



Tópico 01 - Introdução ao Linux e ao Ambiente de Bioinformática



1. Por que Linux é essencial na Bioinformática?

A maior parte das ferramentas usadas em bioinformática (FastQC, MultiQC, Salmon, BWA, samtools, bcftools, GATK, etc.) foi desenvolvida originalmente para **Linux/Unix**, principalmente porque:

- 🧠 **Desempenho muito superior** para análises de grande volume.
- ⚙️ **Automação de tarefas** com Shell Script.
- 🖥️ **Compatibilidade com clusters e servidores HPC.**
- 🧬 **Ferramentas de análise genômica são nativas de Linux.**
- 🔄 **Processamento em lote** de arquivos FASTQ/FASTA, VCF, BAM e GTF.

👉 Em bioinformática, **quase tudo acontece no terminal.**



2. Terminal, Shell e Linha de Comando: o que são?

Terminal

É a tela onde você digita comandos.

Exemplos: Terminal do Ubuntu, WSL, MacOS Terminal.

Shell

É o “intérprete” dos comandos.

O mais usado é o **bash** (Bourne Again SHell).

Linha de Comando (CLI)

É a interface onde você digita comandos como:

```
ls
cd pasta/
mkdir resultados
```

👉 Enquanto bioinformatas usam termos como “rode o script”, “suba no cluster”, “execute o pipeline”, tudo isso significa **usar o shell**.



3. Estrutura básica do sistema Linux

O Linux organiza tudo em uma única estrutura hierárquica.

Os diretórios mais importantes para bioinformática são:

Diretório	Função
/home/usuario	Sua área de trabalho. Aqui ficam seus projetos.
/usr/bin	Onde ficam instalados programas executáveis.
/etc	Arquivos de configuração.
/mnt ou /media	Onde discos externos são montados.
/tmp	Arquivos temporários (geralmente apagados no reboot).



Para bioinformática:

Guarde seus dados em:

```
/home/seu_nome/projeto/
```



4. Primeiro comando: quem sou eu no sistema?

```
whoami
```

5. Onde estou agora? (Comando fundamental)

```
pwd
```

pwd = Print Working Directory

Ele mostra em qual pasta você está — essencial ao trabalhar com vários arquivos FASTQ/FASTA.

6. Listando arquivos e pastas

```
ls
```

Variantes úteis:

Comando	Função
<code>ls -l</code>	Lista arquivos em formato detalhado
<code>ls -lh</code>	Tamanho dos arquivos legível (10K, 2M, 1G)
<code>ls -a</code>	Mostra arquivos ocultos

👉 Em bioinformática, usamos muito `ls -lh` para verificar **tamanho de FASTQs**.

Exemplo: `ls -lh raw_data/`



7. Navegando entre diretórios

Entrar em uma pasta:

```
cd nome_da_pasta
```

Voltar um nível:

```
cd ..
```

Ir para a pasta home:

```
cd ~
```

8. Criando a estrutura inicial de um projeto

Podemos criar uma estrutura organizada para análises de RNA-Seq, WGS, etc.

```
mkdir projeto_bioinfo  
cd projeto_bioinfo  
mkdir raw_data scripts results qc logs
```

👉 Organização é **fundamental** para reprodutibilidade científica.

Exercício para fixação

1 Crie uma pasta chamada `Aulas_LinuxShellScript`:

2 Dentro dela, crie as pastas: `data`, `scripts`, `results`:

3 Descubra onde você está:

4 Liste os arquivos de forma detalhada: