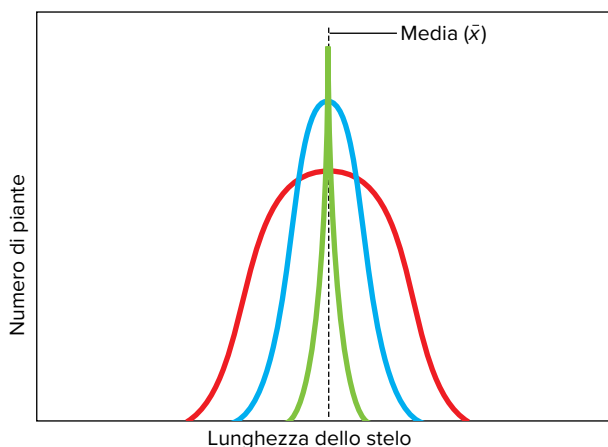


Soluzioni del Capitolo 24

24.1 a. 4; b. 5; c. 7; d. 9; e. 6; f. 1; g. 10; h. 2; i. 3; j. 8.

24.2 a. I denti di leone geneticamente identici ($VG = 0$) coltivati nell'ambiente controllato di una serra (curva verde nel diagramma che segue) varieranno nel fenotipo in misura minore rispetto alle piante geneticamente identiche coltivate in un ambiente variabile (vedi Figura 24.4a; curva azzurra) o alle piante geneticamente diverse coltivate in un ambiente costante (vedi Figura 24.4b; curva rossa). Nella figura che segue, l'ereditabilità di questo carattere nella popolazione è elevata, perché la varianza generata dalla diversità genetica (VG , curva rossa) è maggiore di quella prodotta dalla variabilità ambientale (VE , curva blu). Tuttavia, caratteri diversi in popolazioni diverse possono avere ereditabilità diverse. Se l'ereditabilità fosse bassa, la curva rossa in figura sarebbe invece più stretta della curva blu.



b. La varianza fenotipica non è nulla ($VP > 0$) perché anche in serra l'ambiente non è perfettamente uniforme; piante diverse sperimenteranno microambienti leggermente diversi e possono verificarsi eventi casuali che influenzano lo sviluppo e la crescita dello stelo.

c. Per affinare la stima di VG ottenuta dalla curva rossa in Figura 24.4b [o dalla curva rossa nel diagramma nella parte (a)], si deve sottrarre da questa stima la varianza osservata nell'esperimento descritto nella parte (a), rappresentato con la curva verde.

24.3 a. I cloni genetici hanno alleli identici di tutti i loro geni, come i gemelli identici (MZ). Tutta la variazione fenotipica osservata nei cloni genetici è dovuta agli effetti ambientali. Per ogni carattere specifico, $VG = 0$ e quindi $VP = VE$.

b. Tutta la varianza fenotipica nei gemelli MZ è dovuta a differenze ambientali ($VG = 0$; $VP = VE$). Le differenze fenotipiche nei gemelli DZ sono dovute sia all'ambiente sia ai geni ($VP = VE + VG$). È ragionevole presumere che gli ambienti in cui vengono allevate coppie di gemelli MZ e DZ siano ugualmente simili (i valori VE dei gemelli MZ e dei gemelli DZ sono equivalenti). Pertanto, i confronti di coppie di gemelli MZ e DZ (ogni coppia allevata nella stessa famiglia), possono essere utilizzati per eliminare VE e per stimare l'ereditabilità (VG/VP) di caratteri specifici.

Per un determinato carattere quantitativo, l'ereditabilità è pari al doppio della differenza nei coefficienti di correlazione tra coppie di gemelli MZ e DZ per quel carattere: $\text{ereditabilità} = 2(r_{MZ} - r_{DZ})$. Il Problema 6 esplora questa equazione in modo più dettagliato.

Per un carattere discreto, i valori di concordanza (la frequenza con cui il co-gemello mostra il carattere in questione quando l'altro gemello lo fa) delle coppie di gemelli MZ e DZ possono essere utilizzati per stimare l'ereditabilità:

$$\text{ereditabilità} = \frac{[MZ (\text{concordanza}) - DZ (\text{concordanza})]}{[1 - DZ (\text{concordanza})]}$$

c. Negli studi sull'ereditabilità dei caratteri negli animali, l'adozione incrociata aiuta a randomiz-

zare gli effetti dell'ambiente sulla prole di genitori diversi. L'adozione incrociata è utile perché le misurazioni dell'ereditabilità basate sulla somiglianza fenotipica tra genitori e prole presuppongono che tutti i genitori e la prole nella popolazione vivano nello stesso ambiente.

24.4 I valori di ereditabilità per un carattere raro dipendono dalla quantità totale di variazione fenotipica nella popolazione e da fattori ambientali che potrebbero influenzare il carattere. Inoltre, diverse popolazioni della stessa specie possono avere storie diverse, portando alla presenza di diversi alleli che influenzano il carattere.

24.5 a. Falso. Quando ereditabilità = 1, tutta la variazione fenotipica è dovuta ai geni e i gemelli MZ sono fenotipicamente due volte più simili rispetto ai gemelli DZ. Il motivo è che i gemelli MZ condividono tra loro il doppio degli alleli (100%) rispetto ai gemelli DZ (mediamente 50%). Mano a mano che l'ereditabilità diminuisce, le coppie di gemelli MZ mostrerebbero sempre più lo stesso grado di differenze fenotipiche delle coppie di gemelli DZ (confrontare i grafici per l'ereditabilità 1.0 e 0.0 in Figura 24.9 a e b).

b. Per lo più vero. È sempre vero che i gemelli MZ mostrano pochissime variazioni fenotipiche se l'ereditabilità è alta. Infatti, a meno che ereditabilità = 0, la variabilità fenotipica tra i gemelli DZ sarà sempre superiore a quella dei gemelli MZ (vedi Figura 24.9b). Tuttavia, quando l'ereditabilità è elevata, il grado medio di variazione tra i gemelli DZ dipende in gran parte dalla frequenza del carattere nella popolazione nel suo insieme. Se il carattere è raro, allora è del tutto possibile che un gemello DZ possa ereditare un allele raro che determina il carattere che l'altro gemello non eredita, quindi i gemelli in una coppia potrebbero variare considerevolmente nel fenotipo. Se il carattere è comune, le differenze tra i gemelli DZ saranno in media meno pronunciate.

c. Falso. L'ereditabilità può variare nelle diverse popolazioni per due motivi. Si ricordi che ereditabilità = $VG/VP = VG/(VG + VE)$. Ambienti diversi in Paesi diversi potrebbero influenzare la varianza fenotipica totale (VP) attraverso alterazioni di VE. Allo stesso modo, diversi pool genetici in Paesi diversi potrebbero alterare la varianza fenotipica totale attraverso VG.

