

## INTRODUZIONE ALLA GENOMICA

### GENETICA DI BASE

Le cellule che costituiscono l'organismo animale sono tutte le cellule somatiche, ciascuna delle quali contiene nel suo nucleo un completo corredo di cromosomi. Di ogni cromosoma è possibile individuare un omologo, ossia un cromosoma gemello → i due membri di ogni coppia di cromosomi omologhi provengono ciascuno da un diverso genitore. I cromosomi sono le strutture, visibili talvolta distintamente con un normale microscopio ottico, che veicolano l'informazione genetica caratteristica della specie e specifica di ogni individuo.

Il complesso di questa informazione viene riferito come genoma e, con un'utile approssimazione, possiamo dire che ogni individuo, in ogni sua cellula, possiede due genomi, uno appunto di origine paterna e uno di origine materna.

- uno di origine paterna con cromosoma X o Y
- uno di origine materna con cromosoma X

Ogni genoma è un lungo filamento di DNA di circa 3.000.000.000 di paia di basi (A, T, G, C). Il genoma dei mammiferi è costituito da un lungo filamento di DNA (acido desossiribonucleico) la cui dimensione è ben approssimabile a 3 miliardi di paia di basi. In questo filamento, con le 4 lettere Adenina, Guanina, Timina e Citosina, è scritta tutta l'informazione necessaria per lo sviluppo e il "funzionamento" di ogni individuo.

Il genoma è interrotto in segmenti che in particolari momenti della vita della cellula si vedono anche al microscopio ottico: i cromosomi. I due genomi di ogni individuo sono molto simili ma non identici. Le differenze sono di varia natura, la più semplice è quella di una singola base differente.

Il processo di sviluppo e di differenziazione di ogni individuo trae quindi origine da una cellula originaria (zigote) in cui n cromosomi di origine paterna ed n cromosomi di origine materna si sono fusi nella stessa cellula al momento della fecondazione. Le cellule responsabili di questa fusione sono l'oocita o cellula uovo (femminile) e lo spermatozoo (maschile) che a differenza delle cellule somatiche contengono un numero aploide di cromosomi (n): sono infatti cellule gametiche che derivano da cellule somatiche degli organi sessuali (ovaio e testicolo) attraverso un processo di divisione particolare, differente da una normale mitosi, che prende il nome di meiosi.

I telomeri (la parte terminale del cromosoma) di solito non sono codificanti ma sono strutture di ripetizione.

**GENI:** Segmenti di varia lunghezza nel filamento di DNA costituiscono i geni. Ogni gene è un'unità funzionale complessa che include una regione regolativa che ne modula l'espressione ricevendo e rispondendo a segnali provenienti dall'ambiente circostante. Un gene deve infatti funzionare al momento giusto, nel tessuto giusto e con la giusta intensità. Ogni gene è poi separato dal suo vicino da spazi anche molto ampi di DNA contenenti sequenze magari ripetute e apparentemente prive di qualsiasi utilità funzionale. Anche al suo interno la porzione codificante vera e propria, ossia quella che una volta trascritta verrà effettivamente tradotta in una proteina, include spesso sezioni non codificanti. In un gene vi sono quindi esoni (codificanti) e introni (non codificanti) che si susseguono anche in numero molto elevato. Mediamente un gene di mammifero contiene 7 esoni ed è lungo 15–20 kb di cui solo 2 kb sono effettivamente codificanti.

**SNP:** L'esempio più semplice di allele è forse dato dalla mutazione puntiforme di una singola base o polimorfismo di un singolo nucleotide (SNP = Single Nucleotide Polymorphism). Per esempio, in una certa regione del genoma bovino è presente uno SNP. Questo significa che in quel punto possiamo trovare, nella sequenza codificante, due diverse basi: T (Timina) o G (Guanina). Poiché ogni individuo porta due genomi, questi potranno portare entrambi T oppure G, ma potranno anche essere tra loro diversi e portare l'uno G e l'altro T. Se le due forme sono uguali, TT oppure GG, l'individuo ha genotipo omozigote a quel locus, se sono diverse, TG, l'individuo è eterozigote

Si stima che circa ogni 1.000 paia di basi ci sia una differenza che spesso (ma non sempre) è una base diversa (SNP = Single Nucleotide Polymorphism). Quindi ci sono circa 3.000.000 di SNP in un genoma.

Nella maggior parte dei casi gli SNP si collocano **in regioni non codificanti**. Possono essere quindi definiti dei **MARCATORI (MUTAZIONI) ANONIMI**. Determinano cioè un *polimorfismo* del DNA che non ha alcun riscontro in un polimorfismo del fenotipo (Proteina codificata). Se si trovano in una **regione codificante** possono definire:

- Una **MUTAZIONE CAUSATIVA** cioè che produce un *cambiamento di aminoacido* nella proteina sintetizzata
- Una **MUTAZIONE SINONIMA** cioè che *non produce cambiamenti di aminoacidi* perché la tripletta interessata codifica sempre per lo stesso aminoacido.

Una MUTAZIONE in una regione codificante può determinare anche la formazione di un codone di STOP (TAA, TGA, TAG) che determina l'interruzione della traduzione e la formazione di una proteina incompleta, spesso non funzionante o del tutto mancante

**Traduzione del DNA:** La traduzione del DNA è un processo che avviene nel citoplasma a livello dei ribosomi che scorrono lungo il filamento di mRNA e ne riconoscono la sequenza delle basi a gruppi di tre (triplette).

Il DNA è ridondante (+ triplette per lo stesso aa), all'interno della tripletta se una sostituzione avviene in terza base non ci sono problemi (l'aa rimane lo stesso), se avvenisse in seconda base potrei avere un aa diverso oppure no, in prima base invece avrò sicuramente un aa diverso.

All'interno delle 64 possibilità ci sono dei codoni di STOP o di START, sono specifici per l'attività di produzione della proteina.

**Ridondanza del codice genetico:** Ogni gruppo di 3 basi (codone) viene riconosciuto da un anticodone complementare sito all'estremità di una particolare molecola di RNA (tRNA o RNA transfer) che all'altro estremo porta uno specifico amminoacido.

