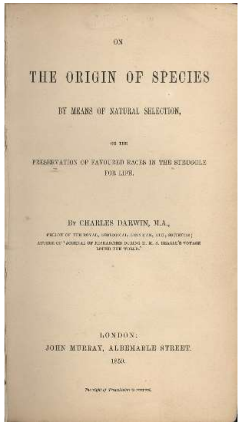
INSTITUT de BIOLOGIA EVOLUTIVA ibe CSIC upf.

instituto de biociências USP Universidade de São Paulo

NATURAL SELECTION AND ADAPTATION

Tábita Hünemeier
tabita.hunemeier@ibe.upf-csic.es
hunemeier@usp.br

Natural Selection



ON

THE ORIGIN OF SPECIES

BY MEANS OF NATURAL SELECTION,

OR THE

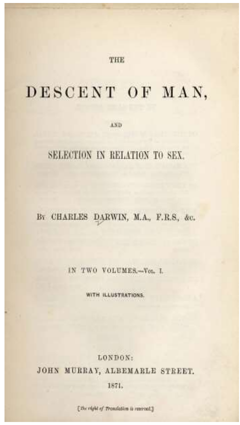
PRESERVATION OF FAVOURED RACES IN THE STRUGGLE FOR LIFE.

BY CHARLES DARWIN, M.A.,

FELLOW OF THE ROYAL, GEOLOGICAL, LINNÆAN, ETC., SOCIETIES;
AUTHOR OF 'JOURNAL OF RESEARCHES DURING H. M. S. BEAGLE'S VOYAGE ROUND THE WORLD.'

LONDON:
JOHN MURRAY, ALBEMARLE STREET.
1859.

The right of Translation is reserved.



THE

DESCENT OF MAN,

AND

SELECTION IN RELATION TO SEX.

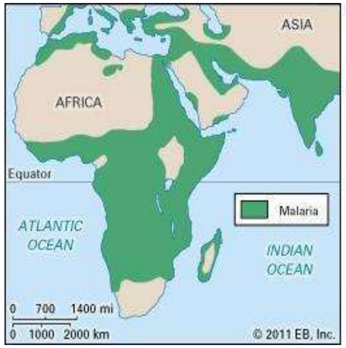
BY CHARLES DARWIN, M.A., F.R.S., &c.

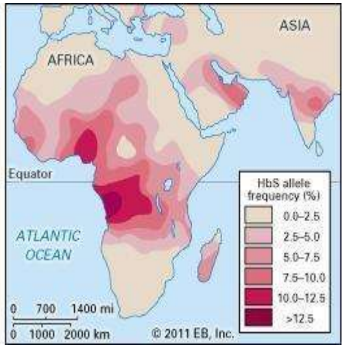
IN TWO VOLUMES.—VOL. I.

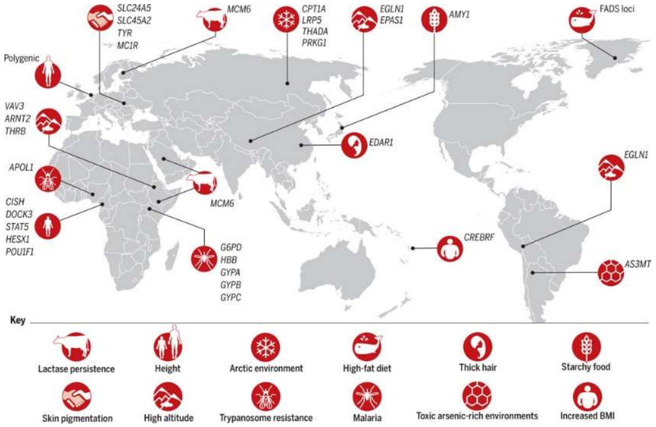
WITH ILLUSTRATIONS.

LONDON:
JOHN MURRAY, ALBEMARLE STREET.
1871.

[The right of Translation is reserved.]







1859

1871

1949

2000

Charles Darwin

i) populations have a phenotypic variation.

ii) the environment presents challenges.

iii) those better able to cope tend to leave more offspring.

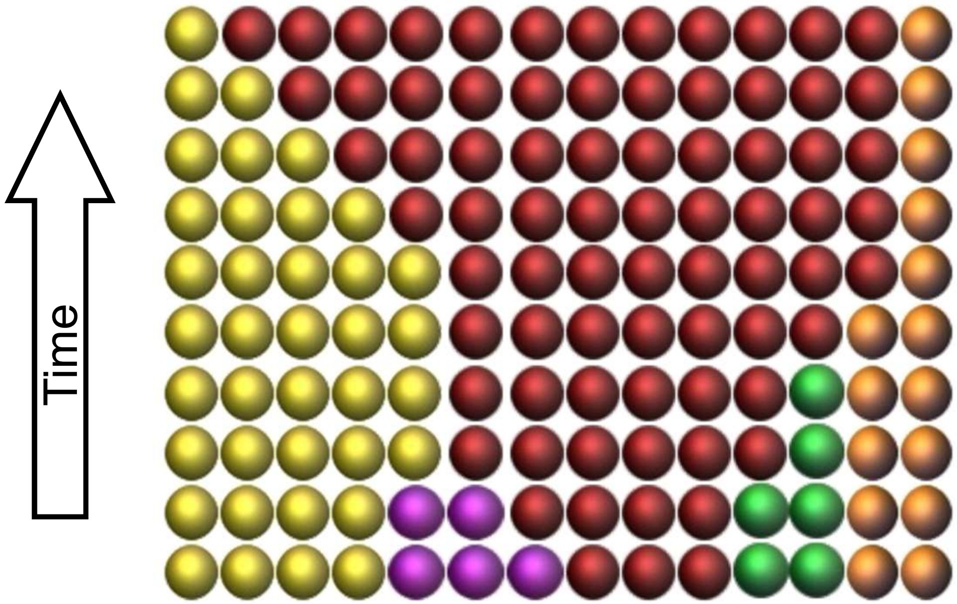
iv) individuals tend to produce more offspring than the environment can support.

Ernst Mayr

i) Variation.

ii) Variation must contribute to survival and differential reproduction.

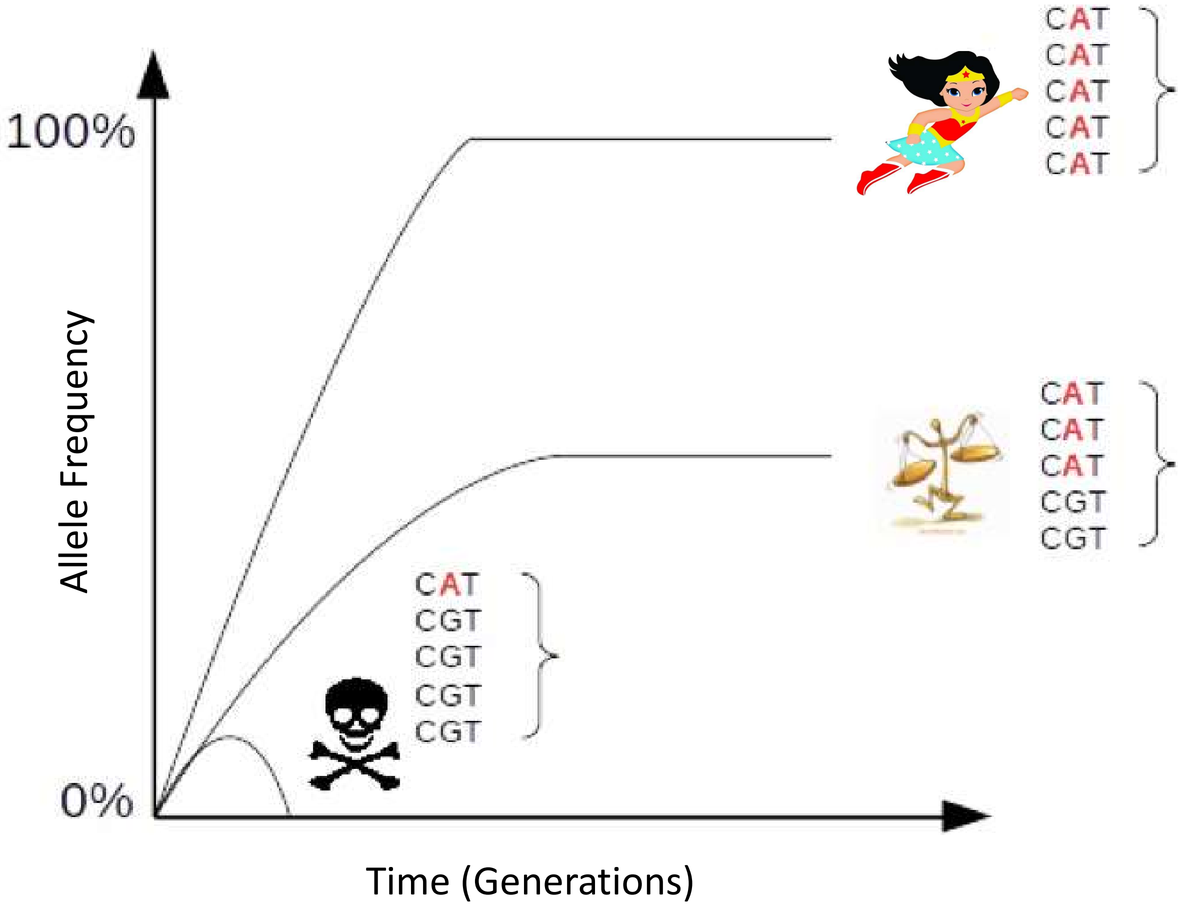
iii) Variation must be inherited.



Natural Selection

- Heritable traits that increase fitness become more common.
- Sites targeted by natural selection are likely to harbor functionality.
- Mutations arise (almost) randomly and evolve according to their effect on the carrier’s fitness.

Natural Selection



Allele Frequency Trajectories

Effect of Selection on Alleles:

- Neutral/Weak: removed, polymorphic or fixed;
- Strong Negative: Removed or polymorphic;
- Strong Positive: removed, polymorphic or fixed.
- Balancing: removed, polymorphic or fixed.

Allele Frequency Trajectories

Effect of Selection on Alleles:

- Neutral/Weak: removed, polymorphic or fixed;
- Strong Negative: Removed or polymorphic;
- Strong Positive: removed, polymorphic or fixed.
- Balancing: removed, polymorphic or fixed.

What is strong? It depends on the effective population size.

Allele Frequency is (frequently) not enough to determine selection.

Allele Frequency Trajectories

If the simple observation of allele frequencies is not enough, what else can we do to detect signals of natural selection?

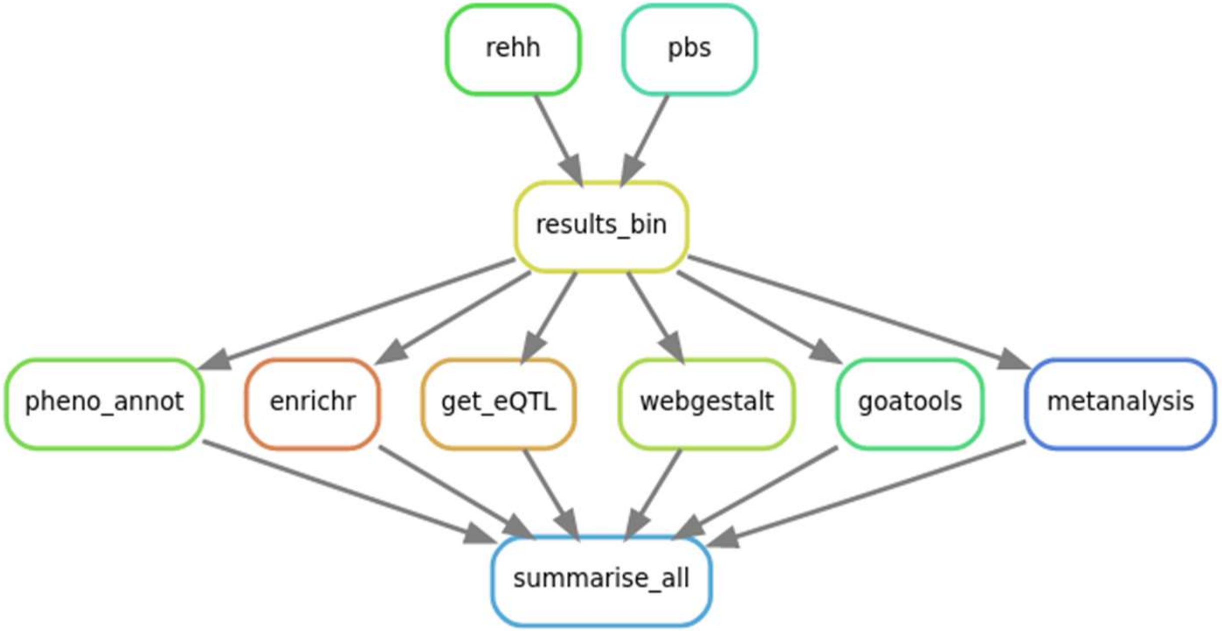
Allele Frequency Trajectories

If the simple observation of allele frequencies is not enough, what else can we do to detect signals of natural selection?

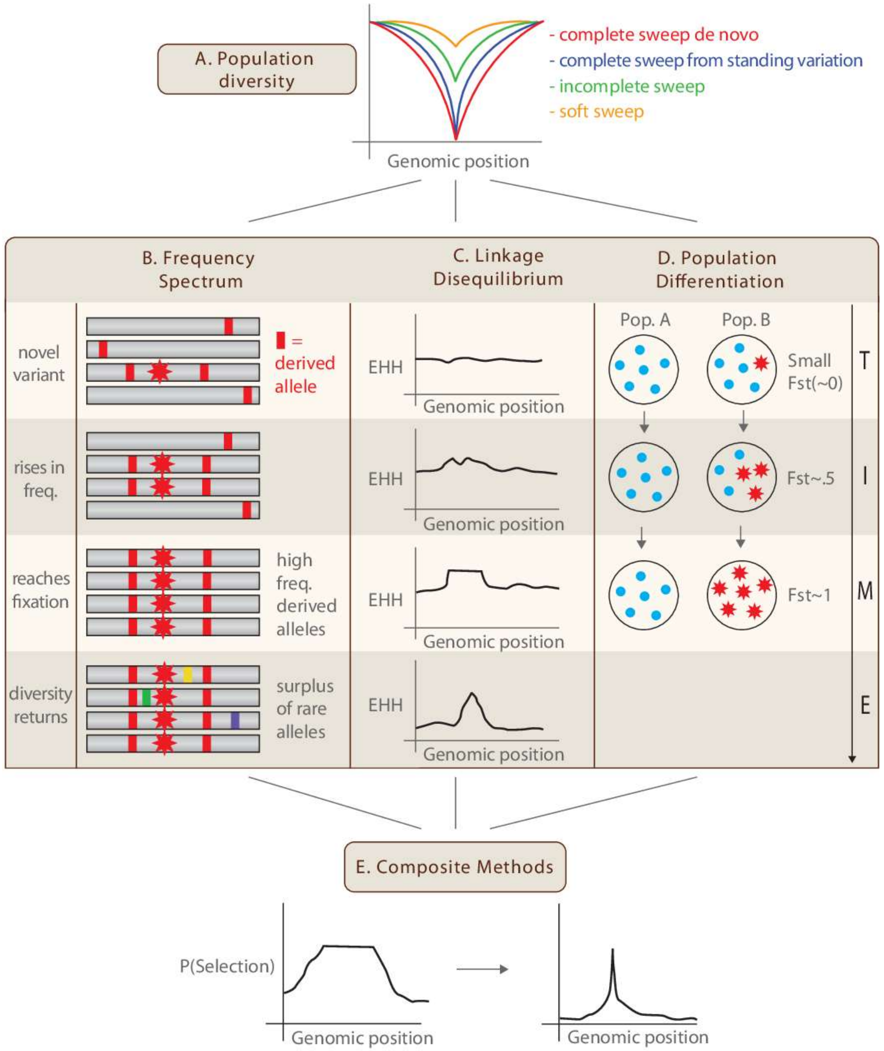
- perform selection experiments;
- use external information: candidate genes/biological knowledge, functional categories, association to phenotypes;
- use information from the surrounding genomic region;
- use information from multiple species/populations;

Allele Frequency Trajectories

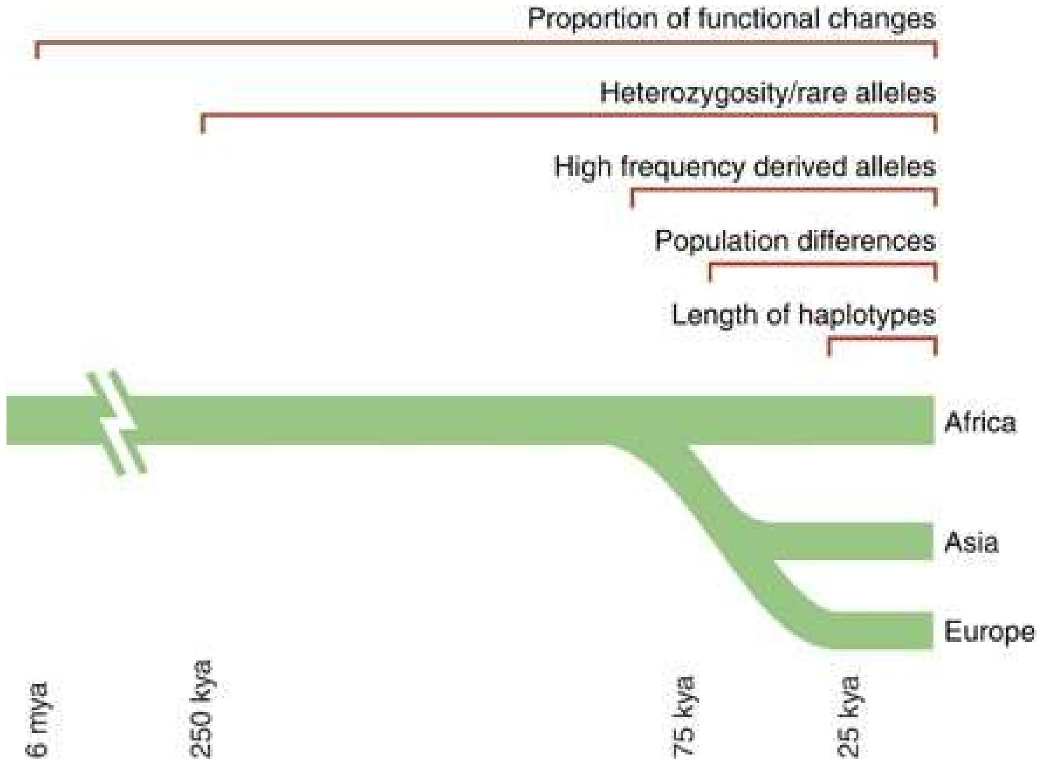
If the simple observation of allele frequencies is not enough, what else can we do to detect signals of natural selection?



