

次世代シーケンサーデータの解析手法 第4回クオリティコントロールとプログラムのイ ンストール: ウェブ資料

東京大学・大学院農学生命科学研究科
孫 建強、湯 敏、清水 謙多郎、門田 幸二
kadota@bi.a.u-tokyo.ac.jp
<http://www.iu.a.u-tokyo.ac.jp/~kadota/>

W1-1

「連載第3回のW25-2の最後」と基本的に同じ。①pwdでカレントディレクトリを表示。②「df -h」でdfコマンドをヒトが判読しやすい(human readable)サイズ単位で表示。このゲストOS環境では全部で146GB確保しており、そのうち65GB使用していることがわかる。③lsコマンドをlong形式かつhuman readableな単位で表示。この作業ディレクトリのみで56GB使用していることもわかる。

```
BioLinux8 [実行中] - Oracle VM VirtualBox
仮想マシン ビュー デバイス ヘルプ
iu@bielinux[~/Documents/srp017156]
① iu@bielinux[srp017156] pwd
/home/iu/Documents/srp017156
② iu@bielinux[srp017156] df -h
Filesystem      Size  Used Avail Use% Mounted on
/dev/sda1       146G   65G   74G   47% /
none            4.0K    0 4.0K   0% /sys/fs/cgroup
udev            991M   4.0K 991M   1% /dev
tmpfs           201M   912K 200M   1% /run
none            5.0M    0 5.0M   0% /run/lock
none           1002M  152K 1001M   1% /run/shm
none            100M   68K  100M   1% /run/user
③ iu@bielinux[srp017156] ls -lh
total 56G
-rw-rw-r-- 1 iu iu  41G 12月 16 01:03 hoge_1.fastq
-rw-rw-r-- 1 iu iu 306M 12月 16 10:50 hoge_subset_1.fastq
-rw-rw-r-- 1 iu iu  7.2G 12月 11 14:51 SRR616268_1.fastq.bz2
-rw-rw-r-- 1 iu iu  6.6G 12月 12 14:59 SRR616268_2.fastq.bz2
-rw-rw-r-- 1 iu iu 306M 12月 16 10:50 subset_1.fastq
-rw-rw-r-- 1 iu iu 278M 12月 16 10:50 subset_2.fastq
iu@bielinux[srp017156] █
```

W1-2

連載第4回の図1で示すようなbzip2圧縮ファイルのみからなる初期状態にするには、赤枠で囲った計4ファイルを削除すればよい。これらはhoge_*とsubset_*のファイル群と表現可能。ここで、アスタリスク(*)は「任意の文字の0回以上の繰り返し」を意味する。他にも「任意の1文字」と意味するドット(.)や、「直前の文字の1回以上の繰り返し」を意味するプラス(+)など、様々なワイルドカードを利用可能。

```
BioLinux8 [実行中] - Oracle VM VirtualBox
仮想マシン ビュー デバイス ヘルプ
iu@bielinux[~/Documents/srp017156]
iu@bielinux[srp017156] pwd
/home/iu/Documents/srp017156
iu@bielinux[srp017156] df -h
Filesystem      Size  Used Avail Use% Mounted on
/dev/sda1       146G   65G   74G   47% /
none            4.0K    0   4.0K    0% /sys/fs/cgroup
udev            991M   4.0K  991M    1% /dev
tmpfs           201M   912K  200M    1% /run
none            5.0M    0   5.0M    0% /run/lock
none           1002M  152K  1001M    1% /run/shm
none            100M   68K   100M    1% /run/user
iu@bielinux[srp017156] ls -lh
total 56G
-rw-rw-r-- 1 iu iu 41G 12月 16 01:03 hoge_1.fastq
-rw-rw-r-- 1 iu iu 306M 12月 16 10:50 hoge_subset_1.fastq
-rw-rw-r-- 1 iu iu 7.2G 12月 11 14:51 SRR616268_1.fastq.bz2
-rw-rw-r-- 1 iu iu 6.6G 12月 12 14:59 SRR616268_2.fastq.bz2
-rw-rw-r-- 1 iu iu 306M 12月 16 10:50 subset_1.fastq
-rw-rw-r-- 1 iu iu 278M 12月 16 10:50 subset_2.fastq
iu@bielinux[srp017156]
```

[6:35午後]

[6:35午後]

[6:35午後]

W1-2

4つ程度であれば1つ1つファイル名を打ち込んでrmコマンドを利用して削除すればよいが、数十~数百のオーダーのファイルを一度に削除する局面もあるため、それを意識した慎重なやり方を伝授します。まずは「hoge_*」ファイル群(この場合hoge_1.fastqとhoge_subset_1.fastaの2つのファイル)の削除から。いきなり「rm -f h*」でも「rm -f hoge_*」と打ち込んでもよいのだが、私はワイルドカードを利用して削除するファイル群を念のためlsで確認する。例えば「ls -lh ho」まで打ち込んでからTabキーを押す(タブ補完)。

```
BioLinux8 [実行中] - Oracle VM VirtualBox
仮想マシン ビュー デバイス ヘルプ
iu@bielinux[~/Documents/srp017156]
iu@bielinux[srp017156] pwd
/home/iu/Documents/srp017156
iu@bielinux[srp017156] df -h
Filesystem      Size  Used Avail Use%
/dev/sda1       146G   65G   74G   47%
none            4.0K    0   4.0K    0%
udev            991M   4.0K  991M    1%
tmpfs           201M   916K  200M    1%
none            5.0M    0   5.0M    0%
none           1002M  152K  1001M    1%
none            100M   68K   100M    1%
iu@bielinux[srp017156] ls -lh
total 56G
-rw-rw-r-- 1 iu iu  41G 12月 16 01:03 hoge_1.fastq
-rw-rw-r-- 1 iu iu 306M 12月 16 10:50 hoge_subset_1.fastq
-rw-rw-r-- 1 iu iu  7.2G 12月 11 14:51 SRR616268_1.fastq.bz2
-rw-rw-r-- 1 iu iu  6.6G 12月 12 14:59 SRR616268_2.fastq.bz2
-rw-rw-r-- 1 iu iu  306M 12月 16 10:50 subset_1.fastq
-rw-rw-r-- 1 iu iu  278M 12月 16 10:50 subset_2.fastq
iu@bielinux[srp017156] ls -lh ho
```

[9:53午前]

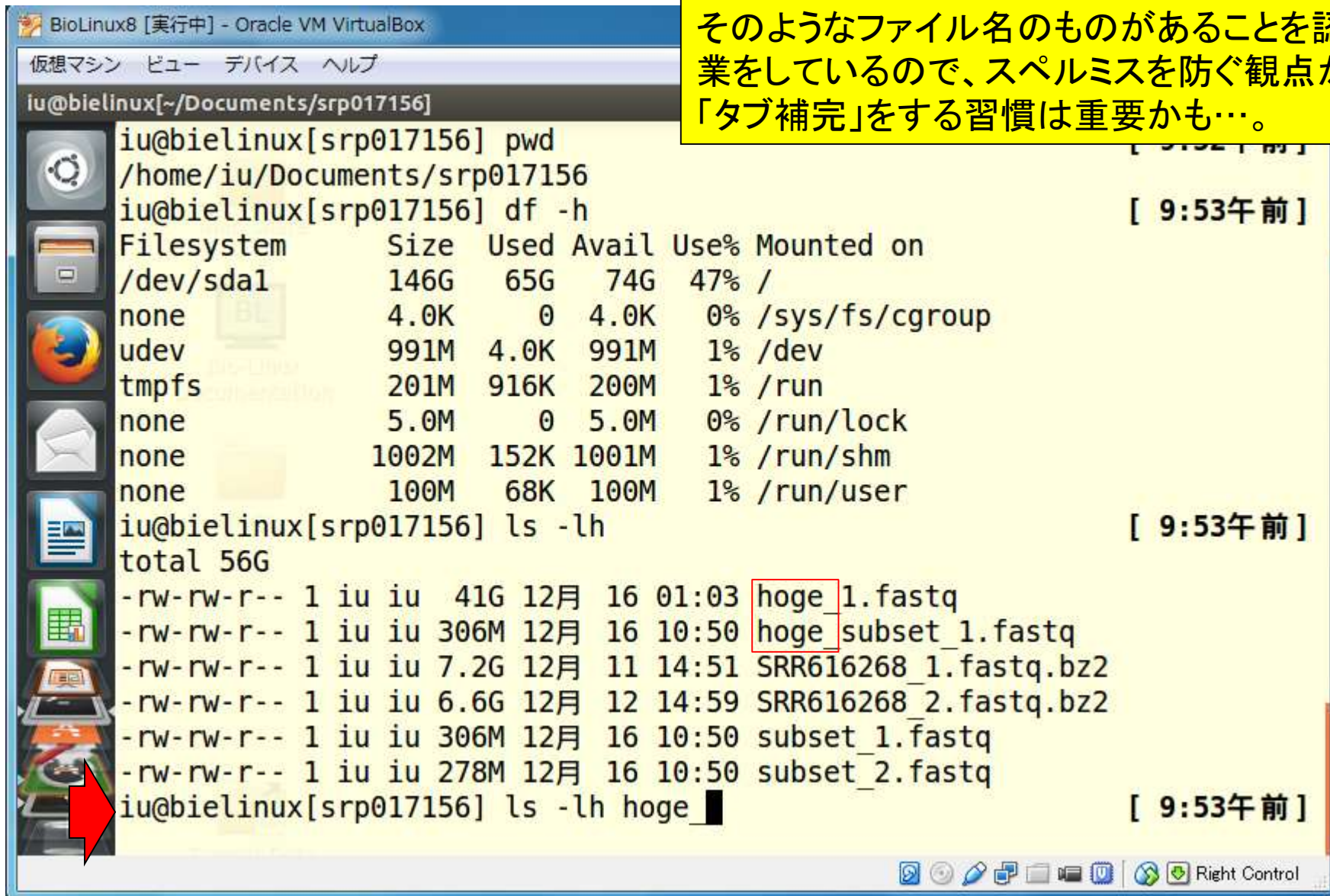
[9:53午前]

[9:53午前]

[9:53午前]

W1-2

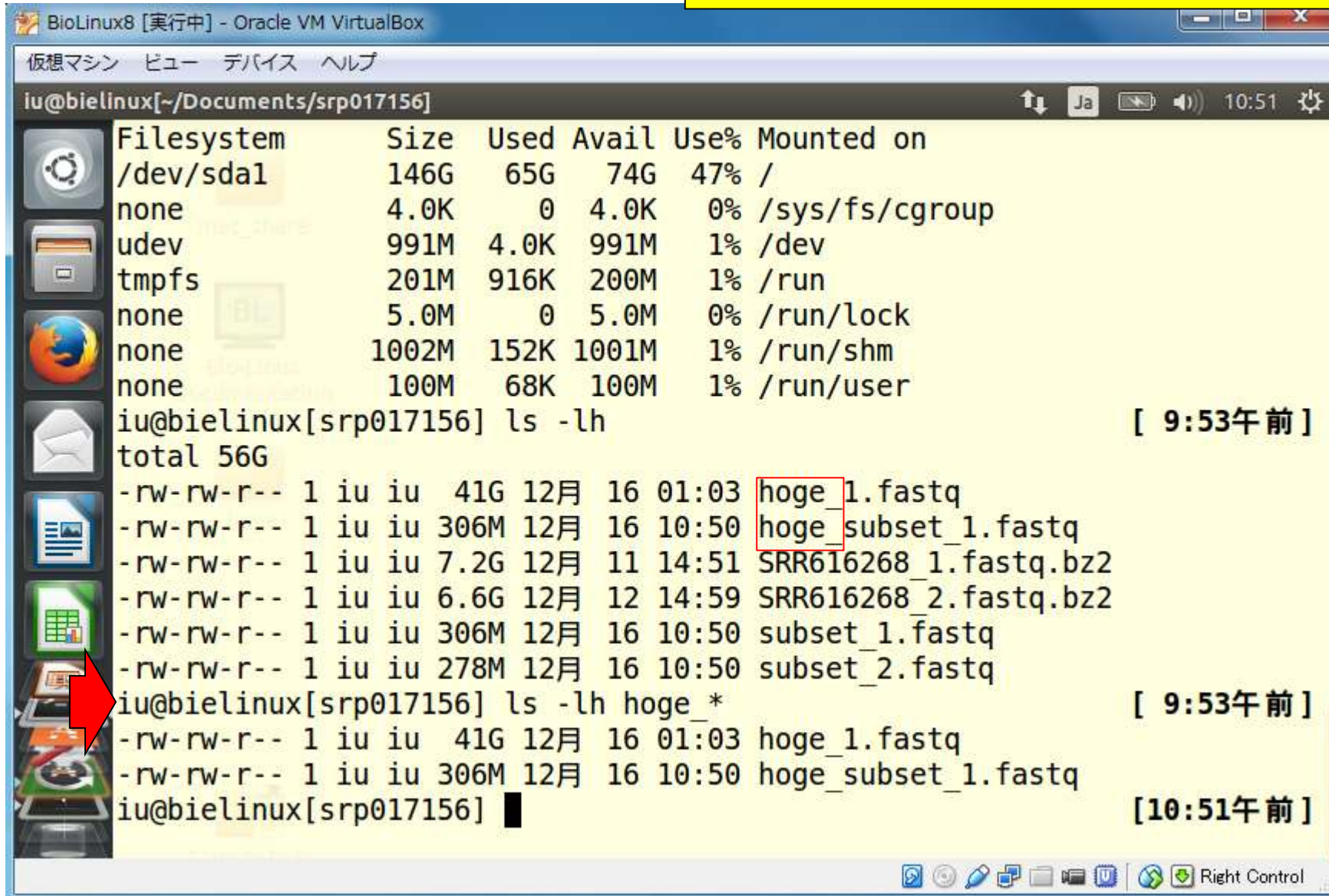
Tabキーを押した直後の状態。「ho」から始まるファイル群の中で共通の文字列部分までを自動的に補完。ただの確認でしかないものの、だいたいカレントディレクトリ中にそのようなファイル名のものがあることを認識した上で作業をしているので、スペルミスを防ぐ観点からも無意識に「タブ補完」をする習慣は重要かも…。



```
BioLinux8 [実行中] - Oracle VM VirtualBox
仮想マシン ビュー デバイス ヘルプ
iu@bielinux[~/Documents/srp017156]
iu@bielinux[srp017156] pwd
/home/iu/Documents/srp017156
iu@bielinux[srp017156] df -h
Filesystem      Size  Used Avail Use% Mounted on
/dev/sda1        146G  65G   74G   47% /
none             4.0K    0   4.0K   0% /sys/fs/cgroup
udev             991M  4.0K 991M   1% /dev
tmpfs            201M  916K 200M   1% /run
none             5.0M    0   5.0M   0% /run/lock
none            1002M 152K 1001M   1% /run/shm
none             100M   68K  100M   1% /run/user
iu@bielinux[srp017156] ls -lh
total 56G
-rw-rw-r-- 1 iu iu  41G 12月 16 01:03 hoge_1.fastq
-rw-rw-r-- 1 iu iu 306M 12月 16 10:50 hoge_subset_1.fastq
-rw-rw-r-- 1 iu iu  7.2G 12月 11 14:51 SRR616268_1.fastq.bz2
-rw-rw-r-- 1 iu iu  6.6G 12月 12 14:59 SRR616268_2.fastq.bz2
-rw-rw-r-- 1 iu iu 306M 12月 16 10:50 subset_1.fastq
-rw-rw-r-- 1 iu iu 278M 12月 16 10:50 subset_2.fastq
iu@bielinux[srp017156] ls -lh hoge_
```

W1-2

アスタリスク(*)をつけてリターンキーを押す。これで削除予定の「hoge_*」から始まるファイル群をリストアップしたことになる。一気に全部を消してよいかどうかを確認。



```
iu@bielinux[~/Documents/srp017156]
filesystem      Size  Used Avail Use% Mounted on
/dev/sda1       146G  65G   74G  47% /
none            4.0K   0    4.0K  0% /sys/fs/cgroup
udev            991M   4.0K 991M   1% /dev
tmpfs           201M  916K 200M   1% /run
none            5.0M   0    5.0M  0% /run/lock
none            1002M 152K 1001M  1% /run/shm
none            100M   68K 100M   1% /run/user

iu@bielinux[srp017156] ls -lh
total 56G
-rw-rw-r-- 1 iu iu  41G 12月 16 01:03 hoge_1.fastq
-rw-rw-r-- 1 iu iu 306M 12月 16 10:50 hoge_subset_1.fastq
-rw-rw-r-- 1 iu iu  7.2G 12月 11 14:51 SRR616268_1.fastq.bz2
-rw-rw-r-- 1 iu iu  6.6G 12月 12 14:59 SRR616268_2.fastq.bz2
-rw-rw-r-- 1 iu iu 306M 12月 16 10:50 subset_1.fastq
-rw-rw-r-- 1 iu iu 278M 12月 16 10:50 subset_2.fastq

iu@bielinux[srp017156] ls -lh hoge_*
-rw-rw-r-- 1 iu iu  41G 12月 16 01:03 hoge_1.fastq
-rw-rw-r-- 1 iu iu 306M 12月 16 10:50 hoge_subset_1.fastq

iu@bielinux[srp017156]
```

W1-3

rmコマンドを用いてファイルの削除。-fは「本当に消してよいか?」的なyes or noを打ち込ませる警告メッセージを表示させないようにするオプション。Linux環境でのrmコマンドの利用は、ゴミ箱への移動ではなく消滅である点に十分注意せよ!。この意味において、-fは一般に危険なオプションである。が、予め「ls -lh hoge_*」として確認していればまあいいだろう、という思想。ディレクトリ削除については連載第3回のW18-4を参照のこと。ここでは「rm -f h」まで打ち込んでからTabキーを押そうとしている。

```
io@bielinux[~/Documents/srp017156]
Filesystem      Size  Used Avail Use% Mounted on
/dev/sda1        146G  65G   74G   45% /
none            4.0K    0  4.0K   0% /boot
udev            991M    4.0K 991M   0% /dev
tmpfs           201M   916K 200M   1% /run
none            5.0M    0  5.0M   0% /run/lock
none           1002M  152K 1001M   1% /run/shm
none            100M   68K  100M   1% /run/user

io@bielinux[srp017156] ls -lh
total 56G
-rw-rw-r-- 1 iu iu  41G 12月 16 01:03 hoge_1.fastq
-rw-rw-r-- 1 iu iu 306M 12月 16 10:50 hoge_subset_1.fastq
-rw-rw-r-- 1 iu iu  7.2G 12月 11 14:51 SRR616268_1.fastq.bz2
-rw-rw-r-- 1 iu iu  6.6G 12月 12 14:59 SRR616268_2.fastq.bz2
-rw-rw-r-- 1 iu iu 306M 12月 16 10:50 subset_1.fastq
-rw-rw-r-- 1 iu iu 278M 12月 16 10:50 subset_2.fastq
io@bielinux[srp017156] ls -lh hoge_*
-rw-rw-r-- 1 iu iu  41G 12月 16 01:03 hoge_1.fastq
-rw-rw-r-- 1 iu iu 306M 12月 16 10:50 hoge_subset_1.fastq
io@bielinux[srp017156] rm -f h
```

[9:53午前]

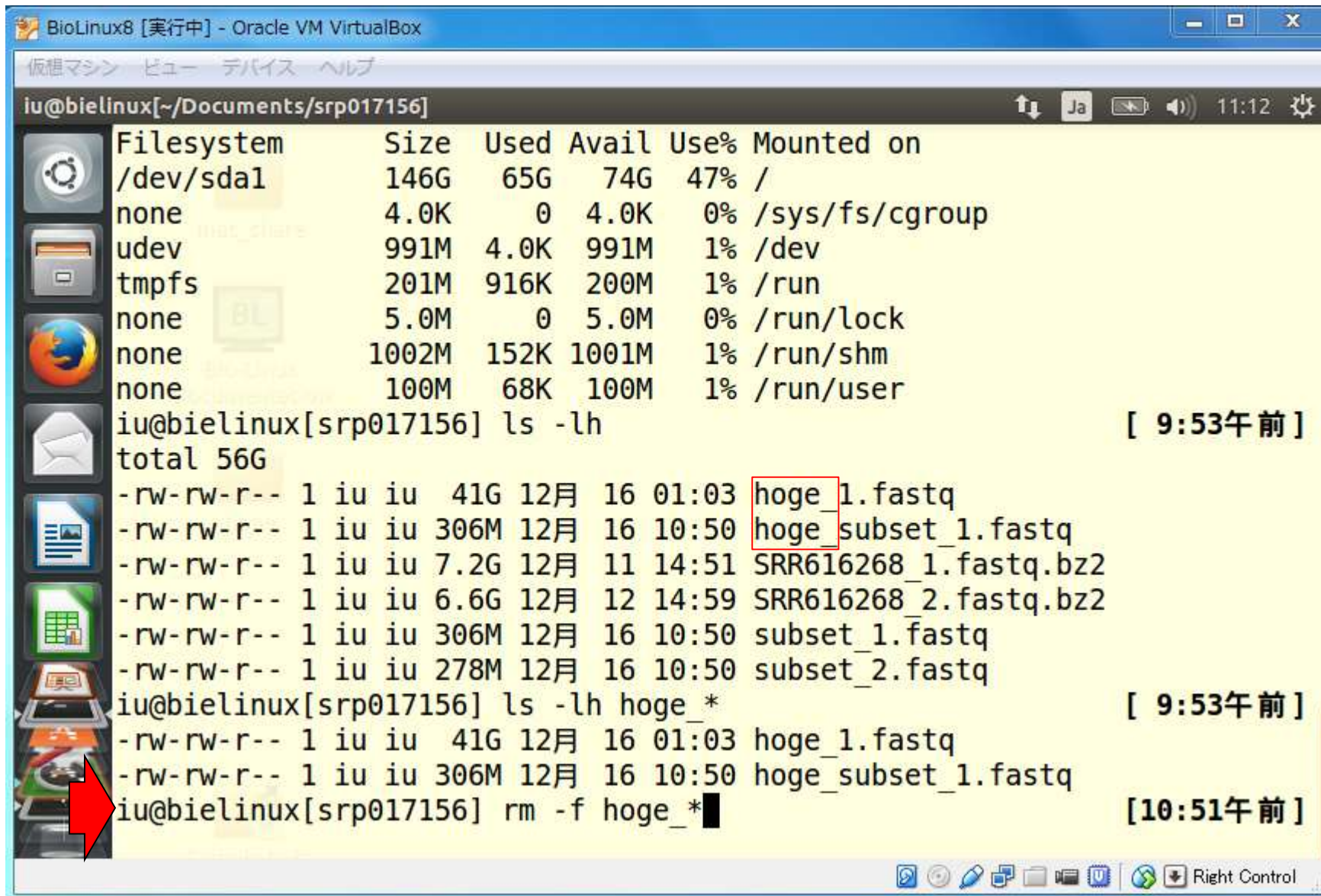
[9:53午前]

[10:51午前]



W1-3

Tabキーを押して「hoge_」まで表示させてから、アスタリスク(*)をつけたところ。この段階でリターンキーを押す。



```
iu@bielinux[~/Documents/srp017156]
filesystem      Size  Used Avail Use% Mounted on
/dev/sda1       146G  65G   74G  47% /
none            4.0K   0    4.0K   0% /sys/fs/cgroup
udev            991M  4.0K  991M   1% /dev
tmpfs           201M  916K  200M   1% /run
none            5.0M   0    5.0M   0% /run/lock
none            1002M 152K 1001M   1% /run/shm
none            100M   68K  100M   1% /run/user

iu@bielinux[srp017156] ls -lh
total 56G
-rw-rw-r-- 1 iu iu  41G 12月 16 01:03 hoge_1.fastq
-rw-rw-r-- 1 iu iu 306M 12月 16 10:50 hoge_subset_1.fastq
-rw-rw-r-- 1 iu iu  7.2G 12月 11 14:51 SRR616268_1.fastq.bz2
-rw-rw-r-- 1 iu iu  6.6G 12月 12 14:59 SRR616268_2.fastq.bz2
-rw-rw-r-- 1 iu iu 306M 12月 16 10:50 subset_1.fastq
-rw-rw-r-- 1 iu iu 278M 12月 16 10:50 subset_2.fastq

iu@bielinux[srp017156] ls -lh hoge_*
-rw-rw-r-- 1 iu iu  41G 12月 16 01:03 hoge_1.fastq
-rw-rw-r-- 1 iu iu 306M 12月 16 10:50 hoge_subset_1.fastq

iu@bielinux[srp017156] rm -f hoge_*
```


W1-3

「rm -f hoge_*」でリターンキーを押したのち、もう一度「ls -lh」した結果を表示。確かに警告メッセージが出ることなくファイル群 hoge_*を削除できていることがわかる。

```
iu@bielinux[~/Documents/srp017156]
none          100M   68K   100M   1% /run/user
iu@bielinux[srp017156] ls -lh
total 56G
-rw-rw-r-- 1 iu iu  41G 12月 16 01:03 hoge_1.fastq
-rw-rw-r-- 1 iu iu 306M 12月 16 10:50 hoge_subset_1.fastq
-rw-rw-r-- 1 iu iu  7.2G 12月 11 14:51 SRR616268_1.fastq.bz2
-rw-rw-r-- 1 iu iu  6.6G 12月 12 14:59 SRR616268_2.fastq.bz2
-rw-rw-r-- 1 iu iu 306M 12月 16 10:50 subset_1.fastq
-rw-rw-r-- 1 iu iu 278M 12月 16 10:50 subset_2.fastq
iu@bielinux[srp017156] ls -lh hoge_*
-rw-rw-r-- 1 iu iu  41G 12月 16 01:03 hoge_1.fastq
-rw-rw-r-- 1 iu iu 306M 12月 16 10:50 hoge_subset_1.fastq
iu@bielinux[srp017156] rm -f hoge_*
iu@bielinux[srp017156] ls -lh
total 15G
-rw-rw-r-- 1 iu iu  7.2G 12月 11 14:51 SRR616268_1.fastq.bz2
-rw-rw-r-- 1 iu iu  6.6G 12月 12 14:59 SRR616268_2.fastq.bz2
-rw-rw-r-- 1 iu iu 306M 12月 16 10:50 subset_1.fastq
-rw-rw-r-- 1 iu iu 278M 12月 16 10:50 subset_2.fastq
iu@bielinux[srp017156]
```

W1-3

一連の作業をファイル群 subset_*についても実行。

```
io@bielinux[~/Documents/srp017156]
-rw-rw-r-- 1 iu iu 41G 12月 16 01:03 hoge_1.fastq
-rw-rw-r-- 1 iu iu 306M 12月 16 10:50 hoge_subset_1.fastq
iu@bielinux[srp017156] rm -f hoge_*
iu@bielinux[srp017156] ls -lh
total 15G
-rw-rw-r-- 1 iu iu 7.2G 12月 11 14:51 SRR616268_1.fastq.bz2
-rw-rw-r-- 1 iu iu 6.6G 12月 12 14:59 SRR616268_2.fastq.bz2
-rw-rw-r-- 1 iu iu 306M 12月 16 10:50 subset_1.fastq
-rw-rw-r-- 1 iu iu 278M 12月 16 10:50 subset_2.fastq
① iu@bielinux[srp017156] ls -lh subset_*
-rw-rw-r-- 1 iu iu 306M 12月 16 10:50 subset_1.fastq
-rw-rw-r-- 1 iu iu 278M 12月 16 10:50 subset_2.fastq
② iu@bielinux[srp017156] rm -f subset_*
iu@bielinux[srp017156] ls -lh ← ③
total 14G
-rw-rw-r-- 1 iu iu 7.2G 12月 11 14:51 SRR616268_1.fastq.bz2
-rw-rw-r-- 1 iu iu 6.6G 12月 12 14:59 SRR616268_2.fastq.bz2
④ iu@bielinux[srp017156] pwd
/home/iu/Documents/srp017156
iu@bielinux[srp017156] █
```

W1-4

連載第4回の図1とほぼ同じ。③のduコマンドでカレントディレクトリのディスク使用量が14Gであることを確認している。ドット(.)は「カレントディレクトリ」という意味。④dfコマンドで全体のディスク使用量を表示させている。赤枠部分のSizeが146GBとなっているのは150GBを割り当てていたから。Usedは全体で23GB使っていることを示す。読者の環境でSizeが48GBあり、Usedが30GB以下程度であれば連載第4回のかんりの解析ができるはず。

```
BioLinux8 [実行中] - Oracle VM VirtualBox
仮想マシン ビュー デバイス ヘルプ
iu@bielinux[~/Documents/srp017156]
① iu@bielinux[srp017156] pwd
/home/iu/Documents/srp017156
iu@bielinux[srp017156] ls -lh ②
total 14G
-rw-rw-r-- 1 iu iu 7.2G 12月 11 14:51 SRR616268_1.fastq.bz2
-rw-rw-r-- 1 iu iu 6.6G 12月 12 14:59 SRR616268_2.fastq.bz2
③ iu@bielinux[srp017156] du -h
14G  .
iu@bielinux[srp017156] df -h ④
Filesystem      Size  Used Avail Use% Mounted on
/dev/sda1       146G  23G  116G  17% /
none            4.0K   0  4.0K   0% /sys/fs/cgroup
udev           991M   4.0K 991M   1% /dev
tmpfs          201M  916K 200M   1% /run
none           5.0M   0  5.0M   0% /run/lock
none          1002M  156K 1001M   1% /run/shm
none          100M   60K  100M   1% /run/user
iu@bielinux[srp017156] █
```

[1:20午後]
[1:20午後]
[1:20午後]

Right Control

W2-1

シェルスクリプトの基本形を示すべく、①テキストエディタgeditを起動し、②(簡単なコマンドならなんでもよいが)3つのコマンドを複数行にわたり書いたものを作成。この際、赤枠のターミナルは、geditを起動して編集中(gedit画面がアクティブな状態)のため、③コマンド入力待ち状態になっていないことがわかる。つまり、geditを終了しない限り他のコマンドを受け付けない。④Saveボタンを押す。

Text Edit File Edit View Search Tools Documents Help

iu@bielinux[~/Documents/srp017156]

iu@bielinux[srp017156] pwd

/home/iu/Documents/srp017156

iu@bielinux[srp017156] ls -lh

total 14G

-rw-rw-r-- 1 iu iu 7.2G 12月 11 14:51 SRR616268_1.fastq.bz2

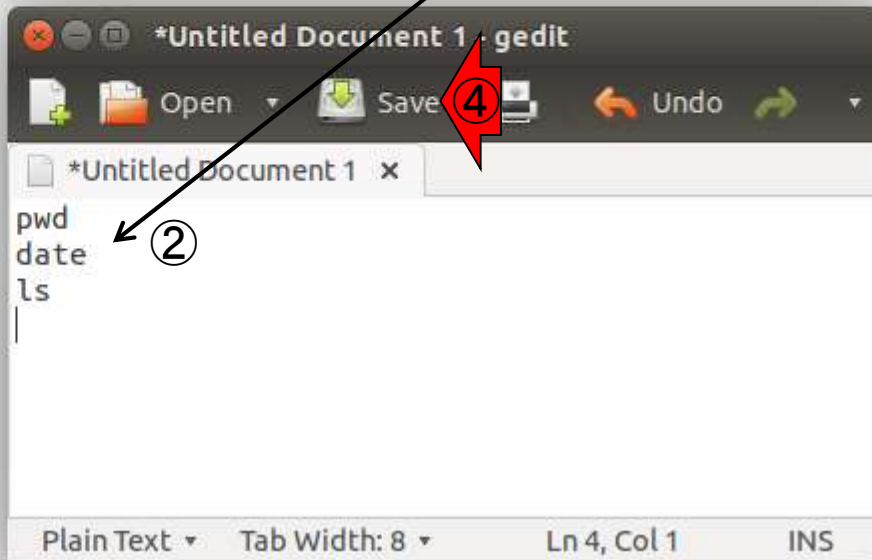
-rw-rw-r-- 1 iu iu 6.6G 12月 12 14:59 SRR616268_2.fastq.bz2

iu@bielinux[srp017156] gedit

[3:14午後]

①

③

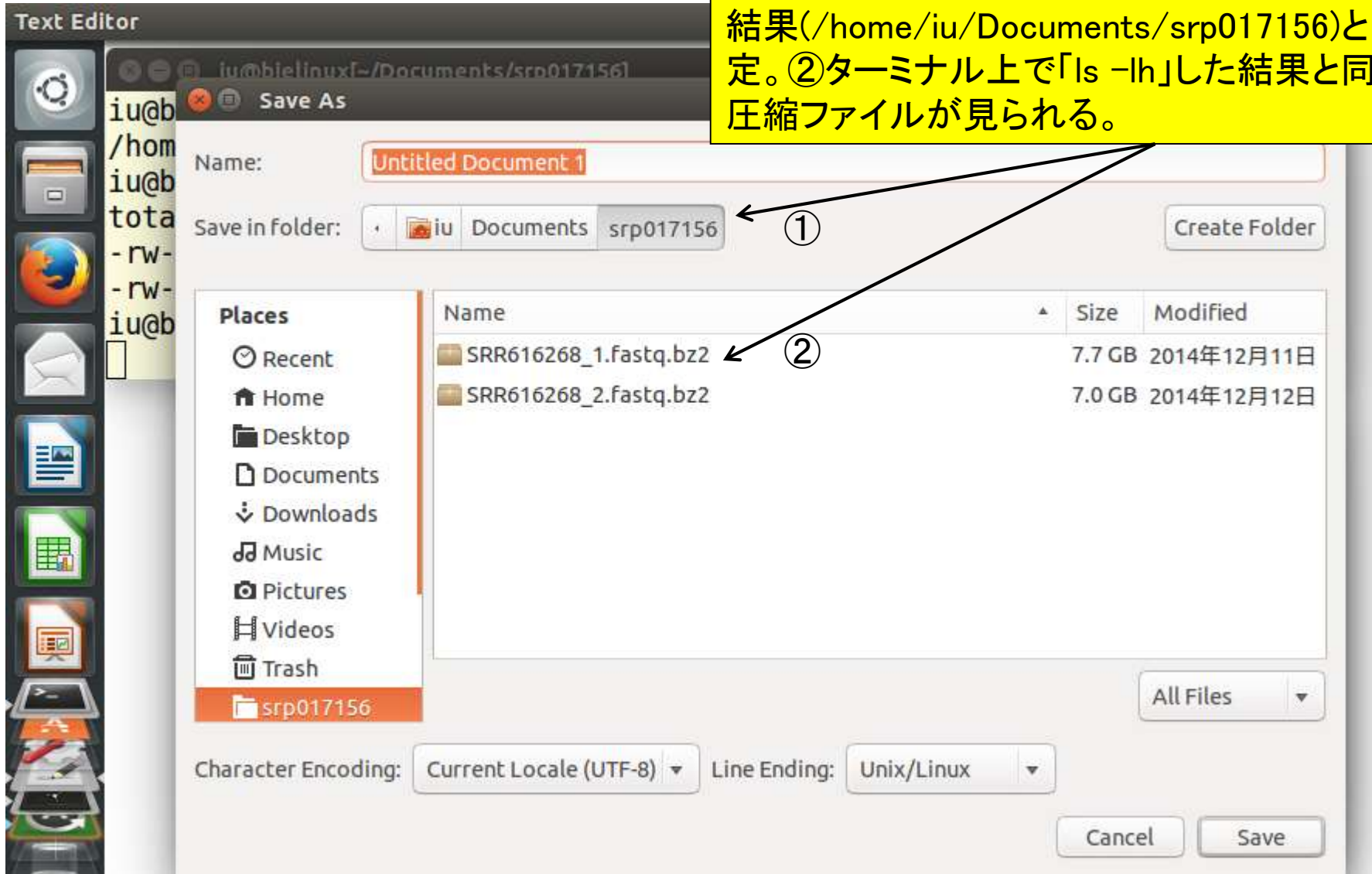


②

④

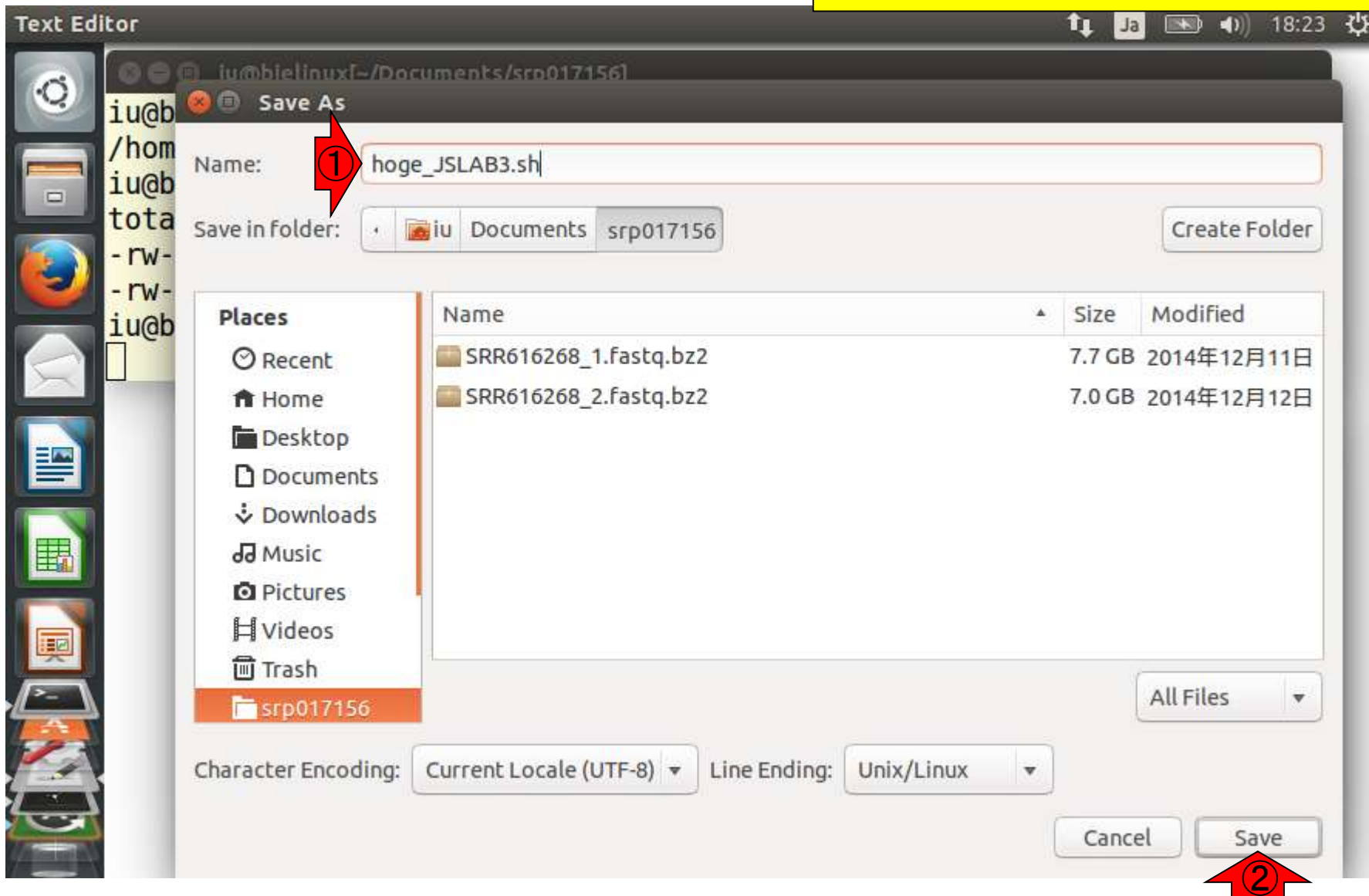
W2-1

①保存するディレクトリのデフォルトは、geditを起動させたターミナルのカレントディレクトリとなっている(少なくとも著者の環境では)。もしそうになっていなければpwd表示結果(/home/iu/Documents/srp017156)と同じ場所を指定。②ターミナル上で「ls -lh」した結果と同じ2つのbzip2圧縮ファイルが見られる。



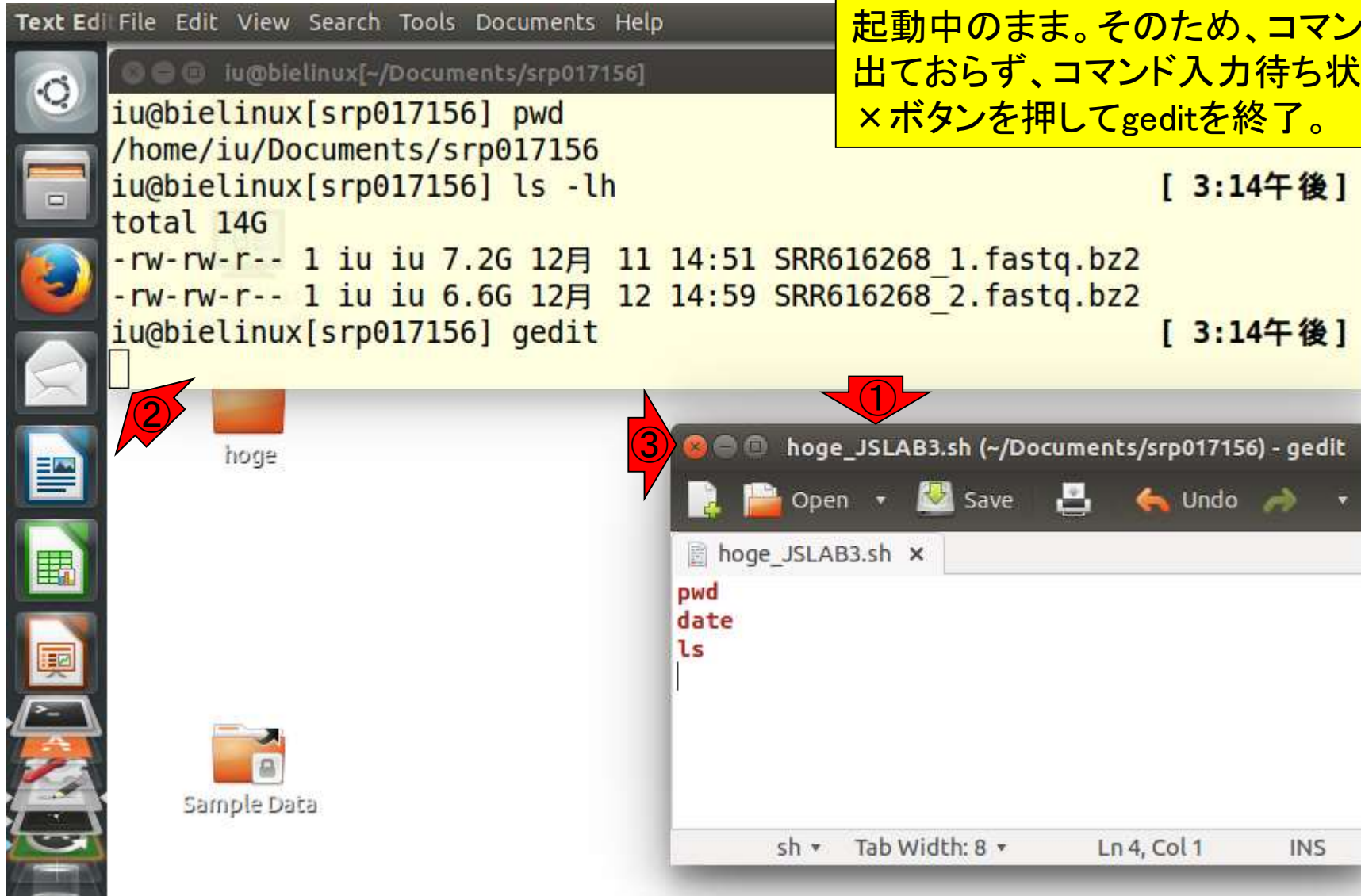
W2-1

ここでは①ファイル名をhoge_JSLAB3.shとして、
②Saveボタンを押している。シェルスクリプトの場合は、拡張子を.shにするのがおそらく一般的。



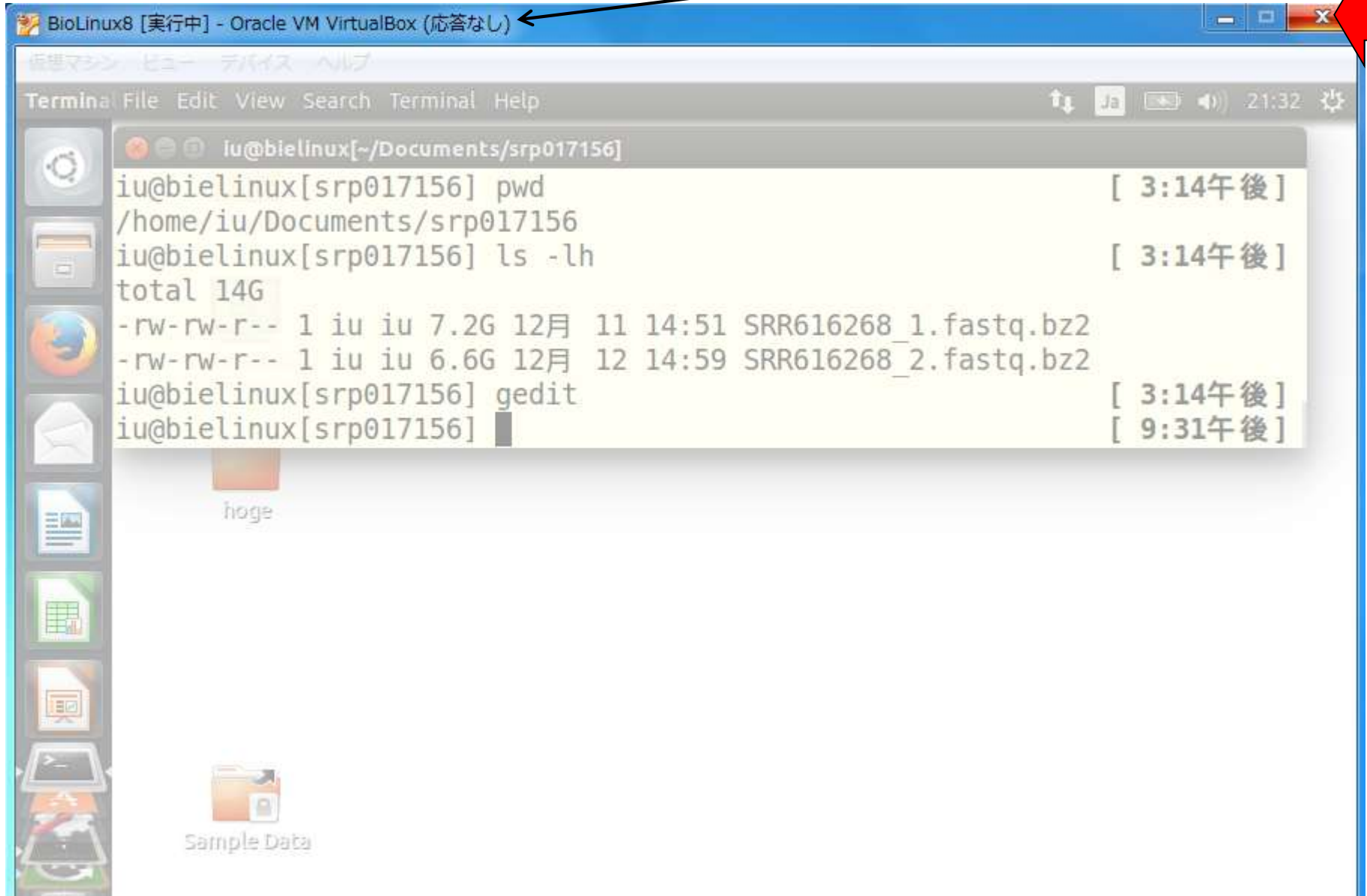
W2-1

Saveボタンを押した後の状態。①ファイル名が `hoge_JSLAB3.sh` に変更されていることが分かる。②ファイルを保存しただけなので、まだ `gedit` は起動中のまま。そのため、コマンドプロンプトは出ておらず、コマンド入力待ち状態ではない。③ × ボタンを押して `gedit` を終了。

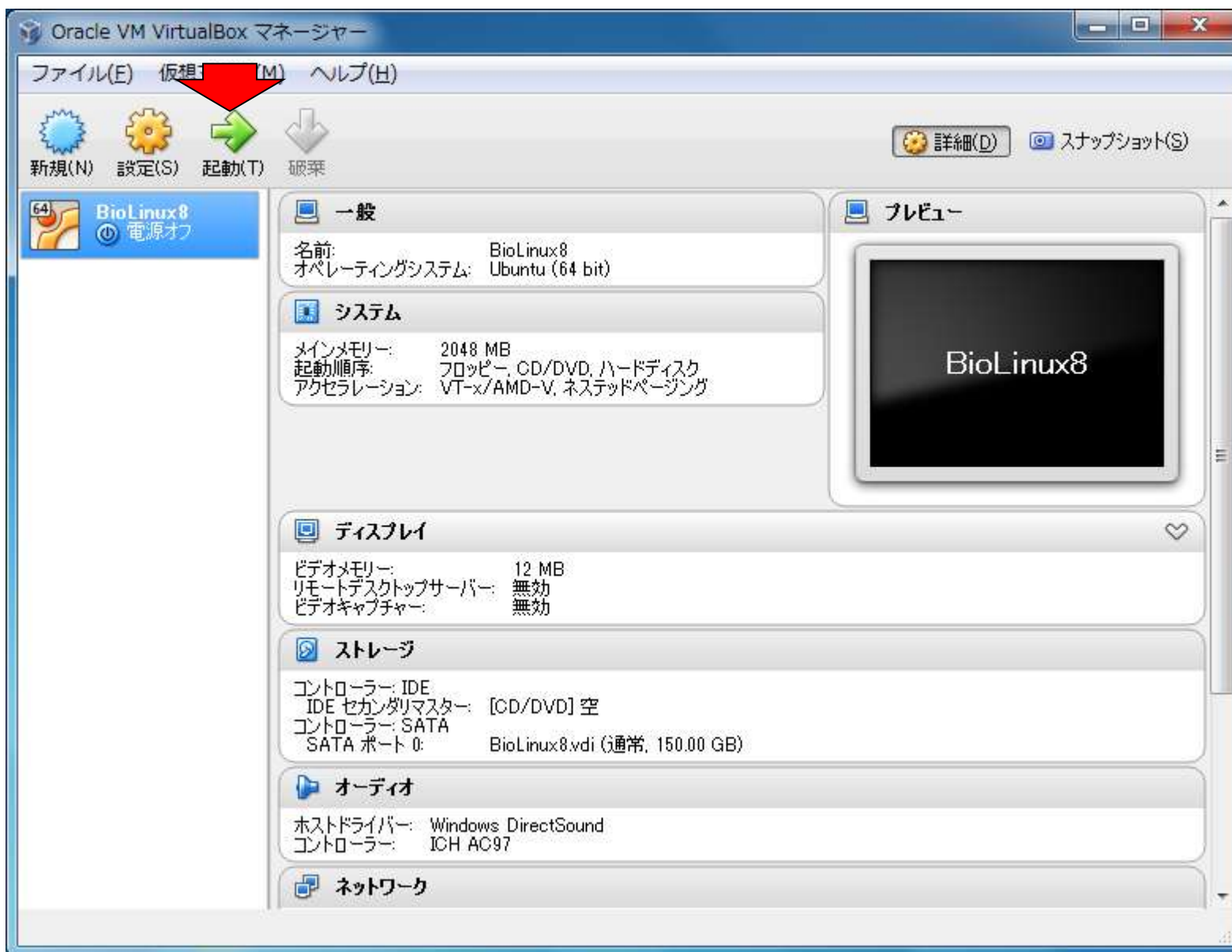


W2-1

「応答なし」という類のフリーズは頻繁に起こります(爆)。こういうときは、数分待って反応がなければ右上の×ボタンを押す以外の選択肢はありません。



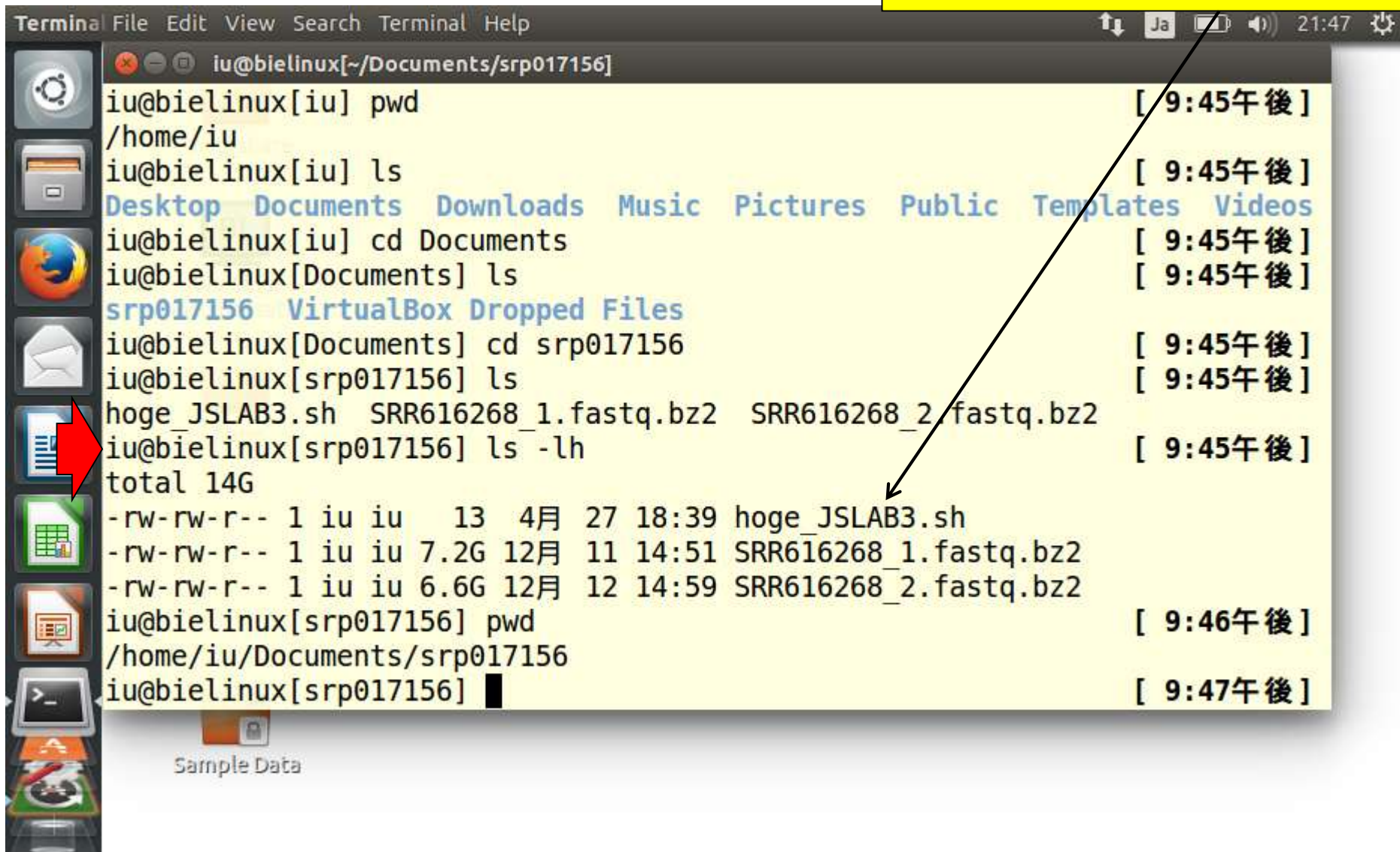
W2-2



W2-2: シェルスクリプト

pwdとlsを駆使して元のディレクトリに移動。
さきほどgeditで作成したhoge_JSLAB3.sh
がlsで見られることがわかる。

```
Terminal File Edit View Search Terminal Help
iu@bielinux[~/Documents/srp017156]
iu@bielinux[iu] pwd [ 9:45午後 ]
/home/iu
iu@bielinux[iu] ls [ 9:45午後 ]
Desktop Documents Downloads Music Pictures Public Templates Videos
iu@bielinux[iu] cd Documents [ 9:45午後 ]
iu@bielinux[Documents] ls [ 9:45午後 ]
srp017156 VirtualBox Dropped Files
iu@bielinux[Documents] cd srp017156 [ 9:45午後 ]
iu@bielinux[srp017156] ls [ 9:45午後 ]
hoge_JSLAB3.sh SRR616268_1.fastq.bz2 SRR616268_2.fastq.bz2
iu@bielinux[srp017156] ls -lh [ 9:45午後 ]
total 14G
-rw-rw-r-- 1 iu iu 13 4月 27 18:39 hoge_JSLAB3.sh
-rw-rw-r-- 1 iu iu 7.2G 12月 11 14:51 SRR616268_1.fastq.bz2
-rw-rw-r-- 1 iu iu 6.6G 12月 12 14:59 SRR616268_2.fastq.bz2
iu@bielinux[srp017156] pwd [ 9:46午後 ]
/home/iu/Documents/srp017156
iu@bielinux[srp017156] [ 9:47午後 ]
```



W2-2: シェルスクリプト

hoge_JSLAB3.shの中身を①moreコマンドや②lessコマンドで確認しているだけ。lessコマンドから抜けるにはqキーを押す。③hoge_JSLAB3.sh中に書き込んだコマンド群(pwd, date, ls)を実行するための最も簡単なやり方は、shコマンドの利用。赤枠内が実行結果。確かにファイル中のコマンド群と対応している。

```
BioLinux8 [実行中] - Oracle VM VirtualBox
仮想マシン ビュー デバイス ヘルプ
Terminal
iu@bielinux[~/Documents/srp017156]
hoge_JSLAB3.sh SRR616268_1.fastq.bz2 SRR616268_2.fastq.bz2
iu@bielinux[srp017156] ls -lh
total 14G
-rw-rw-r-- 1 iu iu 13 4月 27 18:39 hoge_JSLAB3.sh
-rw-rw-r-- 1 iu iu 7.2G 12月 11 14:51 SRR616268_1.fastq.bz2
-rw-rw-r-- 1 iu iu 6.6G 12月 12 14:59 SRR616268_2.fastq.bz2
iu@bielinux[srp017156] pwd [10:01午後]
/home/iu/Documents/srp017156
iu@bielinux[srp017156] more hoge_JSLAB3.sh [10:01午後]
pwd
date
ls
iu@bielinux[srp017156] less hoge_JSLAB3.sh [10:07午後]
iu@bielinux[srp017156] sh hoge_JSLAB3.sh [10:07午後]
/home/iu/Documents/srp017156
2015年 4月 27日 月曜日 22:08:10 JST
hoge_JSLAB3.sh SRR616268_1.fastq.bz2 SRR616268_2.fastq.bz2
iu@bielinux[srp017156] [10:11午後]
```



W2-3: シバン

(ヒトにもよるが) 著者らは、shコマンドを普段利用しない。代わりにhoge_JSLAB3.shファイルがシェルスクリプトであることをファイル中の最初の一行目に明示し、ファイルそのものを実行コマンドのようにして取り扱う。このあたりは何故?と疑問に思うよりは、そういう人達(宗派)も存在するという程度の理解でよい。「gedit ファイル名」で任意のファイルを開くことができる。

The screenshot shows a virtual machine environment with a terminal window and a gedit editor window. The terminal window displays the following commands and output:

```
iu@bielinux[~/Documents/srp017156] pwd
/home/iu/Documents/srp017156
iu@bielinux[srp017156] ls -lh
total 14G
-rw-rw-r-- 1 iu iu 13 4月 27 22:43 hoge_JSLAB3.sh
-rw-rw-r-- 1 iu iu 7.2G 12月 11 14:51 SRR616268_1.fastq.bz2
-rw-rw-r-- 1 iu iu 6.6G 12月 12 14:59 SRR616268_2.fastq.bz2
iu@bielinux[srp017156] gedit hoge_JSLAB3.sh
```

The gedit editor window shows the following content:

```
pwd
date|
ls
```

The terminal window also shows the time [10:43午後] and [10:44午後]. A red arrow points to the terminal window, and a black arrow points to the gedit editor window.

W2-3: シバン

シェルスクリプトの場合は、「#!/bin/sh」と書く。ちなみにPerlプログラムの場合は、「#!/usr/bin/perl」や(システムによっては)「#!/usr/local/bin/perl」と書く。この最初の「#!」は共通であり、シバン(shebang)と呼ぶ。「#!/bin/sh」は、使い方によってはなくても動くので、動かなければつけるという程度の理解でもいいだろう。尚、「#!/bin/bash」などと書くことで動くこともあるので、こうやれば動くという経験を積んでおくとよい。W9-9でも触れる。

The screenshot shows a virtual machine window titled "BioLinux8 [実行中] - Oracle VM VirtualBox". Inside, there is a terminal window and a gedit editor window.

Terminal Window:

```
iu@bielinux[~/Documents/srp017156]
iu@bielinux[srp017156] pwd
/home/iu/Documents/srp017156
iu@bielinux[srp017156] ls -lh
total 14G
-rw-rw-r-- 1 iu iu 13 4月 27 22:43 hoge_JSLAB3.sh
-rw-rw-r-- 1 iu iu 7.2G 12月 11 14:51 SRR616268_1.fastq.bz2
-rw-rw-r-- 1 iu iu 6.6G 12月 12 14:59 SRR616268_2.fastq.bz2
iu@bielinux[srp017156] gedit hoge_JSLAB3.sh [10:44午後]
```

gedit Editor Window:

The gedit window is titled "*hoge_JSLAB3.sh (~/.Documents/srp017156) - gedit". It shows the following content:

```
#!/bin/sh
pwd
date
ls
```

Red arrows with numbers 1 and 2 point to the first line of the script and the Save button in the gedit window, respectively.

W3-1: Permission

geditで編集後に、moreで確認しているだけです。これで基本的にはOKだが、もうひと手間必要。シェルスクリプトに限らず、NGS解析用プログラムなども赤枠部分の実行権限(パーミッション; permission)を適切に変更してやる必要がある。連載第3回のW9-2-5と9-2-6にも言及あり。

The screenshot shows a terminal window titled "BioLinux8 [実行中] - Oracle VM VirtualBox". The terminal output is as follows:

```
iu@bielinux[~/Documents/srp017156]
iu@bielinux[srp017156] ls -lh [10:43午後]
total 14G
-rw-rw-r-- 1 iu iu 13 4月 27 22:43 hoge_JSLAB3.sh
-rw-rw-r-- 1 iu iu 7.2G 12月 11 14:51 SRR616268_1.fastq.bz2
-rw-rw-r-- 1 iu iu 6.6G 12月 12 14:59 SRR616268_2.fastq.bz2
iu@bielinux[srp017156] gedit hoge_JSLAB3.sh [10:44午後]
iu@bielinux[srp017156] more hoge_JSLAB3.sh [11:06午後]
#!/bin/sh
pwd
date
ls
iu@bielinux[srp017156] [11:06午後]
```

A red arrow points to the permissions `-rw-rw-r--` in the first line of the `ls -lh` output. Below the terminal, a folder icon labeled "Sample Data" is visible. The system tray at the bottom right includes a "Right Control" button.

W3-1 : Permission

実行権限が与えられている状態でうまくいくコマンドは「./hoge_JSLAB3.sh」。まずは、このファイルに対して実行権限が与えられていない状態でエラーメッセージを出してみる。①もし実行権限が与えられていたら、「./hoge」まで打ち込んでからTabキーを押してタブ補完できるはずだができない。慣れればこの段階でPermissionを変更する必要があることに気づく。ここではうまくいかないことが分かっている状態だが、めげずに手入力で「./hoge_JSLAB3.sh」と打ち込んでリターンキーを押す。

The screenshot shows a terminal window titled "BioLinux8 [実行中] - Oracle VM VirtualBox". The terminal prompt is "iu@bielinux[~/Documents/srp017156]". The user has run the command "ls" and "ls -lh", showing the following output:

```
iu@bielinux[ srp017156 ] ls
hoge_JSLAB3.sh  SRR616268_1.fastq.bz2
iu@bielinux[ srp017156 ] ls -lh
total 14G
-rw-rw-r-- 1 iu iu   23  4月 27 23:06 hoge_JSLAB3.sh
-rw-rw-r-- 1 iu iu  7.2G 12月 11 14:51 SRR616268_1.fastq.bz2
-rw-rw-r-- 1 iu iu  6.6G 12月 12 14:59 SRR616268_2.fastq.bz2
```

The user is now typing the command "iu@bielinux[srp017156] ./hoge" and pressing the Tab key, which is indicated by a red arrow and the number "1" in a red circle. The terminal shows the command being typed as "iu@bielinux[srp017156] ./hoge" with a cursor at the end. The terminal also shows a timestamp "[8:39午前]".

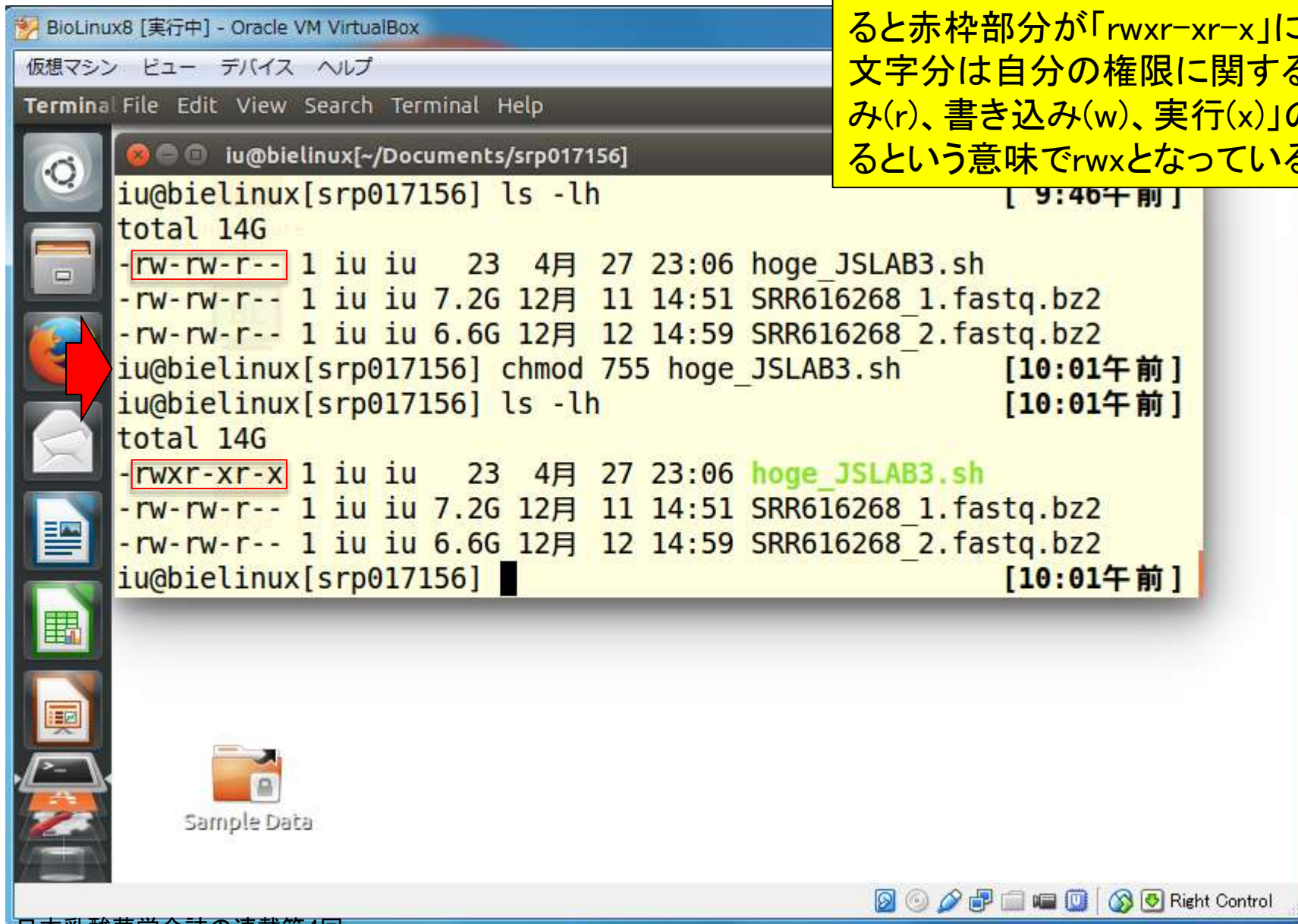
W3-1 : Permission

①「./hoge_JSLAB3.sh」と打ち込んでリターンキーを押すと、「実行権限が与えられていない(permission denied)」と文句を言われる。②「./」が無駄なのだろうと思い、ファイル名のみを指定してもダメ(このあたりはキーボードの上下キーを上手に利用せよ)。「そんなコマンドはない(command not found)」と文句をいわれている。③「/home/iu/Documents/srp017156/hoge_JSLAB3.sh」とやってみても、①と同じ結果になる。

```
BioLinux8 [実行中] - Oracle VM VirtualBox
仮想マシン ビュー デバイス ヘルプ
Terminal
iu@bielinux[~/Documents/srp017156]
iu@bielinux[srp017156] ls -lh
total 14G
-rw-rw-r-- 1 iu iu 23 4月 27 23:06 hoge_JSLAB3.sh
-rw-rw-r-- 1 iu iu 7.2G 12月 11 14:51 SRR616268_1.fastq.bz2
-rw-rw-r-- 1 iu iu 6.6G 12月 12 14:59 SRR616268_2.fastq.bz2
① iu@bielinux[srp017156] ./hoge_JSLAB3.sh [ 8:39午前 ]
zsh: permission denied: ./hoge_JSLAB3.sh
② iu@bielinux[srp017156] hoge_JSLAB3.sh [ 9:44午前 ]
zsh: command not found: hoge_JSLAB3.sh
iu@bielinux[srp017156] pwd [ 9:45午前 ]
/home/iu/Documents/srp017156
③ iu@bielinux[srp017156] /home/iu/Documents/srp017156/hoge_JSLAB3.s
h
zsh: permission denied: /home/iu/Documents/srp017156/hoge_JSLAB3.
sh
iu@bielinux[srp017156] [ 9:46午前 ]
```


W3-1: Permission

パーミッションの変更はchmodコマンドを利用。連載第3回ウェブ資料のMac版ではW9-2-6で説明。chmodコマンドのオプションで「755」とすると赤枠部分が「rwxr-xr-x」になる。最初の3文字分は自分の権限に関する部分。「読み込み(r)、書き込み(w)、実行(x)」の全てを許可するという意味でrwxとなっている。

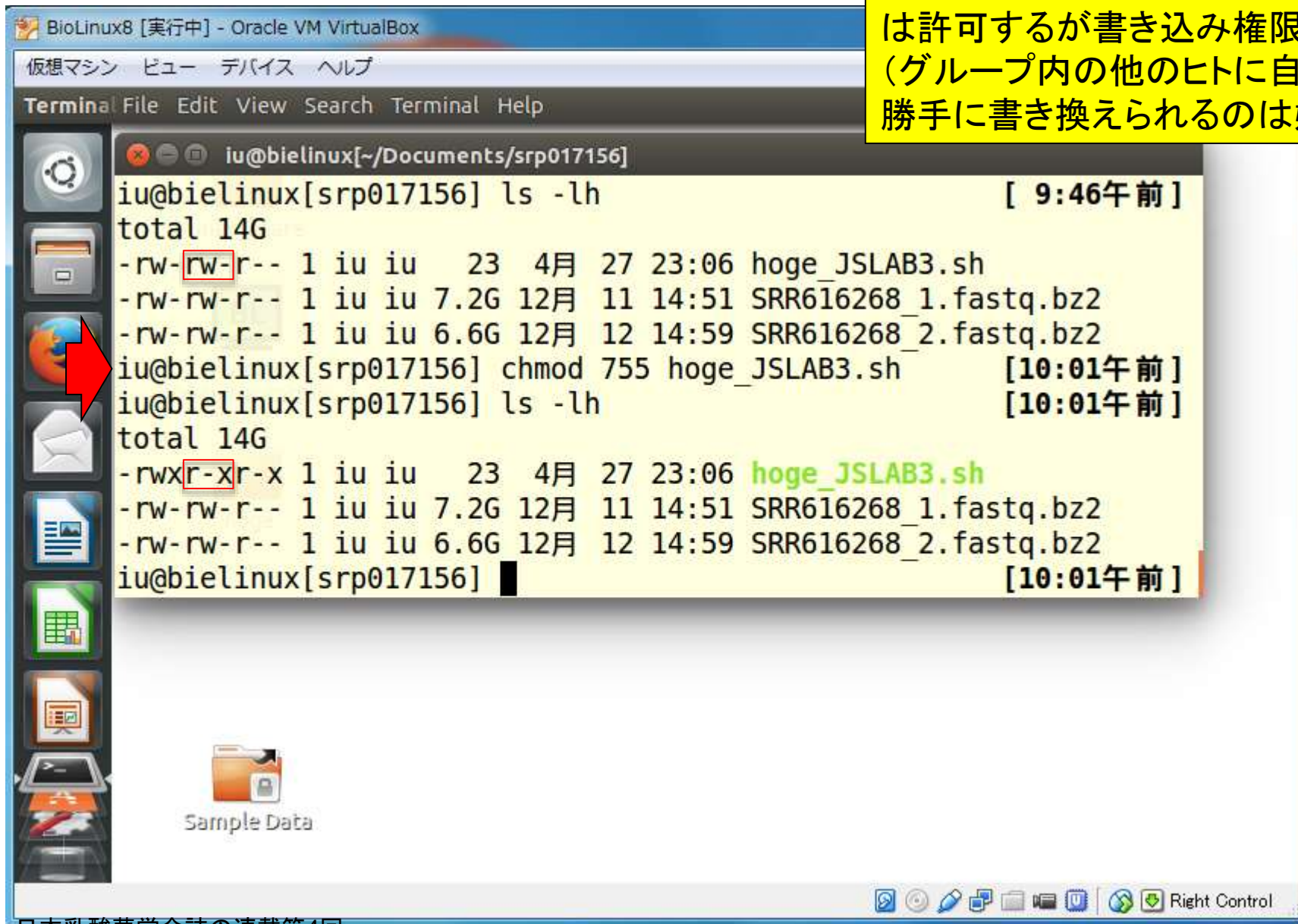


```
BioLinux8 [実行中] - Oracle VM VirtualBox
仮想マシン ビュー デバイス ヘルプ
Terminal File Edit View Search Terminal Help

iu@bielinux[~/Documents/srp017156]
iu@bielinux[srp017156] ls -lh
total 14G
-rw-rw-r-- 1 iu iu 23 4月 27 23:06 hoge_JSLAB3.sh
-rw-rw-r-- 1 iu iu 7.2G 12月 11 14:51 SRR616268_1.fastq.bz2
-rw-rw-r-- 1 iu iu 6.6G 12月 12 14:59 SRR616268_2.fastq.bz2
iu@bielinux[srp017156] chmod 755 hoge_JSLAB3.sh
iu@bielinux[srp017156] ls -lh
total 14G
-rwxr-xr-x 1 iu iu 23 4月 27 23:06 hoge_JSLAB3.sh
-rw-rw-r-- 1 iu iu 7.2G 12月 11 14:51 SRR616268_1.fastq.bz2
-rw-rw-r-- 1 iu iu 6.6G 12月 12 14:59 SRR616268_2.fastq.bz2
iu@bielinux[srp017156]
```

W3-1: Permission

赤枠部分の真ん中のr-xは、自分も属するグループ(研究室や研究プロジェクト単位など)の他のヒトに対する権限。読み込みと実行は許可するが書き込み権限は与えていない(グループ内の他のヒトに自分のファイルを勝手に書き換えられるのは嫌、ということ)。



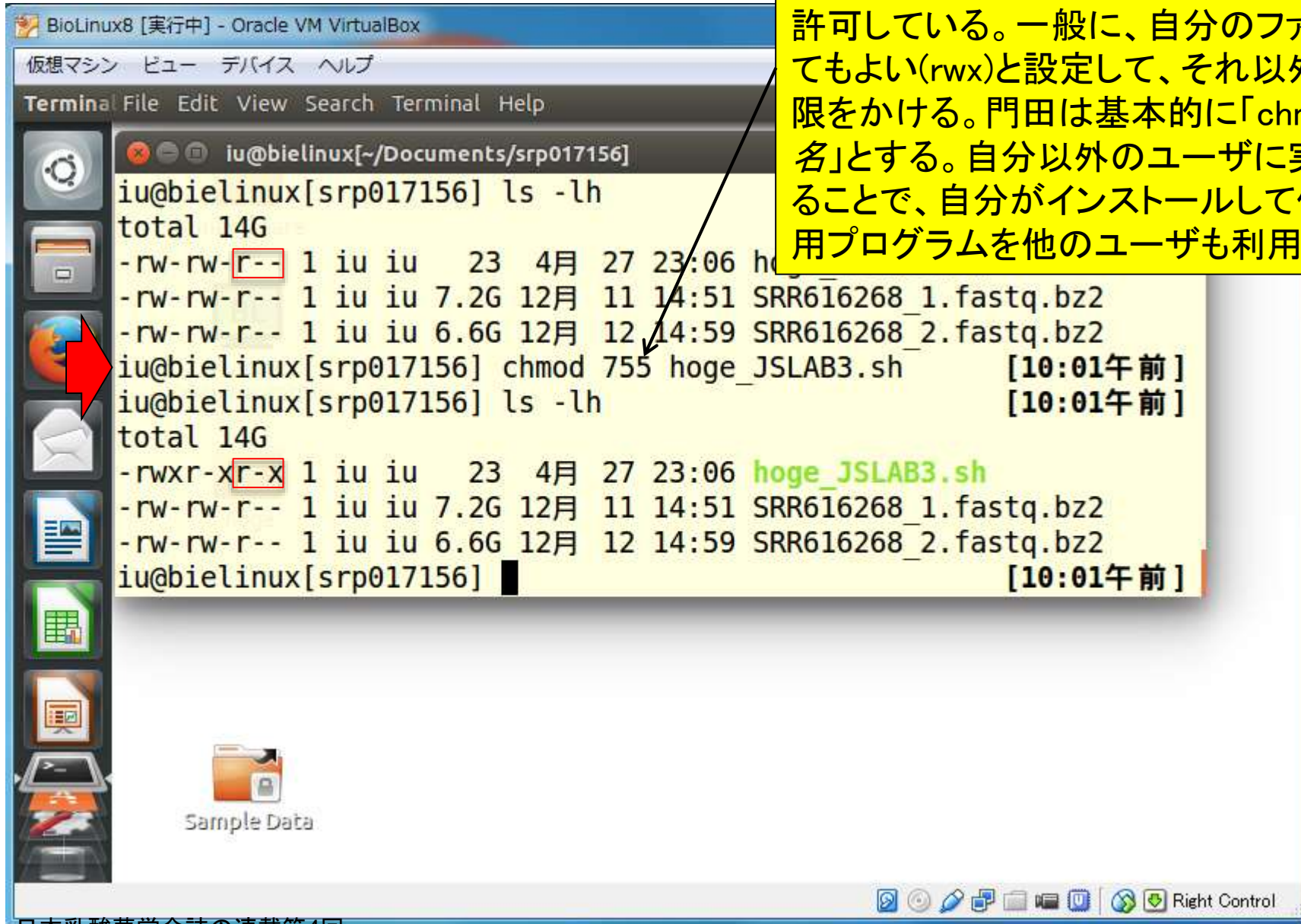
The screenshot shows a terminal window in a virtual machine. The user 'iu' is in the directory ~/Documents/srp017156. The terminal output shows the following sequence of commands and results:

```
iu@bielinux[~/Documents/srp017156] ls -lh [ 9:46午前 ]
total 14G
-rw-rw-r-- 1 iu iu 23 4月 27 23:06 hoge_JSLAB3.sh
-rw-rw-r-- 1 iu iu 7.2G 12月 11 14:51 SRR616268_1.fastq.bz2
-rw-rw-r-- 1 iu iu 6.6G 12月 12 14:59 SRR616268_2.fastq.bz2
iu@bielinux[srp017156] chmod 755 hoge_JSLAB3.sh [10:01午前]
iu@bielinux[srp017156] ls -lh [10:01午前]
total 14G
-rwxr-xr-x 1 iu iu 23 4月 27 23:06 hoge_JSLAB3.sh
-rw-rw-r-- 1 iu iu 7.2G 12月 11 14:51 SRR616268_1.fastq.bz2
-rw-rw-r-- 1 iu iu 6.6G 12月 12 14:59 SRR616268_2.fastq.bz2
iu@bielinux[srp017156] [10:01午前]
```

The permissions for 'hoge_JSLAB3.sh' are highlighted with a red box in the original image, showing the change from 'rw-rw-r--' to 'rwxr-xr-x'. A red arrow points to the 'r-x' part of the permissions.

W3-1: Permission

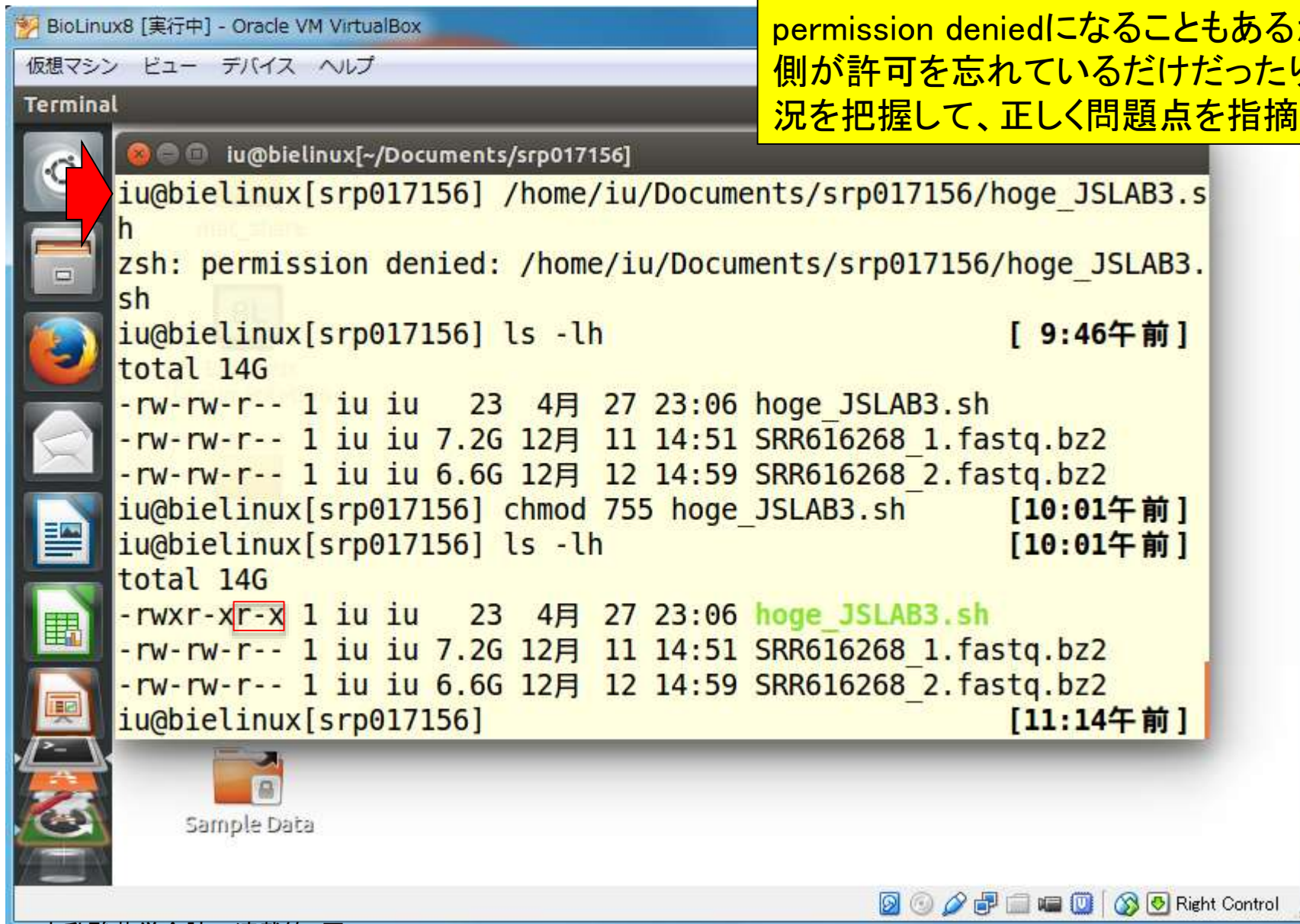
赤枠部分の一番右側のr-xは、自分が所属するグループ以外の全てのユーザに対する権限設定。ここではグループ内と同じく、読み込みと実行のみ許可している。一般に、自分のファイルは何をやってもよい(rwx)と設定して、それ以外には一定の制限をかける。門田は基本的に「chmod 755 ファイル名」とする。自分以外のユーザに実行権限を与えることで、自分がインストールして使っているNGS用プログラムを他のユーザも利用可能となる。



```
BioLinux8 [実行中] - Oracle VM VirtualBox
仮想マシン ビュー デバイス ヘルプ
Terminal File Edit View Search Terminal Help
iu@bielinux[~/Documents/srp017156]
iu@bielinux[srp017156] ls -lh
total 14G
-rw-rw-r-- 1 iu iu 23 4月 27 23:06 hoge_JSLAB3.sh
-rw-rw-r-- 1 iu iu 7.2G 12月 11 14:51 SRR616268_1.fastq.bz2
-rw-rw-r-- 1 iu iu 6.6G 12月 12 14:59 SRR616268_2.fastq.bz2
iu@bielinux[srp017156] chmod 755 hoge_JSLAB3.sh [10:01午前]
iu@bielinux[srp017156] ls -lh [10:01午前]
total 14G
-rwxr-xr-x 1 iu iu 23 4月 27 23:06 hoge_JSLAB3.sh
-rw-rw-r-- 1 iu iu 7.2G 12月 11 14:51 SRR616268_1.fastq.bz2
-rw-rw-r-- 1 iu iu 6.6G 12月 12 14:59 SRR616268_2.fastq.bz2
iu@bielinux[srp017156] [10:01午前]
```

W3-1: Permission

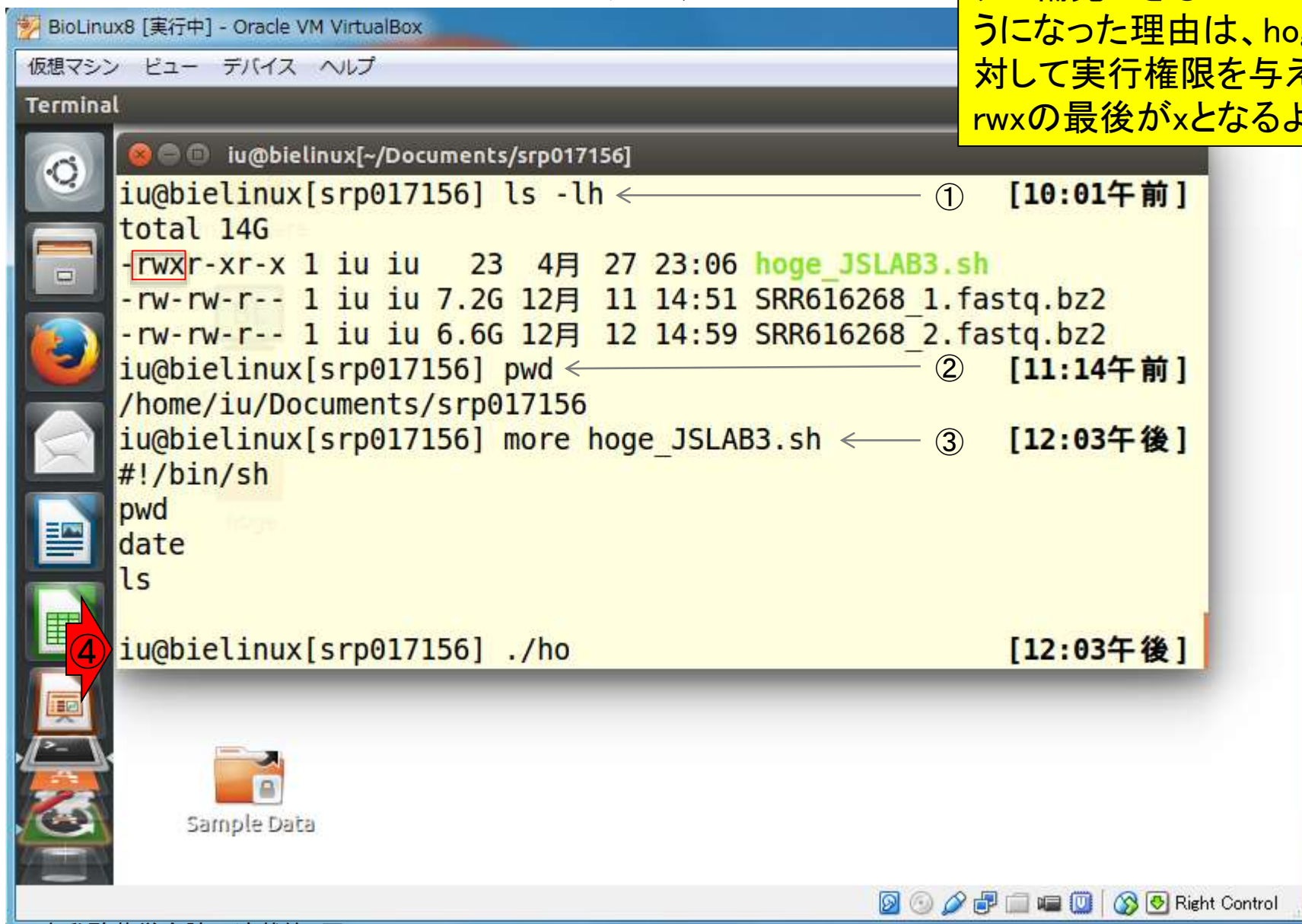
例えば、逆パターンとして、他のユーザから「赤矢印のように指定すればプログラムを実行できるよ」とアドバイスを受ける。で、実際にやってみると permission denied になることもあるが、単純に提供側が許可を忘れていただけだったりします。正しく状況を把握して、正しく問題点を指摘することが重要。



```
iu@bielinux[~/Documents/srp017156]
iu@bielinux[srp017156] /home/iu/Documents/srp017156/hoge_JSLAB3.s
h
zsh: permission denied: /home/iu/Documents/srp017156/hoge_JSLAB3.
sh
iu@bielinux[srp017156] ls -lh                                     [ 9:46午前 ]
total 14G
-rw-rw-r-- 1 iu iu   23  4月 27 23:06 hoge_JSLAB3.sh
-rw-rw-r-- 1 iu iu  7.2G 12月 11 14:51 SRR616268_1.fastq.bz2
-rw-rw-r-- 1 iu iu  6.6G 12月 12 14:59 SRR616268_2.fastq.bz2
iu@bielinux[srp017156] chmod 755 hoge_JSLAB3.sh                 [10:01午前 ]
iu@bielinux[srp017156] ls -lh                                     [10:01午前 ]
total 14G
-rwxr-xr-x 1 iu iu   23  4月 27 23:06 hoge_JSLAB3.sh
-rw-rw-r-- 1 iu iu  7.2G 12月 11 14:51 SRR616268_1.fastq.bz2
-rw-rw-r-- 1 iu iu  6.6G 12月 12 14:59 SRR616268_2.fastq.bz2
iu@bielinux[srp017156]                                          [11:14午前 ]
```

W3-2: シェルスクリプト

①～③は全貌のおさらい。④で「./ho」まで打ち込んでからTabキーを押すとタブ補完できることがわかる。できるようになった理由は、hoge_JSLAB3.shに対して実行権限を与えたから(赤枠rwxの最後がxとなるようにしたから)。



The screenshot shows a terminal window titled "BioLinux8 [実行中] - Oracle VM VirtualBox". The terminal output is as follows:

```
iu@bielinux[~/Documents/srp017156]
iu@bielinux[srp017156] ls -lh ← ① [10:01午前]
total 14G
-rwxr-xr-x 1 iu iu 23 4月 27 23:06 hoge_JSLAB3.sh
-rw-rw-r-- 1 iu iu 7.2G 12月 11 14:51 SRR616268_1.fastq.bz2
-rw-rw-r-- 1 iu iu 6.6G 12月 12 14:59 SRR616268_2.fastq.bz2
iu@bielinux[srp017156] pwd ← ② [11:14午前]
/home/iu/Documents/srp017156
iu@bielinux[srp017156] more hoge_JSLAB3.sh ← ③ [12:03午後]
#!/bin/sh
pwd
date
ls
iu@bielinux[srp017156] ./ho [12:03午後]
```

A red arrow points to the command `./ho` in the terminal, which is labeled with a circled number 4. The file `hoge_JSLAB3.sh` in the `ls` output is highlighted in green, and its permissions `-rwxr-xr-x` are enclosed in a red box.

