



## 次世代シーケンサーデータの解析手法 第5回: ウェブ資料



東京大学・大学院農学生命科学研究科

孫建強、清水謙多郎、門田 幸二

[kadota@bi.a.u-tokyo.ac.jp](mailto:kadota@bi.a.u-tokyo.ac.jp)

<http://www.iu.a.u-tokyo.ac.jp/~kadota/>

# W1-1: FaQCs実行

第4回[W17-3]と基本的に同じ。ここでは①ls実行結果で2つのgzip圧縮ファイル(\*.fastq.gz)のみが見えるように余分なファイルを予め削除している。ファイルサイズが多少違っていても気にしない。②第4回[W9-1]で示すように、2015年10月9日以降にFastQCをインストールしたヒトは、ここがver. 0.11.4以上になっている。③FaQCsの実行。result2ディレクトリに結果を保存。

```
iu@bielinux[srp017156] pwd
/home/iu/Documents/srp017156
iu@bielinux[srp017156] ls -lh
total 139M
-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] fastqc2 -v
FastQC v0.11.4
iu@bielinux[srp017156] FaQCs.pl -v
Version: 1.34
iu@bielinux[srp017156] time FaQCs.pl -adapter -p SRR616268sub_1.fastq.gz SRR6
16268sub_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 Over
all quality 36.37
(99.33, 8.62, 107.0, 107, 50)
FaQCs.pl -adapter -p SRR616268sub_1.fastq.gz SRR616268sub_2.fastq.gz -d 113
4.61s user 11.10s system 96% cpu 19:41.18 total
iu@bielinux[srp017156] █
```

[11:57午前]

[11:57午前]

[12:17午後]

# W1-1: FaQCs実行

④ result2ディレクトリをlsしている。  
QC.1.trimmed.fastqとQC.2.trimmed.fastqが  
FaQCsの主な実行結果ファイル

```
iu@bielinux[srp017156] pwd [11:57午前]
/home/iu/Documents/srp017156
① iu@bielinux[srp017156] ls -lh [11:57午前]
total 139M
-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
② iu@bielinux[srp017156] fastqc2 -v [11:57午前]
FastQC v0.11.4
iu@bielinux[srp017156] FaQCs.pl -v [11:57午前]
Version: 1.34
③ iu@bielinux[srp017156] time FaQCs.pl -adapter -p SRR616268sub_1.fastq.gz SRR6
16268sub_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 Over
all quality 36.37
(99.33, 8.62, 107.0, 107, 50)
FaQCs.pl -adapter -p SRR616268sub_1.fastq.gz SRR616268sub_2.fastq.gz -d 113
4.61s user 11.10s system 96% cpu 19:41.18 total
④ iu@bielinux[srp017156] ls result2 [12:17午後]
fastqCount.txt QC.2.trimmed.fastq QC.stats.txt
QC.1.trimmed.fastq QC_qc_report.pdf QC.unpaired.trimmed.fastq
iu@bielinux[srp017156] [12:25午後]
```

# W1-2: FastQC実行

①forward側、②reverse側のFaQCs実行結果ファイルをFastQC (ver. 0.11.4)の入力として実行

```
iu@bielinux[srp017156] pwd [11:55午前]
/home/iu/Documents/srp017156
iu@bielinux[srp017156] ls result2/*.fastq [11:57午前]
result2/QC.1.trimmed.fastq result2/QC.unpaired.trimmed.fastq
result2/QC.2.trimmed.fastq
① iu@bielinux[srp017156] fastqc2 -q result2/QC.1.trimmed.fastq --outdir=/home/i
u/Desktop/mac_share
② iu@bielinux[srp017156] fastqc2 -q result2/QC.2.trimmed.fastq --outdir=/home/i
u/Desktop/mac_share
iu@bielinux[srp017156] date [11:58午前]
2015年 12月 20日 日曜日 12:04:40 JST
iu@bielinux[srp017156] [12:04午後]
```

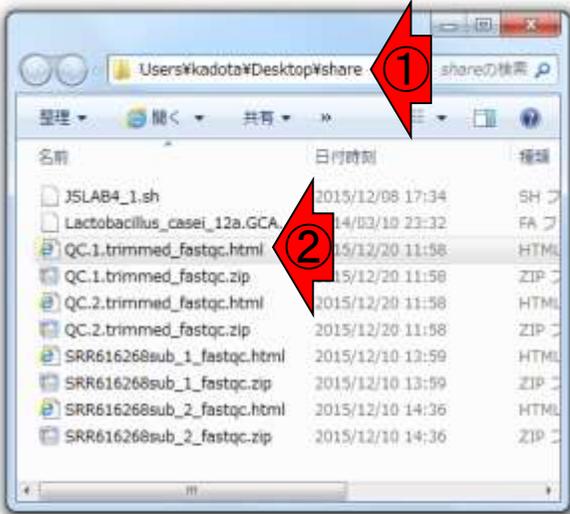
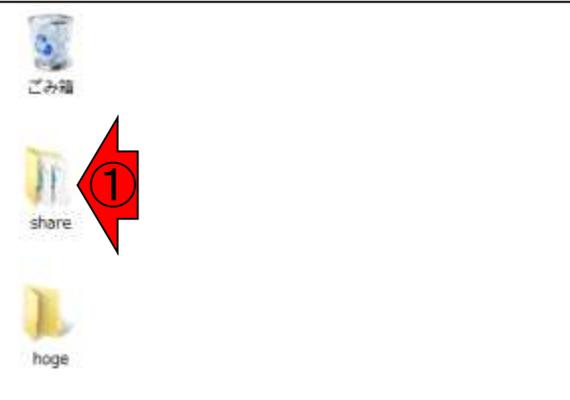
# W1-2: FastQC実行

①dateで日付を表示。②保存先として指定した共有フォルダ(/home/iu/Desktop/mac\_share)中に、確かにFastQC実行結果ファイルが存在することがわかる。

```
iu@bielinux[srp017156] pwd [11:55午前]
/home/iu/Documents/srp017156
iu@bielinux[srp017156] ls result2/*.fastq [11:57午前]
result2/QC.1.trimmed.fastq result2/QC.unpaired.trimmed.fastq
result2/QC.2.trimmed.fastq
iu@bielinux[srp017156] fastqc2 -q result2/QC.1.trimmed.fastq --outdir=/home/i
u/Desktop/mac_share
iu@bielinux[srp017156] fastqc2 -q result2/QC.2.trimmed.fastq --outdir=/home/i
u/Desktop/mac_share
① iu@bielinux[srp017156] date [11:58午前]
2015年 12月 20日 日曜日 12:04:40 JST
② iu@bielinux[srp017156] ls -lh /home/iu/Desktop/mac_share/QC.* [12:04午後]
-rwxrwxrwx 1 iu iu 356K 12月 20 11:58 /home/iu/Desktop/mac_share/QC.1.trimmed
_fastqc.html
-rwxrwxrwx 1 iu iu 404K 12月 20 11:58 /home/iu/Desktop/mac_share/QC.1.trimmed
_fastqc.zip
-rwxrwxrwx 1 iu iu 333K 12月 20 11:58 /home/iu/Desktop/mac_share/QC.2.trimmed
_fastqc.html
-rwxrwxrwx 1 iu iu 373K 12月 20 11:58 /home/iu/Desktop/mac_share/QC.2.trimmed
_fastqc.zip
iu@bielinux[srp017156] [12:07午後]
```

# W1-3: FastQCで確認

①ホストOS上の共有フォルダshare上で、②forward側のFastQC実行結果ファイル(QC.1.trimmed\_fastqc.html)を眺める。



**FastQC Report** Sun 20 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

**Basic Statistics**

Measure	Value
Filename	QC.1.trimmed.fastq
File type	Conventional base calls
Encoding	Sanger / Illumina 1.9
Total Sequences	977202
Sequences flagged as poor quality	0
Sequence length	50-107
%GC	50

**Per base sequence quality**

Quality scores across all bases (Sanger / Illumina 1.9 encoding)

# W1-3: FastQCで確認

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

Sun 20 Dec 2015  
QC.1.trimmed.fastq

## 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](#)

### Overrepresented sequences

Sequence	Count	Percentage	Possible Source
CCCCGGTATATTTTCGGCGCAGTGCCACTCGACTAGTGAGCTATTACGCA	14289	1.4622360576421252	No Hit
GGCCTATTCAGTGGGCTGACCTTGGCGTCAGCACCCCTTCTTCCGAAGT	10984	1.124025534127028	No Hit
GTGCTTTTCACCTTTCCTCACGGTACTGGTTCATATCGGTCAGTAGGG	8746	0.8950043082187716	No Hit
CCCCGGTATATTTTCGGCGCAGTGCCACTCGACTAGTGAGCTATTACGCAC	8446	0.8643044119844209	No Hit
GCCGGCATTCTCACTTCTAAGCGCTCCAGCCGTCCTCAGGATCGACCTTC	8081	0.8269528715659608	No Hit
GTCAGTGGGAGTATTTAGCCTTGGGAGATGGTCTCCCGGATTCCGACG	7943	0.8128309192981594	No Hit
GTCCAGTCTACAACCCCGAGAAGCAAGCTTCTCGGTTTGGGCTCTTCCC	6618	0.6772397109297771	No Hit
GTCGGTTTGGGTCAGGGTAGTTTATTTCTCACTAGAAGCTTTTCTTGGC	6373	0.6521681290050573	No Hit
GGTCACTTGGTTTGGGCTACATCTGCTTACTCATTGCCCCTGTTTCCAG	5461	0.558840444452631	No Hit
GCCGGCATTCTCACTTCTAAGCGCTCCAGCCGTCCTCAGGATCAACCTTC	4804	0.491607671699403	No Hit
CCGGTATATTTTCGGCGCAGTGCCACTCGACTAGTGAGCTATTACGCACT	4378	0.44801381904662496	No Hit
CCCTCCATCGCTTAAACAAAATAAACTAGTGCAGGAATCTCAACCTGCTT	4342	0.44432983149850286	No Hit
CCCGCTGTCGCGCCGCGCCAGCTATGTATTCACTGACAAGCAATACACTG	4335	0.443613500586368	No Hit
CCACAGTTTCCGTATTATGCTTAGCCCGGTATATTTTCGGCGCAGTGCC	4292	0.4392131821261111	No Hit
CTGGGCTGTTCCCTTTTCGACAATGGACCTTATCGCTCACTGTCTGACTC	4087	0.41823491969930476	No Hit
CCGGCATTCTCACTTCTAAGCGCTCCAGCCGTCCTCAGGATCGACCTTCA	3839	0.39285633881224147	No Hit



# W1-3: FastQCで確認

①連載第4回W17-2で見られたアダプター配列がなくなっていることがわかる。これを含め、全てNo Hitになっていたことから、-adapterオプションがうまく機能していることがわかる

## 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](#)

CGCTACTCATGCCGGCATTCTCACTTCTAAGCGCTCCAGCCGCTCCTCACG	1398	0.14306151645207438	No Hit
CCGGGGTGCTTTTACCTTTCCCTCACGGTACTGGTTCATATCGGTCAC	1349	0.13804720006713045	No Hit
GGGCCTATTCAGTGCAGGCTGACCTTGCAGTCCAGCACCCTTCTTCCGAAG	1341	0.13722853616754777	No Hit
CTGGTGATCTGGGCTGTTCCCTTTCCGACAATGGACCTATCGCTCACTG	1336	0.13671687123030857	No Hit
GCCGCCGGCCAGCTATGTATTCAGTGCAGCAAGCAATACACTGATGTGTACT	1319	0.13497721044369537	No Hit
GCCGAGTTCCTTAACGAGAGTTCGCTCGCTCACCTGAGGATACTCTCCTC	1305	0.13354454861942566	No Hit
GTTTGGGCTCTTCCCACTTCGCTCGCCGCTACTATGGGAATCGAGTTTTT	1262	0.12914423015916873	No Hit
CGTCCCTCCATCGCTTAAACAAAATAAACTAGTGCAGGAATCTCAACCTG	1258	0.1287348982093774	No Hit
CACACGGTTTCAGGAACTGTTTCACTCCCTTCCGGGGTGCTTTTACCT	1257	0.12863256522192956	No Hit
CCCTAGTTCAAACAGTGTCTACCTCCATGATCCATCCTCCGAGGCTAAC	1249	0.12781390132234688	No Hit
GTCATTTTGCCGAGTTCCTTAACGAGAGTTCGCTCGCTCACCTGAGGATA	1249	0.12781390132234688	No Hit
ACCTTAACTGGTGATCTGGGCTGTTCCCTTTCCGACAATGGACCTTATCG	1223	0.12515324364870312	No Hit
CACAGTTTTCGGTATTATGCTTAGCCCCGGTATATTTTCGGCGCAGTGCCA	1220	0.12484624468635963	No Hit
GGTCATTTTGCCGAGTTCCTTAACGAGAGTTCGCTCGCTCACCTGAGGAT	1208	0.1236182488369856	No Hit
CCACTTAGCATAAATTTAGGGACCTTAACTGGTGATCTGGGCTGTTCCCT	1207	0.12351591584953776	No Hit
CCGGTTCATTCTACAAAAGGCACGCCATTACCCGTTAACGGGCTTTGACT	1191	0.1218785880503724	No Hit
CTGCCGCCGGCCAGCTATGTATTCAGTGCAGCAAGCAATACACTGATGTGTA	1178	0.12054825921355053	No Hit
CCCTACTCACCATCCTCCACCCACCTTCCACCTTTCCCTTACACCCCT	1174	0.12012892726275017	No Hit



# W2-1 : Rockhopper

①Download。Bio-Linux8の場合は、②「Rockhopper for any platform」でよい



## ROCKHOPPER

HOME

Download

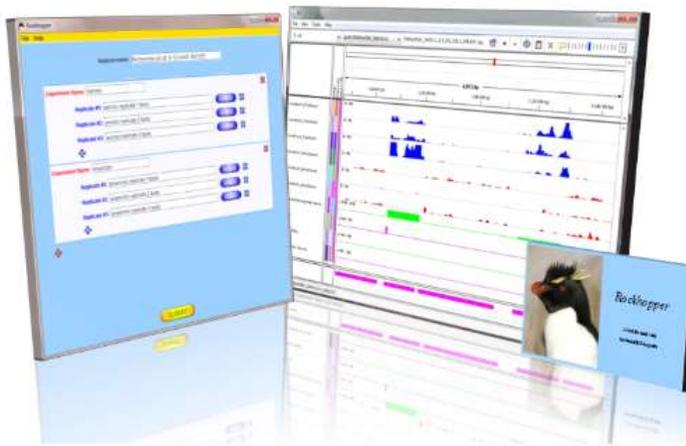
User Guide

FAQ

### A system for analyzing bacterial RNA-seq data

Rockhopper is a comprehensive and user-friendly system for computational analysis. As input, Rockhopper takes RNA sequencing reads output by high-throughput seq (QSEQ, FASTA, SAM, or BAM files). Rockhopper supports the following tasks:

- Reference based transcript assembly (when one or more reference genomes)
  - Aligning reads to genomes
  - Assembling transcripts
  - Identifying transcript boundaries and novel transcripts such as small RNAs
- *De novo* transcript assembly (when reference genomes are unavailable)
- Normalizing data from different experiments
- Quantifying transcript abundance
- Testing for differential gene expression
- Characterizing operon structures
- Visualizing results in a genome browser



### System Requirements

Rockhopper is implemented in Java, so you must have Java installed on your computer to use Rockhopper. To check if Java is installed on your computer, type "java -version" at any shell or terminal window or command prompt:

```
C:\>java -version
java version "1.7.0_11"
Java(TM) SE Runtime Environment (build 1.7.0_11-b21)
Java HotSpot(TM) 64-Bit Server VM (build 23.6-b04, mixed mode)
```

It is recommended that your computer has Java version 1.6 or later and your computer has at least 2 gigabytes of RAM. If you do not have Java or you need to update to a more recent version, you can do so by clicking the Java icon on the right:



### Download Latest Release (Rockhopper version 2.0.3)

Rockhopper for Windows



Rockhopper for Mac



Error opening Rockhopper on Mac?

Rockhopper for any platform



To execute the GUI version of Rockhopper, use the following command:  
java -Xmx1200m -jar Rockhopper.jar

To execute the command line version of Rockhopper, use the following command:  
java -Xmx1200m -cp Rockhopper.jar Rockhopper

Rockhopper source code



To extract the source code from the JAR file, use the following command:  
jar xf Rockhopper.jar



To extract the source code from the compressed TAR archive, use the following command:  
tar xjf Rockhopper-2.0.3.tar.bz2

# W2-1 : Rockhopper

基本的には右クリックで②「対象をファイルに保存」でよいが、wgetコマンドを利用したい場合は③「ショートカットのコピー (Windowsの場合)」でURL情報を取得する



## ROCKHOPPER

HOME

Download

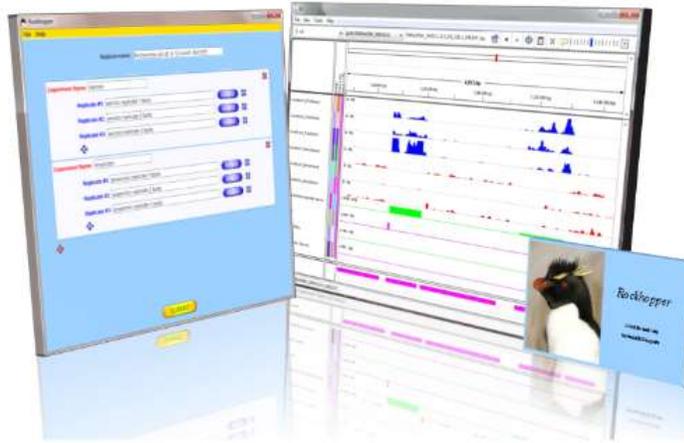
User Guide

FAQ

### A system for analyzing bacterial RNA-seq data

Rockhopper is a comprehensive and user-friendly system for computational analysis. As input, Rockhopper takes RNA sequencing reads output by high-throughput seq (QSEQ, FASTA, SAM, or BAM files). Rockhopper supports the following tasks:

- Reference based transcript assembly (when one or more reference genomes are available)
  - Aligning reads to genomes
  - Assembling transcripts
  - Identifying transcript boundaries and novel transcripts such as small RNAs
- *De novo* transcript assembly (when reference genomes are unavailable)
- Normalizing data from different experiments
- Quantifying transcript abundance
- Testing for differential gene expression
- Characterizing operon structures
- Visualizing results in a genome browser



### System Requirements

Rockhopper is implemented in Java, so you must have Java installed on your computer to use Rockhopper. To check if Java is installed on your computer, type "java -version" at any shell or terminal window or command prompt:

```
C:\>java -version
java version "1.7.0_11"
Java(TM) SE Runtime Environment (build 1.7.0_11-b21)
Java HotSpot(TM) 64-Bit Server VM (build 23.6-b04, mixed mode)
```

It is recommended that your computer has Java version 1.6 or later and your computer has at least 2 gigabytes of RAM. If you do not have Java or you need to update to a more recent version, you can do so by clicking the Java icon on the right:



### Download Latest Release (Rockhopper version 2.0.3)

Rockhopper for Windows



Rockhopper for Mac



Rockhopper for any platform



Rockhopper source code

- 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)
- 貼り付け(P)



- Bing で翻訳
- 電子メール (Windows Live Hotmail)
- すべてのアクセラレータ

from the command: jar

from the use the

c.bz2

# W2-2: Java確認など

①「java -version」実行結果（作業ディレクトリはどこでもよい）。このPCには、ver. 1.7.0\_79がインストールされていることがわかる。②~/Downloadsに移動。③ls実行結果で見えるものはヒトによって異なるが、基本気にしなくてもよい。

```
File Edit View Search Terminal Help
iu@bielinux[iu] pwd
/home/iu
① iu@bielinux[iu] java -version [ 2:14午後]
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)
② iu@bielinux[iu] cd ~/Downloads [ 2:14午後]
iu@bielinux[Downloads] pwd [ 2:14午後]
/home/iu/Downloads
③ iu@bielinux[Downloads] ls [ 2:14午後]
FaQCs FastQC fastqc_v0.11.4.zip IGV_2.3.67 IGV_2.3.67.zip
iu@bielinux[Downloads] █ [ 2:14午後]
```

# W2-3: wgetで取得

①wget実行。赤下線部のURL情報の最後がダウンロードしたいファイル名に相当する。②約13MB (14,039,789 bytes)であることがわかる。

```
iu@bielinux[Downloads] wget -c http://cs.wellesley.edu/~btjaden/Rockhopper/download/current/Rockhopper.jar
--2015-12-20 14:23:49-- http://cs.wellesley.edu/~btjaden/Rockhopper/download/current/Rockhopper.jar
Resolving cs.wellesley.edu (cs.wellesley.edu)... 149.130.136.40
Connecting to cs.wellesley.edu (cs.wellesley.edu)|149.130.136.40|:80... connected.
HTTP request sent, awaiting response... 200 OK
Length: 14039789 (13M) [application/x-java-archive]
Saving to: 'Rockhopper.jar'

100%[=====>] 14,039,789  1.72MB/s  in 9.6s

2015-12-20 14:23:59 (1.39 MB/s) - 'Rockhopper.jar' saved [14039789/14039789]

iu@bielinux[Downloads] ls
FastQC  fastqc_v0.11.4.zip  IGV_2.3.67.zip
FastQC  IGV_2.3.67         Rockhopper.jar
iu@bielinux[Downloads] ls -l Rockhopper.jar
-rw-rw-r-- 1 iu iu 14039789  3月 17  2015 Rockhopper.jar
iu@bielinux[Downloads]
```



# W2-4: GUI版を実行

①RockhopperのGUI版を実行したい場合のやり方が書いてあるので、②その通りに実行。③リターンキーを押すと…

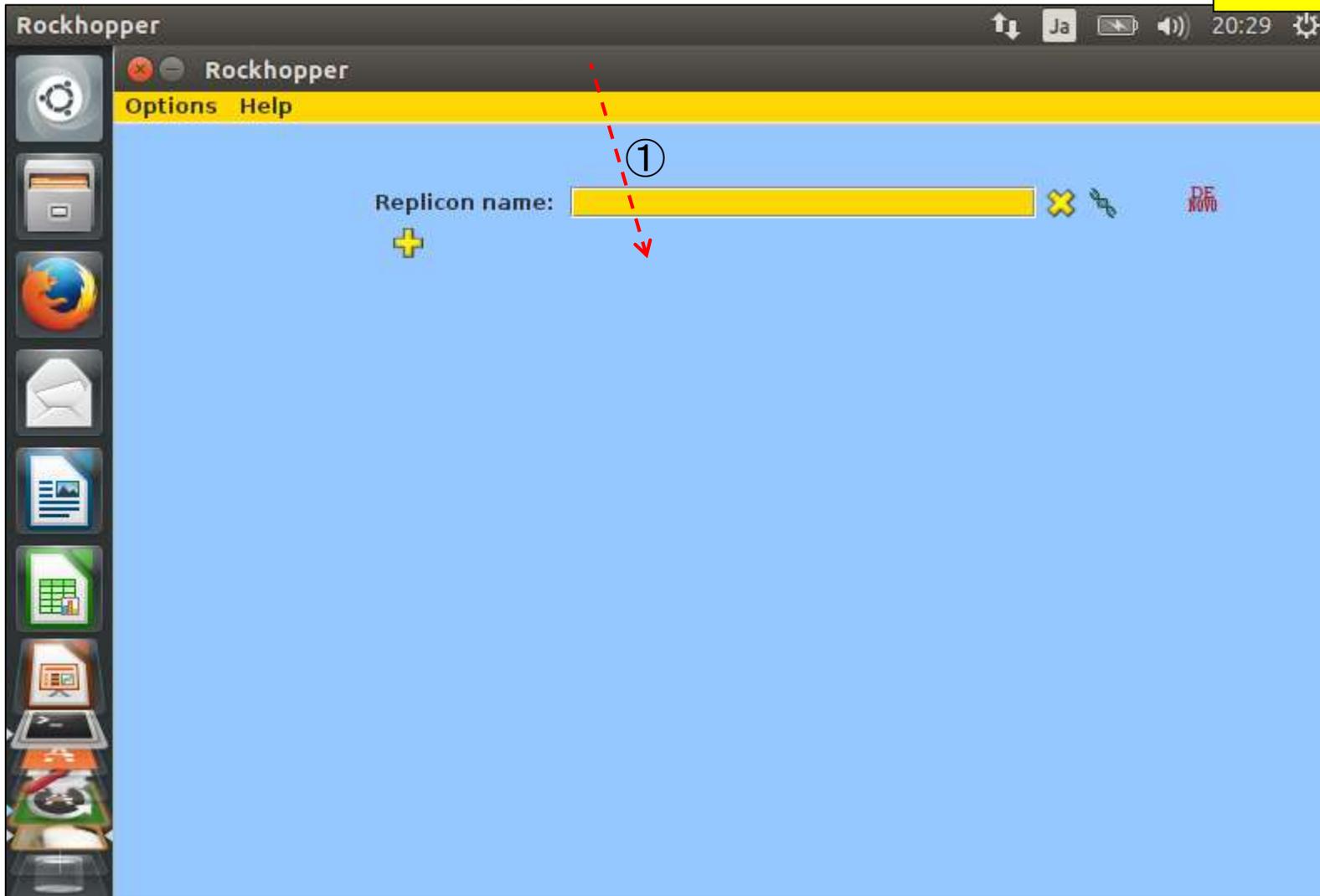
```
iu@bielinux[~/Downloads]
iu@bielinux[Downloads] ls -l Rockhopper.jar [ 8:24午後 ]
-rw-rw-r-- 1 iu iu 14039789 3月 17 05:39 Rockhopper.jar
iu@bielinux[Downloads] java -Xmx1200m -jar Rockhopper.jar
```

### Download Latest Release (Rockhopper version 2.0.3)

<p>Rockhopper for Windows</p> 	<p>Rockhopper for Mac</p>  <p>Error opening Rockhopper on Mac?</p>	<p>Rockhopper for any platform</p>  <p>To execute the GUI version of Rockhopper, use the following command: java -Xmx1200m -jar Rockhopper.jar</p> <p>To execute the command line version of Rockhopper, use the following command: java -Xmx1200m -cp Rockhopper.jar Rockhopper</p>	<p>Rockhopper source code</p>  <p>To extract the source code from the JAR file, use the following command: jar xf Rockhopper.jar</p>  <p>To extract the source code from the compressed TAR archive, use the following command: tar xjf Rockhopper-2.0.3.tar.bz2</p>
---	--	---	--

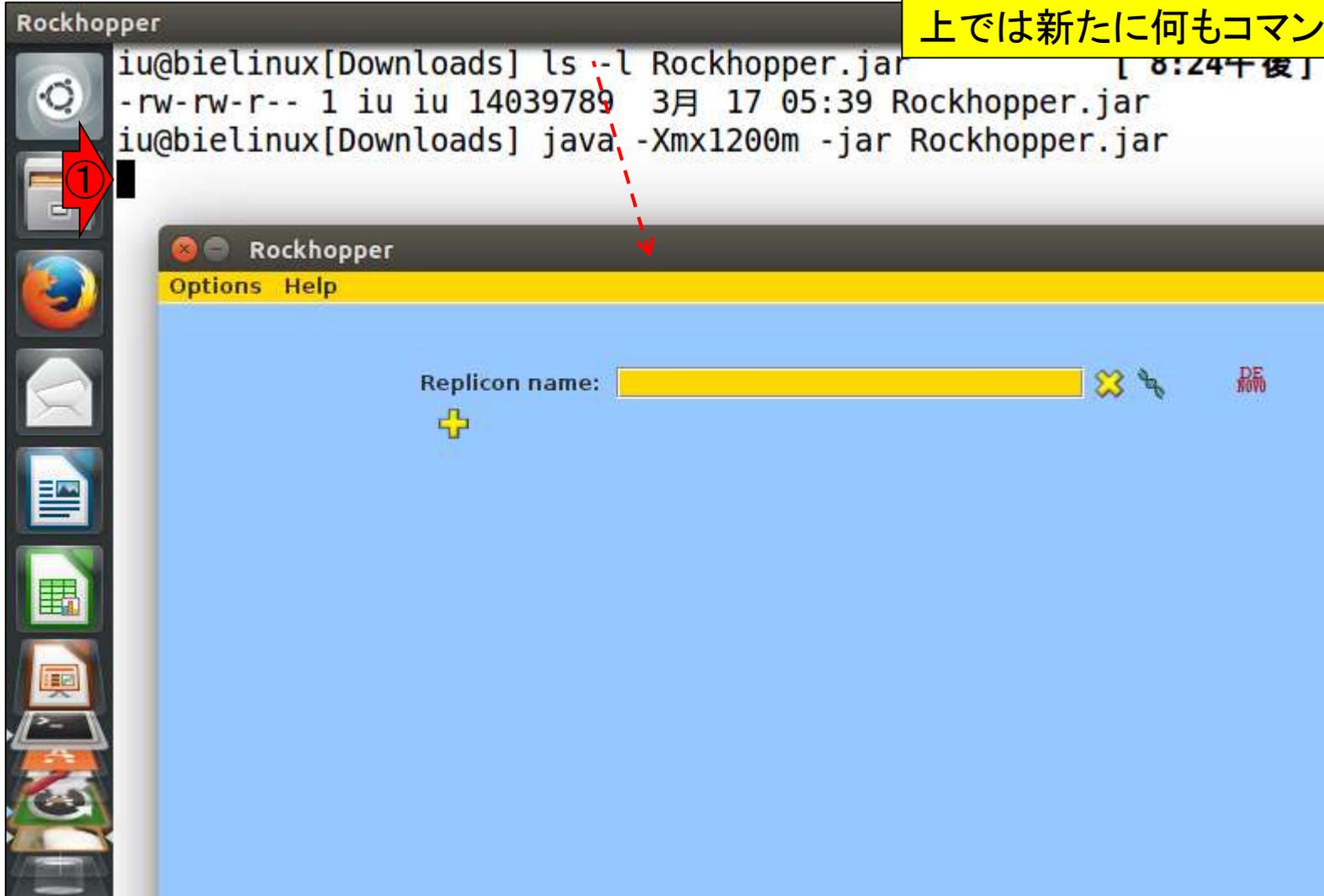
# W2-4: GUI版を実行

RockhopperのGUI版が起動する。①GUIを赤矢印の始点から終点に移動させると…。



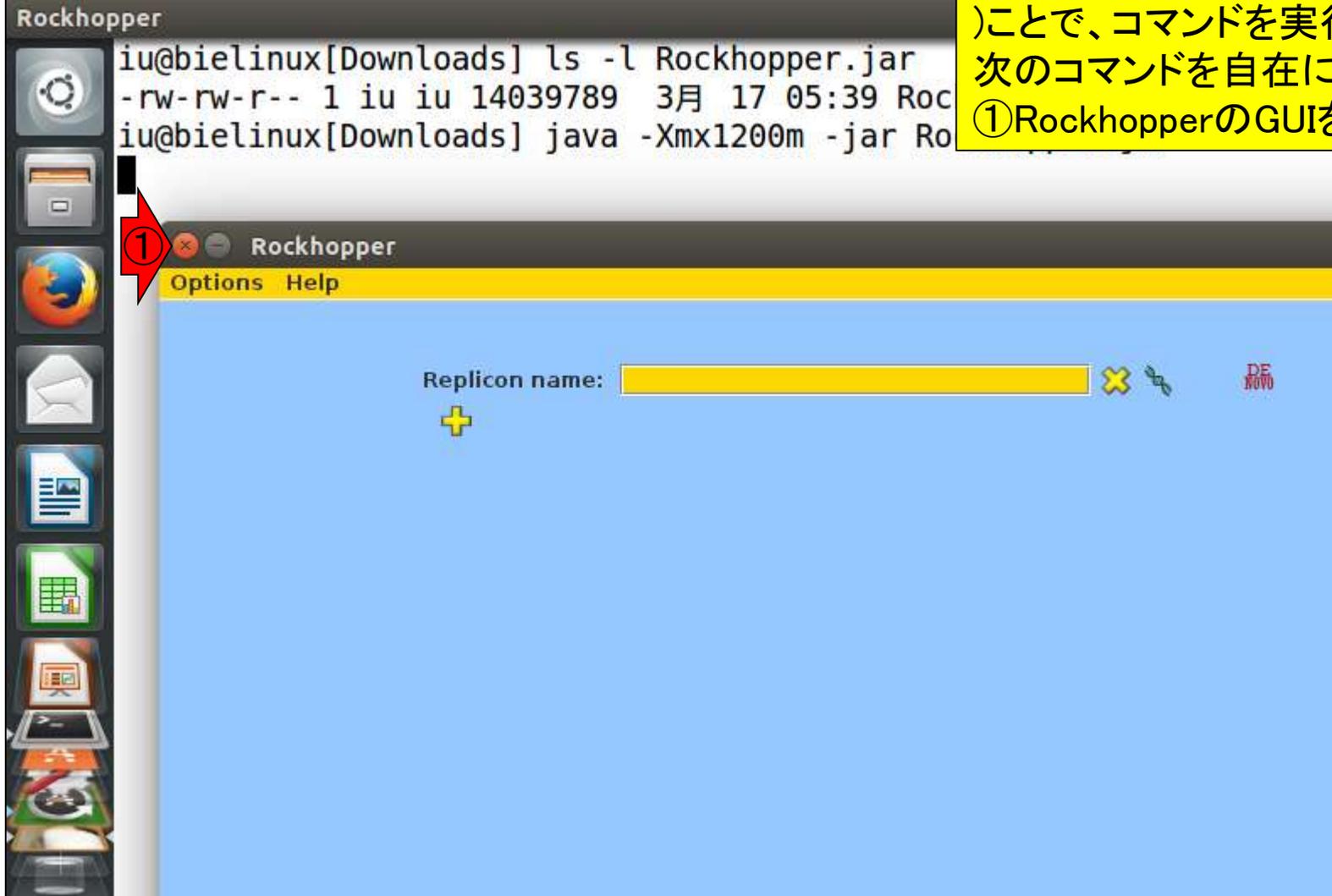
# W2-4: GUI版を実行

RockhopperのGUI版を起動したコマンドが見える。RockhopperのGUI起動中は、①のようにコマンドプロンプトが出ないため、このターミナル上では新たに何もコマンドを打つことができない



