



UNIVERSIDADE FEDERAL DA BAHIA
Instituto Multidisciplinar em Saúde
Campus Anísio Teixeira



Reconstrução Filogenética

Bioinformática

Prof. Dr. Leandro Martins de Freitas



Introdução

A reconstrução filogenética permite a construção de árvores que representam o processo de divergências entre os organismos e suas relações de ancestralidade. A árvore filogenética é um modelo que permite compreendermos como foi o processo de diversificação usando diferentes tipos dados como moléculas de DNA ou proteínas.

A reconstrução filogenética é uma ferramenta fundamental na compreensão da evolução de todos os organismos como vírus e bactérias patogênicas, desempenhando um papel crucial na saúde pública. Por meio da análise filogenética, os cientistas podem traçar a história evolutiva desses microrganismos, identificar padrões de disseminação e entender melhor como surgem novas cepas com potencial de causar doenças (epidemiologia molecular). Esta abordagem utiliza técnicas como a análise de sequências de DNA para reconstruir árvores filogenéticas que representam as relações evolutivas entre diferentes linhagens.

Ao estudar a evolução de vírus (SARS-CoV-2, vírus da dengue, influenza) e bactérias patogênicas, os pesquisadores podem identificar mutações genéticas que conferem resistência a medicamentos, facilitando a seleção de estratégias mais eficazes de tratamento e prevenção. Além disso, a reconstrução filogenética permite rastrear a origem de surtos de doenças infecciosas, ajudando na implementação de medidas de controle e contenção.

A reconstrução filogenética auxilia a saúde pública no estudo da evolução do vírus da influenza. Através da análise filogenética, os cientistas podem monitorar as mudanças genéticas do vírus ao longo do tempo e prever quais cepas são mais prováveis de causar epidemias sazonais ou pandemias. Isso é crucial para o desenvolvimento de vacinas eficazes e estratégias de mitigação.

A reconstrução filogenética desempenha um papel importante no rastreamento de surtos de doenças emergentes, como o COVID-19. Ao sequenciar o genoma do vírus SARS-CoV-2 e compará-lo com outros coronavírus, os cientistas podem entender melhor a origem do vírus, sua transmissão entre espécies e as rotas de propagação global. Isso permite uma resposta mais ágil e direcionada por parte das autoridades de saúde, incluindo a implementação de medidas de distanciamento social, quarentenas e desenvolvimento de vacinas.

Para a reconstrução filogenética devemos fazer a escolha de amostras ou espécies representantes dos táxons (unidades taxonômicas operacionais - OTUs) e sequências de DNA ou proteínas.



Durante o processo de escolha dos táxons você deve escolher aqueles que permitem uma representação das espécies e da diversidade, enquanto a escolha das sequências deve ser adequada para a taxa de evolução da sequência (sequências com muita diversidade e rápida evolução vs. sequências muito conservadas, pouca diversidade e evolução lenta) e distância filogenética das OTUs.

Tutorial

Vamos realizar a reconstrução filogenética seguindo as etapas

1° Realizar o alinhamento das sequências no programa MEGA

Iremos usar o algoritmo de alinhamento MUSCLE dentro do programa MEGA

Para saber mais: Kumar S, Stecher G, Li M, Knyaz C, Tamura K. MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms. *Mol Biol Evol.* 2018 Jun 1;35(6):1547-1549. doi: 10.1093/molbev/msy096. PMID: 29722887; PMCID: PMC5967553.

Para saber mais: Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* 2004 Mar 19;32(5):1792-7. doi: 10.1093/nar/gkh340. PMID: 15034147; PMCID: PMC390337.

