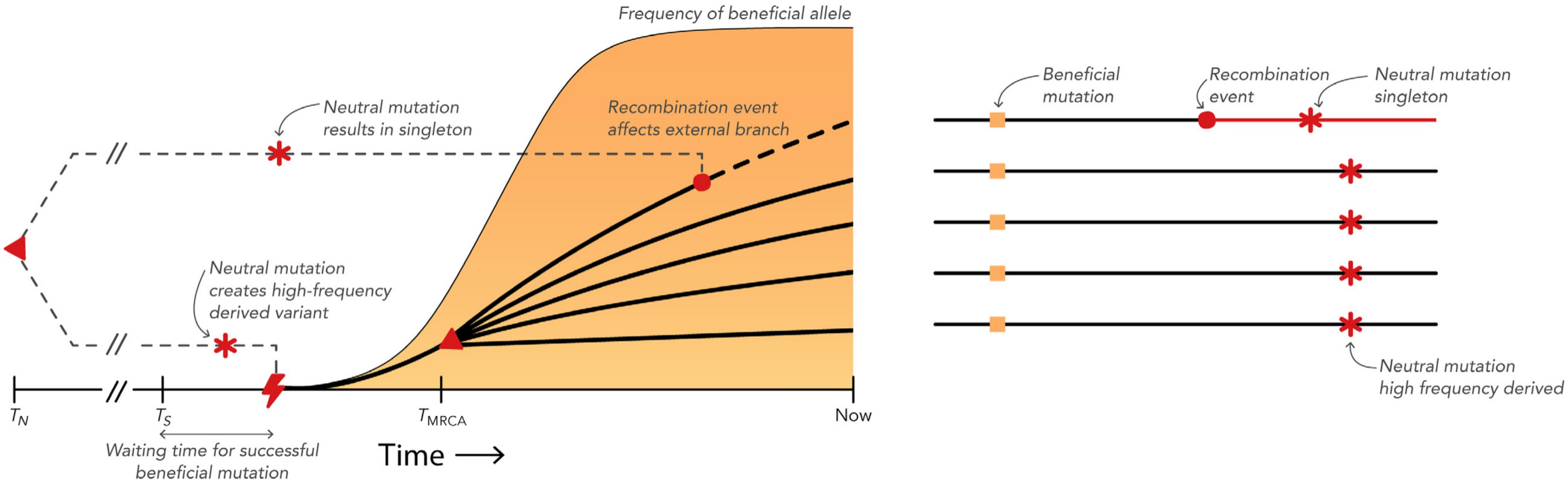
Recall: Classical theory of hard selective sweeps assumes:

Selection acts on a single copy of the beneficial allele, which enters the population as new mutation after the onset of the selection pressure

But: in some cases, the situation is more complex. Selection may act on an allele that was already present in the population (i.e., standing genetic variation) or on multiple alleles at a locus that have similar phenotypic effects.

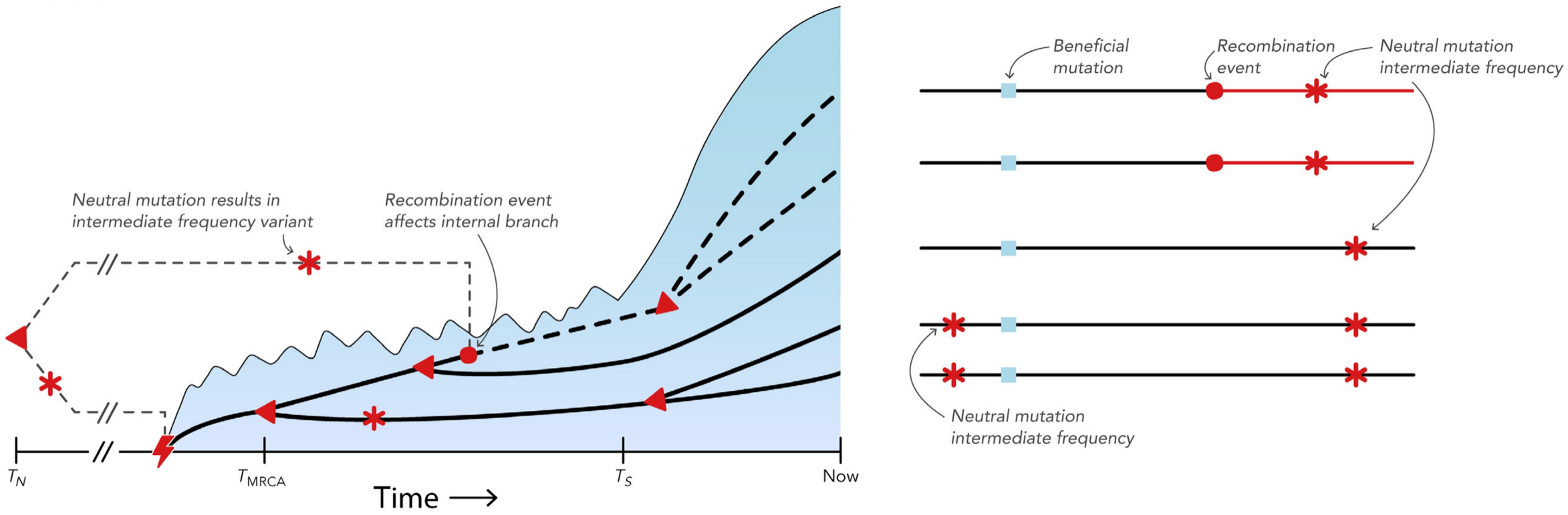
These situations are called ‘soft sweeps’