# W3-1 : background job

それを回避する一つのやり方が「バックグラウンドジョブ」。この場合は、RockhopperのGUIを裏(background)で実行させる(jobを流す)ことで、コマンドを実行したターミナル上で、次のコマンドを自在に打てるようにすること。  
①RockhopperのGUIを一旦終了させると…



# W3-1 : background job

①通常のコマンド打ち込み可能状態となる。バックグラウンドジョブとは、RockhopperのGUIを起動しつつも、このような状態にするテクニックです

```
iu@bielinux[~/Downloads]
iu@bielinux[Downloads] ls -l Rockhopper.jar [ 8:24午後 ]
-rw-rw-r-- 1 iu iu 14039789 3月 17 05:39 Rockhopper.jar
iu@bielinux[Downloads] java -Xmx1200m -jar Rockhopper.jar
iu@bielinux[Downloads] [ 8:57午後 ]
```



①

基本的には①通常のコマンドの最後に「&」をつけるだけ。

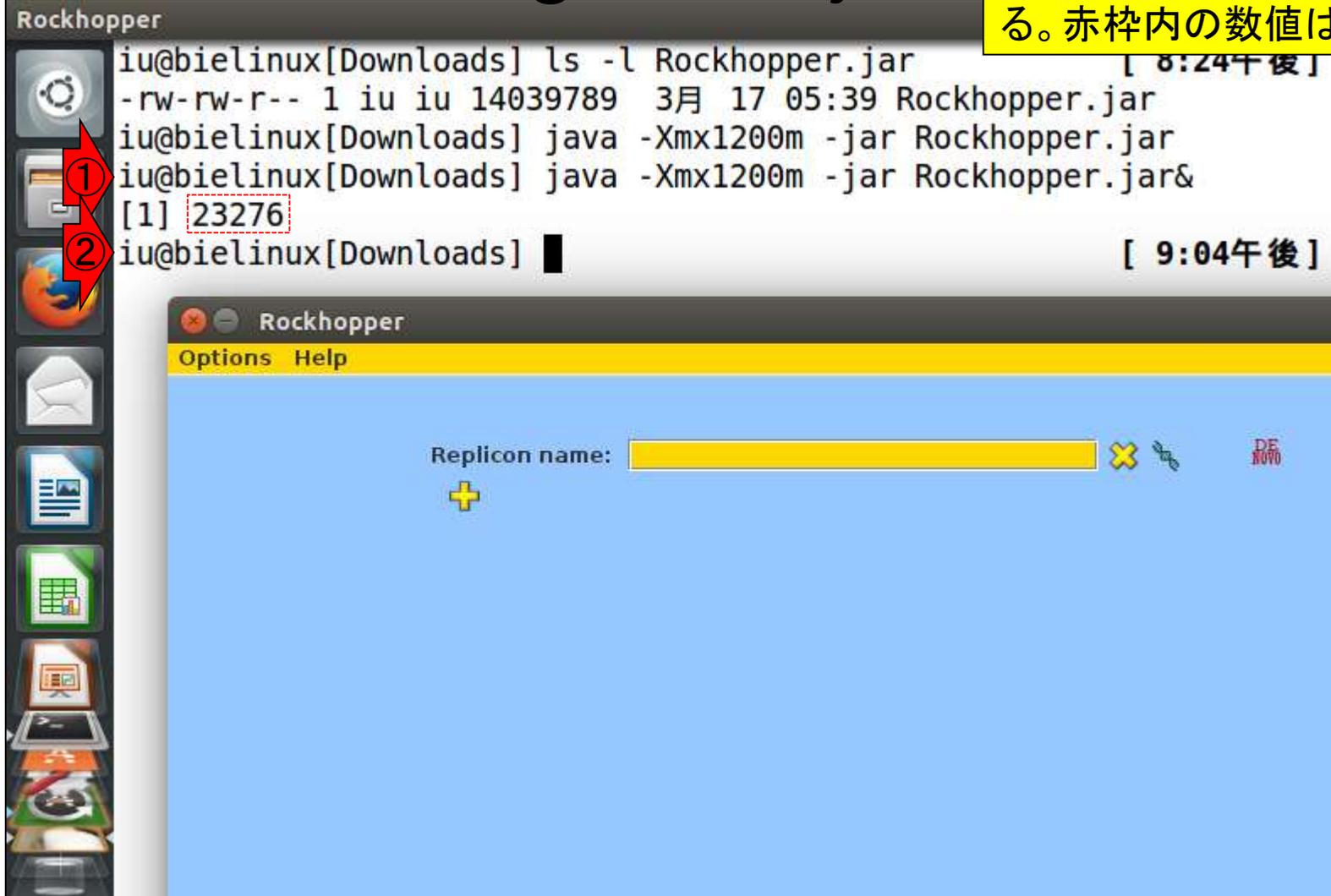
# W3-1 : background job

```
iu@bielinux[~/Downloads] ls -l Rockhopper.jar [ 8:24午後 ]
-rw-rw-r-- 1 iu iu 14039789 3月 17 05:39 Rockhopper.jar
iu@bielinux[Downloads] java -Xmx1200m -jar Rockhopper.jar
iu@bielinux[Downloads] java -Xmx1200m -jar Rockhopper.jar&
```



# W3-2: background job

①リターンキーを押した結果。確かにRockhopper GUIが起動しつつ、②コマンド打ち込み可能状態になっていることがわかる。赤枠内の数値は、ヒトによって異なる。

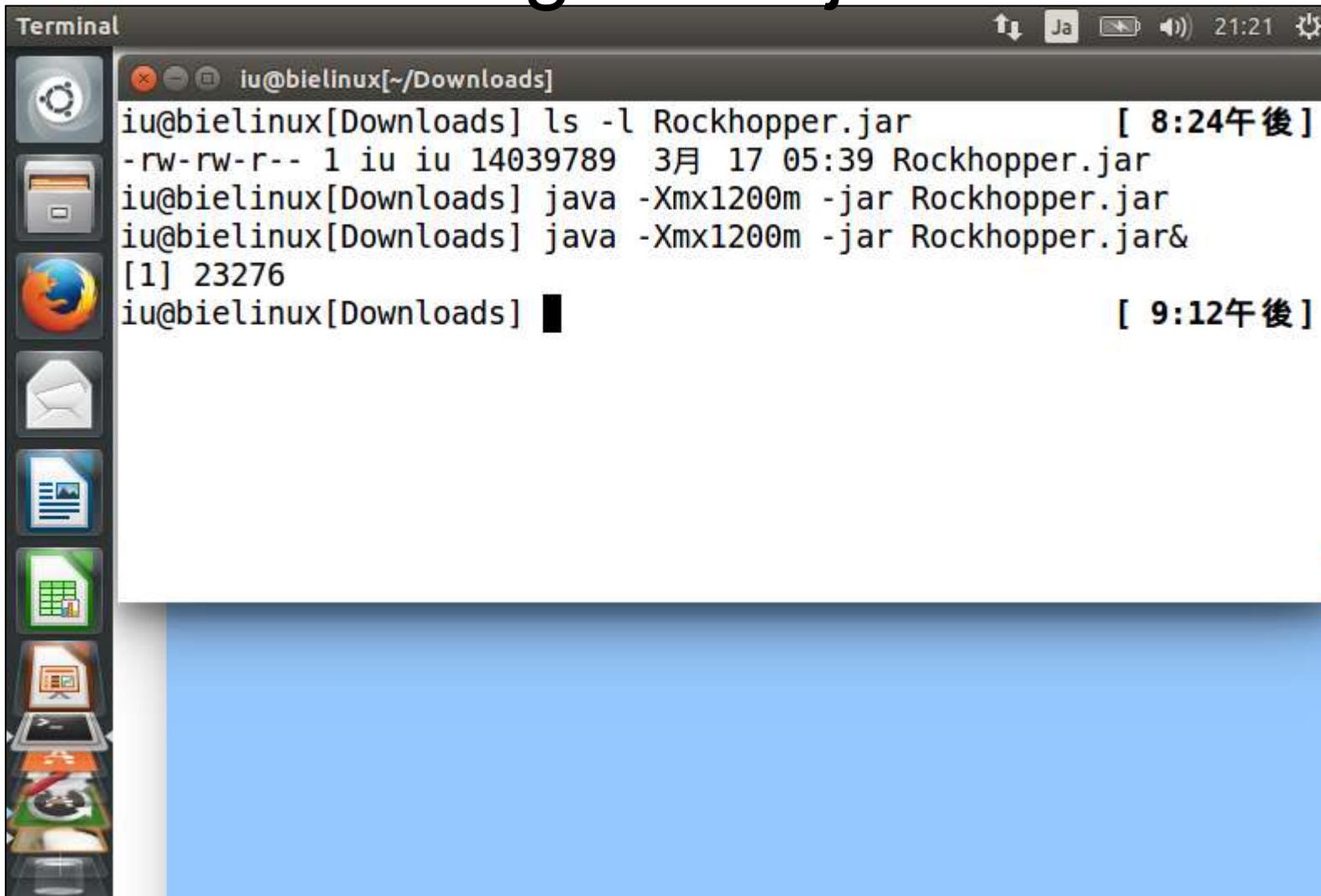


The image shows a terminal window and a Rockhopper GUI window. The terminal window displays the following commands and output:

```
iu@bielinux[Downloads] ls -l Rockhopper.jar [ 8:24午後 ]
-rw-rw-r-- 1 iu iu 14039789 3月 17 05:39 Rockhopper.jar
iu@bielinux[Downloads] java -Xmx1200m -jar Rockhopper.jar
iu@bielinux[Downloads] java -Xmx1200m -jar Rockhopper.jar&
[1] 23276
iu@bielinux[Downloads] [ 9:04午後 ]
```

The Rockhopper GUI window is titled "Rockhopper" and has a yellow header bar with "Options" and "Help" menus. The main area is light blue and contains a text input field labeled "Replicon name:" with a yellow border. Below the input field is a yellow plus sign icon. To the right of the input field are a yellow 'X' icon, a magnifying glass icon, and a red "DE" icon.

# W3-2: background job



The image shows a terminal window titled "Terminal" with a dark theme. The window title bar includes system icons for volume, network, and language (Ja), along with the time 21:21. The terminal content shows the following sequence of commands and output:

```
iu@bielinux[~/Downloads]
iu@bielinux[Downloads] ls -l Rockhopper.jar [ 8:24午後 ]
-rw-rw-r-- 1 iu iu 14039789 3月 17 05:39 Rockhopper.jar
iu@bielinux[Downloads] java -Xmx1200m -jar Rockhopper.jar
iu@bielinux[Downloads] java -Xmx1200m -jar Rockhopper.jar&
[1] 23276
iu@bielinux[Downloads] █ [ 9:12午後 ]
```

The terminal window is part of a desktop environment, with a vertical dock on the left side containing icons for various applications like a file manager, Firefox, and a mail client. The background of the desktop is a solid light blue color.

# W3-3: psコマンド

①psコマンドで実行中のプロセスを表示。プロセスと表現する機会が多いのでそう書いているが、jobやタスクという理解でもよい。Windowsのヒトは、「タスクマネージャー」を開いて眺めているようなものだと思えばよい。

```
Terminal
iu@bielinux[~/Downloads]
iu@bielinux[Downloads] ls -l Rockhopper.jar [ 8:24午後 ]
-rw-rw-r-- 1 iu iu 14039789 3月 17 05:39 Rockhopper.jar
iu@bielinux[Downloads] java -Xmx1200m -jar Rockhopper.jar
iu@bielinux[Downloads] java -Xmx1200m -jar Rockhopper.jar&
[1] 23276
iu@bielinux[Downloads] ps -f [ 9:12午後 ]
UID          PID    PPID  C  STIME TTY          TIME CMD
iu           9274   9254  0  06:53 pts/4        00:00:00 zsh
iu          23276   9274  0  14:09 pts/4        00:00:13 java -Xmx1200m -ja
iu          30731   9274  0  14:38 pts/4        00:00:00 ps -f
iu@bielinux[Downloads] [ 2:38午後 ]
```



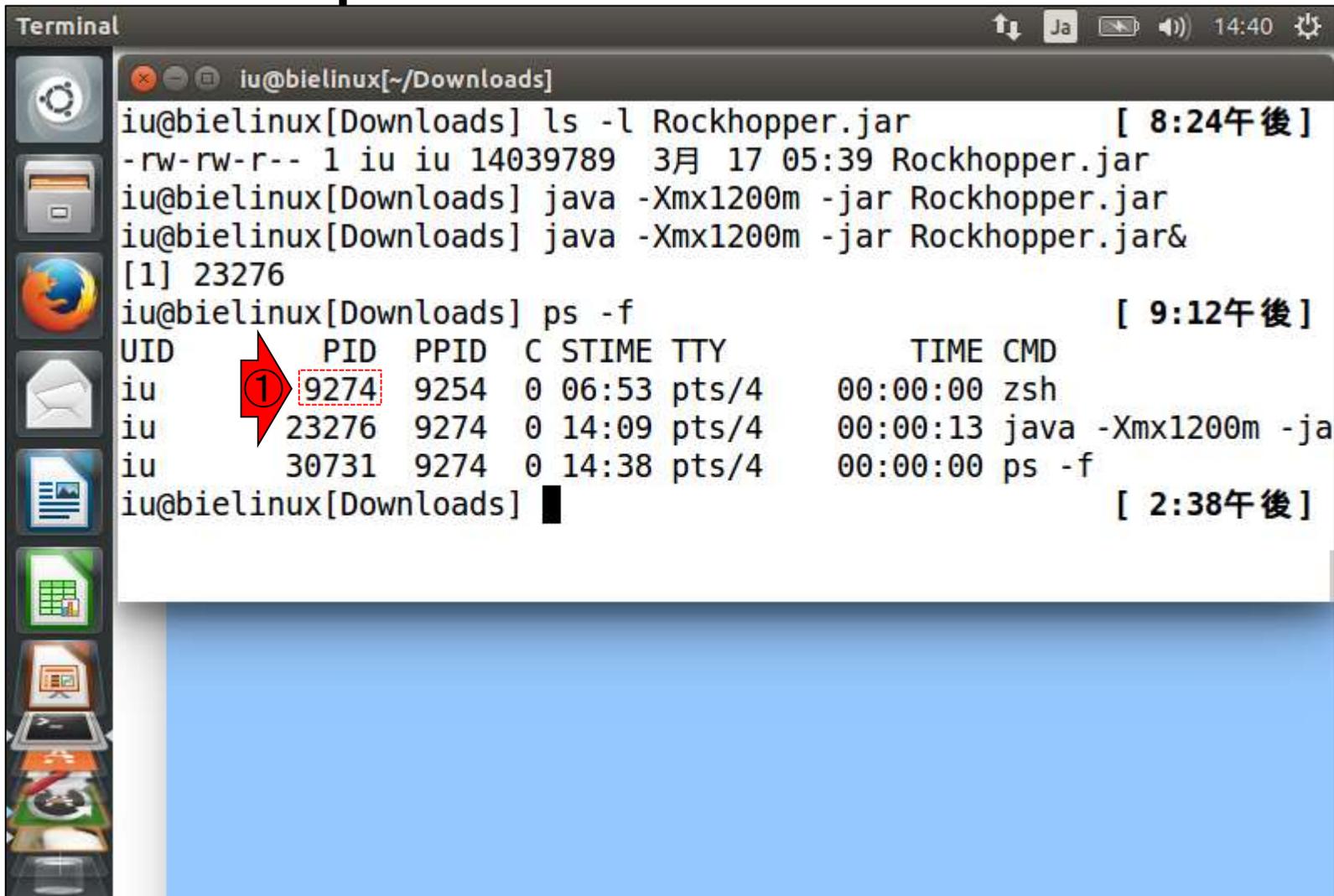
# W3-3: psコマンド

①CMD列が、現在実行中のコマンド。  
②zshのみ打ち込んだ記憶がないだろうが、これは③のターミナルボタンを押して起動中のターミナルそのもの

```
Terminal
iu@bielinux[~/Downloads]
iu@bielinux[Downloads] ls -l Rockhopper.jar [ 8:24午後 ]
-rw-rw-r-- 1 iu iu 14039789 3月 17 05:39 Rockhopper.jar
iu@bielinux[Downloads] java -Xmx1200m -jar Rockhopper.jar
iu@bielinux[Downloads] java -Xmx1200m -jar Rockhopper.jar&
[1] 23276
iu@bielinux[Downloads] ps -f [ 9:12午後 ]
UID      PID  PPID  C  STIME TTY          TIME CMD
iu       9274  9254  0  06:53 pts/4        00:00:00 zsh
iu      23276  9274  0  14:09 pts/4        00:00:13 java -Xmx1200m -ja
iu      30731  9274  0  14:38 pts/4        00:00:00 ps -f
iu@bielinux[Downloads] [ 2:38午後 ]
```

# W3-4: psコマンド

①ターミナルボタンを押して起動したターミナルのプロセスID (PID)は9274。



```
Terminal
iu@bielinux[~/Downloads]
iu@bielinux[Downloads] ls -l Rockhopper.jar [ 8:24午後 ]
-rw-rw-r-- 1 iu iu 14039789 3月 17 05:39 Rockhopper.jar
iu@bielinux[Downloads] java -Xmx1200m -jar Rockhopper.jar
iu@bielinux[Downloads] java -Xmx1200m -jar Rockhopper.jar&
[1] 23276
iu@bielinux[Downloads] ps -f [ 9:12午後 ]
UID          PID  PPID  C  STIME TTY          TIME CMD
iu           9274   9254  0  06:53 pts/4        00:00:00 zsh
iu          23276   9274  0  14:09 pts/4        00:00:13 java -Xmx1200m -ja
iu          30731   9274  0  14:38 pts/4        00:00:00 ps -f
iu@bielinux[Downloads] [ 2:38午後 ]
```

# W3-4: psコマンド

①ターミナルボタンを押して起動したターミナルのプロセスID (PID)は9274。②で打ち込んだコマンドのPIDは23276。

```
Terminal
iu@bielinux[~/Downloads]
iu@bielinux[Downloads] ls -l Rockhopper.jar [ 8:24午後 ]
-rw-rw-r-- 1 iu iu 14039789 3月 17 05:39 Rockhopper.jar
iu@bielinux[Downloads] java -Xmx1200m -jar Rockhopper.jar
iu@bielinux[Downloads] java -Xmx1200m -jar Rockhopper.jar& ②
[1] 23276
iu@bielinux[Downloads] ps -f [ 9:12午後 ]
UID          PID  PPID  C  STIME TTY          TIME CMD
iu           9274   9254  0  06:53 pts/4        00:00:00 zsh
iu          23276   9274  0  14:09 pts/4        00:00:13 java -Xmx1200m -ja
iu          30731   9274  0  14:38 pts/4        00:00:00 ps -f
iu@bielinux[Downloads] [ 2:38午後 ]
```



# W3-4: psコマンド

①ターミナルボタンを押して起動したターミナルのプロセスID (PID)は9274。②で打ち込んだコマンドのPIDは23276。この情報は③のところに相当。

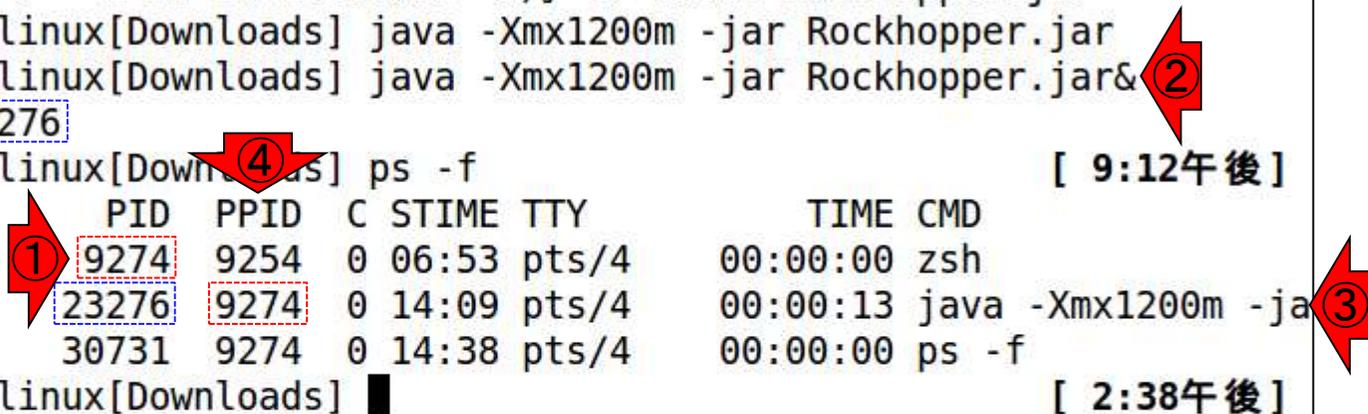
```
Terminal
iu@bielinux[~/Downloads]
iu@bielinux[Downloads] ls -l Rockhopper.jar [ 8:24午後 ]
-rw-rw-r-- 1 iu iu 14039789 3月 17 05:39 Rockhopper.jar
iu@bielinux[Downloads] java -Xmx1200m -jar Rockhopper.jar
iu@bielinux[Downloads] java -Xmx1200m -jar Rockhopper.jar&
[1] 23276
iu@bielinux[Downloads] ps -f [ 9:12午後 ]
UID      PID  PPID  C  STIME TTY          TIME CMD
iu      9274   9254  0  06:53 pts/4        00:00:00 zsh
iu     23276   9274  0  14:09 pts/4        00:00:13 java -Xmx1200m -ja
iu     30731   9274  0  14:38 pts/4        00:00:00 ps -f
iu@bielinux[Downloads] [ 2:38午後 ]
```



# W3-4: psコマンド

①ターミナルボタンを押して起動したターミナルのプロセスID (PID)は9274。②で打ち込んだコマンドのPIDは23276。この情報は③のところに相当。全体像から④のPPIDがPIDの親プロセスIDであることがわかる。

```
Terminal
iu@bielinux[~/Downloads]
iu@bielinux[Downloads] ls -l Rockhopper.jar
-rw-rw-r-- 1 iu iu 14039789  3月 17 05:39 Rockhopper.jar
iu@bielinux[Downloads] java -Xmx1200m -jar Rockhopper.jar
iu@bielinux[Downloads] java -Xmx1200m -jar Rockhopper.jar&
[1] 23276
iu@bielinux[Downloads] ps -f
UID          PID  PPID  C  STIME TTY          TIME CMD
iu           9274   9254  0  06:53 pts/4        00:00:00 zsh
iu          23276   9274  0  14:09 pts/4        00:00:13 java -Xmx1200m -jar Rockhopper.jar&
iu          30731   9274  0  14:38 pts/4        00:00:00 ps -f
iu@bielinux[Downloads]
```



# W3-5: psコマンド

①「ps -f」のプロセスID (PID)は30731。  
この親プロセスID (PPID)が9274なのは  
妥当。理由は、このPID9274のターミナ  
ル上で実行したコマンドだから。

```
Terminal
iu@bielinux[~/Downloads]
iu@bielinux[Downloads] ls -l Rockhopper.jar [ 8:24午後 ]
-rw-rw-r-- 1 iu iu 14039789 3月 17 05:39 Rockhopper.jar
iu@bielinux[Downloads] java -Xmx1200m -jar Rockhopper.jar
iu@bielinux[Downloads] java -Xmx1200m -jar Rockhopper.jar&
[1] 23276
iu@bielinux[Downloads] ps -f [ 9:12午後 ]
UID          PID  PPID  C  STIME TTY          TIME CMD
iu           9274   9254  0  06:53 pts/4        00:00:00 zsh
iu          23276   9274  0  14:09 pts/4        00:00:13 java -Xmx1200m -ja
iu          30731   9274  0  14:38 pts/4        00:00:00 ps -f
iu@bielinux[Downloads] [ 2:38午後 ]
```

# W3-5: psコマンド

①もう一度「ps -f」を実行。このプロセスID (PID)は1596。このように数値はコロコロ変わるものなので、基本的にPIDとPPIDの関係がわかっていればよい。

```
Terminal
iu@bielinux[~/Downloads]
iu@bielinux[Downloads] java -Xmx1200m -jar Rockhopper.jar&
[1] 23276
iu@bielinux[Downloads] ps -f [ 9:12午後 ]
UID          PID    PPID  C STIME TTY          TIME CMD
iu           9274    9254  0 06:53 pts/4        00:00:00 zsh
iu          23276    9274  0 14:09 pts/4        00:00:13 java -Xmx1200m -ja
iu          30731    9274  0 14:38 pts/4        00:00:00 ps -f
① iu@bielinux[Downloads] ps -f [ 2:38午後 ]
UID          PID    PPID  C STIME TTY          TIME CMD
iu           1596    9274  0 16:58 pts/4        00:00:00 ps -f
iu           9274    9254  0 06:53 pts/4        00:00:00 zsh
iu          23276    9274  0 14:09 pts/4        00:00:50 java -Xmx1200m -ja
iu@bielinux[Downloads] [ 4:58午後 ]
```

# W3-6: プロセスの終了

①GUIベースでやる場合は、×ボタンだが、ここでは押さないで。

The screenshot shows a terminal window titled "Rockhopper" with the following content:

```
iu@bielinux[~/Downloads] java -Xmx1200m -jar Rockhopper.jar&
[1] 23276
iu@bielinux[Downloads] ps -f                                     [ 9:12午後 ]
UID          PID  PPID  C  STIME TTY          TIME CMD
iu          9274   9254  0  06:53 pts/4        00:00:00 zsh
```

A red arrow with the number "1" points to the first line of the terminal output. Below the terminal, a GUI window titled "Rockhopper" is open, showing a yellow header bar with "Options" and "Help" buttons. The main area is light blue and contains a text input field labeled "Replicon name:" with a yellow cursor. To the right of the input field are a yellow "X" button, a blue icon, and a red "DE" logo. A yellow plus sign is visible below the input field.

# W3-6: killコマンド

①「kill プロセスID」で終了させることができる。やたらとメモリを消費している意味不明なプロセスが実行されている場合に、このような処理を行って終了させる。

```
Terminal
iu@bielinux[~/Downloads]
iu@bielinux[Downloads] java -Xmx1200m -jar Rockhopper.jar&
[1] 23276
iu@bielinux[Downloads] ps -f                                     [ 9:12午後 ]
UID          PID    PPID  C  STIME TTY          TIME CMD
iu           9274    9254  0   06:53 pts/4        00:00:00 zsh
iu          23276    9274  0  14:09 pts/4        00:00:13 java -Xmx1200m -ja
iu          30731    9274  0  14:38 pts/4        00:00:00 ps -f
iu@bielinux[Downloads] ps -f                                     [ 2:38午後 ]
UID          PID    PPID  C  STIME TTY          TIME CMD
iu           1596    9274  0  16:58 pts/4        00:00:00 ps -f
iu           9274    9254  0   06:53 pts/4        00:00:00 zsh
iu          23276    9274  0  14:09 pts/4        00:00:50 java -Xmx1200m -ja
iu@bielinux[Downloads] kill 23276                             [ 4:58午後 ]
```



# W3-6: killコマンド

①「kill プロセスID」でリターンキーを押した後の状態。RockhopperのGUIが終了していることがわかる。

```
Terminal
iu@bielinux[~/Downloads]
UID      PID  PPID  C  STIME TTY      TIME  CMD
iu       9274 9254  0  06:53 pts/4    00:00:00 zsh
iu      23276 9274  0  14:09 pts/4    00:00:13 java -Xmx1200m -ja
iu      30731 9274  0  14:38 pts/4    00:00:00 ps -f
iu@bielinux[Downloads] ps -f
UID      PID  PPID  C  STIME TTY      TIME  CMD
iu       1596 9274  0  16:58 pts/4    00:00:00 ps -f
iu       9274 9254  0  06:53 pts/4    00:00:00 zsh
iu      23276 9274  0  14:09 pts/4    00:00:50 java -Xmx1200m -ja
iu@bielinux[Downloads] kill 23276
iu@bielinux[Downloads]
[1] + exit 143  java -Xmx1200m -jar Rockhopper.jar
iu@bielinux[Downloads]
```

②「ps -f」で確認。確かにPID23276は存在しない。

# W3-6: killコマンド

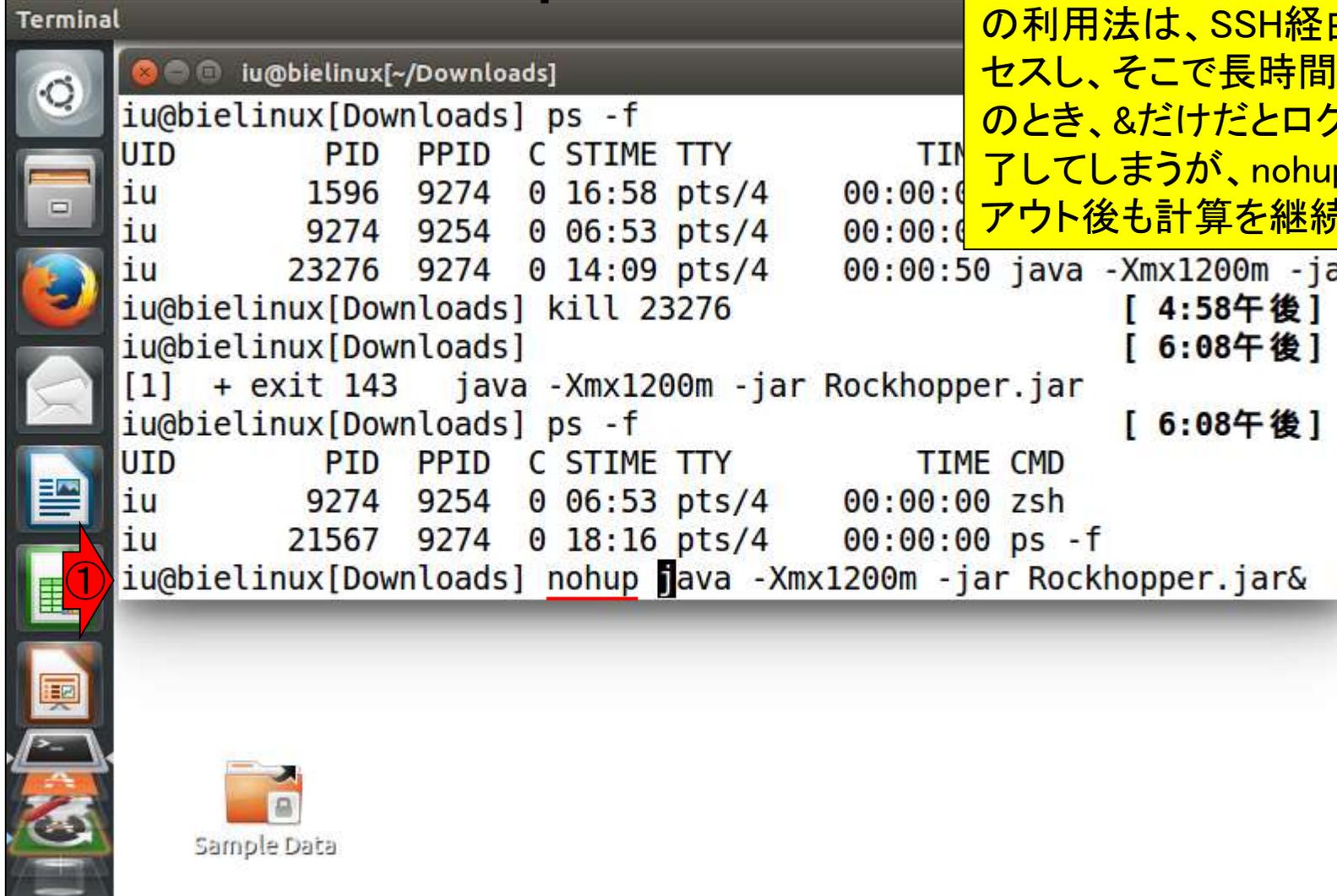
```
Terminal
iu@bielinux[~/Downloads]
iu@bielinux[Downloads] ps -f [ 2:38午後 ]
UID      PID  PPID  C  STIME TTY          TIME CMD
iu       1596  9274  0  16:58 pts/4      00:00:00 ps -f
iu       9274  9254  0  06:53 pts/4      00:00:00 zsh
iu      23276  9274  0  14:09 pts/4      00:00:50 java -Xmx1200m -ja
① iu@bielinux[Downloads] kill 23276 [ 4:58午後 ]
iu@bielinux[Downloads] [ 6:08午後 ]
[1] + exit 143  java -Xmx1200m -jar Rockhopper.jar
② iu@bielinux[Downloads] ps -f [ 6:08午後 ]
UID      PID  PPID  C  STIME TTY          TIME CMD
iu       9274  9254  0  06:53 pts/4      00:00:00 zsh
iu      21567  9274  0  18:16 pts/4      00:00:00 ps -f
iu@bielinux[Downloads] [ 6:16午後 ]
```

