

## Carregando Bioinformáticas do Pipelines na Linha de Comando

Antes de usar a linha de comando do Linux para executar processos de bioinformática, é necessário configurar o computador para poder executar esses comandos. As informações abaixo referem-se especificamente a um processo de metagenômica bacteriana que pode ser utilizado para a montagem de genomas de *Vibrio cholerae*, mas muitos dos mesmos princípios podem ser aplicados a outros processos armazenados no github.

Notas importantes para acompanhar este tutorial:

- O texto com o fundo cinzento e o tipo de letra `monoespaçado` representam comandos a introduzir. Geralmente, os comandos têm uma linha, no entanto, neste documento, os comandos podem passar para a linha seguinte visualmente. Adicionaremos uma linha em branco entre os comandos para indicar a presença de vários comandos.
- O texto azul `monoespaçado com um fundo branco` representa a entrada e/ou saída do terminal. Devido a diferentes programas de terminal, a aparência do seu terminal pode ser diferente.
- O texto negrito rodeado por < > é algo que terá de substituir pelo seu próprio nome de ficheiro, percurso ou nome de amostra.

Este tutorial irá guiá-lo através dos dois principais passos para carregar um pipeline bioinformático no seu computador:

1. Baixando da Internet um pipeline de bioinformática desejado.
2. Criar um ambiente de computação (ou seja, baixar software) necessário para executar o pipeline.

## PASSO 0: Navegue para o seu diretório PESSOAL

A menos que seja indicado o contrário, todos os comandos devem ser executados no seu diretório pessoal. Geralmente, é aqui que a linha de comandos está localizada se abrir uma nova linha de comandos (ou seja, uma janela Terminal).

1. Se não estiver no seu diretório pessoal (ou não tiver a certeza), vá para o seu diretório pessoal com este comando:

```
cd ~/
```

## PASSO 1: Descarregar o Pipeline Bioinformático

Os bioinformáticos disponibilizam frequentemente os seus pipelines de montagem e análise de genomas em websites como [www.github.com](https://www.github.com) (normalmente referido como "GitHub"), que está especificamente configurado para armazenar código e outros tipos de ficheiros. Existem duas formas de descarregar ficheiros e pipelines do GitHub: através do seu navegador Website ou através da linha de comandos. Para realizar a montagem de genómica bacteriana utilizando os pipelines CholGen, utilize um dos métodos abaixo:

### Opção 1: Baixar a partir do GitHub usando a linha de comando

1. Instale o programa de linha de comando git usando o seguinte comando:

```
sudo apt install git-all
```

**Nota:** este comando solicitará a sua senha. Ao digitar sua senha na linha de comando, nenhum caractere aparecerá. No entanto, a linha de comando continuará a receber entradas, por isso certifique-se de que escreve a sua palavra-passe corretamente antes de clicar em Enter.

2. Depois do git estar instalado, execute o seguinte comando no seu diretório inicial para descarregar a pasta do GitHub:

```
git clone https://github.com/CholGen/bacpage.git
```

3. Verifique se o ficheiro foi descarregado corretamente, visualizando o conteúdo do seu diretório inicial:

```
ls
```

Deverá ver um novo diretório no seu diretório inicial com o nome `bacpage/`.

Este procedimento pode ser utilizado para descarregar qualquer outro pipeline armazenado no website do GitHub. Se estiver interessado em carregar um pipeline diferente do que foi usado aqui, basta substituir a URL no Passo 2 pelo link encontrado na página do GitHub do outro pipeline.

## Configuração do processo bioinformático do pipeline

4. Confirme que a pasta de montagem e análise do genoma (bacpage/) contém 5 subpastas: config/, exemplo/, recursos/, teste/ e fluxo de trabalho/.

### Opção 2: Descarregar a partir do GitHub utilizando um navegador Website

1. Colar o seguinte link no seu navegador da Website. Um arquivo chamado bacpage.zip será baixado automaticamente.

Link: <https://github.com/CholGen/bacpage/releases/latest/download/pipeline.zip>

2. Descompacte este ficheiro no seu computador e transfira a pasta resultante (denominada bacpage/) para o diretório PESSOAL do seu computador. Para tal, basta arrastar a pasta bacpage/ da pasta de descarregamentos para o diretório PESSOAL.
3. Confirme que os ficheiros de montagem e análise do genoma (bacpage/) contém 5 subpastas: config/, exemplo/, recursos/, teste/ e fluxo de trabalho/.