A hard selective sweep



Hermisson and Pennings, 2017

A single-origin soft selective sweep

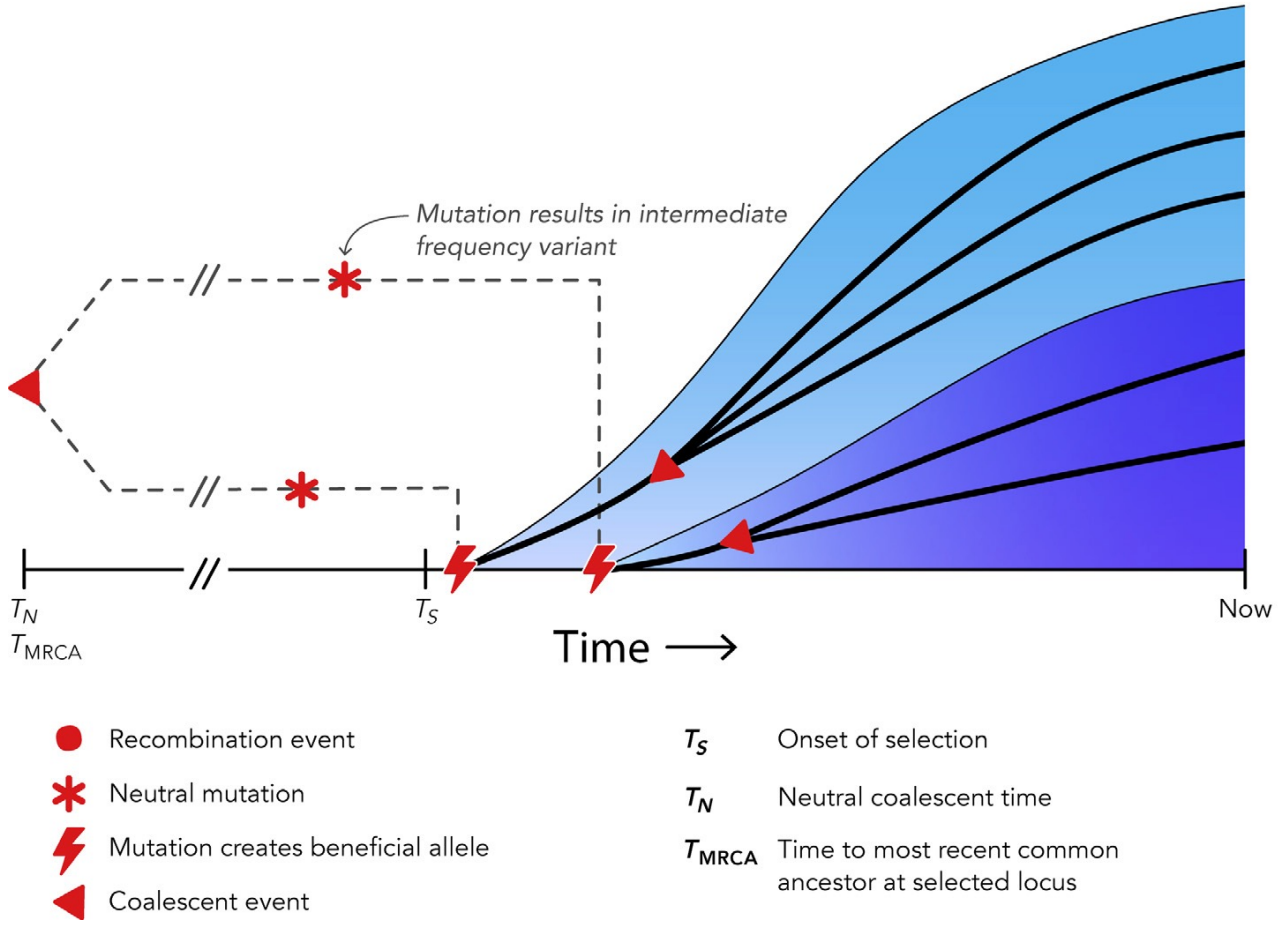


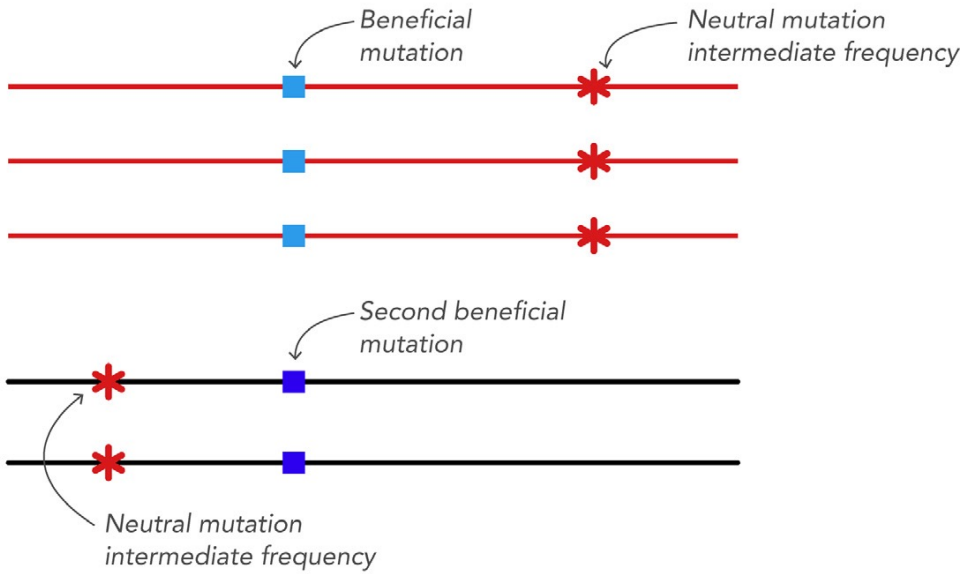
Neutral mutation becomes adaptive after already spending some time in the population

This scenario may be more likely if effects are conditional on environment or other loci

