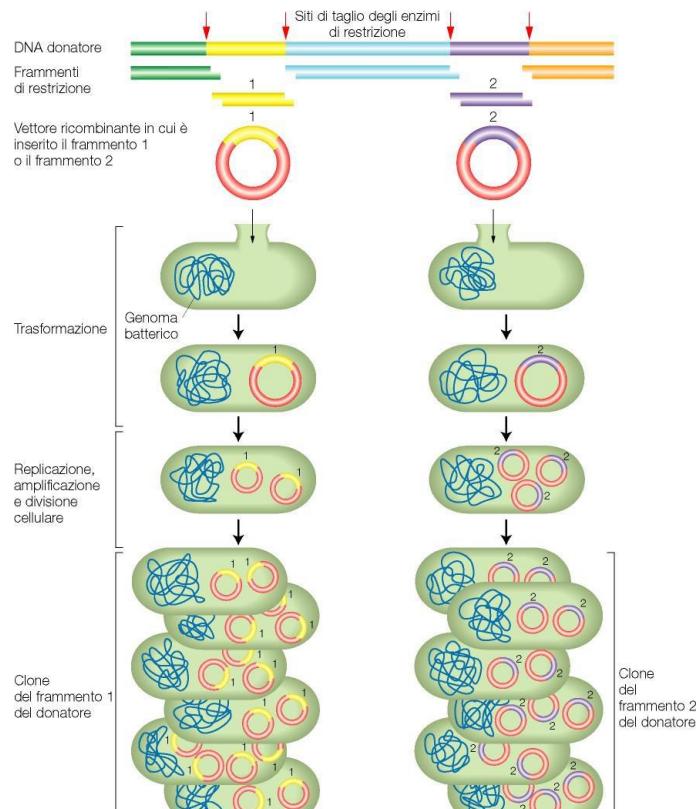
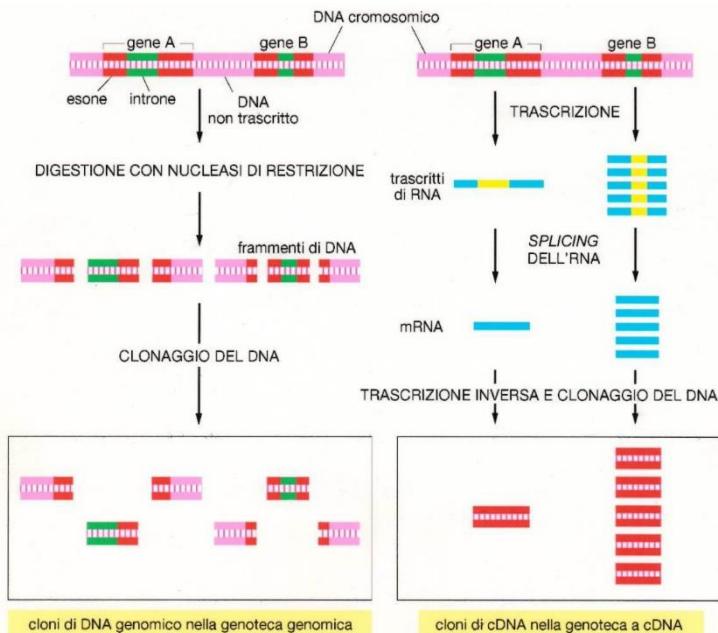


LIBRERIE GENOMICHE E A cDNA

Introduzione

Una *genoteca* rappresenta l'intero genoma (libreria genetica) o l'intero trascrittoma (libreria a cDNA). Si ottiene effettuando la tecnica del clonaggio. Il DNA cromosomico di interesse è clonato all'interno di un ospite, generando una collezione di cloni ricombinanti che rechino frammenti di DNA eterologo clonato.

Le librerie a cDNA presentano solo il DNA genomico espresso, mentre le librerie genetiche presentano tutto il DNA del genoma (promotori, introni, sequenze intergeniche, sequenze regolatorie, geni espressi e non espressi).



1 inserto = 1 clone = 1 colonia

Per selezionare il gene di interesse è necessario effettuare uno *screening*: i singoli cloni di una libreria non sono indicizzati. Con screening si intende una tecnica che permetta di identificare la sequenza che corrisponde al gene di interesse tra milioni di altre sequenze o cloni.

Lo screening può essere basato su:

- sequenza nucleotidica del gene;
- sequenza aa del gene;
- funzione (functional screening).

Librerie genomiche

Per generare librerie genomiche, il DNA di partenza (cromosomale) deve essere reso clonabile n opportuni vettori. I vettori ricombinanti saranno inseriti nell'ospite (solitamente *E.coli*) per generare una popolazione di cloni che rappresenti tutto il genoma di partenza.

Impiego delle librerie genomiche:

- isolamento e caratterizzazione di un tratto di DNA di interesse;
- mappatura e sequenziamento dei genomi.

In generale la scelta del vettore di clonaggio dipende dalla grandezza del DNA da clonare:

- per inserti sino a 10 kb si utilizzano i **plasmidi**
- per inserti sino a 20 kb si utilizzano i **fagi**
- per inserti sino a 45 kb si utilizzano i **cosmidi**
- per inserti sino a 100kb si utilizza il vettore BAC (Bacterial Artificial Cromosome) e oltre i 1000 kb si usa YAC (Yeast Artificial Cromosome).

In particolare, vengono utilizzati i vettori plasmidici o fagici. I vettori fagici hanno le seguenti caratteristiche che li contraddistinguono:

- nel momento in cui infettano *E.coli* generano placche di lisi
- possono essere introdotti nell'ospite come DNA nudo (= plasmidi) o come fago ricostruito
- sono più efficienti dei plasmidi per il clonaggio di DNA di grandi dimensioni e per lo screening su grandi numeri di uno specifico inserto.

Una caratteristica fondamentale della banca è la *rappresentatività*: una libreria è rappresentativa quando contiene almeno una copia di tutte le possibili sequenze. Spesso le banche sono anche *ridondanti* ed è possibile trovare molte copie di alcune sequenze, perciò lo screening può essere complicato.

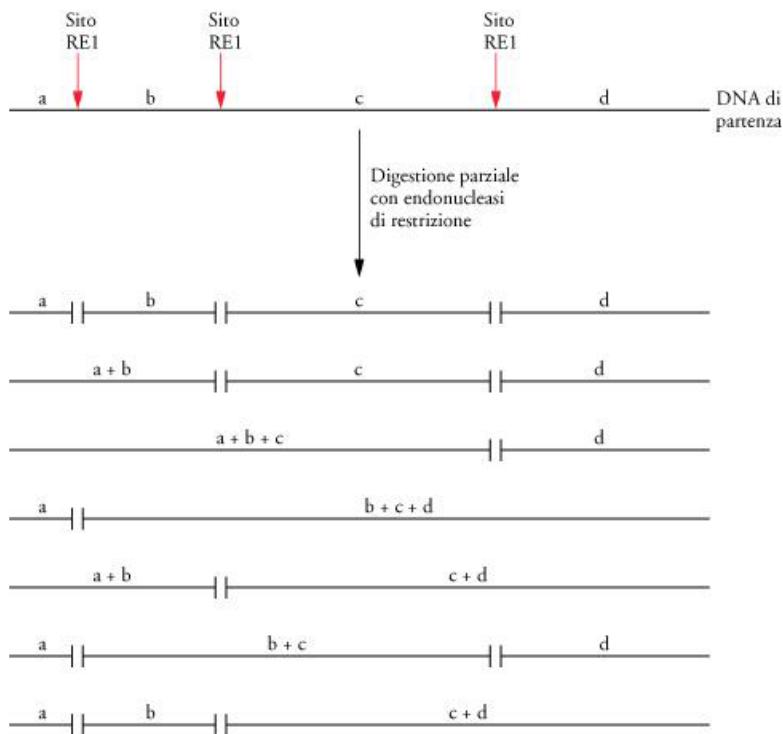
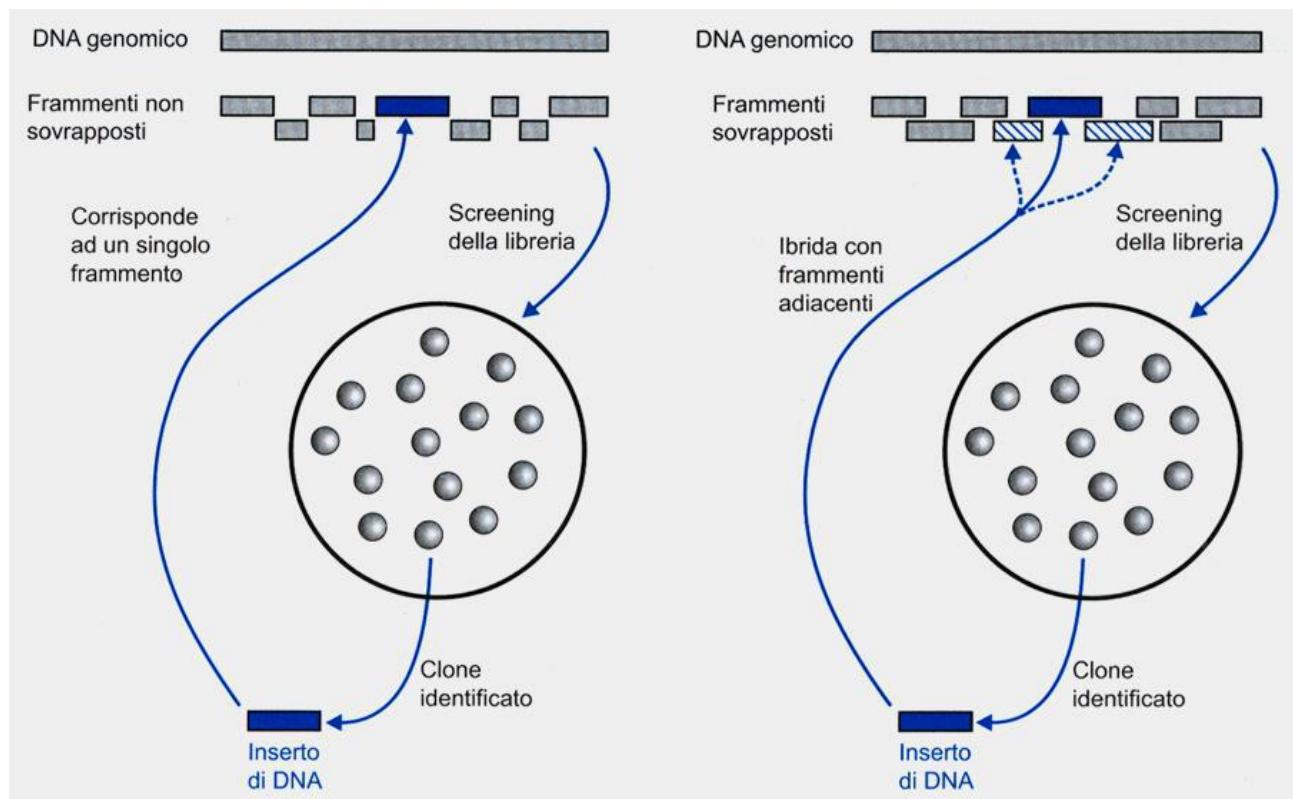
Per frammentare il DNA genomico si utilizzano enzimi di restrizione di tipo 2 che permettano la generazione di estremità coesive per il clonaggio e inoltre non deve esserci perdita di informazione.

Effettuando digestione completa con enzimi di restrizione, i siti di taglio non sono disposti in modo uniforme nel genoma, perciò, si potrebbero generare frammenti lunghi e corti: i frammenti eccessivamente grandi potrebbero dare problemi durante la ligazione e i frammenti piccoli potrebbero andare persi.

Inoltre, effettuando una digestione completa si avrebbe il problema di non riuscire a ricostruire il genoma.

Si effettuano quindi delle **digestioni parziali** in modo tale da ottenere **frammenti sovrapponibili e non sovrapposti**. Nel momento in cui l'enzima è in quantità limitanti rispetto al DNA genomico (oppure in condizioni sfavorevoli), verranno riconosciuti e tagliati i siti con un intorno ottimale e via via meno efficientemente i siti che avranno un contesto non ottimale. Le digestioni parziali del DNA sono riproducibili in funzione della quantità iniziale del DNA di partenza e in funzione della quantità di enzima presente.

Il fatto di avere frammenti sovrapposti fa sì che si possa ricostruire un gene intero o isolare le zone limitrofe.



DIGESTIONE PARZIALE per produrre una collezione di frammenti sovrapposti

Dopo specifici intervalli di tempo si preleva un'aliquota e si verifica su gel di agarosio se il processo ha premesso di ottenere frammenti di DNA di dimensioni opportune. Le bande si potranno visualizzare con EtBr.

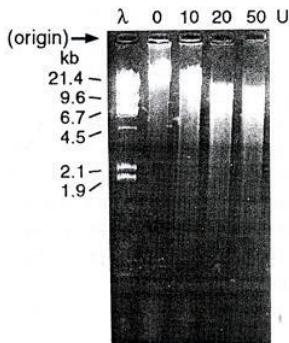


Figure 5.3.2 This 0.9% agarose minigel contains samples of genomic digests performed with increasing concentrations (left to right) of restriction enzyme. As the digest approaches completion, high-molecular-weight fragments become less prominent and faint satellite bands become visible. 10 µg genomic DNA was incubated with 10, 20, and 50 U restriction enzyme for 2 hr, and 40 µl (about 1 µg) DNA was loaded on the gel.

All'aumentare della concentrazione di enzima in uno stesso intervallo di tempo di digestione, si visualizza un DNA genomico sempre più digerito.

Perciò i fattori che influiscono sulla digestione sono:

- *tempi di digestione*
- *quantità di enzima*
- *temperatura di incubazione*

Sapendo che gli enzimi di restrizione tagliano a livello di specifiche sequenze, è possibile calcolare la possibilità che una sequenza sia presente all'interno del genoma:

$$4 \text{ BASI hanno PROBABILITÀ} = (1/4)^4 = 1/256$$

Ciò significa che si ipotizza ci possa essere la sequenza di 4 basi desiderata ogni 256 basi e perciò è molto probabile che avvenga il taglio. Utilizzando enzimi che riconoscono sequenze di 4 pb, in condizioni di digestione limitata si avrà alta probabilità di digestione ma ottenendo frammenti di dimensioni opportune per il clonaggio.

In fine è necessario accertarsi che le estremità siano compatibili, per questo spesso si utilizza l'enzima di restrizione di tipo II Sau3A che riconosce la sequenza a 4 pb /GATC e lascia estremità cohesive 5' protruding. Queste estremità sono inoltre compatibili con i frammenti creati da BamH1 che riconosce la sequenza a 6 pb G/GATCC lasciando estremità 5' protruding.

A questo punto avviene la ligazione con la **ligasi di T4 in condizioni ottimali** e la trasformazione di E.coli con **elettroporazione** per garantire la rappresentatività della banca molto efficientemente.

