

## Metodologie della genetica molecolare

### Enzimi di restrizione

Gli enzimi di restrizione proteggono la cellula batterica dai virus degradando il DNA virale. Gli enzimi di restrizione sono endonucleasi specifiche, che riconoscono ognuno una sequenza specifica di DNA, lunga da 4 a 8 nucleotidi. Le sequenze riconosciute dagli enzimi di restrizione, qualora occorrono nel genoma batterico, sono protette dalla degradazione da una metilazione in A o in C; ove queste sequenze si trovino in un DNA estraneo, non sono in genere metilate e vengono quindi tagliate dalle nucleasi di restrizione.

Le sequenze riconosciute dagli enzimi di restrizione sono palindromiche (come gli operatori e i terminatori).

Es.= ALU I

A G C T  
T C G A

La sequenza riconosciuta da un enzima di restrizione è breve, perfettamente palindromica e altamente conservata, perché altrimenti non viene più riconosciuta dal suo specifico enzima di restrizione: basta il cambio di una base (anche solo metilazione) a impedire il riconoscimento (gli operatori e i terminatori hanno sequenze più lunghe, ma meno perfettamente palindromiche e conservate).

Il sito di taglio non deve necessariamente essere al centro della sequenza di riconoscimento. Il taglio, inoltre, può essere o non essere simmetrico nei due filamenti: dipende dall'enzima di restrizione. Se il taglio non è simmetrico, verranno lasciate brevi code a singolo filamento alle estremità di ciascuna molecola di DNA risultante ⇒ **sticky ends** ("estremità appiccicose o coesive").

Le sticky ends prodotte da un determinato enzima di restrizione sono conservate, perché il sito di riconoscimento dell'enzima di restrizione è conservato a sua volta.

Se si mettono vicine due doppie eliche originate dal taglio di un'unica doppia elica originale con un enzima di restrizione che genera sticky ends, esse si riattaccheranno prontamente nel modo in cui erano prima. Rimangono due nick, saldati in seguito dalla DNA ligasi.

Con questo stesso metodo, si possono attaccare due molecole di DNA diverse, tagliate dallo stesso enzima di restrizione (quindi nella stessa sequenza di riconoscimento ⇒ sticky ends) (le sticky ends, perciò, sono utili nella clonazione del DNA (⇒ DNA ricombinante)).

N.B. = Se non ci sono sticky ends, perché il taglio da parte dell'enzima di restrizione è simmetrico, le due coppie di doppie eliche non avranno tendenza ad attaccarsi.

Ogni molecola di DNA ha dei siti di riconoscimento per enzimi di restrizione. Più lunga è, più ce ne sono.

Si può caratterizzare una regione di DNA tramite la sua mappa di restrizione (mappa dei siti per gli enzimi di restrizione). Per costruire una mappa di restrizione, si osservano le distanze relative tra i siti di restrizione.

**La mappa di restrizione consiste in una sequenza lineare dei siti che sono il bersaglio di determinati enzimi di restrizione.** È possibile costruire una mappa di restrizione

indipendentemente dal fatto che in essa siano state identificate delle mutazioni ed anche dal fatto di conoscerne o meno la funzione.

**La mappa genetica identifica una serie di siti nel DNA al cui livello sono avvenute delle mutazioni.** L'individuazione dei siti nella mappa genetica dipende dal fatto che cambiamenti nella sequenza delle basi hanno prodotto alterazioni del fenotipo.

Per vedere se e dove ci sono siti di restrizione su una molecola di DNA, bisogna metterla in contatto con i vari enzimi di restrizione, uno per volta. Ad es., se c'è un sito per EcoRI, EcoRI taglierà in quel sito.

**Elettroforesi.** La distanza di migrazione è inversamente proporzionale al pM della molecola. Ciò vale per le molecole lineari di DNA. La molecola circolare di DNA, invece, è superavvolta (= più compatta), quindi, rispetto all'equivalente molecola lineare, migrerà più velocemente. Per le molecole circolari di DNA, maggiore è il superavvolgimento, minore il pM apparente, maggiore la distanza di migrazione.

Ad es., una molecola di DNA circolare tagliata con EcoRI migra più velocemente della stessa molecola lineare tagliata con EcoRI.

Con le molecole di DNA circolari, in elettroforesi compariranno più bande, perché le varie molecole tagliate nell'esperimento hanno più o meno superavvolgimenti.

Le molecole di DNA lineare, invece, fanno comparire una sola banda.

**Mappa di restrizione.** Si può usare più di un enzima di restrizione su una molecola di DNA. Facendo vari confronti tra le posizioni dei tagli nei vari casi, si può capire dov'è approssimativamente ogni sito di taglio di ogni enzima di restrizione (mediante l'analisi dei pM dei vari frammenti). Prima si usa il primo enzima da solo, poi il secondo da solo, poi entrambi contemporaneamente.

Bisogna però confrontare il comportamento dei campioni con molecole di controllo a pM noto, che vanno sottoposte a elettroforesi insieme alle molecole da esaminare, in modo da confrontarne i pM.

Un frammento tagliato dall'enzima di restrizione A può venire tagliato dall'enzima di restrizione B oppure no, e viceversa (dipende se contiene o no il sito di riconoscimento appropriato). Alcuni frammenti sono prodotti dal taglio successivo di un enzima e poi dell'altro (o viceversa). Questi frammenti, perciò, si sovrapporranno nello spettro elettroforetico finale, derivante dalla digestione con entrambi gli enzimi. Nello spettro finale, compariranno le bande corrispondenti ai frammenti originati dal taglio con enzima A dei frammenti tagliati prima con B, e viceversa.

Confrontando mappe genetiche e mappe di restrizione, si possono identificare mutazioni puntiformi, qualora queste capitino a livello di siti di restrizione: si avrà un taglio in meno e, quindi, alla fine si troverà un frammento nella molecola di DNA mutante, la cui lunghezza corrisponde esattamente alla somma delle lunghezze dei frammenti tagliati dall'enzima di restrizione nel DNA selvatico (= sano), il cui sito di riconoscimento nel DNA mutato è scomparso e che quindi non ha potuto riconoscere e tagliare il DNA mutato.

## **Ibridazione**

Il DNA può essere denaturato (tramite un aumento di temperatura fino a circa 100°C) e poi rinaturato (tramite una diminuzione della temperatura fino a meno di 65°C): la doppia elica si riforma, grazie all'appaiamento complementare.

Si può vedere se due molecole di DNA sono complementari tramite esperimenti di ibridazione.

Si prende una doppia elica da esaminare e un filamento di DNA conosciuto e marcato radioattivamente (chiamato "**sonda**"). Si denatura il DNA da esaminare e se ne inserisce un filamento su un filtro di nitrocellulosa, sul quale si fissa saldamente (perché la nitrocellulosa forma legami covalenti con DNA a singolo filamento, RNA e proteine). Poi lo si mette a contatto con la sonda, che può venir legata tramite legami labili (legami H). Se l'appaiamento tra la sonda e una regione del filamento è avvenuto, si è verificato un fenomeno di **ibridazione** (non si parla di "rinaturazione", perché i due filamenti coinvolti provengono da molecole diverse).

Dopo che il filamento di DNA da studiare si è legato al filtro, si usa albumina per occupare tutto il resto del filtro, in modo da evitare che la sonda si leghi aspecificamente al filtro.

Sonde:

- Convenzionali ⇒ lunghe; si marcano tutti i nucleotidi
- Oligonucleotidiche ⇒ brevi; si marca solo il primo nucleotide

La marcatura di una sonda convenzionale avviene contestualmente alla sua sintesi: si forniscono alla DNA polimerasi molecole di ATP con  $\alpha$ - $^{32}\text{P}$ .

Ci sono tre modi per fornire un estremo 3'-OH libero alla DNA polimerasi per iniziare a sintetizzare una sonda convenzionale:

1. Nick translation
2. Allungamento a partire dal primer (come nella replicazione)
3. Random priming ⇒ viene sintetizzata una miscela di oligonucleotidi (6 nt) a sequenza casuale. In un lungo tratto di DNA qualsiasi, ci saranno sempre delle sequenze complementari a questi oligonucleotidi (primer). Avverrà così un appaiamento (*annealing*) tra i primer e le sequenze ad essi complementari, evento che ci permetterà di iniziare a sintetizzare le sonde convenzionali.

Per quanto riguarda la marcatura di una sonda oligonucleotidica, essa non può essere compiuta da una DNA polimerasi col metodo sopra descritto, perché questo tipo di sonda è troppo corto per essere sintetizzato dalla DNA polimerasi: bisogna perciò marcare manualmente il primo nucleotide al 5'.

Si utilizza la polinucleotide chinasi, che stacca il  $\gamma\text{-P}_i$  (che è stato marcato utilizzando  $^{32}\text{P}$ ) dall'ATP e lo trasferisce sul ribosio del primo nucleotide in 5'.

Si può marcare una sonda anche con materiale non radioattivo. Si lega dUTP, utilizzato nella sintesi della sonda, ad una molecola facilmente riconoscibile (es., digossigenina). Essa è poi individuata tramite una reazione antigene-anticorpo. L'anticorpo è legato alla fosfatasi alcalina, che riconosce e defosforila una molecola che, se perde un  $\text{P}_i$ , diventa facilmente riconoscibile (diventa colorata o fluorescente).

Alternativamente, al dUTP si può legare la biotina, che forma facilmente complessi con l'avidina, che viene legata dalla fosfatasi alcalina, che defosforila un substrato, modificandone le proprietà ottiche.

Le sonde marcate con dUTP sono denaturate, per poi essere ibridizzate col filamento da esaminare, come le sonde marcate radioattivamente.

Il processo di marcatura non radioattiva è più lungo di quello di marcatura radioattiva, ma i suoi risultati sono più facilmente riconoscibili, soprattutto se il marcatore radioattivo non risalta particolarmente: le sostanze colorate o fluorescenti risaltano molto.

