

PROSPECÇÃO DE GENES EM BIBLIOTECAS DE cDNA

GENE PROSPECTION IN cDNA LIBRARIES

Gaspar Malone¹; Paulo Dejalma Zimmer²; Geri Eduardo Meneghelo³;
Eliseu Binneck⁴; Silmar Teichet Peske⁵

- REVISÃO BIBLIOGRÁFICA -

RESUMO

A prospecção de genes é um dos desafios atuais da ciência. Profissionais da bioinformática, em parceria com biólogos moleculares, precisam analisar grandes volumes de informação. A prospecção de genes, embora tenha avançado bastante nos últimos tempos, ainda constitui-se numa área a ser explorada, principalmente nos organismos eucarióticos. Envolve a combinação de análises *in silico* e de análises moleculares. A estratégia de prospecção de genes inicia com a construção de uma ou mais bibliotecas de cDNA. Posteriormente, é realizada uma busca de seqüências por homologia, dos genes de interesse, nos bancos de dados públicos. Em seguida, das seqüências selecionadas são identificados domínios conservados, através do alinhamento múltiplo. De posse dessa informação pode-se desenhar **primers** degenerados para triagem das bibliotecas de cDNA, analisadas para a localização de um ou mais genes de interesse, utilizando sondas marcadas ou estratégias de amplificação por PCR. O espetacular avanço da bioinformática nos últimos anos, aliado à crescente capacidade computacional e velocidade de acesso a Internet, facilitam a realização de novas análises. Trabalhar no desenvolvimento de metodologias e algoritmos que permitam elucidar todas essas combinações entre genes e proteínas é, sem dúvida, a nova fronteira da bioinformática e da biologia molecular.

Palavras-chave: Bioinformática, biologia molecular, proteína, alinhamento, primers degenerados.

ABSTRACT

Gene prospection is one of the current challenges for science. Molecular biologists and Bioinformatics need to analyze a large amount of information. Gene prospection still constitutes an area to be explored, mainly in eukaryotes. It involves the combination of bioinformatics, with *in silico* analyses, and molecular biology with the construction and screening of cDNA libraries. The strategy of gene prospection begins with the construction of cDNA libraries, followed by a search for homology in public databases. Soon afterwards, conserved regions are identified on the selected sequences, throughout multiple alignments. With this information, degenerate primers may be constructed for cDNA screening, which are analyzed to screen one or more genes, using labeled probes or DNA amplification by PCR. The spectacular progress of bioinformatic in the last few years, and the growing computational capacity and speed access on the Internet, helped to accomplish new analyses. The development of methodologies and algorithms that allow elucidating all genes and proteins combinations is the new border of the bioinformatics and the molecular biology.

Key words: Bioinformatics, molecular biology, protein, alignment, degenerate primers.

INTRODUÇÃO

O século XXI abre as portas para a era da genômica, marcada por acontecimentos históricos, como o sequenciamento de organismos modelos (*Arabidopsis thaliana* e *Drosophila melanogaster*) e espécies de interesse agrícola como o arroz e a cana-de-açúcar (http://igweb.integratedgenomics.com/ERGO_supplement/genomes_all.html #). A disponibilidade dos genomas seqüenciados tem provocado um aumento no volume de informação, gerando dificuldades de armazenamento e manipulação, exigindo recursos computacionais cada vez mais eficientes e pessoas cada vez mais capacitadas na análise de dados. A análise detalhada e a correta interpretação da informação existente têm propiciado condições para o desenvolvimento de estudos que até então eram inusitados. Esses estudos exigem profissionais com conhecimento em bioinformática, com excelente formação em biologia molecular, sólidos conhecimentos em informática e desenho de softwares. Profissionais com estas qualificações têm sido procurados pelas instituições públicas e/ou privadas que apostam na bioinformática como ferramenta promissora na identificação e caracterização de genes de interesse. Por outro lado, ferramentas computacionais adequadas, equipamentos com grande capacidade de armazenamento e devidamente preparados para acesso *on line* aos bancos de dados públicos, se constituem em requisitos indispensáveis.

Tradicionalmente, o enfoque das pesquisas consistia em conhecer uma determinada função e buscar o gene responsável. Atualmente, a situação é inversa, dispõe-se de um enorme número de genes desconhecidos que demandam caracterização. É aqui que surge o termo "Prospecção de genes".

O objetivo deste trabalho é contribuir com informações e estratégias que possam incentivar a pesquisa na área de prospecção de genes. Visa também apresentar as diferentes ferramentas disponíveis para que sua utilização combinada possa ser otimizada na prospecção de genes.

