



INSTITUTO POLITÉCNICO DE BRAGANÇA
Escola Superior Agrária

METODOLOGIAS DE IDENTIFICAÇÃO E ELUCIDAÇÃO DE GENES EM BIBLIOTECAS GENÓMICAS

Maria de Lurdes Antunes Jorge

Bragança, 2012

Lição a apresentar na Escola Superior Agrária do Instituto Politécnico de Bragança, para Provas Públicas de Avaliação de Competência Pedagógica e Técnico-Científica, de acordo com o nº 9 do artigo 6.º e do nº 5 do artigo 8.ºA do Decreto-Lei n.º 207/2009, alterado e aditado pela Lei n.º 7/2010.

1 – ENQUADRAMENTO

1.1 – Objectivos da Unidade Curricular de Biologia Molecular

A lição a apresentar insere-se no conteúdo programático da Unidade Curricular de Biologia Molecular do 2º ano do Curso de Engenharia Biotecnológica da Escola Superior Agrária de Bragança.

Um dos objectivos desta Unidade Curricular (UC) é fornecer aos alunos conhecimentos que lhes permitam compreender as matérias leccionadas noutras UCs do *curriculum* do respectivo curso (Engenharia Genética, Biotecnologia na Produção Animal e Biotecnologia Microbiana), para as quais os assuntos estudados nesta UC são fundamentais.

Adicionalmente, os alunos do Curso de Engenharia Biotecnológica deverão ficar preparados para dar resposta a solicitações de entidades empregadoras em que, directa ou indirectamente, tenham que fazer utilização de conhecimentos na área da Biologia Molecular.

A UC de Biologia Molecular tem uma carga horária semanal de 2 horas de aulas teóricas e de 2 horas de aulas práticas, a que correspondem 6,0 ECTS.

O conteúdo programático das aulas teóricas obedece à seguinte distribuição:

1. Introdução às bases genéticas. Natureza química e estrutura do DNA
2. Diferentes tipos de genoma. Genomas eucariotas, procariotas, virais, organelares e nucleares
3. Replicação. Replicação em procariotas e eucariotas
4. Transcrição. Etapas da transcrição: iniciação, alongamento e terminação
5. Tradução e Código Genético. Síntese de proteínas: fases, factores e enzimas
6. Técnicas básicas de Biologia Molecular utilizadas na medicina
7. Tecnologia de DNA Recombinante. Clonagem. Vectores e estratégia de Clonagem

Em anexo apresenta-se o programa desta unidade curricular.

Na apresentação desta lição será adoptada uma metodologia expositiva de matérias com recurso a meios didácticos diversificados, que irão desde a projecção de uma apresentação em PowerPoint, à visualização de animações de carácter pedagógico e consulta de *sites* na Internet. Sempre que necessário, recorrer-se-á também à escrita no quadro, de forma a facilitar a apreensão de conteúdos e a interligação entre conceitos.

Será privilegiada, sempre que possível, a articulação entre os conteúdos teóricos das matérias leccionadas e os trabalhos de natureza experimental executados nas aulas práticas.

1.2 – Objectivos da lição

No âmbito da unidade curricular de Biologia Molecular esta lição enquadra-se no último capítulo, sobre a Tecnologia do DNA recombinante. Pretende-se integrar conceitos previamente abordados noutras aulas teóricas, com os desta lição, recorrendo à exposição de um caso concreto: a construção e rastreio de bibliotecas de DNA.

Pretende-se que os alunos consigam relacionar conhecimentos previamente adquiridos sobre:

- endonucleases de restrição (modo de acção, tipo de extremos gerados, compatibilidade entre extremos)
- vectores de clonagem
- características de “polylinkers”
- sondas de ácidos nucleicos
- utilização de oligonucleótidos iniciadores específicos e degenerados
- modo de actuação de diversas polimerases: DNA polimerase, Klenow DNA polimerase
- modo de acção de outras enzimas usadas em biologia molecular: ribonucleases (RNases), ligases, nucleases, tranferase terminal, fosfatase alcalina
- selecção de transformantes
- transcrição e RNA “splicing”
- electroforese

com os conceitos a introduzir na presente lição:

- bibliotecas genómicas e de cDNA
- clones de DNA genómico e cDNA
- rastreio de bibliotecas (com sondas, anticorpos, actividade biológica da proteína, por PCR)
- “Expressed Sequence Tags” (ESTs)
- “High-Efficiency Thermal Asymmetric Interlaced PCR” (HE-TAIL PCR)

No final da lição, os alunos deverão ser capazes de:

- distinguir bibliotecas de genómico e de cDNA
- entender as utilizações de bibliotecas de genómico e de cDNA, bem como vantagens e inconvenientes de cada uma
- escolher os vectores de clonagem em função dos objectivos de utilização
- compreender as diferenças de actuação de diversas enzimas usadas no laboratório de biologia molecular
- referir diferentes métodos de rastreio de bibliotecas de DNA
- indicar vantagens e inconvenientes de cada método

2 – SUMÁRIO

Metodologias de identificação e elucidação de genes em bibliotecas genómicas

1. Bibliotecas de DNA

2. Obtenção de uma biblioteca de DNA genómico

2.1 Preparação do DNA

2.2 Escolha dos vectores de clonagem

2.3 Preparação do vector

3. Obtenção de bibliotecas de cDNA

4. Rastreio de bibliotecas de DNA

4. 1 Rastreio por hibridação de DNA

4.1.1 Obtenção de “Expressed Sequence Tags” - ESTs

- 4.2 Rastreio por ensaio imunológico
- 4.3 Rastreio da actividade da proteína
- 4.4 Rastreio baseado em técnicas de PCR

5. Exemplo de aplicação: Metodologias usadas na pesquisa da sequência nucleotídica de novos genes em *Trichoderma harzianum*.

BIBLIOGRAFIA ACONSELHADA

- Alberts, B., Bray, D., Johnson, A., Lewis, J., Raff, M., Roberts, K., & Walter, P. (2002).** *Essential Cell Biology- An Introduction to the Molecular Biology of the Cell*. New York & London: Garland Publishing, Inc. 630pp. ISBN 0-8153-2971-7.
- Correia, I., Moreira, L., & Fialho, A. (2003).** Engenharia Genética. In N. Lima, & M. Mota, *Biotecnologia. Fundamentos e aplicações*: 125-161. LIDEL - Edições Técnicas, Lda. ISBN-978-972-757-197-0.
- Hartl, D., & Jones, E. (1998).** *Genetics. Principles and Analysis* (4th edition ed.). Sudbury, Massachusetts: Jones and Bartlett Publishers. 840pp. ISBN 0-7637-0489-X.
http://www.ornl.gov/sci/techresources/Human_Genome/project/benefits.shtml.
- King, M. (2011).** *Molecular Tools of Medicine*.
<http://themedicalbiochemistrypage.org/molecular-medicine.html#genomic>.
- Lodish, H., Berk, A., Zipursky, S. L., Matsudaira, P., Baltimore, D., & Darnell, J. (2000).** *Molecular Cell Biology*. (4th edition ed.). New York: W. H. Freeman and Company. 1084pp. ISBN 0-7167-3706-X.
- Nicholl, D. (2008).** *An Introduction to Genetic Engineering* (Third Edition ed.). Hong Kong: Cambridge University Press. ISBN 978-0-521-850063.
- Primrose, S., & Twiman, R. (2006).** *Principles of Gene Manipulation and Genomics* (7th Edition ed.). Blackwell Publishing.
- Videira, A. (2001).** *Engenharia Genética. Princípios e aplicações*. Lidel - edições técnicas, Lda. 168pp. ISBN 978-972-757-163-8.

Unidade Curricular	Biologia Molecular	Área Científica	Biologia e Bioquímica
Licenciatura em	Engenharia Biotecnológica	Escola	Escola Superior Agrária de Bragança
Ano Lectivo	2011/2012	Ano Curricular	2
Tipo	Semestral	Semestre	1
Horas totais de trabalho	162	Horas de Contacto	T 30 TP - PL 30 TC - S - E - OT 20 O -
Nível	1-2	Créditos ECTS	6.0
Código	9088-310-2101-00-11		

T - Ensino Teórico; TP - Teórico Prático; PL - Prático e Laboratorial; TC - Trabalho de Campo; S - Seminário; E - Estágio; OT - Orientação Tutoria; O - Outra

Nome(s) do(s) docente(s) Altino Branco Choupina

Resultados da aprendizagem e competências

No fim da unidade curricular o aluno deve ser capaz de:

1. Adquirir sólidos conhecimentos sobre a estrutura e propriedades dos ácidos nucleicos.
2. Conhecer a função do ADN e os mecanismos genéticos fundamentais.
3. Identificar o dogma central da biologia molecular.
4. Compreender a estrutura do genoma.
5. Interpretar os diferentes mecanismos da expressão genética.
6. Compreender os processos que permitem o estabelecimento de mutações e recombinações genéticas.
7. Estabelecer uma visão geral das técnicas do ADN recombinante a partir de conhecimentos adquiridos.
8. Tomar consciência das implicações sociais das aplicações da Biologia Molecular

Pré-requisitos

Antes da unidade curricular o aluno deve ser capaz de:

1. Possuir conhecimentos, gerais, da Biologia, Bioquímica, Genética e Microbiologia.
2. Recomenda-se ainda que possuam conhecimentos de Inglês.

Conteúdo da unidade curricular

Estrutura e função dos ácidos nucleicos, expressão genética, recombinação genética e tecnologia do DNA recombinante.

Conteúdo da unidade curricular (versão detalhada)

1. Teóricos:
 - A estrutura dos ácidos nucleicos;
 - a natureza dos genes e função do ADN;
 - controlo da expressão genética;
 - mutações e recombinação genéticas;
 - ADN recombinante: Visão geral;
2. Práticos:
 - electroforese em géis de agarose;
 - análise de restrição com endonucleases
 - extracção purificação e quantificação de ácidos nucleicos
 - Reacção em cadeia da polimerase (PCR);
 - preparação de células competentes e transformação com plasmídeos autónomos;

Bibliografia recomendada

1. Lehninger Principles of Biochemistry, 3rd ed. Lewin, B. (2000)
2. Genes VII, 7th ed, Oxford University Press. Sambrook J et al (1989)
3. Sambrook J et al (1989) Molecular Cloning: a Laboratory Manual, Cold Spring Harbor Laboratory Press

Métodos de ensino e de aprendizagem

Aulas magistrais com recurso a salas de aula equipadas com datashow; Aulas laboratoriais de manipulação de ácidos nucleicos e transformação genética, de forma a complementar e consolidar os conhecimentos adquiridos nos conteúdos teóricos, com recurso a salas laboratoriais. Pesquisa bibliográfica, usando as suas bibliotecas e a rede wireless existente no Campus de Santa Apolónia.

Alternativas de avaliação

- Exame final - (Ordinário, Trabalhador) (Final, Recurso, Especial)
- Trabalhos Laboratoriais - 25% (Miniteste e avaliação de relatórios)
- Exame Final Escrito - 75%

Língua em que é ministrada

Português

Validação Electrónica

Altino Branco Choupina	Paula Cristina Santos Baptista	Ana Maria Pinto Carvalho
07-11-2011	07-11-2011	15-11-2011

ÍNDICE

ÍNDICE	i
ÍNDICE DE FIGURAS	iii
ÍNDICE DE TABELAS	v
1. Bibliotecas de DNA	1
2. Obtenção de bibliotecas de DNA genómico	5
2.1 Preparação do DNA.....	5
2.2 Escolha dos vectores de clonagem.....	8
2.3 Preparação do vector.....	16
3. Obtenção de bibliotecas de cDNA	17
4. Rastreio de bibliotecas de DNA	25
4. 1 Rastreio por hibridação de DNA.....	29
4.1.1 Obtenção de “Expressed Sequence Tags” - ESTs	33
4.2 Rastreio por ensaio imunológico	34
4. 3 Rastreio da actividade da proteína.....	35
4. 4 Rastreio baseado em técnicas de PCR.....	35
5. Exemplo de aplicação	39
5.1 Metodologias usadas na pesquisa da sequência nucleotídica de novos genes em <i>Trichoderma harzianum</i>	39
5.2 Avaliação da actividade das proteínas codificadas por <i>lip1</i> e <i>lip2</i>	54
6. Considerações finais	57
Referências Bibliográficas	59
ANEXOS	65
ANEXO 1 – Alguns dos vectores usados em bibliotecas de cDNA	A

Lambda Zap.....	A
Fagemídeos.....	A
ANEXO 2 – Alguns dos vectores usados em bibliotecas de genómico	C
Cromossomas artificiais de leveduras (YACs)	C
Cromossomas Bacterianos artificiais (BACs).....	D
Cromossomas artificiais derivados do fago P1 (PACs)	D
ANEXO 3 – Condições de infecção de <i>E. coli</i> (“library screening”)	G

ÍNDICE DE FIGURAS

Figura 1 – Diferenças entre clones de DNA genómico e clones de cDNA derivados da mesma região de DNA.....	2
Figura 2 – Mapas dos vectores pBR322 (A) e pUC19 (B).....	4
Figura 3 – Sobreposição de fragmentos de restrição obtidos por digestão parcial de DNA genómico humano com a enzima de restrição <i>Sau3A</i> I.....	7
Figura 4A – Componentes estruturais do bacteriófago λ	10
Figura 4B – Microfotografia de partículas de fago lambda.	10
Figura 5 – Mapa simplificado do genoma do fago λ	10
Figura 6 – Ciclo de vida do bacteriófago λ	11
Figura 7 – Construção de uma biblioteca de genómico em fago λ	12
Figura 8 – Montagem da partícula viral do fago λ	13
Figura 9 – Mapa do vector lambda gt11.....	14
Figura 10 – Características básicas de um cosmídeo.....	16
Figura 11 – Obtenção de uma biblioteca de cDNA em fago λ	19
Figura 12 – Primers usados na síntese da primeira cadeia de cDNA.....	20
Figura 13 – Processos de síntese da segunda cadeia de cDNA: A) RNase H; B) Self-priming; C) Tailing and priming.	21
Figura 14 – Modo de actuação de enzimas de restrição e de metilases.	23
Figura 15 – Correspondência entre um péptido e as possíveis sequências codificantes na molécula de DNA.....	26
Figura 16 – Moléculas usadas na marcação não radioactiva de sondas de DNA.	28
Figura 17 – Rastreamento de colónias por hibridação de DNA (“colony blot”).	29
Figura 18 – Placas de lise do fago λ em <i>E. coli</i>	32
Figura 19 – Transferência de colónias robotizada no Instituto Sanger.	32
Figura 20 – Microarrays.	33
Figura 21- Caminhar no Cromossoma (“Chromosome Walking”).....	36
Figura 22 – Cromatograma de sequenciação.	41
Figura 23 – Autorradiografia de membranas de hibridação.....	42
Figura 24 – Representação esquemática da amplificação por HE-TAIL PCR do gene <i>lip1</i>	43

Figura 25 – Electroforese dos produtos de reacção da HE-TAIL PCR usada para a elucidação de <i>lip1</i>	45
Figura 26 – Sequência deduzida da proteína lip1 (parte inicial).	46
Figura 27 – Sequência deduzida da proteína lip1 (parte final).	47
Figura 28 – Representação esquemática da amplificação por HE-TAIL PCR do gene <i>lip2</i>	49
Figura 29 – Electroforese dos produtos de reacção da HE-TAIL PCR usada para a elucidação de <i>lip2</i>	51
Figura 30 – Sequência deduzida da proteína lip2 (parte inicial).	52
Figura 31 – Sequência deduzida da proteína lip2 (parte final).	53
Figura 32 – Construção do plasmídeo recombinante pLB2- <i>lip1</i>	55
Figura 33 – Halos de actividade esterásica (A e B) e fluorescência devida a actividade lipolítica (C).	56
Figura 34 – SDS-PAGE obtido com extractos proteicos celulares de levedura GS115 e de GS115 transformada com <i>lip1</i> (crescimento de 48h em YEPD).	56
Figura 35 – Mapa de restrição do vector de inserção Lambda ZAP II	A
Figura 36 – Mapa e características de pBluescript SK ⁺	B
Figura 37 – Representação de um vector YAC.	C
Figura 38 – CopyRight vector pSMART BAC.	E
Figura 39 – Resolução de plasmídeos enredados após replicação por acção da Cre recombinase do fago P1.	E
Figura 40 – Vector pPAC4.	F

ÍNDICE DE TABELAS

Tabela 1- Características de alguns dos vetores de clonagem utilizados na obtenção de bibliotecas de DNA.	8
Tabela 2 – Alfabeto de nucleótidos degenerados	26
Tabela 3 – Etapas e ciclos de HE-TAIL-PCR usados na elucidação de <i>lip1</i>	44
Tabela 4 – Etapas e ciclos de HE-TAIL-PCR usados na elucidação de <i>lip2</i>	49

1. BIBLIOTECAS DE DNA

A sequenciação de genomas tem inúmeras aplicações, nomeadamente em medicina, antropologia, agricultura, melhoramento animal e ambiente, entre outros. Por exemplo, algumas doenças, como as de Alzheimer ou o cancro da mama revelaram uma componente genética até aí desconhecida (Human Genome Project, 2009), tornando mais fácil uma possível despistagem. Também a elucidação de genomas microbianos adaptados a condições adversas extremas poderá ajudar a solucionar alguns problemas para os quais a ciência ainda não encontrou soluções, particularmente em biorremediação, na utilização de energias renováveis, na pesquisa de biosensores ou produção de enzimas (Microbial Genome Program, 2010).

No entanto, para compreender os mecanismos de funcionamento de genes, devemos analisar toda a sua estrutura, que abrange não só a sequência nucleotídica correspondente à região codificante, mas também as sequências de regulação localizadas no promotor, e a sequência do terminador. Isto só é possível se dispusermos de clones de DNA genómico representativos de todo o genoma, e de uma metodologia de rastreio que nos permita identificar os genes de interesse. A maior parte destas metodologias envolve o rastreio de bibliotecas de DNA.

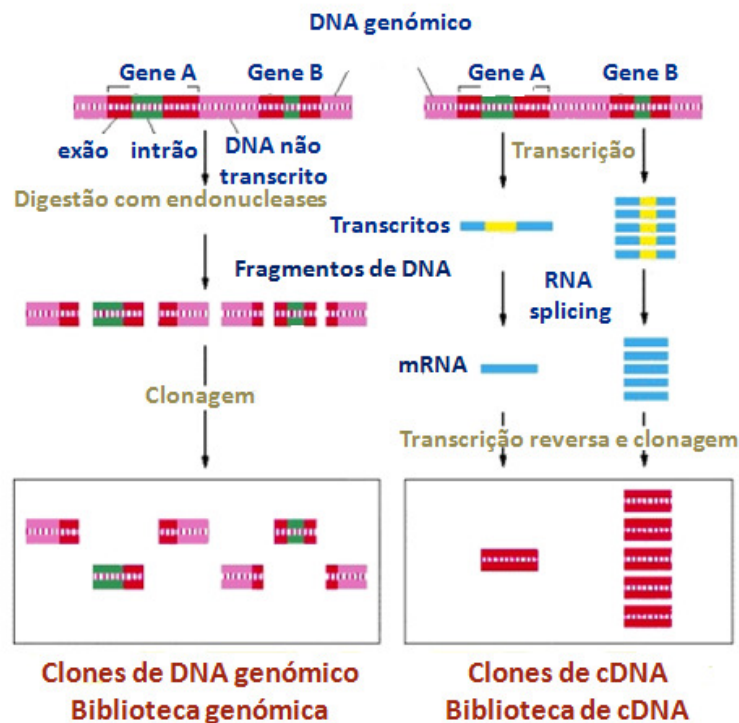
Os bancos ou bibliotecas de DNA são colecções de fragmentos de DNA de um organismo clonados num vector adequado. Estas moléculas de DNA recombinante são armazenadas em células hospedeiras, podendo ser amplificadas e armazenadas indefinidamente. São essencialmente de dois tipos:

- **Bibliotecas de DNA genómico**, em que os fragmentos de DNA clonados são provenientes do genoma, e representativos de todas as sequências de DNA presentes nesse genoma (**Figura 1**). Com raríssimas excepções, contêm sempre as mesmas sequências independentemente do tecido do qual o DNA foi retirado (Alberts, et al., 2002). Numa biblioteca de genómico:
 - todas as regiões do genoma deverão estar igualmente representadas entre si.
 - o número de clones deverá ser de tal forma que uma dada região do genoma tenha elevada probabilidade de ocorrência pelo menos num dos clones (apesar disso, é desejável que cada clone contenha um inserto do

maior tamanho possível, para que o número de clones não seja tão grande que dificulte o manuseamento da biblioteca).

- os fragmentos clonados devem ter tamanho suficiente para poderem abranger as sequências codificantes e flanqueadoras de um gene.

O facto de nas bibliotecas de DNA genómico estarem representadas todas as sequências do genoma, torna-se vantajosa na medida em que permitem a análise da estrutura e função dos genes (existência de intrões, exões, fronteiras intrão-exão, sequências reguladoras), a elucidação de regiões intergénicas, assim como a organização do genoma.



(Alberts, et al., 2000)

FIGURA 1 – DIFERENÇAS ENTRE CLONES DE DNA GENÓMICO E CLONES DE cDNA DERIVADOS DA MESMA REGIÃO DE DNA.

Neste exemplo, o gene A não é frequentemente transcrito enquanto o B é frequentemente transcrito; ambos contêm intrões (a verde). Na biblioteca genómica, quer os intrões quer as sequências não transcritas de DNA (a rosa) são incluídos nos clones e a maioria contém, no máximo apenas uma curta porção da sequência codificante de um gene (vermelho). Nos clones de cDNA, os intrões (a amarelo) foram removidos devido a RNA “splicing”, durante a formação de mRNA (a azul), e assim cada clone apresenta uma sequência codificante contínua. Porque o gene B é mais frequentemente transcrito nas células a partir da qual a biblioteca foi feita, ele vai estar mais representado que o gene A. No entanto, na biblioteca genómica os genes A e B estão em princípio igualmente representados.

As maiores desvantagens destas bibliotecas são:

- o trabalho exigido na sua elaboração e na manutenção de clones.
 - o facto de nem todos os clones poderem ser expressos em células procarióticas.
 - poder haver clones com sequências de genes incompletas, devido ao facto de serem obtidas com o uso de enzimas de restrição.
 - haver necessidade de determinar os codões de iniciação e de terminação nas sequências de genes completas.
- **Bibliotecas de DNA complementar (cDNA)**, que apenas contêm moléculas de DNA copiadas a partir dos mRNAs (cDNAs) existentes num tecido num dado momento e condição ambiental. Nestas bibliotecas há diferente representatividade entre genes, resultante da diferente proporção entre mRNAs na célula (**Figura 1**). Assim, há variações entre bibliotecas de cDNA obtidas de células colocadas em condições ou de estádios de desenvolvimento diferentes, e entre as obtidas de diferentes tecidos de um mesmo organismo.

Uma vantagem das bibliotecas de cDNA é o facto de conterem apenas as regiões codificantes de cada gene. Sendo assim, as bibliotecas de cDNA possibilitam:

- a análise da estrutura e função de exões e a dedução de sequências proteicas.
- a identificação dos genes transcritos em determinada etapa celular ou condição ambiental.
- a produção de proteínas eucarióticas em bactérias (estudos de expressão).
- a obtenção de sondas de cDNA.

Quando comparadas com as de genómico, têm ainda a vantagem de terem um menor tamanho, sendo por isso mais fáceis de manipular e armazenar, e a de maior representatividade de cada uma das sequências transcritas. A sequenciação de moléculas de cDNA completas permite detectar a presença e localização de intrões, por comparação com a sequência nucleotídica do DNA genómico. Têm como desvantagens:

- o facto de só as sequências codificantes poderem ser analisadas.
- a difícil detecção de sequências com baixo nível de transcrição.
- o facto de serem obtidas a partir de mRNA, de fácil degradação.

Os vectores de clonagem usados na construção de bibliotecas são moléculas de DNA capazes de amplificar, em centenas de cópias, a informação genética contida na sua molécula. No caso de conterem um inserto (DNA exógeno), este é copiado em simultâneo. Todos os vectores de clonagem têm uma origem de replicação independente da replicação do DNA da célula hospedeira, e alguns deles codificam as enzimas necessárias à sua própria replicação. Por questões de funcionalidade desenvolveram-se vectores com um local de clonagem múltipla (MCS – “multiple cloning site”, também designado por “polylinker”), ou seja, um local onde há uma sucessão de sequências de reconhecimento para algumas endonucleases de restrição, únicas na cadeia, que possibilitam a clonagem. Na sua sequência nucleotídica encontra-se também a sequência de pelo menos um gene que permite a detecção de fenótipos recombinantes – marcador de selecção. Estão neste caso os genes de resistência a antibióticos (ampicilina, zeocina,..) ou aqueles cuja interrupção provocada pelo inserto leva à detecção das colónias transformadas (β -galactosidase, por exemplo). Na **Figura 2**, a título exemplificativo, estão representados os vectores pBR322 e pUC19, dois dos primeiros mais usados vectores de *Escherichia coli*.

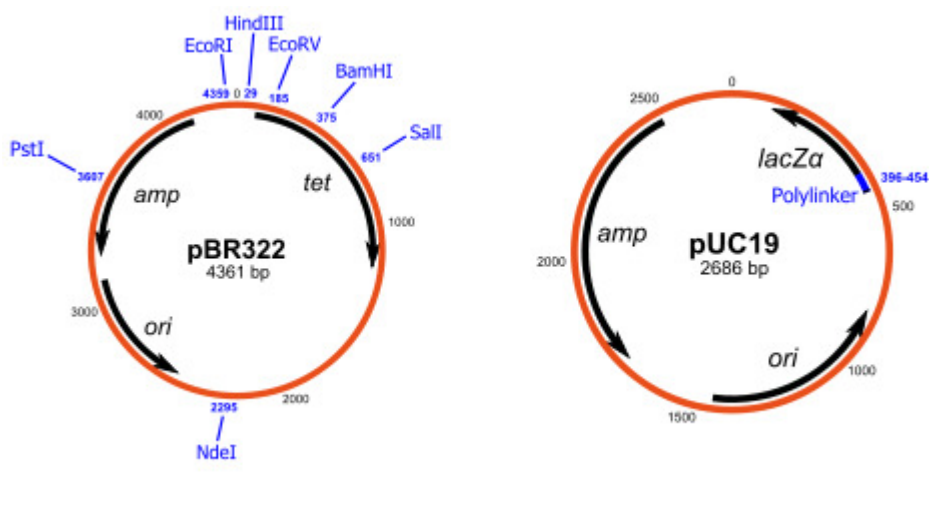


FIGURA 2 – MAPAS DOS VECTORES pBR322 (A) E pUC19 (B).

A - pBR322 foi o primeiro vector de clonagem em *E. coli* a ser amplamente utilizado. Criado por Bolívar e Rodriguez em 1977, tem 4361pb, uma origem de replicação (*ori*), e genes de resistência à ampicilina (*amp^R*) e à tetraciclina (*tet^R*). Onze dos seus 40 locais de restrição situam-se no interior da sequência de gene *tet^R*, seis no meio da sequência *amp^R*.

B – O vector pUC19 foi obtido na Universidade da Califórnia. Além de *amp^R*, contém o fragmento N-terminal do gene da β -galactosidase de *E. coli* (*lacZ*). O “polylinker” está inserido em *lacZ*, pelo que a inserção de DNA exógeno provoca a interrupção do gene. Permite a selecção de recombinantes por “screening” de colónias brancas (fenótipo recombinante).

2. OBTENÇÃO DE BIBLIOTECAS DE DNA GENÓMICO

2.1 PREPARAÇÃO DO DNA

A primeira etapa da construção de uma biblioteca de genómico é a obtenção de DNA de elevado peso molecular.

Para que a biblioteca seja representativa de todos os fragmentos do genoma, estes devem ser obtidos numa forma aleatória. O método que permite obter uma completa aleatoriedade na fragmentação do DNA é o fraccionamento mecânico. No entanto, este método acaba por não ser muito utilizado pela maior dificuldade em clonar os fragmentos assim obtidos num vector, pelo que é frequente recorrer à utilização de endonucleases.

Na construção de bibliotecas de genómico é usual clivar o DNA genómico do organismo de interesse com uma enzima de restrição que assegure uma frequência de clivagem da cadeia nucleotídica elevada.

Numa endonuclease de restrição, a distância média entre pontos de clivagem é dada por:

$$D = 4^n,$$

em que **D** representa a distância entre pontos de clivagem na cadeia; **n** representa o número de bases envolvidas no local de reconhecimento, e **4**, o número de diferentes bases desoxirribonucleotídicas do DNA.