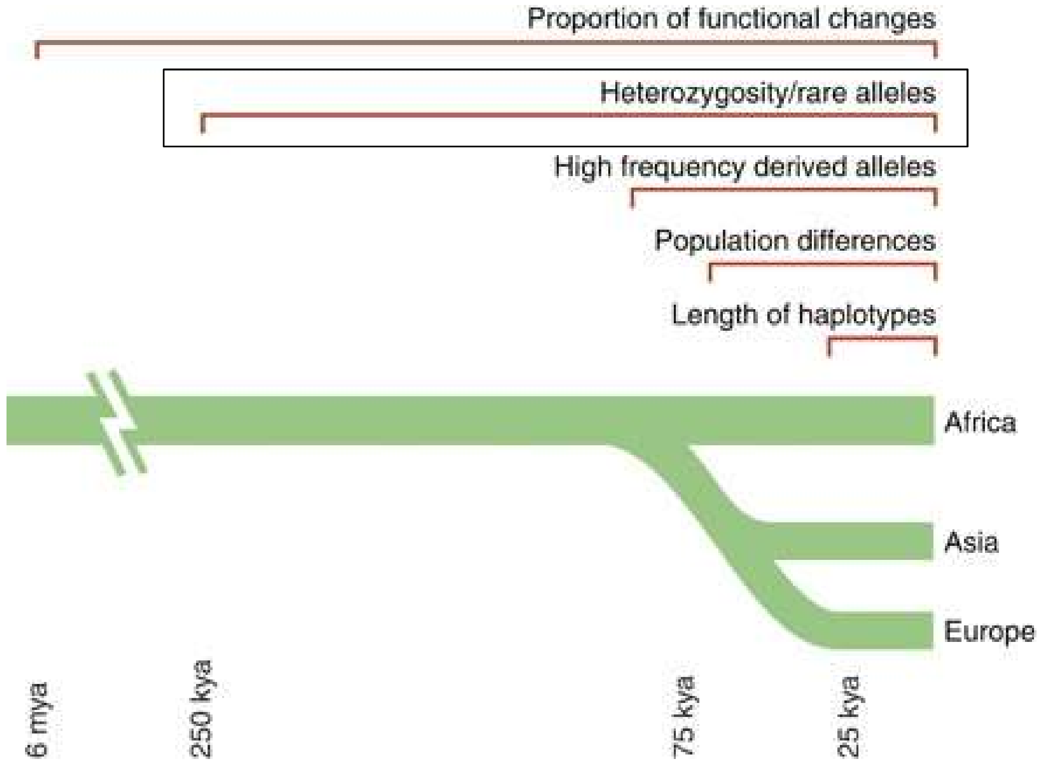
Common Methods to Detect Selection



Common Methods to Detect Selection


Proportion of functional changes
Heterozygosity/rare alleles
High frequency derived alleles
Population differences
Length of haplotypes
Africa
Asia
Europe
6 mya
250 kya
75 kya
25 kya

Common Methods to Detect Selection



Proportion of functional changes
Heterozygosity/rare alleles
High frequency derived alleles
Population differences
Length of haplotypes
Africa
Asia
Europe
6 mya
250 kya
75 kya
25 kya

Tajima's D

ACTAGAGGAT
ACTTGAGGAT
ACTTGACGAT
ACTTGACGAT
ACTTGAGGTT

$$\hat{\theta}_T = \frac{\sum_{i<j} d_{ij}}{n(n-1)/2}$$

$$\hat{\theta}_W = \frac{S}{\sum_{i=1}^{n-1} 1/i}$$

$$D = \frac{\hat{\theta}_T - \hat{\theta}_W}{\sqrt{\hat{V}(\hat{\theta}_T - \hat{\theta}_W)}}$$

Tajima's D

ACTAGAGGAT
ACTTGAGGAT
ACTTGACGAT
ACTTGACGAT
ACTTGAGGTT

$$\hat{\theta}_T = \frac{\sum_{i<j} d_{ij}}{n(n-1)/2}$$

$$\hat{\theta}_W = \frac{S}{\sum_{i=1}^{n-1} 1/i}$$

$$D = \frac{\hat{\theta}_T - \hat{\theta}_W}{\sqrt{\hat{V}(\hat{\theta}_T - \hat{\theta}_W)}}$$

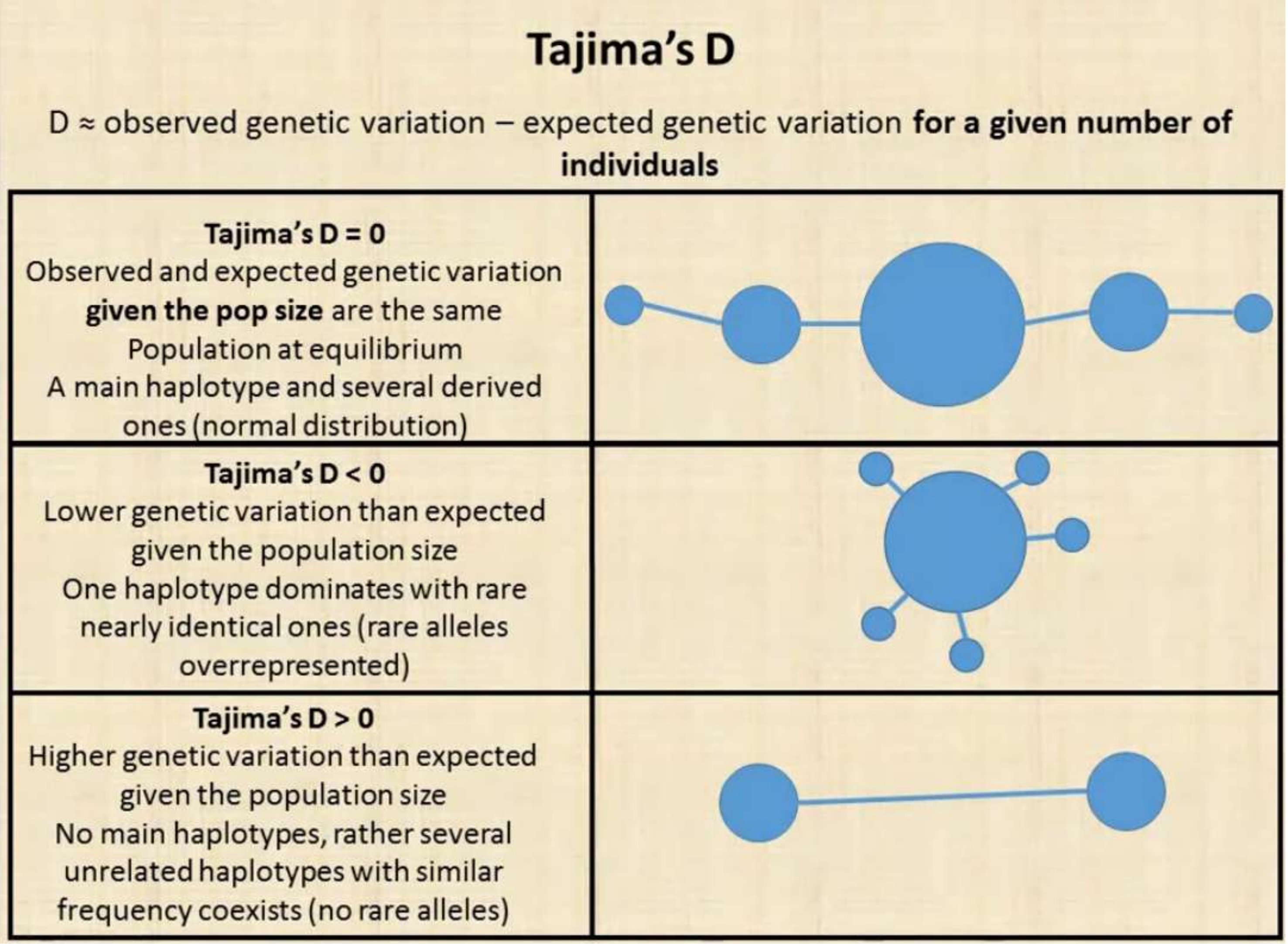
Θ_T = π = average number of differences between pairs of sequences

Θ_T = 1+2+2+2+1+1+1+0+2/10 = 1.2

Θ_W = π expected number under neutrality

Θ_W = 4/(1/1+1/2+1/3+1/4)=1.92

Tajima's D

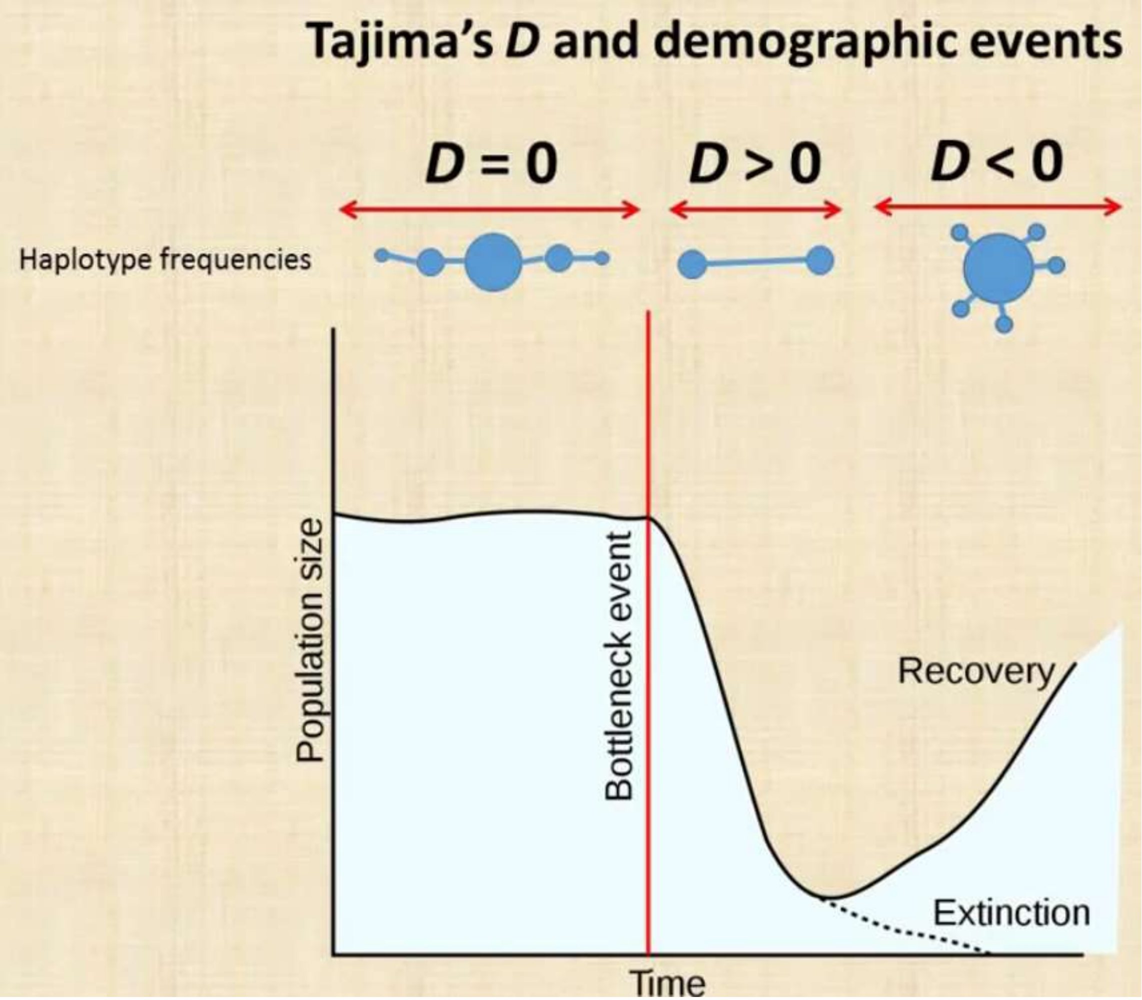


Tajima's D

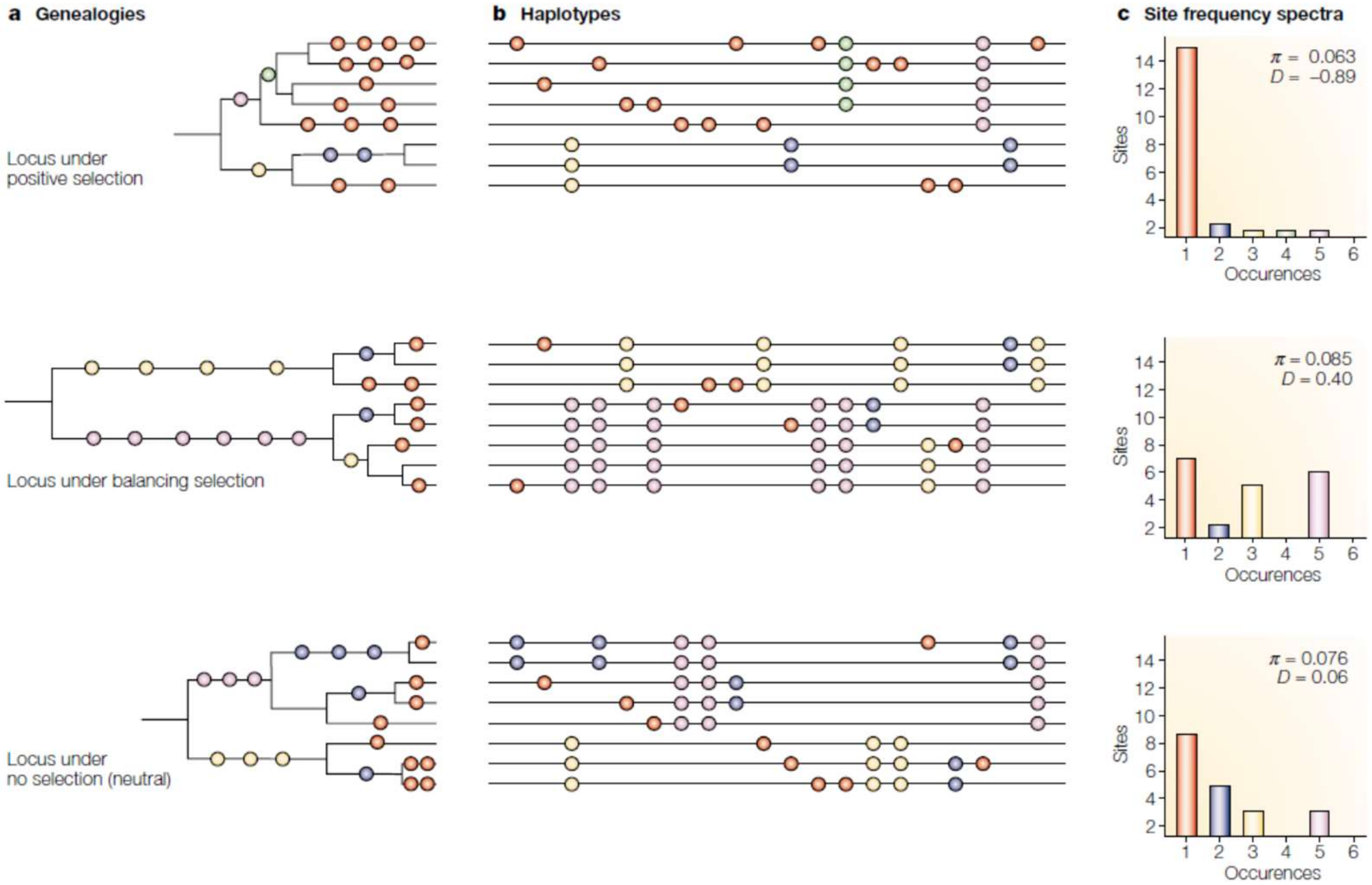
D ≈ observed genetic variation – expected genetic variation **for a given number of individuals**

Tajima's D = 0 Observed and expected genetic variation **given the pop size** are the same Population at equilibrium A main haplotype and several derived ones (normal distribution)	
Tajima's D < 0 Lower genetic variation than expected given the population size One haplotype dominates with rare nearly identical ones (rare alleles overrepresented)	
Tajima's D > 0 Higher genetic variation than expected given the population size No main haplotypes, rather several unrelated haplotypes with similar frequency coexists (no rare alleles)	

Tajima's D


Tajima's *D* and demographic events
D = 0
D > 0
D < 0
Haplotype frequencies
Population size
Bottleneck event
Recovery
Extinction
Time

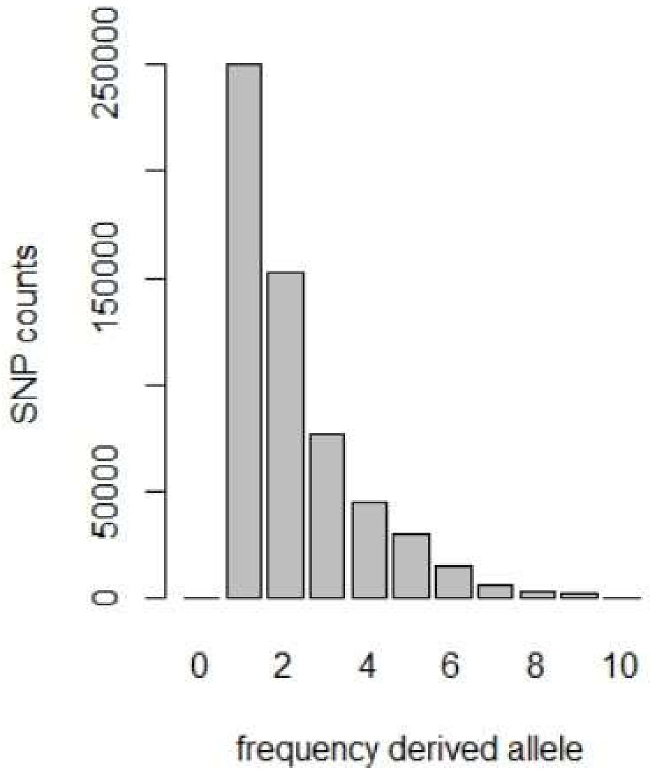
SFS


a Genealogies
b Haplotypes
c Site frequency spectra
Locus under positive selection
π = 0.063
D = –0.89
Locus under balancing selection
π = 0.085
D = 0.40
Locus under no selection (neutral)
π = 0.076
D = 0.06
Sites
14
12
10
8
6
4
2
1
2
3
4
5
6
Occurences

Site Frequency Spectrum (SFS)

Even if we have millions of SNPs we can summarize the genomic data to 10 numbers with the SFS!

The size of the SFS depends on the number of sampled individuals.



