23.
Evoluzione molecolare & Genetica di popolazione

I GENI NELLE POPOLAZIONI.
A) Variabilità genetica

Genetica di popolazione

Lo scopo della Genetica di popolazioni è identificare la composizione genetica di una popolazione e i fattori che determinano e modificano tale composizione

Compiti della genetica di popolazioni:

- misurare la ``quantità di variazione genetica'' esistente nelle popolazioni naturali
- spiegare questa variazione:
- capirne l'origine
- capire come viene mantenuta
- capirne la rilevanza evoluzionistica ed ecologica.

Uno dei principi della genetica di popolazioni è che non ci può essere evoluzione se non c'è variazione genetica


Variabilità genetica
Evoluzione
Adattamento
Biodiversità
Conservazione della specie
Speciazione

Individui, popolazioni

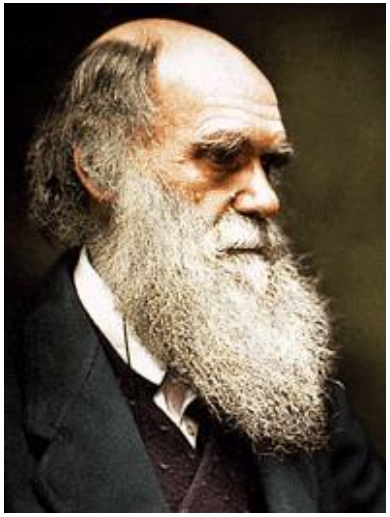
- **Gli individui hanno i loro alleli e genotipi**,
  - **le popolazioni** le loro frequenze alleliche e genotipiche

- Ogni individuo è come è, le frequenze alleliche cambiano nel tempo e questa è l'evoluzione: **genetica evoluzionistica**

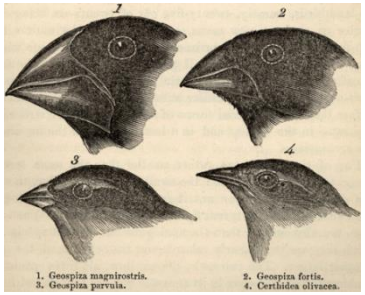
- Gli individui si ammalano e guariscono, le popolazioni sono caratterizzate da frequenze di alleli patologici: **epidemiologia genetica**

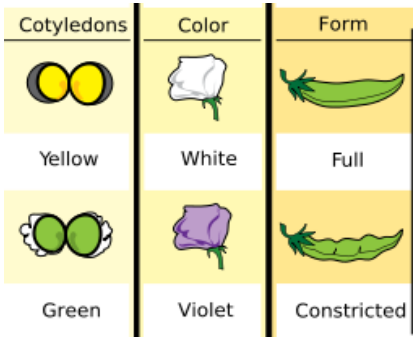
- Gli individui nascono e muoiono, le popolazioni possono espandersi, contrarsi ed estinguersi: **genetica della conservazione**

Un bel problema



- L'**evoluzione**, guidata dalla selezione naturale, modifica i caratteri in maniera graduale ed ereditaria;
  - **la variazione biologica è continua**





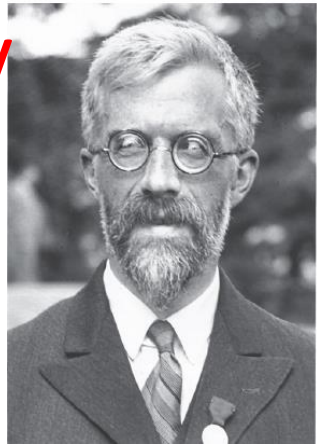


- La trasmissione dei caratteri dipende da fattori discreti (i geni) che, ereditati dai genitori, non si mescolano nella cellula;
- **la variazione biologica è discontinua**

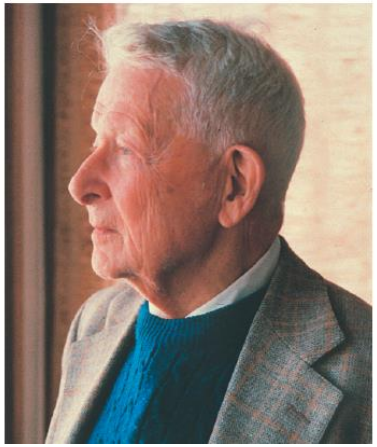
Come è possibile che fattori discreti determinino una variabilità continua?

La sintesi moderna/ neodarwinismo: Fisher, Wright, Dobzhansky, Haldane

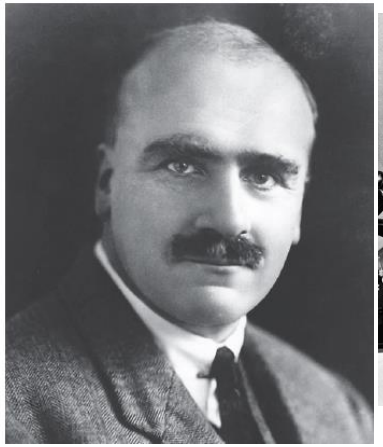
Sir Ronald Fisher



Sewall Wright



John Burdon Sanderson Haldane





- Essa deriva dall'integrazione tra:
  - la **teoria dell'evoluzione delle specie** per selezione naturale di Charles Darwin;
  - la **teoria dell'ereditarietà** di Gregor Mendel sulle basi dell'eredità biologica rivista alla luce della moderna genetica, comprese le mutazioni genetiche casuali come sorgente della variazione;
  - la **forma matematica della genetica delle popolazioni;**
  - **l'analisi dei dati** della paleontologia.
- In sintesi il neodarwinismo consiste nel considerare il **gene come unità fondamentale dell'eredità e bersaglio del meccanismo evoluzionistico della selezione naturale**.
- La sintesi neodarwiniana unifica diverse branche della biologia che in precedenza avevano pochi punti di contatto, in particolare la genetica, la citologia, la sistematica, la botanica e la paleontologia.

Ernst MAYR, Understanding evolution. “Trends in Ecology and Evolution”, v. 14, anno 1999, p. 372



La biodiversità

- La biodiversità è la varietà di
  - ecosistemi,
  - specie e
  - popolazioni all'interno delle specie

- **Perché è importante mantenere la biodiversità?**
  - Per le biorisorse
  - Per i servizi dell'ecosistema
  - Per il piacere estetico
  - Per il diritto all'esistenza degli organismi viventi

Prima di tutto: non c'è genetica senza variabilità

(a)



(b)



- la variabilità è **prodotta dal processo di mutazione**
- La variabiltà della struttura genetica esistente nelle popolazioni è superiore a quella riconducibile alle differenze fenotipiche.
- Essa può essere misurata analizzando i polimorfismi della sequenza di DNA e individuando i diversi genotipi.
- Polimorfismo: proprietà di un gene di avere 2 o più alleli con frequenze maggiori dell' 1%.
- **Nella specie umana due individui differiscono in media per 3 milioni di nucleotidi (0.1 % del genoma)**

Polimorfismo



ˆ Variazione presente nella popolazione con una frequenza superiore a 1%

Variazioni nell'aspetto



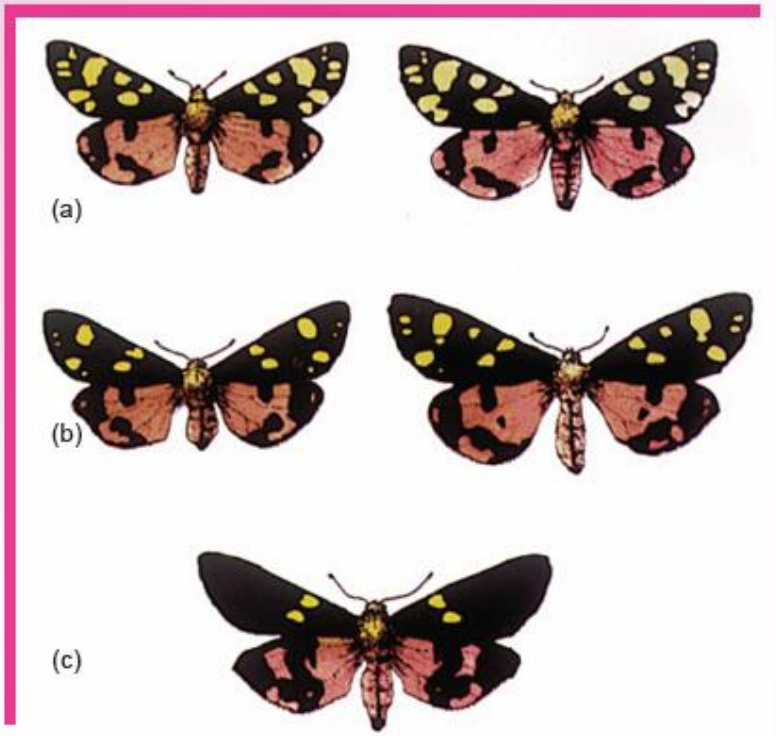




y NA

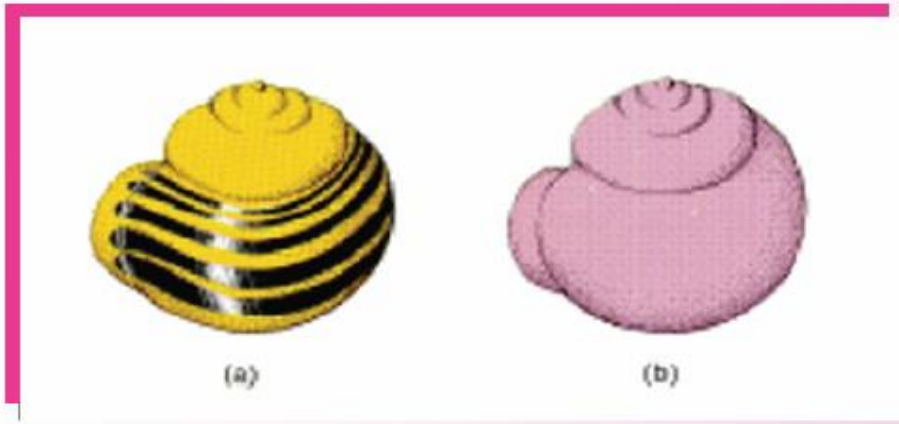
Variabilità a livello morfologico

Panaxia dominula, la falena tigre



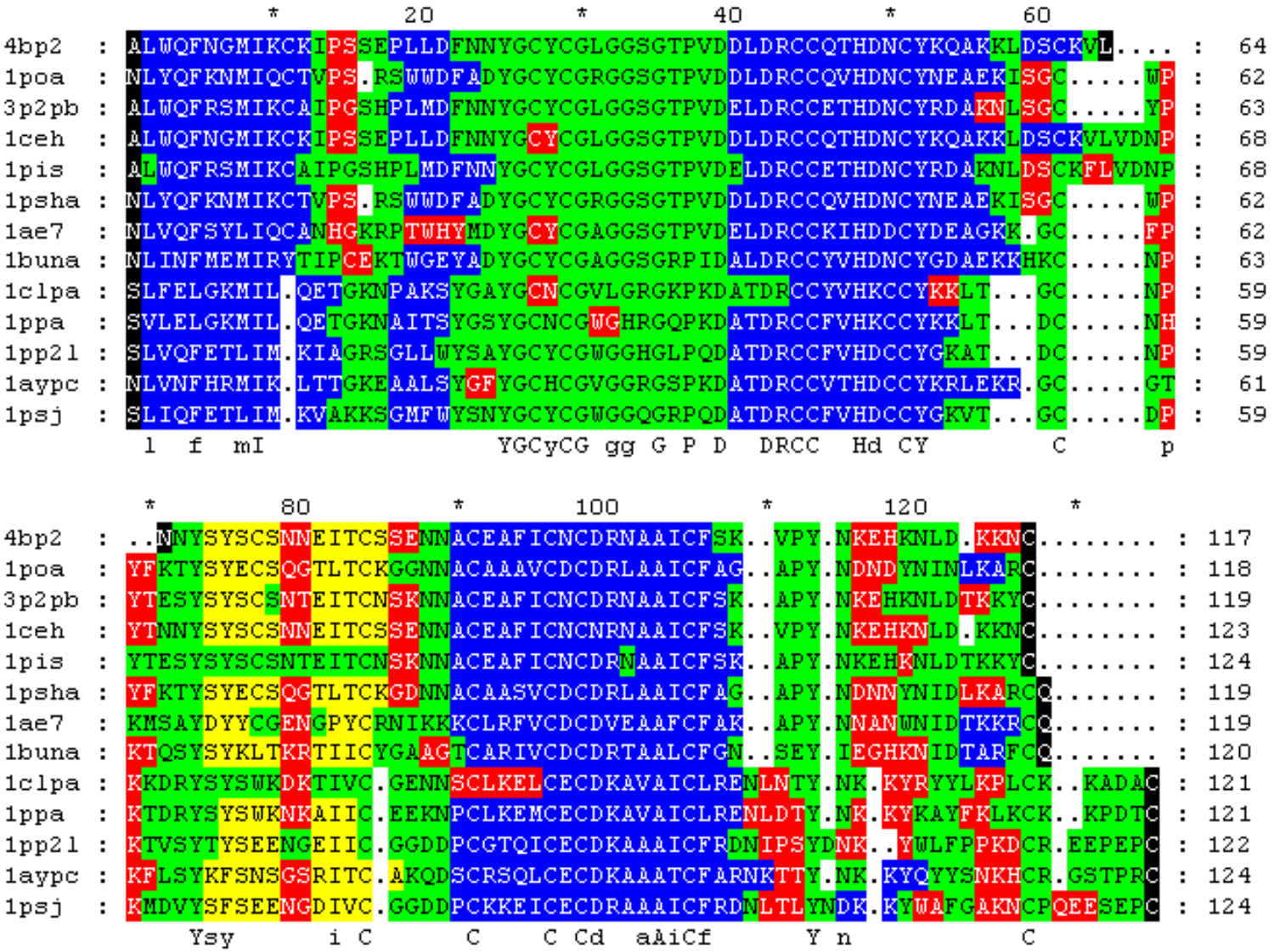
Tre varianti fenotipiche della screziatura: (a) molte chiazze; (b) poche chiazze; (c) pochissime chiazze

Cepea nemoralis, la lumaca terrestre



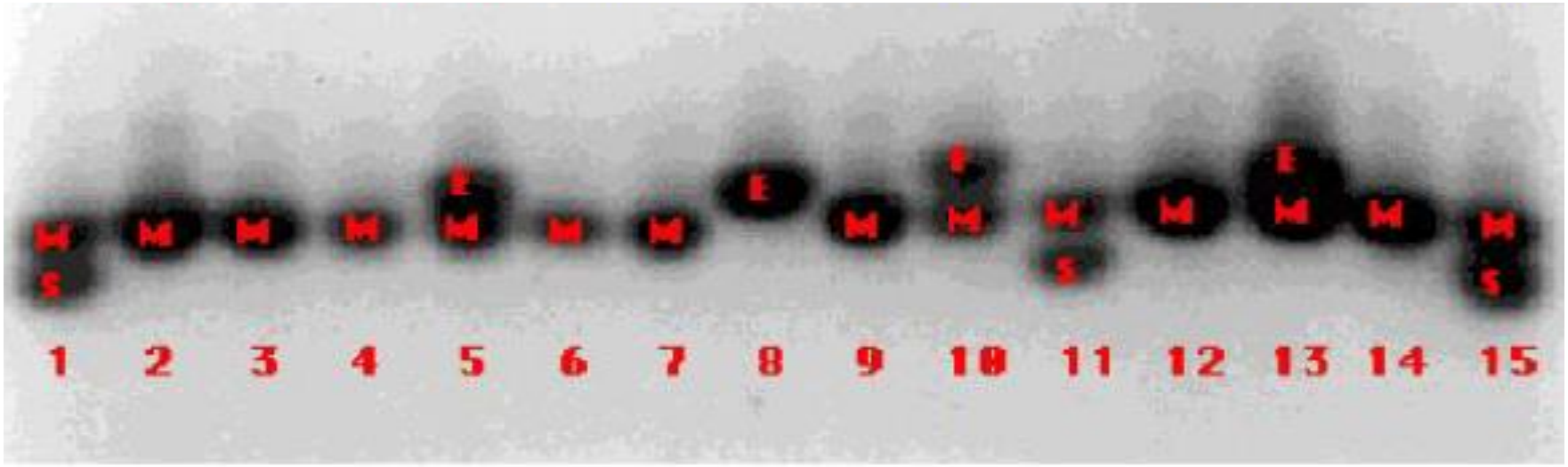
Due varianti fenotipiche del guscio: (a) giallo a strisce; (b) rosa senza strisce

Variabilità genetica: proteine



Polimorfismo proteico

- Variazione presente nella popolazione con una frequenza superiore a 1%

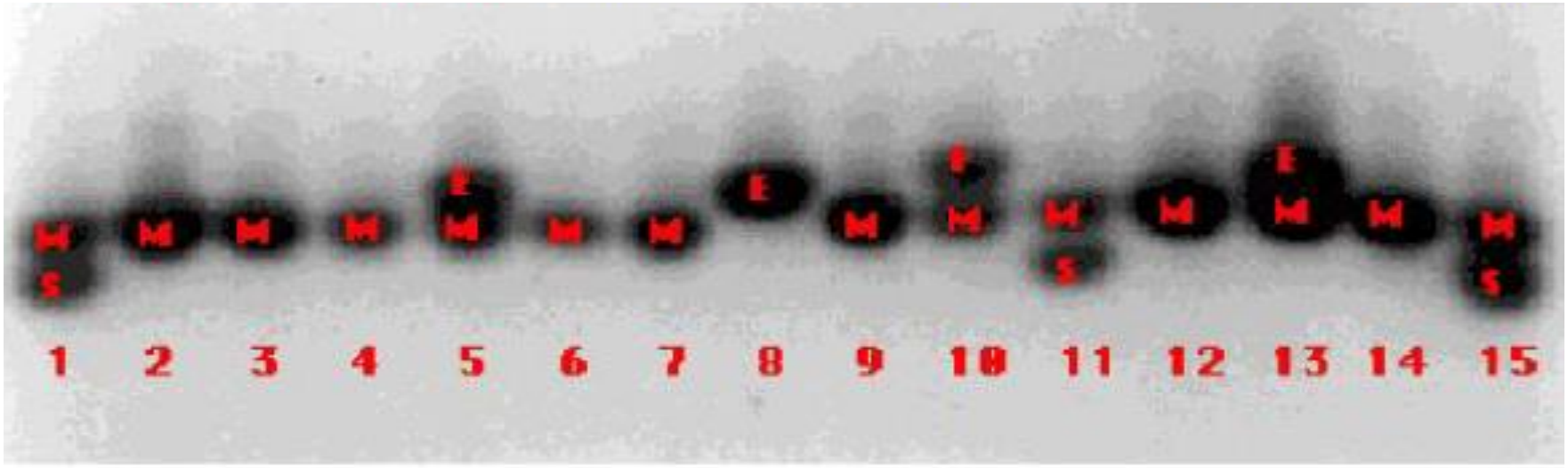


che cosa mi dice questo gel?

- nella popolazione esistono piu' fenotipi di questo enzima: fast, medium, slow: esistono 3 alleli a questo locus.
- La velocita' della corsa dipende dalla carica della molecola, quindi ci sono delle **sostituzioni amminoacidiche** che modificano la carica.
- Queste sostituzioni non coinvolgono siti importanti per la funzione della proteina perche' a meno di fare il gel non e' possibile individuare le molecole varianti.
- Cioe' a livello del fenotipo funzione della proteina, o fenotipo salute dell'individuo gli M, gli F, o MF, MS sono indistinguibili

Polimorfismo proteico

- Variazione presente nella popolazione con una frequenza superiore a 1%



- Perche' trovo individui MF, MS,? Perche' non ho trovato individui solo S?
- HW: dipende dalla frequenza nella popolazione dell'allele S
- Ovviamente la **sostituzione amminoacidica corrisponde** ad una **mutazione nel DNA e** se conosco la sequenza del DNA forse riesco a individuare un fenotipo variante del DNA cioe' un polimorfismo del DNA

- **I geni trascritti in proteine rappresentano negli Eucarioti fra il 2% e il 10% del genoma**

- **Parte del restante 90-98% non è funzionale (DNA ripetitivo), ma un'altra parte contiene importanti siti di regolazione**

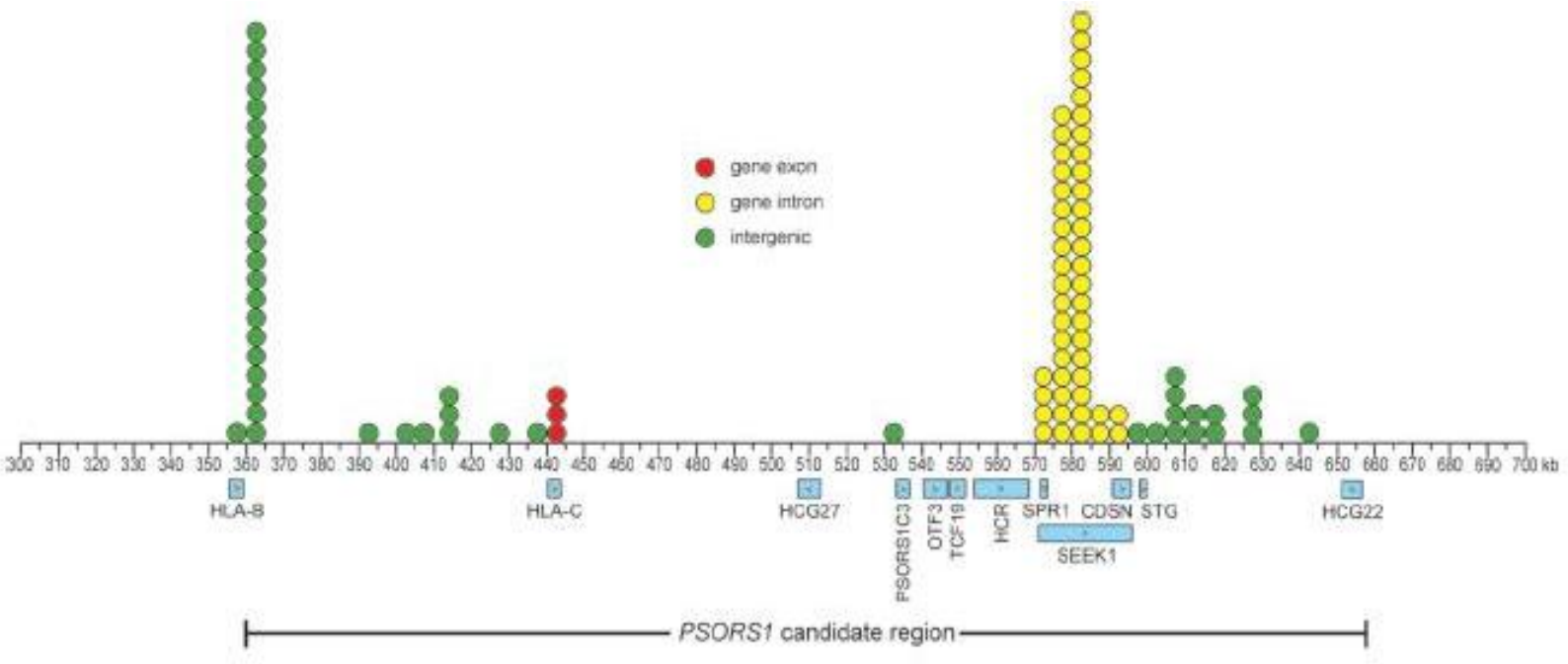


Figure 3 Locations of polymorphisms unique to risk haplotypes within the *PSORS1* candidate region. Polymorphisms with alleles borne by the two sequenced risk haplotypes that differ from the alleles borne by all eight sequenced nonrisk haplotypes are plotted, with circles, to the nearest 5-kb interval. Fill color of the circle indicates whether the polymorphism occurs within a gene exon (*red*), gene intron (*yellow*), or intergenic region (*green*). Known genes and their direction of transcription and the *PSORS1* candidate interval are shown below the coordinate axis.

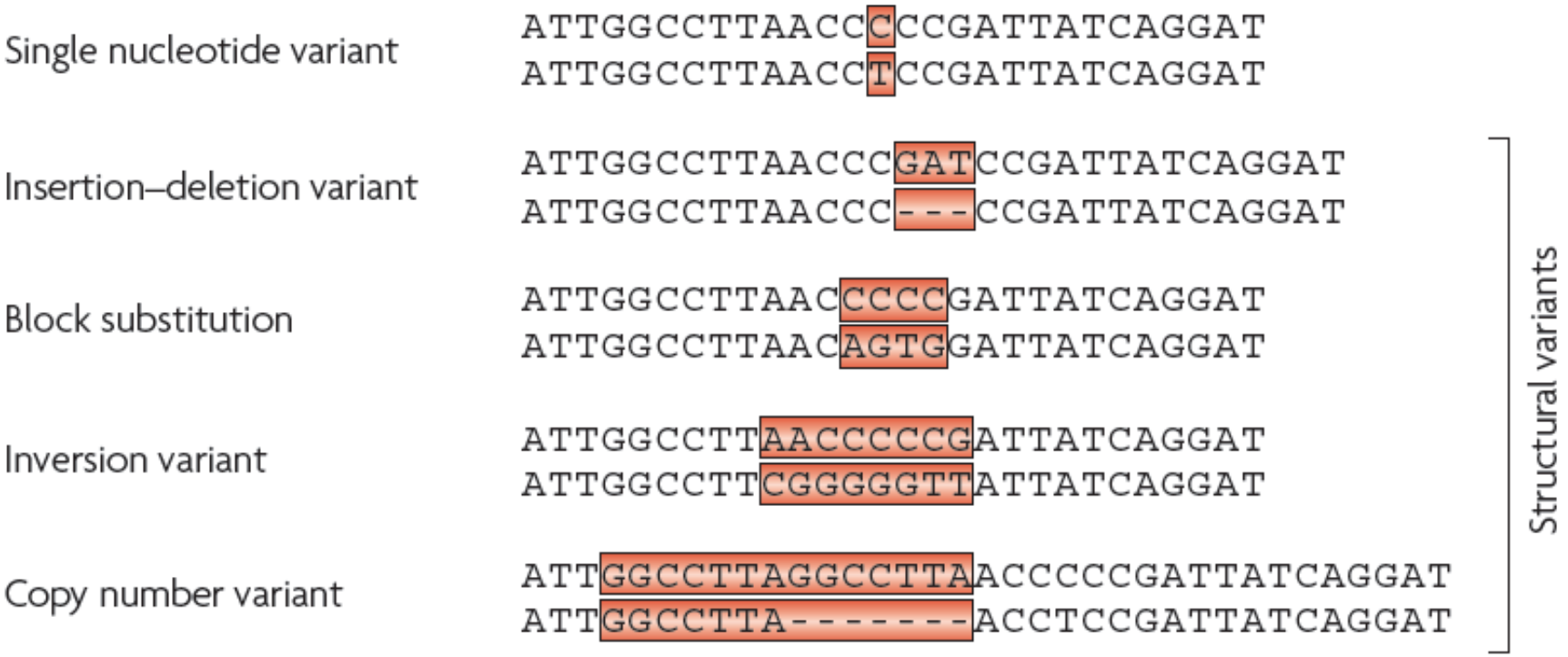
I vantaggi della variabilità a livello del DNA

- Maggiore variabilità: solo il 30% dei cambiamenti del DNA determina un cambiamento nelle proteine quindi la variabilità proteica sottostima il vero livello di diversità genetica.

- Le proteine vengono raccolte da campioni di sangue, fegato o rene per gli animali, di foglie e apici radicali per le piante, poiché questi tessuti contengono abbondanti quantità e varietà di proteine solubili. Di conseguenza gli animali devono essere catturati o uccisi. Il DNA può essere estratto da: peli, pelle, piume, feci, urine, gusci d'uovo, squame di pesce, sangue, saliva, spermi, pertanto il campione biologico può essere prelevato in modo non invasivo.

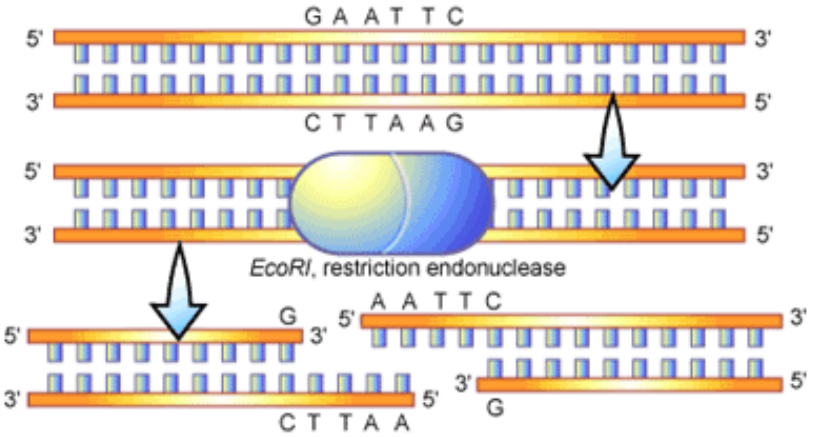
- Il DNA è una macromolecola più stabile delle proteine.