Sendo assim, uma sequência nucleotídica de 4pb tem a probabilidade de aparecer em média na cadeia desoxirribonucleotídica uma vez em cada $4^4 = 256$ pb. Têm sequências de reconhecimento de 4pb as endonucleases *Sau3AI* (/GATC), *MboI* (/GATC), *AluI* (AG/CT), e *HaeIII* (GG/CC), entre outras. Se considerarmos o genoma humano haplóide, com cerca de 3×10^9 pb, uma digestão completa com uma destas enzimas (por exemplo, a *Sau3AI*, como representado na **Figura 3**) originaria qualquer coisa como 10^7 fragmentos não sobreponíveis, demasiado para o manuseamento da biblioteca.

A utilização de enzimas de restrição com 6pb no local de reconhecimento como a *BamHI* (G/GATCC), origina fragmentos médios de $4^6 = 4096$ pb, diminuindo o número de clones necessários. No entanto, verifica-se que, embora os locais de reconhecimento das enzimas se distribuam no genoma ao acaso, na prática podem aparecer zonas de grande concentração de um local de reconhecimento, e outras bastante dispersas. Isto acontece especialmente em enzimas que cortam menos frequentemente (Videira, 2001). Com

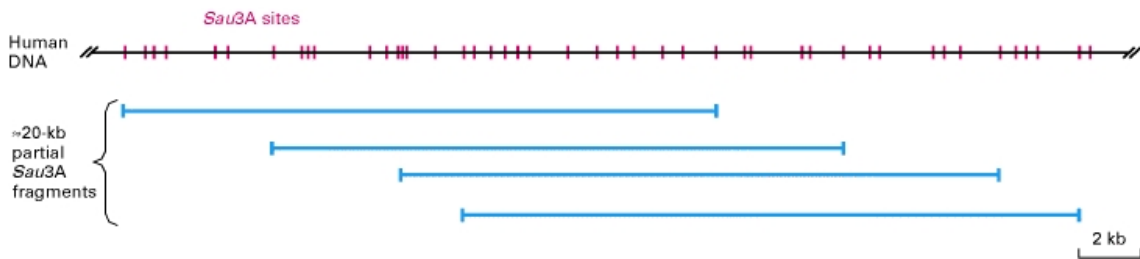
estas últimas, em determinadas regiões do DNA há maior probabilidade de obter fragmentos de dimensões superiores aos limites de capacidade dos vectores utilizados, o que pode levar à não inclusão dessas regiões nas bibliotecas (Videira, 2001).

As endonucleases que reconhecem sequências de 6pb são de utilização menos prática do que as de 4pb, porque obrigam a utilização de diferentes condições de digestão e posterior mistura de fragmentos previamente à clonagem (Wolfe, 2005). Por esta razão, a *Sau3AI* ou o seu isoesquízómero *Mbol* são as enzimas mais utilizadas na clivagem de DNA para obtenção de uma biblioteca de genómico. Estas endonucleases de restrição, para além de terem uma sequência de reconhecimento de 4bp, originam extremos protuberantes 5' compatíveis com os de enzimas de restrição cuja sequência de reconhecimento está presente no local de clonagem múltipla de vectores (*BamHI*, por exemplo).

No entanto, se estas enzimas clivarem todos os locais de reconhecimento existentes na cadeia, vão-se obter inúmeros fragmentos de DNA, pequenos demais para incluir toda a sequência de um gene, incluindo as regiões que flanqueiam o quadro de leitura aberta – ORF (“open reading frame”), muitas vezes com mais de 1000 pb (Pareja, 2005). Nesse caso, o número de clones necessários à construção da biblioteca tornar-se-ia demasiado grande, e a probabilidade de todas as regiões do genoma estarem representadas na biblioteca de genómico diminuiria. Para obviar estes problemas são feitas digestões parciais, usando menos enzima e/ou menos tempo de reacção, de forma a obter fragmentos de restrição de 20Kpb ou mais, dependendo da capacidade de inserção do vector a usar. Na **Figura 3** encontra-se representada a metodologia referida anteriormente.

Após a clivagem, os fragmentos de tamanho desejado (que depende do vector a usar na construção da biblioteca), são seleccionados por electroforese (fragmentos grandes para clonagem em YACs ou BACs¹ poderão ter de ser separados por electroforese em gel de campo pulsado - PFGE – “pulsed field gel electrophoresis”) e posteriormente eluídos. Para o genoma humano e fragmentos com cerca de 20Kpb, clonáveis no fago lambda (λ), são necessários cerca de $1,5 \times 10^5$ clones para construir uma biblioteca completa. Na prática, como a incorporação de fragmentos de restrição nas partículas virais é aleatória, são necessários pelo menos 1×10^6 fagos recombinantes para assegurar a representatividade de cada região do genoma humano, com uma probabilidade de 90-95% (Lodish, et al., 2000).

¹ Cromossomas artificiais de leveduras (YACs) e bacterianos (BACs), usados como vectores de clonagem.



(Lodish, et al., 2000)

FIGURA 3 - SOBREPOSIÇÃO DE FRAGMENTOS DE RESTRIÇÃO OBTIDOS POR DIGESTÃO PARCIAL DE DNA GENÓMICO HUMANO COM A ENZIMA DE RESTRIÇÃO *SAU3A*I.

A enzima reconhece a sequência tetranucleotídica GATC e origina extremos protuberantes 5' em cada uma das cadeias. Uma região hipotética do genoma humano está representada na linha superior, com todos os locais de reconhecimento da enzima (a roxo). A digestão parcial desta região origina fragmentos cerca de 20Kpb (azul), cujas sequências se sobrepõem. Esta técnica permite aumentar a probabilidade de cada fragmento individual da cadeia se encontrar representado.

A fórmula que relaciona o número de clones de uma biblioteca de genómico com o tamanho do inserto clonado e o tamanho do genoma, de forma a representar uma dada região do genoma pelo menos uma vez na biblioteca é:

$$N = \ln(1-P) / \ln[1 - (I/G)]$$

em que N = número de clones, P = probabilidade de uma determinada região do genoma estar representada pelo menos uma vez, I = tamanho do inserto (pb), G = tamanho do genoma (pb) (Clarke & Carbon, A colony bank containing synthetic Col El hybrid plasmids representative of the entire E. coli genome, 1976).

Assim, para detectar uma sequência de interesse no genoma humano, com 99,0% de probabilidade, seria necessário obter uma biblioteca de genómico clonada em fago λ , com um número de clones de cerca de:

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

ou seja, cerca de 7×10^5 clones.

2.2 ESCOLHA DOS VECTORES DE CLONAGEM

Na construção de uma biblioteca de genómico há interesse em utilizar vectores nos quais se consiga inserir o maior fragmento possível da cadeia de DNA, por forma a minimizar o número de clones da biblioteca. Nos últimos anos verificou-se uma grande evolução nas características dos vectores de clonagem. As bibliotecas de genómico de primeira geração utilizavam mais os vectores lambda, posteriormente utilizaram-se sobretudo vectores híbridos fago-plasmídeo, como por exemplo os cosmídeos. Nas mais recentes generalizou-se a utilização de cromossomas artificiais de leveduras, já em fase de substituição por vectores com maior capacidade de acomodar insertos de grandes dimensões, os cromossomas artificiais bacterianos baseados quer na origem de replicação do elemento F quer no bacteriófago P1 (BACs e PACs)(Wolfe, 2005).

Na **Tabela 1** descrevem-se algumas das características de diferentes vectores de clonagem utilizados na obtenção de bibliotecas de DNA, e o número de clones que seriam necessários, em cada um dos casos, para construir uma biblioteca de genómico humano completa.

TABELA 1- CARACTERÍSTICAS DE ALGUNS DOS VECTORES DE CLONAGEM UTILIZADOS NA OBTENÇÃO DE BIBLIOTECAS DE DNA.

Vector	Tamanho máximo do inserto	Número de clones necessários na biblioteca	Vantagens	Inconvenientes
Plasmídeos	10Kpb	10^7	Construção fácil; insertos relativamente estáveis	São necessários muitos clones
Derivados do Fago λ	20 Kpb	5×10^5	Construção fácil; insertos relativamente estáveis	São necessários muitos clones; é difícil obter DNA dos clones
Cosmídeos	45 Kpb	2×10^5	Fácil construção e obtenção de DNA dos clones	Nem sempre são estáveis
YAC	1 Mpb	10^4	São precisos poucos clones	Construção difícil; tendência para rearranjos; Baixa eficiência de transformação
BAC	> 500 Kpb	2×10^4	São precisos poucos clones; muito estáveis	Origem de replicação de cópia única, dificulta a obtenção de DNA em quantidade suficiente
PAC	~120 Kpb	10^5	São necessários menos clones relativamente aos cosmídeos; estáveis	Origem de replicação de cópia única, dificulta a obtenção de DNA em quantidade suficiente

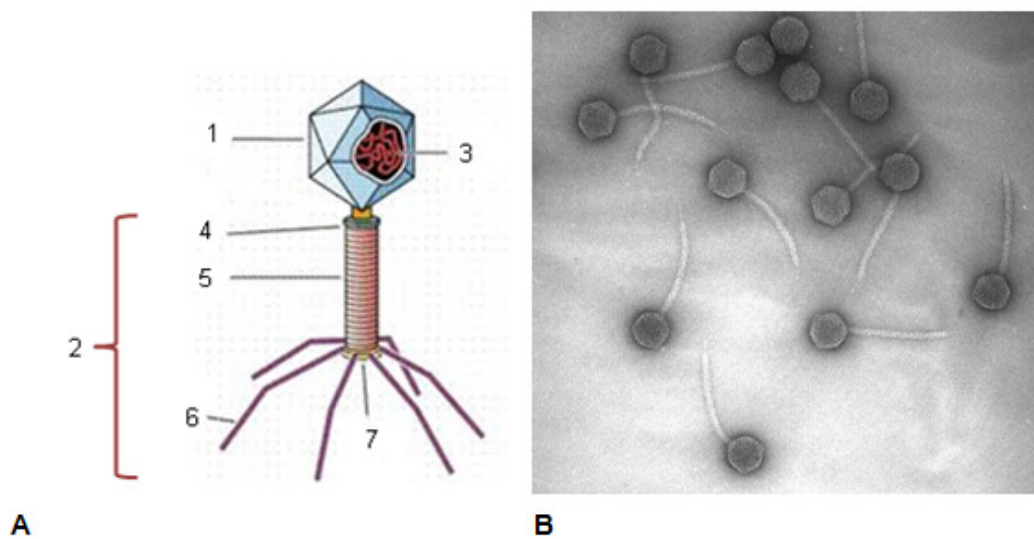
Adaptado de (Wolfe, 2005) e (Lee, sd).

Os plasmídeos derivados de pBR ou pUC (**Figura 2**) não são adequados à construção de uma biblioteca de genómico. Os fragmentos de DNA que conseguem comportar são relativamente pequenos, o que obrigaria a um número elevado de clones. A eficiência de transformação de *E. coli* com plasmídeos é também bastante inferior à capacidade de transdução associada ao fago λ ou vectores derivados (Lodish, et al., 2000).

O fago lambda é uma partícula viral constituída pela cabeça (também designada cápside viral), e pela cauda, onde se localizam as fibras responsáveis pela adsorção à célula bacteriana (**Figura 4**). No interior da cápside encontra-se o material genético – DNA linear de cadeia dupla, com 48490pb, ladeado por dois segmentos de cadeia simples de 12pb em cada um dos extremos 5', complementares, designados “cohesive sites” ou locais *cos*. Funcionam como extremos protuberantes que permitem a circularização do DNA no interior do hospedeiro após a infecção. Na sua forma circular, o genoma do fago atinge as 48502pb. Na **Figura 5** encontra-se o mapa simplificado do fago λ e na **Figura 6** o seu ciclo de vida. Foi muito usado na obtenção de vectores para bibliotecas de genómico porque 20Kpb do seu genoma, correspondentes aos genes necessários para o ciclo lisogénico, e outros, irrelevantes para o ciclo lítico (**Figuras 5 a 7**), podem ser retirados e substituídos por insertos de DNA de interesse, sem perda da capacidade de infecção. Além disso, a eficiência de infecção em *E. coli* é cerca de 1000 vezes superior à transformação obtida com vectores plasmídicos (Lee, sd) (Rinehart, 2005a): cerca de 10^6 colónias/ μ g de DNA plasmídico versus 10^9 placas/ μ g de DNA viral (Anónimo, 2002).

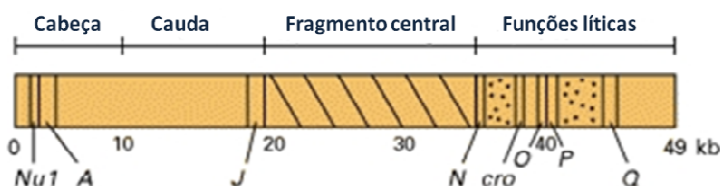
No interior do hospedeiro a montagem das cápsides virais e da cauda (a partir de múltiplas cópias das diferentes proteínas que compõem essas estruturas complexas) são independentes (**Figura 8**). A replicação do DNA origina longas moléculas multiméricas chamadas concatómeros (não são mais do que múltiplas cópias do genoma viral ligadas pelas extremidades através das sequências *cos*, ver **Figuras 7 e 8**). São estas sequências que são reconhecidas pelas proteínas virais levando ao processo de encapsidação. As proteínas *Nu1* e *A* (ver **Figura 5**) ligam-se aos locais *cos* e inserem na cápside viral o DNA localizado entre dois locais *cos* consecutivos, ou seja, cerca de 50kpb. Após a inserção do DNA, dá-se a ligação da cauda, obtendo-se uma nova partícula viral completa.

Para obter viriões lambda com DNA recombinante, o processo de montagem dos componentes do fago (cápside e cauda) decorre “in vitro”. Assim, infecta-se a *E. coli* com um fago λ mutante, defectivo na proteína *A*, uma das duas proteínas necessárias para empacotar o DNA do vírus na cápside viral. Estas células acumulam cápsides vazias.



(Schnos, 2005)

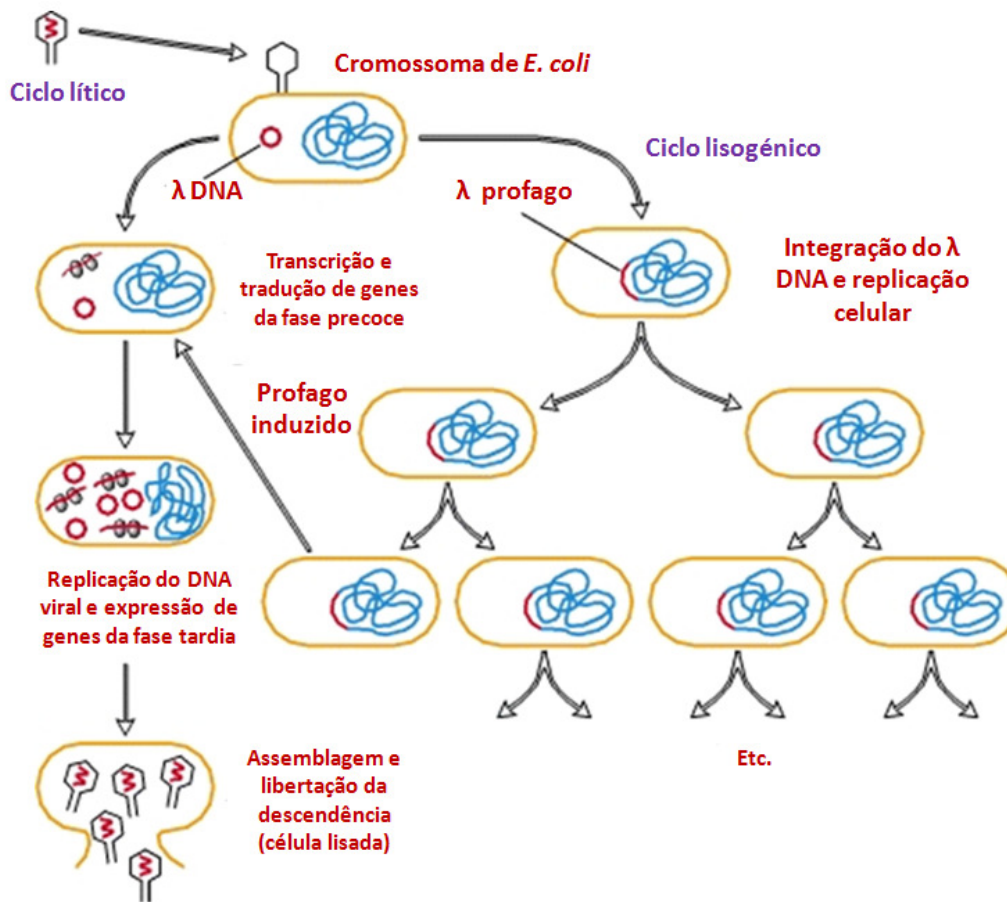
FIGURA 4A - COMPONENTES ESTRUTURAIS DO BACTERIÓFAGO λ : 1. CABEÇA; 2. CAUDA; 3. DNA VIRAL; 4. COLAR; 5. BAINHA; 6. FIBRAS; 7. PLACA BASAL. 4B – MICROFOTOGRAFIA DE PARTÍCULAS DE FAGO LAMBDA.



(Lodish, et al., 2000)

FIGURA 5 - MAPA SIMPLIFICADO DO GENOMA DO FAGO λ

Há cerca de 60 genes no genoma, apenas um pequeno número se encontra representado. Os genes codificantes das proteínas necessárias à montagem da cabeça (cápside viral) e da cauda estão localizados à esquerda, constituindo o braço esquerdo; à direita, formando o braço direito, situam-se os necessários à produção de proteínas do ciclo lítico. As regiões do genoma a tracejado podem ser substituídas por DNA exógeno (correspondendo ao fragmento central), ou retiradas (as pontuadas) sem que a capacidade de infecção do fago λ e a produção de novos viriões seja alterada. Entre os genes J e N podem-se inserir insertos até 24 Kpb (Hartl & Jones, 1998). O braço esquerdo tem cerca de 20Kpb, o direito 9Kpb. A capacidade de empacotamento de DNA na cápside viral do fago λ é de cerca de 37-53Kpb, pelo que o tamanho possível dos insertos variará entre 8Kpb e as 24 Kpb.



(Lodish, et al., 2000)

FIGURA 6 – CICLO DE VIDA DO BACTERÍÓFAGO λ .

O fago reconhece e liga-se ao seu hospedeiro – a *Escherichia coli*, o que leva à injeção do seu DNA no citoplasma da célula bacteriana através da cauda, iniciando-se uma de duas fases do seu ciclo de vida: a lítica ou a lisogénica. Se o estado nutricional do hospedeiro for favorável, inicia-se o ciclo lítico: após a infecção, o DNA linear de cadeia dupla circulariza (à esquerda). Na fase lítica, o DNA do fago λ replica-se originando novas partículas virais e proteínas líticas activas, provocando a lise da célula bacteriana, e a libertação das partículas virais produzidas para o meio extracelular. Se o estado nutricional da *E. coli* não for capaz de suportar a formação de novos viriões, inicia-se o ciclo lisogénico – os genes do fago responsáveis pelo ciclo lítico são reprimidos, e a célula hospedeira produz enzimas fágicas que integram o DNA viral no cromossoma bacteriano (à direita). Esta integração – estágio de profago - ocorre num local específico e não provoca interrupção de genes do hospedeiro. O DNA do profago é replicado ao mesmo ritmo que o DNA cromossomal do hospedeiro. Pode permanecer neste estágio durante várias gerações do ciclo do hospedeiro até que surja uma situação de stress para o hospedeiro (escassez de nutrientes, antibióticos, ou qualquer outra situação prejudicial). Em resposta ao stress, a expressão dos genes da fase lítica é despoletado e o profago activado é excisado do cromossoma de *E. coli* por novas proteínas expressas, passando a replicar-se autonomamente.

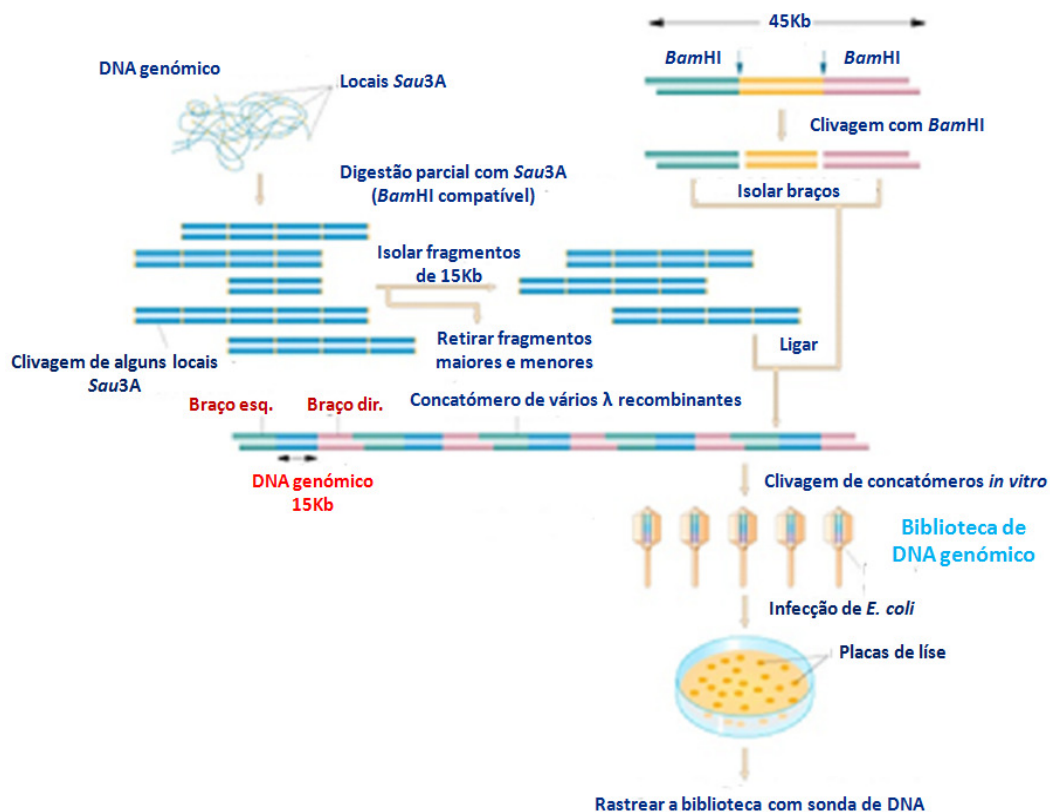
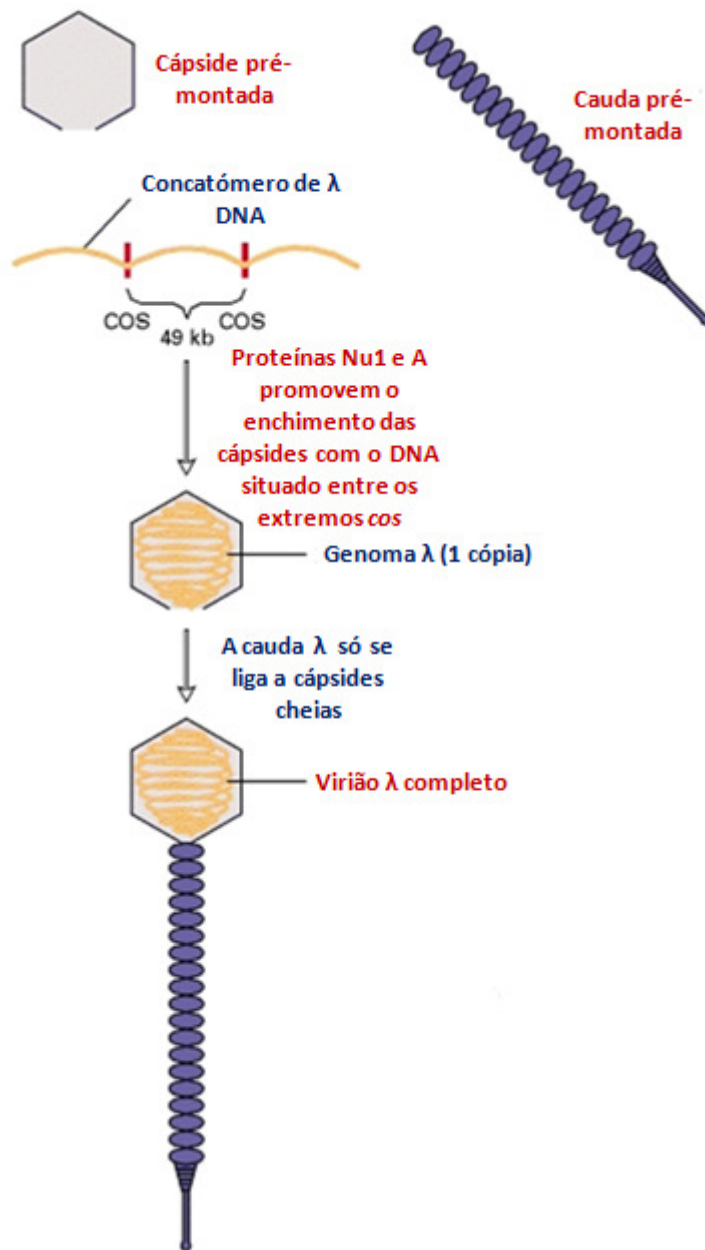


FIGURA 7 –CONSTRUÇÃO DE UMA BIBLIOTECA DE GENÓMICO EM FAGO λ .

A região não essencial do genoma do fago λ (a laranja) é usualmente retirada, de forma a maximizar o tamanho do inserto de DNA. Neste exemplo é retirada com *Bam*HI, e o DNA genómico a inserir é digerido parcialmente com *Sau*3A. As duas enzimas produzem fragmentos coesivos complementares. Os braços de vector λ e cerca de ≈ 15 kpb de DNA genómico são misturados, ligados e empacotados “in vitro” para originar partículas virais λ , com as quais se infecta células de *E. coli*.

Como as caudas pré-montadas só se unem a cápsides “cheias” com DNA, as caudas virais acabam por também se acumular. Com a lise das células obtém-se extractos celulares com elevada concentração de cápsides e de caudas. A mistura destes extractos com proteínas A (obtidas a partir de células infectadas por vírus não mutados e isoladas) e DNA recombinante com locais *cos*, leva ao empacotamento do DNA no interior das cápsides virais e à posterior ligação da cauda, obtendo-se viriões com DNA recombinante.

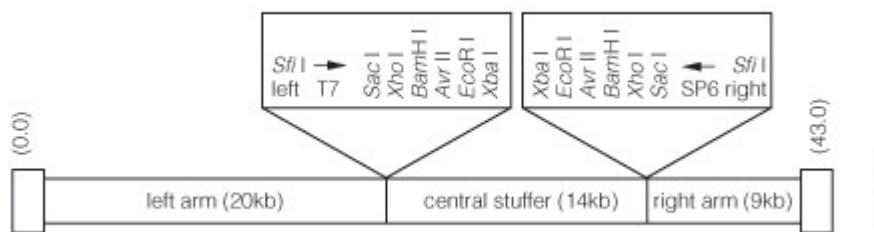


(Lodish, et al., 2000)

FIGURA 8 – MONTAGEM DA PARTÍCULA VIRAL DO FAGO λ.

A partir de várias proteínas do fago são montadas caudas e cápsides vazias. Durante a fase tardia da infecção fágica são formadas longas moléculas de DNA, os concatómeros. Não são mais do que várias cópias do genoma do fago, unidas pelos extremos *cos* (a vermelho). Os locais *cos* são sequências nucleotídicas às quais se ligam as proteínas responsáveis pela encapsidação do material genético, Nu1 e A. Estas proteínas levam à inserção do DNA presente entre dois locais *cos* consecutivos nas cápsides (encapsidação). Só após esta ter ocorrido ocorre a ligação da cauda viral, originando uma partícula viral completa, capaz de infectar a *E. coli*.

Na **Figura 7** está representada a construção de uma biblioteca de genómico em fago λ . A grande desvantagem dos vectores derivados do fago λ reside no facto de não ser fácil obter a partir deles DNA com qualidade suficiente para sequenciar, usar em mutagénesis dirigida e restrição enzimática (Institute of Molecular Development, 2001). Actualmente são usados na construção de bibliotecas de genómico em condições excepcionais (Wolfe, 2005). No entanto, continuam ser usados para a construção de bibliotecas de cDNA (Institute of Molecular Development, 2001), sobretudo pela eficiência de empacotamento do DNA “in vitro” e pela facilidade de manipulação das bibliotecas, quando comparados com os vectores plasmídicos. Alguns dos mais usados em bibliotecas de cDNA de eucariotas são os vectores lambda gt²10 e gt²11 (Kumar & Garg, 2005), que aceitam insertos de 7,6 e 7,2 Kpb, respectivamente (Primrose, et al., 2006). O primeiro é considerado um vector de não-expressão pela ausência de um promotor forte a montante do MCS, pelo que uma biblioteca clonada em λ gt10 só poderá ser rastreada com uma sonda de ácido nucleico. O λ gt11 (**Figura 9**) é um vector de expressão. Neste caso, as bibliotecas que gera podem ser rastreadas com sondas de anticorpos específicas para as sequências de DNA de interesse.



(Promega, 2011)

FIGURA 9 – MAPA DO VECTOR LAMBDA GT11.