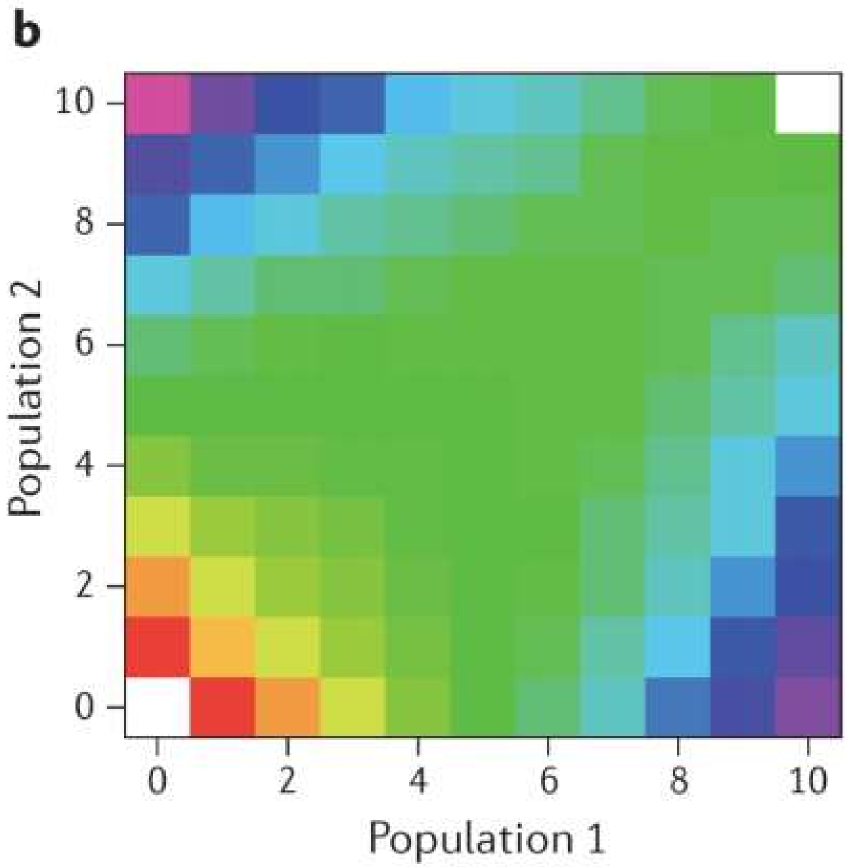
Observed SFS is a vector (1 dimensional SFS):

Frequency	0	1	2	3	4	5	6	7	8	9	10
SNP count	0	250,032	152,300	76,504	45,362	30,210	15,329	5,642	3,524	2,123	0

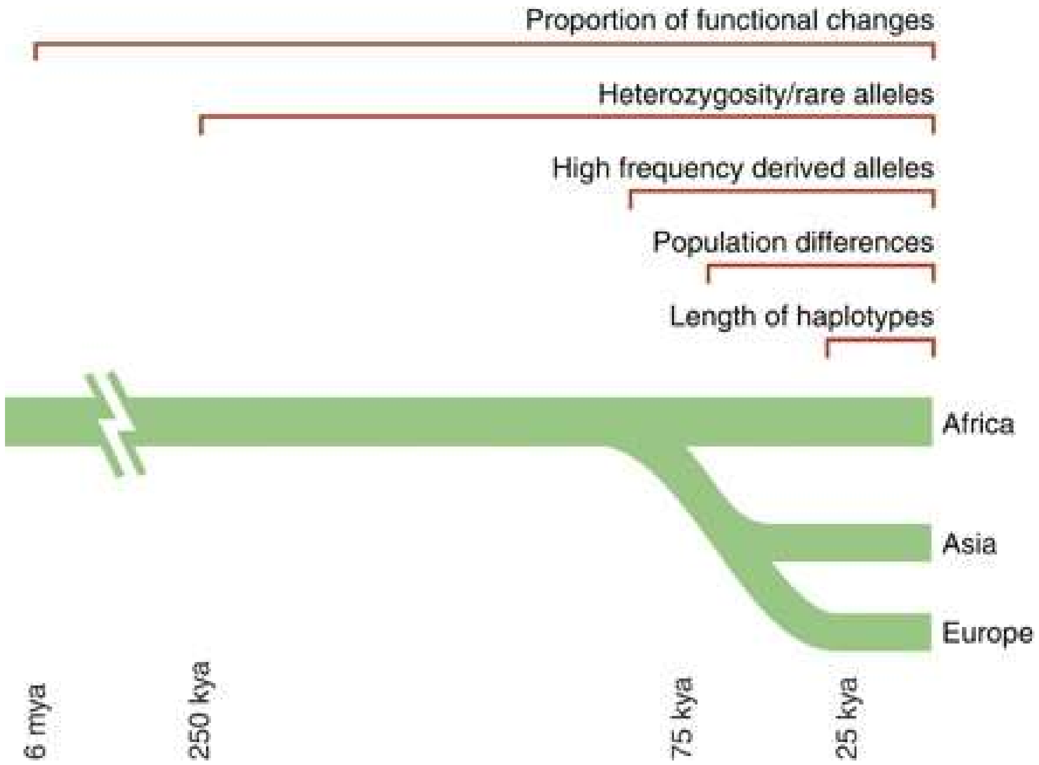
SFS

- For a pair of populations – 2D SFS
  - Count the SNPs have a frequency of the derived allele of *i* in population 1, and of *j* in population 2

- We can extend this to 3D SFS, 4D SFS, etc.

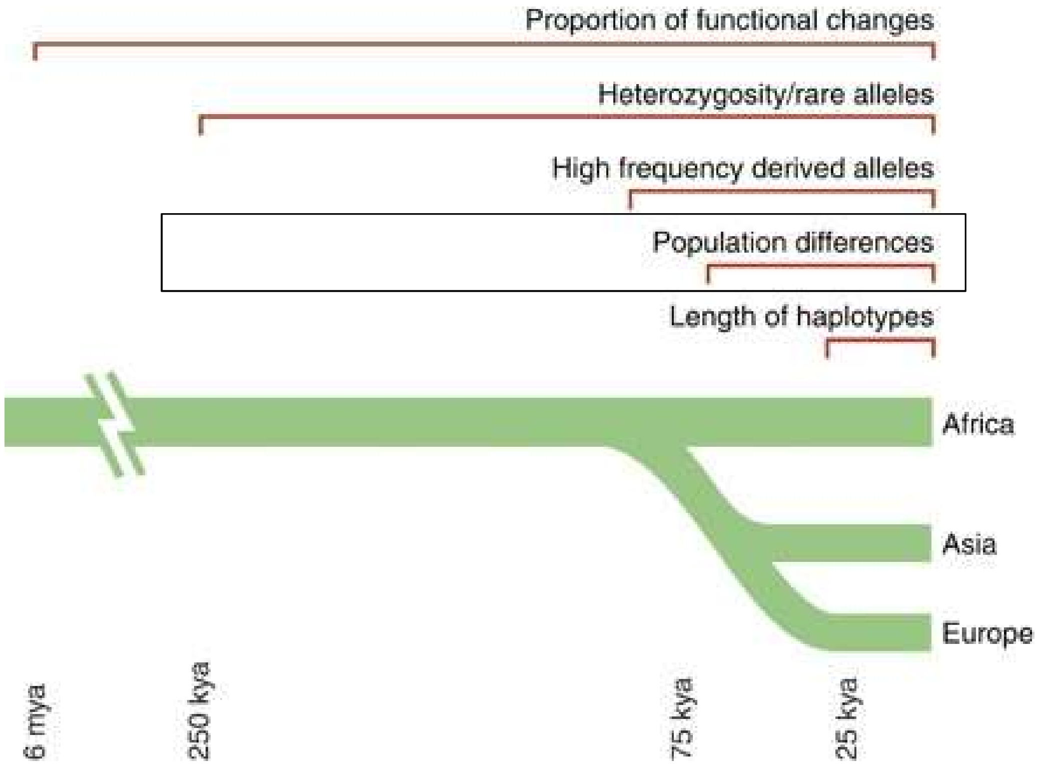


Common Methods to Detect Selection


Proportion of functional changes
Heterozygosity/rare alleles
High frequency derived alleles
Population differences
Length of haplotypes
Africa
Asia
Europe
6 mya
250 kya
75 kya
25 kya

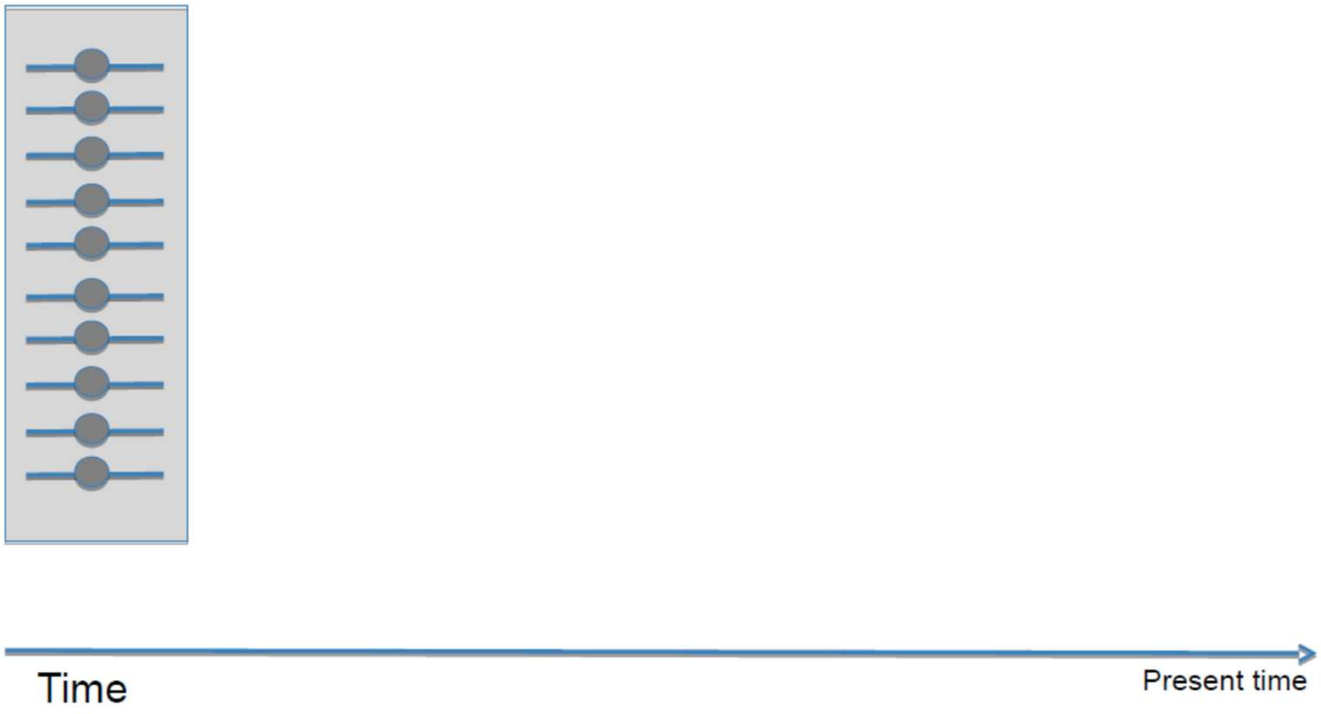
FONTE: Sabeti PC, et al. Positive natural selection in the human lineage. Science. 2006 Jun 16;312(5780):1614-20. Review.

Common Methods to Detect Selection

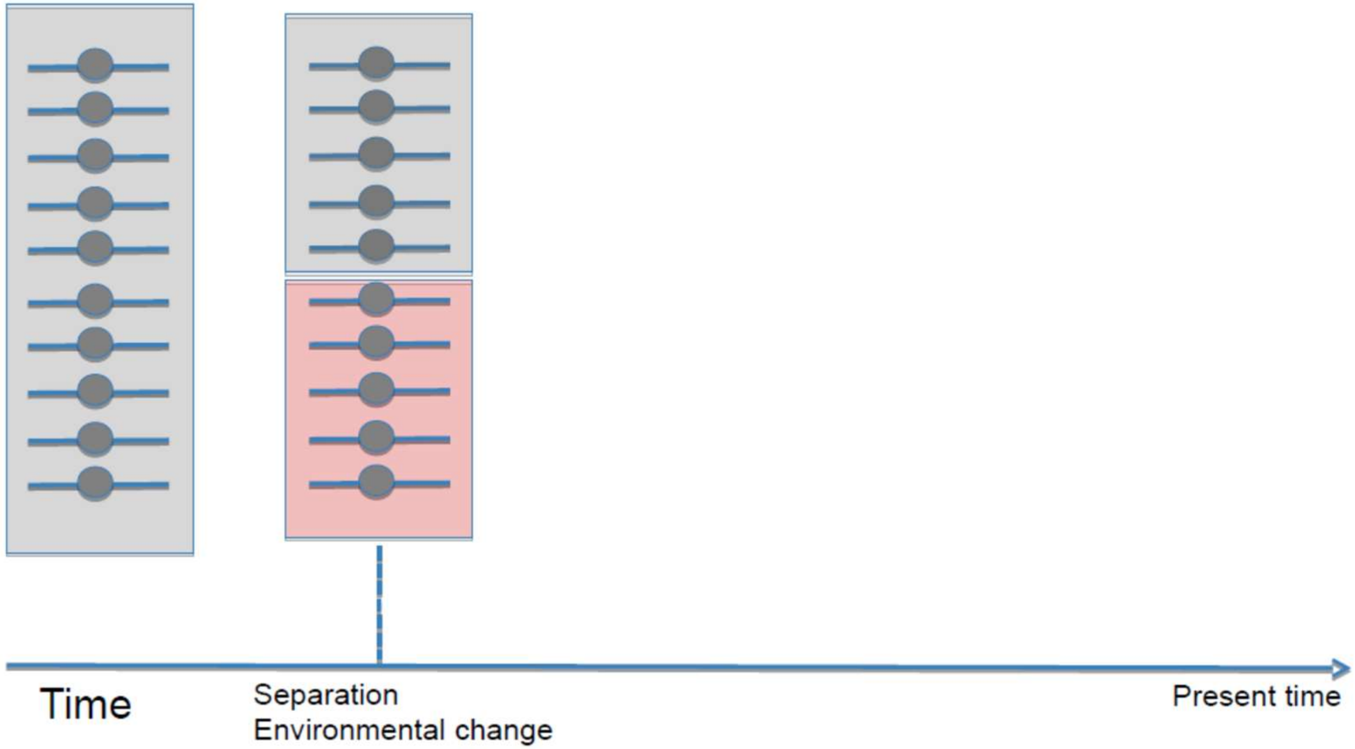

Proportion of functional changes
Heterozygosity/rare alleles
High frequency derived alleles
Population differences
Length of haplotypes
Africa
Asia
Europe
6 mya
250 kya
75 kya
25 kya

FONTE: Sabeti PC, et al. Positive natural selection in the human lineage. Science. 2006 Jun 16;312(5780):1614-20. Review.