A identificação da função e dos fatores responsáveis pela expressão gênica, se constitui no foco principal dos

¹ Geneticista, M.Sc., Doutorando em Ciência e Tecnologia de Sementes. FAEM/UFPEL. C.P. 354. 96001-900 Pelotas - RS E-mail: gmalone@ufpel.edu.br

² Eng Agr. Dr. Professor do PPG em Ciência e Tecnologia de Sementes FAEM/UFPEL. C.P. 354. 96001-900 Pelotas - RS E-mail: djzimmer@ufpel.edu.br

³ Eng. Agr., M.Sc., Doutorando em Ciência e Tecnologia de Sementes. FAEM/UFPEL. C.P. 354. 96001-900 Pelotas - RS E-mail: geriem@ufpel.edu.br

⁴ Eng Agr. Dr. Consultor IICA, Embrapa Soja. CP 231. 86001-970. Londrina - PR. E-mail: binneck@cnpso.embrapa.br

⁵ Eng Agr. PhD. Professor Titular do PPG em Ciência e Tecnologia de Sementes FAEM/UFPEL. C.P. 354. 96001-900 Pelotas - RS E-mail: peske@ufpel.edu.br

“projetos genoma”. Embora a prospecção de genes tenha avançado bastante nos últimos tempos, ainda há um longo caminho a percorrer, principalmente nos organismos eucariotos, onde os genes não estão numa distribuição contínua, pois se encontram distribuídos em diferentes cromossomos e separados por regiões intergênicas. Além

disso, cada gene apresenta-se fragmentado em regiões codificantes (exons), separadas por regiões não codificantes (introns). Na figura 1 é apresentada a estrutura de um gene eucariótico e a produção de um mRNA pela retirada (*splicing*) dos introns (mRNA maduro, pronto para ser traduzido à proteína) durante o processamento da molécula de RNA.

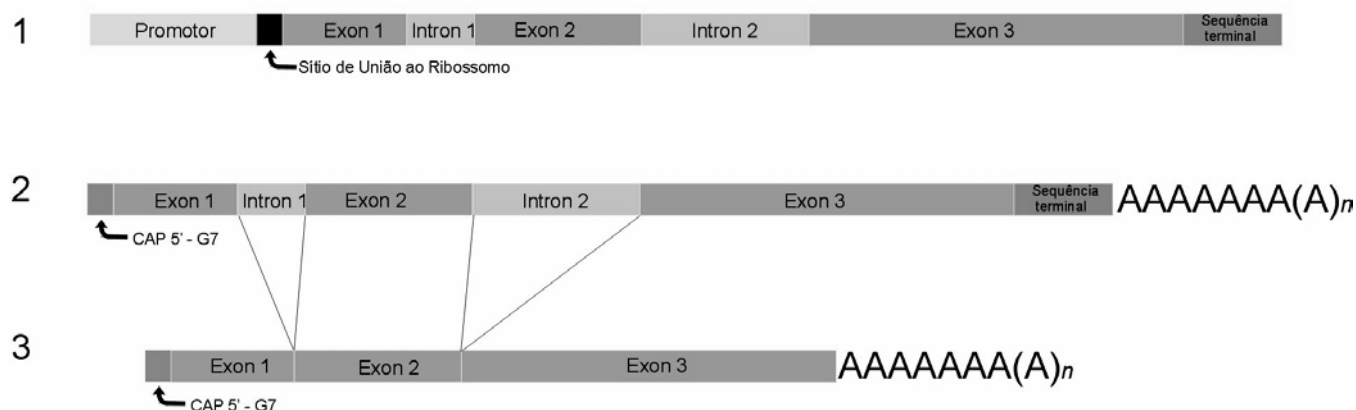


Figura 1 - Esquema geral do processamento de um gene eucarioto. Após a transcrição pela RNA polimerase (1) o RNA heterogêneo deve ser processado (2 e 3) e isso consiste na retirada dos introns, na adição da cauda Poli-A e do CAP 5'. O RNAm é então transferido para fora do núcleo da célula onde será traduzido para uma proteína.

A prospecção de genes está intimamente ligada à bioinformática, uma ferramenta imprescindível para a análise massiva de dados biológicos, que permite identificar genes de

interesse a partir da conjugação de várias áreas da ciência e dados disponibilizados de projetos de sequenciamento de DNA e proteínas (BINNECK, 2004) (Figura 2).

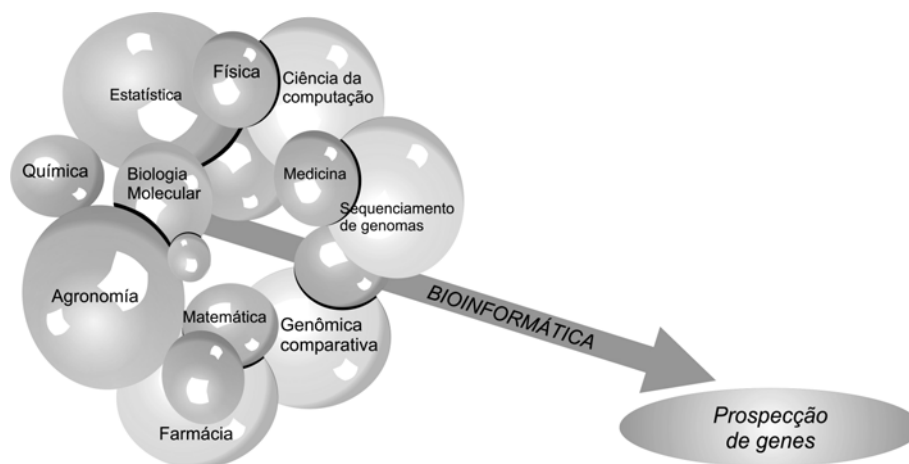


Figura 2 - A Bioinformática como ferramenta integradora de informação.