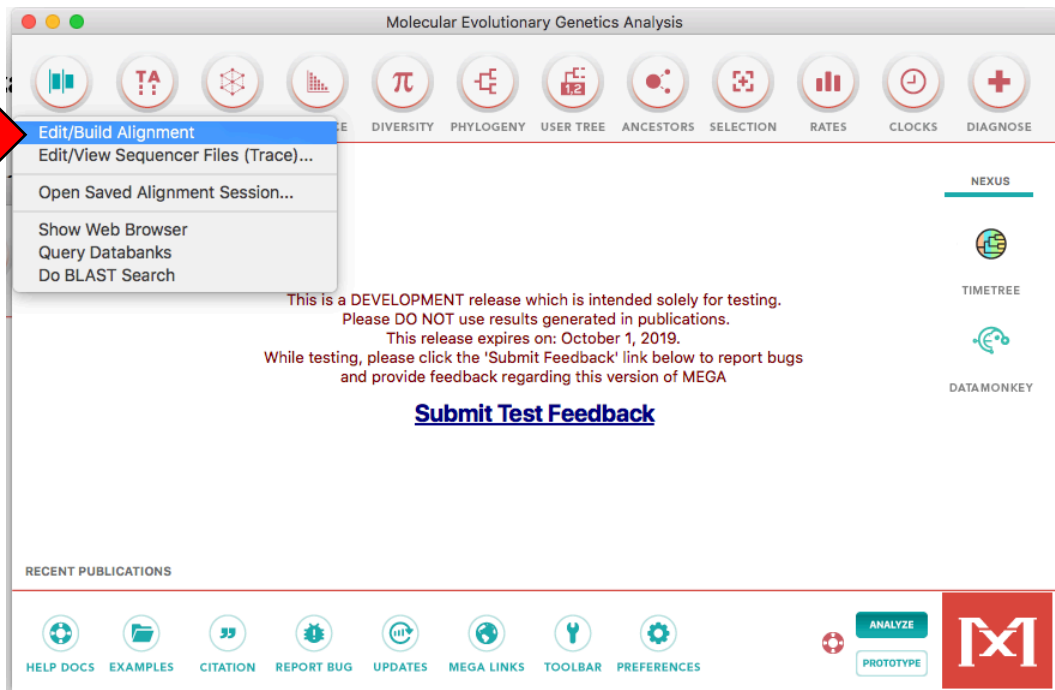
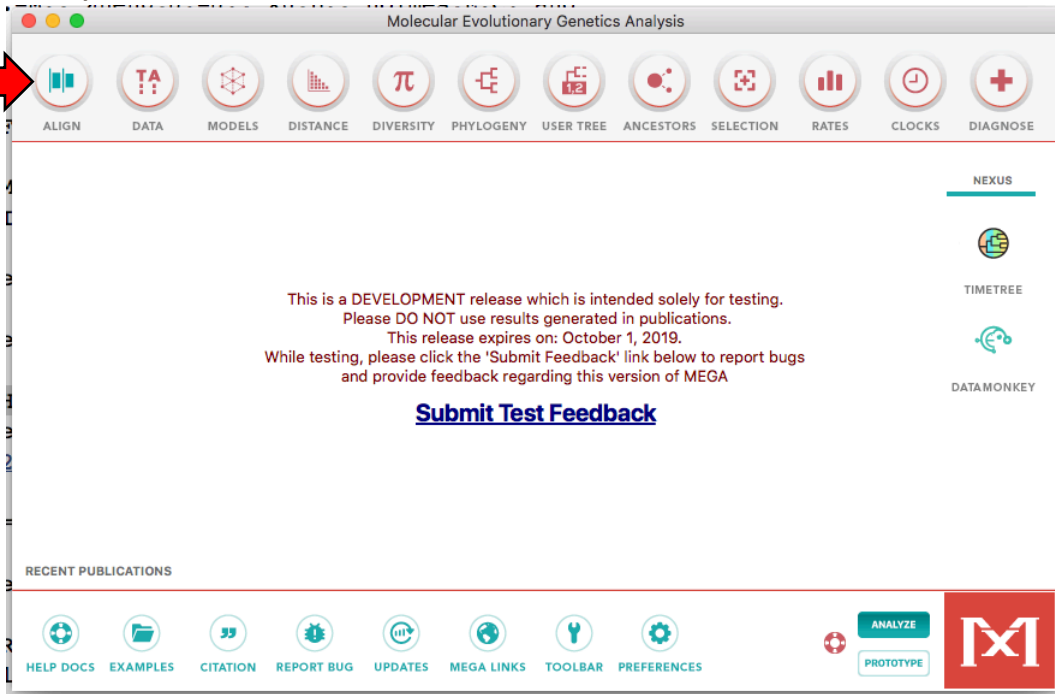
Iremos usar o gene *alcohol dehydrogenase* (Adh) para fazer a reconstrução de algumas espécies do gênero *Drosophila* como exemplo (Observe que muitas linhagens de *Drosophila* são excluídas deste exemplo por uma questão de simplicidade).

