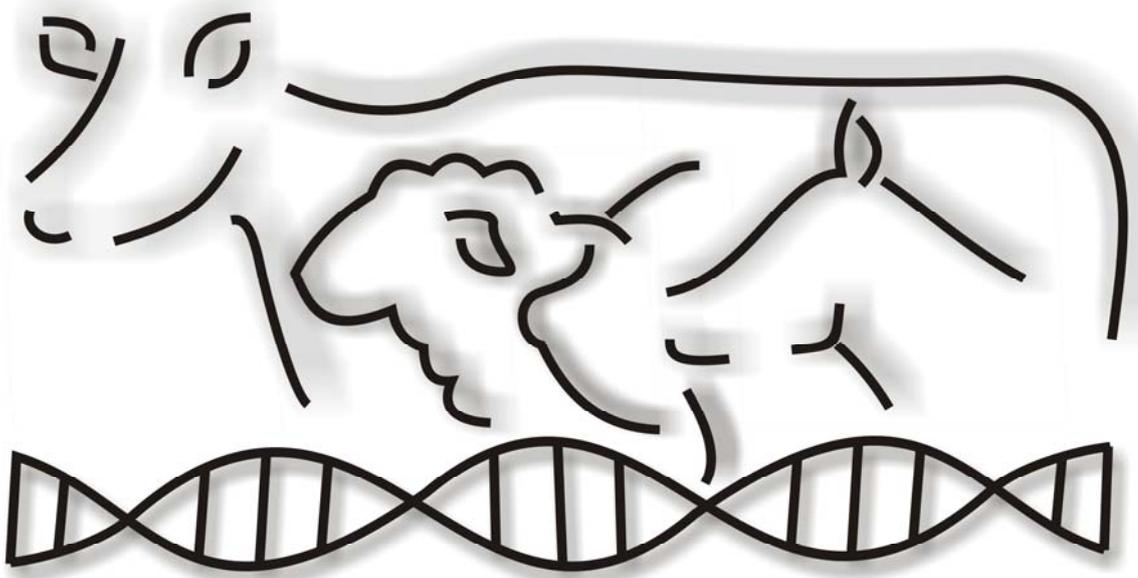


# Manual de Laboratório

---



**Curso:**  
**Técnicas de Biologia Molecular**  
**Aplicadas a Produção Animal**

**2009**

## Organização

Dr<sup>a</sup> Luciana Correia de Almeida Regitano - Embrapa Pecuária Sudeste  
Dr<sup>a</sup> Simone Cristina Méo Niciura - Embrapa Pecuária Sudeste  
Msc. Adriana Mércia Guaratini Ibelli – PPGGEv/UFSCar  
Msc. Gisele Batista Veneroni – PPGGEv/UFSCar  
Polyana Cristine Tizioto - PPGGEv/UFSCar  
Equipe da Área de Comunicação e Negócios Embrapa Pecuária Sudeste

## Colaboração

Dr. Felipe Rodrigues da Silva - Embrapa Recursos Genéticos e Biotecnologia  
Dr. Gustavo Gasparin – Bolsista DTI - CNPq  
Dr. Henrique Nunes de Oliveira - UNESP/Botucatu  
Dr. Luiz Lehmann Coutinho - ESALQ/USP  
Msc. Marcelo Cervini - PPGGEv/UFSCar  
Dr. Marcos Vinicius Gualberto B. Silva – Embrapa Gado de Leite  
Dr. Michel Eduardo Beleza Yamagishi – Embrapa Informática Agropecuária  
Dr. Millor Fernandes do Rosário - ESALQ/USP  
Dr. Andre Carlos Ponce de Leon Ferreira de Carvalho – ICMC/USP – São Carlos  
Dr<sup>a</sup>. Andrea Cristina Basso - In Vitro Brasil Ltda.  
Dr<sup>a</sup>. Christina Ramires Ferreira – LaboratórioThoMson de Espectrometria de Massas, Instituto de Química-UNICAMP  
Dr<sup>a</sup>. Liliana Araceli Soria - Facultad de Ciencias Veterinarias/Universidad de Buenos Aires  
Dr<sup>a</sup>. Márcia Cristina de Sena Oliveira Oliveira - Embrapa Pecuária Sudeste  
Msc. Minos Esperândio Carvalho – ESALQ/USP  
Dr<sup>a</sup>. Natália Martins - Embrapa Recursos Genéticos e Biotecnologia  
Dr<sup>a</sup>. Paula Kuser Falcão – Embrapa Informática Agropecuária  
Rodrigo Giglioti – Programa de Pós-Graduação em Zootecnia – FCAV/UNESP- Jaboticabal

## Editoração

Maria Cristina Campanelli Brito – Embrapa Pecuária Sudeste

## Realização



## Apoio



## Patrocínio



# ÍNDICE

1. Regras de segurança no laboratório	4
2. Extração de DNA	6
3. Extração de DNA de sangue utilizando o kit GFX	10
4. Extração de RNA	11
5. Quantificação em espectrofotômetro	14
6. Reação em cadeia da polimerase (PCR)	17
7. Enzimas de restrição	21
8. Eletroforese de ácidos nucleicos	24
9. Eletroforese de RNA em gel desnaturante	27
10. Eletroforese em sistema capilar	30
11. Seqüenciamento	32
12. Síntese de DNA complementar (cDNA)	40
13. Delineamento de Primers	43
14. PCR quantitativa em tempo real	49
15. Genotipagem de SNPs utilizando sondas Taqman	54
16. Mapas de Ligação - software CRI-MAP (versão 2.4)	58
17. Mapeamento de QTLs – <i>QTL Express</i>	67
18. microRNAs: uma nova classe de moduladores da expressão gênica	73
19. Prática Mineração de dados em Bioinformática	82
20. Preparo de Soluções	88
21. Referências bibliográficas	91
22. Literatura consultada	99

## 1. REGRAS DE SEGURANÇA EM LABORATÓRIOS

*Luciana Correia de Almeida Regitano*

Segurança é prioridade em qualquer laboratório químico.

É importante conhecer o ambiente do laboratório, familiarizar-se com todo equipamento de segurança disponível no laboratório, sabendo onde encontrá-lo e como utilizá-lo:

- lava-olhos;
- chuveiro de emergência;
- extintores;
- saídas de emergência.

Muitas regras de segurança envolvem apenas o bom “senso”:

- não tocar em itens quentes;
- não comer, beber ou fumar no laboratório;
- não fazer brincadeiras;

O uso de Equipamentos de Proteção Individual (EPI's) é obrigatório, assim como o uso de capela de exaustão em caso de solventes.

Usar sempre óculos de segurança e avental de preferências de algodão e de mangas longas.

Não usar saias, bermudas ou calçados abertos.

Manter cabelos longos, sempre presos.

Não trabalhar sozinho, principalmente fora do horário de expediente.

Lavar bem as mãos ao deixar o laboratório

Manter vidrarias sempre em condições adequadas de trabalho.

Manter as soluções sempre identificadas e com os cuidados a serem tomados em caso de acidentes.

Manter os frascos de reagentes sempre limpos, sem respingos para assegurar boa condição de uso por técnicos e estagiários.

Ao executar qualquer experimento estar consciente de como proceder e dos cuidados a serem tomados.

Ler atentamente o protocolo até o fim, verificando se compreendeu todos os procedimentos e se possui todo o material necessário.

Certificar-se da tensão de trabalho dos aparelhos antes de conectá-los à rede elétrica. Quando não estiverem em uso, os aparelhos devem ficar desconectados.

Não usar nenhum equipamento para o qual não tenha sido autorizado ou treinado.

Respeitar seus amigos de trabalho, não os distraindo com conversas paralelas.

Caso perceba algo diferente ou mesmo anormal, recorra ao técnico responsável, para tirar suas dúvidas.

## 2. EXTRAÇÃO DE DNA

*João José de Simoni Gouveia*

*Luciana Correia de Almeida Regitano*

A extração dos ácidos nucleicos (DNA e RNA) é o primeiro passo para a realização da maioria das metodologias de biologia molecular. Pode-se obter DNA a partir de inúmeros tipos de tecidos e células, e existe uma infinidade de protocolos para a realização de tal procedimento. A escolha do protocolo de extração de DNA dependerá de diversos fatores como: tipo de tecido a ser utilizado, grau de pureza e de integridade necessária para a aplicação em que o DNA será utilizado (PCR, seqüenciamento, clonagem gênica, etc.) (Bartlett, 2003).

O primeiro relato sobre isolamento de DNA data de 1869 e foi realizado por Friedrich Miescher, a partir de leucócitos presentes em pus. Miescher lisou essas células com uma solução alcalina e precipitou o material nuclear com solução salina. Esse material nuclear, que posteriormente seria conhecido como DNA, foi chamado de nucleína (<http://www.dbbm.fiocruz.br/helpdesk/mbiology/historico2004.pdf>; <http://www.dnai.org>).

Basicamente, o processo de extração de DNA consiste em duas etapas. A primeira etapa é a extração propriamente dita e consiste no rompimento das membranas celulares e conseqüente exteriorização do DNA. A segunda fase consiste na purificação do DNA em solução, ou seja, na remoção dos outros componentes celulares da solução (restos de membrana, proteínas, RNA) (Romano e Brasileiro, 1999).

O rompimento das membranas celulares geralmente é feito com detergentes (SDS ou CTAB). A utilização de agentes caotrópicos como o tiocianato de guanidina impede o DNA de se ligar a outras moléculas, facilitando sua separação dos outros componentes celulares, na segunda fase do processo. A separação é feita por meio da adição de substâncias que tornam a solução heterogênea de maneira que o DNA fique dissolvido em apenas uma das fases. Por exemplo, quando se utiliza fenol para desnaturar as proteínas, o DNA fica na fase aquosa e as proteínas permanecem na interface entre as fases orgânica e aquosa. Uma alternativa à utilização dos solventes orgânicos como o fenol é fazer a separação das proteínas

utilizando altas concentrações de sal (*salting-out*). Após separar o DNA dos outros componentes celulares, pode-se proceder à precipitação do DNA para garantir a máxima pureza do material. Essa precipitação geralmente é feita com álcool (etanol ou isopropanol) que, em presença de cátions monovalentes, promove uma transição estrutural na molécula de ácido nucléico que resulta em agregação e em precipitação (Regitano, 2001; Azevedo, 2003; <http://www.etall.hpg.ig.com.br/cursopcr.html>).

## 2.1. Extração de DNA de Leucócitos

Os protocolos abordados neste capítulo são utilizados na rotina do laboratório de Biotecnologia Animal do Centro de Pesquisa de Pecuária do Sudeste e são adaptações dos protocolos descritos em Regitano (2001).

O objetivo deste capítulo é fazer com que o aluno entenda os princípios deste processo, a finalidade de cada reagente em um protocolo de extração de DNA e, a partir daí, possa utilizar e adaptar os protocolos já existentes de acordo com sua necessidade e com a disponibilidade de reagentes. A descrição do preparo das soluções utilizadas encontra-se no final do livro.

### A) Obtenção de leucócitos

1. Coletar 5 mL de sangue em tubos contendo EDTA potássico [50 µL de EDTA(K<sub>3</sub>) a 15%]

*O EDTA é uma substância anticoagulante. Existem outras substâncias com essa propriedade (citrato e heparina), porém se o DNA for utilizado para PCR não se deve utilizar heparina, pois essa substância interfere na atividade da enzima Taq DNA polimerase.*

2. Transferir 2,5 mL do sangue para um tubo Falcon de 15 mL
3. Adicionar 10 mL de Solução A e vortexar até homogeneizar
4. Centrifugar por 10 min a 700 xg
5. Dispensar o sobrenadante
6. Ressuspender o pellet em 5 mL de Solução A

7. Vortexar até dissolver completamente o pellet
8. Centrifugar por 10 min a 700 xg
9. Repetir os passos 4 – 9 até obter somente as células brancas (o pellet deve estar branco cremoso). *Atenção para não deixar resíduos de hemoglobina, pois os produtos de oxidação dessa proteína inibem a reação de PCR*
10. Ressuspender o pellet em 500 µL de Solução A e transferir para microtubos de 1,5 mL
11. Centrifugar 5 min a 16.000 xg e descartar o sobrenadante.

*As células brancas assim preparadas podem ser armazenadas entre -20°C e -80°C.*

## **B) Extração e purificação do DNA**

1. Ressuspender o pellet em 500 µL de Solução B
2. Vortexar até o pellet desprender do fundo do tubo
3. Incubar a 55°C por 4-6 horas (ou *overnight*). Vortexar periodicamente durante a incubação para que a dissolução do pellet seja completa

*A Solução B contém Tris HCl (pH 8,0), uma solução tampão cuja finalidade é manter o pH constante. Como o pH ótimo para a ação de DNases endógenas é por volta de 7,0, esse reagente ajuda a evitar a ação dessas nucleases. O SDS é um detergente iônico forte, cuja função é romper as membranas celulares; o EDTA quelata íons  $Mg^{2+}$  e  $Ca^{2+}$  (que são co-fatores de diversas enzimas nucleares como as DNases); e a proteinase K hidrolisa as proteínas.*

*Se houver disponibilidade de um bloco aquecido com agitador (termomixer), deve-se deixar as amostras sob agitação durante a incubação.*

4. Adicionar 210 µL de TE (pH 7,6) e 240 µL de NaCl 5M

*O NaCl fará com que as proteínas precipitem por excesso de íons e o TE é um tampão cuja finalidade é manter estável o pH da solução.*

5. Agitar os tubos por inversão até formar pequenos coágulos

6. Incubar em gelo por 10 min
7. Centrifugar por 15 min a 16.000xg
8. Recolher o sobrenadante dividindo-o em dois microtubos de 1,5 mL limpos (máximo de 500 µL de sobrenadante por tubo)
9. Adicionar 1 mL de etanol 100% (absoluto) gelado em cada tubo e misturar por inversão

*É fundamental que o volume de etanol corresponda à, no mínimo, 02 vezes o volume da solução para que a precipitação do DNA seja eficiente. Alternativamente, pode-se utilizar 01 volume de isopropanol à temperatura ambiente. O álcool torna o meio muito hidrofóbico para o DNA, fazendo com que ele precipite sobre sua própria estrutura.*

10. Centrifugar por 15 min a 16.000 xg
11. Descartar o sobrenadante
12. Secar em papel
13. Adicionar 500 µL de etanol 70% gelado

*O álcool 100% faz com que precipitem muitos sais juntamente com o DNA, além de dificultar a ressuspensão do DNA, por isso utiliza-se uma “lavagem” com álcool 70%.*

14. Centrifugar por 5 min a 16.000 xg
15. Descartar o etanol e secar o pellet (manter o microtubo invertido por 10-15 min)
16. Adicionar 250 µL de TE+RNase (10 µg de RNase por mL de amostra) e incubar por 1h a 37°C (ou deixar na geladeira *overnight*)

*Esse protocolo isola o DNA juntamente com quantidade considerável de RNA, assim, o tratamento com RNase faz-se necessário para a remoção desse contaminante.*

17. Avaliar a integridade (por eletroforese) e a quantidade de DNA (em espectrofotômetro).

### **3. EXTRAÇÃO DE DNA DE SANGUE UTILIZANDO O KIT GFX**

*Márcia Cristina de Sena Oliveira*

1. Colocar 300  $\mu$ L de sangue e 900  $\mu$ L de solução de lise (3 vezes o volume de sangue) em um microtubo de 1,5 mL
2. Homogeneizar, centrifugar a 21.000 xg por 5 minutos e descartar o sobrenadante
3. Homogeneizar e adicionar 500  $\mu$ L de solução de extração
4. Incubar por 5 minutos à temperatura ambiente e transferir o conteúdo do tubo para a coluna de extração acoplada a um tubo coletor
5. Centrifugar a 5.000 xg por 1 minuto
6. Descartar o conteúdo e reaproveitar o tubo coletor
7. Adicionar à coluna 500  $\mu$ L de solução de extração e centrifugar a 5.000 xg por 1 minuto
8. Descartar o conteúdo e reaproveitar o tubo coletor
9. Adicionar à coluna 500  $\mu$ L de solução de lavagem e centrifugar a 21.000 xg por 3 minutos
10. Transferir a coluna para um microtubo de 1,5 mL, limpo e livre de DNAses
11. Adicionar 100  $\mu$ L de água bidestilada (ou TE), autoclavada e aquecida a 70°C. Incubar por 1 minuto à temperatura ambiente
12. Centrifugar a 5.000 xg por 1 minuto.

#### 4. EXTRAÇÃO DE RNA

*Adriana Mércia Guaratini Ibelli*

*Luciana Correia de Almeida Regitano*

*Simone Cristina Méo*

A análise do RNA pode fornecer informações importantes sobre a expressão gênica. Para que isso seja possível, é necessária a purificação eficaz do RNA, mantendo sua integridade e sua qualidade.

Os métodos para isolamento de RNA baseiam-se na lise e na desnaturação das células, que permitem a liberação dos ácidos nucleicos totais, na presença de inibidores que impedem a ação de ribonucleases (RNAses) (Ausubel, 1987).

A principal dificuldade na extração de RNA é a presença de grande quantidade de RNAses estáveis e ativas nos tecidos, que degradam o RNA (altamente instável) rapidamente. Dessa maneira, após a pulverização dos tecidos, a primeira etapa em todos os métodos de isolamento de RNA é a exposição desse material a tampões de extração. Esses tampões possuem substâncias como o cloreto de lítio, que auxilia a precipitação do RNA (o RNA possui maior densidade que o DNA e as proteínas), e o isotiocianato de guanidina, que permite a manutenção do RNA intacto, por meio da degradação das ribonucleases endógenas, nas etapas posteriores da extração (Sambrook e Russel, 2002).

Atualmente, há vários reagentes comerciais como, por exemplo, Trizol (Invitrogen<sup>®</sup>) e Brazol (Lab Trade do Brasil<sup>®</sup>) que possuem em sua composição reagentes combinados, como isotiocianato de guanidina e fenol, possibilitando a extração de RNA mais rápida que a dos protocolos convencionais e garantindo a integridade do material (Sambrook e Russel, 2002). É necessária ainda a adição de clorofórmio ao tampão contendo RNA para a solubilização e a remoção dos lipídios. Após a centrifugação, três fases poderão ser observadas: uma rósea formada pela presença do fenol, uma interfase formada pelo clorofórmio e pelo DNA e uma fase aquosa, onde estará presente o RNA. Nessa técnica, o RNA permanece solúvel em solução com 4M de isotiocianato de guanidina pH 4,0 na presença de fase orgânica de fenol/clorofórmio, enquanto, nessa condição ácida, proteínas e DNA menores (de 50 bp a 10 kb) ficam na fase orgânica e proteínas e DNA maiores, na interfase.

A precipitação do RNA é feita com um álcool, como por exemplo, o isopropanol. Nessa etapa, observa-se uma nuvem branca contendo o RNA. Posteriormente, uma limpeza com álcool permite a retirada de sais que tenham aderido ao precipitado de RNA (Ausubel, 1987).

Na extração de RNA, alguns fatores além da escolha do protocolo utilizado, são extremamente importantes para a obtenção e a manutenção do RNA de qualidade.

Primeiramente, todas as soluções utilizadas deverão ser feitas com água DEPC (dietilporocarbonato) autoclavada. O DEPC atua inativando RNAses, pois degrada as histidinas presentes. Esse tratamento não deve ser realizado em tampões que utilizam Tris, pois o DEPC inativa tal componente.

Para o trabalho com RNA, os almofarizes, os pistilos, as vidrarias e os outros utensílios devem ser previamente esterilizados em estufa a 180°C por um período de quatro horas. As ponteiros e os tubos de polipropileno devem ser novos e livres de RNAses. As micropipetas devem ser utilizadas apenas para os procedimentos com RNA. A bancada deve ser constantemente limpa com SDS 10% e com etanol 70% e, se possível, exclusiva para a manipulação de RNA. E, principalmente, devem ser tomados cuidados com as luvas utilizadas, de forma a mantê-las livres das RNAses, normalmente liberadas abundantemente pelas secreções da pele.

#### **4.1. Protocolo de extração de RNA** (adaptado de Chomczynski & Sacchi, 1987).

1. Macerar o tecido em nitrogênio líquido
2. Para cada 50 a 100 mg de tecido, adicionar 1 mL de trizol (em tubo de polipropileno de 1,5 mL). Homogeneizar em vórtex
3. Incubar 5 min à temperatura ambiente
4. Acrescentar 200 µL de clorofórmio
5. Agitar vigorosamente com as mãos por 15 segundos
6. Incubar durante 5 min à temperatura ambiente
7. Centrifugar a 16.000 xg a 4°C durante 15 min
8. Remover a fase aquosa para microtubo limpo
9. Adicionar 500 µL de isopropanol. Homogeneizar com as mãos
10. Incubar por 10 min à temperatura ambiente
11. Centrifugar a 13.000 xg a 4°C por 10 min

12. Descartar o sobrenadante

13. Lavar com 1 mL de etanol 75% (feito com água DEPC)

*(Neste ponto, pode armazenar em freezer -20°C por até um ano)*

14. Centrifugar a 10.500 xg a 4°C por 5 min

15. Secar o *pellet* durante 15 min à temperatura ambiente

*Obs. Não pode ficar nada de etanol.*

16. Ressuspender o *pellet* em água DEPC (20 µL a 50 µL)

*Obs. Verificar o tamanho do pellet, para avaliar em que volume ressuspender.*

17. Manter a 55°C por 10 min

18. Avaliar a integridade (por eletroforese) e a quantidade de RNA (em espectrofotômetro).

#### **4.2. Protocolo de precipitação de RNA com acetato de sódio**

Uma vez que o RNA em solução é mais instável, após a quantificação e a utilização do RNA extraído, ele pode ser precipitado com acetato de sódio. Esse procedimento visa à maior conservação do RNA durante o período de armazenamento a -80°C.

1. Adicionar 0,1 volume (valor igual a 10% do volume do RNA em solução) de acetato de sódio 3M + 2,5 volumes de etanol absoluto gelado e misturar por inversão

2. Deixar em freezer -20°C por, pelo menos, 2 horas (ou por 1 hora a -80°C)

3. Centrifugar por 30 min a 4°C em velocidade máxima

4. Descartar o etanol sobrenadante

5. Adicionar 500 µL de etanol 75% (preparado com água DEPC) e misturar por inversão

6. Centrifugar por 5 min a 4°C em velocidade máxima

7. Descartar o etanol sobrenadante

8. Secar o *pellet*

Posteriormente, para o RNA ser utilizado novamente:

9. Ressuspender o *pellet* em água DEPC

10. Manter a 55°C por 10 min.

## 5. QUANTIFICAÇÃO EM ESPECTROFOTÔMETRO

*Luciana Correia de Almeida Regitano*

Os ácidos nucleicos absorvem luz no comprimento de onda de 260 nm. Para fazer a leitura no espectrofotômetro, normalmente utiliza-se uma diluição da amostra de ácidos nucleicos em água. Para estimar a concentração de DNA utiliza-se a seguinte relação:  $1 \text{ OD}_{260} = 50 \mu\text{g}$  de DNA dupla-hélice (Regitano, 2001; Sambrook e Russel, 2002).

A concentração de DNA na amostra é obtida pelo seguinte cálculo:

$$[\text{DNA}] = \text{Valor da leitura em O.D.} \times 50 \times \text{Fator de diluição}$$

As proteínas absorvem luz no comprimento de onda de 280 nm. Sendo assim, a relação  $A_{260}/A_{280}$  fornece um parâmetro de avaliação da qualidade das preparações de ácidos nucleicos. Valores inferiores a 1,8 resultam de contaminação com proteína.