A utilização das informações dos bancos de DNA e proteínas é de grande utilidade para a elucidação da função e expressão de novos genes. Os dados disponíveis podem ser utilizados de duas formas, direta ou indiretamente. A utilização direta implica em identificar seqüências similares, ou com propriedades comuns, e estabelecer relações entre elas (por exemplo, filogenia molecular). A aplicação indireta é feita através da predição de regiões e seqüências de genes

relacionados, em espécies desconhecidas, através de alinhamento local (BLAST) e é sobre esse ponto que vamos focar esta abordagem. A Figura 3 ilustra o procedimento utilizado pela conjugação da bioinformática e da biologia molecular para obter informação dos bancos públicos de seqüências, e utilizar essas informações para a predição de seqüências relacionadas em outras espécies, cujas seqüências genômicas ainda não são conhecidas.

A estratégia de prospecção de genes de interesse em genótipos não seqüenciados, através da utilização de informações geradas em projetos genoma ou proteoma, envolve as seguintes etapas: procura de seqüências nos bancos de dados, alinhamento múltiplo das seqüências, identificação dos domínios conservados, desenho de *primers* degenerados, construção das bibliotecas de cDNA, triagem das bibliotecas de cDNA. Estas etapas serão abordadas detalhadamente na seqüência.

Procura de seqüências nos bancos de dados

O primeiro passo na prospecção de genes é analisar detalhadamente as informações já disponíveis para localizar e caracterizar seqüências genômicas relacionadas ao gene ou a proteína da qual desejamos localizar o gene responsável pela sua expressão.

Os programas utilizados para a análise de seqüências podem ser encontrados em pacotes comerciais ou servidores públicos, como o EMBOSS (<http://www.uk.embnnet.org/Software/EMBOSS/>) e outros. Atualmente, a tendência é utilizar programas em servidores públicos, os quais oferecem um sistema de análise de dados integrado através da *Web*. A principal vantagem é o baixo custo na implantação da infraestrutura computacional. Um simples terminal de conexão a Internet com boa velocidade de conexão é suficiente. Além disso, os *sites* onde são disponibilizados programas e bancos de dados, geralmente possibilitam a utilização das versões mais recentes e atualizadas, coisa que não acontece com os pacotes de programas comerciais.

A procura de seqüências de DNA e de proteínas nos bancos de dados públicos é uma metodologia eficiente para a obtenção de informações sobre estrutura e função de genes e moléculas. Como exemplo de bancos de dados em que podem ser encontradas informações sobre genes e proteínas pode-se citar: *GenBank/NCBI – National Center for Biotechnology Information* (<http://www.ncbi.nlm.nih.gov/>), *EBI – European Bioinformatics Institute* (<http://www.ebi.ac.uk/>) *DDBJ – DNA Data Bank of Japan* (<http://www.ddbj.nig.ac.jp/>), *PDB – Protein Data bank* (<http://www.rcsb.org/pdb/>), *GDB – The Genome Database* (<http://www.gdb.org/>), *TIGR Databases – The Institute for Genomic Research* (<http://www.tigr.org/tdb/>), *PIR – Protein Information Resource* (<http://www.nbrf.georgetown.edu/>), *SWISS-PROT – Protein knowledgebase* (<http://us.expasy.org/sprot/>), *KEGG – Kyoto Encyclopedia of Genes and Genomes* (<http://www.genome.ad.jp/kegg/>), *Soybean Genomics Initiative* (<http://soybean.ccgb.umn.edu/>), *Gramene: A Comparative Mapping Resource for Grains* (<http://www.gramene.org/>) e o *Seed Genes Project* (<http://www.seedgenes.org/>) (PROSDOCIMI et al., 2001).

Depois de identificadas as informações existentes sobre o gene ou a proteína alvo, deve-se realizar buscas por similaridade. Os programas mais utilizados atualmente, usam heurísticas computacionais para acelerar o processo de busca, entre os quais estão o BLAST (*Basic Local Alignment Search Tools*) (ALTSCHUL et al., 1990, 1997) e o FASTA (LIPMAN & PEARSON, 1985; PEARSON & LIPMAN, 1988). Ambos procuram seqüências similares à seqüência alvo que poderão ser utilizadas para o subsequente alinhamento

múltiplo. No entanto, a manipulação de um grande número de seqüências requer tempo, conhecimento multidisciplinar e, fundamentalmente, a escolha de ferramentas informáticas apropriadas para cada situação.

