CAPITOLO

12

Le tecnologie del DNA

Per leggere, copiare e riscrivere il DNA abbiamo preso in prestito gli attrezzi che batteri e virus usano da miliardi di anni.

OBIETTIVI

1. Spiegare come gli enzimi di restrizione e la DNA ligasi tagliano e uniscono il DNA, e stimare la lunghezza dei frammenti con 4^n
2. Descrivere i vettori e le tappe del clonaggio, e spiegare perché per esprimere un gene eucariote in un batterio serve il cDNA
3. Descrivere la PCR, calcolare le copie dopo n cicli e interpretare una PCR in tempo reale
4. Interpretare un gel di elettroforesi con una retta di taratura e ricostruire una mappa di restrizione
5. Leggere una sequenza ottenuta con il metodo di Sanger e descrivere il sequenziamento di nuova generazione e la genomica
6. Spiegare il funzionamento di CRISPR-Cas9, le vie di riparazione del taglio e i limiti dell'editing genomico

INDICE DEL CAPITOLO

12.1 Tagliare e incollare il DNA 2
12.2 Il DNA ricombinante e il clonaggio 5
12.3 Copiare il DNA in provetta: la PCR 8
12.4 Separare e riconoscere i frammenti 13
12.5 Leggere il DNA: il sequenziamento e la genomica 16
12.6 Riscrivere il DNA: l'editing genomico 19

Scienze naturali 5 · Parte III · Le biotecnologie · versione 1.0
Capitolo 12: Le tecnologie del DNA · pagine 323–350 del volume completo · estratto per lo studio

Un tampone per SARS-CoV-2 raccoglie poche gocce di muco dal naso. In un laboratorio di analisi, in poco più di un'ora, una macchina dice se nel campione c'è l'RNA del virus, anche quando le particelle virali sono così poche che nessun microscopio potrebbe trovarle. Il segreto non è uno strumento più sensibile, ma un trucco chimico: si copia milioni di volte un breve tratto del genoma del virus finché diventa abbastanza abbondante da emettere un segnale misurabile.

Questa e le altre tecnologie del DNA permettono di tagliare il DNA in punti precisi, di incollarne pezzi di origine diversa, di farli copiare a un batterio o a una provetta, di separarli, di leggerne la sequenza e, da pochi anni, di riscriverla. Quasi nessuno di questi strumenti è stato inventato da zero: sono enzimi che batteri e virus usano da miliardi di anni, presi in prestito e messi al servizio di una domanda. In questo capitolo vedremo come funzionano; nel capitolo 13 vedremo come si usano in medicina, in agricoltura, nell'ambiente e nella giustizia, e quali domande etiche sollevano.

DA RICORDARE

Richiamo · La cassetta degli attrezzi. Dal volume dello scorso anno e da quello precedente ti servono cinque idee (vedi *Scienze naturali 3*, cap. 13–15). 1) Il DNA è una doppia elica di filamenti **antiparalleli**, appaiati per complementarità (A–T, G–C); ogni filamento ha un'estremità 5′ (con il fosfato) e un'estremità 3′ (con l'ossidrile). 2) Le DNA polimerasi allungano un filamento solo in direzione 5′→3′, aggiungendo nucleotidi all'ossidrile 3′ di un **primer** appaiato allo stampo. 3) La **DNA ligasi** chiude un'interruzione formando un legame fosfodiesterico. 4) Molti batteri contengono **plasmidi**, piccoli DNA circolari che si duplicano indipendentemente dal cromosoma, e possono assorbire DNA dall'ambiente (**trasformazione**). 5) I retrovirus copiano il loro RNA in DNA con la **trascrittasi inversa**.

12.1 Tagliare e incollare il DNA

12.1.1 Gli enzimi di restrizione, una difesa dei batteri

Negli anni Cinquanta e Sessanta i microbiologi notarono che alcuni batteriofagi crescono bene su un ceppo di *Escherichia coli* e molto male su un altro: nel secondo ceppo il DNA del fago, appena entrato, viene fatto a pezzi. Lo svizzero Werner Arber ipotizzò che i batteri possiedano enzimi capaci di riconoscere e tagliare il DNA estraneo, e che proteggano il proprio DNA marcandolo con gruppi metile. Nel 1970 lo statunitense Hamilton Smith isolò da *Haemophilus influenzae* il primo enzima che taglia sempre nello stesso punto di una sequenza precisa, e l'anno dopo Daniel Nathans lo usò per tagliare il DNA del virus SV40 in frammenti riproducibili e ricavarne una prima mappa. I tre ricevettero nel 1978 il premio Nobel per la medicina.

Questi enzimi si chiamano **enzimi di restrizione** (o endonucleasi di restrizione), perché «restringono» la crescita dei fagi. Ognuno riconosce una breve sequenza di DNA a doppio filamento, il **sito di riconoscimento**, lunga di solito da 4 a 8 paia di basi, e rompe un legame fosfodiesterico in ciascuno dei due filamenti, sempre nella stessa posizione. Il batterio non taglia il proprio cromosoma perché un secondo enzima, una metilasi, aggiunge un gruppo metile a una base di ogni sito, e il sito metilato non viene più riconosciuto: è il sistema di restrizione-modificazione, una sorta di sistema immunitario primitivo che distingue il «sé» dal «non sé».

Il nome di un enzima di restrizione dice da dove viene: **EcoRI** è il primo enzima (I) isolato dal ceppo RY13 (R) di *Escherichia coli* (Eco); **HindIII** è il terzo di *Haemophilus influenzae* ceppo Rd. Se ne conoscono migliaia, e alcune centinaia, con siti diversi, si comprano purificate (Tabella 12.1).

Tabella 12.1 Alcuni enzimi di restrizione molto usati. Il segno ^ indica il punto di taglio sul filamento scritto (5′→3′); sull'altro filamento il taglio è nella posizione simmetrica.

Enzima	Batterio di origine	Sito (5′→3′)	Estremità prodotte
EcoRI	*Escherichia coli*	G^AATTC	coesive, sporgenza 5′ AATT
BamHI	*Bacillus amyloliquefaciens*	G^GATCC	coesive, sporgenza 5′ GATC
HindIII	*Haemophilus influenzae*	A^AGCTT	coesive, sporgenza 5′ AGCT
PstI	*Providencia stuartii*	CTGCA^G	coesive, sporgenza 3′ TGCA
NotI	*Nocardia otitidis-caviarum*	GC^GGCCGC	coesive, sporgenza 5′ GGCC
SmaI	*Serratia marcescens*	CCC^GGG	piatte
AluI	*Arthrobacter luteus*	AG^CT	piatte

12.1.2 Siti palindromi ed estremità coesive

Il sito di EcoRI è 5′-GAATTC-3′. Se leggi il filamento complementare, anch'esso da 5′ a 3′, trovi la stessa sequenza: GAATTC. Una sequenza a doppio filamento che si legge allo stesso modo sui due filamenti, ciascuno nella sua direzione 5′→3′, si chiama **palindromo** (come «ingegni» in italiano, ma qui la simmetria riguarda i due filamenti). Quasi tutti gli enzimi usati in laboratorio riconoscono siti palindromi, perché lavorano come coppie di subunità identiche, una per filamento.

EcoRI taglia ciascun filamento fra la G e la prima A (G^AATTC). Poiché i due tagli sono sfalsati di quattro basi, ogni frammento termina con un breve tratto a singolo filamento, 5′-AATT, che sporge (Figura 12.1). Le sporgenze di due frammenti tagliati dallo stesso enzima sono complementari e tendono a riappaiarsi con legami a idrogeno: per questo si chiamano **estremità coesive** (in inglese *sticky ends*, «appiccicose»). Altri enzimi, come SmaI (CCC^GGG), tagliano i due filamenti nello stesso punto e lasciano **estremità piatte**, senza basi spaiate.

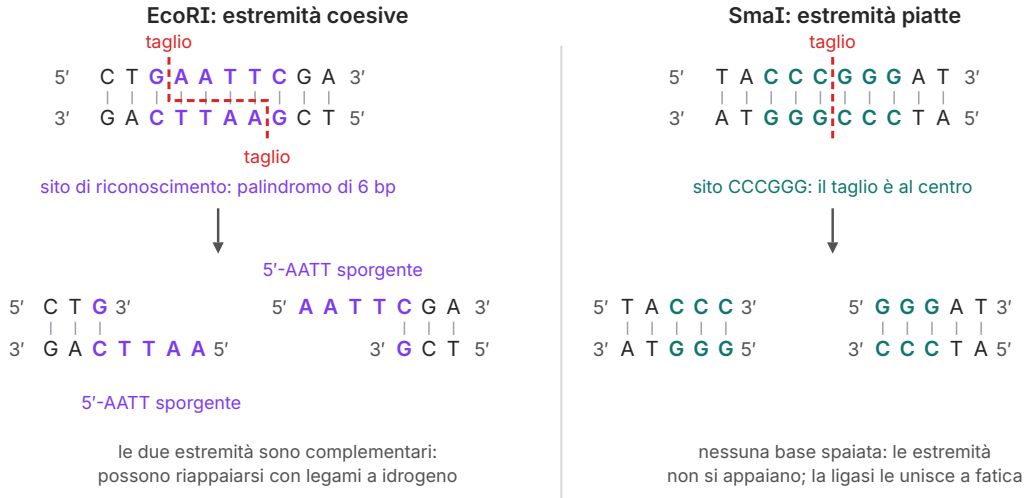


Figura 12.1 Due modi di tagliare. A sinistra, EcoRI riconosce il sito palindromo GAATTC (viola) e taglia ciascun filamento fra G e A: i due tagli (linea rossa tratteggiata) sono sfalsati e lasciano estremità coesive con quattro basi sporgenti al 5′. A destra, SmaI taglia al centro del sito CCCGGG e lascia estremità piatte. Le lineette grigie indicano le coppie di basi appaiate.

12.1.3 Quanto sono lunghi i frammenti?

Un sito di 6 paia di basi compare, in media, una volta ogni 4096 bp: in una sequenza «casuale», in cui le quattro basi si susseguono con la stessa probabilità e in modo indipendente, la probabilità che sei posizioni consecutive formino proprio GAATTC è $(1/4)^6 = 1/4096$. In generale, per un sito di n paia di basi la lunghezza media dei frammenti è $\mathbf{4^n}$: circa 256 bp per un sito di 4 bp, circa 66 kb per un sito di 8 bp. Un enzima con un sito lungo taglia raramente e produce pochi frammenti grandi; uno con un sito corto ne produce molti e piccoli.

ESEMPIO SVOLTO

Tagli attesi in un cromosoma batterico. Il cromosoma di *E. coli* K-12 è lungo circa 4,6 Mb. Quanti frammenti ci si aspetta da una digestione completa con EcoRI, e con AluI (sito di 4 bp)?

Soluzione. Un sito ogni $4^6 = 4096$ bp: $4{,}6 \times 10^6 / 4096 \approx 1{,}1 \times 10^3$ tagli, quindi circa un migliaio di frammenti (in un DNA circolare i frammenti sono tanti quanti i tagli). Con AluI, un sito ogni $4^4 = 256$ bp: $4{,}6 \times 10^6 / 256 \approx 1{,}8 \times 10^4$ frammenti. Il modello è solo una stima: i genomi reali non sono sequenze casuali, e alcuni siti sono più rari o più frequenti del previsto, come vedrai in un esercizio sul DNA del fago λ.

ERRORE FREQUENTE

«Gli enzimi di restrizione tagliano il DNA a caso» oppure «tagliano qualunque DNA, anche quello del batterio che li produce». No: un enzima di restrizione taglia **solo** nel suo sito di riconoscimento e sempre nello stesso punto, per questo due digestioni dello stesso DNA danno gli stessi frammenti. Il batterio protegge il proprio DNA metilando i siti. «Casuale» è semmai la distribuzione dei siti nel genoma, che si può stimare con 4^n.