W3-2: シェルスクリプト

タブ補完してリターンキーを押したあとの状態。③でファイルの中身を示しているが、正しくコマンド群(pwd, date, ls)を実行できていることがわかる。

```
iu@bielinux[~/Documents/srp017156]
-rw-rw-r-- 1 iu iu 6.6G 12月 12 14:59 SRR616268_2.fastq.bz2
iu@bielinux[srp017156] pwd ← ② [11:14午前]
/home/iu/Documents/srp017156
iu@bielinux[srp017156] more hoge_JSLAB3.sh ← ③ [12:03午後]
#!/bin/sh
pwd
date
ls

iu@bielinux[srp017156] ./hoge_JSLAB3.sh [12:03午後]
/home/iu/Documents/srp017156
2015年 4月 28日 火曜日 12:11:32 JST
hoge_JSLAB3.sh SRR616268 1.fastq.bz2 SRR616268 2.fastq.bz2
iu@bielinux[srp017156] [12:11午後]
```

W3-2: シェルスクリプト

①と③は同じ結果。②はhoge_JSLAB3.shに実行権限を与えてもだめだということを示している。必要最小限③のやり方で結果を得られるものの、教えるヒトによってやり方が異なるということは肝に銘じるべき(教わる側中心に世の中回っているわけではない!)。重要なのは様々なやり方を知り、対応する能力を身につけること。

```
date
ls

iu@bielinux[~/Documents/srp017156] ./hoge_JSLAB3.sh ← ① [12:03午後]
/home/iu/Documents/srp017156
2015年 4月 28日 火曜日 12:11:32 JST
hoge_JSLAB3.sh SRR616268 1.fastq.bz2 SRR616268 2.fastq.bz2

iu@bielinux[~/Documents/srp017156] hoge_JSLAB3.sh ← ② [12:11午後]
zsh: command not found: hoge_JSLAB3.sh

iu@bielinux[~/Documents/srp017156] sh hoge_JSLAB3.sh ← ③ [12:15午後]
/home/iu/Documents/srp017156
2015年 4月 28日 火曜日 12:15:52 JST
hoge_JSLAB3.sh SRR616268 1.fastq.bz2 SRR616268 2.fastq.bz2

iu@bielinux[~/Documents/srp017156] [12:15午後]
```

W3-3: その他

①小なり(<)記号をつけるやり方もある。これは、「hoge_JSLAB3.shファイルの中身をshコマンドで受けて実行する」みたいな捉え方でよい。入力ファイルの読み込み時に小なり記号を使う流派も一定数存在する。「sh hoge_JSLAB3.sh」と「sh < hoge_JSLAB3.sh」は実質的に同じものだという理解でよい。②実行結果をリダイレクト(>)でhoge1.txtというファイル名で保存。③moreで中身を確認。

```
iu@bielinux[~/Documents/srp017156]
① iu@bielinux[srp017156] sh < hoge_JSLAB3.sh
/home/iu/Documents/srp017156
2015年 4月 28日 火曜日 13:25:25 JST
hoge_JSLAB3.sh SRR616268 1.fastq.bz2 SRR616268 2.fastq.bz2
② iu@bielinux[srp017156] sh < hoge_JSLAB3.sh > hoge1.txt
iu@bielinux[srp017156] more hoge1.txt ← ③ [ 1:26午後 ]
/home/iu/Documents/srp017156
2015年 4月 28日 火曜日 13:26:17 JST
hoge1.txt
hoge_JSLAB3.sh
SRR616268 1.fastq.bz2
SRR616268 2.fastq.bz2
iu@bielinux[srp017156] [ 1:26午後 ]
```


W3-3: その他

大きな複数ファイルのダウンロードで1つだけ成功し、大部分で失敗したときなど、一時的に特定の行のみ実行したくない局面もある。そのような場合は、③目的行の先頭に#をつけてコメントアウトし、④実行すればよい。例では日付情報を返すdateコマンドを実行させないようにしている。

```
iu@bielinux[~/Documents/srp017156]
iu@bielinux[srp017156] pwd ← ① [ 3:30午後 ]
/home/iu/Documents/srp017156
iu@bielinux[srp017156] ls ← ② [ 3:30午後 ]
hoge1.txt          SRR616268_1.fastq.bz2
hoge_JSLAB3.sh    SRR616268_2.fastq.bz2
iu@bielinux[srp017156] more hoge_JSLAB3.sh ← ③ [ 3:30午後 ]
#!/bin/sh
pwd
#date
ls

iu@bielinux[srp017156] ./hoge_JSLAB3.sh ← ④ [ 3:31午後 ]
/home/iu/Documents/srp017156
hoge1.txt          SRR616268_1.fastq.bz2
hoge_JSLAB3.sh    SRR616268_2.fastq.bz2
iu@bielinux[srp017156] [ 3:31午後 ]
```

W4-1 : wget

NGS解析実務者が納得できるレベルのシェルスクリプトファイル例。ここでは①既存のファイル(hoge_JSLAB3.sh)をコピーしてJSLAB4_1.shを作成し、②geditで2行目以降を編集するやり方の最初のほうを例示。

The screenshot shows a Linux desktop environment with a terminal window and a gedit editor window. The terminal window displays the following commands and their outputs:

```
iu@bielinux[~/Documents/srp017156] ls [10:00午前]
hoge1.txt          SRR616268_1.fastq.bz2
hoge_JSLAB3.sh    SRR616268_2.fastq.bz2
iu@bielinux[srp017156] cp hoge_JSLAB3.sh JSLAB4_1.sh [10:01午前]
iu@bielinux[srp017156] gedit JSLAB4_1.sh [10:01午前]
```

The gedit editor window shows the contents of the file `JSLAB4_1.sh`:

```
#!/bin/sh
wget -c
```

Red arrows with numbers 1 and 2 point to the terminal output for the copy and gedit commands, respectively. The gedit window also has a red dashed box around the script content.

JSLAB4_1.shの完成版はwgetコマンドで以下のURLより取得可能:
http://www.iu.a.u-tokyo.ac.jp/~kadota/book/JSLAB4_1.sh

W4-1 : wget

The screenshot shows a Linux desktop environment with a terminal window and a gedit editor window. The terminal window displays the following commands and their outputs:

```
iu@bielinux[~/Documents/srp017156] ls [10:00午前]
hoge1.txt SRR616268_1.fastq.bz2
hoge_JSLAB3.sh SRR616268_2.fastq.bz2
iu@bielinux[srp017156] cp hoge_JSLAB3.sh JSLAB4_1.sh [10:01午前]
iu@bielinux[srp017156] gedit JSLAB4_1.sh [10:01午前]
```

The gedit editor window shows the file `JSLAB4_1.sh` with the following content:

```
#!/bin/sh
```

A dialog box is displayed over the gedit window, asking to save changes to the document `"JSLAB4_1.sh"` before closing. The dialog box contains the following text:

Save changes to document "JSLAB4_1.sh" before closing?
If you don't save, changes from the last 5 minutes will be permanently lost.

The dialog box has three buttons: `Close without Saving`, `Cancel`, and `Save`. A red arrow labeled "1" points to the gedit window title bar, and another red arrow labeled "2" points to the `Save` button.

W4-1 : wget

ホストOS(Windows)上にダウンロードしたファイル(JSLAB4_1.sh)をEmEditorというエディタで眺めているところ。

書籍 | 日本乳酸菌学会誌 | 第4回クオリティコントロールからカウント情報取得まで

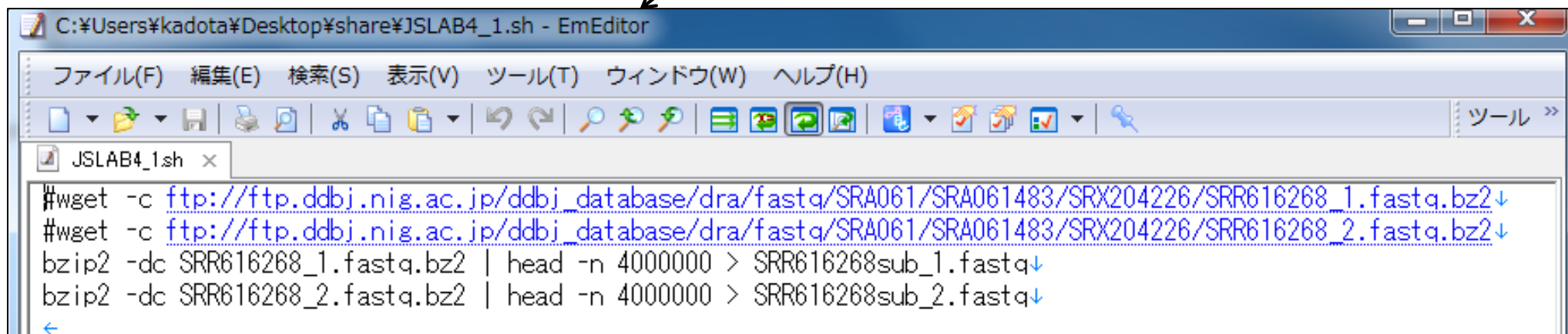
日本乳酸菌学会誌の第4回分です。ウェブ資料はどんどん追加中です。(2015.04.28)

- 原稿PDF
- [ウェブ資料PDF](#)(2015.04.28版; 約3MB)

シェルスクリプト:

- [JSLAB4_1.sh](#) (Windowsの改行コード; CR+LF)

```
wget -c http://www.iu.a.u-tokyo.ac.jp/~kadota/book/JSLAB4_1.sh
```

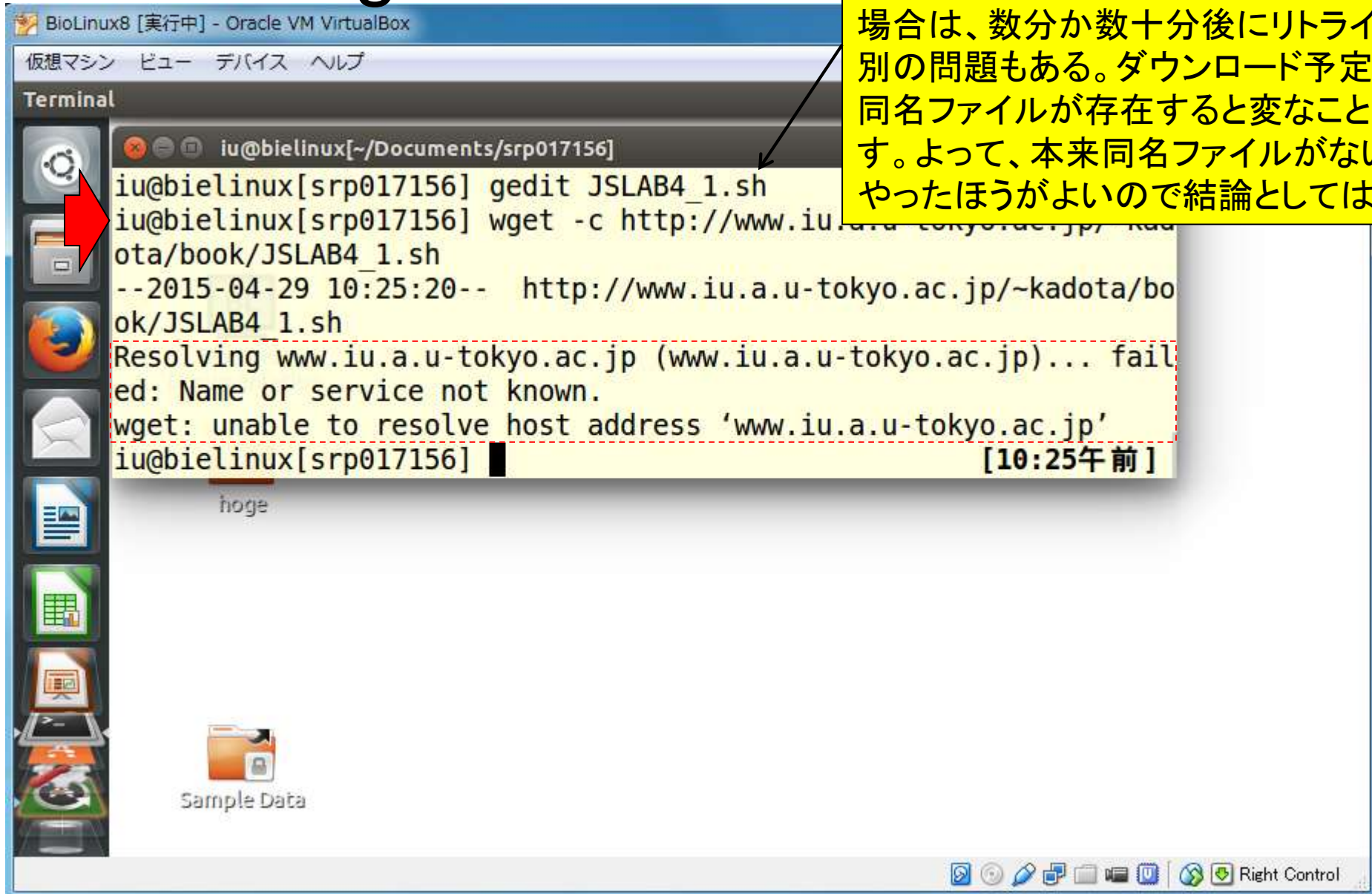


The screenshot shows the EmEditor window titled "C:\Users\kadota\Desktop\share\JSLAB4_1.sh - EmEditor". The menu bar includes "ファイル(F)", "編集(E)", "検索(S)", "表示(V)", "ツール(T)", "ウィンドウ(W)", and "ヘルプ(H)". The toolbar contains various icons for file operations and editing. The main text area shows the following script content:

```
#wget -c ftp://ftp.ddbj.nig.ac.jp/ddbj_database/dra/fastq/SRA061/SRA061483/SRX204226/SRR616268_1.fastq.bz2↓  
#wget -c ftp://ftp.ddbj.nig.ac.jp/ddbj_database/dra/fastq/SRA061/SRA061483/SRX204226/SRR616268_2.fastq.bz2↓  
bzip2 -dc SRR616268_1.fastq.bz2 | head -n 4000000 > SRR616268sub_1.fastq↓  
bzip2 -dc SRR616268_2.fastq.bz2 | head -n 4000000 > SRR616268sub_2.fastq↓
```

W4-1 : wget

「wget -c 任意のURL」で基本的にはダウンロード可能。しかし、赤枠のようにダウンロードに失敗することが頻繁にある。このような場合は、数分か数十分後にリトライ。また、別の問題もある。ダウンロード予定の場所に同名ファイルが存在すると変なことが起きます。よって、本来同名ファイルがない状態でやったほうがよいので結論としては非推奨。

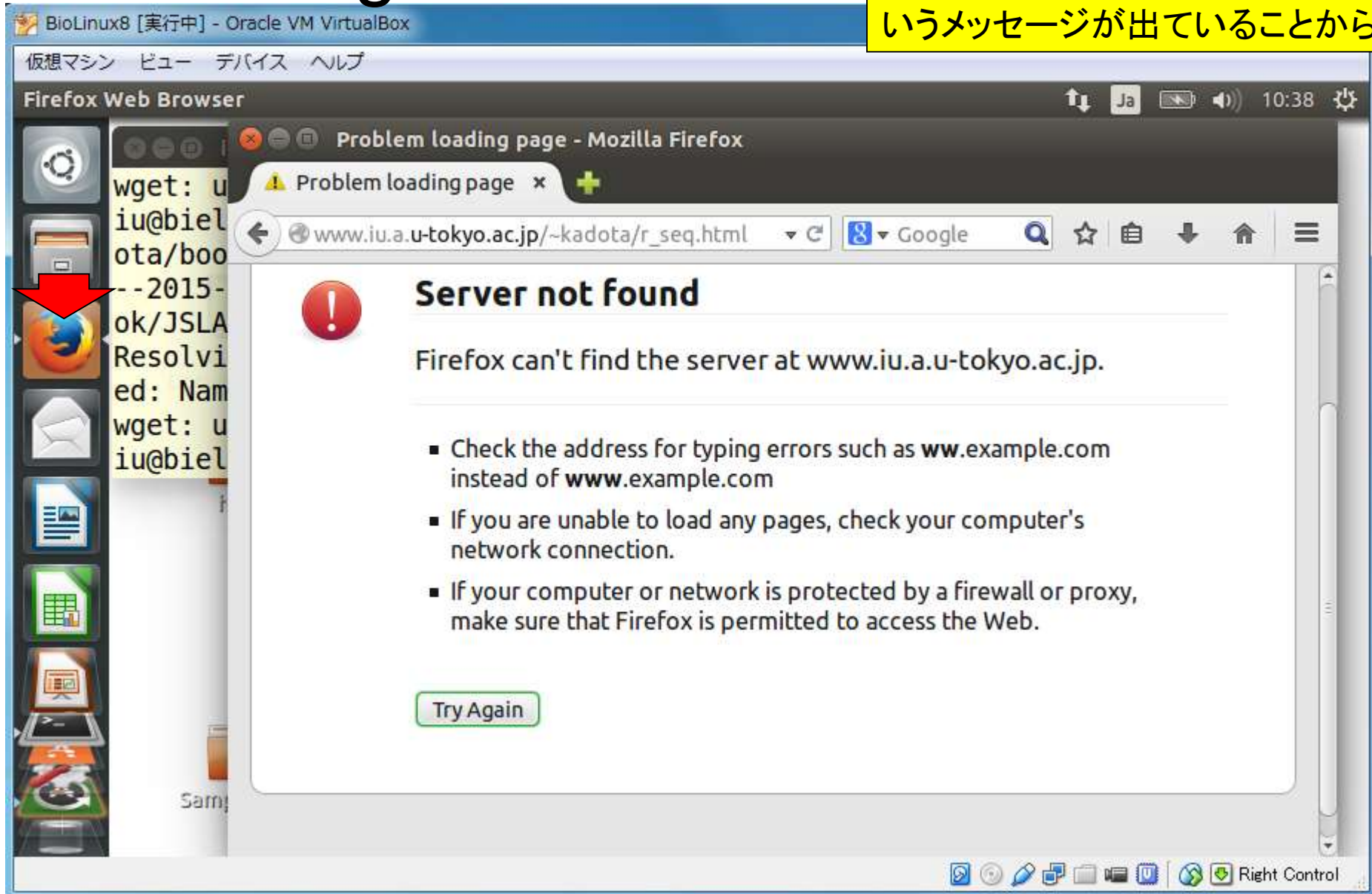


The screenshot shows a terminal window in a virtual machine. The user has run the command `wget -c http://www.iu.a.u-tokyo.ac.jp/~kadota/book/JSLAB4_1.sh`. The output shows the command starting at 10:25:20, but it fails to resolve the host address 'www.iu.a.u-tokyo.ac.jp' with the error message: "Resolving www.iu.a.u-tokyo.ac.jp (www.iu.a.u-tokyo.ac.jp)... failed: Name or service not known. wget: unable to resolve host address 'www.iu.a.u-tokyo.ac.jp'". A red dashed box highlights this error message. A red arrow points to the terminal window icon in the desktop environment.

```
iu@bielinux[~/Documents/srp017156]
iu@bielinux[srp017156] gedit JSLAB4_1.sh
iu@bielinux[srp017156] wget -c http://www.iu.a.u-tokyo.ac.jp/~kadota/book/JSLAB4_1.sh
--2015-04-29 10:25:20-- http://www.iu.a.u-tokyo.ac.jp/~kadota/book/JSLAB4_1.sh
Resolving www.iu.a.u-tokyo.ac.jp (www.iu.a.u-tokyo.ac.jp)... failed: Name or service not known.
wget: unable to resolve host address 'www.iu.a.u-tokyo.ac.jp'
iu@bielinux[srp017156] [10:25午前]
```

W4-1 : wget

wgetがうまく動かないときは、ゲストOS(BioLinux8)上でウェブブラウザFirefoxを起動してもインターネットに接続できないというメッセージが出ていることから納得。



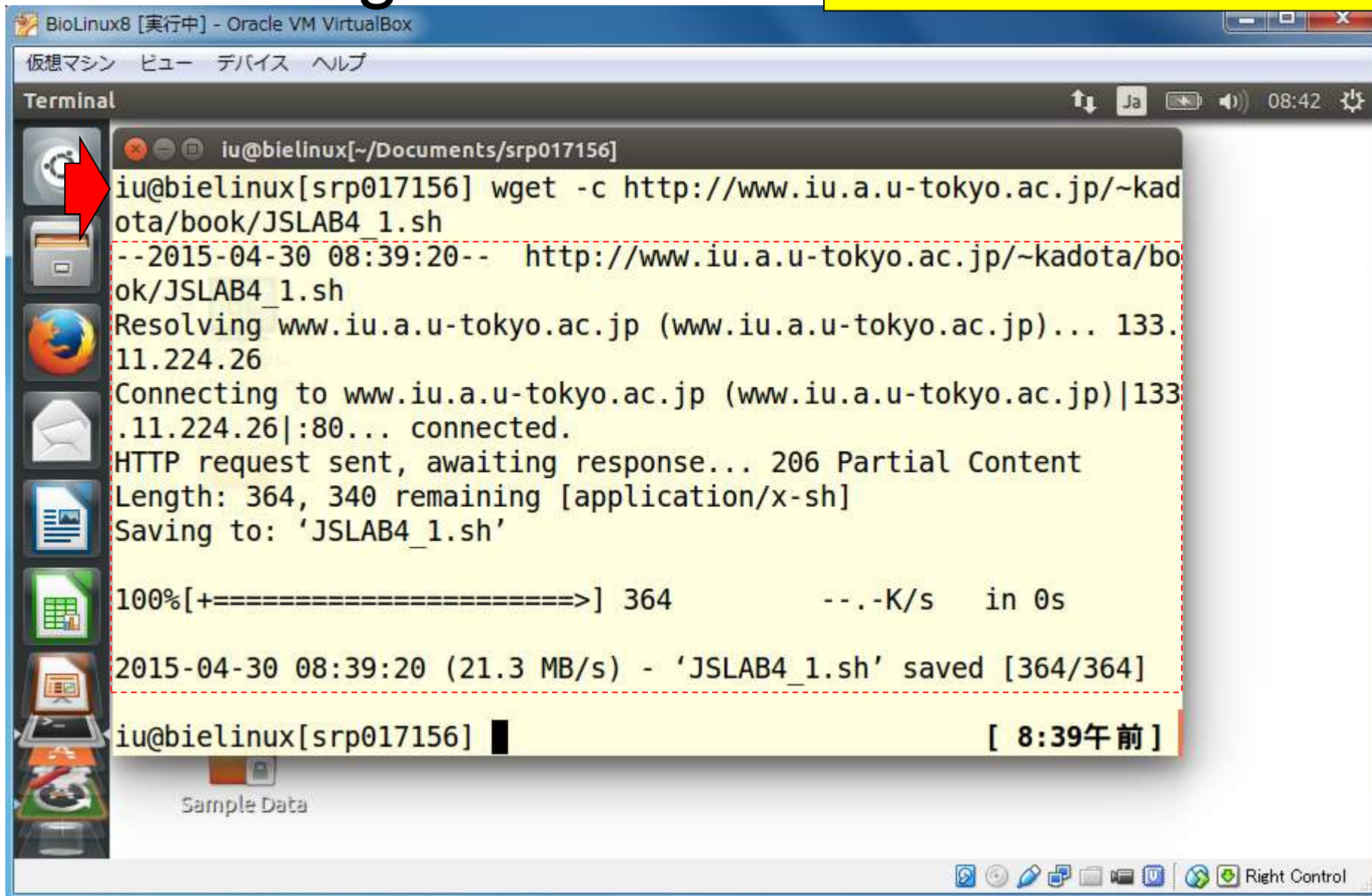
W4-2: wgetリトライ

翌日の起動時のメッセージ。ホストOS(Windows)は無線LANで問題なくネットを見られる状況なので、なぜゲストOS(BioLinux8)でできないのか不明だが、とにかくネットワーク接続できていないのだからしょうがないのだろうと一応納得。

The screenshot shows a terminal window in a BioLinux8 virtual machine. The terminal prompt is `iu@bielinux[~/Documents/srp017156]`. The user has entered the following commands: `pwd`, `#date`, and `ls`. Then, they entered `wget -c http://www.iu.a.u-tokyo.ac.jp/~kadota/book/JSLAB4_1.sh`. The output shows the command was executed on 2015-04-29 at 11:19:42, but it failed with the error: `Resolving www.iu.a.u-tokyo.ac.jp (www.iu.a.u-tokyo.ac.jp)... failed: Name or service not known. wget: unable to resolve host address 'www.iu.a.u-tokyo.ac.jp'`. The user then entered the same command again. A notification box titled "Network service discovery disabled" is overlaid on the terminal, stating: "Your current network has a .local domain, which is recommended and incompatible with the Avahi network service discovery. The service has been disabled." A red arrow points to the terminal window, and another arrow points from the yellow text box to the notification box.

W4-2: wgetリトライ

「Network service discovery disabled」のメッセージが消えて数分以内に、リトライした結果。うまくダウンロードできているようだ!



The screenshot shows a terminal window titled "Terminal" within a "BioLinux8 [実行中] - Oracle VM VirtualBox" environment. The terminal prompt is "iu@bielinux[~/Documents/srp017156]". The user has entered the command "wget -c http://www.iu.a.u-tokyo.ac.jp/~kadota/book/JSLAB4_1.sh". The output shows the file being resolved, connected to, and downloaded successfully. A red arrow points to the terminal window icon in the sidebar. The terminal output is as follows:

```
iu@bielinux[~/Documents/srp017156]
iu@bielinux[srp017156] wget -c http://www.iu.a.u-tokyo.ac.jp/~kadota/book/JSLAB4_1.sh
--2015-04-30 08:39:20-- http://www.iu.a.u-tokyo.ac.jp/~kadota/book/JSLAB4_1.sh
Resolving www.iu.a.u-tokyo.ac.jp (www.iu.a.u-tokyo.ac.jp)... 133.11.224.26
Connecting to www.iu.a.u-tokyo.ac.jp (www.iu.a.u-tokyo.ac.jp)|133.11.224.26|:80... connected.
HTTP request sent, awaiting response... 206 Partial Content
Length: 364, 340 remaining [application/x-sh]
Saving to: 'JSLAB4_1.sh'

100%[+=====>] 364          ---K/s   in 0s

2015-04-30 08:39:20 (21.3 MB/s) - 'JSLAB4_1.sh' saved [364/364]

iu@bielinux[srp017156] [ 8:39午前]
```


W4-2: wgetリトライ

赤枠の中が現在のJSLAB4_1.shの中身。そのうちの青枠分がwgetコマンド実行前の「hoge_JALAB3.shからのコピーで作成したJSLAB4_1.sh」の中身。青枠以外の部分が「http://www.iu.a.u-tokyo.ac.jp/~kadota/book/JSLAB4_1.sh」の中身の全て...ではなく一部(なので変)

```
iu@bielinux[srp017156] ls -la
total 14335148
drwxrwxr-x 2 iu iu          4096  4月 30 08:45 .
drwxr-xr-x 4 iu iu          4096 12月 10 17:55 ..
-rw-rw-r-- 1 iu iu           141  4月 28 13:26 hoge1.txt
-rwxr-xr-x 1 iu iu            24  4月 28 15:48 hoge_JSLAB3.sh
-rwxr-xr-x 1 iu iu           364  4月 28 17:49 JSLAB4_1.sh
-rw-rw-r-- 1 iu iu 7662128101 12月 11 14:51 SRR616268_1.fastq.bz2
-rw-rw-r-- 1 iu iu 7017031734 12月 12 14:59 SRR616268_2.fastq.bz2
iu@bielinux[srp017156] more JSLAB4_1.sh [10:35午前]
#!/bin/sh
pwd
#date
ls
nig.ac.jp/ddbj_database/dra/fastq/SRA061/SRA061483/SRX204226/SRR616268_1.fastq.bz2
#wget -c ftp://ftp.ddbj.nig.ac.jp/ddbj_database/dra/fastq/SRA061/SRA061483/SRX204226/SRR616268_2.fastq.bz2
bzip2 -dc SRR616268_1.fastq.bz2 | head -n 4000000 > SRR616268sub_1.fastq
bzip2 -dc SRR616268_2.fastq.bz2 | head -n 4000000 > SRR616268sub_2.fastq
iu@bielinux[srp017156] [10:35午前]
```

W4-3: wgetリトライ2

一旦JSLAB4_1.shを削除。②rmコマンド実行時に-fオプションをつけていないので「rm: remove regular file 'JSLAB4_1.sh'?」と聞かれている。Yesのyを入力してリターン。尚、デフォルトで本当に削除するかを聞いてくるのは、Bio-Linuxを含む少数派のようです。つまり、システムによっては「-fをつけなければ必ず聞いてくる」とは限らないので注意されたい。

```
File Edit View Search Terminal Help
① iu@bielinux[srp017156] ls -la
total 14335148
drwxrwxr-x 2 iu iu          4096  4月 30 08:55 .
drwxr-xr-x 4 iu iu          4096 12月 10 17:55 ..
-rw-rw-r-- 1 iu iu           141  4月 28 13:26 hoge1.txt
-rwxr-xr-x 1 iu iu            24  4月 28 15:48 hoge_JSLAB3.sh
-rwxr-xr-x 1 iu iu           364  4月 28 17:49 JSLAB4_1.sh
-rw-rw-r-- 1 iu iu 7662128101 12月 11 14:51 SRR616268_1.fastq.bz2
-rw-rw-r-- 1 iu iu 7017031734 12月 12 14:59 SRR616268_2.fastq.bz2
② iu@bielinux[srp017156] rm JSLAB4_1.sh
rm: remove regular file 'JSLAB4_1.sh'? y
③ iu@bielinux[srp017156] ls -la
total 14335144
drwxrwxr-x 2 iu iu          4096  4月 30 10:50 .
drwxr-xr-x 4 iu iu          4096 12月 10 17:55 ..
-rw-rw-r-- 1 iu iu           141  4月 28 13:26 hoge1.txt
-rwxr-xr-x 1 iu iu            24  4月 28 15:48 hoge_JSLAB3.sh
-rw-rw-r-- 1 iu iu 7662128101 12月 11 14:51 SRR616268_1.fastq.bz2
-rw-rw-r-- 1 iu iu 7017031734 12月 12 14:59 SRR616268_2.fastq.bz2
iu@bielinux[srp017156] █
```

[10:49午前]

[10:50午前]

[10:50午前]

W4-3: wgetリトライ2

JSLAB4_1.shがカレントディレクトリにない状態で①再度wgetを実行。②lsで存在確認。③moreでファイルの中身を確認。

```
iu@bielinux[srp017156] wget -c http://www.iu.a.u-tokyo.ac.jp/~kadota/book/JSLAB4_1.sh
--2015-04-30 11:00:24-- http://www.iu.a.u-tokyo.ac.jp/~kadota/book/JSLAB4_1.sh
Resolving www.iu.a.u-tokyo.ac.jp (www.iu.a.u-tokyo.ac.jp)... 133.11.224.26
Connecting to www.iu.a.u-tokyo.ac.jp (www.iu.a.u-tokyo.ac.jp)|133.11.224.26|:80..
. connected.
HTTP request sent, awaiting response... 200 OK
Length: 364 [application/x-sh]
Saving to: 'JSLAB4_1.sh'

100%[=====>] 364      --.-K/s   in 0s

2015-04-30 11:00:24 (67.9 MB/s) - 'JSLAB4_1.sh' saved [364/364]

iu@bielinux[srp017156] ls
hogel.txt      JSLAB4_1.sh      SRR616268_2.fastq.bz2
hoge_JSLAB3.sh SRR616268_1.fastq.bz2

iu@bielinux[srp017156] more JSLAB4_1.sh
#wget -c ftp://ftp.ddbj.nig.ac.jp/ddbj_database/dra/fastq/SRA061/SRA061483/SRX204226/SRR616268_1.fastq.bz2
#wget -c ftp://ftp.ddbj.nig.ac.jp/ddbj_database/dra/fastq/SRA061/SRA061483/SRX204226/SRR616268_2.fastq.bz2
bzip2 -dc SRR616268_1.fastq.bz2 | head -n 4000000 > SRR616268sub_1.fastq
bzip2 -dc SRR616268_2.fastq.bz2 | head -n 4000000 > SRR616268sub_2.fastq
iu@bielinux[srp017156]
```

W4-4: 共有フォルダ

ゲストOS (BioLinux8)上でも忘れたところに接続が復旧されている場合が多いが、別の手段としてホストOS (WindowsやMacintosh)のウェブブラウザ上で目的のファイル(JSLAB4_1.sh)をダウンロードし、共有フォルダ(ホストOSのデスクトップにあるshareフォルダ)に置くのが手っ取り早い。

(Rで)塩基配列解析

~NGS, RNA-seq, ゲノム, トランスクリプトーム, 正規化, 発現変動, 統計, モデル, バイオインフォマティクス~
(last modified 2015/04/28, since 2010)

What's new?

- このウェブページは[インストール](#)についての推奨手順 ([Windows2015.04.04版](#)と[Macintosh2015.04.03版](#))に従ってフリーソフトRと必要なパッケージをインストール済みであるという前提で記述しています。初心者の方は[基本的な利用法](#) ([Windows2015.04.03版](#)と [Macintosh2015.04.03版](#))で自習してください。本ウェブページを体系的にまとめた[書籍](#)もあります。(2015/04/03) **NEW**
- パッケージのインストールのところで、SAFEではなくsafeパッケージであるというご指摘をいただきましたので該当箇所を修正しました。(2015/04/24) **NEW**
- 「アセンブル | トランスクリプトーム(転写物)用」の情報を更新しました。(2015/04/22) **NEW**
- [平成27年度NGSハンズオン講習会](#)を2015年7月22日-8月6日の11日間で実施します。受講申込期間は4月20日(月)-5月15日(金)12時00分です。[申込み受付サイト](#)で受講生とTAの募集の両方を行っております。昨年度の「[NGS速習コース](#)」同様、オブザーバー(TA)募集も行っていますので可能な範囲でご協力をお願いしますm(_)_m(2015/04/21) **NEW**
- [参考資料\(講義、講習会、本など\)](#)の項目を更新しました。(2015/04/21) **NEW**
- [日本乳酸菌学会誌](#)のNGS関連連載の[第3回分PDF](#)を公開しました。(2015/04/20) **NEW**

- [はじめに](#)
- [参考資料](#)
- [過去のお](#)
- [インストー](#)

iu.a.u-tokyo.ac.jp から JSLAB4_1.sh (364 バイト) を開くか、または保存しますか?

ファイルを開く(O)

保存(S)

キャンセル(C)

W4-4: 共有フォルダ

ゲストOS (BioLinux8)上でも忘れたところに接続が復旧されている場合が多いが、別の手段としてホストOS (WindowsやMacintosh)のウェブブラウザ上で目的のファイル(JSLAB4_1.sh)をダウンロードし、共有フォルダ(ホストOSのデスクトップにあるshareフォルダ)に置くのが手っ取り早い。(特にMacユーザは)ファイルの拡張子にも注意!また、連載第3回のW9-2-6を思い出してshareフォルダの実行権限にも注意!

(Rで)塩基配列解析

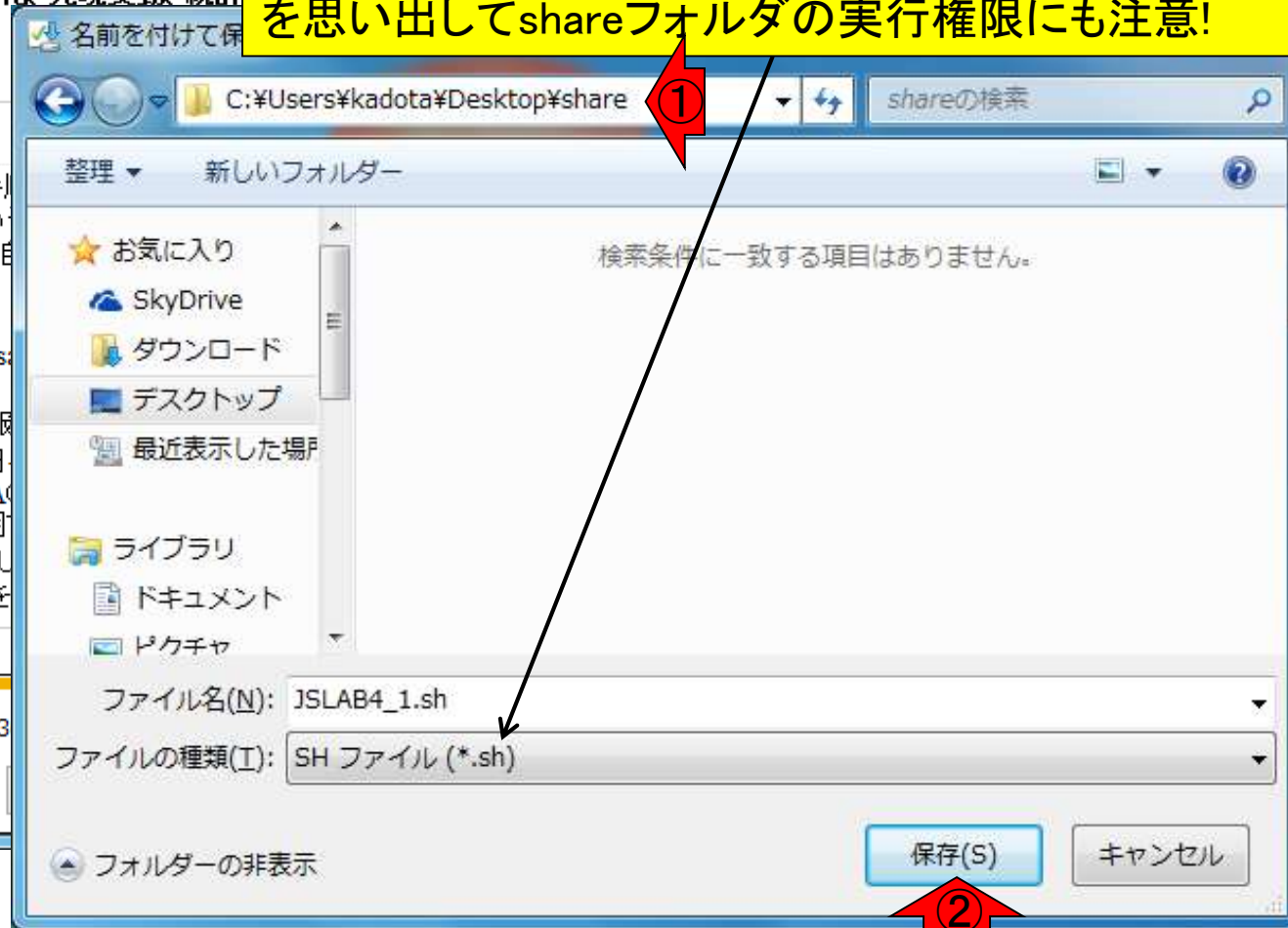
~NGS, RNA-seq, ゲノム, トランスクリプトーム, 正規化, 発現変動, 統計
(last modified 2015/04/28, since 2010)

What's new?

- このウェブページは[インストール](#)についての推奨手順Rと必要なパッケージをインストール済みであるという(Windows2015.04.03版とMacintosh2015.04.03版)で目(2015/04/03) **NEW**
- パッケージのインストールのところで、SAFEではなくsしました。(2015/04/24) **NEW**
- 「アセンブル | トランスクリプトーム(転写物)用」の情報
- 平成27年度NGSハンズオン講習会を2015年7月22日(金)12時00分です。[申込み受付サイト](#)で受講生とTA(ブザーバー(TA)募集も行っていますので可能な範囲
- 参考資料(講義、講習会、本など)の項目を更新しまし
- 日本乳酸菌学会誌のNGS関連連載の[第3回分PDF](#)を

- はじめに
- 参考資料
- 過去のお
- インストー

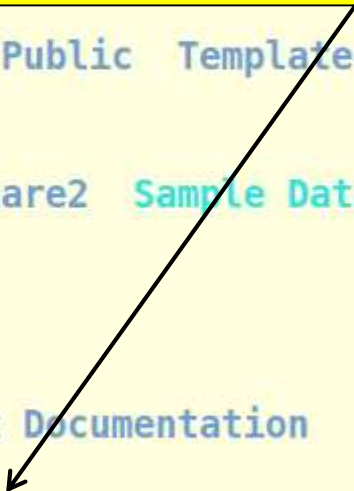
iu.a.u-tokyo.ac.jp から JSLAB4_1.sh (3



W4-4: 共有フォルダ

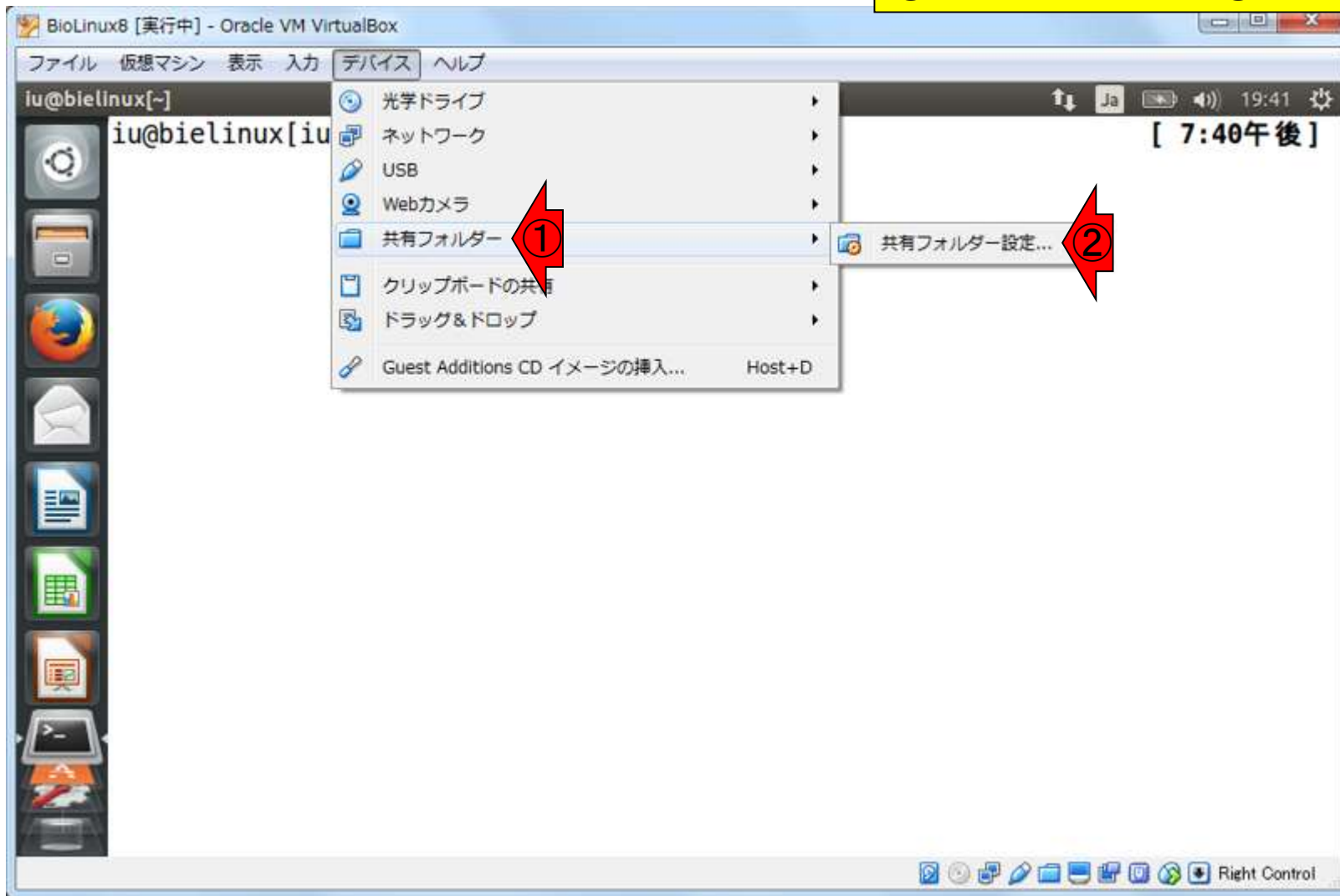
但し、何らかの理由でゲストOSを再起動することが多いが、共有フォルダを利用するにはその都度設定をする必要がある。連載第3回の①W9-2-2のホストOS側フォルダ指定(ホストOS - デスクトップ - share)と②W9-2-5のゲストOS側フォルダ指定(ゲストOS - Desktop - mac_share)が必要。ここでは共有フォルダ設定が無効になっていることを示している。連載第3回のW9-6も参照のこと。

```
iu@bielinux[~/Desktop]
iu@bielinux[srp017156] pwd
/home/iu/Documents/srp017156
iu@bielinux[srp017156] cd
iu@bielinux[iu] pwd
/home/iu
iu@bielinux[iu] ls
Desktop Documents Downloads Music Pictures Public Templates Videos
iu@bielinux[iu] cd Desktop
iu@bielinux[Desktop] ls
Bio-Linux Documentation hoge mac_share mac_share2 Sample Data
iu@bielinux[Desktop] ls -la
total 24
drwx----- 6 iu iu 4096 4月 30 13:47 .
drwxr-xr-x 18 iu iu 4096 4月 30 10:34 ..
drwxr-xr-x 2 iu iu 4096 11月 10 16:03 Bio-Linux Documentation
drwxrwxr-x 2 iu iu 4096 4月 23 20:21 hoge
drwxrwxr-x 2 iu iu 4096 4月 24 13:37 mac_share
drwxrwxr-x 2 iu iu 4096 4月 30 13:47 mac_share2
lrwxrwxrwx 1 iu iu 29 11月 10 16:03 Sample Data -> /usr/local/bioinf/sampledatab/
iu@bielinux[Desktop]
```



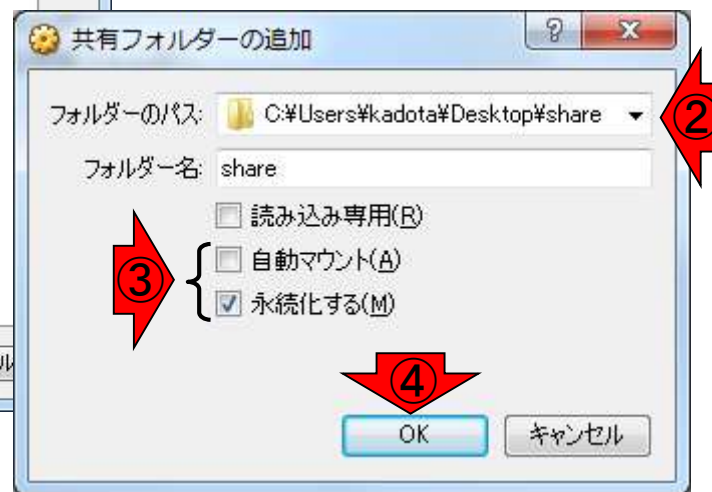
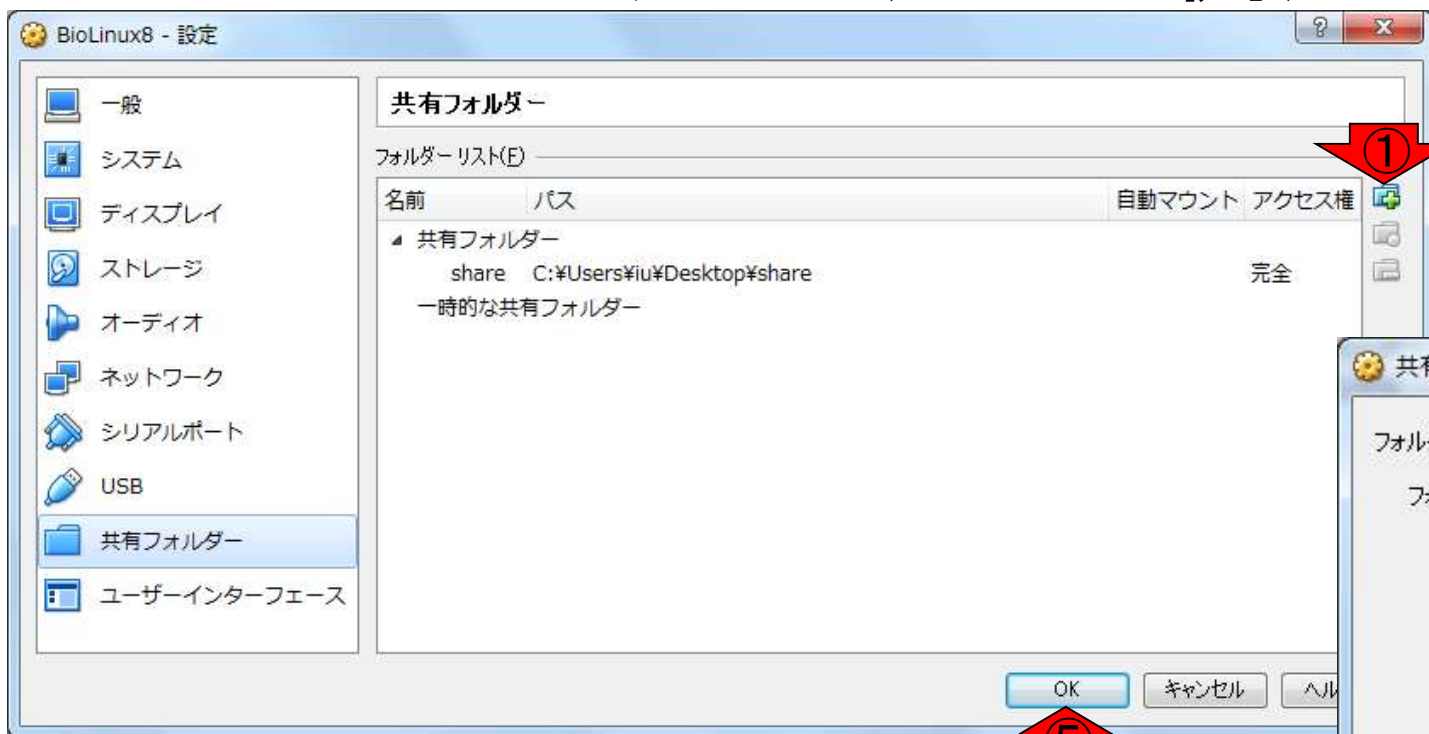
W4-4: 共有フォルダ

ホストOS側フォルダ指定(ホストOS - デスクトップ - share)の一部。連載第3回W9-2-2も参照。
①共有フォルダー、②共有フォルダー設定。



W4-5-1: マウント(ホスト側)

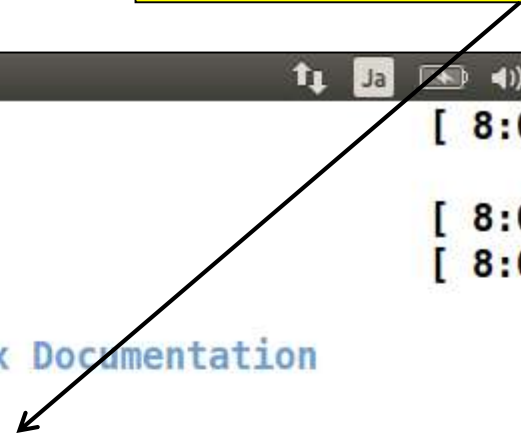
連載第3回W9-2-2。ホストOS側フォルダ指定(ホストOS - デスクトップ - share)の一部を再掲載。③のところで「自動マウント」にチェックは入れず、「永続化する」はチェックを入れる。



W4-5-1: 確認

ゲストOS側フォルダ指定(ゲストOS - デスクトップ - mac_share)の確認

```
iu@bielinux[iu] pwd [ 8:02午後 ]
/home/iu
iu@bielinux[iu] cd Desktop [ 8:02午後 ]
iu@bielinux[Desktop] ls -lh [ 8:02午後 ]
total 12K
drwxr-xr-x 2 iu iu 4.0K 7月 13 15:08 Bio-Linux Documentation
drwxr-x--- 2 iu iu 4.0K 7月 13 17:07 hoge
drwxrwxr-x 2 iu iu 4.0K 7月 13 17:13 mac_share
lrwxrwxrwx 1 iu iu 29 7月 13 15:08 Sample Data -> /usr/local/bioinf/sampled
ata/
iu@bielinux[Desktop] pwd [ 8:02午後 ]
/home/iu/Desktop
iu@bielinux[Desktop] █ [ 8:02午後 ]
```



W4-5-2: dkms

dkmsライブラリのインストール。①「sudo apt-get install dkms」と打ってリターン。パスワードを聞かれたらログインパスワードを打ち込む(推奨手順通りだとpass1409)

```
iu@bielinux[iu] pwd [ 8:02午後 ]
/home/iu
iu@bielinux[iu] cd Desktop [ 8:02午後 ]
iu@bielinux[Desktop] ls -lh [ 8:02午後 ]
total 12K
drwxr-xr-x 2 iu iu 4.0K 7月 13 15:08 Bio-Linux Documentation
drwxr-x--- 2 iu iu 4.0K 7月 13 17:07 hoge
drwxrwxr-x 2 iu iu 4.0K 7月 13 17:13 mac_share
lrwxrwxrwx 1 iu iu 29 7月 13 15:08 Sample Data -> /usr/local/bioinf/sampled
ata/
iu@bielinux[Desktop] pwd [ 8:02午後 ]
/home/iu/Desktop
iu@bielinux[Desktop] sudo apt-get install dkms [ 8:02午後 ]
[sudo] password for iu: █
```



W4-5-2: dkms

無事インストールが完了すると、②のようにコマンド入力待ち状態となる。2015.09.18に別のマシンでやろうとしたら、「Unable to fetch some archives, maybe run apt-get update or try with --fix-missing?」というメッセージが出ました。私の場合は「sudo apt-get update」をやったのち、再度①をやると成功しました。

```
File Edit View Search Terminal Help
iu@bielinux[Desktop] pwd
/home/iu/Desktop
① iu@bielinux[Desktop] sudo apt-get install dkms
[sudo] password for iu:
Reading package lists... Done
Building dependency tree
Reading state information... Done
The following NEW packages will be installed:
  dkms
0 upgraded, 1 newly installed, 0 to remove and 52 not upgraded.
Need to get 64.6 kB of archives.
After this operation, 349 kB of additional disk space will be used.
Get:1 http://jp.archive.ubuntu.com/ubuntu/ trusty-updates/main dkms all 2.2.0.
3-1.1ubuntu5.14.04.1 [64.6 kB]
Fetched 64.6 kB in 0s (201 kB/s)
Selecting previously unselected package dkms.
(Reading database ... 350103 files and directories currently installed.)
Preparing to unpack .../dkms_2.2.0.3-1.1ubuntu5.14.04.1_all.deb ...
Unpacking dkms (2.2.0.3-1.1ubuntu5.14.04.1) ...
Processing triggers for man-db (2.6.7.1-1ubuntu1) ...
Setting up dkms (2.2.0.3-1.1ubuntu5.14.04.1) ...
② iu@bielinux[Desktop] █
```

[8:12午後]

W4-5-3: Guest Additions.

①「デバイス - Guest Additions CDイメージの挿入」。もし「挿入できません。このディスクを強制挿入してみますか?」と出たら「強制マウント解除」を押すなどしてもダメなようです。これは以前挿入したものが残っているためなので、この場合はW4-8を先に行って、「Guest Additions CDイメージ」の除去をやってから、再度ここに戻る

BioLinux8 [実行中] - Oracle VM VirtualBox

ファイル 仮想マシン 表示 入力 デバイス ヘルプ

iu@bielinux[~/Desktop]

- 光学ドライブ
- ネットワーク
- USB
- Webカメラ
- 共有フォルダー
- クリップボードの共有
- ドラッグ&ドロップ
- Guest Additions CD イメージの挿入... Host+D**

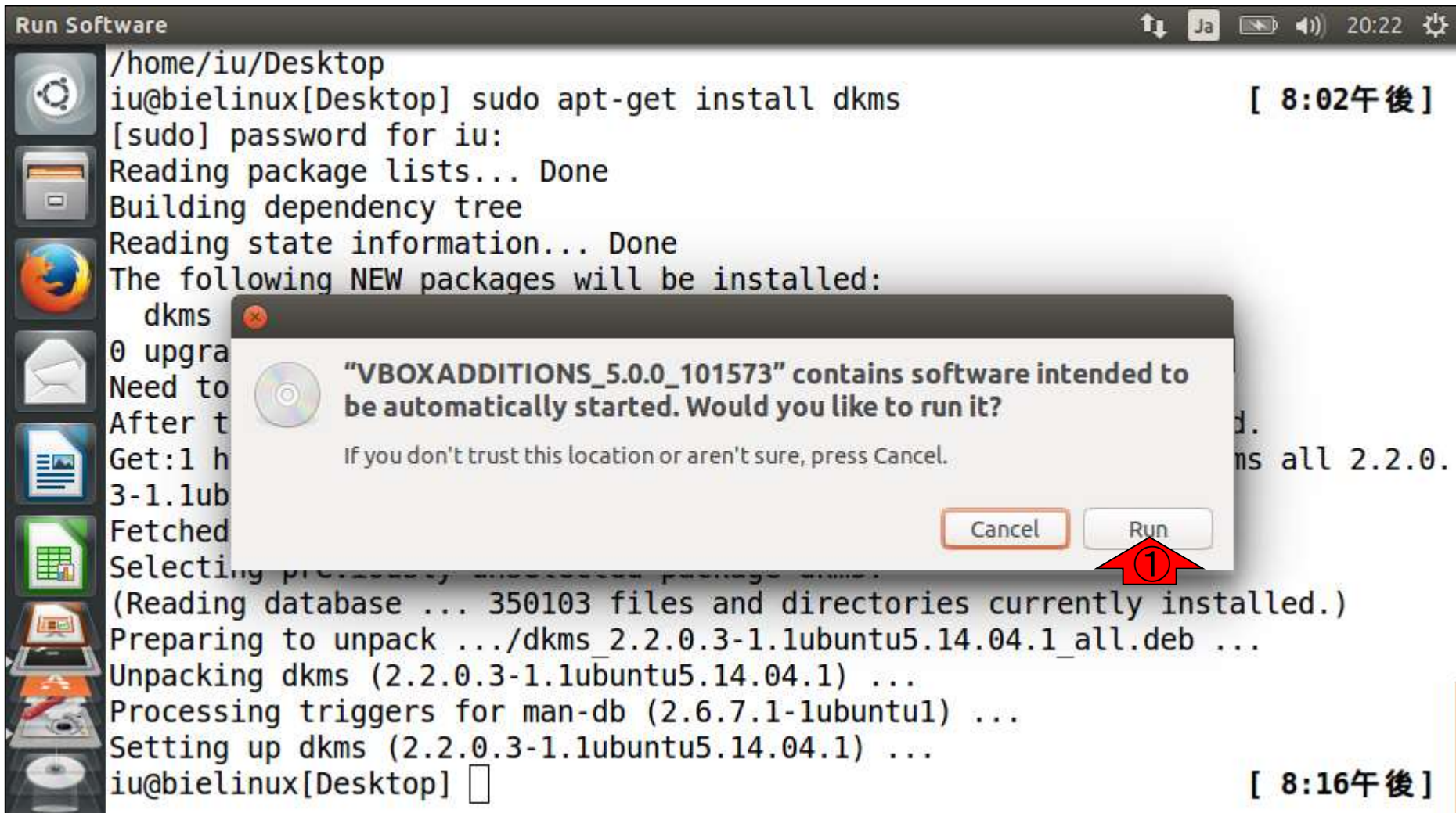
```
iu@bielinux[~/Desktop]
/~/Desktop
iu@bielinux[Desktop]
[sudo] password:
Reading package lists...
Building dependency tree...
Reading state information...
The following packages will be upgraded:
  dkms
0 upgraded, 1 newly installed, 0 to remove and 52 not upgraded.
Need to get 64.6 kB of archives.
After this operation, 349 kB of additional disk space will be used.
Get:1 http://jp.archive.ubuntu.com/ubuntu/trusty-updates/main dkms all 2.2.0.3-1.1ubuntu5.14.04.1 [64.6 kB]
Fetched 64.6 kB in 0s (201 kB/s)
Selecting previously unselected package dkms.
(Reading database ... 350103 files and directories currently installed.)
Preparing to unpack .../dkms_2.2.0.3-1.1ubuntu5.14.04.1_all.deb ...
Unpacking dkms (2.2.0.3-1.1ubuntu5.14.04.1) ...
Processing triggers for man-db (2.6.7.1-1ubuntu1) ...
Setting up dkms (2.2.0.3-1.1ubuntu5.14.04.1) ...
iu@bielinux[Desktop]
```

[8:16午後]

Guest Additions ディスクファイルを仮想光学ドライブに挿入します

Right Control

W4-5-3: Guest Additions...



Run Software

```
/home/iu/Desktop
iu@bielinux[Desktop] sudo apt-get install dkms
[sudo] password for iu:
Reading package lists... Done
Building dependency tree
Reading state information... Done
The following NEW packages will be installed:
 dkms
0 upgra
Need to
After t
Get:1 h
3-1.1ub
Fetched
Selecting previously unselected package dkms.
(Reading database ... 350103 files and directories currently installed.)
Preparing to unpack .../dkms_2.2.0.3-1.1ubuntu5.14.04.1_all.deb ...
Unpacking dkms (2.2.0.3-1.1ubuntu5.14.04.1) ...
Processing triggers for man-db (2.6.7.1-1ubuntu1) ...
Setting up dkms (2.2.0.3-1.1ubuntu5.14.04.1) ...
iu@bielinux[Desktop] █
```

[8:02午後]

"VBOXADDITIONS_5.0.0_101573" contains software intended to be automatically started. Would you like to run it?

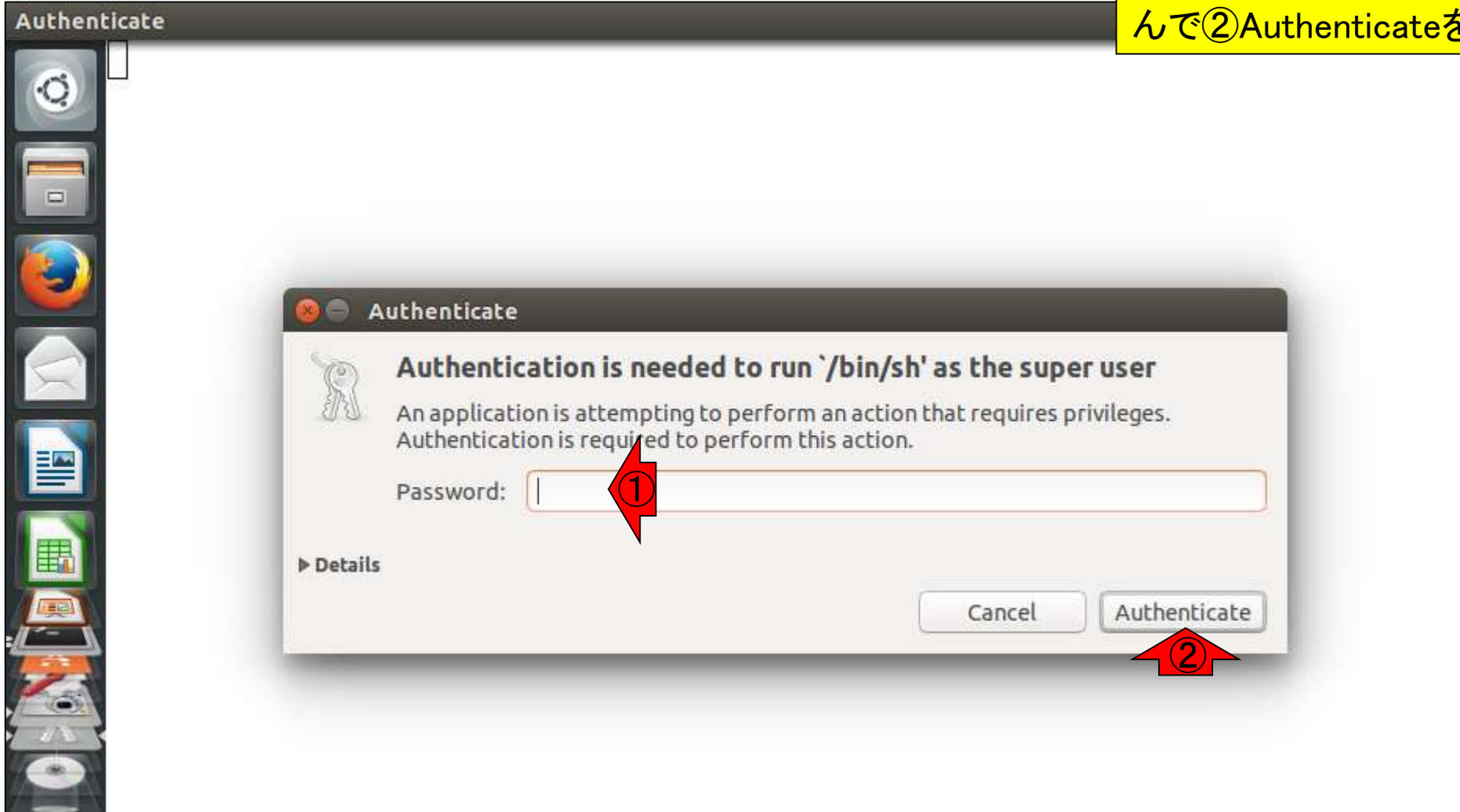
If you don't trust this location or aren't sure, press Cancel.

Cancel Run

[8:16午後]

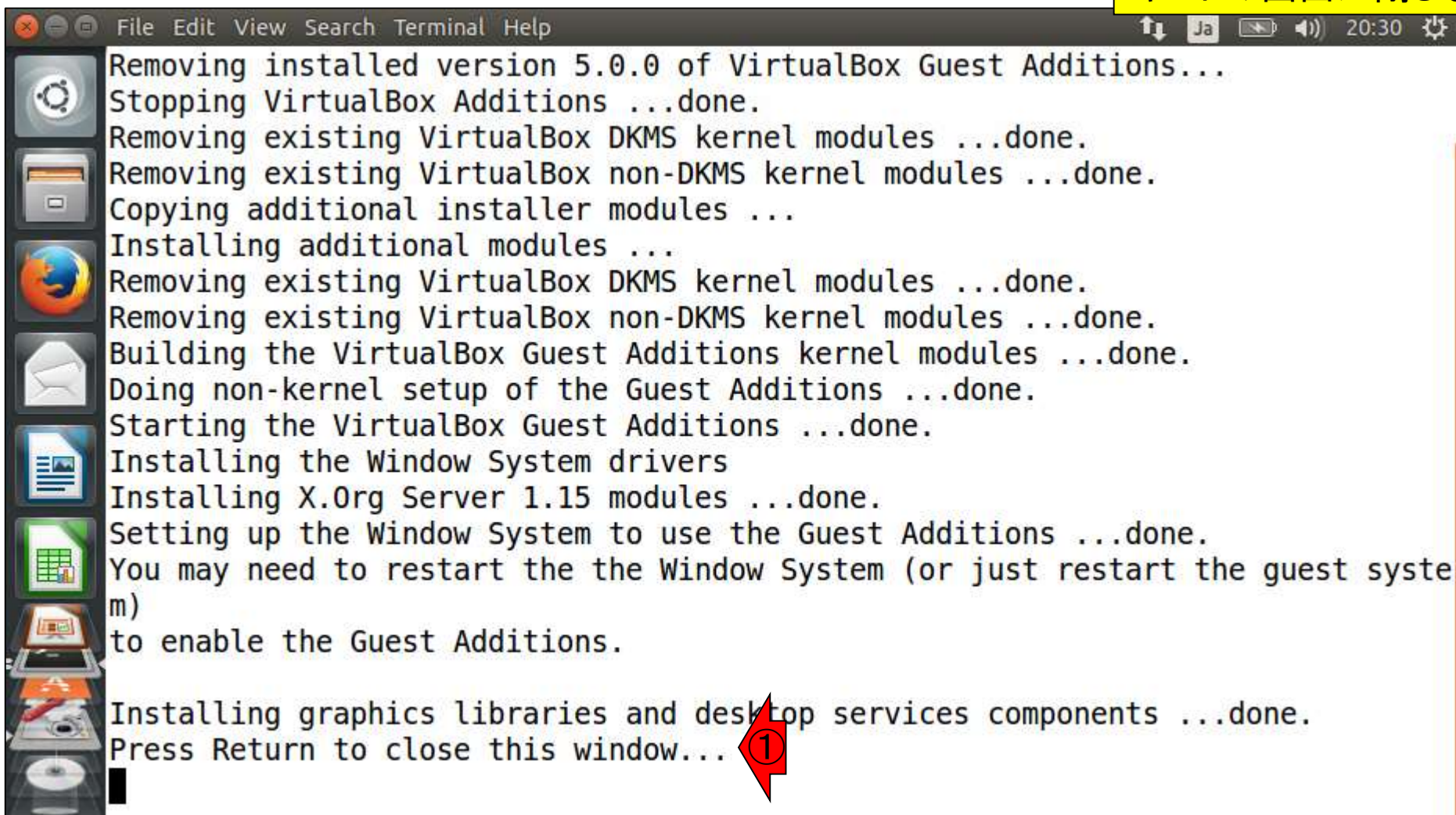
W4-5-3: Guest Additions...

①パスワードの打ち込みを求められているので、ログインパスワード(pass1409)を打ち込んで②Authenticateを押す。



W4-5-3: Guest Additions...

①「Press Return...」と出たらリターンキーを押す。するとテキストの画面が閉じる。

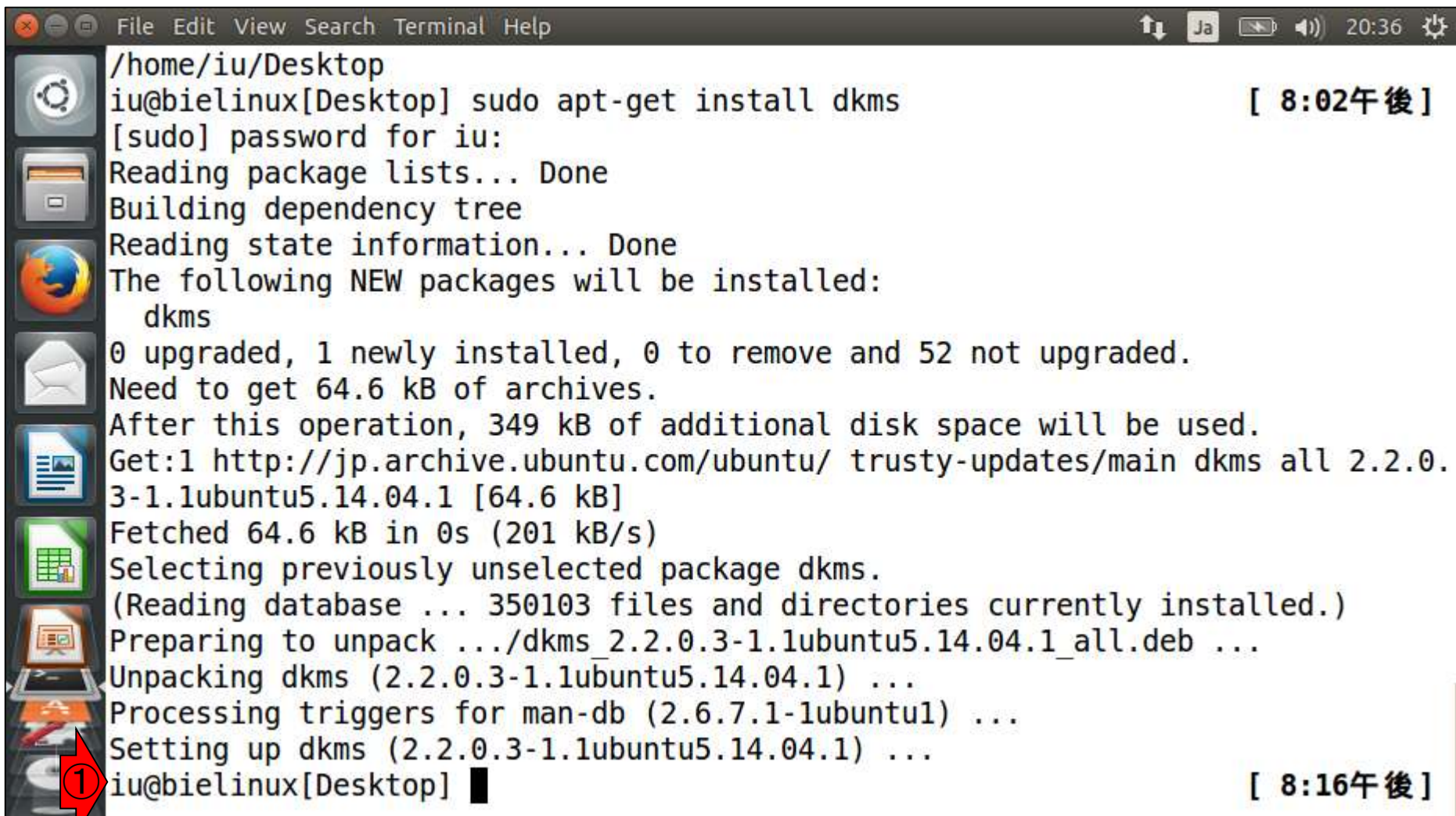


```
File Edit View Search Terminal Help
Removing installed version 5.0.0 of VirtualBox Guest Additions...
Stopping VirtualBox Additions ...done.
Removing existing VirtualBox DKMS kernel modules ...done.
Removing existing VirtualBox non-DKMS kernel modules ...done.
Copying additional installer modules ...
Installing additional modules ...
Removing existing VirtualBox DKMS kernel modules ...done.
Removing existing VirtualBox non-DKMS kernel modules ...done.
Building the VirtualBox Guest Additions kernel modules ...done.
Doing non-kernel setup of the Guest Additions ...done.
Starting the VirtualBox Guest Additions ...done.
Installing the Window System drivers
Installing X.Org Server 1.15 modules ...done.
Setting up the Window System to use the Guest Additions ...done.
You may need to restart the the Window System (or just restart the guest system)
to enable the Guest Additions.

Installing graphics libraries and desktop services components ...done.
Press Return to close this window...
█
```

①こんな感じでコマンド入力待ち状態になっていればOK。

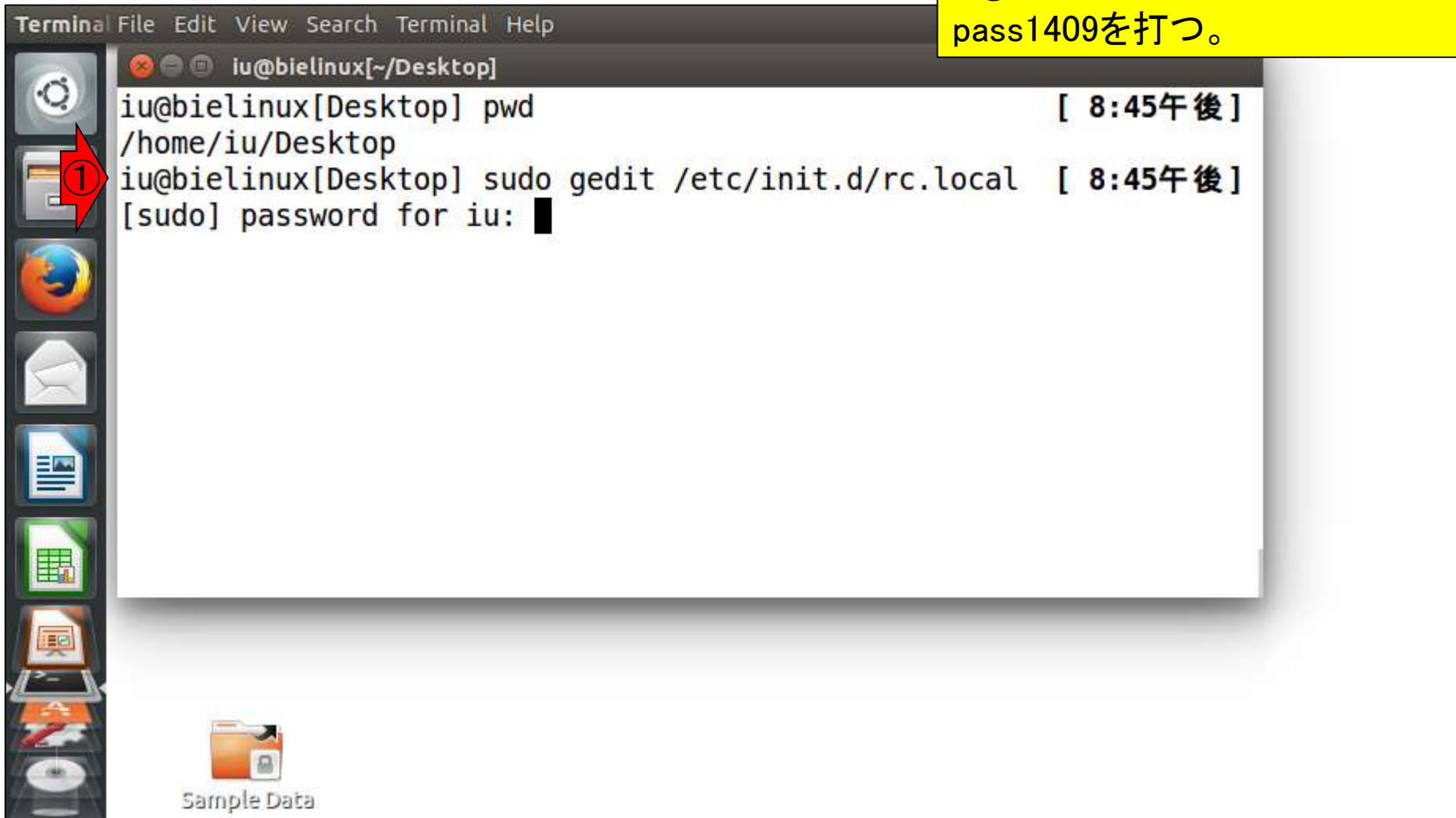
W4-5-3: Guest Additions...



```
File Edit View Search Terminal Help
/home/iu/Desktop
iu@bielinux[Desktop] sudo apt-get install dkms [ 8:02午後 ]
[sudo] password for iu:
Reading package lists... Done
Building dependency tree
Reading state information... Done
The following NEW packages will be installed:
  dkms
0 upgraded, 1 newly installed, 0 to remove and 52 not upgraded.
Need to get 64.6 kB of archives.
After this operation, 349 kB of additional disk space will be used.
Get:1 http://jp.archive.ubuntu.com/ubuntu/ trusty-updates/main dkms all 2.2.0.3-1.1ubuntu5.14.04.1 [64.6 kB]
Fetched 64.6 kB in 0s (201 kB/s)
Selecting previously unselected package dkms.
(Reading database ... 350103 files and directories currently installed.)
Preparing to unpack .../dkms_2.2.0.3-1.1ubuntu5.14.04.1_all.deb ...
Unpacking dkms (2.2.0.3-1.1ubuntu5.14.04.1) ...
Processing triggers for man-db (2.6.7.1-1ubuntu1) ...
Setting up dkms (2.2.0.3-1.1ubuntu5.14.04.1) ...
iu@bielinux[Desktop] █ [ 8:16午後 ]
```


W4-5-4: ゲスト側の設定

カレントディレクトリはどこでもよいが、①
管理者権限(sudo)で「/etc/init.d/rc.local」
をgeditで開く。パスワードを聞かれたら
pass1409を打つ。



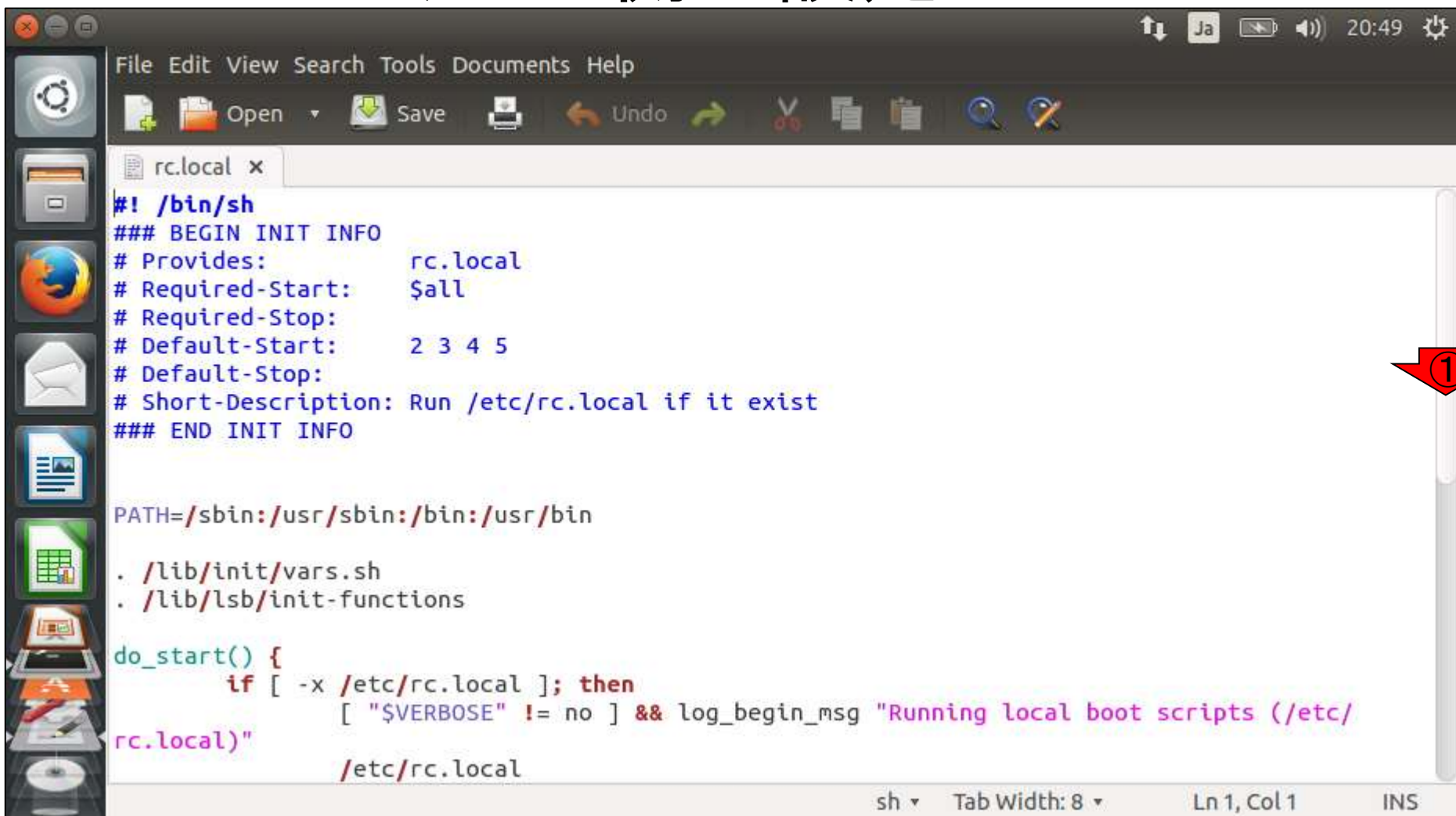
The image shows a terminal window on a Linux desktop environment. The window title is "iu@bielinux[~/Desktop]". The terminal output is as follows:

```
iu@bielinux[Desktop] pwd [ 8:45午後 ]
/home/iu/Desktop
iu@bielinux[Desktop] sudo gedit /etc/init.d/rc.local [ 8:45午後 ]
[sudo] password for iu: █
```

A red arrow with the number "1" points to the second command line in the terminal. On the desktop, there is a folder icon labeled "Sample Data".

geditが起動するので、①
ページが一番下に移動。

W4-5-4: ゲスト側の設定



```
rc.local x
#!/bin/sh
### BEGIN INIT INFO
# Provides:          rc.local
# Required-Start:    $all
# Required-Stop:
# Default-Start:    2 3 4 5
# Default-Stop:
# Short-Description: Run /etc/rc.local if it exist
### END INIT INFO

PATH=/sbin:/usr/sbin:/bin:/usr/bin

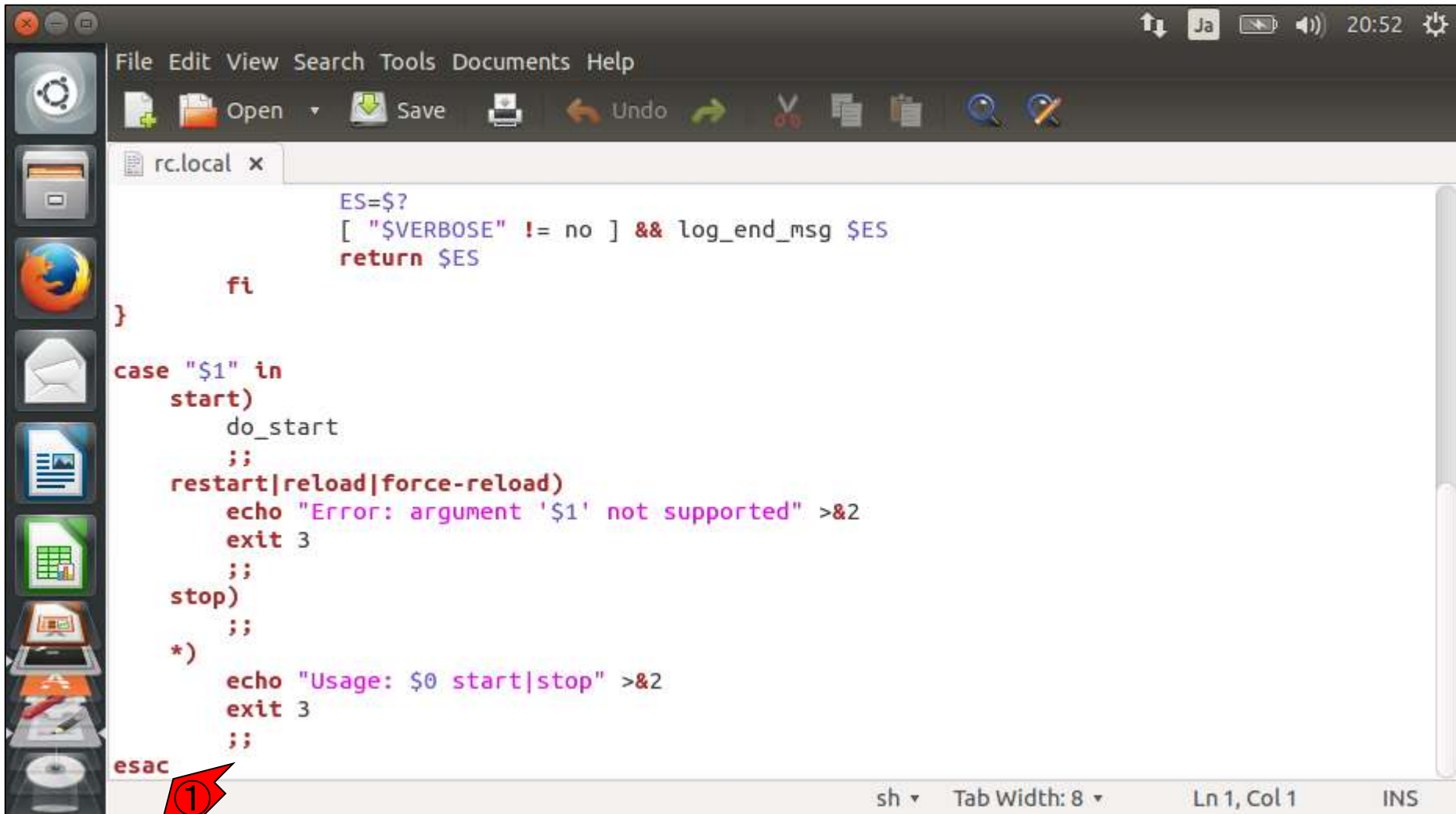
. /lib/init/vars.sh
. /lib/lsb/init-functions

do_start() {
    if [ -x /etc/rc.local ]; then
        [ "$VERBOSE" != no ] && log_begin_msg "Running local boot scripts (/etc/
rc.local)"
        /etc/rc.local
    fi
}
```

sh ▾ Tab Width: 8 ▾ Ln 1, Col 1 INS

私の環境では①のような感じに見えています。

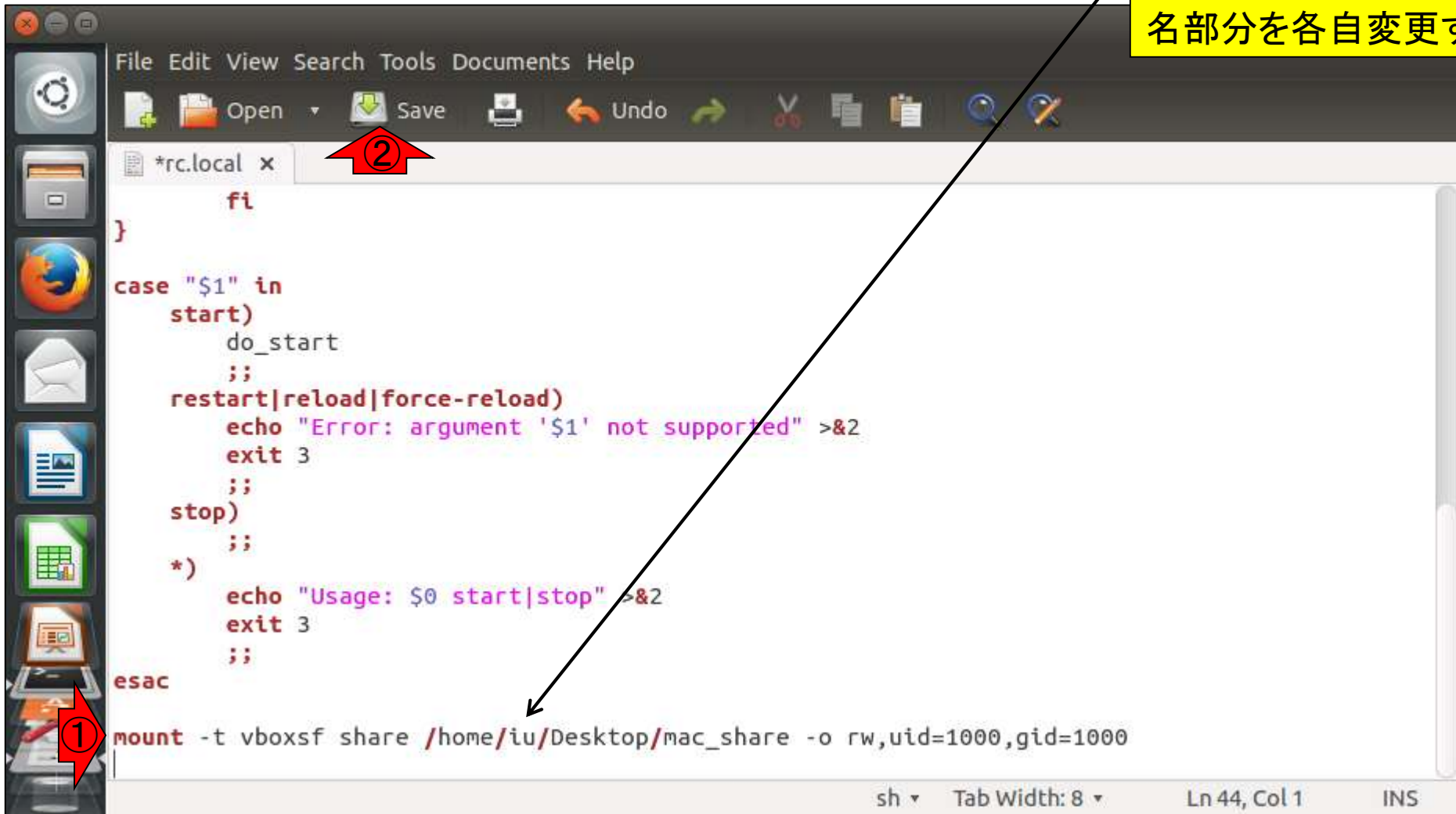
W4-5-4: ゲスト側の設定



```
File Edit View Search Tools Documents Help
Open Save Undo
rc.local x
ES=$?
[ "$VERBOSE" != no ] && log_end_msg $ES
return $ES
fi
}
case "$1" in
start)
do_start
;;
restart|reload|force-reload)
echo "Error: argument '$1' not supported" >&2
exit 3
;;
stop)
;;
*)
echo "Usage: $0 start|stop" >&2
exit 3
;;
esac
sh Tab Width: 8 Ln 1, Col 1 INS
```

W4-5-4: 1行分追加

①の行を追加して、②Save。
これはユーザ名がiuの場合
なので、必要に応じてユーザ
名部分を各自変更すべし。



The screenshot shows a terminal editor window with a menu bar (File, Edit, View, Search, Tools, Documents, Help) and a toolbar (Open, Save, Undo, etc.). The file name is *rc.local. The script content is as follows:

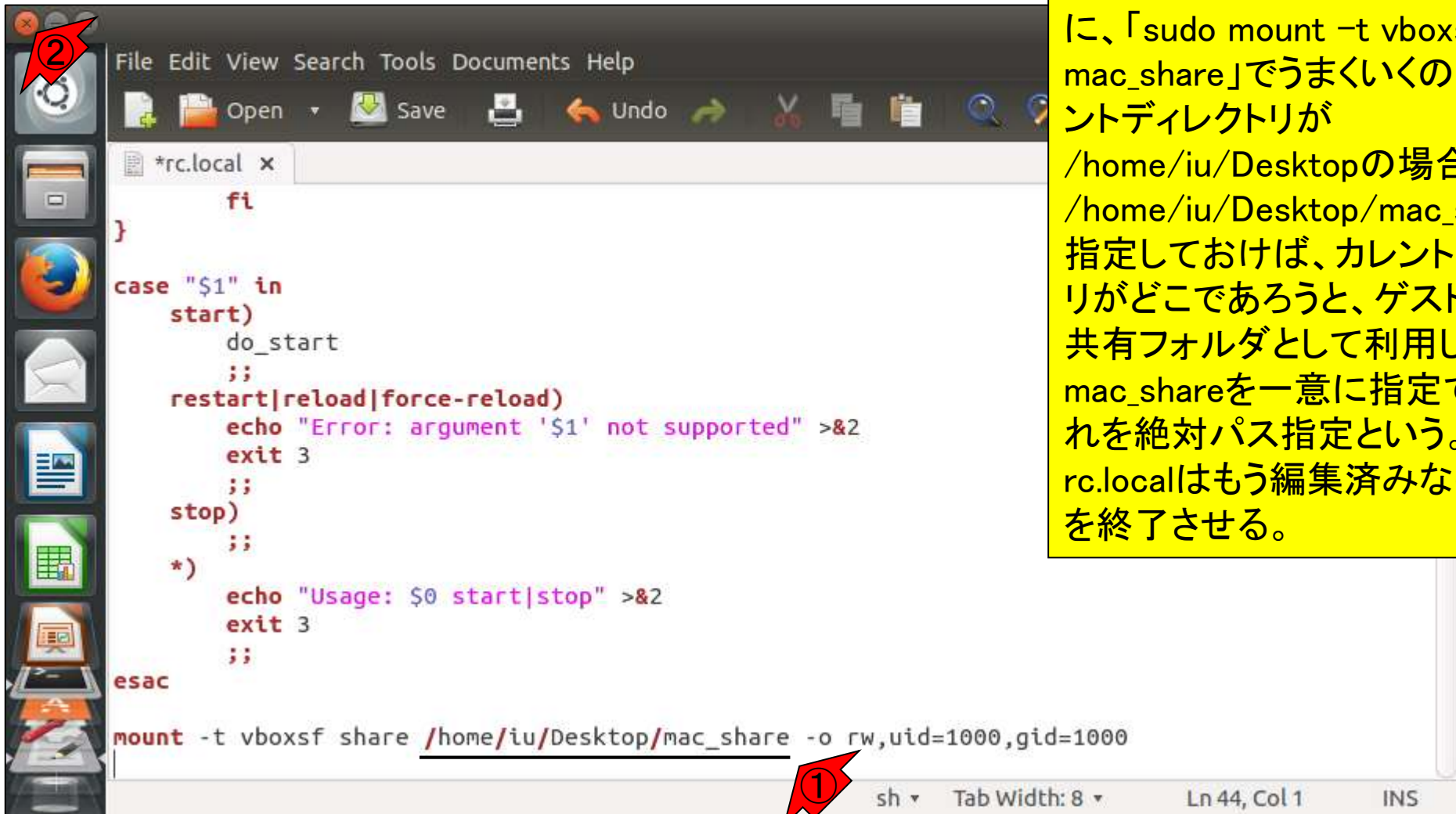
```
fi
}
case "$1" in
start)
do_start
;;
restart|reload|force-reload)
echo "Error: argument '$1' not supported" >&2
exit 3
;;
stop)
;;
*)
echo "Usage: $0 start|stop" >&2
exit 3
;;
esac
mount -t vboxsf share /home/iu/Desktop/mac_share -o rw,uid=1000,gid=1000
```

A red arrow labeled ① points to the line `mount -t vboxsf share /home/iu/Desktop/mac_share -o rw,uid=1000,gid=1000`. Another red arrow labeled ② points to the Save icon in the toolbar. The status bar at the bottom shows `sh`, `Tab Width: 8`, `Ln 44, Col 1`, and `INS`.