A concentração de RNA nas amostras também deve ser avaliada por espectrofotometria. O grau de pureza do RNA será estimado pela razão entre as absorbâncias medidas a 260 nm e a 280 nm e deve ser igual ou superior a 1,75 (Sambrook e Russel, 2002). Para uma amostra de RNA pura tem-se que:  $\text{OD}_{260}/\text{OD}_{280} = 1,8$  a  $2,0$ . Uma densidade óptica igual à 1,0 corresponde a  $40 \mu\text{g}$  de RNA (fita simples) por mL. Dessa forma, a concentração de RNA na amostra pode ser calculada por meio da seguinte relação:

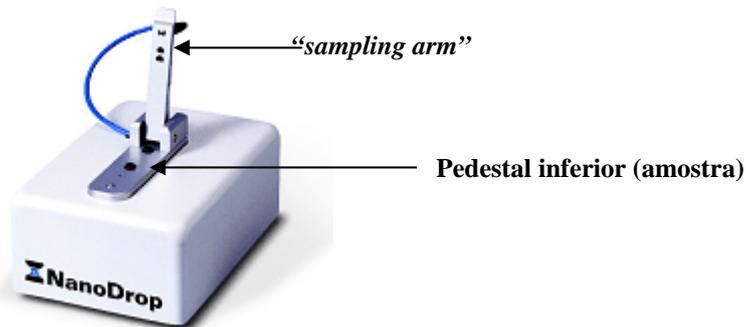
$$\text{Concentração de RNA } (\mu\text{g/mL}) = \text{OD}_{260} \times \text{Fator de diluição (100)} \times 40$$

### 5.1. Quantificação do DNA no NanoDrop®

*Polyana Cristine Tizioto*

O NanoDrop® ND-1000 (Fabricante) é um espectrofotômetro sensível, que possui cobertura de espectro de 220 a 750nm. Este aparelho consegue mensurar a concentração da amostra em volume pequeno ( $1\mu\text{L}$  a  $2\mu\text{L}$ ) com rapidez. Tem a capacidade de quantificar amostras com amplo espectro de concentração, de  $2\text{ng}/\mu\text{L}$  até  $3700\text{ng}/\mu\text{L}$ , não havendo necessidade de diluição nessa faixa de concentração.

O NanoDrop<sup>®</sup> utiliza uma tecnologia de retenção de amostra patenteada na qual a tensão superficial mantém a amostra no local, eliminando assim a necessidade de cubetas. A quantificação da amostra é visualizada no computador, o software do NanoDrop<sup>®</sup> salva os dados em planilhas compatíveis com o Excel (Manual NanoDrop ND-1000).



**Figura 5.1.** Equipamento NanoDrop<sup>®</sup>

(<https://www.pasteur.fr/recherche/genopole/PF2/images/nanodrop.gif>)

Passos para quantificação de ácidos nucleicos no NanoDrop<sup>®</sup>:

- 1- Abra o software do ND-1000, coloque as informações do administrador (nome e senha) e selecione a opção ácidos nucleicos.
- 2- Com o *“sampling arm”* aberto, pipete 2  $\mu\text{L}$  do solvente em que a amostra foi ressuspendida, feche o *“sampling arm”*.
- 3- Selecione no programa aberto no computador a opção *“blank”*. Este passo assegura que o equipamento está funcionando corretamente e o pedestal onde coloca a amostra está limpo.
- 4- Se o campo  $\text{ng}/\mu\text{L}$  do programa mostrar o valor 0 (zero), a quantificação da amostra pode ser iniciada.
- 5- Pipete 2  $\mu\text{L}$  da amostra no pedestal de medida.
- 6- Feche o *“sampling arm”*. Inicie a quantificação pelo software no computador e anote a identificação da sua amostra no campo *“sample ID”*, selecionando a opção *“measure”*.
- 7- Quando o aparelho terminar a quantificação, apresentando no computador os valores da concentração da amostra, o gráfico, a relação 260/280 (pureza de ácidos nucleicos) e demais informações, abra o *“sampling arm”* e limpe as



## 6. REAÇÃO EM CADEIA DA POLIMERASE (PCR)

*Adelita Carolina Santiago*

*Gisele Batista Veneroni*

*Luciana Correia de Almeida Regitano*

A reação em cadeia da polimerase (PCR) nada mais é do que a replicação *in vitro* do DNA. Foi desenvolvida na década de 1980 por Kary Mullis, que conseguiu pela primeira vez amplificar pequenos fragmentos de DNA. Em 1988, foi descoberta uma enzima DNA polimerase termoestável, isolada de *Thermus aquaticus*, permitindo que esse processo ganhasse automação (Saiki et al., 1988).

O impacto desse método foi enorme, pois possibilitou facilidade, rapidez, versatilidade na manipulação de ácidos nucleicos, o que tornou a técnica poderosa e imprescindível em grande parte dos procedimentos realizados atualmente. É utilizada, por exemplo, na realização de seqüenciamento, como diagnóstico de doenças, detecção de mutações pontuais, entre outros.

Essa técnica envolve três etapas: desnaturaç o do DNA pelo calor entre 94-95°C, anelamento de *primers* à seqüência alvo em fita simples, atuando de 3 a 5°C abaixo da temperatura de *melting* (35 a 60°C) e uma etapa de extensão do *primers* anelados por uma DNA polimerase termoestável (72 a 78°C).

Alguns componentes são essenciais para que ocorra a PCR:

1. Presença de uma DNA polimerase termoestável, como por exemplo, *Taq* (*Thermus aquaticus*) ou *Tfl* (*T. flavus*);
2. Presença de um par de *primers*, oligonucleotídeos que funcionam como ponto de início da polimerização;
3. Cátions divalentes, como MgCl<sub>2</sub>;
4. Desoxinucleotídeos trifosfatados (dNTPs), com quantidades iguais de dATP, dCTP, dTTP e dGTP;
5. DNA molde, podendo ser fita simples ou fita dupla;
6. Tampão para manutenção do pH, geralmente entre 8,3 e 9 à temperatura ambiente.

### 6.1. Protocolo de Amplificação do Microssatélite BM6026

O BM6026 é um marcador molecular do tipo microssatélite de repetição “CA”, localizado a cerca de 7,02 cM no cromossomo cinco em bovinos (BTA5). Seus alelos apresentam tamanho entre 148 e 168 bp.

1. Retirar o DNA do freezer e deixar à temperatura ambiente
2. Lavar as mãos e colocar luvas;
3. Limpar a bancada com álcool 70%
4. Marcar os tubos ou identificar a placa. Deve-se criar um registro das amostras que serão amplificadas no caderno de laboratório
5. Preparar o MIX de PCR

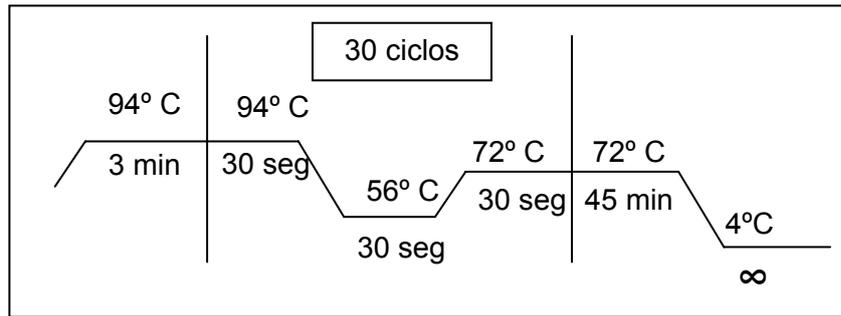
#### MIX de PCR

Reagente	[Estoque]	[Final]	VOLUME (1 amostra)	VOLUME (__ amostras)
H <sub>2</sub> O	-	-	10,95 µL	
Tampão*	10X	1X	1,5 µL	
MgCl <sub>2</sub> *	50 mM	2 mM	0,6 µL	
dNTP	20 mM	0,2 mM	0,15 µL	
Primer F	5 µM	0,1 µM	0,3 µL	
Primer R	5 µM	0,1 µM	0,3 µL	
Taq	5 U/µL	1 U	0,2 µL	
<b>TOTAL</b>			14 µL	

\* Sempre verifique se o tampão contém ou não MgCl<sub>2</sub>. Caso já contenha, não há necessidade de acrescentar. Nesse caso, adicione o volume correspondente ao MgCl<sub>2</sub> no volume de água.

6. Distribuir 14 µL do MIX de PCR sem o DNA em cada tubo
7. Adicionar 1µL de DNA (40 ng/µL) em cada tubo
8. Fazer um tubo controle (branco da reação) com 14 µL do MIX de PCR e 1 µL de água
9. Levar ao termociclador
10. Utilizar o programa de termociclagem descrito abaixo

**Programa**



11. Avaliar os produtos de PCR por eletroforese em sistema capilar e anotar os resultados na tabela abaixo.

Animal	Genótipo

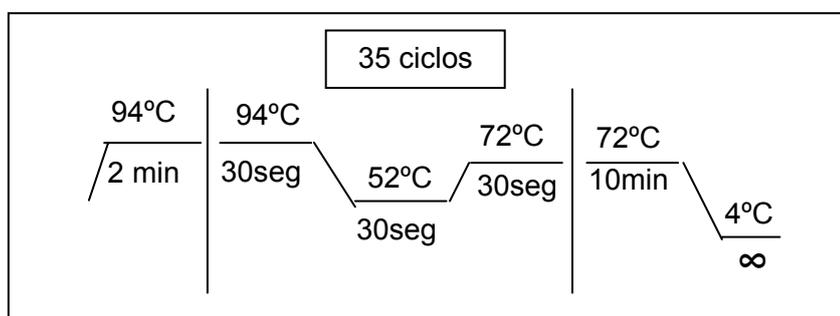
**6.2. Protocolo de amplificação da região promotora do gene liberador de Corticotrofina**

1. Retirar o DNA do freezer
2. Lavar as mãos e colocar luvas novas
3. Limpar a bancada
4. Tirar o gelinho e os reagentes (menos a Taq) do freezer
5. Marcar os tubos ou identificar a placa
6. Preparar o MIX de PCR

**MIX de PCR**

Reagente	[Estoque]	[Final]	Volume (1 amostra)	Volume (– amostras)
H <sub>2</sub> O Milli Q	-	-	12,94 µL	— µL
Tampão de PCR	10X	1X	2 µL	— µL
MgCl <sub>2</sub>	50 mM	1,95 mM	0,78 µL	— µL
dNTP	20 mM	0,20 mM	0,2 µL	— µL
Primer F	5 µM	0,165 µM	0,66 µL	— µL
Primer R	5 µM	0,165 µM	0,66 µL	— µL
Taq	5 U/µL	1.3 U	0,26 µL	— µL
<b>TOTAL</b>			17,5 µL	— µL

7. Distribuir 17,5 µL do MIX de PCR por tubo ou por poço (de placa de PCR), sempre sobre um gelinho
8. Distribuir 2,5 µL do DNA por tubo ou por poço
9. Fazer um tubo controle (branco da reação) com 17,5 µL do MIX de PCR e 2,5 µL de água
10. Colocar as amostras em um termociclador, e escolher o programa de termociclagem descrito abaixo

**Programa**

## 7. ENZIMAS DE RESTRIÇÃO

*Gisele Batista Veneroni*

*Luciana Correia de Almeida Regitano*

As enzimas de restrição, também chamadas de endonucleases de restrição, reconhecem uma seqüência específica de bases em uma molécula de DNA e clivam a molécula naquela seqüência ou próximo dela. A seqüência de reconhecimento é chamada de sítio de restrição. Diferentes enzimas de restrição reconhecem e clivam diferentes sítios de restrição.

Elas são indispensáveis na análise da estrutura dos cromossomos, no isolamento de genes, na criação de moléculas novas de DNA que podem ser clonadas e em técnicas como polimorfismo de comprimento de fragmentos de restrição (RFLP). Uma característica marcante de muitos desses sítios de clivagem é que eles possuem dupla simetria rotacional, isto é, a seqüência de reconhecimento é palindrômica e os sítios de clivagem são simetricamente posicionados.

### 7.1. Digestão do DNA – Região promotora do gene liberador de corticotrofina

1. Preparar o MIX de digestão

#### MIX de digestão

	<b>MIX 1X</b>	<b>MIX X</b>
<b>Água Milli Q</b>	0,87 $\mu\text{L}$	— $\mu\text{L}$
<b>BUFFER 10X</b>	1,0 $\mu\text{L}$	— $\mu\text{L}$
<b><i>HhaI</i> (10 U/<math>\mu\text{L}</math>)</b>	0,13 $\mu\text{L}$	— $\mu\text{L}$
<b>TOTAL</b>	<b>2 <math>\mu\text{L}</math></b>	<b>— <math>\mu\text{L}</math></b>
<b>Concentração final do MIX:</b> 1.3 unidade da endonuclease de restrição <i>HhaI</i> , Buffer 1X [Tris-acetato 33 mM (pH 7.9), acetato de magnésio 10 mM, acetato de potássio 66mM e 0,1 mg/ml de BSA]		

2. Distribuir 2,0  $\mu\text{L}$  do MIX de digestão + 8  $\mu\text{L}$  do DNA amplificado (produto de PCR da amplificação da região promotora do gene liberador de corticotrofina), em uma placa de PCR ou em microtubos de 0,2 mL, sempre sobre o gelinho
3. Incubar essa reação de digestão a 37°C por 3 horas (depende do fabricante da enzima *HhaI*) em aparelho Thermomixer 5436 (Eppendorf)

4. Para analisar os genótipos, preparar um gel de agarose a 3% em tampão de eletroforese TBE 1X (90 mM Tris base, 90 mM ácido bórico e 2 mM EDTA pH 8,0) e com 0,5 µg/mL de brometo de etídeo
5. Colocar o gel polimerizado na cuba de eletroforese coberto com tampão TBE 1X
6. Aplicar, no primeiro poço de cada fileira do gel, uma amostra padrão de tamanho de DNA de 100 pb acrescida de *loading buffer* na proporção 5 de amostra para 1 de *loading buffer* (5:1)
7. Aplicar, em cada poço do gel, 10 µL do produto de digestão acrescidos de 2 µL de *loading buffer* (proporção de 5:1)
8. Submeter à voltagem de 3 V/cm para a separação dos fragmentos
9. Detectar os fragmentos sob iluminação UV e registrar a imagem com uma câmera fotográfica digital
10. Genotipar os animais por meio da estimativa do tamanho dos fragmentos em comparação ao tamanho de DNA de 100 pb em cada gel

Os *primers* utilizados na PCR amplificam uma região de aproximadamente 600 pares de bases da região promotora do gene da corticotrofina, como ilustrado a seguir:

```
ACTTCCTTTGGCCTTTCTAGTAAGAGGCCAGTATGTTTTCACACTTGGGAAATCT
CATTCAAGAATTTTTGTCAATGGACAAGTCATAAGAAGCCCTTCCATTTTAGGGC
TCGTTGACGTCATCAAGGAGGCGATAAATATCTGTTGATATAATTGGATGTGAGA
TTCAGTGTTGAGATAGCAAAAATTCTGCCCTCGTTCCCGGGCAGGGCCCTATG
ATTTATGCAGGAGCAGAGGCAGCGCGCGCAATCCAGCTGTCAAGAGAGCGTCAGC
TTATTAGGCAAATGCTGCGTGGTTTCTGAAGAGGGTTCGACACTATAAAATCCCC
TTCCAGGCTCTGGTGTGGAGAACTCAGAGCCCACGTCCGTGGAGAGACAGAA
GAGGAAGAGAAGAGGAAAAGAACAGAGTGGGAAGAGAAAGGAGAAGGGAAGA
GAACCGCTGAAAAAAAAGCCCCAGAGACTTTCTCTGCAGAGAAGCGCTGCGCC
CCGCTCATCTGCAGAAGCACCTCGG
```

Obs. Em negrito, podem ser observados os sítios de restrição para a enzima *HhaI*. O SNP encontra-se em negrito e sublinhado.



## 8. ELETROFORESE DE ÁCIDOS NUCLÉICOS

*João José de Simoni Gouveia*

*Luciana Correia de Almeida Regitano*

A eletroforese é uma técnica utilizada para separar, identificar e purificar moléculas carregadas (como DNA e proteínas) (Maniatis, 1987). Esse método foi introduzido por Tiselius, no ano de 1937 (Brammer, 2002) e é simples, rápido e capaz de separar fragmentos que não são adequadamente separados por outras técnicas (Sambrook e Russel, 2002).

A técnica apresenta basicamente um sistema de suporte (gel de agarose, por exemplo), um tampão no qual está imerso o gel e os eletrodos nas extremidades da cuba onde estão contidos o tampão e o gel.

Sob a influência de um campo elétrico, moléculas carregadas e partículas migram em direção ao pólo oposto. A carga e a massa das moléculas fazem com que elas migrem em velocidades diferentes e, portanto, propiciam a separação destas (Westermeier, 2005). Como os ácidos nucleicos (DNA e RNA) possuem carga elétrica negativa (devido ao grupamento fosfato), eles sempre migrarão em direção ao pólo positivo. Então, o fator determinante da taxa de migração será a massa da molécula (quando se fala em ácidos nucleicos, a massa é diretamente proporcional ao tamanho da molécula).

Existem vários meios suporte que podem ser utilizados (papel filtro, membrana de celulose, gel de agarose, gel de poliacrilamida). No caso dos géis, a porosidade (que tem uma relação direta com a concentração de agarose ou de poliacrilamida) determina o poder de separação (Brammer, 2002).

Na escolha do tampão, o principal fator a ser considerado é sua capacidade tamponante. Os dois tampões mais utilizados na eletroforese de ácidos nucleicos são o TAE (Tris, Acetato e EDTA) e o TBE (Tris, Borato e EDTA). O TAE é mais utilizado que o TBE, porém é mais facilmente exaurido durante corridas longas ou com alta voltagem. O TBE, apesar da melhor capacidade tamponante, deve ser evitado quando se deseja purificar os ácidos nucleicos do gel (Ausubel, 1994).

Alguns fatores alteram a migração das moléculas através do gel, dentre eles podemos citar: concentração de agarose, conformação do DNA e intensidade da corrente.

A concentração de agarose atua de forma importante na eletroforese, pois ela determina a faixa de tamanho das moléculas de DNA que podem ser separadas. As relações entre concentração de agarose e resolução de moléculas lineares de DNA são apresentadas na Tabela 10.1. (adaptada de Maniatis et al., 1987).

**Tabela 8.1.** Limites de eficiência da separação de DNA em diferentes concentrações de agarose (Maniatis, 1987)

Concentração de agarose no gel (%)	Limites de eficiência para separação de moléculas de DNA (Kb)
0,3	60 – 5,0
0,6	20 – 1,0
0,7	10 – 0,8
0,9	7 – 0,5
1,2	6 – 0,4
1,5	4 – 0,2
2,0	3 – 0,1

Para observar ácidos nucléicos em gel de agarose, deve-se corá-los e submetê-los à luz ultravioleta. O corante mais comum é o brometo de etídeo, que se intercala entre as bases do DNA (Ausubel, 1994).

### 8.1. Protocolo para análise de produtos de amplificação e de fragmentos de digestão em gel de agarose

- **Gel 1%** → 0,3 g de agarose para 30 mL de gel, que deve ser preparado com tampão TBE 1X. Nesse gel, aplicar 5 µL de produto de amplificação + 1 µL de *loading buffer*.
- **Gel 4%** → 1,2 g de agarose de baixo ponto de fusão para 30 mL de gel, que deve ser preparado com tampão TBE 1X. Utilizado para analisar o resultado das digestões com enzimas de restrição. Aplicar todo o produto da digestão (13 µL) + 2,6 µL de *loading buffer*.

1. Pesar a agarose
2. Colocar a agarose em um erlenmeyer contendo o volume necessário (30 mL) de tampão TBE 1X
3. Aquecer no forno de microondas até iniciar a ebulição (aproximadamente 30 segundos em potência média para um gel de 30 mL). Agitar o frasco e retornar ao forno de microondas por mais alguns segundos
4. Aguardar a redução da temperatura para aproximadamente 60 °C e adicionar Brometo de Etídio ou Gel Red, na quantidade indicada para cada gel
5. Verter no molde previamente nivelado e colocar os pentes
6. Após a solidificação, colocar o gel na cuba de eletroforese e cobrir com tampão TBE 1X
7. Aplicar as amostras acrescidas de *loading buffer* nos poços do gel e submeter à corrente elétrica de acordo com o tamanho do gel

\* A eletroforese deve ser realizada em TBE 1X, à voltagem constante na faixa de 2 a 5 V/cm (considerando a distância entre os eletrodos). Ao invés de acrescentar o brometo de etídeo no gel, uma opção é, após a corrida, submergir o gel por 20 minutos em solução de brometo de etídio ou Gel red a 0,5 µg/mL e visualizá-lo sob luz ultra-violeta. Caso o fundo do gel esteja muito corado, pode-se lavar o excesso de brometo por submersão em TBE 1X por 5 a 10 minutos.

\* O brometo de etídio é um potente agente mutagênico. Utilizar luvas e máscara para preparar a solução estoque e para manipular os géis.

\* A luz ultra-violeta produz queimaduras severas. Utilizar máscara ou óculos de proteção adequados.

\*Protocolo para utilização de Gel Red na amostra: diluir 1 µL de Gel Red em 500 µL de água milliQ. Usar 1 µL deste diluído para cada 2 µL de DNA a serem aplicados no gel.

## 9. ELETROFORESE DE RNA EM GEL DESNATURANTE

*Adriana Mércia Guaratini Ibelli*

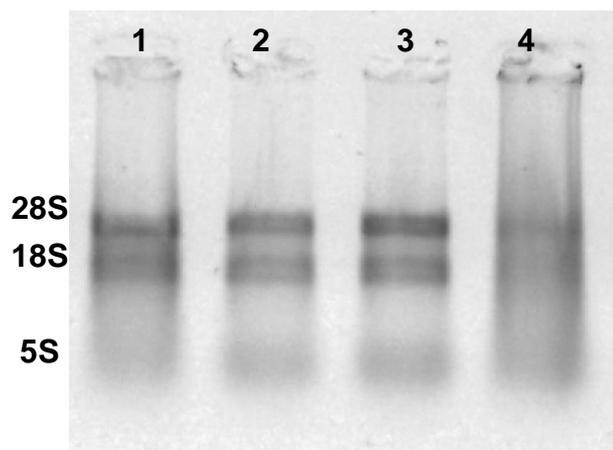
Para analisar moléculas de RNA podem ser utilizados géis de agarose e de poliacrilamida, desnaturantes ou não. As propriedades desta eletroforese são basicamente semelhantes às de eletroforese de DNA (Patel, 1994).

A qualidade das moléculas de RNA pode ser analisada em gel de agarose normal, feito com tampão (TBE, TAE) com água DEPC ou em gel desnaturante. Devido a interações intramoleculares, as moléculas de RNA podem dobrar-se, alterando a estrutura secundária e afetando a migração das moléculas no gel. O gel desnaturante soluciona esse problema, pois sob condições desnaturantes as pontes de hidrogênio são rompidas e o RNA migra como molécula de fita simples. Isto permite avaliar com acurácia a qualidade e o peso molecular do RNA (Sambrook e Russel, 2002; Maseka et al., 2005).

Dentre os desnaturantes existentes, o mais poderoso, além de caro e altamente tóxico, é o hidróxido de metil mercúrio. Porém, o mais freqüentemente utilizado é o formaldeído, mas pode haver variantes de géis utilizando tiocianato de guanidina, formamida e DMSO. A utilização de cada um deles apresenta vantagens e desvantagens de acordo com a escolha do tipo de gel, poliacrilamida ou agarose, níveis de toxicidade e potencial de desnaturação (Patel, 1994).

Os géis polimerizados com formaldeído não coram satisfatoriamente as amostras, de forma que o brometo de etidíó deve ser colocado em cada amostra individualmente, e não no gel, como comumente utilizado (Regitano, 2001).

Outro diferencial no gel desnaturante de agarose é a utilização de MOPS (3-N-morfolino ácido propanosulfônico). O MOPS é um excelente tampão para manutenção de sistemas biológicos em pH neutro. É muito utilizado em biologia molecular e em bioquímica.



**Figura 9.1.** Exemplo de separação eletroforética. Nas colunas 1 a 3 são mostradas as bandas 28S, 18S e 5S do RNA, indicando integridade das amostras. A coluna 4 exemplifica uma amostra de RNA degradada.

