

Lezione 3

I geni nelle popolazioni

Lezione modificata da Rosenblum <http://people.ibest.uidaho.edu/~bree/>

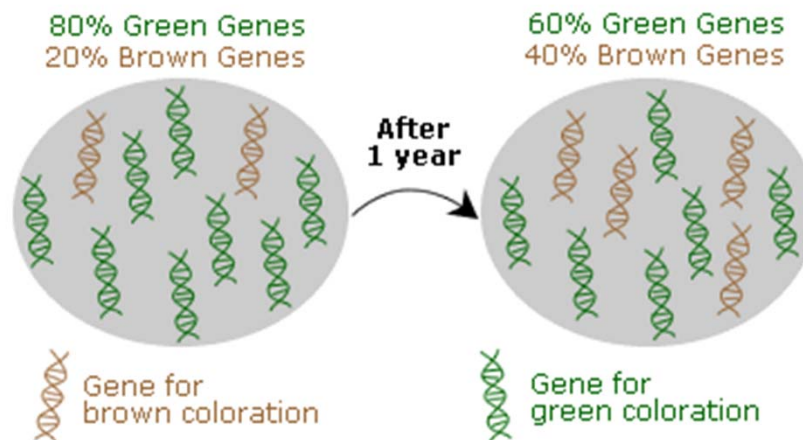
Graur and Li: Capitolo 2

Graur lectures 8-9-10-11-12

L'**evoluzione** è Il processo che risulta dal **cambiamento delle caratteristiche genetiche delle popolazioni**

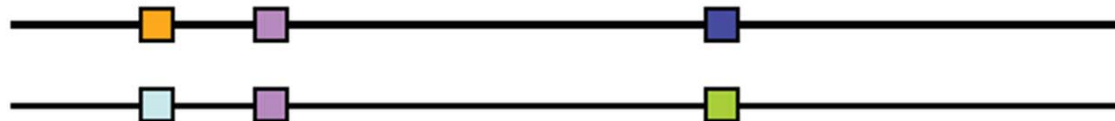
La componente di base del processo evolutivo è il cambiamento nelle **frequenze alleliche nel tempo**

La **genetica di popolazioni** si occupa del cambiamento delle frequenze alleliche che avviene nelle popolazioni



I geni negli individui

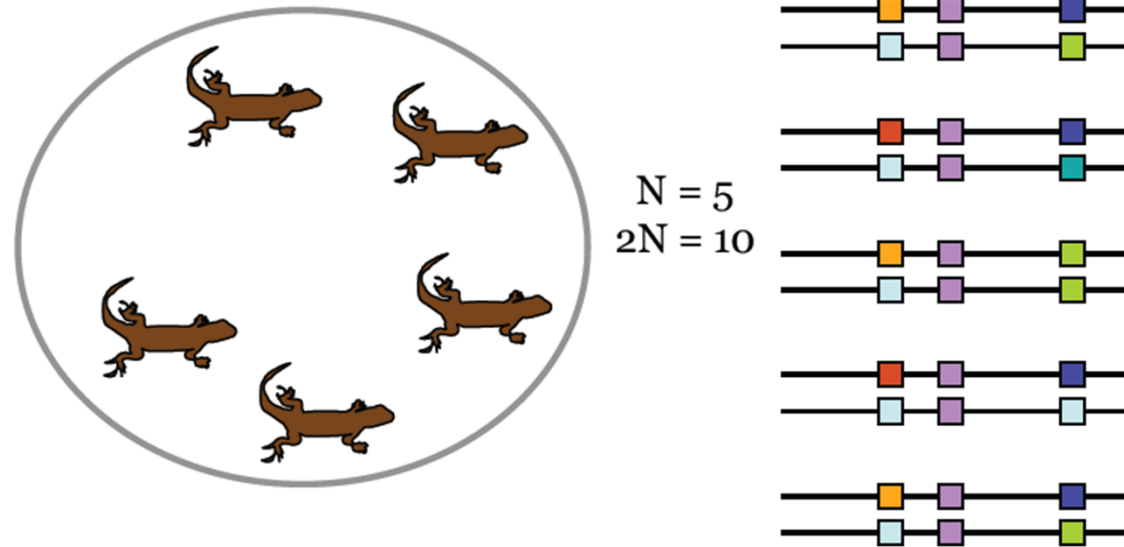
- **Omozigote**: lo stesso allele ad un locus in organismi diploidi (o poliplodi)



- **Eterozigote**: Alleli diversi ad un locus in organismi diploidi (o poliplodi)

I geni nelle popolazioni

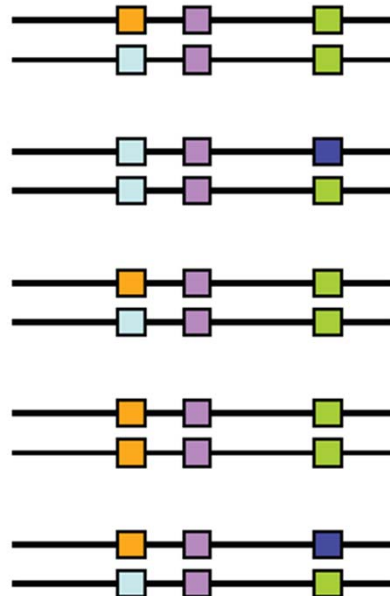
- **Pool genico**: il set di alleli unici a tutti i loci in una popolazione



- **Dimensione della popolazione**: in genere indicata con N per la componente aploide, $2N$ per la diploide

I geni nelle popolazioni

- **Polimorfico**: presenza di >1 tipo genetico distinto nella popolazione (variante, forma alternativa)



- **Monomorfico**: nella popolazione esiste un solo tipo genetico

I geni nelle popolazioni

- **Polimorfico**: presenza di >1 tipo genetico distinto nella popolazione (variante, forma alternativa)
- **Monomorfo**: nella popolazione esiste un solo tipo genetico

Il **polimorfismo** può essere inteso su scala genetica (esempio: single nucleotide polymorphism, SNP)

C	G	A	T	T	C	C	G	A	T	T	C
C	G	A	A	T	C	C	G	A	T	G	C
C	G	A	T	T	C	G	G	A	T	G	C
C	G	A	C	T	C	C	G	A	T	T	C

O su scala popolazionistica (colori, tipi di giaguari in Sud America)



I geni nelle popolazioni

- Frequenza allelica (allele frequency, gene frequency):
proporzione relativa di un allele in una popolazione