Le sonde convenzionali sono resistenti alle alte temperature, perché sono lunghe e quindi contengono molti legami H (2-3 per coppia di basi). Quindi, se la sonda non è perfettamente complementare al filamento da esaminare, l'ibrido si forma comunque. Se

ci sono molte discordanze tra i due filamenti, bisogna abbassare di qualche grado la temperatura, per evitare che avvenga la denaturazione di molti legami H e l'ibrido non si formi. Questa condizione di **ibridazione** è detta a **bassa stringenza**. L'**alta stringenza** prevede una temperatura di 2-3°C inferiore alla temperatura di fusione del DNA.

Le sonde oligonucleotidiche non resistono alle alte temperature, perché sono corte e, quindi, contengono pochi legami H. Se c'è anche solo un mismatch (appaiamento non complementare), in condizioni di alta stringenza, l'ibrido non si forma. In condizione di bassa stringenza, si può tollerare un mismatch, ma non di più. Le sonde oligonucleotidiche possono quindi essere usate per scoprire se un gene è mutato (la sonda è uguale al gene sano): se il gene è mutato anche solo in una coppia di basi, l'ibridazione non avviene.

### **Dot blot**

Questo esperimento è utile per identificare entro una popolazione di sequenze tutte quelle che corrispondono a una data sonda.

Il DNA da esaminare viene messo su un filtro di nitrocellulosa. Il filtro viene poi fatto ibridare con una sonda radioattiva, che rappresenta una data sequenza bersaglio. Se tale sequenza è presente in un determinato tratto del DNA da esaminare, l'autoradiografia evidenzierà come radioattivo il punto che corrisponde al tratto ibridato.

Es. = Dot blot per verificare anemia falciforme (malattia recessiva, in cui Glu  $\Rightarrow$  Val)

Si mette una goccia di DNA da esaminare (denaturato) sul filtro di nitrocellulosa. Si immerge il filtro in una soluzione contenente la sonda (copia dell'allele sano del tratto in cui può avvenire la mutazione dell'anemia falciforme). Si abbassa la temperatura fino alla temperatura di stringenza per quella sonda. L'ibridazione avviene per gli omozigoti sani e per gli eterozigoti; non avviene per gli omozigoti malati.

Per verificare l'esattezza dell'esperimento, si tenta di far ibridare il DNA da esaminare con una sonda contenente l'allele malato dell'anemia falciforme. L'ibridazione non avverrà per gli omozigoti sani, mentre avverrà per gli eterozigoti e gli omozigoti malati. Con questa verifica si riconoscono inoltre gli eterozigoti (ibridano con entrambe le sonde).

### **Southern blot**

È un metodo che sfrutta l'ibridazione e i siti di restrizione.

Si usa un genoma intero o un tratto molto grande di DNA.

Lo si sottopone ad enzimi di restrizione. Vengono generati moltissimi frammenti, che, se sottoposti ad elettroforesi, generano una strisciata continua in cui non si distingue alcuna banda particolare.

Si denaturano i frammenti di DNA, utilizzando una soluzione alcalina (per poterli successivamente far ibridare).

Per riconoscere i frammenti, li si deve sottoporre a ibridazione. Ciò, però, non è possibile su gel (su cui si è fatta l'elettroforesi): bisogna usare un filtro di nitrocellulosa come supporto.

Si trasferiscono quindi sul filtro di nitrocellulosa i vari frammenti, mantenendo le stesse posizioni che essi avevano nello spettro elettroforetico (si sfrutta l'assorbimento per capillarità (*blotting*)). A questo scopo, si procede nel seguente modo: si pone un filtro di carta, che è stato imbevuto di una soluzione salina concentrata, sul gel di agarosio. Quindi il gel viene portato a contatto con un filtro di nitrocellulosa, al di sotto del quale si pone un filtro di carta asciutto. La soluzione salina è attratta dal filtro asciutto, ma per raggiungerlo deve prima passare attraverso il gel di agarosio, poi attraverso il filtro di nitrocellulosa. Il DNA è trascinato insieme alla soluzione salina, ma resta intrappolato nel filtro di nitrocellulosa, nelle stesse posizioni relative che occupava nel gel.

Il DNA immobilizzato nella nitrocellulosa viene poi sottoposto a ibridazione con una sonda radioattiva. Solo i frammenti complementari alla sonda ibrideranno con la sonda stessa.

Il filtro di nitrocellulosa è poi lavato con cura, per farvi rimanere attaccati solo i frammenti che hanno ibridato.

Dal momento che la sonda è radioattiva, l'avvenuta ibridazione diventa visibile con l'autoradiografia. Ogni sequenza complementare dà origine a una banda marcata, la cui posizione dipende dalla grandezza del frammento di DNA.

Per sapere il pM del frammento in esame, ne si confronta la migrazione con quella di una molecola a pM noto.

Nella prima fase (fino a prima del trasferimento su filtro di nitrocellulosa), per vedere il DNA, altrimenti invisibile, si usa bromuro di etidio, che, se si lega al DNA, è fluorescente in ultravioletto. Nella seconda, invece, si può sfruttare l'autoradiografia.

La stessa tecnica può essere applicata all'RNA (*Northern blotting*) e alle proteine (*Western blotting*).

La marcatura del DNA, necessaria al riconoscimento delle bande prodotte dall'autoradiografia, può avvenire tramite coloranti (es., bromuro di etidio: emette fluorescenza se illuminato in luce ultravioletta, mentre è legato al DNA), oppure tramite radioisotopi (es.,  $^{32}\text{P}$ ). La marcatura coi coloranti viene praticata *dopo* l'elettroforesi, quella coi radioisotopi *prima*.

Il Southern blotting è preciso ed efficace, ma lungo da realizzarsi.

Nel caso dell'anemia falciforme, ci indica che c'è anche una mutazione in un sito di restrizione (riconosciuto da Mst II). Nel gene per la  $\beta$  globina di un individuo sano ci sono tre siti di restrizione per Mst II ( $\Rightarrow$  2 frammenti). In quello di un individuo malato, ce ne sono solo due ( $\Rightarrow$  1 frammento).

L'esperimento è condotto introducendo anche tutto il genoma dell'individuo. Si taglia tutto il genoma con Mst II  $\Rightarrow$  migliaia di frammenti (in elettroforesi si nota un'unica strisciata continua). Si trasferiscono i frammenti sul filtro e lo si immerge nella soluzione con la sonda. Essa si lega alla regione della  $\beta$  globina, evidenziando quindi solo quella regione tra tutti i frammenti.

Gli individui malati (omozigoti  $\beta^s\beta^s$ ), hanno tutti e due gli alleli che mancano del sito di restrizione mutato dalla mutazione.

Gli individui sani omozigoti  $\beta^a\beta^a$  ce l'hanno su tutti e due gli alleli.

Gli individui sani eterozigoti  $\beta^s\beta^a$  ce l'hanno su un allele e sull'altro no.

Alcuni **siti** di restrizione sono **polimorfici**: sono cioè siti di restrizione in cui è presente una certa frequenza di mutazione. Il polimorfismo, però, non causa alterazioni del fenotipo, perché questi siti sono localizzati in parti non codificanti del genoma. Questo tipo di polimorfismo è noto come **RFLP** (Polimorfismo della Lunghezza dei Frammenti di Restrizione). Brevi delezioni o inserzioni cambieranno la lunghezza di qualsiasi frammento di restrizione del DNA, se avvengono in un sito di restrizione (per le delezioni) oppure se ne creano uno nuovo (per le inserzioni). Nel caso di una delezione, si avrà un frammento di restrizione in meno (e uno dei frammenti risultanti del mutante sarà lungo come la somma di due dell'allele selvatico). Nel caso di una inserzione, si avrà un frammento di restrizione in più.

Un RFLP è molto utile quando è molto comune nella popolazione, così che esiste un'alta probabilità che i genitori di un individuo portino marcatori distinguibili. Per questa ragione, brevi sequenze ripetute in tandem la cui lunghezza esatta è molto variabile nella popolazione (es., DNA microsatellite) sono la base dei marcatori RFLP più utili. Gli RFLP possono essere usati per la mappatura genetica perché le differenze di dimensione dei frammenti che un individuo eredita sono individuate facilmente mediante un Southern

blot con una sonda di DNA complementare ad una sequenza specifica di DNA di quella regione.

### **PCR (Reazione a Catena della Polimerasi)**

Permette di amplificare un miliardo di volte il DNA di una regione selezionata di un genoma, purché si conosca già almeno una parte della sua sequenza nucleotidica.

Permette di isolare una regione genomica in provetta e clonarla, senza bisogno di cellule animali. È una clonazione in vitro: si usa la DNA polimerasi purificata in provetta.

La PCR si basa sull'allungamento dei primer, tipico della replicazione del DNA.

Il primo passo consiste nel denaturare la molecola di DNA che contiene la regione che vogliamo amplificare (= replicare più volte). Poi si mettono in contatto i due filamenti con due primer, sintetizzati in modo da essere complementari alle estremità delle due regioni (una per filamento) da amplificare. I primer sono orientati in direzione  $5' \Rightarrow 3'$ , verso il centro della regione che ci interessa. Si avrà quindi l'allungamento dei primer e la replicazione della regione.

A questo punto, si denaturano i 4 filamenti (due lunghi quanto la regione da amplificare, due lunghi come tutta la molecola di DNA) e li si fanno appaiare a 4 nuovi primer (uguali ai precedenti). Si ripete questo procedimento più volte.

Il risultato finale sarà l'amplificazione di una regione di genoma, in cui ad ogni ciclo il numero di copie raddoppia (visto che i frammenti di nuova sintesi servono a loro volta da stampo).

La DNA polimerasi utilizzata è purificata da batteri termofili (es., *Thermophilus aquaticus*), quindi non solo sopporta, ma lavora meglio alle alte temperature, necessarie per la denaturazione (più di  $60^{\circ}\text{C}$ ) (una DNA polimerasi comune verrebbe denaturata assieme al DNA dai ripetuti cicli di riscaldamento).