24.6 a. Quando tutta la variazione fenotipica è dovuta ai geni, i gemelli MZ (condividono tutti i loro geni) avrebbero esattamente gli stessi valori del carattere, e quindi $r_{MZ} = 1$. Poiché i gemelli DZ condividono metà dei loro geni, $r_{DZ} = 0,5$. Inse-

rendo questi valori r nell'Equazione (24.8) si ha: ereditabilità = $2(1 - 0,5) = 1$. Un'ereditabilità di 1 significa che tutte le variazioni del valore del carattere sono dovute ai geni e nessuna all'ambiente.

b. Quando nessuna variazione fenotipica è dovuta ai geni, tutta la variazione è dovuta agli effetti ambientali. Le coppie di gemelli MZ e DZ condividono ciascuna gli stessi ambienti e supponiamo che gli effetti di questi ambienti siano simili nei due diversi tipi di gemelli. Pertanto il valore del carattere dovrebbe variare allo stesso modo nelle coppie gemelle MZ e DZ: $r_{MZ} = r_{DZ}$. Nella Figura 24.8b, le pendenze delle rette di regressione per i gemelli MZ e DZ sarebbero identiche. L'Equazione 24.8 diventa quindi: ereditabilità = $2(0) = 0$. Un'ereditabilità pari a 0 significa che tutte le variazioni del valore di quel carattere sono dovute all'ambiente e nessuna ai geni.

24.7 a. Le concordanze dei gemelli MZ sarebbero uguali a 1,0 se l'ereditabilità fosse uguale a 1,0. Una breve occhiata alla Tabella 24.2 rivela che tutti i valori di concordanza sono frazioni di 1,0.

b. L'Equazione 24.9 dovrebbe essere utilizzata per caratteri discreti, in cui vengono misurate le concordanze anziché i coefficienti di correlazione.

c. L'autismo ha la più alta ereditabilità ed è quindi controllato dai geni più di qualsiasi altro carattere nella Tabella 24.2 (almeno per questa popolazione). Le ereditabilità sono: diabete di tipo 1 (0,38); diabete di tipo 2 (0,21); Schizofrenia (0,38); Autismo (0,89); Alzheimer (0,26); Parkinson (0,056); Sclerosi multipla (0,21); Malattia di Crohn (0,37); Cancro colorettale (0,063); Cancro al seno (0,044); Cancro alla prostata (0,15).

24.8 a. Il comportamento docile è stato selezionato attraverso un esperimento di selezione del punto di troncamento, il che implica che i geni controllano il comportamento almeno in misura sostanziale (ereditabilità $\gg 0$). Se tutto il comportamento fosse dovuto all'ambiente e nessuno ai geni, allora ci aspetteremmo che il comportamento dei genitori non abbia alcun impatto sul valore del carattere nella progenie. Per stimare l'ereditabilità sarebbe necessario elaborare un grafico dei valori dei caratteri del genitore medio rispetto ai valori del carattere nella progenie.

b. Una spiegazione è che i geni che controllano il comportamento controllano anche i tratti morfologici associati alla docilità. Un'altra possibilità è che i ricercatori abbiano selezionato inconscia-

mente i tratti morfologici mentre stavano selezionando un comportamento docile. Gli animali con un aspetto associato alla cordialità che erano anche docili avrebbero potuto essere scelti per la riproduzione più frequentemente di animali altrettanto docili, ma che mancavano degli attributi fisici che gli esseri umani associano alla cordialità.

c. Le sequenze del genoma delle volpi addomesticate potrebbero essere confrontate con la sequenza di riferimento di *Vulpes vulpes* selvatiche per identificare le differenze di sequenza del DNA che potrebbero influenzare la funzione genica. Quei geni sarebbero i candidati per rappresentare i geni dell'addomesticamento.

- 24.9** La conoscenza della varianza fenotipica e del numero di loci che influenzano il fenotipo non fornisce informazioni sufficienti per rispondere a questa domanda. Si ricordi dall'Equazione (24.10) che la risposta di un carattere alla selezione (R) è determinata dalla proporzione della varianza fenotipica dovuta alla variazione genetica (cioè, all'ereditabilità, h^2) e dalla forza della selezione (S secondo la relazione $R = h^2 S$ [supponendo che la distribuzione dei fenotipi rientri in una curva normale a campana]). I parametri chiave che regolano la risposta alla selezione non sono la varianza fenotipica totale e il numero di loci, ma l'ereditabilità e la forza della selezione. Questi due set di parametri sono correlati ma non identici. La variazione genetica additiva h^2 è solo una componente della variazione fenotipica totale. Inoltre, una data ereditabilità in senso stretto incorpora le complessità di un numero variabile di loci maggiori e minori, ma la conoscenza del numero e della forza dei QTL coinvolti nei loro caratteri è insufficiente per misurare h^2 .

Le informazioni fornite nel problema non consentono di determinare i parametri chiave di h^2 e S necessari per risolverlo. Se si potesse stimare questi parametri, l'Equazione (24.10) chiarisce che il carattere con la più alta ereditabilità evolverà più velocemente, a parità di condizioni. Se h^2 è lo stesso per i due caratteri in questione e la forza della selezione è la stessa, anche la risposta alla selezione sarà la stessa.

