

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
難易度高い



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. 毎回 `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

コンテナやパッケージ管理ツールの利用

- ・コンテナ(コンテナ型仮想化)
- ・パッケージ管理ツール



パッケージ管理ツールやコンテナの利用

- ・コンテナ

Docker, 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

- ・パッケージ管理ツール

Anaconda (Miniconda, Miniforge), Homebrew

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

コンテナやパッケージ管理ツールの利用

- ・コンテナ

Apptainer

スパコン版のコンテナ

- ・パッケージ管理ツール

Miniconda (Miniforge)

Anaconda (conda-forge) の簡易版

-forgeはオープンソース版

パッケージ管理ツールやコンテナの利用

- 利点

- ✓ インストール不要
- ✓ (コンテナ) 必要なものがすべて入っている
- ✓ (パッケージ管理ツール) 依存アプリも一緒にに入る
- ✓ 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...
- ✓ (コンテナ) ブラックボックスになりがち
- ✓ (パッケージ管理ツール) いろいろ入れていているうちに管理ツール 자체が壊れてしまいがち

とはいっても、利点が多いので
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. コンテナを利用する方法

2. conda仮想環境にインストールする方法

condaを使う方法は昨年の講義資料をご覧ください

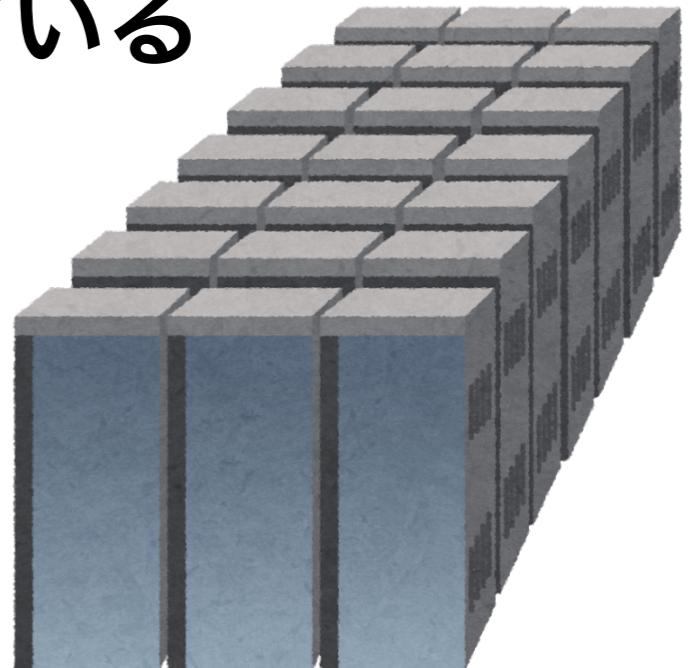
コンテナを利用する方法

Apptainer

Apptainerの利用

Apptainerとは

- 研究やHPC（スパコン）環境で安全・簡単にソフトウェア環境を再現できるコンテナ実行ツール
- コンテナは遺伝研スパコンに用意されている
- インストール、ダウンロード不要



Apptainerの利用

BioContainers Apptainer Images



The screenshot shows a web browser window with the NIG Supercomputer logo at the top. The URL in the address bar is <https://sc.ddbj.nig.ac.jp/guides/software/Container/BioContainers/>. The page title is "BioContainers Apptainer (Singularity) Images の使い方". The main content area features a large heading "BioContainers Apptainer (Singularity) Images の使い方" and a section titled "概要". To the left of the text, there is a screenshot of the BioContainers Flow interface, which shows a flowchart from "SOFTWARE" to "CONTAINER" to "WORKFLOW". The right side contains two columns of text: "Biocontainers Singularity Images" and "各ソフトウェアの内容、使い方の詳細については BioContainers の公式サイトの Registry のページをご参照ください。".

