

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97GB確保しており、そのうち11GB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だと1.1GB使用していることもわかる。

```
File Edit View Search Terminal Help
① iu@bielinux[srp017156] pwd
/home/iu/Documents/srp017156
② iu@bielinux[srp017156] df -h
Filesystem      Size  Used Avail Use% Mounted on
/dev/sda1        97G   11G   81G   13% /
none             4.0K    0   4.0K    0% /sys/fs/cgroup
udev            991M    4.0K 991M    1% /dev
tmpfs           201M   900K 200M    1% /run
none            5.0M    0   5.0M    0% /run/lock
none           1001M   156K 1001M    1% /run/shm
none            100M    32K  100M    1% /run/user

③ iu@bielinux[srp017156] ls -lh
total 1.1G
-rw-rw-r-- 1 iu iu 461M 12月  7 15:42 hoge_1.fastq
-rw-rw-r-- 1 iu iu 306M 12月  7 15:53 hoge_subset_1.fastq
-rw-rw-r-- 1 iu iu  83M 12月  3 14:56 SRR616268_1.fastq.bz2
-rw-rw-r-- 1 iu iu  74M 12月  3 15:51 SRR616268_2.fastq.bz2
-rw-rw-r-- 1 iu iu  74M 12月  7 15:47 subset_1.fastq.gz
-rw-rw-r-- 1 iu iu  66M 12月  7 15:47 subset_2.fastq.gz
iu@bielinux[srp017156] █
```

[1:44午後]

[1:44午後]

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。

```
iu@bielinux[srp017156] pwd
/home/iu/Documents/srp017156
iu@bielinux[srp017156] df -h
Filesystem      Size  Used Avail Use% Mounted on
/dev/sda1        97G   11G   81G   13% /
none             4.0K    0 4.0K    0% /sys/fs/cgroup
udev            991M   4.0K 991M    1% /dev
tmpfs           201M  900K 200M    1% /run
none            5.0M    0 5.0M    0% /run/lock
none           1001M  156K 1001M    1% /run/shm
none            100M   32K  100M    1% /run/user
iu@bielinux[srp017156] ls -lh
total 1.1G
-rw-rw-r-- 1 iu iu 461M 12月  7 15:42 hoge_1.fastq
-rw-rw-r-- 1 iu iu 306M 12月  7 15:53 hoge_subset_1.fastq
-rw-rw-r-- 1 iu iu  83M 12月  3 14:56 SRR616268_1.fastq.bz2
-rw-rw-r-- 1 iu iu  74M 12月  3 15:51 SRR616268_2.fastq.bz2
-rw-rw-r-- 1 iu iu  74M 12月  7 15:47 subset_1.fastq.gz
-rw-rw-r-- 1 iu iu  66M 12月  7 15:47 subset_2.fastq.gz
iu@bielinux[srp017156] █
```

[1:44午後]

[1:44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①「ls -lh ho」まで打ち込んでからTabキーを押す(タブ補完)。

```
iu@bielinux[srp017156] pwd
/home/iu/Documents/srp017156
iu@bielinux[srp017156] df -h
Filesystem      Size  Used Avail Use%
/dev/sda1       97G   11G   81G  13%
none            4.0K    0  4.0K   0%
udev            991M   4.0K 991M   1%
tmpfs           201M  900K 200M   1%
none            5.0M    0  5.0M   0%
none           1001M  156K 1001M   1%
none            100M   32K  100M   1%
iu@bielinux[srp017156] ls -lh
total 1.1G
-rw-rw-r-- 1 iu iu 461M 12月  7 15:42 hoge_1.fastq
-rw-rw-r-- 1 iu iu 306M 12月  7 15:53 hoge_subset_1.fastq
-rw-rw-r-- 1 iu iu  83M 12月  3 14:56 SRR616268_1.fastq.bz2
-rw-rw-r-- 1 iu iu  74M 12月  3 15:51 SRR616268_2.fastq.bz2
-rw-rw-r-- 1 iu iu  74M 12月  7 15:47 subset_1.fastq.gz
-rw-rw-r-- 1 iu iu  66M 12月  7 15:47 subset_2.fastq.gz
iu@bielinux[srp017156] ls -lh ho
```

[1:44午後]

[1:44午後]

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…。

```
iu@bielinux[srp017156] pwd
/home/iu/Documents/srp017156
iu@bielinux[srp017156] df -h
Filesystem      Size  Used Avail Use% Mounted on
/dev/sda1        97G   11G   81G  13% /
none             4.0K    0 4.0K   0% /sys/fs/cgroup
udev            991M   4.0K 991M   1% /dev
tmpfs           201M  900K 200M   1% /run
none            5.0M    0 5.0M   0% /run/lock
none           1001M  156K 1001M   1% /run/shm
none            100M   32K 100M   1% /run/user
iu@bielinux[srp017156] ls -lh
total 1.1G
-rw-rw-r-- 1 iu iu 461M 12月  7 15:42 hoge_1.fastq
-rw-rw-r-- 1 iu iu 306M 12月  7 15:53 hoge_subset_1.fastq
-rw-rw-r-- 1 iu iu  83M 12月  3 14:56 SRR616268_1.fastq.bz2
-rw-rw-r-- 1 iu iu  74M 12月  3 15:51 SRR616268_2.fastq.bz2
-rw-rw-r-- 1 iu iu  74M 12月  7 15:47 subset_1.fastq.gz
-rw-rw-r-- 1 iu iu  66M 12月  7 15:47 subset_2.fastq.gz
iu@bielinux[srp017156] ls -lh hoge_
```

[1:44午後]

[1:44午後]

W1-2

①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[srp017156] pwd [ 1:44午後 ]
/home/iu/Documents/srp017156
iu@bielinux[srp017156] df -h [ 1:44午後 ]
Filesystem      Size  Used Avail Use% Mounted on
/dev/sda1        97G   11G   81G   13% /
none             4.0K    0 4.0K   0% /sys/fs/cgroup
udev            991M   4.0K 991M   1% /dev
tmpfs           201M  900K 200M   1% /run
none            5.0M    0 5.0M   0% /run/lock
none           1001M  156K 1001M   1% /run/shm
none            100M   32K 100M   1% /run/user
iu@bielinux[srp017156] ls -lh [ 1:44午後 ]
total 1.1G
-rw-rw-r-- 1 iu iu 461M 12月  7 15:42 hoge_1.fastq
-rw-rw-r-- 1 iu iu 306M 12月  7 15:53 hoge_subset_1.fastq
-rw-rw-r-- 1 iu iu  83M 12月  3 14:56 SRR616268_1.fastq.bz2
-rw-rw-r-- 1 iu iu  74M 12月  3 15:51 SRR616268_2.fastq.bz2
-rw-rw-r-- 1 iu iu  74M 12月  7 15:47 subset_1.fastq.gz
-rw-rw-r-- 1 iu iu  66M 12月  7 15:47 subset_2.fastq.gz
① iu@bielinux[srp017156] ls -lh hoge_* [ 1:44午後 ]
-rw-rw-r-- 1 iu iu 461M 12月  7 15:42 hoge_1.fastq
-rw-rw-r-- 1 iu iu 306M 12月  7 15:53 hoge_subset_1.fastq
iu@bielinux[srp017156] [ 1:53午後 ]
```


W1-3

①rm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。

```
iu@bielinux[srp017156] pwd
/home/iu/Documents/srp017156
iu@bielinux[srp017156] df -h
Filesystem      Size  Used Avail U
/dev/sda1       97G   11G   81G
none            4.0K    0  4.0K
udev            991M   4.0K  991M
tmpfs           201M  900K  200M   1% /run
none            5.0M    0  5.0M   0% /run/lock
none           1001M  156K 1001M   1% /run/shm
none            100M   32K  100M   1% /run/user
iu@bielinux[srp017156] ls -lh
total 1.1G
-rw-rw-r-- 1 iu iu 461M 12月  7 15:42 hoge_1.fastq
-rw-rw-r-- 1 iu iu 306M 12月  7 15:53 hoge_subset_1.fastq
-rw-rw-r-- 1 iu iu  83M 12月  3 14:56 SRR616268_1.fastq.bz2
-rw-rw-r-- 1 iu iu  74M 12月  3 15:51 SRR616268_2.fastq.bz2
-rw-rw-r-- 1 iu iu  74M 12月  7 15:47 subset_1.fastq.gz
-rw-rw-r-- 1 iu iu  66M 12月  7 15:47 subset_2.fastq.gz
iu@bielinux[srp017156] ls -lh hoge_*
-rw-rw-r-- 1 iu iu 461M 12月  7 15:42 hoge_1.fastq
-rw-rw-r-- 1 iu iu 306M 12月  7 15:53 hoge_subset_1.fastq
iu@bielinux[srp017156] rm -f h
```

[1:44午後]

[1:44午後]

[1:53午後]



W1-3

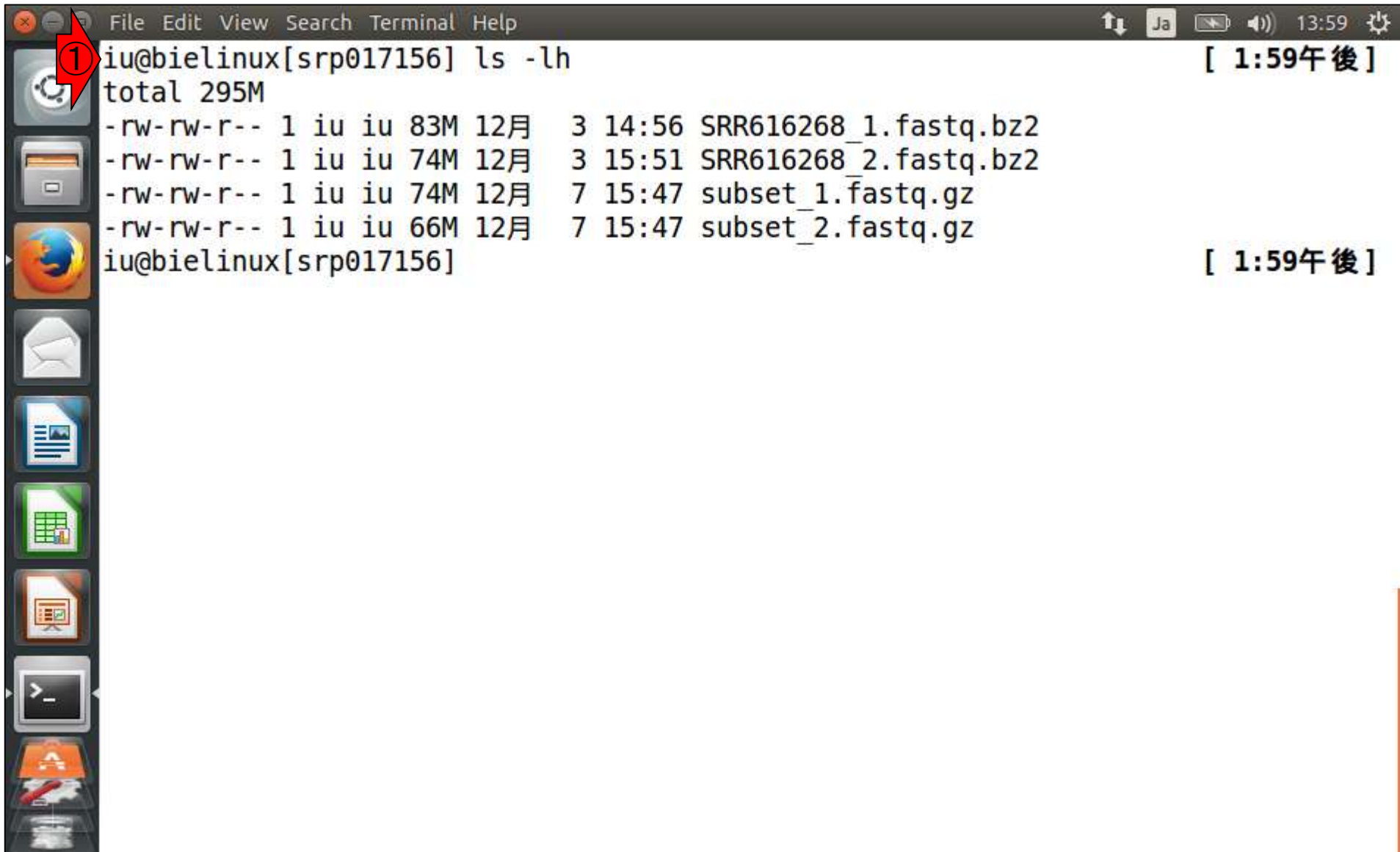
①Tabキーを押して「hoge_」まで表示させてから、アスタリスク(*)をつけたところ。この段階でリターン。

```
iu@bielinux[srp017156] pwd [ 1:44午後 ]
/home/iu/Documents/srp017156
iu@bielinux[srp017156] df -h [ 1:44午後 ]
Filesystem      Size  Used Avail Use% Mounted on
/dev/sda1        97G   11G   81G  13% /
none             4.0K    0 4.0K   0% /sys/fs/cgroup
udev            991M   4.0K 991M   1% /dev
tmpfs           201M  900K 200M   1% /run
none            5.0M    0 5.0M   0% /run/lock
none           1001M  156K 1001M   1% /run/shm
none           100M   32K 100M   1% /run/user
iu@bielinux[srp017156] ls -lh [ 1:44午後 ]
total 1.1G
-rw-rw-r-- 1 iu iu 461M 12月  7 15:42 hoge_1.fastq
-rw-rw-r-- 1 iu iu 306M 12月  7 15:53 hoge_subset_1.fastq
-rw-rw-r-- 1 iu iu  83M 12月  3 14:56 SRR616268_1.fastq.bz2
-rw-rw-r-- 1 iu iu  74M 12月  3 15:51 SRR616268_2.fastq.bz2
-rw-rw-r-- 1 iu iu  74M 12月  7 15:47 subset_1.fastq.gz
-rw-rw-r-- 1 iu iu  66M 12月  7 15:47 subset_2.fastq.gz
iu@bielinux[srp017156] ls -lh hoge_* [ 1:44午後 ]
-rw-rw-r-- 1 iu iu 461M 12月  7 15:42 hoge_1.fastq
-rw-rw-r-- 1 iu iu 306M 12月  7 15:53 hoge_subset_1.fastq
iu@bielinux[srp017156] rm -f hoge_* [ 1:53午後 ]
iu@bielinux[srp017156] [ 1:58午後 ]
```



W1-3

①「ls -lh」。ファイル群hoge_*を削除できていることがわかる。



```
iu@bielinux[srp017156] ls -lh
total 295M
-rw-rw-r-- 1 iu iu 83M 12月  3 14:56 SRR616268_1.fastq.bz2
-rw-rw-r-- 1 iu iu 74M 12月  3 15:51 SRR616268_2.fastq.bz2
-rw-rw-r-- 1 iu iu 74M 12月  7 15:47 subset_1.fastq.gz
-rw-rw-r-- 1 iu iu 66M 12月  7 15:47 subset_2.fastq.gz
iu@bielinux[srp017156]
```

W1-3

一連の作業をファイル群 subset_*についても実行。

```
iu@bielinux[srp017156] ls -lh
total 295M
-rw-rw-r-- 1 iu iu 83M 12月  3 14:56 SRR616268_1.fastq.bz2
-rw-rw-r-- 1 iu iu 74M 12月  3 15:51 SRR616268_2.fastq.bz2
-rw-rw-r-- 1 iu iu 74M 12月  7 15:47 subset_1.fastq.gz
-rw-rw-r-- 1 iu iu 66M 12月  7 15:47 subset_2.fastq.gz
iu@bielinux[srp017156] ls -lh subset_*
-rw-rw-r-- 1 iu iu 74M 12月  7 15:47 subset_1.fastq.gz
-rw-rw-r-- 1 iu iu 66M 12月  7 15:47 subset_2.fastq.gz
iu@bielinux[srp017156] rm -f subset_*
iu@bielinux[srp017156] ls -lh
total 157M
-rw-rw-r-- 1 iu iu 83M 12月  3 14:56 SRR616268_1.fastq.bz2
-rw-rw-r-- 1 iu iu 74M 12月  3 15:51 SRR616268_2.fastq.bz2
iu@bielinux[srp017156] pwd
/home/iu/Documents/srp017156
iu@bielinux[srp017156]
```

[2:02午後]
[2:02午後]
[2:02午後]
[2:02午後]
[2:02午後]
[2:02午後]
[2:02午後]

W1-4

連載第4回の図1とほぼ同じ。①duでカレントディレクトリのディスク使用量が157MB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」の意味。②dfで全体のディスク使用量を表示。赤枠部分のSizeが97GBとなっているのは100GBを割り当てていたから。Usedは全体で11GB使っていることを示す。この軽量版環境は、多くの読者が以後の解析が可能と期待!

```
iu@bielinux[srp017156] pwd
/home/iu/Documents/srp017156
iu@bielinux[srp017156] ls -lh
total 157M
-rw-rw-r-- 1 iu iu 83M 12月  3 14:56 SRR616268_1.fastq.bz2
-rw-rw-r-- 1 iu iu 74M 12月  3 15:51 SRR616268_2.fastq.bz2
iu@bielinux[srp017156] du -h
157M  .
iu@bielinux[srp017156] df -h
Filesystem      Size  Used Avail Use% Mounted on
/dev/sda1       97G   11G   82G   12% /
none            4.0K    0 4.0K   0% /sys/fs/cgroup
udev           991M   4.0K 991M   1% /dev
tmpfs          201M   900K 200M   1% /run
none           5.0M    0 5.0M   0% /run/lock
none          1001M  156K 1001M   1% /run/shm
none           100M   32K  100M   1% /run/user
iu@bielinux[srp017156]
```

[2:20午後]

[2:20午後]

[2:20午後]

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。

W2-1: シェルスクリプト

The screenshot displays a Linux desktop environment. In the foreground, a terminal window titled "iu@bielinux[~/Documents/srp017156]" shows the following commands and their outputs:

```
iu@bielinux[ srp017156 ] pwd [ 2:35午後 ]
/home/iu/Documents/srp017156
iu@bielinux[ srp017156 ] ls -lh [ 2:35午後 ]
total 157M
-rw-rw-r-- 1 iu iu 83M 12月 3 14:56 SRR616268_1.fastq.bz2
-rw-rw-r-- 1 iu iu 74M 12月 3 15:51 SRR616268_2.fastq.bz2
iu@bielinux[ srp017156 ] gedit [ 2:35午後 ]
```

A red circle with the number "1" is placed over the terminal prompt, indicating the step of launching gedit. Below the terminal, a file manager window shows a folder named "Sample Data". In the foreground, a gedit editor window titled "Untitled Document 1 - gedit" is open, showing a blank document with a menu bar (Open, Save, Undo) and a status bar (Plain Text, Tab Width: 8, Ln 1, Col 1, INS).

W2-1: シェルスクリプト

①(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、②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。③Saveボタンを押す。

The image shows a Linux desktop environment with a terminal window and a gedit editor window. The terminal window displays the following commands and output:

```
iu@bielinux[~/Documents/srp017156]
iu@bielinux[ srp017156] pwd
/home/iu/Documents/srp017156
iu@bielinux[ srp017156] ls -lh
total 157M
-rw-rw-r-- 1 iu iu 83M 12月  3 14:56 SRR616268_1.fastq.bz2
-rw-rw-r-- 1 iu iu 74M 12月  3 15:51 SRR616268_2.fastq.bz2
iu@bielinux[ srp017156] gedit
[ 2:35午後]
```

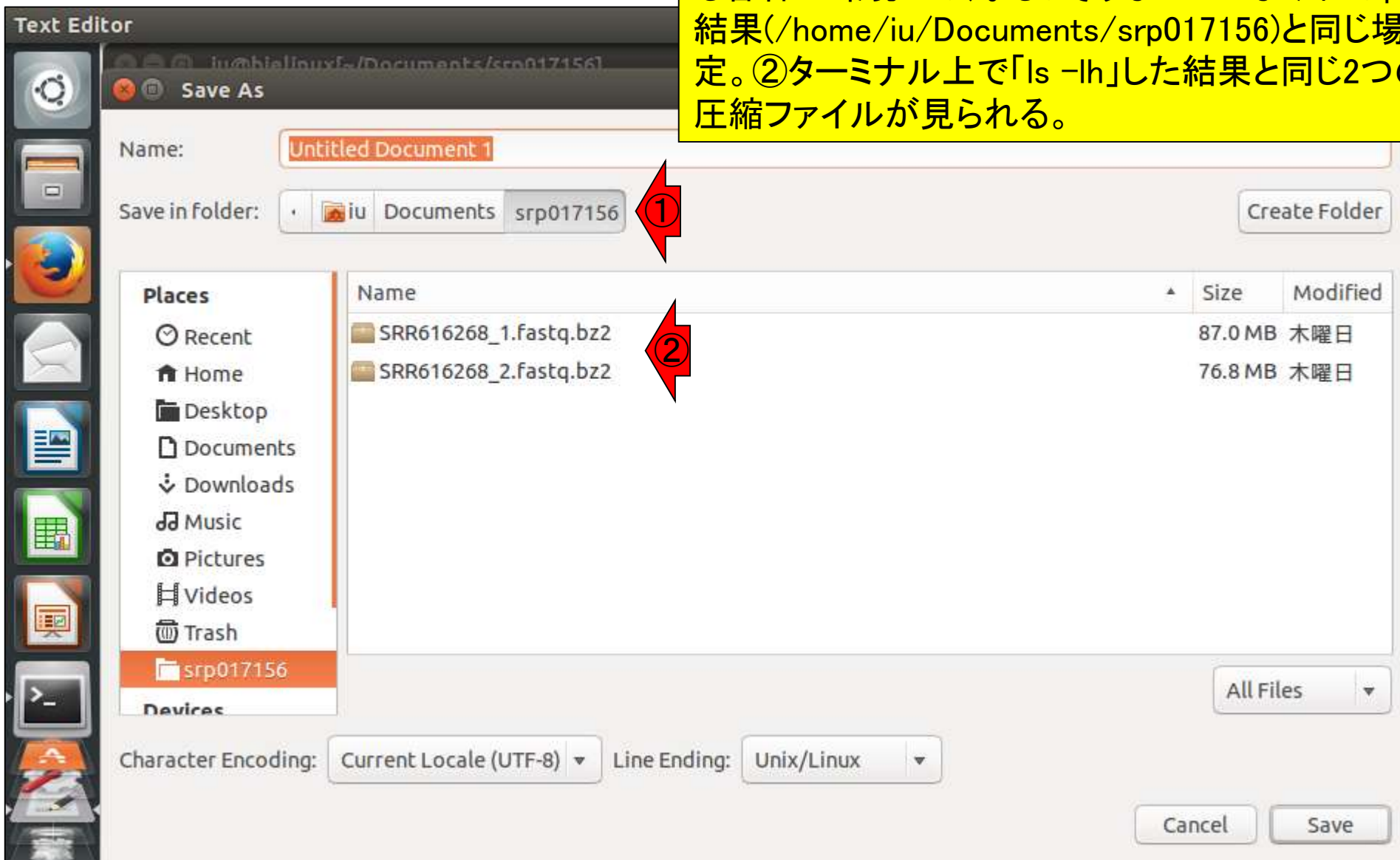
The gedit editor window shows the following content:

```
*Untitled Document 1 - gedit
*Untitled Document 1 x
pwd
date
ls
|
```

Red arrows and numbers indicate the steps: ① points to the first command in the terminal, ② points to the empty prompt in the terminal, and ③ points to the Save button in the gedit editor.

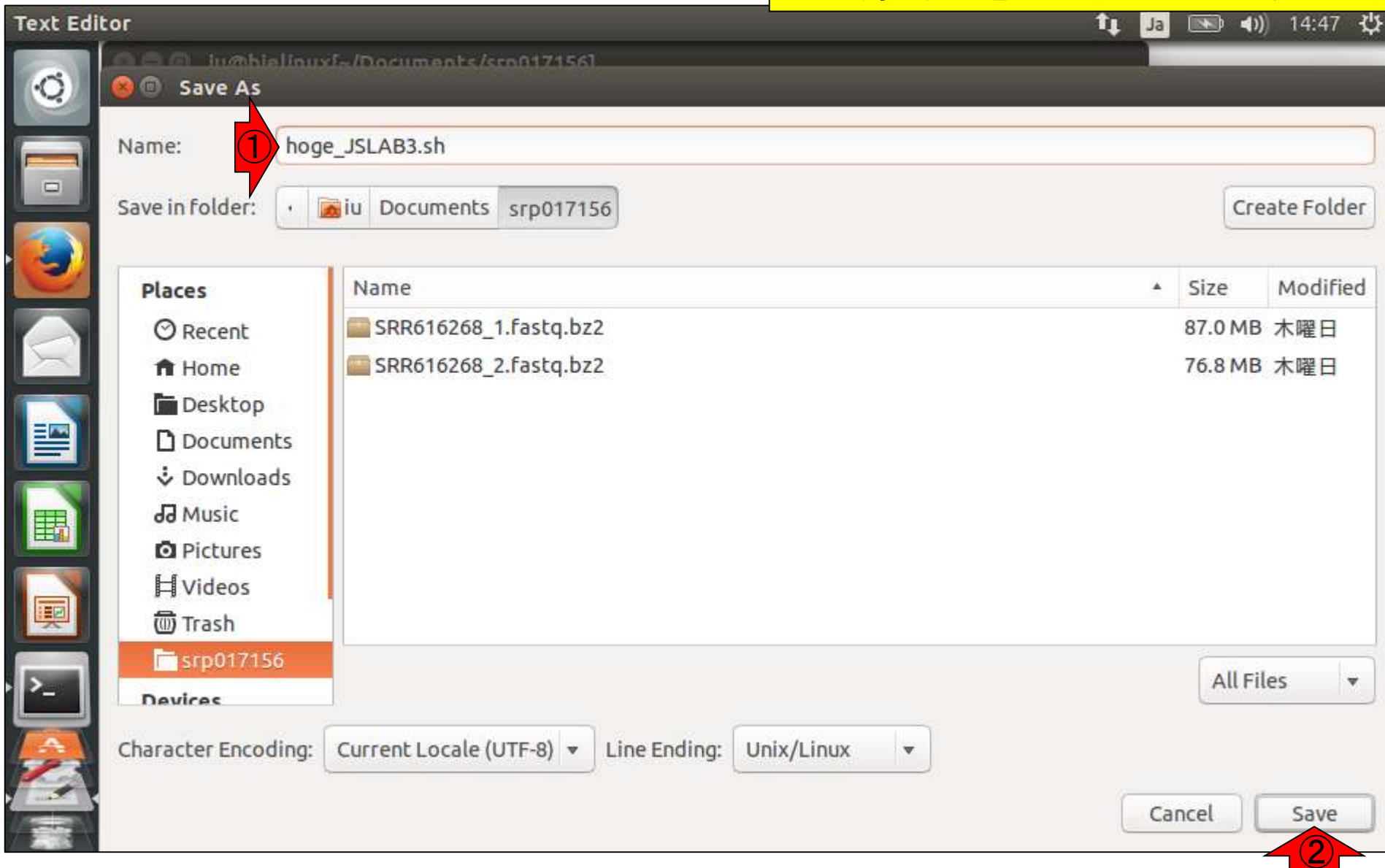
W2-1

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、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。

Terminal output:

```
iu@bielinux[~:/Documents/srp017156]
iu@bielinux[srp017156] pwd [ 2:35午後 ]
/home/iu/Documents/srp017156
iu@bielinux[srp017156] ls -lh [ 2:35午後 ]
total 157M
-rw-rw-r-- 1 iu iu 83M 12月 3 14:56 SRR616268_1.fastq.bz2
-rw-rw-r-- 1 iu iu 74M 12月 3 15:51 SRR616268_2.fastq.bz2
iu@bielinux[srp017156] gedit [ 2:35午後 ]
```

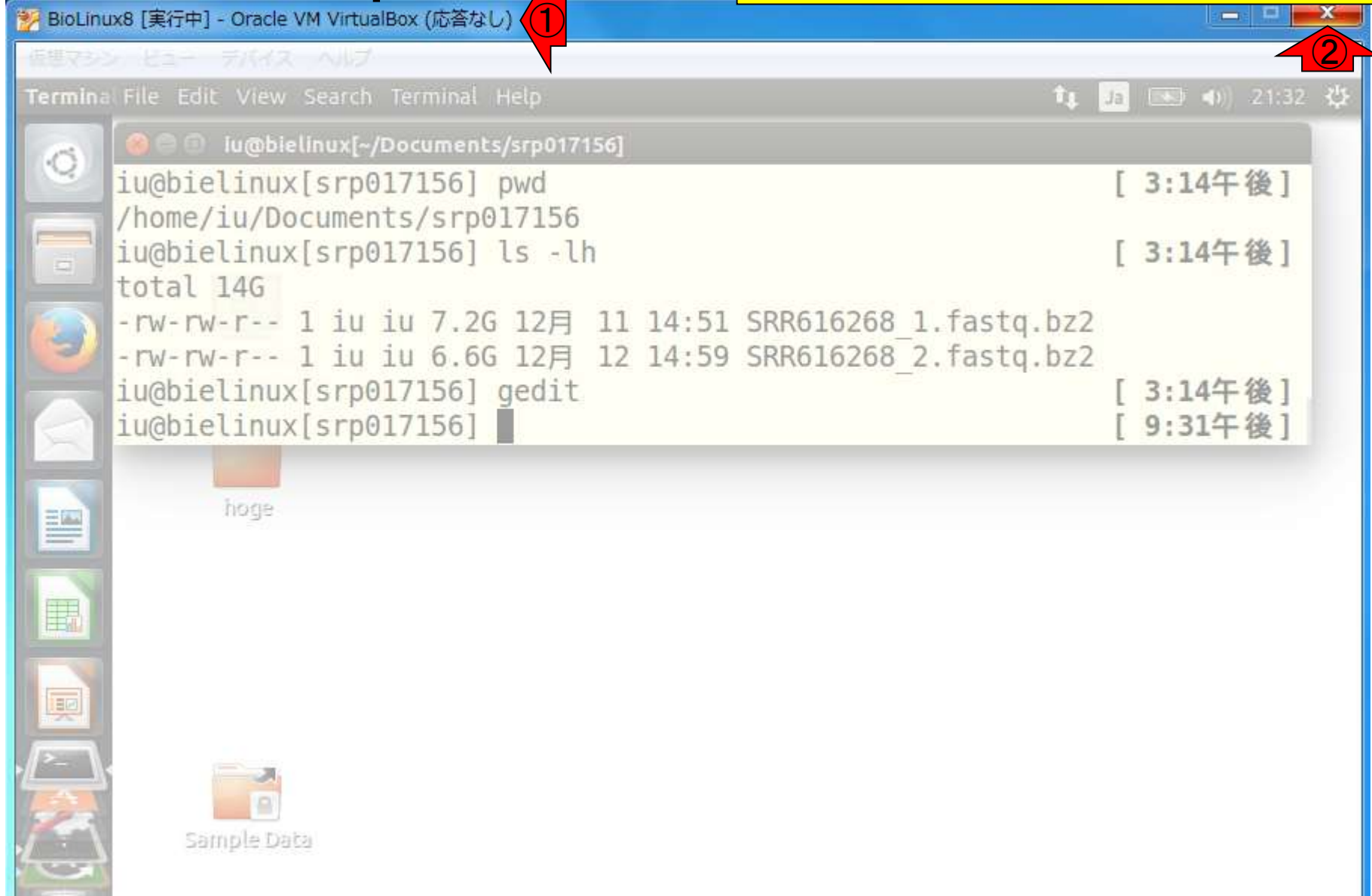
gedit window title: hoge_JSLAB3.sh (~:/Documents/srp017156) - gedit

gedit content:

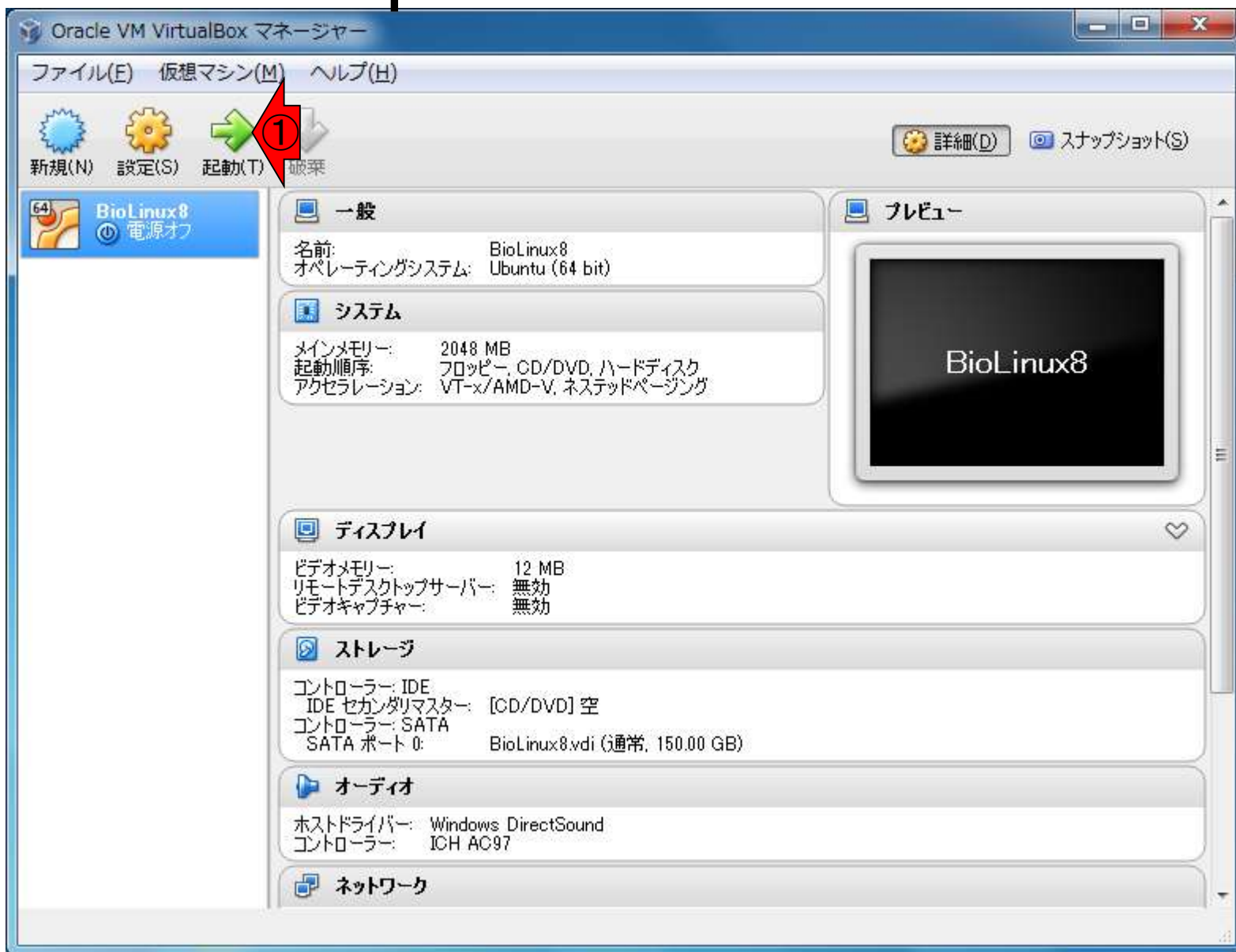
```
pwd
date
ls
```


W2-1: Tips

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②右上の×ボタンを押す以外の選択肢はありません

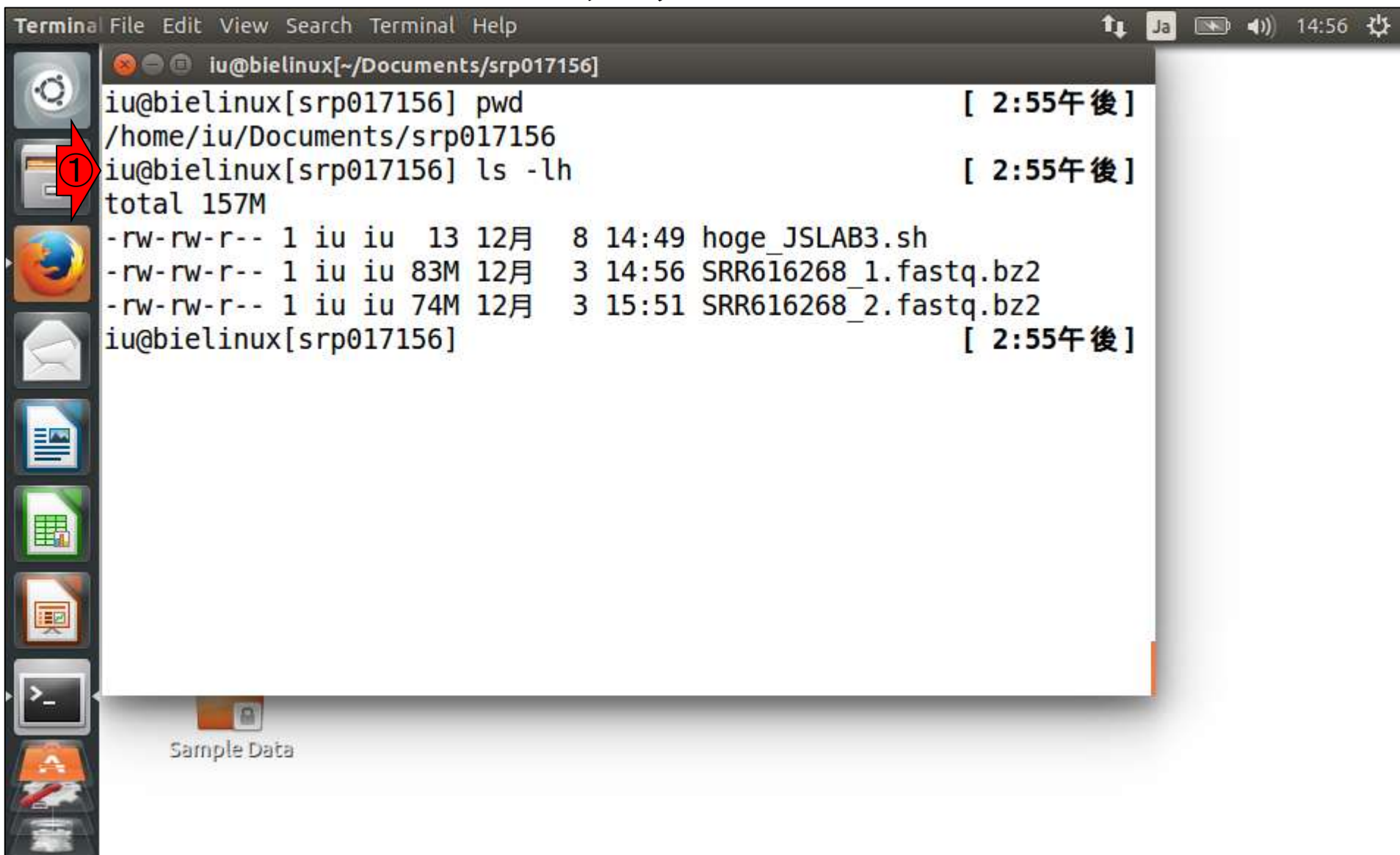


W2-1 : Tips



W2-2: シェルスクリプト

①lsで確認。geditで作成し保存した
hoge_JSLAB3.shで見られることがわかる。



```
Terminal File Edit View Search Terminal Help
iu@bielinux[~/Documents/srp017156]
iu@bielinux[srp017156] pwd [ 2:55午後 ]
/home/iu/Documents/srp017156
iu@bielinux[srp017156] ls -lh [ 2:55午後 ]
total 157M
-rw-rw-r-- 1 iu iu 13 12月 8 14:49 hoge_JSLAB3.sh
-rw-rw-r-- 1 iu iu 83M 12月 3 14:56 SRR616268_1.fastq.bz2
-rw-rw-r-- 1 iu iu 74M 12月 3 15:51 SRR616268_2.fastq.bz2
iu@bielinux[srp017156] [ 2:55午後 ]
```

The terminal window shows the execution of the following commands and their outputs:

- `pwd` returns `/home/iu/Documents/srp017156`.
- `ls -lh` lists the files in the current directory, including `hoge_JSLAB3.sh`, `SRR616268_1.fastq.bz2`, and `SRR616268_2.fastq.bz2`.

A red arrow with the number 1 points to the `ls -lh` command in the terminal output.

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。

```
Terminal File Edit View Search Terminal Help
iu@bielinux[~/Documents/srp017156]
iu@bielinux[srp017156] pwd
/home/iu/Documents/srp017156
iu@bielinux[srp017156] ls -lh
total 157M
-rw-rw-r-- 1 iu iu 13 12月 8 14:49 hoge_JSLAB3.sh
-rw-rw-r-- 1 iu iu 83M 12月 3 14:56 SRR616268_1.fastq.bz2
-rw-rw-r-- 1 iu iu 74M 12月 3 15:51 SRR616268_2.fastq.bz2
① iu@bielinux[srp017156] more hoge_JSLAB3.sh [ 2:55午後 ]
pwd
date
ls
iu@bielinux[srp017156] less hoge_JSLAB3.sh [ 3:01午後 ]
③ iu@bielinux[srp017156] sh hoge_JSLAB3.sh [ 3:01午後 ]
/home/iu/Documents/srp017156
2015年 12月 8日 火曜日 15:01:14 JST
hoge_JSLAB3.sh SRR616268_1.fastq.bz2 SRR616268_2.fastq.bz2
iu@bielinux[srp017156] [ 3:01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。①「gedit ファイル名」で任意のファイルを開くことができる。

The image shows a Linux desktop environment. In the foreground, a terminal window displays the following commands and output:

```
iu@bielinux[~/Documents/srp017156]
iu@bielinux[srp017156] pwd
/home/iu/Documents/srp017156
iu@bielinux[srp017156] ls -lh
total 157M
-rw-rw-r-- 1 iu iu 13 12月 8 14:49 hoge_JSLAB3.sh
-rw-rw-r-- 1 iu iu 83M 12月 3 14:56 SRR616268_1.fastq.bz2
-rw-rw-r-- 1 iu iu 74M 12月 3 15:51 SRR616268_2.fastq.bz2
iu@bielinux[srp017156] gedit hoge_JSLAB3.sh [ 3:08午後 ]
```

A red arrow with the number '1' points to the 'gedit hoge_JSLAB3.sh' command. Below the terminal, a gedit editor window is open, showing the contents of the file 'hoge_JSLAB3.sh':

```
pwd
date
ls
```

The gedit window title is 'hoge_JSLAB3.sh (~/Documents/srp017156) - gedit'. The status bar at the bottom of the gedit window shows 'sh', 'Tab Width: 8', 'Ln 1, Col 1', and 'INS'.

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。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。

```
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 157M
-rw-rw-r-- 1 iu iu 13 12月 8 14:49 hoge_JSLAB3.sh
-rw-rw-r-- 1 iu iu 83M 12月 3 14:56 SRR616268_1.fastq.bz2
-rw-rw-r-- 1 iu iu 74M 12月 3 15:51 SRR616268_2.fastq.bz2
iu@bielinux[srp017156] gedit hoge_JSLAB3.sh [ 3:08午後 ]
```

```
hoge_JSLAB3.sh (~/.Documents/srp017156) - gedit
Open Save Undo
hoge_JSLAB3.sh x
#!/bin/sh
pwd
date
ls
sh Tab Width: 8 Ln 1, Col 10 INS
```


W3-1 : Permission

geditで編集後に、①moreで確認しているだけこれで基本的には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。

Bio-Linux Desktop, powered by Ubuntu

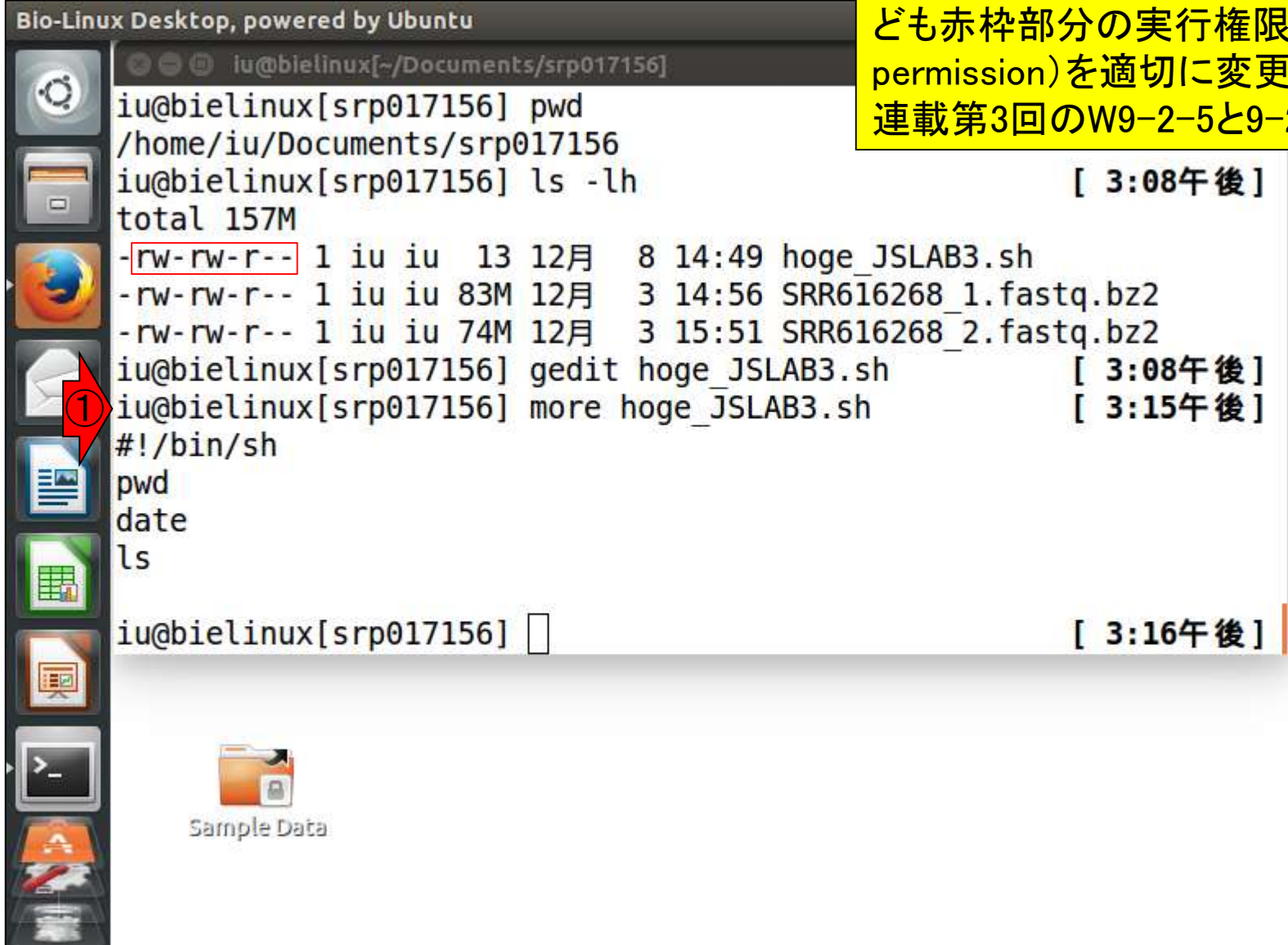
```
iu@bielinux[~/Documents/srp017156]
iu@bielinux[srp017156] pwd
/home/iu/Documents/srp017156
iu@bielinux[srp017156] ls -lh
total 157M
-rw-rw-r-- 1 iu iu 13 12月 8 14:49 hoge_JSLAB3.sh
-rw-rw-r-- 1 iu iu 83M 12月 3 14:56 SRR616268_1.fastq.bz2
-rw-rw-r-- 1 iu iu 74M 12月 3 15:51 SRR616268_2.fastq.bz2
iu@bielinux[srp017156] gedit hoge_JSLAB3.sh
iu@bielinux[srp017156] more hoge_JSLAB3.sh
#!/bin/sh
pwd
date
ls
iu@bielinux[srp017156]
```

[3:08午後]

[3:08午後]

[3:15午後]

[3:16午後]



W3-1 : Permission

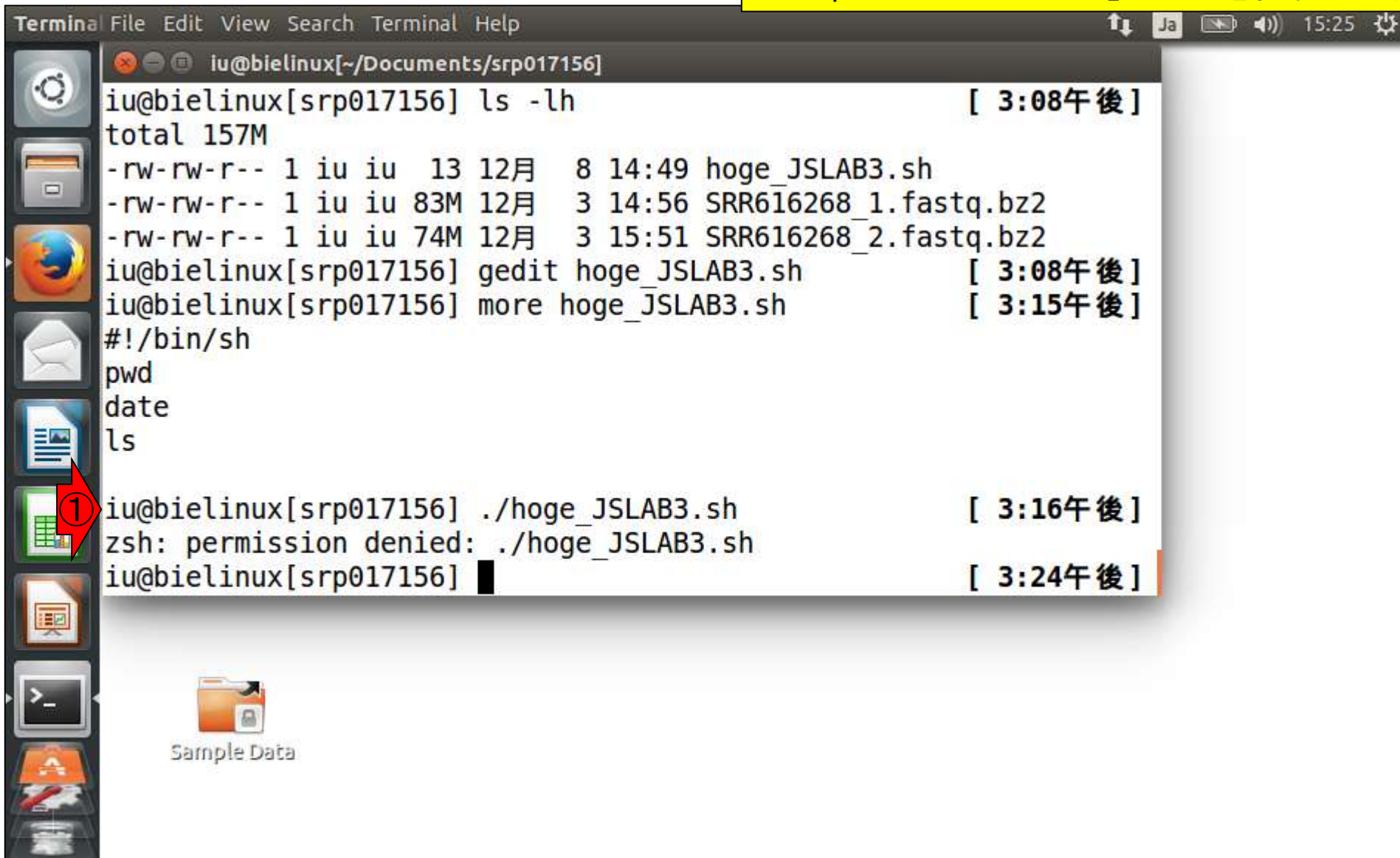
実行権限が与えられている状態でうまくいくコマンドは「./hoge_JSLAB3.sh」。まずは、このファイルに対して実行権限が与えられていない状態でエラーメッセージを出してみる。①もし実行権限が与えられていたら、「./hoge」まで打ち込んでからTabキーを押してタブ補完できるはずだができない。慣れればこの段階でPermissionを変更する必要があることに気づく。ここではうまくいかないことが分かっている状態だが、めげずに手入力で「./hoge_JSLAB3.sh」と打ち込んでリターンキーを押す。

```
Terminal File Edit View Search Terminal Help
iu@bielinux[~/Documents/srp017156]
iu@bielinux[srp017156] pwd
/home/iu/Documents/srp017156
iu@bielinux[srp017156] ls -lh
total 157M
-rw-rw-r-- 1 iu iu  13 12月  8 14:49 ho
-rw-rw-r-- 1 iu iu 83M 12月  3 14:56 SR
-rw-rw-r-- 1 iu iu 74M 12月  3 15:51 SR
iu@bielinux[srp017156] gedit hoge_JSLAB3.sh [ 3:08午後 ]
iu@bielinux[srp017156] more hoge_JSLAB3.sh [ 3:15午後 ]
#!/bin/sh
pwd
date
ls
iu@bielinux[srp017156] ./hoge [ 3:16午後 ]
```

W3-1 : Permission

①手入力で「./hoge_JSLAB3.sh」と打ち込んでリターンキーを押した結果。「実行権限が与えられていない(permission denied)」と文句を言われる。

```
Terminal File Edit View Search Terminal Help
iu@bielinux[~/Documents/srp017156]
iu@bielinux[srp017156] ls -lh [ 3:08午後 ]
total 157M
-rw-rw-r-- 1 iu iu 13 12月 8 14:49 hoge_JSLAB3.sh
-rw-rw-r-- 1 iu iu 83M 12月 3 14:56 SRR616268_1.fastq.bz2
-rw-rw-r-- 1 iu iu 74M 12月 3 15:51 SRR616268_2.fastq.bz2
iu@bielinux[srp017156] gedit hoge_JSLAB3.sh [ 3:08午後 ]
iu@bielinux[srp017156] more hoge_JSLAB3.sh [ 3:15午後 ]
#!/bin/sh
pwd
date
ls
iu@bielinux[srp017156] ./hoge_JSLAB3.sh [ 3:16午後 ]
zsh: permission denied: ./hoge_JSLAB3.sh
iu@bielinux[srp017156] [ 3:24午後 ]
```

A terminal window titled "Terminal File Edit View Search Terminal Help" showing a user "iu" at "bielinux" in the directory "~/Documents/srp017156". The user runs "ls -lh" and lists files including "hoge_JSLAB3.sh". They then run "gedit hoge_JSLAB3.sh" and "more hoge_JSLAB3.sh". After a shell prompt "#!/bin/sh", they run "pwd", "date", and "ls". Finally, they run "./hoge_JSLAB3.sh" which results in the error "zsh: permission denied: ./hoge_JSLAB3.sh". A red arrow with the number "1" points to this error line. The terminal also shows a "Sample Data" folder icon at the bottom.

W3-1 : Permission

「./」が無駄なのだろうと思い、②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。

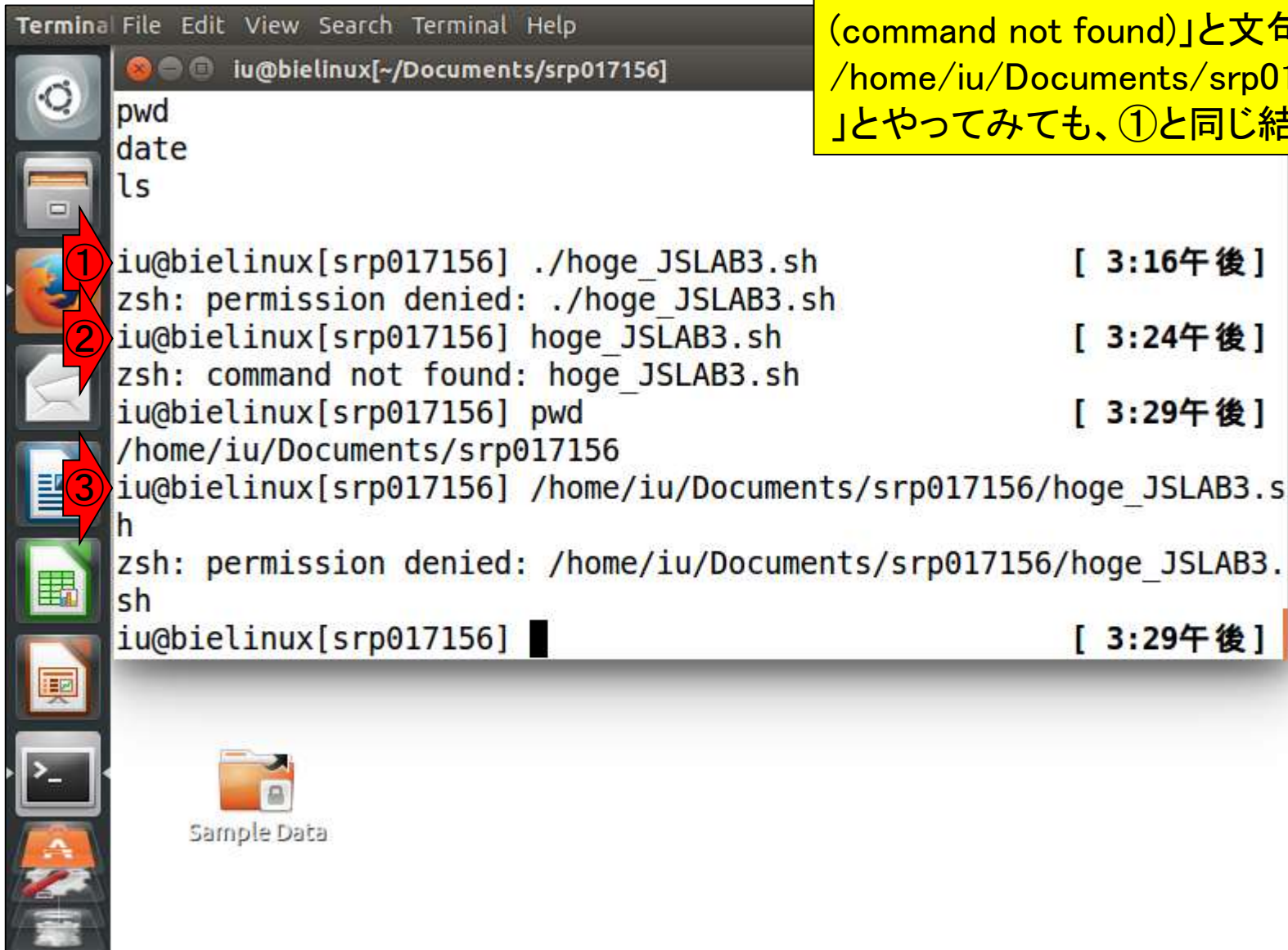
```
Terminal File Edit View Search Terminal Help
iu@bielinux[~/Documents/srp017156]
pwd
date
ls

① iu@bielinux[srp017156] ./hoge_JSLAB3.sh [ 3:16午後 ]
zsh: permission denied: ./hoge_JSLAB3.sh

② iu@bielinux[srp017156] hoge_JSLAB3.sh [ 3:24午後 ]
zsh: command not found: hoge_JSLAB3.sh

iu@bielinux[srp017156] pwd [ 3:29午後 ]
/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] [ 3:29午後 ]
```



W3-1 : Permission

パーミッションの変更は①chmod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「`rw-rw-r--`」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。②テキストエディタ実行後には、よくこのようなファイルの最後に「~(による)」がついたバックアップファイルが勝手に作成されますが、悪さはしないので気にしなくてよい。

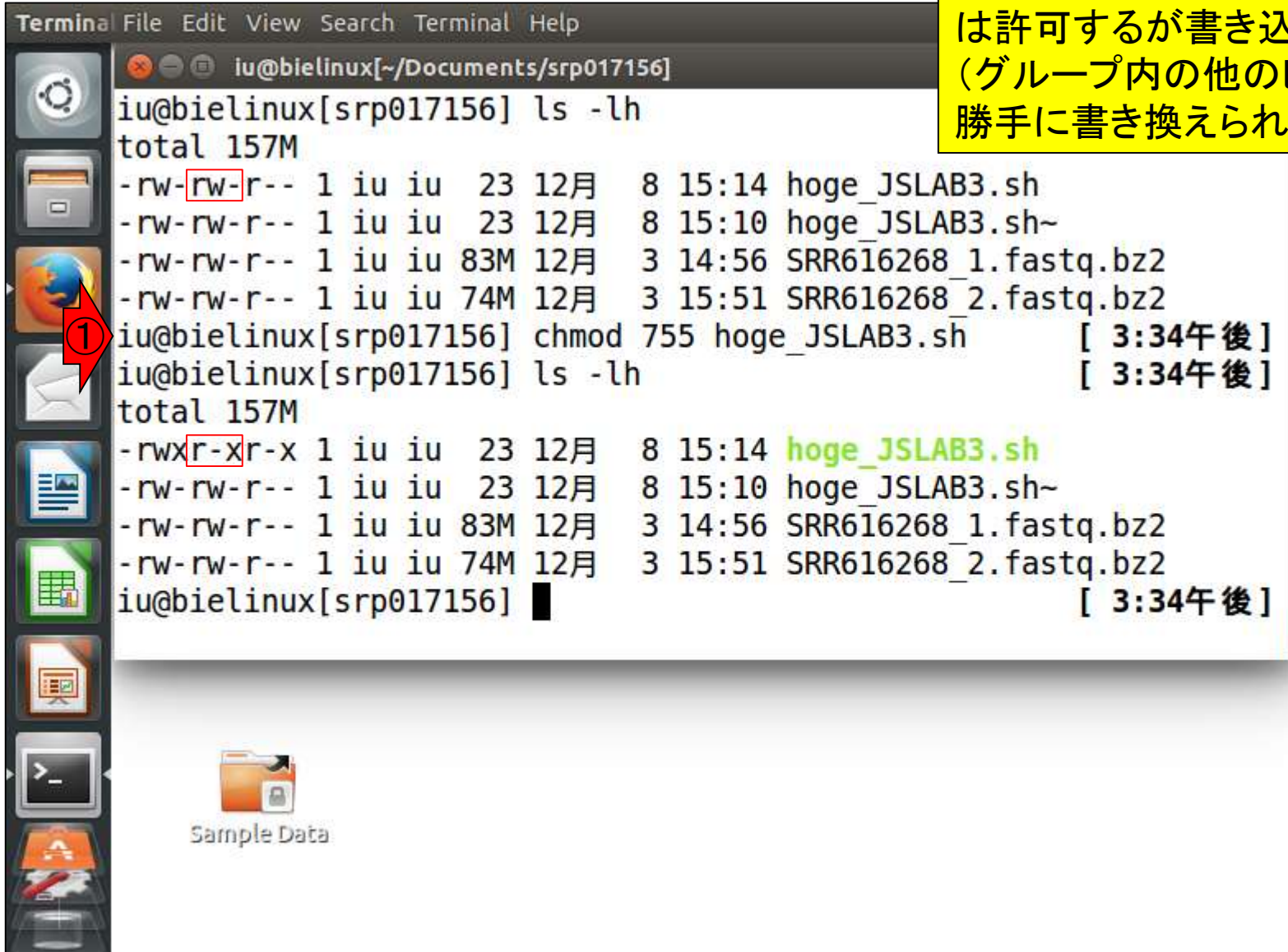
```
Terminal File Edit View Search Terminal Help
iu@bielinux[~/Documents/srp017156]
iu@bielinux[srp017156] ls -lh
total 157M
-rw-rw-r-- 1 iu iu 23 12月 8 15:14
-rw-rw-r-- 1 iu iu 23 12月 8 15:10
-rw-rw-r-- 1 iu iu 83M 12月 3 14:56
-rw-rw-r-- 1 iu iu 74M 12月 3 15:51 SRR616268_2.fastq.bz2
iu@bielinux[srp017156] chmod 755 hoge_JSLAB3.sh [ 3:34午後 ]
iu@bielinux[srp017156] ls -lh [ 3:34午後 ]
total 157M
-rwxr-xr-x 1 iu iu 23 12月 8 15:14 hoge_JSLAB3.sh
-rw-rw-r-- 1 iu iu 23 12月 8 15:10 hoge_JSLAB3.sh~
-rw-rw-r-- 1 iu iu 83M 12月 3 14:56 SRR616268_1.fastq.bz2
-rw-rw-r-- 1 iu iu 74M 12月 3 15:51 SRR616268_2.fastq.bz2
iu@bielinux[srp017156] [ 3:34午後 ]
```



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)。

```
Terminal File Edit View Search Terminal Help
iu@bielinux[~/Documents/srp017156]
iu@bielinux[srp017156] ls -lh
total 157M
-rw-rw-r-- 1 iu iu 23 12月 8 15:14 hoge_JSLAB3.sh
-rw-rw-r-- 1 iu iu 23 12月 8 15:10 hoge_JSLAB3.sh~
-rw-rw-r-- 1 iu iu 83M 12月 3 14:56 SRR616268_1.fastq.bz2
-rw-rw-r-- 1 iu iu 74M 12月 3 15:51 SRR616268_2.fastq.bz2
iu@bielinux[srp017156] chmod 755 hoge_JSLAB3.sh [ 3:34午後 ]
iu@bielinux[srp017156] ls -lh [ 3:34午後 ]
total 157M
-rwxr-xr-x 1 iu iu 23 12月 8 15:14 hoge_JSLAB3.sh
-rw-rw-r-- 1 iu iu 23 12月 8 15:10 hoge_JSLAB3.sh~
-rw-rw-r-- 1 iu iu 83M 12月 3 14:56 SRR616268_1.fastq.bz2
-rw-rw-r-- 1 iu iu 74M 12月 3 15:51 SRR616268_2.fastq.bz2
iu@bielinux[srp017156] [ 3:34午後 ]
```