Ad uno specifico codone corrisponde quindi uno specifico amminoacido che, trainato dal suo tRNA, viene a collocarsi nella giusta posizione lungo la catena proteica che si viene costruendo e stabilisce con l'amminoacido che lo precede quei legami peptidici che definiranno la struttura polimerica primaria della proteina (questo concetto viene indicato col termine di collinearità tra gene e prodotto genico). Poiché le lettere dell'alfabeto genetico sono 4 (A, G, C e T), le combinazioni possibili in gruppi di 3 lettere sono  $4 \times 4 \times 4 = 64$ , mentre gli amminoacidi sono solo 20. Vi è quindi una certa ridondanza di informazione. Si nota infatti che la variazione della terza base di una tripletta non determina in molti casi alcuna sostituzione amminoacidica (questo fatto viene riferito come degenerazione del codice genetico).

Se uno SNP si trova nella regione REGOLATIVA di un gene può determinare un aumento o una diminuzione della trascrizione. Abbiamo in questo caso una variazione quantitativa del PRODOTTO GENICO. Una mutazione causativa determina invece una variazione qualitativa del prodotto genico (variazione della sequenza amminoacidica). In passato si ipotizzava la presenza di un gene osservando il fenotipo: un polimorfismo fenotipico prefigura un sottostante polimorfismo genetico. Al presente, leggendo direttamente il DNA, vediamo numerosissimi polimorfismi genetici che non hanno riscontro nel fenotipo (Marcatori Anonimi). Questi polimorfismi, all'apparenza inutili, trovano invece importanti applicazioni perché tracciano il segmento di DNA nel quale si trovano da un genitore a un figlio. E quel segmento può contenere geni importanti per la sanità veterinaria o la produzione zootecnica.

**FREQUENZE ALLELICHE E GENOTIPICHE:** Supponiamo una popolazione composta da 1.000 individui. Se ogni individuo porta due genomi, ci saranno in tutto 2.000 genomi circolanti. Supponiamo di caratterizzare il genotipo ad un certo SNP (allele T e allele G) nei 1.000 individui. Otteniamo questi risultati:

FREQUENZA	GENOTIPO
490	TT
420	TG
90	GG

Dalle frequenze genotipiche possiamo calcolare le frequenze alleliche:

Allele T:	$(490 \times 2 + 420) / 2.000 =$	$0,7 = p$
Allele G:	$(90 \times 2 + 420) / 2.000 =$	$0,3 = q$

Una legge fondamentale della genetica (Hardy-Weinberg, HW) dice che:

- se una popolazione è sufficientemente ampia:  $N \rightarrow \infty$ ,
- se tutti gli individui si possono accoppiare a caso e senza restrizioni

Le frequenze degli alleli (p e q) restano stabili nel tempo ossia al passare delle generazioni. Inoltre le frequenze dei genotipi sono determinate dall'espansione del binomio  $(p + q)^2$  quindi sono pari a  $p^2$ ,  $2pq$  e  $q^2$ . Quindi, se le frequenze alleliche restano stabili, lo sono anche le frequenze genotipiche. Questa condizione viene definita come: **EQUILIBRIO DI HARDY-WEINBERG (HWE)**

Se questi alleli hanno un effetto qualsiasi su qualche fenotipo, e le due condizioni precedenti sono valide, la media di quel fenotipo non cambia nel tempo (stessa incidenza di malattie, stessa produzione).

VEDI → LEZIONE 2. EQUILIBRIO DI HARDY-WEINBERG (HWE) → CALCOLI

## GENETICA DELLE POPOLAZIONI

## Frequenze geniche e genotipiche: Variabilità Genetica

Quando parliamo di variabilità genetica dobbiamo fare riferimento a qualcosa di ben preciso. In una popolazione osserviamo tre genotipi CC, CG e GG. Se  $p$  = frequenza di C = 0,4 e di conseguenza  $q$  = frequenza di G = 0,6 allora gli eterozigoti saranno  $2pq = 0,48$  ossia il 48% del totale degli animali. Questa quota di eterozigoti ( $2pq$ ) possiamo descriverla con la semplice curva. Questa ci spiega che  $2pq$  raggiunge un massimo (0,50) quando  $p = q$  e arriva a zero quando uno dei due alleli, fissato ( $p = 1$ ) e l'altro estinto ( $q = 0$ ).

La variabilità genetica ad ogni locus precisamente associata a questa curva ed esiste solo in presenza di un polimorfismo (almeno 2 alleli) ed è massima quando i due alleli hanno frequenze bilanciate. Se lo SNP in questione avesse un effetto genetico di qualche natura e questo effetto avrà peso sulla variabilità genetica

In generale possiamo dire che:

1. In una popolazione che si accoppia casualmente e che sia sufficientemente ampia, in assenza di selezione, mutazione e migrazione, le frequenze dei geni restano stabili da una generazione all'altra.
2. Se le frequenze geniche restano stabili da una generazione all'altra, restano anche ugualmente stabili le medie di caratteri quantitativi controllati da questi geni nonché le frequenze delle patologie associate a specifiche mutazioni.
3. La selezione naturale o quella artificiale (ma anche la mutazione, la deriva genetica e la migrazione) possono tuttavia cambiare le frequenze dei geni e di conseguenza l'incidenza delle patologie o la media dei caratteri quantitativi.

Effetto della selezione sulle frequenze geniche: Le frequenze degli alleli presenti in un certo gene tendono a rimanere costanti nel tempo. Le frequenze possono tuttavia cambiare attraverso l'opera della selezione. Selezionare significa scegliere gli animali ai quali è permesso avere dei discendenti (riproduttori).

**Deriva Genetica o Bottleneck:** La deriva in pratica introduce la possibilità che da una generazione all'altra le frequenze cambino non per selezione, ma per caso. E questo avviene soprattutto a fronte di un numero ristretto di riproduttori. Questa variazione, del tutto casuale, può giocare un ruolo drammatico nelle piccole popolazioni che in conseguenza di "colli di bottiglia" demografici possono cambiare anche vistosamente le proprie caratteristiche genetiche.

