

シェルスクリプト入門

I T の チ カ ラ で 研 究 を 支 援



アメリエフ株式会社

Copyright © Amelieff Corporation All Rights Reserved.

本講義にあたって

- 講師の服部について
 - アメリエフ株式会社でバイオインフォマティクスの 解析を行っています
 - LinuxやRやソフトウェアの使い方のトレーニングも
 担当しています
 - 新潟県出身でお酒は飲めますがスキーは下手です
 - 大学でショウジョウバエの進化を研究していました
 - 社会人になってからはずっとSEをやっています
 - 趣味は落語と三味線です

本講義にあたって

- テキストが穴埋めになっています
 埋めて完成させてください
- クイズがたくさんあります

- めざせ全問正解!(賞品は特にありませんが...)

実習もたくさんあります

- とにかく書いてみるのが理解の早道です

- どうしても難しい場合は「sh_answer」ディレクトリ に解答例がありますのでコピーして実行してください



本講義にあたって

クイズの解答など、お手元の資料には 入っていないページがあります



本講義の内容

- プロローグ
- シェルスクリプト
 標準出力と とは
- 変数
- 引数
- 条件付き処理

- ・繰り返し処理
- - 標準エラー出力
- ・シバン
- エピローグ

プロローグ

- あなたは解析担当者です
- 今は朝の10時です
- ・ 突然、 一本の電話がかかってきました

共同研究者の○△だけど 例の解析結果が急に必要になったので 今日の18時までに送ってもらえる? よろしく!

プロローク

(無理だ…) あなたは頭を抱えました

• その解析はA,B,Cという3つのソフトを 順番に実行する必要があるのですが...



プロローグ

今日に限って会議が2つも入っています (18時までに終えるのは無理だ…) (締切を20時まで延ばしてもらえるかな?)

(飲み会は諦めよう…)





プロローク

シェルスクリプトのおかげであなたは 共同研究者の要望に応え飲み会にも行くことができました



Copyright © Amelieff Corporation All Rights Reserved.

シェルスクリプトとは

「Linuxコマンド」をファイルに書いたもの –書かれた内容をLinuxが自動実行 の数・条件付き処理・繰り返し処理 などのプログラミングが可能

シェルスクリプトのメリット

• 効率的に解析できる

- 指定通り自動で実行されるので、解析の待ち時間が減らせる
- 同じ処理を別のデータや異なる条件で繰り返し実行しやすい
- 実行ログを残しやすい



シェルスクリプトの強み

バイオインフォのソフトは毎年変わる しかし Linuxコマンドや、シェルスクリプトの文法 はこの10年ほとんど変わっていない つまり フまり

今身につければ (おそらく) あと10年は使える

シェルとは

ユーザが入力したコマンドをコン ピュータに伝えるプログラムです

再起動したまえ

–bash

-tcsh

–zsh

などがあります

&*>=@,

(再起動しろってさ)

#"!%

(了解!)

シェルの種類

- 本テキストはbashをベースとした
 記述になっています
- BioLinuxはデフォルトがzshです
- 基本的には大きな差はありませんが 特にzshで挙動が異なるところには 注釈を入れています

シェルスクリプトの作成と実行

1. テキストエディタ(vi, gedit等)で 実行内容をファイルに書いて保存

シェルスクリプトファイルは拡張子を「.sh」にします

2. bashコマンドで実行

\$ bash シェルスクリプトファイル名

実習環境

- 仮想環境を起動します
- ホームディレクトリの下の amelieff/shディレクトリに移動
- 実習はすべてここで行います

\$ cd
\$ cd amelieff/sh

実習環境

• テストデータ

- mirbaseからダウンロードしたmiRNA配列
- -hairpin.fa(miRNA前駆体配列)
 - <u>ftp://mirbase.org/pub/mirbase/CURRENT/hairpin.fa.zip</u>
- -mature.fa(成熟miRNA配列)
 - <u>ftp://mirbase.org/pub/mirbase/CURRENT/mature.fa.zip</u>