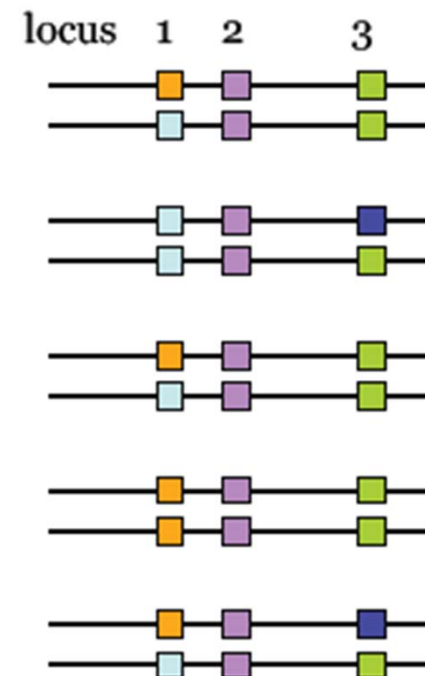


Per un locus con 2 alleli:

p = frequenza dell'allele A

q = frequenza dell'allele a (1-p)

$$p+q = 1$$



I geni nelle popolazioni

Le frequenze alleliche rimangono costanti nel tempo in una popolazione ideale

Questo si chiama
Equilibrio di Hardy Weinberg (HWE)

Che cos'è una popolazione ideale?

1. Generazioni discrete, non sovrapposte
2. Stesse frequenze alleliche in maschi e femmine
3. Panmissia: gli individui si accoppiano in maniera casuale (unione casuale dei gameti)
4. La dimensione della popolazione è molto grande (tendente all'infinito)
5. Non c'è migrazione (la popolazione è chiusa)
6. La mutazione può essere ignorata
7. La selezione non influenza le frequenze alleliche (alleli neutrali)

Popolazione ideale: realistica?

- In genere no, ma HWE fornisce un'importante
IPOTESI NULLA
- Deviazioni dall'HWE ci suggeriscono che
qualche cosa di importante sta avvenendo
nella popolazione (evoluzione!)

Implicazioni dell'HWE

1. Le frequenze alleliche **sono stabili nel tempo**, e non dipendono dal modello di dominanza tra gli alleli
2. Poichè le **frequenze genotipiche** possono essere prdette da quelle alleliche, anche queste **rimangono stabili nel tempo**
3. Date le leggi di Mendel **HWE viene raggiunto dopo una singola generazione di incroci casuali.**
4. Poichè l'evoluzione si definisce come il cambiamento delle frequenze alleliche e genotipiche nel tempo, **una popolazione ideale non evolve**
5. **L'evoluzione avviene perchè vengono violate alcune delle assunzioni di una popolazione ideale**
 - a) Incroci non casuali
 - b) Dimensioni finite (deriva genetica, caso)
 - c) Migrazione
 - d) Mutazioni
 - e) Selezione



Forze evolutive

Se una popolazione è in HWE, quali saranno le sue frequenze genotipiche?

Per un locus con 2 alleli:

p = frequenza dell'allele A

q = frequenza dell'allele a ($1-p$)

$$p+q = 1$$



Così sappiamo quali sono le frequenze alleliche, ma non quelle genotipiche

Con 2 alleli ci sono 3 possibili genotipi:



Se una popolazione è in HWE, quali saranno le frequenze genotipiche?

Genotipi	AA	Aa	aa
Frequenze	$p^2 + 2pq + q^2 = 1$		

Da dove si ottengono queste frequenze genotipiche?

Quadrato di Punnett per popolazione

		Femmine	
		A (p)	a (q)
Maschi	A (p)	AA (p ²)	Aa (pq)
	a (q)	Aa (pq)	aa (q ²)

HW può essere usato per

- Calcolare le frequenze genotipiche per una popolazione in HWE
- Testare per deviazioni statisticamente significative dall'HWE

Esempio: Malattie

1/1700 neonati ha la fibrosi cistica

Qual è la proporzione di portatori asintomatici nella popolazione?

HW can be used to:

- Calculate genotype frequencies for a population in HWE
- Test for statistical deviations from HWE

Example calculation of HWE freq: Disease

1 in 1700 newborns have cystic fibrosis

What proportion of the population
are asymptomatic carriers?

$$q^2 = 1/1700 = 0.00059 = 0.059\%$$

$$q = 0.024$$

$$p = 0.976 \quad p + q = 1$$

$$2pq = 0.047 \quad p^2 + 2pq + q^2 = 1$$

4.7% portatori

Tutti i cambiamenti evolutivi iniziano con cambiamenti intrapopolazione

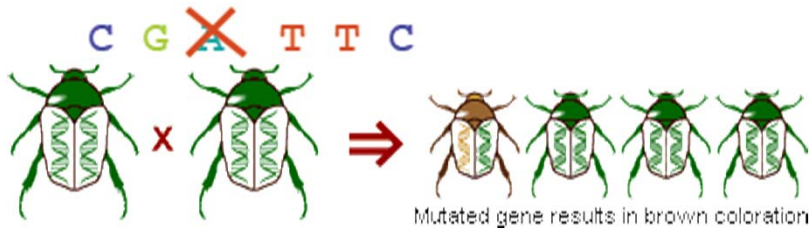
- Domande chiave:
 - Come cambiano le frequenze di un allele mutato nel tempo sotto l'influenza delle diverse forze evolutive?
 - Come viene mantenuta la variabilità nelle popolazioni naturali?
 - Qual è la probabilità che un nuovo mutante rimpiazzhi completamente le varianti esistenti e con che velocità questo accade?

Chance!

N= dimensione della popolazione



Mutazione → nuova variante allelica con frequenza 1/N



La frequenza di *a* cresce nel tempo fino ad arrivare a sostituire *A* quindi a *FISSARSI*



EVOLUZIONE

Deriva genetica



Casuale

Selezione



non casuale

Migrazione

Ricombinazione

Mutazione

Mutazione = variabilità genetica (nessuna direzionalità!):
le frequenze alleliche cambiano grazie alla mutazione

In generale è una forza evolutiva “debole”

Mutazione

Generating variation..

