

# Soluzioni del Capitolo 9

**9.1** a. 10; b. 12; c. 9; d. 7; e. 11; f. 2; g. 8; h. 5; i. 1; j. 4; k. 6; l. 3.

**9.2 a.** *Sau3A* ( $\wedge$ GATC) (I) I siti di riconoscimento *Sau3A* sono lunghi 4 paia di basi (bp) e dovrebbero verificarsi casualmente ogni 44 o 256 bp. Il genoma umano contiene circa  $3 \times 10^9$  bp, ci si aspetterebbe  $3 \times 10^9 / 256 = 1,2 \times 10^7 \sim 12\,000\,000$  frammenti. (II) L'enzima taglierebbe in media una volta ogni 44 o 256 paia di basi, quindi questa dovrebbe essere la dimensione media dei frammenti del genoma umano tagliato con questo enzima di restrizione. (III) Le estremità prodotte dalla digestione con questo enzima hanno estremità appiccicose con una sporgenza al 5'. Si noti che il sito di taglio ( $\wedge$ ) non è nel mezzo della sequenza di riconoscimento, il che significa che l'enzima taglia i due filamenti del DNA bersaglio su coppie di basi diverse, per cui non si produce un'estremità netta. Letto in direzione 5'-3', il sito del taglio è a sinistra del centro della sequenza di riconoscimento, quindi si genera una sporgenza al 5'. (IV) Tutte le estremità dei frammenti di DNA umano prodotti dalla digestione con *Sau3A* sarebbero identiche.

**b.** *BamHI* ( $G\wedge$ GATCC) (I) L'enzima taglia il DNA casualmente una volta ogni 46 o 4.096 bp.  $3 \times 10^9 / 4096 = \sim 732\,400$  frammenti. (II) 4096 coppie di basi =  $\sim 4,1$  kilobasi (kb). (III) Coesive, terminano con una sporgenza al 5'. (IV) Tutte le estremità sono identiche (sono tutte sporgenze 5' GATC).

**c.** *HpaII* ( $C\wedge$ CGG) (I)  $3 \times 10^9 / 256 = \sim 12\,000\,000$  di frammenti. (II) 256 coppie di basi. (III) coesive, terminano con una sequenza sporgente al 5'. (IV) Tutte le estremità sono identiche (sporgenze 5'CG).

**d.** *SphI* (GCATG $\wedge$ C) (I)  $3 \times 10^9 / 4096 = \sim 732\,400$  frammenti. (II) 4096 coppie di basi. (III) Coesive, terminano con una sequenza sporgente al 3'. Il sito tagliato non è nel mezzo della sequenza di

riconoscimento, quindi questa non è un'estremità netta. Tuttavia, in questo caso il sito di taglio è a destra del centro del sito di riconoscimento, quindi si tratta di una sequenza sporgente al 3'. (IV) Tutte le estremità sono identiche (si tratta di sequenze sporgenti CATG 3').

**e.** *NaeI* (GCC $\wedge$ GGC) (I)  $3 \times 10^9 / 4096 = \sim 732\,400$  frammenti. (II) 4096 coppie di basi. (III) Estremità nette. Si noti che il sito di taglio si trova nel mezzo della sequenza di riconoscimento e che l'enzima di restrizione taglia entrambi i filamenti nello stesso punto. (IV) Tutte le estremità sono identiche.

**f.** *BanI* ( $G\wedge$ GYRCC) (I) In questo caso, il frammento di restrizione dovrebbe tagliare il DNA una volta ogni (4) (4) (2) (2) (4) (4) = 1024 bp. (Questo perché le basi Y potrebbero essere C o T, mentre le basi R potrebbero essere A o G). Pertanto,  $3 \times 10^9 / 1024 = \sim 2\,930\,000$  frammenti. (II) 1024 coppie di basi. (III) Coesive, terminano con una sequenza sporgente al 5'. (IV) Tutte le estremità NON sono identiche. Dopo la digestione, alcune estremità appiccicose risultanti saranno 5' GCAC mentre altre saranno 5'GCGC, 5'GTAC o 5'GTGC.

**g.** *BstYI* ( $R\wedge$ GATCY) (I)  $3 \times 10^9 / 1024 = \sim 2\,930\,000$  frammenti. Il valore 1024 è calcolato come nella parte (d). (II) 1024 coppie di basi. (III) Coesive, terminano con una sequenza sporgente al 5'. (IV) Tutte le estremità sono identiche. In contrasto con la parte (d), gli sbalzi effettivi prodotti con questo enzima di restrizione saranno tutti 5'GATC.

**h.** *BsII* (CCNNNNNN $\wedge$ NNGG) (I)  $3 \times 10^9 / 256 = \sim 12\,000\,000$  di frammenti. Anche se questa sequenza di riconoscimento è lunga, sono definite solo 4 coppie di basi, mentre tutte le altre possono essere qualsiasi coppia di basi. Quindi la possibilità che una sequenza casuale di 11 coppie di basi venga riconosciuta da questo en-

zima è  $(1/4)(1/4)(1)(1)(1)(1)(1)(1)(1/4)(1/4) = 1/256$ . (II) 256 coppie di basi. (III) Coesive, terminano con una sequenza sporgente al 3'. (IV) Tutte le estremità NON sono identiche, le sporgenze in 3' sarebbero lunghe 3 basi ma potrebbero essere 3 basi qualsiasi.