In caso di plasmidi: Nel momento in cui vengono ligate le estremità Sau3A e BamH1 si riforma il sito riconosciuto da Sau3A, ma non è detto che si riformi quello di BamH1. Il plasmide viene inserito in un ceppo di E.coli reso competente tramite elettroporazione e poi viene piastrato il tutto su di un terreno selettivo, come per esempio in un terreno contenente Xgal.

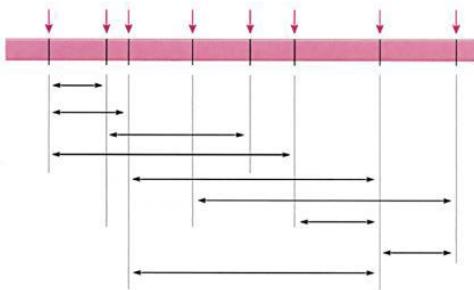
Il piastramento deve avvenire su un numero notevole di piastre (20-30) ed è necessario *valutare la qualità della libreria genomica, ovvero la % di cloni che contiene l'inserto*, che saranno contabili perché le colonie con l'inserto appaiono bianche su Xgal.

Per valutare la dimensione degli inserti occorre costruire una mappa di restrizione: verranno prelevate un certo numero di colonie rappresentative e saranno inserite in un liquido. Verrà estratto il DNA e da qui si farà la mappa di restrizione.

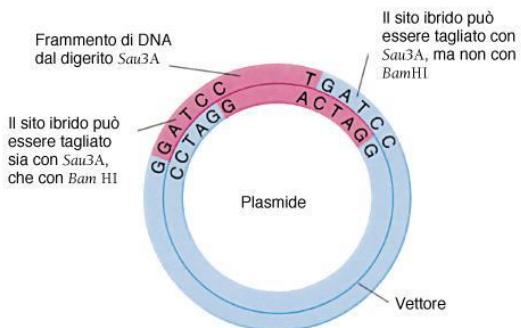
Se la qualità della libreria è buona, le colonie sono trasferite su terreno liquido e congelate a -80° con glicerolo.

Uso di un enzima di restrizione per produrre frammenti di DNA di dimensioni adatte alla costruzione di una banca genomica.

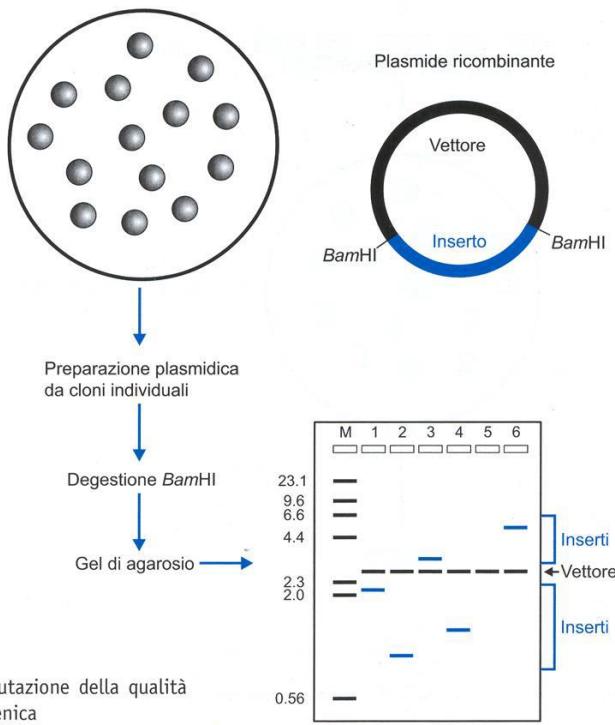
- a) Digestione parziale di DNA con un enzima di restrizione, per es. *Sau3A*, che genera una serie di frammenti sovrapposti, ciascuno con estremità appiccicose identiche 5' GATC



- b) I frammenti risultanti possono essere inseriti nel sito *BamHI* del vettore plasmidico



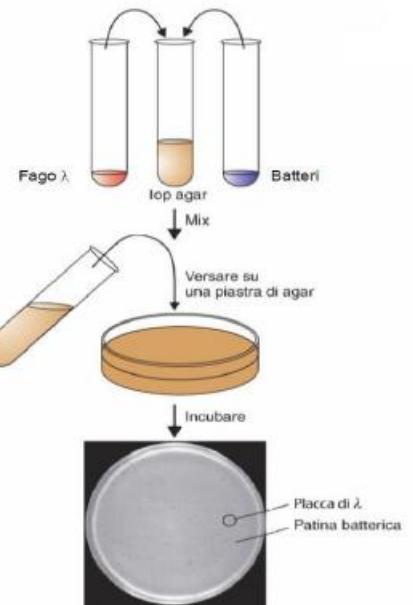
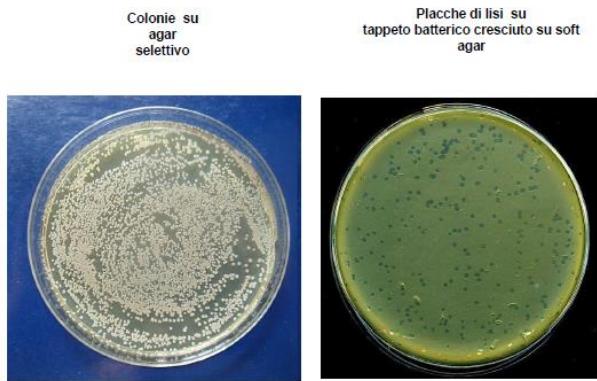
Libreria di frammenti *BamHI* in un vettore plasmidico



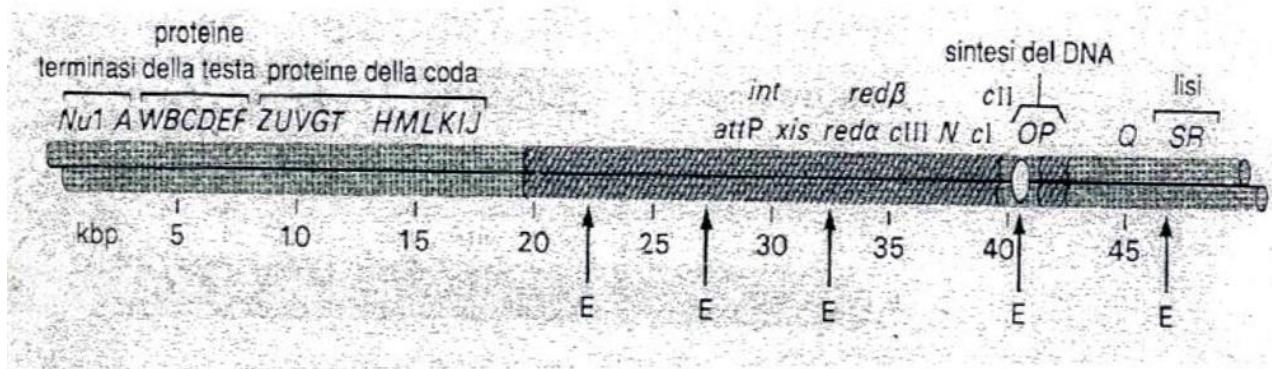
% di cloni effettivamente contenente un inserto e studio delle dimensioni degli stessi

In caso di fagi: I batteriofagi furono descritti per la prima volta intorno alla seconda decade del 900 come placche di lisi osservabili su popolazioni batteriche cresciute a confluenza su terreni solidi. Ciascuna placca rappresenta una popolazione clonale. A partire dalla scoperta iniziale sono stati isolati e caratterizzati

numerosi batteriofagi, ciascuno con differenze genetiche e strutturali. Il batteriofago λ il fago di maggior interesse in Biologia molecolare, può utilizzare due stili di vita all'interno del batterio: il ciclo litico e il ciclo lisogeno. Si utilizza preferenzialmente il fago λ , un batteriofago temperato in grado di infettare *E. coli* ed effettuare ciclo litico e/o ciclo lisogeno, ricombinante. In laboratorio viene fatto crescere per inclusione su una piastra di cellule batteriche e terreno di coltura soft agar. Dopo incubazione overnight è possibile visualizzare le placche di lisi.



Il genoma del fago λ è DNA ds con PM 45,8 Kb e alle estremità presenta 12 basi a ss perfettamente complementari che vengono definite estremità cos. Le estremità cos si appaiano appena il DNA fagico infetta l'ospite. Il DNA circolarizzato all'interno dell'ospite si replica a *rolling circle* formando un *concatamero* di genomi incatenati testa-coda. Il fago sfrutta l'apparato trascrizionale e biosintetico dell'ospite per produrre anche la testa e la coda del fago. Quindi il fago si assembla e in ogni testa si posiziona un unico genoma: il concatamero permea nella testa del fago e l'enzima *terminasi* taglia le estremità cos in modo tale che in una testa sia presente un unico genoma. Si originerà così la particella virale finale, si origineranno poi n^particelle virali che liseranno la cellula.



La testa del fago ha capienza definita per un massimo di 51 kb e un minimo di 38 kb. Pertanto, il genoma ricombinante deve avere dimensioni comprese tra 51 kb e 38 kb. Poiché 30 kb sono indispensabili, l'inserto dovrà essere di grandezza compresa tra 8 – 21 kb.

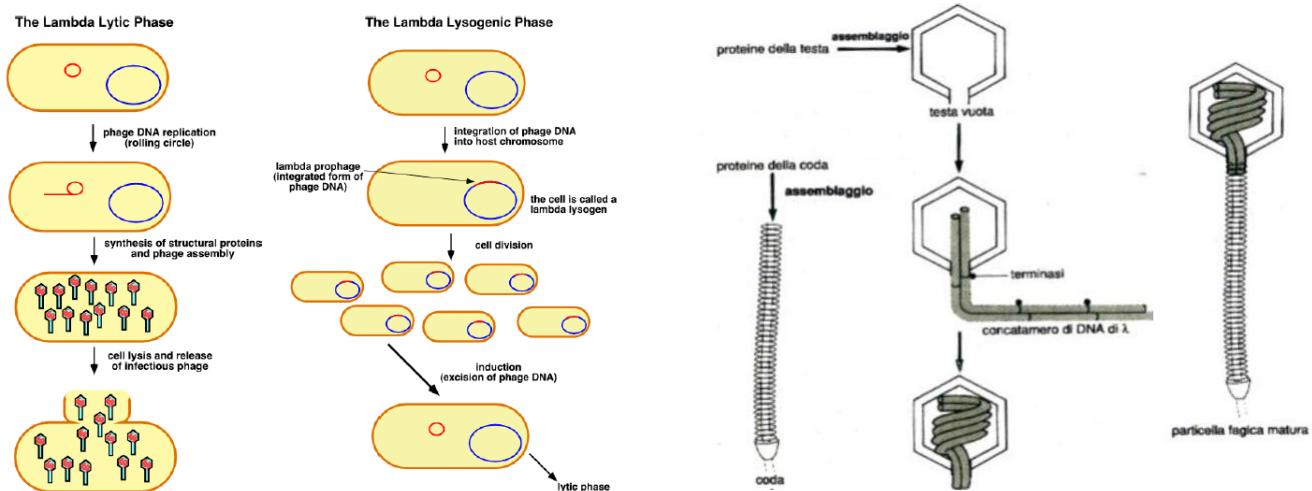
Un vantaggio nell'utilizzo di fagi è la possibilità di effettuare packaging in vitro, creando le particelle virali in vitro.

Partendo dal genoma wild type di λ sono stati costruiti due diversi vettori:

- Vettori di inserzione;
- Vettori di sostituzione;

I vettori di inserzione (λ gt10) danno placche trasparenti se avviene ricombinazione (inattivazione inserzionale), se invece il vettore non è ligato all'inserto si hanno placche turbide. Questo vettore presenta un sito unico di restrizione Eco nella sequenza C', che codifica per un repressore, per l'inserzione. Quando C' è funzionante, il fago va incontro a ciclo lisogeno (stato di profago inattivo e non dannoso per l'ospite batterico), ma una volta inattivato il fago può andare incontro solamente a ciclo litico generando placche chiare.

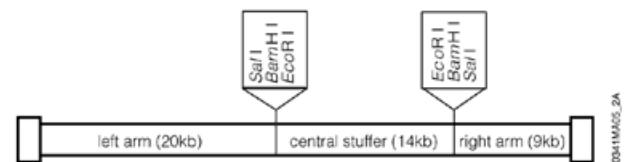
Il vettore λ gt10 è di circa 43,3kb e l'inserto può avere dimensioni massime di 7,6 kb e deve presentare estremità compatibili Eco per essere inserito nella sequenza C'.



Questi vettori sono più facili da utilizzare e possono accettare inserti di dimensioni da 8 fino a 10-12 Kb. Sono generalmente utilizzati per costruire librerie di cDNA.

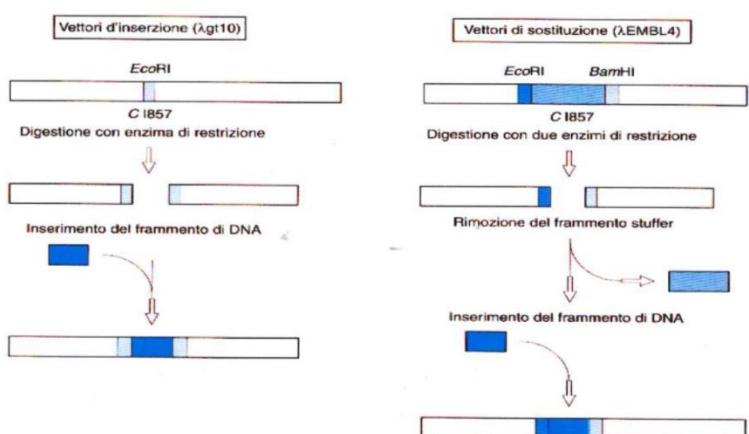
Nei vettori di sostituzione (λ EMBL4) sono stati eliminati tutti i geni non essenziali e presentano un frammento stuffer di 14 kb codificante per l' α -peptide della β -galattosidasi ($lacZ'$) al posto dei geni che codificano per la lisogenia che permette uno screening bianco-blu in E.coli che presentano il frammento ω della galattosidasi.

Il frammento stuffer presenta alle estremità dei siti unici di restrizione, ad esempio BamH1. Questa tipologia di vettori permette il clonaggio di frammenti più grandi perché permette l'excisione dello stuffer che verrà sostituito dai frammenti inseriti di 10 – 20Kpb. La porzione di DNA rimanente è necessaria per la stabilità virale e la produzione di placche di lis.



Il DNA viene corso su gel e vengono quindi purificati i due bracci digeriti BamH1 ciascuno contenente le estremità cos. Le due braccia vengono quindi recuperate e ligate con il DNA genomico frammentato con Sau3A e defosforilato per evitare multimeri. Le due braccia sono in totale di 29Kpb (non sufficiente per il packaging), dunque saranno clonabili solo 22Kpb. Nelle placche prodotte saranno presenti solo fagi ricombinanti.