Tasso di mutazione: frequenza di mutazione per generazione (o per anno) e per *entità biologica* (nucleotide, gamete, individuo)

Alcuni concetti importanti:

Il tasso di mutazione non è casuale:
varia a seconda delle **specie**

<u>Taxon</u>	<u>mut/nucleotide site/year</u>
Mammal	10^{-9}
Flu virus	10^{-2} (more than a million times faster)

Se siamo nel nucleo o negli **organelli**

<u>Taxon</u>	<u>mut/nucleotide site/year</u>
Nuclear DNA	10^{-9}
Mitochondrial DNA	10^{-8}

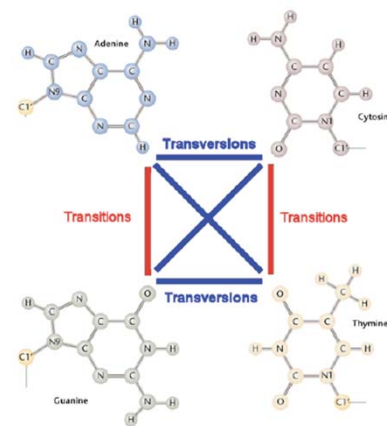
Mutazione

Il tasso di mutazione non è casuale:
varia a seconda delle **regioni genomiche**

Esistono **Hot spot mutazionali** (regione prone alla mutazione)

- sequenze ripetitive
- le transizioni sono più comuni delle trasversioni
- alcuni nucleotidi sono più mutabili (es. G e C nel genoma dei mammiferi)

<u>Mammalian DNA type</u>	<u>mut/nucleotide site/year</u>
Background nuclear DNA	10^{-9}
Microsatellites	10^{-3}



High rate of mutation of
methylated C $\rightarrow$ T

Mutazione

La mutazione è casuale rispetto alle conseguenze (fitness):

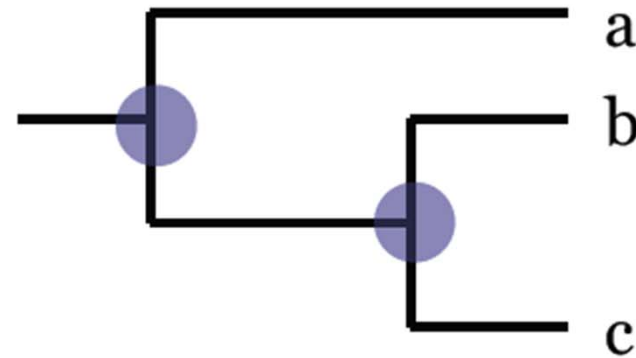
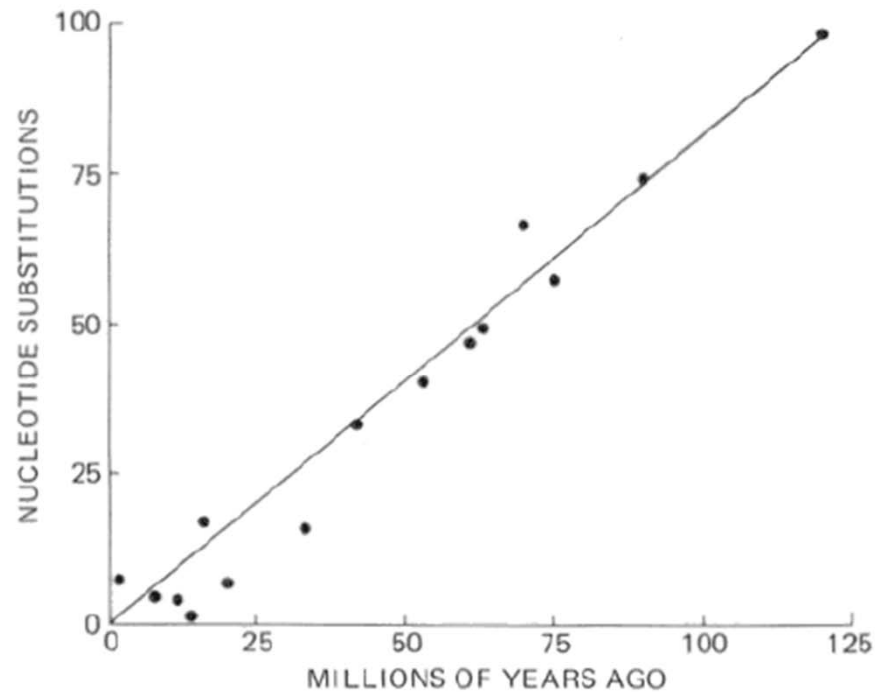
Non è **diretta** a fornire la variazione di cui l'organismo necessita



L'esposizione agli antibiotici seleziona i batteri resistenti ma **non è la causa** della resistenza