12.1.4 La DNA ligasi incolla

Se si mescolano due DNA diversi tagliati con lo stesso enzima, le estremità coesive si appaiano fra loro senza badare all'origine dei frammenti: un pezzo di DNA umano si appaia a un pezzo di DNA batterico, perché le sporgenze sono le stesse. I legami a idrogeno fra quattro coppie di basi, però, sono deboli e a temperatura ambiente si aprono di continuo. A rendere permanente l'unione è la DNA ligasi, la stessa che nella duplicazione salda i frammenti di Okazaki: forma un legame fosfodiesterico fra l'ossidrile 3′ di un filamento e il fosfato 5′ del successivo, su ciascuno dei due filamenti, consumando ATP (la ligasi del fago T4, la più usata in laboratorio). Anche le estremità piatte si possono legare, ma con molta minore efficienza, perché nulla le tiene vicine. Due enzimi diversi possono lasciare sporgenze uguali: BamHI (G^GATCC) e BglII (A^GATCT) lasciano entrambi GATC, e i loro frammenti si possono unire.

12.2 Il DNA ricombinante e il clonaggio

12.2.1 Dal primo esperimento ad Asilomar

Una molecola formata unendo in provetta DNA di origini diverse si chiama **DNA ricombinante**. Nel 1972 il gruppo di Paul Berg, a Stanford, unì il DNA del virus SV40 con quello del fago λ, ma decise di non introdurre il prodotto in un batterio: SV40 provoca tumori nei roditori ed *E. coli* vive nel nostro intestino. Nel 1973 Stanley Cohen e Herbert Boyer tagliarono con EcoRI due plasmidi, ciascuno con un gene di resistenza a un antibiotico diverso (tetraciclina e kanamicina), li unirono con la ligasi e li introdussero in *E. coli*: le cellule trasformate resistevano a entrambi gli antibiotici, e passavano la resistenza alle cellule figlie. Era il primo organismo geneticamente modificato con questa tecnica, e il primo clonaggio di DNA ricombinante.

COLLEGAMENTI

Storia e Filosofia · Asilomar 1975: gli scienziati si danno delle regole. Nel 1974 Berg e altri ricercatori chiesero pubblicamente ai colleghi di sospendere alcuni esperimenti di DNA ricombinante finché non se ne fossero valutati i rischi. Nel febbraio 1975 più di cento scienziati, insieme a giuristi e giornalisti, si riunirono nel centro congressi di Asilomar, in California. Non sapevano se un batterio intestinale con geni di virus tumorali potesse diventare pericoloso, e non avevano dati per escluderlo. Decisero di riprendere le ricerche, ma con livelli di **contenimento** proporzionati al rischio presunto: laboratori sempre più protetti (contenimento fisico) e ceppi batterici indeboliti, incapaci di sopravvivere fuori dal laboratorio (contenimento biologico). Le linee guida dei National Institutes of Health statunitensi, del 1976, nacquero da quelle discussioni.

Asilomar è il primo grande esempio di una comunità scientifica che si ferma da sola davanti a un rischio che non sa misurare. Si può discutere con il *principio responsabilità* di Hans Jonas, che chiede di dare più peso alle previsioni negative quando gli effetti di una tecnica possono essere irreversibili (vedi *Filosofia 5*), e con la domanda, di *educazione civica*, su chi debba decidere le regole di una tecnologia nuova: i ricercatori, i governi o i cittadini. Mezzo secolo dopo, la stessa discussione si è riaperta per l'editing genomico (capitolo 13).

Nel 1980 Paul Berg ricevette il premio Nobel per la chimica per i suoi studi sugli acidi nucleici e in particolare sul DNA ricombinante (con Walter Gilbert e Frederick Sanger, premiati per il sequenziamento).

12.2.2 I vettori

Un frammento di DNA da solo non si duplica in una cellula: per essere copiato deve far parte di una molecola che la cellula sa duplicare. Un **vettore** è una molecola di DNA capace di duplicarsi nella cellula ospite, nella quale si inserisce il frammento che interessa. I vettori più usati sono plasmidi batterici modificati, ottimizzati per il laboratorio. Ogni vettore plasmidico ha almeno:

- un'**origine di replicazione** (*ori*), che permette alla DNA polimerasi della cellula di copiarlo, spesso in decine o centinaia di copie per cellula;
- un **marcatore di selezione**, di solito un gene di resistenza a un antibiotico (per esempio all'ampicillina, amp^R), che permette di riconoscere le cellule che hanno ricevuto il plasmide;
- un sito di clonaggio multiplo, un breve tratto che contiene uno dopo l'altro i siti di molti enzimi di restrizione, ciascuno presente una sola volta nel plasmide: lì si apre il vettore per inserire il frammento.

Per frammenti più lunghi si usano vettori derivati dal fago λ, che accettano fino a circa 20 kb, e cromosomi artificiali batterici o di lievito, che ne accettano centinaia di migliaia. Per portare

geni in cellule animali o umane si usano invece **vettori virali**, virus resi incapaci di replicarsi ma ancora capaci di entrare nelle cellule: sono lo strumento della terapia genica (capitolo 13).

12.2.3 Le tappe del clonaggio

Clonare un gene significa ottenerne molte copie identiche facendolo moltiplicare in una cellula ospite. Le tappe sono quattro (Figura 12.2):

1. **taglio**: il vettore e il DNA che contiene il gene vengono tagliati con lo stesso enzima di restrizione (o con due enzimi diversi, così il frammento si inserisce in un solo orientamento);
2. **legame**: si mescolano i frammenti, le estremità coesive si appaiano e la DNA ligasi le salda; si ottiene, fra gli altri prodotti, il plasmide ricombinante;
3. **trasformazione**: si mettono i plasmidi a contatto con batteri resi permeabili al DNA, con un trattamento con ioni calcio seguito da un rapido riscaldamento («shock termico») oppure con brevi impulsi elettrici (elettroporazione); solo una piccola frazione delle cellule assorbe un plasmide;
4. **selezione**: si seminano i batteri su un terreno con l'antibiotico. Crescono solo le cellule che contengono un plasmide; ciascuna, dividendosi, forma una colonia, cioè un *clone* di cellule identiche che contengono lo stesso plasmide.

L'antibiotico, però, non distingue un plasmide ricombinante da uno che si è richiuso su sé stesso senza inserto. Per riconoscerli molti vettori usano lo **screening blu/bianco**: il sito di clonaggio si trova all'interno di un frammento del gene *lacZ*, che codifica una parte della β-galattosidasi, l'enzima dell'operone *lac* (vedi *Scienze naturali 3*, cap. 15). Sul terreno si aggiunge X-gal, una sostanza incolore che l'enzima trasforma in un prodotto blu. Le colonie con un plasmide vuoto producono l'enzima e diventano blu; quelle con un inserto, che interrompe il gene, restano bianche.

12.2.4 Geni eucarioti in batteri: il cDNA

Se si vuole che un batterio produca una proteina umana, non basta inserirgli il gene umano. I geni eucarioti contengono introni, che nel nucleo vengono rimossi dallo splicing; i batteri non hanno lo splicing e tradurrebbero un mRNA con gli introni ancora dentro, ottenendo una proteina sbagliata. La soluzione è partire dall'mRNA maturo, che gli introni non li ha più: la trascrittasi inversa lo copia in un DNA a singolo filamento, che una DNA polimerasi trasforma poi in doppio filamento. Il prodotto si chiama **cDNA** (DNA complementare) e contiene soltanto le sequenze degli esoni, senza gli introni. Il cDNA ha anche un vantaggio pratico: rispecchia i geni che sono *espressi* in quel tessuto.

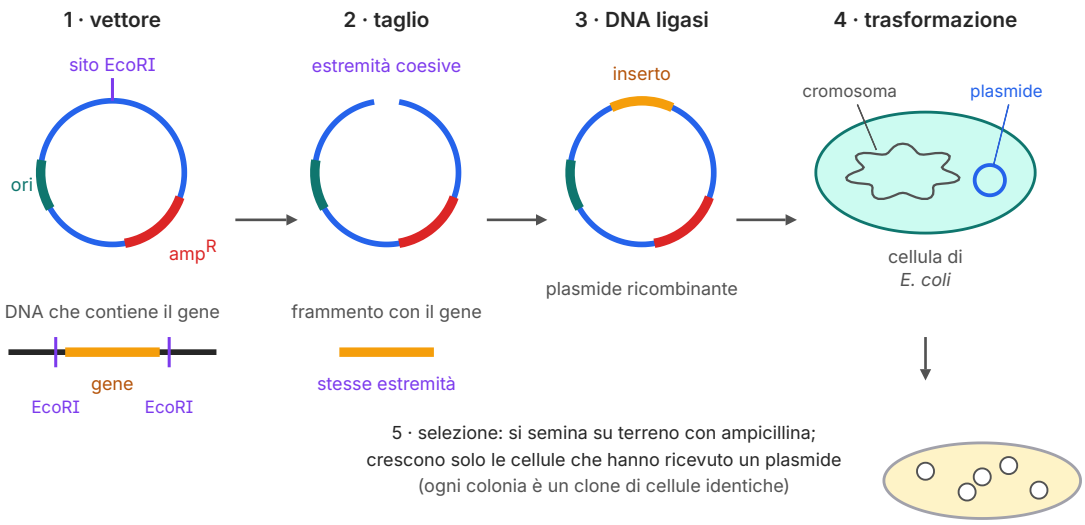


Figura 12.2 Le tappe del clonaggio in un plasmide (schema non in scala). Il vettore ha un'origine di replicazione (*ori*, verde), un gene di resistenza all'ampicillina (amp^R, rosso) e un sito per EcoRI. Vettore e DNA con il gene (arancio) sono tagliati con EcoRI; le estremità coesive si appaiano e la DNA ligasi le salda. Il plasmide ricombinante entra in una cellula di *E. coli*; sul terreno con ampicillina crescono solo le cellule con un plasmide, e ogni colonia è un clone.

Per ottenere la proteina il cDNA si inserisce in un **vettore di espressione**, che contiene a monte del gene un promotore riconosciuto dalla RNA polimerasi dell'ospite e le sequenze per l'inizio della traduzione. Le cellule ospiti si scelgono in base alla proteina: i batteri come *E. coli* crescono in fretta e costano poco, ma non aggiungono catene di zuccheri alle proteine e non sempre le ripiegano bene; i lieviti, eucarioti unicellulari, fanno una parte di queste modifiche; le cellule di mammifero in coltura producono le proteine più simili a quelle umane, a costi più alti. Su questa base si producono l'insulina e gli altri farmaci ricombinanti (capitolo 13).

Quando non si sa ancora quale frammento contiene il gene cercato, si clona tutto: una **libreria genomica** è una raccolta di cloni che, insieme, contengono l'intero genoma di un organismo, tagliato in frammenti; una **libreria di cDNA** contiene invece i cDNA di tutti gli mRNA presenti in un tessuto. Il clone giusto si cerca poi con una sonda, come vedremo.

12.3 Copiare il DNA in provetta: la PCR

12.3.1 Gli ingredienti e il ciclo

Il clonaggio moltiplica un frammento in giorni, con batteri vivi. La **reazione a catena della polimerasi**, o **PCR** (*polymerase chain reaction*), lo fa in poche ore in una provetta. L'hai già incontrata in un Approfondimento (*Scienze naturali 3*, cap. 13): Kary Mullis la ideò nel 1983 e ricevette il

premio Nobel per la chimica nel 1993. Qui ne vediamo il funzionamento in dettaglio. In una provetta di pochi microlitri si mettono:

- il **DNA stampo**, che contiene il tratto da copiare, anche in quantità minime;
- due **primer** di DNA sintetico, lunghi di solito 18–25 nucleotidi, complementari alle estremità 3′ del tratto da copiare sui due filamenti: uno si appaia al filamento superiore, l'altro a quello inferiore, e sono orientati l'uno verso l'altro;
- i quattro **desossiribonucleotidi trifosfato** (dATP, dCTP, dGTP, dTTP, insieme **dNTP**), i mattoni dei nuovi filamenti;
- una DNA polimerasi resistente al calore, la Taq polimerasi, del batterio *Thermus aquaticus*, che vive nelle sorgenti calde;
- una soluzione tampone che mantiene il pH e contiene ioni $\mathbf{Mg^{2+}}$, cofattore indispensabile della polimerasi.

