

Métodos para Encontrar e Expressar Genes Importantes

Antes do advento da biotecnologia, a biologia molecular podia ser dividida em dois campos: bioquímica e genética molecular. A bioquímica é o estudo da química de sistemas vivos, enquanto que a genética molecular é uma área que utiliza organismos simples, tais como bactérias e bacteriófagos, para entender a genética no seu nível mais básico, tal como visto no trabalho de Alfred Hershey, Salvador Luria e Charles Yanofsky. Embora esses dois campos de estudo ainda estejam atualmente muito vivos, a biologia molecular utiliza-se intensamente de cada um deles e, combinada com a biotecnologia, cria uma área nova.

As questões crescentes que os bioquímicos e os geneticistas moleculares enfrentaram no final da década de 1960 e no início da de 1970 tinham enfoque na regulação de genes. A partir do trabalho de François Jacob e Jacques Monod, os cientistas entenderam que havia regiões de genes chamadas de promotores que regulavam a expressão gênica. Os geneticistas moleculares sabiam que os genes estão localizados em cromossomos e deduziram que um gene diferente codifica cada proteína. Entretanto, como o sistema inteiro funciona? Com o que um promotor se parece? Qual é a organização dos genes? Há alguma surpresa na estrutura dos genes? A única maneira de responder essas questões era isolando os genes individualmente.

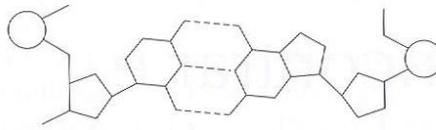
No início da década de 1970, os bioquímicos estavam utilizando novas técnicas para purificar e estudar um número crescente de proteínas. As vias bioquímicas envolvidas em cada processo celular possível foram sendo resolvidas, incluindo a síntese de aminoácidos, nucleotídeos e ácidos graxos. Vias para a produção de moléculas de energia, tais como glicólise e ciclo de Krebs, foram sendo determinadas. Ao mesmo tempo, a genética molecular foi fazendo progressos no entendimento da relação entre DNA e proteínas. No fundo de todos esses estudos estava o gene.

A atenção dos principais bioquímicos e geneticistas moleculares estava presa pela idéia de purificar os genes para estudo adicional. A estratégia inicial foi isolar genes abundantes como aqueles que codificam para RNA ribossômico (rDNA), conhecidos por existirem em centenas ou milhares de cópias por célula. No entanto, era provável que o rDNA fosse diferente dos genes que codificam para proteínas. Sendo assim, qual gene deveria ser estudado primeiro?

O PROBLEMA DOS BIOQUÍMICOS E GENETICISTAS MOLECULARES

A magnitude de crescimento que tem ocorrido na biologia e na genética molecular desde os primeiros experimentos de clonagem de Paul Berg, Stanley Cohen, Herbert

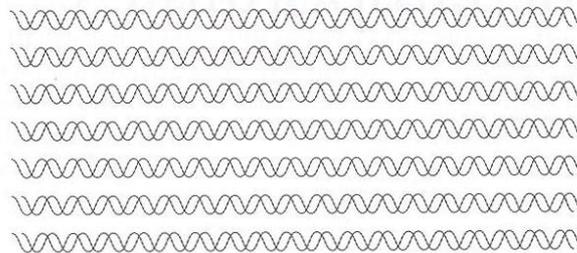
A massa atômica média de um par de bases é 635 dátons (um dáton é 1/12 da massa de um átomo de carbono)



O gene β -globina tem 2.000 pb de comprimento

Assim, a massa atômica do gene β -globina tem

$$2.000 \text{ pb} \times 635 \text{ dátons/pb} = 1,27 \times 10^6 \text{ dátons}$$



A massa atômica média de um par de bases é 635 dátons (um dáton é 1/12 da massa de um átomo de carbono)

O gene β -globina tem 2.000 pb de comprimento

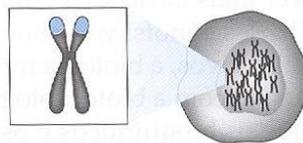
Assim, a massa atômica do gene β -globina tem

$$2000 \text{ pb} \times 635 \text{ dátons/pb} = 1,27 \times 10^6 \text{ dátons}$$

Massa do gene β -globina em um humano adulto

Há duas cópias do gene β -globina por célula

Há 10^{14} células por indivíduo



Assim, a massa atômica total de DNA β -globina por indivíduo é de

$$1,27 \times 10^6 \text{ dátons/gene} \times 2 \text{ genes/célula} \times 10^{14} \text{ células/indivíduo} = 2,54 \times 10^{20} \text{ dátons}$$

Se há $6,02 \times 10^{23}$ dátons por grama, então:

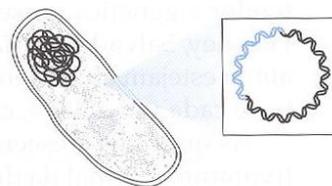
$$\frac{2,54 \times 10^{20} \text{ dátons}}{6,02 \times 10^{23} \text{ dátons/grama}} = 0,00042 \text{ gramas} = 0,42 \text{ mg} = 420 \mu\text{g de DNA } \beta\text{-globina}$$



Massa do gene β -globin em um litro de *E. coli*

Há 500 cópias do gene β -globina (em plasmídeo) por célula

Há 5×10^{11} células por litro

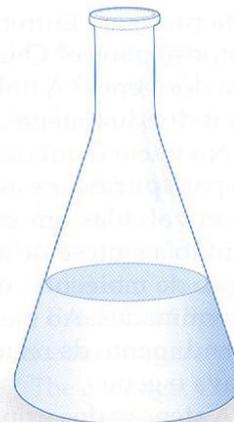


Assim, a massa atômica total de DNA β -globina por litro é:

$$1,27 \times 10^6 \text{ dátons/gene} \times 500 \text{ genes/célula} \times 5 \times 10^{11} \text{ células/litro} = 3,175 \times 10^{20} \text{ dátons}$$

Se há $6,02 \times 10^{23}$ dátons por grama, então:

$$\frac{3,175 \times 10^{20} \text{ dátons}}{6,02 \times 10^{23} \text{ dátons/grama}} = 0,000527 \text{ grama} = 0,527 \text{ mg} = 527 \mu\text{g } \beta\text{-globina DNA}$$



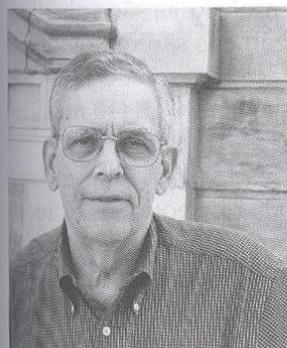
Massa do DNA de β -globina em humano adulto versus cultura de 1 litro de *E. coli* contendo o gene de β -globina inserido no plasmídeo

Boyer e Annie Chang é difícil de compreender. Para apreciar tais experimentos e entender melhor a base da tecnologia de hoje, deve-se considerar como os primeiros genes específicos foram clonados e isolados. Obviamente, a clonagem inicial de genes foi uma espécie de área de treino em que métodos e técnicas foram testados. Sem maiores surpresas, os primeiros genes isolados eram aqueles com produtos abundantes, bem caracterizados e potencialmente fáceis de isolar. Os genes em codificação para hemoglobina estavam entre os primeiros genes de mamíferos a serem isolados usando a tecnologia do DNA recombinante.

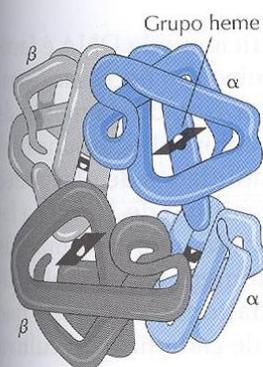
O desenho experimental usado por Philip Leder e seus colaboradores, da Universidade de Harvard, para isolar o gene da β -globina de camundongo a partir do DNA genômico, ilustra como a clonagem molecular resolveu de forma simples dois problemas críticos da análise bioquímica: (1) separando o gene de interesse da massa de DNA genômico e (2) purificando DNA suficiente para análise. Pouco depois do trabalho de Leder, Tom Maniatis, em Harvard, no California Institute of Technology, isolou genes de globina humana usando um protocolo um pouco diferente.

A hemoglobina é a molécula transportadora de oxigênio das células vermelhas do sangue. Ela é composta de quatro proteínas, dois polipeptídeos α -globina e dois β -globina, circundando um grupo heme central. Embora os dois polipeptídeos globina sejam similares em estrutura, cada um é codificado por um gene diferente. A proteína β -globina humana tem 147 aminoácidos de comprimento. É difícil prever o tamanho do gene a partir da proteína. Além dos 441 pares de bases (pb) que codificam aminoácidos (lembrar que há três nucleotídeos por códon), os genes β -globina incluem uma região 5' não-traduzida, que se estende do promotor ao códon de iniciação (51 pb), uma região 3' não-traduzida, estendendo-se do códon de terminação à seqüência sinal poli-A (134 pb), mais dois íntrons (1374 pb). Assim, o gene β -globina tem na realidade 2.000 pb de comprimento. Uma vez que um par de base tem uma massa molecular média de 635 dáltons, o gene β -globina tem uma massa total de cerca de $1,27 \times 10^6$ dáltons. Após a multiplicação desse algarismo pelo número de cópias do gene por célula (2) e pelo número de células no organismo humano adulto (10 trilhões, ou 1×10^{14}), calcula-se uma massa total de 0,00042 g (0,42 mg ou 420 μ g) de DNA de β -globina por indivíduo. Sem dúvida, tentar arrancar genes humanos direto das células não é a melhor maneira. A análise bioquímica poderia necessitar em torno de 1 mg de DNA, que parece pouco, considerando-se que uma colher de chá de açúcar pesa aproximadamente 4.500 mg. No entanto, seria necessário pegar o equivalente à massa celular de 2,5 humanos para extrair 1 mg de DNA de β -globina.

Mesmo se essa tarefa difícil e onerosa fosse completada com sucesso, o problema de separar o gene β -globina do restante do DNA seria quase impossível, pois o gene β -globina representa 1/1.500.000 dos 3 bilhões (3×10^9) pares de bases do DNA que compõem um grupo haplóide de cromossomos humanos (o genoma humano haplóide). Assim, o processo de isolamento do equivalente a 2,5 humanos produziria uma massa total de 1,5 kg de DNA genômico, do qual apenas 1 mg seria de DNA de β -globina. Em função de o DNA ser um polímero de apenas quatro subunidades repetidas, para todos os objetivos e propósitos, bioquimicamente, todo fragmento de 2.000 pb parece idêntico a qualquer outro de 2.000 pb. O problema é ainda pior que a proverbial procura da agulha em um palheiro, no qual ao menos a agulha parece diferente do feno circundante.



Philip Leder



Hemoglobina

(Redesenhada, com permissão, de Rawn J.D. 1989.

Biochemistry, p. 125. Neil Patterson Publishers, Burlington, North Carolina.)

A SOLUÇÃO DOS BIÓLOGOS MOLECULARES: CRIANDO E EXAMINANDO UMA BIBLIOTECA

Ao invés de tentar tirar um gene da massa de DNA genômico, biólogos moleculares propuseram uma estratégia original. A idéia é separar fisicamente cada gene no genoma e, depois, selecionar esses fragmentos para apanhar aquele desejado.

A biblioteca de DNA

Iniciar-se-á com a criação de uma biblioteca genômica, dando esse nome porque, como uma biblioteca, ela contém uma grande quantidade de itens que são procurados com facilidade. Para criar essa biblioteca, começa-se pelo isolamento do DNA genômico (uma vez que Leder estava interessado no isolamento dos genes de globina humana, ele usou células humanas e montou uma biblioteca genômica humana). As células são isoladas e lisadas em uma solução aquosa. Proteínas, lipídeos e outras moléculas são eliminados por extração orgânica na qual tais moléculas são solubilizadas, mas o DNA está insolúvel. Os traços de solventes orgânicos são removidos da fase aquosa pela precipitação do DNA usando-se etanol ou isopropanol. O DNA precipitado é solubilizado em uma solução aquosa tamponada, como TE (10 mM Tris-HCl, 1 mM EDTA).