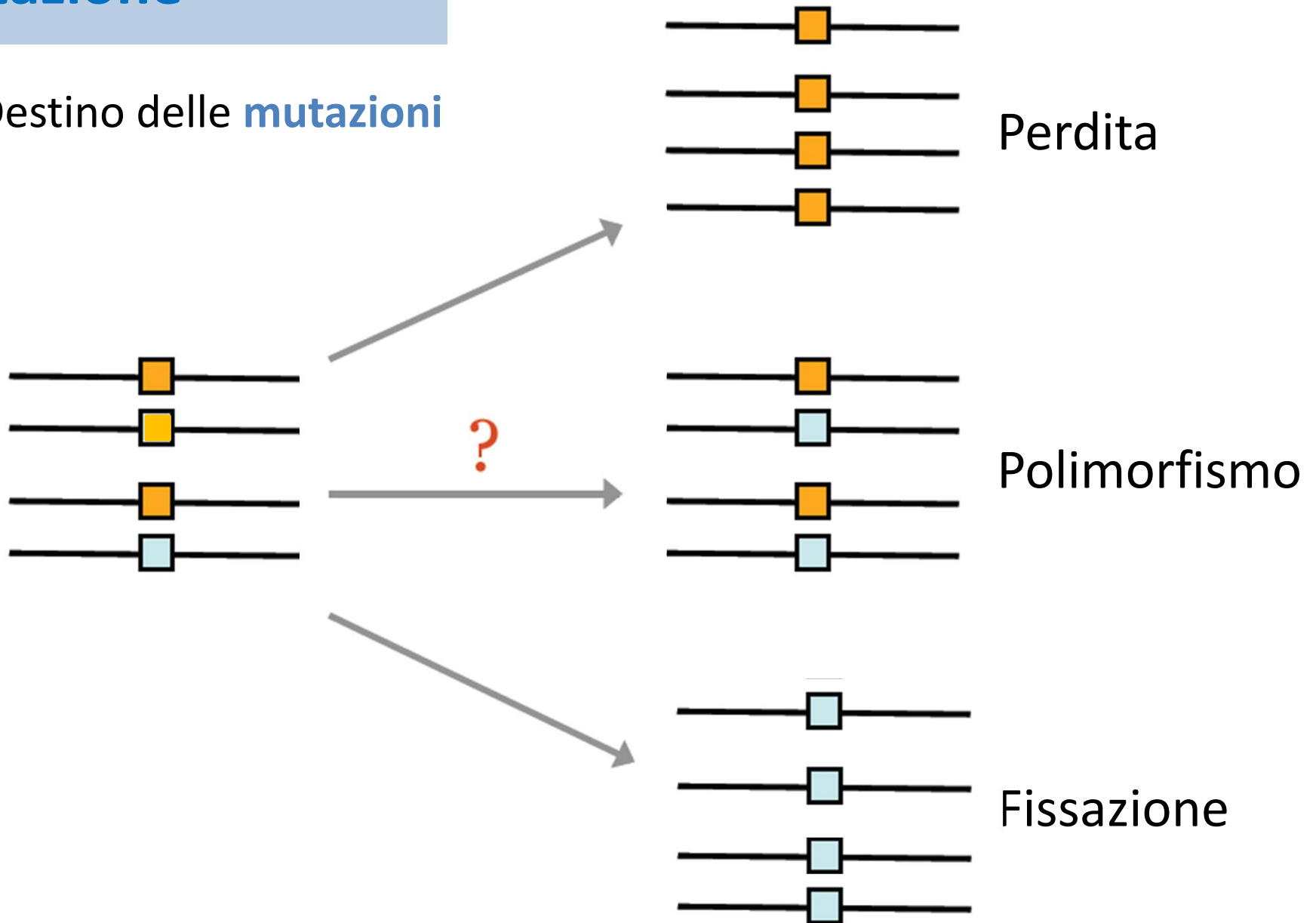
Mutazione

Anche se le mutazioni non sono distribuite equamente nel genoma, approssimando è possibile utilizzare un **orologio molecolare** (con tutti i limiti di questo approccio) specialmente se calibrato con dati fossili



Mutazione

Destino delle **mutazioni**



Approcci per studiare i cambiamenti genetici nelle popolazioni

- Modelli deterministici
 - Più semplici
 - I cambiamenti di generazione in generazione avvengono in un unico modo
 - Questo può essere predetto conoscendo le condizioni iniziali
- Modelli stocastici
 - I cambiamenti avvengono in modo probabilistico
 - Conoscendo lo stato 1 non possiamo predire in modo certo le f alleliche dello stato 2
 - Possiamo solo calcolare delle probabilità

Funzionano solo se

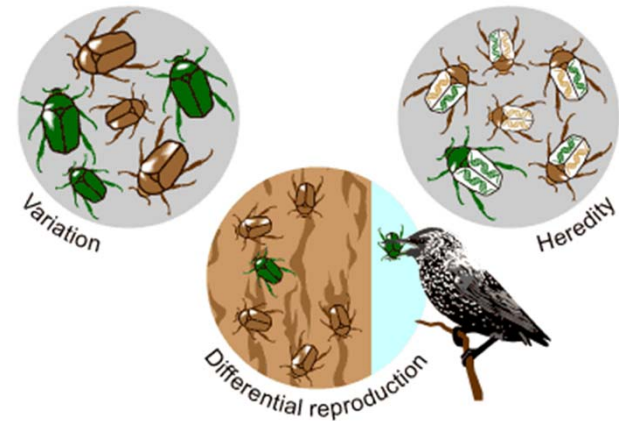
1. N è infinita
2. L'ambiente è costante

Preferibili perché più realistici

Non sempre trattabili matematicamente

Selezione naturale

Un approccio deterministico



<http://evolution.berkeley.edu>

Successo riproduttivo differenziale di alcuni genotipi rispetto ad altri

Dipende da

Mortalità

Fertilità

Fecondità

Successo nell'accoppiamento

Vitalità della prole

$w = \text{fitness}$

misura delle potenzialità di un individuo di sopravvivere e riprodursi

La fitness di un certo genotipo varia nel tempo a seconda dell'ambiente:
è un concetto relativo, ma fissiamola per trattarla matematicamente

Selezione naturale

La fitness (w) di un genotipo è una misura della capacità di un individuo di **sopravvivere e riprodursi**.

Le dimensioni di una popolazione sono determinate dalla capacità portante **dell'ambiente**.

Quindi il **successo evolutivo di un individuo** è determinato non dalla sua fitness assoluta, ma da quella del suo genotipo rispetto agli altri genotipi della popolazione (**fitness relativa**).

Selezione naturale

- La selezione naturale **cambia le frequenze alleliche relative nella popolazione**
- La selezione naturale **è una forza evolutiva di grande efficacia**, il suo effetto è quello di migliorare l'**adattamento** degli individui al loro ambiente
- La selezione non ha uno scopo finale, **non ci sono varianti genetiche buone**

→ Una variante è benefica per il suo portatore in un determinato ambiente e dannosa in un altro



Selezione naturale

Un locus e due alleli

Random mating:

Genotipi	A_1A_1	A_1A_2	A_2A_2
Fitness	w_{11}	w_{12}	w_{22}
Frequenze	p^2	$2pq$	q^2

*Calcolo della frequenza allelica
dalla frequenza genotipica*

Frequenza di $A_1 = p$

$$p = p^2 + \frac{2pq}{2} = p^2 + pq$$

Frequenza di $A_2 = q$ o $1-p$

$$q = q^2 + \frac{2pq}{2} = q^2 + pq$$

Selezione naturale

Un locus e due alleli

Genotipi	A_1A_1	A_1A_2	A_2A_2
Fitness	w_{11}	w_{12}	w_{22}
Frequenze	p^2	$2pq$	q^2
Dopo una generazione e l'azione della selezione			
Frequenze	p^2w_{11}	$2pqw_{12}$	q^2w_{22}

Selezione naturale

Un locus e due alleli

Genotipi	A_1A_1	A_1A_2	A_2A_2
Frequenze	p^2w_{11}	$2pqw_{12}$	q^2w_{22}

Frequenza di A_2 dopo una generazione

$$q' = \frac{pqw_{12} + q^2w_{22}}{p^2w_{11} + 2pqw_{12} + q^2w_{22}}$$

di quanto è cambiata la frequenza del nuovo allele A_2 dopo una generazione?

$$\Delta q = \frac{pq[p(w_{12} - w_{11}) + q(w_{22} - w_{12})]}{p^2w_{11} + 2pqw_{12} + q^2w_{22}}$$

