

Laboratório de Genética Molecular - Nupgen - DBC/Nupélia

Índice

Início	5
1 O Laboratório	6
1.1 Apresentação do Laboratório	6
1.2 Linhas de Pesquisa	6
1.2.1 Taxonomia	7
1.2.2 Sistemática filogenética	7
1.2.3 Genética de populações	7
1.2.4 Biogeografia	7
1.3 Projetos do Nupgen	8
1.3.1 Peixes	8
1.3.2 Macrófitas	8
1.3.3 Parasitos	8
1.3.4 Microorganismos	8
1.3.5 Genes opsins	8
1.3.6 DNA Ambiental	9
2 Genética Molecular e Marcadores	10
2.1 O que é a genética molecular?	10
2.2 Marcadores Moleculares	10
I Protocolos	11
3 Diluições	12
3.1 Etanol	12
3.1.1 Etanol 70%	12
3.1.2 Etanol 80%	12
3.2 GelRed	13
3.3 Ladder 100pb	13
3.4 Lambda	13
3.4.1 Lambda 5ng/ L (para 1000 L)	13
3.4.2 Lambda 10ng/ L (para 1000 L)	13
3.4.3 Lambda 20ng/ L (para 1000 L)	13
3.5 Primer para uso (10 M)	14

3.6	Proteinase K 20 mg/mL	14
3.7	Tampão TBE 5x (estoque)	14
3.8	Tampão TBE 1x (corrida)	14
3.9	PEG 20% NaCl 2,5M (para 75 mL)	15
3.10	Diluição primer liofilizado	15
4	Extração	16
4.1	Protocolo de Extração de DNA - Promega Peixes	16
4.2	Protocolo de Extração de DNA - Promega Caramujos	18
4.3	Protocolo de Extração de DNA - Promega Macrófitas	19
4.4	Protocolo de Extração de DNA com Colunas - QIAGEN Parasitos	20
5	Eletroforese	22
5.1	Ladder	23
5.2	Lambda	24
5.3	Gel agarose	25
5.3.1	Gel 1%	25
5.3.2	Gel 1,4%	25
5.3.3	Gel 3%	25
5.4	Protocolo para quantificação	26
5.5	Protocolo para amplificação	26
5.6	Protocolo para purificação	26
5.7	Preparação para eletroforese	26
5.7.1	Ladder 100 pb	26
5.7.2	Por amostra	26
6	Quantificação	27
6.1	Preparação por amostra	28
6.1.1	GelRed	29
6.1.2	Safer	29
6.2	Diluição do DNA para 5ng/ L	29
7	Amplificação	30
8	Purificação	34
8.1	Purificação com Polietilenoglicol (PEG)	34
9	Sequenciamento	36
II	Análises	39
10	Bancos de Dados Genéticos	40
10.1	GenBank	40

10.2 BOLD Systems	40
10.3 BLAST	41
11 Softwares	42
11.1 BioEdit	42
11.2 MEGA	43
11.3 IQTree	43
11.4 jModelTest	44
11.5 RAxML	44
11.6 PhyML	45
11.7 DNAsp	45
11.8 Delimitadores de espécies	46
11.8.1 PTP	46
11.8.2 ABGD	46
11.8.3 ASAP	47
11.8.4 GMYC	47
11.9 FigTree	47
11.10iTree	48
11.11PopArt	48
12 Pacotes R	49
12.1 Pacote nupgen (R)	49
12.2 Rocc	50
12.3 bold	50
12.4 refdb	50
12.5 Herodotools	51
12.6 ggtree	51
12.7 ggmsa	51
13 Edição de sequências	53
13.1 Passo a passo para edição de sequências	53
14 Alinhamento	55
Referências	56

Início

Este é o livro online do **Nupgen**, o **Laboratório de Genética Molecular do Nupélia (DBC/Nupélia)**. Aqui você poderá encontrar textos explicativos sobre nossa linha de pesquisa, as metodologias utilizadas em nosso laboratório, bem como os protocolos atualizados. Este é um material introdutório destinado principalmente aos alunos do laboratório e pode servir como guia para as tarefas mais técnicas. Esperamos que ele funcione como um Procedimento Operacional Padrão (POP) em nosso laboratório, mantendo as equipes alinhadas, elevando a produtividade e evitando o desperdício de recursos dentro do laboratório. Este livro não precisa ser lido de forma sequencial e cada capítulo é independente de outros, no entanto, organizamos os capítulos seguindo a lógica de trabalho no laboratório para tornar a leitura mais fluida.

O objetivo deste livro não é substituir a colaboração entre alunos do laboratório, mas sim garantir que pequenos detalhes não sejam perdidos com o passar do tempo. Na verdade, esperamos que os tópicos deste livro sejam exaustivamente discutidos entre os alunos, a fim de melhorar este conteúdo e gerar novas perguntas que ajudem no desenvolvimento da pesquisa.

Este livro foi escrito em [Quarto](#).

Organização: Bruno Miotto

Autores: Adrian Silva, Bárbara Scorsim, Bruno Miotto, Eloisa de Páris e Mayara Passere

1 O Laboratório

1.1 Apresentação do Laboratório

O Nupgen é o **Laboratório de Genética Molecular do Nupélia** (Núcleo de Pesquisas em Limnologia, Ictiologia e Aquicultura), na **Universidade Estadual de Maringá (UEM)**, Paraná, Brasil. Coordenado pela Professora Dra. **Alessandra Valéria de Oliveira**, o Nupgen conta com pesquisadores, alunos de graduação e de pós-graduação que desenvolvem pesquisas objetivando detectar polimorfismos genéticos em populações naturais.



1.2 Linhas de Pesquisa

Atualmente, o Nupgen aborda estudos relacionados a polimorfismo genético sob o escopo de uma ampla variedade de organismos: peixes, parasitas, macrófitas, macroinvertebrados,

microorganismos, além de estudos utilizando DNA ambiental. Assim, os pesquisadores e alunos do laboratório conduzem as seguintes linhas de pesquisa:

1.2.1 Taxonomia

Consiste na identificação dos organismos utilizando marcadores moleculares. Projetos nessa linha de pesquisa costumam ocorrer em parceria com outros laboratórios, que contribuem com a identificação morfológica do organismo. A adição da genética molecular facilita a detecção de espécies novas, híbridos e espécies crípticas, por exemplo;

1.2.2 Sistemática filogenética

Diferente da taxonomia, projetos envolvendo Sistemática Filogenética não buscam somente identificar os organismos, mas também levantar hipóteses sobre a relação evolutiva entre eles (entre espécies de um mesmo gênero, por exemplo).

Exemplo de questão: elucidar a relação filogenética em um gênero cuja evolução das espécies seja pouco compreendida (resolução de politomias);

1.2.3 Genética de populações

A Genética de populações busca entender o fluxo gênico em uma população natural, podendo endereçar questões acerca de processos de especiação e adaptação, e podendo observar os efeitos da seleção natural, deriva genética e dos efeitos de gargalo.

Exemplo de questão: diversidade genética em uma população de peixes invasores;

1.2.4 Biogeografia

Biogeografia é uma linha de pesquisa multidisciplinar que busca entender a distribuição de um grupo de organismos em correlação com a evolução ambiental (escala de tempo geológica), gerando resultados importantes para a conservação da biodiversidade.

Exemplo de questão: detectar áreas de endemismo para macrófitas aquáticas.

1.3 Projetos do Nupgen

1.3.1 Peixes

Ictiofaunas de reservatórios aquáticos podem ser formadas por organismos nativos e invasores. O estudo de populações nativas é importante para verificar a presença e interação desses organismos de determinada região. E invasores, quando presentes, modificam o habitat e o sistema onde foram introduzidos, podendo levar a perda de biodiversidade local e alterações genéticas. Assim, estudos moleculares permitem estimar a diversidade e verificar as alterações genéticas que ocorrem nestes animais.

1.3.2 Macrófitas

As macrófitas aquáticas, quando invasoras, constituem uma ameaça em termos ecológicos e também econômicos. A correta identificação e a elucidação da diversidade genética desses organismos podem auxiliar na compreensão dos mecanismos de dispersão e colonização, como também no manejo adequado das espécies.

1.3.3 Parasitos

A relação parasito-hospedeiro possui grande importância nos ambientes naturais. O conhecimento da biodiversidade de parasitos contribui com o levantamento de espécies de hospedeiros no local, além de direcionar métodos de manejo para parasitos com potencial zoonótico.

1.3.4 Microorganismos

Mesmo invisíveis a olho nu, os microorganismos possuem grande importância em ambientes aquáticos, servindo como níveis inferiores na cadeia alimentar. Alguns organismos podem influenciar na qualidade d'água, por isso conhecer a diversidade destes microorganismos é fundamental

1.3.5 Genes opsins

O estudo de proteínas visuais possibilita observar e comparar a influência de fatores ecológicos no processo evolutivo de diversas linhagens, identificando como as espécies se adaptam a diversas pressões seletivas existentes no habitat.

1.3.6 DNA Ambiental

A partir de amostras de água, é possível obter fragmentos de DNA de organismos que estiveram neste ambiente. Peixes, microorganismos, anfíbios, parasitos e muitos outros organismos deixam rastros genéticos e podemos estudá-los indiretamente a partir de técnicas de sequenciamento de nova geração.