Essa non ha però l'attività esonucleasica  $3' \Rightarrow 5'$ , quindi non può correggere i propri errori. Comunque, siccome vengono sintetizzate parecchie molecole, gli errori sono trascurabili. Se è richiesta una alta precisione, si può usare una miscela contenente DNA polimerasi sia di *T. aq.*, che di altri batteri termofili che posseggono l'attività esonucleasica  $3' \Rightarrow 5'$ , ma sono meno efficienti.

La PCR non può però riguardare sequenze ripetute, se no i primer possono appaiarsi a tratti di DNA non desiderati.

Il primo ciclo di amplificazione sarà ben definito a partire dal primer, ma all'altra estremità ci saranno discordanze, perché la DNA polimerasi si ferma in un punto più o meno casuale della restante parte del lungo filamento. Al secondo ciclo e ai successivi, però, questa imperfezione viene eliminata, perché ora ci sono più filamenti stampo ben delimitati. I filamenti originali e le loro (poche) copie dirette sono quindi in netta minoranza rispetto alle copie originate da altre copie.

Tuttavia, bisogna partire da poche molecole di DNA, per non creare troppe copie imperfette.

La PCR è compiuta a temperature diverse: alte temperature per denaturazione ( $95^{\circ}\text{C}$ ), più basse per l'annealing dei primer ( $60^{\circ}\text{C}$ ), intermedie per la replicazione ( $72^{\circ}\text{C}$ ). Il primo ciclo di denaturazione è più lungo degli altri, perché riguarda la doppia elica iniziale di DNA (può essere l'intero genoma).

La PCR non può riguardare filamenti troppo lunghi, perché la DNA polimerasi di T. aq. non può correggere gli errori e, inoltre, il processo sarebbe troppo lungo.

In due ore e mezzo circa, si riescono a creare  $10^9$  copie della regione da amplificare.

Alcuni esperimenti possono essere condotti sia col Southern blot, che con la PCR: con la PCR si lavora di meno, ci si mette meno tempo e serve meno DNA di partenza.

Es., analisi del polimorfismo dei siti di restrizione:

I primer sono stati copiati dalle regioni adiacenti conservate, quindi sono utilizzabili per tutti gli individui (quello che può variare nei vari individui sono i siti di restrizione). Si amplifica il frammento contenente i siti di restrizione. Si sottopone a elettroforesi la copia amplificata, per vedere quanto è lungo il frammento (si usa gel di agarosio per frammenti lunghi, gel di poliacrilammide per frammenti corti). Si misura la lunghezza del frammento, confrontandola con quella delle molecole di controllo a pM noto. La lunghezza è uguale alla somma dei due ipotetici frammenti tagliati con l'enzima di restrizione. Si sottopone questo frammento al taglio dell'enzima di restrizione e ad una nuova elettroforesi.

Se c'è il sito di restrizione su entrambi gli alleli, si avranno due bande corrispondenti a due frammenti, le cui lunghezze sommate tra loro danno la lunghezza del filamento amplificato (omozigote con siti di restrizione).

Se c'è il sito di restrizione su uno solo dei due alleli, si avranno due bande corrispondenti a due frammenti corti (derivanti dall'allele con sito di restrizione) e una corrispondente a uno lungo (derivante dall'allele senza sito di restrizione), la cui lunghezza è uguale a quella del filamento amplificato (eterozigote).

Se non c'è il sito di restrizione su nessuno dei due alleli, si avrà una banda corrispondente a un frammento lungo (omozigote senza siti di restrizione).

Se ci sono mismatch al centro del primer  $\Rightarrow$  esso non si appaia stabilmente (ha solo, ad es., 8 basi da una parte e 10 dall'altra appaiate correttamente). Se ci sono mismatch alla fine del primer  $\Rightarrow$  esso si appaia saldamente comunque (il resto del primer è appaiato correttamente).

Si possono creare *primer allele-specifici* (ASP), utili per identificare mutazioni puntiformi.

Uno dei due primer (ASP1 e ASP2) cade con l'ultimo nucleotide sulla eventuale mutazione. L'altro primer (CON) è normale e posizionato sull'altro filamento. Così, la replicazione dall'altro filamento avviene normalmente, mentre sul filamento contenente la mutazione puntiforme la replicazione non parte neanche, perché, nonostante l'appaiamento del primer (costruito in base all'allele sano) avvenga (è solo l'ultimo nucleotide a fare mismatch), la DNA polimerasi non inizia la replicazione, perché trova un appaiamento errato.

1. Se l'allele *non* contiene la mutazione puntiforme:

- ASP1 (primer per allele sano) si lega perfettamente e permette l'amplificazione;
- ASP2 (primer per allele mutato) ha un appaiamento errato nell'ultimo nucleotide, che impedisce alla DNA polimerasi di iniziare l'amplificazione.

2. Se l'allele contiene la mutazione puntiforme:

- ASP1 ha un appaiamento errato nell'ultimo nucleotide, che impedisce alla DNA polimerasi di iniziare l'amplificazione.
- ASP2 si lega perfettamente e permette l'amplificazione;

In un omozigote (sano o malato), a seconda del primer utilizzato (ASP1 o ASP2), l'amplificazione avverrà o non avverrà per entrambi gli alleli. In un eterozigote, qualsiasi sia il primer utilizzato, un allele sarà amplificato, l'altro no.

Esperimento per verificare le traslocazioni cromosomiche

Traslocazioni cromosomiche  $\Rightarrow$  si trovano pezzi di due cromosomi diversi attaccati; avvengono durante la meiosi e provocano tumori.

Si prendono due primer, e li si fanno appaiare rivolti verso la regione di ipotetica unione dei due pezzi di cromosomi, uno su ciascun filamento. L'amplificazione avviene solo nei soggetti malati, nei quali sono avvenute effettivamente la traslocazione e l'unione dei due pezzi di cromosomi. Nei soggetti sani, invece, l'amplificazione non avviene, perché i due primer si trovano su due cromosomi diversi.

Amplificazione (tramite PCR) di regioni del DNA caratterizzate da polimorfismo riguardante ripetizioni di sequenze molto brevi (DNA microsatelliti)

Se c'è crossing over ineguale durante la meiosi un cromatidio avrà più ripetizioni, l'altro meno. In questo modo, il figlio eredita microsatelliti più o meno lunghi, che lo distinguono dal padre e dagli altri individui.

In una popolazione ci possono essere molti alleli (non solo due, come nel caso delle malattie genetiche).

Per l'analisi è impossibile usare il Southern blot, perché la sonda si confonderebbe, a causa delle sequenze ripetute. Si usa quindi la PCR, a partire dalle regioni adiacenti non ripetute. In questo modo, i primer sono uguali per tutti gli individui (quello che varia nei vari individui è il numero di ripetizioni).

Se tra due persone il numero di sequenze ripetute è diverso, il pM delle due serie di copie amplificate è diverso. Lo si misura facendo correre i frammenti in elettroforesi su gel di poliacrilammide (perché i frammenti differiscono anche di poco  $\Rightarrow$  serve un gel a maglie strette). Si usa, come di consueto, il confronto con molecole a pM noto.

Questo esperimento permette anche di capire quale allele è stato ereditato.

Di solito, queste sequenze ripetute non sono codificanti (a meno che non facciano parte di geni). Quindi, il polimorfismo dei microsatelliti non porta ad alterazioni del fenotipo.

DNA fingerprinting

È un esperimento utile per identificare una persona.

Nel genoma umano ci sono sequenze ripetute in ogni cromosoma (LINEs, SINES, ALU: DNA minisatellite). Si taglia il genoma con un enzima di restrizione che non agisce in mezzo di una sequenza ALU. Si fa il Southern blot e si fanno ibridare i frammenti contenenti le moltissime sequenze ALU con una sonda uguale alla sequenza ALU. Confrontando le bande di un individuo con quelle di un altro nelle regioni in cui le bande sono più rade (e quindi più facilmente identificabili), si riesce a distinguere un individuo dall'altro ( $\Rightarrow$  ogni individuo ha il suo pattern caratteristico).

Analisi delle VNTR (*Variable Number Tandem Repeat*)

Oggi si usa l'analisi delle sequenze ripetute in tandem tipiche di un certo locus (VNTR) (tramite la PCR  $\Rightarrow$  si usano come primer le regioni adiacenti conservate). Queste sequenze sono più lunghe dei microsatelliti; in ogni sito, la sequenza ripetuta è unica e specifica per quel sito. L'enzima di restrizione taglia all'esterno del tratto di sequenze ripetute in tandem. In ogni individuo c'è un numero di ripetizioni diverse. Più numerose sono le ripetizioni, più lungo è il frammento amplificato con la PCR (lo si fa correre in elettroforesi  $\Rightarrow$  banda caratteristica). Il vantaggio è che lo spettro finale è pulito, in quanto contiene solo le bande corrispondenti alle sequenze che ci interessano.

Ogni individuo può avere due bande ( $\Leftarrow$  2 cromosomi). Se compare una sola banda  $\Rightarrow$  omozigote (ha lo stesso allele sui due cromosomi).



L'analisi delle VNTR dà solo due bande, quindi è affidabile fino a un certo punto: possono esistere due persone con bande uguali.

Perciò, nei casi in cui sia necessaria la sicurezza assoluta dell'individuazione dell'individuo, bisogna usare 5 sequenze VNTR diverse e altrettante sonde.

L'enzima di restrizione non deve tagliare all'interno di alcuna VNTR, ma ai margini del locus che la contiene. Si può fare il Southern blot con le 5 sonde mescolate nello stesso gel. Si avranno, quindi, 10 bande, che identificheranno univocamente una persona.

L'analisi delle VNTR può essere eseguita anche con la PCR, non solo col Southern blot. Bisogna scegliere, in questo caso, non le sonde, ma i primer. Per fare ciò, bisogna conoscere le sequenze ai margini delle VNTR. Servono due primer per locus.

