

Comprende versione

**ebook**



T. A. Brown

# GENOMI 4

IV edizione





# GENOMI 4

**T. A. BROWN**



Titolo originale:

Brown, T.A. (Terence A.)

GENOMES 4

Copyright © 2018, 4<sup>a</sup> ed., Garland Science Taylor & Francis Group, LLC

GENOMI 4

Copyright © 2018, 4<sup>a</sup> ed., EdiSES s.r.l. – Napoli

9 8 7 6 5 4 3 2 1 0

2022 2021 2020 2019 2018

Le cifre sulla destra indicano il numero e l'anno dell'ultima ristampa effettuata

*A norma di legge è vietata la  
riproduzione, anche parziale, del  
presente volume o di parte di esso  
con qualsiasi mezzo.*

*L'Editore*

*Fotocomposizione:*

doma book di Di Grazia Massimo – Napoli

*Stampato presso:*

Petruzzi S.r.l.

Via Venturelli 7/B - 06012 Città di Castello (PG)

*per conto della*

EdiSES s.r.l. – Piazza Dante, 89 – Napoli

www.edises.it      info@edises.it

ISBN 978 88 7959 8514



# CURATORI

## *Edizione italiana a cura di:*

**MAURA CARDARELLI**

*Università degli Studi di Roma  
"La Sapienza"*

**VINCENZO CAVALIERI**

*Università degli Studi di Palermo*

**SALVATORE COSTA**

*Università degli Studi di Palermo*

**GIOVANNI DI BERNARDO**

*Università degli Studi della Campania  
"Luigi Vanvitelli"*

**MARIOLINA GULLÌ**

*Università degli Studi di Parma*

**ANGELA MESSINA**

*Università degli Studi di Catania*

**MIRKO PINOTTI**

*Università degli Studi di Ferrara*

**MARINA PISCOPO**

*Università degli Studi di Napoli  
Federico II*

**CARLO PRESUTTI**

*Università degli Studi di Roma  
"La Sapienza"*

**CRISTINA ULIVIERI**

*Università degli Studi di Siena*

*Si ringrazia la Dott.ssa Martina Maione per la  
collaborazione.*

## *Revisione a cura di:*

**PAOLO MARIOTTINI**

*Università degli Studi Roma Tre*

## *Traduzione e revisione della precedente edizione a cura di:*

**Paolo Costantino**

*Università degli Studi di Roma "La Sapienza"*



---

# PREFAZIONE

Ci sono stati notevoli progressi nella conoscenza dei genomi rispetto alla precedente edizione di questo libro che è stata pubblicata dieci anni fa. Nel 2007, il sequenziamento di nuova generazione era agli esordi e si iniziavano a sfruttare i metodi ad alta resa per la trascrittomica e per la proteomica. L'applicazione di questi metodi negli ultimi dieci anni ha determinato un aumento esponenziale del numero di specie per le quali le sequenze e le annotazioni genomiche sono attualmente disponibili e ha permesso più versioni del genoma di una singola specie da esaminare. La profusione di nuove sequenze ha avuto un impatto particolarmente drammatico sulla genomica batterica, con l'introduzione del concetto di pan-genoma e la scoperta di un ampio trasferimento laterale di geni tra specie. La nostra conoscenza dei genomi eucariotici ha subito un cambiamento altrettanto drammatico, con la scoperta di nuovi tipi di RNA non codificanti, inclusi i numerosi RNA lunghi che sono trascritti da regioni presumibilmente intergeniche di molti genomi.

*Genomi 4* mantiene la struttura generale delle precedenti edizioni, con il libro diviso in quattro parti che analizzano il sequenziamento e l'annotazione del genoma, l'anatomia del genoma, l'espressione del genoma e la replicazione e l'evoluzione del genoma. L'ordine dei capitoli rimane invariato, con alcune piccole modifiche. Tuttavia, il testo è stato completamente aggiornato e, in molti capitoli, sostanzialmente rivisto. In particolare, lo sviluppo della trascrittomica e della proteomica ha raggiunto un punto in cui in *Genomi 4* è stato possibile descrivere i processi di trascrizione e di traduzione dalla prospettiva dell'intero genoma, piuttosto che semplicemente attraverso un esame dell'espressione di singoli geni. Questo era il mio obiettivo quando ho scritto la prima edizione di *Genomi* nel lontano 1999, ma le informazioni disponibili in quel momento hanno determinato il fatto che questi capitoli centrali erano trattamenti piuttosto ortodossi del gene piuttosto che dell'espressione del genoma. Siamo ancora lontani dall'essere in grado di descrivere l'intera espressione di un genoma come un singolo processo integrato, ma ci stiamo arrivando e spero di essere stato capace di trasmettere al lettore, in *Genomi 4*, almeno alcuni aspetti della natura complessiva dell'espressione del genoma.

*Genomi 4* è in preparazione da tempo e vorrei ringraziare Liz Owen della Garland Science per il suo instancabile entusiasmo per il libro e per i suoi delicati richiami sull'approssimarsi delle scadenze. Desidero anche ringraziare David Borrowdale e Georgina Lucas per la gestione della produzione del libro alla Garland e Matthew McClements per le sue splendide opere d'arte. Così come nel caso delle precedenti edizioni, *Genomi 4* non sarebbe stato finito senza il supporto di mia moglie Keri. Il riconoscimento presente nella prima edizione che recita "se trovi utile questo libro, dovresti ringraziare Keri, non me, perché è lei che ha assicurato che fosse scritto" è altrettanto vero per la quarta edizione.

## UNA NOTA PER IL LETTORE

Ho cercato di rendere la quarta edizione di *Genomi* quanto più agevole possibile per i lettori. Il libro contiene quindi una serie di strumenti ideati per facilitare l'apprendimento e per aiutare il lettore ad assimilare i contenuti.

### Organizzazione del Libro

*Genomi 4* è diviso in quattro parti:

**Parte I – Studio dei genomi** inizia con un capitolo orientativo che introduce il lettore ai genomi, ai trascrittomi e ai proteomi per passare poi ai metodi, centrati sul clonaggio e sulla PCR, che sono stati utilizzati nell'era pre-genomica per esaminare singoli geni (Capitolo 2). Le tecniche che riguardano in maniera più specifica lo studio dei genomi sono descritte nell'ordine in cui verrebbero utilizzate in un progetto genomico: metodi per costruire mappe genetiche e fisiche (Capitolo 3); metodologie di sequenziamento del DNA e strategie usate per assemblare una sequenza genomica (Capitolo 4). Due capitoli sono dedicati all'analisi delle sequenze genomiche: il Capitolo 5 analizza l'annotazione del genoma mediante l'identificazione di geni e di altre caratteristiche e il Capitolo 6 esamina l'analisi funzionale dei geni che sono stati scoperti.

**Parte II – Anatomia dei genomi** passa in rassegna l'anatomia dei diversi genomi che si ritrovano sul nostro pianeta. Il Capitolo 7 tratta dei genomi nucleari eucariotici con un'enfasi particolare rivolta al genoma umano, in parte a causa dell'importanza del genoma umano in molte aree di ricerca e anche perché il nostro genoma è il migliore rispetto a tutti i genomi per i quali sono disponibili le sequenze. Il Capitolo 8 si occupa dei genomi procariotici e degli organelli eucariotici, inclusi nella stessa sede per le loro origini procariotiche. Il Capitolo 9 descrive i genomi dei virus insieme agli elementi genetici mobili dal momento che alcuni elementi mobili sono correlati ai genomi virali.

**Parte III – Espressione dei genomi** descrive in che modo l'informazione biologica contenuta in un genoma è utilizzata dalla cellula in cui il genoma si trova. Il Capitolo 10 affronta il problema sempre più importante di come la struttura della cromatina influenzi l'espressione delle diverse parti del genoma. Il Capitolo 11 contiene una descrizione dettagliata delle proteine di legame al DNA, che ricoprono un ruolo essenziale nell'espressione di quelle parti del genoma che sono attive in un particolare momento. Il Capitolo 12 esamina il trascrittoma, descrivendo in che modo sono studiati i trascrittomi, la loro composizione e il modo in cui il trascrittoma di una cellula è sintetizzato e mantenuto. Il Capitolo 13 fornisce una descrizione equivalente della proteomica e del proteoma. Il Capitolo 14 conclude questa parte del libro esplorando in che modo il genoma agisce nel contesto di una cellula o di un organismo, rispondendo ai segnali extracellulari e guidando i cambiamenti biochimici che sono alla base del differenziamento e dello sviluppo.

**Parte IV – Replicazione ed evoluzione dei genomi** associa gli eventi di replicazione, mutazione e ricombinazione del DNA alla graduale evoluzione dei genomi nel tempo. Nei Capitoli 15-17 sono descritti i processi molecolari alla base della replicazione, della mutazione e della ricombinazione e nel Capitolo 18 viene spiegato come probabilmente questi processi hanno determinato la struttura e il contenuto dei genomi nel corso dell'evoluzione. Il Capitolo 18 si conclude con alcuni casi studio che illustrano in che modo si stanno utilizzando la filogenetica molecolare e la genetica di popolazione nella ricerca e in biotecnologia.



## OBIETTIVI DI APPRENDIMENTO

Ogni capitolo presenta una serie di Domande a risposta breve e di Problemi di approfondimento così come una lista di Ulteriori letture. Alla fine del libro è presente un esteso Glossario.

Le **Domande a risposta breve** richiedono risposte lunghe da 50 a 500 parole. Le domande, che si riferiscono all'intero contenuto del capitolo, sono rivolte in maniera diretta e la maggior parte delle risposte possono essere verificate semplicemente controllando la parte corrispondente del testo. Lo studente può utilizzare queste domande sistematicamente per rielaborare ogni capitolo o selezionarne alcune per valutare la propria capacità di rispondere a domande su argomenti specifici. Queste domande possono anche essere affrontate senza l'ausilio del libro.

I **Problemi di approfondimento** richiedono una risposta più approfondita, sebbene varino come genere e difficoltà. I più semplici richiedono poco più di un approfondimento della bibliografia nota, con l'intenzione di far apprendere allo studente maggiori dettagli rispetto a quelli forniti da *Genomi 4*. Altri problemi richiedono agli studenti di valutare una fase o un'ipotesi, basandosi su ciò che hanno appreso dal testo e possibilmente ampliando le conoscenze con delle letture sull'argomento. L'intento è quello di stimolare una riflessione critica e generare una maggiore consapevolezza. Alcuni problemi sono realmente difficili, al punto che non esiste una risposta rigorosa alle domande poste. Questi problemi sono stati ideati per stimolare discussioni e speculazioni che rafforzino la conoscenza di ciascuno studente e lo spingano a riflettere attentamente sulle proprie affermazioni. Questi problemi possono essere affrontati singolarmente dagli studenti o in alternativa possono rappresentare il punto di partenza per una discussione di gruppo.

Le **Ulteriori letture** elencano alla fine di ogni capitolo articoli scientifici, rassegne e libri che considero come più utili fonti di materiale addizionale. Il mio proposito è stato quello di rendere la bibliografia di *Genomi 4* il più utile possibile per gli studenti che devono scrivere saggi e dissertazioni su particolari argomenti. È per questo che ho incluso articoli scientifici solo nel caso in cui il contenuto sia comprensibile al lettore medio del libro. L'enfasi è posta su rassegne accessibili, perché questi articoli generali sono spesso più utili per spiegare il contesto e l'importanza di un lavoro. La maggior parte delle letture è suddivisa in sezioni che riflettono l'organizzazione degli argomenti nel capitolo e in alcuni casi ho aggiunto un breve riassunto di una riga che ne descrive il valore, per aiutare il lettore nella scelta. In alcuni casi, le Ulteriori letture includono URL per database e altre risorse online rilevanti per gli argomenti trattati nel capitolo.

Il **Glossario** definisce termini evidenziati in grassetto nel testo, insieme a un numero di termini aggiuntivi che il lettore potrebbe incontrare nella consultazione di libri o articoli presenti nelle liste delle letture. Il Glossario pertanto fornisce un mezzo rapido e conveniente mediante il quale il lettore può ricordare dei termini tecnici rilevanti per lo studio dei genomi e rappresenta un aiuto alla ripetizione per assicurarsi che quelle definizioni siano chiaramente comprese durante i minuti di incertezza che molti studenti sperimentano immediatamente prima un esame.

## MATERIALE DI SUPPORTO PER I DOCENTI

I docenti che utilizzano il testo a scopo didattico possono scaricare dal sito [www.edises.it](http://www.edises.it), previa registrazione all'area docenti, le immagini del libro in formato PowerPoint.

## RINGRAZIAMENTI

L'Autore e gli Editori di *Genomi 4* sono profondamente grati per il contributo dei seguenti revisori nello sviluppo di questa edizione.

David Baillie, Simon Fraser University; Linda Bonen, University of Ottawa; Hugh Cam, Boston College; Yuri Dubrova, University of Leicester; Bart Eggen, University of Groningen; Robert Fowler, San José State University; Sidney Fu, George Washington University; Adrian Hall, Sheffield Hallam University; Lee Hwei Huih, Universiti Tunku Abdul Rahman; Glyn Jenkins, Aberystwyth University; Julian M. Ketley, University of Leicester; Torsten Kristensen, University of Aarhus; Gerhard May, University of Dundee; Mike McPherson, University of Leeds; Isidoro Metón, Universitat de Barcelona; Gary Ogden, St. Mary's University; Paul Overvoorde, Macalester College; John Rafferty, University of Sheffield; Andrew Read, University of Manchester; Joaquin Cañizares Sales, Universitat Politècnica de València; Michael Schweizer, Heriot-Watt University; Eric Spana, Duke University; David Studholme, Exeter University; John Taylor, University of Newcastle; Gavin Thomas, University of York; Matthew Upton, Plymouth University; Guido van den Ackerveken, Utrecht University; Vassie Ware, Lehigh University; Wei Zhang, Illinois Institute of Technology.

