

# 解析環境の構築

国立遺伝学研究所 大量遺伝情報研究室 坂本美佳

# 初心者の壁

インストール

# 初心者の壁

## 解析ツールのインストール

- ・バイナリファイルをどこかに置く（簡単）
- ・ソースコードをコンパイルしてインストール（難易度 高）



# 初心者の壁

## 解析ツールのインストール

- PATHを通す
  1. プログラムをフルパスで指定  
複数プログラムのラッパーの場合うまくうごくかわからん
  2. 每回 export PATH=~ をやる  
忘れるとプログラムを探せなくて動かん
  3. bashrcやbash\_profileに書いておく  
そもそもbashrcってなに？？？という場合どうする

# パッケージ管理ツールやコンテナの利用

- ・ パッケージ管理ツール
- ・ コンテナ(コンテナ型仮想化)



# パッケージ管理ツールやコンテナの利用

- ・ パッケージ管理ツール  
**Anaconda (Miniconda, Miniforge),  
Homebrew (Linuxbrew)**  
インストール、バージョン管理、  
依存関係（ツール相互の利用）を解決
- ・ コンテナ  
**Docker, Singularity (Apptainer)**  
まるごとそのツールを実行するための  
「環境(OS含む）」を提供  
コンピュータの中でもうひとつ別のコン  
ピュータを動かすイメージ  
昨年Singularityプロジェクトの名称がApptainerに変わった

# パッケージ管理ツールやコンテナの利用

- ・ パッケージ管理ツール

**Miniconda (Miniforge)**

Anacondaの簡易版

- ・ コンテナ

**Singularity (Apptainer)**

スパコン版のコンテナ

# パッケージ管理ツールやコンテナの利用

- 利点
  - ✓ makeなどのインストールコマンドを叩かなくてもいい
  - ✓ (パッケージ管理ツール) 依存パッケージも一緒にに入る
  - ✓ (コンテナ) 必要なものがすべて入っている
  - ✓ バージョン管理が楽
  - ✓ 環境の移動や共有が楽

# パッケージ管理ツールやコンテナの利用

- ・ 欠点
  - ✓ 動かないパッケージ...
  - ✓ (パッケージ管理ツール) いろいろ入れているうちに管理ツール 자체が壊れてしまうことも
  - ✓ (コンテナ) ブラックボックスになりがち