Quindi, quanto più la dimensione di una popolazione ridotta, tanto più elevato il rischio che alcuni alleli arrivino all'estinzione, alleli che magari erano tipici ed esclusivi di quella popolazione.

Esempio: Se la popolazione composta solo da 8 individui, ma sono sufficienti solo due riproduttori, un maschio e una femmina, potrebbe capitare che per puro caso i due prescelti siano entrambi CC, con una diretta estinzione nella generazione filiale dell'allele G. La variabilità con cui la frequenza di C oscilla intorno alla media di 0,5 non dipende quindi dal numero complessivo degli animali ( $N = 8$ , numero censito), ma dal numero effettivo,  $N_e = 2$ . Tipicamente nelle popolazioni in selezione vi è spesso un forte sbilanciamento tra maschi (M) e femmine (F) che si riproducono.

**Numero Effettivo:  $N_e$** 

Ad esempio, una popolazione di 1000 animali (numero censito) si riproduce grazie a 50 maschi e 950 femmine.

- Il  $N_e = 190$ .
- **$N_e = \frac{4MF}{M+F} = \frac{4 \cdot 50 \cdot 950}{1000} = 190$ .**

La deriva genetica può quindi avere effetti importanti anche in una popolazione molto numerosa in cui la forte selezione, soprattutto nei maschi, abbia ridotto fortemente la sua dimensione numerica effettiva ( $N_e$ ).

Ad esempio, 990 vacche e 10 tori in FA danno un numero effettivo di  $N_e = 40$ .

## Esercizi

- Calcolare numero effettivo di una popolazione con 100 maschi e 800 femmine.
- $N_e = \frac{4 \times 100 \times 800}{900} = 355$
- Calcolare il numero effettivo con 5 maschi e 15 femmine
- $N_e = \frac{4 \times 5 \times 15}{20} = 15$

## Alcune osservazioni

- L'equazione del numero effettivo assume assenza di selezione, cioè una situazione in cui i riproduttori sono un campione causale dei soggetti nati.
- Se effettuiamo la selezione diminuiamo il numero di riproduttori ed i soggetti migliori avranno più progenie.
- Inoltre con la selezione si sceglieranno pochi soggetti con fenotipo simile che tendono ad avere la stessa genetica.

- Un buon riproduttore tenderà ad avere molti figli e nipoti, e pronipoti aumentando la parentela media della popolazione.

Quindi la selezione: riduce il numero effettivo, Riduce il numero dei riproduttori, Crea disuguaglianza della numerosità della progenie per riproduttore.

Strategie per popolazioni con un numero effettivo molto ridotto

- Massimizzare il numero effettivo della popolazione (aumentando il numero dei riproduttori)
- Minimizzare la parentela tra i riproduttori (con l'uso di genealogie accurate e sufficienti)
- Il numero di riproduttori maschi deve essere il più grande possibile, idealmente il più vicino possibile al numero delle femmine
- Minimizzare la parentela di accoppiamento
- Idealmente cercare di ottenere un numero di progenie uguale per ogni riproduttore.

Naturalmente qui parliamo della selezione artificiale che viene operata dall'uomo che sceglie gli animali (e quindi indirettamente i geni) in funzione del loro valore economico. La selezione naturale invece agisce scegliendo sì gli animali migliori, ma non in considerazione del loro valore economico, bensì premiando i più adatti all'ambiente naturale con la possibilità di lasciare una discendenza più numerosa di quelli meno adatti.

La frequenza di un gene in una popolazione sottoposta a selezione dipende dalla fitness (adattamento, ossia capacità di lasciare discendenti) dei genotipi nella popolazione. La misura assoluta della fitness di un genotipo non è certo semplice, ma per i nostri scopi sarà facile e sufficiente utilizzare una fitness relativa. Con fitness relativa si intende la proporzione di animali di un certo genotipo che si riproduce rispetto ad animali con diverso genotipo.

Fitness da 1 a zero: Una fitness pari a 1 indica una completa capacità di riprodursi di quel genotipo, una fitness inferiore a 1 indica fitness ridotta, per incompleta vitalità o fertilità o altro, di un particolare genotipo.

## SOMIGLIANZA

Quando si parla di somiglianza tra individui ci si riferisce di norma al fatto che riconosciamo delle caratteristiche fenotipiche in comune tra animali diversi, dovute, possiamo facilmente immaginare, alla loro parentela.

In campo umano questa è un'esperienza elementare: i figli assomigliano ai genitori e i fratelli tra loro. Invece cognati, generi o suocere, pur essendo "parenti", non si assomigliano in quanto appartengono a famiglie originarie diverse e non hanno alcuna componente genetica in comune. La parentela, in senso lato, quantifica precisamente il genotipo in comune tra due individui, anche se come vedremo è possibile parlare di parentela di un individuo con sé stesso

### Parentela ( $a_{ij}$ )

Definizione: La parentela esprime la proporzione di geni in comune per discendenza mendeliana tra due animali. Questi geni in comune determinano la somiglianza che caratterizza due individui tra loro parenti.

Poiché i geni possono avere un effetto semplice e diretto sul fenotipo, due animali con una parte dei loro geni in comune avranno anche in comune i relativi effetti semplici (additivi) e questo si tradurrà in una certa somiglianza dei loro fenotipi, siano questi i loro caratteri morfologici o le loro produzioni

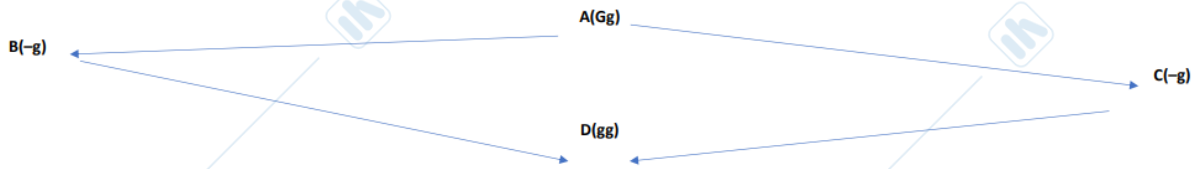
**Parentela additiva:** Si avrà così una parentela additiva ( $a_{ij}$ ) che, per le produzioni animali e la medicina veterinaria, è della massima importanza perché esprime la proporzione di geni in comune tra due individui (i e j) e, quindi, dei relativi effetti genetici. La parentela (in inglese relationship) è quindi un numero associato a una specifica coppia di individui.