I fagi possono quindi essere conservati a 4°C per la libreria. Vengono spesso utilizzati per generare banche genomiche.



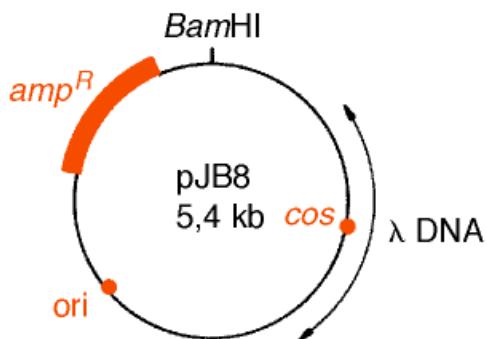
I cosmidi sono plasmidi che presentano delle estremità cos. Tramite i cosmidi possono essere clonati **fino a 45 kb di DNA esogeno**. Così come i plasmidi presentano:

- ori di replicazione
- marcatore di selezione (sfruttando ad esempio la resistenza all'ampicillina)

E come i vettori fagici presentano:

- sequenze cos
- impacchettamento in particelle fagiche (anche in vitro)

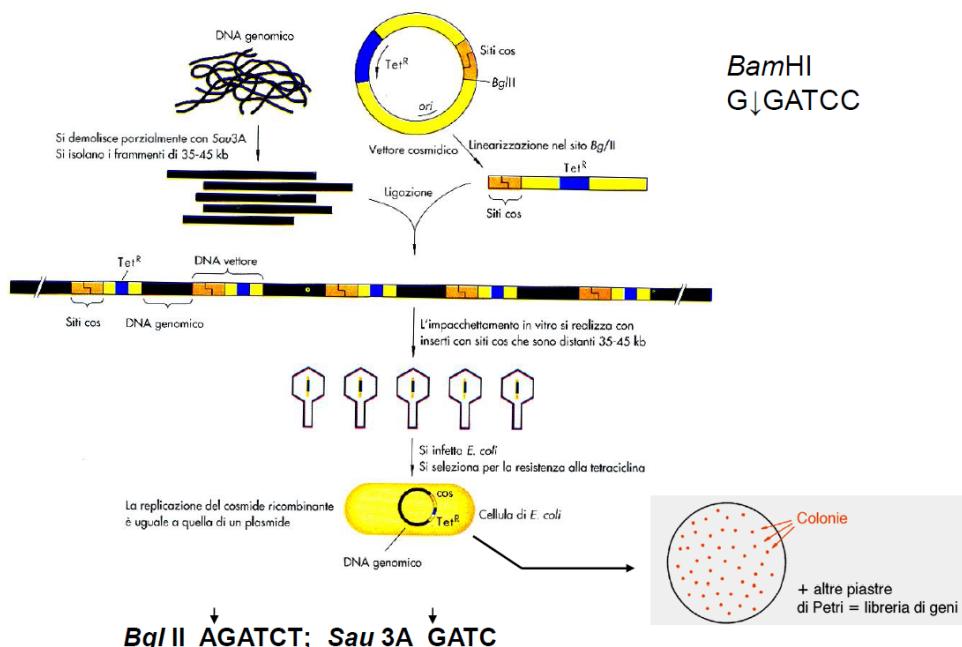
A seguito dell'infezione di *E.coli*, il cosmide viene amplificato come un plasmide e non origina placche di lisi.



BamH1 è il sito di clonaggio. Si potrà poi fare una mappa di restrizione per definire la dimensione degli inserti, vengono inseriti in una provetta con una spatola e messi in glicerolo a -80°, costituiranno quindi la **library primaria** (dovrebbe contenere una copia rappresentativa di ogni regione del genoma).

Conservazione:

- Plasmidi e cosmidi sono conservati a -80°;
- Fagi sono conservati a 4°;



Preparazione di una libreria genomica in un vettore cosmidico

Screening

Per screening si intende il processo di identificazione del clone che contiene il gene/sequenza di interesse tra milioni di altri cloni presenti in una library.

Quando la libreria deve essere sottoposta a screening se ne preleva un'aliquota che deve essere piastrata/messa a infettare E.coli. La restante parte della library viene lasciata congelata.

Prima del piastramento però bisogna determinare il **titolo** della libreria. Il titolo rappresenta le dimensioni della library e la complessità dello screening.

Il titolo è il *numero di colonie o di placche ottenuti per ml* della libreria. Ciò fa sì che si possa determinare il numero di ml della libreria che servono per ogni piastra durante lo screening di modo da ottenere un certo numero di cloni indipendenti. Si vuole infatti un certo numero di cloni/placche sufficiente che permetta di isolare la sequenza di interesse.

Il titolo si determina tramite una serie di diluizioni successive (1:10) da piastrare/da infettare le cellule batteriche (nel caso di fagi).

Calcolo del numero di cloni indipendenti da saggiare:

$$N = \frac{\ln(1-P)}{\ln(1-I/G)}$$

*N = numero di cloni indipendenti da saggiare per isolare una sequenza con probabilità P 99%;
I = dimensione media dei frammenti di DNA inseriti nel vettore in pb;
G = dimensione del genoma in pb;*

Tanto più un genoma avrà dimensioni elevate tanto più sarà maggiore il numero di cloni indipendenti da saggiare.

Se P=99%

genomaumano=3x10⁹bp

I=2x10⁴bp

$$N = \ln(1-0.99) / \ln(1-2 \times 10^4 / 3 \times 10^9) = 690000$$

Box 6.1 Stima delle dimensioni delle librerie genomiche in base al tipo di vettore utilizzato

Organismo	Dimensione del genoma	Tipo di vettore	Dimensione dell'inserto	P	Dimensione della libreria
Batterio	4×10^6 basi	Plasmide	4 kb	0.99	4.6×10^3
		Lambda di sostituzione	18 kb	0.99	1.0×10^3
		Cosmide	40 kb	0.99	458
		BAC	300 kb	0.99	59
Mammifero	3×10^9 basi	Plasmide	4 kb	0.99	3.5×10^6
		Lambda di sostituzione	18 kb	0.99	7.7×10^5
		Cosmide	40 kb	0.99	3.5×10^5
		BAC	300kb	0.99	4.6×10^4

I valori relativi alle dimensioni del genoma batterico e di mammifero sono solo esempi utilizzati ai fini del calcolo. Le dimensioni effettive del genoma variano notevolmente da un organismo all'altro. Sono variabili anche le dimensioni dell'inserto per gli specifici vettori.

La libreria primaria presenta in genere basso titolo, sarà quindi necessaria l'**amplificazione**. La collezione di fagi o delle colonie batteriche viene piastrata nuovamente (o fagi infettano nuovamente), dopodiché vengono raccolte di nuovo tutte le colonie a formare una libreria amplificata.

La library delle colonie verrà messa a -80°, mentre quella dei fagi a 4°. Verrà quindi determinato il titolo per poterla saggiare, questa avrà titolo maggiore e potrà essere usata quindi più volte.

La sequenza di interesse è individuata piastrando il numero di colonie calcolato per ottenere il numero di cloni indipendenti desiderato e si effettua uno screening che si basi su:

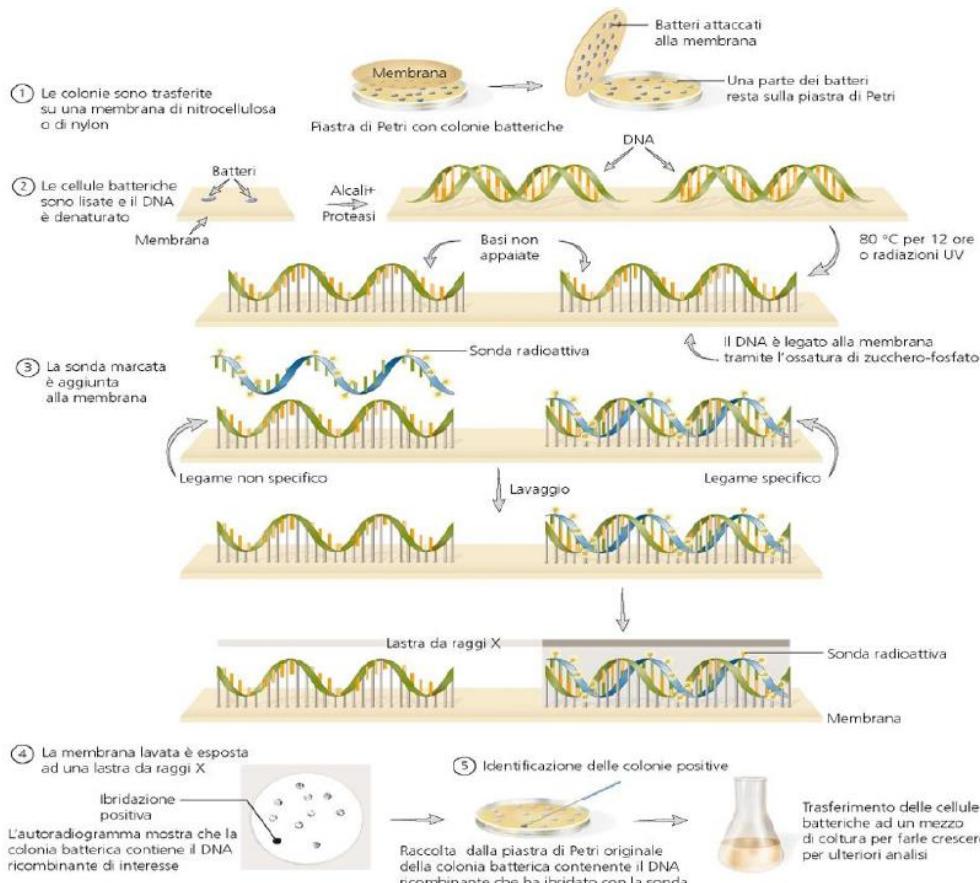
- sequenza nucleotidica del gene;
- sequenza aa del gene;
- funzione (functional screening).

Uno screening basato sulla sequenza nucleotidica si effettua per ibridazione con una sonda. La sonda deve essere un acido nucleico con una sequenza anche parzialmente complementare alla sequenza da riconoscere, in modo che si associno in modo specifico.

Si utilizza una sonda radioattiva *random priming* (elevata attività specifica).

Utilizzando plasmidi e cosmid si procede effettuando *cloning hybridization*, poggiando sulla piastra un filtro di nitrocellulosa oppure di nylon. Premendo leggermente, parte dei cloni si trasferiscono sul filtro. Il filtro viene recuperato con una pinzetta e lo si depone su filtri imbevuti di soda, in grado di lisare la parete batterica e denaturare il DNA. Il DNA si ritroverà libero sul filtro e verrà cross-linkato con raggi UV. Si prosegue quindi con l'ibridazione e l'incubazione a 37°.

COLONY HYBRIDIZATION

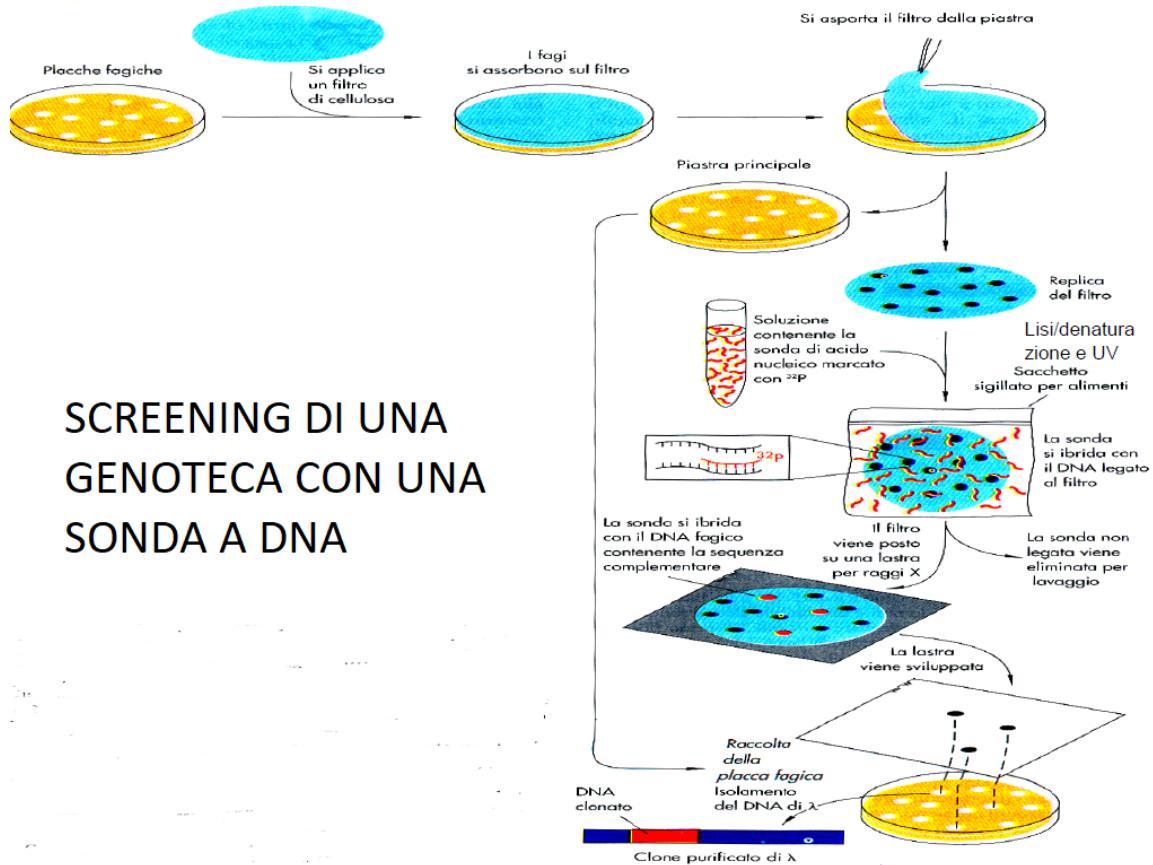


In caso di fagi il processo è lo stesso.