## NOTE SULL'AUTORE

Sono stato molto attratto dal mondo naturale fin da giovane. Ho iniziato la mia carriera di ricercatore studiando gli effetti dell'inquinamento da metalli sui microrganismi e la tolleranza che alcune piante mostrano verso alte concentrazioni di metalli tossici. In seguito, mi ha entusiasmato il DNA e ho lavorato su geni mitocondriali nei funghi per imparare le nuove tecniche (a quei tempi) per il clonaggio genico e per il sequenziamento del DNA. Ho contribuito alla scoperta di introni mitocondriali e allo studio che ha descritto la struttura *base-paired* di questi introni. Successivamente, ho rivolto il mio interesse al DNA antico e sono stato uno dei primi ad effettuare estrazioni di DNA dalle ossa e da resti preservati di piante. Questo studio ha richiesto una stretta collaborazione con gli archeologi e ha determinato il mio attuale impegno verso la paleogenomica, le origini dell'agricoltura e l'evoluzione delle piante domestiche.

Ho conseguito il dottorato di ricerca presso l'UCL (*University College London*) nel 1977 e poi ho lavorato a New York, Oxford, Colchester e Manchester prima di iniziare nel 1984 come docente di Biotecnologie presso l'UMIST (*University of Manchester Institute of Science and Technology*). Sono stato nominato Professore di Archeologia biomolecolare nel 2000 e sono stato Responsabile delle Scienze biomolecolari presso l'UMIST dal 2002 al 2004. Sono stato quindi Preside Associato nella Facoltà di Scienze della vita dell'Università di Manchester fino al 2006, prima di prendere una pausa dagli impegni amministrativi per avere più tempo per fare ricerca.

I miei altri libri di testo universitari comprendono *Genetics*, *A Molecular Approach* (Garland Science).

---

# SOMMARIO

<b>CAPITOLO 1</b>	GENOMI, TRASCRITTOMI E PROTEOMI	1
<b>CAPITOLO 2</b>	STUDIARE IL DNA	27
<b>CAPITOLO 3</b>	MAPPATURA DEI GENOMI	55
<b>CAPITOLO 4</b>	SEQUENZIAMENTO DEI GENOMI	87
<b>CAPITOLO 5</b>	ANNOTAZIONE DEL GENOMA	119
<b>CAPITOLO 6</b>	IDENTIFICARE LE FUNZIONI GENICHE	135
<b>CAPITOLO 7</b>	GENOMI NUCLEARI EUCARIOTICI	155
<b>CAPITOLO 8</b>	GENOMI DEI PROCARIOTI E DEGLI ORGANELLI EUCARIOTICI	181
<b>CAPITOLO 9</b>	GENOMI VIRALI ED ELEMENTI GENETICI MOBILI	203
<b>CAPITOLO 10</b>	COME SI ACCEDE AL GENOMA	219
<b>CAPITOLO 11</b>	RUOLO DELLE PROTEINE CHE LEGANO IL DNA NELL'ESPRESSIONE DEL GENOMA	241
<b>CAPITOLO 12</b>	TRASCRITTOMI	257
<b>CAPITOLO 13</b>	PROTEOMI	293
<b>CAPITOLO 14</b>	ESPRESSIONE DEL GENOMA NEL CONTESTO CELLULARE E DELL'ORGANISMO	329
<b>CAPITOLO 15</b>	REPLICAZIONE DEL GENOMA	357
<b>CAPITOLO 16</b>	MUTAZIONI E RIPARAZIONE DEL DNA	389
<b>CAPITOLO 17</b>	RICOMBINAZIONE E TRASPOSIZIONE	411
<b>CAPITOLO 18</b>	EVOLUZIONE DEI GENOMI	429
<b>GLOSSARIO</b>		G-1
<b>INDICE ANALITICO</b>		I-1

# INDICE GENERALE

## CAPITOLO 1 GENOMI, TRASCRITTOMI E PROTEOMI

### 1.1 DNA

I geni sono fatti di DNA

Il DNA è un polimero di nucleotidi

La doppia elica è stabilizzata dall'appaiamento e dall'impilamento delle basi

La doppia elica è dotata di flessibilità strutturale

### 1.2 RNA E TRASCRITTOMA

Struttura dell'RNA

RNA contenuti nella cellula

Maturazione dell'RNA precursore

Esistono diverse definizioni di trascrittoma

### 1.3 PROTEINE E PROTEOMA

I quattro livelli gerarchici della struttura proteica

La diversità amminoacidica è alla base della diversità proteica

Legame tra trascrittoma e proteoma

Il codice genetico non è universale

Collegamento tra proteoma e biochimica della cellula

### SOMMARIO

### DOMANDE A RISPOSTA BREVE

### PROBLEMI DI APPROFONDIMENTO

### ULTERIORI LETTURE

## CAPITOLO 2 STUDIARE IL DNA

### 2.1 ENZIMI PER LA MANIPOLAZIONE DEL DNA

Meccanismo di azione della DNA polimerasi stampo-dipendente

Tipi di DNA polimerasi utilizzati nella ricerca

Le endonucleasi di restrizione consentono di tagliare le molecole di DNA in posizioni definite

L'elettroforesi su gel viene utilizzata per esaminare i prodotti derivati da digestione con un enzima di restrizione

1	Frammenti di DNA di interesse possono essere identificati mediante ibridazione Southern	35
	Le ligasi uniscono frammenti di DNA	37
	Enzimi di modificazione terminale	38

2	<b>2.2 REAZIONE A CATENA DELLA</b>	
3	<b>POLIMERASI</b>	38
4	Esecuzione di una PCR	39
8	La quantità di prodotto formato può essere seguito durante una PCR	40
9	La PCR ha notevoli e diverse applicazioni	41

11	<b>2.3 CLONAGGIO DEL DNA</b>	41
12	Perché il clonaggio genico è importante?	41
12	I vettori di clonaggio più semplici sono basati su plasmidi di <i>E. coli</i>	43
13	Anche i batteriofagi possono essere usati come vettori di clonaggio	44
15	Vettori per lunghi frammenti di DNA	47
16	Il DNA può essere clonato in organismi diversi da <i>E. coli</i>	48

17	<b>SOMMARIO</b>	50
----	-----------------	----

19	<b>DOMANDE A RISPOSTA BREVE</b>	51
----	---------------------------------	----

20	<b>PROBLEMI DI APPROFONDIMENTO</b>	51
----	------------------------------------	----

22	<b>ULTERIORI LETTURE</b>	52
----	--------------------------	----

23

24

## CAPITOLO 3 MAPPATURA DEI GENOMI

55

24	<b>3.1 PERCHÉ È IMPORTANTE UNA MAPPA GENOMICA</b>	55
----	---	----

25	Le mappe genomiche sono necessarie per sequenziare i genomi più complessi	55
----	---	----

27	Le mappe genomiche non sono solo di ausilio per il sequenziamento	57
----	---	----

28	<b>3.2 MARCATORI PER LA MAPPATURA GENETICA</b>	58
----	--	----

28	I geni sono stati i primi marcatori da utilizzare	58
30	RFLP e SSLP sono esempi di marcatori di DNA	59

32	I polimorfismi di un singolo nucleotide sono il tipo più utile di marcatore di DNA	61
----	--	----

34	<b>3.3 BASI DELLA MAPPATURA GENETICA</b>	63
----	--	----

	Principi dell'ereditarietà e scoperta dell'associazione	63
--	---	----



Il comportamento dei cromosomi durante la meiosi spiega l'associazione genica parziale  
Dall'associazione parziale alla mappatura genetica

### 3.4 ANALISI DI ASSOCIAZIONE IN DIVERSI TIPI DI ORGANISMI

Analisi di associazione quando sono possibili incroci programmati  
Mappatura genetica tramite l'analisi degli alberi genealogici  
Mappatura genetica nei batteri  
Le limitazioni dell'analisi di associazione

### 3.5 MAPPATURA FISICA MEDIANTE ESAME DIRETTO DELLE MOLECOLE DI DNA

La mappatura di restrizione convenzionale è applicabile solo a piccole molecole di DNA  
La mappatura ottica può localizzare i siti di restrizione in molecole di DNA più lunghe  
La mappatura ottica può essere utilizzata per mappare altre caratteristiche in una molecola di DNA

### 3.6 MAPPATURA FISICA MEDIANTE ASSEGNAZIONE DI MARCATORI A FRAMMENTI DI DNA

Qualsiasi sequenza unica di DNA può essere utilizzata come STS  
I frammenti di DNA per la mappatura mediante STS possono essere ottenuti come ibridi da radiazione  
Una libreria di cloni può essere utilizzata come reagente di mappatura

### SOMMARIO

### DOMANDE A RISPOSTA BREVE

### PROBLEMI DI APPROFONDIMENTO

### ULTERIORI LETTURE

## CAPITOLO 4 SEQUENZIAMENTO DEI GENOMI 87

### 4.1 SEQUENZIAMENTO CON IL METODO A TERMINAZIONE DI CATENA

Punti salienti del sequenziamento con il metodo a terminazione di catena  
Non tutte le DNA polimerasi possono essere utilizzate per il sequenziamento  
Sequenziamento con il metodo a terminazione di catena mediante *Taq* polimerasi  
Punti di forza e limiti del sequenziamento a terminazione di catena

### 4.2 SEQUENZIAMENTO DI NUOVA GENERAZIONE

La preparazione di una libreria di sequenziamento è la caratteristica comune dei metodi di sequenziamento di nuova generazione  
Sono stati ideati vari metodi di sequenziamento di nuova generazione  
I metodi di terza generazione e di quarta generazione consentono il sequenziamento in tempo reale

### 4.3 COME SEQUENZIARE UN GENOMA

Il potenziale del metodo shotgun è stato dimostrato dal sequenziamento del genoma di *Haemophilus influenzae*  
Molti genomi procariotici sono stati sequenziati con il metodo shotgun  
Il sequenziamento dei genomi eucariotici con il metodo shotgun richiede sofisticati programmi di assemblaggio  
I genomi più complessi possono essere sequenziati mediante un approccio gerarchico shotgun  
Cos'è una sequenza genomica e perché ne abbiamo sempre bisogno?

### 4.4 PROGETTI DI SEQUENZIAMENTO DEL GENOMA EUCARIOTICO

Progetto Genoma Umano: sequenziamento del genoma nell'età eroica  
Genoma di Neanderthal: assemblaggio di un genoma estinto mediante l'uso della sequenza umana come riferimento  
Genoma del panda gigante: sequenziamento shotgun basato interamente sui dati del sequenziamento di nuova generazione  
Genoma dell'orzo: concetto di spazio genico

### SOMMARIO

### DOMANDE A RISPOSTA BREVE

### PROBLEMI DI APPROFONDIMENTO

### ULTERIORI LETTURE

## CAPITOLO 5 ANNOTAZIONE DEL GENOMA 119

### 5.1 ANNOTAZIONE DEL GENOMA MEDIANTE ANALISI AL COMPUTER DELLA SEQUENZA DEL DNA

Le regioni codificanti dei geni sono schemi di lettura aperti  
L'analisi delle ORF è meno efficace con il DNA di eucarioti superiori

Localizzazione dei geni per RNA non codificanti 122  
 Le ricerche di omologie di sequenza e l'analisi comparata forniscono un'ulteriore dimensione alle analisi di sequenza 123

## 5.2 ANNOTAZIONE DEL GENOMA MEDIANTE ANALISI DEI TRASCRITTI DEI GENI

Saggi di ibridazione possono permettere di determinare se un frammento contiene sequenze espresse 124  
 Esistono metodi per mappare con precisione le estremità dei trascritti 126  
 Le giunzioni esone-introne possono anche essere localizzate con precisione 126

## 5.3 ANNOTAZIONE MEDIANTE MAPPATURA DEGLI RNA DELL'INTERO GENOMA

I tiling array consentono di mappare i trascritti sui cromosomi o su interi genomi 127  
 Le sequenze dei trascritti possono essere mappate direttamente sul genoma 129

## 5.4 BROWSER GENOMICI

### SOMMARIO

### DOMANDE A RISPOSTA BREVE

### PROBLEMI DI APPROFONDIMENTO

### ULTERIORI LETTURE

## CAPITOLO 6 IDENTIFICARE LE FUNZIONI GENICHE

### 6.1 ANALISI AL COMPUTER DELLA FUNZIONE DI UN GENE

L'omologia riflette il grado di parentela evolutiva 135  
 L'analisi di omologia può fornire informazioni sulla funzione di un gene 136  
 L'identificazione di domini proteici può aiutare ad assegnare una funzione a un gene sconosciuto 137  
 L'annotazione della funzione genica richiede una terminologia comune 138

### 6.2 ASSEGNARE UNA FUNZIONE MEDIANTE INATTIVAZIONE O SOVRAESPRESSIONE GENICA

Analisi funzionale per mezzo dell'inattivazione genica 139  
 Singoli geni possono essere inattivati per ricombinazione omologa 140  
 Inattivazione genica senza ricombinazione omologa 142

La sovraespressione di un gene può essere utilizzata per assegnarne la funzione 144  
 Qualche volta è difficile discernere l'effetto fenotipico dell'inattivazione o della sovraespressione di un gene 145

### 6.3 COMPRENDERE LA FUNZIONE DI UN GENE ATTRAVERSO STUDI SUL PATTERN DI ESPRESSIONE E SUL PRODOTTO PROTEICO

I geni reporter e l'immunocitochimica possono essere utilizzati per localizzare dove e quando i geni vengono espressi 146  
 La mutagenesi diretta può essere usata per saggiare la funzione genica in dettaglio 147