Alinhamento múltiplo das seqüências

O alinhamento múltiplo permite identificar domínios de aminoácidos conservados, a partir da seqüência primária das proteínas alinhadas. Para isto, são utilizados programas de formação de *Clusters* de seqüências (CORPET, 1980), os quais, a partir da estimativa de semelhança entre as seqüências par a par, alinham-nas para formar os agrupamentos e identificar as regiões conservadas. A eficiência do alinhamento múltiplo depende do grau de similaridade/identidade entre as seqüências analisadas.

Deve ser destacado que, a análise de alinhamentos múltiplos apresenta enorme complexidade, sendo praticamente inviável a sua execução manual (BARBER & MAIZEL, 1990).

Identificação dos domínios conservados

A identificação de domínios conservados baseia-se no fato de que pequenas regiões idênticas, em proteínas homólogas, podem ser identificadas, caracterizadas e utilizadas para a localização de domínios similares em outras espécies em que este gene e/ou proteína ainda não foram exploradas. Esses domínios permanecem conservados, pois algumas regiões das proteínas são fundamentais para manter as suas propriedades biológicas, sofrendo forte pressão de seleção durante a evolução, o que as manteve praticamente inalteradas ao longo do tempo, facilitando a sua localização por análises de alinhamentos de seqüências.

As seqüências de aminoácidos conservados podem ser utilizadas para o desenho de *primers* degenerados, que serão utilizados posteriormente para o triagem de bibliotecas de cDNA em outras espécies (REPASKE et al., 1992; KOBAYASHI et al., 2002). Observe detalhes do alinhamento múltiplo e identificação de domínios conservados na figura 4.

Desenho de primers degenerados

O termo *primers* degenerados origina-se do fato de que para cada aminoácido do domínio conservado, existem, geralmente (excetuando Metionina e Triptofano), mais de um códon de 3 bases codificando cada aminoácido (ALBERTS et al., 1994) (Tabela 1). O desenho de *primers* é um processo metódico, que requer conhecimento prévio de alguns conceitos da biologia molecular como por exemplo: “*primers dimers*” (pareamento de *primers* por complementaridade de seqüência na extremidade 3’), “*loop*” (porção de fita simples no final de um *hairpin* numa molécula de DNA ou RNA), “*hairpin*” (estrutura secundária ocasionada pela homologia de seqüências internas na fita do DNA), entre outros. Recomenda-se também construir *primers* com conteúdo GC superior a 50% e um comprimento mínimo de 22–24 pares de bases. O que permitirá a utilização de temperaturas de anelamento em torno de 56 °C (maior especificidade de amplificação). Os *primers* são sempre construídos na direção 5’- 3’, sendo que são necessários dois *primers* para a amplificação (LINHART & SHAMIR, 2002).