Este é apenas um exemplo do processo de construção de árvores. As árvores filogenéticas geralmente são baseadas em muito mais caracteres e geralmente envolvem mais linhagens. As reconstruções podem conter dezenas de espécies e caracteres moleculares. Cada variação na sequência será usada pelos métodos de reconstrução como uma característica.

Arquivo: *Drosophila_Adh.fasta*

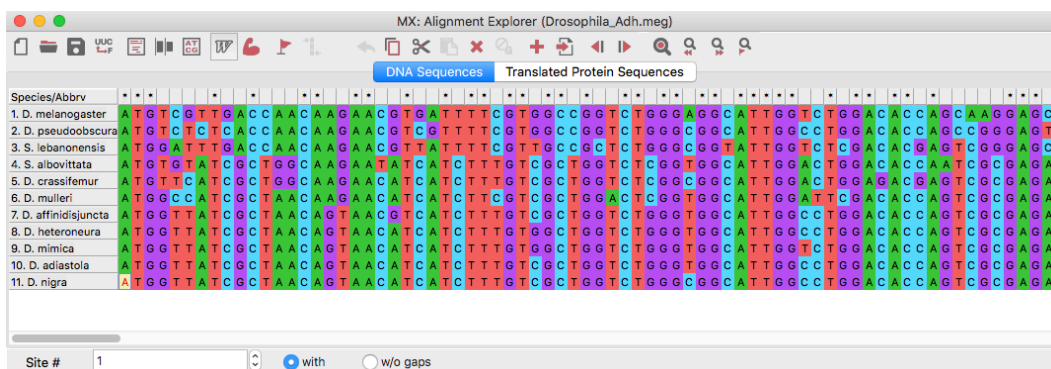
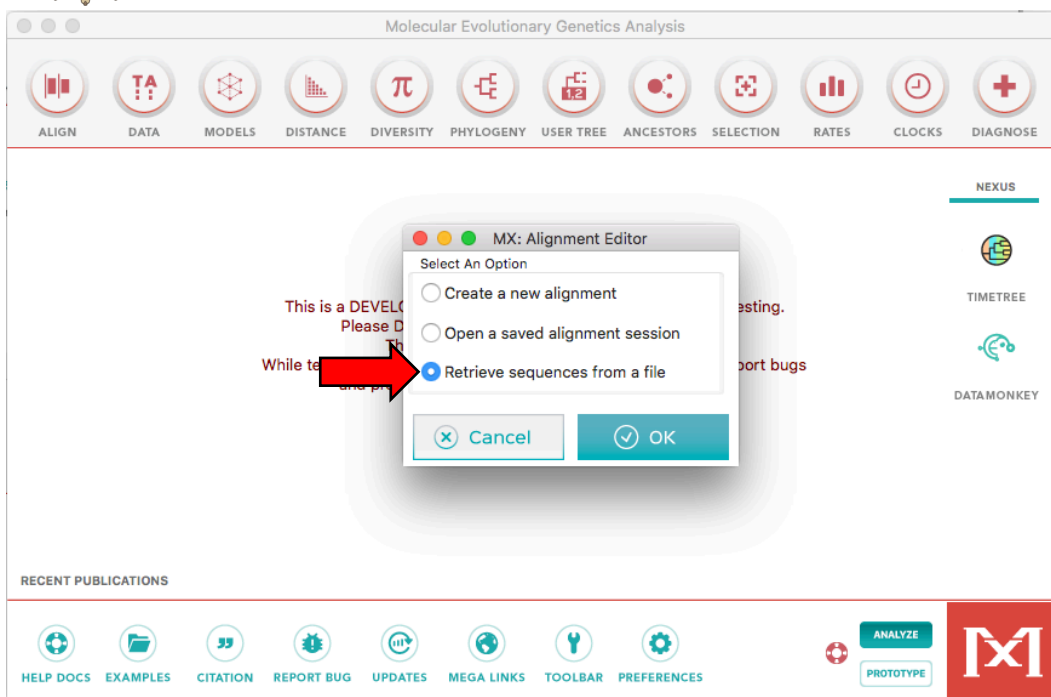


UNIVERSIDADE FEDERAL DA BAHIA
Instituto Multidisciplinar em Saúde
Campus Anísio Teixeira