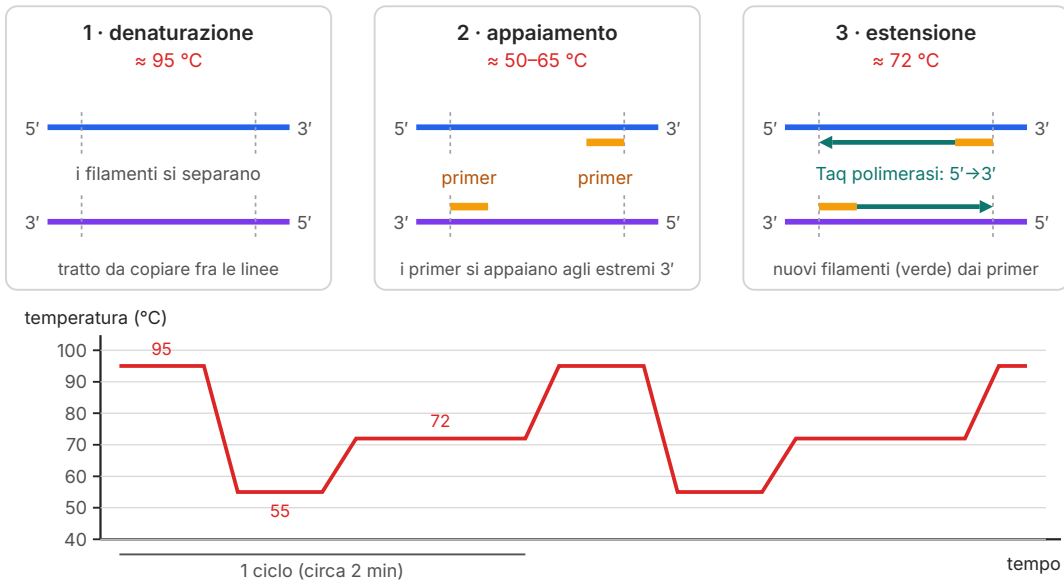


Figura 12.3 Un ciclo di PCR (schema non in scala). 1) A circa 95 °C i due filamenti dello stampo (blu e viola) si separano. 2) A circa 50–65 °C i primer (arancio) si appaiano agli estremi 3′ del tratto da copiare, uno per filamento, orientati l'uno verso l'altro. 3) A circa 72 °C la Taq polimerasi allunga i primer in direzione 5′→3′ (frecce verdi). Sotto, l'andamento della temperatura nel termociclatore per due cicli, con durate tipiche di 30 s, 30 s e 60 s.

La provetta va in un **termociclatore**, uno strumento che cambia la temperatura in pochi secondi e ripete un ciclo di tre fasi (Figura 12.3): 1) **denaturazione** a circa 95 °C, quando i legami a idrogeno si rompono e i due filamenti si separano; 2) **appaiamento** (in inglese *annealing*) a circa 50–65 °C, quando i primer si legano alle sequenze complementari, in eccesso rispetto allo stampo e quindi più rapidi a trovarle dei filamenti lunghi; la temperatura giusta dipende dalla lun-

ghezza e dalla composizione dei primer; 3) **estensione** a circa 72 °C, vicino alla temperatura a cui la Taq è più attiva, quando la polimerasi allunga ciascun primer in direzione 5′→3′ copiando lo stampo. Un ciclo dura da uno a qualche minuto, e una PCR ne prevede di solito da 25 a 35.

La scelta della Taq risolve un problema pratico: a 95 °C quasi tutte le proteine si denaturano (capitolo 8), e nelle prime PCR si doveva aggiungere polimerasi fresca a ogni ciclo. La Taq resiste per decine di cicli: si mette una volta sola e la reazione si automatizza. La Taq, però, non ha l'attività di correzione di bozze (esonucleasica 3′→5′) e sbaglia più spesso delle polimerasi delle cellule; quando serve una copia fedele si usano polimerasi termostabili di altri batteri, che correggono i propri errori.

12.3.2 Una crescita esponenziale, finché dura

All'inizio della reazione lo stampo può essere una molecola lunghissima, ma i primer si appaiano soltanto alle loro sequenze complementari e la polimerasi copia solo a partire da essi. Le copie del primo ciclo cominciano a un primer e proseguono oltre l'altro; dal secondo ciclo, queste copie fanno a loro volta da stampo, e a partire dal **terzo ciclo** compaiono i primi filamenti doppi lunghi *esattamente* quanto il tratto compreso fra i due primer. Da quel momento sono loro a moltiplicarsi più in fretta: dopo una trentina di cicli rappresentano la quasi totalità del DNA nella provetta. Per questo la PCR non copia «tutto il DNA», ma solo il tratto delimitato dai primer, che si chiama *amplicone*.

Poiché ogni ciclo, in condizioni ideali, raddoppia le molecole del tratto bersaglio, dopo n cicli partendo da N_0 copie se ne hanno $N_0 \cdot 2^n$. È una crescita esponenziale, e in scala logaritmica è una retta (Figura 12.4a). Non dura all'infinito: dopo 25–35 cicli si esauriscono i primer e i dNTP, la polimerasi perde attività e i filamenti di amplicone, ormai abbondantissimi, tendono a riappaiarsi fra loro invece che con i primer. La crescita rallenta e la curva arriva a un plateau. È lo stesso andamento del modello logistico di una popolazione che consuma le proprie risorse (vedi *Matematica 5*).

IN SINTESI

La PCR in numeri (efficienza del 100 %, cioè raddoppio a ogni ciclo, nella fase esponenziale):

$$N = N_0 \cdot 2^n \qquad n = \log_2(N/N_0) = \frac{\log(N/N_0)}{\log 2}$$

$$\frac{N_{0,\mathrm{A}}}{N_{0,\mathrm{B}}} = 2^{C_{\mathrm{t,B}} - C_{\mathrm{t,A}}}$$

N_0 = copie iniziali del tratto bersaglio; n = numero di cicli; C_t = ciclo soglia nella PCR in tempo reale. Una differenza di circa 3,3 cicli corrisponde a un fattore 10 (perché $2^{3{,}32} \approx 10$).

ESEMPIO SVOLTO

Quante copie, quanti cicli? a) Un campione contiene 50 copie di un tratto di DNA. Quante copie ci sono dopo 25 cicli ideali? b) Quanti cicli servono, partendo da 50 copie, per superare 10^{10} copie? c) In una PCR in tempo reale il campione A raggiunge la soglia al ciclo 22 e il campione B al ciclo 25. Quante volte più DNA bersaglio conteneva A?

Soluzione. a) $50 \cdot 2^{25} = 50 \cdot 33\,554\,432 \approx 1{,}7 \times 10^{9}$ copie. b) $2^{n} = 10^{10}/50 = 2 \times 10^{8}$, quindi $n = \log(2 \times 10^{8})/\log 2 \approx 8{,}30/0{,}301 \approx 27{,}6$: servono 28 cicli (27 non bastano). In pratica la reazione rallenta prima, e il numero reale di cicli è maggiore. c) A raggiunge la soglia tre cicli prima: conteneva $2^{3} = 8$ volte più copie di B.

12.3.3 RT-PCR e PCR in tempo reale

La Taq copia DNA, non RNA. Per amplificare un RNA, per esempio un mRNA o il genoma di un virus a RNA, prima lo si copia in cDNA con la trascrittasi inversa e poi si fa la PCR sul cDNA: è la **RT-PCR** (da *reverse transcription*, trascrizione inversa). Serve a sapere se un gene è espresso in un tessuto e a rilevare i virus a RNA.

Nella PCR classica si guarda il prodotto alla fine, quando tutte le reazioni sono arrivate al plateau e la quantità finale dice poco su quella iniziale. Nella **PCR in tempo reale** (o quantitativa, qPCR) si aggiunge alla miscela una sostanza che diventa fluorescente quando il prodotto si forma: un colorante che emette luce solo legato al DNA a doppio filamento, oppure una sonda specifica per l'amplicone. Lo strumento misura la fluorescenza a ogni ciclo e registra il **ciclo soglia** (C_t, da *threshold cycle*), il ciclo in cui la fluorescenza supera un valore fissato, ancora nella fase esponenziale (Figura 12.4b). Più copie c'erano all'inizio, prima la soglia viene superata: un C_t basso significa molto bersaglio, un C_t alto poco, e un campione che non raggiunge la soglia dopo 40 cicli si considera negativo.

I test molecolari per SARS-CoV-2 eseguiti sui tamponi sono RT-PCR in tempo reale: l'RNA estratto dal campione viene copiato in cDNA, e la qPCR amplifica tratti di geni del virus scelti dal produttore del test, insieme a un gene umano usato come controllo che il prelievo e l'estrazione abbiano funzionato. Un risultato positivo dice che nel campione c'era RNA virale, non da solo che la persona sia contagiosa: anche frammenti di virus non più attivi possono essere amplificati.

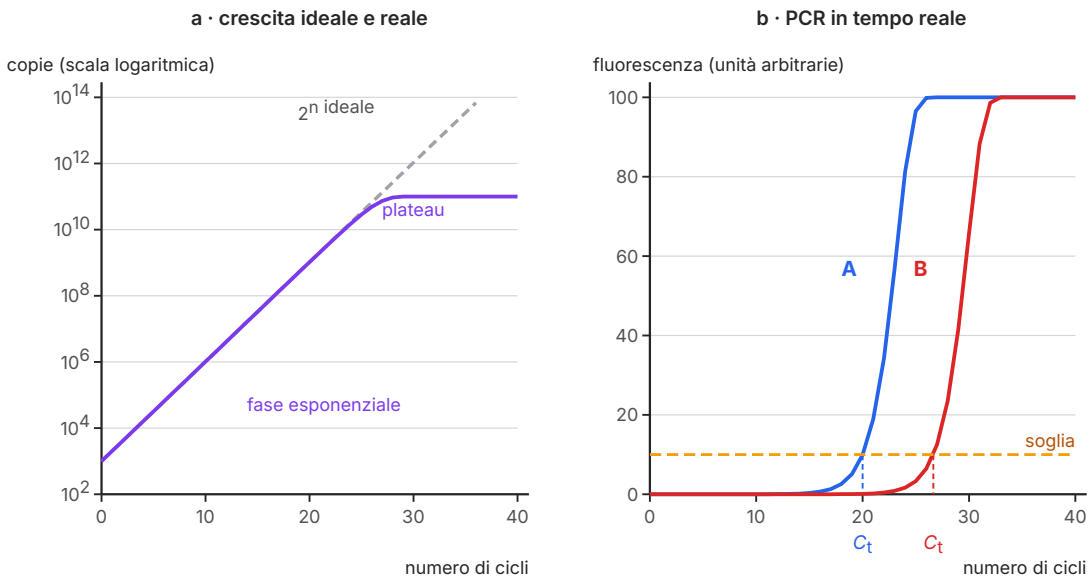


Figura 12.4 La crescita del prodotto in una PCR (modello: raddoppio a ogni ciclo, che rallenta man mano che il prodotto si avvicina al massimo possibile). a) Copie del tratto bersaglio, in scala logaritmica, partendo da 10^3 copie: la crescita ideale 2^n (tratteggiata) è una retta; la reazione reale (viola) la segue nella fase esponenziale e poi si ferma al plateau. b) Fluorescenza in una PCR in tempo reale per due campioni con 10^4 (A) e 10^2 (B) copie iniziali: le curve hanno la stessa forma, spostate di circa 6,6 cicli; la linea arancione è la soglia, e i valori di C_t sono circa 20,0 e 26,7.