Selezione naturale

Un locus e due alleli: A_2 dominante

Genotipi	A_1A_1	A_1A_2	A_2A_2
Fitness (w)	1	$1+s$	$1+s$

s può essere negativo

Un locus e due alleli: A_2 recessivo

Genotipi	A_1A_1	A_1A_2	A_2A_2
Fitness (w)	1	1	$1+s$

s può essere negativo

Selezione naturale

Un locus e due alleli: codominanza

Genotipi	A_1A_1	A_1A_2	A_2A_2
Fitness (w)	1	$1+s$	$1+2s$
s può essere negativo			

$$\Delta q = \frac{spq}{1 + 2spq + 2sq^2}$$

Se s è piccolo (di solito lo è) il denominatore è circa 1

$$\Delta q = spq$$

Si può approssimare con $\frac{dq}{dt} = spq = sq(1 - q)$

Selezione naturale

Un locus e due alleli: codominanza

$$\Delta q = \frac{spq}{1 + 2spq + 2sq^2}$$

Se s è piccolo (di solito lo è) il denominatore è circa 1

$$\Delta q = spq$$

Si può approssimare con

La soluzione è

$$\frac{dq}{dt} = spq = sq(1 - q) \quad \longrightarrow \quad q_t = \frac{1}{1 + \left(\frac{1 - q_0}{q_0}\right)e^{-st}}$$

Selezione naturale

Un locus e due alleli: codominanza

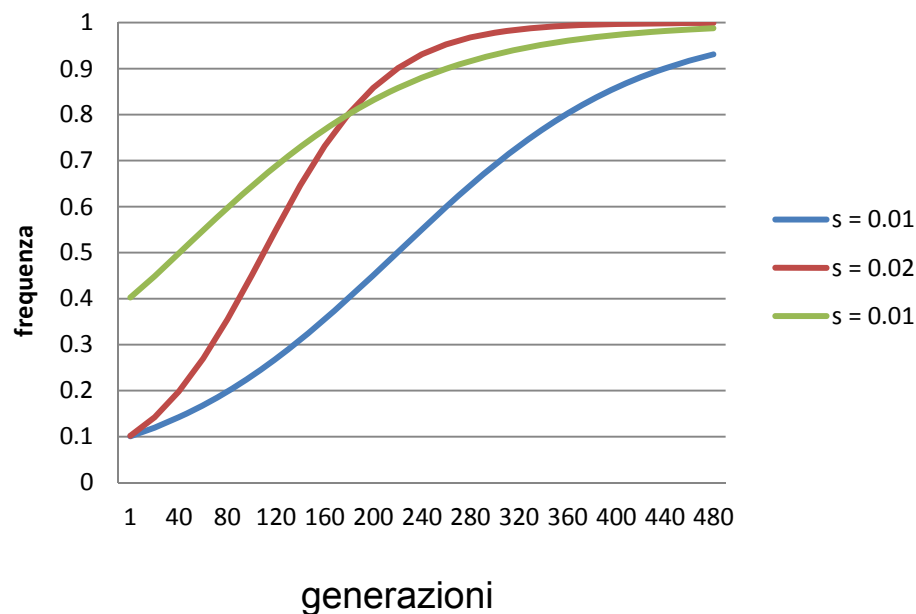
Frequenza di A2 al tempo t

$$q_t = \frac{1}{1 + \left(\frac{1 - q_0}{q_0} \right) e^{-st}}$$

Frequenza di A2 all'inizio

In alternativa si può esprimere t in funzione di q

$$t = \frac{1}{s} \ln \frac{q_t (1 - q_0)}{q_0 (1 - q_t)}$$



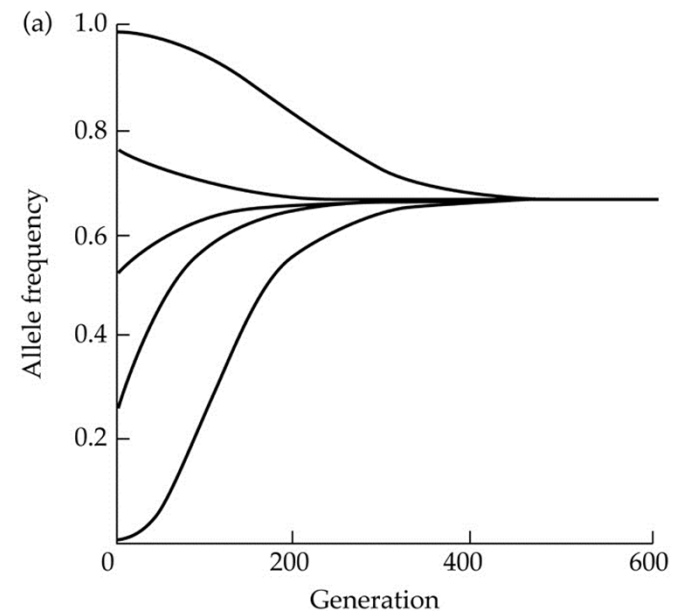
dall'equazione si può calcolare il numero di generazioni necessarie perché la frequenza di A_2 passi da q_0 a q_t