とはいえ、利点がより多いので積極的に使っていきましょう

# RNA-seq解析環境

## このあと利用するツール

- HISAT2 リードをリファレンス配列にマッピング
- StringTie 転写産物isoformの同定
- Samtools sam/bamファイルの操作

1. Minicondaの仮想環境にインストールする方法

2. Singularityコンテナを利用する方法

次はMinicondaの利用

# Minicondaの利用

## 概要

- Minicondaのインストールと初期設定
- 仮想環境の作成 <- condaインストール済みのひとはここから
- 仮想環境に解析ツールをインストール

コマンドは手打ちorハンドアウトからコピペでOK

次はMinicondaのインストール

# Minicondaの利用

## Minicondaのインストール

- AnacondaとMiniconda

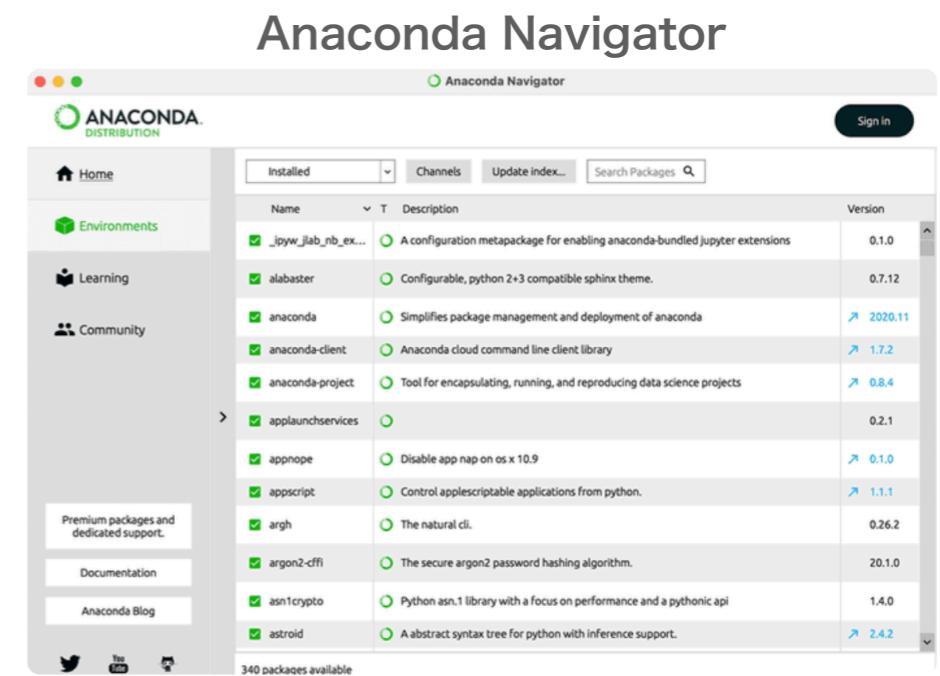
✓ Pythonのディストリビューション

✓ パッケージ管理ツールcondaを含む

✓ AnacondaはGUIツールを含む多くのパッケージ（大きい）

✓ Minicondaは最低限必要なパッケージのみ （小さい）

スパコンで使うのでGUI不要、必要なものを追加インストールする



参考

<https://tooljp.com/windows/chigai/html/Python/conda-anaconda-miniconda-pyenv-pip-chigai.html>

# Minicondaの利用

## Minicondaのインストール

コマンド

出力を確認

入力が必要なところは**文字色を変えて**示します

# Minicondaの利用

## Minicondaのインストール

<https://docs.conda.io/en/latest/miniconda.html>

The screenshot shows the official Miniconda documentation page. The left sidebar has a green header 'conda latest' and sections for Conda, Conda-build, and Miniconda (which is expanded). The main content area has a title 'Miniconda' and a brief description. Below it is a section titled 'System requirements' with a bulleted list of requirements. A large red box highlights the 'Miniconda3 Linux 64-bit' link in the 'Latest Miniconda Installer Links' table. The table lists platforms, names, and SHA256 hashes for various Miniconda versions.

| Latest - Conda 4.12.0 Python 3.9.7 released February 15, 2022 |  |  |
|---|--|--|
| Platform  | Name                                   | SHA256 hash  |
| Windows   | Miniconda3 Windows 64-bit              | 1bc2e8277ddd54a5f724896c7edee112d068529588d944702966c867e7e9cc   |
|   | Miniconda3 Windows 32-bit              | 4fb64e6c9c28b88beab16994bfba4829110ea3145baa60bda5344174ab65d462 |
| macOS   | Miniconda3 macOS Intel x86 64-bit bash | 007baef18dc7b6f2ca6209b5a0c9bd2f283154152f82bef787aac709a51633   |
|   | Miniconda3 macOS Intel x86 64-bit pkg  | cb56184637711685b08f6eba9532cef6985ed7007b38e789613d5dd3f94ccc6b |
|   | Miniconda3 macOS Apple M1 64-bit bash  | 4bd112168cc33f8a4a60d3ef7e72b52a85972d588cd065be803eb21d73b625ef |
| Linux   | Miniconda3 macOS Apple M1 64-bit pkg   | 0cb5165ca751e827d91a4ae6823bdfa24d22c398a0b3b01213e57377a2c54226 |
|   | Miniconda3 Linux 64-bit                | 78f39f9bae971ec1ae7969f0516017f2413f17796670f7040725d83cff5689   |
|   | Miniconda3 Linux-aarch64 64-bit        | 5f4f865812101fdc747cea5b820806f678bb50fe0a61f19dc8aa369c52c4e513 |
|   | Miniconda3 Linux-ppc64le 64-bit        | 1fe3305d0ccc9e55b336b051ae12d82f33af408af4b560625674fa7ad915102b |
|   | Miniconda3 Linux-s390x 64-bit          | ff6fdad3068ab5b15939c6f422ac329fa005d56ee0876c985e22e622d930e424 |

