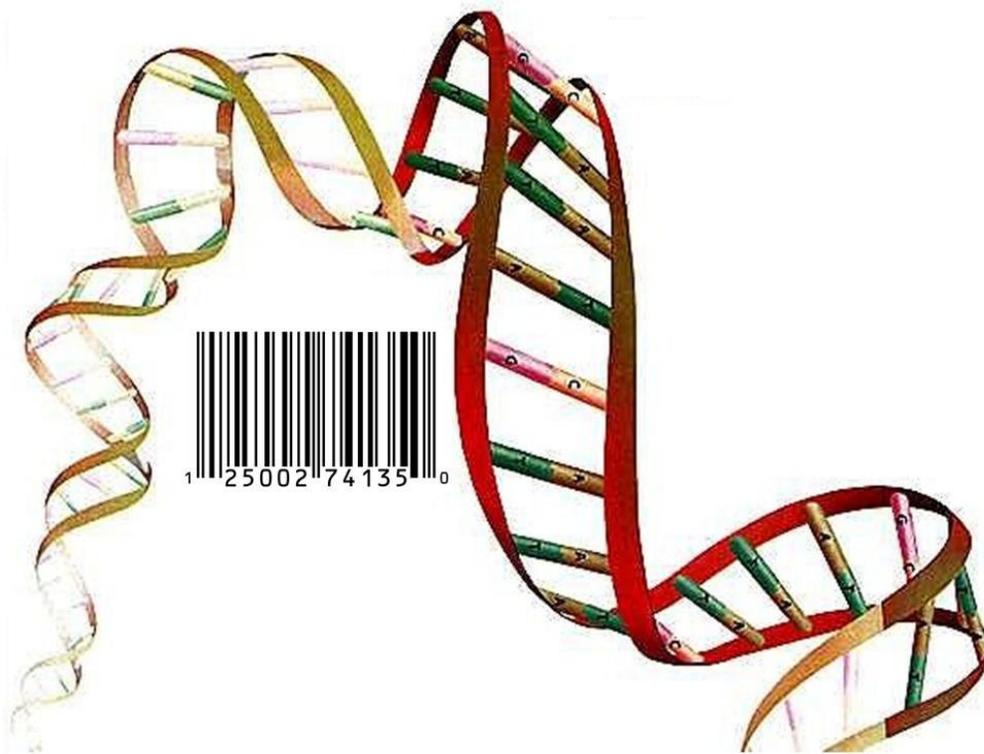




UNIVERSIDADE ESTADUAL DE MARINGÁ
Programa de Pós-Graduação em Ciências Biológicas
Laboratório de Organização Funcional do Núcleo - LORF

Variabilidade genômica: a utilização de DNA barcodes



Prof.^a Dr.^a Maria Aparecida Fernandez
Dr.^a Roxelle Ethienne Ferreira Munhoz
Dr.^a Thaís Souto Bignotto
Doutoranda Cláudia Regina Saez
Doutoranda Verônica Fassina
Naiara Climas Pereira

MARINGÁ
2013

Introdução

A **Taxonomia**, ciência de classificação dos seres vivos, estabelece critérios para classificar todos os animais e plantas em grupos de acordo com suas características fisiológicas, evolutivas, anatômicas e ecológicas. Carl Linnaeus formalizou a classificação biológica com seu sistema de nomenclatura binomial que designa cada organismo a um nome de **gênero** e **espécie**. A taxonomia de Linnaeus é considerada um **sistema natural de classificação** para organismos vivos uma vez que todos os seres vivos são descendentes de um único ancestral comum e foram se diferenciando um do outro por meio de uma série de eventos de especiação, num processo conhecido por evolução. Consequentemente, alguns grupos de espécies são mais relacionados do que outros e podem ser classificados dentro do mesmo gênero, família, ordem, etc. O estudo das relações evolutivas (ou seja, relações filogenéticas) de um grupo de organismos é denominado **filogenia**.

Estima-se que a biodiversidade total da Terra seja de 10 milhões de espécies. Desde o início da taxonomia, entretanto, apenas cerca de 15% da diversidade de espécies de animais e vegetais foram descritas (Figura 1), constatação que ficou conhecida como **impedimento taxonômico** (Solé-Cava, 2008). A identificação de espécies por meio desta abordagem, para ser empregada como rotina, apresenta algumas limitações significativas:

1. Tanto a plasticidade fenotípica quanto a variabilidade genética nos caracteres empregados para o reconhecimento de espécies podem levar a uma identificação incorreta da espécie,
2. Omissão de taxa crípticos morfologicamente, os quais são comuns em vários grupos,
3. Chaves morfológicas freqüentemente são eficientes somente para um estágio do ciclo de vida e assim muitos indivíduos não podem ser identificados,
4. O uso de chaves de identificação muitas vezes demanda a necessidade de recorrer a especialistas para não comprometer a identificação e
5. Falta de estímulos ao trabalho dos taxonomistas.

A limitação herdada no sistema de identificação baseada na morfologia e a escassez de taxonomistas especializados resultaram na baixa taxa de identificação de espécies animais e vegetais até o momento. Dessa maneira, o ritmo de descoberta e descrição de espécies é lento diante da velocidade da taxa de extinção. Portanto, é provável que muitas desapareçam sem terem sido conhecidas.

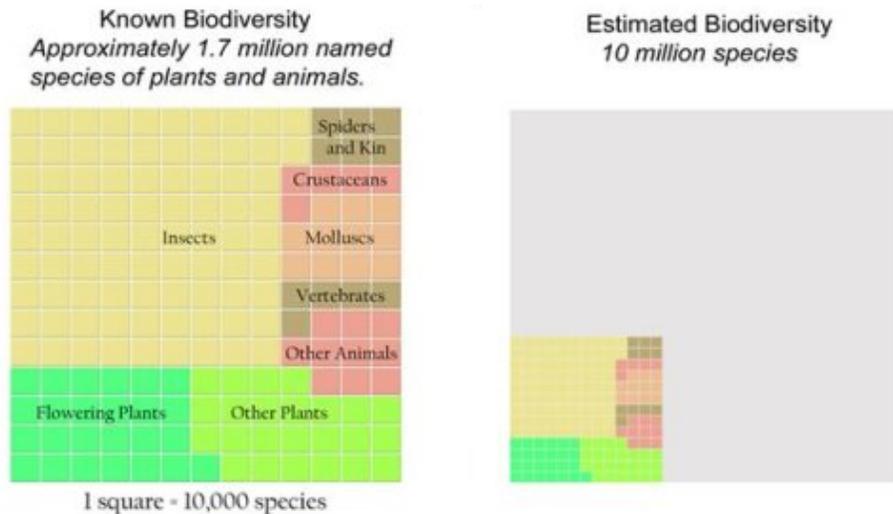


Figura 1. Biodiversidade conhecida e estimada.

A fim de tentar resolver o problema do impedimento taxonômico, Hebert et al., em 2003, propuseram o programa *Consortium for the Barcodes of Life (CBoL)*, que pretende padronizar e automatizar a identificação e descrição de toda biodiversidade apenas com uma sequência de DNA, utilizada de forma semelhante aos códigos de barras, sendo por isso denominada de **DNA barcode**. Para identificar corretamente e rapidamente as espécies de animais, foram selecionados cerca de 700 pares de bases (pb) do gene mitocondrial **citocromo c oxidase I, COI**, cuja sequência completa tem aproximadamente 1.545 pb (Figura 2). Para plantas, os marcadores do cloroplasto, *rbcL* (*rubisco large subunit*) e *matK*, têm sido delimitados como região padrão para código de barras.

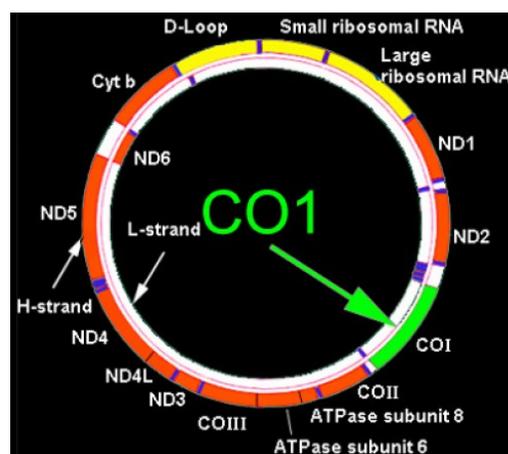


Figura 2. DNA mitocondrial, evidenciando a região do gene citocromo c oxidase I, COI.