In linea generale ogni individuo, nel momento in cui si riproduce, trasmette alla progenie metà dei suoi geni cosicché genitore e figlio hanno in comune metà dei loro geni e dei relativi effetti genetici. Ad ogni generazione, inoltre, i geni trasmessi si dimezzano cosicché un nonno può trasmettere in media al nipote solo il 25% dei suoi geni. Negli animali domestici alcune strutture familiari si ripropongono regolarmente con rapporti di parentela ben precisi che dipendono dalle loro caratteristiche riproduttive (specie prevalentemente unipare o specie multipare).

Rapporto di parentela	$a_{ij}$	Specie
Genitore-figlio ( <i>parent-progeny</i> )	$\frac{1}{2}$	Tutte
Nonno-nipote ( <i>grandparent-grandprogeny</i> )	$\frac{1}{4}$	Tutte
Fratelli pieni ( <i>Full Sibs, F5</i> )	$\frac{1}{2}$	Conigli, suini, cani, gatti
Mezzi fratelli ( <i>Half Sibs, H5</i> )	$\frac{1}{4}$	Bovini, equini, ovini, caprini
Gemelli dizigotici ( <i>fraternal twins</i> )	$\frac{1}{2}$	Tutte
Gemelli monoigotici ( <i>identical twins</i> )	1	Tutte

**Consanguineità (F):** Si parla invece di consanguineità (o inbreeding) quando, riferendosi a un solo individuo, si vuole indicare la proporzione di geni allo stato omozigote per discendenza mendeliana che ci si aspetta siano presenti nel suo genotipo. L'espressione "geni omozigoti per discendenza mendeliana" significa geni che derivano dalla duplicazione di un gene originariamente presente in un antenato comune ai due genitori dell'individuo. Quindi la consanguineità di un individuo è diversa da zero solo se i suoi due genitori sono tra loro parenti.

ESEMPIO: due genitori di D (B e C) possono aver ricevuto lo stesso gene  $g$  da A che era originariamente eterozigote  $Gg$ . D può quindi essere omozigote  $gg$  con una certa probabilità. Questa probabilità, relativa a un gene qualsiasi di D, costituisce il suo coefficiente di consanguineità (F).



Le forme recessive si evidenziano alla fine, e senza analisi genetica non possiamo distinguere tra portatore e sano.

Questa stessa probabilità può anche essere intesa come frazione di geni omozigoti per discendenza complessivamente presenti nel genoma di un individuo. In altri termini, per esempio, una consanguineità del 10% ( $F = 0,1$ ) significa che la probabilità di un gene qualsiasi di essere omozigote per discendenza mendeliana è di 1 su 10, oppure che il 10% dei geni di quell'individuo sono omozigoti per discendenza mendeliana.

Naturalmente un individuo può benissimo essere omozigote  $gg$  perché nella popolazione il gene  $g$  è presente con una certa frequenza ( $q$ ) e quindi individui  $gg$  sono 2 attesi come prodotto di accoppiamenti casuali nella proporzione di  $q$ .

In questo caso i due geni sono identici in stato e vengono anche detti allozigoti. Nel caso precedente invece i due geni  $g$  sono identici per discendenza mendeliana (in inglese, *Identical By Descent*, IBD) e sono anche detti autozigoti.

### Libri genealogici e Registri Anagrafici

La registrazione dei dati genealogici delle popolazioni animali è una pratica molto antica e si affermò inizialmente nella specie equina per il Purosangue Arabo e Inglese con finalità di semplice certificazione commerciale. Oggi tutte le razze domestiche o quasi tengono registrazioni meticolose delle genealogie depositate nel cosiddetto Libro Genealogico (LG). Qui, di ogni animale, vengono registrati la data di nascita, il padre, la madre e altre informazioni. Sebbene gli animali abbiano spesso un nome attribuitogli dall'allevatore, la registrazione di un soggetto è basata anche e soprattutto su una identificazione matricolare numerica (o alfanumerica, composta cioè da numeri e lettere) che individua il soggetto in modo univoco.

La maggior parte dei Libri Genealogici ha iniziato le registrazioni nel XIX o nel XX secolo, ma per alcune razze o popolazioni minori l'attività è molto recente, cosicché la profondità delle registrazioni disponibile può essere alquanto variabile.

Se un animale è iscritto al LG e può così vantare antenali registrali per diverse generazioni, il suo valore commerciale è di solito superiore a quello di un soggetto di ignota genealogia. Alcuni LG detti *aperti* accettano al loro interno, con limitate restrizioni, anche soggetti di ignota genealogia purché presentino i caratteri tipici e distintivi della razza.

Altri, detti *chiusi*, sono invece molto rigorosi nell'escludere dall'iscrizione soggetti che non abbiano registrato almeno un certo numero di generazioni di antenati (per esempio 3 oppure 5). Chi possiede un animale iscritto a un LG può ottenere, dall'Associazione autorizzata alla tenuta del LG, un certificato genealogico dove, oltre a nome, matricola, data di nascita, nome del proprietario e altre informazioni viene riportata la cosiddetta tabella genealogica del soggetto in questione. In questa tabella vengono riportati i 2 genitori del soggetto, di questi a loro volta i genitori e via dicendo fino a un numero definito di generazioni.

**Registri anagrafici:** Accanto ai LG in Italia sono stati istituiti a partire dagli anni '90 dello scorso secolo, anche dei Registri Anagrafici (RA) per le popolazioni zootecniche di ridotto effettivo numerico, magari a rischio di estinzione. Questi RA raccolgono di norma solo delle registrazioni genealogiche finalizzate esclusivamente alla conservazione genetica della popolazione interessata. A differenza del LG, che è uno strumento di selezione, il RA quindi esclude questa attività.

