



Genetica di popolazione

Fabrizio Mafessoni

Universita' di Trieste, 18-11-2024



Sommario

- **Genetica di popolazione**
 - Come definire una popolazione
 - Diversita' nucleotidica
 - Equilibrio di Hardy-Weinberg
 - Principio di Wahlund
 - Varianza delle frequenze alleliche (*Fst* e *AMOVA*)
 - La dimensione effettiva di una popolazione *Ne*
 - Inbreeding, eterozigosita' e deriva genetica
 - Selezione naturale
 - Selezione positiva, negativa e bilanciata
 - Come studiare la selezione positiva a partire da dati genetici
 - Introggressione genetica



Cosa e' una popolazione?

In genetica di popolazione, una popolazione e' generalmente definita come un gruppo di individui della stessa specie che:

- vive in un'area geograficamente definite
- gli individui sono capace di interbreeding/riprodursi tra loro
- gli individui posseggono un pool genetico comune

Non e' sempre facile definire una popolazione

Segregazione temporale

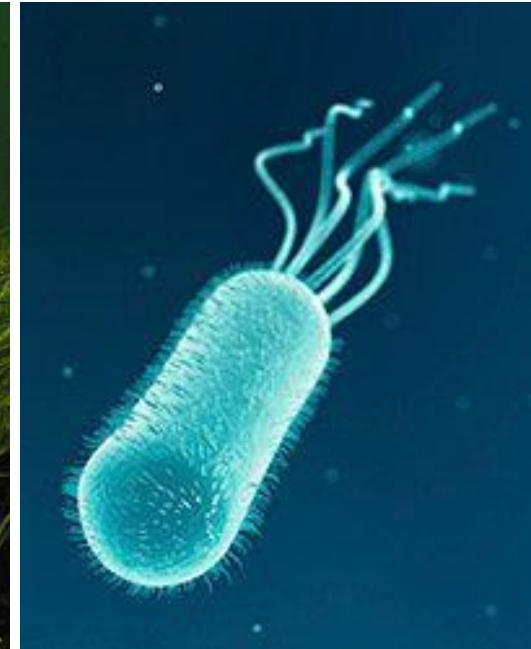


Anax junius

Organismi con riproduzione asexuale



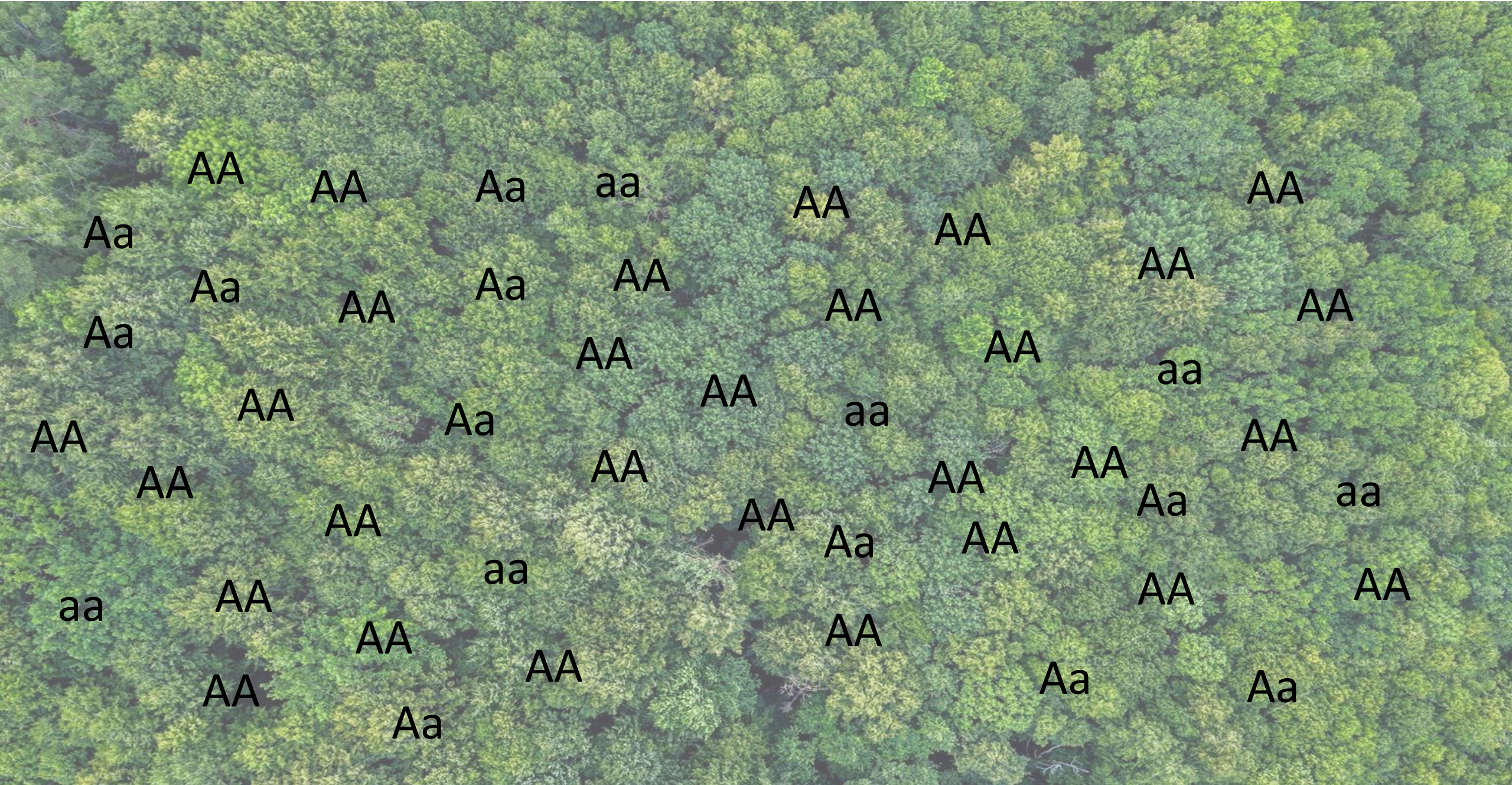
Zostera marina



Escherichia coli

Locus biallelico con due alleli: A e a

E' una popolazione?



AA

AA

Aa

aa

AA

AA

AA

Aa

Aa

AA

Aa

AA

AA

AA

AA

Aa

AA

Aa

AA

AA

aa

AA

aa

AA

AA

AA

AA

Aa

AA

AA

AA

aa

AA

aa

AA

AA

Aa

aa

aa

AA

AA

AA

AA

AA

AA

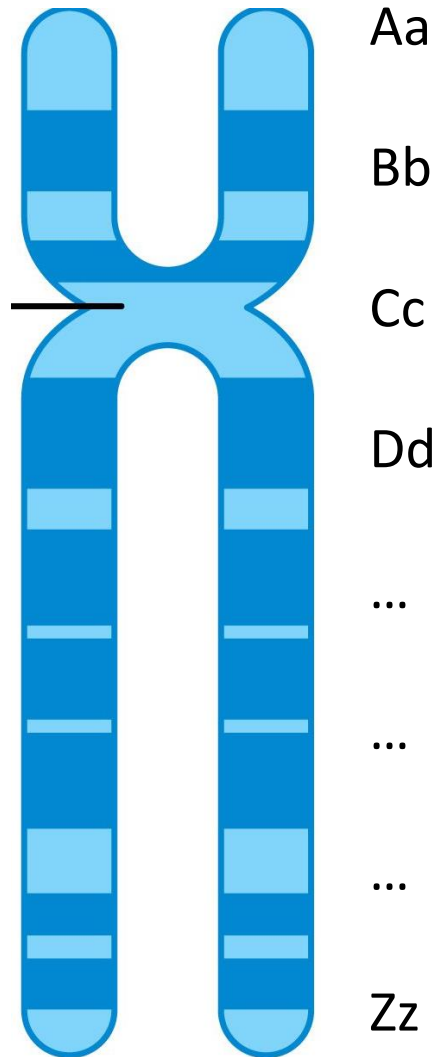
AA

Aa

Aa

Aa

Un genoma ha molti loci



Singolo locus

Eterozigosita'

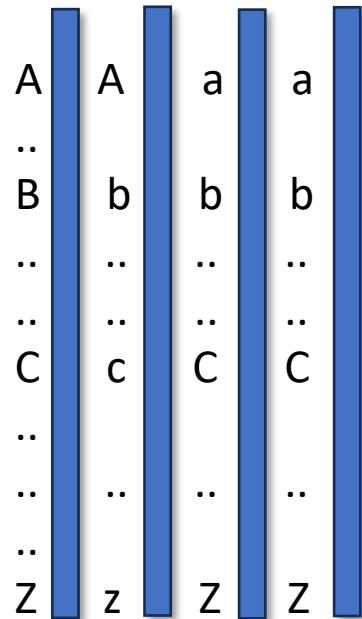
H

Multi-loci

Eterozigosita' media

$E[H]_{\text{loci}}$

Un genoma ha molti loci



Singolo locus

Eterozigosita'

H_0

Frequenza allelica

p

Multi-loci

Eterozigosita' media

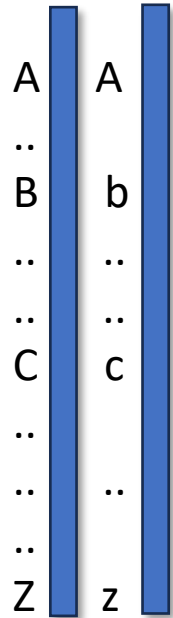
$E[H]_{\text{loci}}$

Diversita' nucleotidica

$$\sum p_i p_j \pi_{ij}$$

dove p_i e p_j sono le frequenze dei genomi i e j , e π_{ij} la divergenza tra loro. Rappresenta la distanza media tra sequenze in una popolazione.

Diversita' nucleotidica



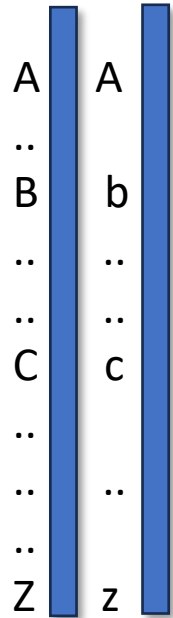
Diversita' nucleotidica

$$\pi = \sum p_i p_j \pi_{ij}$$

dove p_i e p_j sono le frequenze dei genomi i e j ,
e π_{ij} la divergenza tra loro. Rappresenta
la distanza media tra sequenze in una
popolazione.

Esempio: se abbiamo due sequenze con frequenza 10% e
90% e con distanza 1%, quale sara' la diversita'
nucleotidica?

Diversita' nucleotidica



Diversita' nucleotidica

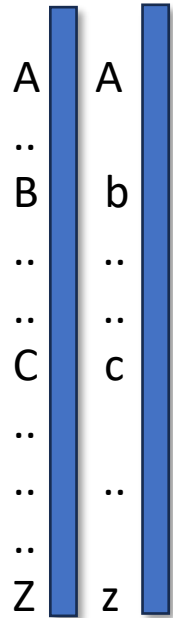
$$\pi = \sum p_i p_j \pi_{ij}$$

dove p_i e p_j sono le frequenze dei genomi i e j ,
e π_{ij} la divergenza tra loro. Rappresenta
la distanza media tra sequenze in una
popolazione.

Esempio: se abbiamo due sequenze con frequenza 10% e
90% e con distanza 1%, quale sara' la diversita'
nucleotidica?

A. ~0.001% B. ~0.01% C. ~0.1% D. ~10%

Diversita' nucleotidica



Diversita' nucleotidica

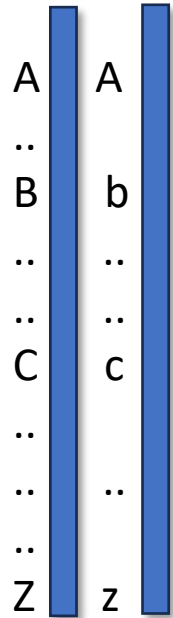
$$\pi = \sum p_i p_j \pi_{ij}$$

dove p_i e p_j sono le frequenze dei genomi i e j ,
e π_{ij} la divergenza tra loro. Rappresenta
la distanza media tra sequenze in una
popolazione.

Esempio: se abbiamo due sequenze con frequenza 10% e
90% e con distanza 1%, quale sara' la diversita'
nucleotidica?

C. ~0.1%: $\pi = 2 \times 0.10 \times 0.90 \times 0.01 = 0.0018$

Diversita' nucleotidica



Diversita' nucleotidica

$$\pi = \sum p_i p_j \pi_{ij}$$

dove p_i e p_j sono le frequenze dei genomi i e j ,
e π_{ij} la divergenza tra loro. Rappresenta
la distanza media tra sequenze in una
popolazione.

Puo' essere vista come la probabilita' che due nucleotide
presi a random differiscano in una popolazione.

Quali fattori fanno sì che una popolazione abbia alta o bassa diversità/eterozigosità?

Quali fattori fanno sì che una popolazione abbia alta o bassa diversità/eterozigosità?

- Deriva genetica
- Dimensione della popolazione di censo
- Presenza di selezione naturale
- Tasso di mutazione
- Separazione in sottopopolazioni
-

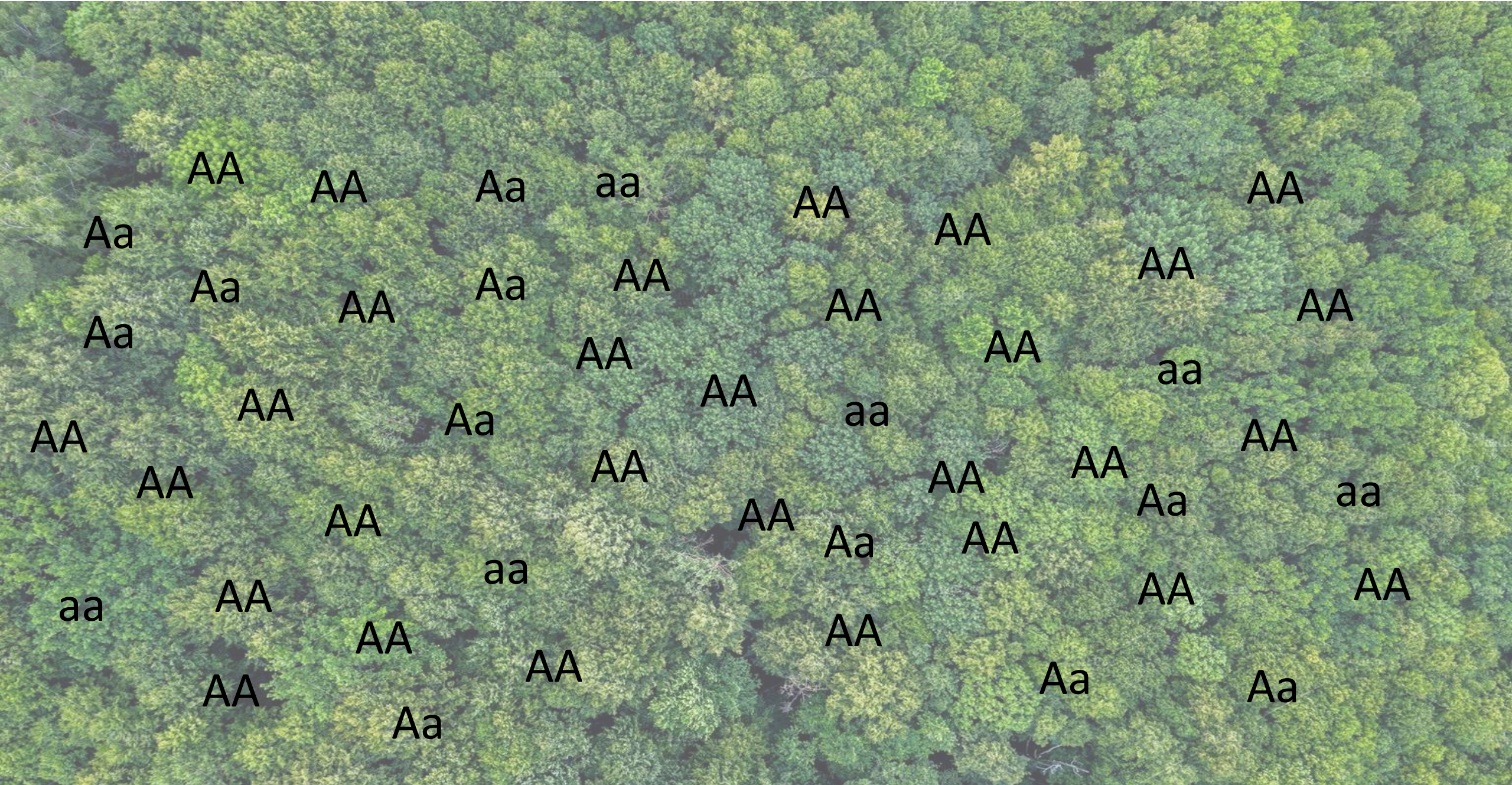
Cosa e' una popolazione?

In genetica di popolazione, una popolazione e' generalmente definita come un gruppo di individui della stessa specie che:

- vive in un'area geograficamente definite
- gli individui sono capace di interbreeding/riprodursi tra loro
- gli individui posseggono un pool genetico comune

Locus biallelico con due alleli: A e a

E' una popolazione?



AA

AA

Aa

aa

AA

AA

AA

Aa

Aa

AA

Aa

AA

AA

AA

AA

Aa

AA

Aa

AA

AA

aa

AA

aa

AA

AA

AA

AA

aa

AA

AA

AA

aa

AA

aa

AA

AA

Aa

AA

Aa

aa

aa

AA

AA

AA

AA

Aa

AA

AA

AA

AA

AA

Aa

AA

AA

Aa

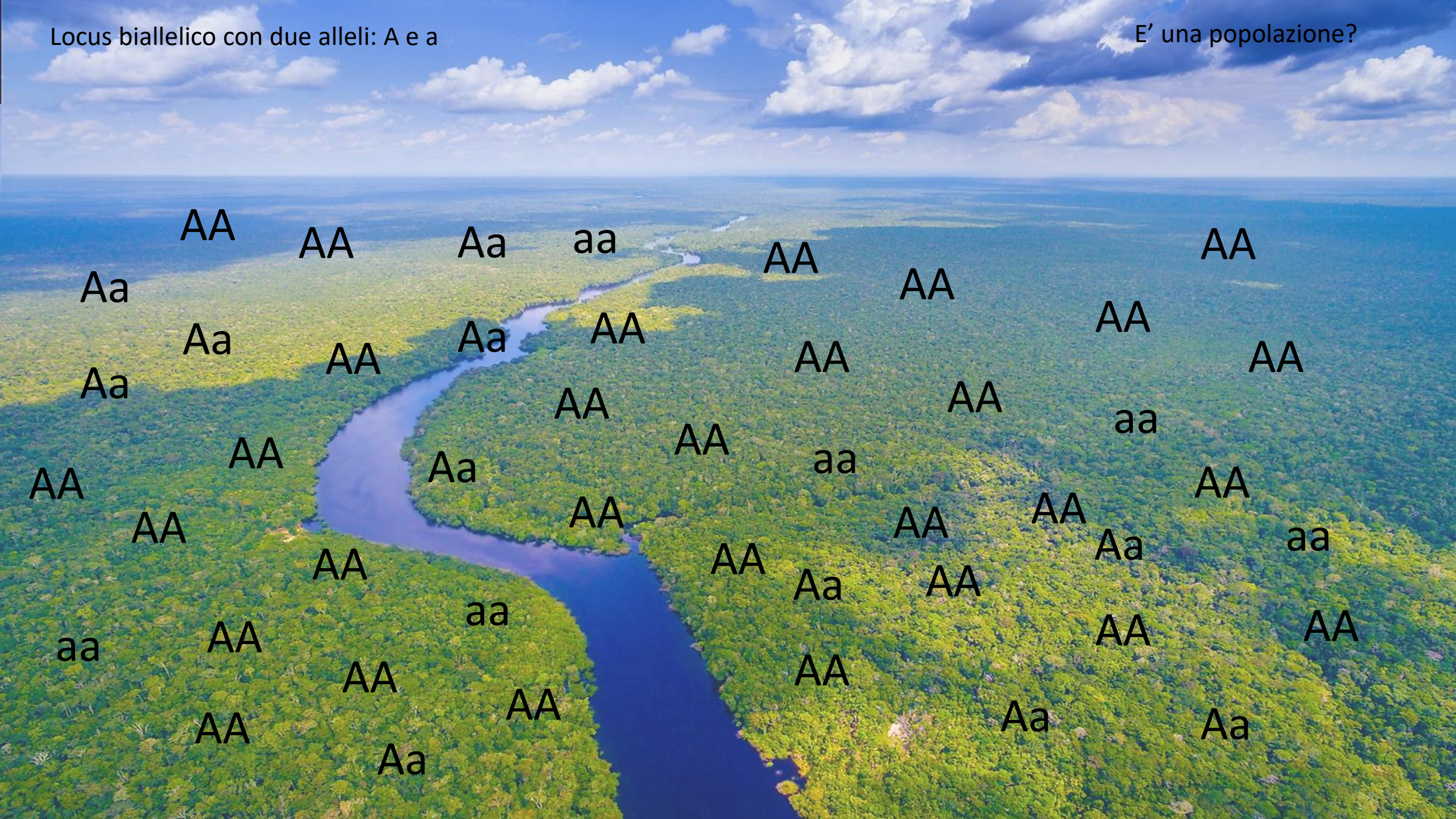
AA

Aa

Aa

Locus biallelico con due alleli: A e a

E' una popolazione?



AA

AA

Aa

aa

AA

AA

AA

Aa

Aa

AA

Aa

AA

AA

AA

AA

Aa

AA

AA

AA

AA

aa

aa

AA

Aa

AA

AA

AA

AA

AA

AA

aa

AA

AA

Aa

AA

Aa

aa

aa

AA

aa

AA

AA

AA

AA

AA

Aa

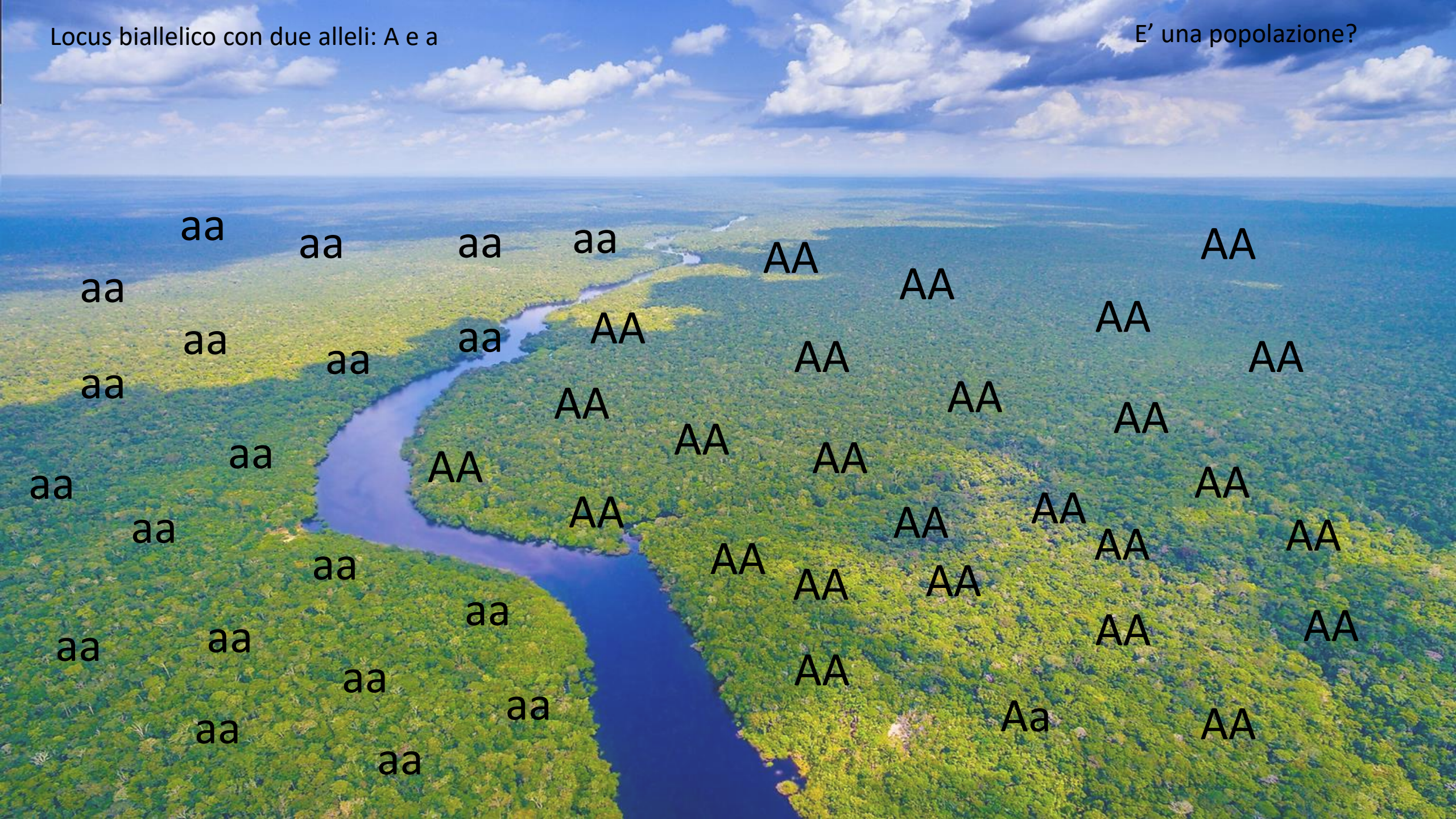
Aa

Aa

AA

Locus biallelico con due alleli: A e a

E' una popolazione?



aa

aa

aa

aa

AA

AA

aa

aa

aa

AA

AA

AA

AA

aa

aa

AA

AA

AA

AA

aa

aa

AA

AA

AA

aa

AA

AA

AA

AA

AA

AA

aa

AA

AA

AA

AA

aa

aa

aa

AA

Aa

AA

aa

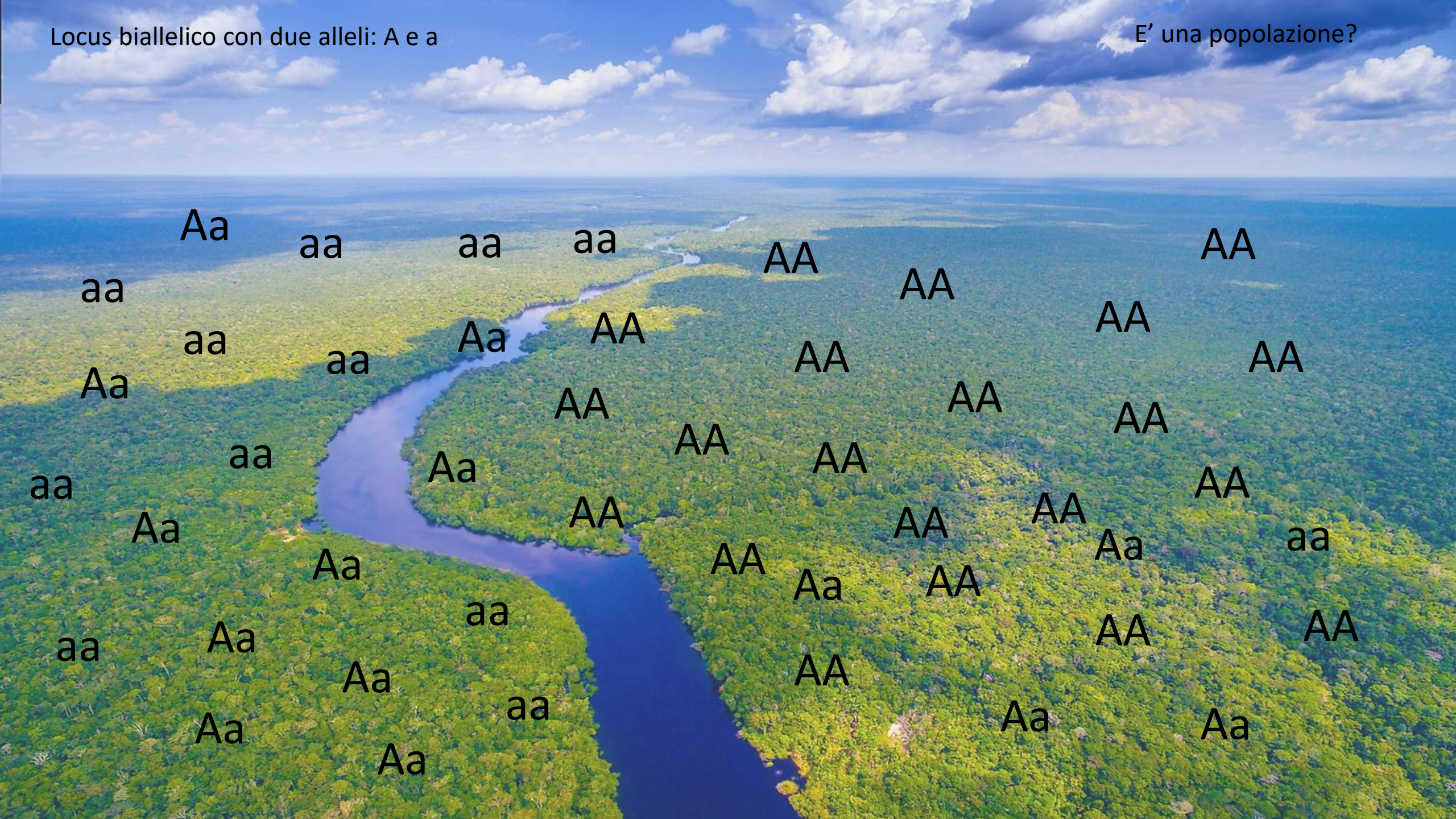
aa

aa

aa

Locus biallelico con due alleli: A e a

E' una popolazione?



Aa

aa

aa

aa

AA

AA

aa

aa

aa

Aa

AA

AA

AA

AA

Aa

aa

Aa

AA

AA

AA

AA

AA

AA

aa

Aa

Aa

AA

AA

Aa

AA

AA

Aa

aa

aa

Aa

aa

Aa

aa

AA

Aa

AA

AA

AA

Aa

Aa

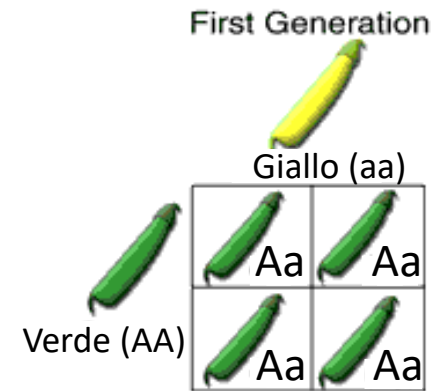
AA

Aa

Aa

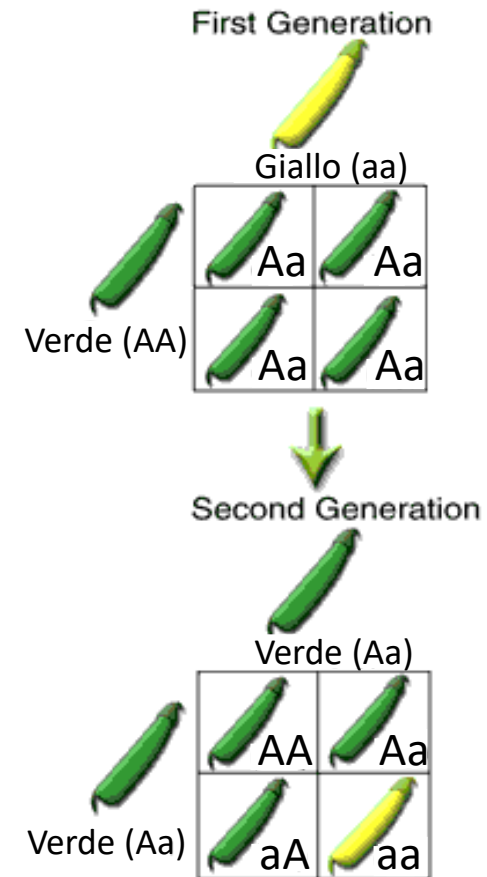
Come si puo' definire se un campione proviene da una popolazione o da piu popolazioni?

Mendel (1865)



Mendel (1865)

$$\left(\frac{1}{2} A + \frac{1}{2} a\right) \times \left(\frac{1}{2} A + \frac{1}{2} a\right) = \left(\frac{1}{2} A + \frac{1}{2} a\right)^2 \text{ ——— } \left(\frac{1}{4} AA + \frac{1}{2} Aa + \frac{1}{4} aa\right)$$



Mendel (1865)

$$(\frac{1}{2} A + \frac{1}{2} a) \times (\frac{1}{2} A + \frac{1}{2} a) = (\frac{1}{2} A + \frac{1}{2} a)^2 \quad \text{_____} \quad (\frac{1}{4} AA + \frac{1}{2} Aa + \frac{1}{4} aa)$$

Hardy&Weinberg (1908)

In una popolazione ideale*, **l'equilibrio delle frequenze alleliche e genotipiche rimane costante** di generazione in generazione. In altre parole, se nessuna di queste forze agisce, le frequenze alleliche p e q sono stabili e distribuite come:

$$p^2 + 2pq + q^2$$

dove:

- p^2 : frequenza degli omozigoti AA,
- $2pq$: frequenza degli eterozigoti Aa,
- q^2 : frequenza degli omozigoti aa.

*Popolazione ideale: popolazione con accoppiamento random e senza selezione, migrazione o deriva genetica.



Mendel (1865)

$$(\frac{1}{2} A + \frac{1}{2} a) \times (\frac{1}{2} A + \frac{1}{2} a) = (\frac{1}{2} A + \frac{1}{2} a)^2 \quad \text{_____} \quad (\frac{1}{4} AA + \frac{1}{2} Aa + \frac{1}{4} aa)$$

Hardy&Weinberg (1908)

In una popolazione ideale*, **l'equilibrio delle frequenze alleliche e genotipiche rimane costante** di generazione in generazione. In altre parole, se nessuna di queste forze agisce, le frequenze alleliche p e q sono stabili e distribuite come: ↓

$$p^2 + 2pq + q^2 = (p+q)^2$$

dove:

- p^2 : frequenza degli omozigoti AA,
- $2pq$: frequenza degli eterozigoti Aa,
- q^2 : frequenza degli omozigoti aa.

*Popolazione ideale: popolazione con accoppiamento random e senza selezione, migrazione o deriva genetica.



Mendel (1865)

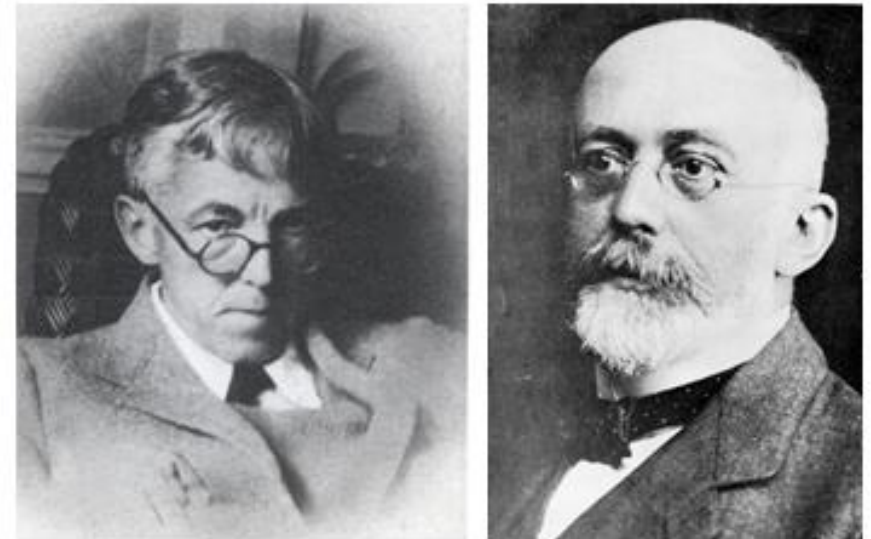
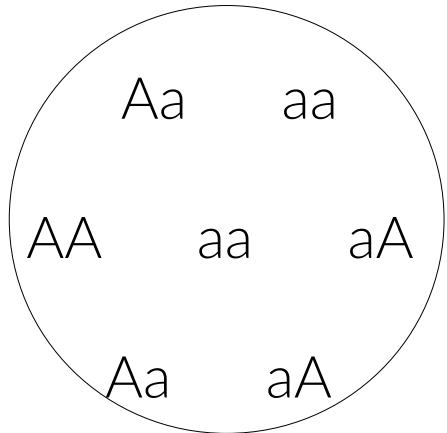
$$(\frac{1}{2} A + \frac{1}{2} a) \times (\frac{1}{2} A + \frac{1}{2} a) = (\frac{1}{2} A + \frac{1}{2} a)^2 \text{ ——— } (\frac{1}{4} AA + \frac{1}{2} Aa + \frac{1}{4} aa)$$

Hardy & Weinberg (1908)

p : frequency of A
 $q=(1-p)$: frequency of a

$f(AA)$ =freq. of AA
 $f(Aa)$ =freq. of Aa and aA
 $f(aa)$ =freq. of aa

$$p=f(AA) + \frac{1}{2} f(Aa)$$



Mendel (1865)

$$\left(\frac{1}{2} A + \frac{1}{2} a\right) \times \left(\frac{1}{2} A + \frac{1}{2} a\right) = \left(\frac{1}{2} A + \frac{1}{2} a\right)^2 \text{ ——— } \left(\frac{1}{4} AA + \frac{1}{2} Aa + \frac{1}{4} aa\right)$$

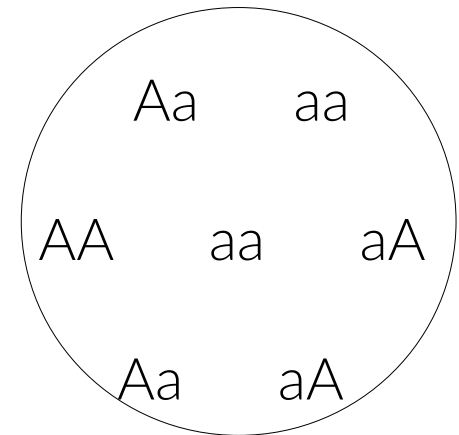
Hardy & Weinberg (1908)

p : frequency of A
 $q=(1-p)$: frequency of a

$f(AA)$ =freq. of AA
 $f(Aa)$ =freq. of Aa and aA
 $f(aa)$ =freq. of aa



$$f_{t+1}(AA) = (f(AA) + f(Aa)/2)^2$$
$$f_{t+1}(Aa) = f(AA) + f(Aa)/2)(f(aa) + f(Aa)/2)$$
$$f_{t+1}(aa) = (f(aa) + f(Aa)/2)^2$$



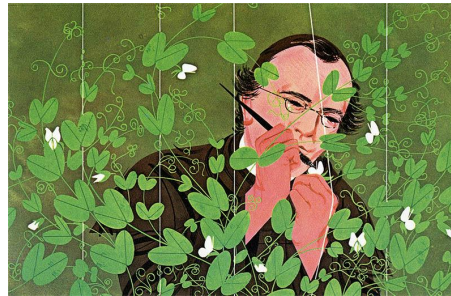
Mendel (1865)

$$(\frac{1}{2} A + \frac{1}{2} a) \times (\frac{1}{2} A + \frac{1}{2} a) = (\frac{1}{2} A + \frac{1}{2} a)^2 \text{ ——— } (\frac{1}{4} AA + \frac{1}{2} Aa + \frac{1}{4} aa)$$

Hardy & Weinberg (1908)

$$(p+q)^2 = \begin{matrix} p^2 & + & 2pq & + & q^2 \\ AA & Aa & & aa \end{matrix}$$

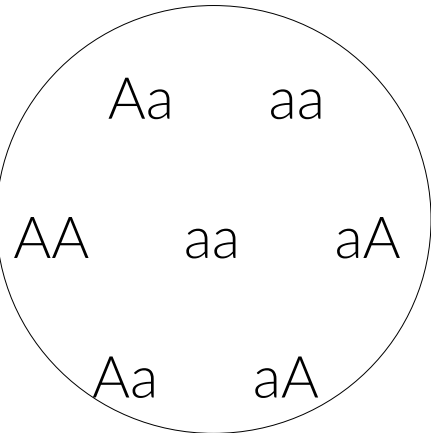
p: frequency of A
q=(1-p) : frequency of a



$$p = f(AA) + \frac{1}{2} f(Aa)$$

$$\begin{aligned} f_{t+1}(AA) &= (f(AA) + f(Aa)/2)^2 \\ f_{t+1}(Aa) &= f(AA) + f(Aa)/2)(f(aa) + f(Aa)/2) \\ f_{t+1}(aa) &= (f(aa) + f(Aa)/2)^2 \end{aligned}$$

f(AA)=freq. of AA
f(Aa)=freq. of Aa and aA
f(aa)=freq. of aa



Testare l'equilibrio di Hardy&Weinberg (HWE)

$$(p+q)^2 = \underset{AA}{p^2} + \underset{Aa}{2pq} + \underset{aa}{q^2}$$

L'equilibrio di Hardy-Weinberg ci da una predizione teorica (ipotesi nulla) per testare se gli alleli in un una popolazione rispettano quelli di una singola popolazione nella quale gli individui si accoppino tra loro in modo random

(e in condizioni di assenza di selezione, alti tassi di mutazione e deriva genetica – nota che questi fenomeni devono essere molto forti per portare a deviazioni important dall'equilibrio).

Come lo fareste?