- 24.10 a.** Se esistono n geni che contribuiscono in egual modo a un carattere complesso e ogni gene ha due alleli codominanti (o non completamente dominanti), il numero di valori diversi di quel carattere è $2n + 1$. Per esempio, quando $n = 1$, i genotipi AA , Aa , e aa hanno ciascuno fenotipi diversi. E quando $n = 2$, i cinque diversi valori del carattere sono spiegati dai seguenti genotipi: (1)

$AA BB$; (2) $Aa BB = AA Bb$; (3) $Aa Bb = aa BB = AA bb$; (4) $Aa bb = aa Bb$; (5) $aa bb$.

b. Dagli esempi forniti, la formula per determinare la frequenza dei fenotipi estremi nella F_2 di un incrocio è $(1/4)^n$, dove n = il numero di geni che determinano il fenotipo. Si può verificare che questa formula ha senso se si considera il caso più semplice in cui $n = 1$. Se l'incrocio parentale era tra AA e aa , allora F_1 è Aa e la progenie è $1/4 AA$; $1/2 Aa$; $1/4 aa$. I genotipi AA e aa corrispondono ai fenotipi estremi. Quando la progenie con i fenotipi estremi è presente ciascuna a una frequenza di $1/256$, allora $(1/4)^n = 1/256$ e $n = 4$. Pertanto, 4 diversi geni controllano il colore del nocciolo; i genotipi corrispondenti ai fenotipi estremi sono $AA BB CC DD$ e $aa bb cc dd$.

- 24.11 a.** Dal Problema 10, parte (a), sappiamo che, quando $n = 4$ come in questa domanda, esisteranno $2n + 1 = 9$ diversi fenotipi alla F_2 . Come nel Problema 10, parte (b), le lunghezze massime e minime delle foglie sono di tipo parentale: $AA BB CC DD$ (lunghezza = 32 cm) a una frequenza di $1/256$, e $A'A' B'B' C'C' D'D'$ (lunghezza = 16 cm) a una frequenza di $1/256$.

Le restanti 7 classi fenotipiche corrispondono a genotipi con 1 allele (') e 7 non primi, 2 alleli (') e 6 non primi, 3 alleli (') e 5 non primi ecc. I valori del carattere per ciascuna classe sono facili da determinare semplicemente sommando i contributi di ciascuno degli otto alleli. Per esempio, per la progenie F_2 con un solo allele (') e sette alleli non primi, la lunghezza della foglia è $1 (2) + 7 (4) = 30$ cm.

La frequenza di ciascun genotipo è più difficile da determinare. Un metodo che richiede molto tempo sarebbe quello di creare un quadrato di Punnett con 256 caselle (ci sono 16 possibili tipi di gameti prodotti da un quadruplo eterozigote e $16 \times 16 = 256$). Sarebbe comunque molto più efficiente usare il teorema binomiale, cioè:

P (che X si verifichi s volte e Y t volte, in n prove) = $[n!/(s! \times t!)] (p^s \times q^t)$

P = la probabilità che si verifichi ciò che è tra parentesi sul lato sinistro dell'equazione.

$$p = P(X)$$

$$q = P(Y)$$

X e Y sono le uniche due possibilità, quindi $p + q = 1$.

Inoltre, $s + t = n$.

Si ricordi che $!$ significa fattoriale: per esempio, $5! = 5 \times 4 \times 3 \times 2 \times 1$

A titolo di esempio, determiniamo il numero di genotipi diversi che contengono 2 alleli (') e 6 alleli non (') (per esempio, $A'A' BB CC DD$ o $AA' BB' CC DD$). Qui, X = alleli (') e Y = alleli non ('). Poiché gli alleli primi e non primi sono ciascuno il 50% di tutti gli alleli, la probabilità di X = la probabilità di $Y = 1/2$, e quindi $p = q = 0,5$. Per genotipi con 2 alleli primi e 6 non primi, $s = 2$, $t = 6$ e $n = 8$. Il $[n!/(s! \times t!)]$, il fattore nelle equazioni binomiali, fornisce il numero di genotipi diversi all'interno di ciascuna classe fenotipica specifica; in questo caso, $8!/(2! \times 6!) = 28$ diversi genotipi che hanno 2 alleli primi e 6 alleli non primi. Il fattore ($p^s \times q^t$) nell'equazione binomiale è la probabilità (frequenza) di ciascuno di questi diversi genotipi: $0,5^2 \times 0,5^6 = (0,5)^8 =$

$(1/2)^8 = 1/256$. Pertanto, la frequenza di progenie con 2 alleli primi e 6 non primi è $28/256$.

Per semplificare i calcoli si può notare che per tutti i possibili genotipi in tutte le classi fenotipiche si verifica sempre che: $p = q = 0,5$, $n = 8$ e $p^s \times q^t = (1/2)^8 = 1/256$; cioè, ciascuno dei 256 diversi possibili genotipi (quadrato di Punnett 16×16) è ugualmente probabile, perché i 16 possibili tipi di gameti sono ugualmente frequenti. L'unica parte del calcolo che differisce in ogni caso è $[n!/(s! \times t!)]$, che è il numero di genotipi diversi che generano ciascuna classe di valori del carattere.

I calcoli della frequenza e della lunghezza delle foglie per tutte le 9 classi fenotipiche della progenie sono mostrati nella tabella che segue. (Si ricordi che $0! = 1$):

Classe di valore del carattere	Numeri primi	Numeri non primi	Frequenza	Lunghezza della foglia
1	0	8	$[8! / (0! \times 8!)] / 256 = 1/256$	$0(4) + 8(2) = 16$ cm
2	1	7	$[8! / (1! \times 7!)] / 256 = 8/256$	$1(4) + 7(2) = 18$ cm
3	2	6	$[8! / (2! \times 6!)] / 256 = 28/256$	$2(4) + 6(2) = 20$ cm
4	3	5	$[8! / (3! \times 5!)] / 256 = 56/256$	$3(4) + 5(2) = 22$ cm
5	4	4	$[8! / (4! \times 4!)] / 256 = 70/256$	$4(4) + 4(2) = 24$ cm
6	5	3	$[8! / (5! \times 3!)] / 256 = 56/256$	$5(4) + 3(2) = 26$ cm
7	6	2	$[8! / (6! \times 2!)] / 256 = 28/256$	$6(4) + 2(2) = 28$ cm
8	7	1	$[8! / (7! \times 1!)] / 256 = 8/256$	$7(4) + 1(2) = 30$ cm
9	8	0	$[8! / (8! \times 0!)] / 256 = 1/256$	$8(4) + 0(2) = 32$ cm

b. È possibile determinare le frequenze genotipiche per ciascun gene utilizzando l'equazione di Hardy-Weinberg: $p^2 + 2pq + q^2 = 1$. Per il gene A , $p^2 = f(AA) = (0,9)^2 = 0,81$; $2pq = f(AA') = (2)(0,9)(0,1) = 0,18$; $q^2 = f(A'A') = (0,1)^2 = 0,01$. Per il gene B , $f(BB) = (0,9)^2 = 0,81$; $f(BB') = (2)(0,9)(0,1) = 0,18$; $f(B'B') = (0,1)^2 = 0,01$. Per il gene C , $f(CC) = (0,1)^2 = 0,01$; $f(CC') = (2)(0,1)(0,9) = 0,18$; $f(CC') = (0,9)^2 = 0,81$. Per il gene D , $f(DD) = (0,5)^2 = 0,25$; $f(DD') = (2)(0,5)(0,5) = 0,50$; $f(D'D') = (0,5)^2 = 0,25$.