La realtà, tuttavia, ben difficilmente è così dicotomica: solo selezione da una parte e solo conservazione dall'altra. Anche una popolazione molto piccola dovrà operare una qualche scelta selettiva per lo meno sui maschi e parimenti anche una popolazione molto ampia a causa della riduzione del numero dei riproduttori dovrà tenere conto di una certa erosione della sua variabilità genetica e conseguentemente della necessità di conservarla.

Generazione Soggetto	1 <sup>a</sup> Genitori	2 <sup>a</sup> Nonni	3 <sup>a</sup> Bisnonni	4 <sup>a</sup> Trisnonni
A	B (p)	E (pp)	I (ppp)	Q (pppp)
			J (mpp)	R (mppp)
			K (pmp)	S (pmpp)
			L (mmp)	T (mmpp)
		F (mp)	U (ppmp)	
			V (mpmp)	
			Q (pmmp)	
			X (mmpm)	
	C (m)	G (pm)	M (ppm)	Y (pppm)
			N (mpm)	Z (mppm)
			Q (pmpm)	
			W (mmpm)	
		H (mm)	O (pmm)	S (ppmm)
			V (mpmm)	
			Q (pmmm)	
			X (mmpm)	
P (mmm)	X (mmpm)			

Osservando la tabella genealogica di esempio, si possono fare alcune osservazioni circa il ricorrere di alcuni identici riproduttori in punti diversi del pedigree.

Per esempio, fra gli antenati di A, si osserva che Q, S, V e X compaiono nella tabella sia tra gli antenati di B sia tra quelli di C.

Questi animali, che vengono indicati come antenati comuni, daranno quindi un certo contributo al grado di parentela tra B e C. Il contributo sarà naturalmente tanto minore quanto più lontano nel passato si trova l'antenato comune

### Linee di Sangue

Gli allevatori appassionati conoscono spesso a memoria molte delle genealogie della razza che allevano e spesso definiscono un animale come appartenente alla "linea di sangue A" oppure B o C ecc. Fanno riferimento in questo caso ad alcuni riproduttori che sono stati dei fondatori famosi della razza o semplicemente del loro allevamento. Per definire la linea di appartenenza da cui un individuo discende basta ricostruire la genealogia seguendo la linea di padre in figlio. Nel caso della tabella riportata il soggetto A potrebbe essere definito di linea Q, perché risalendo di figlio in padre si arriva al fondatore Q. Se gli animali venissero identificati come gli umani si tratterebbe di un vero e proprio cognome, una sorta di gene trasmesso dal cromosoma Y del fondatore. La tabella genealogica è spesso uno strumento molto utile all'allevatore che se ne serve per programmare gli accoppiamenti tra i suoi animali con scopi precisi. Un'elevata parentela tra due possibili genitori darà luogo, per esempio a un figlio consanguineo (inbreed), che potrà manifestare allo stato omozigote patologie indesiderabili.

Misure di base della parentela

Il **coefficiente di kinship** = (f, raramente detto anche coancestry o consanguinity). Il coefficiente di kinship si definisce come la probabilità che, estraendo a caso due geni allo stesso locus in due individui i e j, questi siano identici per discendenza.

**Inbreeding e linebreeding:** Come abbiamo visto, l'accoppiamento di animali parenti produce una progenie inbreed, ossia omozigote a un numero di loci maggiore di quello che ci aspetteremmo in condizione di accoppiamento casuale. Un uso protratto e sistematico della consanguineità produrrà quindi un aumento dell'inbreeding e, conseguentemente, una diminuzione dell'eterozigosità degli animali. In condizioni limite, quando l'inbreeding degli animali sarà pari a 1, tutti i geni saranno omozigoti e gli animali tutti tra loro geneticamente uguali

**Inbreeding:** Di norma gli allevatori evitano di spingere l'inbreeding in condizioni così estreme perché, come vedremo, molti inconvenienti si producono negli animali, tra cui quello già visto della comparsa di caratteri recessivi indesiderati. In campo vegetale invece la pratica è possibile e anche più facile da realizzare, perché molte piante sono auto-impollinanti e l'inbreeding si realizza più in fretta. Famiglie di organismi sottoposte per generazioni a inbreeding possono arrivare a produrre delle linee pure, ossia individui tutti identici tra loro. Alcune linee di topi da laboratorio tendono a queste caratteristiche: animali geneticamente identici garantiscono una precisa misura degli effetti, per esempio, di trattamenti farmacologici diversi, in quanto vengono annullate le differenze dovute all'interazione tra i genotipi e il farmaco.

**Linebreeding:** Se è vero che un eccessivo inbreeding viene evitato dagli allevatori, è anche vero che spesso si vorrebbe fissare negli animali attuali le favorevoli caratteristiche di qualche famoso antenato riproduttore. Si vorrebbe cioè produrre dei soggetti che non abbiano le caratteristiche negative dovute all'elevata consanguineità, e che contemporaneamente siano molto simili al modello cui si aspira. Il tipo di accoppiamenti in questo caso viene indicato come linebreeding.

### FENOTIPI

I fenotipi che rileviamo sugli animali domestici devono essere interpretati da un punto di vista genetico e possono essere per così dire dissezionati in modo da coglierne la componente ereditaria come separata da quella non trasmissibile.

Per i caratteri quantitativi qualsiasi variabile venga misurata su un animale può essere definita come fenotipo (P). Il fenotipo è quindi il risultato della somma di due componenti: il genotipo (G) e l'ambiente (E) secondo l'equazione generale:  $P = G + E$

Sebbene non esista una netta demarcazione tra fenotipi o caratteri quantitativi e qualitativi, forse ricordare alcuni fatti distintivi. I caratteri quantitativi vengono di norma descritti usando un'unità di misura, per esempio cm, litri, kg, punti, percentuali ecc. I caratteri qualitativi vengono invece descritti da un aggettivo qualificativo e permettono l'assegnazione degli individui a classi o gruppi, per esempio presenza o assenza di corna, tipo di pezzatura, colore del mantello, gruppo sanguigno ecc.