Hermisson and Pennings, 2017

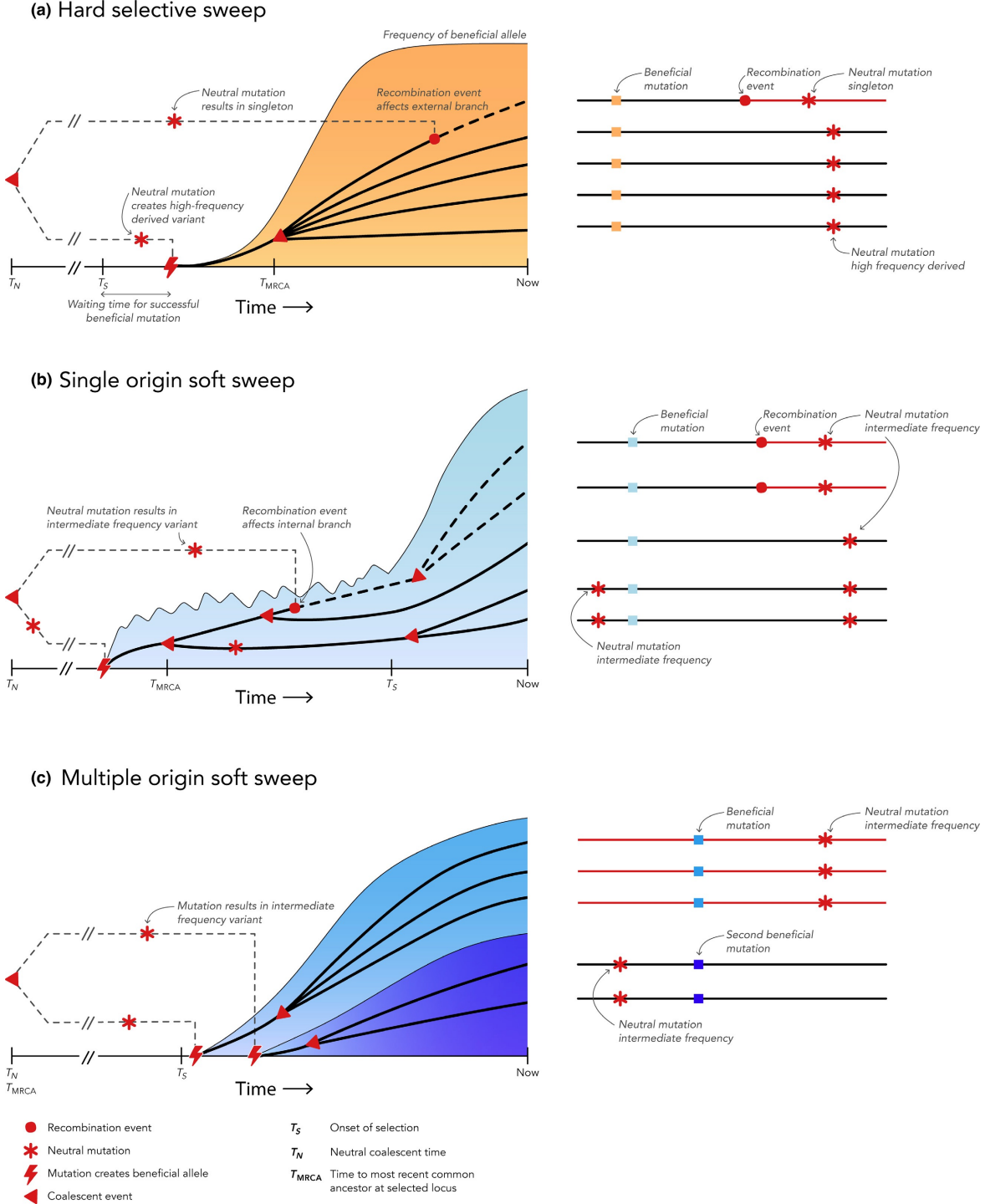
Multiple origin soft selective sweep





Multiple variants arise on different haplotypes and these rise to high frequency as a group

Hermisson and Pennings, 2017



Single adaptive variant arises at a locus and sweeps quickly to high frequency

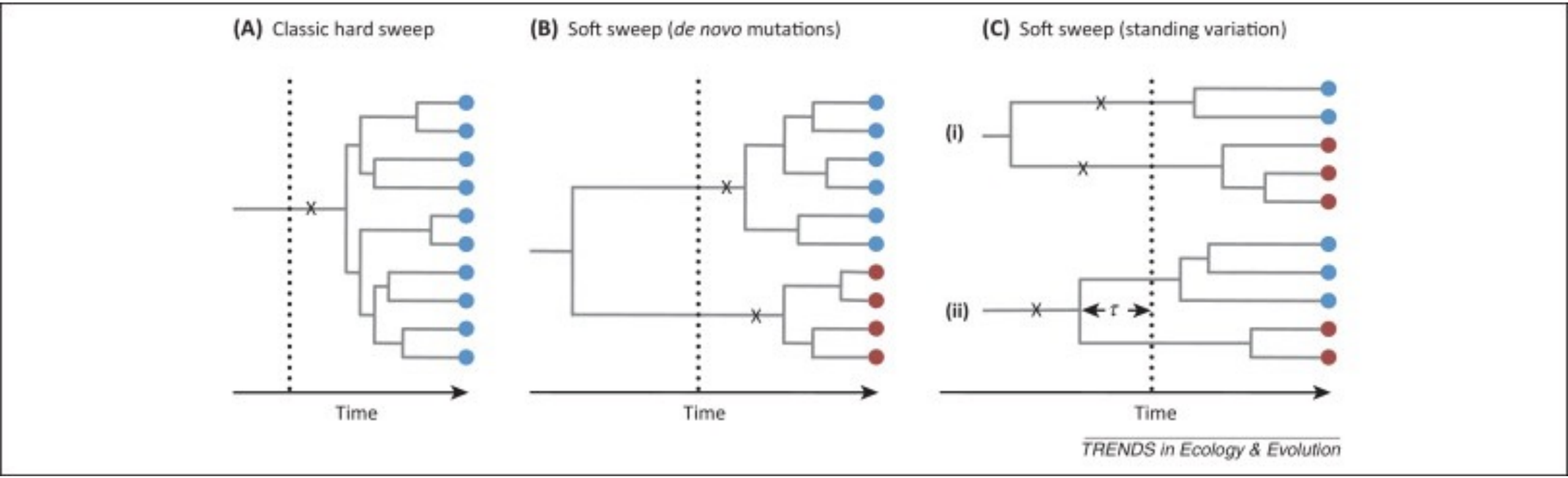
Neutral mutation becomes adaptive after spending some time in the population

Multiple adaptive mutations arise over a short time frame and as a group sweep to high frequency in the population