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。

```
Terminal File Edit View Search Terminal Help
iu@bielinux[~/Documents/srp017156]
iu@bielinux[srp017156] ls -lh
total 157M
-rw-rw-r-- 1 iu iu 23 12月 8 15:14 hoge_JSLAB3.sh
-rw-rw-r-- 1 iu iu 23 12月 8 15:10 hoge_JSLAB3.sh~
-rw-rw-r-- 1 iu iu 83M 12月 3 14:56 SRR616268_1.fastq.bz2
-rw-rw-r-- 1 iu iu 74M 12月 3 15:51 SRR616268_2.fastq.bz2
iu@bielinux[srp017156] chmod 755 hoge_JSLAB3.sh [ 3:34午後 ]
iu@bielinux[srp017156] ls -lh [ 3:34午後 ]
total 157M
-rwxr-xr-x 1 iu iu 23 12月 8 15:14 hoge_JSLAB3.sh
-rw-rw-r-- 1 iu iu 23 12月 8 15:10 hoge_JSLAB3.sh~
-rw-rw-r-- 1 iu iu 83M 12月 3 14:56 SRR616268_1.fastq.bz2
-rw-rw-r-- 1 iu iu 74M 12月 3 15:51 SRR616268_2.fastq.bz2
iu@bielinux[srp017156] [ 3:34午後 ]
```



W3-1 : Permission

例えば、逆パターンとして、他のユーザから「①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。

```
Terminal File Edit View Search Terminal Help
iu@bielinux[~/Documents/srp017156]
iu@bielinux[srp017156] pwd
/home/iu/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] clear [ 3:29午後 ]
iu@bielinux[srp017156] ls -lh [ 3:34午後 ]
total 157M
-rw-rw-r-- 1 iu iu 23 12月 8 15:14 hoge_JSLAB3.sh
-rw-rw-r-- 1 iu iu 23 12月 8 15:10 hoge_JSLAB3.sh~
-rw-rw-r-- 1 iu iu 83M 12月 3 14:56 SRR616268_1.fastq.bz2
-rw-rw-r-- 1 iu iu 74M 12月 3 15:51 SRR616268_2.fastq.bz2
iu@bielinux[srp017156] chmod 755 hoge_JSLAB3.sh [ 3:34午後 ]
iu@bielinux[srp017156] ls -lh [ 3:34午後 ]
total 157M
-rwxr-xr-x 1 iu iu 23 12月 8 15:14 hoge_JSLAB3.sh
-rw-rw-r-- 1 iu iu 23 12月 8 15:10 hoge_JSLAB3.sh~
Sample Data
```



①

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)。

```
Terminal File Edit View Search Terminal Help
iu@bielinux[~/Documents/srp017156]
① iu@bielinux[srp017156] pwd
/home/iu/Documents/srp017156
② iu@bielinux[srp017156] ls -lh
total 157M
-rwxr-xr-x 1 iu iu 23 12月 8 15:14 hoge_JSLAB3.sh
-rw-rw-r-- 1 iu iu 23 12月 8 15:10 hoge_JSLAB3.sh~
-rw-rw-r-- 1 iu iu 83M 12月 3 14:56 SRR616268_1.fastq.bz2
-rw-rw-r-- 1 iu iu 74M 12月 3 15:51 SRR616268_2.fastq.bz2
③ iu@bielinux[srp017156] more hoge_JSLAB3.sh
#!/bin/sh
pwd
date
ls
④ iu@bielinux[srp017156] ./ho
```

[3:51午後]

[3:51午後]

[3:51午後]

Sample Data

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。

```
Terminal File Edit View Search Terminal Help
iu@bielinux[~/Documents/srp017156]
/home/iu/Documents/srp017156
iu@bielinux[srp017156] ls -lh [ 3:51午後 ]
total 157M
-rwxr-xr-x 1 iu iu 23 12月 8 15:14 hoge_JSLAB3.sh
-rw-rw-r-- 1 iu iu 23 12月 8 15:10 hoge_JSLAB3.sh~
-rw-rw-r-- 1 iu iu 83M 12月 3 14:56 SRR616268_1.fastq.bz2
-rw-rw-r-- 1 iu iu 74M 12月 3 15:51 SRR616268_2.fastq.bz2
iu@bielinux[srp017156] more hoge_JSLAB3.sh [ 3:51午後 ]
#!/bin/sh
pwd
date
ls
iu@bielinux[srp017156] ./hoge_JSLAB3.sh [ 3:51午後 ]
/home/iu/Documents/srp017156
2015年 12月 8日 火曜日 15:54:28 JST
hoge_JSLAB3.sh SRR616268_1.fastq.bz2
hoge_JSLAB3.sh~ SRR616268_2.fastq.bz2
iu@bielinux[srp017156] [ 3:54午後 ]
Sample Data
```

④

iu@bielinux[srp017156] ./hoge_JSLAB3.sh
/home/iu/Documents/srp017156
2015年 12月 8日 火曜日 15:54:28 JST
hoge_JSLAB3.sh SRR616268_1.fastq.bz2
hoge_JSLAB3.sh~ SRR616268_2.fastq.bz2

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。

```
Terminal File Edit View Search Terminal Help
iu@bielinux[~/Documents/srp017156]
① iu@bielinux[srp017156] ./hoge_JSLAB3.sh
/home/iu/Documents/srp017156
2015年 12月 8日 火曜日 16:03:36 JST
hoge_JSLAB3.sh SRR616268_1.fastq.bz2
hoge_JSLAB3.sh~ SRR616268_2.fastq.bz2
② iu@bielinux[srp017156] hoge_JSLAB3.sh
zsh: command not found: hoge_JSLAB3.sh
③ iu@bielinux[srp017156] sh hoge_JSLAB3.sh
/home/iu/Documents/srp017156
2015年 12月 8日 火曜日 16:03:43 JST
hoge_JSLAB3.sh SRR616268_1.fastq.bz2
hoge_JSLAB3.sh~ SRR616268_2.fastq.bz2
iu@bielinux[srp017156]
```

[4:03午後]

[4:03午後]

[4:03午後]

Sample Data

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。「sh hoge_JSLAB3.sh」と「sh < hoge_JSLAB3.sh」は実質的に同じという理解でよい。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

```
Terminal File Edit View Search Terminal Help
iu@bielinux[~/Documents/srp017156]
① iu@bielinux[srp017156] sh < hoge_JSLAB3.sh
/home/iu/Documents/srp017156
2015年 12月 8日 火曜日 16:08:33 JST
hoge_JSLAB3.sh SRR616268_1.fastq.bz2
hoge_JSLAB3.sh~ SRR616268_2.fastq.bz2
② iu@bielinux[srp017156] sh < hoge_JSLAB3.sh > hoge1.txt
iu@bielinux[srp017156] more hoge1.txt
/home/iu/Documents/srp017156
2015年 12月 8日 火曜日 16:08:33 JST
hoge1.txt
hoge_JSLAB3.sh
hoge_JSLAB3.sh~
SRR616268_1.fastq.bz2
SRR616268_2.fastq.bz2
iu@bielinux[srp017156]
```

[4:08午後]

[4:08午後]

Sample Data

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、①目的行の先頭に#をつけてコメントアウトし、②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。

```
Terminal File Edit View Search Terminal Help
iu@bielinux[~/Documents/srp017156]
iu@bielinux[srp017156] pwd
/home/iu/Documents/srp017156
iu@bielinux[srp017156] ls
hoge1.txt             hoge_JSLAB3.sh~      SRR616268_2.fastq.bz2
hoge_JSLAB3.sh       SRR616268_1.fastq.bz2
iu@bielinux[srp017156] more hoge_JSLAB3.sh
#!/bin/sh
pwd
#date
ls
iu@bielinux[srp017156] ./hoge_JSLAB3.sh
/home/iu/Documents/srp017156
hoge1.txt             hoge_JSLAB3.sh~      SRR616268_2.fastq.bz2
hoge_JSLAB3.sh       SRR616268_1.fastq.bz2
iu@bielinux[srp017156] █
```

Sample Data

①

②

[4:20午後]

[4:20午後]

[4:20午後]

[4:20午後]

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Windows用のエディタで眺めているところ。①コメントアウトされている最初の2行分は、約1.35億の全リードファイルをダウンロードするときに用いるコマンド。ここに書かれているようなものが実務者がよく使っているシェルスクリプトファイル(の一部)。これをwgetで取得する。

• [W3-3]

```
sh < hoge_JSLAB3.sh
sh < hoge_JSLAB3.sh > hoge1.txt
more hoge1.txt

pwd
ls
more hoge_JSLAB3.sh
./hoge_JSLAB3.sh
```

• [JSLAB4_1.sh](#) (Windowsの改行コード: CR+LF)[W4-1]

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

C:\Users\kadota\Desktop\share\JSLAB4_1.sh - EmEditor

ファイル(F) 編集(E) 検索(S) 表示(V) ツール(T) ウィンドウ(W) ヘルプ(H)

JSLAB4_1.sh x

```
#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
```

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



```
Terminal File Edit View Search Terminal Help
iu@bielinux[~/Documents/srp017156]
iu@bielinux[srp017156] pwd [ 4:34午後 ]
/home/iu/Documents/srp017156
iu@bielinux[srp017156] ls [ 4:34午後 ]
hogel.txt hoge_JSLAB3.sh~ SRR616268_2.fastq.bz2
hoge_JSLAB3.sh SRR616268_1.fastq.bz2
iu@bielinux[srp017156] wget -c http://www.iu.a.u-tokyo.ac.jp/~kadota/book/JSLAB4_1.sh
```


W4-1 : wgetの成功例

①「wget -c 任意のURL」で基本的にはダウンロード可能。これはうまくいったときの見え方

```
Bio-Linux Desktop, powered by Ubuntu
iu@bielinux[~/Documents/srp017156]
hoge1.txt      hoge_JSLAB3.sh~      SRR616268_2.fastq.bz2
hoge_JSLAB3.sh SRR616268_1.fastq.bz2
iu@bielinux[srp017156] wget -c http://www.iu.a.u-tokyo.ac.jp/~kadota/book/JSLAB4_1.sh
--2015-12-08 16:44:05-- 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

2015-12-08 16:44:10 (56.4 MB/s) - 'JSLAB4_1.sh' saved [364/364]

iu@bielinux[srp017156] [ 4:44午後 ]
```



W4-1 : wgetの失敗例

失敗例を示すために、①一旦JSLAB4_1.shを削除。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。(私の環境ではLANケーブルを抜いた状態で)②wgetコマンドを再実行。③失敗したときは赤下線のような結果になります。また、ネットワークの不具合以外にも、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のでやってはいけません

```
Terminal File Edit View Search Terminal Help
iu@bielinux[~/Documents/srp017156]
iu@bielinux[srp017156] ls
hoge1.txt      hoge_JSLAB3.sh~  SRR616268
hoge_JSLAB3.sh JSLAB4_1.sh      SRR616268
① iu@bielinux[srp017156] rm JSLAB4_1.sh
rm: remove regular file 'JSLAB4_1.sh'? y
iu@bielinux[srp017156] ls
hoge1.txt      hoge_JSLAB3.sh~  SRR6
hoge_JSLAB3.sh SRR616268_1.fastq.bz2
② iu@bielinux[srp017156] wget -c http://www.iu.a.u-tokyo.ac.jp/~kadota/book/JSLAB4_1.sh
--2015-12-08 16:50:45-- 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] ls [ 4:51午後 ]
hoge1.txt      hoge_JSLAB3.sh~  SRR616268_2.fastq.bz2
hoge_JSLAB3.sh SRR616268_1.fastq.bz2
iu@bielinux[srp017156] [ 4:58午後 ]
```

Sample Data

W4-1: ネットワーク不調

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。

The image shows a terminal window in the background with the following text:

```
iu@bielinux[~/Documents/srp017156]
hoge1.txt
hoge_JSLAB3.sh
iu@bielinux[srp017156]
rm: remove reg
iu@bielinux[srp017156]
hoge1.txt
hoge_JSLAB3.sh
iu@bielinux[srp017156]
ota/book/JSLAB
--2015-12-08 1
ok/JSLAB4_1.sh
Resolving ww.
ed: Name or se
wget: unable t
iu@bielinux[srp017156]
hoge1.txt
hoge_JSLAB3.sh
iu@bielinux[srp017156]
```

In the foreground, a Firefox browser window is open, displaying a "Problem loading page" error. The address bar shows "Search or enter address". The main content area displays:

Server not found

Firefox can't find the server at start.ubuntu.com.

- Check the address for typing errors such as **ww**.example.com instead of **www**.example.com
- If you are unable to load any pages, check your computer's network connection.
- If your computer or network is protected by a firewall or proxy, make sure that Firefox is permitted to access the Web.

[Try Again](#)

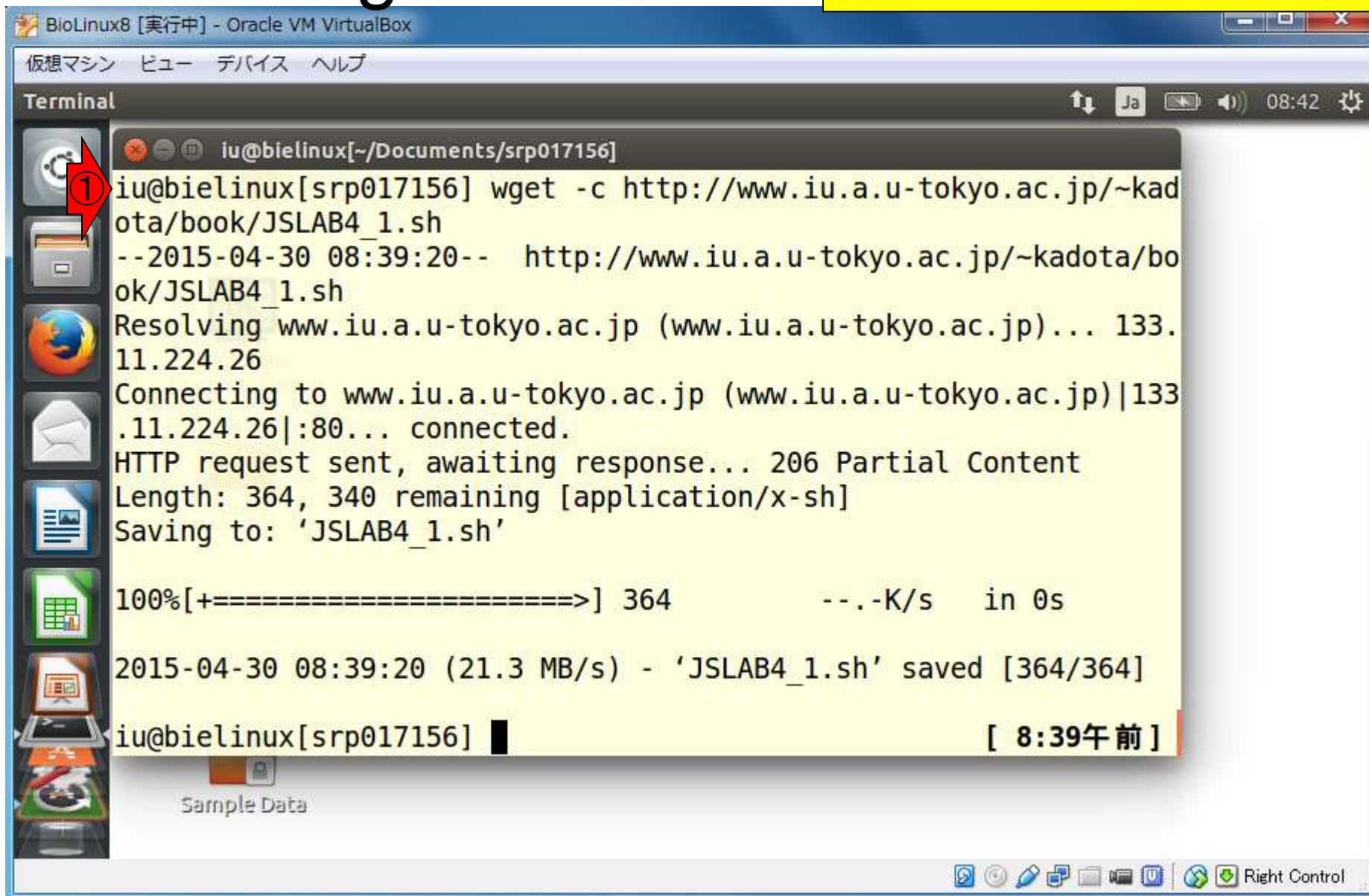
W4-2: wgetリトライ

ホストOS(Windows)は無線LANで問題なくネットを見られるが、ゲストOS(BioLinux8)でできないこともときどきある。理由は不明だが、①のようなメッセージが出ているときはダメです

The screenshot shows a terminal window in a virtual machine. The terminal prompt is `iu@bielinux[~/Documents/srp017156]`. The user has entered `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 file was downloaded successfully. However, when they entered `wget -c http://www.iu.a.u-tokyo.ac.jp/~kadota/book/JSLAB4_1.sh` again, 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'`. A notification box on the right side of the terminal window reads: **Network service discovery disabled**. The message says: "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 with the number 1 points to this 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、①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!



```
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]
```


W4-3: wgetまとめ

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
--2015-12-08 17:28:05-- 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

2015-12-08 17:28:05 (77.8 MB/s) - 'JSLAB4_1.sh' saved [364/364]

iu@bielinux[srp017156] ls                                     [ 5:28午後 ]
hogel.txt          hoge_JSLAB3.sh~  SRR616268_1.fastq.bz2
hoge_JSLAB3.sh    JSLAB4_1.sh     SRR616268_2.fastq.bz2

iu@bielinux[srp017156] more JSLAB4_1.sh                    [ 5:28午後 ]
#wget -c ftp://ftp.ddbj.nig.ac.jp/ddbj_database/dra/fastq/SRA061/SRA061483/SRX20422
6/SRR616268_1.fastq.bz2
#wget -c ftp://ftp.ddbj.nig.ac.jp/ddbj_database/dra/fastq/SRA061/SRA061483/SRX20422
6/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] █                                     [ 5:28午後 ]
```


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**

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

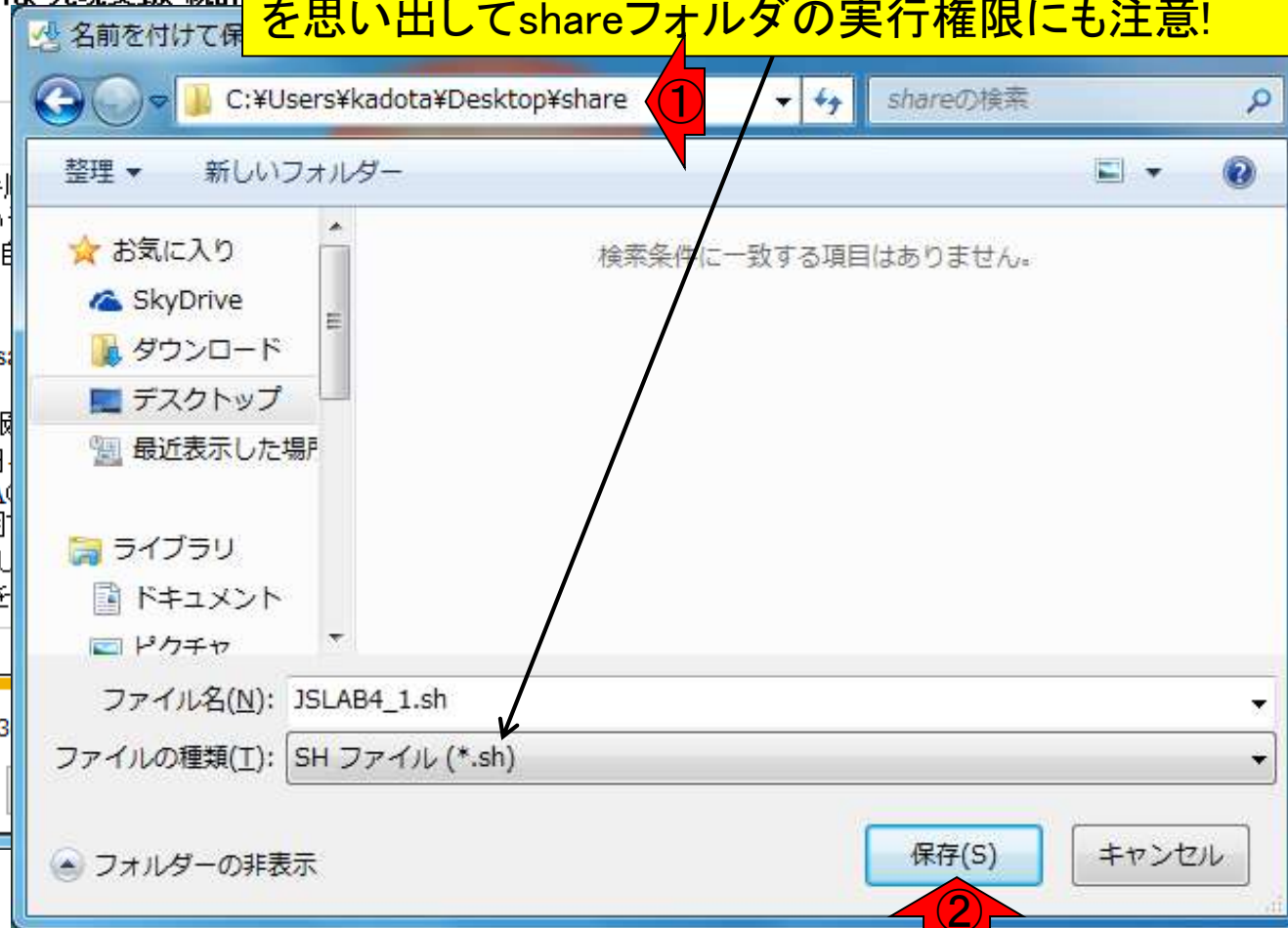
~NGS, RNA-seq, ゲノム, トランスクリプトーム, 正規化, 発現変動, 統計
(last modified 2015/04/28, since 2010)

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W9-2-2のホストOS側フォルダ指定(ホストOS - デスクトップ - share)とW9-2-5のゲストOS側フォルダ指定(ゲストOS - Desktop - mac_share)のどちらか、またはいずれが必要。共有フォルダ設定が無効になっているのは、①赤枠部分のpermissionがrwxrwxrwxになっていないこと、および所有者がrootになっていないことから確認できる。連載第3回のW9-6も参照のこと

```
File Edit View Search Terminal Help
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
iu@bielinux[iu] cd Desktop
iu@bielinux[Desktop] ls
Bio-Linux Documentation hoge mac_share Sample Data
iu@bielinux[Desktop] ls -l
total 12
drwxr-xr-x 2 iu iu 4096 11月 17 13:44 Bio-Linux Documentation
drwxrwxr-x 2 iu iu 4096 11月 30 17:08 hoge
drwxrwxr-x 2 iu iu 4096 11月 26 18:10 mac_share
lrwxrwxrwx 1 iu iu 29 11月 17 13:44 Sample Data -> /usr/local/bioinf/sample
data/
iu@bielinux[Desktop]
```

[5:36午後]

[5:36午後]

[5:36午後]



①

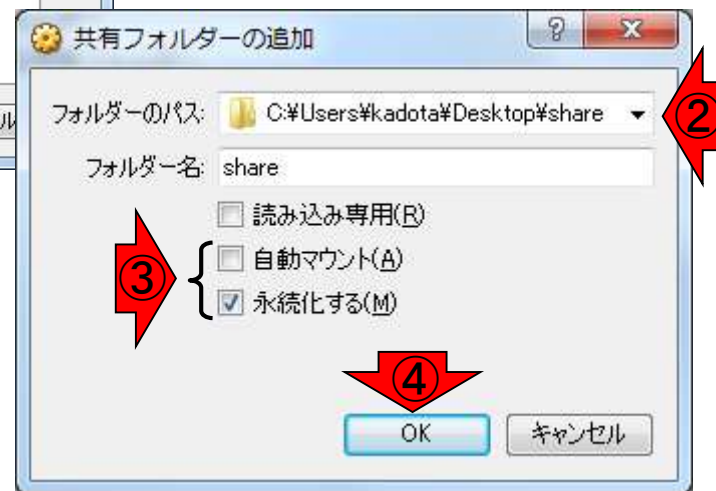
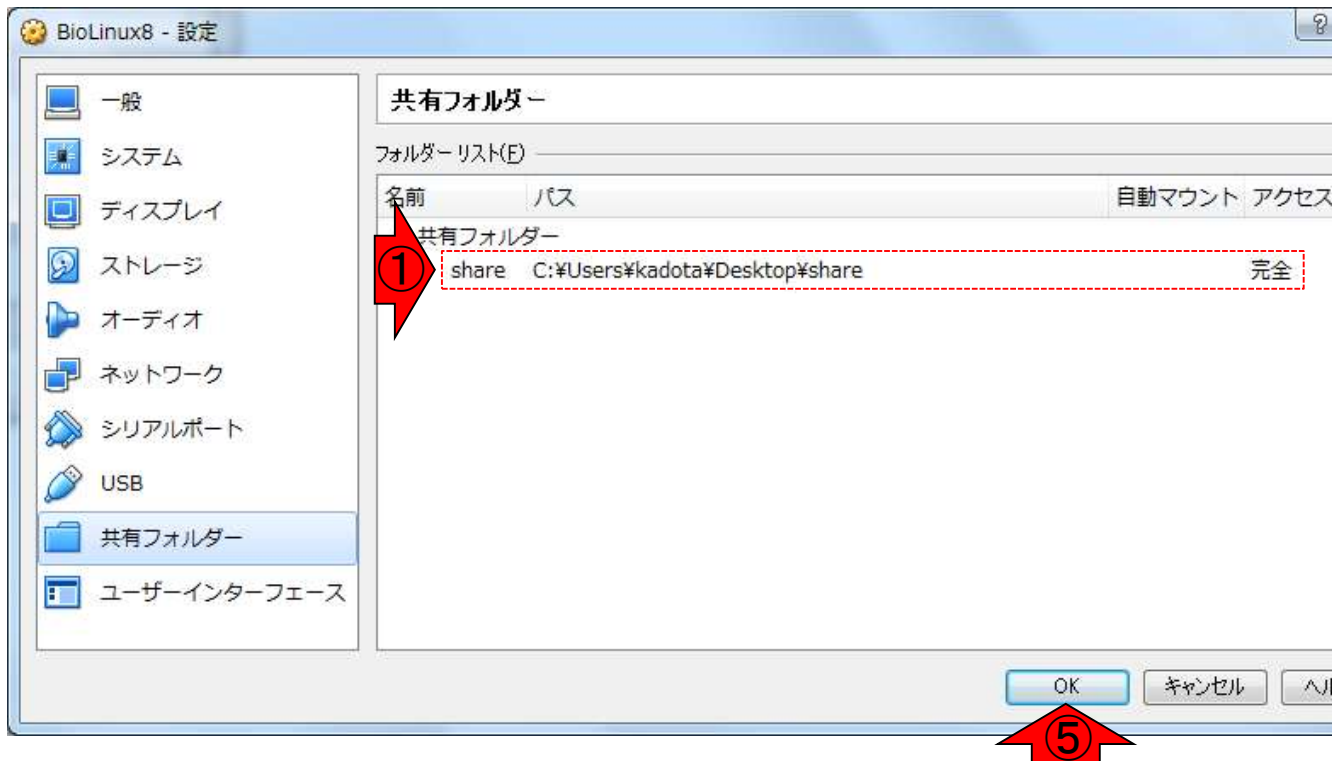
W4-4: 共有フォルダ

ホストOS側の共有フォルダ(ホストOS - デスクトップ - share)指定状況を確認。連載第3回W9-2-2も参照。①デバイス、②共有フォルダー、③共有フォルダー設定。警告メッセージが出るが、とりあえず④OK

共有フォルダーの設定に仮想マシン設定ウィンドウを表示
日本乳酸菌学会誌の連載第4回

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。①赤枠のように見えていなければそうなようにする。パスのところがC¥Users¥kadota¥Desktop¥shareとなっているが、特にユーザ名部分はヒトそれぞれ。③のようになっていることも確認。



W4-5-1: 確認

```

File Edit View Search Terminal Help
iu@bielinux[srp017156] pwd [ 5:36午後]
/home/iu/Documents/srp017156
iu@bielinux[srp017156] cd [ 5:36午後]
iu@bielinux[iu] pwd [ 5:36午後]
/home/iu
iu@bielinux[iu] ls [ 5:36午後]
Desktop Documents Downloads Music Pictures Public Templates Videos
iu@bielinux[iu] cd Desktop [ 5:36午後]
iu@bielinux[Desktop] ls [ 5:36午後]
Bio-Linux Documentation hoge mac_share Sample Data
iu@bielinux[Desktop] ls -l [ 5:36午後]
total 12
drwxr-xr-x 2 iu iu 4096 11月 17 13:44 Bio-Linux Documentation
drwxrwxr-x 2 iu iu 4096 11月 30 17:08 hoge
drwxrwxr-x 2 iu iu 4096 11月 26 18:10 mac_share
lrwxrwxrwx 1 iu iu 29 11月 17 13:44 Sample Data -> /usr/local/bioinf/sample
data/
iu@bielinux[Desktop] [ 5:36午後]

```



1

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)

```
File Edit View Search Terminal Help 18:03
iu@bielinux[srp017156] pwd [ 5:36午後 ]
/home/iu/Documents/srp017156
iu@bielinux[srp017156] cd [ 5:36午後 ]
iu@bielinux[iu] pwd [ 5:36午後 ]
/home/iu
iu@bielinux[iu] ls [ 5:36午後 ]
Desktop Documents Downloads Music Pictures Public Templates Videos
iu@bielinux[iu] cd Desktop [ 5:36午後 ]
iu@bielinux[Desktop] ls [ 5:36午後 ]
Bio-Linux Documentation hoge mac_share Sample Data
iu@bielinux[Desktop] ls -l [ 5:36午後 ]
total 12
drwxr-xr-x 2 iu iu 4096 11月 17 13:44 Bio-Linux Documentation
drwxrwxr-x 2 iu iu 4096 11月 30 17:08 hoge
drwxrwxr-x 2 iu iu 4096 11月 26 18:10 mac_share
lrwxrwxrwx 1 iu iu 29 11月 17 13:44 Sample Data -> /usr/local/bioinf/sample
data/
① iu@bielinux[Desktop] sudo apt-get install dkms [ 5:36午後 ]
[sudo] password for iu:
```

W4-5-2: dkms

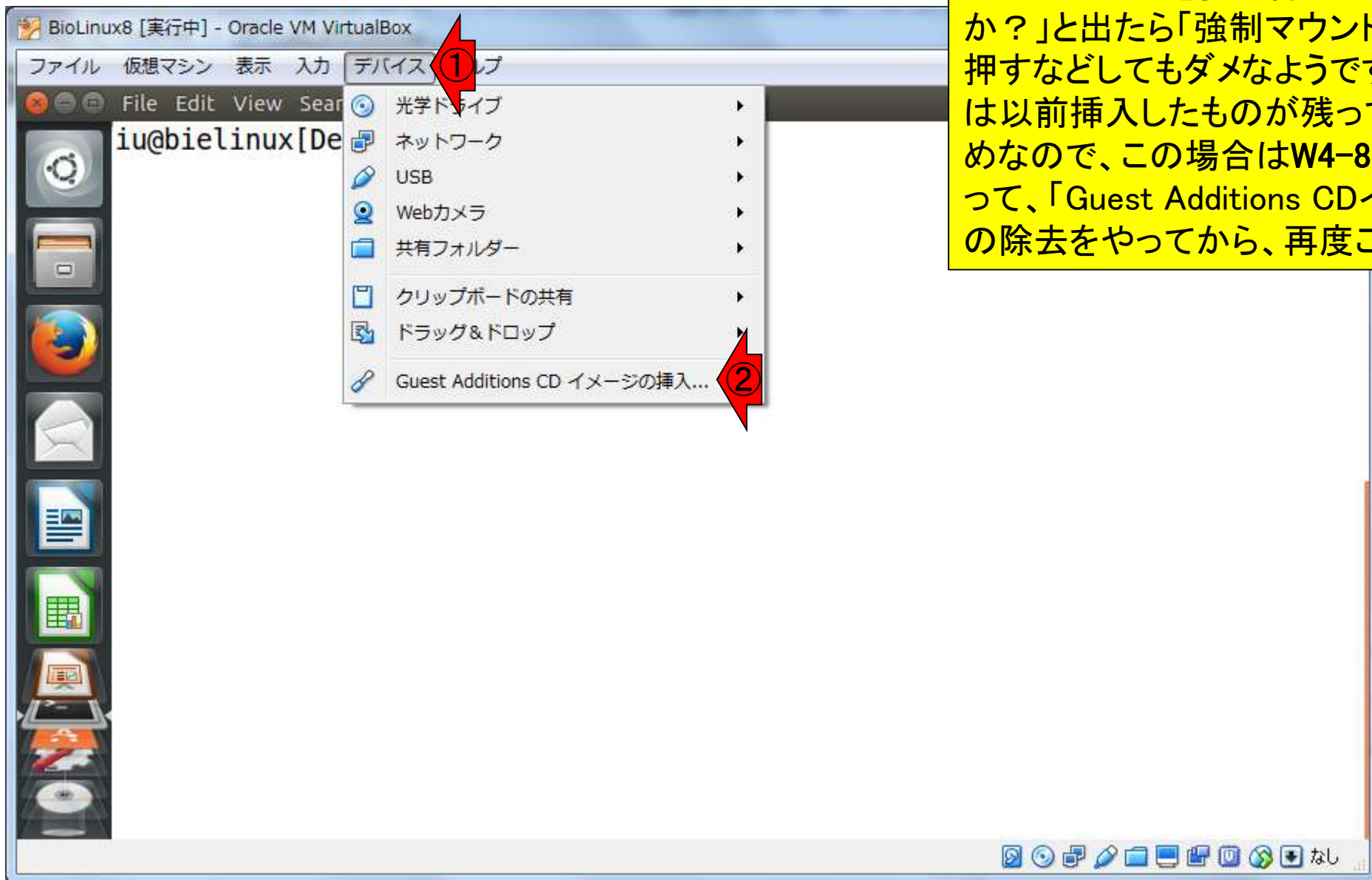
無事インストールが完了すると、②のようにコマンド入力待ち状態となる。2015年9月18日に別のマシンでやろうとしたら、「Unable to fetch some archives, maybe run apt-get update or try with --fix-missing?」というメッセージが出ました。このときは「sudo apt-get update」をやったのち、再度①をやりと成功しました。2015年12月08日に同一マシンで一から動作確認をやり直したときは、9/18のようなメッセージには遭遇しませんでした。

```
iu@bielinux[Desktop] sudo apt-get insta
[sudo] password for iu:
Reading package lists... Done
Building dependency tree
Reading state information... Done
The following packages were automatical
d:
  linux-headers-3.13.0-55 linux-headers-3.13.0-55-generic
  linux-image-3.13.0-55-generic linux-image-extra-3.13.0-55-generic
Use 'apt-get autoremove' to remove them.
The following NEW packages will be installed:
  dkms
0 upgraded, 1 newly installed, 0 to remove and 110 not upgraded.
Need to get 65.4 kB of archives.
After this operation, 351 kB of additional disk space will be used.
Get:1 http://jp.archive.ubuntu.com/ubuntu/ trusty-updates/main dkms all 2.2.0
.3-1.1ubuntu5.14.04.5 [65.4 kB]
Fetched 65.4 kB in 0s (321 kB/s)
Selecting previously unselected package dkms.
(Reading database ... 379556 files and directories currently installed.)
Preparing to unpack .../dkms_2.2.0.3-1.1ubuntu5.14.04.5_all.deb ...
Unpacking dkms (2.2.0.3-1.1ubuntu5.14.04.5) ...
Processing triggers for man-db (2.6.7.1-1ubuntu1) ...
Setting up dkms (2.2.0.3-1.1ubuntu5.14.04.5) ...
iu@bielinux[Desktop]
```

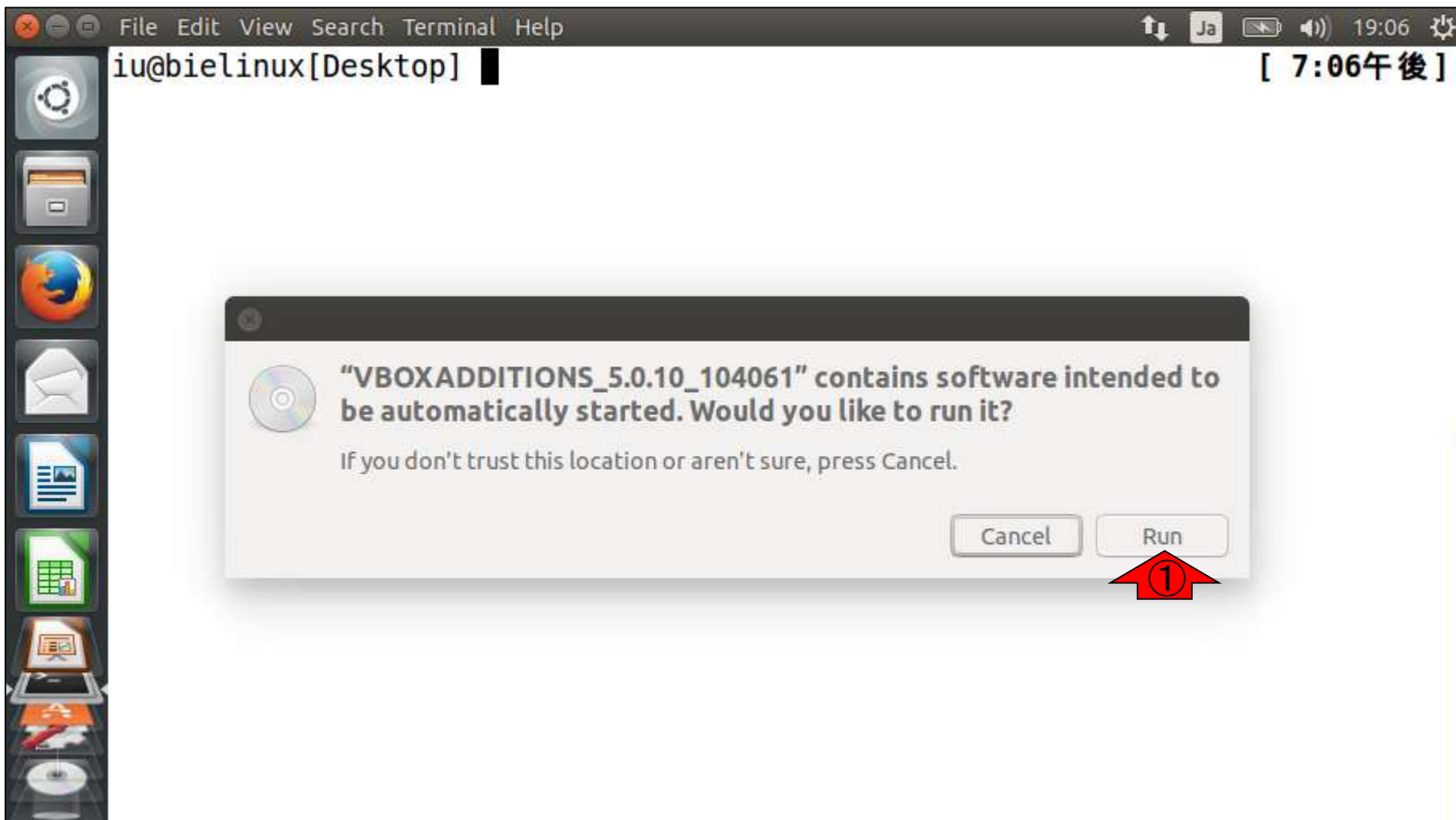
[6:07午後]

W4-5-3: Guest Additions.

①デバイス、②Guest 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

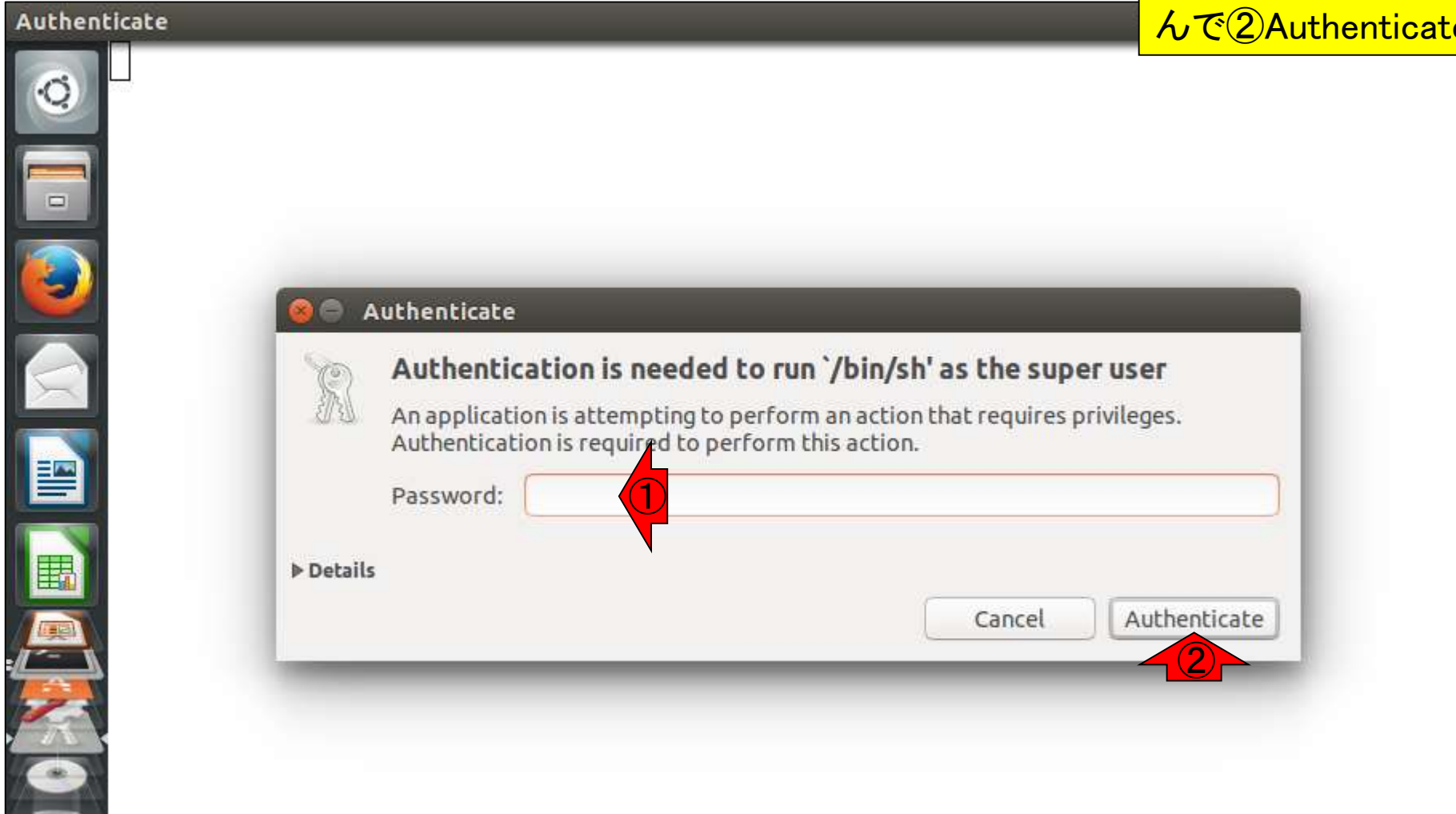


W4-5-3: Guest Additions...



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。

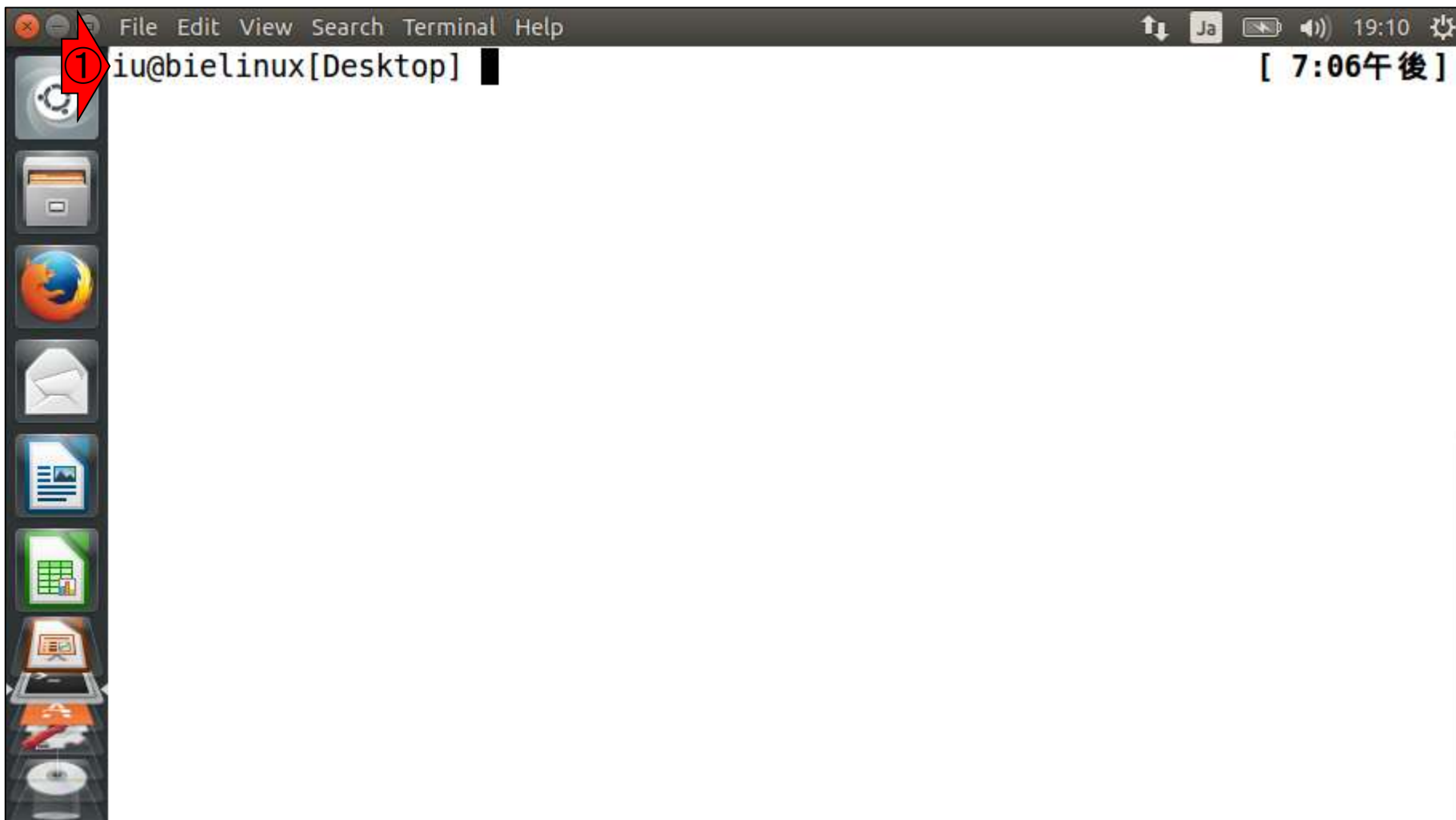
```
File Edit View Search Terminal Help 19:09
Removing installed version 5.0.10 of VirtualBox Guest Additions...
Stopping VirtualBox Additions ...done.
Removing existing VirtualBox DKMS kernel modules ...done.
Removing existing VirtualBox non-DKMS kernel modules ...done.
Copying additional installer 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...
█
```



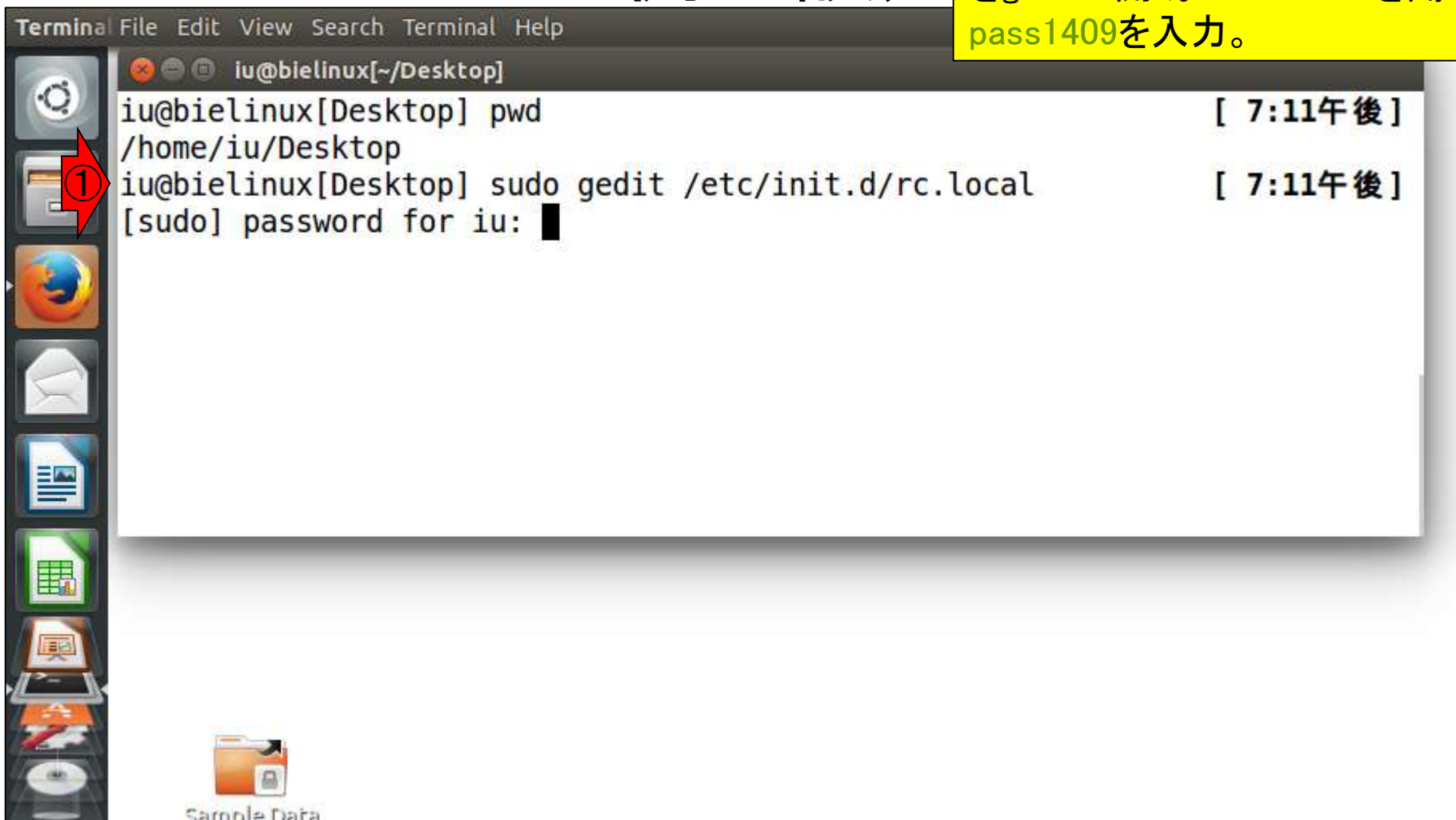
W4-5-3: Guest Additions...

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。



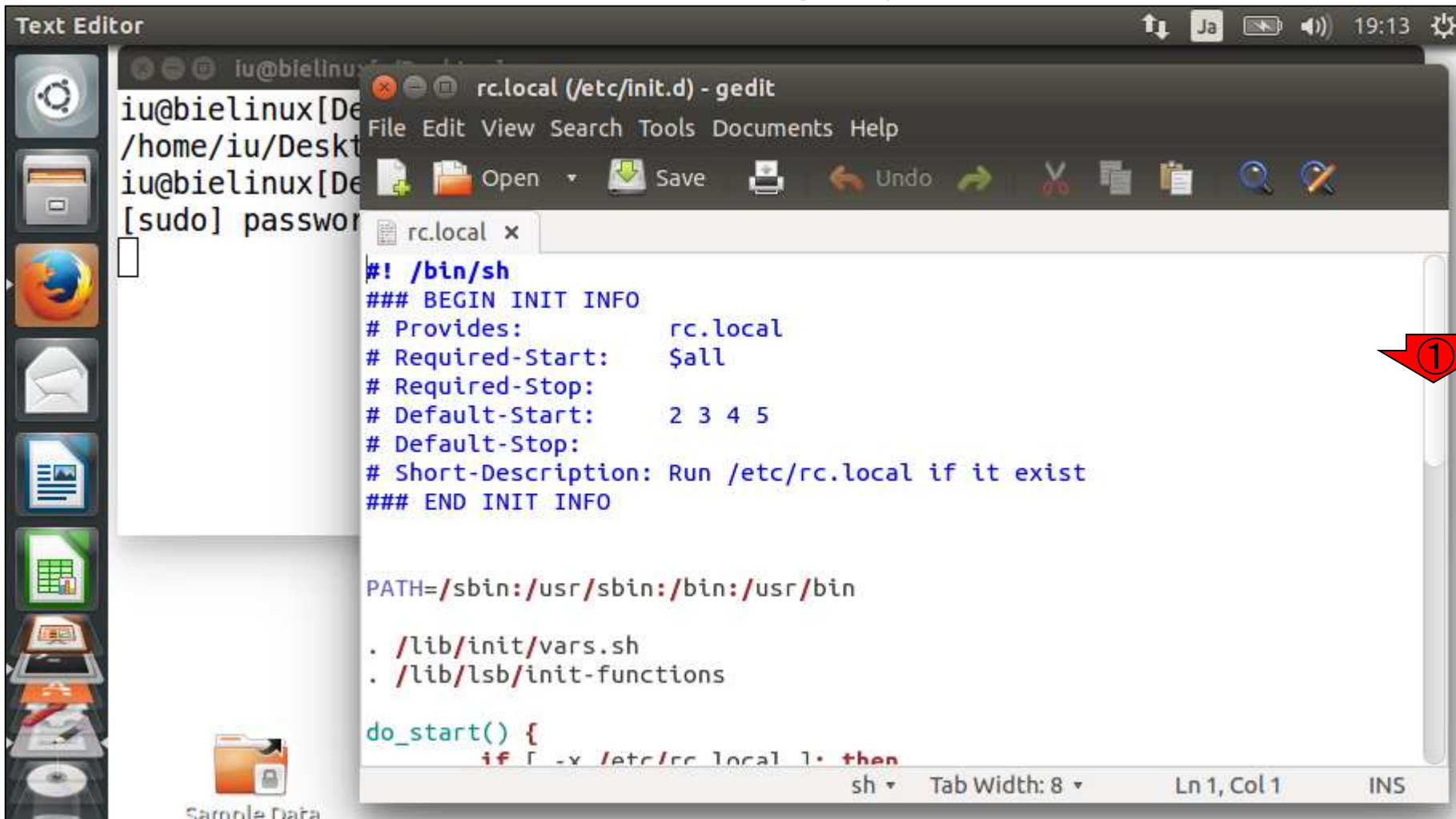
W4-5-4: ゲスト側の設定

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を入力。



geditが起動するので、①
ページの一番下に移動。

W4-5-4: ゲスト側の設定



Text Editor

```
iu@bielinux[De
/home/iu/Desk
iu@bielinux[De
[sudo] passwor
```

```
rc.local (/etc/init.d) - gedit
File Edit View Search Tools Documents Help
Open Save Undo
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

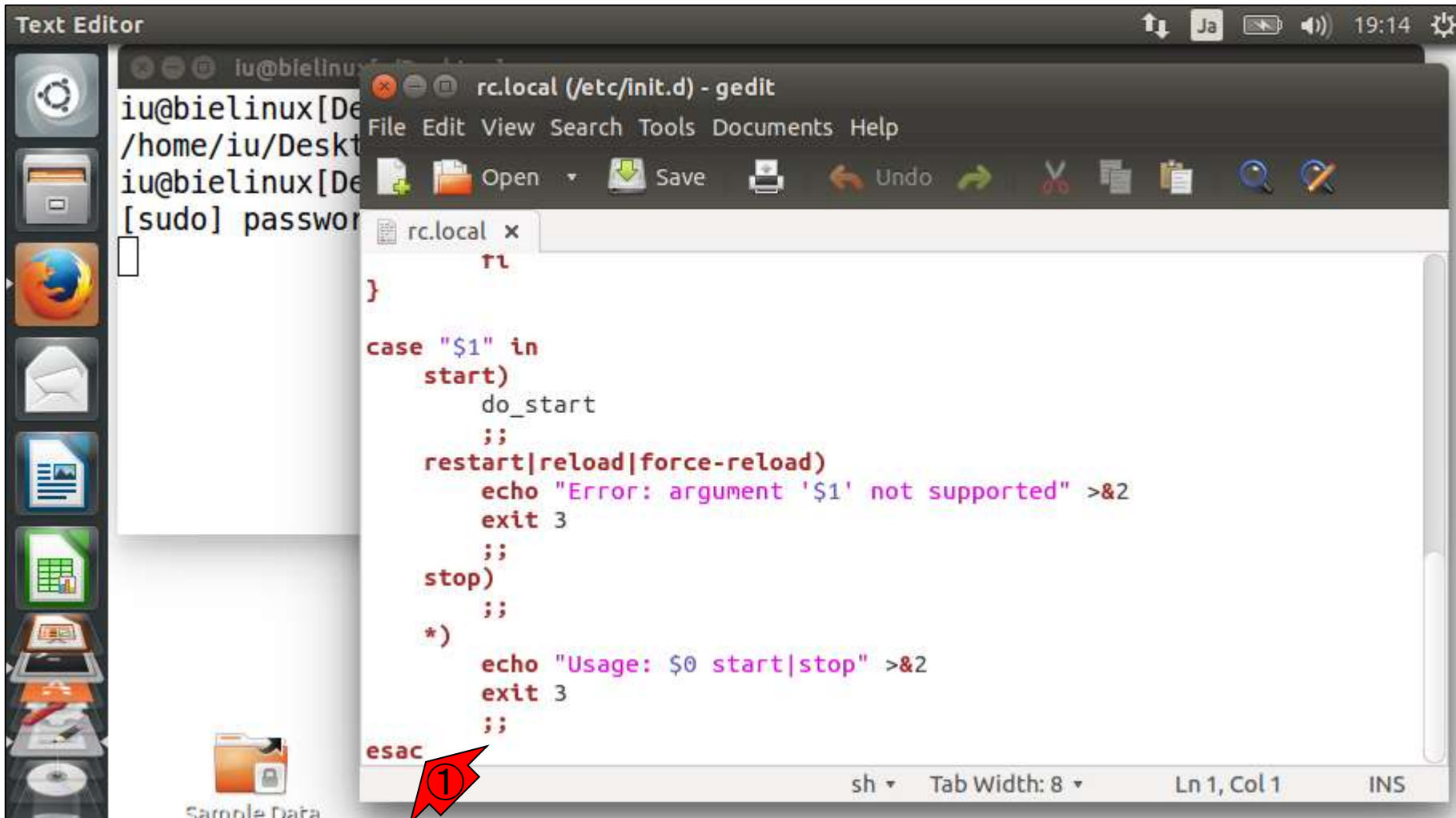
. /lib/init/vars.sh
. /lib/lsb/init-functions

do_start() {
    if [ -x /etc/rc.local ] ; then
```

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



The screenshot shows a Linux desktop environment with a text editor window titled "rc.local (/etc/init.d) - gedit". The editor displays the following configuration:

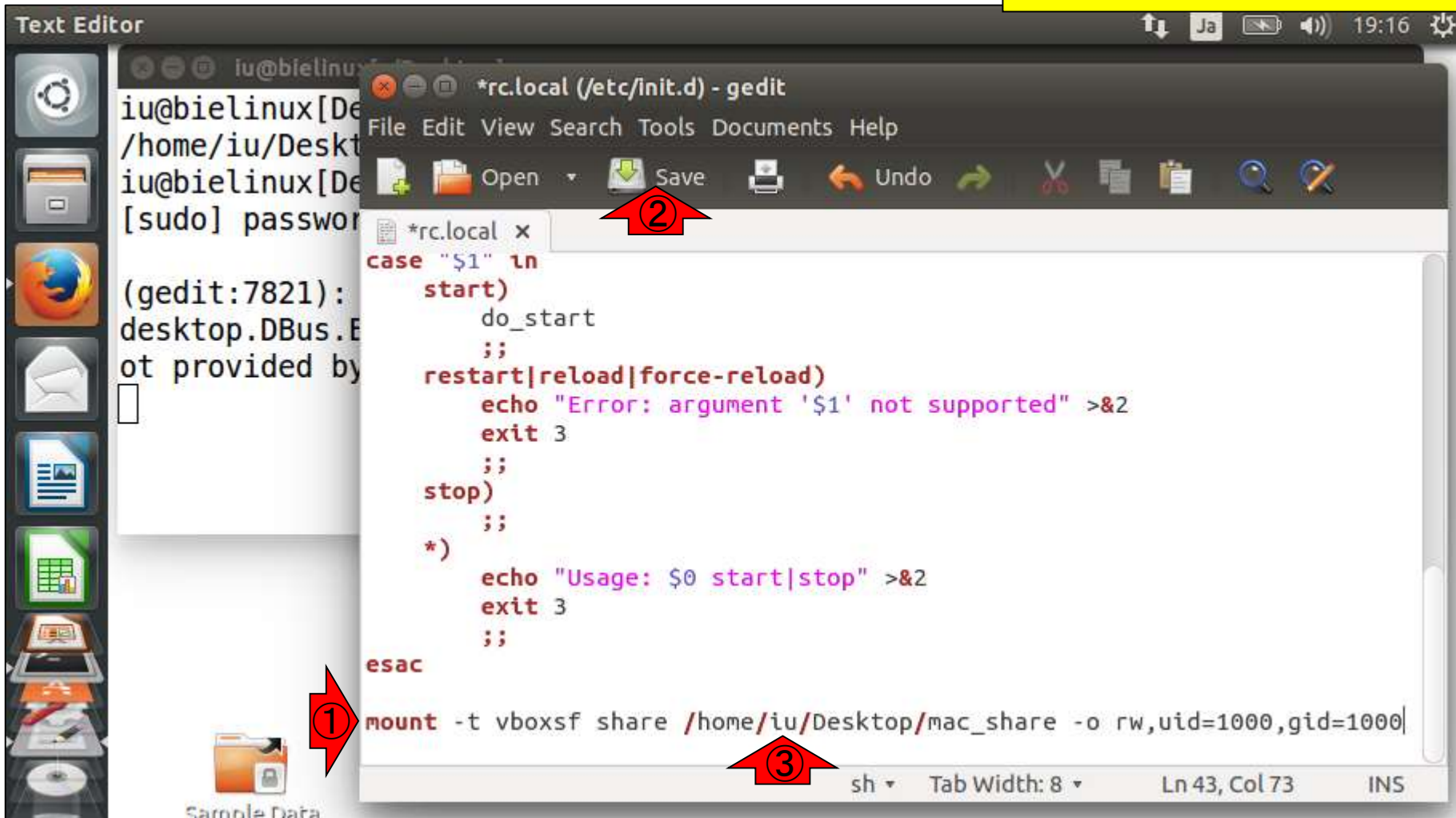
```
rc.local x
}

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
```

A red arrow with the number 1 points to the "esac" keyword at the end of the configuration block.

W4-5-4: 1行分追加

①の行を追加して、②Save。③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



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。

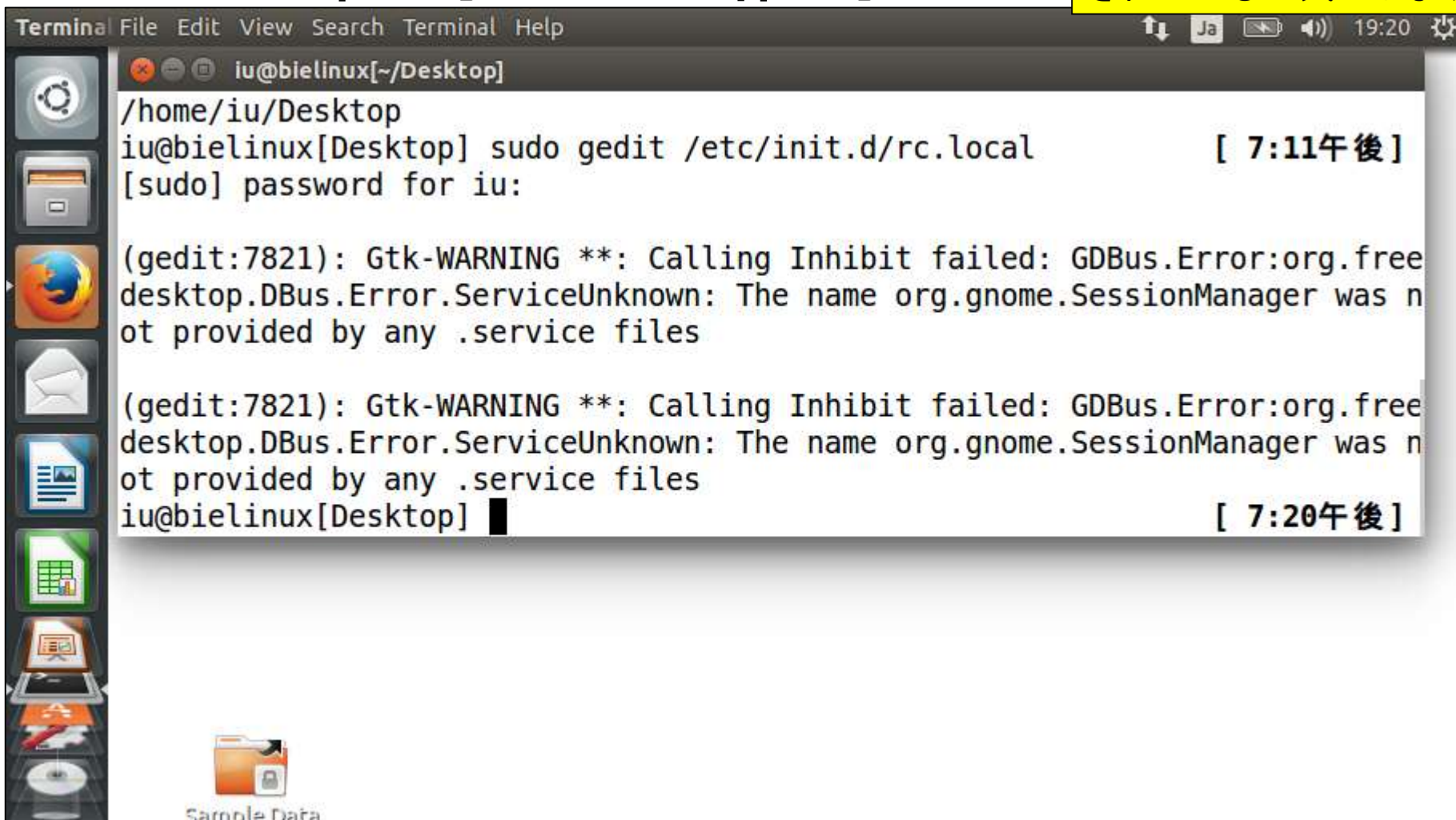
```
iu@bielinux[De
[sudo] passwor
(gedit:7821):
desktop.DBus.E
ot provided by
[