I **caratteri quantitativi** presentano cioè una *variabilità continua*, mentre quelli qualitativi una *variabilità discontinua* o discreta. I caratteri quantitativi risentono largamente delle condizioni ambientali, mentre quelli qualitativi poco o niente. Mentre i caratteri qualitativi hanno un controllo genetico relativamente semplice (ma non sempre, come dimostra per esempio la complessità del controllo genetico del colore del mantello nei mammiferi), uno o pochi geni.

I **quantitativi** hanno un controllo complesso che viene descritto in base a un modello detto infinitesimale. Fondamentalmente, in base a questo modello, molti geni contribuiscono indipendentemente (in quanto dispersi su cromosomi diversi e quindi segreganti indipendentemente) a determinare la componente G.

Ciascun gene inoltre è supposto avere due forme alleliche, una che incrementa e l'altra che decrementa il carattere e gli effetti di questi geni sono singolarmente molto piccoli, infinitesimi appunto. Il numero di geni che controlla un carattere quantitativo non è mai noto, ma talvolta alcuni di questi geni mostrano un effetto non proprio infinitesimale. Si tratta di **geni detti a effetto maggiore** che, se opportunamente identificati e sfruttati, possono contribuire ad accelerare il processo selettivo.

- I caratteri qualitativi, con la loro eredità discreta, hanno contribuito in modo marcato alla differenziazione fenotipica tra le razze.
- I relativamente pochi geni implicati nella definizione del colore e della pezzatura del mantello, per esempio, hanno permesso la fissazione di alleli diversi nelle diverse razze, che si sono rese così molto uniformi e ben differenziate tra loro.

### Caratteri primari e secondari

La morfologia degli animali ha in questo contesto una valenza non univoca. Sicuramente per molti allevatori la vendita di animali da vita di ottima morfologia farebbe collocare questo carattere tra quelli primari (soprattutto animali da affezione come cani o cavalli, ma anche bovini e altre specie più classicamente zootecniche), ma nella maggior parte dei casi l'attenzione verso la morfologia punta sulle ripercussioni che questa ha sui fattori produttivi principali (morfologia funzionale).

È grazie a questi pochi geni con effetto visibile che oggi riconosciamo e distinguiamo le diverse razze. Sono geni che non hanno importanza dal punto di vista della produzione, ma che provocando, per esempio, l'esclusione di un soggetto non conforme dal Libro Genealogico, hanno comunque un impatto economico che non può essere ignorato

### SNP

APPLICAZIONI OMICHE ANIMALI: Da meno di 10 anni è disponibile una tecnologia genomica che permette la lettura simultanea di molte migliaia di SNP anonimi presenti nel genoma di molte specie animali (bovini, ovini, caprini, suini, cani, etc.) Utilizzando questo strumento è quindi possibile genotipizzare gli animali in una molteplicità di loci (3K, 24K, 50K, 150K, 800K). Gli SNP che vengono genotipizzati sono così «fitti» nel genoma da essere in LD a livello di popolazione con praticamente tutti i geni (sconosciuti) di importanza per l'allevamento e la sanità animale.

**Indice genomico:** Diventa così possibile studiare, nel concreto del genoma individuale, la deriva genetica, le distanze genetiche tra popolazioni, l'omozigosi e l'eterozigosi, le tracce lasciate dalla attività selettiva naturale e umana (selection signature), la presenza e l'effetto di geni con significato economico, l'attribuzione di paternità che abbiamo visto in queste lezioni. L'uso zootecnico più di impatto economico di questa tecnologia è sicuramente quello che ha prodotto l'indice genomico.

L'indice genomico per la selezione negli animali domestici è oggi disponibile solo per le specie più importanti in particolare i bovini. Esempio. Supponiamo di avere genotipizzato per uno SNP(A/T) 50 capre di cui conosciamo anche la produzione di latte in kg. I dati mostrano come la produzione ( $y = \text{kg di latte}$ ) cresca in funzione del numero di alleli A del marcatore ( $x = 0, 1, 2$ ). Vogliamo quindi stimare da questi dati l'effetto di sostituzione genica ( $\alpha$ ). Questo può essere stimato dalla regressione di  $y$  su  $x$  che deve essere però «pesata» per la frequenza ( $w$ ) che i tre genotipi manifestano nei dati

### ESERCIZIO dbSNP

- **rs9331896**
- **Position: 27610169**
- **Location: 27,596,917**
- **Lunghezza: position-location= 27610169 - 27,596,917= 13252**

### SELEZIONE GENOMICA

## Applicazioni OMICHE animali

Da meno di 10 anni è disponibile una tecnologia genomica che permette la lettura simultanea di molte migliaia di SNP anonimi presenti nel genoma di molte specie animali (bovini, ovini, caprini, suini, cani, etc.) Utilizzando questo strumento è quindi possibile genotipizzare gli animali in una molteplicità di loci (3K, 24K, 50K, 150K, 800K). Gli SNP che vengono genotipizzati sono così «fitti» nel genoma da essere in LD a livello di popolazione con praticamente tutti i geni (sconosciuti) di importanza per l'allevamento e la sanità animale.

Diventa così possibile studiare, nel concreto del genoma individuale, la deriva genetica, le distanze genetiche tra popolazioni, l'omozigosi e l'eterozigosi, le tracce lasciate dalla attività selettiva naturale e umana (selection signature), la presenza e l'effetto di geni con significato economico, l'attribuzione di paternità che abbiamo visto in queste lezioni. L'uso zootecnico più di impatto economico di questa tecnologia è sicuramente quello che ha prodotto l'indice genomico.

1. Come si fa il miglioramento genetico tradizionale
2. Dal pedigree tradizionale alla genomica

## MIGLIORAMENTO GENETICO

1. Miglioramento in termini genetici di una performance di una data popolazione da una generazione all'altra
2. Fenotipo = Genotipo + Ambiente
  - Fenotipo: performance osservata
  - Genotipo: geni che controllano l'espressione del carattere
  - Ambiente: tutto ciò che non è genetico (alimentazione, clima, zona, età, ...)

Come si fa il miglioramento genetico?