DNA barcode

O DNA *barcode* é uma sequência curta de DNA, facilmente identificada, e característica de cada espécie. A região do gene que está sendo usada por quase todos os grupos de animais é de 648 pb do **gene mitocondrial citocromo c oxidase I, COI** (Figura 2), e está se mostrando altamente eficaz na identificação de aves, insetos, peixes, e outros grupos de animais. A vantagem de usar COI é que a região é curta o suficiente para ser sequenciada de forma rápida e barata e ainda longa o suficiente para identificar variações entre as espécies.

De acordo com o *International Barcode of Life Project, iBOL* (<http://ibol.org/about-us/what-is-dna-barcoding/>), DNA *barcodes* foi criado para chamar a atenção da comunidade científica, quando em 2003 o grupo de pesquisa de Paul Hebert da Universidade de Guelph publicou o artigo científico “*Biological identifications through DNA barcodes*”. Neste artigo, eles propuseram um novo sistema de identificação e descobrimento de espécies utilizando um pequeno trecho de DNA de uma região padronizada do genoma. Essa sequência de DNA pode ser usada para identificar diferentes espécies, da mesma maneira que um scanner de supermercado usa as listras pretas do código de barras UPC (*Universal Product Code*) para identificar suas compras (Figura 3). O código de barras emprega 10 números alternados em 11 posições para gerar 100 bilhões de identificadores únicos. No caso do DNA *barcodes*, pode haver até quatro possibilidades de nucleotídeos (adenina, timina, citosina e guanina) em cada posição, mas com uma cadeia de sítios muito mais longa que 11 posições. Por exemplo, a combinação de apenas 15 dessas posições criaria um bilhão de códigos únicos, um número muito maior do que o de espécies conhecidas (Hebert et al., 2003).

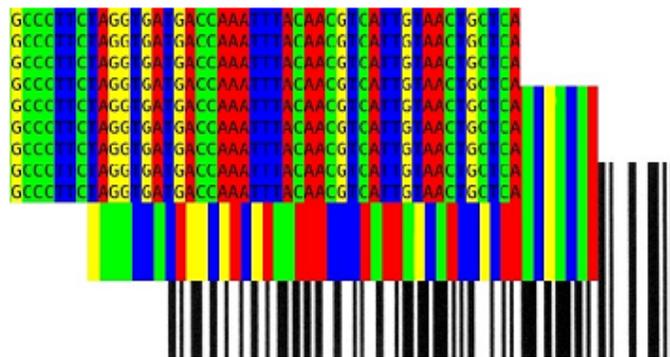


Figura 3. Analogia entre o DNA *barcodes* e o código de barras convencional.

Para armazenar essa enorme quantidade de sequências do DNA *barcodes* para todas as espécies existentes, foi criado o **BOLD**, *The DNA Barcode of Life Data System* (<http://www.boldsystems.org/>). Nesse banco de dados, estão incluídas outras informações além da sequência do DNA *barcodes*, como: fotos do espécime (*voucher*), local de coleta, dados taxonômicos e informações moleculares (eletroferogramas e *primers* utilizados).

A identificação das espécies, no BOLD, é realizada por meio de comparações de uma sequência de interesse com sequências previamente depositadas no banco. Essa comparação é feita por métodos de distância genética, segundo o modelo padrão de substituição molecular **Kimura-2-parâmetros, K2P** (Kimura, 1980), sugerido por Hebert et al. (2003), uma vez que essa medida de distância é a mais eficaz quando distâncias genéticas são baixas. Em seguida, uma árvore *neighbor-joining* (NJ) é construída com as 100 sequências mais similares à sequência de interesse. O procedimento NJ foi considerado como método padrão de inferências filogenéticas nos estudos de DNA *barcodes* (Hebert et al., 2003) devido sua alta eficiência e rapidez em análises de espécies em larga escala (Kumar e Gadagkar, 2000). Finalmente, a espécie é atribuída àquela cuja sequência *barcodes* se mostrou mais similar. De acordo com dados publicados, a identificação de espécies de insetos, aves e mamíferos apenas pode ser aceita se houver similaridade com alguma sequência armazenada no BOLD maior do que 97%, 98% e 98%, respectivamente. Em outras palavras, caso a diferença seja maior que 3% ou 2%, não é possível identificar a amostra e seria necessário realizar análises mais detalhadas (Rubinoff, 2006).

Um critério adotado pelo BOLD para a delimitação de espécies se baseia em **valores limites para as divergências nucleotídicas intra e interespecíficas**. Um dos limites é de **3%** (valor estabelecido para insetos), em que valores de divergência intra-específica abaixo deste limite determinam uma única espécie e valores de divergência interespecíficas acima, apontam para diferentes espécies. O outro limite, que surge como uma atualização do primeiro, sugere que a média da divergência nucleotídica entre espécies pertencentes ao mesmo gênero deve ser **10 vezes** superior à média da divergência intra-específica encontrada para as mesmas espécies. A observação destes critérios permite, assim, determinar se estamos perante a mesma espécie ou espécies diferentes

Características da região *barcodes*:

- Flanqueada por sequências nucleotídicas semelhantes em todas as espécies,
- Fácil isolamento e seqüenciamento,

- Baixa variabilidade intra-específica (todos os indivíduos de uma espécie apresentam sequências idênticas/ semelhantes de DNA),
- Elevados níveis de variação ENTRE espécies,
- Pequena o suficiente para ser sequenciada rápida e economicamente.

Aplicações do DNA *barcodes*

- Ferramenta para **identificação de espécies**:
 - vetores de doenças, pestes agrícolas, espécies invasoras (organismos perigoso)
 - indicadores ambientais, espécies protegidas (espécies ameaçadas)
 - usando exemplares pequenos, danificados, conteúdos estomacais, amostras fecais (identificação de espécie derivada de amostras não-reconhecíveis)
- Ferramenta para complementar **estudos filogenéticos**:
 - Identificação taxonômica em nível de espécies
- Ferramenta para **melhorar a taxonomia** em nível específico:
 - associação de todos os estágios da história de vida, gêneros (determinação da espécie que deu origem a alguma larva específica; inclusão ou não de espécimes similares morfológicamente em uma mesma espécie; identificação de organismos dimórficos)
- Ferramenta de triagem para sinalizar potenciais **novas espécies**:
 - espécies crípticas
 - espécies ainda não descritas
- Ferramenta para **conservação, biossegurança e biopirataria**:
 - preservação da biodiversidade global
 - identificação de espécies invasoras ou introduzidas
 - detecção de fraudes alimentares

Limitações/ Desvantagens do DNA *barcodes*

Apesar de ser uma abordagem que promete identificar rápida e eficientemente todas as espécies, o DNA *barcodes* apresenta algumas limitações que devem ser consideradas. A primeira delas baseia-se no fato de ser uma metodologia que utiliza apenas um marcador molecular para classificar e identificar os organismos, deixando de lado a taxonomia

tradicional. Alguns pesquisadores alertam para o fato de que o gene COI pode não ser informativo o suficiente para descrever toda a diversidade biológica. Isso ocorre porque o DNA mitocondrial apresenta algumas características que restringem a possibilidade de realizar inferências sobre o limite entre as espécies, como a herança materna, retenção de polimorfismo ancestral (subestimação do número de espécies), heteroplasmia e cópias nucleares do gene COI (superestimação do número de espécies). Em alguns casos, portanto, seria necessário incluir o estudo de um marcador nuclear.