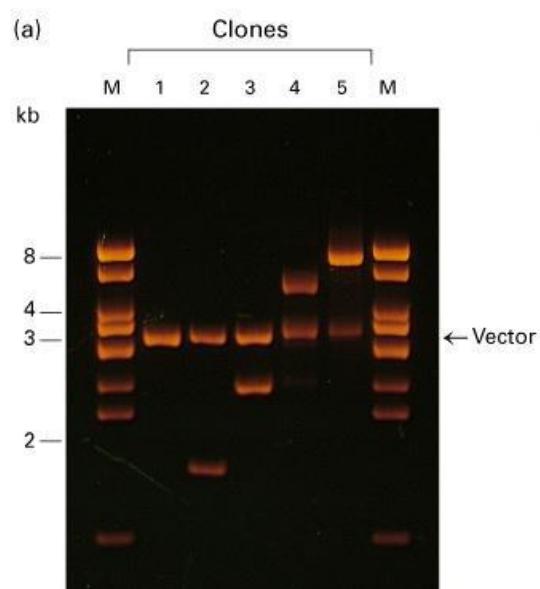
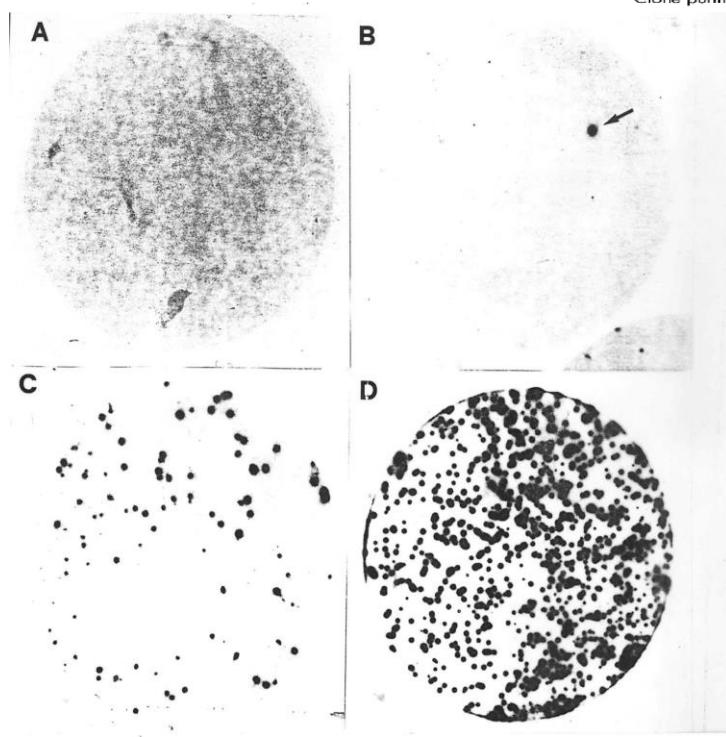
Dopo l'ibridazione viene confrontato il segnale con la piastra per identificare la coltura da prelevare. Si fanno una serie di subculture successive in modo da ottenere colonie isolate e l'*arricchimento del segnale*. Nel caso

di fagi, si individua la placca in cui sono presenti fagi che saranno prelevati e messi in nuovo terreno per reinfettare E.coli.

Dopo aver ottenuto la coltura contenente il frammento di interesse si potrà quindi fare un'*analisi di restrizione e sequenziamento*.



SCREENING DI UNA GENOTECA CON UNA SONDA A DNA

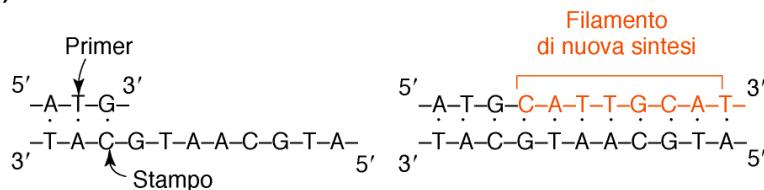


Librerie a cDNA

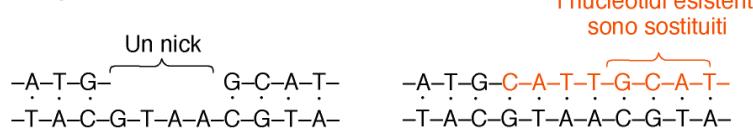
Rappresentano solo i *geni che sono effettivamente espressi* in una particolare cellula o campione di tessuto in un particolare momento. Si basa quindi sull'intero trascrittoma: RNA/mRNA.

Si sfrutta l'attività dell'enzima *trascrittasi inversa* per generare una copia di DNA complementare all'mRNA definito *cDNA*. La trascrittasi inversa è una DNA polimerasi con attività retrotrascrittasicia e necessita di ioni Mg^{2+} , dNTP e un primer. La polimerizzazione prosegue in direzione $5' \rightarrow 3'$ a partire da un primer rappresentato dal 3'OH.

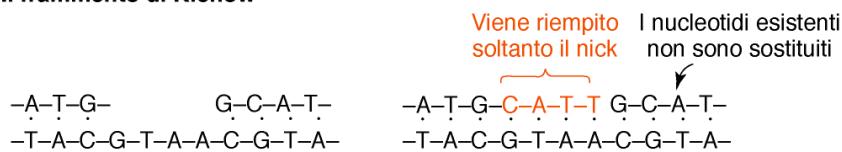
(a) La reazione di base



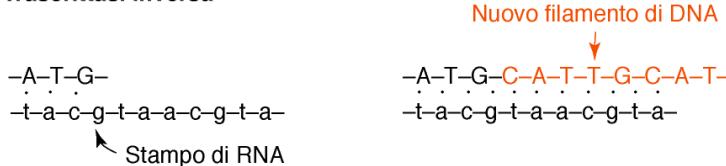
(b) DNA polimerasi I



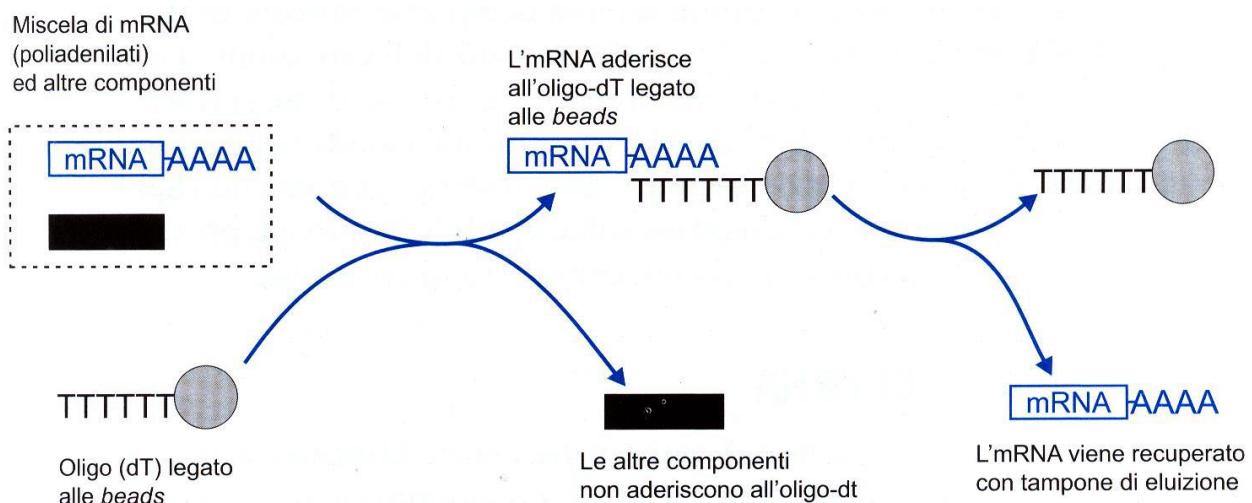
(c) Il frammento di Klenow



(d) Trascrittasi inversa



L'**mRNA** è **purificato** mediante oligo-dT che si appaia con la coda poly-A degli mRNA eucariotici. La purificazione avviene in modo analogo al Northern Blot.



Sulla base del primer usato si possono avere **due metodiche di retrotrascrizione**:

- *Random primer*, utile sia per l'RNA totale che per l'mRNA;
- *Primer Oligo di T*, utile solo per gli mRNA;

I random primer sono una miscela di primer lunghi 6pb che contengono tutte le possibili combinazioni delle 4 basi (4^6 esameric).

I primer si annilano in modo specifico con le sequenze complementari che si trovano su mRNA o RNA. La trascrittasi

prosegue la polimerizzazione retrotrascrizionale a partire da un 3'OH di un primer e terminerà nel momento in cui il primer davanti si arresta (l'enzima non presenta attività eso 5' → 3'). Al termine del processo si aggiunge la ligasi per chiudere i nick.

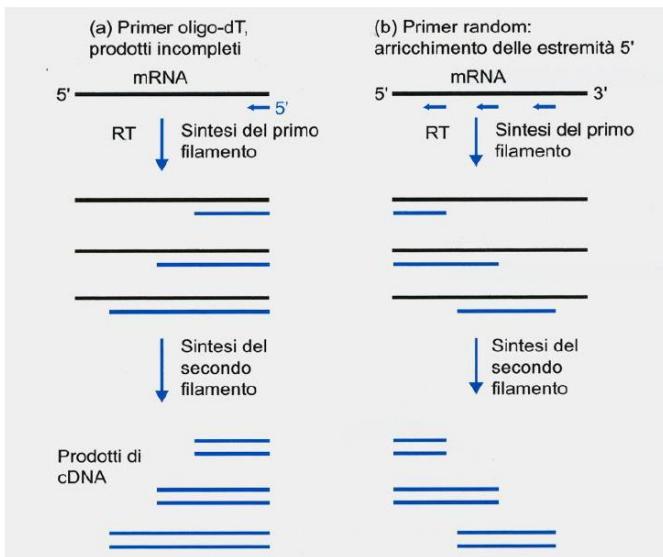
I primer Oligo-dT sono primer costituiti da 18T che si annilano con le code di poly-A, dopo di che si avrà polimerizzazione della trascrittasi inversa. Le code polyA sono lunghe anche 200 pb, perciò si utilizzano primer oligodT che presentano *2 basi aggiunte* AT, CG... ciò fa sì che il primer si annili al termine del messaggero ma prima che inizi la coda. In questo modo non verrà convertita in cDNA anche la coda di polyA, ma solo il messaggero vero e proprio.

Vantaggi/svantaggi:

- Nel caso di messaggeri molto lunghi, con gli oligo dT la trascrittasi inversa può anche non arrivare in fondo e quindi si ottengono dei cDNA incompleti (in base alla processività della trascrittasi), è quindi preferibile utilizzare random primer (aumenta la rappresentatività in 5').
- Però con i random primer può accadere che un primer non si anneali perfettamente in 3', dunque si può perdere rappresentatività in 3'. Esistono però delle metodiche che consentono di recuperare il 3', come la metodica "Race".

Sintesi di cDNA:

utilizzo di random primer per aumentare la rappresentatività delle estremità 5' dell'RNA



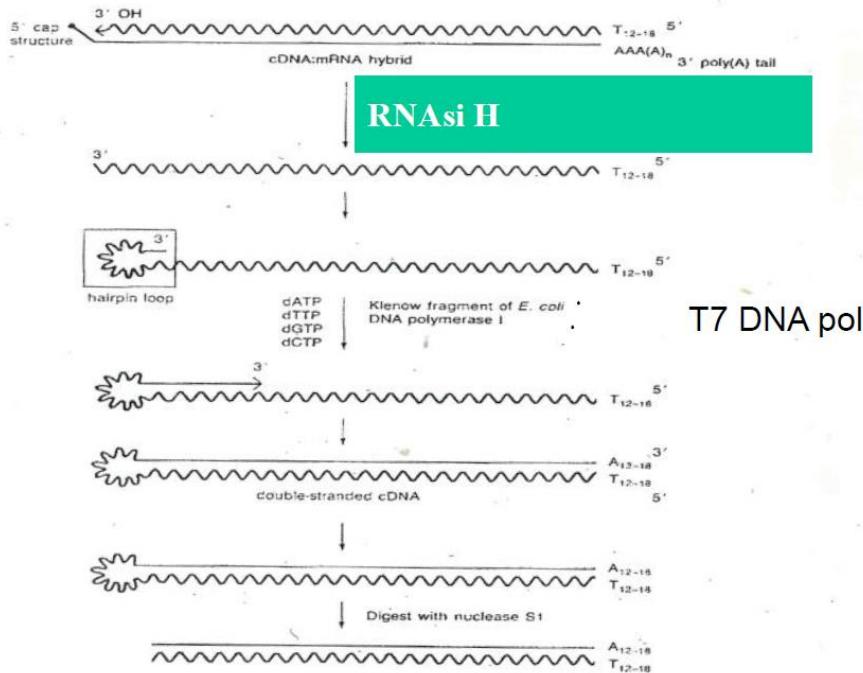
A questo punto è necessario **eliminare l'RNA ibrido**. Si utilizza l'enzima RNAsi H (Hybrid) che degrada in modo specifico l'RNA che si trova in un ibrido RNA/cDNA, si ottiene quindi cDNA ss.

A questo punto per poter clonare il cDNA **occorre renderlo ds**, per farlo ci sono 3 metodiche:

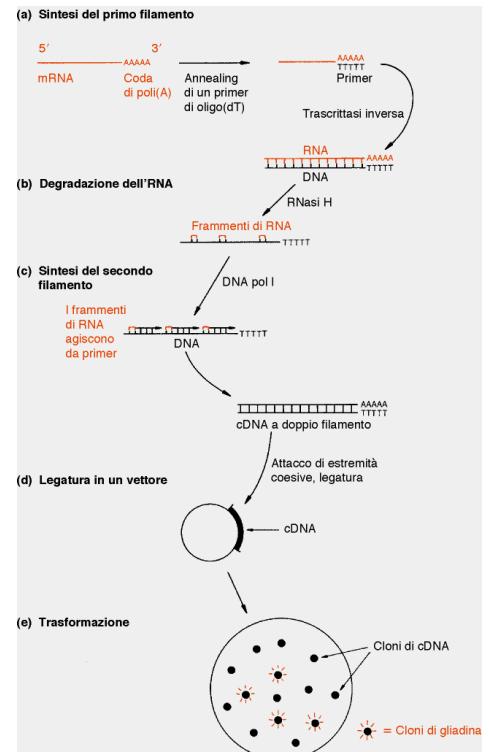
- self-priming method
- RNAsi H
- sintesi mediante coda omopolimerica

Il metodo del self-priming utilizza la *polimerasi di T7* che è particolarmente processiva. In vitro si è visto che quando si elimina l'RNA il cDNA ss fa un loop in 3' la DNA polimerasi può quindi attaccarsi al 3' e polimerizzare (con Mg^{2+} e dNTP). Questo loop va successivamente eliminato per azione della *nucleasi S1* specifica per i ss, che agisce a PH acido. Si ha però come problema la perdita di informazioni contenute nel loop, anche se esistono metodiche che permettono di recuperarla.