Si amplificano le VNTR di ciascuno dei cromosomi omologhi di ciascun individuo. Poi si compie l'elettroforesi.

Se la quantità di DNA a disposizione è abbondante (nella PCR  $\Rightarrow$  poche bande e determinate)  $\Rightarrow$  si rendono visibili le bande con bromuro di etidio.

Se la quantità di DNA a disposizione è scarsa (nel Southern blot  $\Rightarrow$  molte bande e sovrapposte)  $\Rightarrow$  si identificano le bande grazie all'ibridazione con la sonda.

### **Clonazione: plasmidi e fago $\lambda$**

Mentre la PCR è chiamata anche "clonazione in vitro", la clonazione vera e propria è la *clonazione in vivo*.

Nella clonazione, si isola una regione genomica, la si ricombina con un vettore (molecola di DNA capace di entrare in una cellula: fagi, plasmidi), si mette a contatto il vettore con la cellula bersaglio, il vettore entra e inserisce il suo DNA nel genoma della cellula bersaglio (compresa la regione che ci interessa).

Plasmide  $\Rightarrow$  molecola di DNA circolare a doppio filamento, non ricoperta da capsida. È meno efficiente nell'infettare una cellula di un virus, a causa della mancanza del capsida. In natura i plasmidi stanno quasi sempre all'interno del batterio; si trasferiscono in un altro batterio durante i rari eventi di coniugazione tra batteri.

Quando un plasmide è entrato in una cellula, si dice che la cellula è **trasformata**.

Per facilitare la clonazione, in laboratorio, si può alterare la membrana dei batteri, in modo da renderla più permeabile ai plasmidi: si dice che le cellule batteriche diventano "competenti" all'entrata di plasmidi.

Si ottiene questa situazione sottoponendo i batteri a shock elettrici o a soluzioni iposmotiche.

Se dei batteri vengono stesi su un mezzo di coltura ricco di nutrimento, essi si riprodurranno, dando origine a cloni diversi tra loro. Tutte le cellule discendenti da una cellula (= tutti i cloni di una cellula) sono uguali tra loro.

### **Esperimento di clonazione (vettore: plasmide)**

Un plasmide (vettore) può essere tagliato – e quindi aperto (visto che è circolare) – da enzimi di restrizione, dando origine a sticky ends. Alle sticky ends può essere attaccato un pezzo di DNA estraneo al plasmide (DNA ricombinante), a patto che esso sia stato isolato tramite lo stesso procedimento con cui è stato aperto il plasmide (cioè tramite il taglio da parte dello stesso enzima di restrizione). L'inserimento avviene grazie all'annealing delle sticky ends dei due diversi DNA, le quali sono complementari tra loro. A questo punto, la DNA ligasi sigilla i 4 nick (2 per filamento).

Il risultato è l'aggiunta del DNA ricombinante al plasmide. Il plasmide è ora detto "plasmide ricombinante".

In questo modo, è possibile inserire in un batterio qualsiasi tratto di DNA noi vogliamo, tramite il vettore rappresentato dal plasmide.

Il plasmide può entrare con facilità nel batterio solo se questo è stato reso competente alla sua entrata. Il plasmide non viene distrutto dalle DNAsi batteriche perché è circolare (le esonucleasi attaccano le estremità libere di un filamento di DNA). Inoltre, il plasmide viene replicato dal batterio, perché contiene un'origine di replicazione uguale a quelle del batterio.

Per sapere quali cellule batteriche hanno ricevuto il plasmide, si marca quest'ultimo col gene per la resistenza ad un antibiotico (che peraltro è presente anche nei plasmidi naturali). Per selezionare i batteri con un plasmide all'interno, si sottopone la piastra di coltura a un antibiotico, corrispondente a quello dal quale il gene per la resistenza all'antibiotico del plasmide protegge (es., ampicillina). Così, i batteri che non contengono il plasmide che presenta il gene che protegge dall'ampicillina muoiono, e quelli che lo contengono sopravvivono.

All'interno del gene lacZ (codificante per la  $\beta$  galattosidasi dell'operone lac), è presente il cosiddetto *sito polylinker*, che serve per aprire il plasmide e permettere l'integrazione del DNA estraneo. Questo sito contiene siti di riconoscimento per molti enzimi di restrizione: il taglio con uno qualunque di questi enzimi genera sticky ends, a cui può essere congiunto un inserto di DNA estraneo.

Il sito polylinker è stato costruito in modo da rispettare la sequenza amminoacidica del gene lacZ: così la  $\beta$  galattosidasi è sintetizzata normalmente e in più ci sono i siti di restrizione dei vari enzimi.

Se, però, si trasforma il plasmide in plasmide ricombinante, siccome l'inserto di DNA estraneo è inserito nel polylinker  $\Rightarrow$  è inattivato il gene lacZ. Questo fenomeno funge da marcatore dell'inserimento del DNA ricombinante.

La  $\beta$  galattosidasi agisce sui succedanei del lattosio che contengono legami  $\beta$  galattosidici, scindendoli in glucosio e galattosio modificato (X-Gal). Quando sono liberi, alcuni galattosi modificati passano da incolore a colorati (es., blu)  $\Rightarrow$  fungono da marcatori, perché osservandone il comportamento ottico possiamo verificare l'attività della proteina codificata dal gene lac Z ( $\beta$  galattosidasi). In questo modo, si possono distinguere i batteri che contengono plasmidi ricombinanti (utili alla clonazione) dai batteri che contengono plasmidi non ricombinanti (inutili alla clonazione), perché i primi hanno colore bianco ( $\Rightarrow$  lacZ è inattivo  $\Rightarrow$  la  $\beta$  galattosidasi non è sintetizzata) e i secondi blu ( $\Rightarrow$  lacZ è attivo  $\Rightarrow$  la  $\beta$  galattosidasi è sintetizzata) (questa distinzione può avvenire, naturalmente, se nel mezzo di coltura c'è un succedaneo del lattosio, che possa eventualmente venir scisso dalla  $\beta$  galattosidasi).

Caratteristiche plasmide:

- DNA circolare a doppio filamento
- Origine di replicazione
- Gene per la resistenza a un antibiotico
- Sito polylinker (inserito nel gene lacZ), contenente sito unico per l'enzima di restrizione che ha tagliato il plasmide all'inizio (e siti per altri enzimi di restrizione)

Esperimento di clonazione (vettore: fago  $\lambda$ )

Il fago  $\lambda$  è provvisto di capsidi, quindi può infettare le cellule batteriche con facilità, senza la necessità che si rendano le cellule "competenti".

Si possono inserire le regioni che ci interessano, in mezzo ai geni regolatori (geni per la scelta tra lisi e lisogenia  $\Rightarrow$  tanto vogliamo che avvenga esclusivamente la lisi, così il fago infetta molte cellule e diffonde il DNA che ci interessa in molti batteri).

Per inserire il DNA che ci interessa, bisogna togliere dal genoma fagico una regione di lunghezza circa uguale (15-20 kb), perché altrimenti il capsido non riconosce più il genoma fagico e l'infezione non avviene (il capsido riconosce un genoma di circa 45 kb). Per l'inserzione, si taglia con un enzima di restrizione in due siti (distanti tra loro 15-20 kb), si inserisce la regione che ci interessa (isolata con lo stesso metodo) al posto del tratto da sostituire; infine, la DNA ligasi sigilla i nick.

Quando si costruiscono genomi ricombinanti di fago  $\lambda$ , si creano sticky ends nei siti cos, che si attaccano con quelle nei siti cos di altre molecole di fago  $\lambda$ . Si formano così *concatameri*, cioè lunghe doppie eliche contenenti i genomi in tandem di molti fagi  $\lambda$ . In seguito, si impacchetta ogni singolo genoma in un capsido.

Il fago  $\lambda$  può essere ricombinato con frammenti di lunghezza definita (circa uguale alla regione che viene sostituita); questi frammenti possono essere lunghi fino a 15-20 kb.

I plasmidi, invece, possono essere ricombinati con frammenti più corti (max 5 kb), ma di lunghezza variabile, in quanto questi ultimi non devono essere scambiati con tratti di uguale lunghezza.

Non servono marcatori per distinguere i batteri ricombinanti (infettati dal fago  $\lambda$ ) da quelli non ricombinanti, perché le cellule infette si lisano. Si notano, perciò, delle placchette di lisi, dovute alla replicazione dei fagi e all'estensione dell'infezione.

Bisogna disporre di un sistema di impacchettamento in vitro, per costruire un capsido intorno al genoma del fago ricombinante, perché altrimenti, senza capsido, il fago non riesce a infettare i batteri (avrebbe un'efficienza di infezione inferiore a quella dei plasmidi, visto che il suo DNA non è circolare).

### Librerie genomiche (o genoteche)

Mentre la PCR amplifica segmenti lunghi fino a un max di 10 kb, la clonazione è in grado di produrre in grande quantità frammenti di qualsiasi lunghezza.

Una **genoteca** (o **banca genomica**) è una miscela di frammenti originati da uno stesso genoma, che vengono poi amplificati, ed ogni frammento clonato viene impacchettato in un vettore. Più in dettaglio, essa è una miscela di vettori ricombinanti e di batteri trasformati (ricombinanti), che portano il DNA proveniente da uno stesso tratto di DNA diviso in frammenti. Noi possiamo conservare (in freezer) la banca, e poi scegliere al suo interno i tratti di DNA da studiare singolarmente.

Non è facile lavorare su tratti di DNA molto lunghi (più di 100 kb), perciò li si frammenta e si studiano, ad es. tramite mappe di restrizione, i frammenti più piccoli.

È così possibile costruire genoteche, formate da tanti frammenti originati dalla digestione con enzimi di restrizione di un tratto di DNA da studiare.

Ogni frammento è poi fatto ricombinare con un vettore (es., plasmide). Il risultato è un insieme di plasmidi contenenti i vari pezzi di DNA da studiare singolarmente. I plasmidi vengono in seguito inseriti nei batteri, che, replicandosi, replicano i vari frammenti che ci interessano.