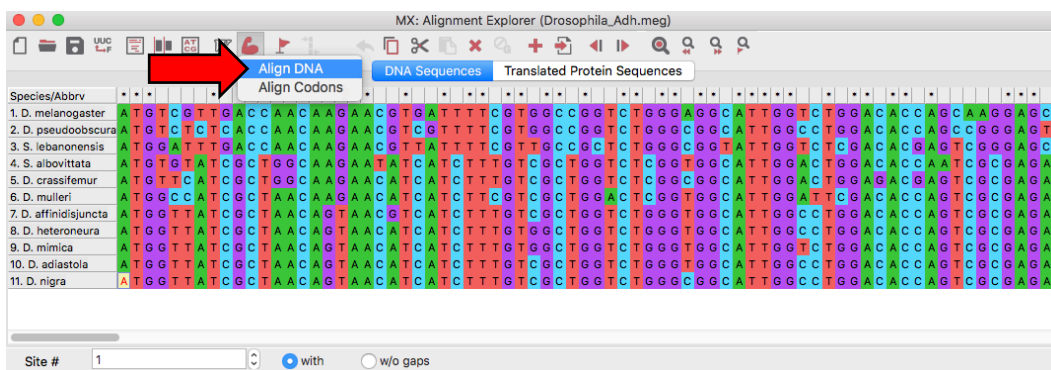


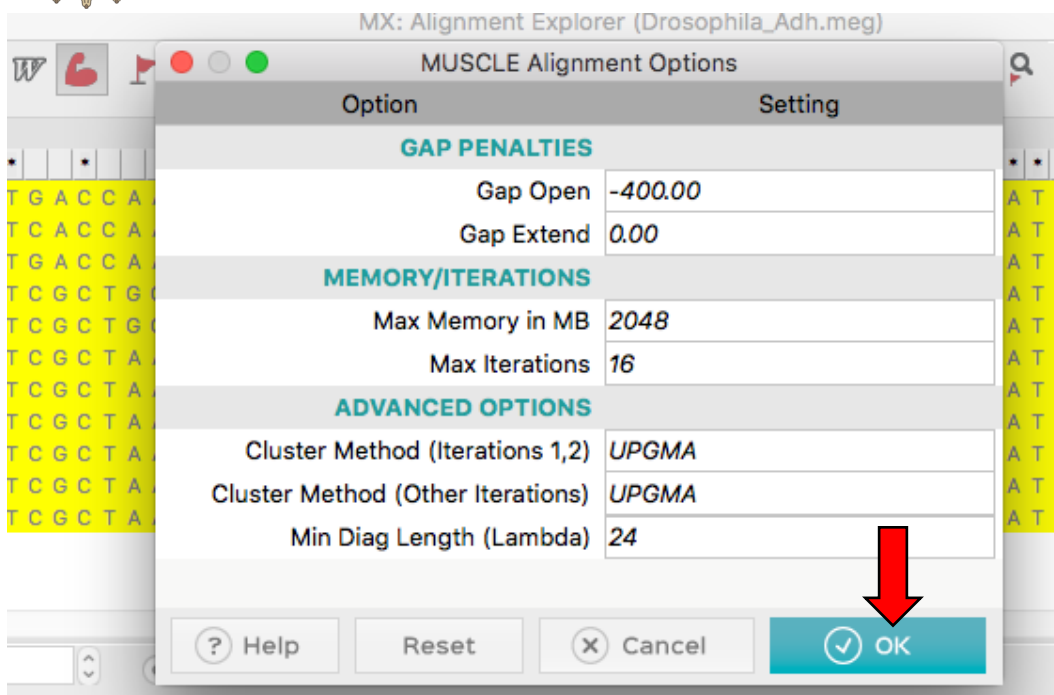


UNIVERSIDADE FEDERAL DA BAHIA
Instituto Multidisciplinar em Saúde
Campus Anísio Teixeira



Se necessário, remover as sequências indesejadas (tamanhos muito diferentes) antes do alinhamento.





2º Após alinhar a sequência salvar no formato MEGA Format (.meg)

Os arquivos no formato MEGA (“.meg”) serão usados nas etapas de análise das sequências.

Salvando no formato MEGA:

File → Export Alignment → MEGA Format

Nome do arquivo: Drosophila_Adh.meg

Agora que realizamos o alinhamento das sequências, podemos fechar a janela de alinhamento. Salve o alinhamento no formato FASTA ou MAS. Salvar o alinhamento é importante para que você não tenha que realizar novamente essa etapa para esse conjunto de sequências.