どちらもFastaフォーマットのファイルです

Fastaフォーマット

- >で始まるID行と配列行(塩基またはアミノ酸)
 から成るフォーマットです
- ゲノムや遺伝子の配列を表すのによく使われます

>cel-let-7 MI0000001 Caenorh Citis elegans let-7 stem-loop UACACUGUGGAUCCGGUGAGGUAGUAG UAUGCAAUUUUCUACCUUACCGGAGACAGAACUCUUCGA

>cel-lin-4 MI0000002 Caenorhabditis elegans lin-4 stem-loop AUGCUUCCGGCCUGUUCCCUGAGACCUCAAGUGUGAGUGUACUAUUGAUGCUUCACACCU GGGCUCUCCGGGUACCAGGACGGUUUGAGCAGAU

達成目標

- ・以下の作業をシェルスクリプトで
 自動実行できるようになりましょう
 - -mature.faの一部の配列を切り出して 別のファイルに書き出す

-書き出したファイルに対して weblogoというソフトウェアを 実行してモチーフ図を描く

ファイルの先頭を表示する

 Linuxのheadコマンドを実行すると、 指定したファイルの先頭数行が表示 されます

\$ head -n 4 mature.fa

>cel-let-7-5p MIMAT0000001 ... UGAGGUAGUAGGUUGUAUAGUU >cel-let-7-3p MIMAT0015091 ... CUAUGCAAUUUUCUACCUUACC head –n *k* ファイル : ファイルの先頭*k*行を出力する

mature.faファイルの 先頭4行が表示される

実習1

・次のシェルスクリプト・test1.shを 書いて実行してみましょう

- mature.faファイルの先頭4行を表示するシェルスクリプトです

\$ gedit test1.sh

test1.shにこの1行を書いて保存します

head -n 4 mature.fa

\$ bash test1.sh





では、別のファイル・hairpin.faの 先頭8行を表示するようにするには スクリプトをどう変更すればよいで しょう?



• 実行内容を以下のように変えます





• でも、何力所もあったら?

echo "mature.fa の先頭4行は?" head -n 4 mature.fa

echo "mature.fa の末尾4行は?" tail -n 4 mature.fa echo:値を出力する tail –n *k*: 末尾*k*行を出力する wc –l:行数を出力する

直すの面倒くさい!

直し忘れがありそう

echo "mature.fa の行数は?" wc -l mature.fa

Copyright © Amelieff Corporation All Rights Reserved.



[<mark>変数</mark>] を使うと値を一元管理できます

FILE="mature.fa" echo "\$FILE の先頭4行は?" head -n 4 \$FILE

echo "\$FILE の末尾4行は?" tail -n 4 \$FILE

echo "\$FILE の行数は?" wc -l \$FILE 変数「FILE」に

ファイル名を入れる

それ以降は

設定した値が

自動で入る!

「\$FILE」と書くと



•「変数」は値を格納するものです -入れた値は変更することができます - 「変数名=値|と書くと、変数に値 を代入できます -「\$変数」と書くと、変数に入って いる値を呼び出すことができます

実習2

次のシェルスクリプト・test2.shを 書いて実行してみましょう

<mark>\$ cp test1.sh test2.sh</mark> ^{cp}

cp FileA FileB: FileAをFileBという名前で複製

\$ gedit test2.sh

test2.shを以下のように変更して保存します

num=4

head -n \$num \$file

\$ bash test2.sh

実習2 解答

・実習1と同じ挙動になります



変数のありがたみがわかる例

• 変数を使わずに書いたスクリプトの例



変数のありがたみがわかる例

• 変数を使って書いたスクリプトの例



変数にしたほうがいいもの

- 実行のたびに変わる可能性のある値
 –例) 「 入力ファイル名 」
- スクリプト内に何度も登場する値
- -例)「 ^{リファレンスゲノム}」 • なるべく変数にしておくと後で修正が しやすい



「<u>引数</u>」を使うとスクリプトの中身を いちいち書き換えなくてよくなります



「引数」は実行時にスクリプト名以降 に入力された値(複数可能)です

\$ bash test3.sh mature.fa 4 ...

- 引数は自動で専用の変数に入ります
 - 変数\$1に1番目の引数の値が、\$2に2番目の引数 の値が(\$3以下同様)入ります

引数

実習3

次のシェルスクリプト・test3.shを 書いて実行してみましょう

以下のように変更して保存



\$ bash test3.sh mature.fa 4

応用: hairpin.faの 先頭10行を 表示するには どう実行したら いいでしょう?