c. Le piante con foglie lunghe 32 cm hanno il genotipo $AA BB CC DD$. Poiché i loci si assortiscono indipendentemente, è possibile trovare la frequenza delle piante con questo genotipo nella popolazione descritta nella parte (b) moltiplicando insieme le frequenze genotipiche rilevanti per ciascun gene: $(0,81) \times (0,81) \times (0,01) \times (0,25) = 0,0016$.

24.12 In tutti e tre i metodi, la genotipizzazione SNP viene utilizzata per trovare SNP associati a geni che controllano un carattere. Le tecniche differiscono in termini di tipi di individui genotipizzati, tempi degli eventi di ricombinazione che si sono

verificati per consentire la mappatura e modi in cui viene rilevata l'associazione.

Nel clonaggio posizionale dei geni di malattie mendeliane, per identificare i loci SNP che sono associati al gene della malattia la genotipizzazione SNP viene eseguita per membri affetti e non affetti della famiglia, in molti pedigree diversi. La mappatura dipende dai crossing-over recenti (durante le ultime generazioni all'interno del pedigree) che si sono verificati tra gli SNP associati e i geni della malattia; i ricercatori stimano la distanza in base alla frequenza di ricombinazione tra l'SNP e il gene della malattia.

Nella mappatura diretta dei QTL, la progenie BC_1 (mappatura approssimativa) o le linee quasi isogeniche (mappatura fine) con valori diversi del carattere sono genotipizzati per gli SNP per determinare se sono omozigoti o eterozigoti per SNP ricevuti da genitori con valori fenotipici ampiamente diversi. Gli SNP in cui omozigoti o eterozigoti hanno valori medi del carattere diversi sono associati a un QTL per quel carattere. Gli eventi di ricombinazione che consentono la mappatura si sono verificati di recente durante gli in-

croci effettuati per generare la progenie BC₁ o per generare le linee quasi isogeniche.

Nel GWAS, un gran numero di individui che condividono un particolare carattere e un gruppo di controllo che non lo presenta sono genotipizzati per gli SNP al fine di identificare SNP che si correlano statisticamente con il carattere. Qui, l'associazione di un SNP con un carattere complesso dipende dalla storia dei crossing-over all'interno della popolazione.

- 24.13** Sia nella mappatura approssimativa sia in quella fine, vengono misurati i valori fenotipici per correlare regioni specifiche del genoma con diversi valori del carattere. Entrambe le tecniche iniziano con un incrocio di due ceppi isogenici con valori di carattere molto diversi, e in entrambi i casi vengono generate piante che hanno genomi ricombinanti. I ricercatori genotipizzano gli SNP per determinare quali regioni genomiche hanno avuto origine da quale ceppo genitore isogenico e misurano anche i valori fenotipici nelle piante ricombinanti.

Le differenze tra i due metodi sono le seguenti: (1) le piante BC₁ utilizzate nella mappatura grossolana sono omozigoti o eterozigoti per ciascun allele del genitore del reincrocio, mentre le linee congeniche utilizzate nella mappatura fine sono omozigoti per tutti gli alleli di tutti i geni; (2) i blocchi di sequenze contigue provenienti dall'altro genitore (non utilizzato per il reincrocio) sono generalmente più piccoli nelle linee congeniche (dove sono chiamate introgressioni) rispetto alle piante BC₁.

Dopo che un gene candidato è stato identificato mediante la mappatura fine, i ricercatori devono convalidarne l'identità. Un metodo per la convalida consiste nel trasferire il gene sospetto da un ceppo con un certo valore del carattere a un ceppo con un valore del carattere molto diverso e determinare se il transgene cambia il fenotipo del ricevente nel modo previsto. Un altro approccio consiste nell'alterare il gene candidato in uno o in entrambi i ceppi (usando CRISPR/Cas9, per esempio) e cercare un effetto sul valore del carattere. Il successo di uno di questi approcci dipende dal contributo del gene candidato al valore del carattere (un QTL con un contributo maggiore è più facile da convalidare) e anche dalla natura degli alleli presenti in ciascun ceppo.

- 24.14** Il gene wild type *fw2.2* produce una proteina che è un regolatore negativo della crescita. La specie di pomodoro piccolo (*S. pennellii*) ha alleli *fw2.2* di tipo selvatico e la specie di pomodoro grande

(*S. lycopersicum*) ha un allele *fw2.2* a perdita di funzione. Poiché il prodotto del gene *fw2.2* ha una forte influenza sul valore del carattere, il trasferimento dell'allele *fw2.2* wild-type di *S. pennellii* in *S. lycopersicum* ha avuto un forte effetto dominante, diminuendo significativamente la dimensione media del pomodoro di *S. lycopersicum*. L'esperimento inverso, il trasferimento dell'allele *S. lycopersicum fw2.2* in *S. pennellii*, avrebbe un effetto scarso o nullo. L'allele *fw2.2* di *S. lycopersicum* ha poca o nessuna funzione e *S. pennellii* è già omozigote per gli alleli *fw2.2* wild-type.

- 24.15** Le persone con cardiopatia a esordio precoce dovrebbero essere utilizzate per il GWAS perché è probabile che abbiano più alleli patogenici per determinati QTL e/o QTL associati a valori del carattere più elevati rispetto a quelle con malattia a esordio tardivo. La popolazione di controllo dovrebbe essere costituita da individui della stessa età del gruppo sperimentale che non hanno malattie cardiache né una storia familiare di malattie cardiache e che vivono nella stessa gamma di ambienti del gruppo con malattie cardiache.