**i. SbfI** (CCTGCA<sup>^</sup>GG): (I) Questo sito di riconoscimento è lungo 8 bp, quindi l'enzima taglierebbe in media ogni 48 o 65 536 bp.  $3 \times 109/65\,536 = \sim 45\,800$  frammenti. (II) 65 536 coppie di basi = 65 kb. (III) Coesive, terminano con una sequenza sporgente di 3'. (IV) Tutte le estremità sono identiche e sarebbero TGCA 3'.

**9.3 a. Sau3A** (<sup>^</sup>GATC): l'enzima taglia con una frequenza di  $(0,204)^2(0,296)^2 = 0,00364$  siti di taglio per coppia di basi. La dimensione media del frammento è quindi  $1/0,00364 \approx 274$  paia di basi.

**b. BamHI** (G<sup>^</sup>GATCC): l'enzima taglia con una frequenza di  $(0,204)^4(0,296)^2 = 0,000152$  siti di taglio per coppia di basi. Il frammento medio è  $1/0,000152 \approx 6590$  coppie di basi.

**c. HpaII** (C<sup>^</sup>CGG): l'enzima taglia con una frequenza di  $(0,204)^4 = 0,00173$  siti di taglio per coppia di basi. Il frammento medio è  $1/0,00173 \approx 577$  paia di basi.

**d. SphI** (GCATG<sup>^</sup>C): l'enzima taglia con una frequenza di  $(0,204)^4(0,296)^2 = 0,000152$  siti di taglio per coppia di basi. Il frammento medio è  $1/0,000152 \approx 6590$  coppie di basi.

**e. NaeI** (GCC<sup>^</sup>GGC): l'enzima taglia con una frequenza di  $(0,204)^6 = 0,0000721$  siti di taglio per coppia di basi. Il frammento medio è  $1/0,0000721 \approx 13\,875$  coppie di basi.

**f. BanI** (G<sup>^</sup>G<sup>^</sup>YRCC): l'enzima taglia con una frequenza di  $(0,204)^4(0,204 + 0,296)^2 = 0,000433$  siti di taglio per coppia di basi. Il frammento medio è  $\approx 2310$  coppie di basi.

**g. BstYI** (R<sup>^</sup>GATCY): l'enzima taglia con una frequenza di  $(0,204 + 0,296)^2(0,204)^2(0,296)^2 = 0,000916$  siti di taglio per coppia di basi. Il frammento medio è  $1/0,000916 \approx 1097$  coppie di basi.

**h. BslI** (CCNNNN<sup>^</sup>NNGG): l'enzima taglia con una frequenza di  $(0,204)^4 = 0,00173$  siti di taglio per coppia di basi. Il frammento medio è  $1/0,00173 \approx 577$  paia di basi.

**i. SbfI** (CCTGCA<sup>^</sup>GG): l'enzima taglia con una frequenza di  $(0,204)^6(0,296)^2 = 0,00000631$  siti di taglio per coppia di basi. Il frammento medio è  $1/0,00000631 \approx 158\,356$  coppie di basi.

**9.4** Si noti che la molecola di DNA raffigurata ha due siti di riconoscimento per l'enzima EcoRI, che produce tre frammenti. Questi frammenti sono mostrati in diversi colori di seguito:

5'AGATG3' 5'AATTCGCTGAAGAACCAAG3' 5' AATTCGATT 3'  
3'TCTACTTAA5' 3'GCGACTTCTGGTTCTAA 5' 3'GCTAA 5'

Poiché il sito di restrizione *EcoRI* è lungo 6 bp (G<sup>^</sup>AATTC), l'enzima digerirebbe in media il DNA di sequenza casuale solo una volta ogni 4<sup>6</sup> o 4096 bp. Il pezzo originale di DNA mostrato in questo problema è molto più piccolo di 4096 bp, quindi è raro trovare due siti di riconoscimento per *EcoRI* così vicini tra loro. Tuttavia, se il DNA che è stato digerito è lungo (come i 3 miliardi di bp nel genoma umano), è probabile che da qualche parte nel genoma due di questi siti siano tanto vicini quanto quelli dell'esempio.

**9.5** Il gel di agarosio attraverso il quale il DNA viene sottoposto a elettroforesi presenta maglie i cui fori hanno all'incirca le dimensioni dei frammenti di DNA da analizzare. I frammenti di DNA più piccoli troveranno più fori attraverso i quali possono viaggiare, quindi si muoveranno più velocemente (i loro percorsi saranno più lineari) dei frammenti di DNA più grandi, i cui movimenti saranno ritardati quando andranno a sbattere contro fori che non sono abbastanza grandi per poterli attraversare (i loro percorsi saranno più tortuosi e quindi più lunghi). Si noti che la forza motrice esercitata dal campo elettrico sui frammenti di DNA è la stessa per tutti i frammenti di DNA indipendentemente dalle loro dimensioni, perché tutti i nucleotidi hanno all'incirca la stessa massa e la stessa carica. Pertanto, la capacità di un frammento di trovare fori che può attraversare è il principale determinante della mobilità di un frammento di DNA nel gel.