É um vector de substituição. O gene funcional da β -galactosidase é inactivado por inserção de DNA exógeno. Os fagos recombinantes podem ser detectados pela formação de placas incolores quando as bibliotecas em hospedeiros *lac*⁻ são plaqueadas na presença de X-gal e IPTG. Os não recombinantes formam placas azuis.

Como o tamanho dos vectores λ dificulta a posterior manipulação do DNA, é usual subclonar os insertos em vectores plasmídicos após isolamento dos clones de interesse (Short, et al., 1988). Foram desenvolvidos vectores de inserção λ que contém um fagemídeo³ de *E. coli* [pBluescriptSK(-)] inserido no genoma. Este pode ser excisado

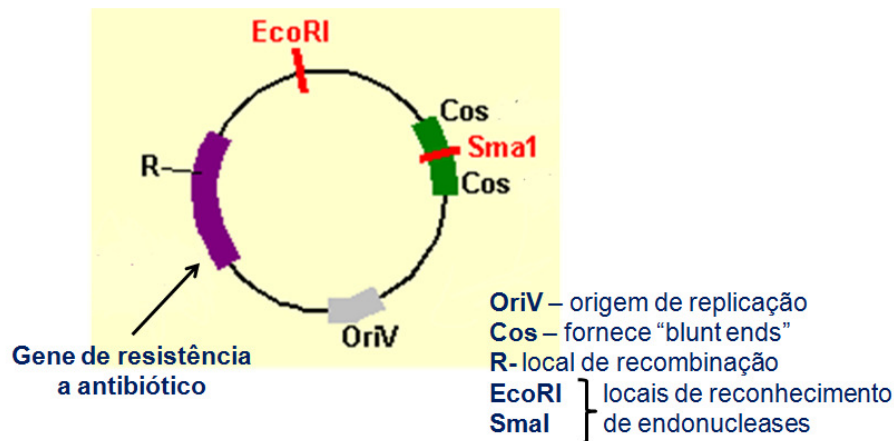
² gt – de “generalized transduction” vector, são vectores de transdução generalizada.

³ Os fagemídeos são vectores híbridos, entre um plasmídeo e o fago M13, um bacteriófago de *E. coli* com DNA de cadeia simples. Os da série pBluescript, cujas características se descrevem no **Anexo 1** são dos mais utilizados como vectores de clonagem em bibliotecas de cDNA.

pelos fagos auxiliares f1 ou M13 “in vivo”, evitando a necessidade de subclonar os insertos de DNA em vectores plasmídicos (Short, et al., 1988). Constituem a série Lambda ZAP (Agilent Technologies ®). Foram desenhados de forma que, com ajuda de um segundo fago (auxiliar – “helper”), se consiga que qualquer inserto clonado no fago seja excisado no interior do plasmídeo ainda em *E. coli*. Evita a sequenciação do DNA do fago, uma vez que cada placa de líse contém um inserto dentro de um plasmídeo pBluescript (Slish, 2010). Aceitam insertos até 10Kpb. No **Anexo 1** estão representados os mapas de um dos vectores Lambda ZAP e o do fagemídeo pBluescriptSK(+/-), assim como algumas das características destes vectores.

Após ter sido constatado que, ao contrário do processo de transformação, a transdução (infecção de bactérias por bacteriófagos) era compatível com a introdução de plasmídeos recombinados com insertos de grandes dimensões (Correia, et al., 2003), desenvolveram-se novos vectores de clonagem para bibliotecas de genómico, os cosmídeos, representados na **Figura 10**. São vectores híbridos obtidos pela inserção de extremos protuberantes *cos* do fago λ num plasmídeo (com origem de replicação, marcador de selecção e MCS). Permitem a clonagem de insertos de maiores dimensões do que os plasmídeos, com cerca de 40-45 Kb, porque os locais *cos* permitem o empacotamento do DNA exógeno (do inserto) no interior das cápsides virais (Wilson, 2003). A sua construção baseia-se no facto das proteínas que promovem o empacotamento do DNA do fago λ apenas necessitarem de reconhecer locais *cos* separados de 37-52Kpb. Os locais *cos* constituem a única parte do fago a ser introduzida no plasmídeo para a obtenção de um cosmídeo. Como não são transferidos genes virais, o vector replica-se em *E. coli* como um plasmídeo. O DNA amplificado pode ser facilmente extraído das células infectadas.

A clonagem em cosmídeos conjuga a elevada eficiência de infecção associada à clonagem no fago λ com a facilidade de manipulação de DNA plasmídico. A selecção de transformantes é feita em placas de Petri onde se adicionou antibiótico ao meio para possibilitar a detecção de clones transformados. Aos cosmídeos de primeira geração, com origem de replicação e genes de resistência a antibióticos derivados dos plasmídeos, e locais *cos*, seguiram-se os de segunda, com genes que permitiam a transfecção e selecção em células eucarióticas (Grosveld, et al., 1982) (Lau & Kan, 1983). Foram muito usados nos anos 80 e início dos anos 90 para clonagens de genómico.



(Wilson, 2003)

FIGURA 10 – CARACTERÍSTICAS BÁSICAS DE UM COSMÍDEO.

Embora alguns genomas como o de *Saccharomyces cerevisiae* tenham sido sequenciados sobretudo com utilização de cosmídeos para a construção das bibliotecas de DNA genómico necessárias, mesmo insertos de 40-45Kb são demasiado pequenos para projectos como os de sequenciação do genoma humano. Sendo assim, o desenvolvimento de vectores para clonagem de extensos fragmentos de DNA foi essencial para ser atingida uma taxa de conclusão de projectos de sequenciação de genomas elevada (Nicholl, 2008). Desenvolveram-se então novos vectores como os cromossomas artificiais de leveduras, YACs ("yeast artificial chromosomes"), os cromossomas bacterianos artificiais (BACs) e os cromossomas artificiais derivados do fago P1 (PACs), descritos no **Anexo 2**. Nestes, os insertos podem variar de 120Kpb a 1Mpb.

2.3 PREPARAÇÃO DO VECTOR

Após ter sido seleccionado, o vector a utilizar na biblioteca é preparado com uma enzima de restrição que produza extremos coesivos complementares aos gerados pela *Sau3AI*. Os dois DNAs são misturados na presença da T4 DNA ligase, de forma a obter ligações covalentes entre cada um dos fragmentos de DNA do organismo de interesse e uma cópia do vector. O resultado são milhares de moléculas recombinantes em que, em cada uma, o vector se encontra ligado a um pedaço diferente do DNA genómico original. Estas moléculas recombinantes são introduzidas num hospedeiro, obtendo-se uma biblioteca genómica: dispomos de milhares de clones independentes, cada um com uma molécula recombinante portadora de um pedaço de DNA genómico do organismo dador.

3. OBTENÇÃO DE BIBLIOTECAS DE cDNA

A primeira etapa da preparação de uma biblioteca de cDNA consiste na obtenção de RNA total existente na célula (tecido) de interesse, e posterior separação dos mRNAs poliadenilados por cromatografia de afinidade, em colunas com oligo-dT. Posteriormente a enzima transcriptase reversa é usada na síntese de cadeias de DNA complementares a cada uma das cadeias das moléculas de mRNA (**Figura 11**). Esta enzima tem a capacidade de efectuar *in vitro* a síntese de DNA a partir da cadeia de RNA, necessitando de um oligonucleótido iniciador (“primer”) com a extremidade 3’OH livre, para poder iniciar a síntese. Normalmente utiliza-se um pequeno oligonucleótido de desoxitimidinas (poli-dT) que hibrida na cauda de poli-A do mRNA, mas a síntese da primeira cadeia de cDNA também pode ser obtida através de “random primers”⁴ ou de primers específicos para o gene de interesse (**Figura 12**). Nesta reacção são formadas moléculas híbridas de DNA/RNA que são posteriormente usadas como molde na síntese da segunda cadeia de cDNA. Esta pode ser obtida de três diferentes maneiras, esquematizadas na **Figura 13**:

- Após a síntese da primeira cadeia de DNA, a enzima ribonuclease (RNase H⁵) vai digerir parcialmente a cadeia de mRNA, deixando alguns fragmentos. Estes vão servir de iniciadores à DNA polimerase I que completa a cópia da segunda cadeia de cDNA. Usa-se a enzima completa (o fragmento Klenow da enzima apenas exerce actividade polimerásica 5’-3’), com actividade exonucleásica 5’-3’, necessária para remover os resíduos de mRNA ainda ligados ao (ss) cDNA⁶ recém-formado. Estes são substituídos por desoxirribonucleótidos – síntese da cadeia de cDNA por substituição (**Figura 13A**). Os diversos fragmentos parciais obtidos são unidos por acção da T4 DNA ligase.
- Como alternativa para a síntese da segunda cadeia, pode-se deixar-se a RNase actuar mais tempo, ou fazer um tratamento com um álcali fraco (NaOH, por exemplo), que hidrolisa a cadeia de mRNA, mas não a de DNA (**Figura 13B e C**), até obter uma hidrólise completa da cadeia. Como qualquer outra molécula de DNA de cadeia simples, também a cadeia de cDNA formada a partir do mRNA molde pode dobrar-se sobre o seu extremo 3’ e originar um “hairpin loop” por emparelhamento casual por

⁴ Os “random primers” são oligodesoxirribonucleótidos (geralmente hexâmeros - d(N)₆) usados para anilhamento com mRNAs para a síntese de cDNA.

⁵ A RNase H é uma endonuclease que hidrolisa especificamente RNA em cadeias híbridas RNA-DNA.

⁶ (ss) cDNA – DNA complementar de cadeia simples.

complementaridade entre uma pequena sequência de bases da própria cadeia (**Figura 13B**). Neste caso, o próprio extremo 3' do "hairpin loop" funciona como "primer" para a síntese da segunda cadeia de DNA - "self-priming". Esta síntese pode ser feita com uma polimerase, ou com a própria transcriptase reversa (Hartl & Jones, 1998). Normalmente usa-se o fragmento Klenow da DNA polimerase I⁷ pelo facto de não ter actividade exonucleásica 5'-3', de modo a evitar a degradação do cDNA recém-formado (Croy, 1998). No final, a nucleasa S1⁸ (de *Aspergillus oryzae*), que cliva regiões de DNA em cadeia simples existentes em (ds) DNA, provoca a eliminação da estrutura em "hairpin loop" e converte a cadeia em cDNA de cadeia dupla característico (Hartl & Jones, 1998).

- O RNA é degradado pela RNase ou por um álcali fraco. Neste caso a síntese da segunda cadeia é precedida pela actuação da transferase terminal⁹, uma DNA polimerase que adiciona vários resíduos de um desoxirribonucleótido (por exemplo dGTP) aos extremos 3' da cadeia simples de cDNA (Lodish, et al., 2000) (Alberts, et al., 2002) (**Figuras 11 e 13C**). Esta enzima não precisa de iniciadores, apenas de extremos 3' livres. Posteriormente usa-se um oligo-dC sintético como iniciador para a DNA polimerase efectuar a cópia da segunda cadeia de cDNA, usando a primeira como molde – o processo é designado por "tailing and priming" O cDNA de cadeia dupla assim obtido - (ds) cDNA é a cópia de cada um dos mRNAs iniciais. Num dos extremos da cadeia apresenta uma região oligo dC-oligo dG, e noutra uma região 5' oligo dT-oligo dA (**Figura 11, etapas 1-5**) (Lodish, et al., 2000).

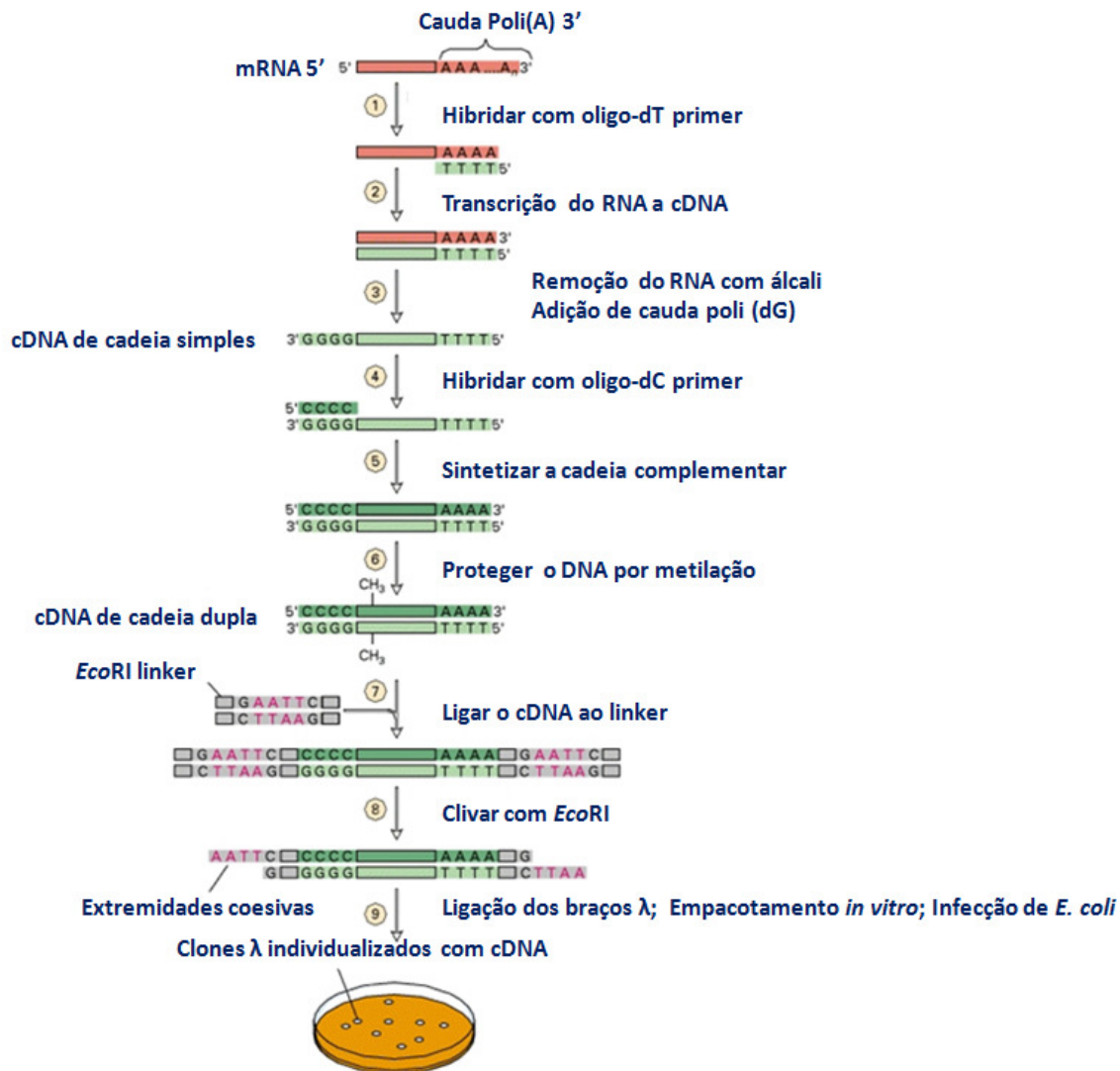
O segundo método (formação de "hairpin loop") tem como desvantagem importante o facto da nucleasa S1 provocar perda de parte da sequência do clone no extremo 5', mas o processo de "tailing and priming" permite a obtenção de moléculas de cDNA de cadeia completa ("full-length" cDNA) (Primrose et al., 2006).

Durante a transcrição reversa completa de uma molécula de mRNA, os cDNAs obtidos ("full-length" cDNA) contêm a sequência codificante completa da proteína de interesse. É então necessário preparar o cDNA para ser inserido num vector de clonagem, processo que depende do vector escolhido: plasmídeo ou fago.

⁷ O fragmento Klenow da DNA polimerase I é o fragmento de 75KDa resultante da proteólise da enzima com subtilisina. A enzima resultante tem actividade exonucleásica 3'-5', mas não a 5'-3' existente na enzima completa.

⁸ A nucleasa S1 degrada especificamente ácidos nucleicos de cadeia simples (DNA ou RNA), libertando 5'mononucleótidos.

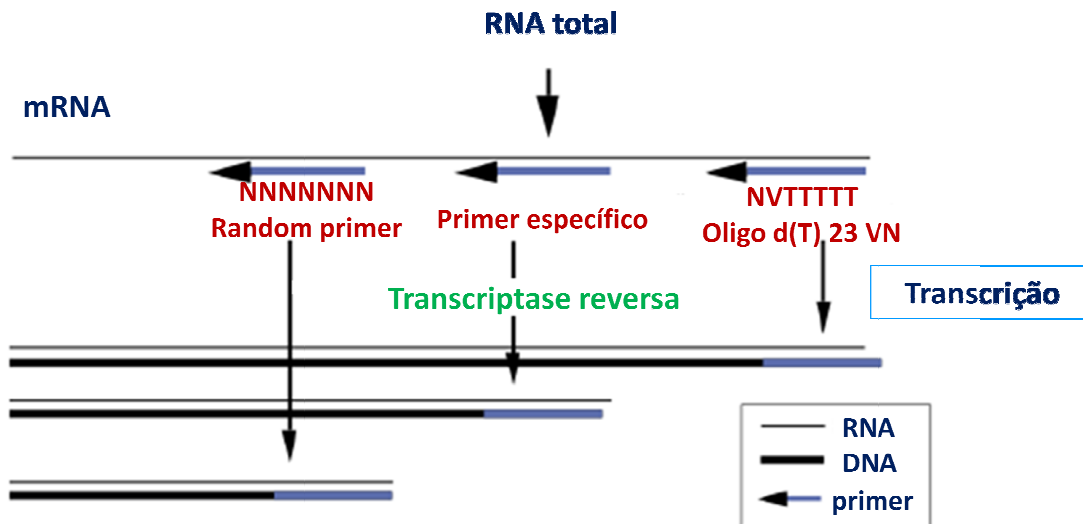
⁹ DNA nucleotidiltransferase (EC 2.7.7.31). Cataliza a adição de dNTPs a extremos 3' das cadeias de DNA; se apenas um dos desoxirribonucleótidos estiver disponível (por ex., dGTP), a enzima forma uma cauda homopolimérica (poli-G) em cada um dos extremos 3' da cadeia de DNA, quer os extremos sejam cegos ou protuberantes (Croy, 1998).



(Alberts, et al., 2002)

FIGURA 11 – OBTENÇÃO DE UMA BIBLIOTECA DE CDNA EM FAGO λ.

É usada uma mistura de mRNAs para obter as cópias de cDNA de cadeia simples correspondentes - (ss) cDNA (**etapas 1-2**). Estas cadeias simples de cDNA (a verde) são convertidas a cadeias duplas, e posteriormente tratadas com a metilase da *EcoRI* para evitar serem clivadas pela enzima (**etapas 3-6**). A cadeia protegida por metilação é ligada a adaptadores com as sequências de reconhecimento de *EcoRI* em cada um dos extremos, e posteriormente clivada com a enzima. Obtém-se extremos protuberantes (letras a vermelho, **etapa 8**), complementares aos do vector, permitindo a obtenção de viriões recombinantes recuperados em *E. coli* (**etapas 7-9**). (Lodish, et al., 2000).



(Wolfe, 2005).

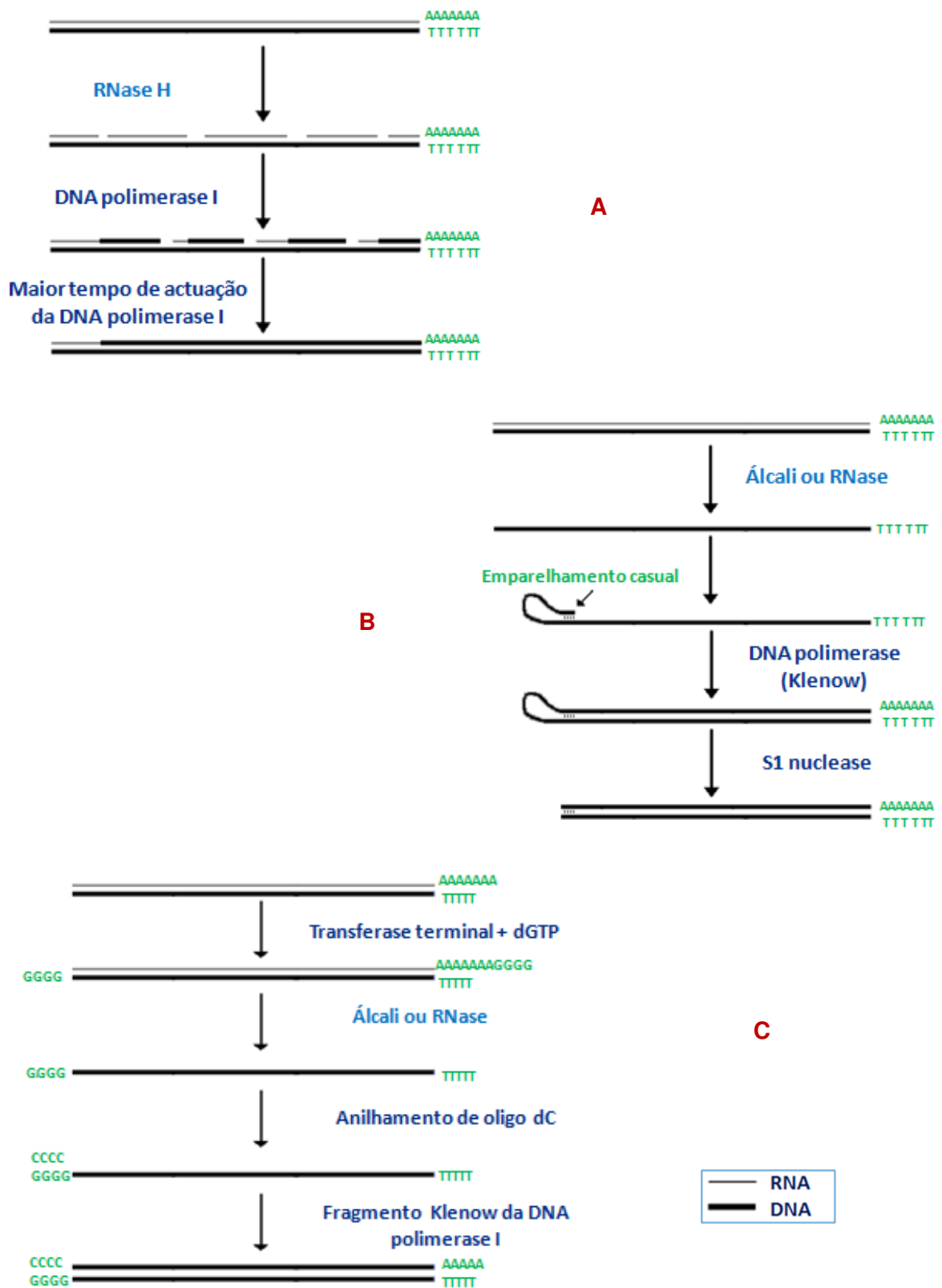
FIGURA 12 – PRIMERS USADOS NA SÍNTESE DA PRIMEIRA CADEIA DE cDNA.

O primer oligo d(T)₂₃ VN é usado para emparelhamento e sequenciação do mRNA adjacente ao extremo 3' (cauda poli-A) (Khan, Wilcox, Hopkins, & Sikela, Efficient double stranded sequencing of cDNA clones containing long poly(A) tails using anchored poly(dT) primers, 1991). V = A,G ou C ; N = A,G,C ou T.

Como os extremos das moléculas de cDNA não contêm sequências de reconhecimento de enzimas, tem que se preparar o cDNA para uma ligação eficiente ao vector escolhido. Em cDNA com extremos cegos (“blunt-ends”) pode fazer-se directamente a ligação ao vector previamente clivado com uma endonuclease que gere extremos cegos (*Sma*I, por exemplo), mas estas ligações são por norma bastante menos eficientes do que as ligações entre extremos protuberantes (Croy, 1998).

Podem-se criar extremos protuberantes em cadeias de cDNA com extremos cegos. Para isso incorporam-se em cada um dos extremos da cadeia uns oligonucleótidos sintéticos com 10-12pb que contêm sequências de reconhecimento de enzimas de restrição que, ao ser clivadas, produzem extremos protuberantes (normalmente escolhe-se uma endonuclease que possa ser usada para clivar ambos os DNAs: o cDNA e o DNA do vector de clonagem). A estas sequências chama-se adaptadores ou “linkers”. São na sua maioria palindrómicas, e acabam por hibridar entre si, formando uma cadeia dupla. Os adaptadores assim gerados são unidos aos extremos do cDNA pela T4 DNA ligase¹⁰ (na presença de ATP) (**Figura 11, etapa 7**), que faz a ligação entre extremidades cegas

¹⁰ Como os oligonucleótidos sintéticos não têm grupos fosfato 5' necessários à ligação, há que tratá-los previamente com T4 polinucleótido cinase, que fosforila os extremos 5' na presença de ATP.



Adaptado de (Wolfe, 2005)

FIGURA 13 – PROCESSOS DE SÍNTESE DA SEGUNDA CADEIA DE CDNA: A) RNASE H; B) SELF-PRIMING; C) TAILING AND PRIMING.

(“blunt-ends”) de cadeias duplas de DNA. Embora esta reacção de ligação não seja muito eficiente, pode ser optimizada usando uma concentração elevada de adaptadores (Alberts, et al., 2002).

A cadeia dupla de cDNA com adaptadores em ambos os extremos é posteriormente digerida com a enzima de restrição específica para o “linker” adaptado, originando moléculas de cDNA com extremos protuberantes (“sticky-ends”), que vão permitir a ligação ao vector por complementaridade entre os extremos, clivados com a mesma enzima. Para evitar que a parte interna da cadeia de cDNA também possa ser clivada pela enzima de restrição¹¹, é usual efectuar a metilação do cDNA previamente à ligação de adaptadores, por recurso a uma metilase específica. Esta enzima faz a modificação de bases específicas no seio da sequência de reconhecimento da endonuclease (**Figura 14**), evitando a clivagem do cDNA.

A ligação entre os extremos protuberantes do cDNA e os do vector (plasmídeo ou fago) é feita pela T4 DNA ligase. Por acção desta enzima ligam-se quaisquer extremos compatíveis, independentemente das moléculas envolvidas. Nos plasmídeos clivados com uma única enzima de restrição há grandes probabilidades de ocorrer religação – há dois extremos compatíveis na mesma molécula. Para evitar a religação do plasmídeo pode efectuar-se um tratamento com fosfatase alcalina (“CIAP” – “calf intestinal alkaline phosphatase”). Ao remover os grupos fosfato 5’, a fosfatase alcalina impede a religação de plasmídeos¹². De acordo com (Croy, 1998) e (Rinehart, 2005b), um método para efectuar a ligação entre vector e cDNA sem recorrer à ligase e a tratamento dos plasmídeos com fosfatase alcalina, consiste em adicionar caudas homopoliméricas¹³ complementares quer no cDNA (poli-dC) quer no vector após a clivagem (poli-dG), recorrendo à transferase terminal. A ligação entre moléculas postas em contacto permanece estável. Após transformação em *E. coli*, quaisquer “gaps¹⁴” ou “nicks¹⁵” existentes no plasmídeo recombinante serão reparados (Croy, 1998).

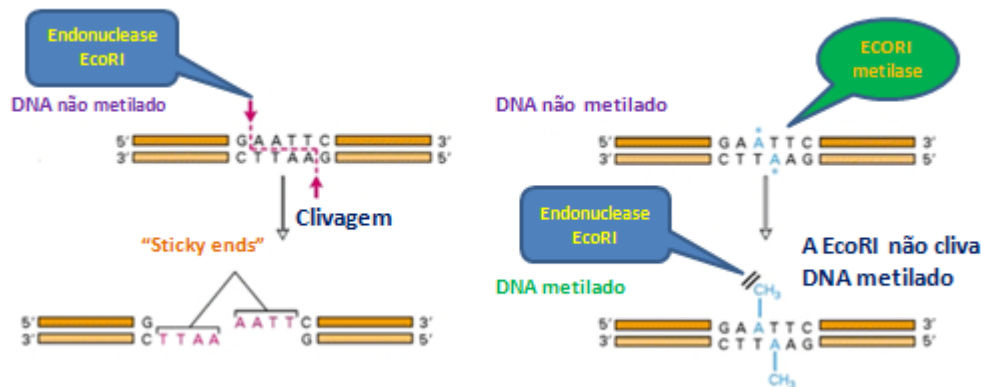
¹¹ A sequência de cDNA pode conter uma ou mais sequências de reconhecimento da endonuclease escolhida para criar “sticky ends”, o que provocaria a cisão da cadeia.

¹² Para haver ligação entre quaisquer extremos compatíveis é necessário haver grupos fosfato 5’ numa das moléculas, e grupos OH 3’ noutras.

¹³ Recorrendo à transferase terminal.

¹⁴ Um “gap” é uma descontinuidade numa das duas cadeias da dupla hélice de DNA, devido à perda de um ou mais nucleótidos.