### 9.1. Protocolo de gel desnaturante de agarose 1% com formaldeído (adaptado de Alvares, 2001).

A vidraria utilizada para a preparação do gel, assim como a cuba de eletroforese devem estar previamente tratadas contra ação de RNases. Todas as soluções empregadas devem ser feitas com água DEPC.

1. Preparar o gel desnaturante de agarose a 1% com formaldeído:

Reagente	Quantidade	[Final]
Agarose	0,3 g	1%
Tampão 5X (1 x)	6,0 mL	1X
Formaldeído*	5,4 mL	2,2 M
Água DEPC	18,6 mL	
<b>TOTAL</b>	<b>30 mL</b>	

\* verificar se o pH do formaldeído é superior a 4

2. Preparar o tampão de corrida 5X, da seguinte maneira:

Reagente	Quantidade	[Final]
MOPS	20,6 g	100 mM
Acetato de sódio 0,05 M	800 mL	40 mM
EDTA 0,5 M pH 8,0	10 mL	5 mM

3. Pesar o MOPS
4. Dissolvê-lo em acetato de sódio 0,05 M. Ajustar o pH para 7,0 com NaOH 2N e completar o volume para um litro de água DEPC
5. Adicionar o EDTA 0,5 M pH 8,0. Guardar protegido da luz
6. Diluir este tampão para concentração final de 1X com água DEPC
7. Preparar o tampão da amostra:

<b>Reagente</b>	<b>Quantidade</b>	<b>[Final]</b>
Tampão de corrida 5X	300 $\mu$ L	1X
Formamida	750 $\mu$ L	50%
Formaldeído	150 $\mu$ L	10%
Azul de bromofenol	0,004 g	0,4%
Água DEPC	300 $\mu$ L	

8. Acrescentar 3  $\mu$ L do tampão da amostra para cada 1  $\mu$ L de RNA
9. Acrescentar 1  $\mu$ L de brometo de etídio (0,05  $\mu$ g/ $\mu$ L)
10. Incubar as amostras por 15 minutos à 65°C
11. Transferir para um recipiente contendo gelo
12. Centrifugar rapidamente para recolher o conteúdo no fundo do tubo
13. Manter em gelo até aplicar as amostras no gel

## 10. ELETROFORESE EM SISTEMA CAPILAR

*Gustavo Gasparin*

Durante décadas, a eletroforese em gel de poliacrilamida ou de agarose foi intensamente utilizada como uma das ferramentas mais importantes dos laboratórios de biotecnologia e bioquímica para a análise de macromoléculas. Nos últimos anos, porém, a eletroforese capilar tem apresentado vantagens em relação à técnica de eletroforese em placas. Para a análise simultânea de amostras, instrumentos de eletroforese capilar com arranjo de capilares são os mais utilizados. Existem aparelhos que possuem desde um único capilar, até 384 capilares, possibilitando a maximização do tempo necessário para as mais diferentes análises, desde detecção de resíduos de explosivos e drogas, até testes de paternidade.

A separação das macromoléculas é conduzida em tubos de dimensões de 15 a 100  $\mu\text{m}$  de diâmetro interno, e 36 a 100 cm de comprimento, preenchidos com um eletrólito. O uso do capilar fornece vantagens sobre as placas de gel devido às razões geométricas, em que há uma elevada relação entre a área superficial e o volume interno, permitindo a dissipação eficiente do calor gerado pela corrente elétrica, e possibilitando a aplicação de campos elétricos elevados (100 a 500 V/cm), o que resulta em separações de alta eficiência, alto poder de resolução, e reduzido tempo de análise.

Em geral, um aparelho de eletroforese capilar básico consiste de uma fonte de alta tensão, capilares (geralmente de sílica fundida), eletrodos (de platina, normalmente), e um detector apropriado. A fonte de alta tensão é aplicada nos eletrodos de platina situados nos reservatórios contendo uma solução eletrolítica apropriada. As extremidades do capilar são imersas nos reservatórios da solução, para completar o contato elétrico. As amostras normalmente são introduzidas no capilar por métodos eletrocinéticos, nos quais uma diferença de potencial é estabelecida entre o capilar e o recipiente que contém a amostra, durante um tempo pré-determinado.

O detector localiza-se em algum ponto do capilar, próximo ao reservatório de saída.

A eletroforese capilar em gel (CGE) é utilizada exclusivamente para separação de macromoléculas, tais como oligonucleotídeos, fragmentos de DNA, e

proteínas. Teve início com a aplicação nas separações de DNA por tamanho molecular, utilizando colunas preenchidas com gel, denominados géis químicos: um capilar é tratado com um reagente que estabiliza o gel junto à parede do capilar por ligações covalentes. Esse método foi descartado dado o grande número de problemas que surgiram, tais como aparecimento de bolhas (perda de condutividade), introdução limitada de amostra, retenção de fragmentos com alto peso molecular, e degradação do gel por hidrólise. Tais problemas levaram à formulação de novos sistemas, que foram denominados géis físicos.

Géis físicos são matrizes poliméricas hidrofílicas, dissolvidas em tampão apropriado, que não são ligados à parede do capilar, podendo assim ser substituídos a cada separação, o que permite o aproveitamento do mesmo capilar por centenas de vezes, sem perda de eficiência. A uma dada concentração de polímero, as fitas poliméricas individuais começam a interagir umas com as outras, formando uma estrutura em rede dentro do capilar, possibilitando a separação dos fragmentos de DNA. A concentração polimérica ótima depende do tamanho do DNA a ser separado.

### **10.1. Protocolo para análise de produtos de amplificação em sistema capilar**

Antes da injeção no capilar, as amostras são desnaturadas em presença de formamida, para evitar que a formação de estruturas secundárias afete a velocidade de migração. Um padrão de tamanhos conhecidos é aplicado junto com as amostras para permitir a estimativa do tamanho dos fragmentos analisados.

1. Para cada amostra, preparar o MIX com 8,85  $\mu\text{L}$  de formamida Hi-Di + 0,15  $\mu\text{L}$  de padrão interno de tamanho de fragmentos (GS 500 ROX)
2. Aplicar esses 9  $\mu\text{L}$  de MIX em um pocinho da placa de corrida
3. Aplicar 1  $\mu\text{L}$  de produto de PCR da sua amostra por pocinho
4. Desnaturar as amostras (já na placa de corrida) durante 5 minutos à 95°C
5. Colocar as amostras imediatamente no gelo após a desnaturação, permanecendo por 5 minutos
6. Preparar a injeção das amostras no computador, através do software *Data Collection*.

## 11. SEQÜENCIAMENTO

*Kerli Ninov*

*Lilian Giotto Zaros*

As duas técnicas mais importantes para o seqüenciamento de DNA são o método químico de degradação de bases desenvolvido por Allan Maxam e Walter Gilbert em 1977 e o método didesoxi ou terminação da cadeia de Fred Sanger e colaboradores em 1978.

Os dois métodos são baseados na produção de um conjunto de fitas simples de DNA que são separadas pelo princípio de eletroforese (Okubo et al., 1992). Se comparado ao seqüenciamento de Maxam e Gilbert, o método de terminação da cadeia de Sanger gera dados que são mais facilmente interpretados. Por isso, essa tem sido a técnica mais utilizada em Projetos Genoma (Sterky e Lundeberg, 2000).

### 11.1. Seqüenciamento didesoxi ou por terminação da cadeia

É baseado na incorporação de desoxinucleotídeos (dNTPs) e de didesoxinucleotídeos (ddNTPs) a uma cadeia de DNA em crescimento, tendo como molde o DNA de interesse.

Quando os ddNTPs são adicionados, a extensão da cadeia é interrompida pois esses didesoxinucleotídeos não apresentam um grupo hidroxila (OH) 3' necessário para a ligação do próximo desoxinucleotídeo (dNTP). Como esses ddNTPs são marcados, podem ser detectados e a seqüência dos nucleotídeos, identificada.

Esse método apresenta duas modalidades:

- Seqüenciamento Manual:

No método manual, a seqüência de bases do DNA é lida manualmente, após a exposição das bases identificadas em um filme de raio-X. Para isso são preparados 4 tubos de reação, cada um contendo o DNA molde, a DNA polimerase e um *primer*. Cada tubo recebe uma pequena quantidade de um dos ddNTP (dATP, ddTTP, ddCTP e ddGTP) junto com um dos dNTPs (dATP, dTTP, dCTP ou dGTP) marcados com fósforo radioativo (P<sup>32</sup>).

Em qualquer um dos 4 tubos serão produzidos vários comprimentos de cadeia, cada um correspondente ao ponto no qual o respectivo ddNTP desse tubo

foi incorporado e ocasionando o término do crescimento da cadeia. As amostras de cada tubo são submetidas à eletroforese e posterior exposição em filme de raio-X para que a posição das bases correspondentes possa ser determinada lendo-se ao longo das 4 colunas do gel ([www.dnai.org](http://www.dnai.org)).

- Seqüenciamento Automático:

O seqüenciamento automático utiliza seqüenciadores com eletroforese vertical em placa ou eletroforese em capilar onde cada ddNTP possui marcação fluorescente, ao contrário do método manual, em que esses apresentam marcação radioativa.

Para a geração de fragmentos durante a reação de seqüenciamento, os dNTPs são adicionados à nova fita e no momento da adição de um ddNTP, a extensão da cadeia é interrompida e conseqüentemente marcada com o último didesoxinucleotídeo incorporado. No final de vários ciclos tem-se várias cadeias de DNA de diferentes tamanhos terminadas com diferentes ddNTPs que posteriormente serão identificados (lidos em seqüenciador automático).

A reação de seqüenciamento ocorre em termociclador e é caracterizada por várias etapas ou ciclos em diferentes temperaturas, repetidos por várias vezes por um determinado período de tempo. Suas etapas são:

1. Desnaturação: é a primeira etapa da reação de seqüenciamento na qual a molécula de DNA é separada em fitas simples. Para ocorrer a desnaturação eleva-se a temperatura a 96°C durante 2 minutos.

2. Anelamento: caracterizada pela utilização de um único *primer* que irá se anelar à fita de DNA na região complementar à sua seqüência. Para ocorrer o anelamento diminui-se a temperatura para 50°C durante 15 segundos.

3. Extensão: após o anelamento do primer, ocorre a extensão do fragmento pela inserção dos dNTPs (dATP, dTTP, dCTP e dGTP) pela Taq DNA Polimerase. Como os ddNTPs estão em concentrações muito menores que os dNTPs, a sua inserção na cadeia é menos freqüente. Essa etapa ocorre a 60°C durante 4 minutos. Quando os ddNTPs são inseridos no lugar dos dNTPs, a extensão pára.

Após a reação de seqüenciamento, os produtos são submetidos à eletroforese capilar no seqüenciador automático. Os fragmentos marcados com fluorescência migram ordenadamente no interior dos capilares e à medida que são

excitados por um feixe de *laser*, emitem luz em diferentes comprimentos de onda que são detectados por uma câmara CCD. Em seguida essa informação é transmitida ao computador e processada para que ao final da corrida, os dados possam ser recuperados em formato de eletroferograma ou de texto (fasta), seguindo-se da análise de bioinformática.

### 11.2. Protocolo de reação de seqüenciamento

Para a realização de uma reação de seqüenciamento são necessários os seguintes componentes:

- DNA *template* na concentração de 100 a 150 ng/ $\mu$ L para DNA plasmidial e de 50 a 80 ng/ $\mu$ L para produto de PCR;
- *Primer*: oligonucleotídeo complementar à determinada região do fragmento de interesse (Ex: *primer* universal M13 R ou F);
- dNTPs: desoxinucleotídeos (dATP, dCTP, dGTP e dTTP);
- ddNTPs: didesoxinucleotídeos (ddATP, ddCTP, ddGTP e ddTTP);
- Tampão contendo magnésio ( $Mg^{+2}$ ) e Tris-HCl, que atuam como co-fator enzimático e como tampão para manutenção do pH, respectivamente;
- Enzima Taq DNA Polimerase

Obs.: A enzima, os dNTPs e os ddNTPs estão contidos no *Kit ABI PRISM® Big Dye terminator v. 3.1 cycle sequencing*.

1. Retirar as amostras e os reagentes do freezer e mantê-los no gelo. Todos devem estar totalmente descongelados.
2. Preparar o seguinte *mix* em microtubo de 1,5 mL:

<b>Mix</b>	<b>1X</b>
Água milliQ	2 $\mu$ L
Buffer Big Dye	2 $\mu$ L
Primer (2 pmol)	2 $\mu$ L
Big dye	2 $\mu$ L
<b>TOTAL</b>	<b>8 <math>\mu</math>L</b>

3. Distribuir 8  $\mu$ L do MIX de seqüenciamento por tubo ou por poço (de placa de PCR), sempre sobre um gelinho.

4. Adicionar a quantidade desejada de produto de PCR já purificado (segundo Tabela 12.1). No caso, distribuir 2  $\mu\text{L}$  (32 ng/ $\mu\text{L}$ ) do produto de PCR por tubo ou por poço.

10. Colocar as amostras em um termociclador, e escolher o programa de termociclagem descrito abaixo. Cobrir a placa com um tapete de borracha e levá-la ao termociclador utilizando o programa adequado, como por exemplo:

→ Pré-incubação: 94 °C por 2 minutos

→ Amplificação:

25 ciclos de

{	96 °C por 20 segundos
	54 °C por 10 segundos
	60 °C por 04 minutos

→ Resfriamento: 4 °C por  $\infty$ .

**Tabela 11.1.** Quantidade de amostra utilizada em uma reação de seqüenciamento.:

Template	Quantity
PCR product:	
100–200 bp	1–3 ng
200–500 bp	3–10 ng
500–1000 bp	5–20 ng
1000–2000 bp	10–40 ng
>2000 bp	40–100 ng
Single-stranded	50–100 ng
Double-stranded	200–500 ng
Cosmid, BAC	0.5–1.0 $\mu\text{g}$
Bacterial genomic DNA	2–3 $\mu\text{g}$

### 11.3. Purificação da reação de seqüenciamento

Após a reação de seqüenciamento é necessário purificar o material para que os ddNTPs, os dNTPs, os *primers* e a enzima não incorporados não interfiram na leitura no seqüenciador.

1. Adicionar 40  $\mu\text{L}$  de isopropanol 65% à temperatura ambiente (TA).
2. Vortexar alguns segundos e incubar no escuro à TA por 15 minutos.
3. Centrifugar por 25 minutos a 16000 xg à TA.
4. Descartar o sobrenadante por inversão e/ou dar um pulso (até 40 xg) em centrífuga com a placa invertida sobre um papel absorvente.

5. Adicionar 200  $\mu$ L de etanol 60% e centrifugar por 5 minutos a 16000 xg à TA.
6. Descartar o sobrenadante por inversão e dar um pulso (até 90 xg) em centrífuga com a placa invertida sobre um papel absorvente (Repetir 1X os passos 5 e 6).
7. Secar no escuro por 1 hora.
8. Caso a aplicação em seqüenciador automático não seja feita imediatamente, selar a placa com adesivo, embalar em papel alumínio e armazenar em freezer a  $-20^{\circ}\text{C}$ .

#### 11.4. Aplicação em ABI 3100

1. Ressuspender as amostras com 10  $\mu$ L de formamida Hi-Di.
2. Desnaturar as amostras em termociclador por 5 minutos a  $95^{\circ}\text{C}$ , seguido por 5 minutos a  $4^{\circ}\text{C}$ .
3. Verificar a validade e a quantidade do polímero (POP6) e do tampão (EDTA 10X). O polímero é válido por cinco dias quando já está no equipamento e o tampão tem validade de 48 horas ou de duas placas corridas. É recomendável a cada 15 dias retirar o capilar, desmontar o aparelho e realizar uma limpeza em seus componentes.
4. Montar a planilha de corrida no software **Data Collection** em *Plate View*
5. Nesse momento, é aberta uma nova planilha:
  - Sample name
  - Dye Set (Z)
  - Mobility File (DT3100POP6- BDV3-V1.mob)
  - Project name (3100Project)
  - Run Module 1 (StdSeq50\_POP6 Default Module)
  - Analysis Module 1 (BC\_3100\_APOP6SR\_SeqOffFtOff.saz)
  - OK
6. Montar a placa nos acessórios próprios do seqüenciador.
7. Colocar a placa no seqüenciador verificando na planilha "*Run View*" se ela está corretamente instalada.
8. Clicar em "*Status View*" e verificar voltagem, *laser* e temperatura.
9. Caso esteja tudo OK, clicar em "*Run*".

10. Aguardar para verificar possíveis erros durante o início da corrida.
11. Após 40 minutos (aproximadamente), é possível o acompanhamento da corrida pela visualização da imagem de cada capilar.
12. Ao término da corrida, analisar os dados pelo *software* **Sequence Analysis**.
13. Adicionar a corrida desejada através do “Add Files” e clicar em “Start”.
14. Aguardar a execução da tarefa.
15. Abrir os eletroferogramas e avaliar os resultados obtidos.

### **11.5. Análise das reações de seqüenciamento com os programas phredPhrap/Consed**

*Gisele B. Veneroni*

*Felipe Rodrigues da Silva*

Após os produtos das reações de sequenciamento terem sido analisados em um seqüenciador ABI Prism 3100 Avant (Applied Biosystems), os eletroferogramas gerados podem ser submetidos ao programa de base calling Phred, que reconhece dados de cromatogramas e identifica a sequência de DNA gerada, atribuindo um valor de qualidade a cada nucleotídeo identificado. Quanto maior o valor de qualidade de Phred, menor a probabilidade de ter ocorrido um erro na designação do nucleotídeo. Em seguida tais seqüências são submetidas ao programa de montagem Phrap (*Phragment Assembly Program*), onde as seqüências são agrupadas e organizadas em *contigs*. A visualização das seqüências geradas é realizada através do programa Consed, onde é possível a observação dos SNPs (PROSDOCIMI *et al.*, 2002). A presença de 2 picos no eletroferograma é interpretada como um polimorfismo (SNP). Esta metodologia permite identificação de polimorfismos somente em indivíduos heterozigotos para o SNP.

### **11.6. Protocolo de utilização dos programas phredPhrap/Consed**

1. Dentro da pasta em que estão instalados os programas phredPhrap/Consed, criar uma pasta com a identificação do animal em análise;
2. Para isso, abra o terminal e, após digitar o caminho referente a pasta onde estão instalados os programas phredPhrap/Consed, exemplo: “cd

- `/usr/local/genome"` , pode-se utilizar o comando `mkdir` para criar a pasta de identificação do animal em análise. Ex. `"mkdir animal_955"`;
3. Dentro da pasta de identificação do animal criar mais 3 pastas cujos nomes impreterivelmente são: `edit_dir`, `chromat_dir`, `phd_dir`. Para tal, pode-se utilizar o mesmo procedimento descrito no passo 2;
  4. Colocar o eletroferograma do animal analisado, gerado pelo sequenciador , dentro da pasta `chromat_dir`, localizada dentro da pasta de identificação do animal. O comando `cp` pode ser utilizado. No sistema linux, se você quiser copiar um arquivo que está em outra pasta para o diretório atual, coloque a localização completa do arquivo e em seguida `"/` (que representa o diretório atual). Ex. `"cp /mnt/cdrom/animal_955.abi ./"`;
  5. Digitar o caminho correspondente a localização da pasta `edit_dir` que se encontra dentro da pasta de identificação do animal que se deseja analisar e, dar `"enter"`;
  6. Digitar o comando de execução do programa `phredPhrap`. Ex. `"phredPhrap"` e, o programa será executado. Observação: O comando pode diferir dependendo de quem instalou o programa;
  7. Em seguida digitar o comando de execução do programa `consed`. Ex. `"consed"` e, o programa será executado. Observação: O comando pode diferir dependendo de quem instalou o programa;
  8. Uma janela cinza irá abrir. No quadro intitulado `" Ace Files"` clicar duas vezes em cima do arquivo criado, geralmente o título desse arquivo contém: hora de execução do programa, data de execução do programa, identificação do animal analisado. `fasta.screen.ace`.
  9. O título descrito no item 8 irá aparecer em um outro quadro de título `" Contig list"`, clicar 2 vezes em cima do título descrito no item 8 ;
  10. Uma janela de fundo preto irá abrir, onde estarão expostos: localização dos nucleotídeos no fragmento sequenciado, a sequência consenso (quando mais de um eletroferograma é analisado, o programa cria uma sequência consenso de todos os eletroferogramas avaliados) e a sequência do fragmento sequenciado;

11. Se a sequência de nucleotídeos aparecer com fundo branco, significa que os nucleotídeos estão com ótima qualidade, a medida que o fundo dos nucleotídeos vai escurecendo em direção ao cinza escuro significa que a qualidade do pico gerado pelo sequenciamento está ruim. Outra forma de saber a qualidade do pico no eletroferograma é clicando apenas uma vez em cima do nucleotídeo. Embaixo, em uma barra cinza aparecerá escrito: quality= “ “ e, o número que aparecer aí é a qualidade phred atribuída àquele nucleotídeo;
12. Clicando com os dois botões do mouse ao mesmo tempo em cima dos nucleotídeos (ou com o botão do meio do mouse), o eletroferograma é aberto e fica possível observar os picos gerados pelo sequenciamento;
13. Cada nucleotídeo possuirá um pico no eletroferograma;
14. Se dois picos se sobreporem é indício de presença de SNP nesse ponto.

## 12. SÍNTESE DE DNA COMPLEMENTAR (cDNA)

*Adriana Mércia Guaratini Ibelli*

*Luciana Correia de Almeida Regitano*

A conversão enzimática de mRNA para cDNA foi descoberta na década de 1970 por Kacia, Ross, Verma e colaboradores (Sambrook e Russel, 2002).

A síntese da primeira fita de cDNA é realizada por uma enzima transcriptase reversa dependente de RNA, geralmente isolada de vírus, que usa amostras de RNA mensageiro (poli A) ou de RNA total como molde. Atualmente, diversas espécies de transcriptases reversas são comercialmente utilizadas, como Mo-MLV, ALV e *SuperScript<sup>®</sup> RT II* (Manual do usuário, 2003).

Para que essa conversão seja feita, são necessários:

- dNTPs: altas concentrações de dNTPs são importantes para a obtenção de boa eficiência na síntese do cDNA. Em baixas concentrações (< 50  $\mu$ M) há decréscimo significativo na produção da primeira fita. Geralmente, são utilizados cerca de 200 a 250  $\mu$ M de cada dNTP (Maniatis, 1987).

- Oligonucleotídeo: o mais utilizado é o oligo dT (12 a 18 bases T), que sintetiza a primeira fita a partir de RNAs de cadeia poli A (Manual do usuário). Podem ser usados também *primers* hexâmeros randômicos ou primer gene específico (GSP).

- Tampão: a presença de tampão adequado é importante, pois a alteração do pH acarreta a diminuição da eficiência de incorporação e de produção de transcritos. O pH ótimo é geralmente 8,3. Já foi observado que pequenas alterações de pH, de até 0,5 resultam em diminuição de até 5X na produção do cDNA (Maniatis, 1987).

-  $MgCl_2$ : os cátions divalentes são importantes na transcrição reversa, pois auxiliam na produção de transcritos completos em concentrações de 6 a 10 mM. Quando fragmentos muito longos serão transcritos, podem ser utilizados cátions monovalentes, que auxiliam na manutenção da eficiência da transcrição, dentre eles, o mais indicado é o potássio.

- *RNAse OUT<sup>®</sup>* (Invitrogen): degrada RNAses e melhora a síntese do cDNA (Manual do usuário).

- DTT ou ditiotreitól: é um agente redutor que tem como principal função evitar a oxidação de enzimas e proteínas (Roche, 2004).

- RNase H: é utilizada no final da reação para remover híbridos RNA:cDNA com a função de aumentar a sensibilidade e a eficiência da PCR. A presença de RNase H durante a síntese do cDNA degrada o mRNA, causando um déficit na produção dessa molécula. Por isso, foram desenvolvidas transcriptases reversas que apresentam atividade reduzida de RNase H durante a produção da primeira fita e que são adicionadas apenas no final da reação, para remoção de moléculas híbridas ( Manual do usuário, 2003).