Hermisson and Pennings, 2017

Sweep patterns

standard, multiple variants, standing variation



Messer and Petrov TREE 2013

Controversy around soft sweeps

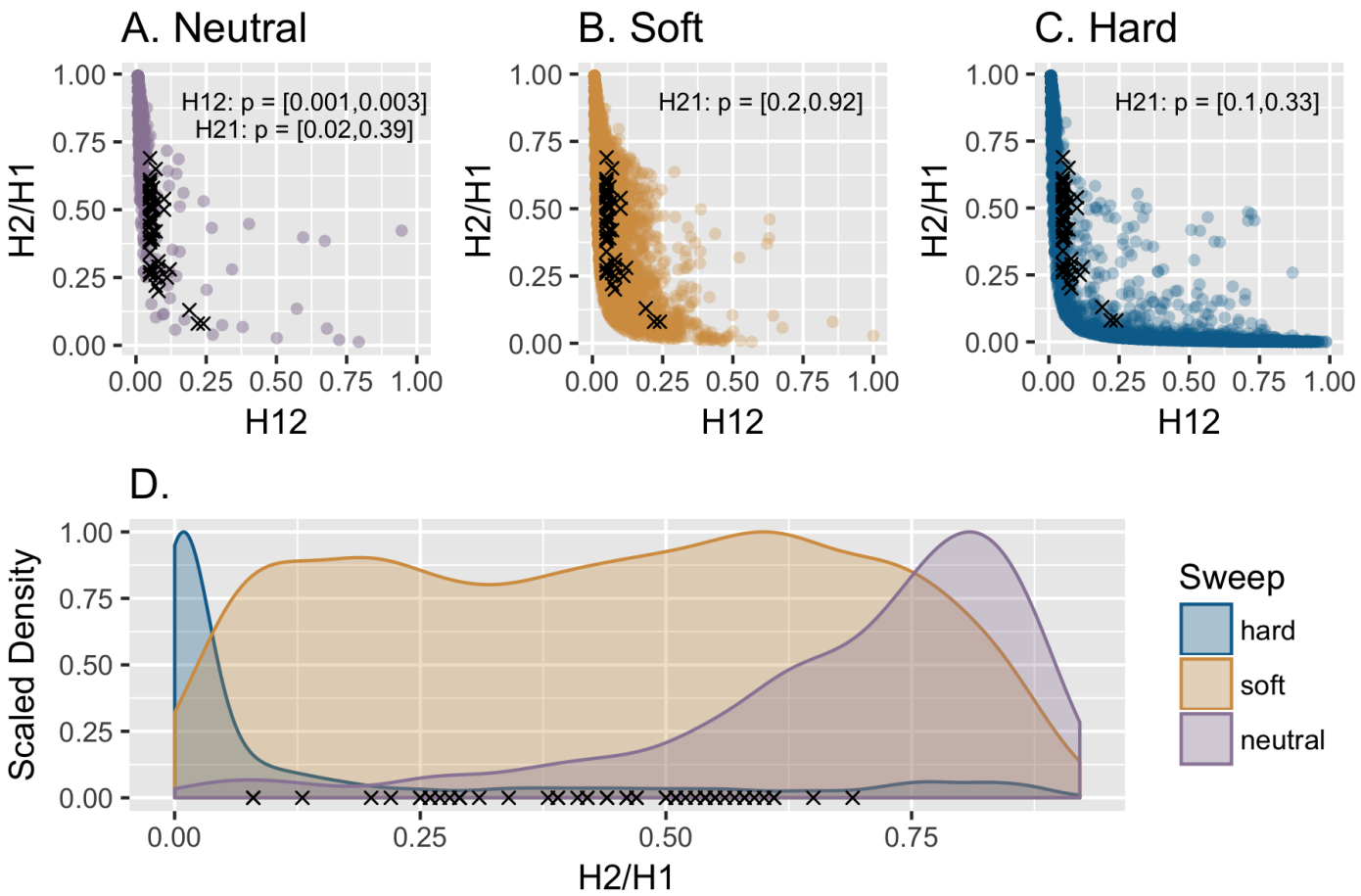
- Soft sweeps from standing variation require a major change in the functional impact of a variant. This might be possible due to cryptic variation that is exposed due to interaction effects like genotype by environment or genotype by genotype (epistatic) interactions
- But we need examples where functional variants are known
- Soft sweeps from standing variation require that multiple variants arise in a short period of time and sweep. This could happen when common assumptions of random mating are not met (e.g., population structure)

Can soft sweeps be detected?

There are methods to detect soft sweeps, but power using selection scans is reduced relative to hard sweeps

There is a statistical test from Garud et al., that uses information about the two predominant haplotypes and can be used to scan for a soft sweep signature.

But given the weaker signature left by soft sweeps, the power is low compared to tests for hard selective sweeps. Soft sweep regions tend to look more like neutral regions.