### 6.4 UTILIZZARE LE ANALISI DI GENETICA CONVENZIONALE PER IDENTIFICARE LA FUNZIONE GENICA

Identificazione di geni umani responsabili di malattie ereditarie 149  
 Gli studi di associazione genomica permettono di identificare i geni per le malattie e per altri tratti 151

### SOMMARIO

### DOMANDE A RISPOSTA BREVE

### PROBLEMI DI APPROFONDIMENTO

### ULTERIORI LETTURE

## CAPITOLO 7 GENOMI NUCLEARI EUCARIOTICI 155

### 7.1 GENOMI NUCLEARI SONO CONTENUTI NEI CROMOSOMI

I cromosomi sono molto più corti delle molecole di DNA che contengono 155  
 Caratteristiche speciali dei cromosomi metafasici 157  
 Interazioni DNA-proteine nei centromeri e nei telomeri 159

### 7.2 COME SONO ORGANIZZATI I GENI IN UN GENOMA NUCLEARE?

I geni non sono distribuiti in modo omogeneo nel genoma 161  
 Un segmento del genoma umano 162  
 Il genoma del lievito è molto compatto 164  
 Organizzazione genica in altri eucarioti 165

### 7.3 QUANTI GENI SONO CONTENUTI IN UN GENOMA NUCLEARE E QUALI SONO LE LORO FUNZIONI?

Le stime del numero di geni possono essere fuorvianti 167  
 I cataloghi genici rivelano le caratteristiche distintive dei diversi organismi 169

Famiglie geniche	172
Pseudogeni e altri relitti evolutivi	174

## 7.4 DNA RIPETITIVO DEI GENOMI NUCLEARI EUCARIOTICI

Nei cromosomi eucariotici il DNA ripetuto in tandem non si trova solo nei centromeri	176
Minisatelliti e microsatelliti	176
Ripetizioni intersperse	177

## SOMMARIO

## DOMANDE A RISPOSTA BREVE

## PROBLEMI DI APPROFONDIMENTO

## ULTERIORI LETTURE

# CAPITOLO 8 GENOMI DEI PROCARIOTI E DEGLI ORGANELLI EUCARIOTICI 181

## 8.1 CARATTERISTICHE FISICHE DEI GENOMI PROCARIOTICI

Visione tradizionale del cromosoma procariotico	181
Alcuni batteri hanno genomi lineari o multipartiti	183

## 8.2 CARATTERISTICHE GENETICHE DEI GENOMI PROCARIOTICI

Organizzazione dei geni nel genoma di <i>E. coli</i> K12	186
Gli operoni sono una caratteristica dei genomi procariotici	188
Le dimensioni del genoma e il numero di geni nei procariotici sono correlati alla complessità biologica	189
Le dimensioni del genoma e il numero di geni variano all'interno delle singole specie	190
Le differenze tra le diverse specie dei procariotici sono attenuate dal trasferimento laterale di geni	192
I metagenomi rappresentano i membri di una comunità	194

## 8.3 GENOMI DEGLI ORGANELLI EUCARIOTICI

La teoria dell'endosimbionte spiega l'origine dei genomi degli organelli	195
I genomi degli organelli sono per la maggior parte circolari	196
Contenuto genetico dei genomi degli organelli	197

## SOMMARIO

## DOMANDE A RISPOSTA BREVE

## PROBLEMI DI APPROFONDIMENTO

## ULTERIORI LETTURE

# CAPITOLO 9 GENOMI VIRALI ED ELEMENTI GENETICI MOBILI 203

## 9.1 GENOMI DEI BATTERIOFAGI E DEI VIRUS EUCARIOTICI

I genomi dei batteriofagi hanno strutture e organizzazioni diverse	203
Strategie di replicazione dei genomi fagici	205
Strutture e strategie replicative dei genomi virali eucariotici	206
Alcuni retrovirus causano il cancro	207
Genomi al limite della vita	209

## 9.2 ELEMENTI GENETICI MOBILI

I trasposoni a RNA con lunghe ripetizioni terminali sono correlati con i retroelementi virali	210
Alcuni trasposoni a RNA sono privi di lunghe ripetizioni terminali	212
I trasposoni a DNA sono comuni nei genomi procariotici	213
I trasposoni a DNA sono meno comuni nei genomi eucariotici	214

## SOMMARIO

## DOMANDE A RISPOSTA BREVE

## PROBLEMI DI APPROFONDIMENTO

## ULTERIORI LETTURE

# CAPITOLO 10 COME SI ACCEDE AL GENOMA 219

## 10.1 ALL'INTERNO DEL NUCLEO

Il nucleo ha una struttura interna altamente ordinata	220
Il DNA all'interno di un nucleo quiescente mostra gradi di impacchettamento differenti	221
La matrice nucleare fornisce una piattaforma di attracco per il DNA cromosomale	222
Ciascun cromosoma occupa un proprio territorio all'interno del nucleo	223
Ciascun cromosoma comprende una serie di domini topologicamente associati	224
Gli isolatori demarcano i confini dei domini topologicamente associati	226

## 10.2 MODIFICAZIONI DEL NUCLEOSOMA ED ESPRESSIONE DEL GENOMA

L'acetilazione degli istoni influenza numerose attività nucleari, compresa l'espressione del genoma	228
La deacetilazione degli istoni reprime le regioni attive del genoma	229

L'acetilazione non è l'unico tipo di modificazione istonica	230
Il riposizionamento dei nucleosomi influenza anch'esso l'espressione genica	231

<b>10.3 MODIFICAZIONE DEL DNA ED ESPRESSIONE DEL GENOMA</b>	234
Silenziamento del genoma mediato dalla metilazione del DNA	234
La metilazione è coinvolta nell'imprinting genomico e nell'inattivazione del cromosoma X	235

<b>SOMMARIO</b>	236
<b>DOMANDE A RISPOSTA BREVE</b>	237
<b>PROBLEMI DI APPROFONDIMENTO</b>	238
<b>ULTERIORI LETTURE</b>	238

## **CAPITOLO 11 RUOLO DELLE PROTEINE CHE LEGANO IL DNA NELL'ESPRESSIONE DEL GENOMA**

241

<b>11.1 METODI PER STUDIARE LE PROTEINE CHE SI LEGANO AL DNA E LORO SITI DI INTERAZIONE</b>	241
---	-----

La cristallografia a raggi X fornisce informazioni sulla struttura di qualsiasi proteina che possa essere cristallizzata

La spettroscopia a risonanza magnetica nucleare è sfruttata per studiare la struttura di proteine di piccole dimensioni

Esperimenti di ritardo su gel identificano frammenti di DNA a cui si legano le proteine

I saggi di protezione identificano con maggiore precisione i siti di legame

L'interferenza per modificazione identifica i nucleotidi essenziali per il legame di una proteina

Scansione dell'intero genoma per l'identificazione di siti di legame per proteine

<b>11.2 CARATTERISTICHE DISTINTIVE DELLE PROTEINE CHE LEGANO IL DNA</b>	249
---	-----

Il motivo elica-giro-elica è presente sia nelle proteine eucariotiche che in quelle procariotiche

I motivi a dito di zinco sono comuni nelle proteine eucariotiche

Altri motivi di legame al DNA

<b>11.3 INTERAZIONE TRA DNA E PROTEINE</b>	252
Lettura diretta della sequenza nucleotidica	252

La sequenza nucleotidica ha alcuni effetti indiretti sulla struttura dell'elica

Interazioni tra DNA e proteine

<b>SOMMARIO</b>	254
<b>DOMANDE A RISPOSTA BREVE</b>	255
<b>PROBLEMI DI APPROFONDIMENTO</b>	256
<b>ULTERIORI LETTURE</b>	256

## **CAPITOLO 12 TRASCRITTOMI**

257

<b>12.1 COMPONENTI DEL TRASCRITTOMA</b>	257
La frazione di mRNA del trascrittoma è piccola ma complessa	257
I piccoli RNA non codificanti hanno diverse funzioni	259
I lunghi RNA non codificanti sono trascritti enigmatici	260
Le analisi con i microarray e il sequenziamento degli RNA sono usati per studiare il contenuto dei trascrittomi	262

<b>12.2 SINTESI DEI COMPONENTI DEL TRASCRITTOMA</b>	263
---	-----

Le RNA polimerasi sono macchine che producono l'RNA

I punti di inizio della trascrizione sono indicati dalle sequenze del promotore

La sintesi dell'RNA batterico è regolata da proteine repressori e attivatori

La sintesi di RNA batterico è regolata anche da un controllo a livello di terminazione della trascrizione

La sintesi degli RNA eucariotici è regolata principalmente da proteine attivatrici

<b>12.3 DEGRADAZIONE DEI COMPONENTI DEL TRASCRITTOMA</b>	275
--	-----

Numerosi processi controllano il turnover di RNA non specifici

RII silenziamento a RNA è stato inizialmente considerato come una opportunità per distruggere gli RNA virali invasori

I microRNA regolano l'espressione del genoma provocando la degradazione di mRNA specifici

<b>12.4 INFLUENZA DELLA MATURAZIONE DELL'RNA SULLA COMPOSIZIONE DEL TRASCRITTOMA</b>	278
--	-----

Il processo di splicing degli introni dei pre-mRNA eucariotici

Il processo di splicing deve essere altamente preciso

Gli elementi enhancer e silencer definiscono i processi di splicing alternativo

<b>12.5 TRASCRITTOMI NELLA RICERCA</b>	284
--	-----

L'analisi del trascrittoma aiuta l'annotazione del genoma



Trascrittomi del cancro	286
Trascrittomi e risposta delle piante allo stress	287

## SOMMARIO

### DOMANDE A RISPOSTA BREVE

### PROBLEMI DI APPROFONDIMENTO

### ULTERIORI LETTURE

## CAPITOLO 13 PROTEOMI

### 13.1 STUDIARE LA COMPOSIZIONE DI UN PROTEOMA

La fase di separazione di un progetto di analisi del profilo proteico

La fase di identificazione di un progetto di analisi del profilo proteico

Comparare la composizione di due proteomi

Gli array analitici di proteine rappresentano un approccio alternativo al profilo di proteine

### 13.2 IDENTIFICARE PROTEINE CHE INTERAGISCONO TRA LORO

Identificare coppie di proteine interagenti

Identificazione dei componenti di un complesso multiproteico

Identificazione di proteine che hanno interazioni funzionali

Le mappe di interazioni proteiche mostrano le interazioni dentro un proteoma

### 13.3 SINTESI E DEGRADAZIONE DEI COMPONENTI DEL PROTEOMA

I ribosomi sono le macchine molecolari che fanno le proteine

Durante lo stress, i batteri non utilizzano i ribosomi allo scopo di ridurre il proteoma

I fattori di inizio mediano il rimodellamento su larga scala del proteoma eucariotico

La traduzione di singoli mRNA può essere regolata

Degradazione dei componenti del proteoma

### 13.4 INFLUENZA DEL PROCESSAMENTO DELLE PROTEINE SULLA COMPOSIZIONE DEL PROTEOMA

La sequenza amminoacidica contiene informazioni per il ripiegamento della proteina

Alcune proteine sono attivate da un taglio proteolitico

Variazioni importanti nell'attività di una proteina possono essere causate da modificazioni chimiche

286	<b>13.5 OLTRE IL PROTEOMA</b>	322
287	Il metaboloma rappresenta l'insieme di tutti i metaboliti di una cellula	322
289	La biologia dei sistemi fornisce una descrizione dettagliata dell'attività cellulare	323
289	<b>SOMMARIO</b>	326
290	<b>DOMANDE A RISPOSTA BREVE</b>	326
290	<b>PROBLEMI DI APPROFONDIMENTO</b>	327
	<b>ULTERIORI LETTURE</b>	327

## 293

## CAPITOLO 14 ESPRESSIONE DEL GENOMA NEL CONTESTO CELLULARE E DELL'ORGANISMO

329

### 14.1 RISPOSTA DEL GENOMA AI SEGNALI ESTERNI

Trasmissione del segnale mediante importazione di un composto-segnale extracellulare

Recettori di natura proteica trasmettono il segnale attraverso le membrane cellulari

Alcune vie di trasduzione del segnale prevedono pochi passaggi tra il recettore e il genoma

Alcune vie di trasduzione del segnale implicano parecchi passaggi tra il recettore e il genoma

Alcune vie di trasduzione del segnale si realizzano mediante secondi messaggeri

### 14.2 CAMBIAMENTI DELL'ATTIVITÀ DEL GENOMA CHE DETERMINANO DIFFERENZIAMENTO CELLULARE

Alcuni processi di differenziamento comportano modifiche alla struttura della cromatina

I gruppi di compatibilità del lievito sono determinati da eventi di conversione genica

I riarrangiamenti del genoma sono responsabili della diversità delle immunoglobuline e dei recettori delle cellule T

### 14.3 CAMBIAMENTI DELL'ATTIVITÀ DEL GENOMA DURANTE LO SVILUPPO

Batteriofago  $\lambda$ : un interruttore genetico consente di scegliere tra percorsi di sviluppo alternativi

Sporulazione in *Bacillus*: coordinamento delle attività in due tipi cellulari distinti

*Caenorhabditis elegans*: base genetica dell'informazione posizionale e determinazione del destino cellulare

Moscerini della frutta: conversione dell'informazione posizionale in uno schema di segmentazione corporea

293

294

297

299

300

301

301

304

305

306

308

308

311

312

313

314

315

315

318

320

330

330

332

333

334

336

336

336

338

339

341

342

343

346

348

I geni selettori omeotici sono una costante dello sviluppo degli eucarioti superiori  
I geni omeotici sono anche alla base dello sviluppo delle piante

**SOMMARIO****DOMANDE A RISPOSTA BREVE****PROBLEMI DI APPROFONDIMENTO****ULTERIORI LETTURE****CAPITOLO 15****REPLICAZIONE DEL GENOMA 357****15.1 PROBLEMA TOPOLOGICO 357**