Sample Data

# W3-7: nohupコマンド

バックグラウンドジョブ時には、通常コマンドの最後に&をつけるだけでなく、①コマンドの最初にnohupをつける。一般的なNGS解析の利用法は、SSH経由で大型計算機にアクセスし、そこで長時間の計算を実行する。このとき、&だけだとログアウト時に計算が終了してしまうが、nohupをつけることで、ログアウト後も計算を継続させることができる。

```
Terminal
iu@bielinux[~/Downloads]
iu@bielinux[Downloads] ps -f
UID          PID    PPID  C  STIME TTY          TIME CMD
iu           1596   9274  0  16:58 pts/4        00:00:00
iu           9274   9254  0   06:53 pts/4        00:00:00
iu          23276   9274  0  14:09 pts/4        00:00:50 java -Xmx1200m -ja
iu@bielinux[Downloads] kill 23276
iu@bielinux[Downloads]
[1] + exit 143  java -Xmx1200m -jar Rockhopper.jar
iu@bielinux[Downloads] ps -f
UID          PID    PPID  C  STIME TTY          TIME CMD
iu           9274   9254  0   06:53 pts/4        00:00:00 zsh
iu          21567   9274  0  18:16 pts/4        00:00:00 ps -f
iu@bielinux[Downloads] nohup java -Xmx1200m -jar Rockhopper.jar&
```



# W3-7: nohupコマンド

①でリターンキーを1回押した直後の状態。ターミナル上では、一見コマンド打ち込み不可能なように見えるが、ターミナル画面をアクティブにしてもう一度リターンキーを押すと、ちゃんとバックグラウンドジョブとしてRockhopper GUIが起動していることが確認できる。

The screenshot displays a terminal window with the following commands and output:

```
iu 23276 9274 0 14:09 pts/4 00:00
iu@bielinux[Downloads] kill 23276
iu@bielinux[Downloads]
[1] + exit 143 java -Xmx1200m -jar Rockhopper.jar
iu@bielinux[Downloads] ps -f
UID      PID  PPID  C  STIME TTY          TIME CMD
iu       9274  9254  0  06:53 pts/4        00:00:00 zsh
iu      21567  9274  0  18:16 pts/4        00:00:00 ps -f
iu@bielinux[Downloads] nohup java -Xmx1200m -jar Rockhopper.jar&
[1] 8785
iu@bielinux[Downloads] nohup: ignoring input and appending output
to 'nohup.out'
```

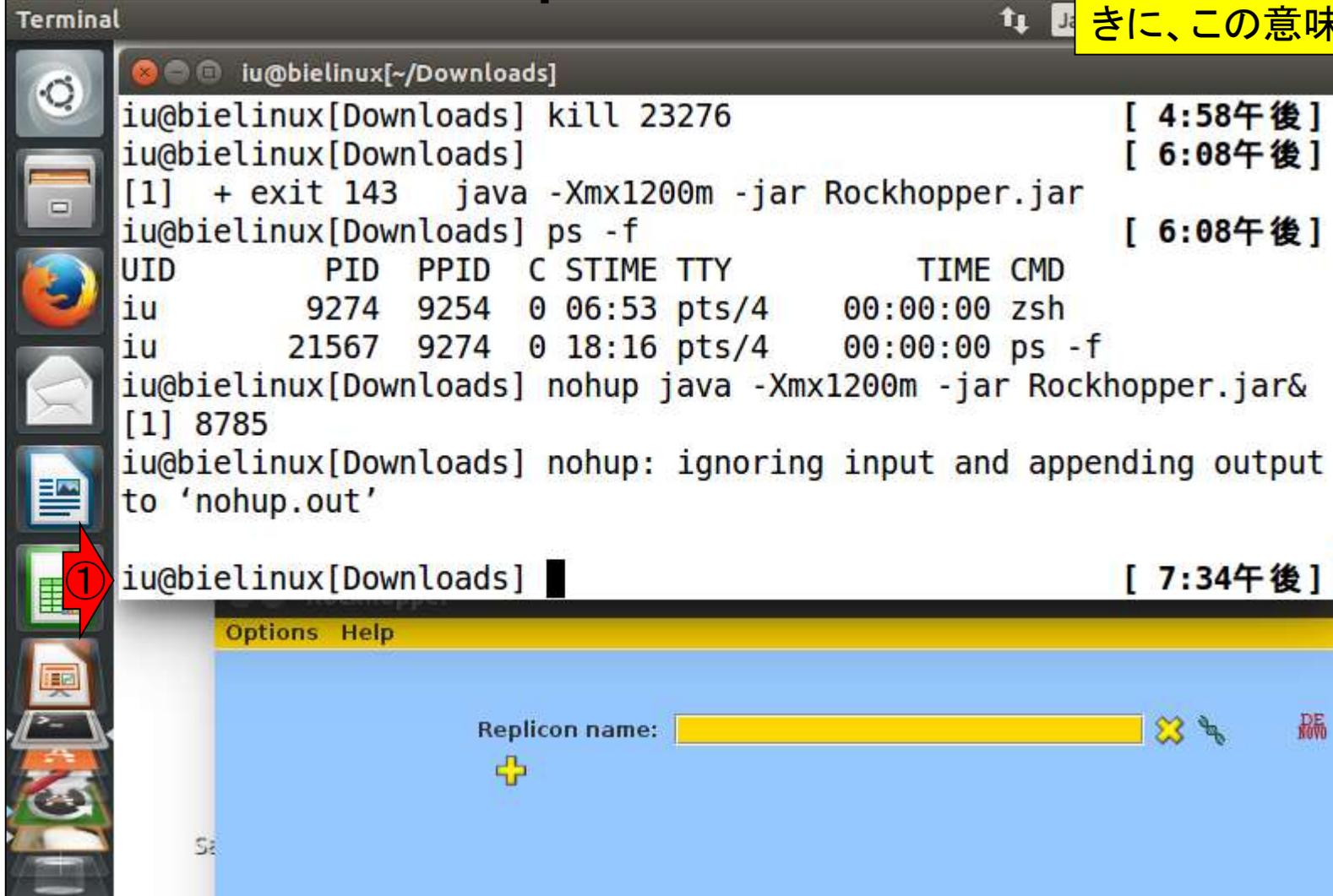
The Rockhopper GUI window is titled "Rockhopper" and has a menu bar with "Options" and "Help". It features a text input field labeled "Replicon name:" with a yellow background, a close button (X), a help icon, and a red "DE" logo. A yellow plus sign is visible below the input field.

Red arrows labeled ① and ② point to the terminal command and the GUI window, respectively.

# W3-7: nohupコマンド

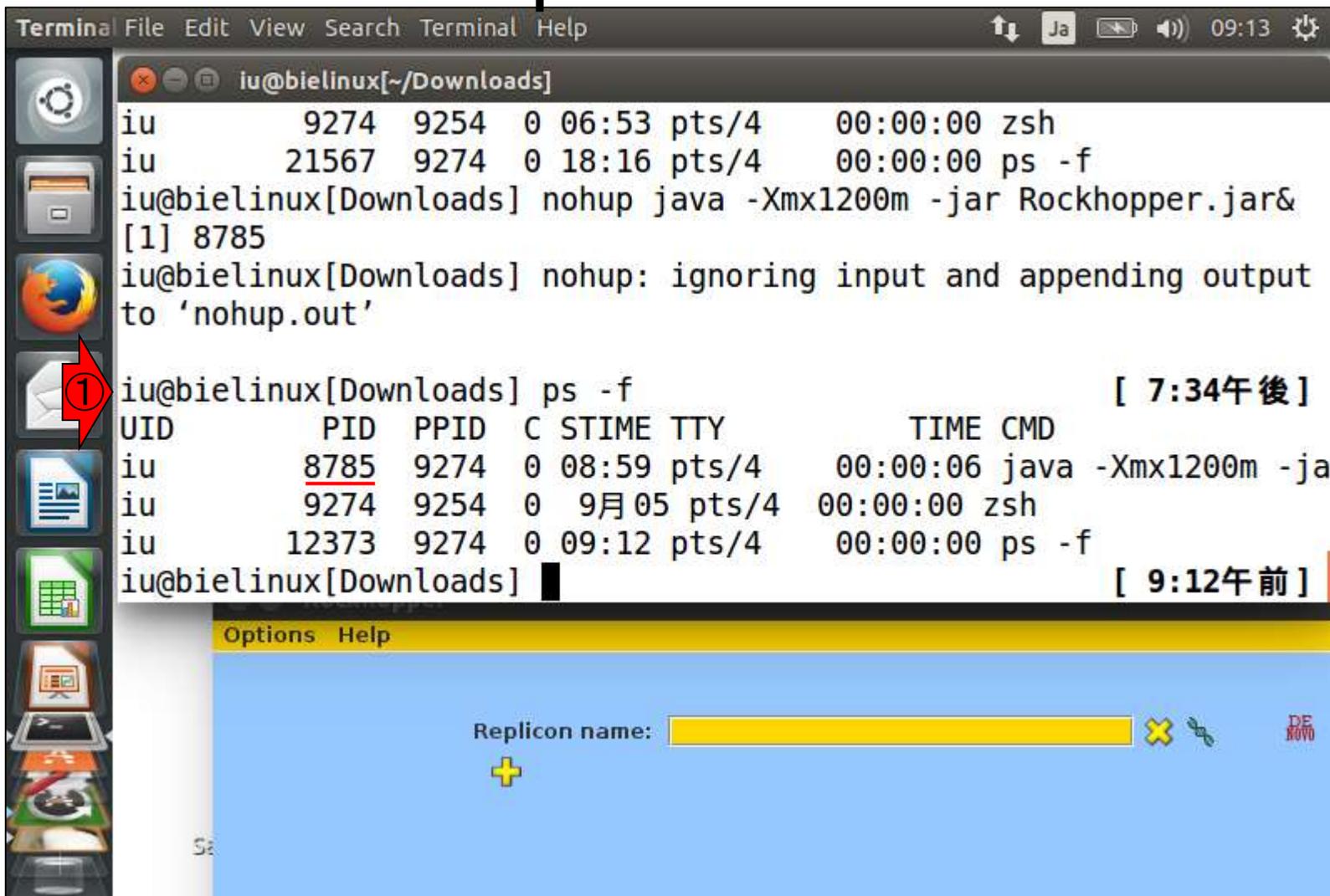
①SSHの話までするとややこしいのでこれ以上深入りしないが、遺伝研スパコンなどを利用するようになったときに、この意味が理解できるでしょう。

```
Terminal
iu@bielinux[~/Downloads]
iu@bielinux[Downloads] kill 23276 [ 4:58午後]
iu@bielinux[Downloads] [ 6:08午後]
[1] + exit 143 java -Xmx1200m -jar Rockhopper.jar
iu@bielinux[Downloads] ps -f [ 6:08午後]
UID      PID  PPID  C  STIME TTY          TIME CMD
iu       9274  9254  0  06:53 pts/4      00:00:00 zsh
iu      21567  9274  0  18:16 pts/4      00:00:00 ps -f
iu@bielinux[Downloads] nohup java -Xmx1200m -jar Rockhopper.jar&
[1] 8785
iu@bielinux[Downloads] nohup: ignoring input and appending output
to 'nohup.out'
iu@bielinux[Downloads] [ 7:34午後]
```



①「ps -f」で確認。確かにPID8785  
が存在することがわかる。

# W3-7: nohupコマンド



The screenshot shows a terminal window with the following commands and output:

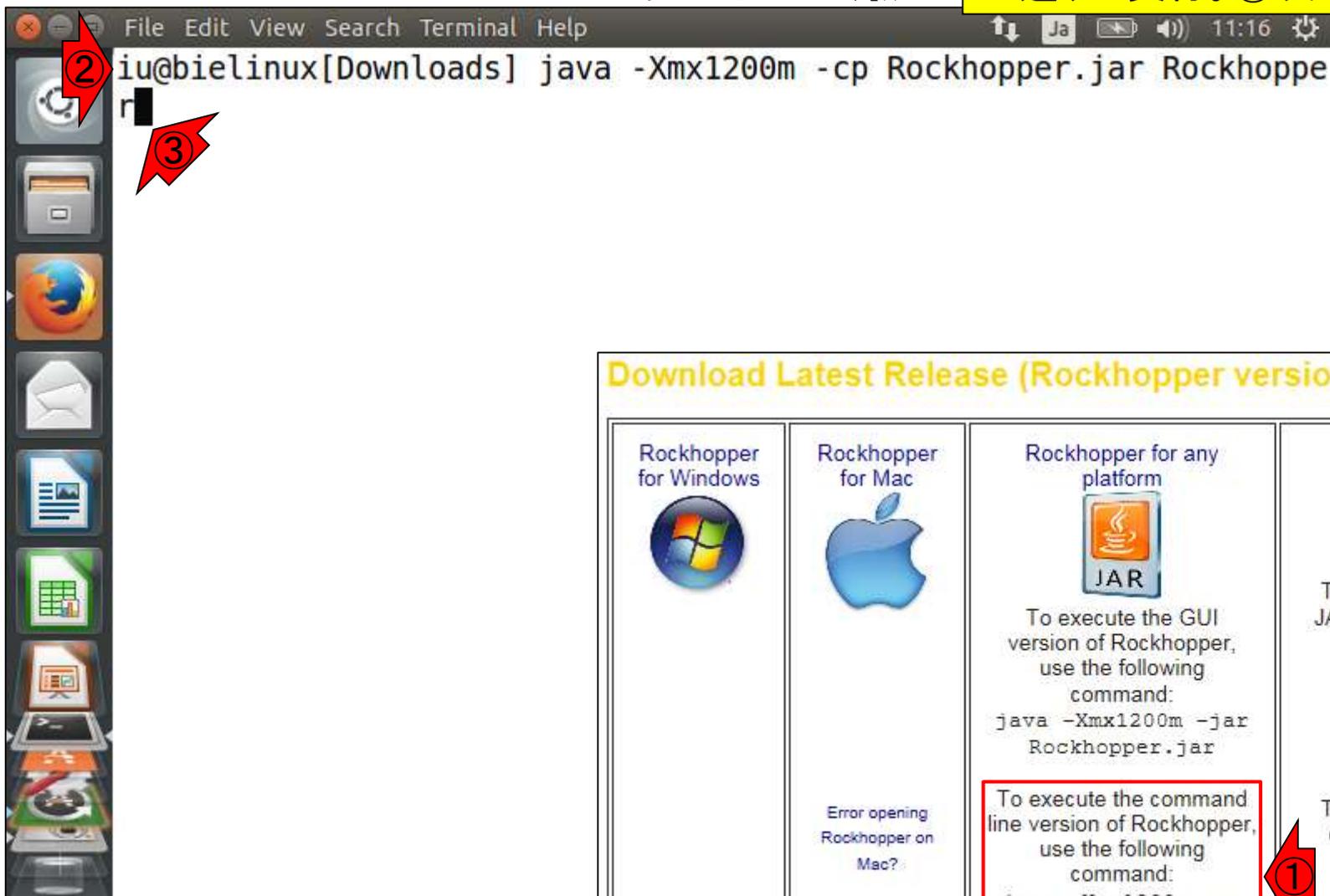
```
iu@bielinux[~/Downloads]
iu      9274  9254  0 06:53 pts/4    00:00:00 zsh
iu     21567  9274  0 18:16 pts/4    00:00:00 ps -f
iu@bielinux[Downloads] nohup java -Xmx1200m -jar Rockhopper.jar&
[1] 8785
iu@bielinux[Downloads] nohup: ignoring input and appending output
to 'nohup.out'

iu@bielinux[Downloads] ps -f                                [ 7:34午後 ]
UID      PID  PPID  C  STIME TTY          TIME CMD
iu       8785  9274  0  08:59 pts/4    00:00:06 java -Xmx1200m -ja
iu       9274  9254  0   9月05 pts/4    00:00:00 zsh
iu      12373  9274  0  09:12 pts/4    00:00:00 ps -f
iu@bielinux[Downloads]                                     [ 9:12午前 ]
```

A red arrow with the number 1 points to the `ps -f` command. Below the terminal output, there is a yellow bar with the text "Options Help" and a blue area with a "Replicon name:" label and a yellow input field.

# W4-1: コマンドライン版

①Rockhopperのコマンドライン版を実行したい場合のやり方が書いてあるので、②その通りに実行。③リターンキーを押すと…



## Download Latest Release (Rockhopper version 2.0.3)

<p>Rockhopper for Windows</p> 	<p>Rockhopper for Mac</p>  <p>Error opening Rockhopper on Mac?</p>	<p>Rockhopper for any platform</p>  <p>To execute the GUI version of Rockhopper, use the following command:</p> <pre>java -Xmx1200m -jar Rockhopper.jar</pre> <p>To execute the command line version of Rockhopper, use the following command:</p> <pre>java -Xmx1200m -cp Rockhopper.jar Rockhopper</pre>	<p>Rockhopper source code</p>  <p>To extract the source code from the JAR file, use the following command:</p> <pre>jar xf Rockhopper.jar</pre>  <p>To extract the source code from the compressed TAR archive, use the following command:</p> <pre>tar xjf Rockhopper-2.0.3.tar.bz2</pre>
---	--	---	--



# W4-1: コマンドライン版

マニュアルが一気に流れる。赤下線で示すように、最後のほうにde novoアセンブリのコマンド実行例があるのでなんとなくわかる。①マニュアルを最初から眺めるべく、「| more」をつけて、直前のコマンドを再実行。

```
java Rockhopper <options> -g genome_DIR1,genome_DIR2
cate1_pairedend1.fastq%aerobic_replicate1_pairedend2
_replicate2_pairedend1.fastq%aerobic_replicate2_paired
naerobic_replicate1_pairedend1.fastq%anaerobic_replicate1_paireden
d2.fastq,anaerobic_replicate2_pairedend1.fastq%anaerobic_replicate
2_pairedend2.fastq
```

EXAMPLE EXECUTION: DE NOVO ASSEMBLY WITH SINGLE-END READS

```
java Rockhopper <options> aerobic_replicate1.fastq,aerobic_replica
te2.fastq anaerobic_replicate1.fastq,anaerobic_replicate2.fastq
```

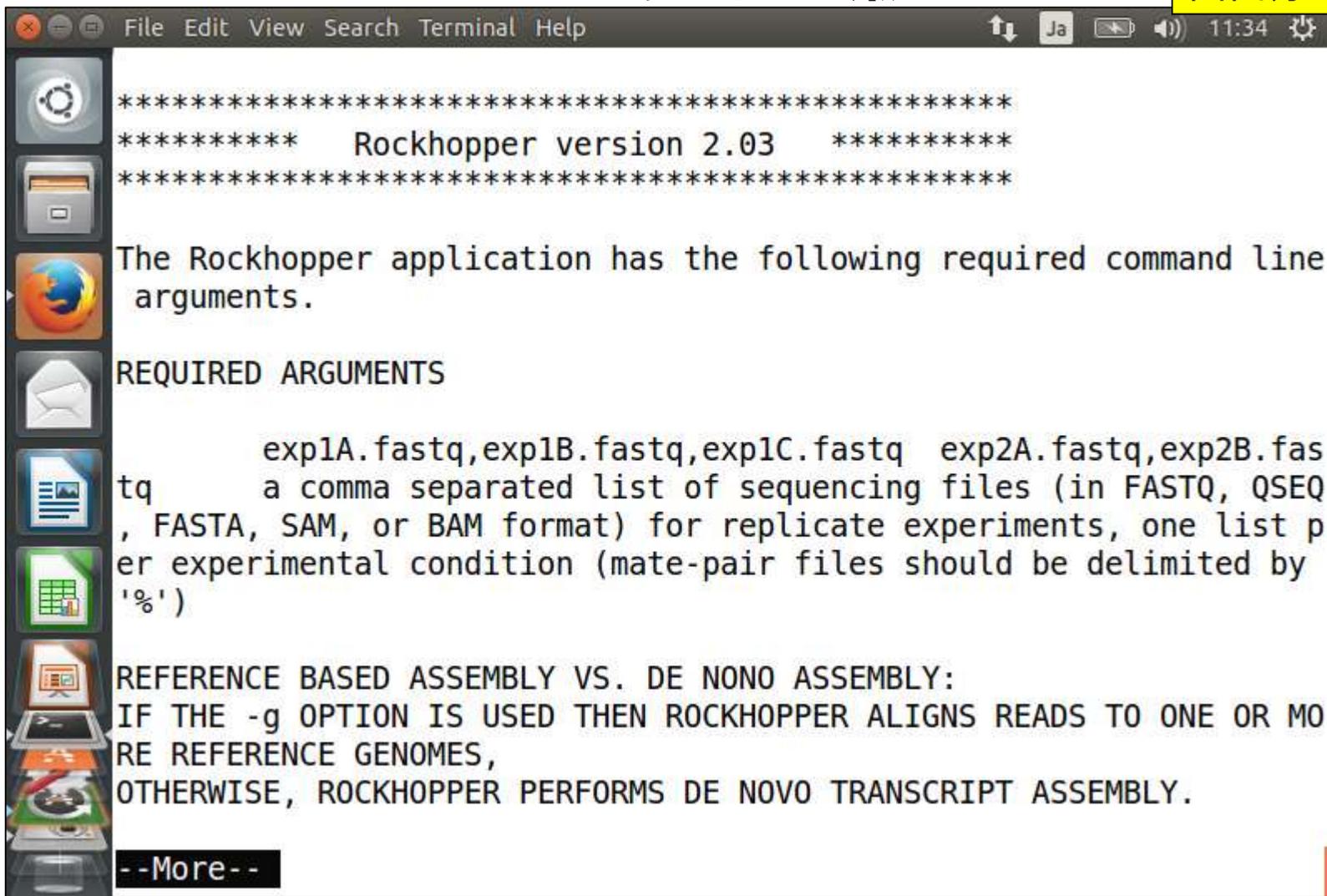
EXAMPLE EXECUTION: DE NOVO ASSEMBLY WITH PAIRED-END READS

```
java Rockhopper <options> aerobic_replicate1_pairedend1.fastq%aero
bic_replicate1_pairedend2.fastq,aerobic_replicate2_pairedend1.fast
q%aerobic_replicate2_pairedend2.fastq anaerobic_replicate1_paireden
d1.fastq%anaerobic_replicate1_pairedend2.fastq,anaerobic_replicat
e2_pairedend1.fastq%anaerobic_replicate2_pairedend2.fastq
```

① `iu@bielinux[Downloads] java -Xmx1200m -cp Rockhopper.jar Rockhoppe  
r | more`

# W4-1: コマンドライン版

moreコマンドは、「Returnキー」  
で1行分づつ、「Spaceキー」で1  
画面分づつスクロールできる



```
File Edit View Search Terminal Help 11:34
*****
*****      Rockhopper version 2.03      *****
*****

The Rockhopper application has the following required command line
arguments.

REQUIRED ARGUMENTS

    exp1A.fastq,exp1B.fastq,exp1C.fastq exp2A.fastq,exp2B.fas
tq      a comma separated list of sequencing files (in FASTQ, QSEQ
, FASTA, SAM, or BAM format) for replicate experiments, one list p
er experimental condition (mate-pair files should be delimited by
'%' )

REFERENCE BASED ASSEMBLY VS. DE NONO ASSEMBLY:
IF THE -g OPTION IS USED THEN ROCKHOPPER ALIGNS READS TO ONE OR MO
RE REFERENCE GENOMES,
OTHERWISE, ROCKHOPPER PERFORMS DE NOVO TRANSCRIPT ASSEMBLY.

--More--
```

# W4-1: オプション

この画面あたりがde novoアセンブリで使うオプションの説明。①kはTrinityと同じく25がデフォルトのようだ。②アセンブル後のコンティグの最低配列長は $2*k = 2*25 = 50$ だと解釈

## OPTIONAL ARGUMENTS FOR DE NOVO ASSEMBLY ONLY

-k <integer> size of k-mer, range of values is 15 to 31  
(default is 25)

-j <integer> minimum length required to use a sequencing read after trimming/processing (default is 35)

-n <integer> size of k-mer hashtable is  $\sim 2^n$  (default is 25). HINT: should normally be 25 or, if more memory is available, 26. WARNING: if increased above 25 then more than 1.2M of memory must be allocated

-b <integer> minimum number of full length reads required to map to a de novo assembled transcript (default is 20)

-u <integer> minimum length of de novo assembled transcripts (default is  $2*k$ )

-w <integer> minimum count of k-mer to use it to seed a new de novo assembled transcript (default is 50)

-x <integer> minimum count of k-mer to use it to extend an existing de novo assembled transcript (default is 5)

EXAMPLE EXECUTION: REFERENCE BASED ASSEMBLY WITH SINGLE-END READS

--More--



# W4-2: 実行コマンド例

オプションは特に指定する必要はなさそうだと解釈し、とりあえず実行例を再確認。これはリターンキーを連打してマニュアルの最後のほうを示している。赤下線部分あたりまでじっくり眺めて、①paired-endの2つのファイルは「%」で連結する、②反復実験データがあれば「,」でつなげていけばよいだろうということを学習する。スペースが入っていないことも記憶にとどめておく。

```
naerobic_replicate1_pairedend1.fastq%anaerobic_
d2.fastq,anaerobic_replicate2_pairedend1.fastq%a
2_pairedend2.fastq

EXAMPLE EXECUTION: DE NOVO ASSEMBLY WITH SINGLE

java Rockhopper <options> aerobic_replicate1.fastq,aerobic_replica
te2.fastq anaerobic_replicate1.fastq,anaerobic_replicate2.fastq

EXAMPLE EXECUTION: DE NOVO ASSEMBLY WITH PAIRED-END READS

java Rockhopper <options> aerobic_replicate1_pairedend1.fastq%aero
bic_replicate1_pairedend2.fastq,aerobic_replicate2_pairedend1.fast
q%aerobic_replicate2_pairedend2.fastq anaerobic_replicate1_pairede
nd1.fastq%anaerobic_replicate1_pairedend2.fastq,anaerobic_replicat
e2_pairedend1.fastq%anaerobic_replicate2_pairedend2.fastq

iu@bielinux[Downloads] [11:48午前]
iu@bielinux[Downloads] [11:48午前]
iu@bielinux[Downloads] [11:48午前]
iu@bielinux[Downloads] [11:48午前]
iu@bielinux[Downloads] [11:48午前]
```



# W4-3: Tips

①何気なしにls。②nohup.outファイルは、nohupコマンドを実行すると自動的に生成される。③Rockhopper\_Resultsディレクトリは、Rockhopperを実行時に自動生成される。これらの何気なしに実行したlsコマンドの結果と、nohupやRockhopperのマニュアルを見比べると、より理解が深まっていくものです。

```
iu@bielinux[~/Downloads]
EXAMPLE EXECUTION: DE NOVO ASSEMBLY WITH SINGLE-READ PAIRED-END READS
java Rockhopper <options> aerobic_replicate1.fastq,aerobic_replicate2.fastq
anaerobic_replicate1.fastq,anaerobic_replicate2.fastq

EXAMPLE EXECUTION: DE NOVO ASSEMBLY WITH PAIRED-END READS
java Rockhopper <options> aerobic_replicate1_pairedend1.fastq,aerobic_replicate1_pairedend2.fastq,aerobic_replicate2_pairedend1.fastq,aerobic_replicate2_pairedend2.fastq,anaerobic_replicate1_pairedend1.fastq,anaerobic_replicate1_pairedend2.fastq,anaerobic_replicate2_pairedend1.fastq,anaerobic_replicate2_pairedend2.fastq

iu@bielinux[Downloads] [11:48午前]
iu@bielinux[Downloads] [11:48午前]
iu@bielinux[Downloads] [11:48午前]
iu@bielinux[Downloads] [11:48午前]
① iu@bielinux[Downloads] ls [11:48午前]
FaQCs fastqc_v0.11.3.zip IGV_2.3.57.zip Rockhopper.jar [11:48午前]
FastQC IGV_2.3.57 nohup.out Rockhopper_Results ③
iu@bielinux[Downloads] █ [12:00午後]
```

# W4-4: クラスパスの設定

①EXAMPLE EXECUTIONのところを眺めると、実行は赤下線のようなコマンドを打つと書いてある。しかし、②それを実際に打ってみてもエラーが出る。理由はクラスパスの設定ができていないから。

```
File Edit View Search Terminal Help
java Rockhopper <options> aerobic_replicate1.fastq,aerobic_replicate2.fastq,anaerobic_replicate1.fastq,anaerobic_replicate2.fastq

① EXAMPLE EXECUTION: DE NOVO ASSEMBLY WITH PAIRED-END READS

java Rockhopper <options> aerobic_replicate1_pairedend1.fastq,aerobic_replicate1_pairedend2.fastq,aerobic_replicate2_pairedend1.fastq,aerobic_replicate2_pairedend2.fastq,anaerobic_replicate1_pairedend1.fastq,anaerobic_replicate1_pairedend2.fastq,anaerobic_replicate2_pairedend1.fastq,anaerobic_replicate2_pairedend2.fastq

iu@bielinux[Downloads] [ 5:31午後]
iu@bielinux[Downloads] ls [ 5:31午後]
FaQCs fastqc_v0.11.3.zip IGV_2.3.57.zip Rockhopper.jar
FastQC IGV_2.3.57 nohup.out Rockhopper_Results
② iu@bielinux[Downloads] java Rockhopper [ 5:31午後]
Error: Could not find or load main class Rockhopper
iu@bielinux[Downloads] [ 5:53午後]
```

# W4-4: クラスパスの設定

①クラスパスの設定は「export CLASSPATH=設定したいjarファイルの絶対パス」。著者らの環境では、Rockhopper.jarの絶対パスは赤下線のように書く。②設定後にもう一度「java Rockhopper」と打つ

EXAMPLE EXECUTION: DE NOVO ASSEMBLY WITH PAIRED-END RE

```
java Rockhopper <options> aerobic_replicate1_pairedend1.fastq%aero  
bic_replicate1_pairedend2.fastq,aerobic_replicate2_pairedend1.fast  
q%aerobic_replicate2_pairedend2.fastq anaerobic_replicate1_pairede  
nd1.fastq%anaerobic_replicate1_pairedend2.fastq,anaerobic_replicat  
e2_pairedend1.fastq%anaerobic_replicate2_pairedend2.fastq
```

```
iu@bielinux[Downloads] [ 5:31午後]
```

```
iu@bielinux[Downloads] ls [ 5:31午後]
```

```
FaQCs fastqc_v0.11.3.zip IGV_2.3.57.zip Rockhopper.jar
```

```
FastQC IGV_2.3.57 nohup.out Rockhopper_Results
```

```
iu@bielinux[Downloads] java Rockhopper [ 5:31午後]
```

```
Error: Could not find or load main class Rockhopper
```

```
iu@bielinux[Downloads] export CLASSPATH=/home/iu/Downloads/Rockhop
```

```
per.jar
```

```
iu@bielinux[Downloads] java Rockhopper [ 6:35午後]
```

①

②

# W4-4: クラスパスの設定

リターンキーを押したあとの状態。  
エラーメッセージではなく、正しくマ  
ニュアルが表示される。

```
File Edit View Search Terminal Help 18:56
java Rockhopper <options> -g genome_DIR1,genome_DIR2 aerobic_repli
cate1_pairedend1.fastq%aerobic_replicate1_pairedend2.fastq,aerobic
_replicate2_pairedend1.fastq%aerobic_replicate2_pairedend2.fastq a
naerobic_replicate1_pairedend1.fastq%anaerobic_replicate1_paireden
d2.fastq,anaerobic_replicate2_pairedend1.fastq%anaerobic_replicate
2_pairedend2.fastq

EXAMPLE EXECUTION: DE NOVO ASSEMBLY WITH SINGLE-END READS

java Rockhopper <options> aerobic_replicate1.fastq,aerobic_replica
te2.fastq anaerobic_replicate1.fastq,anaerobic_replicate2.fastq

EXAMPLE EXECUTION: DE NOVO ASSEMBLY WITH PAIRED-END READS

java Rockhopper <options> aerobic_replicate1_pairedend1.fastq%aero
bic_replicate1_pairedend2.fastq,aerobic_replicate2_pairedend1.fast
q%aerobic_replicate2_pairedend2.fastq anaerobic_replicate1_paireden
d1.fastq%anaerobic_replicate1_pairedend2.fastq,anaerobic_replicat
e2_pairedend1.fastq%anaerobic_replicate2_pairedend2.fastq

iu@bielinux[Downloads] [ 6:56午後 ]
```

# W5-1 : Rockhopper実行

- ① FaQCs実行結果ファイルを含むディレクトリに移動して、
- ② de novoアセンブリを実行。
- ③ OutofMemoryErrorという記述を発見。これはメモリが足りないことに起因するエラー。

```
File Edit View Search Terminal Help
iu@bielinux[Downloads] cd ~/Documents/srp017156/result2
iu@bielinux[result2] ls -lh *.fastq
-rw-rw-r-- 1 iu iu 299M 12月 19 12:16 QC.1.trimmed.fastq
-rw-rw-r-- 1 iu iu 269M 12月 19 12:16 QC.2.trimmed.fastq
-rw-rw-r-- 1 iu iu 5.5M 12月 19 12:17 QC.unpaired.trimmed.fastq
iu@bielinux[result2] java Rockhopper QC.1.trimmed.fastq%QC.2.trimmed.fastq

Assembling transcripts from reads in files:
    QC.1.trimmed.fastq
    QC.2.trimmed.fastq
Exception in thread "main" java.lang.OutOfMemoryError: Java heap space
    at java.util.concurrent.atomic.AtomicLongArray.<init>(AtomicLongArray
.java:80)
    at Table.<init>(Table.java:47)
    at Dictionary.assembleTranscripts(Dictionary.java:171)
    at Assembler.<init>(Assembler.java:219)
    at Rockhopper.<init>(Rockhopper.java:91)
    at Rockhopper.main(Rockhopper.java:908)
iu@bielinux[result2] █
```