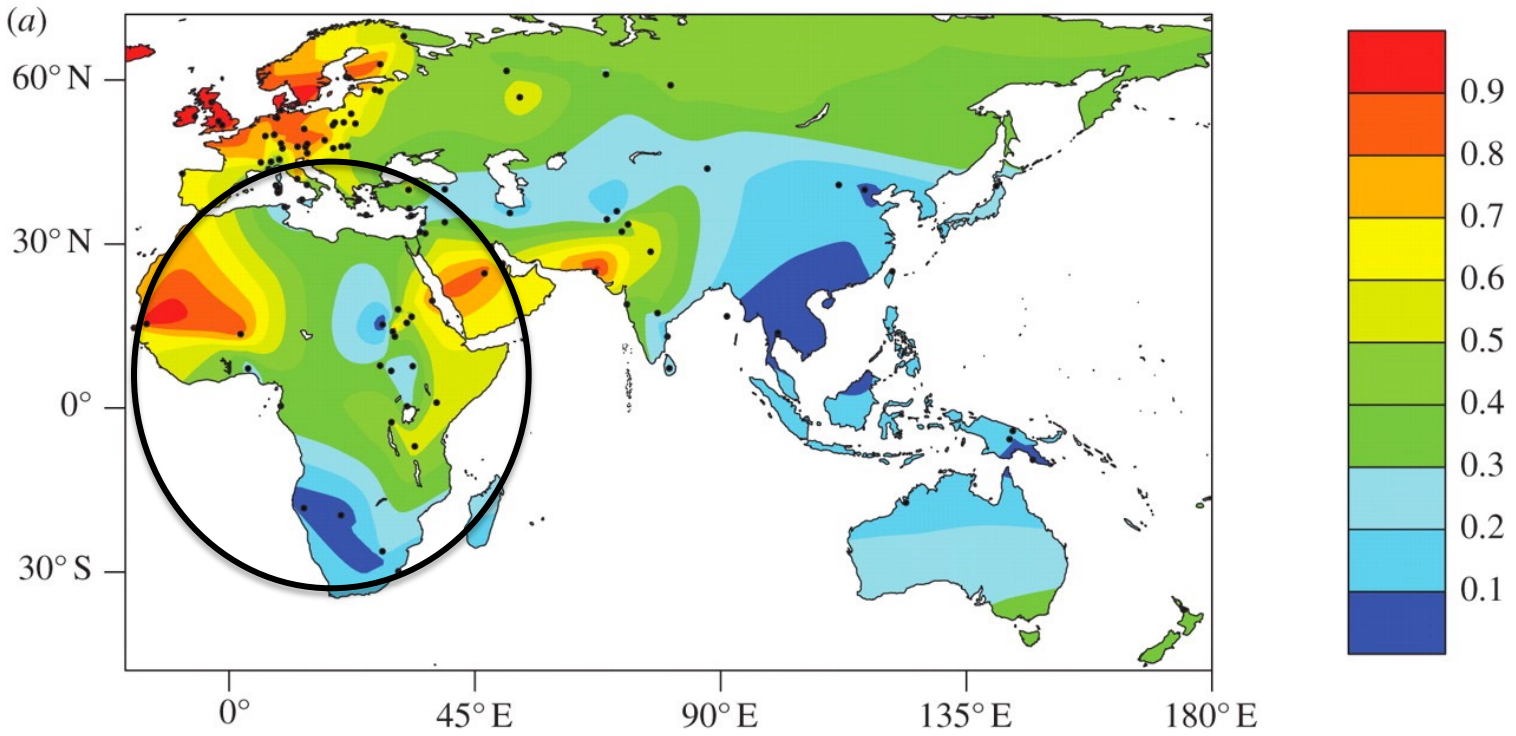
Harris et al., 2018; https://journals.plos.org/plosgenetics/article?id=10.1371/journal.pgen.1007859

What is the impact of soft sweeps
on the expected footprint of selection?

Similar to hard sweep but signal tends to be muted

Some examples of soft sweeps

Worldwide distribution of lactase persistence



High persistence

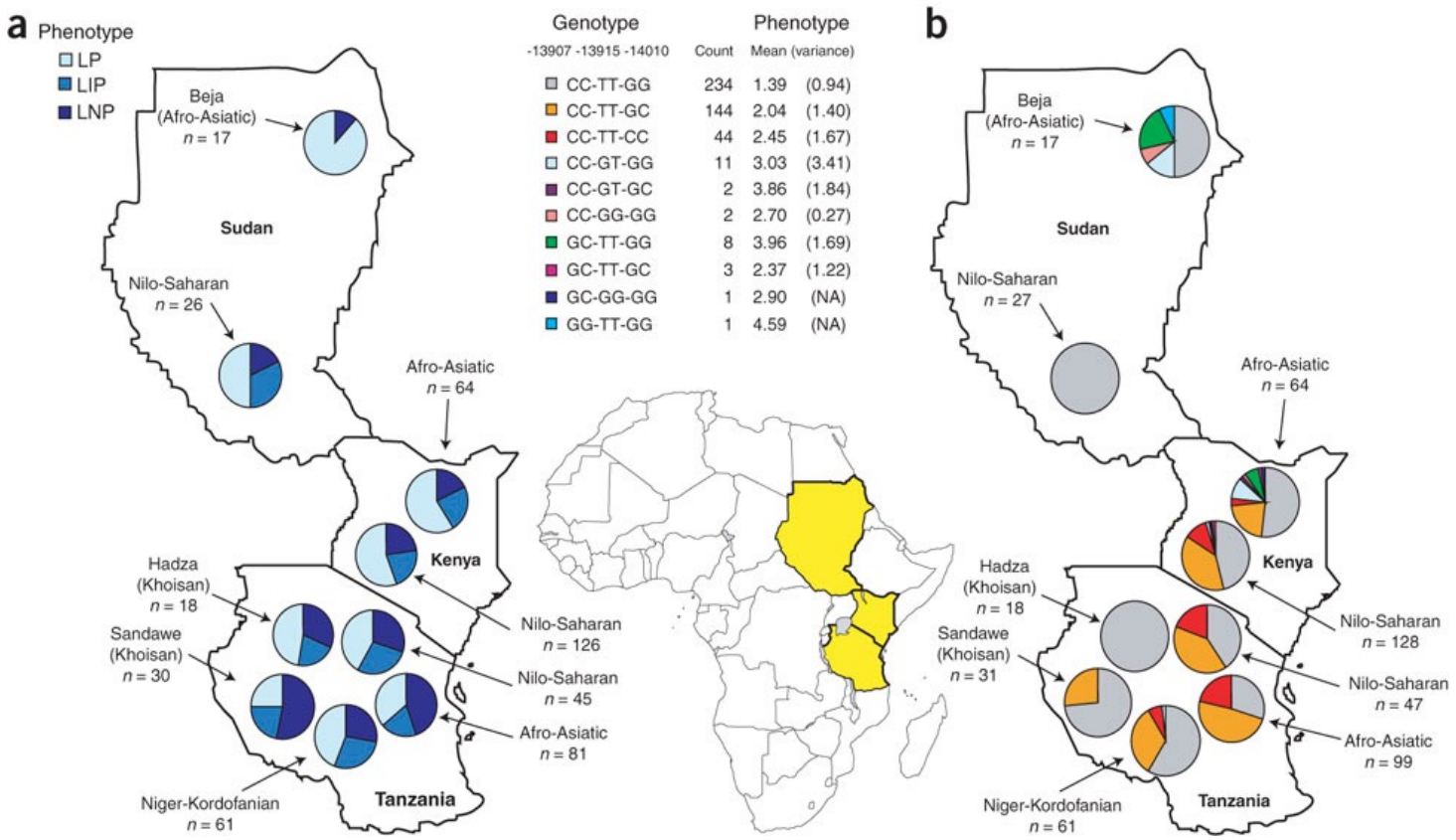
Low persistence

Variants other than *LCT -13910*T* are responsible for lactase persistence in African populations.