Selezione naturale

Un locus e due alleli: overdominance

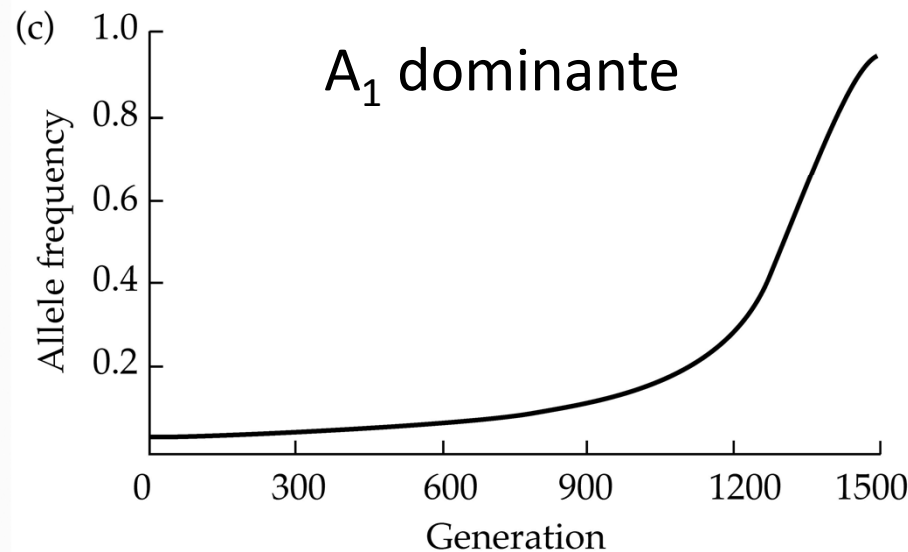
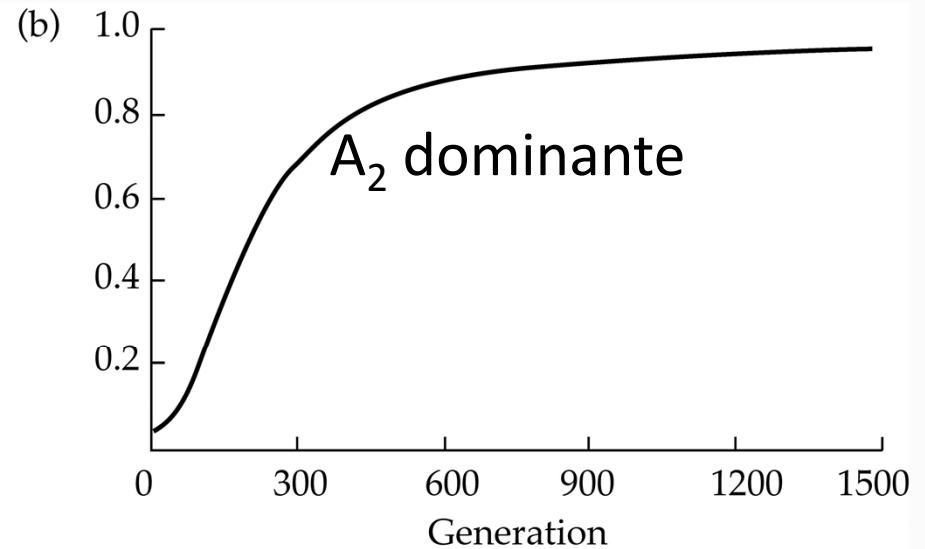
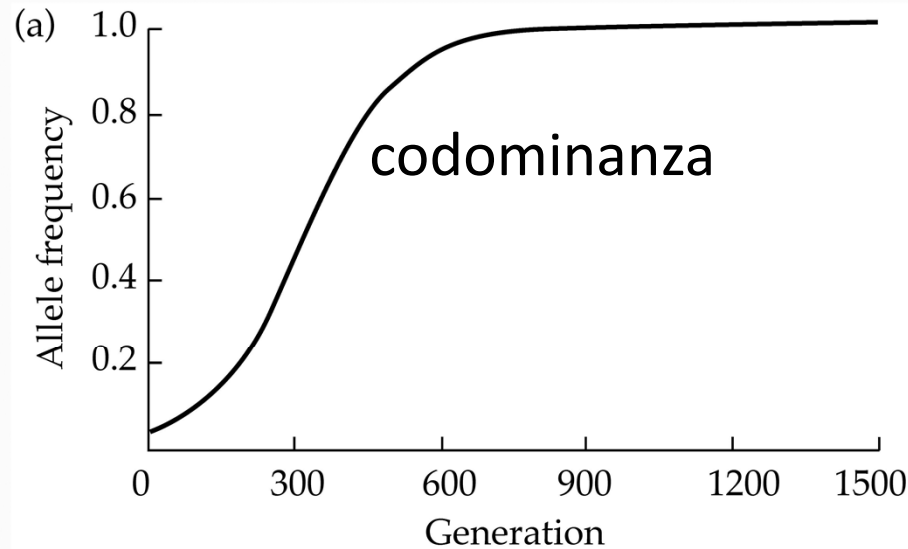
Genotipi	A_1A_1	A_1A_2	A_2A_2
Fitness (w)	1	$1+s_1$	$1+s_2$
in questo caso $s_1 > 0$ e $s_1 > s_2$			

$$\Delta q = \frac{-pq(2s_1q - s_2q - s_1)}{1 + 2s_1pq + s_2q^2}$$



Selezione naturale

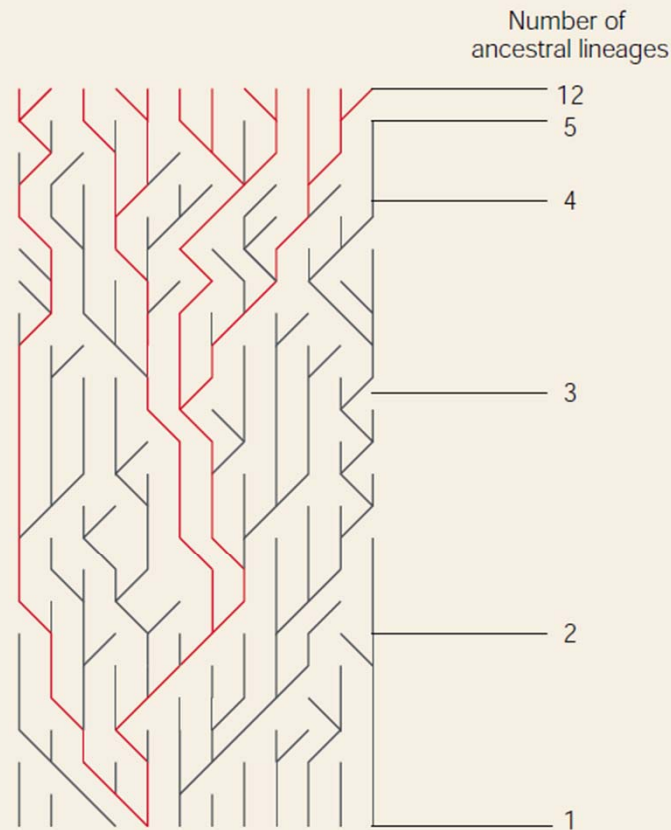
Cambiamento delle frequenze alleliche nel tempo di un allele nuovo A_2 a seconda del modello di dominanza



A_1 = forma ancestrale
 A_2 = forma derivata

Selezione naturale

Box 3 | The coalescent process



If a sample of genes is collected, and the ancestry of those genes is traced backwards in time, common ancestors will be encountered. As these common ancestors are encountered, the number of distinct ancestral lineages decreases from n (the original sample size) to $n - 1$, $n - 2$ and so on, eventually ending in one lineage, which is the common ancestor of the entire sample; this is called the coalescent process. The diagram shows the genealogy of a population of 12 haploid individuals. The red lines trace the ancestry of their 12 lineages back to the most recent common ancestor.