[ 9:59午後 ]

# W5-1: Rockhopper実行

①lsすると、一応 Rockhopper\_Resultsディレクトリはできている。②その中身を眺めている。summary.txtのファイルサイズも0なうえ、コンティグファイルもできていないことがわかる。③ファイルサイズが0ということは中身がないということ。

```
File Edit View Search Terminal Help
Assembling transcripts from reads in files:
  QC.1.trimmed.fastq
  QC.2.trimmed.fastq
Exception in thread "main" java.lang.OutOfMemoryError: Java heap space
  at java.util.concurrent.atomic.AtomicLongArray.<init>(AtomicLongArray.java:80)
  at Table.<init>(Table.java:47)
  at Dictionary.assembleTranscripts(Dictionary.java:171)
  at Assembler.<init>(Assembler.java:219)
  at Rockhopper.<init>(Rockhopper.java:91)
  at Rockhopper.main(Rockhopper.java:908)
iu@bielinux[result2] ls [ 9:59午後 ]
fastqCount.txt      QC_qc_report.pdf      Rockhopper_Results
QC.1.trimmed.fastq QC_stats.txt
QC.2.trimmed.fastq QC.unpaired.trimmed.fastq
iu@bielinux[result2] ls -l Rockhopper_Results [10:02午後 ]
total 4
drwxrwxr-x 2 iu iu 4096 12月 20 21:58 genomeBrowserFiles
-rw-rw-r-- 1 iu iu  0 12月 20 21:58 summary.txt
iu@bielinux[result2] more Rockhopper_Results/summary.txt [10:02午後 ]
iu@bielinux[result2] [10:02午後 ]
```



①最大メモリを2GBまで増やしてリトライ。数分程度

# W5-2: Rockhopper再実行

```
File Edit View Search Terminal Help
QC.1.trimmed.fastq
QC.2.trimmed.fastq
Exception in thread "main" java.lang.OutOfMemoryError: Java heap space
  at java.util.concurrent.atomic.AtomicLongArray.<init>(AtomicLongArray
.java:80)
  at Table.<init>(Table.java:47)
  at Dictionary.assembleTranscripts(Dictionary.java:171)
  at Assembler.<init>(Assembler.java:219)
  at Rockhopper.<init>(Rockhopper.java:91)
  at Rockhopper.main(Rockhopper.java:908)
iu@bielinux[result2] ls [ 9:59午後 ]
fastqCount.txt      QC_qc_report.pdf      Rockhopper_Results
QC.1.trimmed.fastq QC.stats.txt
QC.2.trimmed.fastq QC.unpaired.trimmed.fastq
iu@bielinux[result2] ls -l Rockhopper_Results [10:02午後 ]
total 4
drwxrwxr-x 2 iu iu 4096 12月 20 21:58 genomeBrowserFiles
-rw-rw-r-- 1 iu iu  0 12月 20 21:58 summary.txt
iu@bielinux[result2] more Rockhopper_Results/summary.txt [10:02午後 ]
iu@bielinux[result2] java -Xmx2000m Rockhopper QC.1.trimmed.fastq%QC.2.trimme
d.fastq
```



# W5-2: Rockhopper再実行

①「-Xmx2000m」オプションをつける前は、赤枠の途中経過が出る前にOutOfMemoryErrorとなっていたので、②このようなメッセージが出るのを見られただけでもうれしいものです

```
at Rockhopper.main(Rockhopper.java:908)
iu@bielinux[result2] ls
fastqCount.txt      QC_qc_report.pdf      Rockhopper_Results
QC.1.trimmed.fastq QC_stats.txt
QC.2.trimmed.fastq QC.unpaired.trimmed.fastq
iu@bielinux[result2] ls -l Rockhopper_Results      [10:02午後]
total 4
drwxrwxr-x 2 iu iu 4096 12月 20 21:58 genomeBrowserFiles
-rw-rw-r-- 1 iu iu   0 12月 20 21:58 summary.txt
iu@bielinux[result2] more Rockhopper_Results/summary.txt      [10:02午後]
iu@bielinux[result2] java -Xmx2000m Rockhopper QC.1.trimmed.fastq%QC.2.trimme
d.fastq

Assembling transcripts from reads in files:
    QC.1.trimmed.fastq
    QC.2.trimmed.fastq

Aligning reads to assembled transcripts using files:
    QC.1.trimmed.fastq
    QC.2.trimmed.fastq
```

①

②

# W5-2: Rockhopper再実行

①無事de novoアセンブリが終了し、コマンド入力待ち状態になっている。但し、②アセンブルされた転写物(transcripts)は1つも無いことがわかる。③おそらくこれはバグ。アセンブルされたコンティグ(転写物)が1つも無いのに、35リードがマップされたというのは論理的におかしい。

```
File Edit View Search Terminal Help
Aligning reads to assembled transcripts using files:
  QC.1.trimmed.fastq
  QC.2.trimmed.fastq

Total reads in files:          976468
Perfectly aligned reads:      35      0%

Total number of assembled transcripts: 0
Average transcript length:      0
Median transcript length:       0
Total number of assembled bases: 0

Summary of results written to file:      Rockhopper_Results/su
mary.txt
Details of assembled transcripts written to file:
anscripts.txt

FINISHED.

① iu@bielinux[result2] [10:10午後]
```

# W5-3: 実行結果概観

① Rockhopper\_Resultsディレクトリの中身は、エラーを吐いたとき(W5-1)とは異なることがわかる。②summary.txtの中身は、赤枠でも示されているように、基本的に画面に表示されていたアセンブル結果の要約情報が含まれている

```
iu@bielinux[~/Documents/srp017156/result2]
Median transcript length:
Total number of assembled bases:

Summary of results written to file:                Rockhopper_Results/summary.txt
Details of assembled transcripts written to file:   Rockhopper_Results/transcripts.txt

FINISHED.

iu@bielinux[result2] ls                               [10:10午後]
fastqCount.txt      QC_qc_report.pdf      Rockhopper_Results
QC.1.trimmed.fastq QC.stats.txt
QC.2.trimmed.fastq QC.unpaired.trimmed.fastq

① iu@bielinux[result2] ls -l Rockhopper_Results      [10:17午後]
total 16
drwxrwxr-x 2 iu iu 4096 12月 20 21:58 genomeBrowserFiles
drwxrwxr-x 2 iu iu 4096 12月 20 22:10 intermediary
-rw-rw-r-- 1 iu iu  608 12月 20 22:10 summary.txt
-rw-rw-r-- 1 iu iu   29 12月 20 22:10 transcripts.txt
iu@bielinux[result2] █                               [10:17午後]
```

# W5-3: 実行結果概観

①アセンブルされた転写物配列情報は transcripts.txtファイルに格納される。ただし、この場合は1つもコンティグがないので、②moreでファイルの中身を表示させても「Sequence Length Expression 1」というヘッダ一行しかないことがわかる。

```
File Edit View Search Terminal Help
Total number of assembled bases:

Summary of results written to file:
Results/summary.txt
Details of assembled transcripts written to file:      Rockhopper
Results/transcripts.txt

FINISHED.

iu@bielinux[result2] ls                                [11:02午前]
fastqCount.txt      QC_qc_report.pdf      Rockhopper_Results
QC.1.trimmed.fastq QC.stats.txt
QC.2.trimmed.fastq QC.unpaired.trimmed.fastq
iu@bielinux[result2] ls -l Rockhopper_Results         [11:55午前]
total 16
drwxrwxr-x 2 iu iu 4096  9月  8 11:01 genomeBrowserFiles
drwxrwxr-x 2 iu iu 4096  9月  8 11:02 intermediary
-rw-rw-r-- 1 iu iu  608  9月  8 11:02 summary.txt
-rw-rw-r-- 1 iu iu   29  9月  8 11:02 transcripts.txt
iu@bielinux[result2] more Rockhopper_Results/transcripts.txt
Sequence          Length  Expression 1
iu@bielinux[result2] █                                [11:58午前]
```



# W6-1 : single-endで実行

①single-endとしてforward側のみ  
のファイル(QC.1.trimmed.fastq)を入  
力として実行。nohupをつけてバック  
グラウンドで実行したので、途中経  
過はターミナル画面上には表示さ  
れない。画面出力される内容は、赤  
下線で示すようにnohup.outというフ  
ァイルに追加で書き込まれる。

```
File Edit View Search Terminal Help
summary.txt
Details of assembled transcripts written to file:
anscripts.txt

FINISHED.

iu@bielinux[result2] ls -l Rockhopper_Results [10:43午後]
total 16
drwxrwxr-x 2 iu iu 4096 12月 20 22:42 genomeBrowserFiles
drwxrwxr-x 2 iu iu 4096 12月 20 22:43 intermediary
-rw-rw-r-- 1 iu iu 608 12月 20 22:43 summary.txt
-rw-rw-r-- 1 iu iu 29 12月 20 22:43 transcripts.txt
iu@bielinux[result2] more Rockhopper_Results/transcripts.txt [10:45午後]
Sequence      Length  Expression 1
iu@bielinux[result2] nohup java -Xmx2000m Rockhopper QC.1.trimmed.fastq&
[2] 25122
iu@bielinux[result2] nohup: ignoring input and appending output to 'nohup.out'
[2] + done      nohup java -Xmx2000m Rockhopper QC.1.trimmed.fastq
iu@bielinux[result2] [10:49午後]
```



# W6-1 : single-endで実行

②Rockhopper\_Results中の transcripts.txtのファイルサイズが145 bytesとなっていることから、何かしらアセンブルされた結果があるのだろうと解釈する。③summary.txtをlessで眺める。

```
iu@bielinux[result2] ls -l Rockhopper_Results
total 16
drwxrwxr-x 2 iu iu 4096 12月 20 22:42 genomeBrowserFiles
drwxrwxr-x 2 iu iu 4096 12月 20 22:43 intermediary
-rw-rw-r-- 1 iu iu 608 12月 20 22:43 summary.txt
-rw-rw-r-- 1 iu iu 29 12月 20 22:43 transcripts.txt
iu@bielinux[result2] more Rockhopper_Results/transcripts.txt [10:45午後]
Sequence          Length Expression 1
① iu@bielinux[result2] nohup java -Xmx2000m Rockhopper QC.1.trimmed.fastq&
[2] 25122
iu@bielinux[result2] nohup: ignoring input and appending output to 'nohup.out'
[2] + done          nohup java -Xmx2000m Rockhopper QC.1.trimmed.fastq
② iu@bielinux[result2] ls -l Rockhopper_Results [10:49午後]
total 16
drwxrwxr-x 2 iu iu 4096 12月 20 22:42 genomeBrowserFiles
drwxrwxr-x 2 iu iu 4096 12月 20 22:49 intermediary
-rw-rw-r-- 1 iu iu 582 12月 20 22:49 summary.txt
-rw-rw-r-- 1 iu iu 145 12月 20 22:49 transcripts.txt
③ iu@bielinux[result2] less Rockhopper_Results/summary.txt [10:54午後]
```

# W6-2: single-endで実行

summary.txtのless実行結果。①アセンブル結果として転写物が1つだけ得られたと解釈する。しかし、その長さは107 bp。入力も107 bpなので、どれか1つのリードを出力したのと同じじゃないかと苦笑。②qで抜ける。quitのqです。

```
File Edit View Search Terminal Help
Assembling transcripts from reads in file:
Aligning reads to assembled transcripts using file:
    Total reads in file:          976519
    Perfectly aligned reads:      119      0%
Total number of assembled transcripts: 1
Average transcript length:       107
Median transcript length:        107
Total number of assembled bases: 107
Summary of results written to file: Rockhopper_Results/su
mary.txt
Details of assembled transcripts written to file: Rockhopper_Results/tr
anscripts.txt
FINISHED.
Rockhopper_Results/summary.txt (END)
```



# W6-2: single-endで実行

```

File Edit View Search Terminal Help
total 16
drwxrwxr-x 2 iu iu 4096 12月 20 22:42 genomeBrowserFiles
drwxrwxr-x 2 iu iu 4096 12月 20 22:43 intermediary
-rw-rw-r-- 1 iu iu 608 12月 20 22:43 summary.txt
-rw-rw-r-- 1 iu iu 29 12月 20 22:43 transcripts.txt
iu@bielinux[result2] more Rockhopper_Results/transcripts.txt [10:45午後]
Sequence          Length Expression 1
iu@bielinux[result2] nohup java -Xmx2000m Rockhopper QC.1.trimmed.fastq&
[2] 25122
iu@bielinux[result2] nohup: ignoring input and appending output to 'nohup.out
'

[2] + done          nohup java -Xmx2000m Rockhopper QC.1.trimmed.fastq
iu@bielinux[result2] ls -l Rockhopper_Results [10:49午後]
total 16
drwxrwxr-x 2 iu iu 4096 12月 20 22:42 genomeBrowserFiles
drwxrwxr-x 2 iu iu 4096 12月 20 22:49 intermediary
-rw-rw-r-- 1 iu iu 582 12月 20 22:49 summary.txt
-rw-rw-r-- 1 iu iu 145 12月 20 22:49 transcripts.txt
iu@bielinux[result2] less Rockhopper_Results/summary.txt [10:54午後]
iu@bielinux[result2] [10:58午後]

```

# W6-3: mvでrename

ここまでの作業で、Rockhopperはアセンブル実行結果ファイルを上書き保存していることがわかる。この後に行うreverse側のsingle-endのアセンブリで結果が消えてしまわぬように、forward側の実行結果ファイルの名前を変更しておく。

```
File Edit View Search Terminal Help
[2] + done      nohup java -Xmx2000m Rockhopper QC.1.trimmed.fastq
iu@bielinux[result2] ls -l Rockhopper_Results [10:49午後]
total 16
drwxrwxr-x 2 iu iu 4096 12月 20 22:42 genomeBrowserFiles
drwxrwxr-x 2 iu iu 4096 12月 20 22:49 intermediary
-rw-rw-r-- 1 iu iu  582 12月 20 22:49 summary.txt
-rw-rw-r-- 1 iu iu  145 12月 20 22:49 transcripts.txt
iu@bielinux[result2] less Rockhopper_Results/summary.txt [10:54午後]
iu@bielinux[result2] mv Rockhopper_Results/summary.txt Rockhopper_Results/summary_1.txt
iu@bielinux[result2] mv Rockhopper_Results/transcripts.txt Rockhopper_Results/transcripts_1.txt
iu@bielinux[result2] ls -l Rockhopper_Results [11:05午後]
total 16
drwxrwxr-x 2 iu iu 4096 12月 20 22:42 genomeBrowserFiles
drwxrwxr-x 2 iu iu 4096 12月 20 22:49 intermediary
-rw-rw-r-- 1 iu iu  582 12月 20 22:49 summary_1.txt
-rw-rw-r-- 1 iu iu  145 12月 20 22:49 transcripts_1.txt
iu@bielinux[result2] [11:05午後]
```

①

②

# W6-4: reverse側を実行

①reverse側ファイル(QC.2.trimmed.fastq)を入力としてsingle-endのアセンブリを実行。nohupと&をつけてないので、途中経過(summary.txtと同じもの)がターミナル画面上に出力される。

```
iu@bielinux[result2] java -Xmx2000m Rockhopper QC.2.trimmed.fastq
Assembling transcripts from reads in file: QC.2.trimmed.fastq
Aligning reads to assembled transcripts using file: QC.2.trimmed.fastq

Total reads in file: 977151
Perfectly aligned reads: 706568 72%

Total number of assembled transcripts: 423
Average transcript length: 437
Median transcript length: 228
Total number of assembled bases: 184929

Summary of results written to file: Rockhopper_Results/summary.txt
Details of assembled transcripts written to file: Rockhopper_Results/transcripts.txt

FINISHED.

iu@bielinux[result2] █ [11:10午後]
```

# W6-4: reverse側を実行

①アセンブルされた転写物数は423個!  
②総塩基数は184,929。③入力リード数  
977,151個のうち、72% (706,568個)がマ  
ップされていることがわかる。

```
iu@bielinux[result2] java -Xmx2000m Rockhopper QC.2.trimmed.fastq
Assembling transcripts from reads in file:          QC.2.trimmed.fastq
Aligning reads to assembled transcripts using file: QC.2.trimmed.fastq
Total reads in file:                               977151
Perfectly aligned reads:                           706568 72%
Total number of assembled transcripts:              423
Average transcript length:                         437
Median transcript length:                          228
Total number of assembled bases:                   184929
Summary of results written to file:                Rockhopper_Results/summary.txt
Details of assembled transcripts written to file:   Rockhopper_Results/transcripts.txt
FINISHED.
iu@bielinux[result2] █ [11:10午後]
```

念のため、reverse側の実行結果ファイルの名前を\*\_2.txtに変更しているだけです

# W6-4: mvでrename

```
iu@bielinux[result2] ls -l Rockhopper_Results [11:10午後]
total 208
drwxrwxr-x 2 iu iu 4096 12月 20 22:42 genomeBrowserFiles
drwxrwxr-x 2 iu iu 4096 12月 20 23:09 intermediary
-rw-rw-r-- 1 iu iu 582 12月 20 22:49 summary_1.txt
-rw-rw-r-- 1 iu iu 591 12月 20 23:09 summary.txt
-rw-rw-r-- 1 iu iu 145 12月 20 22:49 transcripts_1.txt
-rw-rw-r-- 1 iu iu 188726 12月 20 23:09 transcripts.txt

① iu@bielinux[result2] mv Rockhopper_Results/summary.txt Rockhopper_Results/summary_2.txt
② iu@bielinux[result2] mv Rockhopper_Results/transcripts.txt Rockhopper_Results/transcripts_2.txt

iu@bielinux[result2] ls -l Rockhopper_Results [11:15午後]
total 208
drwxrwxr-x 2 iu iu 4096 12月 20 22:42 genomeBrowserFiles
drwxrwxr-x 2 iu iu 4096 12月 20 23:09 intermediary
-rw-rw-r-- 1 iu iu 582 12月 20 22:49 summary_1.txt
-rw-rw-r-- 1 iu iu 591 12月 20 23:09 summary_2.txt
-rw-rw-r-- 1 iu iu 145 12月 20 22:49 transcripts_1.txt
-rw-rw-r-- 1 iu iu 188726 12月 20 23:09 transcripts_2.txt

iu@bielinux[result2] [11:15午後]
```

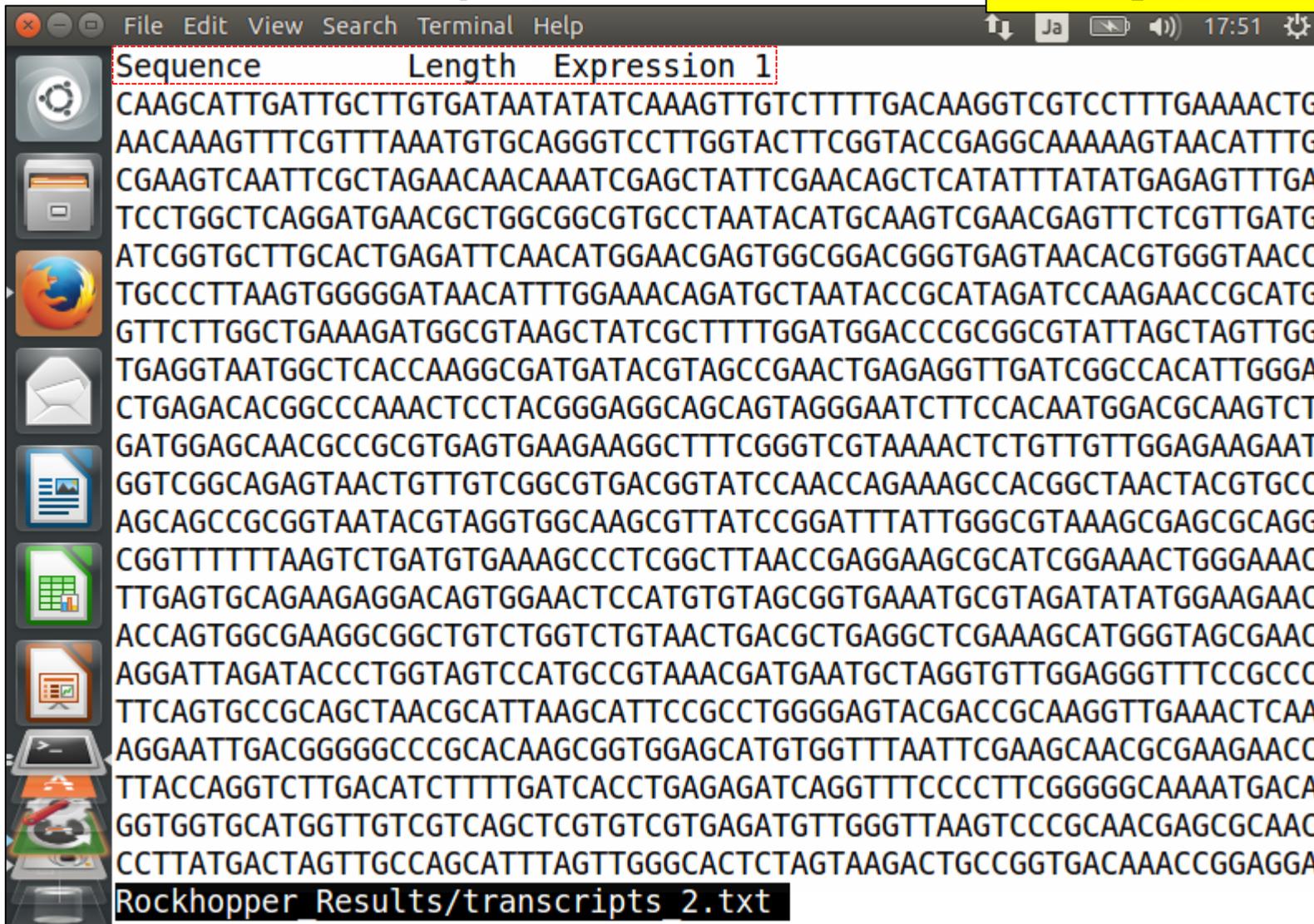
# W6-5: 結果を眺める

①reverse側の実行結果ファイルtranscripts\_2.txtの行数をwcで調べる。424行だったが、最初の1行目はヘッダー行なので、423 transcriptsの結果と矛盾はない。②lessで眺める。

```
iu@bielinux[result2] pwd [11:18午後]
/home/iu/Documents/srp017156/result2
iu@bielinux[result2] ls -l Rockhopper_Results [11:19午後]
total 208
drwxrwxr-x 2 iu iu 4096 12月 20 22:42 genomeBrowserFiles
drwxrwxr-x 2 iu iu 4096 12月 20 23:09 intermediary
-rw-rw-r-- 1 iu iu 582 12月 20 22:49 summary_1.txt
-rw-rw-r-- 1 iu iu 591 12月 20 23:09 summary_2.txt
-rw-rw-r-- 1 iu iu 145 12月 20 22:49 transcripts_1.txt
-rw-rw-r-- 1 iu iu 188726 12月 20 23:09 transcripts_2.txt
① iu@bielinux[result2] wc Rockhopper_Results/transcripts_2.txt [11:19午後]
  424  1273 188726 Rockhopper_Results/transcripts_2.txt
② iu@bielinux[result2] less Rockhopper_Results/transcripts_2.txt [11:19午後]
```

# W6-5: 結果を眺める

①lessで開いた直後の状態。赤枠部分がヘッダ行。ファイル末尾に移動したい場合は「G」、先頭に移動したい場合は「g」。



```
File Edit View Search Terminal Help
Sequence Length Expression 1
CAAGCATTGATTGCTTGTGATAATATATCAAAGTTGTCTTTTGACAAGGTCGTCCTTTGAAAAC TG
AACAAAGTTTCGTTTAAATGTGCAGGGTCCTTGGTACTTCGGTACCGAGGCAAAAAGTAACATTTG
CGAAGTCAATTCGCTAGAACAACAAATCGAGCTATTCGAACAGCTCATATTTATATGAGAGTTTGA
TCCTGGCTCAGGATGAACGCTGGCGGCGTGCCTAATACATGCAAGTCGAACGAGTTCTCGTTGATG
ATCGGTGCTTGCAGTCAACATGGAACGAGTGGCGGACGGGTGAGTAACACGTGGGTAACC
TGCCCTTAAGTGGGGGATAACATTTGAAACAGATGCTAATACCGCATAGATCCAAGAACCGCATG
GTTCTTGGCTGAAAGATGGCGTAAGCTATCGCTTTTGGATGGACCCGCGGCGTATTAGCTAGTTGG
TGAGGTAATGGCTCACCAAGGCGATGATACGTAGCCGAACTGAGAGGTTGATCGGCCACATTGGGA
CTGAGACACGGCCAAACTCCTACGGGAGGCAGCAGTAGGGAATCTTCCACAATGGACGCAAGTCT
GATGGAGCAACGCCGCGTGAGTGAAGAAGGCTTTCGGGTCGTAAACTCTGTTGTTGGAGAAGAAT
GGTCGGCAGAGTAACTGTTGTCGGCGTGACGGTATCCAACCAGAAAGCCACGGCTAACTACGTGCC
AGCAGCCGCGGTAATACGTAGGTGGCAAGCGTTATCCGGATTTATTGGGCGTAAAGCGAGCGCAGG
CGGTTTTTTAAGTCTGATGTGAAAGCCCTCGGCTTAACCGAGGAAGCGCATCGGAAACTGGGAAAC
TTGAGTGCAGAAGAGGACAGTGGAACTCCATGTGTAGCGGTGAAATGCGTAGATATATGGAAGAAC
ACCAGTGGCGAAGGCGGCTGTCTGGTCTGTAAGTACGCTGAGGCTCGAAAGCATGGGTAGCGAAC
AGGATTAGATACCCTGGTAGTCCATGCCGTAAACGATGAATGCTAGGTGTTGGAGGGTTTCCGCC
TTCAGTGCCGAGCTAACGCATTAAGCATTCCGCCTGGGGAGTACGACCGCAAGGTTGAAACTCAA
AGGAATTGACGGGGGCCCGCACAAAGCGGTGGAGCATGTGGTTTAATTCGAAGCAACGCGAAGAACC
TTACCAGGTCTTGACATCTTTTGATCACCTGAGAGATCAGGTTTCCCCTTCGGGGGCAAAATGACA
GGTGGTGCATGGTTGTCGTCAGCTCGTGTGAGATGTTGGGTTAAGTCCCGCAACGAGCGCAAC
CCTTATGACTAGTTGCCAGCATTAGTTGGGCACTCTAGTAAGACTGCCGGTGACAAACCGGAGGA
Rockhopper Results/transcripts 2.txt
```

# W6-5: 結果を眺める

ファイル末尾に移動。赤枠内に見えているのは、計4transcripts分の情報。2列目がLengthなので、赤下線部分が転写物の長さ情報に相当。qで抜ける。

```
File Edit View Search Terminal Help 18:06
128 113
TATGCTGATCGGGAAATTTCTGAATTATCCGGTGGTCAACAACAGCGAGTTGCCATTGCTCGAGCG
ATTGTAICTGAGCCGCAAGTATTGCTGCTAGATGAACCGTTATCAGCACTTGACGCCAAATTGCGT
AAGGATATGCAATATGAATTGCGCGAATTGCAGGAACGGTTGGGGATCACTTTCCTATTTGTGACA
CACGACCAAGAAGAAGCGTTGGCTTTGTGCGGACGAAATTTTGTGATGAACGATGGTGAAGTGCAA
CAAAGTGGTACGCCAGTTGATATTTATGATGAGCCGGTCAATCATTTTGTGGCGGATTTTCATTGGT
GAAAGTAACATCATTCAAGGGCACATGATTAAGGACTTTTTAGTTGAGTTCAATGGCAAACGGTTT
GAATGTGCCGATGCCGGAATGCG 419 138
TCACTCAACCAACCCCGCCAAAGACGGCGAAACGGTTGGCTTGACGTTTGATCCTGAGGACATCCA
TGTCATGCGGCTTAACGAATCTGAAGAAGATTTGACGCTCGGCTGGAACCTACGAAGGGGAATA
ACGTCAGTAAGGATCAGTTGGGGCAGTGTCATCGAAGCGTTTTGACGATTAGATAGACACTGGTCG
GCGCTGATCGATTTTCGAACTGAGGAGGGTCATCGTGAAAAAATCCACCACAAACGCCGCATTCTA
CACACCTTATGTGATGTGGCTGGCATTATTTGTGA 299 162
CTTGCCAATTATCAGACCTATTTTCAATCAGGCACTTATATTATGATGACGATCAATTCCGTCTGG
TATGCGTTCCTGATCACGTTAGCAACTTTGTTAATCAGCTATCCAACAGCATATTTGCTGCATTAT
GCCAAGCATAAACAGTTATGGCTGTTGCTGATTATTCTACCGACTTGGATTAAC 186
139
AATAGTTTCTTGGGAATGTTTGGCATCGCGCCCCAGCAATTCCTGTTACGGATTTTAGTTTCATT
TTTGTTGCTGCTTATATTGAAATCCCGTTCATGATCTTGCCAATTTTCAATGCAATCGAAGAATTA
CCGGAAAACCTTGTCAATGCTGCCAAGATTTAGGGGCCAAAAGCTGGCAGACCTTCACCAAGGTG
ATCTGGCCGTTGACAATTTCTGGCGTGAAATCCGGGGTTCAAGCTGTTTTTCATTCCAAGTTTGAGT
TTATTCATGT 274 121
(END)
```

# W7-1: Rの起動

①Rの起動は、「R」と打ってリターンキーを押すだけ。②Rのバージョンは3.2.0であることがわかる。③「>」となっていれば、コマンド入力待ち状態。

```
iu@bielinux[~/Documents/srp017156/result2]
iu@bielinux[result2] pwd
/home/iu/Documents/srp017156/result2
iu@bielinux[result2] R
R version 3.2.0 (2015-04-16) -- "Full of Ingredients"
Copyright (C) 2015 The R Foundation for Statistical Computing
Platform: x86_64-pc-linux-gnu (64-bit)

R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.

Natural language support but running in an English locale

R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.

Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.

>
```

[ 4:51午後 ]

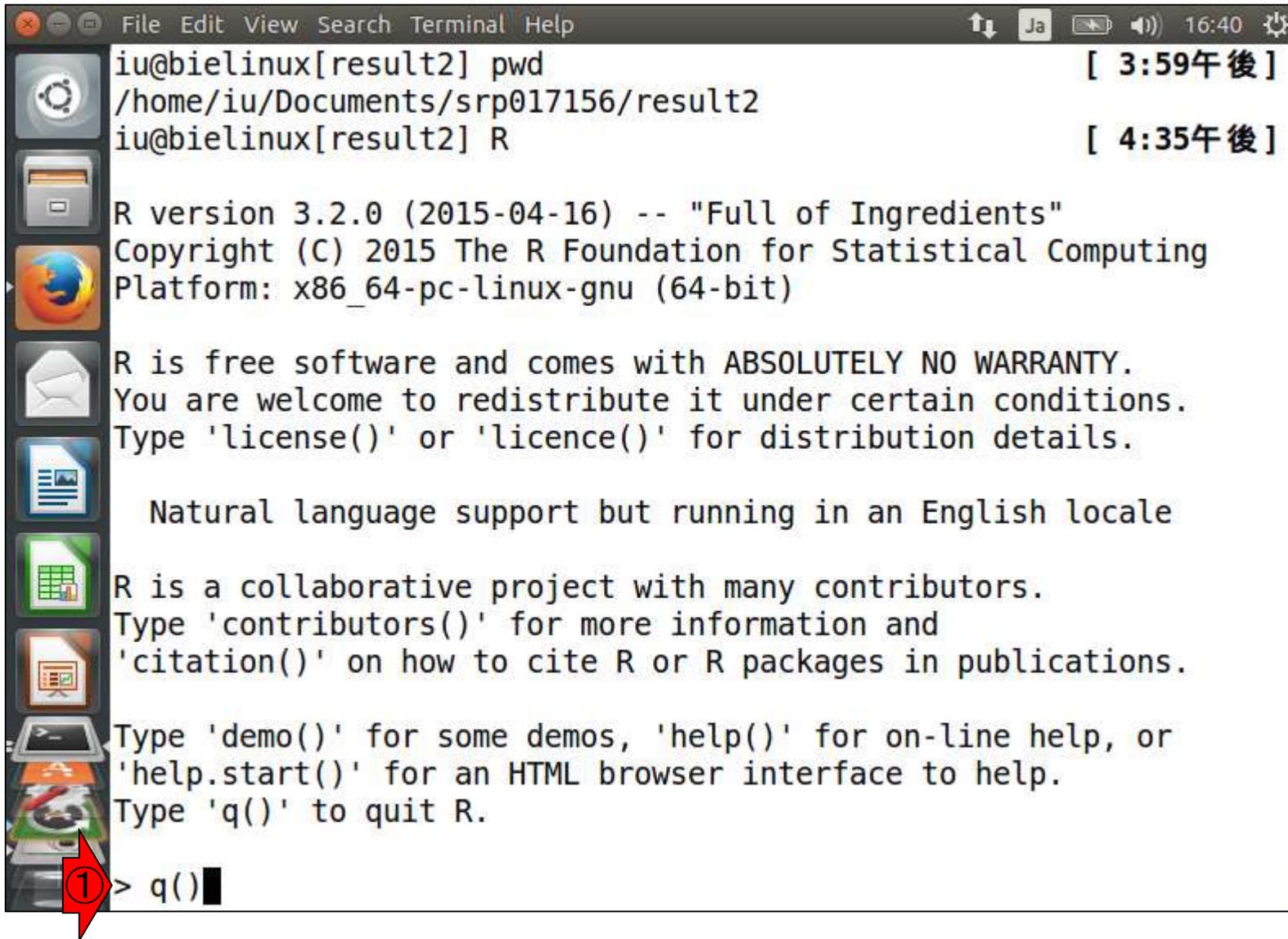
[ 4:51午後 ]