2 Genética Molecular e Marcadores

2.1 O que é a genética molecular?

A Genética Molecular utiliza o DNA como objeto de estudo. O DNA (ácido desoxirribonucleico) é uma biomolécula capaz de armazenar a informação biológica e perpetuá-la pelo processo de auto replicação semiconservativa. Sua estrutura é uma dupla-fita helicoidal de nucleotídeos, que por sua vez é composta por um grupo fosfato, uma pentose e uma base nitrogenada. As duas fitas de DNA são unidas por meio de ligações de hidrogênio entre as bases nitrogenadas. Existem quatro bases nitrogenadas: Adenina (A), Timina (T), Citosina (C) e Guanina (G), sendo a Adenina complementar à Timina, e a Citosina complementar à Guanina. Portanto, uma fita de DNA consiste em uma sequência de nucleotídeos, cuja ordem das bases nitrogenadas (A, T, C, G) determinaram as características biológicas. Embora as células possuam um aparato enzimático complexo que garante uma alta fidelidade nas sequências nucleotídicas durante a replicação do DNA, alguns “erros de replicação” podem ser perpetuados. Esses “erros”, além de fatores ambientais (como a radiação, por exemplo), podem levar a mutações. As mutações são a fonte do polimorfismo genético, que consiste no principal objetivo que permeia as pesquisas na área da Biologia Molecular. Conhecer com profundidade os mecanismos moleculares envolvidos na estrutura e replicação do DNA é importante para garantir o entendimento dos protocolos e o domínio das análises feitas no laboratório, por isso, recomenda-se a leitura de livros-base na área da Biologia Molecular e Genética. Segue abaixo uma lista das principais referências na área:

Bruce Alberts et al. (2017)

Anthony J. F. Griffiths et al. (2016)

Peter Snustad e Michael J. Simmons (2017)

2.2 Marcadores Moleculares

Parte I

Protocolos

3 Diluições

As diluições são um procedimento comum em laboratório que consiste em misturar uma solução concentrada com uma solução diluente para produzir uma solução de concentração conhecida. Elas são utilizadas em diversas aplicações em laboratório, como preparação de amostras para análises químicas ou biológicas, padronização de soluções e muito mais.

Para realizar uma diluição, é necessário conhecer a concentração da solução inicial e a quantidade de solução diluente que será adicionada. A concentração da solução final pode ser calculada usando a fórmula de diluição:

$$C1 * V1 = C2 * V2$$

Onde $C1$ é a concentração da solução inicial, $V1$ é o volume da solução inicial, $C2$ é a concentração da solução final e $V2$ é o volume da solução final.

3.1 Etanol

3.1.1 Etanol 70%

Em um tubo tipo falcon, colocar:

- 70 mL de etanol absoluto
- 30 mL de água destilada

3.1.2 Etanol 80%

Em um tubo tipo falcon, colocar:

- 80 mL de etanol absoluto
- 20 mL de água destilada

3.2 GelRed

Em um tubo de 1,5 mL envolto com fita e papel alumínio, colocar:

- 1 L de GelRed (estoque)
- 500 L de água Milli-Q

3.3 Ladder 100pb

Em um tubo de 1,5 mL, colocar:

- 10 L de Ladder (estoque)
- 90 L de água Milli-Q

3.4 Lambda

3.4.1 Lambda 5ng/ L (para 1000 L)

- 9,27 L de lambda
- 990,73 L de água Milli-Q

3.4.2 Lambda 10ng/ L (para 1000 L)

- 18,56 L de lambda
- 981,44 L de água Milli-Q

3.4.3 Lambda 20ng/ L (para 1000 L)

- 37,11 L de lambda
- 962,89 L de água Milli-Q

3.5 Primer para uso (10 M)

Em um tubo, colocar:

- 10 L do primer estoque
- 50 L de água Milli-Q

3.6 Proteinase K 20 mg/mL

- 20 mg de Proteinase K (estoque)
- 1 mL de água Milli-Q

Modo de preparo: lavar bem uma espátula, esterilizar com álcool 70% e pesar a Proteinase K em um microtubo novo (1,5 mL). Adicionar a água Milli-Q e ressuspender bem com a pipeta. Agitar a solução no vórtex para homogeneizar bem. Manter no freezer.

3.7 Tampão TBE 5x (estoque)

- 54 g de Tris-Base
- 27,5 g de Ácido Bórico
- 20 mL de EDTA 0,5M pH 8,0

Completar para 1000 mL com água destilada

Modo de preparo: em um becker, colocar um pouco da água destilada. Colocar uma pulga no becker e utilizar o agitador para homogeneizar a solução. Ir adicionando aos poucos os reagentes. Ao final, utilizando uma proveta, completar o volume de água que falta e armazenar em frascos de vidro (tampa laranja). Etiquetar e autoclavar.

3.8 Tampão TBE 1x (corrida)

$$C1 * V1 = C2 * V2$$

$$5(\text{estoque}) * V1 = 1 * 1000 \text{ mL}$$

$$V1 = 1000/5$$

$$V1 = 200 \text{ mL de TBE } 5x$$

Completar para 1000 mL de água destilada (= 800 mL)

3.9 PEG 20% NaCl 2,5M (para 75 mL)

- 15 g de Polietilenoglicol
- 10,95 g de NaCl

Completar para 75 mL com água Milli-Q

Modo de preparo: colocar em um becker um pouco da água e misturar o Polietilenoglicol e o NaCl com um bastão de vidro até a solução ficar homogênea. Depois de homogeneizada, completar o volume de água que falta e com o auxílio de uma seringa passar a solução pelo filtro (0,2 μ M – azul). Ir colocando em tubos Falcon para autoclavar. Após autoclavada, fazer alíquotas em microtubos de 1,5 mL. Manter no freezer da geladeira.

3.10 Diluição primer liofilizado

Quantidade em nmol x 1000 / 60 = Quantidade de água Milli-Q

Acrescentar essa quantidade de água no tubo do primer

Obs: 60 porque é a concentração utilizada para diluição dos primers no laboratório (= 60 μ M)

4 Extração

A extração de DNA é um processo fundamental em genética molecular que consiste em separar o material genético de uma célula ou tecido para que possa ser estudado e manipulado. Existem várias técnicas de extração de DNA, mas todas elas envolvem a lise da célula para liberar o DNA e a purificação do DNA para remover contaminantes. Para que seja possível conseguir DNA de boa qualidade, existem protocolos que foram testados, adaptados e otimizados para cada grupo estudado no laboratório.

A primeira etapa da extração de DNA é a lise da célula, que pode ser feita de várias maneiras, como maceração mecânica, lise com detergentes ou lise com enzimas (como a Proteinase K). Na maioria dos protocolos utilizados para extração de DNA no Nupgen os três métodos de lise celular são combinados, ou seja, é primeiramente realizada a maceração mecânica, adicionado tampão que contém detergentes que atuam nos lipídios das membranas promovendo o seu rompimento e posteriormente a enzima Proteinase K também é adicionada.

Após a lise das células, o DNA precisa ser separado dos restos celulares e das proteínas. Dessa maneira, o DNA é purificado por meio de precipitação com etanol ou cloreto de sódio, ou por meio de centrifugação em colunas. Uma vez que o DNA é purificado, ele pode ser utilizado em várias aplicações, como sequenciamento de DNA, clonagem, análise de expressão gênica e muito mais. No entanto, é importante lembrar que o DNA é um material frágil e pode ser danificado facilmente, portanto, é importante tomar cuidado durante todo o processo de extração e manipulação.

i Como fazemos

Atualmente no Nupgen são utilizados dois kits de extração diferentes: **Wizard® Genomic DNA Purification Kit (Promega®)** e **DNeasy® Blood and Tissue Kit (QIAGEN®)**. A escolha do kit irá depender do organismo a ser estudado, por exemplo, para peixes, gastrópodes e macrófitas recomenda-se a utilização do kit da Promega e para organismos menores, como parasitas, o kit em colunas da QIAGEN é o mais ideal.

4.1 Protocolo de Extração de DNA - Promega Peixes

1. Com pinça e bisturi, fracionar aproximadamente 20 mg de músculo.



Figura 4.1: Wizard® Genomic DNA Purification Kit (Promega®)



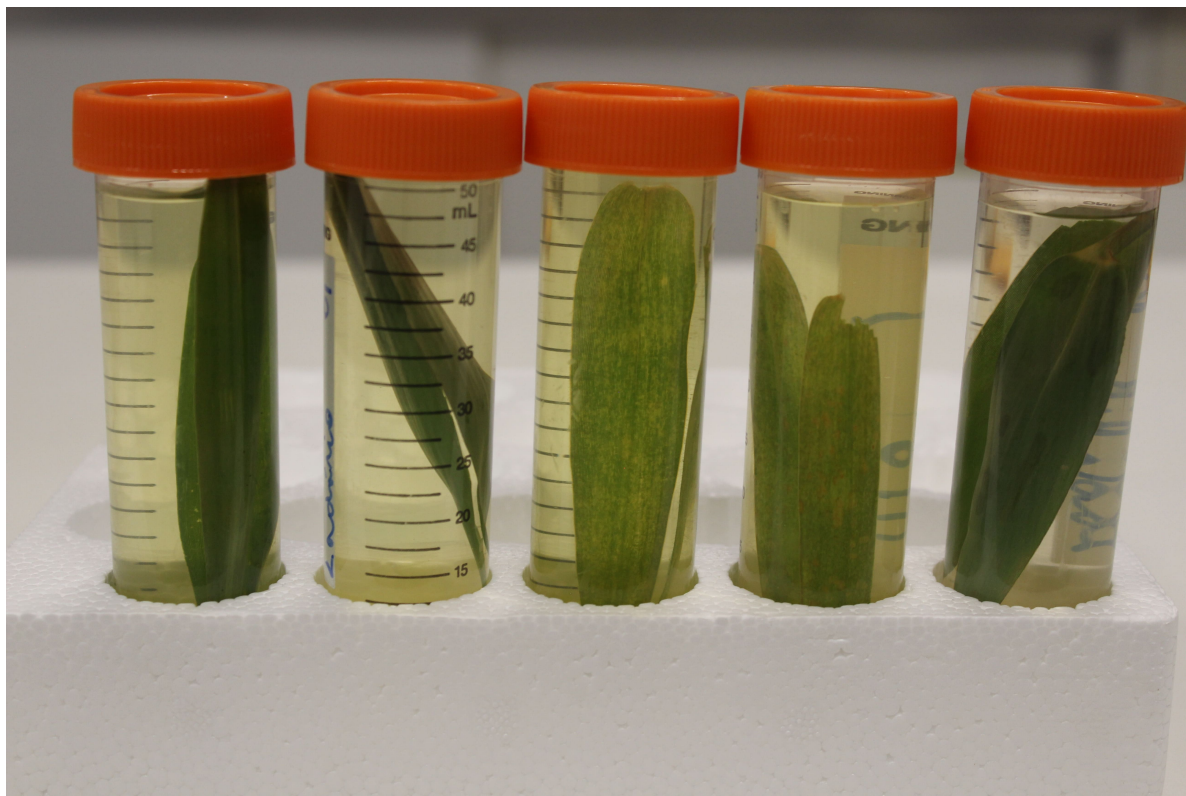
Figura 4.2: DNeasy® Blood and Tissue Kit (QIAGEN®)