24.16 a. Come spiegato nel Capitolo 23, l'effetto del fondatore potrebbe spiegare l'incidenza relativamente alta di alcuni disturbi recessivi e l'incidenza relativamente bassa di altri. Se i finlandesi di oggi discendono quasi esclusivamente da un piccolo numero di persone, gli alleli introdotti da quei particolari fondatori sono ora quelli prevalenti. Inoltre, la frequenza degli alleli di vari geni potrebbe cambiare rapidamente per deriva genetica se la popolazione rimanesse piccola. In effetti, gli studi hanno suggerito che la popolazione fondatrice della Finlandia, circa 4000 anni fa, era piccola, intorno a 3000 individui, e che nel XII secolo (circa 900 anni fa) la popolazione non superava le 50 000 unità.

b. Le popolazioni consanguinee come i finlandesi hanno un numero relativamente piccolo di QTL per un dato carattere, come la schizofrenia, rispetto ad altre popolazioni. Di conseguenza, ciascuno di questi QTL potrebbe apportare un grande contributo al fenotipo; la forte influenza di questi QTL sui valori dei caratteri li rende più facili da identificare. D'altra parte, la scarsa variabilità nella popolazione finlandese implica che alcuni fenotipi non esisteranno affatto nella popolazione e non possono essere studiati. Un altro svantaggio è che alcuni QTL che contribuiscono a determinare il carattere in altre popolazioni non possono essere trovati esaminando i finlandesi.

- 24.17 a.** SNP3, SNP4, SNP5, SNP6 e SNP7 sono in forte linkage disequilibrium con il locus della malat-

tia. L'allele G di SNP3, l'allele T di SNP4, l'allele T di SNP5, l'allele T di SNP6 e l'allele C di SNP7 sono presenti nei genomi dei pazienti ciascuno con una frequenza molto più elevata rispetto a quella dei genomi di controllo.

b. Il gene della malattia di Canavan è molto probabilmente all'interno della regione genomica di circa 400 kb delimitata da SNP3 e SNP7.

c. I dati suggeriscono che nella storia della popolazione esaminata si sono verificate due mutazioni indipendenti del gene della malattia di Canavan. Gli individui affetti 1-4 hanno tutti l'aplotipo GTTTC per SNP3–SNP7; quindi una mutazione patogena nel gene della malattia di Canavan deve essersi verificata in un cromosoma contenente questo aplotipo. L'individuo affetto 5 ha l'aplotipo GCCTG, per cui una mutazione diversa deve essersi verificata in un cromosoma con questo insieme di alleli SNP. Si noti che l'individuo sano 6 (controllo) ha l'aplotipo GTTTC, ma non la malattia. La spiegazione più probabile è che l'individuo 1 sia eterozigote per l'allele della malattia. Gli altri tre individui di controllo (7-9) non hanno nessuno degli aplotipi della malattia.

d. Gli ebrei sefarditi e ashkenaziti condividono l'aplotipo del gene della malattia GTTTC. Pertanto, la mutazione del gene della malattia di Canavan si è verificata su un cromosoma GTTTC prima della separazione sefardita/ashkenazita. I loci SNP3-7 di ulteriori ebrei sefarditi dovrebbero essere genotipizzati per trarre conclusioni sulle origini dell'aplotipo della malattia GCCTG.

e. Quando si studia l'associazione tra aplotipi e una malattia, concentrarsi su una piccola popolazione in cui gli individui affetti sono frequenti presenta due vantaggi: (1) più individui colpiti significa più dati rilevanti, con aumento della probabilità che un'associazione raggiunga significatività statistica; (2) è probabile che gli individui affetti presentino la stessa mutazione, ricevuta da un antenato comune, e quindi lo stesso aplotipo, rendendo l'associazione più evidente.

Concentrarsi su una piccola sottopopolazione presenta due svantaggi: (1) se il carattere è complesso, alcuni loci che possono contribuire al carattere in altre popolazioni non verranno rilevati [vedi anche Problema 16, parte (b)]; (2) se la sottopopolazione si è formata di recente, le regioni di linkage disequilibrium (LD) possono essere molto grandi perché è trascorso meno tempo (meno generazioni) in cui avrebbe potuto verificarsi una ricombinazione. Regioni più grandi di LD rendono più difficile individuare specifici

SNP (aplotipi specifici) associati al gene della malattia, e quindi sarebbe difficile identificare il gene specifico che contribuisce al carattere.

f. La difficoltà è che non si conosce la fase degli alleli di due SNP in un doppio eterozigote. Per determinare particolari aplotipi, i ricercatori dovrebbero esaminare i genotipi dei parenti stretti (genitori, fratelli e figli) degli individui nella tabella. Se un individuo è omozigote per un particolare aplotipo, non è necessario alcun confronto, ma ovviamente non sarà sempre così.

Come esempio molto semplice di questo approccio, supponiamo che un individuo sia eterozigote per gli alleli T e C sia a SNP4 sia a SNP5 [4:(T/C) 5:(T/C)]. Se il genotipo della madre era [4:(C/C) 5:(C/C)] e l'allele del padre era [4:(T/T) 5:(T/T)], allora gli aplotipi nei loro figli devono essere 4:C 5:C/4:T 5:T. Se la genotipizzazione SNP dei genitori non risolve completamente l'aplotipo, utilizzando una logica simile, l'analisi dei genotipi dei fratelli può aiutare.

Esiste un metodo diverso e meno preciso per suggerire gli aplotipi di almeno alcuni individui nel gruppo di pazienti nella regione del gene della malattia. Abbiamo visto nella parte (a) che l'allele T di SNP4 e l'allele T di SNP5 sono in linkage disequilibrium con il gene della malattia, e infatti sono in LD tra loro nel gruppo di pazienti. Una tale scoperta suggerirebbe che molti individui nel gruppo di pazienti potrebbero avere un aplotipo 4:T 5:T.