¹⁵ Um “nick” é uma descontinuidade na molécula de DNA, por faltar uma ligação fosfodiéster entre nucleótidos adjacentes.



(Lodish, et al., 2000)

FIGURA 14 – MODO DE ACTUAÇÃO DE ENZIMAS DE RESTRIÇÃO E DE METILASES.

A) A *EcoRI*, enzima de restrição de *E. coli*, reconhece e cliva a sequência palindrômica de 6-bp GAATTC. Essa clivagem origina fragmentos de extremos protuberantes complementares de cadeia simples - "sticky ends".

B) Como forma de protecção do seu próprio DNA, as células bacterianas para além de produzirem enzimas de restrição produzem também metilases. Estas, ao adicionarem grupos metilo a bases azotadas presentes nas sequências de reconhecimento, impedem a clivagem por endonucleases. No exemplo, as células de *E. coli* contêm a enzima de restrição *EcoRI* mas também a sua metilase, que ao adicionar um grupo metilo às duas adeninas da sequência de reconhecimento, impedem o reconhecimento da sequência pela *EcoRI*, não ocorrendo clivagem.

Os vectores recombinantes, recuperados em células de *E. coli* ou em placas de lise (**Figura 11, etapas 7-9**), constituem uma biblioteca de cDNA, em que cada clone alberga um cDNA derivado de um único mRNA. Estas bibliotecas permitem obter genes diferencialmente expressos. Podem isolar-se genes expressos apenas em determinados tecidos, ou em determinados estádios de desenvolvimento ou em determinadas condições ambientais. Permitem a construção de bibliotecas por hibridação subtractiva.

Podem-se construir bibliotecas subtractivas a partir de duas populações de mRNAs obtidas em condições ligeiramente diferentes (por exemplo, células análogas cultivadas a temperaturas diferentes). A primeira cadeia de cDNA de uma das populações de mRNA é sintetizada e depois hibridada com as moléculas de mRNA da segunda população (em excesso). Há formação de moléculas híbridas entre cadeias de cDNA e mRNA complementares, que são retiradas por cromatografia em camada fina com hidroxapatite. Após a remoção de moléculas híbridas e síntese da segunda cadeia, as moléculas de cDNA específicas da primeira população tem maior representatividade do que inicialmente – há enriquecimento da biblioteca de cDNA em genes específicos. O

processo pode ser repetido. Estas bibliotecas ficam enriquecidas em genes expressos diferencialmente nas condições escolhidas.

Uma biblioteca de cDNA funciona como uma fotografia instantânea da actividade de uma célula num dado momento. Cada célula pode expressar entre 10000 a 50000 genes, e cada transcrito pode variar de uma ou poucas cópias por célula até cerca de 200000 (EcoArray, 2009). Por norma, em cada célula pode haver 10 a 20 genes muito abundantes (vários milhares de mRNAs cópia por célula), várias centenas de genes moderadamente abundantes (com várias centenas de mRNAs cópia por célula), e alguns milhares de genes raros, com apenas uma a algumas dúzias de mRNAs cópia por célula (EcoArray, 2009).

Desta forma, a sequenciação aleatória de uma biblioteca de cDNA “standard” ou não normalizada revela-se ineficiente na detecção de transcritos raros, porque os cDNAs de genes mais abundantes são sequenciados inúmeras vezes, originando redundância.

A normalização da biblioteca, obtida por processo semelhante ao de obtenção da biblioteca subtractiva¹⁶, diminui a prevalência de clones derivados de transcritos abundantes. Sendo assim, a normalização aumenta a eficiência da sequenciação aleatória e é determinante para a detecção de genes raros.

¹⁶ Mas apenas com uma população de mRNA, enquanto nas subtractivas pode haver duas ou mais populações de mRNA.

4. RASTREIO DE BIBLIOTECAS DE DNA

Uma vez disponível a biblioteca, há que adoptar técnicas que permitam identificar os clones recombinantes que contenham o(s) gene(s) de interesse.

Os métodos mais frequentes de rastreio e identificação de clones de interesse numa biblioteca de DNA são a hibridação de DNA com uma sonda marcada, ou a utilização de diversas estratégias de amplificação por PCR. O rastreio imunológico (uso de anticorpos face ao produto de interesse) e o rastreio da actividade biológica da proteína codificada são também métodos utilizados.

Uma sonda é um fragmento de DNA ou RNA usado para detectar sequências específicas de ácidos nucleicos por hibridação (dá-se a ligação entre duas cadeias simples antiparalelas de ácidos nucleicos por emparelhamento entre bases complementares. O DNA de cadeia dupla tem de ser desnaturado para poder formar cadeias simples, que permitam a hibridação).

O tamanho de uma sonda pode ser desde alguns pares de bases até centenas de kilobases (Lee, sd). As sondas curtas de oligonucleótidos são sintéticas e de cadeia simples. Normalmente 20 nucleótidos são suficientes para rastrear uma biblioteca. As “Expressed Sequence Tags” (ESTs) (cf. § 4.1.1) são muito utilizadas como molde para a obtenção de sondas de tamanho intermédio, com algumas centenas de pares de bases. As sondas de maiores dimensões são normalmente obtidas por clonagem e isoladas por PCR ou por restrição enzimática. Seguem-se alguns exemplos de utilização de sondas:

- Foi clonado um gene de um microrganismo (por exemplo, *Trichoderma harzianum*), e pretende-se avaliar se há genes homólogos num outro microrganismo (por exemplo, em *Phytophthora cinnamomi*). A sequência do gene clonado pode ser usada como sonda para detectar o gene homólogo.
- Se for conhecida apenas uma sequência parcial de um gene e se quiser determinar toda a sequência, a sequência parcial conhecida deve ser usada como sonda para detectar o clone de interesse.
- No caso de ser conhecida uma sequência conservada de um gene entre dois ou mais organismos, essa zona conservada pode ser usada para desenhar um oligonucleótido que vai servir de sonda. Esta vai servir para avaliar se há ou não genes análogos num outro microrganismo filogeneticamente próximo.

- Quando a sequência nucleotídica do gene que se pretende localizar na biblioteca é desconhecida, mas se conhece a sequência parcial da proteína por ele codificada, podem-se desenhar oligonucleótidos baseados na sequência conhecida. No entanto, como o código genético é degenerado, um aminoácido pode ser codificado por mais de um tripleto de bases azotadas do DNA - codão, sendo assim necessária a utilização de uma sonda constituída por uma mistura equimolar de oligonucleótidos degenerados¹⁷ que abranjam várias das combinações possíveis (**Figura 15; Tabela 2**).

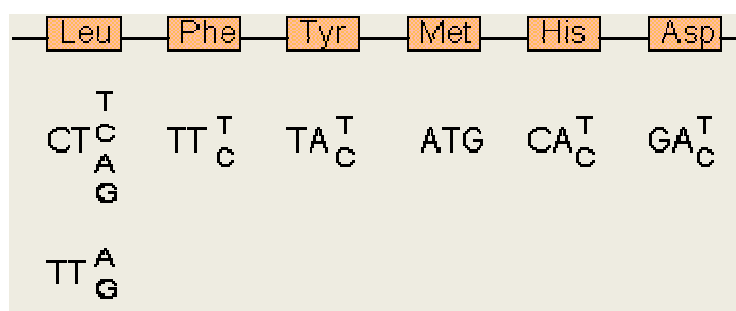


FIGURA 15 – CORRESPONDÊNCIA ENTRE UM PÉPTIDO E AS POSSÍVEIS SEQUÊNCIAS CODIFICANTES NA MOLÉCULA DE DNA(LEE, SD).

A = Adenina; T= Timina; G= Guanina; C= Citosina. No exemplo, à sequência peptídica Leu-Phe-Tyr-Met-His-Asp correspondem 96 (= 6 x 2 x 2 x 1 x 2 x 2) possíveis sequências de primers.

TABELA 2 – ALFABETO DE NUCLEÓTIDOS DEGENERADOS(PRESTON, 2003)

Letra	Especificação
R	Purina (A ou G)
Y	Pirimidina (C ou T)
K	G ou T
M	A ou C
S	G ou C
W	A ou T
B	G, C ou T
D	A, G ou T
H	A, C ou T
V	A, C ou G
N	A, C, G ou T
I	Inosina*

* Embora a inosina não seja um nucleótido, muitos investigadores incluem a sua molécula no desenho de primers degenerados por permitir o emparelhamento com A, T ou C, de uma forma indiscriminada, permitindo a obtenção de oligonucleótidos menos degenerados.

¹⁷ Com variação de bases azotadas numa ou várias posições da sequência.

As sondas de oligonucleótidos que contenham o codão para a metionina ou triptofano são particularmente valiosas – uma vez que cada um destes aminoácidos é codificado por apenas um codão, conseguem-se obter oligonucleótidos menos degenerados (Primrose, et al., 2006), e como tal mais específicos. A utilização da base azotada inosina pode também ser vantajosa – ao emparelhar com A, T ou C, reduz o número de combinações de primers necessárias.

Que tipo de marcação podemos efectuar nas sondas?

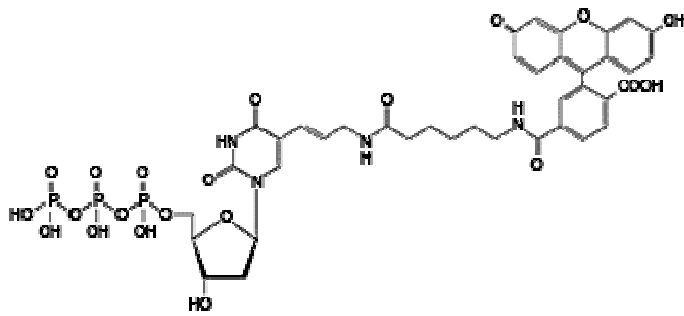
A marcação directa do DNA pode ser feita por isótopos radioactivos (geralmente ^{32}P) ligados a desoxirribonucleótidos. O $\alpha\text{-}^{32}\text{P}$ é incorporado na cadeia de DNA da sonda pelo fragmento Klenow. A sonda marcada ao emparelhar com o DNA de uma colónia emite radiações, detectadas numa película de raios X (de radiografias). Pelo facto de gerar um sinal detectável, o $\alpha\text{-}^{32}\text{P}$ constitui um grupo chamado repórter. Ao fim de um período de contacto de algumas horas a poucos dias entre as membranas e o filme, este é revelado. Se tiver havido hibridação vão aparecer pontos negros, que nos indicam a posição das colónias que continham DNA homólogo à sonda (autorradiografia).

Na marcação não radioactiva de sondas, as moléculas mais frequentemente usadas estão ligadas a uma desoxiuridina trifosfato (dUTP): digoxigenina, fluoresceína e biotina (**Figura 16**), e têm em comum o facto de poderem ser incorporadas no DNA a marcar através de diferentes processos enzimáticos (com a transcriptase reversa, com a *Taq* DNA polimerase, com o fragmento Klenow e com a DNA polimerase I). Nas reacções de síntese das novas cadeias de DNA incorporam-se na cadeia substituindo algumas das moléculas de dTTP. Estes métodos de marcação não radioactiva são indirectos, há ligação da sonda a um grupo modificador (digoxigenina, fluoresceína ou biotina). Este por sua vez, durante o processo de detecção de híbridos é reconhecido por anticorpos (no caso da digoxigenina e da fluoresceína) ou pela estreptavidina¹⁸, no caso da biotina. Estas moléculas encontram-se ligadas a uma enzima, a fosfatase alcalina. A fosfatase alcalina cataliza uma reacção colorimétrica com NBT/BCIP¹⁹ ou de bioluminescência com os substratos CSPD ou CDP-star²⁰. (ROCHE, 2005; ROCHE, 2011; ROCHE, 2010), permitindo a identificação do híbrido.

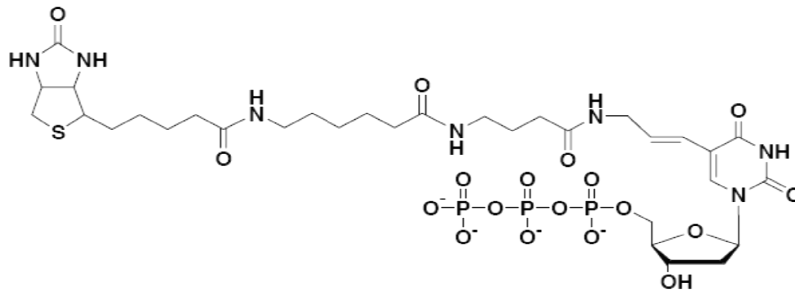
¹⁸ A estreptavidina é uma proteína de origem bacteriana, com capacidade de se ligar à biotina.

¹⁹ NBT/BCIP: substrato colorimétrico. Tem a vantagem de poder ser detectado a olho nú, sem ter de se recorrer a equipamentos de imagem.

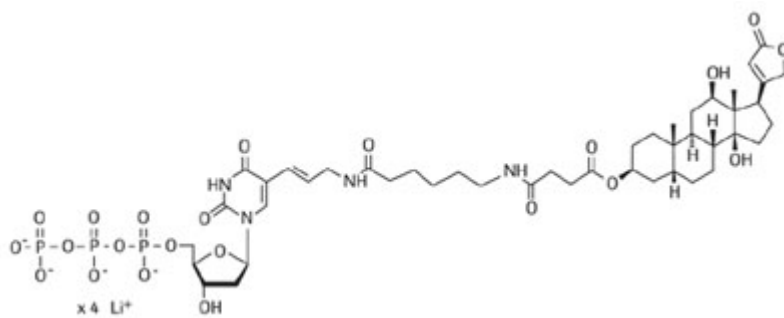
²⁰ A detecção da actividade da fosfatase alcalina com CDP-star ou CSPD resulta na emissão de luz, que pode ser detectada por exposição aos raios X, ou a um equipamento de aquisição de imagem. A reacção ocorre mais rapidamente e de forma mais intensa com o CDP-star.



Fluoresceína-12-dUTP



Biotina-16-dUTP

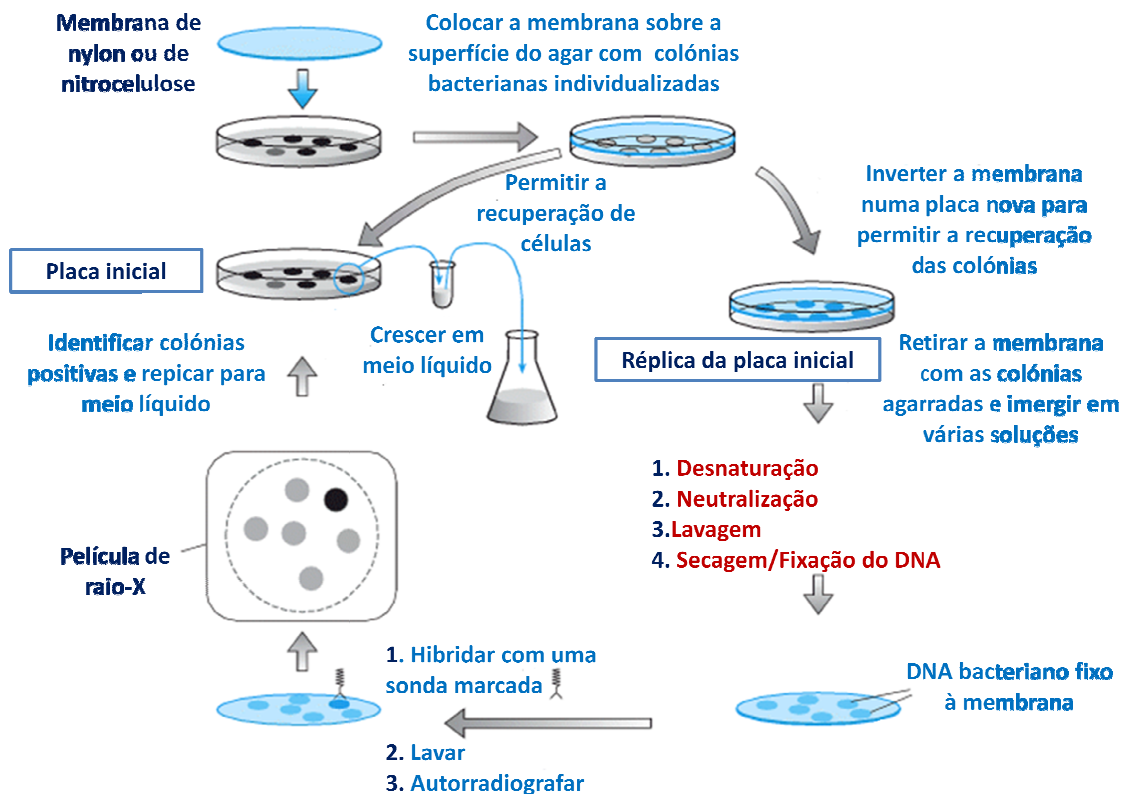


Digoxigenina-11-dUTP

FIGURA 16 - MOLÉCULAS USADAS NA MARCAÇÃO NÃO RADIOACTIVA DE SONDAS DE DNA.

4. 1 Rastreo por hibridação de DNA

O rastreo de colónias (ou placas de lise)²¹ por hibridação de DNA (“colony hybridization” ou “colony blot”) com uma sonda marcada, baseia-se na possibilidade de ligação entre cadeias simples do DNA a pesquisar e cadeias simples do DNA marcado (DNA sonda). Para rastrear uma biblioteca são efectuadas as seguintes etapas, cujo processo se encontra representado na **Figura 17**.



(Strachan, et al.,1999)

FIGURA 17 – RASTREIO DE COLÓNIAS POR HIBRIDAÇÃO DE DNA (“COLONY BLOT”).

²¹ Dependendo do vector utilizado na construção da biblioteca, assim se faz o rastreo de colónias ou placas de lise (“plaque lift”). No entanto a metodologia usada é análoga, pelo que, por uma questão de simplicidade se irá referir apenas uma das hipóteses.

Quando é usado este método de detecção, após a construção da biblioteca é usada a seguinte metodologia:

- São feitas réplicas das colónias obtidas em placa para membranas de nylon ou de nitrocelulose, colocadas sobre as placas.
- As réplicas das colónias são sujeitas a lise *in situ*, que em simultâneo provoca a rotura das células e a desnaturação do DNA.
- O DNA das colónias após ser desnaturado, é neutralizado, lavado e fixado às membranas por acção de radiações ultra-violetas ou por calor.
- Entretanto, fez-se a marcação do fragmento de DNA usado como sonda, e a sua desnaturação.
- Junta-se o DNA da sonda com as membranas que contêm o DNA da biblioteca. Se o DNA sonda tiver homologia com o DNA de algum dos clones, vai emparelhar-se com esse DNA para formar uma dupla hélice. O DNA dos restantes clones não consegue emparelhar-se com a sonda.
- As membranas são reveladas. Se tiver havido hibridação sonda-DNA de interesse encontraremos uma mancha no local correspondente à colónia ou colónias que hibridaram com a sonda. Com esta informação vamos às placas matrizes onde estão as colónias originais de células vivas, e localizamos os clones, com os quais prosseguimos a investigação.

No caso da infecção por bacteriófagos fazem-se “plaque lifts” – transferência de placas de lise para uma membrana e rastreio. O ensaio é análogo, mas são os vestígios de DNA do fago recombinante de cada placa de lise (**Figura 18**) que são transferidos para a membrana de nylon, e posteriormente detectados. Quando um clone é identificado, pode ser isolado e amplificado (**Figura 17**) para determinar a sua sequência.

É o método mais generalizado de rastreio de bibliotecas pela sua rapidez. Em bibliotecas de cDNA permite a identificação de clones com sequências incompletas que, de outra forma não poderiam ser detectados devido à impossibilidade de se expressarem (Primrose & Twiman, Principles of Gene Manipulation and Genomics, 2006).

O rastreio por hibridação revela-se muito versátil, na medida em que tanto se podem usar condições de elevada restringência²² (especificidade), detectando só clones com

²² Podem-se fazer variar as condições de restringência actuando a nível da temperatura usada na hibridação e na concentração salina das soluções usadas.

homologia completa com a sonda, como se podem baixar os níveis de restringência e detectar sequências apenas relacionadas com a sonda (por exemplo, quando se usa uma sonda de uma espécie para detectar genes homólogos noutra). A utilização de sondas desenhadas em domínios funcionais conservados de um gene pode detectar diferentes clones na mesma espécie, em condições de restringência baixa, levando à identificação de famílias de genes ²³(Primrose, et al., 2006). Desde que haja uma sonda disponível, a hibridação pode detectar qualquer sequência em qualquer biblioteca.

Em alternativa a esta metodologia, cada um dos clones da biblioteca pode ser repicado individualmente para placas de microtitulação com 96 micropoços (“microtiter”) e reorganizado numa membrana, de forma a distribuírem-se sob a forma de uma quadrícula regular - “grid”, obtendo-se “gridded libraries”. O processo é actualmente simplificado por recurso à robótica. A recolha de colónias em placa de Petri, colocação em placas de microtitulação, e destas para membranas é feita de uma forma automatizada (**Figura 19**).

Enquanto numa biblioteca tradicional se faz a detecção de sinais positivos por autorradiografia, sendo necessário alinhar as películas de raios-X com as placas iniciais para identificar os clones fornecedores de sinal, nas bibliotecas “gridded”, cada sinal positivo corresponde a uma conjugação de coordenadas (placa, linha, coluna – por exemplo, o clone 255A6 pertence à placa 255 linha A coluna 6), que identificam os clones na placa “microtiter” inicial. Neste caso as réplicas podem também ser facilmente obtidas através do trabalho de robots, facilitando o seu fornecimento a outros laboratórios para “screening” de genes de interesse. Desta forma, uma biblioteca pode ser rastreada por diferentes laboratórios simultaneamente, e os dados podem ser centralizados (Wolfe, 2005).

Os “microarrays” ou “chips” de DNA (em que o DNA da biblioteca foi transferido para um suporte sólido, de vidro ou silicone, de acordo com um padrão de distribuição regular) são muito usados na pesquisa de genes possivelmente envolvidos em doenças. Neste caso o rastreio é feito em simultâneo com sondas de duas origens, cada uma marcada com diferentes fluorocromos. Clones expressos em ambos os tecidos originam uma fluorescência final correspondente à mistura de ambos os fluorocromos, e os diferencialmente expressos fluorescem numa cor próxima da cor pura de uma das sondas (“screening” diferencial). O processo está representado na **Figura 20**.

²³ Grupo de genes estreitamente relacionados, que originam proteínas semelhantes.

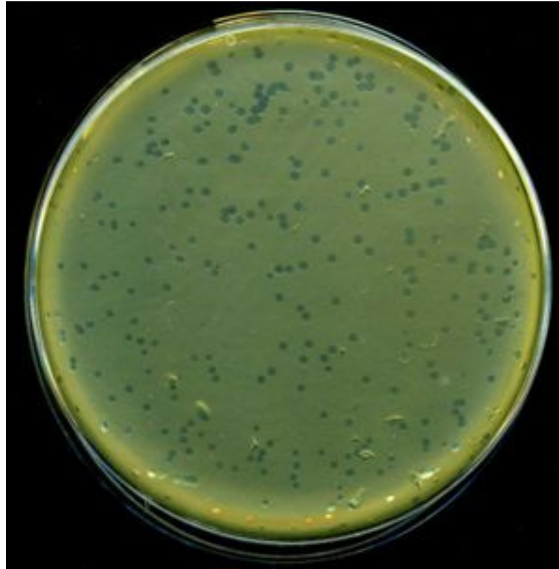


FIGURA 18 – PLACAS DE LÍSE DO FAGO λ EM *E. COLI*.

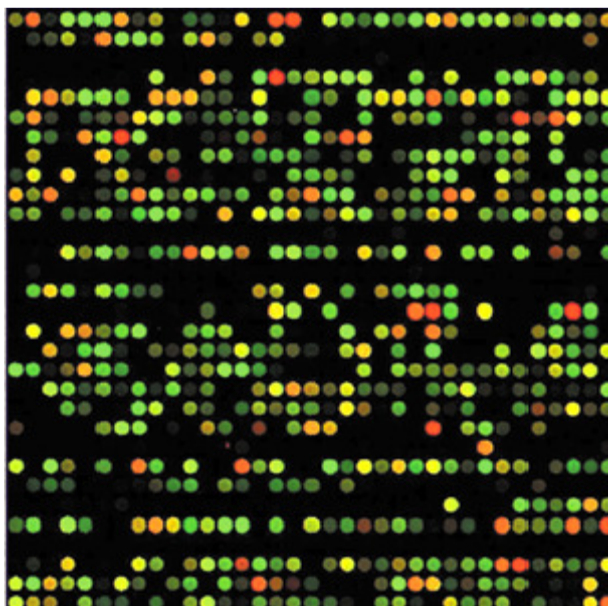
Na ausência do fago o crescimento de *E. coli* forma uma camada contínua, e a superfície do meio de cultura tem um aspecto homogéneo, mais ou menos opaco. Na presença do fago, inicia-se a lise das células infectadas, e também das anexas, provocada pela descendência do fago. Devido ao processo de infecção e lise não há crescimento bacteriano nas zonas à volta das colónias inicialmente infectadas, correspondentes às zonas translúcidas que aparecem na superfície do crescimento em meio semi-sólido. Estas zonas são designadas por placas de lise. Cada placa resulta da infecção inicial de um bacteriófago. (Hartl & Jones, 1998)



(Wolfe, 2005)

FIGURA 19 – TRANSFERÊNCIA DE COLÓNIAS ROBOTIZADA NO INSTITUTO SANGER.

Entre as placas de Petri com colónias (à direita) e as placas “microtiter” com meio nutritivo para onde são individualmente transferidos os clones (à esquerda), há tabuleiros com etanol a 70% para a esterilização das agulhas (situadas num contentor por detrás da lâmpada; este contentor move-se e situa uma agulha por cima de uma colónia) (Wolfe, 2005).



(NCBI, 2007)

FIGURA 20 – MICROARRAYS.

Cada mancha do “microarray” representa um gene. Significado das cores obtidas no rastreamento de um “microarray” por laser.

Verde: representa a cor do cDNA ou DNA controle (por ex. de um tecido sem alterações).

Vermelho: representa a cor do cDNA ou DNA extraídos de um tecido doente, a amostra.

Amarelo: representa a expressão combinada e equitativa em ambos os tecidos.

Preto: representa áreas do “microarray” onde nem o DNA controle, nem a amostra hibridaram com o DNA alvo.

4.1.1 OBTENÇÃO DE “EXPRESSED SEQUENCE TAGS” - ESTS

No rastreamento de bibliotecas de genômico é frequente utilizarem-se sondas de cDNA. A utilização de moléculas de cDNA de cadeia completa (“full-length” cDNA), em que a sequência nucleotídica foi confirmada sem ambiguidades por várias etapas de sequenciação, permite a determinação inequívoca dos códons de início e de stop, e de todas as fronteiras intrão-exão (Primrose, et al., 2001). No entanto, muitas vezes a sua obtenção é difícil. Há análises em que a sua utilização se torna desnecessária - mesmo pequenos fragmentos de cDNA são suficientes para de forma inequívoca identificar genes, estabelecer a sua localização num mapa físico, e proceder à análise da sua expressão (Primrose, et al., 2001).

O desenvolvimento de tecnologias de sequenciação de elevado desempenho, permite que milhares de clones de bibliotecas de cDNA sejam seleccionados aleatoriamente e sujeitos a sequenciação num sentido (5'→3'; 3'→5') ou em ambos, originando fragmentos

de cDNA de 200-500pb conhecidos por “Expressed Sequence Tags” (ESTs) (NCBI, 2004).

A partir de cada cDNA isolado sequenciam-se algumas centenas de bases nos dois extremos da cadeia. A sequenciação do princípio da cadeia origina uma 5’ EST (que por norma faz parte de uma ORF). As regiões 5’ não apresentam grande variabilidade no interior de uma família de genes (NCBI, 2004). A sequenciação da parte terminal do cDNA produz uma 3’ EST (formada a partir da extremidade 3’ de um transcrito, e portanto inclui UTRs – “untranslated regions”), cuja sequência é normalmente menos conservada entre espécies.

As ESTs são uma forma rápida e barata de descobrir novos genes, construir mapas genómicos e obter dados de regulação e expressão génica, tendo-se revelado uma ferramenta muito útil na detecção de alguns genes, nomeadamente os responsáveis pela doença de Alzheimer e pelo cancro do cólon (NCBI, 2004).

4.2 RASTREIO POR ENSAIO IMUNOLÓGICO

Se não dispomos de uma sonda de DNA, e se o gene que estamos a procurar na biblioteca se expressa até ao nível de proteína, podemos recorrer a um ensaio imunológico (desde que tenhamos anticorpos para a proteína em questão). As suas primeiras fases são análogas às da hibridação *in situ*, anteriormente descrita:

- Fazem-se réplicas de todas as colónias da biblioteca em membranas de nylon ou nitrocelulose colocadas na superfície das placas. As colónias de réplica são lisadas *in situ*, e as proteínas libertadas são fixadas à membrana. Esta é tratada com o anticorpo específico que reconhece a proteína - anticorpo primário, não marcado.
- As membranas são lavadas para retirar o excesso de anticorpos e tratadas com um segundo anticorpo. Este anticorpo secundário está unido a uma enzima como a fosfatase alcalina, e reconhece o anticorpo primário. Após a segunda série de lavagens adiciona-se um substrato incolor que é reconhecido pela enzima ligada ao segundo anticorpo, e que origina um composto colorido. Sendo assim, no local de reconhecimento da proteína pelo anticorpo primário, originar-se-á uma mancha de cor devida ao produto gerado pela enzima unida ao anticorpo secundário.