## **PASSO 2: Criar o ambiente de computação**

Ao contrário dos programas simples de linha de comandos, como o cd e o ls, a análise bioinformática requer muitos programas especializados que precisam de ser instalados a partir de uma vasta gama de programadores, com uma vasta gama de requisitos e dependências. Para facilitar a repetição dos seus processos exactos por outros, os bioinformáticos podem criar "ambientes" de computação que contenham todo o software e outros itens de configuração necessárias para seguir uma determinada análise. Podem depois partilhar as instruções para replicar o seu "ambiente" de computação num site como o GitHub, para que outras pessoas possam rapidamente configurar os seus computadores exatamente da mesma forma.

Os ficheiros que descarregou do GitHub acima contêm instruções sobre como criar o ambiente necessário para executar todos os ficheiros que também descarregou. Estas instruções irão instalar automaticamente todo o software necessário para executar os guiões de bioinformática. Para seguir estas instruções (e carregar o ambiente de computação de outra pessoa no seu computador), terá de instalar primeiro um software que possa ler estas instruções.

Existem alguns softwares diferentes que podem ser usados para criar ambientes. O "Mamba" e o "Conda" são opções muito comuns. Iremos utilizar o "Mamba" porque tende a funcionar de forma mais rápida que o "Conda", mas se alguma vez estiver a trabalhar com outro conjunto de instruções que utilize o "Conda", saiba que está a fazer algo muito semelhante!

Siga as instruções abaixo para instalar o Mamba e depois utilize o software Mamba para criar o ambiente de montagem bacteriana no seu computador. Tenha em atenção que vai precisar de uma ligação de internet para completar estes passos, uma vez que vai descarregar o Mamba da Internet e depois utilizar o software Mamba para descarregar outros programas e software da Web.

## Configuração do processo bioinformático do pipeline

1. Navegue até ao seu diretório pessoal com o seguinte comando.

```
cd ~/
```

2. Instale o mamba usando o seguinte comando:

```
curl -L -O
"https://github.com/conda-forge/miniforge/releases/latest/download/Mambaforge-$(uname) -$(uname -m).sh"

bash Mambaforge-$(uname) -$(uname -m).sh
```

Deve ter reparado que a sua linha de comandos mudou de:

```
[seqlaptop@linuxbox seqlaptop]$
```

to:

```
(base) [seqlaptop@linuxbox seqlaptop]$
```

Isto indica que está agora no ambiente base mamba, e que todo o software que carregou nesse ambiente (chamado "base") está acessível.

**Nota:** para sair do ambiente mamba base (por exemplo, se não conseguir aceder a software previamente instalado, execute `mamba deactivate`). Quando sair do ambiente mamba, já não poderá aceder ao software que instalou quando estava dentro do ambiente mamba. Para voltar ao ambiente base do mamba, execute `mamba activate`).

Em vez de instalar tudo no ambiente base, é recomendado criar ambientes específicos de tarefas, contendo apenas o software necessário para uma determinada tarefa. Vamos agora criar um novo ambiente com um nome diferente que terá especificamente o software do pipeline de montagem nele.

3. Navegue até ao diretório onde descarregou o pipeline Bioinformática acima. Esta deve ser uma diretoria intitulada bacpage/ no seu diretório pessoal.

```
cd ~/bacpage
```

Terá de navegar para este diretório sempre que quiser utilizar o pipeline que descarregou do GitHub acima. Execute `pwd` e anote o percurso deste diretório para referência futura.

4. Dentro deste diretório, você encontrará um arquivo chamado `environment.yaml`. Sinta-se à vontade para abri-lo em um editor de texto e observar sua estrutura. O arquivo descreve todo o software necessário para o pipeline. Para instruir o computador a instalar todo esse software, execute o seguinte comando:

## Configuração do processo bioinformático do pipeline

```
mamba env create -f environment.yaml
```

**Nota:** esta instalação pode demorar alguns minutos, dependendo da velocidade da sua Internet.

Este comando irá criar um novo ambiente chamado "bacpage", que é distinto do ambiente "base" por defeito discutido acima. Ele pode ser ativado e desativado de maneira idêntica ao ambiente "base".

5. Para utilizar um ambiente mamba, e todo o software que acabámos de instalar, temos de o ativar. Para o fazer, utilizamos o seguinte comando:

```
mamba activate bacpage
```

O seu comando deve ter mudado novamente para o seguinte:

```
(bacpage) [seq1laptop@linuxbox bacpage]$
```

Isto indica que o seu terminal está agora a utilizar o ambiente bacpage e pode utilizar todo o software necessário. Utilize `mamba deactivate` para retornar ao ambiente padrão "base".

6. Explicaremos este comando com mais detalhe mais tarde, mas execute o seguinte para testar se a instalação foi bem sucedida:

```
snakemake --configfile test/test.yaml --all-temp --cores 8
```

Este comando deve levar alguns minutos para ser executado e depois ser concluído sem problemas.