W4-6: 絶対パスと相対パス

手動マウントとの大きな違いは、`mac_share`が①のようになっている点である。第3回W9-6で示すように、「`sudo mount -t vboxsf share mac_share`」でうまくいくのは、カレントディレクトリが

`/home/iu/Desktop`の場合のみ。
`/home/iu/Desktop/mac_share`と指定しておけば、カレントディレクトリがどこであろうと、ゲストOS側の共有フォルダとして利用したい `mac_share`を一意に指定できる。これを絶対パス指定という。②
`rc.local`はもう編集済みなので `gedit`を終了させる。



```
*rc.local x
}
fi
}
case "$1" in
start)
do_start
;;
restart|reload|force-reload)
echo "Error: argument '$1' not supported" >&2
exit 3
;;
stop)
;;
*)
echo "Usage: $0 start|stop" >&2
exit 3
;;
esac

mount -t vboxsf share /home/iu/Desktop/mac_share -o rw,uid=1000,gid=1000
```

W4-6: 絶対パスと相対パス

gedit終了後の状態。何かErrorという文字を含むメッセージが表示されているが気にしなくてよい。



The image shows a terminal window titled "Terminal" with a system tray at the top right displaying "Ja", a battery icon, a speaker icon, and the time "21:08". The terminal window title bar reads "iu@bielinux[~/Desktop]". The terminal content is as follows:

```
iu@bielinux[Desktop] pwd [ 8:45午後 ]  
/home/iu/Desktop  
iu@bielinux[Desktop] sudo gedit /etc/init.d/rc.local [ 8:45午後 ]  
[sudo] password for iu:  
  
(gedit:25887): Gtk-WARNING **: Calling Inhibit failed: GDBus.Error:  
:org.freedesktop.DBus.Error.ServiceUnknown: The name org.gnome.Ses  
sionManager was not provided by any .service files  
  
(gedit:25887): Gtk-WARNING **: Calling Inhibit failed: GDBus.Error:  
:org.freedesktop.DBus.Error.ServiceUnknown: The name org.gnome.Ses  
sionManager was not provided by any .service files  
iu@bielinux[Desktop] [ 9:08午後 ]
```

Below the terminal window, there is a desktop icon labeled "Sample Data" with a folder icon.

W4-6: 絶対パスと相対パス

①カレントディレクトリが「/etc/init.d」で、②rc.から始まる全てのファイルを表示。③カレントディレクトリをホームディレクトリに移動。④/etc/init.d中のrc.から始まる全てのファイルを表示させることは、絶対パス指定であればどこでも可能であることを示している。絶対パスは/から始まる。

```
iu@bielinux[Desktop] cd
iu@bielinux[iu] pwd
/home/iu
iu@bielinux[iu] cd /etc/init.d
① iu@bielinux[init.d] pwd
/etc/init.d
② iu@bielinux[init.d] ls -lh rc.*
-rwxr-xr-x 1 root root 833  8月 13 20:55 rc.local
-rwxr-xr-x 1 root root 782  3月 13  2014 rc.local~
③ iu@bielinux[init.d] cd
iu@bielinux[iu] pwd
/home/iu
④ iu@bielinux[iu] ls -lh /etc/init.d/rc.*
-rwxr-xr-x 1 root root 833  8月 13 20:55 /etc/init.d/rc.local
-rwxr-xr-x 1 root root 782  3月 13  2014 /etc/init.d/rc.local~
iu@bielinux[iu]
```

[9:11午後]
[9:11午後]
[9:11午後]
[9:11午後]

③「cd ..」は一つ上のディレクトリに移動。

W4-6: 絶対パスと相対パス

```
iu@bielinux[~]
iu@bielinux[iu] pwd ← ① [ 6:09午後 ]
/home/iu
iu@bielinux[iu] ls ← ② [ 7:08午後 ]
Desktop Downloads Pictures Templates
Documents Music Public Videos
iu@bielinux[iu] cd .. ← ③ [ 7:08午後 ]
iu@bielinux[/home] pwd ← ④ [ 7:08午後 ]
/home
iu@bielinux[/home] ls ← ⑤ [ 7:08午後 ]
iu
iu@bielinux[/home] cd iu ← ⑥ [ 7:08午後 ]
iu@bielinux[iu] pwd ← ⑦ [ 7:09午後 ]
/home/iu
iu@bielinux[iu] ls ← ⑧ [ 7:09午後 ]
Desktop Downloads Pictures Templates
Documents Music Public Videos
iu@bielinux[iu] [ 7:09午後 ]
```


W4-6: 絶対パスと相対パス

③「cd .」中の.は現在のディレクトリを意味する。それゆえディレクトリ変更は実質的に行われない。④「cd ./Desktop」は「cd Desktop」と同じ。⑥は1つ上のディレクトリ中にあるDocumentsディレクトリに移動、という意味。これが相対パス指定でのディレクトリ移動に相当。⑦pwdと打ちたかったがスペルミスした例。この場合は意図を汲んで「pwdと打ちたかったんだよね?」と確認してくれている。Yesなのでyを押すと赤枠のようにpwd実行結果が表示される。⑧相対パスでカレントディレクトリ以外の場所のlsも可能。

```
iu@bielinux[~/Documents]
iu@bielinux[iu] ls
Desktop Downloads Pictures Templates
Documents Music Public Videos
iu@bielinux[iu] cd .
iu@bielinux[iu] pwd
/home/iu
iu@bielinux[iu] cd ./Desktop
iu@bielinux[Desktop] pwd
/home/iu/Desktop
iu@bielinux[Desktop] cd ../Documents
iu@bielinux[Documents] pd
zsh: correct 'pd' to 'pwd' [nyae]? y
/home/iu/Documents
iu@bielinux[Documents] ls ../Desktop
Bio-Linux Documentation mac_share Sample Data
hoge mac_share2
iu@bielinux[Documents]
```

W4-6: 絶対パスと相対パス

相対パスの続き。今いる場所からの指定だけでなく、ホームディレクトリからの相対パス指定も可能。ホームディレクトリ(ユーザ名iuの場合は/home/iu)は「~(チルダ、と読む)」で置き換えることが可能。特に②と⑥の比較で違いが一目瞭然。赤枠(~/Documents/srp017156)は、⑦のコマンドプロンプトが出ている最新のカレントディレクトリをホームディレクトリからの相対パスで表示。

```
BioLinux8 [実行中] - Oracle VM VirtualBox
仮想マシン ビュー デバイス ヘルプ
Terminal File Edit View Search Terminal Help

iu@bielinux[~/Documents/srp017156]
iu@bielinux[Documents] pwd ← ①
/home/iu/Documents
iu@bielinux[Documents] ls -la ~/Documents ← ②
total 16
drwxr-xr-x  4 iu iu 4096 12月 10 17:55 .
drwxr-xr-x 20 iu iu 4096  4月 30 17:47 ..
drwxrwxr-x  2 iu iu 4096  4月 30 11:00 srp017156
drwx----- 12 iu iu 4096 12月  1 23:13 VirtualBox Dropped Files
iu@bielinux[Documents] ls ~/Documents/srp017156 ← ③ [ 7:50午後 ]
hogel.txt      JSLAB4_1.sh      SRR616268_2.fastq.bz2
hoge_JSLAB3.sh SRR616268_1.fastq.bz2
iu@bielinux[Documents] cd ~/Documents/srp017156 ← ④ [ 7:50午後 ]
iu@bielinux[srp017156] pwd ← ⑤ [ 7:50午後 ]
/home/iu/Documents/srp017156
iu@bielinux[srp017156] ls /home/iu/Documents ← ⑥ [ 7:50午後 ]
srp017156 VirtualBox Dropped Files
iu@bielinux[srp017156] █ ⑦ [ 8:01午後 ]
```

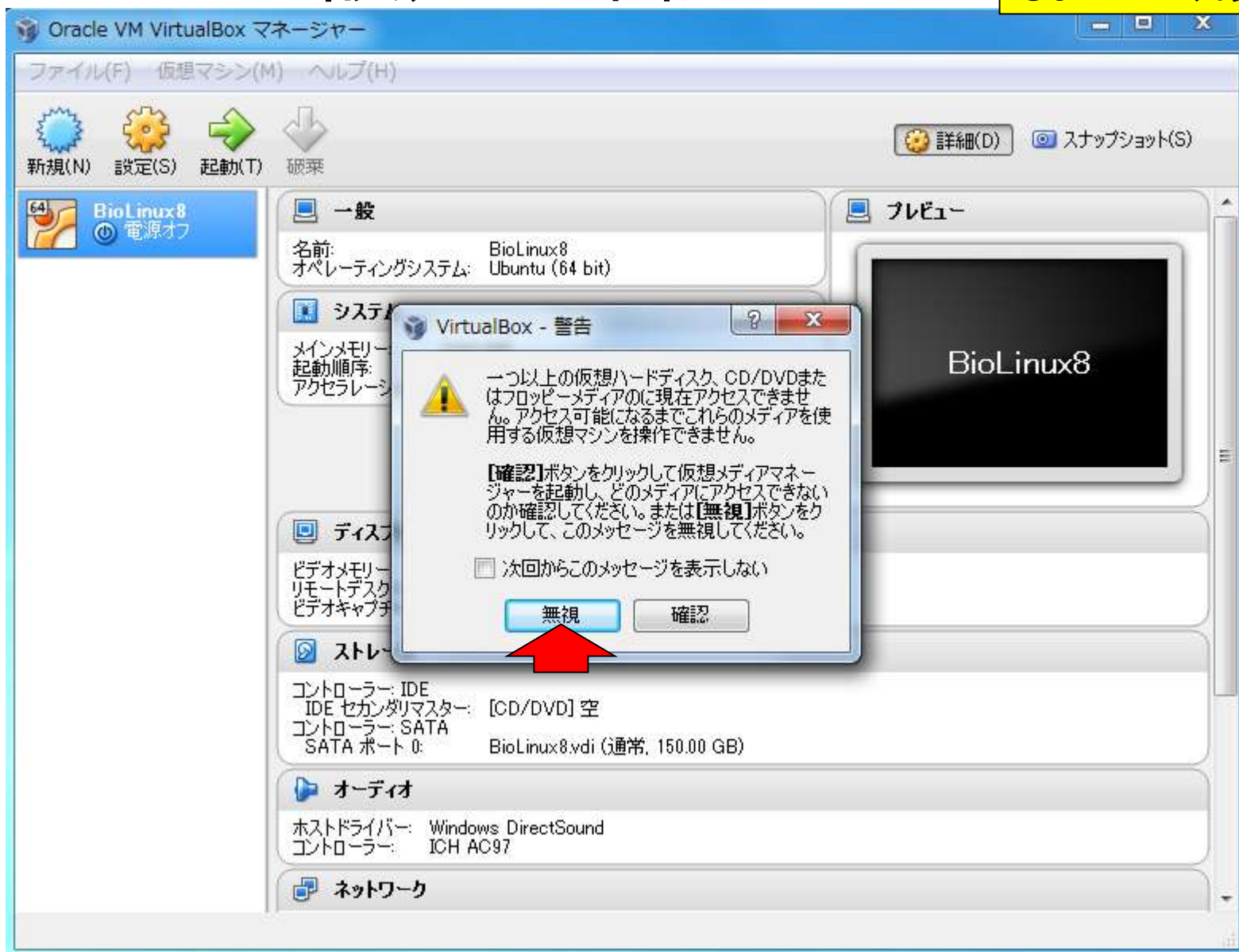
W4-7: 設定の確認

一旦②「仮想マシンの電源オフ」で③ゲストOSを終了。④VirtualBoxも終了して再起動。

The screenshot displays the Oracle VM VirtualBox interface. The main window shows a terminal window for a virtual machine named 'BioLinux8'. The terminal output shows the user 'iu' at the prompt, running 'pwd' and 'ls -la ~/Documents', which returns the path '/home/iu/Documents' and a directory listing. A dialog box titled '仮想マシンを閉じる' (Close Virtual Machine) is overlaid on the terminal. It contains the instruction '操作を選択:' (Select an operation:) and three radio button options: '仮想マシンの状態を保存(S)' (Save the state of the virtual machine), 'シャットダウン シグナル送信(S)' (Send shutdown signal), and '仮想マシンの電源オフ(P)' (Power off the virtual machine). The third option is selected. Below the options are 'OK', 'キャンセル' (Cancel), and 'ヘルプ(H)' (Help) buttons. A red arrow labeled '1' points to the close button of the main window. A red arrow labeled '2' points to the selected 'Power off' option. A red arrow labeled '3' points to the 'OK' button. A red arrow labeled '4' points to the 'Oracle VM VirtualBox マネージャー' (Oracle VM VirtualBox Manager) window, which shows the configuration for the 'BioLinux8' VM, including system, display, storage, audio, and network settings.

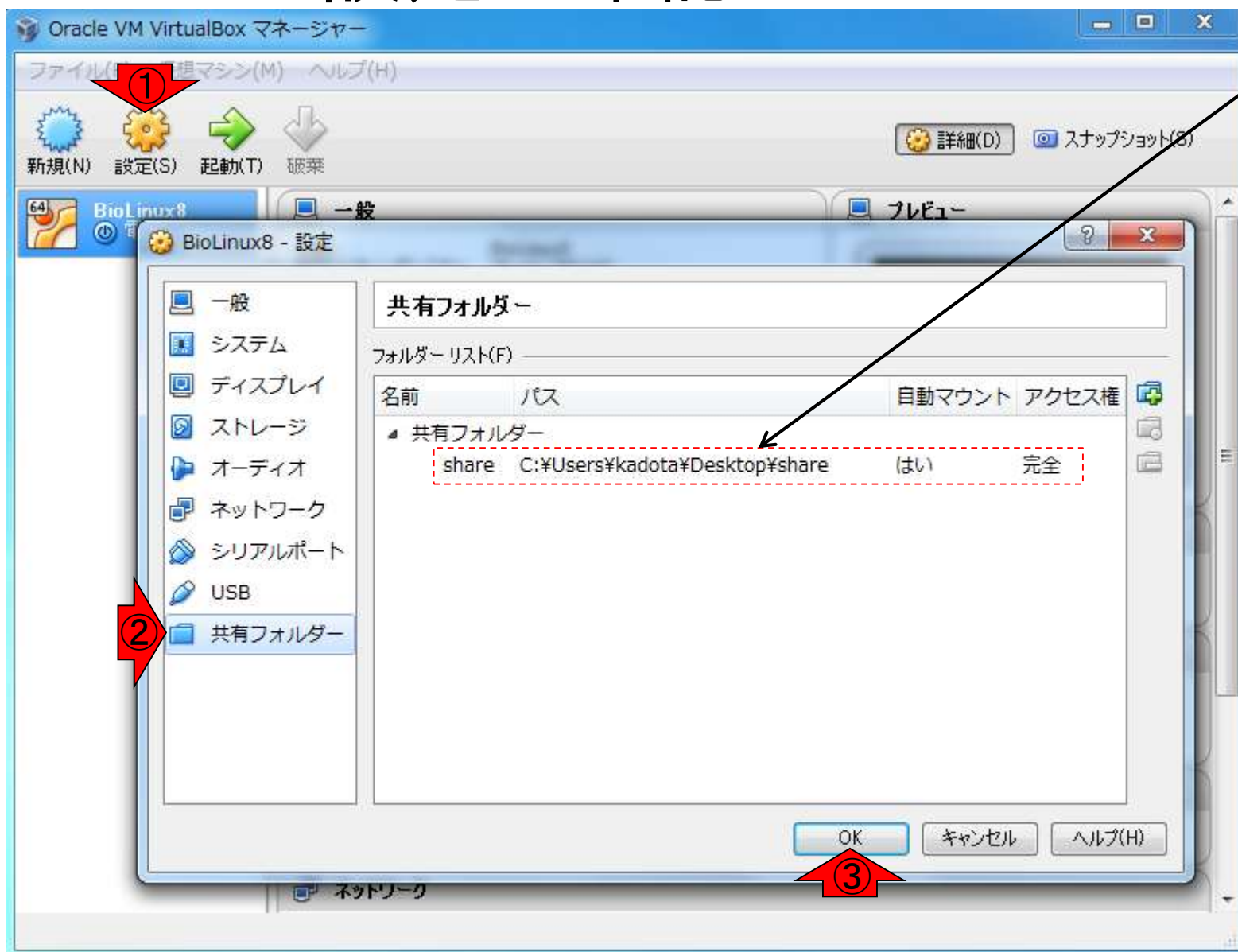
W4-7: 設定の確認

VirtualBox起動時にこのような警告メッセージが出ることもあるが、実害を被るまではとりあえず無視でよいだろう。



W4-7: 設定の確認

ホストOS側の設定はちゃんと保存されているようだ。



ゲストOS側の設定もちゃんと保存されており、起動時にオートマウントされているようだ。

W4-7: 設定の確認

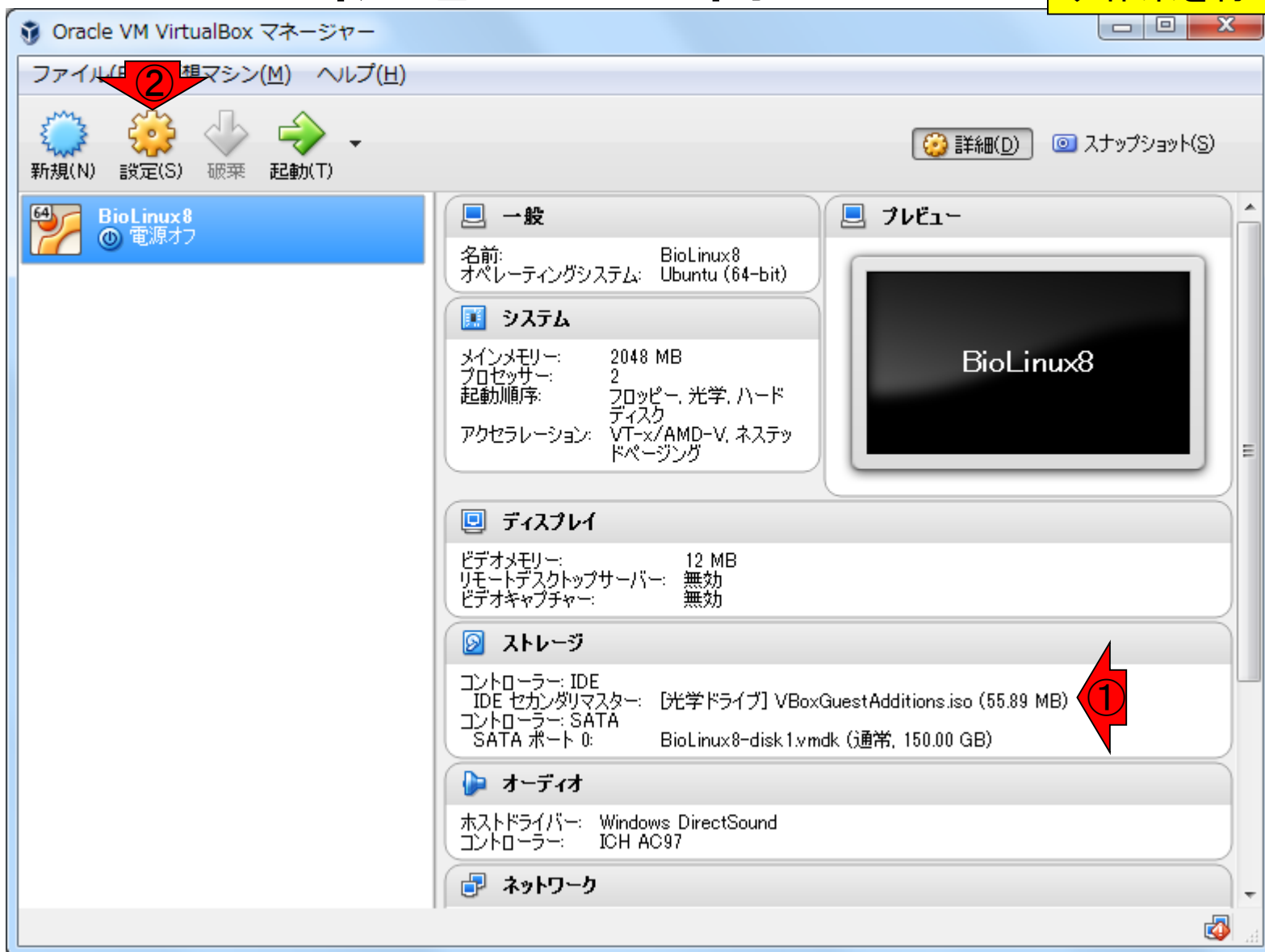
```
iu@bielinux[~/Desktop]
iu@bielinux[iu] ls ← ① [ 9:02午後]
Desktop Downloads Pictures Templates
Documents Music Public Videos
iu@bielinux[iu] cd Desktop/ ← ② [ 9:02午後]
iu@bielinux[Desktop] ls ← ③ [ 9:03午後]
Bio-Linux Documentation hoge mac_share mac_share2 Sample Data
iu@bielinux[Desktop] ls -lh ← ④ [ 9:03午後]
total 16K
drwxr-xr-x 2 iu iu 4.0K 11月 10 16:03 Bio-Linux Documentation
drwxrwxr-x 2 iu iu 4.0K 4月 23 20:21 hoge
drwxrwxrwx 1 root root 4.0K 4月 29 11:20 mac_share
drwxrwxr-x 2 iu iu 4.0K 4月 30 13:47 mac_share2
lrwxrwxrwx 1 iu iu 29 11月 10 16:03 Sample Data -> /usr/local
/bioinf/sampledata/
iu@bielinux[Desktop] ls -lh mac_share ← ⑤ [ 9:03午後]
total 3.5M
-rwxrwxrwx 1 root root 686K 11月 14 17:48 hoge.png
-rwxrwxrwx 1 root root 438 4月 24 13:50 JSLAB3_Lcasei_Linux.txt
-rwxrwxrwx 1 root root 364 4月 29 11:20 JSLAB4_1.sh
-rwxrwxrwx 1 root root 2.8M 12月 1 23:13 Lactobacillus_casei_12a.G
CA_000309565.1.22.dna.toplevel.fa
iu@bielinux[Desktop] [ 9:04午後]
```

W4-8: 仮想CD除去

① ゲストOS(BioLinux8)を一旦終了。② 「仮想マシンの電源オフ」、③ OK。W4-5-3でうまくいかなかったヒトはやる必要あり。そうでないもとりあえずやっておいて、不都合が生じたら残すなどして対応しましょう

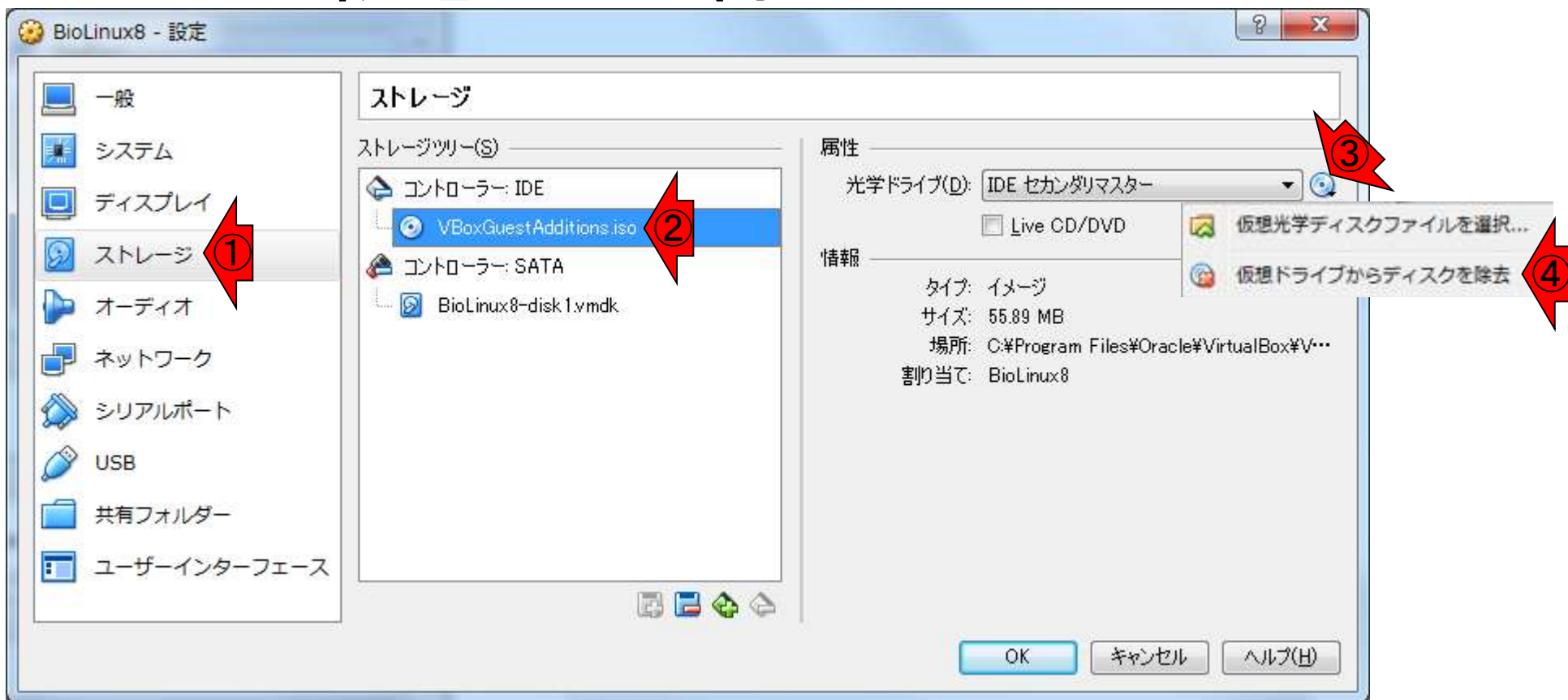
W4-8: 仮想CDの除去

①ここを「空」にするのが目的です。VBoxGuestAdditions.isoを取り出す作業を行います。②設定を押す。

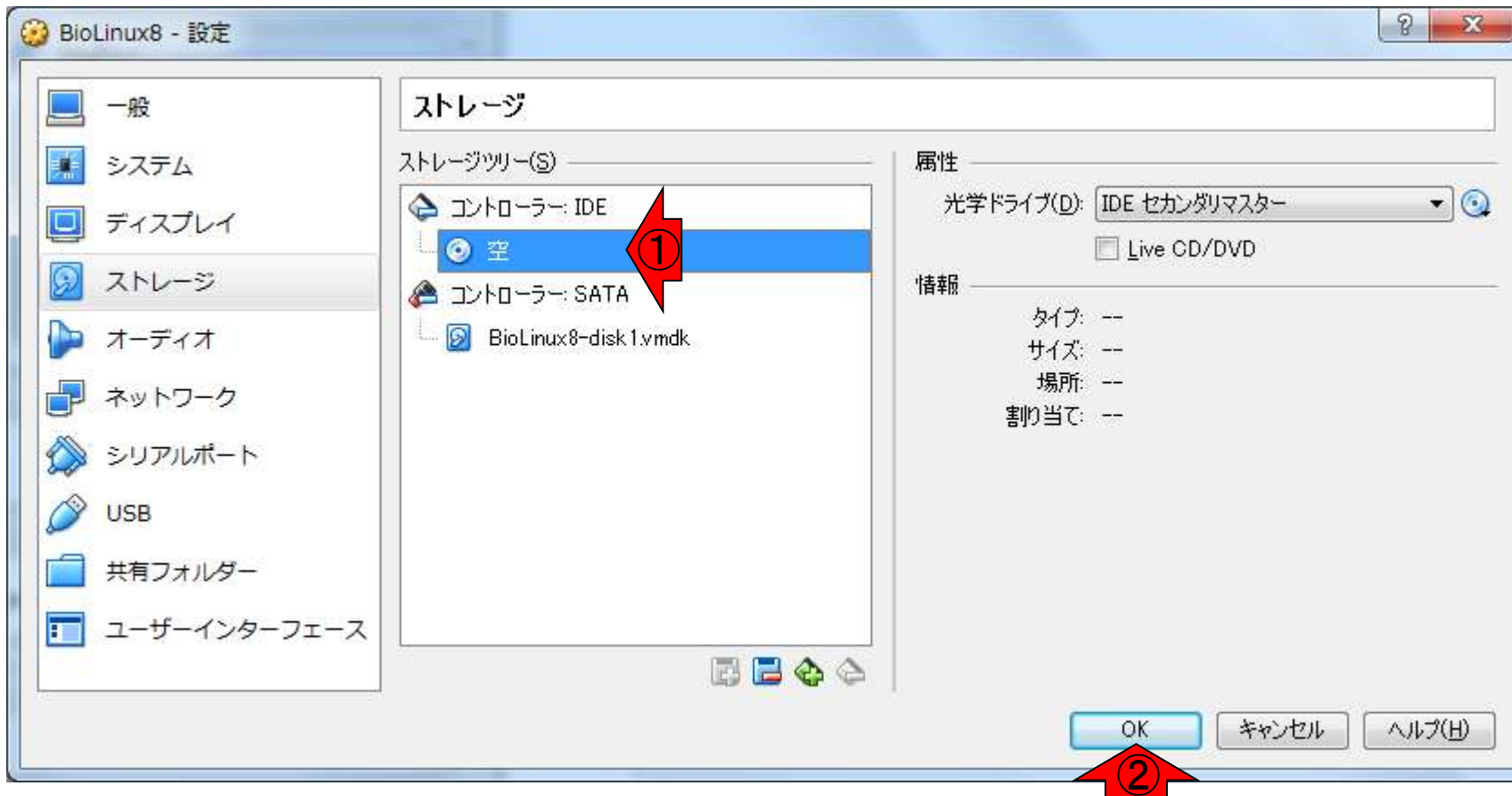


W4-8: 仮想CDの除去

①ストレージ、②.isoを選択、③CDの絵を押す、④仮想ドライブからディスクを除去



W4-8: 仮想CDの除去



W5-1: シェルスクリプト

③共有フォルダ中の
JSLAB4_1.shと⑤異なるディレク
トリ上にある⑦wgetで得た
JSLAB4_1.shの中身が同じであ
ることを確認してるだけです。

```
iu@bielinux[Desktop] pwd ← ①
/home/iu/Desktop
iu@bielinux[Desktop] ls ← ② [ 9:29午後 ]
Bio-Linux Documentation hoge mac share mac_share2 Sample Data
iu@bielinux[Desktop] more mac_share/JSLAB4_1.sh ← ③ [ 9:29午後 ]
#wget -c ftp://ftp.ddbj.nig.ac.jp/ddbj_database/dra/fastq/SRA061/SRA06148
3/SRX204226/SRR616268_1.fastq.bz2
#wget -c ftp://ftp.ddbj.nig.ac.jp/ddbj_database/dra/fastq/SRA061/SRA06148
3/SRX204226/SRR616268_2.fastq.bz2
bzip2 -dc SRR616268_1.fastq.bz2 | head -n 4000000 > SRR616268sub_1.fastq
bzip2 -dc SRR616268_2.fastq.bz2 | head -n 4000000 > SRR616268sub_2.fastq
iu@bielinux[Desktop] cd ~/Documents/srp017156 ← ④ [ 9:29午後 ]
iu@bielinux[srp017156] pwd ← ⑤ [ 9:29午後 ]
/home/iu/Documents/srp017156
iu@bielinux[srp017156] ls ← ⑥ [ 9:29午後 ]
hogel.txt JSLAB4_1.sh SRR616268_2.fastq.bz2
hoge_JSLAB3.sh SRR616268_1.fastq.bz2
iu@bielinux[srp017156] more JSLAB4_1.sh ← ⑦ [ 9:29午後 ]
#wget -c ftp://ftp.ddbj.nig.ac.jp/ddbj_database/dra/fastq/SRA061/SRA06148
3/SRX204226/SRR616268_1.fastq.bz2
#wget -c ftp://ftp.ddbj.nig.ac.jp/ddbj_database/dra/fastq/SRA061/SRA06148
3/SRX204226/SRR616268_2.fastq.bz2
bzip2 -dc SRR616268_1.fastq.bz2 | head -n 4000000 > SRR616268sub_1.fastq
bzip2 -dc SRR616268_2.fastq.bz2 | head -n 4000000 > SRR616268sub_2.fastq
iu@bielinux[srp017156] [ 9:34午後 ]
```

W5-1: シェルスクリプト

⑧「~/Documents/srp017156」中の JSLAB4_1.sh を一旦削除し、⑩相対パス指定で「~/Desktop/mac_share/JSLAB4_1.sh」をカレントディレクトリ(.)にコピーして⑫確認。赤枠の中身のうち、wgetコマンド部分は行頭に#を付加してコメントアウトしているので実行されない。

```
iu@bielinux[srp017156] more JSLAB4_1.sh
#wget -c ftp://ftp.ddbj.nig.ac.jp/ddbj_database
3/SRX204226/SRR616268_1.fastq.bz2
#wget -c ftp://ftp.ddbj.nig.ac.jp/ddbj_database
3/SRX204226/SRR616268_2.fastq.bz2
bzip2 -dc SRR616268_1.fastq.bz2 | head -n 4000000 > SRR616268sub_1.fastq
bzip2 -dc SRR616268_2.fastq.bz2 | head -n 4000000 > SRR616268sub_2.fastq
iu@bielinux[srp017156] rm -f JSLAB4_1.sh ← ⑧ [ 9:34午後 ]
iu@bielinux[srp017156] ls ← ⑨ [ 9:41午後 ]
hoge1.txt hoge_JSLAB3.sh SRR616268_1.fastq.bz2 SRR616268_2.fastq.bz2
iu@bielinux[srp017156] cp ~/Desktop/mac_share/JSLAB4_1.sh . [ 9:41午後 ]
iu@bielinux[srp017156] ls ← ⑩ [ 9:42午後 ]
hoge1.txt JSLAB4_1.sh SRR616268_2.fastq.bz2
hoge_JSLAB3.sh SRR616268_1.fastq.bz2
iu@bielinux[srp017156] more JSLAB4_1.sh ← ⑫ [ 9:42午後 ]
#wget -c ftp://ftp.ddbj.nig.ac.jp/ddbj_database/dra/fastq/SRA061/SRA06148
3/SRX204226/SRR616268_1.fastq.bz2
#wget -c ftp://ftp.ddbj.nig.ac.jp/ddbj_database/dra/fastq/SRA061/SRA06148
3/SRX204226/SRR616268_2.fastq.bz2
bzip2 -dc SRR616268_1.fastq.bz2 | head -n 4000000 > SRR616268sub_1.fastq
bzip2 -dc SRR616268_2.fastq.bz2 | head -n 4000000 > SRR616268sub_2.fastq
iu@bielinux[srp017156] █ [ 9:42午後 ]
```

W5-1: シェルスクリプト

赤矢印部分のシェルスクリプトの実行時に、timeコマンドもつけて実行に要した時間も表示させている。実行時間で主に見るのは赤枠のところ。この場合25.07秒かかっていることがわかる。このコマンドは、連載第3回のW25-2中のもとの基本的と同じ。bzip2圧縮ファイルを解凍し、最初の400万行を青枠右側のファイル名で保存している。約5.4億行から400万行分のサブセットの抽出に30秒程度かかるのだから、xxx行だとこれくらいかかりそう、という感覚的な経験を積むべし。

```
File Edit View Search Terminal Help
iu@bielinux[srp017156] pwd
/home/iu/Documents/srp017156
iu@bielinux[srp017156] ls -lh
total 14G
-rw-rw-r-- 1 iu iu 141 4月 28 13:26 hoge1.txt
-rwxr-xr-x 1 iu iu 24 4月 28 15:48 hoge_JSLAB3.sh
-rwxrwxr-x 1 iu iu 364 4月 30 21:42 JSLAB4_1.sh
-rw-rw-r-- 1 iu iu 7.2G 12月 11 14:51 SRR616268_1.fastq
-rw-rw-r-- 1 iu iu 6.6G 12月 12 14:59 SRR616268_2.fastq
iu@bielinux[srp017156] time ./JSLAB4_1.sh
./JSLAB4_1.sh 25.07s user 3.35s system 64% cpu 44.10s
iu@bielinux[srp017156] ls -lh
total 15G
-rw-rw-r-- 1 iu iu 141 4月 28 13:26 hoge1.txt
-rwxr-xr-x 1 iu iu 24 4月 28 15:48 hoge_JSLAB3.sh
-rwxrwxr-x 1 iu iu 364 4月 30 21:42 JSLAB4_1.sh
-rw-rw-r-- 1 iu iu 7.2G 12月 11 14:51 SRR616268_1.fastq.bz2
-rw-rw-r-- 1 iu iu 6.6G 12月 12 14:59 SRR616268_2.fastq.bz2
-rw-rw-r-- 1 iu iu 306M 5月 1 10:56 SRR616268sub_1.fastq?
-rw-rw-r-- 1 iu iu 278M 5月 1 10:56 SRR616268sub_2.fastq?
iu@bielinux[srp017156] █
```

[10:57午前]

W5-2: 改行コード

①青枠の出力ファイル名末尾の「？」に着目。②この「？」は赤枠中のシェルスクリプト(JSLAB4_1.sh)中の意図したファイル名末尾とは異なる。③気持ち悪いので一旦削除しようとする「？」ではなく「\r」になっている。この「逆スラッシュ」は、Windows環境では「¥r」としか打てない。これはOS間での改行コードの違いによる。

```
iu@bielinux[srp017156] ls -lh
total 15G
-rw-rw-r-- 1 iu iu 141 4月 28 13:26 hoge1.txt
-rwxr-xr-x 1 iu iu 24 4月 28 15:48 hoge_JSLAB3.sh
-rwxrwxr-x 1 iu iu 364 4月 30 21:42 JSLAB4_1.sh
-rw-rw-r-- 1 iu iu 7.2G 12月 11 14:51 SRR616268_1.fastq
-rw-rw-r-- 1 iu iu 6.6G 12月 12 14:59 SRR616268_2.fastq
-rw-rw-r-- 1 iu iu 306M 5月 1 10:56 SRR616268sub_1.fastq?
-rw-rw-r-- 1 iu iu 278M 5月 1 10:56 SRR616268sub_2.fastq?
iu@bielinux[srp017156] more JSLAB4_1.sh
#wget -c ftp://ftp.ddbj.nig.ac.jp/ddbj_database/dra/fastq/SRA061/SRA06148
3/SRX204226/SRR616268_1.fastq.bz2
#wget -c ftp://ftp.ddbj.nig.ac.jp/ddbj_database/dra/fastq/SRA061/SRA06148
3/SRX204226/SRR616268_2.fastq.bz2
bzip2 -dc SRR616268_1.fastq.bz2 | head -n 4000000 > SRR616268sub_1.fastq
bzip2 -dc SRR616268_2.fastq.bz2 | head -n 4000000 > SRR616268sub_2.fastq
iu@bielinux[srp017156] rm SRR616268sub_*
rm: remove regular file 'SRR616268sub_1.fastq\r'?
```

[10:57午前]

[11:48午前]

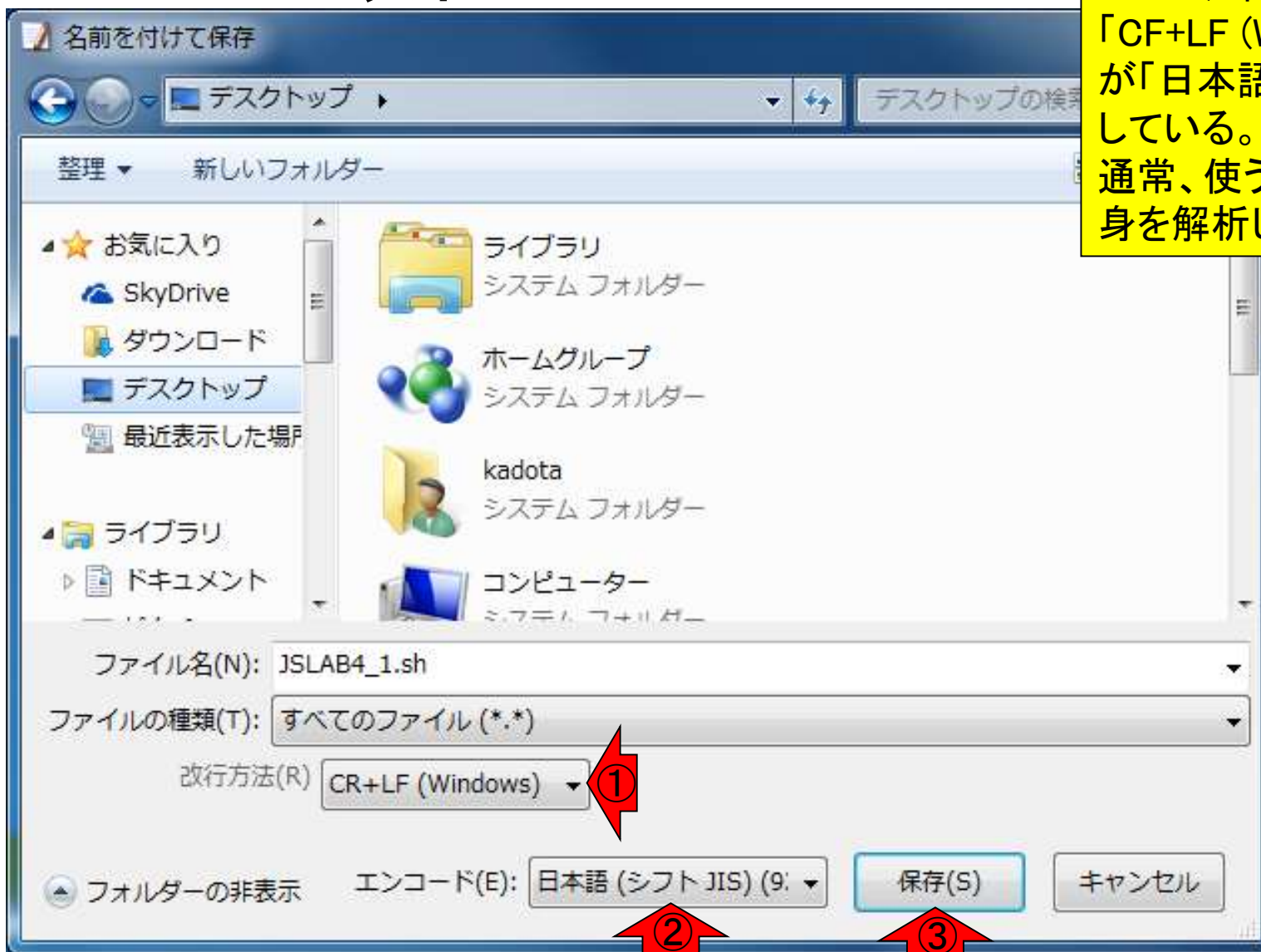
①

②

③

W5-2: 改行コード

ウェブサイトからダウンロード可能な JSLAB4_1.sh は、EmEditor というテキストエディタで作成し、①改行方法が「CR+LF (Windows)」、②文字コードが「日本語(シフトJIS)」として③保存している。改行コードや文字コードは通常、使う側が自分でファイルの中身を解析して確認する必要がある。



W5-2: 改行コード

Linux上でファイルの改行コードを調べるには、odコマンドを利用。多数のオプションが利用可能であるが、制御文字をエスケープ表現で出力する-cオプションを利用するのがおそらく一般的。赤矢印部分の基本形は「od -c 入力ファイル名」。これに「head -n 15」をパイプ(|)で組み合わせることで、出力結果の最初の15行分のみ表示。Windowsの改行コードの場合は、赤枠のように見える。

```
iu@bielinux[srp017156] pwd
/home/iu/Documents/srp017156
iu@bielinux[srp017156] ls
hogel.txt          JSLAB4_1.sh
hoge_JSLAB3.sh    SRR616268_1.fastq.bz2
iu@bielinux[srp017156] od -c JSLAB4_1.sh | head -n 15 [ 2:12午後 ]
0000000 # w g e t - c f t p : / / f
0000020 t p . d d b j . n i g . a c . j p / d d
0000040 p / d d b j _ d a t a b a s e /
0000060 d r a / f a s t q / S R A 0 6 1
0000100 / S R A 0 6 1 4 8 3 / S R X 2 0
0000120 4 2 2 6 / S R R 6 1 6 2 6 8 _ 1
0000140 . f a s t q . b z 2 \r \n # w g e
0000160 t - c f t p : / / f t p . d
0000200 d b j . n i g . a c . j p / d d
0000220 b j _ d a t a b a s e / d r a /
0000240 f a s t q / S R A 0 6 1 / S R A
0000260 0 6 1 4 8 3 / S R X 2 0 4 2 2 6
0000300 / S R R 6 1 6 2 6 8 _ 2 . f a s
0000320 t q . b z 2 \r \n b z i p 2 - d
0000340 c S R R 6 1 6 2 6 8 _ 1 . f a
iu@bielinux[srp017156] [ 2:12午後 ]
```


W5-2: 改行コード

ちなみに、Windowsの改行コードが「¥r¥n」なのに対し、Linuxの改行コードは「¥n」のみである。それゆえWindows → Linuxへの改行コードの変換は「¥r¥n → ¥n」への文字列置換と同義。

```
iu@bielinux[srp017156] pwd
/home/iu/Documents/srp017156
iu@bielinux[srp017156] ls
hogel.txt          JSLAB4_1.sh          SRR616268_2.fastq.bz2
hoge_JSLAB3.sh    SRR616268_1.fastq.bz2
iu@bielinux[srp017156] od -c JSLAB4_1.sh | head -n 15
0000000 # w g e t - c f t p : / / f
0000020 t p . d d b j . n i g . a c . j p / d d
0000040 p / d d b j _ d a t a b a s e /
0000060 d r a / f a s t q / S R A 0 6 1
0000100 / S R A 0 6 1 4 8 3 / S R X 2 0
0000120 4 2 2 6 / S R R 6 1 6 2 6 8 _
0000140 . f a s t q . b z 2 \r \n # w g e
0000160 t - c f t p : / / f t p . d
0000200 d b j . n i g . a c . j p / d d
0000220 b j _ d a t a b a s e / d r a /
0000240 f a s t q / S R A 0 6 1 / S R A
0000260 0 6 1 4 8 3 / S R X 2 0 4 2 2 6
0000300 / S R R 6 1 6 2 6 8 _ 2 . f a s
0000320 t q . b z 2 \r \n b z i p 2 - d
0000340 c S R R 6 1 6 2 6 8 _ 1 . f a
iu@bielinux[srp017156]
```

[2:11午後]

[2:12午後]

[2:12午後]

[2:12午後]

W5-2: 改行コード

perlを用いてJSLAB4_1.sh中の¥r¥nを¥nに置換するコマンド。赤下線部分で、JSLAB4_1.shを小なり記号(<)で入力として与えている。青下線部分で、リダイレクト(>)を用いてperlプログラム実行結果をhoge.shファイルに保存している。

```
iu@bielinux[srp017156] pwd
/home/iu/Documents/srp017156
iu@bielinux[srp017156] ls
hoge1.txt          JSLAB4_1.sh          SRR616268_2.fastq.bz2
hoge_JSLAB3.sh    SRR616268_1.fastq.bz2
iu@bielinux[srp017156] perl -pe 's/\r\n/\n/' < JSLAB4_1.sh > hoge.sh
iu@bielinux[srp017156] od -c hoge.sh | head -n 15
00000000  #   w   g   e   t   -   c           f   t   p   :   /   /   f
00000020  t   p   .   d   d   b   j   .   n   i   g   .   a   c   .   j
00000040  p   /   d   d   b   j   _   d   a   t   a   b   a   s   e   /
00000060  d   r   a   /   f   a   _   s   t   q   /   S   R   A   0   6   1
00000100  /   S   R   A   0   6   1   4   8   3   /   S   R   X   2   0
00000120  4   2   2   6   /   S   R   R   6   1   6   2   6   8   _   1
00000140  .   f   a   s   t   q   .   b   z   2   \n   #   w   g   e   t
00000160  .   -   c           f   t   p   :   /   /   f   t   p   .   d   d
00000200  b   j   .   n   i   g   .   a   c   .   j   p   /   d   d   b
00000220  j   _   d   a   t   a   b   a   s   e   /   d   r   a   /   f
00000240  a   s   t   q   /   S   R   A   0   6   1   /   S   R   A   0
00000260  6   1   4   8   3   /   S   R   X   2   0   4   2   2   6   /
00000300  S   R   R   6   1   6   2   6   8   _   2   .   f   a   s   t
00000320  q   .   b   z   2   \n   b   z   i   p   2   .   -   d   c
00000340  S   R   R   6   1   6   2   6   8   _   1   .   f   a   s   t
iu@bielinux[srp017156]
```

[3:52午後]

[3:52午後]

[3:53午後]

W5-2: 改行コード

改行コードをLinuxの¥nのみに変更した後のhoge.shファイルを入力としてodコマンドで眺め、最初の15行分を表示。確かに青枠のように¥nのみになっていることがわかる。

```
iu@bielinux[srp017156] pwd
/home/iu/Documents/srp017156
iu@bielinux[srp017156] ls
hoge1.txt          JSLAB4_1.sh          SRR616268_2.fastq.bz2
hoge_JSLAB3.sh    SRR616268_1.fastq.bz2
iu@bielinux[srp017156] perl -pe 's/\r\n/\n/' < JSLAB4_1.sh > hoge.sh
iu@bielinux[srp017156] od -c hoge.sh | head -n 15
00000000 # w g e t - c f t p : / / f
00000020 t p . d d b j . n i g . a c . j
00000040 p / d d b j _ d a t a b a s e /
00000060 d r a / f a _ s t q / S R A 0 6 1
00000100 / S R A 0 6 1 4 8 3 / S R X 2 0
00000120 4 2 2 6 / S R R 6 1 6 2 6 8 _ 1
00000140 . f a s t q . b z 2 \n # w g e t
00000160 - c f t p : / / f t p . d d b
00000200 b j . n i g . a c . j p / d d b
00000220 j _ d a t a b a s e / d r a / f
00000240 a s t q / S R A 0 6 1 / S R A 0
00000260 6 1 4 8 3 / S R X 2 0 4 2 2 6 /
00000300 S R R 6 1 6 2 6 8 _ 2 . f a s t
00000320 q . b z 2 \n b z i p 2 - d c t
00000340 S R R 6 1 6 2 6 8 _ 1 . f a s t
iu@bielinux[srp017156]
```

W5-2: 改行コード

¥r¥nを¥nに置換する本質的なところ。
's/置換前/置換後/'が基本形。

```
iu@bielinux[srp017156] pwd [ 3:52午後 ]
/home/iu/Documents/srp017156
iu@bielinux[srp017156] ls [ 3:52午後 ]
hoge1.txt JSLAB4_1.sh SRR616268_2.fastq.bz2
hoge_JSLAB3.sh SRR616268_1.fastq.bz2
iu@bielinux[srp017156] perl -pe 's/\r\n/\n/' < JSLAB4_1.sh > hoge.sh
iu@bielinux[srp017156] od -c hoge.sh | head -n 15 [ 3:52午後 ]
0000000 # w g e t - c f t p : / / f
0000020 t p . d d b j . n i g . a c . j
0000040 p / d d b j _ d a t a b a s e /
0000060 d r a / f a s t q / S R A 0 6 1
0000100 / S R A 0 6 1 4 8 3 / S R X 2 0
0000120 4 2 2 6 / S R R 6 1 6 2 6 8 _ 1
0000140 . f a s t q . b z 2 \n # w g e t
0000160 - c f t p : / / f t p . d d b
0000200 b j . n i g . a c . j p / d d b
0000220 j _ d a t a b a s e / d r a / f
0000240 a s t q / S R A 0 6 1 / S R A 0
0000260 6 1 4 8 3 / S R X 2 0 4 2 2 6 /
0000300 S R R 6 1 6 2 6 8 _ 2 . f a s t
0000320 q . b z 2 \n b z i p 2 - d c t
0000340 S R R 6 1 6 2 6 8 _ 1 . f a s t
iu@bielinux[srp017156] [ 3:53午後 ]
```

W5-2: 改行コード

逆パターンで、Linux → Windowsに改行コードを変換したい場合。赤下線部分が¥nを¥r¥nに置換する本質的なところ。's/置換前/置換後/'が基本形。

```
File Edit View Search Terminal Help
iu@bielinux[srp017156] perl -pe 's/\n/\r\n/' < hoge.sh > hoge2.sh
iu@bielinux[srp017156] od -c hoge2.sh | head -n 15 [ 4:30午後 ]
00000000 # w g e t - c f t p : / / f
00000020 t p . d d b j . n i g . a c . j
00000040 p / d d b j _ d a t a b a s e /
00000060 d r a / f a s t q / S R A 0 6 1
00000100 / S R A 0 6 1 4 8 3 / S R X 2 0
00000120 4 2 2 6 / S R R 6 1 6 2 6 8 _ 1
00000140 . f a s t q . b z 2 \r \n # w g e
00000160 t - c f t p : / / f t p . d
00000200 d b j . n i g . a c . j p / d d
00000220 b j _ d a t a b a s e / d r a /
00000240 f a s t q / S R A 0 6 1 / S R A
00000260 0 6 1 4 8 3 / S R X 2 0 4 2 2 6
00000300 / S R R 6 1 6 2 6 8 _ 2 . f a s
00000320 t q . b z 2 \r \n b z i p 2 - d
00000340 c S R R 6 1 6 2 6 8 _ 1 . f a
iu@bielinux[srp017156] [ 4:30午後 ]
```

W5-2: 改行コード

- WindowsとLinux間のやり取りの際にWinSCPなどのファイル転送ソフトを利用する機会が多いが、目的に応じて改行コードを変換するか否かを要検討
 - 例1: Linux上にある数GBレベルのファイルを、(ディスク容量の豊富な) Windowsマシン付属の外付けHDDに保存しておきたいだけの場合は、Windowsマシンでデータファイルの解析をしないのだから改行コードがかわらない「バイナリ」で転送すべし。もし転送モードを「テキスト」や「自動」にした状態でOS間の移動を行うと、NGSデータの場合、数億行分の改行コードをご丁寧に変換してくれるため、ものすごく時間がかかる。
 - 例2: W5-1ではファイル末尾に?がついた状態ではあるものの、最初の400万行分のサブセットを抽出するという目的自体は一応できている。しかし、?がついたファイルを別のプログラムで実行しようとしたときにエラーが出ることもあるだろう。それゆえ、ものすごく処理に時間がかかるような場合以外は、改行コードの変換はやったほうがいいかもしれない。

W5-3: シェルスクリプト

②Linuxの改行コードにしたhoge.shを実行したが、実行権限がないと怒られている。③ chmodコマンドで実行権限を与えて、④再度実行。約30秒でサブセットの抽出が終了。⑤ SRR616268から始まるファイル群のみ表示。「SRR」まで打ち込んでからタブ補完を利用。赤枠内を見てわかるように、出力ファイル名の末尾に?がついていないことがわかる。

```
iu@bielinux[srp017156] pwd
/home/iu/Documents/srp017156
iu@bielinux[srp017156] ls -lh ← ①
total 14G
-rw-rw-r-- 1 iu iu 141  4月 28 13:26 hoge1.t
-rw-rw-r-- 1 iu iu 364  5月  1 16:30 hoge2.sn
-rwxr-xr-x 1 iu iu  24  4月 28 15:48 hoge_JSLAB3.sh
-rw-rw-r-- 1 iu iu 360  5月  1 15:52 hoge.sh
-rwxrwxr-x 1 iu iu 364  4月 30 21:42 JSLAB4_1.sh
-rw-rw-r-- 1 iu iu 7.2G 12月 11 14:51 SRR616268_1.fastq.bz2
-rw-rw-r-- 1 iu iu 6.6G 12月 12 14:59 SRR616268_2.fastq.bz2
iu@bielinux[srp017156] time ./hoge.sh ← ② [ 4:55午後 ]
zsh: permission denied: ./hoge.sh
./hoge.sh  0.00s user 0.00s system 56% cpu 0.006 total
iu@bielinux[srp017156] chmod 755 hoge.sh ← ③ [ 4:56午後 ]
iu@bielinux[srp017156] time ./hoge.sh ← ④ [ 4:57午後 ]
./hoge.sh 30.83s user 3.96s system 89% cpu 38.711 total
iu@bielinux[srp017156] ls -lh SRR616268* ← ⑤ [ 4:58午後 ]
-rw-rw-r-- 1 iu iu 7.2G 12月 11 14:51 SRR616268_1.fastq.bz2
-rw-rw-r-- 1 iu iu 6.6G 12月 12 14:59 SRR616268_2.fastq.bz2
-rw-rw-r-- 1 iu iu 306M  5月  1 16:58 SRR616268sub_1.fastq
-rw-rw-r-- 1 iu iu 278M  5月  1 16:58 SRR616268sub_2.fastq
iu@bielinux[srp017156] [ 5:03午後 ]
```

W5-3: シェルスクリプト

odコマンド以外にも多数の同様な機能を果たすコマンドがある。例えばfileコマンドは、①Linuxファイルの場合は「ASCII text」という結果を返し、②Windowsファイルの場合は「ASCII text, with CRLF line terminators」という結果を返す。③bzip2圧縮ファイルの判定も可能。④*.fastqはLinuxの改行コードであることが分かる。

```
iu@bielinux[srp017156] ls -lh
total 15G
-rw-rw-r-- 1 iu iu 141  4月 28 13:26 hoge1.txt
-rw-rw-r-- 1 iu iu 364  5月  1 16:30 hoge2.sh
-rwxr-xr-x 1 iu iu  24  4月 28 15:48 hoge_JSLAB3
-rwxr-xr-x 1 iu iu 360  5月  1 15:52 hoge.sh
-rwxrwxr-x 1 iu iu 364  4月 30 21:42 JSLAB4_1.sh
-rw-rw-r-- 1 iu iu 7.2G 12月 11 14:51 SRR616268_1.fastq.bz2
-rw-rw-r-- 1 iu iu 6.6G 12月 12 14:59 SRR616268_2.fastq.bz2
-rw-rw-r-- 1 iu iu 306M  5月  1 16:58 SRR616268sub_1.fastq
-rw-rw-r-- 1 iu iu 278M  5月  1 16:58 SRR616268sub_2.fastq
iu@bielinux[srp017156] file hoge.sh ← ① [ 5:41午後 ]
hoge.sh: ASCII text
iu@bielinux[srp017156] file JSLAB4_1.sh ← ② [ 5:41午後 ]
JSLAB4_1.sh: ASCII text, with CRLF line terminators
iu@bielinux[srp017156] file *.bz2 ← ③ [ 5:41午後 ]
SRR616268_1.fastq.bz2: bzip2 compressed data, block size = 900k
SRR616268_2.fastq.bz2: bzip2 compressed data, block size = 900k
iu@bielinux[srp017156] file *.fastq ← ④ [ 5:41午後 ]
SRR616268sub_1.fastq: ASCII text
SRR616268sub_2.fastq: ASCII text
iu@bielinux[srp017156] [ 5:55午後 ]
```


W5-4: 元に戻す

W1-4と同じ状態に戻します。
②～④のrmコマンドの実行は
かなり乱暴なやり方ですが、
bzip2圧縮ファイルを消さないよ
うに注意すれば大丈夫です。

```
iu@bielinux[srp017156] ls -lh ← ①
total 15G
-rw-rw-r-- 1 iu iu 141 4月 28 13:26 hoge1.txt
-rw-rw-r-- 1 iu iu 364 5月 1 16:30 hoge2.sh
-rwxr-xr-x 1 iu iu 24 4月 28 15:48 hoge_JSLAB3.sh
-rwxr-xr-x 1 iu iu 360 5月 1 15:52 hoge.sh
-rwxrwxr-x 1 iu iu 364 4月 30 21:42 JSLAB4_1.sh
-rw-rw-r-- 1 iu iu 7.2G 12月 11 14:51 SRR616268_1.fastq.bz2
-rw-rw-r-- 1 iu iu 6.6G 12月 12 14:59 SRR616268_2.fastq.bz2
-rw-rw-r-- 1 iu iu 306M 5月 1 16:58 SRR616268sub_1.fastq
-rw-rw-r-- 1 iu iu 278M 5月 1 16:58 SRR616268sub_2.fastq
iu@bielinux[srp017156] rm -f hoge* ← ② [ 6:56午後 ]
iu@bielinux[srp017156] rm -f JSLAB4_1.sh ← ③ [ 6:57午後 ]
iu@bielinux[srp017156] rm -f *.fastq ← ④ [ 6:57午後 ]
iu@bielinux[srp017156] ls -lh ← ⑤ [ 6:57午後 ]
total 14G
-rw-rw-r-- 1 iu iu 7.2G 12月 11 14:51 SRR616268_1.fastq.bz2
-rw-rw-r-- 1 iu iu 6.6G 12月 12 14:59 SRR616268_2.fastq.bz2
iu@bielinux[srp017156] pwd ← ⑥ [ 6:57午後 ]
/home/iu/Documents/srp017156
iu@bielinux[srp017156] █ [ 6:57午後 ]
```

W6-1:ダウンロード

JSLAB4_1_Linux.shをwgetコマンドで取得し、fileコマンドでファイルの種類がLinuxの改行コードとなっている(yesなら「ASCII text」という記述のみ)ことを確認。

```
iu@bielinux[srp017156] wget -c http://www.iu.a.u-tokyo.ac.jp/~kadota/book/JSLAB4_1_Linux.sh
--2015-05-01 19:07:43-- http://www.iu.a.u-tokyo.ac.jp/~kadota/book/JSLAB4_1_Linux.sh
Resolving www.iu.a.u-tokyo.ac.jp (www.iu.a.u-tokyo.ac.jp)... 133.11.224.25
Connecting to www.iu.a.u-tokyo.ac.jp (www.iu.a.u-tokyo.ac.jp)|133.11.224.25|:80... connected.
HTTP request sent, awaiting response... 200 OK
Length: 360 [application/x-sh]
Saving to: 'JSLAB4_1_Linux.sh'

100%[=====>] 360          ---K/s   in 0s

2015-05-01 19:07:48 (64.2 MB/s) - 'JSLAB4_1_Linux.sh' saved [360/360]

iu@bielinux[srp017156] ls                                     [ 7:07午後 ]
JSLAB4_1_Linux.sh  SRR616268_1.fastq.bz2  SRR616268_2.fastq.bz2
iu@bielinux[srp017156] file JSLAB4_1_Linux.sh               [ 7:08午後 ]
JSLAB4_1_Linux.sh: ASCII text
iu@bielinux[srp017156] █                                     [ 7:08午後 ]
```

W6-1: 確認

②パーミッションの変更(実行権限の付与)と④中身の確認。⑤全部で4行からなることがわかる。

```
iu@bielinux[srp017156] ls -lh [ 7:39午後 ]
total 14G
-rw-rw-r-- 1 iu iu 360 5月 1 18:01 JSLAB4_1_Linux.sh
-rw-rw-r-- 1 iu iu 7.2G 12月 11 14:51 SRR616268_1.fastq.bz2
-rw-rw-r-- 1 iu iu 6.6G 12月 12 14:59 SRR616268_2.fastq.bz2
iu@bielinux[srp017156] chmod 755 JSLAB4_1_Linux.sh [ 7:43午後 ]
iu@bielinux[srp017156] ls -lh [ 7:43午後 ]
total 14G
-rwxr-xr-x 1 iu iu 360 5月 1 18:01 JSLAB4_1_Linux.sh
-rw-rw-r-- 1 iu iu 7.2G 12月 11 14:51 SRR616268_1.fastq.bz2
-rw-rw-r-- 1 iu iu 6.6G 12月 12 14:59 SRR616268_2.fastq.bz2
iu@bielinux[srp017156] more JSLAB4_1_Linux.sh [ 7:43午後 ]
#wget -c ftp://ftp.ddbj.nig.ac.jp/ddbj_database/dra/fastq/SRA061/SRA06148
3/SRX204226/SRR616268_1.fastq.bz2
#wget -c ftp://ftp.ddbj.nig.ac.jp/ddbj_database/dra/fastq/SRA061/SRA06148
3/SRX204226/SRR616268_2.fastq.bz2
bzip2 -dc SRR616268_1.fastq.bz2 | head -n 4000000 > SRR616268sub_1.fastq
bzip2 -dc SRR616268_2.fastq.bz2 | head -n 4000000 > SRR616268sub_2.fastq
iu@bielinux[srp017156] wc JSLAB4_1_Linux.sh [ 7:43午後 ]
 4 24 360 JSLAB4_1_Linux.sh
iu@bielinux[srp017156] [ 7:44午後 ]
```

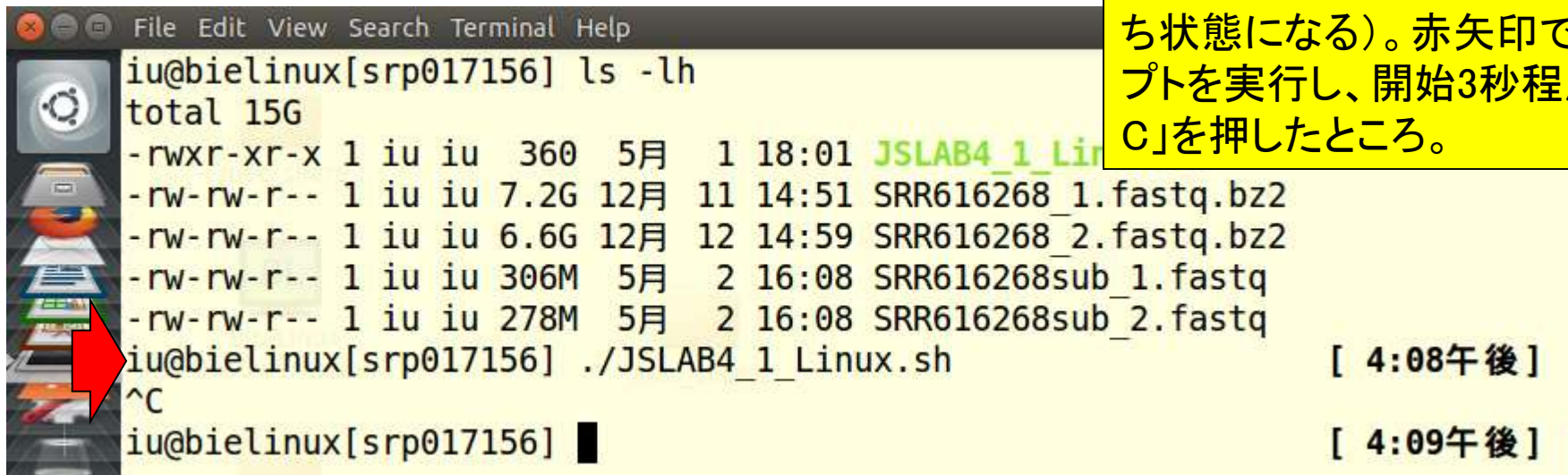
W6-2: シェルスクリプト

第4回原稿の図2と基本的に同じ。① JSLAB4_1_Linux.shの実行権限を確認。②moreで中身を確認。③シェルスクリプトとして実行。④出力ファイルを確認。

```
iu@bielinux[srp017156] ls -lh ← ① [ 7:51午後 ]
total 14G
-rwxr-xr-x 1 iu iu 360 5月 1 18:01 JSLAB4_1_Linux.sh
-rw-rw-r-- 1 iu iu 7.2G 12月 11 14:51 SRR616268_1.fastq.bz2
-rw-rw-r-- 1 iu iu 6.6G 12月 12 14:59 SRR616268_2.fastq.bz2
iu@bielinux[srp017156] more JSLAB4_1_Linux.sh ← ② [ 7:51午後 ]
#wget -c ftp://ftp.ddbj.nig.ac.jp/ddbj_database/dra/fastq/SRA061/SRA06148
3/SRX204226/SRR616268_1.fastq.bz2
#wget -c ftp://ftp.ddbj.nig.ac.jp/ddbj_database/dra/fastq/SRA061/SRA06148
3/SRX204226/SRR616268_2.fastq.bz2
bzip2 -dc SRR616268_1.fastq.bz2 | head -n 4000000 > SRR616268sub_1.fastq
bzip2 -dc SRR616268_2.fastq.bz2 | head -n 4000000 > SRR616268sub_2.fastq
iu@bielinux[srp017156] time ./JSLAB4_1_Linux.sh ← ③ [ 7:51午後 ]
./JSLAB4_1_Linux.sh 30.34s user 3.90s system 71% cpu 47.926 total
iu@bielinux[srp017156] ls -lh ← ④ [ 7:52午後 ]
total 15G
-rwxr-xr-x 1 iu iu 360 5月 1 18:01 JSLAB4_1_Linux.sh
-rw-rw-r-- 1 iu iu 7.2G 12月 11 14:51 SRR616268_1.fastq.bz2
-rw-rw-r-- 1 iu iu 6.6G 12月 12 14:59 SRR616268_2.fastq.bz2
-rw-rw-r-- 1 iu iu 306M 5月 1 19:52 SRR616268sub_1.fastq
-rw-rw-r-- 1 iu iu 278M 5月 1 19:52 SRR616268sub_2.fastq
iu@bielinux[srp017156] [ 7:52午後 ]
```

W6-3: 実行を中断したい

「CTRL + C」(CTRLキーを押しながらCキーを押す)。lessコマンドなどは、qキーを押すと復帰できる(コマンド入力待ち状態になる)。赤矢印でシェルスクリプトを実行し、開始3秒程度で「CTRL + C」を押したところ。



```
File Edit View Search Terminal Help
iu@bielinux[srp017156] ls -lh
total 15G
-rwxr-xr-x 1 iu iu 360 5月 1 18:01 JSLAB4_1_Lin
-rw-rw-r-- 1 iu iu 7.2G 12月 11 14:51 SRR616268_1.fastq.bz2
-rw-rw-r-- 1 iu iu 6.6G 12月 12 14:59 SRR616268_2.fastq.bz2
-rw-rw-r-- 1 iu iu 306M 5月 2 16:08 SRR616268sub_1.fastq
-rw-rw-r-- 1 iu iu 278M 5月 2 16:08 SRR616268sub_2.fastq
iu@bielinux[srp017156] ./JSLAB4_1_Linux.sh [ 4:08午後]
^C
iu@bielinux[srp017156] [ 4:09午後]
```

W6-3: 実行を中断したい

全部で30秒程度で2つのファイルの処理が終わることがW6-2の結果から既知。それゆえSRR616268sub_1.fastqファイル生成段階で計算が中断されている。中断しなかった場合と中断した場合ではファイルサイズが異なることが分かる。この場合は、答えが306MBになっていなければいけないので33MBが変だとわかる。

```
iu@bielinux[srp017156] ls -lh
total 15G
-rwxr-xr-x 1 iu iu 360 5月 1 18:01 JSLAB4_1_Lin
-rw-rw-r-- 1 iu iu 7.2G 12月 11 14:51 SRR616268_1.
-rw-rw-r-- 1 iu iu 6.6G 12月 12 14:59 SRR616268_2.
-rw-rw-r-- 1 iu iu 306M 5月 2 16:08 SRR616268sub
-rw-rw-r-- 1 iu iu 278M 5月 2 16:08 SRR616268sub
iu@bielinux[srp017156] ./JSLAB4_1_Linux.sh
^C
iu@bielinux[srp017156] ls -lh
total 14G
-rwxr-xr-x 1 iu iu 360 5月 1 18:01 JSLAB4_1_Linux.sh
-rw-rw-r-- 1 iu iu 7.2G 12月 11 14:51 SRR616268_1.fastq.bz2
-rw-rw-r-- 1 iu iu 6.6G 12月 12 14:59 SRR616268_2.fastq.bz2
-rw-rw-r-- 1 iu iu 33M 5月 2 16:08 SRR616268sub_1.fastq
-rw-rw-r-- 1 iu iu 278M 5月 2 16:08 SRR616268sub_2.fastq
iu@bielinux[srp017156] █
```

[4:08午後]

[4:09午後]

[4:11午後]

W6-3: 実行を中断したい

仮に、答えがわかっていなくても入力ファイルのサイズ比からも想像がつく。つまり、2つの入力ファイルのサイズ比から大きくかけ離れたサイズ比になることはないだろうという前提に立てばよい。このあたりは多くの実害を被りながら経験値を積む以外にないかも…。

```
File Edit View Search Terminal Help
iu@bielinux[srp017156] ls -lh
total 15G
-rwxr-xr-x 1 iu iu 360 5月 1 18:01 JSLAB4_1_Lin
-rw-rw-r-- 1 iu iu 7.2G 12月 11 14:51 SRR616268_1.
-rw-rw-r-- 1 iu iu 6.6G 12月 12 14:59 SRR616268_2.fastq.bz2
-rw-rw-r-- 1 iu iu 306M 5月 2 16:08 SRR616268sub_1.fastq
-rw-rw-r-- 1 iu iu 278M 5月 2 16:08 SRR616268sub_2.fastq
iu@bielinux[srp017156] ./JSLAB4_1_Linux.sh
^C
iu@bielinux[srp017156] ls -lh
total 14G
-rwxr-xr-x 1 iu iu 360 5月 1 18:01 JSLAB4_1_Linux.sh
-rw-rw-r-- 1 iu iu 7.2G 12月 11 14:51 SRR616268_1.fastq.bz2
-rw-rw-r-- 1 iu iu 6.6G 12月 12 14:59 SRR616268_2.fastq.bz2
-rw-rw-r-- 1 iu iu 33M 5月 2 16:08 SRR616268sub_1.fastq
-rw-rw-r-- 1 iu iu 278M 5月 2 16:08 SRR616268sub_2.fastq
iu@bielinux[srp017156] █
```

[4:08午後]

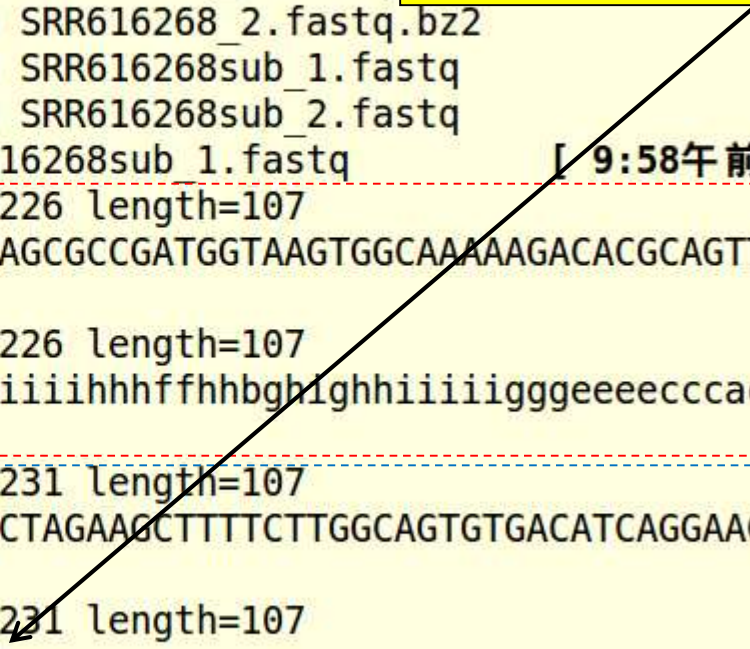
[4:09午後]

[4:11午後]

W6-4: tailで確認・削除

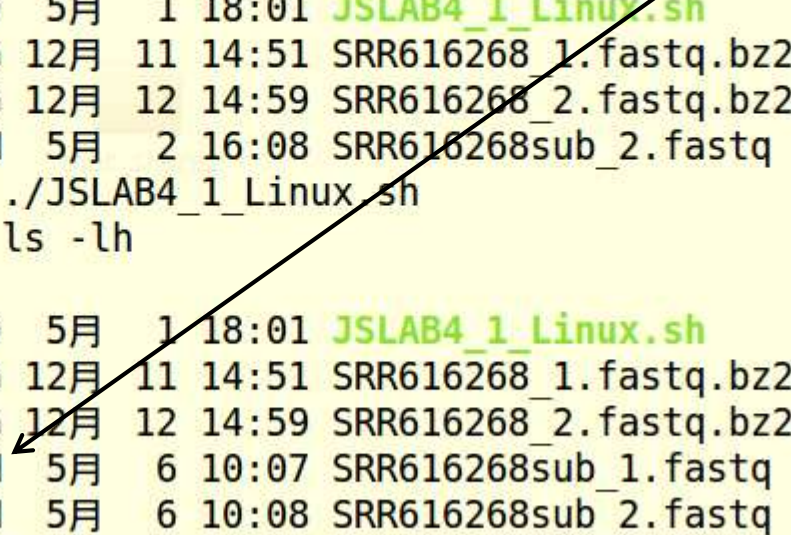
①tailコマンドでファイルの最後の8行分を表示。CTRL + Cで途中で計算を止めたため、(偶然)行数としては4の倍数ではあるものの、一文字表記のクオリティスコア部分が途中で切れていることがわかる。②このファイルを削除。

```
iu@bielinux[srp017156] ls -lh
total 14G
-rwxr-xr-x 1 iu iu 360 5月 1 18:01 JSLAB4_1_Linux.sh
-rw-rw-r-- 1 iu iu 7.2G 12月 11 14:51 SRR616268_1.fastq
-rw-rw-r-- 1 iu iu 6.6G 12月 12 14:59 SRR616268_2.fastq.bz2
-rw-rw-r-- 1 iu iu 33M 5月 2 16:08 SRR616268sub_1.fastq
-rw-rw-r-- 1 iu iu 278M 5月 2 16:08 SRR616268sub_2.fastq
iu@bielinux[srp017156] tail -n 8 SRR616268sub_1.fastq [ 9:58午前 ]
@SRR616268.107756 2291:6:1101:8383:12226 length=107
TCTGGCTCAACATCAACATTGGATACTACAGTTGGATAGCGCCGATGGTAAGTGGCAAAAAGACACGCAGTTCC
TATTAATCACGATGACAACTCAGTCACGGGAT
+SRR616268.107756 2291:6:1101:8383:12226 length=107
bbbeeeeegggggiiiiiiiiiiiiiihiiiihhiighiiiihhffhbbghighhiiiiigggeeecccccacc
cbccddcccccccccccc^bcccc]bbbeeee
@SRR616268.107757 2291:6:1101:8492:12231 length=107
CCTGTGTCGGTTTGCGGTACGGGTAGTTTATTTCTCACTAGAAGCTTTTCTTGGCAGTGTGACATCAGGAACTT
CGCTACTATAATTCGCTCCCATCACAGGAGG
+SRR616268.107757 2291:6:1101:8492:12231 length=107
bbecdcegggggiiihbfhhibfghffgghiiiih%
iu@bielinux[srp017156] rm -f SRR616268sub_1.fastq [ 9:58午前 ]
iu@bielinux[srp017156] █ [ 9:59午前 ]
```



W6-5: サブセット抽出を再実行

```
iu@bielinux[srp017156] pwd [ 9:59午前 ]
/home/iu/Documents/srp017156
iu@bielinux[srp017156] ls -lh [10:07午前 ]
total 14G
-rwxr-xr-x 1 iu iu 360 5月 1 18:01 JSLAB4_1_Linux.sh
-rw-rw-r-- 1 iu iu 7.2G 12月 11 14:51 SRR616268_1.fastq.bz2
-rw-rw-r-- 1 iu iu 6.6G 12月 12 14:59 SRR616268_2.fastq.bz2
-rw-rw-r-- 1 iu iu 278M 5月 2 16:08 SRR616268sub_2.fastq
iu@bielinux[srp017156] ./JSLAB4_1_Linux.sh [10:07午前 ]
iu@bielinux[srp017156] ls -lh [10:08午前 ]
total 15G
-rwxr-xr-x 1 iu iu 360 5月 1 18:01 JSLAB4_1_Linux.sh
-rw-rw-r-- 1 iu iu 7.2G 12月 11 14:51 SRR616268_1.fastq.bz2
-rw-rw-r-- 1 iu iu 6.6G 12月 12 14:59 SRR616268_2.fastq.bz2
-rw-rw-r-- 1 iu iu 306M 5月 6 10:07 SRR616268sub_1.fastq
-rw-rw-r-- 1 iu iu 278M 5月 6 10:08 SRR616268sub_2.fastq
iu@bielinux[srp017156] [10:13午前 ]
```



W7-1 : FastQC

想定外の状態になってしまった場合の対処法。①「fastqc」と打ってから、間違っ
てそのままリターンキーを押すと、②
右下のような画面になってしまう。この
状態から脱出するには…。

```
iu@bielinux[srp017156] pwd
/home/iu/Documents/srp017156
iu@bielinux[srp017156] ls -lh
total 15G
-rwxr-xr-x 1 iu iu 360 5月 1 18:01 JSLAB4_1_Linux.sh
-rw-rw-r-- 1 iu iu 7.2G 12月 11 14:51 SRR616268_1.fastq.bz2
-rw-rw-r-- 1 iu iu 6.6G 12月 12 14:59 SRR616268_2.fastq.bz2
-rw-rw-r-- 1 iu iu 306M 5月 6 10:07 SRR616268sub_1.fastq
-rw-rw-r-- 1 iu iu 278M 5月 6 10:08 SRR616268sub_2.fastq
iu@bielinux[srp017156] fastqc
```

[10:15午前]

[10:16午前]