Salvando o alinhamento no formato “.mas” ou fasta:

Data → Save Session

Nome do arquivo: Drosophila_Adh.mas

ou

Data → Export Alingment → FASTA Format

Nome do arquivo: Drosophila_Adh_aln.fasta

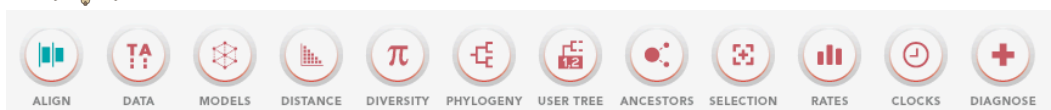
Feche a janela do MEGA “Alignment Explorer”

3º Vamos abrir o arquivo Drosophila_Adh.meg que foi salvo usando análise de filogenia (*Phylogeny*)





UNIVERSIDADE FEDERAL DA BAHIA
Instituto Multidisciplinar em Saúde
Campus Anísio Teixeira



This is a DEVELOPMENT release which is intended solely for testing.
Please DO NOT use results generated in publications.
This release expires on: October 1, 2019.
While testing, please click the 'Submit Feedback' link below to report bugs
and provide feedback regarding this version of MEGA

[Submit Test Feedback](#)

NEXUS



TIMETREE



DATA MONKEY

RECENT PUBLICATIONS



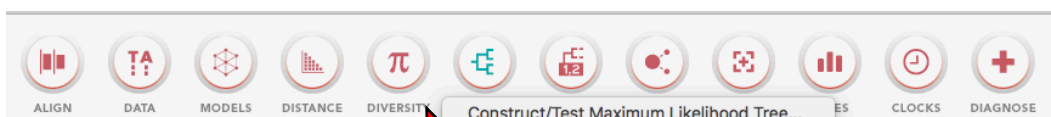
ANALYZE
PROTOTYPE



Vamos usar o método de distância NJ (*neighbor joining*)

Para saber mais: Saitou N, Nei M. The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol Biol Evol. 1987 Jul;4(4):406-25. doi: 10.1093/oxfordjournals.molbev.a040454. PMID: 3447015.

Para saber mais: Tamura K, Nei M, Kumar S. Prospects for inferring very large phylogenies by using the neighbor-joining method. Proc Natl Acad Sci U S A. 2004 Jul 27;101(30):11030-5. doi: 10.1073/pnas.0404206101. Epub 2004 Jul 16. PMID: 15258291; PMCID: PMC491989.



This is a DEVELOPMENT release which is intended solely for testing.
Please DO NOT use results generated in publications.
This release expires on: October 1, 2019.
While testing, please click the 'Submit Feedback' link below to report bugs
and provide feedback regarding this version of MEGA

[Submit Test Feedback](#)

NEXUS



TIMETREE



DATA MONKEY

RECENT PUBLICATIONS



ANALYZE
PROTOTYPE





UNIVERSIDADE FEDERAL DA BAHIA
Instituto Multidisciplinar em Saúde
Campus Anísio Teixeira



Iremos aplicar o teste de filogenia (*Teste of Phylogeny*) usando o método de Bootstrap com 1000 replicas.

Para saber mais: Russo CAM, Selvatti AP. Bootstrap and Rogue Identification Tests for Phylogenetic Analyses. Mol Biol Evol. 2018 Sep 1;35(9):2327-2333. doi: 10.1093/molbev/msy118. PMID: 29878143.

Para saber mais: Felsenstein J. (1985). Confidence limits on phylogenies: An approach using the bootstrap. Evolution 39:783-791.

Após seleccionar o método de teste podemos seguir para a reconstrução (clique em ok).