実習3・解答

• 実習1・2と同じ挙動になります

\$ bash test3.sh mature.fa 4

file=\$1 num=\$2 head -n \$num \$file

–引数を変えると実行内容が変わります

\$ bash test3.sh hairpin.fa 10




love

引数は空白で区切って与えます

ディレクトリを作成する

・以下は「sun」「moon」という
 名前の2つのディレクトリを作成
 するスクリプトです

mkdir "sun" "moon"

mkdir: ディレクトリを作成する

実習4

次のシェルスクリプト・test4.shを 書いて実行してみましょう

- "sun"と"moon"というディレクトリを作成する

- ディレクトリ名は引数で指定する

\$ bash test4.sh sun moon
\$ ls:ファイルとディレクトリ
の一覧を表示
0 一覧を表示
• もう一度実行してみましよう
\$ bash test4.sh sun
moon



• 同じ引数で再実行するとエラー

mkdir: ディレクトリ 'sun' を作成できません: ファイルが存在します mkdir: ディレクトリ 'moon' を作成できません: ファイルが存在します



条件を満たした時だけ処理を実行さ せることができます





• ファイル・ディレクトリの存在確認



実習5

次のシェルスクリプト・test5.shを 書いて実行してみましょう

- 2つのお好きな名前のディレクトリを作成する

- ディレクトリ名は引数で受け取る

- ディレクトリが存在しない場合のみmkdirする

同じ引数で2回実行してエラーが出ない
 ことを確認します







・ 変数の値に応じた処理などが可能です –例) 変数Aが100より大きければ –例) 変数Bが"cancer"でなければ

- 値の比較には「**比較演算子**」を使います
 - 数値の比較演算子

•	文字列の比較演算子
---	-----------

A -eq B	A=Bなら
A -ne B	A≠Bなら
A -lt B	A <bなら< td=""></bなら<>
A -le B	A≦Bなら
A -ge B	A≧Bなら
A -gt B	A>Bなら

A = B	AとBが 同じなら
A != B	AとBが 異なれば

・ 変数を使った条件付き処理





・複数の条件付き処理の例

TEMPERATURE=\$1 TEMPERATUREが if [\$TEMPERATURE -ge 30] 30以上だったら then echo "Hot enough for you?" elif [\$TEMPERATURE -le 10] TEMPERATUREが then 10以下だったら echo "Cold enough for you?" else TEMPERATUREが echo "It's a nice day それ以外だったら today." fi

難易度:★★ Q2.sh mkdir dir3 cd dir3 クイズ if [! -f foo.txt] then スクリプトの実行結果 touch "foo.txt" else はどうなりますか? echo "既に存在します" fi 実行開始時点でdir3は存在しないものと します \$ bash Q2.sh А 「既に存在します」と出力 dir3のみ作成される B dir3と エラーになる foo.txtが作成される

クイズ スクリプトの実行 結果はどうなりま すか?

Q2.sh mkdir dir3 cd dir3 if [! -f foo.txt] then touch "foo.txt" else echo **"既に存在します"** fi

\$ bash Q2.sh

dir3と foo.txtが作成される

右のようにIf文でセミコロン(;) を使うと使用する行数を抑える ことができます

Q2.sh別解

```
mkdir dir3
cd dir3
if [ ! -f foo.txt ];then
touch "foo.txt"
else
echo "既に存在します"
fi
```

実習5の改善点

```
dir1=$1
dir2=$2
if [ !-d $dir1 ]
then
mkdir $dir1
fi
                       なんとなく冗長な
if [ !-d $dir2 ]
                      感じがしませんか?
then
mkdir $dir2
fi
```