2. Adicionar 400 μ L de Nuclei Lysis Solution.
3. Incubar por 30 minutos a 55° C.
4. Adicionar 10 μ L de Proteinase K 20 mg/ml.
5. Incubar a 55° C overnight ou de 2-3 horas até ficar homogêneo.
6. Adicionar 2 μ L de RNase Kit Solution e mexer a amostra por inversão.
7. Incubar 30 minutos a 37° C.
8. Deixar em temperatura ambiente por 5 minutos.
9. Adicionar 200 μ L de Protein Precipitation Solution e agitar por 20 segundos.
10. Deixar no gelo por 5 minutos.
11. Centrifugar por 10 minutos a 12.000 rpm.
12. Remover o sobrenadante transferindo para um tubo novo – descartar o pellet.
13. Adicionar 600 μ L de isopropanol e agitar por inversão.
14. Centrifugar por 2 minutos a 12.000 rpm – descartar o sobrenadante.
15. Adicionar 600 μ L de etanol 70% a temperatura ambiente.
16. Centrifugar por 2 minutos a 12.000 rpm – descartar sobrenadante.
17. Deixar secar – pode ser na estufa.
18. Ressuspender em 30 μ L de DNA Rehydration Solution. Deixar por 1 hora no banho a 65° C ou overnight a 4° C.

4.2 Protocolo de Extração de DNA - Promega Caramujos

1. Separar o tecido (0,5-1 cm) em um tubo de 1,5 ml.
2. Adicionar 500 μ l de Nuclei Lysis Solution.
3. Macerar o tecido com um pistilo.
4. Adicionar 17,5 μ l de Proteinase K 20 mg/ml.
5. Misturar com vórtex por 10 segundos.
6. Incubar a 55° C overnight OU incubar a 55° C por 3 horas, misturando com vórtex a cada hora.
7. Adicionar 3 μ l de RNase Solution à solução e misturar por inversão do tubo 2-5 vezes.
8. Incubar a 37° C por 30 minutos. Deixar as amostras esfriarem a temperatura ambiente por 5 minutos antes da próxima etapa.
9. Adicionar 200 μ l de Protein Precipitation Solution e misturar com o vórtex vigorosamente por 20 segundos.
10. Colocar as amostras no gelo por 5 minutos.
11. Centrifugar a 12.000 rpm por 4 minutos. A proteína precipitada formará um pellet no fundo.
12. Remover cuidadosamente o sobrenadante contendo o DNA (não retirar o pellet) e transferir para um novo tubo de 1,5 ml que já contenha 600 μ l de isopropanol (temperatura ambiente).
13. Misturar cuidadosamente a solução por inversão até os “fios” de DNA formarem uma massa visível.
14. Centrifugar a 12.000 rpm por 1 minuto. O DNA ficará visível como um pequeno pellet branco. Descartar cuidadosamente o sobrenadante.
15. Adicionar 600 μ l de etanol 70% (temperatura ambiente) e misturar cuidadosamente a solução por inversão para lavar o DNA.
16. Centrifugar a 12.000 rpm por 1 minuto.
17. Retirar o etanol cuidadosamente com uma pipeta sem retirar o pellet.
18. Inverter os tubos em papel e deixar secar por 10-15 minutos.
19. Adicionar 30 μ l de DNA Rehydration Solution e reidratar o DNA incubando a 65° C por 1 hora. Periodicamente misturar a solução com cuidado OU incubar a 4° C overnight.
20. Guardar o DNA extraído a 2-8° C.

4.3 Protocolo de Extração de DNA - Promega Macrófitas



1. Tubos de 1,5 mL.
2. Amostras + nitrogênio líquido (macerar).
3. Adicionar 600 μ l de Nuclei Lysis Solution (logo após macerar).
4. Banho de 15 minutos a 65° C.
5. Adicionar 3 μ l de RNase.
6. Virar o tubo 5x.
7. Colocar no banho por 15 minutos a 37° C (pode segurar na mão).
8. Deixar por 5 minutos em temperatura ambiente.
9. Adicionar 200 μ l de Protein Precipitation Solution.
10. Vórtex.
11. Centrifugar por 3 minutos a 16.000 ou 13.000 rpm.
12. Separar sobrenadante e descartar o resto.

13. Adicionar 600 μ l de isopropanol ao sobrenadante.
14. Virar 5x.
15. Centrifugar por 1 minuto a 16.000 ou 13.000 rpm.
16. Descartar sobrenadante.
17. Adicionar 600 μ l de etanol 70%.
18. Virar 5x.
19. Centrifugar por 1 minuto a 16.000 ou 13.000 rpm.
20. Descartar sobrenadante.
21. Deixar o tubo secar por 15 minutos na estufa.
22. Adicionar 50 μ l de DNA Rehydration Solution.

4.4 Protocolo de Extração de DNA com Colunas - QIAGEN Parasitos

1. Deixar o mínimo possível de líquido na amostra antes de começar.
 2. Adicionar 180 μ l de Buffer ATL.
 3. Adicionar 20 μ l de proteinase K. Misturar com vórtex e incubar a 56° C até completa lise celular (30min - 2h). Misturar com vórtex ocasionalmente durante a incubação.
 4. Misturar com vórtex por 15 segundos.
 5. Adicionar 200 μ l de Buffer AL. Misturar bem com vórtex.
 6. Adicionar 200 μ l de etanol (96-100%). Misturar bem com vórtex.
 7. Pipetar a mistura em uma coluna de centrifugação DNeasy Mini colocada em um tubo de coleta. Centrifugar a 8.000 rpm (6.000 x g) por 1 minuto. Descartar o líquido filtrado e o tubo de coleta.
 8. Colocar a coluna em um novo tubo de coleta do kit. Adicionar 500 μ l de Buffer AW1.
- Centrifugar por 1 minuto a 8.000 rpm. Descartar o líquido filtrado e o tubo de coleta.
9. Colocar a coluna em um novo tubo de coleta do kit. Adicionar 500 μ l de Buffer AW2. Centrifugar por 3 minutos a 14.000 rpm (20.000 x g). Descartar o líquido filtrado e o tubo de coleta.
 10. Transferir a coluna para um eppendorf de 1,5 mL.

11. Adicionar 50 μ L de Buffer AE no centro da membrana da coluna para a eluição do DNA. Incubar por 1 minuto a temperatura ambiente (15-25° C). Centrifugar por 1 minuto a 8.000 rpm.
12. Descartar a coluna e guardar o DNA extraído.

5 Eletroforese

A eletroforese consiste em um método de separação de macromoléculas, como o DNA, por meio da aplicação de corrente elétrica. Como as moléculas de DNA apresentam carga negativa, quando são submetidas a um campo elétrico, elas migram para o polo positivo, sendo este o princípio da eletroforese, ou seja, moléculas com cargas distintas irão migrar pelo gel de um polo a outro. Além disso, quanto menor for o tamanho da molécula de DNA (menor a quantidade de nucleotídeos), mais rápido será o seu deslocamento e consequentemente maior será a distância percorrida.

Existem duas maneiras para realizar uma eletroforese em gel, podendo ser vertical ou horizontal. A eletroforese vertical é feita com géis de poliacrilamida e cada polo da cuba está em um recipiente com tampão de corrida (o polo negativo encontra-se acima e o polo positivo abaixo). Já a eletroforese horizontal é realizada geralmente com géis de agarose que pode ser de diferentes concentrações, com uma cuba sendo um recipiente único, ou seja, o tampão preenche todo o recipiente, polos negativo e positivo, inclusive cobrindo o gel. Em ambos os casos a corrente elétrica irá atravessar o gel e induzir a migração das moléculas de DNA do polo negativo para o polo positivo.

A escolha por uma eletroforese vertical ou horizontal vai depender do objetivo que deseja atingir. Géis de poliacrilamida são ideais para fragmentos que necessitam de uma maior resolução, visto que são mais sensíveis quando comparados aos géis de agarose e permitem a separação de moléculas com variações de até 1 pb em seu tamanho. Géis de agarose embora não apresentem uma resolução tão boa para separação de fragmentos com pouca diferença de tamanho, são ideais para separar moléculas e fragmentos com diferenças de tamanho maiores de acordo com a concentração do gel utilizado.

Para a eletroforese horizontal em gel de agarose, o gel é preparado de acordo com a concentração desejada (no laboratório geralmente é utilizada a concentração de **1%**), misturando a quantidade correta de agarose, tampão TBE 5x (**T**ris, **B**órico e **E**DTA) e água destilada. Para que a agarose seja completamente dissolvida é preciso aquecer em microondas até a solução ficar completamente translúcida. Posteriormente, o gel é colocado em um suporte apropriado com os pentes com a finalidade de formar as cavidades (poços) onde as amostras serão aplicadas. Após preparação e aplicação das amostras nas cavidades do gel, este será colocado na cuba juntamente com o tampão de corrida (TBE 1x) e a cuba será ligada por meio de fios em uma fonte de energia, escolhendo a voltagem e o tempo de corrida.