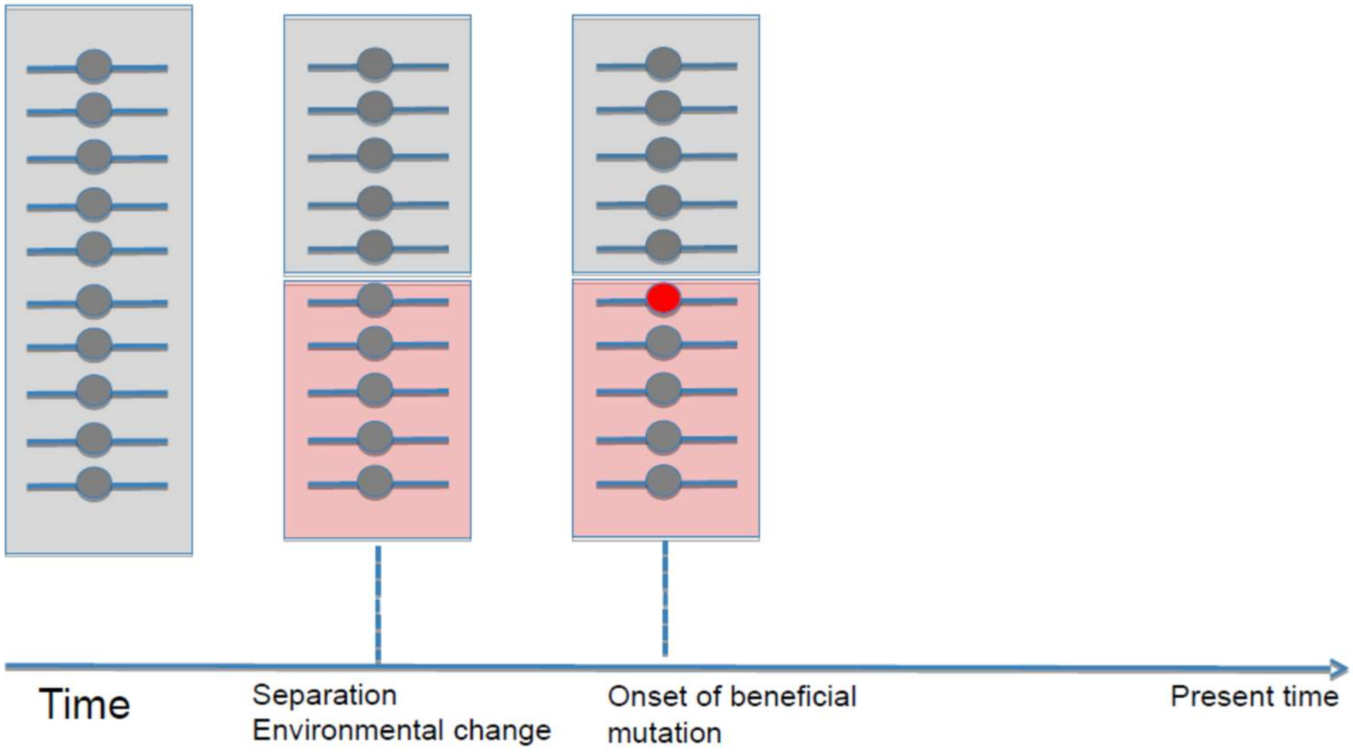
Allele Frequency Differentiation



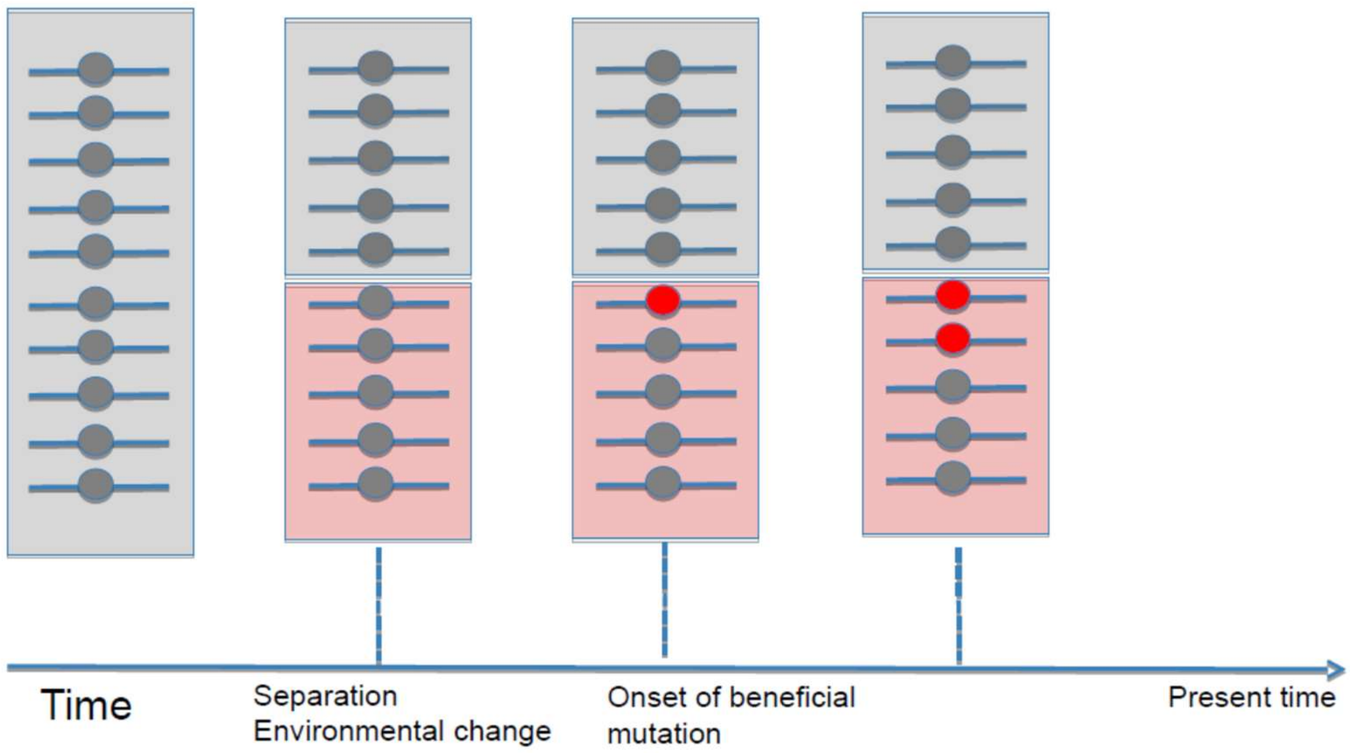
Allele Frequency Differentiation



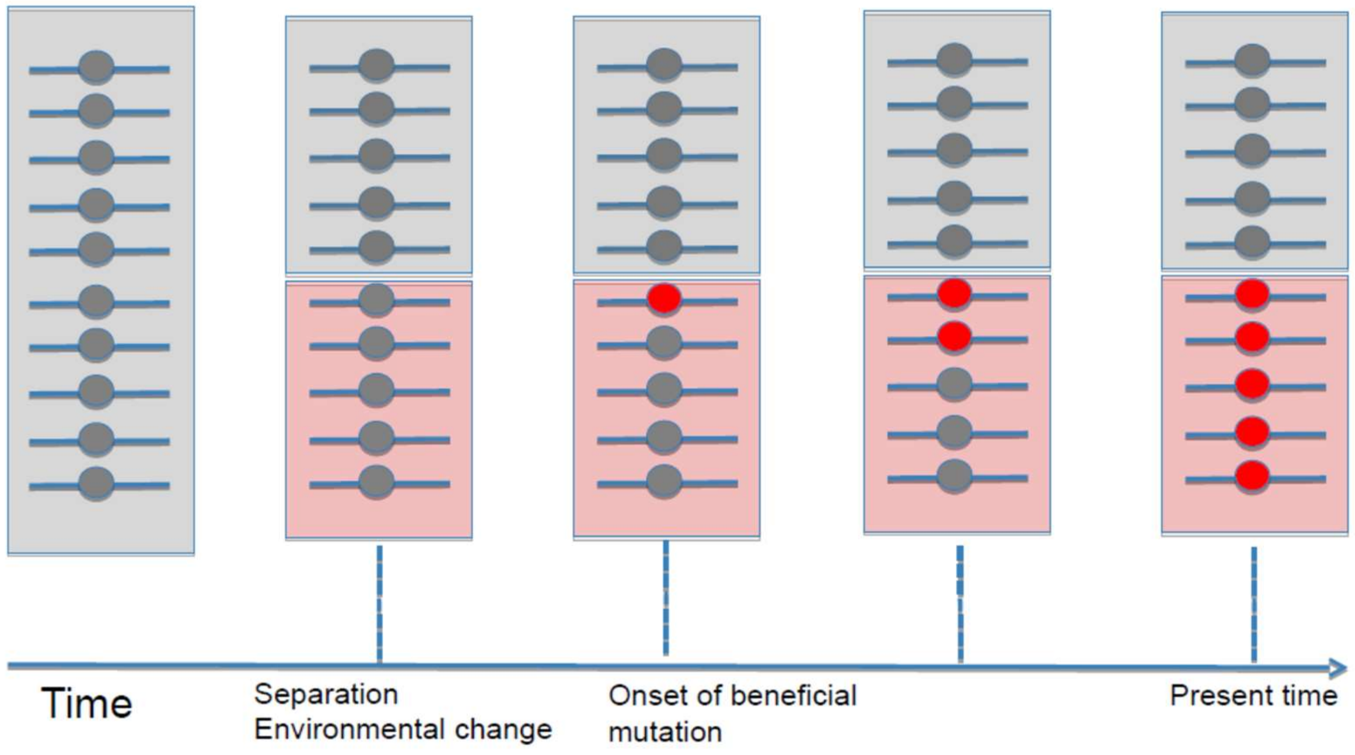
Allele Frequency Differentiation



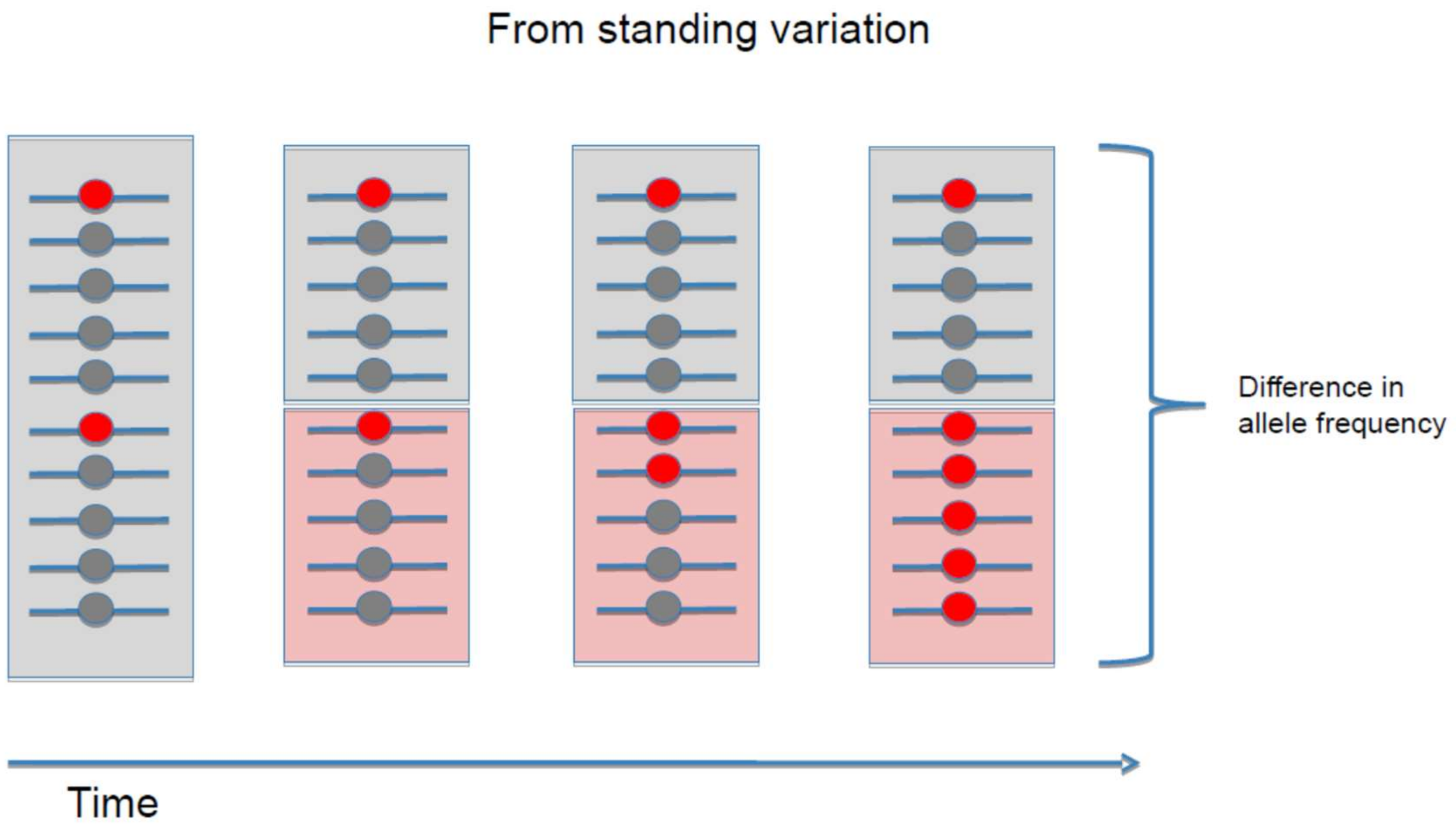
Allele Frequency Differentiation



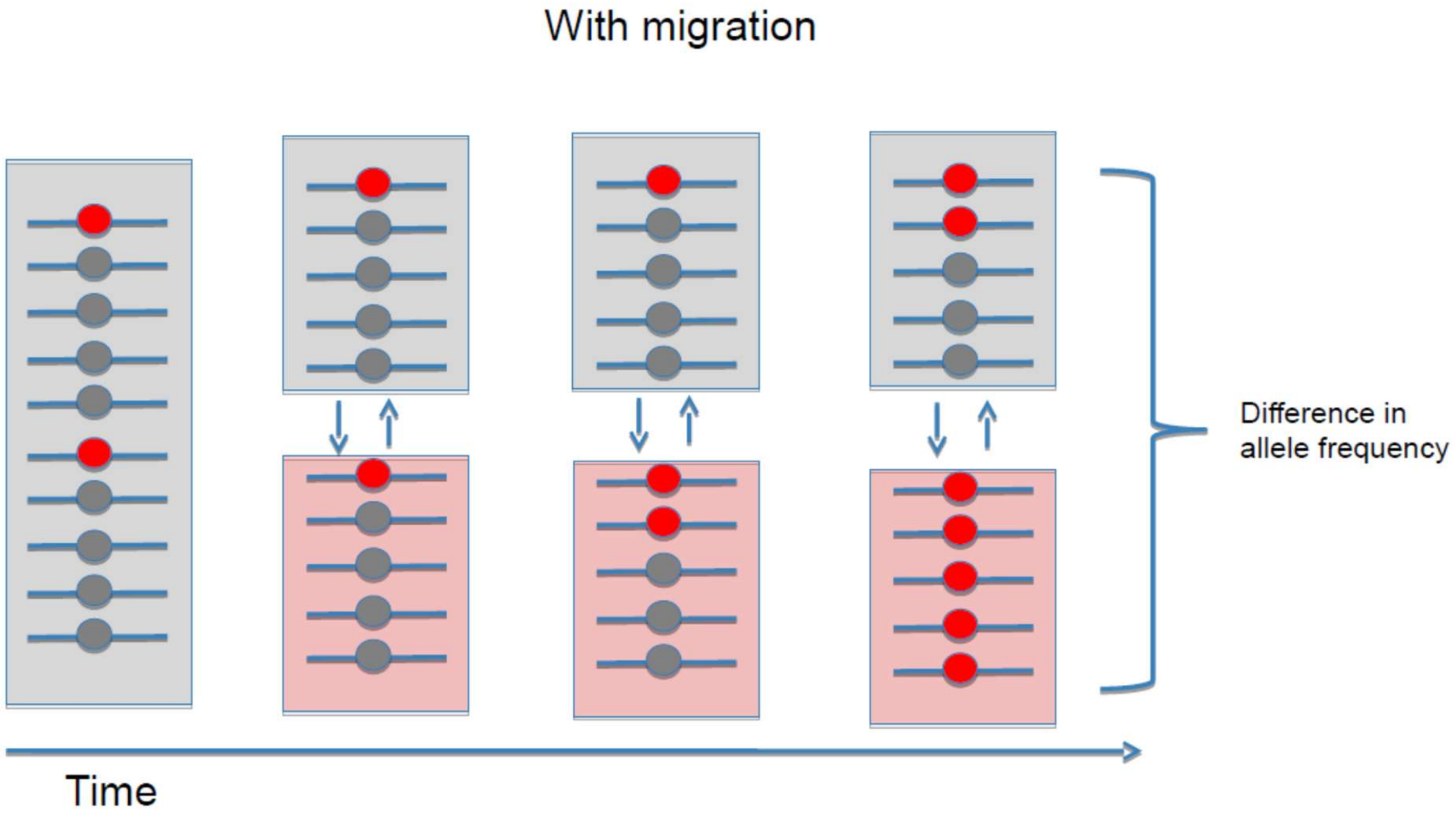
Allele Frequency Differentiation



Allele Frequency Differentiation

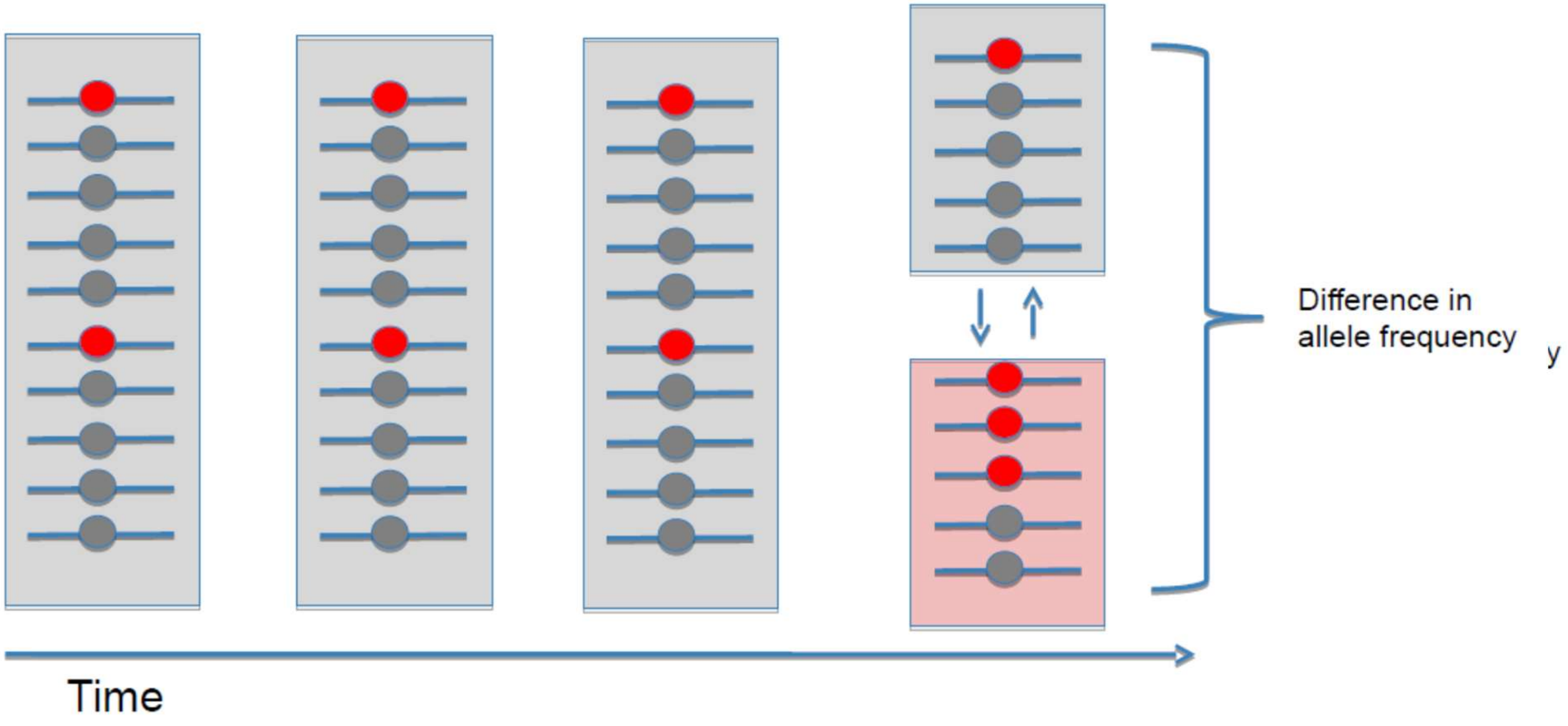


Allele Frequency Differentiation



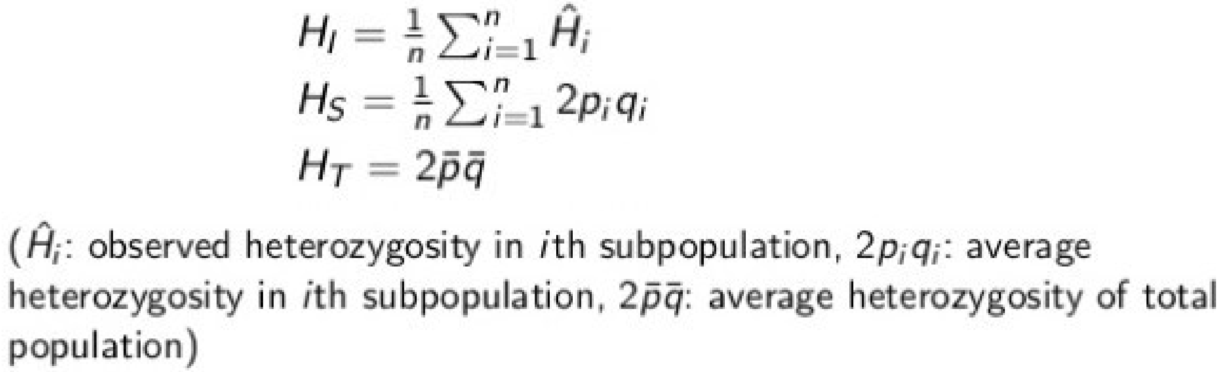
Allele Frequency Differentiation

With recent divergence



FST

Heterozigosity:



$$H_I = \frac{1}{n}\sum_{i=1}^{n}\hat{H}_i$$

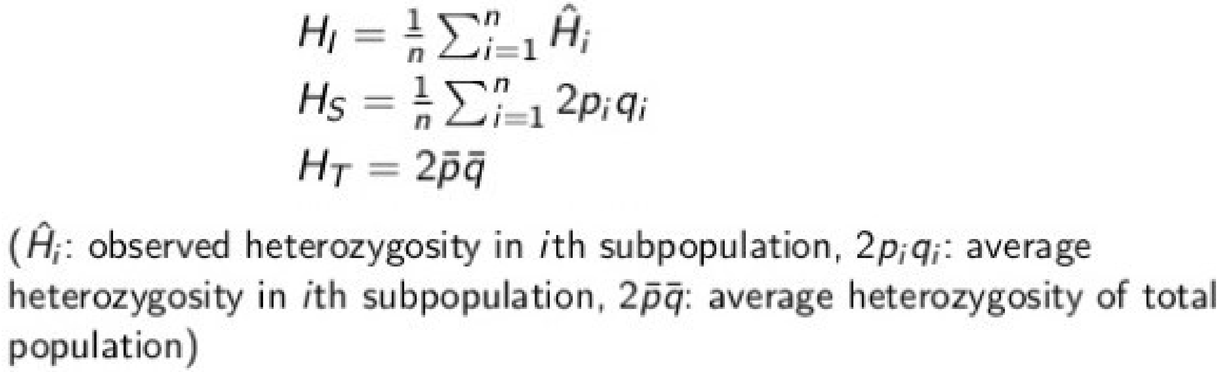
$$H_S = \frac{1}{n}\sum_{i=1}^{n}2p_iq_i$$

$$H_T = 2\bar{p}\bar{q}$$

($\hat{H}_i$: observed heterozygosity in ith subpopulation, $2p_iq_i$: average heterozygosity in ith subpopulation, $2\bar{p}\bar{q}$: average heterozygosity of total population)