Testare l'equilibrio di Hardy&Weinberg (HWE)

$$(p+q)^2 =$$

p^2	$+ 2pq$	$+q^2$	Genotipi Previsti da HWE
AA	Aa	aa	Genotipi Osservati

Tabella di contingenza

L'equilibrio di Hardy-Weinberg ci da una predizione teorica (ipotesi nulla) per testare se gli alleli in un una popolazione rispettano quelli di una singola popolazione nella quale gli individui si accoppino tra loro in modo random

(e in condizioni di assenza di selezione, alti tassi di mutazione e deriva genetica – nota che questi fenomeni devono essere molto forti per portare a deviazioni important dall'equilibrio).

Come lo fareste?

Test di Pearson per deviazioni dal χ^2 !

Box 3.2 Calculating Hardy–Weinberg equilibrium

[Table 3.1](#) is an actual data set on scarlet tiger moths that was collected by the geneticist E. B. Ford.

The data in [Table 3.1](#) tell us that in this sample there is a total of $2(1612) = 3224$ alleles at this particular locus. Of these, 3076 are A alleles ($2938 + 138$), and 148 are a alleles ($138 + 10$). Therefore, the frequency p of the A allele in this population is:

$$p = \frac{3076}{3224} = 0.954$$

and the frequency q of the a allele can be calculated as either:

$$q = \frac{148}{3224} = 0.046$$

or, because $p + q = 1$, can alternatively be calculated as:

$$q = 1 - p = 1 - 0.954 = 0.046$$

If we know p and q , then we can calculate the frequencies of AA ($= p^2$), Aa ($= 2pq$) and aa ($= q^2$) that would be expected if the population is in HWE as follows:

$$p^2 = (0.954)^2 = 0.9101$$

$$2pq = 2(0.954)(0.046) = 0.0878$$

$$q^2 = (0.046)^2 = 0.002$$

We now need to calculate the number of moths in this population that would have each genotype if this population is in HWE. We can do this by multiplying the total number of moths (1612) by each genotype frequency:

$$AA = (0.9101)(1612) = 1467$$

$$Aa = (0.0878)(1612) = 142$$

$$aa = (0.002)(1612) = 3$$

Therefore the Hardy–Weinberg ratio expressed as the number of individuals with each genotype is 1467: 142: 3. This is very close to the actual ratio of genotypes within the population (from [Table 3.1](#)) of 1469 :138 : 5.

We can check whether or not there is a significant difference between the observed and expected genotype frequencies by using a chi-square (χ^2) test. This is based on the difference between the observed (O) number of genotypes, and the number that would be expected (E) under HWE, and is calculated as:

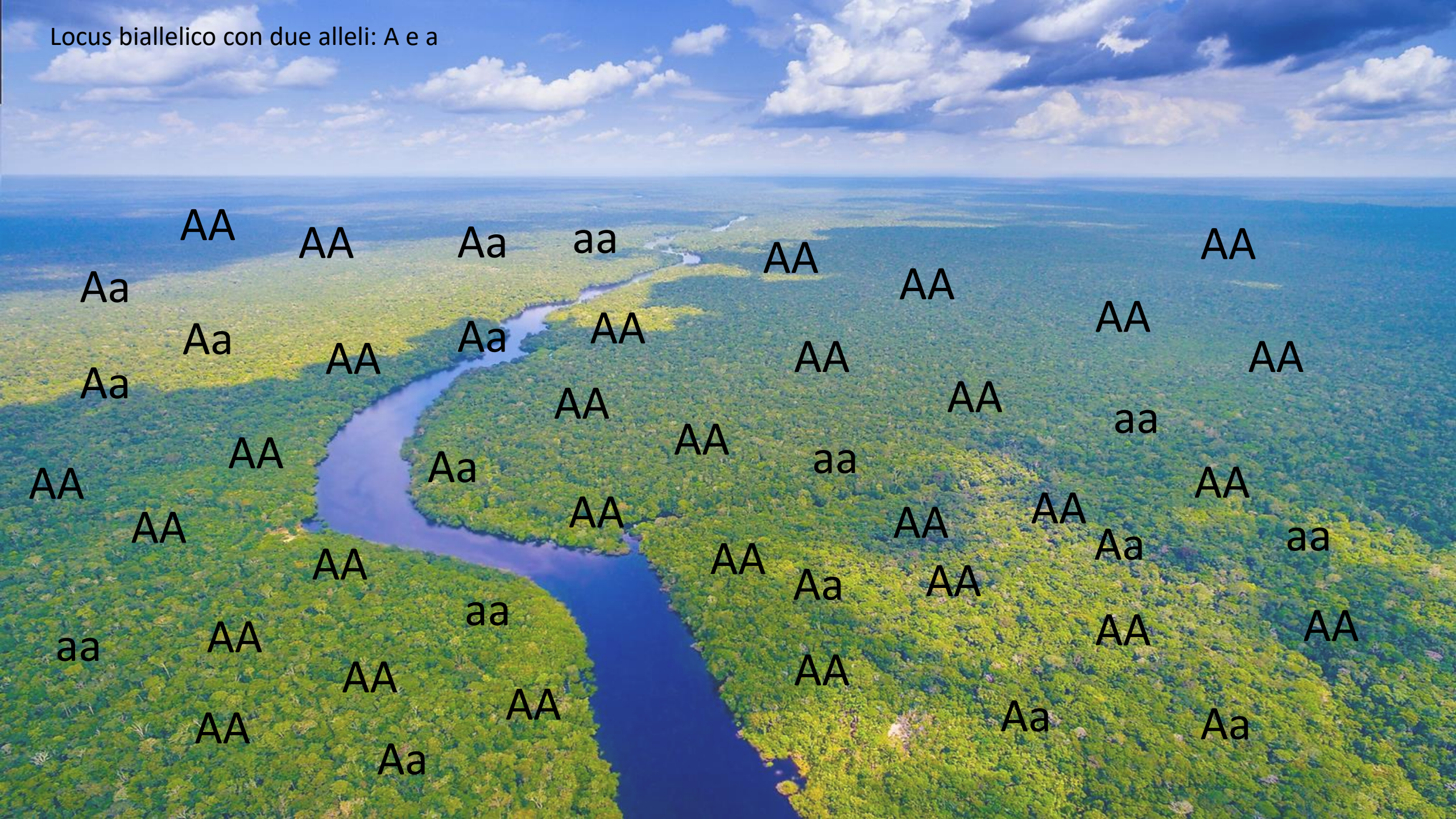
$$(3.5) \quad \chi^2 = \sum \frac{(O - E)^2}{E}$$

The χ^2 value of the scarlet tiger moth example is:

$$\begin{aligned} \chi^2 &= \frac{(1469 - 1467)^2}{1467} + \frac{(138 - 142)^2}{142} + \frac{(5 - 3)^2}{3} \\ &= 0.0027 + 0.11 + 1.33 \\ &= 1.44 \end{aligned}$$

The number of degrees of freedom (df) is determined as 3 (the number of genotypes) minus 1 (because if we know two of the expected genotype frequencies we automatically know the third) minus 1 (the number of independent values we calculated from our observed data to determine our expected values), which leaves $df = 1$. By using a statistical table, we learn that a χ^2 value of 1.44, in conjunction with one df, leaves us with a probability of $P = 0.230$. This means that there is no significant difference between the observed genotype frequencies in the scarlet tiger moth population and the genotype frequencies that are expected under HWE. We would therefore conclude that this population is in HWE.

Locus biallelico con due alleli: A e a



AA

AA

Aa

aa

AA

AA

AA

Aa

Aa

AA

Aa

AA

AA

AA

AA

Aa

AA

AA

AA

Aa

AA

aa

aa

AA

AA

AA

AA

AA

AA

aa

AA

AA

Aa

AA

Aa

aa

aa

AA

aa

AA

AA

AA

AA

AA

AA

AA

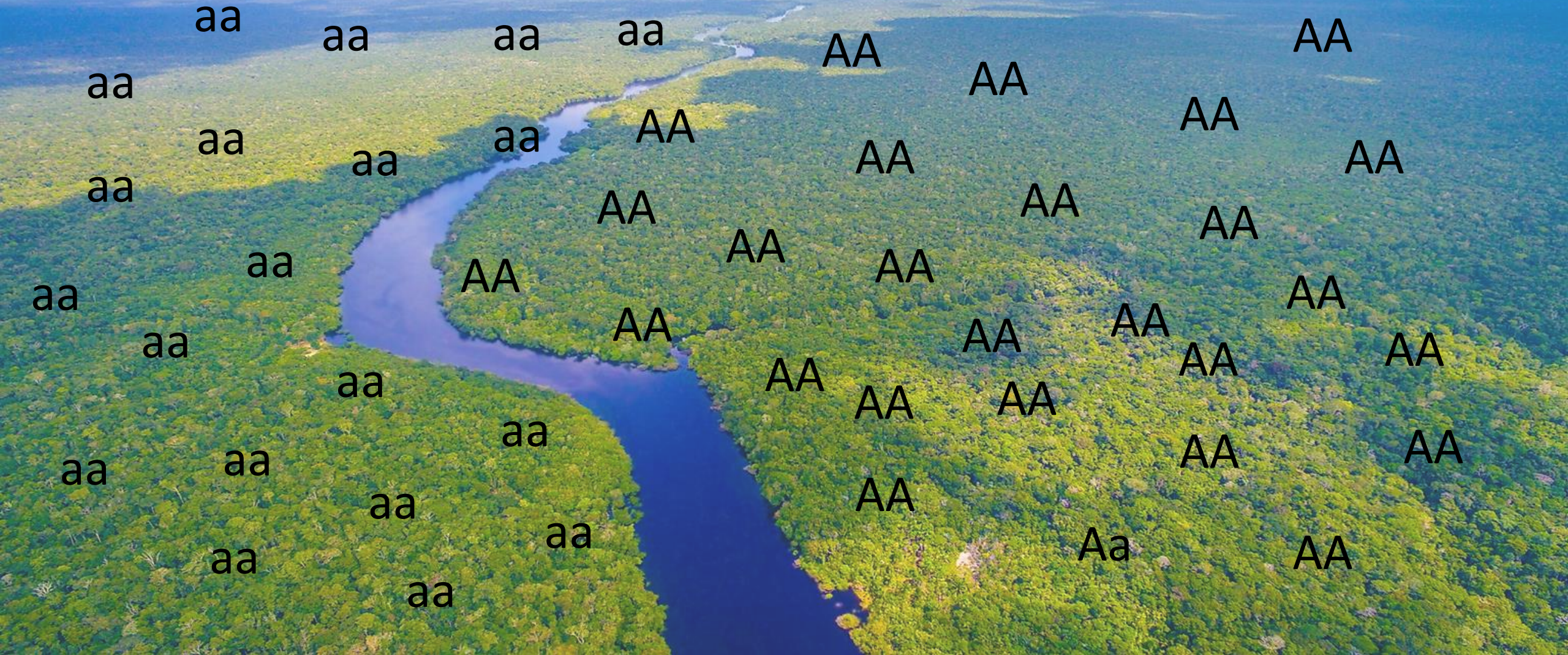
Aa

Aa

Aa

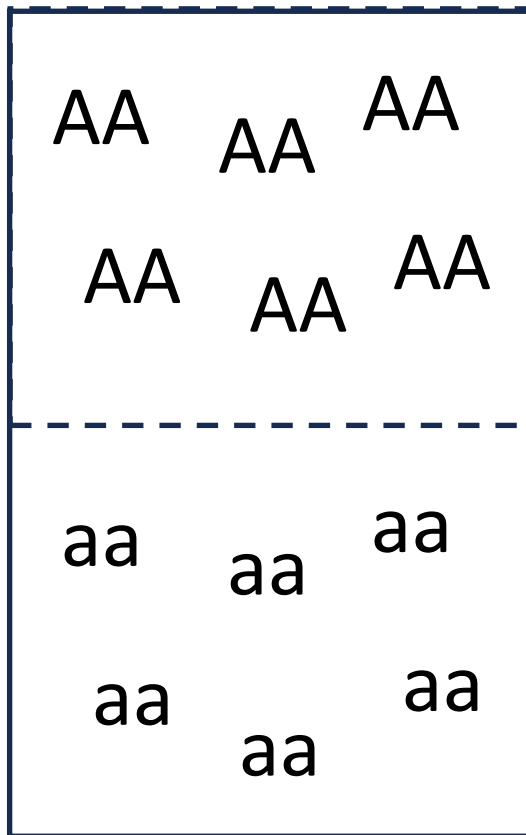
Locus biallelico con due alleli: A e a

Cosa differisce maggiormente nelle proporzioni dei differenti genotipi?



Principio di Wahlund

Popolazioni suddivise mostrano un eccesso di omozigoti/mancanza di eterozigoti rispetto a popolazioni non suddivise (in HWE).



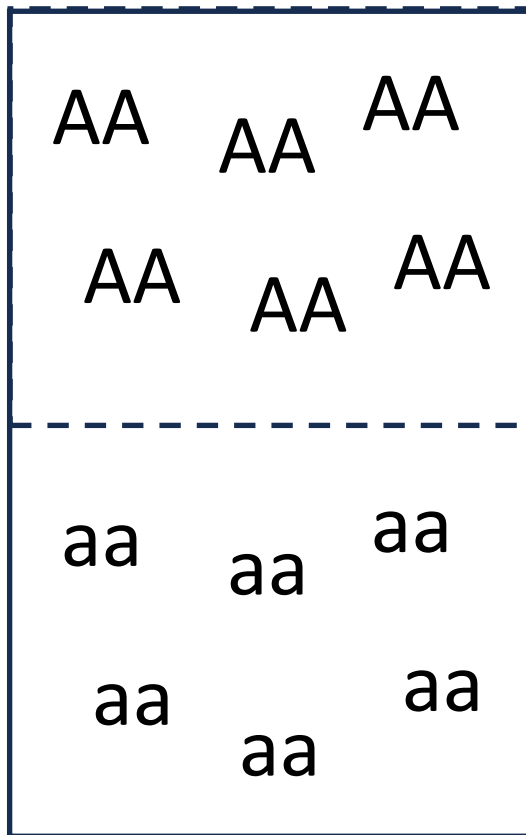
H: eterozigosita' (proporzione di genotipi eterozigosi)

H_o: eterozigosita' osservata: 0% nell'esempio

H_e: eterozigosita' attesa: $2 \times 0.5 \times 0.5 = 50\%$ nell'esempio

Principio di Wahlund

Popolazioni suddivise mostrano un eccesso di omozigoti/mancanza di eterozigoti rispetto a popolazioni non suddivise (in HWE).



H: eterozigosita' (proporzione di genotipi eterozigosi)

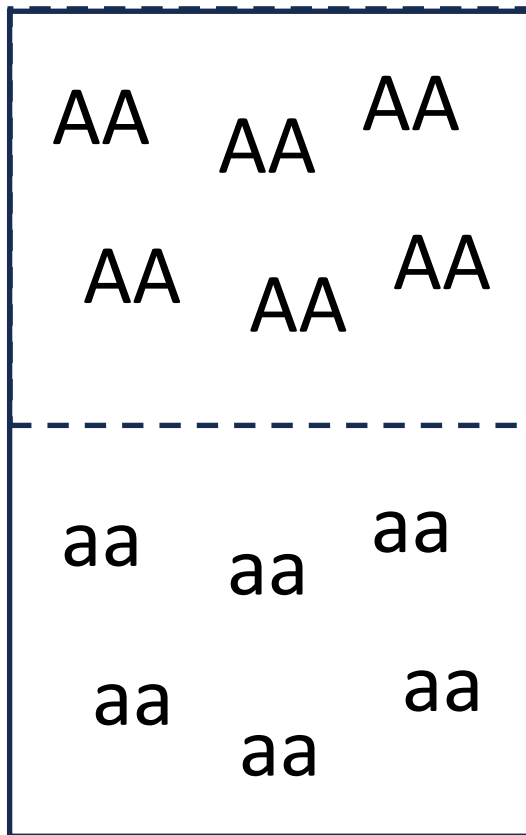
H_o: eterozigosita' osservata: 0% nell'esempio

H_e: eterozigosita' attesa: $2 \times 0.5 \times 0.5 = 50\%$ nell'esempio

Questa osservazione e' alla base di una delle piu importanti statistiche nella genetica di popolazione, l'**F_{st}**.

Principio di Wahlund

Popolazioni suddivise mostrano un eccesso di omozigoti/mancanza di eterozigoti rispetto a popolazioni non suddivise (in HWE).



H: eterozigosita' (proporzione di genotipi eterozigosi)

H_o: eterozigosita' osservata: 0% nell'esempio

H_e: eterozigosita' attesa: $2 \times 0.5 \times 0.5 = 50\%$ nell'esempio

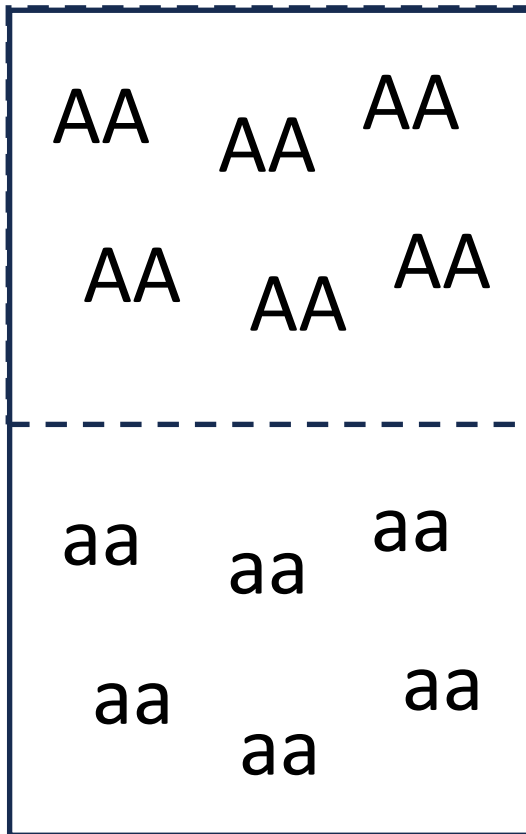
Per loci non biallelici l' *eterozigosita' attesa* e' denominata **gene diversity** e calcolata come sommando su tutti gli m alleli.

$$h = 1 - \sum_{i=1}^m x_i^2$$

Questa terminologia e' spesso usata per organismi non diploidi o marcatori particolari.

Principio di Wahlund

Popolazioni suddivise mostrano un eccesso di omozigoti/mancanza di eterozigoti rispetto a popolazioni non suddivise (in HWE).



H_t : eterozigosita' totale: proporzione di genotipi eterozigoti attesa

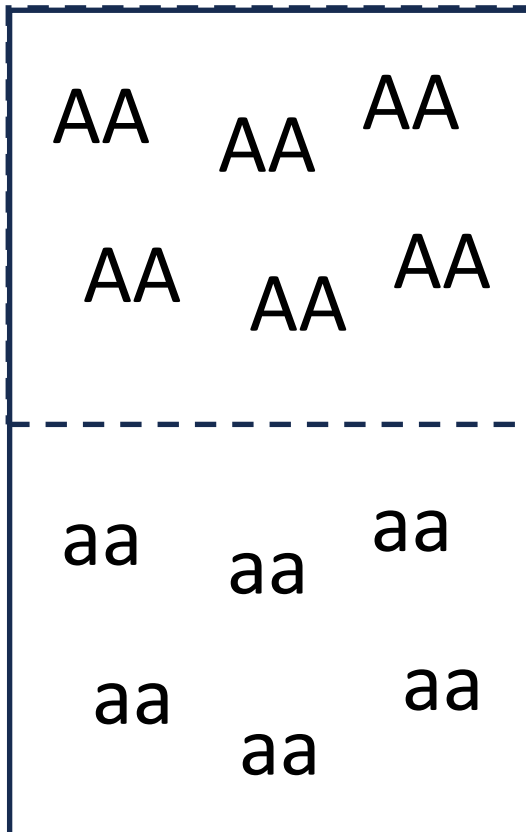
$$H_t = 2p(1 - p)$$

H_s : eterozigosita' delle subpopolazioni: proporzione di genotipi eterozigoti media attesa nelle sottopopolazioni k

$$H_s^{(k)} = 2p_k(1 - p_k)$$

Nota matematica:

La varianza nel campionare un allele
(vedi https://it.wikipedia.org/wiki/Distribuzione_binomiale)
e' $p(1-p)$



H_t : eterozigosita' totale: proporzione di genotipi eterozigoti attesa

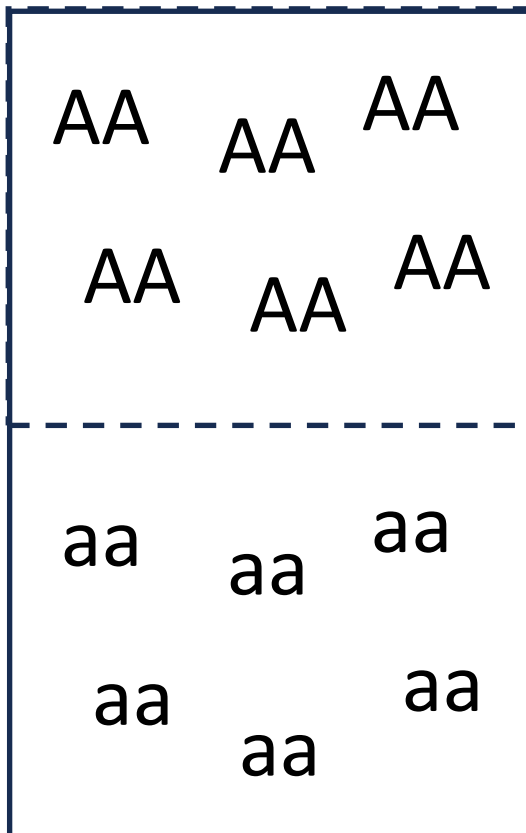
$$H_t = 2p(1 - p)$$

H_s : eterozigosita' delle subpopolazioni: proporzione di genotipi eterozigoti media attesa nelle sottopopolazioni k

$$H_s^{(k)} = 2p_k(1 - p_k)$$

Nota matematica:

La varianza nel campionare un allele
(vedi https://it.wikipedia.org/wiki/Distribuzione_binomiale)
e' $p(1-p)$



H_t : eterozigosita' totale: proporzione di genotipi eterozigoti attesa

$H_t = 2p(1 - p)$ \longrightarrow ~ varianza totale degli alleli

H_s : eterozigosita' delle subpopolazioni: proporzione di genotipi eterozigoti media attesa nelle sottopopolazioni k

$H_s^{(k)} = 2p_k(1 - p_k)$ \longrightarrow ~ varianza degli alleli nelle sottopopolazioni

L'eterozigosita' permette di stimare la varianza allelica dovuta alla:

- suddivisione tra popolazioni (F_{ST})

$$F_{ST} = \frac{\text{varianza tra sottopopolazioni}}{\text{varianza totale}} = \frac{H_T - H_S}{H_T}$$

$F_{ST}=0$: nessuna suddivisione tra sottopop.

$F_{ST}>0.1$ forte suddivisione tra sottopop.

$F_{ST}=1$: fortissima suddivisione tra sottopop.

L'eterozigosita' permette di stimare la varianza allelica dovuta alla:

- suddivisione tra popolazioni (F_{ST})

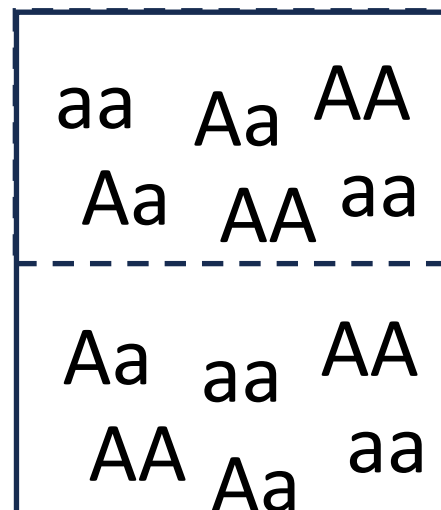
$$F_{ST} = \frac{\text{varianza tra sottopopolazioni}}{\text{varianza totale}} = \frac{H_T - H_S}{H_T}$$

$F_{ST}=0$: nessuna suddivisione tra sottopop.

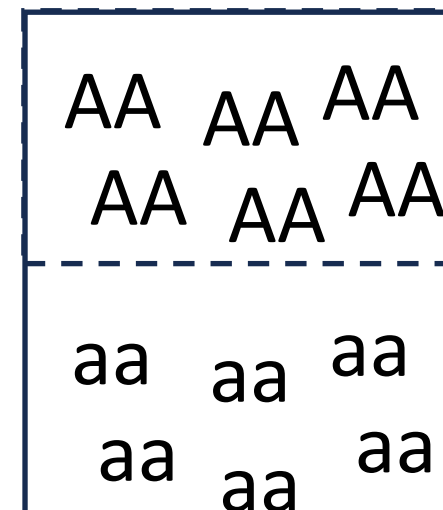
$F_{ST}>0.1$ forte suddivisione tra sottopop.

$F_{ST}=1$: fortissima suddivisione tra sottopop.

Puo' anche essere vista intuitivamente come una "distanza" tra popolazioni basata sulle frequenze alleliche:



$F_{ST} \sim 0$



$F_{ST} \sim 1$

L'eterozigosita' permette di stimare la varianza allelica dovuta alla:

- suddivisione tra popolazioni (F_{ST})

$$F_{ST} = \frac{\text{varianza tra sottopopolazioni}}{\text{varianza totale}} = \frac{H_T - H_S}{H_T}$$

$F_{ST}=0$: nessuna suddivisione tra sottopop.
 $F_{ST}>0.1$ forte suddivisione tra sottopop.
 $F_{ST}=1$: fortissima suddivisione tra sottopop.

Puo' anche essere vista intuitivamente come una "distanza" tra popolazioni basata sulle frequenze alleliche:

F_{ST} *La misura piu' comune.*

Wright (1931)

G_{ST} *Quasi equivalente a F_{ST} . Solo per siti biallelici.*

Nei (1973)

F_{ST} di W&C *Analogo a F_{ST} ma robusto a campioni sbilanciati.*

Weir&Cokerham (1984)

Nei's D *Usato generalmente come distanza genetica.*

Nei (1972)

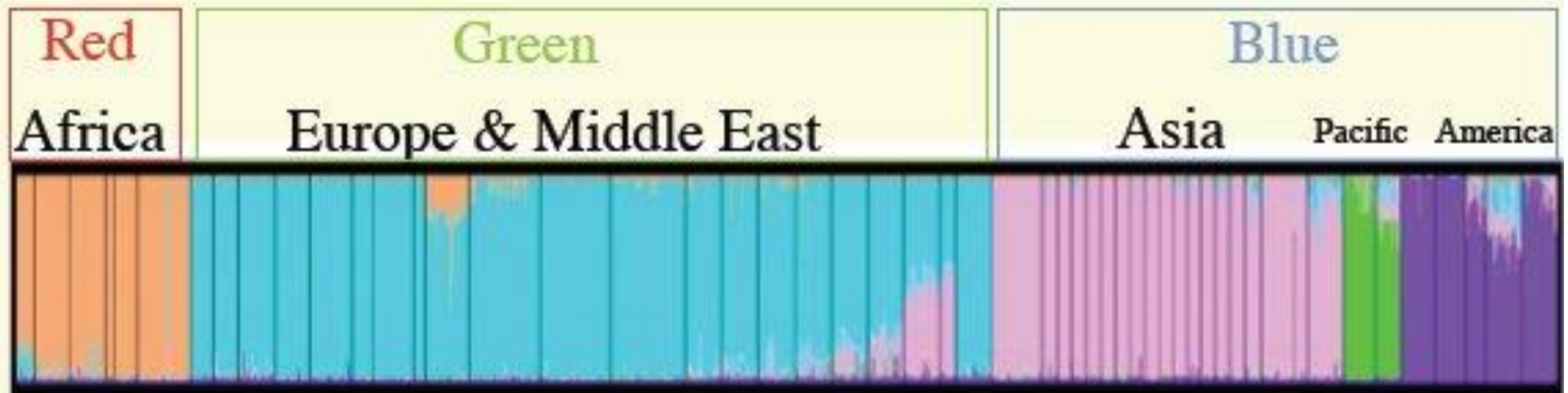
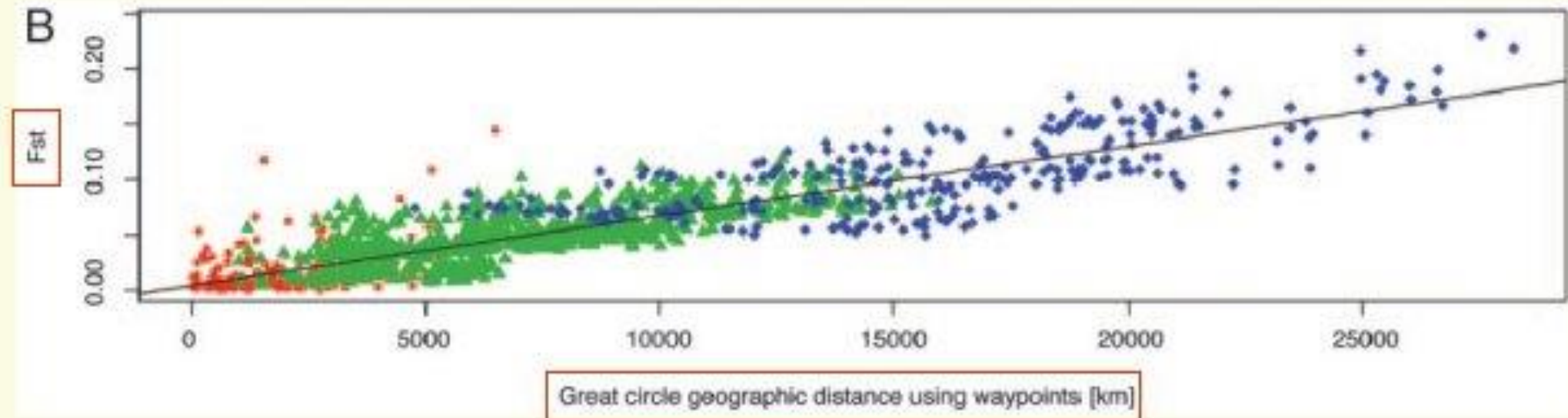
Jost's D *Usato per alleli molto polimorfici.*

Jost (2008)

R_{ST} *Modello mutazionale step-wise per microsatellite*

Slatkin (1995)

Isolamento per distanza



Rosenberg
et al.
Science
298:2381-
2385, 2002

L'Fst rivela differenze tra popolazioni di persico trota

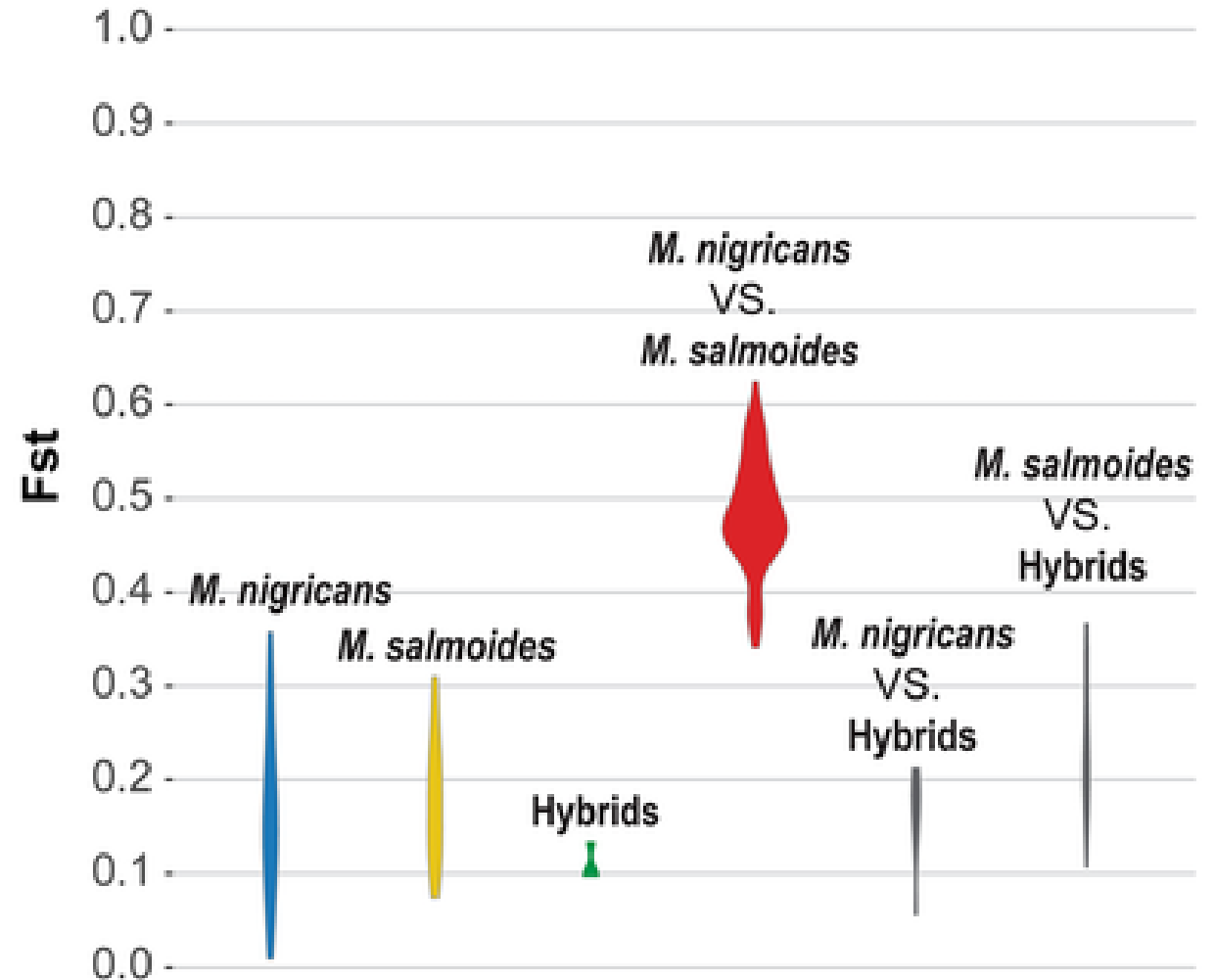
C Pairwise Fst



Micropterus nigricans



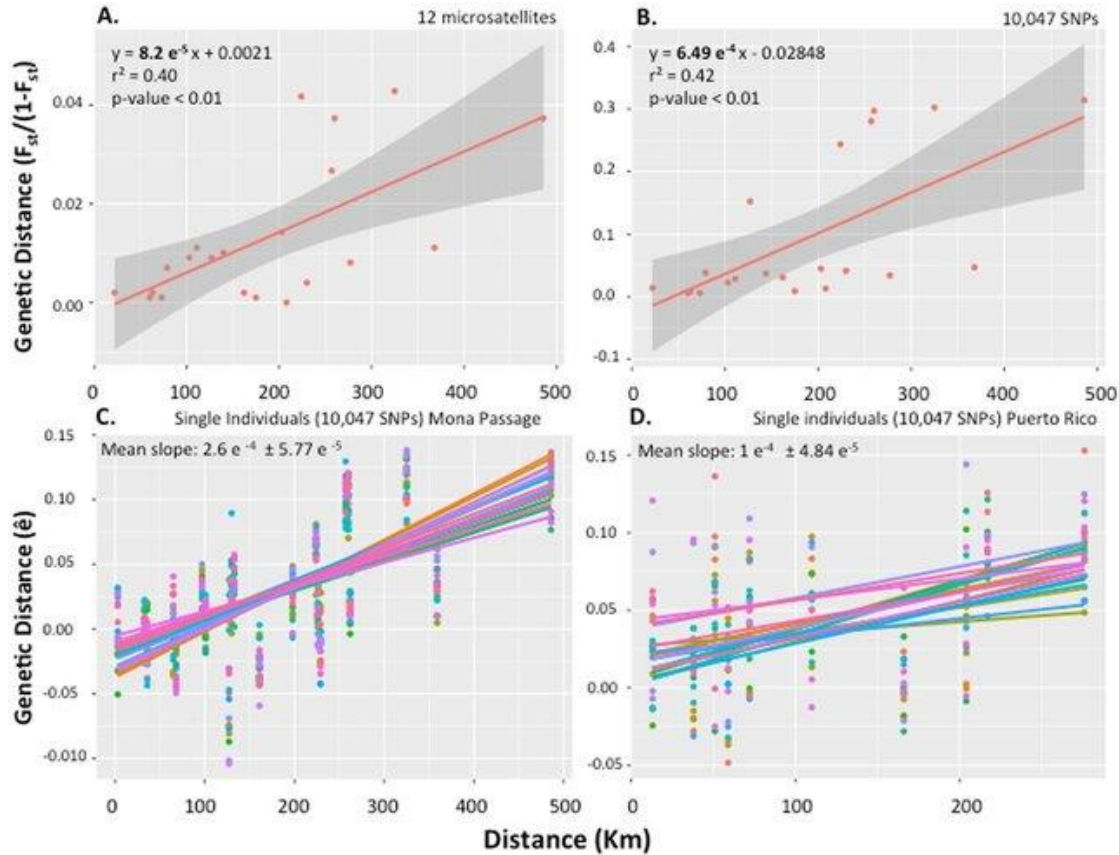
Micropterus salmoides



Effective Dispersal of Caribbean Reef Fish is Smaller than Current Spacing Among Marine Protected Areas

Received: 3 February 2017
 Accepted: 18 May 2017
 Published online: 05 July 2017

Diana M. Beltrán¹, Nikolaos V. Schizas¹, Richard S. Appeldoorn¹ & Carlos Prada²



Pesce minatore
Opistognathus aurifrons

	Tres Palmas	La Parguera	PNNE	Mona Island	La Caleta	Desecheo	Culebra
Tres Palmas	—	0.002	0.002	0.001	0.026	0.002	0.005
Parguera	0.009	—	0.014	0.011	0.041	0.007	0.001
PNNE	0.03	0.043	—	0.009	0.009	0.010	0.011
Mona Island	0.006	0.028	0.022	—	0.040	0.001	0.008
La Caleta	0.219	0.232	0.132	0.196	—	0.036	0.036
Desecheo	0.014	0.037	0.036	0.006	0.229	—	0.004
Culebra	0.013	0.009	0.045	0.033	0.239	0.04	—

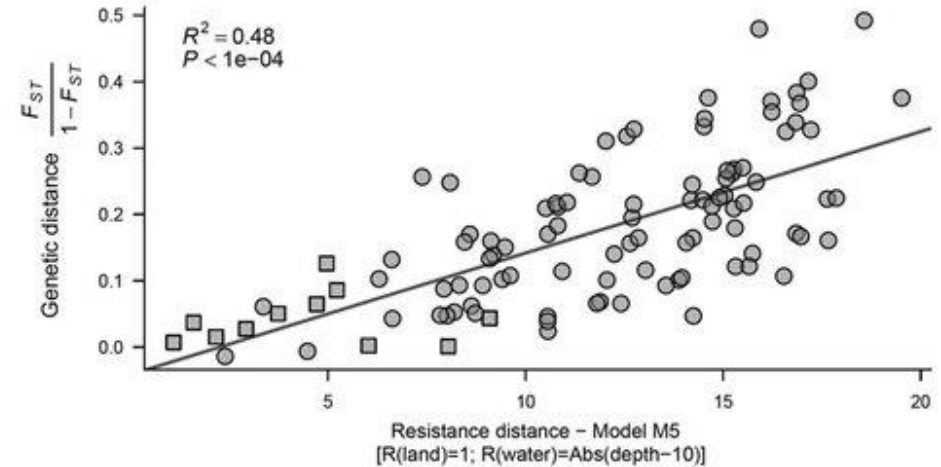
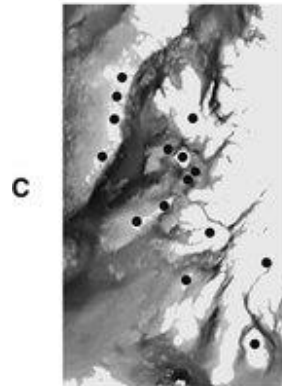
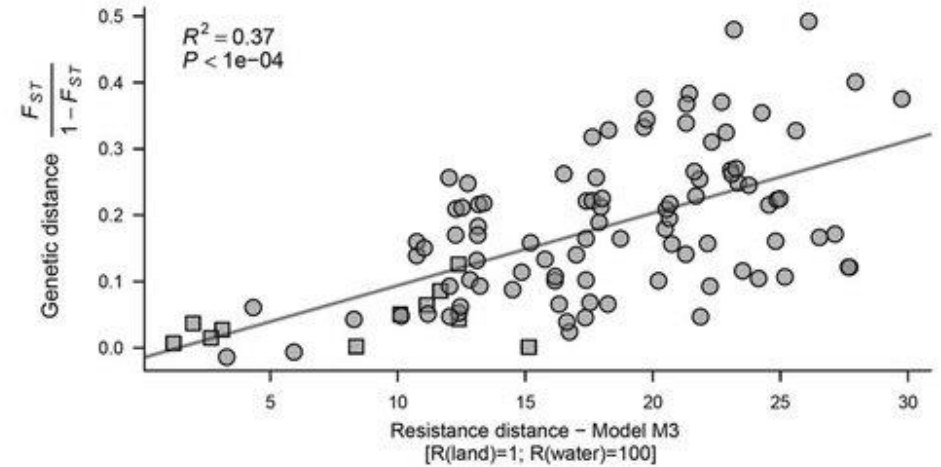
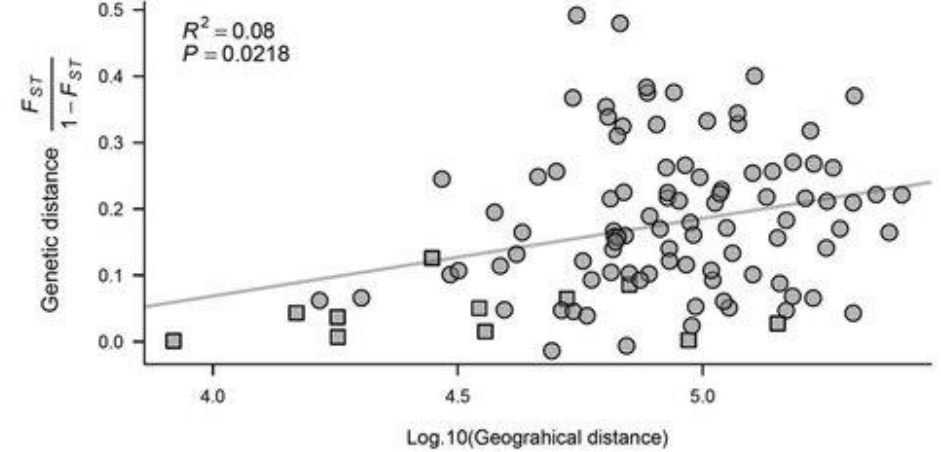
Table 1. Spatial genetic structure (F_{ST}) across the Mona Passage as derived from 12 microsatellites (above diagonal) and 10,047 SNPs (below diagonal). Alpha values were adjusted with a Bonferroni correction ($0.05/21 = 0.023$). Significantly different pairwise comparisons are shown in bold ($p < 0.01$).