NIG Supercomputer

Software > Container > BioContainers Apptainer (Singularity) Images の使い方

On this page

BioContainers Apptainer (Singularity) Images の使い方

概要

Biocontainers Singularity Images

遺伝研スパコンでは、解析ソフトウェアのインストールの手間を軽減するために、BioContainers project が作成した Singularity コンテナイメージ(2千種類を超える解析ソフトウェア、バージョンの違いを含め 9 万個を超える Singularity イメージファイル)を、遺伝研スパコンの `/usr/local/biotools/` ディレクトリ以下に配置してあります。

各ソフトウェアの内容、使い方の詳細については [BioContainers の公式サイト](#) の [Registry](#) のページをご参照ください。

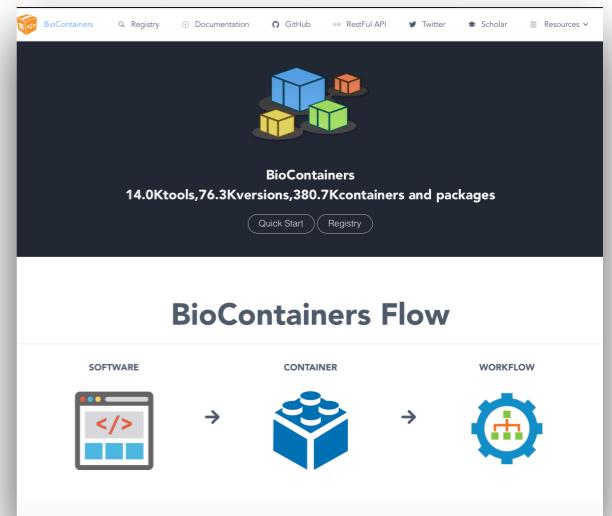
<https://sc.ddbj.nig.ac.jp/guides/software/Container/BioContainers/>

Apptainerの利用

BioContainers Apptainer Images

解析ソフトウェアのインストールの手間を軽減するために、
BioContainers project が作成した Singularity コンテナ
イメージ(2千種類を超える解析ソフトウェア、バージョンの
違いを含め 9 万個を超える Singularity イメージファイル)
を、遺伝研スパコンの

/usr/local/biotools/
ディレクトリ以下に配置してあります。



<https://sc.ddbj.nig.ac.jp/guides/software/Container/BioContainers/>

<https://biocontainers.pro/>

Apptainerの利用

使い方

- ・コンテナの探し方
- ・コンテナの動かし方

Apptainer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)

Apptainerの利用

コンテナの探し方

今回使用するバージョン

```
/usr/local/biotools/h/hisat2:2.2.1--py38he1b5a44_0
```

```
/usr/local/biotools/s/stringtie:3.0.1--h00789bb_0
```

```
/usr/local/biotools/s/samtools:1.22.1--h96c455f_0
```

Apptainer(Singularity)の利用

コンテナの動かし方

基本

```
apptainer 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
```



-B 自分のホームディレクトリ以下「以外」で参照したいディレクトリをバインド

Apptainer(Singularity)の利用

コンテナの動かし方

基本

```
apptainer 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
```



```
/usr/local/biotools/h/hisat2:2.2.1--py38he1b5a44_0
```

Apptainer(Singularity)の利用

コンテナの動かし方

基本

```
apptainer 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  
hisat2 -x genome -1 xxx_1.fq -2 xxx_2.fq -S xxx.sam
```



使ってみましょう

Apptainer(Singularity)の利用

コンテナの動かし方

練習 (StringTieのヘルプメッセージ表示)

```
apptainer exec /usr/local/biotools/s/stringtie:3.0.1--h00789bb_0 stringtie -h
```

lsコマンドで表示されたパスをコピーしてください

Apptainer(Singularity)の利用

コンテナの動かし方

練習その2 seqkitでfastaの統計値を見る

コンテナの場所を探す /usr/local/biotools/ツールの頭文字

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

Apptainer(Singularity)の利用

コンテナの動かし方

練習その2 seqkitでfastaの統計値を見る

GitHub: 4/data/test.fasta をダウンロードしてスパコンにupload

apptainer exec /usr/local/biotoools/s/seqkit:xxx seqkit stats test.fasta

lsコマンドで表示されたものをコピー

The screenshot shows the SeqKit documentation page for the 'stats' command. The URL is <https://bioinf.shenwei.me/seqkit/usage/#stats>. The page includes a table of contents on the right, search and GitHub links at the top, and detailed usage information for the 'stats' command. It lists various metrics like file, format, type, num_seqs, sum_len, min_len, avg_len, max_len, Q1, Q2, Q3, sum_gap, N50, N50_num, Q20_percent, Q30_percent, AvgQual, GC_percent, and sum_n. It also provides tips for handling small files.

