

# Atividade Final de Fiogenia Molecular

Gabriel Rodrigues \*

Universidade Estadual de Santa Cruz,  
gvpina.rodrigues@gmail.com

**Abstract.** Atividade avaliativa realizada com a finalidade de aplicar todos os conceitos teóricos e práticos aprendidos em sala de aule e durante os estudos dirigidos. Essa atividade se propõe a fazer uma análise filogenética de 43 sequências do gene ribossomal mitochondrial 16S de abelhas sem ferrão (tribo Meliponini).

**Keywords.** Filogenia; MSA; Bioinformática; Abelhas

## Introdução

O uso de métodos moleculares é essencial para contar a história dos eventos evolutivos que ocorreram com um táxon em determinado tempo. As evidências necessárias para descrever o processo devem ser coletadas com critérios rígidos, para não gerar dados ruidosos em uma reconstrução filogenética.

Métodos filogenéticos são baseados em inferências, já que não é possível reconstruir os elementos já perdidos na natureza. Na maioria dos trabalhos, as amostras serão parciais. Fatores como a diversidade interna do gênero/família são chave para determinar a completude da filogenia molecular.

Para esse trabalho em particular, iremos usar o arquivo `seqs.fasta`, que possui 43 sequências do gene ribossomal mitochondrial 16S de abelhas sem ferrão (tribo Meliponini).

## Escolha do Caractere

Um caractere, seja molecular ou morfológico, deve ter uma variação intrísccea entre as amostras (táxons estudados), porém sem ser muito divergente ao ponto de não poder ser utilizado como um caractere conservado entre as amostras.

Alguns dos caracteres escolhidos vão estar ausentes, no caso de amostras moleculares pode ser o caso de uma base nitrogenada não sequenciada ou com baixa qualidade.

## Objeto da Análise

A subunidade 16S do RNA Ribossomal é eficaz para resconstruir relações evolutivas se deve à propriedades como distribuição universal, função conservada, taxa de evolução lenta e contém regiões variáveis em sua composição.

---

\*Corresponding author.

## Curadoria das Sequências

Em primeiro lugar, é necessário padronizar os nomes das sequências de input para as análises subsequentes. Como os arquivos foram retirados do GenBank, eles possuem a seguinte estrutura de *header*:

```
>AF343118.1 Lepidotrigona ventralis 16S large subunit ribosomal RNA gene  
TTGTATATTGTATAATGAAATCTGGAATGAAAGGATTAATGAAATAT
```

Em arquivos FASTA, o *header* é indicado como sendo o elemento após o símbolo de > (maior que), indicando o nome da sequência e outras informações relevantes. Para as análises futuras, é vantajoso reduzir o nome para uma melhor visualização da árvore e suporte dos softwares de alinhamento. Podemos reduzir para a seguinte estrutura:

```
>Lepidotrigona_ventralis  
TTGTATATTGTATAATGAAATCTGGAATGAAAGGATTAATGAAATAT
```

## Alinhamento Múltiplo de Sequências (MSA)

MAFFT (Multiple Alignment with Fast Fourier Transform, em inglês) é um programa de alinhamento múltiplo de sequências. O programa MAFFT implementa transformada rápida de Fourier para otimizar os alinhamentos de proteínas com base nas propriedades físicas dos aminoácidos (Katoh et al. 2002). O programa usa o alinhamento progressivo e iterativo. As sequências nucleotídicas e de aminoácidos no formato FASTA podem ser alinhadas. MAFFT é útil para as sequências mais difíceis de se alinhar com outros programas, tais como aquelas que contêm grandes lacunas, (ex: rRNA que contêm as regiões variáveis de laços).

O MAFFT foi instalado no sistema operacional ArchLinux via ambiente conda, através do comando:

```
conda install -c bioconda mafft
```

A partir da sua instalação, podemos executar o alinhamento com o MAFFT a partir do seguindo comando:

```
mafft --maxiterate 1000 --localpair seqs.fasta > alinhamento.fasta
```

## Modelo de Substituição Nucleotídica

Modelo do IQ-TREE: ModelFinder computes the log-likelihoods of an initial parsimony tree for many different models and the Akaike information criterion (AIC) corrected Akaike information criterion (AICc), and the Bayesian information criterion (BIC). Then ModelFinder chooses the model that minimizes the BIC score (you can also change to AIC or AICc by adding the option -AIC or -AICc, respectively).

Checar a descrição dos outros modelos depois.

# **Inferência Filogenética**

## **Visualização e Edição da Árvore**

## **Bibliography**

Katoh, Kazutaka, Kazuharu Misawa, Kei-ichi Kuma, and Takashi Miyata. 2002. “MAFFT: A Novel Method for Rapid Multiple Sequence Alignment Based on Fast Fourier Transform.” *Nucleic Acids Research* 30 (14): 3059–66. <https://doi.org/10.1093/nar/gkf436>.