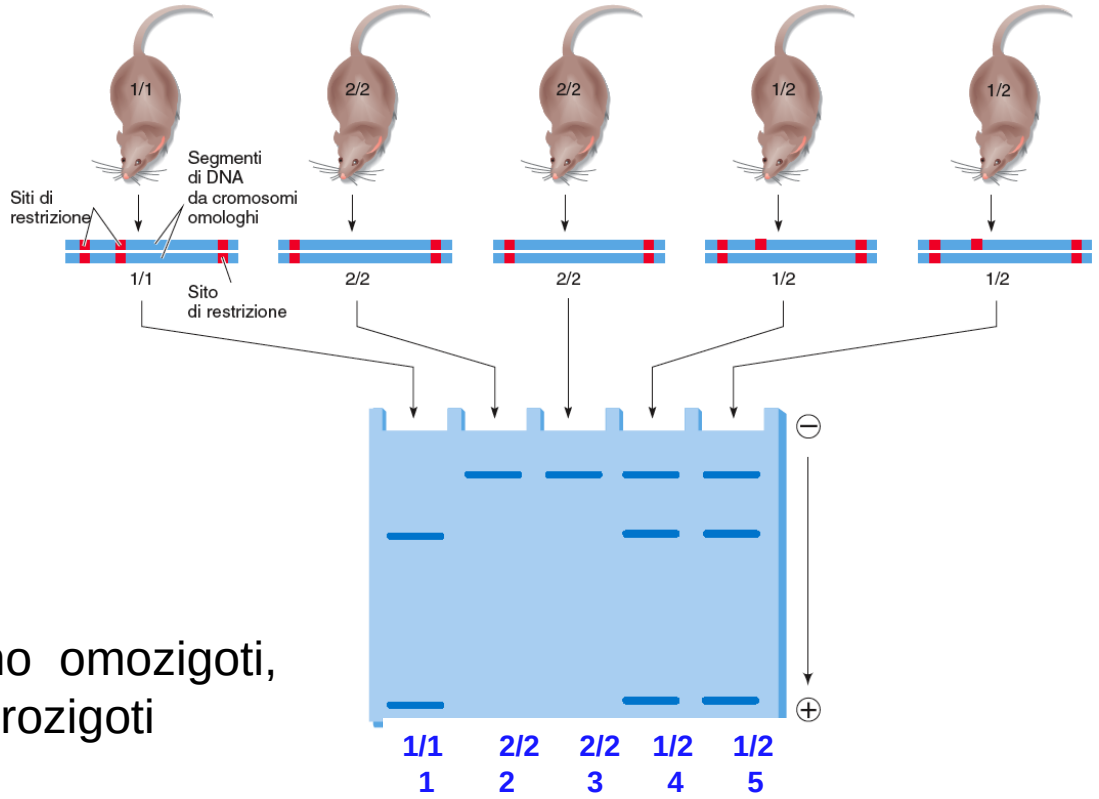
Tipi di polimorfismo studiati nel DNA



1. **Single Nucleotide Polymorphism: SNP**
2. **Inserzione/delezione**
3. **Variazione del numero di copie: STR, VNTR, CNV**

Polimorfismi di restrizione





- I topini 1, 2 e 3 sono omozigoti,
- il 4 e il 5 sono eterozigoti

Polimorfismi di restrizione

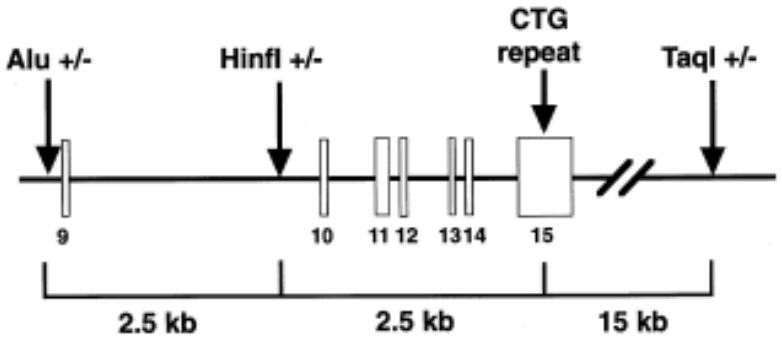
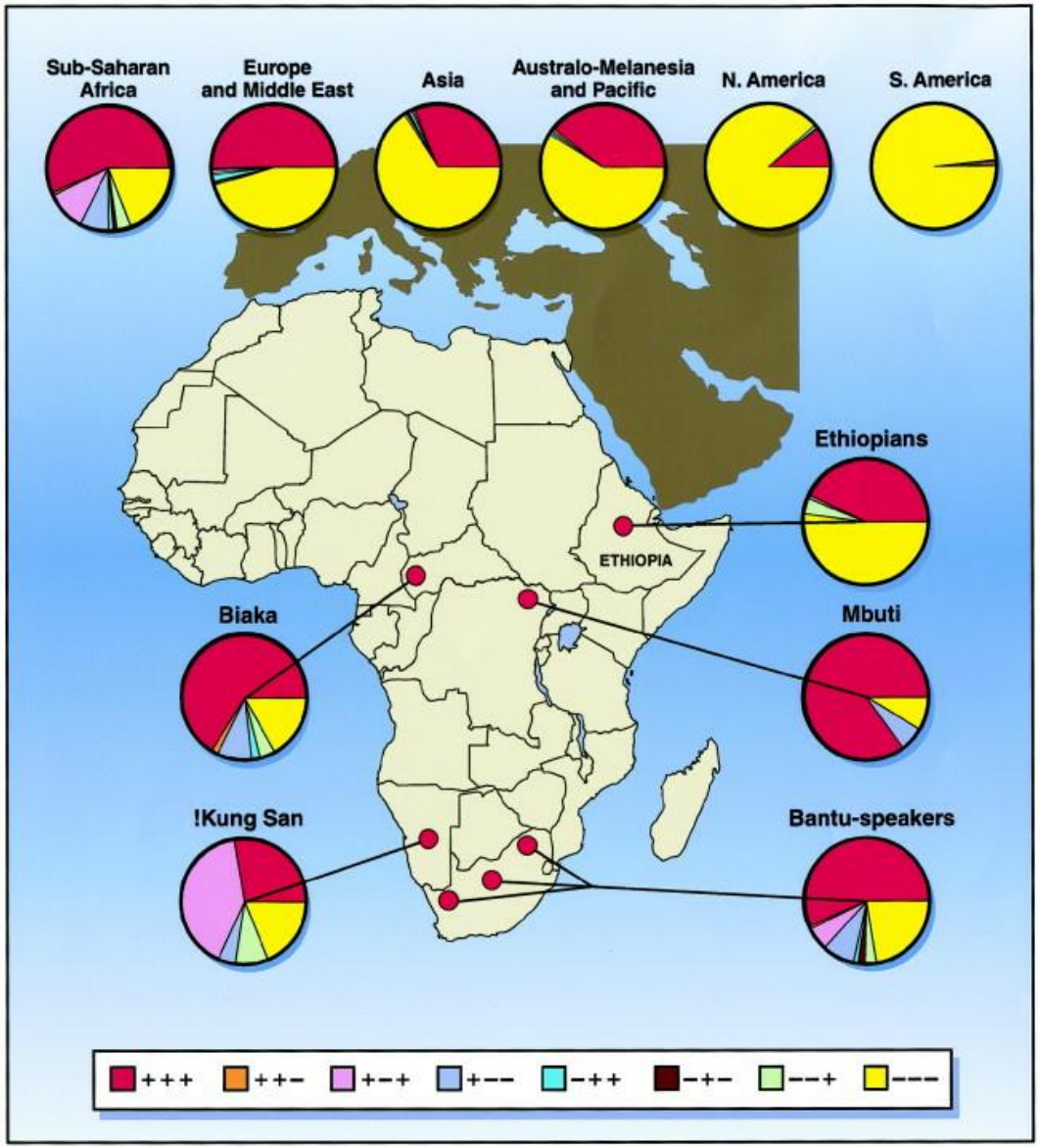
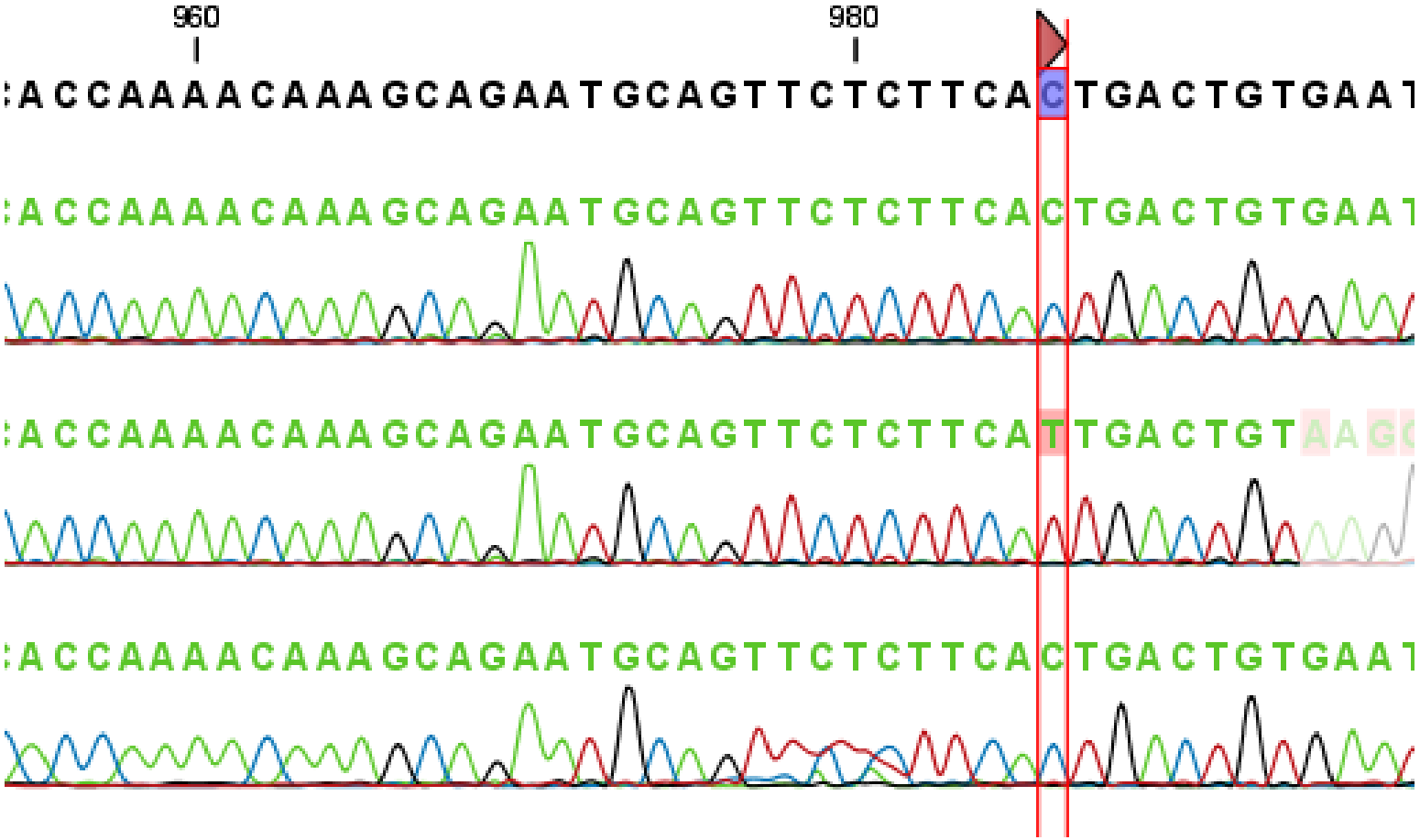


Figure 1 Schematic map of genomic DNA, showing exons 9–15 of the DMPK gene and the relative location of the *Alu*, *Hin*fI, and *Taq*I polymorphisms. "*Alu*(+)" denotes that the full-length *Alu* region is present; and "*Alu*(−)" denotes that it is absent; "*Hin*fI(+)" denotes that the *Hin*fI restriction site is present; and "*Taq*I(+)" denotes that the *Taq*I restriction site is present.

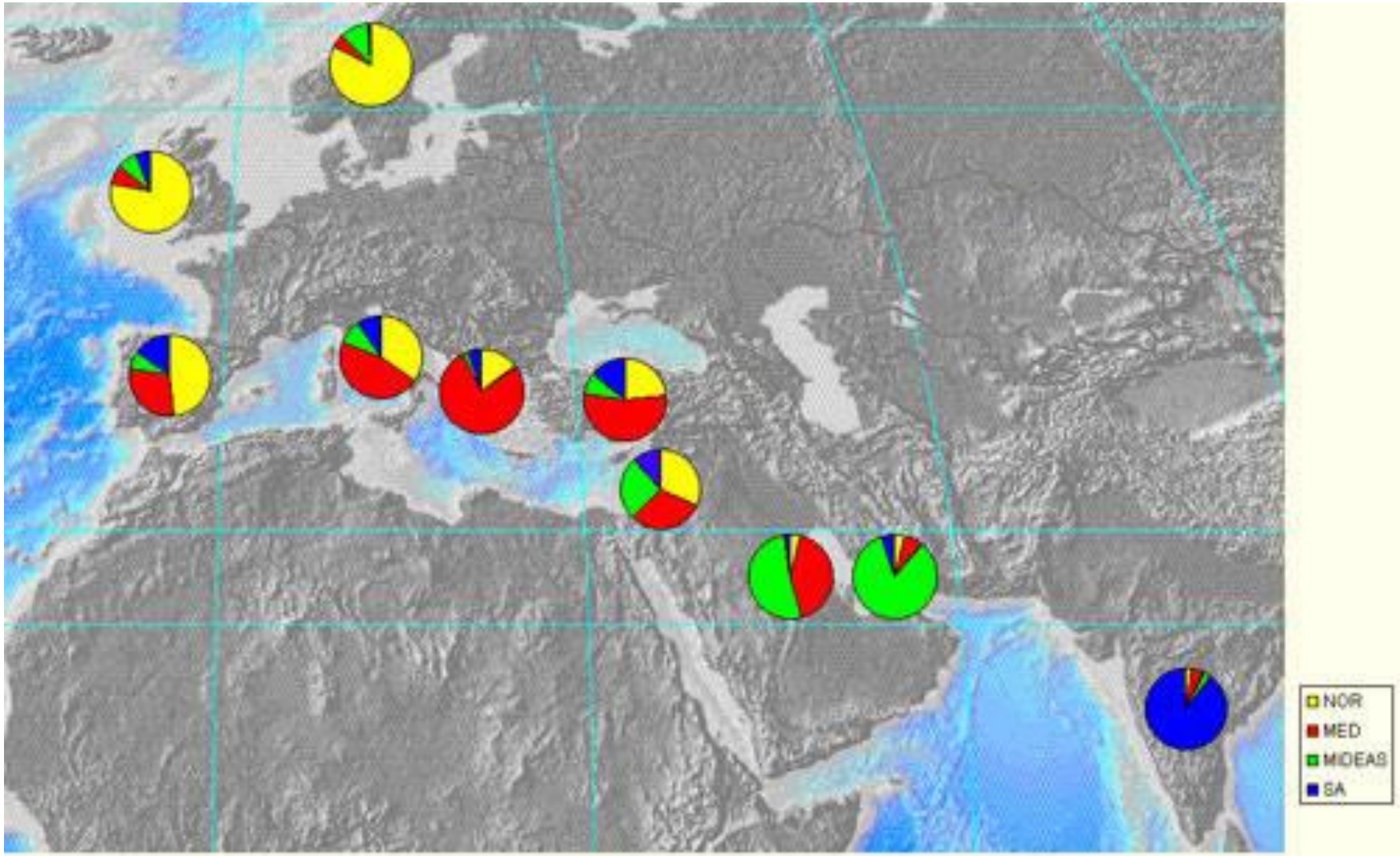


Polimorfismi di sequenza del DNA

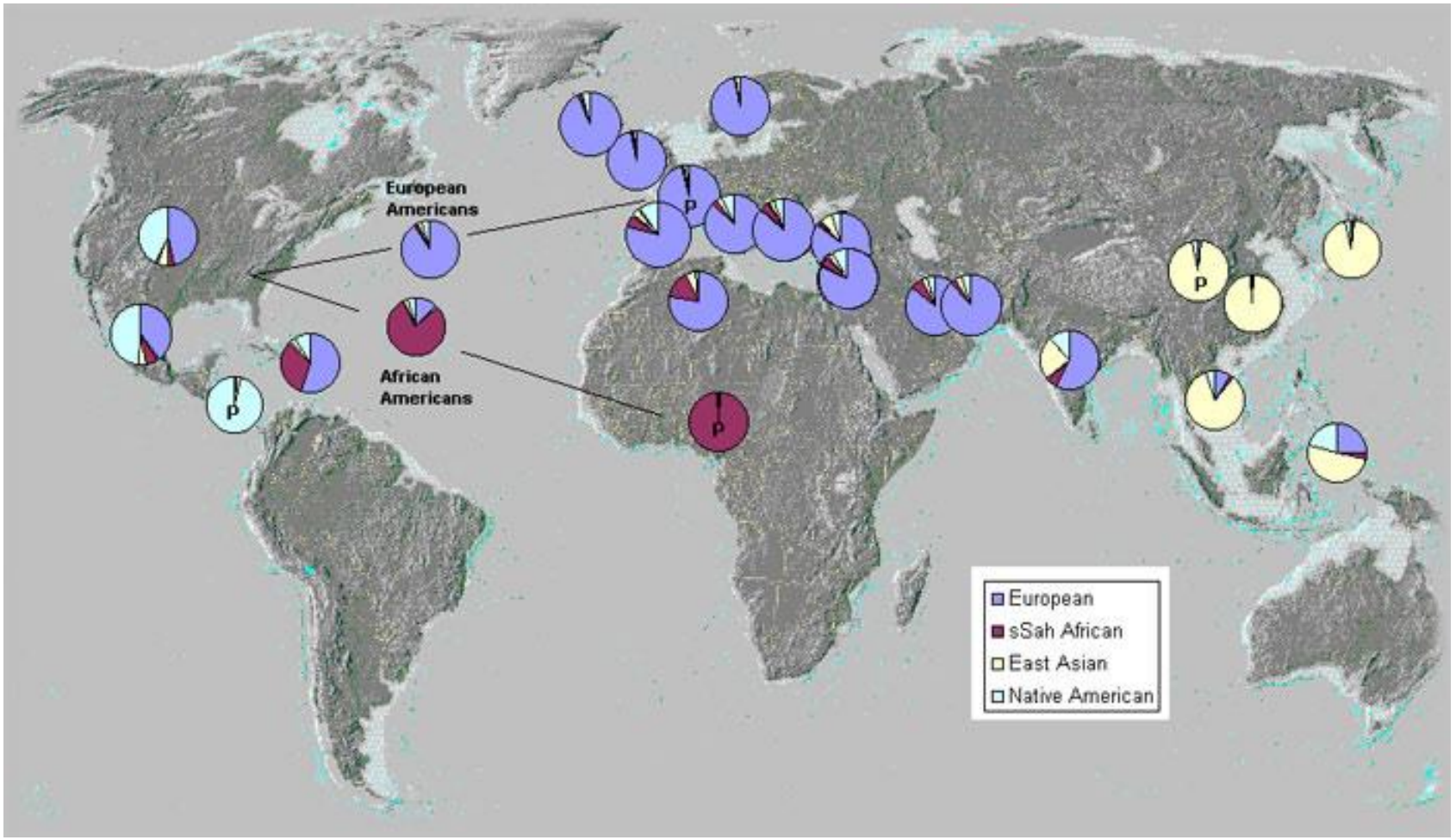


Qual è il vantaggio del sequenziamento rispetto alla restrizione?

Si possono rappresentare le frequenze alleliche con diagrammi a torta



La variabilità delle frequenze alleliche aiuta a definire le origini degli individui



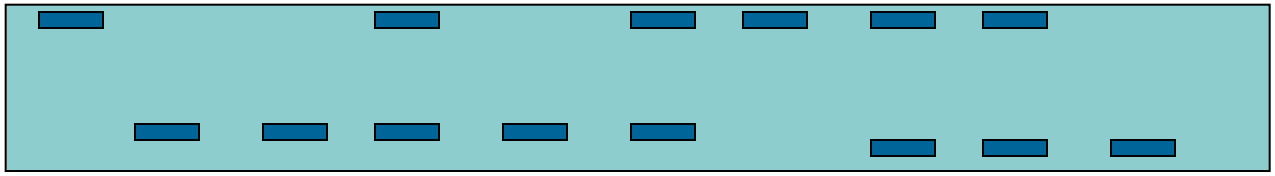
Terminologia per descrivere la diversità genetica

Locus: il sito sul cromosoma in cui è localizzato un particolare gene.
Alleli: differenti varianti di un gene.
Genotipo: la combinazione degli alleli presenti in un locus in un individuo. I genotipi sono eterozigoti o omozigoti.
Omozigote: un individuo con due copie dello stesso allele in un locus. Es. AA o aa.
Eterozigote: un individuo con due differenti alleli in un locus. Es. Aa.
Frequenza allelica: la frequenza relativa di un particolare allele in una popolazione.
Polimorfico: la presenza in una specie di due o più alleli in un locus.
Monomorfico: un locus caratterizzato da un solo allele.

Come si determina la variabilità genetica di una popolazione?

- **Calcolando le frequenze**
  - **dei genotipi**
  - **o degli alleli nei genotipi**
  - **o degli alleli nella popolazione**

Frequenze alleliche



F S S FS S FS F FS FS S **fenotipo osservato**

FF SS SS FS SS FS FF FS FS SS genotipo dedotto

- **Freq. genotipiche: FF=0,2 FS=0,4 SS=0,4**
- **Freq. alleliche: f(F)=p, f(S)=q p+q=1**

$p = (NF + ½\ NH) / NT = (2NF + NH)/2NT$

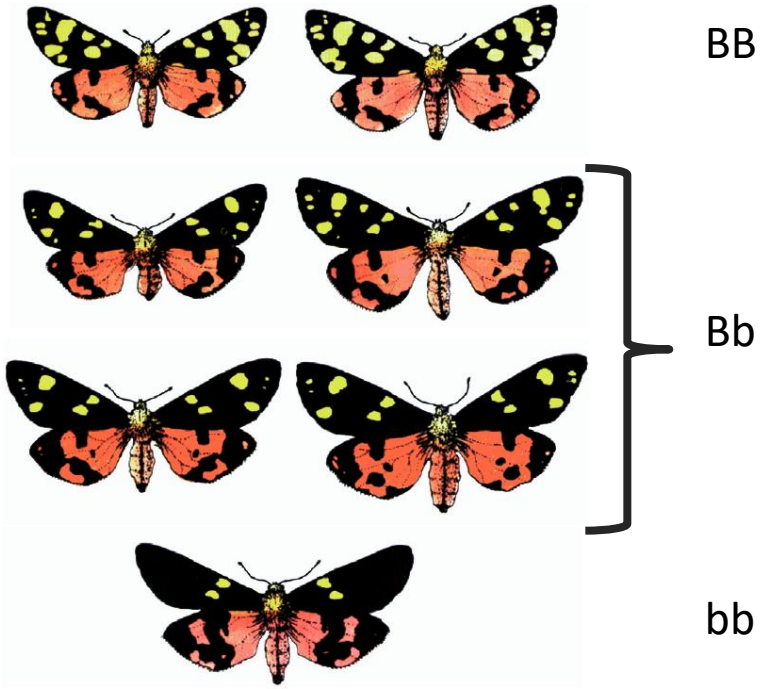
$p = (2 + 2)/10 = 0,4$ $p= (4 + 4)/20 = 0,4$

$q = (4 + 2)/10 = 0,6$ $q = (8 + 4)/20 = 0,6$

$p+q = 0,4 + 0,6 = 1$

La frequenza allelica nella popolazione

- è il **rapporto fra**
  - **il numero di copie di alleli di un certo** allele
  - e il **numero totale di alleli di quel gene nella** popolazione



Pigmentazione nella falena tigre, *Panaxia dominula*

Figura 21.2

Calcolo delle frequenze genotipiche

Panaxia dominula, la falena tigre



$$f(AA) = \frac{\text{numero degli individui } AA}{\text{numero totale degli individui}} = \frac{452}{497} = 0{,}909$$

$$f(Aa) = \frac{\text{numero degli individui } Aa}{\text{numero totale degli individui}} = \frac{43}{497} = 0{,}087$$

$$f(aa) = \frac{\text{numero degli individui } aa}{\text{numero totale degli individui}} = \frac{2}{497} = 0{,}004$$

Una frequenza rappresenta una percentuale o una proporzione ed è di solito espressa come frazione decimale. La somma di tutte le frequenze genotipiche è sempre uguale a 1

Calcolo delle frequenze alleliche

Esistono due metodi per calcolare la frequenza di un allele:

A. Calcolo delle frequenze alleliche sulla base del numero di alleli

$$\text{Frequenza di un allele} = \frac{\text{numero di copie dell'allele}}{\text{numero totale degli alleli di quel locus}}$$

$$p = f(A) = \frac{2\,(\text{individui AA}) + (\text{individui Aa})}{2\,(\text{numero totale degli individui})} = \frac{2\,(452) + 43}{2\,(497)} = 0{,}953$$

$$q = f(a) = \frac{2\,(\text{individui aa}) + (\text{individui Aa})}{2\,(\text{numero totale degli individui})} = \frac{2\,(2) + 43}{2\,(497)} = 0{,}047$$

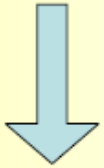
$$p + q = 1$$

B. Calcolo delle frequenze alleliche sulla base delle frequenze genotipiche

$p = f(A) = f\,(AA) + ½\, f\,(Aa) = 0{,}909 + ½\,(0{,}087) = 0{,}953$

$q = f(a) = f\,(aa) + ½\, f\,(Aa) = 0{,}004 + ½\,(0{,}087) = 0{,}047$

$$p + q = 1$$



$q = f(a) = 1 - p = 1 - 0{,}953 = 0{,}047$

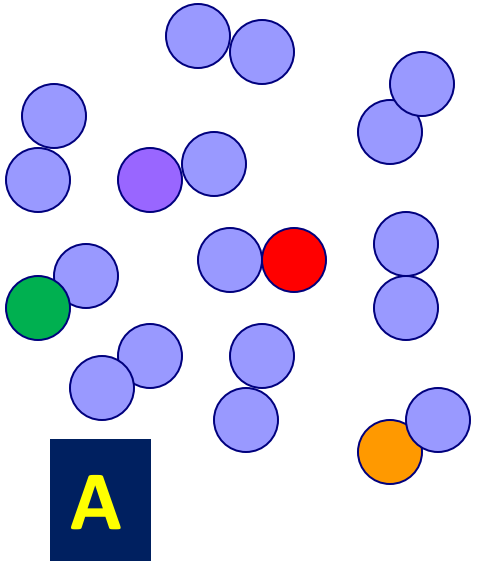
Proporzione di loci polimorfici (*P*): Numero di loci polimorfici / numero totale dei loci campionati. Per esempio, se 3 loci su 10 sono polimorfici e 7 sono monomorfici, allora *P*= 3/10 = 0.3.

Eterozigosità media (*H*): La somma delle proporzioni degli eterozigoti per tutti i loci / numero totale dei loci campionati. Per esempio, le proporzioni di individui eterozigoti per 10 loci in una popolazione sono 0.2, 0.4, 0.1, 0, 0, 0, 0, 0, 0, 0, allora:

H= (0.2+0.4+0.1+0+0+0+0+0+0+0)/10 = 0.07

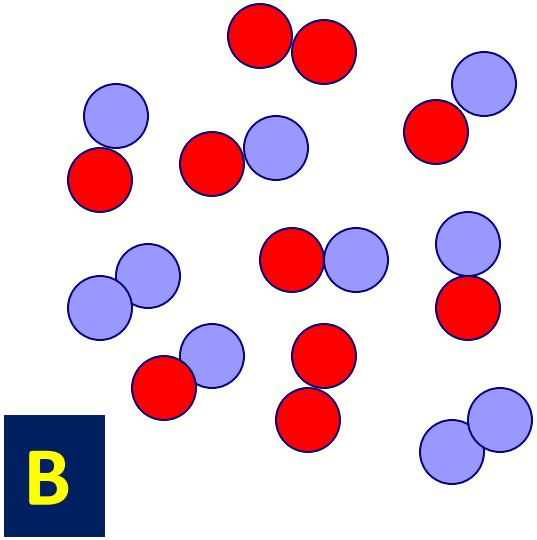
Diversità allelica (*A*): Numero medio di alleli per locus. Per esempio, se il numero di alleli per 10 loci è 2, 3, 2, 1, 1, 1, 1, 1, 1, 1, allora *A*= (2+3+2+1+1+1+1+1+1+1)/10 = 1.4

Quand’è che una popolazione può dirsi variabile?



N alleli = 5

H_O = 0.4



N alleli = 2

H_O = 0.6

- Quando il genotipo individuale è difficile da prevedere

Misure di diversità genetica

- **N di alleli**
- **Eterozigosi osservata:**

  H_o = N genotipi eteroz./N genotipi totali

- **Eterozigosi attesa:**

  $H = 1 - \Sigma p_i^2$

  (la frazione di individui che ci si aspetta siano eterozigoti a un gene sconosciuto)

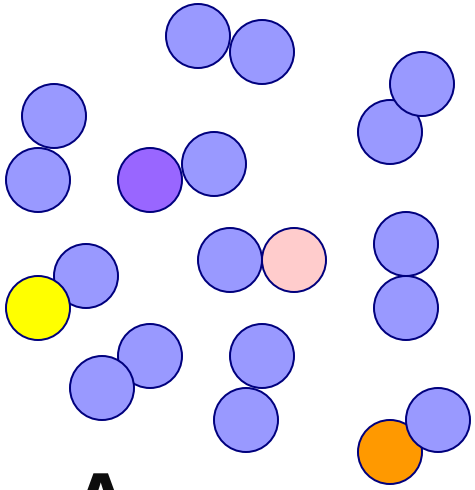
Proporzione di loci polimorfici (*P*): Numero di loci polimorfici / numero totale dei loci campionati. Per esempio, se 3 loci su 10 sono polimorfici e 7 sono monomorfici, allora *P*= 3/10 = 0.3.