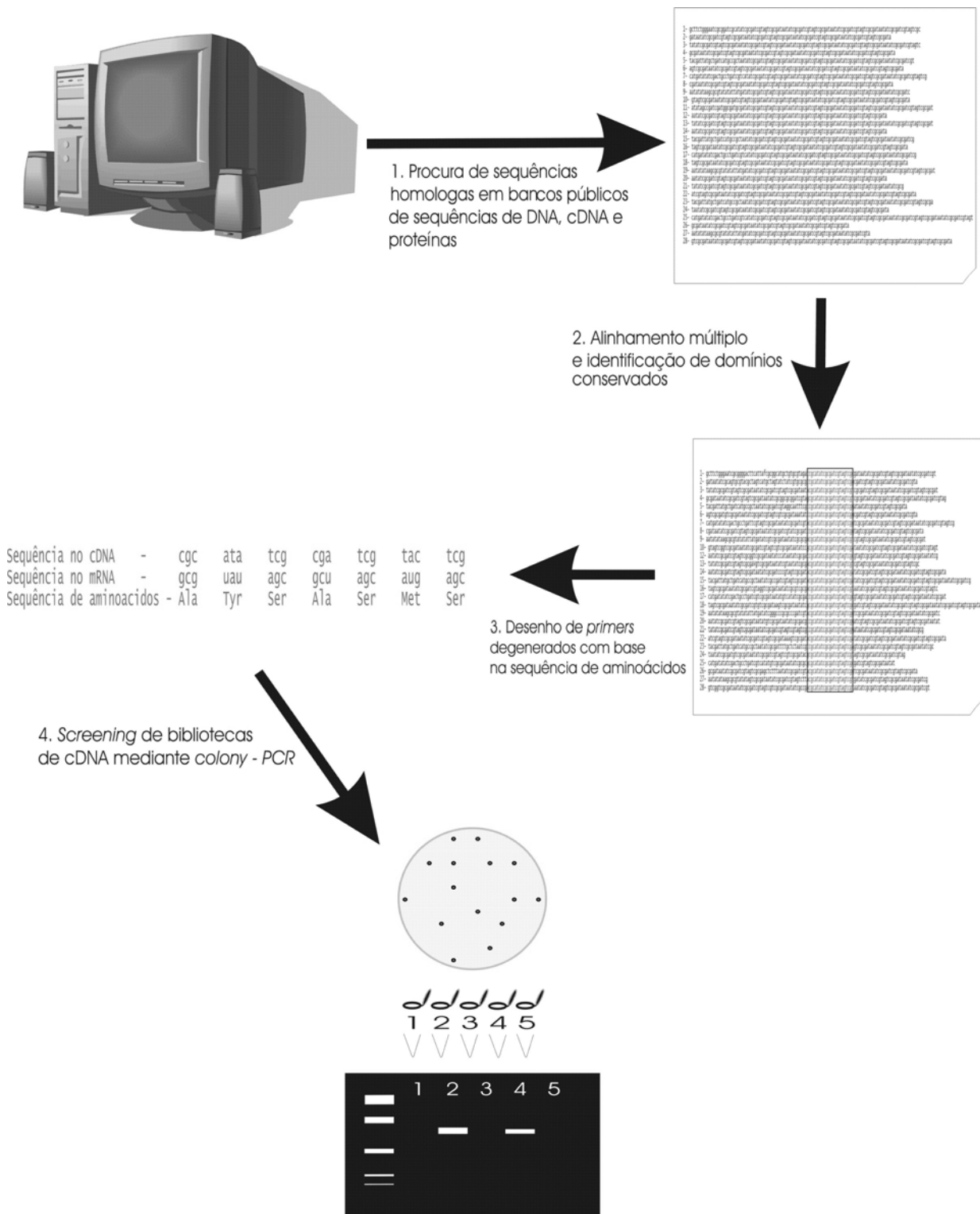


Figura 3 - Diagrama ilustrativo dos passos envolvidos na prospecção de genes em bibliotecas de cDNA a partir de análise *in silico* e desenho de *primers*. Num segundo momento, os *primers* são utilizados em reações de PCR para selecionar os clones de interesse nas bibliotecas de cDNA.

Espécie 1	---ANYGGDKQYG	RETRHTGDYE	NPIHSTGGQY	DQDVRQTDEY	GNPVRRTDEY	50
Espécie 2	---ANYGGDKQYG	RETRHTGDYE	NPIHSTGGQY	DQDVRQTDEY	GNPVRRTDEY	50
Espécie 3	---ANYGGDKQYG	RETRHTGDYE	NPIHSTGGQY	DQDVRQTDEY	GNPVRRTDEY	50
Espécie 4	---ANYGGDKQYG	RETRHTGDYE	NPIHSTGGQY	DQDVRQTDEY	GNPVRRTDEY	50
Espécie 5	---ANYGGDKQYG	RETRHTGDYE	NPIHSTGGQY	DQDVRQTDEY	GNPVRRTDEY	50
51	GNPVHTATGG	TMGDYGSTGL	GQGTGGIGTG	GYGATGHQGL	GTGVGHTTG-	100
51	GNPVHTATGG	TMGDYGSTGL	GQGTGGIGTG	GYGTTGHQGL	GTGVGHTTG-	100
51	GNPVHSTTGG	TMGDYGSTGL	GRGTGGIGTG	GYGTTGHQGL	GTGVGHTTGG	100
51	GNPVHSATGG	TMGDYGSTGL	GQGTGGIGTG	GYGTTGHQGL	GTGVGHTTGG	100
51	GNPVHSTTGG	TMGDYGSTGL	GRGTGGIGTG	GYGTTGHQGL	GTGIGHTTGG	100
101	TGTDYTSGGR	STGQTYQGL	GTESEFGGTT	GTFQNQPSAT	PIGGTGLSSG	150
101	TGTDYTPGGR	STGQTYQGL	GTESEFGGTT	GTFQNQPSAT	PIGGTGLSSG	150
101	TGTDYTSGGR	STGQTYQGL	GTESEFGGTT	GTFQNQPSAT	PIGGTGLGGG	150
101	TGTDYTSGGH	STGQTYQGL	GTESEFGGTT	GTFQNQPSAT	PMGGTGLGSG	150
101	TGTDYTSGGR	STGQTYQGL	GTESEFGGTT	GTFQNQPSAT	PIGGTGLGGG	150
151	TGAGLGGTGT	GTGTGILHRS	GSGSSSSSED	DGQGGRRKKK	GVMQKIKEKL	200
151	TGAGFGGTGT	GTGTGILHRS	GSGSSSSSED	DGQGGRRKKK	GVMQKIKEKL	200
151	TGAGFGGTGT	GTGTGILHRS	GSSSSSSSED	DGQGGRRKKK	GVMQKIKEKL	200
151	TGAGFGGTGT	GTGTGILHRS	GSGSSSSSED	DRQGGRRKKK	GVMQKIKEKL	200
151	TGAGFGGTGT	GTRTGILHRS	GSSSSSSSED	DGQGGRRKKK	GVMQKIKEKL	200
201	PGGHSQEEQY	QSQTTHTTT--	---GYGETHE	KKGMMEKIKE	KLPGHH	246
201	PGGHSQEEQY	QSQTTHTTT--	---GYGETHE	KKGMMEKIKE	KLPGHH	246
201	PGGHSQEGQY	QSQTTHTTT--	GGAGYGETHE	KKGMMEKIKE	KLPGHH	246
201	PGGHSQEEQY	QSQTTHTTT--	---GYGETHE	KKGMMEKIKE	KLPGHH	246
201	PGGHSQEGQY	QSQTTHTTT--	GGAGYGETHE	KKGMMEKIKE	KLPGHH	246

Figura 4 - Alinhamento múltiplo de 5 proteínas LEA's (*Late Embryogenesis Abundant*) pertencentes a cinco espécies diferentes do gênero *Helianthus*. Regiões sombreadas indicam dois domínios conservados nas cinco espécies, possíveis locais para o desenho de *primers* degenerados.