Alcuni siti protetti rimangono isolati tra loro, indicando che la capacità dispersiva del pesce minatore non è sufficiente a connettere questi siti

Pesce minatore
Opistognathus aurifrons

Isolamento per resistenza (barriere al flusso genico)



Modello di resistenza che tiene conto delle possibilità' di migrazione, della dimensione e profondita' dei fiordi, etc.
Goulson et al.,2011, Conservation Genetics

L'Fst permette una stima del tasso di migrazione

$$F_{ST} = \frac{1}{4Nm + 1} \longrightarrow \frac{F_{st}}{1 - F_{st}} = \frac{1}{4Nm}$$

NB: questa stima e' approssimativa in quanto basata su assunzioni particolari (modello a isole, assenza di selezione, etc.)

L'eterozigosita' permette di stimare la varianza allelica dovuta alla:

- suddivisione tra popolazioni (F_{ST})

$$F_{ST} = \frac{\text{varianza tra sottopopolazioni}}{\text{varianza totale}} = \frac{H_T - H_S}{H_T}$$

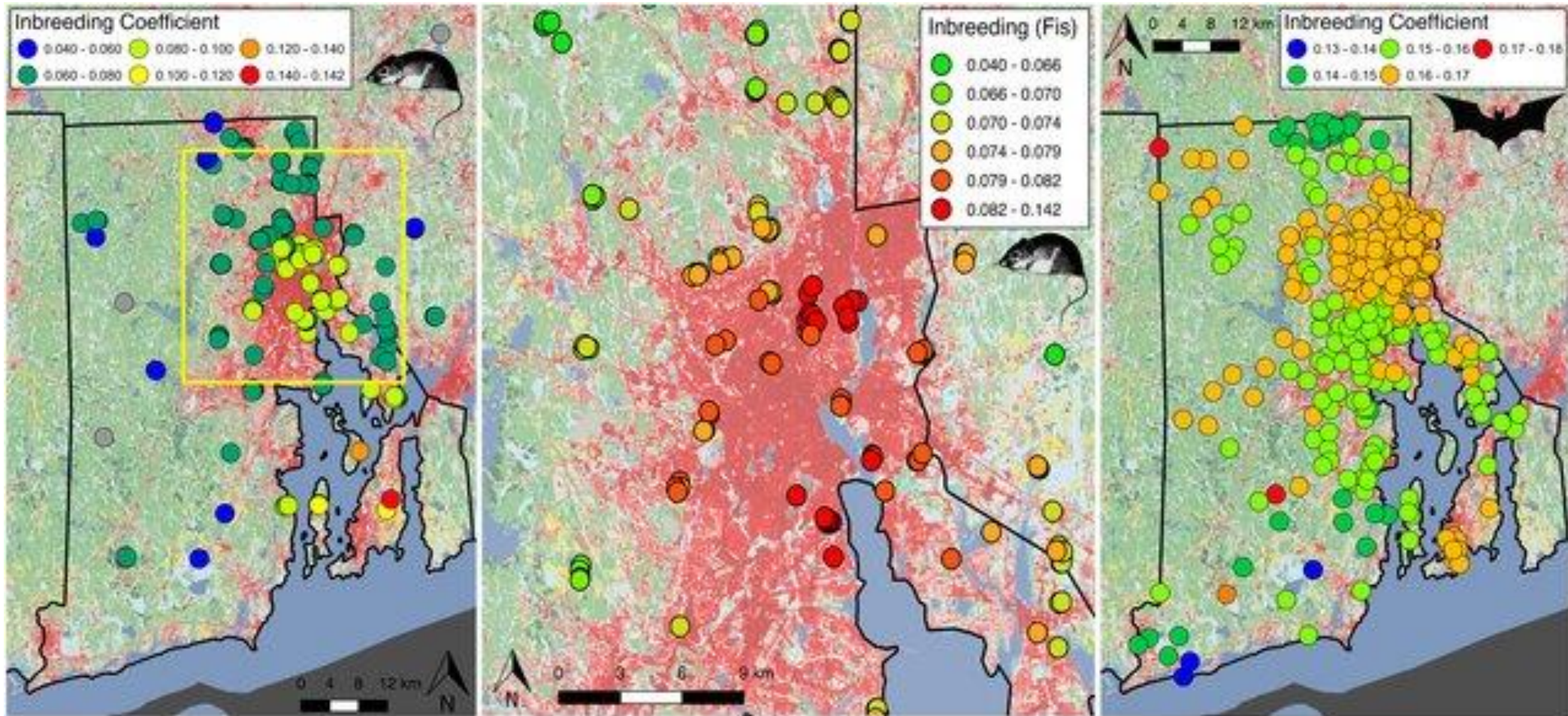
$F_{ST}=0$: nessuna suddivisione tra sottopop.
 $F_{ST}>0.1$ forte suddivisione tra sottopop.
 $F_{ST}=1$: fortissima suddivisione tra sottopop.

- inbreeding (F_{IS})

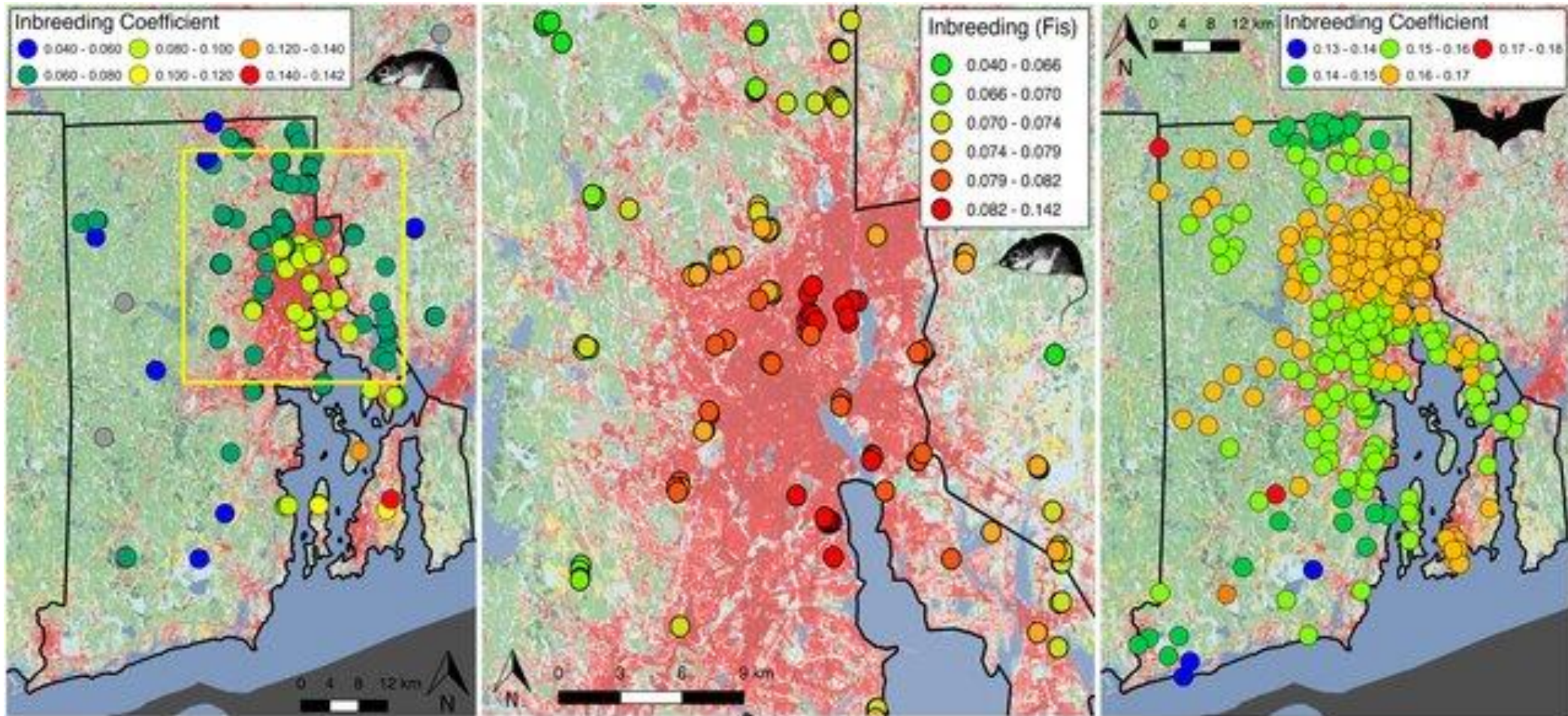
$$F_{IS} = \frac{\text{varianza dovuta all'inbreeding}}{\text{varianza delle sottopopolazioni}} = \frac{H_S - H_I}{H_S}$$

H_I = eterozigosita' individuale
 $F_{IS} \sim 0$: basso inbreeding.
 $F_{IS} \sim 1$: altissimo inbreeding.

Cosa notate nei pattern di inbreeding di topi (*Peromyscus sp.*) e pipistrelli (*Epstesicus fuscus*)?



I topi sono piu' limitati nello spostamento dall'urbanizzazione, i pipistrelli meno.



L'eterozigosita' permette di stimare la varianza allelica dovuta alla:

- suddivisione tra popolazioni (F_{ST})

$$F_{ST} = \frac{\text{varianza tra sottopopolazioni}}{\text{varianza totale}} = \frac{H_T - H_S}{H_T}$$

$F_{ST}=0$: nessuna suddivisione tra sottopop.
 $F_{ST}>0.1$ forte suddivisione tra sottopop.
 $F_{ST}=1$: fortissima suddivisione tra sottopop.

- inbreeding (F_{IS})

$$F_{IS} = \frac{\text{varianza dovuta all'inbreeding}}{\text{varianza delle sottopopolazioni}} = \frac{H_S - H_I}{H_S}$$

H_I = eterozigosita' individuale
 $F_{IS} \sim 0$: basso inbreeding.
 $F_{IS} \sim 1$: altissimo inbreeding.

Analisi della varianza molecolare (AMOVA)

Una estensione dell'ANOVA applicabile a dati genetici, che suddivide la varianza nelle frequenze alleliche in intra-individui (inbreeding), entro-popolazioni, tra-popolazioni, e tra gruppi di popolazioni.

L' AMOVA produce la statistica- Φ , che e' analoga al Wright's **F-statistics** (e.g., F_{st}), e descrive la proporzione di varianza genetica a vari livelli gerarchici.

Il test statistico avviene attraverso un test basato su permutazioni.

ORIGINAL ARTICLE

MOLECULAR ECOLOGY WILEY

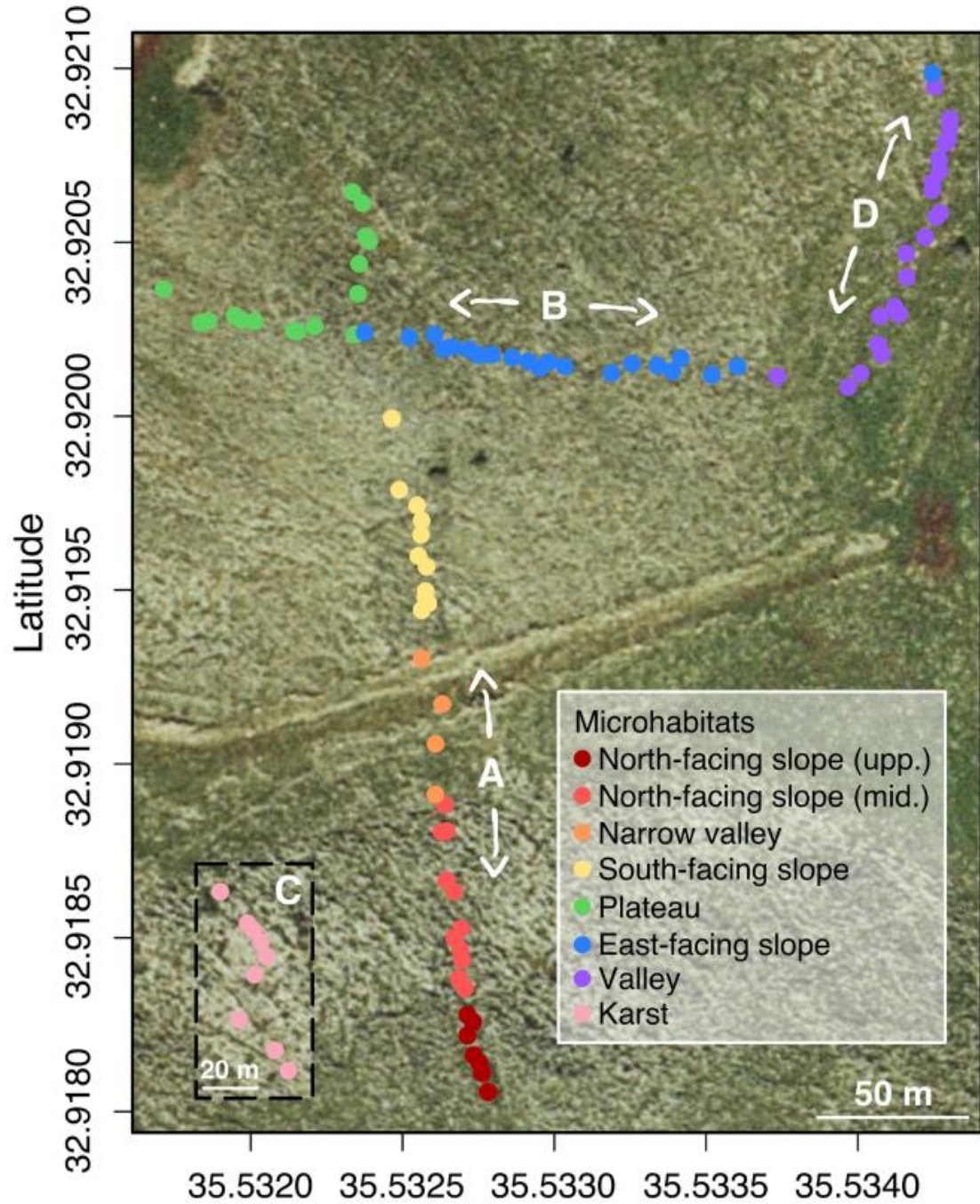
36-year study reveals stability of a wild wheat population across microhabitats

Tal Dahan-Meir^{1,2}  | Thomas James Ellis²  | Fabrizio Mafessoni¹ | Hanan Sela^{3,4} | Ori Rudich¹ | Jacob Manisterski⁴ | Naomi Avivi-Ragolsky¹ | Amir Raz^{1,5} | Moshe Feldman¹ | Yehoshua Anikster^{4,#} | Magnus Nordborg² | Avraham A. Levy¹

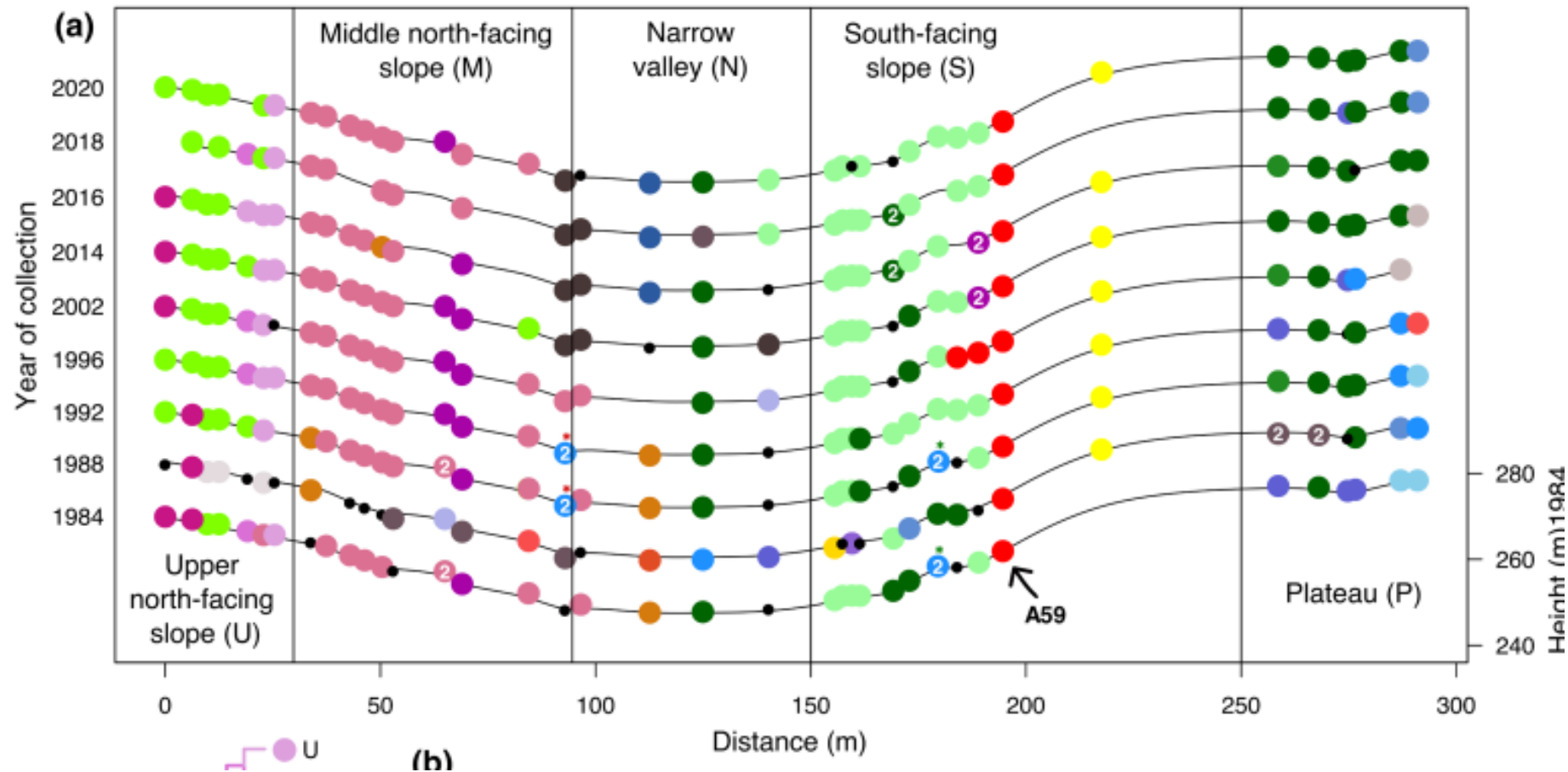


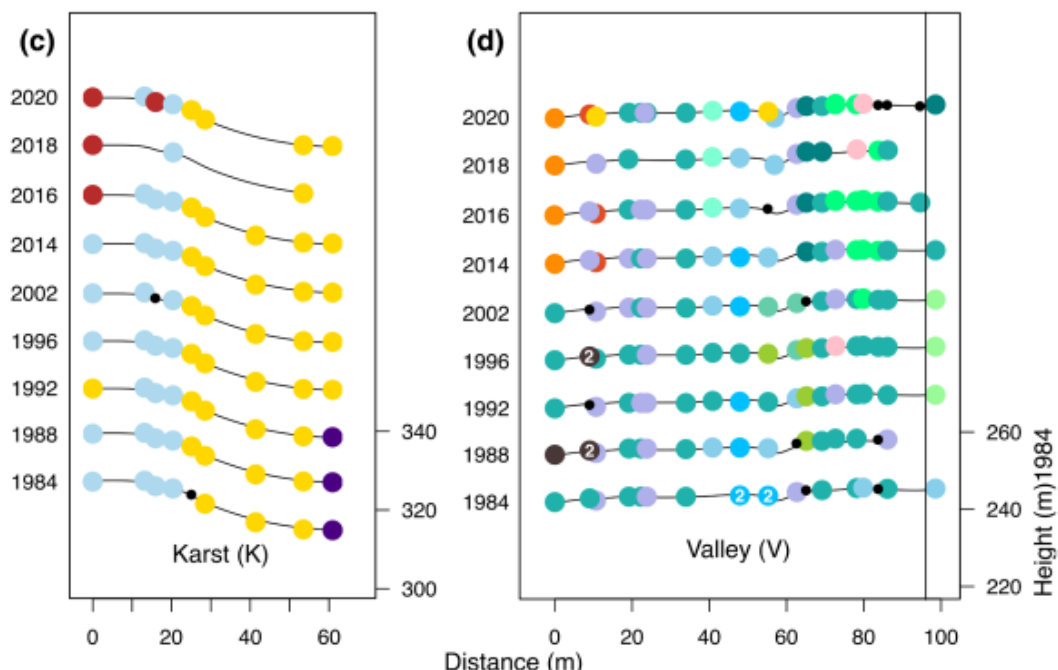
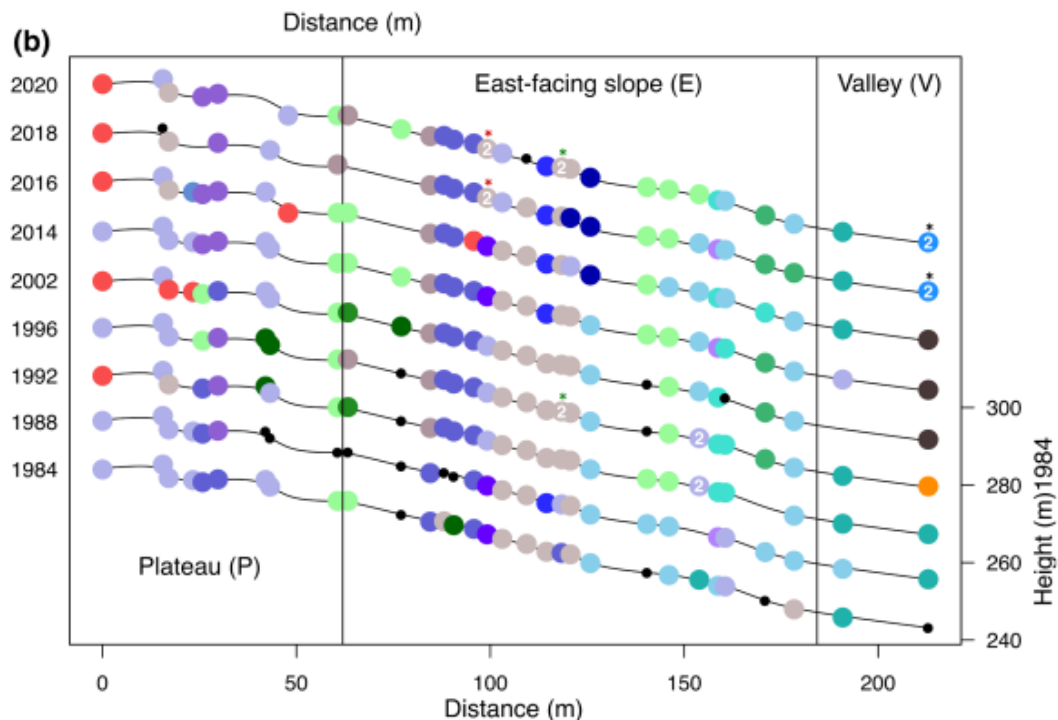
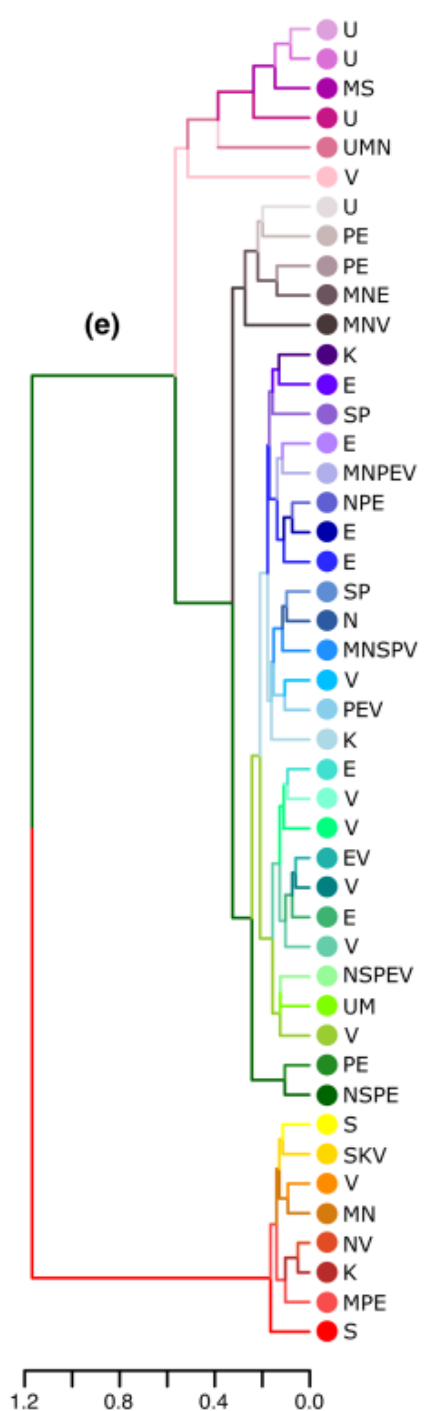
Grano selvatico
Triticum turgidum ssp. *dicoccoides*





Gli stessi genotipi di grano sono ritrovati anno dopo anno negli stessi punti





Cosa accade?

Perche' accade?

Che tipo di analisi fareste?

Table S4. Analysis of molecular variance of genotypes sampled in Ammiad, modelling both habitat and time, using only polymorphic sites within Ammiad. Sum of squared deviations (SSD) and mean square deviations (MSD), variance components (σ^2) and corresponding p-values were computed using the R package *pegas*. Year was modelled as nested within habitat. The Φ statistic for habitat and year was used as an estimate of the proportion of molecular variance explained by the year and habitat structure.

Component	SSD	MSD	df	σ^2	p-value	Φ
Habitat	3.1348	0.4478	7	4.3370e-03	<10 ⁻¹²	0.2831
Year	0.7146	0.0112	64	1.3254e-05	0.7902	0.0012
Error	8.3706	0.0110	824	1.1014e-02	/	/
Total	12.2200	0.0147	831	/	/	/

Quali fattori fanno sì che una popolazione abbia alta o bassa diversità/eterozigosità?

- Deriva genetica
- Dimensione della popolazione di censo
- Presenza di selezione naturale
- Tasso di mutazione
- Separazione in sottopopolazioni
-

La dimensione della popolazione effettiva N_e

La **dimensione effettiva della popolazione N_e** si riferisce al numero di individui in una popolazione ideale (cioè in condizioni teoriche di Hardy-Weinberg) che contribuiscono effettivamente alla prossima generazione in termini di trasmissione genetica.

Questa quantità solitamente differisce dalla dimensione di censo di una popolazione (il reale numero di individui).

Pertanto per una popolazione di censo **N** , **N_e** può essere visto come il corrispettivo numero di individui che genererebbe lo stesso livello di deriva genetica (e quindi di «diversità») in condizioni teoriche ideali di assenza di selezione naturale e assortimento random degli individui.

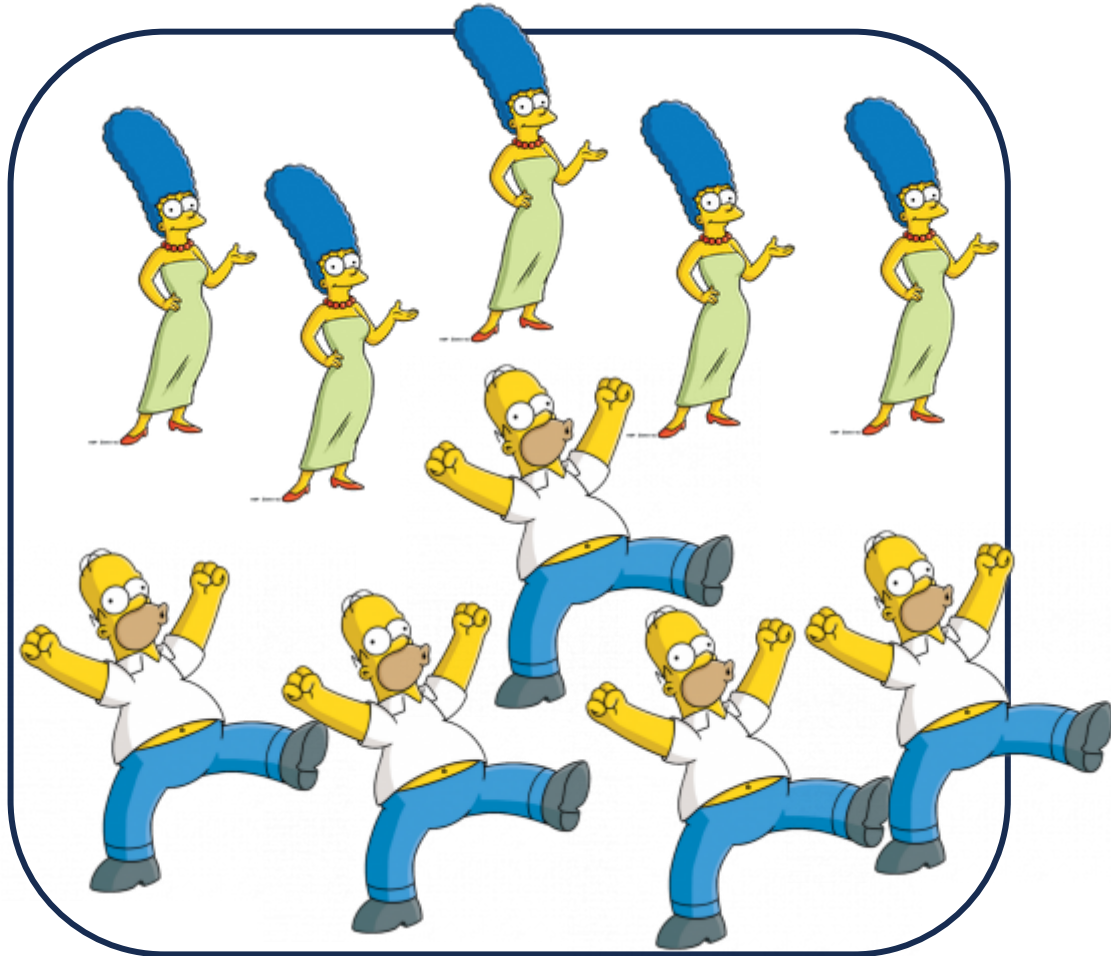


Sewall Wright

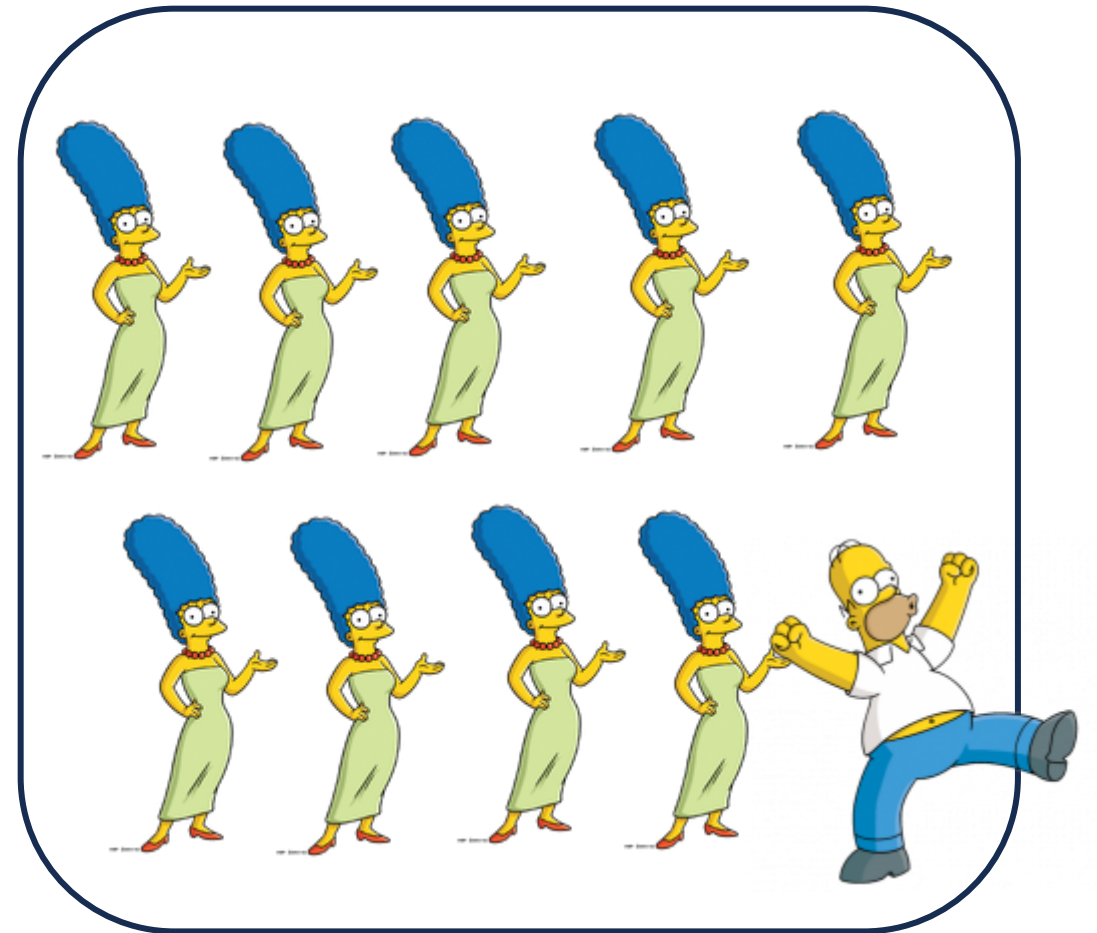


Motoo Kimura

Un esempio di come N e N_e differiscano: effetto del sex ratio su N_e



A



B

Effetto del sex ratio su N_e

$$N_e = 4 (N_{ef} \times N_{em}) / (N_{ef} + N_{em})$$

N_e e' piu alta quando il numero di maschi e femmine e' equilibrato. La spiegazione e' quando uno dei sessi e' piu raro diventa il fattore limitante per la diversita' della popolazione.



Mirounga angustirostris

$$N_e = \frac{4(105 \times 28)}{105 + 28} = 88$$

Effetto del sex ratio su N_e

$$N_e = 4 (N_{ef} \times N_{em}) / (N_{ef} + N_{em})$$

N_e e' piu alta quando il numero di maschi e femmine e' equilibrato. La spiegazione e' quando uno dei sessi e' piu raro diventa il fattore limitante per la diversita' della popolazione.



Mirounga angustirostris

$$N_e = \frac{4(105 \times 28)}{105 + 28} = 88$$

Effetto della variazione nel successo riproduttivo su N_e

$$N_e = \frac{4N_c - 2}{VRS + 2}$$

La varianza nel successo riproduttivo riduce N_e perché un numero di generazioni/individui ristretto contribuisce sproporzionatamente alla popolazione



Geospiza fortis

$$N_e = \frac{4(500) - 2}{7.12 + 2} = 219$$

Effetto delle fluttuazioni in N_e

$$N_e = \frac{t}{(1/N_{e1}) + (1/N_{e2}) + (1/N_{e3}) \dots + (1/N_{et})}$$

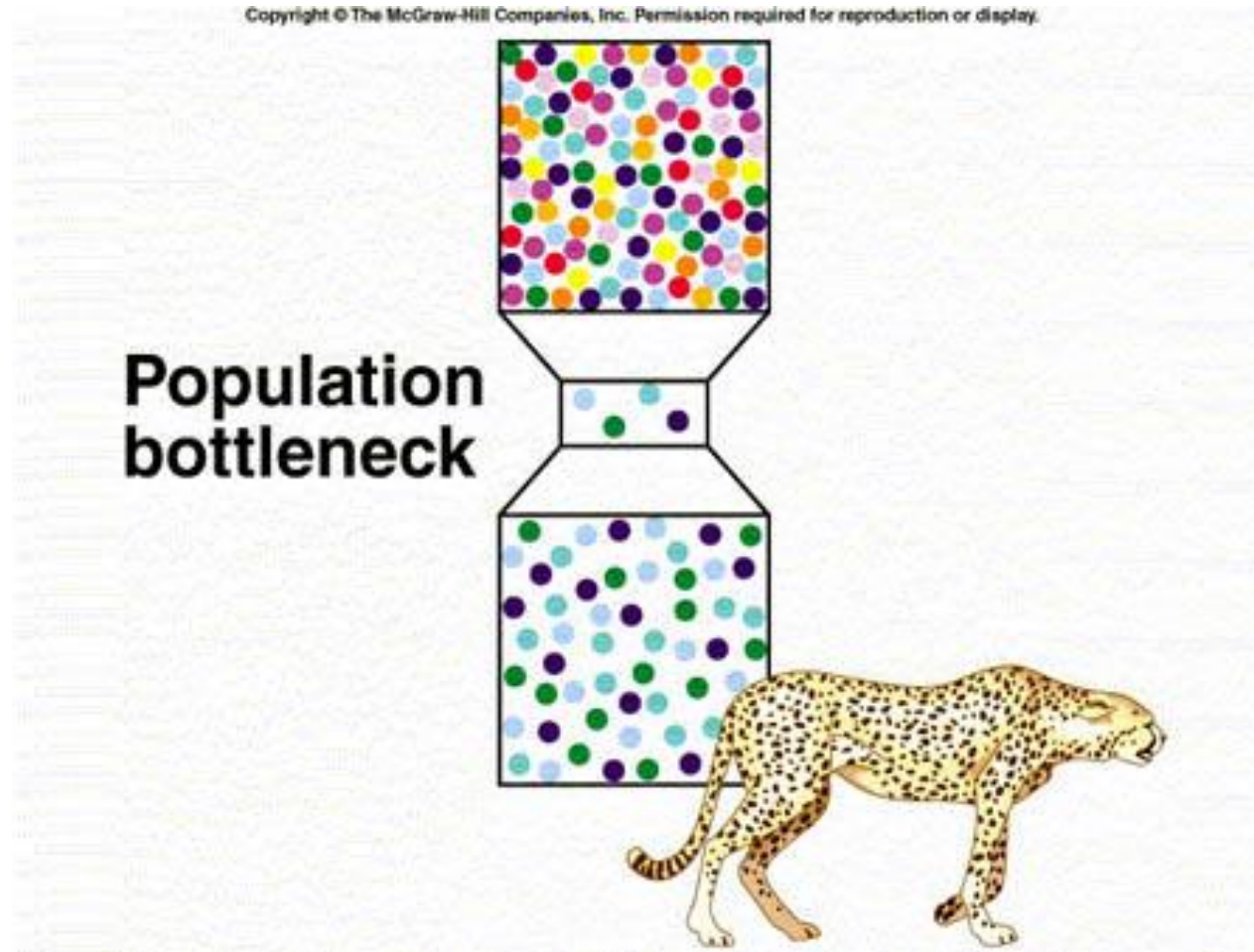
N_e e' circa la media armonica delle popolazioni effettive nel tempo. Nella media armonica i termini con popolazioni effettive piu basse dominano.

$$N_e = \frac{4}{(1/220) + (1/70) + (1/40) + (1/200)}$$

= approx. 82



Colli di bottiglia demografici

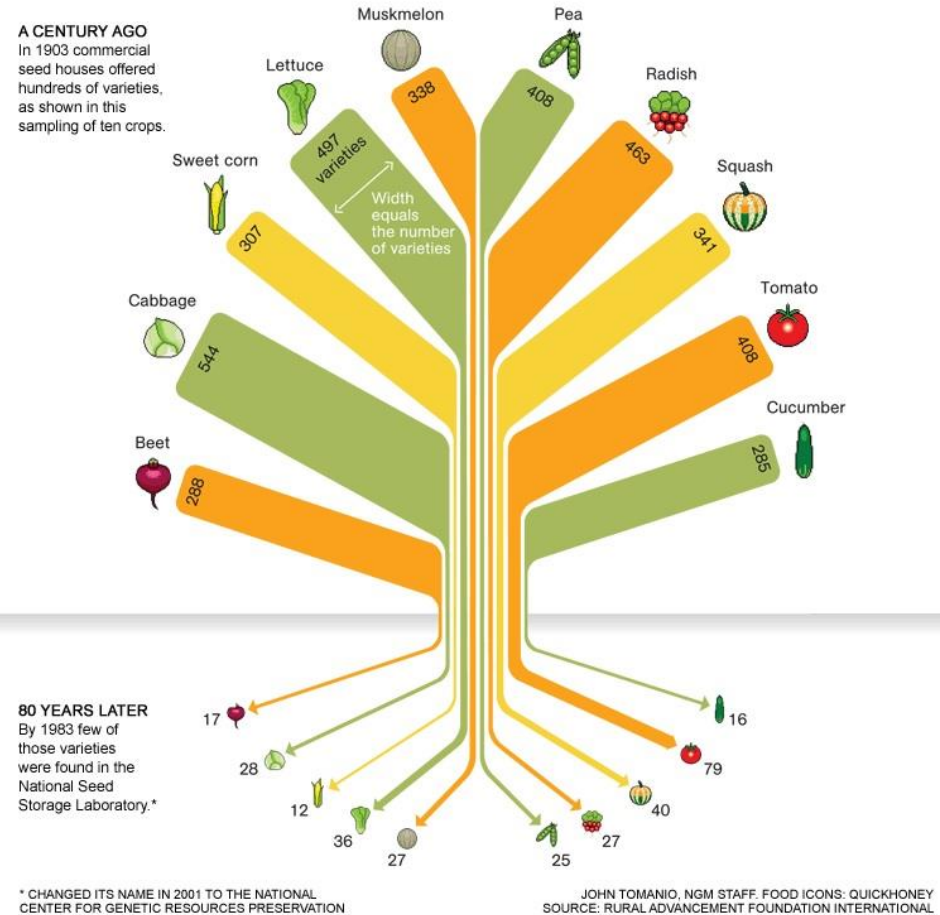


Effetto del fondatore



Tay-Sachs negli ebrei Ashkenazi

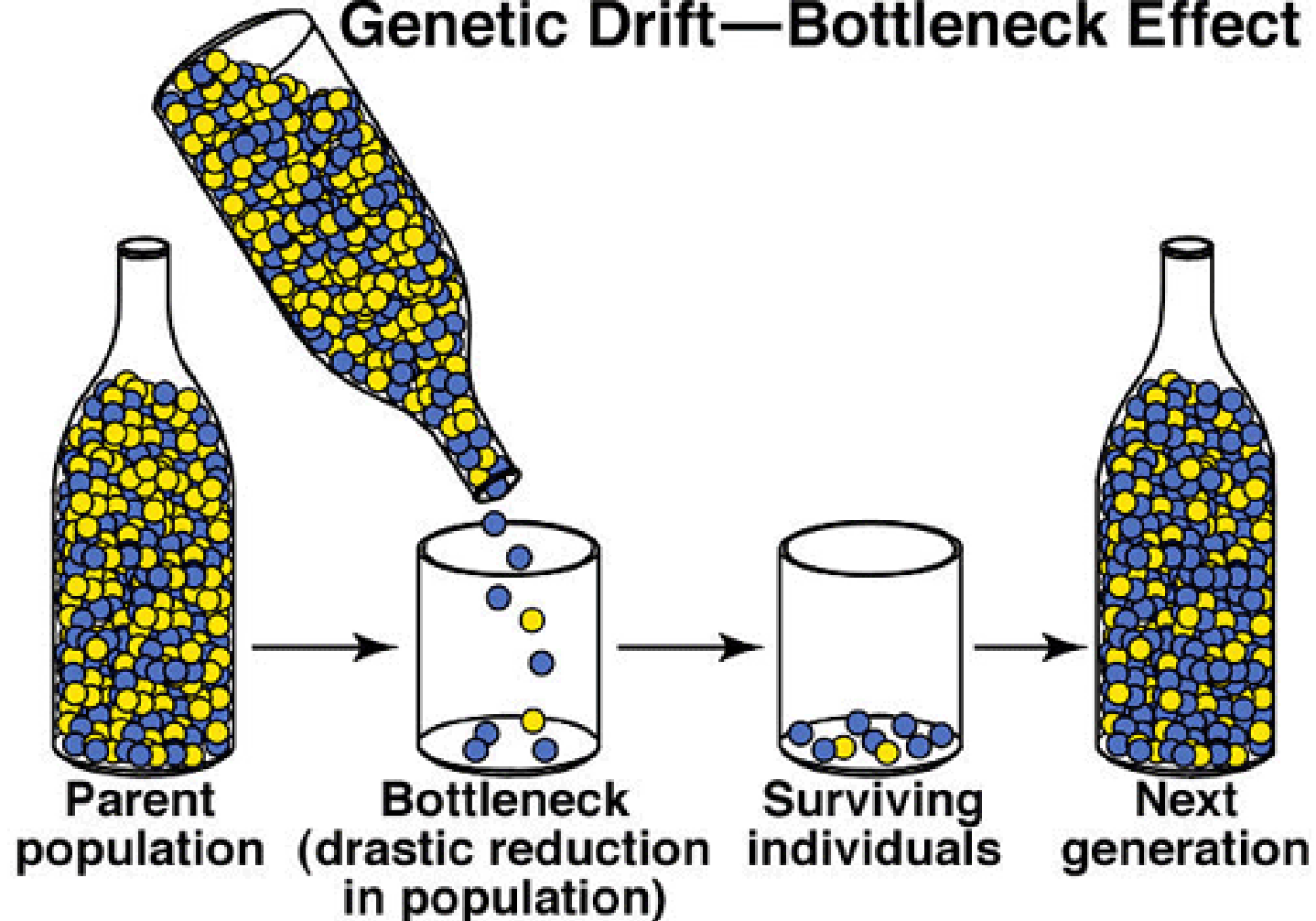
Perdita di alleli vantaggiosi in specie addomesticate



Colli di bottiglia demografici

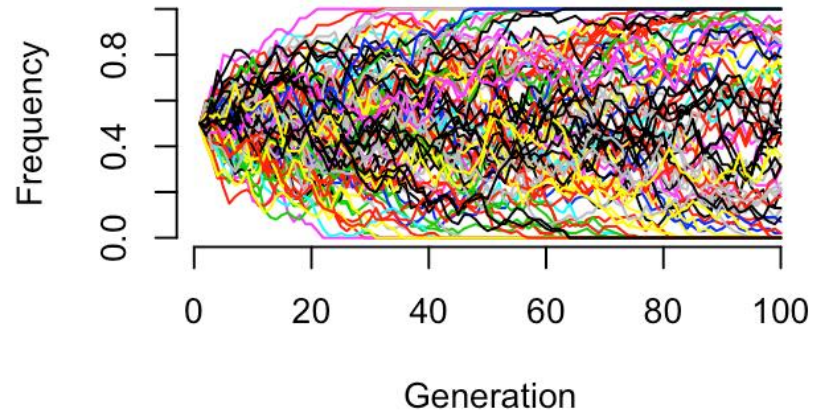
Copyright © The McGraw-Hill Companies, Inc. Permission required for reproduction or display.