*rc.local x
case "$1" in
start)
do_start
;;
restart|reload|force-reload)
echo "Error: argument '$1' not supported"
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 File Edit View Search Terminal Help" with the user "iu@bielinux" in the directory "~/Desktop". The user runs the command "sudo gedit /etc/init.d/rc.local". The terminal shows the password prompt "[sudo] password for iu:" and then two identical error messages from gedit: "(gedit:7821): Gtk-WARNING **: Calling Inhibit failed: GDBus.Error:org.free desktop.DBus.Error.ServiceUnknown: The name org.gnome.SessionManager was not provided by any .service files". The terminal shows the prompt "iu@bielinux[Desktop] █" and the time "[7:20午後]". The desktop background is visible with a "Sample Data" folder icon.

```
Terminal File Edit View Search Terminal Help
iu@bielinux[~/Desktop]
/home/iu/Desktop
iu@bielinux[Desktop] sudo gedit /etc/init.d/rc.local [ 7:11午後]
[sudo] password for iu:
(gedit:7821): Gtk-WARNING **: Calling Inhibit failed: GDBus.Error:org.free
desktop.DBus.Error.ServiceUnknown: The name org.gnome.SessionManager was n
ot provided by any .service files
(gedit:7821): Gtk-WARNING **: Calling Inhibit failed: GDBus.Error:org.free
desktop.DBus.Error.ServiceUnknown: The name org.gnome.SessionManager was n
ot provided by any .service files
iu@bielinux[Desktop] █ [ 7:20午後]
```

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 857 12月 8 19:20 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 857 12月 8 19:20 /etc/init.d/rc.local
-rwxr-xr-x 1 root root 782 3月 13 2014 /etc/init.d/rc.local~
iu@bielinux[iu]
```

[7:24午後]
[7:24午後]
[7:24午後]
[7:24午後]

①「cd ..」は1つ上のディレクトリに移動。

W4-6: 絶対パスと相対パス

```
File Edit View Search Terminal Help
iu@bielinux[iu] pwd [ 7:35午後 ]
/home/iu
iu@bielinux[iu] ls [ 7:35午後 ]
Desktop Documents Downloads Music Pictures Public Templates Videos
iu@bielinux[iu] cd .. [ 7:35午後 ]
iu@bielinux[/home] pwd [ 7:36午後 ]
/home
iu@bielinux[/home] ls [ 7:36午後 ]
iu
iu@bielinux[/home] cd iu [ 7:36午後 ]
iu@bielinux[iu] pwd [ 7:36午後 ]
/home/iu
iu@bielinux[iu] ls [ 7:36午後 ]
Desktop Documents Downloads Music Pictures Public Templates Videos
iu@bielinux[iu] [ 7:36午後 ]
```



W4-6: 絶対パスと相対パス

```
File Edit View Search Terminal Help
iu@bielinux[iu] ls
Desktop Downloads Pictures Templates
Documents Music Public Videos
① iu@bielinux[iu] cd .
iu@bielinux[iu] pwd
/home/iu
② iu@bielinux[iu] cd ./Desktop
iu@bielinux[Desktop] pwd
/home/iu/Desktop
③ iu@bielinux[Desktop] cd ../Downloads
iu@bielinux[Downloads] pd
zsh: correct 'pd' to 'pwd' [nyae]? y
/home/iu/Downloads
④ iu@bielinux[Downloads]
iu@bielinux[Downloads] ls ../Documents
srp017156
iu@bielinux[Downloads]
```

①「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。②「cd ./Desktop」は「cd Desktop」と同じ。③は1つ上のディレクトリ中にあるDownloads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。④pwd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。⑤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。

```
[ 7:40午後]
[ 7:46午後]
[ 7:46午後]
[ 7:46午後]
```

W4-6: 絶対パスと相対パス

```
iu@bielinux[~/Documents/srp017156] ③
iu@bielinux[Documents] pwd [
/home/iu/Documents
① iu@bielinux[Documents] ls -la ~/Documents [
total 12
drwxr-xr-x  3 iu iu 4096 12月  1 22:10 .
drwxr-xr-x 20 iu iu 4096 12月  8 19:44 ..
drwxrwxr-x  2 iu iu 4096 12月  8 17:28 srp017156
iu@bielinux[Documents] ls ~/Documents/srp017156/ [
hogel.txt      hoge_JSLAB3.sh~  SRR616268_1.fastq.bz2
hoge_JSLAB3.sh JSLAB4_1.sh      SRR616268_2.fastq.bz2
iu@bielinux[Documents] cd ~/Documents/srp017156 [ 7:52午後]
iu@bielinux[srp017156] pwd [ 7:52午後]
/home/iu/Documents/srp017156
② iu@bielinux[srp017156] ls /home/iu/Documents [ 7:52午後]
srp017156
④ iu@bielinux[srp017156] █ [ 7:52午後]
```

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。①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。①がDocumentsの相対パス指定で、②が絶対パス指定。③(~/Documents/srp017156)は、④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。

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 image shows a sequence of steps to shut down a virtual machine in Oracle VM VirtualBox:

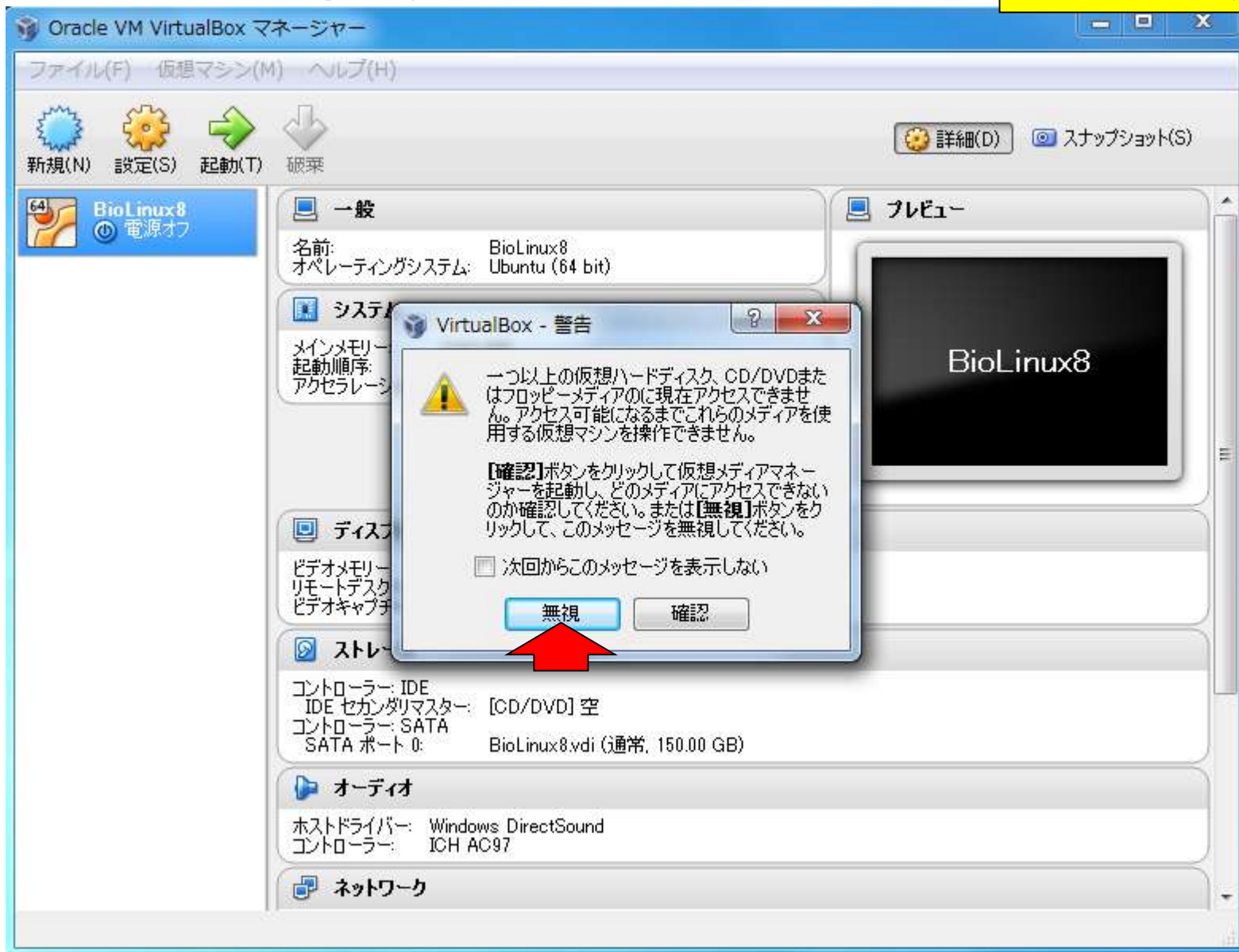
- ① The VirtualBox window is closed.
- ② The '仮想マシンを閉じる' (Close Virtual Machine) dialog box is shown, with '仮想マシンの電源オフ(P)' (Power Off VM) selected.
- ③ The 'OK' button is clicked to confirm the power-off action.
- ④ The VirtualBox Manager window is shown, indicating the VM is powered off.

The terminal window shows the following commands and output:

```
iu@bielinux[~/Documents/srp017156]
iu@bielinux[Documents] pwd
/home/iu/Documents
iu@bielinux[Documents] ls -la ~/Documents
total 16
drwxr-xr-x  4 iu
drwxr-xr-x 20 iu
drwxrwxr-x  2 iu
drwx----- 12 iu
iu@bielinux[Docum]
hogel.txt
hoge_JSLAB3.sh
iu@bielinux[Documents] cd ~/Documents/srp017156
iu@bielinux[srp017156] pwd
/home/iu/Documents/srp017156
iu@bielinux[srp017156] ls /home/iu/Document
srp017156 VirtualBox Dropped Files
iu@bielinux[srp017156]
```

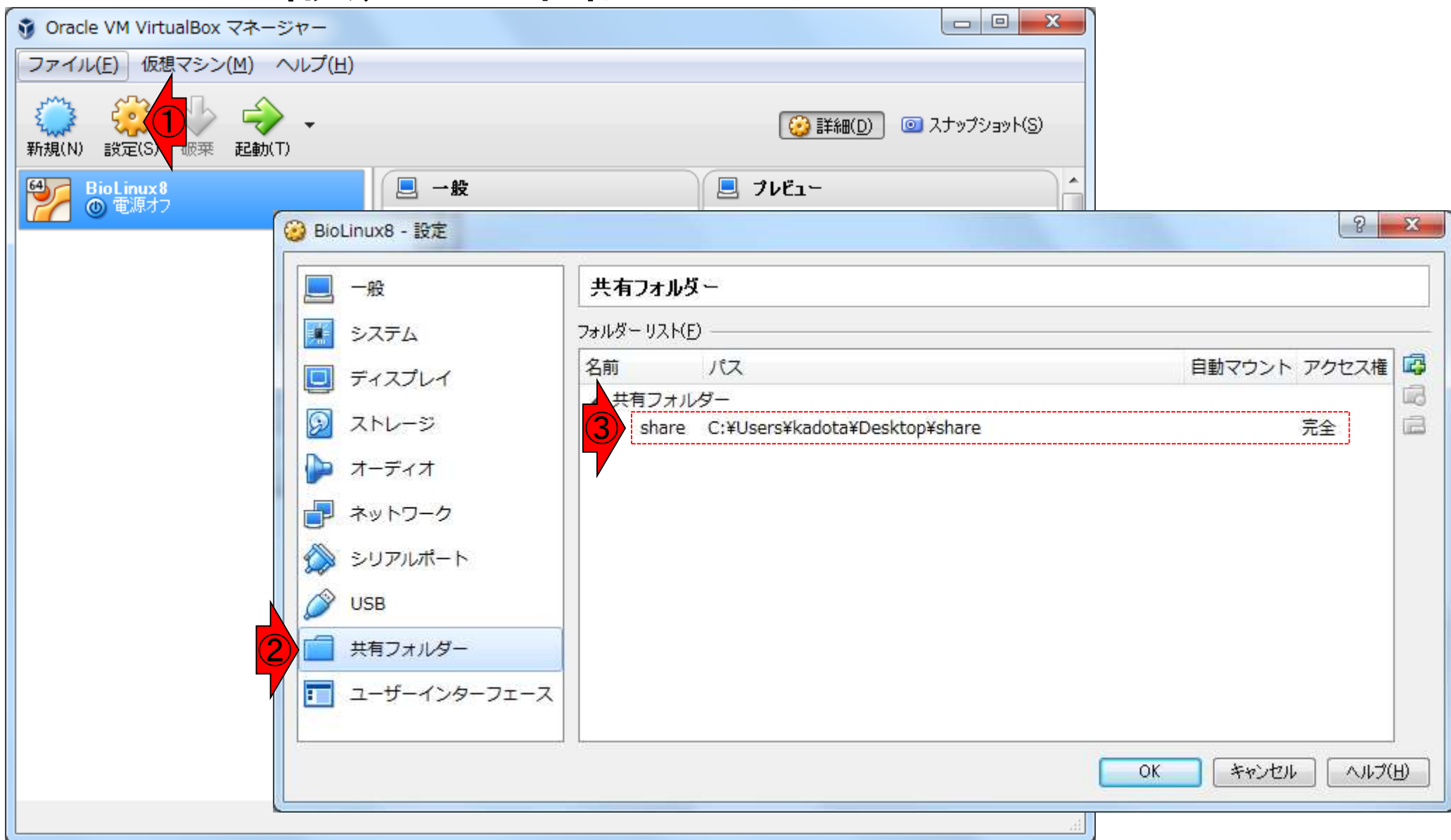

W4-7: 設定の確認

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。



W4-7: 設定の確認

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。①で見えるように、mac_shareが反転されていれば基本的にOK。②実行結果の赤枠でみられるファイルはヒトによって異なる。重要なのは「ホストOS - デスクトップ - share」フォルダ中の中身と同じこと。

```
iu@bielinux[iu] ls
Desktop      Downloads  Pictures   Templates
Documents    Music      Public     Videos

iu@bielinux[iu] cd Desktop
iu@bielinux[Desktop] ls
Bio-Linux Documentation  hoge  mac_share  Sample Data

① iu@bielinux[Desktop] ls -lh
total 12K
drwxr-xr-x 2 iu iu 4.0K 11月 17 13:44 Bio-Linux Documentation
drwxrwxr-x 2 iu iu 4.0K 11月 30 17:08 hoge
drwxrwxrwx 1 iu iu 4.0K 12月  8 17:34 mac_share
lrwxrwxrwx 1 iu iu  29 11月 17 13:44 Sample Data -> /usr/local/
bioinf/sampledata/

② iu@bielinux[Desktop] ls -lh mac_share
total 2.9M
-rwxrwxrwx 1 iu iu  58K 11月 25 19:36 hoge.png
-rwxrwxrwx 1 iu iu   364 12月  8 17:34 JSLAB4_1.sh
-rwxrwxrwx 1 iu iu 2.8M  3月 10 2014 Lactobacillus_casei_12a.GC
A_000309565.1.22.dna.toplevel.fa

iu@bielinux[Desktop]
```


つまり、①「ホストOS -デスクトップ - share」フォルダ中の中身と同じこと。

W4-7: 設定の確認

The screenshot displays the Oracle VM VirtualBox Manager interface. A red arrow labeled '1' points to the 'BioLinux8' VM in the left sidebar. Another red arrow labeled '1' points to the 'share' folder in the Windows File Explorer window. The terminal window shows the following commands and output:

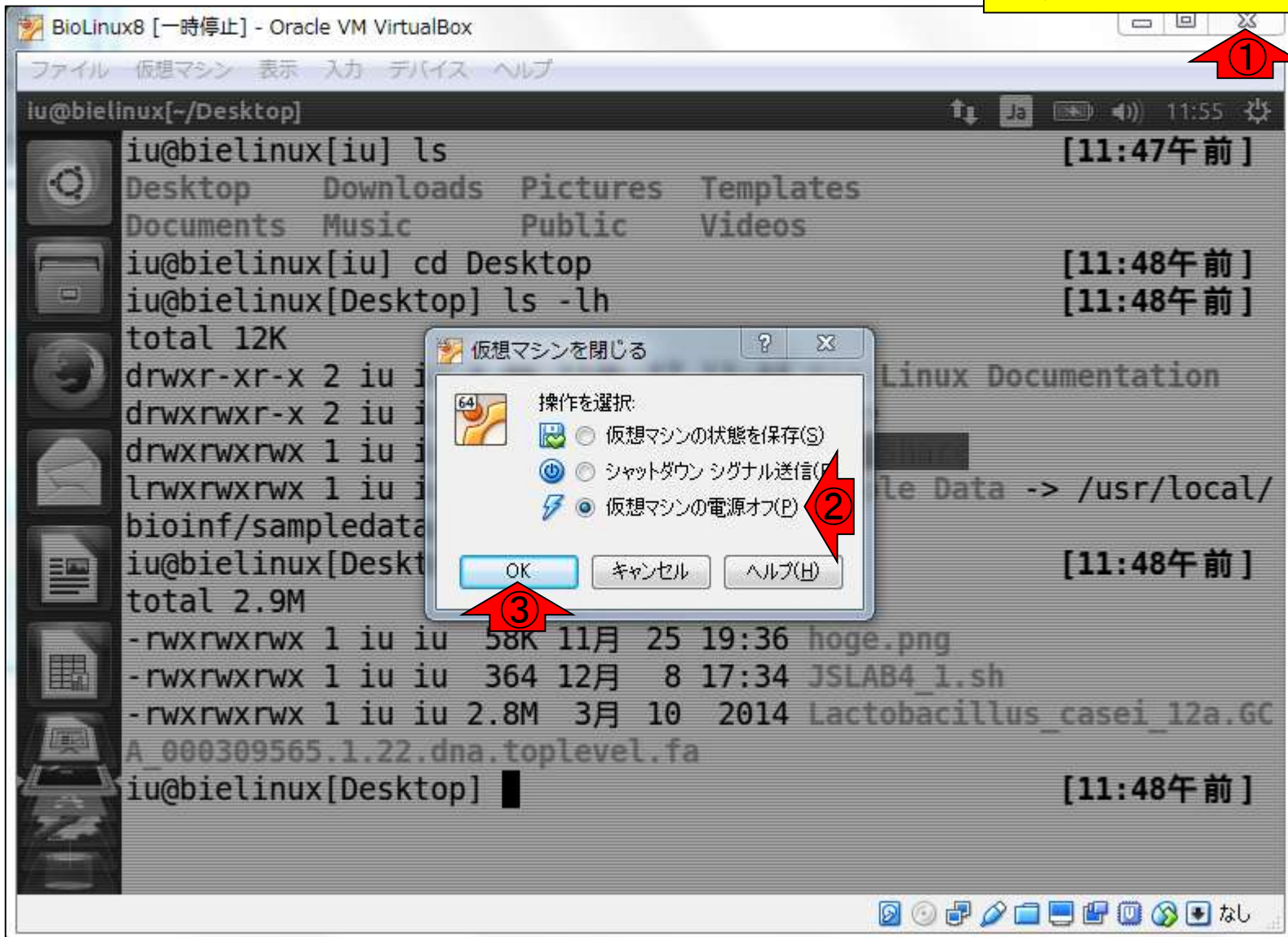
```
iu@bielinux[~/Desktop] ls
Desktop  Downloads  Pictures  Templates
Documents Music      Public    Videos
iu@bielinux[iu] cd Desktop
iu@bielinux[Desktop] ls -lh
total 12K
drwxr-xr-x 2 iu iu 4.0K 11月 17 13:44 Bio-Linear Documentation
drwxrwxr-x 2 iu iu 4.0K 11月 30 17:08 hoge
drwxrwxrwx 1 iu iu 4.0K 12月 8 17:34 hoge.png
lrwxrwxrwx 1 iu iu 29 11月 17 13:44 Sample Data -> /usr/local/
bioinf/sampledata/
iu@bielinux[Desktop] ls -lh mac_share
total 2.9M
-rwxrwxrwx 1 iu iu 58K 11月 25 19:36 hoge.png
-rwxrwxrwx 1 iu iu 364 12月 8 17:34 JSLAB4_1.sh
-rwxrwxrwx 1 iu iu 2.8M 3月 10 2014 Lactobacillus_casei_12a.GC
A_000309565.1.22.dna.toplevel.fa
iu@bielinux[Desktop]
```

The File Explorer window shows the contents of the 'share' folder:

名前	日付時刻	種類	サイズ
hoge.png	2015/11/25 19:36	PNG イメージ	58 KB
JSLAB4_1.sh	2015/12/08 17:34	SH ファイル	1 KB
Lactobacillus_cas...	2014/03/10 22:32	FA ファイル	2,868 KB

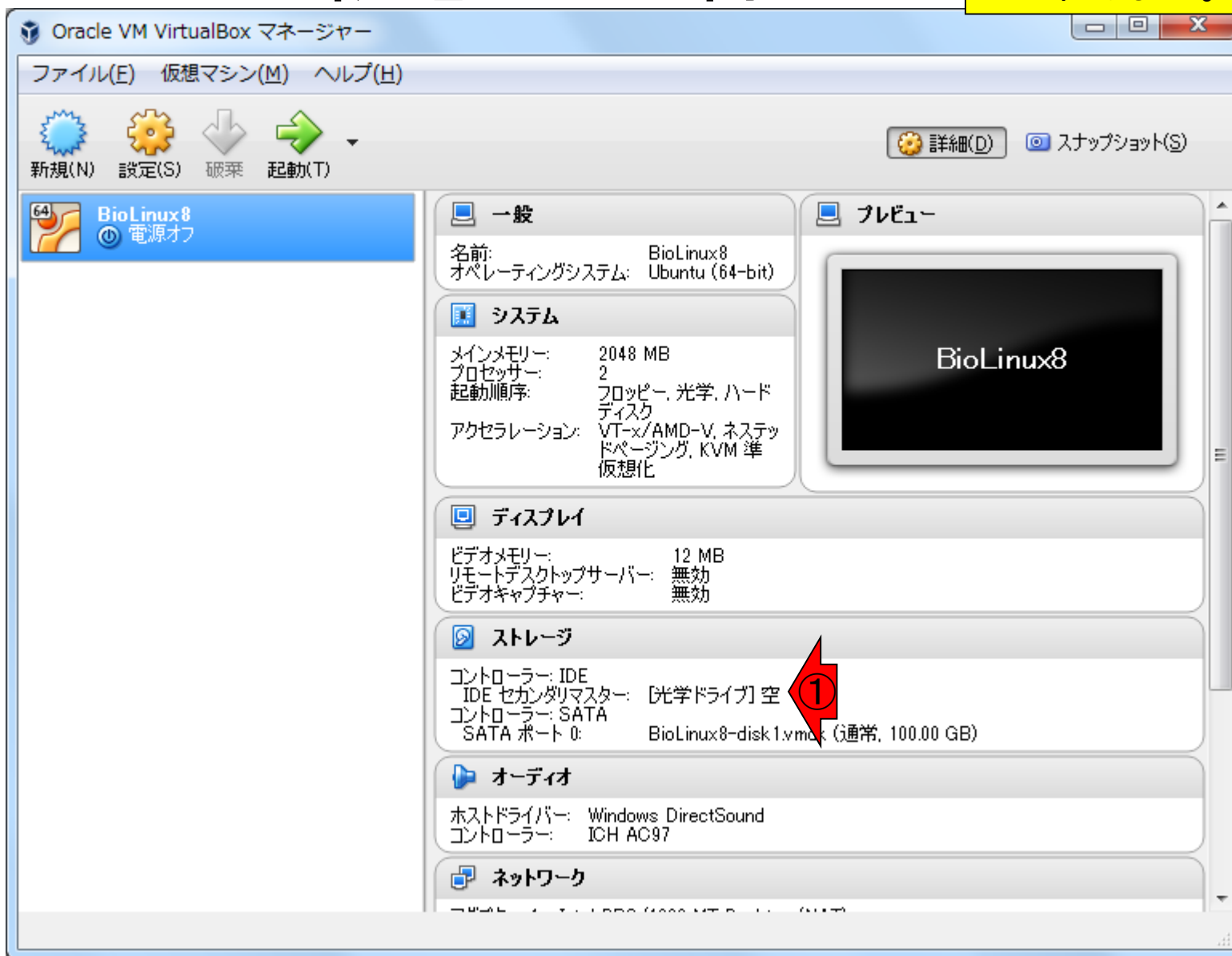
W4-8: 仮想CDの除去

①ゲストOS(BioLinux8)を一旦終了。②「仮想マシンの電源オフ」、③OK。W4-5-3でうまくいかなかったヒトはやる必要あり



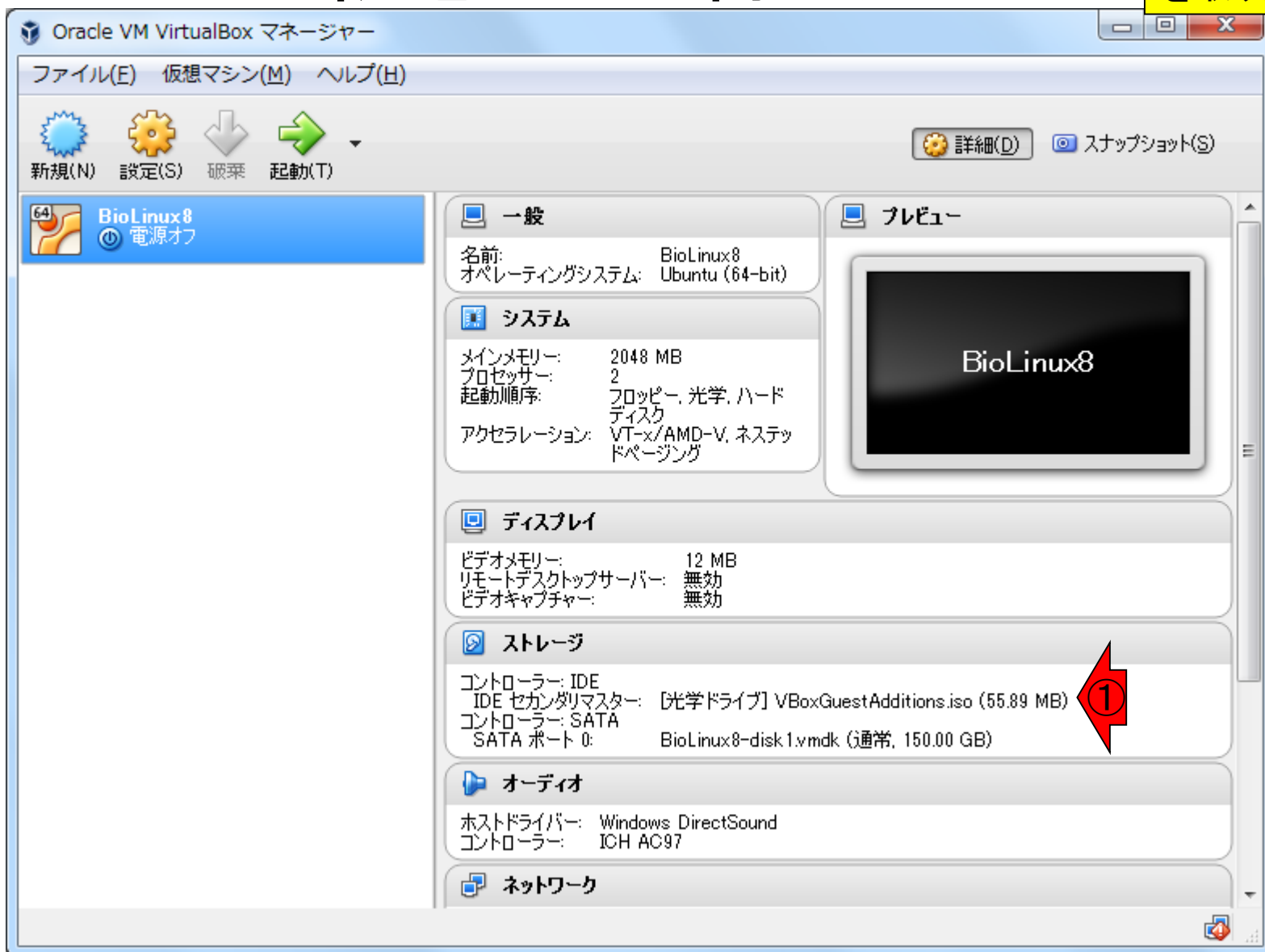
W4-8: 仮想CDの除去

①ここを「空」にするのが目的。ここで示すように、既に「[光学ドライブ]空」になっていればよい。なっていないければ…。



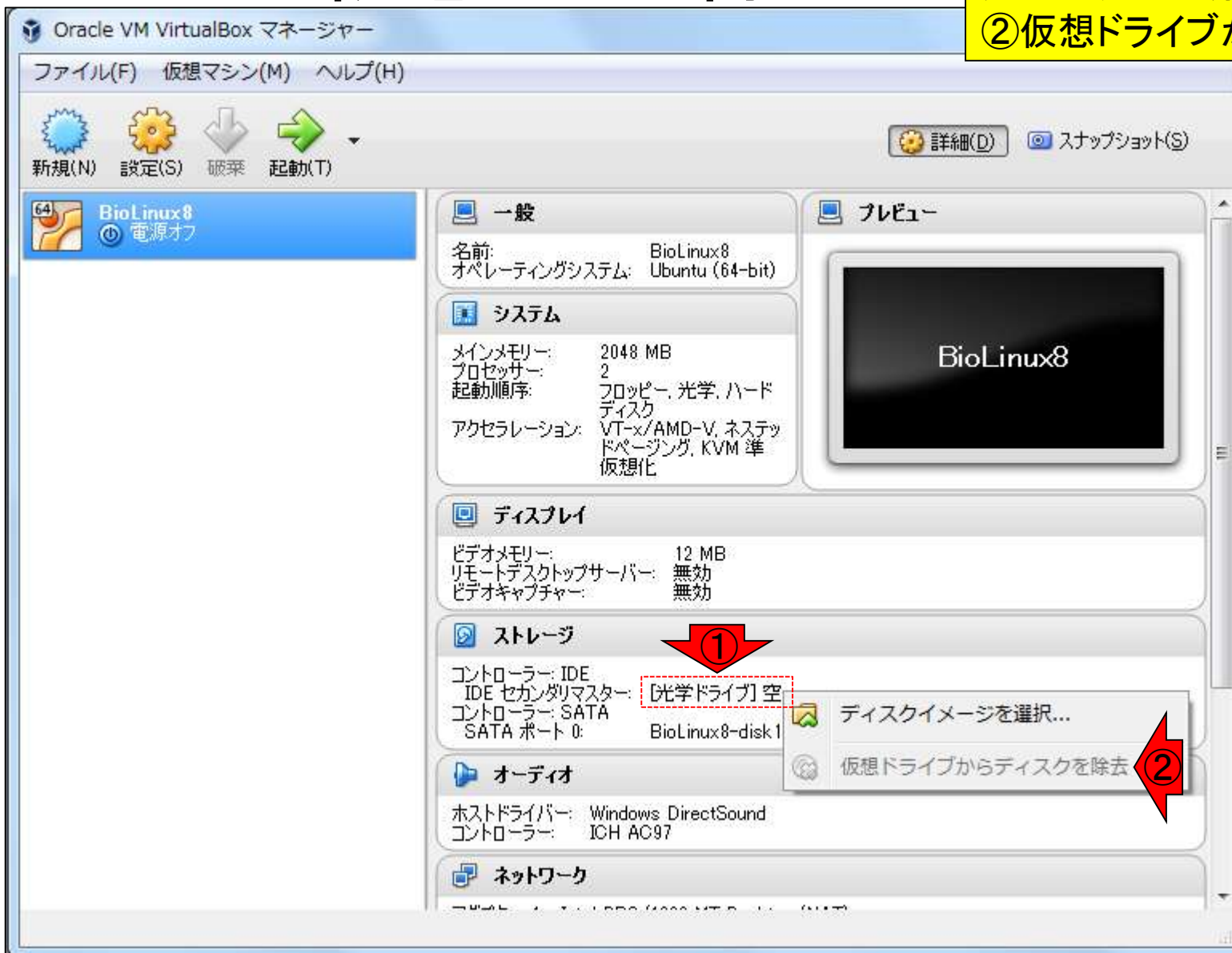
W4-8: 仮想CDの除去

①こんな感じに見えるはず。VBoxGuestAdditions.isoを取り出す作業を行います。



W4-8: 仮想CDの除去

- ① (既に空になっているが、ここに VBoxGuestAdditions.iso があるという想定で行うと...) 赤枠付近で右クリックして
- ② 仮想ドライブからディスクを除去。



W5-1: シェルスクリプト

①共有フォルダ中のJSLAB4_1.shと、②異なるディレクトリ上にある③W4-1で取得したJSLAB4_1.shの中身が同じであることを④diffで確認。

```
iu@bielinux[Desktop] pwd [12:22午後]
/home/iu/Desktop
iu@bielinux[Desktop] ls [12:22午後]
Bio-Linux Documentation hoge mac share Sample Data
① iu@bielinux[Desktop] more mac_share/JSLAB4_1.sh [12:22午後]
#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
② iu@bielinux[Desktop] cd ~/Documents/srp017156 [12:22午後]
iu@bielinux[srp017156] pwd [12:22午後]
/home/iu/Documents/srp017156
iu@bielinux[srp017156] ls [12:22午後]
hogel.txt hoge_JSLAB3.sh~ SRR616268_1.fastq.bz2
hoge_JSLAB3.sh JSLAB4_1.sh SRR616268_2.fastq.bz2
④ iu@bielinux[srp017156] diff ~/Desktop/mac_share/JSLAB4_1.sh JSLAB4_1.sh [12:23午後]
iu@bielinux[srp017156] █
```



W5-1: シェルスクリプト

「~/Documents/srp017156」中のJSLAB4_1.shを①一旦削除。②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。③moreで表示させた赤枠の中身のうち、wgetコマンド部分は行頭に#を付加しているので実行されない。

```
File Edit View Search Terminal Help
iu@bielinux[srp017156] rm -f JSLAB4_1.sh
iu@bielinux[srp017156] ls
hogel.txt      hoge_JSLAB3.sh~      SRR61
hoge_JSLAB3.sh SRR616268_1.fastq.bz2
iu@bielinux[srp017156] cp ~/Desktop/mac_share/JSLAB4_1.sh .      [ 1:23午後]
iu@bielinux[srp017156] ls                                          [ 1:23午後]
hogel.txt      hoge_JSLAB3.sh~      SRR616268_1.fastq.bz2
hoge_JSLAB3.sh JSLAB4_1.sh          SRR616268_2.fastq.bz2
iu@bielinux[srp017156] more JSLAB4_1.sh                          [ 1:23午後]
#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] █                                          [ 1:23午後]
```

W5-1: シェルスクリプト

①シェルスクリプト実行時に、timeコマンドをつけて実行時間も表示。実行時間で主に見るのは赤枠のところ。この場合25.63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中のものと基本的に同じ。bzip2圧縮ファイルを解凍し、最初の400万行を②青枠右側のファイル名で保存している。600万行から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。

```
iu@bielinux[srp017156] pwd
/home/iu/Documents/srp017156
iu@bielinux[srp017156] ls -lh
total 157M
-rw-rw-r-- 1 iu iu 157 12月 8 16:08 hoge1.tx
-rwxr-xr-x 1 iu iu 24 12月 8 16:12 hoge_JSL
-rwxr-xr-x 1 iu iu 24 12月 8 16:12 hoge_JSL
-rwxrwxr-x 1 iu iu 364 12月 9 13:23 JSLAB4_1
-rw-rw-r-- 1 iu iu 83M 12月 3 14:56 SRR616268_1.fastq.bz2
-rw-rw-r-- 1 iu iu 74M 12月 3 15:51 SRR616268_2.fastq.bz2
iu@bielinux[srp017156] time ./JSLAB4_1.sh
./JSLAB4_1.sh 25.63s user 2.77s system 48% cpu 58.116 total
iu@bielinux[srp017156] ls -lh
total 740M
-rw-rw-r-- 1 iu iu 157 12月 8 16:08 hoge1.txt
-rwxr-xr-x 1 iu iu 24 12月 8 16:12 hoge_JSLAB3.sh
-rwxr-xr-x 1 iu iu 24 12月 8 16:12 hoge_JSLAB3.sh~
-rwxrwxr-x 1 iu iu 364 12月 9 13:23 JSLAB4_1.sh
-rw-rw-r-- 1 iu iu 83M 12月 3 14:56 SRR616268_1.fastq.bz2
-rw-rw-r-- 1 iu iu 74M 12月 3 15:51 SRR616268_2.fastq.bz2
-rw-rw-r-- 1 iu iu 306M 12月 9 13:31 SRR616268sub_1.fastq?
-rw-rw-r-- 1 iu iu 278M 12月 9 13:32 SRR616268sub_2.fastq?
iu@bielinux[srp017156]
```

[1:30午後]

[1:32午後]

[1:32午後]



W5-2: 改行コード

①青枠の出力ファイル名末尾の「?」に着目。この「?」は、②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、④「?」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 740M
-rw-rw-r-- 1 iu iu 157 12月 8 16:00
-rwxr-xr-x 1 iu iu 24 12月 8 16:12 hoge_JSLAB3.sh
-rwxr-xr-x 1 iu iu 24 12月 8 16:12 hoge_JSLAB3.sh~
-rwxrwxr-x 1 iu iu 364 12月 9 13:23 JSLAB4_1.sh
-rw-rw-r-- 1 iu iu 83M 12月 3 14:56 SRR616268_1.fastq.bz2
-rw-rw-r-- 1 iu iu 74M 12月 3 15:51 SRR616268_2.fastq.bz2
-rw-rw-r-- 1 iu iu 306M 12月 9 13:31 SRR616268sub_1.fastq?
-rw-rw-r-- 1 iu iu 278M 12月 9 13:32 SRR616268sub_2.fastq?
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] rm SRR616268sub_*
rm: remove regular file 'SRR616268sub_1.fastq\r'? █
```



[1:57午後]



[1:57午後]



W5-2: 改行コード

①青枠の出力ファイル名末尾の「?」に着目。この「?」は、②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、④「?」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。⑤とりあえずyを押して2つのファイルともに消す。

```
iu@bielinux[srp017156] ls -lh
total 740M
-rw-rw-r-- 1 iu iu 157 12月 8 16:0
-rwxr-xr-x 1 iu iu 24 12月 8 16:1
-rwxr-xr-x 1 iu iu 24 12月 8 16:12 hoge_JSLAB3.sh~
-rwxrwxr-x 1 iu iu 364 12月 9 13:23 JSLAB4_1.sh
-rw-rw-r-- 1 iu iu 83M 12月 3 14:56 SRR616268_1.fastq.bz2
-rw-rw-r-- 1 iu iu 74M 12月 3 15:51 SRR616268_2.fastq.bz2
-rw-rw-r-- 1 iu iu 306M 12月 9 13:31 SRR616268sub_1.fastq?
-rw-rw-r-- 1 iu iu 278M 12月 9 13:32 SRR616268sub_2.fastq?
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] rm SRR616268sub_*
rm: remove regular file 'SRR616268sub_1.fastq\r'? y
rm: remove regular file 'SRR616268sub_2.fastq\r'? y
iu@bielinux[srp017156]
```

[1:57午後]

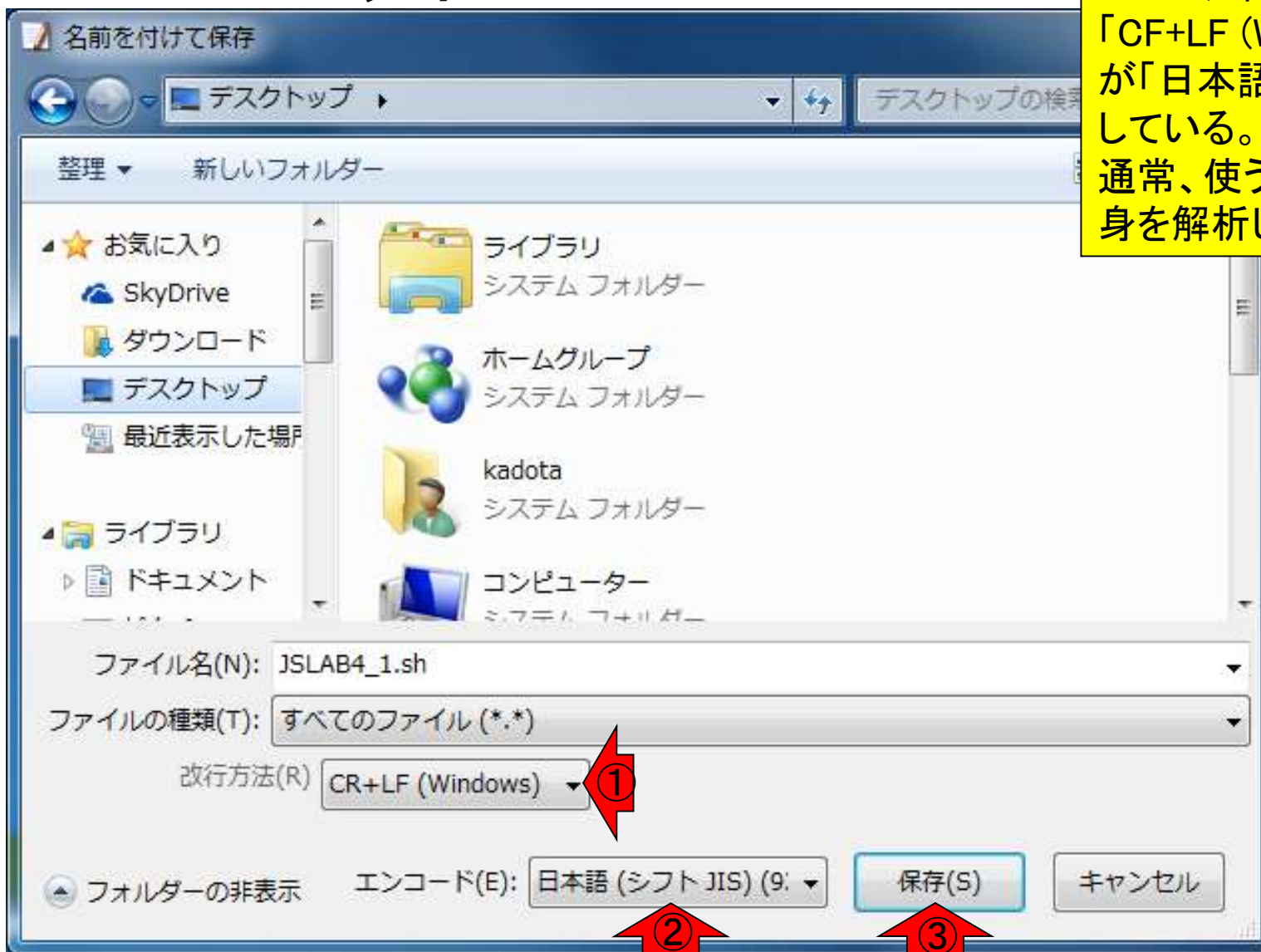
[1:57午後]

[2:05午後]



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。①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      hoge_JSLAB3.sh~ SR
hoge_JSLAB3.sh JSLAB4_1.sh      SRR616268_2.fastq.bz2
iu@bielinux[srp017156] od -c JSLAB4_1.sh | head -n 15 [ 2:07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 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] [ 2:07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             hoge_JSLAB3.sh~  SRR616268_1.fastq.bz2
hoge_JSLAB3.sh       JSLAB4_1.sh     SRR616268_2.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 _ 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 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]
```

W5-2: 改行コード

①perlを用いてJSLAB4_1.sh中の¥r¥nを¥nに置換。**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ファイルhoge.shに保存している。

```
iu@bielinux[srp017156] pwd
/home/iu/Documents/srp017156
iu@bielinux[srp017156] ls
hoge1.txt      hoge_JSLAB3.sh~  SRR616268_1.fastq.bz2
hoge_JSLAB3.sh JSLAB4_1.sh     SRR616268_2.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
00000340 S R R 6 1 6 2 6 8 _ 1 . f a s t
iu@bielinux[srp017156]
```

[2:13午後]

[2:13午後]

[2:13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、①od 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          hoge_JSLAB3.sh~  SRR616268_1.fastq.bz2
hoge_JSLAB3.sh    JSLAB4_1.sh     SRR616268_2.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
00000340 S R R 6 1 6 2 6 8 _ 1 . f a s t
iu@bielinux[srp017156]
```

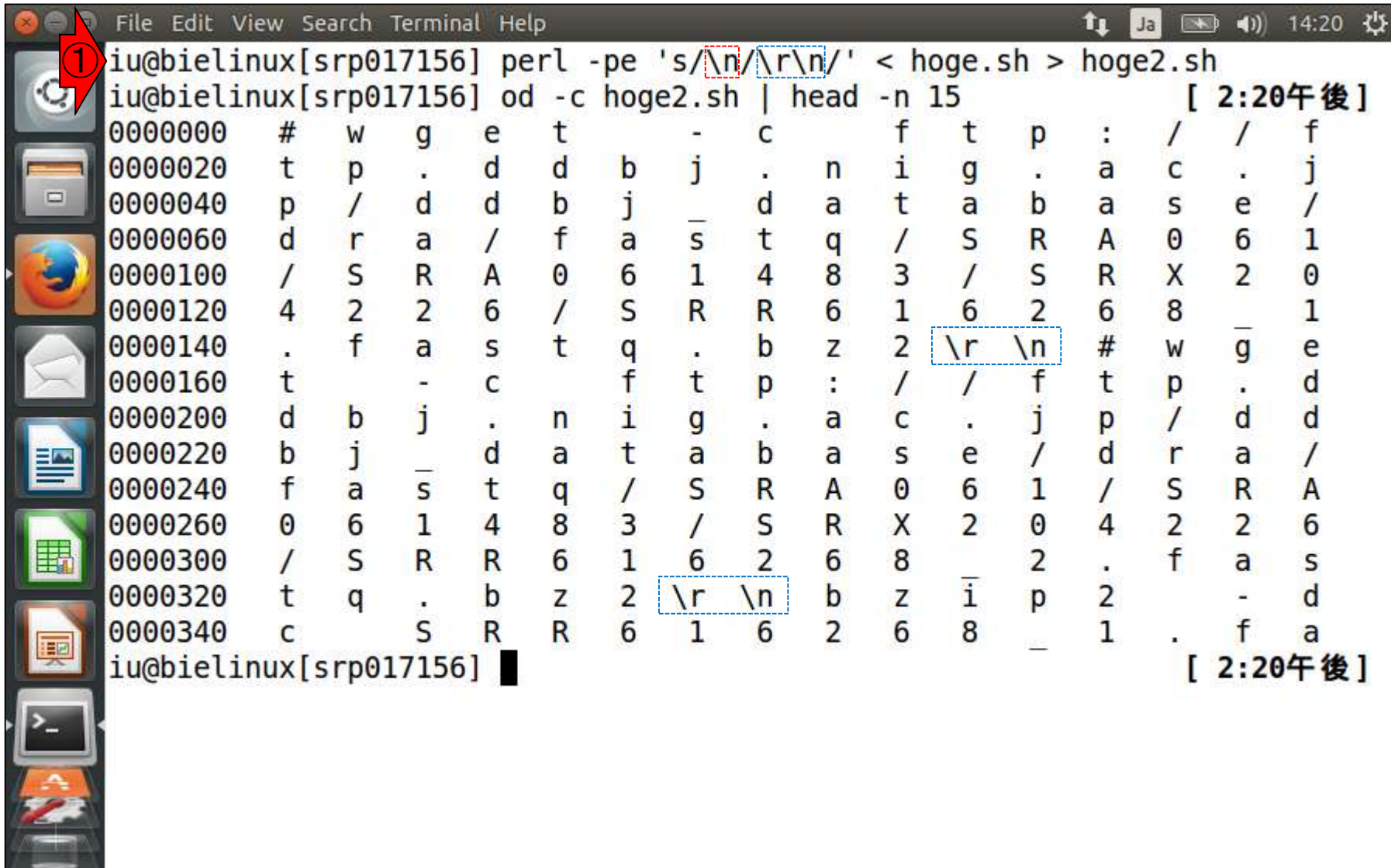

W5-2: 改行コード

① $\$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 [ 2:12午後 ]
/home/iu/Documents/srp017156
iu@bielinux[srp017156] ls [ 2:13午後 ]
hoge1.txt             hoge_JSLAB3.sh~  SRR616268_1.fastq.bz2
hoge_JSLAB3.sh       JSLAB4_1.sh     SRR616268_2.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 [ 2:13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 \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
00000340 S R R 6 1 6 2 6 8 _ 1 . f a s t
iu@bielinux[srp017156] [ 2:13午後 ]
```

逆パターンで、①Linux → Windowsに改行コードを変換したい場合。's/置換前/置換後/'が基本形。

W5-2: 改行コード



```
iu@bielinux[srp017156] perl -pe 's/\n/\r\n/' < hoge.sh > hoge2.sh
iu@bielinux[srp017156] od -c hoge2.sh | head -n 15
[ 2:20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] [ 2:20午後 ]
```

W5-2: 改行コード

改行コードは、見た目上の気持ち悪さという問題だけではない。尚、MacとLinuxは改行コードが同じなので、そのような問題は起きない。

- 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交換してくれるため、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

①Linux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。②chmodコマンドで実行権限を与えて、③再度実行。ほぼ一瞬(0.003秒)でサブセットの抽出が終了。④SRR616268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
total 157M
-rw-rw-r-- 1 iu iu 157 12月 8 16:08 hoge1.
-rw-rw-r-- 1 iu iu 364 12月 9 14:20 hoge2.sh
-rwxr-xr-x 1 iu iu 24 12月 8 16:12 hoge_JSLAB3.sh
-rwxr-xr-x 1 iu iu 24 12月 8 16:12 hoge_JSLAB3.sh~
-rw-rw-r-- 1 iu iu 360 12月 9 14:13 hoge.sh
-rwxrwxr-x 1 iu iu 364 12月 9 13:23 JSLAB4_1.sh
-rw-rw-r-- 1 iu iu 83M 12月 3 14:56 SRR616268_1.fastq.bz2
-rw-rw-r-- 1 iu iu 74M 12月 3 15:51 SRR616268_2.fastq.bz2
① iu@bielinux[srp017156] time ./hoge.sh [ 2:32午後 ]
zsh: permission denied: ./hoge.sh
./hoge.sh 0.00s user 0.00s system 63% cpu 0.003 total
② iu@bielinux[srp017156] chmod 755 hoge.sh [ 2:32午後 ]
iu@bielinux[srp017156] ./hoge.sh [ 2:32午後 ]
④ iu@bielinux[srp017156] ls -lh SRR616268* [ 2:33午後 ]
-rw-rw-r-- 1 iu iu 83M 12月 3 14:56 SRR616268_1.fastq.bz2
-rw-rw-r-- 1 iu iu 74M 12月 3 15:51 SRR616268_2.fastq.bz2
-rw-rw-r-- 1 iu iu 306M 12月 9 14:32 SRR616268sub_1.fastq
-rw-rw-r-- 1 iu iu 278M 12月 9 14:33 SRR616268sub_2.fastq
iu@bielinux[srp017156] [ 2:33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 740M
-rw-rw-r-- 1 iu iu 157 12月 8 16:08 hoge1
-rw-rw-r-- 1 iu iu 364 12月 9 14:20 hoge2
-rwxr-xr-x 1 iu iu 24 12月 8 16:12 hoge_JSLAB3.sh
-rwxr-xr-x 1 iu iu 24 12月 8 16:12 hoge_JSLAB3.sh~
-rwxr-xr-x 1 iu iu 360 12月 9 14:13 hoge.sh
-rwxrwxr-x 1 iu iu 364 12月 9 13:23 JSLAB4_1.sh
-rw-rw-r-- 1 iu iu 83M 12月 3 14:56 SRR616268_1.fastq.bz2
-rw-rw-r-- 1 iu iu 74M 12月 3 15:51 SRR616268_2.fastq.bz2
-rw-rw-r-- 1 iu iu 306M 12月 9 14:32 SRR616268sub_1.fastq
-rw-rw-r-- 1 iu iu 278M 12月 9 14:33 SRR616268sub_2.fastq

① iu@bielinux[srp017156] file hoge.sh [ 2:44午後 ]
hoge.sh: ASCII text

② iu@bielinux[srp017156] file JSLAB4_1.sh [ 2:44午後 ]
JSLAB4_1.sh: ASCII text, with CRLF line terminators

③ iu@bielinux[srp017156] file *.bz2 [ 2:44午後 ]
SRR616268_1.fastq.bz2: bzip2 compressed data, block size = 900k
SRR616268_2.fastq.bz2: bzip2 compressed data, block size = 900k

④ iu@bielinux[srp017156] file *.fastq [ 2:44午後 ]
SRR616268sub_1.fastq: ASCII text
SRR616268sub_2.fastq: ASCII text

iu@bielinux[srp017156] [ 2:44午後 ]
```


W5-4: 元に戻る

W1-4と同じ状態に戻す。①のrmコマンド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
total 740M
-rw-rw-r-- 1 iu iu 157 12月 8 16:08 hoge1.txt
-rw-rw-r-- 1 iu iu 364 12月 9 14:20 hoge2.sh
-rwxr-xr-x 1 iu iu 24 12月 8 16:12 hoge_JSLAB3.sh
-rwxr-xr-x 1 iu iu 24 12月 8 16:12 hoge_JSLAB3.sh~
-rwxr-xr-x 1 iu iu 360 12月 9 14:13 hoge.sh
-rwxrwxr-x 1 iu iu 364 12月 9 13:23 JSLAB4_1.sh
-rw-rw-r-- 1 iu iu 83M 12月 3 14:56 SRR616268_1.fastq.bz2
-rw-rw-r-- 1 iu iu 74M 12月 3 15:51 SRR616268_2.fastq.bz2
-rw-rw-r-- 1 iu iu 306M 12月 9 14:32 SRR616268sub_1.fastq
-rw-rw-r-- 1 iu iu 278M 12月 9 14:33 SRR616268sub_2.fastq
iu@bielinux[srp017156] rm -f hoge*
iu@bielinux[srp017156] rm -f JSLAB4_1.sh
iu@bielinux[srp017156] rm -f *.fastq
iu@bielinux[srp017156] ls -lh
total 157M
-rw-rw-r-- 1 iu iu 83M 12月 3 14:56 SRR616268_1.fastq.bz2
-rw-rw-r-- 1 iu iu 74M 12月 3 15:51 SRR616268_2.fastq.bz2
iu@bielinux[srp017156] pwd
/home/iu/Documents/srp017156
iu@bielinux[srp017156]
```

[2:48午後]

[2:48午後]

[2:48午後]

[2:48午後]

[2:48午後]

[2:48午後]

[2:48午後]



W6-1: ダウンロード

①JSLAB4_1_Linux.shをwgetコマンドで取得し、②file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)ことを確認

```
iu@bielinux[srp017156] wget -c http://www.iu.a.u-tokyo.ac.jp/~kadota/book/J
SLAB4_1_Linux.sh
--2015-12-09 14:56:49-- 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

2015-12-09 14:56:49 (1.81 MB/s) - 'JSLAB4_1_Linux.sh' saved [360/360]

iu@bielinux[srp017156] ls -lh                                     [ 2:56午後 ]
total 157M
-rw-rw-r-- 1 iu iu 360  5月  1  2015 JSLAB4_1_Linux.sh
-rw-rw-r-- 1 iu iu 83M 12月  3 14:56 SRR616268_1.fastq.bz2
-rw-rw-r-- 1 iu iu 74M 12月  3 15:51 SRR616268_2.fastq.bz2
iu@bielinux[srp017156] file JSLAB4_1_Linux.sh                   [ 2:56午後 ]
JSLAB4_1_Linux.sh: ASCII text
iu@bielinux[srp017156] █                                       [ 2:56午後 ]
```

①パーミッションの変更(実行権限の付与)と②more
で中身を表示。⑤全部で4行からなることがわかる。

W6-1: 確認

```
iu@bielinux[srp017156] ls -lh [ 3:01午後 ]
total 157M
-rw-rw-r-- 1 iu iu 360 5月 1 2015 JSLAB4_1_Linux.sh
-rw-rw-r-- 1 iu iu 83M 12月 3 14:56 SRR616268_1.fastq.bz2
-rw-rw-r-- 1 iu iu 74M 12月 3 15:51 SRR616268_2.fastq.bz2
① iu@bielinux[srp017156] chmod 755 JSLAB4_1_Linux.sh [ 3:01午後 ]
iu@bielinux[srp017156] ls -lh [ 3:01午後 ]
total 157M
-rwxr-xr-x 1 iu iu 360 5月 1 2015 JSLAB4_1_Linux.sh
-rw-rw-r-- 1 iu iu 83M 12月 3 14:56 SRR616268_1.fastq.bz2
-rw-rw-r-- 1 iu iu 74M 12月 3 15:51 SRR616268_2.fastq.bz2
② iu@bielinux[srp017156] more JSLAB4_1_Linux.sh [ 3:01午後 ]
#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
③ iu@bielinux[srp017156] wc JSLAB4_1_Linux.sh [ 3:01午後 ]
 4 24 360 JSLAB4_1_Linux.sh
iu@bielinux[srp017156] [ 3:01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。①lsで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赤枠の出力ファイルを確認。

```
File Edit View Search Terminal Help
iu@bielinux[srp017156] ls -lh
total 157M
-rwxr-xr-x 1 iu iu 360 5月 1 2015 JSLAB4_1_Linux.sh
-rw-rw-r-- 1 iu iu 83M 12月 3 14:56 SRR616268_1.fastq.bz2
-rw-rw-r-- 1 iu iu 74M 12月 3 15:51 SRR616268_2.fastq.bz2
iu@bielinux[srp017156] more JSLAB4_1_Linux.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] time ./JSLAB4_1_Linux.sh
./JSLAB4_1_Linux.sh 29.39s user 1.55s system 64% cpu 48.050 total
iu@bielinux[srp017156] ls -lh
total 740M
-rwxr-xr-x 1 iu iu 360 5月 1 2015 JSLAB4_1_Linux.sh
-rw-rw-r-- 1 iu iu 83M 12月 3 14:56 SRR616268_1.fastq.bz2
-rw-rw-r-- 1 iu iu 74M 12月 3 15:51 SRR616268_2.fastq.bz2
-rw-rw-r-- 1 iu iu 306M 12月 9 15:08 SRR616268sub_1.fastq
-rw-rw-r-- 1 iu iu 278M 12月 9 15:08 SRR616268sub_2.fastq
iu@bielinux[srp017156]
```


W6-3: 実行を中断したい

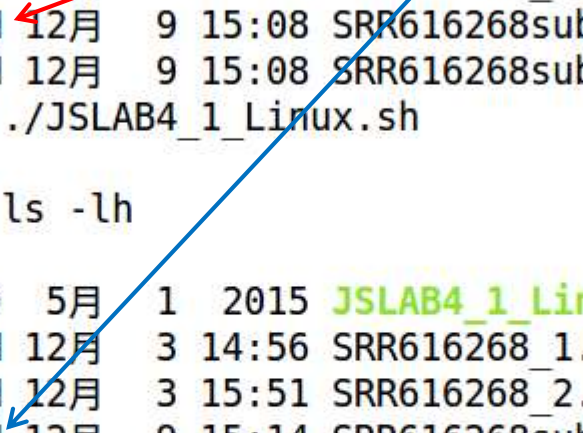
「CTRL + c」(CTRLキーを押しながらcキーを押す)。lessコマンドなどは、qキーを押すと復帰できる(コマンド入力待ち状態になる)。①のシェルスクリプトを実行し、開始5秒程度で「CTRL + c」を押したところ。

```
iu@bielinux[srp017156] ls -lh
total 740M
-rwxr-xr-x 1 iu iu 360 5月 1 2015 JSLAB4_1_Lin
-rw-rw-r-- 1 iu iu 83M 12月 3 14:56 SRR616268_1.fastq.bz2
-rw-rw-r-- 1 iu iu 74M 12月 3 15:51 SRR616268_2.fastq.bz2
-rw-rw-r-- 1 iu iu 306M 12月 9 15:08 SRR616268sub_1.fastq
-rw-rw-r-- 1 iu iu 278M 12月 9 15:08 SRR616268sub_2.fastq
① iu@bielinux[srp017156] ./JSLAB4_1_Linux.sh [ 3:14午後]
^C
iu@bielinux[srp017156] [ 3:14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106MBが変だとわかる。

```
iu@bielinux[srp017156] ls -lh
total 740M
-rwxr-xr-x 1 iu iu 360 5月 1 2015 JSLAB4_1_Lin
-rw-rw-r-- 1 iu iu 83M 12月 3 14:56 SRR616268_1.
-rw-rw-r-- 1 iu iu 74M 12月 3 15:51 SRR616268_2.
-rw-rw-r-- 1 iu iu 306M 12月 9 15:08 SRR616268sub
-rw-rw-r-- 1 iu iu 278M 12月 9 15:08 SRR616268sub
iu@bielinux[srp017156] ./JSLAB4_1_Linux.sh [ 3:14午後 ]
^C [ 3:14午後 ]
iu@bielinux[srp017156] ls -lh
total 540M
-rwxr-xr-x 1 iu iu 360 5月 1 2015 JSLAB4_1_Linux.sh
-rw-rw-r-- 1 iu iu 83M 12月 3 14:56 SRR616268_1.fastq.bz2
-rw-rw-r-- 1 iu iu 74M 12月 3 15:51 SRR616268_2.fastq.bz2
-rw-rw-r-- 1 iu iu 106M 12月 9 15:14 SRR616268sub_1.fastq
-rw-rw-r-- 1 iu iu 278M 12月 9 15:08 SRR616268sub_2.fastq
iu@bielinux[srp017156] [ 3:16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…。

```
iu@bielinux[srp017156] ls -lh
total 740M
-rwxr-xr-x 1 iu iu 360 5月 1 2015 JSLAB4_1_Lin
-rw-rw-r-- 1 iu iu 83M 12月 3 14:56 SRR616268_1.
-rw-rw-r-- 1 iu iu 74M 12月 3 15:51 SRR616268_2.fastq.bz2
-rw-rw-r-- 1 iu iu 306M 12月 9 15:08 SRR616268sub_1.fastq
-rw-rw-r-- 1 iu iu 278M 12月 9 15:08 SRR616268sub_2.fastq
iu@bielinux[srp017156] ./JSLAB4_1_Linux.sh
^C
iu@bielinux[srp017156] ls -lh
total 540M
-rwxr-xr-x 1 iu iu 360 5月 1 2015 JSLAB4_1_Linux.sh
-rw-rw-r-- 1 iu iu 83M 12月 3 14:56 SRR616268_1.fastq.bz2
-rw-rw-r-- 1 iu iu 74M 12月 3 15:51 SRR616268_2.fastq.bz2
-rw-rw-r-- 1 iu iu 106M 12月 9 15:14 SRR616268sub_1.fastq
-rw-rw-r-- 1 iu iu 278M 12月 9 15:08 SRR616268sub_2.fastq
iu@bielinux[srp017156] █
```

[3:14午後]

[3:14午後]

[3:16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、一文字表記の塩基配列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 540M
-rwxr-xr-x 1 iu iu 360 5月 1 2015 JSLAB4_1_Linux.sh
-rw-rw-r-- 1 iu iu 83M 12月 3 14:56 SRR616268_1.fastq.bz2
-rw-rw-r-- 1 iu iu 74M 12月 3 15:51 SRR616268_2.fastq.bz2
-rw-rw-r-- 1 iu iu 106M 12月 9 15:14 SRR616268sub_1.fastq
-rw-rw-r-- 1 iu iu 278M 12月 9 15:08 SRR616268sub_2.fastq
iu@bielinux[srp017156] tail -n 8 SRR616268sub_1.fastq
+SRR616268.346719 2291:6:1101:15095:34145 length=107
_bbecccccgcggifhhegghhiihf[fgghidgfafgiihgfhfgeffhfh__df`gYS^acecc\acbccc^
\WW^[bb_bbcbccbdccbbcca_bbeeee
@SRR616268.346720 2291:6:1101:15238:34177 length=107
CTGGGCTGTTTCATGCTGGTGTCCAATCTAAACCGCTGGATAAGTTCATGTTTCATGGCATGCAGTTTATGCTAC
GCCTGAACACGAAGGACGGGGAGCATTCAAG
+SRR616268.346720 2291:6:1101:15238:34177 length=107
bbbeeeeegggggiiiiiiighiiiiiiiiiiiiiihiiiiiiiiifhiiiiifiiiiiiiiiihighiiiggfgd
_cedcddbcccc`accccV_aBbbbeeee
@SRR616268.346721 2291:6:1101:15049:34182 length=107
GCCGCCGCCAGCTATGTATTCACTGACAAG%
iu@bielinux[srp017156] rm -f SRR616268sub_1.fastq
iu@bielinux[srp017156]
```



[3:19午後]

[3:19午後]

[3:19午後]

[3:19午後]

W6-5: サブセット抽出を再実行