Outra limitação da técnica recai em alguns grupos animais que apresentam baixa diversidade na sequência de DNA, ou seja, mais de uma espécie compartilha a mesma sequência nucleotídica, provavelmente devido a divergência entre as espécies ter ocorrido a relativamente pouco tempo. No entanto, diversos estudos realizados sugeriram que o *barcodes* é capaz de identificar inequivocamente a maioria das espécies animais. Apenas aproximadamente 2-5% das espécies reconhecidas apresentam sequências *barcodes* compartilhadas ou sobrepostas com outras espécies proximamente relacionadas.

O principal pressuposto para a efetividade do DNA *barcodes* é de que as divergências intra-específicas sempre sejam menores que as interespecíficas, o que se chama de *barcoding gap*. Existem casos, no entanto, em que pode ocorrer sobreposição nos níveis de variação intra e interespecíficos de alguns táxons, o que atrapalha a identificação de um organismo pertencente a esse grupo (Figura 4).

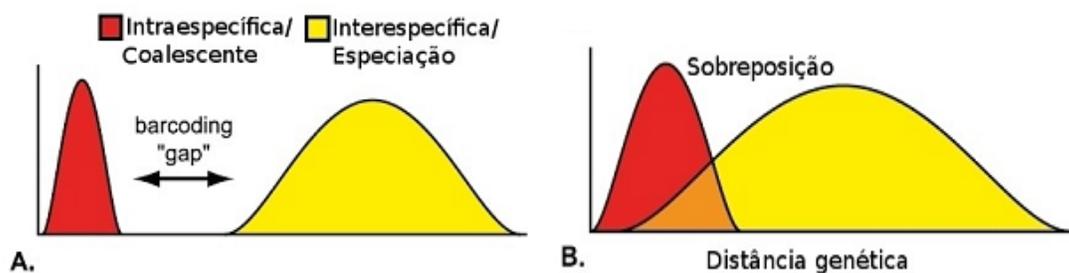


Figura 4. Esquema do *barcoding gap*. A distribuição da variação intra-específica é mostrada em vermelho e da interespecífica, em amarelo. (A) Mundo ideal para DNA *barcodes*, com distribuições discretas e sem sobreposição. (B) Uma versão alternativa com significativa sobreposição, sem *gap*. (Fonte: Meyer e Paulay, 2005).

DNA *barcoding* compreende uma região pequena e altamente variável do genoma mitocondrial, denominada de COI ou citocromo c oxidase I. Com milhares de cópias por célula, essa sequência mitocondrial pode ser rapidamente amplificada por **PCR** (*Polymerase Chain Reaction*, Reação em Cadeia da Polimerase) até mesmo de espécimes muito pequenos ou degradados. Primeiramente, uma amostra do tecido animal deve ser coletada, tomando o cuidado para preservar o espécime sempre que possível e anotando a localização geográfica do local de sua coleta. Um inseto inteiro ou uma amostra de músculo são fontes adequadas. Em seguida, o DNA é extraído da amostra de tecido do animal, e a porção *barcodes* do gene COI é amplificada por PCR com *primers* específicos. A sequência amplificada é então sequenciada em sequenciador automático em uma ou em ambas as direções. As sequências resultantes são utilizadas em busca em bancos de dados de DNA. Uma estreita correspondência entre duas sequências identifica rapidamente uma espécie que já está representada no banco de dados (Figura 5).

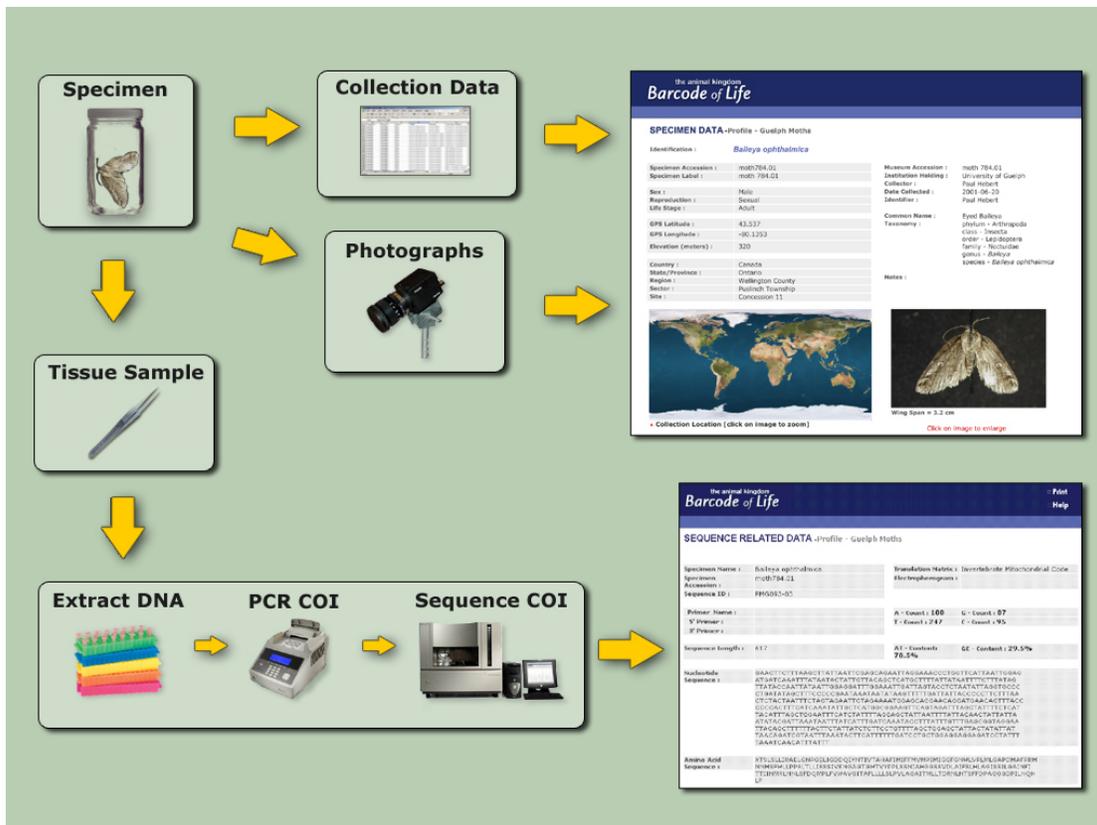


Figura 5. Procedimentos de análise do DNA *barcodes*.

A seguir, serão apresentados os métodos e protocolos utilizados laboratorialmente na obtenção e análise de sequências *barcodes*.

Metodologia experimental



1. Extração de DNA de insetos (*Bombyx mori*)

1. Identificar no tubo a espécie/raça e sua numeração.
2. Macerar a mariposa em nitrogênio líquido (marca de 250 μ L – tubos de 2 mL);
3. Adicionar de 500 μ L a 750 μ L de **tampão de extração**;
4. Centrifugar por 10 minutos a 10.000 rpm a 4°C;
5. Transferir o sobrenadante (~ 750 μ L) para novos tubos;
6. Adicionar ao sobrenadante 750 μ L de **tampão para proteinase K** e 1,5 μ L de **proteinase K**;
7. Incubar as amostras a 37°C por 3 horas ou a 60°C por 1 hora (com agitação);
8. Separar as amostras em dois tubos;
9. Acrescentar a cada tubo 750 μ L de **clorofórmio** (ou o mesmo volume);
10. Centrifugar por 30 minutos a 14.000 rpm a 4°C;
11. Transferir o sobrenadante para novo tubo e adicionar 350 μ L de **fenol** e 350 μ L de **clorofórmio** (1 fenol:1 clorofórmio);
12. Centrifugar por 10 minutos a 14.000 rpm a 4°C;
13. Transferir o sobrenadante para novo tubo (1,5 mL) para precipitação;
14. Para precipitar: usar NaCl 5 M ([final] = 0,2 M) e **etanol** 100% (2x volume) ou **isopropanol** (0,7x volume).
Obs: Nesta fase, se tiver pouco sobrenadante, usar etanol no dobro do volume. Se tiver mais que 500 μ L de sobrenadante, usar isopropanol
15. Fechar os tubos presos em um barbante e mergulhá-los em nitrogênio líquido durante 30 segundos;
16. Centrifugar por 30 minutos a 14.000 rpm a 4°C;
17. Descartar o sobrenadante;
18. Lavar o pellet com **etanol** 70% ou 80% (\approx 70 a 100 μ L);
19. Colocar os tubos para secar;
20. Ressuspender o pellet em TE 1X + RNase;

Purificação do DNA

Obs: completar o volume com TE para 200 μ L antes de iniciar a purificação.

