



Chargement de pipelines bioinformatiques dans la ligne de commande

Avant d'utiliser la ligne de commande de Linux pour réaliser des pipelines bioinformatiques, vous devez configurer votre ordinateur pour qu'il puisse exécuter ces commandes. Les informations ci-dessous renvoient spécifiquement à un pipeline de métagénomique bactérienne qui peut être utilisé pour l'assemblage des génomes de *Vibrio cholerae*, mais de nombreux principes peuvent être appliqués à d'autres pipelines hébergés sur github.

Notes importantes pour suivre ce tutoriel :

- Le texte sur fond gris en `police monospace` représente les commandes à saisir. En général, les commandes comportent une ligne, mais dans ce document, les commandes peuvent visuellement s'étendre jusqu'à la ligne suivante. Nous ajouterons une ligne vide entre les commandes pour indiquer la présence de plusieurs commandes.
- Un `texte bleu monospace sur fond blanc` représente l'entrée et/ou la sortie du terminal. En raison des différentes applications de terminal, l'apparence de votre terminal peut paraître différente.
- Le texte en gras entouré de `< >` est quelque chose que vous devrez remplacer par votre propre nom de dossier, de chemin d'accès ou d'échantillon.

Ce tutoriel vous guidera à travers les deux étapes principales du chargement d'un pipeline bioinformatique sur votre ordinateur :

1. Téléchargement d'un pipeline bioinformatique sur Internet.
2. Création d'un environnement informatique (c'est-à-dire téléchargement de logiciels) nécessaire à l'exécution du pipeline.



ÉTAPE 0 : Naviguez jusqu'à votre répertoire ACCUEIL (HOME)

Sauf indication contraire, toutes les commandes doivent être exécutées dans votre répertoire personnel. En général, c'est là que se trouve la ligne de commande lorsque vous ouvrez une nouvelle invite de commande (c'est-à-dire une fenêtre de Terminal).

1. Si vous n'êtes pas dans votre répertoire personnel (ou si vous n'êtes pas sûr), déplacez-vous dans votre répertoire personnel à l'aide de cette commande :

```
cd ~/
```

ÉTAPE 1 : Télécharger le pipeline bioinformatique

Les bioinformaticiens mettent souvent leurs pipelines d'assemblage et d'analyse des génomes à la disposition du public sur des sites web tels que www.github.com (communément appelé "GitHub"), qui est spécialement configuré pour héberger le code et d'autres types de fichiers. Il existe deux façons de télécharger des fichiers et des pipelines à partir de GitHub : via votre navigateur web ou via la ligne de commande. Pour réaliser un assemblage de génomique bactérienne à l'aide des pipelines CholGen, utilisez l'une des méthodes ci-dessous :

Option 1 : Télécharger depuis GitHub à l'aide de la ligne de commande:

1. Installez le programme de ligne de commande git à l'aide de la commande suivante :

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

Note : Cette commande vous demandera votre mot de passe. Lorsque vous saisissez votre mot de passe sur la ligne de commande, aucun caractère n'apparaît. Cependant, la ligne de commande continuera à recevoir des données, alors assurez-vous de taper votre mot de passe correctement avant d'appuyer sur la touche Entrée.

2. Une fois git installé, exécutez la commande suivante dans votre répertoire HOME pour télécharger le dossier depuis GitHub :

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

3. Vérifiez que le téléchargement s'est déroulé correctement en visualisant le contenu de votre répertoire personnel :

```
ls
```

Un nouveau répertoire devrait apparaître dans votre répertoire personnel sous le nom de `bacpage/`.



Cette procédure peut être utilisée pour télécharger n'importe quel autre pipeline hébergé sur le site GitHub. Si vous souhaitez charger un pipeline différent de celui utilisé ici, remplacez simplement l'URL de l'étape 2 par le lien trouvé sur la page GitHub de l'autre pipeline.

4. Confirmez que le dossier d'assemblage et d'analyse du génome (`bacpage/`) contient 5 sous-dossiers : `config/`, `example/`, `resources/`, `test/`, et `workflow/`.

Option 2 : Télécharger depuis GitHub à l'aide d'un navigateur web :

1. Collez le lien suivant dans votre navigateur web. Un fichier nommé `bacpage.zip` sera automatiquement téléchargé.

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

2. Décompressez ce fichier sur votre ordinateur et déplacez le dossier résultant (nommé `bacpage/`) dans le répertoire HOME de votre ordinateur. Pour ce faire, faites glisser le dossier `bacpage/` de votre dossier Téléchargements vers votre répertoire HOME.
3. Confirmer que le dossier d'assemblage et d'analyse du génome (`bacpage/`) contient 5 sous-dossiers : `config/`, `example/`, `resources/`, `test/`, et `workflow/`.