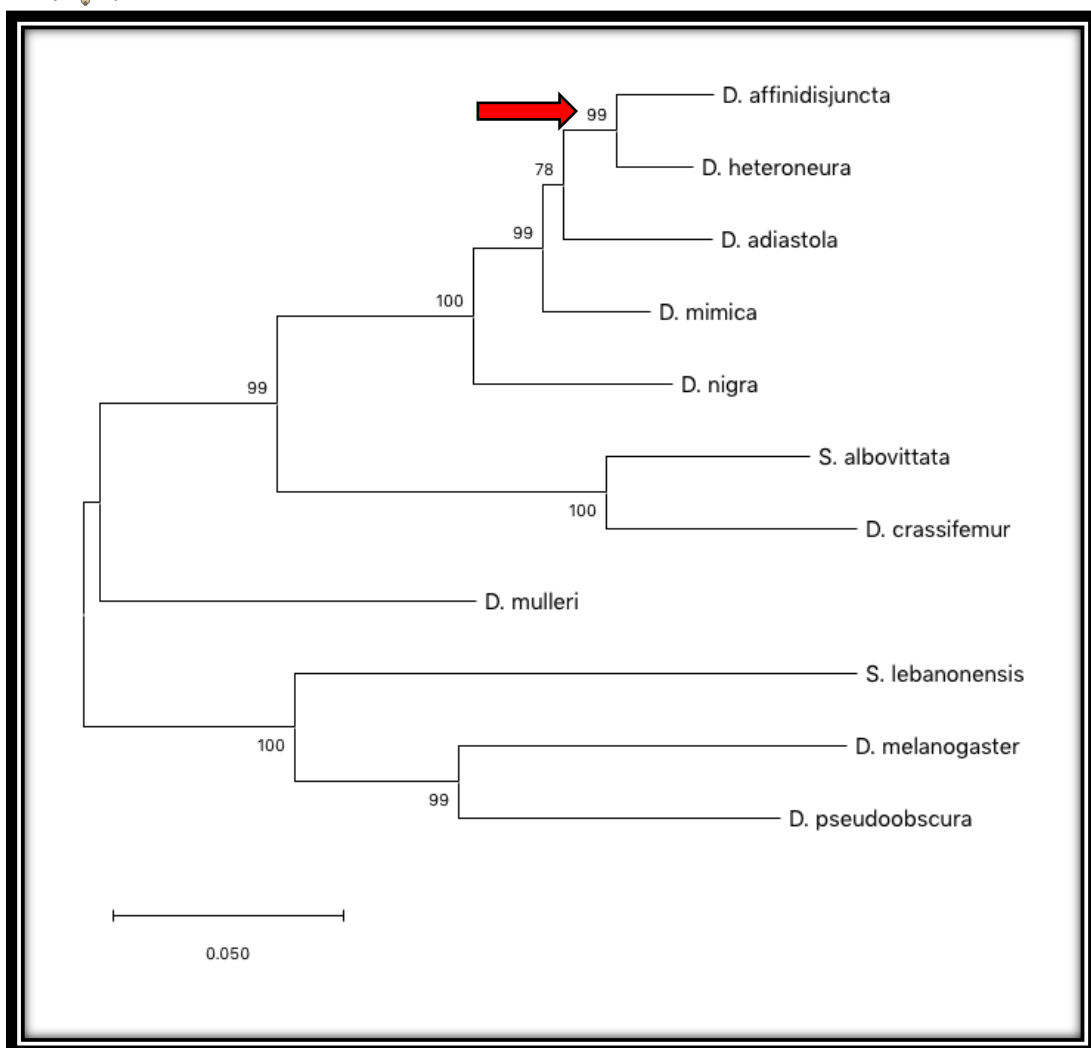
MX: Analysis Preferences

Phylogeny Reconstruction

Option	Setting
ANALYSIS	
Scope	All Selected Taxa
Statistical Method	Neighbor-joining
PHYLOGENY TEST	
Test of Phylogeny	Bootstrap method
No. of Bootstrap Replications	1000
SUBSTITUTION MODEL	
Substitutions Type	Nucleotide
Genetic Code Table	Not Applicable
Model/Method	Maximum Composite Likelihood
Fixed Transition/Transversion Ratio	Not Applicable
Substitutions to Include	d: Transitions + Transversions
RATES AND PATTERNS	
Rates among Sites	Uniform Rates
Gamma Parameter	Not Applicable
Pattern among Lineages	Same (Homogeneous)
DATA SUBSET TO USE	
Gaps/Missing Data Treatment	Pairwise deletion
Site Coverage Cutoff (%)	Not Applicable
Select Codon Positions	<input checked="" type="checkbox"/> 1st <input checked="" type="checkbox"/> 2nd <input checked="" type="checkbox"/> 3rd <input checked="" type="checkbox"/> Encoding Sites
SYSTEM RESOURCE USAGE	
Number of Threads	3

Help Cancel OK

O programa vai retornar uma árvore filogenética parecida com a árvore abaixo.



Os números próximos aos nós da árvore representa a confiança do agrupamento. Esses números estão em porcentagem das 1000 replicas do teste de *bootstrap* que escolhemos. Indico o valor mínimo de 70% para considerar um agrupamento confiável.

