

新学術領域研究「先進ゲノム支援」 2017年度 情報解析講習会

シェルスクリプト

国立遺伝学研究所
生命情報研究センター
森 宙史

シェルスクリプトとは？

- Linuxのコマンドを、実行してほしい一連の処理を実現できるように順番に並べたもの
- 「変数」や「配列」等のデータ構造を使うことができる
- for文やif文等の基礎的なプログラミングの文法が使える（ただし、簡易的なものなので、本格的にプログラミングがしたいのなら、Python等のスクリプト言語を習得すべき）

なぜシェルスクリプトを学ぶ?

- バイオインフォマティクスの解析では、同じ処理を多数のファイルに行うことが多い
- しかも、特にNGSデータは個々のファイルサイズが大きく各処理にわりと時間が掛かる
- 遺伝研スパコン含め大抵のスパコンは、計算を投げるマシンと計算を実行するマシンが別であり、そのようなマシンを使うためにはシェルスクリプトで処理を記述する必要がある

シェルスクリプトの例

- ① FileZillaで遺伝研スパコンのホームディレクトリ上に、
下記のtimes1.shファイルを作成

```
#!/bin/bash  
date  
sleep 7  
date
```

times1.sh

bashシェルを用いると明言

今の日時を表示

7秒間待つ

今の日時を表示

- ② TeraTerm/ターミナル等を用いてsshで遺伝研スパコンにログインし、qlogin後、
下記のコマンドを実行

```
$ chmod u+x times1.sh
```

times1.shを実行可能にする

```
$ ./times1.sh
```

times1.shを実行する

変数

プログラム内で繰り返し使う数値や文字列の入れ物

- ① FileZillaで遺伝研スパコンのホームディレクトリ上に、下記のtimes2.shファイルを作成

```
#!/bin/bash
v=7
date
sleep ${v}
f=`date`
echo ${f}
```

times2.sh

変数vに7を代入

変数v秒間待つ

変数fにdateコマンドの結果を代入

変数fの中身を表示

- ② TeraTerm/ターミナル等を用いてsshで遺伝研スパコンにログインし、qlogin後、下記のコマンドを実行

```
$ chmod u+x times2.sh
$ ./times2.sh
```

変数

プログラム内で繰り返し使う数値や文字列の入れ物

times2.sh

```
#!/bin/bash
v=7
date
sleep ${v}
f=`date`
echo ${f}
```

- 変数の定義と値の代入は、**変数名=値**
=の前後にスペースを入れない
値の代入時には\$や{}は不要
代入する値は文字列でも数値でも良い
コマンドの実行結果を変数に入れたい場合は、
`コマンド名`
変数の呼び出しは、\${変数名}

特殊変数

シェルによって最初から用途が予約されている特殊な変数

- ① FileZillaで遺伝研スパコンのホームディレクトリ上に、下記のtimes3.shファイルを作成

```
#!/bin/bash  
date  
sleep $1  
date
```

times3.sh

特殊変数\$1秒間待つ

- ② TeraTerm/ターミナル等を用いてsshで遺伝研スパコンにログインし、qlogin後、下記のコマンドを実行

```
$ chmod u+x times3.sh  
$ ./times3.sh 7
```

\$1は1番目の引数を受け取る

for文

繰り返し処理に使える文法

- ① FileZillaで遺伝研スパコンのホームディレクトリ上に、下記のec1.shファイルを作成

```
#!/bin/bash
for i in 1 2 a b
do
    echo  ${i}
done
```

ec1.sh

変数iに1, 2, a, bを順に代入し、doからdoneまでの処理をその都度実行。
echoの前に2スペース入れたが、見栄え以外で意味は無い

- ② TeraTerm/ターミナル等を用いてsshで遺伝研スパコンにログインし、qlogin後、下記のコマンドを実行

```
$ chmod u+x ec1.sh
$ ./ec1.sh
```


for文（時間があれば）

繰り返し処理に使える文法

- ① FileZillaで遺伝研スパコンのホームディレクトリ上に、下記のec0.shファイルを作成

```
#!/bin/bash
for i in *.sh
do
    echo ${i}
done
```

ec0.sh

変数iに現在のディレクトリ中の.sh
で終わるファイル名を全て代入

- ② TeraTerm/ターミナル等を用いてsshで遺伝研スパコンにログインし、qlogin後、下記のコマンドを実行

```
$ chmod u+x ec0.sh
$ ./ec0.sh
```

配列

数値や文字列のリストをまとめて扱う

- ① FileZillaで遺伝研スパコンのホームディレクトリ上に、下記のec2.shファイルを作成

```
#!/bin/bash
hoge=(1 2 a b)
echo ${hoge[0]}
echo ${hoge[@]}
echo ${hoge[$1]}
```

ec2.sh

配列hogeに1, 2, a, bのリストを代入

配列hogeの0番目の値を表示

配列hogeの全ての値を表示

配列hogeの引数1個目で指定された値を表示

- ② TeraTerm/ターミナル等を用いてsshで遺伝研スパコンにログインし、qlogin後、下記のコマンドを実行

```
$ chmod u+x ec2.sh
$ ./ec2.sh 3
```

配列とfor文

- ① FileZillaで遺伝研スパコンのホームディレクトリ上に、下記のec3.shファイルを作成

```
#!/bin/bash
hoge=(1 2 a b)
for i in 0 1
do
    echo ${hoge[${i}]}
done
```

ec3.sh

配列hogeの変数i番目の値を表示

- ② TeraTerm/ターミナル等を用いてsshで遺伝研スパコンにログインし、qlogin後、下記のコマンドを実行

```
$ chmod u+x ec3.sh
$ ./ec3.sh
```

コメント

#以降はプログラムとしては無視される

```
#!/bin/bash
date
sleep $1 # $1 is a special value
date
```

- シェルスクリプトは\$や{}、()等の使い方の変数の意味が変わったりするので、後から見てもすぐにわかるようにマメにコメントを残しておくべき
- UGEで計算を投げる時には初めの#以降を特殊な用途に利用するので、注意が必要

シェルスクリプトまとめ

- シェルスクリプトは、Linuxのコマンドを順番に並べたもの
- 「変数」や「配列」等のデータ構造や、for文やif文等の基礎的なプログラミング文法が使える

シェルスクリプトについては: UNIX & Linux コマンド・シェルスクリプト リファレンス

<https://shellscript.sunone.me/>

Linuxコマンドについては: 東大・新領域 笠原先生のコマンドライン講習会資料

<http://cmdline.2016.class.kasahara.ws/>

バッチジョブについて

- <https://sc2.ddbj.nig.ac.jp/index.php/ja-howtouse>