```
iu@bielinux[srp017156] pwd [ 3:23午後 ]
/home/iu/Documents/srp017156
iu@bielinux[srp017156] ls -lh [ 3:23午後 ]
total 434M
-rwxr-xr-x 1 iu iu 360 5月 1 2015 JSLAB4_1_Linux.sh
-rw-rw-r-- 1 iu iu 83M 12月 3 14:56 SRR616268_1.fastq.bz2
-rw-rw-r-- 1 iu iu 74M 12月 3 15:51 SRR616268_2.fastq.bz2
-rw-rw-r-- 1 iu iu 278M 12月 9 15:08 SRR616268sub_2.fastq
iu@bielinux[srp017156] ./JSLAB4_1_Linux.sh [ 3:23午後 ]
iu@bielinux[srp017156] ls -lh [ 3:24午後 ]
total 740M
-rwxr-xr-x 1 iu iu 360 5月 1 2015 JSLAB4_1_Linux.sh
-rw-rw-r-- 1 iu iu 83M 12月 3 14:56 SRR616268_1.fastq.bz2
-rw-rw-r-- 1 iu iu 74M 12月 3 15:51 SRR616268_2.fastq.bz2
-rw-rw-r-- 1 iu iu 306M 12月 9 15:24 SRR616268sub_1.fastq
-rw-rw-r-- 1 iu iu 278M 12月 9 15:24 SRR616268sub_2.fastq
iu@bielinux[srp017156] [ 3:24午後 ]
```



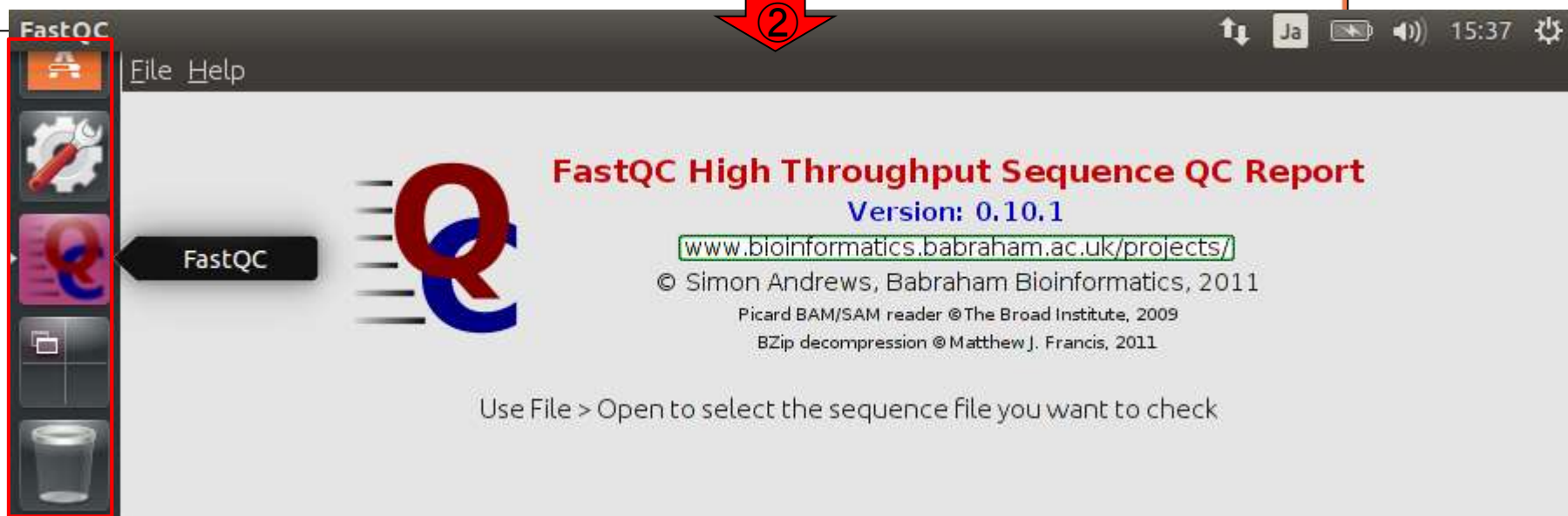
W7-1 : FastQC

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、間違っても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。(もしそうならなかったら赤枠内で起動中のFastQCをクリックすれば、右下のようになる)この状態から脱出するには…

```
File Edit View Search Terminal Help
iu@bielinux[srp017156] pwd
/home/iu/Documents/srp017156
iu@bielinux[srp017156] ls -lh
total 740M
-rwxr-xr-x 1 iu iu 360 5月 1 2015 JSLAB4_1_Linux.sh
-rw-rw-r-- 1 iu iu 83M 12月 3 14:56 SRR616268_1.fastq.bz2
-rw-rw-r-- 1 iu iu 74M 12月 3 15:51 SRR616268_2.fastq.bz2
-rw-rw-r-- 1 iu iu 306M 12月 9 15:24 SRR616268sub_1.fastq
-rw-rw-r-- 1 iu iu 278M 12月 9 15:24 SRR616268sub_2.fastq
iu@bielinux[srp017156] fastqc
```

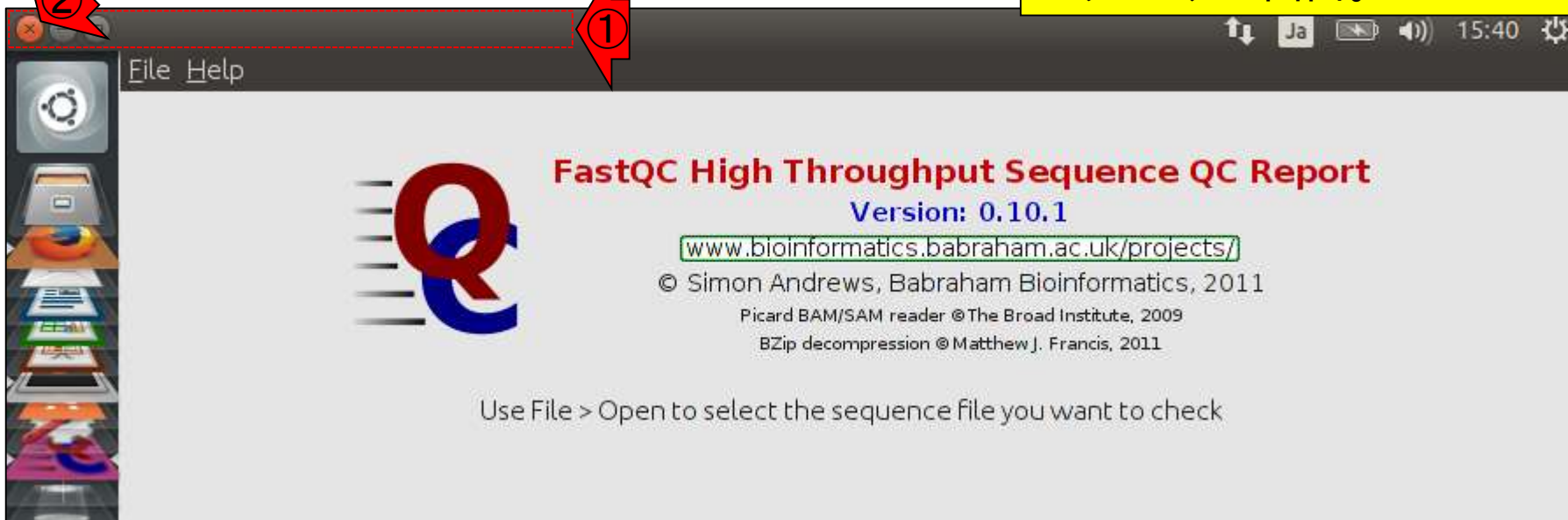
[3:30午後]

[3:30午後]



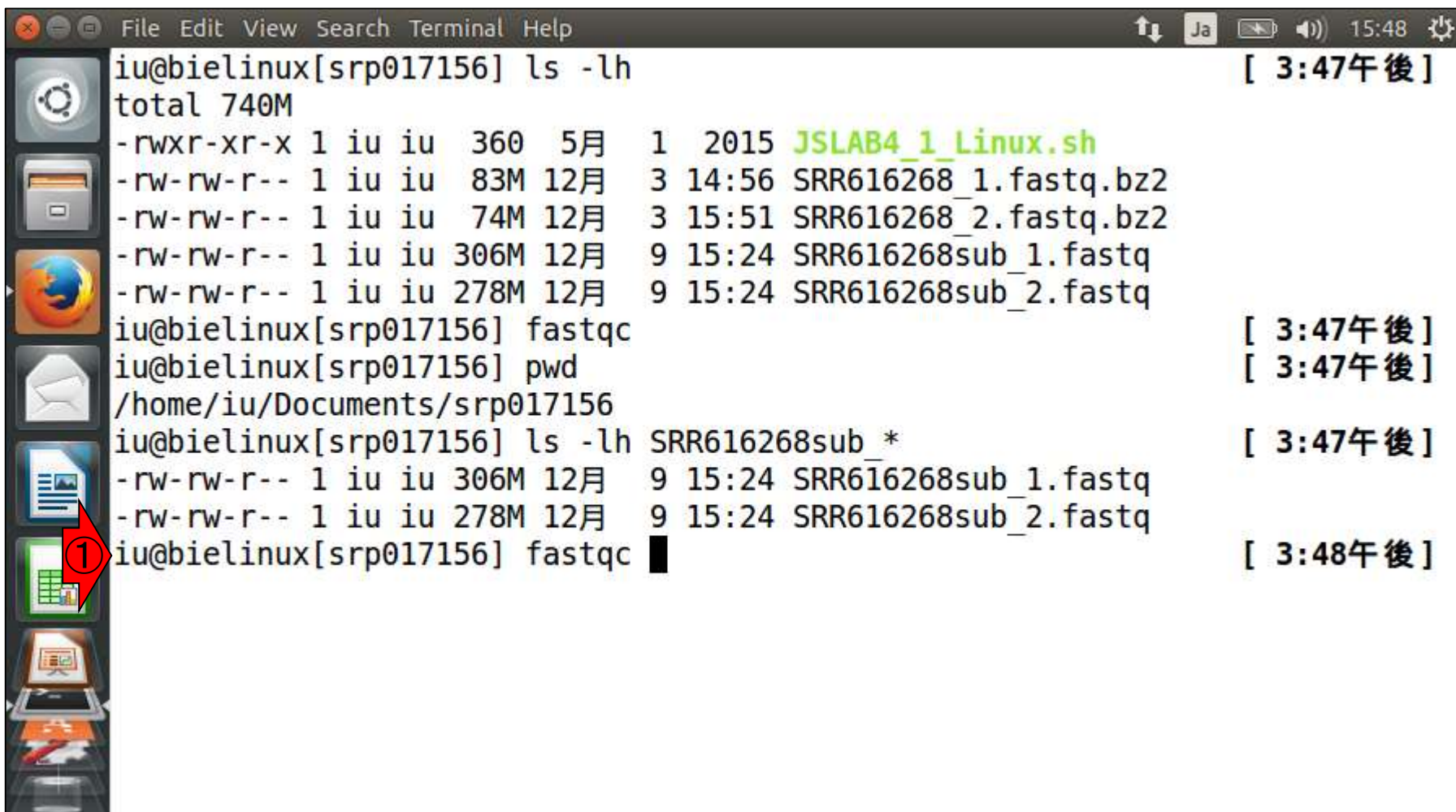
W7-1: FastQC

この状態から脱出するには、①赤枠付近にカーソルを移動させ、②×。右下は復帰後のターミナル画面。



```
File Edit View Search Terminal Help
iu@bielinux[srp017156] pwd [ 3:30午後 ]
/home/iu/Documents/srp017156
iu@bielinux[srp017156] ls -lh [ 3:30午後 ]
total 740M
-rwxr-xr-x 1 iu iu 360 5月 1 2015 JSLAB4_1_Linux.sh
-rw-rw-r-- 1 iu iu 83M 12月 3 14:56 SRR616268_1.fastq.bz2
-rw-rw-r-- 1 iu iu 74M 12月 3 15:51 SRR616268_2.fastq.bz2
-rw-rw-r-- 1 iu iu 306M 12月 9 15:24 SRR616268sub_1.fastq
-rw-rw-r-- 1 iu iu 278M 12月 9 15:24 SRR616268sub_2.fastq
iu@bielinux[srp017156] fastqc [ 3:30午後 ]
iu@bielinux[srp017156] [ 3:42午後 ]
```

W7-2: FastQC

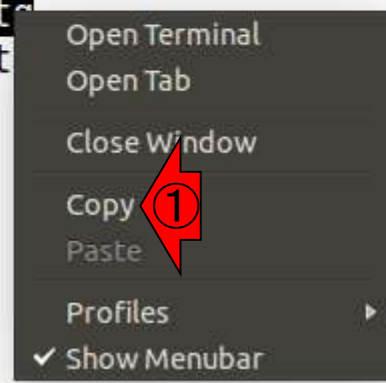
A terminal window showing a series of commands and their outputs. The window title is "File Edit View Search Terminal Help". The user is "iu" on a host named "bielinux" in a shell named "srp017156". The terminal shows the execution of "ls -lh", "fastqc", "pwd", "ls -lh SRR616268sub_*", and "fastqc" again. A red arrow with the number "1" points to the second "fastqc" command. The system tray at the top right shows the time as 15:48 and the language as "Ja".

```
iu@bielinux[srp017156] ls -lh [ 3:47午後 ]
total 740M
-rwxr-xr-x 1 iu iu 360 5月 1 2015 JSLAB4_1_Linux.sh
-rw-rw-r-- 1 iu iu 83M 12月 3 14:56 SRR616268_1.fastq.bz2
-rw-rw-r-- 1 iu iu 74M 12月 3 15:51 SRR616268_2.fastq.bz2
-rw-rw-r-- 1 iu iu 306M 12月 9 15:24 SRR616268sub_1.fastq
-rw-rw-r-- 1 iu iu 278M 12月 9 15:24 SRR616268sub_2.fastq
iu@bielinux[srp017156] fastqc [ 3:47午後 ]
iu@bielinux[srp017156] pwd [ 3:47午後 ]
/home/iu/Documents/srp017156
iu@bielinux[srp017156] ls -lh SRR616268sub_* [ 3:47午後 ]
-rw-rw-r-- 1 iu iu 306M 12月 9 15:24 SRR616268sub_1.fastq
-rw-rw-r-- 1 iu iu 278M 12月 9 15:24 SRR616268sub_2.fastq
iu@bielinux[srp017156] fastqc [ 3:48午後 ]
```

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1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 740M
-rwxr-xr-x 1 iu iu 360 5月 1 2015 JSLAB4_1
-rw-rw-r-- 1 iu iu 83M 12月 3 14:56 SRR616268_1.fastq.bz2
-rw-rw-r-- 1 iu iu 74M 12月 3 15:51 SRR616268_2.fastq.bz2
-rw-rw-r-- 1 iu iu 306M 12月 9 15:24 SRR616268sub_1.fastq
-rw-rw-r-- 1 iu iu 278M 12月 9 15:24 SRR616268sub_2.fastq
iu@bielinux[srp017156] fastqc [ 3:47午後]
iu@bielinux[srp017156] pwd [ 3:47午後]
/home/iu/Documents/srp017156
iu@bielinux[srp017156] ls -lh SRR616268sub * [ 3:47午後]
-rw-rw-r-- 1 iu iu 306M 12月 9 15:24 SRR616268sub_1.fastq
-rw-rw-r-- 1 iu iu 278M 12月 9 15:24 SRR616268sub_2.fastq
iu@bielinux[srp017156] fastqc
```



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1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 740M
-rwxr-xr-x 1 iu iu 360 5月 1 2015 JSLAB4_1
-rw-rw-r-- 1 iu iu 83M 12月 3 14:56 SRR616268_1.fastq.bz2
-rw-rw-r-- 1 iu iu 74M 12月 3 15:51 SRR616268_2.fastq.bz2
-rw-rw-r-- 1 iu iu 306M 12月 9 15:24 SRR616268sub_1.fastq
-rw-rw-r-- 1 iu iu 278M 12月 9 15:24 SRR616268sub_2.fastq
iu@bielinux[srp017156] fastqc [ 3:47午後]
iu@bielinux[srp017156] pwd [ 3:47午後]
/home/iu/Documents/srp017156
iu@bielinux[srp017156] ls -lh SRR616268sub_* [ 3:47午後]
-rw-rw-r-- 1 iu iu 306M 12月 9 15:24 SRR616268
-rw-rw-r-- 1 iu iu 278M 12月 9 15:24 SRR616268
iu@bielinux[srp017156] fastqc [ 3:48午後]
```

- Open Terminal
- Open Tab
- Close Window
- Copy
- Paste
- Profiles
- ✓ Show Menubar

W7-2: FastQC

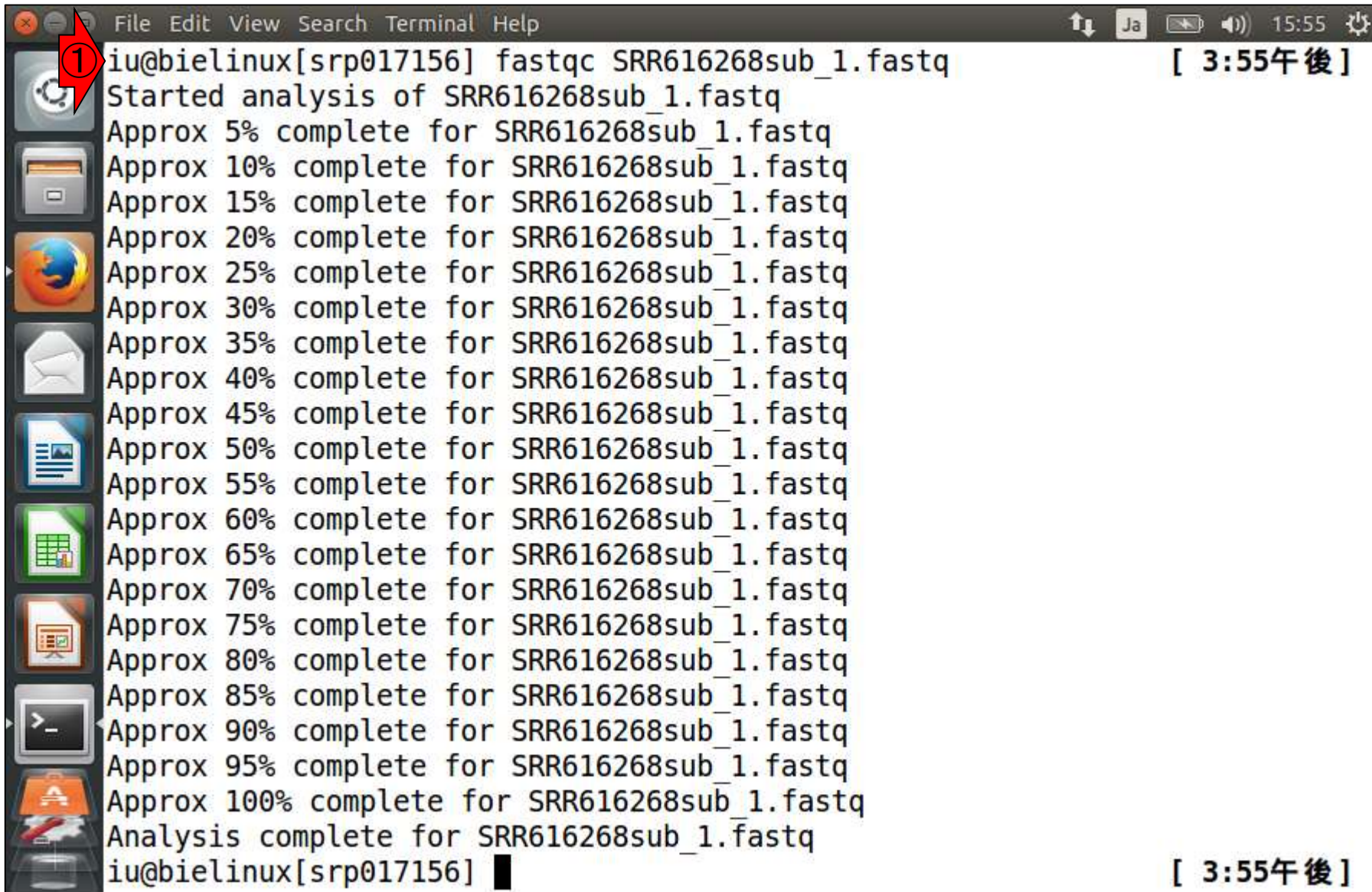
```

iu@bielinux[srp017156] ls -lh
total 740M
-rwxr-xr-x 1 iu iu 360 5月 1 2015 JSLAB4_1_Linux.sh
-rw-rw-r-- 1 iu iu 83M 12月 3 14:56 SRR616268_1.fastq.bz2
-rw-rw-r-- 1 iu iu 74M 12月 3 15:51 SRR616268_2.fastq.bz2
-rw-rw-r-- 1 iu iu 306M 12月 9 15:24 SRR616268sub_1.fastq
-rw-rw-r-- 1 iu iu 278M 12月 9 15:24 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 12月 9 15:24 SRR616268sub_1.fastq
-rw-rw-r-- 1 iu iu 278M 12月 9 15:24 SRR616268sub_2.fastq
iu@bielinux[srp017156] fastqc SRR616268sub_1.fastq

```



W7-3: FastQC



```
iu@bielinux[srp017156] fastqc SRR616268sub_1.fastq [ 3:55午後 ]
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] [ 3:55午後 ]
```


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、③ディレクトリの中身もリストアップしている

```
iu@bielinux[srp017156] ls -lh
total 740M
-rwxr-xr-x 1 iu iu 360 5月 1 2015 JS
-rw-rw-r-- 1 iu iu 83M 12月 3 14:56 SRR616268_1.fastq.bz2
-rw-rw-r-- 1 iu iu 74M 12月 3 15:51 SRR616268_2.fastq.bz2
-rw-rw-r-- 1 iu iu 306M 12月 9 15:24 SRR616268sub_1.fastq
drwxrwxr-x 4 iu iu 4.0K 12月 9 15:55 SRR616268sub_1_fastqc
-rw-rw-r-- 1 iu iu 293K 12月 9 15:55 SRR616268sub_1_fastqc.zip
-rw-rw-r-- 1 iu iu 278M 12月 9 15:24 SRR616268sub_2.fastq

iu@bielinux[srp017156] ls -lh SRR616268sub_*
-rw-rw-r-- 1 iu iu 306M 12月 9 15:24 SRR616268sub_1.fastq
-rw-rw-r-- 1 iu iu 293K 12月 9 15:55 SRR616268sub_1_fastqc.zip
-rw-rw-r-- 1 iu iu 278M 12月 9 15:24 SRR616268sub_2.fastq

SRR616268sub_1_fastqc:
total 88K
-rw-rw-r-- 1 iu iu 26K 12月 9 15:55 fastqc_data.txt
-rw-rw-r-- 1 iu iu 48K 12月 9 15:55 fastqc_report.html
drwxrwxr-x 2 iu iu 4.0K 12月 9 15:55 Icons
drwxrwxr-x 2 iu iu 4.0K 12月 9 15:55 Images
-rw-rw-r-- 1 iu iu 542 12月 9 15:55 summary.txt
```

[4:00午後]

[4:00午後]

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。①「fastqc -h」と打ってリターン

```
iu@bielinux[srp017156] pwd [ 4:05午後 ]
/home/iu/Documents/srp017156
iu@bielinux[srp017156] ls -lh [ 4:05午後 ]
total 740M
-rwxr-xr-x 1 iu iu 360 5月 1 2015 JSLAB4_1_Linux.sh
-rw-rw-r-- 1 iu iu 83M 12月 3 14:56 SRR616268_1.fastq.bz2
-rw-rw-r-- 1 iu iu 74M 12月 3 15:51 SRR616268_2.fastq.bz2
-rw-rw-r-- 1 iu iu 306M 12月 9 15:24 SRR616268sub_1.fastq
drwxrwxr-x 4 iu iu 4.0K 12月 9 15:55 SRR616268sub_1_fastqc
-rw-rw-r-- 1 iu iu 293K 12月 9 15:55 SRR616268sub_1_fastqc.zip
-rw-rw-r-- 1 iu iu 278M 12月 9 15:24 SRR616268sub_2.fastq
iu@bielinux[srp017156] fastqc -h [ 4:05午後 ]
```



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①lessを用いる例を示す。

```
File Edit View Search Terminal Help 16:07
--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. Default length is 5 if not specified.
-q --quiet Suppress all progress messages on stdout and only report errors.
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 [ 4:06午後 ]
```



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。下矢印キーをず〜っと押して一番下まで行ってみよう

```
File Edit View Search Terminal Help

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 on the command line then the program will run with no user interaction required. In this mode it is suitable for inclusion into a standardised analysis pipeline.

:█
```

W7-5: FastQC

①FastQC実行時に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) になったら、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。

```
File Edit View Search Terminal Help

-c Specifies a non-default file which contains the list of
--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 conte
nt module. Specified Kmer length must be between 2 and 10. De
fault length is 5 if not specified.

-q --quiet Supress all progress messages on stdout and only report er
rors.

BUGS

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

(END)
```



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
t.
The file must contain set
form name[tab]sequence.

be ignored.

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

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

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] [ 4:06午後]
[ 4:14午後]
```


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] ls -lh
total 740M
-rwxr-xr-x 1 iu iu 360 5月 1 2015 JSLAB4_1
-rw-rw-r-- 1 iu iu 83M 12月 3 14:56 SRR616268_1.fastq.bz2
-rw-rw-r-- 1 iu iu 74M 12月 3 15:51 SRR616268_2.fastq.bz2
-rw-rw-r-- 1 iu iu 306M 12月 9 15:24 SRR616268sub_1.fastq
drwxrwxr-x 4 iu iu 4.0K 12月 9 15:55 SRR616268sub_1_fastqc
-rw-rw-r-- 1 iu iu 293K 12月 9 15:55 SRR616268sub_1_fastqc.zip
-rw-rw-r-- 1 iu iu 278M 12月 9 15:24 SRR616268sub_2.fastq
iu@bielinux[srp017156] fastqc -q SRR616268sub_2.fastq [ 4:17午後]
iu@bielinux[srp017156] ls -lh [ 4:17午後]
total 740M
-rwxr-xr-x 1 iu iu 360 5月 1 2015 JSLAB4_1_Linux.sh
-rw-rw-r-- 1 iu iu 83M 12月 3 14:56 SRR616268_1.fastq.bz2
-rw-rw-r-- 1 iu iu 74M 12月 3 15:51 SRR616268_2.fastq.bz2
-rw-rw-r-- 1 iu iu 306M 12月 9 15:24 SRR616268sub_1.fastq
drwxrwxr-x 4 iu iu 4.0K 12月 9 15:55 SRR616268sub_1_fastqc
-rw-rw-r-- 1 iu iu 293K 12月 9 15:55 SRR616268sub_1_fastqc.zip
-rw-rw-r-- 1 iu iu 278M 12月 9 15:24 SRR616268sub_2.fastq
drwxrwxr-x 4 iu iu 4.0K 12月 9 16:17 SRR616268sub_2_fastqc
-rw-rw-r-- 1 iu iu 243K 12月 9 16:17 SRR616268sub_2_fastqc.zip
iu@bielinux[srp017156] [ 4:17午後]
```



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*
-rw-rw-r-- 1 iu iu 293K 12月 9 15:55 SRR616268sub_1_fastqc.zip
-rw-rw-r-- 1 iu iu 243K 12月 9 16:17 SRR616268sub_2_fastqc.zip

SRR616268sub_1_fastqc:
total 88K
-rw-rw-r-- 1 iu iu 26K 12月 9 15:55 fastqc_data.txt
-rw-rw-r-- 1 iu iu 48K 12月 9 15:55 fastqc_report.html
drwxrwxr-x 2 iu iu 4.0K 12月 9 15:55 Icons
drwxrwxr-x 2 iu iu 4.0K 12月 9 15:55 Images
-rw-rw-r-- 1 iu iu 542 12月 9 15:55 summary.txt

SRR616268sub_2_fastqc:
total 56K
-rw-rw-r-- 1 iu iu 16K 12月 9 16:17 fastqc_data.txt
-rw-rw-r-- 1 iu iu 26K 12月 9 16:17 fastqc_report.html
drwxrwxr-x 2 iu iu 4.0K 12月 9 16:17 Icons
drwxrwxr-x 2 iu iu 4.0K 12月 9 16:17 Images
-rw-rw-r-- 1 iu iu 542 12月 9 16:17 summary.txt

iu@bielinux[srp017156] ls -lhd *fastqc*
drwxrwxr-x 4 iu iu 4.0K 12月 9 15:55 SRR616268sub_1_fastqc
-rw-rw-r-- 1 iu iu 293K 12月 9 15:55 SRR616268sub_1_fastqc.zip
drwxrwxr-x 4 iu iu 4.0K 12月 9 16:17 SRR616268sub_2_fastqc
-rw-rw-r-- 1 iu iu 243K 12月 9 16:17 SRR616268sub_2_fastqc.zip

iu@bielinux[srp017156]
```



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* [ 4:25午後 ]
drwxrwxr-x 4 iu iu 4.0K 12月 9 15:55 SRR616268sub_1_fastqc
-rw-rw-r-- 1 iu iu 293K 12月 9 15:55 SRR616268sub_1_fastqc.zip
drwxrwxr-x 4 iu iu 4.0K 12月 9 16:17 SRR616268sub_2_fastqc
-rw-rw-r-- 1 iu iu 243K 12月 9 16:17 SRR616268sub_2_fastqc.zip
① iu@bielinux[srp017156] rm -f *fastqc* [ 4:25午後 ]
rm: cannot remove 'SRR616268sub_1_fastqc': Is a directory
rm: cannot remove 'SRR616268sub_2_fastqc': Is a directory
iu@bielinux[srp017156] ls -lhd *fastqc* [ 4:25午後 ]
drwxrwxr-x 4 iu iu 4.0K 12月 9 15:55 SRR616268sub_1_fastqc
drwxrwxr-x 4 iu iu 4.0K 12月 9 16:17 SRR616268sub_2_fastqc
② iu@bielinux[srp017156] rm -rf *fastqc* [ 4:25午後 ]
iu@bielinux[srp017156] ls -lhd *fastqc* [ 4:25午後 ]
zsh: no matches found: *fastqc*
iu@bielinux[srp017156] ls -lh [ 4:25午後 ]
total 740M
-rwxr-xr-x 1 iu iu 360 5月 1 2015 JSLAB4_1_Linux.sh
-rw-rw-r-- 1 iu iu 83M 12月 3 14:56 SRR616268_1.fastq.bz2
-rw-rw-r-- 1 iu iu 74M 12月 3 15:51 SRR616268_2.fastq.bz2
-rw-rw-r-- 1 iu iu 306M 12月 9 15:24 SRR616268sub_1.fastq
-rw-rw-r-- 1 iu iu 278M 12月 9 15:24 SRR616268sub_2.fastq
iu@bielinux[srp017156] [ 4:25午後 ]
```


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 740M
-rwxr-xr-x 1 iu iu 360 5月 1 2015 JSLAB4_1_Linux.sh
-rw-rw-r-- 1 iu iu 83M 12月 3 14:56 SRR616268_1.fastq
-rw-rw-r-- 1 iu iu 74M 12月 3 15:51 SRR616268_2.fastq.bz2
-rw-rw-r-- 1 iu iu 306M 12月 9 15:24 SRR616268sub_1.fastq
-rw-rw-r-- 1 iu iu 278M 12月 9 15:24 SRR616268sub_2.fastq
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_* [ 4:29午後 ]
-rwxr-xr-x 1 iu iu 360 5月 1 2015 JSLAB4_1_Linux.sh
-rw-rw-r-- 1 iu iu 424 5月 5 2015 JSLAB4_2_Linux.sh
② iu@bielinux[srp017156] chmod 755 JSLAB4_2_Linux.sh [ 4:29午後 ]
iu@bielinux[srp017156] ls -lh JSLAB4_* [ 4:29午後 ]
-rwxr-xr-x 1 iu iu 360 5月 1 2015 JSLAB4_1_Linux.sh
-rwxr-xr-x 1 iu iu 424 5月 5 2015 JSLAB4_2_Linux.sh
iu@bielinux[srp017156] █ [ 4:29午後 ]
```

W7-8: シェルスクリプト

①wgetで取得したJSLAB4_2_Linux.shの中身をmoreで表示。②コメントアウトされていない赤枠部分のみが実行される。③shで実行。④生成されるのは水色枠の4つ。

```
iu@bielinux[srp017156] more JSLAB4_2_Linux.sh [ 4:35午後 ]
#wget -c ftp://ftp.ddbj.nig.ac.jp/ddbj_database/dra/fastq/SRA061/SRA061483/SRX
204226/SRR616268_1.fastq.bz2
#wget -c ftp://ftp.ddbj.nig.ac.jp/ddbj_database/dra/fastq/SRA061/SRA061483/SRX
204226/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 [ 4:35午後 ]
iu@bielinux[srp017156] ls -lhd SRR616268sub_* [ 4:36午後 ]
-rw-rw-r-- 1 iu iu 306M 12月 9 15:24 SRR616268sub_1.fastq
drwxrwxr-x 4 iu iu 4.0K 12月 9 16:36 SRR616268sub_1_fastqc
-rw-rw-r-- 1 iu iu 293K 12月 9 16:36 SRR616268sub_1_fastqc.zip
-rw-rw-r-- 1 iu iu 278M 12月 9 15:24 SRR616268sub_2.fastq
drwxrwxr-x 4 iu iu 4.0K 12月 9 16:36 SRR616268sub_2_fastqc
-rw-rw-r-- 1 iu iu 243K 12月 9 16:36 SRR616268sub_2_fastqc.zip
iu@bielinux[srp017156]
```

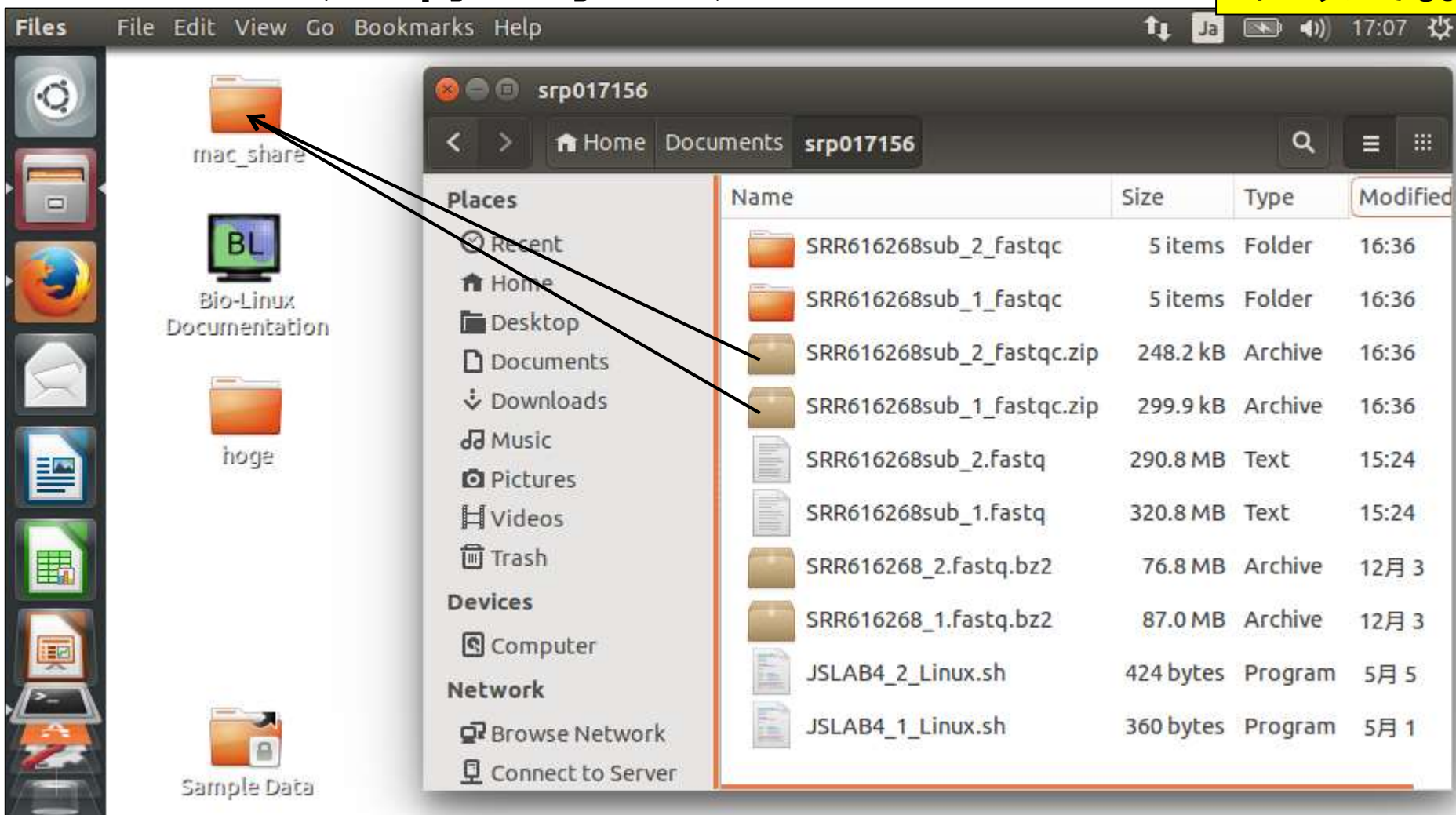

W7-9: 共有フォルダにコピー

①FastQC実行結果のzipファイルを共有フォルダ (~ / Desktop / mac_share) にコピー。
②コピー先フォルダを表示。赤枠のものがコピーされたもの。

```
iu@bielinux[srp017156] ls -lhd SRR616268sub_*
-rw-rw-r-- 1 iu iu 306M 12月 9 15:24 SRR616268sub_1.fastq
drwxrwxr-x 4 iu iu 4.0K 12月 9 16:36 SRR616268sub_1_fastqc
-rw-rw-r-- 1 iu iu 293K 12月 9 16:36 SRR616268sub_1_fastqc.zip
-rw-rw-r-- 1 iu iu 278M 12月 9 15:24 SRR616268sub_2.fastq
drwxrwxr-x 4 iu iu 4.0K 12月 9 16:36 SRR616268sub_2_fastqc
-rw-rw-r-- 1 iu iu 243K 12月 9 16:36 SRR616268sub_2_fastqc.zip
iu@bielinux[srp017156] cp *.zip ~/Desktop/mac_share [ 4:43午後]
iu@bielinux[srp017156] ls -lh ~/Desktop/mac_share [ 4:43午後]
total 3.4M
-rwxrwxrwx 1 iu iu 58K 11月 25 19:36 hoge.png
-rwxrwxrwx 1 iu iu 364 12月 8 17:34 JSLAB4_1.sh
-rwxrwxrwx 1 iu iu 2.8M 3月 10 2014 Lactobacillus_casei_12a.GCA_000309565.1.
22.dna.toplevel.fa
-rwxrwxrwx 1 iu iu 293K 12月 9 2015 SRR616268sub_1_fastqc.zip
-rwxrwxrwx 1 iu iu 243K 12月 9 2015 SRR616268sub_2_fastqc.zip
iu@bielinux[srp017156] [ 4:43午後]
```


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。



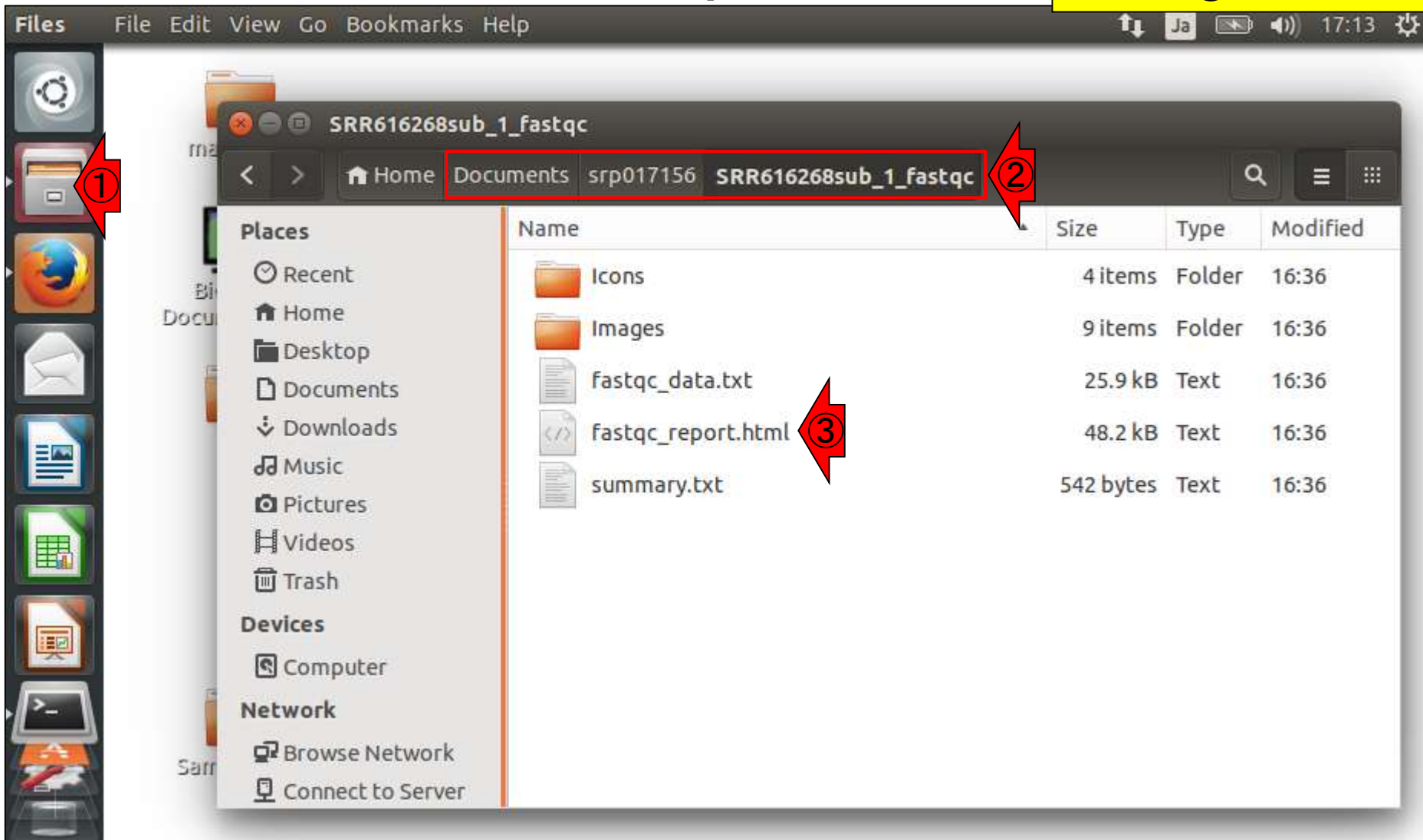
W7-10:バージョン

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 740M
-rwxr-xr-x 1 iu iu 360 5月 1 2015 JSLAB4_2_LINUX.SH
-rwxr-xr-x 1 iu iu 424 5月 5 2015 JSLAB4_2_LINUX.SH
-rw-rw-r-- 1 iu iu 83M 12月 3 14:56 SRR616268_1.fastq.bz2
-rw-rw-r-- 1 iu iu 74M 12月 3 15:51 SRR616268_2.fastq.bz2
-rw-rw-r-- 1 iu iu 306M 12月 9 15:24 SRR616268sub_1.fastq
drwxrwxr-x 4 iu iu 4.0K 12月 9 16:36 SRR616268sub_1_fastqc
-rw-rw-r-- 1 iu iu 293K 12月 9 16:36 SRR616268sub_1_fastqc.zip
-rw-rw-r-- 1 iu iu 278M 12月 9 15:24 SRR616268sub_2.fastq
drwxrwxr-x 4 iu iu 4.0K 12月 9 16:36 SRR616268sub_2_fastqc
-rw-rw-r-- 1 iu iu 243K 12月 9 16:36 SRR616268sub_2_fastqc.zip
① iu@bielinux[srp017156] fastqc -v [ 5:10午後 ]
FastQC v0.10.1
② iu@bielinux[srp017156] fastqc -V [ 5:10午後 ]
FastQC v0.10.1
③ iu@bielinux[srp017156] man fastqc [ 5:10午後 ]
No manual entry for fastqc
See 'man 7 undocumented' for help when manual pages are not available.
④ iu@bielinux[srp017156] fastqc -h [ 5:10午後 ]
```

W8-1: FastQC結果を眺める

ゲストOS上で眺める場合は、② SRR616268sub_1_fastqcディレクトリ中の③htmlファイルを開く。



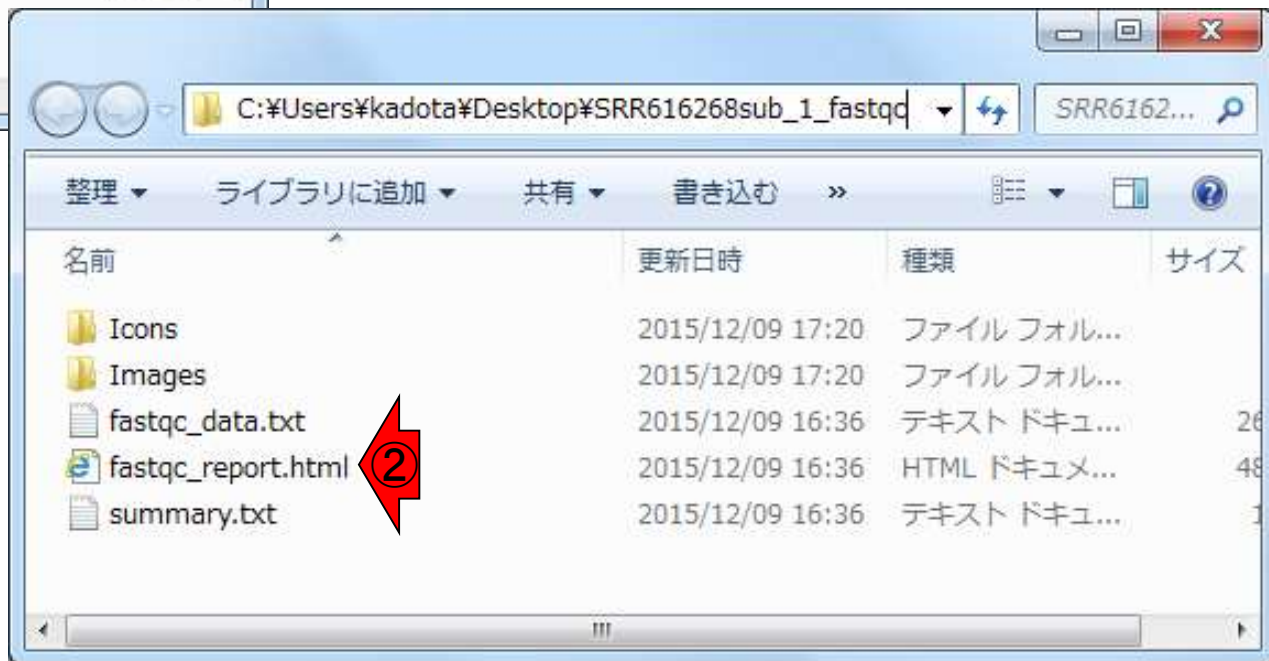
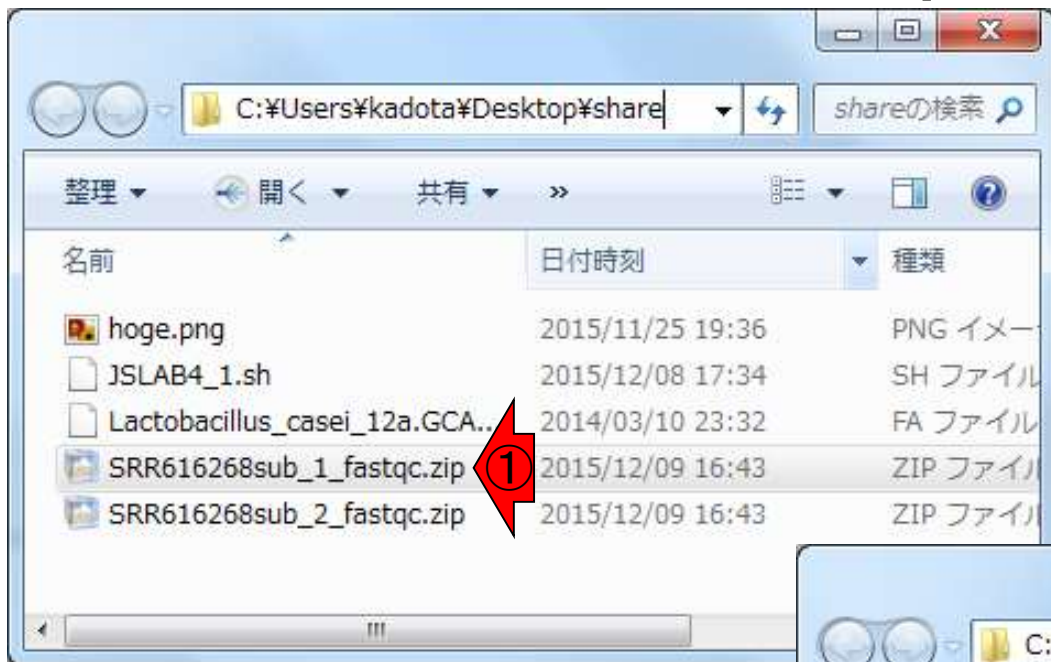
W8-1: FastQC結果を眺める

①SRR616268sub_1_fastqcディレクトリに移動して②firefoxコマンドでhtmlファイルを開くのもよい。