Raiz da árvore - Grupo externo

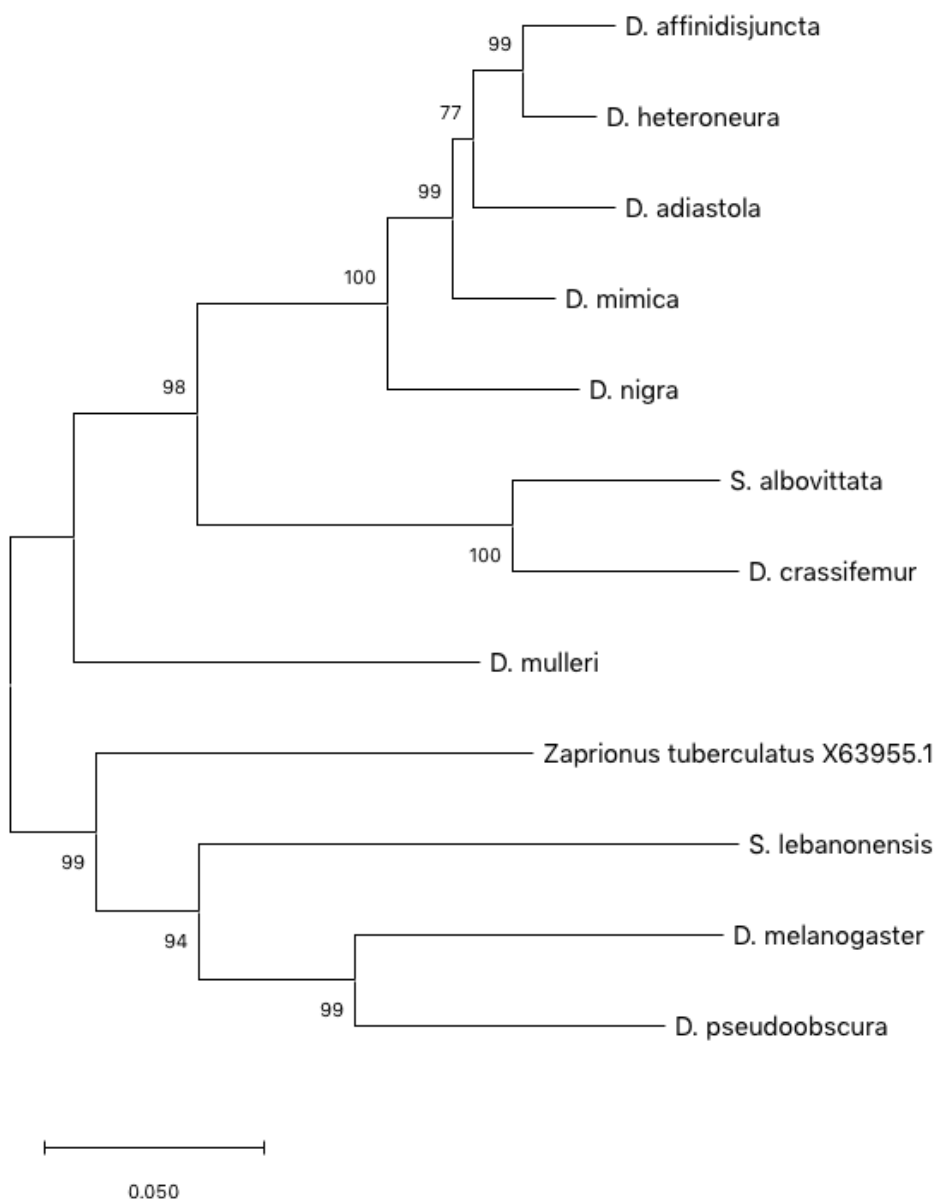
O método de enraizamento da árvore por grupo externo é uma técnica utilizada em análises filogenéticas para determinar a posição das raízes das árvores evolutivas. Consiste em adicionar um grupo externo, que não está diretamente relacionado ao grupo de interesse, à análise. Esse grupo externo é usado como referência para determinar a direção da evolução. Isso permite entender como os grupos evoluíram ao longo do tempo e em que direção as mudanças ocorreram. Vamos agora fazer novamente a reconstrução filogenética usando um arquivo fasta com as mesmas sequências de *Drosophila* e o gene *alcohol dehydrogenase* (*Adh*). Dessa vez vamos adicionar uma sequencia como grupo externo do gene *alcohol dehydrogenase* (*Adh*) da espécie *Zaprionus tuberculatus*.