Miniconda3 Linux 64-bit  
リンクをコピー

# Minicondaの利用

## Minicondaのインストール

スパコンにログイン+ `qlogin`

ホームディレクトリにツール用のディレクトリを作成

```
mkdir tools
```

ツール用ディレクトリに移動

```
cd tools
```

`wget` コマンドでインストールスクリプト入手

1行で (改行しないで) 入力してください

```
wget https://repo.anaconda.com/miniconda/Miniconda3-latest-Linux-x86_64.sh
```

コピーが正しく行われたかどうかを確認

```
sha256sum Miniconda3-latest-Linux-x86_64.sh
```

78f39f9bae971ec1ae7969f0516017f2413f17796670f7040725dd83fcff5689

# Minicondaの利用

## Minicondaのインストール FileZilla利用

<https://docs.conda.io/en/latest/miniconda.html>

The screenshot shows the official Miniconda documentation page. The left sidebar has a navigation menu with sections like Conda, Conda-build, and Miniconda. The main content area is titled "Miniconda" and describes it as a free minimal installer for conda. It includes a "System requirements" section with a bulleted list and a "Latest Miniconda Installer Links" table.

| Platform | Name                                   | SHA256 hash  |
|----------|--|--|
| Windows  | Miniconda3 Windows 64-bit              | 1bc2e8277ddd54a5f724896c7edee112d068529588d944702966c867e7e9cc   |
|          | Miniconda3 Windows 32-bit              | 4fb64e6c9c28b88beab16994bfa4829110ea3145baa60bda5344174ab65d462  |
| macOS    | Miniconda3 macOS Intel x86 64-bit bash | 007baef18dc7b6f2ca6209b5a0c9bd2f283154152f82bef787aac709a51633   |
|          | Miniconda3 macOS Intel x86 64-bit pkg  | cb56184637711685b08f6eba9532cef6985ed7007b38e789613d5dd3f94ccc6b |
|          | Miniconda3 macOS Apple M1 64-bit bash  | 4bd112168cc33f8a4a60d3ef7e72b52a85972d588cd065be803eb21d73b625ef |
| Linux    | Miniconda3 Linux 64-bit                | 0cb5165ca751e827d91a4ae6823bdfa24d22c398a0b3b01213e57377a2c54226 |
|          | Miniconda3 Linux-aarch64 64-bit        | 78f39f9bae971ec1ae7969f0516017f2413f17796670f7040725d83cff5689   |
|          | Miniconda3 Linux-ppc64le 64-bit        | 1fe3305d0ccc9e55b336b051ae12d82f33af408af4b560625674fa7ad915102b |
|          | Miniconda3 Linux-s390x 64-bit          | ff6fdad3068ab5b15939c6f422ac329fa005d56ee0876c985e22e622d930e424 |

Miniconda3 Linux 64-bit の  
リンクから自分のPCにダウンロード

# Minicondaの利用

## Minicondaのインストール **FileZilla**利用

スパコンにログイン+ qlogin、ホームディレクトリにツール用のディレクトリを作成

```
mkdir tools
```

FileZillaを用いてインストールスクリプト **Miniconda3-latest-Linux-x86\_64.sh** を**tools**ディレクトリにコピー

ツール用ディレクトリに移動

```
cd tools
```

コピーが正しく行われたかどうかを確認

```
sha256sum Miniconda3-latest-Linux-x86_64.sh
```

78f39f9bae971ec1ae7969f0516017f2413f17796670f7040725dd83fcff5689

# Minicondaの利用

## Minicondaのインストール

インストールスクリプトの実行

```
bash Miniconda3-latest-Linux-x86_64.sh
```

スクロールして次の文章が表示されたら **yes** と入力

```
Do you accept the license terms? [yes|no]  
[no] >>> yes
```

インストールディレクトリの変更 何もしないと/home/アカウント名/miniconda3にインストールされます

```
Miniconda3 will now be installed into this location:  
/home/アカウント名/miniconda3
```

- Press ENTER to confirm the location
- Press CTRL-C to abort the installation
- Or specify a different location below

```
[/home/アカウント名/miniconda3] >>> /home/アカウント名/tools/miniconda3
```