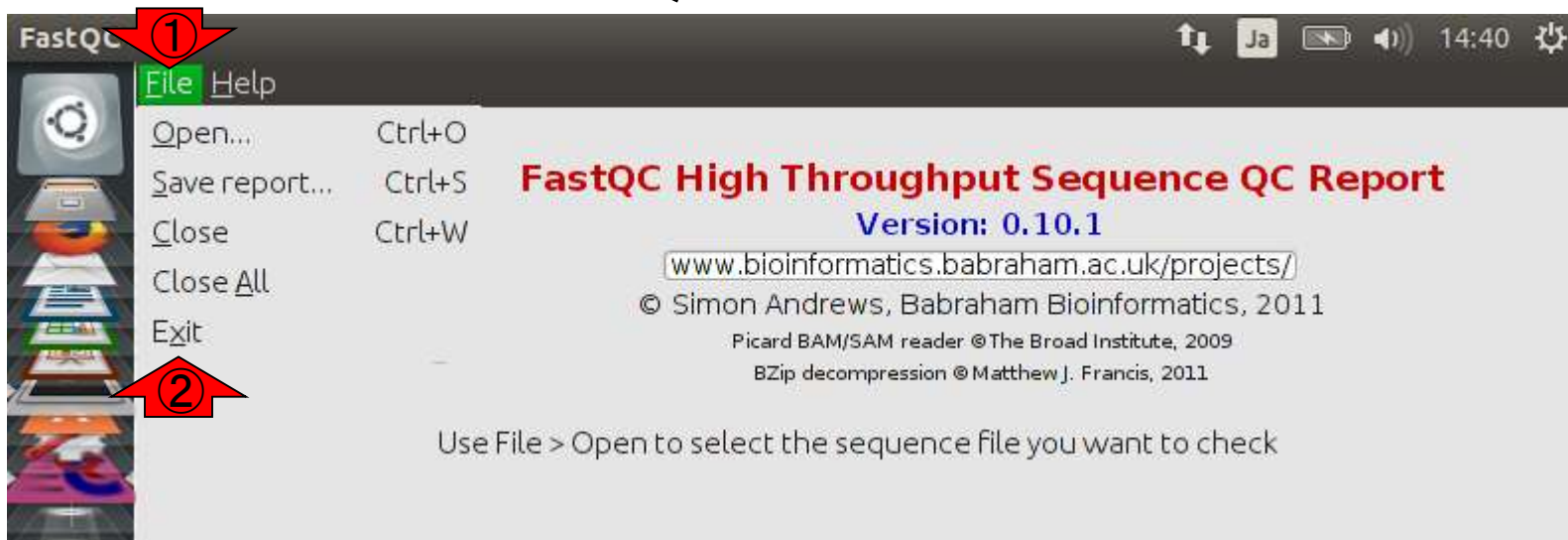
①

②



W7-1 : FastQC

この状態から脱出するには、「File - Exit」。
右下は復帰後のターミナル画面。



```
File Edit View Search Terminal Help
-rw-rw-r-- 1 iu iu 278M  5月  6 10:08 SRR616268sub_2.fastq
iu@bielinux[srp017156] pwd [10:15午前]
/home/iu/Documents/srp017156
iu@bielinux[srp017156] ls -lh [10:15午前]
total 15G
-rwxr-xr-x 1 iu iu  360  5月  1 18:01 JSLAB4_1_Linux.sh
-rw-rw-r-- 1 iu iu 7.2G 12月 11 14:51 SRR616268_1.fastq.bz2
-rw-rw-r-- 1 iu iu 6.6G 12月 12 14:59 SRR616268_2.fastq.bz2
-rw-rw-r-- 1 iu iu 306M  5月  6 10:07 SRR616268sub_1.fastq
-rw-rw-r-- 1 iu iu 278M  5月  6 10:08 SRR616268sub_2.fastq
iu@bielinux[srp017156] fastqc [10:16午前]
iu@bielinux[srp017156] [10:17午前]
```

W7-2: FastQC

```
iu@bielinux[srp017156] ls -lh [10:17午前]
total 15G
-rwxr-xr-x 1 iu iu 360 5月 1 18:01 JSLAB4_1_Linux.sh
-rw-rw-r-- 1 iu iu 7.2G 12月 11 14:51 SRR616268_1.fastq.bz2
-rw-rw-r-- 1 iu iu 6.6G 12月 12 14:59 SRR616268_2.fastq.bz2
-rw-rw-r-- 1 iu iu 306M 5月 6 10:07 SRR616268sub_1.fastq
-rw-rw-r-- 1 iu iu 278M 5月 6 10:08 SRR616268sub_2.fastq
iu@bielinux[srp017156] fastqc [10:18午前]
iu@bielinux[srp017156] pwd [10:18午前]
/home/iu/Documents/srp017156
iu@bielinux[srp017156] ls -lh SRR616268sub_* [10:18午前]
-rw-rw-r-- 1 iu iu 306M 5月 6 10:07 SRR616268sub_1.fastq
-rw-rw-r-- 1 iu iu 278M 5月 6 10:08 SRR616268sub_2.fastq
iu@bielinux[srp017156] fastqc [10:18午前]
```



W7-2: FastQC

基本的な使い方は「fastqc 任意のファイル名」。「SRR」まで打ってからTabキーを押すなどのタブ補完を利用するのも一つだが、他のやり方として、入力したいファイル名を反転させ、①右クリックでコピー、②右クリックでペースト、というのも一般的です。

```
iu@bielinux[~/Documents/srp017156]
iu@bielinux[srp017156] ls -lh
total 15G
-rwxr-xr-x 1 iu iu 360 5月 1 18:01 JSLAB4_1
-rw-rw-r-- 1 iu iu 7.2G 12月 11 14:51 SRR616268_1.fastq.bz2
-rw-rw-r-- 1 iu iu 6.6G 12月 12 14:59 SRR616268_2.fastq.bz2
-rw-rw-r-- 1 iu iu 306M 5月 6 10:07 SRR616268sub_1.fastq
-rw-rw-r-- 1 iu iu 278M 5月 6 10:08 SRR616268sub_2.fastq
iu@bielinux[srp017156] fastqc
iu@bielinux[srp017156] pwd
/home/iu/Documents/srp017156
iu@bielinux[srp017156] ls -lh SRR616268sub *
-rw-rw-r-- 1 iu iu 306M 5月 6 10:07 SRR616268sub_1.fastq
-rw-rw-r-- 1 iu iu 278M 5月 6 10:08 SRR616268sub_2.fastq
iu@bielinux[srp017156] fastqc
```

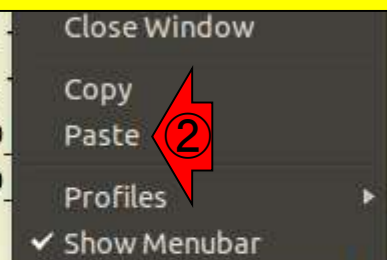


[10:21午前]

W7-2: FastQC

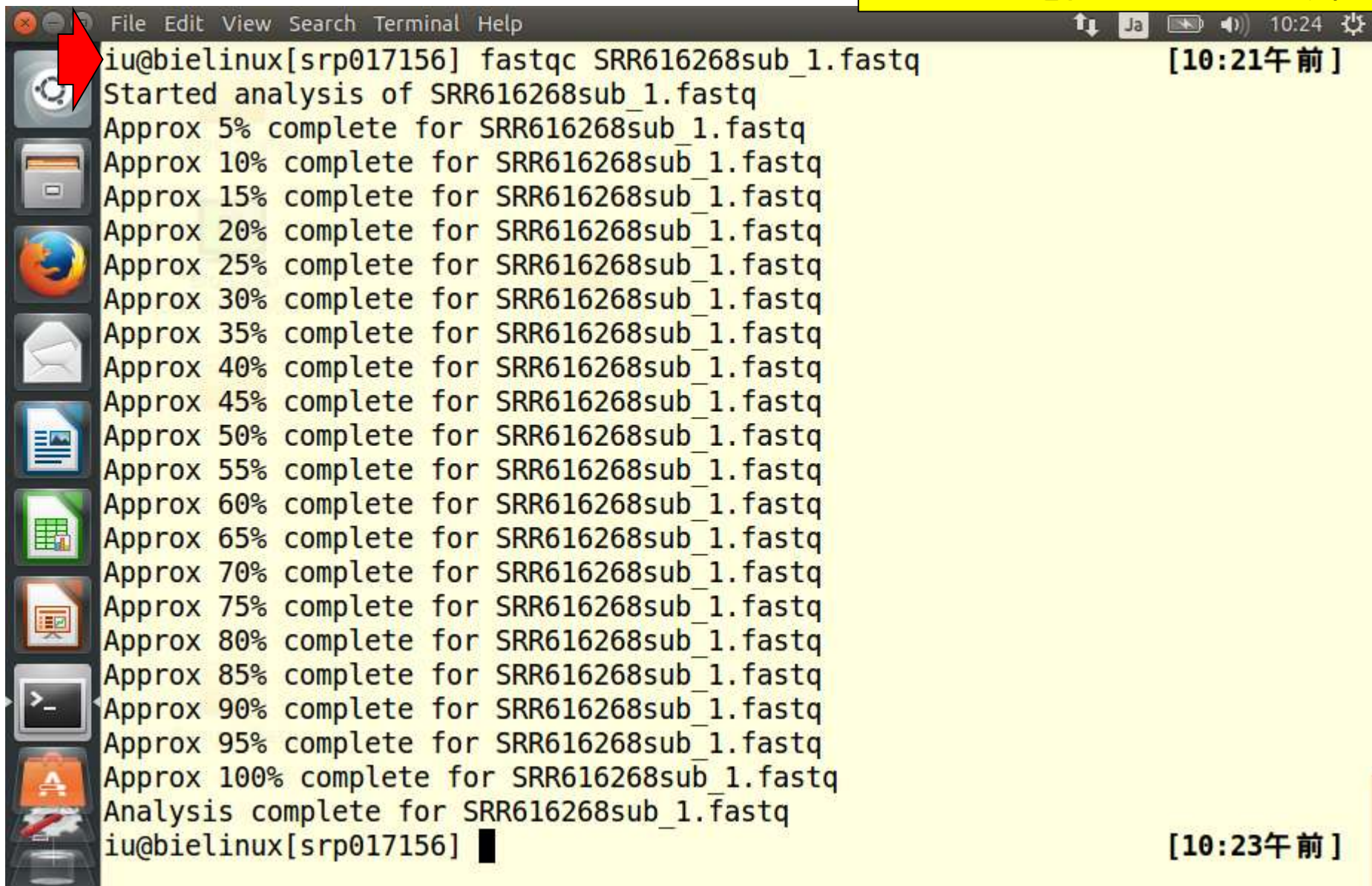
基本的な使い方は「fastqc 任意のファイル名」。「SRR」まで打ってからTabキーを押すなどのタブ補完を利用するのも一つだが、他のやり方として、入力したいファイル名を反転させ、①右クリックでコピー、②右クリックでペースト、というのも一般的です。

```
iu@bielinux[~/Documents/srp017156]
iu@bielinux[srp017156] ls -lh
total 15G
-rwxr-xr-x 1 iu iu 360 5月 1 18:01 JSLAB4_1
-rw-rw-r-- 1 iu iu 7.2G 12月 11 14:51 SRR616268_1.
-rw-rw-r-- 1 iu iu 6.6G 12月 12 14:59 SRR616268_2.
-rw-rw-r-- 1 iu iu 306M 5月 6 10:07 SRR616268sub
-rw-rw-r-- 1 iu iu 278M 5月 6 10:08 SRR616268sub
iu@bielinux[srp017156] fastqc
iu@bielinux[srp017156] pwd
/home/iu/Documents/srp017156
iu@bielinux[srp017156] ls -lh SRR616268sub *
-rw-rw-r-- 1 iu iu 306M 5月 6 10:07 SRR616268sub_1.fastq
-rw-rw-r-- 1 iu iu 278M 5月 6 10:08 SRR616268sub_2.fastq
iu@bielinux[srp017156] fastqc
```



W7-3: FastQC

入力したいファイル名を反転させ、①右クリックでコピー、②右クリックでペースト後にリターンキーを押した後の状態。約20秒。



```
iu@bielinux[srp017156] fastqc SRR616268sub_1.fastq [10:21午前]
Started analysis of SRR616268sub_1.fastq
Approx 5% complete for SRR616268sub_1.fastq
Approx 10% complete for SRR616268sub_1.fastq
Approx 15% complete for SRR616268sub_1.fastq
Approx 20% complete for SRR616268sub_1.fastq
Approx 25% complete for SRR616268sub_1.fastq
Approx 30% complete for SRR616268sub_1.fastq
Approx 35% complete for SRR616268sub_1.fastq
Approx 40% complete for SRR616268sub_1.fastq
Approx 45% complete for SRR616268sub_1.fastq
Approx 50% complete for SRR616268sub_1.fastq
Approx 55% complete for SRR616268sub_1.fastq
Approx 60% complete for SRR616268sub_1.fastq
Approx 65% complete for SRR616268sub_1.fastq
Approx 70% complete for SRR616268sub_1.fastq
Approx 75% complete for SRR616268sub_1.fastq
Approx 80% complete for SRR616268sub_1.fastq
Approx 85% complete for SRR616268sub_1.fastq
Approx 90% complete for SRR616268sub_1.fastq
Approx 95% complete for SRR616268sub_1.fastq
Approx 100% complete for SRR616268sub_1.fastq
Analysis complete for SRR616268sub_1.fastq
iu@bielinux[srp017156] [10:23午前]
```

W7-4: FastQC

①「ls -lh」実行結果。赤枠部分がfastqcコマンドで作成されたもの。1つはSRR616268sub_1_fastqcという名前のディレクトリ。もう1つはそのzip圧縮ファイル。②「ls -lh SRR616268sub_*」実行結果。ls -lh実行時にディレクトリを指定していることになるので、ディレクトリの中身をリストアップしている。

```
File Edit View Search Terminal Help
iu@bielinux[srp017156] ls -lh
total 15G
-rwxr-xr-x 1 iu iu 360 5月 1 18:01 JS
-rw-rw-r-- 1 iu iu 7.2G 12月 11 14:51 SRR616268_1.fastq.bz2
-rw-rw-r-- 1 iu iu 6.6G 12月 12 14:59 SRR616268_2.fastq.bz2
-rw-rw-r-- 1 iu iu 306M 5月 6 10:07 SRR616268sub_1.fastq
drwxrwxr-x 4 iu iu 4.0K 5月 6 10:23 SRR616268sub_1_fastqc
-rw-rw-r-- 1 iu iu 293K 5月 6 10:23 SRR616268sub_1_fastqc.zip
-rw-rw-r-- 1 iu iu 278M 5月 6 10:08 SRR616268sub_2.fastq
iu@bielinux[srp017156] ls -lh SRR616268sub_* [10:24午前]
-rw-rw-r-- 1 iu iu 306M 5月 6 10:07 SRR616268sub_1.fastq
-rw-rw-r-- 1 iu iu 293K 5月 6 10:23 SRR616268sub_1_fastqc.zip
-rw-rw-r-- 1 iu iu 278M 5月 6 10:08 SRR616268sub_2.fastq

SRR616268sub_1_fastqc:
total 88K
-rw-rw-r-- 1 iu iu 26K 5月 6 10:23 fastqc_data.txt
-rw-rw-r-- 1 iu iu 48K 5月 6 10:23 fastqc_report.html
drwxrwxr-x 2 iu iu 4.0K 5月 6 10:23 Icons
drwxrwxr-x 2 iu iu 4.0K 5月 6 10:23 Images
-rw-rw-r-- 1 iu iu 542 5月 6 10:23 summary.txt
iu@bielinux[srp017156] [10:25午前]
```


W7-5: FastQC

プログラムのマニュアルを眺めるのは、基本的にどのディレクトリ上で行ってもよい。「fastqc -h」と打ってリターン。

```
iu@bielinux[srp017156] pwd [10:25午前]
/home/iu/Documents/srp017156
iu@bielinux[srp017156] ls -lh [10:26午前]
total 15G
-rwxr-xr-x 1 iu iu 360 5月 1 18:01 JSLAB4_1_Linux.sh
-rw-rw-r-- 1 iu iu 7.2G 12月 11 14:51 SRR616268_1.fastq.bz2
-rw-rw-r-- 1 iu iu 6.6G 12月 12 14:59 SRR616268_2.fastq.bz2
-rw-rw-r-- 1 iu iu 306M 5月 6 10:07 SRR616268sub_1.fastq
drwxrwxr-x 4 iu iu 4.0K 5月 6 10:23 SRR616268sub_1_fastqc
-rw-rw-r-- 1 iu iu 293K 5月 6 10:23 SRR616268sub_1_fastqc.zip
-rw-rw-r-- 1 iu iu 278M 5月 6 10:08 SRR616268sub_2.fastq
iu@bielinux[srp017156] fastqc -h [10:26午前]
```



W7-5: FastQC

マニュアルが一気に表示されて何が何だかわからない(爆)。そういうときはパイプ(|)でlessやheadに流し込む。ここではlessを用いる例を示す。

```
File Edit View Search Terminal Help
--contaminants contaminants to screen overrepresented sequences against.
The file must contain sets of named contaminants in the
form name[tab]sequence. Lines prefixed with a hash will
be ignored.

-k --kmers Specifies the length of Kmer to look for in the Kmer content
module. Specified Kmer length must be between 2 and 10. Defau
length is 5 if not specified.

-q --quiet Supress all progress messages on stdout and only report error
s.

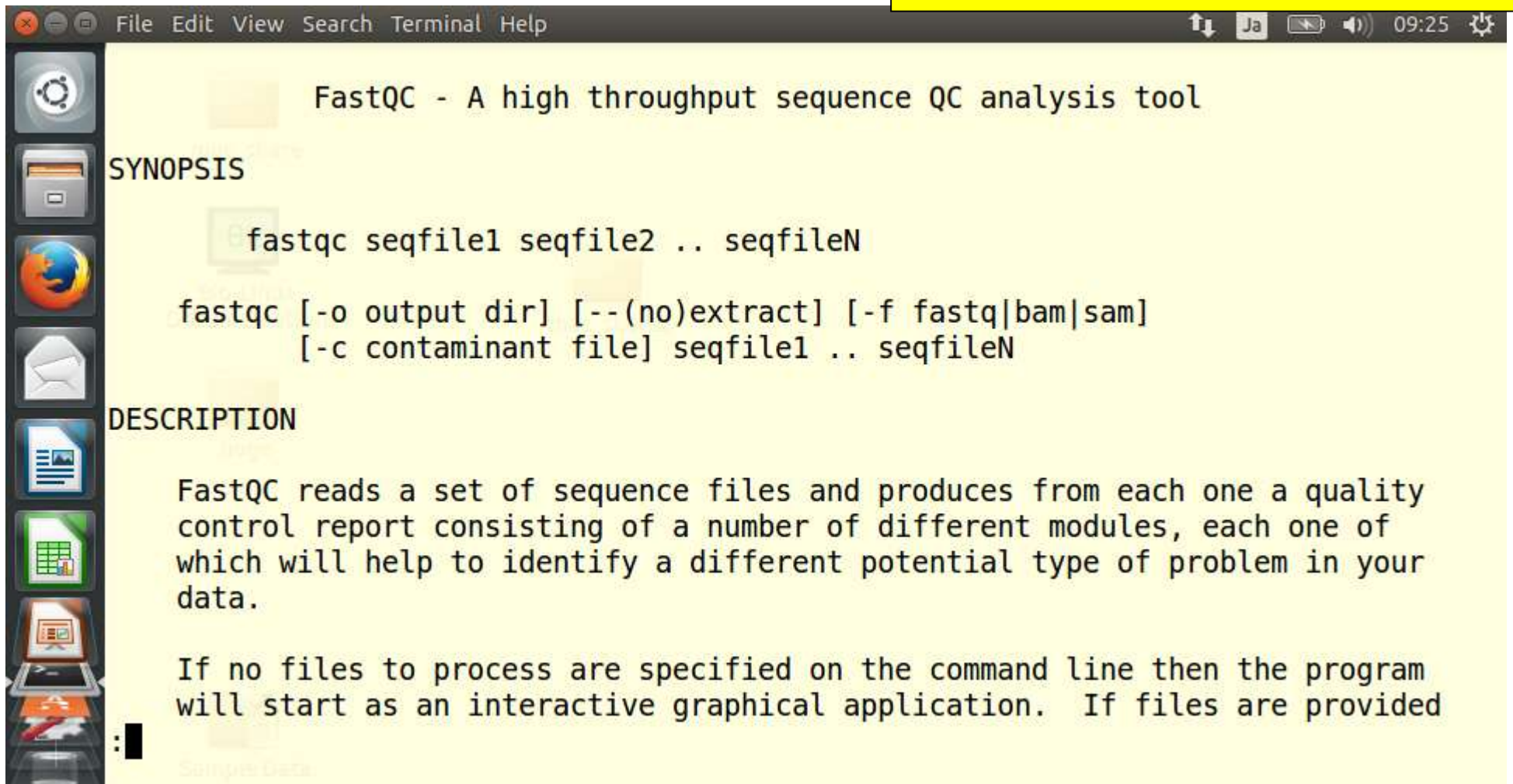
BUGS

Any bugs in fastqc should be reported either to simon.andrews@babraham.ac.uk
or in www.bioinformatics.babraham.ac.uk/bugzilla/

iu@bielinux[srp017156] fastqc -h | less [ 9:23午前]
```

W7-5: FastQC

この状態は、fastqcのマニュアルページをlessで眺めていると解釈すればよい。基本的に、上下矢印キーでページをスクロール。



```
File Edit View Search Terminal Help 09:25

FastQC - A high throughput sequence QC analysis tool

SYNOPSIS

fastqc seqfile1 seqfile2 .. seqfileN

fastqc [-o output dir] [--(no)extract] [-f fastq|bam|sam]
        [-c contaminant file] seqfile1 .. seqfileN

DESCRIPTION

FastQC reads a set of sequence files and produces from each one a quality
control report consisting of a number of different modules, each one of
which will help to identify a different potential type of problem in your
data.

If no files to process are specified on the command line then the program
will start as an interactive graphical application.  If files are provided
:
```

W7-5: FastQC

①進捗状況を非表示にしたい場合は、-qオプションをつければよさそうであることが分かる。②(END)となったら、下矢印キーで一番下まできた状態であることを示す。この状態から抜けるには、qキーを押せばよい。

```
iu@bielinux[~/Documents/srp017156]
--contaminants contaminants to screen over
The file must contain sets of named contaminants in the
form name[tab]sequence. Lines prefixed with a hash will
be ignored.

-k --kmers Specifies the length of Kmer to look for in the Kmer content
module. Specified Kmer length must be between 2 and 10. Defau
length is 5 if not specified.

-q --quiet Supress all progress messages on stdout and only report error
s.

BUGS
Any bugs in fastqc should be reported either to simon.andrews@babraham.ac.uk
or in www.bioinformatics.babraham.ac.uk/bugzilla/

(END) ②
```

①

②

W7-5: FastQC

lessから抜け出した状態。このように、ちょっとした場面で様々なLinuxコマンドを駆使して合理的に作業を行うのが一般的。基本的なコマンドを使いこなせないと対処法を思いつかない。数日程度で習熟できるものではなく、数週間程度はかかるのが一般的。

```
File Edit View Search Terminal Help
The file must contain set
form name[tab]sequence.
be ignored.

-k --kmers          Specifies the length of Kmer to look for in the Kmer content
                    module. Specified Kmer length must be between 2 and 10. Defau
                    length is 5 if not specified.

-q --quiet          Supress all progress messages on stdout and only report error

BUGS

Any bugs in fastqc should be reported either to simon.andrews@babraham.ac.uk
or in www.bioinformatics.babraham.ac.uk/bugzilla/

iu@bielinux[srp017156] fastqc -h | less
iu@bielinux[srp017156] █
```

[9:23午前]

[9:30午前]

W7-6: FastQC

①「-q」オプションをつけてfastqcを実行。確かにW7-3と違って進捗状況を非表示にできていることがわかる。②赤枠のように実行結果として「ディレクトリ」と「同じ内容のzipファイル」がちゃんと生成されている。このようにしてオプション利用の幅を広げていく。

```
iu@bielinux[srp017156] fastqc -h | less
iu@bielinux[srp017156] ls -lh
total 15G
-rwxr-xr-x 1 iu iu 360 5月 1 18:01 JSLAB4_1_Linux.sh
-rw-rw-r-- 1 iu iu 7.2G 12月 11 14:51 SRR616268_1.fastq.bz2
-rw-rw-r-- 1 iu iu 6.6G 12月 12 14:59 SRR616268_2.fastq.bz2
-rw-rw-r-- 1 iu iu 306M 5月 6 10:07 SRR616268sub_1.fastq
drwxrwxr-x 4 iu iu 4.0K 5月 6 10:23 SRR616268sub_1_fastqc
-rw-rw-r-- 1 iu iu 293K 5月 6 10:23 SRR616268sub_1_fastqc.zip
-rw-rw-r-- 1 iu iu 278M 5月 6 10:08 SRR616268sub_2.fastq
iu@bielinux[srp017156] fastqc -q SRR616268sub_2.fastq [10:27午前]
iu@bielinux[srp017156] ls -lh [10:28午前]
total 15G
-rwxr-xr-x 1 iu iu 360 5月 1 18:01 JSLAB4_1_Linux.sh
-rw-rw-r-- 1 iu iu 7.2G 12月 11 14:51 SRR616268_1.fastq.bz2
-rw-rw-r-- 1 iu iu 6.6G 12月 12 14:59 SRR616268_2.fastq.bz2
-rw-rw-r-- 1 iu iu 306M 5月 6 10:07 SRR616268sub_1.fastq
drwxrwxr-x 4 iu iu 4.0K 5月 6 10:23 SRR616268sub_1_fastqc
-rw-rw-r-- 1 iu iu 293K 5月 6 10:23 SRR616268sub_1_fastqc.zip
-rw-rw-r-- 1 iu iu 278M 5月 6 10:08 SRR616268sub_2.fastq
drwxrwxr-x 4 iu iu 4.0K 5月 6 10:28 SRR616268sub_2_fastqc
-rw-rw-r-- 1 iu iu 243K 5月 6 10:28 SRR616268sub_2_fastqc.zip
iu@bielinux[srp017156]
```

①

②

W7-7: Tips

W7-4や①で示すように、ls -lhだと*fastqc*中にディレクトリを含む場合には、ディレクトリの中身まで表示される。②ディレクトリもファイルと同じように表示したい場合は-dオプションをつければよい。

```
iu@bielinux[srp017156] ls -lh *fastqc* [10:28午前]
-rw-rw-r-- 1 iu iu 293K  5月  6 10:23 SRR616268sub_1_fastqc.zip
-rw-rw-r-- 1 iu iu 243K  5月  6 10:28 SRR616268sub_2_fastqc.zip

SRR616268sub_1_fastqc:
total 88K
-rw-rw-r-- 1 iu iu  26K  5月  6 10:23 fastqc_data.txt
-rw-rw-r-- 1 iu iu  48K  5月  6 10:23 fastqc_report.html
drwxrwxr-x 2 iu iu 4.0K  5月  6 10:23 Icons
drwxrwxr-x 2 iu iu 4.0K  5月  6 10:23 Images
-rw-rw-r-- 1 iu iu  542  5月  6 10:23 summary.txt

SRR616268sub_2_fastqc:
total 56K
-rw-rw-r-- 1 iu iu  16K  5月  6 10:28 fastqc_data.txt
-rw-rw-r-- 1 iu iu  26K  5月  6 10:28 fastqc_report.html
drwxrwxr-x 2 iu iu 4.0K  5月  6 10:28 Icons
drwxrwxr-x 2 iu iu 4.0K  5月  6 10:28 Images
-rw-rw-r-- 1 iu iu  542  5月  6 10:28 summary.txt

iu@bielinux[srp017156] ls -lhd *fastqc* [10:30午前]
drwxrwxr-x 4 iu iu 4.0K  5月  6 10:23 SRR616268sub_1_fastqc
-rw-rw-r-- 1 iu iu 293K  5月  6 10:23 SRR616268sub_1_fastqc.zip
drwxrwxr-x 4 iu iu 4.0K  5月  6 10:28 SRR616268sub_2_fastqc
-rw-rw-r-- 1 iu iu 243K  5月  6 10:28 SRR616268sub_2_fastqc.zip
iu@bielinux[srp017156] [10:30午前]
```

W7-7: Tips

おさらい。rmコマンドの-fは、消してよいかを問い合わせる警告メッセージを表示しない、というオプション。①「rm -f」ではディレクトリは削除できない。②ディレクトリの削除は-rオプションをつける。

```
iu@bielinux[srp017156] ls -lhd *fastqc* [10:30午前]
drwxrwxr-x 4 iu iu 4.0K  5月  6 10:23 SRR616268sub_1_fastqc
-rw-rw-r-- 1 iu iu 293K  5月  6 10:23 SRR616268sub_1_fastqc.zip
drwxrwxr-x 4 iu iu 4.0K  5月  6 10:28 SRR616268sub_2_fastqc
-rw-rw-r-- 1 iu iu 243K  5月  6 10:28 SRR616268sub_2_fastqc.zip
① iu@bielinux[srp017156] rm -f *fastqc* [10:30午前]
rm: cannot remove 'SRR616268sub_1_fastqc': Is a directory
rm: cannot remove 'SRR616268sub_2_fastqc': Is a directory
iu@bielinux[srp017156] ls -lhd *fastqc* [10:31午前]
drwxrwxr-x 4 iu iu 4.0K  5月  6 10:23 SRR616268sub_1_fastqc
drwxrwxr-x 4 iu iu 4.0K  5月  6 10:28 SRR616268sub_2_fastqc
② iu@bielinux[srp017156] rm -rf *fastqc* [10:31午前]
iu@bielinux[srp017156] ls -lhd *fastqc* [10:31午前]
zsh: no matches found: *fastqc*
iu@bielinux[srp017156] ls -lh [10:31午前]
total 15G
-rwxr-xr-x 1 iu iu  360  5月  1 18:01 JSLAB4_1_Linux.sh
-rw-rw-r-- 1 iu iu 7.2G 12月 11 14:51 SRR616268_1.fastq.bz2
-rw-rw-r-- 1 iu iu 6.6G 12月 12 14:59 SRR616268_2.fastq.bz2
-rw-rw-r-- 1 iu iu 306M  5月  6 10:07 SRR616268sub_1.fastq
-rw-rw-r-- 1 iu iu 278M  5月  6 10:08 SRR616268sub_2.fastq
iu@bielinux[srp017156] █ [10:31午前]
```


W7-8: シェルスクリプト

①赤下線のURLからfastqcコマンド実行部分を含むシェルスクリプトファイルをwgetでダウンロード。この際、「fastqc -q」と同様に、「wget -q」とすることで進捗状況を非表示にしている。経験上、コマンド名が違っていてもオプションの意味は同じ場合が多い。②実行権限が付与されていない場合は、chmodで付与しておく。

```
iu@bielinux[srp017156] ls -lh
total 15G
-rwxr-xr-x 1 iu iu 360 5月 1 18:01 JSLAB4_1_Linux.sh
-rw-rw-r-- 1 iu iu 7.2G 12月 11 14:51 SRR616268_sub_1.fastqc
-rw-rw-r-- 1 iu iu 6.6G 12月 12 14:59 SRR616268_sub_2.fastqc
-rw-rw-r-- 1 iu iu 306M 5月 6 10:07 SRR616268sub_1.fastqc
-rw-rw-r-- 1 iu iu 278M 5月 6 10:08 SRR616268sub_2.fastqc
① iu@bielinux[srp017156] wget -cq http://www.iu.a.u-tokyo.ac.jp/~kadota/book/JSLAB4_2_Linux.sh
iu@bielinux[srp017156] ls -lh JSLAB4_* [10:38午前]
-rwxr-xr-x 1 iu iu 360 5月 1 18:01 JSLAB4_1_Linux.sh
-rw-rw-r-- 1 iu iu 424 5月 5 15:07 JSLAB4_2_Linux.sh
② iu@bielinux[srp017156] chmod 755 JSLAB4_2_Linux.sh [10:38午前]
iu@bielinux[srp017156] ls -lh JSLAB4_* [10:38午前]
-rwxr-xr-x 1 iu iu 360 5月 1 18:01 JSLAB4_1_Linux.sh
-rwxr-xr-x 1 iu iu 424 5月 5 15:07 JSLAB4_2_Linux.sh
iu@bielinux[srp017156] [10:38午前]
```

W7-8: シェルスクリプト

- ①wgetで取得したJSLAB4_2_Linux.shの中身を表示。赤枠部分のみが実行される。
- ②実行。
- ③生成されるのは水色枠の4つ。

```
iu@bielinux[srp017156] more JSLAB4_2_Linux.sh [10:44午前]
#wget -c ftp://ftp.ddbj.nig.ac.jp/ddbj_database/dra/fastq/SRA061/SRA061483/SR
X204226/SRR616268_1.fastq.bz2
#wget -c ftp://ftp.ddbj.nig.ac.jp/ddbj_database/dra/fastq/SRA061/SRA061483/SR
X204226/SRR616268_2.fastq.bz2
#bzip2 -dc SRR616268_1.fastq.bz2 | head -n 4000000 > SRR616268sub_1.fastq
#bzip2 -dc SRR616268_2.fastq.bz2 | head -n 4000000 > SRR616268sub_2.fastq
fastqc -q SRR616268sub_1.fastq
fastqc -q SRR616268sub_2.fastq
iu@bielinux[srp017156] sh JSLAB4_2_Linux.sh [10:44午前]
iu@bielinux[srp017156] ls -lhd SRR616268sub_* [10:45午前]
-rw-rw-r-- 1 iu iu 306M 5月 6 10:07 SRR616268sub_1.fastq
drwxrwxr-x 4 iu iu 4.0K 5月 6 10:45 SRR616268sub_1_fastqc
-rw-rw-r-- 1 iu iu 293K 5月 6 10:45 SRR616268sub_1_fastqc.zip
-rw-rw-r-- 1 iu iu 278M 5月 6 10:08 SRR616268sub_2.fastq
drwxrwxr-x 4 iu iu 4.0K 5月 6 10:45 SRR616268sub_2_fastqc
-rw-rw-r-- 1 iu iu 243K 5月 6 10:45 SRR616268sub_2_fastqc.zip
iu@bielinux[srp017156] [10:47午前]
```

③

W7-9: 共有フォルダにコピー

FastQC実行結果のzip
ファイルを共有フォルダ
(~/Desktop/mac_share)
にコピー。

```
iu@bielinux[srp017156] ls -lhd SRR616268sub_*
-rw-rw-r-- 1 iu iu 306M  5月  6 10:07 SRR616268sub_1.fastq
drwxrwxr-x 4 iu iu 4.0K  5月  6 10:45 SRR616268sub_1_fastqc
-rw-rw-r-- 1 iu iu 293K  5月  6 10:45 SRR616268sub_1_fastqc.zip
-rw-rw-r-- 1 iu iu 278M  5月  6 10:08 SRR616268sub_2.fastq
drwxrwxr-x 4 iu iu 4.0K  5月  6 10:45 SRR616268sub_2_fastqc
-rw-rw-r-- 1 iu iu 243K  5月  6 10:45 SRR616268sub_2_fastqc.zip
iu@bielinux[srp017156] cp *.zip ~/Desktop/mac_share
iu@bielinux[srp017156] ls -lh ~/Desktop/mac_share
total 4.0M
-rwxrwxrwx 1 root root 686K 11月 14 17:48 hoge.png
-rwxrwxrwx 1 root root 438  4月 24 13:50 JSLAB3_Lcasei_Linux.txt
-rwxrwxrwx 1 root root 360  5月  1 18:01 JSLAB4_1_Linux.sh
-rwxrwxrwx 1 root root 364  4月 29 11:20 JSLAB4_1.sh
-rwxrwxrwx 1 root root 2.8M 12月  1 23:13 Lactobacillus_casei_12a.GCA_0003095
65.1.22.dna.toplevel.fa
-rwxrwxrwx 1 root root 293K  5月  6 10:49 SRR616268sub_1_fastqc.zip
-rwxrwxrwx 1 root root 243K  5月  6 10:49 SRR616268sub_2_fastqc.zip
iu@bielinux[srp017156]
```

[10:45午前]

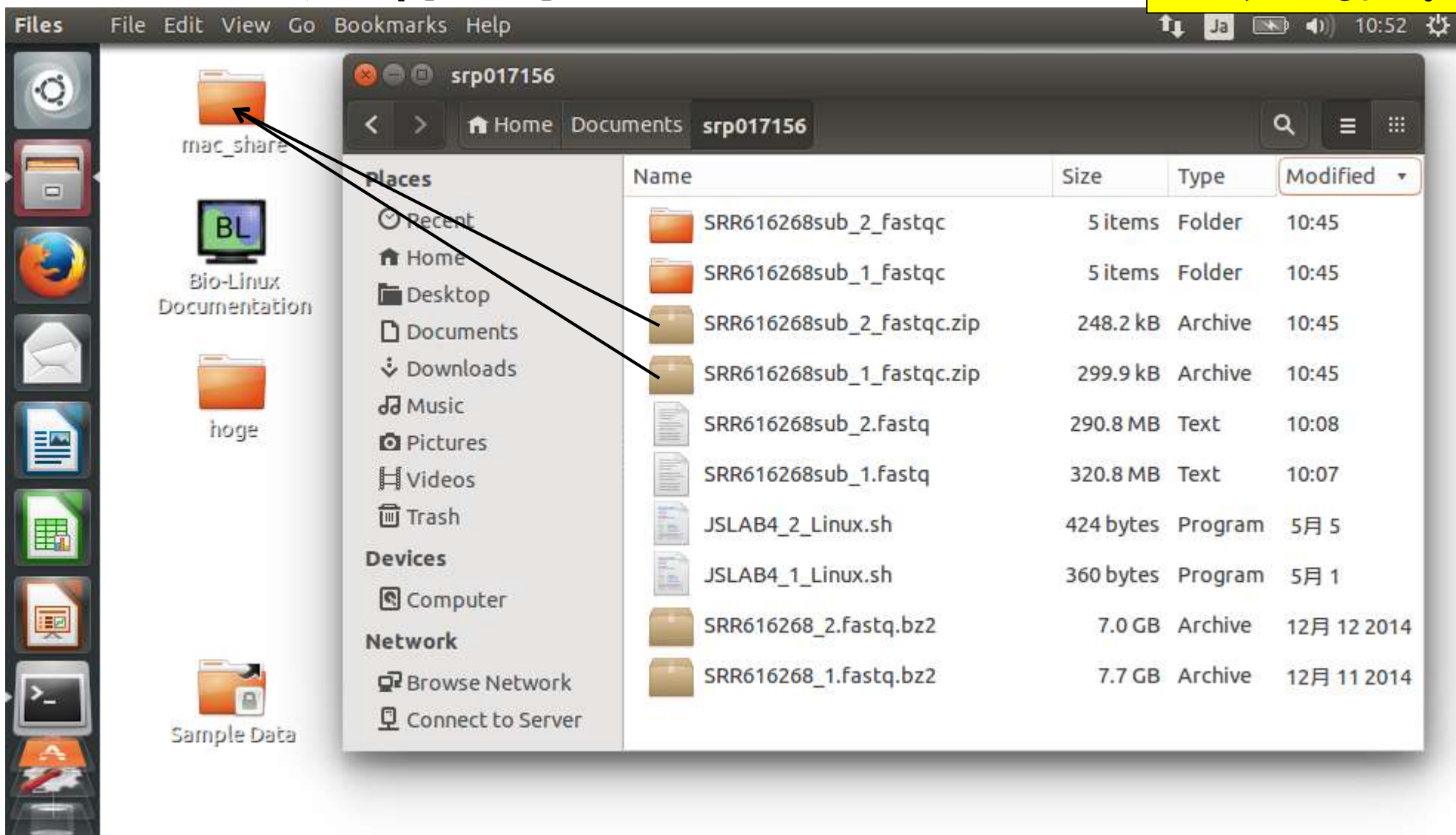
[10:47午前]

[10:49午前]

[10:49午前]

W7-9: 共有フォルダにコピー

別の手段。GUIベースでマウスを用いたドラッグ&ドロップでもよい。



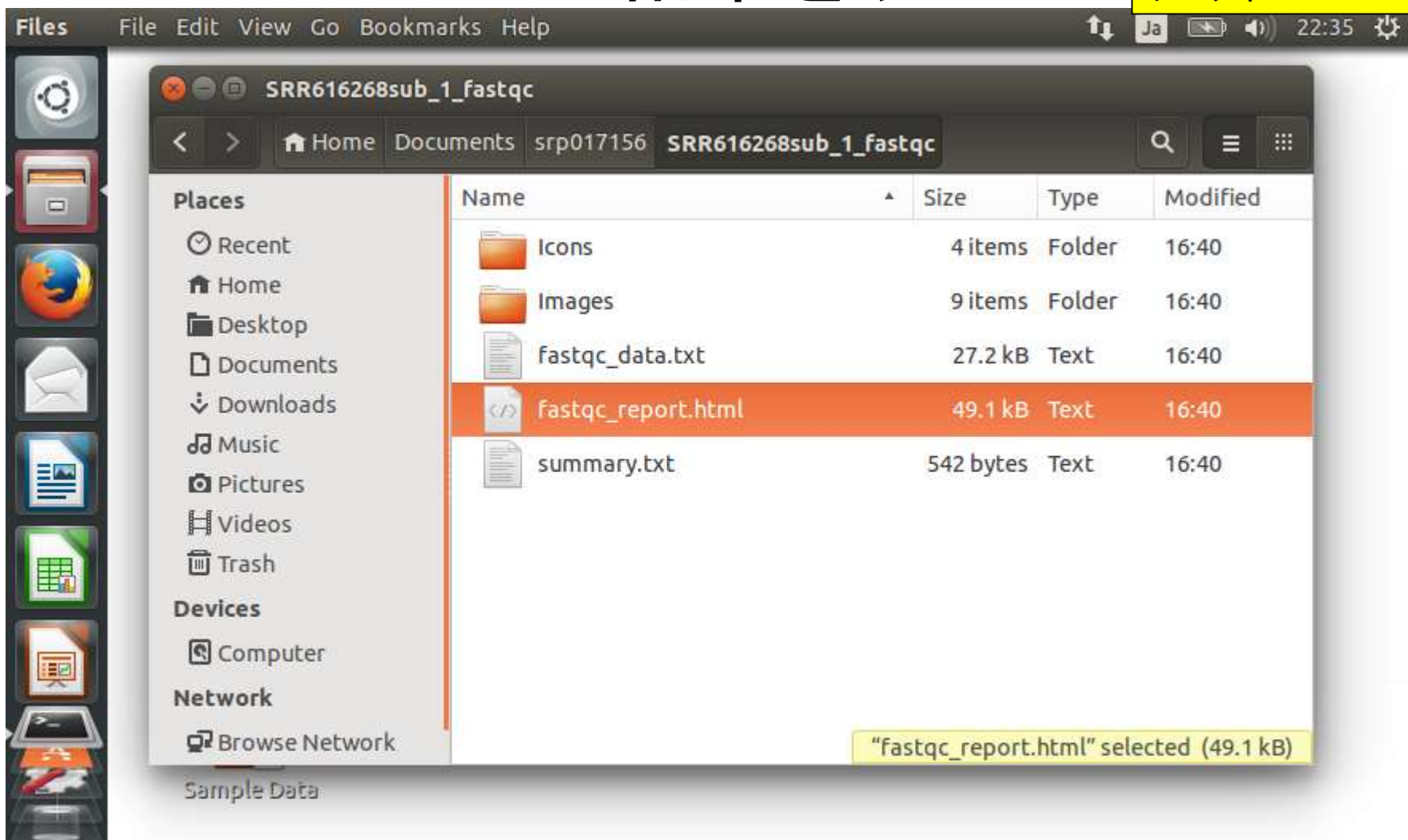
W7-10:バージョン

同じプログラム名(この場合FastQC)でも、様々なバージョンが存在する。「コマンド名 -v」でバージョン情報を見られる場合が多い。FastQCの場合は、①小文字でも②大文字でもどちらでも同じ結果になる。このバージョン情報(FastQC ver. 0.10.1)は忘れずに論文中に記載しましょう。もし -V or -vのいずれでもうまく表示されない場合は③「man コマンド名」または④「コマンド名 -h」でマニュアル中のバージョン情報表示法を調べる

```
iu@bielinux[srp017156] pwd
/home/iu/Documents/srp017156
iu@bielinux[srp017156] ls -lh
total 15G
-rwxr-xr-x 1 iu iu 360 5月 1 18:01 JSLA
-rwxr-xr-x 1 iu iu 424 5月 5 15:07 JSLA
-rw-rw-r-- 1 iu iu 7.2G 12月 11 14:51 SRR6
-rw-rw-r-- 1 iu iu 6.6G 12月 12 14:59 SRR616268_2.fastq.bz2
-rw-rw-r-- 1 iu iu 306M 5月 6 10:07 SRR616268sub_1.fastq
drwxrwxr-x 4 iu iu 4.0K 5月 6 10:45 SRR616268sub_1_fastqc
-rw-rw-r-- 1 iu iu 293K 5月 6 10:45 SRR616268sub_1_fastqc.zip
-rw-rw-r-- 1 iu iu 278M 5月 6 10:08 SRR616268sub_2.fastq
drwxrwxr-x 4 iu iu 4.0K 5月 6 10:45 SRR616268sub_2_fastqc
-rw-rw-r-- 1 iu iu 243K 5月 6 10:45 SRR616268sub_2_fastqc.zip
① iu@bielinux[srp017156] fastqc -v [10:54午前]
FastQC v0.10.1
② iu@bielinux[srp017156] fastqc -V [10:55午前]
FastQC v0.10.1
③ iu@bielinux[srp017156] man fastqc [10:55午前]
No manual entry for fastqc
See 'man 7 undocumented' for help when manual pages are not available.
④ iu@bielinux[srp017156] fastqc -h [10:55午前]
```

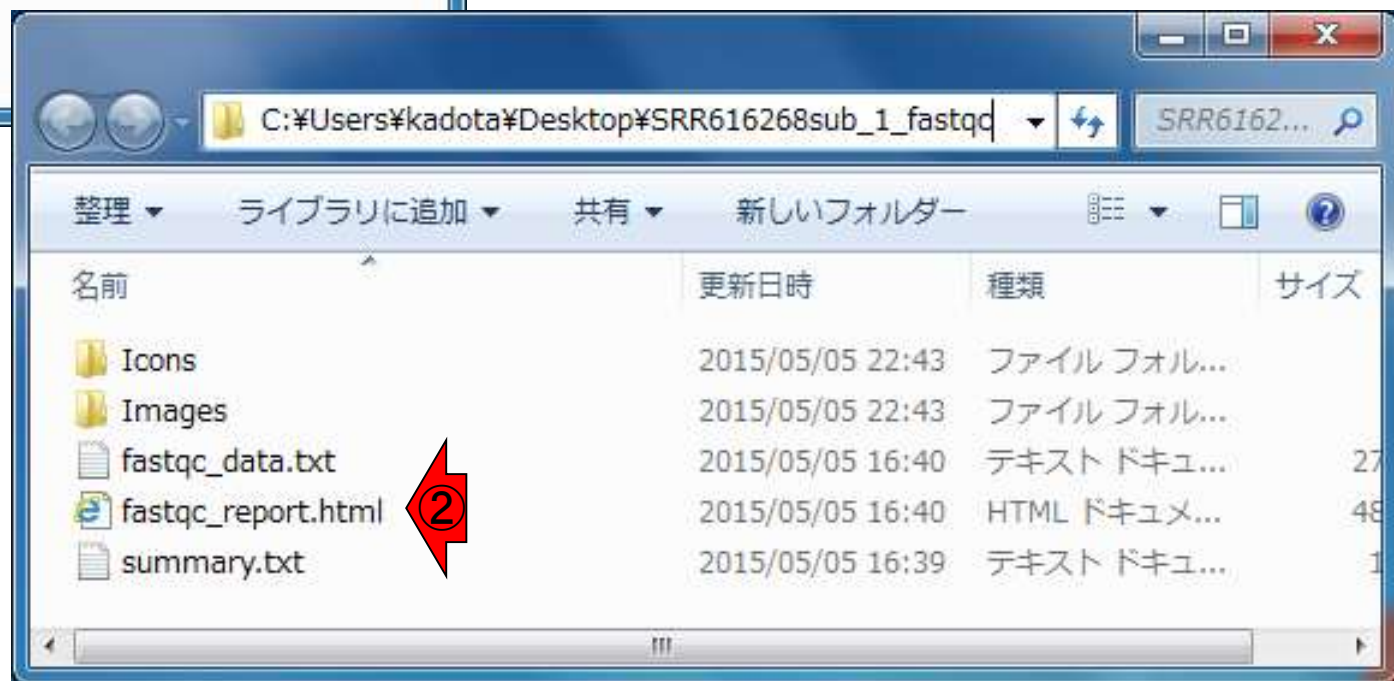
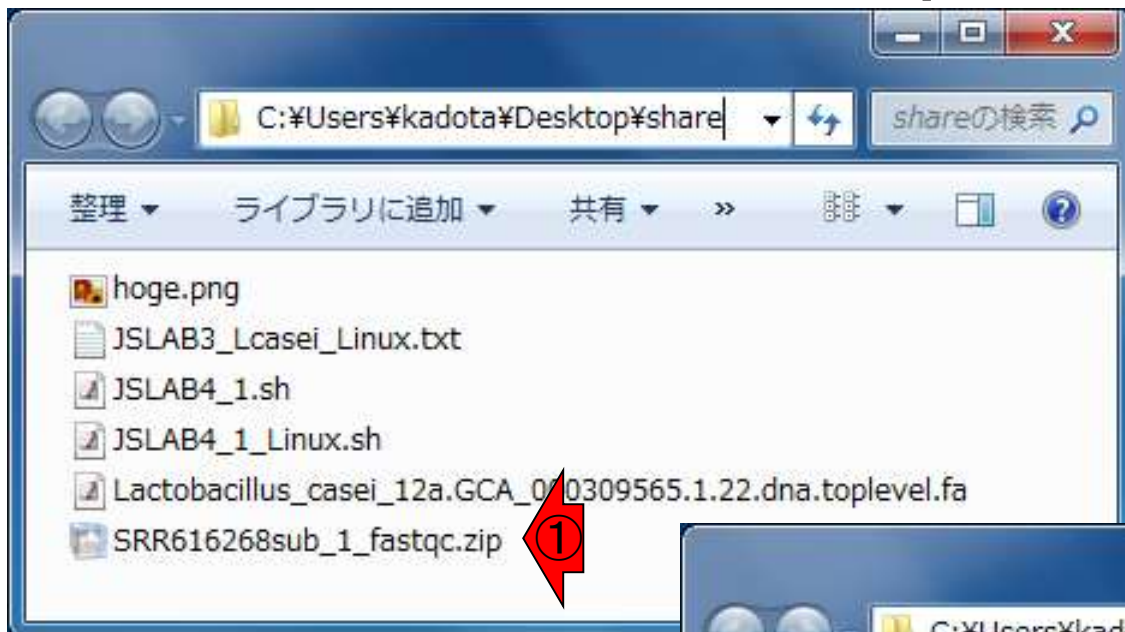
W8-1: FastQC結果を眺める

ゲストOS上で眺める場合は、SRR616268sub_1_fastqcディレクトリ中のhtmlファイルを開く。



W8-2: FastQC結果を眺める

ホストOS上で眺める場合は、共有フォルダ(デスクトップ - share)に置いたzipファイルを①解凍し、②htmlファイルを開けばよい。



W8-2: FastQC結果を眺める

この入力ファイルはオリジナル(SRR616268_1.fastq.bz2)中の最初の100万リード分を抽出したものであるため、①Total Sequencesが1000000になってしかるべき。②配列長には特に言及していなかったが、107塩基であることがわかる。またGC含量が50%であることも分かる。

FastQC Report

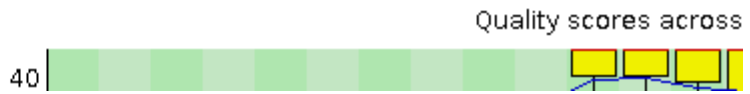
Summary

- ✓ Basic Statistics
- ✓ Per base sequence quality
- ✓ Per sequence quality scores
- ✗ Per base sequence content
- ✗ Per base GC content
- ! Per sequence GC content
- ✓ Per base N content
- ✓ Sequence Length Distribution
- ✗ Sequence Duplication Levels
- ✗ Overrepresented sequences
- ✗ Kmer Content

Basic Statistics

Measure	Value
Filename	SRR616268sub_1.fastq
File type	Conventional base calls
Encoding	Illumina 1.5
Total Sequences	1000000
Filtered Sequences	0
Sequence length	107
%GC	50

Per base sequence quality



W8-3: FastQC結果を眺める

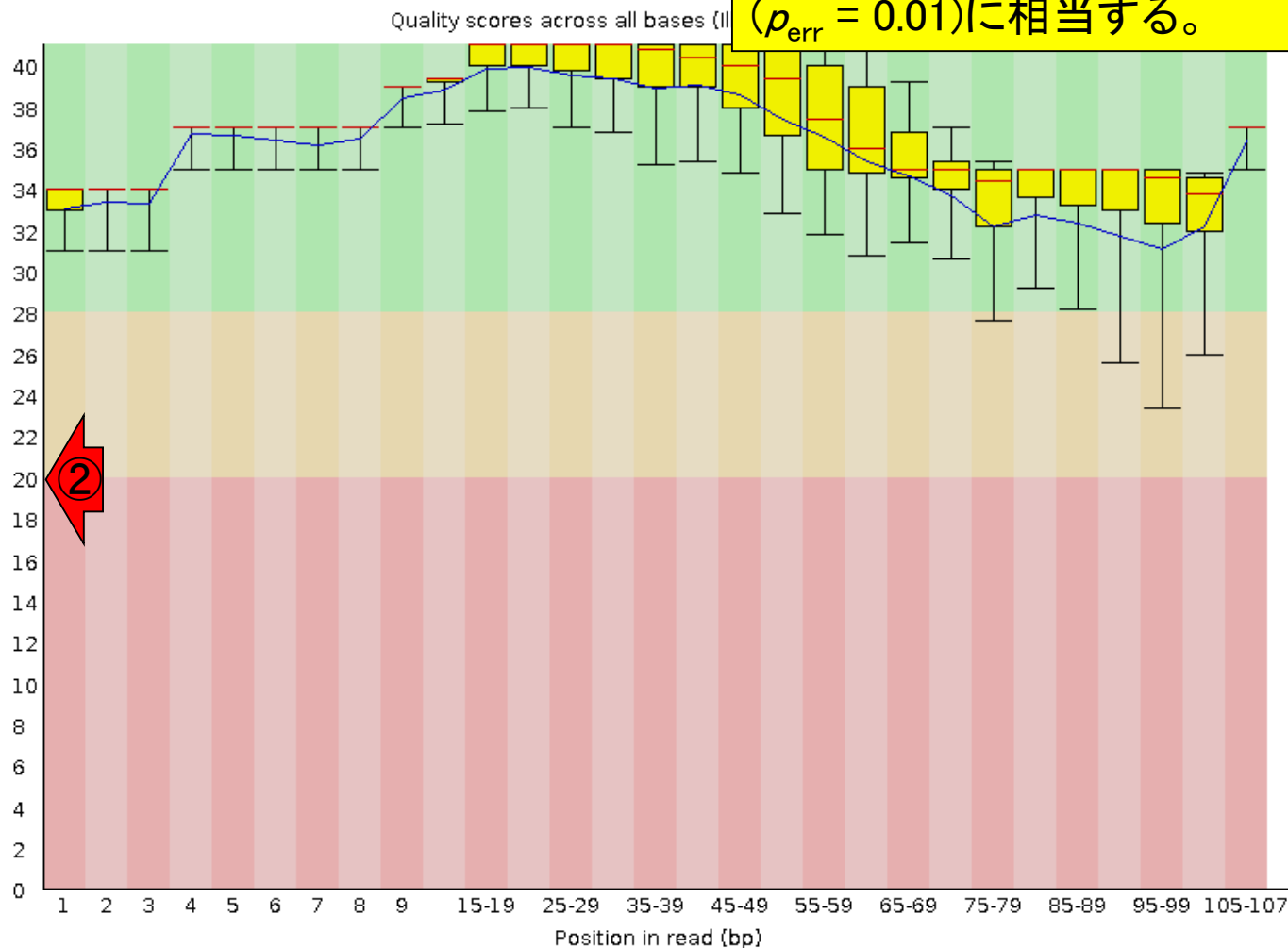
①塩基(ポジション)ごとのクオリティ調査。横軸は塩基のポジション、縦軸はPhredスコア(高いほど塩基配列決定精度が高い)。②スコア20が1つの目安。これは100回中1回のベースコールエラー率($p_{err} = 0.01$)に相当する。

FastQC Report

Summary

- ✓ Basic Statistics
- ✓ Per base sequence quality **①**
- ✓ Per sequence quality scores
- ✗ Per base sequence content
- ✗ Per base GC content
- ! Per sequence GC content
- ✓ Per base N content
- ✓ Sequence Length Distribution
- ✗ Sequence Duplication Levels
- ✗ Overrepresented sequences
- ✗ Kmer Content

✓ Per base sequence quality



W8-3: FastQC結果を眺める

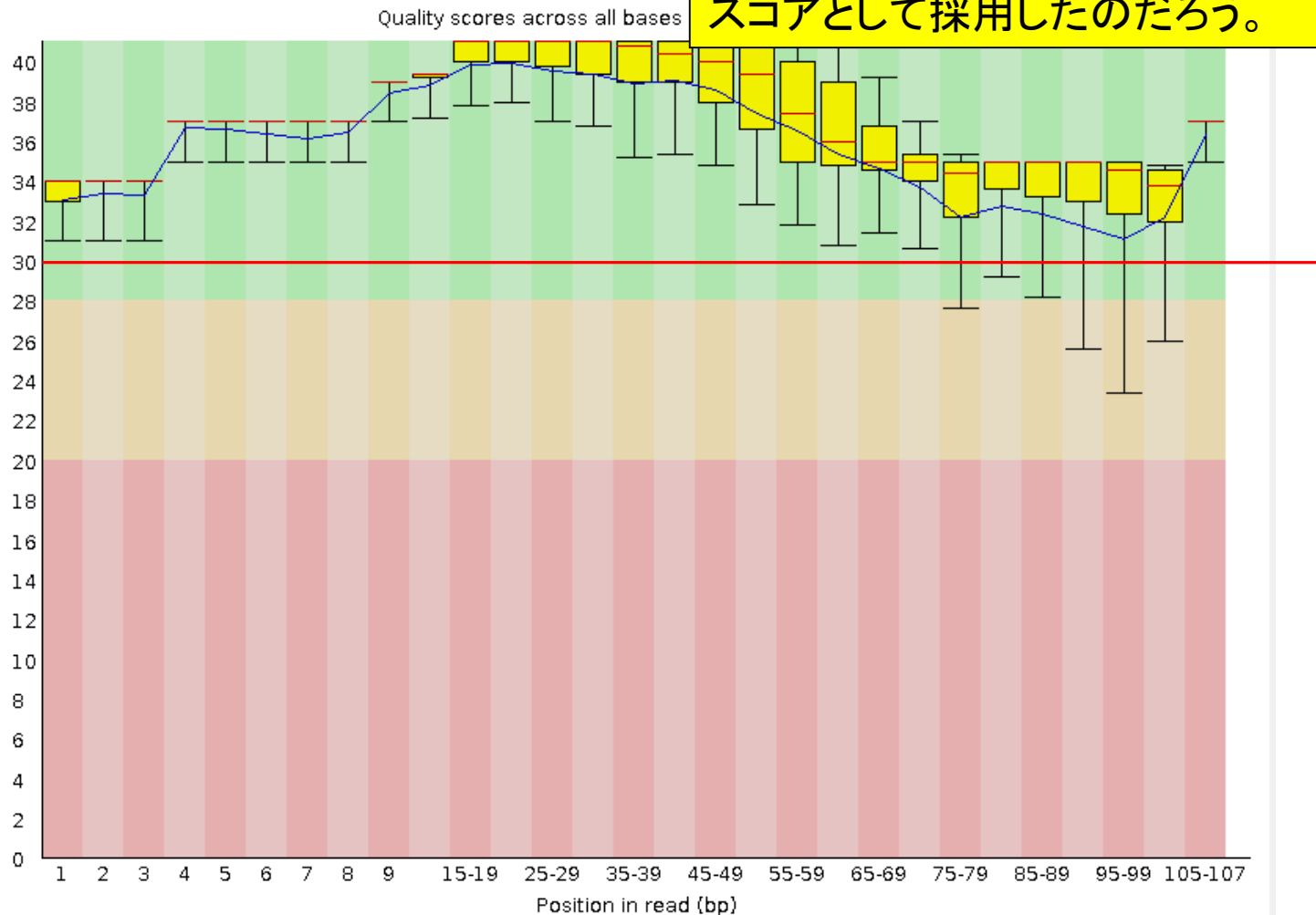
スコア30の赤の水平線は $p_{err} = 0.001$ に相当。 $-\log(p_{err}) = 3$ 。これを10倍した値で表現すると、概ね0-40程度の値の範囲になる。ヒトが認識しやすく、当時のPCの解像度的な観点から、「 $-\log(p_{err}) \times 10$ 」をスコアとして採用したのだろう。

FastQC Report

Summary

- ✓ Basic Statistics
- ✓ Per base sequence quality
- ✓ Per sequence quality scores
- ✗ Per base sequence content
- ✗ Per base GC content
- ! Per sequence GC content
- ✓ Per base N content
- ✓ Sequence Length Distribution
- ✗ Sequence Duplication Levels
- ✗ Overrepresented sequences
- ✗ Kmer Content

✓ Per base sequence quality



W8-3: FastQC結果を眺

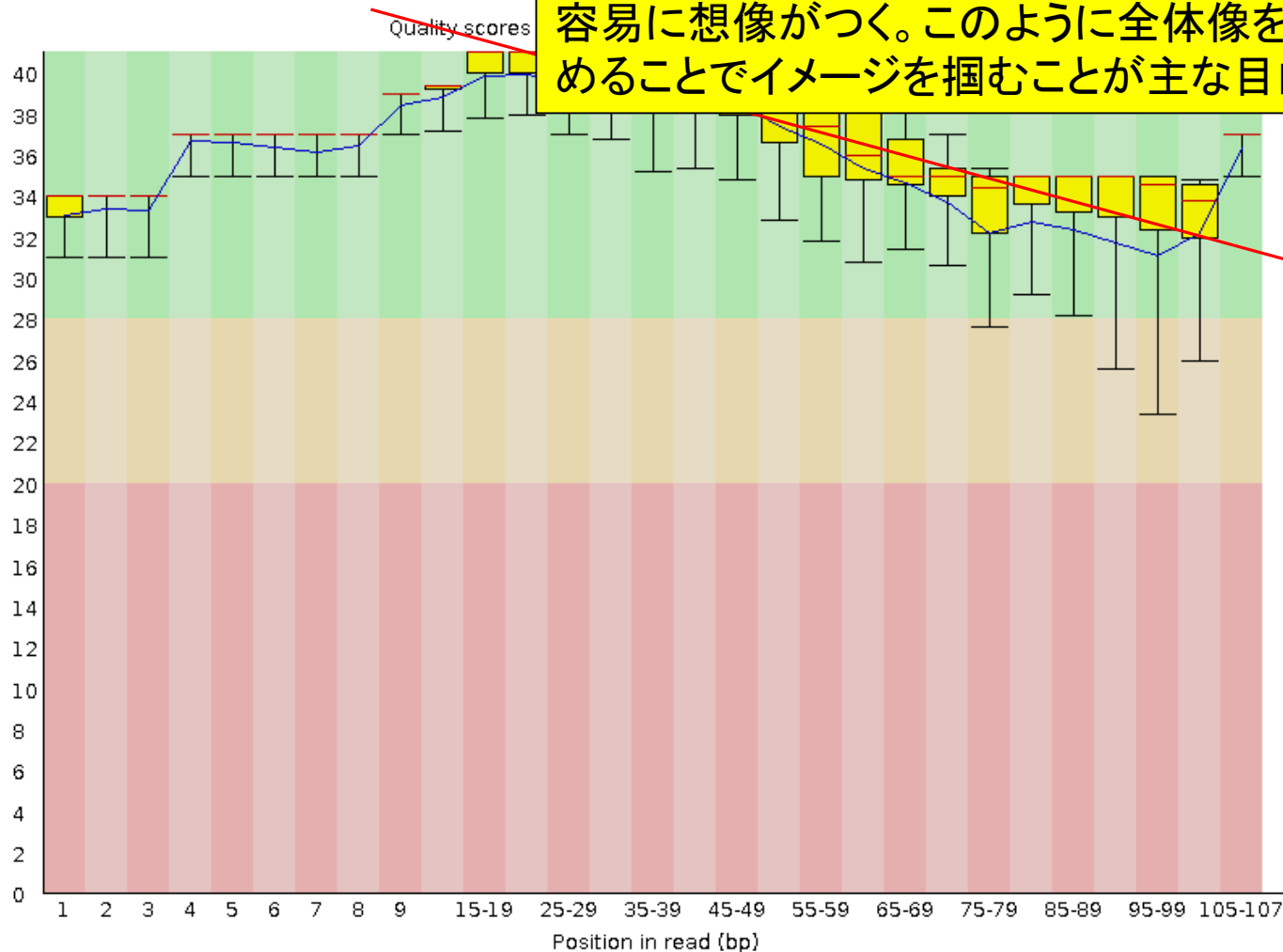
右側ほど塩基配列決定精度が低下傾向にあることがわかる。これが一般的な傾向。このデータは全体的に精度が高いと言える。また、フィルタリング条件として、「スコア20以下の塩基を一定数含むリードの除去」を行っても、ほとんど除去されないということも容易に想像がつく。このように全体像を眺めることでイメージを掴むことが主な目的。

FastQC Report

Summary

- ✓ Basic Statistics
- ✓ Per base sequence quality
- ✓ Per sequence quality scores
- ✗ Per base sequence content
- ✗ Per base GC content
- ! Per sequence GC content
- ✓ Per base N content
- ✓ Sequence Length Distribution
- ✗ Sequence Duplication Levels
- ✗ Overrepresented sequences
- ✗ Kmer Content

✓ Per base sequence quality



W8-4: FastQC結果を眺め

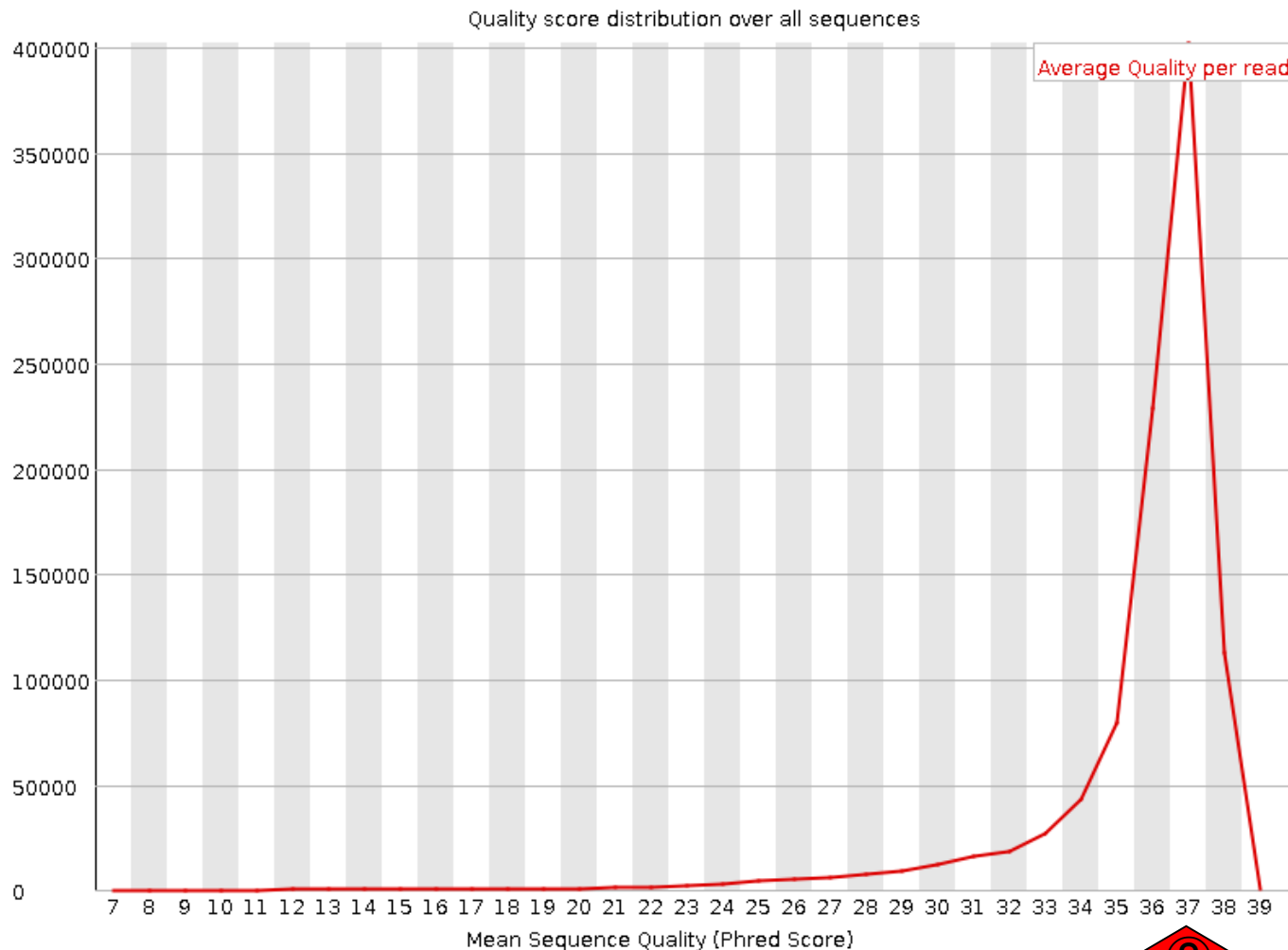
①配列ごとのクオリティスコア。横軸が平均のクオリティスコア。②平均スコアは37。縦軸は配列数。全部で100万リードなので、縦軸が数十万なのは妥当。

FastQC Report

Summary

- ✓ [Basic Statistics](#)
- ✓ [Per base sequence quality](#)
- ✓ [Per sequence quality scores](#)
- ✗ [Per base sequence content](#)
- ✗ [Per base GC content](#)
- ! [Per sequence GC content](#)
- ✓ [Per base N content](#)
- ✓ [Sequence Length Distribution](#)
- ✗ [Sequence Duplication Levels](#)
- ✗ [Overrepresented sequences](#)
- ✗ [Kmer Content](#)

✓ Per sequence quality scores



W8-5: FastQC結果

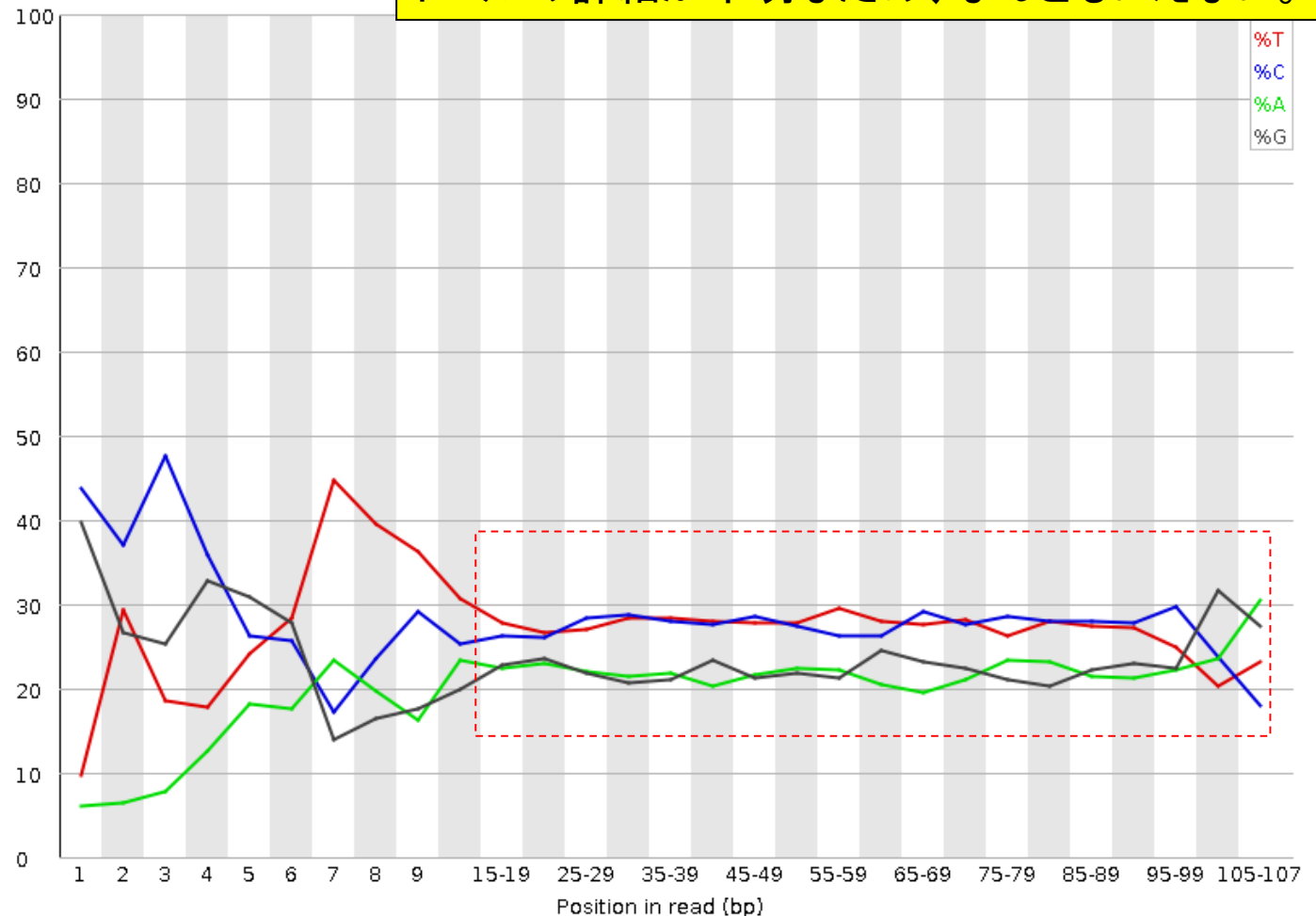
①ポジションごとの塩基の出現確率。赤枠のような塩基ごとのプロファイルがフラットになれば基本的にはOK。1-15番目あたりのプロファイルは、実験プロトコル (oligo dT primer or random hexamer; DNase Iを利用した破碎) に由来するのかもしれない。このデータの場合は、原著論文がなく実験プロトコルの詳細が不明なため、なんともいえない。

FastQC Report

Summary

- ✓ Basic Statistics
- ✓ Per base sequence quality
- ✓ Per sequence quality scores
- ✗ Per base sequence content
- ✗ Per base GC content
- ! Per sequence GC content
- ✓ Per base N content
- ✓ Sequence Length Distribution
- ✗ Sequence Duplication Levels
- ✗ Overrepresented sequences
- ✗ Kmer Content

✗ Per base sequence content















W8-6: FastQC結果を眺め

①頻出する配列をリストアップ。②トップは「CCCCGGTATA…」という50塩基の配列で14,383回出現。Percentageは1.4383%。全部で100万リードなので妥当。後にオリジナル107 bpのうち最初の50 bpで解析していることが判明。

FastQC Report

Summary

-  [Basic Statistics](#)
-  [Per base sequence quality](#)
-  [Per sequence quality scores](#)
-  [Per base sequence content](#)
-  [Per base GC content](#)
-  [Per sequence GC content](#)
-  [Per base N content](#)
-  [Sequence Length Distribution](#)
-  [Sequence Duplication Levels](#)
-  [Overrepresented sequences](#) 
-  [Kmer Content](#)

Overrepresented sequences












Sequence	Count	Percentage	Possible Source
CCCCGGTATATTTTCGGCGCAGTGCCACTCGACTAGTGAGCTATTACGCA	14383	1.4383	No Hit ②
GGCCTATTCACTGCGGCTGACCTTTCGGTACGACCCCTTCTTCCGAAGT	11044	1.1044	No Hit
GTGCTTTTCACCTTTCCCTCACGGTACTGGTTCATCGGTACTAGGG	8892	0.8892000000000001	No Hit
CCCGGTATATTTTCGGCGCAGTGCCACTCGACTAGTGAGCTATTACGCAC	8474	0.8474	No Hit
GTCCTAGGGAGTATTTAGCCTTGGGAGATGGTCTCCCGGATTCGACG	8189	0.8189	No Hit
GCCGGCATTCTCACTTCTAAGCGCTCCAGCCGCTCCTCAGATCGACCTTC	8132	0.8132	No Hit
GTCCAGTCTACAACCCCGAGAAGCAAGCTTCTCGGTTTGGGCTCTTCCC	6663	0.6663	No Hit
GTCGGTTTTCGGTACGGGTAGTTTATTTCTCACTAGAAGCTTTTCTTGGC	6411	0.6411	No Hit
GGTCACTTGGTTTCGGGTCTACATCTGCTTACTCATTGCCCCTGTTTCCAG	5502	0.5502	No Hit
GCCGGCATTCTCACTTCTAAGCGCTCCAGCCGCTCCTCAGATCAACCTTC	4845	0.48450000000000004	No Hit
CCCTCCATCGCTTAAACAATAAACTAGTGCAGGAATCTCAACCTGCTT	4395	0.43949999999999995	No Hit
CCGGTATATTTTCGGCGCAGTGCCACTCGACTAGTGAGCTATTACGCACT	4385	0.4385	No Hit
CCCGGTCTGCCCGCGCCAGCTATGTATTCACTGACAAGCAATACACTG	4366	0.4366	No Hit
CCACAGTTTCGGTATTATGCTTAGCCCCGGTATATTTTCGGCGCAGTGCC	4314	0.4314	No Hit
CTGGGCTGTTCCCTTTTCGACAATGGACCTTATCGCTCACTGTCTGACTC	4113	0.41130000000000005	No Hit
CCGCCGACTCAGGATCCTGGACGGAGGGTTCGACGTTTCGCTTACAGGG	4081	0.4081	No Hit
CCGGCATTCTCACTTCTAAGCGCTCCAGCCGCTCCTCAGATCGACCTTCA	3846	0.3846	No Hit

W8-6: FastQC結果を眺め

①頻出する配列をリストアップ。②ときどき既知のアダプター配列とマッチするものが見つかる。2,415回出現した「GATCGGAA…」という50塩基の配列は、TruSeq Adapter Index 3というものと100%一致していたことを示す。このようなリード中の部分配列はトリムされるべき。

FastQC Report

Summary

-  [Basic Statistics](#)
-  [Per base sequence quality](#)
-  [Per sequence quality scores](#)
-  [Per base sequence content](#)
-  [Per base GC content](#)
-  [Per sequence GC content](#)
-  [Per base N content](#)
-  [Sequence Length Distribution](#)
-  [Sequence Duplication Levels](#)
-  [Overrepresented sequences](#)
-  [Kmer Content](#)



GCCACATCCTTTTCCACTTAGCATAAATTTAGGGACCTTAACTG	2430	0.243	No Hit
CCTCCATCGCTTAAACAAAATAAACTAGTGCAGGAATCTCAACC	2429	0.2429	No Hit
CTTGGGAGATGGTCCTCCCGATTCCGACGGAATTTACGTGTT	2421	0.2421	No Hit
GATCGGAAGAGCACACGTCTGAACTCCAGTCACTTAGGCATCTCGTATGC	2415	0.2415	TruSeq Adapter, Index 3 (100% over 50bp)
CCGGCATTCTCACTTCTAAGCGCTCCAGCCGCTCACGATCAACCTTCA	2393	0.2393	No Hit
GGGAGATTTAGCCTTGGGAGATGGTCCTCCCGATTCCGACGGAATTC	2365	0.2365	No Hit
TGGGCCTATTCAGTGGGCTGACCTTGGGTCAGCACCCCTTCTTCCGAA	2332	0.2332	No Hit
CTCCTGCGTCCCTCCATCGCTTAAACAAAATAAACTAGTGCAGGAATCTC	2331	0.233100000000000003	No Hit
CGCCGCTACTATGGGAATCGAGTTTTCTTCTTCTTCTGCGGGTACTGAG	2255	0.2255	No Hit
CGCCGGTTTCACTTACAAAAGGCAGCCATTACCGTTAACGGCTTTGA	2172	0.217199999999999998	No Hit
CCGCCGGCCAGCTATGATTCACTGACAAGCAATACACTGATGTGACTG	2081	0.2081	No Hit
CCCCTTCGACAATGGACCTTATCGCTCACTGTCTGACTCCCGGAGTAAG	2070	0.207	No Hit
CTTAACCTGCTGATCTGGCTGTTCCCTTTCCGACAATGGACCTTATCGCT	2005	0.200499999999999998	No Hit

②












他にもちらほらと既知のアダプター配列と一致するものが見つかる。

W8-6: FastQC結果を眺める

FastQC Report

Wed 6 May 2015
SRR616268sub_1.fastq

Summary

-  [Basic Statistics](#)
-  [Per base sequence quality](#)
-  [Per sequence quality scores](#)
-  [Per base sequence content](#)
-  [Per base GC content](#)
-  [Per sequence GC content](#)
-  [Per base N content](#)
-  [Sequence Length Distribution](#)
-  [Sequence Duplication Levels](#)
-  [Overrepresented sequences](#)
-  [Kmer Content](#)



GTCCTCTCTGCACTCAAGTTCCAGTTCCGATGGCTTCTCGGTTA	1559	0.1559	No Hit
GATCGGAAGAGCACACGTCTGAACTCCAGTCACCGATGTATCTCGTATGC	1539	0.1539	TruSeq Adapter, Index 2 (100% over 50bp)
CTGCCTTAGATCCCGACTAACCCCTGGGAGGACGAGCCTTCCCCAGGAAA	1499	0.14989999999999998	No Hit
GGGACCTTAACTGGTGATCTGGGCTGTTCCCTTTTCGACAATGGACCTTA	1497	0.1497	No Hit
ATTTAGCCTTGGGAGATGGTCTCCCGATTCCGACGGAATTTACAGTGT	1487	0.1487	No Hit
CATGGGTAGGTCACCTTGGTTTCGGGTCTACATCTGCTTACTCATTGCCC	1464	0.1464	No Hit
CCCTACTGCTGCCTCCCGTAGGAGTTTGGGCCGTGTCTCAGTCCCAATGT	1455	0.1455	No Hit
CTTTCGACAATGGACCTTATCGCTCACTGTCTGACTCCCGGAGTAAGATC	1451	0.1451	No Hit
CGCTACTCATGCCGGCATTCTCACTTCTAAGCGCTCCAGCCGCTCCTCACG	1405	0.1405	No Hit
CCGGGGTGCTTTTACCTTTCCCTCACGGTACTGGTTCACTATCGGTCAC	1361	0.1361	No Hit
CTGGTGATCTGGGCTGTTCCCTTTTCGACAATGGACCTTATCGCTCACTG	1338	0.1338	No Hit
GGGCTATTCACTGCGGCTGACCTTGGGTCAGCACCCCTTCTCCGAAG	1337	0.1337	No Hit
GCCGCCGGCCAGCTATGTATTCACTGACAAGCAATACACTGATGTGTA	1324	0.13240000000000002	No Hit
AGATCGGAAGAGCACACGTCTGAACTCCAGTCACCGATGTATCTCGTATG	1318	0.1318	TruSeq Adapter, Index 2 (100% over 49bp)
GCCGAGTTCCTTAACGAGAGTTGCTCGCTCACCTGAGGATACTCTCCTC	1310	0.131	No Hit
CCCTCCCTCCATCCCTTAAAGAAATTAAGTACTCCAGCAATCTCAAGCTC	1270	0.1270	No Hit

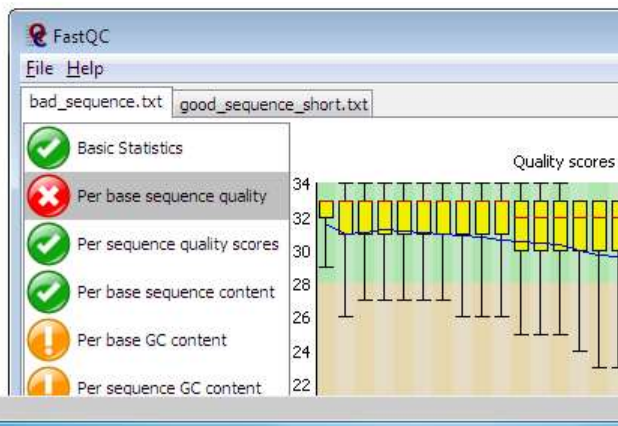
W9-1 : FastQC

FastQCのウェブサイトの場合、Changelogを眺めるとリリースされたバージョンの歴史をたどることができる。2015年5月7日現在のFastQCプログラムの最新版は、ver. 0.11.3。



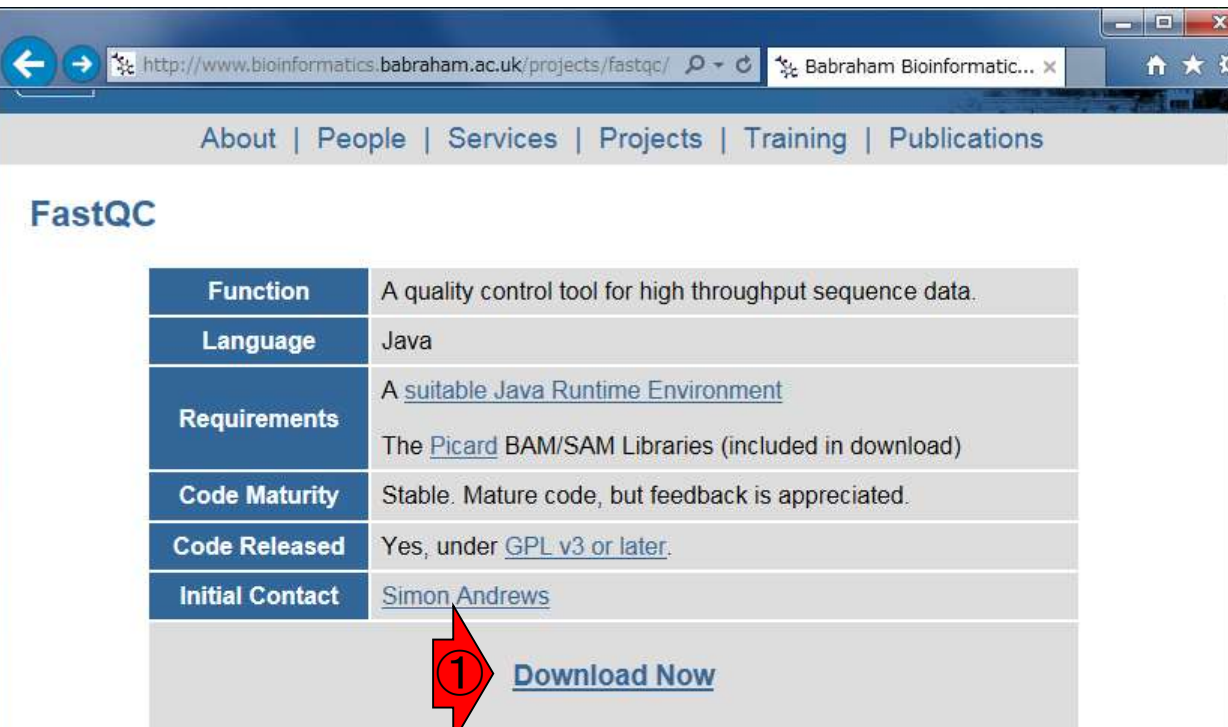
Changelog

- 25-3-15: Version 0.11.3 released
 - Fixed a bug when disabling the per-tile plot from limits.txt
 - Fixed a bug which caused the program to continue when processing of multiple files was actually complete
 - Fixed a bug which meant format selection in the interactive application didn't work
 - Added checks for mis-identifying tile numbers in confusing sample ids
 - Added the SOLID smallRNA adapter to the standard search set
 - Fixed a bug when extracting casava names from uncompressed fastq files
 - Added support for processing files of Oxford Nanopore reads
- 6-6-14: Version 0.11.2 released
 - Fixed incorrect warn/fail defaults for per-seq quality plot
 - Fixed memory leaks in Kmer and per-seq quality modules
 - Added an option to use a custom limits file
 - Fixed a bug in the naming of the folder inside the zip output file
 - Fixed a bug in the --extract option
- 2-6-14: Version 0.11.1 released
 - Added configurable warn/fail thresholds for all modules
 - Allow modules to be selectively turned off



W9-1 : FastQC

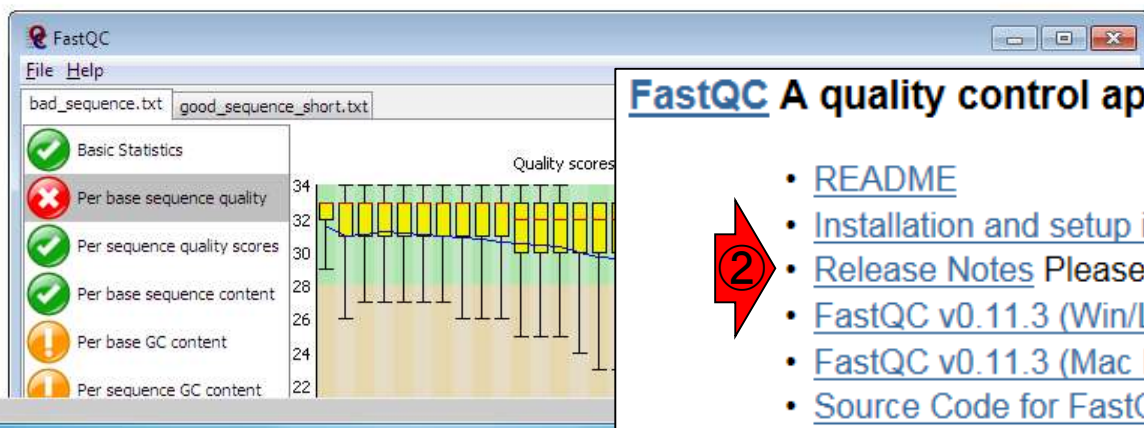
Changelogよりも、②の「Release Notes」が一般的かもしれません。③基本的には、ここを見ながら自力でインストールします。リンク先のマニュアルをざっと眺めても、現時点でもchmodなどいくつかのコマンドや相対パスの意味は分かるようになっているでしょう。



A screenshot of a web browser showing the FastQC project page. The browser address bar shows the URL: <http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>. The page has a navigation menu with links for About, People, Services, Projects, Training, and Publications. Below the menu, the title "FastQC" is displayed. A table provides details about the tool:

Function	A quality control tool for high throughput sequence data.
Language	Java
Requirements	A suitable Java Runtime Environment The Picard BAM/SAM Libraries (included in download)
Code Maturity	Stable. Mature code, but feedback is appreciated.
Code Released	Yes, under GPL v3 or later .
Initial Contact	Simon Andrews

Below the table, there is a red arrow with the number 1 pointing to a blue button labeled "Download Now".



A screenshot of the FastQC application interface. The window title is "FastQC". The menu bar includes "File" and "Help". There are two tabs: "bad_sequence.txt" and "good_sequence_short.txt". On the left, a sidebar shows a list of quality control metrics with status icons: Basic Statistics (green check), Per base sequence quality (red X), Per sequence quality scores (green check), Per base sequence content (green check), Per base GC content (orange exclamation), and Per sequence GC content (orange exclamation). The main area displays a box plot titled "Quality scores" with a y-axis ranging from 22 to 34. The plot shows the distribution of quality scores for each base across the sequence.