APPROFONDIMENTO

Approfondimento · L'efficienza di una PCR quantitativa. Nella realtà un ciclo non raddoppia sempre le copie: le moltiplica per $(1 + E)$, dove l'**efficienza** E vale 1 nel caso ideale e di solito è compresa fra 0,9 e 1. Allora $N = N_0(1 + E)^n$, e al ciclo soglia $N_{\text{soglia}} = N_0(1 + E)^{C_t}$. Prendendo i logaritmi:

$$C_t = \frac{\log N_{\text{soglia}}}{\log(1+E)} - \frac{1}{\log(1+E)} \cdot \log N_0$$

Se si misura il C_t di una serie di campioni a concentrazione nota (diluizioni 1:10 successive) e lo si riporta in funzione di $\log N_0$, si ottiene una retta, detta **curva standard**, con pendenza $m = -1/\log(1 + E)$. Con $E = 1$ la pendenza è $-1/\log 2 \approx -3{,}32$; da una pendenza misurata si ricava $E = 10^{-1/m} - 1$. La stessa retta permette poi di leggere il numero di copie di un campione ignoto dal suo C_t.

12.4 Separare e riconoscere i frammenti

12.4.1 L'elettroforesi su gel

Dopo un taglio con gli enzimi di restrizione o una PCR bisogna vedere che cosa si è ottenuto. Lo strumento è l'**elettroforesi su gel**: la separazione di molecole cariche che si muovono in un gel sotto l'effetto di un campo elettrico. Per il DNA si usa di solito un gel di **agarosio**, un polisaccaride estratto da alghe rosse che, sciolto a caldo in una soluzione tampone e raffreddato, forma una rete di fibre con pori di dimensioni regolabili: più agarosio, pori più piccoli. A un'estremità del gel si scavano dei pozzetti in cui si caricano i campioni.

Ogni nucleotide porta un gruppo fosfato con una carica negativa al pH del tampone, quindi il DNA è carico negativamente e, quando si applica la tensione, si sposta verso il polo positivo (**anodo**). La carica per unità di lunghezza è la stessa per tutti i frammenti, perciò non è la carica a separarli: è il gel, che fa da setaccio. I frammenti corti si muovono fra le fibre più facilmente e arrivano più lontano; quelli lunghi restano indietro. Alla fine, tutte le molecole della stessa lunghezza formano una **banda**.

Il DNA non si vede: per osservarlo si colora il gel con una sostanza che si lega al DNA e diventa fluorescente sotto luce blu o ultravioletta, o con coloranti visibili a occhio nudo, meno sensibili. Per misurare le lunghezze si carica in un pozzetto una scala di riferimento (marcatore di lunghezza, o *ladder*), una miscela di frammenti di lunghezza nota. Entro un certo intervallo, la distanza percorsa è circa proporzionale al **logaritmo** della lunghezza: riportando le bande della scala su un grafico semilogaritmico si ottiene una retta di taratura, da cui si legge la lunghezza di una banda ignota (Figura 12.5).

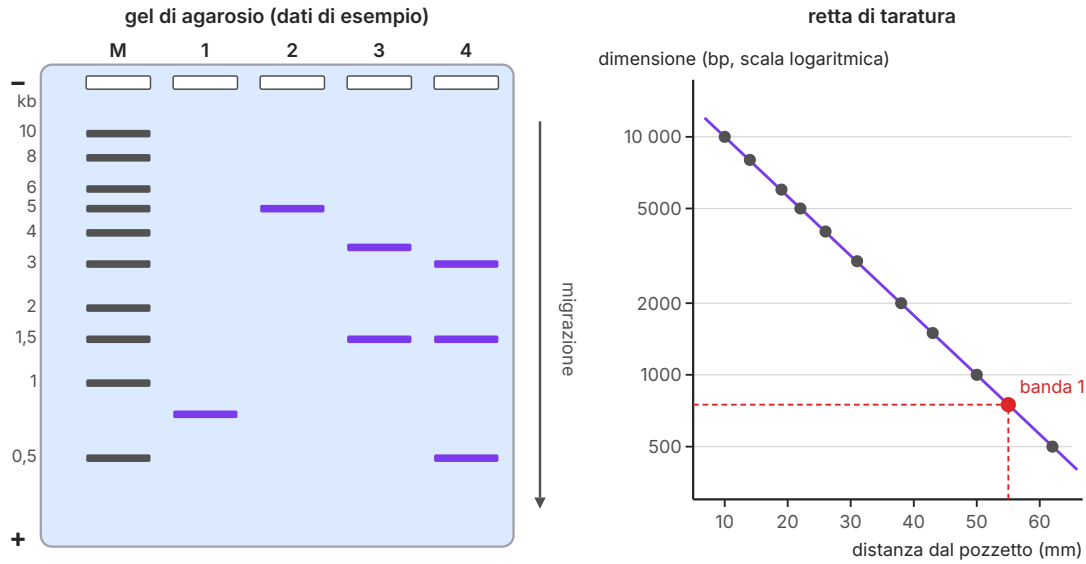


Figura 12.5 Un'elettroforesi su gel di agarosio (dati di esempio). Il DNA entra nel gel dai pozzetti in alto, vicino al polo negativo, e migra verso il polo positivo. Pozzetto M: scala di riferimento con frammenti da 10 a 0,5 kb. Pozzetto 1: il prodotto di una PCR. Pozzetti 2–4: un plasmide circolare tagliato con EcoRI (2), con HindIII (3) e con i due enzimi insieme (4). A destra, la retta di taratura: lunghezza delle bande della scala (scala logaritmica) in funzione della distanza percorsa; il punto rosso è la banda del pozzetto 1, a 55 mm dal pozzetto.

ESEMPIO SVOLTO

Quanto è lunga la banda? Nel gel della Figura 12.5 le bande di 1000 bp e di 500 bp della scala hanno percorso 50,0 mm e 62,0 mm; la banda del pozzetto 1 ha percorso 55,0 mm. Stima la lunghezza del prodotto di PCR.

Soluzione. Fra due bande vicine la relazione fra distanza e logaritmo della lunghezza è lineare, quindi si interpola sul logaritmo, non sulla lunghezza. log 1000 = 3,000; log 500 = 2,699. La banda ignota si trova a (55,0 − 50,0)/(62,0 − 50,0) = 0,417 dell'intervallo, quindi $\log L = 3{,}000 + 0{,}417 \cdot (2{,}699 - 3{,}000) = 2{,}875$ e $L = 10^{2{,}875} \approx 750$ bp. Un'interpolazione lineare sulle lunghezze avrebbe dato 792 bp, un errore del 6 %. Con tutte le bande della scala si può tracciare la retta di regressione, come nel grafico, e il risultato non cambia.

Due errori da evitare quando si legge un gel. Il primo: il DNA migra verso l'**anodo**, il polo positivo, perché è carico negativamente; il gel si orienta sempre con i pozzetti dalla parte del polo negativo. Il secondo: una banda non è una molecola, ma miliardi di molecole della stessa lunghezza; due frammenti diversi lunghi quasi uguali possono finire nella stessa banda, e una banda più intensa può contenerne più di uno.

12.4.2 Sonde e ibridazione

L'elettroforesi separa i frammenti per lunghezza, ma non dice quale contiene un certo gene. Per riconoscerlo si usa una **sonda**: un filamento singolo di DNA o di RNA, marcato con una sostanza fluorescente o radioattiva, complementare alla sequenza cercata. Se si separano i filamenti del campione, la sonda si appaia soltanto alla sequenza complementare: questo appaiamento fra filamenti di origine diversa si chiama **ibridazione**. Nel 1975 Edwin Southern mise a punto il metodo che porta il suo nome: i frammenti separati sul gel vengono denaturati e trasferiti su una membrana, conservando le loro posizioni, e la membrana viene incubata con la sonda. Il Southern blot rivela quali bande contengono la sequenza; per analogia, il metodo che rileva l'RNA si chiama, per scherzo diventato nome ufficiale, *Northern blot*.

Sullo stesso principio funzionano i microarray a DNA: su un vetrino sono fissate, in punti ordinati, migliaia di sonde diverse; il cDNA marcato di un campione si ibrida con quelle complementari, e l'intensità della fluorescenza in ogni punto indica quanto è espresso il gene corrispondente. Così si confronta in un solo esperimento l'attività di migliaia di geni, per esempio fra un tessuto sano e un tumore.

COLLEGAMENTI

Matematica · Esponenziali e logaritmi in laboratorio. La PCR è uno degli esempi più puliti di crescita esponenziale: il numero di copie si moltiplica per un fattore costante a ogni passo, e in scala logaritmica il grafico diventa una retta la cui pendenza è log 2 per ciclo. Il ciclo soglia è in sostanza un logaritmo: C_t diminuisce di circa 3,3 ogni volta che le copie iniziali aumentano di dieci volte. Anche l'elettroforesi usa una scala logaritmica: la distanza percorsa da un frammento dipende in modo approssimativamente lineare dal logaritmo della sua lunghezza. In entrambi i casi un grafico semilogaritmico trasforma una curva in una retta, che si legge e si interpola facilmente.

ESPERIENZA DI LABORATORIO

Una mappa di restrizione su gel. Scopo: separare con l'elettroforesi i frammenti di un DNA tagliato con enzimi di restrizione e ricostruire dalla loro lunghezza la posizione dei siti.

Materiali (kit didattico certificato, preparato e gestito dall'insegnante): campioni di un DNA plasmidico già digeriti con uno o due enzimi e una scala di riferimento; agarosio e tampone di corsa; una camera per elettroforesi con alimentatore a bassa tensione; micropipette; il colorante fornito dal kit; guanti e occhiali. In alternativa si può separare una miscela di coloranti alimentari, che migrano anch'essi perché sono carichi.

Procedimento: 1) L'insegnante prepara il gel all'1 % circa con il pettine per i pozzetti. 2) Si immerge il gel nel tampone, con i pozzetti dal lato del polo negativo. 3) Si caricano lentamente, con la micropipetta, circa 10 µL di ciascun campione e della scala, annotando l'ordine. 4) Si chiude il coperchio, si accende l'alimentatore e si lascia correre finché il colorante di caricamento ha percorso circa tre quarti del gel. 5) A corrente spenta si colora il gel e si misurano le distanze delle bande dai pozzetti. 6) Si costruisce la retta di taratura con la scala su carta semilogaritmica o con un foglio di calcolo, si ricavano le lunghezze e si disegna la mappa.

Sicurezza: l'agarosio fuso è caldo e può ustionare: lo prepara e lo versa l'insegnante. L'alimentatore si accende solo con il coperchio chiuso e si spegne prima di aprire la camera o toccare il gel. Si usano solo coloranti non tossici forniti dal kit, senza transilluminatori a raggi ultravioletti. Guanti e occhiali per tutta l'attività; i materiali si smaltiscono come indicato dall'insegnante.

Risultato atteso e domande: ogni campione dà bande nitide la cui somma è la lunghezza del plasmide. a) Perché la somma delle lunghezze delle bande deve essere sempre la stessa? b) Se un pozzetto dà una banda in più, che cosa può essere successo? c) Perché non si può interpolare linearmente sulle lunghezze? *Laboratori di ricerca*: molte università ed enti di ricerca organizzano per le scuole laboratori di biologia molecolare, in cui si possono svolgere con strumenti professionali una PCR, una digestione e un'elettroforesi.

12.5 Leggere il DNA: il sequenziamento e la genomica

12.5.1 Il metodo di Sanger

Determinare l'ordine delle basi in un tratto di DNA si chiama **sequenziamento**. Nel 1977 il biochimico inglese Frederick Sanger pubblicò il metodo che sarebbe stato usato per trent'anni, e che gli valse nel 1980 il suo secondo premio Nobel per la chimica (il primo, nel 1958, era per la sequenza dell'insulina). L'idea è far copiare il DNA da una polimerasi e interrompere la copia in punti noti.