Depois que a corrida é finalizada, os cabos são desligados e o gel pode ser visualizado em um equipamento chamado transiluminador. A visualização é possível devido a corantes que são adicionados no momento da preparação das amostras (como GelRed e Safer), além de determinar qual o tipo de equipamento que deve ser utilizado (transiluminador UV ou LED). A análise do gel irá depender do objetivo (quantificação, verificar a amplificação ou purificação), apesar dos procedimentos empregados serem os mesmos.

i Como fazemos

É importante lembrar que o tempo de migração das amostras e a voltagem escolhida são parâmetros importantes na eletroforese. Quanto maior o tempo de migração, maior será a separação dos fragmentos. Quanto maior a voltagem, mais rápida será a migração das moléculas no gel. No Nupgen geralmente utiliza-se a voltagem em torno de **90V por aproximadamente uma hora**.

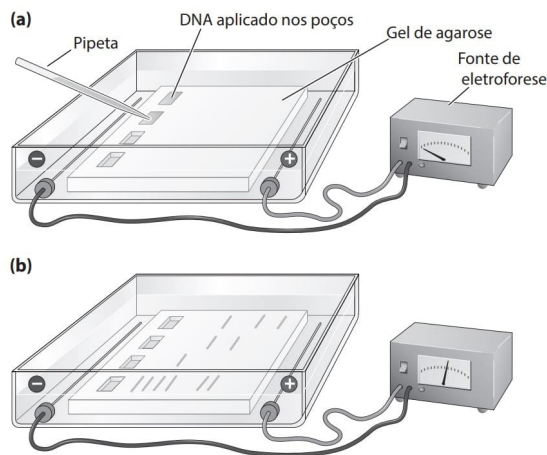


Figura 5.1: Equipamento utilizado na eletroforese



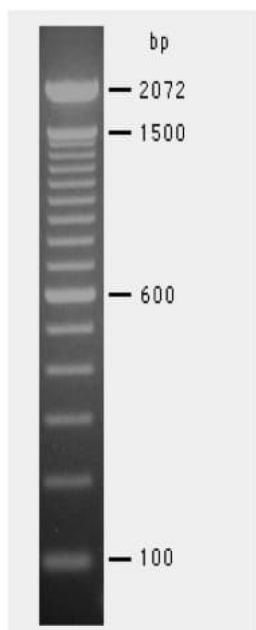
Figura 5.2: Equipamento do NuPage

5.1 Ladder

O Ladder é um marcador de peso molecular utilizado para determinar o tamanho dos fragmentos após a amplificação do DNA. É composto por produtos de PCR e plasmídeos digeridos com enzimas de restrição, sendo ideal para uso como padrão de peso molecular conhecido para eletroforese em gel de agarose. O Ladder apresenta bandas de maior intensidade que irão servir como pontos de referência (geralmente duas bandas para o Ladder **100 pb** e três bandas para o Ladder **1 Kb**).

i Como fazemos

No Nupgen há duas opções de Ladder para serem utilizados como padrão, o Ladder 100 pb e o Ladder 1 Kb. O marcador DNA Ladder 100 pb é o mais utilizado nos protocolos do laboratório e consiste em múltiplas repetições de fragmentos de DNA (12 fragmentos), separados a cada 100 pb até 1 Kb, com dois fragmentos adicionais de 1500 e 2072 pb. Já o marcador de peso molecular DNA Ladder 1 Kb contém 14 fragmentos facilmente identificáveis, comumente utilizado para a identificação de massa molecular de fragmentos de DNA entre 250 bp a 10 Kb.



100 bp DNA Ladder
0.5 µg/lane
2% agarose gel stained

5.2 Lambda

Lambda é um bacteriófago de *Escherichia coli* e seu DNA contém 48.502 pares de nucleotídeos. É utilizado como DNA padrão com concentrações conhecidas (por exemplo 5, 10 e 20 ng/ L) para quantificar o DNA extraído de amostras alvo de estudos por meio da comparação visual pela fluorescência e espessura das bandas. O DNA fago Lambda é adquirido de empresas com uma concentração estoque específica e para ser utilizado na rotina do laboratório é preciso realizar a diluição com base nas concentrações desejadas.

Cuidado

O DNA fago **Lambda** é utilizado para a quantificação em diferentes concentrações, enquanto o **Ladder** é utilizado na eletroforese para verificar o tamanho do fragmento de interesse amplificado após a PCR e nos géis de purificação.

5.3 Gel agarose

texto texto

5.3.1 Gel 1%

Cuba - 60 mL

0,6g de agarose
12 mL de TBE
48 mL de água destilada

Cuba - 120 mL

1,2g de agarose
24 mL de TBE
96 mL de água destilada

5.3.2 Gel 1,4%

Cuba - 120 mL

1,68g de agarose
24 mL de TBE
96 mL de água destilada

5.3.3 Gel 3%

Cuba - 120 mL

3,6g de agarose

Cuba - 120 mL
24 mL de TBE
96 mL de água destilada

5.4 Protocolo para quantificação

texto

5.5 Protocolo para amplificação

texto

5.6 Protocolo para purificação

texto

5.7 Preparação para eletroforese

5.7.1 Ladder 100 pb

2 L de loading + 1 L de GelRed + **2 L de Ladder 100 pb**

5.7.2 Por amostra

2 L de loading + 1 L de GelRed + **3 L da PCR**

! Importante

- Fio preto em cima (polo negativo) e o fio vermelho embaixo (polo positivo).
- Gel agarose 1% (pode ser reutilizado).
- Correr a 90 volts, por aproximadamente 1 hora.
- Colocar o Ladder 100 pb no primeiro poço.

6 Quantificação

Após a etapa de extração, determinar a concentração de DNA, assim como a sua qualidade e pureza, são fatores importantes para o desenvolvimento de estudos em biologia molecular. Dessa maneira, as amostras obtidas podem ser avaliadas quanto à sua concentração e qualidade. Essa avaliação pode acontecer de algumas maneiras, como realizando uma **eletroforese em gel de agarose**, através da **medição da absorbância com o uso de espectrofotômetros** ou utilizando **fluorímetro** por meio das alterações nas características de fluorescência na presença de DNA.

A quantificação irá fornecer informações sobre a concentração de DNA extraído, assim como a integridade das amostras que serão trabalhadas. Na quantificação utilizando gel de agarose, um gel com concentração de 1% é preparado e as amostras de DNA são aplicadas juntamente com amostras de concentrações conhecidas. Para essa finalidade, o DNA do bacteriófago **Lambda** é comumente utilizado como padrão. O gel é normalmente submetido a eletroforese e é realizada uma estimativa visual através da comparação com o DNA padrão para determinar a concentração de DNA em cada amostra, analisando a fluorescência e espessura das bandas, obtendo valores aproximados. Além disso, as amostras consideradas de boa qualidade irão formar bandas íntegras, enquanto amostras degradadas ou com presença de moléculas de RNA apresentarão rastros ao longo do gel.

Por outro lado, a quantificação em espectrofotômetros, como por exemplo o **NanoDrop**, fornece dados um pouco mais exatos sobre a concentração de DNA presente nas amostras. Entretanto, ainda são valores baseados em um calibrante, tomado como padrão, além de ser necessário adquirir aparelhos específicos para realizar a quantificação do DNA, tendo um custo elevado quando comparado à quantificação em gel. Essa quantificação é baseada na medição da quantidade de luz absorvida pelo DNA em solução no comprimento de onda de 260 nm. Quanto maior for a absorção de luz nesse comprimento de onda, maior a concentração de DNA extraído na amostra.

Já o fluorímetro é um dos métodos mais sensíveis e geralmente utilizado para amostras com pouca concentração de DNA ou que tenham contaminantes que absorvem comprimentos de onda de 260 nm. Essa análise acontece pela luz emitida de moléculas fluorogênicas, sendo um método alternativo para avaliar a concentração de DNA e RNA marcando a amostra com um marcador fluorescente (corante fluorescente). Porém, também é um método com um custo elevado e demanda um tempo maior para preparação das amostras.

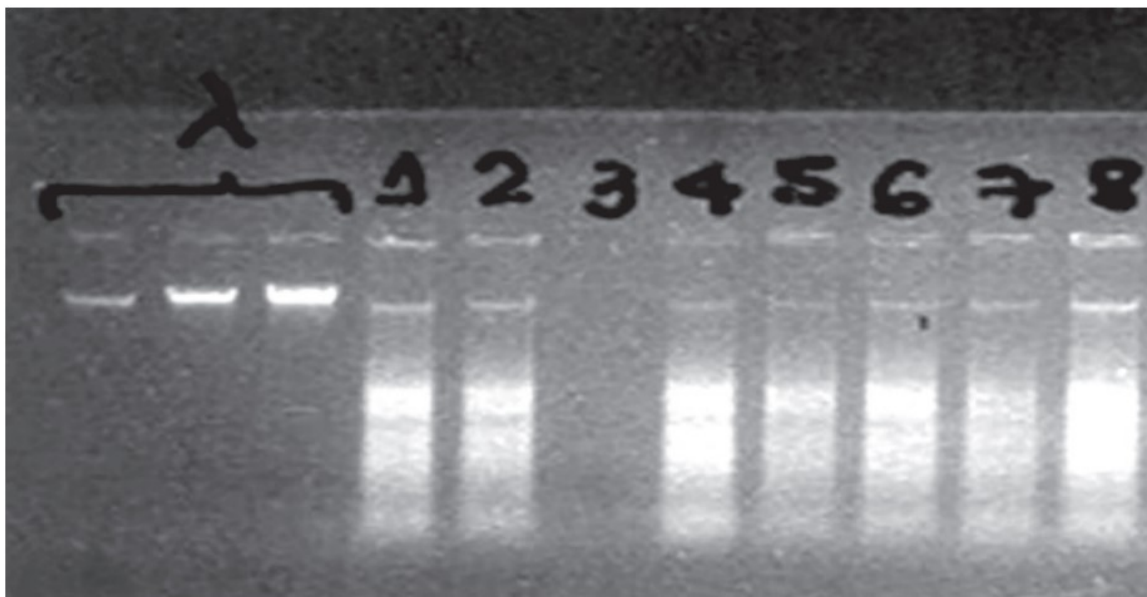


Figura 6.1: Exemplo de quantificação utilizando gel de agarose e Lambda

i Como fazemos

O método de quantificação geralmente utilizado pelos pesquisadores do Nupgen é a **eletroforese em gel de agarose 1%**, visto que tem dado certo para a maioria dos grupos trabalhados no laboratório, exceto parasitas que não passam por essa etapa, indo diretamente para a PCR devido a baixa concentração de DNA obtida nas amostras. Na quantificação são utilizadas três concentrações diferentes de DNA Lambda, sendo elas: 5, 10 e 20 ng/ L.