- Voltamos às placas originais e localizamos a colónia que deu positivo.

4. 3 RASTREIO DA ACTIVIDADE DA PROTEÍNA

Esta estratégia pode-se aplicar quando o que estamos a procurar é um gene que determina uma actividade que não está presente no microrganismo hospedeiro onde se construiu a biblioteca. Alguns exemplos:

- Imaginemos que construímos em *E. coli* uma biblioteca genómica de um organismo produtor de α -amilase, ou de endoglucanase, ou de β -glucosidase (a *E. coli* não produz por si mesmo nenhuma destas enzimas). Pois bem, uma biblioteca completa pode ser rastreada facilmente para estas actividades enzimáticas replicando-a em meios com os substratos adequados, por detecção visual de halos ou alterações colorimétricas correspondentes à degradação desses substratos (Pareja, 2005).
- Se estamos a procurar genes relacionados com biosíntese de algum aminoácido ou base azotada, fazemos uma biblioteca num organismo hospedeiro auxotrófico²⁴ para esse metabolito, e plaqueamos em meio mínimo, de forma a só crescerem os clones com insertos de DNA que codifiquem a enzima biosintética que falta ao hospedeiro, gerando o fenótipo prototrófico.

4. 4 RASTREIO BASEADO EM TÉCNICAS DE PCR

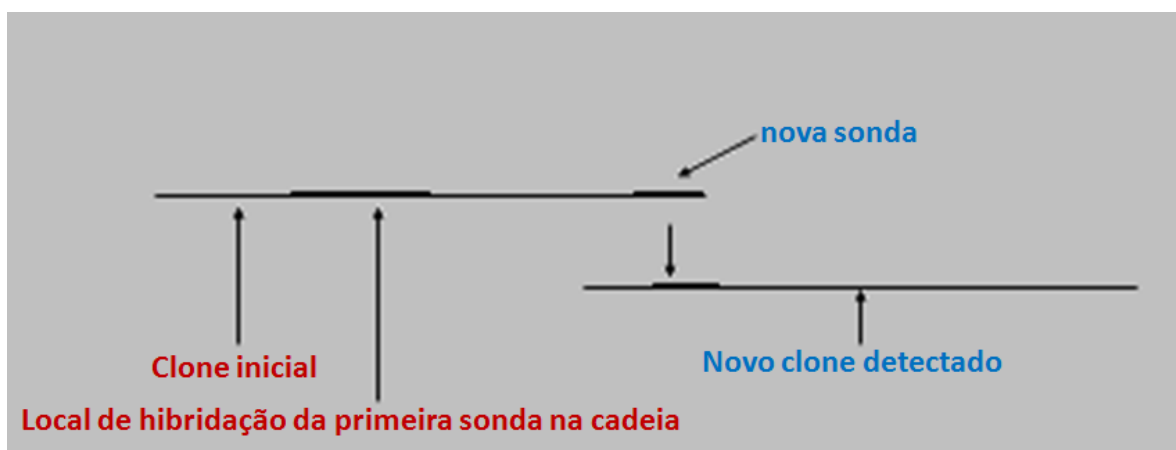
O método de rastreio de bibliotecas de genómico com sondas de cDNA não permite a identificação directa de clones que contenham intrões de tamanho superior à capacidade do vector. Neste caso, tem que se recorrer a “chromosome walking”, ou seja caminhar no cromossoma (**Figura 21**). O princípio do “chromosome walking” baseia-se na sobreposição de clones da biblioteca de genómico, permitindo a obtenção de uma sequência contígua a outra previamente conhecida. Permite o isolamento de genes com função desconhecida, mas dos quais se conhece a localização (Primrose, et al., 2001). Num dos extremos da sequência conhecida de um clone escolhe-se uma sequência que vai funcionar como sonda para a detecção de novos clones, que contenham a sequência

²⁴ Auxotrofia – incapacidade de sintetizar determinado composto orgânico.

da sonda e a adjacente, permitindo alargar o conhecimento da sequência nucleotídica do cromossoma por sequenciação. Se necessário repete-se o processo o número de vezes necessárias para determinar toda a sequência em falta. A maior desvantagem deste método é a necessidade de usar um grande número de sondas para determinar uma sequência longa.

Nos últimos anos surgiram várias metodologias alternativas ao “chromosome walking” baseadas na técnica de PCR, com o objectivo de identificar regiões de genómico adjacentes a outras previamente conhecidas de uma forma mais célere e económica (Chen, et al., 2004). Uma dessas técnicas é a HE-TAIL PCR (“High-Efficiency Thermal Asymmetric Interlaced PCR”), uma técnica de TAIL-PCR descrita por (Michiels, et al., 2003), e que é uma modificação da TAIL-PCR descrita por (Liu, et al., 1995), e por nós utilizada para a elucidação de dois genes em *T. harzianum* (cf. § 5). De acordo com os primeiros autores esta técnica é adequada para determinar sequências de DNA adjacentes a outras previamente conhecidas [para completar sequências pequenas de exões ou de ESTs, permitindo a clonagem de ORFs completas (“full length ORFs – open reading frames”) sem haver necessidade de recorrer ao “screening” de bibliotecas de cDNA, ou para identificar regiões do promotor].

Baseia-se no uso de oligonucleótidos específicos (de cerca de 26bp ou mais, com maior temperatura de anilhamento e estabilidade), desenhados na vizinhança da extremidade da sequência de DNA a completar, e de oligonucleótidos degenerados de 16bp, sujeitos a um programa de ciclos térmicos em várias etapas, em que há alternância de ciclos de



Adaptado de (Wolfe, 2005)

FIGURA 21- CAMINHAR NO CROMOSSOMA (“CHROMOSOME WALKING”).

O fim de uma sequência conhecida é usado como sonda para detectar um fragmento de DNA adjacente. A parte final deste será usada para detectar o próximo, e assim sucessivamente.

baixa e de elevada restringência, que permitem a amplificação das regiões que flanqueiam a região de DNA conhecida. Os ciclos de PCR a temperaturas mais elevadas favorecem o anilhamento dos oligonucleótidos específicos, enquanto os de temperaturas mais baixas permitem o anilhamento de ambos. Para que a técnica seja eficiente, os oligonucleótidos específicos devem ter uma T_m^{25} superior a 70 °C e um conteúdo em (G+C) de 40-50%. Várias outras metodologias de pesquisa de genes baseadas na técnica de PCR foram descritas ao longo dos últimos anos, mas seria exaustivo enumerá-las, assim como referir vantagens e inconvenientes de cada uma delas.

No rastreio de bibliotecas de cDNA a metodologia baseada em PCR mais utilizada é a “colony-PCR”(Otsuka, et al., 2004), ou seja, a PCR feita a partir da colónia bacteriana sem prévia extração de DNA plasmídico. De acordo com (Malone, et al., 2006), nesta técnica todas as colónias de uma biblioteca de cDNA são rastreadas por PCR, com primers desenhados em domínios proteicos conservados. De acordo com estes autores o método apresenta inúmeras vantagens:

- Os clones falso-positivos são facilmente detectados pelo tamanho do produto de amplificação
- Podem ser analisados centenas de clones por dia, reduzindo em muito o tempo necessário para o rastreio completo da biblioteca
- Permite a utilização simultânea de vários conjuntos de primers na mesma reacção de PCR, possibilitando a pesquisa de vários genes em simultâneo.

Os clones positivos são posteriormente sequenciados para avaliar se a sequência do gene está completa. Em caso negativo, poderá ter de se tentar completar a sequência por uma estratégia de PCR inversa.

²⁵ Melting temperature (T_m): temperatura para a qual a dupla hélice de DNA se dissocia em duas cadeias simples.

5. EXEMPLO DE APLICAÇÃO

5.1 METODOLOGIAS USADAS NA PESQUISA DA SEQUÊNCIA NUCLEOTÍDICA DE NOVOS GENES EM *TRICHODERMA HARZIANUM*

O género *Trichoderma* abrange várias espécies usadas no controlo biológico de fungos fitopatogénicos do solo, causadores de prejuízos elevados em diversas culturas de grande importância económica (Vizcaíno, et al., 2006a). A sua acção micoparasítica parece ser devida a uma acção sinérgica envolvendo a produção de antibióticos e de enzimas de degradação das paredes celulares²⁶ dos fungos parasitados (Schirmbock, et al., 1994), nomeadamente quitinases, β -1,3 e β -1,6 glucanases, proteases,.. Uma das espécies mais usadas em biocontrolo é o *T. harzianum*. Para estudar o genoma deste fungo criou-se um projecto de genómica funcional, “TrichoEST”, com o objectivo de detectar genes envolvidos nos mecanismos de controlo biológico.

No âmbito do projecto construíram-se oito bibliotecas de cDNA a partir de diferentes condições de crescimento²⁷ de *T. harzianum* CECT 2413²⁸, na tentativa de simular condições naturalmente existentes no solo (stress de nutrientes e micoparasitismo), e na expectativa de se conseguirem identificar mais genes do que se apenas se tivesse obtido uma única biblioteca de cDNA (Vizcaíno, et al., 2006b). Destas bibliotecas foram geradas 7283 ESTs a partir de 8710 reacções de sequenciação. Cada EST apresentava um tamanho médio de 566 nucleótidos, tendo sido detectadas 3478 sequências únicas (Vizcaíno, et al., 2006b). Estas sequências foram comparadas com as sequências existentes nas bases de dados não redundantes (nr) de proteínas da NCBI²⁹, de acordo com o algoritmo BLASTX(Altschul, et al., 1997). Destas, cerca de 81,4% apresentavam homologia com proteínas existentes no GenBank³⁰.

A tarefa do nosso grupo de trabalho consistiu em pesquisar genes envolvidos na síntese de proteínas com actividade lipolítica ou esterásica (carboxilesterases), para

²⁶ CWDE – cell-wall degrading enzymes.

²⁷ Presença de paredes celulares de *Pseudomonas syringae* e *Botrytis cinerea*; défice de glucose, quitina, presença de paredes celulares de *B. cinerea* e *Penicillium digitatum*, concentração de sais aumentada 100 vezes; glucose; défice de glucose em meio sólido; presença de quitina coloidal (meio sólido); presença de quitina e défice de azoto; em meio de confrontação com *Rhizoctonia solani* (meio sólido); PGA, CMC, pectina, xilano e glucose.

²⁸ CECT - Colección Española de Cultivos Tipo; 2413 foi o biótopo escolhido.

²⁹ <http://blast.ncbi.nlm.nih.gov/Blast.cgi>

³⁰ <http://www.ncbi.nlm.nih.gov/genbank/GenbankOverview.html>

posteriormente se poder avaliar se de alguma forma essas enzimas estão implicadas nos mecanismos de micoparasitismo. Após a sequenciação das ESTs obtidas pela empresa “NBT” (Newbiotechnic, S.A., Sevilla)³¹ foram seleccionadas nove que, de acordo com o programa BLASTX apresentavam alguma homologia com carboxilesterases, triacilglicerol lipases, lipases ou fosfolipases: EST-299.T3, EST-1279, EST-2104, EST-2126, EST-2957, EST-5482 e EST-9632.

Estas ESTs tinham sido obtidas por anilhamento de “primers” universais no sentido directo ou reverso a partir do cDNA clonado em plasmídeos. Estes fragmentos foram sequenciados uma única vez (“singlets”) no sentido 5'→3' (ou no sentido 3'→5') ou sequenciados em ambos os sentidos, embora a maior parte das vezes não o número de vezes suficiente para que as sequências lidas a partir das duas cadeias (directa e reversa complementar) se sobrepusessem efectuando um “contig”³².

Numa primeira etapa as sequências das nove ESTs foram limpas (ou seja, analisadas para a detecção de sequências pertencentes ao vector no “site” da NCBI a partir do programa Vecscreen³³) e editadas usando o programa BioEdit³⁴, tendo-lhes sido retiradas as sequências do vector.

Numa segunda etapa, a partir de uma zona sequenciada sem ambiguidades (**Figura 22**), e situada a cerca de 100pb da extremidade conhecida de cada ESTs, desenhou-se um novo primer em cada sequência, com o objectivo de ampliar a sequência de bases conhecidas, na expectativa de que algumas ESTs contivessem toda a sequência de um gene. Nesse caso não seria necessário efectuar o rastreio da biblioteca de genómico.

Após cada etapa de sequenciação conheceu-se um pouco mais da cadeia o que se traduziu em aumento ou perda de homologia com carboxilesterases ou lipases, de acordo com os resultados dos programas de pesquisa de similaridade BLASTX³⁵ ou FASTX³⁶. Das nove ESTs iniciais, e após algumas etapas de sequenciação/alinhamento de sequências³⁷ prosseguiu-se apenas com duas: a EST-1279 e a EST-9632 (as ESTs 1279 e 1285 revelaram ser a mesma sequência, assim como as ESTs 2104 e 2104; nestas últimas, tal como nas ESTs 2126, 299.T3, 2957 e 5482 verificou-se perda de homologia com carboxilesterases/lipases após elucidação da sequência nucleotídica.

³¹ A partir das bibliotecas de cDNA de *T. harzianum* CECT 2413 clonadas em pBluescript SK+.

³² Abreviatura de sequência contígua.

³³ <http://www.ncbi.nlm.nih.gov/VecScreen/VecScreen.html>.

³⁴ <http://www.mbio.ncsu.edu/bioedit/bioedit.html>

³⁵ <http://blast.ncbi.nlm.nih.gov/>

³⁶ <http://www.ebi.ac.uk/Tools/sss/fasta/>

³⁷ Recorrendo ao programa Sequencher @ 5.0. Gene Codes Corporation.<http://genecodes.com/>



FIGURA 22- CROMATOGRAMA DE SEQUENCIAÇÃO.

As primeiras 88 bases constituem uma zona mal sequenciada. As correspondentes às posições 116, 151 e 174 constituem ambiguidades de sequenciação, identificadas com N. A,C,T,G, correspondem às bases desoxirribonucleotídicas constituídas por adenina, citosina, timina e guanina, respectivamente.

Verificou-se que a EST-1279 possuía 1168bp após duas etapas de sequenciação. Por comparação com sequências existentes na EMBL através do programa FASTA³⁸, observou-se que apresentava um E value³⁹ de $1,8e^{-80}$ e 61,8% de identidade com uma presumível lipase de *Giberella zeae* = *Fusarium graminearum* (Q4ING8-Gibze), de 408 aminoácidos. Por comparação entre as duas sequências, verificou-se que deveriam faltar cerca de 300bp correspondentes ao início da ORF.

A EST-9632 continha uma sequência de 1627pb. A sequência da proteína deduzida da ORF traduzida apresentou 52,1% de identidade (para um E-value de $6,2e^{-105}$) com a proteína Q4WBJ7-ASPFU, uma presumível lipase extracelular de *Aspergillus fumigatus*, de acordo com o algoritmo de FASTX, e 53% de identidade e $1,0e^{-171}$ para a mesma proteína, de acordo com o algoritmo de BLASTX⁴⁰. Também aqui se constatou a falta de cerca de 52pb do início da ORF.

³⁸ <http://www.ebi.ac.uk/Tools/sss/fasta/>

³⁹ “(E) value”, de “Expect value” é o parâmetro que descreve o número de resultados que podem surgir ao acaso numa busca feita numa base de dados de determinada ordem de grandeza. Quanto mais próximo de zero, mais significativo é o resultado.

⁴⁰ Actualmente, devido ao aumento de sequências nucleotídicas identificadas, o melhor score pertence à proteína de D4D500-TRIVH, uma carboxilesterase de *Tricophyton verrucosum*, com 53,2% de identidade (E-value de $1,4e^{-119}$) de acordo com o algoritmo de FASTX, e 53% de identidade a mesma proteína, de acordo com o de acordo com o algoritmo de BLASTX (E-value de $1,1e^{-171}$).

Uma vez que as sequências contidas em cada uma das ESTs não correspondiam a ORFs completas, optou-se por efectuar o rastreio de uma biblioteca de genómico de *Trichoderma harzianum* CECT 2413 clonada no vector λ GEM11 (Promega) e amplificada em *E. coli* LE392, também ela obtida no âmbito do projecto “TrichoEST”. Para isso desenharam-se dois “primers” em cada uma das ESTs (1279 e 9632), e por PCR fez-se a marcação das respectivas sondas com digoxigenina. Para confirmação da marcação efectuou-se um “dot blot”.

As condições de infecção de *E. coli* estão descritas no Anexo 3. Após o aparecimento das placas de líse procedeu-se ao rastreio da biblioteca descrito em § 4.1. Na revelação das membranas detectaram-se vários sinais correspondentes a locais de hibridação com a sonda (**Figura 23**). Após a localização das placas de líse nas placas matrizes iniciais, prosseguiu-se a investigação extraindo DNA a partir dos clones positivos.

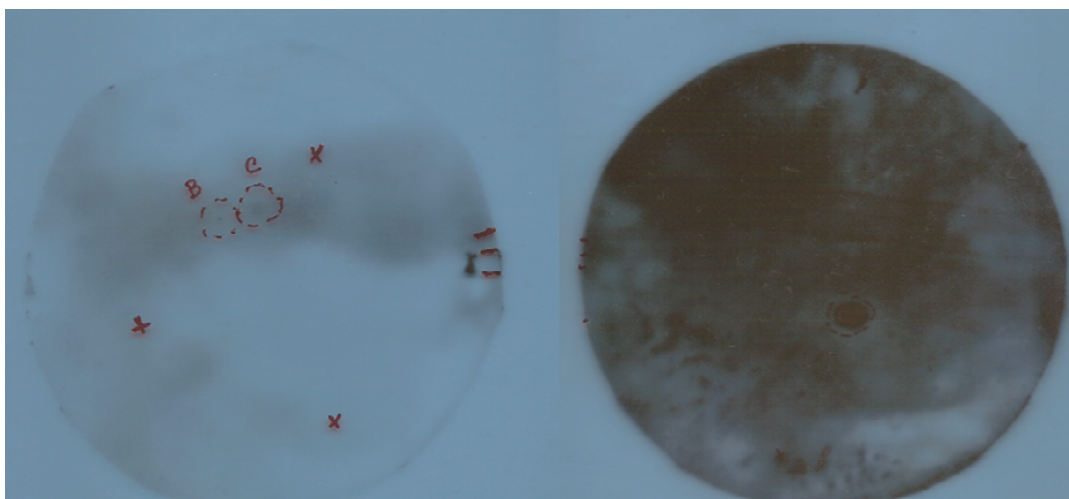


FIGURA 23 – AUTORADIOGRAFIA DE MEMBRANAS DE HIBRIDAÇÃO.

Cada mancha localiza um clone provável. As cruces e as linhas marcadas na periferia do círculo servem apenas para orientação da membrana relativamente à placa original.

No entanto, e após várias tentativas de extracção de DNA dos fagos recombinantes por recurso ao Qiagen Lambda Mini “kit”®, não se conseguiu obter DNA com qualidade suficiente para se obterem cromatogramas de sequenciação com uma qualidade que permitisse a identificação da sequência neles contida. Este é um dos principais problemas apontados à utilização de fagos como vectores de clonagem para bibliotecas

de genómico. Na tentativa de completar as sequências correspondentes aos quadros de leitura aberta dos dois genes, recorreu-se à HE-TAIL PCR descrita no § 4.4.

Para obter a sequência completa do gene *lip1* (obtido a partir da EST-9632) foram desenhados na sequência incompleta da ORF de *lip1* (**Figura 24**) os oligonucleótidos específicos de 26pb **nrev2** (5'- GGA ACC GCT TGA AGC ATC CGA GGG CC -3'), **nrev1** (5'- ATG CTG GTG GTA GGC CAC TGG GAG T -3'), **nrev0a** (5'- ATC CTA CTG GCG GCT GCG CGT AAG G -3') e **nrev0b** (5'- GAA TAG CGC AAG GCC ACG GTT GGA CC -3'), com Tm entre 70 e 72 °C. Foram usados os oligonucleótidos degenerados **R1** (5'- NGT CGA SWG AMA WGA A -3'), **R2** (5'- GTN CGA SWC ANA WGT T -3'), **R3** (5'- WGT GNA GWA NCA NAG A -3') e **R4** (5'- NCA GCT WSC TMT SCT T -3'), descritos por (Michiels, Tucker, van den Ende, & van Laere, 2003).

As três séries de PCR foram realizadas num termociclador MyCycler Thermal Cycler (BIORAD), de acordo com as condições térmicas programadas e ciclos descritos na **Tabela 3**. De referir que há uma única etapa combinada de anilhamento-extensão a 65 °C, nas reacções de PCR primária e secundária. As condições químicas são também referidas na Tabela 3.

A separação e identificação de fragmentos de DNA dos produtos das reacção terciária, incluindo os controlos, foram realizadas por electroforese em géis de agarose (0,8%). Consideraram-se como potencialmente interessantes os fragmentos com amplificação semelhante nas combinações **nrev0a-Ri** e **nrev0b-Ri**, de acordo com a diferença de tamanho prevista entre os seus produtos de amplificação específicos (cerca de 23pb), e com amplificação inexistente nas reacções de controlo Ri-Ri. Os resultados obtidos encontram-se na **Figura 25**.

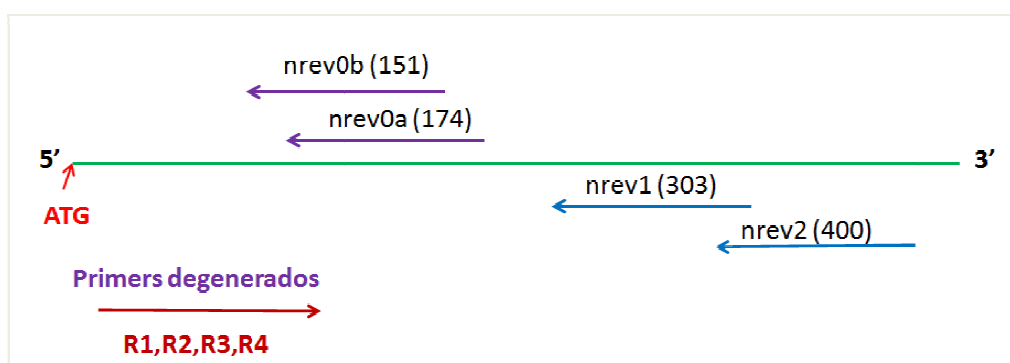


FIGURA 24 - REPRESENTAÇÃO ESQUEMÁTICA DA AMPLIFICAÇÃO POR HE-TAIL PCR DO GENE *LIP1*.

Os números entre parêntesis referem-se à posição da ORF em que cada primer específico foi desenhado. As setas representam o sentido de amplificação na cadeia. Neste caso concreto desconhecia-se a sequência das primeiras 52pb da ORF.

TABELA 3 – ETAPAS E CICLOS DE HE-TAIL-PCR USADOS NA ELUCIDAÇÃO DE *LIP1*.

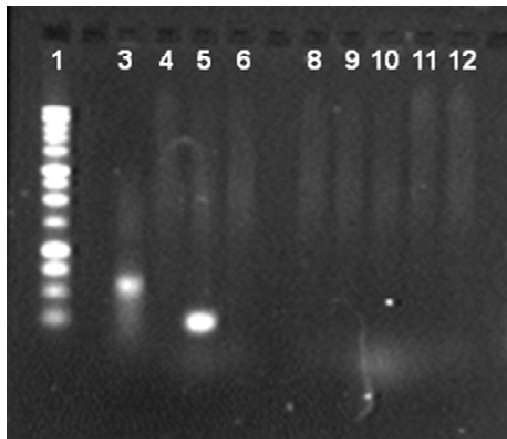
Reacção	Número de ciclos	Condições térmicas programadas
Primária	1	93 °C (1min); 95 °C (5min)
	5	94 °C (0,5min); 62 °C (1min) ; 72 °C (2,5min)
	1	94 °C (0,5min); 25 °C subindo gradualmente até 72 °C em 3min; 72 °C (2,5min)
	15	94 °C (20s) ; 65 °C (3,5min); 94 °C (20s) ; 65 °C (3,5min); 94 °C (0,5min); 42 °C (1min); 72 °C (2,5min)
	1	72 °C (5min); 4 °C. Pausa.
Secundária	12	94 °C (20s) ; 65 °C (3,5min); 94 °C (20s) ; 65 °C (3,5min); 94 °C (0,5min); 42 °C (1min); 72 °C (2,5min)
	1	72 °C (5min); 4 °C. Pausa.
Terciária	30	94 °C (0,5min); 44 °C (1min); 72 °C (2,5min)
	1	72 °C (5min); 4 °C Pausa.

A reacção de PCR primária foi executada em 15µl de volume total com 80ng de DNA genómico; 0,2 µM do primer **nrev2**; 2 µM de um dos primers degenerados (R1,R2,R3 ou R4); 0,2mM de cada dNTP. A reacção secundária foi executada com o primer **nrev1** (0,2 µM) e o mesmo primer degenerado Ri (2 µM) usado na reacção primária. Foi usado 1µl da diluição 1:50 do produto da PCR primária com DNA molde. A reacção terciária foi executada com 1 µl da diluição 1:10 do produto da PCR secundária; 0,2 µM dos primers **nrev0a** e **nrev0b**; 0,2 µM do primer degenerado Ri; 0,2 mM de cada dNTP. Em todas as reacções de PCR reacções foi usada 1U de Taq DNA polimerase (Promega) e tampão da enzima 10X. Para excluir amplificações não específicas, foram usados controlos R-R nas diversas reacções.

Na reacção terciária não foi detectada qualquer amplificação na reacção efectuada apenas com os primers degenerados **R1**. Houve amplificação nas combinações **R1+nrev0a** (carril 2) e **R1+nrev0b** (carril 3), como se pode visualizar na **Figure 25 C**. O facto de não ter sido obtida amplificação na reacção R1-R1 sugere que as bandas obtidas nas interacções de **R1** com os primers **nrev0a** e **nrev0b** foram específicas. A banda de 500bp do carril 2 foi purificada por GeneClean e sequenciada com o primer **nrev0a**, permitindo a elucidação do início da ORF de *lip1*, e de 416pb do promotor.

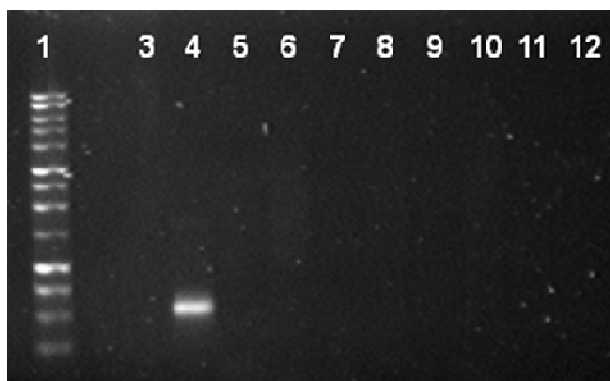
A sequência nucleotídica completa foi submetida às bases de dados da EMBL, onde pode ser consultada com o número de acesso [AM180877.1](https://www.ebi.ac.uk/EMBL/Sequence/Database/seqfetch.html?db=EMBL&acc=AM180877.1). *Lip1*, com uma ORF de 1677pb, codifica uma proteína deduzida de 558 aminoácidos, com 59,3KDa e um ponto isoeléctrico calculado de 4,56 (Jorge, et al., 2008). O número de acesso desta sequência de aminoácidos na base de dados da UniProt⁴¹ é [B0B099_TRIHA](https://www.uniprot.org/entry/B0B099). A sequência deduzida encontra-se representada nas **Figuras 26 e 27**.