In un secondo momento, si cosparge la piastra di coltura con antibiotici e x-gal, in modo da far sopravvivere e riconoscere (appaiono bianchi) solo i batteri trasformati con plasmidi ricombinanti. Così muoiono i batteri non trasformati, e i batteri trasformati con

plasmidi non ricombinanti appaiono blu (perché scindono il succedaneo del lattosio utilizzando la  $\beta$  galattosidasi).

*Trasferasi terminale*  $\Rightarrow$  crea sticky ends artificiali, per la ricombinazione del vettore col DNA di una cellula bersaglio. Essa attacca un nucleotide a nostra scelta (dATP, dGTP, dCTP o dTTP) all'estremità 3'-OH libera di una molecola di DNA, senza utilizzare uno stampo. È utile quando non si può usare lo stesso enzima di restrizione in tutti e due i siti di taglio, oppure quando l'enzima di restrizione compie tagli simmetrici, oppure ancora quando si utilizzano gli ultrasuoni per rompere il DNA (si formano tagli simmetrici).

Come si ottengono pezzi tra 15 e 20 kb da inserire nel DNA del fago  $\lambda$ , visto che i siti di restrizione sono posti a distanza variabile tra loro? Se, infatti, l'enzima di restrizione agisce indiscriminatamente, si avranno molti frammenti più corti della lunghezza necessaria, i quali non verranno incorporati nei fagi e, quindi, la libreria genomica sarà incompleta e casuale.

La risposta consiste nell'utilizzare l'enzima di restrizione in piccola quantità e per poco tempo, in modo tale che solo alcuni dei siti di restrizione siano individuati e tagliati.

Saranno privilegiati i frammenti di una certa lunghezza, a seconda del tempo in cui si lascia agire l'enzima di restrizione e della sua quantità. Poco tempo/poco enzima  $\Rightarrow$  frammenti lunghi; tanto tempo/tanto enzima  $\Rightarrow$  frammenti corti. In questo caso, si cerca di privilegiare i frammenti di 15-20 kb.

Avremo frammenti genomici tagliati in modo casuale, tutti della stessa lunghezza circa. Da ogni molecola originale di DNA si originano dei frammenti che possono essere alcuni uguali, altri diversi, altri ancora in parte uguali a quelli originati da un'altra molecola (uguale) di DNA originale. Molti frammenti, dunque, contengono regioni comuni. Ciò avviene perché un sito di restrizione in un frammento è riconosciuto, in un altro no, e viceversa.

Questa tecnica non crea tuttavia frammenti più corti di 15 kb, quindi alcune regioni non sono rappresentate. Essa è però utile con il fago  $\lambda$ , che accetta frammenti lunghi 15-20 kb.

Se, invece, non servono frammenti di lunghezza fissa (come nei plasmidi), si può digerire completamente il tratto di DNA originale: in questo modo non vengono tralasciati i frammenti brevi.

Anche per la frammentazione con gli ultrasuoni vale lo stesso discorso: si usano gli ultrasuoni per un tempo definito, in modo da creare frammenti casuali (più che con gli enzimi di restrizione) di circa 15-20 kb. Così si otterrà una banca che comprende praticamente tutti i frammenti possibili di quella data lunghezza.

La velocità di replicazione dei plasmidi non è sempre uguale alla velocità di replicazione del genoma batterico. Se le due velocità sono uguali, si parla di *plasmidi a basso numero di copie*; se, invece, la velocità di replicazione del plasmide è molto maggiore, si parla di *plasmidi multicopia*.

Ad un certo momento, noi vogliamo estrarre dalla banca uno specifico frammento (es., frammento contenente il gene per la  $\beta$  globina dalla banca genomica costituita da tutto il genoma umano). Più grande è il genoma di partenza, più numerosi saranno i frammenti possibili tra cui cercare il frammento voluto.

Se il vettore è un plasmide, si stendono su una piastra di coltura gli E. coli trasformati, in una quantità inversamente proporzionale alla probabilità di trovare il frammento voluto. Si può compiere l'ibridazione (o la PCR), su ogni capsula di Petri (che contiene circa 1000 cloni di uno specifico batterio trasformato). Si stende un filtro di nitrocellulosa, su cui si

attaccano i batteri. Si versa sull'altro lato del filtro NaOH, che lisa i batteri e denatura il loro DNA, che si attacca così al filtro. Sul filtro asciugato ci saranno vari filamenti di DNA, ognuno corrispondente al genoma batterico (sempre uguale) + il plasmide ricombinante (variabile). Si mette quindi in contatto il filtro con la sonda copiata dal gene che interessa. La sonda ibrida con uno specifico tipo di frammenti. Infine, si lava il filtro dagli altri frammenti e si compie l'autoradiografia. Poi, si può amplificare (tramite PCR) il frammento trovato, per studiarlo meglio.

Se il vettore è il fago  $\lambda$ , invece, si osservano delle placche di lisi, che fanno riconoscere le zone infettate: in queste placche ci sono cloni del fago  $\lambda$  che stanno per infettare altri batteri. Per il resto l'esperimento continua come prima: lisi dei batteri e denaturazione del DNA, ibridazione con la sonda e autoradiografia.

### **Clonazione: altri vettori**

Un **cosmide** è costituito dall'unione di un plasmide (contenente il gene per la resistenza all'antibiotico, l'origine di replicazione, ecc.) e i siti cos del fago  $\lambda$ . Un cosmide può essere ricombinato con un pezzo di DNA di lunghezza tra le 40 e le 45 kb (più lungo che con il fago  $\lambda$  come vettore), perché il capsido del fago  $\lambda$  accetta genomi di circa 45 kb, e tutto il genoma del fago  $\lambda$  – tranne i siti cos – è stato sostituito. Sono i siti cos ad essere riconosciuti dal capsido del fago  $\lambda$ .

Quando si aprono i siti cos dei cosmidi (tramite taglio con enzimi di restrizione), si formano delle sticky ends, che si uniscono con quelle di altri cosmidi, dando luogo a concatameri (successione di cosmidi attaccati per i siti cos).

Il cosmide si replica all'interno di *E. coli*, dopo averlo infettato facilmente (con lo stesso meccanismo del fago  $\lambda$ ). All'interno del batterio, però, il cosmide si replica come un plasmide (non ha i geni del fago  $\lambda \Rightarrow$  no lisi).

I vantaggi dell'uso di un cosmide, al posto di un plasmide, sono:

- Il frammento inserito è più lungo
- L'infezione è più efficiente

In un **PAC** (*P1 Artificial Chromosome*) si possono inserire frammenti di DNA lunghi fino a 100 kb. Il genoma di PAC si basa su quello del fago P1. Solo alcune sequenze sono conservate dal genoma del fago; le altre sono sostituite con quelle tipiche di un plasmide (ha geni per la resistenza a due antibiotici).

Si taglia il plasmide con gli enzimi di restrizione Sca e BamHI. Si ottengono così due braccia, tra cui si inserisce il frammento di DNA che ci interessa. Ai lati ci sono le sequenze del genoma fagico.

Il vettore viene poi impacchettato nel capsido, che riconosce un genoma di circa 100 kb (bisogna quindi fornire un frammento di DNA di circa 100 kb).

La circolarizzazione del genoma è dovuta ad una ricombinasi che riconosce due sequenze (*cre*), separate da un breve tratto, che viene tagliato via. Il DNA, poi, si chiude ad anello.

In un **BAC** (*Bacterial Artificial Chromosome*) si possono inserire frammenti di DNA lunghi fino a 120 kb. Il genoma di BAC si basa su quello del plasmide naturale della fertilità (F1) di *E. Coli*. Questo plasmide può inserirsi e poi anche escindersi dal genoma batterico.

PAC e BAC sono i vettori più usati per la clonazione del genoma umano. In precedenza si usava YAC.

Se si taglia il DNA originale in pochi siti, si otterranno pochi frammenti e, quindi, pochi saranno i cloni da studiare. È però necessario disporre di un vettore ad alta capacità. In un **YAC** (*Yeast Artificial Chromosome*) si possono inserire frammenti di DNA lunghi fino a

400 kb. Il genoma di YAC si basa su quello del lievito, un eucariote unicellulare che si riproduce molto velocemente (quasi come i batteri). Perciò è adatto per fungere da vettore, nonostante non sia un procariote.

Anche nel lievito sono stati trovati plasmidi, che però non si inseriscono nel genoma del lievito, pur venendo replicati dagli enzimi di questo. Il primo plasmide di questo tipo ad essere scoperto è il plasmide  $\mu$ .

Il lievito non è ucciso dagli antibiotici, perciò non si può utilizzare la stessa tecnica di marcatura utilizzata per i batteri. Si sfrutta invece la marcatura tramite i geni *TRP1* e *URA3*, due geni che codificano per proteine necessarie alla sintesi di triptofano e uracile. Un lievito privo di *URA3* o *TRP1* cresce solo in un terreno ricco di nutrienti. Un lievito selvatico, invece, può crescere anche in un terreno povero di nutrienti. Solo se si ridanno ad un lievito che ne è privo i geni *URA3* e *TRP1* (tramite un plasmide), il lievito torna a poter vivere in un terreno povero. I geni *URA3* e *TRP1* sono geni di *autotrofia*: i lieviti che ne sono privi hanno bisogno che uracile e triptofano siano forniti dall'esterno; i lieviti selvatici sono detti *prototrofici* (se li costruiscono da sé).

Noi vogliamo clonare grandi pezzi di DNA all'interno del cromosoma del lievito, per creare YAC. Bisogna usare sempre lo stesso enzima di restrizione, usandolo a basso dosaggio per poco tempo, in modo da generare frammenti poco numerosi e lunghi. Solo YAC di lunghezza definita possono venir riconosciuti dagli apparati della cellula di lievito.