A quantificação irá mostrar se as amostras de DNA extraídas precisam ser **diluídas** para a realização da PCR, pois é preciso ter uma concentração ideal. Dessa forma, quando as amostras forem quantificadas com valores superiores a 5 ng/ L, elas precisam ser diluídas. A fórmula utilizada para isso é $C1 * V1 = C2 * V2$, onde serão encontradas as quantidades de DNA e água Milli-Q necessárias para que o DNA fique na concentração final de 5 ng/ L.

6.1 Preparação por amostra

- Gel agarose 1%.
- Correr por aproximadamente 40 minutos, a 100 volts.

- DNA lambda (pipetar 5 μ L + 1 μ L de GelRed em cada um):
 - 5 ng/ μ L – 25 ng
 - 10 ng/ μ L – 50 ng
 - 20 ng/ μ L – 100 ng

6.1.1 GelRed

2 μ L de DNA + 1 μ L de loading + 1 μ L de GelRed + 1 μ L de água Milli-Q*

*água Milli-Q é opcional

6.1.2 Safer

3 μ L de DNA + 1 μ L de Safer + 2 μ L de água Milli-Q

6.2 Diluição do DNA para 5ng/ μ L

$$C1 * V1 = C2 * V2$$

C1 – concentração da amostra (inicial)

V1 – quantidade de DNA necessária (descobrir)

C2 – concentração para diluir (vai diluir para 5 ng)

V2 – volume final total (50 μ L)

OBS: Quantidade de água Milli-Q é: 50 μ L - quantidade de DNA

7 Amplificação

Amplificação é o nome dado ao processo de replicação de uma região específica do DNA alvo de um estudo de modo que seja possível produzir um grande número de cópias desse mesmo segmento.

A técnica utilizada para a amplificação é a reação em cadeia da polimerase (**PCR**), para a execução dessa técnica é preciso primeiramente fazer a seleção do **marcador molecular** que será utilizado, ou seja, a região do DNA de interesse a ser replicada, isso irá variar conforme as necessidades do estudo em questão. Com base nessa escolha, serão utilizados reagentes específicos com utilidades diferentes para que a reação aconteça.

Aprendendo

Os reagentes utilizados são:

- O **DNA extraído**;
- Um par de oligonucleotídeos iniciadores, os chamados **primers**, sendo um foward () e um reverse (). Eles irão se ligar na região que deve ser amplificada e permitir a ação de uma enzima polimerizadora;
- A enzima DNA polimerase. Normalmente a enzima utilizada nas reações de PCR é proveniente da bactéria *Thermus aquaticus* (por isso **Taq**), uma espécie termófila adaptada a viver em águas de alta temperatura, dessa forma, a enzima advinda dela é termoestável e consequentemente resistente às altas temperaturas da PCR sem sofrer desnaturação;
- Bases nitrogenadas livres, os desoxirribonucleotídeos (**dNTPs**);
- Tampão de reação (MgCl_2^-), importante para agir como cofator da DNA polimerase;
- Solução **tampão**, com o objetivo de manter o pH da reação estável durante o ciclo de amplificação;
- **Água Milli-Q**, para completar o volume necessário.

A mistura desses componentes é levada ao termociclador, um aparelho com a capacidade de aquecer e resfriar as amostras, isso é importante porque são temperaturas específicas que

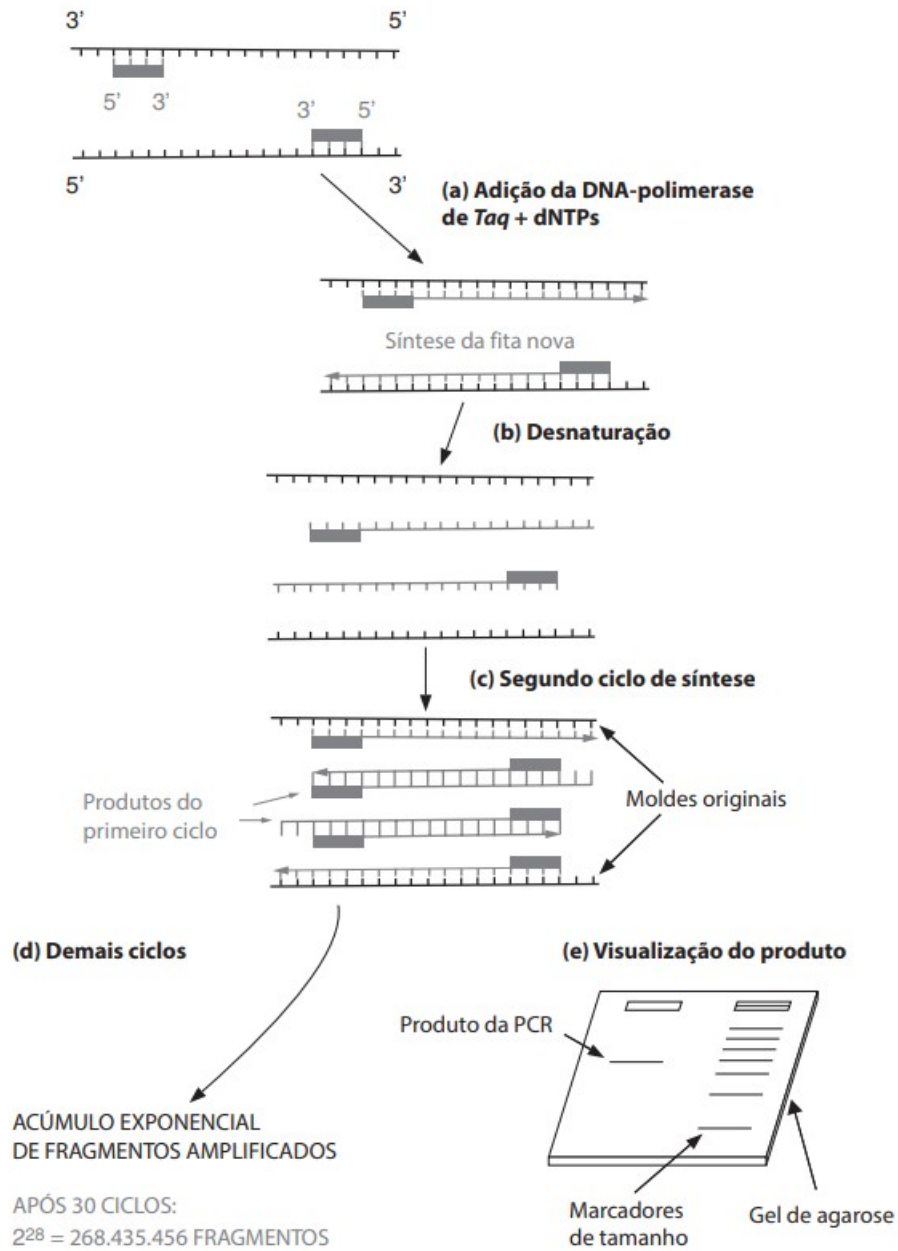


Figura 7.1: Reagentes utilizados na PCR

permitem a abertura das fitas, anelamento dos primers e a ação da enzima DNA-polimerase. Comumente, dividimos a PCR em três etapas, são elas:

1. **Desnaturação:** Nesse momento as fitas de DNA, configuradas em dupla-hélice, terão suas ligações de hidrogênio quebradas, sendo assim separadas. Essa etapa ocorre graças à elevação da temperatura do termociclador a temperaturas acima de 90°C .
2. **Anelamento:** Com as fitas já desnaturadas, a temperatura do aparelho é reduzida comumente para valores entre 45°C e 65°C e o par de primers, chamados de iniciadores, se ligam com as suas sequências complementares na fita molde.
3. **Polimerização:** O termociclador regula a sua temperatura para 72°C , atingindo a temperatura ótima para a ação da enzima DNA polimerase, que começa a síntese de novas fitas utilizando as dNTPs livres como matéria prima. A síntese só é possível graças à disponibilidade de extremidades $3'\text{OH}$ livres pertencentes aos primers.

As três etapas citadas anteriormente são repetidas em cerca de 20 a 30 ciclos, possibilitando que até bilhões de cópias da região selecionada sejam produzidas.



8 Purificação

A purificação de DNA após a amplificação é um passo importante na produção de DNA suficientemente puro para aplicações posteriores, como sequenciamento ou clonagem. A amplificação do DNA, como a reação em cadeia da polimerase (PCR), geralmente produz quantidades muito grandes de DNA, mas também gera muitos contaminantes, como primers, cofator de PCR e outros produtos de reação. Portanto, é importante purificar o DNA amplificado antes de usá-lo em outras aplicações.

Existem várias técnicas de purificação de DNA, como centrifugação em colunas de diálise, precipitação com etanol ou acetato de sódio, extração com fenol-clorofórmio ou precipitação com polietilenoglicol (PEG). Cada método tem suas próprias vantagens e desvantagens e o método ideal dependerá da aplicação específica.

Após confirmar a amplificação do fragmento de interesse, é realizada a purificação das amostras para remover o excesso de primer, nucleotídeos e fragmentos de PCR incompletos de baixo peso molecular.

i Como fazemos

No Nupgen, essa purificação é realizada por meio de uma etapa de precipitação com polietilenoglicol (PEG), protocolo descrito a princípio por Rosenthal, Coutelle, e Craxton (1993) . A mistura de PEG precipita seletivamente DNA de 100 bp a 150 pb, deixando primers residuais, nucleotídeos e produtos de PCR incompletos no sobrenadante. Dessa maneira, o sobrenadante é removido, ficando precipitado apenas o DNA amplificado.