⁴¹ <http://www.uniprot.org>



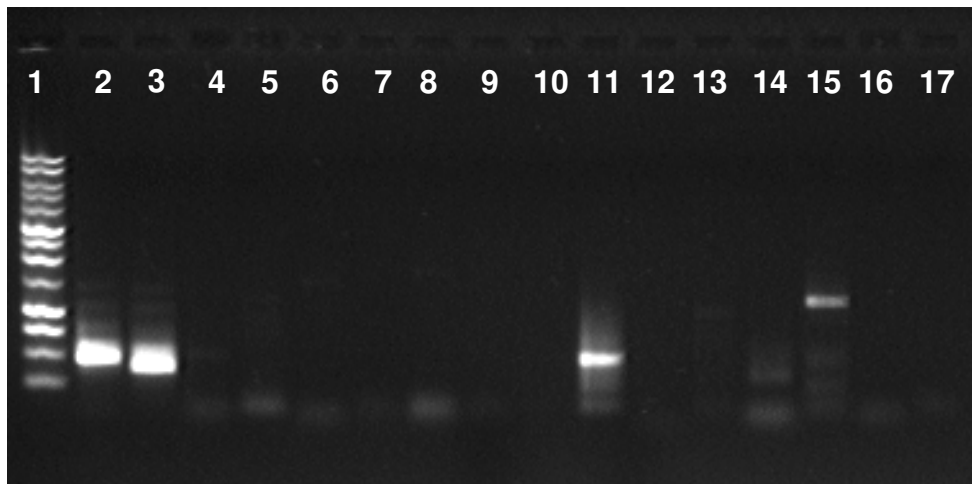
A – TAIL 1

Carril: 1 – Marcador de 1Kb; 3 – R1+nrev2; 4 – R2+nrev2; 5 – R3+nrev2; 6 – R4+nrev2; 8 – R1+R1, 9 – R2+R2; 10 – R3+R3, 11 – R4+R4; 12 – nrev2+nrev2; 13 – controlo negativo.



B – TAIL 2

Carril: 1 – Marcador de 1Kb; 3 – controlo negativo; 4 – R1+nrev1; 5 – R2+nrev1; 6 – R3+nrev1; 7 – R4+nrev1; 8 – R1+R1; 9 – R2+R2; 10 – R3+R3; 11 – R4+R4; 12 – nrev1+nrev1.



C –TAIL 3

Carril: 1 – Marcador de 1Kb; 2 – R1+nrev0a; 3 – R1+nrev0b; 4 – R2+nrev0a; 5 – R2+nrev0b; 6 – R3+nrev0a; 7 – R3+nrev0b; 8 – R4+nrev0a; 9 – R4+nrev0b; 10 – R1+R1; 11 – R2+R2; 12 – R3+R3; 13 – R4+R4; 14 – nrev0a+nrev0a; 15 – nrev0b+nrev0b; 16,17 – controlos negativos.

FIGURA 25- ELECTROFORESE DOS PRODUTOS DE REACÇÃO DA HE-TAIL PCR USADA PARA A ELUCIDAÇÃO DE *LIP1*.

```

>ORF lip1
1   ATG TTC AGC CTC ACA ACA TTA ACG GGA CTC CTT CTT GCC GTT GCC   45
1   Met Phe Ser Leu Thr Thr Leu Thr Gly Leu Leu Leu Ala Val Ala   15
46  TCC AAC GCA CAG CTG GCT GGT GCA GAT GGT CCA ACC GTG GCC TTG   90
16  Ser Asn Ala Gln Leu Ala Gly Ala Asp Gly Pro Thr Val Ala Leu   30
91  CGC TAT TCT ACA GTT GTT GGT TCT AGC AGT AAT GGA GTC GAT AGC  135
31  Arg Tyr Ser Thr Val Val Gly Ser Ser Ser Asn Gly Val Asp Ser   45
136 TTT AGA GGC ATA CCT TAC GCG CAG CCG CCA GTA GGA TCG CTA CGA  180
46  Phe Arg Gly Ile Pro Tyr Ala Gln Pro Pro Val Gly Ser Leu Arg   60
181 TTG AAG CCA CCA CAA CCC ATA ACT TCT TCT TTG GGC GAG GTT CAG  225
61  Leu Lys Pro Pro Gln Pro Ile Thr Ser Ser Leu Gly Glu Val Gln   75
226 GCA ACC GCT ACT CCG AAA GCA TGC CCT CAA TTT TCT TCG CAG TTT  270
76  Ala Thr Ala Thr Pro Lys Ala Cys Pro Gln Phe Ser Ser Gln Phe   90
271 ACT CCC AGT GgC CTA CCA CCA GCA TTA GTG GAT ATC ATC ACG AAT  315
91  Thr Pro Ser Gly Leu Pro Pro Ala Leu Val Asp Ile Ile Thr Asn  105
316 ATA TCG AAT GTA GTA CAA AAC CAA GAC GAG GAT TGC CTG ACA CTG  360
106 Ile Ser Asn Val Val Gln Asn Gln Asp Glu Asp Cys Leu Thr Leu  120
361 AAT GTC CAG AGG CCC TCG GAT GCT TCA AGC GGT TCC AAG CTC CCT  405
121 Asn Val Gln Arg Pro Ser Asp Ala Ser Ser Gly Ser Lys Leu Pro  135
406 GTA GTA TTC TTC ATT TAT GGA GGA GCT TTT GAA TCT GGC GCA ACA  450
136 Val Val Phe Phe Ile Tyr Gly Gly Ala Phe Glu Ser Gly Ala Thr  150
451 CAG GGA GTT GAC TCC ACA AAT CTC ATC CAA GCC TCA ATT GCT TCT  495
151 Gln Gly Val Asp Ser Thr Asn Leu Ile Gln Ala Ser Ile Ala Ser  165
496 GGT ACA CCG ATC ATT TTT GTG GCG GCA AAT TAT CGA TTA GGG GGC  540
166 Gly Thr Pro Ile Ile Phe Val Ala Ala Asn Tyr Arg Leu Gly Gly  180
541 TTC GGT TTC CTC GCT GGC AAA GAG CTC CTC AAC GAT GGA AGT ACC  585
181 Phe Gly Phe Leu Ala Gly Lys Glu Leu Leu Asn Asp Gly Ser Thr  195
586 AAC CTA GGC CTA AGG GAC CAG CGT TTG GCA CTG CAG TGG GTC GCT  630
196 Asn Leu Gly Leu Arg Asp Gln Arg Leu Ala Leu Gln Trp Val Ala  210
631 GAT AAT ATT GAG CAA TTC GGT GGT GAT CCC GAT AAA GTG ACG CTT  675
211 Asp Asn Ile Glu Gln Phe Gly Gly Asp Pro Asp Lys Val Thr Leu  225
676 TGG GGG TTC TCT GCA GGC TCT ATG AGC GTC TTT GAT CAG ACG GCG  720
226 Trp Gly Phe Ser Ala Gly Ser Met Ser Val Phe Asp Gln Thr Ala  240
721 CTC TTT GGA GGC AAT AAC TCT TAT CAC GGA AAG CCC CTT TTT AGG  765
241 Leu Phe Gly Gly Asn Asn Ser Tyr His Gly Lys Pro Leu Phe Arg  255
766 GCA GCA TTG ATG GAG AGC GGC AGC ATC CTG CCT GCG GAG CCT GCC  810
256 Ala Ala Leu Met Glu Ser Gly Ser Ile Leu Pro Ala Glu Pro Ala  270
811 AAT TCT ACA AAG GCT CAA CTC ATC TAT GAC AAG GTT GTT GAC AGC  855
271 Asn Ser Thr Lys Ala Gln Leu Ile Tyr Asp Lys Val Val Asp Ser  285

```

FIGURA 26 - SEQUÊNCIA DEDUZIDA DA PROTEÍNA LIP1 (PARTE INICIAL).

Aparecem a cor os sítios mais característicos das lipases: entre os resíduos **134-144**, a cavidade oxianiónica; de **216-231**, o padrão das carboxilesterases com serina no centro activo (serina activa na posição 229); resíduos de ácido aspártico e de histidina da tríade catalítica nas posições **360** e **474**, respectivamente. De **1-23**, péptido sinal.

856	GCC	GGA	TGC	TCG	ACA	TCC	AGC	GAT	ACT	CTA	GCA	TGT	TTG	AGG	TCC	900
286	Ala	Gly	Cys	Ser	Thr	Ser	Ser	Asp	Thr	Leu	Ala	Cys	Leu	Arg	Ser	300
901	GTA	GAC	TTT	AAC	ACC	TTT	CTG	TTA	GCG	GCA	GAA	TCT	GTT	CCC	ATC	945
301	Val	Asp	Phe	Asn	Thr	Phe	Leu	Leu	Ala	Ala	Glu	Ser	Val	Pro	Ile	315
946	TCA	AGC	AGT	TAT	AAC	TCC	ATC	GCT	TTG	TCA	TAC	CTA	CCA	AGG	CCA	990
316	Ser	Ser	Ser	Tyr	Asn	Ser	Ile	Ala	Leu	Ser	Tyr	Leu	Pro	Arg	Pro	330
991	GAT	GGA	ACA	GTC	CTC	CTT	GAT	TCA	CCG	GAA	ATC	TTG	GCT	TCA	AGA	1035
331	Asp	Gly	Thr	Val	Leu	Leu	Asp	Ser	Pro	Glu	Ile	Leu	Ala	Ser	Arg	345
1036	GGT	CAA	TTT	GCA	AAA	GTC	CCC	TTG	TTG	CTA	GGC	GAT	CAG	GAA	GAT	1080
346	Gly	Gln	Phe	Ala	Lys	Val	Pro	Leu	Leu	Leu	Gly	Asp	Gln	Glu	Asp	360
1081	GAA	GGA	ACT	CTT	TTC	TCG	GTT	TAC	CAA	TTG	AAT	CTT	ACT	AGC	ACA	1125
361	Glu	Gly	Thr	Leu	Phe	Ser	Val	Tyr	Gln	Leu	Asn	Leu	Thr	Ser	Thr	375
1126	CAA	GAC	GTT	GAA	GAA	TAT	CTT	GGC	AGC	CTC	TAT	TTC	CAG	CAA	GCC	1170
376	Gln	Asp	Val	Glu	Glu	Tyr	Leu	Gly	Ser	Leu	Tyr	Phe	Gln	Gln	Ala	390
1171	ACT	GCG	TCC	CAG	GTC	CAA	GAT	CTG	GTG	GCC	ACC	TAT	CCA	GAC	GAT	1215
391	Thr	Ala	Ser	Gln	Val	Gln	Asp	Leu	Val	Ala	Thr	Tyr	Pro	Asp	Asp	405
1216	CCC	TCG	GCT	GGT	TCG	CCT	TTC	AGA	ACA	GGG	TCG	TTG	AAT	AGC	CTT	1260
406	Pro	Ser	Ala	Gly	Ser	Pro	Phe	Arg	Thr	Gly	Ser	Leu	Asn	Ser	Leu	420
1261	TAT	CCA	GAA	TAC	AAA	CGA	CTG	GCC	GCT	ATA	CTT	GGT	GAC	TTT	ATT	1305
421	Tyr	Pro	Glu	Tyr	Lys	Arg	Leu	Ala	Ala	Ile	Leu	Gly	Asp	Phe	Ile	435
1306	TTC	ACA	CTT	CAG	AGG	CGC	AGT	CTT	CTT	GAG	ACA	GTA	GAC	CAG	ATA	1350
436	Phe	Thr	Leu	Gln	Arg	Arg	Ser	Leu	Leu	Glu	Thr	Val	Asp	Gln	Ile	450
1351	GCC	CCA	TCA	GTA	CCG	ACG	TGG	TCT	TTT	CTT	TCG	TCT	TAC	TTA	TAC	1395
451	Ala	Pro	Ser	Val	Pro	Thr	Trp	Ser	Phe	Leu	Ser	Ser	Tyr	Leu	Tyr	465
1396	GGG	ACG	GCT	TTT	TTG	GGA	ACT	TTC	CAC	ACG	ACC	TCA	CAA	ATC	GCT	1440
466	Gly	Thr	Ala	Phe	Leu	Gly	Thr	Phe	His	Thr	Thr	Ser	Gln	Ile	Ala	480
1441	GCT	TTT	GGG	CTG	ATT	CCA	GGC	TAC	GCT	GCC	GCC	GTT	ACA	CAA	GAA	1485
481	Ala	Phe	Gly	Leu	Ile	Pro	Gly	Tyr	Ala	Ala	Ala	Val	Thr	Gln	Glu	495
1486	TTT	TAC	ATC	TCT	TTT	TTC	AAC	ACG	ATG	AAC	CCG	AAC	AGC	AAT	GGC	1530
496	Phe	Tyr	Ile	Ser	Phe	Phe	Asn	Thr	Met	Asn	Pro	Asn	Ser	Asn	Gly	510
1531	TCA	GCA	CTA	CTC	CCT	ACA	TGG	CCA	TGC	TGG	GGA	TCA	GGA	AAA	CAG	1575
511	Ser	Ala	Leu	Leu	Pro	Thr	Trp	Pro	Cys	Trp	Gly	Ser	Gly	Lys	Gln	525
1576	GTG	ATG	CAA	CTT	GGT	GCA	ACT	TCT	TCA	TCT	TTG	TTA	AAT	GAC	AGC	1620
526	Val	Met	Gln	Leu	Gly	Ala	Thr	Ser	Ser	Ser	Leu	Leu	Asn	Asp	Ser	540
1621	TTC	CGA	TCG	GCG	TCC	TTT	GAT	TTT	ATC	CGA	TCA	AAC	TCC	CAG	GCA	1665
541	Phe	Arg	Ser	Ala	Ser	Phe	Asp	Phe	Ile	Arg	Ser	Asn	Ser	Gln	Ala	555
1666	CTA	CAT	ATT	TAG												1677
556	Leu	His	Ile	End												

FIGURA 27 - SEQUÊNCIA DEDUZIDA DA PROTEÍNA LIP1 (PARTE FINAL).

Aparecem a cor os sítios mais característicos das lipases: entre os resíduos **134-144**, a cavidade oxianiónica; de **216-231**, o padrão das carboxilesterases com serina no centro activo (serina activa na posição 229); resíduos de ácido aspártico e de histidina da tríade catalítica nas posições **360** e **474**, respectivamente.

Uma pesquisa efectuada na base integrada InterproScan⁴² integra a proteína codificada por *lip1* nas carboxilesterases com serina no centro activo ("carboxylesterases type-B"⁴³ signatures, serine active site" – PROSITE PS00122) (Sigrist, et al., 2010), com um E-value de $2,9 \times 10^{-128}$. Nas lipases e nas esterases a tríade catalítica é formada por três aminoácidos: uma serina, um ácido glutâmico ou aspártico e uma histidina. A sequência em redor da serina do centro activo está muito conservada, e é usada como "signature pattern" (padrão). No caso das carboxilesterases tipo-B é a sequência **F-[GR]-G-x(4)-[LIVM]-x-[LIV]-x-G-x-S-[STAG]-G**⁴⁴, com serina no centro activo. Considera-se como "signature pattern" secundária a região conservada situada na secção N-terminal, a qual contém uma cisteína implicada numa ligação por pontes dissulfeto (S=S), a sequência **[EDA]-[DG]-C-L-[YTF]-[LIVT]-[DNS]-[LIV]-[LIVFYW]-x-[PQR]**. Em *lip1* estes padrões estão presentes na sequência **FGGDPDKVTLWGFSAAG**, correspondente aos aminoácidos **216** a **231**, e na sequência **EDCLTLNVQRP**, entre as posições **115** e **125** (**Figuras 26 e 27**).

O centro activo da proteína localiza-se entre os resíduos **227** e **231**. A sua estrutura primária coincide com o consenso **G-x-S-x-G**, descrito como sendo o do centro activo de lipases, inserido numa sequência consenso mais ampla, a das carboxilesterases tipo B com serina no centro activo. Quanto aos outros dois resíduos que constituem a tríade catalítica, na posição **360** está localizado o resíduo ácido, que neste caso é de **ácido aspártico**, e na posição **474** o resíduo de **histidina** (**Figuras 26 e 27**).

A HE-TAIL PCR foi também usada para obter a sequência completa do gene *lip2*, a partir da sequência da EST-1279. Neste caso foram usados os oligonucleótidos específicos **lip2a** (5'- CTG GCA GAA CCG ATT CCC GAG CGC TT -3'), **lip2b** (5'- ACG CAA CTA CGA TGG CGC CTT GCT CG -3'), **lip2c** (5'- TGC GAT GAA CCC ACA GCT ATC GCC GA -3') e **lip2d** (5'- GAG AAA GCC TGT ACT CCA CCC TCG TG -3'), de 26bp e T_m de 70-72 °C, desenhados na ORF incompleta de *lip2* nas posições referidas na **Figura 28**, bem como os oligonucleótidos degenerados **R1**, **R2**, **R3** e **R4** descritos anteriormente. As condições químicas e físicas da PCR estão descritas na **Tabela 4**, e foram análogas às usadas na detecção da sequência de *lip1*, com a única excepção de se ter feito o ajustamento de alguns dos valores de temperatura (a negrito na Tabela).

⁴² <http://www.ebi.ac.uk/Tools/pfa/iprscan/>

⁴³ As carboxilesterases podem ser do tipo A, B ou C, de acordo com a sensibilidade aos fosfatos orgânicos.

⁴⁴ Os hífen separa resíduos; x pode ser qualquer resíduo; quando situados entre parêntesis rectos: naquela posição pode ser encontrado um dos diferentes resíduos situados no interior de parêntesis rectos.

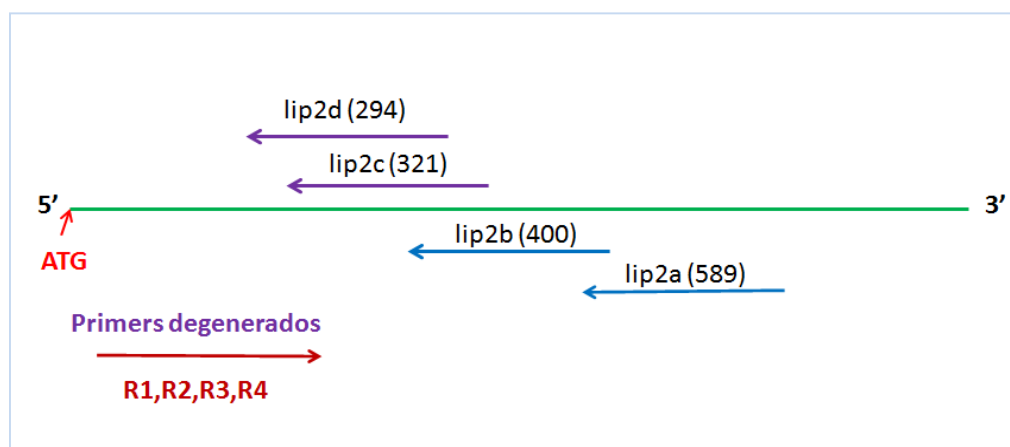


FIGURA 28 - REPRESENTAÇÃO ESQUEMÁTICA DA AMPLIFICAÇÃO POR HE-TAIL PCR DO GENE *LIP2*.

Os números entre parêntesis referem-se à posição da ORF em que cada primer específico foi desenhado. As setas representam o sentido de amplificação na cadeia. Neste caso concreto desconhecia-se a sequência de cerca de 300pb do início da ORF.

TABELA 4 – ETAPAS E CICLOS DE HE-TAIL-PCR USADOS NA ELUCIDAÇÃO DE *LIP2*.

Reacção	Número de ciclos	Condições térmicas programadas
Primária	1	93° C (1min); 95° C (5min)
	5	94 °C (0,5min); 62 °C (1min) ; 72 °C (2,5min)
	1	94 °C (0,5min); 25 °C subindo gradualmente até 72 °C em 3min; 72 °C (2,5min)
	15	94 °C (20s) ; 66 °C (3,5min); 94 °C (20s) ; 66 °C (3,5min); 94 °C (0,5min); 42 °C (1min); 72 °C (2,5min)
	1	72 °C (5min); 4 °C. Pausa.
Secundária	12	94 °C (20s) ; 65 °C (3,5min); 94 °C (20s) ; 65 °C (3,5min); 94 °C (0,5min); 42 °C (1min); 72 °C (2,5min)
	1	72 °C (5min); 4 °C. Pausa.
Terciária	30	94 °C (0,5min); 42 °C (1min); 72 °C (2,5min)
	1	72 °C (5min); 4 °C Pausa.

A reacção de PCR primária foi executada em 15µl de volume total com 80ng de DNA genómico; 0,2 µM do primer **lip2a**; 2 µM de um dos primers degenerados (R1,R2,R3 ou R4); 0,2mM de cada dNTP. A reacção secundária foi executada com o primer **lip2b** (0,2 µM) e o mesmo primer degenerado Ri (2 µM) usado na reacção primária. Foi usado 1µl da diluição 1:50 do produto da PCR primária com DNA molde. A reacção terciária foi executada com 1 µl da diluição 1:10 do produto da PCR secundária; 0,2 µM dos primers **lip2c** e **lip2b**; 0,2 µM do primer degenerado Ri; 0,2 mM de cada dNTP. Em todas as reacções de PCR reacções foi usada 1U de Taq DNA polimerase (Promega) e tampão da enzima 10X. Para excluir amplificações não específicas, foram usados controlos R-R nas diversas reacções.

Na **Figura 29** encontra-se as amplificações obtidas nas TAIL1, 2 e 3, respectivamente, para as diferentes combinações entre primers. Na TAIL1 as bandas mais intensas corresponderam às interações de **R1**, **R3** e **R4** com **lip2a** (carris 4, 6 e 7), e à interacção **lip2a+lip2a** (carril 13). Na TAIL2 a banda mais intensa correspondeu à combinação **R4+lip2b** (carril 6), mas também as interações de **R1**, **R2** e **R3** com **lip2b** originaram amplificação (carris 3, 4 e 5), assim como a combinação **lip2b+lip2b** (carril 11).

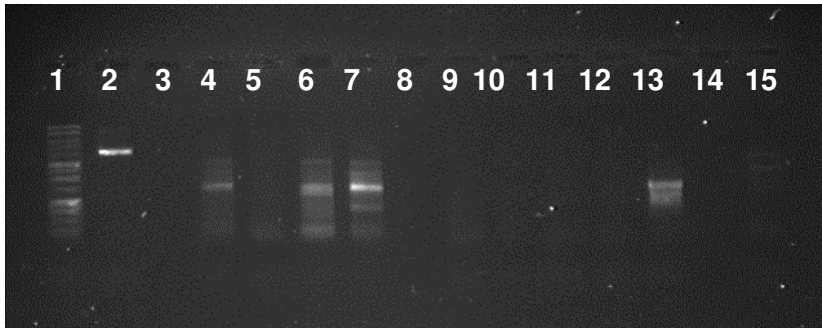
Na TAIL3 só as interações de **lip2c** e **lip2d** com o primer degenerado **R4** originaram bandas que se consideraram poder corresponder ao fragmento do gene que se pretendia isolar (carris 8 e 9), tendo sido seleccionada a banda correspondente a 2000bp que, embora menos intensa, parecia ser a que oferecia mais garantias de não ser devida a amplificação não específica. Essa banda foi purificada por GeneClean e sequenciada com o primer **lip2e** (5'- GCA CAT CGT TC GAC GCT GCT GGT CC - 3'), desenhado numa posição anterior a **lip2c** e **lip2d**, para conseguir simultaneamente efectuar o alargamento da cadeia previamente conhecida e a correcção de eventuais falhas de sequenciação na parte final da sequência pré-existente. A sequenciação permitiu o conhecimento de cerca de 400bp da sequência total do gene. Neste novo troço foi desenhado um novo primer, o **1279promo** (5'-CAA ATT CGT TCG CTG GCG ATG A-3'), para permitir a sequenciação de mais bases na zona do promotor.

No final, a sequência de *lip2* tinha um total de 1992pb, que abarcavam 548pb do promotor, 1215pb do quadro de leitura aberta e um terminador com 227pb. A ORF traduzida apresentava 58,2% de identidade (para um E-value de $4,0e^{-93}$) com a proteína A8W609-GIBZE, uma triacilglicerol lipase de *Giberella zeae* (*Fusarium graminearum*), de acordo com o algoritmo de FASTX, e 61% de identidade e $1e^{-132}$ de homologia para a mesma proteína, de acordo com o algoritmo de BLASTX⁴⁵.

A sequência completa do gene foi submetida à base de dados da EMBL, e pode ser acedida pelo número de acesso [AM774154.1](#). *Lip2* codifica uma proteína deduzida de 404 aminoácidos, ponto isoeléctrico calculado de 5,0 e 44,6KDa (Jorge, et al., 2010), cuja sequência está representada nas **Figuras 30 e 31**. O seu número de acesso na base de dados da UniProt é [B7ZET5_TRIHA](#). Uma pesquisa efectuada na base integrada InterproScan⁴⁶

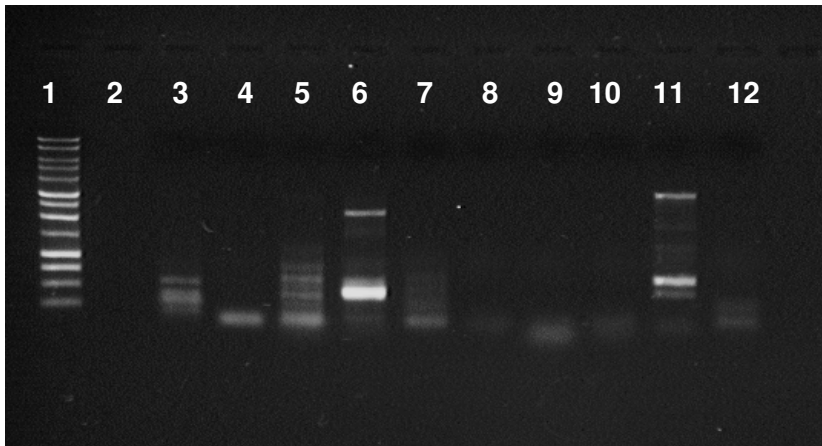
⁴⁵ Actualmente o melhor score pertence à proteína GORMI3-HYPJQ, uma triacilglicerol lipase de *Hypocrea jecorina* (anamorfo: *Trichoderma reesei*), com 80,5% de identidade (E-value de $1,4e^{-120}$) de acordo com o algoritmo de FASTX, e 80% de identidade a mesma proteína, de acordo com o de acordo com o algoritmo de BLASTX.

⁴⁶ <http://www.ebi.ac.uk/Tools/pfa/iprscan/>



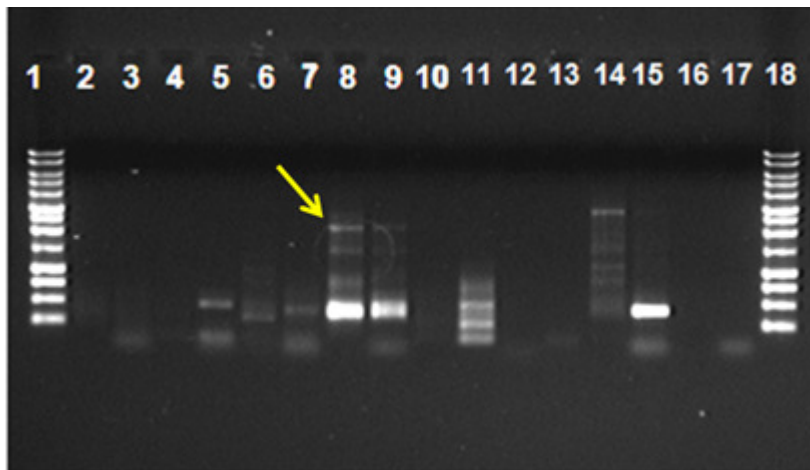
A – TAIL 1

Carril: 1 – Marcador de 1Kb; 2 – Marcador de intensidade; 4 – R1+lip2a; 5 – R2+lip2a; 6 – R3+ lip2a; 7 – R4+ lip2a; 9 – R1+R1, 10 – R2+R2; 11 – R3+R3, 12 – R4+R4; 13 – lip2a + lip2a; 15 – controlo negativo.



B – TAIL 2

Carril: 1 – Marcador de 1Kb; 3 – R1+lip2b; 4 – R2+lip2b; 5 – R3+ lip2b; 6 – R4+ lip2b; 7 – R1+R1, 8 – R2+R2; 9 – R3+R3, 10 – R4+R4; 11 – lip2b + lip2b; 12 – controlo negativo.



C – TAIL 3

Carril: 1 – Marcador de 1Kb; 2 – R1+lip2c; 3 – R1+lip2d; 4 – R2+lip2c; 5 – R2+lip2d; 6 – R3+lip2c; 7 – R3+lip2d; 8 – R4+lip2c; 9 – R4+lip2d; 10 – R1+R1, 11 – R2+R2; 12 – R3+R3, 13 – R4+R4; 14 – lip2c + lip2c; 15 – lip2d + lip2d; 16, 17 – controlos negativos.