24.18 a. Poiché ciascuno dei 3 SNP ha due alleli: $2^3 = 8$ diverse configurazioni possibili.

b. $2^{20} = 1\,048\,576$ diverse configurazioni possibili.

c. No, l'eterozigosi non interferisce con l'identificazione degli aplotipi utilizzando gli SNPs Tag. In effetti, gli SNPs Tag sono stati scelti per il posizionamento su microarray in modo da prendere in considerazione la diploidia. La tabella seguente mostra che si otterrà un risultato unico per ogni possibile genotipo che un individuo possa avere:

Aplotipi	SNP 4	SNP 8	SNP 15
1/1	A	T	C
1/2	A	TC	CG
1/3	AG	T	CG
1/4	A	TC	C
2/2	A	C	G
2/3	AG	TC	CG
2/4	A	C	CG
3/3	G	T	C
3/4	AG	TC	C
4/4	A	C	C

Da notare che questa parte del problema richiede di sapere che esistono solo 4 possibili aplotipi e quali alleli sono associati a ciascun aplotipo. La risposta alla parte (f) del Problema 17 spiega alcuni possibili approcci al problema della determinazione degli aplotipi negli individui diploidi.

d. No; l'insieme di tre SNP non è l'unico che potrebbe essere usato come Tag. Per esempio, SNP1, SNP6 e SNP17 costituirebbero un altro insieme praticabile di SNP Tag per questo blocco di aplotipi:

Aplotipi	SNP 1	SNP 6	SNP 17
1/1	C	A	G
1/2	CT	AT	GA
1/3	C	AT	G
1/4	CT	AT	G
2/2	T	T	A
2/3	CT	T	AG
2/4	T	T	AG
3/3	C	T	G
3/4	CT	T	G
4/4	T	T	G

24.19 a. Alcuni cromosomi nel gruppo della malattia in Figura 24.14 non presentano la variante rossa che causa la malattia perché questa malattia è un carattere complesso e la variante rossa è un singolo QTL. Molti altri geni (QTL) contribuiscono a questa malattia. Pertanto, una persona senza questa particolare variante può avere la malattia a condizione che abbia alleli di altri QTL che possono contribuire al carattere.

b. Alcuni cromosomi nel gruppo di controllo presentano la variante rossa perché altri QTL possono contribuire al fenotipo. Una possibilità è che la variante rossa da sola non sia sufficiente a causare la malattia, ma che invece gli individui affetti abbiano specifici alleli di diversi QTL che insieme contribuiscono al fenotipo. Un'altra possibilità è che la sola variante rossa possa causare la malattia nella maggior parte delle persone, ma altre persone hanno un allele di un QTL diverso che sopprime gli effetti della variante rossa.

c. Gli scienziati cercherebbero alleli di loci anonimi (di solito SNP) le cui frequenze nel gruppo colpito sono maggiori delle loro frequenze nel gruppo di controllo. Nell'esempio di Figura 24.14, una mutazione che ha causato la malattia si è verificata originariamente sul cromosoma blu. Pertanto, gli alleli SNP che erano (I) presenti sul cromosoma blu prima della mutazione che causa la

malattia e (II) strettamente associati alla mutazione saranno in linkage disequilibrium con la mutazione.

d. È probabile che la mutazione che causa la malattia nella regione 1 si sia verificata prima della mutazione nella regione 2. La regione più corta di LD nella regione 1 significa che si è verificata una maggiore ricombinazione tra il cromosoma in cui è stata introdotta originariamente la mutazione della malattia e altri cromosomi nella popolazione che hanno varianti diverse di loci associati. Pertanto, deve essere trascorso più tempo dall'evento della mutazione originale nella regione 1 rispetto all'evento di mutazione della regione 2.

e. Poiché la ricombinazione non è uniforme nel genoma umano (esistono hotspot), la lunghezza di una regione di un cromosoma contenente SNP in LD con un allele della malattia non è un indicatore perfetto di quanto tempo fa si sia verificata la mutazione della malattia. Per esempio, supponiamo che due diversi alleli della malattia vengano introdotti in due diversi cromosomi durante lo stesso anno; uno dei geni è affiancato da hotspot di ricombinazione e l'altro no. La lunghezza della regione degli SNP associati alla malattia sarà molto più breve per il gene della malattia affiancato da hotspot di ricombinazione rispetto al gene della malattia che non lo è, anche se entrambe le mutazioni sono state introdotte nello stesso periodo storico.

24.20 a. Tra i due blocchi LD si trovano uno o più hotspot di ricombinazione.

b. La genotipizzazione limitata ha un valore predittivo a causa dell'esistenza di blocchi LD. Per esempio, un allele di SNP3 prevede gli alleli che si troverebbero in SNP1, SNP2, SNP4 e SNP5, mentre gli alleli di SNP7 predicono SNP8 e SNP9. Va notato che le previsioni fatte in questo modo non sono assolute, perché all'interno dei blocchi LD possono ancora verificarsi rari eventi di ricombinazione.

c. Gli SNP comuni sono tipicamente usati per la genotipizzazione perché per definizione sono quelli che più probabilmente si trovano in qualsiasi popolazione. Se la genotipizzazione viene eseguita con una tecnica che campiona solo un sottoinsieme di tutti i nucleotidi nel genoma (come nel caso dell'analisi di microarray, per esempio), non sarebbe conveniente esaminare gli SNP che sono invarianti nella maggior parte degli individui. Inoltre, se un allele SNP è estremamente raro, potrebbe non essere stato ancora trovato

perché nessun genoma che reca tale SNP è stato ancora caratterizzato.

Un ulteriore motivo per concentrarsi sugli SNP comuni per la creazione di mappe LD è che è probabile che molti SNP rari rappresentino mutazioni avvenute molto di recente. (Se è così, la mutazione non potrebbe essersi diffusa a molti individui nella popolazione odierna perché sono trascorse solo poche generazioni.) Se uno SNP è stato indotto di recente, è probabile che rimanga in linkage disequilibrium con loci anche distanti sullo stesso cromosoma. Pertanto i dati aggiunti, includendo SNP rari, non sarebbero in grado di risolvere piccoli blocchi di LD che indicherebbero le posizioni degli hotspot di ricombinazione.

d. Il motivo principale per cui i genetisti di popolazioni preferiscono avere a disposizione la sequenza completa dei genomi è che queste sequenze includeranno il polimorfismo effettivamente responsabile di un QTL. La mappatura dei QTL sulla base della loro associazione con polimorfismi anonimi in altri loci fornisce solo una posizione approssimativa del QTL. Quando è disponibile l'intera sequenza del genoma, i ricercatori sono in grado di identificare il polimorfismo responsabile e questa conoscenza può potenzialmente fornire informazioni importanti sulle basi biologiche della malattia.

Inoltre, gli SNP a frequenza inferiore, che sfuggono agli studi di genotipizzazione classici ma sono identificati dal sequenziamento completo del genoma, possono fornire ulteriori indizi sulla discendenza delle popolazioni. Se due individui hanno lo stesso SNP raro, devono aver condiviso un antenato comune recente.