FastQC A quality control application for high throughput sequence data

- [README](#)
- [Installation and setup instructions](#)
- [Release Notes](#) Please read these before using the program.
- [FastQC v0.11.3 \(Win/Linux zip file\)](#)
- [FastQC v0.11.3 \(Mac DMG image\)](#)
- [Source Code for FastQC v0.11.3 \(zip file\)](#)

W9-2: Install (FastQC)

FastQC A quality control application for high throughput sequencing data

- [README](#)
- [Installation and setup instructions](#) ①
- [Release Notes](#) Please read these before using the program.
- [FastQC v0.11.3 \(Windows\)](#)
- [FastQC v0.11.3 \(Mac\)](#)
- [Source Code for FastQC](#)

Installing FastQC

FastQC is a java application. In order to run it needs your system to have a suitable Java Runtime Environment (JRE) installed. Before you try to run FastQC you should therefore ensure that you have a suitable JRE. There are a number of different JREs available however the one we have tested is the v1.6 JRE from Oracle. This is available for a number of different platforms from www.java.com (click the download now button at the top).

If you're not sure whether you have java installed then you can test this from a command prompt. To get a command prompt try:

Windows: Select Start > Run, and type 'cmd' (no quotes) in the box which appears, press OK

MaxOSX: Run Applications > Utilities > Terminal

Linux: From your applications menu look for an application called 'Terminal' or 'Konsole'. Either of these will give you a usable shell.

②

At the command prompt type 'java -version' and press enter. You should see something like:

```
java version "1.6.0_17"  
Java(TM) SE Runtime Environment (build 1.6.0_17-b04-248-10M3025)  
Java HotSpot(TM) Client VM (build 14.3-b01-101, mixed mode)
```

③

If you get an error then you don't have java installed. If the version listed on the first line is less than 1.5 then you might have problems running FastQC.

Actually installing FastQC is as simple as unzipping the zip file it comes in into a suitable location. That's it. Once unzipped it's ready to go.

Linux上でのプログラムのインストール作業は、「このプログラムを実行するためにはこれが必要で…」という前もって必要な事柄 (prerequisite) やプログラムの依存関係 (dependency) との格闘。以下は、②ターミナル上で「java -version」と打ち込み、javaが入っていること、および③バージョンが1.5以上であることを確認せよ、と読み取る。

W9-2: Install (FastQC)

作業ディレクトリはどこでもよいので「java -version」と打ち込み、javaが入っており、そのバージョンが1.5以上(ver. 1.7.0_55)であることを確認。

```
iu@bielinux[iu] pwd
/home/iu
iu@bielinux[iu] java -version
java version "1.7.0_55"
OpenJDK Runtime Environment (IcedTea 2.4.7) (7u55-2.4.7-1ubuntu1)
OpenJDK 64-Bit Server VM (build 24.51-b03, mixed mode)
iu@bielinux[iu] java -v
Unrecognized option: -v
Error: Could not create the Java Virtual Machine.
Error: A fatal exception has occurred. Program will exit.
iu@bielinux[iu] cd Desktop
iu@bielinux[Desktop] java -version
java version "1.7.0_55"
OpenJDK Runtime Environment (IcedTea 2.4.7) (7u55-2.4.7-1ubuntu1)
OpenJDK 64-Bit Server VM (build 24.51-b03, mixed mode)
iu@bielinux[Desktop]
```

W9-2: Install (FastQC)

FastQC A quality control application for high throughput sequence data

- [README](#)
- [Installation and setup instructions](#)
- [Release Notes](#) Please read these before using the program.
- ① [FastQC v0.11.3 \(Win/Linux zip file\)](#)
- [FastQC v0.11.3 \(Mac DMG image\)](#)
- [Source Code for FastQC v0.11.3 \(C\)](#)

開く(O)
新しいタブで開く(W)
新しいウィンドウで開く(N)
対象をファイルに保存(A)...
対象を印刷(P)

切り取り
コピー(C)
ショートカットのコピー(T) ②

FastQ Screen A screening application

- [README](#)
- [Release Notes](#) Please read these before using the program.
- [FastQ Screen v0.4.4](#)

FastQC (ver. 0.11.3)のインストール手順。①のzipファイル(を共有フォルダ経由でゲストOSに置いてもいいがここでは)のURL情報を②で取得。③とりあえずDownloadsディレクトリで作業を行う。「wget -cq」と打ち込んだ後に、④でコピーしたURL情報をペースト。⑤ls -lhで確認。

```
File Edit View Search Terminal Help
iu@bielinux[Downloads] pwd ③
/home/iu/Downloads
iu@bielinux[Downloads] wget -cq http://www.bioinformatics.babraham.ac.uk/projects/fastqc/fastqc_v0.11.3.zip ④
iu@bielinux[Downloads] ls -lh ⑤
total 9.6M
-rw-rw-r-- 1 iu iu 9.6M 3月 25 18:40 fastqc_v0.11.3.zip
iu@bielinux[Downloads]
```

W9-2: Install (FastQC)

①(余分な)進捗状況を非表示にして(-qオプションをつけて)zipファイルを解凍するunzipコマンドを実行。②FastQCというディレクトリが作成されていることが分かる。③このディレクトリ中のfastqcが実行コマンド。

```
iu@bielinux[Downloads] pwd
/home/iu/Downloads
iu@bielinux[Downloads] ls -lh
total 9.6M
-rw-rw-r-- 1 iu iu 9.6M  3月 25 18:40 fastqc_v0.11.3.zip
iu@bielinux[Downloads] unzip -q fastqc_v0.11.3.zip [ 2:46午後 ]
iu@bielinux[Downloads] ls -lh [ 2:46午後 ]
total 9.6M
drwxrwxr-x 8 iu iu 4.0K  5月  7 14:46 FastQC
-rw-rw-r-- 1 iu iu 9.6M  3月 25 18:40 fastqc_v0.11.3.zip
iu@bielinux[Downloads] ls FastQC [ 2:46午後 ]
cisd-jhdf5.jar      INSTALL.txt         org                 Templates
Configuration      jbzip2-0.9.jar     README.txt         uk
fastqc              LICENSE_JHDF5.txt  RELEASE_NOTES.txt
fastqc_icon.ico    LICENSE.txt        run_fastqc.bat
Help                net                 sam-1.103.jar
iu@bielinux[Downloads] [ 2:47午後 ]
```



W9-2: Install (FastQC)

解凍後の作業は、chmodで実行権限を付与さえすればよいようだ。これが「You may need to make this file executable」に相当。

FastQC A quality control application for high throughput sequence data

- [README](#)
- [Installation and setup instructions](#) ①
- [Release Notes](#) Please read these before using the program.
- [FastQC v0.11.3 \(Win/Linux zip file\)](#)
- [FastQC v0.11.3 \(Mac DMG image\)](#)
- [Source Code](#)

Running FastQC Interactively

Windows: Simply double click on the run_fastqc bat file. If you want to make a pretty shortcut then we've included an icon file in the top level directory so you don't have to use the generic bat file icon.

MacOSX: There is an application bundle for MacOSX which you can use to install and run FastQC. Just drag the application from the disk image to your Applications folder (or wherever you want to install the program).

Linux: We have included a wrapper script, called 'fastqc' which is the easiest way to start the program. The wrapper is in the top level of the FastQC installation. You may need to make this file executable:

②

```
chmod 755 fastqc
```

..but once you have done that you can run it directly

③

```
./fastqc
```

..or place a link in /usr/local/bin to be able to run the program from any location:

```
sudo ln -s /path/to/FastQC/fastqc /usr/local/bin/fastqc
```

実行権限を付与して確認しているだけです。

W9-2: Install (FastQC)

```
iu@bielinux[Downloads] ls -lh [ 2:46午後 ]
total 9.6M
drwxrwxr-x 8 iu iu 4.0K  5月  7 14:46 FastQC
-rw-rw-r-- 1 iu iu 9.6M  3月 25 18:40 fastqc_v0.11.3.zip
iu@bielinux[Downloads] ls FastQC [ 2:46午後 ]
cisd-jhdf5.jar      INSTALL.txt      org              Templates
Configuration      jbzip2-0.9.jar  README.txt      uk
fastqc              LICENSE_JHDF5.txt  RELEASE_NOTES.txt
fastqc_icon.ico    LICENSE.txt      run_fastqc.bat
Help                net              sam-1.103.jar
iu@bielinux[Downloads] pwd [ 2:47午後 ]
/home/iu/Downloads
iu@bielinux[Downloads] cd FastQC [ 3:11午後 ]
iu@bielinux[FastQC] pwd [ 3:11午後 ]
/home/iu/Downloads/FastQC
iu@bielinux[FastQC] ls -lh fastqc* [ 3:12午後 ]
-rw-rw-r-- 1 iu iu  14K  3月 24 16:23 fastqc
-rw-rw-r-- 1 iu iu  2.2K  3月 21  2012 fastqc_icon.ico
iu@bielinux[FastQC] chmod 755 fastqc [ 3:12午後 ]
iu@bielinux[FastQC] ls -lh fastqc* [ 3:12午後 ]
-rwxr-xr-x 1 iu iu  14K  3月 24 16:23 fastqc
-rw-rw-r-- 1 iu iu  2.2K  3月 21  2012 fastqc_icon.ico
iu@bielinux[FastQC] █ [ 3:12午後 ]
```


W9-3: パス (のイントロ)

①実行権限を付与した後にfastqcのバージョン情報を表示すると、Bio-Linux 8にプレインストールされているFastQC (ver. 0.10.1)のままとなっている[W7-10]。

```
iu@bielinux[FastQC] pwd [ 3:54午後 ]
/home/iu/Downloads/FastQC
iu@bielinux[FastQC] ls -lh fastqc* [ 3:55午後 ]
-rwxr-xr-x 1 iu iu 14K 3月 24 16:23 fastqc
-rw-rw-r-- 1 iu iu 2.2K 3月 21 2012 fastqc_icon.ico
① iu@bielinux[FastQC] fastqc -v [ 3:55午後 ]
FastQC v0.10.1
iu@bielinux[FastQC] [ 3:55午後 ]
```



W9-3:パス(のイントロ)

②「./fastqc -v」とやると、カレントディレクトリ(.)中にある最新版のfastqcコマンドが実行される。これは一種の相対パス指定でのコマンド実行に相当。③「/home/iu/Downloads/FastQC/fastqc」と絶対パス指定にしてもよいし、④ホームディレクトリ(~)からの相対パス指定でもよい。

```
iu@bielinux[FastQC] pwd
/home/iu/Downloads/FastQC
iu@bielinux[FastQC] ls -lh fastqc*
-rwxr-xr-x 1 iu iu 14K  3月 24 16:23 fastqc
-rw-rw-r-- 1 iu iu 2.2K  3月 21  2012 fastqc_icon.
① iu@bielinux[FastQC] fastqc -v
FastQC v0.10.1
② iu@bielinux[FastQC] ./fastqc -v
FastQC v0.11.3
③ iu@bielinux[FastQC] /home/iu/Downloads/FastQC/fastqc -v
FastQC v0.11.3
④ iu@bielinux[FastQC] ~/Downloads/FastQC/fastqc -v
FastQC v0.11.3
iu@bielinux[FastQC] █
```

W9-3: パス (のイントロ)

但し、通常の作業ディレクトリは「~/Downloads/FastQC」ではない。そのため、②の「./fastqc -v」は非現実的。

```
iu@bielinux[FastQC] pwd [ 3:53午後 ]
/home/iu/Downloads/FastQC
iu@bielinux[FastQC] ls -lh fastqc* [ 3:53午後 ]
-rwxr-xr-x 1 iu iu 14K 3月 24 16:23 fastqc
-rw-rw-r-- 1 iu iu 2.2K 3月 21 2012 fastqc_icon.ico
① iu@bielinux[FastQC] fastqc -v [ 3:53午後 ]
FastQC v0.10.1
② iu@bielinux[FastQC] ./fastqc -v [ 3:54午後 ]
FastQC v0.11.3
③ iu@bielinux[FastQC] /home/iu/Downloads/FastQC/fastqc -v [ 3:54午後 ]
FastQC v0.11.3
④ iu@bielinux[FastQC] ~/Downloads/FastQC/fastqc -v [ 3:54午後 ]
FastQC v0.11.3
iu@bielinux[FastQC] [ 3:54午後 ]
```



- ①
- ②
- ③
- ④

W9-3: パス (のイントロ)

①カレントディレクトリを変更し、②最新版のFastQC (ver. 0.11.3)を用いて③SRR616268sub_2.fastqのQCを実行すると、④赤枠内のものが作成される。⑤FastQC (ver. 0.10.1)の結果と違って、ディレクトリではなくhtmlファイルが作成されていることがわかる。中身はほとんど同じだが、バージョンによって出力形態や出力項目が異なる場合があるので注意

```
iu@bielinux[FastQC] cd ~/Documents/srp017156
iu@bielinux[srp017156] pwd
/home/iu/Documents/srp017156
iu@bielinux[srp017156] ls
JSLAB4_1_Linux.sh      SRR616268sub_1.fastq
JSLAB4_2_Linux.sh      SRR616268sub_1_fastqc
SRR616268_1.fastq.bz2  SRR616268sub_1_fastqc.zip
SRR616268_2.fastq.bz2  SRR616268sub_2.fastq
iu@bielinux[srp017156] ~/Downloads/FastQC/fastqc -v      [ 4:24午後 ]
FastQC v0.11.3
iu@bielinux[srp017156] ~/Downloads/FastQC/fastqc -q SRR616268sub_2.fastq
iu@bielinux[srp017156] ls                                [ 4:39午後 ]
JSLAB4_1_Linux.sh      SRR616268sub_1_fastqc ⑤
JSLAB4_2_Linux.sh      SRR616268sub_1_fastqc.zip
SRR616268_1.fastq.bz2  SRR616268sub_2.fastq
SRR616268_2.fastq.bz2  SRR616268sub_2_fastqc.html ④
SRR616268sub_1.fastq   SRR616268sub_2_fastqc.zip
iu@bielinux[srp017156] [ 4:40午後 ]
```

W9-4: パス

①の結果を眺め、Bio-Linux 8にプレインストールされているfastqcコマンドの実体はどこか？なぜ(lsやpwdのような基本コマンドと同様)どのディレクトリ上でもfastqcコマンドを打つとFastQC (ver. 0.10.1)が実行されるのか？が気になる。これがパスの概念を知るスタート地点。

```
iu@bielinux[srp017156] pwd
/home/iu/Documents/srp017156
iu@bielinux[srp017156] ls
JSLAB4_1_Linux.sh          SRR616268sub_1_fastqc
JSLAB4_2_Linux.sh          SRR616268sub_1_fastqc.zip
SRR616268_1.fastq.bz2     SRR616268sub_2.fastq
SRR616268_2.fastq.bz2     SRR616268sub_2_fastqc.html
SRR616268sub_1.fastq      SRR616268sub_2_fastqc.zip
① iu@bielinux[srp017156] fastqc -v
FastQC v0.10.1
iu@bielinux[srp017156]
```

[6:14午後]

[6:14午後]

W9-4: パス

コマンドを実行するプログラムの場所を知りたい場合には、②「whereis コマンド名」、または③「which コマンド名」と打てばよい。FastQC (ver. 0.10.1)の実体であるfastqcコマンドは/usr/local/binというディレクトリ上に存在することがわかる。他にもwhereというコマンドもある。3つのコマンド(whereis, which, and where)の挙動の違いについてはW9-11を参照。

```
iu@bielinux[srp017156] pwd
/home/iu/Documents/srp017156
iu@bielinux[srp017156] ls
JSLAB4_1_Linux.sh          SRR616268sub_1_fastqc
JSLAB4_2_Linux.sh          SRR616268sub_1_fastqc.zip
SRR616268_1.fastq.bz2     SRR616268sub_2.fastq
SRR616268_2.fastq.bz2     SRR616268sub_2_fastqc.html
SRR616268sub_1.fastq      SRR616268sub_2_fastqc.zip
① iu@bielinux[srp017156] fastqc -v
FastQC v0.10.1
② iu@bielinux[srp017156] whereis fastqc
fastqc: /usr/local/bin/fastqc
③ iu@bielinux[srp017156] which fastqc
/usr/local/bin/fastqc
iu@bielinux[srp017156] █
```

[6:14午後]
[6:14午後]
[6:16午後]
[6:16午後]

W9-4:パス

①の「fastqc -v」は、④の「/usr/local/bin/fastqc -v」の省略形。通常、④のように絶対パス指定などでコマンドを実行しなくて済むように、この場所 (/usr/local/bin)に実行コマンドを置くことが多い。

```
iu@bielinux[srp017156] pwd [ 6:12午後 ]
/home/iu/Documents/srp017156
iu@bielinux[srp017156] ls [ 6:14午後 ]
JSLAB4_1_Linux.sh SRR616268sub_1_fastqc
JSLAB4_2_Linux.sh SRR616268sub_1_fastqc.zip
SRR616268_1.fastq.bz2 SRR616268sub_2.fastq
SRR616268_2.fastq.bz2 SRR616268sub_2_fastqc.html
SRR616268sub_1.fastq SRR616268sub_2_fastqc.zip
① iu@bielinux[srp017156] fastqc -v [ 6:14午後 ]
FastQC v0.10.1
② iu@bielinux[srp017156] whereis fastqc [ 6:14午後 ]
fastqc: /usr/local/bin/fastqc
③ iu@bielinux[srp017156] which fastqc [ 6:16午後 ]
/usr/local/bin/fastqc
④ iu@bielinux[srp017156] /usr/local/bin/fastqc -v [ 6:16午後 ]
FastQC v0.10.1
iu@bielinux[srp017156] [ 6:27午後 ]
```

W9-4: パス

⑤のFastQC (ver. 0.11.3)の実体である /home/iu/Downloads/FastQC/fastqc も、/usr/local/binにコマンドのショートカットを置いておけば、プログラムの実体がどこにあるかを覚えておかなくていいので楽ができる。この作業を「パスを通す」という。

```
iu@bielinux[srp017156] pwd
/home/iu/Documents/srp017156
iu@bielinux[srp017156] ls
JSLAB4_1_Linux.sh          SRR616268sub_1_fastqc
JSLAB4_2_Linux.sh          SRR616268sub_1_fastqc.zip
SRR616268_1.fastq.bz2     SRR616268sub_2.fastq
SRR616268_2.fastq.bz2     SRR616268sub_2_fastqc.html
SRR616268sub_1.fastq      SRR616268sub_2_fastqc.zip
① iu@bielinux[srp017156] fastqc -v          [ 5:44午後 ]
FastQC v0.10.1
② iu@bielinux[srp017156] whereis fastqc     [ 5:44午後 ]
fastqc: /usr/local/bin/fastqc
③ iu@bielinux[srp017156] which fastqc      [ 5:44午後 ]
/usr/local/bin/fastqc
④ iu@bielinux[srp017156] /usr/local/bin/fastqc -v [ 5:44午後 ]
FastQC v0.10.1
⑤ iu@bielinux[srp017156] /home/iu/Downloads/FastQC/fastqc -v
FastQC v0.11.3
iu@bielinux[srp017156] █                  [ 6:12午後 ]
```


W9-4: パス

① /usr/local/binに移動し、② ls。
fastqcコマンドは確かに存在する。
これの詳細情報をls -lhで眺める。

```
File Edit View Search Terminal Help
iu@bielinux[srp017156] cd /usr/local/bin [ 6:57午後]
iu@bielinux[bin] pwd [ 6:57午後]
/usr/local/bin
iu@bielinux[bin] ls [ 6:57午後]
ace_contig_coverage.pl  dotter.LIN      LRmapqtl       Qstats
ace_split.pl           Emap            map_db          rbs_finder
act                    Eqtl            mb-multi       Rcross
archaeopteryx         export2sam      melt.pl        Rmap
art                    fasta           mesquite       roche2gap
artemis               fasta36_t       MImapqtl      roche454ace2gap
big_blast              fastf           mrbayes        roche454ace2gap.sh
big-blast             fastf36_t      msatfinder     Rqtl
big_blast.pl          fastm           MSPcrunch.LIN sam2vcf
big-blast.pl          fastm36_t     mspcrunch     soap2sam
blast2sam             fastqc         MultiRegress  splitstree
blixem                fasts           mview          squint
blixem.LIN           fasts36_t     nd_clip        SRmapqtl
bowtie2sam            fastx           novo2sam       ss-count.pl
BTmapqtl              fastx36_t     nrdb           ssearch
caf2ace               fasty          nrdb.linux     ssearch36_t
caf2fastq             fasty36_t     nrdb.linux-x86 stars
caf2gap               fix_qual5      ocount         stars-setup
caf2phrap             formcon
```

W9-4: パス

① /usr/local/bin中のfastq*の条件を満たすファイル(実質的にfastqcのみ)の詳細情報を表示。「lrwxrwxrwx」となっている。この一番左側のl(エル)は、fastqcコマンド自体がシンボリックリンクであることを意味する。実体のないショートカットのようなもの、という捉え方でよい。そして、実体は赤下線部分に存在する。

```
iu@bielinux[bin] pwd
/usr/local/bin
iu@bielinux[bin] ls -lh fastq*
lrwxrwxrwx 1 root root 30 11月 10 15:02 fastqc -> ../bioinf/fastqc/fastqc/fastqc [ 7:23午後]
iu@bielinux[bin] cd ../bioinf/fastqc/fastqc [ 7:24午後]
iu@bielinux[fastqc] pwd
/usr/local/bioinf/fastqc/fastqc [ 7:24午後]
iu@bielinux[fastqc] ls -lh
total 648K
drwxrwxr-x 2 root root 4.0K 7月 5 2014 Contaminants
-rwxr-xr-x 1 root root 8.3K 3月 21 2012 fastqc
-rw-rw-r-- 1 root root 2.2K 3月 21 2012 fastqc_icon.ico
drwxrwxr-x 5 root root 4.0K 7月 5 2014 Help
-rw-rw-r-- 1 root root 49K 3月 21 2012 jbzip2-0.9.jar
-rw-rw-r-- 1 root root 35K 3月 21 2012 LICENSE.txt
-rw-rw-r-- 1 root root 2.1K 3月 21 2012 README.txt
-rw-rw-r-- 1 root root 26K 5月 3 2012 RELEASE_NOTES.txt
-rw-rw-r-- 1 root root 496K 3月 21 2012 sam-1.32.jar
drwxrwxr-x 3 root root 4.0K 7月 5 2014 Templates
drwxrwxr-x 3 root root 4.0K 7月 5 2014 uk [ 7:24午後]
iu@bielinux[fastqc]
```

W9-4:パス

赤下線で書かれているままに、②「../bioinf/fastqc/fastqc」ディレクトリに移動。これは「/usr/local/bin」から1つ上のディレクトリ(..)にあるbioinf/fastqc/fastqcディレクトリに移動するという意味。相対パスの概念 [W4-6]を理解できていれば、③のpwdの結果に納得できるはず。④確かにFastQC (ver. 0.10.1)のfastqcコマンドの実体がある。

```
iu@bielinux[bin] pwd
/usr/local/bin
① iu@bielinux[bin] ls -lh fastqc*
lrwxrwxrwx 1 root root 30 11月 10 15:02 fastqc -> ../bioinf/fastqc/fastqc/fastqc
② iu@bielinux[bin] cd ../bioinf/fastqc/fastqc
iu@bielinux[fastqc] pwd ← ③
/usr/local/bioinf/fastqc/fastqc
iu@bielinux[fastqc] ls -lh
total 648K
drwxrwxr-x 2 root root 4.0K 7月 5 2014 Contaminants
-rwxr-xr-x 1 root root 8.3K 3月 21 2012 fastqc ④
-rw-rw-r-- 1 root root 2.2K 3月 21 2012 fastqc_icon.ico
drwxrwxr-x 5 root root 4.0K 7月 5 2014 Help
-rw-rw-r-- 1 root root 49K 3月 21 2012 jbzip2-0.9.jar
-rw-rw-r-- 1 root root 35K 3月 21 2012 LICENSE.txt
-rw-rw-r-- 1 root root 2.1K 3月 21 2012 README.txt
-rw-rw-r-- 1 root root 26K 5月 3 2012 RELEASE_NOTES.txt
-rw-rw-r-- 1 root root 496K 3月 21 2012 sam-1.32.jar
drwxrwxr-x 3 root root 4.0K 7月 5 2014 Templates
drwxrwxr-x 3 root root 4.0K 7月 5 2014 uk
iu@bielinux[fastqc]
```

[7:23午後]
[7:24午後]
[7:24午後]
[7:24午後]

W9-5: パスを通す

FastQC A quality control application for high throughput sequencing data

- [README](#)
- [Installation and setup instructions](#) ①
- [Release Notes](#) Please read these before using the program.
- [FastQC v0.11.3 \(Win/Linux zip file\)](#)
- [FastQC v0.11.3 \(Mac DMG image\)](#)
- [Source Code](#)

FastQC (ver. 0.11.3)の実行コマンド
/home/iu/Downloads/FastQC/fastqcをこのディレクトリ上からも利用できるようにするためには、/usr/local/binにシンボリックリンクを張っておけばいいことがわかっている。その状態で赤下線部分の説明書きを眺めると、何を言わんとしているのかわかるだろう。当然、②のコマンドをそのまま打ってもダメ！

Running FastQC Interactively

Windows: Simply double click on the run_fastqc bat file. If you want to make a pretty shortcut then we've included an icon file in the top level directory so you don't have to use the generic bat file icon.

MacOSX: There is an application bundle for MacOSX which you can use to install and run FastQC. Just drag the application from the disk image to your Applications folder (or wherever you want to install the program).

Linux: We have included a wrapper script, called 'fastqc' which is the easiest way to start the program. The wrapper is in the top level of the FastQC installation. You may need to make this file executable:

```
chmod 755 fastqc
```

..but once you have done that you can run it directly

```
./fastqc
```

..or place a link in /usr/local/bin to be able to run the program from any location:

```
sudo ln -s /path/to/FastQC/fastqc /usr/local/bin/fastqc
```



W9-5: パスを通す

②は、実行コマンド/path/to/FastQC/fastqcをどこのディレクトリ上からも利用できるようにするために、/usr/local/binにfastqcという名前で、管理者権限で(sudo)シンボリックリンクを張る(ln -s)ということ。

```
./fastqc
```

```
..or place a link in /usr/local/bin to be able to run the progr
```

```
sudo ln -s /path/to/FastQC/fastqc /usr/local/bin/fastqc
```

<http://www.bioinformatics.babraham.ac.uk/projects/fastqc/INSTALL.txt>

W9-5: パスを通す

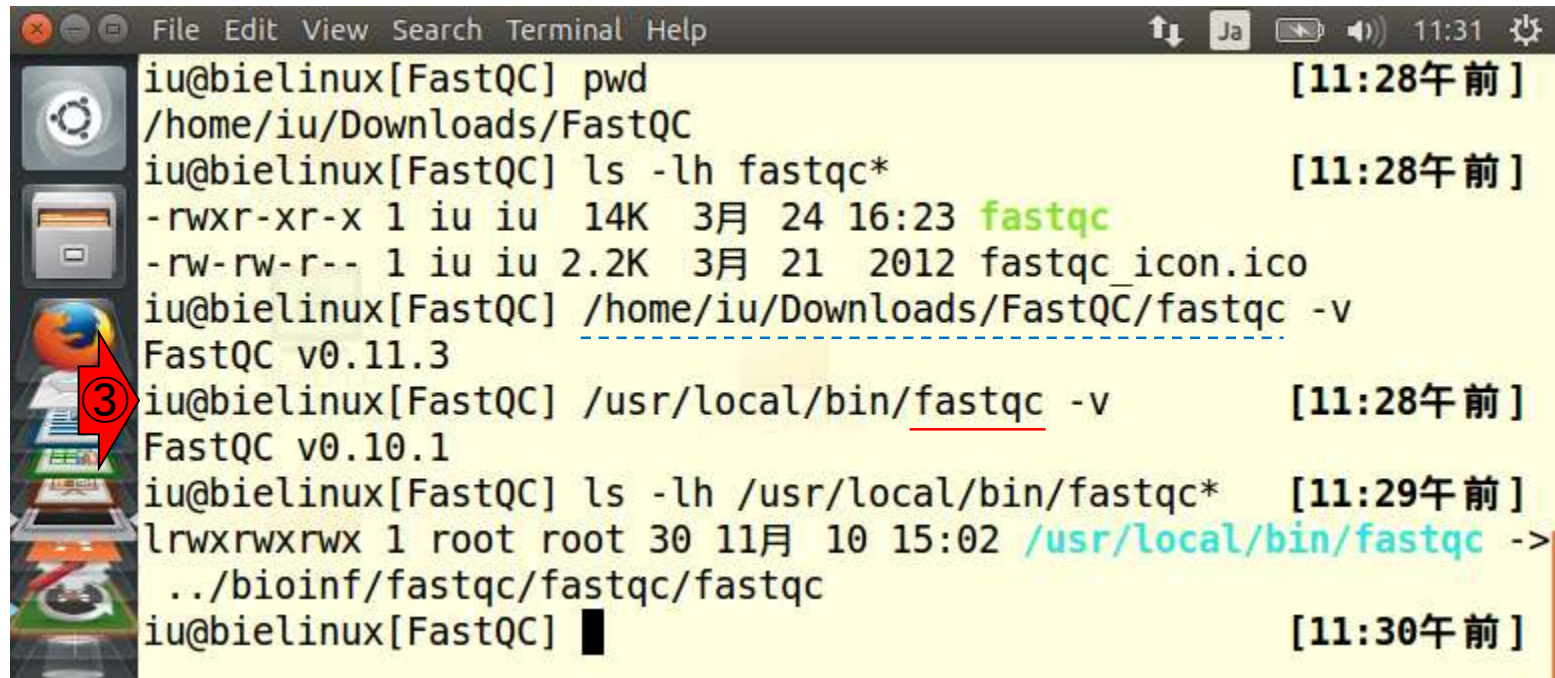
ただし、著者らのゲストOS (Bio-Linux 8)環境では、③ /usr/local/bin/fastqcは、FastQC (ver. 0.10.1)の実行コマンドとして既に存在する。これはただのリマインド。

```
./fastqc
```

```
..or place a link in /usr/local/bin to be able to run the progr
```

```
sudo ln -s /path/to/FastQC/fastqc /usr/local/bin/fastqc
```

<http://www.bioinformatics.babraham.ac.uk/projects/fastqc/INSTALL.txt>



```
iu@bielinux[FastQC] pwd [11:28午前]
/home/iu/Downloads/FastQC
iu@bielinux[FastQC] ls -lh fastqc* [11:28午前]
-rwxr-xr-x 1 iu iu 14K 3月 24 16:23 fastqc
-rw-rw-r-- 1 iu iu 2.2K 3月 21 2012 fastqc_icon.ico
iu@bielinux[FastQC] /home/iu/Downloads/FastQC/fastqc -v
FastQC v0.11.3
iu@bielinux[FastQC] /usr/local/bin/fastqc -v [11:28午前]
FastQC v0.10.1
iu@bielinux[FastQC] ls -lh /usr/local/bin/fastqc* [11:29午前]
lrwxrwxrwx 1 root root 30 11月 10 15:02 /usr/local/bin/fastqc ->
../bioinf/fastqc/fastqc/fastqc
iu@bielinux[FastQC] [11:30午前]
```

それゆえ、(上書きしちゃってもいいが)ここでは
/home/iu/Downloads/FastQC/fastqcをfastqc2という名
前で/usr/local/binにシンボリックリンクを張る。

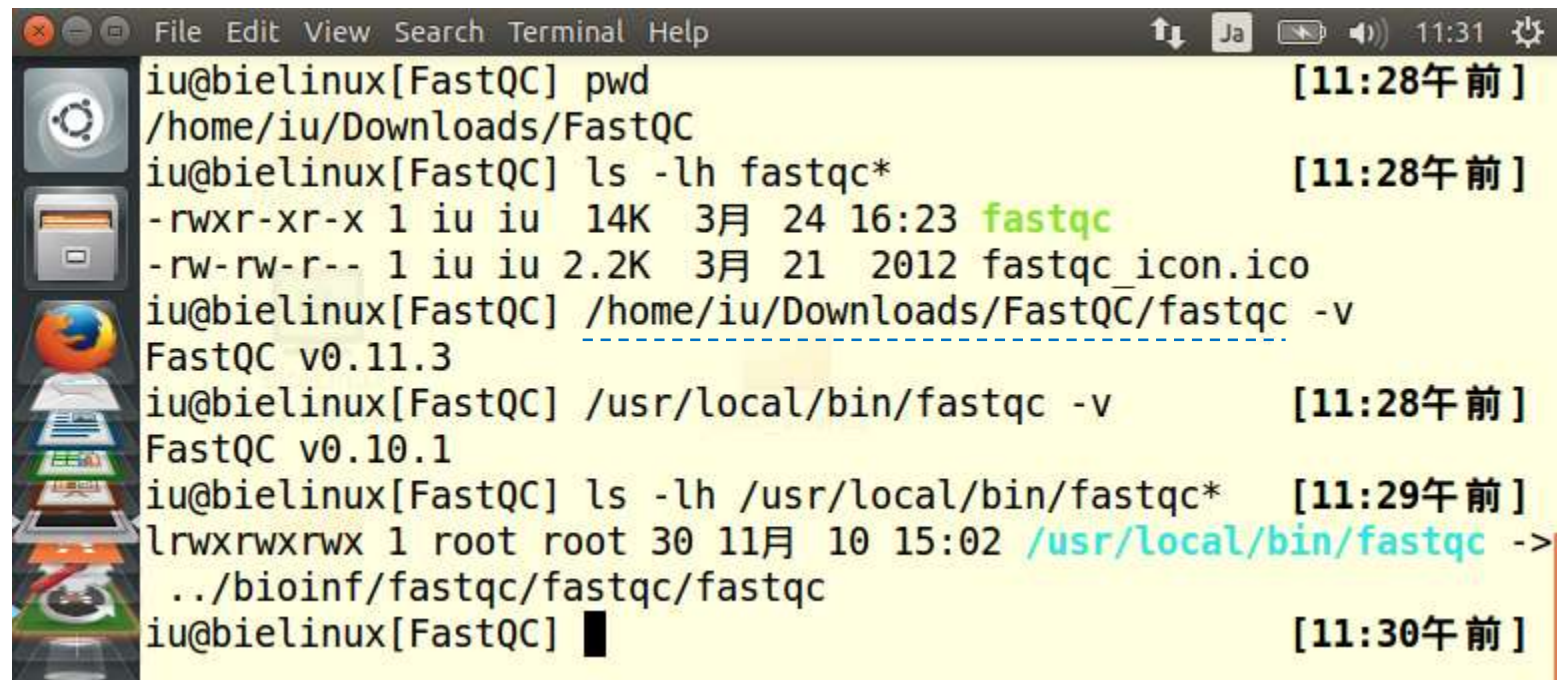
W9-5: パスを通す

```
./fastqc
```

```
..or place a link in /usr/local/bin to be able to run the progr
```

```
sudo ln -s /path/to/FastQC/fastqc /usr/local/bin/fastqc
```

<http://www.bioinformatics.babraham.ac.uk/projects/fastqc/INSTALL.txt>



```
iu@bielinux[FastQC] pwd [11:28午前]
/home/iu/Downloads/FastQC
iu@bielinux[FastQC] ls -lh fastqc* [11:28午前]
-rwxr-xr-x 1 iu iu 14K 3月 24 16:23 fastqc
-rw-rw-r-- 1 iu iu 2.2K 3月 21 2012 fastqc_icon.ico
iu@bielinux[FastQC] /home/iu/Downloads/FastQC/fastqc -v
FastQC v0.11.3
iu@bielinux[FastQC] /usr/local/bin/fastqc -v [11:28午前]
FastQC v0.10.1
iu@bielinux[FastQC] ls -lh /usr/local/bin/fastqc* [11:29午前]
lrwxrwxrwx 1 root root 30 11月 10 15:02 /usr/local/bin/fastqc ->
../bioinf/fastqc/fastqc/fastqc
iu@bielinux[FastQC] [11:30午前]
```

W9-5: パスを通す

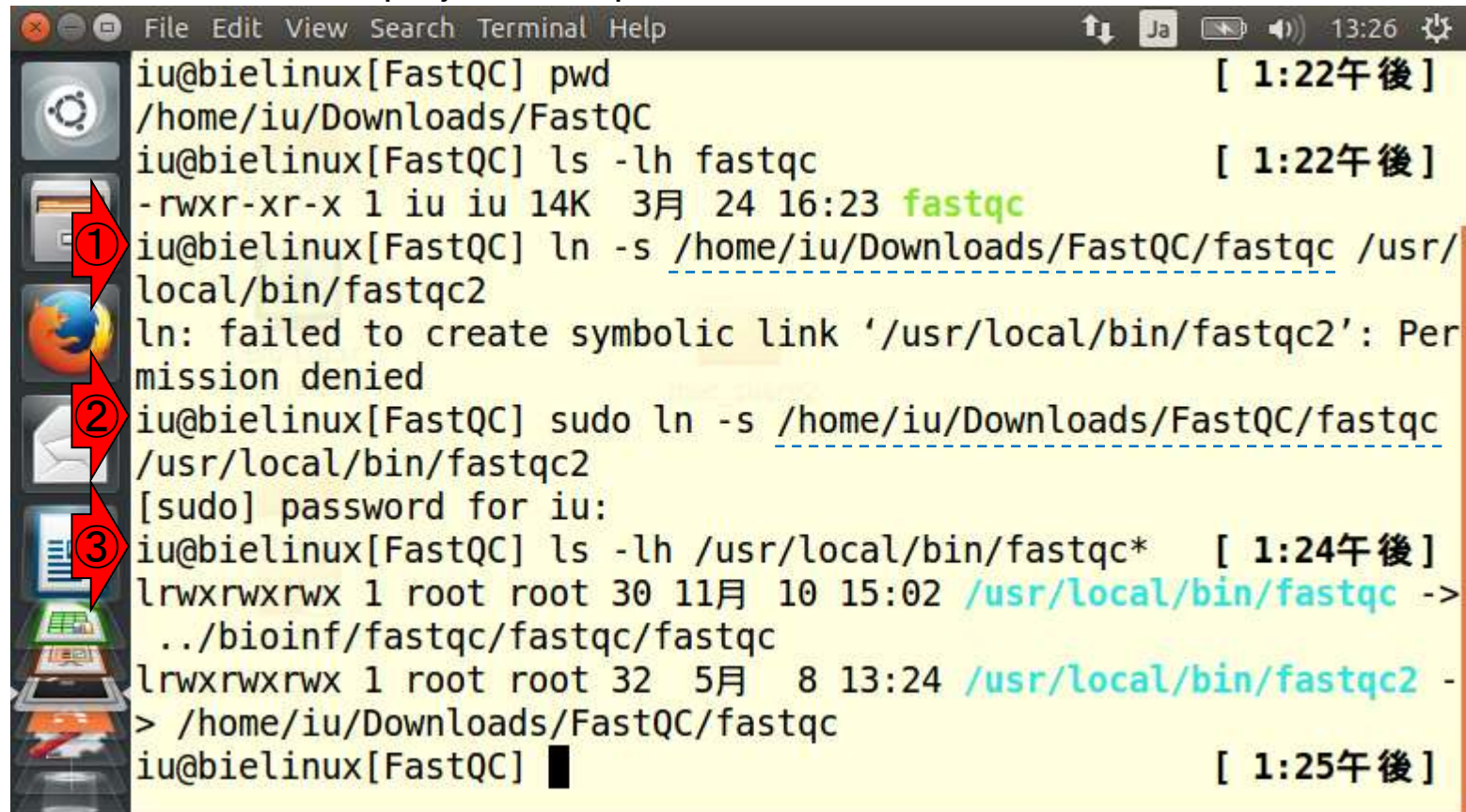
①は管理者権限(sudo)で実行しないと失敗するという例。
②sudoをつけてリトライ。パスワードを要求されたら、「ログインパスワード」を打ち込む。
③/usr/local/bin中のfastqc*の条件を満たすコマンド群を表示。

```
./fastqc
```

```
..or place a link in /usr/local/bin to be able to run the progr
```

```
sudo ln -s /path/to/FastQC/fastqc /usr/local/bin/fastqc
```

<http://www.bioinformatics.babraham.ac.uk/projects/fastqc/INSTALL.txt>



```
iu@bielinux[FastQC] pwd [ 1:22午後 ]
/home/iu/Downloads/FastQC
iu@bielinux[FastQC] ls -lh fastqc [ 1:22午後 ]
-rwxr-xr-x 1 iu iu 14K 3月 24 16:23 fastqc
① iu@bielinux[FastQC] ln -s /home/iu/Downloads/FastQC/fastqc /usr/
local/bin/fastqc2
ln: failed to create symbolic link '/usr/local/bin/fastqc2': Per
mission denied
② iu@bielinux[FastQC] sudo ln -s /home/iu/Downloads/FastQC/fastqc
/usr/local/bin/fastqc2
[sudo] password for iu:
③ iu@bielinux[FastQC] ls -lh /usr/local/bin/fastqc* [ 1:24午後 ]
lrwxrwxrwx 1 root root 30 11月 10 15:02 /usr/local/bin/fastqc ->
../bioinf/fastqc/fastqc/fastqc
lrwxrwxrwx 1 root root 32 5月 8 13:24 /usr/local/bin/fastqc2 -
> /home/iu/Downloads/FastQC/fastqc
iu@bielinux[FastQC] [ 1:25午後 ]
```


W9-6: 確認

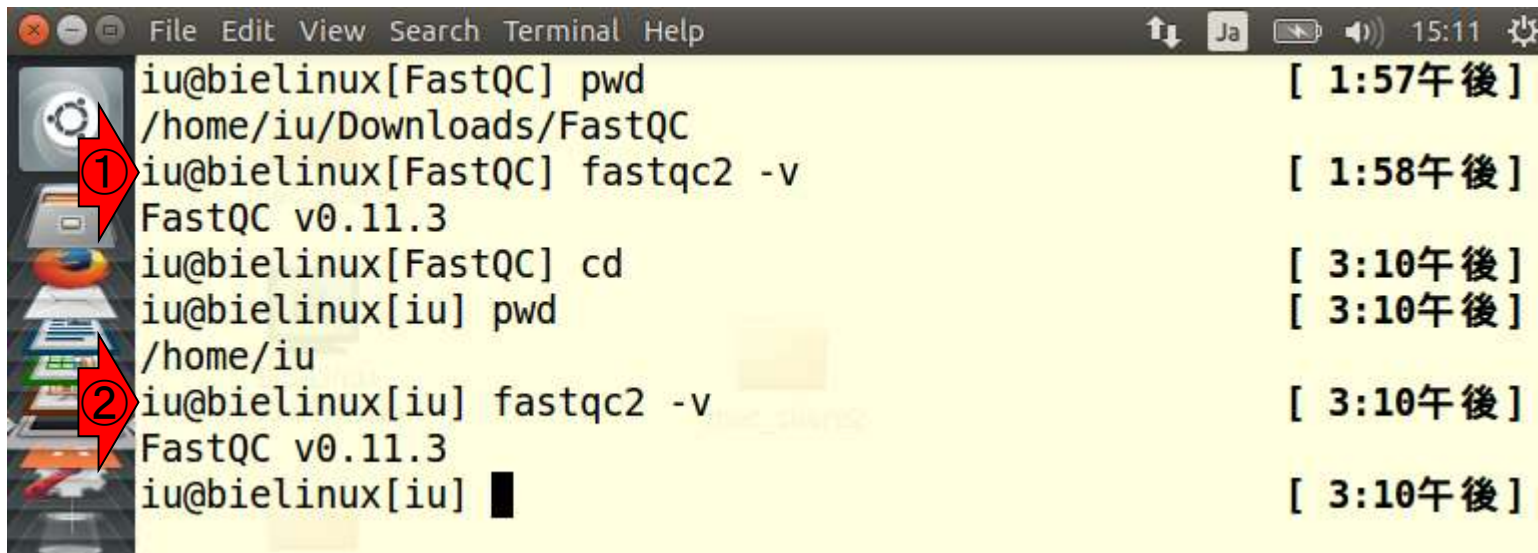
(作業ディレクトリはどこでもよい)①fastqまで打ってからTabキーを押した結果。パスが通っている(コマンド名を打ち込むだけで利用可能な)fastq*という条件を満たす候補がリストアップされる。シンボリックリンクを張ったfastqc2が見られる。

```
iu@bielinux[~/Downloads/FastQC]
iu@bielinux[FastQC] pwd
/home/iu/Downloads/FastQC
iu@bielinux[FastQC] fastq [ 1:58午後 ]
fastqc
fastqc2 ←
fastq_masker
fastq_quality_boxplot_graph.sh
fastq_quality_converter
fastq_quality_filter
fastq_quality_trimmer
fastq_to_fasta
```



W9-6: 確認

(作業ディレクトリはどこでもよい)
という証拠を示しているだけ。



```
File Edit View Search Terminal Help
iu@bielinux[FastQC] pwd [ 1:57午後 ]
/home/iu/Downloads/FastQC
① iu@bielinux[FastQC] fastqc2 -v [ 1:58午後 ]
FastQC v0.11.3
iu@bielinux[FastQC] cd [ 3:10午後 ]
iu@bielinux[iu] pwd [ 3:10午後 ]
/home/iu
② iu@bielinux[iu] fastqc2 -v [ 3:10午後 ]
FastQC v0.11.3
iu@bielinux[iu] [ 3:10午後 ]
```

W9-7: Tips (FastQC)

FastQCプログラム実行結果を任意のディレクトリに保存するように指定することもできる(デフォルトではカレントディレクトリ上に作成)。
①以前作成したFastQC実行結果を削除しておき、確認。FastQC実行結果を共有フォルダ(~/Desktop/mac_share)に保存するように指定すべく、何もない状態にしてから②確認。

```
iu@bielinux[srp017156] pwd
/home/iu/Documents/srp017156
iu@bielinux[srp017156] ls -lh
total 15G
-rwxr-xr-x 1 iu iu 360 5月 1 18:01 JSLAB4_1_Linux.sh
-rwxr-xr-x 1 iu iu 424 5月 5 15:07 JSLAB4_2_Linux.sh
-rw-rw-r-- 1 iu iu 7.2G 12月 11 14:51 SRR616268_1.fastq.bz2
-rw-rw-r-- 1 iu iu 6.6G 12月 12 14:59 SRR616268_2.fastq.bz2
-rw-rw-r-- 1 iu iu 306M 5月 6 10:07 SRR616268sub_1.fastq
-rw-rw-r-- 1 iu iu 278M 5月 6 10:08 SRR616268sub_2.fastq
iu@bielinux[srp017156] ls ~/Desktop/mac_share [ 3:42午後 ]
iu@bielinux[srp017156] [ 3:43午後 ]
```



W9-7: Tips (FastQC)

FastQCプログラム実行時に「--outdir=保存したいディレクトリのパス名」をオプションとして与える。①相対パス指定はうまくいかない、②絶対パス指定でやったらうまくいった。③「~/Desktop/mac_share」の中身を確認。確かにFastQC (ver. 0.11.3)実行結果ファイルが作成されていることがわかる。なぜ相対パスでうまくいかないかと思い悩む暇があったら、絶対パスでうまくいったという記憶のみ残して次に進むべし!

```
iu@bielinux[srp017156] pwd
/home/iu/Documents/srp017156
iu@bielinux[srp017156] ls -lh
total 15G
-rwxr-xr-x 1 iu iu 360 5月 1 18:01 JSLAB4_
-rwxr-xr-x 1 iu iu 424 5月 5 15:07 JSLAB4_
-rw-rw-r-- 1 iu iu 7.2G 12月 11 14:51 SRR6162
-rw-rw-r-- 1 iu iu 6.6G 12月 12 14:59 SRR6162
-rw-rw-r-- 1 iu iu 306M 5月 6 10:07 SRR616268sub_1.fastq
-rw-rw-r-- 1 iu iu 278M 5月 6 10:08 SRR616268sub_2.fastq
iu@bielinux[srp017156] ls ~/Desktop/mac_share [ 3:42午後 ]
iu@bielinux[srp017156] fastqc2 -q SRR616268sub_1.fastq --outdir=
~/Desktop/mac_share
Specified output directory '~/Desktop/mac_share' does not exist
iu@bielinux[srp017156] fastqc2 -q SRR616268sub_1.fastq --outdir=
/home/iu/Desktop/mac_share
iu@bielinux[srp017156] ls ~/Desktop/mac_share [ 3:52午後 ]
SRR616268sub_1_fastqc.html SRR616268sub_1_fastqc.zip
iu@bielinux[srp017156] [ 4:01午後 ]
```

①

②

③

W9-7: Tips (FastQC)

FastQCプログラムの入力は、非圧縮状態のFASTQ形式ファイル(.fastq)でなくてもよい。①gzip圧縮FASTQファイル(.fastq.gz)を作成して、②確認。③.fastq.gzを入力としてfastqc2を実行。④共有フォルダに確かに作成されていることがわかる。

```
iu@bielinux[srp017156] pwd
/home/iu/Documents/srp017156
iu@bielinux[srp017156] ls
JSLAB4_1_Linux.sh  SRR616268_1.fastq.bz2  SRR616268sub_1.fastq
JSLAB4_2_Linux.sh  SRR616268_2.fastq.bz2  SRR616268sub_2.fastq
① iu@bielinux[srp017156] time gzip SRR616268sub_2.fastq
gzip SRR616268sub_2.fastq 45.85s user 3.66s system 90% cpu 54.8
30 total
② iu@bielinux[srp017156] ls [ 4:13午後]
JSLAB4_1_Linux.sh  SRR616268_2.fastq.bz2
JSLAB4_2_Linux.sh  SRR616268sub_1.fastq
SRR616268_1.fastq.bz2  SRR616268sub_2.fastq.gz
③ iu@bielinux[srp017156] fastqc2 -q SRR616268sub_2.fastq.gz --outd
ir=/home/iu/Desktop/mac_share
④ iu@bielinux[srp017156] ls ~/Desktop/mac_share [ 4:14午後]
SRR616268sub_1_fastqc.html  SRR616268sub_2_fastqc.html
SRR616268sub_1_fastqc.zip  SRR616268sub_2_fastqc.zip
iu@bielinux[srp017156] [ 4:18午後]
```

W9-8: Tips (長いコマンド)

```
iu@bielinux[srp017156] pwd
/home/iu/Documents/srp017156
① iu@bielinux[srp017156] ls ~/Desktop/mac_share
SRR616268sub_1_fastqc.html  SRR616268sub_2_fastqc.html
SRR616268sub_1_fastqc.zip  SRR616268sub_2_fastqc.zip
② iu@bielinux[srp017156] ls ~/Desktop/mac_share/*_2_*
/home/iu/Desktop/mac_share/SRR616268sub_2_fastqc.html
/home/iu/Desktop/mac_share/SRR616268sub_2_fastqc.zip
③ iu@bielinux[srp017156] rm -f ~/Desktop/mac_share/*_2_*
iu@bielinux[srp017156] ls ~/Desktop/mac_share
SRR616268sub_1_fastqc.html  SRR616268sub_1_fastqc.zip
iu@bielinux[srp017156] ls
JSLAB4_1_Linux.sh  SRR616268_1.fastq.bz2  SRR616268sub_1.fastq.gz
JSLAB4_2_Linux.sh  SRR616268_2.fastq.bz2  SRR616268sub_2.fastq.gz
④ iu@bielinux[srp017156] fastqc2 -q SRR616268sub_2.fastq.gz\ [ 1:01午後 ]
[ 1:08午後 ]
```

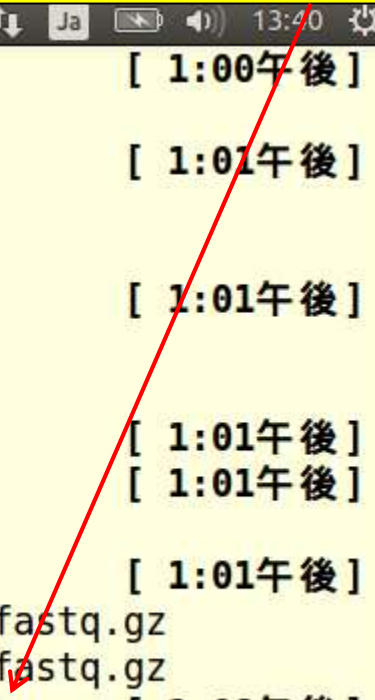
FastQC実行時に保存先ディレクトリ指定まで行くと、通常の画面サイズだと見た目上複数行にわたる長い入力となり、スペースの誤認識や誤入力の原因となる。それを避け、明示的に複数行で1つの長いコマンドを入力するとき用いるのが\`\`(バックスラッシュ)である。W5-2で書いてあるように改行コードの関係で\`\`が¥に見えたり(あるいはその逆)するが同じものなので気にしなくて良い。

Macで「逆スラッシュ」を出したい場合は、「Alt + ¥」で出るらしい

W9-8: Tips (長いコマンド)

リターンキーを押した結果。「>」という入力待ち状態となり、コマンドの続きを入力できる。

```
iu@bielinux[srp017156] pwd [ 1:00午後 ]
/home/iu/Documents/srp017156
iu@bielinux[srp017156] ls ~/Desktop/mac_share [ 1:01午後 ]
SRR616268sub_1_fastqc.html SRR616268sub_2_fastqc.html
SRR616268sub_1_fastqc.zip SRR616268sub_2_fastqc.zip
iu@bielinux[srp017156] ls ~/Desktop/mac_share/*_2_* [ 1:01午後 ]
/home/iu/Desktop/mac_share/SRR616268sub_2_fastqc.html
/home/iu/Desktop/mac_share/SRR616268sub_2_fastqc.zip
iu@bielinux[srp017156] rm -f ~/Desktop/mac_share/*_2_* [ 1:01午後 ]
iu@bielinux[srp017156] ls ~/Desktop/mac_share [ 1:01午後 ]
SRR616268sub_1_fastqc.html SRR616268sub_1_fastqc.zip
iu@bielinux[srp017156] ls [ 1:01午後 ]
JSLAB4_1_Linux.sh SRR616268_1.fastq.bz2 SRR616268sub_1.fastq.gz
JSLAB4_2_Linux.sh SRR616268_2.fastq.bz2 SRR616268sub_2.fastq.gz
iu@bielinux[srp017156] fastqc2 -q SRR616268sub_2.fastq.gz\ [ 1:08午後 ]
>
```



W9-8: Tips (長いコメント)

「--outdir=/home/iu/Desktop/mac_share」と打ち込んでリターンキーを押した結果。英語が苦手なヒトも、否定形の文章であることから、うまくいってないことくらいはわかるでしょう。

```
iu@bielinux[srp017156] pwd
/home/iu/Documents/srp017156
iu@bielinux[srp017156] ls ~/Desktop/mac_share [ 1:01午後 ]
SRR616268sub_1_fastqc.html SRR616268sub_2_fastqc.html
SRR616268sub_1_fastqc.zip SRR616268sub_2_fastqc.zip
iu@bielinux[srp017156] ls ~/Desktop/mac_share/*_2_* [ 1:01午後 ]
/home/iu/Desktop/mac_share/SRR616268sub_2_fastqc.html
/home/iu/Desktop/mac_share/SRR616268sub_2_fastqc.zip
iu@bielinux[srp017156] rm -f ~/Desktop/mac_share/*_2_* [ 1:01午後 ]
iu@bielinux[srp017156] ls ~/Desktop/mac_share [ 1:01午後 ]
SRR616268sub_1_fastqc.html SRR616268sub_1_fastqc.zip
iu@bielinux[srp017156] ls [ 1:01午後 ]
JSLAB4_1_Linux.sh SRR616268_1.fastq.bz2 SRR616268sub_1.fastq.gz
JSLAB4_2_Linux.sh SRR616268_2.fastq.bz2 SRR616268sub_2.fastq.gz
iu@bielinux[srp017156] fastqc2 -q SRR616268sub_2.fastq.gz\ [ 1:08午後 ]
> --outdir=/home/iu/Desktop/mac_share
Skipping 'SRR616268sub_2.fastq.gz--outdir=/home/iu/Desktop/mac_share' which
didn't exist, or couldn't be read
iu@bielinux[srp017156] [ 1:42午後 ]
```



W9-8: Tips (長いコマンド)

W9-7と同じコマンドになっているはずだが…と思ってはイケナイ!!赤矢印部分を見ればわかるが、確かに「スペース」を入力していなかったから、このような結果になってしまったのだろうと学習する。

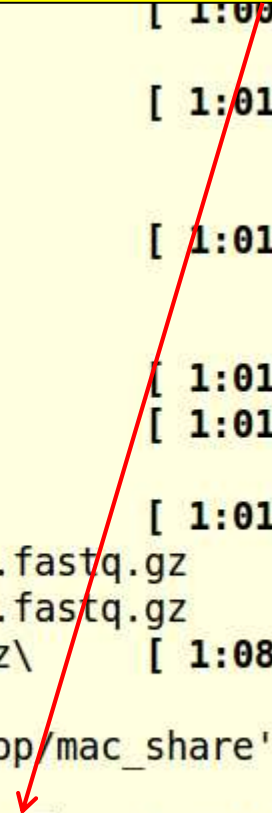
```
iu@bielinux[srp017156] pwd
/home/iu/Documents/srp017156
iu@bielinux[srp017156] ls ~/Desktop/mac_share [ 1:01午後 ]
SRR616268sub_1_fastqc.html SRR616268sub_2_fastqc.html
SRR616268sub_1_fastqc.zip SRR616268sub_2_fastqc.zip
iu@bielinux[srp017156] ls ~/Desktop/mac_share/*_2_* [ 1:01午後 ]
/home/iu/Desktop/mac_share/SRR616268sub_2_fastqc.html
/home/iu/Desktop/mac_share/SRR616268sub_2_fastqc.zip
iu@bielinux[srp017156] rm -f ~/Desktop/mac_share/*_2_* [ 1:01午後 ]
iu@bielinux[srp017156] ls ~/Desktop/mac_share [ 1:01午後 ]
SRR616268sub_1_fastqc.html SRR616268sub_1_fastqc.zip
iu@bielinux[srp017156] ls [ 1:01午後 ]
JSLAB4_1_Linux.sh SRR616268_1.fastq.bz2 SRR616268sub_1.fastq.gz
JSLAB4_2_Linux.sh SRR616268_2.fastq.bz2 SRR616268sub_2.fastq.gz
iu@bielinux[srp017156] fastqc2 -q SRR616268sub_2.fastq.gz\ [ 1:08午後 ]
> --outdir=/home/iu/Desktop/mac_share
Skipping 'SRR616268sub 2.fastq.gz--outdir=/home/iu/Desktop/mac share' which
didn't exist, or couldn't be read
iu@bielinux[srp017156] [ 1:42午後 ]
```



W9-8: Tips (長いコマンド)

上矢印キーを1回押して、直近のコマンドを出し、必要最小限の修正を施す。ヒトにもよるが、我々は「バックスラッシュ」の前に「スペース」を入れる。

```
iu@bielinux[srp017156] pwd [ 1:00午後 ]
/home/iu/Documents/srp017156
iu@bielinux[srp017156] ls ~/Desktop/mac_share [ 1:01午後 ]
SRR616268sub_1_fastqc.html SRR616268sub_2_fastqc.html
SRR616268sub_1_fastqc.zip SRR616268sub_2_fastqc.zip
iu@bielinux[srp017156] ls ~/Desktop/mac_share/*_2_* [ 1:01午後 ]
/home/iu/Desktop/mac_share/SRR616268sub_2_fastqc.html
/home/iu/Desktop/mac_share/SRR616268sub_2_fastqc.zip
iu@bielinux[srp017156] rm -f ~/Desktop/mac_share/*_2_* [ 1:01午後 ]
iu@bielinux[srp017156] ls ~/Desktop/mac_share [ 1:01午後 ]
SRR616268sub_1_fastqc.html SRR616268sub_1_fastqc.zip
iu@bielinux[srp017156] ls [ 1:01午後 ]
JSLAB4_1_Linux.sh SRR616268_1.fastq.bz2 SRR616268sub_1.fastq.gz
JSLAB4_2_Linux.sh SRR616268_2.fastq.bz2 SRR616268sub_2.fastq.gz
iu@bielinux[srp017156] fastqc2 -q SRR616268sub_2.fastq.gz\ [ 1:08午後 ]
> --outdir=/home/iu/Desktop/mac_share
Skipping 'SRR616268sub_2.fastq.gz--outdir=/home/iu/Desktop/mac_share' which
didn't exist, or couldn't be read
iu@bielinux[srp017156] fastqc2 -q SRR616268sub_2.fastq.gz [ 2:15午後 ]
--outdir=/home/iu/Desktop/mac_share
```



実行結果。①と②の比較から、うまく動いていることがわかる。

W9-8: Tips (長いコマンド)

```
iu@bielinux[srp017156] pwd [ 1:00午後 ]
/home/iu/Documents/srp017156
iu@bielinux[srp017156] ls ~/Desktop/mac_share [ 1:01午後 ]
SRR616268sub_1_fastqc.html SRR616268sub_2_fastqc.html
SRR616268sub_1_fastqc.zip SRR616268sub_2_fastqc.zip
iu@bielinux[srp017156] ls ~/Desktop/mac_share/*_2_* [ 1:01午後 ]
/home/iu/Desktop/mac_share/SRR616268sub_2_fastqc.html
/home/iu/Desktop/mac_share/SRR616268sub_2_fastqc.zip
iu@bielinux[srp017156] rm -f ~/Desktop/mac_share/*_2_* [ 1:01午後 ]
iu@bielinux[srp017156] ls ~/Desktop/mac_share [ 1:01午後 ]
SRR616268sub_1_fastqc.html SRR616268sub_1_fastqc.zip
iu@bielinux[srp017156] ls [ 1:01午後 ]
JSLAB4_1_Linux.sh SRR616268_1.fastq.bz2 SRR616268sub_1.fastq.gz
JSLAB4_2_Linux.sh SRR616268_2.fastq.bz2 SRR616268sub_2.fastq.gz
iu@bielinux[srp017156] fastqc2 -q SRR616268sub_2.fastq.gz\ [ 1:08午後 ]
> --outdir=/home/iu/Desktop/mac_share
Skipping 'SRR616268sub_2.fastq.gz--outdir=/home/iu/Desktop/mac_share' which
didn't exist, or couldn't be read
iu@bielinux[srp017156] fastqc2 -q SRR616268sub_2.fastq.gz \ [ 2:15午後 ]
--outdir=/home/iu/Desktop/mac_share
iu@bielinux[srp017156] ls ~/Desktop/mac_share [ 2:20午後 ]
SRR616268sub_1_fastqc.html SRR616268sub_2_fastqc.html
SRR616268sub_1_fastqc.zip SRR616268sub_2_fastqc.zip
iu@bielinux[srp017156] [ 2:21午後 ]
```



W9-9:パスの確認

①fastqc2、②fastqc、③perl、④ls、⑤pwd、⑥whereis、の存在場所をwhereisコマンドを用いて調べている。①fastqc2と②fastqcが/usr/local/binにあるのは、W9-4でも言及しており理解できる。

```
iu@bielinux[srp017156] pwd
/home/iu/Documents/srp017156
iu@bielinux[srp017156] whereis fastqc2 ← ① [ 3:34午後 ]
fastqc2: /usr/local/bin/fastqc2
iu@bielinux[srp017156] whereis fastqc ← ② [ 3:34午後 ]
fastqc: /usr/local/bin/fastqc
iu@bielinux[srp017156] whereis perl ← ③ [ 3:34午後 ]
perl: /usr/bin/perl /etc/perl /usr/lib/perl /usr/bin/X11/perl
/usr/share/perl /usr/share/man/man1/perl.1.gz
iu@bielinux[srp017156] whereis ls ← ④ [ 3:34午後 ]
ls: /bin/ls /usr/share/man/man1/ls.1.gz
iu@bielinux[srp017156] whereis pwd ← ⑤ [ 3:34午後 ]
pwd: /bin/pwd /usr/include/pwd.h /usr/share/man/man1/pwd.1.gz
iu@bielinux[srp017156] whereis whereis ← ⑥ [ 3:34午後 ]
whereis: /usr/bin/whereis /usr/bin/X11/whereis /usr/share/man
/man1/whereis.1.gz
iu@bielinux[srp017156] [ 3:58午後 ]
```

W9-9: パスの確認

④lsと⑤pwdは/binにあることがわかる。普通に使えるコマンドなので、/binにもデフォルトでパスが通っているんだな、と学習する。

```
iu@bielinux[srp017156] pwd [ 3:34午後 ]
/home/iu/Documents/srp017156
iu@bielinux[srp017156] whereis fastqc2 ← ① [ 3:34午後 ]
fastqc2: /usr/local/bin/fastqc2
iu@bielinux[srp017156] whereis fastqc ← ② [ 3:34午後 ]
fastqc: /usr/local/bin/fastqc
iu@bielinux[srp017156] whereis perl ← ③ [ 3:34午後 ]
perl: /usr/bin/perl /etc/perl /usr/lib/perl /usr/bin/X11/perl
/usr/share/perl /usr/share/man/man1/perl.1.gz
iu@bielinux[srp017156] whereis ls ← ④ [ 3:34午後 ]
ls: /bin/ls /usr/share/man/man1/ls.1.gz
iu@bielinux[srp017156] whereis pwd ← ⑤ [ 3:34午後 ]
pwd: /bin/pwd /usr/include/pwd.h /usr/share/man/man1/pwd.1.gz
iu@bielinux[srp017156] whereis whereis ← ⑥ [ 3:34午後 ]
whereis: /usr/bin/whereis /usr/bin/X11/whereis /usr/share/man
/man1/whereis.1.gz
iu@bielinux[srp017156] [ 3:58午後 ]
```

W9-9: パスの確認

③perlと⑥whereisは/usr/binにあることがわかる。普通に使えるコマンドなので、/usr/binにもデフォルトでパスが通っているんだな、と学習する。W2-3で、「#!/usr/bin/perl」と書いているのは、perlの絶対パスが/usr/bin/perlだと分かっているから。逆に、「whereis perl」実行結果が/usr/local/bin/perlとなっていれば、W2-3で書くのは「#!/usr/local/bin/perl」となる。

```
iu@bielinux[srp017156] pwd [
/home/iu/Documents/srp017156
iu@bielinux[srp017156] whereis fastqc2 ← ① [
fastqc2: /usr/local/bin/fastqc2
iu@bielinux[srp017156] whereis fastqc ← ② [
fastqc: /usr/local/bin/fastqc
iu@bielinux[srp017156] whereis perl ← ③ [
perl: /usr/bin/perl /etc/perl /usr/lib/perl /usr/bi
/usr/share/perl /usr/share/man/man1/perl.1.gz
iu@bielinux[srp017156] whereis ls ← ④ [ 3:34午後 ]
ls: /bin/ls /usr/share/man/man1/ls.1.gz
iu@bielinux[srp017156] whereis pwd ← ⑤ [ 3:34午後 ]
pwd: /bin/pwd /usr/include/pwd.h /usr/share/man/man1/pwd.1.gz
iu@bielinux[srp017156] whereis whereis ← ⑥ [ 3:34午後 ]
whereis: /usr/bin/whereis /usr/bin/X11/whereis /usr/share/man
/man1/whereis.1.gz
iu@bielinux[srp017156] [ 3:58午後 ]
```

W9-9: パスの確認

③perlの実行結果として、perlの実体が /usr/bin/perl 以外に /etc/perl や /usr/lib/perl など複数存在することがわかる。となると、例えばW5-2で用いたperlプログラムは一体どこに実体があるものを使ったことになるのだろうかとなり始める。

```
iu@bielinux[srp017156] pwd [
/home/iu/Documents/srp017156
iu@bielinux[srp017156] whereis fastqc2 ← ① [
fastqc2: /usr/local/bin/fastqc2
iu@bielinux[srp017156] whereis fastqc ← ② [ 3:34午後 ]
fastqc: /usr/local/bin/fastqc
iu@bielinux[srp017156] whereis perl ← ③ [ 3:34午後 ]
perl: /usr/bin/perl /etc/perl /usr/lib/perl /usr/bin/X11/perl
/usr/share/perl /usr/share/man/man1/perl.1.gz
iu@bielinux[srp017156] whereis ls ← ④ [ 3:34午後 ]
ls: /bin/ls /usr/share/man/man1/ls.1.gz
iu@bielinux[srp017156] whereis pwd ← ⑤ [ 3:34午後 ]
pwd: /bin/pwd /usr/include/pwd.h /usr/share/man/man1/pwd.1.gz
iu@bielinux[srp017156] whereis whereis ← ⑥ [ 3:34午後 ]
whereis: /usr/bin/whereis /usr/bin/X11/whereis /usr/share/man
/man1/whereis.1.gz
iu@bielinux[srp017156] [ 3:58午後 ]
```

W9-9:パスの確認

パスが通っている場所を確認するには、①「echo \$PATH」または②「env | grep PATH」と打つ。②は分かりにくいかもしれないが、envコマンド実行結果をパイプ(|)でgrepコマンドの入力として用い、PATHという文字列を含む行を出力しているだけである。この場合、envの出力結果中にPATHという文字列を含む行が全部で6行分存在していたと解釈する。③は、行頭がPATHから始まる行のみを出力するやり方。

```
iu@bielinux[srp017156] pwd
/home/iu/Documents/srp017156
iu@bielinux[srp017156] whereis perl
perl: /usr/bin/perl /etc/perl /usr/lib
/usr/share/perl /usr/share/man/man1/p
iu@bielinux[srp017156] echo $PATH [ 5:00午後 ]
/usr/local/sbin:/usr/local/bin:/usr/sbin:/usr/bin:/sbin:/bin:
/usr/games:/usr/local/games:/usr/lib/cd-hit:/usr/lib/cd-hit
iu@bielinux[srp017156] env | grep PATH [ 5:06午後 ]
DEFAULTS PATH=/usr/share/gconf/ubuntu.default.path
PATH=/usr/local/sbin:/usr/local/bin:/usr/sbin:/usr/bin:/sbin:
/bin:/usr/games:/usr/local/games:/usr/lib/cd-hit:/usr/lib/cd-
hit
XDG_SESSION_PATH=/org/freedesktop/DisplayManager/Session0
MANDATORY_PATH=/usr/share/gconf/ubuntu.mandatory.path
XDG_SEAT_PATH=/org/freedesktop/DisplayManager/Seat0
COMPIZ_BIN_PATH=/usr/bin/
iu@bielinux[srp017156] env | grep ^PATH [ 5:06午後 ]
PATH=/usr/local/sbin:/usr/local/bin:/usr/sbin:/usr/bin:/sbin:
/bin:/usr/games:/usr/local/games:/usr/lib/cd-hit:/usr/lib/cd-
hit
iu@bielinux[srp017156] [ 5:31午後 ]
```


W9-9: パスの確認

赤枠の「echo \$PATH」出力結果は、パスが通っている場所をコロン(:)で区切って示している。また、同じコマンドがあったとしても、優先度は左上の表示順になっている。つまり、優先度1は/usr/local/sbin、優先度2は/usr/local/bin、優先度3は/usr/sbinという具合。

```
File Edit View Search Terminal Help
iu@bielinux[srp017156] pwd
/home/iu/Documents/srp017156
iu@bielinux[srp017156] whereis perl [ 5:35午後 ]
perl: /usr/bin/perl /etc/perl /usr/lib/perl /usr/bin/X11/perl
      /usr/share/perl /usr/share/man/man1/perl.1.gz
iu@bielinux[srp017156] echo $PATH [ 5:35午後 ]
/usr/local/sbin:/usr/local/bin:/usr/sbin:/usr/bin:/sbin:/bin:
/usr/games:/usr/local/games:/usr/lib/cd-hit:/usr/lib/cd-hit
iu@bielinux[srp017156] [ 5:35午後 ]
```



W9-9:パスの確認

①「whereis perl」実行結果として、/usr/bin/perl、
/etc/perl、/usr/lib/perl、/usr/bin/X11/perl、
/usr/share/perlの5つが使える。このうち、②「echo
\$PATH」実行結果として見られるパスが通っていると
ころは/usr/binのみ。それゆえ、単にperlと打ち込むと
、実際に実行されるコマンドは/usr/bin/perlと解釈。

```
File Edit View Search Terminal Help
iu@bielinux[srp017156] pwd
/home/iu/Documents/srp017156
① iu@bielinux[srp017156] whereis perl
perl: /usr/bin/perl /etc/perl /usr/lib/perl /usr/bin/X11/perl
   /usr/share/perl /usr/share/man/man1/perl.1.gz
② iu@bielinux[srp017156] echo $PATH [ 5:35午後 ]
/usr/local/sbin:/usr/local/bin:/usr/sbin:/usr/bin:/sbin:/bin:
/usr/games:/usr/local/games:/usr/lib/cd-hit:/usr/lib/cd-hit
iu@bielinux[srp017156] [ 5:35午後 ]
```

W9-9:パスの確認

よく?!やりがちなミスは、新しいバージョンのプログラムをインストールしたのち、`/usr/bin`にシンボリックリンク[W9-5]を張ったものの、同名のプログラムがより優先度の高い`/usr/local/bin`に存在していたため、古いバージョンのプログラムを実行し続けていたというパターン(爆)。

```
File Edit View Search Terminal Help
iu@bielinux[srp017156] pwd
/home/iu/Documents/srp017156
① iu@bielinux[srp017156] whereis perl
perl: /usr/bin/perl /etc/perl /usr/lib/perl /usr/bin/X11/perl
   /usr/share/perl /usr/share/man/man1/perl.1.gz
② iu@bielinux[srp017156] echo $PATH [ 5:35午後 ]
/usr/local/sbin:/usr/local/bin:/usr/sbin:/usr/bin:/sbin:/bin:
/usr/games:/usr/local/games:/usr/lib/cd-hit:/usr/lib/cd-hit
iu@bielinux[srp017156] [ 5:35午後 ]
```

W9-9: パスの確認

/usr/bin/perlのバージョン情報を念のため確認すべく
①「/usr/bin/perl -v」。出力結果が長いので必要最小限のバージョン情報部分を含む行のみ出力させたい。全体を眺め、versionという文字を含む行のみgrepすればいいという判断を下す。②や③がその結果。

```
iu@bielinux[~]
① iu@bielinux[iu] /usr/bin/perl -v
This is perl 5, version 18, subversion 2 (v5.18.2) built for x86_64-linux-gnu
-thread-multi
(with 41 registered patches, see perl -V for more detail)

Copyright 1987-2013, Larry Wall

Perl may be copied only under the terms of either the Artistic License or the
GNU General Public License, which may be found in the Perl 5 source kit.

Complete documentation for Perl, including FAQ lists, should be found on
this system using "man perl" or "perldoc perl".  If you have access to the
Internet, point your browser at http://www.perl.org/, the Perl Home Page.

② iu@bielinux[iu] /usr/bin/perl -v | grep version [ 7:46午後 ]
This is perl 5, version 18, subversion 2 (v5.18.2) built for x86_64-linux-gnu
-thread-multi

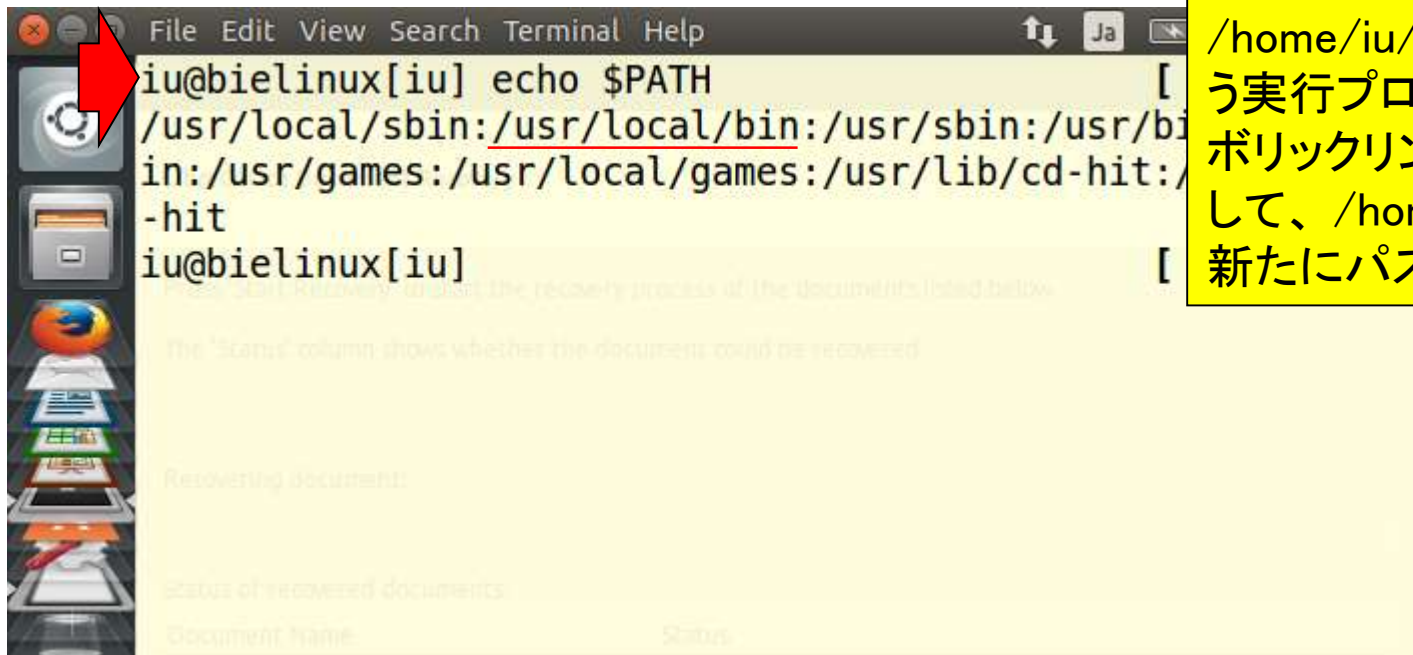
③ iu@bielinux[iu] /usr/bin/perl -v | grep "version" [ 7:46午後 ]
This is perl 5, version 18, subversion 2 (v5.18.2) built for x86_64-linux-gnu
-thread-multi
iu@bielinux[iu] █ [ 7:46午後 ]
```

W9-9: パスの確認

念のため、①「whereis perl」で見られるperlプログラムのバージョンを確認。②、③、⑤のように実行権限がない(permission denied)と怒られるものもある。④も同じバージョンであり、古いものを使っていなくて安心。

```
iu@bielinux[iu] whereis perl [ 7:46午後]
perl: /usr/bin/perl /etc/perl /usr/lib/perl /usr/bin/X11/perl /usr/share/perl
      /usr/share/man/man1/perl.1.gz
iu@bielinux[iu] /etc/perl -v | grep version [ 7:53午後]
zsh: permission denied: /etc/perl
iu@bielinux[iu] /usr/lib/perl -v | grep version [ 7:53午後]
zsh: permission denied: /usr/lib/perl
iu@bielinux[iu] /usr/bin/X11/perl -v | grep version [ 7:53午後]
This is perl 5, version 18, subversion 2 (v5.18.2) built for x86_64-linux-gnu
-thread-multi
iu@bielinux[iu] /usr/share/perl -v | grep version [ 7:53午後]
zsh: permission denied: /usr/share/perl
iu@bielinux[iu] [ 7:54午後]
```

W9-10: パスの設定



```
File Edit View Search Terminal Help
iu@bielinux[iu] echo $PATH
/usr/local/sbin:/usr/local/bin:/usr/sbin:/usr/bin:/usr/games:/usr/local/games:/usr/lib/cd-hit:/usr/bin/cd-hit
iu@bielinux[iu]
```

「echo \$PATH」では、パスが通っている場所を表示している。ここに新たにパスを通したい局面もある。W9-5では、
/home/iu/Downloads/FastQC/fastqcという実行プログラムを/usr/local/binにシンボリックリンクを貼った。これ以外の方法として、/home/iu/Downloads/FastQC/へ新たにパスを通すというやり方を伝授。

W9-10:パスの設定

②で現在通っているパスを表示。④で現在通っているパスに/home/iu/Downloads/FastQCを追加している。「PATH=\$PATH:/home/iu/Downloads/FastQC」と打ってから「export PATH」と打ってもよい。

```
iu@bielinux[iu] pwd
/home/iu
iu@bielinux[iu] ls -lh /home/iu/Downloads/FastQC/fast* ← ① [10:57午前]
-rwxr-xr-x 1 iu iu 14K 3月 24 16:23 /home/iu/Downloads/FastQC/fastqc
-rw-rw-r-- 1 iu iu 2.2K 3月 21 2012 /home/iu/Downloads/FastQC/fastqc_icon.ico
iu@bielinux[iu] echo $PATH ← ② [10:57午前]
/usr/local/sbin:/usr/local/bin:/usr/sbin:/usr/bin:/sbin:/bin:/usr/games:/usr/local/g
ames:/usr/lib/cd-hit:/usr/lib/cd-hit
iu@bielinux[iu] whereis fastqc ← ③ [10:57午前]
fastqc: /usr/local/bin/fastqc
iu@bielinux[iu] export PATH=PATH:/home/iu/Downloads/FastQC ← ④ [10:57午前]
iu@bielinux[iu] echo $PATH ← ⑤ [10:58午前]
/usr/local/sbin:/usr/local/bin:/usr/sbin:/usr/bin:/sbin:/bin:/usr/games:/usr/local/g
ames:/usr/lib/cd-hit:/usr/lib/cd-hit:/home/iu/Downloads/FastQC
iu@bielinux[iu] whereis fastqc ← ⑥ [10:58午前]
fastqc: /usr/local/bin/fastqc
iu@bielinux[iu] /home/iu/Downloads/FastQC/fastqc -v ← ⑦ [10:58午前]
FastQC v0.11.3
iu@bielinux[iu] /usr/local/bin/fastqc -v ← ⑧ [10:58午前]
FastQC v0.10.1
iu@bielinux[iu] fastqc -v ← ⑨ [10:59午前]
FastQC v0.10.1
iu@bielinux[iu] █ [10:59午前]
```

W9-10: パスの設定

⑥の結果は個人的には疑問。予想される結果はW9-9で「whereis perl」出力結果のように、「fastqc: /usr/local/bin/fastqc /home/iu/Downloads/FastQC/fastqc」であったが、そうはなっていない。結論としては、whereisではなくwhereでそうなる。

```
iu@bielinux[iu] pwd
/home/iu
iu@bielinux[iu] ls -lh /home/iu/Downloads/FastQC
-rwxr-xr-x 1 iu iu 14K  3月 24 16:23 /home/iu/Downloads/FastQC/fastqc
-rw-rw-r-- 1 iu iu 2.2K  3月 21  2012 /home/iu/Downloads/FastQC/fastqc_icon.ico
iu@bielinux[iu] echo $PATH ← ② [10:57午前]
/usr/local/sbin:/usr/local/bin:/usr/sbin:/usr/bin:/sbin:/bin:/usr/games:/usr/local/g
ames:/usr/lib/cd-hit:/usr/lib/cd-hit
iu@bielinux[iu] whereis fastqc ← ③ [10:57午前]
fastqc: /usr/local/bin/fastqc
iu@bielinux[iu] export PATH=$PATH:/home/iu/Downloads/FastQC ← ④ [10:57午前]
iu@bielinux[iu] echo $PATH ← ⑤ [10:58午前]
/usr/local/sbin:/usr/local/bin:/usr/sbin:/usr/bin:/sbin:/bin:/usr/games:/usr/local/g
ames:/usr/lib/cd-hit:/usr/lib/cd-hit:/home/iu/Downloads/FastQC
iu@bielinux[iu] whereis fastqc ← ⑥ [10:58午前]
fastqc: /usr/local/bin/fastqc
iu@bielinux[iu] /home/iu/Downloads/FastQC/fastqc -v ← ⑦ [10:58午前]
FastQC v0.11.3
iu@bielinux[iu] /usr/local/bin/fastqc -v ← ⑧ [10:58午前]
FastQC v0.10.1
iu@bielinux[iu] fastqc -v ← ⑨ [10:59午前]
FastQC v0.10.1
iu@bielinux[iu]
```


W9-11: Tips

コマンドの存在場所を調べるコマンドは複数ある (whereis, which, where)。このスライドの結果を眺めると、むしろwhereisの挙動が変なのだろう。特に⑤と⑥の結果を踏まえ、whichはパスが通っているものの中で優先度が高いもののみ表示、whereはパスが通っているコマンド全てを表示しているのだろうと解釈。

```
iu@bielinux[iu] echo $PATH
/usr/local/sbin:/usr/local/bin:/usr/sbin:/usr/bin:/sbin:/bin:/usr/games:/usr/local/games:/usr/lib/cd-hit:/usr/lib/cd-hit:/home/iu/Downloads/FastQC

iu@bielinux[iu] whereis perl ←———— ① [11:26午前]
perl: /usr/bin/perl /etc/perl /usr/lib/perl /usr/bin/X11/perl /usr/share/perl /usr/share/man/man1/perl.1.gz

iu@bielinux[iu] which perl ←———— ② [11:26午前]
/usr/bin/perl

iu@bielinux[iu] where perl ←———— ③ [11:26午前]
/usr/bin/perl

iu@bielinux[iu] whereis fastqc ←———— ④ [11:26午前]
fastqc: /usr/local/bin/fastqc

iu@bielinux[iu] which fastqc ←———— ⑤ [11:26午前]
/usr/local/bin/fastqc

iu@bielinux[iu] where fastqc ←———— ⑥ [11:26午前]
/usr/local/bin/fastqc
/home/iu/Downloads/FastQC/fastqc

iu@bielinux[iu] whereis fastqc2 ←———— ⑦ [11:27午前]
fastqc2: /usr/local/bin/fastqc2

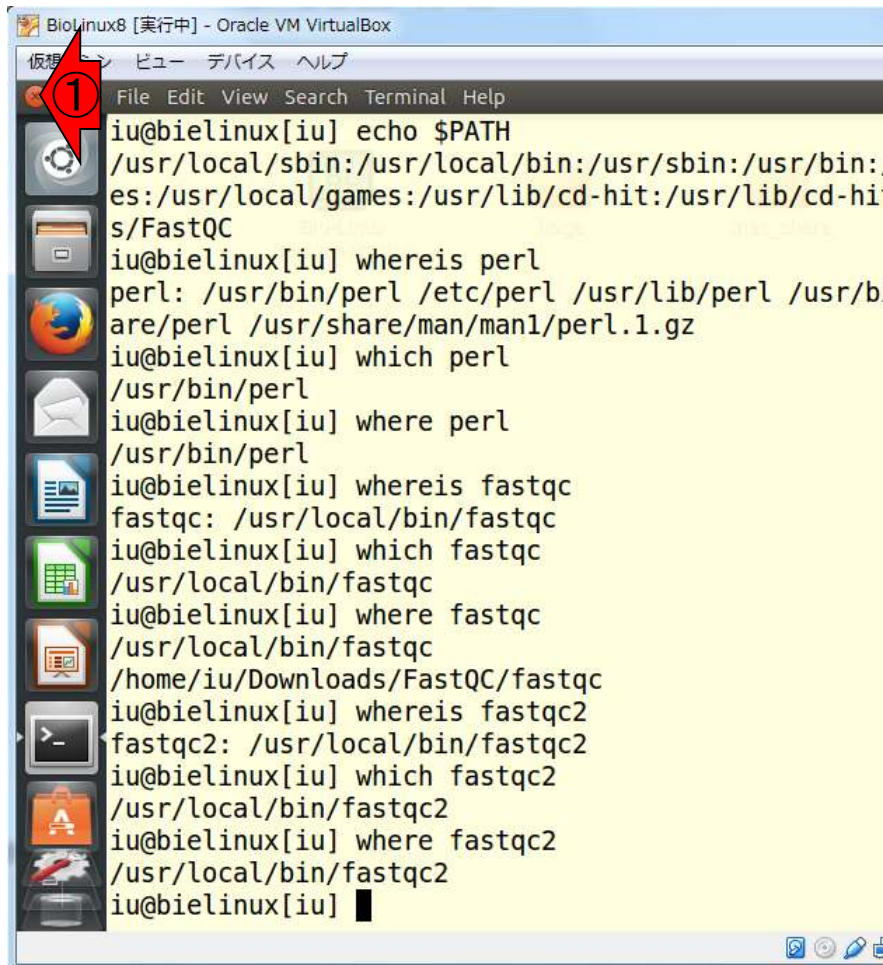
iu@bielinux[iu] which fastqc2 ←———— ⑧ [11:27午前]
/usr/local/bin/fastqc2

iu@bielinux[iu] where fastqc2 ←———— ⑨ [11:27午前]
/usr/local/bin/fastqc2

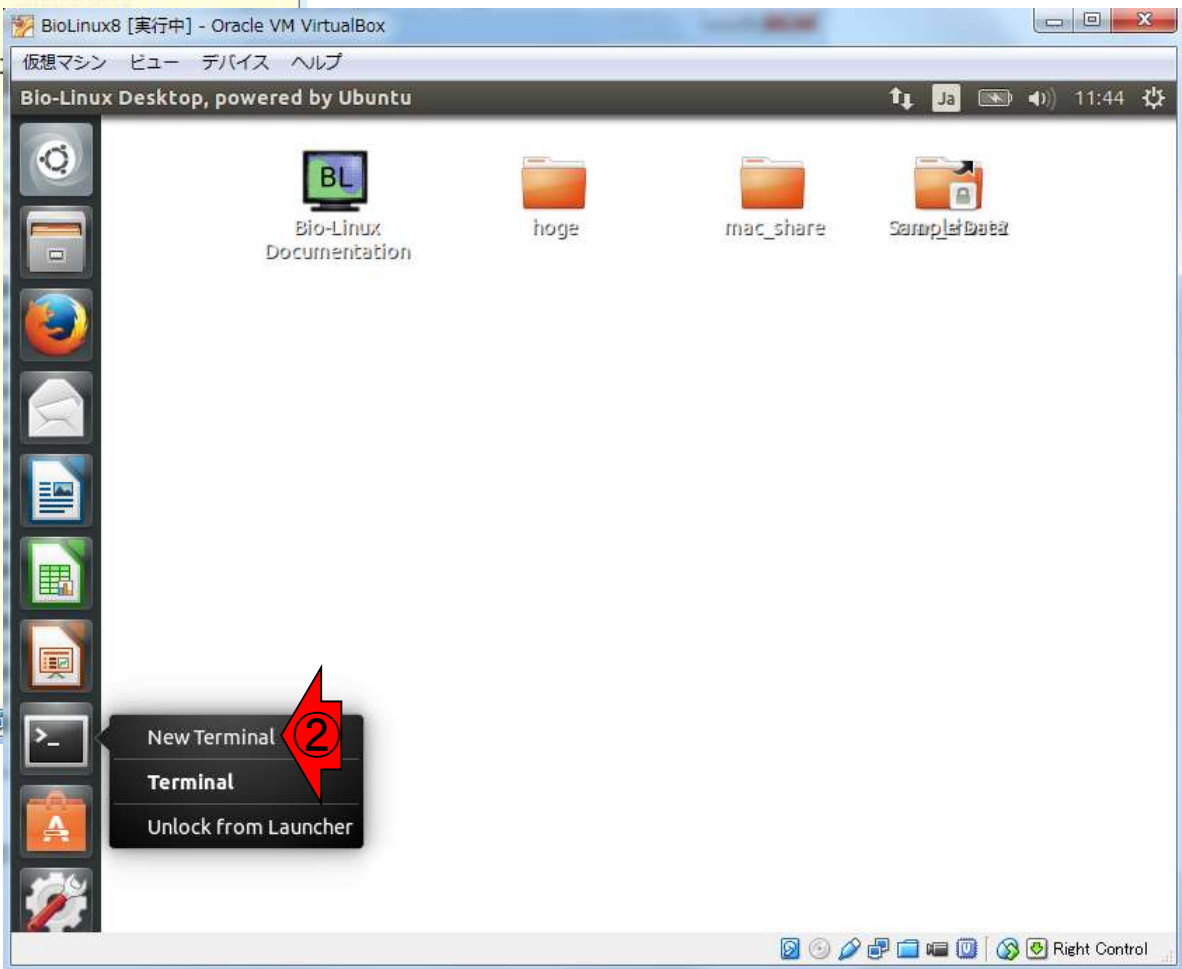
iu@bielinux[iu] █ [11:27午前]
```