The coalescent process is a point of key interest in efforts to detect natural selection in human genes. Because natural selection can affect the rate at which common ancestors are encountered (for example, compare the time at which the common ancestor is encountered in each panel of FIG. 1), the shape of gene genealogies can inform us about the selective processes that have been involved. Selective sweeps, for example, result in 'shallow' gene genealogies with a few, rare mutations. By contrast, balancing natural selection results in 'deep' gene genealogies in which many mutations are found at intermediate frequencies.

Selezione naturale

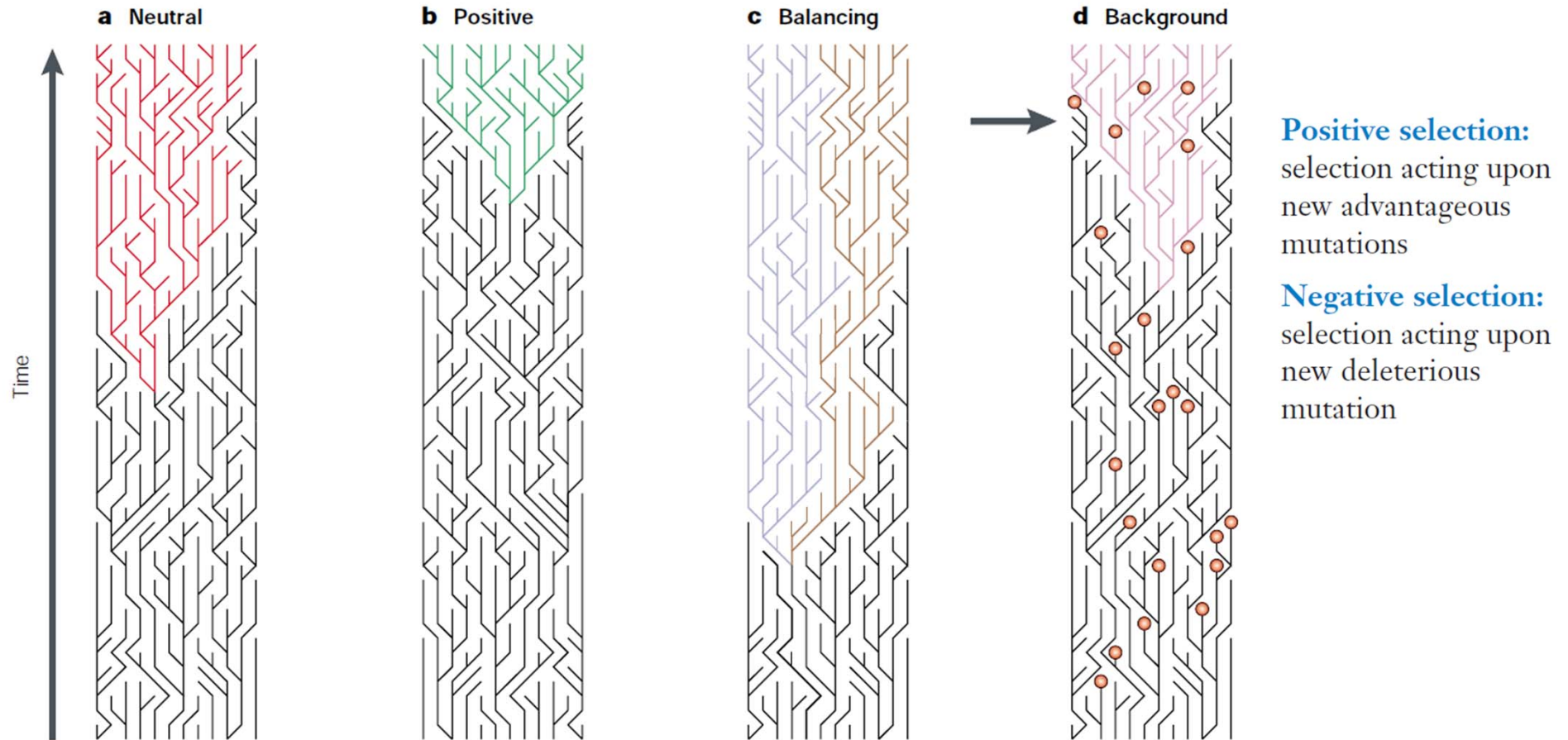


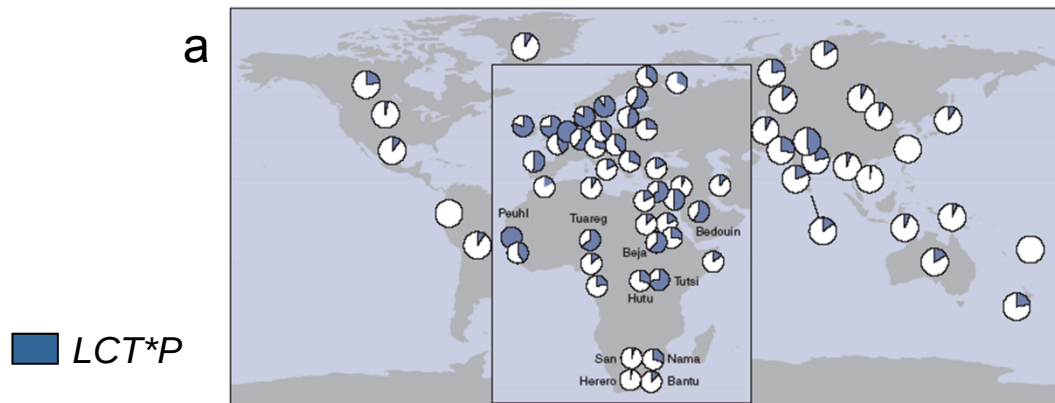
Figure 1 | **Effects of natural selection on gene genealogies and allele frequencies.** Bamshad and Wooding. *Nat Rev Gen* 2003

d | The genealogy of an allele (purple) that drifts to fixation under the influence of BACKGROUND SELECTION. Each circle represents the elimination of a deleterious (that is, lethal) mutation by background selection. The coalescence of the lineage is more recent than expected under a neutral model because a linked deleterious mutation caused the extinction of one lineage (arrow) more quickly than would be predicted.

