

Aula prática: Linux.

Explorando dados de NGS.

06/07/2017

Professor: Jorge Estefano Santana de Souza, jorge@imd.ufrn.br;

Monitores: Danilo Lopes Martins, danilolmartins@gmail.com;
Luan Pereira, luanpereira00@outlook.com.

Objetivos:

Este tutorial tem como objetivo fazer com que o aluno conheça um pouco mais sobre o Linux, bem como habilitar o usuário para trabalhar com as principais ferramentas de bioinformática.

Ferramentas:

- 1- SSH.
- 2- shell.
- 3- Comandos básicos do Linux.

Comandos Básicos:

Durante a execução do tutorial iremos abordar os comandos essenciais, no entanto é fortemente recomendável que os alunos expandam os seus conhecimentos aprendendo outros comandos básicos do Linux. Podem procurar mais informação no site:

<http://wiki.ubuntu-br.org/ComandosBasicos>

Login Servidor:

Inicialmente vamos fazer o login no servidor, abra um terminal no linux e digite:

```
ssh -p 4422 bif@10.7.5.38
```

Irá pedir uma senha, digite:

```
bif0003
```

*ps. não aparece a digitação, o teclado não quebrou não!

Regras para login no servidor:

Interno à UFRN:

```
ssh -p 4422 bif@10.7.5.38
```

Senha: bif0003

Externo à UFRN:

```
ssh -p 4422 bif@177.20.147.141
```

Senha: bif0003

Dados brutos (raw data):

Durante a execução dos tutoriais necessitaremos de alguns dados iniciais, em via de regra estarão disponíveis no diretório:

```
/home/treinamento/
```

No terminal LINUX (ou Mac):

Quando estiver conectado por ssh e quiser saber quem mais está trabalhando na máquina, use o comando **who**.

Como nos conectamos a várias máquinas com vários logins, as vezes precisamos digitar **whoami** para lembrar como nos conectamos.

Mais dois comandos precisam ser usados sempre:

ls (dá um ls) - lista o conteúdo da pasta ou diretório;

pwd - diz o caminho do presente diretório (path of working directory);

1) Vamos começar pelo básico, certifique-se de que a pasta atual é:

```
/home/bif
```

Para isso digite o comando:

```
pwd
```

*ps. vc vai ver onde está

2) Nesse momento pode não haver o que listar, mas sempre é importante ver o que tem, ou não tem, na pasta:

```
ls
```

3) Crie um diretório nomeando-o com o seu nome:

```
mkdir SeuNome
```

4) dá um ls:

```
ls
```

5) Entre no diretório com cd SeuNome:

```
cd SeuNome
```

dando um pwd vc deve ver /home/bif/SeuNome:

```
pwd
```

6) suba de volta um diretório com cd .. - esses dois pontos significam diretório acima, então esse comando manda mudar para o diretório acima (change directory to upper directory):

```
cd ..
```

Outros comandos:

7) Ver o conteúdo do diretório: /home/treinamento:

```
ls /home/treinamento/
```

8) Copiar o arquivo lyrics para o diretório presente, perceba o ponto que representa o presente diretório:

```
cp /home/treinamento/lyrics .
```

9) Criar o diretório teste:

```
mkdir teste
```

10) Mover o arquivo lyrics para o diretório teste:

```
mv lyrics /teste
```

11) Entrar no diretório teste:

```
cd teste
```

12) Trocar o nome do arquivo:

```
mv lyrics letra
```

13) Copiar o arquivo:

```
cp letra lyrics
```

14) Remove o arquivo letra:

```
rm letra
```

15) Listar com mais informações:

```
ls -l
```

16) Mostrar o manual para o programa ls:

```
man ls
```

q (interrompe o output do manual)

17) Imprimir na tela o conteúdo do arquivo:

```
more lyrics
```

q (interrompe o comando more)

idem, mas tem que dar "q" para sair/quit:

```
less lyrics
```

q (interrompe o output do manual, do comando more e less do Linux)

Com o comando **more** move-se com barra de espaço ou "B", com o comando **less** com as setinhas do teclado.

18) Imprimir as primeiras linhas do arquivo:

```
head /home/treinamento/ERR844339.fastq
```

19) Imprimir as últimas linhas do arquivo:

```
tail /home/treinamento/ERR844339.fastq
```

20) **tabulador(tab) e asterisco**: são usados para nomes compridos. O tabulador completa o nome, o asterisco funciona como coringa, por exemplo **more ly*** imprimirá o conteúdo de lyrics:

```
more ly*
```

21) Listar o conteúdo do diretório /home/treinamento/blast_aula/FASTAS:

```
ls /home/treinamento/blast_aula/FASTAS/
```

22) Imprimir na tela o conteúdo do arquivo GAPDH:

```
more /home/treinamento/blast_aula/FASTAS/GAPDH
```

23) Copiar o arquivo GAPDH para o diretório atual:

```
cp /home/treinamento/blast_aula/FASTAS/GAPDH .
```

24) dá um ls:

```
ls
```

Editor de texto:

Nesse momento vc já está se virando bem, falta aprender a editar arquivos no servidor

25) Abrir o arquivo GAPDH no editor de texto vi:

```
vi GAPDH
```

- 1 - Para entrar no INSERT MODE e poder editar o texto digite " i "
- 2 - Troque o nome da sequência para >hsa
- 3 - Para salvar tecle **ESC** depois : (dois pontos) depois **x!** [enter]

26) Vamos ver o resultado:

```
more GAPDH
```

27) Vamos renomear o arquivo: :

```
mv GAPDH gapdh.hsa
```

28) Agora pegue uma sequência de mioglobina no NCBI:

1 - Entrar no navegador

2 - Entrar no site:

<https://www.ncbi.nlm.nih.gov/nuccore/1049011000?report=fasta>

3 - Copiar

29) Digite no terminal:

```
vi mioglobina
```

Cole o conteúdo do fasta. Botão da direita no terminal cola!!!!!!!

Saia como antes **ESC** depois **:** depois **x!** **[enter]**

30) Troque o nome da sequência para:

```
>myoglobin
```

*Ps. Não digite direto no terminal, antes digite "vi mioglobina"

31) Vamos ver o resultado:

```
more mioglobina
```

O comando grep:

O comando `grep` é bastante utilizado para realizar buscas em textos/arquivos, a ideia é procurar texto em uma string ou dentro de arquivos e mostrar as linhas de ocorrências, vamos testar:

32) Primeiro vamos ver o conteúdo do diretório `/home/treinamento/blast_aula/CDS`:

```
ls /home/treinamento/blast_aula/CDS
```

33) Agora vamos ver o conteúdo do arquivo `h.sapiens.nuc`:

```
less /home/treinamento/blast_aula/CDS/h.sapiens.nuc
```

q (interrompe o comando)

34) Vamos identificar as ocorrências da palavra aldolase:

```
cat /home/treinamento/blast_aula/CDS/h.sapiens.nuc | grep aldolase
```

35) Vamos quantificar o numero de sequencias do arquivo fasta:

```
cat /home/treinamento/blast_aula/CDS/h.sapiens.nuc | grep ">" -c
```

Referências:

- 1- ubuntu-br: <http://wiki.ubuntu-br.org/ComandosBasicos;>
- 2- NCBI: <https://www.ncbi.nlm.nih.gov/>
- 3- Linuxbr: <http://br-linux.org/>