21. Adicionar 1 fenol :1 clorofórmio: 100 µL de **fenol** e 100 µL de **clorofórmio**;
22. Centrifugar por 10 minutos a 14.000 rpm a 4°C;
23. Transferir o sobrenadante para novo tubo e adicionar **clorofórmio** no mesmo volume (200 µL);
24. Centrifugar por 10 minutos a 14.000 rpm a 4°C;
25. Transferir o sobrenadante para novo tubo e precipitar com NaCl 5 M (14 µL) e **etanol** 100% (2x volume) ou ISOPROPANOL (0,7x volume);
26. Mergulhar as amostras em nitrogênio líquido durante 30 segundos;
27. Centrifugar por 30 minutos a 14.000 rpm a 4°C;
28. Descartar o sobrenadante;
29. Lavar com etanol 70% e deixar secar;
30. Ressuspender o pellet com TE 1X (autoclavado 2x de 45' e filtrado);
31. Manter as amostras em freezer.

A. Tampão de extração

Soluções	[inicial]	[final]
200 µL de Tris-HCl pH 7,5	1 M	10 mM
240 µL de NaCl	5 M	60mM
400 µL de EDTA pH 8,0	0,5 M	10 mM
6.250 µL de sacarose	16%	5%
24 µL de espermidina	125mM	0,15mM
60 µL de espermina	50mM	0,15mM
Completar volume com água mili-Q para 20 mL		

B. Tampão para proteinase K

Soluções	[inicial]	[final]
4.000 µL de Tris-HCl pH 9,0	1 M	0,2 M
1.200 µL de EDTA pH 8,0	0,5 M	30 mM
2.000 µL de SDS	20%	2%
6.250 µL de sacarose	16%	5%
Completar volume com água mili-Q para 20 mL		

2. Amplificação do DNA genômico por PCR (Reação em cadeia da polimerase)

Introdução

Em 1985, foi descrita a metodologia da reação em cadeia da polimerase (PCR, de *polymerase chain reaction*). Seu idealizador, Kary Mullis, recebeu, por isso, o Prêmio Nobel de Medicina e Fisiologia em 1993. A PCR permite a produção de grandes quantidades de um determinado segmento de DNA *in vitro* a partir de uma pequena quantidade de DNA-molde, evitando assim a necessidade de introdução (clonagem) do DNA de interesse em bactérias.

Objetivo: amplificar a região *barcodes* do gene mitocondrial COI, por meio de PCR.

Material:

- Oligonucleotídeos (*primers*) específicos (5 mM)
- DNA-molde (DNA de *Bombyx mori*) contendo o alvo
- dNTPs (desoxirribonucleotídeos): dATP, dCTP, dGTP, dTTP) (1 mM)
- Tampão para *Taq* DNA-polimerase (10X)
- *Taq* DNA-polimerase (1 U/ μ l)
- MgCl₂ (50 mM)

Os iniciadores ou *primers* são oligonucleotídeos complementares a seqüências do DNA. Nesse caso, serão utilizados *primers* descritos por Hebert et al. (2004) que flanqueiam a região padrão do *barcodes* (Figura 1). O segmento de DNA amplificado inclui a seqüência nucleotídica situada entre dois *primers* de aproximadamente 650 pb.

LepF1: 5'-ATTCAACCAATCATAAAGATATTGG-3'

LepR1: 5'-TAAACTTCTGGATGTCCAAAAAATCA-3'

A temperatura de anelamento dos iniciadores ao DNA-molde na PCR é definida com base na T_m. Geralmente, ela deve ser de cerca de 20 a 25°C abaixo da T_m dos iniciadores.

Estimativa da Tm de iniciadores

A seguinte fórmula pode ser utilizada para estimar a temperatura média de fusão (Tm) de um iniciador: $T_m (^{\circ}\text{C}) = 4 \times (\text{G} + \text{C}) + 2 \times (\text{A} + \text{T})$

Procedimentos:

1. Pipetar as seguintes soluções em um tubo de centrífuga de 0,5 mL:

Quantidades para uma reação:

Água miliQ	8,25 μL
Tampão (10X)*	1,5 μL
dNTPs (1 mM)	0,5 μL
Primer LepF1 (5 mM)	0,5 μL
Primer LepR1 (5 mM)	0,5 μL
MgCl ₂ (50 mM)	0,75 μL
Taq DNA-polimerase (1 U/ μl)	1,5 μL
<u>DNA molde (60 ng/μl)</u>	<u>1,5 μL</u>
Volume final	15 μL

* 20 mM Tris-HCl (pH 8.0), 0.1 mM EDTA, 1 mM DTT, 50% glicerol, estabilizantes, da marca Invitrogen.

Nota: Uma unidade enzimática da DNA-polimerase é definida como a quantidade de enzima necessária para catalisar a incorporação de 10 nmol de dNTP durante 30 min à 74°C.

2. Colocar a reação em um termociclador com a seguinte programação de ciclos de temperatura:

Passo	Temperatura	Tempo	
1	94°C	1 min	Temperatura Inicial/Denaturação do DNA
2	94°C	40 seg	Denaturação
3	60°C	40 seg	Temperatura de Anelamento dos <i>primers</i>
4	72°C	1 min	Extensão ou Amplificação do DNA
5	Voltar para o passo 2 - 35 X		
6	72°C	30 min	Extensão Final
7	4°C	indefinidamente	Armazenamento

3. Preparar 5 μL do produto da reação de PCR para eletroforese em gel de agarose.
4. Submeter as amostra a eletroforese em gel de agarose 1,5% .

3. Eletroforese em gel de agarose

Objetivo: preparar gel de agarose para análise de ácidos nucleicos previamente amplificados por PCR.

Materiais:

- *TBE 5X (Composição para 1 litro)*

54 g de tris-HCl

27,5 g de ácido bórico

20 mL de EDTA 0,5 M pH 8,0

- *Gel de agarose 1,5%, 100 mL*

100 mL de TBE 1X

1,5 g agarose

- *Brometo de etídeo (1 mg/ml)*

Usar 1 μL para cada 10 mL de gel.

Cuidado! O brometo de etídeo é mutagênico e cancerígeno. Não tocar nos géis corados com brometo de etídeo sem a devida proteção. Use sempre luvas para manusear o frasco e o gel com brometo de etídeo.

- *Transiluminador ultravioleta e máscaras de proteção*

Procedimentos

- Preparação do gel:

Preparar a solução de agarose em frasco de Erlenmeyer e aquecê-la (até a fervura) em forno de microondas até a sua completa solubilização. O frasco deve ser agitado freqüentemente para homogeneização e para evitar transbordamento.