Por fim, é feita uma eletroforese em gel de agarose para verificar se a purificação foi bem-sucedida e se as amostras podem ser utilizadas para o sequenciamento.

8.1 Purificação com Polietilenoglicol (PEG)

1. Adicionar 50 μ L de PEG 20% NaCl 25m. em um tubo de 1,5 mL.
2. Transferir a PCR para o tubo e misturar.
3. Incubar o conteúdo do tubo a 37°C por 15 minutos.

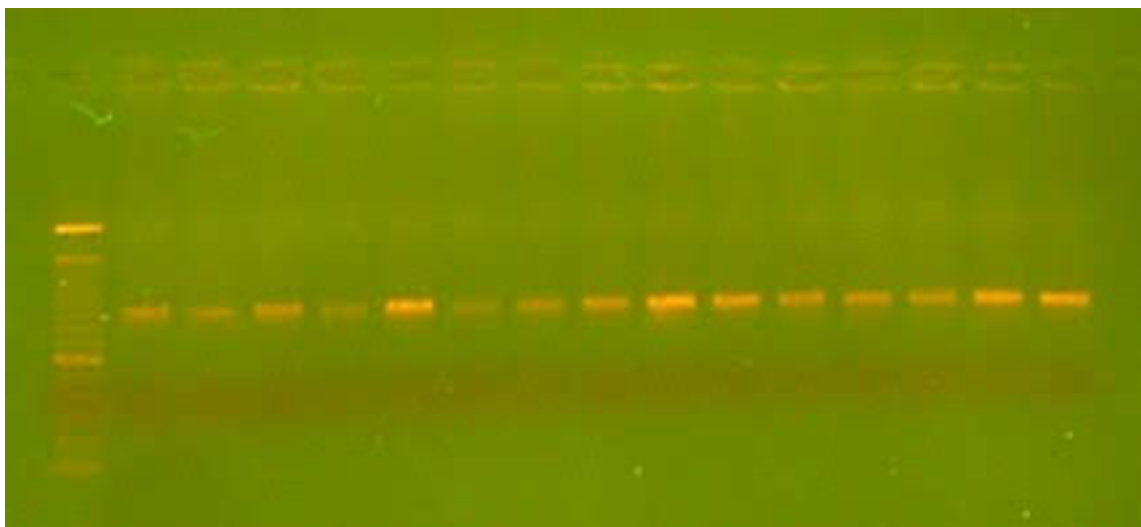


Figura 8.1: Foto de um gel de purificação

4. Centrifugar a reação + PEG por 15 minutos a 12000 rpm.
5. Retirar o sobrenadante sem tocar na parede do tubo e descartar – usar pipeta.
6. Adicionar 62,5 μ L de etanol 80% (álcool Merck) gelado no tubo – colocar na parede.
7. Centrifugar por 2 minutos.
8. Virar o tubo, descartando o sobrenadante.
9. Deixar secar até não ter mais gotas – pode ser na estufa.
10. Ressuspender com 10 μ L de água Milli-Q lavando a parede do tubo de 5 a 6 vezes.

9 Sequenciamento

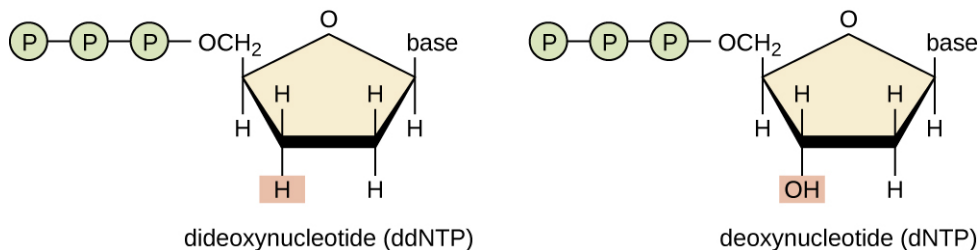
Depois de realizada a purificação, o fragmento purificado pode ser sequenciado. O sequenciamento de DNA consiste no processo de determinação da sequência de nucleotídeos (**A, T, G e C**) em um fragmento de DNA. A informação contida nessas sequências de nucleotídeos possibilita diversos tipos de aplicações, como, por exemplo, identificar o organismo ao qual pertence uma amostra, diagnosticar uma doença genética, aplicar em estudos de filogenia e diversidade genética, entre outros.

💡 Aprendendo

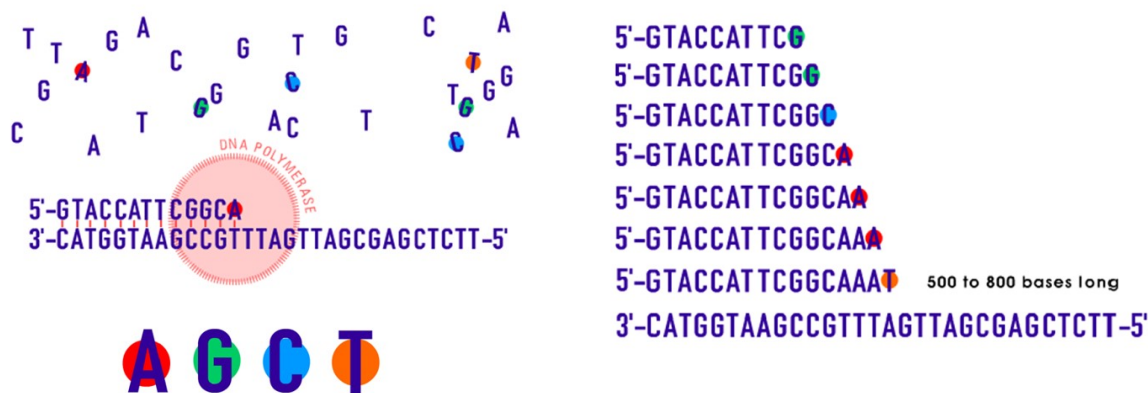
Existem vários métodos para o sequenciamento do DNA, tais como:

- O método enzimático (**1ª Geração**) de Frederik Sanger (1980);
- O método de pirosequenciamento (**2ª Geração**) de Mostafa Ronaghi (1996);
- O de **3ª Geração**, Single Molecule Real Time (SMRT) - DNA Sequencing;
- Métodos mais modernos conhecidos como **Sequenciamento de Nova Geração (NGS)**.

A técnica desenvolvida por Sanger e colaboradores possibilitou a determinação da sequência nucleotídica e se tornou mais acessível devido aos avanços recentes na fabricação de equipamentos automatizados. Além disso, é um dos métodos mais utilizados (inclusive nos trabalhos do Nupgen) e consiste na adição de um reagente único: **os dideoxirribonucleotídeos trifosfatados (ddNTPs)**. É realizada uma reação de sequenciamento, contendo os mesmos reagentes utilizados na [Reação em Cadeia da Polimerase](#), com a presença dos ddNTPS, ou seja, terminadores de cadeia para os quatro nucleotídeos (**ddATP, ddTTP, ddCTP, ddGTP**), cada um deles marcado com um corante fluorescente de cores diferentes.



A variação no tamanho dos fragmentos acontece por essa adição de nucleotídeos modificados (ddNTPs) juntamente com nucleotídeos normais (dNTPs). Os ddNTPs não possuem o radical hidroxila (OH) no carbono 3' da ribose, então ao serem incorporados na fita que está sendo sintetizada, haverá a interrupção dessa síntese. É uma adição aleatória e produzirá fragmentos terminados em cada uma das posições dos nucleotídeos da sequência de DNA que se deseja analisar. Dessa maneira, o método de Sanger consiste na produção de fragmentos de DNA com tamanhos que diferem em um nucleotídeo, utilizando como molde uma fita simples de DNA.



Os equipamentos automatizados utilizados hoje em dia para o sequenciamento possuem capilares extremamente finos que por meio da migração por eletroforese capilar, irão alinhar os fragmentos com base em seu tamanho (do menor para o maior), detectando por um feixe de laser a marcação de cada um dos quatro ddNTPs com a molécula fluorescente. Isso permite identificar qual foi o nucleotídeo adicionado em cada posição terminal dos fragmentos, construindo a sequência da molécula de DNA alvo dos estudos. Essa técnica é geralmente utilizada para o sequenciamento de moléculas com aproximadamente 100 e 1000 pb. Como resultado do sequenciamento, são obtidos os **eletroferogramas**, que apresentam picos representando cada um dos nucleotídeos e consequentemente a sequência do fragmento de interesse.

i Como fazemos

Atualmente, o Nupgen envia as amostras para o sequenciamento para a empresa ACTGene Análises Moleculares LTDA que utiliza o sequenciador automatizado Applied Biosystems® AB 3500 Genetic Analyzer (sequenciamento de Sanger). Essa empresa é do Rio Grande do Sul, então as amostras são preparadas e devidamente acondicionadas para serem enviadas pelos correios.