Eterozigosità media (*H*): La somma delle proporzioni degli eterozigoti per tutti i loci / numero totale dei loci campionati. Per esempio, le proporzioni di individui eterozigoti per 10 loci in una popolazione sono 0.2, 0.4, 0.1, 0, 0, 0, 0, 0, 0, 0, allora:

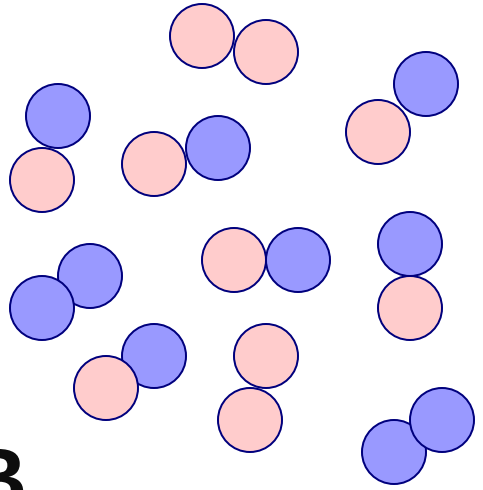
H= (0.2+0.4+0.1+0+0+0+0+0+0+0)/10 = 0.07

Diversità allelica (*A*): Numero medio di alleli per locus. Per esempio, se il numero di alleli per 10 loci è 2, 3, 2, 1, 1, 1, 1, 1, 1, 1, allora *A*= (2+3+2+1+1+1+1+1+1+1)/10 = 1.4

Quand’è che una popolazione può dirsi variabile?



A



B

Quando una popolazione può dirsi variabile?

- Quando molti siti del DNA sono variabili

  diversità nucleotidica:

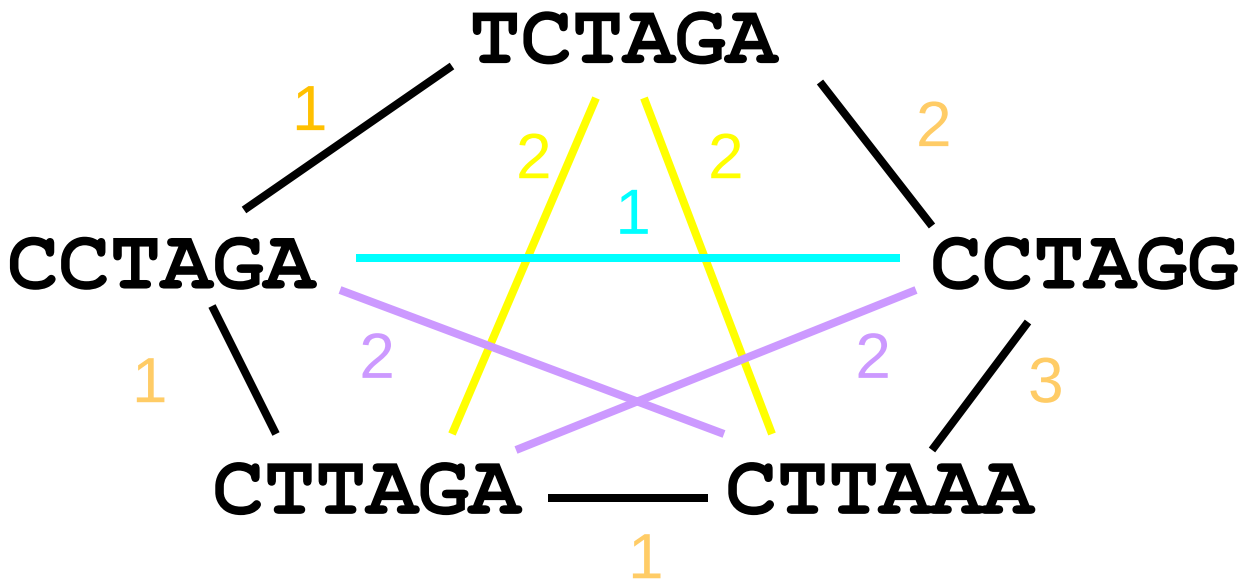
  $$\pi = N \text{ siti polimorfici} / N \text{ totale siti}$$

- Quando ci sono grandi differenze molecolari fra i suoi membri

  mismatch medio:

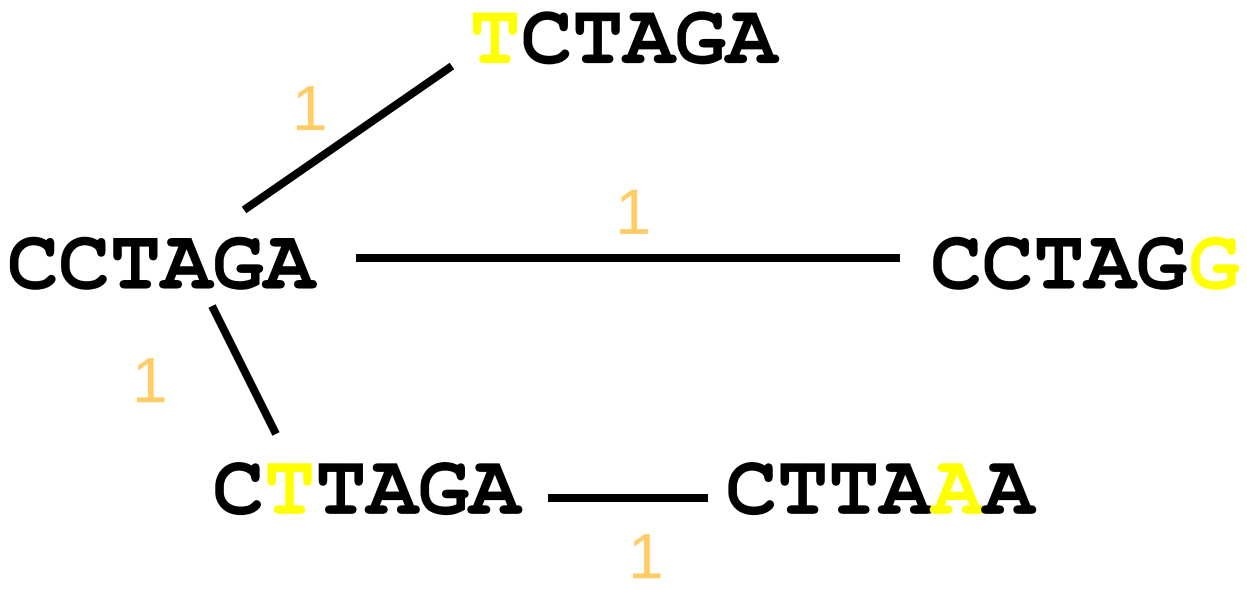
  $$k = \Sigma d_{ij} / [N (N-1) / 2]$$

Il mismatch è il numero di sostituzioni fra coppie di individui



$\Sigma d_{ij} = 17$ $k = 1.7$

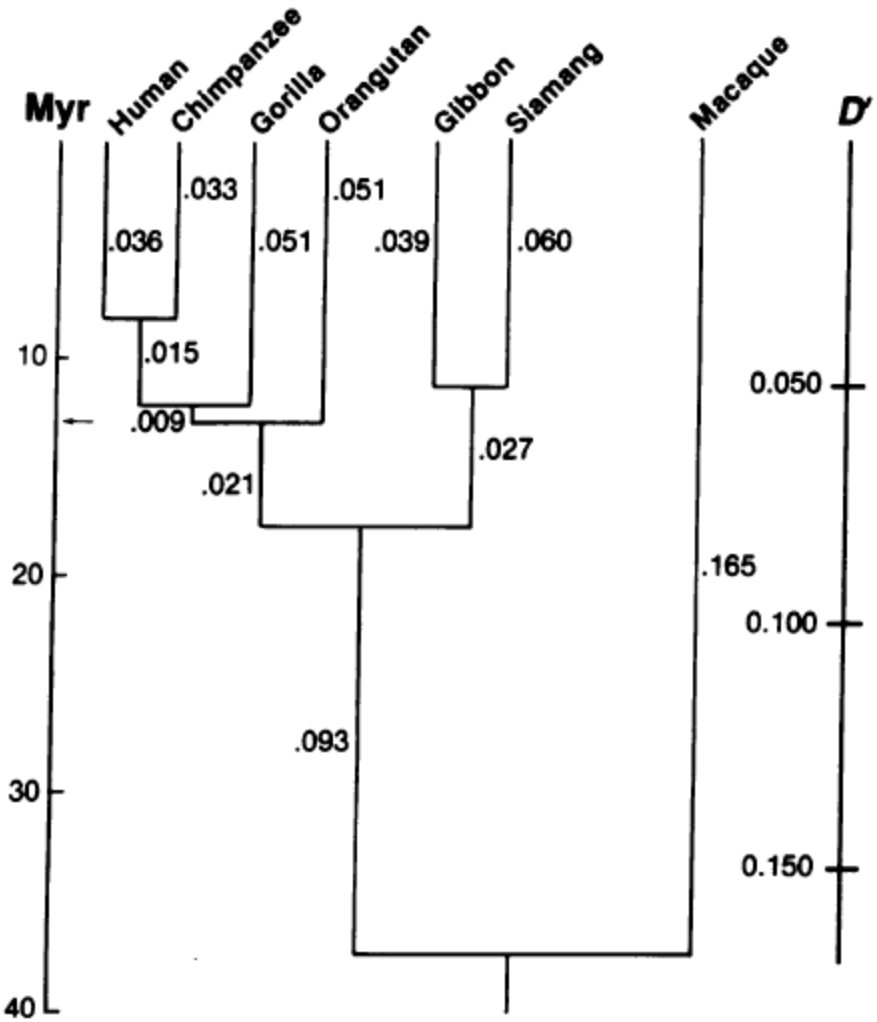
(Ricostruzioni parsimoniose)

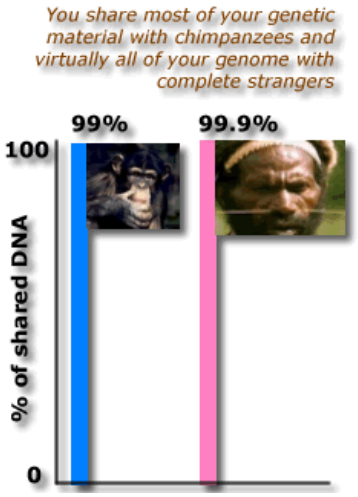


$\Sigma d_{ij} = 17$ $k = 1.7$

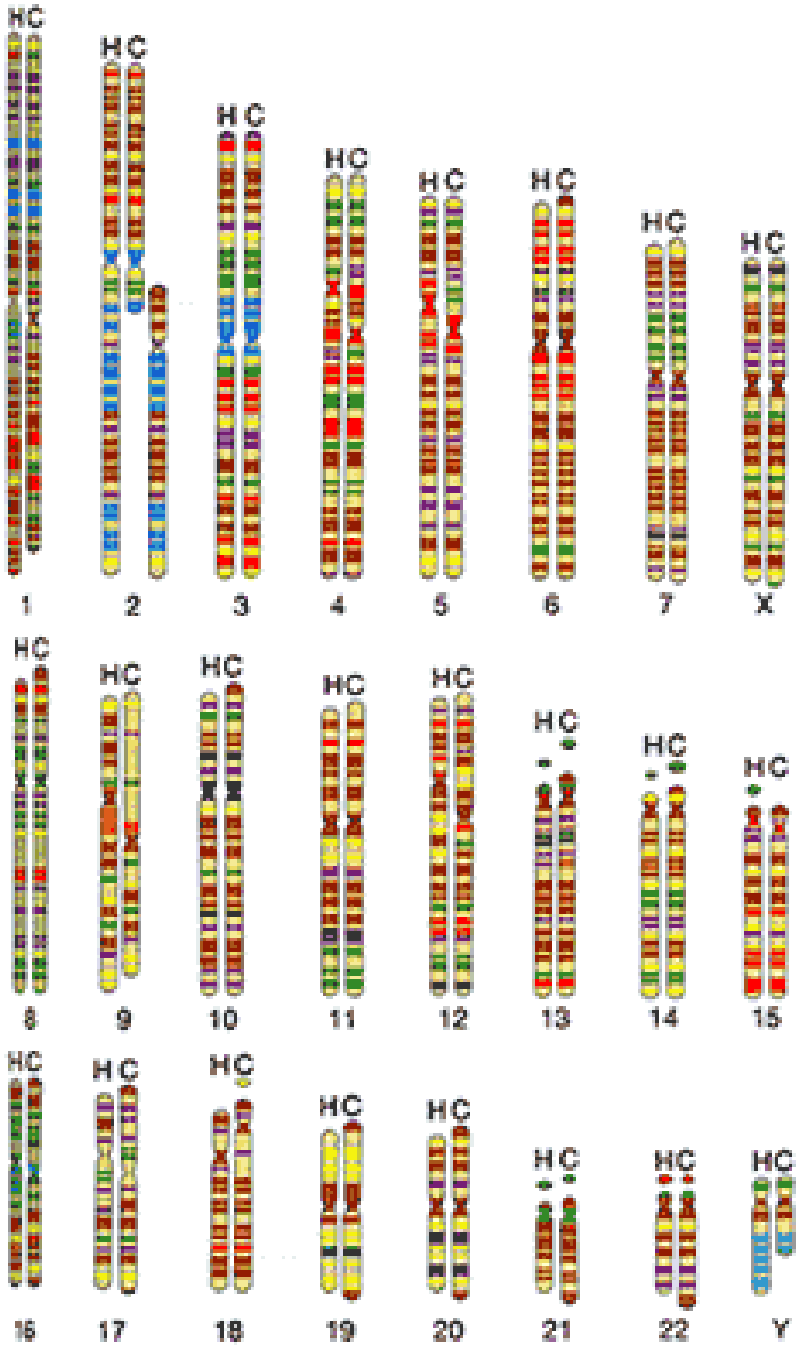
Quanto polimorfismo c'è nel DNA della nostra specie?

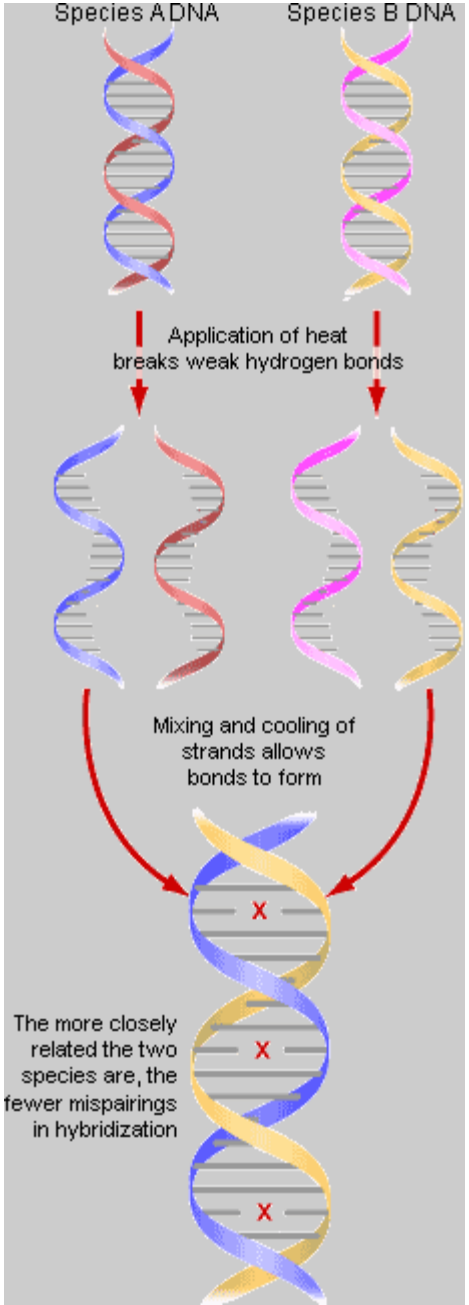






Goldman et al. (1987) PNAS

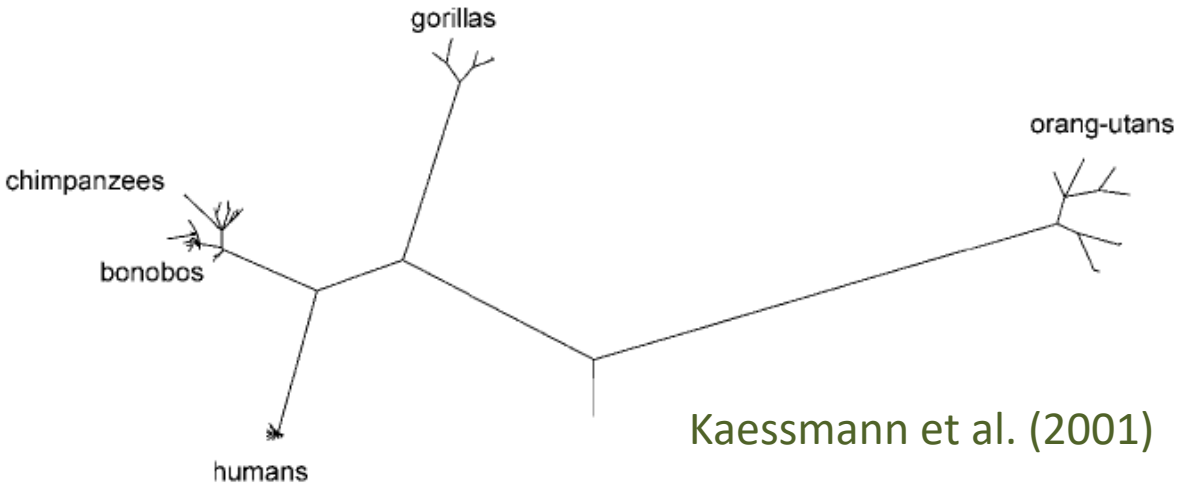




Livelli di diversità fra uomo e scimpanzè

Tempi di dissociazione indicano una differenza pari all'1.76% (Sibley & Ahlquist, 1984)

Livelli di polimorfismo nell'uomo: SNP



Kaessmann et al. (2001)

Fig. 1 Phylogenetic tree[10] of human (n=70; ref. 2), chimpanzee (n=30), bonobo (n=5; ref. 1), gorilla (n=11) and orang-utan (n=14) Xq13.3 sequences. Gorilla DNA samples (n=11) were obtained from zoos and primate research institutes. Orang-utan samples were from skin fibroblast cell lines (n=9; collected using remote biopsy darts) of wild orang-utans from both Borneo and Sumatra[4] as well as from zoos and primate research institutes (n=5). We performed PCR amplification and sequencing as described[1]. We used a gibbon sequence as an out-group. The maximum likelihood tree reconstruction was performed with PUZZLE 4.0 (ref. 10) assuming a Tamura-Nei model with γ-distributed rates.

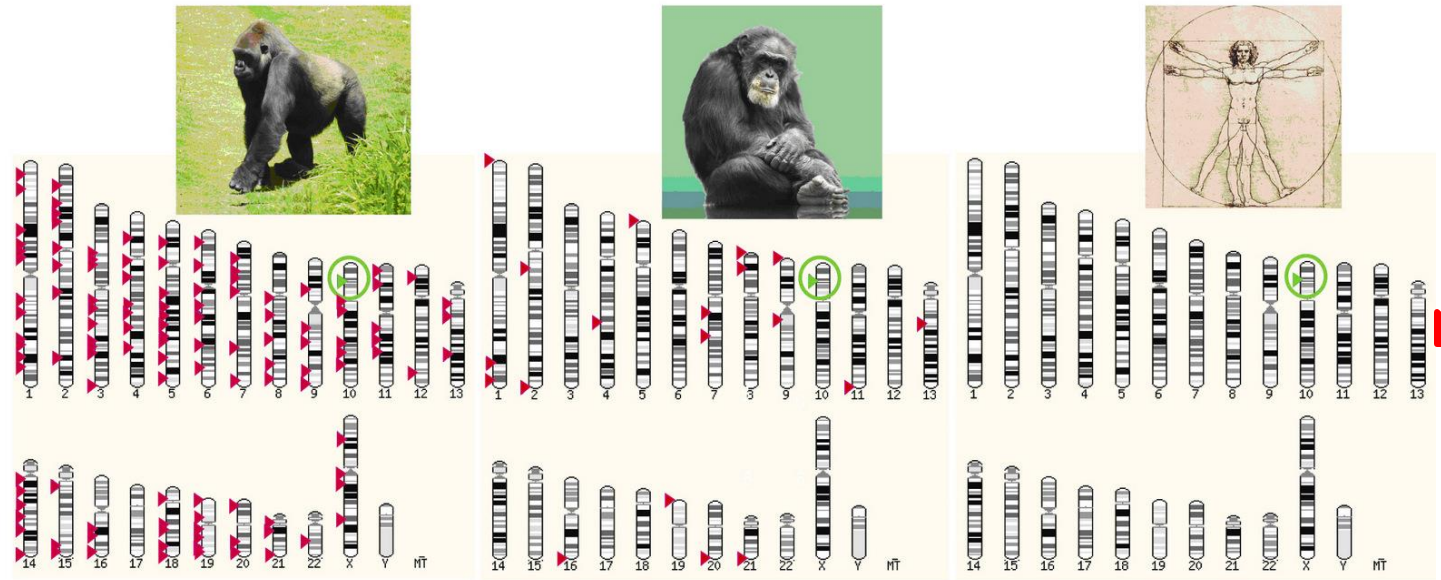


35 milioni di SNP fra uomo e scimpanzè. Su un totale di 3 miliardi di basi (genoma aploide) → 1.23% (Chimpanzee Genome Sequencing Consortium, 2005)

Di questi SNP, 1.06% fissati fra specie
0.17% polimorfici nell'uomo

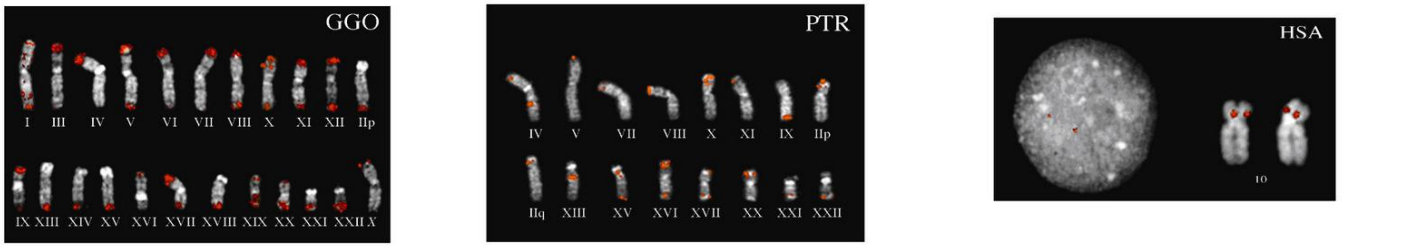
Livelli di polimorfismo nell'uomo: Indel

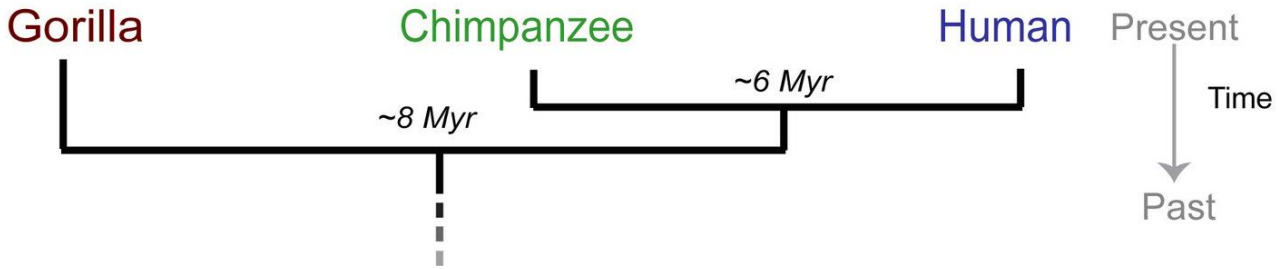
Marquès-Bonet et al. (2009)



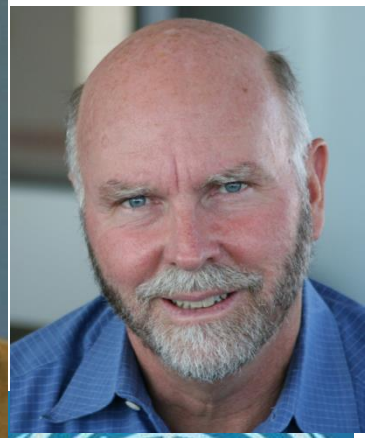
Origine della duplicazione

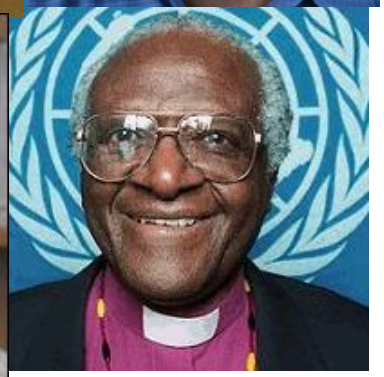
Siti di inserzione





Livelli di polimorfismo nell'uomo: genomi completi



Nel 2010 erano 14 i genomi interamente sequenziati (6 miliardi di paia di basi)

Genome sequenced (publication year)	HGP (2003)	Venter (2007)	Watson (2008)
Time taken (start to finish)	13 years	4 years	4.5 months
Number of scientists listed as authors	> 2,800	31	27
Cost of sequencing (start to fi nish)	$2.7 billion	$100 million	< $1.5 million
Coverage	8–10 ×	7.5 ×	7.4 ×
Number of institutes involved	16	5	2
Number of countries involved	6	3	1

Nel 2010 erano 14 i genomi interamente sequenziati (6 miliardi di paia di basi)

Table 1 | Number of SNPs in the genome and the sequenced exome-containing regions

	Individual	Genomic SNPs	Novel SNPs	Coding-exon SNPs
Africa del sud	KB1	4,053,781	743,714	22,119
	NB1	1,181,663	181,427	19,593
	MD8	125,848	25,485	17,739
	TK1	136,985	30,963	19,226
	ABT SOLiD	3,624,334	412,754	17,342
	ABT exome	121,383	20,294	18,994
Americani bianchi	NA18507	2,639,169	115,843	16,431
	NA19240	3,586,490	216,968	17,268
	J. Watson	2,060,544	98,926	11,868
	J. C. Venter	3,074,574	160,370	15,079
Yoruba	NA12891	2,968,312	35,575	13,375
	NA12892	2,972,120	36,120	13,317
Asiatici	Chinese	3,074,061	84,786	15,759
	Korean (SJK)	3,439,097	130,566	16,637

Novel means that an SNP is not in dbSNP126, other personal genomes (except that it can be shared between our participants), PhenCode, ENCODE resequencing, the Environmental Genome Project, or the 1000 Genomes data set. NA18507 and NA19240 are Yoruban; NA12891 and NA12892 are of Caucasian origin.

Polimorfici almeno lo 0.13% dei siti

Fattori di rischio genetico per malattie

- Malattie causate da una combinazione di multipli fattori genetici e fattori ambientali.
- Numerosi studi di associazione hanno permesso di identificare i fattori di rischio genetico per alcune comuni malattie:
  - Diabete mellito di tipo I → HLA, gene dell'insulina, gene CTLA4;
  - Malattia di Alzheimer → gene dell'APOE;
  - Trombosi venosa profonda → mutazione del fattore V;
  - Morbo di Crohn → mutazione di NOD2, variazione genetica nel cromosoma 5q31.