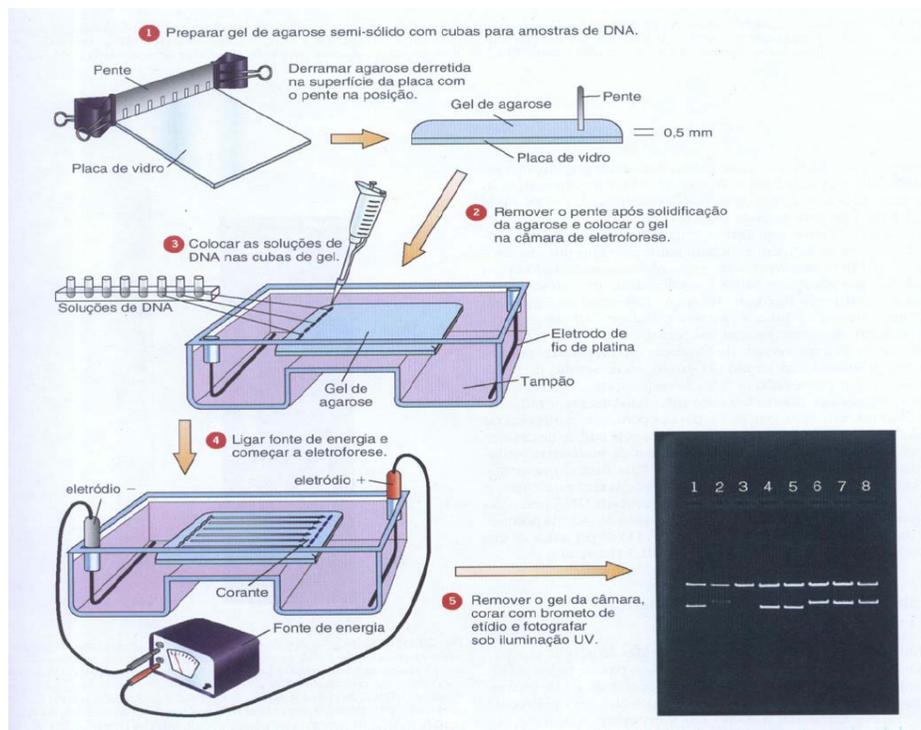
Após a solubilização, a solução de agarose deve ser resfriada por alguns minutos e a ela devem ser adicionados solução de brometo de etídio (1 mg/ml). Depois disso, agitar a solução e vertê-la sobre uma placa-molde de gel com o pente já montado. Aguardar a polimerização a temperatura ambiente.

• Preparação das amostras:

Em parafilme misturar 5 µl de cada amostra amplificada de DNA mais 3 µl de loading buffer 6X.

• Eletroforese:

1. Colocar o gel de agarose na cuba de eletroforese e submergi-lo em TBE 1X.
2. Aplicar as amostras no gel com uma micropipeta
4. Submeter as amostras a eletroforese a uma voltagem de 100V, por aproximadamente 60 min.
5. Visualizar as amostras de DNA resolvidas eletroforéticamente colocando o gel sobre transluminador com iluminação ultravioleta (Cuidado! O ultravioleta pode causar queimaduras e é mutagênico. Não visualizar os géis sem as devidas precauções de segurança).



4. Clonagem e transformação

Introdução

A clonagem molecular é o processo de construção de moléculas de DNA recombinantes e da sua propagação em hospedeiros apropriados que possibilitam a seleção do DNA recombinante. A clonagem molecular permite isolar e amplificar uma sequência particular de DNA a partir de uma mistura complexa de fragmentos, porque se formam moléculas de DNA recombinante independentes que replicam numerosas vezes no hospedeiro.

Objetivo: aumentar a quantidade de DNA amplificado por PCR, no caso uma região do gene COI, para posterior seqüenciamento.

Materiais:

- *Meio LB sólido, 500 mL:*

- 5 g NaCl;
- 2,5 g Extrato de levedura;
- 5 g Triptona-bacto;
- 7,5 g Agar Select.

Diluir em 400mL de água destilada com agitador em um becker. Completar volume para 500mL com água destilada.

- *Meio LB líquido, 250mL:*

- 2,5 g NaCl;
- 1,25 g Extrato de levedura;
- 2,5 g Triptona-bacto.

- *Meio SEC:*

- 1 mL de meio LB líquido
- 10 µL de MgCl₂ 1M
- 10 µL de MgSO₄ 1M
- 100 µL de glicose 1M

- *Meio LB sólido para placa:*

- 40 mL de meio LB sólido
- 400 µL de ampicilina 5 mg/mL
- 80 µL de X-gal 20 mg/mL
- 20 µL de IPTG 20 mg/mL

Verter 20 mL em duas placas de Petri autoclavadas.

Procedimentos:

1. Ligação do produto de PCR ao vetor de clonagem pGEM:

* Kit: pGEM®-T Easy Vector Systems, Promega.

5 µL de tampão*

1 µL de ligase*

1 µL de vetor pGEM*

2µL do produto PCR

1µL de água miliQ

2. Transformação por choque térmico

- Descongelar bactérias competentes no gelo por 10 minutos;
- Colocar em um tubo 5µL da ligação em 50µL das bactérias. Misturar 2x com o pipetador;
- Manter no gelo por 30 minutos (nesse tempo fazer o meio SEC)
- Efetuar choque térmico: 2 minutos a 42°C e em seguida 2 minutos no gelo;
- Adicionar 450µL de meio SEC;
- Incubar durante 1 hora a 37°C sob leve agitação (70 rpm) (Nesse tempo derreter o LB sólido e preparar o meio sólido para placa);
- Plaquear em placa de Petri contendo meio sólido e incubar a 37°C durante aproximadamente 16 horas. Colocar 100 µL de cultura em cada placa contendo 20mL de meio sólido;
- Coletar colônias brancas e transferi-las para tubos individuais contendo 5 mL de LB líquido e 50 µL de ampicilina;
- Incubar a 37°C por 70 rpm durante aproximadamente 16 horas.

3. Preparação do DNA plasmidial por STET e CTAB

- Fazer conserva e estocar no freezer -80°C

Conserva: 850 µL da cultura + 150 µL de glicerol

- Centrifugar 1,5 mL da cultura por 2 min a 8.000 rpm. Descartar o sobrenadante. Repetir o processo adicionando mais 1,5 mL da cultura no mesmo tubo;
- Ressuspender em 300 µL de STET no vórtex;
- Adicionar 6 µL de lisozima (50 mg/mL) e incubar por 5 min a temperatura ambiente;
- Manter as amostras por 1 ½ min em 100°C;
- Centrifugar 10 min a 14.000 rpm;
- Transferir o sobrenadante para novo tubo;

- Adicionar 4 μL de CTAB 5%. Misturar gentilmente até que a solução fique turva. Obs: deixar o CTAB 5% no banho-maria a 37°C por aproximadamente 10 min para eliminar os cristais;
- Centrifugar 5 min a 14.000 rpm;
- Descartar o sobrenadante e deixar secar
- Ressuspender em 300 μL de NaCl 1,2M no vórtex e aquecer a 37°C por 1 min;
- Precipitar o DNA com 750 μL de etanol absoluto;
- Centrifugar por 10 min a 14.000 rpm e descartar o sobrenadante;
- Lavar com 100 μL de etanol 70% gelado;
- Secar em temperatura ambiente
- Ressuspender em 100 μL de TE com RNase;
- Deixar por no mínimo 1 hora ou no máximo 2 horas;
- Aplicar as amostras em gel de agarose 0,7%.

5. Amplificação do DNA clonado por PCR

Procedimentos:

1. Amplificação por PCR:

Quantidades para uma reação:

Água miliQ	17,65 μL
Tampão (10X)*	2,5 μL
dNTPs (1 mM)	0,8 μL
<i>Primer M13F</i> (10 mM)	0,8 μL
<i>Primer M13R</i> (10 mM)	0,8 μL
MgCl ₂ (50 mM)	1,25 μL
<i>Taq</i> DNA-polimerase (1 U/ μl)	0,2 μL
<u>DNA molde (60 ng/μl)</u>	<u>1,0 μL</u>
Volume final	25 μL

2. Termociclador:

Passo	Temperatura	Tempo
1	94°C	4 min
2	94°C	30 seg
3	55°C	30 seg
4	72°C	1 min
5	Voltar para o passo 2 - 25 X	
6	72°C	10 min
7	4°C	indefinidamente

6. Purificação do produto de PCR por PEG (Polietilenoglicol)

Objetivo: retirar impurezas, como excesso de primers e dNTPs, da reação de PCR para posterior uso em sequenciamento de DNA.

Materiais:

• *PEG, 50 mL:*

- 10 g de polietilenoglicol 8.000 (MW=6.000 a 8.000) (20% PEG)
- 7,3 g de NaCl (NaCl 2,5 M)
- Água Milli-Q para 45 mL

Misturar e deixar o PEG entrar na solução. PEG leva 20 min para entrar na solução.

Após os 20 min, complete com água para 50 mL.

