

Durante o primeiro ciclo de encontros abordaremos os bancos de dados biológicos, que são muito relevantes, seja para a sua aquisição de dados no início do estudo quanto para que você submeta seus dados ao final da sua pesquisa.

Para os primeiros encontros deste ciclo, vamos abordar um tipo específico de banco de dados: aqueles que contêm dados vindos do sequenciamento de DNA e RNA, especialmente como podemos adquirir tais dados e filtrar aqueles que não são relevantes para o nosso estudo.

Além dos materiais de apoio de cada encontro, vamos disponibilizar também um Desafio da Semana. Caso você já se sinta confortável com algum dos tópicos da seção “Materiais de Apoio”, pode ignorá-lo sem problemas, a única parte obrigatória é a instalação dos programas da parte prática antes do encontro em si.

Primeiro encontro (24/09/2020)

TEMA:

O que são e pra que usar banco de dados biológicos de sequenciamento?

DESAFIO DA SEMANA:

Para esse primeiro encontro, propomos que cada um busque o banco de dados biológicos mais aleatório que conseguir, se quiser, um bom ponto de partida é a página da Wiki listando alguns bancos de dados biológicos: [List of biological databases](#)

PROGRAMAS DA PARTE PRÁTICA:

Durante a prática usaremos o pacote de programas SRA-Tools, cujo download pode ser feito [aqui](#). Caso você utilize o canal [Bioconda](#) do pacote conda, pode instalar também com o comando, disponível em <https://anaconda.org/bioconda/sra-tools>:

```
conda install -c bioconda sra-tools
```

Dentre os comandos desse pacote vamos abordar o *prefetch*, o *fastq-dump* e sua versão mais atual, *fasterq-dump*.

Você pode ir se acostumando com tais comandos, sendo que as flags (parâmetros) mais relevantes, junto com uma explicação de seu uso, do *fastq-dump*, estão disponíveis [aqui](#), e as flags do *fasterq-dump*, junto com uma explicação do programa podem ser encontradas [aqui](#).

Além disso, veremos algumas funcionalidades dos sites Galaxy, disponível em <https://usegalaxy.org/>, e a função Run Selector do SRA, disponível em <https://www.ncbi.nlm.nih.gov/Traces/study/?>.

MATERIAIS DE APOIO

1. Materiais sobre sequenciamento

[Vídeo curto](#) e [texto](#) que explicam o processo de sequenciamento de DNA pela técnica de Sanger, uma das primeiras do ramo.

Artigo em português explicando sobre as diferentes plataformas de sequenciamento: <https://medium.com/omixdata/entendendo-o-ngs-parte-i-plataformas-de-sequenciamento-66017d6d38d3>

2. O que são banco de dados biológicos
[Aula sobre bancos de dados biológicos](#)

3. NCBI

Site do [National Center for Biotechnology Information](#), um dos grandes agregadores de bancos de dados e outras informações de relevância médica e biotecnológica, desde artigos, via [PubMed](#), até ferramentas de análise e comparação de sequências biológicas, como o [BLAST](#).

4. SRA

Site do [Sequence Read Archive](#), o banco de dados (primário) que será abordado no primeiro encontro e que conta com sequências vindas de sequenciamentos, especialmente sequências curtas (com menos de 1000 pares de bases) vindas de sequenciadores de nova geração.

Esse banco conta com o [SRA-Tools](#), que é um conjunto de ferramentas (scripts) para download e transformações básicas desses dados em formatos mais usáveis.

5. Artigo de exemplo da parte prática

O seguinte artigo: [Small RNA Transcriptomes of Two Types of Exosomes in Human Whole Saliva Determined by Next Generation Sequencing](#), será utilizado como base para a realização da parte prática do encontro. Em específico, os dados do artigo no banco de dados estão disponíveis no seguinte link:
<https://www.ncbi.nlm.nih.gov/bioproject/247214>