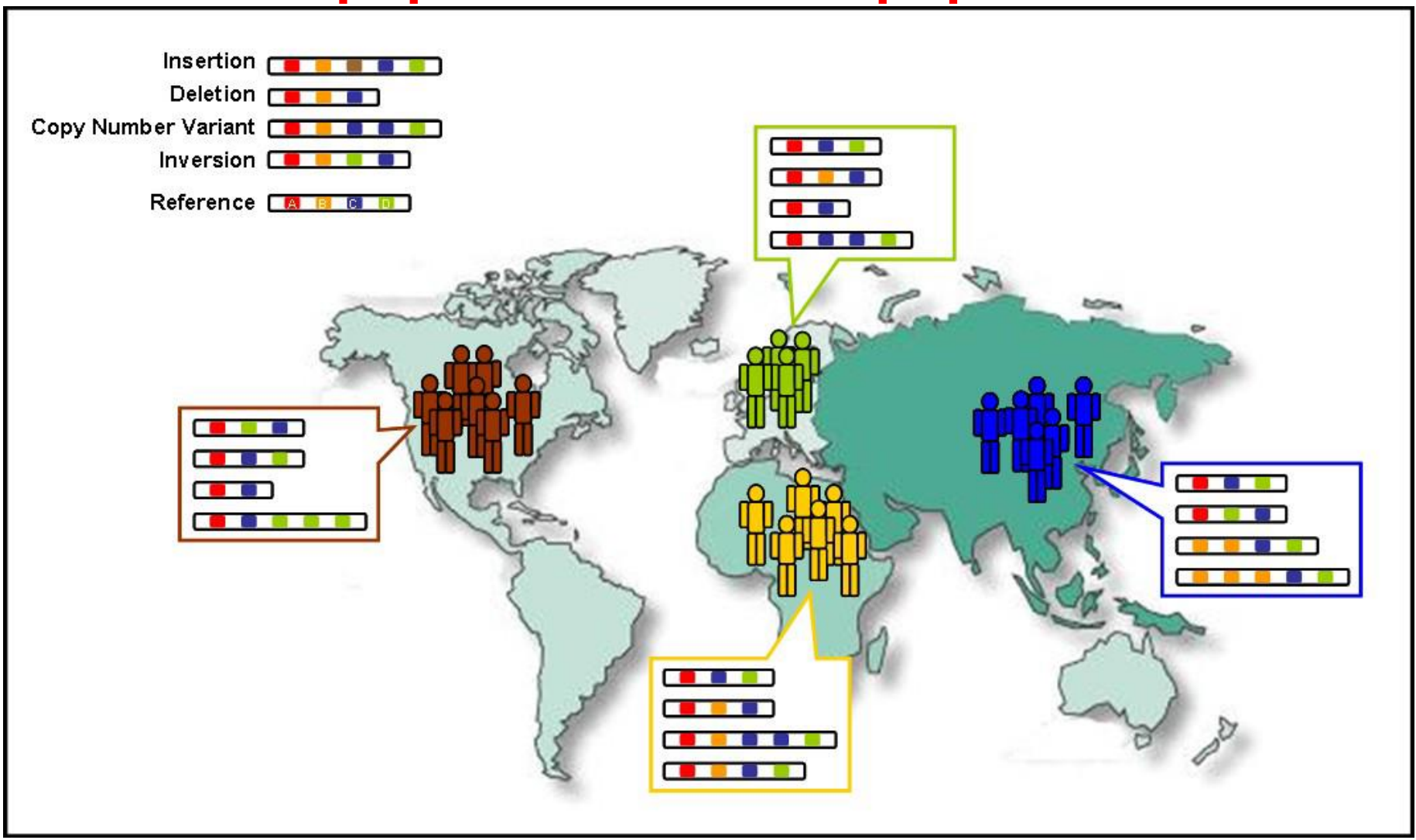
Progetto HAPMAP

- Nel genoma umano ci sono probabilmente tra 10 ed i 30 milioni di SNPs, circa uno ogni 100 – 300 basi.
- Di questi SNP, circa 4 milioni sono SNPs comuni, con entrambi gli alleli di ogni SNP che ha una frequenza sopra il 20%.
- La presenza di un particolare allele di SNP in un individuo viene determinata con il saggio (tipizzazione genetica) di un campione di DNA genomico.
- La coeredità di alleli SNP degli aplotipi produce associazioni tra tali alleli nella popolazione

Progetto HAPMAP

- Lo scopo dell' "International HapMap Project" è stato quello di determinare i "patterns" comuni della variazione di sequenza del DNA nel genoma umano;
- cio è stato conseguito caratterizzando le varianti di sequenza, le loro frequenze e le correlazioni tra esse, in campioni di DNA da popolazioni con antenati da parti dell'Africa, dell'Asia e dell'Europa.
- Lo scopo finale di questo progetto internazionale è quel-lo di sviluppare uno strumento di ricerca che possa aiutare i ricercatori a scoprire i fattori genetici che contribuiscono a:
  - suscettibilità alle malattie,
  - protezione contro le malattie
  - risposta ai farmaci.

Cambiamenti nel numero e nell'ordine dei geni crea la diversità genetica entro le popolazioni e tra le popolazioni.

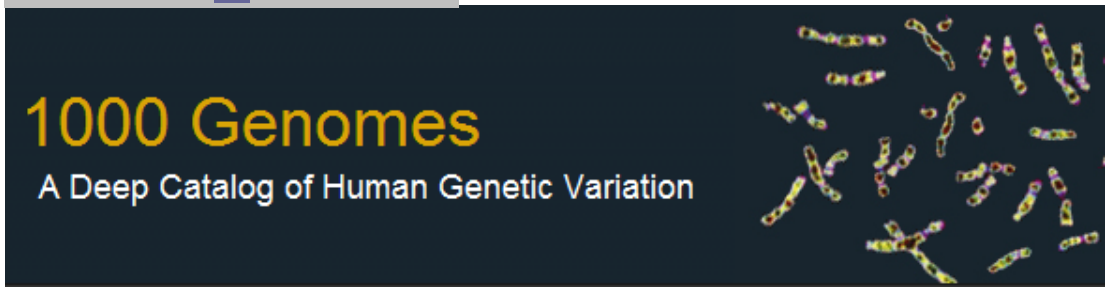


‘’1000 Genomes project”

- Il “1000 Genomes project”, iniziato nel 2008, è una ricerca internazionale che ha lo scopo di creare un catalogo molto dettagliato della variazione genetica umana. Gli scienziati hanno pianificato di sequenziare i genomi di almeno 1000 partecipanti anonimi di diversi gruppi etnici.
- Nel 2010 il progetto ha terminato la fase preliminare, mentre il sequenziamento di 1092 genomi è stato riportato in una pubblicazione su Nature dell’ottobre 2012.
- Il progetto ha coinvolto numerosi gruppi di ricerca interdisciplinari in tutto il mondo (Cina, Italia, Giappone, Kenia, Nigeria, Perù, Gran Bretagna e Stati Uniti).

Nel 2013 i genomi umani interamente sequenziati sono migliaia

Population	Number of Individual DNA Samples
African Caribbean in Barbados [ACB]	79
African Ancestry in SW USA [ASW]	62
British From England and Scotland [GBR]	100
Chinese Dai in Xishuangbanna, China [CDX]	100
Colombian in Medellín, Colombia [CLM]	105
Finnish in Finland [FIN]	100
Han Chinese in Beijing, China [CHB]	120
Han Chinese South [CHS]	150
Iberian Populations in Spain [IBS]	150
Japanese in Tokyo, Japan [JPT]	120
Kinh in Ho Chi Minh City, Vietnam [KHV]	100
Luhya in Webuye, Kenya [LWK]	120
Mexican Ancestry in Los Angeles [MXL]	71
Peruvian in Lima Peru [PEL]	105
Puerto Rican in Puerto Rico [PUR]	105
Toscani in Italia [TSI]	114
Yoruba in Ibadan, Nigeria [YRI]	62



Livelli di polimorfismo nell'uomo: genomi completi

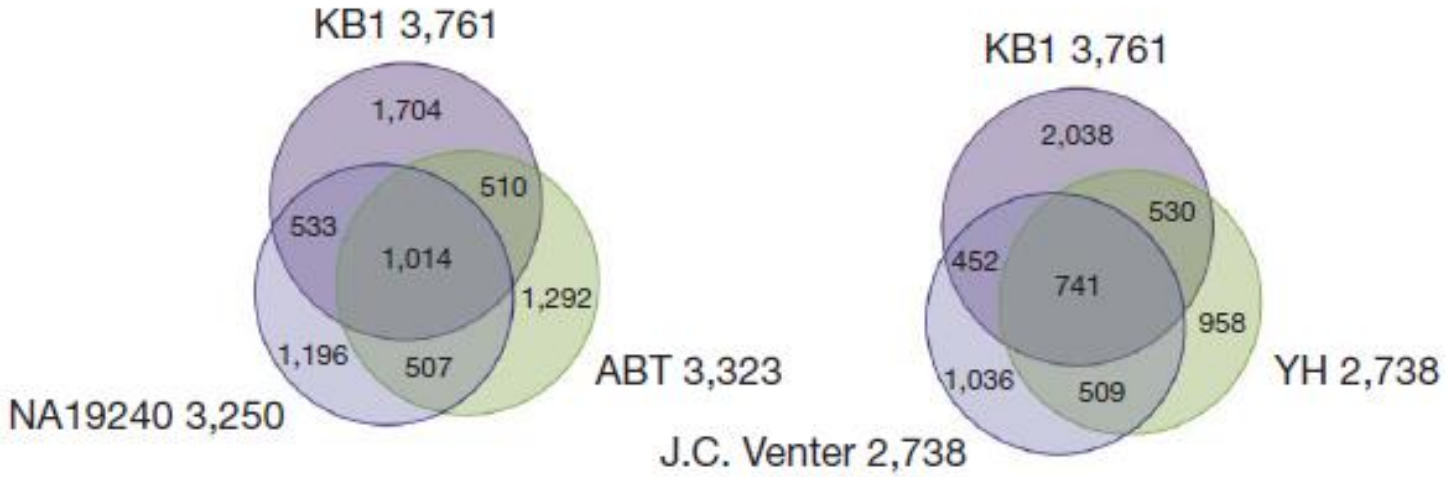


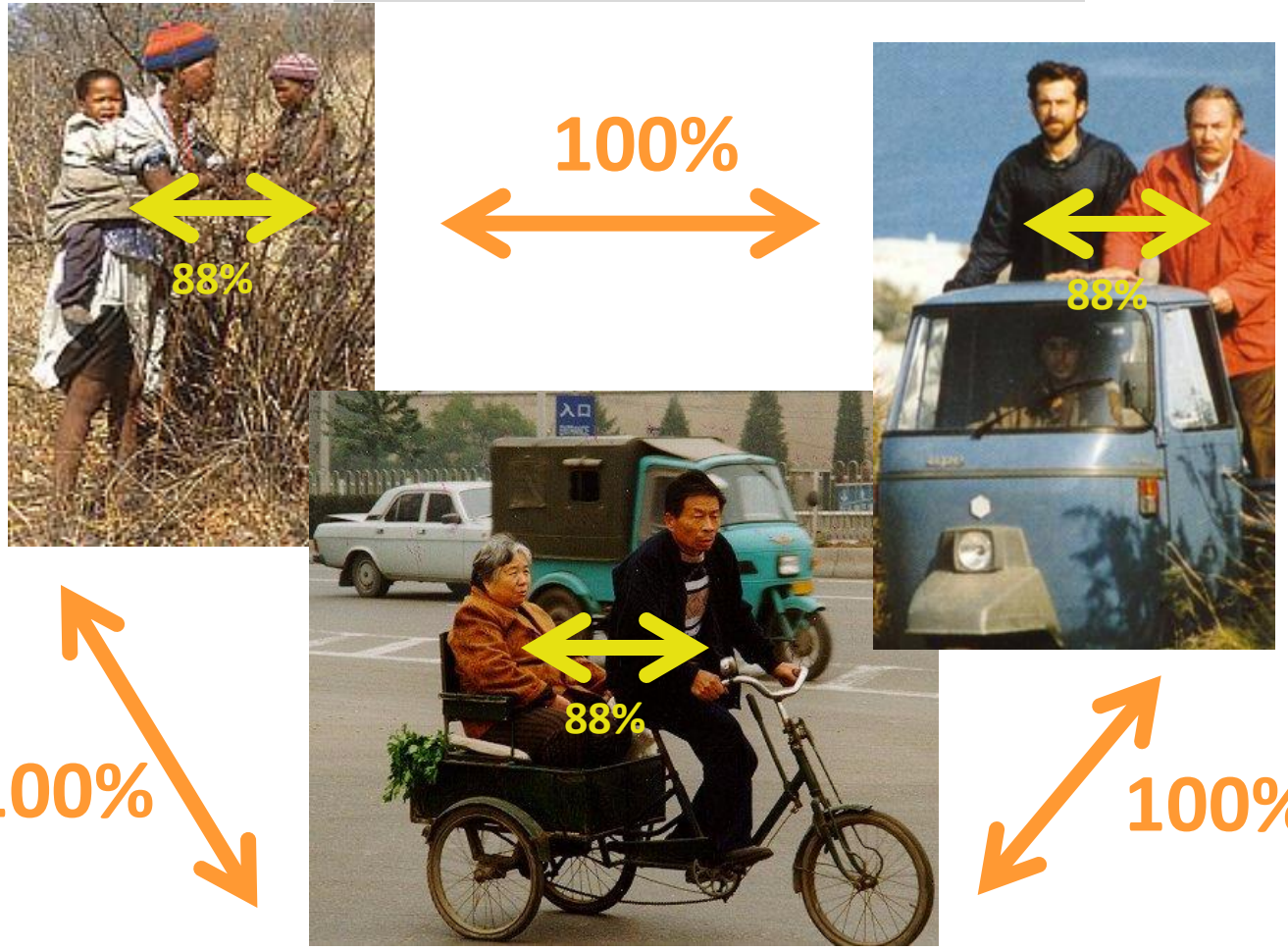
Figure 2 | Three-way relationships among SNPs. SNPs from KB1 are compared with those of the Yoruban NA19240 and ABT (left panel), and with an American of European descent (J. C. Venter) and a Chinese individual (YH) (right panel). Numbers are given in thousands. Variant positions that

Le stime genomiche di F_{ST} per la popolazione umana globale sono ~ 0.12

N of markers	Samples	F_{ST}	Reference
599,356 SNPs	209 individuals from 4 populations: Caucasian, Chinese, Japanese, Yoruba	**0.13**	Weir et al. 2005
1,034,741 SNPs	71 individuals from 4 populations: Caucasian, Chinese, Japanese, Yoruba	**0.10**	Weir et al. 2005
1,007,329 SNPs	269 individuals from 4 populations: Caucasian, Chinese, Japanese, Yoruba	**0.12**	International HapMap Consortium 2005
443,434 SNPs	3845 worldwide distributed individuals	**0.052**	Auton et al. 2009
2,841,354 SNPs	210 individuals from 4 populations: Caucasian, Chinese, Japanese, Yoruba	**0.11**	Barreiro et al. 2008
243,855 SNPs	554 individuals from 27 worldwide populations	**0.123**	Xing et al. 2009
100 Alu insertions	710 individuals from 23 worldwide populations	**0.095**	Watkins et al. 2008
67 CNVs	270 individuals from 4 populations with ancestry in Europe, Africa or Asia	**0.11**	Redon et al. 2006

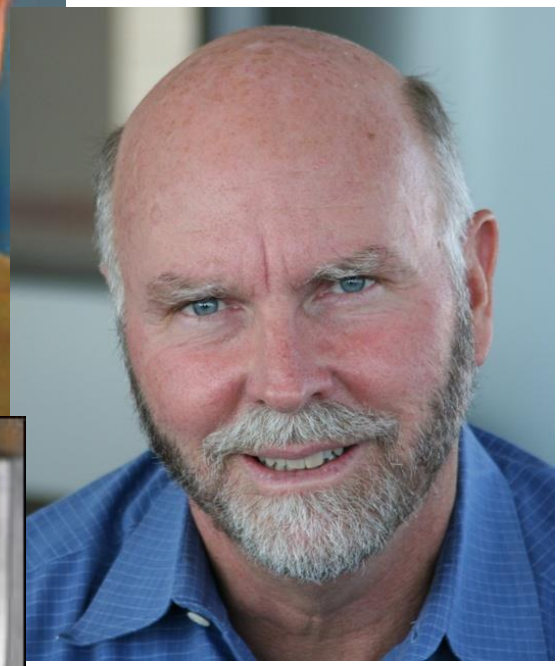
Le popolazioni umane esprimono ~ 12% della varianza massima possibile, date le loro frequenze alleliche

Cosa vuol dire?



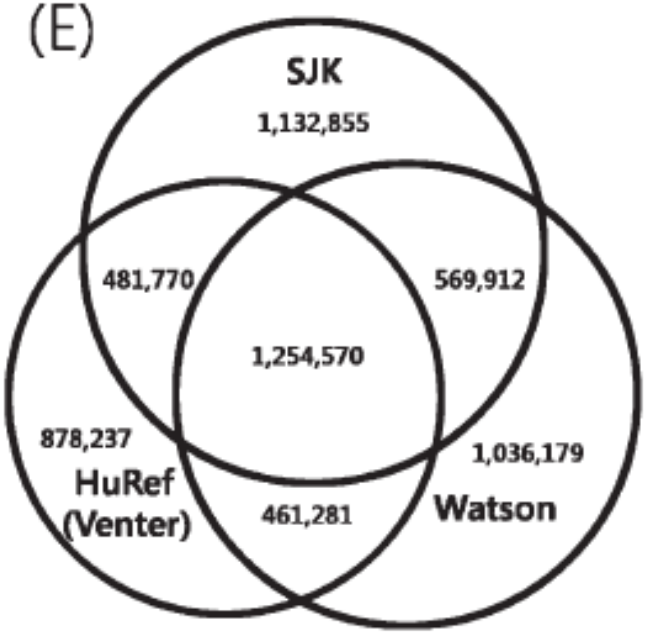
I membri della **nostra comunità sono solo un po' più simili a** noi dei membri di comunità molto lontane



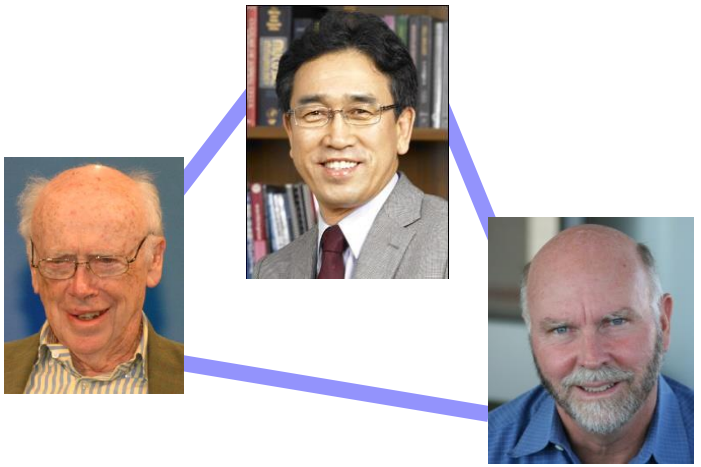




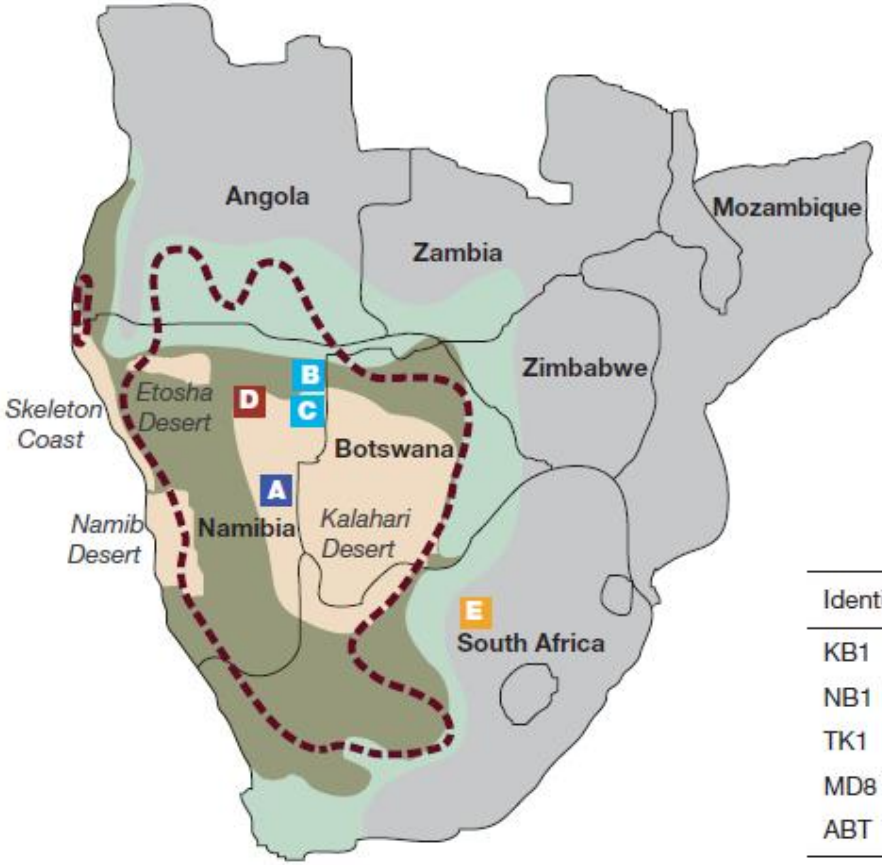
(E)

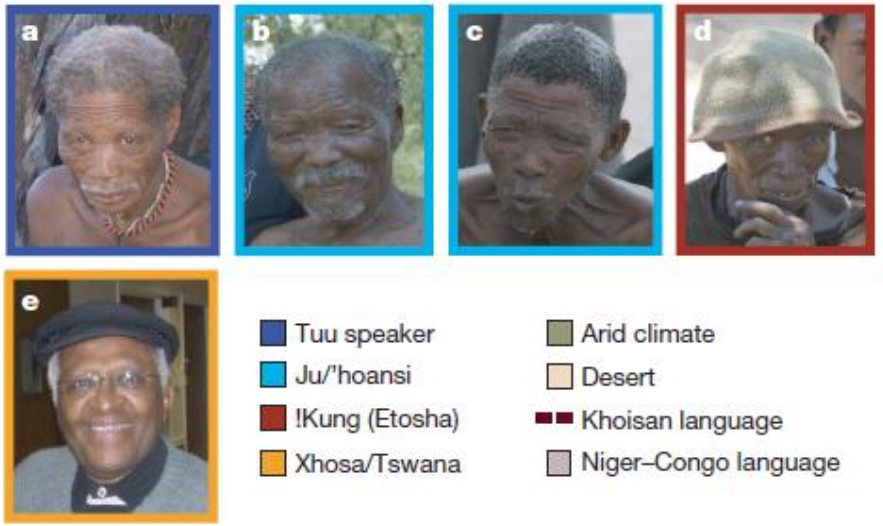


Due persone dello stesso continente possono essere geneticamente più distanti di persone di continenti diversi.



In **media, ci sono differenze più grandi fra due africani che fra europei e asiatici**



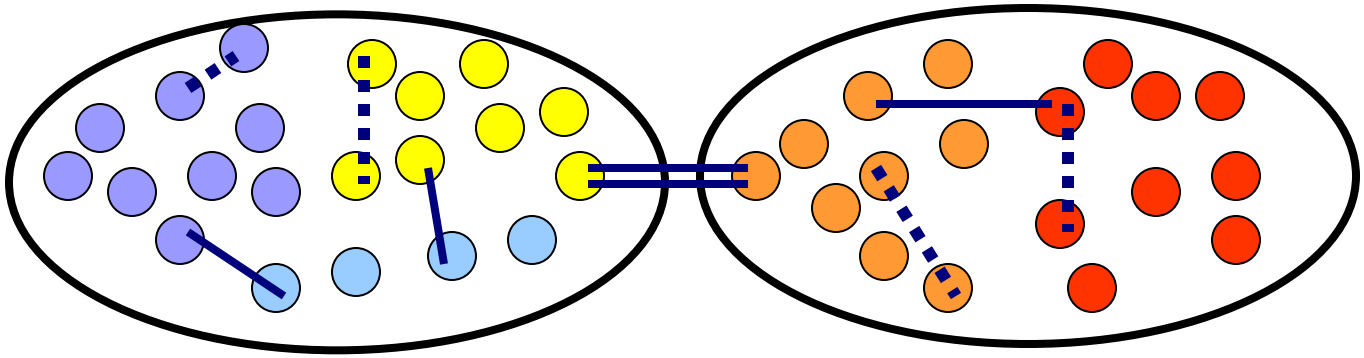


Identifier	Name	Location of origin	Linguistic group	Y chromosome
KB1	!Gubi	Southern Kalahari	Tuu-speaker	B2b
NB1	G/aq'o	Northern Kalahari	Juu (Ju/'hoansi)	A3b1
TK1	D#kgao	Northern Kalahari	Juu (Ju/'hoansi)	A2
MD8	!Aî	Northern Kalahari	Juu (!Kung)	E1b1b1
ABT	Tutu	South Africa	Bantu (Xhosa/Tswana)	E1b1a8a *

In the 117 megabases (Mb) of sequenced exome-containing intervals, the average rate of nucleotide difference between a pair of the Bushmen was 1.2 per kb, compared to an average of 1.0 per kb between a European and Asian individual. Schuster et al. (2010)

Nota bene

- La variabilità interna di una popolazione è solo uno degli aspetti della variabilità genetica:
- Variabilità tra individui della stessa popolazione
- Variabilità tra individui di popolazioni diverse
- Variabilità tra individui di gruppi di popolazioni diverse
- eccetera



Varianze genetiche

- F_{it} = varianza di ciascun individuo rispetto alla media totale della popolazione
- F_{is} = varianza di ciascun individuo rispetto alla media della sua sottopopolazione
- F_{st} = varianza di ciascuna sottopopolazione rispetto alla media totale della popolazione

$$(1 - F_{it}) = (1 - F_{is})(1 - F_{st})$$

Riassunto

- La genetica di popolazioni è un ramo della genetica,
- quindi si **occupa di studiare la variabilità**
- Si può descrivere la variabilità o diversità genetica a partire da:
- fenotipo (morfologia o proteina)
- o genotipo (DNA)
- La diversità del DNA è dovuta a polimorfismi di singolo nucleotide (SNP), alla presenza di elementi ripetuti (STR, VNTR) e alla presenza di delezioni, inserzioni o duplicazioni (CNV)
- Per quantificare la diversità genetica si ricorre a statistiche, fra cui:

(1) numero di alleli; (2) eterozigosi osservata; (3) eterozigosi attesa; (4) polimorfismo nucleotidico; (5) differenze a coppie o mismatch

I GENI NELLE POPOLAZIONI.
Struttura genetica

Struttura genetica di una popolazione

- **Popolazione naturale:** modello è una popolazione panmittica, cioè una comunità di individui a riproduzione incrociata e ad incrocio casuale.

➢ In questo caso l'analisi genetica deve tener conto che:

- non si ha a che fare con singoli individui, ma con gruppi di individui, con una popolazione, che costituisce l'unità evolutiva;
- la popolazione non è omogenea: per ogni locus, sono presenti tutti i possibili genotipi;
- non si hanno schemi di incrocio controllati: avvengono casualmente tutti i tipi di incrocio;
- non si considera una sola generazione, ma molte generazioni successive

Implicazioni dallo studio della genetica di popolazione

- Lo studio della genetica di popolazione ha diverse importanti implicazioni.

- **Consente di studiare i meccanismi base dell'evoluzione:**
  - Questi consistono essenzialmente in sostituzioni alleliche e riassortimento delle combinazioni geniche.

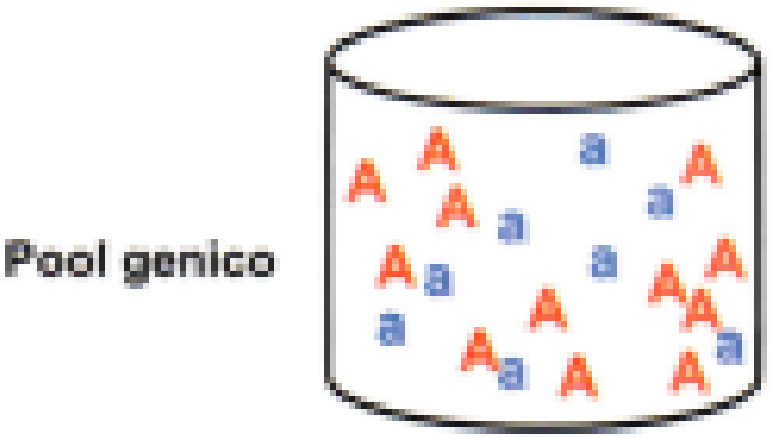
- **L'unità evolutiva è la popolazione**: una comunità di individui appartenenti alla stessa specie, a riproduzione sessuale allogama, che si incrociano, o possono potenzialmente incrociarsi tra loro, in modo casuale.
  - Ciò significa che ciascun membro della popolazione ha la stessa probabilità di incrociarsi con qualsiasi altro individuo di sesso opposto.
  - Una popolazione di questo tipo è detta **panmittica.**

Struttura genetica di una popolazione

- **Popolazione mendeliana in equilibrio: una popolazione teorica, il cui modello presenta le seguenti caratteristiche:**

- La popolazione è molto grande, è costituita da un numero molto elevato di individui.
- Gli incroci tra gli individui che la compongono avvengono completamente a caso (***popolazione panmittica).***
- I genotipi non differiscono tra loro per il tasso di riproduzione e la capacità riproduttiva (***assenza di selezione).***
- Gli alleli sono del tutto stabili (***assenza di mutazione).***
- Non si verifica scambio di individui con altre popolazioni (***assenza di migrazione).***

Pool genico.