É importante considerar que para que este procedimento tenha sucesso, o RNA das amostras deve ser intacto e sua quantificação, precisa, evitando que variações na qualidade e quantidade do RNA molde afete o resultado das análises a serem feitas com o cDNA , como por exemplo, em arranjos de DNA ou em PCR quantitativo.

### 12.1. Protocolo de síntese de cDNA

1. Descongelar e homogeneizar brevemente cada componente abaixo e preparar o *mix* RNA/*primer*.

Componente	Para 1 amostra
Até 5 µg de RNA total	Máximo de 8 µL*
dNTP <i>mix</i> 10 mM	1 µL
Oligo dT 0,5 µg/µL	1 µL
Água DEPC (do kit)	Completar para 10 µL*

\*O volume desse mix deve ser de 10 µL. Assim, o volume máximo de RNA a ser utilizado é 8 µL contendo de 1 ng a 5 µg de RNA total. Se for utilizado 8 µL, a água DEPC do kit não é adicionada ao mix.

2. Incubar a 65°C por 5 minutos.
3. Esfriar em gelo por 1 minuto.
4. Preparar o *mix* de reação adicionando cada componente a partir da primeira linha da tabela. Esse *mix* pode ser preparado em um tubo separado, junto com a preparação do *mix* anterior. Se isso for feito, mantê-lo sempre em gelo:

<b>Componente</b>	<b>Para 1 amostra</b>
Buffer RT 10X	2 $\mu$ L
MgCl <sub>2</sub> 25 mM	4 $\mu$ L
DTT 0,1 M	2 $\mu$ L
RNase Out 40 U/ $\mu$ L	1 $\mu$ L
<b>TOTAL</b>	<b>9 <math>\mu</math>L</b>

5. Adicionar os 9  $\mu$ L do *mix* acima em cada tubo contendo o RNA/*primer*. Dar um *spin*.

6. Incubar a 42°C por 2 minutos.

7. Adicionar 1  $\mu$ L (50 U) da enzima *SuperScript RTII* em cada tubo.

*Obs.: Os tubos não podem ser colocados em gelo para adição da enzima RTII. A temperatura deve continuar de 42 °C;*

8. Incubar a 42°C por 50 minutos.

9. Terminar a reação a 70°C por 15 minutos.

10. Esfriar no gelo e dar um *spin*.

11. Adicionar 1  $\mu$ L de RNase H e incubar por 20 minutos a 37°C.

12. Estocar em freezer a – 20°C.

### 13. DELINEAMENTO DE PRIMERS

*Gisele Batista Veneroni*

*Adriana Mércia Guaratini Ibelli*

A PCR é uma técnica muito utilizada na Biologia Molecular para inúmeras finalidades. Condições de reação ótimas são provenientes de uma boa otimização dos componentes e, para essa otimização, a obtenção de um bom par de iniciadores (primers) é primordial.

Segundo Dieffenbach et al. (1993) o desenho de primer visa principalmente 2 fatores: especificidade e eficiência de amplificação. Especificidade no que diz respeito à amplificação de uma única região do genoma (isto é, a região de interesse de estudo). A eficiência se refere a capacidade do par de primers propiciar a formação do dobro de produto de amplificação a cada ciclo de PCR.

A especificidade é afetada principalmente pelo tamanho e temperatura de anelamento do primer. Primers que contenham entre 18 e 24 bases tendem a ser específicos, anelando apenas na região a ser amplificada, e proporcionam temperaturas de anelamento pouco variáveis da temperatura de *melting* (temperatura na qual metade das fitas de DNA está na forma de fitas simples e a outra metade na forma de dupla hélice) (Dieffenbach et al. 1993). Primers mais curtos podem aumentar a chance de anelamento em outras regiões do genoma provocando a amplificação de produtos não específicos. Estes últimos podem ser utilizados apenas em alguns estudos específicos como mapeamentos de genomas simples ou bibliotecas subtrativas.

A temperatura de *melting* é influenciada pela composição do DNA, de modo que um aumento do conteúdo de G/C no DNA provoca aumento na temperatura de *melting*, pois as ligações GC precisam de mais energia para serem quebradas por possuírem 3 pontes de hidrogênio unindo-as, enquanto que ligações AT são unidas por apenas 2 pontes de hidrogênio. A temperatura de anelamento é a temperatura na qual os primers se pareiam ao DNA molde. Geralmente a temperatura de anelamento de um primer é 5 graus mais baixa do que a temperatura de *melting* (Bock 1997). A temperatura de anelamento é específica de cada primer, pois depende da quantidade de bases, principalmente G/C, que ele possui. O ideal é que a porcentagem de conteúdo GC de um primer fique entre 45 e 55%. Temperaturas

de anelamento abaixo da ideal tendem a causar menor estringência, ou seja, o primer pode se anelar em qualquer lugar. Em contrapartida, temperaturas de anelamento muito altas tendem a provocar maior estringência e o primer pode nem se anelar. Por isso a determinação da temperatura de anelamento adequada de cada primer é primordial (Dieffenbach et al. 1993). Além disso, o par de primers escolhido para amplificação de uma região deve possuir temperaturas de anelamento similares.

A extremidade 3' de um primer é um ponto crítico para constituição de condições ideais de PCR, pois é essa extremidade que se liga primeiro a fita de DNA para dar início a replicação. Estudos demonstram que se essa extremidade possuir um nucleotídeo G ou C ou CG ou GC a reação será mais eficiente, pois possuem ligações mais estáveis (Kwok 1994). No entanto devem ser evitadas seqüências de 3 ou mais G ou C, pois podem promover pareamento espúrio em regiões ricas em G ou C (devido a estabilidade do emparelhamento).

Pares de primer não devem ser autocomplementares (podem resultar na formação de grampos ou *self-dimers*) e nem complementares entre si (podendo formar dímeros), principalmente na extremidade 3'. Esse tipo de ocorrência provoca uma competição do anelamento entre primers com o anelamento primer/DNA molde, diminuindo a eficiência da reação (Bock 1990).

Quando não se consegue evitar a escolha de primers que formem grampo ou dímeros, o ideal é escolher o par de primers cuja temperatura para formação de tal estrutura não seja semelhante à temperatura de anelamento. Outros fatores também devem ser observados quando ocorre esse tipo de situação como os valores de  $\Delta G$  (energia livre de Gibbs) e  $\Delta H$  (entalpia), pois se o  $\Delta G$  for menor que zero, a reação é termodinamicamente favorável e é considerada espontânea, isto é, a formação de grampos e dímeros pode ocorrer espontaneamente. Assim, deve-se optar por primers que possuam  $\Delta G$  próximo de zero ou positivos.  $\Delta H$  se refere a energia absorvida ou liberada em pressão constante. Se o  $\Delta H$  for menor que zero significa que a reação é exotérmica, não precisa de energia pra acontecer, sendo assim a escolha de primers que possuam  $\Delta H$  maior que zero é mais adequada.

Existem muitos programas e páginas web que podem nos auxiliar no desenho de primers, dentre eles podemos destacar Primer Premier (disponível em <http://www.premierbiosoft.com/primerdesign/index.html>), Primer Blast (disponível em

[http://www.ncbi.nlm.nih.gov/tools/primer-blast/index.cgi?LINK\\_LOC=BlastHomeAd](http://www.ncbi.nlm.nih.gov/tools/primer-blast/index.cgi?LINK_LOC=BlastHomeAd)) e Primer3 e Primer3 Plus (disponíveis em <http://frodo.wi.mit.edu/>). Geralmente tais interfaces levam em consideração na construção do primer alguns fatores primordiais como temperatura de melting, tamanho, conteúdo GC, estabilidade 3', probabilidade de anelamento ou amplificação de sequências não desejáveis, formação de dímeros entre outros.

Após o desenho de um primer é necessária a verificação da qualidade e especificidade do primer. Isto pode ser realizado pelo próprio programa ou interface web (como no caso do Primer Blast) ou por programas específicos. Um bom site para verificação de qualidade e especificidade é o <http://www.idtdna.com/analyzer/Applications/OligoAnalyzer/>. Nesse site pode-se avaliar se o primer forma grampos, se é autocomplementar ou se ele se liga a outros locais do genoma que não são de interesse. A especificidade de um par de primers pode ser verificada também usando a ferramenta BLAST do banco de dados do ncbi (<http://blast.ncbi.nlm.nih.gov/Blast.cgi>), usando a opção “nucleotide BLAST”. Os resultados do BLAST são apresentados de acordo com dois parâmetros: o valor do escore (Score bits) e o valor E (e-value). O valor de escore depende do tamanho do alinhamento, do número de matches/mismatches/gaps e da matriz de comparação de seqüências utilizada e é normalizado através de variáveis estatísticas. Quanto maior o score melhor. No que se refere ao e-value quanto menor o valor, melhor o alinhamento, de forma que um valor de E igual a zero significa que não há chance de que um alinhamento entre as duas seqüências tenha ocorrido por mero acaso. Porém, no caso de pequenas seqüências, como as em questão, esses valores são apenas indicativos.

### **13.1. Desenho de primers para análise de expressão gênica**

O desenho de primers para quantificação de mRNA apresenta muitas semelhanças quando comparado com os desenhos de primers tradicionais. Avaliar critérios como proporção de conteúdo GC, temperatura de anelamento, formação de dímeros de primers, de heterodímeros, de *hairpins*, valores de  $\Delta G$  (energia livre de Gibbs) e  $\Delta H$  (entalpia) são essenciais para o desenho de um bom par de primers. No entanto, outras características que devem ser consideradas para esta aplicação em particular estão descritas abaixo (TaqMan® EZ RT-PCR Kit, 2002; Dorak, 2009).

1. Os primers devem conter aproximadamente 15 a 30 bases, sendo que tamanhos de 18 a 24 bases são ideais.

2. O conteúdo GC deve estar idealmente entre 50 a 80 %. Se um conteúdo GC maior for inevitável, cosolventes como glicerol ou DMSO podem ser necessários na reação de PCR.

3. O amplicon deve conter entre 50 a 150 bases. Amplicons menores dão uma consistência maior, pois são mais eficientes nas condições da reação.

4. Nos últimos 5 nucleotídeos da extremidade 3' dos primers, deve-se evitar a presença de mais de 2 ou 3 Gs ou Cs. Isso ajuda a diminuir a estabilidade da extremidade 3'.

5. Indica-se que a temperatura de anelamento de cada primer seja de aproximadamente 60 °C. Se a plataforma Taqman está sendo utilizada, é necessário o desenho de sondas, internas ao produto que será amplificado, que deve ter a temperatura de anelamento 10 °C maior que a temperatura dos primers (TaqMan® EZ RT-PCR Kit, 2002).

6. Como o interesse é a amplificação de produtos de mRNA expressos, é indicado que os primers desenhados que não amplifiquem regiões de DNA genômico, interferindo nas análises. Para isso, é necessário ter conhecimento do mRNA de interesse, em que regiões estão contidos os exons, para assim, desenhar primers que estejam dentro de exons diferentes. Por exemplo, o primer F está desenhado no exon 1 de um gene, e o primer R, está localizado no exon 2. Uma outra alternativa, é desenhar os primers nas junções exon-exon, o que impede totalmente a amplificação a partir de DNA. Nesse caso, o primer F é parcialmente desenhado no exon 1 e no exon 2, e o primer R no exon 2 ou no 3.

As seqüências de interesse podem ser encontradas nas bases de dados: Genbank do NCBI (<http://www.ncbi.nlm.nih.gov/sites/entrez?db=nucleotide>), que permite o acesso às seqüências de mRNA e ENSEMBL (<http://www.ensembl.org/index.html>), que disponibiliza informações mais acuradas do transcrito, nesse caso, encontrada como cDNA. Nessa última, são encontradas as informações exatas como o número de exons, onde começam e onde terminam cada um, possibilitando dessa maneira o desenho nas junções exon-exon. É sempre conveniente “blaster” as seqüências obtidas nas duas bases de dados, para verificar a consistência entre elas.

Após a obtenção das seqüências nas bases de dados, podem ser usadas plataformas de desenho de primers específicas como a disponível no site da IDT DNA (<http://www.idtdna.com/Scitools/Applications/RealTimePCR/>), o programa Primer 3 Plus, com a opção para qPCR (<http://www.bioinformatics.nl/cgi-bin/primer3plus/primer3plus.cgi>) e o Primer-BLAST ([http://www.ncbi.nlm.nih.gov/tools/primer-blast/index.cgi?LINK\\_LOC=BlastHomeAd](http://www.ncbi.nlm.nih.gov/tools/primer-blast/index.cgi?LINK_LOC=BlastHomeAd)).

Para análise da qualidade dos primers desenhados, podem ser utilizados: o software online Netprimer (<http://www.premierbiosoft.com/netprimer/index.html>), o Oligoanalyzer (<http://www.idtdna.com/analyzer/Applications/OligoAnalyzer/>) e o Beacon design (<http://www.premierbiosoft.com/qpcr/index.html>), que realiza a análise de qualidade de primers e presença de estruturas secundárias dos dois primers (F e R) desenhados.

### 13.2. Aula Prática - Desenho de primer

**1. A sequência abaixo contém os exons 5 e 6 do gene IGFBP3 bovino. Os nucleotídeos em preto correspondem aos introns e os nucleotídeos grifados, em destaque, aos exons.**

**a-** Desenhe um par de primers, utilizando o primer3 plus, que amplifique o exon 5 do gene IGFBP3 bovino formando um amplicon de no máximo 400 pb.

**b-** Verifique a qualidade do primer desenhado (<http://www.idtdna.com/analyzer/Applications/OligoAnalyzer/>) e faça um blast (<http://blast.ncbi.nlm.nih.gov/Blast.cgi>) para verificar especificidade, observando o score e E-value. Escolha o melhor par de primers.

GGACAACCTCCTACCACAGTACCTACAGAATTAATCACAGAACAGCACAGCCATGTTCAAAGTGTT  
GGTAGTAGTGACCTCAATCAGAGGGCCCATCCTGTTTCAGTGATCACTGAACAGATCTCGTTGGC  
CTCTTTGTATATTCTATTAATCACTTATTTAGCAGACTTACCCTGAGAAGGTACCACGTACACAGC  
TCAGGGCGTACGCTGAAAGACCCAGATGTGACAAGCACAGGGAAGATGGGGCATTGTCATGCCT  
CCCCATTTCAGCAGATAGACTTTGCAGGGCCAACAGGGTCCAGTGTCCACCCCTGTGAGACCCT  
GAGGGCACAAAACCACAGGTCCCCTCGTTTTGGGGGGCACTCACATCCACCCATCACAGCCCC  
AGACACACGCTGTGCACACAAGGCCTTAATCAGTTGTGCTTTCTGAGGGTAGCAGAGACCACACT  
GGGAAGCCATCCTGGGTGCTGACCACCTGCCCATGATAGGACCCTCAAGGCAGGGCCTCCTC  
AGGGCTCATTTAATCTGTGCTCTGCCTGCAGTGCCGCCCTTCCAAGGGCAGGAAGCGGGGTTTC

**TGCTGGTGTGTGGATAAGTACGGGCAGCCCCTCCCGGGCTTCGACGTGAAGGGGAAAGGGGAC**  
**GTGCACTGCTACAGCATGGAGAGCAAGTAGACTGTGGCCAC**GTAAAGTGCTGGGGACTGGCAGAT  
CCACAGGGGCACAGGACCCGGCCCTGGTGGGAGGGGACACCTGTAGATCCCAGGGGTGGTGG  
CCTACACACAAGACAGAGCTCTCAATCTGCCCGGACAAGAGCAGGAGGGAGGGTCTGTGGCCA  
GGTTGTCTGAGACACACCTCAC

**c-** Desenhe um par de primers para quantificação de mRNA, utilizando o primer3 plus, que amplifique parte do mRNA do gene IGFBP3 bovino formando um amplicon de no máximo 200 pb.

**d-** Verifique a qualidade do primer desenhado no software net primer (<http://www.premierbiosoft.com/netprimer/index.html>) e faça um blast (<http://blast.ncbi.nlm.nih.gov/Blast.cgi>) para verificar especificidade, observando o score e E-value. Escolha o melhor par de primers.

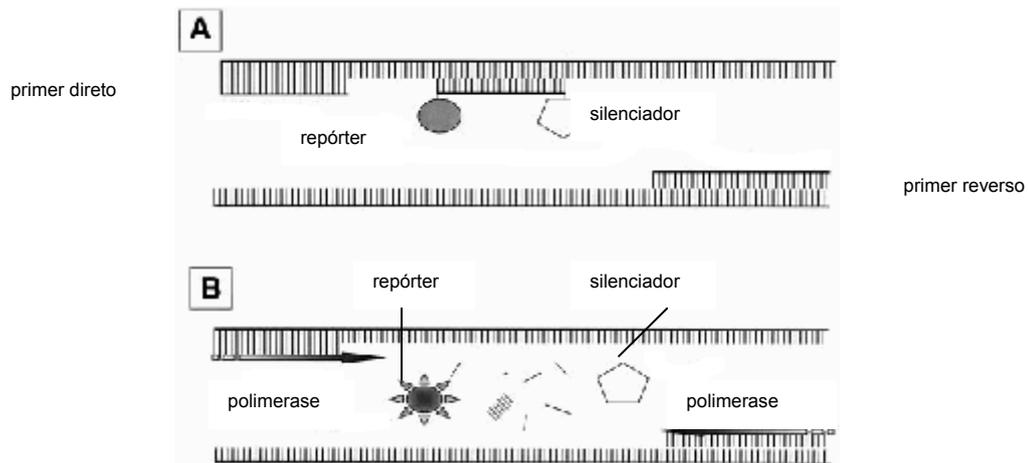
## 14. PCR QUANTITATIVA EM TEMPO REAL

*Helena Javiel Alves*

O método de PCR quantitativa em tempo real é muito sensível e pode ser usado quando é necessário quantificar a expressão de RNAs mensageiros em baixos níveis. Esse procedimento permite a detecção direta dos produtos da PCR durante a fase exponencial da reação, combinando amplificação e detecção em um só passo.

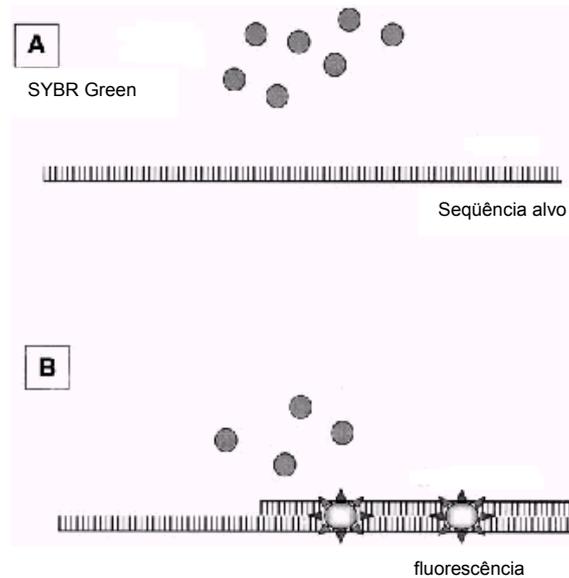
O princípio da técnica de PCR em tempo real pode se utilizar de duas descobertas importantes: a primeira é a atividade exonuclease 5'→3' da enzima *Taq* DNA polimerase e a segunda é a construção de sondas de oligonucleotídeos marcadas duplamente. Essas sondas são baseadas no princípio FRET (do inglês, *Fluorescence Resonance Energy Transfer*) e emitem sinal de fluorescência somente quando clivadas. Com o objetivo de detectar o sinal de fluorescência durante as reações foram desenvolvidas várias técnicas. Em PCR em tempo real utilizando sondas *TaqMan* (Figura 15.1.), a sonda, específica para o gene de interesse, é marcada duplamente com um fluoróforo repórter em uma extremidade e um fluoróforo “silenciador” na outra. Na forma íntegra, a transferência de energia fluorescente ocorre de forma que a emissão pelo repórter é absorvida pelo silenciador. Quando ocorre a degradação da sonda pela atividade exonuclease da enzima *Taq* DNA polimerase, durante a PCR, os fluoróforos repórter e silenciador são separados e a emissão de fluorescência do repórter não será mais absorvida pelo silenciador resultando em aumento de emissão de fluorescência pelo repórter que será detectada e quantificada.

Mais recentemente, outros sistemas sofisticados têm sido desenvolvidos como, por exemplo, *molecular beacons*, *scorpions* e sondas para hibridização. Esses sistemas são baseados no princípio FRET, porém sem a necessidade de hidrólise por atividade exonuclease.



**Figura 14.1.** Sistema *TaqMan*. (A) Após a desnaturação, os *primers* e a sonda anelam-se à seqüência alvo. A emissão de fluorescência não ocorre devido à proximidade entre o repórter e o silenciador. (B) Durante a fase de extensão, a sonda é clivada (5'→3') pela atividade exonuclease da enzima *Taq* DNA polimerase. Então, repórter e silenciador são separados, permitindo a emissão de fluorescência pelo repórter.

Outra forma de detecção é a utilização de corantes como *SYBR Green* (Figura 15.2.), que se liga ao DNA dupla fita emitindo fluorescência. O uso desse método de detecção tem sido bastante empregado em ensaios de PCR quantitativa em tempo real, pois tem a vantagem de ser mais barato em relação à construção de sondas marcadas. Um aspecto importante no uso de *SYBR Green* é que os *primers* utilizados na PCR devem ser desenhados cuidadosamente para evitar a amplificação de produtos inespecíficos e a formação de dímeros, uma vez que esse corante se liga a qualquer DNA dupla-fita.



**Figura 14.2.** Sistema *SYBR Green*. (A) O corante livre na solução não emite fluorescência. (B) Quando o corante se liga ao DNA fita-dupla ocorre a emissão de fluorescência.

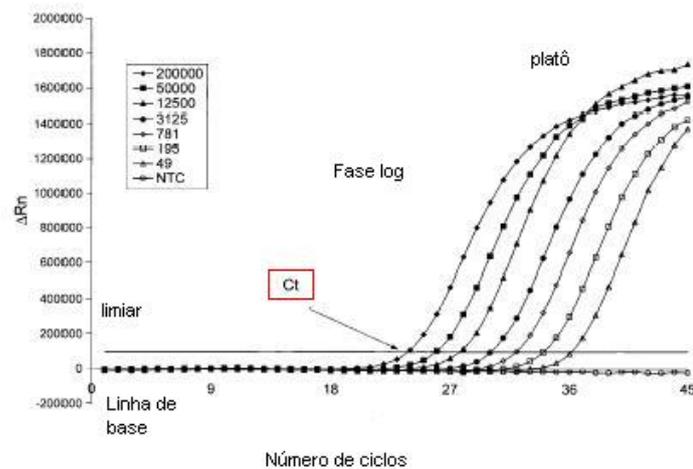
Em qualquer uma das químicas apresentadas, o aumento na emissão de fluorescência pode ser lido por um detector em tempo real, durante a PCR e é uma consequência direta da amplificação da seqüência de DNA de interesse. Um programa de computador calcula um  $\Delta Rn$  por meio da seguinte equação:

$$\Delta Rn = Rn^+ - Rn^-$$

Onde:  $Rn^+$  = emissão de fluorescência em cada ponto;

$Rn^-$  = emissão de fluorescência na linha de base.

O computador constrói os *plots* de amplificação usando os dados de emissão de fluorescência durante a PCR. Os valores de  $\Delta Rn$  são plotados vezes o número de ciclos de amplificação. Durante os ciclos iniciais da amplificação por PCR, os valores de  $\Delta Rn$  não excedem a linha de base. Portanto, um ciclo limiar ( $Ct$ ) arbitrário é escolhido baseado na variação da linha de base. Os valores de  $Ct$  são então calculados por determinação do ponto no qual a fluorescência excede esse limiar escolhido. O  $Ct$  é definido como o número de ciclos nesse ponto (Figura 15.3.).



**Figura 14.3.** Exemplo de construção dos plots de amplificação, onde  $\Delta Rn$  é plotado contra o número de ciclos.

Existem dois métodos mais comumente utilizados para analisar dados de experimentos de PCR quantitativa em tempo real, são eles: o método de quantificação absoluta (método da curva padrão) e o método de quantificação relativa.