W10-1: 環境設定

パスについても、共有フォルダの自動マウント設定(W4-5)で示すように、毎回設定をしなくて済むように環境設定ファイルの内容を書き換えるヒトのほうが多いかもしれないので、そのお話。①/home/iu/Downloads/FastQCのパスを追加したターミナル画面を一旦終了。②再度右クリックでNew Terminalを開く。

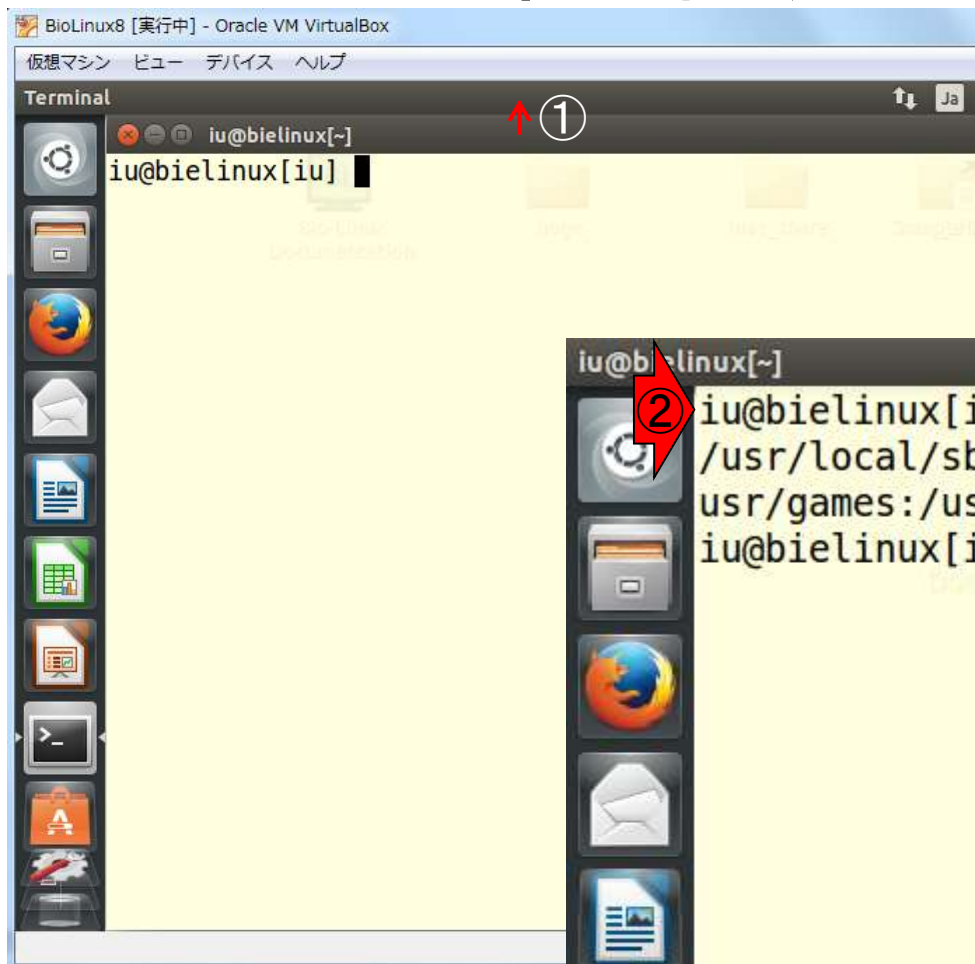


```
iu@bielinux[iu] echo $PATH
/usr/local/sbin:/usr/local/bin:/usr/sbin:/usr/bin:/s
es:/usr/local/games:/usr/lib/cd-hit:/usr/lib/cd-hit:/home/iu/Downloa
s/FastQC
iu@bielinux[iu] whereis perl
perl: /usr/bin/perl /etc/perl /usr/lib/perl /usr/b
are/perl /usr/share/man/man1/perl.1.gz
iu@bielinux[iu] which perl
/usr/bin/perl
iu@bielinux[iu] where perl
/usr/bin/perl
iu@bielinux[iu] whereis fastqc
fastqc: /usr/local/bin/fastqc
iu@bielinux[iu] which fastqc
/usr/local/bin/fastqc
iu@bielinux[iu] where fastqc
/usr/local/bin/fastqc
/home/iu/Downloads/FastQC/fastqc
iu@bielinux[iu] whereis fastqc2
fastqc2: /usr/local/bin/fastqc2
iu@bielinux[iu] which fastqc2
/usr/local/bin/fastqc2
iu@bielinux[iu] where fastqc2
/usr/local/bin/fastqc2
iu@bielinux[iu]
```

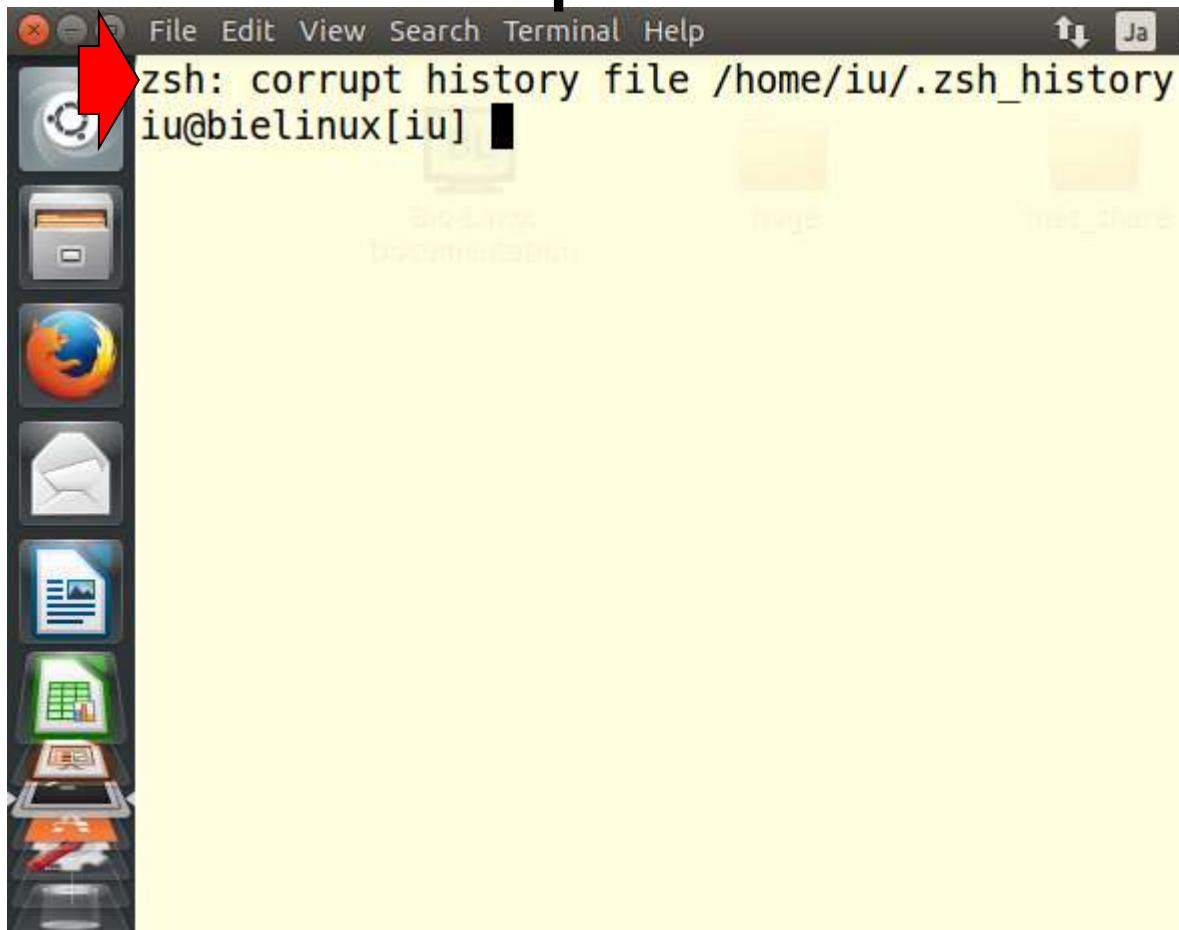


W10-1: 環境設定

①は矢印の下から上にマウสดラッグすると右下のような画面になる、というただのおさらい(連載第3回のW9-2-3)。②の結果から、以前打ち込んだ「export PATH=\$PATH:/home/iu/Downloads/FastQC」が無効になっていることがわかる。このことから、「W9-10のパス追加方法は開いているターミナル内でのみ有効」であることが分かる。

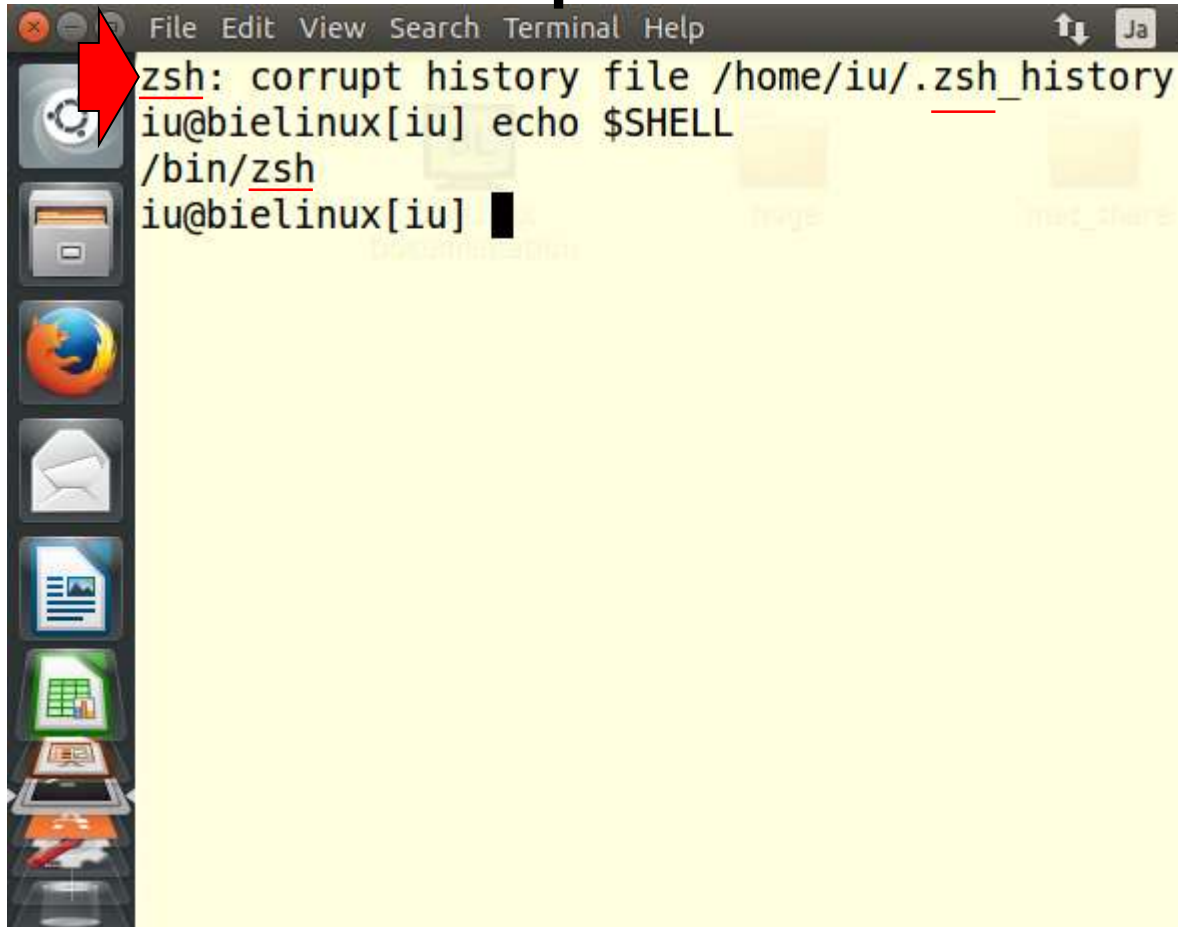


W10-2: Tips



W2-1の最後でも示したように、ときどきBio-Linuxがお亡くなりになります。しょうがないので、ゲストOSを再起動してターミナルを立ち上げたときに、「zsh: corrupt history file /home/**ユーザ名**/.zsh_history」となったりすることがあります。これはユーザ名iuのホームディレクトリ直下にある.zsh_historyというファイルが破損していることを意味します。基本無視でいいのですが、気になるヒト用に解決策を伝授します

W10-2: Tips



```
File Edit View Search Terminal Help
zsh: corrupt history file /home/iu/.zsh_history
iu@bielinux[iu] echo $SHELL
/bin/zsh
iu@bielinux[iu]
```

A screenshot of a Linux terminal window. The window title is "Terminal" and it has a menu bar with "File", "Edit", "View", "Search", "Terminal", and "Help". The terminal output shows an error message: "zsh: corrupt history file /home/iu/.zsh_history". Below that, the user enters the command "echo \$SHELL" and the output is "/bin/zsh". A red arrow points to the first line of the terminal output.

赤矢印の最初の部分は、ユーザiuが使っているログインシェルは、zshだということ。基本的にBio-Linux 8のデフォルトはzshというものらしい、という程度の理解でよい。将来どこか別の研究所のスパコンなどを使うことがあるかもしれないが、そこではbashというものかもしれないし、tcshというものかもしれない。このあたりについては、ウェブ上で検索したときに流派(zsh, bash, tcsh)ごとの情報があるので、適宜「echo \$SHELL」実行結果から自分が利用しているシェルの情報を入手してうまく置き換えるとよい。例えば、「.bash_profile ↔ .zsh_profile」

W10-2: Tips

```
zsh: corrupt history file /home/iu/.zsh_history
iu@bielinux[iu] echo $SHELL
/bin/zsh
① iu@bielinux[iu] ls -lh
total 32K
drwx----- 6 iu iu 4.0K  4月 30 13:47 Desktop
drwxr-xr-x  4 iu iu 4.0K 12月 10 17:55 Documents
drwxr-xr-x  3 iu iu 4.0K  5月  7 14:46 Downloads
drwxr-xr-x  2 iu iu 4.0K 11月 10  2014 Music
drwxr-xr-x  2 iu iu 4.0K 11月 29 12:06 Pictures
drwxr-xr-x  2 iu iu 4.0K 11月 10  2014 Public
drwxr-xr-x  2 iu iu 4.0K 11月 10  2014 Templates
drwxr-xr-x  2 iu iu 4.0K 11月 10  2014 Videos
iu@bielinux[iu] pwd
/home/iu
② iu@bielinux[iu] ls -la
```

.zsh_historyファイルを含む、.(どっと)から始まるファイルは①ls -lhで表示されない。これは、どっと(.)から始まるファイルは設定ファイル関連のものであり、(よく分かっていないエンドユーザにいじられないよう)通常は表示されないようになっているからである。どっと(.)から始まるファイルも表示するためには②lsコマンド実行時に-aオプションをつける。もちろん「ls -lha」や「ls -a」などでもよい。

[3:09午後]

[3:09午後]

W10-2: Tips

「ls -la」実行結果。画面がざっと流れて全部表示されないくらい、どっと(.)から始まるファイルが多数存在することがわかる。そして、ターミナル起動時に文句を言われていた.zsh_historyファイルが確かに存在することがわかる。

```
File Edit View Search Terminal Help
drwxr-xr-x  2 iu  iu   4096 11月 10  2014 Musi
drwxr-xr-x  2 iu  iu   4096 11月 29 12:06 Pict
-rw-r--r--  1 iu  iu    676  5月 14 14:19 .pro
-rw-r--r--  1 iu  iu    675 11月 10  2014 .profile~
drwxr-xr-x  2 iu  iu   4096 11月 10  2014 Public
drwxr-xr-x  2 iu  iu   4096 11月 10  2014 Templates
-rw-r----- 1 iu  iu     5  5月 14 14:23 .vboxclient-clipboard.pid
-rw-r----- 1 iu  iu     5  5月 14 14:23 .vboxclient-display.pid
-rw-r----- 1 iu  iu     5  5月 14 14:23 .vboxclient-draganddrop.pid
-rw-r----- 1 iu  iu     5  5月 14 14:23 .vboxclient-seamless.pid
drwxr-xr-x  2 iu  iu   4096 11月 10  2014 Videos
-rw-----  1 iu  iu     0  5月 14 14:19 .viminfo
-rw-----  1 iu  iu    53  5月 14 14:23 .Xauthority
-rw-----  1 iu  iu   352  5月 14 14:23 .xsession-errors
-rw-----  1 iu  iu   352  5月 14 14:20 .xsession-errors.old
-rw-rw-r--  1 iu  iu 35211 11月 10  2014 .zcompdump
-rw-----  1 iu  iu  8114  5月 14 14:40 .zsh_history
-rw-----  1 iu  iu   296  5月 14 14:14 .zshrc
iu@bielinux[iu]
```

[3:18午後]

W10-2: Tips

①見やすいようにls -la出力結果からhistoryというキーワードを含む行のみを表示。②基本的に.zsh_historyファイルの中身は「連載第2回のW16」で示すようなコマンド履歴情報のようなもの。それゆえ消しても基本的に差支えない。③は消えていることの確認。

```
File Edit View Search Terminal Help
-rw-r----- 1 iu iu      5  5月 14 14:23 .vboxc
-rw-r----- 1 iu iu      5  5月 14 14:23 .vboxc
-rw-r----- 1 iu iu      5  5月 14 14:23 .vboxc
drwxr-xr-x  2 iu iu    4096 11月 10  2014 Videos
-rw-----  1 iu iu       0  5月 14 14:19 .viminfo
-rw-----  1 iu iu      53  5月 14 14:23 .Xauthority
-rw-----  1 iu iu     352  5月 14 14:23 .xsession-errors
-rw-----  1 iu iu     352  5月 14 14:20 .xsession-errors.old
-rw-rw-r--  1 iu iu   35211 11月 10  2014 .zcompdump
-rw-----  1 iu iu    8114  5月 14 14:40 .zsh_history
-rw-----  1 iu iu     296  5月 14 14:14 .zshrc
① iu@bielinux[iu] ls -la | grep history           [ 3:35午後 ]
-rw-----  1 iu iu    8114  5月 14 14:40 .zsh_history
iu@bielinux[iu] pwd                               [ 3:35午後 ]
/home/iu
② iu@bielinux[iu] rm ~/.zsh_history               [ 3:36午後 ]
rm: remove regular file '/home/iu/.zsh_history'? y
③ iu@bielinux[iu] ls -la | grep history           [ 3:36午後 ]
iu@bielinux[iu] █                                 [ 3:36午後 ]
```


W10-2: Tips

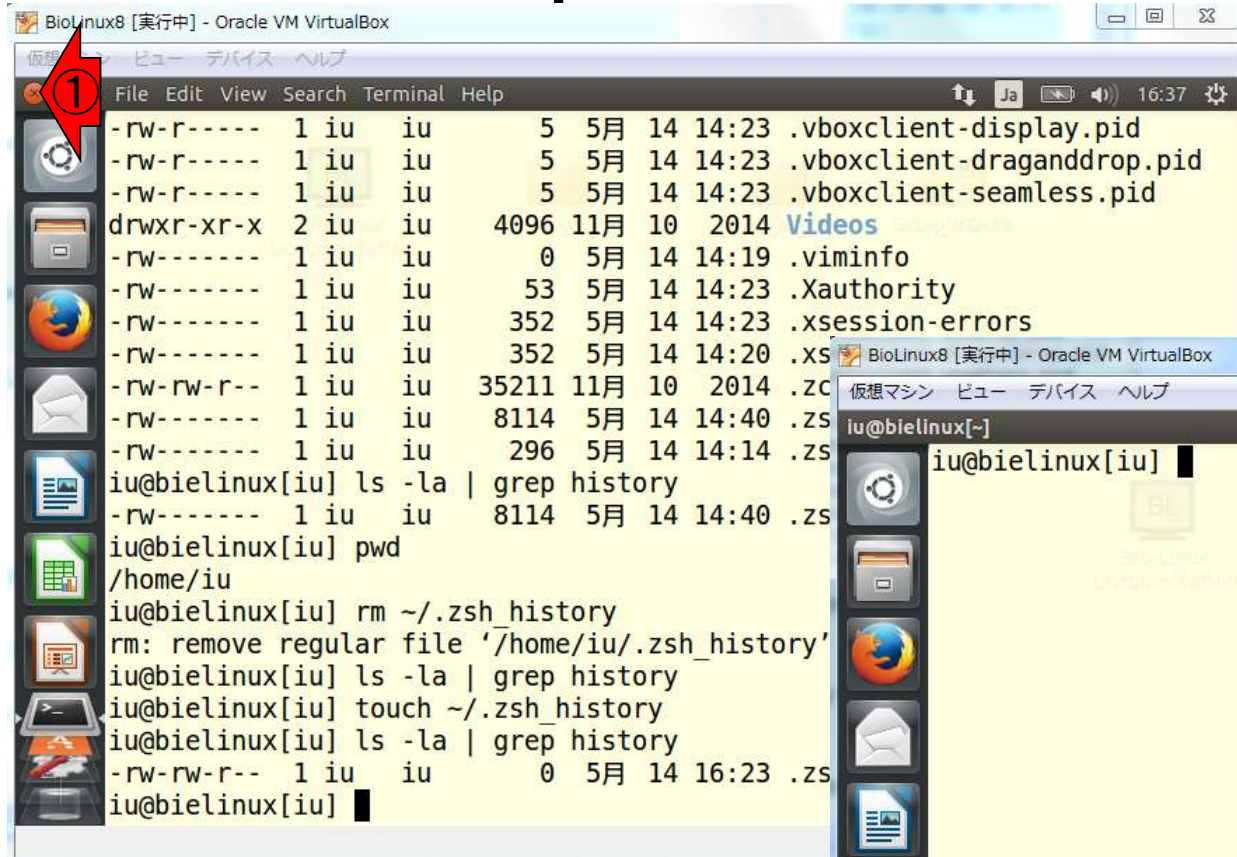
④のtouchコマンドはタイム・スタンプ変更を行うというもの。ここでやっている作業は、「現在時刻が最終更新日の .zsh_history という中身が空のファイル」の作成である。

```
File Edit View Search Terminal Help
-rw-r----- 1 iu iu      5 5月 14 14:23 .vboxc
-rw-r----- 1 iu iu      5 5月 14 14:23 .vboxclient-druganddrop.p
-rw-r----- 1 iu iu      5 5月 14 14:23 .vboxclient-seamless.pid
drwxr-xr-x  2 iu iu    4096 11月 10  2014 Videos
-rw-----  1 iu iu      0 5月 14 14:19 .viminfo
-rw-----  1 iu iu     53 5月 14 14:23 .Xauthority
-rw-----  1 iu iu    352 5月 14 14:23 .xsession-errors
-rw-----  1 iu iu    352 5月 14 14:20 .xsession-errors.old
-rw-rw-r--  1 iu iu   35211 11月 10  2014 .zcompdump
-rw-----  1 iu iu    8114 5月 14 14:40 .zsh_history
-rw-----  1 iu iu     296 5月 14 14:14 .zshrc
iu@bielinux[iu] ls -la | grep history           [ 3:35午後 ]
-rw-----  1 iu iu    8114 5月 14 14:40 .zsh_history
iu@bielinux[iu] pwd                             [ 3:35午後 ]
/home/iu
iu@bielinux[iu] rm ~/.zsh_history                [ 3:36午後 ]
rm: remove regular file '/home/iu/.zsh_history'? y
iu@bielinux[iu] ls -la | grep history           [ 3:36午後 ]
iu@bielinux[iu] touch ~/.zsh_history            [ 3:36午後 ]
iu@bielinux[iu] ls -la | grep history           [ 4:23午後 ]
-rw-rw-r--  1 iu iu      0 5月 14 16:23 .zsh_history
iu@bielinux[iu] █                               [ 4:26午後 ]
```

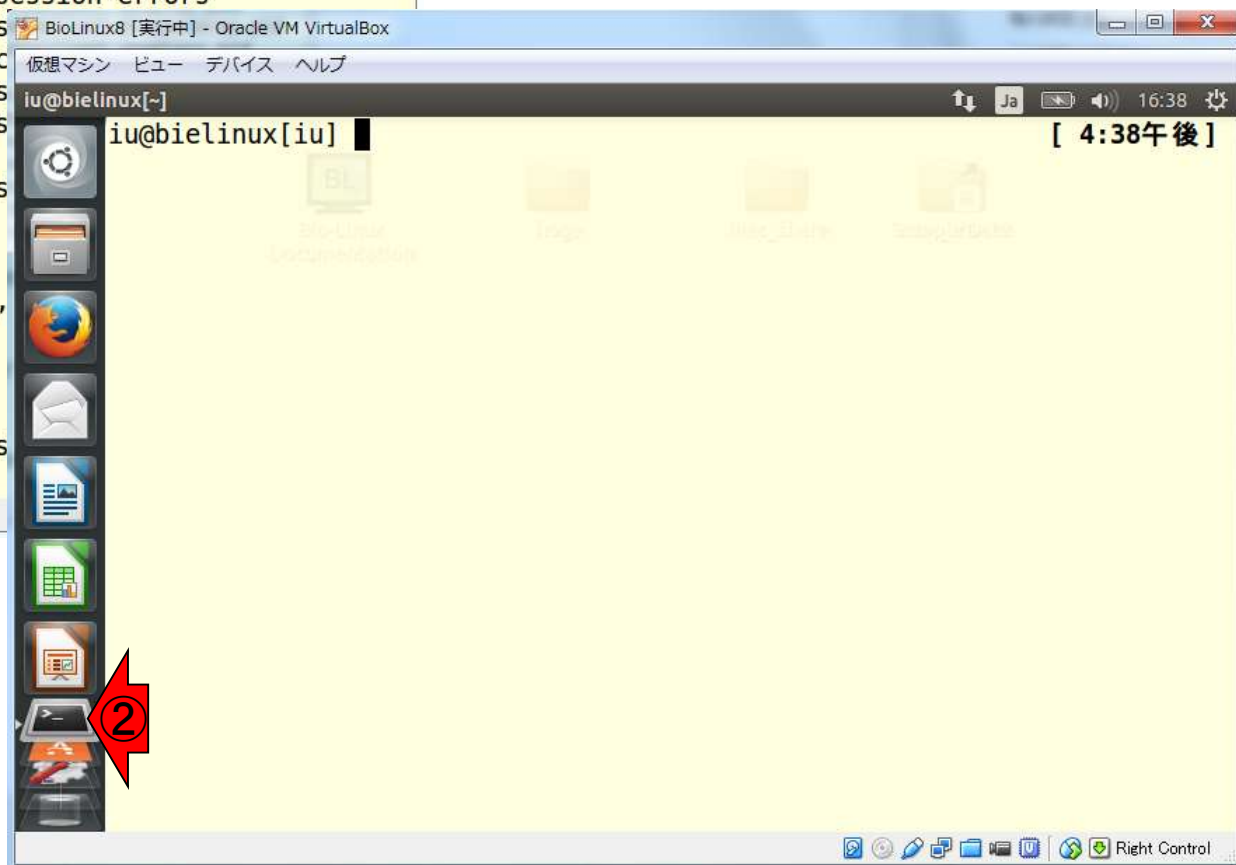


W10-2: Tips

①一旦ターミナルを閉じて、②New Terminalを開くと、.zsh_historyが破損しているというメッセージが無事消えた。



```
File Edit View Search Terminal Help
-rw-r----- 1 iu iu 5 5月 14 14:23 .vboxclient-display.pid
-rw-r----- 1 iu iu 5 5月 14 14:23 .vboxclient-draganddrop.pid
-rw-r----- 1 iu iu 5 5月 14 14:23 .vboxclient-seamless.pid
drwxr-xr-x 2 iu iu 4096 11月 10 2014 Videos
-rw----- 1 iu iu 0 5月 14 14:19 .viminfo
-rw----- 1 iu iu 53 5月 14 14:23 .Xauthority
-rw----- 1 iu iu 352 5月 14 14:23 .xsession-errors
-rw----- 1 iu iu 352 5月 14 14:20 .X
-rw-rw-r-- 1 iu iu 35211 11月 10 2014 .ZC
-rw----- 1 iu iu 8114 5月 14 14:40 .ZS
-rw----- 1 iu iu 296 5月 14 14:14 .ZS
iu@bielinux[iu] ls -la | grep history
-rw----- 1 iu iu 8114 5月 14 14:40 .ZS
iu@bielinux[iu] pwd
/home/iu
iu@bielinux[iu] rm ~/.zsh_history
rm: remove regular file '/home/iu/.zsh_history'
iu@bielinux[iu] ls -la | grep history
iu@bielinux[iu] touch ~/.zsh_history
iu@bielinux[iu] ls -la | grep history
-rw-rw-r-- 1 iu iu 0 5月 14 16:23 .ZS
iu@bielinux[iu]
```



```
iu@bielinux[~]
iu@bielinux[iu]
```

W10-3: 環境設定

W10-1の続き。①ログインシェル(ターミナルを起動したときのデフォルトのシェル)はzsh。この場合、どっと(.)から始まり、ログインシェル名の後ろにrcがついた.zshrcというファイルを開き、「export PATH=\$PATH:/home/iu/Downloads/FastQC」という行を追加すればよい。

```
File Edit View Search Terminal Help
iu@bielinux[iu] pwd
/home/iu
iu@bielinux[iu] echo $PATH
/usr/local/sbin:/usr/local/bin:/usr/sbin:/usr/bin:/sbin:/bin:/usr/games:/usr/local/games:/usr/lib/cd-hit:/usr/lib/cd-hit
iu@bielinux[iu] echo $SHELL
/bin/zsh
iu@bielinux[iu] ls -lha .*rc
-rw-r--r-- 1 iu iu 3.6K  5月 14 13:56 .bashrc
-rw-r--r-- 1 iu iu  25 11月 10 2014 .dmrc
-rw----- 1 iu iu  296  5月 14 14:14 .zshrc
iu@bielinux[iu]
```



[4:56午後]
[4:56午後]
[4:56午後]

W10-3: 環境設定

- ①編集前後の状況を把握すべく、.zshrcファイルの最後の5行分を表示。
- ②geditで.zshrcファイルを編集

```
iu@bielinux[iu] pwd [ 4:56午後 ]
/home/iu
iu@bielinux[iu] echo $PATH [ 4:56午後 ]
/usr/local/sbin:/usr/local/bin:/usr/sbin:/usr/bin:/sbin:/bin:/usr/games:/usr/local/games:/usr/lib/cd-hit:/usr/lib/cd-hit
iu@bielinux[iu] echo $SHELL [ 4:56午後 ]
/bin/zsh
iu@bielinux[iu] ls -lha .*rc [ 4:56午後 ]
-rw-r--r-- 1 iu iu 3.6K 5月 14 13:56 .bashrc
-rw-r--r-- 1 iu iu 25 11月 10 2014 .dmrc
-rw----- 1 iu iu 296 5月 14 14:14 .zshrc
iu@bielinux[iu] tail -n 5 .zshrc [ 4:56午後 ]
#
# if [[ -n "" ]] && [[ -z "" ]]; then
#   screen -xRR
# fi
iu@bielinux[iu] gedit .zshrc [ 5:01午後 ]
```



W10-3: 環境設定

①パスを通したい/home/iu/Downloads/FastQCを追加して、②Save、③geditを終了。export行の前後に反射的に空行を入れているが、おそらくなくても大丈夫。

```
.zshrc x
#Add custom zsh initialisation here
#System-wide settings can go into /etc/zsh/zshrc.d/customrc.zsh

#These lines enable automatic screen sessions for SSH logins.
#See http://nrc.nrc.ac.uk/tools/bio-linux/other-bl-docs/using-screen
#
# if [[ -n "" ]] && [[ -z "" ]]; then
#   screen -xRR
# fi

export PATH=$PATH:/home/iu/Downloads/FastQC
```

W10-3: 環境設定

①再びtailコマンドを実行し、最後の5行分を表示して確認。

```
iu@bielinux[iu] pwd [ 4:56午後 ]
/home/iu
iu@bielinux[iu] echo $PATH [ 4:56午後 ]
/usr/local/sbin:/usr/local/bin:/usr/sbin:/usr/bin:/sbin:/bin:/usr/games:/usr/local/games:/usr/lib/cd-hit:/usr/lib/cd-hit
iu@bielinux[iu] echo $SHELL [ 4:56午後 ]
/bin/zsh
iu@bielinux[iu] ls -lha .*rc [ 4:56午後 ]
-rw-r--r-- 1 iu iu 3.6K 5月 14 13:56 .bashrc
-rw-r--r-- 1 iu iu 25 11月 10 2014 .dmrc
-rw----- 1 iu iu 296 5月 14 14:14 .zshrc
iu@bielinux[iu] tail -n 5 .zshrc [ 4:56午後 ]
#
# if [[ -n "" ]] && [[ -z "" ]]; then
#   screen -xRR
# fi
iu@bielinux[iu] gedit .zshrc [ 5:01午後 ]
iu@bielinux[iu] tail -n 5 .zshrc [ 5:11午後 ]
#   screen -xRR
# fi
export PATH=$PATH:/home/iu/Downloads/FastQC
iu@bielinux[iu] [ 5:11午後 ]
```

①

screen -xRR
fi
export PATH=\$PATH:/home/iu/Downloads/FastQC

W10-4: sourceコマンド

①.zshrcの中身を書き換えても、②現在開いているターミナル自体は、起動時の.zshrcファイルのパス設定情報のまま。それゆえ③sourceコマンドを用いて、今現在開いているターミナルに対し、さきほど変更した.zshrcファイルの設定情報を読み込ませる作業を行っている。ウェブブラウザ上で再読み込みを行うようなもの。そうすることで、④現在開いているターミナル上でも、更新された.zshrcファイル中の設定情報を利用可能。

```
iu@bielinux[iu] gedit .zshrc
iu@bielinux[iu] tail -n 5 .zshrc
# screen -xRR
# fi
export PATH=$PATH:/home/iu/Downloads/FastQC
iu@bielinux[iu] echo $PATH
/usr/local/sbin:/usr/local/bin:/usr/sbin:/usr/bin:/sbin:/bin:/usr/games:/usr/local/games:/usr/lib/cd-hit:/usr/lib/cd-hit
iu@bielinux[iu] source .zshrc [ 5:17午後]
iu@bielinux[iu] echo $PATH [ 5:17午後]
/usr/local/sbin:/usr/local/bin:/usr/sbin:/usr/bin:/sbin:/bin:/usr/games:/usr/local/games:/usr/lib/cd-hit:/usr/lib/cd-hit:/home/iu/Downloads/FastQC
iu@bielinux[iu] [ 5:17午後]
```

W10-5: 確認

①一旦今のターミナルを閉じて、②もう一度ターミナルを新たに立ち上げる。そして③「echo \$PATH」を実行。確かに、/home/iu/Downloads/FastQCのパスが通っており、一安心。

```
iu@bielinux[iu] gedit .zshrc [ 5:01午後 ]
iu@bielinux[iu] tail -n 5 .zshrc [ 5:11午後 ]
# screen -xRR
# fi

export PATH=$PATH:/home/iu/Downloads/FastQC

iu@bielinux[iu] echo $PATH [ 5:16午後 ]
/usr/local/sbin:/usr/local/bin:/usr/sbin:/usr/bin:/sbin:/bin:/
usr/games:/usr/local/games:/usr/lib/cd-hit:/usr/lib/cd-hit:/ho
me/iu/Downloads/FastQC
iu@bielinux[iu] source .zshrc
iu@bielinux[iu] echo $PATH
/usr/local/sbin:/usr/local/bin:/usr/sbin:/usr/bin:/sbin:/bin:/
usr/games:/usr/local/games:/usr/lib/cd-hit:/usr/lib/cd-hit:/ho
me/iu/Downloads/FastQC
iu@bielinux[iu]
```

```
iu@bielinux[~]
iu@bielinux[iu] echo $PATH [ 5:38午後 ]
/usr/local/sbin:/usr/local/bin:/usr/sbin:/usr/bin:/sbin:/bin:/
usr/games:/usr/local/games:/usr/lib/cd-hit:/usr/lib/cd-hit:/ho
me/iu/Downloads/FastQC
iu@bielinux[iu] [ 5:38午後 ]
```


W10-6: Tips(rcとprofile)

```
iu@bielinux[iu] echo $PATH
/usr/local/sbin:/usr/local/bin:/usr/sbin:/usr/bin:/s
usr/games:/usr/local/games:/usr/lib/cd-hit:/usr/lib/
me/iu/Downloads/FastQC
iu@bielinux[iu] ls -la .*rc .*profile
-rw-r--r-- 1 iu iu 3637  5月 14 13:56 .bashrc
-rw-r--r-- 1 iu iu   25 11月 10  2014 .dmrc
-rw-r--r-- 1 iu iu  676  5月 14 14:19 .profile
-rw----- 1 iu iu  341  5月 14 17:07 .zshrc
iu@bielinux[iu]
```

W10-3では、**zsh**を用いてrcがついた設定ファイルに「`export PATH=...`」という行を追加した。**.zshrc**に追加する以外のやり方として、「**.zsh_profile**」というファイルに追加するのもよい。ここではデフォルトでは**.zsh_profile**は存在しないようだが、ない場合には「`gedit .zsh_profile`」とやって自作してもよい。また、一般には**zsh**の代わりに**bash**が使われるようです。それゆえ、パスの設定に関するウェブ上の情報は**.bashrc**をいじるものや、**.bash_profile**をいじるものがあります。適宜自分のログインシェルで置き換えて解釈しましょう。

W11-1: QC用プログラム

(Rで)塩基配列解析

~NGS, RNA-seq, ゲノム, トランスクリプトーム, 正規化, 発現変動, 統計, モ
(last modified 2015/05/15, since 2010)

What's new

- このウ...
- てフリ...
- 的な利...
- まとめ...
- 参考資...
- 「アセン...

- イントロ | ファイル形式の変換 | [qseq --> FASTA](#) (last modified 2013/06/17)
- イントロ | ファイル形式の変換 | [qseq --> Illumina FASTQ](#) (last modified 2013/06/17)
- イントロ | ファイル形式の変換 | [qseq --> Sanger FASTQ](#) (last modified 2013/06/17)
- **前処理 | クオリティコントロール | について** (last modified 2015/05/04) **NEW**
- 前処理 | クオリティチェック | [qrc](#) (last modified 2014/07/17)
- 前処理 | クオリティチェック | [PHREDスコアに変換](#) (last modified 2013/06/18)
- 前処理 | クオリティチェック | [配列長分布を調べる](#) (last modified 2013/06/18)
- 前処理 | フィルタリング | [PHREDスコアが低い塩基をNに置換](#) (last modified 2013/06/18)
- 前処理 | フィルタリング | [PHREDスコアが低い配列\(リード\)を除去](#) (last modified 2013/06/18)
- 前処理 | フィルタリング | [ACGTのみからなる配列を抽出](#) (last modified 2014/07/17)
- 前処理 | フィルタリング | [ACGT以外の character "-" をNに変換](#) (last modified 2014/07/17)
- 前処理 | フィルタリング | [ACGT以外の文字数が閾値以下の配列を抽出](#) (last modified 2014/07/17)

前処理 | クオリティコントロール | について **NEW**

数億~数十億リードからなるNGSデータの全体的な精度チェック、クオリティの低いリードのフィルタリング、リードに含まれるアダプター配列やクオリティの低い配列部分の除去(トリミング)などを実行する様々な方法をリストアップします。Krakenなどアダプター配列除去などが行えるものも含まれます。FaQCs (Lo and Chain, 2014) など比較的最近のものはpaired-endリードの処理にデフォルトで対応しています。

R用:

- [qrc](#): 原著論文なし
- [PIQA](#): [Martinez-Alcantara et al., Bioinformatics, 2009](#)
- [ShortRead](#): [Morgan et al., Bioinformatics, 2009](#)
- [girafe](#): [Toedling et al., Bioinformatics, 2010](#)
- [QuasR](#): [Gaidatzis et al., Bioinformatics, 2014](#)

R以外:

- [FastQC](#): 原著論文なし
- [FASTX-Toolkit](#): 原著論文なし
- [SolexaQA](#): [Cox et al., BMC Bioinformatics, 2010](#)
- [Quake](#): [Kelley et al., Genome Biol., 2010](#)
- [NGSQC](#): [Dai et al., BMC Genomics, 2010](#)
- [Cutadapt](#): [Martin, M., EMBnet journal, 2011](#)
- [PRINSEQ](#): [Schmieder and Edwards, Bioinformatics, 2011](#)
- [ECHO](#): [Kao et al., Genome Res., 2011](#)
- [Btrim](#): [Kong Y., Genomics, 2011](#)
- [Hammer](#): [Medvedev et al., Bioinformatics, 2011](#)
- [ConDeTri](#): [Smeds et al., PLoS One, 2011](#)
- [BIGpre](#): [Zhang et al., Genomics Proteomics Bioinformatics, 2011](#)
- [NGS QC Toolkit](#): [Patel et al., PLoS One, 2012](#)
- [RobiNA](#): [Lohse et al., Nucleic Acids Res., 2012](#)
- [SEQuel](#): [Ronen et al., Bioinformatics, 2012](#)
- [AdapterRemoval](#): [Lindgreen S., BMC Res Notes, 2012](#)
- [Slim-Filter](#): [Golovko et al., BMC Bioinformatics, 2012](#)
- [HTQC](#): [Yang et al., BMC Bioinformatics, 2013](#)
- [QC-Chain](#): [Zhou et al., PLoS One, 2013](#)
- [Kraken](#): [Davis et al., Methods, 2013](#)
- [AlienTrimmer](#): [Criscuolo and Brisse, Genomics, 2013](#)
- [NextClip](#): [Leggett et al., Bioinformatics, 2014](#)
- [QTrim](#) (Roche/454などの long read用): [Shrestha et al., BMC Bioinformatics, 2014](#)
- [Trimmomatic](#): [Bolger et al., Bioinformatics, 2014](#)
- [Skewer](#): [Jiang et al., BMC Bioinformatics, 2014](#)
- [FaQCs](#): [Lo and Chain, BMC Bioinformatics, 2014](#)

Review:

- [Paszkiwicz et al., Front Genet., 2014](#)

W11-2: FASTX-toolkit

FASTX-toolkitのウェブサイト。
アダプター配列除去を行うツールがあることが分かる。

FASTX-Toolkit

FASTQ/A short-read

Available Tools

- FASTQ-to-FASTA converter
Convert FASTQ files to FASTA files.
- FASTQ Information
Chart Quality Statistics and Nucleotide Distribution
- FASTQ/A Collapser
Collapsing identical sequences in a FASTQ/A file into a single sequence (while maintaining reads counts)
- FASTQ/A Trimmer
Shortening reads in a FASTQ or FASTQ files (removing barcodes or noise).
- FASTQ/A Renamer
Renames the sequence identifiers in FASTQ/A file.
- FASTQ/A Clipper
Removing sequencing adapters / linkers
- FASTQ/A Reverse-Complement
Producing the Reverse-complement of each sequence in a FASTQ/FASTA file.
- FASTQ/A Barcode splitter
Splitting a FASTQ/FASTA files containing multiple samples
- FASTA Formatter
changes the width of sequences line in a FASTA file
- FASTA Nucleotide Changer
Convets FASTA sequences from/to RNA/DNA
- FASTQ Quality Filter
Filters sequences based on quality
- FASTQ Quality Trimmer
Trims (cuts) sequences based on quality
- FASTQ Masker
Masks nucleotides with 'N' (or other character) based on quality

Home | Download & In

Introduction

The FASTX-Toolkit is a

Next-Generation sequen
quality information).

The main processing of
specialized programs. E

However,
It is sometimes more pr
sequences to produce b

The FASTX-Toolkit tool

Available Tools

- FASTQ-to-FASTA

W11-2: FASTX-toolkit

FASTX-Toolkit

FASTQ/A short-reads pre-processing tools



実際に利用するコマンド名がよくわからないという現実的な問題があるが、どうにかしてマニュアルを探せばよい。例えば、①を押して② FASTQ/A Clipperを押すと、コマンド名がfastx_clipperであることなどが分かる。しかし、よく見ると我々が利用しているバージョンはver. 0.0.14でここに記されているのはver. 0.0.6。

[Home](#) | [Download & Installation](#) | [Galaxy Usage](#) | [Command-line Usage](#) | [License](#) | [Useful Links](#) | [Contact](#)

Here you'll find a short description and examples of how to use the FASTX-toolkit from the command line.

- [Command Line Arguments](#)
 - [FASTQ-to-FASTA](#)
 - [FASTQ/A Quality Statistics](#)
 - [FASTQ Quality chart](#)
 - [FASTQ/A Nucleotide Distribution](#)
 - [FASTQ/A Clipper](#)
 - [FASTQ/A Rename](#)
 - [FASTQ/A Trimmer](#)
 - [FASTQ/A Collapser](#)
 - [FASTQ/A Artifacts Filter](#)
 - [FASTQ Quality Filter](#)
 - [FASTQ/A Reverse Complement](#)
 - [FASTA Formatter](#)
 - [FASTA nucleotides changer](#)
 - [FASTA Clipping Histogram](#)
 - [FASTX Barcode Splitter](#)
- [Example: FASTQ Information](#)
- [Example: FASTQ/A manipulation](#)

FASTA/Q Clipper

```
$ fastx_clipper -h
usage: fastx_clipper [-n] [-a ADAPTER] [-D] [-l N] [-n] [-d N] [-c] [-C] [-o] [-v] [-z] [-i INFILE] [-o OUTFILE]

version 0.0.6
[-h] = This helpful help screen.
[-a ADAPTER] = ADAPTER string. default is CCTTAAGG (dummy adapter).
[-l N] = discard sequences shorter than N nucleotides. default is 5.
[-d N] = Keep the adapter and N bases after it.
         (using '-d 0' is the same as not using '-d' at all. which is the default).
[-c] = Discard non-clipped sequences (i.e. - keep only sequences which contained the adapter).
[-C] = Discard clipped sequences (i.e. - keep only sequences which did not contained the adapter).
[-k] = Report Adapter-Only sequences.
[-n] = keep sequences with unknown (N) nucleotides. default is to discard such sequences.
[-v] = Verbose - report number of sequences.
         If [-o] is specified, report will be printed to STDOUT.
         If [-o] is not specified (and output goes to STDOUT),
         report will be printed to STDERR.
[-z] = Compress output with GZIP.
[-D] = DEBUG output.
[-i INFILE] = FASTA/Q input file. default is STDIN.
[-o OUTFILE] = FASTA/Q output file. default is STDOUT.
```

W11-2: FASTX-toolkit

赤枠内をクリックして目的のコマンド名を探してもよい。他の手段として、「FASTQ/A Clipperがfastx_clipper」を頼りに、Bio-Linux上でfastxまで打ってからTabキーを押す。

FASTX-Toolkit

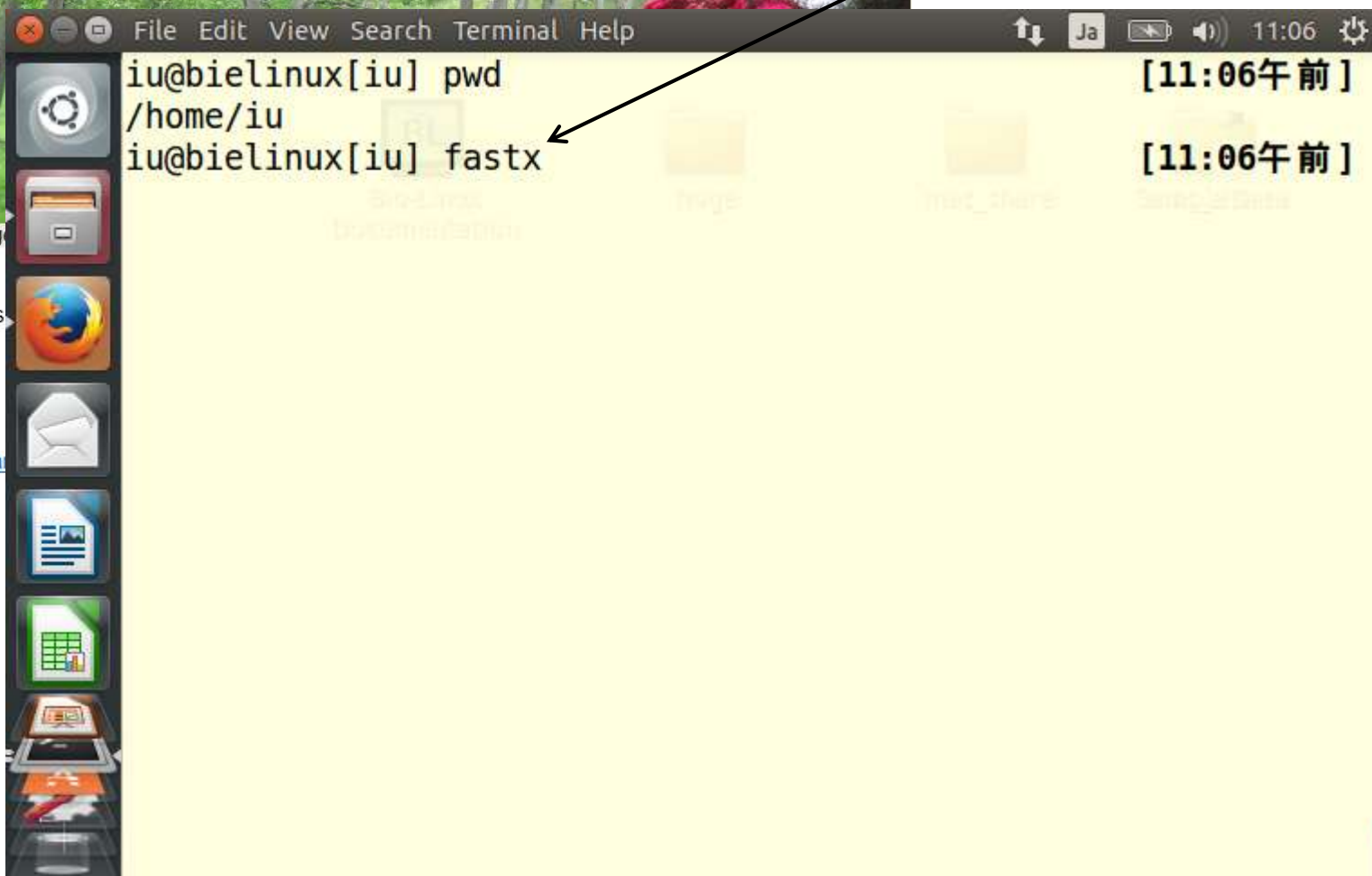
FASTQ/A short-reads pre-processing tools

• [Command Line Arguments](#)

- [FASTQ-to-FASTA](#)
- [FASTQ/A Quality Statistics](#)
- [FASTQ Quality chart](#)
- [FASTQ/A Nucleotide Distribution chart](#)
- [FASTQ/A Clipper](#)
- [FASTQ/A Renamer](#)
- [FASTQ/A Trimmer](#)
- [FASTQ/A Collapser](#)
- [FASTQ/A Artifacts Filter](#)
- [FASTQ Quality Filter](#)
- [FASTQ/A Reverse Complement](#)
- [FASTA Formatter](#)
- [FASTA nucleotides changer](#)
- [FASTA Clipping Histogram](#)
- [FASTX Barcode Splitter](#)

• [Example: FASTQ Information](#)

• [Example: FASTQ/A manipulation](#)

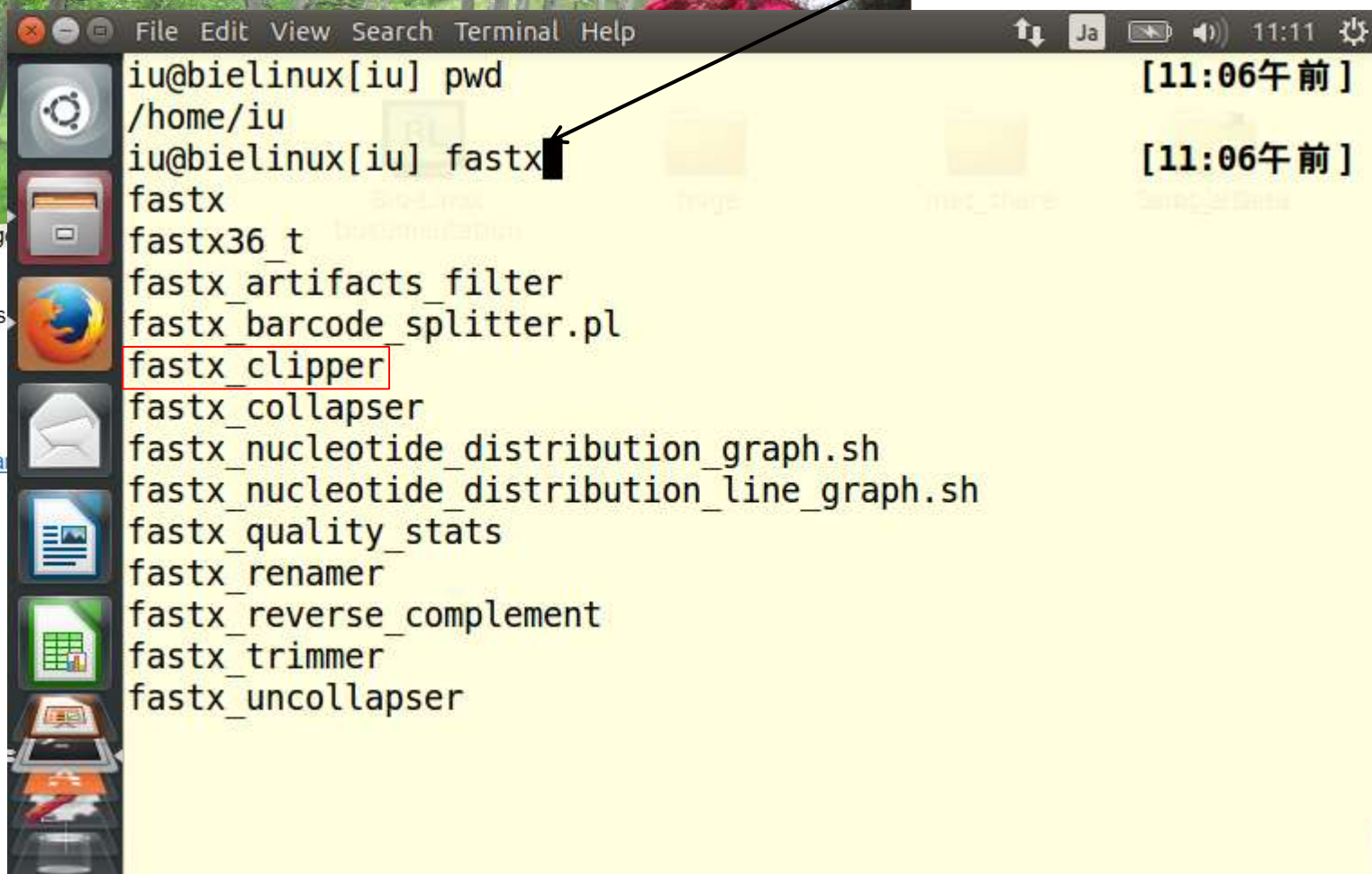


W11-2: FASTX-toolkit

そうすると、fastxから始まるパスが通っているコマンド候補が表示される。fastx_clipperを含む、FASTX-toolkitで提供されているコマンド群っぽいものを概観できる。

FASTX-Toolkit

FASTQ/A short-reads pre-processing tools



```
iu@bielinux[iu] pwd
/home/iu
iu@bielinux[iu] fastx
fastx
fastx36_t
fastx_artifacts_filter
fastx_barcode_splitter.pl
fastx_clipper
fastx_collapser
fastx_nucleotide_distribution_graph.sh
fastx_nucleotide_distribution_line_graph.sh
fastx_quality_stats
fastx_renamer
fastx_reverse_complement
fastx_trimmer
fastx_uncollapser
```

• [Command Line Arguments](#)

- [FASTQ-to-FASTA](#)
- [FASTQ/A Quality Statistics](#)
- [FASTQ Quality chart](#)
- [FASTQ/A Nucleotide Distribution chart](#)
- [FASTQ/A Clipper](#)
- [FASTQ/A Renamer](#)
- [FASTQ/A Trimmer](#)
- [FASTQ/A Collapser](#)
- [FASTQ/A Artifacts Filter](#)
- [FASTQ Quality Filter](#)
- [FASTQ/A Reverse Complement](#)
- [FASTA Formatter](#)
- [FASTA nucleotides changer](#)
- [FASTA Clipping Histogram](#)
- [FASTX Barcode Splitter](#)

• [Example: FASTQ Information](#)

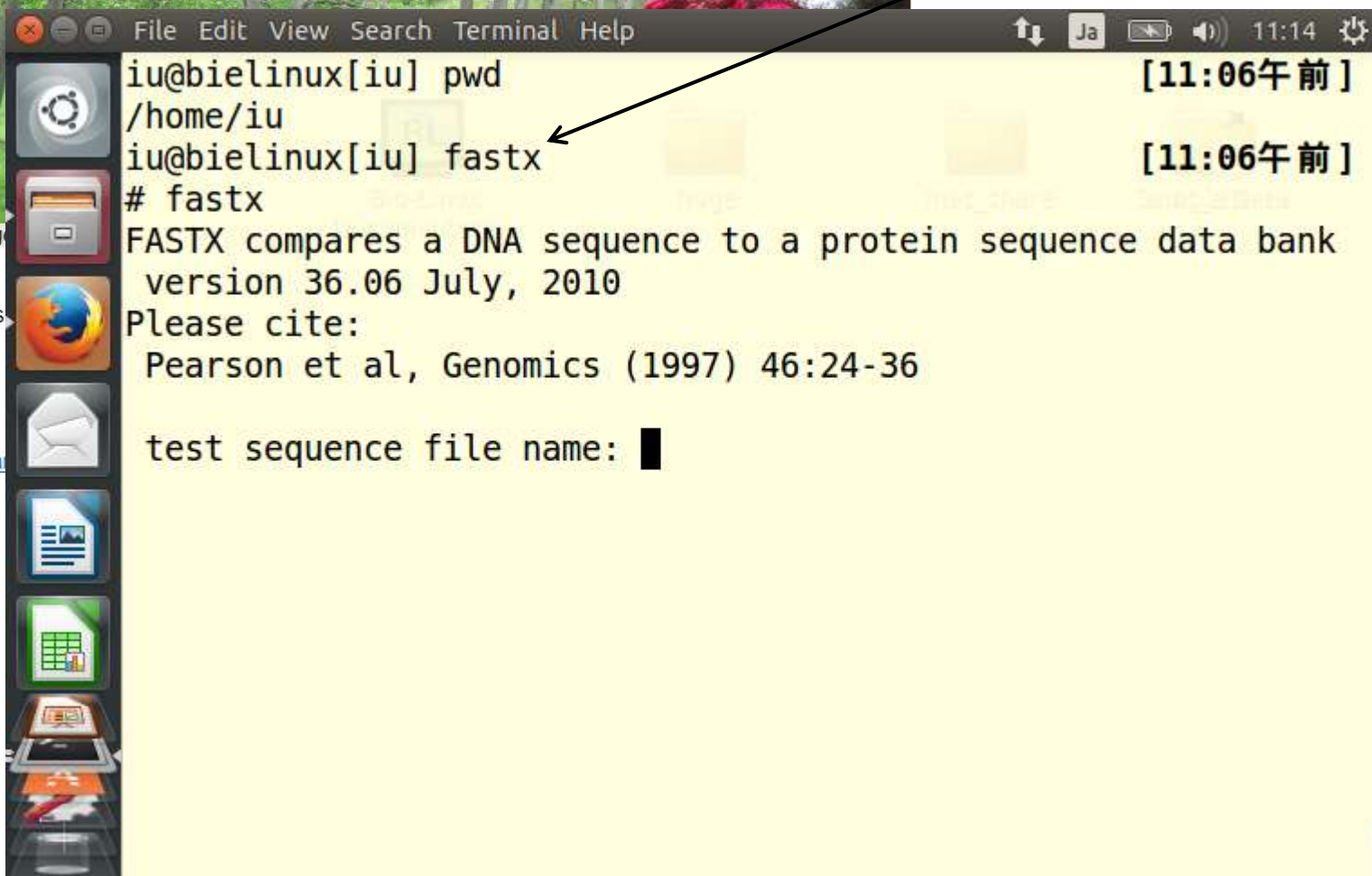
• [Example: FASTQ/A manipulation](#)

W11-3: Tips

間違っってリターンキーを押すとfastxというコマンドが実行される。使い方もよくわからないので、とりあえずこの状況から抜けるべく、「CTRL + C」。「CTRL + C」で抜けられない場合は、多用しないようがよいが、「CTRL + D」や「CTRL + Z」もある。

FASTX-Toolkit

FASTQ/A short-reads pre-processing tools



```
iu@bielinux[iu] pwd
/home/iu
iu@bielinux[iu] fastx
# fastx
FASTX compares a DNA sequence to a protein sequence data bank
version 36.06 July, 2010
Please cite:
Pearson et al, Genomics (1997) 46:24-36

test sequence file name: █
```

• [Command Line Arguments](#)

- [FASTQ-to-FASTA](#)
 - [FASTQ/A Quality Statistics](#)
 - [FASTQ Quality chart](#)
 - [FASTQ/A Nucleotide Distribution chart](#)
 - [FASTQ/A Clipper](#)
 - [FASTQ/A Renamer](#)
 - [FASTQ/A Trimmer](#)
 - [FASTQ/A Collapser](#)
 - [FASTQ/A Artifacts Filter](#)
 - [FASTQ Quality Filter](#)
 - [FASTQ/A Reverse Complement](#)
 - [FASTA Formatter](#)
 - [FASTA nucleotides changer](#)
 - [FASTA Clipping Histogram](#)
 - [FASTX Barcode Splitter](#)
- [Example: FASTQ Information](#)
- [Example: FASTQ/A manipulation](#)

W11-3: Tips

無事復帰。画面をリフレッシュしたい場合は「CTRL + L」。「clear」というコマンドでもいいらしい。

FASTX-Toolkit

FASTQ/A short-reads pre-processing tools

```
iu@bielinux[iu] pwd [11:06午前]
/home/iu
iu@bielinux[iu] fastx [11:06午前]
# fastx
FASTX compares a DNA sequence to a protein sequence data bank
version 36.06 July, 2010
Please cite:
Pearson et al, Genomics (1997) 46:24-36
test sequence file name: ^C
iu@bielinux[iu] [11:17午前]
```

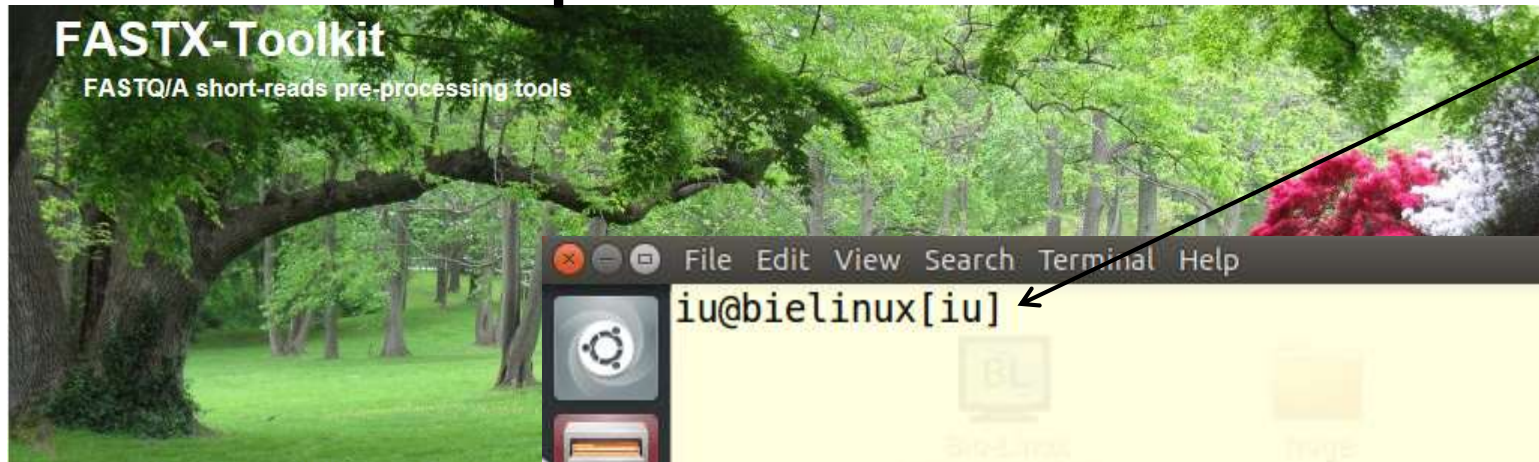
• [Command Line Arguments](#)

- [FASTQ-to-FASTA](#)
 - [FASTQ/A Quality Statistics](#)
 - [FASTQ Quality chart](#)
 - [FASTQ/A Nucleotide Distribution chart](#)
 - [FASTQ/A Clipper](#)
 - [FASTQ/A Renamer](#)
 - [FASTQ/A Trimmer](#)
 - [FASTQ/A Collapser](#)
 - [FASTQ/A Artifacts Filter](#)
 - [FASTQ Quality Filter](#)
 - [FASTQ/A Reverse Complement](#)
 - [FASTA Formatter](#)
 - [FASTA nucleotides changer](#)
 - [FASTA Clipping Histogram](#)
 - [FASTX Barcode Splitter](#)
- [Example: FASTQ Information](#)
- [Example: FASTQ/A manipulation](#)

W11-3: Tips

FASTX-Toolkit

FASTQ/A short-reads pre-processing tools



iu@bielinux[iu]

[11:17午前]

Home | Download & Installation | Galaxy Usage

Here you'll find a short description and examples

- [Command Line Arguments](#)
 - [FASTQ-to-FASTA](#)
 - [FASTQ/A Quality Statistics](#)
 - [FASTQ Quality chart](#)
 - [FASTQ/A Nucleotide Distribution chart](#)
 - [FASTQ/A Clipper](#)
 - [FASTQ/A Renamer](#)
 - [FASTQ/A Trimmer](#)
 - [FASTQ/A Collapser](#)
 - [FASTQ/A Artifacts Filter](#)
 - [FASTQ Quality Filter](#)
 - [FASTQ/A Reverse Complement](#)
 - [FASTA Formatter](#)
 - [FASTA nucleotides changer](#)
 - [FASTA Clipping Histogram](#)
 - [FASTX Barcode Splitter](#)
- [Example: FASTQ Information](#)
- [Example: FASTQ/A manipulation](#)

W11-4: FastQC Report

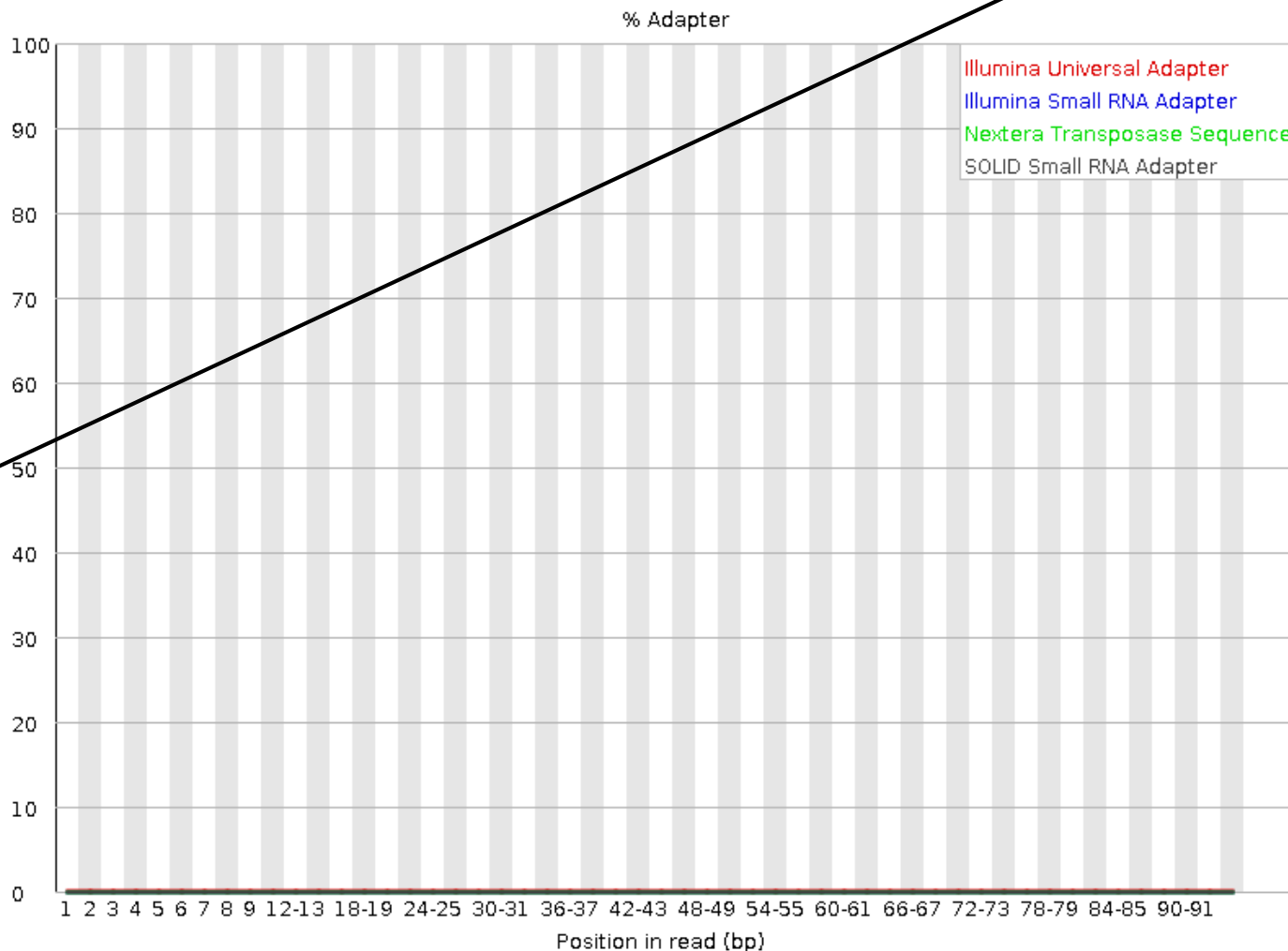
SRR616268sub_1.fastqのFastQC (ver 0.11.1)実行結果。Overrepresented sequencesで多少アダプター配列が含まれていることはわかるが、リードに占めるアダプター配列の割合はほぼ0であることがわかる。

FastQC Report

Summary

- ✓ Basic Statistics
- ✓ Per base sequence quality
- ✓ Per tile sequence quality
- ✓ Per sequence quality scores
- ✗ Per base sequence content
- ! Per sequence GC content
- ✓ Per base N content
- ✓ Sequence Length Distribution
- ✗ Sequence Duplication Levels
- ✗ Overrepresented sequences
- ✓ Adapter Content **①**
- ✗ Kmer Content

✓ Adapter Content



W11-5: FASTX-toolkit

バージョンを確認。Bio-Linux 8にプレインストールされているのは、ver. 0.0.14であることがわかる。赤下線を眺めてオプションのイメージを掴む。

```
iu@bielinux[iu] fastx_clipper -h
usage: fastx_clipper [-h] [-a ADAPTER] [-D] [-l N] [-n] [-d N] [-c] [-C] [-o] [-v] [-z] [-i
INFILE] [-o OUTFILE]
Part of FASTX Toolkit 0.0.14 by A. Gordon (assafgordon@gmail.com)

[-h]           = This helpful help screen.
[-a ADAPTER]  = ADAPTER string. default is CCTTAAGG (dummy adapter).
[-l N]        = discard sequences shorter than N nucleotides. default is 5.
[-d N]        = Keep the adapter and N bases after it.
                (using '-d 0' is the same as not using '-d' at all. which is the default).
[-c]          = Discard non-clipped sequences (i.e. - keep only sequences which contained
the adapter).
[-C]          = Discard clipped sequences (i.e. - keep only sequences which did not contain
the adapter).
[-k]          = Report Adapter-Only sequences.
[-n]          = keep sequences with unknown (N) nucleotides. default is to discard such se
quences.
[-v]          = Verbose - report number of sequences.
                If [-o] is specified, report will be printed to STDOUT.
                If [-o] is not specified (and output goes to STDOUT),
                report will be printed to STDERR.
[-z]          = Compress output with GZIP.
[-D]          = DEBUG output.
[-M N]        = require minimum adapter alignment length of N.
                If less than N nucleotides aligned with the adapter - don't clip it. [-i
INFILE] = FASTA/Q input file. default is STDIN.
[-o OUTFILE] = FASTA/Q output file. default is STDOUT.
```

W11-5: FASTX-toolkit

FASTX-Toolkit

FASTQ/A short-reads pre-processing tools



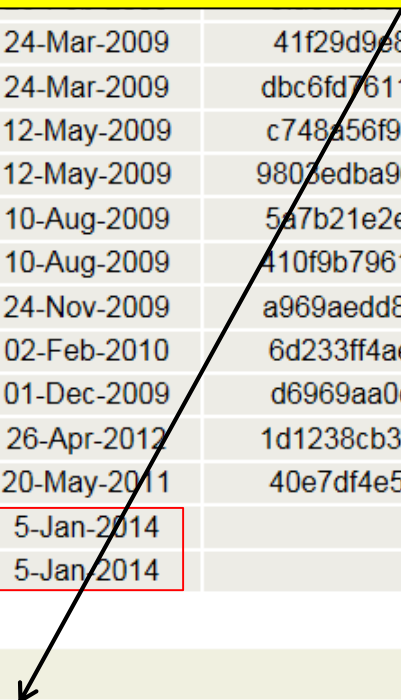
Download

File	Version		
fastx_toolkit-0.0.6.tar.bz2	0.0.6		
fastx_toolkit-0.0.7.tar.bz2	0.0.7	24-Mar-2009	41f29d9e848fc5d8b06e3a11f4fe0006
libgtextutils-0.1.tar.gz	0.1	24-Mar-2009	dbc6fd761103bb17d512bed796288de9
fastx_toolkit-0.0.8.tar.bz2	0.0.8	12-May-2009	c748a56f906cf7436ea6ee20c024ebdd
libgtextutils-0.3.tar.bz2	0.3	12-May-2009	9802edba96036a32b3b7256d4869d38b
fastx_toolkit-0.0.10.tar.bz2	0.0.10	10-Aug-2009	5a7b21e2eb467a28f6b14aabc7223f61
libgtextutils-0.5.tar.bz2	0.5	10-Aug-2009	410f9b79618799ee146587a2b71ab7b0
fastx_toolkit-0.0.11.tar.bz2	0.0.11	24-Nov-2009	a969aedd8050f509cb836209ac9cd391
fastx_toolkit-0.0.13.tar.bz2	0.0.13	02-Feb-2010	6d233ff4ae3d52c457d447179f073a56
libgtextutils-0.6.tar.bz2	0.6	01-Dec-2009	d6969aa0d31cc934e1fedf3fe3d0dc63
fastx_toolkit-0.0.13.2.tar.bz2	0.0.13.2	26-Apr-2012	1d1238cb3029cb1c5d1a3eea7d7d2ca8
libgtextutils-0.6.1.tar.bz2	0.6.1	20-May-2011	40e7df4e5a72efe50aa789af8caeb935
fastx_toolkit-0.0.14.tar.bz2	0.0.14	5-Jan-2014	
libgtextutils-0.7.tar.gz	0.7	5-Jan-2014	

Requirements

- Fastx-toolkit version 0.0.13 requires **libgtextutils-0.6** (available here for download)
- A recent g++ compiler (tested with GNU G++ 4.1.2 and later).
- The **fasta_clipping_histogram** tool requires two perl modules: **PerlIO::gzip** and **GD::Graph::bars**.
- The **fastx_barcode_splitter** tool requires **GNU sed**.
- The **fastq_quality_boxplot** tool requires **gnuplot** version 4.2 or newer.

バージョンを確認。2015年5月21日現在の最新版はver. 0.0.14であることがわかる。Requirementの記述内容のバージョンが古いが、ver. 0.0.14の場合はlibgtextutils-0.7も必要であることがわかる。他にもいろいろと必要らしいが、Bio-Linux8にプレインストールされているバージョンが最新なのでここでは気にしない。



W11-6: fastx_clipper

```
iu@bielinux[srp017156] pwd [ 8:31午後 ]
/home/iu/Documents/srp017156
iu@bielinux[srp017156] wget -cq http://www.iu.a.u-tokyo.ac.jp/~kadota/book/J
SLAB4_3_Linux.sh
iu@bielinux[srp017156] head -n 11 JSLAB4_3_Linux.sh [ 8:33午後 ]
### fastx clipper with default options ###
time gunzip SRR616268sub_1.fastq.gz
wc SRR616268sub_1.fastq
time fastx_clipper \
-a GATCGGAAGAGCACACGTCTGAACTCCAGTCACTTAGGCATCTCGTATGC \
-i SRR616268sub_1.fastq \
-o hogel.fastq
fastqc2 -q hogel.fastq \
--outdir=/home/iu/Desktop/mac_share
wc hogel.fastq

iu@bielinux[srp017156] [ 8:34午後 ]
```

W11-6: fastx_clipper

連載第4回の図4とほぼ同じ。赤矢印部分までで、fastx_clipper実行のための非圧縮FASTQファイル(SRR616268sub_1.fastq)を作成し、行数が400万行であることを確認している。

```
iu@bielinux[~/Documents/srp017156]
iu@bielinux[srp017156] pwd
/home/iu/Documents/srp017156
iu@bielinux[srp017156] ls -lh *sub* [ 7:51午後 ]
-rw-rw-r-- 1 iu iu 72M 5月 6 10:07 SRR616268sub_1.fastq.gz
-rw-rw-r-- 1 iu iu 65M 5月 6 10:08 SRR616268sub_2.fastq.gz
iu@bielinux[srp017156] time gunzip SRR616268sub_1.fastq.gz [ 7:51午後 ]
gunzip SRR616268sub_1.fastq.gz 2.24s user 0.44s system 56% cpu 4.698 total
iu@bielinux[srp017156] wc SRR616268sub_1.fastq [ 7:51午後 ]
4000000 8000000 320760716 SRR616268sub_1.fastq
iu@bielinux[srp017156] time fastx_clipper \ [ 7:52午後 ]
> -a GATCGGAAGAGCACACGTCTGAACTCCAGTCACTTAGGCATCTCGTATGC \
> -i SRR616268sub_1.fastq \
> -o hoge1.fastq
fastx_clipper -a GATCGGAAGAGCACACGTCTGAACTCCAGTCACTTAGGCATCTCGTATGC -i -o
126.96s user 9.97s system 85% cpu 2:39.97 total
iu@bielinux[srp017156] fastqc2 -q hoge1.fastq \ [ 7:54午後 ]
> --outdir=/home/iu/Desktop/mac_share
iu@bielinux[srp017156] wc hoge1.fastq [ 7:55午後 ]
3915596 7831192 286011042 hoge1.fastq
iu@bielinux[srp017156] [ 7:55午後 ]
```

W11-6: fastx_clipper

赤枠部分がfastx_clipper実行コマンド。長いコマンドなので、視認性向上のため「¥(逆スラッシュ)」を用いて明示的に複数行で記述している[W9-8]。-aでアダプター配列(ここではTruSeq Adapter Index 3)を指定、-iは入力ファイル(SRR616268sub_1.fastq)、-oは出力ファイル名(hoge1.fastq)を指定する。fastx_clipperの前にtimeコマンドをつけて実行時間も計測している。赤矢印で示すように2分強かかる。

```
iu@bielinux[~/Documents/srp017156]
iu@bielinux[srp017156] pwd
/home/iu/Documents/srp017156
iu@bielinux[srp017156] ls -lh *sub*
-rw-rw-r-- 1 iu iu 72M  5月  6 10:07 SRR616268sub_1.fastq
-rw-rw-r-- 1 iu iu 65M  5月  6 10:08 SRR616268sub_2.fastq
iu@bielinux[srp017156] time gunzip SRR616268sub_1.fastq.gz
gunzip SRR616268sub_1.fastq.gz  2.24s user 0.44s system 85% cpu 2:39.97 total
iu@bielinux[srp017156] wc SRR616268sub_1.fastq
 4000000   8000000 320760716 SRR616268sub_1.fastq
iu@bielinux[srp017156] time fastx_clipper \
> -a GATCGGAAGAGCACACGTCTGAACTCCAGTCACTTAGGCATCTCGTATGC \
> -i SRR616268sub_1.fastq \
> -o hoge1.fastq
fastx_clipper -a GATCGGAAGAGCACACGTCTGAACTCCAGTCACTTAGGCATCTCGTATGC -i -o
126.96s user 9.97s system 85% cpu 2:39.97 total
iu@bielinux[srp017156] fastqc2 -q hoge1.fastq \
> --outdir=/home/iu/Desktop/mac_share
iu@bielinux[srp017156] wc hoge1.fastq
 3915596   7831192 286011042 hoge1.fastq
iu@bielinux[srp017156]
```

W11-6: fastx_clipper

hoge1.fastqを入力としてFastQC (ver. 0.11.3)を実行[W9-5; W9-6]。結果は共有フォルダである/home/iu/Desktop/mac_shareに出力するように指定[W9-7]。共有フォルダにhoge1_fastqc.htmlが作成される。

```
iu@bielinux[~/Documents/srp017156]
iu@bielinux[srp017156] pwd
/home/iu/Documents/srp017156
iu@bielinux[srp017156] ls -lh *sub* [ 7:51午後 ]
-rw-rw-r-- 1 iu iu 72M 5月 6 10:07 SRR616268sub_1.fastq.gz
-rw-rw-r-- 1 iu iu 65M 5月 6 10:08 SRR616268sub_2.fastq.gz
iu@bielinux[srp017156] time gunzip SRR616268sub_1.fastq.gz [ 7:51午後 ]
gunzip SRR616268sub_1.fastq.gz 2.24s user 0.44s system 56% cpu 4.698 total
iu@bielinux[srp017156] wc SRR616268sub_1.fastq [ 7:51午後 ]
4000000 8000000 320760716 SRR616268sub_1.fastq
iu@bielinux[srp017156] time fastx_clipper \ [ 7:52午後 ]
> -a GATCGGAAGAGCACACGTCTGAACTCCAGTCACTTAGGCATCTCGTATGC \
> -i SRR616268sub_1.fastq \
> -o hoge1.fastq
fastx_clipper -a GATCGGAAGAGCACACGTCTGAACTCCAGTCACTTAGGCATCTCGTATGC -i -o
126.96s user 9.97s system 85% cpu 2:39.97 total
iu@bielinux[srp017156] fastqc2 -q hoge1.fastq \ [ 7:54午後 ]
> --outdir=/home/iu/Desktop/mac_share
iu@bielinux[srp017156] wc hoge1.fastq [ 7:55午後 ]
3915596 7831192 286011042 hoge1.fastq
iu@bielinux[srp017156] [ 7:55午後 ]
```


W11-6: fastx_clipper

hoge1.fastqの行数を出力。入力は107 bpの長さの100万リード(FASTQは1リードを4行で表すので、400万行)。これに50塩基の長さのアダプター配列を除去すべく実行すると、なぜか出力ファイルが3,915,596行となっていて変。

```
iu@bielinux[~/Documents/srp017156]
iu@bielinux[srp017156] pwd
/home/iu/Documents/srp017156
iu@bielinux[srp017156] ls -lh *sub*
-rw-rw-r-- 1 iu iu 72M  5月  6 10:07 SRR616268sub_1.fastq.gz
-rw-rw-r-- 1 iu iu 65M  5月  6 10:08 SRR616268sub_2.fastq.gz
iu@bielinux[srp017156] time gunzip SRR616268sub_1.fastq.gz [ 7:51午後 ]
gunzip SRR616268sub_1.fastq.gz 2.24s user 0.44s system 56% cpu 4.698 total
iu@bielinux[srp017156] wc SRR616268sub_1.fastq [ 7:51午後 ]
4000000 ← 8000000 320760716 SRR616268sub_1.fastq
iu@bielinux[srp017156] time fastx_clipper [ 7:52午後 ]
> -a GATCGGAAGAGCACACGTCTGAACTCCAGTCACTTAGGCATCTCGTATGC \
> -i SRR616268sub_1.fastq \
> -o hoge1.fastq
fastx_clipper -a GATCGGAAGAGCACACGTCTGAACTCCAGTCACTTAGGCATCTCGTATGC -i -o
126.96s user 9.97s system 85% cpu 2:39.97 total
iu@bielinux[srp017156] fastqc2 -q hoge1.fastq [ 7:54午後 ]
> --outdir=/home/iu/Desktop/mac_share
iu@bielinux[srp017156] wc hoge1.fastq [ 7:55午後 ]
3915596 ← 7831192 286011042 hoge1.fastq
iu@bielinux[srp017156] [ 7:55午後 ]
```

W11-7: FastQC

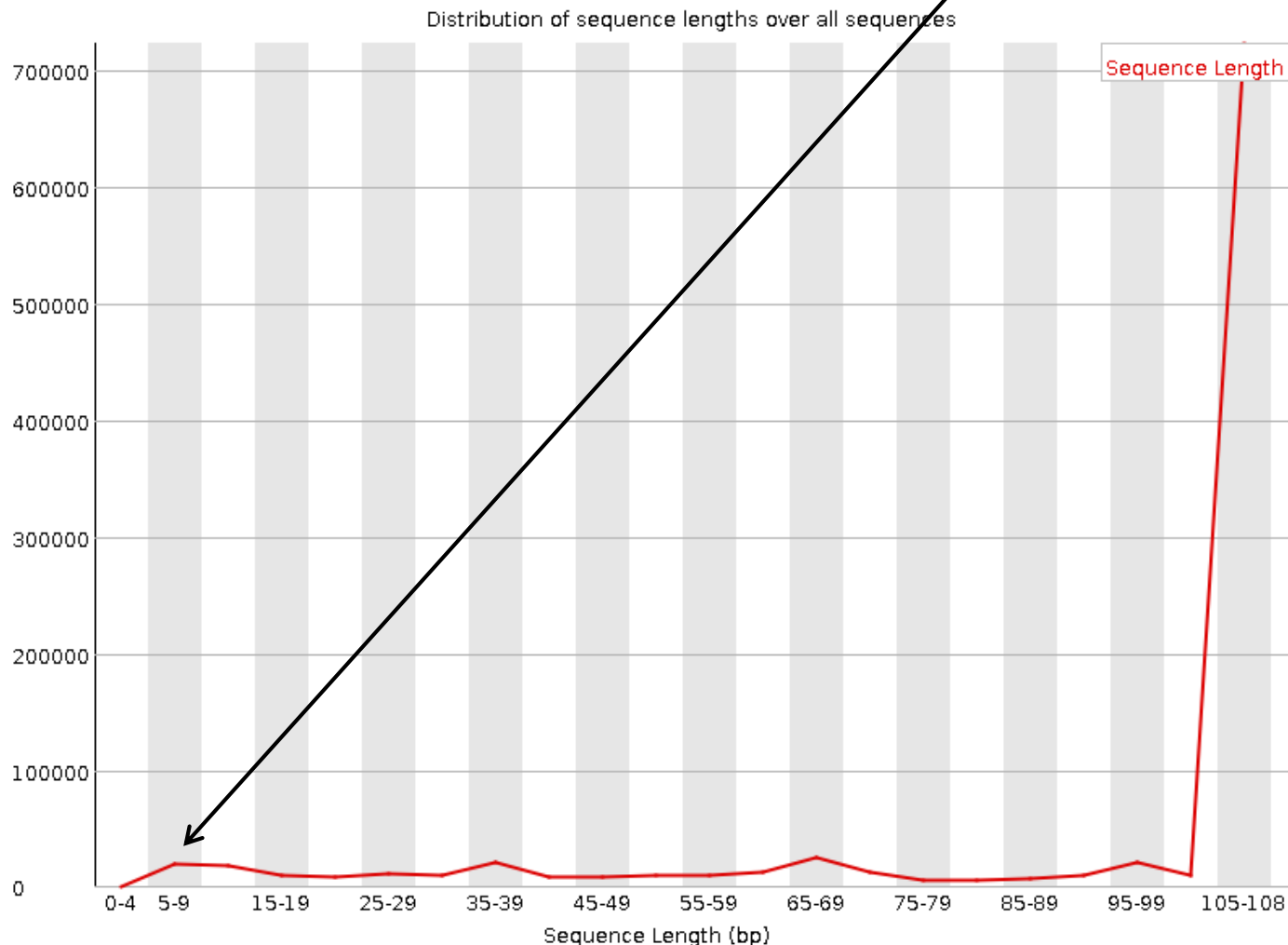
hoge1.fastqを入力としてFastQC (ver. 0.11.3)を実行した結果。配列長分布の項目を表示。50 bpのアダプター配列除去後で107 bpから5 bpまでトリムされるリードがあるのでやはり変

FastQC Report

Summary

- ✓ Basic Statistics
- ✓ Per base sequence quality
- ✓ Per tile sequence quality
- ✓ Per sequence quality scores
- ✗ Per base sequence content
- ✗ Per sequence GC content
- ✓ Per base N content
- ! Sequence Length Distribution
- ✗ Sequence Duplication Levels
- ✗ Overrepresented sequences
- ✓ Adapter Content
- ✗ Kmer Content