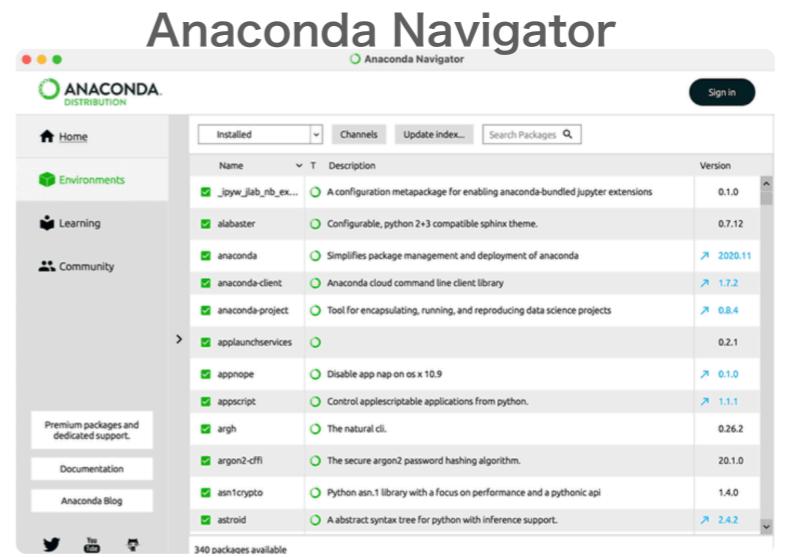
<https://bioinf.shenwei.me/seqkit/usage/#stats>

conda (Miniforge)

Miniforgeの利用

Anaconda/Miniconda/Miniforge

- ✓ 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
- ✓ Miniconda(Miniforge)は最低限必要なパッケージのみ
- より高速なパッケージマネージャ
✓ Miniforgeはmambaが初めから含まれている (2023.9~)



参考

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

Miniforgeの利用

Miniforgeとは

- 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

次は仮想環境の作成

Miniforgeの利用

Miniforgeインストーラ

<https://github.com/conda-forge/miniforge>

The screenshot shows the GitHub repository page for `conda-forge/miniforge`. The repository is public and has 115 issues, 1 pull request, and 323 forks. The main tab is selected, showing a list of recent commits from `hmaarrf`. The commits include merges from other repositories like `amorehead/patch-1` and `jaimergp/system-requiremen...`, as well as updates to scripts and documentation. The sidebar on the right provides an 'About' summary, links to releases (the latest being 24.7.1-2), and a 'Sponsor this project' section for NumFOCUS.

About

A conda-forge distribution.

[conda-forge.org/miniforge](#)

[Readme](#) [View license](#) [Activity](#) [Custom properties](#) [6.2k stars](#) [55 watching](#) [323 forks](#) [Report repository](#)

Releases 86

[Release 24.7.1-2 \(Latest\)](#) last week [+ 85 releases](#)

Sponsor this project

[numfocus NumFOCUS](#) [https://numfocus.org/donate-to-conda...](#)

Learn more about GitHub Sponsors

Packages

No packages published

Code Issues 115 Pull requests 1 Discussions Actions Security Insights

main 6 Branches 87 Tags Go to file Code

hmaarrf Merge pull request #652 from amorehead/patch-1 e00f412 · 2 days ago 717 Commits

.github Merge pull request #626 from jaimergp/system-requiremen... last month

Miniforge3 Use powershell + WScript 2 weeks ago

docs restore deeper docs path 4 years ago

scripts update micromamba version 2 months ago

.gitignore consolidate scripts 5 years ago

LICENSE Fix typo 4 months ago

README.md revert changes to the filename preview. 2 days ago

build_miniforge.sh Do pass \$MINIFORGE_LICENSE_OVERRIDE 3 months ago

build_miniforge_osx.sh Try setup-miniconda for macOS last year

build_miniforge_win.sh Fix shellcheck findings 3 years ago

[README](#) [License](#)

Miniforge

Build miniforge passing downloads 18M

This repository holds the minimal installers for [Conda](#) and [Mamba](#) specific to [conda-forge](#), with the following features pre-configured:

Miniforgeの利用

Miniforgeインストーラ

<https://github.com/conda-forge/miniforge>

The screenshot shows the GitHub repository page for Miniforge. The 'Download' section contains a table of installers for different OS and architectures. A red box highlights the first two rows of the table, which correspond to Linux x86_64 and aarch64 architectures. A red arrow points from the text 'このリンクをコピーして wgetまたはcurlでダウンロード' to the 'Miniforge3-Linux-x86_64' link in the highlighted row.