**9.6** Si possono preparare facilmente gel con pori di diverse dimensioni, utilizzando diverse concentrazioni di agarosio. Maggiore è la concentrazione di agarosio, minore è la dimensione media dei pori, perché l'agarosio occuperà una parte maggiore del volume. I gel sono generalmente compresi tra lo 0,5 e il 2,5% di agarosio, il che significa da 0,5 a 2,5 g di agarosio disciolti in un volume finale di 100 mL di soluzione tampone. Un gel allo 0,5% può risolvere frammenti di DNA tra ~1 kb e 3 kb. Un gel al 2,5% può risolvere frammenti da ~50 bp a un massimo di ~2 kb.

**9.7** Corsia A: genoma del batteriofago digerito con *EcoRI*. Nel gel si vedono solo poche bande, quindi il genoma digerito deve essere piccolo. Ci sono meno bande di dimensioni maggiori rispetto a quelle osservate nella corsia C (l'altro campione in cui è stato digerito il genoma del batteriofago, ma in quel caso con un enzima che riconosce

una sequenza di 4 bp anziché la sequenza di 6 bp per *EcoRI*).

Corsia B: DNA genomico umano digerito con *EcoRI*. Viene prodotto una strisciata composta da molte bande, quindi il genoma digerito deve essere molto grande. Questi frammenti sono di dimensioni medie maggiori rispetto a quelli nella corsia D (dove il genoma umano è stato digerito con l'enzima *HpaII*).

Corsia C: genoma del batteriofago digerito con *HpaII*. Si vedono solo poche bande (quindi questo deve essere DNA del batteriofago), ma queste sono più piccole di quelle viste nella corsia A dove lo stesso DNA è stato digerito con *EcoRI*, che ha un sito di riconoscimento a sei basi.

Corsia D: DNA genomico umano digerito con *HpaII*. Viene prodotta una strisciata di molte bande, quindi questo deve essere una digestione del DNA genomico umano. Questi sono in media più piccoli di quelli visti nella corsia B, quindi il sito di riconoscimento per l'enzima utilizzato nella corsia D deve essere più piccolo del sito di riconoscimento per l'enzima utilizzato per digerire lo stesso DNA nella corsia B.

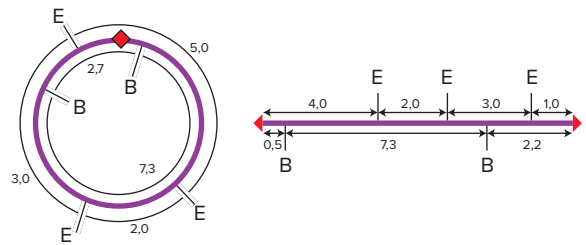
**9.8 a.** Il campione A rappresenta la forma circolare. Confrontare le digestioni con *EcoRI* del campione A e del campione B. Il campione A mostra 3 bande mentre il campione B mostra 4 bande. Se si digerisce una molecola circolare in 3 punti, si ottengono 3 bande, mentre digerendo una molecola lineare negli stessi 3 punti, si ottengono 4 bande. Lo stesso schema si osserva per tutti le altre digestioni: il campione A ha sempre 1 banda in meno rispetto al campione B quando i due campioni vengono tagliati con lo stesso(i) enzima(i).

**b.** Si possono sommare le dimensioni dei frammenti di restrizione ottenuti dalla digestione del Campione B (la forma lineare del DNA del batteriofago) per trovare la lunghezza totale di questa molecola. Per esempio, nella digestione con *EcoRI*, questa sarebbe  $4 + 3 + 2 + 1 = 10$  kb. Si troverà lo stesso valore per qualsiasi altra digestione del campione B. (Nota: il genoma del batteriofago  $\lambda$  è in realtà 48,5 kb.)

**c.** Utilizzando la stessa procedura di addizione della parte (b), la lunghezza totale della forma circolare del DNA del batteriofago è  $5 + 3 + 2 = 10$  kb. La dimensione del DNA del batteriofago non cambia quando viene circolarizzato.

**d.** La figura seguente illustra le posizioni di tutti i siti di riconoscimento degli enzimi di restrizione (E = *EcoRI*; B = *BamHI*) e le dimensioni dei frammenti che verrebbero prodotti dalla dige-

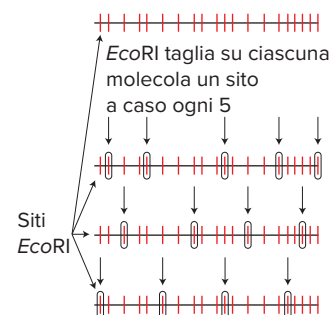
stione della forma circolare (sinistra) o linearizzata (destra) del cromosoma. I triangoli rossi e il diamante rosso indicano le posizioni delle estremità coesive del cromosoma del batteriofago  $\lambda$ .



**9.9 a.** Con la digestione di restrizione parziale, l'enzima non taglia il DNA genomico in ogni sito di riconoscimento presente nel genoma. Pertanto, la digestione parziale produrrebbe un minor numero di frammenti di dimensioni medie maggiori rispetto a quelli che verrebbero prodotti digerendo totalmente lo stesso campione di DNA con lo stesso enzima.