Oltre ai normali dNTP, nella reazione si aggiunge una piccola quantità di **dideossinucleotidi** (ddNTP): nucleotidi il cui zucchero non ha l'ossidrile né in 2′ né in 3′. La polimerasi li inserisce come gli altri, appaiandoli alla base complementare dello stampo, ma dopo un ddNTP non può

aggiungere altro, perché manca l'ossidrile 3′ a cui legare il nucleotide successivo: la catena termina. Poiché i ddNTP sono pochi rispetto ai dNTP, la terminazione avviene a caso in una molecola o nell'altra, e alla fine nella provetta ci sono frammenti di tutte le lunghezze, ciascuno terminato da un ddNTP che corrisponde alla base in quella posizione. Nella versione originale si facevano quattro reazioni, una per ciascun ddNTP, e i prodotti si separavano su quattro corsie di un gel capace di distinguere frammenti che differiscono di un solo nucleotide (Figura 12.6).

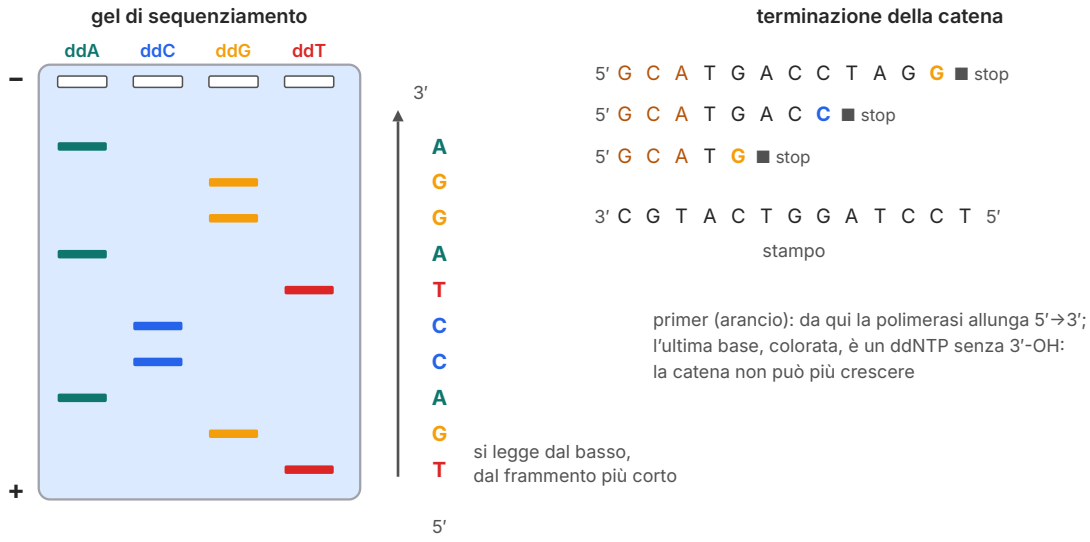


Figura 12.6 Il sequenziamento con il metodo di Sanger (schema). A destra: la polimerasi allunga il primer (arancio) copiando lo stampo; ogni frammento termina dove è stato inserito un dideossinucleotide (in colore), privo dell'ossidrile 3′. A sinistra: i prodotti delle quattro reazioni (ddA, ddC, ddG, ddT) separati su un gel che distingue frammenti che differiscono di un solo nucleotide. Leggendo le bande dal basso, cioè dal frammento più corto, si ottiene la sequenza del nuovo filamento da 5′ a 3′: TGACCTAGGA.

ESEMPIO SVOLTO

Leggere un gel di Sanger. Nella Figura 12.6 leggi la sequenza del filamento sintetizzato e ricava quella del filamento stampo.

Soluzione. Il frammento più corto è quello che ha percorso più strada: è in fondo, nella corsia T, quindi il primo nucleotide aggiunto dopo il primer è T. Risalendo una banda alla volta, qualunque sia la corsia, si legge T, G, A, C, C, T, A, G, G, A. Il nuovo filamento, letto nella direzione in cui è stato sintetizzato, è 5′-TGACCTAGGA-3′. Lo stampo è complementare e antiparallelo: 3′-ACTGGATCCT-5′, cioè, scritto come si usa da 5′ a 3′, 5′-TCCTAGGTCA-3′. L'errore tipico è leggere dall'alto in basso, cioè dalla fine della sequenza all'inizio, o dimenticare che si legge il filamento *copiato*, non lo stampo.

Oggi i quattro ddNTP sono marcati con quattro sostanze fluorescenti di colore diverso, e la reazione si fa in una sola provetta. I frammenti passano in un capillare sottile pieno di gel, in ordine di lunghezza, davanti a un laser e a un rivelatore che registra il colore di ciascuno: il risultato è un grafico a picchi colorati, l'*elettroferogramma*, letto da un programma. Con questi strumenti si leggono fino a circa mille basi per reazione; il metodo di Sanger è ancora il riferimento per controllare un clone o un singolo gene.

12.5.2 Il sequenziamento di nuova generazione

Per leggere un genoma intero, mille basi alla volta sono poche. Il **sequenziamento di nuova generazione** (NGS, da *next-generation sequencing*), diffuso dalla metà degli anni Duemila, cambia strategia: il DNA viene frammentato in milioni di pezzi corti, ciascuno fissato in un punto di una superficie e amplificato sul posto, e tutti vengono sequenziati **in parallelo**, leggendo a ogni ciclo, con una fotografia, quale base è stata aggiunta in ciascun punto. Si ottengono milioni o miliardi di letture brevi, di qualche centinaio di basi al massimo, che un programma ricompone come un puzzle sovrapponendo i tratti in comune. Le tecnologie a **letture lunghe**, come quelle che fanno passare un singolo filamento attraverso un nanoporo e ne riconoscono le basi dalle variazioni di una corrente elettrica, leggono invece molecole di decine di migliaia di basi, utili per le regioni ripetute che le letture corte non riescono a ricomporre (Tabella 12.2). In vent'anni il costo di un genoma umano è sceso di molti ordini di grandezza e il tempo da anni a giorni.

Tabella 12.2 Tre famiglie di metodi di sequenziamento a confronto (valori indicativi, che cambiano rapidamente con la tecnologia).

Metodo	Lunghezza di una lettura	Letture per corsa	Usi tipici
Sanger (capillare)	fino a circa 1000 basi	una per capillare	controllo di cloni e di singoli geni
nuova generazione, letture corte	da circa 50 a qualche centinaio di basi	da milioni a miliardi	genomi interi, espressione genica, diagnosi
letture lunghe (per esempio nanopori)	da decine a centinaia di migliaia di basi	da migliaia a milioni	regioni ripetute, genomi completi

12.5.3 Il genoma umano e la bioinformatica

Il Progetto Genoma Umano, un consorzio pubblico internazionale nato nel 1990, pubblicò una prima bozza della sequenza nel 2001, in contemporanea con l'azienda privata Celera Genomics, e la dichiarò completata nel 2003. «Completa» voleva dire circa il 92 % del genoma: mancavano i tratti più ripetuti, come i centromeri. La prima sequenza davvero continua, da un telomero all'altro di ogni cromosoma (*telomere-to-telomere*, T2T), è stata pubblicata nel 2022, grazie alle letture lunghe; mancava solo il cromosoma Y, completato l'anno dopo. Il genoma umano aploi-

de conta circa 3,1 miliardi di paia di basi e circa 20 000 geni che codificano proteine; le sequenze codificanti ne occupano solo l'1–2 %.

Una sequenza, da sola, è una lunga fila di lettere. Per darle un significato servono la genomica, lo studio dei genomi interi, della loro organizzazione e della loro evoluzione, e la **bioinformatica**, che usa l'informatica per conservare, confrontare e interpretare le sequenze. Le sequenze pubblicate sono raccolte in banche dati liberamente consultabili; un programma come BLAST confronta in pochi secondi una sequenza nuova con milioni di sequenze note e trova le più simili, allineandole base per base. La **genomica comparativa** confronta i genomi di specie diverse: i geni conservati da batteri a esseri umani indicano funzioni fondamentali, e le differenze fra genomi di specie vicine aiutano a ricostruire la storia evolutiva. Oggi si può sequenziare il genoma di una singola persona: un'informazione preziosa per la medicina e delicatissima per la riservatezza, come vedremo nel capitolo 13.

12.6 Riscrivere il DNA: l'editing genomico

12.6.1 Forbici programmabili

Con gli strumenti visti finora si può tagliare il DNA in provetta e inserire geni in un batterio. Modificare con precisione un punto scelto del genoma di una cellula viva, una base fra miliardi, è un'altra impresa. L'**editing genomico** è la modifica mirata della sequenza del DNA in un punto prestabilito del genoma di una cellula viva. La strategia comune è provocare un **taglio a doppio filamento** esattamente nel punto voluto, e lasciare che la cellula lo ripari.

I primi strumenti, le nucleasi a dita di zinco (ZFN) e le TALEN, sono proteine che riconoscono il DNA con una serie di moduli, ciascuno per una o poche basi, legate a un dominio che taglia. Funzionano, ma per ogni nuovo bersaglio bisogna progettare e produrre una proteina nuova, un lavoro di mesi.

12.6.2 CRISPR: da difesa dei batteri a strumento

Nei genomi di molti batteri e archei si trovano brevi sequenze ripetute, separate da tratti diversi l'uno dall'altro: è il locus **CRISPR** (da *clustered regularly interspaced short palindromic repeats*). I tratti fra le ripetizioni sono frammenti di DNA di fagi che hanno infettato la cellula o i suoi antenati, conservati come in un archivio. Quando lo stesso fago ritorna, la cellula trascrive l'archivio in brevi RNA, che guidano proteine Cas (*CRISPR-associated*) verso il DNA del fago con la sequenza complementare, e lo tagliano. È un sistema immunitario adattativo, con una memoria che si trasmette alle cellule figlie.

Nel 2012 Emmanuelle Charpentier e Jennifer Doudna mostrarono che la proteina Cas9 del batterio *Streptococcus pyogenes* taglia in provetta qualsiasi DNA indicato da un RNA, e che i due RNA naturali del sistema si possono unire in un'unica molecola progettata a piacere. Dall'anno

successivo il metodo fu applicato con successo in cellule umane e di molti altri organismi. Nel 2020 Charpentier e Doudna ricevettero il premio Nobel per la chimica.

12.6.3 Come funziona Cas9

Il sistema ha due componenti (Figura 12.7). La proteina **Cas9** è la nucleasi, che taglia. L'**RNA guida** (sgRNA, *single-guide RNA*) ha una parte fissa, che si lega a Cas9, e una parte variabile di 20 nucleotidi che il ricercatore sceglie uguale alla sequenza del bersaglio. Il complesso scorre lungo il DNA e si ferma dove trova una breve sequenza detta **PAM** (*protospacer adjacent motif*), che per la Cas9 di *S. pyogenes* è 5′-NGG-3′ (N = una base qualsiasi). Lì apre la doppia elica: se i 20 nucleotidi dell'RNA guida si appaiano al filamento di DNA complementare, Cas9 taglia entrambi i filamenti tre paia di basi prima della PAM, lasciando estremità piatte. Senza una PAM accanto, Cas9 non taglia nemmeno una sequenza perfettamente complementare: per i batteri è il modo di non tagliare il proprio archivio CRISPR, che non contiene PAM.