Frequenze alleliche: p = frequenza di A = 12/20 = 0.6

q = frequenza di a = 8/20 = 0.4

- **La struttura genetica di una popolazione può essere completamente definita per ogni locus dalle frequenze alleliche al locus stesso.**

la frequenza dei diversi tipi di zigoti dipende esclusivamente dai valori di p *e q, secondo la distribuzione binomiale*

$f(A) = p \qquad f(a) = q \qquad p + q = 1$

Combinazioni	Probabilità
AA	p^2
Aa	pq } 2pq
aA	pq } 2pq
aa	q^2

Conoscendo i valori di ***p*** e di ***q,*** è possibile prevedere esattamente la frequenza dei diversi genotipi nella popolazione**.**

le frequenze alleliche possono essere dedotte dalle frequenze genotipiche:

GENOTIPI	AA	Aa	aa	
f(GENOTIPICHE)	P	H	Q	(P + H + Q) = 1
espresse come f(ALLELICHE)	p^2	2pq	q^2	

Da cui: $p = P + 1/2\ H$ infatti: $p^2 + pq = p^2 + p(1 - p) = p^2 + p - p^2 = p$

$q = Q + 1/2\ H$ infatti: $q^2 + pq = q^2 + q(1 - q) = q^2 + q - q^2 = q$

Legge di Hardy- Weinberg

- in base al rapporto atteso tra le frequenze alleliche e le frequenze genotipiche, in condizioni di equilibrio, è possibile:

1. **calcolare le frequenze alleliche** nella popolazione, sulla base delle frequenze genotipiche;
2. **fare previsioni circa le frequenze genotipic**he, a partire dalle frequenze alleliche

- **Legge di Hardy e Weinberg** è il teorema fondamentale della genetica di popolazione,
  - questo permette di estendere le teorie mendeliane alle popolazioni e di elaborare modelli per eseguire previsioni circa le caratteristiche genetiche delle generazioni successive.
- La Legge afferma che:
  - In una popolazione mendeliana in equilibrio sia le frequenze geniche e sia le frequenze genotipiche restano costanti da una generazione all'altra.
  - Le **frequenze genotipiche** della progenie sono d**eterminate esclusivamente dalle frequenze geniche della generazione parentale**, quindi:
  - la relazione tra frequenze geniche e genotipiche è espressa da $(p+q)^2$ in ogni generazione;
  - questa relazione si **stabilisce dopo una sola generazione di incrocio casuale.**

Teorema di Hardy-Weinberg: associazione casuale (random mating) tra individui

		♂ AA P	♂ Aa H	♂ aa Q
♀	AA P	P^2	PH	PQ
♀	Aa H	PH	H^2	HQ
♀	aa Q	PQ	HQ	Q^2

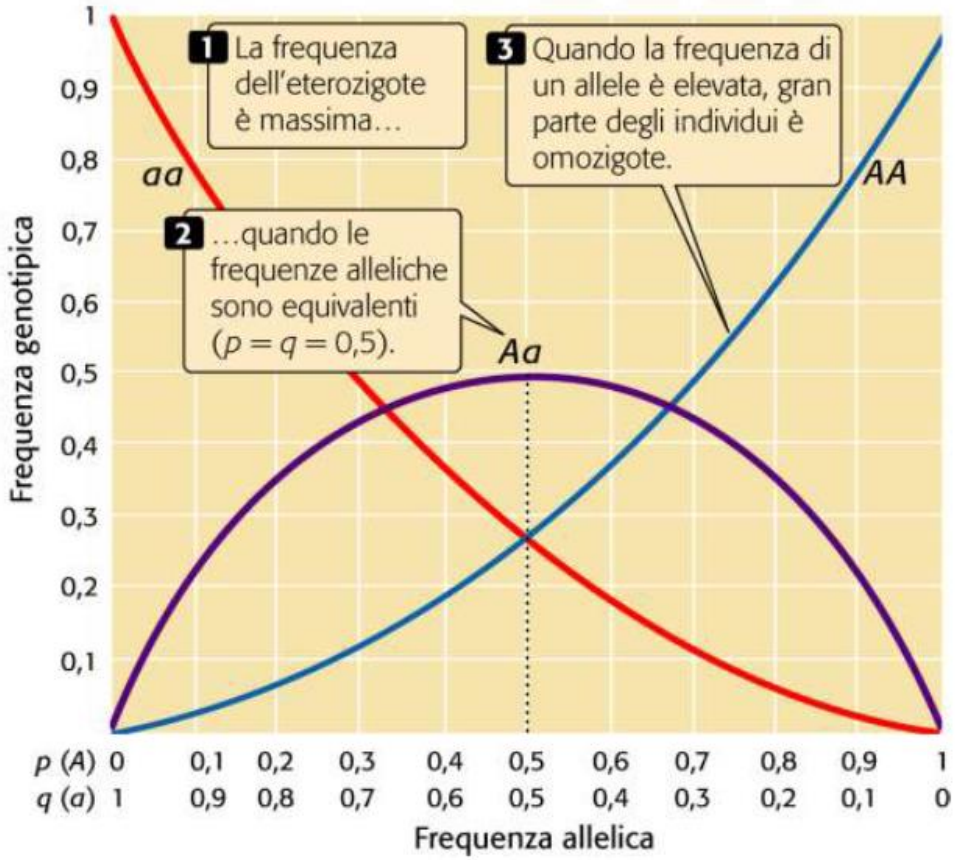
Verifica per un singolo locus con due alleli

Frequenze genotipiche dopo una generazione di incrocio casuale.

INCROCI	P(INCROCIO)	PROGENIE		
		P(AA)	P(Aa)	P(aa)
AA x AA	$p^2 * p^2 = p^4$	p^4		
AA x Aa	$2(p^2 * 2pq) = 4p^3q$	$\frac{1}{2}\, 4p^3q$	$\frac{1}{2}\, 4p^3q$	
AA x aa	$2(p^2 * q^2) = 2p^2q^2$		$2p^2q^2$	
Aa x Aa	$2pq * 2pq = 4p^2q^2$	$\frac{1}{4}\, 4p^2q^2$	$\frac{1}{2}\, 4p^2q^2$	$\frac{1}{4}\, 4p^2q^2$
Aa x aa	$2(2pq * q^2) = 4pq^3$		$\frac{1}{2}\, 4pq^3$	$\frac{1}{2}\, 4pq^3$
aa x aa	$q^2 * q^2 = q^4$			q^4
		$p^4+2p^3q+p^2q^2$ $=p^2(p^2+2pq*+q^2)$ $= p^2$	$2p^3q+4p^2q^2+2pq^3$ $=2pq(p^2+2pq+q^2)$ $= 2pq$	$p^2q^2+2pq^3+q^4$ $=q^2(p^2+2pq+q^2)$ $= q^2$

Relazione tra frequenze geniche e genotipiche

- **Quando una popolazione è in equilibrio di H.W., le proporzioni dei genotipi sono stabilite dalle frequenze alleliche**



Per un locus con 2 alleli, la frequenza dell'eterozigote è più alta delle frequenze degli omozigoti quando le frequenze dei due alleli sono comprese fra 0,33 e 0,66. E' massima (0,5) quando entrambe le frequenze sono = 0,5

Frequenza allele bassa: omozigoti rari **Frequenza allele alta: omozigoti numerosi**

Pierce, GENETICA, Zanichelli editore S.p.A. Copyright © 2005

Stima delle frequenze alleliche di una popolazione

1. geni che mostrano codominanza: locus MN

GENOTIPI:	MM	MN	NN	
	76	92	32	(N = 200)
FREQUENZA	0.38	0.46	0.16	

$$f(M) = \frac{(76 * 2) + 92}{200 * 2} = 0.61$$

ovvero: 0.38 + 1/2 0.46, cioè la proporzione degli omozigoti più metà di quella degli eterozigoti.

Allo stesso modo:

$$f(N) = \frac{(32 * 2) + 92}{200 * 2} = 0.39$$

oppure: 0.16 + 1/2 0.46, o anche, essendo p + q = 1, 1 − 0.61.

Stima delle frequenze alleliche di una popolazione

2. per un gene che mostra DOMINANZA

GUSTO DELLA FENILTIOCARBAMMIDE (PTC)





DOMINANTE E **RECESSIVO**: RIFERITO AL COMPORTAMENTO DEGLI **ALLELI**, NON ALLA FREQUENZA DEL TRATTO NELLA POPOLAZIONE!!!!

- *T (taster):* l'allele dominante conferisce la capacità di percepire il sapore amaro della fenil-tio-carbamide, mentre gli omozigoti recessivi (*tt)* non lo sentono. Si hanno due categorie fenotipiche:

✓ ***Tasters,*** *che includono soggetti TT e Tt,*

✓ ***non-Tasters,*** *costituiti da soggetti di genotipo tt.*

✓ *Si supponga di aver classificato*

- ***140 Tasters e 60 non-Tasters***
- ***campione di 200 individui***

$$\sqrt{60/200} = \sqrt{0.3} = 0.55 = q$$

$$p = 1 - 0.55 = 0.45$$

Stima delle frequenze alleliche di una popolazione

3. un gene legato al sesso.

♀♀: AA Aa aa

♂♂: AY aY

1 Se l'espressione del gene in studio è codominante, il calcolo delle frequenze consisterà semplicemente nel "contare" il numero di alleli dei due tipi nei maschi e nelle femmine:

$$f(A) = \frac{\text{♀♀ AA} \cdot 2 + \text{♀♀ Aa} + \text{♂♂ A}}{\text{n ♀♀} \cdot 2 + \text{n ♂♂}}$$

$$f(a) = \frac{\text{♀♀ aa} \cdot 2 + \text{♀♀ Aa} + \text{♂♂ a}}{\text{n ♀♀} \cdot 2 + \text{n ♂♂}}$$

2 Qualora invece il gene mostri dominanza, è possibile calcolare le frequenze utilizzando solo i maschi. Ad esempio, la frequenza dell'allele *a* sarà stimata come:

$$f(a) = \frac{\text{♂♂ a}}{\text{totale dei ♂♂}}$$

IL PRINCIPIO DI HARDY-WEINBERG

(per geni legati al cromosoma X)

- Poiché il cromosoma Y dell'uomo è essenzialmente privo di geni, per i geni che si trovano nel cromosoma X le frequenze genotipiche sono diverse nei due sessi:
  - Nelle femmine seguono il principio di HW ($p^2+2pq+q^2$)
  - Nei maschi sono uguali alle frequenze alleliche ($p+q$)

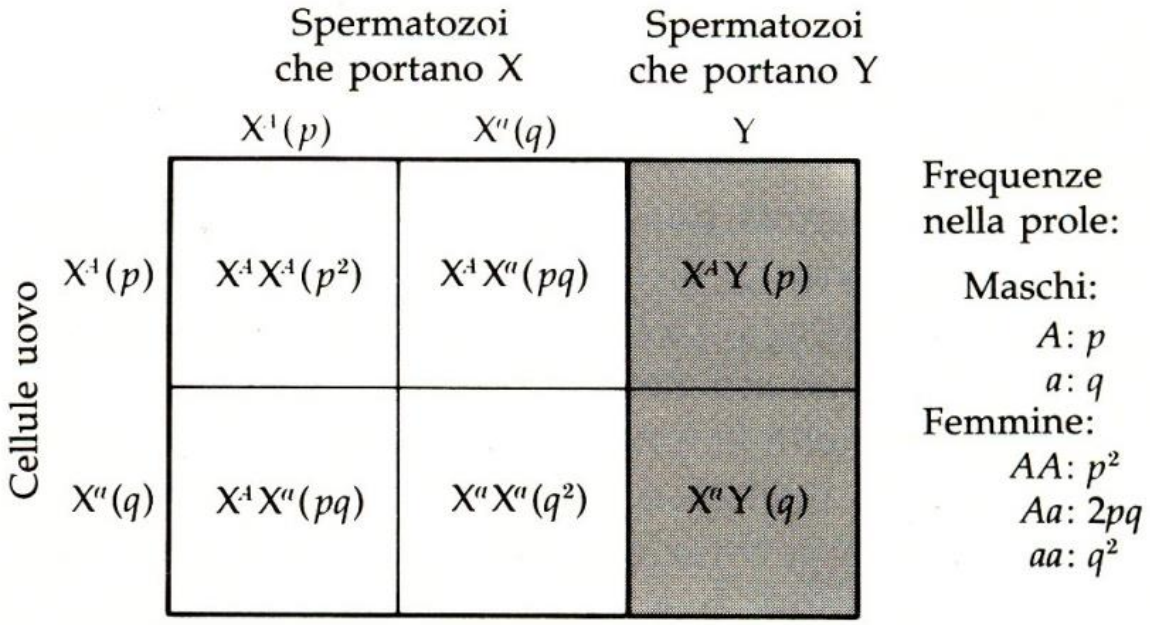


Figura 1.21. Conseguenza dell'accoppiamento casuale per loci legati all' X. Le frequenze genotipiche nelle femmine corrispondono alle frequenze di Hardy-Weinberg; le frequenze genotipiche nei maschi corrispondono alle frequenze alleliche.

Stima delle frequenze alleliche di una popolazione

4. alleli multipli: Es. Sistema AB0

Alleli:	I^A	I^B	i	
Frequenza:	p	q	r	$(p + q + r = 1)$
Fenotipi:	A	B	AB	0
Genotipi:	$\overbrace{p^2\ I^AI^A}$ + $2pr\ I^Ai$	$\overbrace{q^2\ I^BI^B}$ + $2qr\ I^Bi$	$\overbrace{2pq\ I^AI^B}$	$\overbrace{r^2\ ii}$

$$f(A) + f(0) = (p + r)^2$$

e quindi $\sqrt{f(A) + f(0)} = p + r$

Poiché, essendo $p + q + r = 1$, $p + r = 1 - q$, allora $\sqrt{f(A) + f(0)} = 1 - q$

cioè $q = 1 - \sqrt{f(A) + f(0)}$

e anche $r = \sqrt{f(0)}$

Pertanto $p = 1 - q - r$

Si supponga ad esempio di disporre dei seguenti dati:

$A = 56$; $B = 11$; $AB = 8$; $0 = 25$; $N = 100$

$\sqrt{f(A) + f(0)} = p + r = 1 - q$

$\sqrt{0.56 + 0.25} = 1 - q = \sqrt{0.81} = 0.9$

$q = 1 - 0.9 = 0.1$

$r = \sqrt{f(0)} = \sqrt{0.25} = 0.5$

$p = 1 - 0.5 - 0.1 = 0.4$

IL PRINCIPIO DI HARDY-WEINBERG

(per due geni)

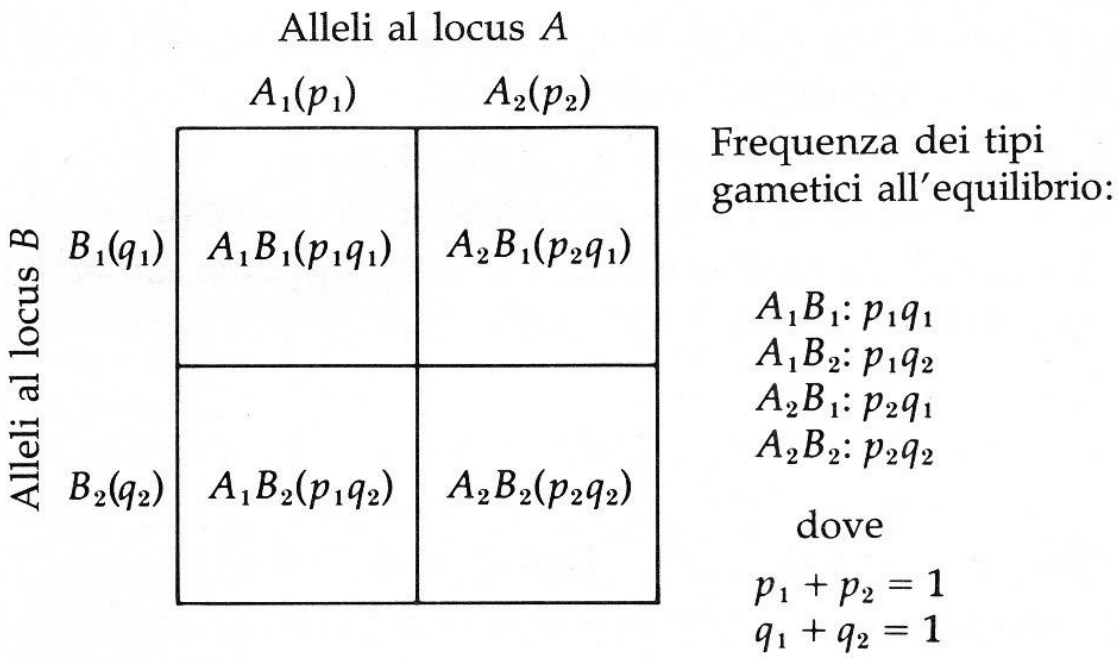


Figura 1.22. Associazione casuale tra gli alleli di due loci, che mostra le frequenze gametiche attese quando i loci sono in equilibrio.

- NB: però la combinazione degli A e B nei quadrati si riferisce ai gameti e non ai genotipi

- Per semplicità è opportuno considerare due loci, ciascuno con due alleli *(A, a; B, b) aventi le frequenze p, q, r e s*

  (p + q = t+ s = 1).

- Lo stato di equilibrio viene raggiunto quando le combinazioni gameti che si formano dall'assortimento casuale degli alleli:

gameti	*A B*	*A b*	*a B*	*a b*
frequenze	*p r*	*p s*	*q r*	*q s*

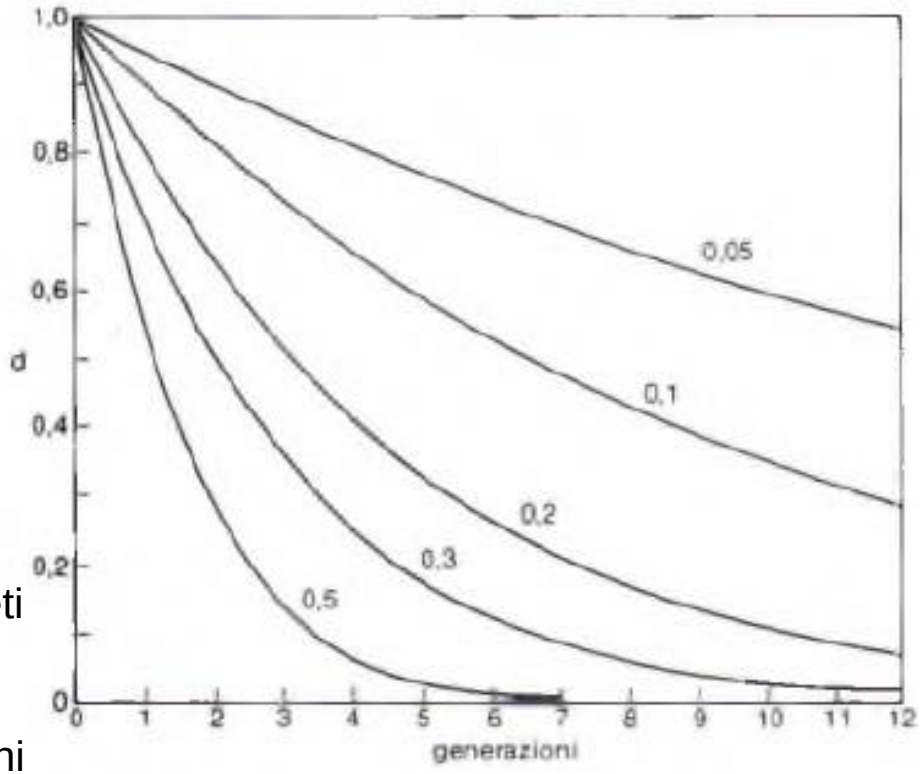
- Si noti che se *p = q = r = s = 1/2*
  si ottengono i classici rapporti genotipici di un diibrido in F2.

- **Se la popolazione non è in equilibrio,** per esempio è composta da
  50% di individui *AABB e* 50% di individui *aabb,*
- dopo una generazione di incrocio casuale (combinazioni casuali tra gameti *AB e gameti ab,* entrambi presenti con frequenza del 50%) la popolazione sarà composta da:
  25% AABB, 50% AaBb e 25% aabb
  pertanto, la successiva popolazione gametica sarà costituita da:
  3/8 AB, 1/8 A b , 1/8 aB, 3/8 ab .
- Dopo un certo numero di generazioni, i rapporti gametici si approssimano a quelli attesi in condizioni di equilibrio.

- Il tempo (numero di generazioni) con cui viene raggiunto l'equilibrio aumenta se i geni sono concatenati, ed è tanto più lungo quanto più stretto è il grado di concatenazione In equilibrio.

- Raggiungimento dell'equilibrio a seguito di *random mating* per le combinazioni genotipiche di due loci, ciascuno con due alleli, in una popolazione costituit inizialmente da eterozigoti in combinazioni *trans.*

- ***Le diverse curve sono riferite a valori diversi di frequenza di ricombinazione; indica la differenza tra le frequenze dell combinazioni in trans e quelle delle co****mbinazioni in cis.*



- In equilibrio, è atteso che le due fasi di concatenazione (gameti in cis AB, ab e gameti in trans Ab e aB) siano ugualmente frequenti (infatti: *pr* x *qs* = *ps* x *qr)* come nel caso di geni non associati.

- Nel caso risultasse che una delle due fasi è più frequente dell'altra (**linkage disequilibrium**): è possibile che la popolazione non abbia ancora raggiunto lo stato di equilibrio o che una combinazione allelica sia più vantaggiosa delle altre.

Dinamica delle popolazioni

- Legge H.W. :

1. popolazione grandissima
2. random mating
3. assenza di forze evolutive quali: mutazioni, selezione migrazione.

- Però le freq. geniche non restano costanti:

✓ Fattori sistematici (mutazioni, selezione migrazione).

✓ Effetto campionamento.

Fattori che possono disturbare l'equilibrio di Hardy-Weinberg

- Vari fattori possono disturbare l'equilibrio di H. W.:

1. Accoppiamento non casuale
2. Mutazione
3. Migrazione
4. Inincrocio
5. Selezione
6. Dimensioni ridotte della popolazione

- Uno o piu' fattori che violano le condizioni di validità della legge di H.W. possono promuovere cambiamenti nelle frequenze alleliche e genotipiche -> EVOLUZIONE

Fattori che cambiano frequenze geniche

- **Mutazione**
- **Selezione**
- **Migrazione**

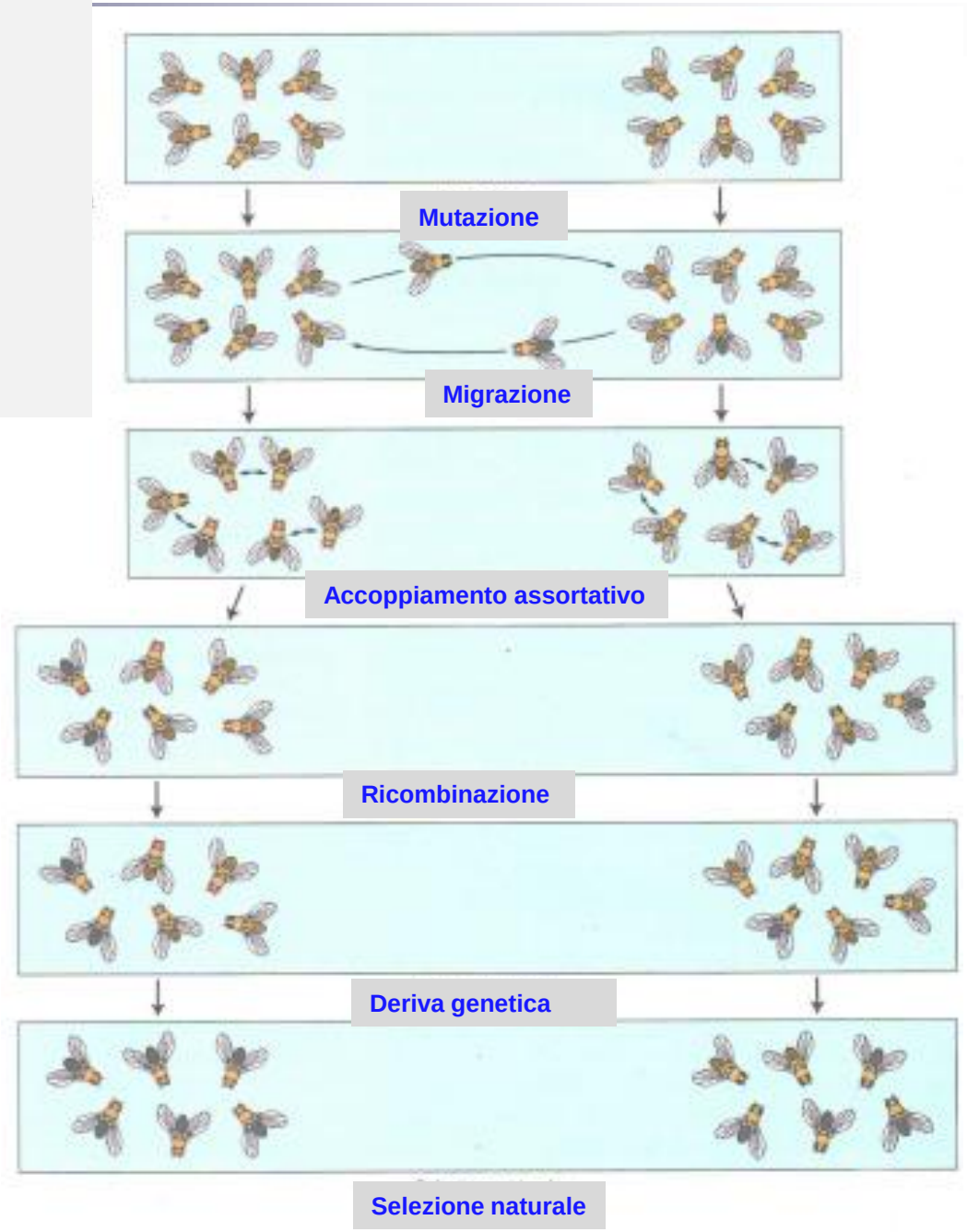
} Introducono variabilità nel pool genico

- **Dimensioni limitate di una popolazione (<1000 ii)**
- **Unioni non casuali (p.es. inincrocio)**

} **Riducono la variabilità del pool genico**

L'*inincrocio* è l'incrocio fra individui strettamente imparentali o consanguinei

Quadro d’insieme dei fenomeni che producono cambiamenti genetici nelle popolazioni.



CAMBIAMENTI NELLE FREQUENZE GENICHE DELLE POPOLAZIONI

1) ***mutazione,*** il processo che porta alla formazione di nuovi alle*li;*
2) ***selezione,*** responsabile della trasmissione differenziale dei vari alleli;
3) ***migrazione*** *di individui tra popolazioni diverse;*
4) ***deriva genetica*** *(genetic drift),* variazioni casuali delle frequenze geniche determinate dal campionamento (dal caso) in piccole popolazioni;
5) ***sistema di riproduzione****,* che nelle sue varie forme può determinare deviazioni dall'incrocio casuale.

1. Mutazione

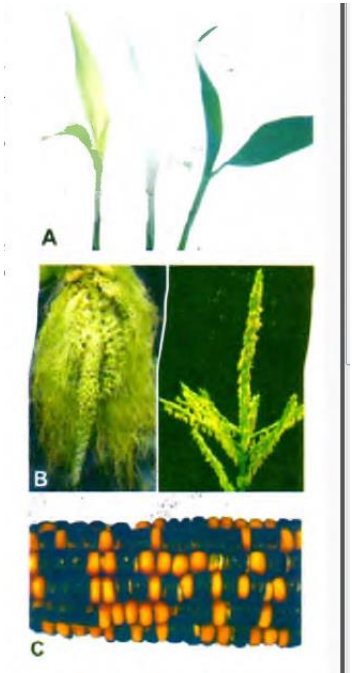
Tassi di mutazione puntiforme in vari organismi

Organismo	Gene	Tasso di mutazione per generazione
Batteriofago	Gamma di ospiti	$2{,}5 \times 10^{-9}$
Escherichia coli	Resistenza a fagi	2×10^{-8}
Zea mais (granturco)	*R* (fattore di colore)	$2{,}9 \times 10^{-4}$
	Y (semi gialli)	2×10^{-6}
Drosophila melanogaster	Letale medio	$2{,}6 \times 10^{-5}$

- **Mutazioni geniche evidenti a livello fenotipico:**
  (A) mutante Clorofilliano I (*luteus),*

(B) mutante del pennacchio *ts1 (tasse/ seed),*

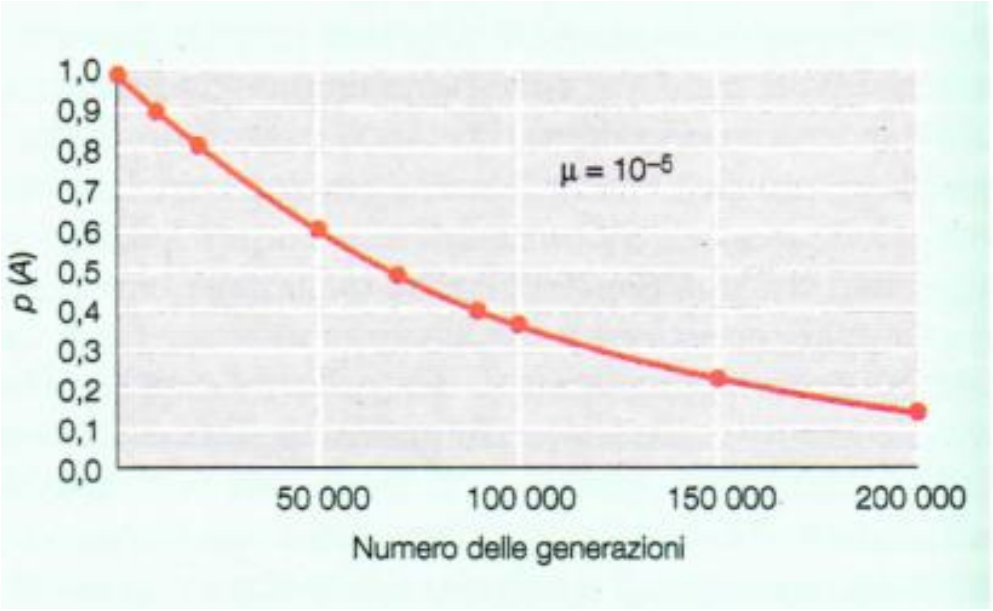
(C) mutante per cariossidi difettive *dek2*
(defective kernels)



Effetto delle mutazioni sulle frequenze alleliche

Allele:	A	a
Frequenza iniziale:	p_0	q_0
Tasso di mutazione:	u	v
	$\rightarrow$	$\leftarrow$

$\Delta q = 0$ quando up = vq.



- Variazione nelle generazioni della frequenza dell'allele *A* dovuta alla mutazione da *A ad a* con tasso di mutazione (u) costante e pari a $10'^{-5}$.

2,Selezione naturale :

motore principale del cambiamento delle frequenze alleliche

- **selezione: l'insieme dei meccanismi responsabili della modificazione della efficienza riproduttiva di un genotipo**:

- è la selezione che determina l'ammontare del contributo che i diversi genotipi nella popolazione daranno alla generazione successiva.

- La Legge di Hardy e Weinberg assume che tutti i genotipi siano egualmente efficienti nella trasmissione degli alleli alla successiva generazione, ma questo, come si è detto, difficilmente si realizza nella realtà, in quanto individui che differiscono in vitalità e fertilità contribuiranno con diversa quantità di progenie alla successiva generazione.