**b.** I frammenti di DNA prodotti dai genomi di diverse cellule mediante digestione di restrizione parziale sarebbero diversi l'uno dall'altro. In contrasto con una digestione completa, in cui ogni sito di riconoscimento *EcoRI* verrebbe tagliato nelle molecole di DNA di tutte le cellule, in una digestione parziale si otterrebbe una gamma di frammenti diversi prodotti da molecole di DNA genomico diverse.

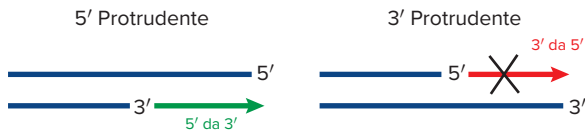
La figura che segue illustra che 3 molecole di DNA genomico provenienti da tre diverse cellule produrrebbero in effetti frammenti diversi dopo una digestione parziale. (In questo caso, le condizioni sono state regolate in modo tale che in media fosse tagliato solo un sito su cinque; digerendo il DNA per tempi più o meno lunghi con l'enzima, i ricercatori possono produrre frammenti con dimensioni medie diverse.)



La digestione parziale del DNA con enzimi di restrizione ha importanti implicazioni pratiche. Poiché diverse molecole di DNA genomico pro-

ducono frammenti diversi, gli scienziati possono utilizzare la digestione a restrizione parziale per creare librerie genomiche con frammenti sovrapposti che consentono la ricostruzione della sequenza del DNA genomico.

- 9.10 a.** No; ciascuna delle due estremità di un frammento di DNA genomico che è stato tagliato è il prodotto di un diverso evento di rottura.
- b.** Le estremità non sono coesive, il che significa che le estremità non sono tutte di sequenze complementari e quindi non possono ibridarsi tutte (rinaturare) tra loro o con un vettore tagliato con uno stesso enzima di restrizione. Innanzitutto, molti dei frammenti hanno estremità nette. In secondo luogo, i frammenti con estremità sporgenti non hanno tutti la stessa sequenza; le sporgenze possono essere di qualsiasi lunghezza, o di qualsiasi sequenza di base e 5' o 3'.
- c.** Solo le sporgenze 5' possono essere rese nette dalla DNA polimerasi. Poiché la DNA polimerasi sintetizza il DNA nella direzione 5'-3', la sporgenza 5' può essere utilizzata come stampo per l'estensione dell'altro filamento di DNA.



Entrambe le sporgenze 5' e 3' possono essere rese nette con la nucleasi S1, perché l'enzima degraderà il DNA a filamento singolo indipendentemente dalla polarità.

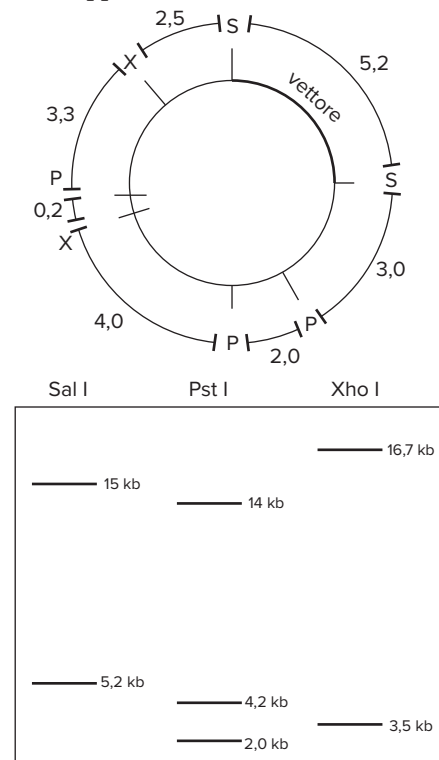
- 9.11 a.** Lo scopo del clonaggio molecolare è isolare il DNA corrispondente a una piccola regione di un genoma complesso. Con il clonaggio, gli scienziati possono purificare una quantità sufficiente di questa piccola regione di DNA per studiarla in dettaglio (per esempio, per determinare la sequenza nucleotidica della regione).
- b.** I marcatori selezionabili nei vettori forniscono un mezzo per determinare quali cellule nel mix di trasformazione hanno incorporato il vettore. Questi marcatori sono spesso geni di resistenza ai farmaci, quindi un farmaco può essere aggiunto al mezzo di coltura e solo le cellule che hanno ricevuto e mantenuto il vettore cresceranno in colonie.
- c.** L'origine della replicazione in un vettore plasmidico consente al vettore (e a qualsiasi molecola di DNA ricombinante che incorpora il vettore) di replicarsi in una cellula batterica ospite, indipendentemente dal cromosoma batterico. Questa proprietà consente al plasmide di essere trasferito

to alle cellule figlie quando le cellule si dividono e consente ai ricercatori di purificare il plasmide dal cromosoma dell'ospite batterico. Alcune origini della replicazione dei plasmidi sono così efficienti che una singola cellula batterica può contenere centinaia di copie del plasmide, consentendo ai ricercatori di preparare grandi quantità di DNA del plasmide.

**d.** I polylinker consentono di aprire (linearizzare) i vettori con uno (o due) diversi enzimi di restrizione in modo che i frammenti di DNA con uno (o due) tipi di estremità possano essere ligati nel vettore.