La struttura a doppia elica complica il processo di replicazione

L'esperimento di Meselson e Stahl dimostra che la replicazione è semiconservativa

Le DNA topoisomerasi forniscono una soluzione al problema topologico

Variazioni sul tema della replicazione semiconservativa

**15.2 INIZIO DELLA REPLICAZIONE DEL GENOMA 364**

Inizio all'origine di replicazione di *E. coli*

Le origini di replicazione sono state identificate anche in lievito

Le origini di replicazione degli eucarioti superiori sono state identificate con maggiori difficoltà

**15.3 EVENTI CHE SI REALIZZANO A LIVELLO DELLA FORCINA DI REPLICAZIONE 367**

Le DNA polimerasi sono enzimi capaci di sintetizzare (e degradare) il DNA

La replicazione è complicata da limitazioni delle DNA polimerasi

I frammenti di Okazaki devono essere uniti assieme per completare la replicazione del filamento lagging

**15.4 TERMINAZIONE DELLA REPLICAZIONE 372**

La replicazione del genoma di *E. coli* termina in corrispondenza di regioni definite

La terminazione della replicazione negli eucarioti è nota solo in parte

In alcune cellule, le telomerasi completano la replicazione del DNA cromosomale

La lunghezza dei telomeri è implicata nella senescenza e nel cancro

Unicità delle estremità telomeriche in *Drosophila*

**15.5 REGOLAZIONE DELLA REPLICAZIONE DEL GENOMA EUCARIOTICO 380**

Coordinazione tra la replicazione del genoma e la divisione cellulare

La formazione del complesso di pre-replicazione rappresenta il prerequisito per il superamento del punto di controllo G1-S

Le origini di replicazione non si attivano tutte nello stesso momento

Le cellule mostrano varie opzioni qualora il DNA venga danneggiato

**SOMMARIO 384****DOMANDE A RISPOSTA BREVE 385****PROBLEMI DI APPROFONDIMENTO 385****ULTERIORI LETTURE 386****CAPITOLO 16****MUTAZIONI E RIPARAZIONE DEL DNA 389****16.1 CAUSE DELLE MUTAZIONI 389**

Gli errori di replicazione danno origine a mutazioni puntiformi

Gli errori di replicazione possono determinare anche inserzione e delezione di nucleotidi

Le mutazioni sono causate anche da mutageni chimici e fisici

**16.2 MECCANISMI DI RIPARAZIONE DELLE MUTAZIONI E ALTRI TIPI DI DANNO AL DNA 398**

I sistemi di riparazione diretta riempiono le interruzioni e correggono alcuni tipi di modificazione dei nucleotidi

L'escissione di basi ripara molti tipi di nucleotidi danneggiati

La riparazione per escissione nucleotidica corregge una gamma di estesi danni

Riparazione dei mismatch: correzione degli errori di replicazione

Le rotture del singolo o doppio filamento nel DNA possono essere riparate

Se necessario, il danno al DNA può essere aggirato durante la replicazione del genoma

Difetti nei meccanismi di riparazione del DNA sono alla base di diverse malattie umane, tra cui il cancro

**SOMMARIO 406****DOMANDE A RISPOSTA BREVE 407****PROBLEMI DI APPROFONDIMENTO 407****ULTERIORI LETTURE 408**

## CAPITOLO 17 RICOMBINAZIONE E TRASPOSIZIONE

### 17.1 RICOMBINAZIONE OMOLOGA

Modelli di Holliday e di Meselson-Radding per la ricombinazione omologa

Modello di ricombinazione omologa mediante taglio del doppio filamento

Il sistema di ricombinazione mediato da RecBCD è il più importante nei batteri

*E. coli* può effettuare la ricombinazione omologa anche con il sistema mediato da RecFOR

Ricombinazione omologa negli eucarioti

Il ruolo principale della ricombinazione omologa sembra essere la riparazione del DNA

### 17.2 RICOMBINAZIONE SITO-SPECIFICA

Il batteriofago  $\lambda$  utilizza la ricombinazione sito-specifica durante il ciclo di infezione lisogenico

Ricombinazione sito-specifica a supporto della genesi di piante geneticamente modificate

### 17.3 TRASPOSIZIONE

Trasposizione replicativa e trasposizione conservativa dei trasposoni a DNA

I retroelementi si traspongono replicativamente attraverso un intermedio a RNA

### SOMMARIO

### DOMANDE A RISPOSTA BREVE

### PROBLEMI DI APPROFONDIMENTO

### ULTERIORI LETTURE

## CAPITOLO 18 EVOLUZIONE DEI GENOMI

### 18.1 GENOMI: I PRIMI 10 MILIARDI DI ANNI

I primi sistemi biochimici si basavano sull'RNA

Primi genomi a DNA

Quanto è unica la vita?

### 18.2 EVOLUZIONE DI GENOMI SEMPRE PIÙ COMPLESSI

Le sequenze genomiche forniscono un'ampia testimonianza di duplicazioni geniche avvenute in passato

Diversi processi sono responsabili della duplicazione genica

È anche possibile la duplicazione di un intero genoma

Duplicazioni più limitate sono identificabili nel genoma umano e in altri genomi

Procarioti ed eucarioti possono acquisire geni da altre specie

L'evoluzione dei genomi implica anche il riarrangiamento dei geni esistenti

Sono state formulate ipotesi contrastanti tra per spiegare l'origine degli introni

Evoluzione dell'epigenoma

### 18.3 GENOMI: GLI ULTIMI 6 MILIONI DI ANNI

Il genoma umano è molto simile a quello dello scimpanzé

La paleogenomica ci aiuta a capire l'evoluzione recente del genoma umano

### 18.4 GENOMI OGGI: DIVERSITÀ NELLE POPOLAZIONI

Origine dell'HIV/AIDS

Prime migrazioni degli uomini fuori dall'Africa

La diversità dei genomi nelle piante è un aiuto nel miglioramento genetico di specie coltivate

### SOMMARIO

### DOMANDE A RISPOSTA BREVE

### PROBLEMI DI APPROFONDIMENTO

### ULTERIORI LETTURE

## GLOSSARIO

## INDICE ANALITICO

G-1

I-1





# MAPPATURA DEI GENOMI

## CAPITOLO

# 3

In questo capitolo studieremo i vari modi in cui vengono costruite le **mappe genomiche**. Una mappa genomica, come qualsiasi altro tipo di mappa, indica le posizioni di caratteristiche interessanti e altri punti di riferimento importanti. In una mappa genomica, queste caratteristiche e i punti di riferimento sono i geni e altre sequenze distintive del DNA. Sebbene una varietà di tecniche possa essere utilizzata per mappare i geni e altri punti di riferimento del DNA, per convenzione la mappatura genomica viene effettuata con due approcci sperimentali complementari:

- La **mappatura genetica** (Sezioni 3.2–3.4), chiamata anche **analisi di associazione** o **linkage**, che è basata sull'uso di tecniche genetiche, inclusi esperimenti di riproduzione programmati o, nel caso della specie umana, l'esame delle storie familiari (anche chiamate **pedigree**).
- La **mappatura fisica** (Sezioni 3.5 e 3.6) che utilizza tecniche di biologia molecolare per esaminare direttamente le molecole di DNA al fine di identificare le posizioni di caratteristiche sequenze, inclusi i geni.

Prima di esplorare le varie tecniche coinvolte nella mappatura genetica e fisica, dobbiamo prima capire perché sono importanti le mappe genomiche.

## 3.1 PERCHÉ È IMPORTANTE UNA MAPPA GENOMICA

Lo studio dei genomi è spesso visto come un'area moderna e intensa della ricerca biologica, molto lontana dal lavoro dei genetisti della vecchia era come Gregor Mendel. Eppure molte delle tecniche utilizzate per costruire le mappe genomiche sono basate proprio sulle scoperte di Mendel e degli altri primi genetisti. Dobbiamo quindi dedicare qualche minuto per capire perché la mappatura del genoma, nonostante sia un vecchio tipo di biologia, è ancora importante nel ritmo frenetico della ricerca nell'era genomica.

### Le mappe genomiche sono necessarie per sequenziare i genomi più complessi

Durante i primi tempi della ricerca sul genoma, si riteneva che acquisire una mappa dettagliata sarebbe stato un prerequisito essenziale per l'assemblaggio della corretta sequenza di un genoma. Questo perché il sequenziamento del DNA ha una limitazione importante: solo con la tecnologia più sofisticata e recentemente introdotta è possibile ottenere una sequenza di oltre 750 bp (*base pair*, cioè coppie di basi) in un singolo esperimento. Questo significa che la sequenza di una lunga molecola di DNA deve essere ricostruita da una serie di sequenze più corte. Ciò si ottiene tagliando la molecola di DNA in frammenti, che vengono sequenziati singolarmente e poi, mediante il computer, con appositi software, si cercano i punti di sovrapposizione per ricostruire la sequenza originaria (Figura 3.1). Questo metodo, chiamato **shotgun**, rappresenta l'approccio standard per il sequenziamento del genoma, ma presenta due problemi. Il primo è che, specialmente con genomi più grandi, potrebbe non essere possibile ottenere sequenze

### 3.1 PERCHÉ È IMPORTANTE UNA MAPPA GENOMICA

### 3.2 MARCATORI PER LA MAPPATURA GENETICA

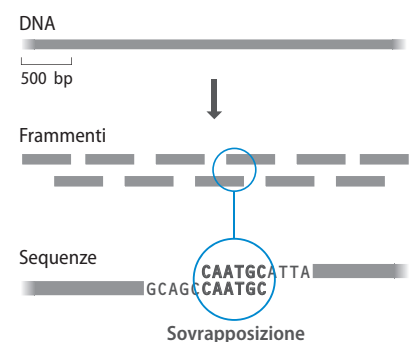
### 3.3 BASI DELLA MAPPATURA GENETICA

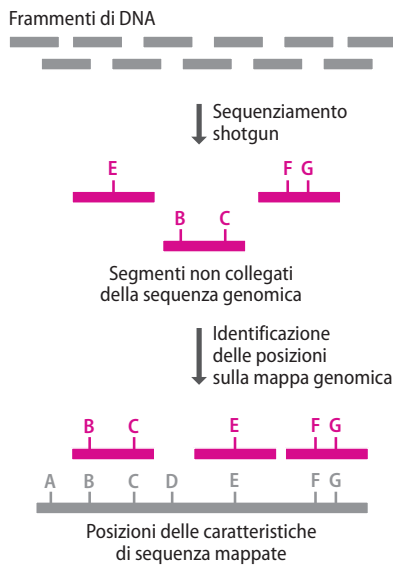
### 3.4 ANALISI DI ASSOCIAZIONE IN DIVERSI TIPI DI ORGANISMI

### 3.5 MAPPATURA FISICA MEDIANTE ESAME DIRETTO DELLE MOLECOLE DI DNA

### 3.6 MAPPATURA FISICA MEDIANTE ASSEGNAZIONE DI MARCATORI A FRAMMENTI DI DNA

**Figura 3.1 Metodo shotgun per l'assemblaggio delle sequenze.** La molecola di DNA viene spezzata in piccoli frammenti, ognuno dei quali viene sequenziato. La sequenza originaria viene ricostruita attraverso la sovrapposizione tra le sequenze dei singoli frammenti.





**Figura 3.2 Utilizzo di una mappa genomica per l'assemblaggio della sequenza.** Un genoma è stato spezzato in brevi frammenti di DNA, che sono stati sequenziati mediante il metodo shotgun. Quando le sequenze sono assemblate, si ottengono una serie di segmenti del genoma non collegati. I segmenti contengono geni e altre caratteristiche di sequenza (A, B, C, ecc.) le cui posizioni nel genoma sono state mappate. La mappa può essere quindi utilizzata per identificare le posizioni dei segmenti nella sequenza del genoma.

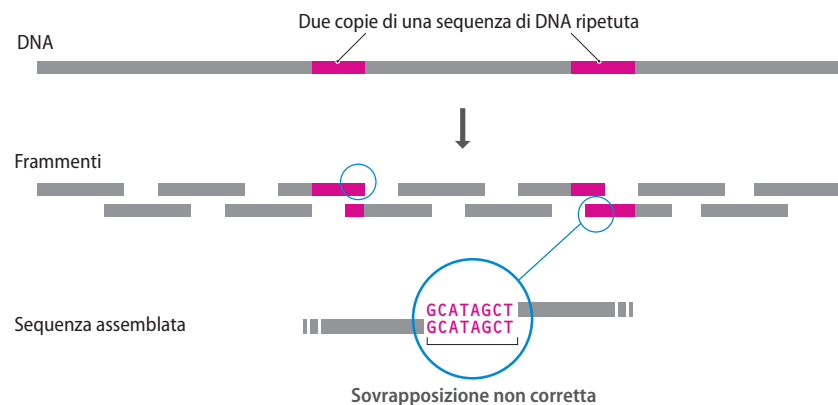
brevi sufficienti per produrre una sequenza contigua di DNA per l'intero genoma e quindi la sequenza del genoma potrebbe essere composta da molti segmenti brevi separati da lacune che rappresentano parti del genoma che, casualmente, non sono stati coperti dalle sequenze che sono state ottenute (Figura 3.2). Se questi segmenti non sono collegati, come possono essere posizionati correttamente l'uno rispetto all'altro per costruire la sequenza del genoma? La risposta è identificare all'interno di quei segmenti caratteristiche che si trovano sulla mappa genomica. Sistemando i segmenti sulla mappa, può essere ottenuta la sequenza corretta del genoma, anche se quella sequenza contiene ancora alcune lacune.