- Non tutti i genotipi hanno lo stesso **valore adattativo (o fitness):** con questo termine, indicato come **W**, si definisce il contributo relativo di un genotipo alla generazione successiva.

- **valore adattativo di un genotipo:** la fitness di quello meglio adattato in un certo ambiente è posta pari ad 1,

- gli altri genotipi hanno una **fitness ridotta rispetto al primo di una frazione s, s:coefficiente di selezione**: misura la riduzione proporzionale del contributo gametico di un particolare genotipo, comparato con il genotipo standard, normalmente il favorito.

- il genotipo meno adattato ha una fitness: W = 1 — s.
- Un genotipo letale avrà W = 0 (s = 1),
- un semiletale W = 0.5 (s = 0.5), e così via.

Calcolo della fitness:

contributo gametico alla generazione successiva

		AA	Aa	aa	Totale
	Prima della selezione	3100	4000	1800	8900
	Dopo una generazione di selezione	3600	4600	1600	9800
1.	*Frequenze alleliche nella 1ª generazione*	$p = \frac{2(3100) + 4000}{2(8900)} = \mathbf{0.57}$		$q = 1 - p = \mathbf{0.43}$	
2.	*Frequenze genotipiche attese*	$N(0.57)^2$ 3184	$N(2)(0.57)(0.43)$ 4804	$N(0.43)^2$ 1812	
3.	*Rapporto tra f_o e f_a*	f_o	f_a		
		$\frac{3600}{3184} = 1.13$	$\frac{4600}{4804} = 0.96$	$\frac{1600}{1812} = 0.88$	
4.	*Fitness relativa W*	$\frac{1.13}{1.13} = \mathbf{1}$	$\frac{0.96}{1.13} = \mathbf{0.85}$	$\frac{0.88}{1.13} = \mathbf{0.78}$	
5.	*Coefficiente di selezione S*	**0**	**0.15**	**0.22**	

Calcolo di *w* e *di s*

	AA	Aa	aa
Prima della selezione (semi)	4100	5000	2200
Dopo la selezione (piante)	3900	4000	1200
Tasso di sopravvivenza	3900 / 4100	4000 / 5000	1200 / 2200
	0.95	**0.80**	**0.55**
Fitness relativa W	0.95 / 0.95	0.80 / 0.95	0.55 / 0.95
	1	**0.84**	**0.58**
Coefficiente di selezione S	1 - 1	1 - 0.84	1 - 0.58
	0	**0.16**	**0.42**

calcolo della fitness: quota di sopravvivenza

Selezione contro un recessivo

Selezione contro un recessivo: variazione delle frequenze

Genotipi	AA	Aa	aa	tot
Frequenze iniziali	p^2	2pq	q^2	1
W = contributo gametico relativo	1	1	1 - s	
Contributo gametico effettivo (genotipi risultanti)	$p^2 * 1$	$2pq * 1$	$q^2(1 - s)$	$1 - q^2 s$

$$q_1 = \frac{q^2(1 - s) + pq}{1 - sq^2} = \dots \frac{q\,(1 - sq)}{1 - sq^2}$$

$$\Delta q = q_1 - q = \frac{q\,(1 - sq)}{1 - sq^2} - q = \frac{sq^2\,(1 - q)}{1 - sq^2}$$

Variazioni della frequenza di un allele recessivo con s = 1

Generazioni

1 p = q = **0.5**

	AA	*Aa*	*aa*	
f genotipiche:	0.25	0.50	0.25	
W	1	1	0	
contributo gametico:	0.25	0.50	0	
	q_1 = 0.25/(0.25+0.5) = **0.33**			Δq = 0.17

2 p = 0.66 q = 0.33

	AA	*Aa*	*aa*	
f genotipiche:	0.43	0.43	0.11	
W	1	1	0	
contributo gametico:	0.43	0.43	0	
	q_2 = 0.22/(0.43+0.43) = **0.25**			Δq = 0.08

3 p = 0.75 q = 0.25

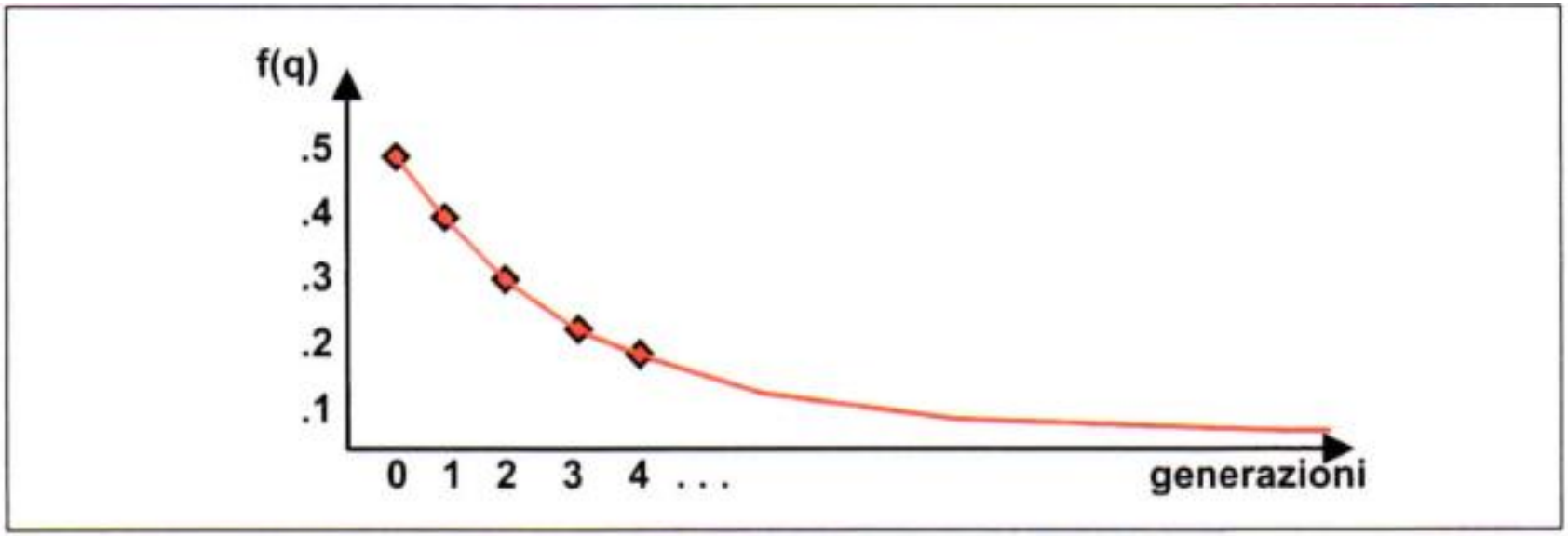
	AA	*Aa*	*aa*	
f genotipiche:	0.56	0.37	0.06	
W	1	1	0	
contributo gametico:	0.56	0.37	0	
	q_3 = 0.18/(0.56+0.37) = **0.20**			Δq = 0.05

4 p = 0.8 q = 0.2

	AA	*Aa*	*aa*	
f genotipiche:	0.64	0.32	0.04	
W	1	1	0	
contributo gametico:	0.64	0.32	0	
	q_4 = 0.16/(0.64+0.32) = **0.17**			Δq = 0.03

...

Efficacia della selezione contro un recessivo



Selezione contro un dominante

- ***Variazione della frequenza di un allele dominante con s = 0.5***

Generazioni p = q = **0.5**

1

	AA	*Aa*	*aa*	
Frequenze genotipiche	0.25	0.5	0.25	
W	0.5	0.5	1	
Contributo gametico	0.125	0.25	0.25	Σ=0.625

p_1 = (0.125 + 0.125)/0.625 = **0.4** Δp = **0.1**

2

	AA	*Aa*	*aa*	
Frequenze genotipiche	0.16	0.48	0.36	
W	0.5	0.5	1	
Contributo gametico	0.08	0.24	0.36	Σ=0.68

p_2 = (0.08+0.12)/0.68 = **0.29** Δp = **0.1**

3

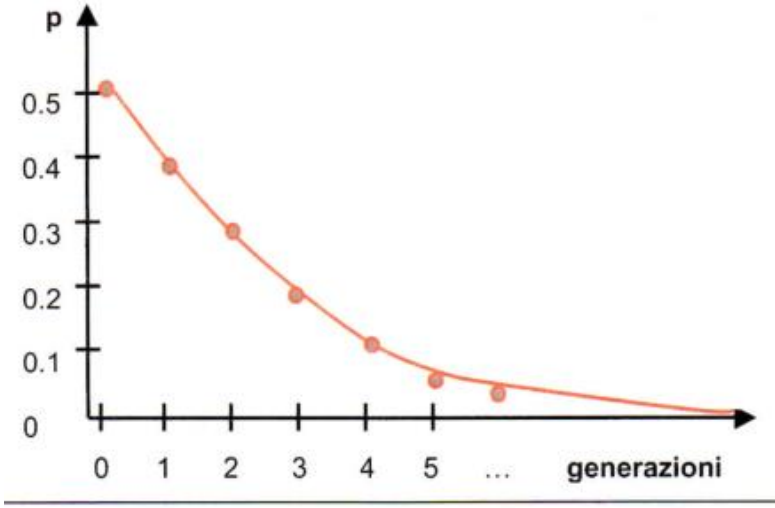
	AA	*Aa*	*aa*	
Frequenze genotipiche	0.09	.42	0.49	
W	0.5	0.5	1	
Contributo gametico	0.045	0.21	0.49	Σ=0.745

p_3 = (0.045+0.10)/0.745 = **0.19** Δp = **0.1**

4

	AA	*Aa*	*aa*	
Frequenze genotipiche	0.04	0.32	0.64	
W	0.5	0.5	1	
Contributo gametico	0.02	0.16	0.64	Σ=0.82

p_4 = (0.02+0.08)/0.82 = **0.012** Δp = **0.1**



Efficacia della selezione contro

un dominante

Selezione contro o a favore dell'eterozigote

	AA	Aa	aa		
				p = q = 0.5	
Frequenze genotipiche	0.25	0.5	0.25		
W	1	0	1		
Contributo gametico	0.25	0	0.25	Totale = 0.5	

$p_1 = 0.25/0.5 = \mathbf{0.5}$ $q_1 = \mathbf{0.5}$

	AA	Aa	aa		
				p = 0.8	**q = 0.2**
Frequenze genotipiche	0.64	0.32	0.04		
W	1	0	1		
Contributo gametico	0.64	0	0.04	Totale = 0.68	

$p_1 = 0.64/0.68 = \mathbf{0.9}$ $q_1 = \mathbf{0.1}$

- **Effetto della selezione contro l'eterozigote in relazione alle frequenze alleliche**

Selezione a favore dell'eterozigote

- **ha una fitness maggiore di entrambi gli omozigoti→**
  porta ad un *polimorfismo bilanciato,* dove per "polimorfismo" si intende la presenza di due (o più) alleli di un gene a frequenza apprezzabile.

Genotipi	AA	Aa	aa
Frequenze iniziali	p^2	$2pq$	q^2
W	$1 - s_1$	1	$1 - s_2$
Contributo gametico	$p^2(1 - s_1)$	$2pq$	$q^2(1 - s_2)$

$$\Sigma = p^2 - p^2 s_1 + 2pq + q^2 - q^2 s_2$$
$$= 1 - p^2 s_1 - q^2 s_2$$

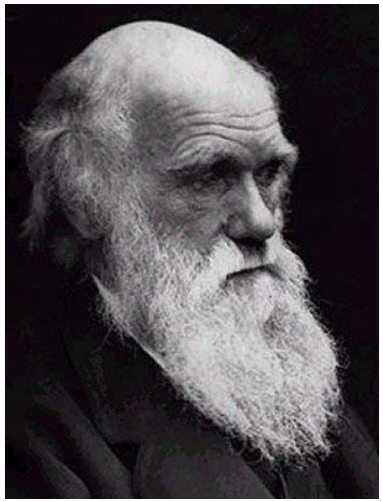
Selezione a favore dell'eterozigote

La popolazione tende ad una situazione di equilibrio, in cui entrambi gli alleli vengono mantenuti nel pool genico, equilibrio che si verificherà quando la proporzione di alleli *A* persi ($p^2 s_1/p$) eguaglierà la proporzione di alleli *a* persi ($q^2 s_2/q$), per ogni generazione: si avrà cioè una situazione stabile ($\Delta q = 0$), quando $ps_1 = qs_2$,

La selezione naturale

- **Selezione naturale**: sopravvivenza e riproduzione differenziale dei genotipi caratterizzati da fitness diversa
- **Fitness darwiniana**: valore che indica la probabilità relativa di sopravvivenza e di riproduzione media di un genotipo in un determinato ambiente

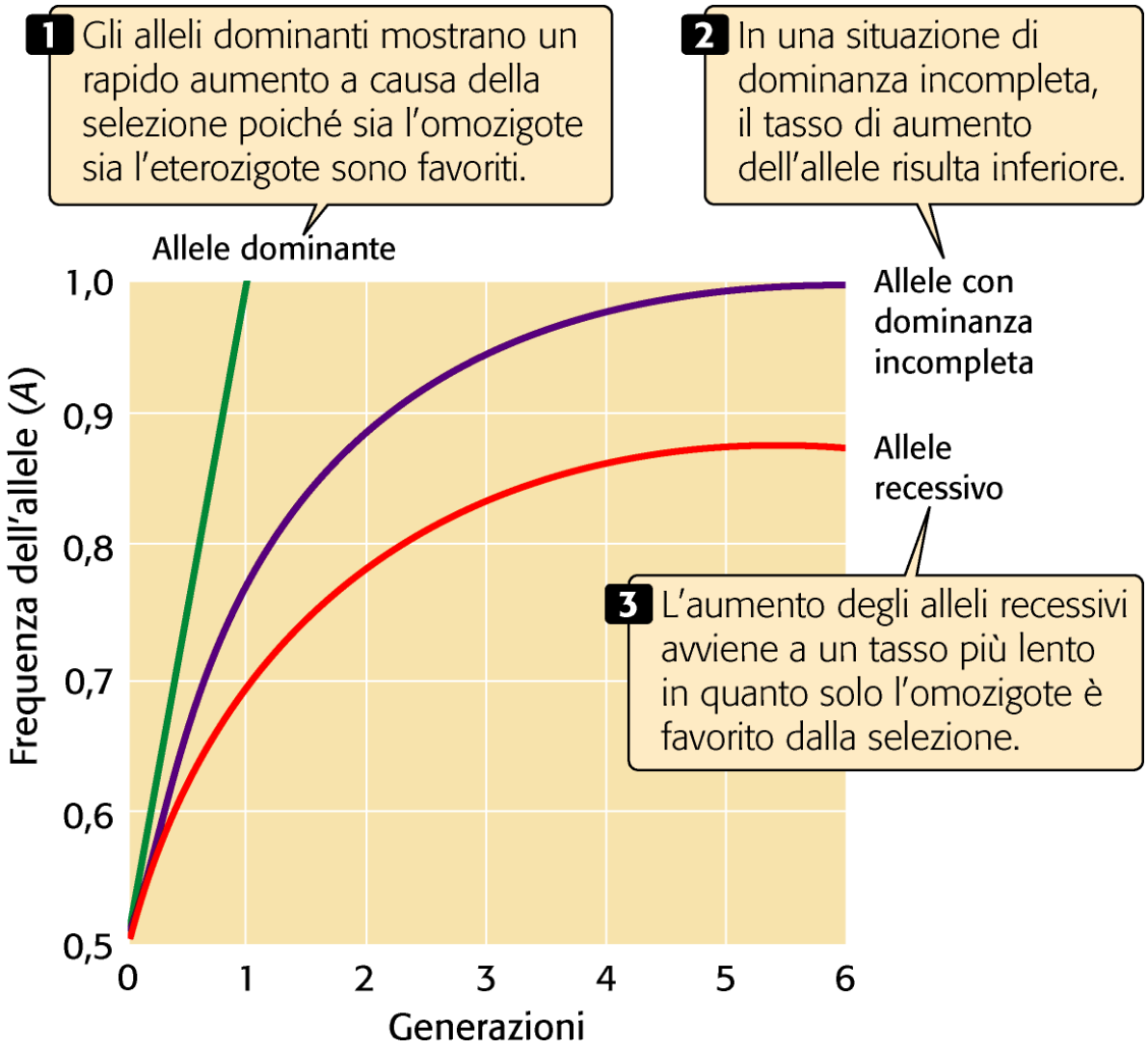
Teoria dell’evoluzione (C. Darwin)



C. Darwin

1. **Principio della variabilità**
2. **Principio dell’ereditarietà**
3. **Principio della selezione**

Effetto della selezione di una mutazione vantaggiosa sulla frequenza allelica



Esempio di selezione naturale e ruolo dell'ambiente

Figura 22.19

***Biston betularia*, la falena screziata, e la sua forma scura *carbonaria* (a) sul tronco di un albero coperto di licheni nella campagna non inquinata e (b) sul tronco di un albero con corteccia annerita.** La forma scura è chiaramente visibile sul tronco chiaro, mentre la forma chiara risulta ben camuffata. Sull'albero scuro, è la forma scura ad essere ben mimetizzata.

a)



b)



Polimorfismo bilanciato (vantaggio dell'eterozigote)

- Si verifica quando la fitness dell'eterozigote non è intermedia ma **più estrema** rispetto a quelle degli omozigoti.
  - In questo caso la selezione non porta alla scomparsa di un allele
  - ma tende a raggiungere una frequenza allelica **intermedia**

Vantaggio dell'eterozigote

- si è verificato in africa occ. in ambiente malarico.

- gli eterozigoti $\beta^a\beta^s$ hanno *fitness* maggiore degli omozigoti $\beta^a\beta^a$, perché risultano resistenti alla malaria (i loro globuli rossi sono inospitali al protozoo *plasmodium*) e anche degli omozigoti $\beta^s\beta^s$, gravemente colpiti dalla **anemia falciforme**.

Quindi l'allele dannoso viene mantenuto nella popolazione perché porta un vantaggio: **polimorfismo bilanciato**

Polimorfismo bilanciato

In una popolazione, mantenimento di più di un allele in conseguenza della superiorità adattativa degli eterozigoti rispetto agli omozigoti. Rappresenta il risultato della pressione selettiva su una mutazione sfavorevole, che ne permette la persistenza in una popolazione. Per esempio, la persistenza del gene per l'anemia falciforme nelle popolazioni africane grazie alla protezione contro la malaria manifestata dagli eterozigoti.

La diffusione della malaria rende la situazione eterozigote molto favorevole dal punto di vista evolutivo. In questo modo le frequenze dei fenotipi omozigoti ed eterozigoti si mantengono relativamente stabili.

Anemia falciforme e malaria



Distribuzione della malaria



Distribuzione dell'allele HbS

- Un esempio di vantaggio dell'eterozigote è quello che ha causato l'espansione dell'anemia falciforme nelle zone in cui la malaria è endemica

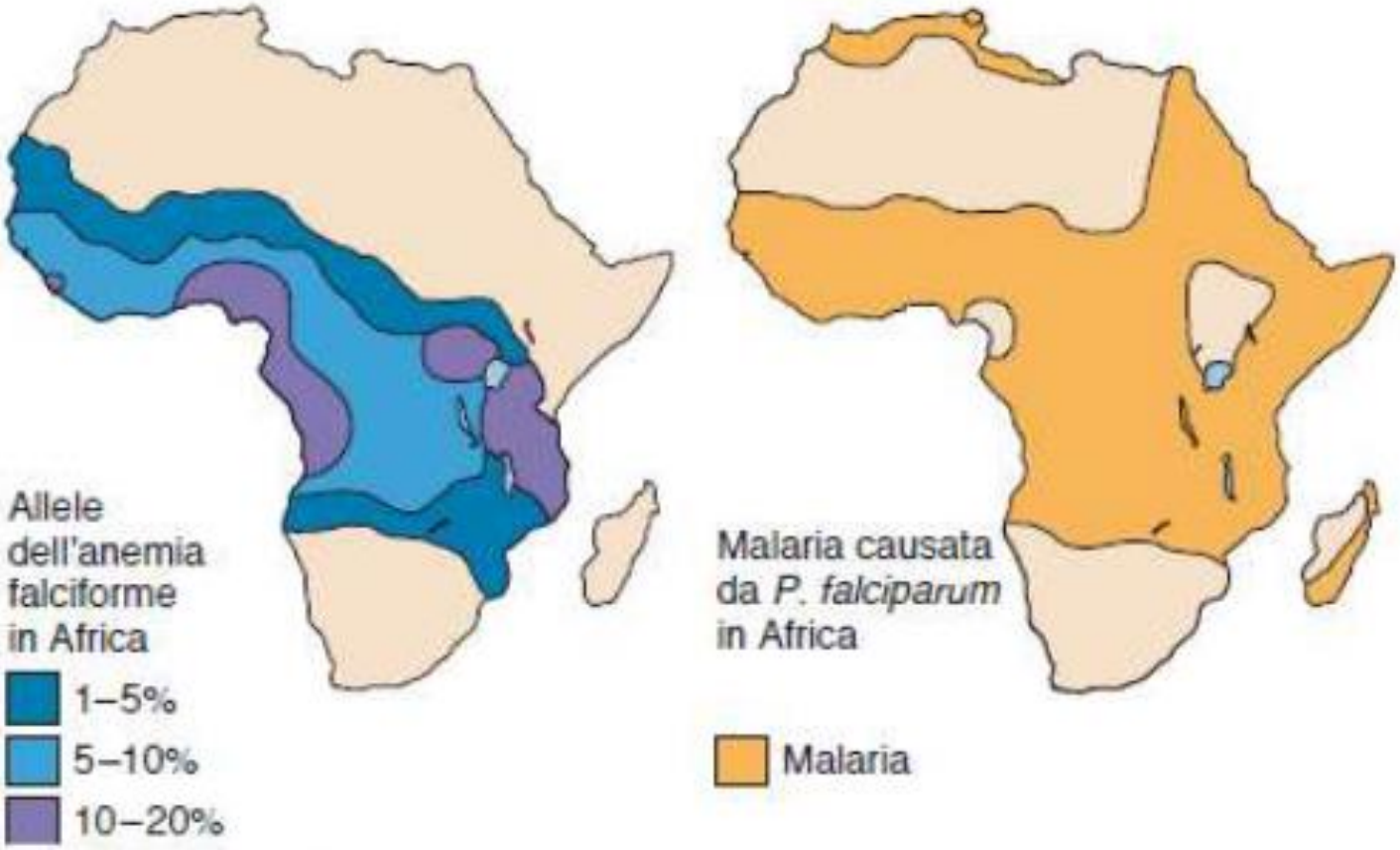


Figura 15.13 **Polimorfismo bilanciato.** Paragonare la distribuzione dei malati di anemia falciforme e di malaria in Africa rivela un polimorfismo bilanciato. I portatori dell'anemia falciforme sono resistenti alla malaria, perché i cambiamenti nel sangue causati dall'allele non sono abbastanza gravi per compromettere la salute, ma abbastanza per inibire il parassita della malaria.

Tabella 15.5 Polimorfismo bilanciato

Malattia 1 (ereditaria, portatore)	Protegge contro →	Malattia 2	Perché →	Meccanismo	Referenze
Anemia falciforme		Malaria		Cellule del sangue anomale non possono sostenere i parassiti	Sezione 12.2
Favismo		Malaria		I parassiti non si possono riprodurre nelle cellule del sangue anormali	Sezione 12.5
Fenilchetonuria		Infezioni fungine nel feto		Elevati livelli di fenilalanina inattivano le tossine fungine	Sezioni 5.2, 14.1, 15.6
Mutazioni prioniche nelle proteine		Encefalopatia spongiforme trasmissibile		La proteina prionica non può cambiare conformazione in presenza di una proteina prionica infettiva	Figura 10.22, Sezione 12.5, Lettura 10.1
Fibrosi cistica		Malattie che causano diarrea (colera, tifo)		Pochi canali del cloro nelle cellule intestinali impediscono la perdita dell'acqua	Sezioni 14.1, 14.3, Lettura 2.2
Sindrome di Smith-Lemli-Opitz		Malattie cardiovascolari		Colosterolo nel siero più basso	MIM 270400 (deformazioni multiple alla nascita, ritardo mentale)

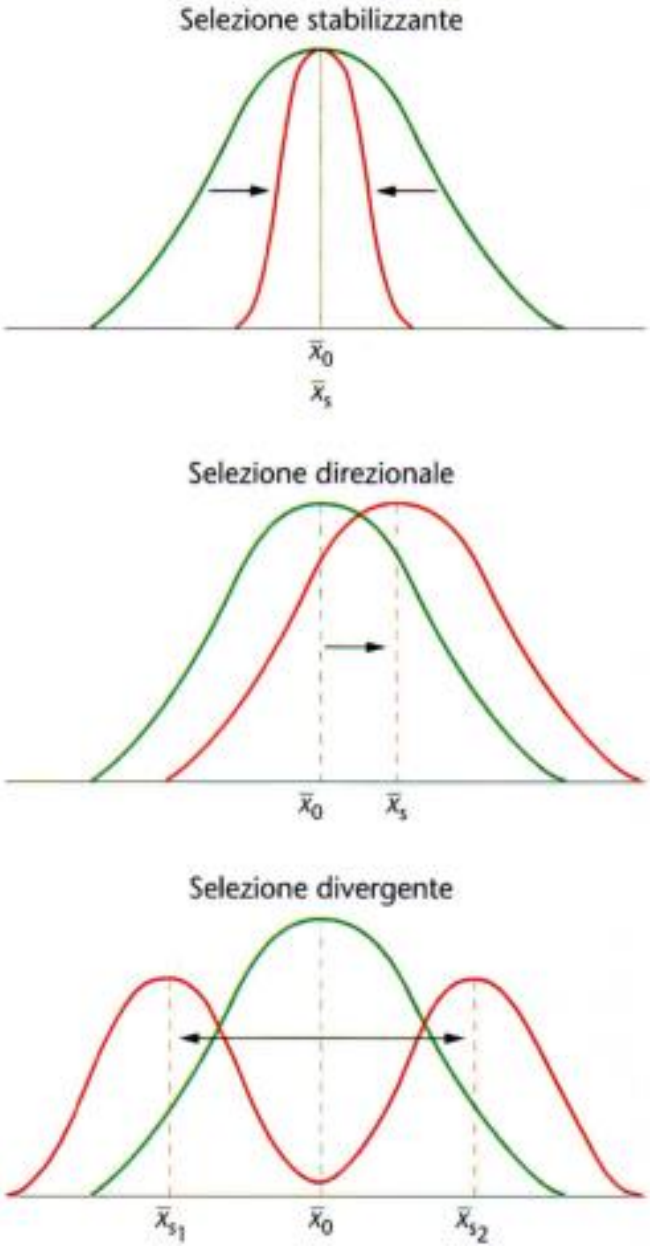
Selezione naturale e caratteri quantitativi

- **Impatto della selezione**
  - **stabilizzante,**
  - **direzionale**
  - **divergente.**

➢ Per ciascun caso, è mostrata

✓ la media di una popolazione originaria (in verde)

✓ la media della popolazione in seguito a selezione (in rosso).



Equilibrio tra selezione e mutazione

- le variazioni delle frequenze geniche, determinate dalla selezione e dalla mutazione, **dipendono entrambe dalle frequenze iniziali**, ma in **modo opposto:**

✓ la mutazione è particolarmente efficace quando l'allele mutato è molto raro (perché ci sono più elementi suscettibili di mutazione),
✓ mentre la selezione è meno efficace quanto più l'allele è raro (perché presente sostanzialmente solo in eterozigosi).

✓ **L'equilibrio** si stabilisce **quando la variazione dovuta alla selezione bilancia quella dovuta alla mutazione**,

- cioè quando la quota di alleli persi a causa della selezione, sq^2, equivale alla proporzione di alleli guadagnati per mutazione, pu.
- Essendo p molto grande (p —> 1), sq^2 —> u, cioè:

$$q = \sqrt{\frac{u}{s}}$$

3.Migrazione

- Effetto della migrazione dipende da:
  - **distanza genetica, cioè della differenza tra le frequenze alleliche del gene in esame nelle due popolazioni (d)**
  - **proporzione di alleli introdotti per generazione** (m).