*Método de quantificação absoluta* – O número de cópias do gene de interesse é determinado relacionando-se o sinal da PCR com uma curva padrão. Para a construção da curva padrão, são utilizadas diluições seriadas de uma amostra de concentração conhecida. Esse método tem a vantagem de que a curva padrão incluída em cada PCR fornece um controle extra e correção na eficiência da PCR de cada amostra individual, tornando mais confiáveis as comparações entre experimentos.

*Método de quantificação relativa* – Esse método é mais usado quando existem muitos genes a serem testados em muitas amostras. Nesse caso, é feita a relação entre o sinal da PCR de um gene de interesse em um grupo tratado e o sinal de uma amostra controle (por exemplo, uma amostra não-tratada). Esse procedimento tem a vantagem de não necessitar da construção de curvas padrão. Por outro lado, as eficiências de amplificação dos genes de interesse e controle têm que ser similares para a obtenção de resultados confiáveis.

Em ambos os métodos são utilizados genes constitutivos para a normalização das amostras. Tais genes devem ter expressão constante em todos os tecidos e em todas as fases de desenvolvimento, bem como não serem afetados pelos

tratamentos experimentais. O gene mais utilizado com esta finalidade é o gene da  $\beta$ -actina.

O método de PCR quantitativa em tempo real tem se mostrado uma ferramenta de extrema importância na quantificação de RNAs mensageiros em baixos níveis, tornando a quantificação mais rápida e acurada.

#### 14.1. Protocolo de uma PCR em tempo real

1. Preparar o *mix* de PCR em tempo real:

Reagente	[Inicial]	[Final]	Volume 1 amostra	Volume (__ amostras)
<b>Power Sybr Green Master Mix</b>	2,5 X	0,2 mM	7,5 $\mu$ L	
<b>Primer F</b>	2,5 $\mu$ M	0,125 $\mu$ M	1,0 $\mu$ L	
<b>Primer R</b>	2,5 $\mu$ M	0,125 $\mu$ M	1,0 $\mu$ L	
<b>Água ultrapura</b>			3,5 $\mu$ L	
<b>cDNA</b>	15 ng/ $\mu$ L		2 $\mu$ L	
<b>TOTAL</b>			15 $\mu$ L	

\*\* Manter a solução contendo *Sybr Green* ao abrigo da luz

3. Colocar 13  $\mu$ L do *mix* de PCR em cada pocinho da placa.

4. Colocar 2  $\mu$ L do cDNA (15 ng/ $\mu$ L) de cada amostra no respectivo pocinho da placa e homogeneizar bem a mistura (amostra + *mix*).

5. Dar um pulso cuidadosamente (por centrifugação a 3.000 rpm).

6. Colocar a placa no aparelho de PCR em tempo real.

7. Escolher o programa de amplificação do gene de interesse. Para o gene MCP-1, utiliza-se o seguinte programa.

→ Pré-incubação: 95 °C por 10 minutos

→ Amplificação:

45 ciclos de  $\left\{ \begin{array}{l} 95 \text{ }^\circ\text{C por 10 segundos} \\ 60 \text{ }^\circ\text{C por 60 segundos} \end{array} \right.$

→ Curva de *melting*:

95 °C por 0 minuto

60 °C por 15 minutos

95 °C por 0 minuto

→ Resfriamento: 40 °C por 30 segundos

## 15. ANÁLISE DE POLIMORFISMOS PELO SISTEMA TaqMan

*Minos Esperândio Carvalho*

*Luiz Lehmann Cutinho*

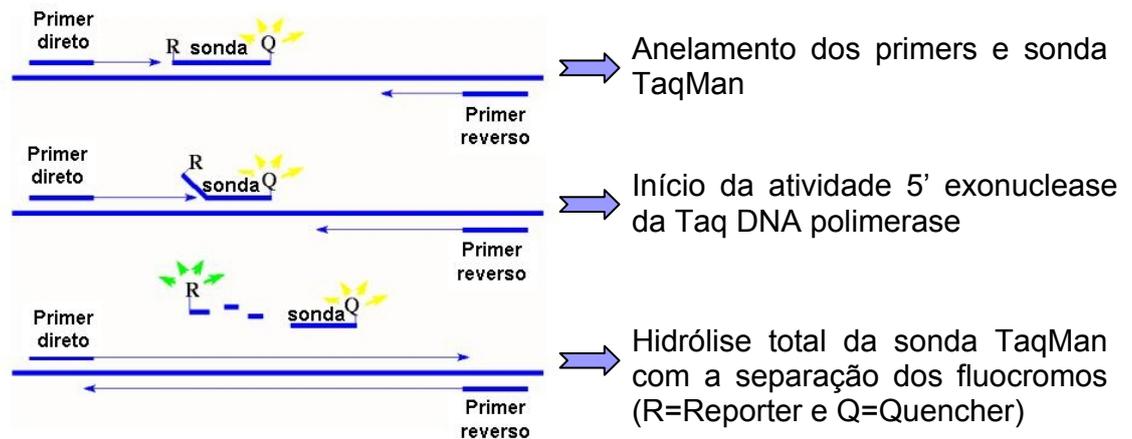
O Sistema para detecção de polimorfismos pelo método TaqMan permite a detecção do produto da PCR à medida que este vai sendo sintetizado na reação. O engenhoso sistema é baseado no uso de uma sonda, dirigida contra uma região interna da sequência que se deseja amplificar.

Para isto, é utilizado o sistema ou também conhecido, ensaio de detecção TaqMan, que contém *primers* e sondas. Os primers amplificam a região genômica de interesse, enquanto que a sonda permite a identificação das variantes alélicas. Para estes sistemas que são sempre específicos podem ser sintetizadas duas sondas de forma a parear seletivamente no DNA molde onde se encontra o polimorfismo de interesse.

O princípio da técnica é que na extremidade 5' há um fluorocromo que só fluoresce se estiver distante fisicamente do fluorocromo na posição 3'. Este segundo fluorocromo funciona como capturador de energia (*quencher*) e não deixa com que a energia luminosa usada para excitar a sonda chegue em quantidade suficiente para excitar o primeiro fluorocromo. Estes dois fluorocromos estão representados como R e Q (para *quencher*) na Figura 1. Quando o primer hibridiza na região 5', a sonda também o faz no meio da sequência. À medida em que a Taq polimerase avança sintetizando a fita nova, ela vai degradando a sonda à sua frente, liberando o fluorocromo R da sonda e permitindo que absorva energia e emita luz. A energia para a excitação dos fluorocromos provem de um feixe de laser que atravessa a amostra e o método que faz isto se denomina PCR em tempo real (*real time PCR*) ou Taqman, capturando os sinais de fluorescência em cada ciclo da PCR.

Como exemplo, pode-se citar um sistema em que uma das sondas é marcada com o fluoróforo FAM™ na porção 5' do oligonucleotídeo e sintetizada de forma a parear perfeitamente com um dos alelos do polimorfismo de interesse (ex: FAM™-NNNNNANNNNN). A outra sonda será marcada com VIC® na porção 5' do oligonucleotídeo e sintetizada de forma a parear perfeitamente com o outro alelo do polimorfismo de interesse (ex: VIC®-NNNNNGNNNNN). Se, pelo exemplo dado, um indivíduo é homocigoto para o alelo A apenas a sonda marcada com FAM™ formará

par com a seqüência de DNA e será degradada pela enzima DNA polimerase (que tem também atividade exonuclease 5'-3'), durante a amplificação do DNA molde, liberando, portanto, a fluorescência do FAM™ que será captada pelo sistema de detecção do termociclador. A fluorescência VIC® não é liberada porque a sonda não se pareia ao DNA e não é degradada. Se, pelo exemplo dado, um indivíduo é heterozigoto (apresenta os dois alelos) as duas sondas formarão pares com as respectivas seqüências e ambas serão degradadas pela DNA polimerase, portanto, haverá sinal dos fluoróforos FAM™ e VIC® na leitura termociclador. Se, pelo mesmo exemplo dado, um indivíduo é agora homozigoto para o alelo G apenas a sonda marcada com VIC® formará par com a seqüência de DNA e será degradada pela enzima DNA polimerase liberando, portanto, a fluorescência do VIC® que será captada pelo sistema de detecção do termociclador. Neste caso, a fluorescência FAM™ não é liberada porque a sonda não se pareia ao DNA e não é degradada, assim como também está exemplificado na Figura 1.



A intensidade de emissão de luz fluorescente é proporcional à quantidade de produto formado

**Figura 15.1.** PCR em Tempo Real – Sistema de detecção Taqman.

### 15.1. Protocolo de reação para amplificação do polimorfismo SNP do gene da calpastatina posição C1645G

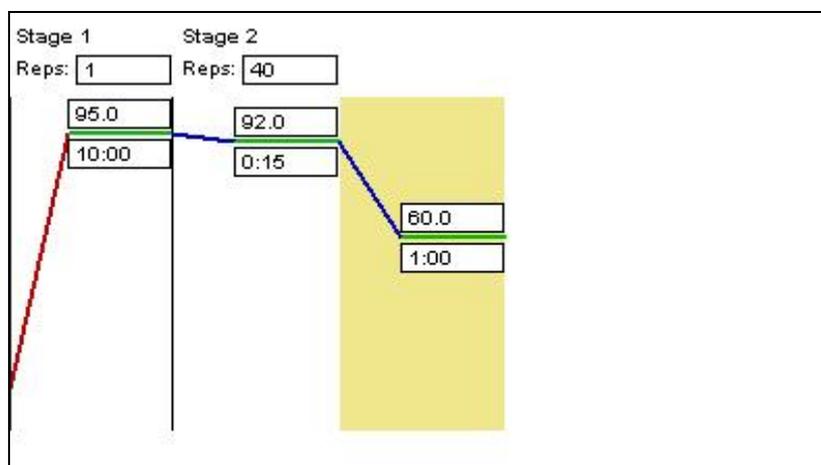
Para a PCR em tempo real são comumente utilizados kits comerciais de concentração e condições otimizadas. Como exemplo de reação para um kit da *Applied Biosystem*, pode-se considerar para uma reação de volume final de 10 µl, o uso de aproximadamente 15 ng de DNA, contendo 0,25 µl de Assay Mix® 1X

(Applied Biosystems), 5,0  $\mu\text{L}$  de Taqman<sup>®</sup> Master Mix Universal PCR 2X (Applied Biosystems). A reação ocorre com uma primeira e única etapa por 10 minutos a 95°C, seguidas com 45 ciclos de 15 segundos a 92°C e 1 minuto a 60°C.

**Tabela 15.1.** Mix de PCR em tempo real para o Sistema TaqMan:

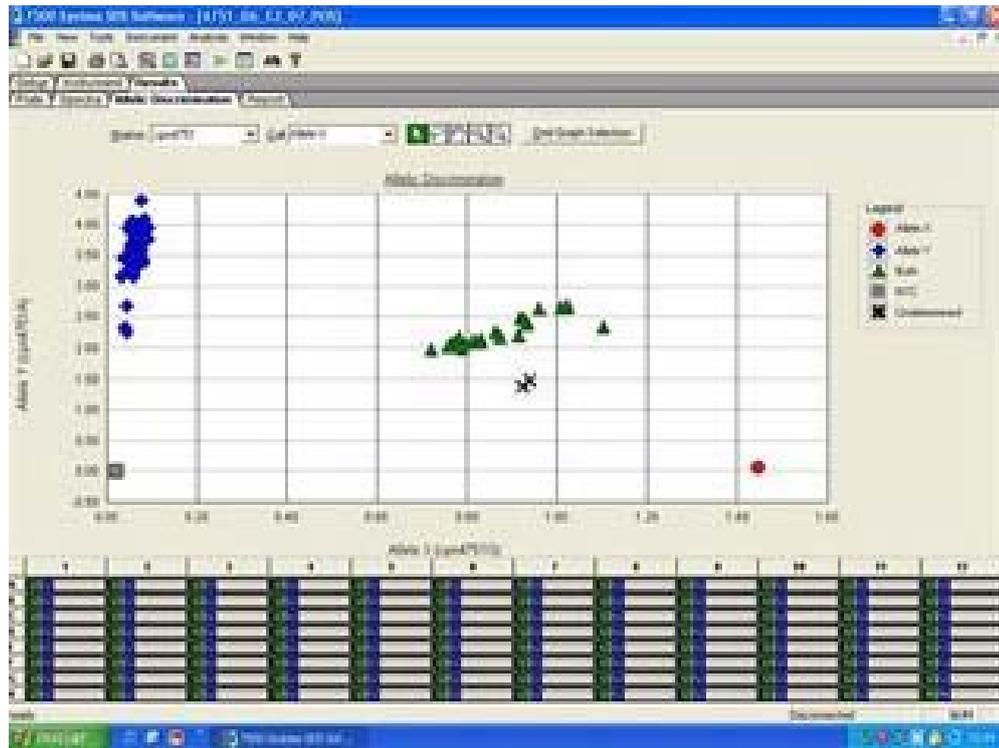
Reagente	<sup>1</sup> Volume 1 amostra	<sup>2</sup> Volume 1 amostra	Volume (__ amostras)
TaqMan Universal PCR MASTER MIX	5 $\mu\text{L}$	25 $\mu\text{L}$	
Assay	0,25 $\mu\text{L}$	1,25 $\mu\text{L}$	
Água ultrapura	3,75 $\mu\text{L}$	23,75 $\mu\text{L}$ contendo de 1-20 ng de DNA	
DNA	1 $\mu\text{L}$		
<b>TOTAL</b>	10 $\mu\text{L}$	50 $\mu\text{L}$	

<sup>1</sup> Reação otimizada para menor volume, <sup>2</sup> Indicação da empresa fabricante.



**Figura 15.2.** Protocolo de ciclos para amplificação com sistema Taqman.

Os equipamentos que permitem a amplificação em tempo real também são capazes de, a partir dos resultados obtidos, por capturarem fluorescência, genotipar automaticamente as amostras com seus próprios programas. A Figura 2 apresenta um exemplo de resultados de genotipagem automática pelo programa 7500 system software do equipamento ABI Prism<sup>®</sup> 7500 Sequence Detection System (Applied Biosystems) utilizado para PCR em tempo real.



**Figura 15.3.** Resultado de discriminação alélica por classificação automática. Os retângulos verdes representam indivíduos heterozigotos para o polimorfismo em questão, as bolas vermelhas e os losangos azuis representam os indivíduos homozigotos para ambas as formas alélicas. Os sinais em X representam amostras que não amplificaram suficientemente e o retângulo cinza representa o teste negativo NTC.

## 16. MAPAS DE LIGAÇÃO - SOFTWARE CRI-MAP (versão 2.4)

*Millor Fernandes do Rosário*

*Erica Elias Baron*

Mapa de ligação é uma ordenação linear de marcadores (locos) em grupos de ligação que devem corresponder ao número de cromossomos da espécie que está sendo estudada (Falconer & Mackay, 1996). Com o advento dos marcadores moleculares, a partir da década de 1960, tem sido possível obter mapas genéticos cada vez mais saturados (milhares, milhões de locos), a fim de permitir o mapeamento de QTLs e de genes candidatos, estudos de sintenia, dentre outros.

O software comumente utilizado para construir mapas de ligação em espécies de animais domésticos é o CRI-MAP (Green et al., 1990). O programa utiliza genótipos de indivíduos de uma população a fim de determinar a origem dos alelos, possibilitando a ordenação e determinação das distâncias entre os locos – se estiverem em um mesmo grupo de ligação – utilizando as taxas de recombinação. Este programa transforma as taxas de recombinação em distâncias centiMorgans (cM) através da função de mapeamento de Kosambi (Kosambi, 1944), considerando a interferência. De maneira geral, 1% de taxa de recombinação equivale a aproximadamente 1 cM.

### 16.1. Obtenção do programa CRI-MAP e instalação:

Para se obter o arquivo do programa CRI-MAP (versão 2.4) e o tutorial de instalação é necessário acessar a página <http://linkage.rockefeller.edu/soft/crimap/>. O CRI-MAP foi desenvolvido pelo Dr. Phil Green e colaboradores da University of Washington em linguagem computacional “C”. É necessário utilizar um compilador para linguagem “C” no computador onde será instalado. A compilação é feita para gerar um arquivo executável em sistema LINUX. Outros detalhes podem ser obtidos diretamente no tutorial.

### 16.2. Documentação do CRI-MAP

A documentação do programa compreende os seguintes itens: *i)* uso do CRI-MAP para pedigrees gerais e para locos de doenças; *ii)* iniciando o uso do programa;

*iii)* estruturas dos arquivos; *iv)* opções de comando; *v)* documentação técnica; *vi)* mudanças incorporadas na versão 2.4; *vii)* bibliografia.

### 16.3. Procedimentos

Para utilizar o programa é necessário um arquivo contendo os genótipos dos indivíduos organizados por famílias seguindo o formato exigido pelo programa. Para ser reconhecido pelo CRI-MAP o arquivo deve ser salvo com a extensão “.gen”. Por exemplo, se o foco do estudo for o cromossomo 30, então esse arquivo deverá ser nomeado "chr30.gen" em um editor de texto e estruturado como se segue:

1ª linha: número de famílias

2ª linha: número de marcadores

3ª linha: nome dos marcadores

4ª linha: nome da primeira família

5ª linha: número de indivíduos da primeira família

A partir da 6ª linha, deve ser inserido sem interrupção: identificação do indivíduo, número da mãe, número do pai e sexo (0 se fêmea, 1 se macho, 3 se desconhecido)

7ª linha: genótipos dos indivíduos para os marcadores

8ª linha: identificação do próximo indivíduo, repetindo o procedimento das 6ª e 7ª linhas, até que o último indivíduo.

Exemplo do arquivo de genótipos:

1ª linha	20									
2ª linha	6									
3ª linha	SSR1	SSR2	SSR3	SSR4	SSR5					
4ª linha	FAM01									
5ª linha	88									
6ª linha	5649	0	0	1						
7ª linha	208	208	289	303	259	263	171	173	272	272
8ª linha	886	0	0	0						
	208	289	301	301	265	265	171	173	272	272
	6232	0	0	1						
	208	208	289	297	263	265	173	185	264	270
	332	0	0	0						
	208	208	291	301	265	265	171	173	272	274

7716	886	5649	1						
208	289	289	301	259	265	171	173	272	272
7810	332	6232	0						
208	208	297	301	263	265	173	185	270	274
786	7810	7716	1						
208	208	301	301	259	265	171	185	272	274
787	7810	7716	1						
208	208	289	297	259	265	171	185	270	272
788	7810	7716	1						

Os pais desconhecidos são identificados pelo número 0 após o número do indivíduo.

Uma vez criado o arquivo “chr30.gen”, inicia-se o programa, a partir de uma “janela” da plataforma DOS :

```
> cd CRIMAP ↵
CRIMAP> { prompt esperando comando}
```

Para iniciar o programa é necessário usar o comando “prepare” para criar os arquivos “.dat”, “.ord”, “.par” e “.loc”. Esses arquivos serão utilizados posteriormente em outras opções de comando.

O arquivo “.dat” possui duas estruturas de dados, com as informações sobre as fases dos locos no cromossomo para cada marcador e indivíduo, denominadas de fase conhecida e desconhecida. As fases são classificadas como “0” se o alelo é herdado da linhagem materna, 1 se é da linhagem paterna e X quando a fase é desconhecida.

O arquivo “.loc” traz a informação do número de meioses informativas para cada marcador.

O arquivo “.ord” traz informação sobre a ordem em que os marcadores estão sendo avaliados.

O arquivo “.par” é um arquivo de parâmetros e já está criado dentro do programa (default); se necessário ele pode ser modificado em um editor de texto. Alguns dos parâmetros contidos nesse arquivo são: memória inicial a ser utilizada;

se a porcentagem de recombinação calculada será igual ou diferente para cada sexo; tolerância para determinação para conversão do algoritmo EM; o número máximo de ordens permitidas para a construção do mapa para as fases conhecidas e desconhecidas; tolerância para descartar ordens de locos; a ordem inicial dos marcadores e os marcadores que serão inseridos.

Exemplificando:

Na tela do DOS, na linha de comando, deve ser escrito crimap “nome do cromossomo” e chamar a opção “prepare” para gerar os arquivos acima citados:

```
CRIMAP> ./crimap 30 prepare ↵
```

A saída desta opção de comando é a seguinte:

```
millor@lpa-27:~$ cd CRIMAP
millor@lpa-27:~/CRIMAP$ cd TESTE
millor@lpa-27:~/CRIMAP/TESTE$ ./crimap 30 prepare

chromosome 30

50400 bytes allocated in orders_morecore
Use existing .dat file chr30.dat? (y/n) n
Are you sure you want to delete file chr30.dat? (y/n) y
504000 bytes allocated in morecore
Creating .dat file chr30.dat from .gen file chr30.gen
504000 bytes allocated in morecore

family id FAM01
family id FAM02
family id FAM03
family id FAM04

Writing file chr30.dat
Finished writing chr30.dat
Writing locus names to chr30.loc
```

Current values for parameters:

```
par_file = chr30.par
dat_file = chr30.dat
gen_file = chr30.gen
ord_file = chr30.ord
nb_our_alloc = 3000000  [# bytes reserved for our_alloc]
SEX_EQ = 1  [0 = sex specific analysis, 1 = sex equal]
TOL = 0.010000
PUK_NUM_ORDERS_TOL = 6
PK_NUM_ORDERS_TOL = 8
PUK_LIKE_TOL = 3.000
PK_LIKE_TOL = 3.000
use_ord_file = 0
write_ord_file = 1
use_haps = 1
Do you wish to change any of these values? (y/n) n
```

The loci and their indices are:

```
0  SSR1          1  SSR2          2  SSR3
3  SSR4          4  SSR5
```

Do you wish to enter any new haplotyped systems? (y/n) n

Do you wish to hold any additional recombination fractions fixed (NB these will only be used with the options FIXED and CHROMPIC, and only when the loci in question are adjacent)? (y/n) n

The crimap options are:

```
[1] build [2] instant [3] quick [4] fixed
[5] flips [6] all [7] twopoint [8] chrompic
```

Enter the number of the option you will be running next: 1

```
Do you wish to build the map incorporating ALL loci in decreasing order of
their informativeness (the usual procedure)? (y/n) y
```

```
The loci and their indices are:
```

```
1  SSR2          3  SSR4
2  SSR3          4  SSR5          0  SSR1
```

```
Creating orders file chr30.ord
```

```
Done.
```

```
OK to set up new parameter file? (y/n) y
```

```
chr30.par has been created; use text editor for further modifications, if needed
```

```
millor@lpa-27:~/CRIMAP/TESTE$
```

Das oito opções apresentadas pelo programa recomenda-se que a primeira empregada seja a opção de comando “build”. Com esta opção é possível a construção de um mapa de ligação pela incorporação sequencial dos marcadores. No início, o arquivo de ordens está vazio e o mapa será constituído apenas pelos dois marcadores mais informativos, que serão introduzidos automaticamente na linha de comando “ordered\_loci” do arquivo “.par”.

```
CRIMAP> ./crimap 30 build > build30 ↵
CRIMAP> {prompt esperando comando}
```

Obs: após a estrutura de comando “crimap {número do cromossomo} {opção de comando}” coloca-se o símbolo “>” seguido do nome do arquivo de saída. Essa nova estrutura dá o comando para que a saída da análise seja gravada no mesmo diretório em que está sendo rodado o programa.