Site da ACTGene: <https://actgene.com.br/>

Informações sobre envio/preparo das amostras: https://actgene.com.br/wp-content/uploads/2022/06/ACTGene_Guia-preparo-de-amostras.pdf

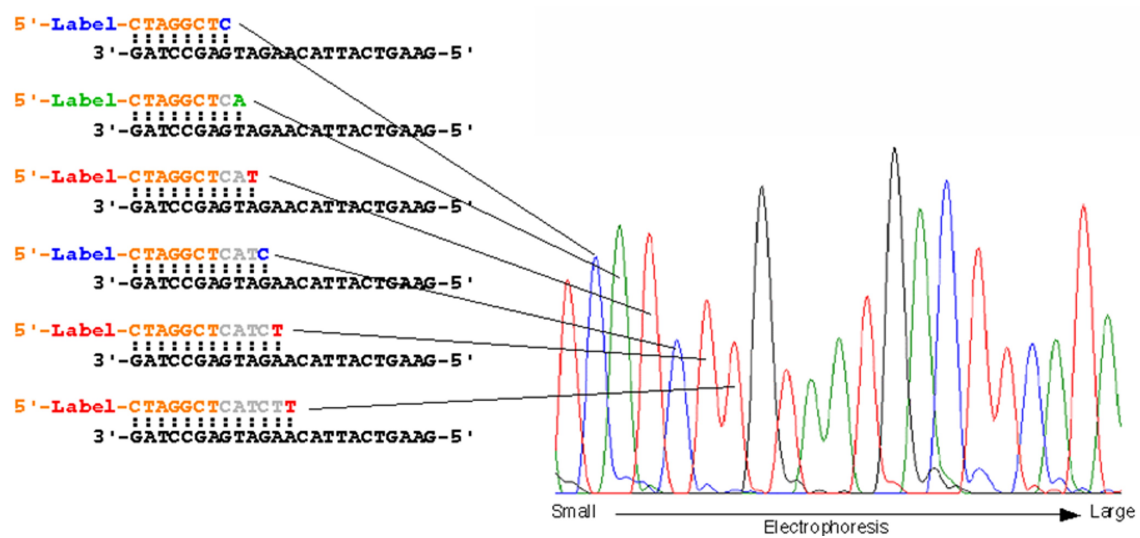


Figura 9.1: Eletroferograma

Parte II

Análises

10 Bancos de Dados Genéticos

10.1 GenBank

Link: <https://www.ncbi.nlm.nih.gov/genbank/>

O GenBank é um banco de dados de sequências de DNA mantido pelo National Center for Biotechnology Information (NCBI), um instituto do National Institutes of Health (NIH) dos Estados Unidos. Ele armazena sequências de DNA de diversas fontes, incluindo genomas, sequências de expressão gênica, sequências de proteínas e etc.

O GenBank é um recurso valioso para os pesquisadores de genética molecular porque permite o acesso a uma ampla variedade de sequências de DNA, tornando mais fácil a realização de análises e pesquisas. Além disso, o GenBank permite a submissão de novas sequências de DNA pelos usuários, o que permite o compartilhamento de dados entre os pesquisadores.

Todo trabalho ao ser publicado deve submeter as sequências utilizadas para o GenBank, dessa forma outros pesquisadores podem utilizar em suas próprias análises.

10.2 BOLD Systems

Link: <http://boldsystems.org/>

O BOLD Systems (Barcode of Life Data Systems) é um projeto de ciência cidadã que visa coletar e armazenar sequências de DNA barcodes (COI) de espécies de todo o mundo. O DNA barcode é um trecho curto de DNA que pode ser usado para identificar espécies de maneira rápida e precisa. Ele é amplamente utilizado em biologia evolutiva e conservação de espécies e é uma ferramenta valiosa para a identificação de espécies em ambientes complexos, como ecossistemas tropicais.

O BOLD Systems é mantido pelo Instituto de Biodiversidade do Canadá e é uma iniciativa internacional que conta com a participação de cientistas, conservacionistas e cidadãos do mundo todo. Ele possui uma base de dados online que armazena as sequências de DNA barcodes coletadas e permite o acesso a essas sequências pelos usuários.

10.3 BLAST

Link: <https://blast.ncbi.nlm.nih.gov/Blast.cgi>

O BLAST (Basic Local Alignment Search Tool) é uma ferramenta amplamente utilizada pelos pesquisadores de genética molecular para realizar buscas de sequências de DNA em bancos de dados de sequências. Ele é mantido pelo National Center for Biotechnology Information (NCBI) e está disponível gratuitamente na internet.

O BLAST funciona comparando uma sequência de entrada com as sequências armazenadas em um banco de dados de referência e encontrando as sequências mais similares. Ele utiliza uma técnica de alinhamento de sequências chamada alinhamento local, que permite o alinhamento apenas de trechos similares das sequências, ignorando trechos com pouca similaridade.

O BLAST é amplamente utilizado em diversas análises de genética molecular, como a identificação de genes e/ou espécies.

11 Softwares

Em nosso laboratório dependemos bastante do uso de diversos softwares para realizar diferentes análises. Enquanto o início dos trabalhos é feito com aparelhos em nosso laboratório, após receber os resultados do sequenciamento, dependemos de análises realizadas no próprio computador.

Para isso, precisamos entender como as análises funcionam e também qual a melhor forma que realizá-las, evitando dores de cabeça por resultados que não refletem a realidade.

A seguir vamos apresentar os principais softwares utilizados, bem como algumas dicas que podem facilitar o trabalho. Lembre-se, é muito mais fácil aprender com outros alunos do que apenas com os manuais de cada programa. Peça ajuda, principalmente com as análises, que podem gerar resultados errados e atrapalhar toda sua pesquisa.

11.1 BioEdit

Útil para: Edição

BioEdit é um editor de sequências de DNA e proteínas muito utilizado na genética molecular. Ele é especialmente útil na visualização de eletroferogramas, que são gráficos gerados por sequenciadores de DNA que mostram a sequência de base de uma amostra de DNA.

O BioEdit permite a visualização de eletroferogramas de maneira fácil e rápida, além de fornecer uma ampla gama de ferramentas para análise de sequências, como alinhamento de sequências, análise de substituição de nucleotídeos e proteínas, e muito mais. Ele também possui uma interface intuitiva e é compatível com Windows, MacOS e Linux.

Além da visualização de eletroferogramas, o BioEdit é amplamente utilizado em diversas outras tarefas de genética molecular, como edição de sequências, anotação de genes e análise de expressão gênica.

Usos

Prós

Contras

- A última versão disponível é de 2005.

11.2 MEGA

Útil para: Edição / Alinhamento / Modelo / Árvore

O MEGA (Molecular Evolutionary Genetics Analysis) é um software de análise de genética molecular utilizado para estudar a evolução e a relação filogenética entre sequências de DNA, RNA ou proteínas. Ele é amplamente utilizado por pesquisadores em diversas áreas da genética molecular, incluindo biologia evolutiva, genômica e epidemiologia de doenças.

O MEGA possui uma ampla gama de recursos, incluindo alinhamento de sequências, inferência de árvores filogenéticas, análise de substituição de nucleotídeos e proteínas, e muito mais. Ele também permite a visualização e análise de dados de diversas maneiras, como gráficos e tabelas.

Ele um software fácil de usar, com uma interface intuitiva e uma ampla base de usuários e recursos on-line disponíveis. Ele é compatível com Windows, MacOS e Linux e está disponível para download gratuito em seu site oficial.

Aviso

Atualmente o MEGA está na versão **11**, mas contém um ou outro bug que pode atrapalhar bastante. O MEGA **7** possui quase todas as ferramentas que precisamos e é mais estável. Recomendamos usar essa versão para coisas básicas.

Usos

Prós

- É simples e fácil de aprender.

Contras

- Roda no computador e algumas análises (como árvores ML) podem ser bem demoradas.

11.3 IQTree

Útil para: Modelo / Árvore

Usos

Prós

Contras

11.4 jModelTest

Útil para: Modelo

Usos

Prós

- Bastante reconhecido e utilizado em outros trabalhos.

Contras

- É bem demorado.

11.5 RAxML

Útil para: Árvore

Link: <https://raxml-ng.vital-it.ch>

RAxML (Randomized Axelerated Maximum Likelihood) é um software de análise filogenética utilizado para inferir árvores filogenéticas a partir de sequências de DNA, RNA ou proteínas. Ele é especialmente útil para análises com grandes conjuntos de dados, pois é capaz de processar grandes quantidades de dados de maneira rápida e eficiente.

O RAxML possui vários métodos de inferência de árvores filogenéticas, como o método de máxima verossimilhança (ML) e o método de Bayesiano. Ele também permite a realização de análises bootstrap, que são usadas para avaliar a confiabilidade das árvores filogenéticas inferidas.

Prós

- Fácil uso;
- Não requer uso de memória do computador, ele roda em servidor.

Contras

- Tem um limite de bootstrap de 100

11.6 PhyML

Útil para: Modelo / Árvore

Link: <http://www.atgc-montpellier.fr/phyml/>

Phylogenetic Maximum Likelihood (PhyML) é um servidor online para gerar inferências filogenéticas pelo método de máxima verossimilhança. O servidor é intuitivo, portanto, de fácil uso. Para gerar uma árvore filogenética no PhyML, é preciso fornecer um arquivo de alinhamento no formato PHYLIP (é possível exportar alinhamentos neste formato na versão do MEGA 10). Após o término das análises, o PhyML fornece um arquivo zip que contém, além da árvore, outros relatórios estatísticos (como a lista dos modelos evolutivos, por exemplo)

Prós:

- Fácil uso;
- Permite realizar várias customizações nos parâmetros estatísticos que não são possíveis de personalizar em outros softwares, como o MEGA;
- Possui bootstrap de 1000;
- Não requer uso de memória do computador, ele roda em servidor.

Contras:

- Utiliza o formato PHYLIP, diferente de outros que utilizam FASTA, por isso é possível que o usuário tenha que formatar os terminais da árvore, um trabalho manual desgastante. Recomendamos o uso do pacote [nupgen](#) no R para corrigir este problema;
- As análises demoram em média 8 horas (podendo variar de acordo com a quantidade de sequências). Por isso, preste atenção antes de rodar as análises.

11.7 DNAsp

Útil para: Haplótipos

11.8 Delimitadores de espécies

Útil para: Delim. Espécies

Análises de delimitação de espécies são utilizadas para determinar os limites entre espécies e definir a diversidade de espécies em um determinado grupo. Elas são muito importantes em biologia evolutiva e conservação de espécies, pois permitem o entendimento da distribuição de espécies e como elas se relacionam entre si.