Consideriamo due popolazioni x e y

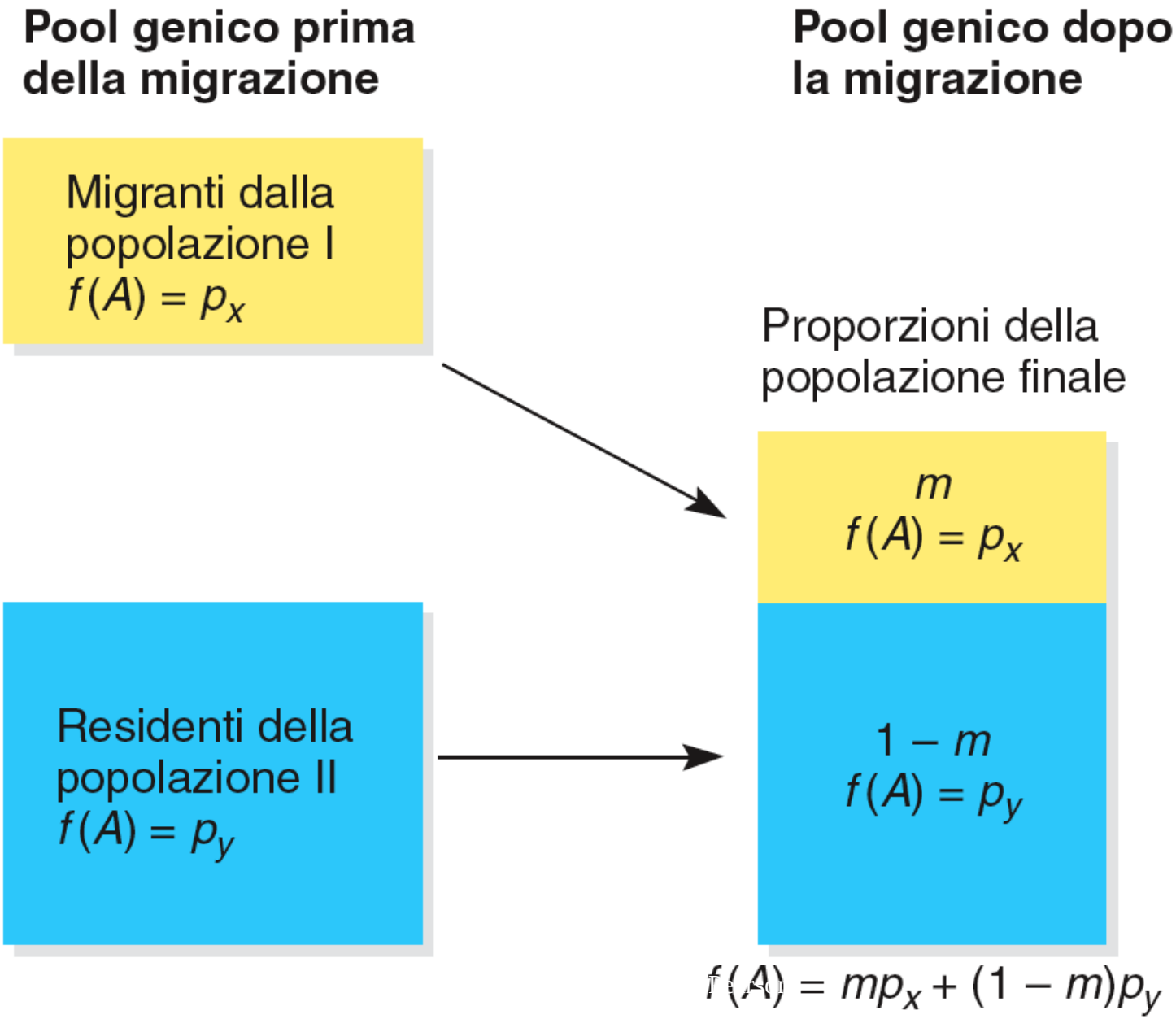
Una parte "m" degli individui della popolazione x migra alla popolazione y

Se la frequenza dell'allele A nella popolazione x è px e nella popolazione y è py

La variazione nella frequenza dell'allele A nella popolazione ricevente sarà

$$\Delta p = m\,(p_x - p_y)$$

- m: proporzione dei migranti nella popolazione finale
- $(p_x - p_y)$: differenza nelle frequenze alleliche

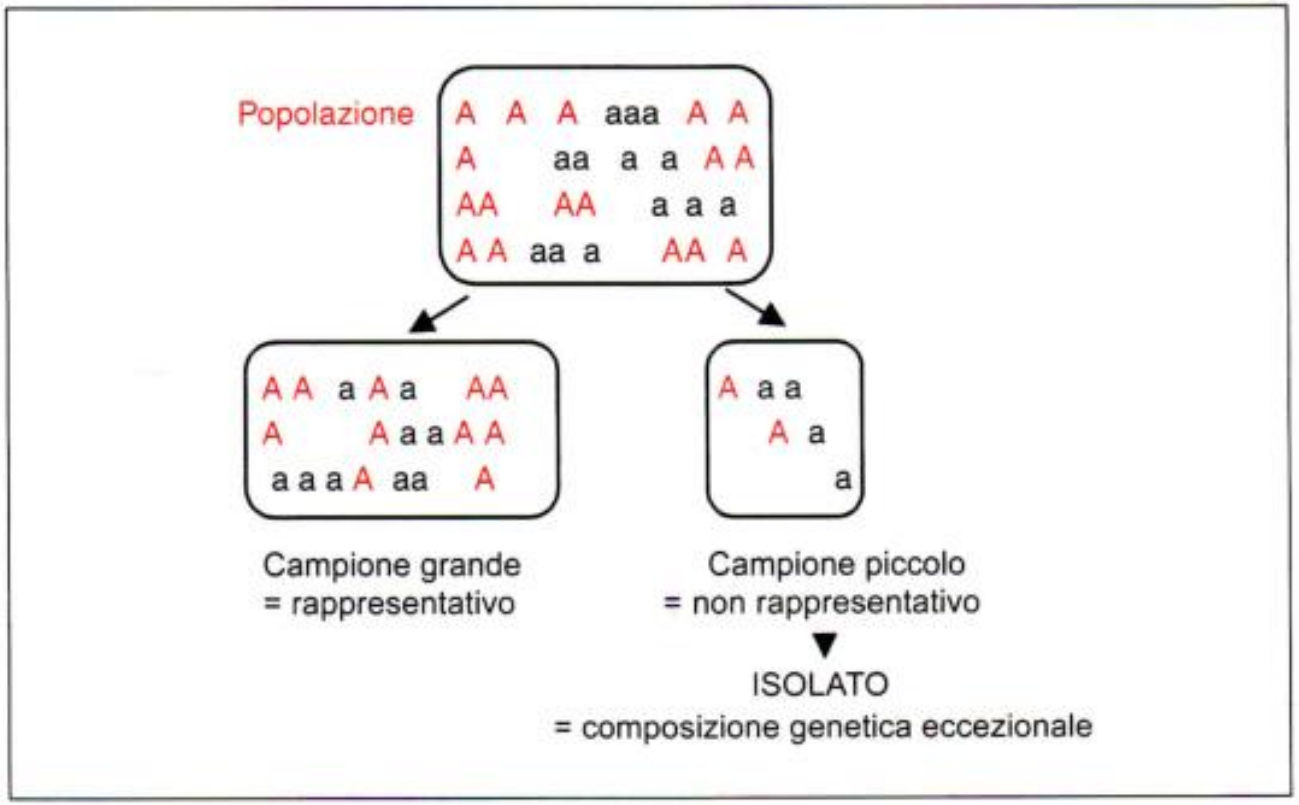

Pool genico prima della migrazione
Pool genico dopo la migrazione
Migranti dalla popolazione I
$f(A) = p_x$
Proporzioni della popolazione finale
m
$f(A) = p_x$
Residenti della popolazione II
$f(A) = p_y$
$1 - m$
$f(A) = p_y$
$f(A) = mp_x + (1 - m)p_y$

La migrazione causa:

- aumento delle dimensioni di una popolazione
- riduce la divergenza tra popolazioni diverse
- si oppone agli effetti della deriva genetica

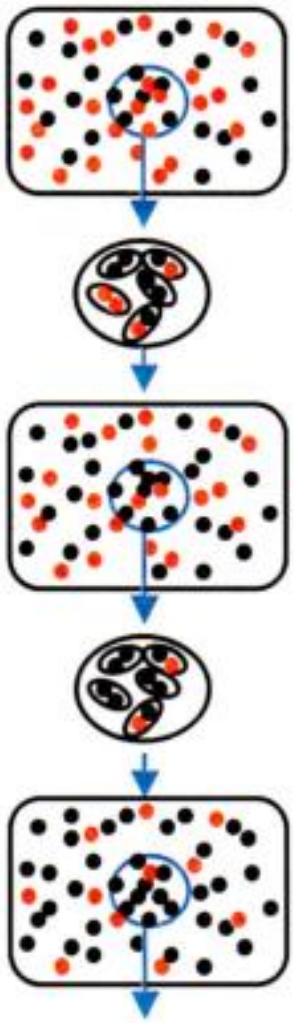
4. Deriva genetica (random drift)

- **Variazione, dovuta al caso, delle frequenze alleliche in una piccola popolazione**



- *Effetto della riduzione della dimensione della popolazione sulle frequenze alleliche*

Random drift



popolazione gametica
q = 0.5

popolazione zigotica
q = 0.4

popolazione gametica
q = 0.4

popolazione zigotica
q = 0.2

popolazione gametica
q = 0.2

...

Random drift

- ***Variazione delle frequenze alleliche in relazione a σ***

$$0.50 \pm \sqrt{\frac{0.50 \cdot 0.50}{10}} = 0.5 - 0.158 = 0.34$$

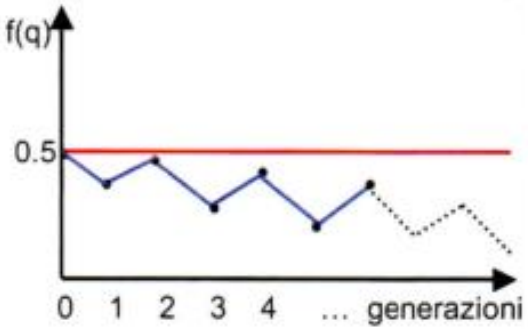
$$0.34 \pm \sqrt{\frac{0.34 \cdot 0.66}{10}} = 0.34 + 0.149 = 0.489$$

$$0.49 \pm \sqrt{\frac{0.49 \cdot 0.51}{10}} = 0.49 - 0.158 = 0.33$$

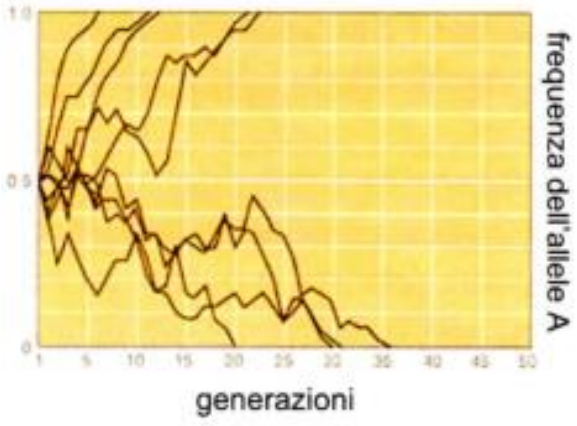
$$0.33 \pm \sqrt{\frac{0.33 \cdot 0.66}{10}} = 0.33 + 0.148 = 0.478$$

$$0.48 \pm \sqrt{\frac{0.48 \cdot 0.52}{10}} = 0.48 - 0.157 = 0.32$$

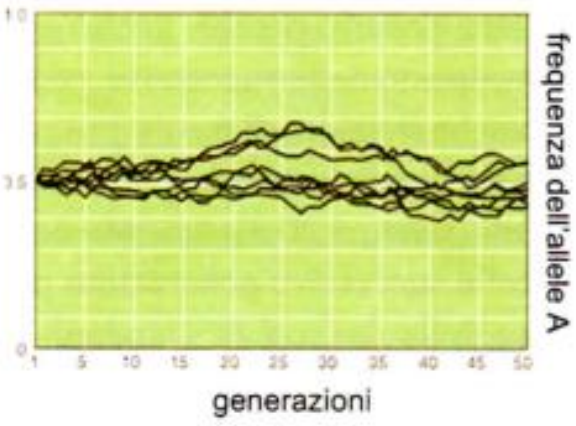
...



- *Effetto del drift sulle frequenze alleliche.*
  a) Popolazione piccola,
  (b) Popolazione grande

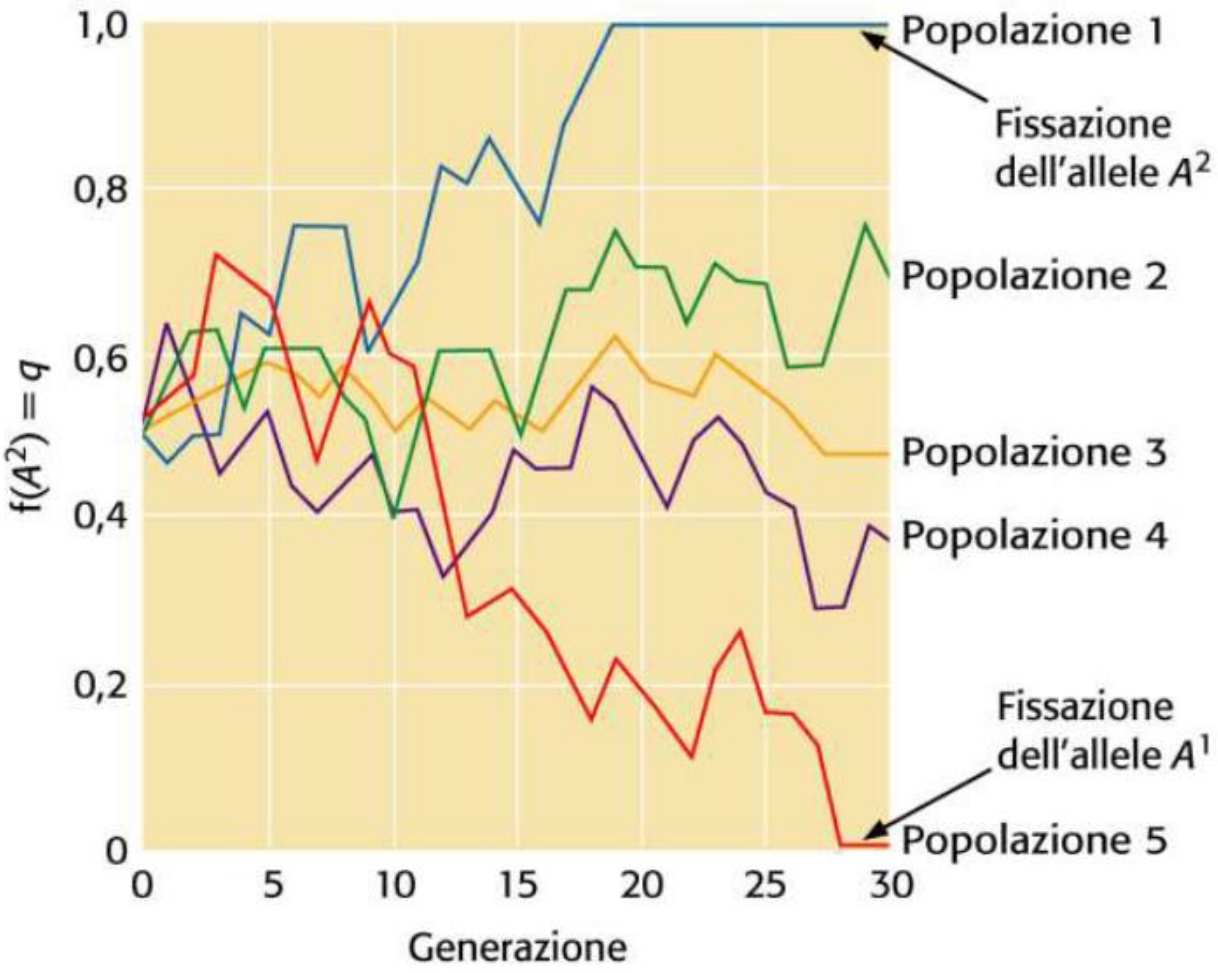


(a) Popolazione piccola



(b) Popolazione grande

La deriva genetica fa variare le frequenze alleliche all'interno delle popolazioni, portando a una riduzione di variabilità genetica attraverso la fissazione e la divergenza tra le popolazioni



La deriva genetica causa:

- perdita di variabilità genetica all'interno delle popolazioni
- divergenza tra popolazioni diverse
- fluttuazione delle frequenze alleliche di una popolazione nel tempo

Cause della deriva genetica

Effetto del fondatore



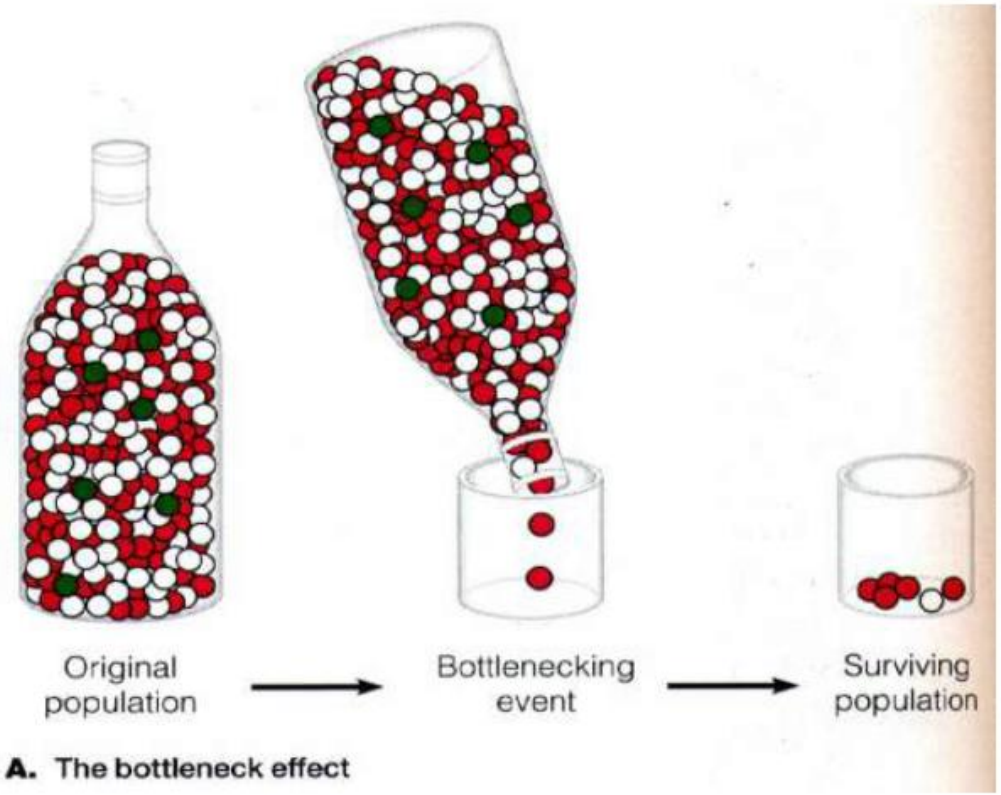
Isola di Tristan da Cuna

- Colonizzata da W. Glass nel 1817
- Arrivo di marinai scampati a naufragio e donne trasferite da S. Elena Popolazione di 100 abitanti nel 1855
- 1856, Morte di Glass.
- Emigrazione e riduzione popolazione a 33 abitanti
- 1885, dopo espansione, nuova riduzione per morte di 15 uomini ed emigrazione delle vedove
- Attualmente, 300 abitanti.
- Elevata incidenza di asma.

Cause della deriva genetica

b. EFFETTO COLLO DI BOTTIGLIA (Bottleneck)

- Si ha quando una popolazione si trova in condizioni sfavorevoli che riducono drasticamente il numero di individui
- Determina variazioni casuali nelle frequenze alleliche simili a quelle dovute all'effetto del fondatore



A. The bottleneck effect

5. Unione/Accoppiamento non casuale

- Può essere la conseguenza di matrimoni fra membri di comunità ristrette e geograficamente isolate
- Può essere la conseguenza di matrimoni combinati fra consanguinei (p.es. primi cugini) per ragioni religiose, culturali, patrimoniali etc.
- Aumenta la probabilità di omozigosi nella prole per alleli recessivi rari→ Comparsa di malattie AR (artrite reumatoide) rare

Unione non casuale

- Quando la scelta del partner riproduttivo non è casuale rispetto al suo genotipo si parla di unione assortativa
- L'unione assortativa è
  - **positiva** quando si scelgono preferenzialmente partner geneticamente affini,
  - **negativa** quando avviene il contrario

- Può essere la conseguenza di matrimoni fra membri di comunità ristrette e geograficamente isolate
- Può essere la conseguenza di matrimoni combinati fra consanguinei (p.es. primi cugini) per ragioni religiose, culturali, patrimoniali etc.
- Aumenta la probabilità di omozigosi nella prole per alleli recessivi rari→ Comparsa di malattie AR (Artrite reumatoide) rare

Unione non casuale

- L'unione assortativa positiva provoca un **deficit di etero-zigoti** rispetto alle attese di Hardy-Weinberg
- Il deficit di eterozigoti viene misurato dal coefficiente **F di inbreeding**
- Coefficienti di inbreeding possono essere stimati dalle frequenze genotipiche o dagli alberi genealogici
- L'inbreeding è conseguenza anche del fatto che il numero di antenati di ognuno raddoppia ad ogni gene-razione, mentre le popolazioni hanno dimensioni finite

. Accoppiamento non casuale

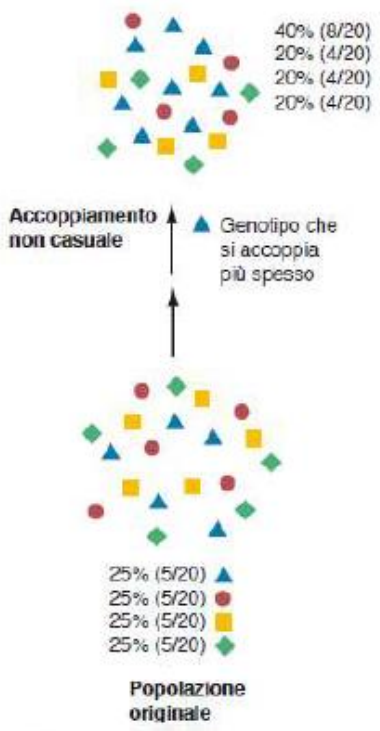


Figura 15.2 **L'accoppiamento non casuale altera le frequenze alleliche.** Le forme colorate diversamente rappresentano individui con genotipi distinti. Se esiste l'equilibrio di Hardy-Weinberg per quei geni in questa popolazione, allora le percentuali rimarranno le stesse da generazione a generazione. Tuttavia, il genotipo del triangolo blu ha più successo riproduttivo, alterando le frequenze alleliche nella generazione successiva.



Figura 15.1 **La scelta del partner influisce sulla diversità genetica nelle popolazioni.** La maggior parte di noi si sposa con persone della stessa popolazione che quindi sono geneticamente simili a noi stessi.

Unione assortativa positiva: autofecondazione

f(AA) = ¼ f(Aa) = ½ f(aa) = ¼

¼ AA x AA → 100% AA

½ Aa x Aa → ¼ AA, ½ Aa, ¼ aa

¼ aa x aa → 100% aa

f(AA) = ¼ + (½ x ¼) f(Aa) = ½ f(aa) = ¼ + (½ x ¼)

f(AA) = 3/8 f(Aa) = ¼ f(aa) = 3/8

Unione assortativa positiva: autofecondazione

f(AA) = 3/8 f(Aa) = ¼ f(aa) = 3/8

3/8 AA x AA → 100% AA
¼ Aa x Aa → ¼ AA, ½ Aa, ¼ aa
3/8 aa x aa → 100% aa

f(AA) = 3/8 + (¼ x ¼) f(Aa) = ¼ f(aa) = 3/8 + (¼ x ¼)

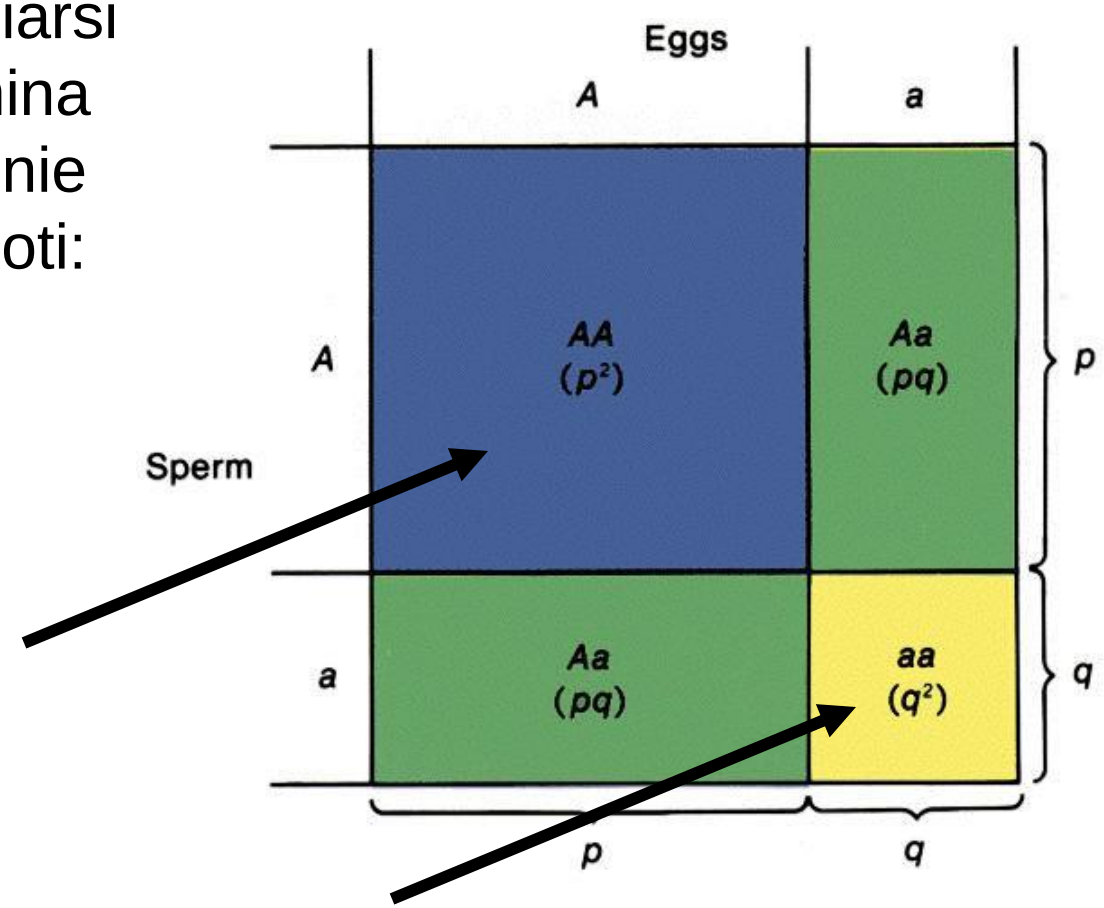
f(AA) = 7/16 f(Aa) = 1/8 f(aa) = 7/16

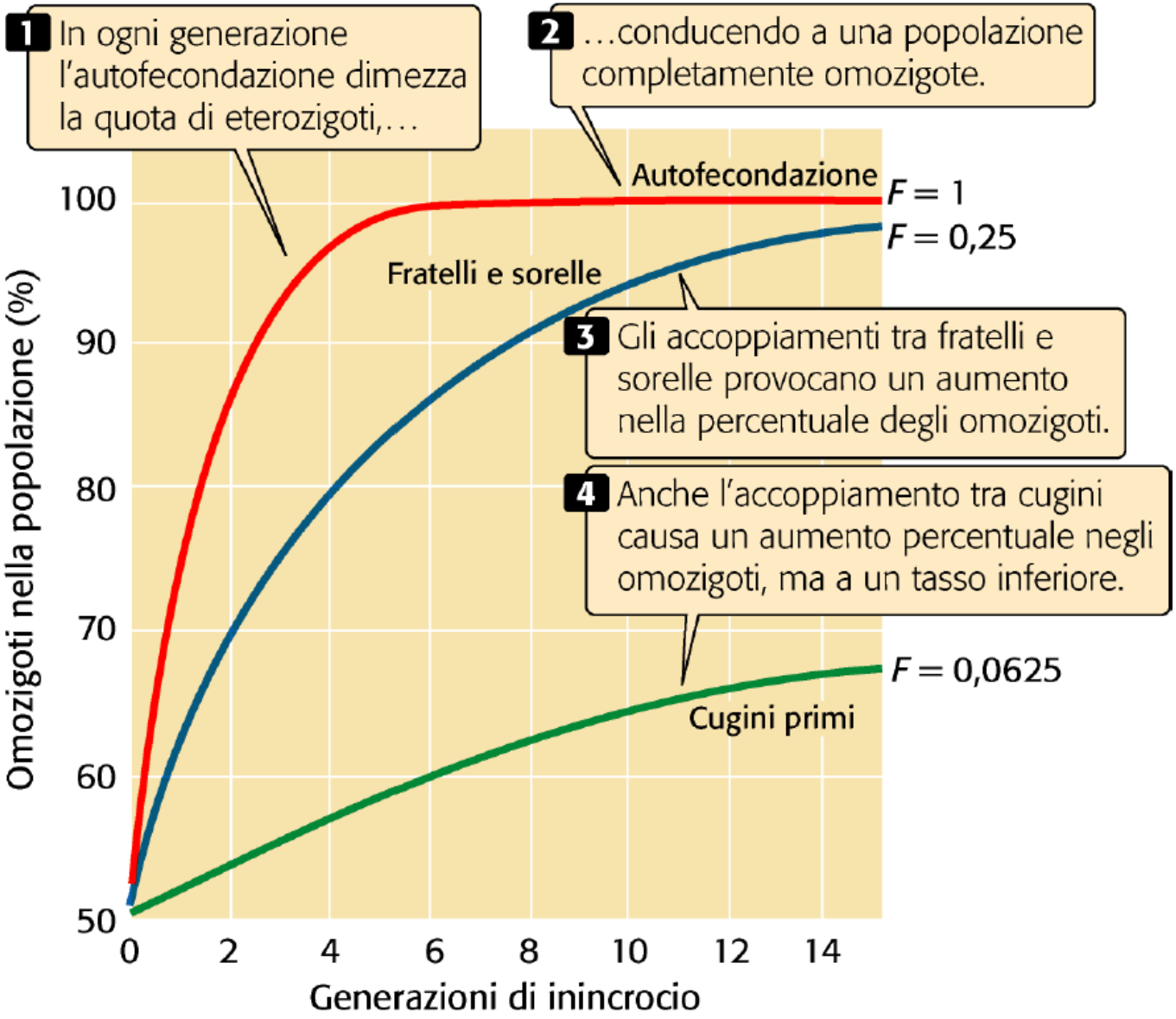
Unione assortativa positiva: autofecondazione

Generazione	AA	Aa	aa
1	1/4	1/2	1/4
2	3/8	1/4	3/8
3	7/16	1/8	7/16
4	15/32	1/16	15/32
N		$1/2^N$	