```
iu@bielinux[srp017156] pwd [ 5:16午後 ]
/home/iu/Documents/srp017156
iu@bielinux[srp017156] ls [ 5:16午後 ]
JSLAB4_1_Linux.sh SRR616268sub_1_fastq SRR616268sub_2_fastqc
JSLAB4_2_Linux.sh SRR616268sub_1_fastqc SRR616268sub_2_fastqc.zip
SRR616268_1_fastq.bz2 SRR616268sub_1_fastqc.zip
SRR616268_2_fastq.bz2 SRR616268sub_2_fastq
① iu@bielinux[srp017156] cd SRR616268sub_1_fastqc [ 5:16午後 ]
iu@bielinux[SRR616268sub_1_fastqc] pwd [ 5:16午後 ]
/home/iu/Documents/srp017156/SRR616268sub_1_fastqc
iu@bielinux[SRR616268sub_1_fastqc] ls [ 5:16午後 ]
fastqc_data.txt fastqc_report.html Icons Images summary.txt
② iu@bielinux[SRR616268sub_1_fastqc] firefox fastqc_report.html [ 5:16午後 ]
```

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: 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

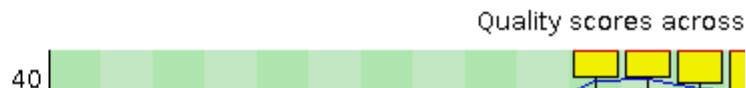
Summary

- ✓ 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: 結果を眺める

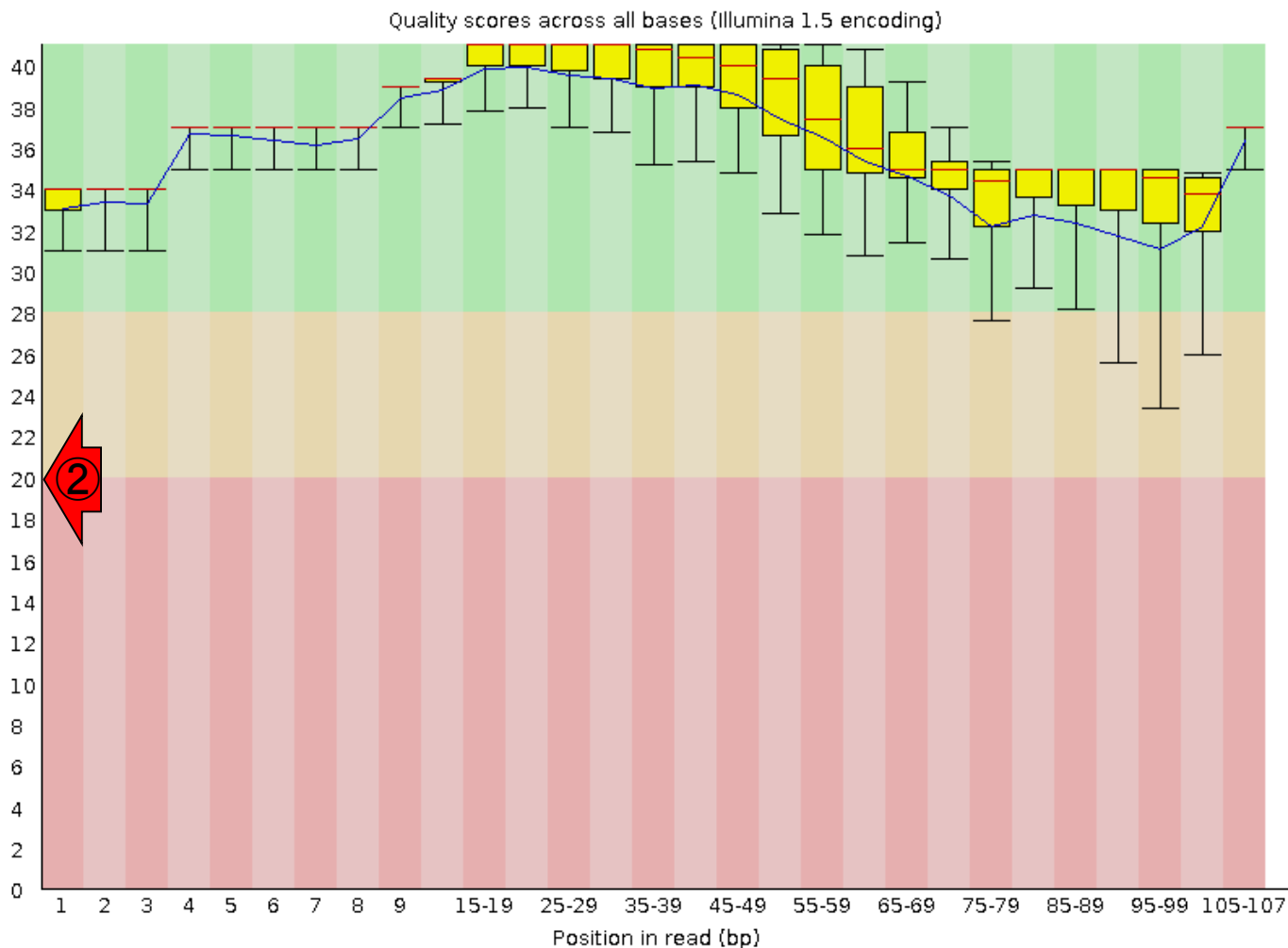
①塩基(ポジション)ごとのクオリティ調査。横軸は塩基のポジション、縦軸はPhredスコア(高いほど塩基配列決定精度が高い)。②スコア20が1つの目安。これは100回中1回のベースコールエラー率($p_{err} = 0.01$)に相当する。

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: 結果を眺める

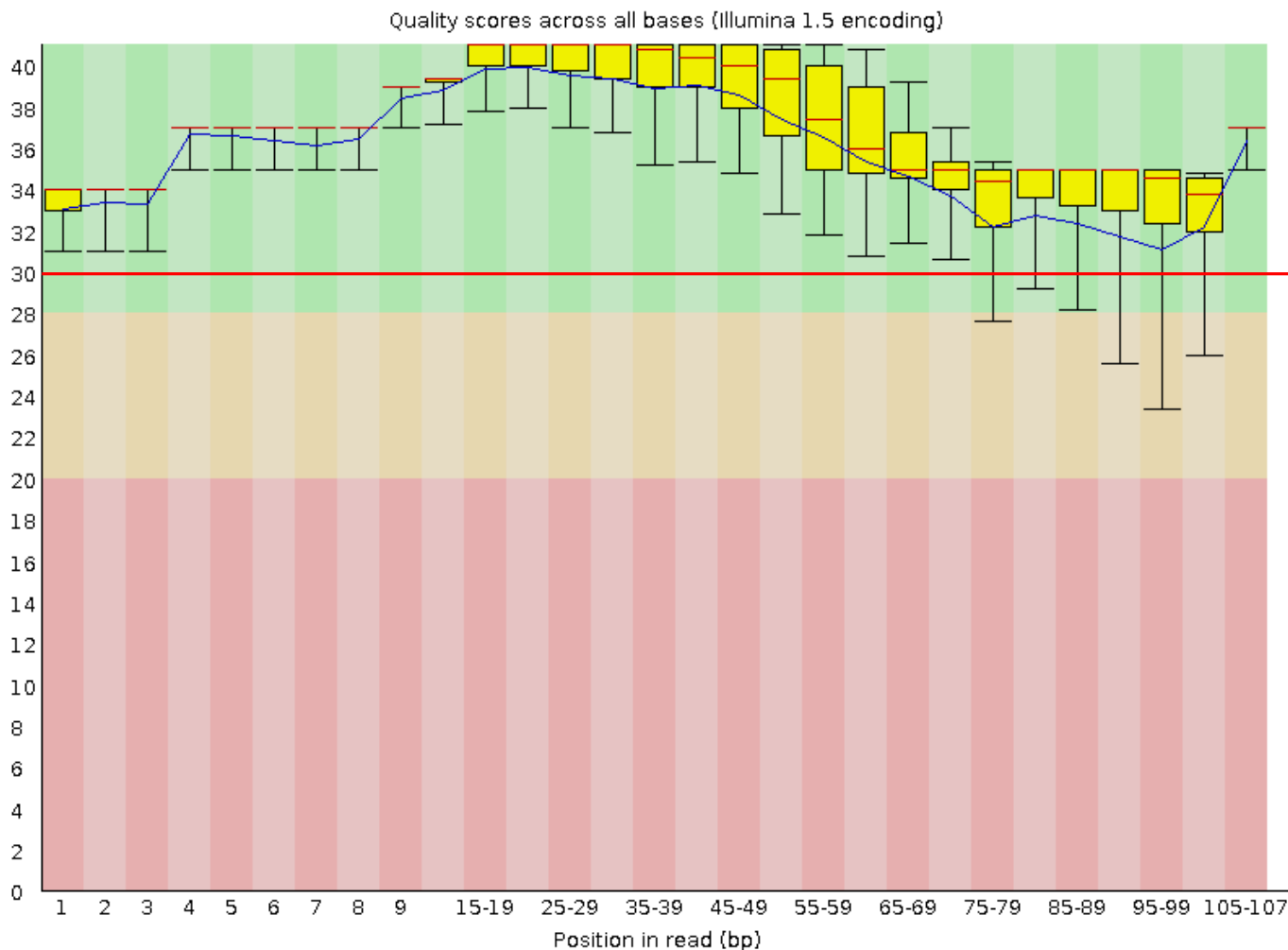
①スコア30の赤の水平線は $p_{err} = 0.001$ に相当。
 $-\log(p_{err}) = 3$ 。これを10倍した値で表現すると、
概ね0-40程度の値の範囲になる。ヒトが認識し
やすく、当時のPCの解像度的な観点から、「 $-\log(p_{err}) \times 10$ 」
をスコアとして採用したのだろう。

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: 結果を眺める

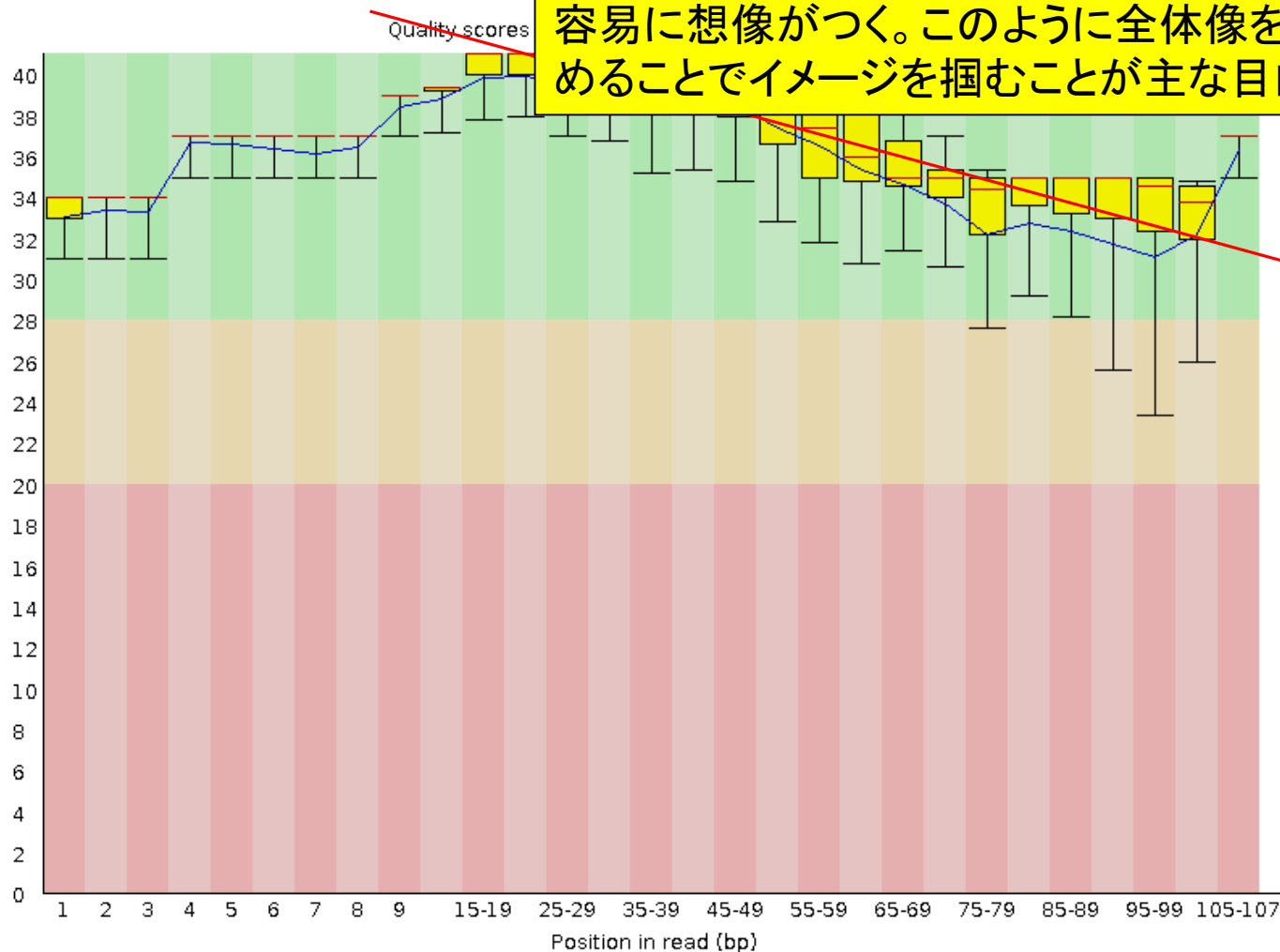
右側ほど塩基配列決定精度が低下傾向にあることがわかる。これが一般的な傾向。このデータは全体的に精度が高いと言える。また、フィルタリング条件として、「スコア20以下の塩基を一定数含むリードの除去」を行っても、ほとんど除去されないということも容易に想像がつく。このように全体像を眺めることでイメージを掴むことが主な目的。

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: 結果を眺める

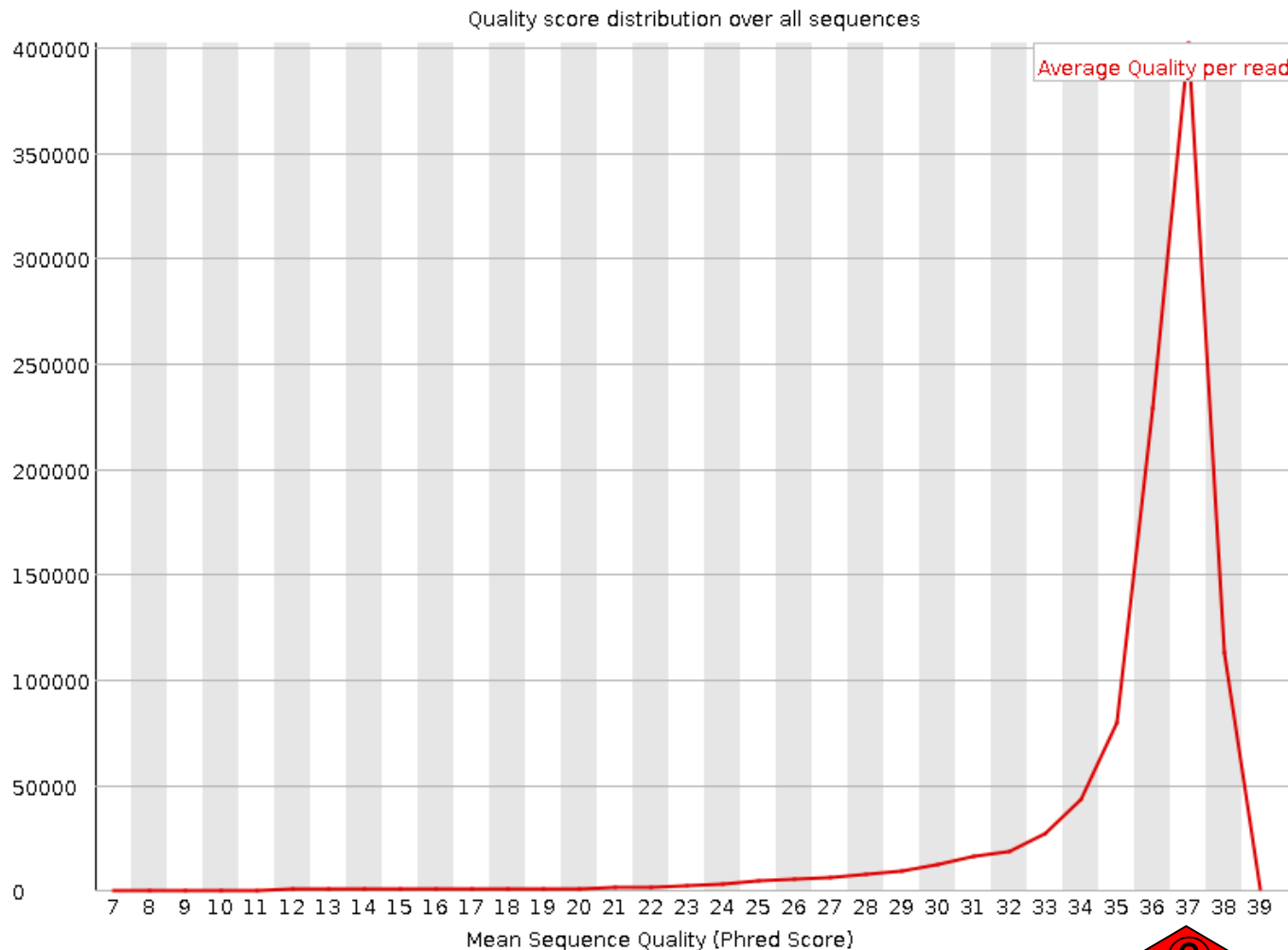
①配列ごとのクオリティスコア。横軸が平均のクオリティスコア。②平均スコアは37。縦軸は配列数。全部で100万リードなので、縦軸が数十万なのは妥当。

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: 結果を眺める

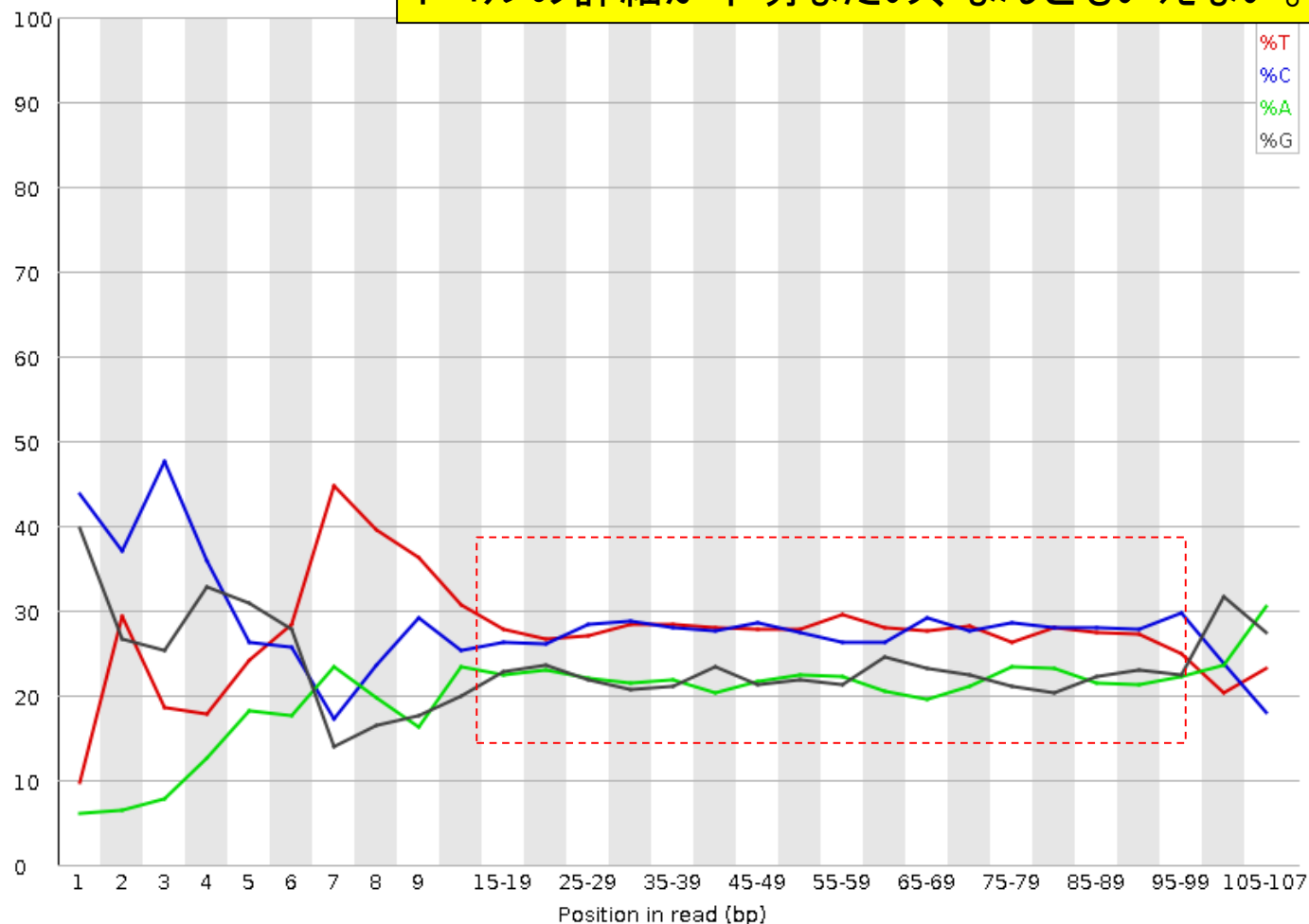
①ポジションごとの塩基の出現確率。赤枠のような塩基ごとのプロファイルがフラットになっていれば基本的にはOK。1-15番目あたりのプロファイルは、実験プロトコル (oligo dT primer or random hexamer; DNase Iを利用した破碎) に由来するのかもしれない。このデータの場合は、原著論文がなく実験プロトコルの詳細が不明なため、なんともいえない。

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: 結果を眺める

FastQC Report

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で解析してることが判明。

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 ①
- ✗ Kmer Content

✗ Overrepresented sequences












Sequence	Count	Percentage	Possible Source
CCCCGGTATATTTTCGGCGCAGTGCCACTCGACTAGTGAGCTATTACGCA	14383	1.4383	No Hit ②
GGCCTATTCAGTGCAGGCTGACCTTGCAGTACAGCCCTTCTTCCGAAGT	11044	1.1044	No Hit
GTGCTTTTCACCTTTCCCTCACGGTACTGGTTCAGTACTCGGTCAGTACGGG	8892	0.8892000000000001	No Hit
CCCGGTATATTTTCGGCGCAGTGCCACTCGACTAGTGAGCTATTACGCAC	8474	0.8474	No Hit
GTCAGTACGGGAGTATTTAGCCTTGGGAGATGGTCTCCCGGATTCGACG	8189	0.8189	No Hit
GCCGGCATTCTCACTTCTAAGCGCTCCAGCCGCTCAGATCGACCTTC	8132	0.8132	No Hit
GTCCAGTCTACAACCCCGAGAAGCAAGCTTCTCGGTTTGGGCTCTTCCC	6663	0.6663	No Hit
GTCGGTTTGGGTCAGGGTAGTTTATTTCTCACTAGAAGCTTTTCTTGGC	6411	0.6411	No Hit
GGTCACTTGGTTTCCGGTCTACATCTGCTTACTCATTGCCCCTGTTTCCAGA	5502	0.5502	No Hit
GCCGGCATTCTCACTTCTAAGCGCTCCAGCCGCTCAGATCAACCTTC	4845	0.48450000000000004	No Hit
CCCTCCATCGCTTAAACAATAAACTAGTGCAGGAATCTCAACCTGCTT	4395	0.43949999999999995	No Hit
CCGGTATATTTTCGGCGCAGTGCCACTCGACTAGTGAGCTATTACGCACT	4385	0.4385	No Hit
CCCGGTCTGCCGCGCCAGCTATGTATTCAGTACAAGCAATACACTG	4366	0.4366	No Hit
CCACAGTTTCCGGTATTATGCTTAGCCCCGGTATATTTTCGGCGCAGTGCC	4314	0.4314	No Hit
CTGGGCTGTTCCCTTTTCGACAATGGACCTTATCGCTCACTGTCTGACTC	4113	0.41130000000000005	No Hit
CCGCGTACTCAGGATCCTGGACGGAGGGTTCGACGTTTTCGCTTACAGGG	4081	0.4081	No Hit
CCGGCATTCTCACTTCTAAGCGCTCCAGCCGCTCAGATCGACCTTCA	3846	0.3846	No Hit

W8-6: 結果を眺める

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](#)

GCCACATCCTTTTCCACTTAGCATAAATTTAGG			
CCTCCATCGCTTAAACAAAATAAACTAGTGCAGGAATCTCAACCTGCTTG	2451	0.2451	No Hit
CTTGGGAGATGGTCTCCCGGATTCCGACGGAATTTACGTGTTCCGCCG	2438	0.24380000000000002	No Hit
CTTGGTTTCGGGTCTACATCTGCTTACTCATTCCGCCCTGTTTCAGACTCGC	2430	0.243	No Hit
GTCATGGGTAGGTCACCTTGGTTTCGGGTCTACATCTGCTTACTCATTCCGC	2429	0.2429	No Hit
CTAGGGAGTATTTAGCCTTGGGAGATGGTCTCCCGGATTCCGACGGAAT	2421	0.2421	No Hit
② GATCGGAAGAGCACACGTCTGAACTCCAGTCACTTAGGCATCTCGTATGC	2415	0.2415	TruSeq Adapter, Index 3 (100% over 50bp)
CCGGCATTCTCACTTCTAAGCGCTCCAGCCGCTCACGATCAACCTTCA	2393	0.2393	No Hit
GGGAGTATTTAGCCTTGGGAGATGGTCTCCCGGATTCCGACGGAATTTTC	2365	0.2365	No Hit
TGGGCCTATTCACTGCGGTGACCTTGGGTCAGCACCCCTTCTTCCGAA	2332	0.2332	No Hit
CTCCTGCGTCCCTCCATCGCTTAAACAAAATAAACTAGTGCAGGAATCTC	2331	0.23310000000000003	No Hit
CGCCGCTACTATGGGAATCGAGTTTTCTTCTTCTTCTGCGGGTACTGAG	2255	0.2255	No Hit
CGCCGGTTTATTCTACAAAAGGCAGCCATTACCGTTAACGGGCTTTGA	2172	0.21719999999999998	No Hit
CCGCCGGCCAGCTATGTATTCAGTACGCAAGCAATACACTGATGTGACTG	2081	0.2081	No Hit
CCCCTTTCGACAATGGACCTTATCGCTCACTGTCTGACTCCCGGAGTAAG	2070	0.207	No Hit
CTTAACCTGCTGATCTGCTCTTCCCTTTCCGACAATGACCTTATCGCT	2005	0.20049999999999998	No Hit














W8-6: 結果を眺める

②や③のように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。

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](#)

	GTCCTCTCTGCACTCAAGTTCCCAGTTCGGATGGGCTTCCTCGGTTA	1559	0.1559	No Hit
	GATCGGAAGAGCACACGTCTGAACTCCAGTCACCGATGTATCTCGTATGC	1539	0.1539	TruSeq Adapter, Index 2 (100% over 50bp)
	CTCGCCTTAGATCCCGACTAACCCCTGGGAGGACGAGCCTTCCCCAGGAAA	1499	0.14989999999999998	No Hit
	GGGACCTTAACTGGTGATCTGGGCTGTCCCTTTTCGACAATGGACCTTA	1497	0.1497	No Hit
	ATTTAGCCTTGGGAGATGGTCTCCCGATTCCGACGGAATTTACAGTGT	1487	0.1487	No Hit
	CATGGGTAGGTCACCTTGGTTTCGGGTCTACATCTGCTTACTCATTCGCC	1464	0.1464	No Hit
	CCCTACTGCTGCCTCCCGTAGGAGTTTGGGCCGTGTCTCAGTCCCAATGT	1455	0.1455	No Hit
	CTTTCGACAATGGACCTTATCGCTCACTGTCTGACTCCCGGAGTAAGATC	1451	0.1451	No Hit
	CGCTACTCATGCCGGCATTCTCACTTCTAAGCGCTCCAGCCGCTCCTCACG	1405	0.1405	No Hit
	CCGGGGTGCTTTTACCTTTCCCTCACGGTACTGGTTCCTACTATCGGTCAC	1361	0.1361	No Hit
	CTGGTGATCTGGGCTGTCCCTTTTCGACAATGGACCTTATCGCTCACTG	1338	0.1338	No Hit
	GGGCCTATCACTGCGGCTGACCTTGGCGTACGACCCCTTCTCCGAAG	1337	0.1337	No Hit
	GCCGCCGGCCAGCTATGTATCACTGACAAGCAATACACTGATGTGTACT	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、ページ下部の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12月9日現在のFastQCプログラムの最新版は、2015年10月9日にリリースされたver. 0.11.4。第4回原稿はver. 0.11.3ですが、ここではver. 0.11.4で進めます。



About | People | Services | Project

FastQC

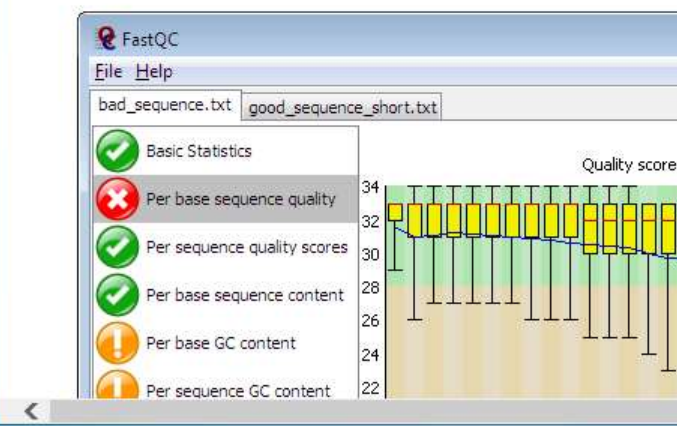
Function	A quality control tool for high-throughput sequencing data
Language	Java
Requirements	A suitable Java Runtime Environment The Picard BAM/SAM Library
Code Maturity	Stable. Mature code, but few tests
Code Released	Yes, under GPL v3 or later .
Initial Contact	Simon Andrews

[Download Now](#)

Changelog

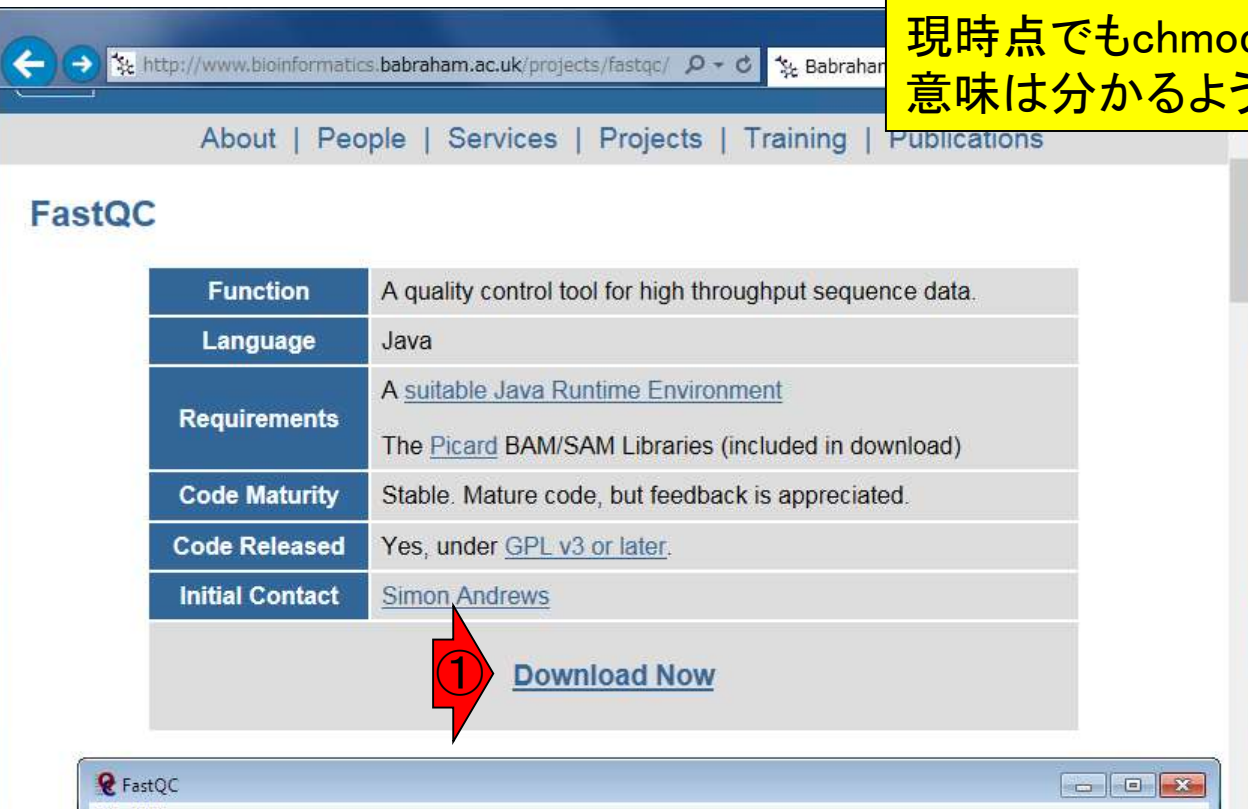


- 09-10-15: Version 0.11.4 released
 - Changed the OSX launcher to not rely on the internal JVM framework, but use any command line java which is found
 - Fixed a typo in one of the adapter sequences
 - Fixed a bug which meant that some file extensions weren't removed from report names in non-interactive mode
 - Made the per-tile module not collect any stats if it's disabled in limits.txt
 - Fixed a bug in the calculation of duplication for highly duplicated, ordered files with very small numbers of sequences
 - Fixed an incorrect error flag in the per-base quality module where there were less than 100 observations in a read group
- 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



W9-1 : FastQC

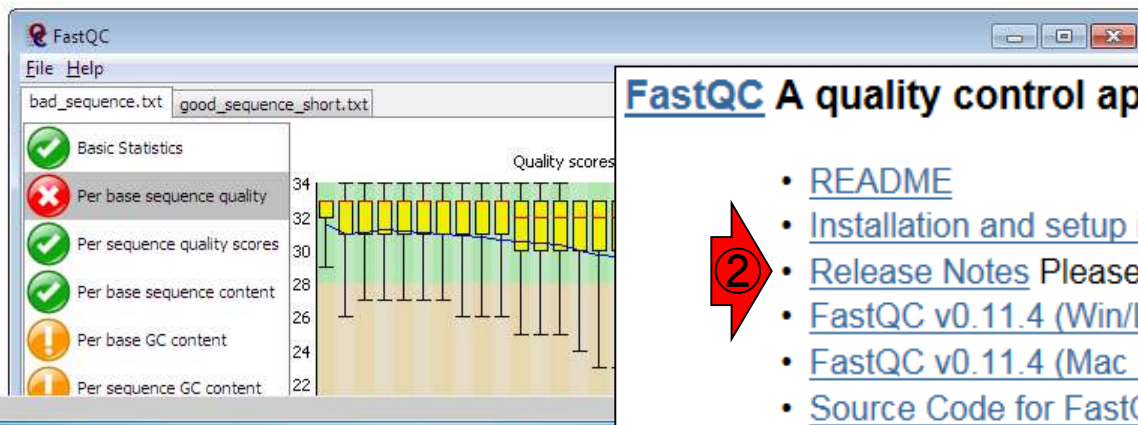
Changelogよりも、②の「Release Notes」が一般的かもしれません。③基本的には、ここを見ながら自力でインストールします。リンク先のマニュアルをざっと眺めても、現時点でもchmodなどいくつかのコマンドや相対パスの意味は分かるようになっているでしょう。③をクリック。



FastQC

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

1 [Download Now](#)



FastQC

File Help

bad_sequence.txt good_sequence_short.txt

- Basic Statistics
- Per base sequence quality
- Per sequence quality scores
- Per base sequence content
- Per base GC content
- Per sequence GC content

Quality scores

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 v0.11.4 \(Win/Linux zip file\)](#)
- [FastQC v0.11.4 \(Mac DMG image\)](#)
- [Source Code for FastQC v0.11.4 \(zip file\)](#)

W9-2: Install (FastQC)

Installing FastQC

FastQC is a java application. In order to run it needs your system to have Java Runtime Environment (JRE) installed. Before you try to run FastQC ensure that you have a suitable JRE. There are a number of different JREs however the ones we have tested are the v1.6-v1.8 JREs from Oracle. The following are instructions for a number of different platforms.

Windows/Linux: Go to java.com - click on Free Java Download - DON'T click on the big red button but choose the smaller link to "See all java downloads". Find your operating system and click on the appropriate offline installer. If you are using a 64bit operating system (and nearly everyone is these days), then make sure you select the 64bit version of the the installer.


OSX: On newer versions of OSX you need to install the Java Development Kit. The normal Java runtime environment IS NOT enough. To get this go to java.com, click "Free java download", then IGNORE the big red button, and select "See all java downloads", on the next screen select "Looking for the JDK?" from the left hand menu and select the link to "JDK downloads" in the first paragraph. You can then click the "Download" button underneath JDK in the page you are taken to. Sorry this is such a pain!

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.

At the command prompt type 'java -version' and press enter.  ② should see something like:

```
java version "1.8.0_60"  
Java(TM) SE Runtime Environment (build 1.8.0_60-b27)  
Java HotSpot(TM) 64-Bit Server VM (build 25.60-b23, mixed mode)
```

If you get an error then you don't have java installed. If the version listed on the first line is less than 1.6 then you might have problems running FastQC.  ③

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1.6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1.6以上(ver. 1.7.0_79)であることを確認。

```
File Edit View Search Terminal Help
iu@bielinux[iu] pwd
/home/iu
iu@bielinux[iu] java -version
java version "1.7.0_79"
OpenJDK Runtime Environment (IcedTea 2.5.5) (7u79-2.5.5-0ubuntu0.14.04.2)
OpenJDK 64-Bit Server VM (build 24.79-b02, 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_79"
OpenJDK Runtime Environment (IcedTea 2.5.5) (7u79-2.5.5-0ubuntu0.14.04.2)
OpenJDK 64-Bit Server VM (build 24.79-b02, mixed mode)
iu@bielinux[Desktop]
```


W9-2: Install (FastQC)

FastQC (ver. 0.11.4)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ととりあえずDownloadsディレクトリで作業を行う。④「wget -cq」と打ち込んだ後に、②で取得したURL情報をペースト。⑤ls -lhで確認。

FastQC A quality control application for high throughput sequen

- [README](#)
- [Installation and setup instructions](#)
- [Release Notes](#) Please read these before using the program.
- [FastQC v0.11.4 \(Win/Linux zip file\)](#)
- [FastQC v0.11.4 \(Mac DMG image\)](#)
- [Source Code for FastQC v0.11.4 \(z](#)



FastQ Screen A screening application

- [README](#)
- [Release Notes](#) Please read these b
- [FastQ Screen v0.5.2](#)

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)



```
iu@bielinux[Downloads] pwd  
/home/iu/Downloads  
iu@bielinux[Downloads] wget -cq http://www.bioinformatics.babraham.ac.uk/proje  
cts/fastqc/fastqc_v0.11.4.zip  
iu@bielinux[Downloads] ls -lh  
total 9.6M  
-rw-rw-r-- 1 iu iu 9.6M 10月 9 20:55 fastqc_v0.11.4.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赤下線で示す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 10月  9 20:55 fastqc_v0.11.4.zip
iu@bielinux[Downloads] unzip -q fastqc_v0.11.4.zip
iu@bielinux[Downloads] ls -lh
total 9.6M
drwxrwxr-x 8 iu iu 4.0K 12月 10 10:09 FastQC
-rw-rw-r-- 1 iu iu 9.6M 10月  9 20:55 fastqc_v0.11.4.zip
iu@bielinux[Downloads] ls FastQC
cisd-jhdf5.jar      Help          LICENSE.txt    RELEASE_NOTES.txt  uk
Configuration     INSTALL.txt   net            run_fastqc.bat
fastqc             jbzip2-0.9.jar org            sam-1.103.jar
fastqc           LICENSE_JHDF5.txt README.txt     Templates
fastqc_icon.ico   LICENSE_
iu@bielinux[Downloads]
```

[10:09午前]

[10:09午前]

[10:09午前]

[10:09午前]

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 v0.11.4 \(Win/Linux zip file\)](#)
- [FastQC v0.11.4 \(Mac DMG image\)](#)
- [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
```


①fastqcに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] pwd [11:09午前]
/home/iu/Downloads
iu@bielinux[Downloads] ls -lh [11:09午前]
total 9.6M
drwxrwxr-x 8 iu iu 4.0K 12月 10 10:09 FastQC
-rw-rw-r-- 1 iu iu 9.6M 10月 9 20:55 fastqc_v0.11.4.zip
iu@bielinux[Downloads] cd FastQC [11:09午前]
iu@bielinux[FastQC] pwd [11:09午前]
/home/iu/Downloads/FastQC
iu@bielinux[FastQC] ls -lh fastqc* [11:09午前]
-rw-rw-r-- 1 iu iu 14K 10月 9 11:59 fastqc
-rw-rw-r-- 1 iu iu 2.2K 3月 21 2012 fastqc_icon.ico
iu@bielinux[FastQC] chmod 755 fastqc [11:09午前]
iu@bielinux[FastQC] ls -lh fastqc* [11:09午前]
-rwxr-xr-x 1 iu iu 14K 10月 9 11:59 fastqc
-rw-rw-r-- 1 iu iu 2.2K 3月 21 2012 fastqc_icon.ico
iu@bielinux[FastQC] █ [11:09午前]
```



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]。

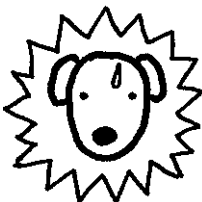
```
iu@bielinux[FastQC] pwd
/home/iu/Downloads/FastQC
iu@bielinux[FastQC] ls -lh fastqc*
-rwxr-xr-x 1 iu iu 14K 10月 9 11:59 fastqc
-rw-rw-r-- 1 iu iu 2.2K 3月 21 2012 fastqc_icon.ico
iu@bielinux[FastQC] fastqc -v
FastQC v0.10.1
iu@bielinux[FastQC] █
```

[11:13午前]

[11:13午前]

[11:13午前]

[11:13午前]



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。

```
File Edit View Search Terminal Help
iu@bielinux[FastQC] pwd
/home/iu/Downloads/FastQC
iu@bielinux[FastQC] ls -lh fastqc*
-rwxr-xr-x 1 iu iu 14K 10月 9 11:59 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 v0.11.4
③ iu@bielinux[FastQC] /home/iu/Downloads/FastQC/fastqc -v
FastQC v0.11.4
④ iu@bielinux[FastQC] ~/Downloads/FastQC/fastqc -v
FastQC v0.11.4
iu@bielinux[FastQC] █
```

[11:15午前]

[11:13午前]

[11:15午前]

[11:15午前]

[11:15午前]

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 [11:13午前]
/home/iu/Downloads/FastQC
iu@bielinux[FastQC] ls -lh fastqc* [11:13午前]
-rwxr-xr-x 1 iu iu 14K 10月 9 11:59 fastqc
-rw-rw-r-- 1 iu iu 2.2K 3月 21 2012 fastqc_icon.ico
① iu@bielinux[FastQC] fastqc -v [11:13午前]
FastQC v0.10.1
② iu@bielinux[FastQC] ./fastqc -v [11:13午前]
FastQC v0.11.4
③ iu@bielinux[FastQC] /home/iu/Downloads/FastQC/fastqc -v [11:15午前]
FastQC v0.11.4
④ iu@bielinux[FastQC] ~/Downloads/FastQC/fastqc -v [11:15午前]
FastQC v0.11.4
iu@bielinux[FastQC] █ [11:15午前]
```

W9-3: パス (のイントロ)

おさらい。①FastQC (ver. 0.10.1)の結果はディレクトリとzipファイルだった。FastQC (ver. 0.11.4)の結果がディレクトリではなくhtmlファイルが作成されることを示すべく、②一旦削除。

```
iu@bielinux[FastQC] cd ~/Documents/srp017156
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
SRR616268sub_1.fastq     SRR616268sub_2_fastqc.zip
iu@bielinux[srp017156] rm -rf SRR616268sub_2_*
iu@bielinux[srp017156] ls
JSLAB4_1_Linux.sh          SRR616268_2.fastq.bz2  SRR616268sub_1_fastqc.zip
JSLAB4_2_Linux.sh          SRR616268sub_1.fastq   SRR616268sub_2.fastq
SRR616268_1.fastq.bz2     SRR616268sub_1_fastqc
iu@bielinux[srp017156] █
```

[11:36午前]

[11:36午前]

[11:36午前]

[11:36午前]

[11:36午前]



W9-3: パス (のイントロ)

③最新版のFastQC (ver. 0.11.4)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。①FastQC (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

```
File Edit View Search Terminal Help
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
SRR616268sub_1.fastq SRR616268sub_2_fastqc.zip
iu@bielinux[srp017156] rm -rf SRR616268sub_2_*
iu@bielinux[srp017156] ls
JSLAB4_1_Linux.sh SRR616268_2.fastq.bz2 SRR616268sub_1_fastqc.zip
JSLAB4_2_Linux.sh SRR616268sub_1.fastq SRR616268sub_2.fastq
SRR616268_1.fastq.bz2 SRR616268sub_1_fastqc
iu@bielinux[srp017156] ~/Downloads/FastQC/fastqc -v [11:36午前]
FastQC v0.11.4
iu@bielinux[srp017156] ~/Downloads/FastQC/fastqc -q SRR616268sub_2.fastq [11:42午前]
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
iu@bielinux[srp017156] [11:42午前]
```


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と打つと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
iu@bielinux[srp017156] fastqc -v
FastQC v0.10.1
iu@bielinux[srp017156] █
```

[11:48午前]

[11:48午前]

[11:48午前]



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」、③「which コマンド名」、または④「where 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。3つのコマンド(whereis, which, and where)の挙動の違いはW9-11にもあり。

```
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
④ iu@bielinux[srp017156] where fastqc
/usr/local/bin/fastqc
iu@bielinux[srp017156] █
```

[11:48午前]
[11:48午前]
[11:52午前]
[11:52午前]
[11:52午前]

W9-4:パス

①の「fastqc -v」は、⑤の「/usr/local/bin/fastqc -v」の省略形。
⑤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 [11:48午前]
/home/iu/Documents/srp017156
iu@bielinux[srp017156] ls [11:48午前]
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 [11:48午前]
FastQC v0.10.1
② iu@bielinux[srp017156] whereis fastqc [11:48午前]
fastqc: /usr/local/bin/fastqc
③ iu@bielinux[srp017156] which fastqc [11:52午前]
/usr/local/bin/fastqc
④ iu@bielinux[srp017156] where fastqc [11:52午前]
/usr/local/bin/fastqc
⑤ iu@bielinux[srp017156] /usr/local/bin/fastqc -v [11:52午前]
FastQC v0.10.1
iu@bielinux[srp017156] [11:57午前]
```


W9-4: パス

①のFastQC (ver. 0.11.4)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
iu@bielinux[srp017156] fastqc -v
FastQC v0.10.1
iu@bielinux[srp017156] whereis fastqc
fastqc: /usr/local/bin/fastqc
iu@bielinux[srp017156] which fastqc
/usr/local/bin/fastqc
iu@bielinux[srp017156] where fastqc
/usr/local/bin/fastqc
iu@bielinux[srp017156] /usr/local/bin/fastqc -v
FastQC v0.10.1
iu@bielinux[srp017156] /home/iu/Downloads/FastQC/fastqc -v
FastQC v0.11.4
iu@bielinux[srp017156] █
```

[11:48午前]

[11:48午前]

[11:52午前]

[11:52午前]

[11:52午前]

[11:57午前]

[12:01午後]



① /usr/local/binに移動し、②ls。③fastqcコマンドは確かに存在する。この詳細情報をls -lhで眺める。

W9-4: パス

```
iu@bielinux[srp017156] cd /usr/local/bin [12:05午後]
iu@bielinux[bin] pwd [12:06午後]
/usr/local/bin
iu@bielinux[bin] ls [12:06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       sam2vcf
big-blast.pl            fastm36_t       MSPcrunch.LIN  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       stars
```

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 17 13:25 fastqc -> ../bioinf/fastqc/fastqc/fastqc
iu@bielinux[bin]
```

[12:17午後]

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
iu@bielinux[bin] ls -lh fastqc*
lrwxrwxrwx 1 root root 30 11月 17 13:25 fastqc -> ../bioinf/fastqc/fastqc/fastqc
iu@bielinux[bin] cd ../bioinf/fastqc/fastqc
iu@bielinux[fastqc] pwd
/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
-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]
```

[12:17午後]
[12:18午後]
[12:18午後]
[12:18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 software
- [FastQC v0.11.4 \(Win/Linux zip file\)](#)
- [FastQC v0.11.4 \(Mac DMG image\)](#)
- [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:

```
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
```

FastQC (ver. 0.11.4)の実行コマンド

/home/iu/Downloads/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！

②

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(sudo)で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。それゆえ、(上書きしちゃってもいいが)ここでは④ `/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。

```
./fastqc
```

```
..or place a link in /usr/local/bin
```

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

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

```
iu@bielinux[fastqc] cd ~/Downloads/FastQC [12:29午後]
iu@bielinux[FastQC] pwd [12:30午後]
/home/iu/Downloads/FastQC
iu@bielinux[FastQC] ls -lh fastqc* [12:30午後]
-rwxr-xr-x 1 iu iu 14K 10月 9 11:59 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 [12:30午後]
FastQC v0.11.4
iu@bielinux[FastQC] /usr/local/bin/fastqc -v [12:30午後]
FastQC v0.10.1
iu@bielinux[FastQC] ls -lh /usr/local/bin/fastqc* [12:30午後]
lrwxrwxrwx 1 root root 30 11月 17 13:25 /usr/local/bin/fastqc -> ../bioin
f/fastqc/fastqc/fastqc
iu@bielinux[FastQC] █ [12:30午後]
```

W9-5: パスを通す

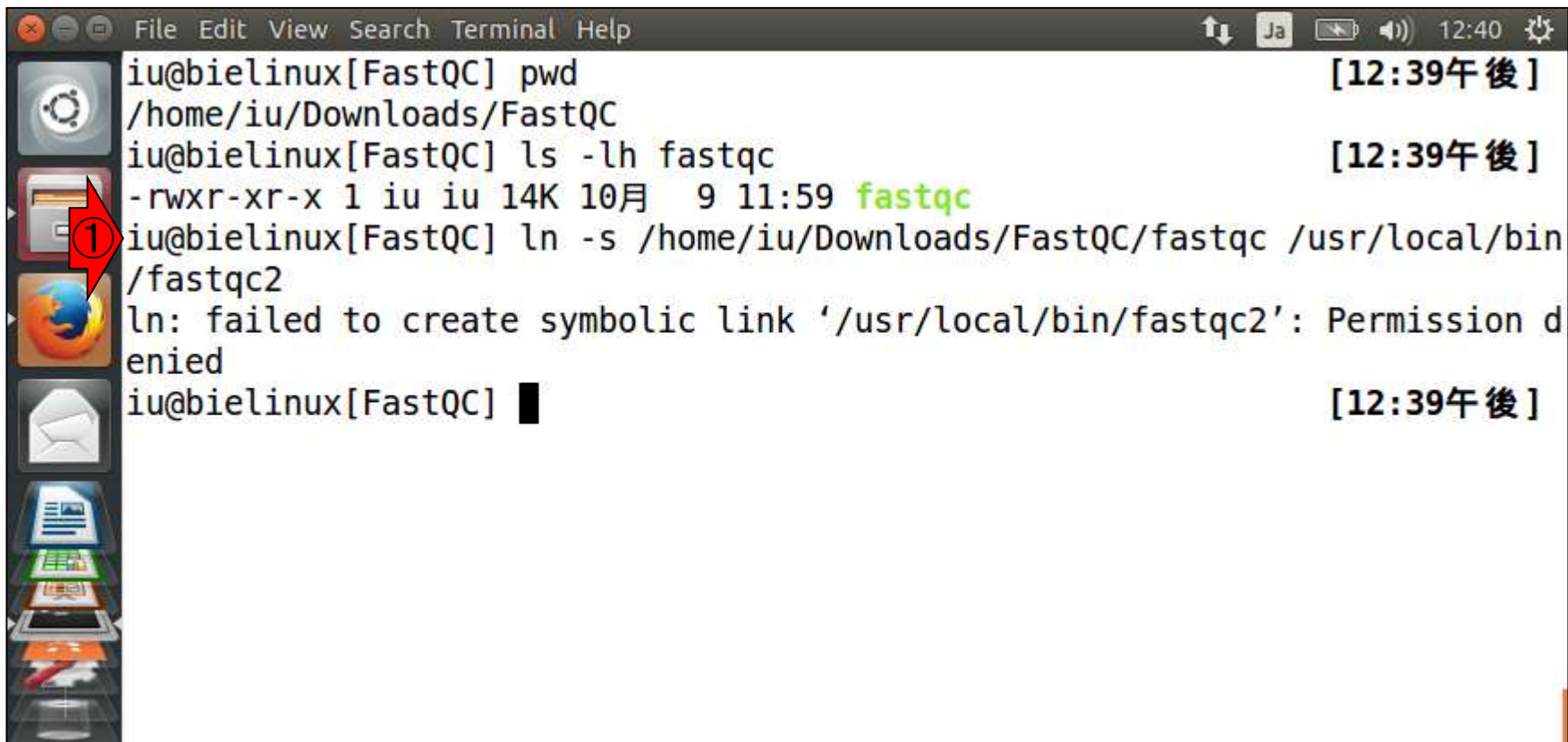
Tips。①管理者権限(sudo)で実行しないと失敗します。理由は、/usr/local/binは管理者のみが書き込める場所であり、ユーザiuにはここにfastqc2というファイルを置く権限がないためです。

```
./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 [12:39午後]
/home/iu/Downloads/FastQC
iu@bielinux[FastQC] ls -lh fastqc [12:39午後]
-rwxr-xr-x 1 iu iu 14K 10月 9 11:59 fastqc
iu@bielinux[FastQC] ln -s /home/iu/Downloads/FastQC/fastqc /usr/local/bin
/fastqc2
ln: failed to create symbolic link '/usr/local/bin/fastqc2': Permission d
enied
iu@bielinux[FastQC] █ [12:39午後]
```

W9-5: パスを通す

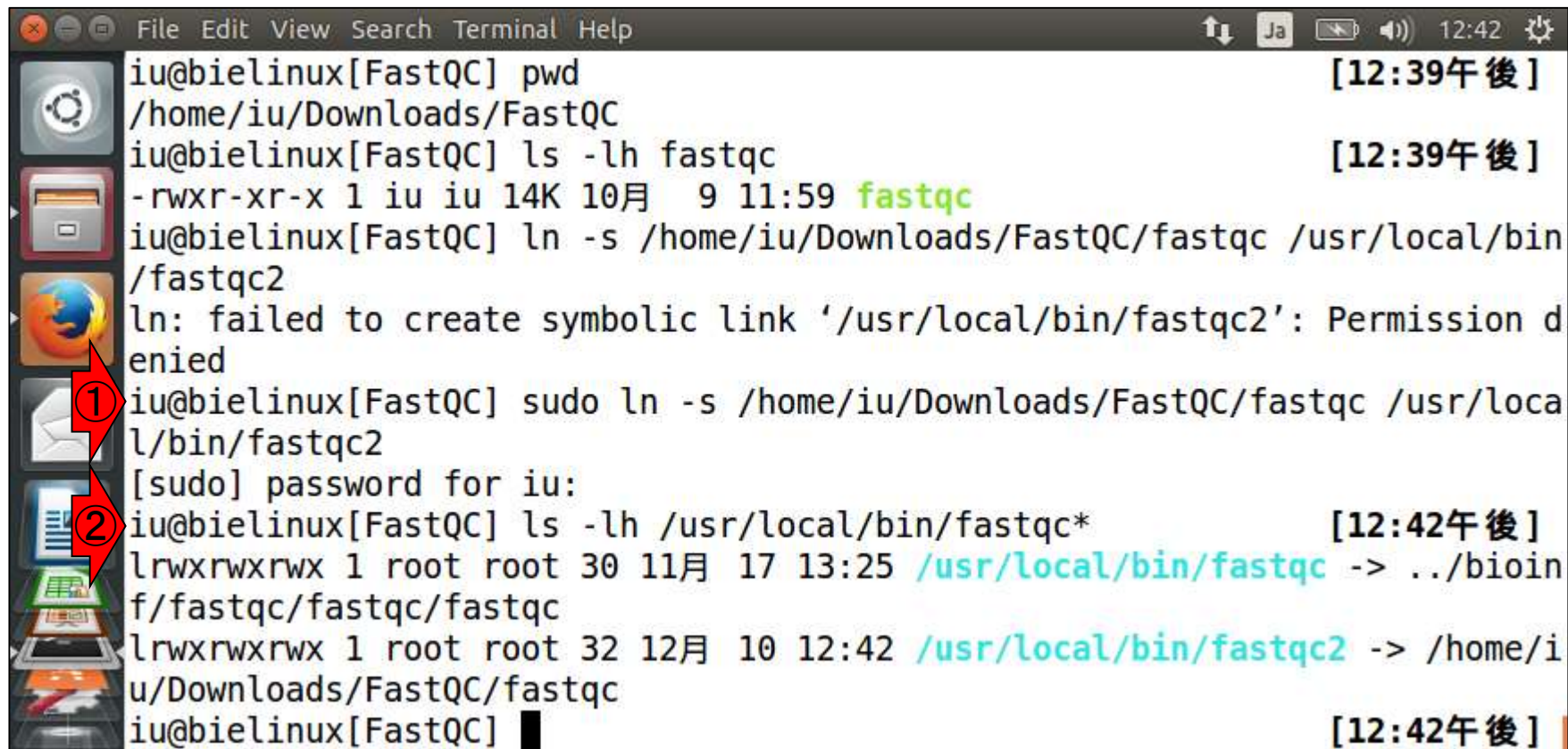
①sudoをつけてリトライ。パスワードを要求されたら、「ログインパスワード(pass1409)」を打ち込む。②/usr/local/bin中のfastqc*の条件を満たすコマンド群を表示。

```
./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 [12:39午後]
/home/iu/Downloads/FastQC
iu@bielinux[FastQC] ls -lh fastqc [12:39午後]
-rwxr-xr-x 1 iu iu 14K 10月 9 11:59 fastqc
iu@bielinux[FastQC] ln -s /home/iu/Downloads/FastQC/fastqc /usr/local/bin
/fastqc2
ln: failed to create symbolic link '/usr/local/bin/fastqc2': Permission d
enied
① iu@bielinux[FastQC] sudo ln -s /home/iu/Downloads/FastQC/fastqc /usr/loca
l/bin/fastqc2
[sudo] password for iu:
② iu@bielinux[FastQC] ls -lh /usr/local/bin/fastqc* [12:42午後]
lrwxrwxrwx 1 root root 30 11月 17 13:25 /usr/local/bin/fastqc -> ../bioin
f/fastqc/fastqc/fastqc
lrwxrwxrwx 1 root root 32 12月 10 12:42 /usr/local/bin/fastqc2 -> /home/i
u/Downloads/FastQC/fastqc
iu@bielinux[FastQC] [12:42午後]
```


W9-6: 確認

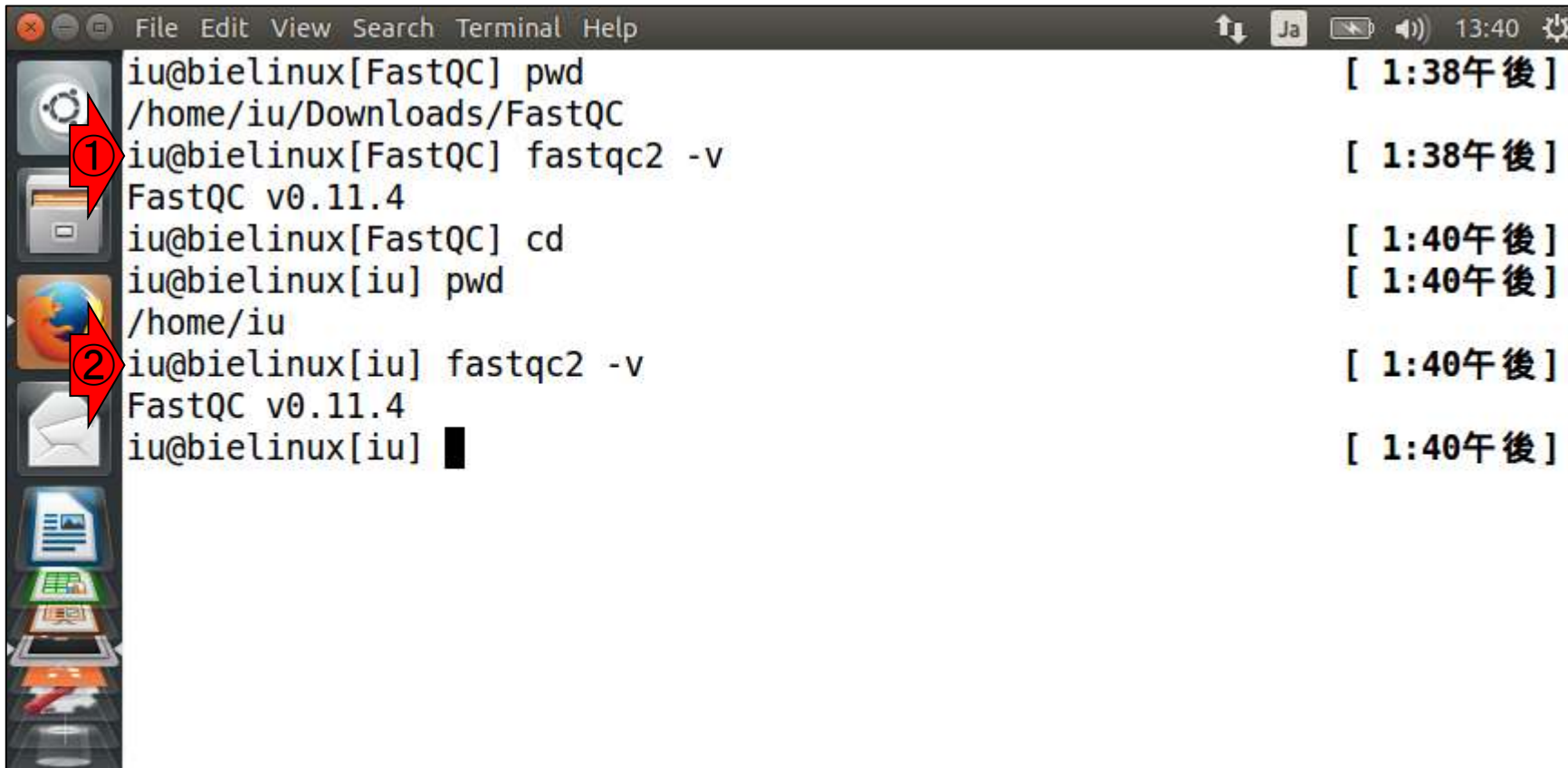
(作業ディレクトリはどこでもよい)①fastqまで打ってからTabキーを押した結果。パスが通っている(コマンド名を打ち込むだけで利用可能な)fastq*という条件を満たす候補がリストアップされる。シンボリックリンクを張ったfastqc2が見られる。

```
iu@bielinux[FastQC] pwd
/home/iu/Downloads/FastQC
iu@bielinux[FastQC] fastq
fastqc
fastqc2
fastqc-clipper
fastq-join
fastq_masker
fastq-mcf
fastq-multx
fastq_quality_boxplot_graph.sh
fastq_quality_converter
fastq_quality_filter
fastq_quality_trimmer
fastq-stats
fastq_to_fasta
```

[1:38午後]

(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。

W9-6: 確認



```
File Edit View Search Terminal Help
iu@bielinux[FastQC] pwd [ 1:38午後 ]
/home/iu/Downloads/FastQC
① iu@bielinux[FastQC] fastqc2 -v [ 1:38午後 ]
FastQC v0.11.4
iu@bielinux[FastQC] cd [ 1:40午後 ]
iu@bielinux[iu] pwd [ 1:40午後 ]
/home/iu
② iu@bielinux[iu] fastqc2 -v [ 1:40午後 ]
FastQC v0.11.4
iu@bielinux[iu] [ 1:40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)。以前作成したFastQC実行結果を削除しておき、①lsで確認。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。どこにもFastQC実行結果がないようにしておくことが重要であり、削除手段は問わない。

```
iu@bielinux[iu] cd ~/Documents/srp017156
iu@bielinux[srp017156] pwd
/home/iu/Documents/srp017156
iu@bielinux[srp017156] rm -rf SRR616268sub_
iu@bielinux[srp017156] ls -lh
total 740M
-rwxr-xr-x 1 iu iu 360 5月 1 2015 JSLAB_1.sh
-rwxr-xr-x 1 iu iu 424 5月 5 2015 JSLAB4_2_Linux.sh
-rw-rw-r-- 1 iu iu 83M 12月 3 14:56 SRR616268_1.fastq.bz2
-rw-rw-r-- 1 iu iu 74M 12月 3 15:51 SRR616268_2.fastq.bz2
-rw-rw-r-- 1 iu iu 306M 12月 9 15:24 SRR616268sub_1.fastq
-rw-rw-r-- 1 iu iu 278M 12月 9 15:24 SRR616268sub_2.fastq
iu@bielinux[srp017156] ls ~/Desktop/mac_share
JSLAB4_1.sh Lactobacillus_casei_12a.GCA_000309565.1.22.dna.toplevel.fa
iu@bielinux[srp017156]
```

[1:52午後]
[1:52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. 0.11.4)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] rm -rf SRR616268sub_1
iu@bielinux[srp017156] ls -lh
total 740M
-rwxr-xr-x 1 iu iu 360 5月 1 2015 JSLAB4_
-rwxr-xr-x 1 iu iu 424 5月 5 2015 JSLAB4_
-rw-rw-r-- 1 iu iu 83M 12月 3 14:56 SRR6162
-rw-rw-r-- 1 iu iu 74M 12月 3 15:51 SRR616268_2.fastq.bz2
-rw-rw-r-- 1 iu iu 306M 12月 9 15:24 SRR616268sub_1.fastq
-rw-rw-r-- 1 iu iu 278M 12月 9 15:24 SRR616268sub_2.fastq
iu@bielinux[srp017156] ls ~/Desktop/mac_share [ 1:52午後 ]
JSLAB4_1.sh Lactobacillus_casei_12a.GCA_000309565.1.22.dna.toplevel.fa
① iu@bielinux[srp017156] fastqc2 -q SRR616268sub_1.fastq --outdir=~/Desktop
/mac_share
Specified output directory '~/Desktop/mac_share' does not exist
② iu@bielinux[srp017156] fastqc2 -q SRR616268sub_1.fastq --outdir=/home/iu/
Desktop/mac_share
③ iu@bielinux[srp017156] ls ~/Desktop/mac_share [ 1:59午後 ]
JSLAB4_1.sh
Lactobacillus_casei_12a.GCA_000309565.1.22.dna.toplevel.fa
SRR616268sub_1_fastqc.html
SRR616268sub_1_fastqc.zip
iu@bielinux[srp017156] [ 2:00午後 ]
```

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          [ 2:06午後 ]
gzip SRR616268sub_2.fastq 17.88s user 0.16s system 99% cpu 18.168 total
② iu@bielinux[srp017156] ls                                       [ 2:06午後 ]
JSLAB4_1_Linux.sh  SRR616268_1.fastq.bz2  SRR616268sub_1.fastq
JSLAB4_2_Linux.sh  SRR616268_2.fastq.bz2  SRR616268sub_2.fastq.gz
③ iu@bielinux[srp017156] fastqc2 -q SRR616268sub_2.fastq.gz --outdir=/home/
iu/Desktop/mac_share
④ iu@bielinux[srp017156] ls ~/Desktop/mac_share                    [ 2:06午後 ]
JSLAB4_1.sh
Lactobacillus_casei_12a.GCA_000309565.1.22.dna.toplevel.fa
SRR616268sub_1_fastqc.html
SRR616268sub_1_fastqc.zip
SRR616268sub_2_fastqc.html
SRR616268sub_2_fastqc.zip
iu@bielinux[srp017156] █                                         [ 2:06午後 ]
```


W9-8: Tips (長いコマンド)

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だ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④\
(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。

```
iu@bielinux[srp017156] pwd
/home/iu/Documents/srp017156
iu@bielinux[srp017156] ls ~/Desktop/mac_share
JSLAB4_1.sh
Lactobacillus_casei_12a.GCA_000309565.1.22.dna.topl
SRR616268sub_1_fastqc.html
SRR616268sub_1_fastqc.zip
SRR616268sub_2_fastqc.html
SRR616268sub_2_fastqc.zip
iu@bielinux[srp017156] ls ~/Desktop/mac_share/*_2_* [ 2:12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_* [ 2:12午後 ]
iu@bielinux[srp017156] ls ~/Desktop/mac_share [ 2:12午後 ]
JSLAB4_1.sh
Lactobacillus_casei_12a.GCA_000309565.1.22.dna.toplevel.fa
SRR616268sub_1_fastqc.html
SRR616268sub_1_fastqc.zip
iu@bielinux[srp017156] fastqc2 -q SRR616268sub_2.fastq.gz\ [ 2:12午後 ]
```

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)

①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 [ 2:11午後 ]
/home/iu/Documents/srp017156
iu@bielinux[srp017156] ls ~/Desktop/mac_share [ 2:12午後 ]
JSLAB4_1.sh
Lactobacillus_casei_12a.GCA_000309565.1.22.dna.toplevel.fa
SRR616268sub_1_fastqc.html
SRR616268sub_1_fastqc.zip
SRR616268sub_2_fastqc.html
SRR616268sub_2_fastqc.zip
iu@bielinux[srp017156] ls ~/Desktop/mac_share/*_2_* [ 2:12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_* [ 2:12午後 ]
iu@bielinux[srp017156] ls ~/Desktop/mac_share [ 2:12午後 ]
JSLAB4_1.sh
Lactobacillus_casei_12a.GCA_000309565.1.22.dna.toplevel.fa
SRR616268sub_1_fastqc.html
SRR616268sub_1_fastqc.zip
iu@bielinux[srp017156] fastqc2 -q SRR616268sub_2.fastq.gz\ [ 2:12午後 ]
> █
```



W9-8: Tips (長いコマン

①「--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 [ 2:33午後 ]
JSLAB4_1.sh
Lactobacillus_casei_12a.GCA_000309565.1.22.dna.toplevel.fa
SRR616268sub_1_fastqc.html
SRR616268sub_1_fastqc.zip
SRR616268sub_2_fastqc.html
SRR616268sub_2_fastqc.zip
iu@bielinux[srp017156] ls ~/Desktop/mac_share/*_2_* [ 2:33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_* [ 2:33午後 ]
iu@bielinux[srp017156] ls ~/Desktop/mac_share [ 2:33午後 ]
JSLAB4_1.sh
Lactobacillus_casei_12a.GCA_000309565.1.22.dna.toplevel.fa
SRR616268sub_1_fastqc.html
SRR616268sub_1_fastqc.zip
iu@bielinux[srp017156] fastqc2 -q SRR616268sub_2.fastq.gz\ [ 2:33午後 ]
> --outdir=/home/iu/Desktop/mac_share
Skipping 'SRR616268sub_2.fastq.gz--outdir=/home/iu/Desktop/mac_share' whi
ch didn't exist, or couldn't be read
iu@bielinux[srp017156] [ 2:33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…とってはイケナイ!!①の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 [ 2:33午後 ]
JSLAB4_1.sh
Lactobacillus_casei_12a.GCA_000309565.1.22.dna.toplevel.fa
SRR616268sub_1_fastqc.html
SRR616268sub_1_fastqc.zip
SRR616268sub_2_fastqc.html
SRR616268sub_2_fastqc.zip
iu@bielinux[srp017156] ls ~/Desktop/mac_share/*_2_* [ 2:33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_* [ 2:33午後 ]
iu@bielinux[srp017156] ls ~/Desktop/mac_share [ 2:33午後 ]
JSLAB4_1.sh
Lactobacillus_casei_12a.GCA_000309565.1.22.dna.toplevel.fa
SRR616268sub_1_fastqc.html
SRR616268sub_1_fastqc.zip
iu@bielinux[srp017156] fastqc2 -q SRR616268sub_2.fastq.gz\ [ 2:33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
iu@bielinux[srp017156] [ 2:33午後 ]
```



W9-8: Tips (長いコマンド)

①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。

```
File Edit View Search Terminal Help 14:35
/home/iu/Documents/srp017156
iu@bielinux[srp017156] ls ~/Desktop/mac_share [ 2:33午後 ]
JSLAB4_1.sh
Lactobacillus_casei_12a.GCA_000309565.1.22.dna.toplevel.fa
SRR616268sub_1_fastqc.html
SRR616268sub_1_fastqc.zip
SRR616268sub_2_fastqc.html
SRR616268sub_2_fastqc.zip
iu@bielinux[srp017156] ls ~/Desktop/mac_share/*_2_* [ 2:33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_* [ 2:33午後 ]
iu@bielinux[srp017156] ls ~/Desktop/mac_share [ 2:33午後 ]
JSLAB4_1.sh
Lactobacillus_casei_12a.GCA_000309565.1.22.dna.toplevel.fa
SRR616268sub_1_fastqc.html
SRR616268sub_1_fastqc.zip
iu@bielinux[srp017156] fastqc2 -q SRR616268sub_2.fastq.gz\ [ 2:33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\ [ 2:33午後 ]
--outdir=/home/iu/Desktop/mac_share
```



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、②我々は「バックスラッシュ」の前に「スペース」を入れる。上下矢印キーでカーソルを移動できます。

```
File Edit View Search Terminal Help
/home/iu/Documents/srp017156
iu@bielinux[srp017156] ls ~/Desktop/mac_share
JSLAB4_1.sh
Lactobacillus_casei_12a.GCA_000309565.1.22.dna.toplevel.fa
SRR616268sub_1_fastqc.html
SRR616268sub_1_fastqc.zip
SRR616268sub_2_fastqc.html
SRR616268sub_2_fastqc.zip
iu@bielinux[srp017156] ls ~/Desktop/mac_share/*_2_* [ 2:33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_* [ 2:33午後]
iu@bielinux[srp017156] ls ~/Desktop/mac_share [ 2:33午後]
JSLAB4_1.sh
Lactobacillus_casei_12a.GCA_000309565.1.22.dna.toplevel.fa
SRR616268sub_1_fastqc.html
SRR616268sub_1_fastqc.zip
iu@bielinux[srp017156] fastqc2 -q SRR616268sub_2.fastq.gz\ [ 2:33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 \ [ 2:33午後]
--outdir=/home/iu/Desktop/mac_share
```



②

W9-8: Tips (長いコマンド)

①リターンキーを押して実行した結果。②と③のls結果の比較から、うまく動いていることがわかる。

```
iu@bielinux[srp017156] ls ~/Desktop/mac_share/*_2_* [ 2:33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_* [ 2:33午後]
iu@bielinux[srp017156] ls ~/Desktop/mac_share [ 2:33午後]
JSLAB4_1.sh
Lactobacillus_casei_12a.GCA_000309565.1.22.dna.toplevel.fa
SRR616268sub_1_fastqc.html
SRR616268sub_1_fastqc.zip
iu@bielinux[srp017156] fastqc2 -q SRR616268sub_2.fastq.gz\ [ 2:33午後]
> --outdir=/home/iu/Desktop/mac_share
Skipping 'SRR616268sub_2.fastq.gz--outdir=/home/iu/Desktop/mac_share' whi
ch didn't exist, or couldn't be read
iu@bielinux[srp017156] fastqc2 -q SRR616268sub_2.fastq.gz \ [ 2:33午後]
--outdir=/home/iu/Desktop/mac_share
iu@bielinux[srp017156] ls ~/Desktop/mac_share [ 2:36午後]
JSLAB4_1.sh
Lactobacillus_casei_12a.GCA_000309565.1.22.dna.toplevel.fa
SRR616268sub_1_fastqc.html
SRR616268sub_1_fastqc.zip
SRR616268sub_2_fastqc.html
SRR616268sub_2_fastqc.zip
iu@bielinux[srp017156] [ 2:37午後]
```


W9-9: パスの確認

whereisコマンドを用いて①fastqc2、②fastqc、③perl、④ls、⑤pwd、⑥whereisの存在場所を調べている。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
① iu@bielinux[srp017156] whereis fastqc2 [ 2:42午後 ]
fastqc2: /usr/local/bin/fastqc2
② iu@bielinux[srp017156] whereis fastqc [ 2:42午後 ]
fastqc: /usr/local/bin/fastqc
③ iu@bielinux[srp017156] whereis perl [ 2:42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
④ iu@bielinux[srp017156] whereis ls [ 2:42午後 ]
ls: /bin/ls /usr/share/man/man1/ls.1.gz
⑤ iu@bielinux[srp017156] whereis pwd [ 2:42午後 ]
pwd: /bin/pwd /usr/include/pwd.h /usr/share/man/man1/pwd.1.gz
⑥ iu@bielinux[srp017156] whereis whereis [ 2:42午後 ]
whereis: /usr/bin/whereis /usr/bin/X11/whereis /usr/share/man/man1/wherei
s.1.gz
iu@bielinux[srp017156] [ 2:42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 [ 2:42午後 ]
/home/iu/Documents/srp017156
① iu@bielinux[srp017156] whereis fastqc2 [ 2:42午後 ]
fastqc2: /usr/local/bin/fastqc2
② iu@bielinux[srp017156] whereis fastqc [ 2:42午後 ]
fastqc: /usr/local/bin/fastqc
③ iu@bielinux[srp017156] whereis perl [ 2:42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
④ iu@bielinux[srp017156] whereis ls [ 2:42午後 ]
ls: /bin/ls /usr/share/man/man1/ls.1.gz
⑤ iu@bielinux[srp017156] whereis pwd [ 2:42午後 ]
pwd: /bin/pwd /usr/include/pwd.h /usr/share/man/man1/pwd.1.gz
⑥ iu@bielinux[srp017156] whereis whereis [ 2:42午後 ]
whereis: /usr/bin/whereis /usr/bin/X11/whereis /usr/share/man/man1/wherei
s.1.gz
iu@bielinux[srp017156] [ 2:42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。

```
File Edit View Search Terminal Help
iu@bielinux[srp017156] pwd
/home/iu/Documents/srp017156
① iu@bielinux[srp017156] whereis fastqc2
fastqc2: /usr/local/bin/fastqc2
② iu@bielinux[srp017156] whereis fastqc
fastqc: /usr/local/bin/fastqc
③ iu@bielinux[srp017156] whereis perl [ 2:42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
④ iu@bielinux[srp017156] whereis ls [ 2:42午後 ]
ls: /bin/ls /usr/share/man/man1/ls.1.gz
⑤ iu@bielinux[srp017156] whereis pwd [ 2:42午後 ]
pwd: /bin/pwd /usr/include/pwd.h /usr/share/man/man1/pwd.1.gz
⑥ iu@bielinux[srp017156] whereis whereis [ 2:42午後 ]
whereis: /usr/bin/whereis /usr/bin/X11/whereis /usr/share/man/man1/wherei
s.1.gz
iu@bielinux[srp017156] [ 2:42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と気になり始める。

```
File Edit View Search Terminal Help
iu@bielinux[srp017156] pwd
/home/iu/Documents/srp017156
① iu@bielinux[srp017156] whereis fastqc2
fastqc2: /usr/local/bin/fastqc2
② iu@bielinux[srp017156] whereis fastqc
fastqc: /usr/local/bin/fastqc
③ 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
④ iu@bielinux[srp017156] whereis ls
ls: /bin/ls /usr/share/man/man1/ls.1.gz
⑤ iu@bielinux[srp017156] whereis pwd
pwd: /bin/pwd /usr/include/pwd.h /usr/share/man/man1/pwd.1.gz
⑥ iu@bielinux[srp017156] whereis whereis
whereis: /usr/bin/whereis /usr/bin/X11/whereis /usr/share/man/man1/wherei
s.1.gz
iu@bielinux[srp017156]
```

[2:42午後]

[2:42午後]

[2:42午後]

[2:42午後]

[2:42午後]

[2:42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
perl /usr/share/man/man1/perl.1.gz
① iu@bielinux[srp017156] 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
② iu@bielinux[srp017156] env | grep PATH
DEFAULTS PATH=/usr/share/gconf/ubuntu.default.path
PATH=/usr/local/sbin:/usr/local/bin:/usr/sbin:/usr/bin:/sbin:/bin:/usr/ga
mes:/usr/local/games:/usr/lib/cd-hit:/usr/lib/cd-hit
XDG_SESSION_PATH=/org/freedesktop/DisplayManager/Session2
MANDATORY_PATH=/usr/share/gconf/ubuntu.mandatory.path
XDG_SEAT_PATH=/org/freedesktop/DisplayManager/Seat0
COMPIZ_BIN_PATH=/usr/bin/
③ iu@bielinux[srp017156] env | grep ^PATH
PATH=/usr/local/sbin:/usr/local/bin:/usr/sbin:/usr/bin:/sbin:/bin:/usr/ga
mes:/usr/local/games:/usr/lib/cd-hit:/usr/lib/cd-hit
iu@bielinux[srp017156]
```