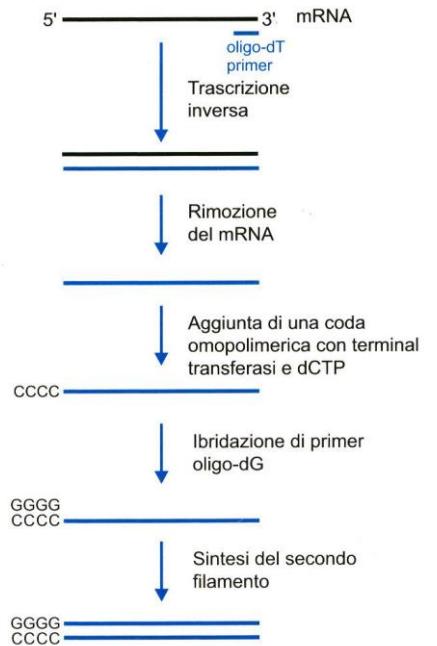
Synthesis of double-stranded cDNA by self-priming method



Utilizzando l'RNAsi H in condizioni limitanti è possibile eliminare tratti di RNA ibridato con il cDNA neosintetizzato, lasciando piccoli frammenti appaiati che saranno utilizzati come primer. Per la polimerizzazione si utilizza l'*oloenzima di E.coli*, che presenta attività di nick translation. Questa metodica *previene la formazione di loop e la perdita di informazioni*.



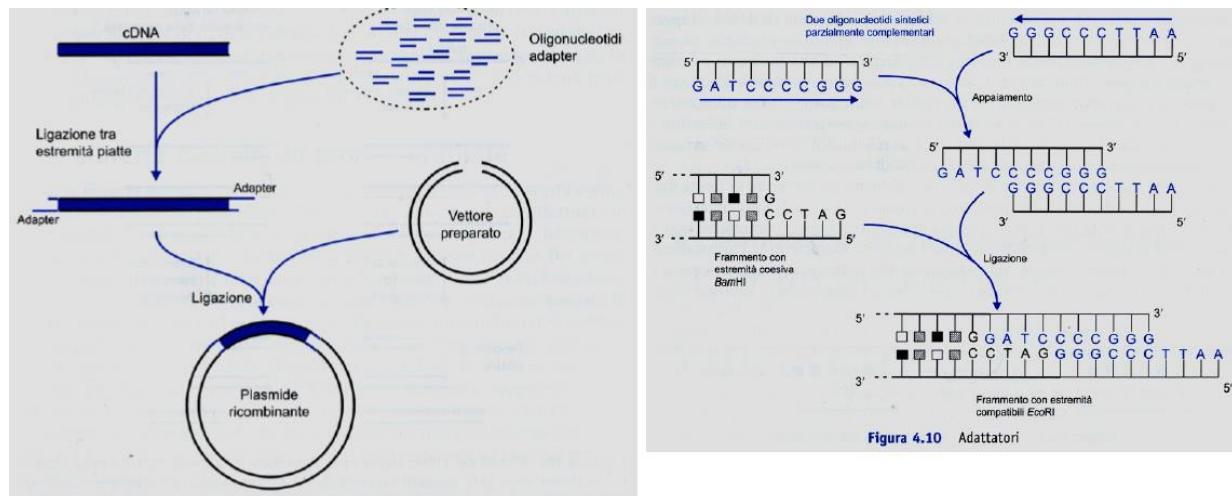
La sintesi mediante coda omopolimerica si effettua a partire dal ss di cDNA a cui viene aggiunto una terminal trasferasi in presenza solo di citosine. La *terminal trasferasi* non necessita di uno stampo e pertanto aggiungerà una *coda di C in 5'*. A questo punto verrà usato un oligo di G come primer e la polimerasi T7 sintetizza il secondo strand.



A questo punto si effettua il **clonaggio** con i cDNA ds ottenuti. I cDNA devono essere *resi clonabili* con:

- adattatori:
- linker ds con sito di restrizione

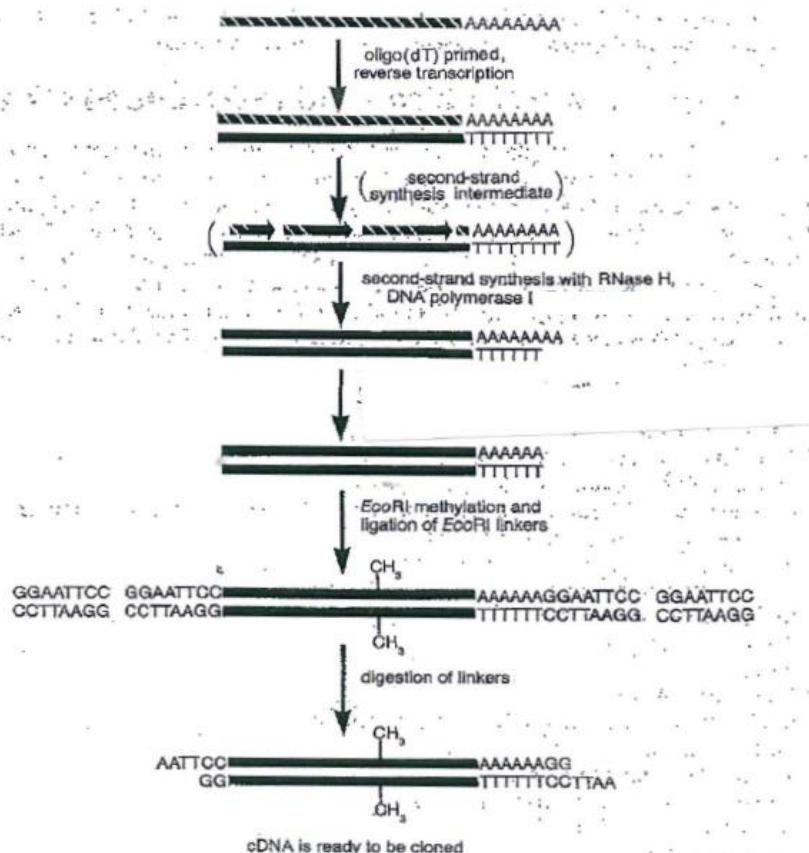
Gli adattatori sono molecole con estremità coesive preformate. Dopo aver costruito il cDNA ds viene aggiunta una miscela di questi adattatori in presenza di ligasi (*non necessitano di digestione*).



I *linker* sono una sequenza di poche basi *contenenti un sito di restrizione* e poche basi aggiuntive per permettere un miglior riconoscimento per l'enzima. Sono aggiunti alla miscela di cDNA ds assieme alla ligasi. I linker si leggeranno alle estremità del cDNA ds e successivamente sarà necessario effettuare una digestione con enzimi di restrizione per generare le estremità compatibili. Non conoscendo la sequenza del cDNA è possibile che siano presenti siti di restrizione all'interno del genoma. Dunque, è necessario

un passaggio di metilazione dei siti Eco sul cDNA e si potrà quindi aggiungere il linker contenente siti Eco non metilato che verrà riconosciuto e tagliato.

Preparazione di cDNA



I frammenti con estremità compatibili al vettore così ottenuti possono quindi essere inseriti in λ gt10 che verrà piastrato su soft agar, verranno eseguite quindi le procedure classiche precedentemente descritte e verrà eseguita l'ibridazione con una sonda radioattiva ad alta attività specifica in condizioni di stringenza. Verrà quindi effettuata un'autoradiografia, verrà isolata la placca di interesse fino all'ottenimento di segnali 100% positivi per isolare il DNA e costruire la mappa di restrizione.

Una volta costruita la mappa di restrizione si può usare il DNA isolato, viene fatto il saggio di Northern e vengono viste le sequenze di DNA. Verrà poi fatto il sequenziamento e la caratterizzazione dei cDNA.

In caso di banche a cDNA è possibile effettuare screening tramite anticorpi di banche ad espressione. A partire da cDNA ds, un vettore opportuno (λ gt11) ed E.coli, si effettua clonaggio dei cDNA ds con estremità EcoRI.

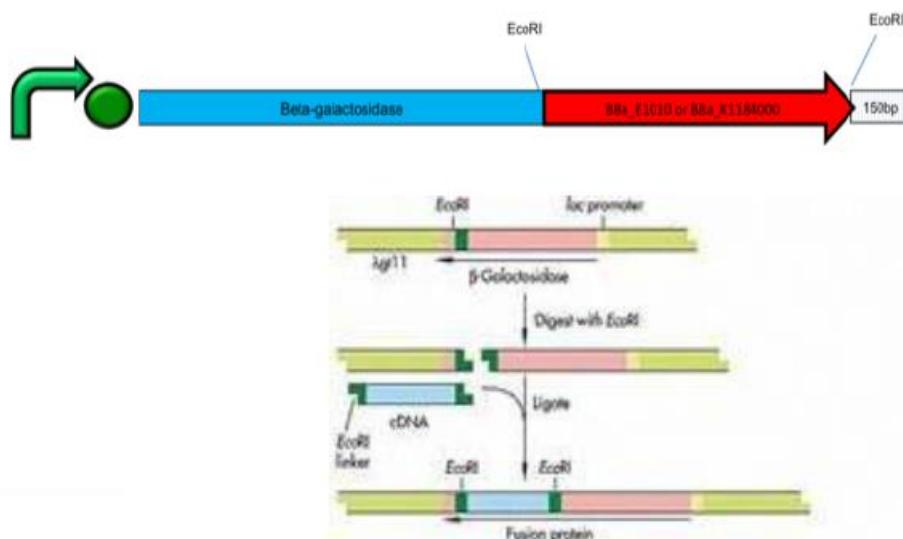
Nel vettore sono stati sostituiti i geni per la lisogenia con il gene LACZ. Quindi λ gt11 darà luogo solo a ciclo litico e, grazie alla β – galattosidasi, in presenza di X gal si potranno discriminare placche bianche da placche blu. Inoltre, i piastramenti devono essere fatti su glucosio, occorre quindi effettuare delle modificazioni di modo che LACZ del vettore non risenta della catabolyte repression.

Clonando con il vettore tal quale si otterranno **placche di lisi blu** perché il substrato IPTG verrà idrolizzato dalla β – galattosidasi attiva. Nel momento in cui invece **si clona** all'interno di LACZ nel sito di restrizione EcoRI si otterranno **placche di lisi bianche** perché verrà prodotta β – galattosidasi non funzionante.

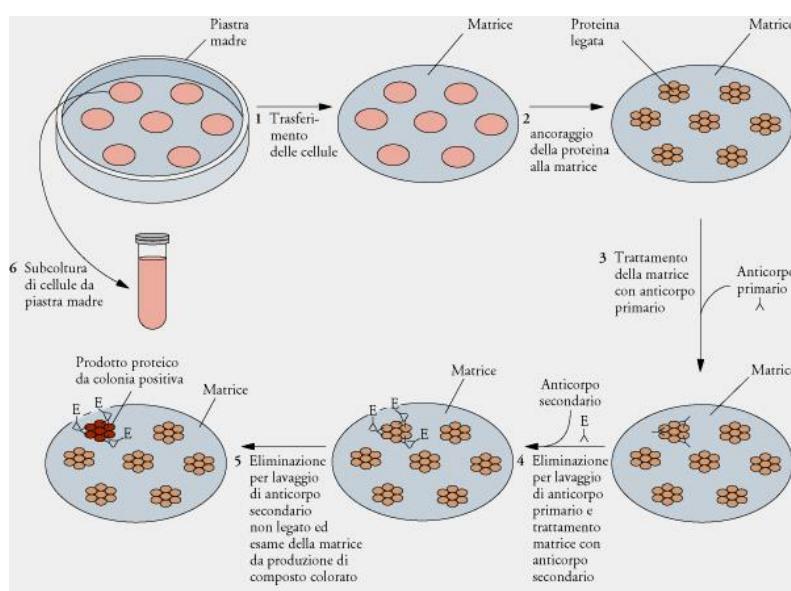
Il sito Eco è posto al 3', in fondo al gene e dopo l'ultimo codone senso e prima del codone di STOP. La situazione è quindi ben diversa dal caso in cui si abbia il sito Eco in 5' che causa rottura del frame e disattivazione inserzionale. In questo caso la β - galattosidasi viene prodotta, ma non sarà funzionale. Quando viene clonato cDNA con estremità EcoEco si avrà una **proteina di fusione**, con β - galattosidasi all'N terminale e la proteina del cDNA al C terminale.

La proteina di funzione può essere usata come target per gli anticorpi per lo screening di proteine ricombinanti.

Le dimensioni di λ gt11 sono di 37Kpb, quindi si può **clonare DNA al massimo di 14Kpb**. Si possono defosforilare le estremità del vettore in modo da impedire che le braccia del DNA del fago si richiudano su sé stesse, viene quindi effettuata la ligasi con tutti i cDNA a cui è stato addizionato il *linker* Eco a destra e a sinistra.



Si avrà formazione di **concatameri** e **packaging** in vitro. Con questi fagi verrà quindi infettato E.coli mutato. Vengono raccolti tutti i fagi di questo **piastramento** e si otterrà così la banca a **cDNA ad espressione primaria** che viene conservata. Di questa, attraverso diluizioni seriali e piastramenti, verrà fatto il **titolo** che darà informazioni sulle quantità da piastrare in assenza di X-gal.



Viene appoggiato sulle piastre un foglio di ***nitrocellulosa imbevuto di IPTG*** che indurrà la trascrizione. Viene lasciato il tutto ad incubare per un po' di tempo, dopodichè vengono recuperati i fogli di nitrocellulosa e le piastre vengono messe da parte.

A questo punto si avranno le proteine sul filtro di nitrocellulosa, non sono necessarie né denaturazioni né fissaggi. Il filtro viene trattato come un ***Western Blot*** (Saturazione con latte o BSA; Lavaggio; Anticorpo primario; Lavaggio; Anticorpo secondario; Lavaggio; con il rilevamento si otterrà un segnale relativo a una certa placca con β -galattosidasi + proteina X), viene quindi inserito l'anticorpo specifico per la proteina X.

Anche in questo caso occorre isolare la placca pura tramite vari piastramenti e rilevazioni fino ad ottenere una placca di lisi omogenea. Da questa si potrà estrarre il fago e quindi il suo DNA. Estratto il DNA viene eseguita la ***mappa di restrizione***, ci si aspetterà di trovare tutta la parte relativa al gene LACZ e DNA esogeni con estremità EcoEco prima della terminazione.

A questo punto si può ***amplificare*** tramite PCR il frammento endogeno per poterlo ***analizzare e caratterizzare***:

- Analisi di Northern, che consente di capire di che messaggero si tratta;
- Analisi di Shoutern, che consente di mappare tutto il locus;
- Sequenziamento;
- Analisi della proteina di fusione su elettroforesi.

RT-PCR concenzionale

L'amplificazione può essere effettuata per RT-PCR convenzionale, tecnica preparativa o analitica utilizzata anche per fare i tamponi SarsCovid2 e rilevare la presenza di DNA patogeno.

- Retrotrascrizione

Sfrutta la retrotrascrizione dell'RNA in cDNA utilizzando:

- retrotrascrittasi inversa
- random primer oppure con primer specifici
- oligo dT

Il primer si annealerà solo con l'mRNA per cui è specifico. Viene poi retroscritto il tutto con trascrittasi inversa in presenza di nucleotidi ed Mg²⁺; si otterrà un eteroduplex RNA/cDNA. Viene trattato l'eteroduplex con RNAsi H per eliminare l'RNA.