Filtrar com filtros 0,2 µm

• *Etanol 80% (Merck)*

Procedimentos:

(Obs: considerando um volume de 25 µL de produto de PCR)

1. Verifique a qualidade do produto de PCR em gel de agarose;
2. Adicione 25 µL de solução de PEG (PEG 20% NaCl 2,5 M) em um tubo de microcentrífuga de 1,5 mL;

3. Transfira o produto de PCR para os tubos contendo PEG e misture muito bem com o pipetador;
 4. Incube a 37°C por 15 minutos.
 5. Centrifugue a 12.000 rpm por 20 minutos a temperatura ambiente. Coloque os tubos fechados com a alça da tampa voltada para a parte externa do rotor (assim você saberá onde sedimentará o *pellet* já que não será possível vê-lo) ;
 6. Retire o sobrenadante SUAVEMENTE com o auxílio de um micropipetador P200 e descarte-o;
 7. Adicione 40 µL de etanol (Merck) 80% gelado. Centrifugue por 2 minutos a 12.000 rpm;
 8. Retire o sobrenadante SUAVEMENTE com o auxílio de um micropipetador P200 e descarte-o;
 9. Repita os passos 7 e 8.
 10. Deixe evaporar o etanol residual utilizando o banho seco, estufa ou deixe sobre a bancada. Não deve existir nenhum traço (visível ou cheiro) de etanol;
 11. Ressuspenda o *pellet* entre 5 e 15 µl* de água Milli-Q. Pipete várias vezes!!!
- *O volume de água dependerá da concentração inicial do PCR;
12. Quantifique o produto purificado em gel de agarose (1-2µl da amostra).

7. Sequenciamento

Introdução

Seqüenciamento de fragmentos de genes permite a avaliação direta de polimorfismos de DNA, fornecendo o melhor tipo de dado para inferências filogenéticas e determinação de relações de parentesco entre indivíduos e populações. Basicamente, dois métodos têm sido utilizados para determinar a seqüência de nucleotídeos de DNA: 1) o método de degradação química de DNA, desenvolvido por Maxam e Gilbert e 2) o método enzimático de Sanger ou de terminação da cadeia com didesoxirribonucleosídeo trifosfato. Embora o primeiro método tenha tido ampla aplicação inicialmente, o segundo tem sido o mais utilizado atualmente.

Ambos os procedimentos de sequenciamento de DNA dependem da produção de uma população de fragmentos de DNA que tenham uma ponta em comum (todos começam

exatamente no mesmo nucleotídeo) e terminem em todas as posições possíveis (cada nucleotídeo consecutivo) na outra ponta. Esses fragmentos são então separados com base no tamanho por eletroforese em gel. Nos dois casos, quatro reações bioquímicas separadas são feitas simultaneamente, cada uma gerando um conjunto de fragmentos que terminam em uma das quatro bases (A, G, C ou T) no DNA.

Método de Maxam e Gilbert ou Método Químico

Criado por Allan Maxam e Walter Gilbert, esse método envolve a degradação química do DNA, marcado em uma extremidade, de forma que uma população de fragmentos é gerada e analisada em gel de poliacrilamida.

As pontas 5' do DNA são marcadas com ^{32}P , o DNA é clivado, os fragmentos são isolados pelo tamanho, e então o duplex de DNA é separado, com um filamento específico sendo descartado, para produzir uma população de filamentos idênticos marcados em uma ponta. A mistura é então dividida em quatro amostras, sendo cada uma delas submetida a um reagente químico diferente que destrói uma ou duas bases específicas. Os quatro reagentes destroem (1) só G, (2) A e G, (3) T e C, ou (4) apenas C. A perda de uma base torna a sequência açúcar-fosfato da cadeia de DNA mais sensível a quebrar nesse ponto. As condições dos tratamentos químicos para modificações das bases e clivagem são feitos das seguintes formas:

(1) Para G, um tratamento com dimetil sulfato metila o nitrogênio 7 da guanina, que se abre entre os carbonos 8 e 9. Um tratamento com piperidina remove a base modificada;

(2) Para G+A, um tratamento com ácido fórmico enfraquece as ligações glicosídicas. Protonando os nitrogênios dos anéis de purina. Esses são, então, removidos com piperidina;

(3) Para T+C, um tratamento com hidrazina cliva os anéis de timina e citosina. Um tratamento com piperidina remove as bases modificadas;

(4) Para C, hidrazina, na presença de NaCl, faz com que só a citosina seja modificada e possa ser deslocada por tratamento com piperidina

A concentração do reagente é ajustada, de modo que apenas cerca de 1 em 50 bases-alvo são destruídas. Isso significa que dentro de uma determinada amostra em alguns filamentos apenas um nucleotídeo será destruído. Para tais filamentos, os comprimentos dos fragmentos serão determinados pela distância da ponta 5' marcada até o nucleotídeo em particular que foi perdido na sequência. Como este é um evento aleatório em cada um dos

agentes destruidores de base, cada ocorrência dessa base em particular irá produzir um fragmento marcado de tamanho distinto. Quando esses pedaços são separados em colunas diferentes de um gel de poliacrilamida, de acordo com o tipo de reagente no qual estavam, eles são dispostos por ordem de comprimento, e a base destruída em cada sítio pode ser determinada notando-se em que coluna(s) a banda aparece. Portanto, a sequência de bases no filamento pode ser lida com facilidade pelo padrão de bandas no gel (Figura 1).

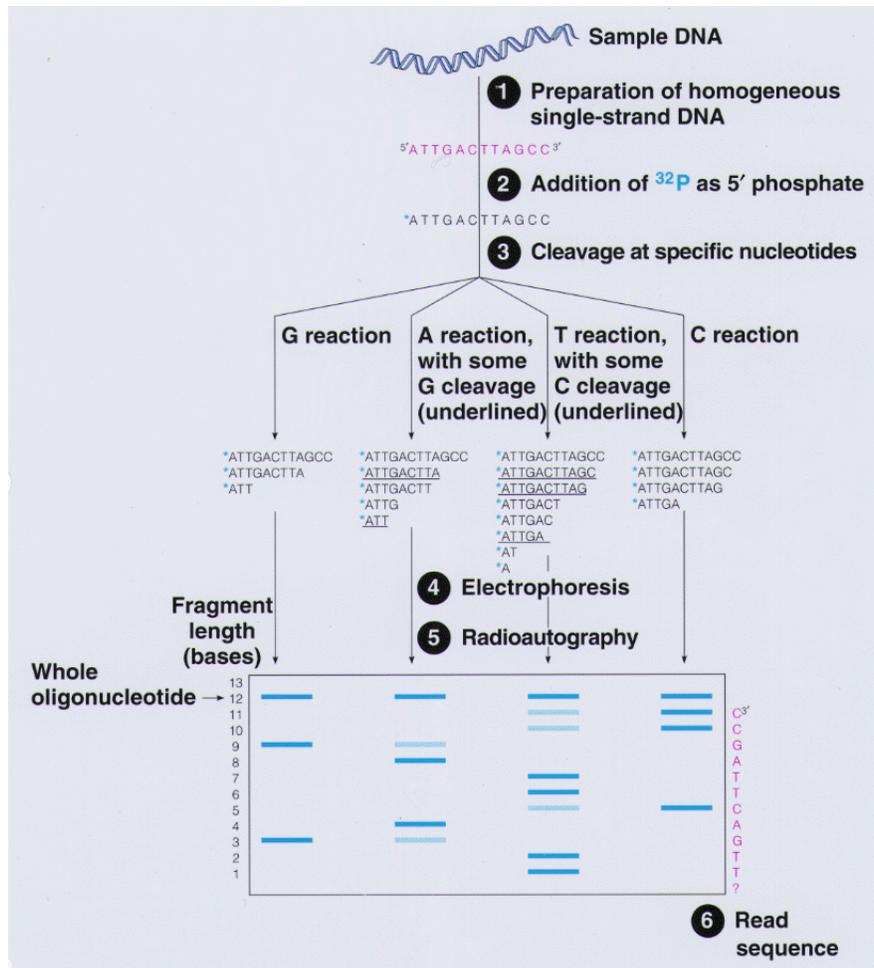


Figura 1. Sequenciamento pelo método de Maxam-Gilbert.