Após aparecer na tela a linha de comando acima, num editor de texto, abre-se o arquivo chr30.loc e verifica-se quais os dois marcadores com maior número de meioses informativas de fase conhecida. Em seguida no arquivo chr30.par, modifica-se a linha “ordered\_loci” colocando nela o número dos dois marcadores escolhidos do arquivo chr30.loc. Na linha de comando “inserted\_loci” é colocado o terceiro

melhor marcador. E em seguida, feito um novo ciclo com a opção “build”. O resultado terá inserido o terceiro marcador numa determinada ordem, com um determinado valor de “likelihood ratio”. Retorna-se então ao arquivo chr30.par, e acrescenta-se na linha de comando “ordered\_loci” o marcador ordenado, e coloca-se o quarto melhor marcador na linha de comando “inserted\_loci”. Dessa maneira o mapa “multipoint” vai sendo construído através de ciclos de adição de locos, seguido por avaliação do valor de “likelihood ratio” de cada ordem gerada, usando a opção de comando “build”. Cada novo loco é testado em todos os possíveis intervalos do mapa corrente. Quando o valor de LOD é igual ou superior a 3 o marcador é posicionado no intervalo e, assim, esse último mapa passa a ser usado para a adição de novos locos. Em seguida com o uso da opção de comando “chrompic” é feita uma verificação da fase gamética de cada indivíduo analisado, para cada um dos dois alelos rastreando prováveis erros de “crossover locations”.

Tem-se até o momento a seguinte seqüência de trabalho:

1. Estruturar arquivo
2. Abrir programa
3. Rodar o comando “prepare”
4. Abrir arquivo “.dat”
5. Modificar arquivo “.par”
6. Rodar comando “build”
7. Verificar saída em arquivo de texto
8. Repetir vários ciclos dos passos 5, 6 e 7 até incorporação de todos os locos
9. Verificar qual ordem pelos arquivos de saída tem melhor LOD
10. Verificar a fase gamética dos indivíduos pelo comando “chrompic”

Se após a verificação da fase gamética, forem encontradas regiões com mais de uma recombinação cromossômica num indivíduo, é importante a verificação dos genótipos e até mesmo voltar ao arquivo do eletroferograma para conferência dos alelos do loco.

Se todos os indivíduos estiverem corretos, podemos abrir o arquivo de saída para verificar a ordem dos locos e suas respectivas distâncias em cM (centiMorgan).

Abaixo segue um exemplo da saída do último ciclo, com todos os locos inseridos no mapa de ligação:

```
Sex_averaged map (recomb. frac., Kosambi cM):
 4  SSR5          0.0
      0.37  48.3
 1  SSR2          48.3
      0.39  53.3
 3  SSR4          101.6
      0.25  27.0
 0  SSR1          128.6
      0.09  8.9
 2  SSR3          137.4
* denotes recomb. frac. held fixed in this analysis
log10_like = -779.21

Sex-specific map (recomb. frac., Kosambi cM -- female, male ):
 4  SSR5          0.0          0.0
      0.38  49.6          0.36  45.0
 1  SSR2          49.6          45.0
      0.39  51.6          0.40  54.5
 3  SSR4          101.2          99.5
      0.32  37.5          0.21  21.9
 0  SSR1          138.7          121.4
      0.07  6.7          0.09  9.3
 2  SSR3          145.4          130.7
* denotes recomb. frac. held fixed in this analysis
log10_like = -773.90
```

Além da metodologia apresentada acima, o mapa pode ser construído de outras formas. Uma delas é utilizando o comando “twopoint” em ciclos sucessivos para a adição de loco, um a um, em um intervalo de dois marcadores apenas. No entanto, com este comando não é possível construir o mapa propriamente dito; podemos somente inferir os valores de “likelihood ratio” para cada posição de um marcador, dentro dos diferentes intervalos. Em seguida com os melhores valores, é

possível rodar o comando “build”, com todos os marcadores presentes na linha de comando “ordered\_loci”. Essa metodologia é mais usada quando não se sabe a qual grupo de ligação pertencem os locos.

Sugere-se também empregar os comandos “all” que testa todas as possíveis ordens para um dado loco e o comando “flips” que promove combinações 2 a 2 (“flips2”), 3 a 3 (“flips3”), etc, buscando por combinações com melhor “likelihood ratio”, ou seja, quanto mais próximo de zero for o valor desta estatística, mais verossímil ele será.

## 17. MAPEAMENTO DE QTLs – SOFTWARE QTL EXPRESS

*Millor Fernandes do Rosário*

*Kátia Nones*

QTL (*quantitative trait loci* ou loco que controla características quantitativas) é um segmento do cromossomo que está associado a uma determinada característica quantitativa, não sendo necessariamente um único gene (Falconer & Mackay, 1996). O propósito do mapeamento de QTLs é inferir sobre os seus genótipos, estimando seus efeitos e sua localização, a partir de associações estatísticas com marcadores genéticos conhecidos (Liu, 1998).

O software mais utilizado para mapear QTLs em populações parcialmente endogâmicas, como nas espécies de animais domésticos é o QTL EXPRESS (Seaton et al., 2002). Este programa pode ser acessado através de: <http://qtl.cap.ed.ac.uk/>. Entretanto, a partir do segundo semestre de 2009, ele será substituído pela plataforma GridQTL (Seaton et al., 2006; Hernández-Sánchez et al., 2009), que pode ser acessado através de: <http://www.gridqtl.org.uk/>. O GridQTL permite realizar análises de desequilíbrio de ligação e análise de ligação.

O QTL EXPRESS implementa o mapeamento por intervalo (Lander e Botstein, 1989) através da regressão por quadrados mínimos (Haley et al., 1994), realizando uma varredura dos intervalos entre dois locos adjacentes, testando a hipótese inicial de que não há QTL no intervalo considerado.

O princípio básico deste programa é baseado na fixação dos alelos contrastantes do possível QTL nas linhagens fundadoras da população utilizada para mapear os QTLs. Por exemplo, em aves, na linhagem de corte os alelos QQ estão fixados enquanto que na linhagem de postura são os alelos qq ou vice-versa; em bovinos, na raça Gir são os alelos QQ e na raça Holandesa os alelos qq ou vice-versa. Também se baseia na estimação das probabilidades de identidade por descendência (IBD) dos alelos em cada loco considerando um conjunto de dados multilocos, além de testar um modelo estatístico, podendo incluir efeitos fixos e covariáveis para as observações (fenótipos) e os coeficientes IBD.

As Figuras 1, 2 e 3 abaixo foram extraídas de Haley et al. (1994). Elas ilustram um esquema de população experimental F<sub>2</sub> com duas linhagens para

cálculo das probabilidades condicionais dos genótipos dos QTLs dadas as probabilidades dos marcadores flanqueadores.

**TABLE 1**  
**Example F<sub>2</sub> pedigree from a cross between outbred lines with four markers (A, B, C and D) and possible line origin combinations of marker alleles**

Line 1: sire of the sire (SS) <i>A<sub>1</sub>A<sub>1</sub>B<sub>2</sub>B<sub>2</sub>C<sub>2</sub>D<sub>1</sub>D<sub>1</sub></i>		Line 2: dam of the sire (DS) <i>A<sub>2</sub>A<sub>2</sub>B<sub>1</sub>B<sub>1</sub>C<sub>2</sub>D<sub>2</sub>D<sub>2</sub></i>		Line 1: sire of the dam (SD) <i>A<sub>1</sub>A<sub>1</sub>B<sub>2</sub>B<sub>2</sub>C<sub>2</sub>D<sub>1</sub>D<sub>2</sub></i>		Line 2: dam of the dam (DD) <i>A<sub>2</sub>A<sub>2</sub>B<sub>2</sub>B<sub>2</sub>C<sub>1</sub>C<sub>1</sub>D<sub>1</sub>D<sub>2</sub></i>	
Sire (S) <i>A<sub>1</sub>A<sub>2</sub>B<sub>1</sub>B<sub>2</sub>C<sub>1</sub>C<sub>2</sub>D<sub>1</sub>D<sub>2</sub></i>				Dam (D) <i>A<sub>1</sub>A<sub>2</sub>B<sub>2</sub>B<sub>2</sub>C<sub>1</sub>C<sub>2</sub>D<sub>1</sub>D<sub>2</sub></i>			
F <sub>2</sub> offspring (O) <i>A<sub>2</sub>A<sub>2</sub>B<sub>1</sub>B<sub>1</sub>C<sub>2</sub>C<sub>2</sub>D<sub>2</sub>D<sub>2</sub></i>							

Line 1: sire of the sire	Line 2: dam of the sire	Line 1: sire of the dam	Line 2: dam of the dam	Line origin combination	Putative QTL	Marker			
						A	B	C	D
×		×		11	QQ				
×			×	12	Qq				
	×	×		21	qQ		*	*	*
	×		×	22	qq	*	*	*	*

Possible line origins of alleles are indicated by ×. For each line origin combination the putative QTL genotype at a locus fixed for allele Q in line 1 and allele q in line 2 is shown. The possible line origin combinations for the F<sub>2</sub> individual in the pedigree shown above are shown by asterisks in the right-hand side of the table.

**Figura 17.1.** Combinações das possíveis origens entre as linhagens dos alelos dos locos.

**TABLE 2**  
**The probability of the line origin combination of a locus conditional upon the line origin combination of a linked locus and the recombination fractions between the two loci**

Line origin combination	Line origin combination			
	11	12	21	22
11	$(1-r_m)(1-r_f)$	$(1-r_m)r_f$	$r_m(1-r_f)$	$r_mr_f$
12	$(1-r_m)r_f$	$(1-r_m)(1-r_f)$	$r_mr_f$	$r_m(1-r_f)$
21	$r_m(1-r_f)$	$r_mr_f$	$(1-r_m)(1-r_f)$	$(1-r_m)r_f$
22	$r_mr_f$	$r_m(1-r_f)$	$(1-r_m)r_f$	$(1-r_m)(1-r_f)$

Conditional probabilities for the line origin combinations given in Table 1. The recombination frequencies between the pair of loci in the male and female parent are  $r_m$  and  $r_f$ , respectively.

**Figura 17.2.** Probabilidades das combinações de origem entre as linhagens dos alelos entre dois locos.

**TABLE 3**

**Example calculation of the probabilities of line origin combinations of a putative QTL conditional upon the possible line origin combinations of flanking markers**

Line origin combination of alleles (A-B-Q-C)	Conditional probability	Example: $r_{AB} = r_{BQ} = r_{QC} = 0.1$
22-21-11-21	$r_{AB}(1-r_{AB})r_{BQ}(1-r_{BQ})r_{QC}(1-r_{QC})/P$	0.004
22-22-11-21	$(1-r_{AB})^2r_{BQ}^2r_{QC}(1-r_{QC})/P$	0.004
22-21-12-21	$r_{AB}(1-r_{AB})r_{BQ}^2r_{QC}^2/P$	0.000
22-22-12-21	$(1-r_{AB})^2r_{BQ}(1-r_{BQ})r_{QC}^2/P$	0.004
22-21-21-21	$r_{AB}(1-r_{AB})(1-r_{BQ})^2(1-r_{QC})^2/P$	0.328
22-22-21-21	$(1-r_{AB})^2r_{BQ}(1-r_{BQ})(1-r_{QC})^2/P$	0.328
22-21-22-21	$r_{AB}(1-r_{AB})r_{BQ}(1-r_{BQ})r_{QC}(1-r_{QC})/P$	0.004
22-22-22-21	$(1-r_{AB})^2(1-r_{BQ})^2r_{QC}(1-r_{QC})/P$	0.328

For this individual the line origin combination (as defined in Table 1) of the first marker (A) is 22, that of the second marker (B) may be either 21 or 22 and that of the third marker (C) is 21. The putative QTL (Q) is placed between B and C. The recombination frequencies are assumed to be the same in both sexes and that between A and B is  $r_{AB}$ , that between B and the putative position of the QTL is  $r_{BQ}$  and that between the QTL and C is  $r_{QC}$ . P is the sum of the numerators of the conditional probabilities. An example is given for  $r_{AB} = r_{BQ} = r_{QC} = 0.1$ . For this individual the predicted coefficient for a would be -0.324 (= 0.004 + 0.004 - 0.004 - 0.328) and for d would be 0.660 (= 0.000 + 0.004 + 0.328 + 0.328).

**Figura 17.3.** Probabilidades das combinações de origem entre as linhagens de um suposto QTL condicional às combinações de origem entre as linhagens dos marcadores flanqueadores.

O programa possui diferentes opções de análises de acordo com o delineamento utilizado na população a ser estudada ( $F_2$  e half-sib são as mais usadas). Em cada uma dessas opções são encontradas instruções sobre os formatos dos arquivos a serem utilizados no programa.

A seguir é descrito o formato dos três arquivos (genótipos, mapa e fenótipos) que devem ser criados para realizar as análises no QTL EXPRESS utilizando a opção “ $F_2$  ANALYSIS”. Todos os arquivos devem ser tipo texto, onde os dados são separados por espaços ou tabulação.

#### ARQUIVO DE GENÓTIPOS:

1ª linha – Número de marcadores

2ª linha – Lista dos nomes dos marcadores – qualquer caracter alfanumérico.

3ª linha – Nome dado a cada geração (nome das linhagens parentais,  $F_1$  e  $F_2$ )

4ª linha – Indicação de sexo (1-macho e 2-fêmea, ou M-macho e F-fêmea)

5ª linha – Símbolo que representa dados perdidos (\*, ou -, ou 0)

6ª linha – Dados de cada indivíduo, na seguinte ordem:

1) Identificação do indivíduo, qualquer caracter alfanumérico.

2) Identificação do pai; no caso dos parentais onde a identificação do pai é desconhecida utilizar “\*, ou -, ou 0”.

- 3) Identificação da mãe, igual a do pai.
- 4) Sexo, de acordo com o definido na linha 4.
- 5) Linhagem ou geração, como definido na linha 3
- 6) Genótipos dos marcadores, cada dois valores correspondem aos alelos de um marcador; os genótipos devem estar na mesma ordem da lista dos marcadores (linha 2)

Este arquivo pode ser obtido reformatando o arquivo “.gen” do CRIMAP ou criado através do EXCEL.

#### EXEMPLO:

```

14
MK1 MK2 MK3 MK4 MK5 MK6 MK7 MK8 MK9 MK10 MK11 MK12 MK13 MK14
LW MS F1 F2
1 2
0
1039 1529 1530 1 F1 4 1 4 7 1 2 1 2 5 2 1 1 2 5 4 2 1 2 1 3 2 4 6 3 1 2 5 2
1040 1529 1530 2 F1 4 1 4 7 1 2 1 2 7 2 1 1 2 5 2 5 1 2 1 3 2 4 1 3 4 2 0 0

```

#### ARQUIVO DO MAPA DE LIGAÇÃO:

1ª linha – Número de cromossomos.

2ª linha – Intervalo para cálculo das probabilidades genotípicas e coeficientes em cM.

3ª linha – Nome do primeiro cromossomo, número de marcadores deste cromossomo, determinar se o mapa será construído para os dois sexos (=1) ou não (=2).

4ª linha – Para o primeiro cromossomo ou grupo de ligação, indicar os nomes dos marcadores em ordem, separados pelas distâncias entre eles em cM. Os nomes dos marcadores devem ser os mesmos utilizados no arquivo de genótipos. Se for feita a opção pela utilização do mesmo mapa para os dois sexos, esta linha aparece apenas uma vez, se os dois sexos tiverem mapas separados esta linha aparece duas vezes com diferentes distâncias.

Repetir as linhas 3 e 4 para cada cromossomo.

## EXEMPLO:

```
2
1
SSC1 7 1
MK2 42 MK1 28 MK7 8 MK10 28 MK3 3 MK4 22 MK5
SSC7 7 1
MK6 35 MK8 25 MK9 17 MK11 22 MK12 25 MK13 21 MK14
```

## ARQUIVO DE FENÓTIPOS:

1ª linha – Número de efeitos fixos, covariáveis e características, respectivamente.

2ª linha – Nome de cada efeito fixo, covariável e característica.

3ª linha – Código para dados perdidos; os indivíduos podem ter dados perdidos para as características fenotípicas, mas todos devem ter valores para efeitos fixos e covariáveis.

4ª linha – Identificação do indivíduo (mesma do arquivo de genótipos), seguida dos valores de efeitos fixos, covariáveis e características.

## EXEMPLO:

```
2 1 1
Sexo Grupo Peso Esp_Toucinho
-999
1117 1 1 67800 19
1118 2 1 75400 24
1119 2 1 69800 29
1124 1 1 64800 18
1125 1 1 65800 19
```

Depois dos arquivos organizados conforme os modelos acima, faz-se a opção do delineamento a ser utilizado na análise (Ex: F<sub>2</sub> Analysis). Nesta janela temos acesso aos arquivos através de um “browse” para cada um dos arquivos (genótipos, mapa e fenótipos) clicando em seguida no botão “SUBMIT”. Neste momento os genótipos são checados para erros, pela comparação dos alelos nos parentais e nas progênies, em famílias de irmãos completos e meio-irmãos. Os erros são indicados em uma janela (“Fatal errors”) para análise e correções.

Se não forem encontrados erros, a análise avançará e uma nova janela estará disponível, indicando o número de famílias de irmãos completos e o número total da progênie que será utilizada na análise. Estará também disponível uma tabela indicando o nome e número de marcadores, suas distâncias e o número de alelos observados em cada marcador.

Para continuar o procedimento de análise deve ser clicado o botão “CONTINUE”; uma nova janela aparecerá.

Nesta nova janela temos vários campos para a escolha dos parâmetros e modelos a serem utilizados na análise, juntamente com as características que se deseja analisar. Depois de escolhidos todos os parâmetros, clicar no botão “SUBMIT”.

Uma nova janela se abrirá apresentando o resultado do mapeamento por intervalos (em cM), da análise estatística (F, razão de verossimilhança e LOD) e estimativas do efeito aditivo e de dominância, além dos gráficos do mapeamento por intervalo, dos efeitos de aditividade e dominância e das permutações.

As permutações são implementadas de acordo com a metodologia proposta por Churchill e Doerge (1994) e têm por objetivo determinar os níveis de significância ou *threshold* no cromossomo ou no genoma. O QTL EXPRESS também permite que os intervalos de confiança de cada QTL mapeado sejam estimados de acordo com a proposta de Visscher et al. (1996) por meio do procedimento de *bootstrap* com reamostragem.

## 18. MicroRNA: UMA NOVA CLASSE DE MODULADORES DA EXPRESSÃO GÊNICA

*Luiz Lehmann Coutinho*

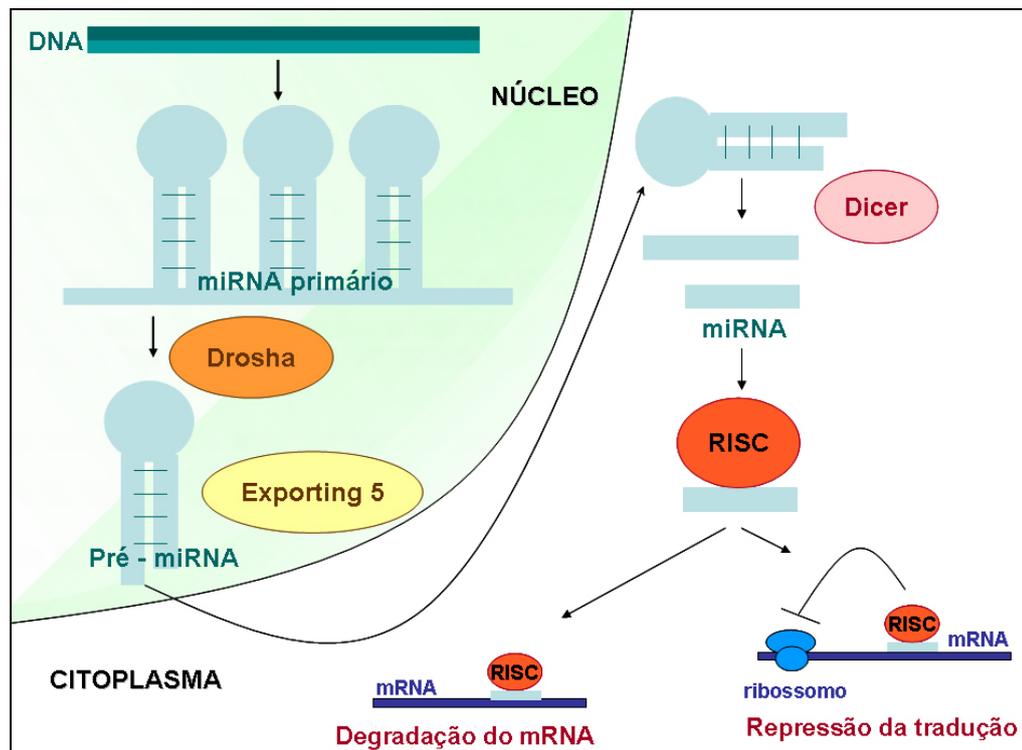
*Ana Paula Dini Andreote*

*Érika Cristina Jorge*

MicroRNAs (miRNAs) são uma classe abundante de moléculas pequenas de RNA não-codante, de aproximadamente 22 nucleotídeos de comprimento, com capacidade de regular a expressão gênica por meio de controle da tradução ou degradação do RNA mensageiro (RNAm) (Ambros et al., 2003). Estes microRNAs foram originalmente descobertos em *Caenorhabditis elegans* como reguladores pós-transcricionais de genes envolvidos com o processo de desenvolvimento (Ambros et al., 2003; Lee RC et al., 1993) e agora são identificados em todos os organismos multicelulares, com capacidade de modular a expressão de centenas de genes (Lim et al., 2005; Shingara et al., 2005), atuando em vários processos biológicos, entre eles: morfogênese do cérebro (Giraldez et al., 2005), degradação de RNAm materno durante o início do desenvolvimento (Giraldez et al., 2006), secreção de insulina (Poy et al., 2004), proliferação e diferenciação de cardiomiócito (Zhao et al., 2005), tumorigênese, defesa viral (Lecellier et al., 2005), diferenciação hematopoiética (Chen and Lodish, 2005), desenvolvimento embrionário (Tang F et al., 2007) e desenvolvimento muscular (McCarthy JJ and Esser KA, 2007).

Os microRNAs são transcritos na forma de miRNA-primário pela enzima RNA polimerase II, contendo um *cap* 5' e uma cauda poli-A característicos (Cai et al., 2004; Lee et al., 2002; Smalheiser, 2003). MicroRNAs também podem ser transcritos como parte do íntron de um RNAm primário. Enquanto ainda no núcleo, o miRNA-primário apresenta uma estrutura secundária tipo grampo, que é reconhecida e clivada pela enzima Drosha, uma RNase que reconhece RNA fita dupla (Lee et al., 2003). A molécula resultante é um RNA fita dupla de aproximadamente 70 nucleotídeos, na forma de grampo, e com terminais protuberantes na extremidade 5' ou 3'. Esta estrutura é denominada pré-miRNA (Figura 1). Este pré-microRNA é então, transportado do núcleo para o citoplasma por uma proteína conhecida com Exportin-5 (Lund et al., 2004; Yi et al., 2003) (Figura 1). No citoplasma, a enzima Dicer, membro da superfamília de RNase III, cliva este pré-miRNA (Lee et al., 2003;

Yi et al., 2003) em uma molécula de RNA fita dupla, com extremidades 5' e 3' protuberantes (Lund et al., 2004). Uma das fitas de RNA (de aproximadamente 22 nucleotídeos) se associa com o complexo de silenciamento de RNA-induzido (do inglês, *RNA-induced silencing complex*, RISC) e se transforma na forma ativa e madura do miRNA. Esta forma madura modula a expressão gênica por meio de interação com regiões específicas da porção 3' dos RNAm (Schwarz et al., 2003) (Figura 18.1).



**Figura 18.1:** Biogênese e modo de ação dos microRNAs.

A maneira exata como os miRNAs modulam a expressão gênica ainda não está totalmente elucidada. Estudos recentes apontam que os mecanismos são pós-transcricionais, envolvendo a inibição da tradução nas fases de iniciação e alongação, ou levando diretamente à degradação do RNAm alvo. Estudos sugerem que os microRNAs reconhecem o RNAm alvo por interações de pareamento de bases entre a extremidade 5' do microRNA e seqüências complementares na região 3'UTR do RNAm alvo (Nilsen, 2007).