Tabela 1 - Diferentes combinações de nucleotídeos obtidas a partir de uma seqüência de sete aminoácidos. A natureza degenerada do código genético pode inviabilizar o desenho de *primers* quando a seqüência de aminoácidos for demasiadamente grande.

Met	Ile	His	Trp	Cys	Pro	Leu	TOTAL						
ATG	ATT	CAT	TGG	TGT	CCT	TTT							
	ATC	CAC		TGC	CCC	TTG							
	ATA				CCA	CTT							
					CCG	CTC							
						CTA							
						CTG							
1	x	3	x	2	x	1	x	2	x	4	x	6	288

O desenho do *primers* degenerados, considerando o exemplo anterior, inclui a utilização de 288 combinações, cada uma delas representando uma seqüência possível. Ou seja, seriam necessários 288 *primers* em quantidades equimolares para se ter uma segurança de 100% de que a seqüência genômica pertencente ao domínio conservado vai ser amplificada. Indubitavelmente, é um procedimento inviável. A alternativa para isso é a utilização de bases modificadas que apresentam a particularidade de se parear com mais de uma base. Por exemplo, a base modificada Inosina pode parear com Adenina, Timina e Citosina, o que reduz notavelmente o número de *primers* necessários (KWOK et al., 1994).

Existem vários programas disponíveis para auxiliar o

desenho de *primers*, dentre eles o CODEHOP (ROSE et al., 1998), e páginas *Web* gratuitas que possibilitam trabalhar *on line*, sem a necessidade de instalação dos programas no computador. O programa *Primer3*, disponível de forma gratuita (http://frodo.wi.mit.edu/cgi-bin/primer3/primer3_www.cgi), é uma alternativa interessante, pois apresenta grande simplicidade na utilização e não requer nenhum investimento para a aquisição de programas licenciados. Outros programas gratuitos para o desenho de *primers* podem ser obtidos na *Web*: <http://bioinformatics.org/annhyb/>; <http://alces.med.umn.edu/rawtm.html>; <http://www.ncbi.nlm.nih.gov/>.

Construção das Bibliotecas de cDNA

Esse termo faz referência a uma coleção de fragmentos de cDNA (DNA complementar a um mRNA), representando as seqüências expressas em um determinado tecido ou órgão, sob influência de uma determinada condição do ambiente, num determinado momento. Esses fragmentos são normalmente armazenados em vetores de clonagem que, devido a sua estabilidade podem ser armazenados por períodos prolongados de tempo (BIRREN et al., 1999).

Ao contrário de uma biblioteca genômica, que contém em princípio qualquer fragmento de DNA do organismo doador, a biblioteca de cDNA contém os genes expressos no tempo e local específicos. As moléculas de mRNA (altamente instáveis) são convertidas à cDNA para a sua clonagem e armazenamento.

As bibliotecas de cDNA apresentam várias vantagens quando o objetivo é a identificação de genes diferencialmente expressos ou quando se está prospectando um determinado gene, do qual se têm apenas algumas seqüências do gene homólogo de outra espécie. A principal vantagem é a ausência

de introns nas seqüências clonadas, o que facilita a identificação de seqüências expressas, pois somente a seqüência codificante (exons) é clonada.

Uma biblioteca de cDNA devidamente construída deverá: i) representar todos os mRNA expressos num determinado tecido e estágio de desenvolvimento; ii) conter clones com seqüências de cDNAs provenientes de mRNAs completos; iii) possuir número suficiente de clones que garantam a representatividade de todos os mRNA e; iv) ser estável, para que possibilite um correto armazenamento por períodos prolongados de tempo.

A geração de uma biblioteca de cDNA inclui, primeiramente, a extração de mRNA do tecido ou organismo de interesse, a síntese de cDNA a partir do mRNA, a clonagem das moléculas de cDNA em um vetor de clonagem e a incorporação do vetor em uma bactéria para a sua multiplicação (transformação bacteriana). Na figura 5 é ilustrado, de forma resumida, os passos envolvidos na construção de uma biblioteca de cDNA.