Effetti dell'inbreeding

- La tendenza ad accoppiarsi fra consanguinei determina la comparsa nella progenie di un eccesso di omozigoti:





Pierce, GENETICA, Zanichelli editore S.p.A. Copyright © 2005

Unione assortativa positiva: inbreeding

- Se $F_{oss}(Aa) = H$

  $F_{att}(Aa) = H_0 = 2pq$

$(H_0 - H) / H_0 = F$ **coefficiente di inbreeding**

$FH_0 = H_0 - H$

$H = H_0 - FH_0$, ma $H_0 = 2pq$

$H = 2pq - 2pqF = 2pq(1-F)$

- Un coefficiente di inbreeding pari a F porta a un deficit di eterozigoti pari a (1-F): metà AA e metà aa

Effetto dell'inbreeding

Genotipo	Hardy-Weinberg	con inbreeding
AA	p^2	$p^2 + pqF$
Aa	2pq	2pq (1-F)
aa	q^2	$q^2 + pqF$

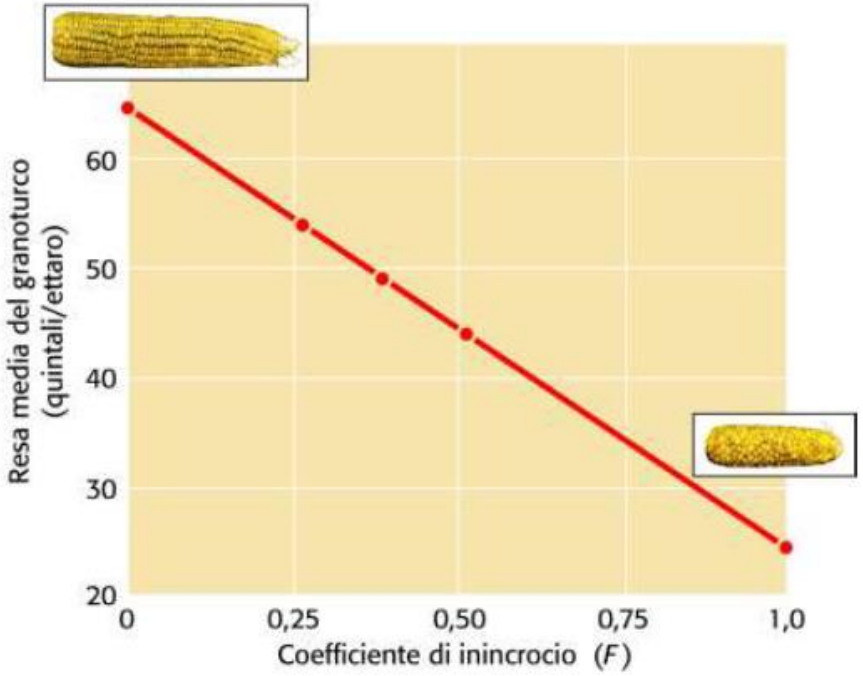
- L'inbreeding **non altera le frequenze alleliche**

Genotipo	F = 0	F = 0,15	F = 0,50	F = 0,85	F = 1
A_1A_1	0,160	0,196	0,280	0,364	0,400
A_1A_2	0,480	0,408	0,240	0,072	0,000
A_2A_2	0,360	0,396	0,480	0,564	0,600

Depressione da inincrocio

Riduzione della *fitness* conseguente alla pratica sistematica dell'*inbreeding*
Dovuta all'omozigosi per alleli recessivi con effetto deleterio

Si osservano ad es. riduzione di fertilità, ridotto tasso di sopravvivenza, comparsa di malformazioni, aumentata suscettibilità alle infezioni



Per la maggior parte delle specie a fecondazione incrociata, l'inincrocio è dannoso in quanto aumenta la proporzione degli omozigoti favorendo quindi la comparsa di omozigoti recessivi deleteri o addirittura letali

Incroci non casuali

- **Inbreeding o Inincrocio**: accoppiamento tra con-sanguinei

- **Esoincrocio o Esogamia**: accoppiamento prefe-renzialmente tra non consanguinei

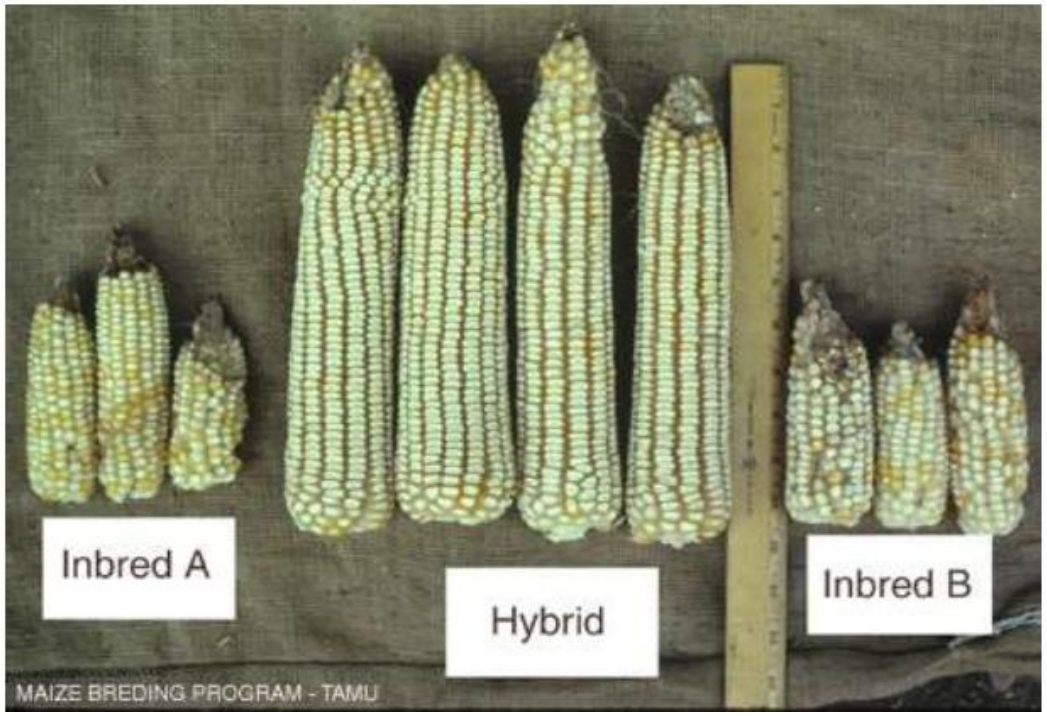
  - **L'inincrocio aumenta l'omozigosità.**
  - **L'esogamia aumenta l'eterozigosità**.

Eterosi (o vigore) dell'ibrido

L'eterosi si riferisce all'incrocio tra individui non imparentati. Il termine è l'opposto di *inbreeding* che si riferisce all'incrocio tra consanguinei.

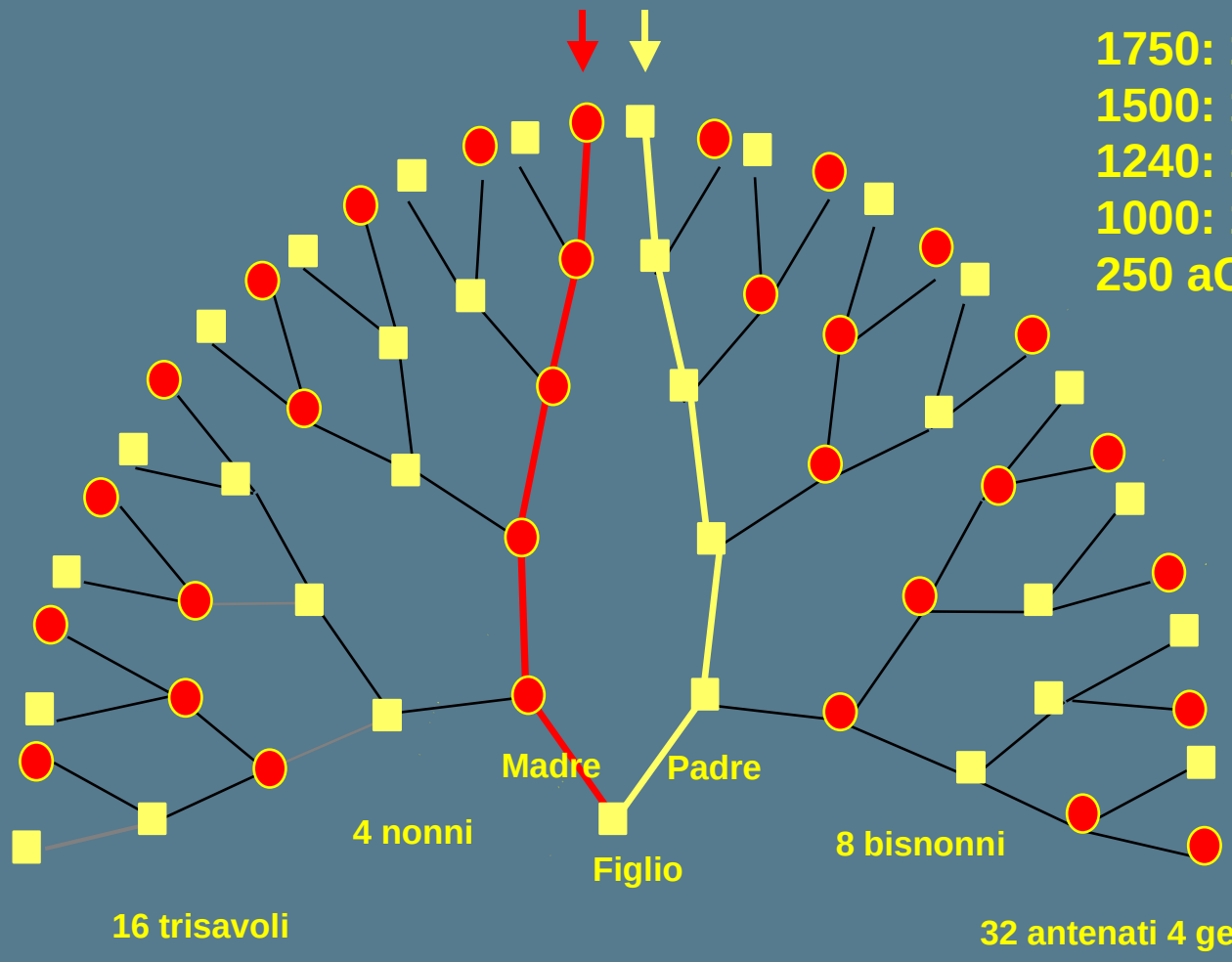
La popolazione prodotto di eterosi risulta con un genotipo in cui aumenta la frequenza di eterozigosi. Al contrario nell'inbreeding aumenta la omozigosi cioè la presenza di alleli identici nello stesso locus.

All'eterosi si associa il fenomeno osservato detto vigore dell'ibrido, in cui l'individuo prodotto dell'accoppiamento risulta con caratteristiche di fenotipo particolarmente vigorose



Abbiamo tanti antenati

6 miliardi di nucleotidi nel genoma umano



1750: 1024 antenati
1500: 1 milione
1240: 1 miliardo
1000: 1000 miliardi
250 aC: 10^{30}

e ciascuno ci ha trasmesso un pezzetto del suo genoma

Nessuno è immune dall'inbreeding

- 40 generazioni fa (1000 dC): 1 000 000 000 000 antenati
- Popolazione stimata della terra: 100 000 000

- 80 generazioni fa: 10^{30} antenati
- Popolazione stimata della terra: 100 000 000

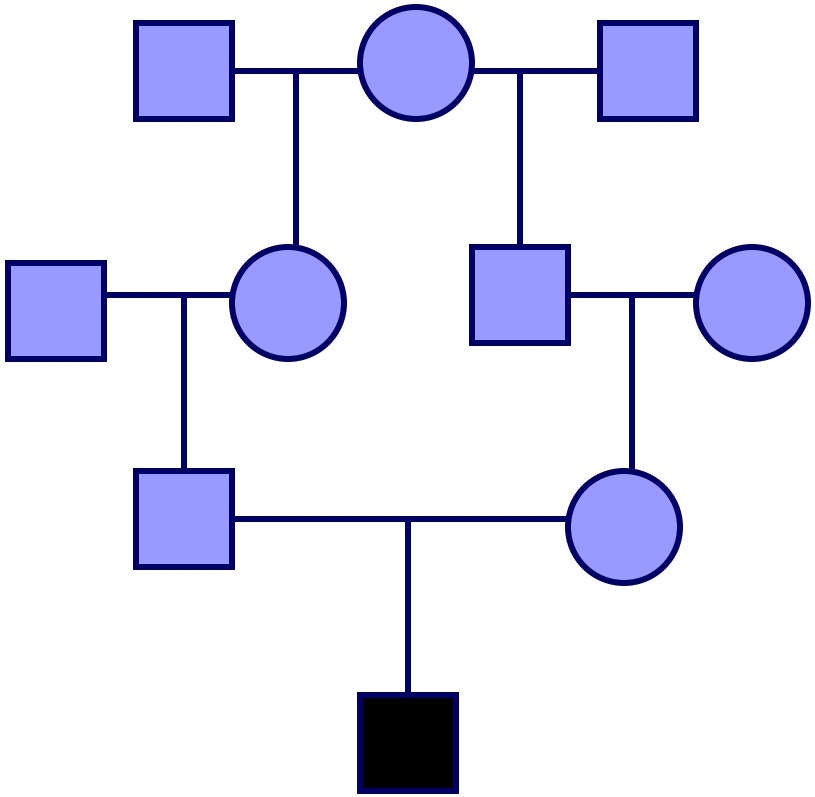
- 1000 generazioni fa: 10^{300} antenati
- Popolazione stimata della terra: 1 000 000

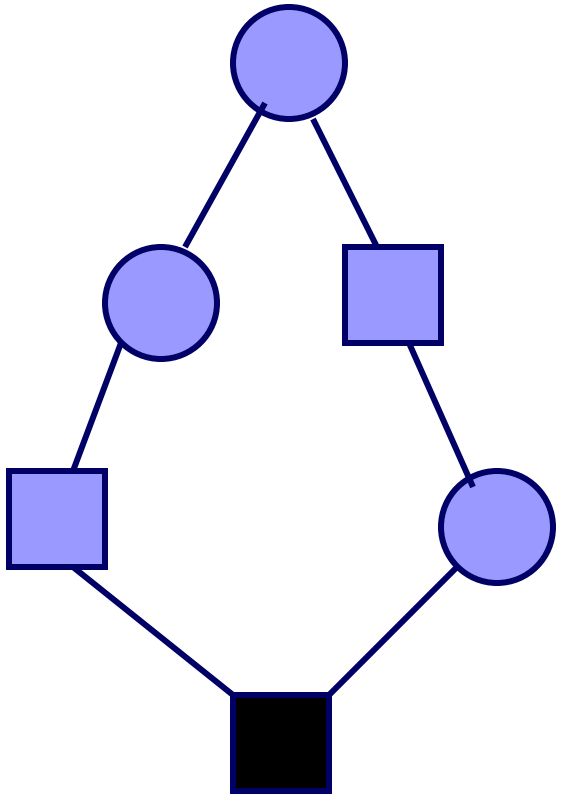
Quindi:

✓ Del milione di individui presenti 25 000 anni fa, molti non hanno lasciato discendenti, molti non sono nostri antenati, altri lo sono miliardi di volte

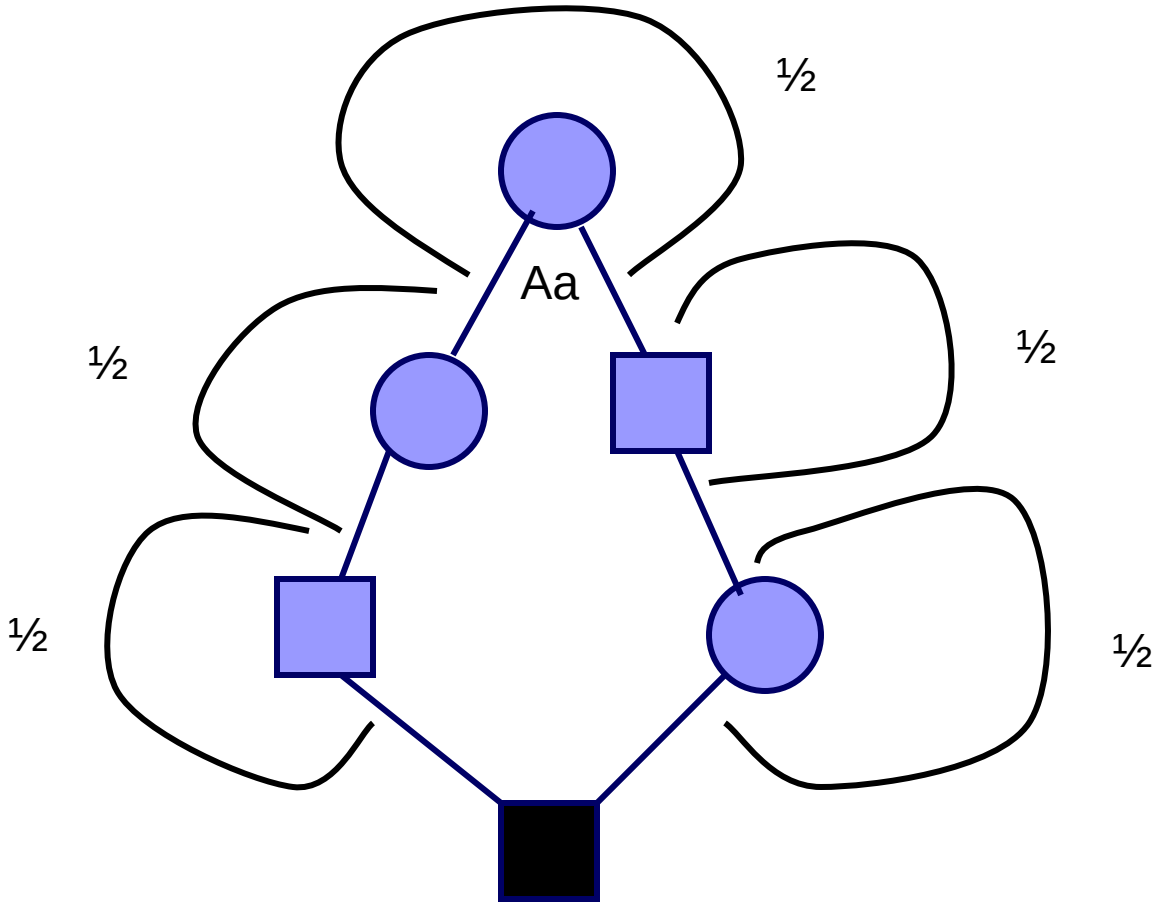
✓ Le nostre genealogie sono tutte fortemente intrecciate

Stima del coefficiente di inbreeding da pedigree



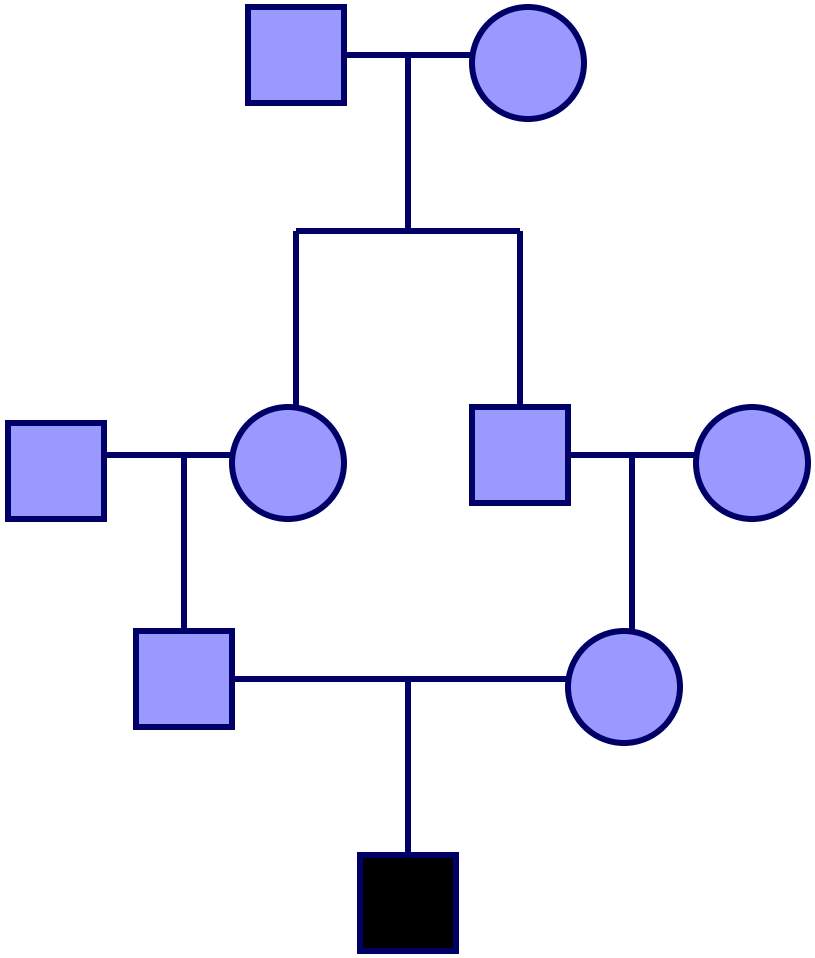


Stima del coefficiente di inbreeding da pedigree



$F = (½)^5 = 1/32$

Stima del coefficiente di inbreeding da pedigree



$F = (½)^5 \times 2 = 1/32 \times 2 = 1/16$

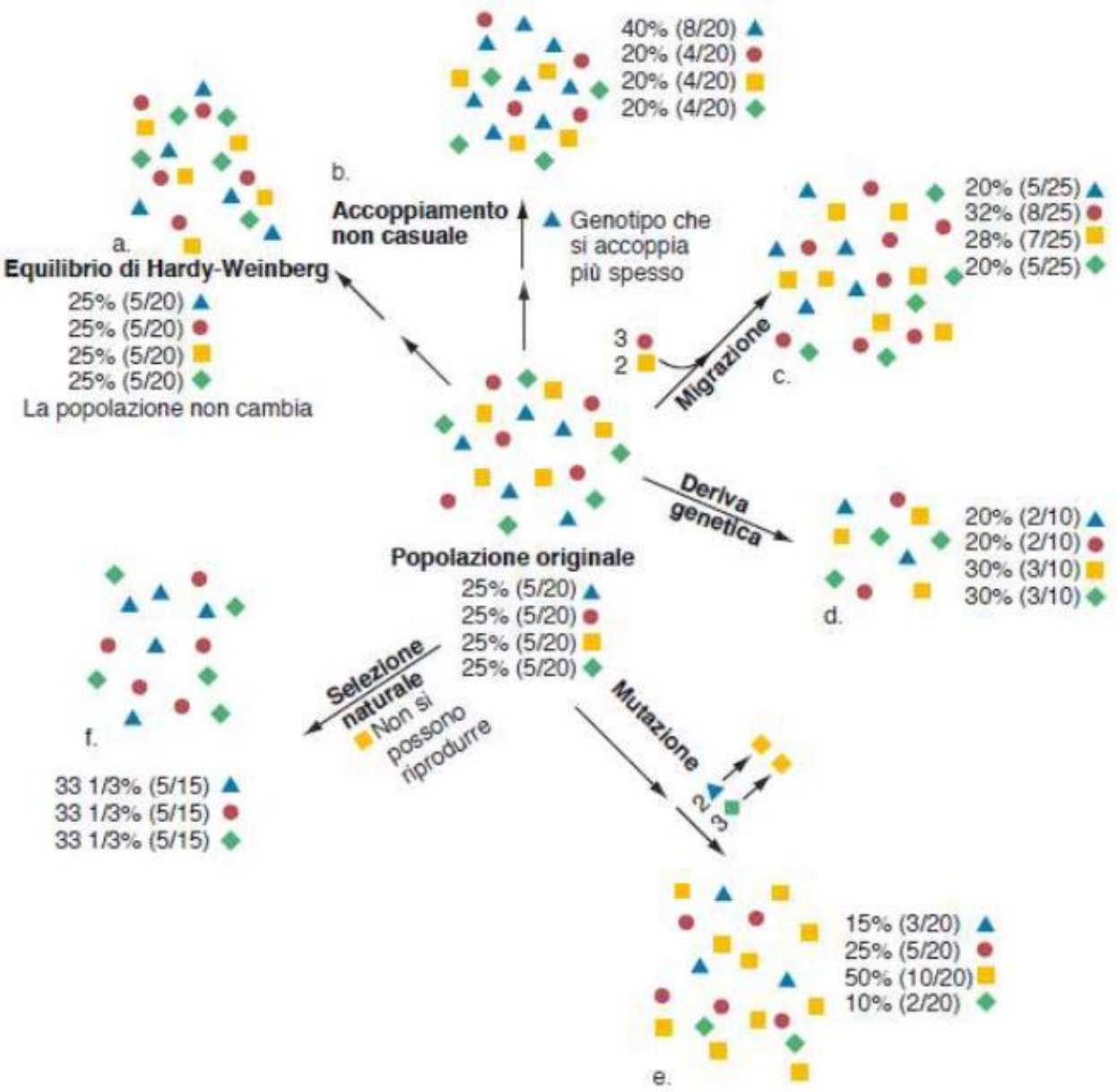
Stima del coefficiente di inbreeding da pedigree

Il valore di F è pari a ½ elevato a una potenza pari al numero di passaggi nel pedigree.

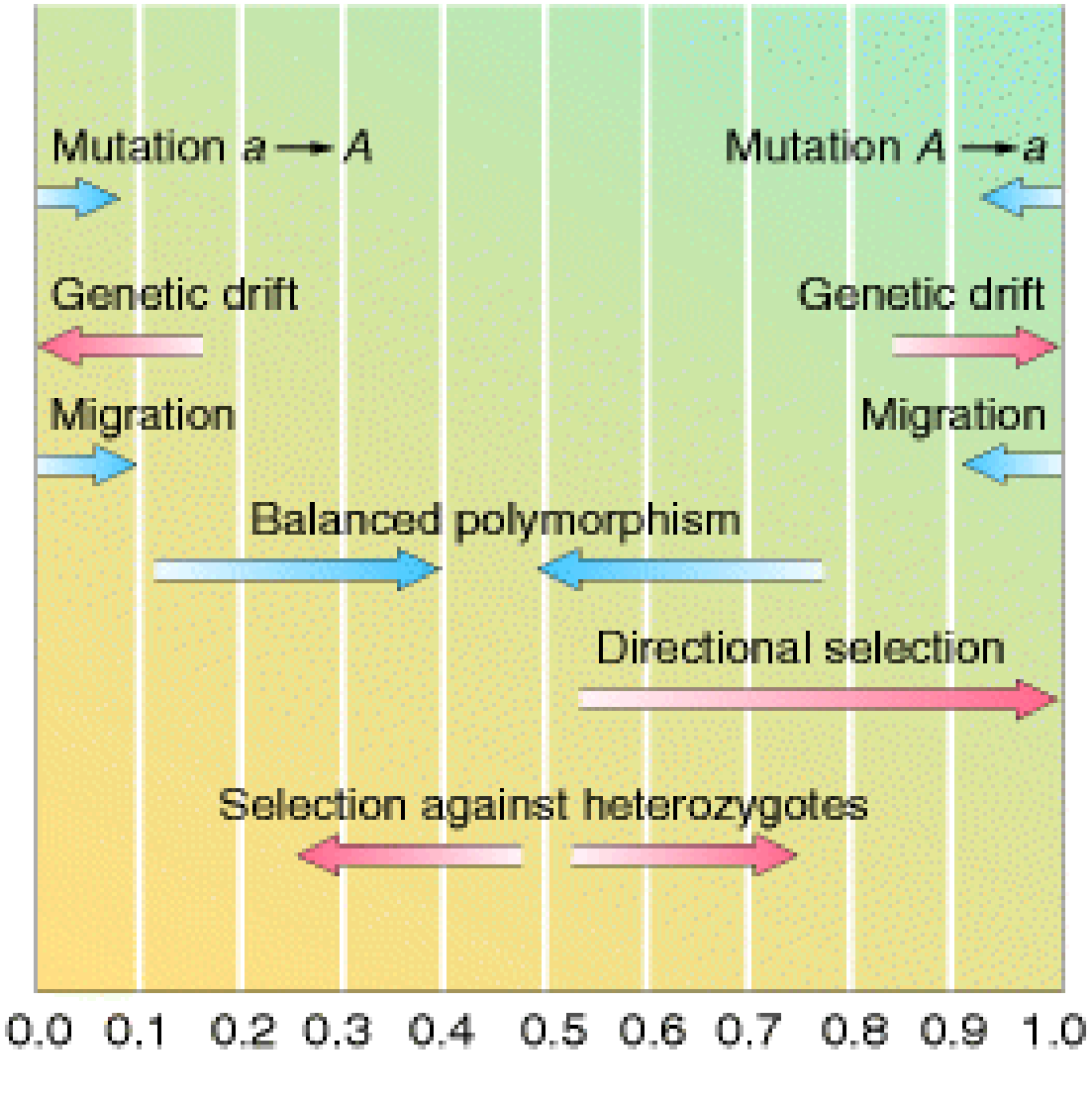
Valore di F nella progenie di varie unioni consanguinee:

Autofecondazione: ½
Fra fratello e sorella: ¼
Fra zio e nipote: 1/8
Fra cugini primi: 1/16
Fra cugini 1 e ½: 1/32
Fra cugini secondi: 1/64
…

Forze che cambiano le frequenze geniche



Riassunto degli effetti delle forze evolutive sulla variabilità genetica




BASTA!