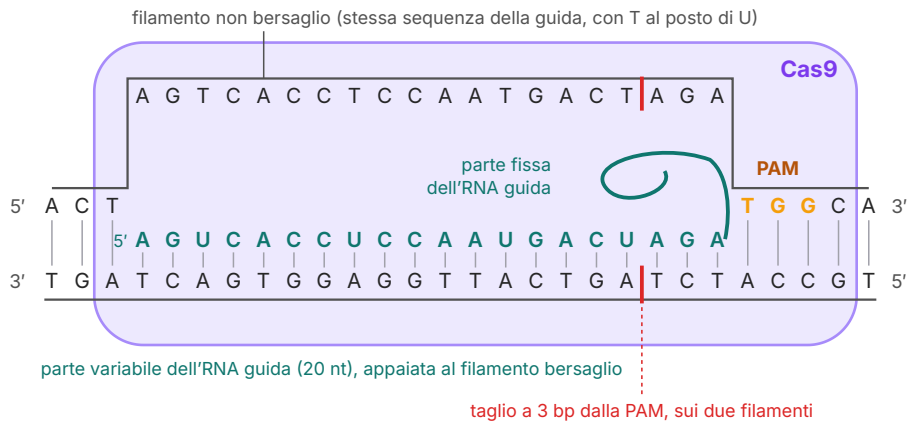


Figura 12.7 Il complesso CRISPR-Cas9 su un DNA bersaglio (schema non in scala). La parte variabile dell'RNA guida (verde), 20 nucleotidi, si appaia al filamento bersaglio, mentre l'altro filamento resta spostato in alto; ha la stessa sequenza dell'RNA guida, con T al posto di U, ed è seguito dalla PAM 5′-TGG-3′ (una sequenza NGG, in arancio). Cas9 (viola) taglia i due filamenti tre paia di basi prima della PAM (linee rosse).

A riparare il taglio è la cellula, con uno dei suoi due sistemi. Il più rapido, l'**unione delle estremità non omologhe** (NHEJ, da *non-homologous end joining*), riattacca le estremità direttamente, ma spesso aggiunge o toglie qualche base: se il taglio cade in un gene, la finestra di lettura si sposta e la proteina non si forma più. È il modo più semplice per «spegnere» un gene (*knock-out*) e scoprirne la funzione. Se invece si fornisce alla cellula un DNA stampo con la sequenza desiderata, affiancata da tratti uguali a quelli intorno al taglio, la cellula può usarlo per la **riparazione per ricombinazione omologa** (HDR, da *homology-directed repair*) e copiare la modifica

voluta: correggere una mutazione, inserire un gene. L'HDR è precisa ma meno frequente, soprattutto nelle cellule che non si dividono.

12.6.4 Limiti e nuove versioni

Cas9 non è infallibile. Può tagliare anche in punti del genoma con una sequenza simile, ma non identica, a quella della guida: sono gli **effetti fuori bersaglio** (*off-target*), che si riducono scegliendo guide uniche nel genoma e usando varianti di Cas9 più selettive, e che si cercano sequenziando le cellule modificate. Inoltre in un tessuto o in un embrione non tutte le cellule vengono modificate allo stesso modo (mosaicismo). Per evitare il taglio a doppio filamento sono nati l'*editing di base*, che trasforma direttamente una base in un'altra, e il *prime editing*, che riscrive un breve tratto usando un RNA guida che porta anche lo stampo della modifica. Le applicazioni mediche e agricole, con le questioni etiche che pongono, sono l'argomento del capitolo 13.

ERRORE FREQUENTE

«CRISPR modifica il DNA con precisione assoluta: si scrive la correzione e la cellula la esegue». No: Cas9 si limita a **tagliare** dove trova la sequenza della guida accanto a una PAM, a volte anche in punti simili (effetti fuori bersaglio); il risultato dipende da come la cellula ripara il taglio, e con la NHEJ è un piccolo inserimento o una delezione non controllati. Per questo ogni esperimento di editing si conclude sequenziando il DNA delle cellule modificate.

SINTESI

- Gli **enzimi di restrizione** sono una difesa dei batteri contro i fagi: tagliano il DNA a doppio filamento in siti di riconoscimento specifici, di solito palindromi di 4–8 bp; il batterio protegge il proprio DNA metilandolo.
- Un taglio sfalsato lascia **estremità coesive**, uno al centro del sito **estremità piatte**. Per un sito di n bp la lunghezza media dei frammenti, in una sequenza casuale, è 4^n. La DNA ligasi salda i frammenti con legami fosfodiesterici.
- Il **DNA ricombinante** (Berg 1972; Cohen e Boyer 1973) unisce DNA di origini diverse; ad Asilomar, nel 1975, gli scienziati stabilirono regole di contenimento proporzionate al rischio.
- Un **vettore** plasmidico ha un'origine di replicazione, un marcatore di selezione e un sito di clonaggio multiplo. Clonare un gene: taglio, legame, trasformazione, selezione; lo screening blu/bianco distingue i plasmidi con l'inserto.
- Per esprimere un gene eucariote in un batterio si usa il **cDNA**, copiato dall'mRNA maturo con la trascrittasi inversa, perché i batteri non eliminano gli introni; la proteina si produce con un vettore di espressione in batteri, lieviti o cellule di mammifero.
- La **PCR** copia il tratto compreso fra due primer con cicli di denaturazione (≈ 95 °C), appaiamento (≈ 50–65 °C) ed estensione (≈ 72 °C) grazie alla Taq polimerasi termostabile; le copie crescono come $N_0 \cdot 2^n$ fino al plateau.
- La **RT-PCR** amplifica un RNA dopo averlo copiato in cDNA; la **PCR in tempo reale** misura la fluorescenza a ogni ciclo, e il ciclo soglia C_t è tanto più basso quanto più bersaglio c'era all'inizio (3,3 cicli ≈ fattore 10).
- Nell'**elettroforesi su gel** il DNA, carico negativamente, migra verso l'anodo; i frammenti corti vanno più lontano, e la lunghezza si ricava dalla distanza con una retta di taratura in scala logaritmica.
- Una **sonda** marcata riconosce per ibridazione una sequenza complementare (Southern blot, microarray).
- Il **metodo di Sanger** (1977) usa dideossinucleotidi senza ossidrile 3′ che terminano la catena; leggendo i frammenti dal più corto si ottiene la sequenza del filamento copiato da 5′ a 3′. Il sequenziamento di nuova generazione legge in parallelo milioni di frammenti.
- Il genoma umano (bozza 2001, completamento 2003, sequenza continua T2T 2022) conta circa 3,1 miliardi di bp; la **bioinformatica** confronta e interpreta le sequenze.
- Nell'**editing genomico** con CRISPR-Cas9 (Charpentier e Doudna 2012) un RNA guida di 20 nucleotidi porta Cas9 su un bersaglio adiacente a una PAM NGG; il taglio a doppio filamento è riparato per NHEJ (knock-out) o per HDR con uno stampo (modifica precisa); restano gli effetti fuori bersaglio.

ESERCIZI

Comprendere e ricordare

1. ● Nei batteri che li producono, gli enzimi di restrizione servono a:
   - **A.** duplicare i plasmidi prima della divisione
   - **B.** tagliare il DNA dei batteriofagi che entra nella cellula
   - **C.** riparare le rotture del cromosoma
   - **D.** tagliare il cromosoma batterico durante la divisione

2. ●● Il cromosoma di un batterio non viene tagliato dai suoi stessi enzimi di restrizione perché:
   - **A.** non contiene sequenze palindrome
   - **B.** è a singolo filamento
   - **C.** nei siti di riconoscimento alcune basi sono metilate
   - **D.** è avvolto intorno agli istoni

3. ● Quale di queste sequenze, scritte da 5′ a 3′ su un filamento di un DNA a doppio filamento, è un sito palindromo?
   - **A.** 5′-GGATCG-3′
   - **B.** 5′-AAGCCT-3′
   - **C.** 5′-GTTAAG-3′
   - **D.** 5′-AAGCTT-3′

4. ● In un ciclo di PCR la fase a circa 72 °C serve a:
   - **A.** separare i due filamenti dello stampo
   - **B.** far appaiare i primer allo stampo
   - **C.** allungare i primer con la Taq polimerasi
   - **D.** inattivare la polimerasi alla fine del ciclo

5. ● Nell'elettroforesi su gel di agarosio i frammenti di DNA:
   - **A.** migrano verso il polo positivo, e i più corti arrivano più lontano
   - **B.** migrano verso il polo negativo, e i più corti arrivano più lontano
   - **C.** migrano verso il polo positivo, e i più lunghi arrivano più lontano
   - **D.** si separano in base alla carica, diversa per ogni sequenza

6. ●● Nel metodo di Sanger un dideossinucleotide (ddNTP) interrompe l'allungamento del filamento perché:
   - **A.** non ha il gruppo fosfato
   - **B.** manca dell'ossidrile sul carbonio 3′ dello zucchero
   - **C.** non si appaia ad alcuna base dello stampo
   - **D.** non contiene la base azotata

7. ●● Per far produrre a *E. coli* una proteina umana si inserisce nel vettore il cDNA del gene e non il gene preso dal genoma perché:
   A. il cDNA è più stabile del DNA genomico
   B. il gene genomico è troppo lungo per entrare in un plasmide
   C. il cDNA codifica una proteina più attiva
   D. i batteri non hanno lo splicing e non potrebbero eliminare gli introni

8. ●● Nel sistema CRISPR-Cas9 la PAM è:
   A. una breve sequenza del DNA, accanto al tratto riconosciuto dalla guida, senza la quale Cas9 non taglia
   B. la parte dell'RNA guida che si lega a Cas9
   C. il punto in cui Cas9 taglia il DNA
   D. lo stampo usato dalla cellula per la riparazione omologa

9. ●●● In una PCR in tempo reale il campione A raggiunge la soglia al ciclo 20 e il campione B al ciclo 26. Se l'efficienza è del 100 %:
   A. A conteneva circa 6 volte meno bersaglio di B
   B. B conteneva circa 64 volte più bersaglio di A
   C. A conteneva circa 64 volte più bersaglio di B
   D. i due campioni contenevano la stessa quantità di bersaglio

Vero o falso

10. ● La DNA ligasi unisce due frammenti con estremità coesive formando legami a idrogeno fra le basi delle sporgenze.

11. ● In una PCR ideale il numero di copie del tratto bersaglio raddoppia a ogni ciclo.

12. ●● Nel metodo di Sanger il frammento più corto termina con la base che si trova all'estremità 3′ del filamento sintetizzato.

13. ●● Un vettore plasmidico senza origine di replicazione non viene copiato nella cellula ospite e si perde con le divisioni.

14. ●● Cas9 guidata da un RNA taglia qualunque sequenza di DNA complementare alla guida, anche se non è seguita da una PAM.

15. ● Con la RT-PCR si può rilevare un virus il cui genoma è fatto di RNA.

Applicare

16. ● a) Calcola la lunghezza media dei frammenti prodotti, in una sequenza casuale, da AluI (sito di 4 bp), EcoRI (6 bp) e NotI (8 bp). b) Quale dei tre enzimi useresti per tagliare un genoma in pochi grandi frammenti?

17. ●● Un campione contiene 200 copie di un gene. a) Quante copie ci sono dopo 28 cicli di una PCR ideale? b) Quanti cicli servono, nel caso ideale, per superare 10^9 copie? c) Se ogni ciclo dura 2,5 minuti, quanto dura una reazione di 30 cicli?

18. ●●● Un DNA lineare di 10,0 kb viene tagliato con EcoRI e dà due frammenti di 3,0 e 7,0 kb; con BamHI dà due frammenti di 4,0 e 6,0 kb; con i due enzimi insieme dà tre frammenti di 3,0, 1,0 e 6,0 kb. Disegna la mappa di restrizione, indicando la distanza di ciascun sito da un'estremità, e spiega perché l'altra posizione possibile per BamHI va scartata.

19. ●● Vuoi amplificare con la PCR l'intero tratto seguente (filamento superiore, 54 bp):
5′-ATGGCTAGC AAGGAGGAA GTTCTGACC GATCATGCT GGTCAAGCC TTAGGTACG-3′
Scrivi, da 5′ a 3′, la sequenza dei due primer di 18 nucleotidi e indica la lunghezza del prodotto.