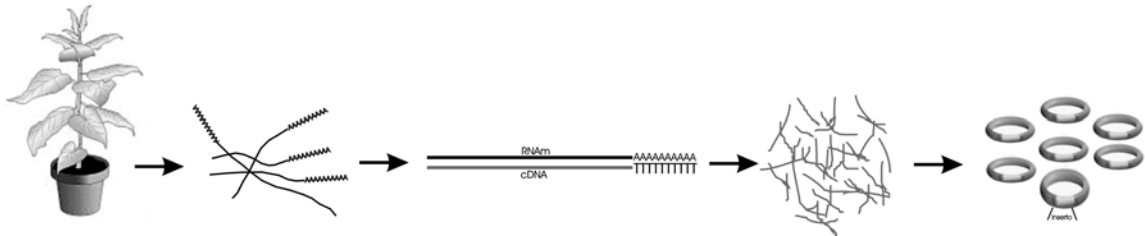


Figura 5 - Esquema representativo das principais etapas na geração de bibliotecas de cDNA. 1 – Extração de RNA mensageiro; 2 – Síntese reversa utilizando RT-PCR; 3 – Síntese de cDNA dupla fita; 4 – Clonagem dos cDNAs.

Triagem da biblioteca de cDNA

As bibliotecas de cDNA podem ser acessadas para a identificação de um ou mais clones de interesse, o que pode ser feito utilizando-se sondas (marcadas química ou fluorescentemente) ou *primers*.

As sondas são provenientes de uma seqüência de cDNA já conhecida, um produto de PCR (*Polymerase Chain Reaction* ou Reação em Cadeia da Polimerase) ou fragmentos genômicos. Também é possível construir uma sonda a partir da seqüência de aminoácidos de uma proteína de interesse. Neste caso a sonda é obtida por tradução reversa da seqüência de aminoácidos (Oligo degenerado). No entanto, a triagem de bibliotecas utilizando sondas de cDNA, de produtos de PCR ou de seqüências genômicas apresenta algumas limitações, dentre elas, problemas relacionados ao anelamento inespecífico da sonda de hibridização e ao volume de trabalho requerido.

A triagem de bibliotecas utilizando PCR é uma ferramenta muito eficiente na identificação e isolamento de clones de interesse dentro de uma biblioteca (ISRAEL, 1993; ALFANDARI & DARRIBERE, 1994; MUNROE et al., 1995). Dentre as principais vantagens da triagem por PCR podem ser destacadas as seguintes: i) Os clones falso-positivos podem ser facilmente reconhecidos em função do tamanho do fragmento amplificado; ii) A análise de centenas de clones por dia, o que reduz notavelmente o tempo para a triagem completa da biblioteca, e; iii) Permite utilizar vários conjuntos de *primers* na mesma reação de PCR, o que possibilita realizar a triagem de vários genes simultaneamente.

Em função do elevado número de triagens necessárias para estudo de uma biblioteca, é necessário que as condições

de PCR sejam extremamente padronizadas. Dependendo da abundância de expressão do gene de interesse, o número de clones analisados pode facilmente superar os 2000.

A metodologia baseada em PCR mais amplamente utilizada na triagem de bibliotecas de cDNA é a *Colony – PCR* (ou PCR diretamente da colônia bacteriana) (BOTTOLI et al., 1999; SANTOS et al., 2000; OTSUKA et al., 2004). O procedimento é relativamente simples, eficiente e rápido. Uma única colônia transformante é conjugada com os componentes comuns a uma reação de PCR e submetida à amplificação. Os *primers* desenhados com base nos domínios protéicos conservados são utilizados em todas as colônias da biblioteca de cDNA. Após amplificação, os produtos da reação são visualizados mediante eletroforese em géis de agarose ou poliacrilamida. Dessa forma, os clones contendo a seqüência de cDNA complementar ao domínio conservado serão identificados, podendo ser seqüenciados e caracterizados em estudos posteriores. Dependendo da qualidade da biblioteca de cDNA, pode ser difícil identificar um clone com a seqüência completa do gene. É por isso que se faz necessário o sequenciamento dos clones positivos, para obter a seqüência completa, ou pelo menos, uma porção do gene alvo. Estratégias de clonagem e PCR inversa poderão auxiliar na obtenção da obtenção da seqüência completa do gene.

CONSIDERAÇÕES FINAIS

Foi apresentada uma aplicação da bioinformática, combinada com técnicas moleculares que está sendo utilizada para começar a desvendar, caracterizar e melhor aproveitar a informação contida nos bancos públicos de seqüências de

DNA e proteínas. O espetacular avanço da bioinformática nos últimos anos, aliado à crescente capacidade computacional e velocidade de acesso a *Internet*, facilitam a realização de análises complexas a partir de qualquer parte do mundo, evidenciando que a preocupação atual dos cientistas já não é a geração de informação, mas sim a análise e interpretação rápida do que já está disponível.

O impressionante desenvolvimento de novos métodos experimentais, a possibilidade de combinar a informação contida em seqüências e estruturas, para obter novos dados sobre níveis de expressão gênica de todos os genes (genômica funcional) e de todas as funções e interações entre proteínas (proteômica), abrem a porta para uma nova era, cheia de incógnitas e desafios. A era da pós genômica.