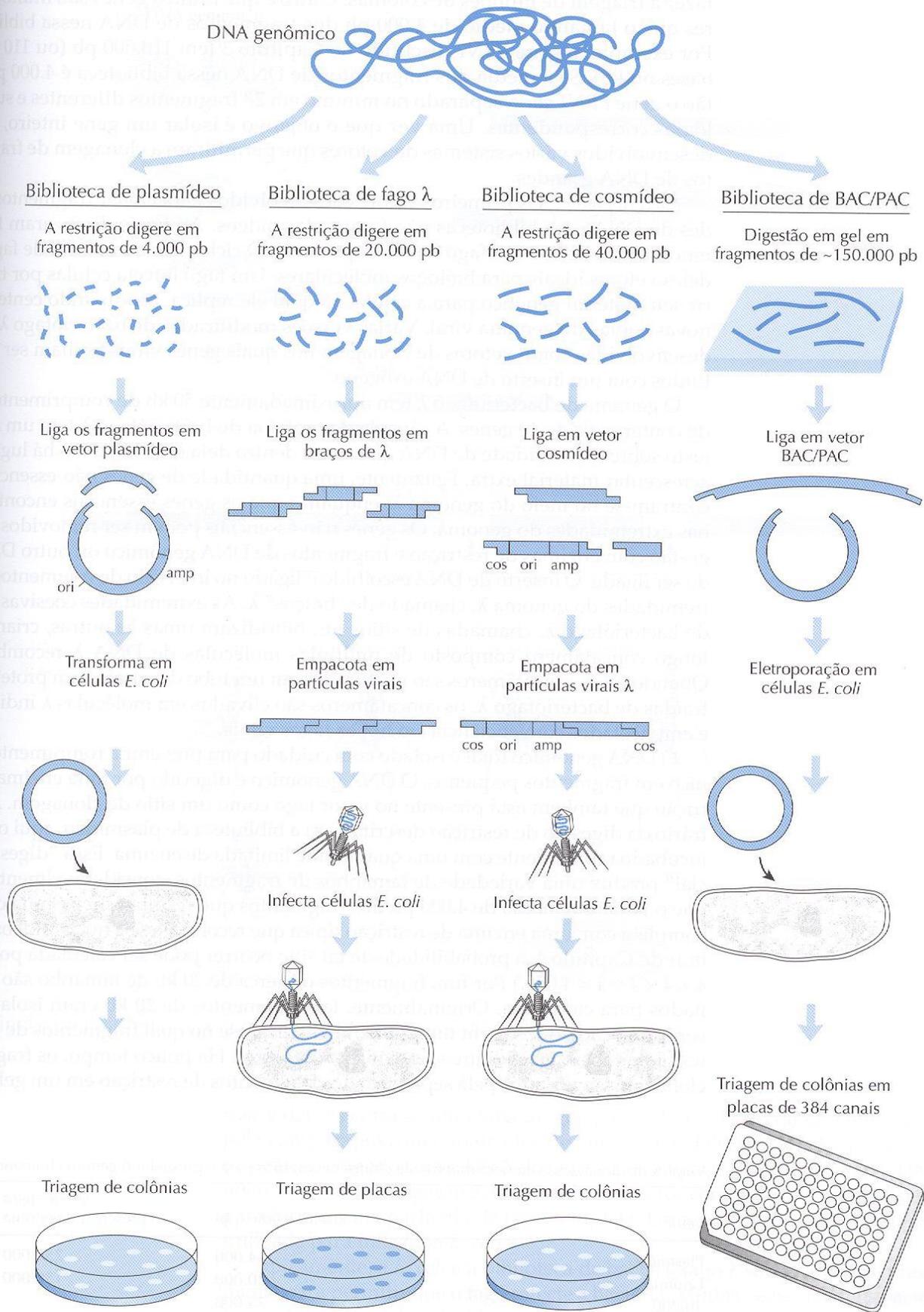
Uma amostra de DNA genômico humano é cortada (clivada) com uma endonuclease de restrição, que corta em média uma vez a cada 4.000 nucleotídeos e produz cerca de 750.000 fragmentos de DNA. Tais fragmentos genômicos de restrição são associados com moléculas de plasmídeos que foram digeridas com a mesma enzima de restrição. O vetor plasmídeo e os fragmentos de DNA genômico são unidos com a enzima DNA ligase (conforme descrito no Capítulo 4). A coleção de 750.000 plasmídeos recombinantes cria uma biblioteca genômica. Então, os plasmídeos recombinantes são transformados em células competentes de *Echerichia coli*. Teoricamente, cada célula capta um plasmídeo diferente que carrega um fragmento diferente de DNA humano. As bactérias transformadas são diluídas e distribuídas em placas como células individuais. Cada célula se reproduz até formar uma colônia visível. Cada colônia contém milhões de células idênticas, cada uma com muitas cópias de um único plasmídeo recombinante.

O problema de separar o gene de interesse de todo o outro DNA genômico agora está resolvido. Em média, uma de cada 750.000 colônias de bactérias deveria conter um gene β -globina intacto. Os métodos para fazer a triagem das colônias e identificar aquela que contém o gene β -globina são discutidos mais adiante deste capítulo.

Agora, também o problema de purificar quantidade suficiente de DNA é facilmente resolvido. Apenas em torno de 50 μ g de DNA genômico são necessários para construir a biblioteca genômica, quantidade que pode ser isolada de menos de 1 g de tecido ou células crescendo em cultivo. Desde que a colônia bacteriana contendo o gene β -globina tenha sido identificada, grandes quantidades de plasmídeo podem ser preparadas a partir de cultivos líquidos; 1 L de cultivo de *E. coli* na fase estacionária contém aproximadamente 5×10^{11} células, cada uma com mais de 500 cópias de plasmídeo recombinante. Multiplicando o número total de plasmídeo por litro, vezes a massa do gene β -globina ($1,27 \times 10^6$ dáltons), calcula-se que mais de 527 μ g de DNA de β -globina podem ser extraídos de cada litro de cultivo bacteriano.

Vetores melhores fazem bibliotecas melhores

Na "biblioteca de plasmídeo" descrita, estimou-se que o genoma humano inteiro poderia ser representado por aproximadamente 750.000 clones. Na verdade, uma biblioteca de 750.000 clones representaria apenas cerca de 90% do genoma, devido à probabilidade associada com qualquer fragmento particular de DNA tornar-se ligado em um plasmídeo. Por conseguinte, para aumentar a chance de que cada gene seja representado, uma biblioteca deve conter moléculas equivalentes a várias vezes a massa do genoma. Isso significa que uma biblioteca de plasmídeo que representa verdadeiramente o genoma inteiro deve ter na ordem de vários milhões de colônias. Entretanto, uma biblioteca desse tamanho cria problemas. Um problema é que o processo de triagem para encontrar um gene específico significa



Construindo uma biblioteca genômica

fazer a triagem de milhões de colônias. Outro é que muitos genes são muito maiores que o tamanho médio de 4.000 pb dos fragmentos de DNA nessa biblioteca. Por exemplo, o gene *FBN1* discutido no Capítulo 3 tem 110.000 pb (ou 110 quilobases ou kb). Se a média dos fragmentos de DNA nessa biblioteca é 4.000 pb, então o gene *FBN1* seria separado no mínimo em 28 fragmentos diferentes e suas colônias correspondentes. Uma vez que o objetivo é isolar um gene inteiro, foram desenvolvidos vários sistemas de vetores que permitiram a clonagem de fragmentos de DNA grandes.

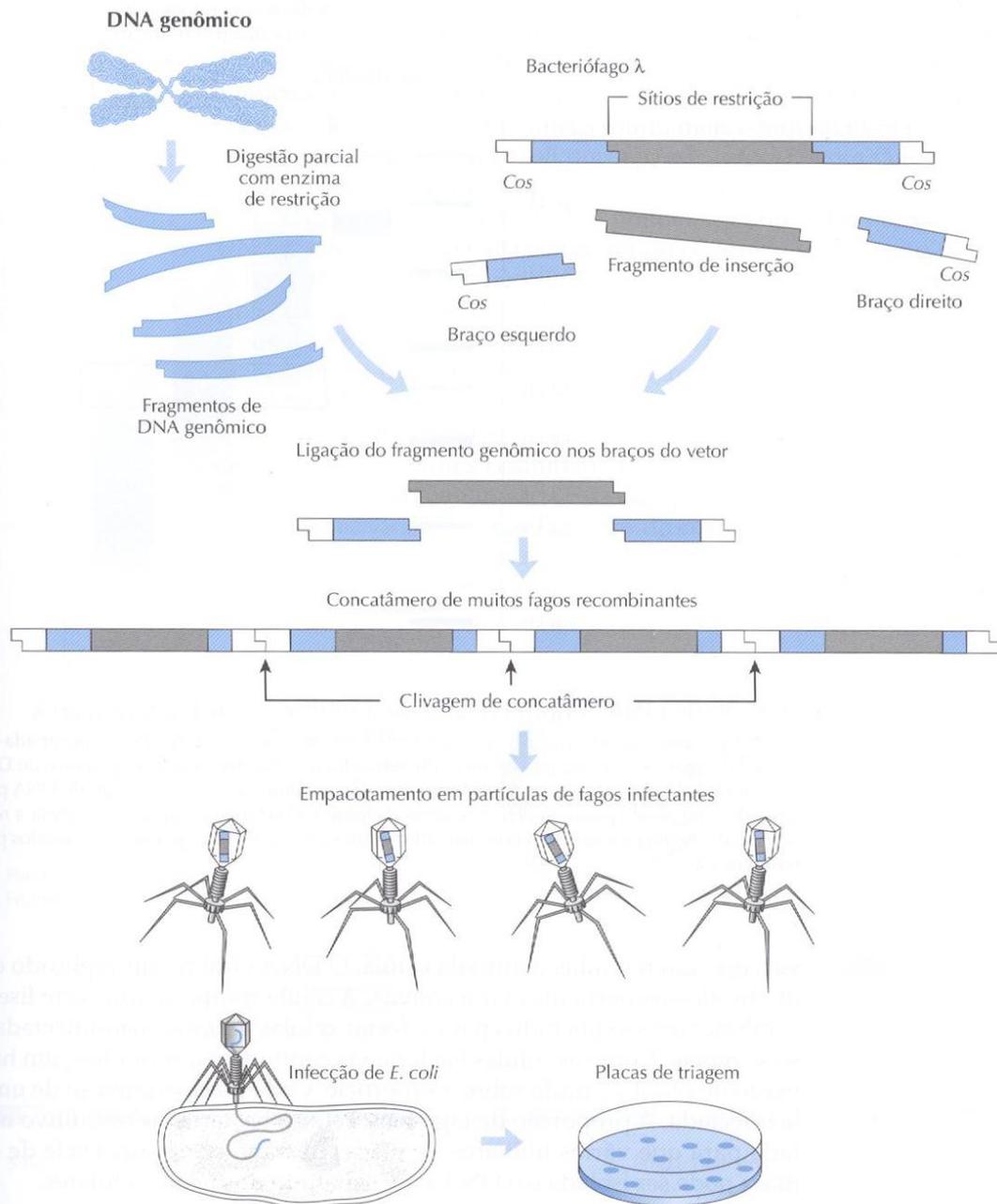
Bacteriófago λ . Os primeiros vetores desenvolvidos para clonar fragmentos grandes de DNA para bibliotecas não foram plasmídeos. Ao invés disso, eram formas modificadas de bacteriófago λ (ver Capítulo 1). O ciclo de vida natural de fagos faz deles vetores ideais para biólogos moleculares. Um fago infecta células por transferir seu material genético para a célula, na qual ele replica, produzindo centenas de novas cópias do genoma viral. Várias versões modificadas de bacteriófago λ foram desenvolvidas como vetores de clonagem nos quais genes virais podiam ser substituídos com um inserto de DNA exógeno.

O genoma do bacteriófago λ tem aproximadamente 50 kb de comprimento, além de conter mais de 40 genes. A cápsula de proteína do bacteriófago λ tem um arranjo justo sobre a quantidade de DNA que ficará dentro dela (~55 kb). Não há lugar para acrescentar material extra. Felizmente, uma quantidade de genes não-essenciais encontram-se no meio do genoma λ , enquanto que os genes essenciais encontram-se nas extremidades do genoma. Os genes não-essenciais podem ser removidos por digestão com enzimas de restrição e fragmentos de DNA genômico ou outro DNA pode ser ligado. O inserto de DNA escolhido é ligado no intervalo de fragmentos de extremidades do genoma λ , chamado de "braços" λ . As extremidades coesivas (*sticky*) de bacteriófago λ , chamadas de sítios *cos*, hibridizam umas às outras, criando um longo concatâmero composto de múltiplas moléculas de DNA λ recombinante. Quando esses concatâmeros são misturados em um tubo de ensaio com proteínas extraídas de bacteriófago λ , os concatâmeros são clivados em moléculas λ individuais e empacotados com eficiência em novos fagos virais.