Del genoma del lievito si conservano il centromero e i due telomeri (oltre ai geni *URA3* e *TRP1*).

Tuttavia, si è dimostrato che negli YAC possono avvenire frequenti fenomeni di ricombinazione (più che negli altri vettori), che alterano i risultati degli esperimenti. Quindi sono stati preferiti allo YAC PAC e BAC, in cui questi fenomeni sono molto meno frequenti.

## **Genoteca di cDNA**

È una genoteca che contiene solo i geni espressi nelle cellule di un certo tessuto o in un certo tipo cellulare. È una banca di DNA a doppia elica, ma, per ottenerla, non si frammenta una grossa molecola di DNA di partenza: si parte dagli mRNA presenti nel tipo cellulare che ci interessa, li si retrotrascrive con la trascrittasi inversa, si fanno ricombinare i cDNA con altrettanti vettori, creando così la banca di cDNA.

I cloni genomici rappresentano un campione casuale di tutte le sequenze di DNA di un organismo e, salvo rarissime eccezioni, saranno uguali indipendentemente dal tipo cellulare usato per prepararle. I cloni di cDNA, invece, contengono soltanto quelle regioni del genoma che sono state trascritte in mRNA; poiché le cellule di tessuti diversi producono serie distinte di molecole di mRNA, da ciascun tipo di cellula usato per preparare la genoteca si otterrà una genoteca di cDNA diversa.

In una genoteca di cDNA non si trovano geni normali: essi infatti hanno perso gli introni e certe sequenze al 3' e al 5'.

Innanzitutto, si estraggono tutti gli RNA da un tipo cellulare in fase solubile. Si separano poi gli mRNA dagli altri RNA, tramite il riconoscimento della sequenza di poli(A) (tail). Si sfrutta per questo scopo la cromatografia di affinità, nella quale in una colonna si utilizza una resina che funge da filtro che trattiene (legandovisi tramite legami H) tutte le molecole con una certa caratteristica (in questo caso il poli(A) al 3') e fa passare tutte le molecole che ne sono prive (gli altri RNA). Per riconoscere la coda di poli(A), si sfruttano degli oligo(dT), che vengono legati covalentemente dalla resina.

Avremo quindi purificato gli mRNA, rispetto agli altri RNA. Gli mRNA si staccano poi facilmente dalla resina, rompendo i labili legami H.

In seguito, si sottopongono gli mRNA a trascrittasi inversa, che li trascrive in cDNA, che possono essere ricombinati con vettori per creare poi una genoteca di cDNA di un certo tipo cellulare.

La trascrittasi inversa è una DNA polimerasi particolare, perché usa come stampo RNA e non DNA. Essa, come tutte le DNA polimerasi, ha bisogno di primer: le si forniscono perciò oligo(dT). In questo modo inizia la trascrizione inversa dell'mRNA, a partire dalla coda di poli(A). Il secondo passaggio è la distruzione parziale del filamento di mRNA da parte della RNasi H. Essa lascia pezzetti di mRNA che fungono da primer per una DNA polimerasi che replica il filamento di DNA in precedenza copiato, rimuovendo contemporaneamente i primer. Il risultato finale è una doppia elica di cDNA, che però non ha sticky ends per la clonazione: si usa perciò la trasferasi terminale per procurarglieli (le si danno, ad es., dCTP per il cDNA e dGTP per il vettore  $\Rightarrow$  le sticky ends artificiali si riattaccano  $\Rightarrow$  ricombinazione cDNA-vettore).

Le banche di cDNA sono molto meno estese delle banche genomiche, perché il numero di mRNA di una cellula è molto inferiore a quello dei frammenti di DNA che si possono ottenere dalla digestione di un genoma intero (visto che solo circa il 10% del genoma codifica per mRNA).

Si possono utilizzare vettori a capacità ridotta (plasmidi e fago  $\lambda$ ), perché gli mRNA di partenza sono corti.

La banca di cDNA può essere utilizzata normalmente per immagazzinare geni, ma può anche essere sfruttata per esprimere proteine in organismi eterologhi (es., batteri). Bisogna, a questo proposito, che il gene non sia semplicemente immagazzinato, ma venga anche espresso: ci vuole un promotore batterico. Il promotore va inserito nel plasmide (insieme a lacZ, all'origine di replicazione, al gene per la resistenza all'antibiotico), a monte del sito polylinker. Al promotore vanno anche aggiunti un ribosome binding site e, spesso, un operatore, in quanto può essere necessario che l'espressione del gene inserito sia modulata (ad es., qualora la proteina estranea interferisca con le attività del batterio in certi momenti). Gli operatori più usati a questo proposito sono quelli degli operoni lac e trp e del repressore del fago  $\lambda$ . Un vettore con queste caratteristiche (promotore, operatore e ribosome binding site batterici + cDNA di un gene eterologo) è detto **vettore di espressione**. Per costruire vettori di espressione, si possono usare dei plasmidi o il fago  $\lambda$ , ma non PAC, YAC o BAC.

I vettori di espressione servono anche a produrre in grande quantità una proteina che si vuole studiare, ma che è presente solo in minima parte nelle cellule. In questo caso, bisogna usare un promotore forte.

I discendenti delle cellule batteriche trasformate in questo modo hanno immagazzinato stabilmente il cDNA, e lo replicano come parte integrante del proprio genoma.

In una banca di cDNA ci sono vari tipi di batteri contenenti ognuno un cDNA diverso: ogni cDNA non ha sequenze comuni a quelle di un altro cDNA, perché non ci sono sovrapposizioni dovute a tagli con enzimi di restrizione.

Per riconoscere, all'interno della banca, un batterio con un certo cDNA, si può usare l'ibridazione (come con le banche genomiche), oppure cercare quale proteina estranea ogni batterio trasformato esprime, tramite una reazione antigene-anticorpo. Si tratta la nitrocellulosa in modo che esponga la proteina attaccata all'anticorpo. Le cellule sono però rotte non con NaOH (che denaturerebbe anche la proteina), ma con vapori di cloroformio.

## Proteine di fusione e organismi transgenici

Il gene eterologo da inserire nel plasmide che agisce come vettore di espressione va inserito nel gene lacZ, vicino al polylinker. Il promotore/operatore è quello dell'operone lac  $\Rightarrow$  la trascrizione del gene della proteina di fusione verrà regolata dall'operatore dell'operone lac. Si avrà il gene di una **proteina di fusione**, formato da un pezzo di gene della  $\beta$  galattosidasi e dal gene della proteina eterologa da inserire. È utile mantenere all'estremità 5' del gene della proteina di fusione un pezzo di gene della proteina batterica, in modo da difendere la proteina di fusione dalle proteasi del batterio.

Spesso si può isolare la proteina di fusione dalle altre proteine prodotte dal batterio, inserendo nel gene della proteina di fusione una serie di 6 codoni codificanti per His, che ha alta affinità per il nichel. Così, mediante una cromatografia di affinità, il nichel si attaccherà alla proteina di fusione, rendendola riconoscibile. Se le His interferiscono, è comunque possibile staccarle in un secondo momento.

La tecnica delle proteine di fusione può essere utilizzata anche per far produrre ad animali (resi transgenici) una proteina che ci interessa, unendo il suo gene a quello di una proteina facilmente prelevabile (es., proteina secreta nel latte di una pecora).

Alcune proteine eucariotiche funzionano solo se hanno subito modifiche post-traduzionali (es., glicosilazione, fosforilazione, ecc.), che però non possono avvenire nei procarioti. Queste proteine non riuscirebbero quindi a venir espresse nei procarioti; perciò le si fanno esprimere nel lievito (bisogna usare un plasmide di espressione per lievito con promotore per lievito).

Spesso i plasmidi di espressione per lievito contengono anche un'origine di replicazione per batteri e un gene per la resistenza ad un antibiotico. Questi plasmidi possono perciò essere replicati (più velocemente) nei batteri, per poi essere utilizzati normalmente nel lievito (*vettori navetta*).

L'*Agrobacterium tumefaciens* è un batterio che può indurre il tumore del colletto in alcune piante. Si riproduce se riesce ad entrare nel tessuto della pianta (es., tramite una ferita del tronco). Ha un plasmide (plasmide TI  $\Rightarrow$  induce il tumore) che porta la **sequenza DNA T**, che è responsabile dell'induzione del tumore. La sequenza DNA T si integra nel genoma delle cellule della pianta, provocandone l'ipertrofia ( $\Rightarrow$  tumore).

Le proteine codificate dal batterio (tra cui quelle della sequenza DNA T) non possono essere utilizzate dalle cellule della pianta (e per di più sono dannose), quindi è possibile sostituirne i geni in laboratorio con dei geni a nostra scelta (geni che migliorano la vita della pianta), mantenendone però il promotore (che è forte  $\Rightarrow$  quindi i geni che introduciamo saranno sempre espressi), sfruttando così l'*Agrobacterium tumefaciens* come vettore. Questa tecnica è largamente utilizzata nella creazione di piante transgeniche.

Alternativamente, tramite l'*Agrobacterium tumefaciens* è possibile inattivare i geni della pianta che hanno effetti non graditi per i nostri scopi (es., gene che fa raggrinzire i pomodori maturi). Per fare ciò, si inserisce, dopo il promotore per DNA T, l'RNA antisenso del gene che vogliamo inattivare  $\Rightarrow$  pianta transgenica.

C'è anche un altro metodo per inattivare dei geni nelle cellule di un organismo: provocare mutazioni o delezioni in parte o in tutto il gene.

Nei geni del lievito è particolarmente facile provocare mutazioni costruendo un gene alterato (es., con delezioni), inserendolo poi in un plasmide di lievito, che viene fatto ricombinare (tramite ricombinazione omologa) col gene residente  $\Rightarrow$  il gene alterato del plasmide sostituisce quello residente, inattivando così l'espressione della proteina da quest'ultimo codificata.



Questa tecnica può servire per sostituire e inattivare il gene residente, oppure per ottenere la proteina mutata facendola esprimere in un normale vettore d'espressione.