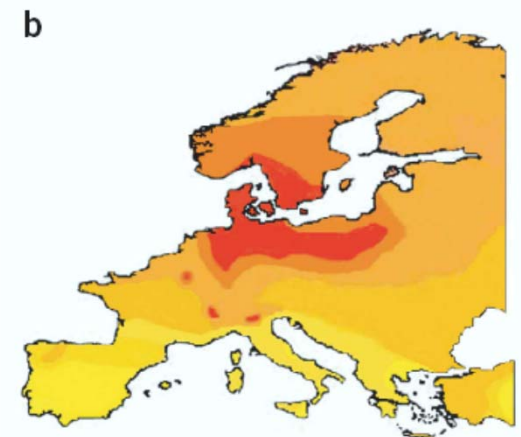
Positive selection in humans: Lactase persistence

- All infants have high lactase enzyme activity to digest the sugar lactose in milk
- In most humans, activity declines after weaning, but in some it persists:

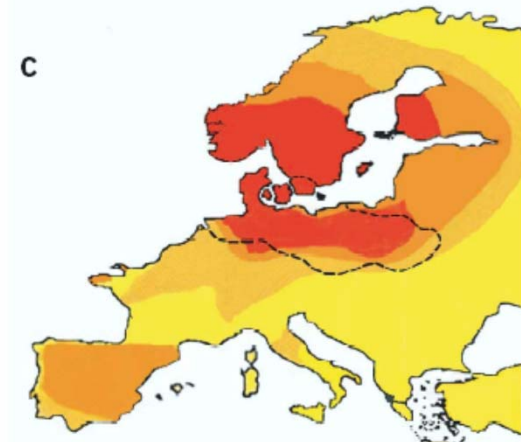
From Jobling 2003



a: Frequency of the lactase persistency (LCT^*P) in different populations



b: Synthetic map showing the first principal component resulting from the allele frequencies at the cattle genes (genes important for milk production)

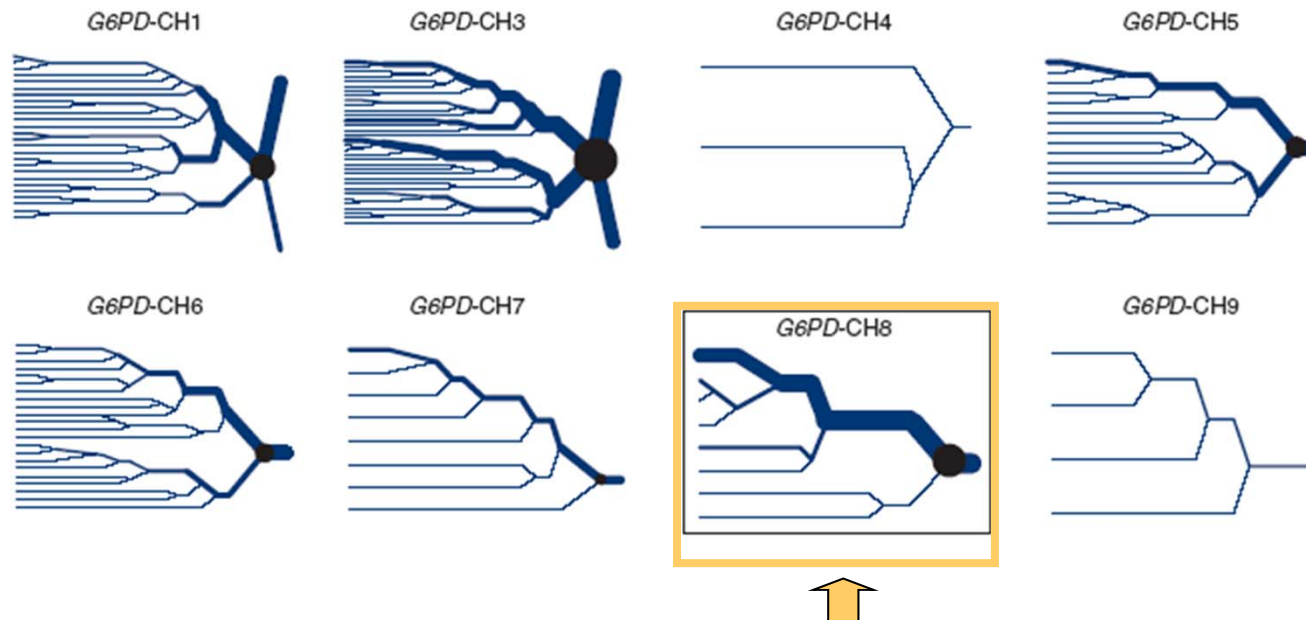


c: Geographic distribution of the lactase persistence allele in contemporary Europeans

Beja-Pereira et al. 2005

Positive selection in humans: Selection at the *G6PD* gene by malaria

- Reduced G6PD enzyme activity (e.g. A allele) confers some resistance to *falciparum* malaria



Extended haplotype homozygosity at the A allele

Haplotter - explore the evidence for recent positive selection in the human genome
<http://hg-wen.uchicago.edu/selection/haplotter.htm>

OPEN ACCESS Freely available online

PLOS BIOLOGY

A Map of Recent Positive Selection in the Human Genome

Benjamin F. Voight¹, Sridhar Kudaravalli¹, Xiaoquan Wen, Jonathan K. Pritchard^{*}

Department of Human Genetics, University of Chicago, Chicago, Illinois, United States of America

The identification of signals of very recent positive selection provides information about the adaptation of modern humans to local conditions. We report here on a genome-wide scan for signals of very recent positive selection in favor of variants that have not yet reached fixation. We describe a new analytical method for scanning single nucleotide polymorphism (SNP) data for signals of recent selection, and apply this to data from the International HapMap Project. In all three continental groups we find widespread signals of recent positive selection. Most signals are region-specific, though a significant excess are shared across groups. Contrary to some earlier low resolution studies that suggested a paucity of recent selection in sub-Saharan Africans, we find that by some measures our strongest signals of selection are from the Yoruba population. Finally, since these signals indicate the existence of genetic variants that have substantially different fitnesses, they must indicate loci that are the source of significant phenotypic variation. Though the relevant phenotypes are generally not known, such loci should be of particular interest in mapping studies of complex traits. For this purpose we have developed a set of SNPs that can be used to tag the strongest ~250 signals of recent selection in each population.

Heterozygote advantage – Overdominance in humans: Sickle-cell disease

	Genotypes		
	A_1A_1	A_1A_2	A_2A_2
Relative fitness	w_{11}	w_{12}	w_{22}
Heterozygote advantage	$1 - s_1$	1	$1 - s_2$
6 Glu>Val	AA	AS	SS
	HbA HbA	HbA HbS	HbS HbS



Red blood cells: -
Malaria susceptibility: +

Red blood cells: +
Malaria susceptibility: -