O DNA genômico total é isolado com cuidado para prevenir a rompimento mecânico em fragmentos pequenos. O DNA genômico é digerido por uma enzima de restrição que também está presente no vetor fago como um sítio de clonagem. Ao contrário da digestão de restrição descrita para a biblioteca de plasmídeo, aqui o DNA é incubado rapidamente com uma quantidade limitada de enzima. Essa "digestão parcial" produz uma variedade de tamanhos de fragmentos, consideravelmente maior que o tamanho médio de 4.000 pb dos fragmentos que resultariam de uma digestão completa com uma enzima de restrição típica que reconhece seis nucleotídeos. (Lembrar do Capítulo 4: a probabilidade de tal sítio ocorrer pode ser calculada por $4 \times 4 \times 4 \times 4 \times 4 \times 4 = 4.056$.) Por fim, fragmentos de cerca de 20 kb de tamanho são selecionados para clonagem. Originalmente, tais fragmentos de 20 kb eram isolados por centrifugação do DNA em um gradiente de sacarose no qual fragmentos de diferentes tamanhos estabeleciam-se em níveis diferentes. Há pouco tempo, os fragmentos clonáveis são isolados pela separação dos fragmentos de restrição em um gel de aga-

Vetores de clonagem e número mínimo de clones necessários para representar o genoma humano haplóide

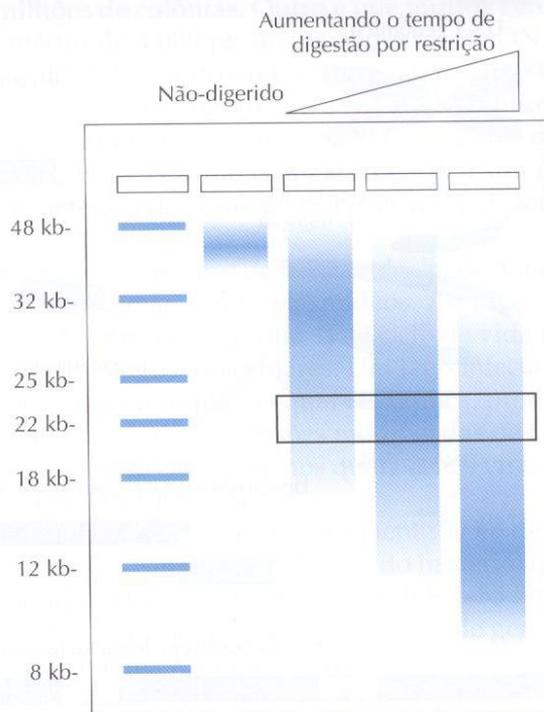
Vetor	Inserto médio (pb)	Clones para representar o genoma
Plasmídeo	4.000	750.000
Lambda	20.000	150.000
40.000	75.000	
Cromossomo bacteriano artificial (BAC)	150.000	20.000



Construindo uma biblioteca genômica com vetor bacteriófago λ

rose e depois corta-se uma fatia do gel que contém o DNA do tamanho apropriado, pela comparação com o marcador de tamanho de DNA. O DNA é extraído da fatia de agarose e ligado com o DNA do fago. Após a ligação do vetor λ com o DNA genômico, a mistura de ligação é adicionada a extratos contendo as proteínas necessárias para formar uma partícula de fago completa. Uma vez "empacotada", a biblioteca é utilizada para infectar a *E. coli*.

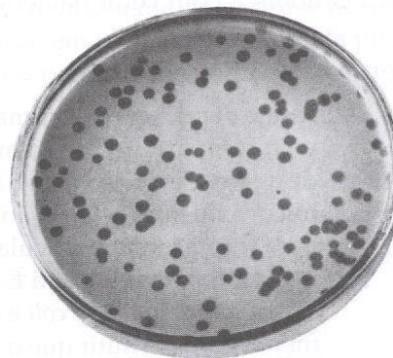
Tipicamente, a *E. coli* e a biblioteca de bacteriófago λ são incubadas juntas brevemente para permitir que o fago infecte a bactéria. Muitas vezes, há mais bactérias que fagos. A mistura é distribuída na superfície de placas com ágar. O fago injeta seu material genético nas células *E. coli* e, então, é replicado muitas vezes, além de codificar também as proteínas da cápsula necessária para empacotar as partículas virais no-



Seleção de DNA genômico para uma biblioteca de bacteriófago λ

O DNA genômico total é digerido parcialmente com uma enzima de restrição apropriada e separado em um gel de agarose. A primeira coluna é um marcador de DNA mostrando fragmentos de DNA de tamanhos conhecidos. A segunda coluna é DNA não-digerido. As últimas três colunas são de DNA parcialmente digerido e incubado por períodos crescentes de tempo. O retângulo interno representa a região do gel que contém os fragmentos de DNA com tamanho apropriado, que serão, por sua vez, usados para produzir a biblioteca λ .

vas, que são reunidas dentro da célula. O DNA viral recém-replicado é empacotado dentro dessas partículas virais novas. A célula rompe-se (ou sofre lise), os vírus recombinantes são liberados para infectar células vizinhas não-infectadas, e o processo se repete. Como as células bacterianas continuam a sofrer lise, um halo claro, chamado de placa, é criado sobre a superfície. Cada placa origina-se de uma única célula infectada. A proporção de fago para células bacterianas no cultivo original é ajustada para que vários milhares de placas formem-se na superfície de cultivo. Cada placa pode ser isolada e o DNA viral separado dos restos celulares.

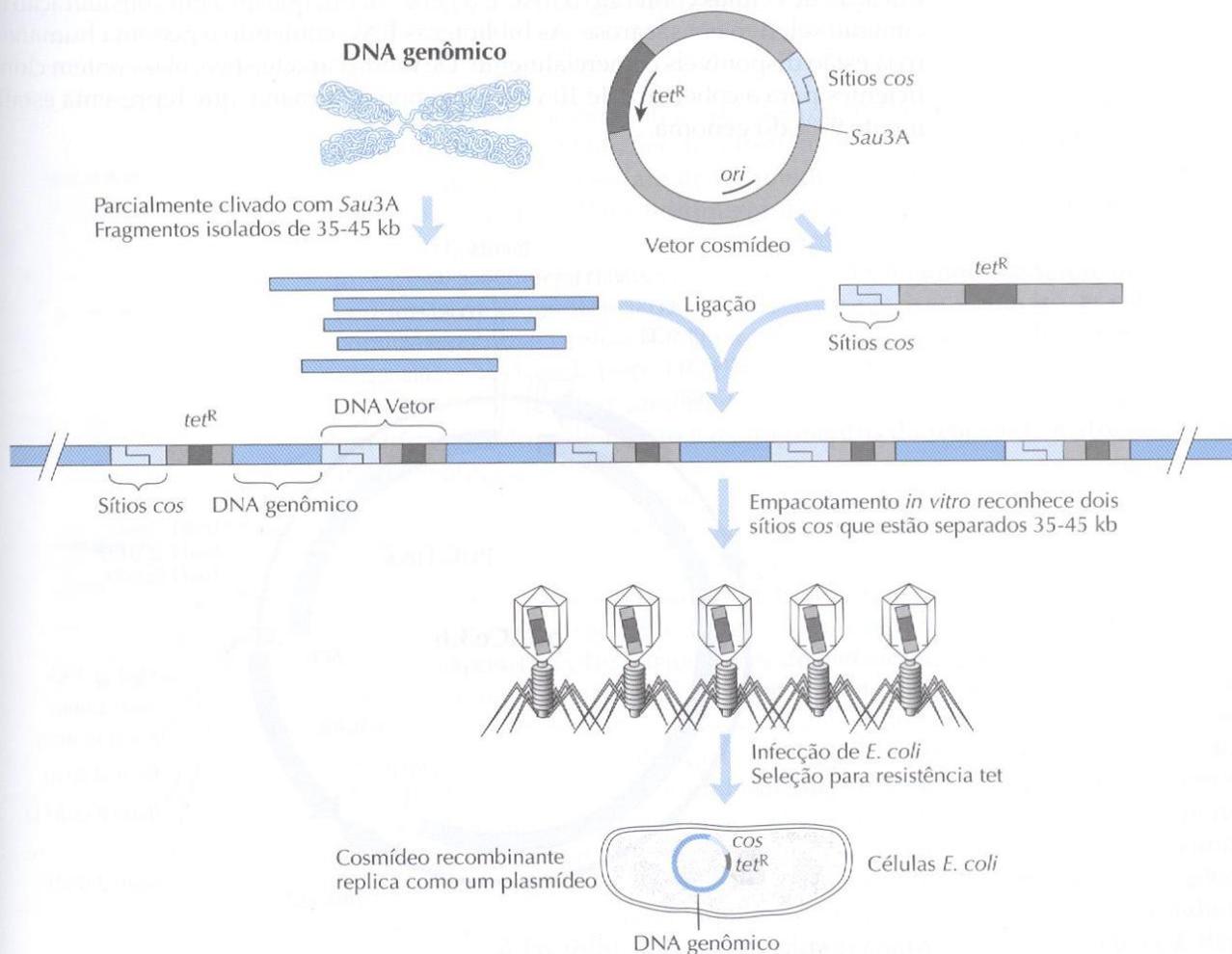


Placas de fago em uma sementeira de bactérias *E. coli*

(Reimpressa, com permissão, de Stent G.S. 1963. *Molecular biology of bacterial viruses*, p. 41. W.H. Freeman, San Francisco.)

Duas características de vetores λ promoveram o avanço na construção de bibliotecas de DNA genômico. Primeiro, caracteristicamente os vetores λ podem acomodar insertos que têm em média 20.000 pb (20 kb); portanto, um quinto das moléculas recombinantes é necessário para representar o genoma humano inteiro. Isso torna o trabalho de triagem da biblioteca muito mais controlável. Segundo, a introdução de DNA recombinante em *E. coli* por meio de infecção com fagos é muitas vezes mais eficiente que a transformação de *E. coli* com plasmídeos.

Cosmídeos. Os cosmídeos, que podem ser utilizados para clonar fragmentos de DNA tendo em média 40.000 pb, foram o avanço seguinte no desenvolvimento de vetores. Os referidos vetores híbridos são essencialmente plasmídeos que contêm os sítios *cos* do bacteriófago λ . Para criar um cosmídeo, um fragmento de DNA λ contendo o sítio *cos* é ligado em um plasmídeo contendo uma origem de replicação e um marcador capaz de ser selecionado (marcador selecionável). O DNA genômico é digerido em fragmentos de 40 kb e ligado em vetores cosmídeo. Os vetores são, então, misturados com extratos protéicos de bacteriófago λ , que reconhecem o sítio *cos* e empacotam os vetores cosmídeo em partículas de fago, criando vírus infecciosos. Tais vírus recombinantes são utilizados para infectar células *E. coli*. Uma vez que todos os genes virais foram removidos do cosmídeo, os novos vírus λ não são replicados – as células infectadas não são lisadas e não formam placas. Ao contrário, o cosmídeo replica como um plasmídeo dentro de uma célula bacteriana e as células infec-



Clonagem de cosmídeo

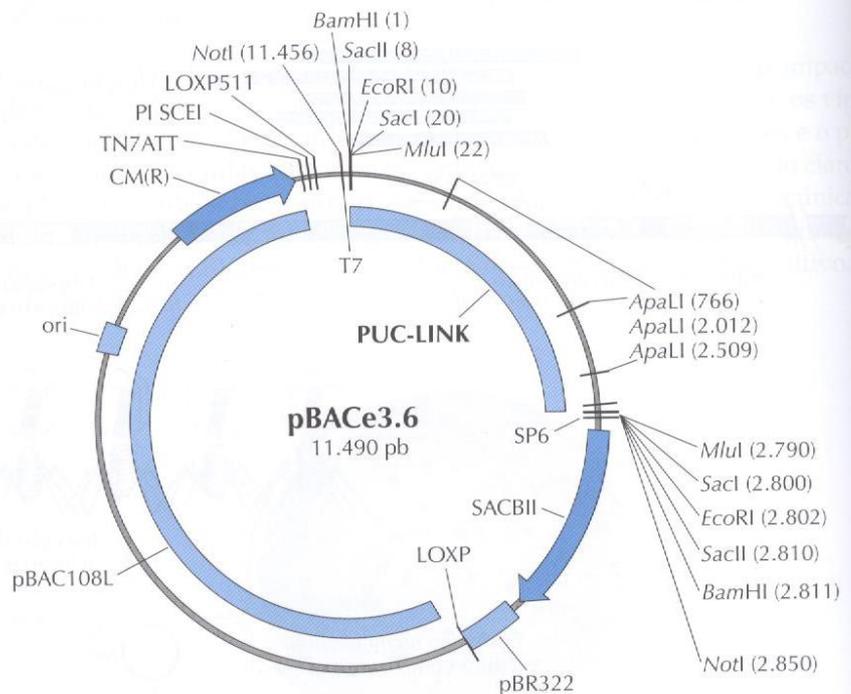
Nessa estratégia, são formados cosmídeos com insertos ligados *in tandem* que serão reduzidos a cosmídeos simples durante o empacotamento em partículas de fago.

tadas crescem em colônias bacterianas normais sobre meios seletivos. Os cosmídeos podem ser isolados com facilidade pelas técnicas-padrão de isolamento de plasmídeos (discutidas no Capítulo 4).