• 「繰り返し処理」を用いれば、何度も 実行する処理を1回書くだけで よくなります

繰り返し処理

• 繰り返し処理の構文

for _ **変数** _ in _ 值1 值2 值3 ... do ~処理~ done

繰り返し処理

・ 繰り返し処理の例

- 「1」「2」...「100」という名前のファイルをtouchコマンドで作成するスクリプト



m:n-mからnまでの1刻みの数 touch ファイルを作成する

> `seq n m` n**から**mまで1刻みの数



\$ bash Q3.sh hairpin.fa mature.fa Out

Α

hairpin.fa**の先頭2行が** Out**に出力される**

В

hairpin.faの先頭2行と mature.faの先頭2行が Outに出力される

С

mature.fa**の先頭2行が** Out**に出力される**

D エラーになる



ちなみに、>を>>にすると ファイル書き出しが追記になり hairpin.faの先頭2行の次に、 mature.faのf2の先頭2行が 出力されます Q3.sh修正版 f1=\$1 f2=\$2 out=\$3 for f in \$f1 \$f2 do head -n 2 \$f >> \$out done

実習6

次のシェルスクリプト・test5.shを 書いて実行してみましょう

- 3つのお好きな名前のディレクトリを作成する
- ディレクトリ名は引数で受け取る
- ディレクトリが存在しない場合のみmkdirする
- ディレクトリを作成する手順はfor文を使って 1回だけ記述する

実習6・解答例

```
dir1=$1
dir2=$2
dir3=$3
for dir in $dir1 $dir2 $dir3
do
 if [ ! -d $dir ]
                     何度も実行する処理だが
 then
                    書くのは一回だけなので楽
  mkdir $dir
 fi
                    処理内容に変更があっても
done
                     ここだけ変更すればよい
```



実行コマンドをechoで出力すると 結果がわかりやすくなります

file=\$1

echo "\$file のマッピング開始" bwa mem genome \$file >out.sam

echo "\$file のマッピング終了"

\$ bash bwa.sh B.fastq B.fastqのマッピング開始 B.fastqのマッピング終了

標準出力と標準エラー出力

正常時の出力と、エラー時の出力を 区別して出すことができます

\$ bash miso_soup.sh ネギを切りました 豆腐を切りました お湯が沸きました ネギと豆腐を投入しました エラー!味噌が見つかりません 終了します
エラー時の出力は したい

標準出力と標準エラー出力

通常のechoの結果は「標準出力」へ、
 末尾に「>&2」をつけてechoした結果
 は「標準エラー出力」へ出力されます



実習7

・次のシェルスクリプト・test7.shを 書いて実行してみましょう

echo "I'm file."

echo "Something wrong." >&2

• 実行結果の違いを確認します



何をやっているかわかりやすくするため
 スクリプトに「コメント」を入れましょう

コメント

#で始まる行はコメント扱いとなり、
 処理に影響しません



シバン

 スクリプトの1行目に以下を記述すると このファイルがシェルスクリプトであ ることが明示的になります
 #!/bin/bash

 これにより、bashコマンドなしでも 実行できるようになります

\$ chmod a+x test8.sh
\$./test8.sh

chmod a+x:実行権限をつける

nts Reserved.



- ・ 複数の配列を入力すると上のような組 ・ 成割合の図を描くソフトウェアです
 - http://weblogo.threeplusone.com
- ・以下のように実行します

- 入力する配列は長さが揃っている必要があります

\$ weblogo -F jpeg < fastaファイル > 結果.jpg

最終課題

次のシェルスクリプト・test8.shを 書いて実行します

【基本】mkdir, headを実行する

- inとoutというディレクトリがあるか確認し、無ければ作る
- mature.faから先頭2行のみを抜き出し、in/mature_hsa.faに書き出す
 - ・ ヒント: head -n 2 mature.fa > in/mature_hsa.fa
- ディレクトリ名やファイル名は変数を使わずに直書きで結構です

余裕のある方は次のページへ

最終課題

• test8.shを以下のように書き換えます

【やや難】mkdir, headを実行する

- inとoutというディレクトリがあるか確認し、無ければ作る
- mature.faから先頭2行のみを抜き出し、in/mature_hsa.faに書き出す
 - ・ ヒント: head -n 2 mature.fa > in/mature_hsa.fa
- ディレクトリ名やファイル名を引数で指定できるようにしましょう

余裕のある方は次のページへ
最終課題

- test8.shを以下のように書き換えます
 - 【やや難】mkdir, grepを実行する
 - inとoutというディレクトリがあるか確認し、無ければ作る
 - mature.faから各生物種の「miR390-5p」のIDと配列のみを抜き出し、in/mature_hsa.faに書き出す
 - ヒント: grep -A 1 'miR390-5p' mature.fa | grep -v '^--\$' > in/mature_hsa.fa