# Minicondaの利用

## Minicondaのインストール

必要最低限パッケージのインストールが終わると `conda init` を実行するか聞かれる  
ので `yes` と入力

### WARNING:

You currently have a PYTHONPATH environment variable set. This may cause unexpected behavior when running the Python interpreter in Miniconda3.  
For best results, please verify that your PYTHONPATH only points to directories of packages that are compatible with the Python interpreter in Miniconda3: /home/アカウント名/tools/miniconda3

Do you wish the installer to initialize Miniconda3  
by running `conda init`? [yes|no]  
[no] >>> yes

~/.bashrcに設定が書き込まれます

いったん`exit`し、ふたたび`qlogin`

プロンプトに(`base`)と表示されていることを確認

(`base`) [アカウント名@ノード名 ~]\$

### 参考URL

<https://conda.io/projects/conda/en/latest/user-guide/install/index.html>

<https://conda.io/projects/conda/en/latest/user-guide/install/linux.html#install-linux-silent>

次は初期設定

# Minicondaの利用

## 初期設定

Anaconda/Minicondaが既にインストールされていた場合、ここから始めましょう

パッケージリストの表示

```
conda list
```

conda-forge チャンネルの追加 パッケージの取得元をconda-forge優先にします

```
conda config --add channels conda-forge
```

チャネルの優先度の確認

```
conda config --get channels
```

以下のように出力されるはず

```
--add channels 'defaults' # lowest priority
```

```
--add channels 'conda-forge' # highest priority
```

conda-forge優先になつていないと、  
このあとツールをインストールする時に不具合が出ます

# Minicondaの利用

## 初期設定

- Anaconda/Miniconda
  - ✓ 個人、公認教育機関の学生・教員、オープンソース開発者、中小企業などは無料（それ以外は有償）
- conda-forge
  - ✓ コミュニティベースで開発
  - ✓ オープンソース conda-forgeをdefault チャンネルにしたMiniforgeインストーラもあります

今回は上記理由のほか、Minicondaのdefaultチャンネルを最優先にすると、  
解析ツールのバージョン指定インストールがうまくいきませんでした...

次は仮想環境の作成

# Minicondaの利用

## 仮想環境の作成

- ・ ツール（ソフトウェア）間に依存関係があるため、
  - ✓ あるツールを動かすためにほかのツールが必要
  - ✓ あるツールをインストールすると他のツールが動かなくなることもある

これを防ぐために

- 仮想環境を作成し、その中にツールをインストール
- うまく動かなかったら仮想環境を破棄して作り直すことは簡単

# Minicondaの利用

## 仮想環境の作成

toolsディレクトリに移動

ここまで流れではhomeディレクトリにいるはず

```
cd tools
```

pags\_rnaseqという名の仮想環境を作成

```
conda create -n pags_rnaseq
```

Proceed ([y]/n)? と聞かれたら y と入力

```
## Package Plan ##
```

```
environment location: /home/xxxxx/tools/miniconda3/envs/pags_rnaseq
```

Proceed ([y]/n)? y

# Minicondaの利用

## 仮想環境の作成

完了すると以下のように表示される

```
#  
# To activate this environment, use  
#  
#     $ conda activate pags_rnaseq    仮想環境に入るコマンド  
#  
# To deactivate an active environment, use  
#  
#     $ conda deactivate    仮想環境から出るコマンド
```

# Minicondaの利用

## 仮想環境の作成

仮想環境の有効化

```
conda activate pags_rnaseq
```

プロンプトが以下のように表示されていることを確認

```
(pags_rnaseq) [アカウント名@ノード名 tools] $
```

↑ 仮想環境名

仮想環境の終了

```
conda deactivate
```

プロンプトが以下のように表示されていることを確認

```
(base) [アカウント名@ノード名 tools] $
```

↑ baseに戻る

次は解析ツールのインストール

# Minicondaの利用

## 解析ツールのインストール

- HISAT2 v2.2.1
- StringTie v2.2.1
- Samtools v1.16.1

# Minicondaの利用

## 解析ツールのインストール

- インストールコマンドの調べ方



# Minicondaの利用

## 解析ツールのインストール

The screenshot shows a web browser window displaying the Anaconda.org website. The URL in the address bar is `anaconda.org`. The main navigation menu includes links for `Gallery`, `About`, `Anaconda`, `Help`, `Download Anaconda`, and `Sign In`. The search bar contains the text `Samtools`.