- Amplificazione

Si ottiene il cDNA ss, questo viene amplificato tramite PCR tradizionale multiciclica a tre fasi aggiungendo un altro primer, uno era già presente perché è stato usato per la RT – PCR. Viene cambiata anche la polimerasi, si usa la Taq.

- Separazione elettroforetica dei prodotti di PCR (ampliconi)

L'amplicone ottenuto dopo uno specifico numero di cicli di amplificazione può essere caricato sul gel di agarosio ed essere visualizzato e caratterizzato.

SEQUENZIAMENTO

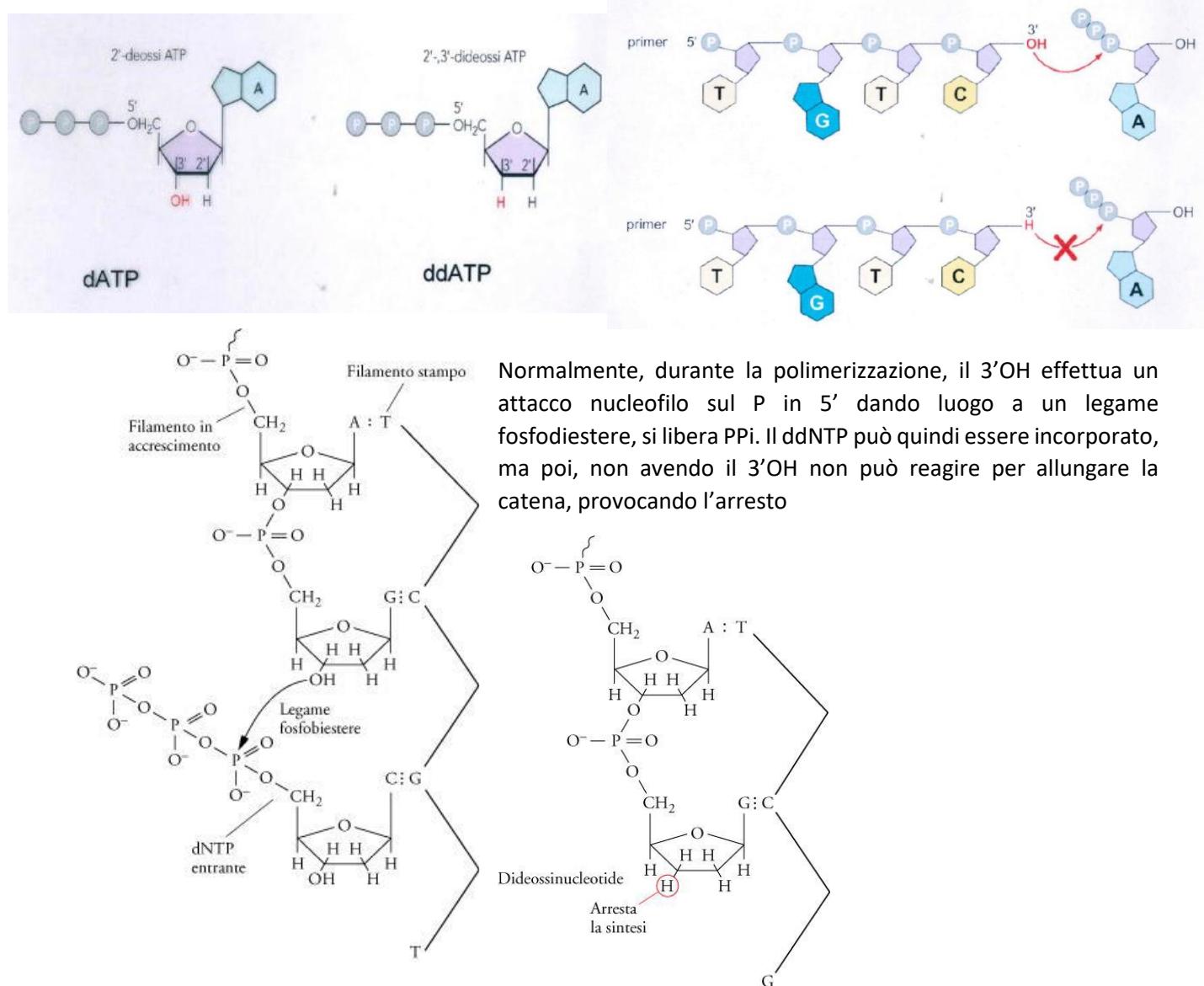
Metodo Sanger

Si tratta di un metodo enzimatico, anche detto metodo dei dideossi o a terminazione di catena. Si basa sulla sintesi in vitro di copie del DNA stampo che differiscono in lunghezza per un solo nucleotide per poi effettuare un'analisi elettroforetica ad alta risoluzione.

In soluzione, per far avvenire la reazione si avrà quindi

- DNA stampo;
- Oligo complementare (primer);
- DNA polimerasi;
- Deossinucleotidi (dNTP);
- Dideoossinucleotidi (ddNTP);

Il principio del metodo di sequenziamento enzimatico è di far copiare l'elica complementare del DNA da una polimerasi in una miscela contenente i 4 dNTP e una piccola percentuale di uno dei 4 ddNTP. Il ddNTP può essere incorporato ma blocca la sintesi successiva in corrispondenza delle posizioni dove si trova l'analogo nucleotide.



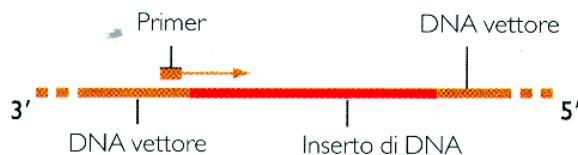
Questo metodo è alla base degli attuali metodi di sequenziamento automatizzato.

Il DNA stampo è solitamente presente in più copie e può essere ss se deriva da clonaggio in vettori fagici oppure ds se proviene da PCR, plasmidi fino 10kb o cosmidì e necessita quindi di denaturazione.

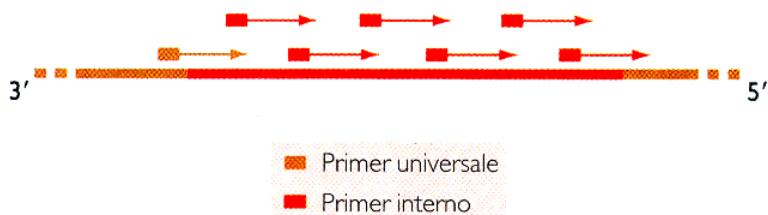
Il primer è unico. Per ogni reazione si ottiene la sequenza di un solo filamento di DNA. Il primer è un oligonucleotide complementare alla regione a monte della sequenza che si vuole ottenere, quindi per sequenziare un frammento di DNA bisogna conoscere la sequenza delle regioni fiancheggianti.

Se si ha un frammento derivato da PCR vengono semplicemente riciclati i primer, negli altri casi sono disponibili dei *primer universali* per sequenziare il frammento di interesse. I primer universali però hanno dei limiti perché la polimerasi potrebbe non essere abbastanza processiva per sequenziare l'intero DNA. La soluzione a questo problema sta nella creazione di *primer interni*.

(A) Un primer universale



(B) Primer interni



La DNA polimerasi utilizzata per il sequenziamento deve presentare particolari caratteristiche:

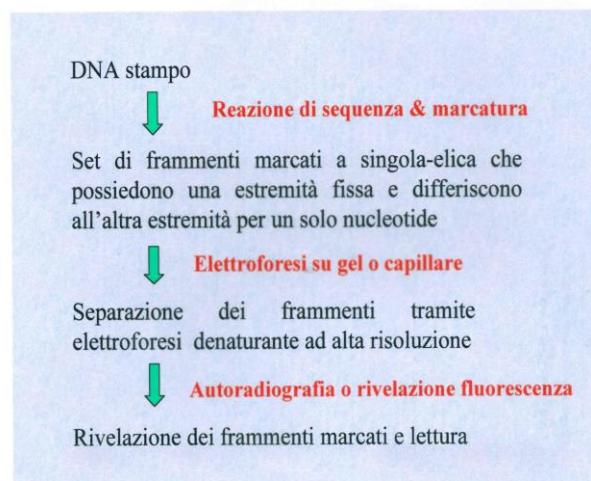
- alta processività
- alta velocità di reazione
- endo 5' → 3'
- scarsa o nulla attività di proofreading eso3' → 5' e eso5' → 3'
- non deve discriminare tra dNTP e ddNTP

Per cui le DNA polimerasi viste sono state modificate in modo tale che non discriminino tra ddNTP e dNTP. A seconda che il sequenziamento sia di tipo manuale o automatico abbiamo due tipi di polimerasi diverse;

- T7 polimerasi manipolata geneticamente, lavora a 37° e prende il nome di "sequenase" nel sequenziamento manuale;
- Taq manipolata geneticamente, lavora a 72° e prende il nome di "Thermosequenase" nel sequenziamento automatico. Durante il sequenziamento viene anche aggiunta una *pirofosfatasi* inorganica termostabile per eliminare il fenomeno della pirofosforolisi, quindi la reazione di assimilazione dei dNTP e ddNTP viene resa irreversibile.

La reazione si svolge in 4 provette, in ognuna delle quali viene aggiunto un solo ddNTP in qualità tali da ottenere reazioni parziali. L'incorporazione del ddNTP deve essere casuale. Si ha quindi una famiglia di frammenti che partono tutti dallo stesso primer, ma che si fermano a lunghezze differenti fino ad ottenere l'intera sequenza.

Per visualizzare i frammenti occorre *marcarli ed effettuare autoradiografia oppure utilizzare un sistema di fluorescenza*. Viene fatto correre il tutto su un *gel altamente risolutivo* (anche elettroforesi su capillare) in condizioni denaturanti.



La marcatura radioattiva può essere effettuata utilizzando:

- P^{32} che emette radiazioni β ad alta energia, ma decade velocemente poiché ha un tempo di dimezzamento di 15 giorni.
- S^{36} emette radiazioni β a bassa energia ma permette di visualizzare bande nette. Permette maggiore risoluzione delle bande e tempi di dimezzamento più lunghi. Permette la conservazione delle reazioni a -20° anche per settimane.
- P^{33} emette radiazioni β due volte più forti di S^{36} ma sempre meno forti di P^{32} però permette buona risoluzione e tempi brevi di esposizione autoradiografica.

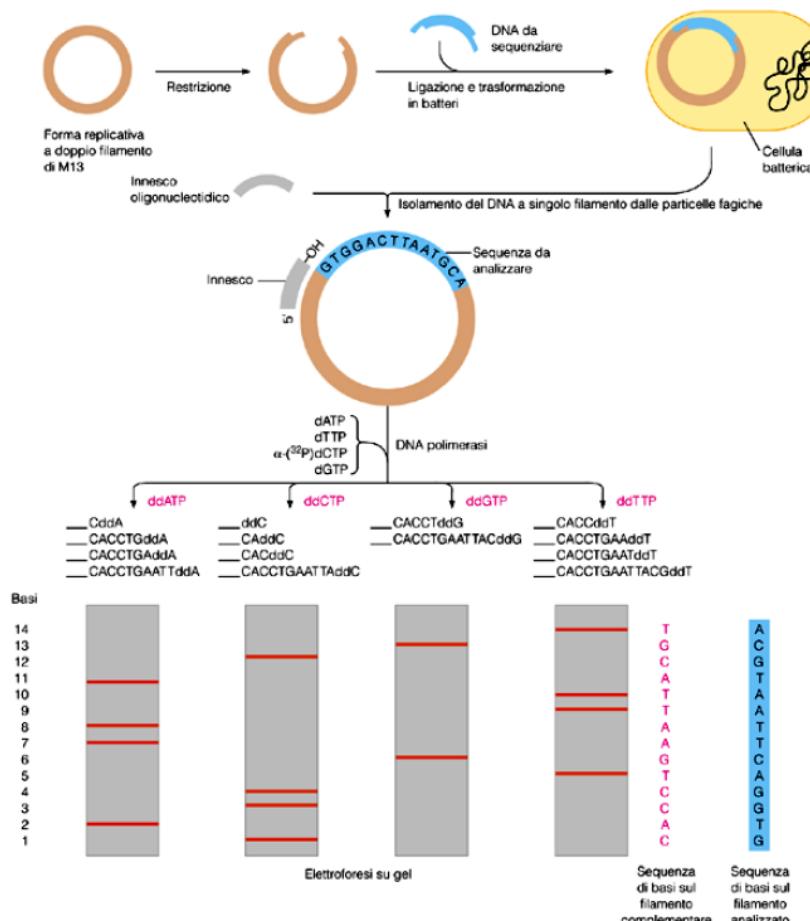
I siti di marcatura sono i seguenti:

- i primer
- i dNTP
- ddNTP

Nei dNTP e ddNTP va marcato il P_α , anche il ^{35}S viene incorporato nel P in α sostituendo un O non coinvolto nel legame.

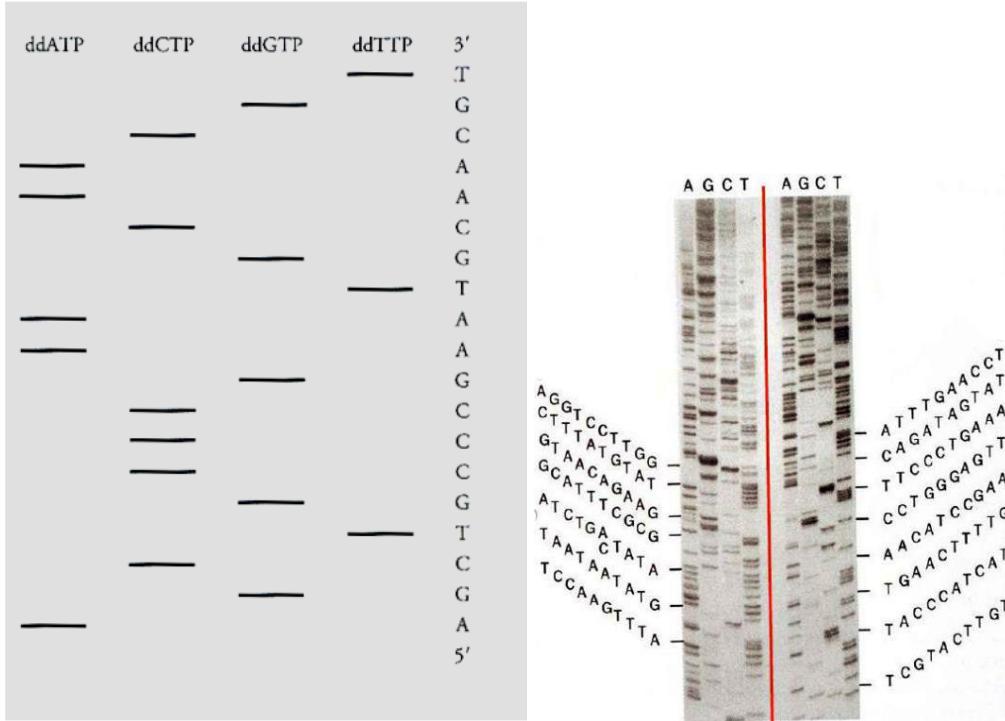
Per quanto riguarda il primer, questo va invece marcato in 5' tramite trattamento con fosfatasi e poi chinasi che aggiunge P^* in γ al 5'. Anche l' S^* al posto dell'O, sempre anche questo in γ per il primer.