24.21 a. Non tutti gli individui malati hanno lo stesso genotipo SNP perché la malattia è un carattere complesso. Come spiegato nelle parti (a) e (b) del Problema 19, il QTL per la malattia rilevata da SNP1 potrebbe dare solo un piccolo contributo al fenotipo della patologia. Inoltre, è altamente improbabile che SNP1 stesso sia il QTL (la mutazione che causa la malattia); invece, SNP1 è strettamente associato alla mutazione. Pertanto, in alcuni individui, la particolare variante SNP1 associata alla mutazione che contribuisce alla malattia potrebbe aver ricombinato e non essere più associata nei cromosomi degli individui odierni che mostrano il fenotipo della malattia.

b. Si ricordi che la Figura 24.17 mostra il conteggio degli alleli, non degli individui. Poiché ogni

individuo ha 2 alleli per ogni locus, il numero di alleli è il doppio del numero di persone. Pertanto, $3848/2 = 1924$ individui erano i Casi e $5872/2 = 2936$ individui erano i Controlli.

24.22 a. In generale, maggiore è la frequenza di un allele SNP nella popolazione, maggiore può essere il valore p per la sua associazione con un carattere. Il valore p per gli alleli SNP ad alta frequenza associati a un carattere dipenderà dal valore del carattere del locus contrassegnato dall'allele SNP. Se un allele SNP è raro, anche se contrassegna un locus con un valore elevato del carattere, può avere un valore p inferiore alla soglia di significatività perché poche persone con il carattere porteranno quell'allele SNP.

b. Quando si confrontano i loci SNP con frequenze simili nella popolazione, maggiore è il valore del carattere, maggiore è il valore p . Tuttavia, gli alleli SNP rari con valori elevati del carattere possono avere valori p inferiori rispetto agli alleli SNP comuni con valori bassi del carattere.

c. Maggiore è l'estensione di LD tra un allele SNP e il locus che contrassegna, maggiore è il valore p . Il motivo è che il particolare allele SNP è associato al carattere più spesso se il LD è alto che se il LD è basso.

24.23 a. Gli alleli SNP possono essere rari o perché sono deleteri (per esempio, gli alleli di una malattia) o perché sono mutazioni recenti che non hanno ancora avuto il tempo di diffondersi nella popolazione.

b. Gli alleli SNP sono comuni quando sono antichi (hanno avuto il tempo di diffondersi in tutta la popolazione) e quando sono benefici o neutri.

24.24 a. L'allele C dell'SNP sembra essere correlato all'ipertensione: l'allele C appare più frequentemente nei genomi delle persone con la malattia ($1025/1927 = 53,2\%$ dei Casi) rispetto ai genomi delle persone senza di essa ($725/1647 = 44,0\%$ dei Controlli). Per determinare se la correlazione è statisticamente significativa, dobbiamo eseguire un test del chi quadro per l'indipendenza. Possiamo testare i dati contro l'ipotesi nulla che le distribuzioni di C e T nei casi e nei controlli siano le stesse; in altre parole, l'ipotesi nulla è che C e T, e Casi e Controlli, si associno tra loro in modo casuale: la presenza di un particolare allele SNP e l'ipertensione sono indipendenti.

Utilizzare i passaggi descritti nel riquadro Gli strumenti della genetica intitolato "Il test del chi quadro di indipendenza" per impostare una tabella di contingenza per il set di dati:

	Casi	Controlli	
C	O = 1927	O = 725	C totali = 1750
	$E = [(1750/3574) \times 1927] = 943,6$ $\chi^2 = [(1025 - 943,6)^2 / 943,6] = \mathbf{7,02}$	$E = [(1750/3574) \times 1647] = 806,4$ $\chi^2 = [(725 - 806,4)^2 / 806,4] = \mathbf{8,22}$	
T	O = 902	O = 922	T totali = 1824
	$E = [(1824/3574) \times 1927] = 983,4$ $\chi^2 = [(902 - 983,4)^2 / 983,4] = \mathbf{6,74}$	$E = [(1824/3574) \times 1647] = 840,6$ $\chi^2 = [(922 - 840,6)^2 / 840,6] = \mathbf{7,88}$	
	Casi totali = 1927	Controlli totali = 1647	Genomi totali = 3574

Il valore totale di χ^2 è $(7,02 + 8,22 + 6,74 + 7,88) = 29,86$. Un valore χ^2 di 29,86 con $df = 1$ dà $p = 5 \times 10^{-8}$; la probabilità che l'associazione osservata tra SNP C e ipertensione sia dovuta a un errore di campionamento casuale è solo 0,00000005, o 1/20 000 000. Se questo fosse l'unico SNP analizzato, i risultati indicherebbero un'associazione altamente significativa.

b. Nel contesto di un GWAS, in cui lo SNP nella parte (a) era 1 su 1 000 000 di SNP analizzati, il valore p di 5×10^{-8} pone l'SNP al limite della significatività. Come descritto nel testo, $p < 5 \times 10^{-8}$ è il limite di significatività quando viene analizzato 1 000 000 di SNP in un GWAS (perché $0,05/1\,000\,000 = 5 \times 10^{-8}$). Se invece il cutoff fosse $p < 0,05$ (come sarebbe appropriato con un singolo SNP), allora circa 1/20 del 1 000 000 di SNP analizzati cioè 50 000 SNP) risulterebbe casualmente associato al carattere per errore di campionamento anche se nessuno di essi fosse effettivamente associato al QTL responsabile del carattere.

c. All'allele C è associato all'ipertensione. Come affermato nella parte (a), C compare più frequentemente nei genomi delle persone con la condizione ($1025/1927 = 53,2\%$ dei Casi) che nei genomi delle persone senza di essa ($725/1647 = 44,0\%$ dei Controlli). Come discusso nelle parti (a) e (b), queste differenze sarebbero considerate altamente significative se questo fosse l'unico SNP analizzato, ma solo marginalmente significative se questi dati fossero ottenuti in un esperimento in cui vengono considerati 1 000 000 di SNP.

d. L'odds ratio allelico è l'aumento del rischio di avere l'ipertensione conferito dal possedere l'allele C dell'SNP. Il rapporto è $[(\text{Casi con C/Casi con T})/(\text{Controlli con C/Controlli con T})]$, che è $[(1025/725)/(902/922)] = 1,41/0,978 = 1,44$. L'odds ratio allelico di 1,44 significa che un individuo con un allele C dell'SNP ha 1,44 volte più probabilità di avere la pressione alta rispetto a un individuo senza un allele C.