20. ●● Le quattro reazioni di Sanger di un frammento danno, dopo il primer, prodotti terminati alle posizioni: ddA: 2, 6, 7; ddC: 1, 9; ddG: 3, 5, 10; ddT: 4, 8 (la posizione 1 è il primo nucleotide aggiunto). Scrivi la sequenza del filamento sintetizzato e quella del filamento stampo, entrambe da 5′ a 3′.

21. ●●● Questo è un filamento di un tratto di DNA che vuoi tagliare con CRISPR-Cas9 (considera solo questo filamento):
5′-ATTCAGTACG TTAGCATCTA ACTGTGGATC-3′
a) Trova la PAM. b) Scrivi da 5′ a 3′ la parte variabile (20 nucleotidi) dell'RNA guida. c) Fra quali basi di questo filamento avverrà il taglio?

22. ●●● In un gel le bande della scala di 3000, 2000 e 1000 bp hanno percorso 28,3 mm, 34,5 mm e 45,0 mm. Una banda ignota ha percorso 38,0 mm. Stima la sua lunghezza interpolando sul logaritmo.

Leggere dati e figure

23. ●● Osserva la Figura 12.4b. a) Stima il ciclo soglia dei campioni A e B. b) Di quanti cicli differiscono, e quale rapporto fra le copie iniziali ne ricavi? Confrontalo con i valori indicati nella didascalia. c) Perché le due curve arrivano allo stesso plateau, anche se i campioni erano molto diversi? Che cosa ti dice questo sulla PCR classica letta solo alla fine?

24. ●●● Nella Figura 12.5 i pozzetti 2, 3 e 4 contengono lo stesso plasmide circolare tagliato con EcoRI, con HindIII e con i due enzimi insieme. a) Con la scala, stima le lunghezze delle bande. b) Quanto è lungo il plasmide, e quanti siti ha ciascun enzima? (Ricorda: in un DNA circolare i frammenti sono tanti quanti i tagli.) c) Disegna la mappa del plasmide.

25. ●●● Il DNA del fago λ è una molecola lineare di 48 502 bp. Digerito con EcoRI dà sei frammenti, di 21 226, 7421, 5804, 5643, 4878 e 3530 bp. a) Verifica che la loro somma è la lunghezza del genoma. b) Quanti siti di EcoRI ha il DNA di λ, e quanti ne prevederebbe il modello 4^n? c) Proponi una spiegazione della differenza. d) Quali due frammenti sarebbe difficile separare su un gel di agarosio, e perché?

26. ●● Osserva la Figura 12.7. a) Su quale filamento si trova la PAM, e quale sequenza ha? b) Scrivi da 5′ a 3′ la sequenza di DNA del filamento bersaglio appaiato all'RNA guida. c) Il taglio di Cas9 lascia estremità coesive o piatte? d) Che cosa succederebbe se una mutazione trasformasse la PAM in TCG?

Indagare e argomentare

27. ●● Un laboratorio esegue RT-PCR in tempo reale per un virus a RNA. Accanto ai campioni dei pazienti, in ogni seduta si analizzano un campione senza alcun acido nucleico (solo acqua e reagenti), un campione con una quantità nota di RNA virale e, in ogni campione, un gene umano. Spiega a che cosa serve ciascuno dei tre controlli e quale errore impedisce di commettere.

28. ●●● Nelle prime PCR, a metà degli anni Ottanta, si usava una polimerasi di *E. coli* e bisognava aggiungerne di nuova dopo ogni denaturazione. Spiega perché, usando ciò che sai sulla struttura delle proteine e sulla denaturazione, e perché la scoperta di una polimerasi in un batterio delle sorgenti calde ha cambiato la tecnica. Discuti anche perché un batterio che vive a 70 °C ha enzimi stabili a quella temperatura (un argomento evolutivo).

29. ●●● Un gruppo di ricerca vuole sapere a che cosa serve un gene in cellule umane in coltura. Progetta l'esperimento con CRISPR-Cas9: quale via di riparazione sfrutti e perché ottieni la perdita di funzione, come scegli la guida per ridurre gli effetti fuori bersaglio, quale controllo usi, e come verifichi con le tecniche del capitolo che il gene è stato davvero modificato.

Verso l'esame

30. ●●● **Dal documento al colloquio.** *Documento* (testo della redazione): «Nel febbraio 1975 più di cento biologi si chiusero per quattro giorni in un centro congressi sulla costa della California. Avevano imparato da poco a tagliare e incollare il DNA e a farlo moltiplicare nei batteri, e non sapevano dire se quei batteri potessero diventare un pericolo. Non c'erano dati per rispondere. Ne uscirono con una decisione insolita: la ricerca poteva continuare, ma con precauzioni proporzionate al rischio immaginato, finché i dati non avessero detto di più.» a) Quali tecniche, descritte nel capitolo, avevano reso possibili quegli esperimenti, e quale timore suscitavano? b) Che cosa significano «contenimento fisico» e «contenimento biologico»? c) *Filosofia*: confronta la scelta di Asilomar con il principio responsabilità di Hans Jonas. d) *Educazione civica*: chi dovrebbe decidere oggi le regole per l'editing genomico, e con quali criteri?

31. ●●● **Collegare.** «La PCR è la duplicazione del DNA senza cellula.» Sviluppa questa frase in un breve percorso orale che colleghi: la duplicazione semiconservativa e i suoi enzimi (*Scienze naturali 3*, cap. 13), le proprietà degli enzimi e la denaturazione delle proteine (capitolo 8), la crescita esponenziale (Matematica) e almeno un'applicazione della PCR fra quelle del capitolo 13. Indica che cosa, nella cellula, fanno l'elicasi, la primasi e la DNA polimerasi, e che cosa le sostituisce nella provetta.

Approfondimento

32. ●●● (Approfondimento) In una PCR quantitativa la curva standard, cioè il ciclo soglia in funzione del logaritmo delle copie iniziali, ha pendenza −3,45. a) Calcola l'efficienza della reazione. b) Di quanti cicli differiscono, con questa efficienza, due campioni le cui copie iniziali stanno nel rapporto 10 : 1?

PER APPROFONDIRE

- [BRO] T. A. Brown, *Gene Cloning and DNA Analysis: An Introduction*: i capitoli sugli enzimi di restrizione, sui vettori, sulla PCR e sul sequenziamento, con molti schemi di esperimenti.
- [ALB] B. Alberts et al., *Molecular Biology of the Cell*: il capitolo sui metodi per studiare i geni, dalla clonazione all'editing con CRISPR.
- [STR] J. M. Berg, J. L. Tymoczko, L. Stryer, *Biochemistry*: il capitolo sulle tecnologie del DNA, con la chimica dei dideossinucleotidi e la PCR quantitativa.

Glossario del capitolo

Bioinformatica
Disciplina che usa metodi informatici per conservare, confrontare e interpretare sequenze di DNA, RNA e proteine (ingl. *bioinformatics*).

Clonaggio genico
Produzione di molte copie identiche di un frammento di DNA inserendolo in un vettore e facendolo moltiplicare in cellule ospiti (ingl. *gene cloning*).

CRISPR-Cas9
Sistema derivato dall'immunità adattativa dei batteri in cui un RNA guida porta la nucleasi Cas9 su una sequenza complementare, che viene tagliata a doppio filamento (ingl. *CRISPR-Cas9*).

Dideossinucleotide
Nucleotide privo dell'ossidrile in 3′ (e in 2′): inserito in un filamento, ne interrompe l'allungamento; è la base del metodo di Sanger (ingl. *dideoxynucleotide*).

DNA complementare (cDNA)
DNA copiato da un mRNA maturo con la trascrittasi inversa e reso a doppio filamento da una DNA polimerasi; contiene le sequenze degli esoni, senza gli introni (ingl. *complementary DNA*).

DNA ricombinante
Molecola di DNA ottenuta in laboratorio unendo tratti di DNA di origini diverse (ingl. *recombinant DNA*).

Editing genomico
Modifica mirata della sequenza del DNA in un punto prestabilito del genoma di una cellula viva (ingl. *genome editing*).

Elettroforesi su gel
Separazione di molecole cariche in un gel sotto l'azione di un campo elettrico; i frammenti di DNA migrano verso l'anodo con velocità che diminuisce al crescere della lunghezza (ingl. *gel electrophoresis*).

Enzima di restrizione
Endonucleasi batterica che riconosce una breve sequenza specifica del DNA a doppio filamento, di solito palindroma, e la taglia sempre nella stessa posizione; nei batteri è una difesa contro i fagi (ingl. *restriction enzyme*).

Estremità coesive / piatte
Estremità di un DNA tagliato con un breve tratto a singolo filamento sporgente, capace di appaiarsi a una sporgenza complementare (coesive), oppure senza basi spaiate (piatte) (ingl. *sticky ends / blunt ends*).

Libreria genomica
Raccolta di cloni che contengono, nel loro insieme, tutti i frammenti del genoma di un organismo; una libreria di cDNA contiene i cDNA degli mRNA di un tessuto (ingl. *genomic library*).

Marcatore di selezione
Gene del vettore, spesso di resistenza a un antibiotico, che permette di far crescere solo le cellule che hanno ricevuto il vettore (ingl. *selectable marker*).

PAM
Breve sequenza del DNA adiacente al bersaglio (NGG per la Cas9 più usata), necessaria perché Cas9 riconosca e tagli il sito (ingl. *protospacer adjacent motif*).

PCR
Reazione a catena della polimerasi: amplificazione in provetta del tratto di DNA compreso fra due primer, con cicli di denaturazione, appaiamento ed estensione (ingl. *polymerase chain reaction*).

PCR in tempo reale
PCR in cui si misura la fluorescenza del prodotto a ogni ciclo; il ciclo soglia, in cui la fluorescenza supera un valore fissato, indica la quantità iniziale di bersaglio (ingl. *real-time PCR, qPCR*).

RT-PCR
PCR preceduta dalla copia di un RNA in cDNA con la trascrittasi inversa (ingl. *reverse-transcription PCR*).

Sito di riconoscimento
Sequenza di DNA, di 4–8 bp per gli enzimi più usati, riconosciuta e tagliata da un enzima di restrizione (ingl. *recognition site*).

Sonda
Filamento singolo di DNA o RNA marcato, complementare a una sequenza cercata, che la riconosce per ibridazione (ingl. *probe*).

Vettore
Molecola di DNA capace di duplicarsi in una cellula ospite (plasmide, fago, virus, cromosoma artificiale), usata per introdurvi e moltiplicarvi un frammento di DNA (ingl. *vector*).

Vettore di espressione
Vettore che contiene, accanto al gene inserito, le sequenze necessarie perché la cellula ospite lo trascriva e lo traduca (ingl. *expression vector*).

Soluzioni

1. B. Gli enzimi di restrizione tagliano il DNA estraneo, soprattutto quello dei fagi, nei loro siti di riconoscimento: per questo «restringono» la crescita dei fagi.

2. C. La metilasi del sistema di restrizione-modificazione aggiunge un metile a una base di ogni sito del DNA del batterio, che così non viene più riconosciuto.

3. D. Il complementare di AAGCTT, letto da 5′ a 3′, è ancora AAGCTT (è il sito di HindIII). Per le altre: GGATCG → CGATCC, AAGCCT → AGGCTT, GTTAAG → CTTAAC.

4. C. A 72 °C la Taq allunga i primer da 5′ a 3′; a 95 °C i filamenti si separano, a 50–65 °C si appaiano i primer.

5. A. I gruppi fosfato rendono il DNA negativo: migra verso l'anodo. La carica per nucleotide è la stessa per tutti i frammenti, e il gel rallenta di più quelli lunghi.