Il secondo problema con l'approccio shotgun è che può portare a errori se il genoma contiene sequenze **ripetitive di DNA**. Queste sono sequenze, lunghe molte chilobasi, che si ripetono in due o più punti in un genoma. Quando un genoma contenente DNA ripetitivo, viene tagliato in frammenti, alcuni di questi conterranno gli stessi motivi di sequenza. Sarebbe molto facile riassemblare queste sequenze in modo tale che una parte del DNA tra le ripetizioni è lasciato fuori, o collegare insieme due tratti abbastanza separati dello stesso o di diversi cromosomi (Figura 3.3A). Ancora una volta, una mappa genomica consente di evitare errori di questo tipo. Se le caratteristiche di sequenza su entrambi i lati di una regione ripetitiva corrispondono alla mappa genomica, vuol dire che la sequenza in quella regione è stata assemblata in maniera corretta. Se la sequenza e la mappa non coincidono, vuol dire che è stato fatto un errore e l'assemblaggio deve essere rivisto (Figura 3.3B).

Nel corso degli anni, la tecnologia di sequenziamento ha acquisito maggiore potenza e ha permesso di generare un numero sempre crescente di brevi sequenze da un singolo genoma: ciò significa che la probabilità che la sequenza finale contenga molte lacune è molto bassa. Allo stesso tempo, gli algoritmi informatici usati per assemblare le sequenze in segmenti contigui sono diventati più sofisticati. Gli algoritmi più recenti sono in grado di riconoscere quando l'assemblaggio raggiunge una regione ripetitiva di DNA e assicurano che la sequenza attorno a queste regioni non sia messa assieme in modo errato (Sezione 4.3). Le mappe sono diventate quindi meno importanti. Molti genomi procariotici (che sono relativamente piccoli e contengono poco DNA ripetitivo) sono stati sequenziati senza riferimento a una mappa e un numero crescente di progetti di genomi eucariotici sta facendo a meno di esse. Tuttavia, le mappe non sono ancora del tutto ridondanti come aiuto al sequenziamento del genoma. Una delle più grandi sfide oggi è ottenere sequenze genomiche per importanti colture vegetali. Molte specie vegetali hanno genomi di grandi dimensioni con un cospicuo contenuto di DNA ripetitivo. Il girasole, *He-*

**Figura 3.3 Possibili errori nell'assemblaggio della sequenza a causa del DNA ripetitivo.** (A) La molecola di DNA contiene due copie di una sequenza ripetuta. Quando vengono esaminate le sequenze mediante shotgun, i due frammenti sembrano sovrapposti ma un frammento contiene la parte sinistra di una ripetizione e l'altro frammento quella destra dell'altra ripetizione. Se l'errore di assemblaggio non viene riconosciuto, il segmento di DNA tra le due ripetizioni non verrà inserito nella sequenza finale. Se le due ripetizioni fossero su cromosomi diversi, le sequenze di questi cromosomi verrebbero erroneamente fuse assieme. (B) L'errore nell'assemblaggio della sequenza è riconosciuto perché le posizioni relative delle caratteristiche mappate (A, B, C, ecc.) nella sequenza assemblata non corrispondono alle posizioni corrette nella mappa genomica.

(A) Errore nell'assemblaggio della sequenza causato da DNA ripetitivo



(B) La mappa del genoma consente di riconoscere l'errore



*lianthus annuus*, che è una fonte di olio vegetale usata sia come alimento che come biocarburante, ne è un esempio. Il suo genoma è solo di poco più grande del genoma umano (3600 Mb per *H. annuus* rispetto a 3235 Mb per quello umano), ma l'80% del genoma del girasole è costituito da DNA ripetitivo, rispetto a solo il 44% per il genoma umano. Il genoma dell'orzo ha anche circa l'80% di DNA ripetitivo ed è molto più grande, 5100 Mb. Una sfida ancora più grande viene presentata dal pane di grano, che è un **esaploide**: ciò significa che ha tre genomi, chiamati A, B e D. Ciascuno è di circa 5500 Mb (un totale complessivo di 16.500 Mb) con un contenuto di DNA ripetitivo simile a quello dell'orzo. I progetti genoma per queste e altre colture importanti sono ancora in corso e, a causa della complessità dei loro genomi, le mappe sono essenziali per assemblare le sequenze. Questa è un'area critica di ricerca: la comprensione di tutti gli aspetti della biologia delle colture è essenziale per affrontare il problema della fame globale nei prossimi decenni.

### Le mappe genomiche non sono solo di ausilio per il sequenziamento

Le mappe potrebbero aver acquisito meno rilievo come ausili nell'assemblaggio di sequenze del genoma, ma il loro valore in altri aspetti della ricerca sulla genomica non è diminuito. È importante riconoscere che il completamento della sequenza nucleotidica di un genoma non è fine a se stesso. Infatti, ogni genoma è semplicemente una serie di A, C, G e T, e l'elaborazione di queste lettere non ci dice molto, se non riguardo al modo in cui un genoma funge da riserva di informazioni biologiche o come tali informazioni vengono utilizzate per specificare le caratteristiche della specie da studiare. Come vedremo nei Capitoli 5 e 6, la prima fase della comprensione della sequenza del genoma è identificare i geni che sono contenuti e assegnare le funzioni a quanti più geni possibili. Molti dei metodi utilizzati per assegnare le funzioni iniziano con un gene e ci si chiede cosa fa questo gene; tuttavia il processo inverso, cioè quando si parte da una funzione e ci chiediamo qual è il gene responsabile, è altrettanto importante. Come vedremo nella Sezione 6.4, una mappa genomica è essenziale per rispondere a questa seconda domanda, perché l'approccio utilizzato inizialmente comporta l'identificazione della posizione del gene di interesse rispetto ad altri geni o caratteristiche di sequenza di cui sono già note le posizioni sulla mappa. Questo processo è stato e continua ad essere la chiave per l'identificazione di geni responsabili di malattie umane come la fibrosi cistica e il cancro al seno. Metodi simili sono usati per identificare gruppi di geni, possibilmente dispersi nel genoma, che non causano direttamente una malattia ma conferiscono diversi gradi di suscettibilità a quella malattia. Un passo successivo consiste in metodi usati per identificare **loci di caratteri quantitativi (QTL, Quantitative Trait Loci)**, che sono regioni di un genoma, ognuna possibilmente contenente diversi geni, che controllano tratti variabili come la produttività della carne negli animali da allevamento e la resistenza ai parassiti nelle piante coltivate.

L'informazione fornita da una mappa genomica sulle posizioni dei geni e i QTL che controllano caratteri commercialmente importanti nelle piante coltivate sono anche utilizzate nei programmi di miglioramento genetico finalizzati allo sviluppo di nuove varietà con proprietà agricole ottimizzate. Generalmente questi programmi di coltivazione generano migliaia di piantine, le cui precise caratteristiche biologiche sono sconosciute a causa della casualità del processo di ereditarietà. Una piantina potrebbe combinare le migliori caratteristiche dei due genitori e potenzialmente rappresentare una nuova importante varietà, o potrebbe combinare le proprietà meno utili di entrambi i genitori e non essere di valore commerciale. Molti caratteri di interesse per i coltivatori vengono mostrati verso la fine del ciclo vitale delle piante – per esempio la produzione di semi o di frutti – che possono essere dosati solo crescendo ogni pianticella fino alla maturità. Ciò richiede tempo e uno spazio per la crescita di grandi dimensioni. Vedremo nella Sezione 18.4 in che modo il metodo definito **selezione assistita da marcatori** consente di utilizzare lo screening del DNA per identificare piantine che possiedono una caratteristica benefica, affinché queste ultime possano essere mantenute e quelle che invece presentano caratteristiche meno interessanti possano essere scartate. La selezione assistita da marcatori è possibile solo se è disponibile una mappa genomica. Se una mappa è disponibile, allora può essere eseguita con successo anche se la sequenza completa del genoma è sconosciuta, come nel caso delle colture dell'orzo e del grano.

### 3.2 MARCATORI PER LA MAPPATURA GENETICA

Come per qualsiasi tipo di mappa, una mappa genetica deve mostrare le posizioni di caratteristiche particolari. Su una mappa geografica questi marcatori sono rappresentati da elementi riconoscibili del paesaggio come fiumi, strade e costruzioni. Quali marcatori possiamo usare in un “paesaggio” genetico?

#### I geni sono stati i primi marcatori da utilizzare

Le prime mappe genetiche, costruite nei primi decenni del XX secolo per organismi come il moscerino della frutta, usavano i geni come **marcatori genetici**. Per essere utile nell'analisi genetica, una caratteristica ereditaria deve esistere in almeno due forme alternative o **alleli**, che specificano ciascuno un **fenotipo** diverso, come, ad esempio, i fusti alti o bassi delle piante di pisello originariamente studiate da Gregor Mendel. Inizialmente, i soli geni che potevano essere studiati erano quelli che specificavano fenotipi distinguibili ad un esame visivo. Così, per esempio, le prime mappe del moscerino della frutta mostravano la posizione dei geni per il colore del corpo, per il colore degli occhi, per la forma delle ali e simili: tutti fenotipi visibili semplicemente a occhio nudo o con un microscopio a bassa potenza. All'inizio questo approccio era adeguato, ma fu presto chiaro ai genetisti che esistevano solo un numero limitato di fenotipi visibili la cui ereditarietà potesse essere studiata e che in molti casi l'analisi era resa più difficoltosa dal fatto che più di un gene regola un singolo fenotipo. Per esempio, fino al 1922 erano stati mappati più di 50 geni sui quattro cromosomi del moscerino della frutta, ma nove di questi geni codificavano per il colore degli occhi. I genetisti che hanno studiato successivamente il moscerino della frutta hanno dovuto imparare a distinguere occhi di colore rosso, rosso pallido, vermiglio, granata, garofano, cannella, rubino, seppia, scarlatto, rosa, rosso cardinale, bordeaux, viola e marrone. Per ampliare le mappe geniche sarebbe stato necessario identificare caratteristiche più distintive e meno complesse di quelle visive.

La soluzione fu quella di utilizzare la biochimica per distinguere i fenotipi. Ciò si è dimostrato particolarmente importante per due tipi di organismi: i microrganismi e l'uomo. I microrganismi, come i batteri e i lieviti, hanno pochissime caratteristiche visibili, quindi la loro mappatura genica si deve basare su fenotipi biochimici come quelli riportati nella **Tabella 3.1**. Nel caso dell'uomo è possibile sfruttare caratteristiche visibili, ma sin dagli anni '20 del secolo scorso gli studi sulla variazione genetica umana sono stati basati ampiamente su fenotipi biochimici evidenziabili tramite tipizzazione del sangue. Questi fenotipi includono non solo i gruppi sanguigni, come il sistema ABO, ma anche le varianti delle proteine del siero e le proteine del sistema immunitario come gli antigeni leucocitari umani (il sistema HLA, Human Leukocyte Antigen). Un vantaggio sostanziale di questi marcatori è che molti dei geni importanti hanno **alleli multipli**. Per esempio, il gene *HLA-DRB1* ha oltre 1800 alleli e *HLA-B* ne ha 4200. Ciò è molto importante per il modo in cui viene realizzata la mappatura genica nell'uomo (Sezione 3.4).

**TABELLA 3.1 TIPICI MARCATORI BIOCHIMICI USATI PER L'ANALISI GENETICA DI *SACCHAROMYCES CEREVISIAE***

Marcatore	Fenotipo	Metodo con cui si identificano le cellule che contengono il marcatore
<i>ADE2</i>	Richiede adenina	Cresce solo se è presente adenina nel mezzo di coltura
<i>CAN1</i>	Resistente alla canavanina	Cresce in presenza di canavanina
<i>CUP1</i>	Resistente al rame	Cresce in presenza di rame
<i>CYH1</i>	Resistente alla cicloesimide	Cresce in presenza di cicloesimide
<i>LEU2</i>	Richiede leucina	Cresce solo se è presente la leucina nel mezzo di coltura
<i>SUC2</i>	Capace di fermentare il saccarosio	Cresce se il saccarosio è l'unico carboidrato presente nel mezzo di coltura
<i>URA3</i>	Richiede uracile	Cresce solo se l'uracile è presente nel mezzo di coltura

Piuttosto che pianificare esperimenti in cui si effettuano incroci mirati, che è la procedura normalmente adottata in organismi sperimentali quali il moscerino della frutta e il topo, i dati sulla ereditarietà dei geni nell'uomo devono essere dedotti dall'esame dei fenotipi dei membri di famiglie in cui in cui i genitori si sono uniti in matrimonio per motivi personali e non per convenienza dei genetisti. Se tutti i membri di una famiglia presentano lo stesso allele per il gene di interesse, non sarà possibile ottenere alcuna informazione utile. Per la mappatura genica è quindi necessario individuare famiglie in cui i genitori abbiano casualmente alleli diversi. Questa evenienza naturalmente è molto più probabile se il gene che si sta studiando presenta 1800 alleli invece di 2.

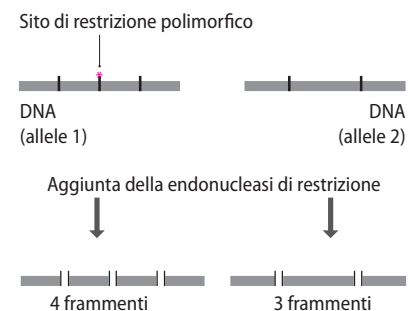
### RFLP e SSLP sono esempi di marcatori di DNA

I geni sono marcatori molto utili ma non ideali. Il problema, in particolare con i genomi più grandi come quelli dei vertebrati e delle piante da fiore, è che una mappa basata unicamente sui geni non è molto dettagliata. Questo accadrebbe anche se venisse mappato ogni singolo gene perché, nella maggior parte dei genomi eucariotici, i geni sono spazati gli uni dagli altri da ampie interruzioni. Il problema viene reso ancora più complicato dal fatto che solo una frazione del numero totale dei geni è presente in forme alleliche che possano essere distinte in modo semplice. Le mappe geniche non sono quindi molto dettagliate e si ha bisogno di altri tipi di marcatori.