Existem várias abordagens para realizar análises de delimitação de espécies, como análises de distância genética, análises de diferenças genéticas e análises filogenéticas. Cada método tem suas próprias vantagens e desvantagens, e o método ideal dependerá da aplicação específica. Além disso, é importante lembrar que a delimitação de espécies é um campo em constante evolução e novas abordagens são desenvolvidas constantemente.

Como estes são métodos complementares, serão apresentados em conjunto.

11.8.1 PTP

Link: <https://species.h-its.org/ptp/>

texto

O arquivo gerado pode ser convertido em tabela utilizando o pacote [nupgen](#).

11.8.2 ABGD

Link: <https://bioinfo.mnhn.fr/abi/public/abgd/abgdweb.html>

Artigo: <https://onlinelibrary.wiley.com/doi/10.1111/j.1365-294X.2011.05239.x>

texto

! Importante

“The **ASAP** analysis uses pairwise genetic distance to automatically find the barcode gap to divide the dataset into a list of several partitions ranked by a score. This new algorithm overcomes the limitations of its predecessor, **ABGD** ([Puillandre et al., 2012](#)), by selecting the best partition using the ‘ASAP score’ calculated by the probabilities of groups to be panmictic species and the barcode gap widths.” ([Mendoza-Ramírez et al., 2022](#))

11.8.3 ASAP

Link: <https://bioinfo.mnhn.fr/abi/public/asap/>

Artigo: <https://onlinelibrary.wiley.com/doi/10.1111/1755-0998.13281>

texto

11.8.4 GMYC

Link: <https://species.h-its.org/gmyc/>

texto

Os resultados são gerados utilizando um servidor que apenas roda um código em R na nuvem. Como o site não permite mais baixar os resultados em forma de tabela, fizemos um aplicativo em Shiny para uso interno que permite isso: https://brunomiotto.shinyapps.io/shiny_gmyc/

11.9 FigTree

Útil para: Árvore - Edição

FigTree é um software de visualização de árvores filogenéticas utilizado para exibir árvores inferidas a partir de sequências de DNA, RNA ou proteínas. Ele é compatível com Windows, MacOS e Linux e permite a visualização de árvores filogenéticas de maneira clara e intuitiva, tornando mais fácil a interpretação dos resultados.

O FigTree possui uma ampla gama de recursos de visualização, como opções de estilo de linha, rótulos de nós, legendas e muito mais. Ele também permite a exportação de árvores para vários formatos, incluindo imagens, PDF e arquivos Newick. Além disso, o FigTree possui uma interface intuitiva e é fácil de usar, mesmo para usuários sem experiência prévia em análises filogenéticas.

Prós

Contras

11.10 iTOL

Útil para: **Árvore - Edição**

Link: <https://itol.embl.de/>

Prós

Contras

11.11 PopArt

Útil para: **Haplótipos / Árvore - Edição**

A capacidade de análise de milhões de regiões de sequência de DNA em uma só vez é exigida em diversas disciplinas, tais como genômica populacional e ecologia molecular. Para tanto, o DnaSP, derivado do inglês “DNA Sequence Polymorphism” (Polimorfismo de Sequência de DNA) é uma ferramenta bioinformática concebida para a análise abrangente da variação dos dados da sequência de DNA, utilizando uma interface gráfica de utilizador de fácil uso. A versão 6 do software apresenta um PDF disponível para consulta disponibilizado no próprio site e pode ser acessado pelo link: http://www.ub.edu/dnasp/DnaSP6_Documentation_6.12.pdf.

O programa permite a caracterização detalhada dos níveis e padrões de variação da sequência de DNA em diferentes escalas de tempo, utilizando variantes polimórficas (dados intra-específicos), dados de divergência (dados interespecíficos ou interpopulação), ou uma combinação de ambos. É possível calcular diversidade haplotípica, nucleotídica, bem como a quantidade distintas de haplótipos (quando analisadas duas ou mais sequências de DNA).

12 Pacotes R

Ao realizar pesquisas ecológicas, acabamos nos deparando com quantidades absurdas de dados, com planilhas e mais planilhas. Por esse motivo, é muito comum utilizarmos a linguagem de programação **R** em pesquisas ecológicas.

Utilizamos o R para organizar dados, tratar tabelas, criar índices, realizar testes estatísticos e finalizar com lindos gráficos. Tudo isso é possível através do uso das ferramentas certas, que geralmente são distribuídas em **pacotes**, que nada mais são que conjuntos de funções específicas.

Obviamente é inviável e até foge do escopo deste livro trazer uma introdução ao R e ecologia quantitativa, mas recentemente foi lançado um livro on-line escrito por pesquisadores brasileiros com toda uma introdução neste tópico, é o livro [Análises Ecológicas no R](#) (que foi escrito no R como este aqui). Caso precise de ajuda com o R ou afins, estamos à disposição para ajudar (praticamente todos os laboratórios do Nupélia utilizam o R para análises).

Para a pesquisa com Genética Molecular, existem alguns pacotes que podem ajudar.

12.1 Pacote nupgen (R)

Link: <https://brunomioto.github.io/nupgen/>

Este é um pacote criado com algumas funções internas do nosso laboratório, são elas:

- Renomear arquivos **PHYLIP** com `label_phy()`
- Obter informações de sequências do **GenBank** com `info_genbank()`

As informações buscadas são:

- Código do GenBank
- Nome do organismo
- País de origem (pode conter mais informações)
- Latitude
- Longitude

– Gene

- Formatar resultados da [Delimitação de Espécies](#) do PTP (ou bPTP) com `ptp_results()`

Instalação

```
if (!requireNamespace("remotes", quietly = TRUE)){
  install.packages("remotes")
}
remotes::install_github("brunomiotto/nupgen")
```

12.2 Rocc

Link: <https://docs.ropensci.org/bold/>

Baixa dados do specieslink

Instalação

```
if (!requireNamespace("remotes", quietly = TRUE)){
  install.packages("remotes")
}
remotes::install_github("liibre/Rocc")
```

12.3 bold

Link: <https://docs.ropensci.org/bold/>

Baixa dados do bold

Instalação

```
install.packages("bold")
```

12.4 refdb

Link: <https://fkeck.github.io/refdb/index.html>

Baixa dados do genbank

Instalação

```
if (!requireNamespace("remotes", quietly = TRUE)){  
  install.packages("remotes")  
}  
remotes::install_github("fkeck/refdb")
```

12.5 Herodotools

Link: <https://gabrielnakamura.github.io/Herodotools/index.html>

Biogeografia

Instalação

```
if (!requireNamespace("remotes", quietly = TRUE)){  
  install.packages("remotes")  
}  
remotes::install_github("GabrielNakamura/Herodotools")
```

12.6 ggtree

Link: <https://guangchuangyu.github.io/ggtree-book/chapter-ggtree.html>

Edição de árvores

Instalação

```
install.packages("ggtree")
```

12.7 ggmsa

Link: <http://yulab-smu.top/ggmsa/>

Visualização de sequências

Instalação

```
if (!requireNamespace("remotes", quietly = TRUE)){  
  install.packages("remotes")  
}
```

```
remotes::install_github("YuLab-SMU/ggmsa")
```

13 Edição de sequências

Após enviar nossas amostras para o sequenciamento, recebemos um conjunto de arquivos com os resultados. Mas antes de utilizá-los em análises, precisamos garantir a qualidade dos dados recebidos. É normal e esperado que sejam encontrados erros de leitura do sequenciador, mas devemos tomar cuidado para tornar esse processo livre de vieses e erros de interpretação.

Os arquivos geralmente são enviados em 2 formatos, um deles é responsável pela sequência de nucleotídeos de forma básica, geralmente no formato **.fasta/.fas** (é neste formato que baixamos as sequências do [GenBank](#)) ou **.seq**, e o outro arquivo é o responsável pelos dados do eletroferograma, onde podemos observar os picos do sinal do detector do sequenciador, recebido geralmente em formato **.ab1**.

Os arquivos FASTA podem ser abertos no programa [MEGA](#) (ou mesmo no bloco de notas para uma edição rápida no nome das sequências, por exemplo), enquanto os eletroferogramas podem ser visualizados com o programa [BioEdit](#).

13.1 Passo a passo para edição de sequências

Para realizar uma edição correta e livre de vieses, vamos utilizar papel e caneta além do BioEdit e MEGA. Peça ajuda para alguém que já fez isso antes.

1. Abra o eletroferograma com o BioEdit e anote cada observação da sequência em um papel, destacando o início e final da sequência, junto da numeração do nucleotídeo e a modificação a ser feita, por exemplo:

Seq1 - A04

Início

34

----->

ATGCTA

pb 134 145 167

Antes	Y	K	W

Agora C G A

Fim

673

----->

TGCCTA

Faça isso para a sequência inteira. O ideal é que você faça isso com todas as suas sequências de uma vez.

2. Abra o MEGA e realize as edições na seguinte ordem: 1. Correções de pb, 2. Corte do final da sequência, 3. Corte do início da sequência. Isso vai garantir que os números relativos a cada nucleotídeo não mude durante a edição.

14 Alinhamento

MEGA, raxML, phyML

Adicionar referências para cada um

MEGA 7:

Referências

- Anthony J. F. Griffiths, Susan R. Wessler, Sean B. Carroll, e John Doebley. 2016. *Introdução à Genética*. 11^o ed. Guanabara Koogan.
- Bruce Alberts, Alexander Johnson, Julian Lewis, David Morgan, Martin Raff, Keith Roberts, Peter Walter, John Wilson, e Tim Hunt. 2017. *Biologia Molecular da Célula*. 6^o ed. Porto Alegre: Artmed.
- Peter Snustad, e Michael J. Simmons. 2017. *Fundamentos da Genética*. 7^o ed. Rio de Janeiro: Guanabara Koogan.
- Rosenthal, André, Oliver Coutelle, e Molly Craxton. 1993. "Large-scale production of DNA sequencing templates by microtitre format PCR". *Nucleic acids research* 21 (1): 173–74. <http://www.ncbi.nlm.nih.gov/pubmed/8441614>.