Trabalhar em cima do desenvolvimento de metodologias e algoritmos que permitam elucidar todas essas combinações entre genes e proteínas é, sem dúvida, a nova fronteira da ciência.

REFERÊNCIAS

- ALBERTS, B.; BRAY, D.; LEWIS, J. et al. **Molecular Biology of the Cell**. New York, Garland Publishing, 1994. 1294 p.
- ALFANDARI, D.; DARRIBERE, T. A simple PCR method for screening cDNA libraries. **Genome Research**, Stanford, v. 4, p. 46-49, 1994.
- ALTSCHUL, S.F.; GISH, W.; MILLER, W. et al. Basic local alignment search tool. **Journal of Molecular Biology**. London, v. 98, p.12473-12478. 1990.
- ALTSCHUL, S. F.; THOMAS, L. M.; ALEJANDRO, A. S. et al. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. **Nucleic Acids Research**, Oxford, v. 25, p. 3389-3402. 1997.
- BARBER, A.M.; MAIZEL, J.V.J Sequence editing aligner : A multiple sequence editor and aligner, **Genetic Analysis: Techniques and Applications**, New York, v. 7, p. 39-45. 1990.
- BINNECK, E. As Ômicas: integrando a bioinformação. **Biotecnologia Ciência e Desenvolvimento**. Brasília, N. 32. p. 28-37, 2004.
- BIRREN, B.; GREEN, E.D.; MYERS, R.M. et al. **Genome Analysis: a Laboratory Manual**. Cold Spring Harbor Laboratory Press. v.2. 1999. 528 p.
- BOTTOLI, A.P.; KERTESZ-CHALOUPOVA, K.; BOULIANNE R.P. et al. Rapid isolation of genes from an indexed genomic library of *C. cinereus* in a novel pab1+ cosmid. **Journal of Microbiological Methods**, Amsterdam, v.35, n.2, p.129 -141, 1999.
- CORPET, F. Multiple sequence alignments with hierarchical clustering. **Nucleic Acids Research**, Oxford, v. 16, p.10881-10890, 1980.
- ISRAEL, D. I. A PCR-based method for high stringency screening of DNA libraries **Nucleic Acids Research**, Oxford, v. 21, p. 2627-2631, 1993.
- KOBAYASHI, M.; NAKAGAWA, H.; SUDA, I. et al. Purification and cDNA Cloning of UDP-D-Glucuronate Carboxy-lyase (UDP-D-xylose Synthase) from Pea Seedlings. **Plant and Cell Physiology**, Oxford, v. 43, p.1259-1265, 2002.
- KWOK, S.; CHANG, S. Y.; SNINSKY, J. J. et al. A guide to the design and use of mismatched and degenerated primers. **PCR Methods and Applications**, New York, v. 3, n. 4, p. 39-47, 1994.
- LIPMAN D.J.; PEARSON W.R. Rapid and sensitive protein similarity search, **Science**, London, v. 227, p. 1435-1444, 1985.
- LINHART, C.; SHAMIR, R. The degenerated primer desing problem. **Bioinformatics**, Oxford, v. 18, Supl. 1, p. 172-180, 2002.
- MUNROE, D. J; LOEBBERT, R; BRIC, E. et al. Systematic screening of an arrayed cDNA library by PCR. **Proceedings of the National Academy of Sciences**, Washington, v. 92, n. 6, p. 2209 - 2213. 1995
- OTSUKA, M.; ICHINOSE, K.; FUJII, I. et al. 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**, Bethesda, v. 48, n.9, p. 3468 – 3476, 2004
- PEARSON, W.R.; LIPMAN D.J. Improved tools for biological sequence comparison, **Proceedings of the National Academy of Sciences**, Washington, v. 85, n. 8, p. 2444-2448, 1988.
- PROSDOCIMI F.; CERQUEIRA, G. C.; BINNECK, E. et al. Bioinformática: Manual do usuário. **Biotecnologia Ciência & Desenvolvimento**. Brasília, n. 29. p. 12-25, 2001.
- REPASKE, D. R.; SWINNEN, J. V.; JIN, S. L. et al. A polymerase chain reaction strategy to identify and clone cyclic nucleotide phosphodiesterase cDNAs. Molecular cloning of the cDNA encoding the 63-kDa calmodulin-dependent phosphodiesterase. **Journal of Biological Chemistry**, Stanford, v. 267, p. 18683 – 18688, 1992.
- ROSE, T. M.; SCHULTZ, E. R.; HENIKOFF, J. G.; et al. Consensus-degenerated hybrid oligonucleotides primers for amplification of distantly related sequence. **Nucleic Acids Research**, Oxford, v. 26, p. 1628-1635, 1998.
- SANTOS, W. G. dos; METCHEVA, I.; BUCK, G. A. Colony polymerase chain reaction of stably transfected *Trypanosoma cruzi* grown on solid médium. **Memórias do Instituto Oswaldo Cruz**, Rio de Janeiro, v. 95, n. 1, p. 111-114, 2000.