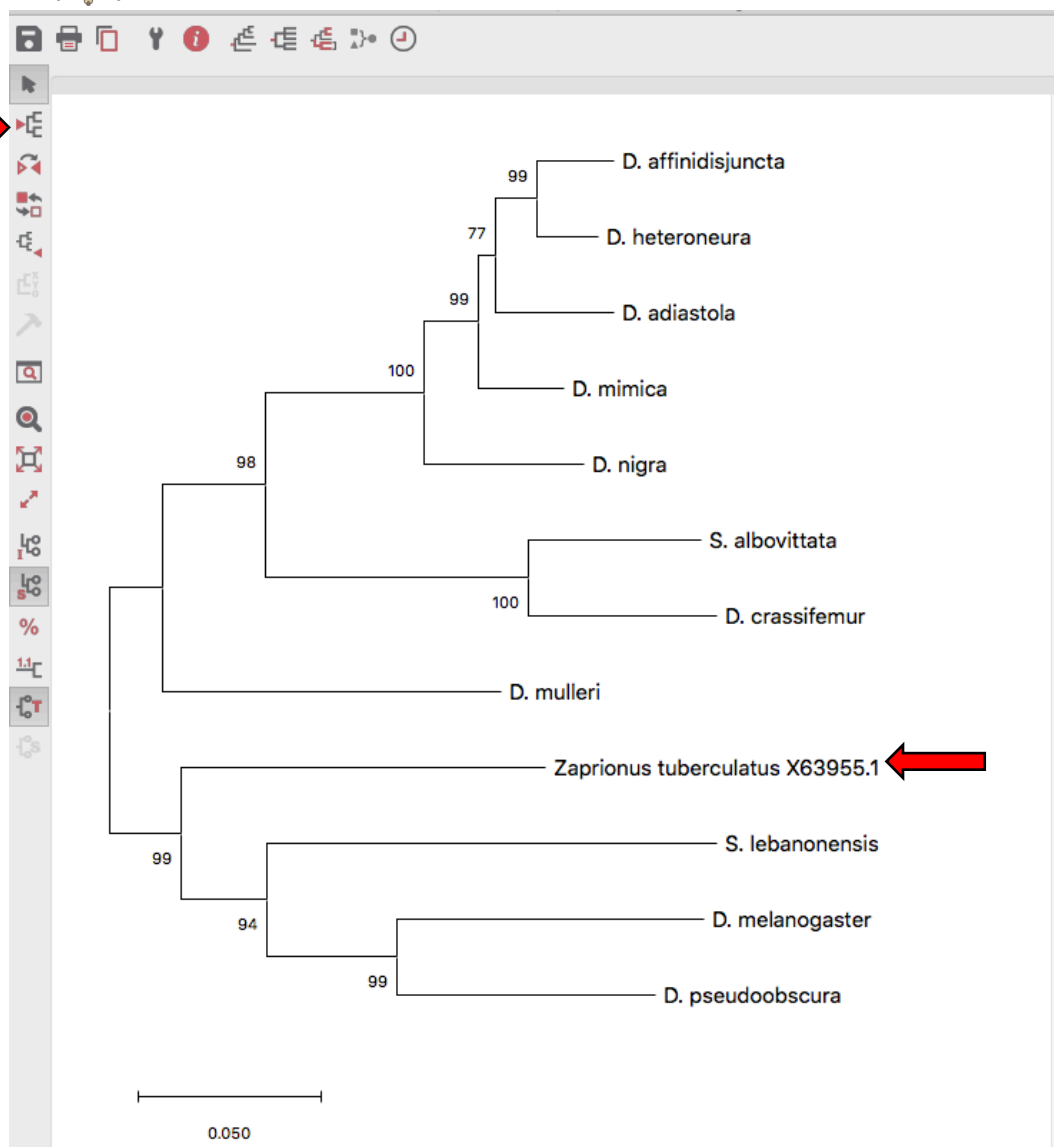
UNIVERSIDADE FEDERAL DA BAHIA
Instituto Multidisciplinar em Saúde
Campus Anísio Teixeira



Após a reconstrução você terá uma árvore parecida com essa.



Agora escolha a espécie *Zaprionus tuberculatus* como grupo externo como indicado na figura abaixo.



A árvore abaixo será a reconstrução com a espécie *Zaprionus tuberculatus* como grupo externo indicando o direcionamento da árvore.