Caratteristiche di sequenza mappate, diverse dai geni, sono definite **marcatori di DNA**. Come nel caso dei marcatori genici, un marcatore di DNA, per essere utile, deve essere presente in almeno due forme alleliche. Due esempi di marcatori del DNA sono le sequenze chiamate **polimorfismi della lunghezza dei frammenti di restrizione (RFLP, Restriction Fragment Length Polymorphisms)** e i **polimorfismi della lunghezza di sequenze semplici (SSLP, Simple Sequence Length Polymorphisms)**.

I primi marcatori di DNA studiati sono stati gli RFLP. Si ricordi che gli enzimi di restrizione tagliano le molecole di DNA a livello di sequenze di riconoscimento specifiche (Sezione 2.1). Per specificità di sequenza si intende che il trattamento di una molecola di DNA con un enzima di restrizione dovrebbe produrre sempre lo stesso insieme di frammenti. Ciò non è sempre vero nel caso del DNA genomico poiché alcuni siti di restrizione sono polimorfici, cioè esistono in due forme alleliche diverse, di cui una con la sequenza di riconoscimento corretta, quindi in grado di essere tagliata dall'enzima, e l'altra con un'alterazione che non ne permette il riconoscimento. Questa alterazione della sequenza ha come risultato il fatto che i due frammenti di restrizione adiacenti resteranno uniti dopo il trattamento con l'enzima, producendo un polimorfismo di lunghezza (Figura 3.4). Questo è un RFLP e la sua posizione sulla mappa genomica può essere dedotta seguendo l'ereditarietà dei suoi alleli, esattamente come quando vengono utilizzati i geni come marcatori. Si pensa che ci siano circa  $10^5$  RFLP in un genoma di mammifero.

Nel caso di molecole di DNA di piccole dimensioni, si possono distinguere i due alleli di un RFLP semplicemente tagliando con l'enzima di restrizione appropriato e identificando le dimensioni dei frammenti risultanti mediante elettroforesi su gel di agarosio. Caratterizzare un RFLP nel DNA genomico invece è più difficile. Un enzima come *EcoRI*, che riconosce una sequenza di sei nucleotidi, dovrebbe tagliare circa una volta ogni  $4^6 = 4096$  bp e così genererebbe quasi 800.000 frammenti se usato sul DNA umano. Dopo la separazione mediante elettroforesi su gel di agarosio, questi 800.000 frammenti producono una unica scia (smear) di DNA e gli RFLP non si possono distinguere gli uni dagli altri. L'ibridazione Southern fornirebbe un modo per visualizzare gli RFLP (Figura 3.5A) utilizzando una sonda che comprende il sito di restrizione polimorfico. Questo però è un processo lungo ed è difficile esaminare più di circa 12 campioni di DNA in un singolo esperimento. La caratterizzazione (tipizzazione) RFLP è una delle tante procedure che sono state rese più semplici da quando è stata inventata la PCR. Mediante PCR, infatti un RFLP può essere caratterizzato in un campione di DNA genomico senza la necessità di tagliare quel DNA con l'enzima di restrizione. I primer per la PCR invece sono progettati in modo da appaiarsi ai due lati del sito polimorfico e l'RFLP viene rivelato trattando il frammento amplificato con l'enzi-

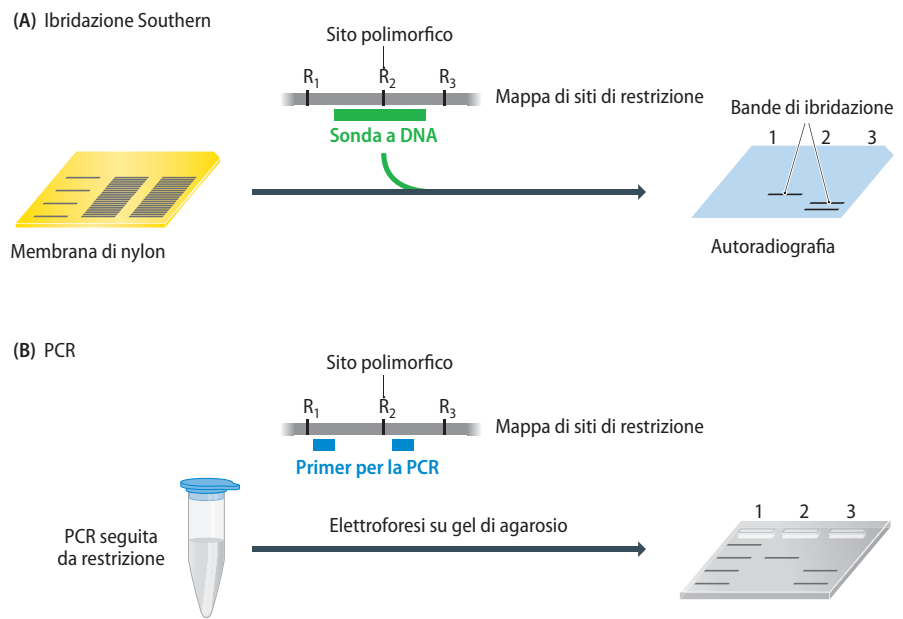


**Figura 3.4 Polimorfismo della lunghezza di un frammento di restrizione (RFLP).** La molecola di DNA sulla sinistra ha un sito di restrizione polimorfico (indicato con l'asterisco) che non è presente nella molecola mostrata a destra. L'RFLP è rivelato dopo trattamento con l'enzima di restrizione perché una delle molecole viene tagliata in quattro frammenti mentre l'altra in tre.



**Figura 3.5 Due metodi per individuare un RFLP.**

Gli RFLP possono essere individuati mediante ibridazione Southern. Il DNA viene digerito con l'enzima di restrizione appropriato e successivamente separato su gel di agarosio. L'insieme dei frammenti di restrizione viene trasferito su una membrana di nylon e ibridato con una regione di DNA che contiene il sito di restrizione polimorfico. Se il sito è assente, viene rivelato un solo frammento di restrizione (corsia 2); se il sito è presente, vengono rivelati due frammenti (corsia 3). (B) L'RFLP può essere anche caratterizzato mediante PCR, utilizzando primer che si appaiano ad entrambi i lati del sito di restrizione polimorfico. Dopo la PCR, i prodotti vengono trattati con l'enzima di restrizione appropriato e successivamente analizzati attraverso elettroforesi su gel d'agarosio. Se il sito è assente, si vede una sola banda su gel d'agarosio (corsia 2); se il sito è presente, si osserveranno due bande (corsia 3).

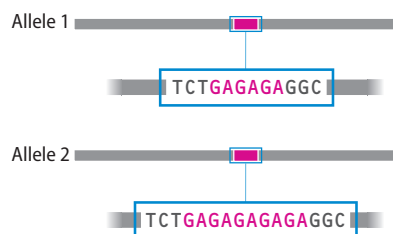
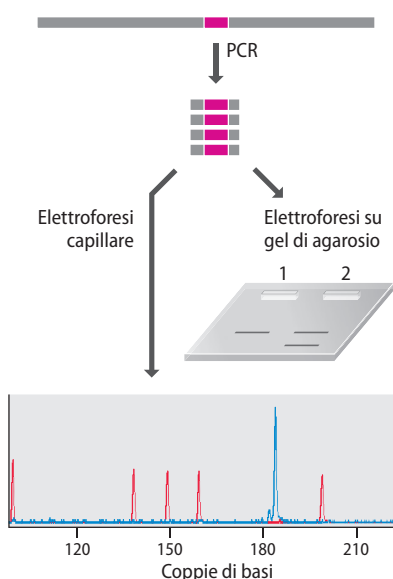


ma di restrizione (Figura 3.5B) e poi analizzandone un campione su gel di agarosio. Reazioni di PCR multiple possono essere facilmente allestite in piastre multi-pozzetto, in modo da analizzare fino a 96 campioni di DNA in un'unica analisi.

Gli SSLP sono piuttosto diversi rispetto agli RFLP. Gli SSLP sono insiemi di sequenze ripetute che mostrano variazioni di lunghezza e i cui differenti alleli contengono numeri diversi delle unità ripetute (Figura 3.6A). Diversamente dagli RFLP, gli SSLP possono essere multiallelici poiché ogni SSLP può avere diverse varianti di lunghezza. Esistono due tipi di SSLP:

- I **minisatelliti**, conosciuti anche come **ripetizioni in tandem a numero variabile (VNTR, Variable Number of Tandem Repeats)**, in cui la lunghezza dell'unità ripetuta è massimo 25 bp.
- I **microsatelliti**, o **ripetizioni in tandem semplici (STR, Short Tandem Repeats)**, le cui ripetizioni sono più corte, solitamente 13 bp o meno.

I microsatelliti sono più usati dei minisatelliti come marcatori di DNA, per due motivi. Innanzitutto, i minisatelliti non sono distribuiti uniformemente nel genoma, bensì tendono ad essere localizzati più frequentemente nelle regioni telomeriche alle estremità dei cromosomi. In termini geografici, questo equivale a cercare di utilizzare la mappa delle posizioni dei fari marini per orientarsi su un'isola. I microsatelliti sono, più vantaggiosamente, dispersi in tutto il genoma. In secondo luogo, il modo più rapido per trovare un polimorfismo di lunghezza è tramite la PCR, ma la caratterizzazione mediante PCR è molto più veloce e affidabile per sequenze di lunghezza inferiore a 300 bp. La maggior parte degli alleli

**(A) Due varianti di un SSLP****(B) Caratterizzazione di un STR mediante PCR****Figura 3.6 SSLP e modo in cui essi sono individuati.**

(A) Due alleli di un SSLP. Questo particolare esempio è una breve ripetizione in tandem (STR), chiamata anche microsatellite. Il motivo "GA" è ripetuto tre volte nell'allele 1 e cinque volte nell'allele 2. (B) Caratterizzazione di un STR mediante PCR. L'STR e parte della sequenza circostante vengono amplificate e la lunghezza dei prodotti viene determinata mediante elettroforesi su gel di agarosio o elettroforesi capillare. Nel gel di agarosio il pozzetto 1 contiene il prodotto della PCR e il pozzetto 2 contiene marcatori di DNA che mostrano le dimensioni delle bande che si ottengono in seguito alla PCR sui due alleli. La dimensione della banda nella corsia 1 corrisponde alla taglia maggiore dei marcatori di DNA, indicando che il DNA analizzato conteneva l'allele 2. I risultati dell'elettroforesi capillare vengono visualizzati come elettroferogramma, dove la posizione del picco blu indica la dimensione del prodotto di PCR. L'elettroferogramma viene calibrato automaticamente rispetto ai marcatori di peso molecolare (picchi rossi), in modo che è possibile calcolare con precisione la lunghezza del prodotto di PCR.



dei minisatelliti è più lunga, poiché le unità ripetute sono piuttosto lunghe e spesso ce ne sono molte in un singolo insieme, quindi per evidenziarli sono necessari prodotti di PCR lunghi diverse kb. I microsatelliti tipici consistono di 10–30 copie di una ripetizione che generalmente non è più lunga di 6 bp e sono quindi ideali per l'analisi mediante PCR. Sono presenti  $2,86 \times 10^6$  microsatelliti con unità ripetute di 2–6 coppie di basi nel genoma umano.

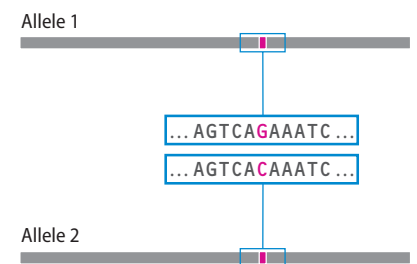
L'allele presente in una STR, quando analizzato mediante PCR, è rivelato dalla precisa lunghezza del prodotto di PCR (Figura 3.6B). Le variazioni nella lunghezza possono essere visualizzate mediante elettroforesi su gel di agarosio, ma l'elettroforesi su gel convenzionale è una procedura difficile da automatizzare e quindi non adatta alle analisi su larga scala su cui si basa la ricerca genomica moderna. Per lo più, le STR vengono caratterizzate mediante **elettroforesi capillare** in un gel di poliacrilammide. I gel di poliacrilammide hanno pori di dimensioni più piccoli e rispetto a quelli dei gel di agarosio e offrono una maggiore precisione nella separazione di molecole di diverse lunghezze. La maggior parte dei sistemi di elettroforesi capillare utilizza la fluorescenza come metodo di rivelazione, per cui un segnale fluorescente viene legato a uno o ad entrambi i primer prima di eseguire la PCR. Dopo aver eseguito la PCR, il prodotto ottenuto viene caricato nel sistema capillare dove passa attraverso un rivelatore di fluorescenza. Un computer collegato al rivelatore misura il tempo di percorrenza del prodotto di PCR confrontandolo con quello di marcatori a peso molecolare noto, e quindi identifica l'esatta lunghezza del prodotto di PCR.