Genetic Drift—Bottleneck Effect

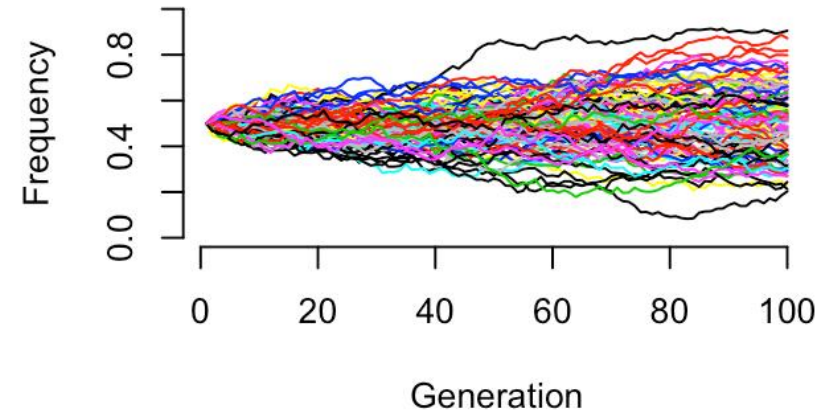


La deriva genetica in piccole popolazioni accentua le fluttuazioni stocastiche nelle frequenze alleliche

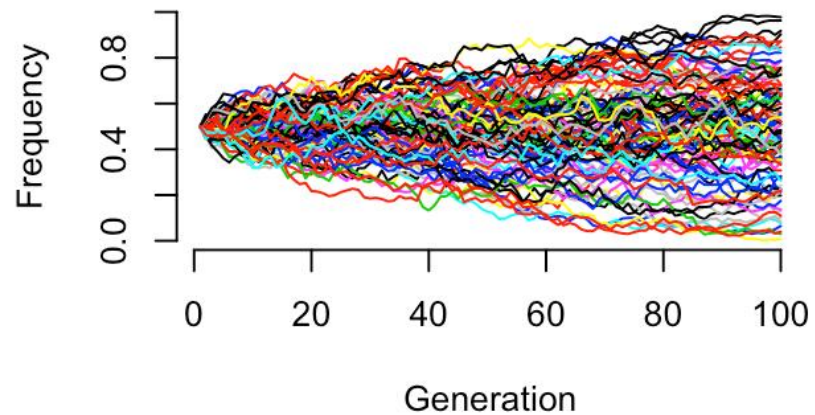
Population size 50



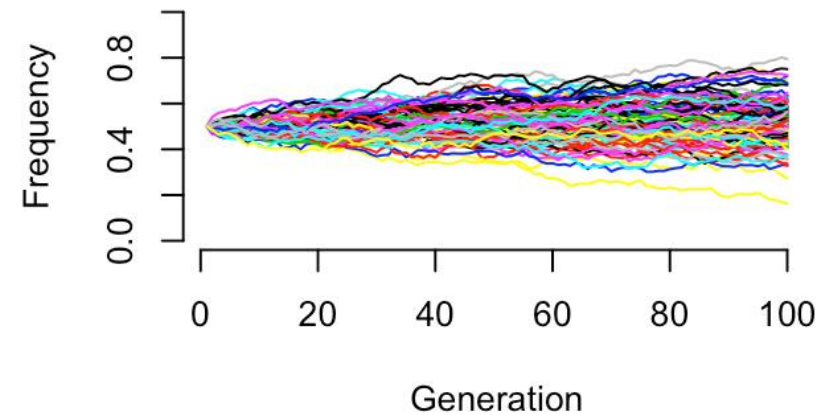
Population size 500



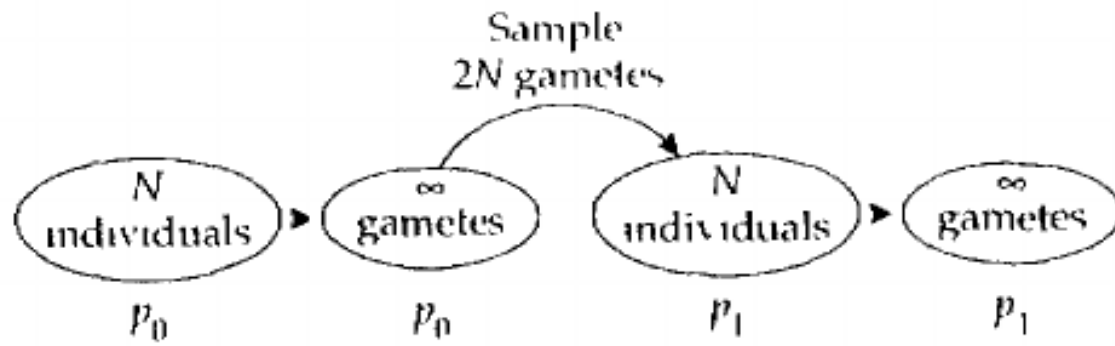
Population size 200



Population size 1000

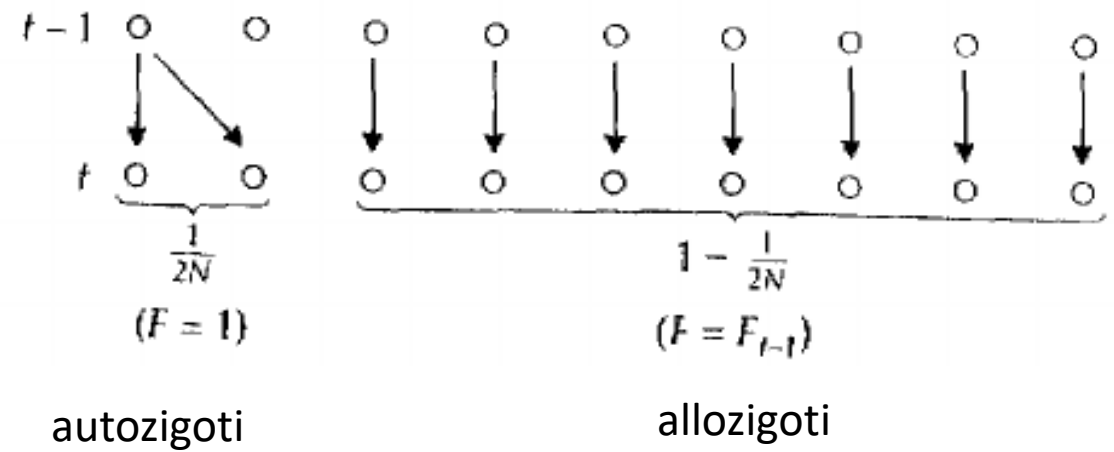


Il concetto di Identity by descent aiuta a spiegare la connessione tra N_e , H e deriva genetica



Il concetto di Identity by descent aiuta a spiegare la connessione tra N_e , H e deriva genetica

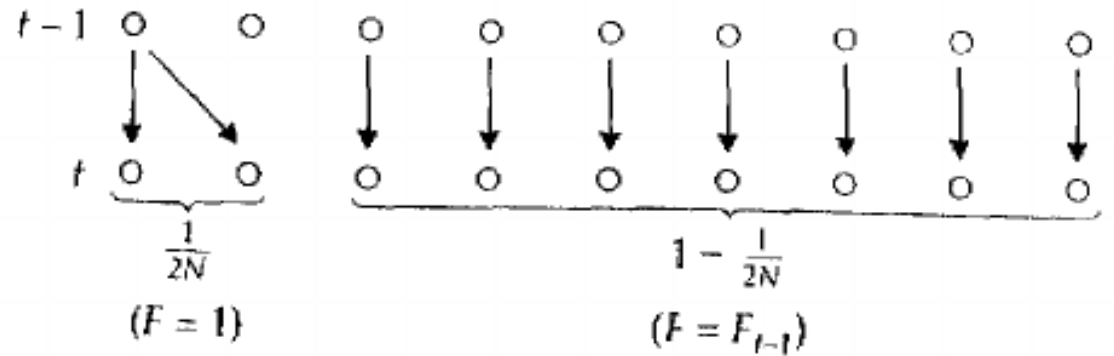
F=Identity by descent: probabilita' che due individui abbiano avuto lo stesso antenato in una generazione precedente arbitraria. (qui $t-1$)



$$F_t = \frac{1}{2N} + \left(1 - \frac{1}{2N}\right)F_{t-1}$$

Il concetto di Identity by descent aiuta a spiegare la connessione tra Ne, H e deriva genetica

F=Identity by descent: probabilita' che due individui abbiano avuto lo stesso antenato alla generazione precedente



$$F_t = \frac{1}{2N} + \left(1 - \frac{1}{2N}\right) F_{t-1}$$

Multiplying both sides by -1 and adding 1 leads to

$$1 - F_t = 1 - \frac{1}{2N} - \left(1 - \frac{1}{2N}\right) F_{t-1} = \left(1 - \frac{1}{2N}\right) (1 - F_{t-1})$$

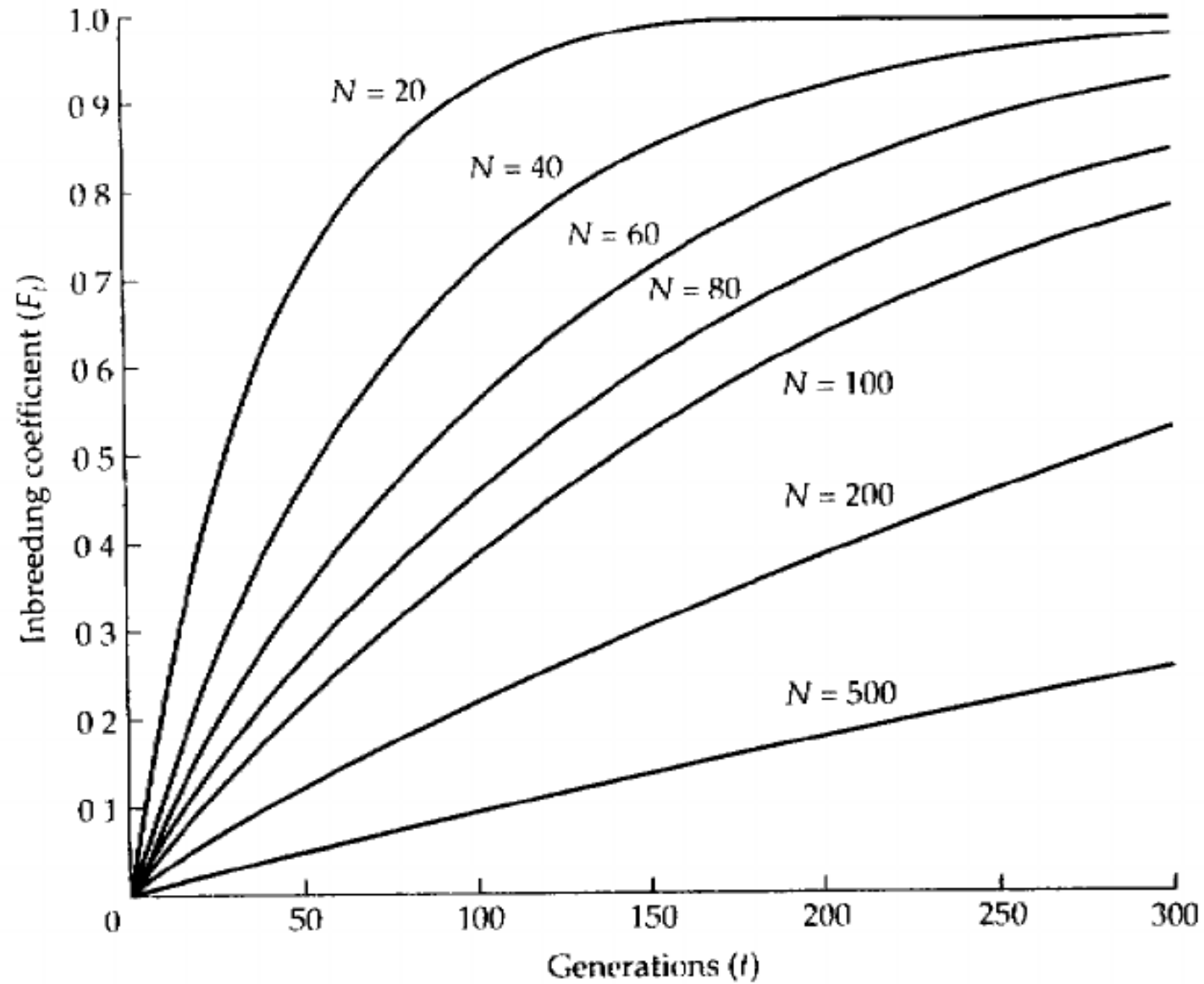
and so

$$1 - F_t = \left(1 - \frac{1}{2N}\right)^t (1 - F_0)$$

or, when $F_0 = 0$,

$$F_t = 1 - \left(1 - \frac{1}{2N}\right)^t$$

Il coefficiente di inbreeding sale piu rapidamente in popolazioni piu piccole



L'eterozigosita' cala piu rapidamente in popolazioni piu piccole

Per definizione se due alleli
non hanno antenati in
comune
sono diversi (H)

$$H_t = 1 - F_t$$

$$1 - F_t = \left(1 - \frac{1}{2N}\right)(1 - F_{t-1})$$

$$H_t = \left(1 - \frac{1}{2N}\right)H_{t-1}$$

$$H_t = \left(1 - \frac{1}{2N}\right)^t H_0 \approx H_0 e^{-t/2N}$$

L'eterozigosita' cala piu rapidamente in popolazioni piu piccole

Per definizione se due alleli non hanno antenati in comune sono diversi (H)

$$H_t = 1 - F_t$$

$$1 - F_t = \left(1 - \frac{1}{2N}\right)(1 - F_{t-1})$$

$$H_t = \left(1 - \frac{1}{2N}\right)H_{t-1}$$

$$H_t = \left(1 - \frac{1}{2N}\right)^t H_0 \approx H_0 e^{-t/2N}$$

A causa della connessione intima tra H ed F possiamo stimare F con la formula

$$F = (H_s - H_i) / H_s$$

Dove H_s e H_i sono le eterozigosita' attese dalle frequenze tra e all'interno di un gruppo di sottopopolazioni, rispettivamente

L'eterozigosita' cala piu rapidamente in popolazioni piu piccole

Per definizione se due alleli
non hanno antenati in
comune
sono diversi (H)

$$H_t = 1 - F_t$$

$$1 - F_t = \left(1 - \frac{1}{2N}\right)(1 - F_{t-1})$$

$$H_t = \left(1 - \frac{1}{2N}\right)H_{t-1}$$

$$H_t = \left(1 - \frac{1}{2N}\right)^t H_0 \approx H_0 e^{-t/2N}$$

H in tempi diversi puo' essere usata come estimatore di Ne:

$$N_e = \frac{1}{2 \left(1 - \frac{H_t}{H_0}\right) / t}$$

La dimensione della popolazione effettiva N_e

La **dimensione effettiva della popolazione N_e** si riferisce al numero di individui in una popolazione ideale (cioè in condizioni teoriche di Hardy-Weinberg) che contribuiscono effettivamente alla prossima generazione in termini di trasmissione genetica.

Questa quantità solitamente differisce dalla dimensione di censo di una popolazione (il reale numero di individui).

Pertanto per una popolazione di censo **N** , **N_e** può essere visto come il corrispettivo numero di individui che genererebbe lo stesso livello di deriva genetica (e quindi di «diversità») in condizioni teoriche ideali di assenza di selezione naturale e assortimento random degli individui.



Sewall Wright



Motoo Kimura

La dimensione della popolazione effettiva N_e

La **dimensione effettiva della popolazione N_e** si riferisce al numero di individui in una popolazione ideale (cioè in condizioni teoriche di Hardy-Weinberg) che contribuiscono effettivamente alla prossima generazione in termini di trasmissione genetica.

Questa quantità solitamente differisce dalla dimensione di censo di una popolazione (cioè il reale numero di individui).

**Possiamo
misurarla in molti
modi!**

Pertanto per una popolazione di censo N , N_e può essere visto come il corrispondente numero di individui che genererebbe lo stesso livello di deriva genetica (e quindi di «diversità») in condizioni teoriche ideali di assenza di selezione naturale e assortimento random degli individui.



Sewall Wright



Motoo Kimura

Ne puo' essere stimata con diversi estimatori basati su caratteristiche diverse

Estimatori basati sulla diversita' genetica:

semplici, molto usati anche in combinazione, approssimativi

Estimatori basati su serie temporali:

raramente usati data la mancanza di dati diacronici

Estimatori basati su *linkage disequilibrium*:

relativamente robusti

$$N_e = \frac{1 - r^2}{4cr^2}$$

C: tasso di ricombinazione
r²: linkage disequilibrium

Estimatori basati su metodi di coalescenza (e.g. PSMC):

Largamente usati quando genomi interi sono disponibili

Neanderthal

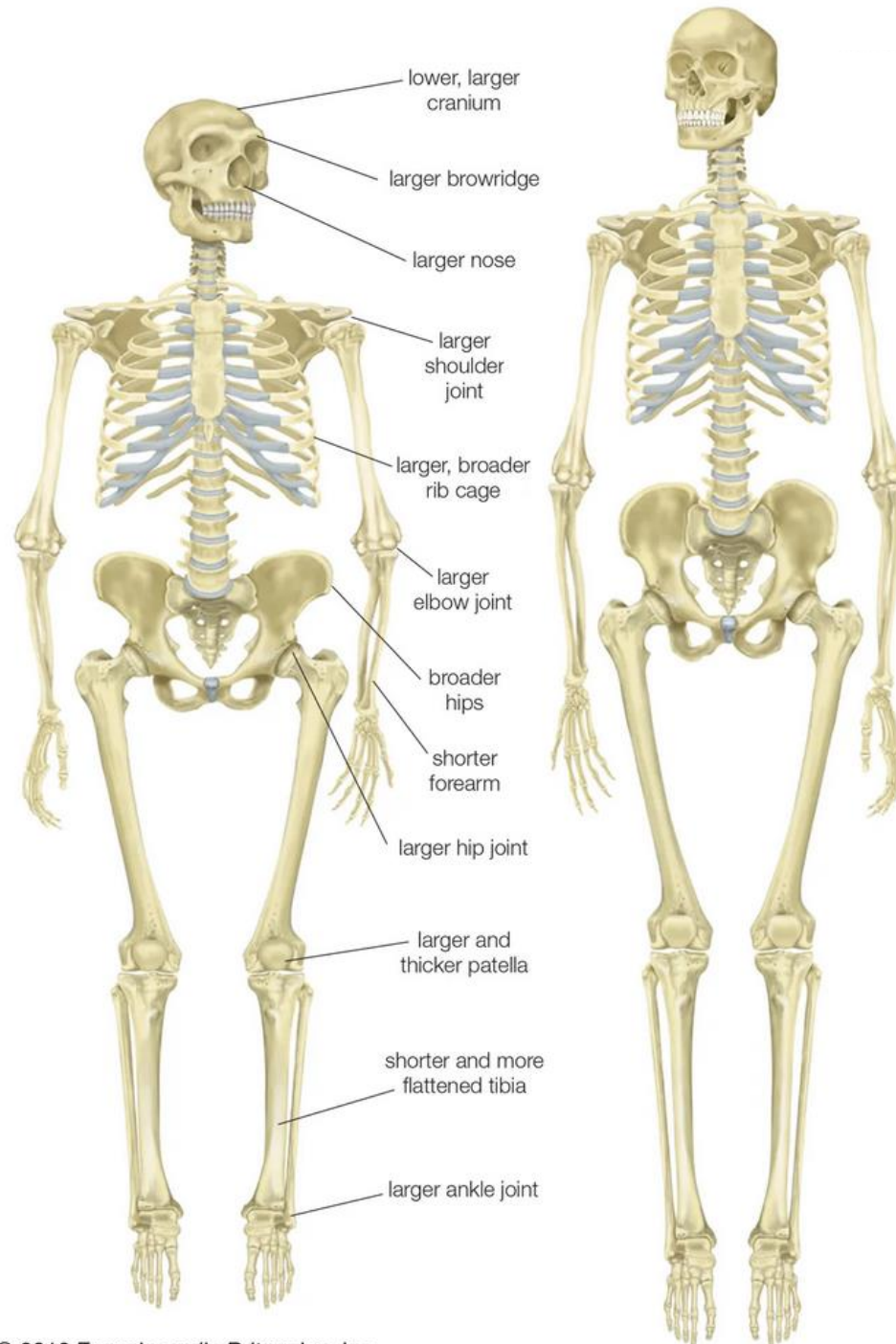


HOMO SAPIENS
NEANDERTHALENSIS

Umani moderni



HOMO SAPIENS



- lower, larger cranium
- larger browridge
- larger nose
- larger shoulder joint
- larger, broader rib cage
- larger elbow joint
- broader hips
- shorter forearm
- larger hip joint
- larger and thicker patella
- shorter and more flattened tibia
- larger ankle joint

Tre genomi di Neandertal ad alta' qualita' (coverage > 25x)

Neandertal

H.sapiens

Prof. Svante Pääbo

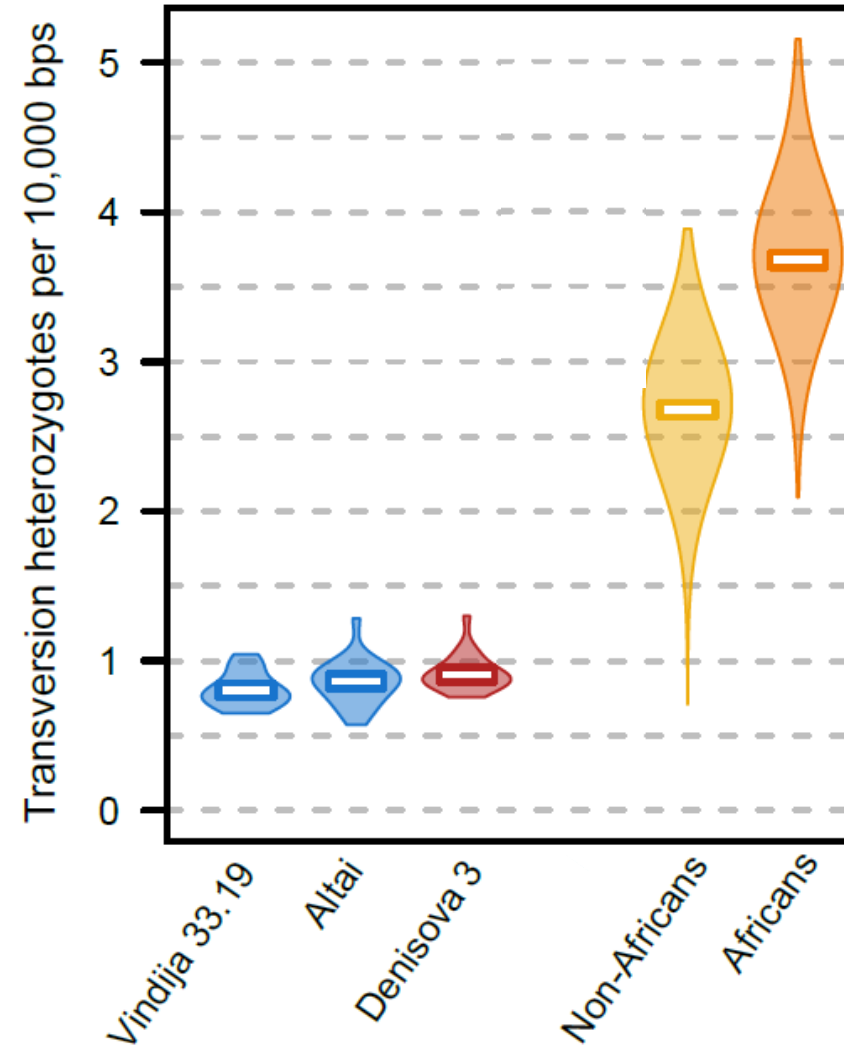
Max Planck Institute for
Evolutionary
Anthropology

Premio Nobel 2022 per
la Fisiologia e Medicina

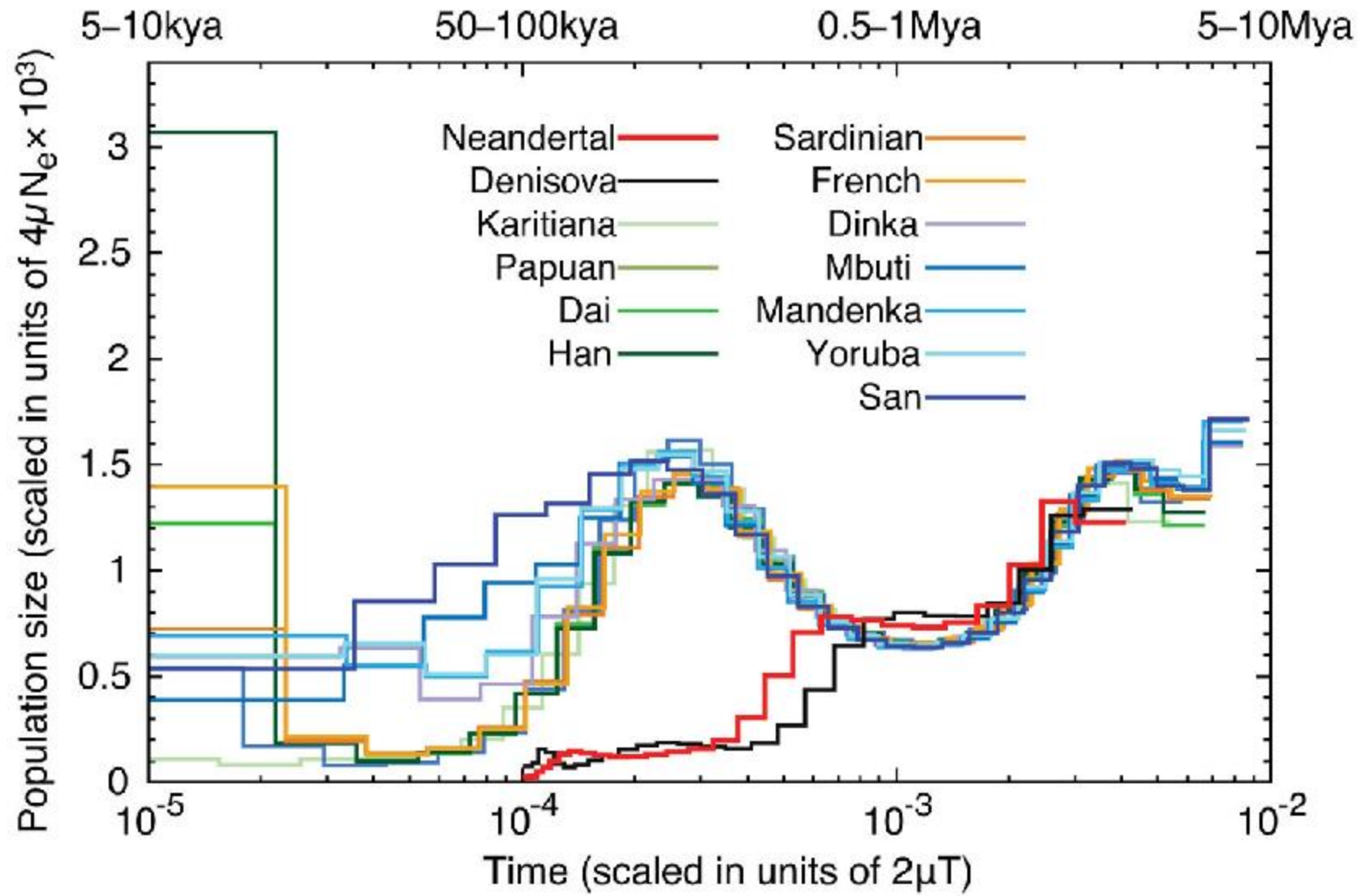
Adattato da Mafessoni 2018, *Nat.Ecol.Evol.*

Green et al.,2010. *Science*
Prüfer et al.,2014,2017. *Nature*
Mafessoni et al.,2020. *PNAS*

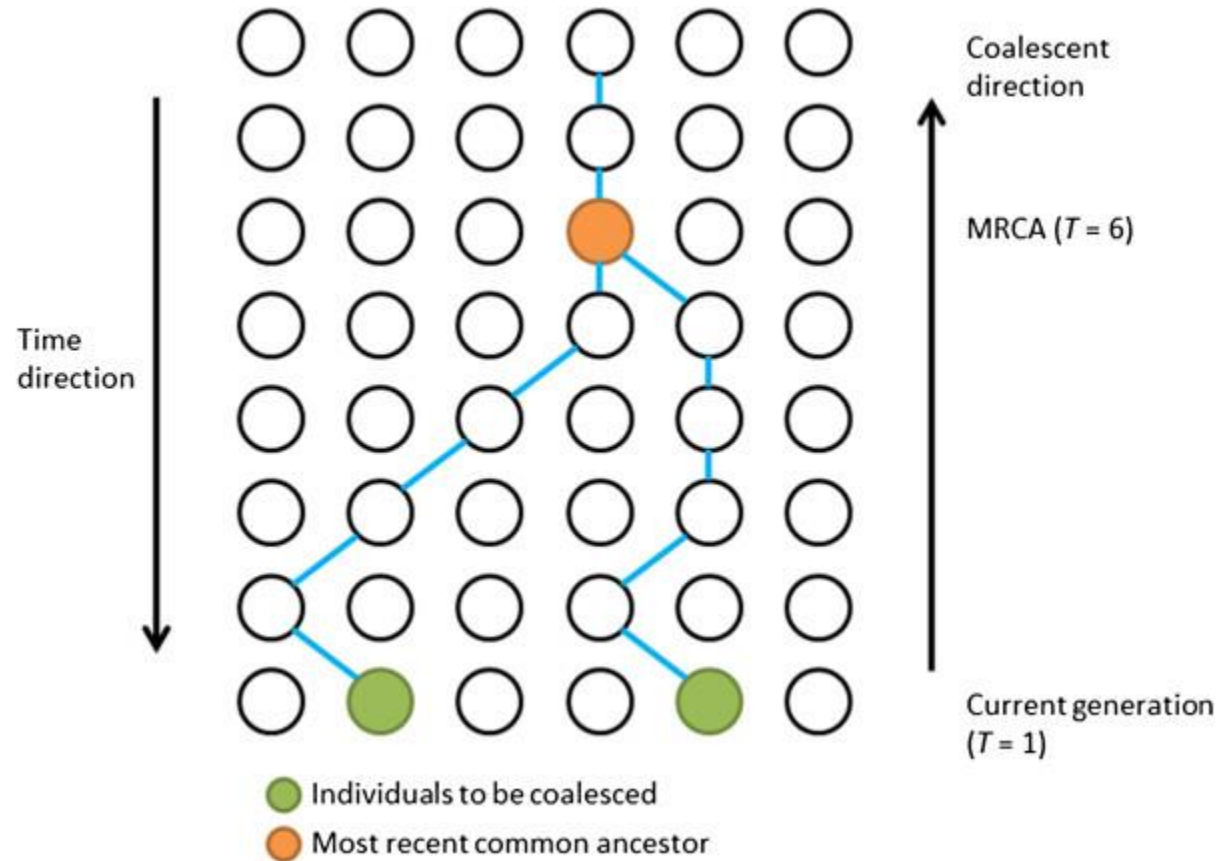
I genomi degli umani arcaici (Neandertal, Denisova) hanno bassissima eterozigosita' e bassa diversita' nucleotidica. Perche'?



I genomi degli umani arcaici indicano che vissero per ~200 anni in piccole popolazioni

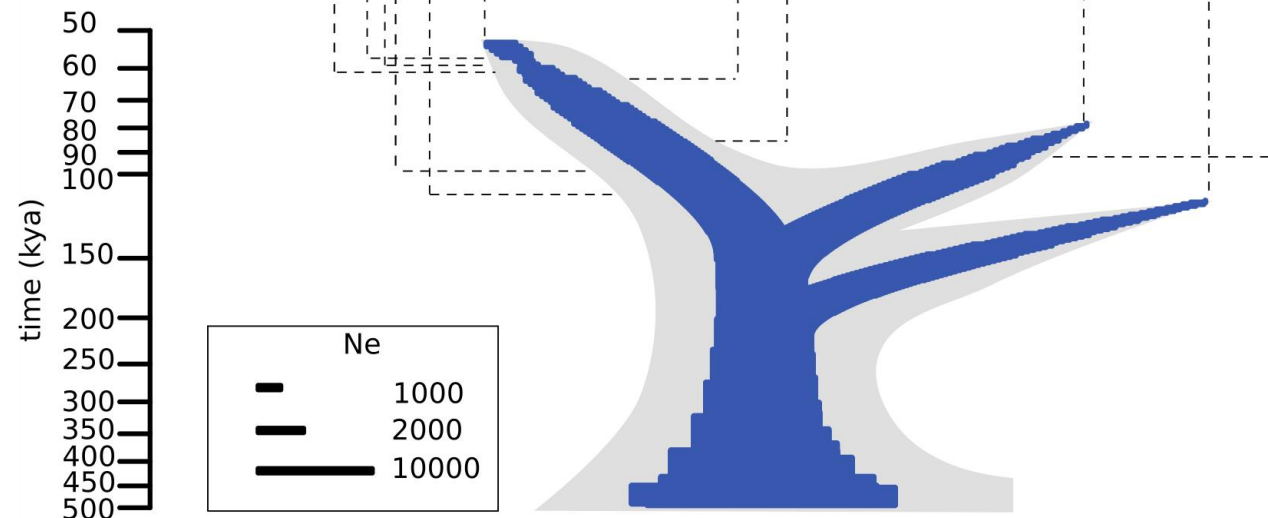


Teoria della coalescenza



Ad ogni generazione la probabilita' che i geni di due individui "coalescano" e' $1/2N_e$.