Método de Sanger ou Método Enzimático

O método de sequenciamento mais usado hoje em dia foi desenvolvido por Fred Sanger, e utiliza a síntese de DNA *in vitro* na presença de nucleotídeos radioativos e finalizadores específicos de cadeia para gerar quatro populações de fragmentos marcados

radioativamente que terminam em As, Gs, Cs e Ts. Os finalizadores de cadeia mais frequentemente usados no procedimento de seqüenciamento são os **didesoxirribonucleosídeo trifosfatos** (ddATP, ddGTP, ddCTP e ddTTP). Os didesoxirribonucleosídeo trifosfatos não apresentam a hidroxila no carbono 3' (Fig. 4). Os respectivos trifosfatos de nucleotídeos de didesoxi (ddNTPs) podem ser incorporados a uma cadeia em crescimento, mas terminam a síntese, pois não tem a 3'-hidroxila necessária para se ligar ao trifosfato do nucleotídeo seguinte. Lembre-se que as DNA polimerases necessitam de uma ponta 3'-OH livre para fazer o alongamento da nova fita de DNA. Se um didesoxirribonucleotídeo é adicionado à extremidade de uma cadeia, ele irá bloquear a extensão desta cadeia devido à falta de 3'-OH. Usando ddATP, ddGTP, ddCTP e ddTTP como finalizadores de cadeia em quatro reações separadas de síntese de DNA, podem ser geradas quatro populações de fragmentos, e cada população conterá cadeias que terminam na mesma base (A, G, C ou T)

Em uma reação, a proporção de dNTP:ddNTP é mantida em aproximadamente 100:1. Em qualquer um dos tubos, serão produzidos vários comprimentos de cadeia, cada um correspondendo ao ponto no qual o ddNTP respectivo do tubo foi incorporado, terminando o crescimento da cadeia.

O DNA a ser seqüenciado é hibridizado com um *primer*. A fita complementar é sintetizada na presença dos quatro desoxirribonucleosídeos normais (dATP, dGTP, dCTP e dTTP), mais um único tipo de didesoxirribonucleosídeo (ddATP, ddGTP, ddCTP ou ddTTP), o qual não possui a 3'-hidroxila, e de DNA polimerase. A nova fita de DNA sintetizada é marcada radioativamente ou por marcação da extremidade do *primer* ou por incorporação de um desoxirribonucleosídeo marcado durante a síntese. Quatro reações separadas são feitas, uma para cada didesoxirribonucleosídeo trifosfato.

Após os fragmentos de DNA gerados nas quatro reações paralelas serem liberados dos filamentos-molde por desnaturação na presença de formamida, eles podem ser separados por eletroforese em gel de poliacrilamida e suas posições no gel podem ser detectadas por auto-radiografia. Lendo a escada produzida pela auto-radiografia dos géis de poliacrilamida, pode-se determinar a seqüência completa de nucleotídeos de uma cadeia de DNA (Figura 2).

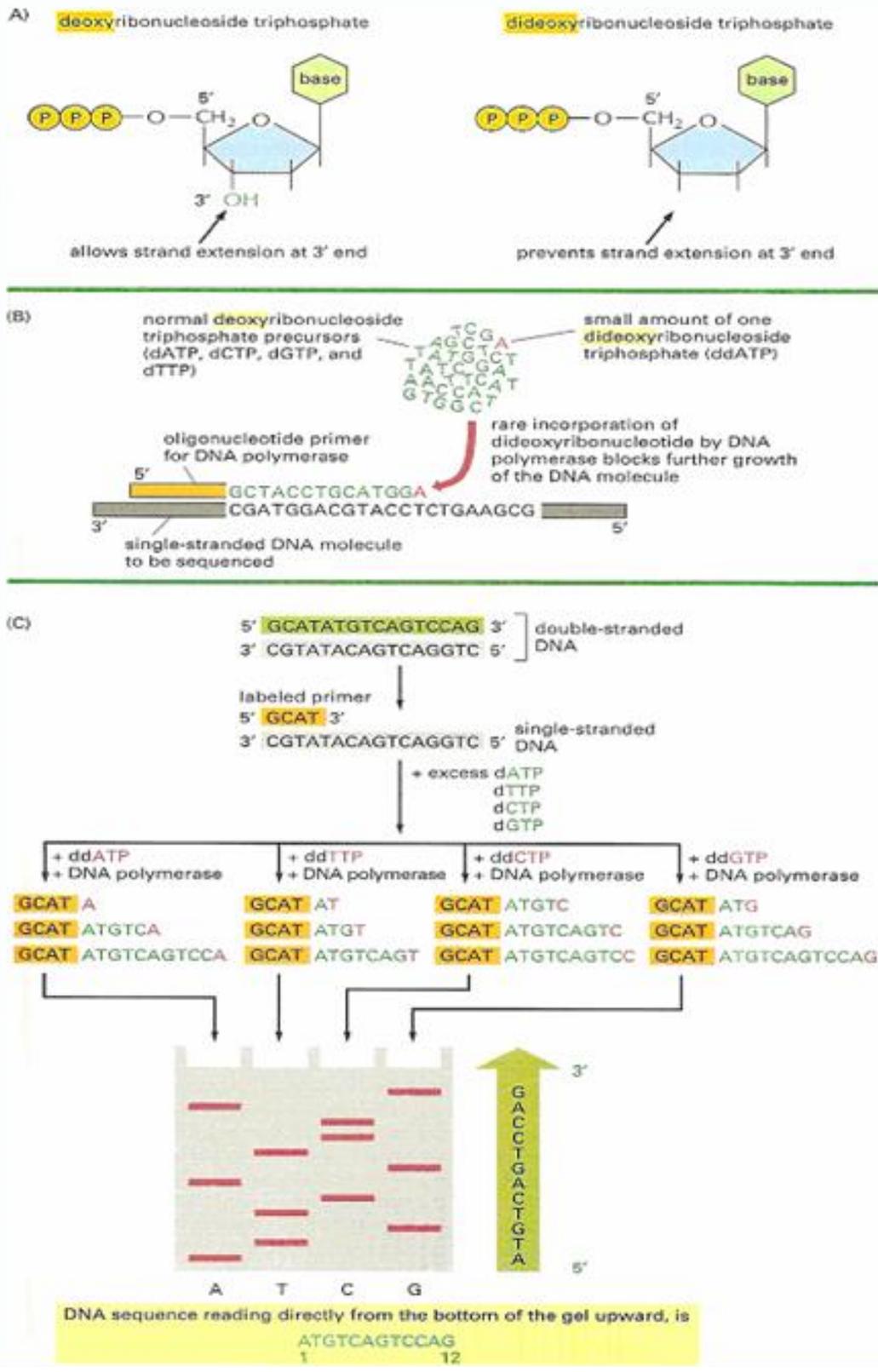


Figura 2. Sequenciamento pelo método de Sanger.

Sequenciamento automático

Hoje em dia, todo seqüenciamento de DNA em larga escala é feito por máquinas de seqüenciamento automatizado de DNA que usam o procedimento de finalização de cadeia didesoxi descrito anteriormente, mas com algumas modificações: (1) uso de corantes fluorescentes no lugar de isótopos radioativos para detectar cadeias de DNA; (2) a separação dos produtos de todas as quatro reações didesoxi de finalização de cadeia por eletroforese em um único gel ou tubo capilar; (3) uso de fotocélulas para detectar a fluorescência dos corantes à medida que eles passam através do gel ou do tubo capilar; e (4) a transferência direta da saída da fotocélula para um computador, que analisa automaticamente, registra e imprime os resultados.

Os seqüenciadores de DNA analisam automaticamente as fitas de DNA que são marcadas com qualquer um dos quatro didesoxirribonucleosídeos terminadores (ddATP, ddCTP, ddGTP, ddTTP) marcados com compostos fluorescentes, que são usados para identificar cada uma das posições dos diferentes nucleotídeos na cadeia. Utilizando-se terminadores marcados, uma única reação é realizada e aplicada em um gel (placa ou capilar). Os fragmentos de DNA marcados são separados em gel de acordo com o tamanho. Um feixe de laser rastreia o gel, excitando os marcadores, que emitem luz em um comprimento de onda específico, de acordo com o corante utilizado. A luz é coletada e o padrão de espectro analisado com ajuda de programas que permitem gerar as seqüências de DNA.