BACs e PACs. Cosmídeos são limitados como vetores pela quantidade de DNA que fisicamente pode caber dentro da cápsula do fago. Pelo mesmo motivo, o método de transformação química de Mandel e Higa são limitados para plasmídeos relativamente pequenos. A eletroporação, popularizada na década de 1990, criou um método para introduzir moléculas de DNA realmente grandes (100.000 a 300.000 pb) direto em *E. coli*. Moléculas de DNA com mais de um milhão de pb têm sido eletroporadas e não há limite teórico para o tamanho da molécula estudada por esse método. A eletroporação também eliminou a preparação de extratos complicados para empacotamento de fagos ou de células *E. coli* altamente competentes. A eletroporação mudou a estratégia no desenho de vetores. Os vetores utilizados atualmente para clonar fragmentos grandes de DNA precisam conter apenas uma origem de replicação bacteriana e um marcador selecionável.

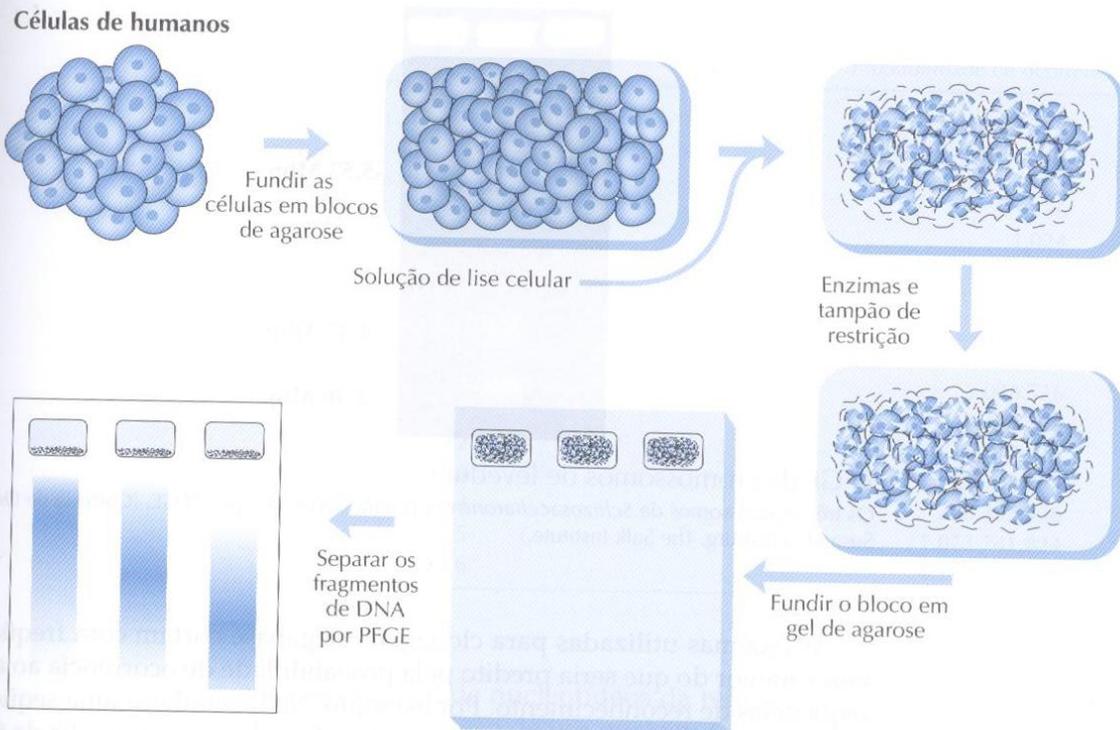
Surgiram dois tipos de vetores de clonagem que levariam em conta a clonagem de fragmentos de DNA de 100 a 300 kb: BACs (cromossomos artificiais de bactéria, do inglês *bacterial artificial chromosomes*) e PACs (cromossomos artificiais derivados de P1, do inglês *P1-derived artificial chromosomes*). Como o nome indica, BACs contêm uma origem de replicação do episossomo F da *E. coli* (descrito no Capítulo 3) e PACs encerra a origem de replicação do bacteriófago P1.

O pBACe3.6 é um vetor BAC típico. O marcador selecionável que permite a identificação de células contendo o BAC é o gene *SacBII*, que leva em consideração o crescimento seletivo em sacarose. As bibliotecas BAC contendo o genoma humano inteiro já estão disponíveis comercialmente. De modo característico, elas contêm clones suficientes para a cobertura de 10 vezes o genoma humano, que representa estatisticamente 99% do genoma.

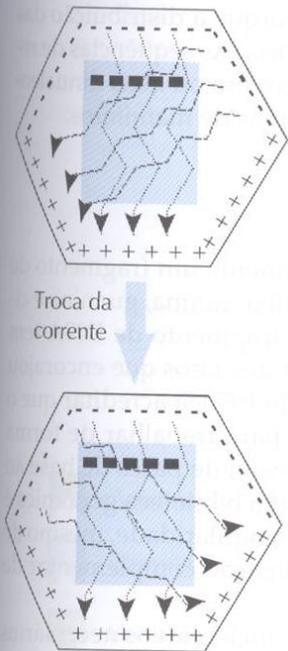


Mapa do plasmídeo BAC pBACe3.6

O gene *SacBII* codifica para levansacarose, a qual converte sacarose em levan que é tóxica para as células hospedeiras. O inserto genômico é colocado no meio de *SacBII*, dividindo-o e prevenindo a produção da proteína levansacarose completa. Os plasmídeos que não têm o inserto produzirão a proteína levansacarose, deixando de crescer no meio contendo sacarose.



Isolamento de fragmentos longos de DNA para clonagem de BAC

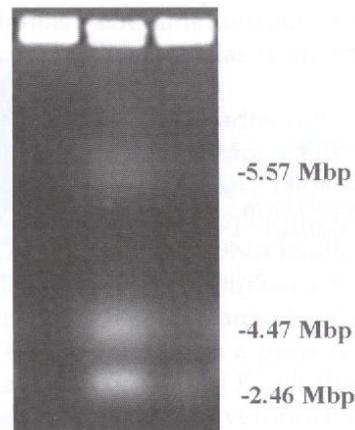


PFGE

Trocando a corrente entre os eletrodos laterais permite-se uma boa separação de fragmentos grandes de DNA.

Os métodos convencionais de isolamento de DNA cortam-no em fragmentos menores do que 100 kb. Sendo assim, a construção de bibliotecas "megabase" de BAC e PAC, que necessitam de fragmentos de DNA de 100 a 300 kb, exigem manipulação cuidadosa. Para minimizar a quebra do DNA, as amostras celulares são embebidas em pequenos blocos de agarose, cada um com volume de cerca de 100 μ l. Então, reações para liberar e digerir o DNA genômico são realizadas *in situ* pela difusão de reagentes dentro do bloco de agarose. Primeiro, as células são lisadas com detergente e proteases. Após uma lavagem, enzima de restrição e tampão são difundidos no bloco. Depois da incubação, EDTA é difundido para parar a reação de restrição e produzir uma digestão parcial. Um gel de agarose maior é fundido em torno do bloco de reação e os fragmentos da reação de restrição são separados por eletroforese.

A eletroforese convencional (de campo fixo) não pode separar fragmentos de DNA maiores que 50.000 pb. No entanto, várias técnicas novas podem separar fragmentos de acima de 10 milhões de pares de base de comprimento. Na eletroforese em gel de campo pulsátil (PFGE, do inglês *pulse-field gel electrophoresis*), utilizada em primeiro lugar por Charles Cantor, na Universidade de Columbia, a corrente é alternada entre vários pares de eletrodos opostos. A cuba de eletroforese é montada em um hexágono, com eletrodos na parte de cima e de baixo do gel e dois grupos de eletrodos em cada lado do gel em um ângulo de 120°. A corrente é aplicada de maneira constante entre o eletrodo de cima e o de baixo. A corrente aplicada a partir dos eletrodos laterais é trocada de um lado para o outro entre os dois grupos. Isso altera a trajetória do DNA enquanto ele migra pelo gel, permitindo a melhor separação de fragmentos grandes. Na eletroforese em gel de campo invertido ou de campo com alternância ortogonal (OFAGE, do inglês *orthogonal-field-alternation gel electrophoresis*), uma corrente longa para frente é alternada com uma corrente reversa curta. A capacidade de resolução de PFGE e OFAGE depende da duração de tempo que as duas correntes alternadas são aplicadas. Fatias de gel contendo fragmentos de DNA do tamanho apropriado são isolados suavemente e ligados em vetores BAC ou PAC.



PFGE de cromossomos de levedura

Os três cromossomos de *Schizosaccharomyces pombe* separados por PFGE. (Cortesia de Debbie T. Liang e Susan L. Forsburg, The Salk Institute.)

As enzimas utilizadas para clonagem megabase cortam com freqüência 15 a 75 vezes menor do que seria predito pela probabilidade de ocorrência ao acaso de suas seqüências de reconhecimento. Por exemplo, *NotI* reconhece uma seqüência de 8 pb e seria esperado que produzisse fragmentos de restrição com média de 65.536 pb (4^8). De fato, ela produz fragmentos com média de um milhão de pares de base. *MluI*, *NruI* e *PvuI* reconhecem seqüências de seis nucleotídeos com probabilidade de ocorrência ao acaso de 4.096 nucleotídeos (4^6). Contudo, cada uma produz fragmentos de restrição com média de 300.000 nucleotídeos. Isso ocorre porque a distribuição das seqüências de nucleotídeos no genoma humano não é aleatória. As seqüências de reconhecimento de todas as enzimas de corte megabase contêm um ou mais dinucleotídeos CG, uma seqüência que se torna muito rara em genomas de mamíferos.

Fazendo a triagem de bibliotecas de DNA

Tem sido discutida a dificuldade de distinguir bioquimicamente um fragmento de DNA de outro. Uma vez que a abundância relativa de adenina, timina, guanina e citosina é constante dentro de um polímero de DNA, cada fragmento de DNA tem mais ou menos a mesma composição química. Esse foi um dos fatos que encorajou cientistas, durante a década de 1940 e o início da década de 1950, a acreditar que o DNA era incapaz de ser o material genético. No entanto, para trabalhar de forma criativa com ele, deve-se pensar em termos da seqüência exata de pares de base ao longo de um fragmento específico do ácido. A triagem de uma biblioteca genômica é feita com base em dois atributos da molécula de DNA: (1) a singularidade, mesmo de fragmentos relativamente curtos de nucleotídeos, e (2) a ligação complementar de pares de nucleotídeos dentro de uma seqüência.

É preciso considerar o comprimento de uma cadeia de nucleotídeos necessários para ser uma seqüência única de DNA humano – uma seqüência que estatisticamente seria encontrada apenas uma vez no genoma humano inteiro. Para ser única, a seqüência deve ter uma probabilidade de ocorrer menor que uma vez em 3 bilhões de nucleotídeos.

Para o propósito dessa discussão, será considerado que, de um modo geral, as quatro bases estão distribuídas aleatoriamente em uma molécula de DNA. (Para simplificar, uma seqüência de DNA normalmente leva em conta apenas o arranjo de nucleotídeo de uma fita porque a fita complementar oposta pode ser deduzida.