- 9.12** Per creare una molecola di DNA ricombinante, sono necessari (oltre al vettore e il DNA da inserire) i seguenti enzimi: (c) un enzima di restrizione per tagliare il vettore e il DNA genomico in modo che i pezzi risultanti abbiano estremità coesive compatibili; (d) la DNA ligasi per legare insieme il vettore e l'inserito.
- 9.13** Sì, questo è possibile se gli enzimi riconoscono la stessa sequenza ma tagliano quella sequenza in un modo diverso. Tali coppie di enzimi sono chiamate *neoschizomeri*. Per esempio, sia *ApaI* sia *Bsp120I* riconoscono la sequenza 5'GGGCCC3', ma uno lascia un'estremità protrudente al 5' (*Bsp120I*: 5'G^GGCCCC3') e l'altro lascia un'estremità protrudente al 3' (*ApaI*: 5'GGGCC^C3'). Queste estremità coesive non sono compatibili perché non possono appaiarsi tra loro.

**9.14**





**9.15** Dopo la digestione, il vettore (rosso) apparirà come segue (qui rappresentato come molecola lineare; i punti alle due estremità sono collegati tra loro.)

5'...CGGATCCCTAAGATG3' 5'AATTCGCGCGCATCGGC...3'  
3'...GCCTAGGGGATTCTACTTAA5' 3'GGCGCGCGTAGCCG...5'

Il frammento dell'inserto, che è il frammento blu del Problema 4 con estremità coesive *EcoRI* su entrambi i lati, è il seguente:

5' AATTCGCTGAAGAACCAAG 3'  
3' GCGACTTCTTGTTCTTAA 5'

**a.** Le due possibili molecole ricombinanti sono:

5'...CGGATCCCTAAGATG  
AATTCGCTGAAGAACCAAG  
AATTCGCGCGCATCGGC...3'  
3'...GCCTAGGGGATTCTACTTAA  
GCGACTTCTTGTTCTTAA  
GGCGCGCGTAGCCG...5'

o

5'...CGGATCCCTAAGATG  
AATTCGCTGAAGAACCAAG  
AATTCGCGCGCATCGGC...3'  
3'...GCCTAGGGGATTCTACTTAA  
GAACCAAGAAGTCGTTAA  
GGCGCGCGTAGCCG...5'

**b.** Il frammento potrebbe essere inserito nel vettore in una delle due orientazioni che conservano ancora la polarità da 5' a 3' nelle risultanti molecole ricombinanti. Relativamente l'uno all'altro, i due orientamenti sono entrambi capovolti (in termini di filamenti) e ruotati di 180°.

**c.** La molecola di DNA ricombinante prodotta (indipendentemente dall'orientamento dell'inserto) avrà un solo sito per *BamHI*. Questo deriva dal vettore ed è sottolineato nelle risposte alla parte (a). L'inserto blu non ha siti di riconoscimento per questo enzima.

**d.** Dopo che il DNA ricombinante è stato digerito con *BamHI*, esisterà un solo frammento (la molecola ricombinante circolare sarà linearizzata). La dimensione di questo frammento coincide con la lunghezza del vettore pMBG36 (4271 bp) più la dimensione dell'inserto (19 bp), cioè 4290 bp.

**e.** La molecola di DNA ricombinante avrebbe due siti per *EcoRI*; uno alla giunzione sinistra e l'altro alla giunzione destra tra vettore e inserto. Questo è un punto importante: dopo la saldatura, l'inserto sarà affiancato da siti *EcoRI* su entrambi i lati, perché sia il vettore sia l'inserto contribui-

scono alle sequenze che formeranno questi siti nel DNA ricombinante.

**f.** Dopo che il DNA ricombinante è stato digerito con *EcoRI*, esisteranno due frammenti; uno è il vettore e l'altro è l'inserto. Un frammento sarà quindi 4271 bp e l'altro sarà 19 bp. (Qui, contiamo le estremità adesive a singolo filamento come se fossero a doppio.)

**9.16** In primo luogo, si lavora attraverso la digestione e la saldatura dei frammenti di DNA e del vettore. Il vettore (rosso) viene tagliato con *BamHI*, lasciando una molecola lineare (nel plasmide integro i punti si collegano tra loro) e le seguenti estremità:

5'...G3' 5'GATCC...3' nel  
3'...CCTAG5' 3'G...5'

Il DNA dell'inserto (blu) viene tagliato con *MboI*, lasciando le seguenti estremità adesive. La N indica qualsiasi base; ovviamente le basi nei due filamenti sono complementari.

5' GATCN — N 3' 5' GATCN — N 3'  
3' N — NCTAG 5' 3' N — NCTAG 5'

La saldatura di un frammento *MboI* a un'estremità adesiva *BamHI* creerà solo occasionalmente una sequenza che può essere digerita da *BamHI*. Dipende dall'esatta sequenza di basi alle estremità del frammento *MboI*. La N nella sequenza seguente indica questa ambiguità. In tutti i casi si troverà la seguente sequenza:

5'...GGATCN — NGATCC...3'  
3'...CCTAGN — NCTAGG...5'

**a.** Il 100% delle giunzioni può essere digerito con *MboI* perché contengono tutte la sequenza di riconoscimento 5'GATC3'.