Con la PCR si può sostituire un codone per creare una mutazione puntiforme nella proteina da esprimere. In *E. coli* esistono metilasi che metilano il DNA in certe basi in determinate sequenze (es., G A T C  $\Rightarrow$  metilazione di A).

Il plasmide originale viene fatto replicare in degli *E. coli* in cui le metilasi provocano la metilazione  $\Rightarrow$  il plasmide diventa foriero di un gene mutante. In questo modo si distinguono i plasmidi di partenza da quelli mutanti.

In seguito, si denaturano i filamenti dei plasmidi (mutanti e non) e si usano due primer che portano a loro volta la base modificata, così la mutazione può venire amplificata con la PCR. Si avranno così alcuni plasmidi con entrambi i filamenti metilati, altri con uno dei due filamenti metilato (perché hanno ricevuto il primer metilato, ma non erano stati metilati nei batteri; la replicazione, in questo caso, avviene a bassa stringenza, per ignorare il mismatch), altri ancora con nessuno dei filamenti metilato.

C'è un enzima (Dpn I) che riconosce la sequenza G A T C, solo se ha la A metilata. Così si possono selezionare i plasmidi con entrambi i filamenti mutanti. A questo punto abbiamo plasmidi mutanti che possono essere usati per mutare un gene in un organismo.

## **Sequenziamento del DNA**

Ci sono due tipi di sequenziamento: "normale" e ciclico.

Il **sequenziamento "normale"** è il più antico e si basa sulla terminazione controllata della replicazione. Si usano frammenti di DNA clonati in un plasmide, così si può usare un primer universale, che si appaia con il tratto di DNA del plasmide a monte del tratto di DNA inserito da sequenziare.

Prima di tutto, si isola il tratto della doppia elica di DNA da sequenziare e lo si fa ricombinare col plasmide. Poi, si fa appaiare a una delle due regioni del plasmide adiacenti a quella in cui è stato inserito il tratto di DNA da sequenziare, il primer universale e si fa iniziare la replicazione.

La chiave del meccanismo è somministrare alla DNA polimerasi, oltre ai soliti 4 deossinucleotidi (dNt), anche un dideossinucleotide (ddNt) (un tipo di ddNt e tutti i 4 tipi di dNt). La DNA polimerasi riconosce e attacca alla catena nucleotidica in crescita il ddNt, ma questo, essendo privo di un gruppo OH anche in 3' (non solo in 2', come i dNt), non permette l'ulteriore allungamento della catena in crescita.

In questo esperimento, si utilizzano 4 provette, in ognuna delle quali si inseriscono i 4 dNt in grande quantità e uno dei 4 ddNt (ddATP, ddGTP, ddCTP o ddTTP) in piccola concentrazione e avviene la reazione di allungamento della catena in crescita. Quando la DNA polimerasi "pesca" e incorpora il ddNt, si ha l'arresto della replicazione. Siccome il sequenziamento coinvolge molte molecole uguali di DNA (non solo una) e siccome, in ciascuna delle 4 provette, la terminazione della catena avverrà a caso in una delle molte posizioni contenenti la base in questione (in modo casuale, perché [ddNt] è molto bassa), avremo in ogni provetta tutti i possibili filamenti che terminano con quel dato ddNt.

In questo modo, complessivamente avremo tutti i possibili filamenti allungati a partire dal primer universale, ognuno lungo una base più di un altro, in quanto ogni filamento è originato dalla replicazione con conseguente terminazione controllata in corrispondenza di uno specifico nucleotide del filamento stampo. Guardando il ddNt terminale di uno di questi filamenti si riesce a individuare il dNt corrispondente del filamento che si sta sequenziando (es., se c'è ddATP, vuol dire che esso si sarà appaiato con un dTTP del filamento stampo).

Compiendo un'analisi di tutti i filamenti così originati, riusciremo a scoprire base per base la sequenza del filamento di DNA che stiamo sequenziando.

È importante notare che il filamento da sequenziare rimane sempre lo stesso in tutte le fasi dell'esperimento, mentre ogni filamento originato con la terminazione controllata della replicazione di quello stampo può essere usato solo una volta.

È possibile che siano presenti frammenti replicati in toto, i quali vengono però facilmente individuati e quindi scartati.

Per distinguere i vari frammenti (grazie alla determinazione dei loro pM), li si sottopone a elettroforesi su gel di poliacrilammide (gel a maglie strette e sensibile alle differenze anche minime di pM) in 4 diverse corsie (una per ogni provetta), previa marcatura dei primer, dei ddNt, o dei dNt. In ciascuna corsia compariranno le bande corrispondenti a tutti i filamenti terminanti col ddNt relativo a quella corsia. In questo modo, per ogni base del filamento sequenziato c'è una banda corrispondente a uno dei 4 ddNt (quindi, ciascuno dei 4 ddNt presenta più bande). Dal profilo di bande si può quindi leggere la sequenza del DNA. Se manca una banda, vuol dire che c'è stato un errore nel sequenziamento (una base del filamento in esame non è stata determinata). È importante sottolineare che ogni banda differisce dalla banda adiacente per solo una base (per quanto riguarda il pM).

La marcatura dei 4 ddNt è stata in seguito perfezionata: i ddNt vengono marcati con colori fluorescenti diversi, in modo tale da poter far procedere le 4 reazioni insieme e poter utilizzare una sola corsia in elettroforesi (le bande hanno colori diversi, che sono riconosciuti da un raggio laser in base alla lunghezza d'onda). Con questa tecnica, e grazie all'ausilio di computer, il risultato del sequenziamento ci perviene in modo automatico.

Il punto fondamentale del sequenziamento del DNA è che i ddNt sono incorporati casualmente, in modo da produrre tutti i filamenti possibili.

È importante segnalare, infine, che è necessario verificare il risultato del sequenziamento del filamento in esame, sequenziando anche il filamento complementare e controllando che possa avvenire un appaiamento perfetto tra ciascuna delle coppie di basi individuate nei due filamenti.

Oggi, tuttavia, il sequenziamento "normale" è stato perfezionato e perciò si fa uso del cosiddetto **sequenziamento ciclico**.

Il sequenziamento ciclico è un metodo di sequenziamento simile alla PCR (DNA polimerasi termostabile e cicli di denaturazione-appaiamento-sintesi di DNA). La differenza sta nel fatto che il sequenziamento ciclico fa uso solo di un primer e include nella reazione un ddNt (terminatore della catena). L'uso di un solo primer implica che il prodotto si accumula con un incremento lineare, non esponenziale (come invece accade con la PCR).

Questa tecnica fa uso della DNA polimerasi termostabile di T. aq. (vedi prima), oltre che dei 4 dNt e di un ddNt diverso per ogni reazione. Questa variazione è determinata dal fatto che il DNA deve essere denaturato in molti cicli successivi ed è quindi più pratico lavorare con una DNA polimerasi resistente alle alte temperature.

Si hanno molti cicli di replicazione, per l'amplificazione della stessa molecola di DNA da sequenziare. In questo modo, aumenta la quantità di molecole interessate dal sequenziamento: ogni molecola di DNA copiata dal filamento stampo viene utilizzata più volte, venendo copiata a sua volta.

Nel sequenziamento ciclico, la quantità di DNA di partenza richiesta è minore che nel sequenziamento "normale". Nel sequenziamento "normale", ogni volta che un filamento stampo è usato per la copia necessaria al sequenziamento, non può più venir utilizzato (in quanto questo metodo non prevede cicli di denaturazione e riappaiamento). Il sequenziamento ciclico, invece, è più evoluto, perché ogni volta che viene fatta una copia del filamento da sequenziare, la doppia elica così formata viene denaturata e il filamento da sequenziare viene riutilizzato.

Il risultato viene anche in questo caso verificato tramite il sequenziamento (con la stessa tecnica) del filamento complementare.

## Progetto genoma

Dna microsatellite  $\Rightarrow$  utile per costruire mappe genetiche

Inizialmente, si marcavano certi tratti del DNA dei 2 gameti e li si faceva unire, poi si analizzava la disposizione dei marcatori nella cellula figlia. Si localizzava la posizione (e quindi la distanza da altri marcatori) di un marcatore (di solito un gene) tramite la analisi della frequenza di ricombinazione con un altro marcatore: più vicini sul cromosoma sono i marcatori, minori sono le probabilità che, durante il crossing over, avvenga un evento di ricombinazione tra di essi.

Questo procedimento per l'uomo era innapplicabile (troppo lungo e complicato). Perciò si impiegò l'analisi dei DNA microsatelliti, sequenze ripetute in tandem che variano da individuo a individuo (ci sono molti alleli diversi all'interno di una popolazione), ma che non comportano alterazioni del fenotipo (come invece prevedeva la tecnica precedente, che si basava su mutazioni di geni con ripercussioni sul fenotipo). Nei vari individui, il numero di ripetizioni di sequenze microsatelliti in un locus è determinato dal crossing over ineguale. Siccome ci sono molti alleli, è più facile costruire mappe genetiche, ricostruendo gli eventi di ricombinazione avvenuti nella genesi dell'individuo dai genitori.

DNA minisatelliti  $\Rightarrow$  sequenze di 30-60 bp ripetute in tandem  $\Rightarrow$  sono usati per distinguere un individuo da un altro (DNA fingerprinting); sono presenti quasi solo nei telomeri (non sono perciò utili per costruire mappe genetiche)

DNA microsatelliti  $\Rightarrow$  sequenze di poche bp ripetute in tandem  $\Rightarrow$  riconosciuti solo con la PCR (non col Southern blot); sono presenti in tutto il genoma (utili per costruire mappe genetiche)

A partire dalla mappa genetica (che tiene traccia delle mutazioni avvenute nei geni), si è arrivati ad una mappa fisica del DNA (es., mappa di restrizione, mappa della sequenza di basi).