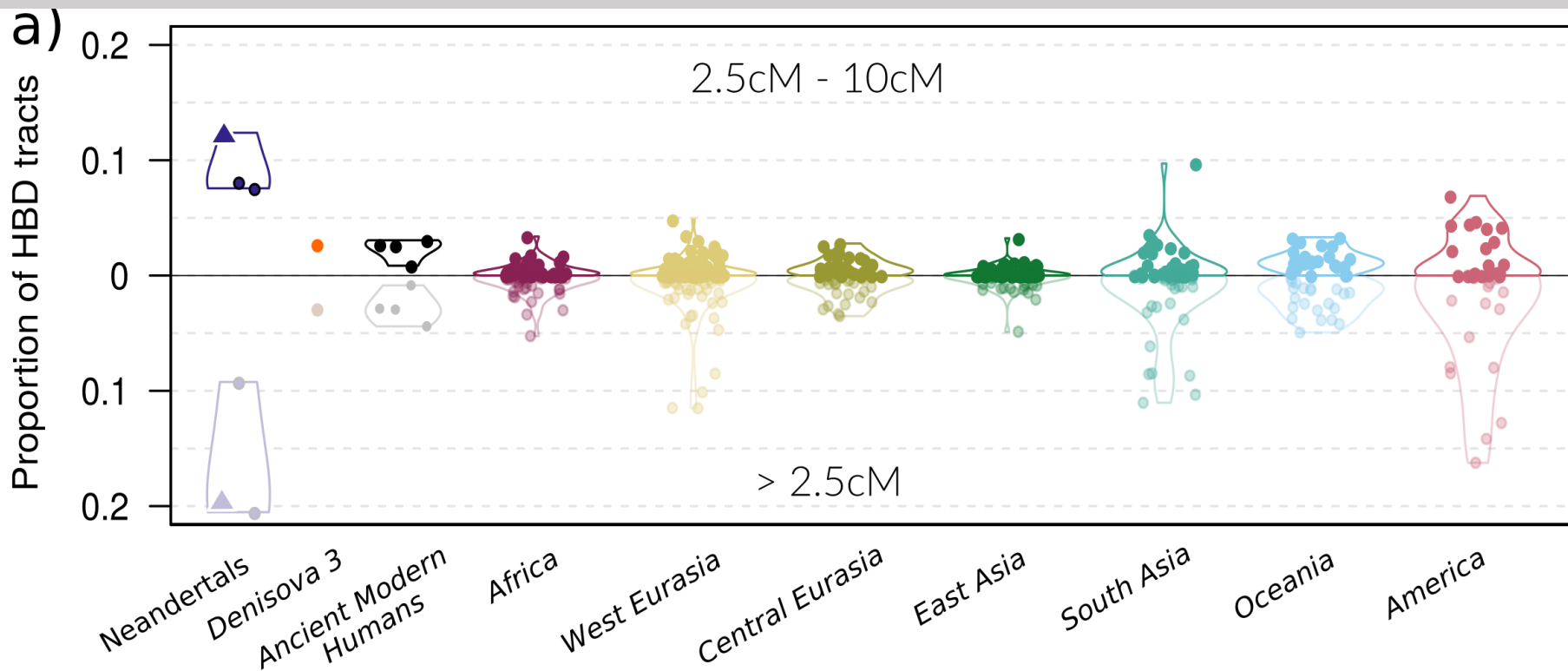
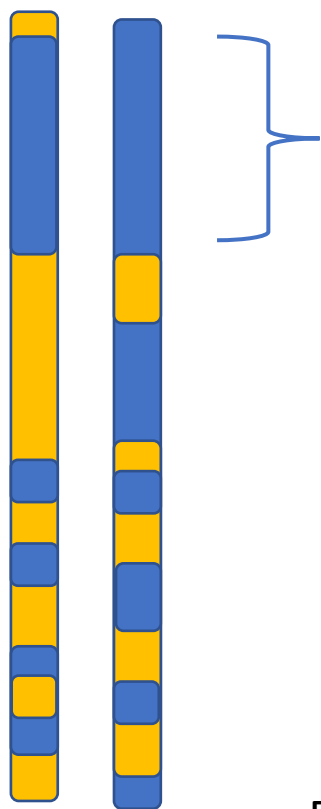
I Neanderthals vissero in piccole popolazioni per >100.000 years



- List of Neanderthal genome shown
 Pruefer et al., 2014, *Nature*
 Pruefer, Grote*, De Filippo*, **Mafessoni***, et al., 2017, *Science*
 Slon*, **Mafessoni***, Vernot* et al., 2018, *Nature*
 Hajdinjak, Fu, Hubner, Petr, **Mafessoni**,.. et al., 2018, *Nature*
 Peyregne, Slon, **Mafessoni** et al., 2019, *Science Advances*
Mafessoni et al., 2020, *PNAS*

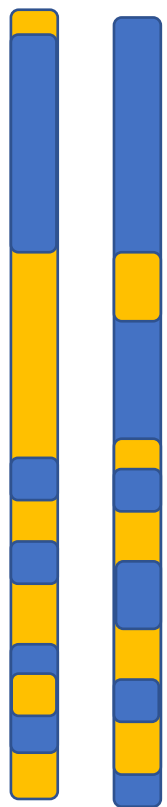
Mafessoni and Pruefer, 2017, *PNAS*
 Mafessoni et al., 2020, *PNAS*

I Neanderthal mostrano forti segnali di inbreeding



Regioni di omozigotia' per discendenza (Homozigosity by Descent)

I Neanderthal mostrano forti segnali di inbreeding

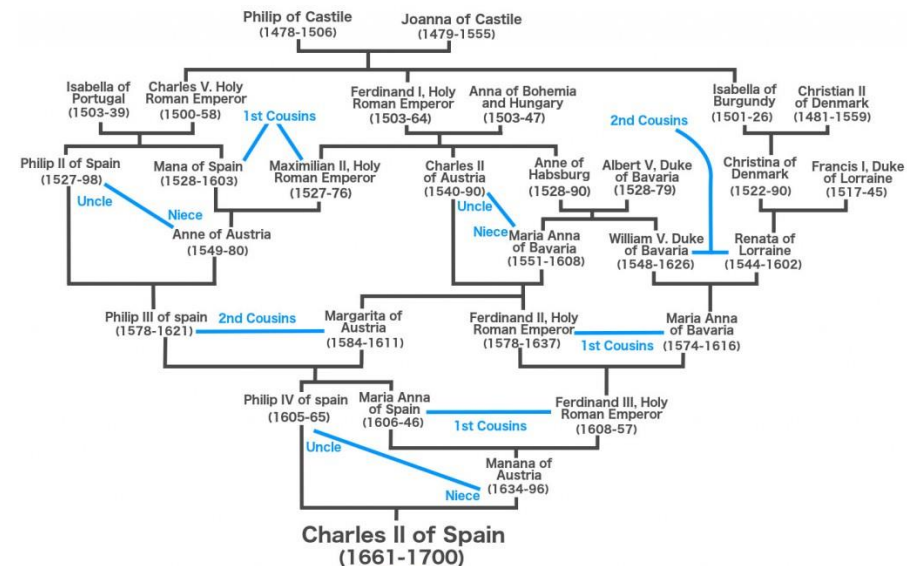
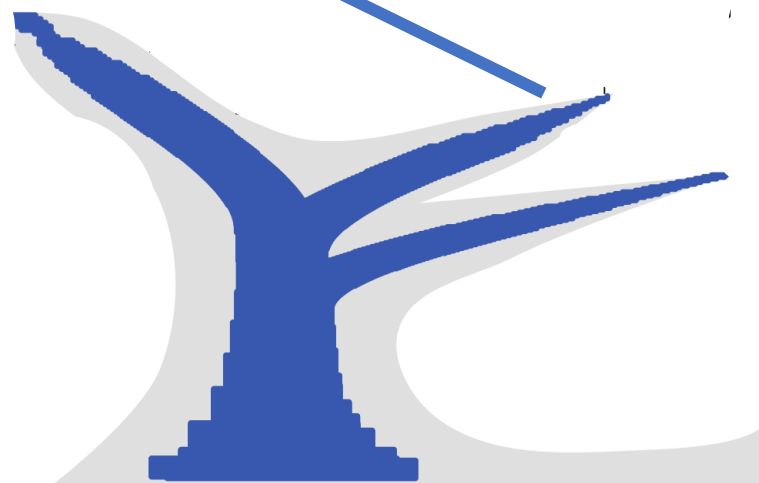
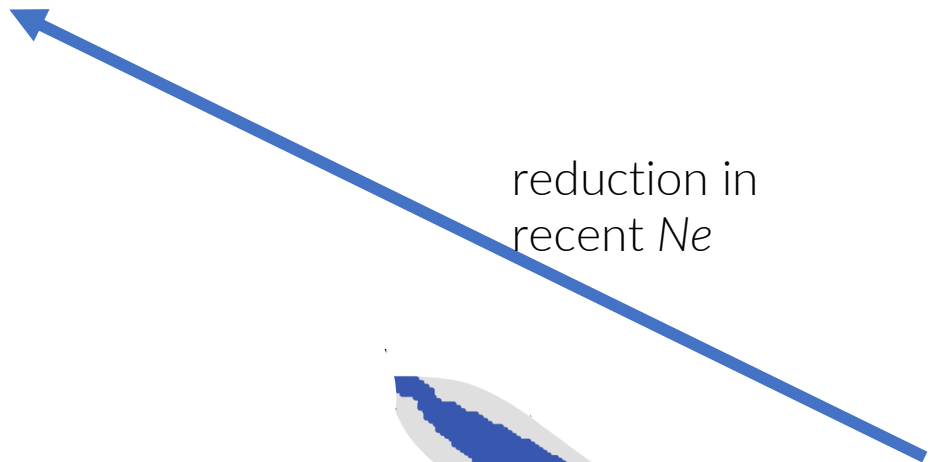


HBD tracts

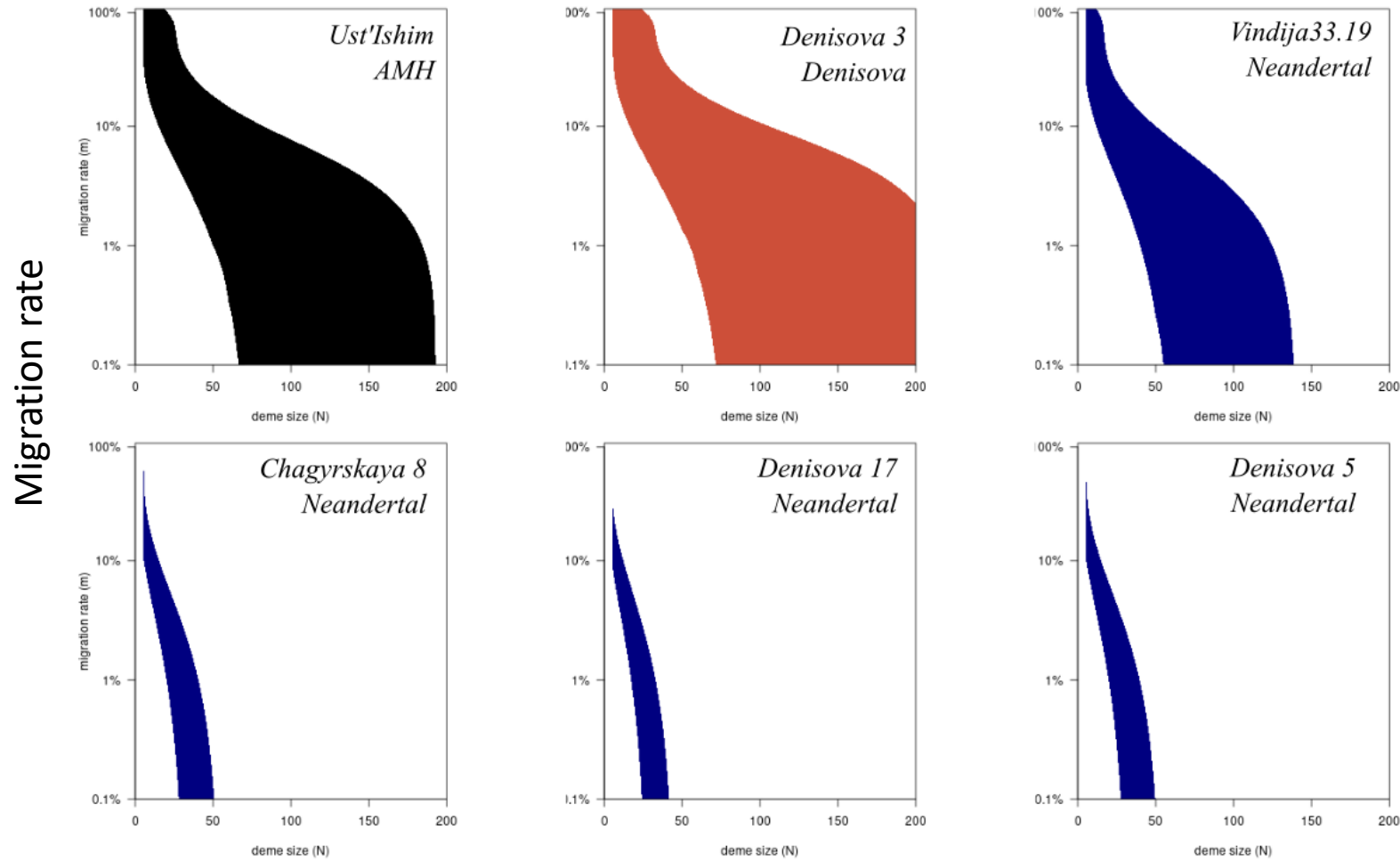
inbreeding



reduction in recent N_e

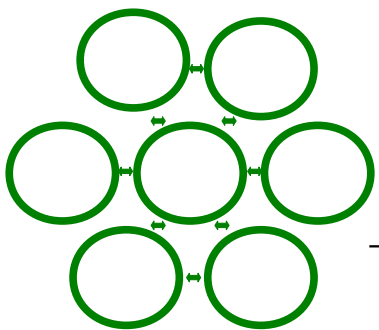


Possiamo usare le regioni in omozigosita' per che i Neanderthal vissero in piccoli gruppi di <50 individuali



Late western Neandertal and others

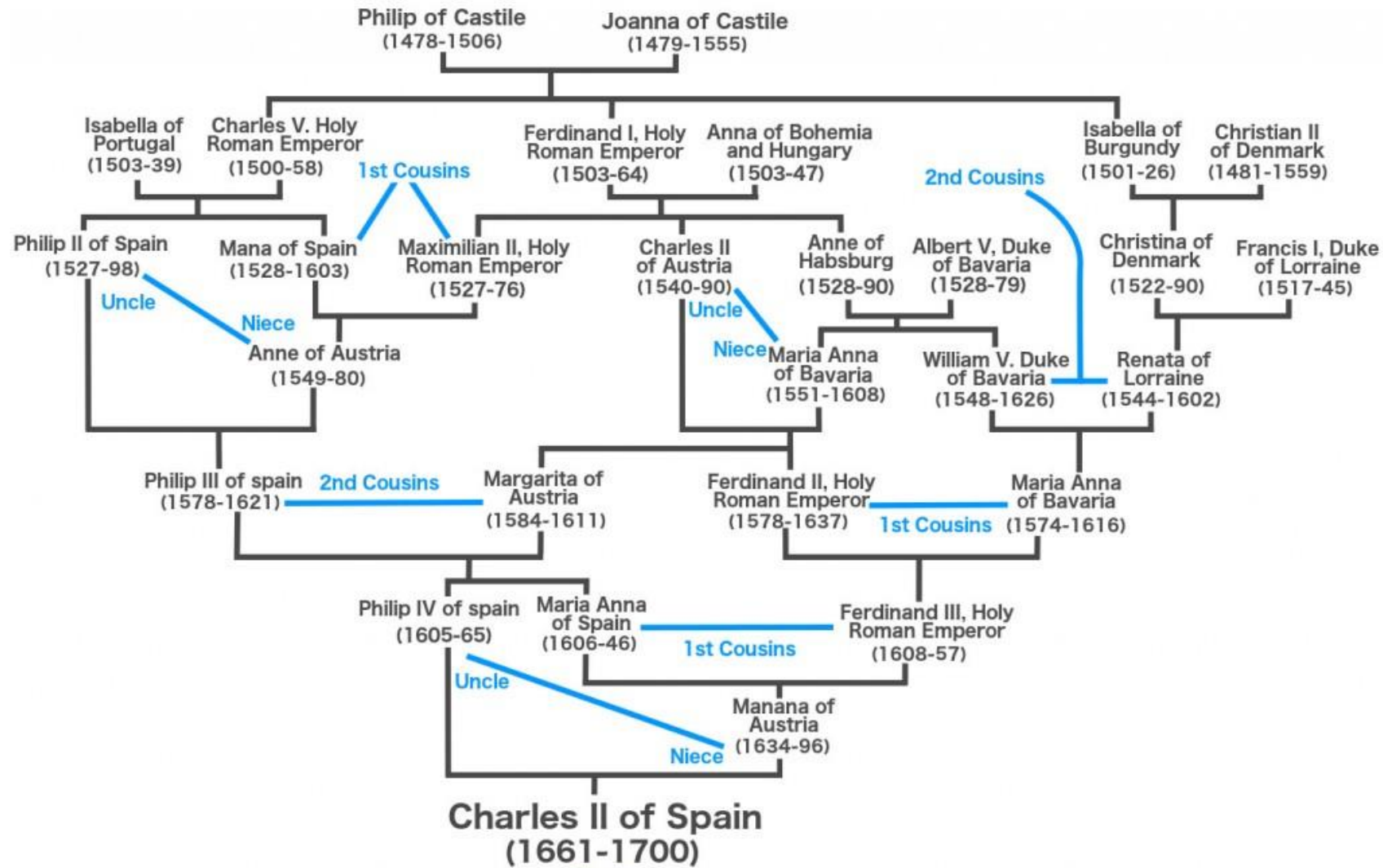
Siberian Neandertals

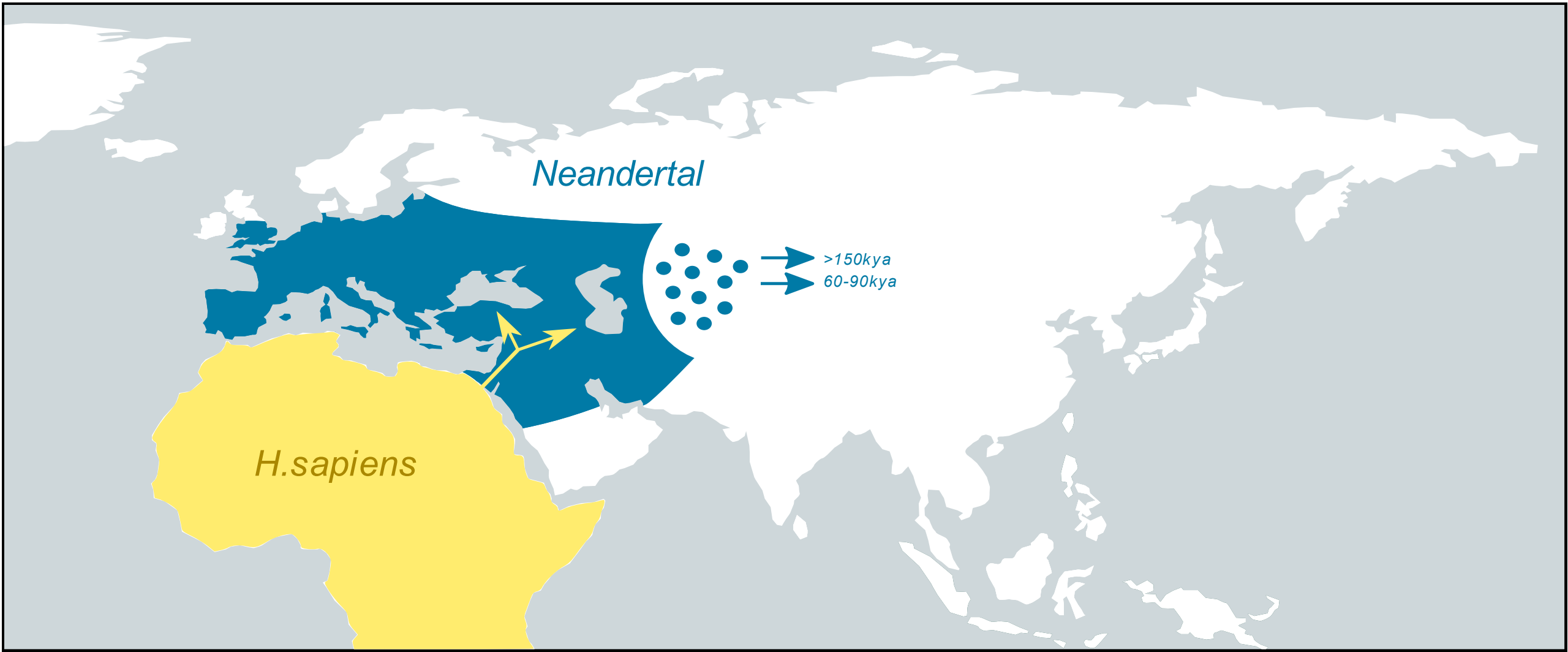


Tasso di migrazione m

Demi di dimensione N individui

Number of individuals per deme

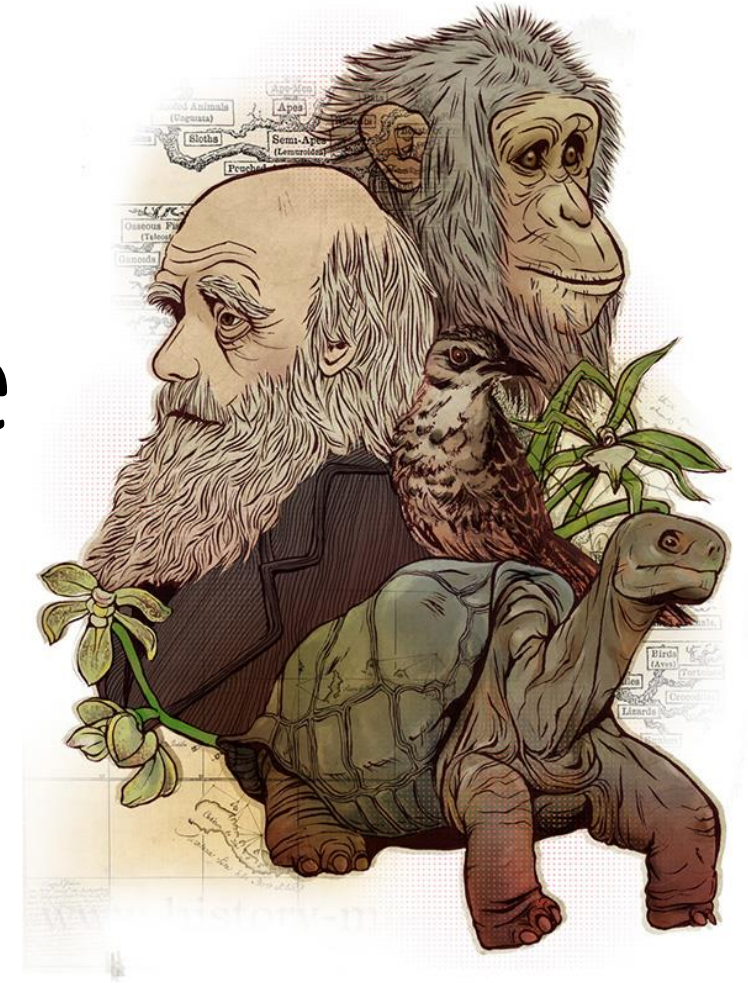




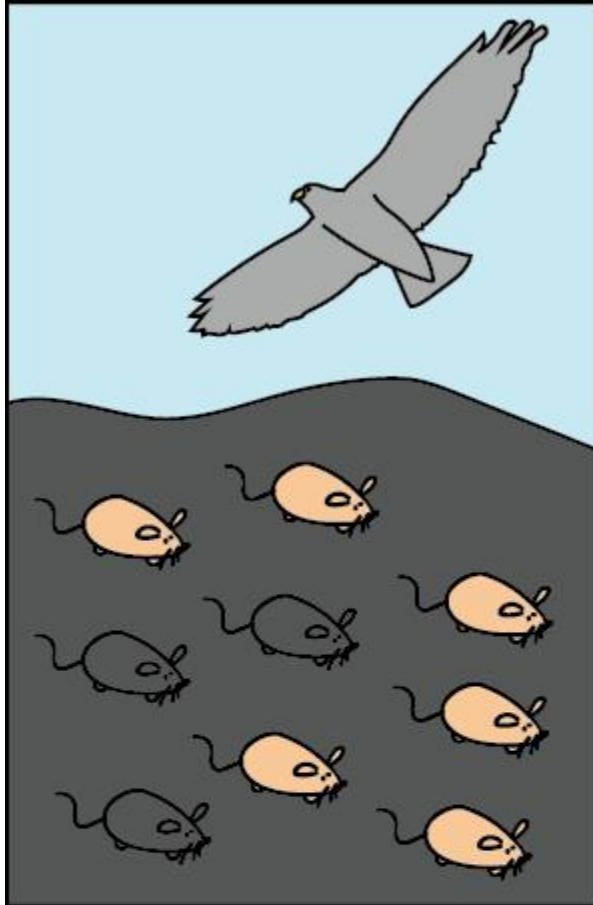
Adapted from Mafessoni 2018, *Nat.Ecol.Evol.*



Selezione naturale



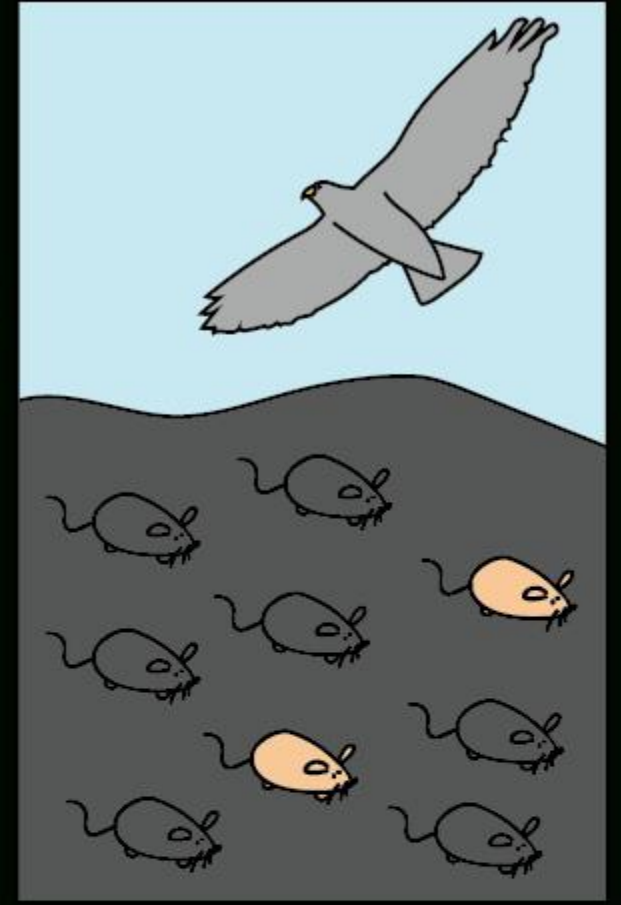
I requisiti della selezione naturale



Variazione nei tratti



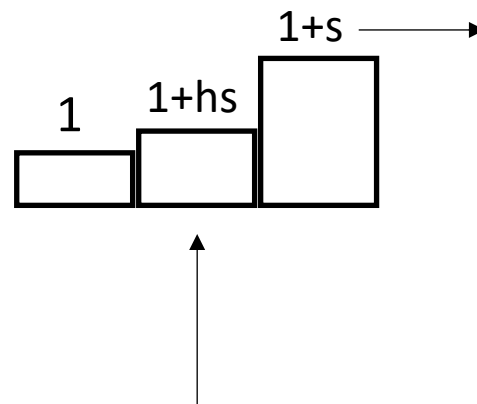
Contributo differenziale alle nuove generazioni (nascita e sopravvivenza)



Ereditarietà dei tratti

Un semplice modello di selezione

Genotipi: BB AB AA



Coefficiente di selezione s : *piu e' alto, Maggiore e' il vantaggio dell'omozigote AA*

Coefficiente di dominanza h :

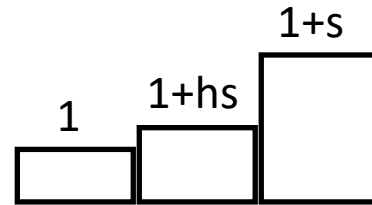
se $h=0$, A e' recessiva (l'eterozigote AB ha la stessa fitness di BB)

se $h=1$, A e' dominante (l'eterozigote AB ha stessa fitness di AA)

se $0 < h < 1$ la fitness di A e' intermedia tra AA e AB (e.g. $h=0.5$, codominante)

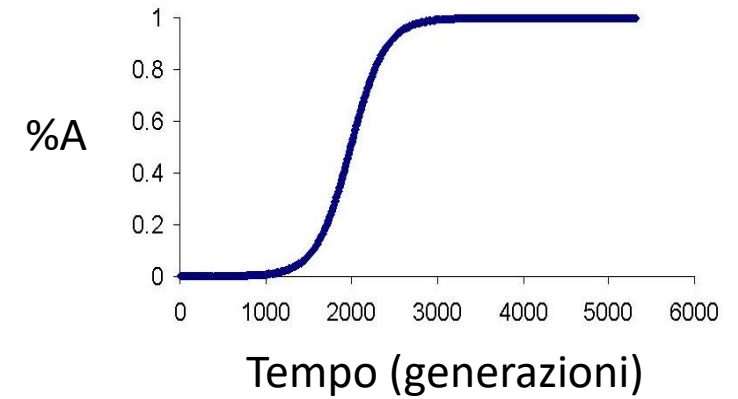
Tre principali modalita' di selezione naturale per un allele A

Genotipi: BB AB AA



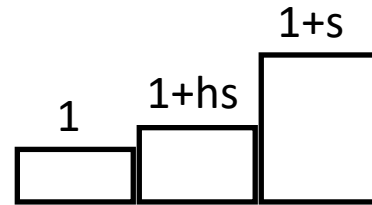
A e' vantaggioso

(selezione positiva: $s>0$, $0<h<1$)



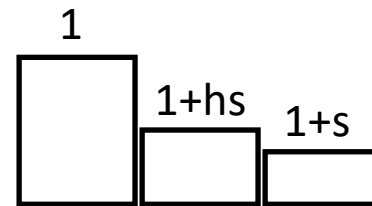
Tre principali modalita' di selezione naturale per un allele A

Genotipi: BB AB AA



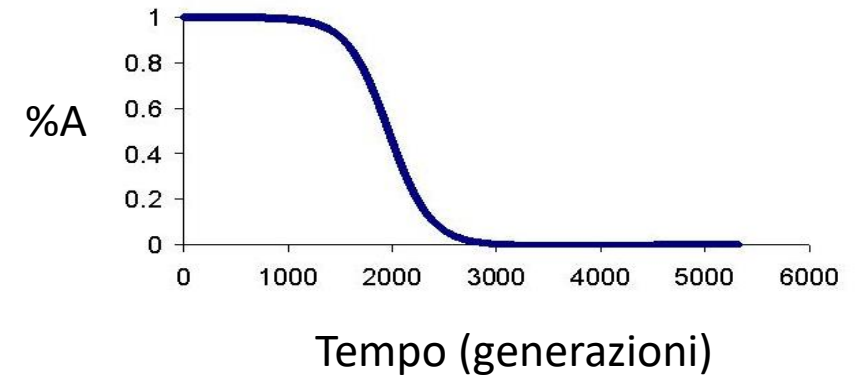
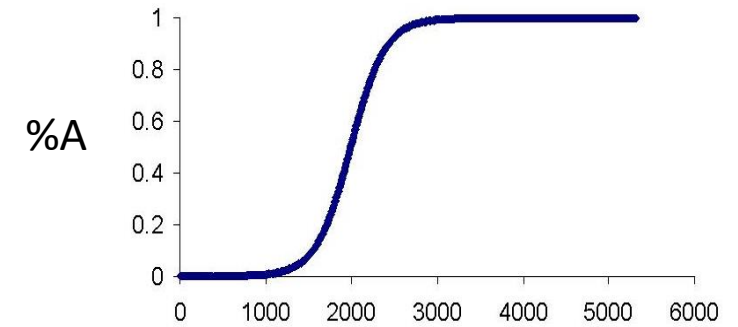
A e' vantaggioso

(selezione positiva: $s > 0$, $0 < h < 1$)



A e' deleterio

(selezione negativa/purificante: $s < 0$ e $0 < h < 1$)

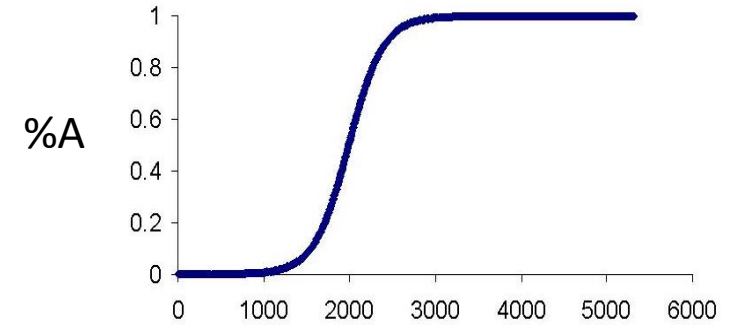
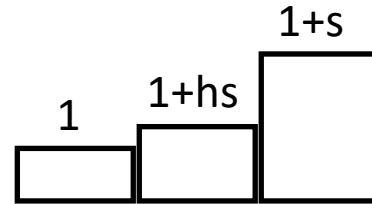


Tre principali modalita' di selezione naturale per un allele A

Genotipi: BB AB AA

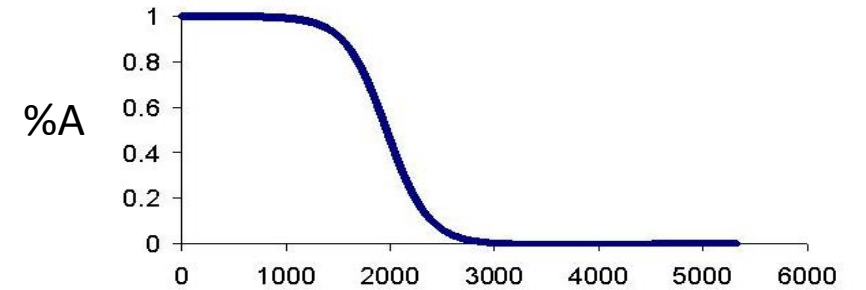
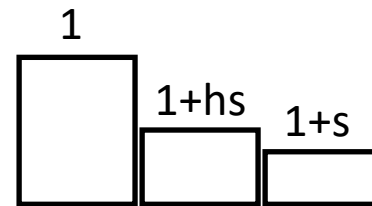
A e' vantaggioso

(selezione positiva: $s > 0$, $0 < h < 1$)



A e' deleterio

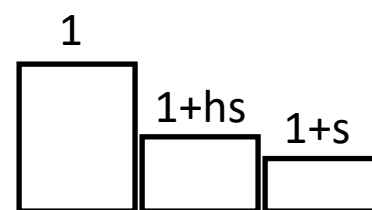
(selezione negativa/purificante: $s < 0$ e $0 < h < 1$)



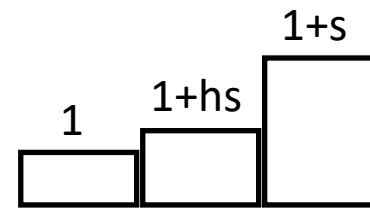
**A e' a volte vantaggioso
e a volte deleterio**

(selezione bilanciata: $s > 0$, $h > 1$)

Contesto 1



Contesto 2

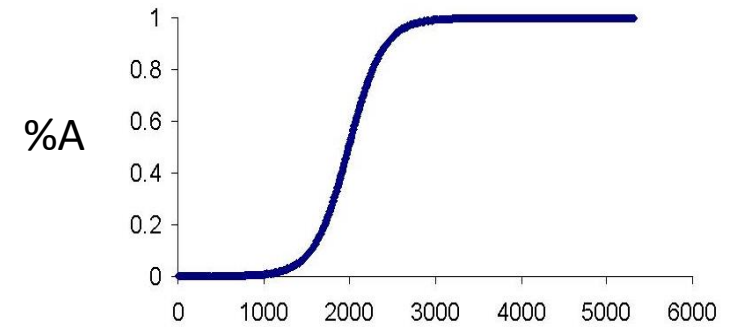
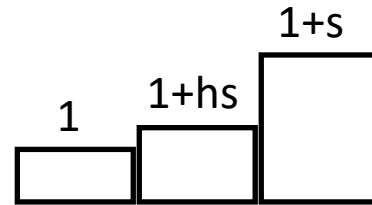


Tre principali modalita' di selezione naturale per un allele A

Genotipi: BB AB AA

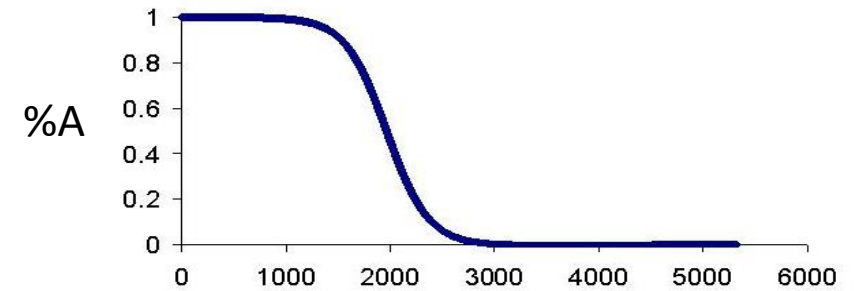
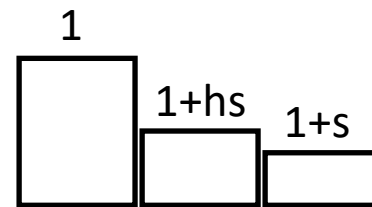
A e' vantaggioso

(selezione positiva: $s > 0$, $0 < h < 1$)



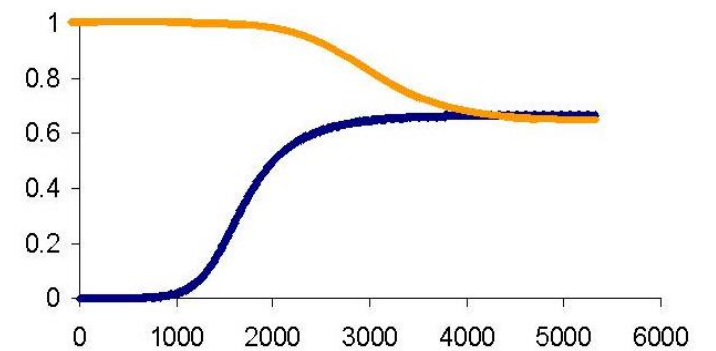
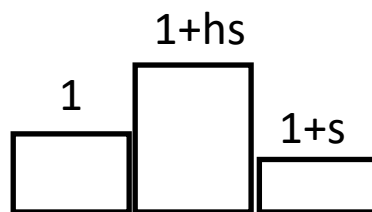
A e' deleterio

(selezione negativa/purificante: $s < 0$ e $0 < h < 1$)



**A e' a volte vantaggioso
e a volte deleterio**

(selezione bilanciata: $s > 0$, $h > 1$)



Tempo (generazioni)

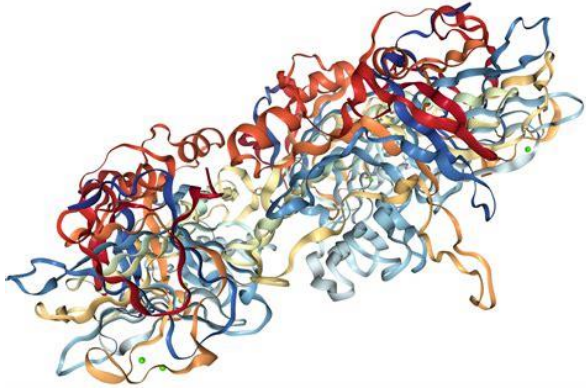
Quale pensate siano le (nuove) mutazioni piu frequenti/la modalita' di selezione piu comune?

Selezione negativa/purificante (su *nuove varianti* deleterie)

Selezione positiva (su *nuove varianti* vantaggiose)

Selezione bilanciata

Quale pensate siano le (nuove) mutazioni piu frequenti/la modalita' di selezione piu comune?



Selezione negativa/purificante (su nuove varianti deleterie)



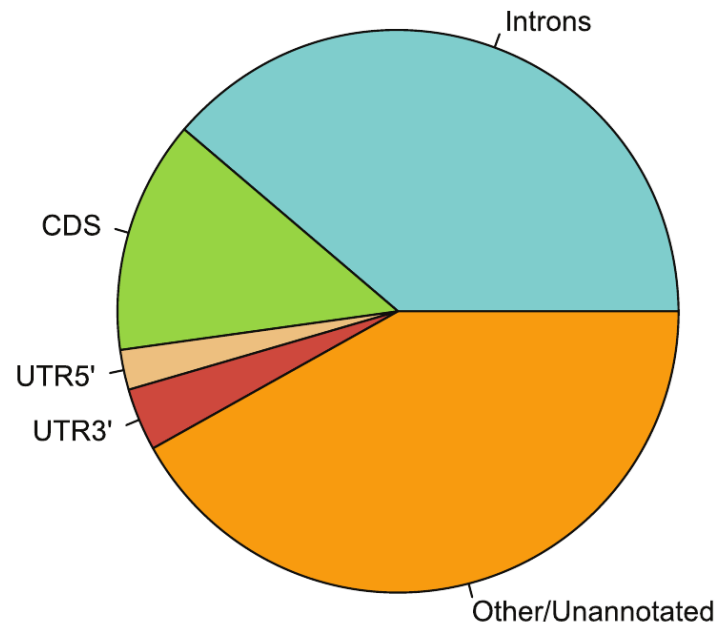
Selezione positiva (su nuove varianti vantaggiose)



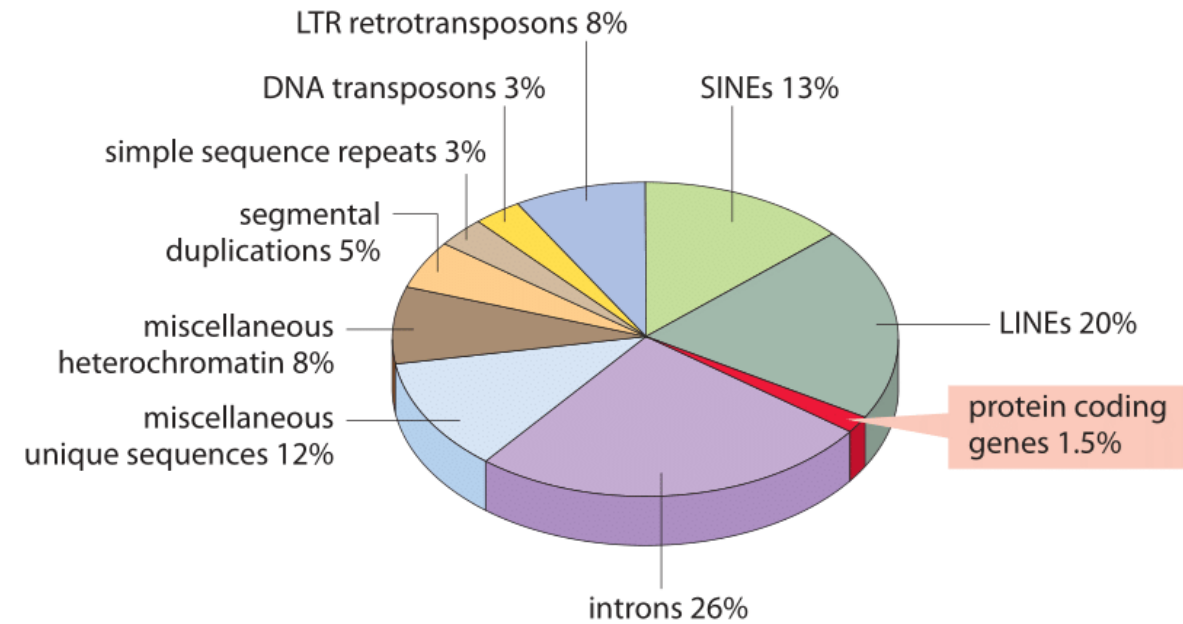
Selezione bilanciata



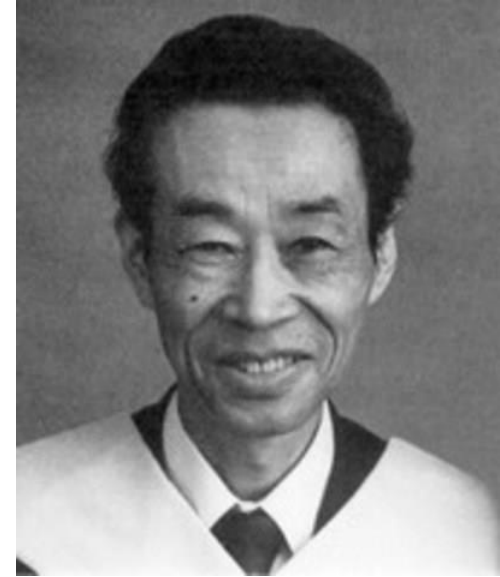
Regioni del genoma identificate come funzionali/constrained



main components of the human genome



The neutral theory of molecular evolution



La maggior parte della variazione all'interno delle singole popolazioni e tra popolazioni e' causata da mutazioni neutrali (o quasi neutrali).