- Identificazione dell'obiettivo di selezione: Latte (produzione e/o qualità), Funzionalità, benessere
- Registrazione del dato: identificazione dell'animale, raccolta dati eventi e di performance

**BANCA DATI E LIBRO GENEALOGICO:** Ogni animale in Italia è iscritto in una Banca dati nazionale (BDN) deOa anche anagrafe zootecnica nazionale che fornisce ad ogni animale un identificativo «matricola». Se un animale appartiene ad una razza definita può richiedere di essere incluso nel libro genealogico

Ad esempio l'ANAFIBJ: Associazione che detiene i libri genealogici della Frisone Italiana

**Controlli funzionali** → La finalità dei Controlli Funzionali è quella di realizzare in modo sistematico il rilevamento, la registrazione, l'elaborazione, la pubblicazione e la divulgazione dei dati; tecnici necessari all'attività di incremento e miglioramento della produttività animale ed alla valorizzazione economica delle produzioni secondo norme stabilite e riconosciute a livello internazionale. L'esecuzione dei controlli funzionali consente il rilevamento di dati indispensabili ai fini della realizzazione dei programmi di miglioramento genetico che vengono imposti; e realizza; dalle singole Associazioni Nazionali di razza (ANA) che detengono l'Ufficio Centrale del Libro Genealogico (UCLG). Negli allevamenti ad indirizzo produttivo latte, mensilmente vengono misurate le produzioni di latte individuali, di ciascuna vacca o capra o bufala, prelevando un campione di latte per ciascun soggetto per l'analisi di grasso proteine e cellule somatiche. I dati raccolti con i controlli funzionali vengono elaborati dalle Associazioni di Razza e dall'AIA, ed agli allevatori vengono rese molte informazioni ed elaborazioni utili per la gestione dell'allevamento.

Obiettivi di una qualsiasi valutazione genetica

1. Stimare il valore genetico di ogni individuo della popolazione (EBV = Indice genetico)
2. Utilizzare l'indice per fare selezione (scegliere) e per accoppiare gli animali
3. Migliorare le produzioni (progresso genetico)

Requisito fondamentale: **Ereditabilità ( $h_2$ )** Quanto della variabilità che osservo è dovuta alla genetica?

Cerco nel mio chip gli alleli che portano un valore maggiore o minore dello standard

L'**ereditabilità  $h_2$**  che va da 0 a 1 indica quando il carattere è condizionato dalla genetica. Se è 1 è totalmente influenzato dalla genetica. Ci permette di dire se il carattere è ereditabile.

L'indice genetico: è lo strumento che permette ai tecnici ed agli allevatori di scegliere gli animali superiori geneticamente:

- I padri di toro
- Le madri di toro
- I tori da utilizzare in azienda

Calcolare gli indici = ripulire il dato misurato dagli effetti ambientali = stimare la componente GENETICA trasmissibile alla progenie  $I = bX = h_2X$

Si determina il valore di un toro aggiustando la produzione delle figlie entro allevamento per i maggiori fattori ambientali:

- Età
- Mese di parto
- Ordine di parto
- Intervallo parto concepimento

L'inserimento delle parentele: Un ulteriore perfezionamento del modello tiene conto della correlazione genetica che esiste fra individui discendenti dallo stesso progenitore

A	B	C	D	E	F
1	0	0	.5	.5	.5
0	1	0	.5	0	.25
0	0	1	0	.5	.25
.5	.5	0	1	.25	.625
.5	0	.5	.25	1	.625
.5	.25	.25	.625	.625	1.125

I primi passi verso la previsione del valore riproduttivo dalle informazioni sul genotipo sono stati compiuti nei bovini da latte, dove i marcatori del DNA dell'intero genoma sono stati utilizzati per stimare un "valore genomico" per i tori. Questo approccio è interessante in quanto prevedere un valore riproduttivo dal genoma significa che i migliori tori possono essere identificati e utilizzati per la riproduzione in età molto più giovane, quindi:

Con la selezione tradizionale si scelgono tori con figlie che fanno molto latte, passano 5 anni (intervallo di generazione). Adesso utilizziamo la selezione genomica dove in poco tempo riesco ad ottenere già l'indice genetico e genomico dell'individuo

- L'indice genetico: non prevede l'utilizzo di SNP
- L'indice genomico: la selezione utilizza gli SNP. Ogni SNP è un allele di interesse.

La fattibilità dell'utilizzo di informazioni genomiche, derivate dall'uso di sequenze adesso e di pannelli come quello di Illumina con 50.000 SNP, per prevedere il merito genetico dei tori è stata dimostrata nella razza Holstein. Questa applicazione delle informazioni genomiche ha avuto successo: con informazioni su un insieme di riferimento sufficientemente ampio di animali per i quali sono disponibili dati fenotipici e genotipici (per "addestrare" il sistema di modellazione) è possibile stimare i valori di riproduzione di altri animali privi di fenotipi con alta affidabilità per alcuni caratteri

#### LA RICOMBINAZIONE

Supponiamo che questi siano i due genomi di un riproduttore (supponiamo un toro). I due colori indicano la provenienza paterna e materna dei genomi.



Nella produzione di oociti e spermatozoi il riproduttore ricombina i suoi due genomi formando gameti (aploidi) tutti diversi tra loro ma che contengono porzioni complementari dei suoi due genomi originari.



Per ogni gamete di questo tipo ce ne sarà infatti un altro «complementare». Consideriamo ora i due punti (loci) indicati che chiamiamo A e B

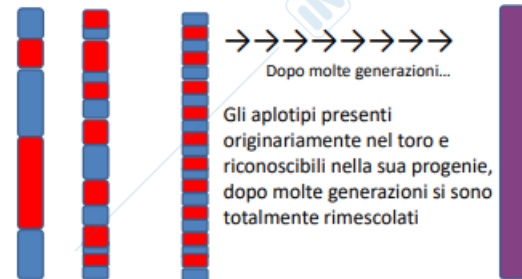


A e B sono detti in LINKAGE (sono vicini nello stesso cromosoma). Possiamo ora distinguere i due concetti di GENOTIPO e APLOTIPO. I pedici 1 e 2 indicano ALLELI diversi allo stesso locus.