The main content area shows a package page for `bioconda / packages / hisat2 2.2.1`. The package summary states: "Graph-based alignment of next generation sequencing reads to a population of genomes." Below this, there are tabs for `Conda`, `Files`, `Labels`, and `Badges`. The `Conda` tab is selected.

The `Conda` section displays the following metadata:

- License: [GPL-3.0](#)
- Home: <http://daehwankimlab.github.io/hisat2>
- </> Development: <https://github.com/DaehwanKimLab/hisat2>
- Documentation: <https://daehwankimlab.github.io/hisat2/manual/>
- 173791 total downloads
- Last upload: 7 months and 17 days ago

The `Installers` section contains a note: "Info: This package contains files in non-standard labels." It provides instructions for installing the package using Conda:

`conda install` ⓘ

[linux-64 v2.2.1](#) 最新バージョン  
[osx-64 v2.2.1](#)

To install this package run one of the following:

```
conda install -c bioconda hisat2
conda install -c "bioconda/label/cf201901" hisat2
```

A pink box highlights the command `conda install -c "bioconda/label/cf201901" hisat2`.

# Minicondaの利用

## 解析ツールのインストール

HISAT2のインストール

仮想環境の有効化

```
conda activate pags_rnaseq
```

パッケージ名=バージョンとするとバージョンを指定してインストールできる

```
conda install -c bioconda hisat2=2.2.1
```

インストールの可否について聞かれるので **y** と入力

```
Proceed ([y]/n)? y
```

ヘルプメッセージの表示（インストールの確認）

```
hisat2 -h
```

# Minicondaの利用

## 解析ツールのインストール

StringTieのインストール

-y を付けると、インストールの可否について聞かれずに進む

```
conda install -c bioconda stringtie=2.2.1 -y
```

ヘルプメッセージの表示（インストールの確認）

```
stringtie -h
```

Samtoolsのインストール

```
conda install -c bioconda samtools=1.16.1 -y
```

ヘルプメッセージの表示（インストールの確認）

```
samtools
```

# Minicondaの利用

## パッケージリストの作成

仮想環境 `pags_rnaseq` のパッケージリスト

```
conda env export > pags_rnaseq.yml
```

`pags_rnaseq.yml` を使うと同じ環境（依存パッケージやバージョン）が作成可能

今回は行いません

仮想環境から出る

```
conda deactivate
```

プロンプトが以下のように表示されていることを確認

```
(base) [アカウント名@ノード名 tools] $
```

次はSingularityコンテナの利用

# Singularityの利用

## 概要

- ・ コンテナの探し方
- ・ コンテナの動かし方

# Singularityの利用

## コンテナの探し方

コンテナの場所 `/usr/local/biotools/ツールの頭文字`

```
ls /usr/local/biotools/h/hisat2*
```

```
ls /usr/local/biotools/s/stringtie*
```

```
ls /usr/local/biotools/s/samtools*
```

- ・一番新しいバージョンを使う
- ・(ただし動かないものもあるので、そのときは別のバージョンを使う)

# Singularityの利用

## コンテナの動かし方

### 基本

singularity exec オプション コンテナのフルパス ツールのコマンド ツールのオプション

### 練習 (StringTieのヘルプメッセージ表示)

singularity exec /usr/local/biotools/s/stringtie:2.2.1--hecb563c\_2 stringtie -h

ハンドアウトから、または lsコマンドで表示されたパスをコピーしてください

実行すると以下のヘルプメッセージが出力される

StringTie v2.2.1 usage:

```
stringtie <in.bam ..> [-G <guide_gff>] [-l <prefix>] [-o <out.gtf>] [-p <cpus>]
[-v] [-a <min_anchor_len>] [-m <min_len>] [-j <min_anchor_cov>] [-f <min_iso>]
[-c <min_bundle_cov>] [-g <bdist>] [-u] [-L] [-e] [--viral] [-E <err_margin>]
[--ptf <f_tab>] [-x <seqid,...>] [-A <gene_abund.out>] [-h] {-B|-b <dir_path>}
[--mix] [--conservative] [--rf] [--fr]
```

Assemble RNA-Seq alignments into potential transcripts.

(以下略)

では実際に使ってみましょう