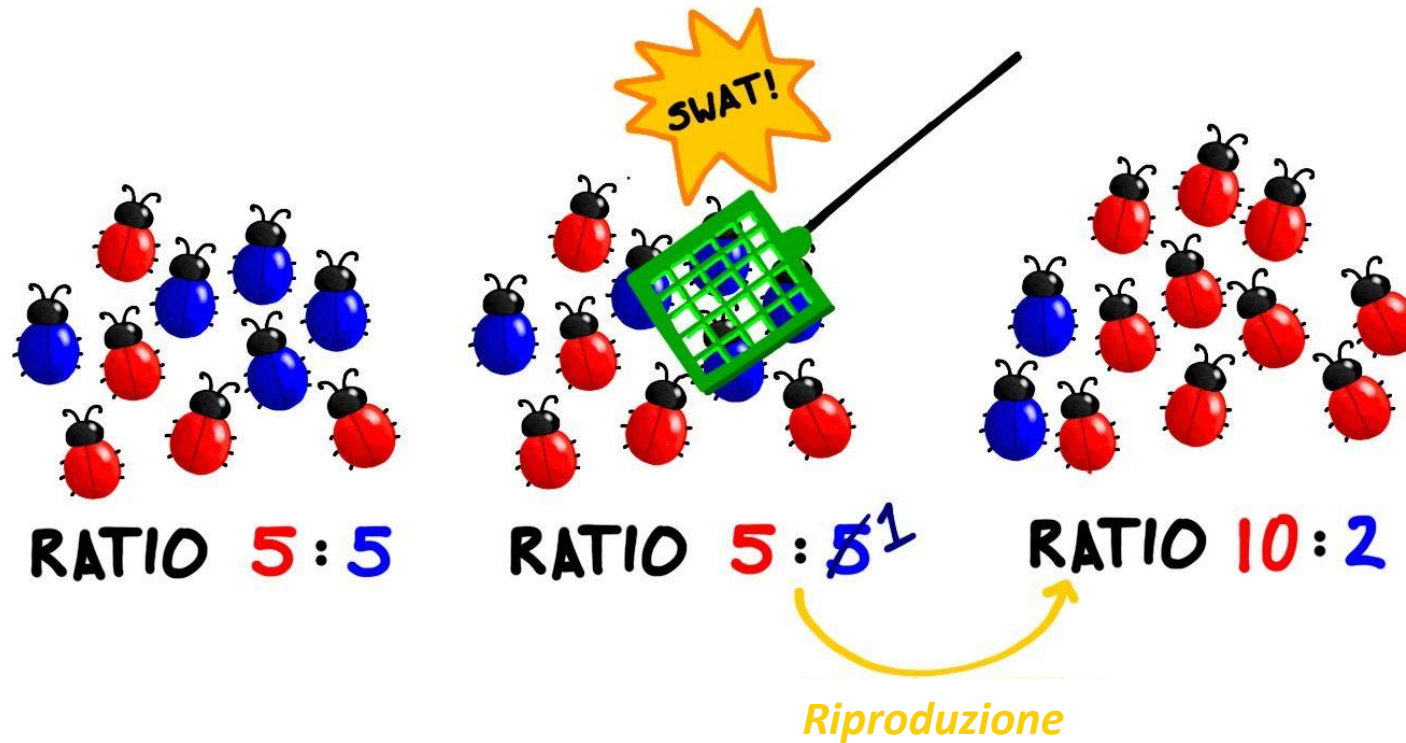
Alla base della Teoria degli orologi molecolari. Suggerisce che il grosso del genoma puo' essere considerato (con eccezioni) un background neutrale per test statistici, e che l'ipotesi nulla per un allele sotto studio e' che sia neutrale.

Motoo Kimura

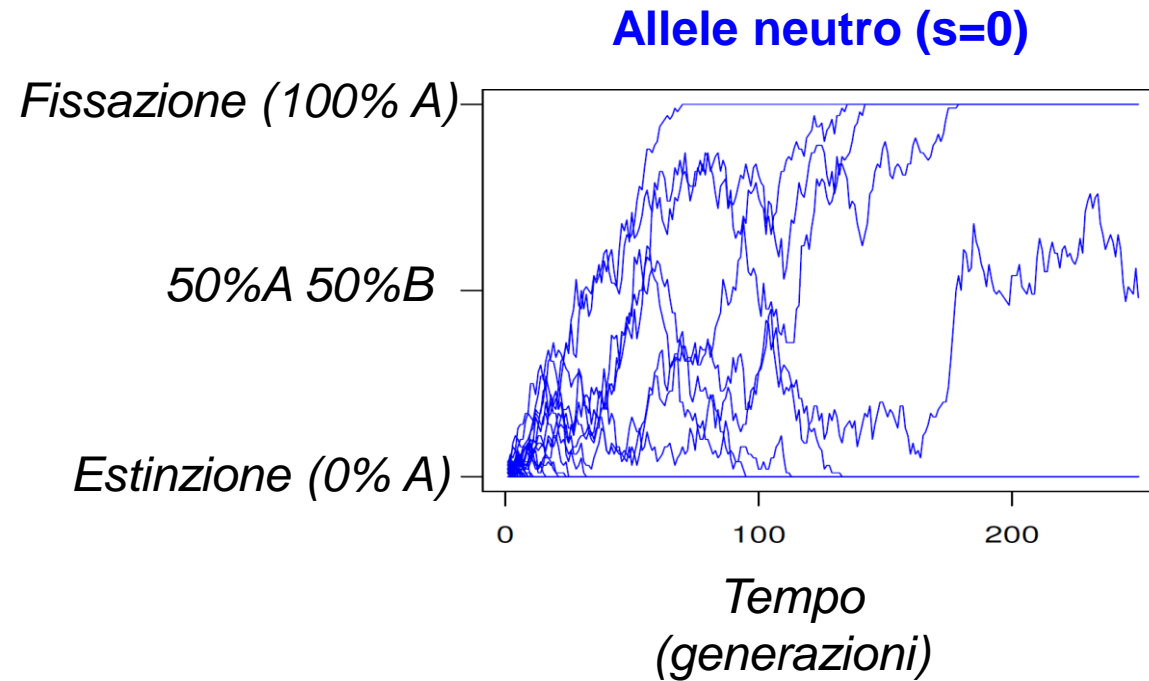
La selezione naturale non agisce da sola

Deriva genetica

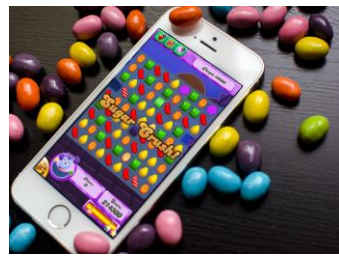
un cambiamento nelle frequenze alleliche dovuto ad eventi casuali



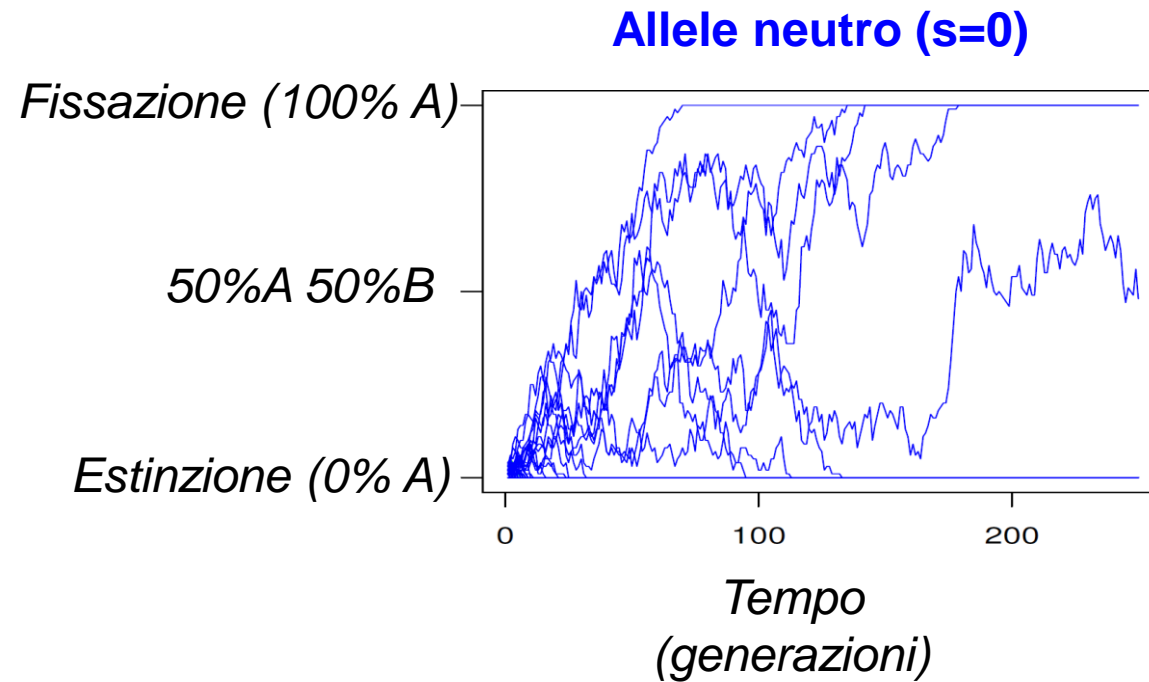
Alleli possono raggiungere la fissazione ed estinguersi per deriva genetica



Se volete esplorare l'effetto della deriva genetica con simulazioni Wright-Fisher
<https://chrisnajman.github.io/genetic-drift/>
(invece di Candy Crush)

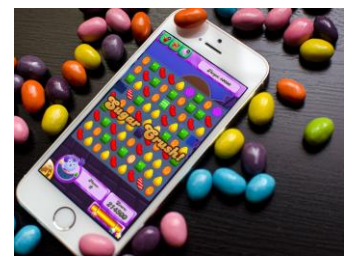


Qual'è la probabilità di fissazione per una singola nuova mutazione (frequenza $1/2N$) in una popolazione di N individui?

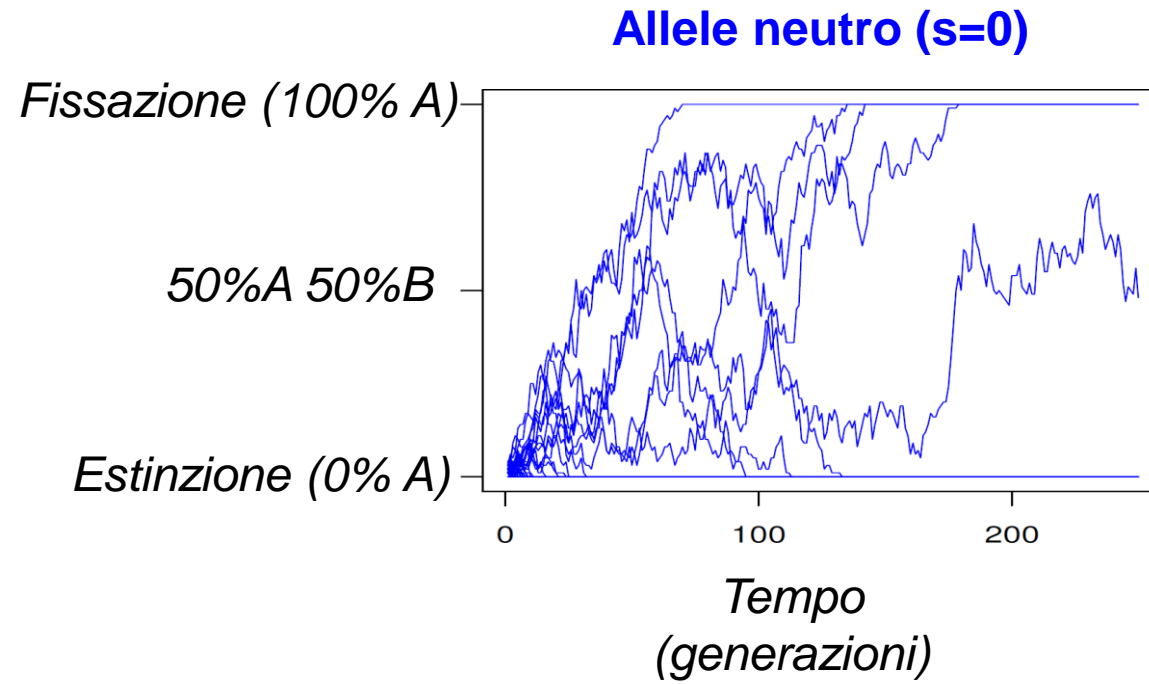


- $1/2N$
- $1/N$
- 1%

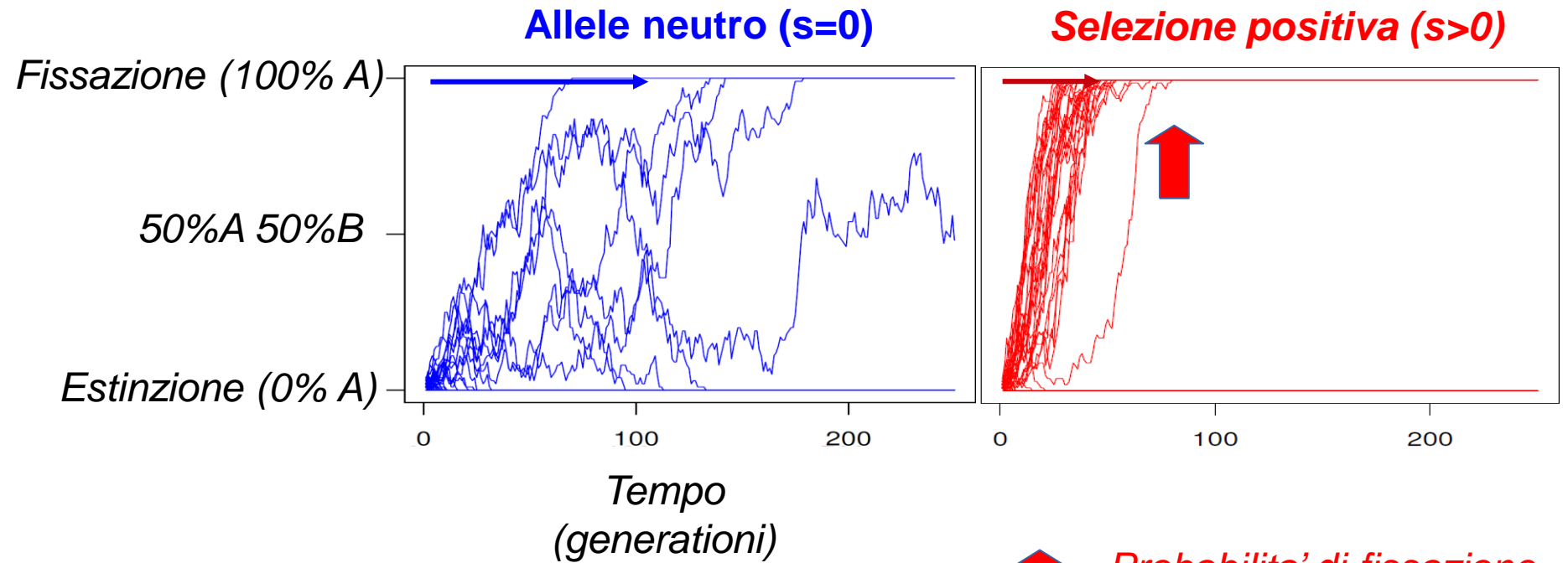
Se volete esplorare l'effetto della deriva genetica con simulazioni Wright-Fisher
<https://chrisnajman.github.io/genetic-drift/>
(invece di Candy Crush)



Qual'è la probabilità di fissazione per una singola nuova mutazione (frequenza $1/2N$) in una popolazione di N individui?

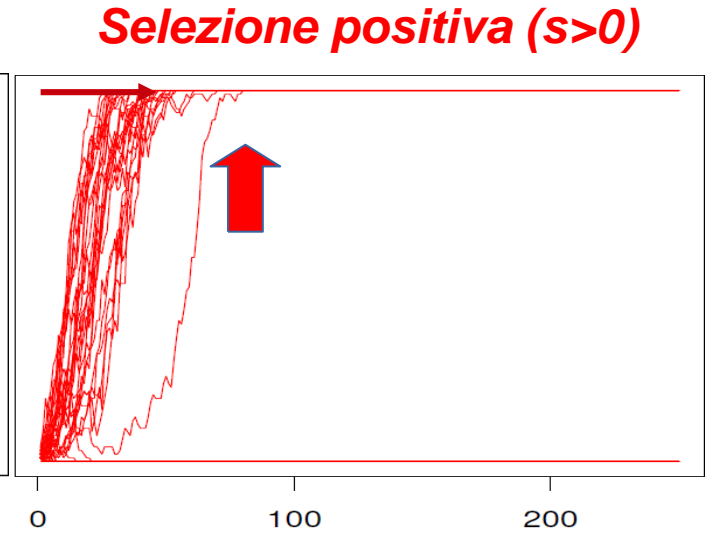
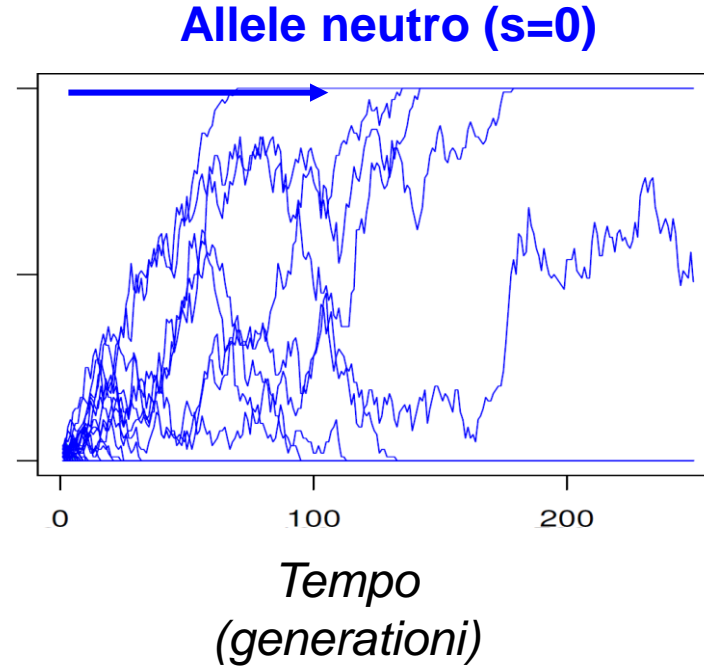
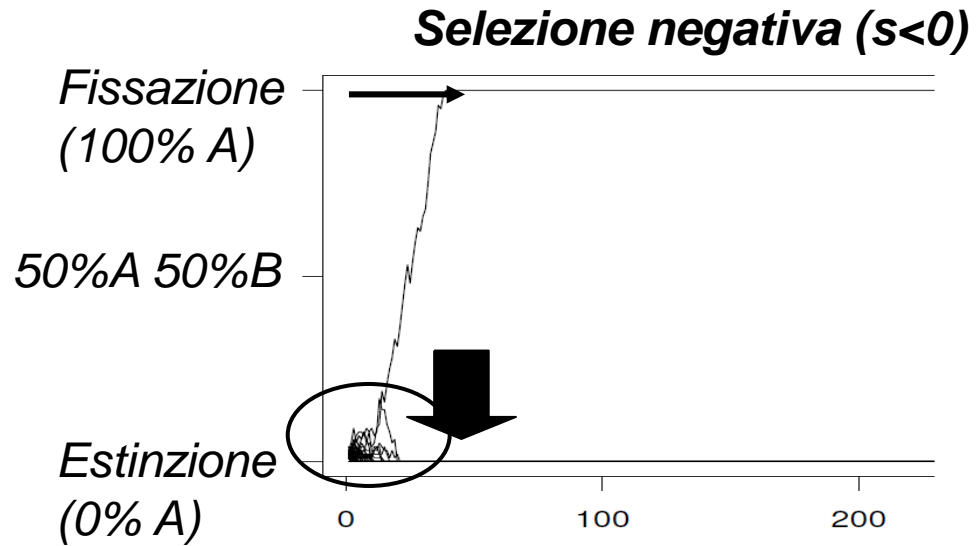


- $1/2N$
- $1/N$
- 1%



↑ Probabilità di fissazione più alta

→ Tempo di fissazione più rapido



↑ Probabilità di estinzione più alta

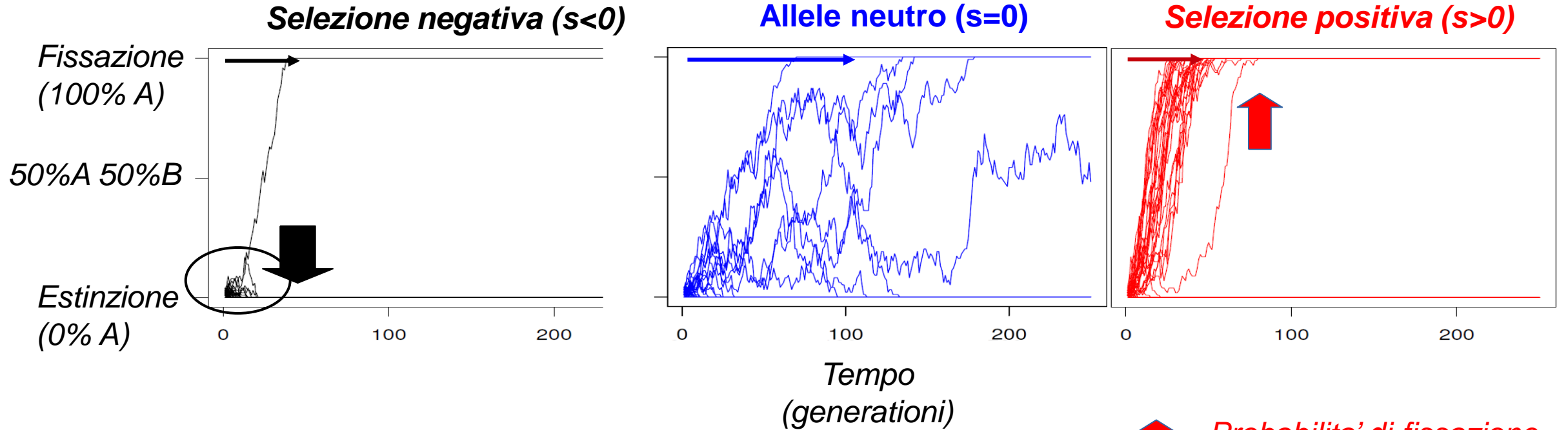
↓

○ Frequenze di A più basse

→ Tempo di fissazione più rapido

↑ Probabilità di fissazione più alta

→ Tempo di fissazione più rapido



↑ Probabilità di estinzione più alta
↓
○ Frequenze di A più basse
→ Tempo di fissazione più rapido

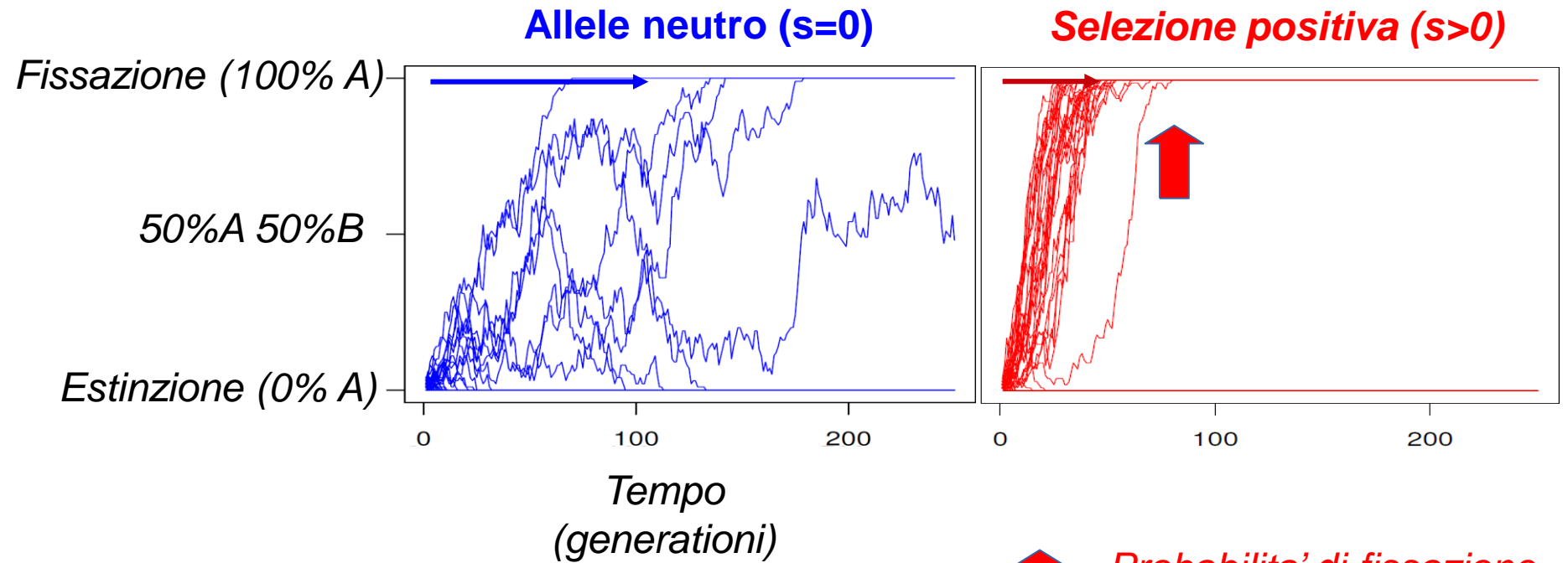
↑ Probabilità di fissazione più alta
→ Tempo di fissazione più rapido

Se aveste delle sequenze genetiche di una popolazione, come indaghereste se questa popolazione ha subito selezione positiva?

Selezione naturale positiva

E come trovereste le regioni del genoma che sono state oggetto di selezione naturale?

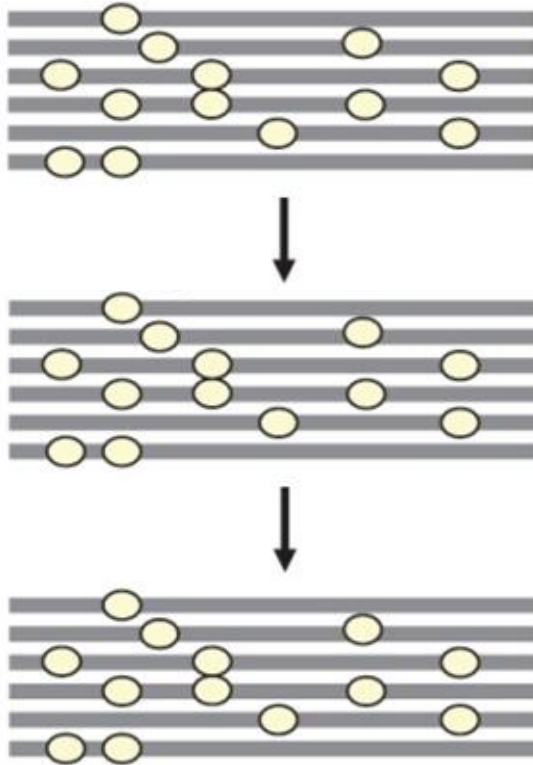
Dinamiche della selezione positiva



↑ Probabilità di fissazione più alta

→ Tempo di fissazione più rapido

Regione neutrale/lontano
dalla mutazione vantaggiosa

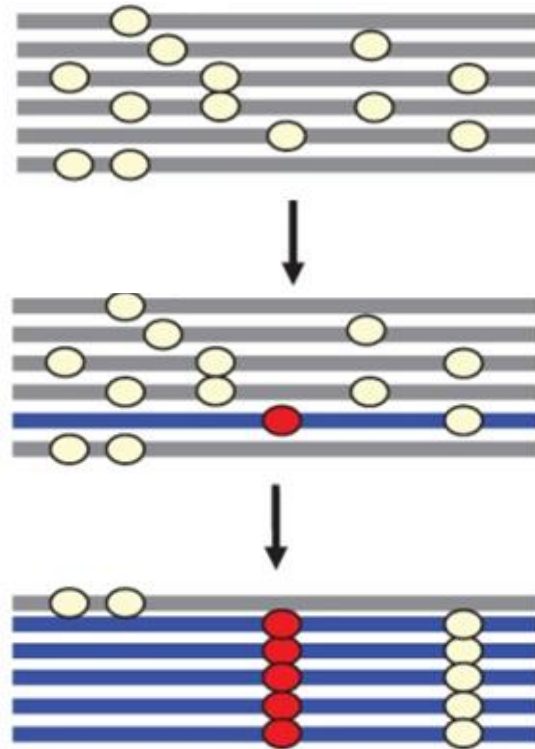


Basso linkage.

Diversita' genetica
non ridotta.

No haplotype
homozigosity

Regione sotto selezione



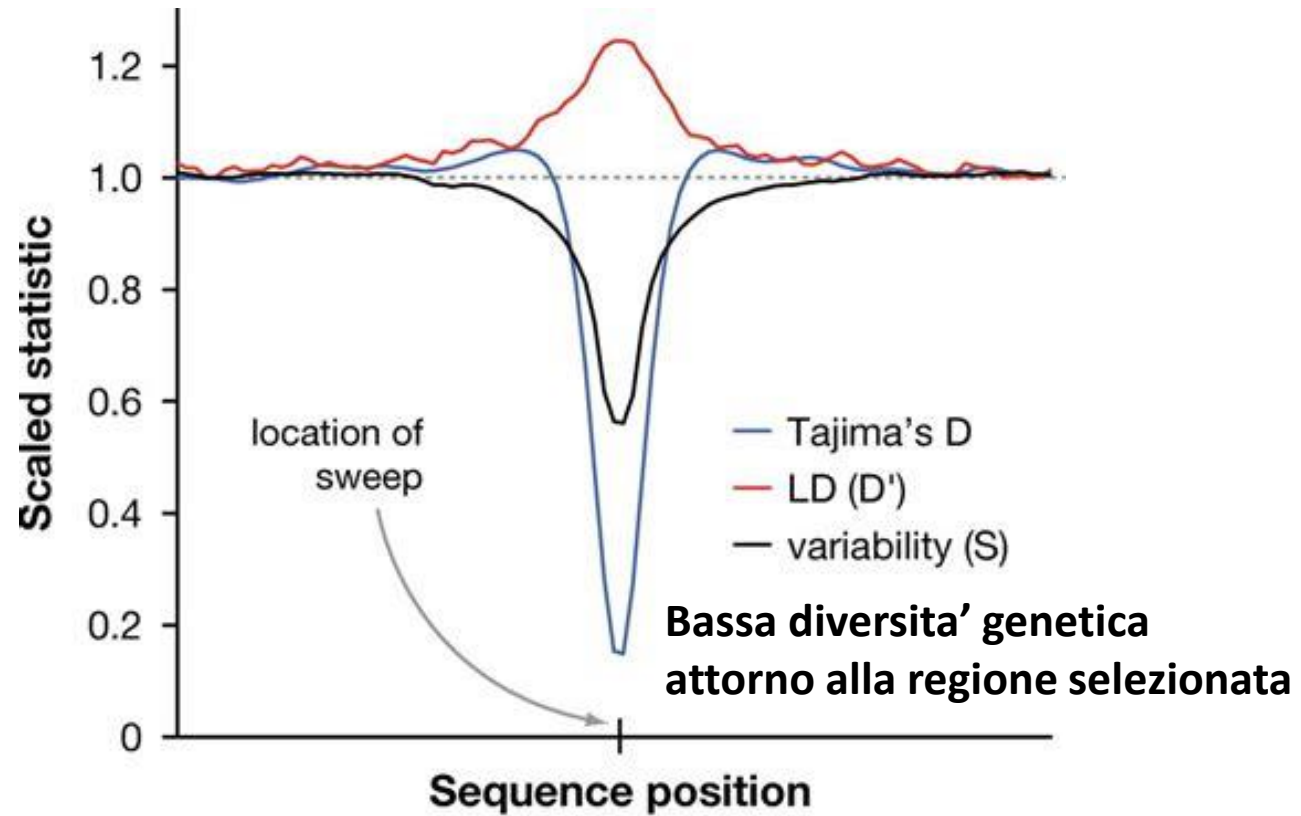
Alto linkage con aplotipo
con variante sotto selezione

Ridotta diversita' genetica

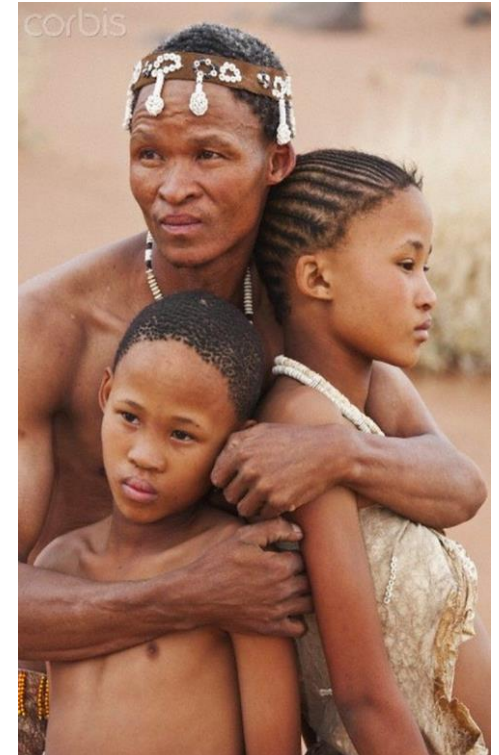
Extended haplotype
homozygosity
(EHH)

Selective sweep

La variante ● si diffonde
rapidamente nella popolazione
senza che ricombinazione e
nuove mutazioni abbiano il
tempo di mantenere diversita'
genetica.

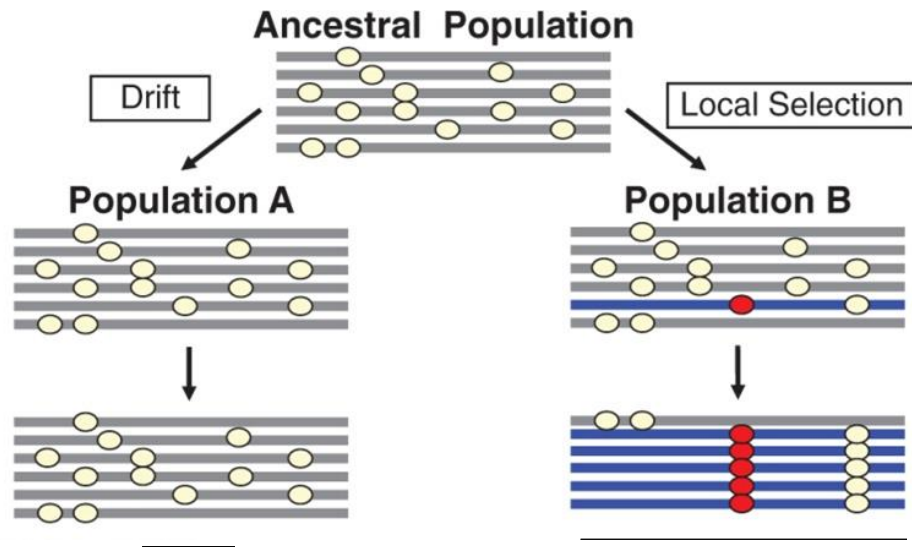


Nielsen R. 2005.
Annu. Rev. Genet. 39:197–218



Gene *SLC24A5* nei Khoe-San
Lin et al., 2018, *PNAS*

Selezione positiva e selective sweeps possono determinare la differenziazione tra popolazioni



Basso linkage.

Diversita' genetica non ridotta.

No haplotype homozigosity

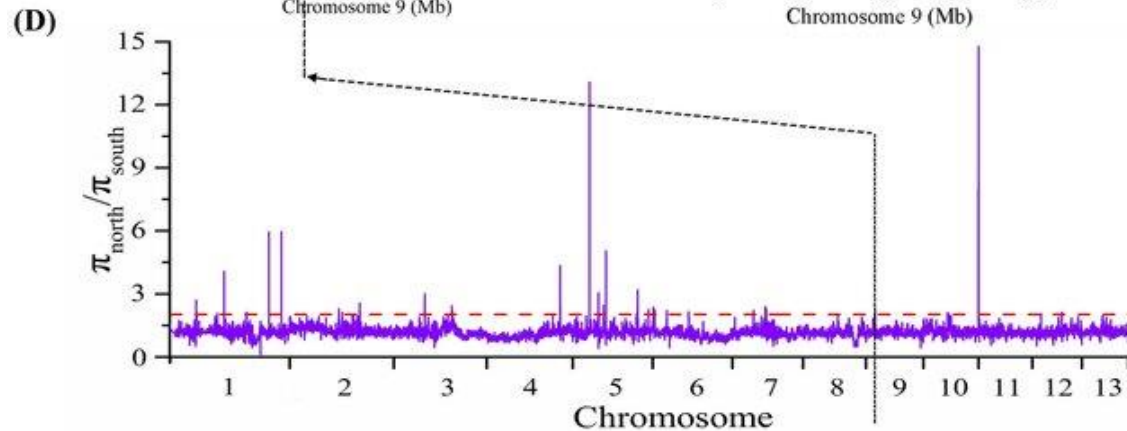
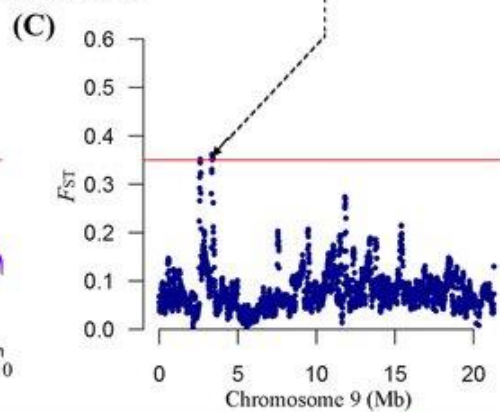
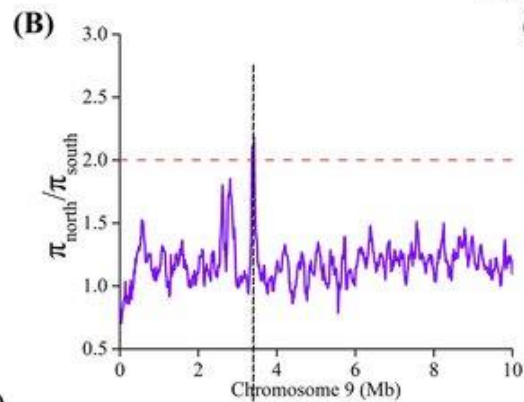
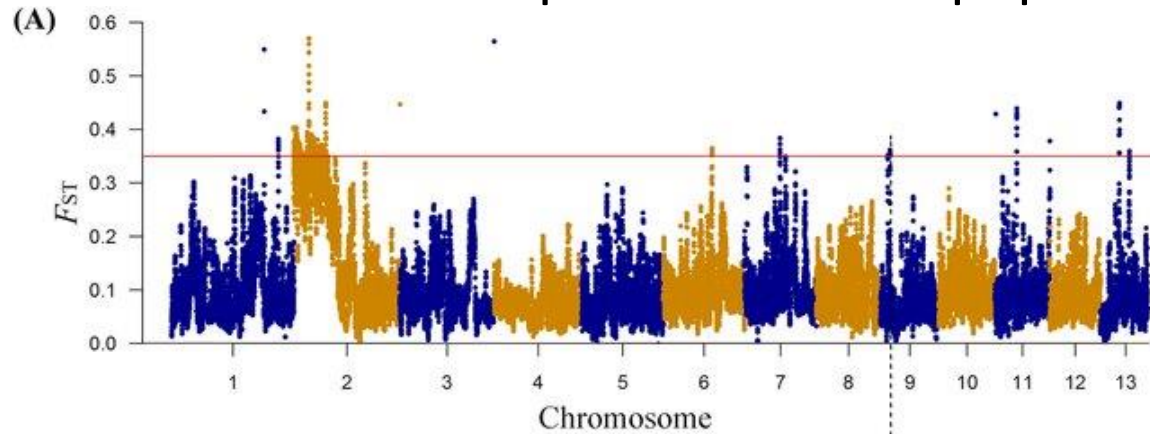
Alto linkage con aplotipo con variante sotto selezione

Ridotta diversita' genetica

Extended haplotype homozigosity (EHH)

Klassman et al., Plos One, 2022 per una review sui vari metodi (XP-EHH, etc.)