①

②

③

# W7-1: Rの終了



```
iu@bielinux[result2] pwd [ 3:59午後 ]
/home/iu/Documents/srp017156/result2
iu@bielinux[result2] R [ 4:35午後 ]

R version 3.2.0 (2015-04-16) -- "Full of Ingredients"
Copyright (C) 2015 The R Foundation for Statistical Computing
Platform: x86_64-pc-linux-gnu (64-bit)

R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.

Natural language support but running in an English locale

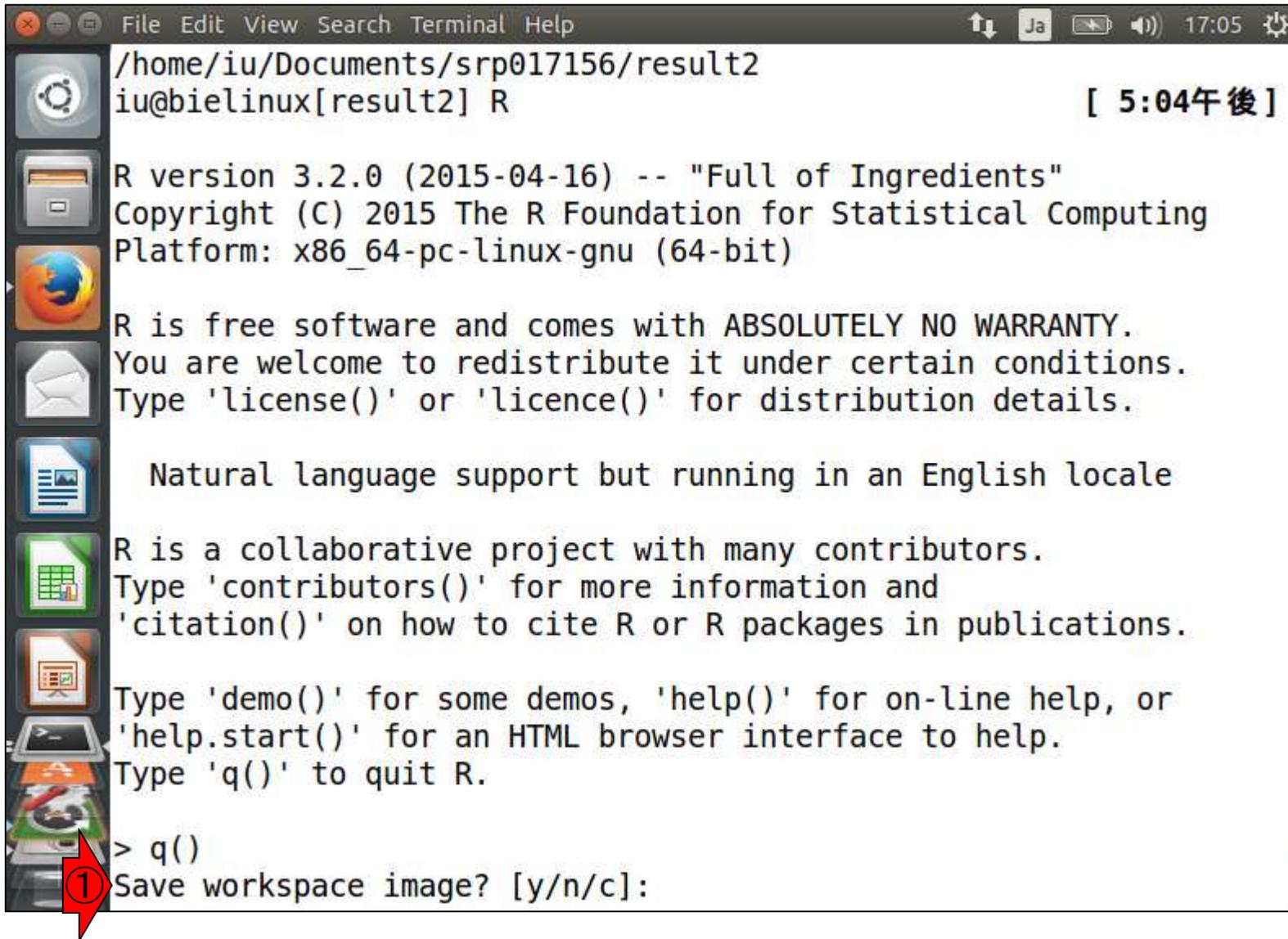
R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.

Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.

① > q() █
```

①「Save workspace image?」と聞かれる。この意味がよくわからないうちは、Noに相当する「n」を打つ。

# W7-1: Rの終了



The image shows a terminal window with the following content:

```
File Edit View Search Terminal Help
/home/iu/Documents/srp017156/result2
iu@bielinux[result2] R [ 5:04午後 ]

R version 3.2.0 (2015-04-16) -- "Full of Ingredients"
Copyright (C) 2015 The R Foundation for Statistical Computing
Platform: x86_64-pc-linux-gnu (64-bit)

R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.

Natural language support but running in an English locale

R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.

Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.

> q()
Save workspace image? [y/n/c]:
```

A red arrow with the number 1 points to the prompt "Save workspace image? [y/n/c]:".

# W7-1: Rの終了

①「n」と打ってリターンした直後の状態。②通常のBio-Linuxのコマンド入力待ち状態に戻ったことがわかる。

```
iu@bielinux[result2] R [ 5:04午後 ]
R version 3.2.0 (2015-04-16) -- "Full of Ingredients"
Copyright (C) 2015 The R Foundation for Statistical Computing
Platform: x86_64-pc-linux-gnu (64-bit)

R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.

Natural language support but running in an English locale

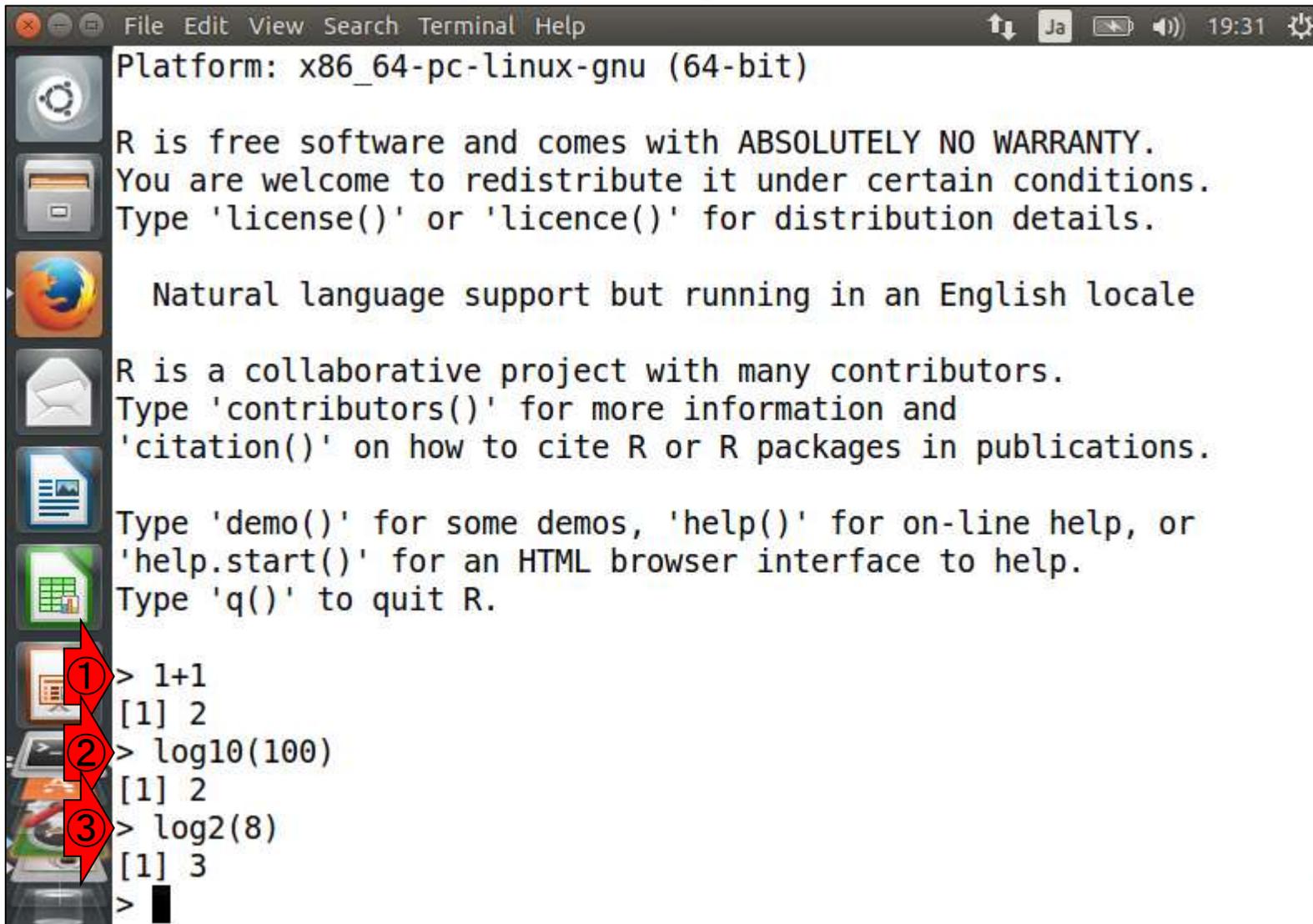
R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.

Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.

> q()
Save workspace image? [y/n/c]: n
iu@bielinux[result2] [ 5:08午後 ]
```



# W7-2: 基本的な利用法



```
Platform: x86_64-pc-linux-gnu (64-bit)

R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.

Natural language support but running in an English locale

R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.

Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.

> 1+1
[1] 2
> log10(100)
[1] 2
> log2(8)
[1] 3
>
```

The terminal window displays the R startup screen with various instructions. Three red arrows on the left side of the terminal point to the first three lines of code: `> 1+1`, `> log10(100)`, and `> log2(8)`.

# W7-3: パッケージのロード

①QuasRパッケージを利用したい場合はlibraryという関数を用いてロードする。Bio-Linux8にはまだQuasRがインストールされていないので、「そのようなパッケージはない」と文句を言われていることがわかる。②一旦終了。

```
File Edit View Search Terminal Help
Type 'license()' or 'licence()' for distribution details
Natural language support but running in an English locale
R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.
Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.

> 1+1
[1] 2
> log10(100)
[1] 2
> log2(8)
[1] 3
① > library(QuasR)
② Error in library(QuasR) : there is no package called 'QuasR'
> q()
Save workspace image? [y/n/c]: n
iu@bielinux[result2] [ 7:23午後 ]
```

# W7-4: パッケージインストール



①作業ディレクトリはどこでもいいので「sudo R」。これはroot (管理者)権限でRを実行するという指令。パッケージのインストール時に書き込み権限がないことに起因するエラー回避が目的。②パスワードを聞かれたらログインパスワード(推奨手順通りだとpass1409)を打ち込む。

# W7-4: パッケージインストール

①「source(“…”)」を打ち込む。これは赤下線部分で示す biocLite というインストール用の関数を利用できるようにするためのおまじないのようなもの。ネットワーク経由でのインストール作業になる。有線LAN環境が望ましい。

```
iu@bielinux[result2] sudo R [ 7:42]
[sudo] password for iu:
R version 3.2.0 (2015-04-16) -- "Full of Ingredients"
Copyright (C) 2015 The R Foundation for Statistical Computing
Platform: x86_64-pc-linux-gnu (64-bit)

R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.

Natural language support but running in an English locale

R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.

Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.

> source("http://bioconductor.org/biocLite.R")
```



# W7-4: QuasRインストール

①biocLite関数を用いてQuasRをインストール。赤下線部分を変えることで、同じノリで他の任意のパッケージをインストール可能。東大有線LAN環境でインストール完了まで約20分。

```
File Edit View Search Terminal Help
R version 3.2.0 (2015-04-16) -- "Full of Ingredients"
Copyright (C) 2015 The R Foundation for Statistical Computing
Platform: x86_64-pc-linux-gnu (64-bit)

R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.

Natural language support but running in an English locale

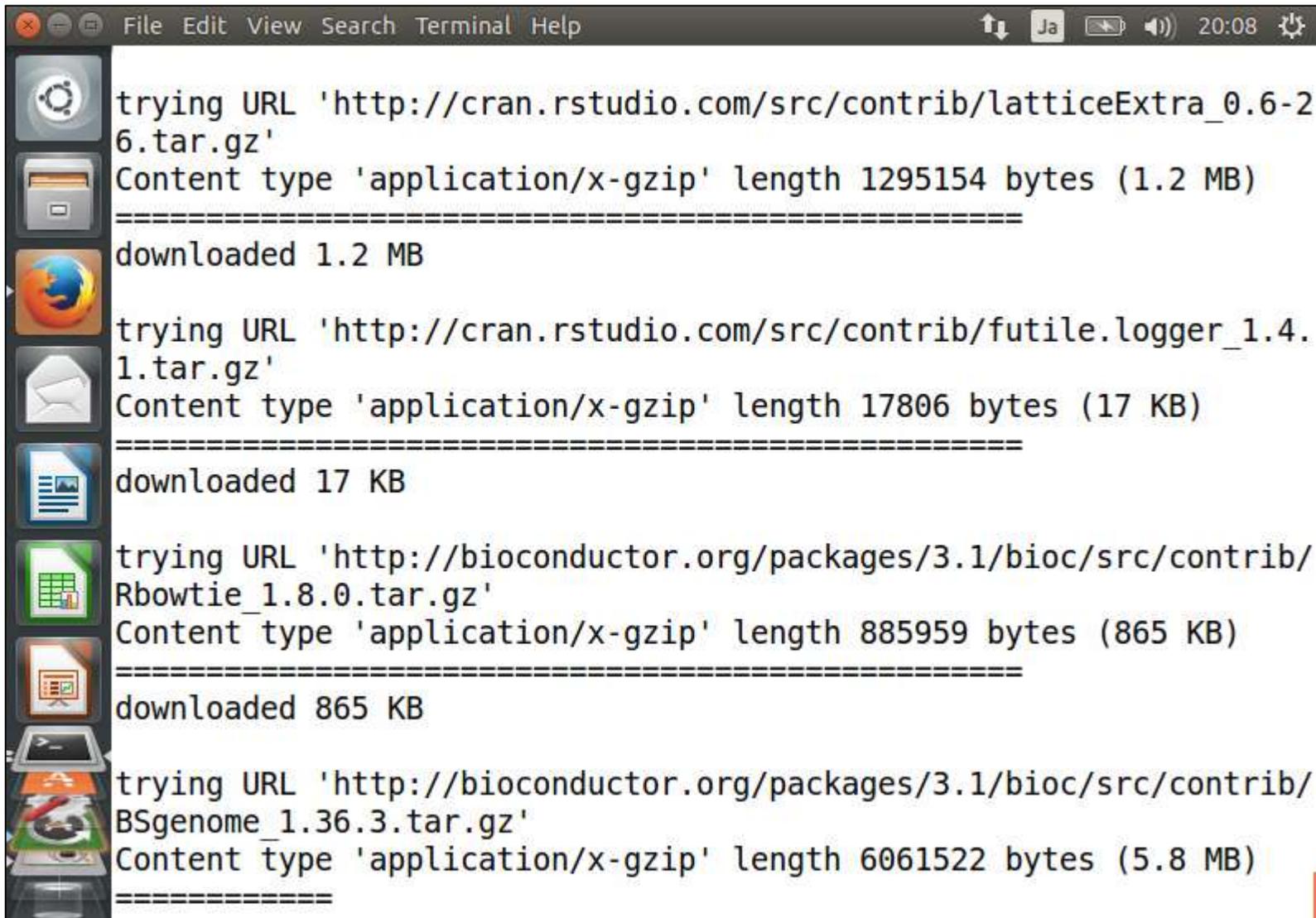
R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.

Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.

> source("http://bioconductor.org/biocLite.R")
Bioconductor version 3.1 (BiocInstaller 1.18.3), ?biocLite for help
> biocLite("QuasR")
```



# W7-5: 途中経過1



```
File Edit View Search Terminal Help 20:08
trying URL 'http://cran.rstudio.com/src/contrib/latticeExtra_0.6-26.tar.gz'
Content type 'application/x-gzip' length 1295154 bytes (1.2 MB)
=====
downloaded 1.2 MB

trying URL 'http://cran.rstudio.com/src/contrib/futile.logger_1.4.1.tar.gz'
Content type 'application/x-gzip' length 17806 bytes (17 KB)
=====
downloaded 17 KB

trying URL 'http://bioconductor.org/packages/3.1/bioc/src/contrib/Rbowtie_1.8.0.tar.gz'
Content type 'application/x-gzip' length 885959 bytes (865 KB)
=====
downloaded 865 KB

trying URL 'http://bioconductor.org/packages/3.1/bioc/src/contrib/BSgenome_1.36.3.tar.gz'
Content type 'application/x-gzip' length 6061522 bytes (5.8 MB)
=====
```

# W7-5: 途中経過2

①リターンキーを押してから約5分後にこのような状態になる。古いパッケージのアップデートをするかどうかを聞かれている。基本はすべてアップデートの「a」か、アップデートしないの「n」。②ここでは「a」と打ってリターン。

```
File Edit View Search Terminal Help
** building package indices
** installing vignettes
** testing if installed package can be loaded
* DONE (QuasR)

The downloaded source packages are in
      '/tmp/RtmpS1T00B/downloaded_packages'
Old packages: 'annotate', 'biomaRt', 'Biostrings', 'BitSeq', 'DESeq', 'edgeR',
              'evaluate', 'gdata', 'gee', 'GenomeInfoDb', 'GenomicRanges', 'HilbertVis',
              'IRanges', 'limma', 'lme4', 'matrixStats', 'pcaMethods', 'plotrix',
              'prettyR', 'qvalue', 'Rcpp', 'RcppEigen', 'RCurl', 'S4Vectors',
              'scales',
              'scatterplot3d', 'sp', 'stringi', 'tcltk2', 'testthat', 'XML', 'xtable',
              'zlibbioc', 'boot', 'class', 'cluster', 'codetools', 'foreign',
              'KernSmooth',
              'lattice', 'MASS', 'Matrix', 'mgcv', 'nlme', 'nnet', 'rpart', 'spatial',
              'survival'
Update all/some/none? [a/s/n]: a
```



# W7-5: 途中経過3

特にエラーを吐くことなく順調にインストールが進んでいるようだ。この間は、基本的に画面が流れているかどうかには注意を払っていただければよい。もし止まっているようだったら、「何か聞かれているかエラーかも」という視点でメッセージを見る

```
File Edit View Search Terminal Help
* DONE (prettyR)
* installing *source* package 'qvalue' ...
** R
** data
** inst
** preparing package for lazy loading
** help
*** installing help indices
** building package indices
** installing vignettes
** testing if installed package can be loaded
* DONE (qvalue)
* installing *source* package 'Rcpp' ...
** package 'Rcpp' successfully unpacked and MD5 sums checked
** libs
g++ -I/usr/share/R/include -DNDEBUG -I../inst/include/ -fpic
-g -O2 -fstack-protector --param=ssp-buffer-size=4 -Wformat -Werror=format-security -D_FORTIFY_SOURCE=2 -g -c Date.cpp -o Date.o
g++ -I/usr/share/R/include -DNDEBUG -I../inst/include/ -fpic
-g -O2 -fstack-protector --param=ssp-buffer-size=4 -Wformat -Werror=format-security -D_FORTIFY_SOURCE=2 -g -c Module.cpp -o Module.o
0
```

# W7-5: 終了後の状態

①コマンド入力待ち状態になれば基本的にOK。パッと見、エラーメッセージが出ていないようだ

```
File Edit View Search Terminal Help
installing to /usr/lib/R/library/mgcv/libs
** R
** data
** inst
** byte-compile and prepare package for lazy loading
** help
*** installing help indices
** building package indices
** testing if installed package can be loaded
* DONE (mgcv)

The downloaded source packages are in
      '/tmp/RtmpS1T00B/downloaded_packages'
Updating HTML index of packages in '.Library'
Making 'packages.html' ... done
Warning messages:
1: In install.packages(update[instlib == 1, "Package"], 1, contrib
url = contriburl, :
  installation of package 'RCurl' had non-zero exit status
2: In install.packages(update[instlib == 1, "Package"], 1, contrib
url = contriburl, :
  installation of package 'XML' had non-zero exit status
>
```



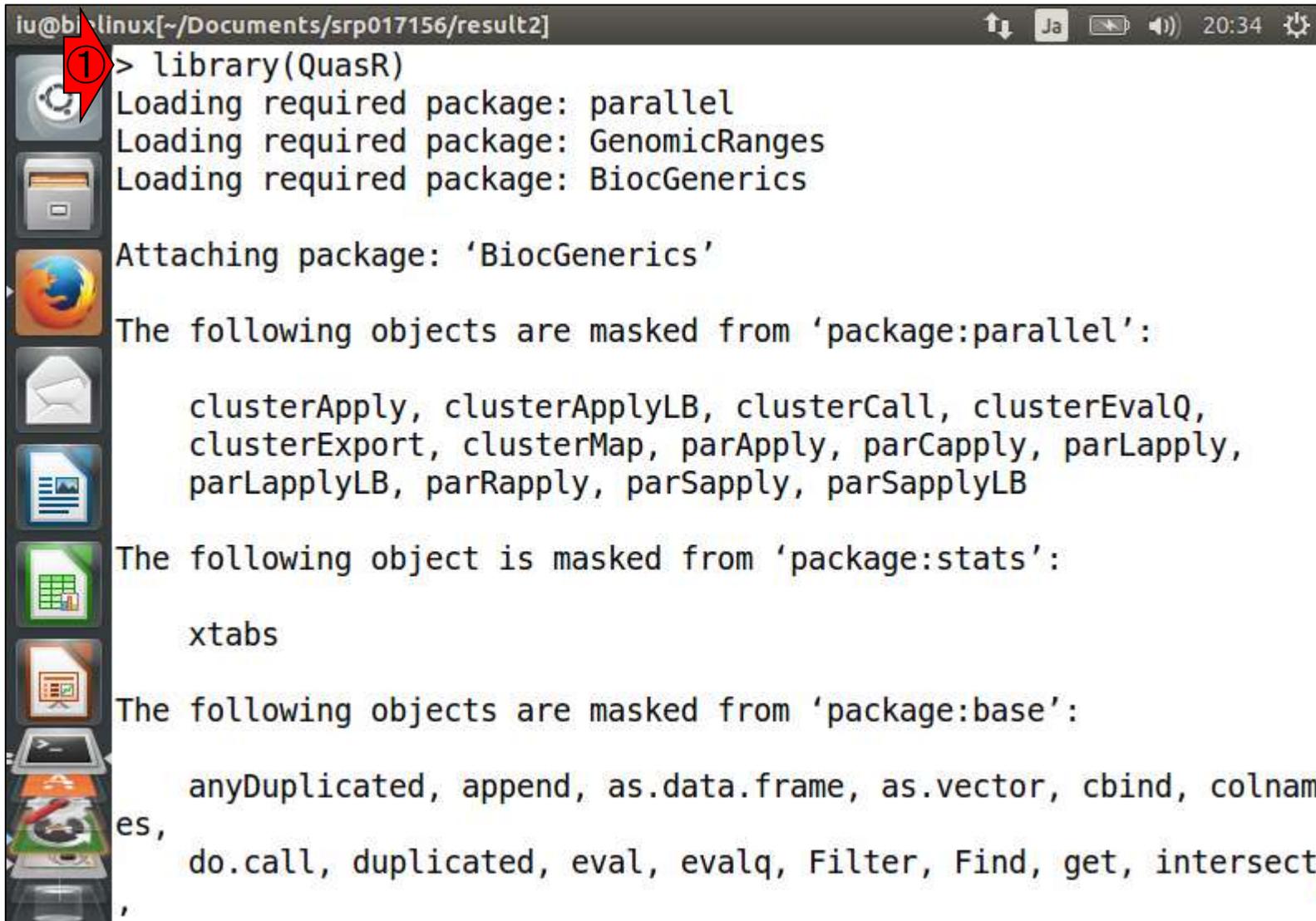
# W7-6: インストール確認

```
File Edit View Search Terminal Help 20:32
installing to /usr/lib/R/library/mgcv/libs
** R
** data
** inst
** byte-compile and prepare package for lazy loading
** help
*** installing help indices
** building package indices
** testing if installed package can be loaded
* DONE (mgcv)

The downloaded source packages are in
      '/tmp/RtmpS1T00B/downloaded_packages'
Updating HTML index of packages in '.Library'
Making 'packages.html' ... done
Warning messages:
1: In install.packages(update[instlib == l, "Package"], l, contrib
url = contriburl, :
  installation of package 'RCurl' had non-zero exit status
2: In install.packages(update[instlib == l, "Package"], l, contrib
url = contriburl, :
  installation of package 'XML' had non-zero exit status
> library(QuasR)
```



# W7-6: インストール確認



```
iu@biolinux[~/Documents/srp017156/result2]
> library(QuasR)
Loading required package: parallel
Loading required package: GenomicRanges
Loading required package: BiocGenerics

Attaching package: 'BiocGenerics'

The following objects are masked from 'package:parallel':

  clusterApply, clusterApplyLB, clusterCall, clusterEvalQ,
  clusterExport, clusterMap, parApply, parCapply, parLapply,
  parLapplyLB, parRapply, parSapply, parSapplyLB

The following object is masked from 'package:stats':

  xtabs

The following objects are masked from 'package:base':

  anyDuplicated, append, as.data.frame, as.vector, cbind, colnames,
  do.call, duplicated, eval, evalq, Filter, Find, get, intersect
```

# W7-6: インストール確認

リターンキーを押した最後のほうの画面。特にエラーメッセージが出ていないことがわかる。

```
File Edit View Search Terminal Help 20:36
xtabs
The following objects are masked from 'package:base':
  anyDuplicated, append, as.data.frame, as.vector, cbind, colnames,
  do.call, duplicated, eval, evalq, Filter, Find, get, intersect,
  is.unsorted, lapply, Map, mapply, match, mget, order, paste, pmax,
  pmax.int, pmin, pmin.int, Position, rank, rbind, Reduce, rep.int,
  rownames, sapply, setdiff, sort, table, tapply, union, unique,
  unlist, unsplit
Loading required package: S4Vectors
Loading required package: stats4
Creating a generic function for 'nchar' from package 'base' in package 'S4Vectors'
Loading required package: IRanges
Loading required package: GenomeInfoDb
Loading required package: Rbowtie
>
```



# W7-6: インストール確認

画面がばーっと流れてエラーの確認がしづらいときは、もう一度同じコマンドを実行するとよい。このとき、一般的なLinuxのTipsと同様に、キーボードの上矢印キーを押すと直前に打ったコマンドが表示される。有効利用すべし

```
File Edit View Search Terminal Help
xtabs
The following objects are masked from 'package:base':
  anyDuplicated, append, as.data.frame, as.vector, cbind, colnames,
  do.call, duplicated, eval, evalq, Filter, Find, get, intersect,
  is.unsorted, lapply, Map, mapply, match, mget, order, paste, pmax,
  pmax.int, pmin, pmin.int, Position, rank, rbind, Reduce, rep.int,
  rownames, sapply, setdiff, sort, table, tapply, union, unique,
  unlist, unsplit
Loading required package: S4Vectors
Loading required package: stats4
Creating a generic function for 'nchar' from package 'base' in package 'S4Vectors'
Loading required package: IRanges
Loading required package: GenomeInfoDb
Loading required package: Rbowtie
> library(QuasR)
```



# W7-6: インストール確認

2回目は、特に何も表示されない。このような場合は、QuasRパッケージのロードに成功していることを意味する。何らかのエラーに遭遇していれば、その旨表示がなされる。②一旦Rを終了

```
File Edit View Search Terminal Help
anyDuplicated, append, as.data.frame, as.vector, cbind, cchain
es,
do.call, duplicated, eval, evalq, Filter, Find, get, intersect
,
is.unsorted, lapply, Map, mapply, match, mget, order, paste, p
max,
pmax.int, pmin, pmin.int, Position, rank, rbind, Reduce, rep.i
nt,
rownames, sapply, setdiff, sort, table, tapply, union, unique,
unlist, unsplit

Loading required package: S4Vectors
Loading required package: stats4
Creating a generic function for 'nchar' from package 'base' in pac
kage 'S4Vectors'
Loading required package: IRanges
Loading required package: GenomeInfoDb
Loading required package: Rbowtie
① > library(QuasR)
> q()
② Save workspace image? [y/n/c]: n
iu@bielinux[result2] [ 8:48午後 ]
```

# W7-7: QuasRウェブページ

BioconductorのQuasRウェブページ。QuasRパッケージのインストール手順と実際に行ったこととの対応関係がよくわかるでしょう。

Bioconductor  
OPEN SOURCE SOFTWARE FOR BIOINFORMATICS

Home Install Help Developers About

Search:

Home » Bioconductor 3.1 » Software Packages » QuasR

## QuasR

platforms: some downloads: top 5% posts: 3 / 3 / 3 / 0 in BioC: 2.5 years  
build: ok commits: 4.00 test coverage: 31%

### Quantify and Annotate Short Reads in R

Bioconductor version: Release (3.1)

This package provides a framework for the quantification and analysis of Short Reads. It covers a complete workflow starting from raw sequence reads, over creation of alignments and quality control plots, to the quantification of genomic regions of interest.

Author: Anita Lerch, Dimos Gaidatzis and Michael Stadler  
Maintainer: Michael Stadler <michael.stadler@fmh.ch>

Citation (from within R, enter `citation("QuasR")`):

Gaidatzis D, Lerch A, Hahne F and Stadler MB (2015). "QuasR: Quantification and annotation of short reads in R." *Bioinformatics*, **31**(7), pp. 1130–1132. <http://dx.doi.org/10.1093/bioinformatics/btu761>, PMID:25417205.

Langmead B, Trapnell C, Pop M and Salzberg SL (2009). "Ultrafast and memory-efficient alignment of short DNA sequences to the human genome." *Genome Biology*, **10**(3), pp. R25. <http://dx.doi.org/10.1186/gb-2009-10-3-r25>, PMID:19261174.

Au KF, Jiang H, Lin L, Xing Y and Wong WH (2010). "Detection of splice junctions from paired-end RNA-seq data by SpliceMap." *Nucleic Acids Research*, **38**(14), pp. 4570–4578. <http://dx.doi.org/10.1093/nar/gkq211>, PMID:20371516.