Um corante fluorescente diferente é usado para marcar os produtos de cada uma das quatro reações de seqüenciamento com término de cadeia por didesoxi. Como resultado os produtos das quatro reações podem ser distinguidos por sua fluorescência à medida que passam por um gel ou por um tubo capilar. Os corantes fluorescentes podem ser acoplados aos *primers* ou diretamente aos didesoxirribonucleosídeos trifosfatos.

Objetivo: determinar a seqüência de DNA da região *barcode* por meio de seqüenciador automático.

Procedimentos:

1. Preparar reação de seqüenciamento:

4 µL do mix (DYEnamic ET Dye Terminator, MegaBace, GE Healthcare)

1 µL do *primer* M13F ou M13R

5 µL do produto de PCR (DNA) purificado com PEG

2. Colocar os tubos contendo as amostras em termociclador programado para a seguinte ciclagem:

Passo	Temperatura	Tempo
1	95°C	30 seg
2	55°C	15 seg
3	60°C	1 min
4	Voltar para o passo 1 – 25 X	
5	4°C	indeterminado

8. Purificação da reação de sequenciamento

Objetivo: purificar as reações de seqüenciamento, de modo a ficarem livres de *primers* e dNTPs, para serem lidas em seqüenciador automático

Procedimento:

1. Em um tubo de microcentrífuga (1,5 mL) adicionar:
 - 10 µL da amostra de DNA amplificada para o sequenciamento,
 - 10 µL de água Milli-Q
 - 2 µL de acetato de amônio 7,5 M
 - 55 µL de etanol 100% (Merck; gelado)
2. Misturar por inversão 15X ;
3. Deixar em repouso por 15 minutos em temperatura ambiente e no escuro;
4. Centrifugar a 12.000 RPM por 20 minutos a 4°C.
5. Descartar sobrenadante VAGAROSAMENTE em papel;
6. Adicionar 150 µL de etanol 70% (gelado);
7. Centrifugar a 12.000 RPM por 10 minutos a 4°C;
8. Descartar sobrenadante;
9. Repetir os passos 6, 7 e 8;
10. Secar em temperatura ambiente no escuro;

11. Adicionar 10 μ L de água Milli-Q;
12. Transferir o DNA para placa de seqüenciamento;
13. Guardar em freezer -20°C até o momento do seqüenciamento.
14. Sequenciar em seqüenciador automático MegaBace

9. Análise das seqüências nucleotídicas e bioinformática

Introdução

O seqüenciamento nos fornece a seqüência de bases nucleotídicas de uma determinada região do genoma que está sendo estudada. No entanto, para manipular e analisar estas seqüências faz-se necessário o uso de ferramentas da bioinformática.

Procedimentos:

1. BLAST (*Basic Local Alignment Search Tool*)

BLAST é provavelmente a ferramenta computacional mais utilizada em biologia molecular e bioinformática. Tem a função de buscar seqüências armazenadas nos bancos de dados pela similaridade entre a estrutura primária da seqüência *query* e as seqüências armazenadas no banco.

É uma ferramenta online, disponibilizada no site <http://blast.ncbi.nlm.nih.gov/>. Para nucleotídeos, a opção “**nucleotide blast**” (**blastn**) busca uma seqüência nucleotídica no banco de dados com base na seqüência *query*. O programa libera resultados com representação gráfica dos alinhamentos, sumário com uma descrição em uma linha de cada *hit* bem como os alinhamentos com seus respectivos parâmetros calculados.

2. Edição das seqüências nucleotídicas

Nem sempre o seqüenciamento fornece seqüências com 100% de confiança. Para reparar erros advindos do seqüenciamento, recomenda-se fazer a edição dessas seqüências, o que compreende verificar o cromatograma gerado pelo seqüenciador, buscando por picos duvidosos e com erros (N). O programa **BioEdit** é um editor de seqüências que roda em ambiente Windows e que pode ser utilizado para editar, alinhar e manipular seqüências.

Pode ser obtido gratuitamente em <http://www.mbio.ncsu.edu/BioEdit/bioedit.html>. Devido sua interface gráfica, facilita a edição manual das seqüências.

3. Alinhamento de múltiplas seqüências

O alinhamento de três ou mais seqüências nucleotídicas pode ser realizado por meio do programa **ClustalW**, disponível online em <http://www.ebi.ac.uk/Tools/msa/clustalw2/>. É utilizado para identificar regiões conservadas e sítios polimórficos ao longo da seqüência.

4. Análises de distância genética

Quando as seqüências alinhadas são comparadas baseadas em modelos que reflitam sua semelhança, é possível quantificar as diferenças entre elas por meio de valores numéricos. Baseado nesse princípio, a distância genética entre pares de seqüências pode ser estimada a fim de identificar as espécies, utilizando o modelo de evolução de Kimura-2-parâmetros e o algoritmo de *Neighbor-Joining*.

Foi padronizado para estudos de DNA *barcodes* o emprego do modelo de substituição molecular **Kimura-2-parâmetros, K2P** (Kimura, 1980). K2P é o melhor modelo a ser adotado quando as distâncias genéticas dos organismos em estudo são baixas. Nesse modelo é considerada a razão de transições e transversões, levando em conta o desvio na direção das substituições mais comuns, as transições.

5. Construção da árvore filogenética

O procedimento *Neighbor-Joining, NJ*, foi considerado como método padrão de inferências filogenéticas nos estudos de DNA *barcodes* (Hebert et al., 2003) devido sua alta eficiência e rapidez em análises de espécies em larga escala (Kumar e Gadagkar, 2000). A vantagem desse algoritmo é não assumir que a taxa de evolução entre as seqüências que compõem o conjunto de dados seja constante.

Tanto para a obtenção dos valores de distância genética quanto para a construção da árvore filogenética *Neighbor-Joining*, utiliza-se o programa **MEGA (Molecular Evolutionary Genetics Analysis)**, o qual pode ser obtido em <http://www.megasoftware.net/>.

Leituras sugeridas

GONÇALVES, P.F.M. O potencial do *DNA barcode* na identificação de espécies de aves neotropicais. (2009). Dissertação (Pós-graduação em Ciências). Universidade de São Paulo.

HEBERT P.D., CYWINSKA A., BALL S.L., DEWAARD J.R. (2003). Biological identifications through DNA barcodes. *Proceedings of the Royal Society B: Biological Sciences* 270(1512): 313-21.

HEBERT P.D.N., PENTON E.H., BURNS J.M., JANZEN D.H., HALLWACHS W. (2004). Ten species in one: DNA barcoding reveals cryptic species in the neotropical skipper butterfly *Astraptes fulgerator*. *Proc Natl Acad Sci U S A*. 101(41):14812-7.

KIMURA, M.A. (1980). Simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *Journal of Molecular Evolution*, 16: 111-120.

KUMAR, S.; GADAGKAR, S.R.. (2000). Efficiency of the neighbor-joining method in reconstructing deep and shallow evolutionary relationships in large phylogenies. *Journal of Molecular Evolution*, 51: 544-553.

MEYER, C.P.; PAULAY, G. (2005). DNA barcoding: error rates based on comprehensive sampling. *PLoS Biol.*, 3: e422.

RATNASINGHAM, S., HEBERT, P.D.N. (2007). Barcoding BOLD: The Barcode of Life Data System. *Molecular Ecology Notes* 7(3): 355-64.

RUBINOFF, D. (2006). Utility of mitochondrial DNA barcodes in species conservation. *Conserv. Biol.*, 4: 1026-1033.

SOLÉ-CAVA, A. M. (2008). Código de barras de DNA: o rabo que abana o cachorro. *Ciência Hoje*, 41(245): 65-67.

STOECKLE M. (2003). Taxonomy, DNA, and the Bar Code of Life. *BioScience* 53(9): 2-3.