! Sequence Length Distribution



W11-8-1: 動作確認

-cオプション(アダプターを含むリードのみ出力)をつけて実行した結果[W11-5]。テンプレートコードはJSLAB4_3_Linux.shにあります。指定したアダプター配列を含むリードは $1,021,860 / 4 = 255,465$ 個ということになる。

```
iu@bielinux[~/Documents/srp017156]
iu@bielinux[srp017156] pwd
/home/iu/Documents/srp017156
iu@bielinux[srp017156] time fastx_clipper \ [11:34午前]
> -a GATCGGAAGAGCACACGTCTGAACTCCAGTCACTTAGGCATCTCGTATGC \
> -i SRR616268sub_1.fastq \
> -c \ ←
> -o hoge2.fastq
fastx_clipper -a GATCGGAAGAGCACACGTCTGAACTCCAGTCACTTAGGCATCTCGTATGC -i
-c -o 119.52s user 9.16s system 91% cpu 2:21.21 total
iu@bielinux[srp017156] wc hoge2.fastq [11:36午前]
1021860 ← 2043720 53961976 hoge2.fastq
iu@bielinux[srp017156] █ [11:36午前]
```

W11-8-2: 動作確認

-Cオプション(アダプターを含まないリードのみ出力)をつけて実行した結果[W11-5]。テンプレートコードはJSLAB4_3_Linux.shにあります。指定したアダプター配列を含まないリードは $2,893,736 / 4 = 723,434$ 個ということになる。

```
iu@bielinux[~/Documents/srp017156]
iu@bielinux[srp017156] pwd
/home/iu/Documents/srp017156
iu@bielinux[srp017156] time fastx_clipper \                               [11:49午前]
> -a GATCGGAAGAGCACACGTCTGAACTCCAGTCACTTAGGCATCTCGTATGC \
> -i SRR616268sub_1.fastq \
> -C \ ←
> -o hoge3.fastq
fastx_clipper -a GATCGGAAGAGCACACGTCTGAACTCCAGTCACTTAGGCATCTCGTATGC -i
-C -o 123.52s user 9.57s system 90% cpu 2:26.64 total
iu@bielinux[srp017156] wc hoge3.fastq                                     [11:51午前]
2893736 ← 5787472 232049066 hoge3.fastq
iu@bielinux[srp017156] █                                               [11:51午前]
```

W11-8-3: 動作確認

W11-5を眺めると、-kオプションがアダプター配列のみからなるリードを出力するものらしいので実行してみる。テンプレートコードはJSLAB4_3_Linux.shにあります。指定したアダプター配列のみからなるリードは34,052 / 4 = 8,513個ということになる。

```
iu@bielinux[~/Documents/srp017156]
iu@bielinux[srp017156] pwd
/home/iu/Documents/srp017156
iu@bielinux[srp017156] time fastx_clipper \
> -a GATCGGAAGAGCACACGTCTGAACTCCAGTCACTTAGGCATCTCGTATGC \
> -i SRR616268sub 1.fastq \
> -k \
> -o hoge4.fastq
fastx_clipper -a GATCGGAAGAGCACACGTCTGAACTCCAGTCACTTAGGCATCTCGTATGC -i
-k -o 117.28s user 8.46s system 91% cpu 2:17.65 total
iu@bielinux[srp017156] wc hoge4.fastq
34052 ← 68104 2730404 hoge4.fastq
iu@bielinux[srp017156] █
```

[11:57午前]

[11:57午前]

W11-8-4: 動作確認

```
iu@bielinux[~/Documents/srp017156]
iu@bielinux[srp017156] pwd
/home/iu/Documents/srp017156
iu@bielinux[srp017156] time fastx_clipper \
> -a GATCGGAAGAGCACACGTCTGAACTCCAGTCACTTAGG \
> -i SRR616268sub_1.fastq \
> -n \ ←
> -o hoge5.fastq
fastx_clipper -a GATCGGAAGAGCACACGTCTGAACTCCAGTCACTTAGG -i SRR616268sub_1.fastq -n -o hoge5.fastq
-n -o 125.16s user 9.76s system 90% cpu 2m15s
iu@bielinux[srp017156] wc hoge5.fastq
3927924 ← 7855848 286951162 hoge5.fastq
iu@bielinux[srp017156] █
```

W11-5を眺めると、-nオプションをつけるとW11-6のデフォルトオプションで得られた4,000,000行から3,915,596行になった理由がつかめるのではと思いやってみる。テンプレートコードはJSLAB4_3_Linux.shにあります。実行結果は3,927,924行となっており、3,915,596行から増えている。このことからfastx_clipperはデフォルトでNを含むリードを除去すること、そしてそれも残したい場合は-nをつけなければいけないことを学習する。ただし、マニュアル中の記述だけでは、107 bp全てがNの場合のみのリードが該当するのかなど、Nをどの程度含むのかについての言及はないので気持ち悪い。尚、-nに該当するリードは $(3,927,924 - 3,915,596) / 4 = 3,082$ 個と計算する。

W11-8-5: 動作確認

```
iu@bielinux[~/Documents/srp017156]
iu@bielinux[srp017156] pwd
/home/iu/Documents/srp017156
iu@bielinux[srp017156] time fastx_clipper \
> -a GATCGGAAGAGCACACGTCTGAACTCCAGTCACTTAGC
> -i SRR616268sub_1.fastq \
> -n -l 0 \ ←
> -o hoge6.fastq
fastx_clipper -a GATCGGAAGAGCACACGTCTGAACTCC
-n -l 124.69s user 9.84s system 86% cpu 2
iu@bielinux[srp017156] wc hoge6.fastq
3965948 ← 7931896 288015124 hoge6.fastq
iu@bielinux[srp017156] █
```

W11-7の配列長分布から、50 bpのアダプター配列除去後で5 bpになってしまっているリードが存在すること、そしてW11-5から、-lオプションをつけるとトリム後に5 bp以下になってしまっているリードも出力させられそうだと判断し、赤下線部分の-l 0を追加して実行。3,965,948行になっており、-nのみの3,927,924行に比べて $(3,965,948 - 3,927,924) / 4 = 9,506$ リード分だけ増えていることがわかる。つまり、9,506リードがアダプター配列除去後に5 bp未満になっていたと判断できる。しかし、これでもまだ出力結果が4,000,000行にならないのは気持ち悪いので、自分の感性と合う別のプログラムを探すという結論を下す。W11-8-3の-kオプションで得られた行数分を足すと400万行になることが後にわかった。

W11-8-6: 動作確認

その後、念のためW11-5から、-Mオプションをつけるとアダプター配列と一致する文字数を与えられることを知る。ここでは50 bpとしてアダプター配列と完全一致するリードのみ-cオプションをつけて出力している。全部で548リードが該当するとなっているが、W8-6の結果と明らかに矛盾する。なぜならW8-6の結果が正しいと仮定すると、最低でも2,415リードはこのアダプター配列と完全一致する領域をもつはずだからである。

```
iu@bielinux[~/Documents/srp017156]
iu@bielinux[srp017156] pwd
/home/iu/Documents/srp017156
iu@bielinux[srp017156] time fastx clipper \
> -a GATCGGAAGAGCACACGTCTGAACTCCAGTCACTTAGC
> -i SRR616268sub_1.fastq \
> -c -M 50 \
> -o hoge7.fastq
fastx_clipper -a GATCGGAAGAGCACACGTCTGAACTCC
-c -M 118 29s user 8.59s system 91% cpu 2:18.61 total
iu@bielinux[srp017156] wc hoge7.fastq
548 1096 22572 hoge7.fastq
iu@bielinux[srp017156] head -n 8 hoge7.fastq
@SRR616268.2681 2291:6:1101:11058:2427 length=107
GTCAACTGATAGTTAAGGAAGCGCGAACA
+SRR616268.2681 2291:6:1101:11058:2427 length=107
bbbeeeeeggggghiiiiihiiiiifggf
@SRR616268.4052 2291:6:1101:1412:2732 length=107
AGATCGGAAGAGCACACGTCTGAACTCCAGTCACCGATGTATCATTA
+SRR616268.4052 2291:6:1101:1412:2732 length=107
^^Zac\`cCHQ[\`bdZ[H`adeddddbdcachdhccc0accba]cc
iu@bielinux[srp017156]
```

[1:49午後]
[1:49午後]
[1:50午後]

W11-9: 動作確認

図5と基本的に同じ。W9-9でも利用している、文字列検索を行うgrepコマンドで検証。①は赤線で示したアダプター配列をどこかに完全一致で含む行数。②は赤線で示したアダプター配列を行頭で完全一致で含む行数。^は正規表現で行頭の意味。

```
iu@bielinux[srp017156] pwd
/home/iu/Documents/srp017156
iu@bielinux[srp017156] wc SRR616268sub_1.fastq [ 2:25午後 ]
4000000 8000000 320760716 SRR616268sub_1.fastq
① iu@bielinux[srp017156] grep -c GATCGGAAGAGCACACGTCTGAACTCCAGTCACTTAGGC
ATCTCGTATGC SRR616268sub_1.fastq
2812
② iu@bielinux[srp017156] grep -c ^GATCGGAAGAGCACACGTCTGAACTCCAGTCACTTAGG
CATCTCGTATGC SRR616268sub_1.fastq
2415
iu@bielinux[srp017156] grep ^GATCGGAAGAGCACACGTCTGAACTCCAGTCACTTAGGCAT
CTCGTATGC SRR616268sub_1.fastq | head -n 3
GATCGGAAGAGCACACGTCTGAACTCCAGTCACTTAGGCATCTCGTATGCCGTCTTCTGCTTGAAAAGAA
TCTGAGCAAACCCCTGGTTGTCGGGGCACGGAATAC
GATCGGAAGAGCACACGTCTGAACTCCAGTCACTTAGGCATCTCGTATGCCGTCTTCTGCTTGAAAAAAA
AAATAAAAGAAAAAGACAACACTCAAGCATGGGGGATGG
GATCGGAAGAGCACACGTCTGAACTCCAGTCACTTAGGCATCTCGTATGCCGTCTTCTGCTTGAAAAAAA
CAAAAACCTTAGTGAACAAACATCGGTCACGGGTCCG
iu@bielinux[srp017156] grep ^GATCGGAAGAGCACACGTCTGAACTCCAGTCACTTAGGCAT
CTCGTATGC SRR616268sub_1.fastq > age.txt
iu@bielinux[srp017156] less age.txt [ 2:52午後 ]
```

W11-9: 動作確認

③では、②の結果を行数ではなく(つまり-cオプションをつけずに)、行そのものを出力させている。「| head -n 3」をつけることで最初の3行分のみ表示。確かに^をつけているから水色下線で示すようにアダプター配列と完全一致のものが見られる。

```
iu@bielinux[srp017156] pwd
/home/iu/Documents/srp017156
iu@bielinux[srp017156] wc SRR616268sub_1.fastq
4000000 8000000 320760716 SRR616268sub_1.fastq
iu@bielinux[srp017156] grep -c GATCGGAAGAGCACACGTCTGAACTCCAGTCACTTAGGC
ATCTCGTATGC SRR616268sub_1.fastq
2812
iu@bielinux[srp017156] grep -c ^GATCGGAAGAGCACACGTCTGAACTCCAGTCACTTAGG
CATCTCGTATGC SRR616268sub_1.fastq
2415
iu@bielinux[srp017156] grep ^GATCGGAAGAGCACACGTCTGAACTCCAGTCACTTAGGCAT
CTCGTATGC SRR616268sub_1.fastq | head -n 3
GATCGGAAGAGCACACGTCTGAACTCCAGTCACTTAGGCATCTCGTATGCCGTCTTCTGCTTGAAAAGAA
TCTGAGCAAACCCCTGGTTGTCGGGGCACGGAATAC
GATCGGAAGAGCACACGTCTGAACTCCAGTCACTTAGGCATCTCGTATGCCGTCTTCTGCTTGAAAAAAA
AAATAAAAGAAAAAGACAACACTCAAGCATGGGGGATGG
GATCGGAAGAGCACACGTCTGAACTCCAGTCACTTAGGCATCTCGTATGCCGTCTTCTGCTTGAAAAAAA
CAAAAACCTTAGTGAACAAACATCGGTCACGGGTCCG
iu@bielinux[srp017156] grep ^GATCGGAAGAGCACACGTCTGAACTCCAGTCACTTAGGCAT
CTCGTATGC SRR616268sub_1.fastq > age.txt
iu@bielinux[srp017156] less age.txt
```

[2:52午後]

W11-9: 動作確認

もちろん④のようにgrepの結果をage.txtなどのファイルに書きだしたのち、⑤less(やmore)コマンドなどでざっと眺めるのもよい

```
iu@bielinux[srp017156] pwd [ 2:21午後 ]
/home/iu/Documents/srp017156
iu@bielinux[srp017156] wc SRR616268sub_1.fastq [ 2:25午後 ]
4000000 8000000 320760716 SRR616268sub_1.fastq
① iu@bielinux[srp017156] grep -c GATCGGAAGAGCACACGTCTGAACTCCAGTCACTTAGGC
ATCTCGTATGC SRR616268sub_1.fastq
2812
② iu@bielinux[srp017156] grep -c ^GATCGGAAGAGCACACGTCTGAACTCCAGTCACTTAGG
CATCTCGTATGC SRR616268sub_1.fastq
2415
③ iu@bielinux[srp017156] grep ^GATCGGAAGAGCACACGTCTGAACTCCAGTCACTTAGGCAT
CTCGTATGC SRR616268sub_1.fastq | head -n 3
GATCGGAAGAGCACACGTCTGAACTCCAGTCACTTAGGCATCTCGTATGCCGTCTTCTGCTTGAAAAGAA
TCTGAGCAAACCCCTGGTTGTCGGGGCACGGAATAC
GATCGGAAGAGCACACGTCTGAACTCCAGTCACTTAGGCATCTCGTATGCCGTCTTCTGCTTGAAAAAAA
AAATAAAAGAAAAAGACAACACTCAAGCATGGGGGATGG
GATCGGAAGAGCACACGTCTGAACTCCAGTCACTTAGGCATCTCGTATGCCGTCTTCTGCTTGAAAAAAA
CAAAAACCTTAGTGAACAAACATCGGTCACGGGTCCG
④ iu@bielinux[srp017156] grep ^GATCGGAAGAGCACACGTCTGAACTCCAGTCACTTAGGCAT
CTCGTATGC SRR616268sub_1.fastq > age.txt
⑤ iu@bielinux[srp017156] less age.txt [ 2:52午後 ]
```

W11-10: 動作確認

重要なのは、「W8-6の出現回数」と「②の結果」が一致し、①の結果と一致しないところ。W8-6の Overrepresented sequencesがなぜ50 bpなのかは(マニュアルに書いているのかもしれないが見ていないので)よくわからない。しかし、少なくとも「FastQC実行結果である黒矢印部分の出現回数は最初の50 bpのみで評価している」と判断。

```
iu@bielinux[srp017156] pwd
/home/iu/Documents/srp017156
iu@bielinux[srp017156] wc SRR616268sub_1.
 4000000  8000000 320760716 SRR616268sub_1.fastq
① iu@bielinux[srp017156] grep -c GATCGGAAGAGCACACGTCTGAACTCCAGTCACTTAGGC
ATCTCGTATGC SRR616268sub_1.fastq
2812
② iu@bielinux[srp017156] grep -c ^GATCGGAAGAGCACACGTCTGAACTCCAGTCACTTAGG
CATCTCGTATGC SRR616268sub_1.fastq
2415
iu@bielinux[srp017156] grep ^GATCGGAAGAGCACACGTCTGAACTCCAGTCACTTAGGCAT
CTCGTATGC SRR616268sub_1.fastq | head -n 3
GATCGGAAGAGCACACGTCTGAACTCCAGTCACTTAGGCATCTCGTATGCCGTCCTTCTGCTTGAAAAGAA
TCTGAGCACTAGGGAGTATTTAGCCTTGGGAGATGGTCCTCCGGATTCCGACGGAAT 2421 0.2421 No Hit
GATCGGAACTAGGGAGTATTTAGCCTTGGGAGATGGTCCTCCGGATTCCGACGGAAT 2415 0.2415 TruSeq Adapter, Index 3
AAATAAAA GATCGGAAGAGCACACGTCTGAACTCCAGTCACTTAGGCATCTCGTATGC 2415 0.2415 (100% over 50bp)
GATCGGAACTAGGGAGTATTTAGCCTTGGGAGATGGTCCTCCGGATTCCGACGGAAT 2393 0.2393 No Hit
CAAAAACC CCGGCATTCTCACTTCTAAGCGCTCCAGCCGTCCTCAGATCAACCTTCA 2393 0.2393 No Hit
iu@bielinux[srp017156] grep ^GATCGGAAGAGCACACGTCTGAACTCCAGTCACTTAGGCAT
CTCGTATGC SRR616268sub_1.fastq > age.txt
iu@bielinux[srp017156] less age.txt
```

TCTGAGCACTAGGGAGTATTTAGCCTTGGGAGATGGTCCTCCGGATTCCGACGGAAT	2421	0.2421	No Hit
GATCGGAACTAGGGAGTATTTAGCCTTGGGAGATGGTCCTCCGGATTCCGACGGAAT	2415	0.2415	TruSeq Adapter, Index 3 (100% over 50bp)
AAATAAAA GATCGGAAGAGCACACGTCTGAACTCCAGTCACTTAGGCATCTCGTATGC	2415	0.2415	
GATCGGAACTAGGGAGTATTTAGCCTTGGGAGATGGTCCTCCGGATTCCGACGGAAT	2393	0.2393	No Hit
CAAAAACC CCGGCATTCTCACTTCTAAGCGCTCCAGCCGTCCTCAGATCAACCTTCA	2393	0.2393	No Hit

[2:52午後]

W11-10: 動作確認

念のため出現回数上位の配列も調べ、判断の妥当性を確認。このことからOverrepresented sequences欄は、リードの最初のほうのみ(5'側のみ)で判断しているため、アダプターがリードの最後のほうに含まれるsmall RNAの場合にはこの結果のみで判断するのは危険だろうと予想する。ただし、small RNAの場合は読めるリード長が短いNGS機器を用いるので事実上気にしなくてもいいかもしれない。

```
iu@bielinux[~/Documents/srp017156]
iu@bielinux[srp017156] pwd
/home/iu/Documents/srp017156
iu@bielinux[srp017156] grep -c CCCCCG
GCTATTACGCA SRR616268sub_1.fastq
29372
iu@bielinux[srp017156] grep -c ^CCCCGGTATATTTTCGGCGCAGTGCCACTCGACTAGTG
AGCTATTACGCA SRR616268sub_1.fastq
14383
iu@bielinux[srp017156] grep -c GGCCTATTCAGTGCAGGCTGACCTTGCGGTCAGCACCCCT
TCTTCCGAAGT SRR616268sub_1.fastq
20429
iu@bielinux[srp017156] grep -c ^GGCCTATTCAGTGCAGGCTGACCTTGCGGTCAGCACCCC
TTCTTCCGAAGT SRR616268sub_1.fastq
11044
iu@bielinux[srp017156]
```

Overrepresented sequences

Sequence	Count
CCCCGGTATATTTTCGGCGCAGTGCCACTCGACTAGTGAGCTATTACGCA	14383
GGCCTATTCAGTGCAGGCTGACCTTGCGGTCAGCACCCCTTCTTCCGAAGT	11044
GTGCTTTTCACCTTTCCTCACGGTACTGGTTCAGTATCGGTCAGTACAGGG	8892
CCCGGTATATTTTCGGCGCAGTGCCACTCGACTAGTGAGCTATTACGCAC	8474
GTCAGTACAGGGAGTATTTAGCCTTGGGAGATGGTCCCTCCCGGATTCCGACG	8189

W11-11: Tips

①の結果は2,812行。②の結果は2,415行。③
grep -vで指定した文字列とマッチしない行を出力
させることもできる。2,812行の中で、行頭から検
索文字列であるアダプター配列を含まないものは
397行だけということ。

```
iu@bielinux[~/Documents/srp017156]
iu@bielinux[srp017156] pwd
/home/iu/Documents/srp017156
① iu@bielinux[srp017156] grep GATCGGAAGAGCACACGTCTGAACTCCAGTCACTTAGGCATC
TCGTATGC SRR616268sub_1.fastq > uge1.txt
iu@bielinux[srp017156] wc uge1.txt [ 4:24午後 ]
 2812  2812 303696 uge1.txt
② iu@bielinux[srp017156] grep -c ^GATCGGAAGAGCACACGTCTGAACTCCAGTCACTTAGG
CATCTCGTATGC uge1.txt
2415
③ iu@bielinux[srp017156] grep -v ^GATCGGAAGAGCACACGTCTGAACTCCAGTCACTTAGG
CATCTCGTATGC uge1.txt > uge2.txt
iu@bielinux[srp017156] wc uge2.txt [ 4:24午後 ]
  397   397 42876 uge2.txt
iu@bielinux[srp017156] head -n 4 uge2.txt [ 4:24午後 ]
AGATCGGAAGAGCACACGTCTGAACTCCAGTCACTTAGGCATCTCGTATGCCGTCTTCTGCTTGAAAAAA
ACACAATGCCACTCGAATAGTGAGCTTATAAGATCGG
AGATCGGAAGAGCACACGTCTGAACTCCAGTCACTTAGGCATCTCGTATGCCGTCTTCTGCTTGAAAAAA
AAAAAAAAACAAACCGCAGGCAAGAGAAGAAGATCGG
ACCGATATGCTTCGGGAGATCGGAAGAGCACACGTCTGAACTCCAGTCACTTAGGCATCTCGTATGCCGT
CTTCTGCTTGAAAAAAAAAAAAAAAAACAAAATCCCGAAG
AGATCGGAAGAGCACACGTCTGAACTCCAGTCACTTAGGCATCTCGTATGCCGTCTTCTGCTTGAAAAAA
AAAAAATAAAATACAAACGCACGCACGCGGAGATCGG
iu@bielinux[srp017156] less uge2.txt [ 4:24午後 ]
```

W11-11: Tips

ただし、どこかにその文字列を含むことは、①の結果を入力としていることから保証されている。④ uge2.txtの最初の4行分を表示。赤下線部分がアダプター配列。この程度なら「目grep」でどうにかなるが、⑤のlessコマンドで開いてからアダプター配列を検索するほうがよいかも[W2-2; W7-5; 連載第3回のW14-6]。

```
iu@bielinux[~/Documents/srp017156]
① iu@bielinux[srp017156] pwd
/home/iu/Documents/srp017156
iu@bielinux[srp017156] grep GATCGGAAGAGC
TCGTATGC SRR616268sub_1.fastq > uge1.txt
iu@bielinux[srp017156] wc uge1.txt
2812 2812 303696 uge1.txt
② iu@bielinux[srp017156] grep -c ^GATCGGAAGAGCACACGTCTGAACTCCAGTCACTTAGG
CATCTCGTATGC uge1.txt
2415
③ iu@bielinux[srp017156] grep -v ^GATCGGAAGAGCACACGTCTGAACTCCAGTCACTTAGG
CATCTCGTATGC uge1.txt > uge2.txt
iu@bielinux[srp017156] wc uge2.txt
397 397 42876 uge2.txt
④ iu@bielinux[srp017156] head -n 4 uge2.txt
AGATCGGAAGAGCACACGTCTGAACTCCAGTCACTTAGGCATCTCGTATGCCGTCTTCTGCTTGAAAAAA
ACACAATGCCACTCGAATAGTGAGCTTATAAGATCGG
AGATCGGAAGAGCACACGTCTGAACTCCAGTCACTTAGGCATCTCGTATGCCGTCTTCTGCTTGAAAAAA
AAAAAAAAACAAACCGCAGGCAAGAGAAGAAGATCGG
ACCGATATGCTTCGGGAGATCGGAAGAGCACACGTCTGAACTCCAGTCACTTAGGCATCTCGTATGCCGT
CTTCTGCTTGAAAAAAACAAAAATCCCGAAG
AGATCGGAAGAGCACACGTCTGAACTCCAGTCACTTAGGCATCTCGTATGCCGTCTTCTGCTTGAAAAAA
AAAAAATAAAATACAAACGCACGCACGCGGAGATCGG
⑤ iu@bielinux[srp017156] less uge2.txt
```

[4:24午後]

[4:24午後]

[4:24午後]

[4:24午後]

「less uge2.txt」実行直後。連載第3回のW14-6-2を参考にして、アダプター配列を検索する。

W11-12: Tips

```
iu@bielinux[~/Documents/srp017156] 16:38
AGATCGGAAGAGCACACGTCTGAACTCCAGTCACTTAGGCATCTCGTATGCCGTCTTCTGCTTGAAAAAA
ACACAATGCCACTCGAATAGTGAGCTTATAAGATCGG
AGATCGGAAGAGCACACGTCTGAACTCCAGTCACTTAGGCATCTCGTATGCCGTCTTCTGCTTGAAAAAA
AAAAAAAAACAAACCGCAGGCAAGAGAAGAAGATCGG
ACCGATATGCTTCGGGAGATCGGAAGAGCACACGTCTGAACTCCAGTCACTTAGGCATCTCGTATGCCGT
CTTCTGCTTGAAAAAAAAAAAAAAAAACAAAATCCCGAAG
AGATCGGAAGAGCACACGTCTGAACTCCAGTCACTTAGGCATCTCGTATGCCGTCTTCTGCTTGAAAAAA
AAAAAATAAAATACAAACGCACGCACGCGGAGATCGG
GGGACGATCACCCAGATCGGAAGAGCACACGTCTGAACTCCAGTCACTTAGGCATCTCGTATGCCGTCTT
CTGCTTGAAAAAAAAAAAAAAAAACACACGGGTGAT
AGATCGGAAGAGCACACGTCTGAACTCCAGTCACTTAGGCATCTCGTATGCCGTCTTCTGCTTGAAAAAA
CTGTAAAAGCACCATCCTTTTCACTTAGAGATCGG
AGATCGGAAGAGCACACGTCTGAACTCCAGTCACTTAGGCATCTCGTATGCCGTCTTCTGCTTGAAAAAA
ACAAAAAAAAACCAAGACGAACACCACGGAGATCGG
AGATCGGAAGAGCACACGTCTGAACTCCAGTCACTTAGGCATCTCGTATGCCGTCTTCTGCTTGAAAAAA
ACAATCGAACGCTTCAGTGCGCACTCCCCGAGATCGG
AGATCGGAAGAGCACACGTCTGAACTCCAGTCACTTAGGCATCTCGTATGCCGTCTTCTGCTTGAAAAAA
AAACAAAAAAAAAATATTGTAGAGATACATAGATCGG
AGATCGGAAGAGCACACGTCTGAACTCCAGTCACTTAGGCATCTCGTATGCCGTCTTCTGCTTGAAAC
GCAAAGAGAGTTTCGTGTTGCTTCTTTAGTAGATCGG
GCGCACCTTACGGGTGCACCCACAGTTTCGGAGATCGGAAGAGCACACGTCTGAACTCCAGTCACTTAG
GCATCTCGTATGCCGTCTTCTGCTTGAAACCGAAAC
uge2.txt
```


スラッシュ (/) の後に予めコピーしておいたアダプター配列をペーストした直後の状態。リターンキーを押す。

W11-12: Tips

```
iu@bielinux[~/Documents/srp017156] 16:51
AGATCGGAAGAGCACACGTCTGAACTCCAGTCACTTAGGCATCTCGTATGCCGTCTTCTGCTTGAAAAAA
ACACAATGCCACTCGAATAGTGAGCTTATAAGATCGG
AGATCGGAAGAGCACACGTCTGAACTCCAGTCACTTAGGCATCTCGTATGCCGTCTTCTGCTTGAAAAAA
AAAAAAAAACAAACCGCAGGCAAGAGAAGAAGATCGG
ACCGATATGCTTCGGGAGATCGGAAGAGCACACGTCTGAACTCCAGTCACTTAGGCATCTCGTATGCCGT
CTTCTGCTTGAAAAAAAAAAAAAAAAACAAAATCCCGAAG
AGATCGGAAGAGCACACGTCTGAACTCCAGTCACTTAGGCATCTCGTATGCCGTCTTCTGCTTGAAAAAA
AAAAAATAAAATACAAACGCACGCACGCGGAGATCGG
GGGACGATCACCCAGATCGGAAGAGCACACGTCTGAACTCCAGTCACTTAGGCATCTCGTATGCCGTCTT
CTGCTTGAAAAAAAAAAAAAAAAACACACGGGTGAT
AGATCGGAAGAGCACACGTCTGAACTCCAGTCACTTAGGCATCTCGTATGCCGTCTTCTGCTTGAAAAAA
CTGTAAAAGCACCATCCTTTTCAACTTAGAGATCGG
AGATCGGAAGAGCACACGTCTGAACTCCAGTCACTTAGGCATCTCGTATGCCGTCTTCTGCTTGAAAAAA
ACAAAAAAAAAACCAAGACGAACACCACGGAGATCGG
AGATCGGAAGAGCACACGTCTGAACTCCAGTCACTTAGGCATCTCGTATGCCGTCTTCTGCTTGAAAAAA
ACAATCGAACGCTTCAGTGCGCACTCCCCGAGATCGG
AGATCGGAAGAGCACACGTCTGAACTCCAGTCACTTAGGCATCTCGTATGCCGTCTTCTGCTTGAAAAAA
AAACAAAAAAAAAATATTGTAGAGATACATAGATCGG
AGATCGGAAGAGCACACGTCTGAACTCCAGTCACTTAGGCATCTCGTATGCCGTCTTCTGCTTGAAAC
GCAAAGAGAGTTTCGTGTTGCTTCTTTAGTAGATCGG
GCGCACCTTACGGGTGCACCCACAGTTTCGGAGATCGGAAGAGCACACGTCTGAACTCCAGTCACTTAG
GCATCTCGTATGCCGTCTTCTGCTTGAAACCGAAAC
/GATCGGAAGAGCACACGTCTGAACTCCAGTCACTTAGGCATCTCGTATGC
```

W11-12: Tips

アダプター配列の位置がハイライトされる。これらがFastQC実行結果である「W8-6の出現回数」に反映されていないものたち。パッと見て、ハイライトされている50bpのアダプター配列は全体の一部であり、赤下線部分程度の長さは本当はありそうだと閃く。自由に眺めてqで終了。

iu@bielinux[~/Documents/srp017156]

```
AGATCGGAAGAGCACACGTCTGAACTCCAGTCA
ACACAATGCCACTCGAATAGTGAGCTTATAAGATCG
AGATCGGAAGAGCACACGTCTGAACTCCAGTCACTTAGGCATCTCGTATGCCGTCTTCTGCTTGAAAAA
AAAAAAAAACAAACCGCAGGCAAGAGAAGAAGATCGG
ACCGATATGCTTCGGGAGATCGGAAGAGCACACGTCTGAACTCCAGTCACTTAGGCATCTCGTATGCCGT
CTTCTGCTTGAAAAAAAAAAAAAAAAACAAAATCCCGAAG
AGATCGGAAGAGCACACGTCTGAACTCCAGTCACTTAGGCATCTCGTATGCCGTCTTCTGCTTGAAAAA
AAAAAATAAAATACAAACGCACGCACGCGGAGATCGG
GGGACGATCACCCAGATCGGAAGAGCACACGTCTGAACTCCAGTCACTTAGGCATCTCGTATGCCGTCTT
CTGCTTGAAAAAAAAAAAAAAAAAACACACGGGTGAT
AGATCGGAAGAGCACACGTCTGAACTCCAGTCACTTAGGCATCTCGTATGCCGTCTTCTGCTTGAAAAA
CTGTAAAAGCACCATCCTTTTCACTTAGAGATCGG
AGATCGGAAGAGCACACGTCTGAACTCCAGTCACTTAGGCATCTCGTATGCCGTCTTCTGCTTGAAAAA
ACAAAAAAAAAACCAAGACGAACACCACGGAGATCGG
AGATCGGAAGAGCACACGTCTGAACTCCAGTCACTTAGGCATCTCGTATGCCGTCTTCTGCTTGAAAAA
ACAATCGAACGCTTCAGTGCGCACTCCCCGAGATCGG
AGATCGGAAGAGCACACGTCTGAACTCCAGTCACTTAGGCATCTCGTATGCCGTCTTCTGCTTGAAAAA
AAACAAAAAAAAAATATTGTAGAGATACATAGATCGG
AGATCGGAAGAGCACACGTCTGAACTCCAGTCACTTAGGCATCTCGTATGCCGTCTTCTGCTTGAAAC
GCAAAGAGAGTTTCGTGTTGCTTCTTTAGTAGATCGG
GCGCACCTTACGGGTGCACCCACAGTTTCGGAGATCGGAAGAGCACACGTCTGAACTCCAGTCACTTAG
GCATCTCGTATGCCGTCTTCTGCTTGAAACCGAAAC
:
```

W11-13: アダプター配列

GOOGLE TruSeq Adapter, Index 3

ウェブ ショッピング 画像 動画 ニュース もっと見る ▼ 検索ツール

約 12,600 件 (0.18 秒)

ヒント: 日本語の検索結果のみ表示します。検索言語は [表示設定] で指定できます。

[PDF] [PDF] Illumina Customer Sequence Letter
supportres.illumina.com/.../illumina-customer-sequenc... ▼ このページを訳す
2014/07/08 - 5'
GATCGGAAGAGCACACGTCTGAACTCCAGTCACCGATGTATCTCGTATGCCGTCTTCTGCTTG.
TruSeq Adapter, Index 3. 5'
GATCGGAAGAGCACACGTCTGAACTCCAGTCACTTAGGCATCTCGTATGCCGTCTTCTGCTTG.

Illumina Single End Apapter 1 ...
www.researchgate.net/publictopics.PublicPostFileLoad... ▼ このページを訳す
... PCR Primer Index 2
CAAGCAGAAGACGGCATAACGAGATACATCGGTGACTGGAGTTC Illumina PCR
Primer Index 3 ...
AATGATACGGCGACCACCGAGATCTACTCTTTCCCTACACGACGCTCTTCCGATCT
TruSeq Adapter, Index 1 ...

[PDF] NGSの新たな利用法 Tailed PCR法を用いたライブラリー調製
www.illumina.co.jp/documents/.../2014_techsupport_session5.pdf ▼

AGATCGGAAGAGCACACGTCT
GAACTCCAGTCACTTAGGCATC
TCGTATGCCGTCTTCTGCTTGと
いう赤下線のアダプター配列候補、
またはW8-6から得られる「TruSeq
Adapter, Index 3」でウェブ検索し、
一番左側のAは違うようだが、残りは
予想通りであることがわかる。

W12-1

①Index 1-5のアダプター配列を含むリード数、およびIndex 1-27を含むリード数を一度に調べるシェルスクリプトファイル(JSLAB4_6_Linux.sh)を用意し、②実行。Index 1を含むものが619リード、Index 2が2,625リード、Index 3が2,603リード、Index 1-27を含むものが5,906リードであることがわかる。赤下線部分がインデックス配列

```
iu@bielinux[~/Documents/srp017156]
iu@bielinux[srp017156] pwd
/home/iu/Documents/srp017156
iu@bielinux[srp017156] ls -lh *.fastq
-rw-rw-r-- 1 iu iu 306M 5月 6 10:07 SRR616268sub_1.fastq
iu@bielinux[srp017156] wget -cq http://www.iu.a.u-tokyo.ac.jp/~kadota/book/JSLAB4_6_Linux.sh
iu@bielinux[srp017156] more JSLAB4_6_Linux.sh [ 2:04午後 ]
### TruSeq Adapter, Index 1
grep -c GATCGGAAGAGCACACGTCTGAACTCCAGTCACATCACGATCTCGTATGCCGTCTTCTGCTTG SRR616268sub_1.fastq
### TruSeq Adapter, Index 2
grep -c GATCGGAAGAGCACACGTCTGAACTCCAGTCACCGATGTATCTCGTATGCCGTCTTCTGCTTG SRR616268sub_1.fastq
### TruSeq Adapter, Index 3
grep -c GATCGGAAGAGCACACGTCTGAACTCCAGTCACTTAGGCATCTCGTATGCCGTCTTCTGCTTG SRR616268sub_1.fastq
### TruSeq Adapter, Index 4
grep -c GATCGGAAGAGCACACGTCTGAACTCCAGTCACTGACCAATCTCGTATGCCGTCTTCTGCTTG SRR616268sub_1.fastq
### TruSeq Adapter, Index 5
grep -c GATCGGAAGAGCACACGTCTGAACTCCAGTCACTGACCAATCTCGTATGCCGTCTTCTGCTTG SRR616268sub_1.fastq
### TruSeq Adapter, Index 1-27
grep -c GATCGGAAGAGCACACGTCTGAACTCCAGTCAC.....ATCTCGTATGCCGTCTTCTGCTTG SRR616268sub_1.fastq
iu@bielinux[srp017156] sh JSLAB4_6_Linux.sh [ 2:05午後 ]
619
2625
2603
0
0
5906
iu@bielinux[srp017156] █ [ 2:05午後 ]
```

W12-2: Tips

```
iu@bielinux[~/Documents/srp017156] [ 3:23午後 ]
iu@bielinux[srp017156] pwd
/home/iu/Documents/srp017156
iu@bielinux[srp017156] wget -cq http://www.iu.a.u-tokyo.ac.jp/~kadota/book/JSLAB4_7_Linux.sh
iu@bielinux[srp017156] more JSLAB4_7_Linux.sh [ 3:24午後 ]
### TruSeq Adapter, Index 1
grep GATCGGAAGAGCACACGTCTGAACTCCAGTCACATCACGATCTCGTATGCCGTC
TTCTGCTTG SRR616268sub_1.fastq > hoge.txt
wc hoge.txt
### TruSeq Adapter, Index 2
grep GATCGGAAGAGCACACGTCTGAACTCCAGTCACCGATGTATCTCGTATGCCGTC
TTCTGCTTG SRR616268sub_1.fastq >> hoge.txt
wc hoge.txt
### TruSeq Adapter, Index 3
grep GATCGGAAGAGCACACGTCTGAACTCCAGTCACTTAGGCATCTCGTATGCCGTC
TTCTGCTTG SRR616268sub_1.fastq >> hoge.txt
wc hoge.txt
iu@bielinux[srp017156] ls -lh hoge* [ 3:24午後 ]
zsh: no matches found: hoge*
iu@bielinux[srp017156] sh JSLAB4_7_Linux.sh [ 3:25午後 ]
619 619 66852 hoge.txt
3244 3244 350352 hoge.txt
5847 5847 631476 hoge.txt
iu@bielinux[srp017156] [ 3:26午後 ]
```

①Index 1-3のアダプター配列を含むリード情報をhoge.txtファイルにどんどん追加保存するやり方。Index 1の結果を保存するときには通常のリダイレクト(>)。Index 2以降は既に存在するhoge.txtに追加保存をすべく、本連載では初出の追加書き込み(>>)を用いている。③で①のシェルスクリプトを実行し、hoge.txtの行数が619行、3244行、5847行と増えていることを確認。

W12-3: Tips

W11-6ではgzip圧縮ファイルを解凍してからfastx_clipperを実行した。しかし、連載第3回のW25に示すようにbzip2コマンドの-cオプション同様、①gzipコマンドでも-cオプションとパイプを併用することでfastx_clipperに限らずgzipファイルのまま作業を行うことができる。②の結果はW11-6と同じ。

```
iu@bielinux[~/Documents/srp017156]
iu@bielinux[srp017156] pwd
/home/iu/Documents/srp017156
iu@bielinux[srp017156] gzip SRR616268sub_1.fastq [ 5:37午後]
iu@bielinux[srp017156] ls -lh *sub* [ 5:38午後]
-rw-rw-r-- 1 iu iu 72M 5月 6 10:07 SRR616268sub_1.fastq.gz
-rw-rw-r-- 1 iu iu 65M 5月 6 10:08 SRR616268sub_2.fastq.gz
① iu@bielinux[srp017156] gunzip -c SRR616268sub_1.fastq.gz | fastx_clipper -a GATCGGAAGAGCACACGTCTGAACTCCAGTCACTTAGGCATCTCGTATGC -o hoge8.fastq
② iu@bielinux[srp017156] wc hoge8.fastq [ 5:41午後]
3915596 7831192 286011042 hoge8.fastq
iu@bielinux[srp017156] ls -lh *sub* [ 5:42午後]
-rw-rw-r-- 1 iu iu 72M 5月 6 10:07 SRR616268sub_1.fastq.gz
-rw-rw-r-- 1 iu iu 65M 5月 6 10:08 SRR616268sub_2.fastq.gz
iu@bielinux[srp017156] [ 5:45午後]
```

W13-1: 文字コード変換

「Linux 文字コード変換」での検索結果。iconvやnkfコマンドを利用するやり方が上位にきている。

Google Linux 文字コード変換

ウェブ 画像 ニュース ショッピング 動画 もっと見る ▼ 検索ツール

約 307,000 件 (0.15 秒)

他のキーワード: linux 文字コード 確認 linux 文字コード変換 一括

Linuxコマンド集 - 【iconv】文字コードを変換して出力する:ITpro
itpro.nikkeibp.co.jp > 情報システム > オープンソース/Linux ▼
2013/12/13 - 文字コードを変換して出力する。変換前のファイルをINPUTFILEに(もしくは標準入力に)、その文字コードを-tオプションで、変換先の文字コードを-Tオプションで指定する。変換結果は標準出力に出力される。ファイルに出力するには、-o ...

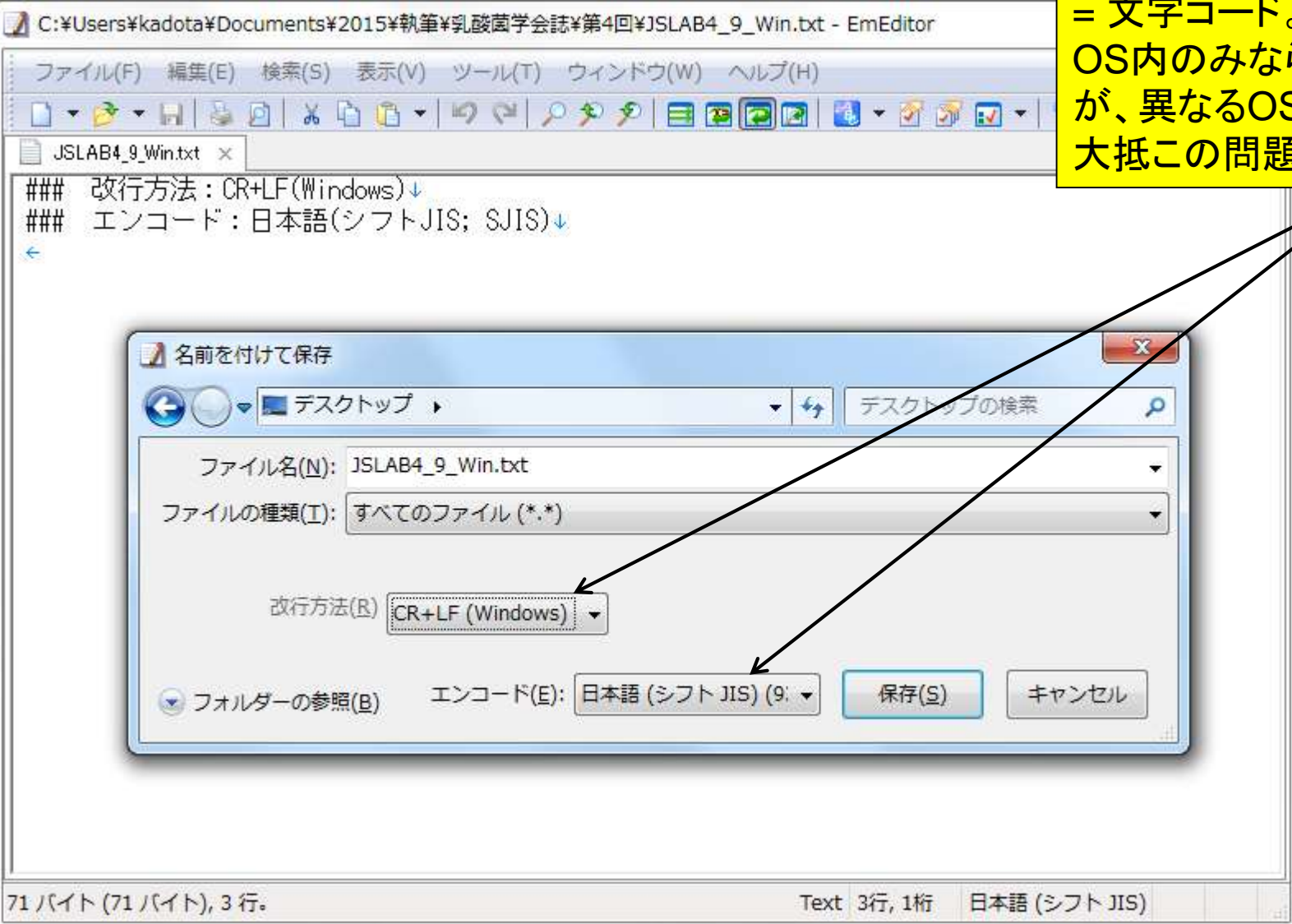
Linuxコマンド集 - 【nkf】文字コードを変換する:ITpro
itpro.nikkeibp.co.jp > 情報システム > オープンソース/Linux ▼
2006/02/27 - -b, バッファリング出力を行う。-u, 出力時にバッファリングを行わない。-j, JISコードに変換する。-e, EUCコードに変換する。-s, シフトJISコードに変換する。-w, UTF8コードに変換する。-i? JIS漢字を指示するシーケンスとして ESC-'\$'-?を使用する。

Linuxコマンド【iconv】ファイルの文字コードを変換 - Linux入門 ...
webkaru.net > Linux入門 > Linuxコマンド集 > テキスト ▼
iconvコマンドは、指定した入力ファイルの文字コードを変換するコマンドです。iconvコマンド書式 iconv cf. » nkfコマンド - 文字コードを変換 オプション -c変換できなかった文字を出力から除きます。-f 文字コード変換前の文字コード(エンコーディング)を指定し ...

Linuxコマンド【nkf】文字コードを変換 - Linux入門 - Webkaru
webkaru.net > Linux入門 > Linuxコマンド集 > テキスト ▼
nkfコマンドは、JIS、Shift_JIS、EUC、UTFといった文字コードを変換するコマンドです。nkfコマンド 書式 nkf オプション -bバッファリング出力を行います。-eEUCコードを出力します。-g文字コードを自動判別し、その結果を出力します。-jJISコードを出力します。

W13-2: 文字コード変換

JSLAB4_9_Win.txtは、Windowsの改行コード(CR+LF)、エンコードはシフトJISで保存したもの。エンコード = 文字コード。全ての作業が同じOS内のみなら気にする必要がないが、異なるOS間を行き来していると大抵この問題に悩まされる[W5-2]。



W13-3: iconv

①JSLAB4_9_Win.txtをBio-Linux上でmoreすると文字化けする。Bio-Linuxには②iconvのパスは通っているが、③nkfコマンドはないようだ。

```
iu@bielinux[~/Desktop]
iu@bielinux[Desktop] pwd [ 7:01午後 ]
/home/iu/Desktop
iu@bielinux[Desktop] wget -cq http://www.iu.a.u-tokyo.ac.jp
/~kadota/book/JSLAB4_9_Win.txt
iu@bielinux[Desktop] more JSLAB4_9_Win.txt [ 7:01午後 ]
### 000s000@FCR+LF(Windows)
### 0G000R0[0h0F000{00(0V0t0gJIS; SJIS)
iu@bielinux[Desktop] where iconv [ 7:01午後 ]
/usr/bin/iconv
iu@bielinux[Desktop] where nkf [ 7:02午後 ]
nkf not found
iu@bielinux[Desktop] [ 7:02午後 ]
```



W13-3: iconv

とりあえず①iconvを利用してLinux環境で読めるように文字コードを変換したものをhoge.txtというファイル名で保存。文字化けしているときは赤下線のオプションを無条件で打つと基本的にうまくいく。昔はLinux環境の日本語の文字コードはEUCだったらしいが、最近はUTF8らしい。③水色下線のように明示的にUTF8に変換するように指定したのちhoge2.txtに保存。このことから、-tオプションを指定しなければUTF8に自動的に変換するようだ。

```
iu@bielinux[Desktop] pwd
/home/iu/Desktop
iu@bielinux[Desktop] wget -cq http://www.iu
/~kadota/book/JSLAB4_9_Win.txt
iu@bielinux[Desktop] more JSLAB4_9_Win.txt
### 000s000@FCR+LF(Windows)
### 0G000R0[0h0F000{00(0V0t0gJIS; SJIS)
iu@bielinux[Desktop] where iconv
/usr/bin/iconv
iu@bielinux[Desktop] where nkf [ 7:02午後 ]
nkf not found
① iu@bielinux[Desktop] iconv -f SJIS JSLAB4_9_Win.txt > hoge.
txt
② iu@bielinux[Desktop] more hoge.txt [ 7:06午後 ]
### 改行方法 : CR+LF(Windows)
### エンコード : 日本語 (シフト JIS; SJIS)
③ iu@bielinux[Desktop] iconv -f SJIS -t UTF8 JSLAB4_9_Win.txt
> hoge2.txt
④ iu@bielinux[Desktop] more hoge2.txt [ 7:20午後 ]
### 改行方法 : CR+LF(Windows)
### エンコード : 日本語 (シフト JIS; SJIS)
iu@bielinux[Desktop] [ 7:20午後 ]
```

W13-4: nkf使いたい

JSLAB4_9_Win.txtは自分でShift JISで保存したことが分かっているので、iconv利用時に「-f SJIS」が正しいと分かる。文字コードが何かを調べたい場合には、nkfコマンドを利用するやり方が示されているので、nkfをどうにかして利用したい!

```
iu@bielinux[Desktop] pwd [
/home/iu/Desktop
iu@bielinux[Desktop] wget -cq http://www.iu.a.u-t
/~kadota/book/JSLAB4_9_Win.txt
iu@bielinux[Desktop] more JSLAB4_9_Win.txt [ 7:01午後 ]
### 000s000@0FCR+LF(Windows)
### 0G000R0[0h0F000{00(0V0t0gJIS; SJIS)
iu@bielinux[Desktop] where iconv [ 7:01午後 ]
/usr/bin/iconv
iu@bielinux[Desktop] where nkf [ 7:02午後 ]
nkf not found
iu@bielinux[Desktop] iconv -f SJIS JSLAB4_9_Win.txt > hoge.
txt
iu@bielinux[Desktop] more hoge.txt [ 7:06午後 ]
### 改行方法 : CR+LF(Windows)
### エンコード : 日本語 (シフト JIS; SJIS)
iu@bielinux[Desktop] iconv -f SJIS -t UTF8 JSLAB4_9_Win.txt
> hoge2.txt
iu@bielinux[Desktop] more hoge2.txt [ 7:20午後 ]
### 改行方法 : CR+LF(Windows)
### エンコード : 日本語 (シフト JIS; SJIS)
iu@bielinux[Desktop] [ 7:20午後 ]
```

W13-5:nkfインストール

```
File Edit View Search Terminal Help 19:58
① iu@bielinux[Desktop] pwd [ 7:56午後 ]
/home/iu/Desktop
② iu@bielinux[Desktop] where nkf [ 7:56午後 ]
nkf not found
③ iu@bielinux[Desktop] sudo apt-get install nkf [ 7:56午後 ]
[sudo] password for iu:
Reading package lists... Done
Building dependency tree
Reading state information... Done
The following NEW packages will be installed:
  nkf
0 upgraded, 1 newly installed, 0 to remove and 694 not upgraded.
Need to get 123 kB of archives.
After this operation, 357 kB of additional disk space will be used.
Get:1 http://jp.archive.ubuntu.com/ubuntu/ trusty/universe
nkf amd64 2.13-1 [123 kB]
Fetched 123 kB in 0s (289 kB/s)
Selecting previously unselected package nkf.
(Reading database ... 314160 files and directories currently installed.)
Preparing to unpack .../archives/nkf_2.13-1_amd64.deb ...
Unpacking nkf (2.13-1) ...
Processing triggers for man-db (2.6.7.1-1) ...
Setting up nkf (2.13-1) ...
④ iu@bielinux[Desktop] where nkf [ 7:57午後 ]
/usr/bin/nkf
iu@bielinux[Desktop] [ 7:57午後 ]
```

①作業ディレクトリはどこでもよい。
②インストール前はnkfのパスは通っていない。
③nkfコマンドのインストール。「sudo apt-get install プログラム名」が基本形。ログインパスワード(推奨手順通りだとpass1409)を聞かれたら打ち込む。インストール自体は1分程度で終わる。特にエラーメッセージも出ていないが念のため④/usr/bin/nkfとパスが無事通っていることが分かる。

W13-6:nkfを利用

①～③の--guess、あるいは④～⑥の-gオプションをつけてnkfコマンドを実行すると、入力として与えたファイルの文字コードがわかる。

```
iu@bielinux[Desktop] pwd [ 8:11午後 ]
/home/iu/Desktop
iu@bielinux[Desktop] ls -lh *.txt [ 8:11午後 ]
-rw-rw-r-- 1 iu iu 88 5月 27 19:20 hoge2.txt
-rw-rw-r-- 1 iu iu 88 5月 27 19:06 hoge.txt
-rw-rw-r-- 1 iu iu 71 5月 27 18:06 JSLAB4_9_Win.txt
① iu@bielinux[Desktop] nkf --guess JSLAB4_9_Win.txt
Shift_JIS (LF)
② iu@bielinux[Desktop] nkf --guess hoge.txt [ 8:11午後 ]
UTF-8 (LF)
③ iu@bielinux[Desktop] nkf --guess hoge2.txt [ 8:11午後 ]
UTF-8 (LF)
④ iu@bielinux[Desktop] nkf -g JSLAB4_9_Win.txt [ 8:12午後 ]
Shift_JIS
⑤ iu@bielinux[Desktop] nkf -g hoge.txt [ 8:15午後 ]
UTF-8
⑥ iu@bielinux[Desktop] nkf -g hoge2.txt [ 8:15午後 ]
UTF-8
iu@bielinux[Desktop] [ 8:15午後 ]
```

W13-6:nkfを利用

①nkfコマンドを利用した文字コード変換。②iconvと違って、入力ファイルの文字コードを自動的に認識してくれるため、よほどのことがない限りここで示すようにシンプルで便利。

```
iu@bielinux[Desktop] pwd
/home/iu/Desktop
iu@bielinux[Desktop] ls -lh *.txt
-rw-rw-r-- 1 iu iu 88  5月 27 19:20 hoge2.txt
-rw-rw-r-- 1 iu iu 88  5月 27 19:06 hoge.txt
-rw-rw-r-- 1 iu iu 71  5月 27 18:06 JSLAB4_9_Win.txt
iu@bielinux[Desktop] more JSLAB4_9_Win.txt
###  ???s???@FCR+LF(Windows)
###  ?G???R?[?h?F???{??(??V?t?gJIS; SJIS)
① iu@bielinux[Desktop] nkf JSLAB4_9_Win.txt > hoge3.txt
iu@bielinux[Desktop] more hoge3.txt
###  改行方法 : CR+LF(Windows)
###  エンコード : 日本語 (シフト JIS; SJIS)
② iu@bielinux[Desktop] iconv JSLAB4_9_Win.txt > hoge4.txt
iconv: illegal input sequence at position 5
iu@bielinux[Desktop] more hoge4.txt
###
iu@bielinux[Desktop]
```

[8:33午後]

[8:33午後]

[8:34午後]

[8:38午後]

[8:39午後]

W13-7: アンインストール

```
iu@bielinux[~/Desktop]
① iu@bielinux[Desktop] nkf -v
Network Kanji Filter Version 2.1.3 (2013-11-22)
Copyright (C) 1987, FUJITSU LTD. (I.Ichikawa).
Copyright (C) 1996-2013, The nkf Project.
② iu@bielinux[Desktop] where nkf
/usr/bin/nkf
③ iu@bielinux[Desktop] apt-get remove nkf
E: Could not open lock file /var/lib/dpkg/lock - open
tion denied)
E: Unable to lock the administration directory (/var/
are you root?
④ iu@bielinux[Desktop] sudo apt-get remove nkf
Reading package lists... Done
Building dependency tree
Reading state information... Done
The following packages will be REMOVED:
  nkf
0 upgraded, 0 newly installed, 1 to remove and 694 not upgraded.
After this operation, 357 kB disk space will be freed.
Do you want to continue? [Y/n] y
(Reading database ... 314168 files and directories currently inst
alled.)
Removing nkf (2.13-1) ...
Processing triggers for man-db (2.6.7.1-1) ...
⑤ iu@bielinux[Desktop] where nkf
nkf not found
iu@bielinux[Desktop] sudo apt-get install nkf ⑥ [ 8:21午後 ]
[ 8:22午後 ]
```

①と②はnkfが使える状態であることを確認してるだけ。③で示す「apt-get remove パッケージ名」がアンインストール(プログラムを削除)するコマンドの基本形。sudoを入れ忘れたので権限が与えられていないと文句を言われている。④気を取り直してsudoをつけてやり直し。ログインパスワードを聞かれたら打ち込む。途中「本当にアンインストール作業を進めていいか?」と聞かれているのでyと打っている。⑤nkfコマンドがなくなったことを確認。一通り確認したら、⑥で再インストールしておきましょう。

W14-1: pipインストール

pipコマンドのインストール。①作業ディレクトリはどこでもよい。②と③でpipがないことを確認。nkfのときは「sudo apt-get install nkf」だったが、④pipの場合は「**sudo apt-get install pip**」ではない点に注意。ログインパスワードを聞かれたら打ち込む。

```
File Edit View Search Terminal Help
① iu@bielinux[Desktop] pwd
/home/iu/Desktop
② iu@bielinux[Desktop] where pip
pip not found
③ iu@bielinux[Desktop] which pip
pip not found
④ iu@bielinux[Desktop] sudo apt-get install python-pip [ 1:12午後]
[sudo] password for iu:
Reading package lists... Done
Building dependency tree
Reading state information... Done
The following extra packages will be installed:
 python-colorama python-distlib python-html5lib python-setuptools
 python-wheel
Suggested packages:
 python-genshi
Recommended packages:
 python-dev-all
The following NEW packages will be installed:
 python-colorama python-distlib python-html5lib python-pip python-
n-setuptools
 python-wheel
0 upgraded, 6 newly installed, 0 to remove and 694 not upgraded.
Need to get 589 kB of archives.
```

[1:12午後]

[1:13午後]

W14-1: pipインストール

途中「本当に作業を進めていいか？」と聞かれるのでyと打っている箇所。この作業はうっとおしいが、「sudo apt-get -y install python-pip」と-yオプションをつけておけば聞かれずにすむ。

```
python-colorama python-distlib python-html5lib py
python-wheel
Suggested packages:
python-genshi
Recommended packages:
python-dev-all
The following NEW packages will be installed:
python-colorama python-distlib python-html5lib python-pip pytho
n-setuptools
python-wheel
0 upgraded, 6 newly installed, 0 to remove and 694 not upgraded.
Need to get 589 kB of archives.
After this operation, 2,745 kB of additional disk space will be u
sed.
Do you want to continue? [Y/n] y
Get:1 http://jp.archive.ubuntu.com/ubuntu/ trusty-updates/univers
e python-colorama all 0.2.5-0.1ubuntu2 [18.4 kB]
Get:2 http://jp.archive.ubuntu.com/ubuntu/ trusty-updates/univers
e python-distlib all 0.1.8-1ubuntu1 [113 kB]
Get:3 http://jp.archive.ubuntu.com/ubuntu/ trusty-updates/main py
thon-html5lib all 0.999-3~ubuntu1 [83.5 kB]
Get:4 http://jp.archive.ubuntu.com/ubuntu/ trusty/main python-set
uptools all 3.3-1ubuntu1 [230 kB]
Get:5 http://jp.archive.ubuntu.com/ubuntu/ trusty-updates/univers
```



W14-2: pip確認

インストール作業完了後に、「pip -h」でマニュアルを表示。pipもapt-getと同様にPythonパッケージのインストール(install)とアンインストール(uninstall)ができるようだ。

```
File Edit View Search Terminal Help
Selecting previously unselected package python-wheel.
Preparing to unpack .../python-wheel_0.24.0-1~ubuntu1_all.deb ...
Unpacking python-wheel (0.24.0-1~ubuntu1) ...
Processing triggers for man-db (2.6.7.1-1) ...
Setting up python-colorama (0.2.5-0.1ubuntu2) ...
Setting up python-distlib (0.1.8-1ubuntu1) ...
Setting up python-html5lib (0.999-3~ubuntu1) ...
Setting up python-setuptools (3.3-1ubuntu1) ...
Setting up python-pip (1.5.4-1ubuntu1) ...
Setting up python-wheel (0.24.0-1~ubuntu1) ...
iu@bielinux[Desktop] pip -h [ 1:17午後 ]

Usage:
  pip <command> [options]

Commands:
  install          Install packages.
  uninstall        Uninstall packages.
  freeze           Output installed packages in requirem
  format.
  list             List installed packages.
  show            Show information about installed pa
  ckages.
  search          Search PyPI for packages.
  wheel           Build wheels from your requirements
```

W14-3: cutadapt

①pipコマンドを用いてアダプター配列除去プログラムcutadaptをインストール。2015年9月18日に別のPCでこの作業を行うとエラーが出ました。私の場合は、この後のW14-4とW14-5をやってから、再度①を実行すると何事もなかったかのようにうまくいきました。

```
iu@bielinux[Desktop] pwd
/home/iu/Desktop
iu@bielinux[Desktop] where pip
/usr/bin/pip
iu@bielinux[Desktop] sudo pip install cutadapt [ 1:42午後 ]
Downloading/unpacking cutadapt
  Downloading cutadapt-1.8.1.tar.gz (484kB): 484kB downloaded
  Running setup.py (path:/tmp/pip_build_root/cutadapt/setup.py) e
gg_info for package cutadapt
  At least one C source file is missing or out of date.
  Compiling cutadapt/_align.pyx because it changed.
  Cythonizing cutadapt/_align.pyx

Installing collected packages: cutadapt
  Running setup.py install for cutadapt
    building 'cutadapt._align' extension
      x86_64-linux-gnu-gcc -pthread -fno-strict-aliasing -DNDEBUG -
g -fwrapv -O2 -Wall -Wstrict-prototypes -fPIC -I/usr/include/pyth
on2.7 -c cutadapt/_align.c -o build/temp.linux-x86_64-2.7/cutadap
t/_align.o
      x86_64-linux-gnu-gcc -pthread -shared -Wl,-O1 -Wl,-Bsymbolic-
functions -Wl,-Bsymbolic-functions -Wl,-z,relro -fno-strict-alias
ing -DNDEBUG -g -fwrapv -O2 -Wall -Wstrict-prototypes -D_FORTIFY_
SOURCE=2 -g -fstack-protector --param=ssp-buffer-size=4 -Wformat
-Werror=format-security build/temp.linux-x86_64-2.7/cutadapt/_ali
```



W14-3: cutadapt

無事インストール完了。「cutadapt -v」
でバージョン情報を把握。ver. 1.8.1で
あることがわかる。

```
functions -Wl,-Bsymbolic-functions -Wl,-z,relro -fno-strict-aliasing -DNDEBUG -g -fwrapv -O2 -Wall -Wstrict-prototypes -D_FORTIFY_SOURCE=2 -g -fstack-protector --param=ssp-buffer-size=4 -Wformat -Werror=format-security build/temp.linux-x86_64-2.7/cutadapt/_seqio.o -o build/lib.linux-x86_64-2.7/cutadapt/_seqio.so
changing mode of build/scripts-2.7/cutadapt from 644 to 755

changing mode of /usr/local/bin/cutadapt to 755
Successfully installed cutadapt
Cleaning up...
iu@bielinux[Desktop] cutadapt -v [ 1:42午後 ]
cutadapt version 1.8.1
Copyright (C) 2010-2015 Marcel Martin <marcel.martin@scilifelab.se>

cutadapt removes adapter sequences from high-throughput sequencing reads.

Usage:
  cutadapt -a ADAPTER [options] [-o output.fastq] input.fastq

For paired-end reads:
  cutadapt -a ADAPT1 -A ADAPT2 [options] -o out1.fastq -p out2.fastq in1.fastq in2.fastq
```

W14-3: cutadapt

しかしよく見ると「-vというオプションはない」と文句を言われている。「cutadapt --help」と打ち込んで正しいオプション名を把握するのが王道だが、経験的に-vでだめなら--versionでよいと分かっている。

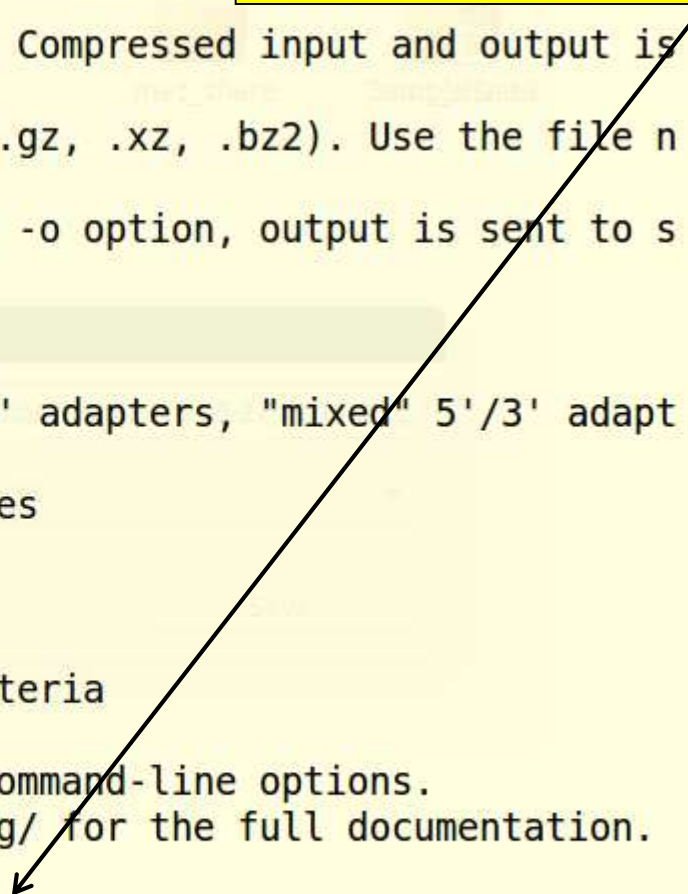
```
File Edit View Search Terminal Help
adapter will be removed.

Input may also be in FASTA format. Compressed input and output is
supported and
auto-detected from the file name (.gz, .xz, .bz2). Use the file n
ame '-' for
standard input/output. Without the -o option, output is sent to s
tandard output.

Some other available features are:
* Various other adapter types (5' adapters, "mixed" 5'/3' adapt
ers etc.)
* Trimming a fixed number of bases
* Quality trimming
* Trimming paired-end reads
* Trimming colorspace reads
* Filtering reads by various criteria

Use "cutadapt --help" to see all command-line options.
See http://cutadapt.readthedocs.org/ for the full documentation.

cutadapt: error: no such option: -v
iu@bielinux[Desktop] cutadapt --version [ 1:46午後 ]
1.8.1
iu@bielinux[Desktop] [ 1:56午後 ]
```



W14-4: HTSeq

PyPIのHTSeqパッケージのウェブページから①作者のページに飛んでいる。②Prerequisite and installationへ。

python™ search

» Package Index > HTSeq > 0.6.1

PACKAGE INDEX >>

Browse packages
Package submission
List trove classifiers
List packages
RSS (latest 40 updates)
RSS (newest 40 packages)
Python 3 Packages
PyPI Tutorial
PyPI Security
PyPI Support
PyPI Bug Reports
PyPI Discussion
PyPI Developer Info

ABOUT >>
NEWS >>
DOCUMENTATION >>
DOWNLOAD >>
COMMUNITY >>
FOUNDATION >>
CORE DEVELOPMENT >>

HTSeq 0.6.1

A framework to process and analyze data from high-throughput sequencing (HTS) assays

Downloads ↓

Not Logged In
Login
Register
Lost Login?
Use OpenID & Ip

For instructions, see:

① <http://www-huber.embl.de/users/anders/HTSeq/>

File	Type
HTSeq-0.6.1.tar.gz (md5)	Source
HTSeq-0.6.1.win32-py2.7.exe (md5)	MS Windows ins
HTSeq-0.6.1p1.tar.gz (md5)	Source

Downloads (All Versions):
13229 downloads in the last day
30649 downloads in the last week
36180 downloads in the last month

Author: Simon Anders
Home Page: <http://www-huber.embl.de/users/anders/>
Requires: python (>=2.5, <3.0), numpy
Categories:
[Development Status :: 5 - Production/Stable](#)
[Intended Audience :: Developers](#)
[Intended Audience :: Science/Research](#)
[License :: OSI Approved :: GNU General Public License](#)
[Operating System :: POSIX](#)
[Programming Language :: Python](#)
[Topic :: Scientific/Engineering :: Bio-Informatics](#)

Package Index Owner: sanders, quetzcoatl
Package Index Maintainer: quetzcoatl
DOAP record: [HTSeq-0.6.1.xml](#)

HTSeq 0.6.1p2 documentation » previous | next | index

Table Of Contents

HTSeq: Analysing high-throughput sequencing data with Python

- Paper
- Documentation overview
- Author
- License

Previous topic

HTSeq: Analysing high-throughput sequencing data with Python

Next topic

Prerequisites and installation

This Page

Show Source

Quick search

Enter search terms or a module, class or function name.

HTSeq: Analysing high-throughput sequencing data with Python

HTSeq is a Python package that provides infrastructure to process data from high-throughput sequencing assays.

- Please see the chapter *A tour through HTSeq* first for an overview on the kind of analysis you can do with HTSeq and the design of the package, and then look at the reference documentation.
- While the main purpose of HTSeq is to allow you to write your own analysis scripts, customized to your needs, there are also a couple of stand-alone scripts for common tasks that can be used without any Python knowledge. See the *Scripts* section in the overview below for what is available.
- For downloads and installation instructions, see *Prerequisites and installation*.

Paper

HTSeq is described in the following publication:

Simon Anders, Paul Theodor Pyl, Wolfgang Huber
HTSeq — A Python framework to work with high-throughput sequencing data
Bioinformatics (2014), in print, online at doi:10.1093/bioinformatics/btu638

If you use HTSeq in research, please cite this paper in your publication.

Documentation overview

- Prerequisites and installation*

Download links and installation instructions can be found here

W14-4: HTSeq

Prerequisites and installation

HTSeq is available from the Python Package Index (PyPI). Download HTSeq from the

[HTSeq package download page on PyPI](#),

where you will find various version of the package.

To use HTSeq, you need at least version 2.5 of Python (Python 3 does not work yet), together with NumPy, a commonly used Python package for numerical calculations, and matplotlib, a plotting library.

HTSeq can be installed like any other Python package. For users unfamiliar with this, more detailed instructions are given below.

Installation on Linux

Make sure that you the standard GNU build environment installed, as well as Python together with its development headers and numpy and matplotlib. Users of Ubuntu Linux simply type:

```
sudo apt-get install build-essential python2.7-dev python-numpy python-matplotlib
```

This command installs the required packages as listed (or simply does nothing if they are already installed).

For users of RedHat or RedHat-derived distros (Fedora, CentOS), the equivalent command seems to be (untested):

```
sudo yum groupinstall "Development Tools"  
sudo yum install python-devel numpy python-matplotlib
```

To install HTSeq itself, download the *source* package from the HTSeq PyPI page, unpack the tarball, go into the directory with the unpacked files and type there

```
python setup.py install --user
```

to install HTSeq for the user currently logged in. To make HTSeq available to all users, use instead:

```
python setup.py build  
sudo python setup.py install
```

To test the installation, change to another director than the build directory, start Python (by typing `python`) and then try whether typing `import HTSeq` causes an error meesage.

基本的にはここに書いてある手順通りに進めていく。赤枠と青枠は、連載第1回で述べたようにLinuxには様々な種類(ディストリビューション)がある。apt-getなど知っているコマンドを頼りに赤下線の情報でBio-Linux 8がUbuntuベースのものであることを思い出し、HTSeqをインストールする前段階に必要な環境(Prerequisites)を整えておく。要するに赤矢印部分のコマンドをコピーで実行すればいい、ということ。

W14-4: HTSeq

コピペ。OS間でコピペがうまくいかない場合は、ゲストOS (Bio-Linux 8) 内のFirefoxを起動をして行えば大抵うまくいく。作業ディレクトリはどこでもよい。

Installation on Linux

Make sure that you the standard GNU build environment installed, as well as Python together with its development headers and numpy and matplotlib. Users of Ubuntu Linux simply type:

```
sudo apt-get install build-essential python2.7-dev python-numpy python-matplotlib
```

This command installs the required

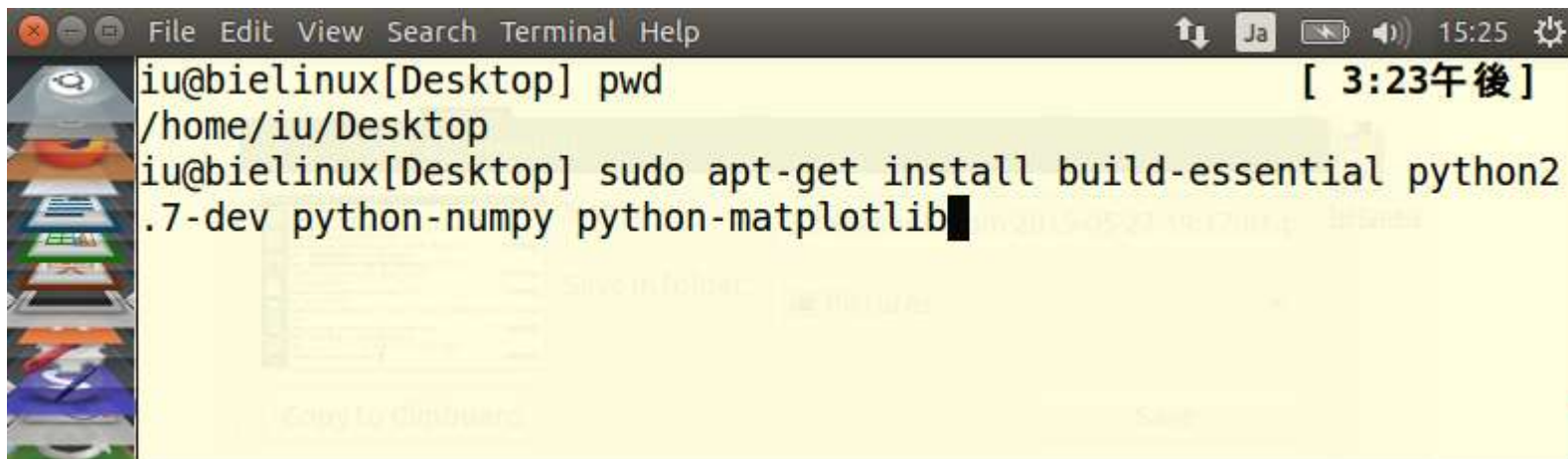
For users of RedHat or RedHat-d
(untested):

```
sudo yum groupinstall "Dev  
sudo yum install python-de
```



thing if they are already installed.

the equivalent command seems to be



W14-4: HTSeq

コピー後の全貌。特にエラーも出ていないので、HTSeq本体をインストールできる状況になった。

```
File Edit View Search Terminal Help 15:28
iu@bielinux[Desktop] sudo apt-get install build-essential python2
.7-dev python-numpy python-matplotlib
[sudo] password for iu:
Reading package lists... Done
Building dependency tree
Reading state information... Done
build-essential is already the newest version.
build-essential set to manually installed.
python2.7-dev is already the newest version.
python2.7-dev set to manually installed.
python-matplotlib is already the newest version.
python-matplotlib set to manually installed.
Suggested packages:
  python-nose python-numpy-dbg python-numpy-doc
The following packages will be upgraded:
  python-numpy
1 upgraded, 0 newly installed, 0 to remove and 693 not upgraded.
Need to get 1,603 kB of archives.
After this operation, 2,048 B of additional disk space will be us
ed.
Do you want to continue? [Y/n] y
Get:1 http://jp.archive.ubuntu.com/ubuntu/ trusty-updates/main py
thon-numpy amd64 1:1.8.2-0ubuntu0.1 [1,603 kB]
Fetched 1,603 kB in 0s (4,632 kB/s)
(Reading database ... 314497 files and directories currently inst
alled.)
Preparing to unpack .../python-numpy_1%3a1.8.2-0ubuntu0.1_amd64.d
eb ...
Unpacking python-numpy (1:1.8.2-0ubuntu0.1) over (1:1.8.1-1ubuntu
1) ...
Processing triggers for man-db (2.6.7.1-1) ...
Setting up python-numpy (1:1.8.2-0ubuntu0.1) ...
iu@bielinux[Desktop] [ 3:27午後 ]
```

W14-5: HTSeq

Prerequisites and installation

HTSeq is available from the [Python Package Index \(PyPI\)](#). Download HTSeq from the

[HTSeq package download page on PyPI](#),

where you will find various version of the package.

To use HTSeq, you need at least version 2.5 of [Python](#) (Python 3 does not work yet), together with [NumPy](#), a commonly used Python package for numerical calculations, and [matplotlib](#), a plotting library.

HTSeq can be installed like any other Python package. For users unfamiliar with this, more detailed instructions are given below.

Installation on Linux

Make sure that you the standard GNU build environment installed, as well as Python together with its development headers and numpy and matplotlib. Users of Ubuntu Linux simply type:

```
sudo apt-get install build-essential python2.7-dev python-numpy python-matplotlib
```

This command installs the required packages as listed (or simply does nothing if they are already installed).

For users of RedHat or RedHat-derived distros (Fedora, CentOS), the equivalent command seems to be (untested):

```
sudo yum groupinstall "Development Tools"  
sudo yum install python-devel numpy python-matplotlib
```

To install HTSeq itself, download the *source* package from the [HTSeq PyPI page](#), unpack the tarball, go into the directory with the unpacked files and type there

```
python setup.py install --user
```

to install HTSeq for the user currently logged in. To make HTSeq available to all users, use instead:

```
python setup.py build  
sudo python setup.py install
```

To test the installation, change to another director than the build directory, start Python (by typing `python`) and then try whether typing `import HTSeq` causes an error meesage.

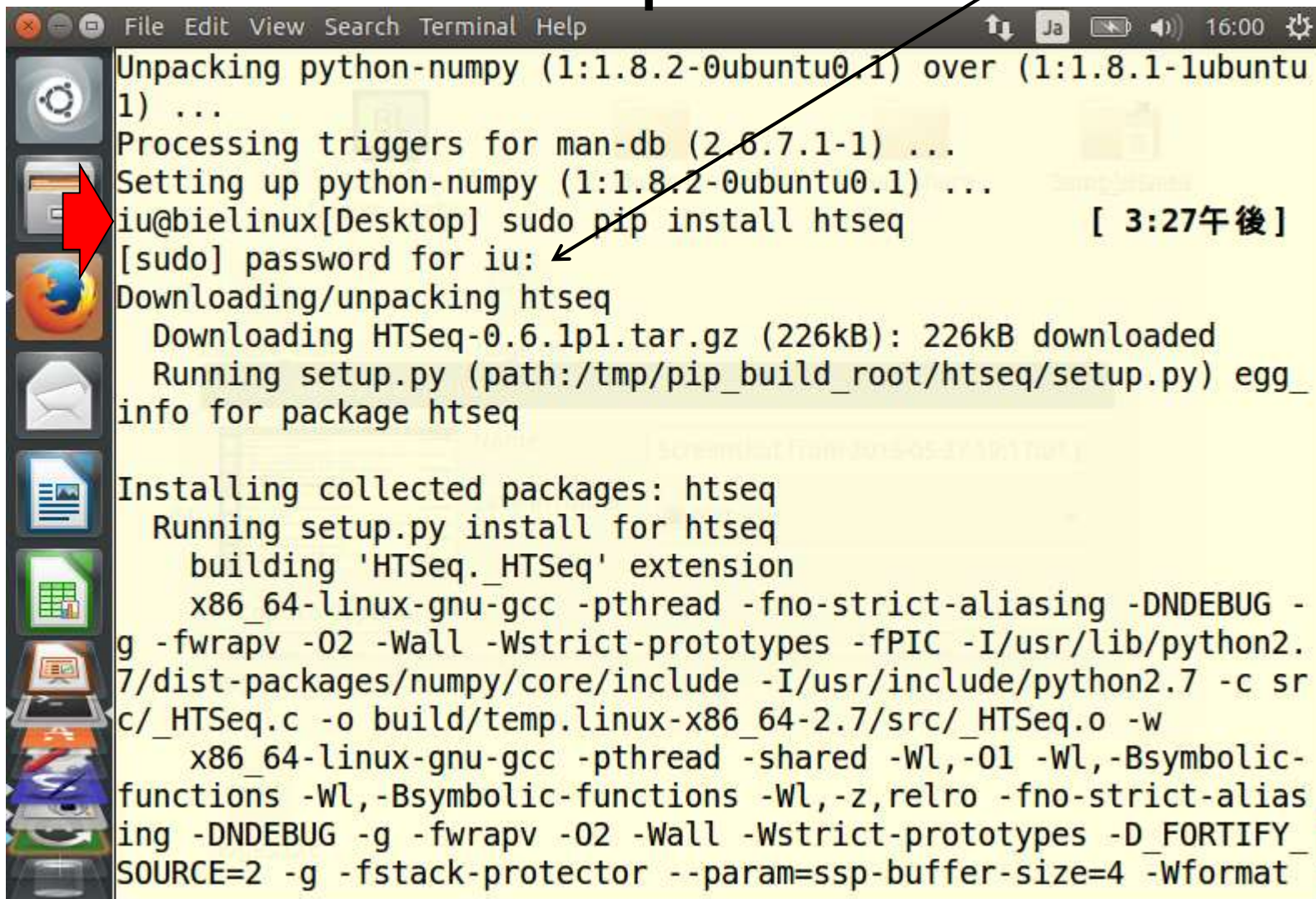
①が現在のログインユーザ(この場合iu)のみにHTSeqをインストールするやり方。本稿でsudoコマンドを用いていることから想像できるように、②が全ユーザが使えるようにHTSeqをインストールするやり方。尚、両方ともpipコマンドを利用しないやり方。インストール手順書の書き方は開発者次第である。読み手側のリテラシーが問われる場合が多い。

①

②

W14-5: HTSeq

HTSeqパッケージ本体のインストールは「sudo pip install htseq」。



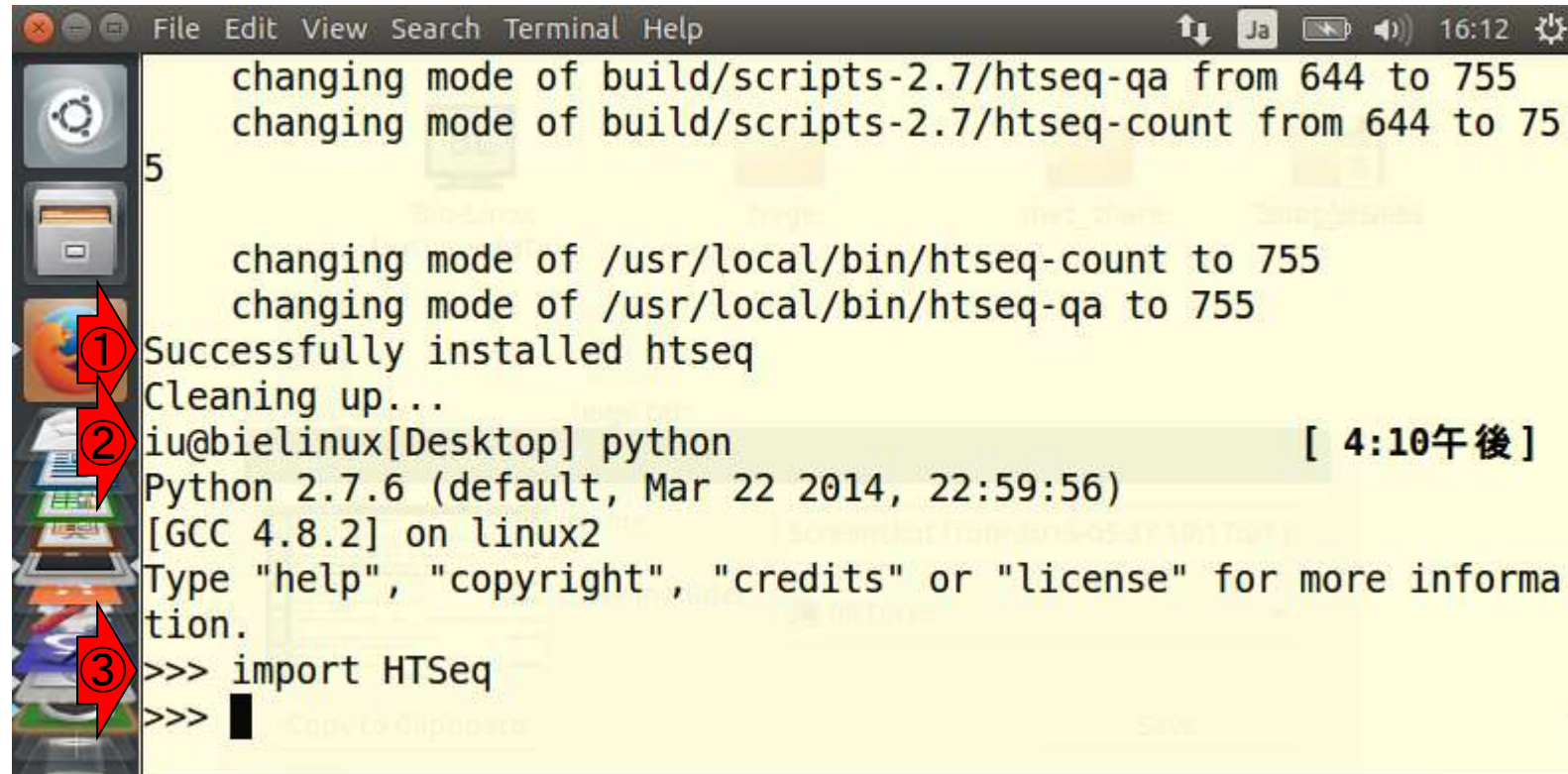
```
File Edit View Search Terminal Help 16:00
Unpacking python-numpy (1:1.8.2-0ubuntu0.1) over (1:1.8.1-1ubuntu1) ...
Processing triggers for man-db (2.6.7.1-1) ...
Setting up python-numpy (1:1.8.2-0ubuntu0.1) ...
iu@bielinux[Desktop] sudo pip install htseq [ 3:27午後 ]
[sudo] password for iu:
Downloading/unpacking htseq
  Downloading HTSeq-0.6.1p1.tar.gz (226kB): 226kB downloaded
  Running setup.py (path:/tmp/pip_build_root/htseq/setup.py) egg_info for package htseq

Installing collected packages: htseq
  Running setup.py install for htseq
    building 'HTSeq._HTSeq' extension
    x86_64-linux-gnu-gcc -pthread -fno-strict-aliasing -DNDEBUG -g -fwrapv -O2 -Wall -Wstrict-prototypes -fPIC -I/usr/lib/python2.7/dist-packages/numpy/core/include -I/usr/include/python2.7 -c src/_HTSeq.c -o build/temp.linux-x86_64-2.7/src/_HTSeq.o -w
    x86_64-linux-gnu-gcc -pthread -shared -Wl,-O1 -Wl,-Bsymbolic-functions -Wl,-Bsymbolic-functions -Wl,-z,relro -fno-strict-aliasing -DNDEBUG -g -fwrapv -O2 -Wall -Wstrict-prototypes -D_FORTIFY_SOURCE=2 -g -fstack-protector --param=ssp-buffer-size=4 -Wformat
```

W14-5: HTSeq

①無事インストール完了していることがわかるが、念のため、HTSeqのウェブページに書かれている手順通りに②と③を実行してエラーが出ないことを確認している。

To test the installation, change to another director than the build directory, start Python (by typing `python`) and then try whether typing `import HTSeq` causes an error message.



The image shows a terminal window with the following output:

```
File Edit View Search Terminal Help
changing mode of build/scripts-2.7/htseq-qa from 644 to 755
changing mode of build/scripts-2.7/htseq-count from 644 to 755
5
changing mode of /usr/local/bin/htseq-count to 755
changing mode of /usr/local/bin/htseq-qa to 755
① Successfully installed htseq
Cleaning up...
② iu@bielinux[Desktop] python [ 4:10午後 ]
Python 2.7.6 (default, Mar 22 2014, 22:59:56)
[GCC 4.8.2] on linux2
Type "help", "copyright", "credits" or "license" for more information.
③ >>> import HTSeq
>>> █
```

Red arrows with circled numbers 1, 2, and 3 point to the corresponding lines in the terminal output.

W14-5: HTSeq

To test the installation, change to another director than the build director then try whether typing `import HTSeq` causes an error message.