6. B. Senza l'ossidrile 3′ la polimerasi non può formare il legame fosfodiesterico con il nucleotide successivo.

7. D. Il cDNA, copiato dall'mRNA maturo, non contiene introni; i batteri non hanno il macchinario dello splicing.

8. A. La PAM (NGG per la Cas9 di *S. pyogenes*) sta nel DNA, subito dopo il tratto uguale alla guida; Cas9 taglia tre paia di basi prima di essa.

9. C. A raggiunge la soglia 6 cicli prima: $2^6 = 64$, quindi A conteneva circa 64 volte più bersaglio.

10. F. I legami a idrogeno fra le sporgenze si formano da soli; la ligasi forma i legami **fosfodiesterici** fra l'ossidrile 3′ e il fosfato 5′ dei nucleotidi vicini, su ciascun filamento.

11. V. In condizioni ideali ogni molecola del tratto bersaglio dà due molecole a ogni ciclo: $N = N_0 \cdot 2^n$. Nella realtà l'efficienza è un po' minore e alla fine si arriva al plateau.

12. F. Il frammento più corto contiene solo il primo nucleotide aggiunto dopo il primer, che è vicino all'estremità **5′** del filamento sintetizzato; la sequenza si legge dal più corto al più lungo, da 5′ a 3′.

13. V. Senza origine di replicazione la DNA polimerasi della cellula non inizia a copiarlo: il plasmide non si moltiplica e viene diluito fra le cellule figlie fino a scomparire.

14. F. Senza una PAM accanto al bersaglio Cas9 non taglia, anche se la guida è perfettamente complementare; d'altra parte può tagliare sequenze simili alla guida accanto a una PAM (effetti fuori bersaglio).

15. V. La trascrittasi inversa copia l'RNA del virus in cDNA, che poi la PCR amplifica: è il principio dei test molecolari per SARS-CoV-2.

16. a) AluI: $4^4 = 256$ bp; EcoRI: $4^6 = 4096$ bp ≈ 4,1 kb; NotI: $4^8 = 65\,536$ bp ≈ 66 kb. b) NotI: il sito più lungo è il più raro e dà pochi frammenti grandi.

17. a) $200 \cdot 2^{28} = 200 \cdot 268\,435\,456 \approx 5{,}4 \times 10^{10}$ copie. b) $2^n = 10^9/200 = 5 \times 10^6$; $n = \log(5 \times 10^6)/\log 2 \approx 22{,}3$: servono 23 cicli. c) 30 · 2,5 min = 75 min, cioè 1 h 15 min (più qualche minuto di denaturazione iniziale ed estensione finale).

18. EcoRI taglia a 3,0 kb da un'estremità (chiamiamola sinistra). BamHI taglia a 4,0 kb da una delle due estremità. Se il sito di BamHI fosse a 4,0 kb da sinistra, la doppia digestione darebbe 3,0 (da 0 a 3,0), 1,0 (da 3,0 a 4,0) e 6,0 kb (da 4,0 a 10,0): è il risultato osservato. Se fosse a 4,0 kb da destra, cioè a 6,0 kb da sinistra, darebbe 3,0, 3,0 e 4,0 kb, che non si osserva. Mappa: EcoRI a 3,0 kb e BamHI a 4,0 kb dalla stessa estremità, a 1,0 kb l'uno dall'altro (o la stessa mappa letta dall'altra estremità: 6,0 e 7,0 kb).

19. Primer diretto = le prime 18 basi del filamento superiore: 5′-ATGGCTAGCAAGGAGGAA-3′. Primer inverso = complementare inverso delle ultime 18 basi: le ultime 18 basi sono 5′-GGTCAAGCCTTAGGTACG-3′, e il primer che si appaia a esse è 5′-CGTACCTAAGGCTTGACC-3′. Il prodotto comprende anche i primer: 54 bp.

20. Ordinando le posizioni da 1 a 10: C A G T G A A T C G. Filamento sintetizzato: 5′-CAGTGAATCG-3′. Stampo (complementare e antiparallelo): 3′-GTCACTTAGC-5′, cioè 5′-CGATTCACTG-3′.

21. a) Nel filamento l'unica sequenza NGG è TGG (basi 25–27). b) I 20 nucleotidi che la precedono sono AGTACGTTAGCATCTAACTG; l'RNA guida ha la stessa sequenza con U al posto di T: 5′-AGUACGUUAGCAUCUAACUG-3′. c) Tre paia di basi prima della PAM: fra la A e la C di ...TAA|CTGTGG... (fra le basi 21 e 22); anche l'altro filamento è tagliato nella posizione corrispondente, con estremità piatte.

22. log 2000 = 3,301; log 1000 = 3,000. La banda è a (38,0 − 34,5)/(45,0 − 34,5) = 0,333 dell'intervallo: log L = 3,301 − 0,333 · 0,301 = 3,201; $L = 10^{3,201} \approx 1{,}6 \times 10^3$ bp, cioè circa 1,6 kb.

23. a) A ≈ 20, B ≈ 27 (dalla didascalia 20,0 e 26,7). b) Circa 6,6 cicli: $2^{6{,}6} \approx 100$, cioè A conteneva circa 100 volte più copie, come le 10^4 e 10^2 copie della didascalia. c) Il plateau dipende dall'esaurimento dei reagenti e dall'attività della polimerasi, non dalla quantità iniziale: alla fine tutte le reazioni producono una quantità simile. Per questo la PCR classica, letta solo alla fine, dice se il bersaglio c'era ma non quanto ce n'era; per misurarlo bisogna guardare la fase esponenziale.

24. a) Pozzetto 2: una banda di 5,0 kb; pozzetto 3: 3,5 e 1,5 kb; pozzetto 4: 3,0, 1,5 e 0,5 kb. b) Il plasmide è lungo 5,0 kb (somma costante: 3,5 + 1,5 = 3,0 + 1,5 + 0,5 = 5,0). EcoRI ha un sito (un taglio, un frammento lineare), HindIII due (due frammenti). c) Nella doppia digestione il frammento di 1,5 kb resta intero, mentre quello di 3,5 kb diventa 3,0 + 0,5: il sito di EcoRI cade nel frammento HindIII di 3,5 kb, a 0,5 kb da uno dei due siti HindIII. Mappa (su un cerchio di 5,0 kb): HindIII a 0, HindIII a 1,5 kb, EcoRI a 2,0 kb (o, simmetricamente, a 4,5 kb).

25. *Traccia.* a) 21 226 + 7421 + 5804 + 5643 + 4878 + 3530 = 48 502 bp. b) Sei frammenti di un DNA lineare richiedono cinque tagli: cinque siti. Il modello prevede 48 502/4096 ≈ 11,8, cioè una dozzina di siti. c) Il genoma di λ non è una sequenza casuale: la frequenza delle basi e delle loro combinazioni non è uniforme, e un'ipotesi è che la selezione abbia sfavorito i siti di restrizione nei genomi dei fagi, che devono sopravvivere agli enzimi dei batteri. d) I frammenti di 5804 e 5643 bp differiscono di meno del 3 %: in un gel di agarosio corrono quasi alla stessa distanza e appaiono come un'unica banda più intensa.

26. a) Sul filamento non bersaglio, quello in alto, con la stessa sequenza della guida: 5′-TGG-3′. b) Il filamento bersaglio è complementare e antiparallelo ai 20 nucleotidi del filamento non bersaglio che precedono la PAM: 5′-TCTAGTCATTGGAGGTGACT-3′. c) Piatte: Cas9 taglia i due filamenti nella stessa posizione. d) Senza la sequenza NGG Cas9 non riconosce il sito e non taglia, anche se la guida è complementare.

27. *Traccia.* Il campione con sola acqua (controllo negativo) deve restare negativo: se si amplifica, i reagenti o lo strumento sono contaminati da DNA, e tutti i positivi della seduta sono sospetti (evita i falsi positivi). Il campione con RNA virale noto (controllo positivo) deve dare un C_t atteso: se non si amplifica, la reazione non ha funzionato e i negativi della seduta non valgono (evita i falsi negativi). Il gene umano (controllo interno) dimostra che nel singolo tampone c'erano cellule e che estrazione e retrotrascrizione hanno funzionato: un campione negativo per il virus e anche per il gene umano non è un vero negativo, ma un campione da ripetere.

28. *Traccia.* A circa 95 °C le interazioni deboli che mantengono la struttura terziaria di una proteina si rompono: la polimerasi di *E. coli* si denatura e perde il sito attivo, quindi ogni ciclo ne richiedeva di nuova (lavoro manuale, costi, contaminazioni). La Taq resiste a decine di cicli: si aggiunge una volta e il termociclatore automatizza la reazione; inoltre,

con una polimerasi termostabile, si possono usare temperature di appaiamento più alte, che rendono più specifico il legame dei primer. Argomento evolutivo: in un ambiente a 70 °C sopravvivono e si riproducono solo i batteri le cui proteine restano ripiegate; la selezione naturale ha favorito sequenze con più interazioni stabilizzanti (legami ionici, nuclei idrofobici compatti).

29. *Traccia.* Si sfrutta la NHEJ: si sceglie una guida all'inizio della sequenza codificante, vicina a una PAM; le piccole inserzioni o delezioni spostano la finestra di lettura e producono un codone di stop precoce (knock-out). La guida deve essere unica nel genoma (si controlla con la bioinformatica, confrontandola con la sequenza del genoma) e si può usare una Cas9 ad alta selettività. Controllo: cellule trattate con Cas9 e una guida che non riconosce alcun gene, nelle stesse condizioni. Verifica: PCR del tratto intorno al taglio e sequenziamento di Sanger dei cloni cellulari; si sequenziano anche i principali siti fuori bersaglio previsti; infine si controlla che la proteina non sia più prodotta e si confronta il comportamento delle cellule modificate con quello dei controlli.

30. *Traccia di risposta.* a) Enzimi di restrizione e ligasi per unire DNA di origini diverse, plasmidi come vettori, trasformazione di *E. coli*; il timore era che batteri intestinali con geni di virus tumorali o di resistenza agli antibiotici potessero sfuggire dal laboratorio e diffondersi. b) Contenimento fisico: laboratori con barriere crescenti al crescere del rischio; biologico: ceppi ospiti indeboliti, incapaci di sopravvivere fuori dal laboratorio. c) Jonas: quando le conseguenze sono potenzialmente irreversibili e incerte, la previsione negativa deve pesare più di quella positiva; Asilomar applica una prudenza di questo tipo, ma la affida agli scienziati stessi e la rivede quando arrivano i dati (le regole furono poi allentate). d) Collegamenti attesi: ruolo degli scienziati, delle istituzioni e dei cittadini; principio di precauzione; regole internazionali e il caso dell'editing della linea germinale (capitolo 13); distinguere le domande scientifiche (quali rischi?) da quelle etiche e politiche (quali rischi accettare?).

31. *Traccia di risposta.* Nella cellula l'elicasi apre la doppia elica, la primasi sintetizza un primer di RNA, la DNA polimerasi allunga 5′→3′, la ligasi salda i frammenti di Okazaki; nella PCR il calore (95 °C) sostituisce l'elicasi, i primer di DNA sintetico sostituiscono la primasi e definiscono il tratto copiato, la Taq termostabile sostituisce la polimerasi cellulare, e la duplicazione resta semiconservativa. Le proprietà degli enzimi (temperatura ottimale, denaturazione, cofattori come Mg^{2+}) spiegano la scelta della Taq. La crescita è esponenziale, 2^n, e poi rallenta fino al plateau, come il modello logistico. Applicazioni del capitolo 13: test molecolari, profili STR forensi, diagnosi di malattie genetiche.

32. a) $E = 10^{1/3{,}45} - 1 = 1{,}949 - 1 \approx 0{,}95$, cioè circa il 95 %. b) Con una pendenza di −3,45, un fattore 10 nelle copie iniziali corrisponde a 3,45 cicli (contro 3,32 con efficienza del 100 %): circa 3,5 cicli.