La marcatura a fluorescenza è permessa dalla fluorescenza sull' α dei dNTP e ddNTP, oppure il primer può essere marcato con un fluoroforo al 5' attraverso uno *spacer*.



Clonaggio in M13 e sequenziamento con il metodo di Sanger

Il sequenziamento manuale prevede il caricamento del contenuto delle 4 provette di reazione su gel di poliacrilammide a maglie molto strette con urea (non SDS) per mantenere le condizioni denaturanti. Dopo la corsa elettroforetica altamente risolutiva, la lastra sul gel viene esposta ad autoradiografia (viene utilizzata marcatura radioattiva). I frammenti migreranno in base al PM e quindi in relazione al numero di pb dei frammenti. *Per convenzione le bande si leggono dal basso verso l'alto.*

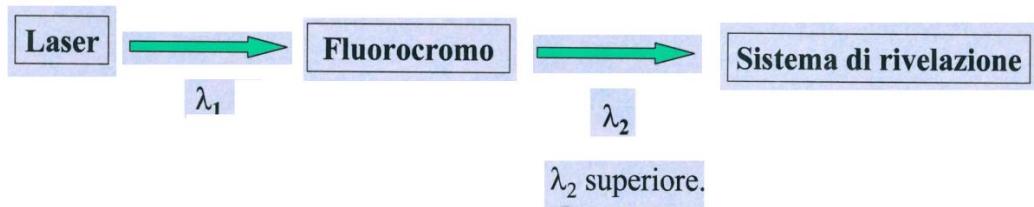


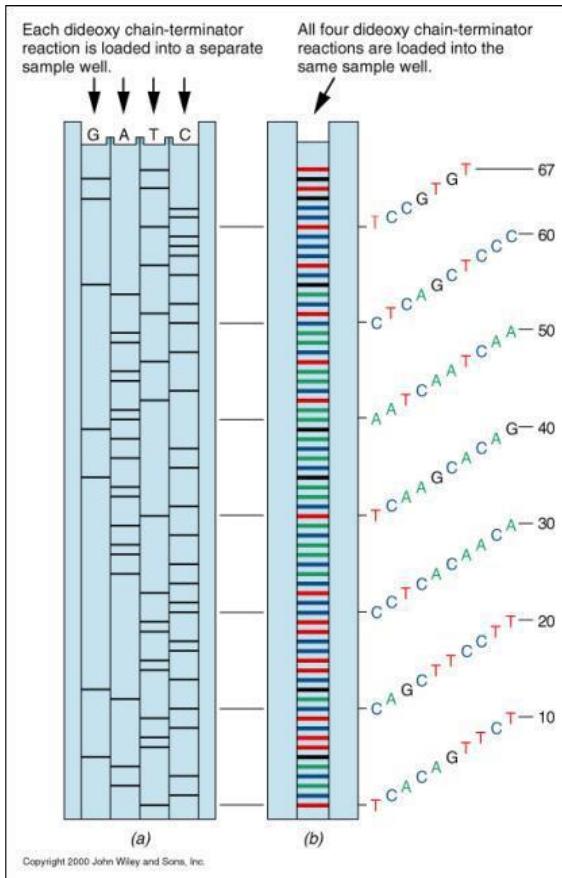
Il sequenziamento automatizzato è meno laborioso e veloce. Nei processi automatizzati si usano fluorofori e laser che eccitano il fluoroforo a una determinata λ , il detector rileverà l'emissione.

Viene usata molto la marcatura dei ddNTP nel caso di fluorocromi, ci sono due metodi:

- One – dye sistem \rightarrow ciascun ddNTP è marcato con lo stesso fluorocromo, vengono quindi usate 4 provette diverse;
- Four – dyes sistem \rightarrow i ddNTP sono tutti marcati con fluorocromi diversi, si può quindi usare una sola provetta/pozzetto per reazione. Il sistema è un po' più complicato.

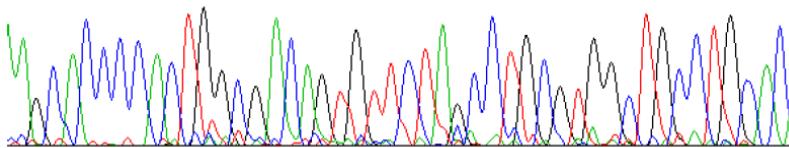
Il computer leggerà un elettrofluorogramma e restituirà la sequenza.





File: 01-M1 Sequence Name: M1 Run ended: Sep 18, 2001

300	310	320	330	340
. A G C A C C C C A C T G G C G A C A G T G T T C T A G C C T G C G T G G C T G C C T G C A C				



Il computer leggerà un *elettrofluorogramma* e restituirà la sequenza.

Pirosequenziamento

Fa parte delle tecniche di Next Generation sequencing che consentono di sequenziare molti frammenti di DNA contemporaneamente. Rispetto a Sanger presentano *efficienza minore* in termini di numero di basi sequenziate.

Sono tutte applicabili al sequenziamento di DNA genomici amplificato prima del sequenziamento: non necessitano di clonaggio in vettori batterici o fagici e non è necessaria la migrazione su gel.

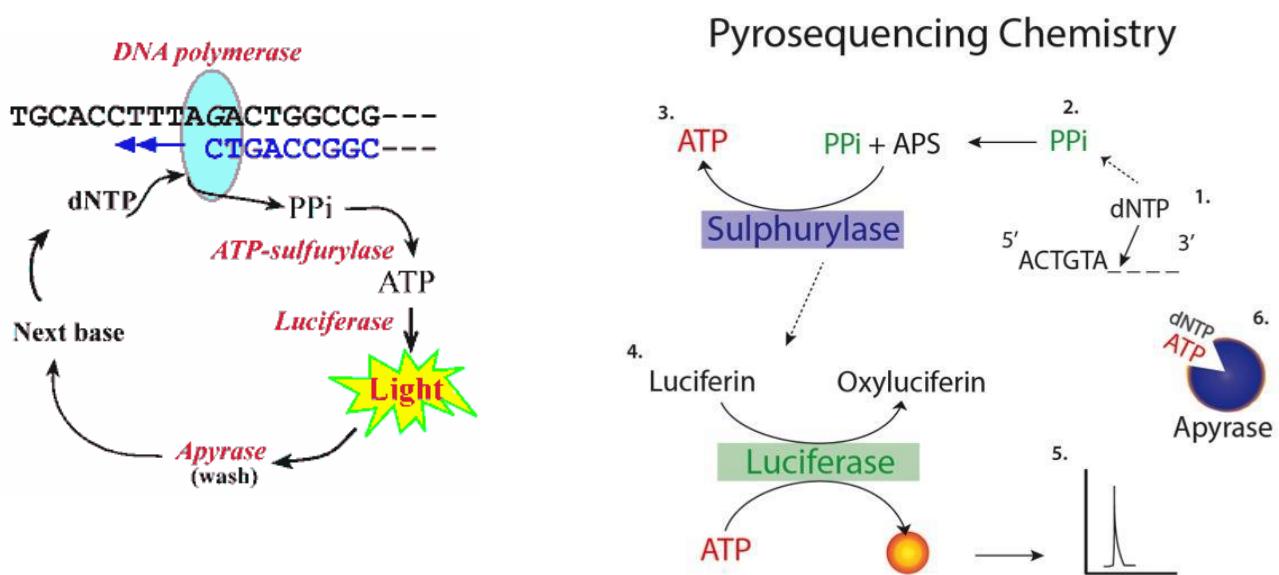
Il punto di partenza è dunque DNA genomico che viene frammentato attraverso metodi chimici/fisici/enzimatici. I frammenti devono poi essere uniti ad adattatori e poi, attraverso particolari PCR (PCR a breccia) possono essere fissati su vari supporti di natura diversa. Su questi supporti vengono poi sequenziati.

Quello che è importante è che la lettura viene effettuata *step by step*, dunque si copia e si legge in simultanea.

Il pirosequenziamento si basa sul rilevamento del PP_i liberato in seguito all'attacco di un dNTP al filamento polimerizzato. Prevede 4 passaggi principali che vengono ripetuti ciclicamente:

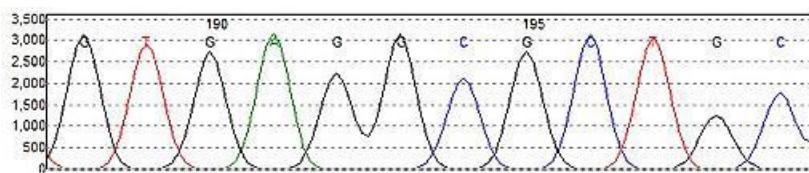
1. La sequenza da analizzare viene amplificata per PCR e fissata sul supporto come singola elica insieme a primers adeguati, enzimi (ATP fosforilasi, luciferasi, apirasi) e due substrati degli enzimi (APS e luciferina).
2. Viene aggiunto un nucleotide, se questo è complementare la polimerasi catalizzerà l'aggiunta e si avrà liberazione di un pirofosfato. Nel caso in cui il nucleotide non venga incorporato l'apirasi lo degrada.
3. Se il nucleotide viene incorporato, il pirofosfato liberato viene trasformato in ATP ad opera dell'ATP sulfurylasi $APS + PPi \rightarrow ATP$. L'ATP prodotto viene usato dalla luciferasi per ridurre la luciferina ad ossiluciferina; verrà emesso un segnale luminoso che verrà rilevato da una camera fotosensibile.

Si aggiungono ciclicamente tutti e 4 i d(NTP) fino alla deduzione completa della sequenza.

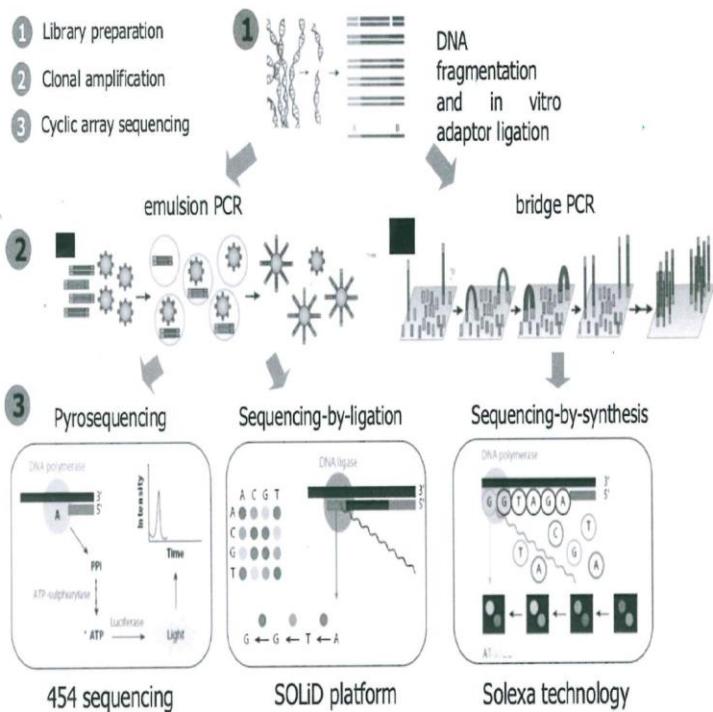


Il segnale luminoso prodotto ogni volta dalla luciferina viene registrato in un apposito "pirogramma". Il segnale sarà proporzionale all'ATP prodotto e quindi al nucleotide inglobato:

- un picco di intensità doppia, ad esempio, rileva che nello stesso ciclo sono stati inglobati 2 dNTP (ripetizione della stessa base sul template),
- un segnale nullo indica che il dNTP aggiunto in quel ciclo non è complementare.



Next-generation DNA sequencing



MUTAGENESI

Strategia che consente di indurre mutazioni nel genoma di un organismo.

Si distingue la genetica classica e la genetica inversa:

- genetica classica: dall'osservazione di una alterazione fenotipica spontanea o indotta da agenti chimico-fisici *in vivo*, all'analisi genetica che porta alla individuazione del gene mutato,
- genetica inversa: dall'introduzione *in vitro* di una mutazione in uno specifico frammento di DNA alla successiva analisi degli eventi biologici.

La mutagenesi *in vitro* permette di modificare una sequenza nucleotidica, può essere:

- casuale,
- localizzata,
- sito diretta.

La mutagenesi permette di studiare *regioni regolative* e fare *ingegneria proteica*.

es. modificaione sito diretta dei promotori -10 e -35 del promotore di Lac per rendere il sistema meno leaky, modificaione del genoma di E.coli in modo che esprimesse solo l'ω-peptide.

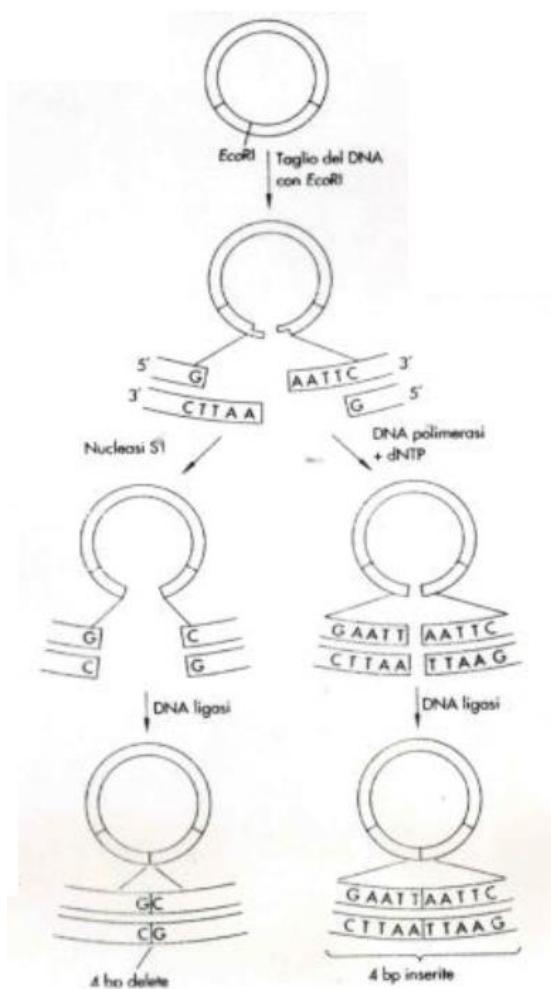
Ingegneria proteica

Il Protein engineering implica l'uso di tecniche di manipolazione genetica per alterare la sequenza codificante di un gene (clonato) e quindi della proteina da esso codificata.