Pythonの終了の仕方は特に記述されていない。①とりあえずqと打ってみたがダメっぽいので、「Python 終了コマンド」でウェブ検索してexit()またはCTRL+Dであることを知った。②ここではexit()と打っている。

```
File Edit View Search Terminal Help 16:18
changing mode of build/scripts-2.7/htseq-qa from 644 to 755
changing mode of build/scripts-2.7/htseq-count from 644 to 755
5
changing mode of /usr/local/bin/htseq-count to 755
changing mode of /usr/local/bin/htseq-qa to 755
Successfully installed htseq
Cleaning up...
iu@bielinux[Desktop] python [ 4:10午後 ]
Python 2.7.6 (default, Mar 22 2014, 22:59:56)
[GCC 4.8.2] on linux2
Type "help", "copyright", "credits" or "license" for more information.
>>> import HTSeq
>>> q
Traceback (most recent call last):
  File "<stdin>", line 1, in <module>
NameError: name 'q' is not defined
>>> exit()
iu@bielinux[Desktop] [ 4:18午後 ]
```



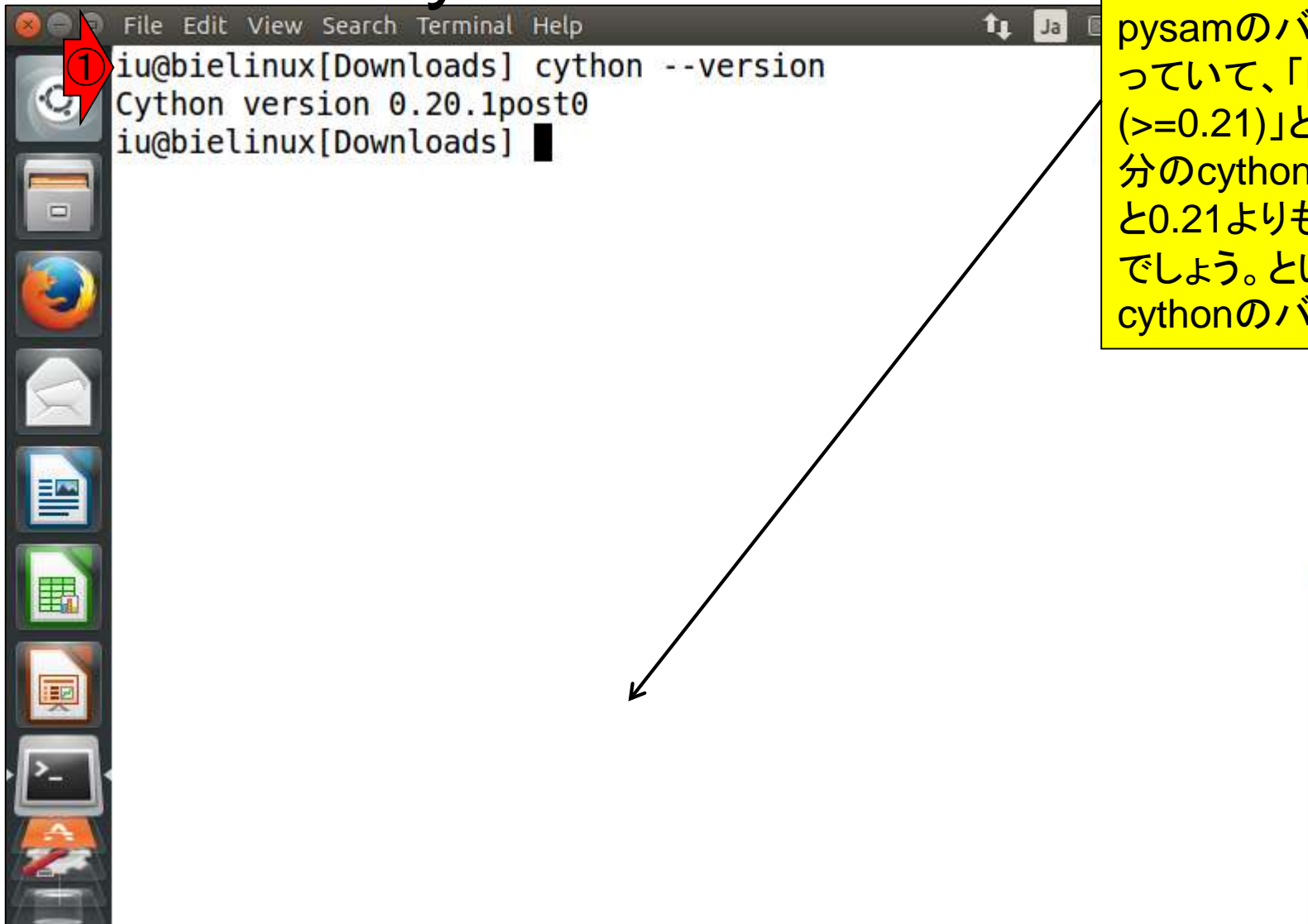
W14-6: pysam

pysamパッケージのインストールは「sudo pip install pysam」。数分程度かかるが、無事インストール完了していることが分かる。このパッケージはBAM形式ファイルを読み込む際に必要なものらしい。

```
linux-x86_64-2.7/htslib/cram/cram_encode.o build/temp.l
-2.7/htslib/cram/pooled_alloc.o build/temp.linux-x86_6
b/cram/string_alloc.o build/temp.linux-x86_64-2.7/htsl
_header.o build/temp.linux-x86_64-2.7/htslib/cram/cram_codecs.o b
uild/temp.linux-x86_64-2.7/htslib/cram/cram_decode.o build/temp.l
linux-x86_64-2.7/htslib/cram/md5.o build/temp.linux-x86_64-2.7/hts
lib/cram/rANS_static.o build/temp.linux-x86_64-2.7/htslib/cram/cr
am_io.o build/temp.linux-x86_64-2.7/htslib/cram/cram_samtools.o b
uild/temp.linux-x86_64-2.7/htslib/cram/cram_stats.o build/temp.li
linux-x86_64-2.7/htslib/cram/mFILE.o build/temp.linux-x86_64-2.7/ht
slib/cram/open_trace_file.o build/temp.linux-x86_64-2.7/htslib/cr
am/cram_index.o build/temp.linux-x86_64-2.7/htslib/cram/files.o b
uild/temp.linux-x86_64-2.7/htslib/cram/vlen.o -Lpysam -lz -o buil
d/lib.linux-x86_64-2.7/pysam/cfaidx.so

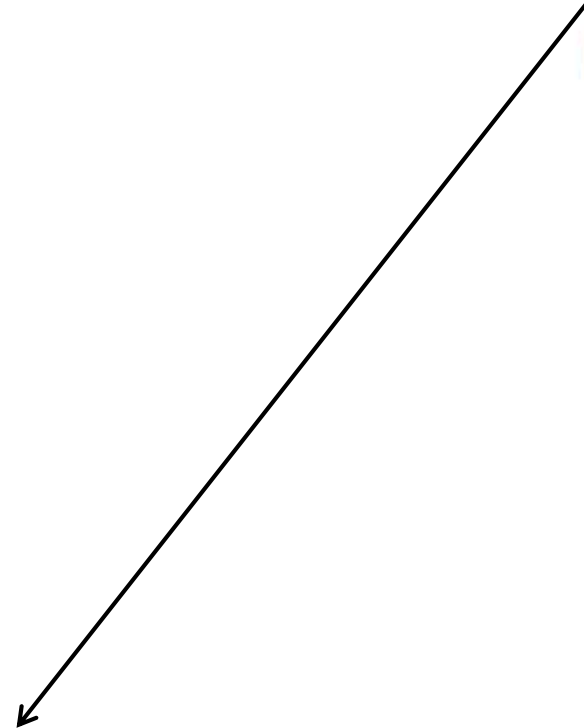
warning: no files found matching 'tests/pysam_data'
warning: no files found matching 'tests/tabix_data'
Successfully installed pysam
Cleaning up...
iu@bielinux[Desktop] [ 4:27午後]
```

W14-7: cython



```
File Edit View Search Terminal Help
iu@bielinux[Downloads] cython --version
Cython version 0.20.1post0
iu@bielinux[Downloads] █
```

2015年9月18日に別のPCで「`sudo pip install pysam`」を実行しようとしたらエラーが出ました。pysamのバージョンが0.8.3に上がっていて、「Requires cython (>=0.21)」と書かれていました。自分のcythonのバージョンを調べると0.21よりも低いのでそれが原因でしょう。というわけで、まずはcythonのバージョンを上げます。



W14-7: cython

①「`sudo pip install cython`」を実行。しかし既にインストールされているので、`--upgrade`をつけるように指示されている。この場合は②のように`--upgrade`オプションをつける。③`ver. 0.23.2`がインストールされるようで`pysam`の要求は満たせるので安心。

```
iu@bielinux[Downloads] cython --version
Cython version 0.20.1post0
① iu@bielinux[Downloads] sudo pip install cython
[sudo] password for iu:
Requirement already satisfied (use --upgrade to upgrade): cython in
/usr/lib/python2.7/dist-packages
Cleaning up...
② iu@bielinux[Downloads] sudo pip install --upgrade cython
Downloading/unpacking cython from https://pypi.python.org/packages
/source/C/Cython/Cython-0.23.2.tar.gz#md5=188e18c826dae4514ee06281
44df6c3e
      ③ Downloading Cython-0.23.2.tar.gz (1.6MB): 18% 307kB
```


W14-7: cython

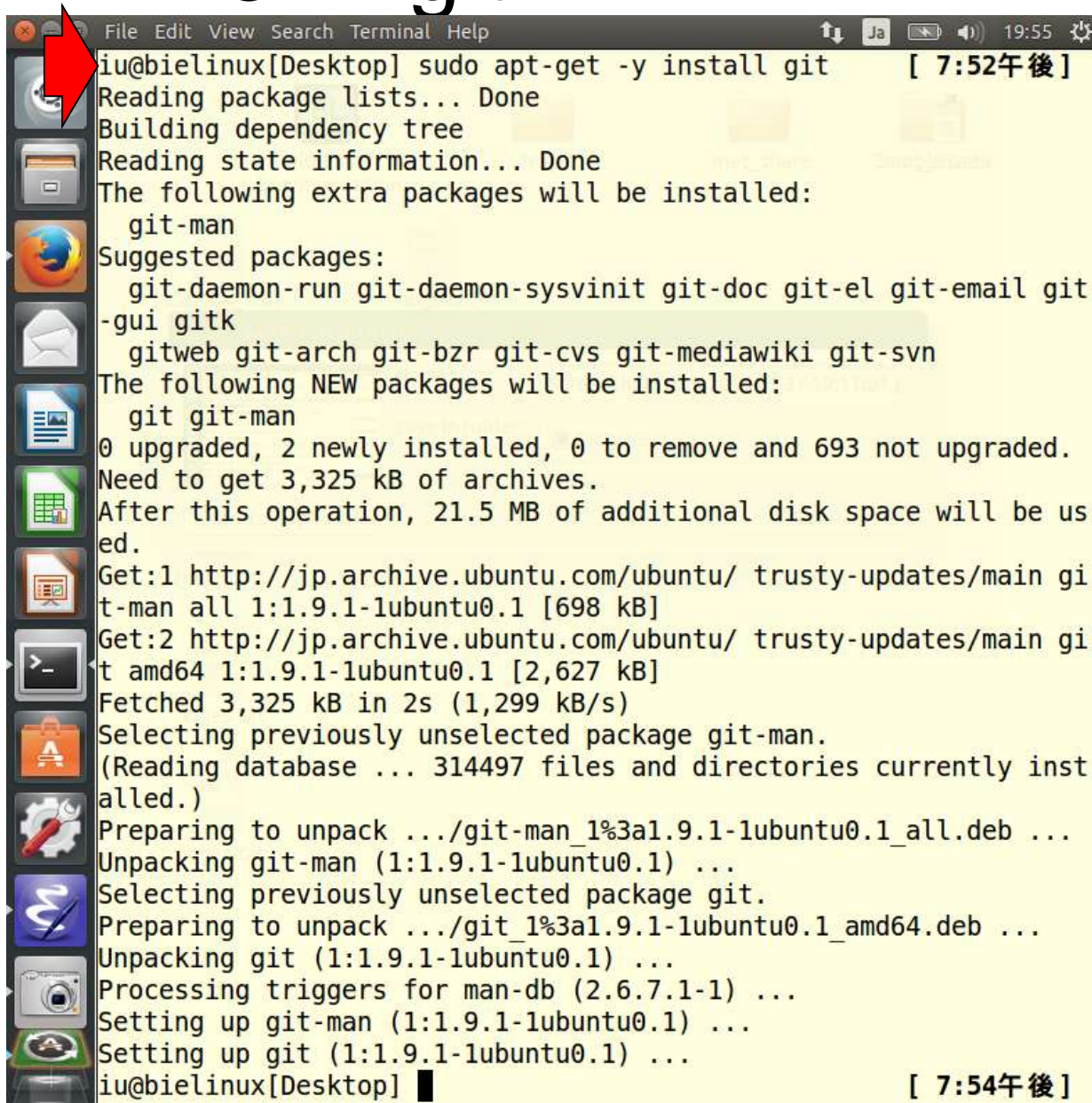
しかし、このようなエラーが出てインストールに失敗した。とりあえず放置(爆)。

```
3, in _download_url
    for chunk in resp_read(4096):
        File "/usr/lib/python2.7/dist-packages/pip/download.py", line 42
1, in resp_read
    chunk_size, decode_content=False):
    File "/usr/share/python-wheels/urllib3-1.7.1-py2.py3-none-any.wh
l/urllib3/response.py", line 225, in stream
        data = self.read(amt=amt, decode_content=decode_content)
    File "/usr/share/python-wheels/urllib3-1.7.1-py2.py3-none-any.wh
l/urllib3/response.py", line 174, in read
        data = self._fp.read(amt)
    File "/usr/lib/python2.7/httplib.py", line 573, in read
        s = self.fp.read(amt)
    File "/usr/lib/python2.7/socket.py", line 380, in read
        data = self._sock.recv(left)
    File "/usr/lib/python2.7/ssl.py", line 341, in recv
        return self.read(buflen)
    File "/usr/lib/python2.7/ssl.py", line 260, in read
        return self._sslobj.read(len)
SSLError: The read operation timed out

Storing debug log for failure in /home/iu/.pip/pip.log
iu@bielinux[Downloads] [ 5:50午後 ]
```

W15-1 : git

gitのインストール。-yオプションをつけることで「yes or no?」で毎回yと打つ手間を省いている。



```
iu@bielinux[Desktop] sudo apt-get -y install git [ 7:52午後 ]
Reading package lists... Done
Building dependency tree
Reading state information... Done
The following extra packages will be installed:
  git-man
Suggested packages:
  git-daemon-run git-daemon-sysvinit git-doc git-el git-email git
-gui gitk
  gitweb git-arch git-bzr git-cvs git-mediawiki git-svn
The following NEW packages will be installed:
  git git-man
0 upgraded, 2 newly installed, 0 to remove and 693 not upgraded.
Need to get 3,325 kB of archives.
After this operation, 21.5 MB of additional disk space will be us
ed.
Get:1 http://jp.archive.ubuntu.com/ubuntu/ trusty-updates/main gi
t-man all 1:1.9.1-1ubuntu0.1 [698 kB]
Get:2 http://jp.archive.ubuntu.com/ubuntu/ trusty-updates/main gi
t amd64 1:1.9.1-1ubuntu0.1 [2,627 kB]
Fetched 3,325 kB in 2s (1,299 kB/s)
Selecting previously unselected package git-man.
(Reading database ... 314497 files and directories currently inst
alled.)
Preparing to unpack .../git-man_1%3a1.9.1-1ubuntu0.1_all.deb ...
Unpacking git-man (1:1.9.1-1ubuntu0.1) ...
Selecting previously unselected package git.
Preparing to unpack .../git_1%3a1.9.1-1ubuntu0.1_amd64.deb ...
Unpacking git (1:1.9.1-1ubuntu0.1) ...
Processing triggers for man-db (2.6.7.1-1) ...
Setting up git-man (1:1.9.1-1ubuntu0.1) ...
Setting up git (1:1.9.1-1ubuntu0.1) ...
iu@bielinux[Desktop] [ 7:54午後 ]
```

W15-2: FaQCs

<https://github.com/LANL-Bioinformatics/FaQCs>

LANL-Bioinformatics / FaQCs
forked from chienchi/FaQCs

👁 Watch 5
★ Star 2
🍴 Fork 2

📄 45 commits
🌿 1 branch
📦 1 release
👤 1 contributor

↶
branch: **master** ▾
FaQCs / +
☰

This branch is 23 commits ahead, 15 commits behind chienchi:master
Pull Request Compare

fix typo

chienchi authored on Mar 31 latest commit 6aed8354ec 📄

📁 example	adjust margin of pdf plots	3 months ago
📁 galaxy_module	Update FaQCs.xml	6 months ago
📁 lib	Add option "-5trim_off <bool> Turn off trimming from 5'end."	4 months ago
📄 COPYRIGHT	add LICENSE	a year ago
📄 FaQCs.pl	patch for jellyfish v2	3 months ago
📄 LICENSE	add LICENSE	a year ago
📄 README.md	fix typo	2 months ago

📄 README.md

FaQCs: Quality Control of Next
Generation Sequencing Data

↔ Code

🔔 Issues 1

🔗 Pull requests 0

📡 Pulse

📊 Graphs

HTTPS clone URL

https://github.com/l
📄

You can clone with [HTTPS](#) or [Subversion](#).

🖨 Clone in Desktop

📦 Download ZIP

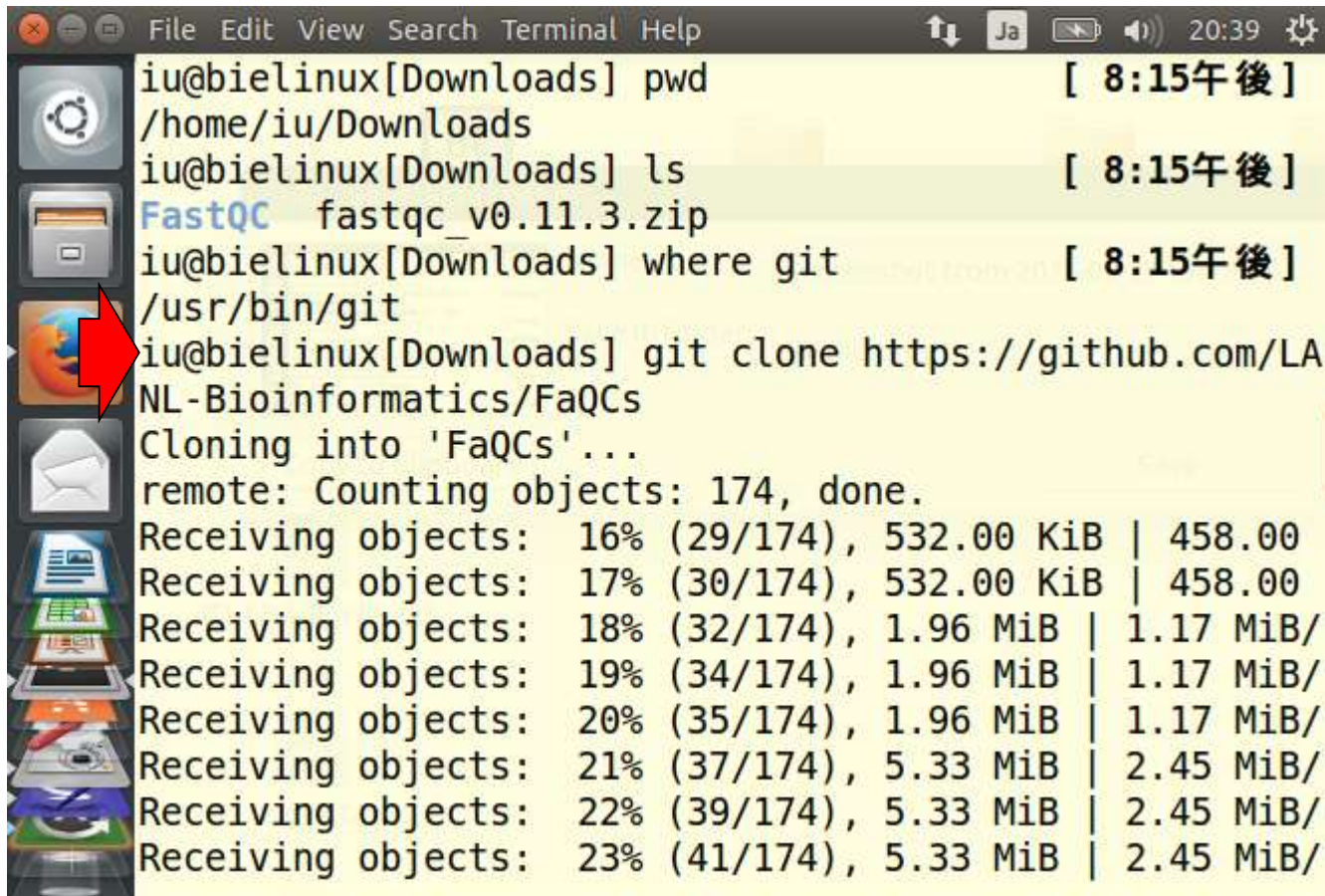
W15-2: FaQCs

「git clone 欲しいプログラムのURL」が基本形。
「git clone 欲しいプログラムのURL.git」としてもよい。後者の記述形式のほうが一般的らしい。

A terminal window with a dark title bar containing 'File Edit View Search Terminal Help', 'Ja', and '20:15'. The terminal text shows a user at 'iu@bielinux[Downloads]' running 'pwd' (output: '/home/iu/Downloads'), 'ls' (output: 'FastQC fastqc_v0.11.3.zip'), and 'where git' (output: '/usr/bin/git'). A red arrow points to the final command: 'git clone https://github.com/LA-NL-Bioinformatics/FaQCs'.

```
iu@bielinux[Downloads] pwd [ 8:15午後 ]
/home/iu/Downloads
iu@bielinux[Downloads] ls [ 8:15午後 ]
FastQC fastqc_v0.11.3.zip
iu@bielinux[Downloads] where git [ 8:15午後 ]
/usr/bin/git
iu@bielinux[Downloads] git clone https://github.com/LA
NL-Bioinformatics/FaQCs
```

W15-2: FaQCs



```
iu@bielinux[Downloads] pwd [ 8:15午後 ]
/home/iu/Downloads
iu@bielinux[Downloads] ls [ 8:15午後 ]
FastQC fastqc_v0.11.3.zip
iu@bielinux[Downloads] where git [ 8:15午後 ]
/usr/bin/git
iu@bielinux[Downloads] git clone https://github.com/LA
NL-Bioinformatics/FaQCs
Cloning into 'FaQCs'...
remote: Counting objects: 174, done.
Receiving objects: 16% (29/174), 532.00 KiB | 458.00
Receiving objects: 17% (30/174), 532.00 KiB | 458.00
Receiving objects: 18% (32/174), 1.96 MiB | 1.17 MiB/
Receiving objects: 19% (34/174), 1.96 MiB | 1.17 MiB/
Receiving objects: 20% (35/174), 1.96 MiB | 1.17 MiB/
Receiving objects: 21% (37/174), 5.33 MiB | 2.45 MiB/
Receiving objects: 22% (39/174), 5.33 MiB | 2.45 MiB/
Receiving objects: 23% (41/174), 5.33 MiB | 2.45 MiB/
```

W15-2: FaQCs

1分足らずでダウンロードが終了する。「git clone」自体はインストールを行うものではなくwgetコマンドのようなもの。(FastQCと名前が似ていてややこしいが)確かに目的のFaQCsというディレクトリが作成されている

```
File Edit View Search Terminal Help 20:41
Receiving objects: 94% (164/174), 5.33 MiB | 2.45 MiB
Receiving objects: 95% (166/174), 5.33 MiB | 2.45 MiB
Receiving objects: 96% (168/174), 5.33 MiB | 2.45 MiB
Receiving objects: 97% (169/174), 5.33 MiB | 2.45 MiB
Receiving objects: 98% (171/174), 5.33 MiB | 2.45 MiB
Receiving objects: 99% (173/174), 5.33 MiB | 2.45 MiB
Receiving objects: 100% (174/174), 5.33 MiB | 2.45 MiB
Receiving objects: 100% (174/174), 7.26 MiB | 2.45 MiB
/s, done.
Resolving deltas: 100% (82/82), done.
Checking connectivity... done.
iu@bielinux[Downloads] ls -lh [ 8:16午後 ]
total 9.6M
drwxrwxr-x 6 iu iu 4.0K 5月 28 20:16 FaQCs
drwxrwxr-x 8 iu iu 4.0K 5月 7 14:46 FastQC
-rw-rw-r-- 1 iu iu 9.6M 3月 25 18:40 fastqc_v0.11.3.z
ip
iu@bielinux[Downloads] [ 8:40午後 ]
```

[8:16午後]

[8:40午後]



W15-3: FaQCs

GitHub | This repository | Search | Explore | Features | Enterprise | Blog | Sign up | Sign in

LANL-Bioinformatics / FaQCs
forked from chienchi/FaQCs

45 commits | 1 branch | 1 release | 1 contributor

branch: master - FaQCs / +

This branch is 22 commits ahead, 15 commits behind chienchi/master | Pull Request

fix type

chienchi authored on Mar 31 | latest commit: faed855

- example: adjust margin of pdf plots (3 months ago)
- galaxy_module: Update FaQCs.yml (6 months ago)
- lib: Add option "-trim_off <bool>" Turn off trimming from 5' end. (4 months ago)
- COPYRIGHT: add LICENSE (8 years ago)
- FaQCs.pl: patch for jellyfish v2 (3 months ago)
- LICENSE: add LICENSE (8 years ago)
- README.md: fix typo (2 months ago)

README.md

FaQCs: Quality Control of Next Generation Sequencing Data

Quality 3D plot. (Position vs. Score vs. Frequency)

PREREQUISITES

1. The main program is developed in Perl v 5.8.8.
2. Parallel::ForkManager module from CPAN
(<http://search.cpan.org/~dlux/Parallel-ForkManager-0.7.9/lib/Parallel/ForkManager.pm>)
3. String::Approx module from CPAN
(<http://search.cpan.org/~jhi/String-Approx-3.27/Approx.pm>)
4. R for plotting
(<http://www.r-project.org/>)
5. Jellyfish for kmer counting (Optional) (<http://www.cbcu.umd.edu/software/jellyfish/>)

Note: The two Perl modules can be installed by INSTALL.sh script in the lib directory.

```
cd lib
./INSTALL.sh
```

BASIC USAGE

- Trimming by quality 5 and filtering reads with any ambiguous base or low complexity.

```
$ perl FaQCs.pl -p reads1.fastq reads2.fastq -d out_directory
```

W15-4: FaQCs

①FaQCsディレクトリ中に、確かにlibというディレクトリがあることを確認。②libディレクトリに移動し、③マニュアルの記述通りに./INSTALL.shを実行。

```
iu@bielinux[Downloads] ls -lh [ 8:16午後 ]
total 9.6M
drwxrwxr-x 6 iu iu 4.0K  5月 28 20:16 FaQCs
drwxrwxr-x 8 iu iu 4.0K  5月  7 14:46 FastQC
-rw-rw-r-- 1 iu iu 9.6M  3月 25 18:40 fastqc_v0.11.3.z
ip
iu@bielinux[Downloads] pwd [ 8:40午後 ]
/home/iu/Downloads
iu@bielinux[Downloads] cd FaQCs [ 8:49午後 ]
iu@bielinux[FaQCs] pwd [ 8:49午後 ]
/home/iu/Downloads/FaQCs
① iu@bielinux[FaQCs] ls [ 8:49午後 ]
COPYRIGHT FaQCs.pl lib README.md
example galaxy_module LICENSE
② iu@bielinux[FaQCs] cd lib [ 8:49午後 ]
iu@bielinux[lib] pwd [ 8:50午後 ]
/home/iu/Downloads/FaQCs/lib
③ iu@bielinux[lib] ./INSTALL.sh [ 8:50午後 ]
```


W15-4: FaQCs

```
File Edit View Search Terminal Help 21:01
chmod 755 blib/arch/auto/String/Approx/Approx.so
cp Approx.bs blib/arch/auto/String/Approx/Approx.bs
chmod 644 blib/arch/auto/String/Approx/Approx.bs
Manifying blib/man3/String::Approx.3pm

-----
String-Approx-3.27 Installed
-----

R is found.
All done!

iu@bielinux[lib] [ 9:00午後 ]
```

インストールが無事できた模様。「R is found.」となっている。これはW15-3でRもprerequisiteにリストアップされているから。Bio-Linux 8はデフォルトでRがインストールされているので、問題なくFaQCsのインストールが成功する。このあたりがBio-Linux 8のありがたいところ。

W15-5: パスを通す

FaQCsプログラムの実体は、FaQCs.plというperlプログラムであることが「基本的な利用法 (basic usage)」中の記述内容からわかる。

```
cd lib
./INSTALL.sh
```

BASIC USAGE

- Trimming by quality 5 and filtering reads with any ambiguous base or low complexity.

```
$ perl FaQCs.pl -p reads1.fastq reads2.fastq -d out_directory
```

- Quality check only on subsamples of input, no trimming and filtering.

```
$ perl FaQCs.pl -p reads1.fastq reads2.fastq -d out_directory -qc_only
```

Full USAGE

```
Usage: perl FaQCs.pl [options] [-u unpaired.fastq] -p reads1.fastq reads2.fastq -d out_directory
Version 1.34
Input File: (can use more than once)
    -u          <Files> Unpaired reads
```

W15-5: パスを通す

FaQCsプログラムの実体は、
/home/iu/Downloads/FaQCs/FaQCs.pl。
どのディレクトリ上からもこのプログラムを
利用できるように「パスを通す」。手段は2
つ。1つはシンボリックリンク(W9-5)、そし
てもう1つは.zshrcファイル中に
/home/iu/Downloads/FaQCsディレクトリ
のパスを追加するやり方(W10-3)。

```
File Edit View Search Terminal Help
R is found.
All done!

iu@bielinux[lib] ls [ 9:00午後 ]
auto Parallel-ForkManager-1.03.tar.gz
INSTALL.sh String
Parallel String-Approx-3.27.tar.gz
iu@bielinux[lib] cd .. [ 9:04午後 ]
iu@bielinux[FaQCs] pwd [ 9:05午後 ]
/home/iu/Downloads/FaQCs
iu@bielinux[FaQCs] ls [ 9:05午後 ]
COPYRIGHT FaQCs.pl lib README.md
example galaxy_module LICENSE
iu@bielinux[FaQCs] [ 9:05午後 ]
```

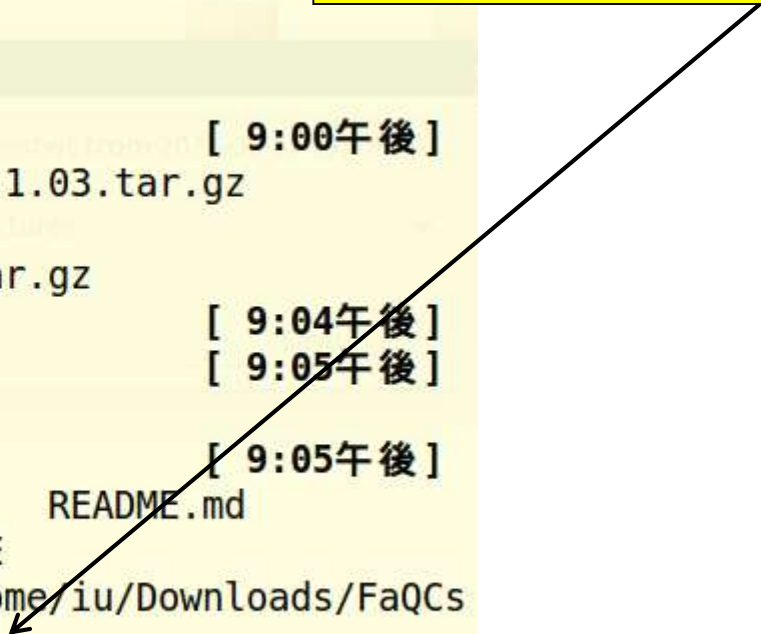


W15-5: パスを通す

ここでは、/usr/local/binに /home/iu/Downloads/FaQCs/FaQCs.pl のシンボリックリンクを貼る(W9-5)。混乱をきたさないように、FaQCs.plと同じファイル名にしている。

```
File Edit View Search Terminal Help
All done!

iu@bielinux[lib] ls [ 9:00午後 ]
auto Parallel-ForkManager-1.03.tar.gz
INSTALL.sh String
Parallel String-Approx-3.27.tar.gz
iu@bielinux[lib] cd .. [ 9:04午後 ]
iu@bielinux[FaQCs] pwd [ 9:05午後 ]
/home/iu/Downloads/FaQCs
iu@bielinux[FaQCs] ls [ 9:05午後 ]
COPYRIGHT FaQCs.pl lib README.md
example galaxy_module LICENSE
① iu@bielinux[FaQCs] sudo ln -s /home/iu/Downloads/FaQCs
/FaQCs.pl /usr/local/bin/FaQCs.pl
[sudo] password for iu:
iu@bielinux[FaQCs] [ 9:25午後 ]
```



W15-6: 確認

①ヘルプが表示されているかどうかで確認しようとしたがこけてる。。。W15-4で./Install.shを実行するとParallel::ForkManagerが自動的にインストールされるとW15-3のGitHubマニュアルには書いてあったが...。②と③でパスは確かに通っていることを確認。

```
iu@bielinux[FaQCs] pwd
/home/iu/Downloads/FaQCs
① iu@bielinux[FaQCs] FaQCs.pl -h
Can't locate Parallel/ForkManager.pm in @INC (you may need to install the Parallel::ForkManager module) (@INC contains: /usr/local/bin/./lib /usr/local/bin/lib /etc/perl /usr/local/lib/perl/5.18.2 /usr/local/share/perl/5.18.2 /usr/lib/perl5 /usr/share/perl5 /usr/lib/perl/5.18 /usr/share/perl/5.18 /usr/local/lib/site_perl .) at /usr/local/bin/FaQCs.pl line 44.
BEGIN failed--compilation aborted at /usr/local/bin/FaQCs.pl line 44.
② iu@bielinux[FaQCs] where FaQCs.pl
/usr/local/bin/FaQCs.pl
③ iu@bielinux[FaQCs] which FaQCs.pl
/usr/local/bin/FaQCs.pl
iu@bielinux[FaQCs] █
```

[9:34午後]
[11:58午前]
[11:58午前]
[11:58午前]



W15-7: 確認

念のため~/Downloads/FaQCs/libの中身を確認。②確かにINSTALL.sh実行によって、今問題だと言われているParallel ForkManagerというものは存在する。③(これは長年の経験がものをいうが、)もしやと思い、パスが通っている /usr/local/bin/FaQCs.plではなく、①の場所に存在するFaQCs.pl (つまり./FaQCs.pl)を指定するとエラーなくヘルプが表示される!

```
iu@bielinux[FaQCs] pwd
/home/iu/Downloads/FaQCs
iu@bielinux[FaQCs] ls
COPYRIGHT  example  FaQCs.pl  galaxy_module  lib  LICENSE  README.md
[ 4:01午後]
iu@bielinux[FaQCs] ls -lh lib
total 80K
drwxrwxr-x 4 iu iu 4.0K  5月 28 21:00 auto
-rwxrwxr-x 1 iu iu 2.3K  5月 28 20:16 INSTALL.sh
drwxr-xr-x 2 iu iu 4.0K  5月 28 21:00 Parallel
-rw-rw-r-- 1 iu iu 14K   5月 28 20:16 Parallel-ForkManager-1.03.tar.gz
drwxr-xr-x 2 iu iu 4.0K  5月 28 21:00 String
-rw-rw-r-- 1 iu iu 45K   5月 28 20:16 String-Approx-3.27.tar.gz
[ 4:01午後]
iu@bielinux[FaQCs] ./FaQCs.pl -h
Usage: perl ./FaQCs.pl [options] [-u unpaired.fastq] -p reads1.fas
tq reads2.fastq -d out_directory
Version 1.34
Input File: (can use more than once)
    -u          <Files> Unpaired reads
    -p          <Files> Paired reads in two files and separat
e by space
```

W15-7: 確認

①同じディレクトリ上で、②ホームディレクトリからの相対パス指定でもエラーなくヘルプが表示されるのを確認することで、最低限FaQCs.plを動作させる手段は確保。

```
File Edit View Search Terminal Help [ 4:02午後 ]
iu@bielinux[FaQCs] pwd
/home/iu/Downloads/FaQCs
iu@bielinux[FaQCs] ls [ 4:13午後 ]
COPYRIGHT example FaQCs.pl galaxy_module lib LICENSE README.md
iu@bielinux[FaQCs] ~/Downloads/FaQCs/FaQCs.pl -h | head [ 4:13午後 ]
Usage: perl /home/iu/Downloads/FaQCs/FaQCs.pl [options] [-u unpaired.fastq] -p reads1.fastq reads2.fastq -d out_directory
Version 1.34
Input File: (can use more than once)
-u <Files> Unpaired reads
-p <Files> Paired reads in two files and separate by space
Trim:
-mode "HARD" or "BWA" or "BWA_plus" (default BWA_plus)
BWA trim is NOT A HARD cutoff! (see bwa's bwa_trim_read() function in bwaseqio.c)
iu@bielinux[FaQCs] [ 4:13午後 ]
```

W15-8: 考察

```
File Edit View Search Terminal Help
iu@bielinux[FaQCs] pwd
/home/iu/Downloads/FaQCs
iu@bielinux[FaQCs] ls
COPYRIGHT example FaQCs.pl galaxy_module lib
iu@bielinux[FaQCs] ls -lh lib
total 80K
drwxrwxr-x 4 iu iu 4.0K  5月 28 21:00 auto
-rwxrwxr-x 1 iu iu 2.3K  5月 28 20:16 INSTALL.sh
drwxr-xr-x 2 iu iu 4.0K  5月 28 21:00 Parallel
-rw-rw-r-- 1 iu iu 14K  5月 28 20:16 Parallel-F
drwxr-xr-x 2 iu iu 4.0K  5月 28 21:00 String
-rw-rw-r-- 1 iu iu 45K  5月 28 20:16 String-App
iu@bielinux[FaQCs] ./FaQCs.pl -h
Usage: perl ./FaQCs.pl [options] [-u unpair
tq reads2.fastq -d out_directory
Version 1.34
Input File: (can use more than once)
-u <Files> Unpaired reads
-p <Files> Paired reads in two files and separat
e by space
```

今回の原因であるParallel-ForkManagerというものは
~/Downloads/FaQCs/lib/Parallelとして確かに存在する(これは事実)。おそらくFaQCs.plのプログラム中では、これを認識する手段として相対パス指定をしているからだろうと推測する。つまり、FaQCs.plからみて、./lib/Parallelという参照をしている。それゆえ、パスを通したFaQCs.plを実行しようとする、/usr/local/bin上から、./lib/Parallelという実体のないものを参照しようとするためにそんなものはない!と文句を言われているのだろうと妄想するが、それが真実かどうかはそれ以上追及はしない。だって、使える手段は既に確保しているのだから。。原因を追究するエフォートはFaQCsのマニュアルを読んだりexamplesの実行に費やすべし!

W16-1: Perlモジュール

それでも、一部のバイオインフォーマティションはある程度はあがく。例えば、`./INSTALL.sh`では必要な2つのPerlモジュールとやらをインストールしているのではなく、`~/Downloads/FaQCs/lib`上に置いていただけなのではと推測する。となると、まずやることはW15-3の「Parallel::ForkManager module from CPAN」の記述からこれを通常の方法でインストールしてみようと考え

```
iu@bielinux[FaQCs] pwd
/home/iu/Downloads/FaQCs
iu@bielinux[FaQCs] ls
COPYRIGHT example FaQCs.pl galaxy_module l
iu@bielinux[FaQCs] ls -lh lib
total 80K
drwxrwxr-x 4 iu iu 4.0K  5月 28 21:00 auto
-rwxrwxr-x 1 iu iu 2.3K  5月 28 20:16 INSTALL.sh
drwxr-xr-x 2 iu iu 4.0K  5月 28 21:00 Parallel
-rw-rw-r-- 1 iu iu 14K  5月 28 20:16 Parallel-ForkManager-1.03.tar.gz
drwxr-xr-x 2 iu iu 4.0K  5月 28 21:00 String
-rw-rw-r-- 1 iu iu 45K  5月 28 20:16 String-Approx-3.27.tar.gz
iu@bielinux[FaQCs] ./FaQCs.pl -h [ 4:01午後 ]
Usage: perl ./FaQCs.pl [options] [-u unpaired.fastq] -p reads1.fas
tq reads2.fastq -d out_directory
Version 1.34
Input File: (can use more than once)
    -u          <Files> Unpaired reads
    -p          <Files> Paired reads in two files and separat
e by space
```

W16-1: Perlモジュール

①「sudo cpan ForkManager」を試している。ログインパスワードを聞かれたら打つ (pass1409)。②よくわからないが、できるだけ自動でセッティングしてくれるようだ。基本的に言われるがままリターンキーを押す。

```
iu@bielinux[FaQCs] ls
COPYRIGHT example FaQCs.pl galaxy_module lib LICENSE README.md
iu@bielinux[FaQCs] FaQCs.pl -h [ 5:03午後 ]
Can't locate Parallel/ForkManager.pm in @INC (you may need to install t
he Parallel::ForkManager module) (@INC contains: /usr/local/bin/./lib
/usr/local/bin/lib /etc/perl /usr/local/lib/perl/5.18.2 /usr/local/shar
e/perl/5.18.2 /usr/lib/perl5 /usr/share/perl5 /usr/lib/perl/5.18 /usr/s
hare/perl/5.18 /usr/local/lib/site_perl .) at /usr/local/bin/FaQCs.pl l
ine 44.
BEGIN failed--compilation aborted at /usr/local/bin/FaQCs.pl line 44.
iu@bielinux[FaQCs] sudo cpan ForkManager [ 5:04午後 ]
CPAN.pm requires configuration, but most of it can be done automaticall
y.
If you answer 'no' below, you will enter an interactive dialog for each
configuration option instead.
Would you like to configure as much as possible automatically? [yes] █
```



W16-1: Perlモジュール

```
File Edit View Search Terminal Help [ 5:03午後 ]
iu@bielinux[FaQCs] ls
COPYRIGHT example FaQCs.pl galaxy_module lib LICENSE README.md
iu@bielinux[FaQCs] FaQCs.pl -h [ 5:03午後 ]
Can't locate Parallel/ForkManager.pm in @INC (you may need to install t
he Parallel::ForkManager module) (@INC contains: /usr/local/bin/./lib
/usr/local/bin/lib /etc/perl /usr/local/lib/perl/5.18.2 /usr/local/shar
e/perl/5.18.2 /usr/lib/perl5 /usr/share/perl5 /usr/lib/perl/5.18 /usr/s
hare/perl/5.18 /usr/local/lib/site_perl .) at /usr/local/bin/FaQCs.pl l
ine 44.
BEGIN failed--compilation aborted at /usr/local/bin/FaQCs.pl line 44.
① iu@bielinux[FaQCs] sudo cpan ForkManager [ 5:04午後 ]
CPAN.pm requires configuration, but most of it can be done automaticall
y.
If you answer 'no' below, you will enter an interactive dialog for each
configuration option instead.
Would you like to configure as much as possible automatically? [yes] ②
Would you like me to automatically choose some CPAN mirror
sites for you? (This means connecting to the Internet) [yes] ③
```

W16-1: Perlモジュール

yesの状態ではリターンキーを押したので、何やらインストールが始まったようだ。

```
iu@bielinux[~/Downloads/FaQCs]
Would you like me to automatically choose some CPAN mirror sites for you? (This means connecting to the Internet) [yes]
Fetching with LWP:
http://mirrors-ru.go-parts.com/cpan/authors/01mailrc.txt.gz
Reading '/home/iu/.cpan/sources/authors/01mailrc.txt.gz'
.....
.....DONE
Fetching with LWP:
http://mirrors-ru.go-parts.com/cpan/modules/02packages.details.txt.gz
Reading '/home/iu/.cpan/sources/modules/02packages.details.txt.gz'
Database was generated on Fri, 29 May 2015 06:41:02 GMT
.....
New CPAN.pm version (v2.10) available.
[Currently running version is v2.00]
You might want to try
  install CPAN
  reload cpan
to both upgrade CPAN.pm and run the new version without leaving
the current session.
```



3

W16-1: Perlモジュール

結論として「sudo cpan ForkManger」は失敗。
①「sudo cpan Parallel::ForkManager」でリトライ

```
File Edit View Search Terminal Help
[Currently running version is v2.00]
You might want to try
  install CPAN
  reload cpan
to both upgrade CPAN.pm and run the new version without leaving
the current session.
.....DONE
Fetching with LWP:
http://mirrors-ru.go-parts.com/cpan/modules/03modlist.data.gz
Reading '/home/iu/.cpan/sources/modules/03modlist.data.gz'
DONE
Writing /home/iu/.cpan/Metadata
Warning: Cannot install ForkManager, don't know what it is.
Try the command

  i /ForkManager/

to find objects with matching identifiers.
iu@bielinux[FaQCs] sudo cpan Parallel::ForkManager [ 5:10午後 ]
```



W16-2: Perlモジュール

警告メッセージなどもなく、パッと見 successfulなどという記述しかないのでもうまくいったと判断。①ついでに、W15-3で記載されているもう一つの「String::Approx」というPerlモジュールのインストールしておく。

```
File Edit View Search Terminal Help
child process '24744' disappeared. A call to `waitpid`
perl::ForkManager might have reaped it.
t/waitpid-conflict.t .. ok
All tests successful.

Test Summary Report
-----
t/waitpid-conflict.t (Wstat: 0 Tests: 2 Failed: 0)
  TODO passed: 1-2
Files=6, Tests=18, 57 wallclock secs ( 0.05 usr  0.02 sys + 0.61 cusr
 0.37 csys = 1.05 CPU)
Result: PASS
  YANICK/Parallel-ForkManager-1.14.tar.gz
  /usr/bin/make test -- OK
Running make install
Installing /usr/local/share/perl/5.18.2/Parallel/ForkManager.pm
Installing /usr/local/man/man3/Parallel::ForkManager.3pm
Appending installation info to /usr/local/lib/perl/5.18.2/perllocal.pod
  YANICK/Parallel-ForkManager-1.14.tar.gz
  /usr/bin/make install -- OK
iu@bielinux[FaQCs] sudo cpan String::Approx
```



[5:18午後]

W16-3: FaQCs.pl

「String::Approx」もうまくインストールできた模様。Perlのモジュール名が違う程度なら「sudo cpan モジュール名」でうまくいく成功体験を得た。もしかしてと思い①「FaQCs.pl -h | more」を実行すると...

```
File Edit View Search Terminal Help
t/arindex.t .. ok
t/aslice.t ... ok
t/asubst.t ... ok
t/user.t ..... ok
All tests successful.
Files=7, Tests=140, 0 wallclock secs ( 0.04 usr 0.01 sys + 0.15 cusr
0.03 csys = 0.23 CPU)
Result: PASS
JHI/String-Approx-3.27.tar.gz
/usr/bin/make test -- OK
Running make install
Files found in blib/arch: installing files in blib/lib into architectur
e dependent library tree
Installing /usr/local/lib/perl/5.18.2/auto/String/Approx/Approx.bs
Installing /usr/local/lib/perl/5.18.2/auto/String/Approx/Approx.so
Installing /usr/local/lib/perl/5.18.2/String/Approx.pm
Installing /usr/local/man/man3/String::Approx.3pm
Appending installation info to /usr/local/lib/perl/5.18.2/perllocal.pod
JHI/String-Approx-3.27.tar.gz
/usr/bin/make install -- OK
① iu@bielinux[FaQCs] FaQCs.pl -h | more [ 5:29午後 ]
```

W16-3: FaQCs.pl

```

File Edit View Search Terminal Help
Installing /usr/local/lib/perl/5.18.2/auto/String/Approx/Approx.bs
Installing /usr/local/lib/perl/5.18.2/auto/String/Approx/Approx.so
Installing /usr/local/lib/perl/5.18.2/String/Approx.pm
Installing /usr/local/man/man3/String::Approx.3pm
Appending installation info to /usr/local/lib/perl/5.18.2/perllocal.pod
JHI/String-Approx-3.27.tar.gz
/usr/bin/make install -- OK
① iu@bielinux[FaQCs] FaQCs.pl -h | more [ 5:29午後 ]
Usage: perl /usr/local/bin/FaQCs.pl [options] [-u unpaired.fastq]
-p reads1.fastq reads2.fastq -d out_directory
Version 1.34
Input File: (can use more than once)
-u <Files> Unpaired reads
-p <Files> Paired reads in two files and separat
e by space
Trim:
-mode "HARD" or "BWA" or "BWA_plus" (default BWA_pl
us)
BWA trim is NOT A HARD cutoff! (see bwa's bwa

```


W16-3: 確認

①念のためホームディレクトリに移動したのち、再度②「FaQCs.pl -h」。完璧！

```
iu@bielinux[FaQCs] pwd [ 5:42午後 ]
/home/iu/Downloads/FaQCs
iu@bielinux[FaQCs] ls [ 5:43午後 ]
COPYRIGHT example FaQCs.pl galaxy_module lib LICENSE README.md
iu@bielinux[FaQCs] FaQCs.pl -h | head -n 4 [ 5:43午後 ]
Usage: perl /usr/local/bin/FaQCs.pl [options] [-u unpaired.fastq]
-p reads1.fastq reads2.fastq -d out_directory
Version 1.34
Input File: (can use more than once)
-u <Files> Unpaired reads
① iu@bielinux[FaQCs] cd [ 5:43午後 ]
iu@bielinux[iu] pwd [ 5:43午後 ]
/home/iu
② iu@bielinux[iu] FaQCs.pl -h | head -n 4 [ 5:43午後 ]
Usage: perl /usr/local/bin/FaQCs.pl [options] [-u unpaired.fastq]
-p reads1.fastq reads2.fastq -d out_directory
Version 1.34
Input File: (can use more than once)
-u <Files> Unpaired reads
iu@bielinux[iu] [ 5:43午後 ]
```

W17-1: FaQCs

FaQCsの利用例。例題では明記されていないが、gzip圧縮ファイルにも対応している。但し、出力が非圧縮fastqなのがオシイところ。

GitHub | This repository | Search | Explore | Features | Enterprise | Blog | Sign up | Sign in

LANL-Bioinformatics / FaQCs
forked from chianchi/FaQCs

45 commits | 1 branch | 1 release | 1 contributor

branch: master | FaQCs / +

This branch is 22 commits ahead, 15 commits behind chianchi/master

fix type

chianchi authored on Mar 31 | latest commit: 6aed8354ec

- example: adjust margin of pdf plots (3 months ago)
- galaxy_module: Update FaQCs.yml (6 months ago)
- lib: Add option "-Trim_off <bool>" Turn off trimming from 5' end. (4 months ago)
- COPYRIGHT: add LICENSE (a year ago)
- FaQCs.pl: patch for jellyfish v2 (3 months ago)
- LICENSE: add LICENSE (a year ago)
- README.md: fix typo (2 months ago)

FaQCs: Quality Control of Next Generation Sequencing Data

Quality 3D plot. (Position vs. Score vs. Frequency)

BASIC USAGE

- Trimming by quality 5 and filtering reads with any ambiguous bases or low complexity.

```
$ perl FaQCs.pl -p reads1.fastq reads2.fastq -d out_directory
```



- Quality check only on subsamples of input, no trimming and filtering.

```
$ perl FaQCs.pl -p reads1.fastq reads2.fastq -d out_directory -qc_only
```

Full USAGE

```
Usage: perl FaQCs.pl [options] [-u unpaired.fastq] -p reads1.fastq reads2.fastq -d out_directory
Version 1.34
Input File: (can use more than once)
-u <Files> Unpaired reads

Trim:
-p <Files> Paired reads in two files and separate by space
-mode "HARD" or "BWA" or "BWA_plus" (default BWA_plus)
      BWA trim is NOT A HARD cutoff! (see bwa's bwa_trim_read() function in bwa)
-q <INT> Targets # as quality level (default 5) for trimming
-5end <INT> Cut # bp from 5 end before quality trimming/filtering
-3end <INT> Cut # bp from 3 end before quality trimming/filtering
-adapter <bool> Filter reads with illumina adapter/primers (default: no)
        -rate <FLOAT> Mismatch ratio of adapters' length (default: 0.2, allow 2)
Filters:
-min_L <INT> Trimmed read should have to be at least this minimum length (default: 1)
-avg_q <NUM> Average quality cutoff (default:0, no filtering)
```

W17-1: FaQCs

①乳酸菌paired-end RNA-seqデータを入力として実行。「-d result」オプションをつけているので、resultディレクトリが作成されて、そこに出力結果ファイルが生成される。赤下線で示すように、実行時間を計測している。理由は、最初計算が全然終わらなくて、CTRL + Cで計算を途中でやめたら、resultディレクトリがちゃんと作成されてファイルができてそうな雰囲気を感じたから、もう一度timeコマンドを追加して再実行した

```
iu@bielinux[~/Documents/srp017156]
iu@bielinux[srp017156] pwd
/home/iu/Documents/srp017156
iu@bielinux[srp017156] ls -la
-rw-rw-r-- 1 iu iu 74906576 5月 6 10:08 SRR616268sub_2.fastq.gz
-rw-rw-r-- 1 iu iu 67158462 5月 6 10:08 SRR616268sub_1.fastq.gz
iu@bielinux[srp017156] time FaQCs.pl -p SRR616268sub_1.fastq.gz SRR616268sub_2.fastq.gz -d result
Bwa extension trimming algorithm is used.
Processing SRR616268sub_1.fastq.gz SRR616268sub_2.fastq.gz file
Processed 2000000/2000000
Post Trimming Length(Mean, Std, Median, Max, Min) of 1976760 reads
with Overall quality 36.37
(99.33, 8.65, 107.0, 107, 50)
FaQCs.pl -p SRR616268sub_1.fastq.gz SRR616268sub_2.fastq.gz -d result
lt 1070.94s user 78.65s system 91% cpu 20:56.11 total
iu@bielinux[srp017156] [ 1:33午後]
```

W17-1: FaQCs

片側100万リードなので、入力は計200万リード。そのうち1,976,760リードが出力されている。それらの平均クオリティスコアは36.37。リード長の平均(99.33)、最短(50)などとなっているが、②中央値(Median = 107.0)は直観的にオカシイ。理由は入力ファイルの配列長はForward側が107 bp、Reverse側が93 bp。出力は同じリード数(1,976,760/2 = 988,380)のはずなので、どう転んでもmedianは(107 + 93)/2 = 100 bp 以下になるべきという判断である。おそらくバグ。

```
iu@bielinux[~/Documents/srp017156]
iu@bielinux[srp017156] pwd
/home/iu/Documents/srp017156
iu@bielinux[srp017156] ls -la *sub*
-rw-rw-r-- 1 iu iu 74906576  5月  6
-rw-rw-r-- 1 iu iu 67158462  5月  6
iu@bielinux[srp017156] time FaQCs.pl
616268sub_2.fastq.gz -d result
Bwa extension trimming algorithm is used.
Processing SRR616268sub_1.fastq.gz SRR616268sub_2.fastq.gz file
Processed 2000000/2000000
Post Trimming Length(Mean, Std, Median, Max, Min) of 1976760 reads
with Overall quality 36.37
(99.33, 8.65, 107.0, 107, 50)
FaQCs.pl -p SRR616268sub_1.fastq.gz SRR616268sub_2.fastq.gz -d resu
lt 1070.94s user 78.65s system 91% cpu 20:56.11 total
iu@bielinux[srp017156] [ 1:33午後]
```

W17-1: FaQCs

①FaQCs.pl実行時にtimeコマンドをつけているので、②実行時間に関する表示がある。1070.94秒(約18分)かかっていることがわかる。サブセットの100万リードでこれだけかかるのだから、100倍以上あるオリジナルの1.35億リードだと、単純計算で $1070.94 \times 135 = 40.1625$ 時間かかることになる。

```
iu@bielinux[~/Documents/srp017156]
iu@bielinux[ srp017156] pwd
/home/iu/Documents/srp017156
iu@bielinux[ srp017156] ls -la *sub*
-rw-rw-r-- 1 iu iu 74906576  5月  6 10:07 SRR616268sub_1.fastq.gz
-rw-rw-r-- 1 iu iu 67158462  5月  6 10:08 SRR616268sub_2.fastq.gz
iu@bielinux[ srp017156] time FaQCs.pl -p SRR616268sub_1.fastq.gz SRR616268sub_2.fastq.gz -d result
Bwa extension trimming algorithm is used.
Processing SRR616268sub_1.fastq.gz SRR616268sub_2.fastq.gz file
Processed 2000000/2000000
Post Trimming Length(Mean, Std, Median, Max, Min) of 1976760 reads
with Overall quality 36.37
(99.33, 8.65, 107.0, 107, 50)
FaQCs.pl -p SRR616268sub_1.fastq.gz SRR616268sub_2.fastq.gz -d result
lt 1070.94s user 78.65s system 91% cpu 20:56.11 total
iu@bielinux[ srp017156] [ 1:33午後]
```

W17-1: FaQCs

```
iu@bielinux[~/Documents/srp017156] [ 1:52午後 ]
iu@bielinux[srp017156] ls -la *sub* result
-rw-rw-r-- 1 iu iu 74906576 5月 6 10:07 SRR616268sub_1.fastq.gz
-rw-rw-r-- 1 iu iu 67158462 5月 6 10:08 SRR616268sub_2.fastq.gz

result:
total 587908
drwxrwxr-x 2 iu iu 4096 6月 25 12:57 .
drwxrwxr-x 3 iu iu 4096 6月 25 12:36 ..
-rw-rw-r-- 1 iu iu 96 6月 25 12:37 fastqCount.txt
-rw-rw-r-- 1 iu iu 313487572 6月 25 12:56 QC.1.trimmed.fastq
-rw-rw-r-- 1 iu iu 281569978 6月 25 12:56 QC.2.trimmed.fastq
-rw-rw-r-- 1 iu iu 200216 6月 25 12:57 QC_qc_report.pdf
-rw-rw-r-- 1 iu iu 833 6月 25 12:57 QC.stats.txt
-rw-rw-r-- 1 iu iu 6734366 6月 25 12:57 QC.unpaired.trimmed.fastq
iu@bielinux[srp017156] [ 1:53午後 ]
```

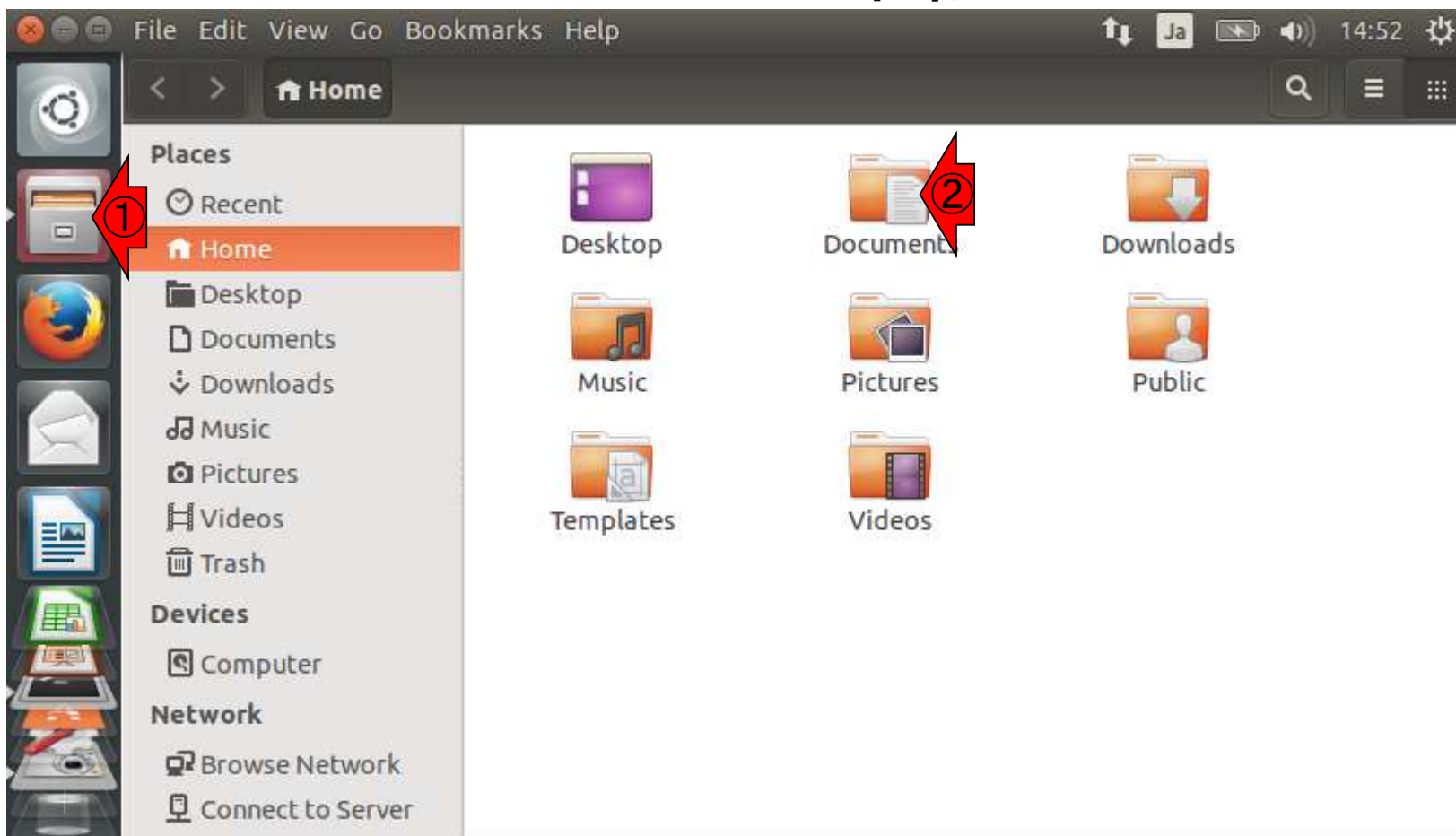
W17-1: FaQCs

①作業ディレクトリをresultに変更。②赤枠内がトリム後のfastqファイル。③QC.1.trimmed.fastqファイルに対してfastqc2コマンド(FastQC ver. 0.11.3)を実行。④FastQC実行結果ファイル。

```
iu@bielinux[~/Documents/srp017156/result]
① iu@bielinux[srp017156] cd result
iu@bielinux[result] ls -la
total 587908
drwxrwxr-x 2 iu iu      4096  6月 25 12:57 .
drwxrwxr-x 3 iu iu      4096  6月 25 12:36 ..
-rw-rw-r-- 1 iu iu        96  6月 25 12:37 fastqCount.txt
-rw-rw-r-- 1 iu iu 313487572  6月 25 12:56 QC.1.trimmed.fastq
-rw-rw-r-- 1 iu iu 281569978  6月 25 12:56 QC.2.trimmed.fastq
-rw-rw-r-- 1 iu iu   200216  6月 25 12:57 QC_qc_report.pdf
-rw-rw-r-- 1 iu iu     833  6月 25 12:57 QC.stats.txt
-rw-rw-r-- 1 iu iu  6734366  6月 25 12:57 QC.unpaired.trimmed.fastq
③ iu@bielinux[result] fastqc2 -q QC.1.trimmed.fastq      [ 2:32午後 ]
iu@bielinux[result] ls -la                                [ 2:35午後 ]
total 588672
drwxrwxr-x 2 iu iu      4096  6月 25 14:33 .
drwxrwxr-x 3 iu iu      4096  6月 25 12:36 ..
-rw-rw-r-- 1 iu iu        96  6月 25 12:37 fastqCount.txt
-rw-rw-r-- 1 iu iu 313487572  6月 25 12:56 QC.1.trimmed.fastq
-rw-rw-r-- 1 iu iu   363785  6月 25 14:33 QC.1.trimmed_fastqc.html
-rw-rw-r-- 1 iu iu  416251  6月 25 14:33 QC.1.trimmed_fastqc.zip
-rw-rw-r-- 1 iu iu 281569978  6月 25 12:56 QC.2.trimmed.fastq
-rw-rw-r-- 1 iu iu   200216  6月 25 12:57 QC_qc_report.pdf
-rw-rw-r-- 1 iu iu     833  6月 25 12:57 QC.stats.txt
-rw-rw-r-- 1 iu iu  6734366  6月 25 12:57 QC.unpaired.trimmed.fastq
iu@bielinux[result] █ [ 2:35午後 ]
```

W17-2: FastQCで確認

FastQC実行結果ファイルを眺めるべく、ファイルが存在するディレクトリまで移動。



W17-2: FastQCで確認

FastQC実行結果ファイルを眺めるべく、ファイルが存在するディレクトリまで移動。①で示すようにディレクトリ移動に伴い赤枠で表示されるのでそれを頼りにする。②目的のhtmlファイルをゲストOS上で開く。もちろんこれをDesktopのmac_shareにコピーしてホストOS上で眺めてもよい。



W17-2: FastQCで確認

①IlluminaのTruSeq Adapter Index 2が見つかった。このことからデフォルトではアダプター配列除去までは行えないことを学ぶ。

FastQC Report		Thu 25 Jun 2015		
		QC.1.trimmed.fastq		
Summary <ul style="list-style-type: none"> Basic Statistics Per base sequence quality Per tile sequence quality Per sequence quality scores Per base sequence content Per sequence GC content Per base N content Sequence Length Distribution Sequence Duplication Levels Overrepresented sequences Adapter Content Kmer Content 	ATTTAGCCTTGGGAGATGATCCCTCCCGGATTCGGACCGAATTTCCAGTGT	1420	0.1430003223103410	No Hit
	CGCTACTCATGCCGGCATTCTCACTTCTAAGCGCTCCAGCCGTCCTCACG	1398	0.14304234073749686	No Hit
	CCGGGGTGTCTTTTACCTTTCCCTCACGGTACTGGTTCACTATCGGTCAC	1349	0.13802869646272048	No Hit
	GGGCCTATTCAGTGGCTGACCTTGGCGTCAGCACCCCTTCTCCGAAG	1341	0.13721014229541006	No Hit
	CTGGTGATCTGGGCTGTTCCCTTTGACAATGGACCTTATCGCTCACTG	1336	0.13669854594084105	No Hit
	GCCGCCGGCCAGCTATGTATTCAGTACAAGCAATACACTGATGTGTA	1319	0.13495911833530638	No Hit
	GCCGAGTTCCTTAACGAGAGTTCGCTCGCTCACCTGAGGATACTCTCCTC	1305	0.13352664854251314	No Hit
	AGATCGGAAGAGCACACGTCTGAACTCCAGTCACCGATGTATCTCGTATG	1296	0.13260577510428892	TruSeq Adapter, Index 2 (100% over 49bp)
	GTTTGGGCTCTTCCCACTTCGCTCGCCGCTACTATGGGAATCGAGTTTT	1262	0.12912691989321962	No Hit
	CGTCCCTCCATCGCTTAAACAAAATAAACTAGTGCAGGAATCTCAACCTG	1258	0.12871764280956438	No Hit
	CACACGGTTTCAGGAAGTTCCTTCCCTTCCGGGTGCTTTTACCT	1257	0.1286153235386506	No Hit
	CCCTAGTTCAAACAGTGTCTACCTCCATGATCCATCCTCCGAGGCTAAC	1249	0.12779676937134016	No Hit
	GTCATTTTGCCGAGTTCCTTAACGAGAGTTCGCTCGCTCACCTGAGGATA	1249	0.12779676937134016	No Hit
	ACCTTAACTGGTGATCTGGGCTGTTCCCTTTGACAATGGACCTTATCG	1223	0.1251364683275813	No Hit
	CACAGTTTCGGTATTATGCTTAGCCCCGGTATATTTTCGGCGCAGTGCCA	1220	0.12482951051483988	No Hit
	GGTCATTTTGCCGAGTTCCTTAACGAGAGTTCGCTCGCTCACCTGAGGAT	1208	0.12360167926387423	No Hit
	CCACTTAGCATAAATTTAGGGACCTTAACTGGTGATCTGGGCTGTTCCCT	1207	0.12349935999296043	No Hit
	CCGGTTCACTTCTACAAAAGGCACGCCATTACCCGTTAACGGGCTTTGACT	1191	0.12186225165833958	No Hit
	CTGCCGCCGGCCAGCTATGTATTCAGTACAAGCAATACACTGATGTGTA	1178	0.12053210113646014	No Hit
	GCCGTAATCAGGATCCTGGACGGAGGTTTCGACGTTTCGCTTACAGGGCT	1174	0.12012282405280492	No Hit



①

W17-3: マニュアル

マニュアル(Full USAGE)をよく見ると、Illuminaデータの場合は「-adapter」をつけることでよさそうだと気付くが、「Filter reads with ...」という記述内容から「アダプター配列を含むリードごと除去されるのでは?!」という一抹の不安を覚える。

GitHub | This repository | Search | Explore | Features | Enterprise | Blog | Sign up | Sign in

LANL-Bioinformatics / FaQCs
forked from chianchi/FaQCs

45 commits | 1 branch | 1 release | 1 contributor

branch: master | FaQCs / +

This branch is 22 commits ahead, 15 commits behind chianchi/master | Pull Request | Compare

fix typo

chianchi authored on Mar 31 | Latest commit: 6aed9354ec

- example | adjust margin of pdf plots | 3 months ago
- galaxy_module | Update FaQCs.yml | 6 months ago
- lib | Add option "-Trim_off <bool>" Turn off trimming from 5' end. | 4 months ago
- COPYRIGHT | add LICENSE | a year ago
- FaQCs.pl | patch for jellyfish v2 | 3 months ago
- LICENSE | add LICENSE | a year ago
- README.md | fix typo | 2 months ago

FaQCs: Quality Control of Next Generation Sequencing Data

Quality 3D plot. (Position vs. Score vs. Frequency)

BASIC USAGE

- Trimming by quality

```
$ perl FaQCs.pl -p reads1.fastq reads2.fastq -d out_directory
```

- Quality check only on subsamples of input, no trimming and filtering.

```
$ perl FaQCs.pl -p reads1.fastq reads2.fastq -d out_directory -qc_only
```

Full USAGE

```
Usage: perl FaQCs.pl [options] [-u unpaired.fastq] -p reads1.fastq reads2.fastq -d out_directory
Version 1.34
Input File: (can use more than once)
-u <Files> Unpaired reads

Trim:
-p <Files> Paired reads in two files and separate by space
-mode "HARD" or "BWA" or "BWA_plus" (default BWA_plus)
      BWA trim is NOT A HARD cutoff! (see bwa's bwa_trim_read() function in bwa)
-q <INT> Targets # as quality level (default 5) for trimming
-5end <INT> Cut # bp from 5 end before quality trimming/filtering
-3end <INT> Cut # bp from 3 end before quality trimming/filtering
-adapter <bool> Filter reads with illumina adapter/primers (default: no)
      -rate <FLOAT> Mismatch ratio of adapters' length (default: 0.2, allow 2)
-artifactFile <File> additional artifact (adapters/primers/contaminations) reference
Filters:
-min_L <INT> Trimmed read should have to be at least this minimum length (default:
-avg_q <NUM> Average quality cutoff (default:0, no filtering)
```



W17-3: FaQCs

①「-adapter 1」オプションをつけて実行した結果をresult2ディレクトリに保存。出力ファイル中の総リード数(1,972,635)は、「-adapter 1」オプションをつけなかったとき(1,976,760)に比べて減っているのが妥当。しかし、真実は不明。

```
iu@bielinux[result] cd ..
iu@bielinux[srp017156] pwd
/home/iu/Documents/srp017156
iu@bielinux[srp017156] ls
hoge.txt          JSLAB4_6_Linux.sh  SRR616268_1.fastq.bz2
JSLAB4_1_Linux.sh JSLAB4_7_Linux.sh  SRR616268_2.fastq.bz2
JSLAB4_2_Linux.sh JSLAB4_8_Linux.sh  SRR616268sub_1.fastq.gz
JSLAB4_3_Linux.sh result              SRR616268sub_2.fastq.gz
iu@bielinux[srp017156] time FaQCs.pl -adapter 1 -p SRR616268sub_1.fastq.gz SRR616268sub_2.fastq.gz -d result2
Bwa extension trimming algorithm is used.
Processing SRR616268sub_1.fastq.gz SRR616268sub_2.fastq.gz file
Processed 2000000/2000000
Post Trimming Length(Mean, Std, Median, Max, Min) of 1972635 reads with Overall quality 36.37
(99.33, 8.62, 107.0, 107, 50)
FaQCs.pl -adapter 1 -p SRR616268sub_1.fastq.gz SRR616268sub_2.fastq.gz -d 1531.12s user 124.22s system 92% cpu 29:41.93 total
iu@bielinux[srp017156] ls
hoge.txt          JSLAB4_7_Linux.sh  SRR616268_2.fastq.bz2
JSLAB4_1_Linux.sh JSLAB4_8_Linux.sh  SRR616268sub_1.fastq.gz
JSLAB4_2_Linux.sh result              SRR616268sub_2.fastq.gz
JSLAB4_3_Linux.sh result2
JSLAB4_6_Linux.sh SRR616268_1.fastq.bz2
iu@bielinux[srp017156]
```



W18-1: IGVダウンロード

IGV (Integrative Genomics Viewer)のインストール手順を示します。①Registrationをしてから、②Downloads。



- Home
- Downloads ②
- Documents
- Hosted Genomes
- FAQ
- IGV User Guide
- File Formats
- Release Notes
- IGV for iPad
- Credits
- Contact

Search website

search
[Broad Home](#)
[Cancer Program](#)



Home



What's New

NEWS September 2014. The IGV iPad app can now be installed from the Apple App Store. *IGV for iPad* is a lightweight genomic data viewer that provides some of the functionality available in our regular desktop IGV. See the [IGV for iPad documentation](#) for details.

Overview

The Integrative Genomics Viewer (IGV) is a high-performance visualization tool for interactive exploration of large, integrated genomic datasets. It supports a wide variety of data types, including array-based and next-generation sequence data, and genomic annotations.

Downloads

Please [register](#) ① to download IGV. After registering, you can log in at any time using your email address. Permission to use IGV is granted under the GNU [LGPL license](#).

Citing IGV

To cite your use of IGV in your publication:

James T. Robinson, Helga Thorvaldsdóttir, Wendy Winckler, Mitchell Guttman, Eric S. Lander, Gad Getz, Jill P. Mesirov. [Integrative Genomics Viewer](#). *Nature Biotechnology* 29, 24–26 (2011).

Helga Thorvaldsdóttir, James T. Robinson, Jill P. Mesirov. [Integrative Genomics Viewer \(IGV\): high-performance genomics data visualization and exploration](#). *Briefings in Bioinformatics* 14, 178-192 (2013).

Funding

Development of IGV is made possible by funding from the [National Cancer Institute](#), the [National Institute of General Medical Sciences](#) of the [National Institutes of Health](#), and the [Starr Cancer Consortium](#).

IGV participates in the [GenomeSpace](#) initiative, which is funded by the [National Human Genome Research Institute](#).



W18-1: IGVダウンロード

①「Download Binary Distribution」をダウンロードすべく、右クリックで「ショートカットのコピー」などでURL情報を取得。

Home > Downloads

Downloads

Integrative Genomics Viewer (IGV) (Version 2.3)

Install IGV

Options for installing and running IGV:

1. (Mac only) Download and run the Mac application; or
2. (All systems) Use the Java Web Start buttons (Mac users: see below for limitations); or
3. (All systems) Download the binary distribution and run IGV from the command line.

1. Mac Application

Download and unzip the Mac App archive, then double-click the IGV application. The application can be moved to the "Applications" folder, or anywhere else. **Note: The 64-bit Mac OS X version of IGV requires Java 6 (JRE 1.6). Mac users with Java 6 (JRE 1.6) should use the binary distribution archive. Use the Java Web Start buttons below.**

2. Java Web Start

The buttons below use Java Web Start (JWS) to install and launch IGV directly from our web site.

***Mac Users:** The Java Web Start option is not recommended for Mac OSX Mountain Lion or higher. Using it requires that you set Gatekeeper security to its lowest level, and it is possible that even this will not be enough.

Chrome: Chrome does not automatically launch the Java Webstart files by default. Instead, the launch buttons below will download a "jnlp" file. This should appear in the lower left corner of the browser. Double-click the downloaded file to run.

Windows users: To run with more than 1.2 GB of memory you must install 64-bit Java. **Most Windows installs do not include 64-bit Java by default, even if the operating system is 64-bit.** Attempting to use the 2GB or greater launch options with 32-bit Java will result in the error "could not create virtual machine".

 Launch with 750 MB	 Launch with 1.2 GB Maximum usable memory for Windows OS with 32-bit Java.	 Launch with 2 GB Maximum usable memory for 32-bit MacOS.
--	---	--




2. Java Web Start

The buttons below use Java Web Start (JWS) to install and launch IGV directly from our web site.

***Mac Users:** The Java Web Start option is not recommended for Mac OSX Mountain Lion or higher. Using it requires that you set Gatekeeper security to its lowest level, and it is possible that even this will not be enough.


Chrome: Chrome does not automatically launch the Java Webstart files by default. Instead, the launch buttons below will download a "jnlp" file. This should appear in the lower left corner of the browser. Double-click the downloaded file to run.

Windows users: To run with more than 1.2 GB of memory you must install 64-bit Java. **Most Windows installs do not include 64-bit Java by default, even if the operating system is 64-bit.** Attempting to use the 2GB or greater launch options with 32-bit Java will result in the error "could not create virtual machine".

 Launch with 750 MB	 Launch with 1.2 GB Maximum usable memory for Windows OS with 32-bit Java.	 Launch with 2 GB Maximum usable memory for 32-bit MacOS.	 Launch with 10 GB For large memory machines with 64-bit Java.
---	---	--	---

3. Binary Distribution

Download and unzip the binary distribution archive in a folder of your choosing. IGV is launched from a command prompt -- follow instructions in the "readme" file. To launch igv on Mac or Linux platforms use the shell script "igv.sh". On Windows use "igv.bat".

 ①

igvtools

Utilities for preprocessing data files.

- [igvtools_2.3.57.zip](#)

Other IGV Versions

W18-1: IGVダウンロード

①wgetコマンドを用いてzipファイルのダウンロード。リターンキーを押す。



The image shows a terminal window with the following commands and output:

```
iu@bielinux[Downloads] pwd [ 7:34午後 ]
/home/iu/Downloads
iu@bielinux[Downloads] ls [ 7:34午後 ]
FaQCs FastQC fastqc_v0.11.3.zip
iu@bielinux[Downloads] wget -c http://data.broadinstitute.org/igv/projects/downloads/IGV_2.3.57.zip
```

A red arrow with the number '1' points to the 'wget' command line.

W18-1: IGVダウンロード

```
iu@bielinux[Downloads] pwd [ 7:34午後 ]
/home/iu/Downloads
iu@bielinux[Downloads] ls [ 7:34午後 ]
FaQCs FastQC fastqc_v0.11.3.zip
iu@bielinux[Downloads] wget -c http://data.broadinstitute.org/igv/pro
jcts/downloads/IGV_2.3.57.zip
--2015-06-25 19:35:58-- http://data.broadinstitute.org/igv/projects/
downloads/IGV_2.3.57.zip
Resolving data.broadinstitute.org (data.broadinstitute.org)... 69.173
.92.29
Connecting to data.broadinstitute.org (data.broadinstitute.org)|69.17
3.92.29|:80... connected.
HTTP request sent, awaiting response... 307 Moved Temporarily
Location: https://data.broadinstitute.org/igv/projects/downloads/IGV_
2.3.57.zip? [following]
--2015-06-25 19:35:59-- https://data.broadinstitute.org/igv/projects
/downloads/IGV_2.3.57.zip?
Connecting to data.broadinstitute.org (data.broadinstitute.org)|69.17
3.92.29|:443... connected.
HTTP request sent, awaiting response... 200 OK
Length: 30359749 (29M) [application/zip]
Saving to: 'IGV_2.3.57.zip'

15% [====> ] 4,702,208 327KB/s eta 77s
```

W18-1: IGVダウンロード

この環境では、ダウンロードに91秒かかったことがわかる。IGV_2.3.57.zipは約30MB。

```
File Edit View Search Terminal Help
--2015-06-25 19:35:59-- https://data.broadinstitute.org/igv/projects
/downloads/IGV_2.3.57.zip?
Connecting to data.broadinstitute.org (data.broadinstitute.org)|69.17
3.92.29|:443... connected.
HTTP request sent, awaiting response... 200 OK
Length: 30359749 (29M) [application/zip]
Saving to: 'IGV_2.3.57.zip'

100%[=====>] 30,359,749 327KB/s in 91s

2015-06-25 19:37:31 (327 KB/s) - 'IGV_2.3.57.zip' saved [30359749/303
59749]

iu@bielinux[Downloads] ls -la [ 7:37午後 ]
total 39460
drwxr-xr-x 4 iu iu 4096 6月 25 19:36 .
drwxr-xr-x 23 iu iu 4096 6月 25 17:55 ..
drwxrwxr-x 6 iu iu 4096 5月 29 17:03 FaQCs
drwxrwxr-x 8 iu iu 4096 5月 7 14:46 FastQC
-rw-rw-r-- 1 iu iu 10024552 3月 25 18:40 fastqc_v0.11.3.zip
-rw-rw-r-- 1 iu iu 30359749 6月 24 01:17 IGV_2.3.57.zip
iu@bielinux[Downloads] ls -lh IGV_2.3.57.zip [ 7:37午後 ]
-rw-rw-r-- 1 iu iu 29M 6月 24 01:17 IGV_2.3.57.zip
iu@bielinux[Downloads] █ [ 7:38午後 ]
```

①

②

③

W18-2: IGV解凍

- ①zip圧縮ファイル(IGV_2.3.57.zip)を解凍。
- ③IGV_2.3.57というディレクトリが作成されていることがわかります。

```
iu@bielinux[Downloads] pwd [ 7:38午後 ]
/home/iu/Downloads
iu@bielinux[Downloads] ls [ 7:41午後 ]
FaQCs FastQC fastqc_v0.11.3.zip IGV_2.3.57.zip
iu@bielinux[Downloads] unzip IGV_2.3.57.zip [ 7:41午後 ]
Archive:  IGV_2.3.57.zip
  creating:  IGV_2.3.57/
  inflating:  IGV_2.3.57/igv.bat
  inflating:  IGV_2.3.57/igv.command
  inflating:  IGV_2.3.57/igv.sh
  inflating:  IGV_2.3.57/batik-codec__V1.7.jar
  inflating:  IGV_2.3.57/goby-io-igv__V1.0.jar
  inflating:  IGV_2.3.57/igv.jar
  inflating:  IGV_2.3.57/readme.txt
iu@bielinux[Downloads] ls -la [ 7:41午後 ]
total 39464
drwxr-xr-x  5 iu iu      4096  6月 25 19:41 .
drwxr-xr-x 23 iu iu      4096  6月 25 17:55 ..
drwxrwxr-x  6 iu iu      4096  5月 29 17:03 FaQCs
drwxrwxr-x  8 iu iu      4096  5月  7 14:46 FastQC
-rw-rw-r--  1 iu iu 10024552  3月 25 18:40 fastqc_v0.11.3.zip
drwxr-xr-x  2 iu iu      4096  6月 23 12:17 IGV_2.3.57
-rw-rw-r--  1 iu iu 30359749  6月 24 01:17 IGV_2.3.57.zip
iu@bielinux[Downloads] [ 7:43午後 ]
```

W18-3: IGVパスを通す

①igv.shのシンボリックリンクを
/usr/local/binに置く。パスワードを聞か
れたら、ログインパスワードを打ち込む。

```
iu@bielinux[Downloads] cd IGV_2.3.57 [ 7:49午後 ]
iu@bielinux[IGV_2.3.57] pwd [ 7:49午後 ]
/home/iu/Downloads/IGV_2.3.57
iu@bielinux[IGV_2.3.57] ls -la [ 7:49午後 ]
total 32524
drwxr-xr-x 2 iu iu 4096 6月 23 12:17 .
drwxr-xr-x 5 iu iu 4096 6月 25 19:41 ..
-rw-r--r-- 1 iu iu 177245 6月 23 12:17 batik-codec_V1.7.jar
-rw-r--r-- 1 iu iu 2120141 6月 23 12:17 goby-io-igv_V1.0.jar
-rwxr-xr-x 1 iu iu 258 6月 23 12:17 igv.bat
-rwxr-xr-x 1 iu iu 696 6月 23 12:17 igv.command
-rw-r--r-- 1 iu iu 30975040 6月 23 12:17 igv.jar
-rwxr-xr-x 1 iu iu 493 6月 23 12:17 igv.sh
-rw-r--r-- 1 iu iu 1238 6月 23 12:17 readme.txt
① iu@bielinux[IGV_2.3.57] sudo ln -s /home/iu/Downloads/IGV_2.3.57/igv.
sh /usr/local/bin
[sudo] password for iu:
iu@bielinux[IGV_2.3.57] [ 7:53午後 ]
```

W18-3: IGVパスを通す

①rehashは、設定変更(この場合シンボリックリンク)を現在開いているターミナル上で反映させるためコマンドです。ウェブブラウザのリロードのようなものです。②「igv.sh」と打ってリターンキーを押す。IGVが無事起動すればOKです。

```
iu@bielinux[Downloads] cd IGV_2.3.57
iu@bielinux[IGV_2.3.57] pwd
/home/iu/Downloads/IGV_2.3.57
iu@bielinux[IGV_2.3.57] ls -la
total 32524
drwxr-xr-x 2 iu iu      4096  6月 23 12:17 .
drwxr-xr-x 5 iu iu      4096  6月 25 19:41 ..
-rw-r--r-- 1 iu iu    177245  6月 23 12:17 batik-codec__V1.7.jar
-rw-r--r-- 1 iu iu   2120141  6月 23 12:17 goby-io-igv__V1.0.jar
-rwxr-xr-x 1 iu iu       258  6月 23 12:17 igv.bat
-rwxr-xr-x 1 iu iu       696  6月 23 12:17 igv.command
-rw-r--r-- 1 iu iu  30975040  6月 23 12:17 igv.jar
-rwxr-xr-x 1 iu iu       493  6月 23 12:17 igv.sh
-rw-r--r-- 1 iu iu     1238  6月 23 12:17 readme.txt
iu@bielinux[IGV_2.3.57] sudo ln -s /home/iu/Downloads/IGV_2.3.57/igv.sh /usr/local/bin
[sudo] password for iu:
iu@bielinux[IGV_2.3.57] rehash
iu@bielinux[IGV_2.3.57] igv.sh
```

[7:53午後]
[8:00午後]



W18-4: IGV起動

リターンキーを押して、少し時間が経過(十秒程度)ときの状態です。

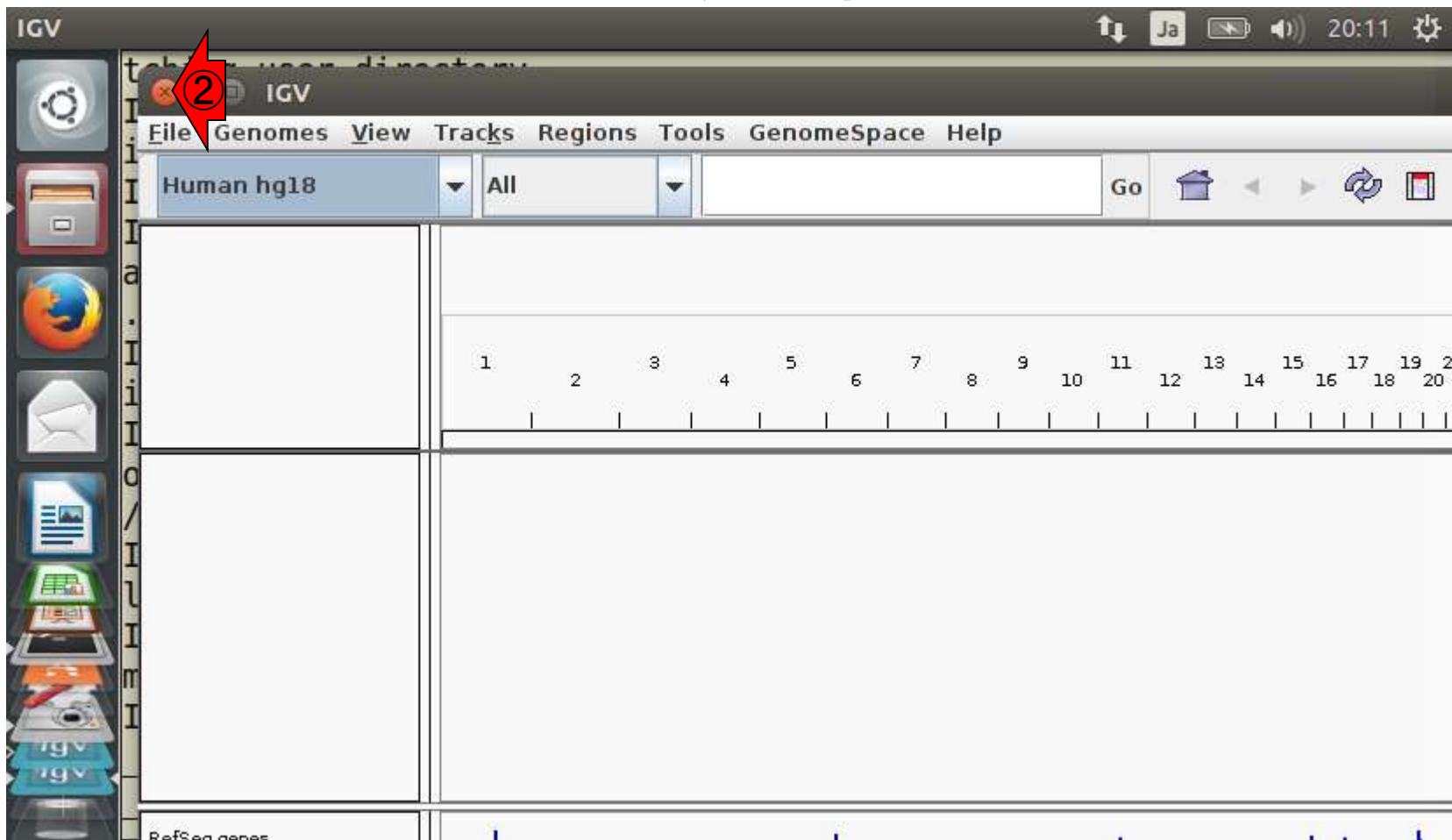
```
File Edit View Search Terminal Help 20:09
sh /usr/local/bin
[sudo] password for iu:
iu@bielinux[IGV_2.3.57] rehash [ 7:53午後]
iu@bielinux[IGV_2.3.57] igv.sh [ 8:00午後]
Jun 25, 2015 8:08:53 PM java.util.prefs.FileSystemPreferences$1 run
INFO: Created user preferences directory.
INFO [2015-06-25 20:08:53,729] [DirectoryManager.java:149] [main] I
GV Directory: /home/iu/igv
INFO [2015-06-25 20:08:53,729] [DirectoryManager.java:149] IGV Direc
tory: /home/iu/igv
INFO [2015-06-25 20:08:53,767] [Main.java:85] [main] Startup IGV V
ersion 2.3.57 (84)06/23/2015 12:09 PM
INFO [2015-06-25 20:08:53,775] [Main.java:86] [main] Java 1.7.0_55
INFO [2015-06-25 20:08:53,775] [DirectoryManager.java:58] [main] Fe
tching user directory...
INFO [2015-06-25 20:08:54,829] [Main.java:87] [main] Default User D
irectory: /home/iu
INFO [2015-06-25 20:08:54,830] [Main.java:88] [main] OS: Linux
```



②

W18-4: IGV起動と終了

こんな感じでIGVが起動すればOK。
①×を押して閉じる。



コマンド入力待ち状態
になっているはずですが。

W18-5: IGV終了後

```
File Edit View Search Terminal Help 20:14
tching user directory...
INFO [2015-06-25 20:08:54,829] [Main.java:87] [main] Default User Directory: /home/iu
INFO [2015-06-25 20:08:54,830] [Main.java:88] [main] OS: Linux
INFO [2015-06-25 20:09:25,728] [GenomeManager.java:1002] [main] java.io.FileNotFoundException: /home/iu/igv/genomes/user-defined-genomes.txt (No such file or directory)
INFO [2015-06-25 20:09:28,022] [GenomeManager.java:130] [main] Loading genome: http://igvdata.broadinstitute.org/genomes/hg18.genome
INFO [2015-06-25 20:09:28,086] [URLDownloader.java:64] [main] Downloading http://igvdata.broadinstitute.org/genomes/hg18.genome to /home/iu/igv/genomes/hg18.genome
INFO [2015-06-25 20:09:28,872] [URLDownloader.java:105] [main] Download complete. Total bytes downloaded = 4998108
INFO [2015-06-25 20:09:33,060] [GenomeManager.java:177] [main] Genome loaded. id= hg18
INFO [2015-06-25 20:09:33,779] [CommandListener.java:91] [Thread-6] Listening on port 60151
iu@bielinux[IGV_2.3.57] [ 8:12午後]
```