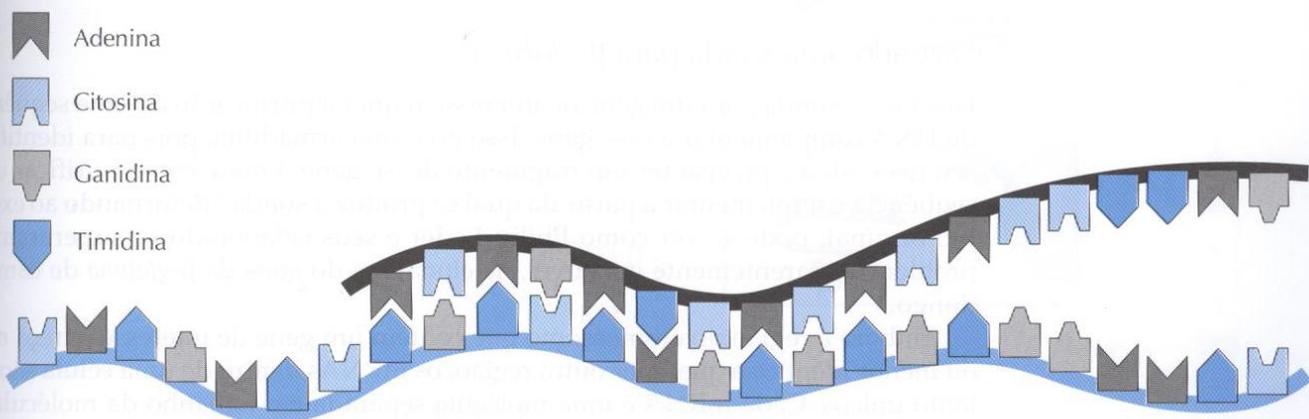
A probabilidade de qualquer seqüência de DNA ocorrer pode ser calculada pela fórmula $1/4^n$, na qual 4 é igual ao número de nucleotídeos diferentes e n é igual

Probabilidade de ocorrência de seqüências de DNA de número crescente de nucleotídeos

Nucleotídeos	Probabilidade de ocorrência (1/)
1	4
2	14
3	64
4	256
5	1.024
6	4.096
7	16.384
8	65.536
9	262.144
10	1.048.576
11	4.194.304
12	16.777.216
13	67.108.864
14	268.435.456
15	1.073.741.824
16	4.294.967.296

ao número total de nucleotídeos da referida seqüência. Por exemplo, começando em qualquer posição de um nucleotídeo ao longo da fita de DNA, há uma chance em quatro ($1/4$) de que este será adenina. Há também uma chance em quatro ($1/4$) de que o nucleotídeo vizinho será citosina. Portanto, a chance combinada de encontrar a seqüência de dinucleotídeo adenina-citosina (AC) é $1/4 \times 1/4 = 1/4^2 = 1/16$. Todas as outras seqüências de dinucleotídeo também teriam uma probabilidade de ocorrência de 1 em 16, AA, TT, GG, CC, AT, AG, TA, TG, TC, GA, GT, GC, CA, CT e CG.

Um exercício simples com uma calculadora mostra que a chance de encontrar uma seqüência específica de 10 nucleotídeos é de $1/4^{10}$ ou de 1 em 1.048.576. Estatisticamente, apenas 3.000 cópias de uma seqüência específica de 10 nucleotídeos deveriam ser encontradas no genoma humano. Uma seqüência de 15 nucleotídeos tem uma probabilidade de ocorrência de $1/4^{15}$ ou de um em 1.073.741.824, enquanto uma seqüência de 16 nucleotídeos tem uma probabilidade de ocorrência de $1/4^{16}$ ou de um em 4.294.967.296. Uma seqüência de 15 nucleotídeos, por sua vez, deveria ocorrer três vezes no genoma humano, enquanto uma seqüência de 16 nucleotídeos deveria ocorrer apenas uma vez.



Hibridização

A sonda radioativa (preto) encontra sua seqüência complementar no DNA genômico (azul) e estabelece pontes de hidrogênio, formando uma molécula duplex.

Esse cálculo sugere a regra geral que uma seqüência de 16 nucleotídeos ou uma seqüência mais longa provavelmente seja única no genoma humano. No entanto, deve-se lembrar que seqüências de DNA repetitivo compõem uma percentagem significativa dos cromossomos humanos. Obviamente, uma seqüência de 16 nucleotídeos dentro de um elemento repetitivo seria comum a cada uma das unidades repetitivas. Algumas seqüências repetitivas longas incluem as centenas de cópias de genes de RNA ribossomal, porém, em geral, seqüências repetitivas são curtas – variando de dois a poucas centenas de nucleotídeos de comprimento.

Identificando seqüências específicas de DNA

Em 1961, Sol Spiegelman e Benjamin Hall descobriram que DNA fita simples estabeleceria pontes de hidrogênio com sua seqüência de RNA complementar para formar uma molécula fita dupla (duplex) estável. A mesma ligação pode também ser realizada entre moléculas de DNA fita simples. Esse emparelhamento complementar entre bases proporciona uma ferramenta poderosa para a investigação de seqüências de DNA únicas na biblioteca genômica.

A incubação de DNA fita dupla a uma temperatura acima de 90°C, em um pH maior que 10,5 ou com vários compostos orgânicos (tais como uréia ou formamida), rompe as pontes de hidrogênio entre os pares de base, causando a dissociação das fitas complementares. Esse processo é chamado de desnaturação. Sob condições adequadas de sal, temperatura e pH, as duas moléculas de fita simples podem ser unidas novamente para formar a molécula de DNA duplex original. Esse processo de alinhamento de moléculas complementares fita simples e formação de moléculas fita dupla é conhecido como *hibridização*. (Anelamento é o termo aplicado para hibridização de seqüências de DNA curtas.)

Sob condições de reação de “alta estringência”, DNAs duplex estáveis formam-se apenas quando o emparelhamento das bases complementares é totalmente perfeito ao longo do comprimento inteiro das fitas de DNA. Sob condições de “baixa estringência”, ocorre hibridização parcial entre as fitas que têm graus menores de complementaridade. Ambas as condições são úteis na triagem de bibliotecas.

A hibridização é o método usado para identificar e isolar uma colônia ou uma placa contendo um gene específico de uma biblioteca. A idéia é criar uma “sonda”, normalmente um fragmento de DNA fita simples marcado radioativamente, que contenha uma seqüência complementar ao gene de interesse e, depois, incubá-lo com o DNA genômico desnaturado. A sonda é adicionada em excesso, de forma que é mais provável que o gene hibridizará a sonda do que a sua fita complementar original.

Fazendo uma sonda para β -globina

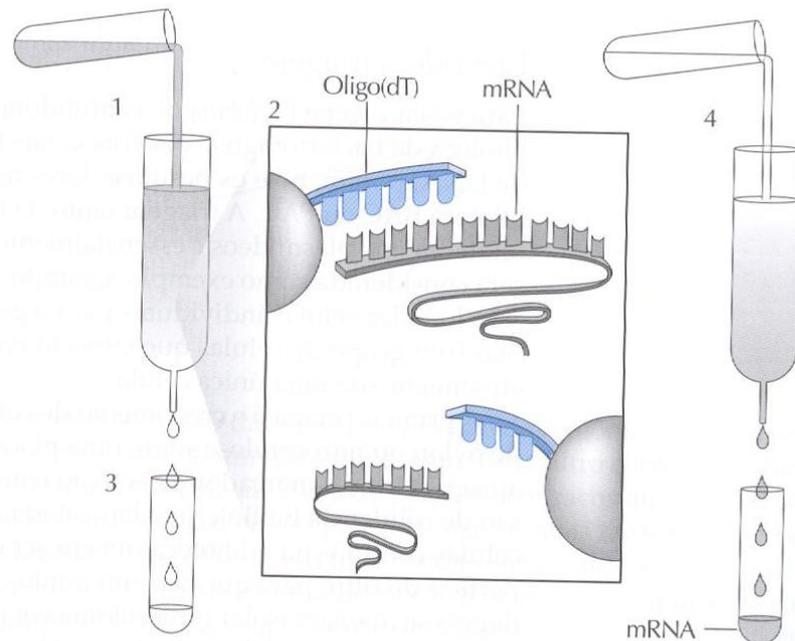
Fazer uma sonda para um gene de interesse requer a purificação de uma seqüência de DNA complementar a esse gene. Isso cria uma armadilha, pois para identificar seu gene-alvo é preciso ter um fragmento desse gene. Então, como purificar uma seqüência complementar a partir da qual se produz a sonda? Retornando ao exemplo original, pode-se ver como Philip Leder e seus colaboradores superaram tal problema, aparentemente insolúvel, na clonagem do gene de β -globina de camundongo.

Embora a região do cromossomo que contém um gene de interesse pareça mais ou menos idêntica a qualquer outra região, os mRNAs dentro de uma célula são um tanto únicos. Cada mRNA é uma molécula separada e o tamanho da molécula de mRNA reflete o tamanho do gene. Dessa forma, mRNAs diferentes podem ser separados pelo tamanho. O mRNA é complementar à seqüência de DNA do gene a par-

tir do qual é transcrito; e ele hibridizará de imediato essa sequência de DNA. Leder pensou que ele poderia tirar vantagem da hibridização, em combinação com as propriedades únicas das células vermelhas do sangue (hemácias) e do mRNA eucariótico, para fazer uma sonda para o gene β -globina.

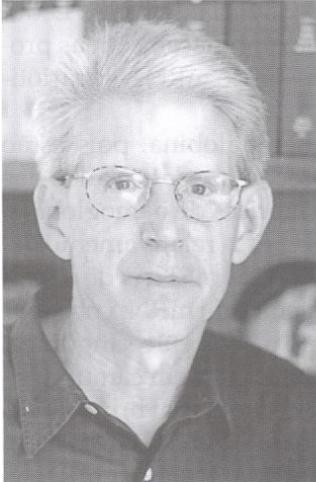
As hemácias podem ser imaginadas como sacos de hemoglobina, pois estão preenchidas com essa proteína e poucas mais. Assim, os precursores nucleados das hemácias, os reticulócitos, funcionam essencialmente como fábricas de hemoglobina. Enquanto a maioria das células de mamíferos contém quantidades relativamente pequenas de milhares de tipos diferentes de moléculas de mRNA, os mRNAs de α - e β -globina são responsáveis pela maior porção de todas as moléculas de mRNA nos reticulócitos. Os reticulócitos não são encontrados normalmente na corrente sanguínea, mas podem abranger mais de 50% das células sanguíneas circulantes em camundongos com anemia induzida. Sendo assim, os reticulócitos foram o material bruto a partir do qual Leder e seus colaboradores criaram uma sonda para o gene β -globina.

Isolando de mRNA de globina. Conforme descrito no Capítulo 3, alguns mRNAs eucarióticos contêm uma extensão longa de 100 a 200 resíduos de adenina nas suas extremidades 3'. Tal fragmento poli(A) torna possível isolar mRNA em uma etapa usando cromatografia de afinidade em oligo(dT)-celulose. Nesse protocolo, uma coluna é preenchida com partículas de celulose, as quais são ligadas a seqüências sintéticas pequenas (10 a 20 nucleotídeos) de deoxitimina (dT). A extensão de resíduos T hibridizará a extensão de A na seqüência poli(A) de mRNA sob condições de alta concentração de sal. Isso irá permitir a captura de mRNA poli(A) quando ele se liga à coluna. Quando um lisado de reticulócitos é passado por uma coluna de oligo(dT)-celulose em um tampão de alta concentração salina contendo cloreto de sódio, as seqüências poli(A) do mRNA ligam-se aos resíduos T complementares. Assim, as moléculas de mRNA são retidas na coluna enquanto todas as outras moléculas passam por meio dela. Depois, o sal é enxaguado da coluna, que libera o mRNA ligado. Uma vez que α e β -globina representam 98% dos mRNAs contidos nos reticulócitos, não é necessária uma purificação adicional.



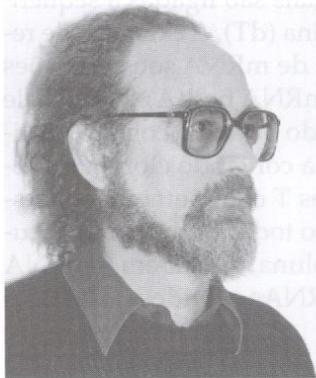
Purificando mRNA de globina

(1) Reticulócitos lisados em tampão com alta concentração de sal são derramados por meio de uma coluna de oligo(dT)-celulose. (2) As seqüências poli(A) do mRNA ligam-se às moléculas de oligo(dT) nas partículas de celulose. (3) Moléculas diferentes de mRNA passam pela coluna. (4) A lavagem com tampão de baixa concentração salina libera o mRNA da coluna.