**b.** Una giunzione che può essere digerita con *BamHI* deve avere una C all'estremità 3' della sequenza di riconoscimento *MboI*; in altre parole, la N che segue il 5'GATC all'incrocio dovrebbe essere C. Ciò avverrebbe 1/4 o il 25% delle volte.

**c.** Nessuna delle giunzioni sarà scindibile da *XorII* perché nessuna delle possibili giunzioni ha la sequenza 5'CGATCG3'.

**d.** Leggendo su un filamento, una giunzione che può essere digerita da *BstYI* deve avere una purina prima del 5'GATC e una pirimidina dopo il 5'GATC. Nel diagramma precedente, il nucleotide che precede il 5'GATC è sempre una purina (è una G originariamente derivata dal vettore), mentre il nucleotide che segue il 5'GATC sarà una pirimidina il 50% delle volte nel DNA. Pertanto, il 50% delle giunzioni può essere tagliato da *BstYI*.

e. Nel genoma umano, un sito *MboI* può essere scritto 5'...N<sub>1</sub>GATCN<sub>2</sub>...3' su un filamento. Dopo la scissione con *MboI*, il prodotto sarebbe 5'GATCN<sub>2</sub>...3'. Se questo ora viene inserito nel vettore digerito da *Bam*HI e il sito risultante può essere tagliato con *Bam*HI, allora N<sub>2</sub> doveva essere un C. Quindi, il sito originale nel genoma umano era 5'...N<sub>1</sub>GATCC...3'. Affinché il sito di restrizione sia un sito *Bam*HI nel genoma umano, N<sub>1</sub> deve essere un G. Nel DNA, la possibilità che N<sub>1</sub> NON fosse un G = 3/4.

**9.17 a.** Avrete bisogno di 4-5 equivalenti del genoma per raggiungere un livello di confidenza del 95% di trovare una sequenza di DNA univoca specifica.

**b.** Il numero di cloni necessari dipende dalla dimensione totale del genoma dell'organismo di ricerca e dalla dimensione media dell'inserito nel vettore. I vettori plasmidici normalmente hanno inserti più piccoli di 15 kb, mentre gli inserti nei vettori costituiti da cromosomi batterici artificiali (BAC) possono essere di 300 kb e quelli nei cromosomi artificiali di lievito (YAC) possono avere inserti grandi fino a 2000 kb (= 2 Mb). Si divide il numero di coppie di basi del genoma per la dimensione media dell'inserito, e si moltiplica per cinque, per ottenere il numero di cloni di cinque equivalenti del genoma.

**9.18 a.** Il numero minimo teorico di BAC ricombinanti per coprire l'intero genoma è la dimensione del genoma divisa per la dimensione media dell'inserito in ciascun BAC: 3 000 000 000 paia di basi per genoma/100 000 paia di basi per BAC = 30 000 BAC/genoma. Si noti che per coprire l'intero genoma, ciascuno dei 30 000 BAC dovrebbe contenere un segmento completamente diverso (non sovrapposto) del genoma che finirebbe alla coppia di basi prima che inizi il successivo. È praticamente impossibile che si possa mai costruire una tale genoteca usando DNA genomico frammentato casualmente.

**b.**  $f = 100\,000\text{ bp per BAC} / 3\,000\,000\,000\text{ bp per genoma} = 0,000034\text{ genomi per BAC}$ . (In alternativa, ogni inserto BAC contiene  $\sim 1/30\,000$  del genoma.)

**c.** Per una probabilità del 99% che sia rappresentata una specifica regione di 100 kb:

$$N = [\ln(1 - 0,99)] / [\ln(1 - 0,000034)] = 135\,444\text{ BAC.}$$

Per una probabilità del 99,9% che sia rappresentata una specifica regione di 100 kb:

$$N = [\ln(1 - 0,999)] / [\ln(1 - 0,000034)] = 203\,170\text{ BAC.}$$

**d.** Nella parte (a) si è calcolato che 30 000 BAC è 1 equivalente genomico. Pertanto, 135 444 BAC

sono  $135\,444/30\,000 \approx 4,5$  equivalenti genomici; 203 170 BAC sono  $203\,170/30\,000 \approx 6,8$  equivalenti genomici.

**e.** Possiamo rispondere a questa domanda usando l'equazione fornita e calcolare  $P$ . In questo caso,  $N = 30\,000$ .

$$30\,000 = [\ln(1 - P)] / [\ln(1 - 0,000034)]$$

$$30\,000 \times \ln(1 - 0,000034) = \ln(1 - P)$$

$$-1,02 = \ln(1 - P)$$

$$e^{-1,02} = 1 - P$$

$$0,36 = 1 - P$$

$$P = 1 - 0,36 = 0,64$$

Pertanto, la probabilità è del 64% che qualsiasi regione specifica di 100 000 bp del genoma sia rappresentata in un BAC ricombinante. Da notare che questa è la stessa genoteca che hai costruito nella parte (a).