OS	Architecture	Minimum Version	Download
Linux	x86_64 (amd64)	glibc >= 2.17	Miniforge3-Linux-x86_64
Linux	aarch64 (arm64) (***)	glibc >= 2.17	Miniforge3-Linux-aarch64
Linux	ppc64le (POWER8/9)	glibc >= 2.17	Miniforge3-Linux-ppc64le
macOS	x86_64	macOS >= 10.13	Miniforge3-MacOSX-x86_64
macOS	arm64 (Apple Silicon) (***)	macOS >= 11.0	Miniforge3-MacOSX-arm64
Windows	x86_64	Windows >= 7	Miniforge3-Windows-x86_64

(*) The Python version is specific only to the base environment. Conda can create new environments with different Python versions and implementations.

(**) For Raspberry Pi that include a 64 bit processor, you must also use a 64-bit operating system such as [Raspberry Pi OS 64-bit](#) or [Ubuntu for Raspberry Pi](#). The versions listed as "System: 32-bit" are not compatible with the installers on this website.

(***) Apple silicon builds are experimental and haven't had testing like the other platforms.

Miniforge-pypy3

Latest installers with PyPy 3.9 in the base environment:

OS	Architecture	Minimum Version	Download
----	--------------	-----------------	----------

このリンクをコピーして
wgetまたはcurlでダウンロード

Miniforgeの利用

Miniforgeのインストール

<https://github.com/conda-forge/miniforge>

The screenshot shows the GitHub page for the Miniforge repository. At the top, there is a link to the [README.md](#) file. Below it, a note states: "It can be compared to the [Miniconda](#) installer." A large red rectangular box highlights the "Install" section. This section contains two code snippets for Unix-like platforms (Mac OS & Linux). The first snippet uses curl:

```
curl -L -O "https://github.com/conda-forge/miniforge/releases/latest/download/Miniforge3-$(uname)-$uname-m.sh" Copy
```

The second snippet uses wget:

```
wget "https://github.com/conda-forge/miniforge/releases/latest/download/Miniforge3-$(uname)-$uname-m.sh" Copy
```

(**) For Raspberry Pi that include a 64 bit processor, you must also use a 64-bit operating system such as [Raspberry Pi OS 64-bit](#) or [Ubuntu for Raspberry Pi](#). The versions listed as "System: 32-bit" are not compatible with the installers on this website.

(***) Apple silicon builds are experimental and haven't had testing like the other platforms.

Miniforge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

Miniforge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



Miniforgeの利用

解析ツールのインストール

The screenshot shows the Anaconda.org website interface. At the top, there is a navigation bar with the Anaconda logo, a search bar containing 'stringtie', and links for 'About', 'Anaconda', 'Help', 'Download Anaconda', 'Sign In', and 'Anaconda.org'. Below the header, the main content area displays the package details for 'bioconda / packages / hisat2' version 2.2.1. The title 'bioconda / packages / hisat2 2.2.1' is in green. To the right of the title is a green action bar with icons for home, code, file, star, and a number '7'. A sub-header below the title reads 'Graph-based alignment of next generation sequencing reads to a population of genomes.' Below this, there are tabs for 'Conda', 'Files', 'Labels', and 'Badges', with 'Conda' being the active tab. The 'Conda' section contains the following information:

- License: GPL-3.0
- Home: <http://daehwankimlab.github.io/hisat2>
- </> Development: <https://github.com/DaehwanKimLab/hisat2>
- Documentation: <https://daehwankimlab.github.io/hisat2/manual/>
- Downloads: 297950 total downloads
- Last upload: 1 year and 4 months ago

Below this section is a large 'Installers' box. It contains a note: 'Info: This package contains files in non-standard labels.' It shows two available builds: 'linux-64 v2.2.1' and 'osx-64 v2.2.1'. To the right of these, the text '最新バージョン' (Latest Version) is displayed in red. Below this, there is a 'conda install' command section with the text 'To install this package run one of the following:' followed by two command examples:

```
conda install bioconda::hisat2
conda install bioconda/label/cf201901::hisat2
```

The entire 'Installers' box is highlighted with a red border. To the right of the 'Installers' box, the text 'インストールコマンド' (Install Command) is displayed in red.

では次の講義で
実際に使ってみましょう