O padrão de expressão dos miRNAs é diferente nos diversos tecidos (Barad et al., 2004; Farh et al., 2005; Lim et al., 2005; Liu et al., 2004; Shingara et al., 2005) e a

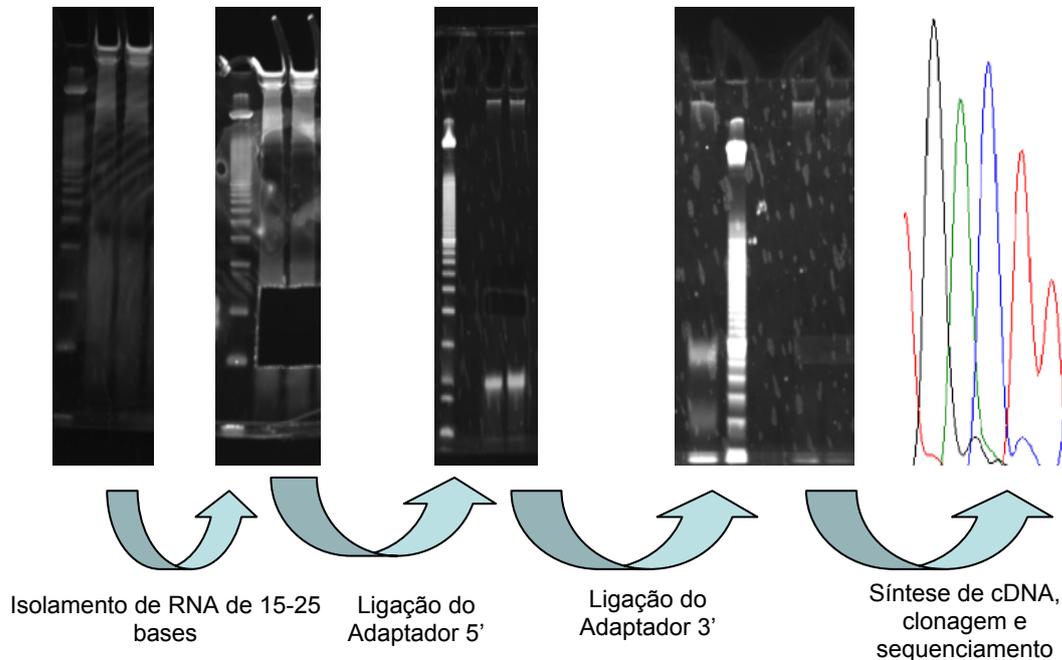
abundância de um miRNA específico varia de algumas cópias a mais de 30.000 moléculas por célula (Chen et al., 2005).

Até o momento, 4.321 miRNA já foram descritos, sendo: 474 deles em humanos, 98 em bovinos, 149 em aves e 54 em suínos (<http://microrna.sanger.ac.uk/> ; release 9.1, fevereiro de 2007). Programas de bioinformática e experimentos de clonagem desenvolvidos para identificar miRNAs no genoma, sugerem a existência de aproximadamente 800 miRNAs no genoma humano. Estes programas buscam em bancos de dados genômicos por seqüências de aproximadamente 70 bases, que apresentem potencial de formar um RNA fita dupla estável e que seja conservado entre espécies (uma lista dos programas pode ser encontrado em (Kim and Nam, 2006).

A comprovação da existência de um novo miRNA deve ser feita utilizando-se estratégias independentes e complementares. Ou seja, é preciso que o miRNA seja detectado (por *Northen blo* ou clonagem, por exemplo) e que a seqüência do miRNA primário, obtido da seqüência genômica da espécie em estudo, tenha capacidade de formar a estrutura de RNA fita dupla característica dos miRNAs (Ambros et al., 2003).

A clonagem de miRNAs pode ser feita com a construção de bibliotecas de cDNA a partir da fração de RNA de baixo peso molecular, de tamanho entre 15 e 25 bases. Estas bibliotecas permitem a clonagem dos microRNAs expressos em um tecido de interesse. Neste caso, um cuidado especial deve ser observado no processo de extração de RNA total, pois alguns métodos de purificação do RNA podem eliminar a fração de baixo peso molecular. Uma vez obtido o RNA total, diferentes procedimentos podem ser utilizados para enriquecer a amostra com RNA de baixo peso molecular. As estratégias para esta etapa variam desde a precipitação do RNA de alto peso molecular com PEG, até a utilização de colunas ou sistema de eletroforese para o isolamento das moléculas menores. A fração enriquecida de RNA com fragmentos de 15-25 bases é isolada em gel de acrilamida. Após a eluição do gel, adaptadores são ligados aos RNAs pela ação de uma RNA ligase. Estes adaptadores utilizados são compostos de oligonucleotídeos modificados com o objetivo de facilitar a ligação dos adaptadores com o RNA e impedir a ligação dos adaptadores entre si. Após a ligação, uma transcrição reversa é realizada a partir de um *primer* complementar ao adaptador 3'. Finalmente é feita uma PCR a partir de

*primers* complementares aos adaptadores, com posterior clonagem e seqüenciamento. A figura 18.2 ilustra as etapas empregadas para a clonagem de miRNAs.



**Figura 18.2:** Metodologia utilizada para clonar e identificar microRNAs.

Uma vez identificado um novo microRNA, este precisa ser validado pelos critérios de formação de fita dupla e conservação entre espécies. Considerando que o miRNA clonado representa o miRNA maduro, de aproximadamente 22 bases, é necessário identificar as regiões genômicas que flanqueiam o miRNA. Isto pode ser feito com o auxílio dos bancos do NCBI (<http://www.ncbi.nlm.nih.gov/>) ou da UCSC (<http://genome.ucsc.edu/>). Para exemplificar, apresentamos abaixo as etapas necessárias para validar a identidade de um miRNA. Como exemplo, utilizamos o hsa-mir-26a, um miRNA identificado em seres humanos (Figura 18.3).

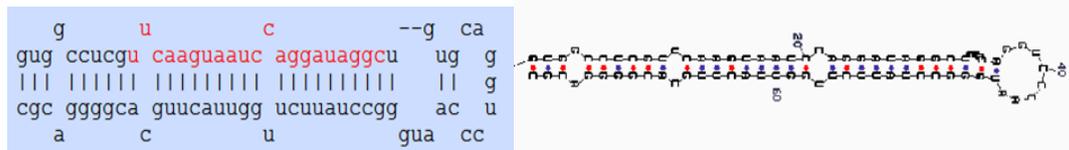
1- Sequencia do miRNA obtida de uma biblioteca de cDNA

hsa-miR-26a UUCAAGUAAUCCAGGAUAGGC

2- Sequencia genômica flanqueando o hsa-mir-26a

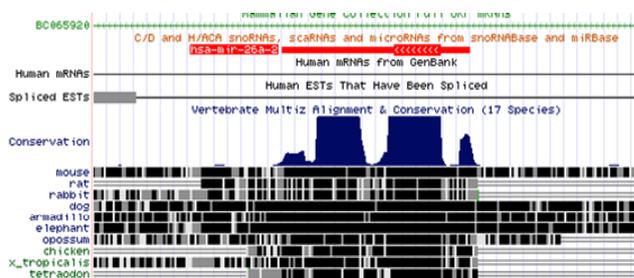
GUGGCCUCGUUCAAGUAAUCCAGGAUAGGCUGUGCAGGUCCAAUGGGCCUAUUCUUGGUUACUUGCACGGGGACGC  
(obtida por alinhamento da sequencia clonada com sequencia genômica. Pode ser realizado por Blat no site da UCSC).

3- Estrutura secundário do hsa-mir-26a



(pode ser obtida com o programa mfold (<http://bioweb.pasteur.fr/seqanal/interfaces/mfold-simple.html>))

4- Conservação entre espécies da região genômica que contem a sequencia do hsa-mir-26a



(obtido no site da UCSC)

**Figura 18.3:** Etapas envolvendo bioinformática para a validação de um novo microRNA.

A importância desta nova classe de pequenos RNAs não codantes para a produção animal foi elegantemente comprovado em um trabalho realizado pela equipe do Dr. Michel George, da Bélgica (Clop et al., 2006). Neste estudo, o QTL que controla o fenótipo de musculosidade observado em ovinos da raça Texel foi mapeado no cromossomo 2, em um intervalo que continha o gene GDF-8. No entanto, o seqüenciamento da região codante do GDF-8 não revelou polimorfismos nesta região. A mutação foi identificada na região 3' não traduzida (3'UTR) do GDF-8 e associada a característica de musculosidade observada em animais Texel. Os autores observaram que a mutação nesta região criava um sítio de interação com os microRNAs mir1 e mir206, normalmente expressos no tecido muscular. Estudos de imuno-deteção revelaram que animais portadores da mutação possuem um terço do nível de miostatina (GDF-8) observado no soro de animais normais. Estudos de expressão do GDF-8 em animais heterozigotos para a mutação revelaram que a abundância do RNAm do alelo normal da miostatina era 1,5 vezes superior a do

alelo mutante. Estes dois resultados estão de acordo com o mecanismo de ação proposto para os miRNA, ou seja, na regulação da tradução e da alteração da estabilidade do RNAm de um gene alvo.

Esta nova classe de RNA adiciona mais um nível de complexidade no já intrincado mecanismo de controle da expressão gênica, e ao mesmo tempo abre novas perspectivas para que possamos avançar o nosso conhecimento sobre as bases moleculares que controlam a variabilidade fenotípica observada nos seres vivos.

### **18.1. Protocolo de Isolamento de microRNAs (kit miRCat – Promega)**

#### ***1º/ 2º Dias – precipitação do RNA de alto peso molecular***

A partir de cerca de 500 µg de RNA total, fazer a precipitação do RNA de alto peso molecular.

Inicie com RNA total isolado com Trizol

Para precipitar o mRNA e rRNA (de alto peso molecular) adicione 50% PEG (MW 8000) para uma concentração final de 10% e 5M NaCl para uma concentração final de 0.5 M. Misture bem e coloque no gelo por 30 min.

Para 588 uL de RNA adicione 168 ul 50% PEG e 84 ul 5M NaCl.

Centrifugue na velocidade máxima (~12000 rpm) por 10 minutos a 4°C para precipitar o RNA de alto peso molecular.

Transfira o sobrenadante para dois tubos de microcentrífuga (420 ul para cada tubo) e adicione 2.5 volumes de etanol 100%. Incube a -80C por ao menos uma hora.

#### ***2º Dia – Isolamento dos microRNAs***

Aplicar em gel desnaturante de uréia/poliacrilamida 15%.

Eluir do gel o RNA contido na banda referente ao tamanho de cerca de 20 bases utilizando o *Flashpage*.

#### ***3º Dia – ligação do adaptador 3´ e gel***

##### **A. Ligação do adaptador 3´**

*\* Diluir a enzima na hora do uso em 10X Ligation Buffer (tubo 8).*

Em tubo 0,2, adicionar:

- 10  $\mu$ l do RNA recuperado
- 1  $\mu$ l do 3' RNA linker (50 $\mu$ M)
- 2  $\mu$ l 10X Ligation Buffer - **tubo 8**
- 6  $\mu$ l Ligation Enhancer - **tubo 9**
- 1  $\mu$ l T4 RNA Ligase (1U/ $\mu$ l)\*

Volume final = 20  $\mu$ l

Incubar a reação à 22°C por duas horas, então:

1. Adicionar 80 $\mu$ l IDTE (pH7.5) – **tubo 7**
2. Transferir todo o volume para tubo de 1.5ml
3. Adicionar 1,5  $\mu$ l glicogênio (20mg/ml)
4. Adicionar 1/10 volume (10 $\mu$ l) 3.0M NaOAc - **tubo 12**
5. Adicionar 2.5 volumes (250 $\mu$ l) -20°C 100% EtOH
6. Misturar bem por inversão
7. – 80°C, 30 minutos.
8. Centrifugar a 14.000 rpm, 4°C, 10 minutos.
9. Descartar o sorenadante.
10. Secar o pellet.
11. Ressuspender em 10 $\mu$ l IDT DNase/RNase free water.

*Se não for continuar a próxima etapa imediatamente, secar o pellet e estocar a -20°C.*

## **B. Gel**

Aplicar em gel desnaturante de uréia/poliacrilamida 10%.

Eluir do gel o RNA contido na banda referente ao tamanho de cerca de 40 bases utilizando o *Flashpage*.

## **4º Dia – ligação do adaptador 5' e RT**

### **A. Ligação do adaptador 5'**

Em tubo 0,2, adicionar:

- 10  $\mu$ l do RNA recuperado
- 1  $\mu$ l do 5' RNA linker (50 $\mu$ M)

- 2  $\mu$ l 10X Ligation Buffer - **tubo 8**
- 6  $\mu$ l Ligation Enhancer - **tubo 9**
- 1  $\mu$ l T4 RNA Ligase (5U/ $\mu$ L)
- 2  $\mu$ l ATP (10 mM)

Volume final = 20  $\mu$ L

Incubar a reação à 22°C por duas horas, então:

1. Adicionar 80  $\mu$ l IDTE (pH7.5) – **tubo 7**
2. Transferir todo o volume para tubo de 1.5ml
3. Adicionar 1,5  $\mu$ l glicogênio (20 mg/ml)
4. Adicionar 1/10 volume (10  $\mu$ l) 3.0 M NaOAc - **tubo 12**
5. Adicionar 2.5 volumes (250  $\mu$ l) -20°C 100% EtOH
6. Misturar bem por inversão
7. Incubar em – 80°C, 30 minutos.
8. Centrifugar a 14.000 rpm, 4°C, 10 minutos.
9. Descartar o sobrenadante.
10. Secar o pellet.
11. Ressuspender em 10 $\mu$ l água livre de DNase/RNase.

*Se não for continuar a próxima etapa imediatamente, secar o pellet e estocar a -20°C.*

## **B. Transcrição Reversa**

Adicionar em um tubo 0,2:

- 10  $\mu$ l do RNA recuperado.
- 1.0  $\mu$ l dNTPs (10mM).
- 1.0  $\mu$ l RT primer (10 $\mu$ M).
- 1.0  $\mu$ l de água.

Volume Final = 13.0  $\mu$ l

Incubar a 65°C, 5 minutos.

Colocar em gelo e adicionar:

- 4  $\mu$ l de 5X First Strand Buffer.

- 1  $\mu\text{l}$  0.1 M DTT.
- 1  $\mu\text{l}$  de RNase-OUT (40 U/ $\mu\text{l}$ ).
- 1  $\mu\text{l}$  de SuperScript III RT (200 U/ $\mu\text{l}$ ).

Volume Final = 20.0  $\mu\text{l}$

Incubar a 50 °C por uma hora e 15 minutos a 70 °C.

### **C. Reação em Cadeia da Polimerase (PCR)\***

*\*Fazer no dia em que for clonar!*

Adicionar:

- 3  $\mu\text{l}$  do produto da RT
- 5.0  $\mu\text{l}$  de 10x PCR Buffer
- 3.0  $\mu\text{l}$  de  $\text{MgCl}_2$  (3 mM)
- 1.0  $\mu\text{l}$  de Forward Primer (10 pmol)
- 1.0  $\mu\text{l}$  de Reverse Primer (10 pmol)
- 1.0  $\mu\text{l}$  de dNTPs (10 mM)
- 35.5  $\mu\text{l}$  de água
- 1  $\mu\text{l}$  de Taq polimerase (5 U/ $\mu\text{l}$ )

Volume Final = 50.0  $\mu\text{l}$

### **Condições da PCR:**

95.0° C por 10 minutos

35 ciclos { 95.0° C por 30 segundos  
52.0° C por 30 segundos  
72.0° C por 30 segundos

72.0° C por 10 minutos

## 19. PRÁTICA MINERAÇÃO DE DADOS EM BIOINFORMÁTICA

*André C P L F de Carvalho*

*Felipe R. da Silva*

A Bioinformática ou Biologia Computacional, área de pesquisa em grande expansão, utiliza técnicas e ferramentas computacionais para resolver problemas da Biologia. Nesta prática será mostrado como a Informática pode ser utilizada para resolver problemas de Biologia Molecular. A prática será dividida em duas partes. Na primeira parte, serão utilizadas técnicas de Mineração de Dados baseadas em Aprendizado de Máquina para analisar dados biológicos. Na segunda parte será dada uma visão completa do processo de se utilizar dados de projetos EST (*Expressed Sequence Tag*) na busca de genes diferencialmente expressos.

### 19.1. Mineração de Dados baseada em Aprendizado de Máquina

Mineração de Dados, em especial, técnicas de Aprendizado de Máquina, têm sido utilizadas com sucesso para resolver diversos problemas de Bioinformática, como reconhecimento de genes, previsão da estrutura de proteínas e análise de expressão gênica. Nesta prática, será mostrado como técnicas de Aprendizado de Máquina podem ser utilizadas para resolver problemas de Bioinformática, em particular reconhecimento de genes e análise de expressão gênica.

Neste módulo, 3 técnicas de Aprendizado de Máquina devem ser investigadas: Redes Neurais Artificiais (RNs), Máquinas de Suporte Vetorial (SVMs) e o algoritmo C4.5 para indução de árvores de decisão (AD).

As SVMs foram propostas na década de 60, porém suas primeiras aplicações surgiram apenas nos anos 90 com a proposta da Teoria do Aprendizado Estatístico. As SVMs pertencem ao paradigma de aprendizado estatístico e têm recebido grande atenção nos últimos anos na área de Bioinformática.

As RNs são modelos computacionais distribuídos baseados na estrutura e funcionamento do sistema nervoso. Elas têm sido aplicadas em uma enorme variedade de domínios, entre eles, a Bioinformática.

As Árvores de Decisão, outra técnica de AM, também têm sido utilizadas em problemas de classificação na área de Bioinformática. ADs tem sido muito utilizada

em problemas da biologia. Um dos trabalhos pioneiros foi o reconhecimento de regiões de splicing.

Neste módulo da prática, o seguinte roteiro será seguido:

- Baixar os dados do site do curso, os dados da UCI promoters e os três conjuntos de dados no arquivo fornecido;
- Preparar os dados para os experimentos. Para as técnicas que utilizam valores numéricos, converter as sequências de nucleotídeos em vetores numéricos, usando a codificação canônica ou topológica;
- Nos experimentos, particionar os dados utilizando a metodologia 10-fold cross-validation;
- Nos experimentos com as Redes Neurais do tipo MLP, treinar com os algoritmos de treinamento backpropagation e backpropagation momentum.
- Nos experimentos com ADs, selecionar algoritmo J48.
- Nos experimentos com as três técnicas de Aprendizado de Máquina (SVMs, ADs e Redes Neurais do tipo MLP):
  - Particionar os dados utilizando a metodologia 10-fold cross-validation;
  - Variar valores dos parâmetros para cada técnica de Aprendizado de Máquina e a arquitetura da Rede Neural;
    - Definir melhor conjunto de valores para os parâmetros.
  - Analisar medidas de acurácia preditiva e matriz de confusão;

Ao final dos experimentos, comparar desempenho das diferentes técnicas de Aprendizado de Máquina.

## **19.2. Busca de Genes diferencialmente expressos em projetos EST.**

Projetos EST amostram os genes expressos no instante em que o mRNA foi extraído. Usualmente estes projetos contêm diversas bibliotecas, cada uma delas feita a partir de mRNA extraído em uma determinada condição. Estas “condições”

incluem um tecido determinado, uma variedade/raça, um estágio de desenvolvimento, a exposição a fatores geradores de stress.

Uma propriedade simples que advém da metodologia é a de que só podemos ter amostrado determinado mRNA nos tecidos em que ele de fato é expresso. A recíproca, entretanto, não é verdadeira: o fato de não termos amostrado determinado mRNA em um tecido não é uma indicação inequívoca de que ele não se expressa nas condições em que o mRNA foi extraído.

Um problema real na análise de dados de um projeto EST deve-se ao fato de que (ao menos ao longo do projeto) gera-se apenas uma sequência de cada clone. Na maior parte das vezes, apenas parte da sequência total do inserto é revelada pela reação de sequenciamento (quando o inserto tem mais de 1000 pb, *e.g.*). Outro problema advém do fato de que nem todo clone corresponde à sequência completa do mRNA que o originou. Para completar, o processo de sequenciamento não gera sequências totalmente corretas. Há, sempre, uma porcentagem de erro em cada sequência.

Idealmente, o processo preliminar de análise de dados de um projeto EST lograria agrupar todas as sequências oriundas de um mesmo mRNA. Os problemas listados no parágrafo anterior, associados ainda a outros, como a existência de alelos, famílias multigênicas, domínios conservados e quimeras tornam muito difícil a realização desta tarefa de maneira perfeita. Precisa-se assumir, portanto, o resultado desta análise preliminar como uma boa aproximação.

Admitindo-se que o agrupamento de sequências tenha sido bem feito, as análises estatísticas da frequência de amostragem de transcritos esbarram ainda em (i) desvios de amostragem e (ii) tamanhos muito diferentes dos grupos amostrais.

### **19.3. Teste Exato de Fisher (Northern Blot eletrônico) implementado no WebEST**

Apesar dos gráficos de expressão relativa serem úteis para se ter uma idéia da diferença de expressão de um *Cluster* nas Bibliotecas, eles podem mostrar diferenças que não tem significância estatística, ou seja, para o tamanho da amostragem, aquela diferença pode ser apenas um artefato de amostragem, e a expressão das seqüências ser na verdade muito próxima. Para lidar com estes desvios de amostragem, inerentes a este tipo de projeto, foi desenvolvida uma forma

de aplicar o Teste Exato de Fisher (Fisher, 1922) para determinar quais diferenças são significativas.

Além disso, tratar diversas bibliotecas como se fossem uma só pode ser extremamente vantajoso para responder algumas perguntas biológicas. Por exemplo, o projeto genoma café tem 9 bibliotecas de *C. arábica* feitas a partir de folhas. Algumas induzidas por hormônios, outras não. Esta implementação do Teste Exato de Fisher permite que diversas bibliotecas sejam agrupadas. Desta forma, além de encontrar expressões significativamente distintas entre 2 bibliotecas, é possível encontrar expressões distintas entre 2 grupos de bibliotecas.

Ao ser acessada pela 1ª vez, a página do Teste Exato de Fisher mostra uma lista com todas as bibliotecas do projeto. Para aplicar o Teste, é necessário antes criar 2 grupos (algo como controle e tratamento). Para criar o 1º grupo, selecione uma ou mais bibliotecas da lista, digite um nome para o grupo no campo no topo da lista e clique no botão “Criar Grupo” no final da lista. O nome do grupo recém criado será mostrado no quadro do lado esquerdo. Use o mesmo procedimento para criar o 2º grupo. É possível a criação de vários grupos, mas apenas 2 podem ser comparados de cada vez. Os ícones à direita do nome do grupo permitem (i) visualizar o grupo (listando as bibliotecas incluídas e gerando um relatório como se o grupo todo fosse uma única biblioteca), (ii) editar o grupo (alterando seu nome ou as bibliotecas que o compõem) e (iii) descartar o grupo (apagar este grupo de sua lista de grupos).

Para realizar o Teste Exato de Fisher, selecione “Comparar Grupos” no quadro da esquerda e selecione quais 2 grupos serão comparados. O campo “Divisor:” permite escolher a significância da comparação. Quanto menor o número, mais significativa a diferença. Após escolher os 2 grupos clique no botão “Comparar Grupos” para efetivamente iniciar o Teste.

O resultado do Teste são os *Clusters* nos quais a distribuição dos *reads* dos 2 grupos foge da “regra geral”. Ou, no jargão biológico, quais genes estão super-expressos em um grupo quando comparado com o outro.

Note, por favor, que dependendo do tamanho dos grupos podem ser necessários mais de 10 minutos para completar a operação. Parece muito tempo? Pois dê uma olhada na fórmula do Teste. Lembre-se que ela tem que ser calculada para cada um dos clusters. Ainda não parece muito? Tente calcular o

fatorial de 30.000 na calculadora do seu computador e veja o tempo necessário para esta operação! Enfim, tenha consciência de que este Teste consome muito recurso da máquina (que é compartilhada por você e todos os outros usuários). Aguarde a resposta do seu teste. Uma vez com o resultado na sua frente, use etiquetas para identificar os clusters que são interessantes para você. Evite, sempre que possível, repetir o mesmo teste.

Por último, os grupos criados, modificados ou apagados nesta ferramenta devem ser salvos antes do encerramento da sessão (antes de fechar o navegador), ou serão perdidos. Utilize a opção *Gravar Grupos* e confirme sua opção na nova página.