Tom Maniatis

(Cortesia de Kris Snibbe,
Harvard News Office.)



Argiris Efstratiadis

Fazendo uma sonda de cDNA a partir de mRNA. O Capítulo 2 discutiu o dogma central da biologia molecular – o DNA é transcrito para fazer RNA que é traduzido para fazer proteína. No entanto, deve-se lembrar que existem algumas exceções para essa regra. A fim de fazer a sonda para triagem de uma biblioteca, pode-se levar vantagem de uma dessas exceções para fazer DNA a partir de RNA. Para tanto, é necessária uma DNA polimerase especial chamado de DNA polimerase dependente de RNA ou transcriptase reversa (RT, do inglês *reverse transcriptase*).

A RT é encontrada em *retrovírus*, que armazena seu material genético na forma de RNA. Quando um retrovírus infecta uma célula, ele deve converter seu genoma RNA em DNA para começar seu ciclo de vida. A RT do retrovírus utiliza o RNA como um molde para construir uma nova fita de DNA complementar (essa nova fita é denominada “cDNA”), que é convertida em uma molécula de DNA fita dupla.

Na metade da década de 1970, vários grupos descobriram métodos para utilizar RT em ensaios para fazer cDNA a partir de mRNA. Os métodos utilizados hoje são derivados do trabalho realizado em Harvard por Tom Maniatis, Argiris Efstratiadis, Fotis Kafatos e Allan Maxam. Utilizando RT em ensaios, uma sonda de cDNA marcada radioativamente pode ser criada a partir de mRNA de globina purificado. Primeiro, um primer curto composto de oligo(dT) é hibridizado à seqüência poli(A) 3' do mRNA purificado. Como todos os DNAs polimerases, a RT inicia a síntese de DNA no primer. Então, ela constrói uma cadeia de DNA que incorpora um nucleotídeo complementar para cada posição na molécula de mRNA. Como é verdade, para todas as polimerases (tanto DNA como RNA polimerases) a RT sintetizada extremidade 5' para 3' – adicionando sempre um nucleotídeo novo à OH 3' do nucleotídeo adicionado anteriormente. A síntese da sonda de cDNA ocorre em uma mistura de reação contendo RT e os quatro deoxinucleotídeos trifosfatos, um ou mais dos quais contém P³² (um isótopo radioativo do fósforo).

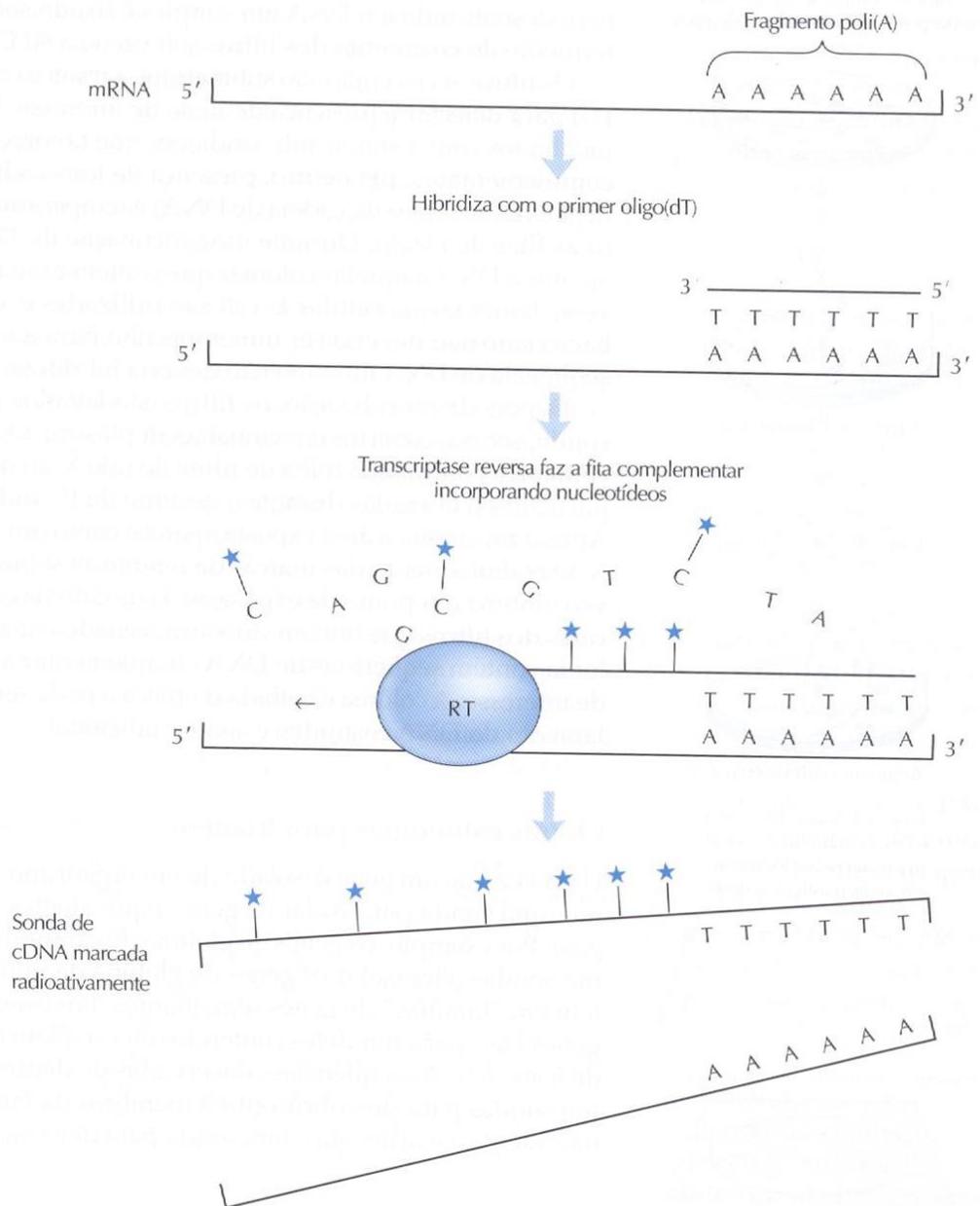
Leder utilizou essa reação para produzir sondas de cDNA marcadas com P³² a partir de mRNA de α - e β -globina que ele isolou de reticulócitos. Como mRNAs para α e β -globina estavam presentes, a triagem da biblioteca identificou colônias contendo ambos os genes. A caracterização adicional dos clones isolados permitiu que ele distinguísse entre os dois genes de globina.

Fazendo a triagem

Para isolar o gene β -globina de camundongo, Leder e colaboradores triaram uma biblioteca de bacteriófago λ . Na época, tais bibliotecas continham os maiores insertos de DNA, porém, hoje os pesquisadores muito provavelmente iniciam com uma biblioteca BAC ou PAC. A triagem tanto da biblioteca de bacteriófago λ quanto das bibliotecas tipo plasmídeos é essencialmente a mesma, assim, a biblioteca de cosmídeo será considerada como exemplo. Quando se faz a triagem de uma biblioteca, não são selecionadas células individuais para o gene, mas, ao contrário, selecionam-se colônias (um grupo de células que crescem em uma superfície sólida e que resultam do crescimento de uma única célula).

A primeira etapa é o crescimento de colônias. Inicia-se pela colocação de um filtro de nylon ou nitrocelulose sobre uma placa de Petri com agar contendo os nutrientes apropriados e o marcador de seleção como, por exemplo, ampicilina. Uma suspensão de células da biblioteca é depositada sobre o filtro e espalhada uniformemente. As células contidas na biblioteca devem ser diluídas e espalhadas bastante sobre a superfície do filtro para que haja um mínimo de sobreposição de colônias (lembrar que depois se desejará isolar uma colônia em particular). Caracteristicamente, vários milhares de células são depositados na placa e os filtros múltiplos são preparados a fim de fazer a triagem de insertos suficientes para cobrir o genoma várias vezes a mais. Após a incubação durante a noite, colônias aparecem sobre os filtros.

O processo de triagem de biblioteca requer a destruição das células sobre o filtro. Para se obter uma colônia viável após a triagem, um grupo idêntico de filtros (dupli-

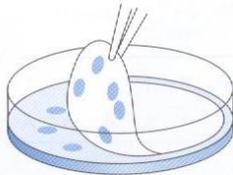
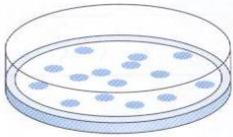


Sintetizando cDNA

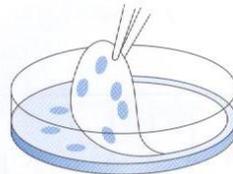
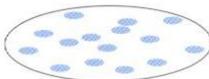
cata) deve ser feito por meio da réplica da posição das colônias de cada filtro. Isso é realizado pela colocação de um filtro novo sobre cada filtro contendo colônias. Marcas de registro são feitas pela abertura de pequenos buracos através das bordas dos filtros, de modo que os filtros possam ser realinhados mais tarde e, então, separados. Algumas células de cada colônia são transferidas para o filtro novo, o qual é colocado em uma placa nova, incubando-se ambas as placas até que apareçam colônias de tamanho completo. (Essa é a adaptação de uma técnica para fazer replicação de colônias bacterianas, publicada em 1952 por Joshua e Eswther Lederberg, na Universidade de Wisconsin.)

Um grupo-réplica de filtros é deixado na placa de Petri e depois refrigerado. O outro filtro-réplica é utilizado para triagem. Os filtros são molhados em uma solução alcalina de hidróxido de sódio, que simultaneamente faz a lise das células e desnatu-

Colocar células sobre um filtro sobreposto em uma placa de ágar

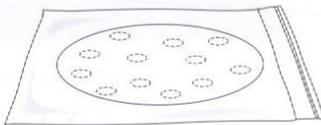


Fazer um filtro réplica

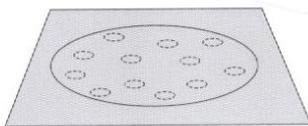


Remover o filtro réplica

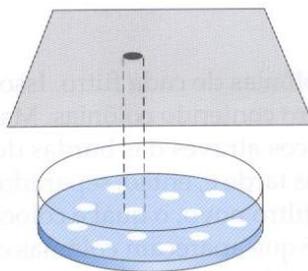
Lisar as células; desnaturar o DNA e hibridizar a sonda marcada radioativamente em sacos plásticos vedados



Sobrepôr o filtro com um filme de raio X



A revelação do auto-radiograma mostra a localização da colônia que a sonda hibridizou



Fazendo a triagem de uma biblioteca de cosmídeo

Uma auto-radiografia indica a localização de colônias em uma placa-padrão hibridizada por uma sonda radioativa.

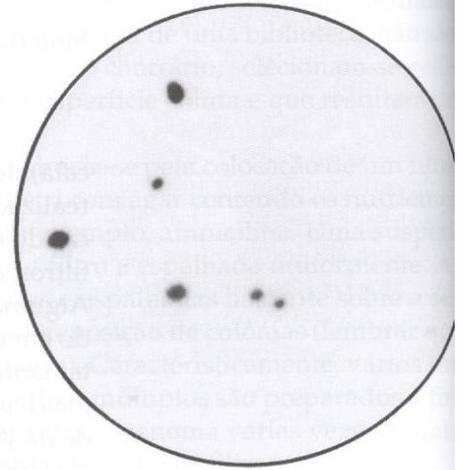
ra o DNA. Após, os filtros são molhados em tampão em pH neutro. O DNA permanece desnaturado e o DNA fita simples é fixado sobre o nylon/nitrocelulose por intermédio do cozimento dos filtros sob vácuo a 80°C.

Os filtros secos então são submetidos à triagem com a sonda (cDNA marcado com P^{32}) para detectar a presença do gene de interesse. Para conseguir isso, os filtros são incubados com a sonda sob condições que favoreçam a hibridização de seqüências complementares: pH neutro, presença de ions sódio (que neutralizam as alterações negativas ao longo da cadeia de DNA) e temperatura elevada (65°C, que desembaraça as fitas de DNA). Durante uma incubação de 12 a 24 horas, a sonda hibridizará apenas o DNA daquelas colônias que contêm todo ou parte do gene de interesse. Deve-se lembrar que células *E. coli* são utilizadas e, desse modo, a presença do DNA bacteriano não deveria ser um empecilho para a triagem – a sonda para uma única seqüência de DNA humano não deveria hibridizar o DNA bacteriano.