Prima di tutto, si sono separati tra loro i cromosomi. Non si è potuta usare l'elettroforesi, perché essi sono troppo grossi.

A questo scopo, si sono creati **eterocarionti** tra cellule umane e cellule di topo: si crea un nucleo ibrido instabile (i cromosomi umani vengono espulsi gradualmente); in alcuni eterocarionti, però, dopo un po' si stabilizzerà un eterocarionte con un solo cromosoma umano a caso (gli altri vengono espulsi).

Un altro metodo di separare i cromosomi è la **citometria a flusso**: si separano i diversi cromosomi grazie alla quantità di una sostanza fluorescente che si lega al DNA, in quantità proporzionale alla grandezza della molecola di DNA (e quindi dei cromosomi). Inoltre, ci sono coloranti a diversa affinità per coppie A-T o C-G, che si attaccano più o meno ad ogni regione del cromosoma, in base alla sua composizione in basi. Con la citometria a flusso si separano i cromosomi in base alla quantità e alla localizzazione del colorante legato.

L'**elettroforesi a campo pulsante** è utilizzata per separare i cromosomi più piccoli (es., cromosomi di lievito o frammenti molto grossi di cromosomi umani) (cioè molecole di DNA molto grosse). La direzione del campo elettrico viene continuamente variata, forzando le molecole a riorientarsi, prima di continuare a muoversi attraverso il gel. Questo riorientamento richiede tanto più tempo, quanto più grande è la molecola in questione. È

così possibile distinguere (in base al pM) i vari cromosomi, che in un'elettroforesi normale non si separerebbero.

In secondo luogo, è necessario rompere un cromosoma in frammenti più piccoli, per analizzarlo. A questo proposito, si può far uso dell'**ibridazione da irradiazione**: si frammenta il cromosoma di partenza tramite l'irradiazione, per poi ibridare con cromosomi di topo i vari frammenti, per analizzarne la composizione.

Alternativamente, si usano enzimi di restrizione che tagliano raramente, perché il loro sito è poco diffuso nel genoma, originando così grossi pezzi di cromosoma.

Dopo aver compiuto la frammentazione del cromosoma, si procede con la costruzione di una banca genomica. Il vettore che si utilizza a questo scopo è lo YAC (perché è l'unico con una capacità sufficiente a incorporare grossi frammenti).

Per distinguere i diversi frammenti di cromosoma, li si identifica localizzando sequenze particolari. A questo proposito, non si può utilizzare una mappa di restrizione, perché i frammenti sono troppo lunghi e verrebbero originati troppi siti di restrizione.

Si può invece utilizzare l'ibridazione con geni già conosciuti di quel cromosoma; oppure si può riconoscere siti detti *STS* (Sequence-Tagged Sites), che sono identificabili con la PCR (si usano primer complementari a sequenze conosciute); oppure ancora si può far uso della cosiddetta *ibridazione in situ* (che, a differenza di quella sopracitata, può essere osservata al microscopio direttamente sul cromosoma, vedendo anche la localizzazione del tratto che ibrida), utilizzando la marcatura della sonda in fluorescenza, tecnica che prende il nome di *FISH* (Fluorescence In Situ Hybridization).

Per determinare la mappa fisica di un (pezzo di) cromosoma, ci si serve di genoteche.

Si taglia un (pezzo di) cromosoma con enzimi di restrizione (digestione parziale per creare frammenti di determinata lunghezza). Si procede poi alla ricombinazione dei vari frammenti con altrettanti vettori, costituendo così la banca genomica.

È necessario poi riordinare i vari frammenti presenti nei vettori, costruendo un **continuo di cloni (contig)**, che contiene ordinati uno dopo l'altro tutti i frammenti. L'ordinamento avviene grazie al riconoscimento di parti comuni a due frammenti, che perciò sono adiacenti (frammenti con queste caratteristiche esistono solo se i tagli operati dall'enzima di restrizione sono stati casuali).

Per ordinare i frammenti si può far uso di due metodi: il **chromosome walking** ("camminata sul cromosoma") (che fa uso dell'ibridazione e delle mappe di restrizione), oppure un'altra tecnica che si serve della PCR inversa.

**Chromosome walking.** Si prende un frammento clonato per cercare di identificare quelli adiacenti, per poi risalire al continuo di cloni. Si isola la parte terminale di uno dei frammenti e se ne fa una sonda, che può così ibridare con una regione ad essa complementare, in un altro (o in altri) frammenti all'interno della banca.

Poi, si procede allo stesso modo coi frammenti successivi. Di ogni frammento trovato si fa la mappa di restrizione, per vedere se è più lungo o più corto di altri che contengono la stessa regione comune. Tra i vari frammenti con una regione terminale comune alla sonda, si sceglie sempre quello più lungo.

N.B. = Dal primo frammento analizzato si procede sia un senso che nell'altro, per costruire il continuo di cloni corrispondente al filamento originale.

Per studiare poi ogni frammento, bisogna risalire alla colonia batterica da cui esso proviene e poi isolare all'interno di essa uno o più cloni contenenti quel frammento.

**Tecnica con la PCR inversa.** Permette di identificare la parte finale di un frammento da sequenziare, nella regione adiacente al punto di inserzione di esso nel vettore.

Si usa un'endonucleasi di restrizione che taglia in tutto il vettore ricombinante (non solo nell'inserito). Tra i vari sub-frammenti originati da questa operazione, se ne creerà uno comprendente il tratto a cavallo tra vettore e inserto.

Si fa circularizzare il frammento misto vettore-inserito (tramite le sticky ends). Si sottopone il frammento a PCR inversa. I primer sono collocati alle estremità della regione nota (regione del vettore). In questo tipo di PCR, i primer, al contrario che nella PCR tradizionale, sono divergenti, altrimenti verrebbe amplificata la regione già nota. Siccome il frammento è circularizzato e i primer sono divergenti, viene amplificata la parte non nota (regione dell'inserito).

Si sottopongono poi in varie tornate tutti i frammenti a saggi di PCR normale (usando come primer le estremità della regione comune al frammento precedente), per vedere di volta in volta quale (o quali) frammento viene amplificato, costruendo così il contiguo di cloni.

Ogni frammento (clone) conterrà un certo numero di STS. Si possono identificare quali cloni contengono STS comuni, in modo da costruire il contiguo di cloni, come si faceva col chromosome walking.

N.B. = I frammenti di questo esperimento devono essere corti (la PCR funziona solo con frammenti corti).

A questo proposito, prima di tutto da ogni cromosoma umano si fa una banca con YAC. Di ogni frammento costituente la banca YAC si fa una sottobanca, ad es. col fago  $\lambda$ . Però servono frammenti ancora più piccoli: perciò si sottopongono i frammenti della sottobanca a digestione parziale, così da creare corti frammenti che si sovrappongono parzialmente. A questo punto si sequenziano questi cloni per poter poi ordinarli nel contiguo di cloni specifico di quella sottobanca. Terminata una sottobanca, si procede a compiere lo stesso passaggio con un'altra sottobanca, fino a esaurire la banca. Terminato questo procedimento, si ripetono tutti i suddetti passaggi sugli altri cromosomi.

Questo approccio di tipo sequenziale è stato applicato dal consorzio pubblico, nel Progetto Genoma. Il consorzio privato, invece, ha frammentato direttamente tutto il cromosoma in un unico passaggio. Questo metodo, tuttavia, si è rivelato meno efficiente dell'altro, in quanto i "buchi" sono maggiori.

In uno stadio successivo, bisogna scoprire dove sono localizzati i geni, all'interno dei cromosomi già sequenziati.

Per fare ciò, si possono identificare:

- gli esoni
- le giunzioni tra esoni ed introni
- le isole CpG (citosina-fosfato-guanina) (si trovano al 5' dei geni molto espressi e housekeeping (geni di base di tutte le cellule)  $\Rightarrow$  indicano che siamo nei pressi di un gene)
- omologie o con proteine codificate da geni simili (che codificano quindi proteine già note), o con proteine simili in altri organismi.

Quando si è poi localizzato il gene, si cerca il modulo di lettura codificante (Open Reading Frame (ORF)) tra i 3 possibili, su entrambi i filamenti (perché non sappiamo qual è il filamento codificante) (tot. = 6 moduli di lettura possibili). I due moduli di lettura non corretti fanno comparire troppo presto un codone di stop.

Alternativamente, è possibile cercare la posizione di un gene in un cromosoma anche senza averlo sequenziato, grazie alle EST.

La prossima frontiera nello studio del genoma umano è l'identificazione delle diversità tra gli individui di una popolazione. A questo proposito, si utilizza il fenomeno dell'SNP (Single Nucleotide Polymorphism), piuttosto del DNA microsatellite, perché l'SNP è più facilmente automatizzabile.

### **Riepilogo sul mappaggio fisico del genoma**

Avviene tramite marcatori ( $\Rightarrow$  geni), oppure sequenze intergeniche come il DNA microsatellite, oppure i siti di restrizione, oppure regioni di DNA conosciute (STS) (di cui si è fatto un saggio di identificazione con la PCR).

Un STS può essere DNA microsatellite, oppure un pezzo di gene espresso (EST: Expressed Sequence Tagged, ottenuto grazie al sequenziamento casuale di cDNA di una banca (il primer è l'oligo(dT) ( $\Rightarrow$  ci sono banche di EST).

N.B. = Per verificare polimorfismi non si usa il mappaggio fisico di tutto il DNA del genoma, ma l'analisi della sola regione di interesse col Southern Blot, ad es.

Per il sequenziamento del genoma, da ogni cromosoma bisogna fare una banca con YAC (o BAC, o PAC), eventualmente frazionando prima il cromosoma, se è grosso. Bisogna poi ordinare gli YAC sul cromosoma, definendo il frammento di cromosoma che ogni determinato YAC contiene, in relazione a quelli contenuti negli altri YAC (ad es., tramite chromosome walking).

Si crea un contiguo di cloni, che rappresenta base per base tutto il genoma contenuto in quel cromosoma.