Se A e B sono abbastanza vicini saranno prodotti solo gameti A1 – B1 oppure A2 – B2 in uguale proporzione (GAMETI (aplotipi) PARENTALI). Se la distanza tra A e B è più ampia potranno essere prodotti anche gameti A1 – B2 e A2 – B1 (GAMETI (aplotipi) RICOMBINANTI)

Ad ogni generazione la ricombinazione continua ad agire per cui col tempo i segmenti rossi e quelli blu si frammentano sempre di più determinando bande sottili finemente mescolate. Dopo un sufficiente numero di generazioni vedremo un misto di rosso e di blu a determinare un uniforme colore «violetto».

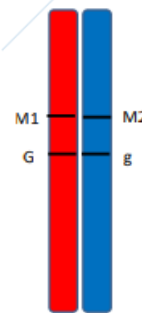


Supponiamo che A e B siano molto vicini. Nella progenie di un toro troveremo (quasi) solo le combinazioni parentali e (quasi) mai quelle ricombinanti. Si dice che tra i due loci vi è DISEQUILIBRIO DA LINKAGE (LD). Infatti, invece di trovare le 4 possibili combinazioni in uguale proporzione (¼ ciascuna):

**A1B1** Parentale  
**A1B2** Ricombinante  
**A2B1** Ricombinante  
**A2B2** Parentale

troviamo la prima, l'ultima e solo raramente quelle ricombinanti (questo è il disequilibrio). Se lasciamo passare molte generazioni (arriviamo al «genoma violetto») troveremo sicuramente tutte e 4 le combinazioni in perfetto EQUILIBRIO a causa delle continue ricombinazioni che si sono verificate nel corso delle generazioni (Linkage Equilibrium). I due loci sono sempre in Linkage, ma non vediamo più il disequilibrio.

Supponiamo che ad un certo locus l'allele G determini una produzione di +50 kg di latte, mentre l'allele g di -50 kg di latte. Un toro è doppio eterozigote al gene (Gg) e al vicino marcatore M1M2. Il gene è sconosciuto, ma il marcatore M è noto. Metà delle figlie del toro riceveranno M1 con G, metà M2 con g. La differenza tra le produzioni delle figlie M1M1 e quelle M2M2 sarà di 200 kg. Conclusione. Entro la famiglia del toro M e G sono in LD e posso utilizzare M per fare la selezione assistita da marcatori (MAS)



Ma a livello di popolazione M e G sono in LE quindi in un altro toro l'assetto potrebbe essere: **M1 – g/M2 – G** Conclusione: In una popolazione «outcrossing» il marcatore è inutilizzabile per una Marker-assisted Selection. A meno che marcatore e gene siano così vicini da essere in LD anche a livello di popolazione (marcatori genetici molto fitti)

## LINKAGE

Il linkage genetico descrive il modo in cui due geni che si trovano uno vicino all'altro su un cromosoma vengono spesso ereditati insieme. Nel 1905, William Bateson, Edith Rebecca Saunders e Reginald C. Punnett notarono che i tratti del colore dei fiori e della forma del polline nelle piante di pisello dolce sembravano essere collegati tra loro. Pochi anni dopo, nel 1911, Thomas Hunt Morgan, che studiava l'ereditarietà nei moscerini della frutta, notò che il colore degli occhi di una mosca era associato al sesso della mosca e ipotizzò che i due tratti fossero collegati tra loro.

Queste osservazioni hanno portato al concetto di linkage genetico, che descrive come due geni strettamente associati sullo stesso cromosoma vengano spesso ereditati insieme. Infatti, più due geni sono vicini tra loro su un cromosoma, maggiori sono le loro possibilità di essere ereditati insieme o collegati. Al contrario, i geni situati più distanti tra loro sullo stesso cromosoma hanno maggiori probabilità di essere separati durante la ricombinazione, il processo che ricombina il DNA durante la meiosi. La forza del legame tra due geni, quindi, dipende dalla distanza tra i geni sul cromosoma



**Linkage disequilibrium:** Se i geni sono presenti su cromosomi diversi, di solito vengono distribuiti alla progenie in modo casuale. Questo assortimento indipendente di alleli (forme di un gene) è ciò che di solito vediamo nella genetica mendeliana. Tuttavia, quando i geni si trovano uno vicino all'altro sullo stesso cromosoma, di solito si segregano insieme durante la meiosi. In questi casi, le combinazioni parentali di alleli su un cromosoma appariranno più frequentemente nella prole. I geni o gli alleli che segregano insieme sono collegati

**Linkage e distanze genetiche:** Il linkage è molto utile ai genetisti perché può dirci quali geni si trovano sullo stesso cromosoma e l'ordine di questi geni. L'ordine dei geni può essere determinato osservando il numero di ricombinanti che compaiono nella progenie. Se i geni si trovano molto distanti l'uno dall'altro su un cromosoma, a volte si verificano eventi di crossover (ricombinazione di alleli) tra di loro durante la meiosi. Più due geni sono distanti su un cromosoma, più eventi di crossover ci saranno. La quantità di ricombinazione non può dirci l'esatta distanza fisica, ma mostrerà la posizione relativa dei geni. La distanza tra geni o alleli è misurata in unità di mappa. Un'unità di mappa è definita come una frequenza di ricombinazione dell'1%.

**LINKAGE DISEQUILIBRIUM** è l'associazione non casuale di alleli in due o più loci.

Il linkage disequilibrium viene utilizzato come strumento per descrivere quando alcune combinazioni di marcatori genetici o alleli si verificano più o meno spesso in una popolazione di quanto ci si aspetterebbe se il processo fosse il risultato di un rimescolamento casuale, basato sulle frequenze di questi geni nella popolazione. Fondamentalmente, il linkage disequilibrium misura le associazioni non casuali tra polimorfismi in diversi loci.

Perché l'analisi del linkage disequilibrium è importante per la genetica? La risposta rapida è che i ricercatori possono cercare il linkage disequilibrium associato a marcatori molecolari per aiutarli a scoprire geni legati a particolari malattie o tratti.