#### Installation

To install this package, start R and enter:

```
## try http if https is not available
source("https://bioconductor.org/biocLite.R")
biocLite("QuasR")
```

#### Documentation

To view documentation for the version of this package installed in your system, start R and enter:

```
browseVignettes("QuasR")
```

To install this package, start R and enter:

```
## try http if https is not available
source("https://bioconductor.org/biocLite.R")
biocLite("QuasR")
```

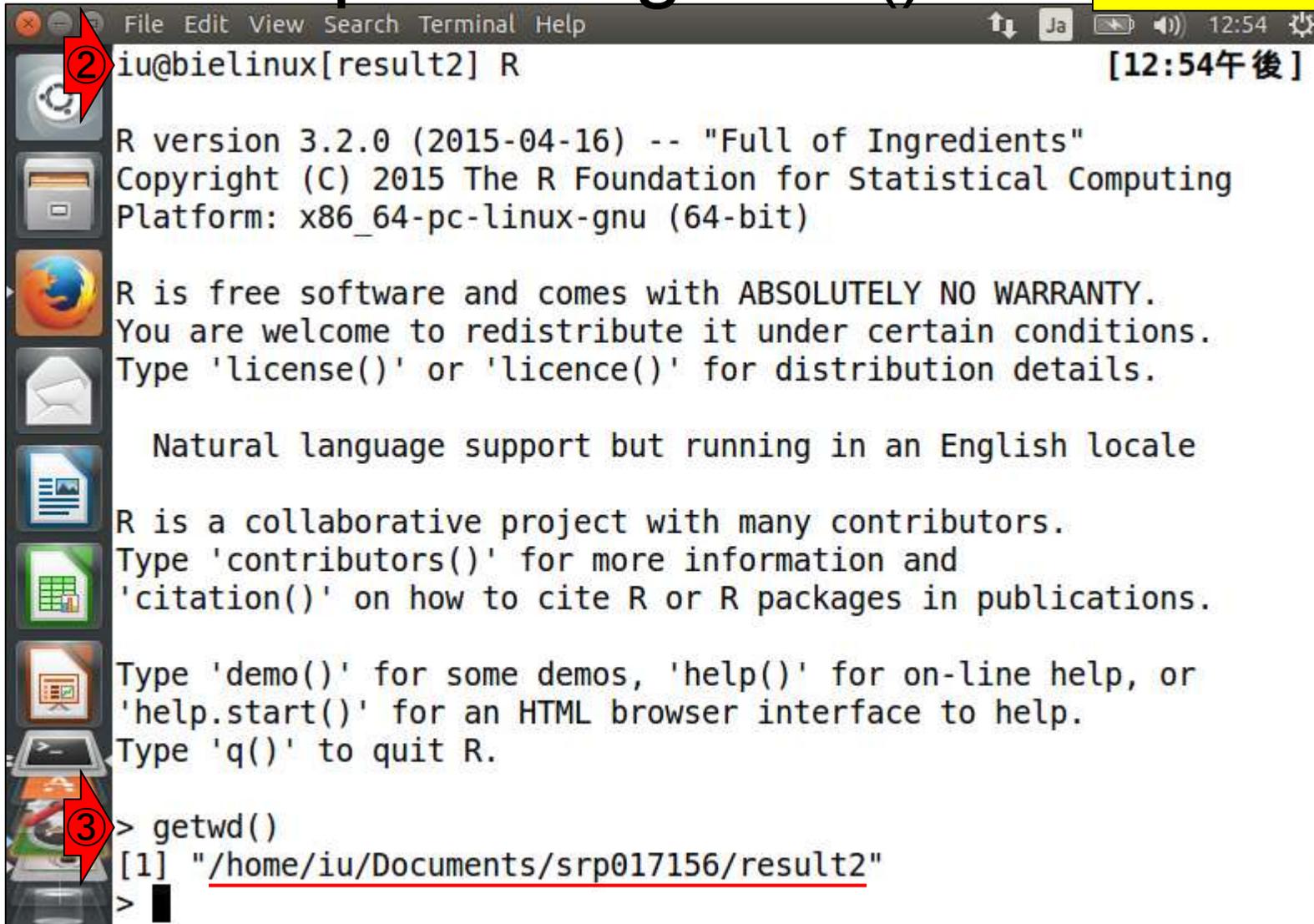
# W8-1 : pwd ⇔ getwd()

①現在の作業ディレクトリは赤下線で示したところ。②R起動後に、③Linuxのpwdコマンドに対応するR上での作業ディレクトリ表示コマンドは「getwd()」

```
iu@bielinux[~/Documents/srp017156/result2]
① iu@bielinux[result2] pwd [ 8:52午後 ]
   /home/iu/Documents/srp017156/result2
② iu@bielinux[result2] R [ 8:52午後 ]
R version 3.2.0 (2015-04-16) -- "Full of Ingredients"
Copyright (C) 2015 The R Foundation for Statistical Computing
Platform: x86_64-pc-linux-gnu (64-bit)
R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.
Natural language support but running in an English locale
R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.
Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.
③ > getwd()█
```

# W8-1 : pwd ⇔ getwd()

ホストOS (WinやMac)上のR GUI版との違いは、起動時の作業ディレクトリが、Rを起動した場所になる点。



A terminal window showing the R startup process. The prompt is `iu@bielinux[result2] R`. The output includes the R version (3.2.0), copyright information, and a list of help topics. A red arrow labeled '2' points to the prompt. At the bottom, the command `> getwd()` is entered, and the output is `[1] "/home/iu/Documents/srp017156/result2"`. A red arrow labeled '3' points to the command.

```
File Edit View Search Terminal Help 12:54 [12:54午後]
iu@bielinux[result2] R
R version 3.2.0 (2015-04-16) -- "Full of Ingredients"
Copyright (C) 2015 The R Foundation for Statistical Computing
Platform: x86_64-pc-linux-gnu (64-bit)

R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.

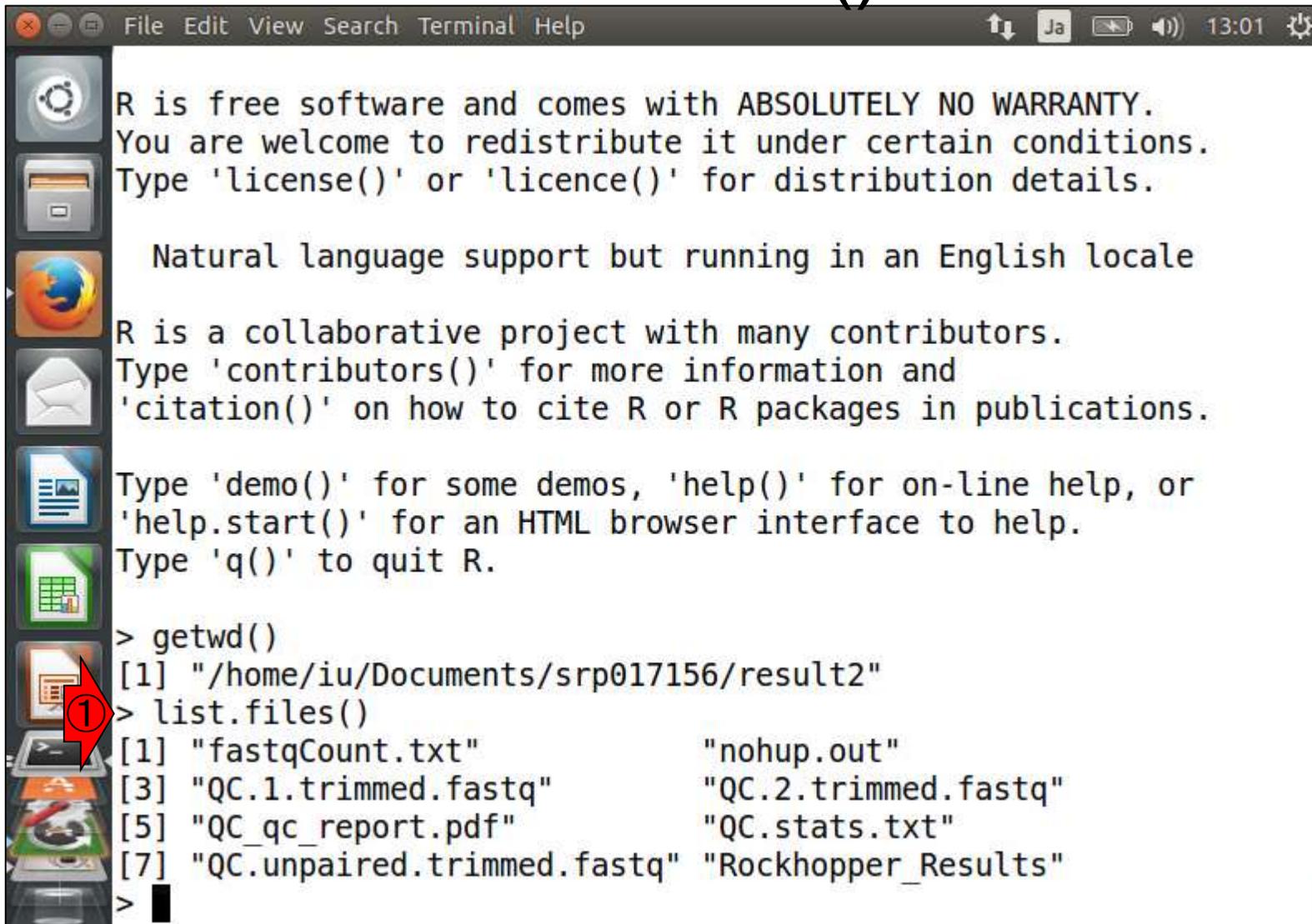
Natural language support but running in an English locale

R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.

Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.

> getwd()
[1] "/home/iu/Documents/srp017156/result2"
>
```

# W8-2: ls ⇔ list.files()



```
File Edit View Search Terminal Help 13:01
R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.

Natural language support but running in an English locale

R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.

Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.

> getwd()
[1] "/home/iu/Documents/srp017156/result2"
> list.files()
[1] "fastqCount.txt"          "nohup.out"
[3] "QC.1.trimmed.fastq"     "QC.2.trimmed.fastq"
[5] "QC_qc_report.pdf"      "QC.stats.txt"
[7] "QC.unpaired.trimmed.fastq" "Rockhopper_Results"
>
```



# W8-3: cd ⇔ setwd()

```
File Edit View Search Terminal Help 13:06
R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.

Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.

> getwd()
[1] "/home/iu/Documents/srp017156/result2"
> list.files()
[1] "fastqCount.txt"          "nohup.out"
[3] "QC.1.trimmed.fastq"     "QC.2.trimmed.fastq"
[5] "QC_qc_report.pdf"       "QC.stats.txt"
[7] "QC.unpaired.trimmed.fastq" "Rockhopper_Results"
① > setwd("/home/iu/Documents/srp017156")
> getwd()
[1] "/home/iu/Documents/srp017156"
> list.files()
[1] "result2"                "SRR616268sub_1.fastq.gz"
[3] "SRR616268sub_2.fastq.gz"
>
```

# W8-4: Rは閉じた世界

①作業ディレクトリ変更後にRを終了。  
Linuxの世界に戻ったのち②pwd。R起動前のディレクトリと同じ。このことから、Rの中は閉じた世界であることがわかる。

```
File Edit View Search Terminal Help
> getwd()
[1] "/home/iu/Documents/srp017156/result2"
> list.files()
[1] "fastqCount.txt"          "nohup.out"
[3] "QC.1.trimmed.fastq"     "QC.2.trimmed.fastq"
[5] "QC_qc_report.pdf"       "QC.stats.txt"
[7] "QC.unpaired.trimmed.fastq" "Rockhopper_Results"
① > setwd("/home/iu/Documents/srp017156")
> getwd()
[1] "/home/iu/Documents/srp017156"
> list.files()
[1] "result2"                 "SRR616268sub_1.fastq.gz"
[3] "SRR616268sub_2.fastq.gz"
> q()
Save workspace image? [y/n/c]: n
② iu@bielinux[result2] pwd [ 1:07午後 ]
/home/iu/Documents/srp017156/result2
iu@bielinux[result2] ls [ 1:07午後 ]
fastqCount.txt      QC.2.trimmed.fastq  QC.unpaired.trimmed.fastq
nohup.out           QC_qc_report.pdf   Rockhopper_Results
QC.1.trimmed.fastq  QC.stats.txt
iu@bielinux[result2] [ 1:08午後 ]
```

# W9-1: ゲノム配列取得

①作業ディレクトリは「~/Documents/genomes」。  
②wget実行時にqオプションをつけているので途中経過が表示されなくてスッキリ。  
③ls

```
iu@bielinux[result2] cd ~/Documents [ 4:28午後 ]
iu@bielinux[Documents] mkdir genomes [ 4:28午後 ]
iu@bielinux[Documents] cd genomes [ 4:28午後 ]
① iu@bielinux[genomes] pwd [ 4:28午後 ]
/home/iu/Documents/genomes
② iu@bielinux[genomes] wget -cq ftp://ftp.ensemblgenomes.org/pub/ba
cteria/release-22/fasta/bacteria_15_collection/lactobacillus_case
i_12a/dna/Lactobacillus_casei_12a.GCA_000309565.1.22.dna.toplevel
.fa.gz
③ iu@bielinux[genomes] ls -l [ 4:28午後 ]
total 884
-rw-rw-r-- 1 iu iu 902847 12月 21 16:28 Lactobacillus_casei_12a.G
CA_000309565.1.22.dna.toplevel.fa.gz
iu@bielinux[genomes] █ [ 4:28午後 ]
```

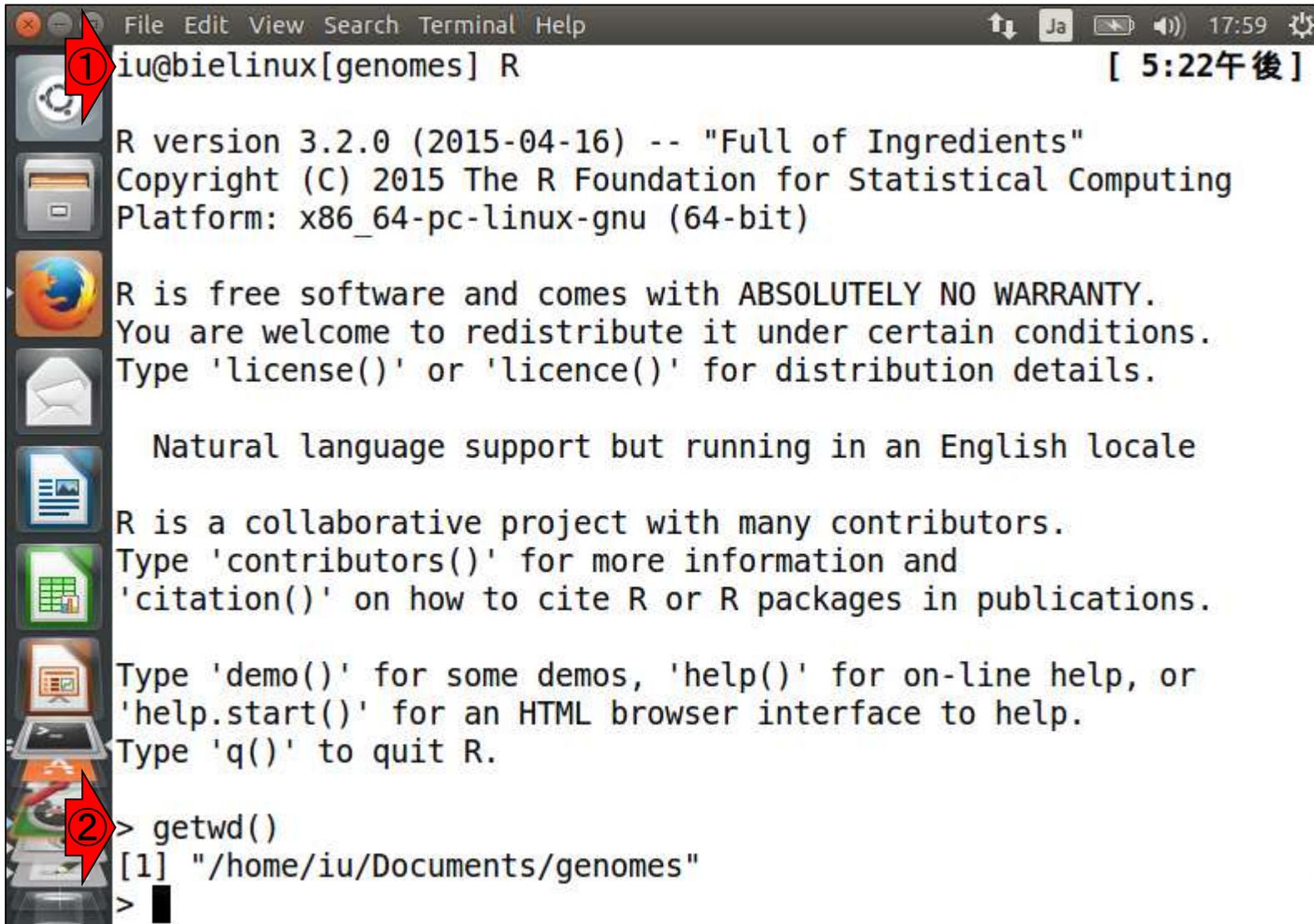
# W9-1: ゲノム配列取得

①gunzipでgzファイルを解凍。解凍後のファイルサイズは2,935,945 bytes (約2.8MB)。

```
iu@bielinux[result2] cd ~/Documents [ 4:28午後 ]
iu@bielinux[Documents] mkdir genomes [ 4:28午後 ]
iu@bielinux[Documents] cd genomes [ 4:28午後 ]
iu@bielinux[genomes] pwd [ 4:28午後 ]
/home/iu/Documents/genomes
iu@bielinux[genomes] wget -cq ftp://ftp.ensemblgenomes.org/pub/ba
cteria/release-22/fasta/bacteria_15_collection/lactobacillus_case
i_12a/dna/Lactobacillus_casei_12a.GCA_000309565.1.22.dna.toplevel
.fa.gz
iu@bielinux[genomes] ls -l [ 4:28午後 ]
total 884
-rw-rw-r-- 1 iu iu 902847 12月 21 16:28 Lactobacillus_casei_12a.G
CA_000309565.1.22.dna.toplevel.fa.gz
① iu@bielinux[genomes] gunzip Lactobacillus_casei_12a.GCA_000309565
.1.22.dna.toplevel.fa.gz
iu@bielinux[genomes] ls -l [ 4:30午後 ]
total 2868
-rw-rw-r-- 1 iu iu 2935947 12月 21 16:28 Lactobacillus_casei_12a.
GCA_000309565.1.22.dna.toplevel.fa
iu@bielinux[genomes] ls -lh [ 4:30午後 ]
total 2.9M
-rw-rw-r-- 1 iu iu 2.8M 12月 21 16:28 Lactobacillus_casei_12a.GCA
_000309565.1.22.dna.toplevel.fa
```

「~/Documents/genomes」上で、① Rを起動。②getwd()はただの確認。

# W9-2: Rを起動



The image shows a terminal window with a dark background and a light-colored font. The window title is "iu@bielinux[genomes] R" and the time is "17:59". The terminal output shows the R startup sequence, including the version number (3.2.0), copyright information, and a list of help topics. A red arrow with the number "1" points to the first prompt character ">". The second prompt character ">" is followed by the command "getwd()", and the output is "[1] \"/home/iu/Documents/genomes\"". A second red arrow with the number "2" points to the "getwd()" command.

```
iu@bielinux[genomes] R [ 5:22午後 ]
R version 3.2.0 (2015-04-16) -- "Full of Ingredients"
Copyright (C) 2015 The R Foundation for Statistical Computing
Platform: x86_64-pc-linux-gnu (64-bit)

R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.

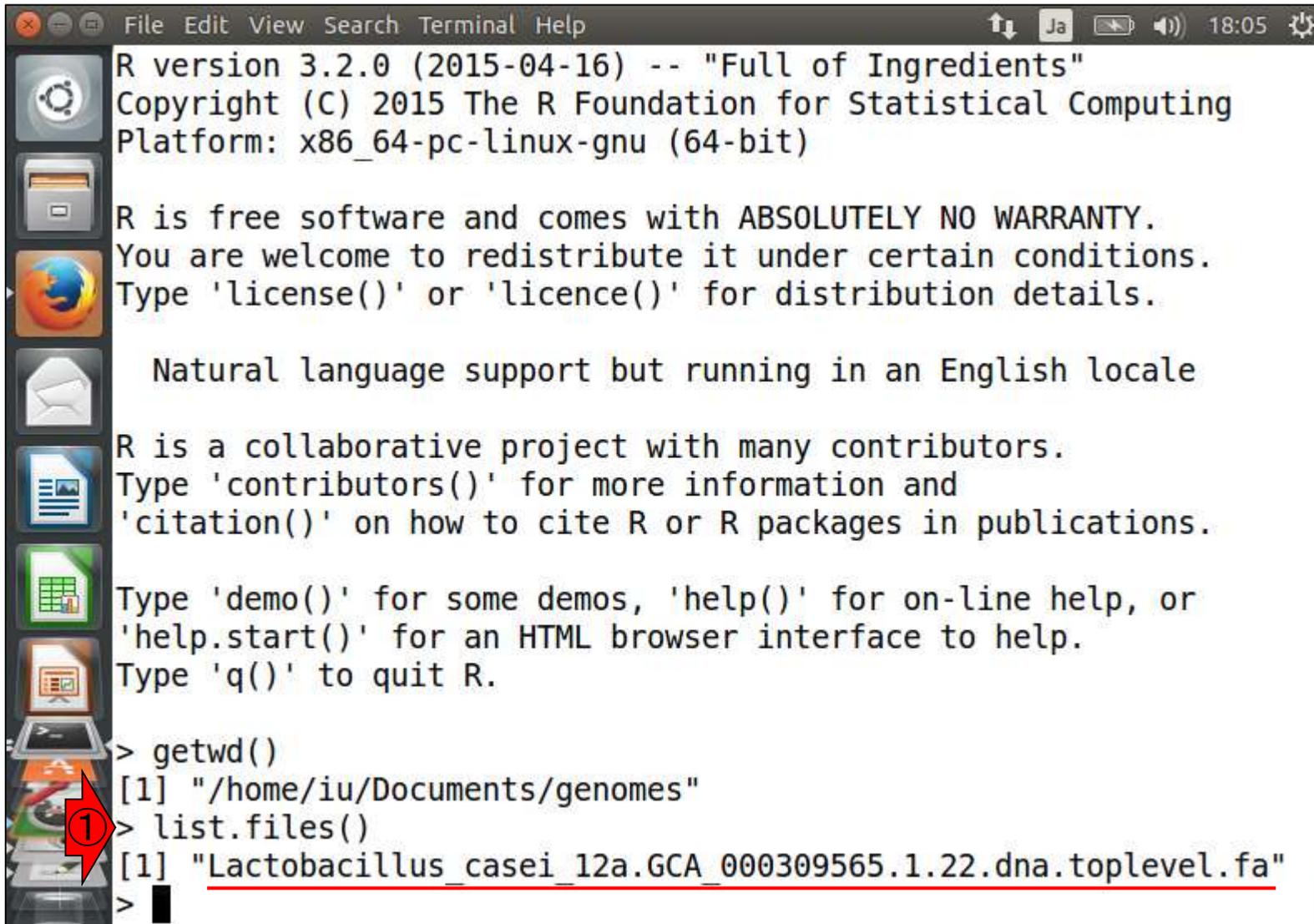
Natural language support but running in an English locale

R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.

Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.

> getwd()
[1] "/home/iu/Documents/genomes"
>
```

# W9-3: 入力ファイルの確認



The screenshot shows an R terminal window with the following content:

```
R version 3.2.0 (2015-04-16) -- "Full of Ingredients"
Copyright (C) 2015 The R Foundation for Statistical Computing
Platform: x86_64-pc-linux-gnu (64-bit)

R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.

Natural language support but running in an English locale

R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.

Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.

> getwd()
[1] "/home/iu/Documents/genomes"
> list.files()
[1] "Lactobacillus_casei_12a.GCA_000309565.1.22.dna.toplevel.fa"
>
```

A red arrow with the number 1 points to the `list.files()` command and its output.

# W9-4: コピペ用コード

入出力の関係。①入力ファイル名、②出力ファイル名。③ここでは、入力ファイル名をin\_fで、出力ファイル名をout\_fとして取り扱っている

- 連載第1回の「Rでゲノム解析」と同じコード。

このウェブページ中の「イントロ | NGS | 読み込み | FASTA形式 | [基本情報を取得](#)」の5と基本的に同じ。

以下の内容は [JSLABS 1.R](#) の中身と同じです。

```
in_f <- "Lactobacillus_casei_12a.GCA_000309565.1.22.dna.toplevel.fasta" #①入力ファイル名を指
out_f <- "result_JSLAB1.txt" #出力ファイル名を指定してout_fに格納

#③Rのパッケージをロード
library(Biostrings) #②パッケージの読み込み

#入力ファイルの読み込み
fasta <- readDNAStringSet(in_f, format="fasta") #in_fで指定したファイルの読み込み

#本番(基本情報取得)
Total_len <- sum(width(fasta)) #コンティグの「トータルの長さ」を取得
Number_of_contigs <- length(fasta) #「コンティグ数」を取得
Average_len <- mean(width(fasta)) #コンティグの「平均長」を取得
Median_len <- median(width(fasta)) #コンティグの「中央値」を取得
Max_len <- max(width(fasta)) #コンティグの長さの「最大値」を取得
Min_len <- min(width(fasta)) #コンティグの長さの「最小値」を取得

#本番(N50情報取得)
sorted <- rev(sort(width(fasta))) #長さ情報を降順にソートした結果をsortedに格納
obj <- (cumsum(sorted) >= Total_len*0.5) #条件を満たすかどうかを判定した結果をobjに格納(長
N50 <- sorted[obj][1] #objがTRUEとなる1番最初の要素のみ抽出した結果をN50
```

黒枠内(つまりコード全部)を全選択して①コピー。

# W9-5: コピペ

- 連載第1回の「Rでゲノム解析」と同じコード。  
このウェブページ中の「イントロ | NGS | 読み込み | FASTA形式 | [基本情報を取得](#)」の5と基本的に同じ。  
以下の内容は [JSLABS 1.R](#) の中身と同じです。

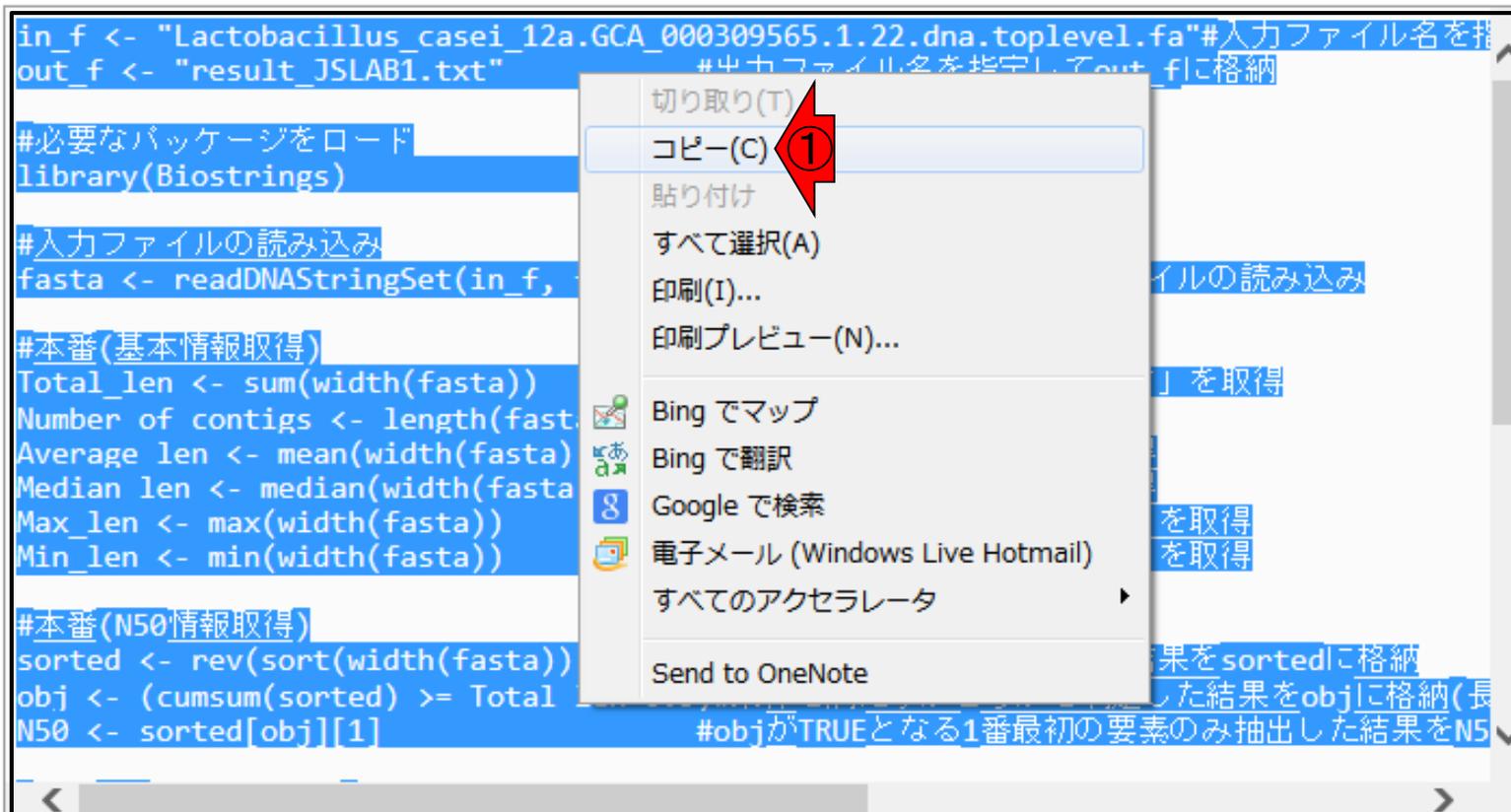
```
in_f <- "Lactobacillus_casei_12a.GCA_000309565.1.22.dna.toplevel.fa" #入力ファイル名を指
out_f <- "result_JSLAB1.txt" #出力ファイル名を指定してout_fに格納

#必要なパッケージをロード
library(Biostrings)

#入力ファイルの読み込み
fasta <- readDNASTringSet(in_f,

#本番(基本情報取得)
Total_len <- sum(width(fasta))
Number of contigs <- length(fasta)
Average len <- mean(width(fasta))
Median len <- median(width(fasta))
Max_len <- max(width(fasta))
Min_len <- min(width(fasta))

#本番(N50情報取得)
sorted <- rev(sort(width(fasta)))
obj <- (cumsum(sorted) >= Total
N50 <- sorted[obj][1]
```



# W9-5: コピペ

Rの画面上で①ペースト。ホスト - ゲスト間でコピペがうまくできないときは、②のfirefoxを起動してコードのコピーをすればよい。

```
iu@bielinux[~/Documents/genomes]
R version 3.2.0 (2015-04-16) -- "Full of Ingredients"
Copyright (C) 2015 The R Foundation for Statistical Computing
Platform: x86_64-pc-linux-gnu (64-bit)

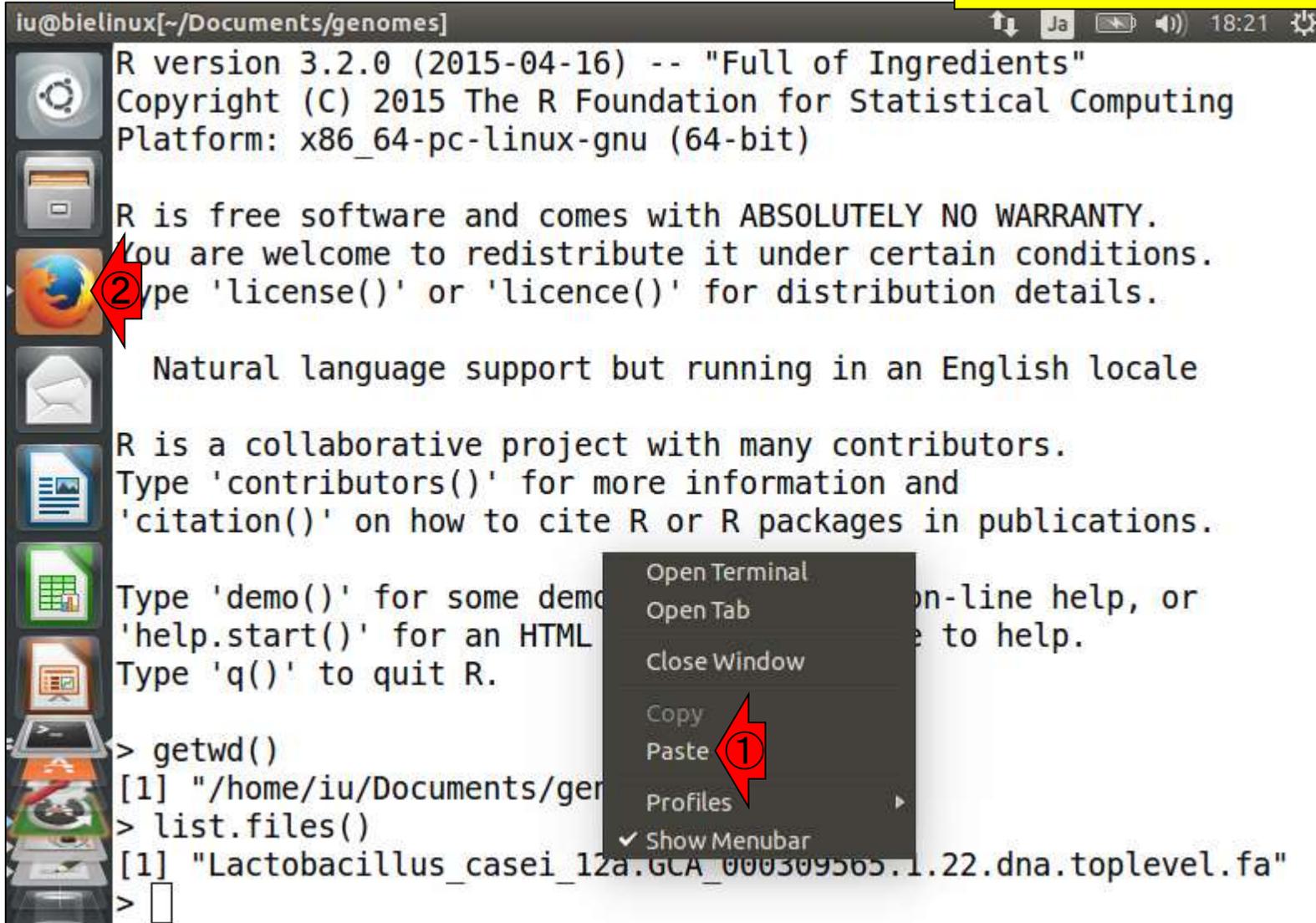
R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.

Natural language support but running in an English locale

R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.

Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.

> getwd()
[1] "/home/iu/Documents/genomes"
> list.files()
[1] "Lactobacillus_casei_12a.GCA_000309565.1.22.dna.toplevel.fa"
>
```