FST

Heterozigose:



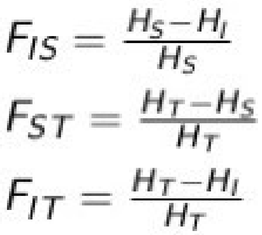
$$H_I = \frac{1}{n}\sum_{i=1}^{n}\hat{H}_i$$

$$H_S = \frac{1}{n}\sum_{i=1}^{n}2p_iq_i$$

$$H_T = 2\bar{p}\bar{q}$$

($\hat{H}_i$: observed heterozygosity in ith subpopulation, $2p_iq_i$: average heterozygosity in ith subpopulation, $2\bar{p}\bar{q}$: average heterozygosity of total population)

Estatística F (Wright)



$$F_{IS} = \frac{H_S - H_I}{H_S}$$

$$F_{ST} = \frac{H_T - H_S}{H_T}$$

$$F_{IT} = \frac{H_T - H_I}{H_T}$$

FST

Heterozigose:

$$H_I = \frac{1}{n}\sum_{i=1}^{n}\hat{H}_i$$
$$H_S = \frac{1}{n}\sum_{i=1}^{n}2p_iq_i$$
$$H_T = 2\bar{p}\bar{q}$$

($\hat{H}_i$: observed heterozygosity in ith subpopulation, $2p_iq_i$: average heterozygosity in ith subpopulation, $2\bar{p}\bar{q}$: average heterozygosity of total population)

Estatística F (Wright)

$$F_{IS} = \frac{H_S - H_I}{H_S}$$
$$F_{ST} = \frac{H_T - H_S}{H_T}$$
$$F_{IT} = \frac{H_T - H_I}{H_T}$$

FST

F_{ST}

Common measure for quantifying population subdivision.

$$F_{ST} = H_B / (H_W + H_B)$$

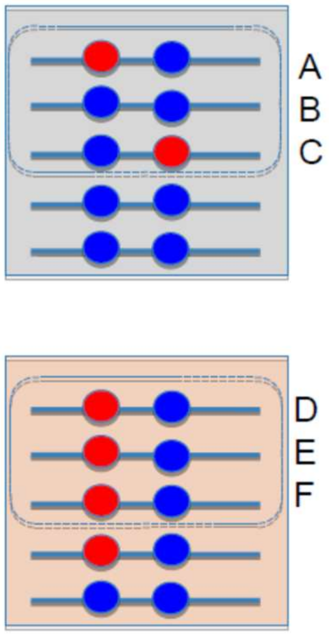
H_B: between populations

H_W: average within populations

- if $H_W << H_B 0$ then $F_{ST} \sim 1$
- if $H_B = 0$ then $F_{ST} = 0$

FST

F_{ST} based on haplotype differentiation between populations



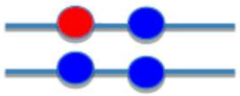
$$F_{ST} = 1 – (H_W / H_B)$$

Within populations

Between populations

What is the variation within populations?

e.g. A vs B



The differ by 1 site

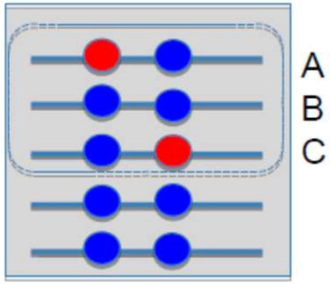
Hudson et al. 1992.

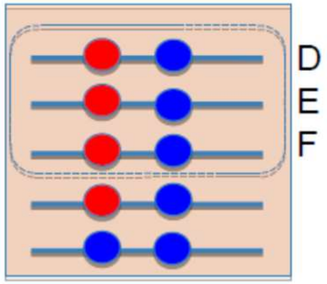
FST

$$F_{ST} = 1 - (H_W / H_B)$$

Within populations

Between populations





What is the variation within populations?

A	B
A	C
B	C

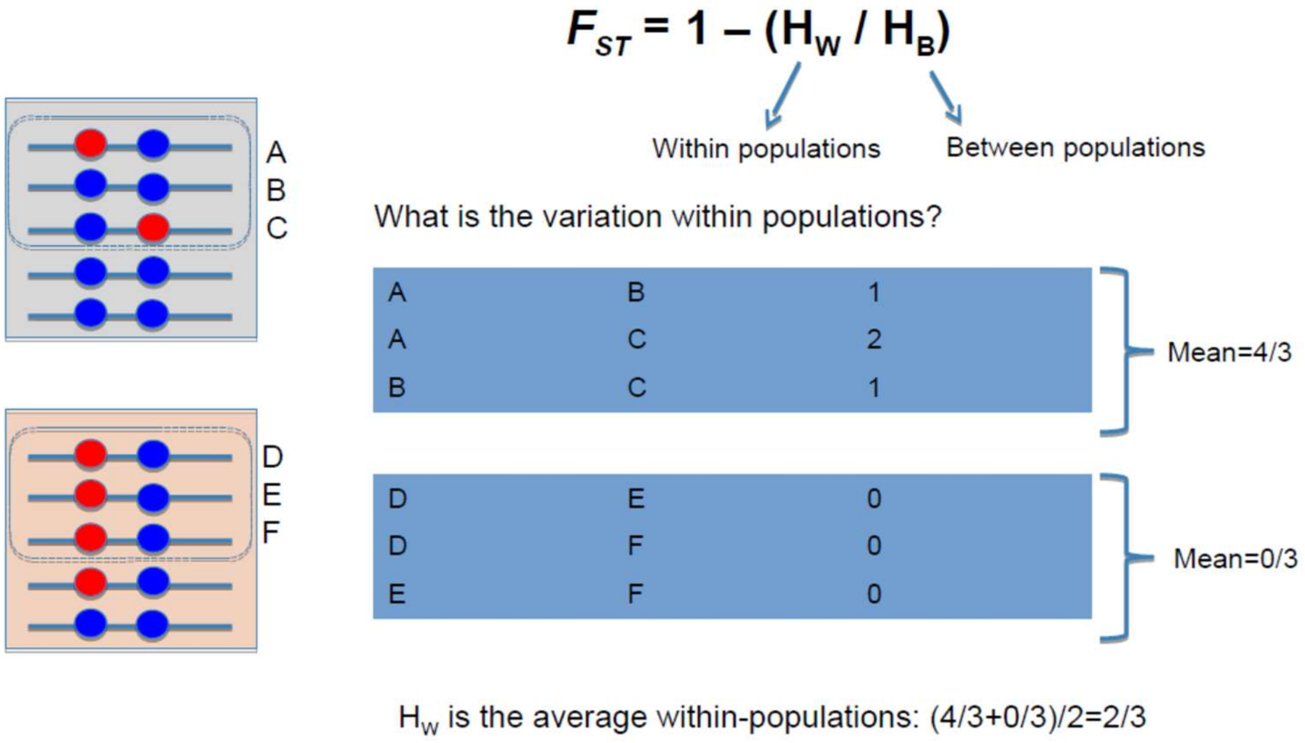
Mean=?

D	E
D	F
E	F

Mean=?

H_W is the average within-populations: ?

FST



$$F_{ST} = 1 - (H_W / H_B)$$

Within populations

Between populations

What is the variation within populations?

A	B	1
A	C	2
B	C	1

Mean=4/3

D	E	0
D	F	0
E	F	0

Mean=0/3

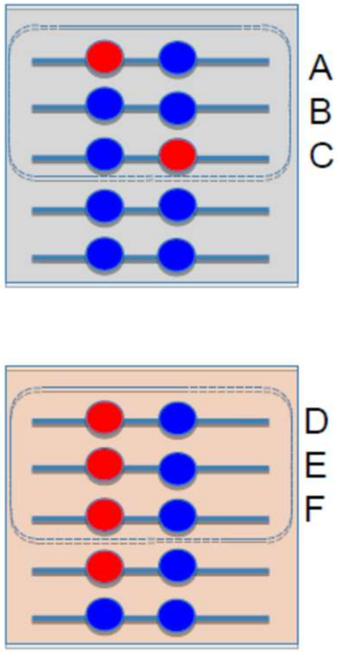
H_W is the average within-populations: (4/3+0/3)/2=2/3

FST

$$F_{ST} = 1 - (H_W / H_B)$$

Within populations

Between populations



What is the variation between populations?

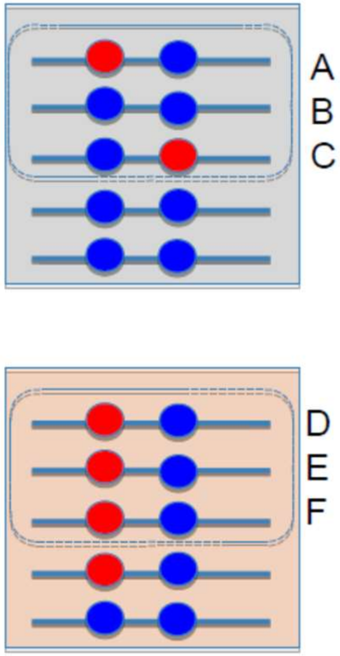
A	D	0
A	E	0
A	F	0
B	D	1
B	E	1
B	F	1
C	D	2
C	E	2
C	F	2

Mean=9/9

H_B is the average between-populations: 9/9=1

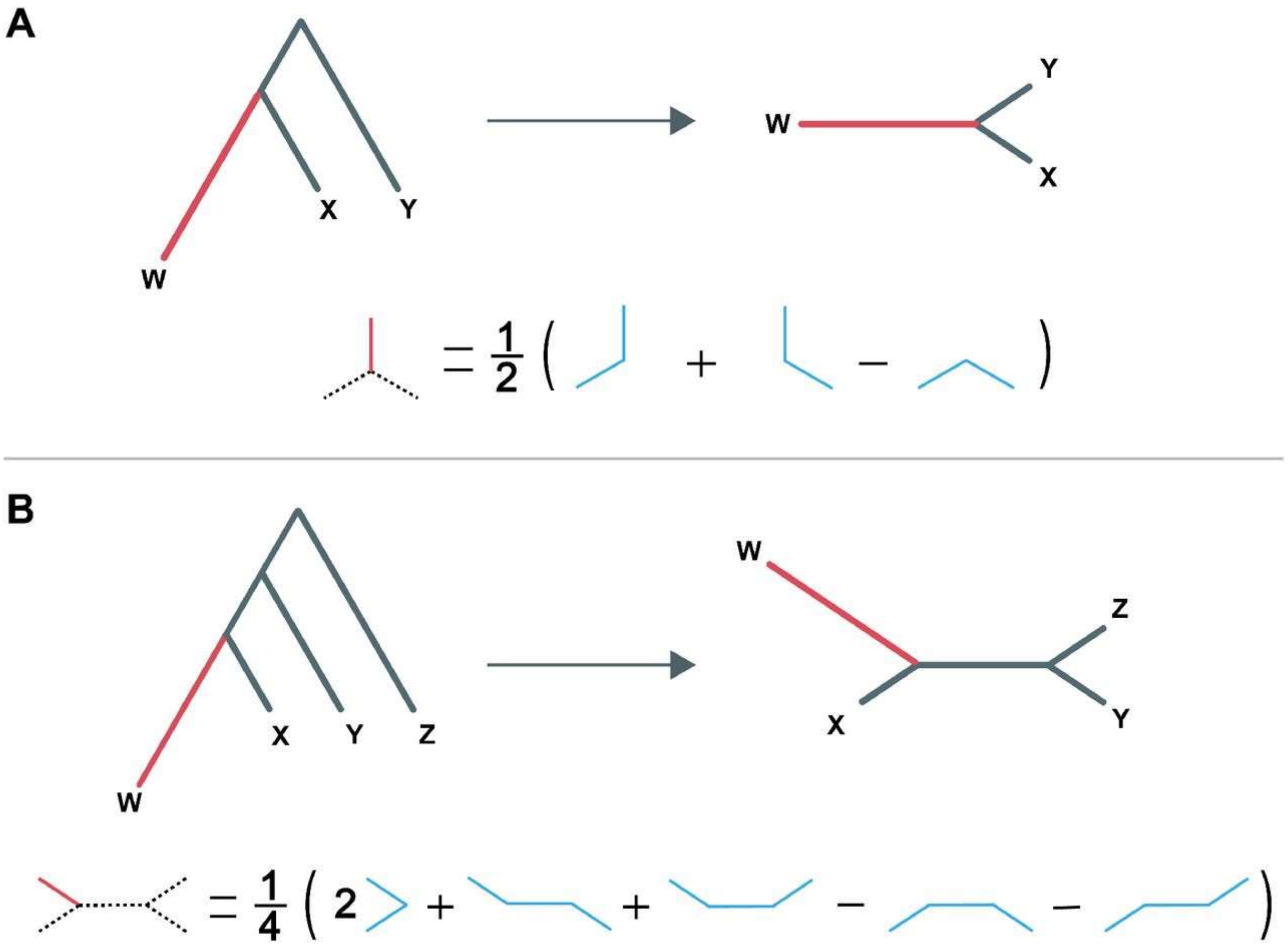
FST

F_{ST} based on haplotype differentiation between populations

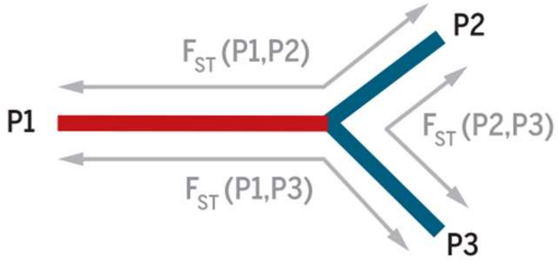


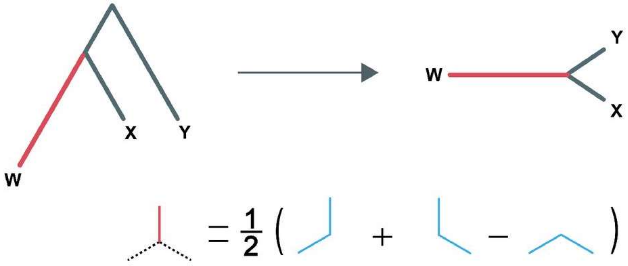
$$F_{ST} = 1 - (H_W / H_B) = 1 - ((2/3)/1) = 1/3 \sim 0.33$$

Population Branch Statistic (PBS)

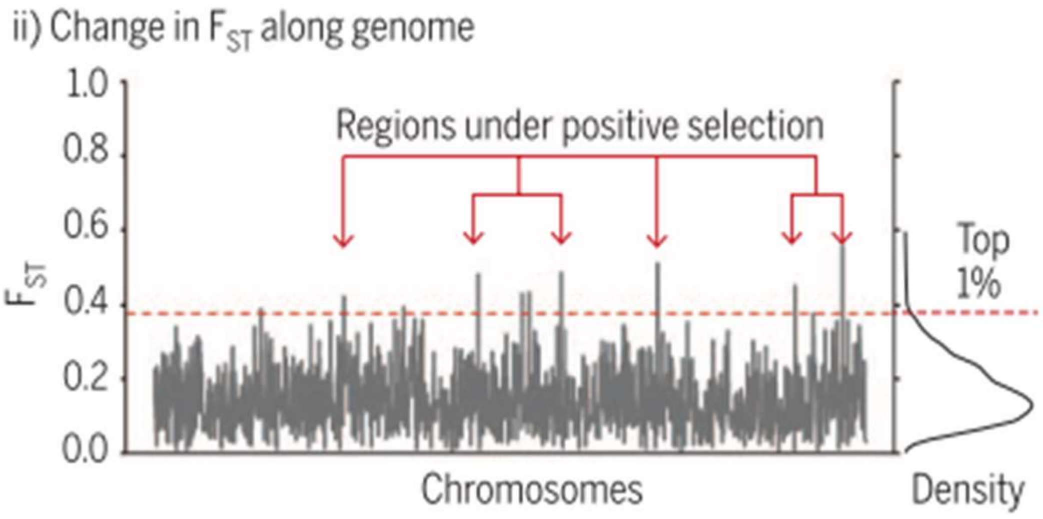


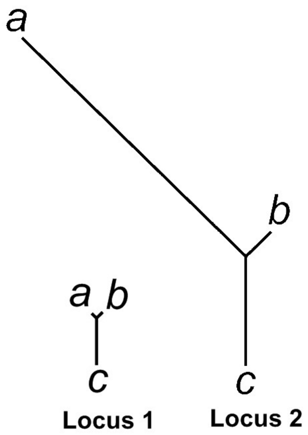
Population Branch Statistic (PBS)



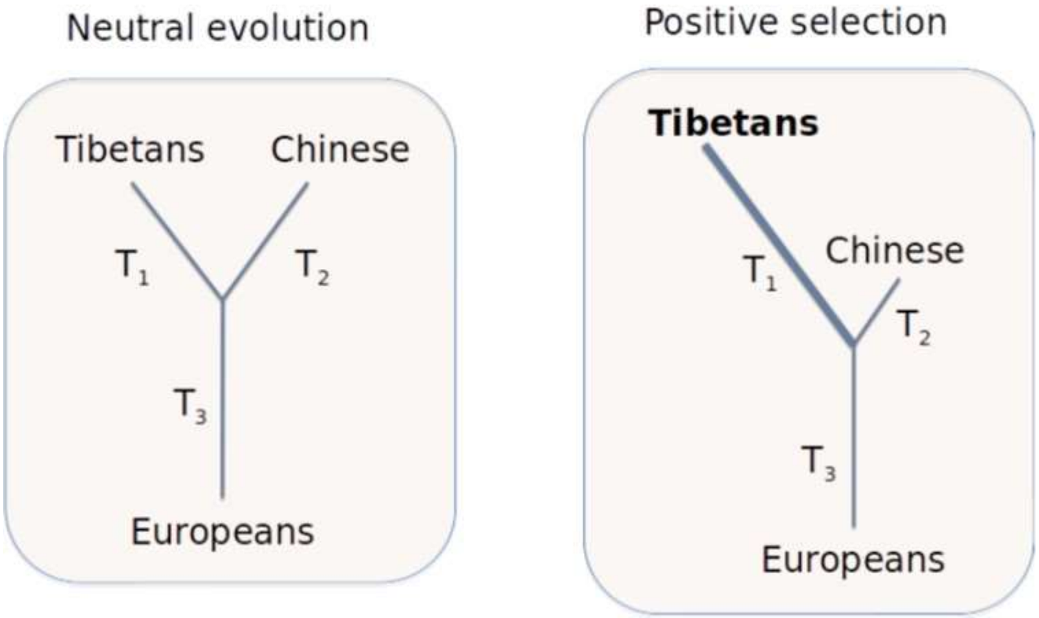


$$\text{PBS} = \frac{\text{Fst(P1;P2)} + \text{Fst(P1;P3)} - \text{Fst(P2;P3)}}{2}$$