ÉTAPE 2 : Construire l'environnement informatique

Contrairement aux programmes de ligne de commande simples tels que `cd` et `ls`, l'analyse bioinformatique nécessite de nombreux programmes spécialisés qui doivent être installés par un large éventail de développeurs, avec un large éventail d'exigences et de dépendances. Pour permettre à d'autres de reproduire facilement leurs processus exacts, les bioinformaticiens peuvent créer des "environnements" informatiques contenant tous les logiciels et autres éléments d'installation nécessaires pour suivre une analyse particulière. Ils peuvent ensuite partager les instructions pour reproduire leur "environnement" informatique sur un site tel que GitHub, afin que d'autres personnes puissent rapidement configurer leurs ordinateurs de la même manière.

Les fichiers que vous avez téléchargés sur GitHub ci-dessus contiennent des instructions sur la façon de créer l'environnement nécessaire à l'exécution de tous les scripts que vous avez également téléchargés. Ces instructions installeront automatiquement tous les logiciels dont vous avez besoin pour exécuter les scripts bioinformatiques. Pour suivre ces instructions (et charger l'environnement informatique de quelqu'un d'autre sur votre ordinateur), vous devez d'abord installer un logiciel capable de lire ces instructions.

Il existe plusieurs logiciels différents qui peuvent être utilisés pour créer des environnements. « Mamba » et « Conda » sont des options très courantes. Nous utiliserons « Mamba » parce qu'il a tendance à être plus rapide que « Conda », mais si vous travaillez avec un autre jeu d'instructions qui utilise « Conda » à la place, sachez que vous faites quelque chose de très similaire !

Suivez les instructions ci-dessous pour installer Mamba et utiliser le logiciel Mamba pour créer l'environnement d'assemblage bactérien sur votre ordinateur. Veuillez noter que vous aurez besoin d'une connexion réseau pour



mener à bien ces étapes, car vous téléchargerez Mamba à partir d'Internet et utiliserez ensuite le logiciel Mamba pour télécharger d'autres programmes et logiciels à partir du Web.

1. Naviguez vers votre répertoire personnel à l'aide de la commande suivante.

```
cd ~/
```

2. Installez mamba en utilisant la commande suivante :

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

Vous avez dû remarquer que l'invite de votre ligne de commande a changé :

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

à :

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

Cela indique que vous êtes maintenant dans l'environnement mamba de base, et que tous les logiciels que vous avez chargés dans cet environnement (appelé « base ») sont accessibles.

Note : pour sortir de l'environnement mamba de base (par exemple, si vous ne pouvez pas accéder à un logiciel précédemment installé, exécutez `mamba deactivate`). Lorsque vous quittez l'environnement mamba, vous ne pourrez plus accéder aux logiciels que vous avez installés lorsque vous étiez dans l'environnement mamba. Pour revenir dans l'environnement mamba de base, exécutez `mamba activate`).

Plutôt que de tout installer dans l'environnement de base, il est recommandé de créer des environnements spécifiques à une tâche, contenant uniquement les logiciels nécessaires à une tâche donnée. Nous allons maintenant créer un nouvel environnement avec un nom différent qui contiendra spécifiquement le logiciel du pipeline d'assemblage.

3. Naviguez jusqu'au répertoire où vous avez téléchargé le pipeline bioinformatique ci-dessus. Il doit s'agir d'un répertoire intitulé `bacpage/` dans votre répertoire personnel.

```
cd ~/bacpage
```

Vous devrez naviguer jusqu'à ce répertoire à chaque fois que vous voudrez utiliser le pipeline que vous avez téléchargé sur GitHub ci-dessus. Exécutez `pwd` et notez le chemin de ce répertoire pour toute référence future.



4. Dans ce répertoire, vous trouverez un fichier appelé `environment.yaml`. N'hésitez pas à l'ouvrir dans un éditeur de texte et à examiner sa structure. Ce fichier décrit tous les logiciels requis pour le pipeline. Pour demander à l'ordinateur d'installer tous ces logiciels, exécutez la commande suivante :

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

Note : cette installation peut prendre quelques minutes en fonction de la vitesse de votre internet.

Cette commande crée un nouvel environnement appelé « `bacpage` », qui est distinct de l'environnement « de base » par défaut mentionné ci-dessus. Il peut être activé et désactivé de la même manière que l'environnement « de base ».

5. Pour utiliser un environnement mamba, et tous les logiciels que nous venons d'installer, nous devons l'activer. Faites-le avec la commande suivante :

```
mamba activate bacpage
```

L'invite devrait avoir changé à nouveau pour se transformer en ce qui suit :

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

Ceci indique que votre terminal utilise maintenant l'environnement `bacpage` et peut utiliser tous les logiciels requis. Utilisez `mamba deactivate` pour revenir à l'environnement « base » par défaut.

6. Nous expliquerons cette commande plus en détail ultérieurement, mais exécutez ce qui suit pour vérifier si l'installation a réussi :

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

L'exécution de cette commande devrait prendre quelques minutes et se terminer sans problème.