L' F_{ST} puo' essere utilizzato per individuare regioni sotto selezioni positiva in sottopopolazioni

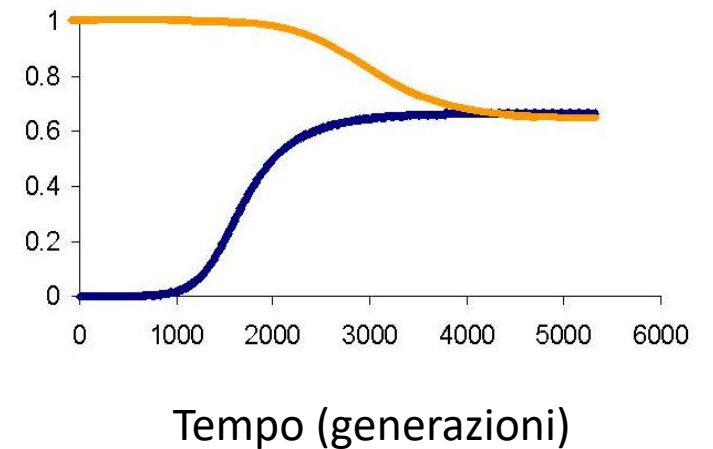
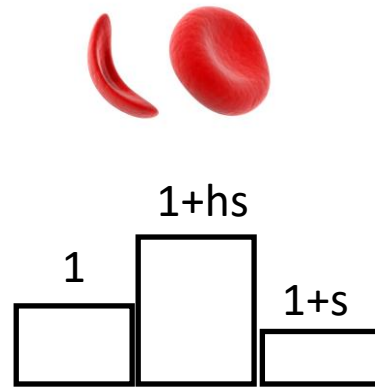


Regioni con alto F_{ST} e differenza in diversita' nucleotidica tra sottopopolazioni indicano adattamento a differenti temperature in popolazioni di gelso da carta

Selezione bilanciata

- Mantiene diversita' vantaggiosa nelle popolazioni

A e' a volte vantaggioso
e a volte deleterio
(selezione bilanciata: $s > 0$, $h > 1$)



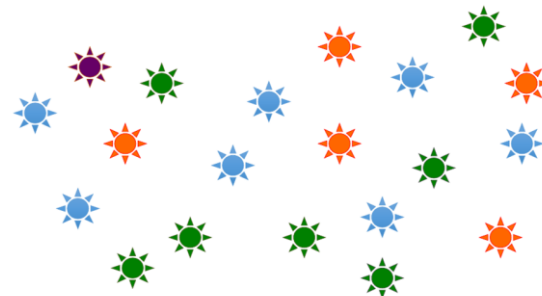
Selezione bilanciata

- Mantiene diversita' vantaggiosa nelle popolazioni
- Molti potenziali meccanismi:



- Vantaggio dell'eterozigote
(anemia falciforme)

Selezione frequenza dipendente
(piu una variante e' frequente, meno il vantaggio selettivo)

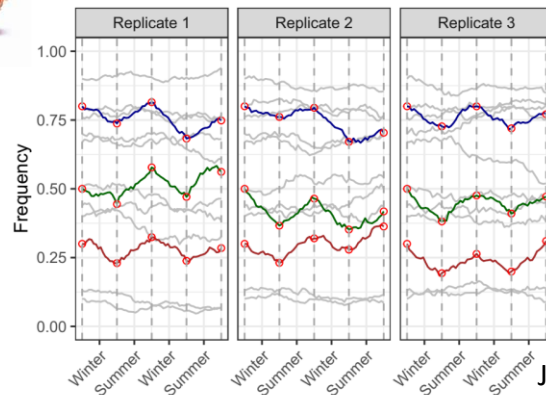


S-locus per la
autoincompatibilita'

Castric et al.,2014, Mol.Ecol
Ioerger et al,1990,PNAS



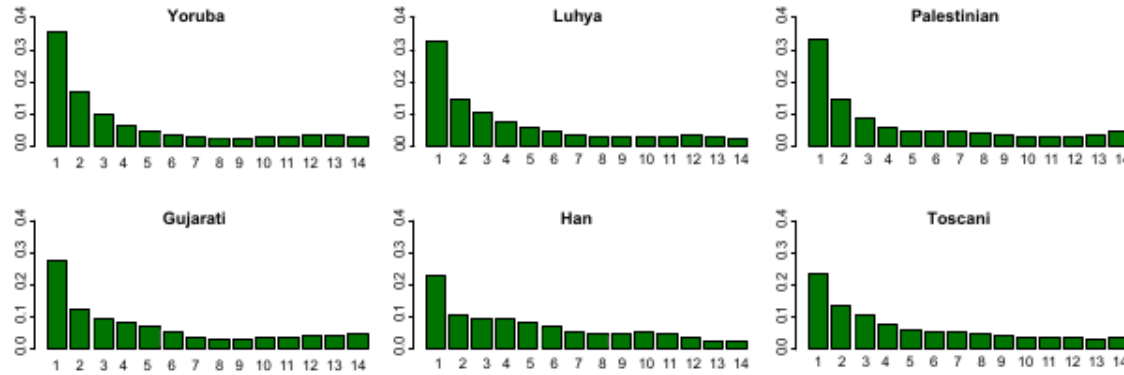
- Selezione fluttuante



Johnson et al.,2023, Trends in Genetics

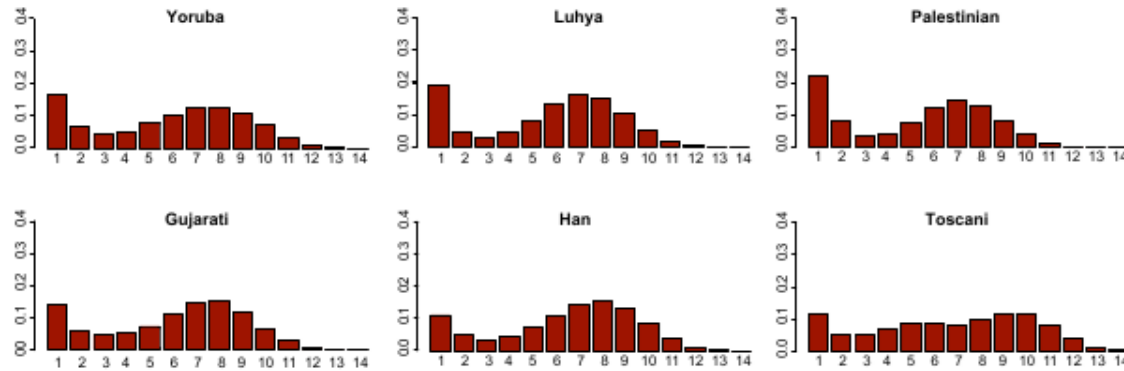
Selezione bilanciata

**CONTROL
REGIONS**



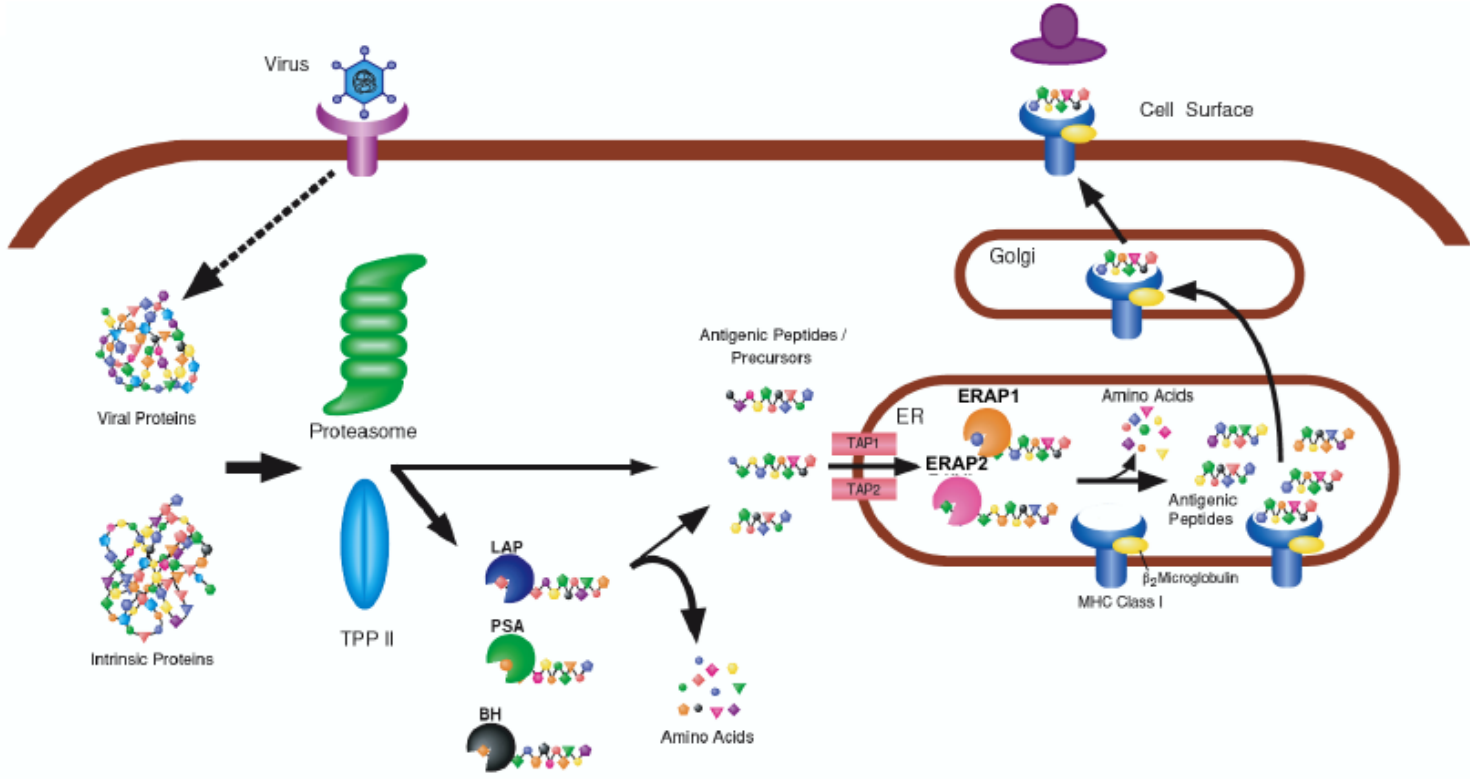
0.82 SNPs per
fixed difference

ERAP2

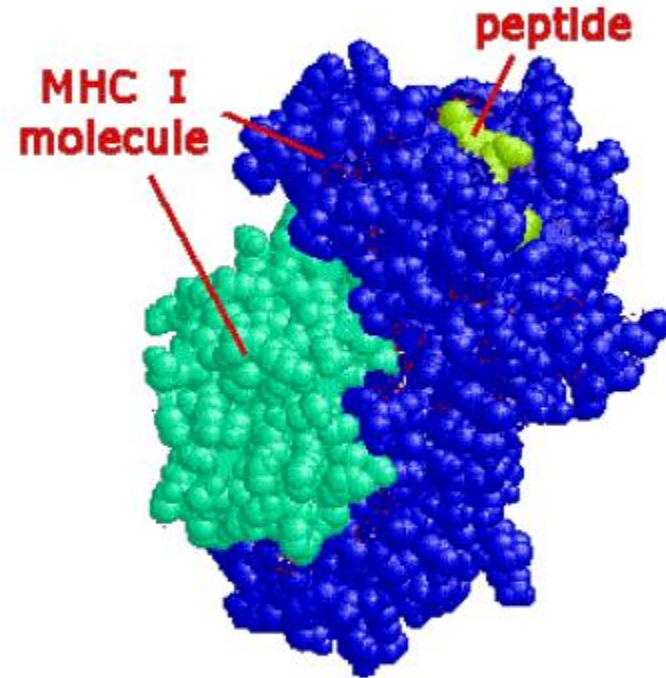


2.2 SNPs per
fixed difference

Selezione bilanciata – complesso Maggiore di istocompatibilita' (MHC)



Selezione bilanciata – complesso Maggiore di istocompatibilita' (MHC)

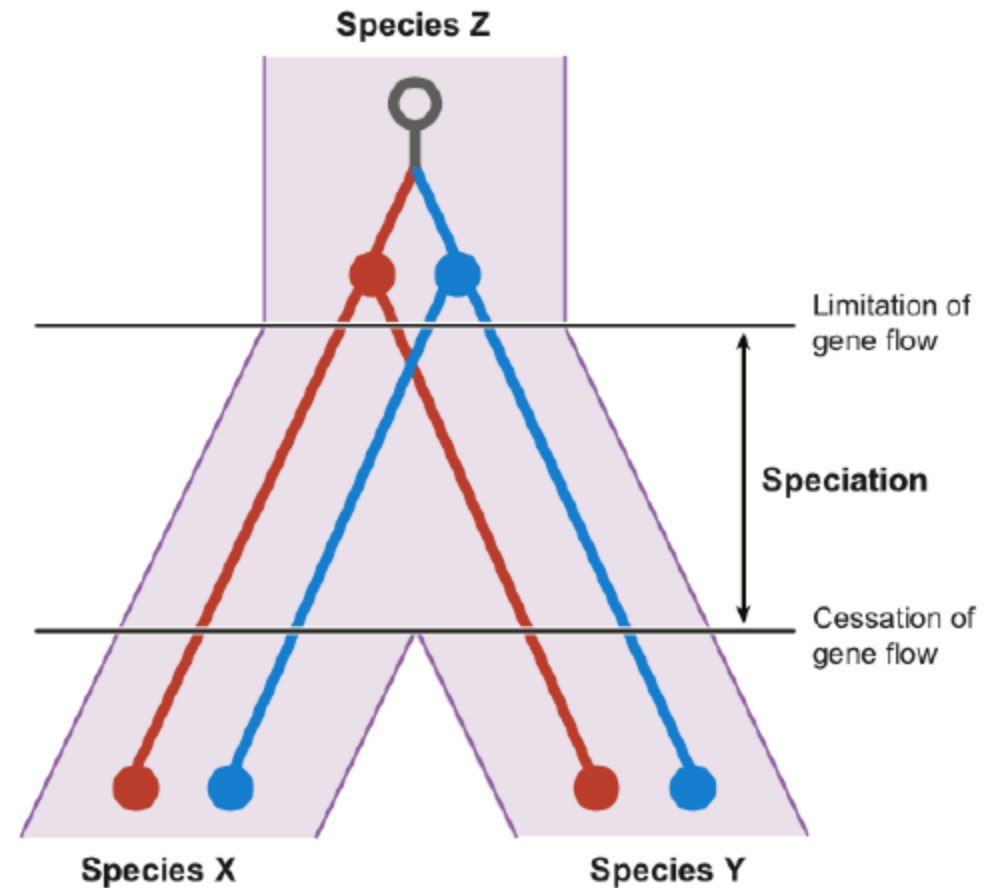


MHC (or HLA in humans)

Thousands of different haplotypes identified in humans

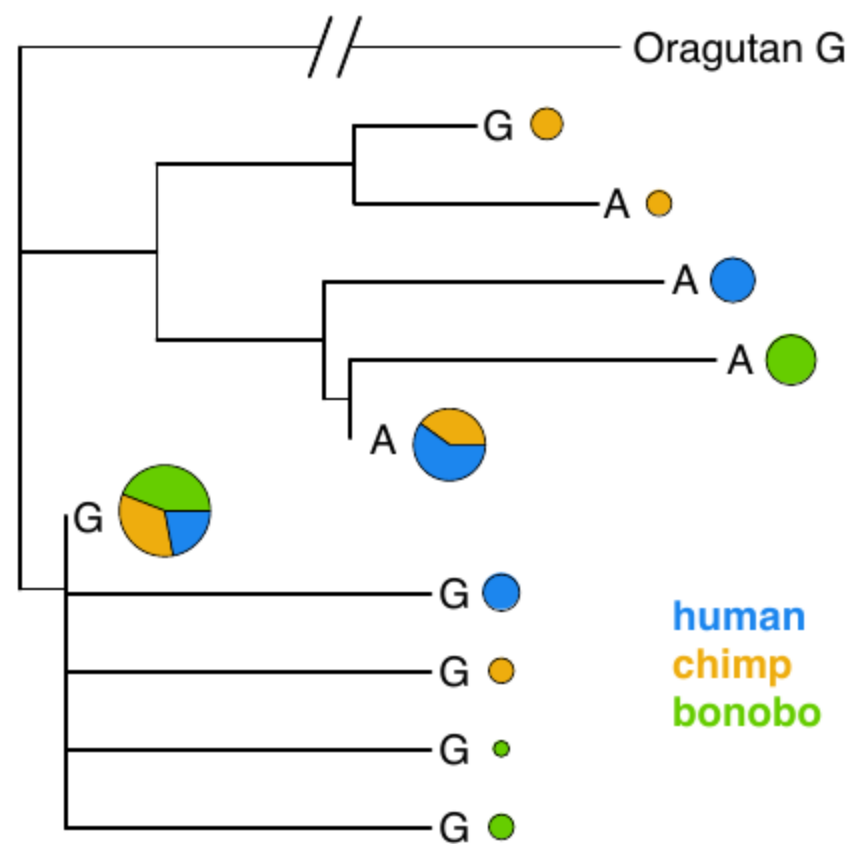
Extreme variability in virtually all species

Trans-species polymorphisms

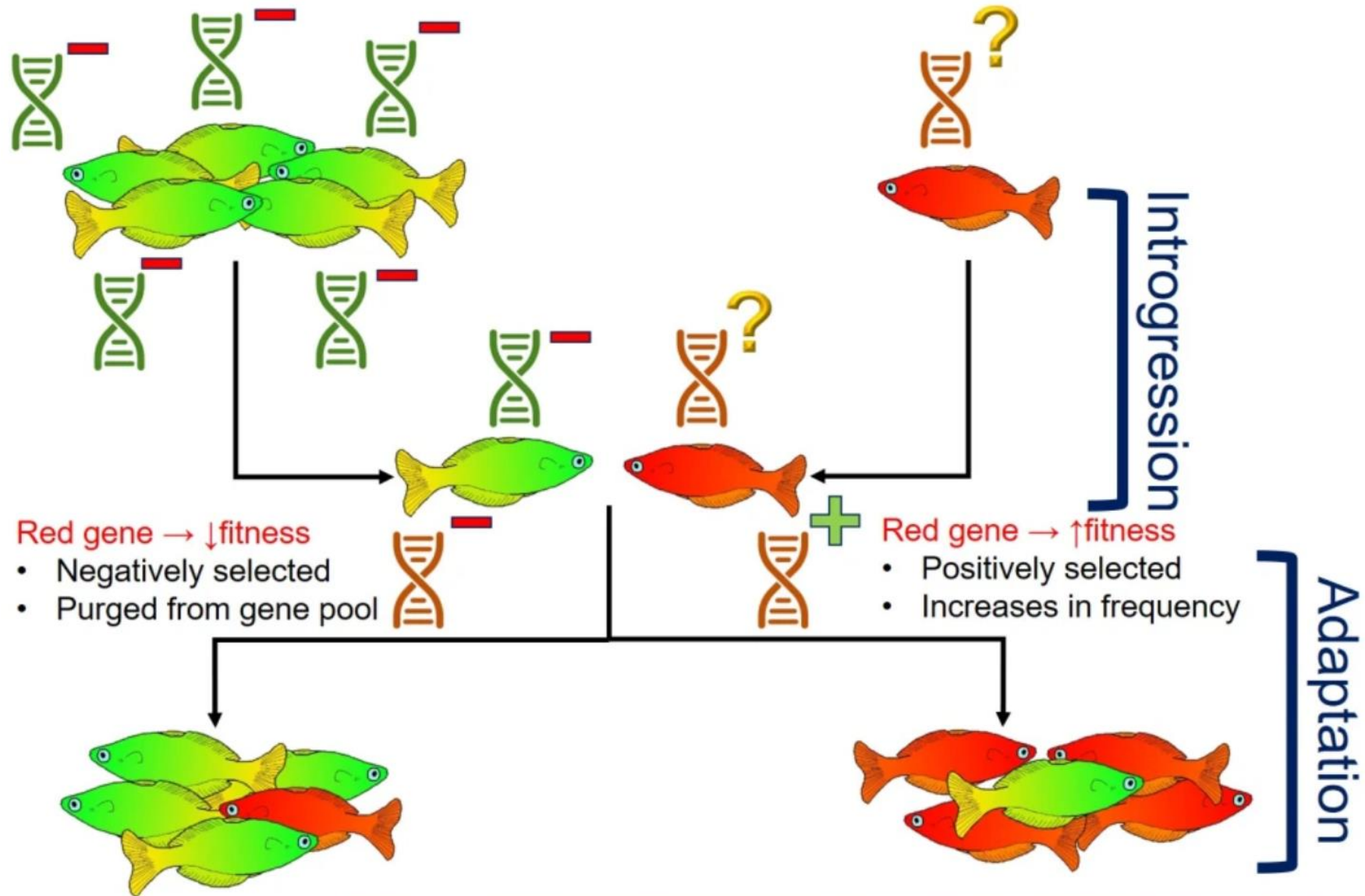


Trans-species polymorphism

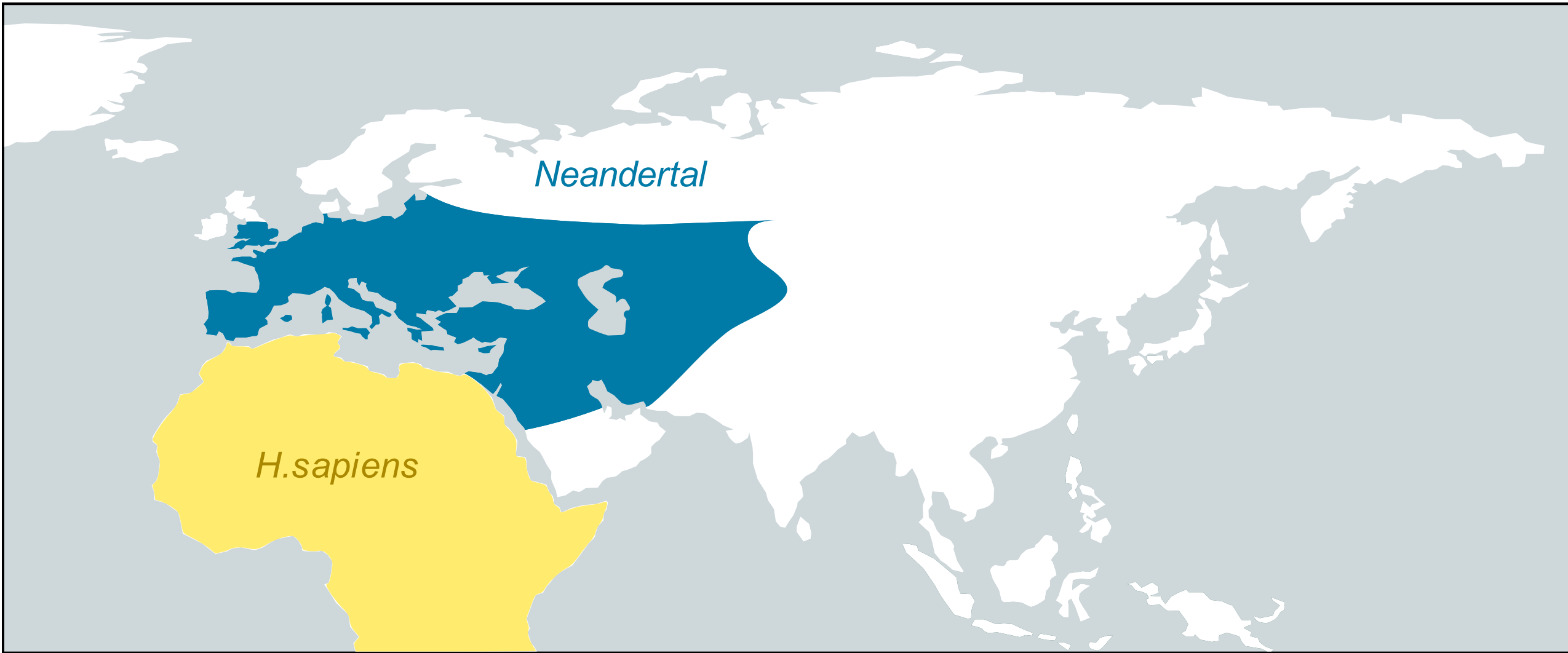
Trans-species polymorphisms – *LAD1*



Flusso genico e selezione



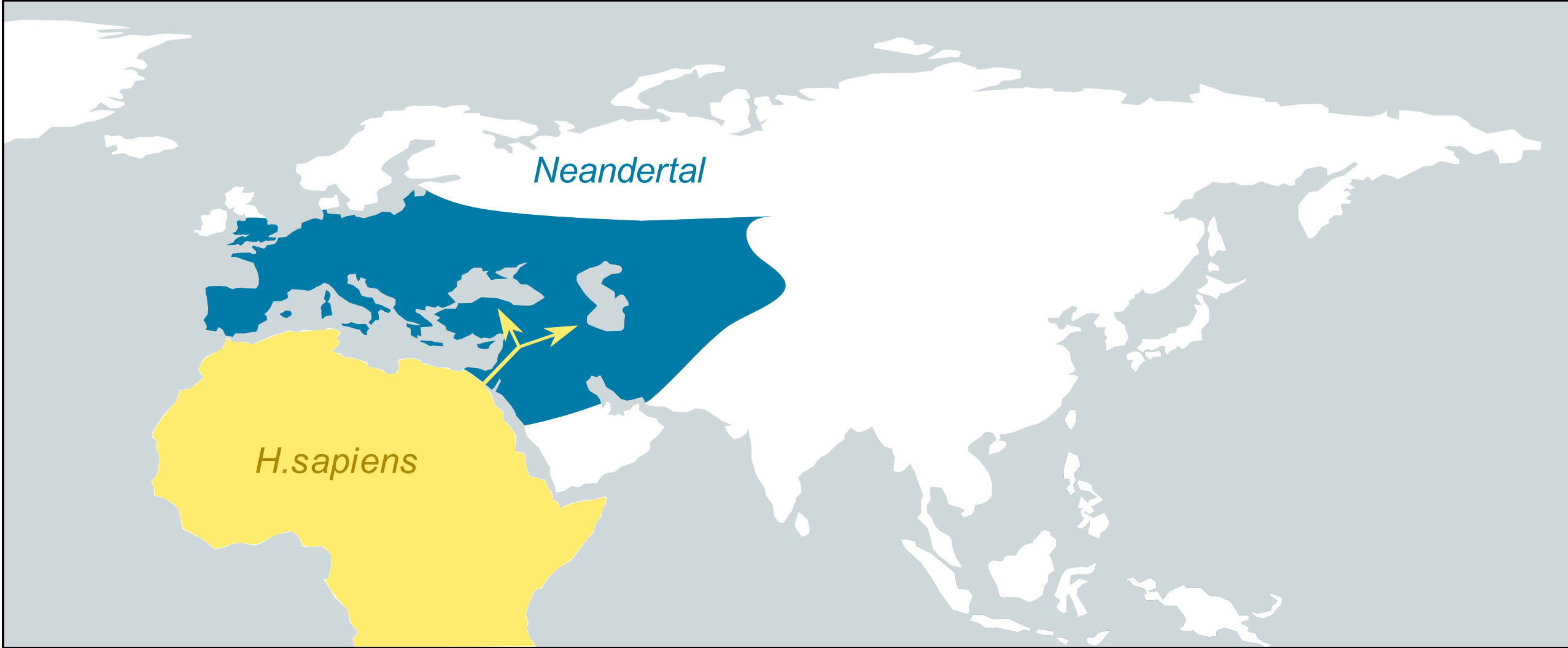
150.000 BC



Adapted from Mafessoni 2018, *Nat.Ecol.Evol.*

Green et al., 2010, *Science*

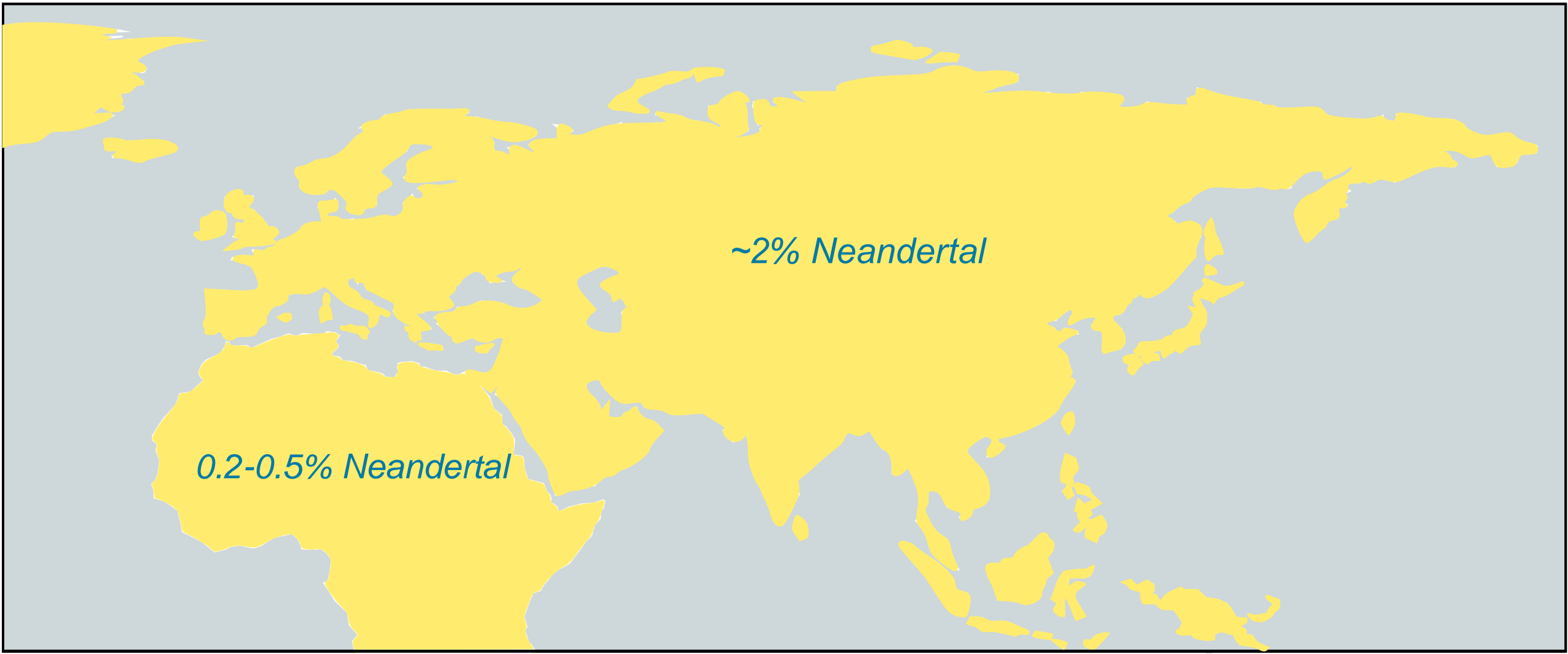
45-100.000 BC



Adapted from Mafessoni 2018, *Nat.Ecol.Evol.*

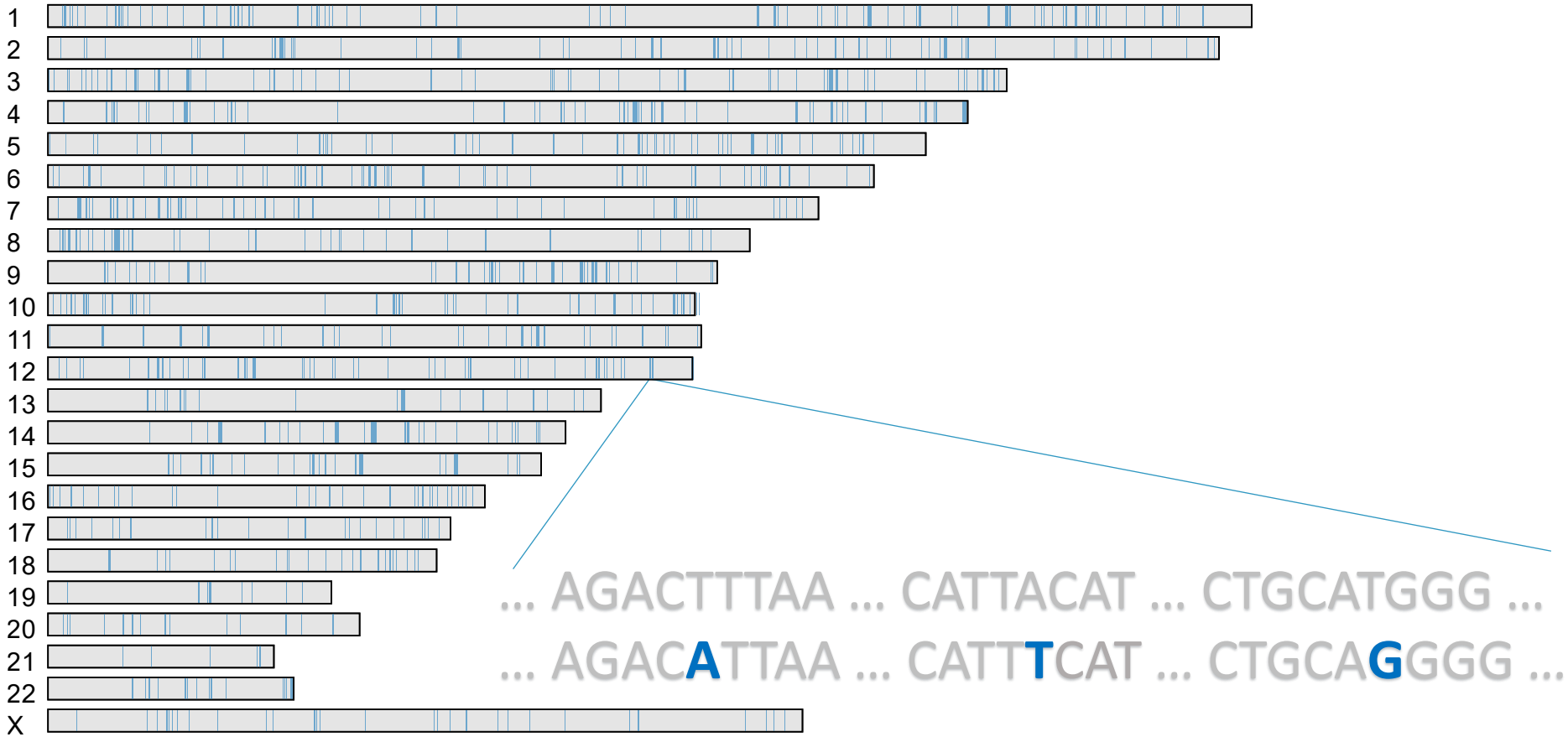
Green et al., 2010, *Science*

today



Adapted from Mafessoni 2018, *Nat.Ecol.Evol.*

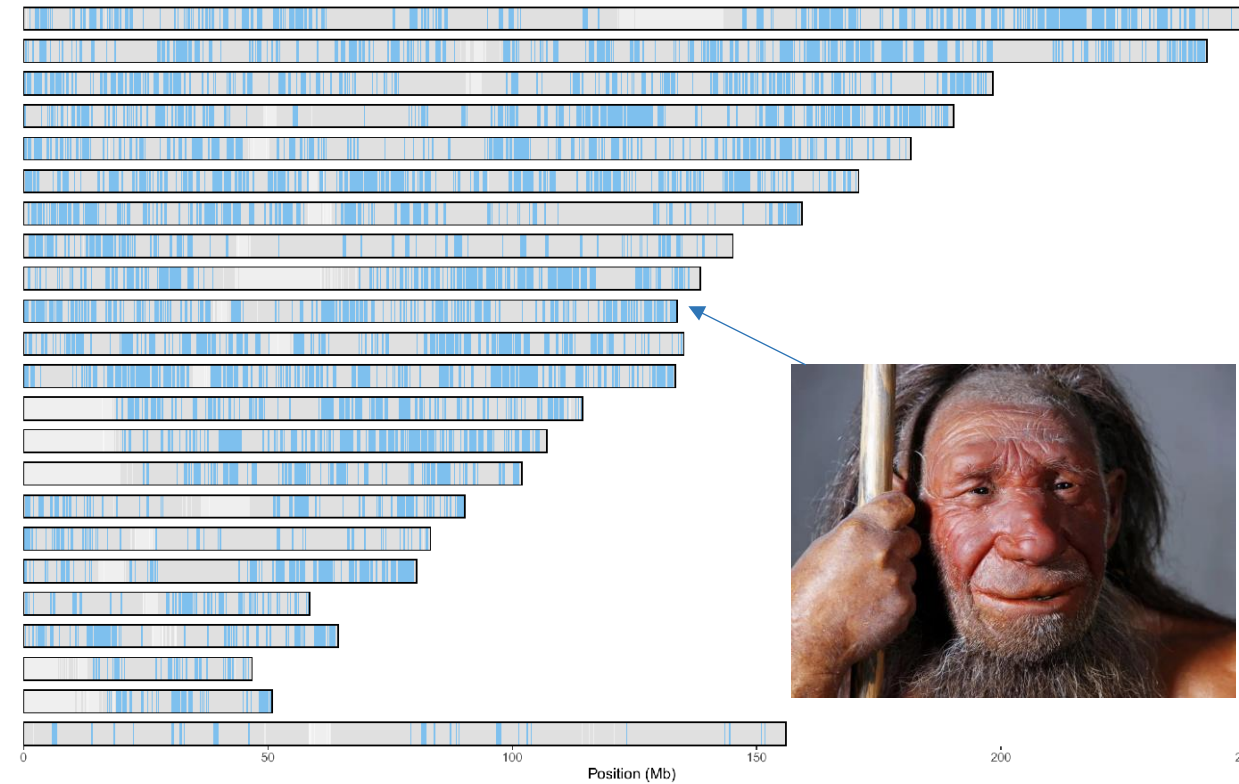
Chen et al., 2020, *Cell*
Schaefer et al., 2021, *Sci.Ad.*



... AGAC**A**TTAA ... CATT**T**CAT ... CTGCAG**G**GGG ...
 ... AGAC**A**TTAA ... CATT**T**CAT ... CTGCAG**G**GGG ...