Lactase persistence in Africa is due to soft sweeps – multiple causal haplotypes

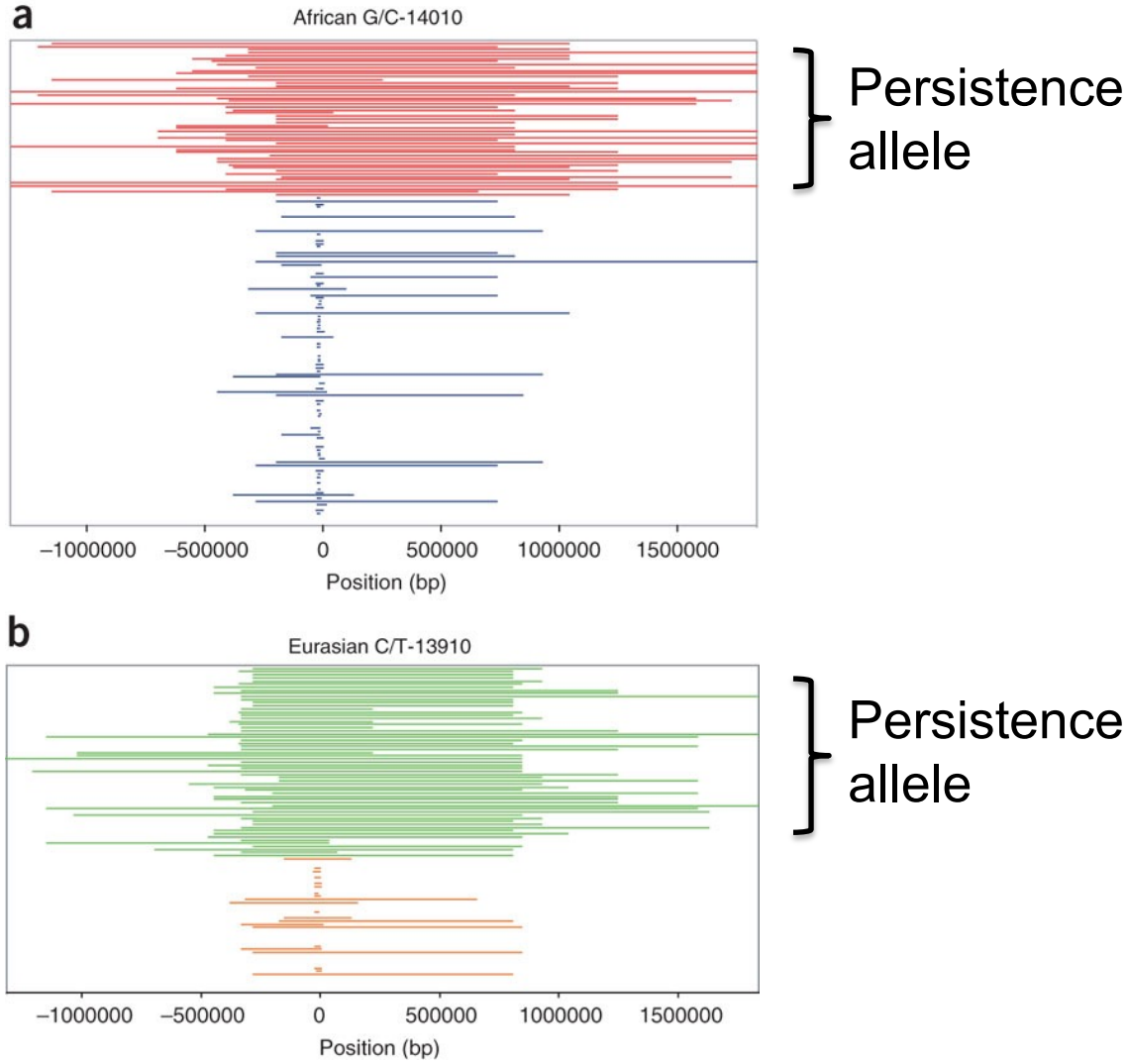
Proportion of population with LP

Proportion compound genotypes



Reed et al., 2007

Haplotype structure is stronger for the persistence alleles

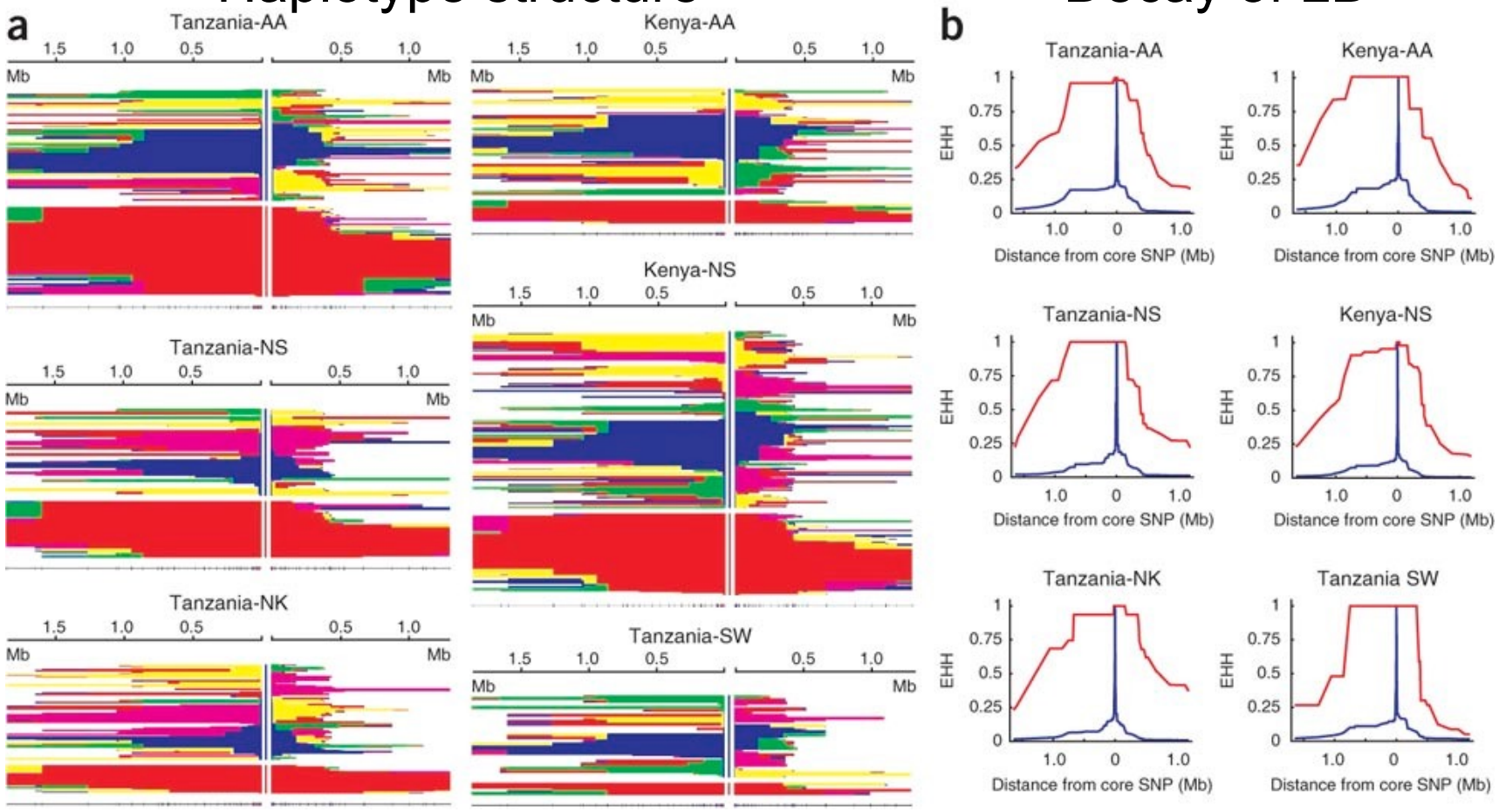


Reed et al., 2007

LD is more extensive for derived (persistence) alleles