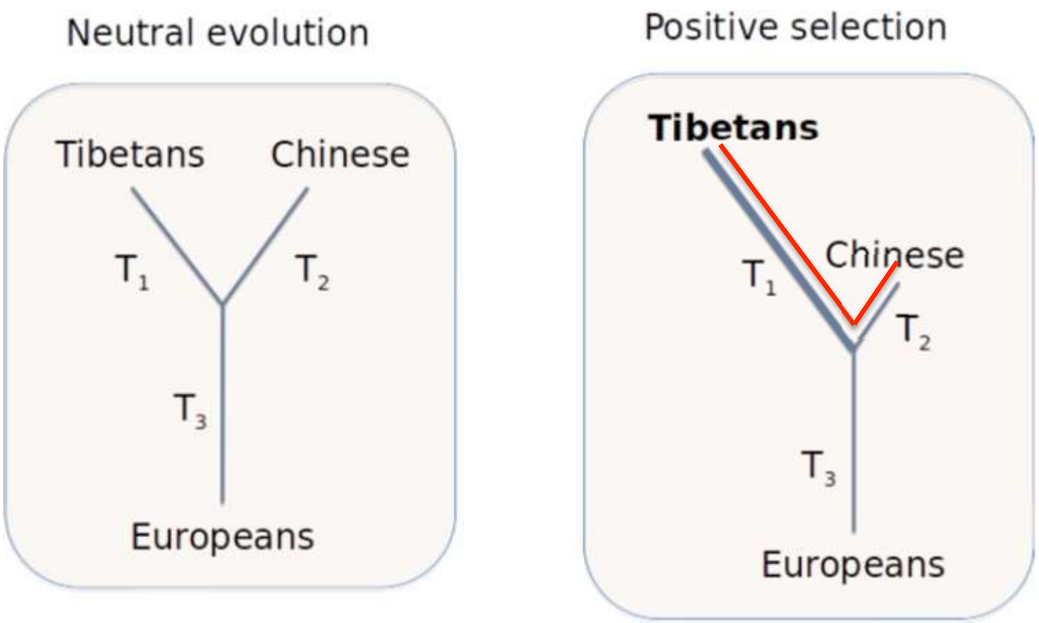


Population Branch Statistic (PBS)



Yi X, et al. Sequencing of 50 human exomes reveals adaptation to high altitude. Science. 2010 Jul 2;329(5987):75-8.

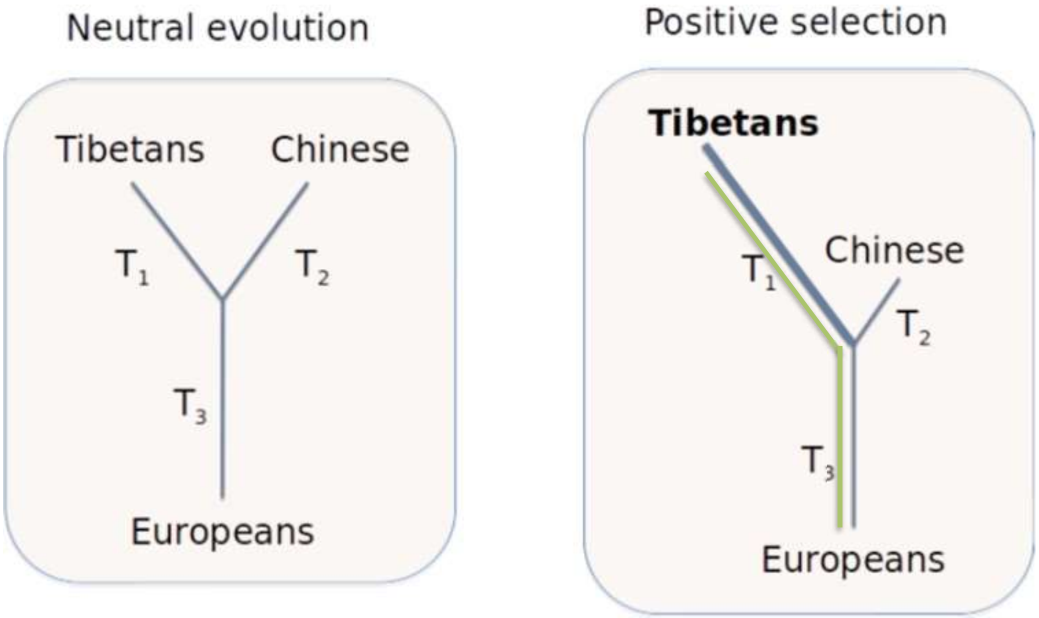
Population Branch Statistic (PBS)



PBS = T1T2+T1T3-T2T3

Yi X, et al. Sequencing of 50 human exomes reveals adaptation to high altitude. Science. 2010 Jul 2;329(5987):75-8.

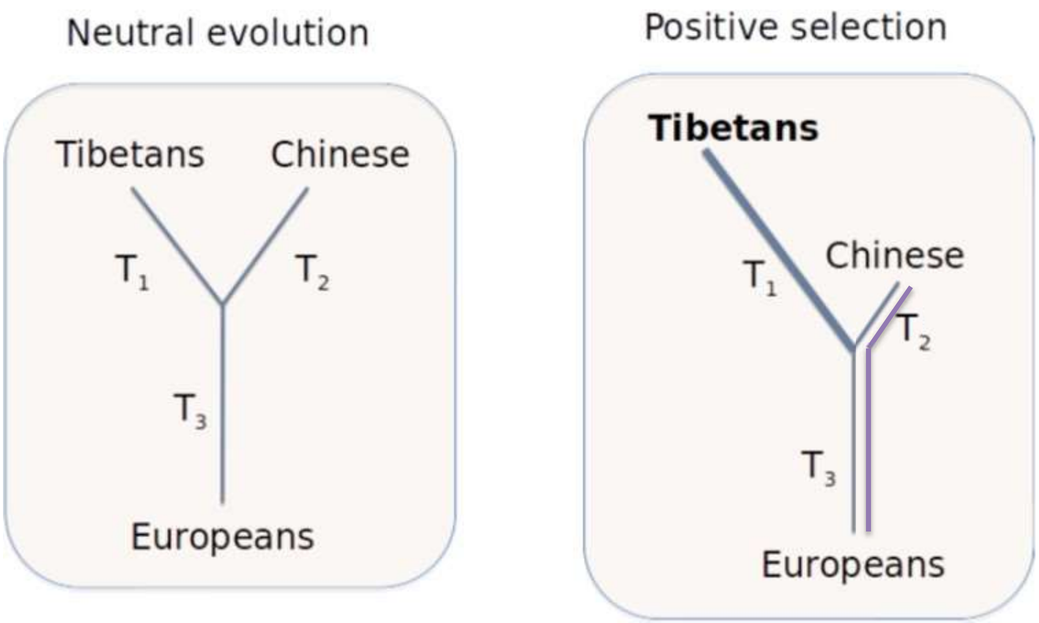
Population Branch Statistic (PBS)



PBS = T1T2+T1T3-T2T3

Yi X, et al. Sequencing of 50 human exomes reveals adaptation to high altitude. Science. 2010 Jul 2;329(5987):75-8.

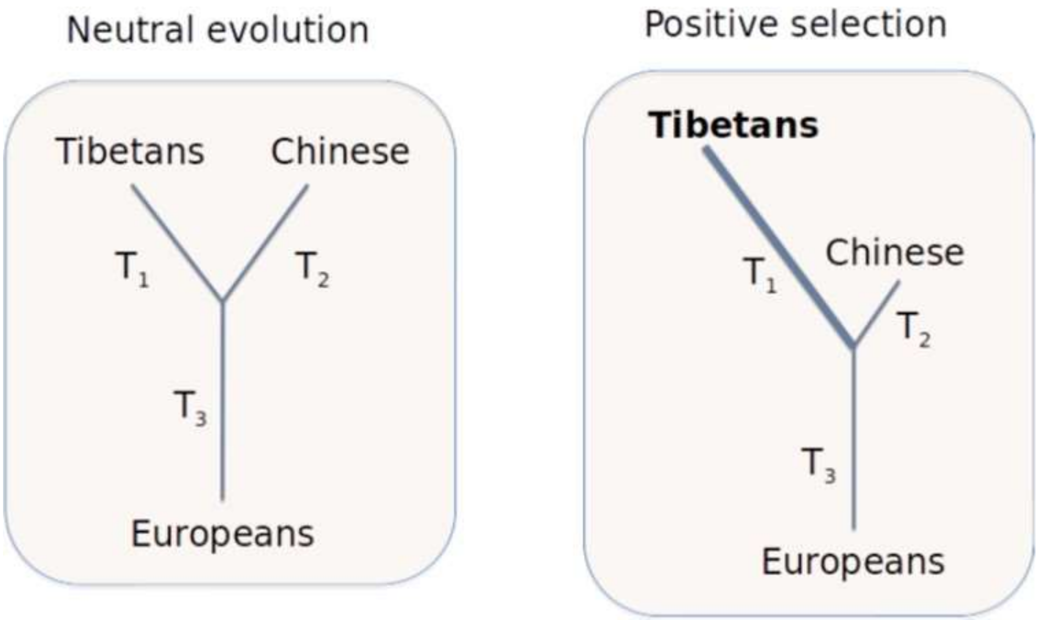
Population Branch Statistic (PBS)



PBS = T1T2+T1T3-T2T3

Yi X, et al. Sequencing of 50 human exomes reveals adaptation to high altitude. Science. 2010 Jul 2;329(5987):75-8.

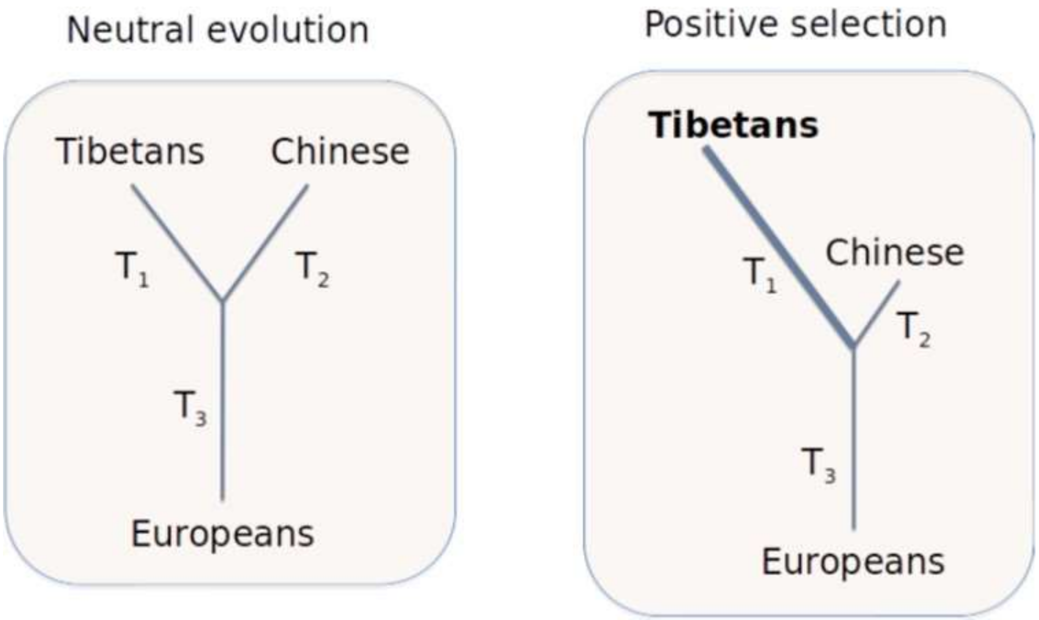
Population Branch Statistic (PBS)



PBS = T1T2+T1T3-T2T3/2

Yi X, et al. Sequencing of 50 human exomes reveals adaptation to high altitude. Science. 2010 Jul 2;329(5987):75-8.

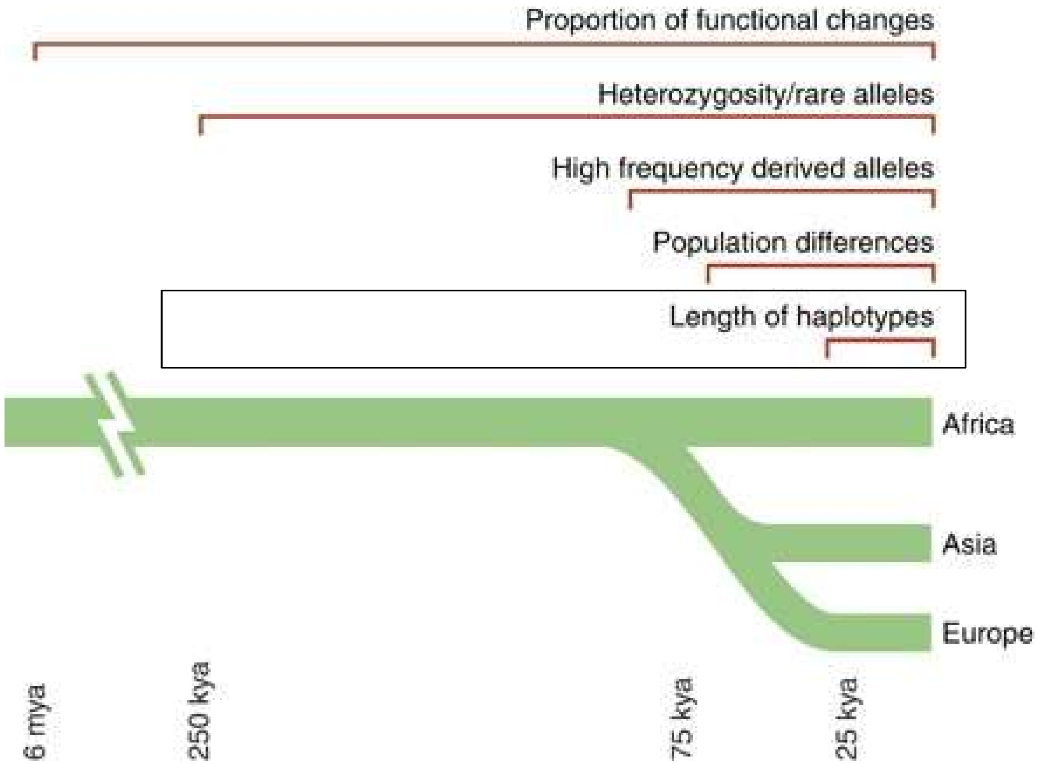
Population Branch Statistic (PBS)



$$PBS = FST_{TIB\_CHB} + FST_{TIB\_EUR} - FST_{CHB\_EUR}/2$$

Yi X, et al. Sequencing of 50 human exomes reveals adaptation to high altitude. Science. 2010 Jul 2;329(5987):75-8.

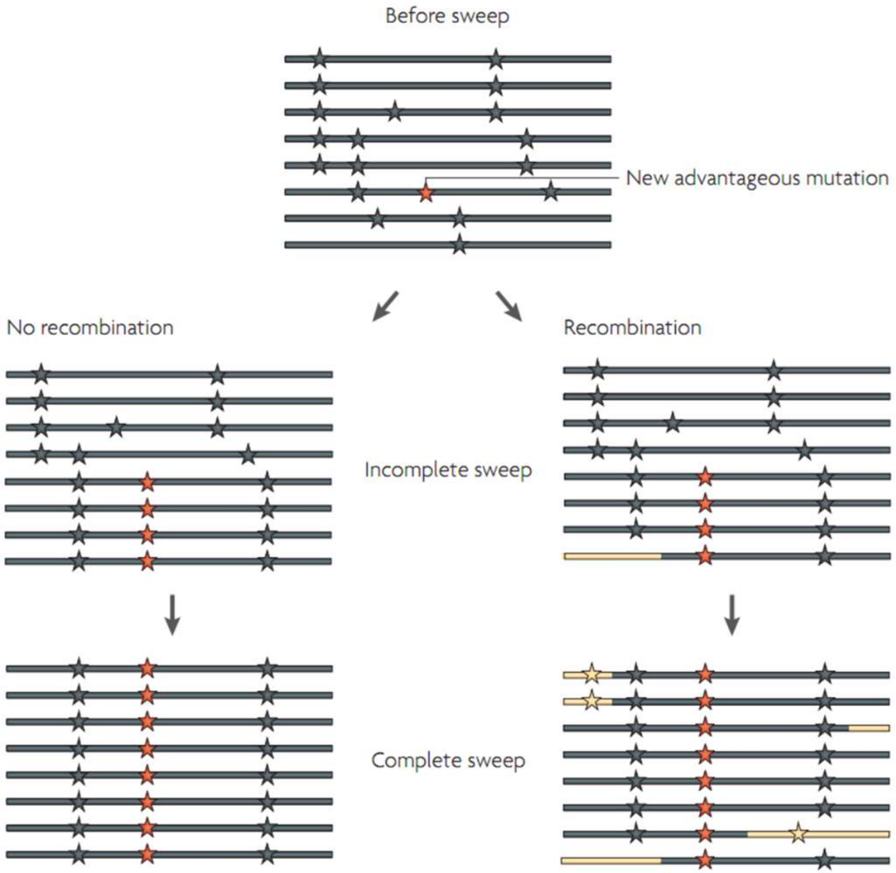
Testes de Seleção intra-específicos


Proportion of functional changes
Heterozygosity/rare alleles
High frequency derived alleles
Population differences
Length of haplotypes
Africa
Asia
Europe
6 mya
250 kya
75 kya
25 kya

FONTE: Sabeti PC, et al. Positive natural selection in the human lineage. Science. 2006 Jun 16;312(5780):1614-20. Review.