#### 19.4. Explicação técnica do Teste Exato de Fisher.

Uma descrição sucinta do Teste Exato de Fisher pode ser encontrada na Wikipedia: [http://en.wikipedia.org/wiki/Fisher's\\_exact\\_test](http://en.wikipedia.org/wiki/Fisher's_exact_test). Outra descrição curta, porém mais técnica, encontra-se em <http://www.quantitativeskills.com/sisa/statistics/fishrhlp.htm>. O trecho a seguir foi extraído da monografia de conclusão de curso de Raphael Melo de Oliveira Bastos Sales, que implementou o teste empregado pelo sistema.

O Teste de Fisher é um teste de significância estatística usado para analisar dados categóricos quando o tamanho da amostra é pequena. Dada uma tabela A:

	Grupo A	Grupo B	Total
Condição A	a	b	a + b
Condição B	c	d	c + d
Totais	a + c	b + d	n

E usando a fórmula: 
$$p = \frac{(a+b)!(c+d)!(a+c)!(b+d)!}{n!a!b!c!d!}$$

Obtém-se a probabilidade da diferença se repetir na hipótese nula de 'a', 'b', 'c', e 'd' serem iguais. Quanto menor essa probabilidade, maior a diferença real entre esses valores, permitindo determinar diferenças significativas em condições de amostragem pequena ou insuficiente. As probabilidades aceitas são aquelas que

estão abaixo de uma linha de corte. No caso, essa linha de corte é calculada através da seguinte fórmula:

C = total de *Clusters* na análise

P = Total de *Pools* na análise

$$p \leq \frac{0,05}{n} \text{ onde } G = \frac{C \times P \times (P-1)!}{2}$$

Se tivermos, por exemplo, um *Cluster* com a seguinte distribuição de expressão:

	# de <i>Reads</i> desse <i>Cluster</i>	Outros <i>Reads</i>	Total
Biblioteca A	192	2578	2770
Biblioteca B	5	2842	2847
Totais	197	5420	5617

A probabilidade de o *Cluster* ter expressão igual em ambas as bibliotecas é:

$$a = 192$$

$$c = 5$$

$$n = 5617$$

$$b = 2578$$

$$d = 2842$$

$$G = 2698$$

$$p = \frac{(192 + 2578)! (5 + 2842)! (192 + 5)! (2578 + 2842)!}{5617! 192! 2578! 5! 2842!} \cong 3e^{-68} < \frac{0,05}{2698} \cong 2e^{-5}$$

Como p é menor que a linha de corte, esse *Cluster* é admitido como tendo diferença de expressão estatisticamente significativa entre as duas bibliotecas consideradas.

## 20. PREPARO DAS SOLUÇÕES

### Tris.HCl [tris(hydroxymethyl)aminomethane] - 1M

Dissolva 121g de Tris Base em 800 mL de água MiliQ

Ajuste o pH desejado com HCl concentrado

Adicione água até 1L

Autoclave

Cerca de 70 mL de HCl são necessários para pH de 7,4 ; e cerca de 42 mL para pH de 8,0.

*Nota importante: O pH dos tampões com Tris muda significativamente com a temperatura, caindo cerca de 0,028 para elevação de cada 1°C. O pH das soluções tamponadas com Tris deve ser ajustado na temperatura na qual as soluções serão usadas. Como o pK do Tris é 8,08, o tampão não deve ser usado a pH abaixo 7,2 nem acima de 9,0. Não adicione DEPC (Diethyl Pyrocarbonate) em soluções com Tris, porque o DEPC inativa essa substância.*

### EDTA (ethylenediamine tetraacetic acid) - 0,5M (pH 8,0)

Dissolva 186,1g de Na<sub>2</sub>EDTA.2H<sub>2</sub>O em 700 mL de água

Ajuste o pH para 8,0 com NaOH 10M (cerca de 50 mL)

Adicione água MiliQ para 1 L.

Autoclave

### TE pH 7,6 (tampão Tris/EDTA)

Reagente	Concentração final
Tris HCl pH 7,6	10mM
EDTA pH 8,0	1mM
<b>Volume para uma amostra</b>	<b>510 µL</b>

### SOLUÇÃO DE RNase (10mg/mL)

Dissolver a enzima Ribonuclease A livre de DNase (Sigma R- 6513) em solução Tris.HCl 10mM (pH 7.5), NaCl 15mM.

Armazenar a -20°C.

**TBE(Tris/Borate/EDTA) 10X – Tampão de Eletroforese****Solução Estoque**

Reagente	Quantidade
Tris Base (890 mM)	108 g
Ácido bórico (890 mM)	55 g
EDTA 0,5M pH 8,0 (20 mM)	40 mL
Água deionizada	Completar para 1L
<b>Total</b>	<b>1L</b>

**Solução A (Tampão de Hemólise)**

Para 1 amostra (25mL):

Reagente	Concentração final	X 1 amostra
Tris.HCl pH 7,6 1M	10mM	250 µL
MgCl <sub>2</sub> 0,5M	5mM	250 µL
NaCl 5M	10mM	50 µL
Água deionizada		Completar para 25 mL

**Solução B (Solução de lise)**

Reagente	Volume x 1 amostra	Concentração final
Tris.HCl pH 8,0 1 M	5µL	10mM
NaCl 5 M	10µL	100mM
EDTA pH 8,0 0,5 M	10µL	10mM
SDS 20%	12,5µL	0,5%
Proteinase K (20 mg/mL)	2,5µL	
Água deionizada	460,5 µL	
<b>Volume final</b>	<b>500 µL</b>	

**Água DEPC (dietilpirocarbonato):**

Adicionar 100 µL de DEPC para cada 100 mL de água deionizada.

Agitar bem e incubar a 37°C por 12 horas.

Autoclavar a água DEPC por 45 minutos a 1 atm.

### **SOLUÇÃO DE PROTEINASE K (20mg/mL)**

Dissolver 100 mg da enzima em 5 mL de solução Tris-HCl 10mM (pH 7.5),

Cloreto de Cálcio 20mM, Glicerol 50%.

Armazenar a -20°C.

### **Loading Buffer 6 x**

Bromophenol Blue 0,25%

Xileno Cianol 0,25%

Glicerol 30% em água (estocar a 4°C).

## 21. REFERÊNCIAS BIBLIOGRÁFICAS

ALVARES, L. E. Aspectos práticos da RT-PCR: extração de RNA, síntese de cDNA e PCR. In: **Biologia molecular aplicada à produção animal**. Embrapa. 2001.

AMBROS,V., BARTEL,B., BARTEL,D.P., BURGE,C.B., CARRINGTON,J.C., CHEN,X., DREYFUSS,G., EDDY,S.R., GRIFFITHS-JONES,S., MARSHALL,M., MATZKE,M., RUVKUN,G., AND TUSCHL,T.. A uniform system for microRNA annotation. *RNA*. 9, 277-279, 2003.

AUSUBEL, F. M. **Current protocols in molecular biology**. John Wiley & Sons, v. 1, 1987.

AUSUBEL, F. M. **Current protocols in molecular biology**. John Wiley & Sons. 1994.

AZEVEDO, M. O.; FELIPE, M. S. S.; BRÍGIDO, M. M.; MARANHÃO, A. Q.; DE SOUZA, M. T. **Técnicas básicas em biologia molecular**. Brasília: UnB, 2003. 211p.

BARAD,O., MEIRI,E., AVNIEL,A., AHARONOV,R., BARZILAI,A., BENTWICH,I., EINAV,U., GILAD,S., HURBAN,P., KAROV,Y., LOBENHOFER,E.K., SHARON,E., SHIBOLETH,Y.M., SHTUTMAN,M., BENTWICH,Z., AND EINAT,P.. MicroRNA expression detected by oligonucleotide microarrays: system establishment and expression profiling in human tissues. **Genome Res**. 14, 2486-2494, 2004.

BARTLETT, J. M. S.; STIRLING, D. Methods in molecular biology. In: **PCR Protocols**, v. 226, 2<sup>a</sup>. ed. Humana Press Inc., 2003.

BOCK, R. Biolistic transformation of plants. **Qiagen News**. Issue N. 5 . 1997

BRAMMER, A. P. **Atualizações em técnicas celulares e moleculares aplicadas ao melhoramento genético vegetal**. 1<sup>a</sup> ed. Passo fundo: Embrapa Trigo, 2002.

CAI,X., HAGEDORN,C.H., AND CULLEN,B.R. Human microRNAs are processed from capped, polyadenylated transcripts that can also function as mRNAs. **RNA**. 10, 1957-1966, 2004.

CALSA JUNIOR, T.; BENEDITO, V. A.; FIGUEIRA, A. V. O. Análise serial da expressão gênica. **Revista Biotecnologia Ciência e Desenvolvimento**, Brasília, n. 33, p. 86-98, 2004.

CHEN,C., RIDZON,D.A., BROOMER,A.J., ZHOU,Z., LEE,D.H., NGUYEN,J.T., BARBISIN,M., XU,N.L., MAHUVAKAR,V.R., ANDERSEN,M.R., LAO,K.Q., LIVAK,K.J., AND GUEGLER,K.J.. Real-time quantification of microRNAs by stem-loop RT-PCR. **Nucleic Acids Res**. 33, e179, 2005

CHEN,C.Z. AND LODISH,H.F. MicroRNAs as regulators of mammalian hematopoiesis. **Semin. Immunol**. 17, 155-165, 2005.

CHOMCZYNSKI, P.; SACCHI, N. Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. **Analytical Biochemistry**, v. 162, n. 1, p. 156-159, 1987.

CHURCHILL, G. A.; DOERGE, R. W. Empirical threshold values for quantitative trait mapping. **Genetics**, v. 138, p. 963-971, 1994.

CLOP,A., MARCQ,F., TAKEDA,H., PIROTTIN,D., TORDOIR,X., BIBE,B., BOUIX,J., CAIMENT,F., ELSEN,J.M., EYCHENNE,F., LARZUL,C., LAVILLE,E., MEISH,F., MILENKOVIC,D., TOBIN,J., CHARLIER,C., AND GEORGES,M.. A mutation creating a potential illegitimate microRNA target site in the myostatin gene affects muscularity in sheep. **Nat. Genet**. 38, 813-818, 2006.

DIEFFENBACH, C.W. ; LOWE, T.M.J.; DVEKSLER, G.S. General Concepts for PCR Primer **Design**. **Genome Res**. v. 3: p. S30-S37. 1993

DORAK, M.T. Real time PCR. <http://www.dorak.info/genetics/realtime.html>. acesso em 08 de maio de 2009.

DUGGAN, D. J.; BITTNER, M.; CHEN, Y.; MELTZER, P.; TRENT, J. M. Expression profiling using cDNA microarrays. **Nature Genetics**, London, v. 21, sup. 1, p. 10-14, 1999.

FALCONER, D.S.; MACKAY, T.F.C. **Introduction to Quantitative Genetics**. 4 ed. Longman: New York. 1996. 464p.

FARH,K.K., GRIMSON,A., JAN,C., LEWIS,B.P., JOHNSTON,W.K., LIM,L.P., BURGE,C.B., AND BARTEL,D.P.. The widespread impact of mammalian MicroRNAs on mRNA repression and evolution. **Science**, 310, 1817-1821, 2005.

FISHER, R.A.. "On the interpretation of  $\chi^2$  from contingency tables, and the calculation of P". **Journal of the Royal Statistical Society**. 85(1):87-94, 1922.

GIRALDEZ,A.J., CINALLI,R.M., GLASNER,M.E., ENRIGHT,A.J., THOMSON,J.M., BASKERVILLE,S., HAMMOND,S.M., BARTEL,D.P., AND SCHIER,A.F. MicroRNAs regulate brain morphogenesis in zebrafish. **Science**, 308, 833-838, 2005.

GIRALDEZ,A.J., MISHIMA Y, RIHEL J, GROCOCK,R.J., VAN DONGEN,S., INOUE K, ENRIGHT,A.J., AND SCHIER,A.F. Zebrafish MiR-430 promotes deadenylation and clearance of maternal mRNAs. **Science**, 312, 75-79, 2006.

GREEN, P.; FALLS, K.; CROOKS, S. **CRI-MAP Documentation version 2.4**. University of Washington. 1990. Disponível em: <http://linkage.rockefeller.edu/soft/crimap/>.

HAL, N. L. W. van; VORST, O.; VAN HOUWELINGEN, A.; KOK, E. J.; PEIJNENBURG, A.; AHARONI, A.; VAN TUNEN, A.; KEIJER, J. The application of DNA microarrays in gene expression analysis. **Journal of Biotechnology**, Amsterdam, v. 78, n. 3, p. 271-280, 2000.

HALEY, C. S.; KNOTT, S. A.; ELSEN, J. M. Mapping quantitative trait loci in crosses between outbred lines using least squares. **Genetics**, v. 136, p. 1195-1207, 1994.

HARRINGTON, C. A.; ROSENOW, C.; RETIEF, J. Monitoring gene expression using DNA microarrays. **Current Opinion in Microbiology**, Oxford, v. 3, n. 3, p. 285-291, 2000.

HERNÁNDEZ-SÁNCHEZ, J.; GRUNCHEC, J.A.; KNOTT, S. A web application to perform linkage disequilibrium and linkage analyses on a computational grid. **Bioinformatics**, 25: 1377–1383, 2009.

<http://www.dbbm.fiocruz.br/helpdesk/mbiology/historico2004.pdf>

<http://www.dnai.org>

<http://www.etail.hpg.ig.com.br/cursopcr.html>

KIM,V.N. AND NAM,J.W. Genomics of microRNA. **Trends Genet.** 22, 165-173, 2006.

KOSAMBI, D.D. The estimation of map distance from recombination values. **Annals of Eugenics**, 12: 172-175, 1944.

KWOK, S., D.E. KELLOG, N. MCKINNEY, D. SPASIC, L. GODA, C. LEVENSON, AND J.J. SNINSKY. Effects of primer-template mismatches on the polymerase chain reaction: Human immunodeficiency virus 1 model studies. **Nucleic Acids Res.** 18: 999-1005. 1990.

LANDER, E. S.; BOTSTEIN, D. Mapping Mendelian factors underlying quantitative traits using RFLP linkage maps. **Genetics**, v. 121, p. 185-199, 1989.

LECELLIER,C.H., DUNOYER,P., ARAR,K., LEHMANN-CHE,J., EYQUEM,S., HIMBER,C., SAIB,A., AND VOINNET,O. A cellular microRNA mediates antiviral defense in human cells. **Science**, 308, 557-560, 2005.

LEE RC, FEINBAUM,R., AND AMBROS,V. Heterochromic gene lin-4 encodes small RNAs with antisense complementarity to lin-14. **Cell**, 75, 843-854, 1993.

LEE,Y., AHN,C., HAN,J., CHOI,H., KIM,J., YIM,J., LEE,J., PROVOST,P., RADMARK,O., KIM,S., AND KIM,V.N. The nuclear RNase III Drosha initiates microRNA processing. **Nature**, 425, 415-419, 2003.

LEE,Y., JEON,K., LEE,J.T., KIM,S., AND KIM,V.N. MicroRNA maturation: stepwise processing and subcellular localization. **EMBO J.**, 21, 4663-4670, 2002.

LIM,L.P., LAU,N.C., GARRETT-ENGELE,P., GRIMSON,A., SCHELTER,J.M., CASTLE,J., BARTEL,D.P., LINSLEY,P.S., AND JOHNSON,J.M. Microarray analysis shows that some microRNAs downregulate large numbers of target mRNAs. **Nature**, 433, 769-773, 2005.

LIU, B.H. **Statistical Genomics: Linkage, Mapping and QTL Analysis**. Boca Raton: CRC, 1998. 648p.

LIU,C.G., CALIN,G.A., MELOON,B., GAMLIEL,N., SEVIGNANI,C., FERRACIN,M., DUMITRU,C.D., SHIMIZU,M., ZUPO,S., DONO,M., ALDER,H., BULLRICH,F., NEGRINI,M., AND CROCE,C.M. An oligonucleotide microchip for genome-wide microRNA profiling in human and mouse tissues. **Proc. Natl. Acad. Sci. U. S. A.**, 101, 9740-9744, 2004 .

LUND,E., GUTTINGER,S., CALADO,A., DAHLBERG,J.E., AND KUTAY,U. Nuclear export of microRNA precursors. **Science**, 303, 95-98, 2004.

MANIATIS, T.; FRISCH, E. F.; SAMBROOCK, J. **Molecular cloning: a laboratory manual**. Cold Spring Harbor Laboratory, 1987.

**Manual do usuário.** *SuperScript® First-Strand Synthesis System for RT-PCR.* Catálogo 11904-018. version E. Invitrogen, 2003.

MASEKA, T.; VOPALENSKYA, V.; SUCHOMELOVAB, P.; POSPISEKA, M. Denaturing RNA electrophoresis in TAE agarose gels. **Analytical Biochemistry**, v. 336, p. 46-50, 2005.

MCCARTHY JJ AND ESSER KA. MicroRNA-1 and MicroRNA-133a expression are decreased during skeletal muscle hypertrophy. **J. Appl Physiol.**, 102, 306-313, 2007.

MEABURN, E.; BUTCHER, L. M.; LIU, L.; FERNANDES, C.; HANSEN, V.; AL-CHALABI, A.; PLOMIN, R.; CRAIG, I.; SCHALKWYK, L. C. Genotyping DNA pools on microarrays: tackling the QTL problem of large samples and large numbers of SNPs. **BMC Genomics**, v. 6, p. 52, 2005.

NILSEN, T.W. Mechanisms of microRNA-mediated gene regulation in animal cells. **Trends Genet.** 23, 243-249, 2007.

OKUBO, K.; HORI, N.; MATOBA, R.; NIIYAMA, T.; FUKUSHIMA, A.; KOJIMA, Y.; MATSUBARA, K. Large scale cDNA sequencing for analysis of quantitative and qualitative aspects of gene expression. **Nature Genetics**, v. 2, p. 173-179, 1992.

PATEL, D. **Gel eletrophoresis: essential data.** Wiley, 1994

POY, M.N., ELIASSON, L., KRUTZFELDT, J., KUWAJIMA, S., MA, X., MACDONALD, P.E., PFEFFER, S., TUSCHL, T., RAJEWSKY, N., RORSMAN, P., AND STOFFEL, M. A pancreatic islet-specific microRNA regulates insulin secretion. **Nature**, 432, 226-230, 2004.

REGITANO, L. C. A. Extração de DNA para aplicação em reação em cadeia da polimerase (PCR). In: **Biologia molecular aplicada à produção animal.** Embrapa. 2001.

ROCHE. 1,4-Dithiothreitol. **Roche Applied Science**, 2004.

ROMANO, E.; BRASILEIRO, A. C. M. Extração de DNA de plantas. **Biotecnologia, Ciência e Desenvolvimento**, p. 40-43, 1999.

SAIKI, R. K.; GELFAND, D. H.; STOFFEL, S.; SCHARF, S. J.; HIGUCHI, R.; HORN, G. T.; MULLIS, K. B.; ERLICH, H. A. Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. **Science**, v. 239, p. 487-491, 1988.

SAMBROOK, J. E.; RUSSEL, D. W. **Molecular cloning: a laboratory manual**. 3 ed. New York: Cold Spring Harbor Laboratory Press, 2002. 2700 p.

SCHWARZ,D.S., HUTVAGNER,G., DU,T., XU,Z., ARONIN,N., AND ZAMORE,P.D. Asymmetry in the assembly of the RNAi enzyme complex. **Cell**, 115, 199-208, 2003.

SEATON, G.; HALEY, C. S.; KNOTT, S. A.; KEARSEY, M.; VISSCHER, P. M. QTL Express: mapping quantitative trait loci in simple and complex pedigrees. **Bioinformatics**, v. 18, p. 339-340, 2002.

SEATON, G.; HERNANDEZ, J.; GRUNCHEC, J.A.; WHITE, I.; ALLEN, J.; DE KONING, D.J.; WEI, W.; BERRY, D.; HALEY, C.; KNOTT, S. GridQTL: a grid portal for QTL mapping of compute intensive datasets. 8<sup>th</sup> World Congress on Genetics Applied to Livestock Production. **Proceedings**. Belo Horizonte, Brazil, 2006. 1 CD-ROM.

SHAFFER, L. G.; BEJJANI, B. A. Medical applications of array CGH and the transformation of clinical cytogenetics. **Cytogenetical Genome Research**, v. 115, n. 3-4, p. 303-309, 2006.

SHINGARA,J., KEIGER,K., SHELTON,J., LAOSINCHAI-WOLF,W., POWERS,P., CONRAD,R., BROWN,D., AND LABOURIER,E. An optimized isolation and labeling platform for accurate microRNA expression profiling. **RNA**. 11, 1461-1470, 2005.

SMALHEISER, N.R. EST analyses predict the existence of a population of chimeric microRNA precursor-mRNA transcripts expressed in normal human and mouse tissues. **Genome Biol.** 4, 403, 2003.

SOUTHERN, E.; MIR, K.; SHCHEPINOV, M. Molecular interactions on microarrays. **Nature Genetics**, v. 21, p. 5-9, 1999.

STERKY, F.; LUNDEBERG, J. Sequence analysis of genes and genomes. **Journal of Biotechnology**, v. 76, p. 1-31, 2000.

TANG F, KANEDA M, O'CARROLL D, HAJKOVA P, BARTON SC, SUN YA, LEE C, TARAKHOVSKY A, LAO, K.Q., AND SURANI MA. Maternal microRNAs are essential for mouse zygotic development. **Genes Dev.** 21, 644-648, 2007.

TaqMan® EZ RT-PCR Kit, Applied Biosystems, 2002.  
[http://www3.appliedbiosystems.com/cms/groups/mcb\\_support/documents/generaldocuments/cms\\_040931.pdf](http://www3.appliedbiosystems.com/cms/groups/mcb_support/documents/generaldocuments/cms_040931.pdf) . acesso em 08 de maio de 2009.

VISSCHER, P. M.; THOMPSON, R.; HALEY, C. S. Confidence intervals in QTL mapping by bootstrapping. **Genetics**, v. 143, p. 1013-1020, 1996.

WESTERMEIER, R. **Electrophoresis in practice: a guide to methods and applications of DNA and protein separations**. Wiley-VCH, 2005.

YI, R., QIN, Y., MACARA, I.G., AND CULLEN, B.R.. Exportin-5 mediates the nuclear export of pre-microRNAs and short hairpin RNAs. **Genes Dev.** 17, 3011-3016, 2003.

ZHAO, Y., SAMAL, E., AND SRIVASTAVA, D.. Serum response factor regulates a muscle-specific microRNA that targets Hand2 during cardiogenesis. **Nature**, 436, 214-220, 2005.

## 22. LITERATURA CONSULTADA

ALCAMO, I. E. **DNA technology: the awesome skill**. 2<sup>a</sup>. ed. San Diego: Elsevier Academic Press, 2001. 348 p.

**ANSWERS.COM health**. Disponível em:< <http://www.answers.com/topic/molecular-biology?cat=health>>. Acesso em: 25 de junho de 2007.

CALLADINE, C. R.; DREW, H. R.; LUISI, B. F.; TRAVERS, A. A. **Understanding DNA: the molecule & how it works**. 3<sup>a</sup>. ed. San Diego: Elsevier Academic Press, 2004. 352 p.

EUN, H. M. **Enzymology primer for recombinant DNA technology**. San Diego: Elsevier Academic Press, 1996. 702 p.

MANIATIS, T.; FRITSCH, E. F.; SAMBROCK, J. **Molecular cloning: a laboratory manual**. New York: Cold Spring Harbor Laboratory, 2001. 545 p.

SILVA JÚNIOR, J. G. da **Eletroforese de proteínas: guia teórico e prático**. Rio de Janeiro: Interciência Ltda, 2001.

**WIKIPEDIA a enciclopédia livre**. Disponível em:

<[http://pt.wikipedia.org/wiki/Biologia\\_molecular](http://pt.wikipedia.org/wiki/Biologia_molecular)>. Acesso em: 25 de junho de 2007.