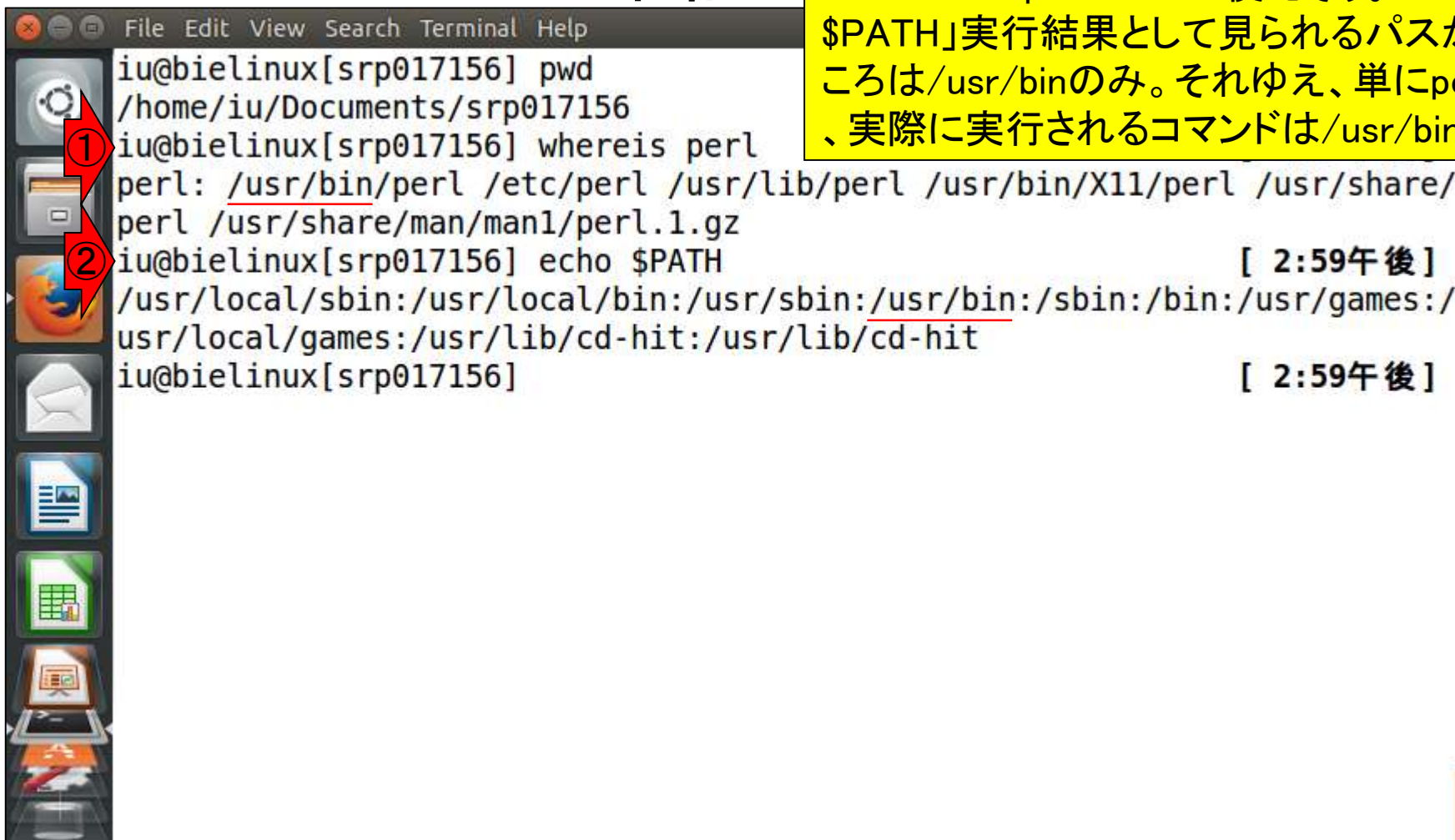

W9-9:パスの確認

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 [ 2:59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 [ 2:59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] [ 2:59午後 ]
```


W9-9:パスの確認

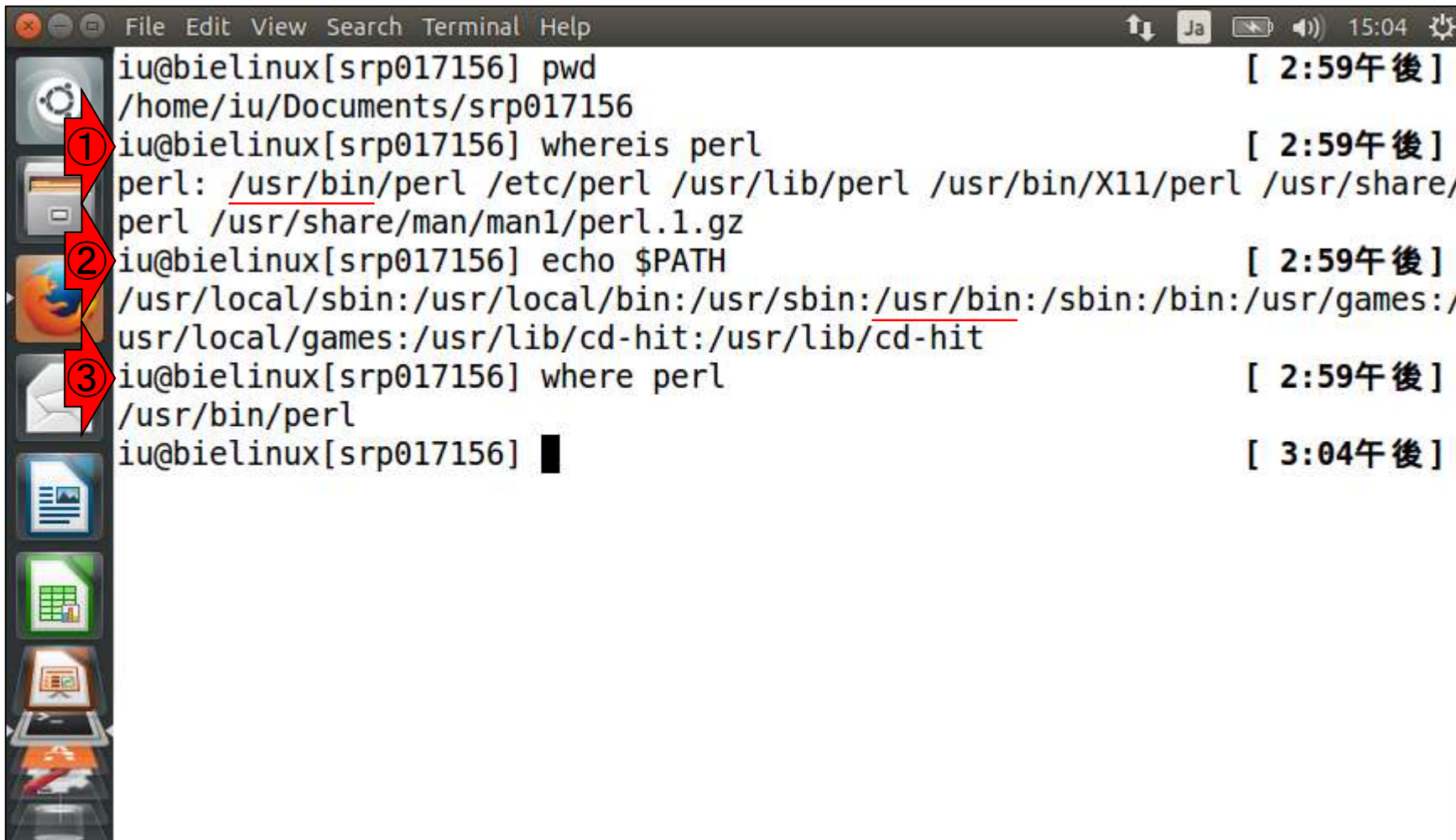
①「whereis perl」実行結果として、`/usr/bin/perl`、`/etc/perl`、`/usr/lib/perl`、`/usr/bin/X11/perl`、`/usr/share/perl`の5つが使いそう。このうち、②「echo \$PATH」実行結果として見られるパスが通っているところは`/usr/bin`のみ。それゆえ、単にperlと打ち込むと、実際に実行されるコマンドは`/usr/bin/perl`と解釈。



```
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 [ 2:59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] [ 2:59午後 ]
```

な~んて、面倒なことは普通やりませんw。
③whereコマンドでサクッと調べます

W9-9:パスの確認



A terminal window titled "Terminal" with a menu bar (File, Edit, View, Search, Terminal, Help) and system icons (up/down arrow, Ja, battery, volume, 15:04, settings). The terminal shows the following sequence of commands and outputs:

```
iu@bielinux[srp017156] pwd [ 2:59午後 ]
/home/iu/Documents/srp017156
① iu@bielinux[srp017156] whereis perl [ 2:59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
② iu@bielinux[srp017156] echo $PATH [ 2:59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
③ iu@bielinux[srp017156] where perl [ 2:59午後 ]
/usr/bin/perl
iu@bielinux[srp017156] [ 3:04午後 ]
```

Red arrows on the left side of the terminal point to the three numbered steps: ①, ②, and ③.

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、①/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②/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
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
iu@bielinux[srp017156] echo $PATH [ 2:59午後 ]
/usr/local/sbin:/usr/local/bin:/usr/sbin:/usr/bin:/sbin:/bin:/usr/games:/
usr/local/games:/usr/①/cd-hit:/usr/lib/cd-②
iu@bielinux[srp017156] where perl [ 2:59午後 ]
/usr/bin/perl
iu@bielinux[srp017156] █ [ 3:04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、赤下線の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

```
File Edit View Search Terminal Help
iu@bielinux[srp017156] cd
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.

iu@bielinux[iu] /usr/bin/perl -v | grep version [ 3:19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/bin/perl -v | grep "version" [ 3:19午後 ]
This is perl 5, version 18, subversion 2 (v5.18.2) built for x86_64-linux-gnu-
thread-multi
iu@bielinux[iu] [ 3:19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/usr/bin/perlと同じバージョンであり、古いものを使っていなくて安心。

```
File Edit View Search Terminal Help
① iu@bielinux[iu] 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[iu] /etc/perl -v | grep version [ 3:23午後 ]
zsh: permission denied: /etc/perl
③ iu@bielinux[iu] /usr/lib/perl -v | grep version [ 3:23午後 ]
zsh: permission denied: /usr/lib/perl
④ iu@bielinux[iu] /usr/bin/X11/perl -v | grep version [ 3:23午後 ]
This is perl 5, version 18, subversion 2 (v5.18.2) built for x86_64-linux-gnu-
thread-multi
⑤ iu@bielinux[iu] /usr/share/perl -v | grep version [ 3:23午後 ]
zsh: permission denied: /usr/share/perl
iu@bielinux[iu] [ 3:23午後 ]
```


W9-10:パスの設定

①「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のシンボリックリンクを/usr/local/binに貼った。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。

```
iu@bielinux[iu] pwd
/home/iu
iu@bielinux[iu] where fastqc
/usr/local/bin/fastqc
iu@bielinux[iu] where fastqc2
/usr/local/bin/fastqc2
iu@bielinux[iu] ls -lh /home/iu/Downloads/FastQC/fast*
-rwxr-xr-x 1 iu iu 14K 10月  9 11:59 /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
/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]
```

[3:38午後]

[3:38午後]

[3:38午後]

[3:38午後]



W9-10: パスの設定

②で現在通っているパスを表示。③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」としてもよい。④で、もう一度echo \$PATH。確かに追加されていることがわかる。追加前後の①と⑤のwhere実行結果の違いにも注目

```
File Edit View Search Terminal Help
① iu@bielinux[iu] where fastqc
   /usr/local/bin/fastqc
② iu@bielinux[iu] echo $PATH
   /usr/local/sbin:/usr/local/bin:/usr/sbin:/usr/bin:/sbin:/bin:/usr/games:/usr/local/games:/usr/lib/cd-hit:/usr/lib/cd-hit
④ iu@bielinux[iu] export PATH=PATH=$PATH:/home/iu/Downloads/FastQC [ 3:42午後 ]
④ iu@bielinux[iu] echo $PATH [ 3:42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
⑤ iu@bielinux[iu] where fastqc [ 3:42午後 ]
   /usr/local/bin/fastqc
   /home/iu/Downloads/FastQC/fastqc
iu@bielinux[iu] [ 3:42午後 ]
```

W9-10: パスの設定

①echo \$PATHの結果(優先順位)からも想像がつくように、②同じfastqcコマンドでバージョンの異なるものが共存する状態になってはいる。⑤単にfastqcと打つと、デフォルトで/usr/local/bin/fastqcに入っていたver. 0.10.1のプログラムが実行される。

```
iu@bielinux[iu] where fastqc
/usr/local/bin/fastqc
iu@bielinux[iu] echo $PATH
/usr/local/sbin:/usr/local/bin:/usr/sbin:/usr/bin:/sbin:/bin:/usr/games:/usr/local/games:/usr/lib/cd-hit:/usr/lib/cd-hit
iu@bielinux[iu] export PATH=$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:/home/iu/Downloads/FastQC
iu@bielinux[iu] where fastqc
/usr/local/bin/fastqc
/home/iu/Downloads/FastQC/fastqc
iu@bielinux[iu] /usr/local/bin/fastqc -v
FastQC v0.10.1
iu@bielinux[iu] /home/iu/Downloads/FastQC/fastqc -v
FastQC v0.11.4
iu@bielinux[iu] fastqc -v
FastQC v0.10.1
iu@bielinux[iu]
```

[3:42午後]

[3:42午後]

[3:42午後]

[3:42午後]

[3:42午後]

[3:42午後]

[3:53午後]

[3:53午後]

[3:56午後]

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)。①which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、②where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。③whereisは個人的にはよくわからないので非推奨

```
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 fastqc [ 4:05午後 ]
fastqc: /usr/local/bin/fastqc
iu@bielinux[iu] which fastqc [ 4:05午後 ]
/usr/local/bin/fastqc
iu@bielinux[iu] where fastqc [ 4:05午後 ]
/usr/local/bin/fastqc
/home/iu/Downloads/FastQC/fastqc
iu@bielinux[iu] [ 4:05午後 ]
```



W10-1: 環境設定

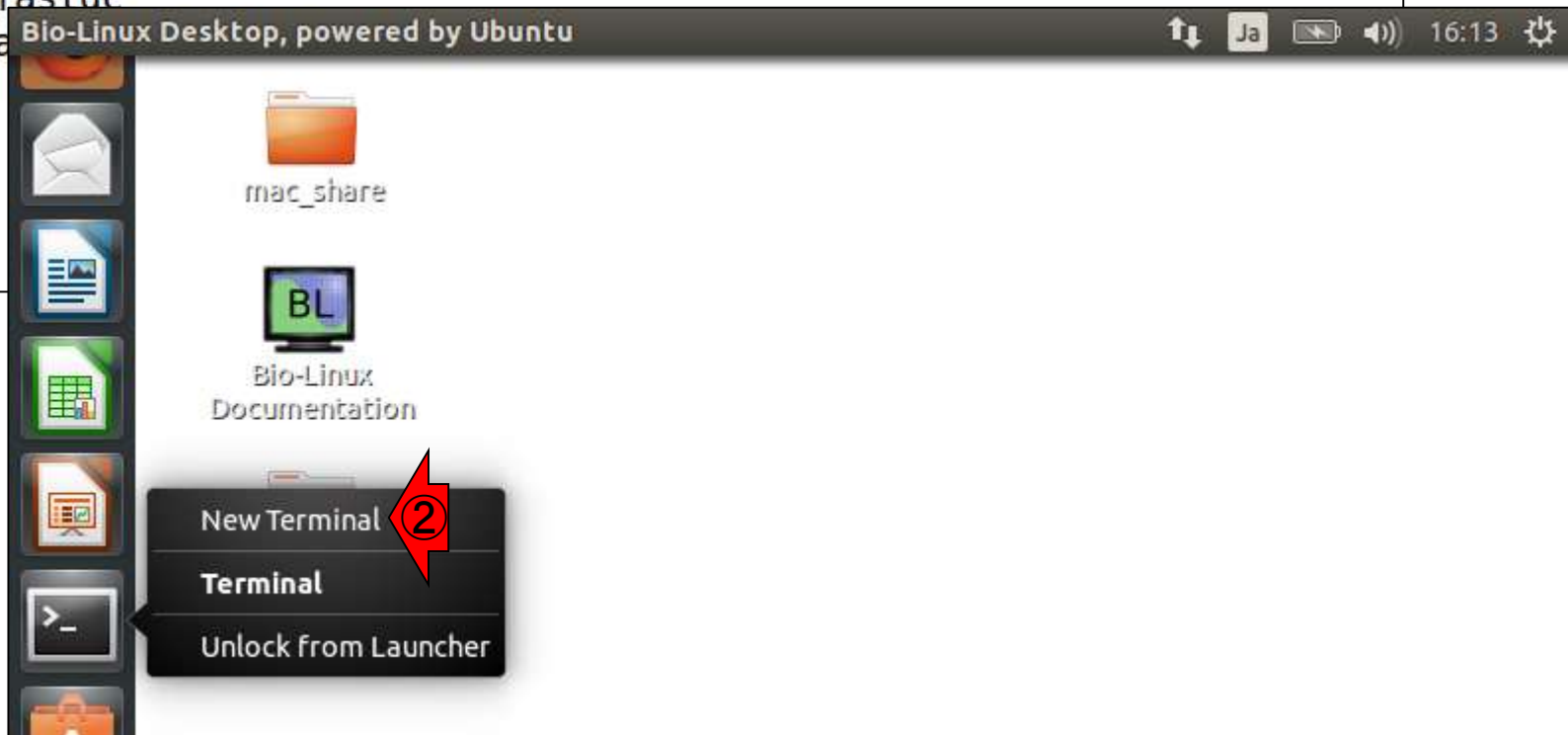
パスについても、共有フォルダの自動マウント設定(W4-5)で示すように、毎回設定をしなくて済むように環境設定ファイルの内容を書き換えるヒトのほうが多いかもしれないので、そのお話。①/home/iu/Downloads/FastQCのパスを追加したターミナル画面を一旦終了。②再度右クリックでNew Terminalを開く。

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

[4:05午後]

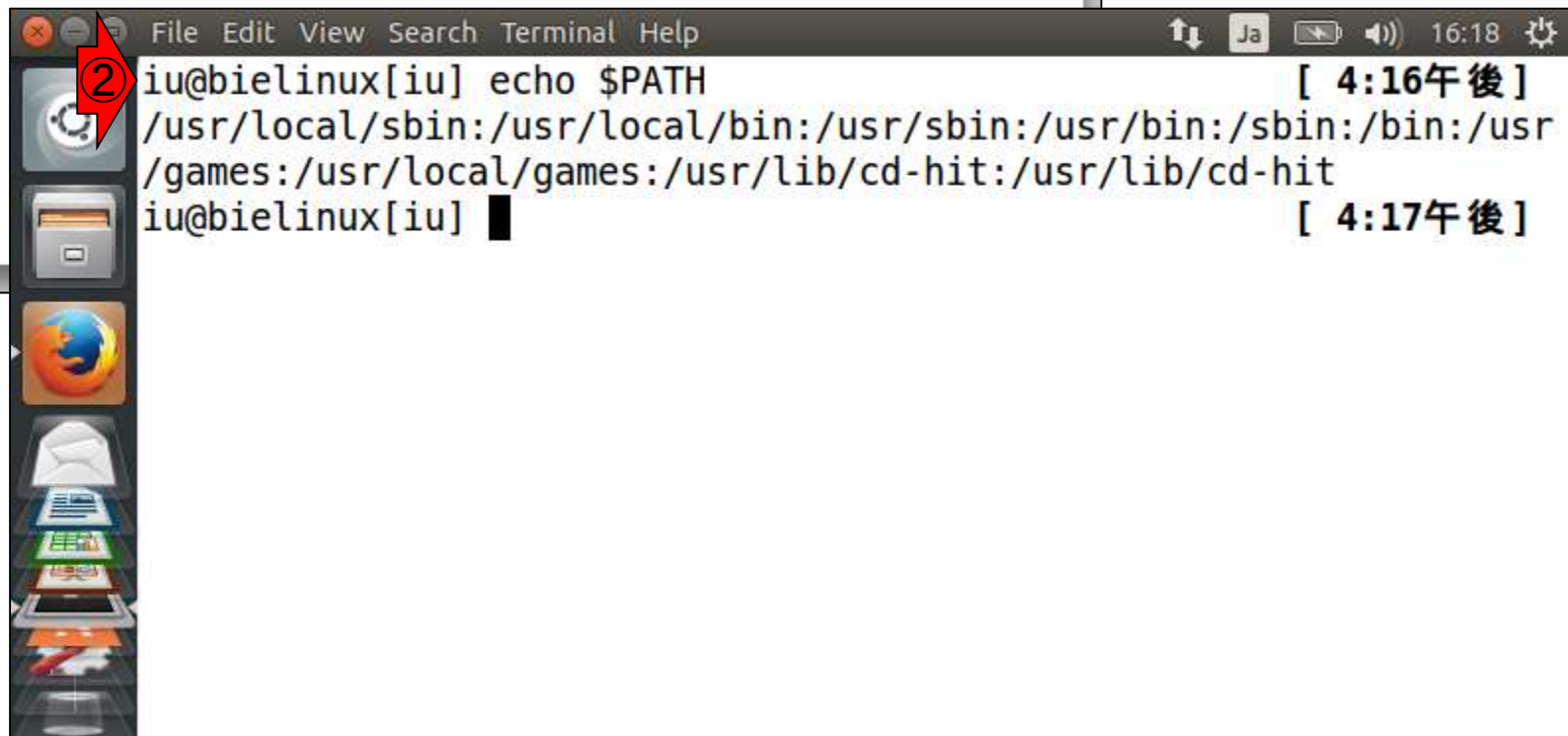
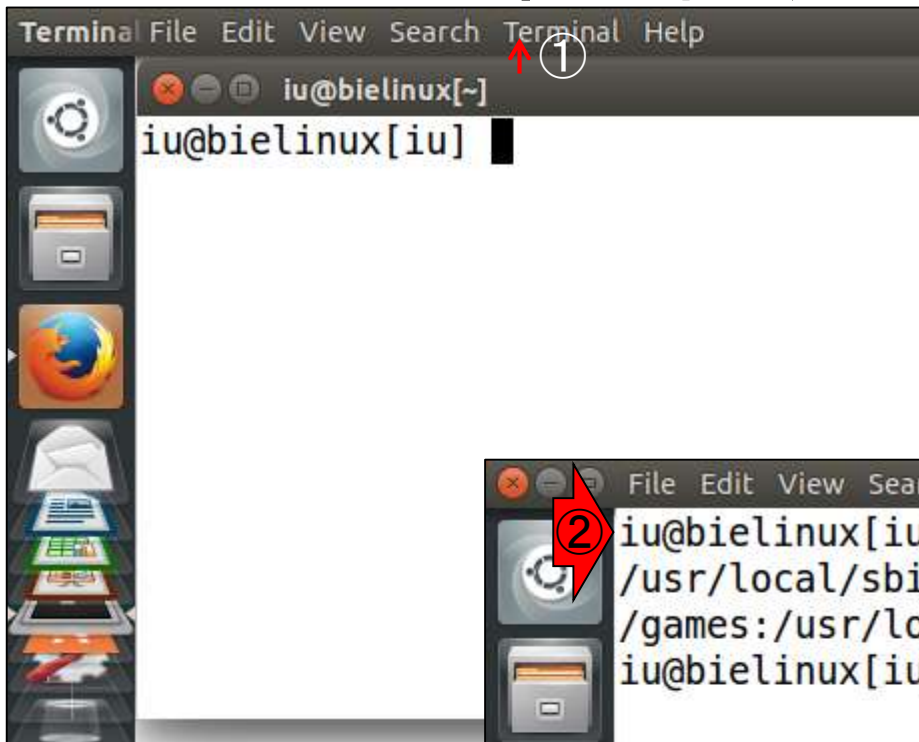
[4:05午後]

[4:05午後]

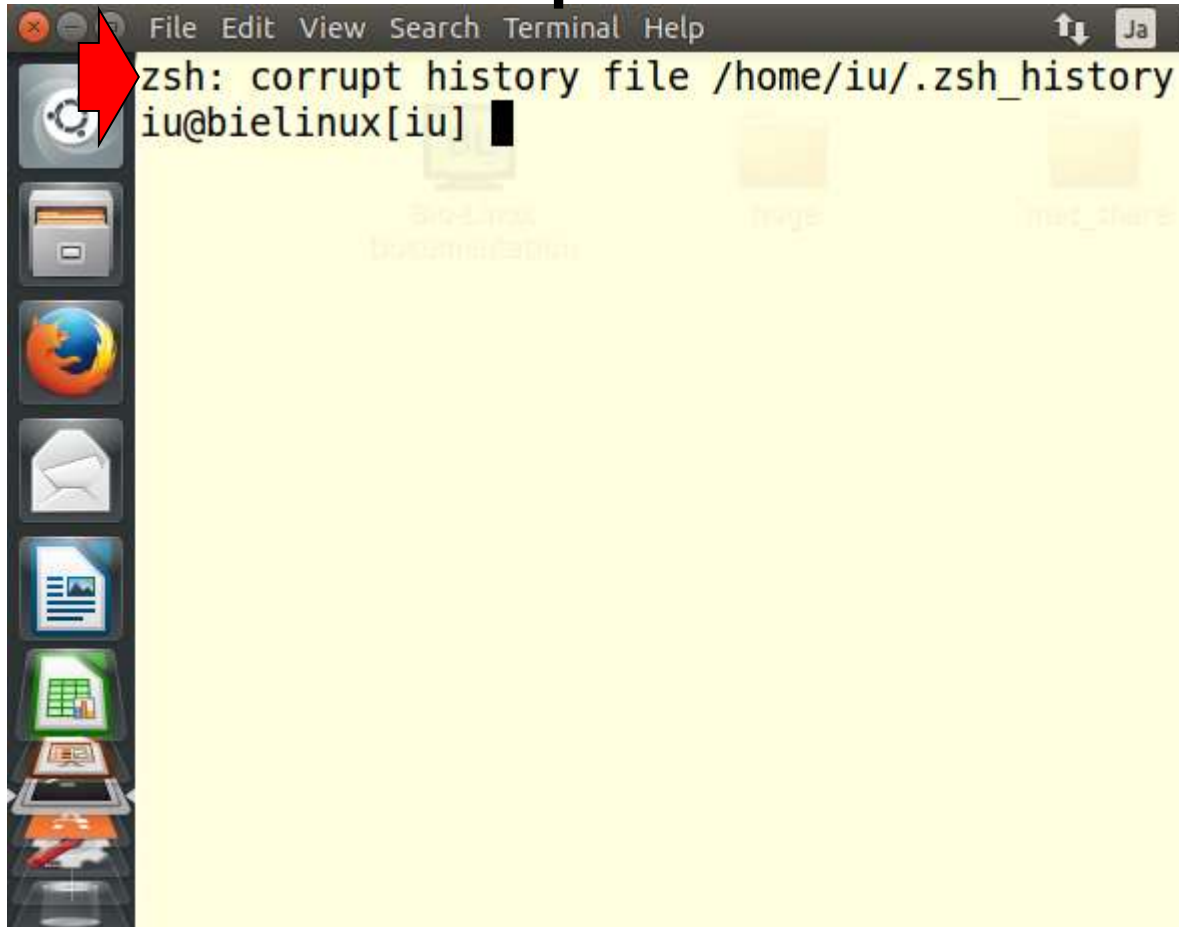


W10-1: 環境設定

おさらい(連載第3回のW9-2-3)。①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。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」なのだと学習する。

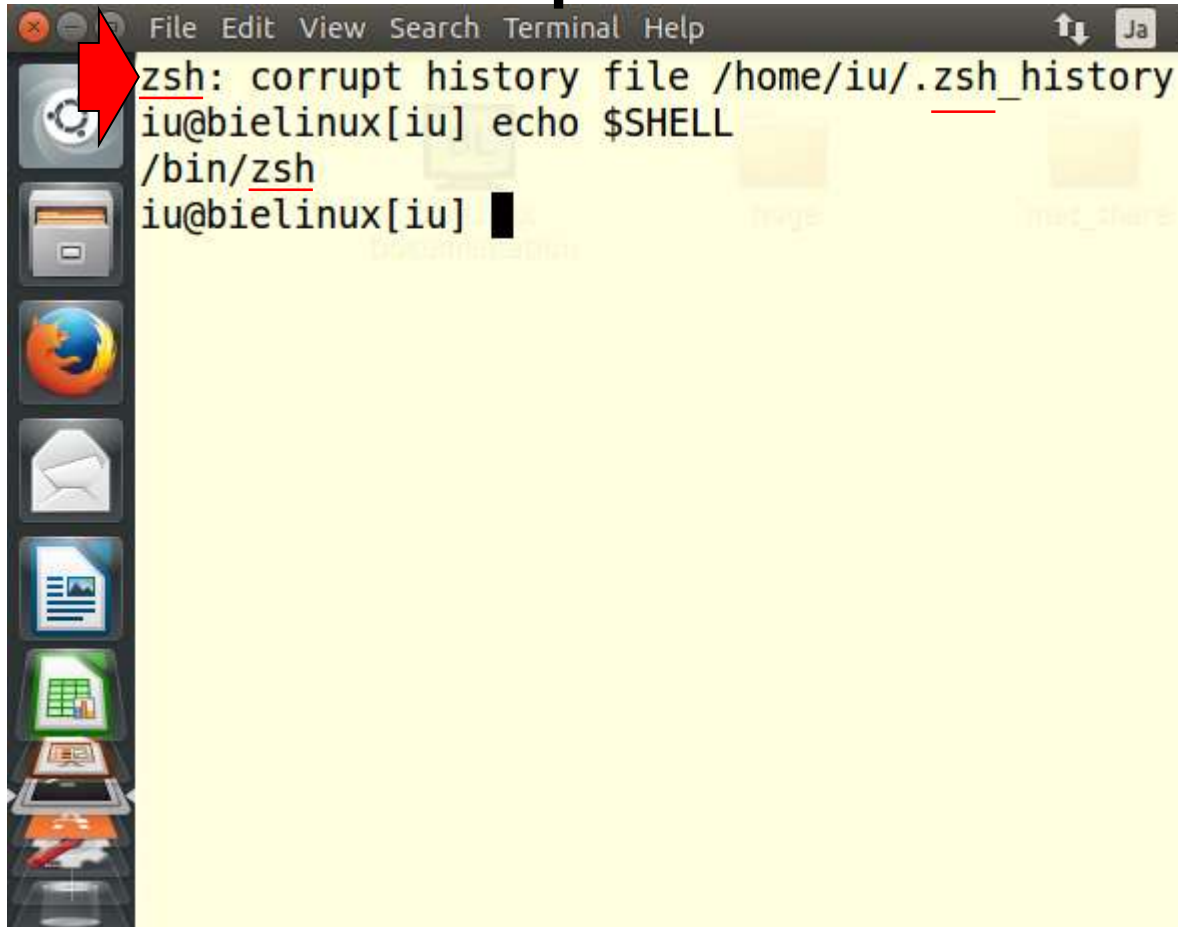


W10-2: Tips



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 terminal window with a dark title bar containing 'File Edit View Search Terminal Help' and a language indicator 'Ja'. The terminal text shows an error message 'zsh: corrupt history file /home/iu/.zsh_history', followed by the command 'iu@bielinux[iu] echo \$SHELL' and its output '/bin/zsh'. The prompt 'iu@bielinux[iu]' is shown again. 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

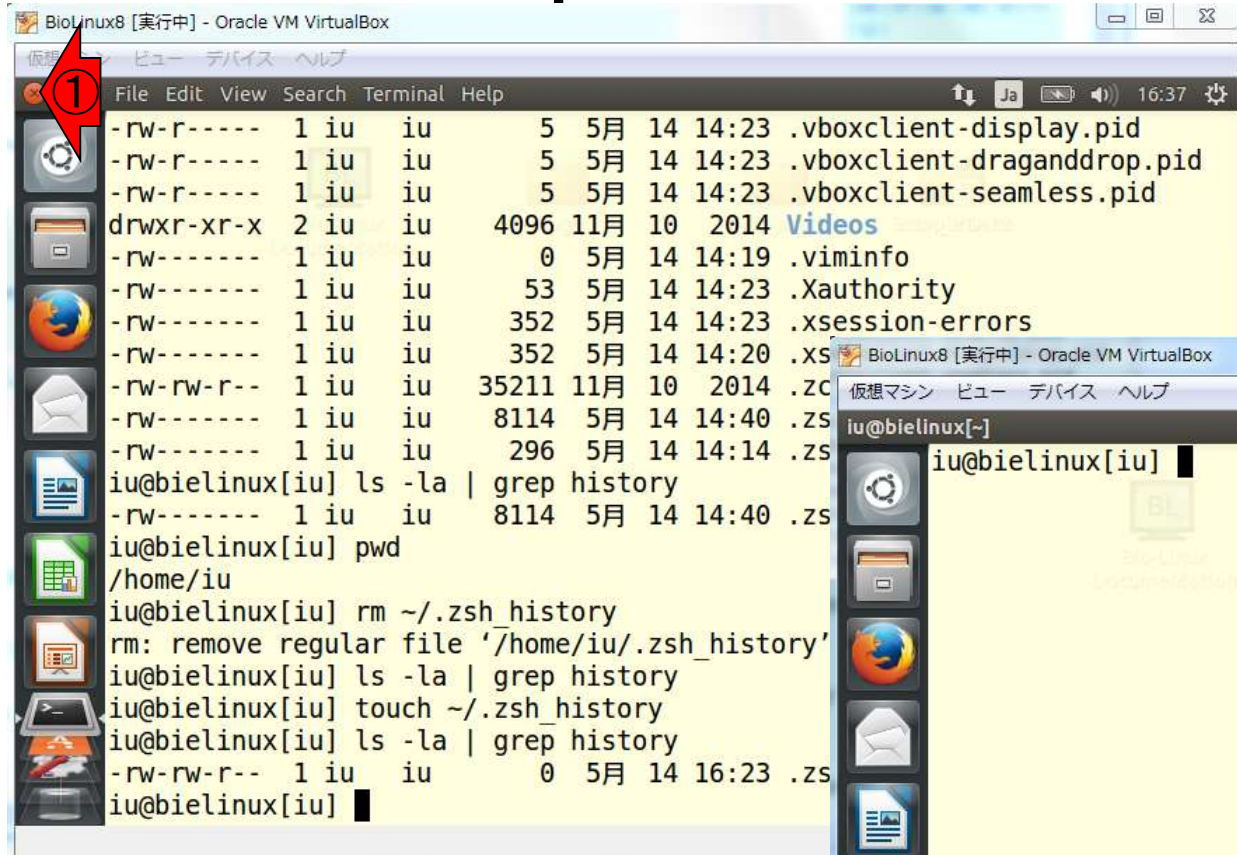
④のtouchコマンドはタイム・スタンプ変更を行うというもの。ここでやっている作業は、「現在時刻が最終更新日の .zsh_history という中身が空のファイル」の作成である。

```
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 ]
```



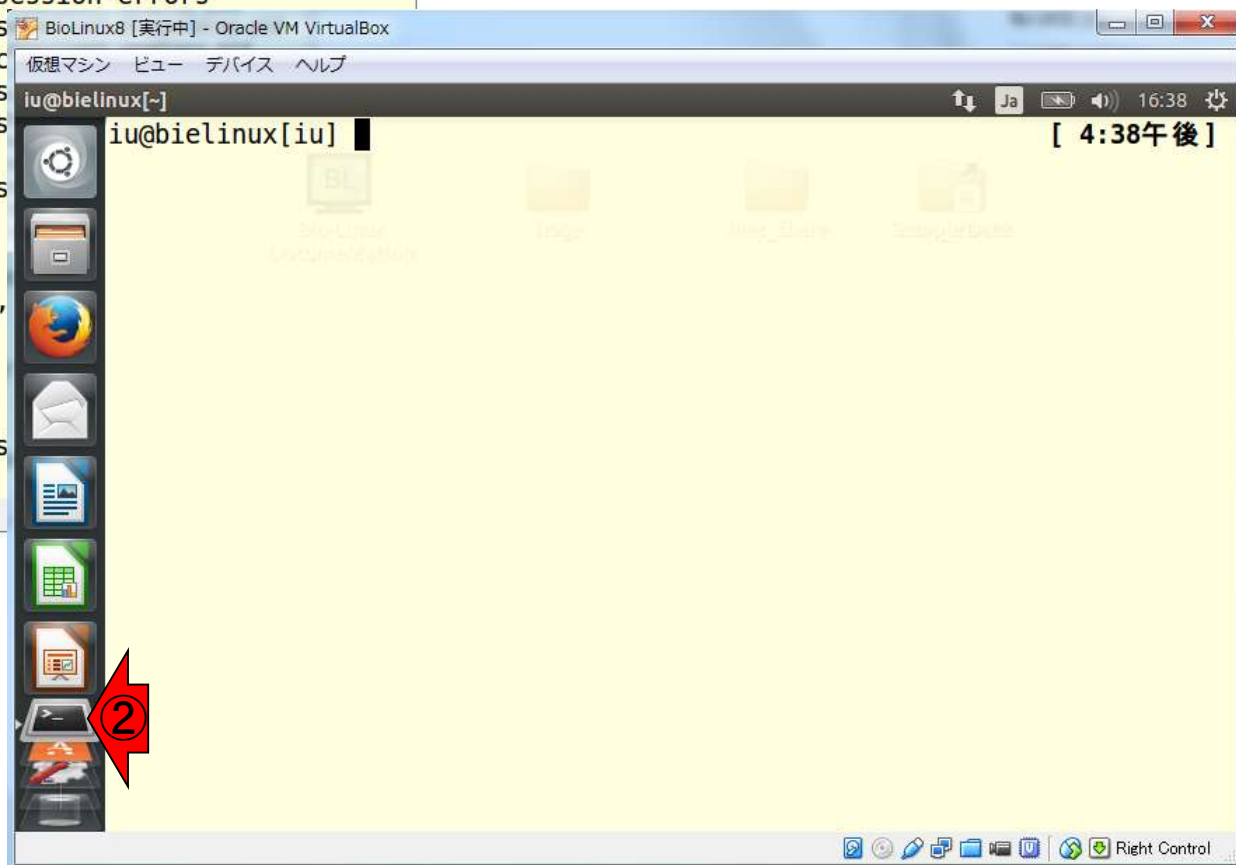
W10-2: Tips

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。



1

```
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]
```



2

```
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。②確かにホームディレクトリ上に.zshrcファイルが存在する。

```
File Edit View Search Terminal Help
iu@bielinux[iu] cd
iu@bielinux[iu] pwd
/home/iu
iu@bielinux[iu] echo $PATH
/usr/local/sbin:/usr/local/bin:/usr/sbin:/usr/bin:/sbin:/usr
/games:/usr/local/games:/usr/lib/cd-hit:/usr/lib/cd-hit
iu@bielinux[iu] echo $SHELL [ 4:25午後 ]
/bin/zsh
iu@bielinux[iu] ls -lha .*rc [ 4:25午後 ]
-rw-r--r-- 1 iu iu 3.6K 11月 17 13:44 .bashrc
-rw-r--r-- 1 iu iu 25 11月 17 14:07 .dmrc
-rw----- 1 iu iu 295 11月 17 13:44 .zshrc
iu@bielinux[iu] [ 4:25午後 ]
```



W10-3: 環境設定

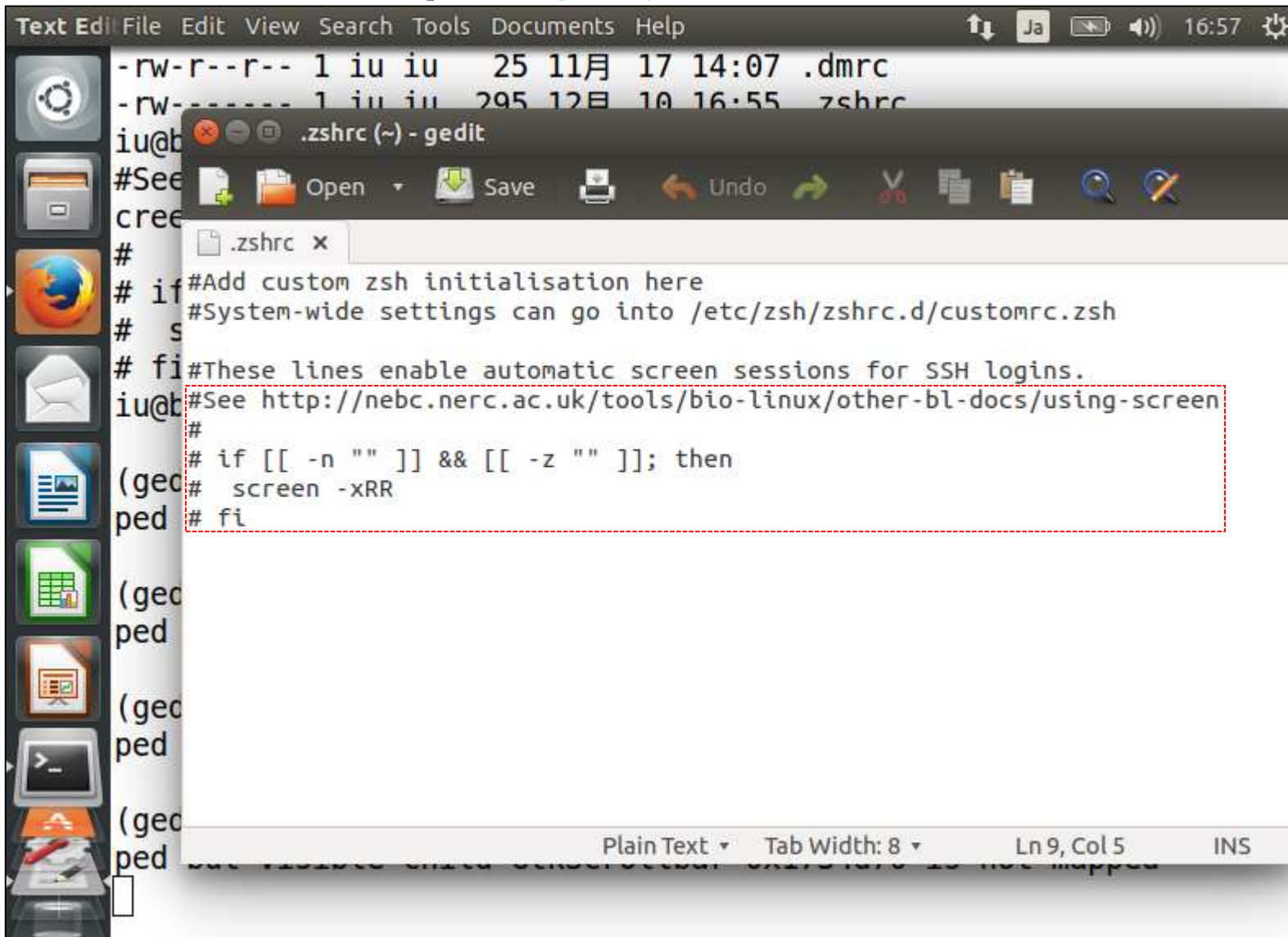
- ①編集前後の状況を把握すべく、.zshrcファイルの最後の5行分を表示。
- ②geditで.zshrcファイルを編集

```
iu@bielinux[iu] cd [ 4:28午後 ]
iu@bielinux[iu] pwd [ 4:28午後 ]
/home/iu
iu@bielinux[iu] echo $PATH [ 4:28午後 ]
/usr/local/sbin:/usr/local/bin:/usr/sbin:/usr/bin:/sbin:/usr
/games:/usr/local/games:/usr/lib/cd-hit:/usr/lib/cd-hit
iu@bielinux[iu] echo $SHELL [ 4:28午後 ]
/bin/zsh
iu@bielinux[iu] ls -lha .*rc [ 4:28午後 ]
-rw-r--r-- 1 iu iu 3.6K 11月 17 13:44 .bashrc
-rw-r--r-- 1 iu iu 25 11月 17 14:07 .dmrc
-rw----- 1 iu iu 295 11月 17 13:44 .zshrc
iu@bielinux[iu] tail -n 5 .zshrc [ 4:28午後 ]
#See http://nebc.nerc.ac.uk/tools/bio-linux/other-bl-docs/using-screen
#
# if [[ -n "" ]] && [[ -z "" ]]; then
# screen -xRR
# fi
iu@bielinux[iu] gedit .zshrc [ 4:29午後 ]
```

①

②

W10-3: 環境設定



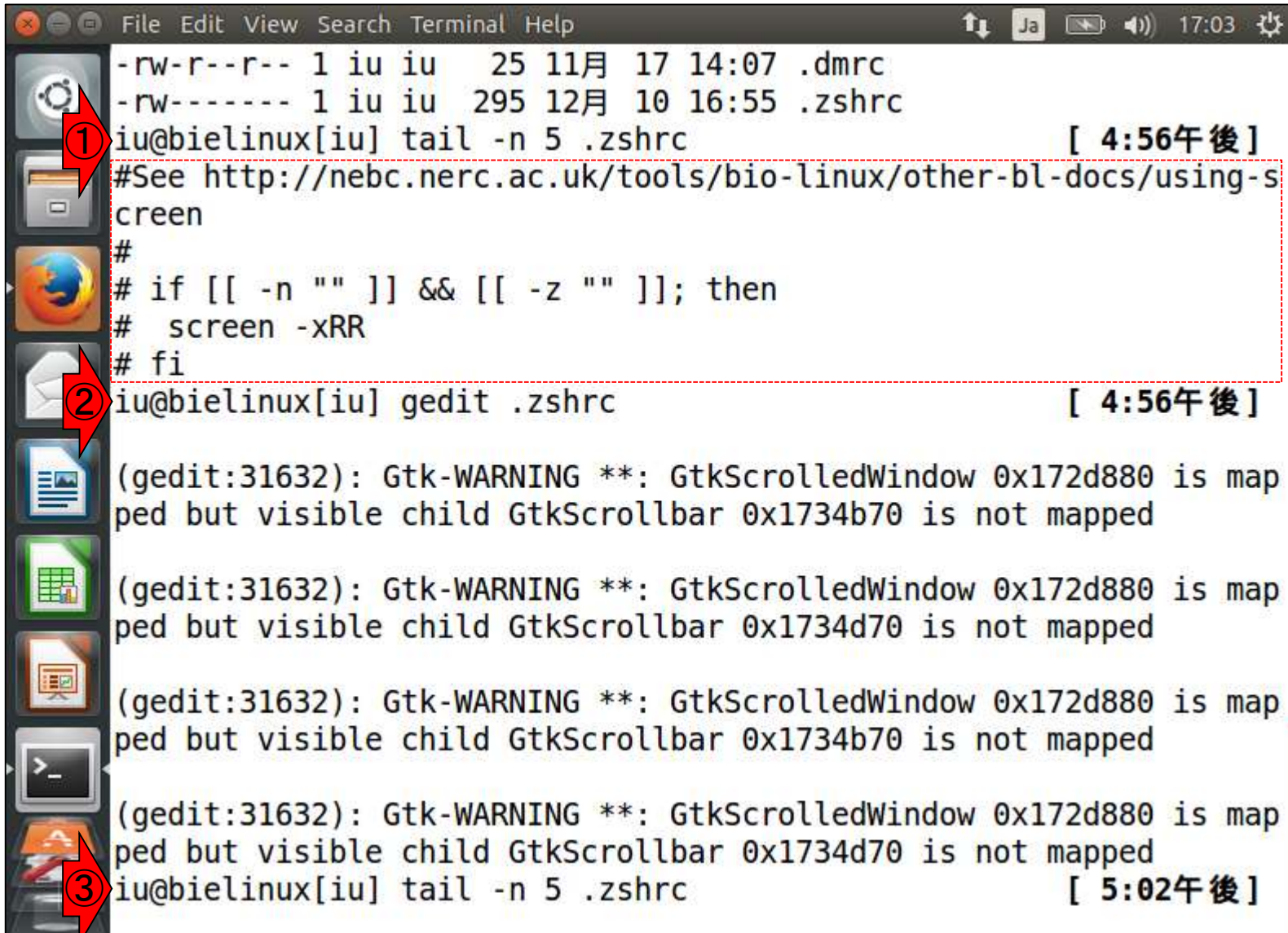
W10-3: 環境設定

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

```
-rw-r--r-- 1 iu iu  25 11月 17 14:07 .dmrc
-rw-r--r-- 1 iu iu 295 12月 10 16:55  .zshrc
iu@b
#See
cre
#
# if #Add custom zsh initialisation here
# s #System-wide settings can go into /etc/zsh/zshrc.d/customrc.zsh
# fi#These lines enable automatic screen sessions for SSH logins.
iu@ #See http://nebc.nerc.ac.uk/tools/bio-linux/other-bl-docs/using-screen
#
# if [[ -n "" ]] && [[ -z "" ]]; then
ged # screen -xRR
ped # fi
ged export PATH=$PATH:/home/iu/Downloads/FastQC
ped
ged
ped
ged
ped
```

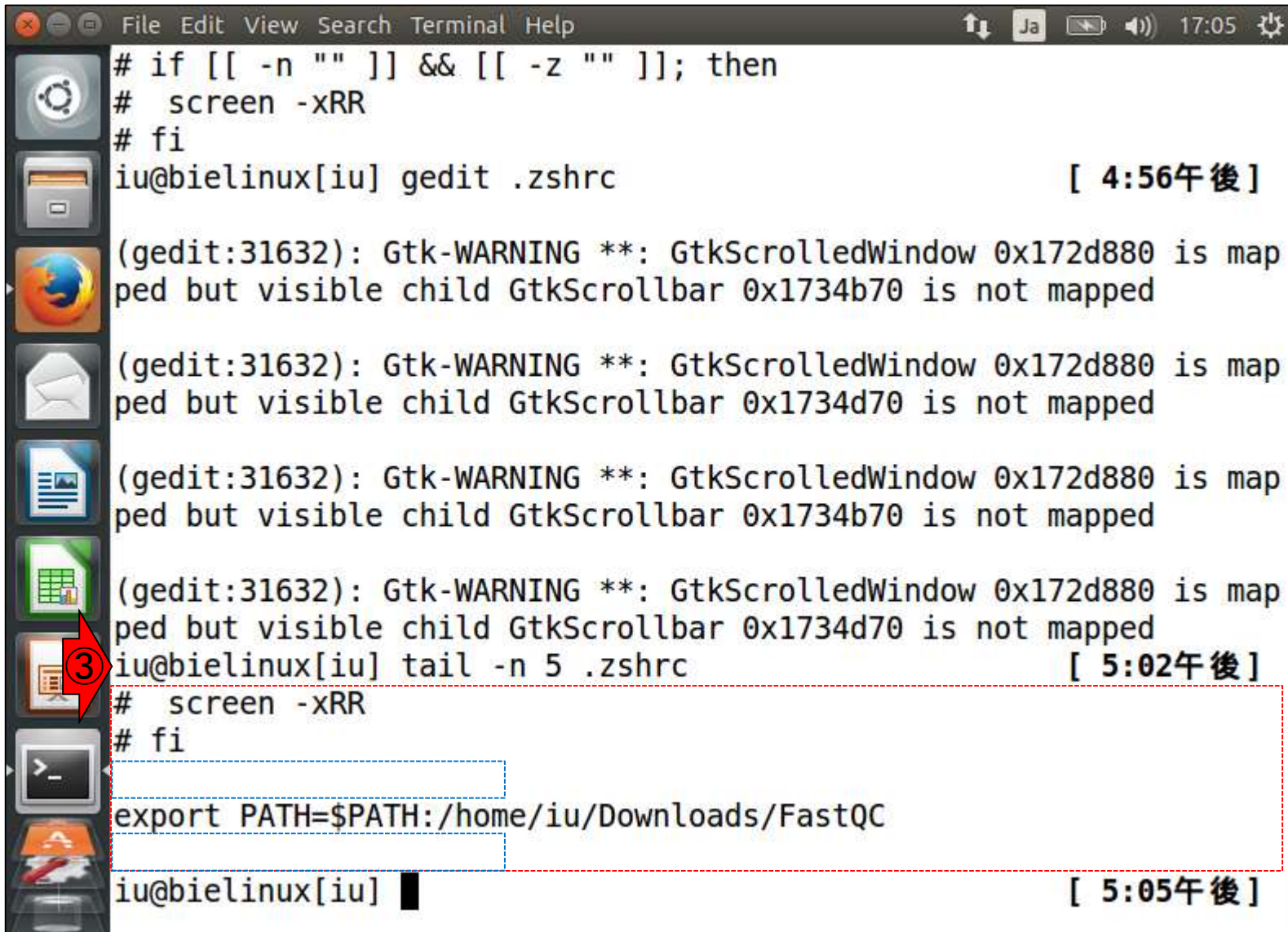
gedit終了直後の状態。③再びtailコマンドを実行し、最後の5行分を表示して確認する。

W10-3: 環境設定



```
File Edit View Search Terminal Help 17:03
-rw-r--r-- 1 iu iu 25 11月 17 14:07 .dmrc
-rw----- 1 iu iu 295 12月 10 16:55 .zshrc
iu@bielinux[iu] tail -n 5 .zshrc [ 4:56午後 ]
#See http://nebc.nerc.ac.uk/tools/bio-linux/other-bl-docs/using-s
screen
#
# if [[ -n "" ]] && [[ -z "" ]]; then
# screen -xRR
# fi
iu@bielinux[iu] gedit .zshrc [ 4:56午後 ]
(gedit:31632): Gtk-WARNING **: GtkScrolledWindow 0x172d880 is map
ped but visible child GtkScrollbar 0x1734b70 is not mapped
(gedit:31632): Gtk-WARNING **: GtkScrolledWindow 0x172d880 is map
ped but visible child GtkScrollbar 0x1734d70 is not mapped
(gedit:31632): Gtk-WARNING **: GtkScrolledWindow 0x172d880 is map
ped but visible child GtkScrollbar 0x1734b70 is not mapped
(gedit:31632): Gtk-WARNING **: GtkScrolledWindow 0x172d880 is map
ped but visible child GtkScrollbar 0x1734d70 is not mapped
iu@bielinux[iu] tail -n 5 .zshrc [ 5:02午後 ]
```

W10-3: 環境設定



The image shows a terminal window with a dark theme. The window title is "Terminal" and it contains a shell script. The script starts with a conditional check for empty strings, followed by a loop that runs "gedit .zshrc". The output shows several "Gtk-WARNING" messages. A red arrow points to the "tail" command, which is highlighted with a red dashed box. Below the "tail" command, the "export" command is shown, and the terminal prompt is visible.

```
File Edit View Search Terminal Help 17:05
# if [[ -n "" ]] && [[ -z "" ]]; then
#   screen -xRR
# fi
iu@bielinux[iu] gedit .zshrc [ 4:56午後 ]

(gedit:31632): Gtk-WARNING **: GtkScrolledWindow 0x172d880 is mapped but visible child GtkScrollbar 0x1734b70 is not mapped
(gedit:31632): Gtk-WARNING **: GtkScrolledWindow 0x172d880 is mapped but visible child GtkScrollbar 0x1734d70 is not mapped
(gedit:31632): Gtk-WARNING **: GtkScrolledWindow 0x172d880 is mapped but visible child GtkScrollbar 0x1734b70 is not mapped
(gedit:31632): Gtk-WARNING **: GtkScrolledWindow 0x172d880 is mapped but visible child GtkScrollbar 0x1734d70 is not mapped
iu@bielinux[iu] tail -n 5 .zshrc [ 5:02午後 ]
#   screen -xRR
# fi
export PATH=$PATH:/home/iu/Downloads/FastQC
iu@bielinux[iu] [ 5:05午後 ]
```


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。

```
File Edit View Search Terminal Help
(gedit:31632): Gtk-WARNING **: GtkScrolledWin
ped but visible child GtkScrollbar 0x1734d70
(gedit:31632): Gtk-WARNING **: GtkScrolledWin
ped but visible child GtkScrollbar 0x1734b70
(gedit:31632): Gtk-WARNING **: GtkScrolledWin
ped but visible child GtkScrollbar 0x1734d70
iu@bielinux[iu] tail -n 5 .zshrc
# screen -xRR
# fi
export PATH=$PATH:/home/iu/Downloads/FastQC
iu@bielinux[iu] echo $PATH [ 5:05午後 ]
/usr/local/sbin:/usr/local/bin:/usr/sbin:/usr/bin:/sbin:/usr
/games:/usr/local/games:/usr/lib/cd-hit:/usr/lib/cd-hit
iu@bielinux[iu] source .zshrc [ 5:10午後 ]
iu@bielinux[iu] echo $PATH [ 5:10午後 ]
/usr/local/sbin:/usr/local/bin:/usr/sbin:/usr/bin:/sbin:/usr
/games:/usr/local/games:/usr/lib/cd-hit:/usr/lib/cd-hit:/home/iu/
Downloads/FastQC
iu@bielinux[iu] [ 5:10午後 ]
```



[5:02午後]

[5:05午後]

[5:10午後]

[5:10午後]

[5:10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のパス設定が保存されており、一安心。

The image shows a sequence of terminal operations in a Linux environment. The first terminal window shows the following commands and output:

```
# fi
export PATH=$PATH:/home/iu/Downloads/FastQC
iu@bielinux[iu] echo $PATH [ 5:05午後 ]
/usr/local/sbin:/usr/local/bin:/usr/sbin:/usr/bin:/sbin:/usr
/games:/usr/local/games:/usr/lib/cd-hit:/usr/lib/cd-hit
iu@bielinux[iu] source .zshrc [ 5:10午後 ]
iu@bielinux[iu] echo $PATH [ 5:10午後 ]
/usr/local/sbin:/usr/local/bin:/usr/sbin:/usr/bin:/sbin:/usr
/games:/usr/local/games:/usr/lib/cd-hit:/usr/lib/cd-hit:/home/iu/
Downloads/FastQC
iu@bielinux[iu]
```

A second terminal window is then opened, and the following command and output are shown:

```
iu@bielinux[~]
iu@bielinux[iu] echo $PATH [ 5:21午後 ]
/usr/local/sbin:/usr/local/bin:/usr/sbin:/usr/bin:/sbin:/usr
/games:/usr/local/games:/usr/lib/cd-hit:/usr/lib/cd-hit:/home/iu/
Downloads/FastQC
iu@bielinux[iu] [ 5:21午後 ]
```

A context menu is shown over the terminal icon in the dock, with the following options:

- New Terminal
- Terminal
- Unlock from Launcher
- Quit

W10-6: Tips(rcとprofile)

```
iu@bielinux[iu] echo $PATH
/usr/local/sbin:/usr/local/bin:/usr/sbin:/usr/bin:/s
/games:/usr/local/games:/usr/lib/cd-hit:/usr/lib/cd-
Downloads/FastQC
iu@bielinux[iu] ls -la .*rc .*profile
-rw-r--r-- 1 iu iu 3637 11月 17 13:44 .bashrc
-rw-r--r-- 1 iu iu 25 11月 17 14:07 .dmrc
-rw-r--r-- 1 iu iu 675 11月 17 13:44 .profile
-rw----- 1 iu iu 341 12月 10 17:02 .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②FASTQ/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(「fastx_clipper -h」などでわかる)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 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](#)

FASTA/Q Clipper

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

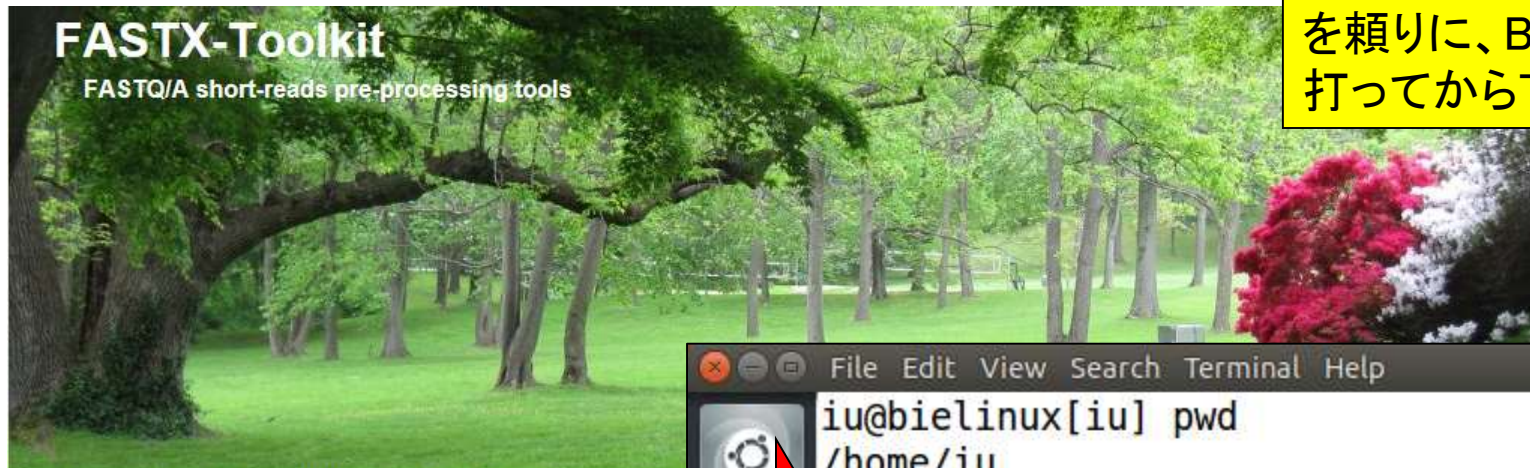
version 0.0.6
[-h] = This helpful help screen.
[-a ADAPTER] = ADAPTER string. default is CCTTAAGG (dummy adapter).
[-l N] = discard sequences shorter than N nucleotides. default is 5.
[-d N] = Keep the adapter and N bases after it.
         (using '-d 0' is the same as not using '-d' at all. which is 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

①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②fastxまで打ってからTabキーを押す。

FASTX-Toolkit

FASTQ/A short-reads pre-processing tools



Home | Download & Installation | Galaxy Usage | Command

Here you'll find a short description and examples of how to use

• [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](#)

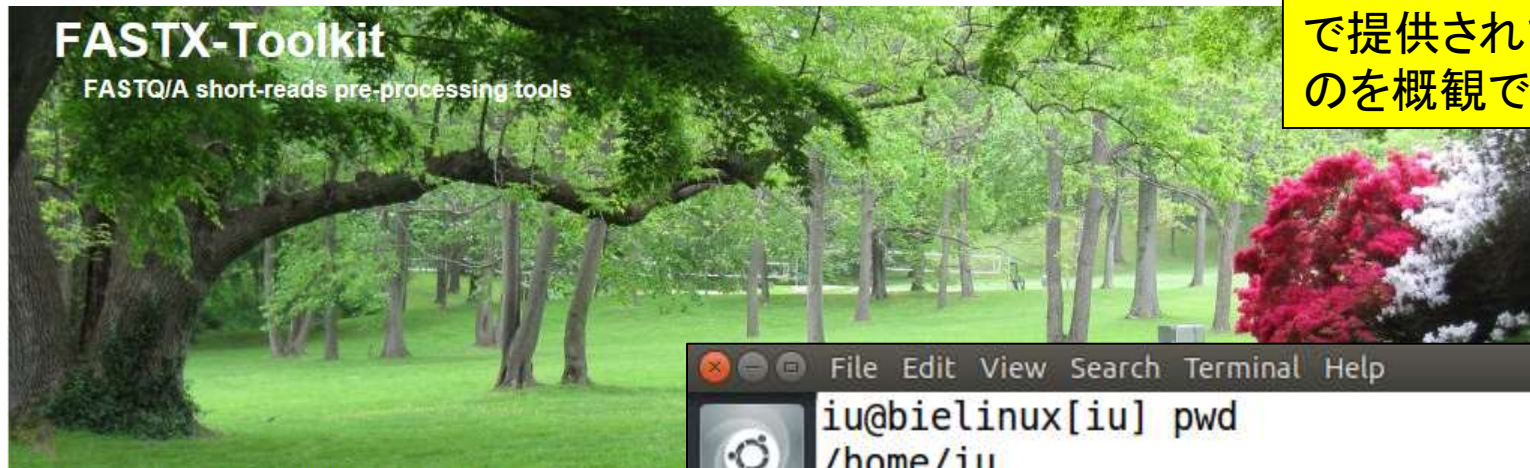
```
iu@bielinux[iu] pwd [ 5:32午後 ]
/home/iu
iu@bielinux[iu] fastx [ 5:32午後 ]
```

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



Home | Download & Installation | Galaxy Usage | Command

Here you'll find a short description and examples of how to use

- [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](#)

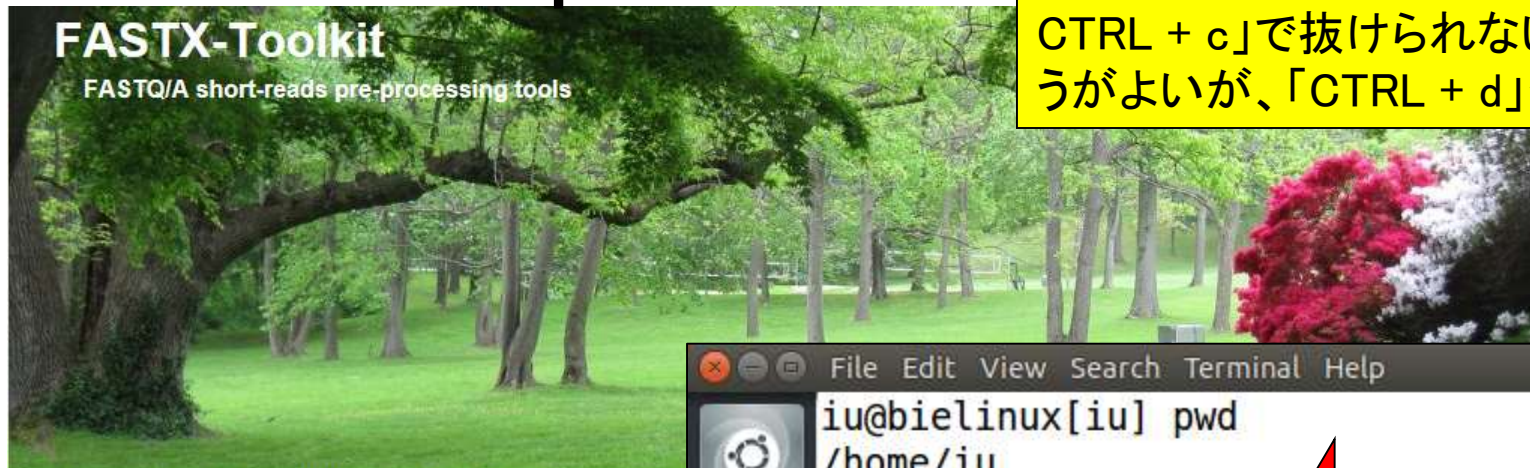
```
iu@bielinux[iu] pwd [ 5:33午後 ]
/home/iu
iu@bielinux[iu] fastx [ 5:33午後 ]
fastx
fastx36_t
fastx_artifacts_filter
fastx_barcode_splitter.pl
fastx_clipper
fastx_collapser
fastx-graph
fastx_nucleotide_distribution_graph.sh
fastx_nucleotide_distribution_line_graph.sh
fastx_quality_stats
fastx_renamer
fastx_reverse_complement
fastx_trimmer
fastx_uncollapser
```


W11-3: Tips

①間違っ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、多用しないほうがよいが、「CTRL + d」や「CTRL + z」もある。

FASTX-Toolkit

FASTQ/A short-reads pre-processing tools



Home | Download & Installation | Galaxy Usage | Command

Here you'll find a short description and examples of how to use

- [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](#)

```
iu@bielinux[iu] pwd [ 5:33午後 ]
/home/iu
iu@bielinux[iu] fastx [ 5:33午後 ]
# 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: █
```



W11-3: Tips

① 「CTRL + c」を押して無事復帰したところ。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



Home | Download & Installation | Galaxy Usage | Command

Here you'll find a short description and examples of how to use

- [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](#)

```
iu@bielinux[iu] pwd [ 5:33午後 ]
/home/iu
iu@bielinux[iu] fastx [ 5:33午後 ]
# 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

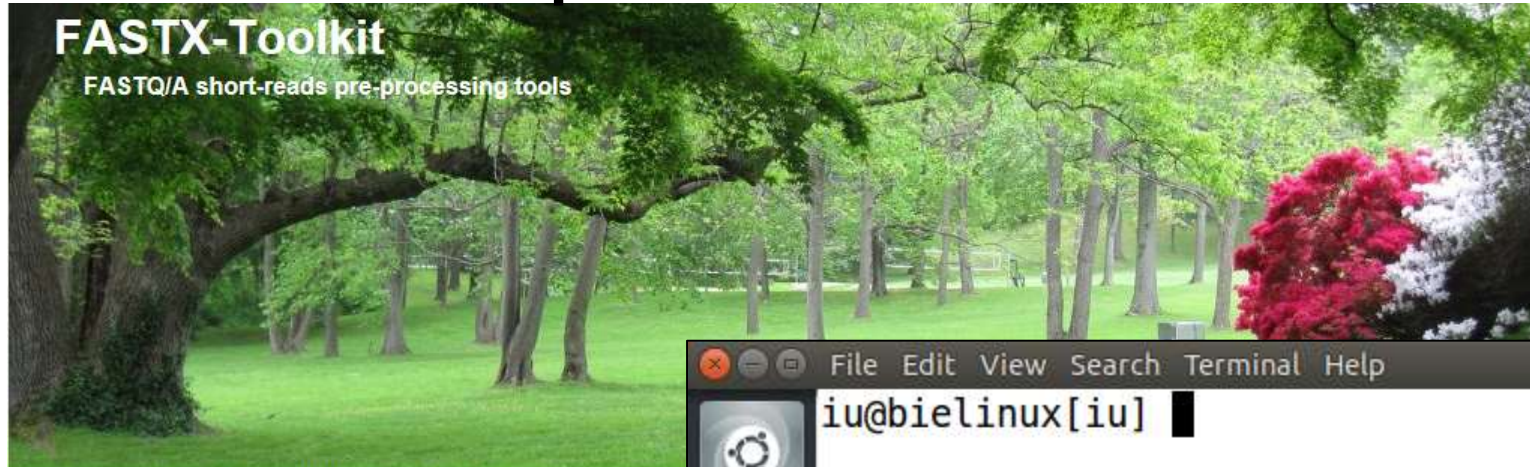
test sequence file name: ^C [ 5:38午後 ]
iu@bielinux[iu] █
```



W11-3: Tips

FASTX-Toolkit

FASTQ/A short-reads pre-processing tools



File Edit View Search Terminal Help

iu@bielinux[iu] [5:38午後]

System tray icons: Up/Down arrow, Ja, Network, Volume, 17:39, Settings.