FIGURA 29- ELECTROFORESE DOS PRODUTOS DE REACÇÃO DA HE-TAIL PCR USADA PARA A ELUCIDAÇÃO DE *LIP2*.

```

>ORF lip2
1   ATG GGG TGG ACA TTT CTG GGC CGC TTC GGG CTG TTG CTG CTT GTA   45
1   Met Gly Trp Thr Phe Leu Gly Arg Phe Gly Leu Leu Leu Leu Val   15
46  CAG CTC GTC GCT GCT GTC TTT GTC GAG GCC GCT TAC ATC AAT CAA   90
16  Gln Leu Val Ala Ala Val Phe Val Glu Ala Ala Tyr Ile Asn Gln   30
91  CAA GAA CCG TTG GGC GTC AAA GAG CCA GCG TCT GTT TCG AGC ATC   135
31  Gln Glu Pro Leu Gly Val Lys Glu Pro Ala Ser Val Ser Ser Ile   45
136 CCC GTG GCC CTG TTC GCG GAT CTC GAA CGA CTG GCG AGG CTC GTT   180
46  Pro Val Ala Leu Phe Ala Asp Leu Glu Arg Leu Ala Arg Leu Val   60
181 GAC GTC TCA TAT TGC CTT GGC ACC ACA GGC ATT CGC AAG CCG TTC   225
61  Asp Val Ser Tyr Cys Leu Gly Thr Thr Gly Ile Arg Lys Pro Phe   75
226 CAA TGC GTC TCT CGA TGC GAT GAA TTC CCC AAT GTG ACG CTG GCC   270
76  Gln Cys Val Ser Arg Cys Asp Glu Phe Pro Asn Val Thr Leu Ala   90
271 TCT ACG TGG AGT ACA GGC TTT CTC TTC GGC GAT AGC TGT GGG TTC   315
91  Ser Thr Trp Ser Thr Gly Phe Leu Phe Gly Asp Ser Cys Gly Phe   105
316 ATC GCA GTT GAT CAC GGC TCG GAC CAG CAG CGT CGA AAC GAT GTG   360
106 Ile Ala Val Asp His Gly Ser Asp Gln Gln Arg Arg Asn Asp Val   120
361 CTT ATA GGA GAT GAC GAG CAA GGC GCC ATC GTA GTT GCG TTC CGT   405
121 Leu Ile Gly Asp Asp Glu Gln Gly Ala Ile Val Val Ala Phe Arg   135
406 GGG ACC TAC AGC ATC ACC AAT ACC ATT ATT GAT CTG AGC ACC ATG   450
136 Gly Thr Tyr Ser Ile Thr Asn Thr Ile Ile Asp Leu Ser Thr Met   150
451 CCG CAA AAA TAC GTG CCA TAC CCG TCG CCA GAT CAC GGA GGA GAA   495
151 Pro Gln Lys Tyr Val Pro Tyr Pro Ser Pro Asp His Gly Gly Glu   165
496 TCA CCA GAA AAG CCT AGC CAT GAG TGC ACA AAT TGC ACC GTA CAC   540
166 Ser Pro Glu Lys Pro Ser His Glu Cys Thr Asn Cys Thr Val His   180
541 AGT GGG TTT CTC GAG TCT TGG AAA AGC GCT CGG GAA TCG GTT CTG   585
181 Ser Gly Phe Leu Glu Ser Trp Lys Ser Ala Arg Glu Ser Val Leu   195
586 CCA GAA CTC AAG GCA CTG AGG GCC AAG TAT CCG TCT CAC CCC GTC   630
196 Pro Glu Leu Lys Ala Leu Arg Ala Lys Tyr Pro Ser His Pro Val   210
631 CAT CTC ATC GGC CAC AGC CTT GGA GGC GCC GTG GCT TGT CTT GCG   675
211 His Leu Ile Gly His Ser Leu Gly Gly Ala Val Ala Cys Leu Ala   225
676 GCA CTC GAA CTG AAA GTG TCT CTT GGC TGG GAT GAT GTG ACG GTT   720
226 Ala Leu Glu Leu Lys Val Ser Leu Gly Trp Asp Asp Val Thr Val   240
721 ACC ACG TTT GGA GAG CCT CGT GTC GGC AAC TCT GAG TTC GCC CAC   765
241 Thr Thr Phe Gly Glu Pro Arg Val Gly Asn Ser Glu Phe Ala His   255
766 TTC GTT GAC GAT GTT TTT TAC CTC GAC GGC ATC ATT GAC CCT GAA   810
256 Phe Val Asp Asp Val Phe Tyr Leu Asp Gly Ile Ile Asp Pro Glu   270
811 AAG AGG ACC TAT CGA CGA GTT ACC CaC GCC GAC GAC CCA GTA CCC   855
271 Lys Arg Thr Tyr Arg Arg Val Thr His Ala Asp Asp Pro Val Pro   285

```

FIGURA 30 - SEQUÊNCIA DEDUZIDA DA PROTEÍNA LIP2 (PARTE INICIAL).

Aparecem a cor os sítios mais característicos das lipases: entre os resíduos **130-144**, a cavidade oxianiónica; de **210-219**, o padrão das lipases com serina no centro activo (serina activa na posição 216); resíduos de ácido aspártico e de histidina da tríade catalítica nas posições **282** e **373**, respectivamente. De **1-25**, péptido sinal.

856	TTA	CTT	CCA	CCG	GGC	GAG	TTT	GGT	TAC	CAG	TCA	CAT	GGC	GGC	GAG	900
286	Leu	Leu	Pro	Pro	Gly	Glu	Phe	Gly	Tyr	Gln	Ser	His	Gly	Gly	Glu	300
901	ATA	TTC	ATC	TCC	AAG	TCT	GCC	CTC	TCG	CCA	TCT	GAG	ACG	GAT	GTC	945
301	Ile	Phe	Ile	Ser	Lys	Ser	Ala	Leu	Ser	Pro	Ser	Glu	Thr	Asp	Val	315
946	CAG	CTA	TGC	GTC	GGC	GAC	GCA	GAT	CCT	AAC	TGC	AGT	GCC	AGG	GAT	990
316	Gln	Leu	Cys	Val	Gly	Asp	Ala	Asp	Pro	Asn	Cys	Ser	Ala	Arg	Asp	330
991	GAC	AGC	TCA	ATG	GAA	GGT	CTG	CTC	AAT	CGC	CTT	CTC	CAT	TTC	TGG	1035
331	Asp	Ser	Ser	Met	Glu	Gly	Leu	Leu	Asn	Arg	Leu	Leu	His	Phe	Trp	345
1036	GGA	ACG	ACT	GCC	TCT	TTG	GAA	GAG	TAC	ACG	GAG	AGG	ATG	AGC	ATC	1080
346	Gly	Thr	Thr	Ala	Ser	Leu	Glu	Glu	Tyr	Thr	Glu	Arg	Met	Ser	Ile	360
1081	CCG	ACG	CGG	TTC	AAA	TTG	TGG	CAG	CTG	TTT	TTT	GCT	CAT	CGA	GAT	1125
361	Pro	Thr	Arg	Phe	Lys	Leu	Trp	Gln	Leu	Phe	Phe	Ala	His	Arg	Asp	375
1126	TAC	TTC	TGG	AGA	CTC	GGA	CTA	TGT	GTG	CCC	GGC	GGC	GAC	CCT	ACA	1170
376	Tyr	Phe	Trp	Arg	Leu	Gly	Leu	Cys	Val	Pro	Gly	Gly	Asp	Pro	Thr	390
1171	AAC	TGG	GGT	CGG	CCT	CCG	TAT	GCG	CCT	CAT	GAC	GAG	GAG	TTG	TAA	1215
391	Asn	Trp	Gly	Arg	Pro	Pro	Tyr	Ala	Pro	His	Asp	Glu	Glu	Leu	End	

FIGURA 31 - SEQUÊNCIA DEDUZIDA DA PROTEÍNA LIP2 (PARTE FINAL).

Aparecem a cor os sítios mais característicos das lipases: entre os resíduos **130-144**, a cavidade oxianiónica; de **210-219**, o padrão das lipases com serina no centro activo (serina activa na posição 216); resíduos de ácido aspártico e de histidina da tríade catalítica nas posições **282** e **373**, respectivamente.

incluiu a sequência de aminoácidos de lip2 na família das lipases com serina no centro activo (PROSITE PS00120, centro activo na posição 210-219) (Sigrist, et al., 2010), e na das lipases_classe3 (InterPro IPR002921, Pfam PF01746, intervalo 131-290) com uma probabilidade de $1,0.e^{-34}$.

Na estrutura primária é detectável o consenso **G-x-S-x-G**, descrito como centro activo de lipases. Este pentapéptido encontra-se inserido na sequência consenso das lipases com serina no centro activo, com a "signature pattern": **[LIV]-{KG}-[LIVFY]-[LIVMST]-G-[HYWV]-S-{YAG}-G-[GSTAC]**. Em lip2 está presente na sequência **VHLIGHSLGG**, correspondente aos aminoácidos **210** a **219**. Os outros dois resíduos que constituem a tríade catalítica encontram-se na posição **282** (resíduo ácido, neste caso é o ácido aspártico), e na posição **373** (histidina).

5.2 AVALIAÇÃO DA ACTIVIDADE DAS PROTEÍNAS CODIFICADAS POR *LIP1* E *LIP2*

Para confirmação do sucesso dos procedimentos anteriores, após a identificação das respectivas ORFs, procedeu-se à clonagem de *lip1* e *lip2* em plasmídeos de expressão tendo-se obtido clones de *lip1* expressos em *Pichia pastoris* GS115 e de *lip2* expressos em *E. coli* DH5 α .

Para a clonagem de *lip1* e *lip2* em plasmídeos de expressão adoptou-se a seguinte estratégia:

- Fez-se o mapa de restrição de cada uma das ORFs, com o objectivo de seleccionar enzimas com sequências de reconhecimento inexistentes nas ORFs, para haver a garantia de não interromper a sequência da ORF inadvertidamente.
- De entre as enzimas seleccionadas na fase anterior, escolheram-se aquelas cuja sequência de reconhecimento apresentava maiores semelhanças com a sequência nucleotídica de cada um dos genes na posição imediatamente anterior ao início da ORF ou na posição a seguir ao códon de stop, e que simultaneamente fizessem parte do MCS do plasmídeo a utilizar, de forma a permitir a ligação posteriormente.
- A partir da sequência de reconhecimento das enzimas escolhidas na fase anterior desenharam-se oligonucleótidos que funcionam como “linkers” (incluem além da sequência de reconhecimento da enzima a sequência do início ou do fim da ORF, consoante se trata do primer directo ou reverso, respectivamente).
- Com estes oligonucleótidos obteve-se a sequência total de cada ORF por PCR.
- Cada produto de PCR foi digerido com as respectivas endonucleases de restrição, assim como os vectores de expressão utilizados.
- Foi efectuada a ligação entre vectores e insertos.

As sequências de reconhecimento de *EcoRI* e de *XhoI* (a sublinhado), necessárias para a clonagem no vector pIB2 foram incluídas na sequência de *lip1* por PCR com o primer directo EcoLip (5'- GGC GAA TTC ATG TTC AGC CTC AC -3') e o primer reverso XhoLip (5'- TAC CTC GAG CTA AAT ATG TAG TGC CTG G-3'). A sequência resultante foi clonada “in frame”⁴⁷ no vector pIB2 digerido com as mesmas enzimas e purificado por “gene clean”, tendo-se obtido o vector de expressão constitutiva em *P. pastoris* pIB2-*lip1* (**Figura 32**). A sequência do inserto incluído em pIB2-*lip1* foi confirmada por sequenciação com os primers EcoLip e XhoLip.

⁴⁷ No sentido de leitura.

Transformaram-se células electrocompetentes de *P. pastoris* GS115 por electroporação com o plasmídeo recombinante pIB2-lip1 linearizado com *Kpn2I* (ao nível do gene *his4*) a 1500V e 5mS. A transformação de *P. pastoris* com o plasmídeo pIB2 funcionou como controlo positivo. Após extracção de DNA das colónias transformadas, confirmou-se a presença de *lip1*, e a sua correcta orientação por PCR, utilizando dois primers desenhados em pIB2 (F514 e R756), e os primers nrev2 e lj11, desenhados na sequência de *lip1*, de acordo com a Figura 34.

A estratégia de clonagem de *lip2* no vector pET28 e o processo de confirmação foram análogos: os locais de reconhecimento das enzimas *SacI* e *Sall*, necessários para clonar no vector foram incluídos em *lip2* por PCR usando o primer directo lip2-dir-*SacI* (5'- TGA GCT CAT CAT CGC CAG CGA -3') e o primer reverso lip2-rev-*Sall* (5'- CGT TGG TCG ACT TTA CAA CTC CTC G -3'). A transformação de *E. coli* DH5α foi feita por choque térmico. A presença e correcta orientação de *lip2* nos transformantes foram confirmadas por PCR.

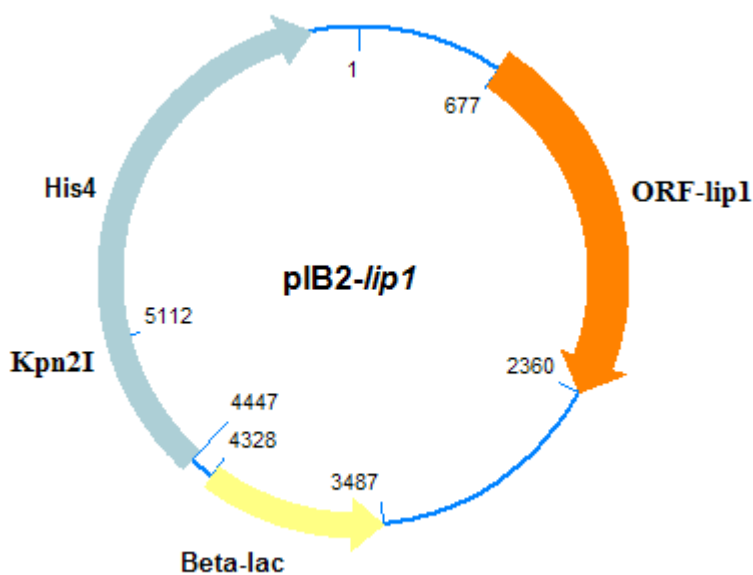


FIGURA 32 – CONSTRUÇÃO DO PLASMÍDEO RECOMBINANTE PIB2-LIP1.

Localização dos primers: **nrev2** (5'- GGA ACC GCT TGA AGC ATC CGA GGG CC -3'), desenhado na posição 371-395 da ORF, sentido reverso; **lj11** (5'- GCT TCA AGA GGT CAA TTT GC -3'), desenhado na posição 1027-1046 da ORF, sentido directo. **F514** e **R756**, desenhados em pIB2, no sentido directo e reverso, respectivamente nas posições 514 e 756.

A expressão da proteína codificada por *lip1* foi detectada por actividade em placas de Petri usando meios como a tributirina ou como a Rodamina B lipase agar (**Figura 33**), substratos utilizados na detecção de actividade esterásica e lipolítica, respectivamente. Em SDS-PAGE foi detectada uma proteína com cerca de 59kDa, inexistente na levedura não transformada (**Figura 34**). A avaliação da actividade de *lip2* encontra-se numa fase mais atrasada.

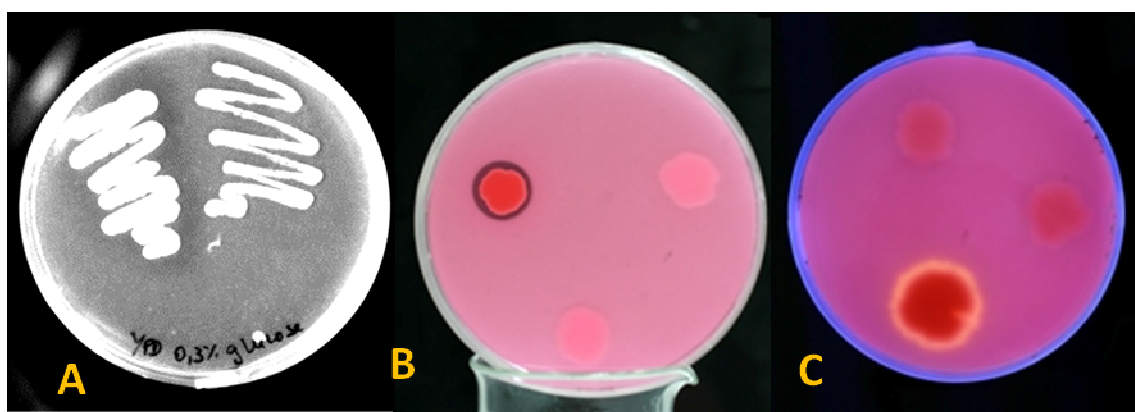


FIGURA 33 – HALOS DE ACTIVIDADE ESTERÁSICA (A E B) E FLUORESCÊNCIA DEVIDA A ACTIVIDADE LIPOLÍTICA (C).

A – meio YEPD modificado com 1% de tributirina; **B** – meio YEPD modificado com 1% de tributirina e rodamina; **C** – meio rodamina B lipase agar(Alken-Murray Corporation, 2001) modificado, com azeite como substrato lipídico. Os crescimentos verificados em placa que não apresentam halos ou fluorescência correspondem à levedura não transformada, usada como controlo negativo.

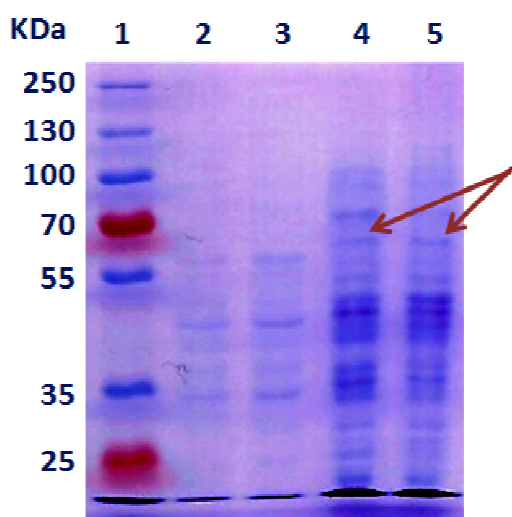


FIGURA 34 – SDS-PAGE OBTIDO COM EXTRACTOS PROTEICOS CELULARES DE LEVEDURA GS115 E DE GS115 TRANSFORMADA COM *LIP1* (CRESCIMENTO DE 48H EM YEPD)

Carril 1- Marcador de proteínas peq-Gold V; **Carris 2 e 3-** GS115; **Carris 4 e 5-** GS115-pIB2-lip1.

6. CONSIDERAÇÕES FINAIS

O rastreio de bibliotecas de DNA permite a identificação e caracterização de genes. As bibliotecas de genómico são utilizadas para analisar a sequência de um gene em toda a sua extensão: o promotor, os intrões, as fronteiras intrão-exão, exões e a região do terminador. No entanto, devido elevado número de clones que é necessário usar, são mais difíceis de manipular, e de manter. As de cDNA permitem a identificação directa da região codificante e a dedução das sequências de proteínas, sendo conseqüentemente de mais fácil manipulação. No entanto, muitas vezes a obtenção de moléculas de cDNA de cadeia completa (“full-length” cDNA) é difícil, pois alguns dos métodos de síntese de cDNA podem levar à obtenção de cópias incompletas (cf. § 3).

A hibridação de sondas de ácidos nucleicos é o método mais generalizado de rastreio de bibliotecas. Nas de cDNA permite a identificação de clones com sequências incompletas que, de outra forma não poderiam ser detectados devido à impossibilidade de se expressarem.

Embora no rastreio de bibliotecas de genómico tenha sido frequente a utilização de sondas de cDNA, a evolução das tecnologias de sequenciação levou ao recrudescimento da utilização de ESTs (“Expressed Sequence Tags”), que não são mais do que pequenas porções da molécula de cDNA, de obtenção mais fácil e barata. As ESTs revelaram-se uma ferramenta muito útil na detecção de genes, nomeadamente os responsáveis pela doença de Alzheimer e pelo cancro do cólon.

Um dos problemas que se pode colocar à elucidação das sequências nucleotídicas contidas nalgumas bibliotecas de DNA está relacionado com a utilização de vectores derivados do fago lambda, pois muitas vezes não é fácil a obtenção de DNA de fago com qualidade suficiente para sequenciação, tal como se verificou neste trabalho. A solução por nós encontrada foi recorrer a uma metodologia de rastreio por PCR, utilizando a HE-TAIL PCR (“High-Efficiency Thermal Asymmetric Interlaced PCR”), que se revelou eficaz na elucidação dos genes *lip1* e *lip2* de *T. harzianum*, permitindo determinar as sequências de DNA adjacentes às previamente conhecidas.

REFERÊNCIAS BIBLIOGRÁFICAS

- Abremski, K., & Hoess, R. (1984).** Bacteriophage P1 site-specific recombination. Purification and properties of the Cre recombinase protein. *J. Biol. Chem.*, 259(3), pp. 1509-1514.
- Agilent Technologies. (2010).** *pBlueScript II Phagemid System. Instruction Manual.* Obtido em 23 de Novembro de 2011, de <http://www.genomics.agilent.com>: <http://www.genomics.agilent.com/files/Manual/212205.pdf>
- Agilent Technologies. (2011).** Lambda ZAP II Undigested Vector Kit. Instructions Manual.
- Alberts, B., Bray, D., Johnson, A., Lewis, J., Raff, M., Roberts, K., & Walter, P. (2002).** *Essential Cell Biology- An Introduction to the Molecular Biology of the Cell.* New York & London: Garland Publishing, Inc. 630pp. ISBN 0-8153-2971-7.
- Alken-Murray Corporation. (2001).** Obtido de <http://www.alken-murray.com/>.
- Altschul, S., Madden, T., Schäffer, A., Zhang, J., Zhang, Z., Miller, W., & Lipman, D. (1997).** Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.*, 25:3389-3402. <http://blast.ncbi.nlm.nih.gov/Blast.cgi>.
- Anónimo. (2002).** *Genomic libraries.* Obtido em 7 de Dezembro de 2011, de <http://homepages.strath.ac.uk/~dfs99109/BB211/GenomicLibrary.html>
- Anónimo. (sd).** *Vector PPAC4 Information/Map .* Obtido de BACPAC Resources Center (BPRC): <http://bacpac.chori.org/ppac4.htm>
- BookRags. (2005).** *Bacterial Artificial Chromosome (bac).* Obtido em 18 de Novembro de 2011, de BookRags: <http://www.bookrags.com/research/bacterial-artificial-chromosome-bac-wmi/>
- Brem, G., Brenig, B., Muller, M., & Springmann, K. (1989).** *Ex situ cryoconservation of genomes and genes of endangered cattle breeds by means of modern biotechnological methods.* (Department of Molecular Animal Breeding (University of Munich)). Obtido em 14 de Novembro de 2011, de Food and Agriculture Organization of the United Nations: <http://www.fao.org/DOCREP/004/T0094E/T0094E00.HTM>
- Chasteen, L., Ayriss, J., Pavlik, P., & Bradbury, A. (2006).** *Nucleic Acids Research*, 34 (21), p. 11pp.
- Chen, B., Sun, C., Wang, Y., Hu, Y., & Lin, Z. (2004).** Anchored-PCR (A-PCR): a new method for chromosome walking. *Chinese Science Bulletin*, 49 (16), pp. 1772-1774.
- Clarke, L., & Carbon, J. (1976).** A colony bank containing synthetic ColE1 hybrid plasmids representative of the entire E. coli genome. *Cell*, 9, pp. 91-99.

Correia, I., Moreira, L., & Fialho, A. (2003). Engenharia Genética. In N. Lima, & M. Mota, *Biotecnologia. Fundamentos e aplicações*: 125-161. LIDEL - Edições Técnicas, Lda. ISBN-978-972-757-197-0.

Croy, R. (1998). *cDNA synthesis and cloning*. Obtido em 3 de Dezembro de 2011, de David W. Brooks Teaching and Research : <http://dwb4.unl.edu/Chem/CHEM869N/CHEM869NLinks/www.dur.ac.uk/~dbl0www/Staff/Croy/cDNAfigs.htm>

EcoArray. (2009). *Library Construction*. Obtido em 10 de Dezembro de 2011, de ecoarray.com: <http://ecoarray.com/Service/Library%20Constr.htm>

Frengen, E., Zhao, B., Howe, S., Weichenhan, D., Osoegawa, K., Gjernes, E., de Jong, P. (2000). Modular bacterial artificial chromosome vectors for transfer of large inserts into mammalian cells. *Genomics*, 68(2), pp. 118-126.

Grosveld, F., Lund, T., Murray, E., Mellor, A., Dahl, H., & Flavell, R. (1982). The construction of cosmid libraries which can be used to transform eukaryotic cells. *Nucleic Acid Research*, 10, pp. 6715-6732.

Hartl, D., & Jones, E. (1998). *Genetics. Principles and Analysis* (4th edition ed.). Sudbury, Massachusetts: Jones and Bartlett Publishers. 840pp. ISBN 0-7637-0489-X.

Hoess, R., & Abremski, K. (1984). Interaction of the bacteriophage P1 recombinase Cre with the recombining site loxP. *Proc. Natl. Acad. Sci.*, 81, pp. 1026-1029.

Houghton, J. (2011). *Lecture 3 - Recombinant DNA Technology- DNA Structure Function*. Obtido em 20 de Novembro de 2011, de Advanced Genetics. Bio4564 Summer'11: <http://biology200.gsu.edu/houghton/4564%20%2711/lecture3.html>

Human Genome Project. (2009). *Potential Benefits of Human Genome Project Research*. Obtido em 8 de Dezembro de 2011, de www.ornl.gov/hgmis: http://www.ornl.gov/sci/techresources/Human_Genome/project/benefits.shtml

Institute of Molecular Development. (2001). *Construction of cDNA Library into Lambda Zap Vector*. Obtido em Dezembro de 2011, de Molecular Techniques and Methods: <http://www.molecularinfo.com/MTM/A/A2/A2-1/A2-1-1.html>

Khan, A., Wilcox, A., Hopkins, J., & Sikela, J. (1991). Efficient double stranded sequencing of cDNA clones containing long poly(A) tails using anchored poly(dT) primers. *Nucleic Acid Research*, 19, p. 1715.

King, M. (2011). *Molecular Tools of Medicine*. Obtido em 19 de Novembro de 2011, de themedicalbiochemistrypage.org: <http://themedicalbiochemistrypage.org/molecular-medicine.html#genomic>

Kumar, A., & Garg, N. (2005). *Genetic Engineering*. New York: Nova Science Publishers. 207pp.

- Jorge, L., González, F., Monte, E. & Choupina, A. (2008).** HE-TAIL PCR, a powerful tool for identification of *Trichoderma harzianum lip1* gene. XVI National Congress of Biochemistry. University of the Azores, Ponta Delgada. 22-25 October 2008.
- Jorge, L., González, F., Monte, E. & Choupina, A. (2010).** Identification of *Trichoderma harzianum lip2* gene by HE-TAIL PCR. Actas do XVII National Congress of Biochemistry. Porto. 15-17 December 2010.
- Lau, Y., & Kan, Y. (1983).** Versatile cosmid vectors for the isolation, expression and rescue of gene sequences: Studies with human globin gene cluster. *Proc. Natl. Acad. Sci.*, 80, pp. 5225-5229.
- Lee, F. (sd).** *Molecular Biology Web book. Chapter 9: Biotechnology and Bioinformatics.* (W.-b. publisher, Editor) Obtido em 17 de Novembro de 2011, de Web-books publisher: <http://www.web-books.com/MoBio/Free/Ch9A4.htm>
- Liu, Y., & Whittier, R. (1995).** Thermal asymmetric interlaced PCR: automatable amplification and sequencing of insert end fragments from P1 and YAC clones for chromosome walking. 25, pp. 674-681.
- Lodish, H., Berk, A., Zipursky, S. L., Matsudaira, P., Baltimore, D., & Darnell, J. (2000).** *Molecular Cell Biology.* (4th edition ed.). New York: W. H. Freeman and Company. 1084pp. ISBN 0-7167-3706-X. Obtido em 15 de Novembro de 2011, de The National Center for Biotechnology Information: <http://www.ncbi.nlm.nih.gov/books/NBK21696/>
- Lucigen.com. (2011).** *CopyRight® v2.0 BAC Cloning Kits (pSMART® BAC and pEZ™ BAC).* Obtido em 23 de Novembro de 2011, de Lucigen.com: <http://lucigen.com/store/CopyRight-v2.0-BAC-Cloning-Kits-pSMART-BAC-and-pEZ-BAC>
- Malone, G., Zimmer, P., Meneghello, G., Binneck, E., & Peske, S. (2006).** Prospecção de genes em bibliotecas de cDNA. *R. Bras. Agrociência*, 12(1), pp. 7-13.
- Metzenberg, S. (2002a).** *Lecture 18- High capacity vectors.* Obtido em 24 de Novembro de 2011, de escience: <http://escience.ws/b572/L18/L18.htm>
- Metzenberg, S. (2002b).** *Lecture 17 - Virulent lambda vectors.* Obtido em 3 de Dezembro de 2011, de <http://www.escience.ws/b572/L17/L17.htm>
- Michiels, A., Tucker, M., van den Ende, W., & van Laere, A. (2003).** Chromosomal walking of flanking regions from short known sequences in GC-rich plant genomic DNA. *Plant Molecular Biology Reporter*, 21(3), pp. 295-302. DOI: 10.1007/BF02772805 .
- Microbial Genome Program . (2010).** *Microbial Genomics.* Obtido em 8 de Dezembro de 2011, de <http://microbialgenomics.energy.gov/>: <http://microbialgenomics.energy.gov/>
- NCBI. (2004).** *A science primer. ESTs: GENE DISCOVERY MADE EASIER.* Obtido em 16 de Dezembro de 2011, de <http://www.ncbi.nlm.nih.gov>: <http://www.ncbi.nlm.nih.gov/About/primer/est.html>

NCBI. (2007). *A Science Primer*. (National Center for Biotechnology Information). Obtido em 14 de Dezembro de 2011, de <http://www.ncbi.nlm.nih.gov/About/primer/microarrays.html>

Nicholl, D. (2008). *An Introduction to Genetic Engineering* (Third Edition ed.). Hong Kong: Cambridge University Press. ISBN 978-0-521-850063.

openwetware.org. (2007). *Bacterial artificial chromosomes*. Obtido em 22 de Novembro de 2011, de [openwetware.org](http://openwetware.org/wiki/Bacterial_artificial_chromosomes):
http://openwetware.org/wiki/Bacterial_artificial_chromosomes

Otsuka, M., Ichinose, K., Fujii, I., & Ebizuka, Y. (2004). Cloning, sequencing, and functional analysis of an iterative type I polyketide synthase gene cluster for biosynthesis of the antitumor chlorinated polyenone neocarzilin in *Streptomyces carzinostaticus*. *Antimicrobial Agents and Chemotherapy*, 48(9), pp. 3468-3476.