余裕のある方は次のページへ

最終課題

• test8.shを以下のように書き換えます

【より難】mkdir, grep, weblogoを実行する

- inとoutというディレクトリがあるか確認し、無ければ作る
- mature.faから各生物種の「miR390-5p」のIDと配列のみを抜き出し、in/mature_hsa.faに書き出す
 - ・ ヒント: grep -A 1 miR390-5p mature.fa | grep -v '^--\$' > in/mature_hsa.fa
- in/mature_hsa.faをweblogoに入力して結果を out/mature_hsa.jpgに出力する

余裕のある方は次のページへ

最終課題

test8.shについて

【難】hairpin.faに対して実行するにはどう変更したら よいか考えてみる

エピローク

- hairpin.faでは塩基配列が複数行になっている場合があるため、mature.faのスクリプトがそのまま使えません
- ・このような場合には、Perlが必要です





	Q4.sh
	mkdir dirl
解答	cd dirl
	touch a.txt b.txt c.txt
	ls -l *.txt > tmp.txt
dir1内に作成したテキストファイル の数が返ってくる	wc -l tmp.txt

 mkdir でdir1ディレクトリを作成
cd で作成したディレクトリへ移動
touch でテキストファイルを作成
ls -lで拡張子が「.txt」のファイルを全てリスト表示し、 その結果を「>」で tmp.txtへ記載
wc -lで tmp.txtに記載された行数をカウント

tmp.txtの記述を確認する場合 \$ less tmp.txt

Q5.sh

mkdir dir2

cd dir2

echo "ok" > f1.txt 2> f2.txt

head f3.txt > f1.txt 2> f2.txt

\$ bash Q5.sh

実行開始時点でdir2は 存在しないものとします

f2.txtには標準

エラー出力が

何行記載され

るでしょう?

クイズ



難易度:★ ★ ★

		Q5.sh	
		mkdir dir2	
 		cd dir2	
正解は、	Bı	echo "ok" > f1.txt 2> f2.txt	
]	1	head f3.txt > f1.txt 2> f2.txt	
	\perp		

1は標準出力、2は標準エラー出力を表します

1と2の結果をそれぞれf1.txtとf2.txtに振り分けたいときは、 \$ コマンド 1>f1.txt 2>f2.txt と入力します

- ① mkdir でdir2ディレクトリを作成
- ② cd で作成したディレクトリへ移動
- ③ echo "ok"の実行結果をf1.txtへ、エラーをf2.txtへ出力します
 - エラーは出ないので、この時点でf1.txtは「ok」1行、f2.txtは空行です
- ④ head f3.txtを実行すると、f3.txtが存在しないのでエラーがf2.txtに 書き出されます

Linuxのテキストエディタ

- ➢ GUIのエディタとCUIのエディタがあります
 - ➤ GUI: Windows/Macソフトのように、マウスで操作する
 - ▶ 長所:Linux初心者にも操作が容易
 - ➤ 短所: GUIがない環境では使えない
 - ➤ CUI:キーボードからコマンドで操作する
 - ▶ 長所: GUIがない環境でも使える
 - ▶ 短所:操作コマンドを覚える必要がある

gedit

- ➤ CentOSにはデフォルトでgeditというGUIエディタが入っています
- ➤ geditを起動するには \$ gedit

コマンドを実行します



➤ CentOSにはデフォルトでviというCUIエディタが入っています

▶ viを起動するには <mark>\$ vi</mark> コマンドを実行します

▶ viには2つのモードがあり、モードを切り替えながら操作します

▶ 入力モード:文字を入力する

▶ コマンドモード:編集する(切り貼り、ファイルの保存など)



vi

▶ 入力モードのコマンド

Escキー コマンドモードに移行

> コマンドモードのコマンド

а	入力モードに移行(カーソルの右から入力)
0	入力モードに移行(次の行の行頭から入力)
x	1文字カット
dd	今いる行をカット
уу	1行コピー
р	カットした行をペースト
[数字]g	[数字]行に移動
G	最終行に移動
:%s/foo/bar/	文字列置換(fooをbarに置換)