Home | Download & Installation | Galaxy Usage | Command

Here you'll find a short description and examples of how to use

- [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

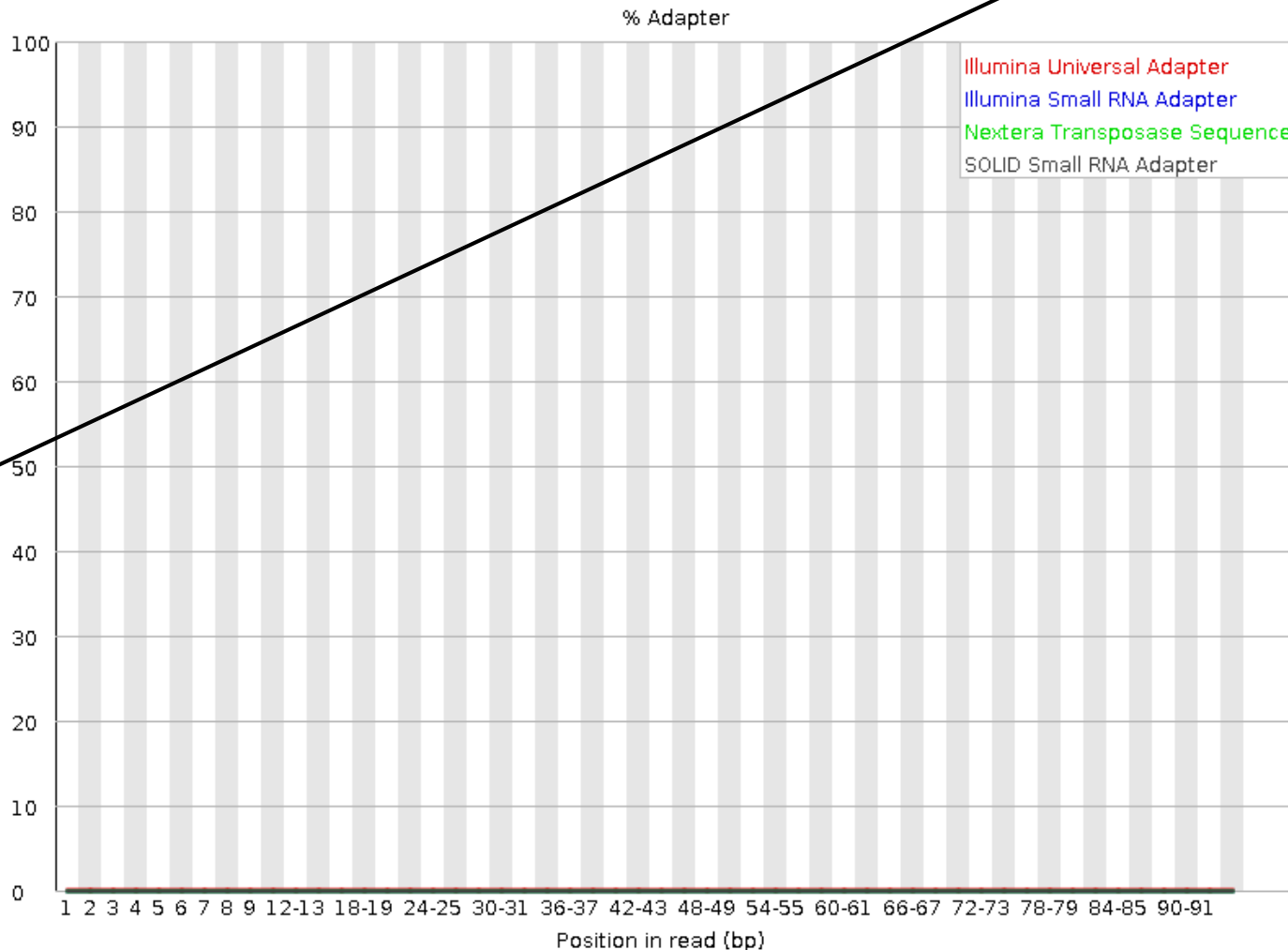
SRR616268sub_1.fastqのFastQC (ver 0.11.4)実行結果。Overrepresented sequencesで多少アダプター配列が含まれていることはわかるが、リードに占めるアダプター配列の割合はほぼ0であることがわかる。

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
- ✗ Kmer Content

✓ Adapter Content



①

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 IN
FILE] [-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 th
e adapter).
[-C] = Discard clipped sequences (i.e. - keep only sequences which did not containe
d the adapter).
[-k] = Report Adapter-Only sequences.
[-n] = keep sequences with unknown (N) nucleotides. default is to discard such sequ
ences.
[-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 I
NFILE] = FASTA/Q input file. default is STDIN.
[-o OUTFILE] = FASTA/Q output file. default is STDOUT.

iu@bielinux[iu]
```

[8:33午後]

W11-5: FASTX-toolkit

FASTX-Toolkit

FASTQ/A short-reads pre-processing tools



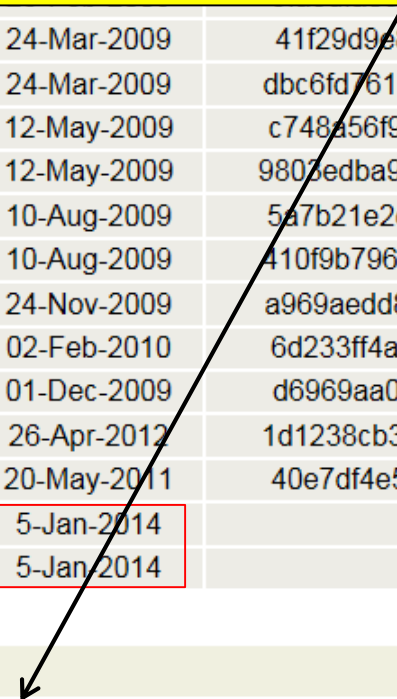
Download

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12月10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

連載第4回の図4は赤枠部分のコマンドをコピーで実行している。大まかには、①gzip圧縮ファイルの解凍、②解凍後のfastqファイルの行数をカウント、③fastq_clipperを用いたアダプター除去部分、④アダプター除去後のファイルに対してFastQC ver. 0.11.4を実行、⑤行数カウント。

```
iu@bielinux[iu] cd ~/Documents/srp017156
iu@bielinux[srp017156] pwd
/home/iu/Documents/srp017156
iu@bielinux[srp017156] wget -cq http://www.iu.a.u-tokyo.ac.jp/~kadota/book/JSLAB4_3_Linux.sh
iu@bielinux[srp017156] head -n 11 JSLAB4_3_Linux.sh
### 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]
```

[8:42午後]

[8:42午後]

W11-6: fastx_clipper

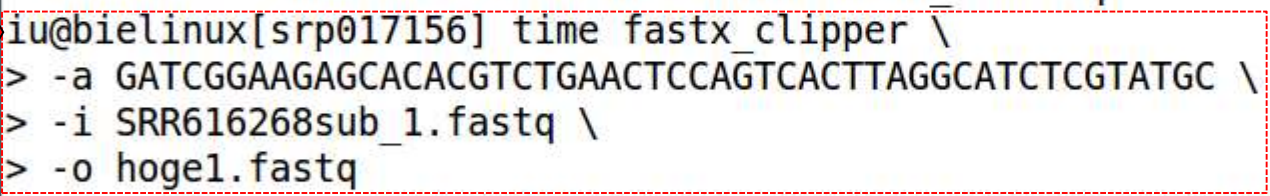
連載第4回の図4とほぼ同じ。①は図4の最初の状態にすべく圧縮しているだけ。fastx_clipper実行のための②非圧縮FASTQファイル(SRR616268sub_1.fastq)を作成し、③行数が400万行であることを確認している。

```
File Edit View Search Terminal Help
/home/iu/Documents/srp017156
iu@bielinux[srp017156] ls -lh *sub*
-rw-rw-r-- 1 iu iu 306M 12月  9 15:24 SRR616268sub_1.fastq
-rw-rw-r-- 1 iu iu  66M 12月  9 15:24 SRR616268sub_2.fastq.gz
iu@bielinux[srp017156] gzip SRR616268sub_1.fastq
iu@bielinux[srp017156] ls -lh *sub*
-rw-rw-r-- 1 iu iu 74M 12月  9 15:24 SRR616268sub_1.fastq.gz
-rw-rw-r-- 1 iu iu 66M 12月  9 15:24 SRR616268sub_2.fastq.gz
iu@bielinux[srp017156] time gunzip SRR616268sub_1.fastq.gz
gunzip SRR616268sub_1.fastq.gz 2.77s user 0.78s system 7% cpu 44.515 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 hogel.fastq
fastx_clipper -a GATCGGAAGAGCACACGTCTGAACTCCAGTCACTTAGGCATCTCGTATGC -i -o 101.0
ls user 0.65s system 96% cpu 1:45.48 total
iu@bielinux[srp017156] fastqc2 -q hogel.fastq
> --outdir=/home/iu/Desktop/mac_share
iu@bielinux[srp017156] wc hogel.fastq
3915596  7831192 286011042 hogel.fastq
iu@bielinux[srp017156]
```


W11-6: fastx_clipper

①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。赤下線で示すように2分弱かかる。

```
File Edit View Search Terminal Help
/home/iu/Documents/srp017156
iu@bielinux[srp017156] ls -lh *sub*
-rw-rw-r-- 1 iu iu 306M 12月 9 15:24 SRR616268sub_1.fastq.gz
-rw-rw-r-- 1 iu iu 66M 12月 9 15:24 SRR616268sub_2.fastq.gz
iu@bielinux[srp017156] gzip SRR616268sub_1.fastq
iu@bielinux[srp017156] ls -lh *sub*
-rw-rw-r-- 1 iu iu 74M 12月 9 15:24 SRR616268sub_1.fastq.gz
-rw-rw-r-- 1 iu iu 66M 12月 9 15:24 SRR616268sub_2.fastq.gz
iu@bielinux[srp017156] time gunzip SRR616268sub_1.fastq.gz [ 9:00午後 ]
gunzip SRR616268sub_1.fastq.gz 2.77s user 0.78s system 7% cpu 44.515 total
iu@bielinux[srp017156] wc SRR616268sub_1.fastq [ 9:01午後 ]
4000000 8000000 320760716 SRR616268sub_1.fastq
iu@bielinux[srp017156] time fastx_clipper \ [ 9:01午後 ]
> -a GATCGGAAGAGCACACGTCTGAACTCCAGTCACTTAGGCATCTCGTATGC \
> -i SRR616268sub_1.fastq \
> -o hoge1.fastq
fastx_clipper -a GATCGGAAGAGCACACGTCTGAACTCCAGTCACTTAGGCATCTCGTATGC -i -o 101.0
ls user 0.65s system 96% cpu 1:45.48 total
iu@bielinux[srp017156] fastqc2 -q hoge1.fastq \ [ 9:03午後 ]
> --outdir=/home/iu/Desktop/mac_share
iu@bielinux[srp017156] wc hoge1.fastq [ 9:03午後 ]
3915596 7831192 286011042 hoge1.fastq
iu@bielinux[srp017156] █ [ 9:03午後 ]
```



W11-6: fastx_clipper

①hoge1.fastqを入力としてFastQC (ver. 0.11.4)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htmlファイルが作成される。

```
File Edit View Search Terminal Help
/home/iu/Documents/srp017156
iu@bielinux[srp017156] ls -lh *sub*
-rw-rw-r-- 1 iu iu 306M 12月  9 15:24 SRR616268sub_1.fastq
-rw-rw-r-- 1 iu iu  66M 12月  9 15:24 SRR616268sub_2.fastq.gz
iu@bielinux[srp017156] gzip SRR616268sub_1.fastq
iu@bielinux[srp017156] ls -lh *sub*
-rw-rw-r-- 1 iu iu 74M 12月  9 15:24 SRR616268sub_1.fastq.gz
-rw-rw-r-- 1 iu iu 66M 12月  9 15:24 SRR616268sub_2.fastq.gz
iu@bielinux[srp017156] time gunzip SRR616268sub_1.fastq.gz
gunzip SRR616268sub_1.fastq.gz 2.77s user 0.78s system 7% cpu 44.515 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 101.0
ls user 0.65s system 96% cpu 1:45.48 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

```
File Edit View Search Terminal Help
/home/iu/Documents/srp017156
iu@bielinux[srp017156] ls -lh *sub*
-rw-rw-r-- 1 iu iu 306M 12月  9 15:24 SRR616268sub_1.fastq
-rw-rw-r-- 1 iu iu  66M 12月  9 15:24 SRR616268sub_2.fastq.gz
iu@bielinux[srp017156] gzip SRR616268sub_1.fastq
iu@bielinux[srp017156] ls -lh *sub*
-rw-rw-r-- 1 iu iu 74M 12月  9 15:24 SRR616268sub_1.fastq.gz
-rw-rw-r-- 1 iu iu 66M 12月  9 15:24 SRR616268sub_2.fastq.gz
iu@bielinux[srp017156] time gunzip SRR616268sub_1.fastq.gz
gunzip SRR616268sub_1.fastq.gz 2.77s user 0.78s system 7% cpu 44.515 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 101.0
ls user 0.65s system 96% cpu 1:45.48 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] █
```



W11-7: FastQC

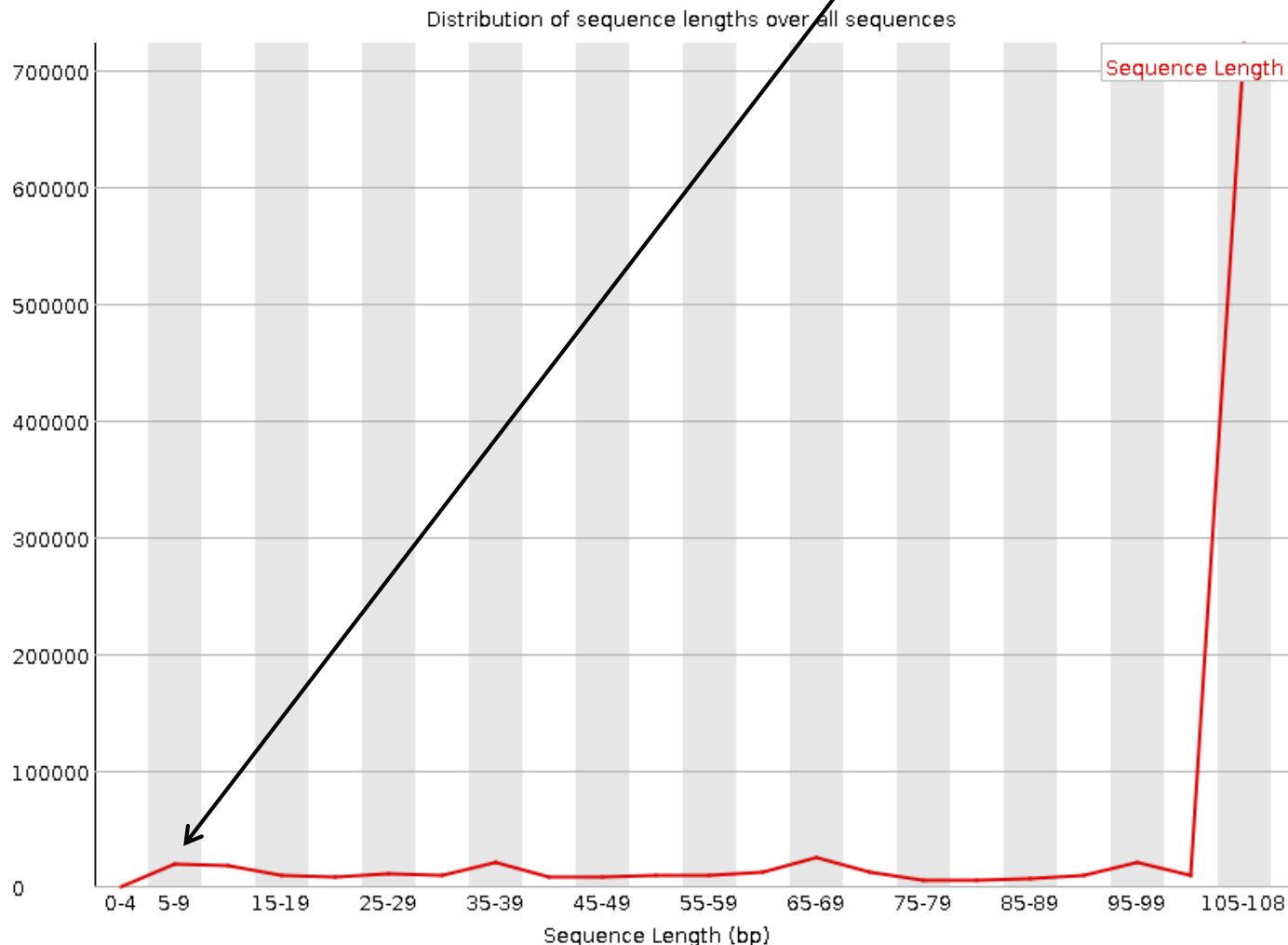
hoge1.fastqを入力としてFastQC (ver. 0.11.4)を実行した結果。①配列長分布の項目を表示。50 bpのアダプター配列除去後で107 bpから5 bpまでトリムされるリードがあるのでやはり変

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

①-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。②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[srp017156] pwd
/home/iu/Documents/srp017156
iu@bielinux[srp017156] time fastx_clipper \
> -a GATCGGAAGAGCACACGTCTGAACTCCAGTCACTTAGGCATCTCGTATGC \
> -i SRR616268sub_1.fastq \
> -c \
> -o hoge2.fastq
fastx_clipper -a GATCGGAAGAGCACACGTCTGAACTCCAGTCACTTAGGCATCTCGTATGC -i -c -o 98.77s user 0.58s system 99% cpu 1:39.75 total
iu@bielinux[srp017156] wc hoge2.fastq
1021860 2043720 53961976 hoge2.fastq
iu@bielinux[srp017156]
```

[10:45午前]

[10:47午前]

[10:48午前]

W11-8-2: 動作確認

① -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。
②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[srp017156] pwd
/home/iu/Documents/srp017156
iu@bielinux[srp017156] time fastx_clipper \
> -a GATCGGAAGAGCACACGTCTGAACTCCAGTCACTTAGGCATCTCGTATGC \
> -i SRR616268sub_1.fastq \
> -C \
> -o hoge3.fastq
fastx_clipper -a GATCGGAAGAGCACACGTCTGAACTCCAGTCACTTAGGCATCTCGTATGC -i -C -o 101
.27s user 0.64s system 94% cpu 1:48.41 total
iu@bielinux[srp017156] wc hoge3.fastq
2893736 5787472 232049066 hoge3.fastq
iu@bielinux[srp017156] █
```

[10:51午前]
[10:53午前]
[10:53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①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。②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[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 97.05s user 0.34s system 98% cpu 1:38.89 total
iu@bielinux[srp017156] wc hoge4.fastq
34052 68104 2730404 hoge4.fastq
iu@bielinux[srp017156]
```

[10:57午前]

[10:57午前]

W11-8-4: 動作確認

```
iu@bielinux[srp017156] pwd
/home/iu/Documents/srp017156
iu@bielinux[srp017156] time fastx_clipper \
> -a GATCGGAAGAGCACACGTCTGAACTCCAGTCACTTAGGC
> -i SRR616268sub_1.fastq \
> -n \
> -o hoge5.fastq
fastx_clipper -a GATCGGAAGAGCACACGTCTGAACTCC
.91s user 0.64s system 96% cpu 1:46.04 total
iu@bielinux[srp017156] wc hoge5.fastq
3927924 7855848 286951162 hoge5.fastq
iu@bielinux[srp017156]
```

①W11-5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、②-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[srp017156] pwd
/home/iu/Documents/srp017156
iu@bielinux[srp017156] time fastx_clipper \
> -a GATCGGAAGAGCACACGTCTGAACTCCAGTCACTTAGGC
> -i SRR616268sub_1.fastq \
> -n -l 0 \
> -o hoge6.fastq
fastx_clipper -a GATCGGAAGAGCACACGTCTGAACTCC
.20s user 0.73s system 93% cpu 1:49.43 total
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、①赤下線部分の-l 0を追加して実行。②3,965,948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①50 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。②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[srp017156] pwd
/home/iu/Documents/srp017156
iu@bielinux[srp017156] time fastx clipper \
> -a GATCGGAAGAGCACACGTCTGAACTCCAGTCACTTAGGC
> -i SRR616268sub_1.fastq \
> -c -M 50 \
> -o hoge7.fastq
fastx_clipper -a GATCGGAAGAGCACACGTCTGAACTCC
49s user 0.24s system 99% cpu 1:37.96 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]
```

[11:09午前]

[11:09午前]

[11:09午前]

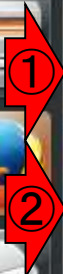
W11-9: 動作確認

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
4000000 8000000 320760716 SRR616268sub_1.fastq
iu@bielinux[srp017156] grep -c GATCGGAAGAGCACACGTCTGAACTCCAGTCACTTAGGCATCTCGTATGC
SRR616268sub_1.fastq
2812
iu@bielinux[srp017156] grep -c ^GATCGGAAGAGCACACGTCTGAACTCCAGTCACTTAGGCATCTCGTATGC
SRR616268sub_1.fastq
2415
iu@bielinux[srp017156] █
```

[11:15午前]

[11:15午前]



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
① iu@bielinux[srp017156] grep -c GATCGGAAGAGCACACGTCTGAACTCCAGTCACTTAGGCATCTCGTATGC
SRR616268sub_1.fastq
2812
② iu@bielinux[srp017156] grep -c ^GATCGGAAGAGCACACGTCTGAACTCCAGTCACTTAGGCATCTCGTATGC
SRR616268sub_1.fastq
2415
③ iu@bielinux[srp017156] grep ^GATCGGAAGAGCACACGTCTGAACTCCAGTCACTTAGGCATCTCGTATGC SR
R616268sub_1.fastq | head -n 3
GATCGGAAGAGCACACGTCTGAACTCCAGTCACTTAGGCATCTCGTATGCCGTCTTCTGCTTGAAAAGAATCTGAGCAAACC
CCCTGGTTGTCGGGGCACGGAATAC
GATCGGAAGAGCACACGTCTGAACTCCAGTCACTTAGGCATCTCGTATGCCGTCTTCTGCTTGAAAAAAAAAATAAAAGAAA
AAGACAACCTCAAGCATGGGGGATGG
GATCGGAAGAGCACACGTCTGAACTCCAGTCACTTAGGCATCTCGTATGCCGTCTTCTGCTTGAAAAAACAAAAACCTTAG
TGAACAAACATCGGTCACGGGTCCG
iu@bielinux[srp017156] █
```

[11:17午前]

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 [11:15午前]
/home/iu/Documents/srp017156
iu@bielinux[srp017156] wc SRR616268sub_1.fastq [11:15午前]
4000000 8000000 320760716 SRR616268sub_1.fastq
① iu@bielinux[srp017156] grep -c GATCGGAAGAGCACACGTCTGAACTCCAGTCACTTAGGCATCTCGTATGC
SRR616268sub_1.fastq
2812
② iu@bielinux[srp017156] grep -c ^GATCGGAAGAGCACACGTCTGAACTCCAGTCACTTAGGCATCTCGTATGC
SRR616268sub_1.fastq
2415
③ iu@bielinux[srp017156] grep ^GATCGGAAGAGCACACGTCTGAACTCCAGTCACTTAGGCATCTCGTATGC SR
R616268sub_1.fastq | head -n 3
GATCGGAAGAGCACACGTCTGAACTCCAGTCACTTAGGCATCTCGTATGCCGTCTTCTGCTTGAAAAGAATCTGAGCAAACC
CCCTGGTTGTCGGGGCACGGAATAC
GATCGGAAGAGCACACGTCTGAACTCCAGTCACTTAGGCATCTCGTATGCCGTCTTCTGCTTGAAAAAAAAAATAAAAGAAA
AAGACAACCTCAAGCATGGGGGATGG
GATCGGAAGAGCACACGTCTGAACTCCAGTCACTTAGGCATCTCGTATGCCGTCTTCTGCTTGAAAAAACAAAAACCTTAG
TGAACAAACATCGGTCACGGGTCCG
④ iu@bielinux[srp017156] grep ^GATCGGAAGAGCACACGTCTGAACTCCAGTCACTTAGGCATCTCGTATGC SR
R616268sub_1.fastq > age.txt
⑤ iu@bielinux[srp017156] less age.txt [11:20午前]
```


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 (黒矢印部分) の出現回数は最初の50 bpのみで計算しており、妥当

```
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 GATCGGAAGAGCACACGTCTGAACTCCAGTCACTTAGGCATCTCGTATGC
SRR616268sub_1.fastq
2812
iu@bielinux[srp017156] grep -c ^GATCGGAAGAGCACACGTCTGAACTCCAGTCACTTAGGCATCTCGTATGC
SRR616268sub_1.fastq
2415
iu@bielinux[srp017156] grep ^GATCGGAAGAGCACACGTCTGAACTCCAGTCACTTAGGCATCTCGTATGC SR
R616268sub_1.fastq | head -n 3
GATCGGAAGAGCACACGTCTGAACTCCAGTCACTTAGGCATCTCGTATGCCCGTCTTCTGCTTGAAAAGAATCTGAGCAAACC
CCCTGGTCTAGGGAGTATTTAGCCTTGGGAGATGGTCCTCCGGATTCCGACGGAAT 2421 0.2421 No Hit
GATCGGAAGAGCACACGTCTGAACTCCAGTCACTTAGGCATCTCGTATGCCCGTCTTCTGCTTGAAAAGAATCTGAGCAAACC
AAGACAACGATCGGAAGAGCACACGTCTGAACTCCAGTCACTTAGGCATCTCGTATGC 2415 0.2415 TruSeq Adapter, Index 3
(100% over 50bp)
GATCGGAAGAGCACACGTCTGAACTCCAGTCACTTAGGCATCTCGTATGCCCGTCTTCTGCTTGAAAAGAATCTGAGCAAACC
TGAACAAACCGGCATTCTCACTTCTAAGCGCTCCAGCCGTCTCAGCATCAACCTTCA 2393 0.2393 No Hit
iu@bielinux[srp017156] grep ^GATCGGAAGAGCACACGTCTGAACTCCAGTCACTTAGGCATCTCGTATGC SR
R616268sub_1.fastq > age.txt
iu@bielinux[srp017156] less age.txt
```

CCCTGGTCTAGGGAGTATTTAGCCTTGGGAGATGGTCCTCCGGATTCCGACGGAAT	2421	0.2421	No Hit
AAGACAACGATCGGAAGAGCACACGTCTGAACTCCAGTCACTTAGGCATCTCGTATGC	2415	0.2415	TruSeq Adapter, Index 3 (100% over 50bp)
TGAACAAACCGGCATTCTCACTTCTAAGCGCTCCAGCCGTCTCAGCATCAACCTTCA	2393	0.2393	No Hit

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[srp017156] pwd
/home/iu/Documents/srp017156
iu@bielinux[srp017156] grep -c CCCCCG
SRR616268sub_1.fastq
29372
iu@bielinux[srp017156] grep -c ^CCCCGGTATATTTTCGGCGCAGTGCCACTCGACTAGTGAGCTATTACGCA
SRR616268sub_1.fastq
14383
iu@bielinux[srp017156] grep -c GGCCTATTCAGTGCAGGCTGACCTTGCAGGTCAGCACCCCTTCTTCCGAAGT
SRR616268sub_1.fastq
20429
iu@bielinux[srp017156] grep -c ^GGCCTATTCAGTGCAGGCTGACCTTGCAGGTCAGCACCCCTTCTTCCGAAGT
SRR616268sub_1.fastq
11044
iu@bielinux[srp017156]
```

Overrepresented sequences

Sequence	Count
CCCCGGTATATTTTCGGCGCAGTGCCACTCGACTAGTGAGCTATTACGCA	14383
GGCCTATTCAGTGCAGGCTGACCTTGCAGGTCAGCACCCCTTCTTCCGAAGT	11044
GTGCTTTTCACCTTTCCTCACGGTACTGGTTCAGTATCGGTCAGTACAGGG	8892
CCCGGTATATTTTCGGCGCAGTGCCACTCGACTAGTGAGCTATTACGCAC	8474
GTCAGTACAGGGAGTATTTAGCCTTGGGAGATGGTCTCCCGGATTCCGACG	8189

W11-11: Tips

①の結果は2,812行。②の結果は2,415行。③grep -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行だということ

```
iu@bielinux[srp017156] pwd
/home/iu/Documents/srp017156
① iu@bielinux[srp017156] grep GATCGGAAGAGCACACGTCTGAACTCCAGTCACTTAGGCATCTCGTATGC SRR
616268sub_1.fastq > uge1.txt
iu@bielinux[srp017156] wc uge1.txt
2812 2812 303696 uge1.txt
② iu@bielinux[srp017156] grep -c ^GATCGGAAGAGCACACGTCTGAACTCCAGTCACTTAGGCATCTCGTATGC
uge1.txt
2415
③ iu@bielinux[srp017156] grep -v ^GATCGGAAGAGCACACGTCTGAACTCCAGTCACTTAGGCATCTCGTATGC
uge1.txt > uge2.txt
iu@bielinux[srp017156] wc uge2.txt
397 397 42876 uge2.txt
iu@bielinux[srp017156]
```


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[srp017156] pwd
/home/iu/Documents/srp017156
① iu@bielinux[srp017156] grep GATCGGAAGAGC
616268sub_1.fastq > uge1.txt
iu@bielinux[srp017156] wc uge1.txt
2812 2812 303696 uge1.txt
② iu@bielinux[srp017156] grep -c ^GATCGGAAGAGCACACGTCTGAACTCCAGTCACTTAGGCATCTCGTATGC
uge1.txt
2415
③ iu@bielinux[srp017156] grep -v ^GATCGGAAGAGCACACGTCTGAACTCCAGTCACTTAGGCATCTCGTATGC
uge1.txt > uge2.txt
iu@bielinux[srp017156] wc uge2.txt
397 397 42876 uge2.txt
④ iu@bielinux[srp017156] head -n 4 uge2.txt
AGATCGGAAGAGCACACGTCTGAACTCCAGTCACTTAGGCATCTCGTATGCCGTCTTCTGCTTGAAAAAACACAATGCCAC
TCGAATAGTGAGCTTATAAGATCGG
AGATCGGAAGAGCACACGTCTGAACTCCAGTCACTTAGGCATCTCGTATGCCGTCTTCTGCTTGAAAAAAAAAAAAAAAAACAA
ACCGCAGGCAAGAGAAGAAGATCGG
ACCGATATGCTTCGGGAGATCGGAAGAGCACACGTCTGAACTCCAGTCACTTAGGCATCTCGTATGCCGTCTTCTGCTTGAA
AAAAAAAAAACAAAAATCCCGAAG
AGATCGGAAGAGCACACGTCTGAACTCCAGTCACTTAGGCATCTCGTATGCCGTCTTCTGCTTGAAAAAAAAAAAAATAAAAT
ACAAACGCACGCACGCGGAGATCGG
⑤ iu@bielinux[srp017156] less uge2.txt
```

[11:43午前]
[11:43午前]
[11:43午前]
[11:43午前]
[11:43午前]

「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



```
AGATCGGAAGAGCACACGTCTGAACTCCAGTCACTTAGGCATCTCGTATGCCGTCTTCTGCTTGAAAAAACACAATGCCAC
TCGAATAGTGAGCTTATAAGATCGG
AGATCGGAAGAGCACACGTCTGAACTCCAGTCACTTAGGCATCTCGTATGCCGTCTTCTGCTTGAAAAAAAAAAAAAAAAACAA
ACCGCAGGCAAGAGAAGAAGATCGG
ACCGATATGCTTCGGGAGATCGGAAGAGCACACGTCTGAACTCCAGTCACTTAGGCATCTCGTATGCCGTCTTCTGCTTGAA
AAAAAAAAAACAAAAATCCCGAAG
AGATCGGAAGAGCACACGTCTGAACTCCAGTCACTTAGGCATCTCGTATGCCGTCTTCTGCTTGAAAAAAAAAAAAATAAAAT
ACAAACGCACGCACGCGGAGATCGG
GGGACGATCACCCAGATCGGAAGAGCACACGTCTGAACTCCAGTCACTTAGGCATCTCGTATGCCGTCTTCTGCTTGAAAA
AAAAAAAAAAAAACACACGGGTGAT
AGATCGGAAGAGCACACGTCTGAACTCCAGTCACTTAGGCATCTCGTATGCCGTCTTCTGCTTGAAAAACTGTAAAAAGCA
CCATCCTTTTCACTTAGAGATCGG
AGATCGGAAGAGCACACGTCTGAACTCCAGTCACTTAGGCATCTCGTATGCCGTCTTCTGCTTGAAAAAACAAAAAAAA
CCAAGACGAACACCACGGAGATCGG
AGATCGGAAGAGCACACGTCTGAACTCCAGTCACTTAGGCATCTCGTATGCCGTCTTCTGCTTGAAAAACAATCGAACGC
TTCAGTGCGCACTCCCCGAGATCGG
AGATCGGAAGAGCACACGTCTGAACTCCAGTCACTTAGGCATCTCGTATGCCGTCTTCTGCTTGAAAAAAAAAACAAAAAA
AATATTGTAGAGATACATAGATCGG
AGATCGGAAGAGCACACGTCTGAACTCCAGTCACTTAGGCATCTCGTATGCCGTCTTCTGCTTGAAAACGCAAAGAGAGTT
TCGTGTTGCTTCTTTAGTAGATCGG
GCGCACCTTACGGGTGCACCCACAGTTTCGGAGATCGGAAGAGCACACGTCTGAACTCCAGTCACTTAGGCATCTCGTATG
CCGTCTTCTGCTTGAAAACCGAAAC
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

```
AGATCGGAAGAGCACACGTCTGAACTCCAGTCACTTAGGCATCTCGTATGCCGTCTTCTGCTTGAAAAAACACAATGCCAC
TCGAATAGTGAGCTTATAAGATCGG
AGATCGGAAGAGCACACGTCTGAACTCCAGTCACTTAGGCATCTCGTATGCCGTCTTCTGCTTGAAAAAAAAAAAAAAAAACAA
ACCGCAGGCAAGAGAAGAAGATCGG
ACCGATATGCTTCGGGAGATCGGAAGAGCACACGTCTGAACTCCAGTCACTTAGGCATCTCGTATGCCGTCTTCTGCTTGAA
AAAAAAAAAACAAAAATCCCGAAG
AGATCGGAAGAGCACACGTCTGAACTCCAGTCACTTAGGCATCTCGTATGCCGTCTTCTGCTTGAAAAAAAAAAAAATAAAAT
ACAAACGCACGCACGCGGAGATCGG
GGGACGATCACCCAGATCGGAAGAGCACACGTCTGAACTCCAGTCACTTAGGCATCTCGTATGCCGTCTTCTGCTTGAAAA
AAAAAAAAAAAAACACACGGGTGAT
AGATCGGAAGAGCACACGTCTGAACTCCAGTCACTTAGGCATCTCGTATGCCGTCTTCTGCTTGAAAAACTGTAAAAAGCA
CCATCCTTTTCACTTAGAGATCGG
AGATCGGAAGAGCACACGTCTGAACTCCAGTCACTTAGGCATCTCGTATGCCGTCTTCTGCTTGAAAAAACAAAAAAAA
CCAAGACGAACACCACGGAGATCGG
AGATCGGAAGAGCACACGTCTGAACTCCAGTCACTTAGGCATCTCGTATGCCGTCTTCTGCTTGAAAAACAATCGAACGC
TTCAGTGCGCACTCCCCGAGATCGG
AGATCGGAAGAGCACACGTCTGAACTCCAGTCACTTAGGCATCTCGTATGCCGTCTTCTGCTTGAAAAAAAACAAAAAAA
AATATTGTAGAGATACATAGATCGG
AGATCGGAAGAGCACACGTCTGAACTCCAGTCACTTAGGCATCTCGTATGCCGTCTTCTGCTTGAAAACGCAAAGAGAGTT
TCGTGTTGCTTCTTTAGTAGATCGG
GCGCACCTTACGGGTGCACCCACAGTTTCGGAGATCGGAAGAGCACACGTCTGAACTCCAGTCACTTAGGCATCTCGTATG
CCGTCTTCTGCTTGAAAACCGAAAC
/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。パッと見で、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。

File Edit View Search Terminal Help

```
AGATCGGAAGAGCACACGTCTGAACTCCAGTCA
TCGAATAGTGAGCTTATAAGATCGG
AGATCGGAAGAGCACACGTCTGAACTCCAGTCACTTAGGCATCTCGTATGCCGTCTTCTGCTTGAAAAA
ACCGCAGGCAAGAGAAGAAGATCGG
ACCGATATGCTTCGGGAGATCGGAAGAGCACACGTCTGAACTCCAGTCACTTAGGCATCTCGTATGCCGTCTTCTGCTTGAA
AAAAAAAAAAACAAAATCCCGAAG
AGATCGGAAGAGCACACGTCTGAACTCCAGTCACTTAGGCATCTCGTATGCCGTCTTCTGCTTGAAAAAATAAAAT
ACAAACGCACGCACGCGGAGATCGG
GGGACGATCACCCAGATCGGAAGAGCACACGTCTGAACTCCAGTCACTTAGGCATCTCGTATGCCGTCTTCTGCTTGAAAA
AAAAAAAAAAACACACGGGTGAT
AGATCGGAAGAGCACACGTCTGAACTCCAGTCACTTAGGCATCTCGTATGCCGTCTTCTGCTTGAAAAACTGTAAAAGCA
CCATCCTTTTCACTTAGAGATCGG
AGATCGGAAGAGCACACGTCTGAACTCCAGTCACTTAGGCATCTCGTATGCCGTCTTCTGCTTGAAAAACA
CCAAGACGAACACCACGGAGATCGG
AGATCGGAAGAGCACACGTCTGAACTCCAGTCACTTAGGCATCTCGTATGCCGTCTTCTGCTTGAAAAACAATCGAACGC
TTCAGTGCGCACTCCCCGAGATCGG
AGATCGGAAGAGCACACGTCTGAACTCCAGTCACTTAGGCATCTCGTATGCCGTCTTCTGCTTGAAAAACA
AATATTGTAGAGATACATAGATCGG
AGATCGGAAGAGCACACGTCTGAACTCCAGTCACTTAGGCATCTCGTATGCCGTCTTCTGCTTGAAAACGCAAAGAGAGTT
TCGTGTTGCTTCTTTAGTAGATCGG
GCGCACCTTACGGGTGCACCCACAGTTTCGGAGATCGGAAGAGCACACGTCTGAACTCCAGTCACTTAGGCATCTCGTATG
CGTCTTCTGCTTGAAACCGAAAC
```


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 ...
AATGATACGGCGACCACCGAGATCTACTCTTCCCTACACGACGCTCTTCCGATCT
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。

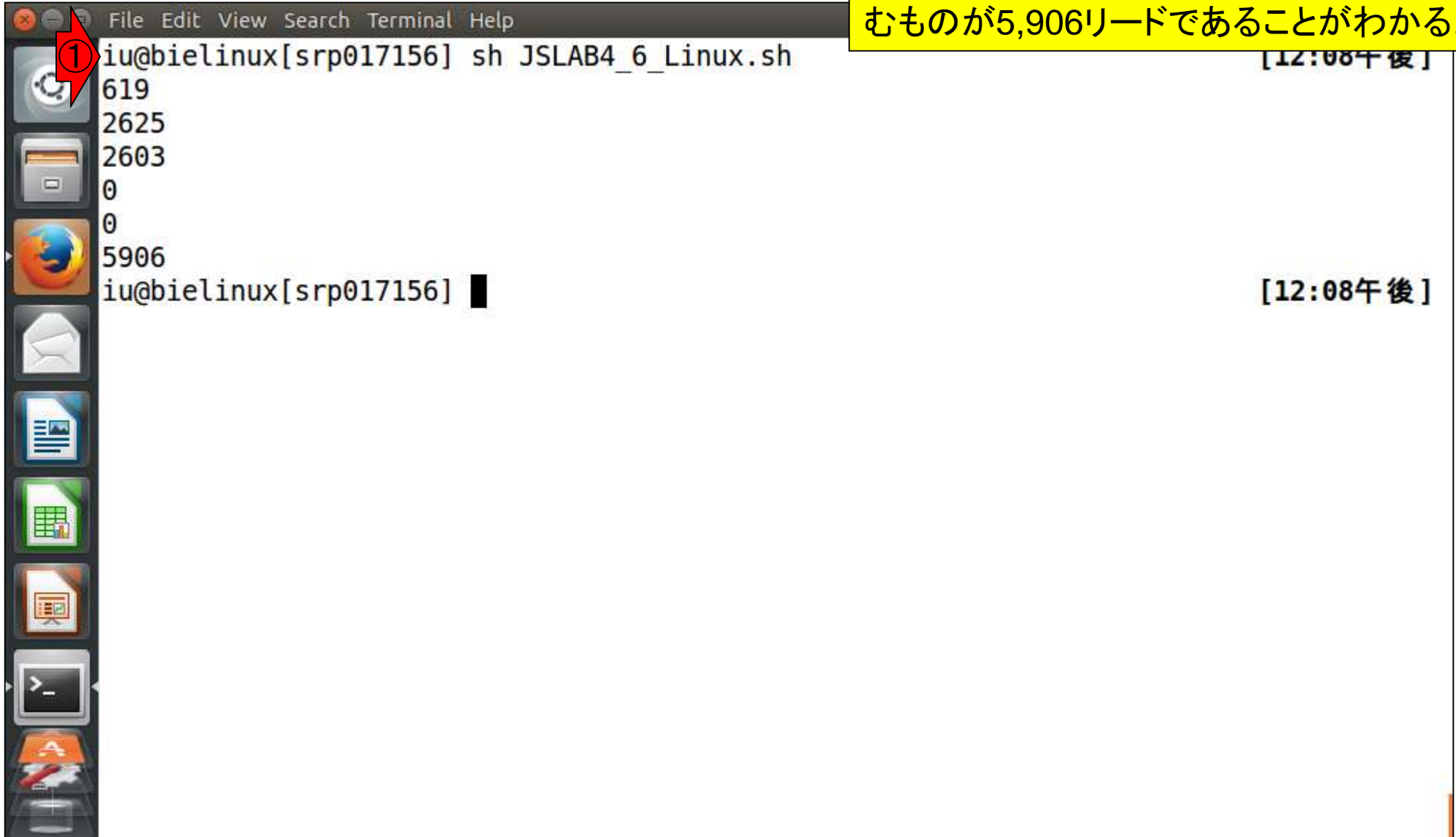
W12-1

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)をダウンロードし、②moreで眺める。赤下線部分がインデックス配列

```
iu@bielinux[srp017156] pwd [12:01午後]
/home/iu/Documents/srp017156
iu@bielinux[srp017156] rm -f hoge*.fastq [12:01午後]
zsh: no matches found: hoge*.fastq
iu@bielinux[srp017156] ls -lh *.fastq [12:01午後]
-rw-rw-r-- 1 iu iu 306M 12月 9 15:24 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 [12:01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] [12:01午後]
```

W12-1

①shでシェルスクリプトを実行。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。



The image shows a terminal window with a menu bar (File, Edit, View, Search, Terminal, Help) and a sidebar with application icons. A red arrow with the number '1' points to the command prompt. The terminal displays the following text:

```
iu@bielinux[srp017156] sh JSLAB4_6_Linux.sh  
619  
2625  
2603  
0  
0  
5906  
iu@bielinux[srp017156] █
```

Timestamps are shown as [12:08午後] on the right side of the terminal output.

W12-2: Tips

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。②Index 1の結果を保存するときには通常のリダイレクト(>)。③Index 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。

```
iu@bielinux[srp017156] pwd
/home/iu/Documents/srp017156
iu@bielinux[srp017156] ls
age.txt                JSLAB4_3_Linux.sh      SRR616268_2.fastq.bz2  uge1.txt
JSLAB4_1_Linux.sh      JSLAB4_6_Linux.sh      SRR616268sub_1.fastq   uge2.txt
JSLAB4_2_Linux.sh      SRR616268_1.fastq.bz2 SRR616268sub_2.fastq.gz
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
### TruSeq Adapter, Index 1
grep GATCGGAAGAGCACACGTCTGAACTCCAGTCACATCACGATCTCGTATGCCGTCTTCTGCTTG SRR616268sub_1.
fastq > hoge.txt
wc hoge.txt
### TruSeq Adapter, Index 2
grep GATCGGAAGAGCACACGTCTGAACTCCAGTCACCGATGTATCTCGTATGCCGTCTTCTGCTTG SRR616268sub_1.
fastq >> hoge.txt
wc hoge.txt
### TruSeq Adapter, Index 3
grep GATCGGAAGAGCACACGTCTGAACTCCAGTCACTTAGGCATCTCGTATGCCGTCTTCTGCTTG SRR616268sub_1.
fastq >> hoge.txt
wc hoge.txt
iu@bielinux[srp017156]
```



[12:20午後]

[12:20午後]

[12:20午後]

W12-2: Tips

①シェルスクリプトを実行。hoge.txtの行数が619行, 3244行, 5847行と増えていることがわかる。②シェルスクリプト実行後のhoge.txtの行数は5847。

```
iu@bielinux[srp017156] ls -lh hoge* [12:26午後]
zsh: no matches found: hoge*
① iu@bielinux[srp017156] sh JSLAB4_7_Linux.sh [12:26午後]
   619   619 66852 hoge.txt
   3244  3244 350352 hoge.txt
   5847  5847 631476 hoge.txt
iu@bielinux[srp017156] ls -lh hoge* [12:26午後]
-rw-rw-r-- 1 iu iu 617K 12月 11 12:26 hoge.txt
② iu@bielinux[srp017156] wc hoge.txt [12:26午後]
   5847   5847 631476 hoge.txt
iu@bielinux[srp017156] [12:26午後]
```


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[srp017156] pwd
/home/iu/Documents/srp017156
iu@bielinux[srp017156] ls -lh *sub*
-rw-rw-r-- 1 iu iu 306M 12月  9 15:24 SRR616268sub_1.fastq
-rw-rw-r-- 1 iu iu  66M 12月  9 15:24 SRR616268sub_2.fastq.gz
iu@bielinux[srp017156] gzip SRR616268sub_1.fastq
iu@bielinux[srp017156] ls -lh *sub*
-rw-rw-r-- 1 iu iu 74M 12月  9 15:24 SRR616268sub_1.fastq.gz
-rw-rw-r-- 1 iu iu 66M 12月  9 15:24 SRR616268sub_2.fastq.gz
① iu@bielinux[srp017156] gunzip -c SRR616268sub_1.fastq.gz | fastx_clipper -a GATCGGAA
GAGCACACGTCTGAACTCCAGTCACTTAGGCATCTCGTATGC -o hoge8.fastq
② iu@bielinux[srp017156] wc hoge8.fastq
 3915596  7831192 286011042 hoge8.fastq
iu@bielinux[srp017156] ls -lh *sub*
-rw-rw-r-- 1 iu iu 74M 12月  9 15:24 SRR616268sub_1.fastq.gz
-rw-rw-r-- 1 iu iu 66M 12月  9 15:24 SRR616268sub_2.fastq.gz
iu@bielinux[srp017156] █
```


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 ...

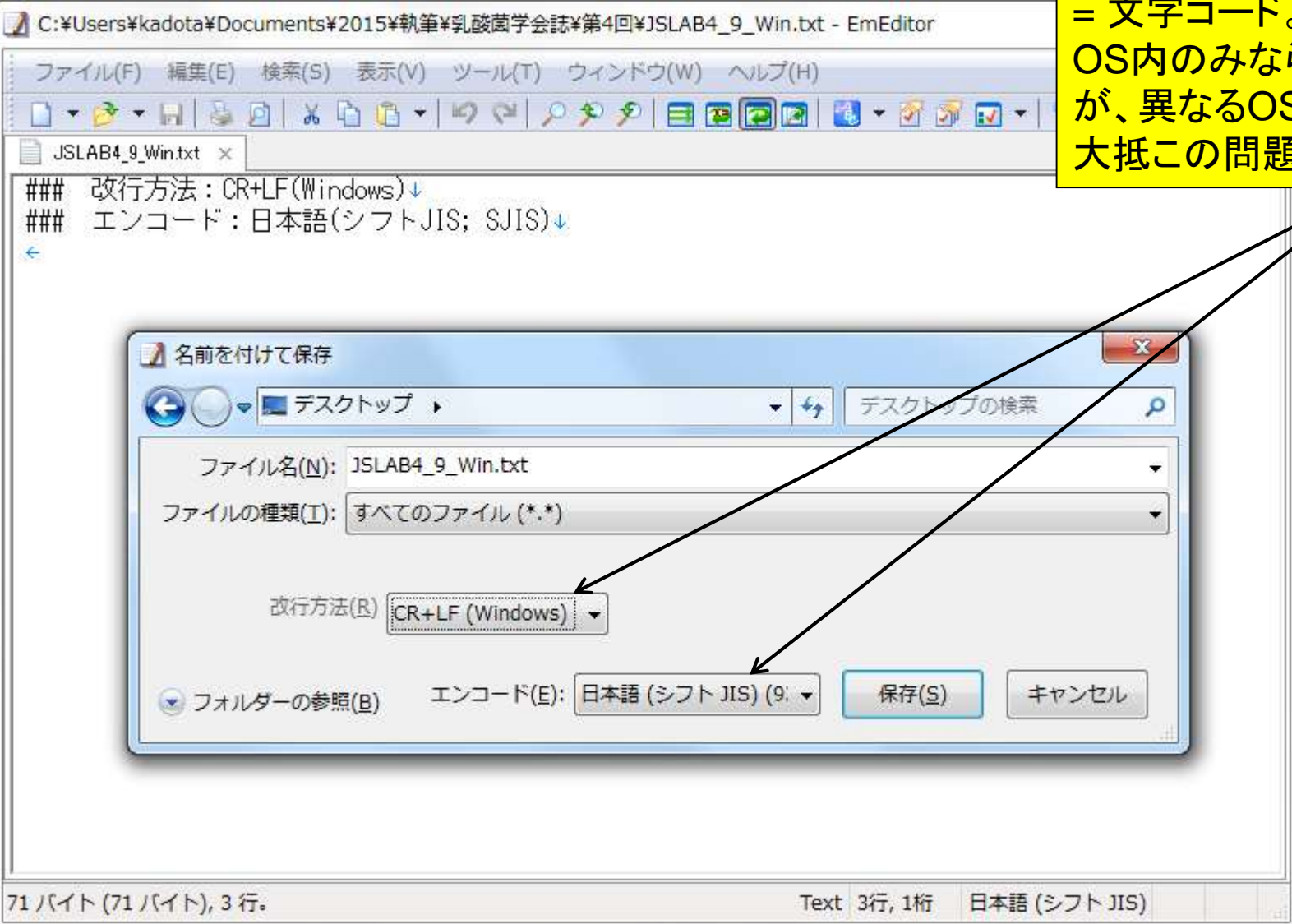
Linuxコマンド集 - 【nkf】文字コードを変換する:ITpro
itpro.nikkeibp.co.jp > 情報システム > オープンソース/Linux ▼
2006/02/27 - -b, バッファリング出力を行う。-u, 出力時にバッファリングを行わない。-j, JISコードに変換する。-e, EUCコードに変換する。-s, シフトJISコードに変換する。-w, UTF8コードに変換する。-i? JIS漢字を指示するシーケンスとして ESC-'\$'-?を使用する。

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。

```
File Edit View Search Terminal Help
iu@bielinux[srp017156] pwd [ 1:26午後 ]
/home/iu/Documents/srp017156
iu@bielinux[srp017156] wget -cq http://www.iu.a.u-tokyo.ac.jp/~kadota/book/JSLAB4_9_
Win.txt
iu@bielinux[srp017156] more JSLAB4_9_Win.txt [ 1:26午後 ]
### 000s000@0FCR+LF(Windows)
### 0G000R0[0h0F000{00(0V0t0gJIS; SJIS)
iu@bielinux[srp017156] where iconv [ 1:26午後 ]
/usr/bin/iconv
iu@bielinux[srp017156] where nkf [ 1:26午後 ]
nkf not found
iu@bielinux[srp017156] [ 1:26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。②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いデフォルトはUTF8変換だと学習する。

```
iu@bielinux[srp017156] pwd
/home/iu/Documents/srp017156
iu@bielinux[srp017156] wget -cq ht
Win.txt
iu@bielinux[srp017156] more JSLAB4
### 000s000@FCR+LF(Windows)
### 0G000R0[0h0F000{00(0V0t0gJIS; SJIS)
iu@bielinux[srp017156] where iconv [ 1:26午後]
/usr/bin/iconv
iu@bielinux[srp017156] where nkf [ 1:26午後]
nkf not found
iu@bielinux[srp017156] iconv -f SJIS JSLAB4_9_Win.txt > hoge.txt [ 1:26午後]
iu@bielinux[srp017156] more hoge.txt [ 1:30午後]
### 改行方法 : CR+LF(Windows)
### エンコード : 日本語 (シフト JIS; SJIS)
iu@bielinux[srp017156] iconv -f SJIS -t UTF8 JSLAB4_9_Win.txt > hoge2.txt [ 1:30午後]
iu@bielinux[srp017156] more hoge2.txt [ 1:30午後]
### 改行方法 : CR+LF(Windows)
### エンコード : 日本語 (シフト JIS; SJIS)
iu@bielinux[srp017156] [ 1:30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[srp017156] pwd
/home/iu/Documents/srp017156
iu@bielinux[srp017156] wget -cq http://www.iu.a.u-tokyo.ac.jp/~kadota/book/JSLAB4_9_Win.txt
iu@bielinux[srp017156] more JSLAB4_9_Win.txt [ 1:26午後 ]
### 000s000@FCR+LF(Windows)
### 0G000R0[0h0F000{00(0V0t0gJIS; SJIS)
iu@bielinux[srp017156] where iconv [ 1:26午後 ]
/usr/bin/iconv
iu@bielinux[srp017156] where nkf [ 1:26午後 ]
nkf not found
iu@bielinux[srp017156] iconv -f SJIS JSLAB4_9_Win.txt > hoge.txt [ 1:26午後 ]
iu@bielinux[srp017156] more hoge.txt [ 1:30午後 ]
### 改行方法 : CR+LF(Windows)
### エンコード : 日本語 (シフト JIS; SJIS)
iu@bielinux[srp017156] iconv -f SJIS -t UTF8 JSLAB4_9_Win.txt > hoge2.txt
iu@bielinux[srp017156] more hoge2.txt [ 1:30午後 ]
### 改行方法 : CR+LF(Windows)
### エンコード : 日本語 (シフト JIS; SJIS)
iu@bielinux[srp017156] █ [ 1:30午後 ]
```


W13-5: nkfインストール

作業ディレクトリはどこでもよい。①インストール前はnkfのパスは通ってない。②nkf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。

```
File Edit View Search Terminal Help
① iu@bielinux[serp017156] where nkf
nkf not found
② iu@bielinux[serp017156] sudo apt-get install nkf
[sudo] password for iu:
Reading package lists... Done
Building dependency tree
Reading state information... Done
The following packages were automatically installed and are no longer required:
  linux-headers-3.13.0-55 linux-headers-3.13.0-55-generic
  linux-image-3.13.0-55-generic linux-image-extra-3.13.0-55-generic
Use 'apt-get autoremove' to remove them.
The following NEW packages will be installed:
  nkf
0 upgraded, 1 newly installed, 0 to remove and 110 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 (179 kB/s)
Selecting previously unselected package nkf.
(Reading database ... 379604 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 (2.6.7.1-1ubuntu1) ...
```


W13-5:nkfインストール

特にエラーメッセージも出ていないが念のため①where nkfで確認。/usr/bin/nkfとパスが無事通っていることが分かる。

```
File Edit View Search Terminal Help
Reading package lists... Done
Building dependency tree
Reading state information... Done
The following packages were automatically installed and are no longer required:
  linux-headers-3.13.0-55 linux-headers-3.13.0-55-generic
  linux-image-3.13.0-55-generic linux-image-extra-3.13.0-55-generic
Use 'apt-get autoremove' to remove them.
The following NEW packages will be installed:
  nkf
0 upgraded, 1 newly installed, 0 to remove and 110 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 (179 kB/s)
Selecting previously unselected package nkf.
(Reading database ... 379604 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 (2.6.7.1-1ubuntu1) ...
Setting up nkf (2.13-1) ...
iu@bielinux[srp017156] where nkf                                     [ 1:39午後 ]
/usr/bin/nkf
iu@bielinux[srp017156] █                                           [ 1:42午後 ]
```



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[srp017156] pwd [ 1:48午後 ]
/home/iu/Documents/srp017156
iu@bielinux[srp017156] ls -lh *.txt [ 1:49午後 ]
-rw-rw-r-- 1 iu iu 88 12月 11 13:30 hoge2.txt
-rw-rw-r-- 1 iu iu 88 12月 11 13:30 hoge.txt
-rw-rw-r-- 1 iu iu 71 5月 27 2015 JSLAB4_9_Win.txt
① iu@bielinux[srp017156] nkf --guess JSLAB4_9_Win.txt [ 1:49午後 ]
Shift_JIS (LF)
② iu@bielinux[srp017156] nkf --guess hoge.txt [ 1:49午後 ]
UTF-8 (LF)
③ iu@bielinux[srp017156] nkf --guess hoge2.txt [ 1:49午後 ]
UTF-8 (LF)
④ iu@bielinux[srp017156] nkf -g JSLAB4_9_Win.txt [ 1:49午後 ]
Shift_JIS
⑤ iu@bielinux[srp017156] nkf -g hoge.txt [ 1:49午後 ]
UTF-8
⑥ iu@bielinux[srp017156] nkf -g hoge2.txt [ 1:49午後 ]
UTF-8
iu@bielinux[srp017156] █ [ 1:49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、nkfのほうが便利

```
iu@bielinux[srp017156] pwd [ 1:55午後 ]
/home/iu/Documents/srp017156
iu@bielinux[srp017156] ls -lh *.txt [ 1:55午後 ]
-rw-rw-r-- 1 iu iu 88 12月 11 13:30 hoge2.txt
-rw-rw-r-- 1 iu iu 88 12月 11 13:30 hoge.txt
-rw-rw-r-- 1 iu iu 71 5月 27 2015 JSLAB4_9_Win.txt
iu@bielinux[srp017156] more JSLAB4_9_Win.txt [ 1:55午後 ]
### 000s000@FCR+LF(Windows)
### 0G000R0[0h0F000{00(0V0t0gJIS; SJIS)
① iu@bielinux[srp017156] nkf JSLAB4_9_Win.txt > hoge3.txt [ 1:55午後 ]
iu@bielinux[srp017156] more hoge3.txt [ 1:55午後 ]
### 改行方法 : CR+LF(Windows)
### エンコード : 日本語 (シフトJIS; SJIS)
② iu@bielinux[srp017156] iconv JSLAB4_9_Win.txt > hoge4.txt [ 1:55午後 ]
iconv: illegal input sequence at position 5
iu@bielinux[srp017156] more hoge4.txt [ 1:55午後 ]
###
iu@bielinux[srp017156] [ 1:55午後 ]
```


W13-7: アンインストール

①と②はnkfが使える状態であることを確認するだけ。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。⑤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と打つ。

```
iu@bielinux[srp017156] nkf -v
Network Kanji Filter Version 2.1.3 (2013-11-11)
Copyright (C) 1987, FUJITSU LTD. (I.Ichikawa)
Copyright (C) 1996-2013, The nkf Project.

iu@bielinux[srp017156] where nkf
/usr/bin/nkf

iu@bielinux[srp017156] apt-get remove nkf
E: Could not open lock file /var/lib/dpkg/lock - open (13: Permission denied)
E: Unable to lock the administration directory (/var/lib/dpkg/), are you root?

iu@bielinux[srp017156] sudo apt-get remove nkf
[sudo] password for iu:
Reading package lists... Done
Building dependency tree
Reading state information... Done
The following packages were automatically installed and are no longer required:
 linux-headers-3.13.0-55 linux-headers-3.13.0-55-generic
 linux-image-3.13.0-55-generic linux-image-extra-3.13.0-55-generic
Use 'apt-get autoremove' to remove them.
The following packages will be REMOVED:
 nkf
0 upgraded, 0 newly installed, 1 to remove and 110 not upgraded.
After this operation, 357 kB disk space will be freed.
Do you want to continue? [Y/n]
```

[1:59午後]

W13-7: アンインストール

⑥nkfコマンドがなくなったことを確認。一通り確認したら、⑦再インストールしておきましょう。

```
File Edit View Search Terminal Help
iu@bielinux[srp017156] apt-get remove nkf [ 1:59午後 ]
E: Could not open lock file /var/lib/dpkg/lock - open (13: Permission denied)
E: Unable to lock the administration directory (/var/lib/dpkg/), are you root?
iu@bielinux[srp017156] sudo apt-get remove nkf [ 1:59午後 ]
[sudo] password for iu:
Reading package lists... Done
Building dependency tree
Reading state information... Done
The following packages were automatically installed and are no longer required:
  linux-headers-3.13.0-55 linux-headers-3.13.0-55-generic
  linux-image-3.13.0-55-generic linux-image-extra-3.13.0-55-generic
Use 'apt-get autoremove' to remove them.
The following packages will be REMOVED:
  nkf
0 upgraded, 0 newly installed, 1 to remove and 110 not upgraded.
After this operation, 357 kB disk space will be freed.
⑤ Do you want to continue? [Y/n] y
(Reading database ... 379612 files and directories currently installed.)
Removing nkf (2.13-1) ...
Processing triggers for man-db (2.6.7.1-1ubuntu1) ...
⑥ iu@bielinux[srp017156] where nkf [ 2:02午後 ]
nkf not found
⑦ iu@bielinux[srp017156] sudo apt-get install nkf [ 2:03午後 ]
```


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[srp017156] pwd
/home/iu/Documents/srp017156
② iu@bielinux[srp017156] where pip
pip not found
③ iu@bielinux[srp017156] which pip
pip not found
④ iu@bielinux[srp017156] sudo apt-get install python-pip [ 2:11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。

```
File Edit View Search Terminal Help
Reading package lists... Done
Building dependency tree
Reading state information... Done
The following packages were automatically installed and are no longer required:
 linux-headers-3.13.0-55 linux-headers-3.13.0-55-generic
 linux-image-3.13.0-55-generic linux-image-extra-3.13.0-55-generic
Use 'apt-get autoremove' to remove them.
The following extra packages will be installed:
 python-chardet-whl python-colorama python-colorama-whl python-distlib
 python-distlib-whl python-html5lib-whl python-pip-whl python-requests-whl
 python-setuptools python-setuptools-whl python-six-whl python-urllib3-whl
 python-wheel
Recommended packages:
 python-dev-all
The following NEW packages will be installed:
 python-chardet-whl python-colorama python-colorama-whl python-distlib
 python-distlib-whl python-html5lib-whl python-pip python-pip-whl
 python-requests-whl python-setuptools python-setuptools-whl python-six-whl
 python-urllib3-whl python-wheel
0 upgraded, 14 newly installed, 0 to remove and 110 not upgraded.
Need to get 1,599 kB of archives.
After this operation, 3,626 kB of additional disk space will be used.
Do you want to continue? [Y/n] y
```



インストール作業完了後に、①「pip -h」で
マニュアルを表示。

W14-2: pip確認

```
File Edit View Search Terminal Help
Unpacking python-setuptools (3.3-1ubuntu2) ...
Selecting previously unselected package python-pip.
Preparing to unpack .../python-pip_1.5.4-1ubuntu3_all.deb ...
Unpacking python-pip (1.5.4-1ubuntu3) ...
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 (2.6.7.1-1ubuntu1) ...
Setting up python-chardet-whl (2.2.1-2~ubuntu1) ...
Setting up python-colorama (0.2.5-0.1ubuntu2) ...
Setting up python-colorama-whl (0.2.5-0.1ubuntu2) ...
Setting up python-distlib (0.1.8-1ubuntu1) ...
Setting up python-distlib-whl (0.1.8-1ubuntu1) ...
Setting up python-html5lib-whl (0.999-3~ubuntu1) ...
Setting up python-six-whl (1.5.2-1ubuntu1) ...
Setting up python-urllib3-whl (1.7.1-1ubuntu4) ...
Setting up python-requests-whl (2.2.1-1ubuntu0.3) ...
Setting up python-setuptools-whl (3.3-1ubuntu2) ...
Setting up python-pip-whl (1.5.4-1ubuntu3) ...
Setting up python-setuptools (3.3-1ubuntu2) ...
Setting up python-pip (1.5.4-1ubuntu3) ...
Setting up python-wheel (0.24.0-1~ubuntu1) ...
iu@bielinux[srp017156] pip -h
```

[2:14午後]



W14-2: pip確認

インストール作業完了後に、①「pip -h」でマニュアルを表示。②pip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
iu@bielinux[srp017156] pip -h
Usage:
  pip <command> [options]

Commands:
  install          Install packages.
  uninstall       Uninstall packages.
  freeze          Output installed packages in requirements format.
  list            List installed packages.
  show           Show information about installed packages.
  search         Search PyPI for packages.
  wheel         Build wheels from your requirements.
  zip           DEPRECATED. Zip individual packages.
  unzip        DEPRECATED. Unzip individual packages.
  bundle       DEPRECATED. Create pybundles.
  help        Show help for commands.

General Options:
  -h, --help          Show help.
  -v, --verbose       Give more output. Option is additive, and can be
                     used up to 3 times.
  -V, --version       Show version and exit.
```



[2:14午後]

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。2015年12月11日に再度別のPCで行うと普通にインストールできました

```
iu@bielinux[srp017156] pwd
/home/iu/Documents/srp017156
iu@bielinux[srp017156] where pip
/usr/bin/pip
iu@bielinux[srp017156] sudo pip install cutadapt [ 2:21午後 ]
Downloading/unpacking cutadapt
  Downloading cutadapt-1.9.1.tar.gz (595kB): 595kB downloaded
  Running setup.py (path:/tmp/pip_build_root/cutadapt/setup.py) egg_info for package
  cutadapt

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/python2.7 -c cutadapt/_align.c -o build/te
      mp.linux-x86_64-2.7/cutadapt/_align.o
      x86_64-linux-gnu-gcc -pthread -shared -Wl,-O1 -Wl,-Bsymbolic-functions -Wl,-Bsym
      bolic-functions -Wl,-z,relro -fno-strict-aliasing -DNDEBUG -g -fwrapv -O2 -Wall -Wst
      rict-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/_align.o -o bui
      ld/lib.linux-x86_64-2.7/cutadapt/_align.so
    building 'cutadapt._qualtrim' extension
      x86_64-linux-gnu-gcc -pthread -fno-strict-aliasing -DNDEBUG -g -fwrapv -O2 -Wall
```

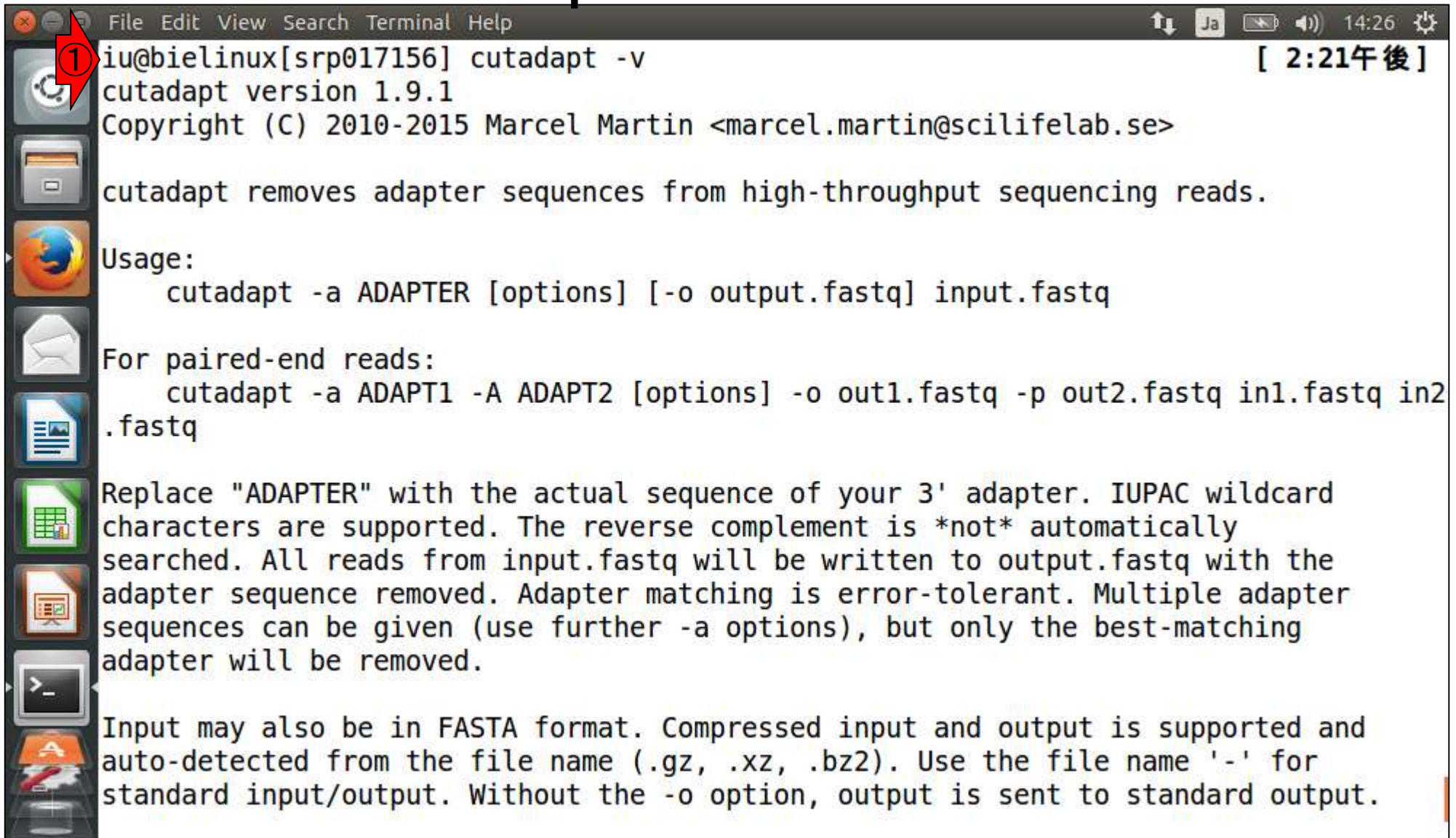

W14-3: cutadapt

```
File Edit View Search Terminal Help 14:24
x86_64-linux-gnu-gcc -pthread -fno-strict-aliasing -DNDEBUG -g -fwrapv -O2 -Wall
-Wstrict-prototypes -fPIC -I/usr/include/python2.7 -c cutadapt/_qualtrim.c -o build
/temp.linux-x86_64-2.7/cutadapt/_qualtrim.o
x86_64-linux-gnu-gcc -pthread -shared -Wl,-O1 -Wl,-Bsymbolic-functions -Wl,-Bsym
bolic-functions -Wl,-z,relro -fno-strict-aliasing -DNDEBUG -g -fwrapv -O2 -Wall -Wst
rict-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/_qualtrim.o -o
build/lib.linux-x86_64-2.7/cutadapt/_qualtrim.so
building 'cutadapt._seqio' extension
x86_64-linux-gnu-gcc -pthread -fno-strict-aliasing -DNDEBUG -g -fwrapv -O2 -Wall
-Wstrict-prototypes -fPIC -I/usr/include/python2.7 -c cutadapt/_seqio.c -o build/te
mp.linux-x86_64-2.7/cutadapt/_seqio.o
x86_64-linux-gnu-gcc -pthread -shared -Wl,-O1 -Wl,-Bsymbolic-functions -Wl,-Bsym
bolic-functions -Wl,-z,relro -fno-strict-aliasing -DNDEBUG -g -fwrapv -O2 -Wall -Wst
rict-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 bui
ld/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
① Successfully installed cutadapt
Cleaning up...
iu@bielinux[serp017156] █ [ 2:21午後 ]
```


無事インストール完了。①「cutadapt -v」でバージョン情報を把握。ver. 1.9.1であることがわかる

W14-3: cutadapt



```
File Edit View Search Terminal Help
iu@bielinux[srp017156] cutadapt -v [ 2:21午後 ]
cutadapt version 1.9.1
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

Replace "ADAPTER" with the actual sequence of your 3' adapter. IUPAC wildcard
characters are supported. The reverse complement is *not* automatically
searched. All reads from input.fastq will be written to output.fastq with the
adapter sequence removed. Adapter matching is error-tolerant. Multiple adapter
sequences can be given (use further -a options), but only the best-matching
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 name '-' for
standard input/output. Without the -o option, output is sent to standard output.
```


W14-3: cutadapt

しかしよく見ると①「-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、②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
characters are supported. The reverse complement is not automatically
searched. All reads from input.fastq will be written to output.fastq with the
adapter sequence removed. Adapter matching is error-tolerant. Multiple adapter
sequences can be given (use further -a options), but only the best-matching
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 name '-' for
standard input/output. Without the -o option, output is sent to standard output.