Pareja, E. (2005). *Universidad de Granada*. Obtido em 16 de Novembro de 2011, de http://www.ugr.es/~eianez/Biotecnologia/IngGenet.htm#_Toc30851506

Pearson, W., Wood, T., Zhang, Z. & Miller, W. (1997). Comparison of DNA sequences with protein sequences. *Genomics*, 46(1), pp.24-36.

Preston, G. (2003). Cloning gene family members using PCR with degenerate oligonucleotide primers. In D. S. John M. S. Bartlett (Ed.), *Methods in Molecular Biology* (2nd ed., Vol. 226, pp. 485-498). Totowa, New Jersey: Humana Press Inc.

Primrose, S., Twyman, R., & Old, R. (2001). *Principles of gene manipulation* (6th edition ed.). Wiley-Blackwell. 390pp. ISBN 0-632-059540.

Primrose, S., & Twyman, R. (2006). *Principles of Gene Manipulation and Genomics* (7th Edition ed.). Blackwell Publishing. 644pp. ISBN-1.4051-3544-1.

Promega. (2011). *Discontinued vectors*. Obtido em 7 de Dezembro de 2011, de <http://www.promega.com/es-es/products/pm/discontinued-vectors/>

Rinehart, C. (2005a). *Introduction to Recombinant Genetics- Biology 350*. Obtido em 4 de Dezembro de 2011, de <http://www.wku.edu/biology/>:
<http://bioweb.wku.edu/courses/biol350/Transformation10/Review.html>

Rinehart, C. (2005b). *Introduction to Recombinant Genetics- Biology 350*. (W. K. University, Ed.). Obtido em 4 de Dezembro de 2011, de <http://www.wku.edu/biology/>:
<http://bioweb.wku.edu/courses/biol350/RestrictionEnz3/Review.html>

ROCHE. (2005). <http://www.roche-applied-science.com>. Obtido em 12 de Janeiro de 2012, de http://www.roche-applied-science.com/proddata/gpip/3_6_13_6_1_1.html

ROCHE. (2010). <https://www.roche-applied-science.com>. Obtido em 12 de Janeiro de 2012, de <https://www.roche-applied-science.com>

science.com/PROD_INF/index.jsp?id=iforu&iforu_page=search&catalogNumber=11585649910

ROCHE. (2011). <http://www.roche-applied-science.com>. Obtido em 12 de Janeiro de 2012, de https://www.roche-applied-science.com/PROD_INF/index.jsp?id=iforu&iforu_page=search&catalogNumber=11636090910

Schirmbock, M., Lorito, M., Wang, Y., Hayes, C., Arisan-atac, I., Scala, F., Kubicek, C. (1994). Paralell formation and synergismof hydrolytic enzymes and peptaboil antibiotics, molecular mechanisms involved in the antagonistic action of *Trichoderma harzianum* against phytopathogenic fungi. *Appl. Environ.Microbiol.*, 60, pp. 4364-4370.

Schnos, M. (2005). <http://www.biochem.wisc.edu/faculty/inman/empics/virus.htm>. Obtido de University of Wisconsin-Madison. Biochemistry Department: <http://www.biochem.wisc.edu/faculty/inman/empics/0022a.jpg>

Seed, B., Parker, R., & Davidson, N. (1982). Representation of DNA sequences in recombinant DNA libraries prepared by restriction enzyme partial digestion. *Gene*, 19, pp. 201-209.

Short, J., Fernández, J., Sorge, F., & Huse, W. (1988). λ ZAP: a bacteriophage λ expression vector with "in vivo" excision properties. *Nucleic Acid Research*, 16(15), pp. 7583-7600.

Sigrist, C., Cerrutti, L., de Castro, E., Langendijk-Genevaux, P., Bulliard, V., Bairoch, A., & Hulo, N. (2010). PROSITE, a protein domain database for functional characterization and annotation. *Nucleic Acids Res.*, 38, pp. 161-166.

Strachan, T., & Read, A. (1999). *Human Molecular Genetics* (2nd edition ed.). New York: Bios Scientific Publishers, an imprint of Taylor & Francis Group. ISBN-10: 1-85996-202-5. Disponível em <http://www.ncbi.nlm.nih.gov/books/NBK7580/>.

Videira, A. (2001). *Engenharia Genética. Princípios e aplicações*. Lidel - edições técnicas, lda. 168pp. ISBN 978-972-757-163-8.

Vizcaíno, J., Cardoza, R., Dubost, L., Bodo, B., Gutiérrez, S., & Monte, E. (2006a). Detection of peptaboils and partial cloning of a peptaboil synthetase gene from *T. harzianum* CECT 2413. *Folia Microbiol.*, 51(2), pp. 114-120.

Vizcaíno, J., González, F., Suárez, M., Redondo, J., Heinrich, J., Delgado-Jarana, J., Rey, M. (2006b). Generation, annotation and analysis of ESTs from *Trichoderma harzianum* CECT 2413. *BMC Genomics*, 7:193. Disponível em <http://www.biomedcentral.com/1471-2164/7/193>.

Wilson, E. (2003). *Bacteriophages*. Obtido em 14 de Novembro de 2011, de Davidson College Molecular Biology Home Page:

<http://www.bio.davidson.edu/Courses/Molbio/MolStudents/spring2003/WilsonE/home.cosmids.html>

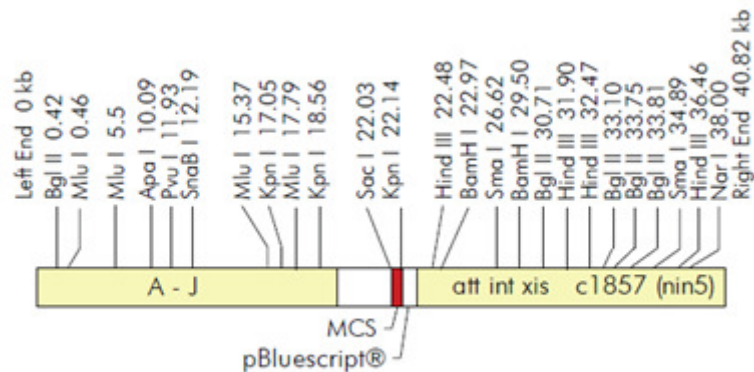
Wolfe, J. (2005). *Genome Project - lecture one*. Obtido em 16 de Novembro de 2011, de London's Global University: http://www.ucl.ac.uk/~ucbhjow/b200/lecture_1.html

ANEXOS

ANEXO 1 – ALGUNS DOS VECTORES USADOS EM BIBLIOTECAS DE CDNA

LAMBDA ZAP

O vector Lambda ZAP II tem seis locais de restrição únicos (*Sac* I, *Not* I, *Xba* I, *Spe* I, *Eco*R I e *Xho* I). Comporta insertos até DNA 10 kpb. Os clones inseridos no Lambda ZAP II podem ser rastreados com sondas de DNA ou com anticorpos. A rápida excisão “in vivo” do fagemídeo pBluescript permite a caracterização dos insertos em vectores plasmídicos.



(Agilent Technologies, 2011)

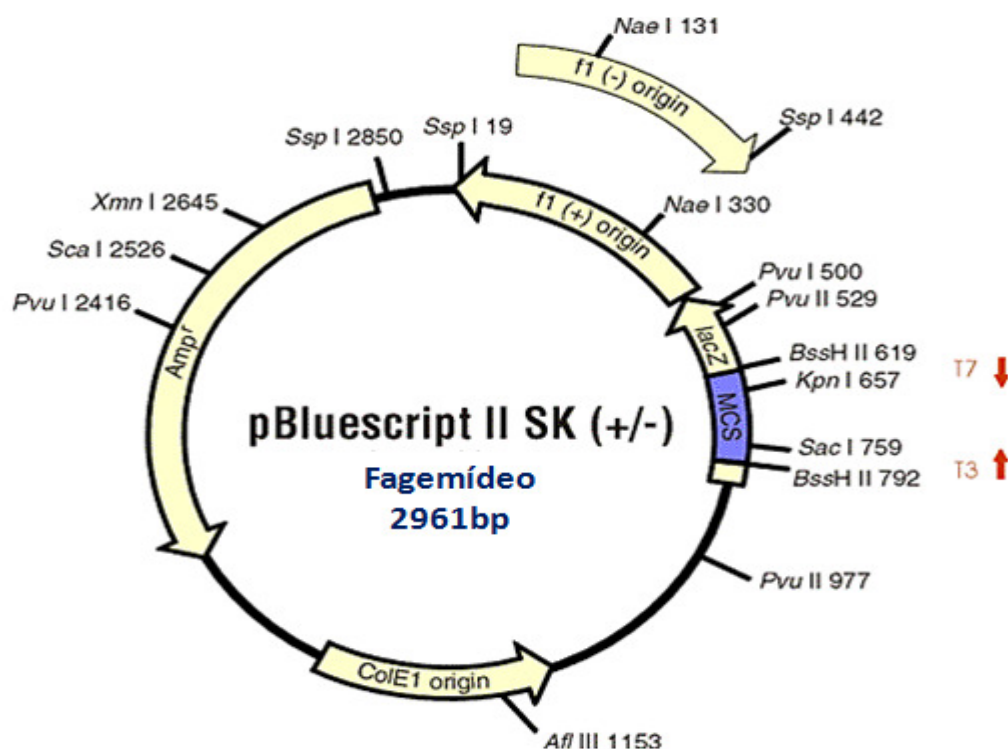
FIGURA 35 – MAPA DE RESTRIÇÃO DO VECTOR DE INSERÇÃO LAMBDA ZAP II

FAGEMÍDEOS

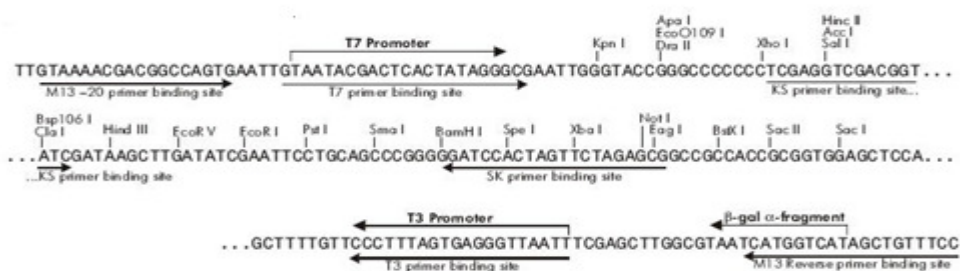
Os fagemídeos são vectores híbridos entre um plasmídeo e o fago M13, um bacteriófago de *E. coli* com DNA de cadeia simples. A origem de replicação do fago é inserida num plasmídeo (como o pUC19, que apresenta os genes *lacZ* e *amp^r*). Os fagemídeos da série pBluescript são dos mais utilizados como vectores de clonagem em bibliotecas de cDNA. Estes vectores são derivados do pUC19, e têm a origem de replicação f1 do fago filamentososo f1, relacionado com M13. Podem ser usados para sequenciação, transcrição *in vitro* e mutagénese *in vitro*. A título exemplificativo, na **Figura 35** estão representados o mapa e o local de clonagem múltipla do vector pBluescript II SK (+/-), um fagemídeo com 2961pb.

Como no fagemídeo não foi inserida a informação relativa à síntese das proteínas da cápside viral, quando se querem obter partículas virais tem que se co-infectar as células hospedeiras com um fago auxiliar⁴⁸ (“phage helper”). O fago auxiliar M13KO7, derivado do fago M13, é capaz de se replicar na ausência de DNA fagemídico. Em presença de fagemídeos com origem de replicação f1, verifica-se que o DNA fagemídico é encapsidado preferencialmente ao de M13KO7.

⁴⁸ Os fagos auxiliares são indispensáveis para a replicação, empacotamento e montagem das partículas fagemídicas. Também podem ser usados extractos celulares bacterianos que produzam componentes idênticos (Chasteen, Ayriss, Pavlik, & Bradbury, 2006).



Local de clonagem múltipla de pBlueScript SK (+/-)
(bases representadas, de 601-826)



(Agilent Technologies, 2010)

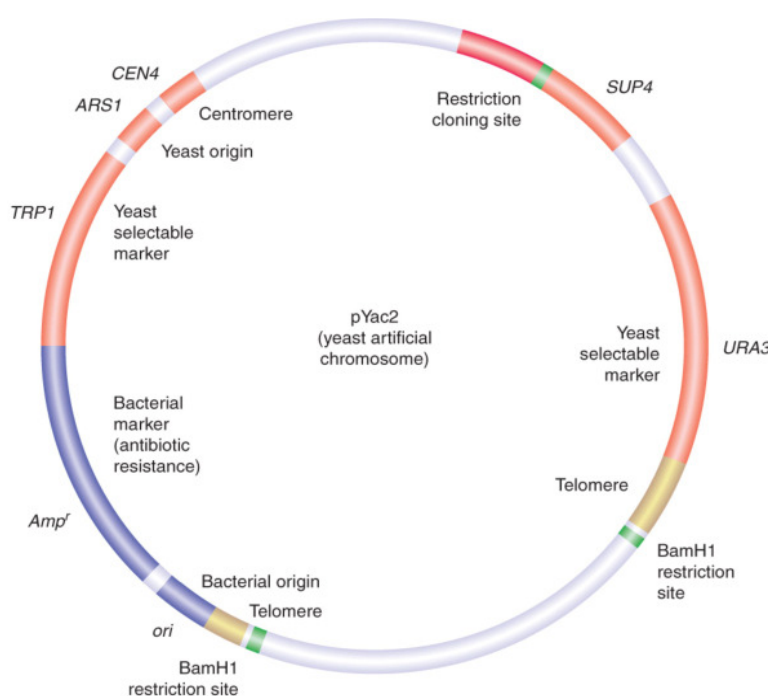
FIGURA 36 – MAPA E CARACTERÍSTICAS DE PBLUESCRIPT SK⁺.

É um fagemídeo (plasmídeo com origem de replicação de fago – f1) autorreplicativo derivado de pUC19, desenhado para facilitar a clonagem e a sequenciação de DNA, a obtenção de transcritos de RNA *in vitro*, a mutagénese e o mapeamento de genes. Tem origem de replicação relaxada de *E. coli*, um fragmento do operão *lac* de *E. coli*, que codifica o péptido α da β -galactosidase (*lac Z*). O “polylinker” tem 21 sequências de reconhecimento únicas para enzimas de restrição. A flaquear o “polylinker” encontram-se as sequências promotoras da T3 e da T7 RNA polimerases, que possibilitam a síntese de RNA *in vitro*. A localização do “polylinker” na região N-terminal do gene *lac Z* leva a que os transformantes possam ser detectados em placa por localização de colónias brancas. A flaquear os promotores da T3 e da T7 RNA polimerases há locais de restrição *BssHI*, raros, que permitem a excisão do inserto e dos promotores do fago T e o uso em mapeamento de genes. A região intergénica F1 (relacionada com o fago M13), inclui 307pb da origem de replicação f1 e permite a recuperação de DNA de cadeia simples (ssDNA) pelo fago *helper* (M13K07), podendo ser usado em sequenciação ou mutagénese. A região intergénica F1 ter duas orientações: (+), no mesmo sentido de *lac Z*, ou (-), no sentido inverso.

ANEXO 2 – ALGUNS DOS VECTORES USADOS EM BIBLIOTECAS DE GENÓMICO

CROMOSSOMAS ARTIFICIAIS DE LEVEDURAS (YACs)

Os cromossomas artificiais de leveduras, YACs (“yeast artificial chromosomes”), desenvolvidos para a clonagem em leveduras e representados na **Figura 36**, permitem a clonagem de fragmentos de genómico com cerca de 1Mpb. Contêm elementos característicos dos cromossomas de leveduras: um centrómero (CEN), telómeros (TEL), e uma origem de replicação funcional em leveduras e autónoma (ARS, “autonomously replicating sequence”).



(Houghton, 2011)

FIGURA 37 – REPRESENTAÇÃO DE UM VECTOR YAC.

O vector contém os seguintes elementos estruturais de levedura: telómeros (TEL), um centrómero (CEN) e uma origem de replicação autónoma (ARS). Tem 3 marcadores de selecção em levedura (*URA3*, *TRP1* e *SUP4*), e um marcador de selecção (*Amp^r*) e uma origem de replicação para *E.coli*.

Na clonagem de DNA genómico num vector do tipo YAC, o DNA é parcialmente digerido com *EcoRI*. Os fragmentos da ordem dos 400-500Kpb são separados por PFGE. O vector é digerido com a *EcoRI*, e também com a *BamHI*, ficando os telómeros nas extremidades do vector linearizado. O fragmento situado entre os dois locais de restrição de *BamHI* é removido por electroforese. O DNA genómico é ligado ao vector e usado para a transformação de leveduras. Devido à sua inserção, o gene *SUP4* é inativado, deixando de haver expressão proteica, o que leva ao aparecimento de colónias vermelhas (King, 2011).

A ARS funciona de forma independente da origem de replicação cromossomal da levedura hospedeira. Têm pelo menos um marcador de selecção funcional em leveduras (normalmente o gene URA3, ligado à síntese de uracilo) e dois locais únicos de reconhecimento para as enzimas *EcoRI* e *BamHI*. Para permitir a sua propagação e manipulação em *E. coli* previamente à introdução de DNA genómico, estes vectores têm também uma origem de replicação bacteriana e um marcador de selecção (normalmente resistência a um antibiótico), que permite a detecção de transformantes em *E. coli*.

Estes vectores foram criados para conseguir clonar alguns genes eucariotas para os quais a clonagem em bactérias não originava expressão proteica, por falta de modificações pós-traducionais. No entanto, há sequências de DNA exógeno não toleradas por leveduras, e que acabam por ser removidas. Não são muito utilizados devido aos problemas de instabilidade genómica (Wolfe, 2005).

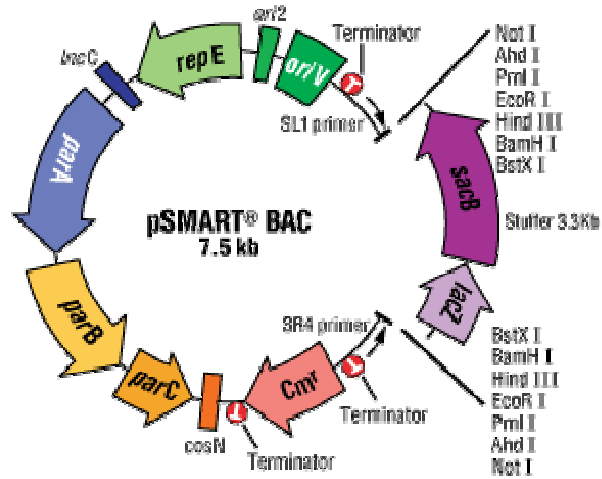
CROMOSSOMAS BACTERIANOS ARTIFICIAIS (BACs)

Os cromossomas bacterianos artificiais (BACs) baseiam-se no plasmídeo F (ou factor F, de fertilidade) de *E. coli*, que contém a informação necessária à ocorrência do processo de conjugação, pelo qual duas bactérias se ligam fisicamente, podendo ocorrer trocas de DNA. São plasmídeos muito maiores do que os vulgarmente usados como vectores de clonagem, podendo clonar fragmento de maiores dimensões, até 500Kpb (Nicholl, 2008). São muito estáveis, a informação contida no DNA do inserto mantém-se ao longo dos ciclos de replicação. Os insertos são facilmente sequenciados. O facto de serem utilizados na clonagem de fragmentos de genómico humano obtidos por bombardeamento de partículas (“shotgun”) pelo TIGR (The Institute for Genomic Research⁴⁹) levou a que só o número de fragmentos do genoma humano clonados em *E. coli* através de BACs seja de aproximadamente um milhão (BookRags, 2005). Na **Figura 37** está representado um destes vectores.

CROMOSSOMAS ARTIFICIAIS DERIVADOS DO FAGO P1 (PACs)

Os PACs (cromossomas artificiais derivados do fago P1) são vectores análogos aos anteriores, mas derivados do fago P1. Este bacteriófago tem a particularidade única de não ser integrado no genoma do hospedeiro durante o ciclo lisogénico, permanecendo no interior da célula bacteriana como plasmídeo. Têm cerca de 16Kpb. A origem de replicação é o replicação de P1, originando um plasmídeo com baixo número de cópias por célula. Quando inicia o seu ciclo lítico faz várias cópias do seu genoma, as quais ficam enredadas umas nas outras (**Figura 38**), tendo de ser posteriormente separadas (Metzenberg, 2002b).

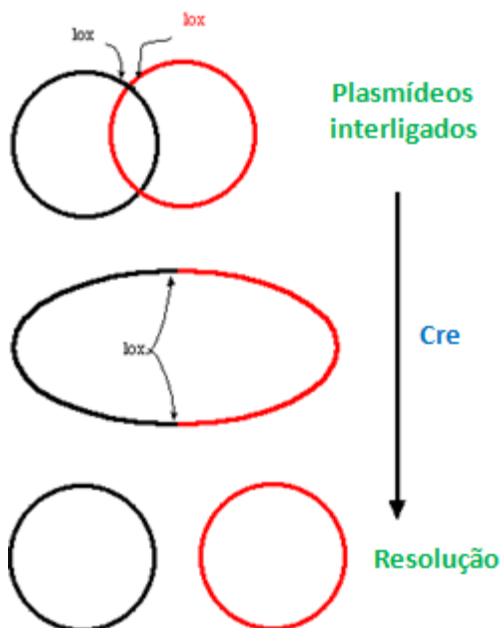
⁴⁹ Integrado em 2006 no J. Craig Venter Institute (<http://www.jcvi.org>)



(Lucigen.com, 2011)

FIGURA 38 - COPYRIGT VECTOR PSMART BAC.

CopyRight vector pSMART BAC. *ori2*, *repE*, *IncC* – origem de replicação (cópia única); *oriV* – origem de replicação indutível; *par A,B,C* – genes de partição; *Cm^r* – gene de resistência ao cloranfenicol; *cosN* – sinais de empacotamento em lambda; T – terminadores de transcrição; *sacB* – gene da sacarase. A posição aproximada dos primers para sequenciação e dos terminadores de transcrição está indicada.

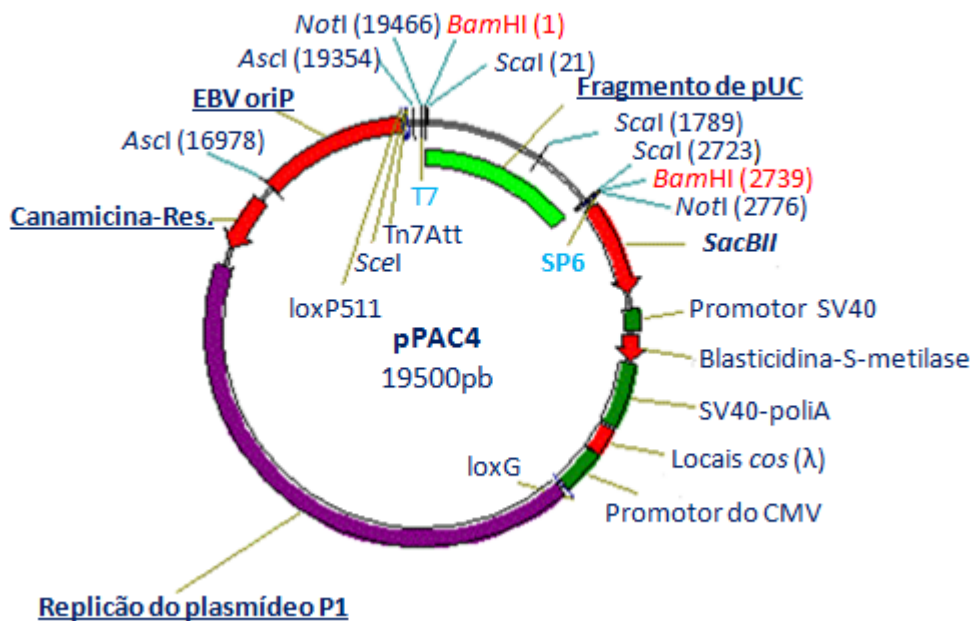


(Metzenberg, 2002b)

FIGURA 39 – RESOLUÇÃO DE PLASMÍDEOS ENREDADOS APÓS REPLICAÇÃO POR ACÇÃO DA CRE RECOMBINASE DO FAGO P1.

A Cre recombinase detecta os locais *lox* em cada um dos genomas e provoca a recombinação. Após duas voltas efectua a separação de monómeros.

O bacteriófago P1 usa um sistema de recombinação local específica (“site-specific recombination”) responsável pela partição do genómico sintetizado durante a replicação (Hoess & Abremski, 1984). De acordo com estes autores, este sistema é composto pela Cre recombinase que medeia a partição entre os dois locais *loxP* durante a recombinação. Na **Figura 39** está representado um vector deste tipo, o pPAC 4.



(Anónimo, sd)

FIGURA 40 - VECTOR PPAC4.

A selecção positiva dos clones é feita através do gene *sacBII*, responsável pela conversão de sacarose a levana, uma polifrutose tóxica para *E. coli*. No vector um fragmento derivado do vector pUC (vector multi-cópia) está inserido no “polylinker”, inactivando *SacBII*, e permitindo a purificação de grandes quantidades de vector. Locais de reconhecimento de *BamHI* flanqueiam o fragmento “stuffer”. O inserto, com adaptadores *BamHI*, substitui este fragmento, e o vector regressa a cópia em reduzido número. Nem o vector nem os clones recombinantes expressam o gene *SacBII*, só os que perderam o “stuffer” são sensíveis à sacarose, não formando colónias nesse meio. Tem um gene de resistência à canamicina, para selecção em *E. coli*. Sequências dos promotores da SP6 e T7 RNA polimerases do fago a flanquear o “stuffer”, ladeados por locais de reconhecimento para *NotI*. Local de reconhecimento para *Sccl*, extremamente raro e inexistente no genoma humano, para permitir a linearização do vector. Origem de replicação funcional nalgumas células eucarióticas - replicação do vírus Epstein Barr (EBV oriP). Um local *loxG* que, em conjunto com *loxP511* e *Cre* permite a integração de DNA no genoma de mamíferos de uma forma local específica. Gene de resistência à blasticidina para selecção em células eucarióticas (cultura de tecidos). Local SV40 de poliadenilação, que funciona como promotor de transcrição do citomegalovírus (CMV). Sítios *cos* do fago lambda. (Frengen, et al., 2000).

ANEXO 3 – CONDIÇÕES DE INFECÇÃO DE *E. COLI* (“LIBRARY SCREENING”)

Para a infecção de *E. coli* LE392 iniciou-se o crescimento a partir de uma colónia em LB líquido (suplementado com maltose e MgSO₄, que estimulam a posterior capacidade de infecção do fago), até ser atingida uma densidade óptica de 0,8 e 1,0 a 600nm, correspondente à fase exponencial de crescimento. As células foram centrifugadas e ressuspensas numa solução de 10mM de MgSO₄ até se atingir um valor de DO de 0,5 a 600nm. Adicionaram-se alguns microlitros da biblioteca de DNA de forma a obter a formação de cerca de 15000 pfu –“plaque forming units” - por placa de 150mm (o título do fago era de $\sim 2,2 \times 10^5$ pfu/ μ l. Incubaram-se os tubos de Eppendorf contendo o fago e as células hospedeiras a 37 °C durante 15 minutos para permitir que o fago se ligasse às células. Posteriormente as bactérias infectadas foram misturadas com NZY “top agar” (previamente fundido e arrefecido até aos 49 °C) e a mistura colocada em cada placa de 150mm com NZY “bottom agar”, rodando a placa para distribuir a mistura aleatoriamente por toda a superfície. As placas de NZY “bottom agar” foram pré-aquecidas a 37 °C em estufa antes de ser usadas, para facilitar a distribuição do NZY “top agar” infectado em todas as direcções e para evitar que este solidificasse logo ao entrar em contacto com a placa. As placas foram invertidas e incubadas a 37 °C durante a noite.