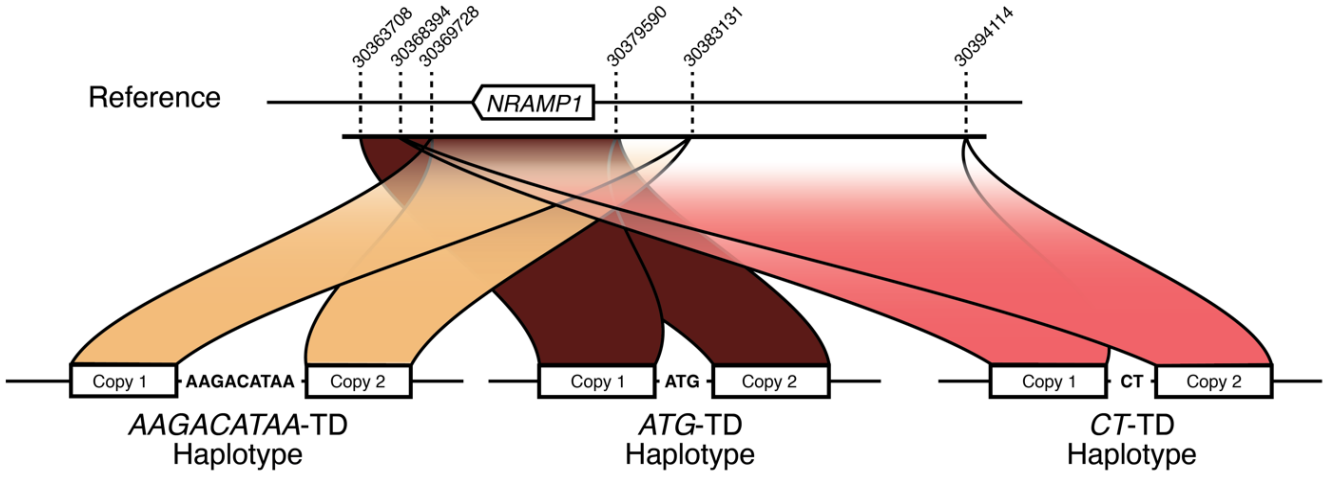
Haplotype structure

Decay of LD

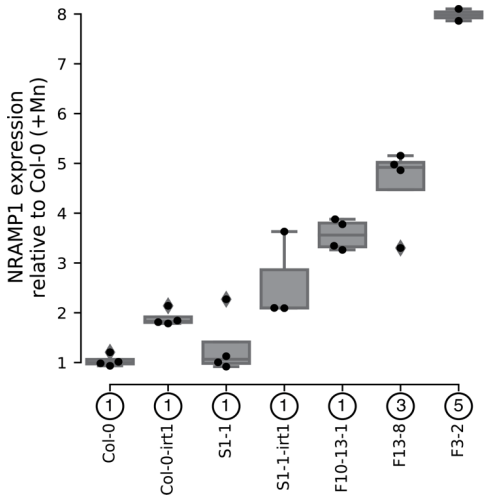


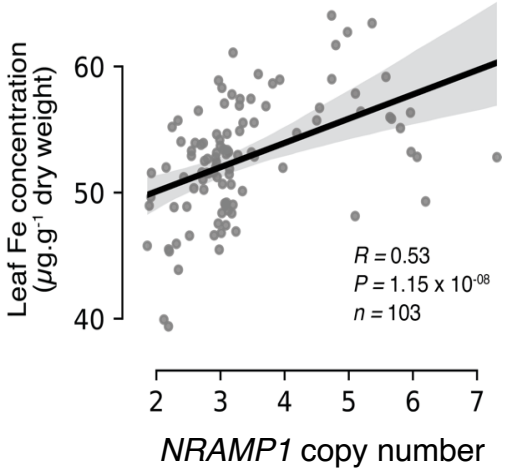
Multiple tandem duplications arose and swept to increase mineral nutrient transport

Variation in leaf elemental content results in variation in plant health



Multiple tandem duplications arose and swept to near fixation (98%) in a Cape Verdean volcanic island.