Depois da hibridização, os filtros são lavados para que se remova o excesso de sonda, secos e cobertos em embalagem plástica. Os filtros cobertos são pressionados com força contra uma folha de filme de raio X, ao qual é exposto por várias horas. As partículas β liberadas durante o decaimento do P^{32} radioativo na sonda revelam o filme. Após a revelação, a área exposta aparece como um ponto escuro sobre o filme de raio X. O realinhamento das marcas de referência sobre o filme, o filtro e a placa de cultivo combina um ponto de exposição à sua colônia correspondente em uma das duplicatas dos filtros que tinham sido armazenados no refrigerador. Sendo assim, essa colônia contém seqüências de DNA complementar à sonda e, portanto, contém o gene de interesse. A colônia é colhida da placa e pode ser crescida em quantidade para isolamento do DNA cosmídeo e análise adicional.

Outras estratégias para triagem

Uma vez que um gene é isolado de um organismo, muitas vezes ele poderá servir como uma sonda para isolar os genes equivalentes de organismos diferentes (ortólogos). Por exemplo, os genes de globina de camundongo e de coelho são utilizados como sondas para isolar os genes de globina de humanos. Muitos genes também existem em "famílias" de genes semelhantes. Por exemplo, no Capítulo 3, discutiu-se os genes *Hox*, cada um deles contendo uma seqüência altamente conservada, chamada de *homeobox*. As seqüências conservadas de dentro do *homeobox* foram utilizadas como sondas para descobrir outros membros da família *Hox*. Quando uma seqüência não-idêntica é utilizada como sonda para detectar outro gene, o processo é chamado



Hibridização de colônias

Placa de LB com colônias transformadas (esquerda). Filtro com clones positivos (direita).

de “hibridização cruzada”. Condições de menor estringência (temperatura baixa e concentração de sal alta) são utilizadas em triagem por hibridização cruzada, permitindo a ocorrência de pontes de hidrogênio, apesar de poucas diferenças entre a sonda e a seqüência-alvo.

Oligonucleotídeos Sintéticos. O progresso na bioquímica de proteínas, com o começo da era da clonagem do gene, resultou em que muitas proteínas tenham sido total ou parcialmente seqüenciadas. O advento dos sintetizadores automáticos de DNA, no início dos anos 80, tornou factível a utilização de praticamente qualquer seqüência de aminoácidos para produzir sondas curtas de oligonucleotídeos (oligo = alguns) para o gene correspondente.

Uma vez que cada aminoácido é codificado por um códon, uma seqüência de seis aminoácidos pode ser utilizada para derivar uma sonda de DNA de 18 nucleotídeos, que deveria hibridizar a uma única seqüência no genoma humano. (Lembrar o cálculo para o comprimento de uma seqüência única.) No entanto, a maioria dos aminoácidos é codificada por mais de um códon. Por causa dessa degeneração, ou “hesitação” do código genético, a seqüência de DNA extraída não pode ser precisamente derivada de uma seqüência de aminoácidos.

Esse problema pode ser evitado pela seleção de uma seqüência de aminoácidos com a mínima degeneração – os ideais são aqueles codificados por apenas um ou dois códons diferentes. Então, são sintetizados oligonucleotídeos que representam cada combinação de códon possível. A mistura de sondas é marcada radioativamente usando ATP P^{32} e polinucleotídeo quinase, que transfere o fosfato γ do ATP para a extremidade 5' do oligonucleotídeo. A sonda é utilizada para a triagem de uma biblioteca genômica ou de cDNA, enquanto que a seqüência completa de aminoácidos da proteína é predita a partir da seqüência do DNA do gene clonado.

Fazendo a Triagem de Biblioteca BAC e PAC. Uma vez que bibliotecas BAC e PAC contenham fragmentos bastante grandes de DNA, é necessária a triagem de muito poucos clones para encontrar um gene em particular. Caracteristicamente, a triagem dessas bibliotecas não é feita por propagação de células em placas de cultivo. Ao invés disso, colônias individuais de BAC ou PAC são propagadas primeiro em placas microtitulação com 384 poços. Um braço mecânico autômato espalha as colônias em um filtro de alta densidade. Depois, é feita a triagem do filtro utilizando métodos padrão de hibridização. A posição das colônias positivas sobre o filtro são correlacionadas com os poços específicos na placa microtitulação.

Uma vez que um clone genômico é identificado como contendo um gene-alvo, ele deve estar localizado dentro de uma BAC ou PAC muito grande. A abordagem mais fácil é digerir a BAC com uma endonuclease de restrição e “subclonar” tais fragmentos em outro vetor como, por exemplo, um plasmídeo ou um bacteriófago λ . Essa minibiblioteca é submetida a uma segunda triagem com a mesma sonda usada para identificar a BAC. Hoje, pelo fato de o seqüenciamento de DNA ser tão rápido e barato, normalmente os subclones são seqüenciados diretamente para identificar o gene-alvo.

Arg-Lys-Met-Val-His-Asn-Cys-Trp-Gly-Leu-Leu-Met-Ser-Gln-Pro-Tyr

ATG GTA CAC AAC TGCTGG

C T T T

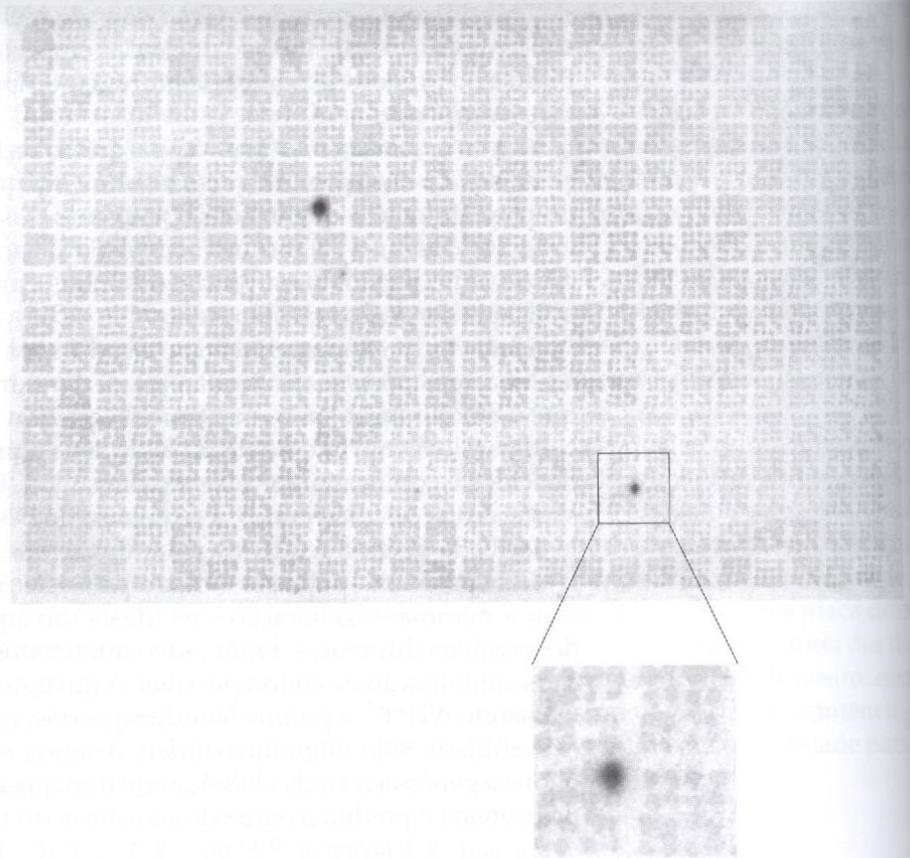
G

T

$4 \times 2 \times 2 \times 2 \times 1 = 32$

Escolhendo a melhor seqüência para uma sonda

Na seqüência protéica hipotética aqui apresentada, foi feita a triagem para uma região na qual os aminoácidos têm o mínimo de degeneração ou hesitação. A seqüência sublinhada tem poucos aminoácidos degenerados; e uma mistura de 32 sondas considerará cada seqüência que pode ser codificada.



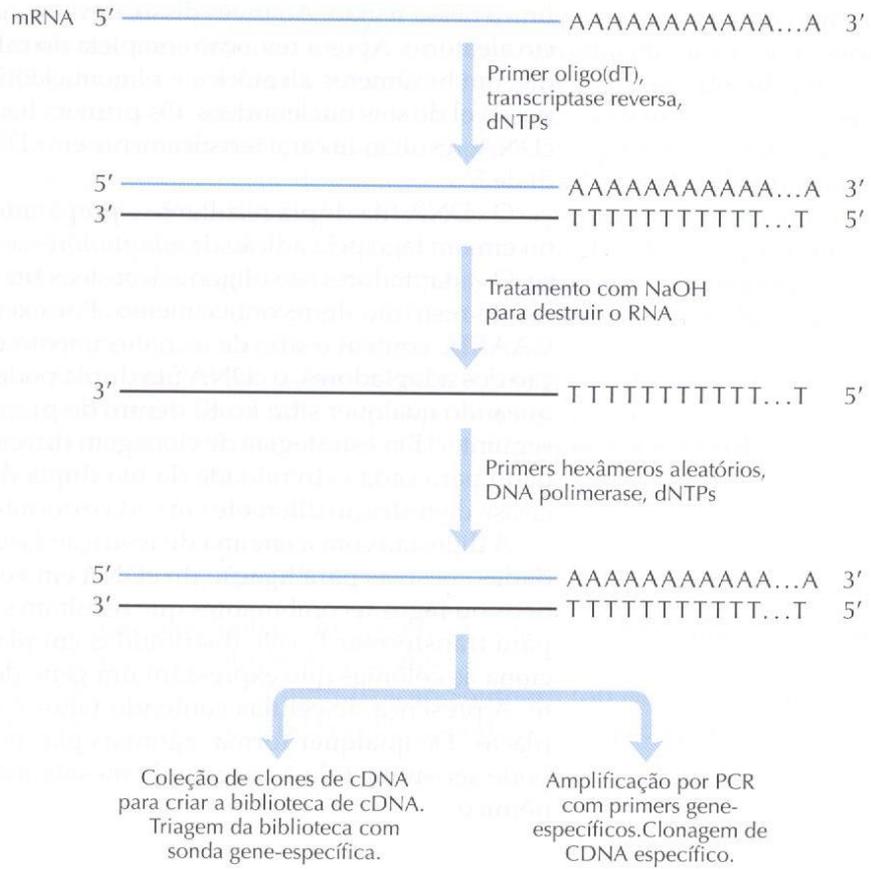
Hibridização de colônias

Placa de LB com colônias transformadas (*esquerda*). Filtro com clones positivos (*direita*).