Può essere usata per:

- Aumentare la *stabilità* della proteina;
- Aumentare la *purezza* della proteina durante l'estrazione;
- Aumentare *l'espressione*;
- Modificare *l'interazione* con i cofattori (DNA polimerasi per i sequenziamenti);
- Aumentare *l'attività* di un enzima;
- Modificare la *specificità* di un enzima;
- Studiare la *funzione* di una proteina.

Un esempio semplice di mutagenesi localizzata lo si può avere **manipolando un sito di restrizione**, nel momento in cui viene tagliato e si generano estremità "blunt" viene eliminata parte della sequenza e vengono quindi creati mutanti in modo localizzato, perché si sa dove è stata modificata la sequenza.



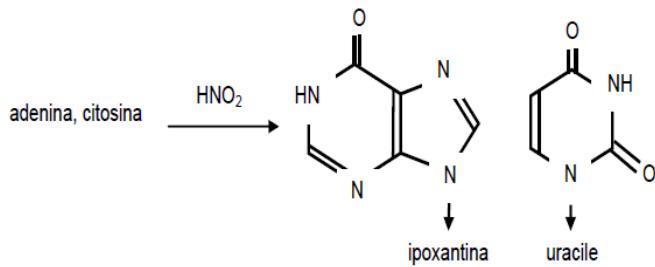
Si possono ottenere mutazioni anche **sostituendo le basi**. Ciò può essere ottenuto attraverso:

- mutageni chimici che agiscono sulla catena di DNA già formata
- incorporazione di analoghi di basi;
- errato accoppiamento di basi, solo durante la sintesi di uno strand;

Mutagenesi casuale

I **mutageni chimici** sono ad esempio l'acido nitroso e il bisolfato di sodio. L'**acido nitroso** ha come target A e C, le quali vengono deamminate ottenendo ipoxantina e uralcile. L'ipoxantina si appaia con C e l'uracile con A. Il **biossido di sodio** ha come target C, che viene deaminata in U, la quale si appaia con A.

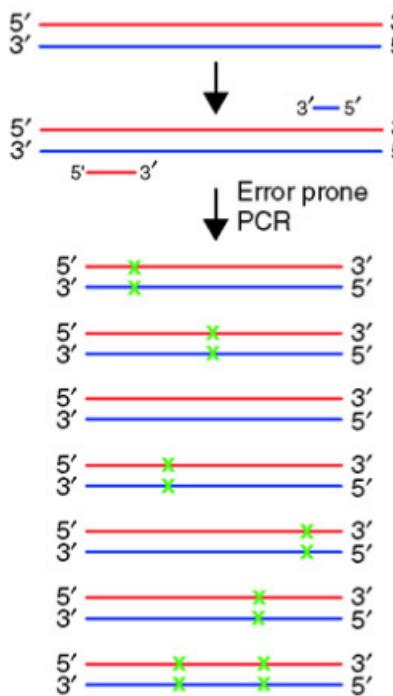
Sono mutazioni random perché avranno luogo lungo tutta la sequenza, non è quindi necessario conoscere la sequenza per provocare queste mutazioni; si possono creare mutanti inattesi con proprietà nuove che possono essere interessanti, verranno create delle librerie di mutanti.



La mutazione può essere introdotta durante PCR (*error prone PCR*). Anche questa è una mutazione random. Consiste di un'amplificazione in cui si verifica, con frequenza opportuna, l'introduzione di errori di incorporazione delle basi.

A tale scopo verranno utilizzati:

- DNA polimerasi senza attività di proofreading;
- Alte concentrazioni di Mg²⁺;
- Concentrazioni di dNTP sbilanciate;
- Temperature di annealing di molto inferiori alla T_m degli oligo;
- Alta forza ionica;
- Aumento del numero di cicli;



Mutagenesi sito diretta

Può avvenire sempre tramite PCR. Si basa su oligonucleotidi a sequenza nota in cui le mutazioni sono state introdotte in posizioni predeterminate.

Si hanno diversi tipi di procedimento in base alla grandezza delle mutazioni che si vogliono introdurre:

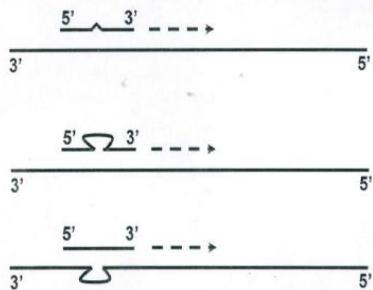
- Possono essere mutazioni puntiformi o di poche basi < 10;
- Possono esserci inserzioni/delezioni molto estese, fino a 100pb;

Se si vogliono inserire delle piccole mutazioni verranno usati oligo non perfettamente complementari al templato, per i quali verranno studiate le condizioni di annealing.

Avranno infatti una formula diversa per la T_m rispetto a quella che si usa per i filamenti perfettamente complementari; si userà la formula:

$$T_m = [81,5 + 0,41 \cdot (\%G + C) - \frac{650}{\text{numero di basi} - \% \text{ mismatch}}] ^\circ\text{C}$$

Singoli o brevi mismatches, inserzioni o delezioni in mezzo al primer



Gli oligonucleotidi utilizzati devono avere le seguenti caratteristiche:

- lunghezza compresa tra 18 e 30 pb, spesso 20-25
- T_m compresa tra 52 °C e 64 °C, omogenea tra i primer utilizzati

N.B. Per gli oligonucleotidi che hanno basi non omologhe alle estremità il calcolo della T_m viene fatto solo sulla sequenza completamente omologa, mentre per gli oligo contenenti mismatch la formula è modificata come segue:

$$T_m = [81,5 + 0,41 \cdot (\%G + C) - \frac{650}{N \text{ basi} - \% \text{ mismatch}}] ^\circ\text{C}$$

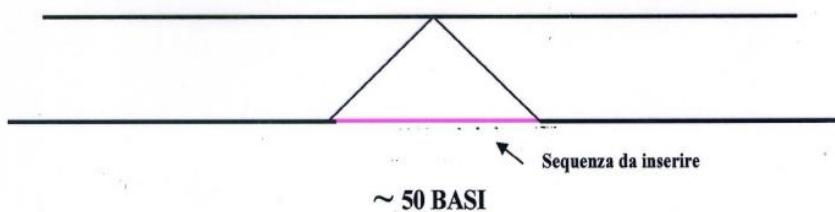
- self annealing assente
- dimerizzazione assente
- stabilità dell'annealing con la sequenza bersaglio
- specificità
- eventuale presenza di siti di restrizione.

Il Quick Change utilizza la trasformazione di un ospite.

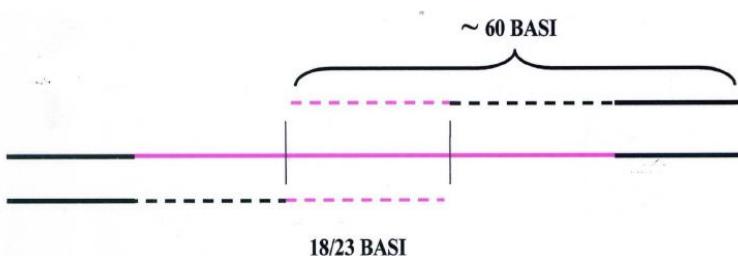
Viene estratto il DNA da mutagenizzare da un ceppo *dam*⁺. Per ottenere la mutazione si utilizza l'oligonucleotide mutato, si denatura e si effettua l'annealing. La polimerizzazione è effettuata dalla polimerasi *Pfu* o *Tu*. Alla fine della polimerizzazione verrà trattato il tutto con *Dpn*1 che frammenterà il templato parentale. *Dpn*1 è una endonucleasi specifica per il DNA metilato nella sequenza GATC.

In vitro si avrà quindi DNA nikkato che conterrà la mutazione, avviene quindi la trasformazione. Verrà poi prelevato il DNA e verrà sequenziato per verificare che abbia la mutazione.

Un'altra tecnica di mutagenesi che utilizza la PCR e che permette l'inserzione di un'intera sequenza di 50 pb prevede l'attuazione di 2 PCR separate e successivamente una terza PCR che utilizzerà come templati i prodotti di amplificazione delle due PCR iniziali.

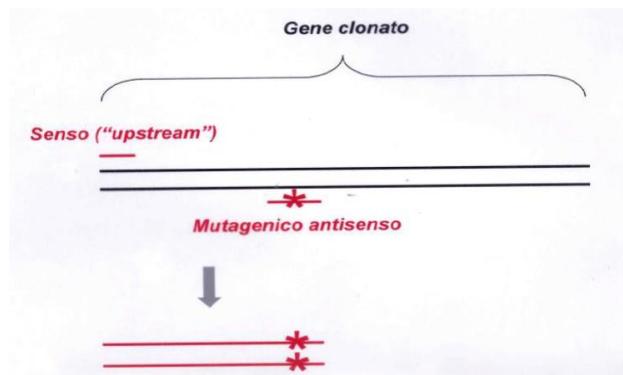


Si devono quindi generare gli oligonucleotidi mutagenici, uno senso e uno antisenso. La lunghezza totale di questi oligonucleotidi può essere di 60 o più basi e devono avere una parte perfettamente complementare al templato. La porzione 5' conterrà la sequenza da inserire.

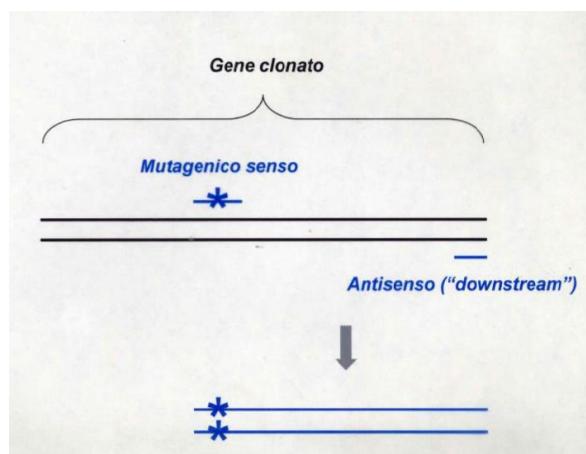


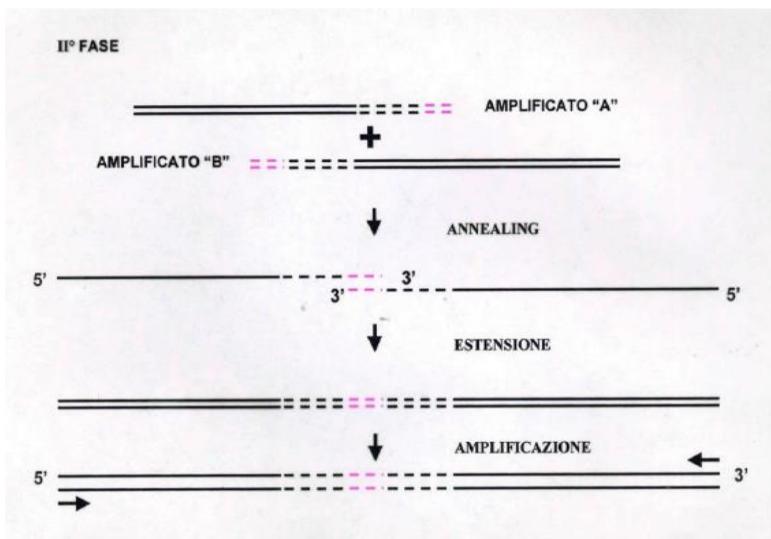
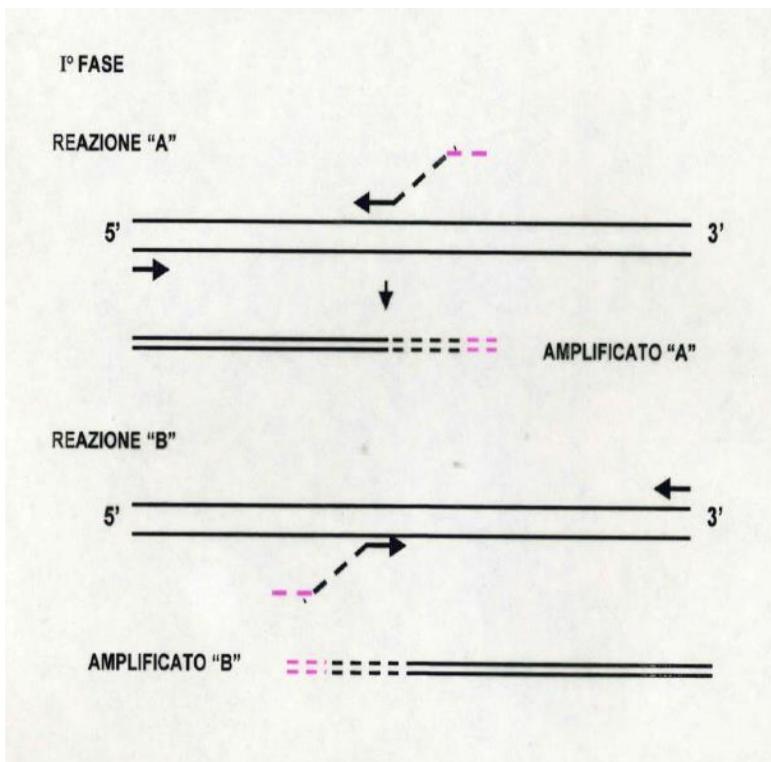
Questi due primer saranno utilizzati separatamente per effettuare 2 PCR abbinati ad altri 2 primer.

Prima PCR: amplificazione di una porzione del gene clonato.



Seconda PCR: effettuata in parallelo.





I primer utilizzati sono quello senso e antisenso normali utilizzati nelle prime reazioni di PCR abbinati ai primer mutagenici. Questo permette l'amplificazione della sequenza contenente la mutazione desiderata.

Mutagenesi mediante PCR

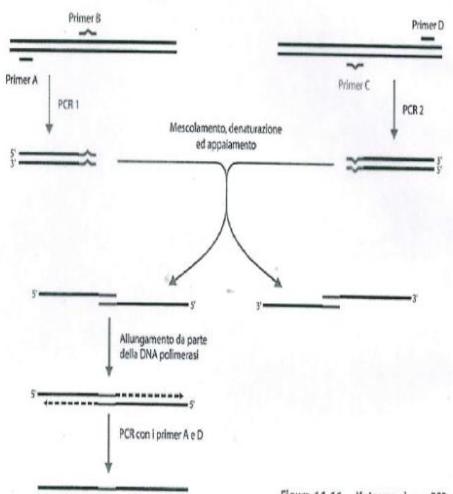


Figura 11.11 Mutagenesi con PCR