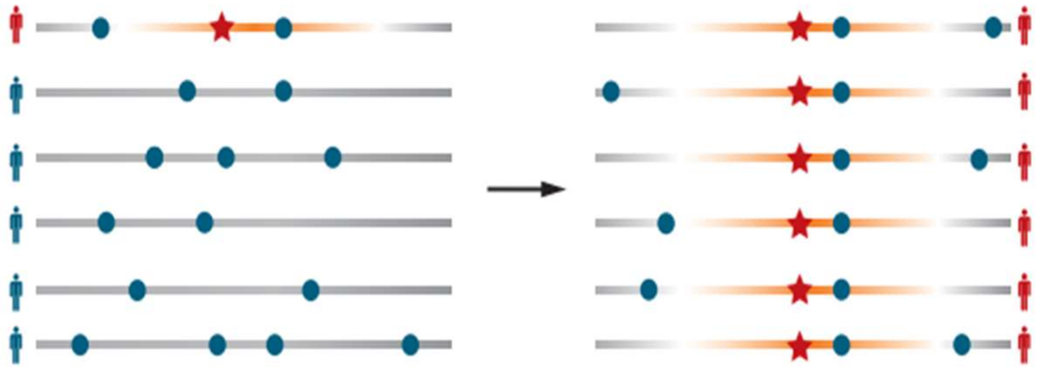
Extended Haplotype Homozygosity (EHH)

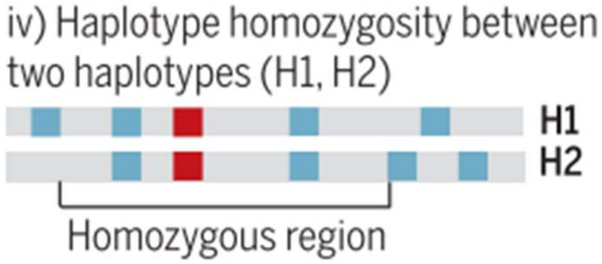
Extended haplotype homozygosity (EHH): EHH at distance x from the core region is the probability that two randomly chosen chromosomes carry a tested core haplotype are homozygous at all SNPs for the entire interval from the core region to the distance x.

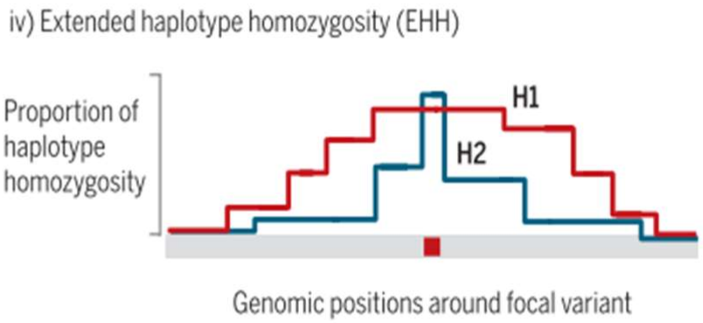


Extended Haplotype Homozygosity (EHH)

Hard sweep

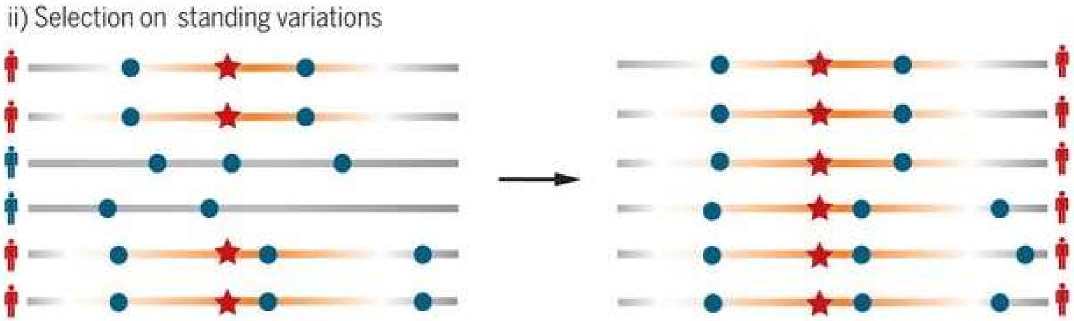


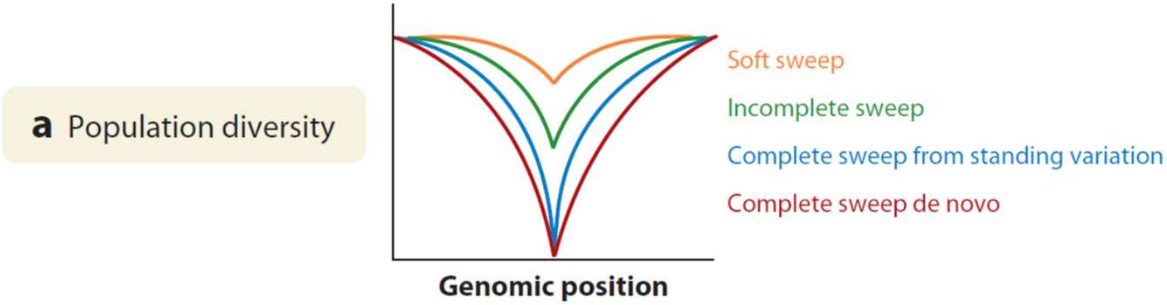




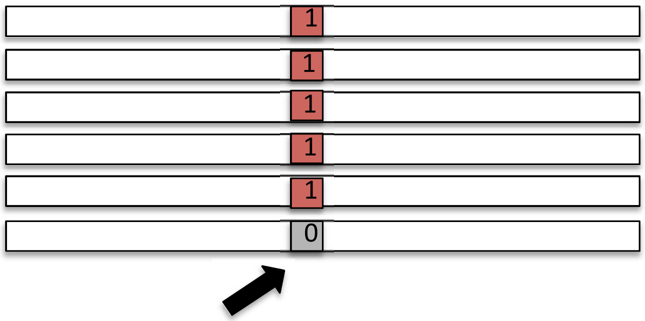
Extended Haplotype Homozygosity (EHH)

Soft sweep





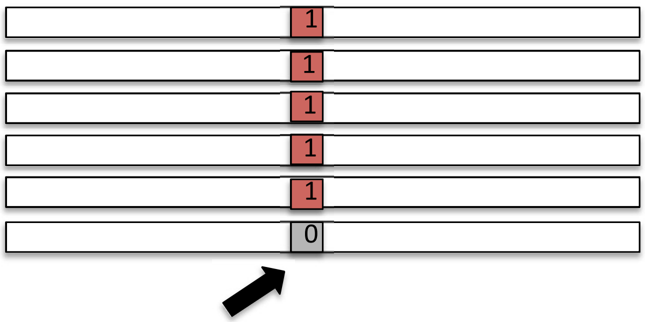
Extended Haplotype Homozygosity (EHH)



Core allele 1 : biallelic loci; 0 is the ancestral allele and 1 is the derived allele.

Adapted from Mateus Fumagalli.

Extended Haplotype Homozygosity (EHH)

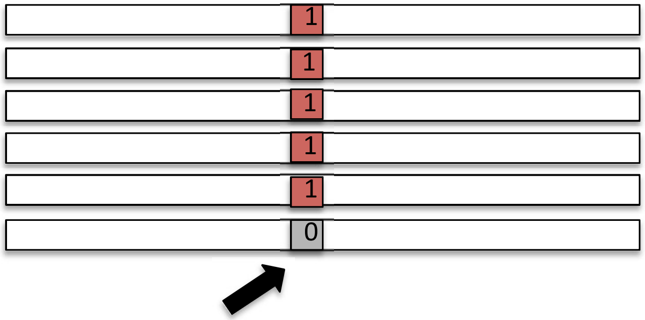


Core allele 1 : biallelic loci; 0 is the ancestral allele and 1 is the derived allele.

$$EHH_c(x_i) = \sum_{h \in H_c(x_i)} \frac{\binom{n_h}{2}}{\binom{n_c}{2}}$$

Adapted from Mateus Fumagalli.

Extended Haplotype Homozygosity (EHH)



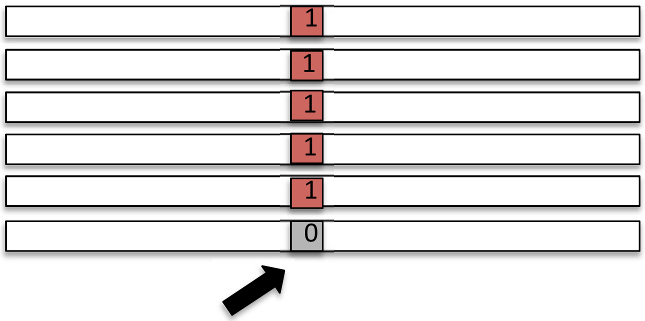
Core allele 1 : biallelic loci; 0 is the ancestral allele and 1 is the derived allele.

$$EHH_c(x_i) = \sum_{h \in H_c(x_i)} \frac{\binom{n_h}{2}}{\binom{n_c}{2}}$$

From $x_{0\ to}\ x_i$

Adapted from Mateus Fumagalli.

Extended Haplotype Homozygosity (EHH)



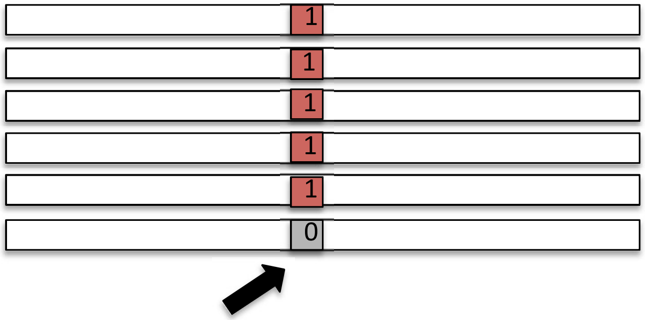
Core allele 1 : biallelic loci; 0 is the ancestral allele and 1 is the derived allele.

$$EHH_c(x_i) = \sum_{h \in H_c(x_i)} \frac{\binom{n_h}{2}}{\binom{n_c}{2}}$$

Sum of all haplotypes containing the allele of interest (core allele)

Adapted from Mateus Fumagalli.

Extended Haplotype Homozygosity (EHH)



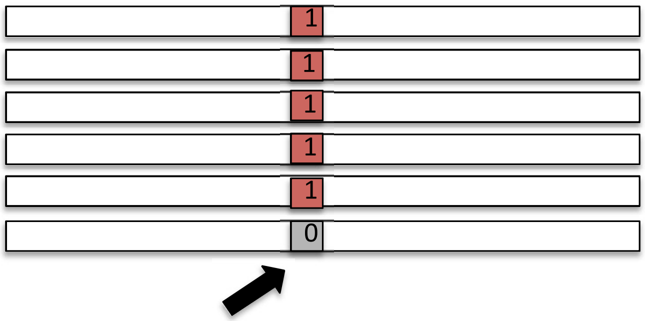
Core allele 1 : biallelic loci; 0 is the ancestral allele and 1 is the derived allele.

$$EHH_c(x_i) = \sum_{h \in H_c(x_i)} \frac{\binom{n_h}{2}}{\binom{n_c}{2}}$$

n_h haplotype Frequency of h

Adapted from Mateus Fumagalli.

Extended Haplotype Homozygosity (EHH)



Core allele 1 : biallelic loci; 0 is the ancestral allele and 1 is the derived allele.

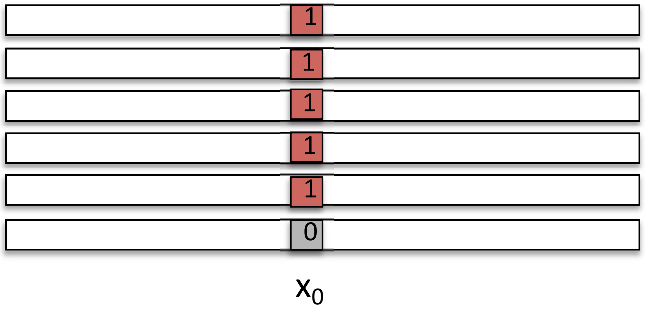
$$EHH_c(x_i) = \sum_{h \in H_c(x_i)} \frac{\binom{n_h}{2}}{\binom{n_c}{2}}$$

n_h haplotype Frequency of h

N_c haplotype Frequency with core SNP

Adapted from Mateus Fumagalli.

Extended Haplotype Homozygosity (EHH)



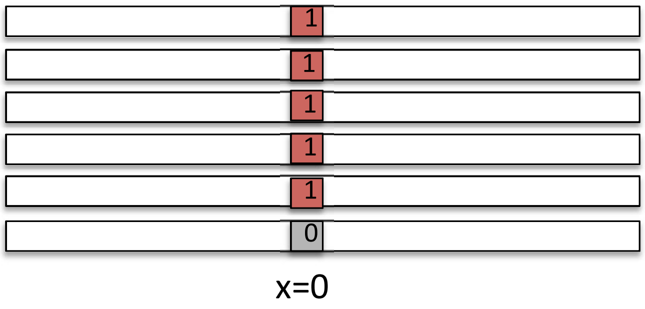
What is the EHH to x_0?

$$EHH_c(x_i) = \sum_{h \in H_c(x_i)} \frac{\binom{n_h}{2}}{\binom{n_c}{2}}$$

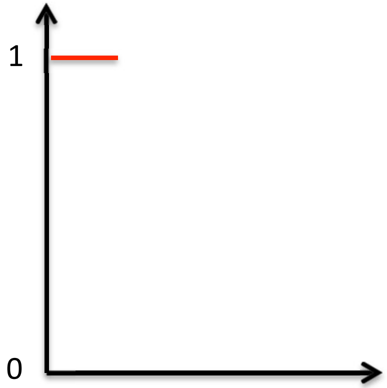
$$EHH_c(x_i = 0) = \frac{\binom{5}{2}}{\binom{5}{2}} = 1$$

FONTE: Mateus Fumagalli slides modificado

Extended Haplotype Homozygosity (EHH)



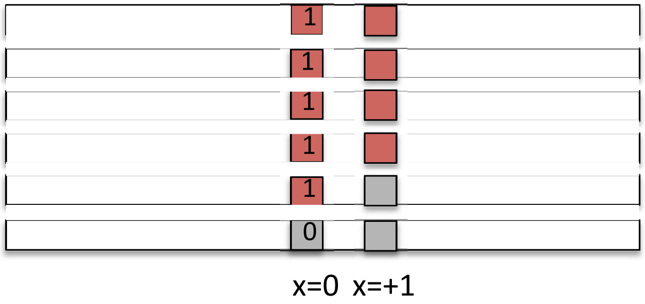
$$EHH_c(x_i) = \sum_{h \in H_c(x_i)} \frac{\binom{n_h}{2}}{\binom{n_c}{2}}$$



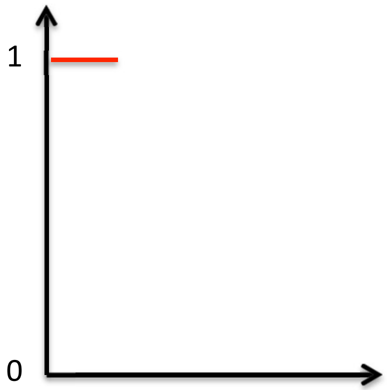
$$EHH_c(x_i = 0) = \frac{\binom{5}{2}}{\binom{5}{2}} = 1$$

Adapted from Mateus Fumagalli.

Extended Haplotype Homozygosity (EHH)



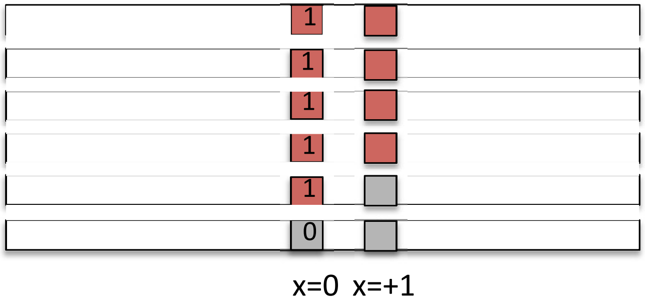
$$EHH_c(x_i)=\sum_{h\in H_c(x_i)}\frac{\binom{n_h}{2}}{\binom{n_c}{2}}$$



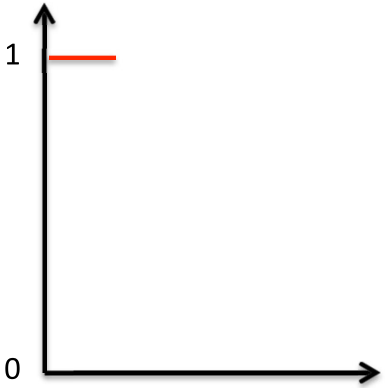
$$EHH_c(x_i=+1)=\ ?$$

Adapted from Mateus Fumagalli.

Extended Haplotype Homozygosity (EHH)



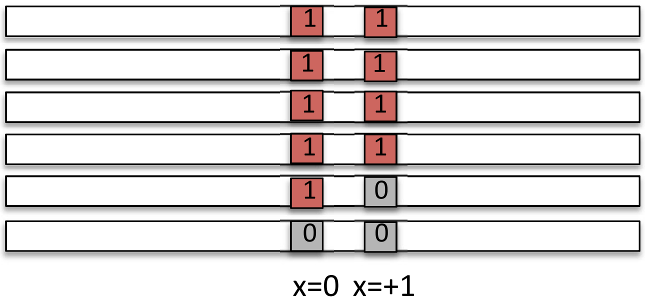
$$EHH_c(x_i) = \sum_{h \in H_c(x_i)} \frac{\binom{n_h}{2}}{\binom{n_c}{2}}$$



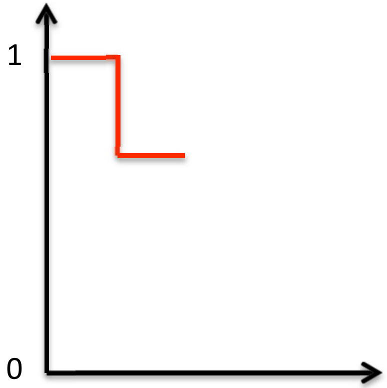
$$EHH_c(x_i = +1) = \frac{\binom{4}{2} + \binom{1}{2}}{\binom{5}{2}} = \frac{6+0}{10} = 0.60$$

Adapted from Mateus Fumagalli.

Extended Haplotype Homozygosity (EHH)



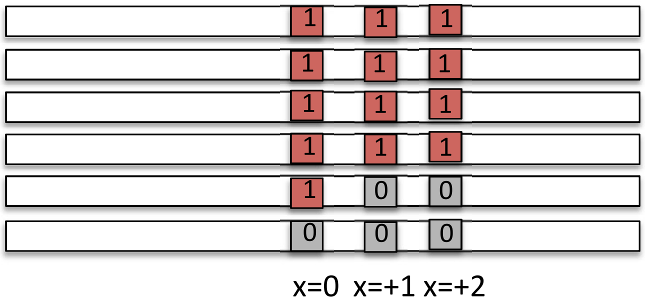
$$EHH_c(x_i) = \sum_{h \in H_c(x_i)} \frac{\binom{n_h}{2}}{\binom{n_c}{2}}$$



$$EHH_c(x_i = +1) = \frac{\binom{4}{2} + \binom{1}{2}}{\binom{5}{2}} = \frac{6+0}{10} = 0.60$$

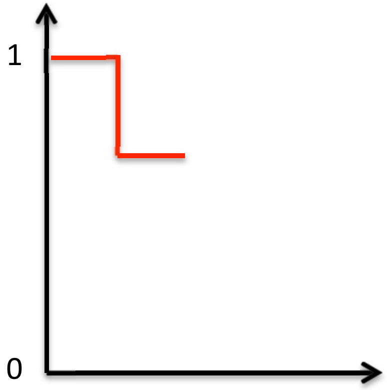
Adapted from Mateus Fumagalli.