### I polimorfismi di un singolo nucleotide sono il tipo più utile di marcatore di DNA

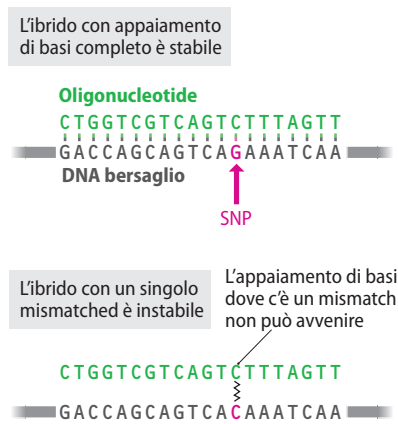
Gli RFLP e gli SSLP sono utili in alcuni tipi di ricerca genomica, ma la maggior parte dei moderni progetti di mappatura genetica utilizza un diverso tipo di marcatore di DNA. Questi sono chiamati **polimorfismi di un singolo nucleotide** (SNP, **Single-Nucleotide Polymorphisms**). Un SNP è una posizione in un genoma in cui alcuni individui hanno un nucleotide (ad esempio una G) e altri hanno un diverso nucleotide (ad es., una C) (Figura 3.7). Ci sono un gran numero di SNP in ogni genoma (circa 10 milioni nel genoma umano), alcuni dei quali danno anche origine agli RFLP, ma molti di essi non lo fanno perché la sequenza in cui risiedono non è riconosciuta da alcun enzima di restrizione.

Dal momento che in ogni posizione del genoma può essere presente uno qualsiasi dei quattro nucleotidi, si potrebbe immaginare che ciascun SNP possa avere quattro alleli. Teoricamente è possibile, ma in pratica la maggior parte degli SNP esiste in due sole varianti. Questo perché ogni SNP ha origine quando si verifica una **mutazione puntiforme** (Capitolo 16) nel genoma, che converte un nucleotide in un altro. Se la mutazione avviene nelle cellule riproduttive di un individuo, allora uno o più discendenti potrebbero ereditare la mutazione e, dopo molte generazioni, l'SNP potrebbe essersi fissato nella popolazione. Tuttavia, esistono solo due alleli: la sequenza originale e la versione mutata. Perché si origini un terzo allele, nella stessa posizione del genoma deve avvenire una nuova mutazione in un altro individuo, e questo individuo e la sua prole si devono riprodurre in modo tale da fissare il nuovo allele. Tale situazione non è impossibile, ma è improbabile: di conseguenza la grande maggioranza degli SNP è biallelica. Tale svantaggio è controbilanciato dall'enorme numero di SNP presenti in ciascun genoma: almeno uno ogni 1000 bp di DNA nella maggior parte degli eucarioti. Gli SNP permettono dunque di costruire mappe genomiche molto dettagliate.

La frequenza degli SNP in un genoma indica che questi marcatori hanno assunto notevole importanza nei progetti che utilizzano mappe genomiche per identificare geni o QTL che specificano caratteristiche particolari (Sezione 6.4), nonché nei programmi di miglioramento delle colture che utilizzano una mappa come ausilio alla selezione assistita da marcatori (Sezione 18.4). Queste applicazioni hanno portato allo sviluppo di metodi per la caratterizzazione rapida sia di singoli SNP sia di grandi gruppi di SNP. Alcuni di questi metodi si basano sull'**analisi di ibridazione di oligonucleotidi**. Un oligonucleotide è una corta molecola di DNA a singolo filamento, in genere di lunghezza inferiore ai 50 nucleotidi, che viene sintetizzata in provetta. Nelle giuste condizioni, un oligonucleotide for-



**Figura 3.7** Polimorfismo di un singolo nucleotide (SNP).



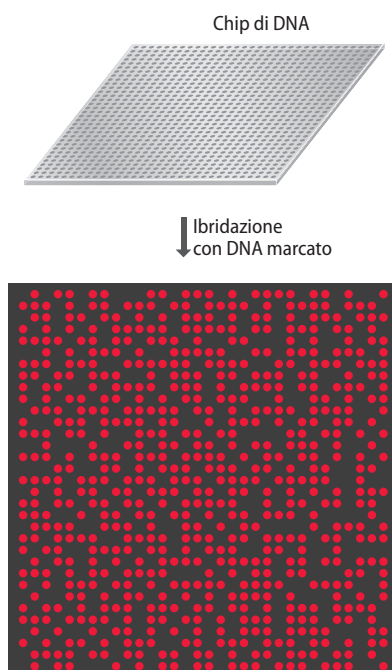
**Figura 3.8 Caratterizzazione di un SNP mediante l'ibridazione di un oligonucleotide.** In condizioni molto stringenti di ibridazione, un ibrido stabile si forma solo se un oligonucleotide è in grado di formare con il DNA bersaglio una struttura con appaiamento di basi completo. L'ibrido non si forma se c'è anche un solo appaiamento errato. Per raggiungere questo livello di stringenza, la temperatura di incubazione deve essere appena al di sotto della **temperatura di fusione ( $T_m$ )** dell'oligonucleotide. A temperature superiori rispetto alla  $T_m$  anche l'ibrido perfettamente appaiato risulta instabile. A temperature inferiori alla  $T_m$  di oltre 5°C, anche gli ibridi appaiati male possono essere stabili. La  $T_m$  dell'oligonucleotide mostrato in figura è di circa 58°C. La  $T_m$  (espressa in gradi Celsius) viene calcolata con la seguente formula:  $T_m = (4 \times \text{numero di nucleotidi G + C}) + (2 \times \text{numero di nucleotidi A + T})$ . Questa formula fornisce un'indicazione approssimativa della  $T_m$  per oligonucleotidi di lunghezza compresa tra 15 e 30 nucleotidi.

ma un ibrido con un'altra molecola di DNA solo quando si ha un completo appaiamento di basi. Se c'è anche un singolo appaiamento non corretto – una sola posizione in cui l'oligonucleotide non forma una coppia di basi – l'ibridazione non ha luogo (**Figura 3.8**). L'ibridazione con un oligonucleotide può quindi discriminare tra i due alleli di un SNP. Tra le varie strategie di analisi sviluppate basate sull'ibridazione di oligonucleotidi ci sono le seguenti:

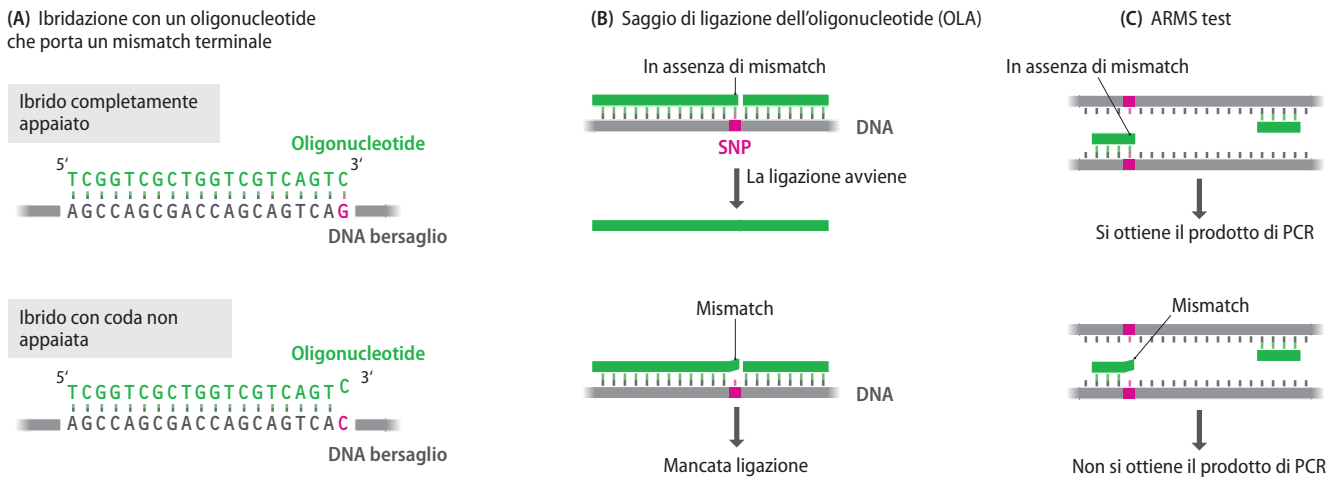
- La tecnologia del **chip di DNA** utilizza un "wafer" di vetro o silicio con un'area di 2 cm<sup>2</sup> o meno su cui sono disposti molti oligonucleotidi diversi che formano un array ad alta densità. Il DNA da testare è marcato con un fluoroforo e viene posto sulla superficie del chip con una pipetta. L'ibridazione viene rilevata esaminando il chip con un microscopio a fluorescenza. Le posizioni in cui viene emesso un segnale fluorescente indicano quali oligonucleotidi hanno formato ibrido con il DNA in esame (**Figura 3.9**). L'ibridazione richiede una corrispondenza completa tra un oligonucleotide e la sua sequenza complementare nel DNA in esame, e così indica quale delle due versioni di un SNP è presente nel DNA in esame. Sulla superficie del chip sono possibili densità fino a 300.000 oligonucleotidi/cm<sup>2</sup>, quindi un chip di 2 cm<sup>2</sup> può analizzare 300.000 SNP in un singolo esperimento, se il chip porta gli oligonucleotidi per entrambi gli alleli di ogni SNP.
- Le tecniche di **ibridazione in soluzione** vengono eseguite nei pozzetti di una piastra di microtitolazione, utilizzando un sistema di rivelazione in grado di discriminare tra il DNA a singolo filamento non ibridato e il DNA a doppio filamento che si forma quando un oligonucleotide si ibrida al campione di DNA. Il sistema di rivelazione maggiormente utilizzato si basa sullo smorzamento della fluorescenza, che abbiamo già incontrato nella Sezione 2.2 come base per il modo in cui una sonda reporter viene utilizzata per seguire la formazione del prodotto durante la real-time PCR (vedi Figura 2.21). Nella caratterizzazione degli SNP, il colorante è attaccato ad un'estremità dell'oligonucleotide e all'altra estremità al composto che smorza il segnale fluorescente. L'ibridazione tra l'oligonucleotide e il DNA in esame è indicata dalla generazione del segnale fluorescente. Quando viene utilizzato in questo contesto, la tecnica di smorzamento della fluorescenza è a volte chiamata **tecnica dei fari molecolari (molecular beacons)**.

Altri metodi utilizzano un oligonucleotide in cui il mismatch, cioè la mancata corrispondenza, con l'SNP si presenta all'estremità 5' o 3'. In condizioni appropriate, un oligonucleotide di questo tipo ibriderà con lo stampo di DNA mantenendo una piccola "coda" non appaiata (**Figura 3.10A**). Questa caratteristica viene sfruttata in due modi diversi:

- Il **saggio di ligazione degli oligonucleotidi (OLA, Oligonucleotide Ligation Assay)** che utilizza due oligonucleotidi che si appaiano l'uno accanto all'altro, con l'estremità 3' di uno di questi oligonucleotidi posizionata esattamente a livello dell'SNP. Questo oligonucleotide si appaierà completamente se è presente il giusto SNP nello stampo di DNA e in tal



**Figura 3.9 Caratterizzazione degli SNP mediante chip di DNA.** Gli oligonucleotidi sono immobilizzati in un array sulla superficie del chip. Il DNA marcato viene aggiunto e le posizioni in cui avviene l'ibridazione sono determinate da scansione laser o microscopia confocale a fluorescenza.



**Figura 3.10 Metodi per caratterizzare un SNP.** (A) In condizioni appropriate, un oligonucleotide il cui mismatch è localizzato ad una delle due estremità (5' o 3') ibriderà con il DNA stampo mantenendo una piccola coda non appaiata. (B) Caratterizzazione di un SNP mediante il saggio di ligazione degli oligonucleotidi (OLA). (C) ARMS test.

caso avverrà la ligazione con l'oligonucleotide consecutivo (**Figura 3.10B**). Se il DNA in esame contiene l'altro allele dell'SNP, il nucleotide al 3' della sonda oligonucleotidica non potrà appaiarsi allo stampo e la ligazione non avviene. L'allele è quindi caratterizzato in base al risultato della ligazione. Se è stato testato un singolo SNP, la formazione del prodotto della ligazione può essere analizzata semplicemente correndo un campione della reazione in un sistema di elettroforesi capillare, come descritto precedentemente per la caratterizzazione degli STR.

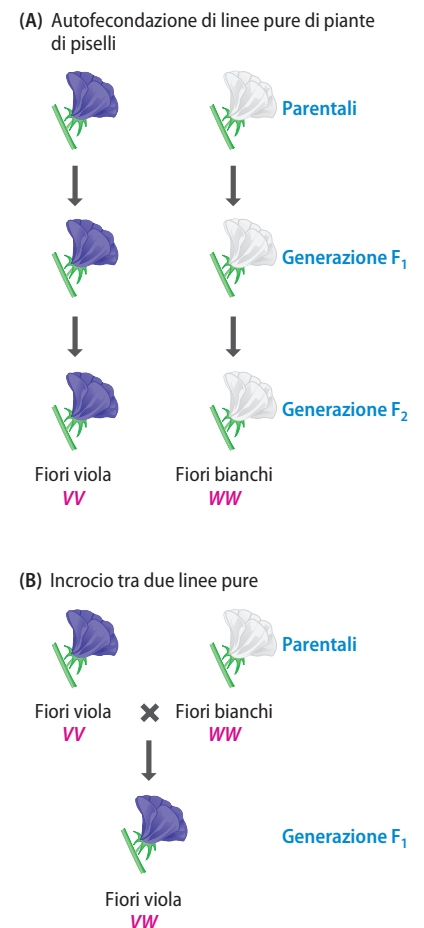
- Nel **sistema di amplificazione refrattario alle mutazioni** o **ARMS test** (**Amplification Refractory Mutation System test**) l'oligonucleotide sonda è uno dei due primer per la PCR. Se il nucleotide al 3' del primer sonda si appaia all'SNP, poi può essere esteso dalla *Taq* polimerasi e la PCR può aver luogo, ma in caso contrario, cioè se è presente l'altra versione dell'SNP, non si genera alcun prodotto di PCR (**Figura 3.10C**).