**f.** Una delle principali lezioni di questo problema è che, poiché  $\ln(0)$  non è definito, non è possibile creare una genoteca in cui la probabilità che un segmento sia rappresentato sia del 100%. Invece, operativamente bisognerebbe tendere sempre a costruire una genoteca con una probabilità predefinita  $P$  che una regione specifica sia rappresentata. Se si risolve per  $N$ , e se  $P$  è già definito in questo modo, l'unica altra variabile è  $f$ , la frazione del genoma in un singolo clone ricombinante.

Per un dato genoma,  $f$  può essere variato solo alterando la dimensione dell'inserito: più grande è l'inserito medio, meno cloni sarebbero necessari per creare la genoteca. BAC e YAC sono apprezzati come vettori per la costruzione di librerie genomiche perché possono ospitare inserti di grandi dimensioni.

**9.19** Uno dei problemi che si affrontano nel clonaggio molecolare (la costruzione di molecole di DNA ricombinante) è che dopo che il vettore è stato tagliato da un enzima di restrizione, le due estremità del vettore possono tornare insieme ed essere risigillate dalla DNA ligasi senza che vi sia alcun inserto incorporato. Tali vettori risaldati sono di intralcio perché le cellule batteriche trasformate con questi DNA formerebbero colonie che non potrebbero essere distinte (tranne per il sequenziamento del DNA che richiede tempo) dalle colonie contenenti il DNA ricombinante che si stava cercando di creare. Un modo per affrontare questo problema è utilizzare vettori come quello mostrato nella figura.

Se il vettore si richiudesse senza alcun inserto (questi a volte sono chiamati vettori vuoti), le colonie risultanti che crescono su piastre di ampi-

cillina integrate con X-gal sarebbero di colore blu, perché queste cellule produrrebbero l'enzima  $\beta$ -galattosidasi e questo enzima trasformerebbe l'incolore X-gal in un composto blu. Se il vettore contenesse un inserto, le colonie risultanti che crescono sulla stessa piastra sarebbero bianche, perché non verrebbe prodotto alcun enzima e quindi alcun composto blu. Questa proprietà del vettore consente ai ricercatori di identificare le colonie bianche che hanno molecole di DNA ricombinante e di ignorare le colonie blu che hanno risalato i vettori senza inserto.

- 9.20 a.** (1) 3,1, 6,9 kb; (2) 4,3, 4,0, 1,7 kb; (3) 1,5, 0,6, 1,0, 6,9 kb; (4) 4,3, 2,1, 1,9, 1,7 kb; (5) 3,1, 1,2, 4,0, 1,7 kb.

**b.** Il frammento di 6,9 kb della digestione *EcoRI* + *HindIII*; i frammenti di 2,1 e 1,9 kb di *BamHI* + *PstI* e il frammento di 4,0 kb del digerito di *EcoRI* + *BamHI*.

- 9.21** Assumendo di avere già disponibile la molecola di DNA da sequenziare, il solo enzima necessario per l'operazione è la DNA polimerasi (a).

- 9.22 a.** Il filamento neosintetizzato è complementare al filamento stampo. Leggendo dal gel la sequenza del filamento di nuova sintesi si può risalire alla sequenza del filamento stampo.

Filamento neosintetizzato:

5' TAGCTAGGCTAGCCCTTTATCG 3'

Filamento stampo:

3' ATCGATCCGATCGGGAAATAGC 5'

**b.** Nella sequenza ci sono dei codoni di stop in frame, quindi è improbabile che questa sia la sequenza di un esone di una regione codificante.

- 9.23** Il punto chiave del sequenziamento genomico è che i frammenti da sequenziare devono sovrapporsi; altrimenti non si avrebbe modo di capire la loro relazione reciproca nel genoma originale. Questo problema esplora diversi metodi per ottenere tale sovrapposizione.

**a.** Il taglio meccanico frammenterà il DNA in posizioni casuali, assicurando che si ottenga la sovrapposizione se vengono analizzati frammenti sufficienti. Poiché il campione di DNA genomico è ottenuto da molte cellule, i frammenti di cellule diverse non avranno gli stessi punti di inizio e fine; quindi, si sovrapporranno.

**b.** In questo caso, la sovrapposizione si ottiene realizzando diverse librerie dello stesso genoma, ciascuna libreria costruita con un diverso enzima di restrizione (per esempio, una libreria realizzata con l'enzima *EcoRI* e una seconda libreria con l'enzima *BamHI*). I siti di riconoscimento

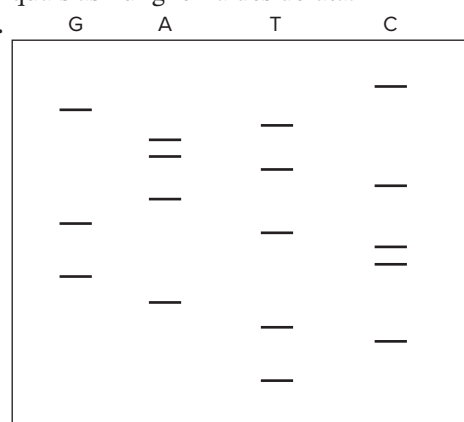
per diversi enzimi si trovano in posizioni diverse nel genoma da analizzare, in modo che i frammenti in qualsiasi regione ottenuta dalle librerie si sovrappongano.