24.25 (I) Il secondo studio (in blu) potrebbe aver coinvolto meno casi e controlli rispetto al primo studio (in rosso). Per ottenere valori di p molto bassi, è necessario genotipizzare un numero molto elevato di individui. (II) Le due popolazioni sono diverse. A seconda del background genetico, il QTL evidenziato dallo SNP *rs1333049* può avere effetti diversi sul fenotipo in studio.

Si noti che la figura non chiarisce se il valore soglia di $-\log_{10} [\text{valore di } p] = 7$ si riferisce a uno o entrambi gli studi. Se questo rappresenta il limite per entrambi gli studi, l'implicazione è che entrambi gli studi hanno preso in considerazione lo stesso numero di SNPs ($= 500\,000$). In questo caso, l'associazione dello SNP *rs1333049* con la malattia coronarica nello studio blu non è significativa. Tuttavia, se lo studio blu esaminasse un minor numero di SNP, il limite di significatività sarebbe inferiore ed è possibile che l'associazione di questo SNP con la condizione sarebbe significativa in entrambi gli studi.

24.26 Una possibilità è che l'SNP non sia comune e quindi o non sia stato testato in esperimenti precedenti oppure la sua associazione sia risultata non significativa dato che pochi Casi o Controlli possiedono l'allele. Un'altra possibilità è che *TBK1* possa avere un basso valore di impatto sul carattere (cioè un'influenza relativamente piccola sul fenotipo SLA) e quindi l'associazione nel GWAS potrebbe essere stata insignificante. I valori dei tratti per *TBK1* potrebbero anche essere abbastanza diversi nelle popolazioni esaminate in diversi studi.

24.27 È probabile che il gene *FTO* interagisca con l'ambiente. Per esempio, la dieta media delle persone negli Stati Uniti potrebbe essere cambiata in modo significativo dopo il 1945, rendendo, solo dopo tale data, la variante di *FTO* molto importante per mantenere un peso adeguato.

24.28 I GWAS in generale ricercano l'associazione tra SNP comuni (frequenti) e altezza, quindi l'eredità

tabilità mancante potrebbe essere spiegata da varianti più rare causate da mutazioni più recenti. Questa idea è particolarmente interessante per spiegare la situazione umana a causa della recente esplosione demografica dall'inizio della Rivoluzione Industriale (vedi Figura 24.2). Questa ipotesi potrebbe essere verificata sequenziando l'intero genoma di molti più esseri umani di quanti ne siano stati sequenziati fino a oggi, rilevando contemporaneamente l'altezza di questi individui. È quindi possibile utilizzare i dati generati per analizzare l'associazione con l'altezza di eventuali nuovi alleli SNP rari trovati. I dati potrebbero anche essere usati per inferenza, eseguendo un GWAS in cui si aggiungono nei microarray i rari SNP trovati (in altre parole, si potrebbe genotipizzare ancora più persone per questi SNP, oltre il numero di individui con genomi completamente sequenziati).

24.29 Identificare un gene associato a un particolare carattere complesso può essere difficile (vedi Figura 24.20 come esempio) e non esiste una risposta semplice a questa domanda. Tuttavia, in generale, si può esaminare il blocco LD per i geni le cui funzioni dedotte suggeriscono una connessione con la preferenza. Si dovrebbero quindi confrontare le sequenze dei geni candidati negli individui che preferiscono il cioccolato con quelle di coloro che preferiscono la vaniglia.

24.30 a. I genotipi SNP possono indicare se si possiede un allele per una malattia genetica mendeliana solo se il gene e le mutazioni sono note, così come il modello di eredità del carattere (l'allele mutante è dominante o recessivo?) e se il gene della malattia presenta una penetranza del 100%. (Il termine carattere mendeliano viene solitamente applicato solo se tutte queste ipotesi sono vere. Molti caratteri mendeliani sono caratteri a singolo locus come la fibrosi cistica o la malattia di Huntington, sebbene alcuni genetisti chiamino carattere anche quello che può essere causato da alleli mutanti in uno qualsiasi di alcuni geni mendeliani, purché la presenza di un allele mu-

tante in uno qualsiasi dei geni sia completamente penetrante.)

b. Per una malattia che rappresenta un carattere complesso, la risposta dipende dal numero di QTL noti per essere associati alla malattia e dai valori d'impatto di ciascuno di questi QTL sul carattere. Una malattia si potrebbe sviluppare solo se si possiedono diversi alleli QTL con valori di carattere elevati. Per la maggior parte delle malattie che sono caratteri complessi, la somma totale dei contributi dei QTL identificati fino a oggi è troppo bassa per avere un grande valore predittivo.

c. In primo luogo, le società potrebbero aver esaminato insieme di SNP diversi. In secondo luogo, potrebbero aver condotto i loro studi su popolazioni diverse. In terzo luogo, i modelli utilizzati dalle aziende per spiegare la base genetica del carattere potrebbero essere stati diversi e tutti questi modelli sono quasi certamente basati su una conoscenza incompleta (per esempio, molti QTL per il carattere probabilmente non sono stati ancora identificati).

d. L'accuratezza delle stime del rischio per i caratteri complessi migliorerà senza dubbio con il tempo, ma rimarrà tutt'altro che perfetta per il prossimo futuro. Il problema dell'ereditabilità mancante potrebbe essere risolto in parte dal sequenziamento dell'intero genoma e con GWAS aggiuntivi concentrati su SNP molto rari, come esposto in dettaglio nella risposta al Problema 30. Tuttavia, alcuni SNP rari nella popolazione umana non possono essere trovati, a meno che i ricercatori non sequenziano l'intero genoma di ogni persona sulla terra. Un ulteriore problema fondamentale è che l'ereditabilità di quasi tutti i caratteri complessi è inferiore a 1,0, perché anche l'ambiente e gli eventi casuali possono influenzare queste caratteristiche. Quindi, anche se si conoscessero tutti i nucleotidi in ogni essere umano e i contributi relativi di ogni QTL coinvolto nel carattere, si dovrebbe tenere conto del modo in cui il genoma di ogni persona interagisce con l'ambiente in cui vive.