Introggressione Neanderthal in 27,566 genomi Islandesi



Circa 50% del genoma di Neanderthal puo' essere Ricostruito usando 27,566 genomi islandesi

Neanderthal Introgression at Chromosome 3p21.31 ...
Natural Selection in East Asians
The Strength of Self
Neanderthal Introgression reintroduced functional
ancestral alleles lost in Eurasian populations
nature
ecology & evolution

610 | Nature | Vol 587 | 26 November 2020

Impact and Evolutionary Det
of Neanderthal Introgression on Transcriptional
and Post-Transcriptional Regulation

The major genetic risk factor for severe COVID-19 is inherited from Neanderthals

Genetics

Neanderthal DNA raises risk of severe COVID

Neanderthal-Derived Genetic Variation is Associated with Functional Connectivity in the Brains of Living Humans

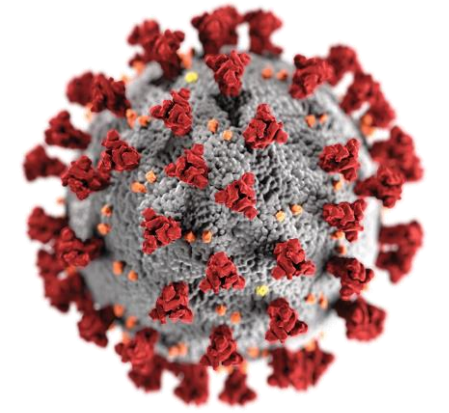
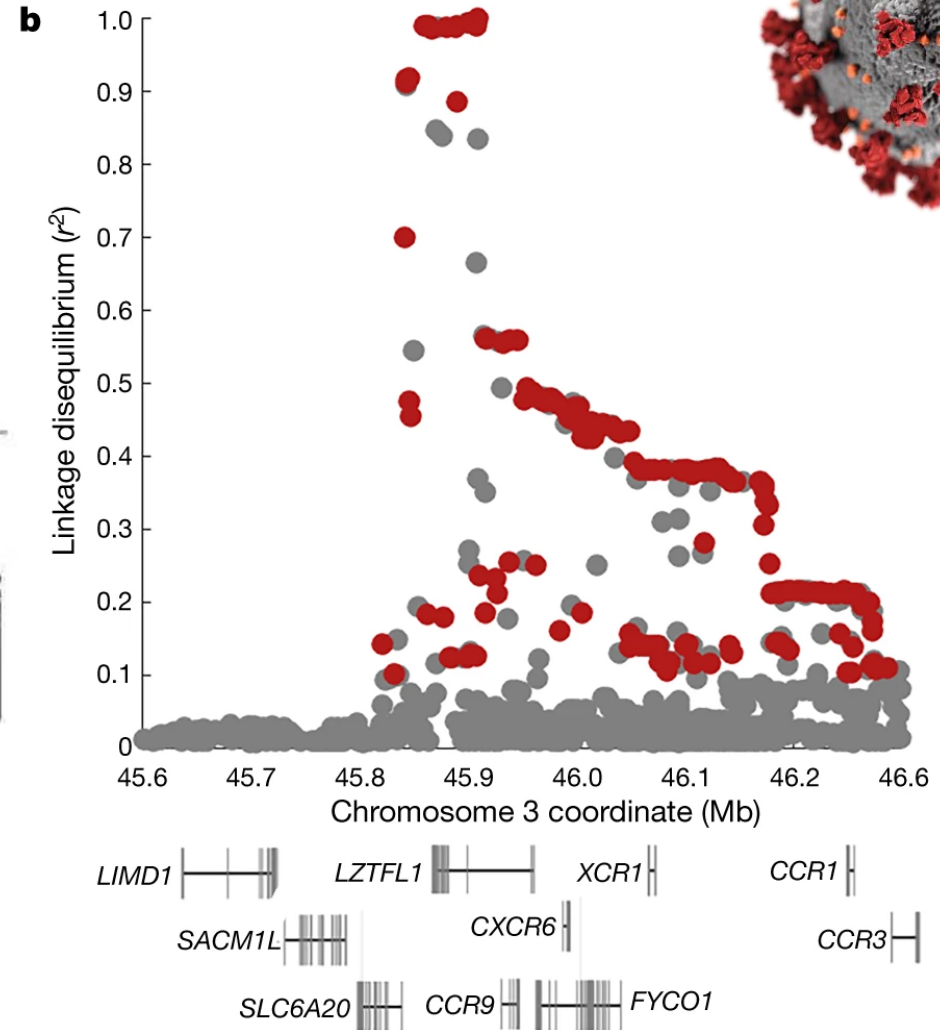
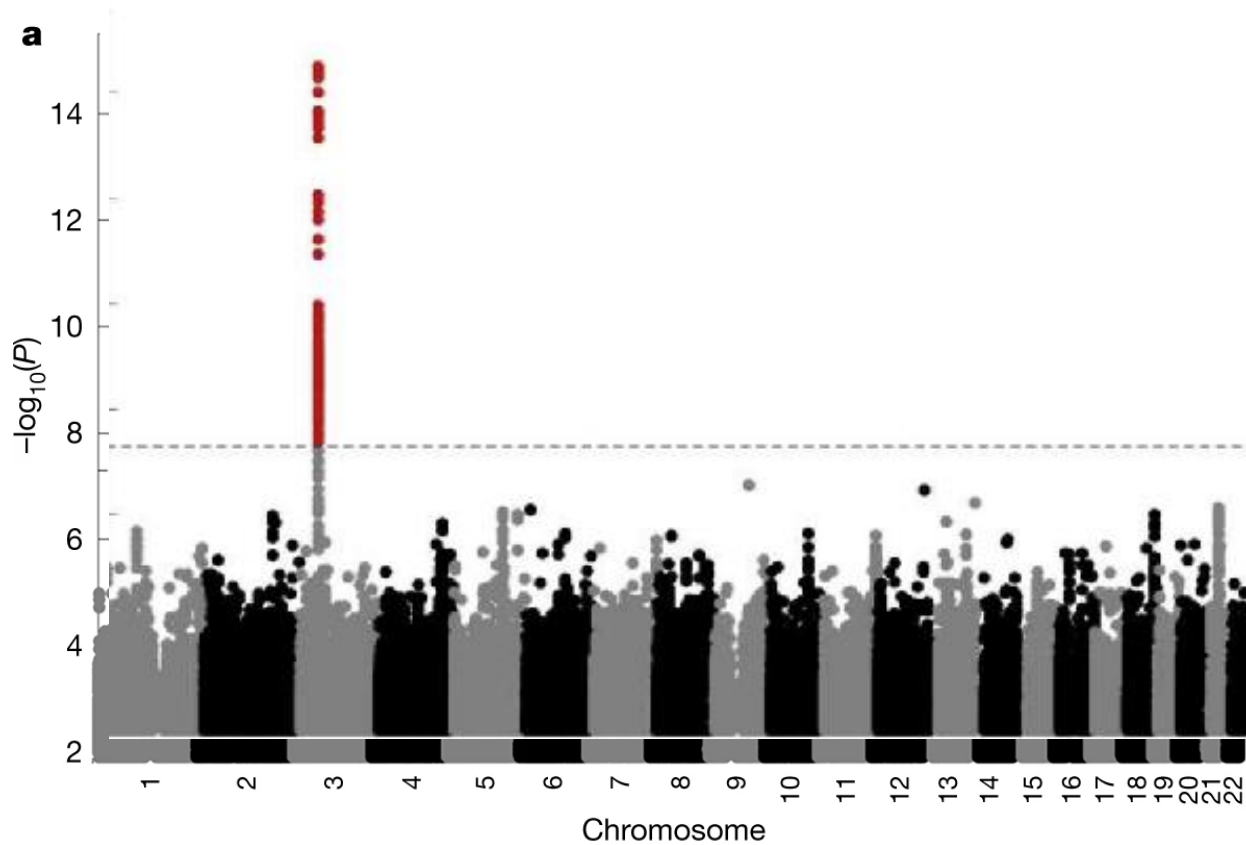
The nature of Neanderthal introgression revealed by 27,566 Icelandic genomes

Nature | Vol 582 | 4 June 2020

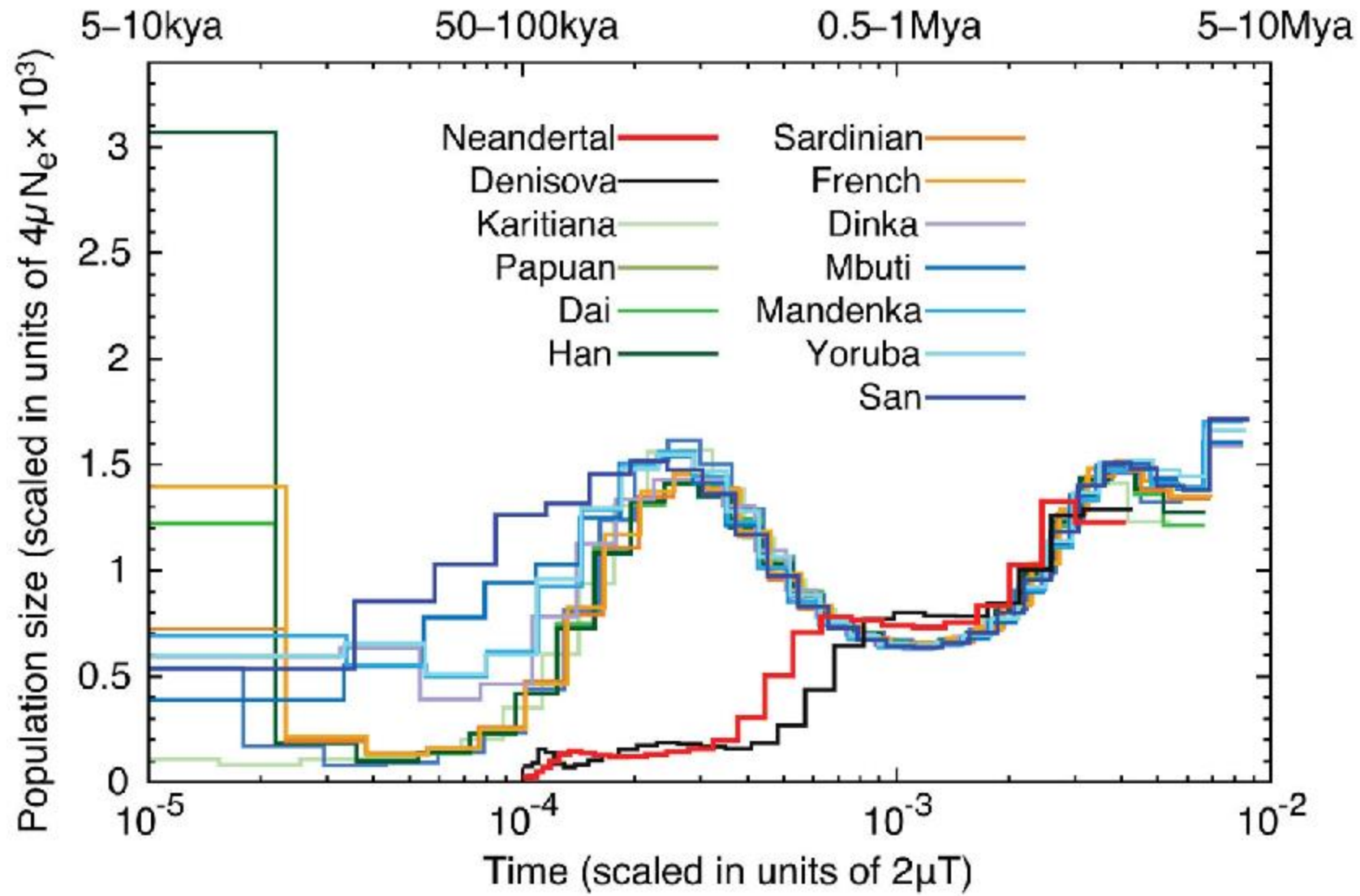
A Neanderthal Sodium Channel Increases Pain Sensitivity in Present-Day Humans

The major genetic risk factor for severe COVID-19 is inherited from Neanderthals

Hugo Zeberg^{1,2} & Svante Pääbo^{1,3}



I genomi degli umani arcaici indicano che vissero per ~200 anni in piccole popolazioni



Introgressione adattativa

Denisova genome



Huerta-Sanchez et al.2014, *Nature*

EPAS1 (HIF-2alpha) gene





EPAS1 also involved in high-altitude adaptation in Himalayan wolf, Tibetan mastiff and Tibetan horses

Liu et al., 2019, *Mol.Biol.Evol.*

Conclusioni

L'equilibrio di Hardy-Weinberg fornisce un punto di partenza per verificare se una popolazione è panmittica

Popolazioni suddivise mostrano una riduzione nel numero degli eterozigoti

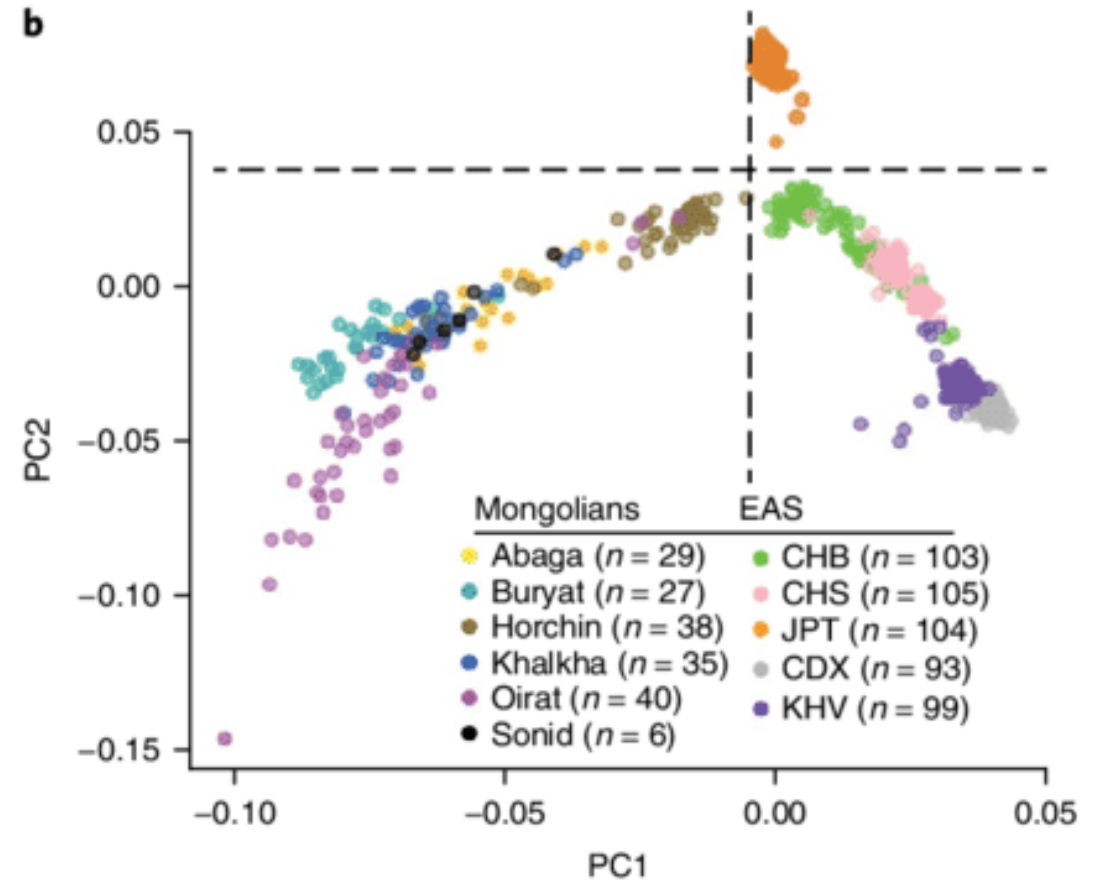
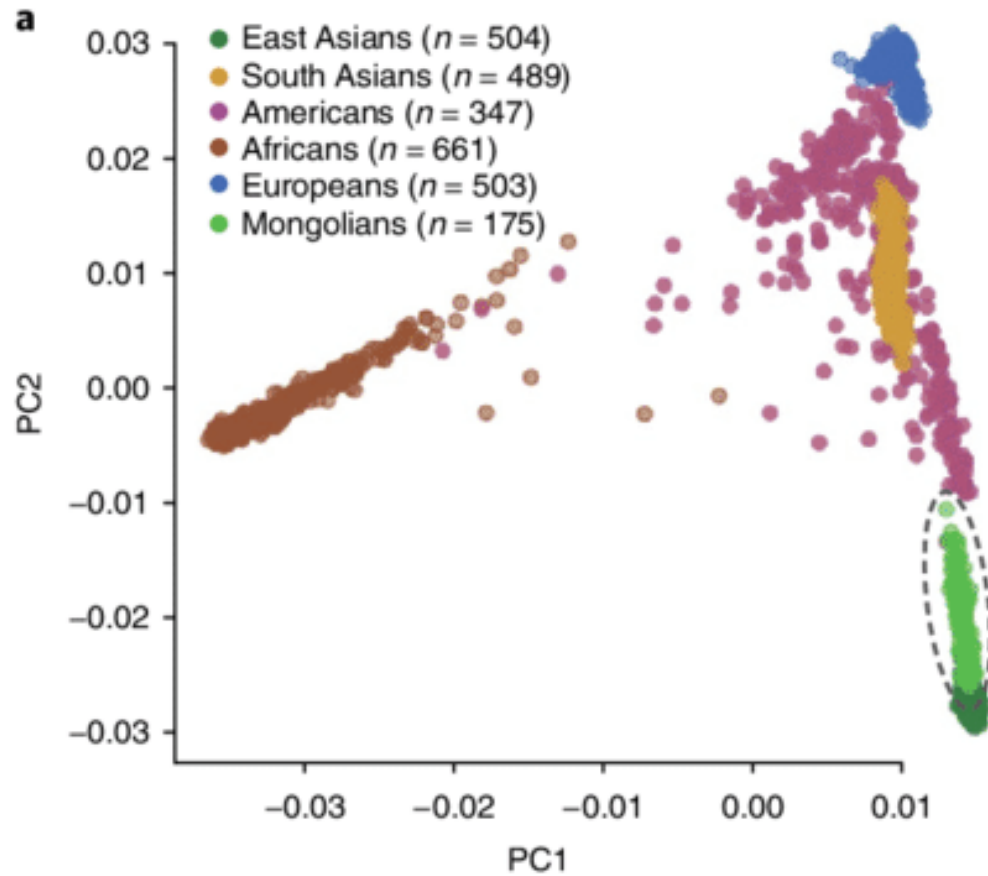
Eterozigosità e diversità nucleotidica sono due dei descrittori principali di una singola popolazione, e sono intimamente connesse ad N_e

N_e può essere influenzata da molti fattori biologici, molti dei quali la riducono in confronto alla popolazione di censo

Un allele può essere neutrale, o sotto tre principali tipi di selezione naturale: positiva, negativa o bilanciata

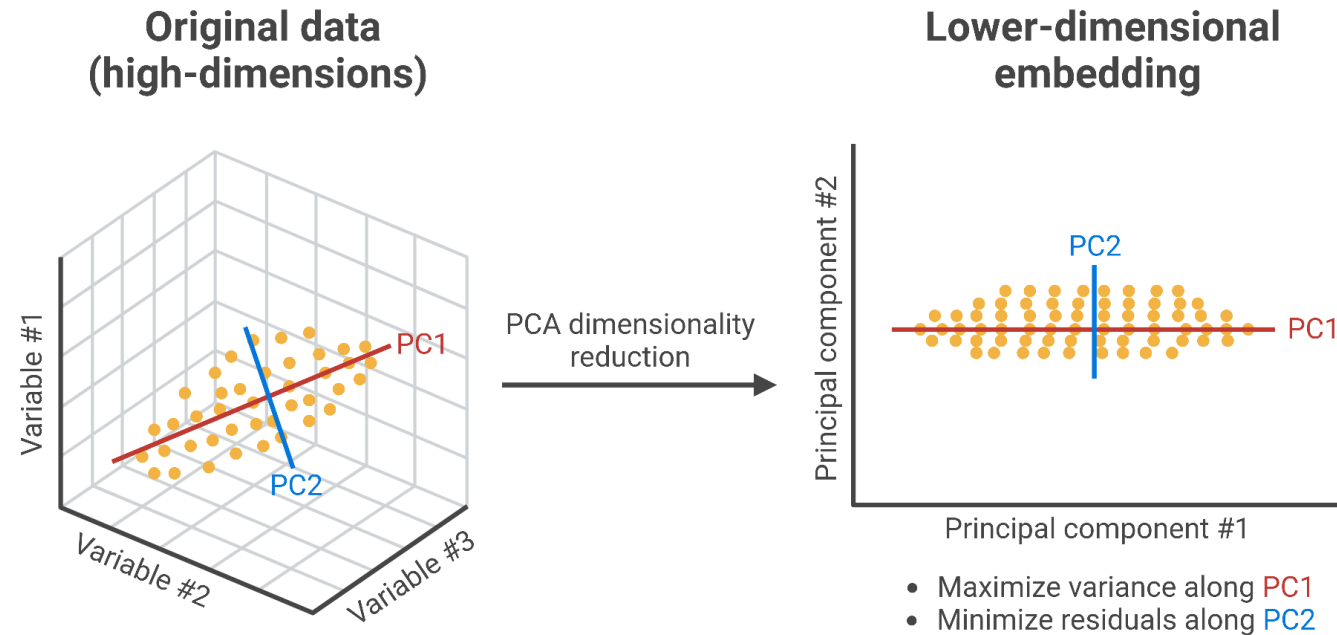
La selezione positiva determina specifici segnali nel genoma (riduzione della diversità, differenziazione tra popolazioni), opposti a quelli della selezione bilanciata.

Analisi delle Componenti Principali (PCA)



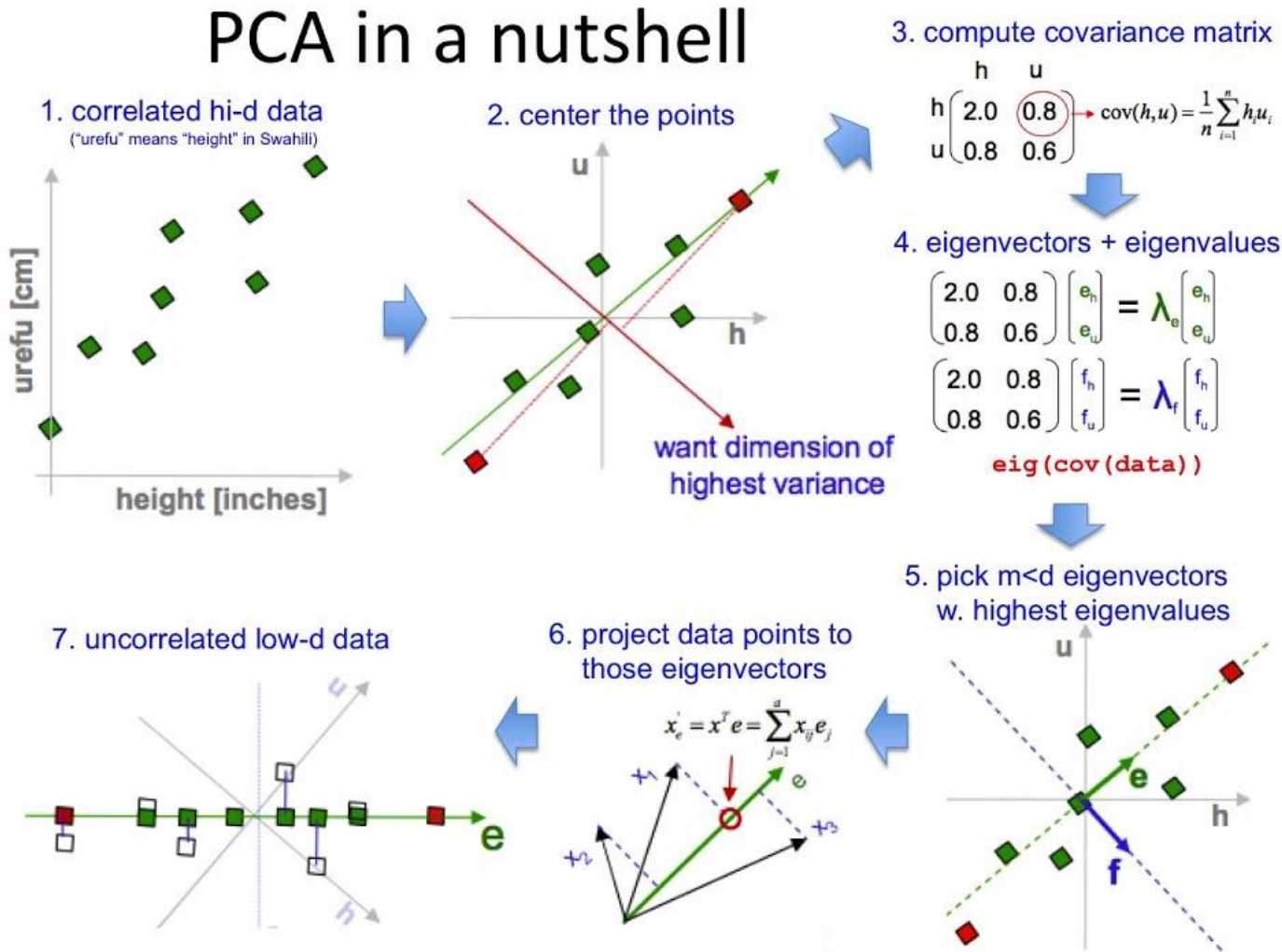
Analisi delle Componenti Principali (PCA)

Principal Component Analysis (PCA) Transformation



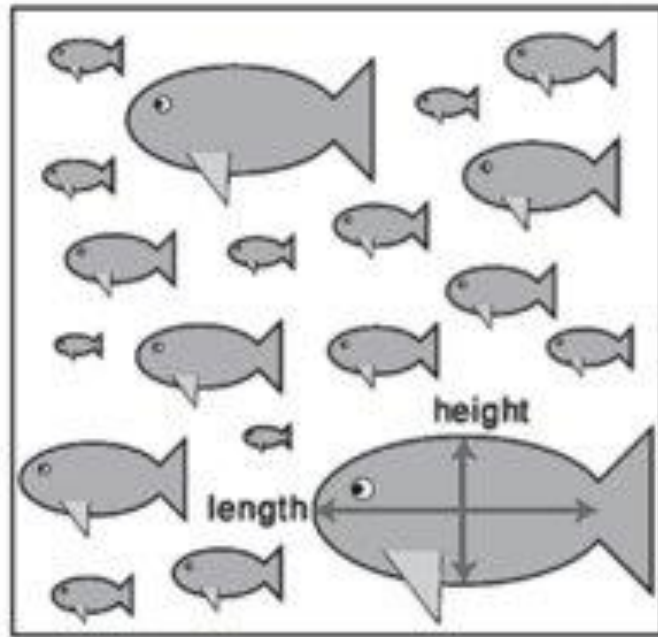
Analisi delle Componenti Principali (PCA)

PCA in a nutshell

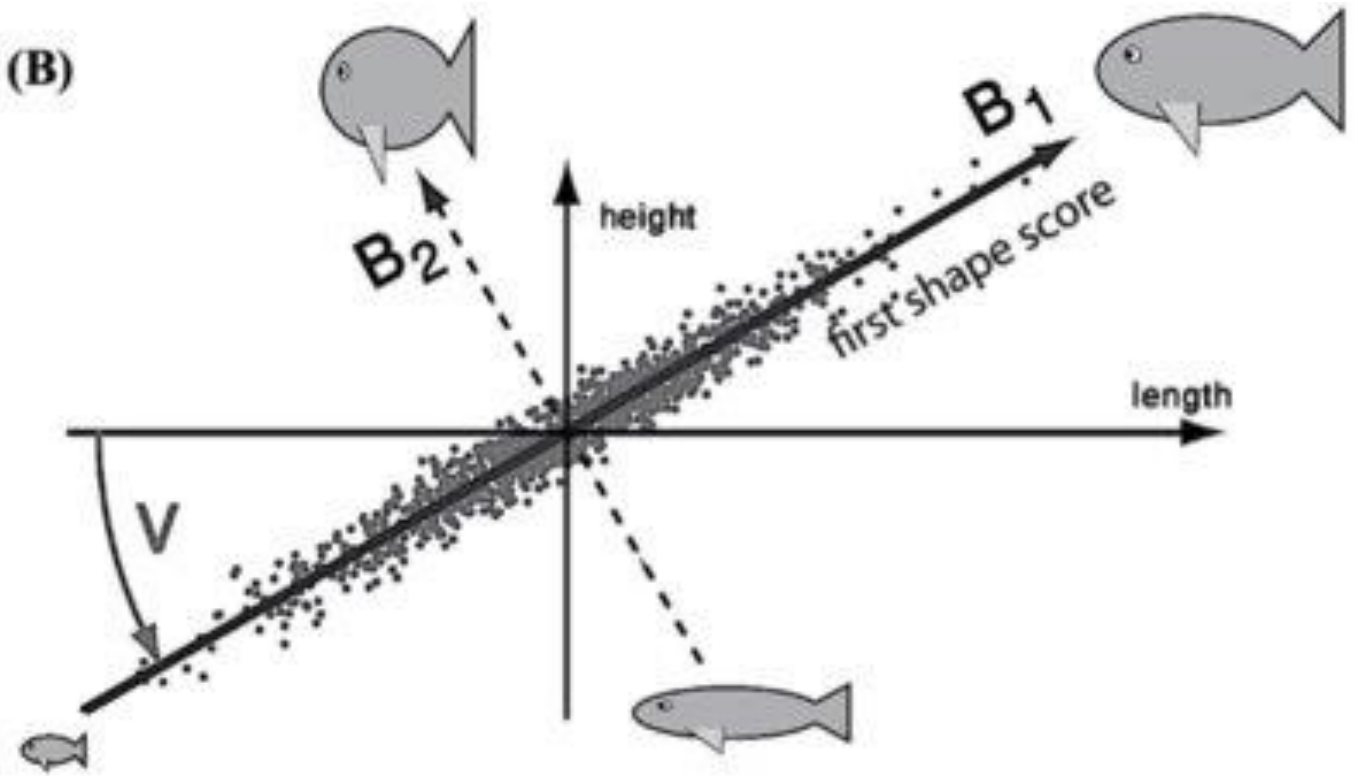


Analisi delle Componenti Principali (PCA)

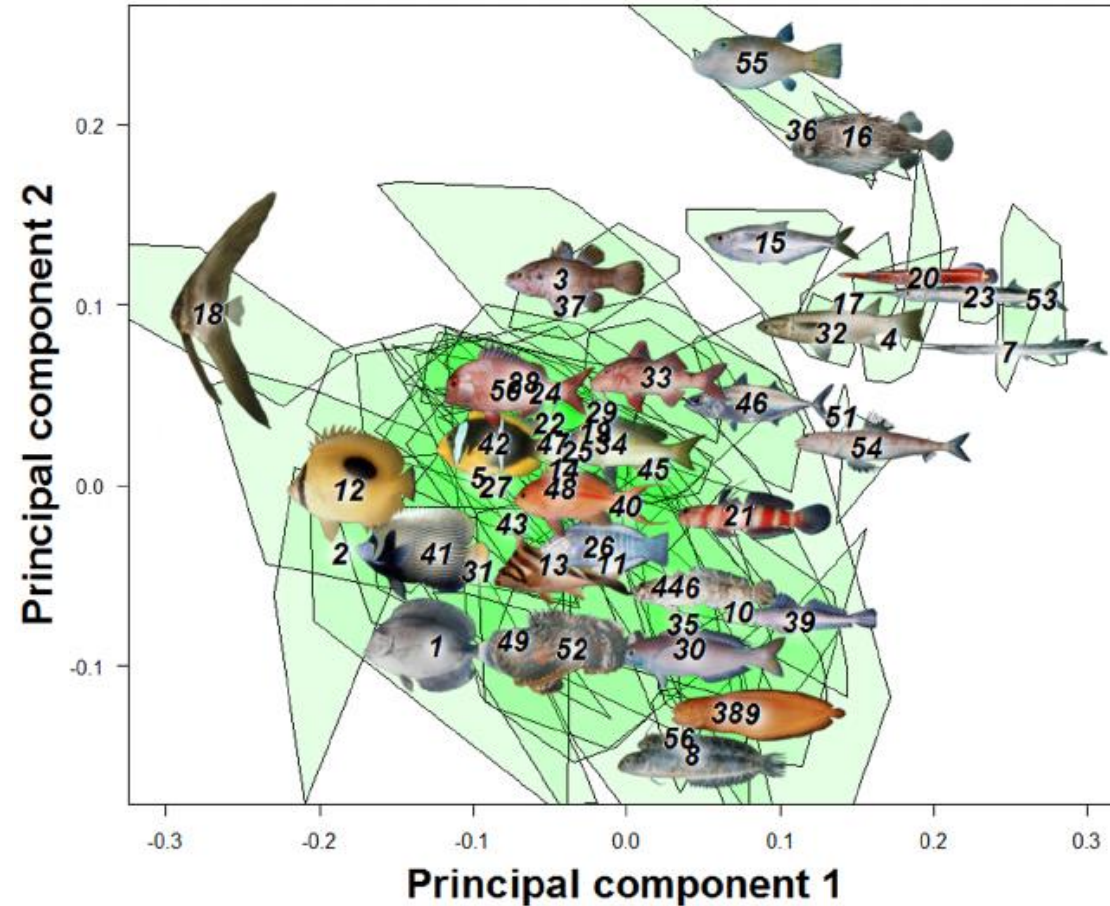
(A)



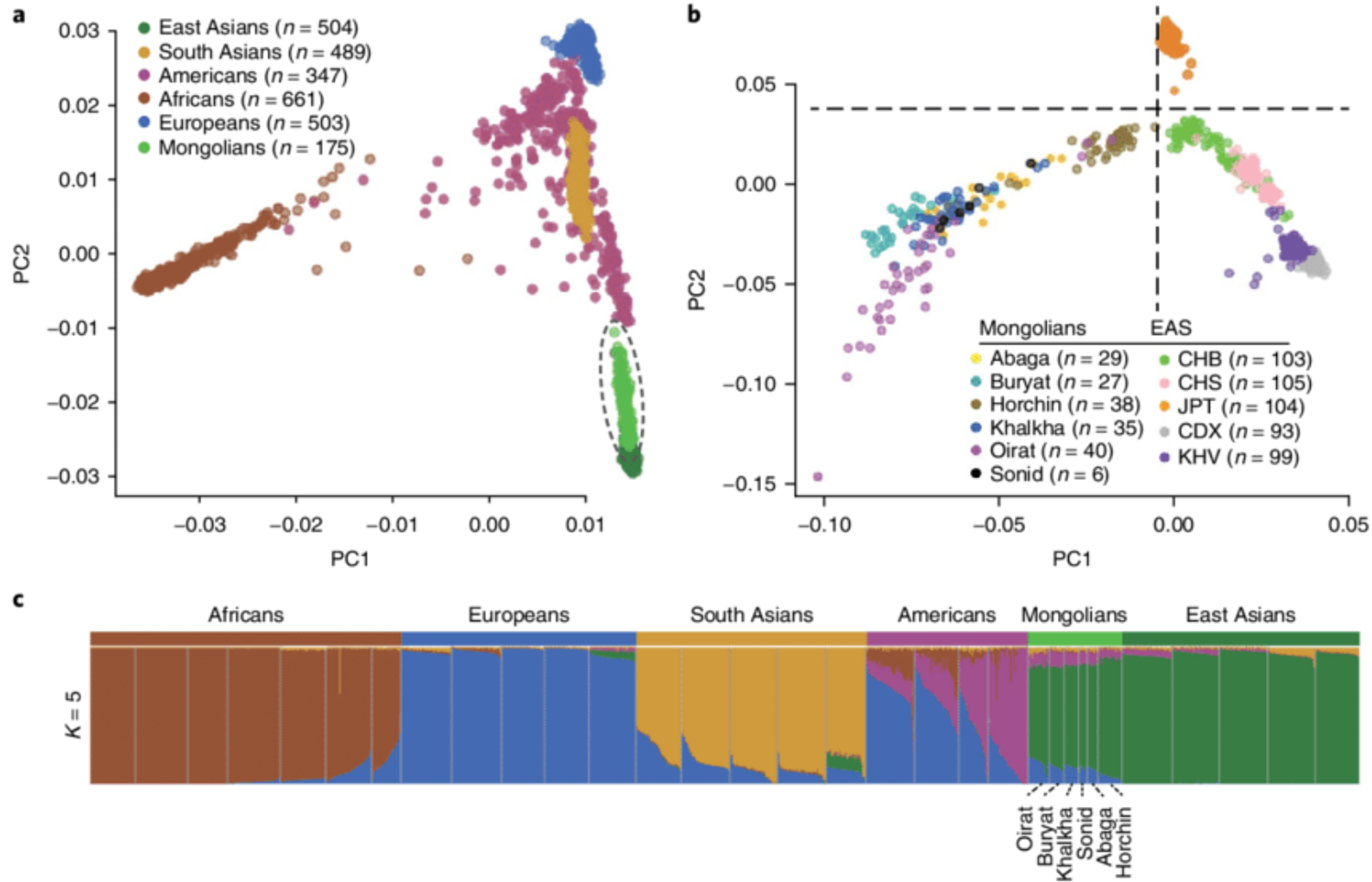
(B)



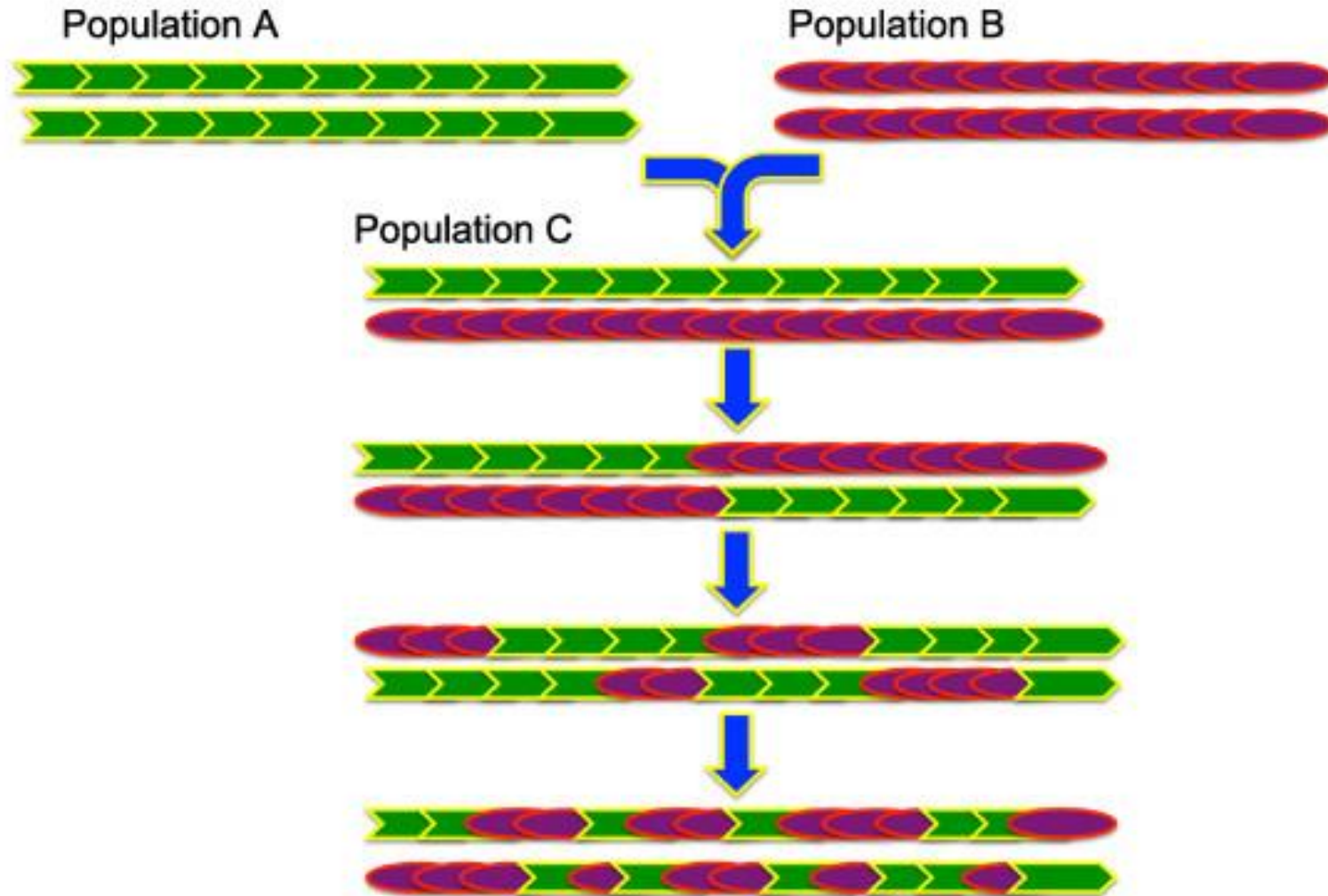
Analisi delle Componenti Principali (PCA)

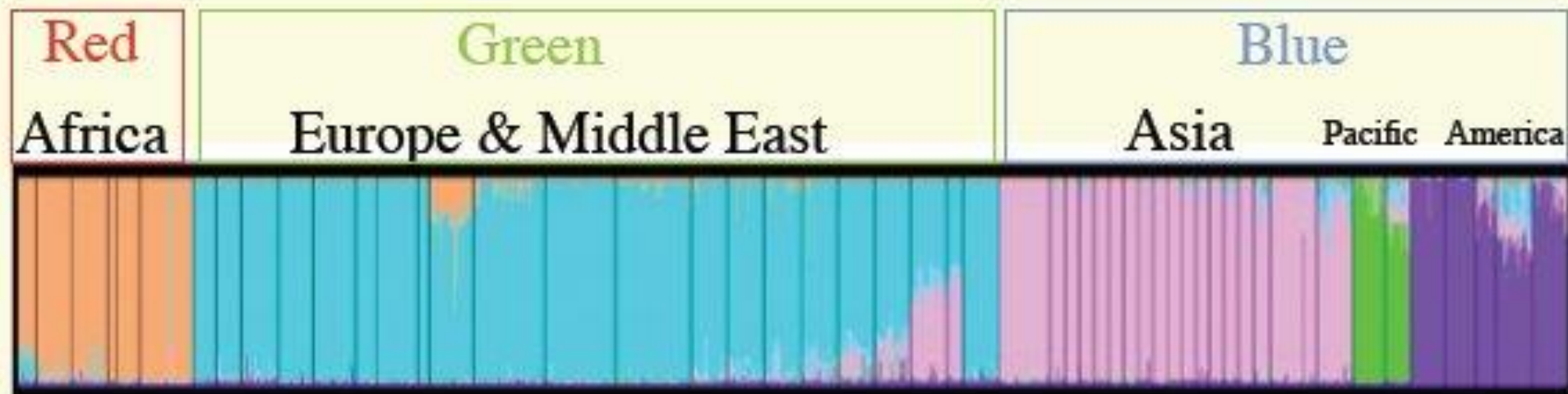
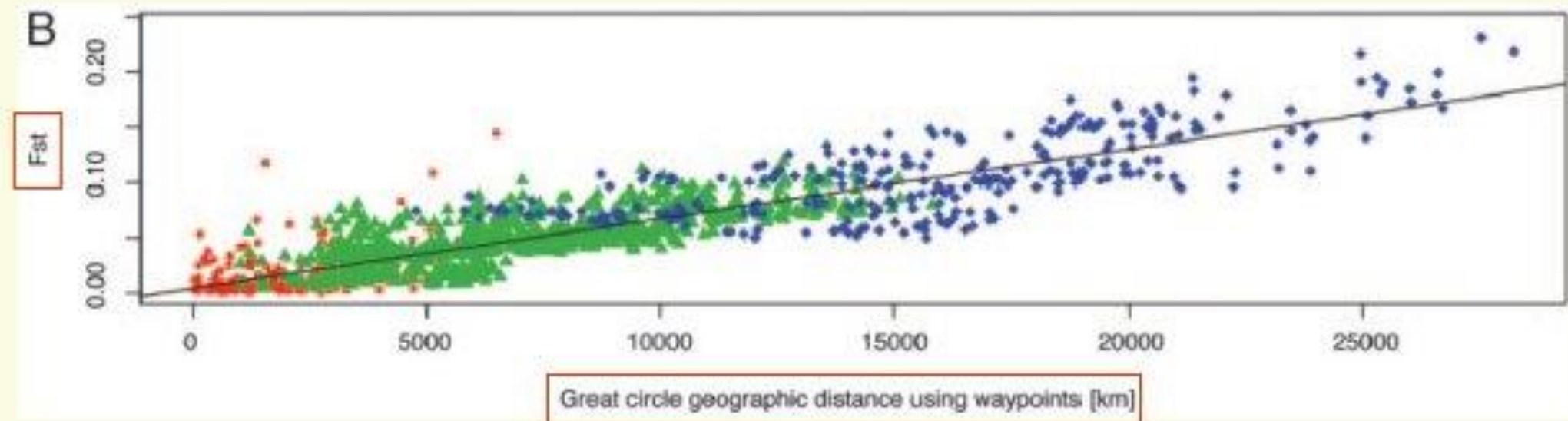


Analisi delle Componenti Principali (PCA)



STRUCTURE and ADMIXTURE





Rosenberg
et al.
Science
298:2381-
2385, 2002

Un'applicazione pratica di STRUCTURE

I salmoni rinvenuti negli stomaci di merluzzi atlantici provengono da acquacultura e non da popolazioni locali (e sono malati di Reovirus)