Extended Haplotype Homozygosity (EHH)



1	1	1
1	1	1
1	1	1
1	1	1
1	0	0
0	0	0
x=0	x=+1	x=+2

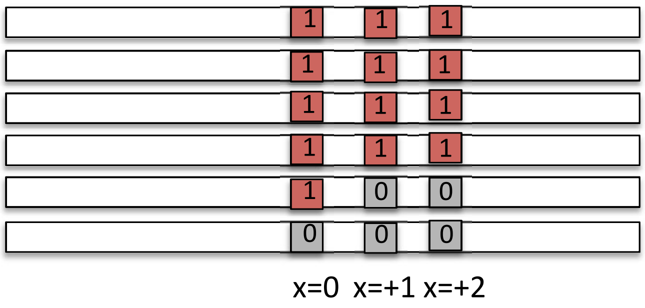
$$EHH_c(x_i) = \sum_{h \in H_c(x_i)} \frac{\binom{n_h}{2}}{\binom{n_c}{2}}$$



$$EHH_c(x_i = +2) = ?$$

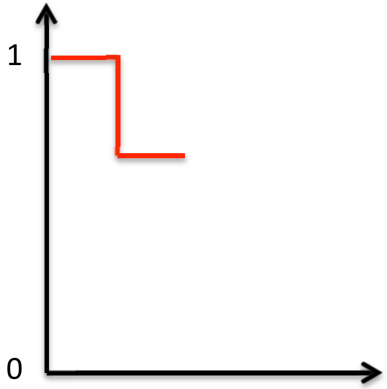
Adapted from Mateus Fumagalli.

Extended Haplotype Homozygosity (EHH)



x=0	x=+1	x=+2
1	1	1
1	1	1
1	1	1
1	1	1
1	0	0
0	0	0

$$EHH_c(x_i) = \sum_{h \in H_c(x_i)} \frac{\binom{n_h}{2}}{\binom{n_c}{2}}$$

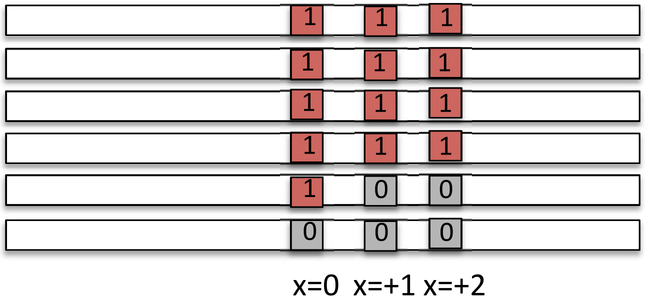


$$EHH_c(x_i = +2) = ?$$

How many unique haplotypes carriyng the core SNP?
What is their frequency?

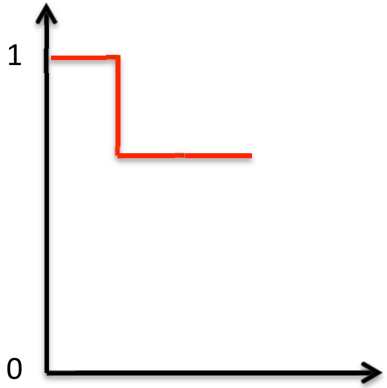
Adapted from Mateus Fumagalli.

Extended Haplotype Homozygosity (EHH)



x=0	x=+1	x=+2
1	1	1
1	1	1
1	1	1
1	1	1
1	0	0
0	0	0

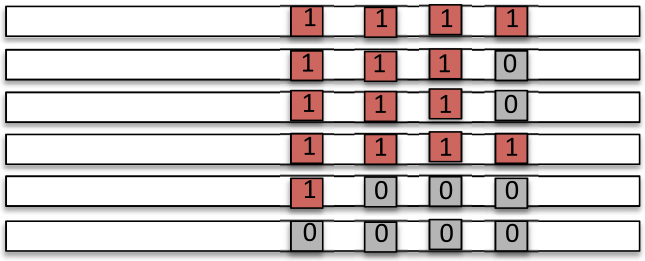
$$EHH_c(x_i) = \sum_{h \in H_c(x_i)} \frac{\binom{n_h}{2}}{\binom{n_c}{2}}$$



$$EHH_c(x_i = +2) = EHH_c(x_i = +1) = 0.60$$

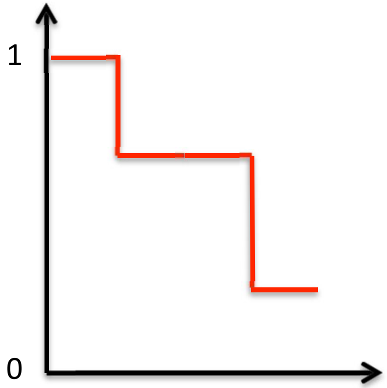
Adapted from Mateus Fumagalli.

Extended Haplotype Homozygosity (EHH)



x=0	x=+1	x=+2	x=+3
1	1	1	1
1	1	1	0
1	1	1	0
1	1	1	1
1	0	0	0
0	0	0	0

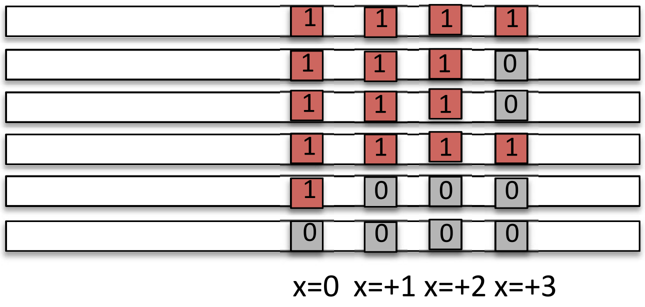
$$EHH_c(x_i) = \sum_{h \in H_c(x_i)} \frac{\binom{n_h}{2}}{\binom{n_c}{2}}$$



How many unique haplotypes carriyng the core SNP?
What is their frequency?

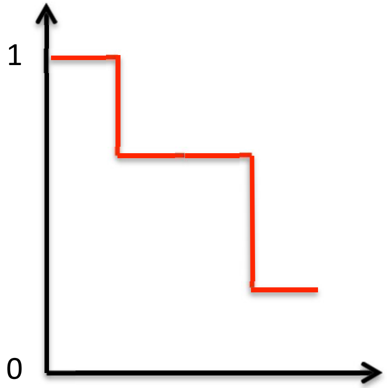
Adapted from Mateus Fumagalli.

Extended Haplotype Homozygosity (EHH)



x=0	x=+1	x=+2	x=+3
1	1	1	1
1	1	1	0
1	1	1	0
1	1	1	1
1	0	0	0
0	0	0	0

$$EHH_c(x_i) = \sum_{h\in H_c(x_i)} \frac{\binom{n_h}{2}}{\binom{n_c}{2}}$$



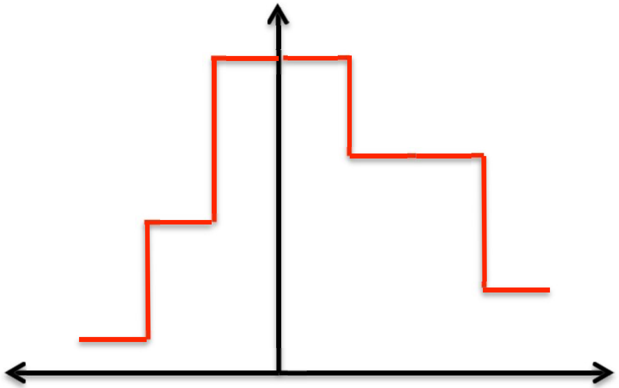
$$EHH_c(x_i = +3) = \frac{\binom{2}{2}+\binom{2}{2}+\binom{1}{2}}{\binom{5}{2}} = \frac{1+1+0}{10} = 0.20$$

Adapted from Mateus Fumagalli.

Extended Haplotype Homozygosity (EHH)

x=-2	x=-1	x=0	x=+1	x=+2	x=+3
1	1	1	1	1	1
0	0	1	1	1	0
0	1	1	1	1	0
1	1	1	1	1	1
1	0	1	0	0	0
0	1	0	0	0	0

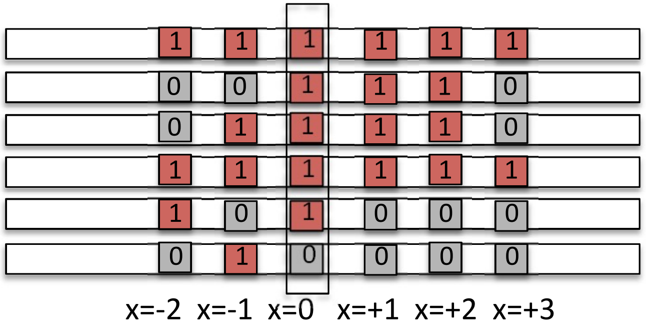
$$EHH_c(x_i) = \sum_{h \in H_c(x_i)} \frac{\binom{n_h}{2}}{\binom{n_c}{2}}$$



$$EHH_c(x_i = -1) = \frac{\binom{3}{2} + \binom{2}{2}}{\binom{5}{2}} = \frac{3+1}{10} = 0.4$$

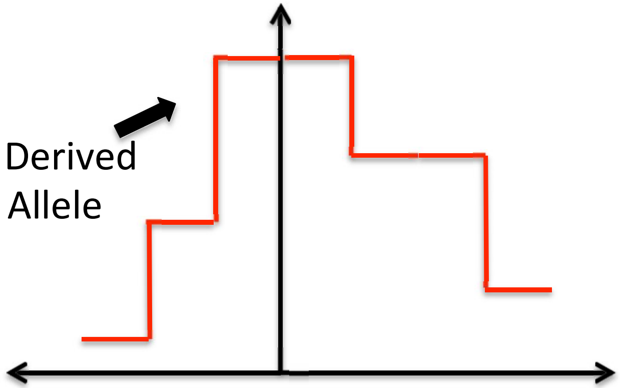
Adapted from Mateus Fumagalli.

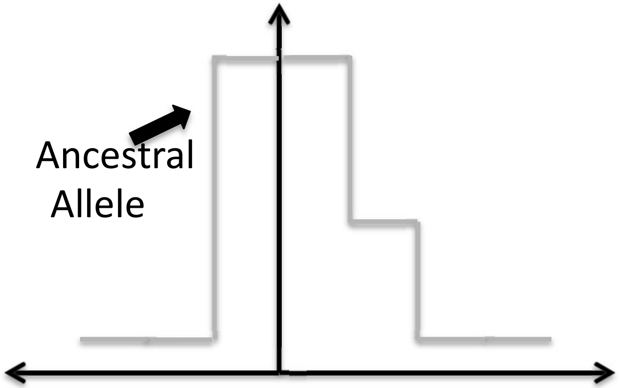
Extended Haplotype Homozygosity (EHH)



x=-2	x=-1	x=0	x=+1	x=+2	x=+3
1	1	1	1	1	1
0	0	1	1	1	0
0	1	1	1	1	0
1	1	1	1	1	1
1	0	1	0	0	0
0	1	0	0	0	0

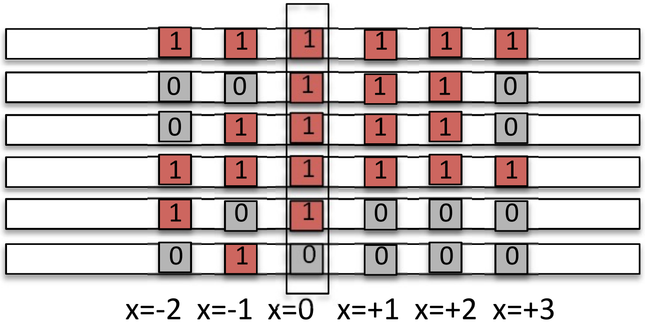
$$EHH_c(x_i) = \sum_{h \in H_c(x_i)} \frac{\binom{n_h}{2}}{\binom{n_c}{2}}$$





Adapted from Mateus Fumagalli.

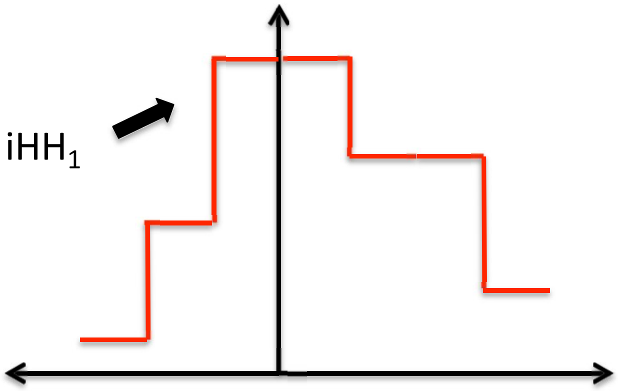
Integrated Haplotype Score (iHS)

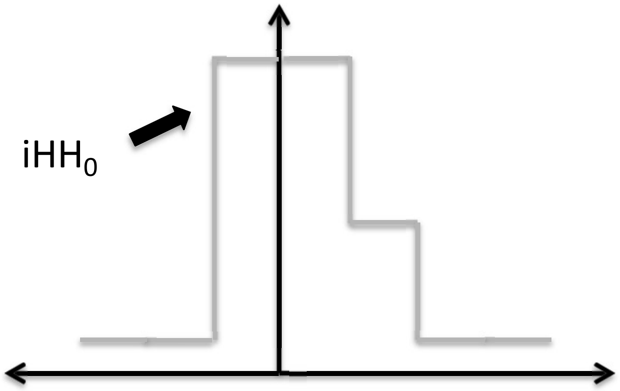


x=-2	x=-1	x=0	x=+1	x=+2	x=+3
1	1	1	1	1	1
0	0	1	1	1	0
0	1	1	1	1	0
1	1	1	1	1	1
1	0	1	0	0	0
0	1	0	0	0	0

Integrated Haplotype Homozigosity (iHH)

iHS= $\ln(iHH_1/iHH_0)$



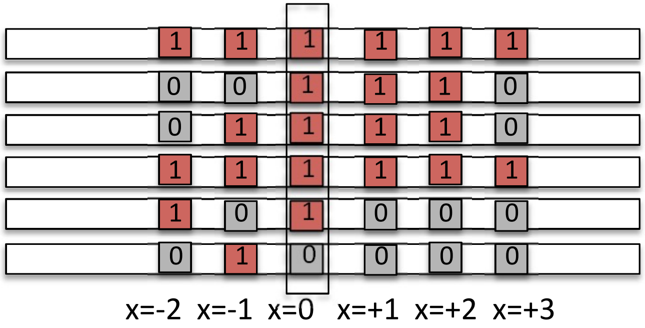


< -2 = derived allele

> 2 = ancestral allele

Adapted from Mateus Fumagalli.

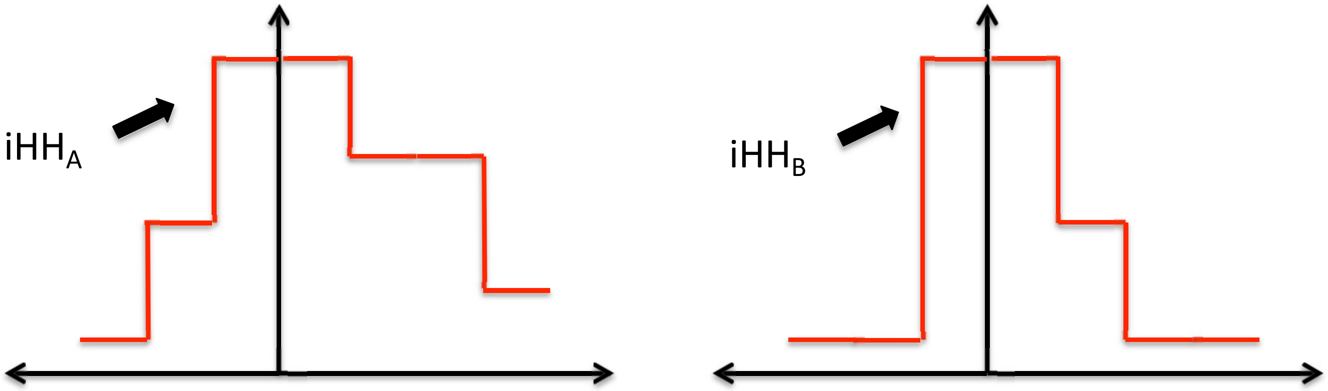
Cross Population Extended Homozygosity Haplotype (xpEHH)



x=-2	x=-1	x=0	x=+1	x=+2	x=+3
1	1	1	1	1	1
0	0	1	1	1	0
0	1	1	1	1	0
1	1	1	1	1	1
1	0	1	0	0	0
0	1	0	0	0	0

Integrated Haplotype Homozygosity (iHH) for A and B

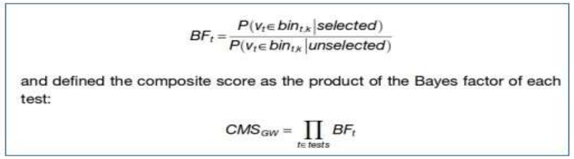
xpEHH= ln (iHH_A/iHH_B)



Adapted from Mateus Fumagalli.

Recent Advances to Detect Selection

1. Composite scores (Grossman et al. 2013)

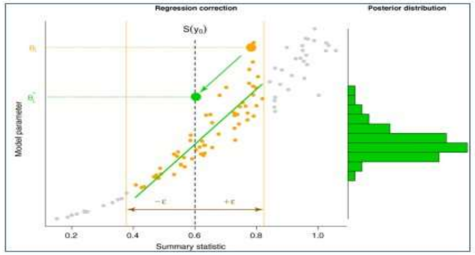


$$BF_t = \frac{P(v_t \in bin_{t,x} | selected)}{P(v_t \in bin_{t,x} | unselected)}$$

and defined the composite score as the product of the Bayes factor of each test:

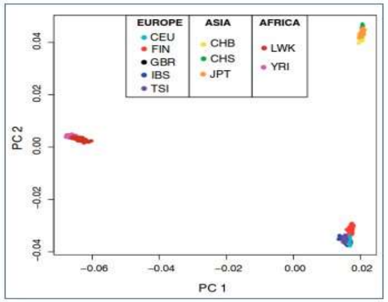
$$CMS_{GW} = \prod_{t \in tests} BF_t$$

2. Simulations-based (rejection, ABC)



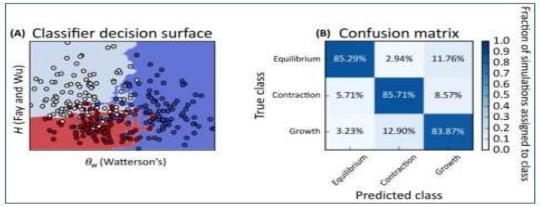
3. Unsupervised machine learning

(PCA, Duforet-Frebourg et al. 2016)



4. Supervised machine learning

(SVM, Schrider & Kern 2018)



Adapted from Mateus Fumagalli.

Thank you!