# W9-6: コピペ後

エラーなく実行できたときの全貌(っていうか最後のほう)。①最後の行は、write.tableという関数を用いてtmpの中身をout\_fで指定したファイルに保存するコマンド。

```
> #本番 (GC含量情報取得)
> hoge <- alphabetFrequency(fasta)           #A,C,G,T,..の数を配列ごと
にカウントした結果をhogeに格納
> CG <- rowSums(hoge[,2:3])                  #C,Gの総数を計算してCGに
格納
> ACGT <- rowSums(hoge[,1:4])                #A,C,G,Tの総数を計算してA
CGTに格納
> GC_content <- sum(CG)/sum(ACGT)            #トータルGC含量の情報を
取得
>
> #ファイルに保存
> tmp <- NULL
> tmp <- rbind(tmp, c("Total length (bp)", Total_len))
> tmp <- rbind(tmp, c("Number of contigs", Number_of_contigs))
> tmp <- rbind(tmp, c("Average length", Average_len))
> tmp <- rbind(tmp, c("Median length", Median_len))
> tmp <- rbind(tmp, c("Max length", Max_len))
> tmp <- rbind(tmp, c("Min length", Min_len))
> tmp <- rbind(tmp, c("N50", N50))
> tmp <- rbind(tmp, c("GC content", GC_content))
> write.table(tmp, out_f, sep="\t", append=F, quote=F, row.names=F
, col.names=F)#tmpの中身を指定したファイル名で保存
> █
```



# W9-6: コピペ後

それゆえ、out\_fで指定した出力ファイル(result\_JALAB1.txt)をイチイチ開いて確認しなくても、①tmpと打って、tmpの中身をR画面上で確認するのもよい。

```
File Edit View Search Terminal Help
> #本番 (GC含量情報取得)
> hoge <- alphabetFrequency(fasta)           #A,C,G,T,..の数を配列ごと
にカウントした結果をhogeに格納
> CG <- rowSums(hoge[,2:3])                  #C,Gの総数を計算してCGに
格納
> ACGT <- rowSums(hoge[,1:4])               #A,C,G,Tの総数を計算してA
CGTに格納
> GC_content <- sum(CG)/sum(ACGT)           #トータルGC含量の情報を
取得
>
> #ファイルに保存
> tmp <- NULL
> tmp <- rbind(tmp, c("Total length (bp)", Total_len))
> tmp <- rbind(tmp, c("Number of contigs", Number_of_contigs))
> tmp <- rbind(tmp, c("Average length", Average_len))
> tmp <- rbind(tmp, c("Median length", Median_len))
> tmp <- rbind(tmp, c("Max length", Max_len))
> tmp <- rbind(tmp, c("Min length", Min_len))
> tmp <- rbind(tmp, c("N50", N50))
> tmp <- rbind(tmp, c("GC content", GC_content))
> write.table(tmp, out_f, sep="\t", append=F, quote=F, row.names=F
, col.names=F)#tmpの中身を指定したファイル名で保存
> tmp
```



# W9-6: コピペ後

①tmpの中身を表示させた結果。これが「連載第1回の図2」と基本的に同じもの

```
File Edit View Search Terminal Help 19:40
> #ファイルに保存
> tmp <- NULL
> tmp <- rbind(tmp, c("Total length (bp)", Total_len))
> tmp <- rbind(tmp, c("Number of contigs", Number_of_contigs))
> tmp <- rbind(tmp, c("Average length", Average_len))
> tmp <- rbind(tmp, c("Median length", Median_len))
> tmp <- rbind(tmp, c("Max length", Max_len))
> tmp <- rbind(tmp, c("Min length", Min_len))
> tmp <- rbind(tmp, c("N50", N50))
> tmp <- rbind(tmp, c("GC content", GC_content))
> write.table(tmp, out_f, sep="\t", append=F, quote=F, row.names=F,
, col.names=F)#tmpの中身を指定したファイル名で保存
> tmp
      [,1]      [,2]
[1,] "Total length (bp)" "2885619"
[2,] "Number of contigs" "28"
[3,] "Average length"    "103057.821428571"
[4,] "Median length"     "58047"
[5,] "Max length"        "472701"
[6,] "Min length"        "899"
[7,] "N50"               "222389"
[8,] "GC content"        "0.464045193416998"
>
```



①確かにW9-3のコピペ前には存在しなかったresult\_JSLAB1.txtがある

# W9-7: 存在確認

```
File Edit View Search Terminal Help 19:50
> tmp <- rbind(tmp, c("Number of contigs", Number_of_contigs))
> tmp <- rbind(tmp, c("Average length", Average_len))
> tmp <- rbind(tmp, c("Median length", Median_len))
> tmp <- rbind(tmp, c("Max length", Max_len))
> tmp <- rbind(tmp, c("Min length", Min_len))
> tmp <- rbind(tmp, c("N50", N50))
> tmp <- rbind(tmp, c("GC content", GC_content))
> write.table(tmp, out_f, sep="\t", append=F, quote=F, row.names=F,
, col.names=F)#tmpの中身を指定したファイル名で保存
> tmp
      [,1]      [,2]
[1,] "Total length (bp)" "2885619"
[2,] "Number of contigs" "28"
[3,] "Average length"    "103057.821428571"
[4,] "Median length"     "58047"
[5,] "Max length"        "472701"
[6,] "Min length"        "899"
[7,] "N50"               "222389"
[8,] "GC content"        "0.464045193416998"
> list.files()
[1] "Lactobacillus_casei_12a.GCA_000309565.1.22.dna.toplevel.fa"
[2] "result_JSLAB1.txt"
>
```



①Rを終了させて、②lsで存在確認しているだけです

# W9-7: 存在確認

```
File Edit View Search Terminal Help 19:53
> tmp <- rbind(tmp, c("N50", N50))
> tmp <- rbind(tmp, c("GC content", GC_content))
> write.table(tmp, out_f, sep="\t", append=F, quote=F, row.names=F,
, col.names=F)#tmpの中身を指定したファイル名で保存
> tmp
      [,1]          [,2]
[1,] "Total length (bp)" "2885619"
[2,] "Number of contigs" "28"
[3,] "Average length"   "103057.821428571"
[4,] "Median length"    "58047"
[5,] "Max length"       "472701"
[6,] "Min length"       "899"
[7,] "N50"              "222389"
[8,] "GC content"       "0.464045193416998"
> list.files()
[1] "Lactobacillus_casei_12a.GCA_000309565.1.22.dna.toplevel.fa"
[2] "result_JSLAB1.txt"
> q()
Save workspace image? [y/n/c]: n
iu@bielinux[genomes] ls [ 7:53午後 ]
Lactobacillus_casei_12a.GCA_000309565.1.22.dna.toplevel.fa
result_JSLAB1.txt
iu@bielinux[genomes] [ 7:53午後 ]
```



# W9-8: R起動時のTips

①R起動時に-qオプションをつけることで、スタートアップメッセージを非表示にすることができる。

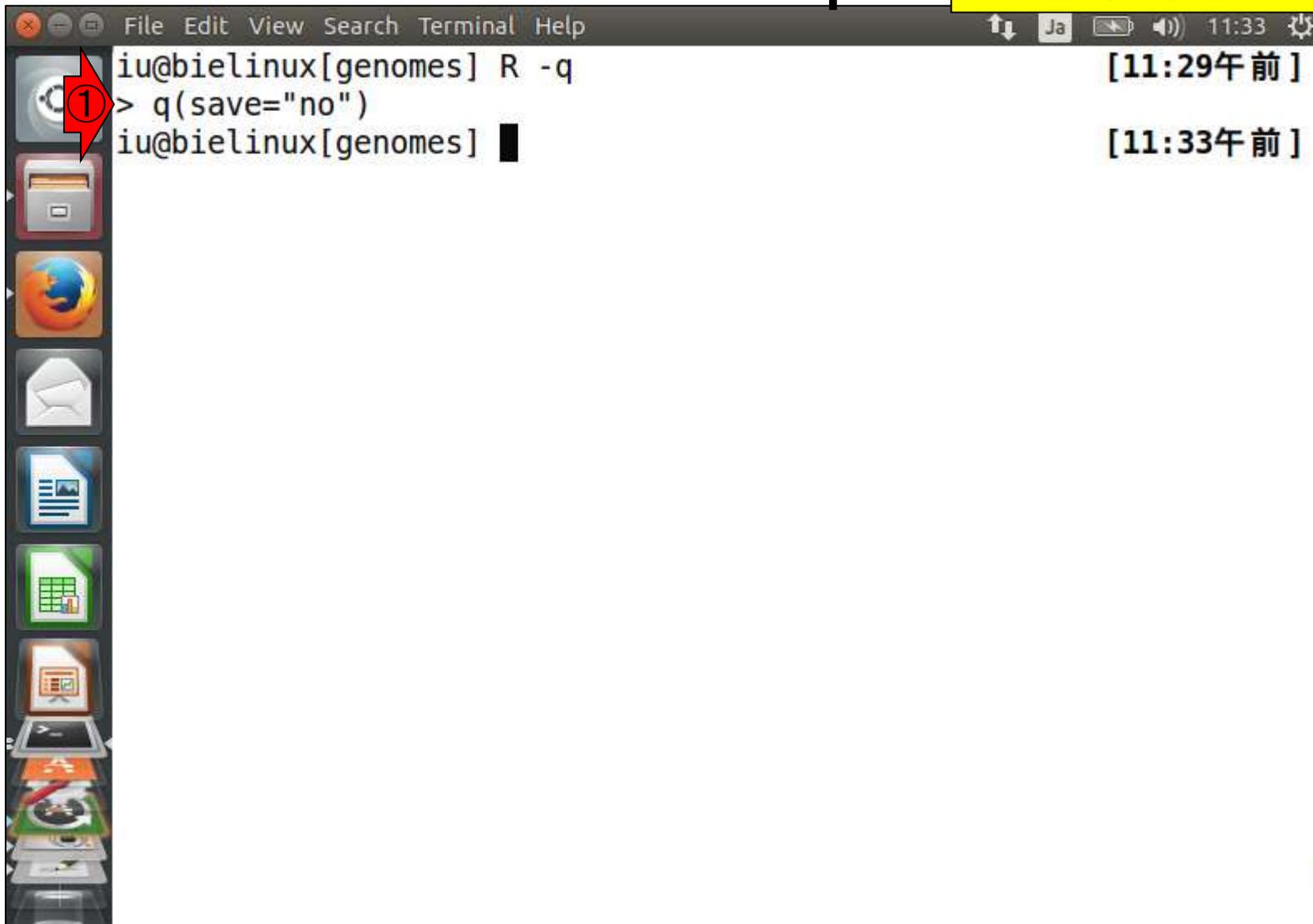


A terminal window titled "Terminal" with a menu bar containing "File", "Edit", "View", "Search", "Terminal", and "Help". The window title bar also shows "iu@bielinux[genomes] R -q" and the time "11:30". The terminal content shows the command "R -q" being entered, with a red arrow pointing to the "1" in the command. The terminal output is currently blank, with a cursor on the line below the command. The window has a sidebar on the left with various application icons.

```
iu@bielinux[genomes] R -q  
> █
```

# W9-8: R終了時のTips

①R終了時に「save="no"」オプションをつけることで、「Save workspace image? [y/n/c]」と毎回聞かれる苦行から解放される



The screenshot shows a terminal window with the following content:

```
iu@bielinux[genomes] R -q [11:29午前]
> q(save="no")
iu@bielinux[genomes] [11:33午前]
```

A red arrow with the number '1' points to the command `q(save="no")`.

①一旦result\_JSLAB1.txtを削除。②  
wgetでJSLAB5\_1.Rファイルを取得。

# W10-1 : source関数

```
iu@bielinux[genomes] pwd [ 5:31午後 ]
/home/iu/Documents/genomes
iu@bielinux[genomes] ls [ 5:32午後 ]
Lactobacillus_casei_12a.GCA_000309565.1.22.dna.toplevel.fa
result_JSLAB1.txt
① iu@bielinux[genomes] rm -f result_JSLAB1.txt [ 5:32午後 ]
iu@bielinux[genomes] ls [ 5:32午後 ]
Lactobacillus_casei_12a.GCA_000309565.1.22.dna.toplevel.fa
② iu@bielinux[genomes] wget -cq http://www.iu.a.u-tokyo.ac.jp/~kado
ta/book/JSLAB5_1.R
iu@bielinux[genomes] ls [ 5:32午後 ]
JSLAB5_1.R
Lactobacillus_casei_12a.GCA_000309565.1.22.dna.toplevel.fa
iu@bielinux[genomes] [ 5:32午後 ]
```

# W10-1 : source関数

①headで最初の5行分を表示。文字化けしているが、結果に影響しないコメント部分なので、ここでは気にしない

```
iu@bielinux[genomes] pwd [ 5:31午後 ]
/home/iu/Documents/genomes
iu@bielinux[genomes] ls [ 5:32午後 ]
Lactobacillus_casei_12a.GCA_000309565.1.22.dna.toplevel.fa
result_JSLAB1.txt
iu@bielinux[genomes] rm -f result_JSLAB1.txt [ 5:32午後 ]
iu@bielinux[genomes] ls [ 5:32午後 ]
Lactobacillus_casei_12a.GCA_000309565.1.22.dna.toplevel.fa
iu@bielinux[genomes] wget -cq http://www.iu.a.u-tokyo.ac.jp/~kado
ta/book/JSLAB5_1.R
iu@bielinux[genomes] ls [ 5:32午後 ]
JSLAB5_1.R
Lactobacillus_casei_12a.GCA_000309565.1.22.dna.toplevel.fa
iu@bielinux[genomes] head -n 5 JSLAB5_1.R [ 5:32午後 ]
in_f <- "Lactobacillus_casei_12a.GCA_000309565.1.22.dna.toplevel.
fa"#000t0@0C00000000所00in_f0Qi0[
out_f <- "result_JSLAB1.txt" #0t0@0C00000000所00out_f0Qi
0[
#0K0v0âp0b0P0[0W00000[0h
library(Biostrings) #0p0b0P0[0W0#v00
iu@bielinux[genomes] [ 5:33午後 ]
```

①

①スタートアップメッセージ  
を非表示にしてRを起動。

# W10-1 : source関数

```
iu@bielinux[genomes] pwd [ 5:31午後 ]
/home/iu/Documents/genomes
iu@bielinux[genomes] ls [ 5:32午後 ]
Lactobacillus_casei_12a.GCA_000309565.1.22.dna.toplevel.fa
result_JSLAB1.txt
iu@bielinux[genomes] rm -f result_JSLAB1.txt [ 5:32午後 ]
iu@bielinux[genomes] ls [ 5:32午後 ]
Lactobacillus_casei_12a.GCA_000309565.1.22.dna.toplevel.fa
iu@bielinux[genomes] wget -cq http://www.iu.a.u-tokyo.ac.jp/~kado
ta/book/JSLAB5_1.R
iu@bielinux[genomes] ls [ 5:32午後 ]
JSLAB5_1.R
Lactobacillus_casei_12a.GCA_000309565.1.22.dna.toplevel.fa
iu@bielinux[genomes] head -n 5 JSLAB5_1.R [ 5:32午後 ]
in_f <- "Lactobacillus_casei_12a.GCA_000309565.1.22.dna.toplevel.
fa"#000t0@0C00000000所00in_f0Qi0[
out_f <- "result_JSLAB1.txt" #0t0@0C00000000所00out_f0Qi
0[
#0K0v0âp0b0P0[0W00000[0h
library(Biostrings) #0p0b0P0[0W0#v00
iu@bielinux[genomes] R -q [ 5:33午後 ]
>
```



# W10-1 : source関数

目的は「source(“JSLAB5\_1.R”)」と打ち込むこと。Rの対話モードでもタブ補完が有効なので、①のような状態でタブキーを押すと…



A terminal window titled "Terminal" with a menu bar (File, Edit, View, Search, Terminal, Help) and a system tray (Ja, 17:36, settings). The prompt is "iu@bielinux[genomes] R -q". The command "source("JS" is entered, with a red circle containing the number "1" and an arrow pointing to the cursor. The terminal also shows "[ 5:35午後]" in the top right corner. On the left side of the terminal, there is a vertical dock with various application icons.

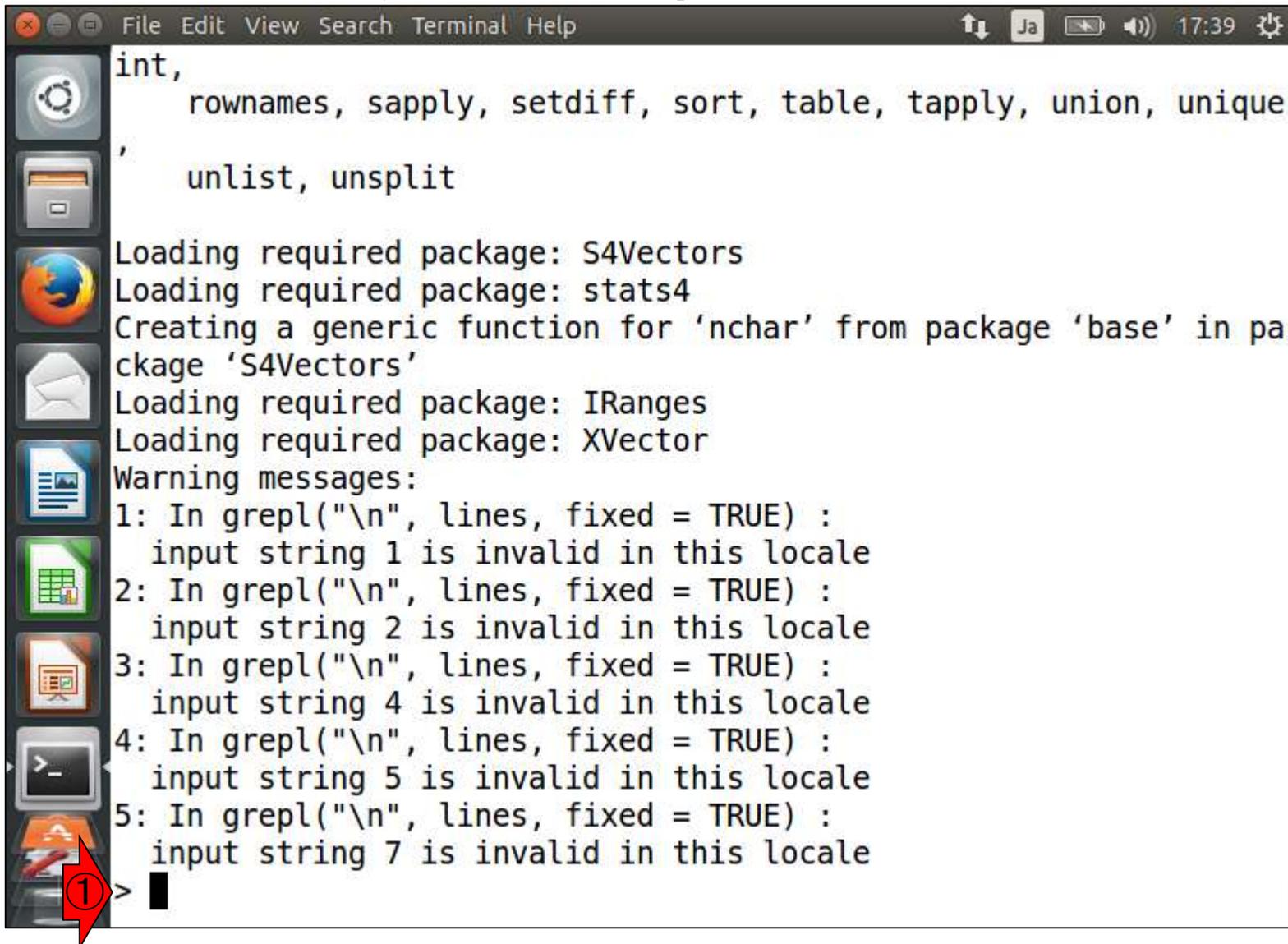
```
iu@bielinux[genomes] R -q  
> source("JS
```

# W10-1 : source関数



```
File Edit View Search Terminal Help
iu@bielinux[genomes] R -q [ 5:35午後 ]
> source("JSLAB5_1.R")
```

# W10-2: 実行結果



```
File Edit View Search Terminal Help 17:39
int,
  rownames, sapply, setdiff, sort, table, tapply, union, unique
,
  unlist, unsplit

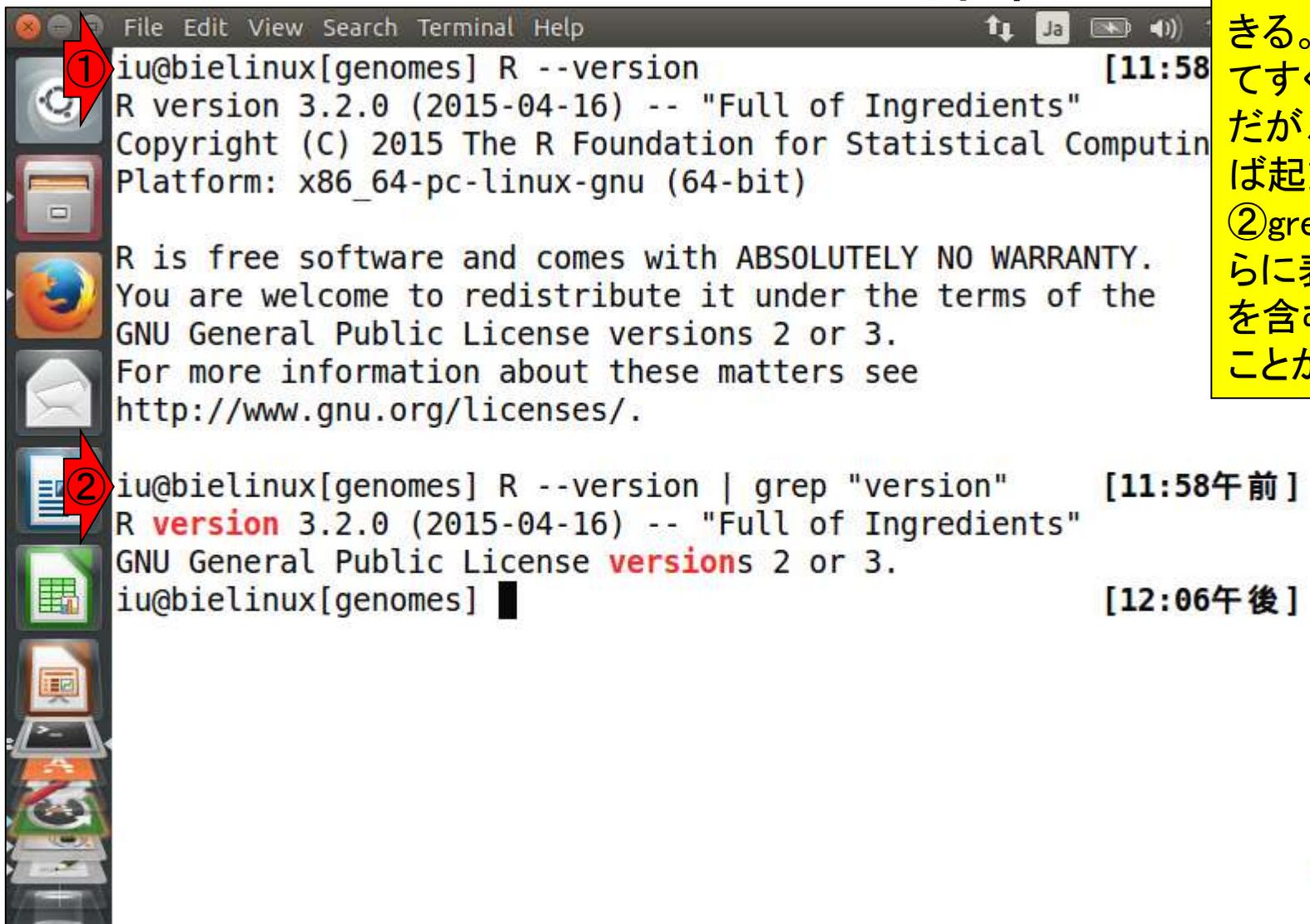
Loading required package: S4Vectors
Loading required package: stats4
Creating a generic function for 'nchar' from package 'base' in pa
ckage 'S4Vectors'
Loading required package: IRanges
Loading required package: XVector
Warning messages:
1: In grepl("\n", lines, fixed = TRUE) :
  input string 1 is invalid in this locale
2: In grepl("\n", lines, fixed = TRUE) :
  input string 2 is invalid in this locale
3: In grepl("\n", lines, fixed = TRUE) :
  input string 4 is invalid in this locale
4: In grepl("\n", lines, fixed = TRUE) :
  input string 5 is invalid in this locale
5: In grepl("\n", lines, fixed = TRUE) :
  input string 7 is invalid in this locale
>
```

# W10-3: 確認

①list.files()で確認。確かに出力ファイルとして指定したresult\_JSLAB1.txtが生成されている。②Rを終了。③lsで念のため確認。④moreでファイルの中身を表示。妥当な結果である

```
File Edit View Search Terminal Help
4: In grepl("\n", lines, fixed = TRUE) :
  input string 5 is invalid in this locale
5: In grepl("\n", lines, fixed = TRUE) :
  input string 7 is invalid in this locale
> list.files()
[1] "JSLAB5_1.R"
[2] "Lactobacillus_casei_12a.GCA_000309565.1.22.dna.toplevel.fa"
[3] "result_JSLAB1.txt"
> q(save="no")
iu@bielinux[genomes] ls [ 5:44午後 ]
JSLAB5_1.R
Lactobacillus_casei_12a.GCA_000309565.1.22.dna.toplevel.fa
result_JSLAB1.txt
iu@bielinux[genomes] more result_JSLAB1.txt [ 5:44午後 ]
Total length (bp)      2885619
Number of contigs      28
Average length        103057.821428571
Median length         58047
Max length            472701
Min length            899
N50                   222389
GC content             0.464045193416998
iu@bielinux[genomes] [ 5:45午後 ]
```

# W10-4: Rのバージョン確認



The image shows a terminal window with a sidebar on the left containing icons for various applications. Two red arrows point to specific terminal commands. The terminal output shows the R version and license information.

```
iu@bielinux[genomes] R --version [11:58]
R version 3.2.0 (2015-04-16) -- "Full of Ingredients"
Copyright (C) 2015 The R Foundation for Statistical Computing
Platform: x86_64-pc-linux-gnu (64-bit)

R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under the terms of the
GNU General Public License versions 2 or 3.
For more information about these matters see
http://www.gnu.org/licenses/.

iu@bielinux[genomes] R --version | grep "version" [11:58午前]
R version 3.2.0 (2015-04-16) -- "Full of Ingredients"
GNU General Public License versions 2 or 3.
iu@bielinux[genomes] [12:06午後]
```

①「R --version」と打つことで、Rを起動することなくバージョン確認をすることができます。正確には一旦起動しただけで終了している模様だが、エンドユーザからすれば起動していないのと同じ。

②grepを併用することで、さらに表示結果から”version”を含む行のみに限定させることができる。

# W10-5: バッチモード

①一旦result\_JSLAB1.txtを削除。②lsで出力予定ファイル(result\_JSLAB1.txt)がないことを確認して、③バッチモードの基本形を実行

```
iu@bielinux[genomes] pwd [ 5:51午後 ]
/home/iu/Documents/genomes
iu@bielinux[genomes] ls [ 5:51午後 ]
JSLAB5_1.R
Lactobacillus_casei_12a.GCA_000309565.1.22.dna.toplevel.fa
result_JSLAB1.txt
iu@bielinux[genomes] rm -f result_JSLAB1.txt ① [ 5:51午後 ]
iu@bielinux[genomes] ls [ 5:51午後 ]
JSLAB5_1.R
Lactobacillus_casei_12a.GCA_000309565.1.22.dna.toplevel.fa
iu@bielinux[genomes] R --vanilla < JSLAB5_1.R [ 5:51午後 ]
```

# W10-5: バッチモード

①リターンを押して実行した結果。確かにq()などでRを終了させる手間なく、通常のLinuxコマンド入力待ち状態になっていることがわかる。

```
File Edit View Search Terminal Help 17:52
> #0{00(GC0mY00000)
> hoge <- alphabetFrequency(fasta) #A,C,G,T,..00000z0[BJ0E
000g0000000000hoge0Qi0[
> CG <- apply(as.matrix(hoge[,2:3]), 1, sum)#C,G00000000v0000CG0Qi
0[
> ACGT <- apply(as.matrix(hoge[,1:4]), 1, sum)#A,C,G,T00000000v0000
0ACGT0Qi0[
> GC_content <- sum(CG)/sum(ACGT) #0g0[0^0000GC0m$00000000
>
> #0t0@0C0000c_0
> tmp <- NULL
> tmp <- rbind(tmp, c("Total length (bp)", Total_len))
> tmp <- rbind(tmp, c("Number of contigs", Number_of_contigs))
> tmp <- rbind(tmp, c("Average length", Average_len))
> tmp <- rbind(tmp, c("Median length", Median_len))
> tmp <- rbind(tmp, c("Max length", Max_len))
> tmp <- rbind(tmp, c("Min length", Min_len))
> tmp <- rbind(tmp, c("N50", N50))
> tmp <- rbind(tmp, c("GC content", GC_content))
> write.table(tmp, out_f, sep="\t", append=F, quote=F, row.names=
F, col.names=F)#tmp000g0000 000t0@0C00000r_0
>
iu@bielinux[genomes] [ 5:52午後]
```



# W10-5: 確認

①lsで確認。確かにresult\_JSLAB1.txtが出力結果として得られており、②その中身も妥当。

```
File Edit View Search Terminal Help 17:54
> tmp <- rbind(tmp, c("Average length", Average_len))
> tmp <- rbind(tmp, c("Median length", Median_len))
> tmp <- rbind(tmp, c("Max length", Max_len))
> tmp <- rbind(tmp, c("Min length", Min_len))
> tmp <- rbind(tmp, c("N50", N50))
> tmp <- rbind(tmp, c("GC content", GC_content))
> write.table(tmp, out_f, sep="\t", append=F, quote=F, row.names=
F, col.names=F)#tmp000g000所000t000C00000r0
>
① iu@bielinux[genomes] ls [ 5:52午後 ]
JSLAB5_1.R
Lactobacillus_casei_12a.GCA_000309565.1.22.dna.toplevel.fa
result_JSLAB1.txt
② iu@bielinux[genomes] more result_JSLAB1.txt [ 5:53午後 ]
Total length (bp)          2885619
Number of contigs          28
Average length  103057.821428571
Median length   58047
Max length      472701
Min length      899
N50             222389
GC content      0.464045193416998
iu@bielinux[genomes] █ [ 5:53午後 ]
```





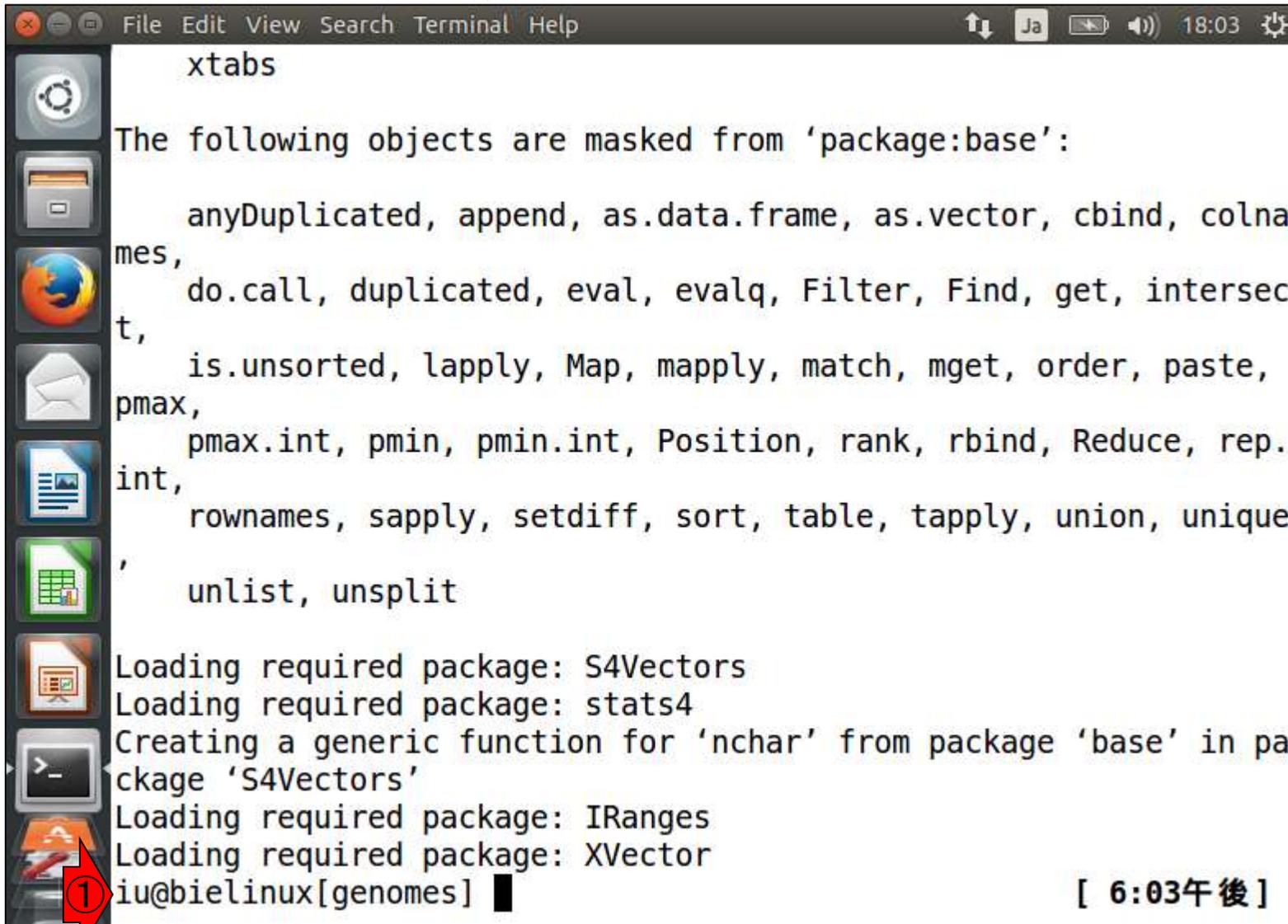
# W12-1: 発展形1

①--slaveオプションをつけて実行させると、実行中に画面表示されるものが減るので、多少見づらさが緩和される

```
iu@bielinux[genomes] pwd [ 6:01午後 ]
/home/iu/Documents/genomes
iu@bielinux[genomes] ls [ 6:01午後 ]
hoge.R
JSLAB5_1.R
Lactobacillus_casei_12a.GCA_000309565.1.22.dna.toplevel.fa
result_JSLAB1.txt
iu@bielinux[genomes] rm -f hoge.R result_JSLAB1.txt [ 6:01午後 ]
iu@bielinux[genomes] ls [ 6:01午後 ]
JSLAB5_1.R
Lactobacillus_casei_12a.GCA_000309565.1.22.dna.toplevel.fa
iu@bielinux[genomes] R --vanilla --slave < JSLAB5_1.R
```