### 3.3 BASI DELLA MAPPATURA GENETICA

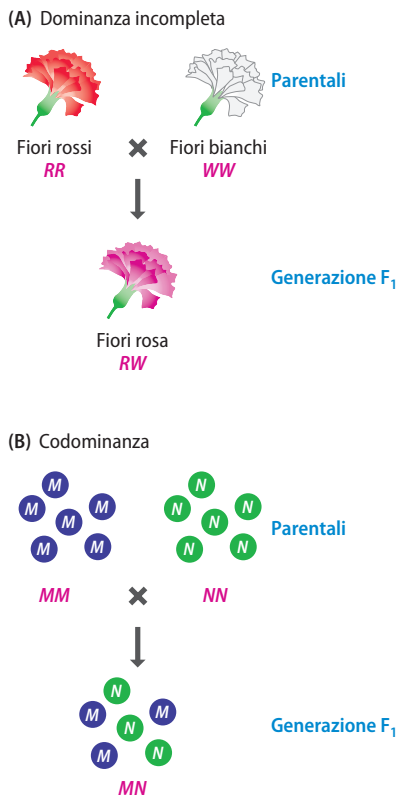
Ora che abbiamo identificato una serie di marcatori con cui costruire una mappa genetica, possiamo dedicarci alla descrizione delle tecniche di mappatura. Queste tecniche si basano sull'**associazione genetica**, che a sua volta deriva dalle scoperte effettuate da Gregor Mendel nella metà del XIX secolo.

#### Principi dell'ereditarietà e scoperta dell'associazione

La mappatura genetica si basa sui principi dell'ereditarietà descritti per la prima volta da Gregor Mendel nel 1865. Dai risultati dei suoi esperimenti di incrocio con i piselli, Mendel concluse che ogni pianta di pisello possiede due alleli per ogni gene, ma mostra un solo fenotipo. È facile capire se la pianta è una linea pura, o **omozigote**, per una caratteristica particolare, perché in tal caso possiede due alleli identici e ne mostra il fenotipo relativo (**Figura 3.11A**). Mendel dimostrò che,



**Figura 3.11 Omozigosi ed eterozigosi.** Mendel studiò sette coppie di caratteri alternativi nelle piante di pisello, una delle quali aveva i fiori di colore viola e l'altra bianco, come qui mostrato. (A) Linee pure di piante che si autofecondano danno sempre origine a fiori del colore parentale. Queste piante sono omozigoti, cioè posseggono una coppia di alleli identici, indicata da VV per i fiori viola e WW per i fiori bianchi. (B) Quando le due linee pure vengono incrociate tra loro, si osserva uno solo dei due fenotipi nella generazione F<sub>1</sub>. Mendel dedusse che il **genotipo** delle piante F<sub>1</sub> era VW, per cui V era l'allele dominante e W era l'allele recessivo.



**Figura 3.12** Due tipi di interazione allelica non affrontate da Mendel. (A) Dominanza incompleta del colore dei fiori nei garofani. (B) Codominanza degli alleli dei gruppi sanguigni  $M$  e  $N$ .

se due linee pure di piante con fenotipi diversi vengono incrociate, tutta la progenie (la generazione  $F_1$ ) ha lo stesso fenotipo. Queste piante  $F_1$  devono essere **eterozigoti**, nel senso che possiedono due diversi alleli, uno per ciascun fenotipo, ereditati rispettivamente dalla madre e dal padre. Mendel postulò che in questa condizione eterozigote un allele annulla gli effetti dell'altro. Mendel quindi descrisse il fenotipo espresso dalle piante  $F_1$  come **dominante** sul secondo, detto **recessivo** (Figura 3.11B).

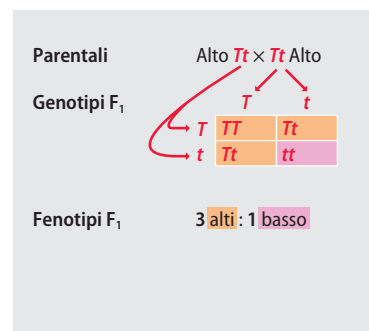
Per la coppia di alleli studiati da Mendel, questa interpretazione è totalmente corretta, ma noi ora sappiamo che la semplice regola dominante-recessivo può essere complicata da situazioni che lui non riscontrò. Queste includono:

- La **dominanza incompleta**, che si verifica quando il fenotipo eterozigote è intermedio tra le due forme omozigoti. Il colore del fiore in piante come i garofani (ma non nei piselli) ne è un esempio. Se i garofani rossi vengono incrociati con quelli bianchi, gli eterozigoti  $F_1$  non sono né bianchi né rossi, ma rosa (Figura 3.12A).
- La **codominanza**, in cui la forma eterozigote mostra entrambi i fenotipi omozigoti. I gruppi sanguigni umani forniscono diversi esempi di codominanza. Per esempio, le due forme omozigoti della serie MN sono M e N e tali individui sintetizzano rispettivamente la glicoproteina ematica M o N. Gli eterozigoti invece sintetizzano entrambe le glicoproteine e vengono definiti MN (Figura 3.12B).

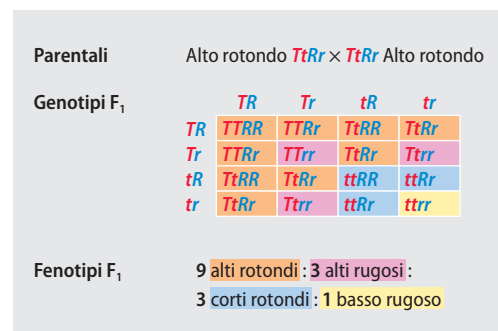
Oltre a scoprire la dominanza e la recessività, Mendel effettuò ulteriori incroci, che gli permisero di stabilire le due leggi della Genetica. La prima legge stabilisce che *gli alleli segregano in modo casuale*. In altre parole, se gli alleli parentali sono  $A$  e  $a$ , allora un membro della generazione  $F_1$  ha le stesse probabilità di ereditare  $A$  o  $a$ . La seconda legge afferma che *le coppie di alleli segregano in maniera indipendente*, quindi l'ereditarietà degli alleli del gene  $A$  è indipendente da quella degli alleli del gene  $B$ . Grazie a queste leggi, i prodotti degli incroci genetici sono prevedibili (Figura 3.13).

Quando il lavoro di Mendel è stato riscoperto nel 1900, la sua seconda legge ha preoccupato i primi genetisti perché si comprese presto che i geni sono localizzati sui cromosomi e che tutti gli organismi hanno molti più geni che cromosomi. I cromosomi vengono ereditati come unità intere, quindi si riteneva ragionevole che gli alleli di alcune coppie di geni venissero ereditati insieme poiché si trovava-

#### INCROCIO MONOIBRIDO



#### INCROCIO DIIBRIDO



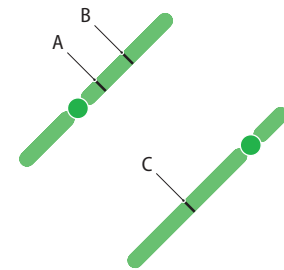
**Figura 3.13** Le leggi di Mendel hanno permesso di predire i risultati degli incroci genetici.

Sono mostrati due incroci con i risultati attesi. In un **incrocio monoibrido**, seguiamo gli alleli di un singolo gene, in questo caso l'allele  $T$  per piante di pisello alte e l'allele  $t$  per piante di pisello basse.  $T$  è dominante,  $t$  è recessivo. Il quadrato di Punnett mostra i genotipi e i fenotipi attesi nella generazione  $F_1$  in base alla prima legge di Mendel, secondo cui gli alleli segregano in maniera casuale. Quando Mendel eseguì questo incrocio ottenne 787 piante di pisello alte e 277 piante basse, un rapporto di 2,84:1. Nell'**incrocio diibrido**, vengono seguiti due geni. Il secondo gene determina la forma dei piselli; gli alleli sono  $R$  (rotondo, l'allele dominante) e  $r$  (rugoso, l'allele recessivo). I genotipi e i fenotipi mostrati sono quelli attesi in base alla prima e alla seconda Legge di Mendel; quest'ultima stabilisce che coppie di alleli segregano indipendentemente l'uno dall'altro.

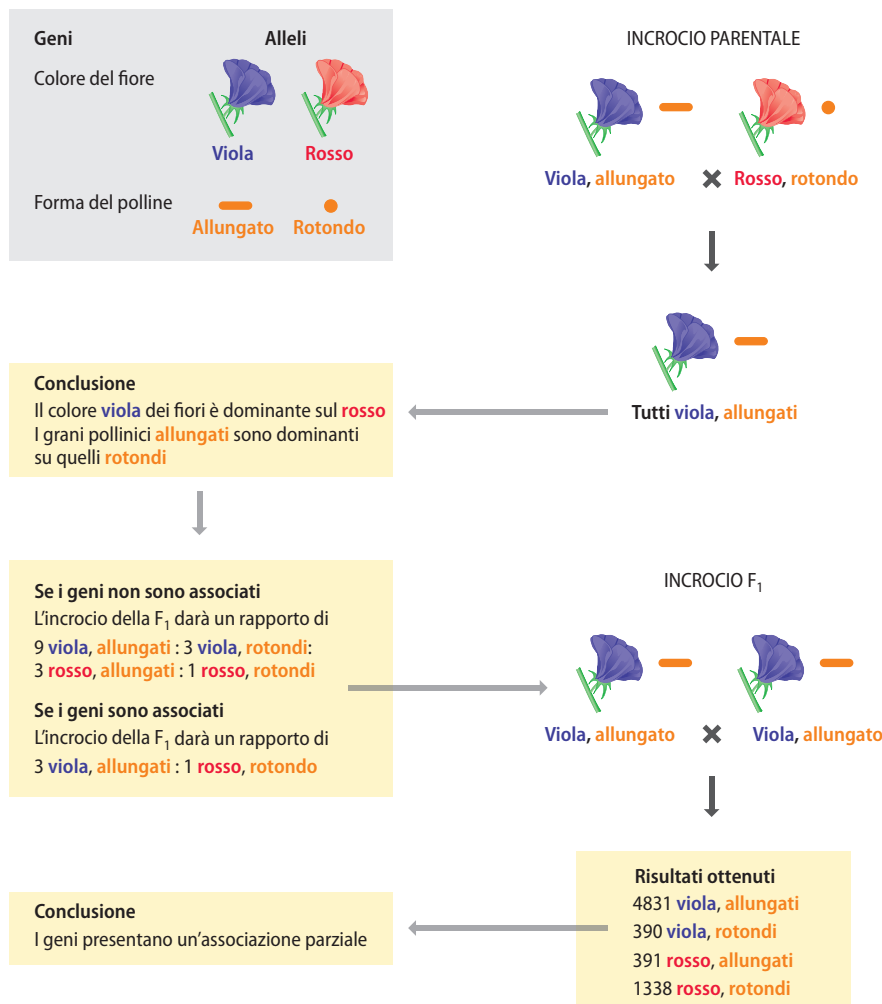
no sullo stesso cromosoma (Figura 3.14). Questo è il principio dell'associazione genetica, che si dimostrò presto corretto sebbene i risultati non si rivelarono esattamente come ci si aspettava. L'associazione completa che era stata ipotizzata tra numerose coppie di geni non si riscontrò. Le coppie di geni venivano ereditate in maniera indipendente, come ci si aspettava da geni presenti su cromosomi diversi, oppure, quando mostravano **associazione**, questa era solamente **parziale**: a volte venivano ereditati assieme e altre volte no (Figura 3.15). La soluzione di questa contraddizione tra predizione e osservazione ha rappresentato lo stadio critico nello sviluppo delle tecniche di mappatura genetica.

### Il comportamento dei cromosomi durante la meiosi spiega l'associazione genica parziale

La scoperta decisiva fu fatta da Thomas Hunt Morgan, che correlò l'associazione parziale al comportamento dei cromosomi quando il nucleo di una cellula è in divisione. I citologi, alla fine del XIX secolo, avevano distinto due tipi di divisione nucleare: la **mitosi** e la **meiosi**. La mitosi è più comune, perché è il processo con cui il nucleo diploide di una cellula somatica si divide per formare due nuclei figli,



**Figura 3.14** I geni presenti sullo stesso cromosoma dovrebbero mostrare **associazione**. I geni A e B sono sullo stesso cromosoma e dovrebbero quindi essere ereditati insieme. La seconda legge di Mendel non si può applicare all'ereditarietà di A e B. Il gene C si trova su un altro cromosoma e quindi la seconda legge di Mendel è valida per l'ereditarietà di A e C, o di B e C. Mendel non scoprì l'associazione dal momento che i sette geni che aveva studiato erano ognuno su un cromosoma diverso delle piante di pisello.



**Figura 3.15** Associazione parziale. L'associazione parziale è stata scoperta all'inizio del XX secolo. L'incrocio qui mostrato è stato effettuato da Bateson, Saunders e Punnett nel 1905 su piante di piselli dolci. L'incrocio parentale dà il tipico risultato di ibrido (Figura 3.13) in cui tutte le piante F<sub>1</sub> mostrano lo stesso fenotipo, indicando che gli alleli dominanti sono fiori viola e grani pollinici lunghi. L'incrocio della F<sub>1</sub> dà risultati inaspettati poiché la progenie non mostra né il rapporto 9:3:3:1 (atteso nel caso di geni presenti su cromosomi diversi), né quello 3:1 (atteso nel caso di geni completamente associati). Un rapporto insolito è tipico della associazione parziale.



T. A. Brown

# GENOMI 4

Accedi all'ebook e ai contenuti digitali

➤ Espandi le tue risorse

➤ con un libro che **non pesa** e si **adatta** alle dimensioni del tuo **lettore**



All'interno del volume il **codice personale** e le istruzioni per accedere alla versione **ebook** del testo e agli ulteriori servizi. L'accesso alle risorse digitali è **gratuito** ma limitato a **18 mesi dalla attivazione del servizio**.



[www.edises.it](http://www.edises.it)



€ 49,00