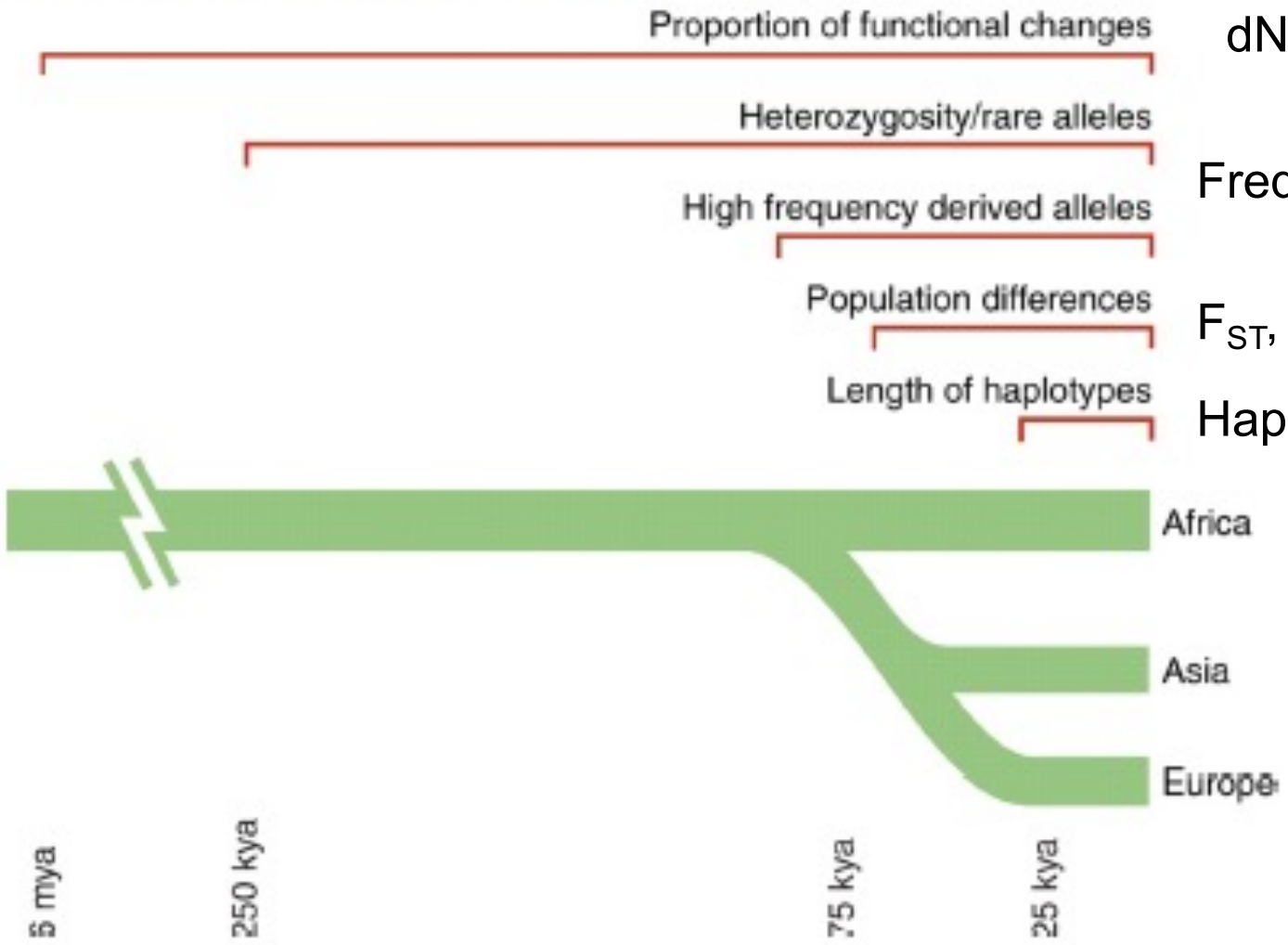
These increase copies of the *NRAMP1* transporter as well as expression (mRNA) and Fe transport

Tergemina et al., 2023

Summary: types of neutrality tests

- Divergence-based
  - dN/dS and MK-type tests
- Differentiation
  - F_{ST}, environmental associations)
- Inter-locus comparison of divergence relative to diversity
  - HKA test
- SFS based tests
  - Tajima's D, Fay and Wu's H test
- Haplotype / LD based tests
  - EHH and variants, number of haplotypes, frequency of major haplotype
- Haplotype differentiation tests use signatures of different haplotype homozygosity across populations
  - XP-EHH

Time scale for signatures of selection



dN/dS

Frequency spectrum-based tests

F_{ST}, environmental correlations

Haplotype-based tests

From Rogers lecture notes

Lack of strong concordance between selection scans across the human genome

	(SFS + LD)	EHH	Sweepfinder Tajima	EHH	MK test	MK test
	Williamson *et al.*[5]	Voight *et al.*[40]	Carlson *et al.*[45]	Wang *et al.*[6]	Bustamante *et al.*[3] (PS $p < 0.025$)	Busta
Williamson *et al.*[5]*	179	12	20	0	0	4
Voight *et al.*[40]*	13	713	6	7	22	32
carlson *et al.*[45]*	23	7	59	5	3	10
Wang *et al.*[6]*	0	7	3	90	3	1
Bustamante *et al.*[3] (PS $p < 0.025$)‡	0	22	3	3	301	#
Bustamante *et al.*[3] (NS $p > 0.975$)‡	3	30	10	2	#	802

… no test uses all patterns: different tests pick up different signals

False positives and negatives are also expected to contribute to disparity

Summary: neutrality tests

Which test to use? – Power of tests for selection

- depends on time / frequency of adaptive fixations
  - dN/dS and MK type tests need many substitutions over long time
  - Tajima‘s test, tests based on singletons: up to about 0.1 N generations
  - haplotype and LD based tests: up to about 0.01 N generations
  - EHH and similar tests: incomplete sweeps
- depends on the underlying demography
  - worst-case scenario: bottlenecks (no problem for divergence tests)
- depends on the selection scenario
  - Soft sweeps, polygenic adaptation, local adaptation, partial sweeps