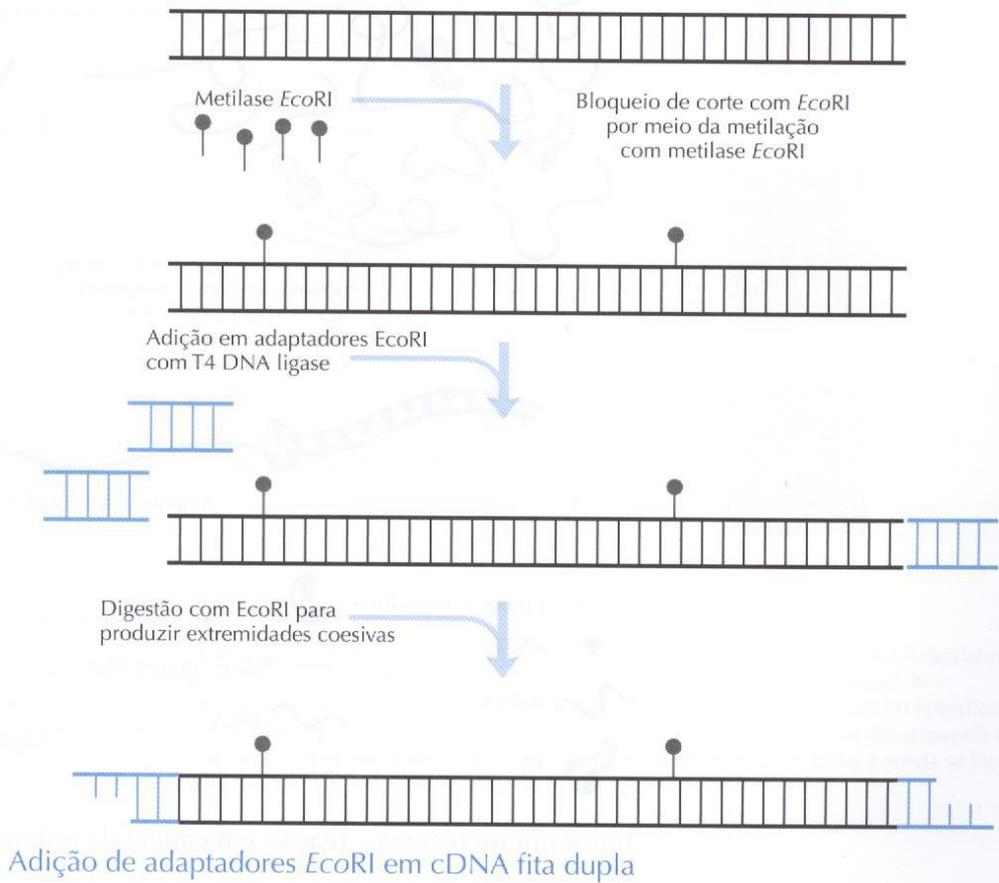
Biblioteca de cDNA. Até agora, tem sido descrita a clonagem de DNA genômico. Deve-se lembrar que muitas vezes uma seqüência de DNA genômico caracteristicamente contém tanto a seqüência íntron quanto a seqüência éxon. Isso significa que um único gene pode se estender por uma grande distância, dificultando seu isolamento e sua análise. Outro problema com a triagem de uma biblioteca genômica é que ela contém todos os genes presentes no genoma de um organismo. Muitas vezes um cientista deseja olhar para um gene específico que está ativo em um tecido ou tipo celular especial.

Bibliotecas de cDNA proporcionam a solução para os dois problemas. Os cDNAs são produzidos a partir de mRNAs que, quando processados, não contêm quaisquer íntrons. Isso significa que qualquer cDNA isolado irá conter apenas as seqüências codificantes para esse gene, o que reduz a quantidade desse DNA a ser isolado e analisado. Além disso, uma biblioteca de cDNA representa apenas aqueles genes que são expressos em um tipo celular específico, reduzindo, dessa forma, o número de genes a ser submetido à triagem.

Para produzir uma biblioteca de cDNA, o mRNA total é isolado de um tecido de células de interesse por meio da utilização da cromatografia de afinidade a oligo(dT). E cópias de cDNA são sintetizadas utilizando oligo(dT) como primer para RT, conforme já descrito. Essa síntese produz uma molécula híbrida de DNA-RNA. A etapa seguinte é sintetizar a segunda fita de DNA para produzir um cDNA fita dupla (ds cDNA, do inglês *double-stranded cDNA*). A síntese de DNA requer um primer. O ideal seria poder utilizar um primer que inicie o mais próximo possível da extremidade 5' do mRNA para sintetizar a molécula inteira, como foi realizado na extremidade 3' utilizando oligo(dT). No entanto, é muito difícil começar a síntese no início da extremidade 5' do mRNA, uma vez que cada mRNA tem sua própria seqüência



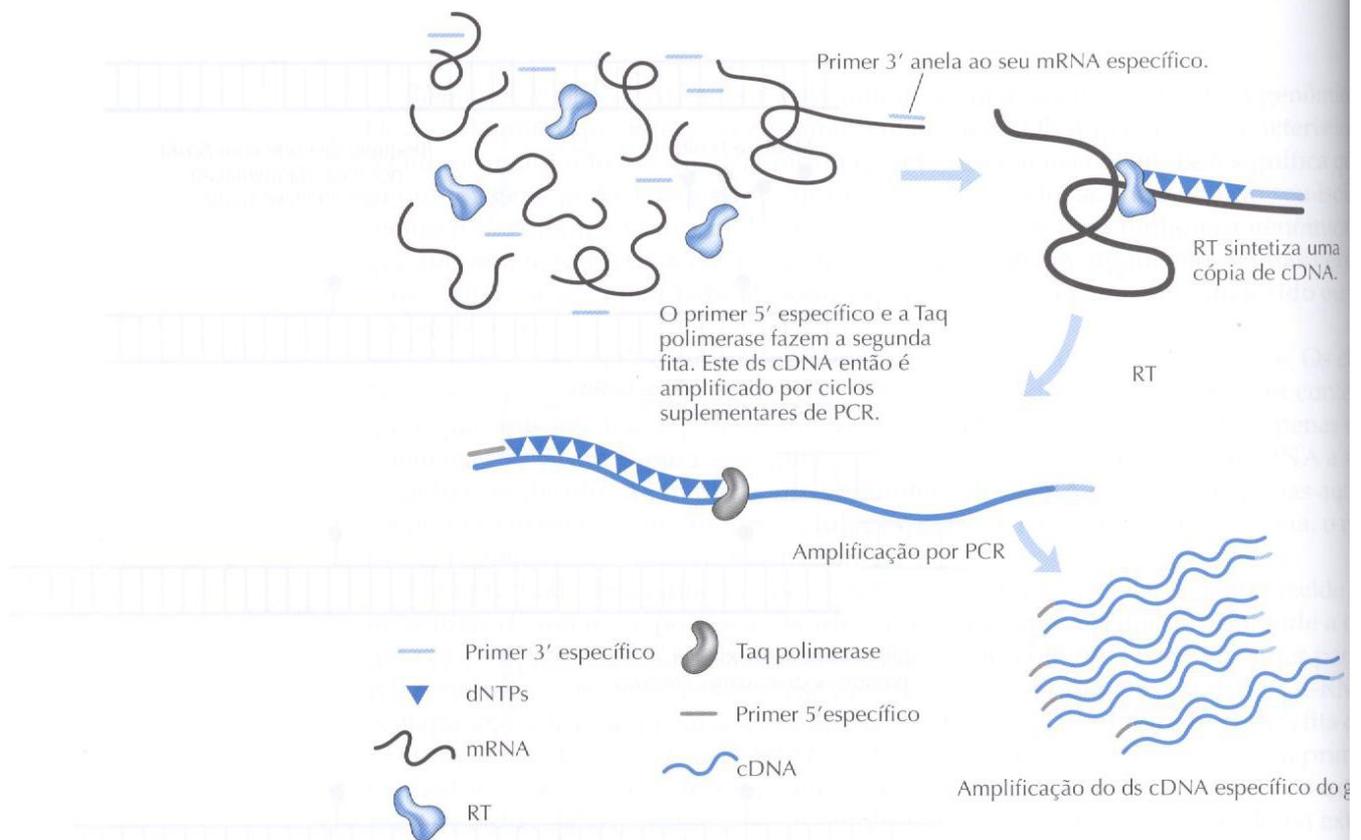
Criando um cDNA fita dupla para clonagem



única nessa região. Ao invés disso, deve-se contar com um método chamado de início aleatório. Após a remoção completa do mRNA, a fita de cDNA restante é incubada com hexâmeros aleatórios – oligonucleotídeos que representam cada seqüência possível de seis nucleotídeos. Os primers ligam-se iniciando seqüências a partir do cDNA, resultando caracteristicamente em cDNAs fita dupla que carecem da extremidade 5'.

O cDNA fita dupla resultante é preparado para ligação em um vetor plasmídico ou em um fago pela adição de adaptadores às extremidades utilizando T4 DNA ligase. Os adaptadores são oligonucleotídeos fita dupla curtos que contêm uma seqüência de restrição de reconhecimento. Por exemplo, um adaptador com a seqüência GAATTC contém o sítio de reconhecimento para *EcoRI*. Se necessário, antes da adição dos adaptadores, o cDNA fita dupla pode ser incubado com metilase *EcoRI*, bloqueando qualquer sítio *EcoRI* dentro do próprio cDNA de clivagem durante a etapa seguinte. (Em estratégias de clonagem direcional, um adaptador diferente é adicionado para cada extremidade da fita dupla de cDNA, criando um sítio de endonuclease de restrição diferente em cada extremidade.)

A digestão com a enzima de restrição *EcoRI* corta os adaptadores e cria extremidades coesivas para ligação do cDNA em vetores digeridos com *EcoRI*. Os plasmídeos ou fagos recombinantes que resultam são transformados em *E. coli* ou usados para transformar *E. coli*, distribuídas em placas de ágar. Antibiótico no meio seleciona as colônias que expressam um gene de resistência de um vetor recombinante. A presença de células contendo fagos é visualizada por meio da formação de placas. De qualquer forma, agora as placas contêm uma biblioteca de cDNA que pode ser submetida à triagem da mesma maneira que uma biblioteca de DNA genômico.



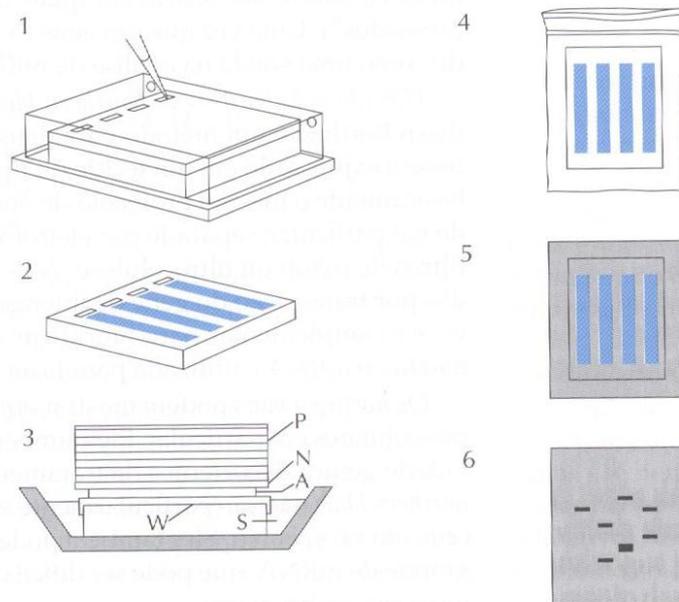
Transcriptase reversa – reação em cadeia da polimerase (RT-PCR)

Triagem Baseada em Análise Computacional. Com as seqüências do genoma de muitos organismos completas ou quase completas, geralmente não é necessário isolar um cDNA de uma biblioteca. Um gene de interesse pode ser identificado pela triagem de um banco de dados. Usando tal informação, podem ser feitos primers específicos para a extremidade 5' e 3' do gene. O primer 3' é pareado ao mRNA isolado de células que expressam o gene de interesse, e um cDNA é produzido utilizando-se RT. A seguir, o primer 5' é adicionado e a síntese da segunda fita é realizada com a utilização do DNA polimerase da *E. coli*, produzindo um cDNA fita dupla específico. Essa estratégia tem um ponto fraco. Se o mRNA está presente em pequena quantidade, pode ser difícil ter cDNA fita dupla suficiente para efetuar o trabalho. A alternativa é fazer a transcrição reversa do mRNA para produzir o cDNA e, então, adicionar o primer 5' e usar a reação em cadeia da polimerase (PCR, do inglês *Polymerase Chain Reaction*) para amplificar o cDNA em várias cópias, conforme a necessidade (ver Capítulo 6). Esse método é denominado PCR transcriptase reversa (RT-PCR, do inglês *reverse transcriptase PCR*).

Trabalhando com genes clonados

Uma vez que um gene tenha sido clonado de uma biblioteca, uma grande quantidade de possibilidades se abrem para análise.

Hibridização Southern. A hibridização Southern (ou *blotting*), assim denominada após sua descoberta por Ed Southern, pode ser utilizada para estudar a organi-



Southern blotting

(1) Aplicar o DNA digerido no gel de agarose e separar por eletroforese; (2) visualizar o DNA e desnaturar no gel; (3) transferir o DNA para filtro de nitrocelulose por capilaridade (P = toalhas de papel, N = nitrocelulose, A = gel de agarose, W = esponja, S = solução com alta concentração de sal); (4) hibridizar a sonda radioativa ao filtro que, se corado, será uma réplica do gel no espelho; (5) lavar o filtro e revestir com filme de raio X e (6) o auto-radiograma realizado revela a localização de bandas às quais a sonda se ligou. (Adaptada a partir do projeto de arte de Lisa Shoemaker.)