Un problema con la costruzione di una genoteca mediante la digestione completa di un genoma con un dato enzima di restrizione è che, per caso, alcuni frammenti possono essere troppo grandi per essere clonati o analizzati, mentre altri sono troppo piccoli. Pertanto, un altro vantaggio dell'utilizzo di più librerie realizzate con diversi enzimi di restrizione è ridurre al minimo la possibilità che una regione non venga coperta. Per esempio, un frammento troppo lungo con siti *EcoRI* alle sue estremità può avere siti interni per *BamHI* o qualche altro enzima.

**c.** Si potrebbe eseguire una digestione parziale con il singolo enzima di restrizione. Come mostrato nella figura che accompagna la risposta al Problema 9.9, la digestione parziale garantirà la sovrapposizione perché diverse molecole di DNA genomico verranno tagliate in diversi gruppi di siti di riconoscimento.

Se l'enzima di restrizione riconosce siti che sono relativamente distanti nel genoma (per esempio, una sequenza di riconoscimento di 6 bp verrà trovata in media ogni 4 kb, mentre una sequenza di 8 bp verrà trovata in media ogni 64 kb), qualsiasi approccio utilizzando un singolo enzima ne risentirà perché alcuni frammenti sono troppo grandi per essere clonati. Sarà preferibile utilizzare un enzima che riconosca una sequenza di 4 bp perché i siti saranno più ravvicinati. Più sono i siti potenziali di digestione, più i frammenti assomiglierebbero a una genoteca realizzata mediante taglio meccanico con tagli eseguiti in posizioni casuali. Se si regolano correttamente le condizioni di digestione parziale, è possibile costruire una genoteca che abbia inserti con una dimensione media di qualsiasi lunghezza desiderata.

- 9.24 a.**



**b.** Il filamento di nuova sintesi contiene codoni di stop in tutti e tre i frame di lettura (sottolineati) e pertanto non può essere una sequenza codificante (esone). Sulla sequenza di DNA del filamento stampo il registro di lettura che inizia con il primo nucleotide non contiene codoni di stop, per cui questo filamento è quello RNA-like.

Filamento sintetizzato:  
5' TCTAGCCTGAACTAATGC 3'

Sequenza di DNA stampo:  
3' AGATCGGACTTGATTACG 5'

**c.** La sequenza amminoacidica dell'esapeptide è N-Ala-Leu-Val-Gln-Ala-Arg-C. L'estremità ammino-terminale corrisponde all'estremità 5' della sequenza di DNA dell'mRNA-like.

**9.25** Il sequenziamento shotgun dell'intero genoma ridotto in piccoli frammenti è ora una strategia praticabile per il sequenziamento del genoma di un individuo umano perché abbiamo già la sequenza di riferimento umana; le letture paired-end di inserti più grandi non sono più necessarie. Usando BLAST, ogni breve sequenza letta può essere abbinata a una posizione sulla sequenza di riferimento umana.

**9.26 a.** Leggendo i picchi da sinistra a destra (corrispondenti alla direzione 5-3' perché la DNA polimerasi aggiunge nucleotidi all'estremità 3'), la sequenza sarebbe:

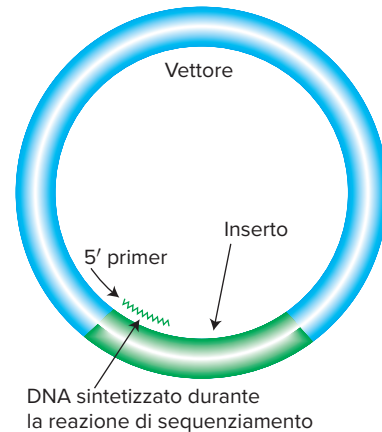
5' TTTGCTTTGTGAGCGGATAACAA 3'

[Da notare che questa NON è la sequenza del primer, ma la sequenza di una parte dell'inserto di DNA umano.]

**b.** La sequenza scritta nella parte (a) è del filamento neosintetizzato nella reazione di sequenziamento.

**c.** Si può progettare il primer perché: (1) l'inserto è stato aggiunto nel vettore in una posizione nota, definita da uno specifico enzima

di restrizione; (2) si conosce la sequenza del vettore che fiancheggia la posizione dell'inserto.



**d.** La più piccola molecola sintetizzata in una reazione di sequenziamento che contiene dideoossiG sarebbe:

5' GCCTCGAATCGGGTACCTTG\* 3'

L'asterisco indica il dideoossiG, che deve trovarsi all'estremità 3' perché la catena termina quando questo dideoossi-nucleotide è stato incorporato. Il primer è sottolineato. (Questa risposta presuppone che il picco rosso più piccolo nell'analisi della sequenza rappresenti il DNA più piccolo possibile che potrebbe essere prodotto nella reazione di sequenziamento.)

**9.27** Se il genoma materno e paterno differiscono a un determinato locus, supponiamo che l'allele materno sia una G e l'allele paterno una T. Una singola read, che copre il locus, conterrà uno solo dei possibili alleli, o la G o la T. Quindi la ricostruzione del genoma di quell'individuo mostrerà in quel locus circa il 50% di read con la G e il 50% di read con la T, fornendo l'indicazione che l'individuo è eterozigote. Allo stesso modo, si produrranno due contig diversi, che coprono la regione cromosomica, uno con l'allele G e uno con quello T.