Citation:

Marcel Martin. Cutadapt removes adapter sequences from high-throughput
sequencing reads. EMBnet.Journal, 17(1):10-12, May 2011.
http://dx.doi.org/10.14806/ej.17.1.200

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

cutadapt: error: no such option: -v
iu@bielinux[srp017156] cutadapt --version
1.9.1
iu@bielinux[srp017156]
```

[2:26午後]

[2:27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で必要な環境(Prerequisites)を整えておく。要するに①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
```

切り取り(T)
コピー(C) ①
貼り付け
すべて選択(A)
印刷(I)...
印刷プレビュー(N)...

thing if they are already installed.

e equivalent command seems to be

File Edit View Search Terminal Help

iu@bielinux[srp017156] 14:45 [2:44午後]

```
iu@bielinux[srp017156] pwd  
/home/iu/Documents/srp017156  
iu@bielinux[srp017156] sudo apt-get install build-essential python2.7-dev  
python-numpy python-matplotlib
```

W14-4: HTSeq

②実行後の全貌。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

```
iu@bielinux[srp017156] pwd [ 2:44午後 ]
/home/iu/Documents/srp017156
② iu@bielinux[srp017156] 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.
python-matplotlib is already the newest version.
python-matplotlib set to manually installed.
python-numpy is already the newest version.
python-numpy set to manually installed.
python2.7-dev is already the newest version.
python2.7-dev set to manually installed.
The following packages were automatically installed and are no longer required:
  linux-headers-3.13.0-55 linux-headers-3.13.0-55-generic
  linux-image-3.13.0-55-generic linux-image-extra-3.13.0-55-generic
Use 'apt-get autoremove' to remove them.
0 upgraded, 0 newly installed, 0 to remove and 110 not upgraded.
iu@bielinux[srp017156] [ 2:58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
```

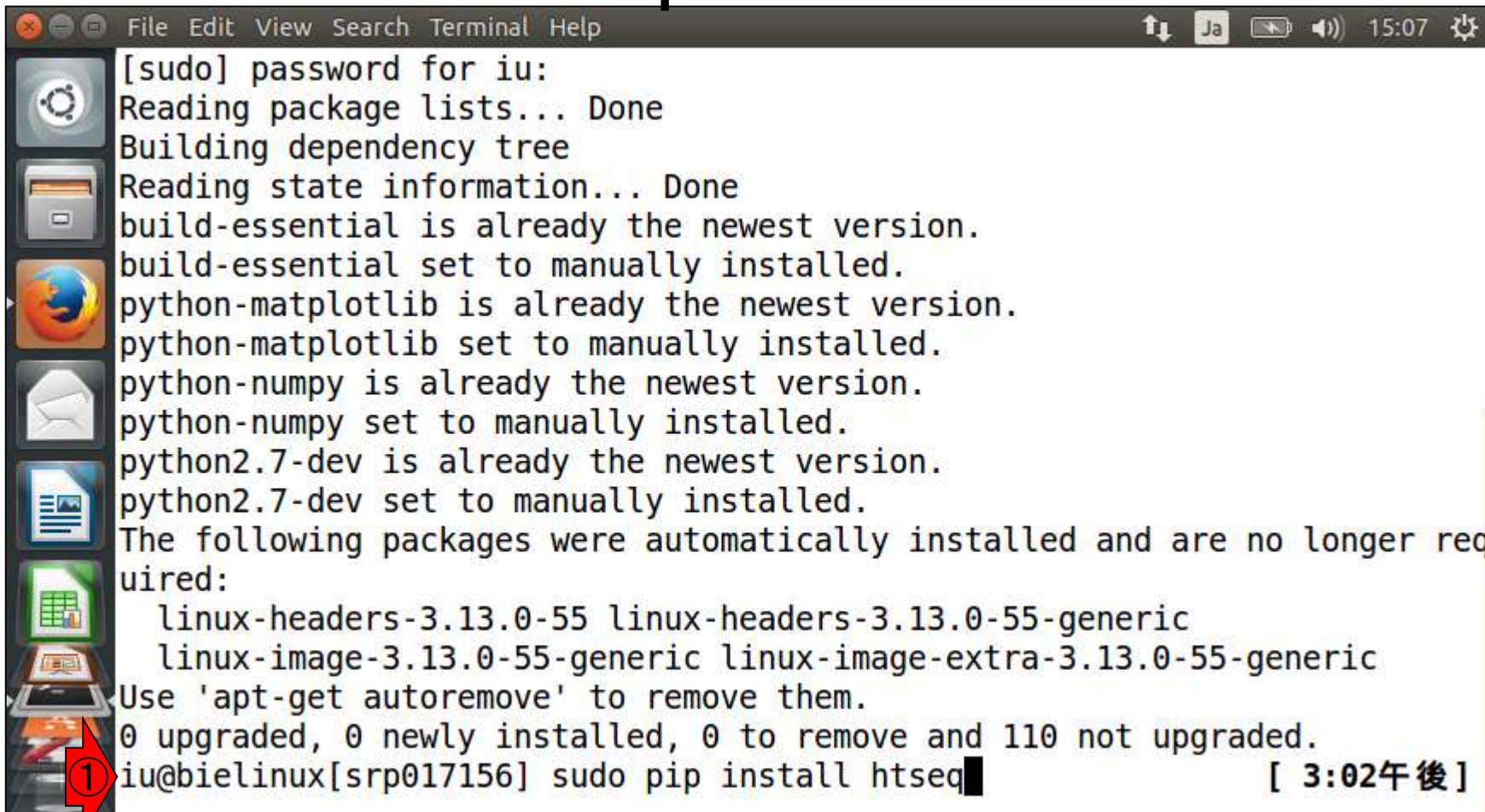
To test the installation, change to another director than the build directory, start Python (by typing `python`) and then try whether typing `import HTSeq` causes an error meesage.

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



```
[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.
python-matplotlib is already the newest version.
python-matplotlib set to manually installed.
python-numpy is already the newest version.
python-numpy set to manually installed.
python2.7-dev is already the newest version.
python2.7-dev set to manually installed.
The following packages were automatically installed and are no longer required:
  linux-headers-3.13.0-55 linux-headers-3.13.0-55-generic
  linux-image-3.13.0-55-generic linux-image-extra-3.13.0-55-generic
Use 'apt-get autoremove' to remove them.
0 upgraded, 0 newly installed, 0 to remove and 110 not upgraded.
iu@bielinux[srp017156] sudo pip install htseq [ 3:02午後 ]
```

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.



```
File Edit View Search Terminal Help
c++ -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 -Werror=format-security build/temp.linux-x86_64-2
.7/src/StepVector_wrap.o -o build/lib.linux-x86_64-2.7/HTSeq/_StepVector.
so
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

changing mode of /usr/local/bin/htseq-qa to 755
changing mode of /usr/local/bin/htseq-count to 755

① Successfully installed htseq
Cleaning up...
② iu@bielinux[srp017156] python
Python 2.7.6 (default, Jun 22 2015, 17:58:13)
[GCC 4.8.2] on linux2
Type "help", "copyright", "credits" or "license" for more information.
③ >>> import HTSeq
>>> █
```

[3:08午後]

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
s0
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

changing mode of /usr/local/bin/htseq-qa to 755
changing mode of /usr/local/bin/htseq-count to 755
Successfully installed htseq
Cleaning up...
iu@bielinux[srp017156] python [ 3:08午後 ]
Python 2.7.6 (default, Jun 22 2015, 17:58:13)
[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[srp017156] [ 3:12午後 ]
```


W14-6: pysam

pysamパッケージのインストールは①「sudo pip install pysam」。数分程度かかる。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。



```
File Edit View Search Terminal Help
iu@bielinux[serp017156] sudo pip install pysam [ 3:14午後]
```

W14-6: pysam

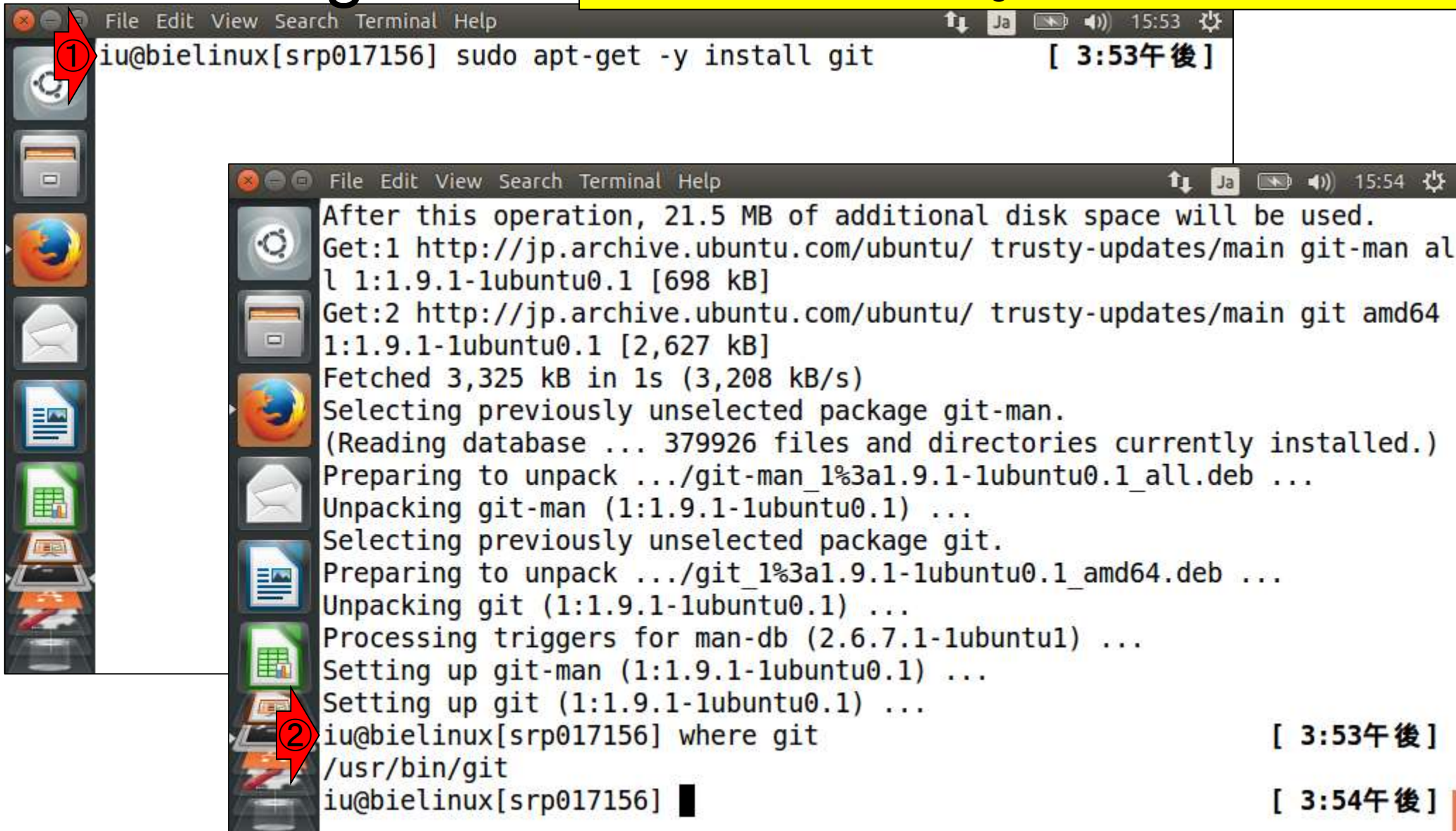
無事インストール完了。2015年9月18日に別のPCでこの作業を行うと、cython関係でエラーが出ました。①2015年12月11日に再度別のPCでやるとうまくいきました。

```
File Edit View Search Terminal Help 15:18
ce_file.o build/temp.linux-x86_64-2.7/htslib/cram/mFILE.o build/temp.linux-x86_64-2.7/htslib/cram/vlen.o build/temp.linux-x86_64-2.7/htslib/cram/cram_samtools.o build/temp.linux-x86_64-2.7/htslib/cram/cram_index.o build/temp.linux-x86_64-2.7/htslib/cram/cram_encode.o build/temp.linux-x86_64-2.7/htslib/cram/md5.o build/temp.linux-x86_64-2.7/htslib/cram/zfio.o build/temp.linux-x86_64-2.7/htslib/cram/thread_pool.o build/temp.linux-x86_64-2.7/htslib/cram/rANS_static.o build/temp.linux-x86_64-2.7/htslib/cram/cram_stats.o build/temp.linux-x86_64-2.7/htslib/cram/cram_decode.o build/temp.linux-x86_64-2.7/htslib/cram/string_alloc.o build/temp.linux-x86_64-2.7/htslib/cram/cram_io.o build/temp.linux-x86_64-2.7/htslib/cram/sam_header.o build/temp.linux-x86_64-2.7/htslib/cram/files.o -Lpysam -lz -o build/lib.linux-x86_64-2.7/pysam/cutils.so

warning: no files found matching 'tests/pysam_data'
warning: no files found matching 'tests/tabix_data'
① Successfully installed pysam
Cleaning up...
iu@bielinux[srp017156] [ 3:18午後 ]
```


W15-1 : git

①git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。無事インストール完了したようなので、②where gitでパスが通っているかを確認



```
iu@bielinux[srp017156] sudo apt-get -y install git [ 3:53午後]
```

```
After this operation, 21.5 MB of additional disk space will be used.  
Get:1 http://jp.archive.ubuntu.com/ubuntu/ trusty-updates/main git-man all  
1:1.9.1-1ubuntu0.1 [698 kB]  
Get:2 http://jp.archive.ubuntu.com/ubuntu/ trusty-updates/main git amd64  
1:1.9.1-1ubuntu0.1 [2,627 kB]  
Fetched 3,325 kB in 1s (3,208 kB/s)  
Selecting previously unselected package git-man.  
(Reading database ... 379926 files and directories currently installed.)  
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 (2.6.7.1-1ubuntu1) ...  
Setting up git-man (1:1.9.1-1ubuntu0.1) ...  
Setting up git (1:1.9.1-1ubuntu0.1) ...  
iu@bielinux[srp017156] where git [ 3:53午後]  
/usr/bin/git  
iu@bielinux[srp017156] [ 3:54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。
②ときどきこのようにコマンドを勝手に修正しようとするが、まあ気にしない。ここではとりあえずyとやったが、本来やりたかったコマンドではないので、git cloneはうまく動いていない。

```
iu@bielinux[srp017156] cd ~/Downloads
iu@bielinux[Downloads] pwd
/home/iu/Downloads
iu@bielinux[Downloads] ls
FastQC fastqc_v0.11.4.zip
iu@bielinux[Downloads] git clone https://github.com/LANL-Bioinformatics/F
aQCs
zsh: correct 'git' to ' git' [nyae]? y
iu@bielinux[Downloads] █
```

[4:05午後]

再度①をやり直し。よく見るとgitを_gitに変更しようとしているので、ここでは②nを押した。うまくダウンロードできたようだ。

W15-2: FaQCs

```
iu@bielinux[srp017156] cd ~/Downloads [ 4:03午後 ]
iu@bielinux[Downloads] pwd [ 4:03午後 ]
/home/iu/Downloads
iu@bielinux[Downloads] ls [ 4:03午後 ]
FastQC fastqc_v0.11.4.zip
iu@bielinux[Downloads] git clone https://github.com/LANL-Bioinformatics/F
aQCs
zsh: correct 'git' to '_git' [nyae]? y
iu@bielinux[Downloads] git clone https://github.com/LANL-Bioinformatics/F
aQCs
zsh: correct 'git' to '_git' [nyae]? n
Cloning into 'FaQCs'...
remote: Counting objects: 193, done.
remote: Total 193 (delta 0), reused 0 (delta 0), pack-reused 193
Receiving objects: 100% (193/193), 7.27 MiB | 2.10 MiB/s, done.
Resolving deltas: 100% (94/94), done.
Checking connectivity... done.
iu@bielinux[Downloads] [ 4:09午後 ]
```



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)②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
FastQC fastqc_v0.11.4.zip
iu@bielinux[Downloads] git clone https://github.com/LANL-Bioinformatics/F
aQCs
zsh: correct 'git' to '_git' [nyae]? y
iu@bielinux[Downloads] git clone https://github.com/LANL-Bioinformatics/F
aQCs
zsh: correct 'git' to '_git' [nyae]? n
Cloning into 'FaQCs'...
remote: Counting objects: 193, done.
remote: Total 193 (delta 0), reused 0 (delta 0), pack-reused 193
Receiving objects: 100% (193/193), 7.27 MiB | 2.10 MiB/s, done.
Resolving deltas: 100% (94/94), done.
Checking connectivity... done.
iu@bielinux[Downloads] ls -lh
total 9.6M
drwxrwxr-x 6 iu iu 4.0K 12月 11 16:09 FaQCs
drwxrwxr-x 8 iu iu 4.0K 12月 10 10:09 FastQC
-rw-rw-r-- 1 iu iu 9.6M 10月 9 20:55 fastqc_v0.11.4.zip
iu@bielinux[Downloads]
```

[4:09午後]

[4:12午後]

①

②

W15-3: FaQCs

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

fix typo

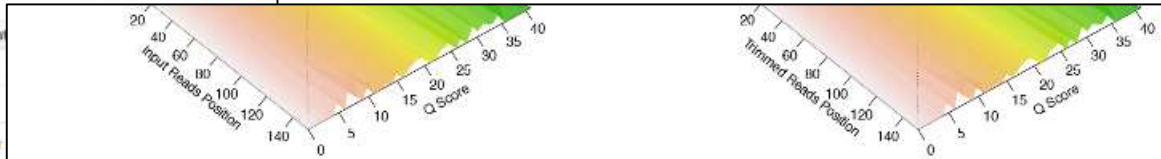
chianchi authored on Mar 31 | latest commit: faed85

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

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②libというディレクトリがあることを確認。③libディレクトリに移動し、④インストールマニュアル通りに./INSTALL.shを実行。

```
iu@bielinux[Downloads] pwd [ 4:42午後 ]
/home/iu/Downloads
iu@bielinux[Downloads] cd FaQCs [ 4:42午後 ]
iu@bielinux[FaQCs] pwd [ 4:42午後 ]
/home/iu/Downloads/FaQCs
iu@bielinux[FaQCs] ls [ 4:42午後 ]
COPYRIGHT example FaQCs.pl galaxy_module lib LICENSE README.md
iu@bielinux[FaQCs] cd lib [ 4:42午後 ]
iu@bielinux[lib] pwd [ 4:42午後 ]
/home/iu/Downloads/FaQCs/lib
iu@bielinux[lib] ls [ 4:42午後 ]
INSTALL.sh Parallel-ForkManager-1.03.tar.gz String-Approx-3.27.tar.gz
iu@bielinux[lib] ./INSTALL.sh [ 4:42午後 ]
```



W15-4: FaQCs

インストールが無事できた模様。①「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。

```
File Edit View Search Terminal Help
chmod 755 blib/arch/auto/String/Approx/Approx.150
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] [ 4:45午後 ]
```



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

①FaQCs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)。

```
String-Approx-3.27
-----
R is found.

All done!

iu@bielinux[lib] ls [ 4:45午後]
auto Parallel String
INSTALL.sh Parallel-ForkManager-1.03.tar.gz String-Approx-3.27.tar.gz
iu@bielinux[lib] cd .. [ 4:50午後]
iu@bielinux[FaQCs] pwd [ 4:50午後]
/home/iu/Downloads/FaQCs
iu@bielinux[FaQCs] ls [ 4:50午後]
COPYRIGHT example FaQCs.pl galaxy_module lib LICENSE README.md
iu@bielinux[FaQCs] [ 4:50午後]
```



W15-5: パスを通す

ここでは、①/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。

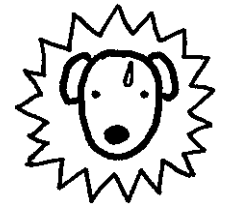
```
File Edit View Search Terminal Help
R is found.
All done!
iu@bielinux[lib] ls [ 4:45午後]
auto Parallel String
INSTALL.sh Parallel-ForkManager-1.03.tar.gz String-Approx-3.27.tar.gz
iu@bielinux[lib] cd .. [ 4:50午後]
iu@bielinux[FaQCs] pwd [ 4:50午後]
/home/iu/Downloads/FaQCs
iu@bielinux[FaQCs] ls [ 4:50午後]
COPYRIGHT example FaQCs.pl galaxy_module lib LICENSE README.md
iu@bielinux[FaQCs] sudo ln -s /home/iu/Downloads/FaQCs/FaQCs.pl /usr/local/bin/FaQCs.pl
[sudo] password for iu:
iu@bielinux[FaQCs] [ 4:54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 [ 4:58午後 ]
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/per
l/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 [ 4:58午後 ]
/usr/local/bin/FaQCs.pl
③ iu@bielinux[FaQCs] which FaQCs.pl [ 5:09午後 ]
/usr/local/bin/FaQCs.pl
iu@bielinux[FaQCs] █ [ 5:10午後 ]
```



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ParallelForkManager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!

```
File Edit View Search Terminal Help
① iu@bielinux[FaQCs] pwd
/home/iu/Downloads/FaQCs
iu@bielinux[FaQCs] ls
COPYRIGHT example FaQCs.pl galaxy_module lib LICENSE README.md
② iu@bielinux[FaQCs] ls -lh lib
total 80K
drwxrwxr-x 4 iu iu 4.0K 12月 11 16:45 auto
-rwxrwxr-x 1 iu iu 2.3K 12月 11 16:09 INSTALL.sh
drwxr-xr-x 2 iu iu 4.0K 12月 11 16:45 Parallel
-rw-rw-r-- 1 iu iu 14K 12月 11 16:09 Parallel-ForkManager-1.03.tar.gz
drwxr-xr-x 2 iu iu 4.0K 12月 11 16:45 String
-rw-rw-r-- 1 iu iu 45K 12月 11 16:09 String-Approx-3.27.tar.gz
③ iu@bielinux[FaQCs] ./FaQCs.pl -h
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
```


W15-7: 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 [ 5:20午後 ]
iu@bielinux[FaQCs] pwd [ 5:20午後 ]
/home/iu/Downloads/FaQCs
iu@bielinux[FaQCs] ls [ 5:20午後 ]
COPYRIGHT example FaQCs.pl galaxy_module lib LICENSE README.md
iu@bielinux[FaQCs] ~/Downloads/FaQCs/FaQCs.pl -h | head [ 5:20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_t
rim_read() function in bwaseqio.c)
iu@bielinux[FaQCs] [ 5:21午後 ]
```

W15-8: 考察

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、パスを通じた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!

```
iu@bielinux[FaQCs] pwd
/home/iu/Downloads/FaQCs
iu@bielinux[FaQCs] ls
COPYRIGHT example FaQCs.
iu@bielinux[FaQCs] ls -lh
total 80K
drwxrwxr-x 4 iu iu 4.0K 12月 11 16:09 Parallel-ForkManager-1.03.tar.gz
-rwxrwxr-x 1 iu iu 2.3K 12月 11 16:45 String
drwxr-xr-x 2 iu iu 4.0K 12月 11 16:09 String-Approx-3.27.tar.gz
-rw-rw-r-- 1 iu iu 14K 12月 11 16:09 String-Approx-3.27.tar.gz
drwxr-xr-x 2 iu iu 4.0K 12月 11 16:45 String
-rw-rw-r-- 1 iu iu 45K 12月 11 16:09 String-Approx-3.27.tar.gz
iu@bielinux[FaQCs] ./FaQCs.pl -h [ 5:15午後 ]
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
```


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クソ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上に置いている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「Parallell::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する

```
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 12月 11 16:45 auto
-rwxrwxr-x 1 iu iu 2.3K 12月 11 16:09 INSTALL.sh
drwxr-xr-x 2 iu iu 4.0K 12月 11 16:45 Parallel
-rw-rw-r-- 1 iu iu 14K 12月 11 16:09 Parallel-ForkManager-1.03.tar.gz
drwxr-xr-x 2 iu iu 4.0K 12月 11 16:45 String
-rw-rw-r-- 1 iu iu 45K 12月 11 16:09 String-Approx-3.27.tar.gz
iu@bielinux[FaQCs] ./FaQCs.pl -h [ 5:15午後 ]
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
```


W16-1: Perlモジュール

作業ディレクトリはどこでもよいので、
①「sudo cpan Parallel::ForkManager」。
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、②と③リターンキーを
押す。インストールが進んでいく。

```
File Edit View Search Terminal Help
iu@bielinux[FaQCs] sudo cpan Parallel::ForkManager
[sudo] password for iu:

CPAN.pm requires configuration, but most of it can be done automatically.
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]
Would you like me to automatically choose some CPAN mirror
sites for you? (This means connecting to the Internet) [yes]
Fetching with LWP:
http://www.perl.com/CPAN/authors/01mailrc.txt.gz
Reading '/home/iu/.cpan/sources/authors/01mailrc.txt.gz'
.....
...DONE
Fetching with LWP:
http://www.perl.com/CPAN/modules/02packages.details.txt.gz
```



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
All tests successful.

Test Summary Report
-----
t/waitpid-conflict.t      (Wstat: 0 Tests: 2 Failed: 0)
  TODO passed:    1-2
t/waitpid-waitonechild.t (Wstat: 0 Tests: 3 Failed: 0)
  TODO passed:    1-3
Files=10, Tests=24, 61 wallclock secs ( 0.05 usr  0.01 sys +  0.71 cusr
0.44 csys =  1.21 CPU)
Result: PASS
  YANICK/Parallel-ForkManager-1.17.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-1.17.tar.gz
  /usr/bin/make install -- OK
① iu@bielinux[FaQCs] sudo cpan String::Approx [ 5:34午後 ]
```


W16-3: FaQCs.pl

①「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②「FaQCs.pl -h | more」を実行すると...

```
File Edit View Search Terminal Help
t/aslice.t ... ok
t/asubst.t ... ok
t/user.t ..... ok
① All tests successful.
Files=7, Tests=140, 0 wallclock secs ( 0.03 usr  0.01 sys +  0.12 cusr
0.04 csys =  0.20 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 architecture
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
② iu@bielinux[FaQCs] FaQCs.pl -h | more [ 5:37午後 ]
```


W16-3: FaQCs.pl

```

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 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_t
rim_read() function in bwaseqio.c)
    -q          <INT> Targets # as quality level (default 5) fo
r trimming
    -5end       <INT> Cut # bp from 5 end before quality trimmi
ng/filtering
--More--

```

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 [ 5:42午後 ]
COPYRIGHT example FaQCs.pl galaxy_module lib LICENSE README.md
iu@bielinux[FaQCs] FaQCs.pl -h | head -n 4 [ 5:42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 [ 5:42午後 ]
iu@bielinux[iu] pwd [ 5:42午後 ]
/home/iu
② iu@bielinux[iu] FaQCs.pl -h | head -n 4 [ 5:42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] [ 5:42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

README.md

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

-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 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)

W17-1: 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、20分くらいかかるから。画面は実行直後の計算途中の状態。

```
File Edit View Search Terminal Help
iu@bielinux[iu] cd ~/Documents/srp017156
iu@bielinux[srp017156] pwd
/home/iu/Documents/srp017156
iu@bielinux[srp017156] ls -l *sub*
-rw-rw-r-- 1 iu iu 76659501 12月  9 15:24 SRR616268sub_1.fastq.gz
-rw-rw-r-- 1 iu iu 68682959 12月  9 15:24 SRR616268sub_2.fastq.gz
iu@bielinux[srp017156] time FaQCs.pl -p SRR616268sub_1.fastq.gz SRR616268
sub_2.fastq.gz -d result
Bwa extension trimming algorithm is used.
Processing SRR616268sub_1.fastq.gz SRR616268sub_2.fastq.gz file
```

[5:48午後]



①

W17-1: FaQCs

片側100万リードなので、①入力は計200万リード。②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。

```
File Edit View Search Terminal Help
iu@bielinux[iu] cd ~/Documents/srp01
iu@bielinux[srp017156] pwd
/home/iu/Documents/srp017156
iu@bielinux[srp017156] ls -l *sub*
-rw-rw-r-- 1 iu iu 76659501 12月 9
-rw-rw-r-- 1 iu iu 68682959 12月 9
iu@bielinux[srp017156] time FaQCs.pl
sub_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 77
7.24s user 10.61s system 89% cpu 14:37.72 total
iu@bielinux[srp017156] [ 6:03午後]
```

①

②

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。777.24秒(約13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、約135倍のオリジナルの1.35億リードだと、単純計算で $777.24 \times 135 = \text{約}30\text{時間}$ かかることになる。

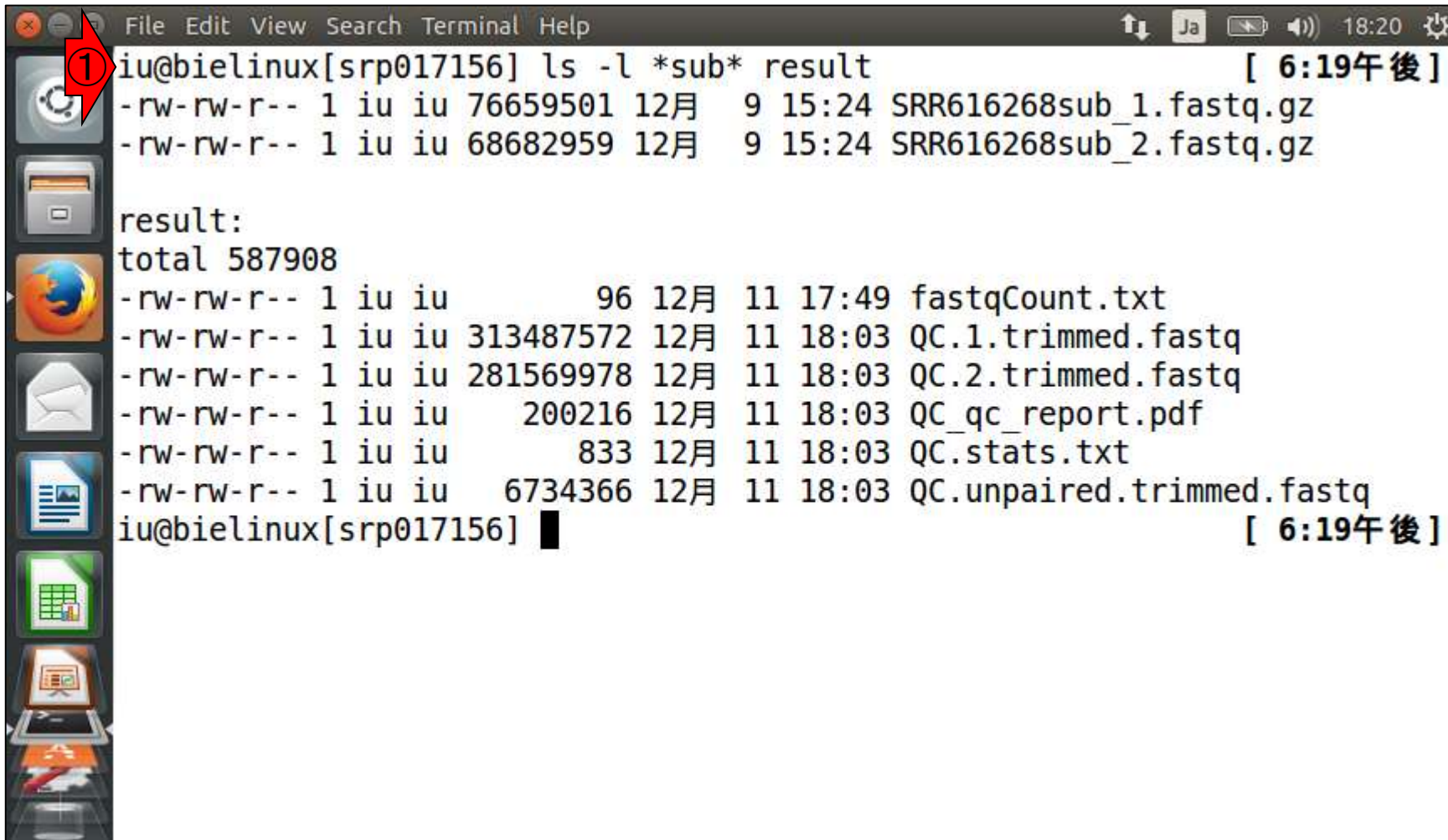
```
File Edit View Search Terminal Help
iu@bielinux[iu] cd ~/Documents/srp017156
iu@bielinux[srp017156] pwd
/home/iu/Documents/srp017156
iu@bielinux[srp017156] ls -l *sub*
-rw-rw-r-- 1 iu iu 76659501 12月  9 15:24 SRR616268sub_1.fastq.gz
-rw-rw-r-- 1 iu iu 68682959 12月  9 15:24 SRR616268sub_2.fastq.gz
iu@bielinux[srp017156] time FaQCs.pl -p SRR616268sub_1.fastq.gz SRR616268
sub_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 77
7.24s user 10.61s system 89% cpu 14:37.72 total
iu@bielinux[srp017156]
```

[5:48午後]

[6:03午後]



W17-1: FaQCs



```
iu@bielinux[srp017156] ls -l *sub* result [ 6:19午後 ]
-rw-rw-r-- 1 iu iu 76659501 12月 9 15:24 SRR616268sub_1.fastq.gz
-rw-rw-r-- 1 iu iu 68682959 12月 9 15:24 SRR616268sub_2.fastq.gz

result:
total 587908
-rw-rw-r-- 1 iu iu          96 12月 11 17:49 fastqCount.txt
-rw-rw-r-- 1 iu iu 313487572 12月 11 18:03 QC.1.trimmed.fastq
-rw-rw-r-- 1 iu iu 281569978 12月 11 18:03 QC.2.trimmed.fastq
-rw-rw-r-- 1 iu iu   200216 12月 11 18:03 QC_qc_report.pdf
-rw-rw-r-- 1 iu iu     833 12月 11 18:03 QC.stats.txt
-rw-rw-r-- 1 iu iu 6734366 12月 11 18:03 QC.unpaired.trimmed.fastq
iu@bielinux[srp017156] [ 6:19午後 ]
```

W17-1: FaQCs

- ①作業ディレクトリをresultに変更。
- ②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. 0.11.4)を実行。
- ④FastQC実行結果ファイル。FastQC実行結果をホストOS上で眺める場合は⑤のようにしてもよい。

```
iu@bielinux[srp017156] cd result
iu@bielinux[result] ls -l
total 587908
-rw-rw-r-- 1 iu iu          96 12月 11 17:49 fastqCount.txt
-rw-rw-r-- 1 iu iu 313487572 12月 11 18:03 QC.1.trimmed.fastq
-rw-rw-r-- 1 iu iu 281569978 12月 11 18:03 QC.2.trimmed.fastq
-rw-rw-r-- 1 iu iu   200216 12月 11 18:03 QC_qc_report.pdf
-rw-rw-r-- 1 iu iu     833 12月 11 18:03 QC.stats.txt
-rw-rw-r-- 1 iu iu 6734366 12月 11 18:03 QC.unpaired.trimmed.fastq
iu@bielinux[result] fastqc2 -q QC.1.trimmed.fastq [ 6:22午後 ]
iu@bielinux[result] ls -l [ 6:22午後 ]
total 588672
-rw-rw-r-- 1 iu iu          96 12月 11 17:49 fastqCount.txt
-rw-rw-r-- 1 iu iu 313487572 12月 11 18:03 QC.1.trimmed.fastq
-rw-rw-r-- 1 iu iu   363785 12月 11 18:22 QC.1.trimmed_fastqc.html
-rw-rw-r-- 1 iu iu   416251 12月 11 18:22 QC.1.trimmed_fastqc.zip
-rw-rw-r-- 1 iu iu 281569978 12月 11 18:03 QC.2.trimmed.fastq
-rw-rw-r-- 1 iu iu   200216 12月 11 18:03 QC_qc_report.pdf
-rw-rw-r-- 1 iu iu     833 12月 11 18:03 QC.stats.txt
-rw-rw-r-- 1 iu iu 6734366 12月 11 18:03 QC.unpaired.trimmed.fastq
iu@bielinux[result] cp QC.1.trimmed_fastqc.html ~/Desktop/mac_share/
```

W17-2: FastQCで確認

① Overrepresented sequencesを眺めて Possible Sourceのところが全てNo Hitになっているかどうかをチェック

FastQC Report

Fri 11 Dec 2015
QC.1.trimmed.fastq

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](#)

✗ Overrepresented sequences

Sequence	Count	Percentage	Possible Source
CCCCGGTATATTTTCGGCGCAGTGCCACTCGACTAGTGAGCTATTACGCA	14289	1.4620400620873335	No Hit
GGCCTATTCACTGCGGCTGACCTTGCGGTCAGCACCCCTTCTTCCGAAGT	10984	1.123874871717214	No Hit
GTGCTTTTCACCTTTCCCTCACGGTACTGGTTCACTATCGGTCACTAGGG	8746	0.8948843434121226	No Hit
CCCCGGTATATTTTCGGCGCAGTGCCACTCGACTAGTGAGCTATTACGCAC	8447	0.8642908814088954	No Hit
GCCGGCATTCTCACTTCTAAGCGCTCCAGCCGCTCCTCAGATCGACCTTC	8081	0.8268420282544434	No Hit
GTCCTAGGGAGTATTTAGCCTTGGGAGATGGTCTCCCGGATCCGACG	7943	0.8127219688683386	No Hit
GTCCAGTCTACAACCCGAGAAGCAAGCTTCTCGGTTTGGGCTCTTCCC	6618	0.6771489349075495	No Hit
GTCGGTTTGGGTACGGTAGTTTATTTCTCACTAGAAGCTTTTCTTGGC	6373	0.6520807135336677	No Hit
GGTCACTTGGTTTCGGGTCTACATCTGCTTACTCATTCCGCCTGTTGAGA	5461	0.5587655384602792	No Hit
GCCGGCATTCTCACTTCTAAGCGCTCCAGCCGCTCCTCAGATCAACCTTC	4804	0.49154177746991046	No Hit
CCGGTATATTTTCGGCGCAGTGCCACTCGACTAGTGAGCTATTACGCACT	4378	0.44795376806063025	No Hit
CCCTCCATCGCTTAAACAAAATAAACTAGTGAGGAATCTCAACCTGCTT	4342	0.44427027430773336	No Hit
CCCGCGTCTGCCCGGCCAGCTATGTATTCAGTACAGCAATACACTG	4335	0.44355403941133675	No Hit
CCACAGTTTCGGTATTATGCTTAGCCCCGGTATATTTTCGGCGCAGTGCC	4292	0.4391543107620432	No Hit
CTGGGCTGTTCCCTTTTCGACAATGGACCTTATCGCTCACTGTCTGACTC	4088	0.41828117949562743	No Hit
CCGGCATTCTCACTTCTAAGCGCTCCAGCCGCTCCTCAGATCGACCTTCA	3839	0.3928036810380904	No Hit
CCGCCGTAATCAGGATCCTGGACGGAGGGTTCGACGTTTCGCTTACAGGG	3837	0.3925990424962628	No Hit
GTAGGTCACTTGGTTTCGGGTCTACATCTGCTTACTCATTCCGCCTGTTT	3801	0.38891554874336587	No Hit
GGAGTATTTAGCCTTGGGAGATGGTCTCCCGGATTCGACGGAATTTCA	3634	0.37182823050076075	No Hit

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		Fri 11 Dec 2015 QC.1.trimmed.fastq		
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				
	CTTTCGACAATGGACCTTATCGCTCACTGTCTGACTCCCGGAGTAAGATC	1444	0.14774902719953178	No Hit
	ATTTAGCCTTGGGAGATGGTCTCCCGGATTCGACGGAATTTACAGTGT	1423	0.1456003225103419	No Hit
	CGCTACTCATGCCGGCATTCTCACTTCTAAGCGCTCCAGCCGCTCCTCACG	1398	0.14304234073749686	No Hit
	CCGGGGTGCTTTTTCACCTTTCCCTCACGGTACTGGTTCACTATCGGTCAC	1349	0.13802869646272048	No Hit
	GGGCCTATTCAGTCCGGCTGACCTTCCGGTCAGCACCCCTTCTCCGAAG	1341	0.13721014229541006	No Hit
	CTGGTGATCTGGGCTGTTCCCTTTTCGACAATGGACCTTATCGCTCACTG	1336	0.13669854594084105	No Hit
	GCCGCCGCCAGCTATGTATTCAGTACAAGCAATACACTGATGTGTACT	1319	0.13495911833530638	No Hit
	GCCGAGTTCCTTAACGAGAGTTCGCTCGCTCACCTGAGGATACTCTCCTC	1305	0.13352664854251314	No Hit
	AGATCGGAAGAGCACACGTCTGAACTCCAGTCACCGATGTATCTCGTATG	1296	0.13260577510428892	TruSeq Adapter, Index 2 (100% over 49bp)
	GTTTGGGCTCTTCCCACTTCGCTCGCCGCTACTATGGGAATCGAGTTTTT	1262	0.12912691989321962	No Hit
	CGTCCCTCCATCGTTAAACAAAATAAACTAGTGCAGGAATCTCAACCTG	1258	0.12871764280956438	No Hit
	CACACGGTTTCAGGAACTGTTTCACTCCCTTCCGGGGTGCTTTTTCACCT	1257	0.1286153235386506	No Hit
	CCCTAGTTCAAACAGTGTCTACCTCCATGATCCATCCTCCGAGGCTAAC	1249	0.12779676937134016	No Hit
	GTCATTTTGGCGAGTTCCTTAACGAGAGTTCGCTCGCTCACCTGAGGATA	1249	0.12779676937134016	No Hit
	ACCTTAACTGGTGATCTGGGCTGTTCCCTTTTCGACAATGGACCTTATCG	1223	0.12513646883275813	No Hit
	CACAGTTTCGGTATTATGCTTAGCCCCGGTATATTTTCGGCGCAGTGCCA	1220	0.12482951051483988	No Hit
	GGTCATTTTGGCGAGTTCCTTAACGAGAGTTCGCTCGCTCACCTGAGGAT	1208	0.12360167926387423	No Hit
	CCACTTAGCATAAATTTAGGGACCTTAACTGGTGATCTGGGCTGTTCCCT	1207	0.12349935999296043	No Hit
	CCGGTTCATTCTACAAAAGGCACGCCATTACCCGTTAACGGGCTTTGACT	1191	0.12186225165833958	No Hit
	CTGCCGCCGCCAGCTATGTATTCAGTACAAGCAATACACTGATGTGTA	1178	0.12053210113646014	No Hit



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

fix typo

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

```
$ 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 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 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で妥当。しかし、真実は不明

```
iu@bielinux[result] cd ..
iu@bielinux[srp017156] pwd
/home/iu/Documents/srp017156
iu@bielinux[srp017156] ls
hoge2.txt          JSLAB4_2_Linux.sh  SRR616268_1.fastq.bz2
hoge3.txt          JSLAB4_3_Linux.sh  SRR616268_2.fastq.bz2
hoge4.txt          JSLAB4_6_Linux.sh  SRR616268sub_1.fastq.gz
hoge8.fastq        JSLAB4_7_Linux.sh  SRR616268sub_2.fastq.gz
hoge.txt           JSLAB4_9_Win.txt
JSLAB4_1_Linux.sh result
① iu@bielinux[srp017156] time FaQCs.pl -adapter -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 -p SRR616268sub_1.fastq.gz SRR616268sub_2.fastq.gz -d
1196.86s user 10.76s system 93% cpu 21:35.82 total
iu@bielinux[srp017156] █
```

[6:39午後]

[7:01午後]



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](#)

BROAD INSTITUTE
© 2013 Broad Institute

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 prompt.

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 users with Java 6 (JRE 1.6) should use the binary distribution archive and the Java Web Start buttons below.**

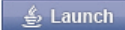
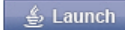
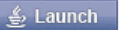
2. Java Web Start

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。東大有線LAN環境でも30秒程度かかる。
- ②IGV_2.3.67.zipは約27MB。

```
iu@bielinux[srp017156] cd ~/Downloads [ 7:29午後 ]
iu@bielinux[Downloads] ls [ 7:29午後 ]
FaQCs FastQC fastqc_v0.11.4.zip
① iu@bielinux[Downloads] wget -cq http://data.broadinstitute.org/igv/projec
ts/downloads/IGV_2.3.67.zip
iu@bielinux[Downloads] ls -lh [ 7:30午後 ]
total 37M
drwxrwxr-x 6 iu iu 4.0K 12月 11 16:09 FaQCs
drwxrwxr-x 8 iu iu 4.0K 12月 10 10:09 FastQC
-rw-rw-r-- 1 iu iu 9.6M 10月 9 20:55 fastqc_v0.11.4.zip
-rw-rw-r-- 1 iu iu 27M 12月 1 12:10 IGV_2.3.67.zip ②
iu@bielinux[Downloads] █ [ 7:30午後 ]
```


W18-2: 解凍

①zipファイルの解凍はunzip。②IGV_2.3.67というディレクトリが作成されたことがわかる。

```
iu@bielinux[Downloads] pwd [ 7:32午後 ]
/home/iu/Downloads
① iu@bielinux[Downloads] unzip IGV_2.3.67.zip [ 7:32午後 ]
Archive:  IGV_2.3.67.zip
  creating:  IGV_2.3.67/
  inflating: IGV_2.3.67/igv.bat
  inflating: IGV_2.3.67/igv.command
  inflating: IGV_2.3.67/igv.sh
  inflating: IGV_2.3.67/batik-codec_V1.7.jar
  inflating: IGV_2.3.67/goby-io-igv_V1.0.jar
  inflating: IGV_2.3.67/igv.jar
  inflating: IGV_2.3.67/readme.txt
② iu@bielinux[Downloads] ls -lh [ 7:32午後 ]
total 37M
drwxrwxr-x 6 iu iu 4.0K 12月 11 16:09 FaQCs
drwxrwxr-x 8 iu iu 4.0K 12月 10 10:09 FastQC
-rw-rw-r-- 1 iu iu 9.6M 10月  9 20:55 fastqc_v0.11.4.zip
drwxr-xr-x 2 iu iu 4.0K 11月 30 22:10 IGV_2.3.67
-rw-rw-r-- 1 iu iu 27M 12月  1 12:10 IGV_2.3.67.zip
iu@bielinux[Downloads] [ 7:32午後 ]
```

W18-3: 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_2.3.67 [ 7:36午後]
iu@bielinux[IGV_2.3.67] pwd [ 7:36午後]
/home/iu/Downloads/IGV_2.3.67
iu@bielinux[IGV_2.3.67] ls -l [ 7:36午後]
total 30052
-rw-r--r-- 1 iu iu 178588 11月 30 22:10 batik-codec_V1.7.jar
-rw-r--r-- 1 iu iu 2121484 11月 30 22:10 goby-io-igv_V1.0.jar
-rwxr-xr-x 1 iu iu 258 11月 30 22:10 igv.bat
-rwxr-xr-x 1 iu iu 630 11月 30 22:10 igv.command
-rw-r--r-- 1 iu iu 28454734 11月 30 22:10 igv.jar
-rwxr-xr-x 1 iu iu 427 11月 30 22:10 igv.sh
-rw-r--r-- 1 iu iu 1238 11月 30 22:10 readme.txt
① iu@bielinux[IGV_2.3.67] sudo ln -s /home/iu/Downloads/IGV_2.3.67/igv.sh /
usr/local/bin
[sudo] password for iu:
iu@bielinux[IGV_2.3.67] [ 7:36午後]
```

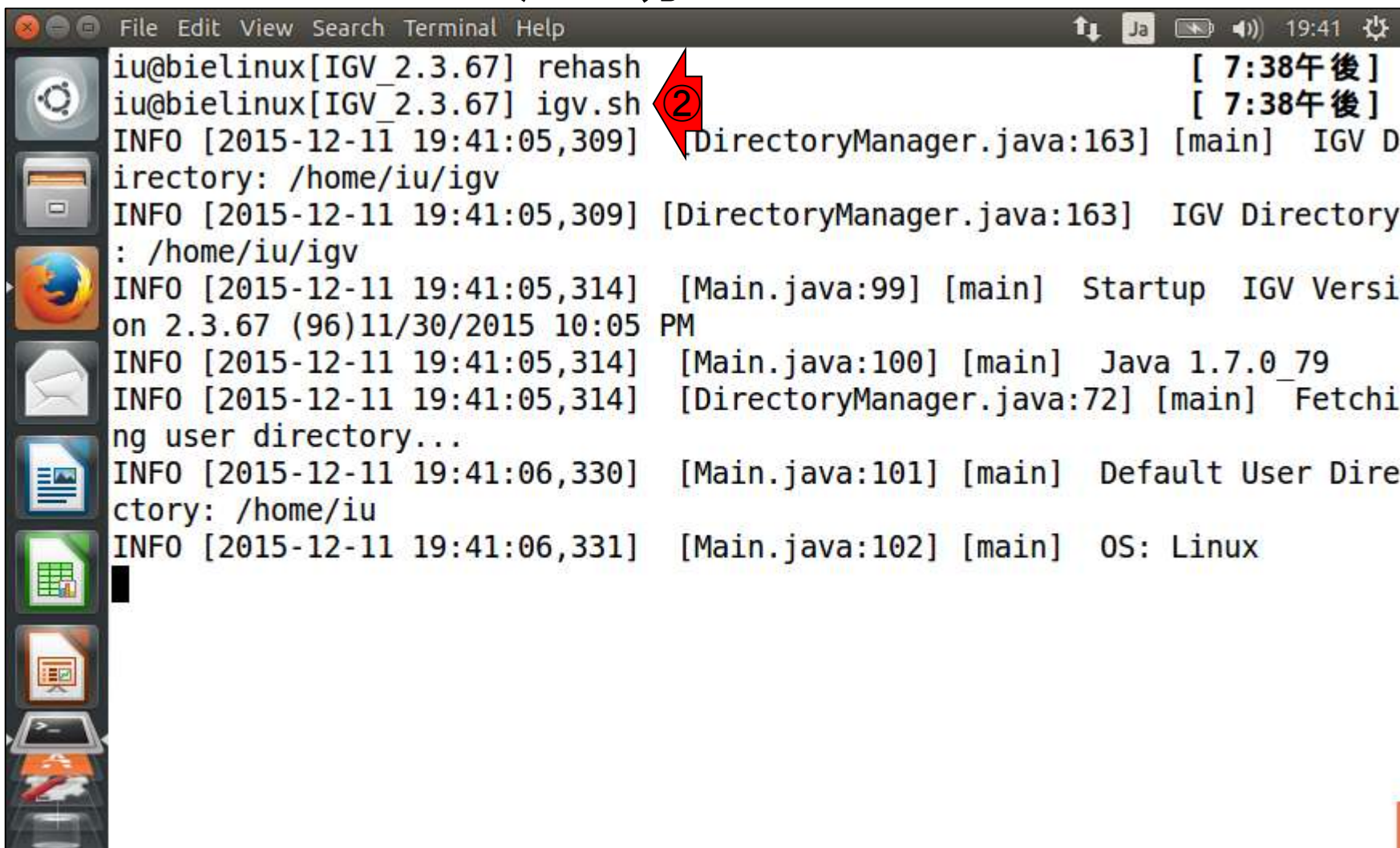
W18-3: IGVパスを通す



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。「igv.sh」と打ってダメだったらrehashしてみる的な軽いノリでいいです。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。

②リターンキーを押して、少し経過
(数秒程度)したときの状態です

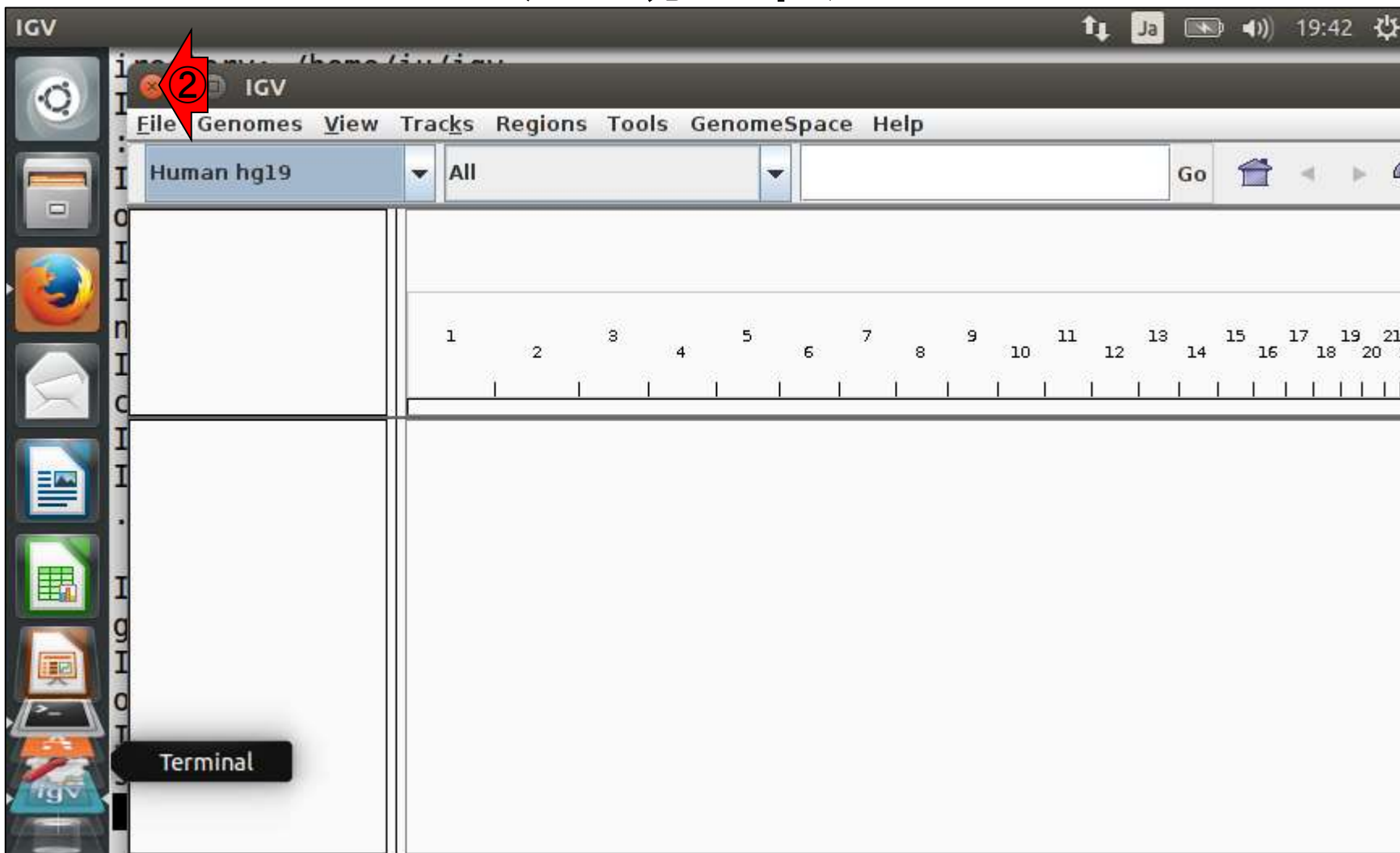
W18-4: IGV起動



```
iu@bielinux[IGV_2.3.67] rehash [ 7:38午後]
iu@bielinux[IGV_2.3.67] igv.sh [ 7:38午後]
INFO [2015-12-11 19:41:05,309] [DirectoryManager.java:163] [main] IGV D
irectory: /home/iu/igv
INFO [2015-12-11 19:41:05,309] [DirectoryManager.java:163] IGV Directory
: /home/iu/igv
INFO [2015-12-11 19:41:05,314] [Main.java:99] [main] Startup IGV Versi
on 2.3.67 (96)11/30/2015 10:05 PM
INFO [2015-12-11 19:41:05,314] [Main.java:100] [main] Java 1.7.0_79
INFO [2015-12-11 19:41:05,314] [DirectoryManager.java:72] [main] Fetchi
ng user directory...
INFO [2015-12-11 19:41:06,330] [Main.java:101] [main] Default User Dire
ctory: /home/iu
INFO [2015-12-11 19:41:06,331] [Main.java:102] [main] OS: Linux
```

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。



W18-5: IGV終了後

```
File Edit View Search Terminal Help 19:42
: /home/iu/igv
INFO [2015-12-11 19:41:05,314] [Main.java:99] [main] Startup IGV Versi
on 2.3.67 (96)11/30/2015 10:05 PM
INFO [2015-12-11 19:41:05,314] [Main.java:100] [main] Java 1.7.0_79
INFO [2015-12-11 19:41:05,314] [DirectoryManager.java:72] [main] Fetchi
ng user directory...
INFO [2015-12-11 19:41:06,330] [Main.java:101] [main] Default User Dire
ctory: /home/iu
INFO [2015-12-11 19:41:06,331] [Main.java:102] [main] OS: Linux
INFO [2015-12-11 19:41:24,112] [GenomeManager.java:1017] [main] java.io
.FileNotFoundException: /home/iu/igv/genomes/user-defined-genomes.txt (No
such file or directory)
INFO [2015-12-11 19:41:24,939] [GenomeManager.java:145] [main] Loading
genome: /home/iu/igv/genomes/hg19.genome
INFO [2015-12-11 19:41:26,900] [GenomeManager.java:192] [main] Genome l
oaded. id= hg19
INFO [2015-12-11 19:41:27,215] [CommandListener.java:106] [Thread-6] Li
stening on port 60151
INFO [2015-12-11 19:42:37,761] [ShutdownThread.java:51] [Thread-2] Shut
ting down
iu@bielinux[IGV_2.3.67] [ 7:42午後]
```