# W12-1: 発展形1



```
xtabs
The following objects are masked from 'package:base':
  anyDuplicated, append, as.data.frame, as.vector, cbind, colnames,
  do.call, duplicated, eval, evalq, Filter, Find, get, intersect,
  is.unsorted, lapply, Map, mapply, match, mget, order, paste,
  pmax,
  pmax.int, pmin, pmin.int, Position, rank, rbind, Reduce, rep.int,
  rownames, sapply, setdiff, sort, table, tapply, union, unique
'
  unlist, unsplit

Loading required package: S4Vectors
Loading required package: stats4
Creating a generic function for 'nchar' from package 'base' in package 'S4Vectors'
Loading required package: IRanges
Loading required package: XVector
iu@bielinux[genomes] █ [ 6:03午後]
```

# W12-1: 発展形1

```
anyDuplicated, append, as.data.frame, as.vector, cbind, colna
mes,
do.call, duplicated, eval, evalq, Filter, Find, get, intersec
t,
is.unsorted, lapply, Map, mapply, match, mget, order, paste,
pmax,
pmax.int, pmin, pmin.int, Position, rank, rbind, Reduce, rep.
int,
rownames, sapply, setdiff, sort, table, tapply, union, unique
,
unlist, unsplit

Loading required package: S4Vectors
Loading required package: stats4
Creating a generic function for 'nchar' from package 'base' in pa
ckage 'S4Vectors'
Loading required package: IRanges
Loading required package: XVector
iu@bielinux[genomes] ls [ 6:03午後 ]
JSLAB5_1.R
Lactobacillus_casei_12a.GCA_000309565.1.22.dna.toplevel.fa
result JSLAB1.txt
iu@bielinux[genomes] █ [ 6:04午後 ]
```

# W12-2: 発展形2

①JSLAB5\_1.Rの最初の2行分を表示。  
nkf実行結果とパイプさせているのは、  
文字化け対策。JSLAB5\_1.Rが正常動作  
するのは、②作業ディレクトリ上に、in\_f  
で指定した入力ファイルが存在するから

```
iu@bielinux[genomes] rm -f result_JSLAB1.txt
iu@bielinux[genomes] ls
JSLAB5_1.R
Lactobacillus_casei_12a.GCA_000309565.1.22.dna.toplevel.fa
iu@bielinux[genomes] nkf JSLAB5_1.R | head -n2 [ 4:09午後 ]
in_f <- "Lactobacillus_casei_12a.GCA_000309565.1.22.dna.toplevel.f
a"#入力ファイル名を指定してin_fに格納
out_f <- "result_JSLAB1.txt" #出力ファイル名を指定してou
t_fに格納
iu@bielinux[genomes] pwd [ 4:09午後 ]
/home/iu/Documents/genomes
iu@bielinux[genomes] [ 4:09午後 ]
```



# W12-2: 発展形2

①result2ディレクトリにJSLAB5\_1.Rをコピー。  
②移動先にはJSLAB5\_1.Rが入力として読み込む乳酸菌ゲノムファイルは存在しない。③result2上でJSLAB5\_1.Rを実行してみると…。

```
iu@bielinux[genomes] rm -f result_JSLAB1.txt [ 4:09午後]
iu@bielinux[genomes] ls [ 4:09午後]
JSLAB5_1.R
Lactobacillus_casei_12a.GCA_000309565.1.22.dna.toplevel.fa
iu@bielinux[genomes] nkf JSLAB5_1.R | head -n2 [ 4:09午後]
in_f <- "Lactobacillus_casei_12a.GCA_000309565.1.22.dna.toplevel.f
a"#入力ファイル名を指定してin_fに格納
out_f <- "result_JSLAB1.txt" #出力ファイル名を指定してou
t_fに格納
iu@bielinux[genomes] pwd [ 4:09午後]
/home/iu/Documents/genomes
① iu@bielinux[genomes] cp JSLAB5_1.R ~/Documents/srp017156/result2
iu@bielinux[genomes] cd ~/Documents/srp017156/result2 [ 4:21午後]
② iu@bielinux[result2] pwd [ 4:21午後]
/home/iu/Documents/srp017156/result2
iu@bielinux[result2] ls [ 4:21午後]
fastqCount.txt QC.1.trimmed.fastq QC.stats.txt
JSLAB5_1.R QC.2.trimmed.fastq QC.unpaired.trimmed.fastq
nohup.out QC_qc_report.pdf Rockhopper_Results
③ iu@bielinux[result2] R --vanilla --slave < JSLAB5_1.R [ 4:21午後]
```

# W12-2: 発展形2

①cannot open file…や②Execution haltedというネガティブなメッセージからも、実行失敗の想像がつく。

```
File Edit View Search Terminal Help 16:39
do.call, duplicated, eval, evalq, Filter, Find, get, intersect
,
is.unsorted, lapply, Map, mapply, match, mget, order, paste, p
max,
pmax.int, pmin, pmin.int, Position, rank, rbind, Reduce, rep.i
nt,
rownames, sapply, setdiff, sort, table, tapply, union, unique,
unlist, unsplit

Loading required package: S4Vectors
Loading required package: stats4
Creating a generic function for 'nchar' from package 'base' in pac
kage 'S4Vectors'
Loading required package: IRanges
Loading required package: XVector
Error in .Call2("new_input_ExternalFilePtr", fp, PACKAGE = "Biostr
ings") :
  cannot open file 'Lactobacillus_casei_12a.GCA_000309565.1.22.dna
.toplevel.fa'
Calls: readDNAStringSet ... .open_input_files -> lapply -> lapply
-> FUN -> .Call2 -> .Call
Execution halted
iu@bielinux[result2] [ 4:32午後 ]
```

①

②

# W12-2: 発展形2

③lsした結果。result\_JSLAB1.txtが生成されていないことがわかる。④当然  
~/Documents/genomes上にもない。理由はシンプル。JSLAB5\_1.Rは、入力ファイルをカレントディレクトリ上でのみ探索しているから

```
File Edit View Search Terminal Help
unlist, unsplit

Loading required package: S4Vectors
Loading required package: stats4
Creating a generic function for 'nchar' from package 'base' in package 'S4Vectors'
Loading required package: IRanges
Loading required package: XVector
Error in .Call2("new_input_ExternalFilePtr", fp, PACKAGE = "Biostrings") :
  cannot open file 'Lactobacillus_casei_12a.GCA_000309565.1.22.dna.toplevel.fa'
Calls: readDNAStringSet ... .open_input_files -> lapply -> lapply
-> FUN -> .Call2 -> .Call
Execution halted

iu@bielinux[result2] ls [ 4:32午後 ]
fastqCount.txt QC.1.trimmed.fastq QC.stats.txt
JSLAB5_1.R QC.2.trimmed.fastq QC.unpaired.trimmed.fastq
nohup.out QC_qc_report.pdf Rockhopper_Results

iu@bielinux[result2] ls ~/Documents/genomes [ 4:48午後 ]
JSLAB5_1.R
Lactobacillus_casei_12a.GCA_000309565.1.22.dna.toplevel.fa
iu@bielinux[result2] [ 4:52午後 ]
```



# W12-3: 発展形3

- ①wgetでJSLAB5\_2.Rをダウンロード。②最初の2行分を表示。赤下線で示すように、入力ファイルを絶対パスで指定している。
- ③JSLAB5\_2.Rをバッチモードで実行。

```
iu@bielinux[result2] pwd [ 7:04午後 ]
/home/iu/Documents/srp017156/result2
iu@bielinux[result2] ls [ 7:05午後 ]
fastqCount.txt QC.1.trimmed.fastq QC.stats.txt
JSLAB5_1.R QC.2.trimmed.fastq QC.unpaired.trimmed.fastq
nohup.out QC_qc_report.pdf Rockhopper_Results
① iu@bielinux[result2] wget -cq http://www.iu.a.u-tokyo.ac.jp/~kadota/book/JSLAB5_2.R
iu@bielinux[result2] ls [ 7:05午後 ]
fastqCount.txt QC.1.trimmed.fastq QC.unpaired.trimmed.fastq
JSLAB5_1.R QC.2.trimmed.fastq Rockhopper_Results
JSLAB5_2.R QC_qc_report.pdf
nohup.out QC.stats.txt
② iu@bielinux[result2] nkf JSLAB5_2.R | head -n 2 [ 7:05午後 ]
in_f <- "/home/iu/Documents/genomes/Lactobacillus_casei_12a.GCA_00309565.1.22.dna.toplevel.fa"#入力ファイル名を指定してin_fに格納
out_f <- "result_JSLAB1.txt" #出力ファイル名を指定してout_fに格納
③ iu@bielinux[result2] R --vanilla --slave < JSLAB5_2.R [ 7:05午後 ]
```

# W12-3: 発展形3

```
File Edit View Search Terminal Help 19:19
xtabs
The following objects are masked from 'package:base':
  anyDuplicated, append, as.data.frame, as.vector, cbind, colnames,
  do.call, duplicated, eval, evalq, Filter, Find, get, intersect,
  is.unsorted, lapply, Map, mapply, match, mget, order, paste, pmax,
  pmax.int, pmin, pmin.int, Position, rank, rbind, Reduce, rep.int,
  rownames, sapply, setdiff, sort, table, tapply, union, unique,
  unlist, unsplit
Loading required package: S4Vectors
Loading required package: stats4
Creating a generic function for 'nchar' from package 'base' in package 'S4Vectors'
Loading required package: IRanges
Loading required package: XVector
iu@bielinux[result2] [ 7:19午後 ]
```

# W12-3: 発展形3

①lsで確認。確かに出力ファイルが存在し、②その中身も正しい。

```
File Edit View Search Terminal Help 19:21
unlist, unsplit
Loading required package: S4Vectors
Loading required package: stats4
Creating a generic function for 'nchar' from package 'base' in package 'S4Vectors'
Loading required package: IRanges
Loading required package: XVector
iu@bielinux[result2] ls [ 7:19午後 ]
fastqCount.txt QC.1.trimmed.fastq QC.unpaired.trimmed.fastq
JSLAB5_1.R QC.2.trimmed.fastq result_JSLAB1.txt
JSLAB5_2.R QC_qc_report.pdf Rockhopper_Results
nohup.out QC.stats.txt
iu@bielinux[result2] more result_JSLAB1.txt [ 7:21午後 ]
Total length (bp) 2885619
Number of contigs 28
Average length 103057.821428571
Median length 58047
Max length 472701
Min length 899
N50 222389
GC content 0.464045193416998
iu@bielinux[result2] [ 7:21午後 ]
```

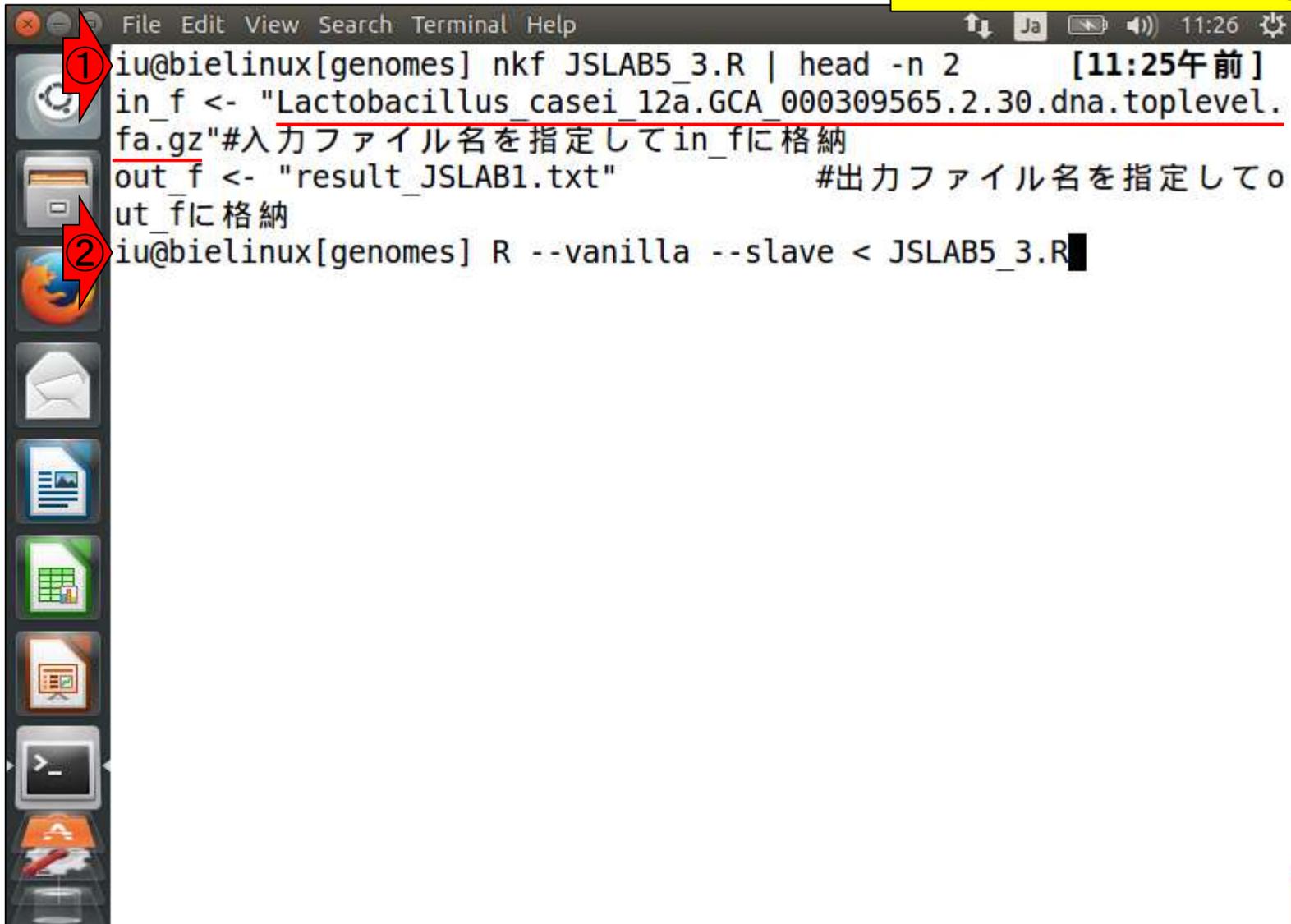
# W13-1: 最新版で解析

①ゲノムファイルと②Rスクリプトファイルのダウンロード。赤下線で示すように、正しく取得できてるっぽいことがわかる

```
iu@bielinux[result2] cd ~/Documents/genomes [11:23午前]
iu@bielinux[genomes] pwd [11:24午前]
/home/iu/Documents/genomes
iu@bielinux[genomes] ls [11:24午前]
JSLAB5_1.R
Lactobacillus_casei_12a.GCA_000309565.1.22.dna.toplevel.fa
① iu@bielinux[genomes] wget -cq ftp://ftp.ensemblgenomes.org/pub/ba
cteria/release-30/fastq/bacteria_15_collection/lactobacillus_case
i_12a/dna/Lactobacillus_casei_12a.GCA_000309565.2.30.dna.toplevel
.fa.gz
② iu@bielinux[genomes] wget -cq http://www.iu.a.u-tokyo.ac.jp/~kado
ta/book/JSLAB5_3.R
iu@bielinux[genomes] ls -l [11:24午前]
total 3764
-rw-rw-r-- 1 iu iu 2042 9月 11 10:35 JSLAB5_1.R
-rw-rw-r-- 1 iu iu 2045 12月 22 11:17 JSLAB5_3.R
-rw-rw-r-- 1 iu iu 2935947 12月 21 16:28 Lactobacillus_casei_12a.
GCA_000309565.1.22.dna.toplevel.fa
-rw-rw-r-- 1 iu iu 906567 12月 22 11:24 Lactobacillus_casei_12a.
GCA_000309565.2.30.dna.toplevel.fa.gz
iu@bielinux[genomes] [11:24午前]
```

# W13-1: 最新版で解析

①Rスクリプトファイルの最初の2行分を表示。  
赤下線で示すようにgzip圧縮ファイルのまま  
取り扱うことができる。②JSLAB5\_3.Rを実行。

A terminal window titled "Terminal" with a menu bar (File, Edit, View, Search, Terminal, Help) and system status (Ja, 11:26). The terminal shows the execution of an R script. The first two lines of the script are highlighted with a red underline. Red arrows with numbers 1 and 2 point to the first and second lines of the script, respectively. The first line is `in_f <- "Lactobacillus_casei_12a.GCA_000309565.2.30.dna.toplevel.fa.gz"` with a red underline. The second line is `out_f <- "result_JSLAB1.txt"`. The third line is `iu@bielinux[genomes] R --vanilla --slave < JSLAB5_3.R`.

```
iu@bielinux[genomes] nkf JSLAB5_3.R | head -n 2 [11:25午前]
in_f <- "Lactobacillus_casei_12a.GCA_000309565.2.30.dna.toplevel.
fa.gz"#入力ファイル名を指定してin_fに格納
out_f <- "result_JSLAB1.txt" #出力ファイル名を指定してo
ut_fに格納
iu@bielinux[genomes] R --vanilla --slave < JSLAB5_3.R
```

# W13-1: 最新版で解析

```
File Edit View Search Terminal Help 11:28
xtabs
The following objects are masked from 'package:base':
  anyDuplicated, append, as.data.frame, as.vector, cbind, colnames,
  do.call, duplicated, eval, evalq, Filter, Find, get, intersect,
  is.unsorted, lapply, Map, mapply, match, mget, order, paste,
  pmax,
  pmax.int, pmin, pmin.int, Position, rank, rbind, Reduce, rep.int,
  rownames, sapply, setdiff, sort, table, tapply, union, unique
  ,
  unlist, unsplit
Loading required package: S4Vectors
Loading required package: stats4
Creating a generic function for 'nchar' from package 'base' in package 'S4Vectors'
Loading required package: IRanges
Loading required package: XVector
iu@bielinux[genomes] [11:27午前]
```

# W13-1: 最新版で解析

①lsで確認。出力ファイルresult\_JSLAB1.txt  
が確かにできている。②moreで中身を表示  
。1 contig、2,907,892 bpであることがわかる

```
File Edit View Search Terminal Help
Loading required package: S4Vectors
Loading required package: stats4
Creating a generic function for 'nchar' from package 'base' in pa
ckage 'S4Vectors'
Loading required package: IRanges
Loading required package: XVector
iu@bielinux[genomes] ls [11:27午前]
JSLAB5_1.R
JSLAB5_3.R
Lactobacillus_casei_12a.GCA_000309565.1.22.dna.toplevel.fa
Lactobacillus_casei_12a.GCA_000309565.2.30.dna.toplevel.fa.gz
result_JSLAB1.txt
iu@bielinux[genomes] more result_JSLAB1.txt [11:28午前]
Total length (bp)      2907892
Number of contigs     1
Average length        2907892
Median length         2907892
Max length            2907892
Min length            2907892
N50                   2907892
GC content             0.464441595492542
iu@bielinux[genomes] [11:28午前]
```



# W13-2: Ensembl

①をクリックした結果。連載第1回当時はコンテンツレベルだったが、第5回執筆時には②染色体レベルになっていることがわかる。③トータル塩基数は2,907,892 bp。

**Lactobacillus casei 12A**  
*Lactobacillus casei* 12A

Provider [European Nucleotide Archive](#) | Taxonomy ID [1051650](#)

Search *Lactobacillus casei* 12A...

e.g. *spaB* or **Chromosome:502057-502915** or *synthetase*

### About *Lactobacillus casei* 12A

**1** Information and statistics

Genome assembly: [GCA\\_000309565.2](#)

More information and statistics

Download DNA sequence (FASTA)

Display your data in Ensembl Bacteria

View karyotyp

Example regio

### Comparative genomics

What can I find? Gene families based on HAMAP and PANTHER classification.

More about comparative analyses

Ensembl Bacteria release 28 - August 2015 © EBI

## Lactobacillus casei 12A Assembly and Gene Annotation

### *Lactobacillus casei* 12A

#### Organism

Taxonomy ID [1051650](#)

Name *Lactobacillus casei* 12A

[Wikipedia](#)

Aliases *Lactobacillus casei* str. 12A  
*Lactobacillus casei* strain 12A

#### Classification

- cellular organisms
- Bacteria
- Firmicutes
- Bacilli
- Lactobacillales
- Lactobacillaceae
- Lactobacillus
- Lactobacillus casei group
- Lactobacillus casei
- Lactobacillus casei 12A

#### European Nucleotide Archive Records

[CP006690.1](#)

#### References

1. Analysis of the *Lactobacillus casei* supragenome and its influence in species evolution and lifestyle adaptation. Broadbent J.R., Neeno-Eckwall E.C., Stahl B., Tandee K., Cai H., Morovic W., Horvath P., Heidenreich J., Perna N.T., Barrangou R., Steele J.L. - *BMC Genomics* 2012, 13:533 PubMed: [23035691](#)

#### Ensembl Genomes API Example

This example Perl script shows how to create a database adaptor for this species. For more information see the [Ensembl Bacteria documentation](#).

#### Statistics

##### Summary

Assembly:	ASM30956v2, INSDC Assembly <a href="#">GCA_000309565.2</a>
Database version:	81.1
Base Pairs:	2,907,892
Golden Path Length:	2,907,892
Data source:	<a href="#">European Nucleotide Archive</a>
Genebuild method:	Generated from ENA annotation

##### Gene counts

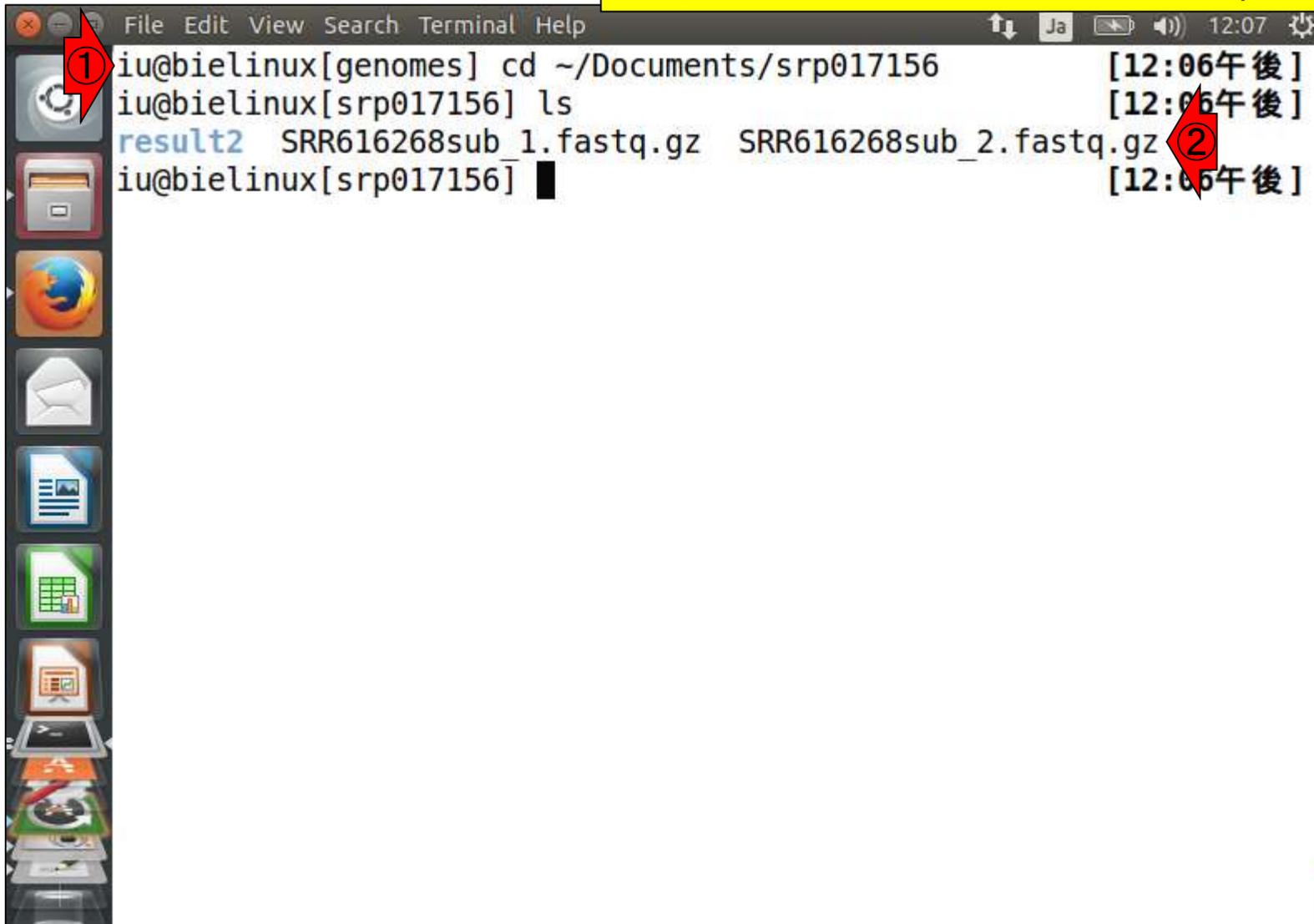
Coding genes:	2,681
Non coding genes:	72
Small non coding genes:	72
Pseudogenes:	46
Gene transcripts:	2,799

##### Coordinate Systems

chromosome	1 sequence
	<input type="text" value="Filter"/>
Sequence	Length (bp)
<a href="#">Chromosome</a>	2907892

# W14-1: 準備

①作業ディレクトリは「~/Documents/srp017156」とする。②ここで  
見えている2つのgzファイルは、連載第3回W25あたりで作成した、  
最初の100万リード(400万行)からなるpaired-end RNA-seqデータ



The image shows a terminal window with a dark background and a light-colored text area. The window title bar includes 'File Edit View Search Terminal Help', a language indicator 'Ja', and the time '12:07'. The terminal content is as follows:

```
iu@bielinux[genomes] cd ~/Documents/srp017156 [12:06午後]
iu@bielinux[srp017156] ls [12:06午後]
result2 SRR616268sub_1.fastq.gz SRR616268sub_2.fastq.gz [12:06午後]
iu@bielinux[srp017156] █ [12:06午後]
```

Two red arrows with white numbers '1' and '2' are overlaid on the terminal. Arrow '1' points to the 'cd' command, and arrow '2' points to the two gz files listed in the 'ls' output.

# W14-1: 準備

①「~/Documents/srp017156/result2」ディレクトリ上にある\*.fastqファイルがFaQCs実行結果なので、これを確認。②このディレクトリ中の\*.fastqを満たすファイル(赤線の3つ)全てをgzip圧縮。数分

```
iu@bielinux[genomes] cd ~/Documents/srp017156 [12:06午後]
iu@bielinux[srp017156] ls [12:06午後]
result2 SRR616268sub_1.fastq.gz SRR616268sub_2.fastq.gz
iu@bielinux[srp017156] ls result2 [12:06午後]
fastqCount.txt QC.1.trimmed.fastq QC.unpaired.trimmed.fastq
JSLAB5_1.R QC.2.trimmed.fastq result_JSLAB1.txt
JSLAB5_2.R QC_qc_report.pdf Rockhopper_Results
nohup.out QC.stats.txt
iu@bielinux[srp017156] gzip result2/*.fastq [12:13午後]
iu@bielinux[srp017156] ls result2 [12:14午後]
fastqCount.txt QC_qc_report.pdf
JSLAB5_1.R QC.stats.txt
JSLAB5_2.R QC.unpaired.trimmed.fastq.gz
nohup.out result_JSLAB1.txt
QC.1.trimmed.fastq.gz Rockhopper_Results
QC.2.trimmed.fastq.gz
iu@bielinux[srp017156] [12:14午後]
```

# W14-1: 準備

①②マッピングに用いたいのは赤下線の2つのファイルのみなので、これらをカレントディレクトリにコピー。cpコマンドの最後のピリオド(.)はコピー先をカレントディレクトリにするという意味

```
iu@bielinux[genomes] cd ~/Documents/srp017156 [12:06午後]
iu@bielinux[srp017156] ls [12:06午後]
result2 SRR616268sub_1.fastq.gz SRR616268sub_2.fastq.gz
iu@bielinux[srp017156] ls result2 [12:06午後]
fastqCount.txt QC.1.trimmed.fastq QC.unpaired.trimmed.fastq
JSLAB5_1.R QC.2.trimmed.fastq result_JSLAB1.txt
JSLAB5_2.R QC_qc_report.pdf Rockhopper_Results
nohup.out QC.stats.txt
iu@bielinux[srp017156] gzip result2/*.fastq [12:13午後]
iu@bielinux[srp017156] ls result2 [12:14午後]
fastqCount.txt QC_qc_report.pdf
JSLAB5_1.R QC.stats.txt
JSLAB5_2.R QC.unpaired.trimmed.fastq.gz
nohup.out result_JSLAB1.txt
QC.1.trimmed.fastq.gz Rockhopper_Results
QC.2.trimmed.fastq.gz
① iu@bielinux[srp017156] cp result2/QC.1.trimmed.fastq.gz .
iu@bielinux[srp017156] cp result2/QC.2.trimmed.fastq.gz . ②
iu@bielinux[srp017156] ls [12:22午後]
QC.1.trimmed.fastq.gz SRR616268sub_1.fastq.gz
QC.2.trimmed.fastq.gz SRR616268sub_2.fastq.gz
result2
iu@bielinux[srp017156] █ [12:22午後]
```

# W14-1: 行数確認

-cオプションをつけて元ファイルを残したままでgzip圧縮ファイルを解凍。パイプ(|)でそのまま行数をカウントするwcコマンドに流すことで、元ファイルを変更することなくgzファイルの行数情報を得ることができる。FaQCs実行前(pre)のファイルは4,000,000行、実行後(post)のファイルは3,908,808行であることがわかる。

```
iu@bielinux[srp017156] gzip result2/*.fastq
iu@bielinux[srp017156] ls result2
fastqCount.txt          QC_qc_report.pdf
JSLAB5_1.R              QC_stats.txt
JSLAB5_2.R              QC.unpaired.trimmed.fastq.gz
nohup.out               result_JSLAB1.txt
QC.1.trimmed.fastq.gz  Rockhopper_Results
QC.2.trimmed.fastq.gz
iu@bielinux[srp017156] cp result2/QC.1.trimmed.fastq.gz .
iu@bielinux[srp017156] cp result2/QC.2.trimmed.fastq.gz .
iu@bielinux[srp017156] ls [12:22午後]
QC.1.trimmed.fastq.gz  SRR616268sub_1.fastq.gz
QC.2.trimmed.fastq.gz  SRR616268sub_2.fastq.gz
result2
① iu@bielinux[srp017156] gunzip -c SRR616268sub_1.fastq.gz | wc
4000000 8000000 320760716
② iu@bielinux[srp017156] gunzip -c SRR616268sub_2.fastq.gz | wc
4000000 8000000 290760716
③ iu@bielinux[srp017156] gunzip -c QC.1.trimmed.fastq.gz | wc
3908808 7817616 313285638
④ iu@bielinux[srp017156] gunzip -c QC.2.trimmed.fastq.gz | wc
3908808 7817616 281534300
iu@bielinux[srp017156] [12:28午後]
```

# W14-2: リストファイル

①リストファイルの作成(正確にはダウンロード)と確認。QuasRは複数サンプルのマッピングが可能。ここでは、FaQCs実行前(pre)と実行後(post)のpaired-endファイルをリストとして与えてマッピングを実行するつもり。

```
File Edit View Search Terminal Help
iu@bielinux[srp017156] pwd
/home/iu/Documents/srp017156
iu@bielinux[srp017156] ls
QC.1.trimmed.fastq.gz  SRR616268sub_1.fastq.gz
QC.2.trimmed.fastq.gz  SRR616268sub_2.fastq.gz
result2
iu@bielinux[srp017156] wget -cq http://www.iu.a.u-tokyo.ac.jp/~kado
ota/book/JSLAB5_4.txt
iu@bielinux[srp017156] ls
JSLAB5_4.txt
QC.1.trimmed.fastq.gz  SRR616268sub_1.fastq.gz
QC.2.trimmed.fastq.gz  SRR616268sub_2.fastq.gz
iu@bielinux[srp017156] █
```

[ 2:03午後 ]

[ 2:03午後 ]

[ 2:03午後 ]



# W14-2: リストファイル

②リストファイルの中身を確認。paired-endの場合は、1行目(FileName1 FileName2 SampleName)の部分はこのように書く(固定)。2行目以降にマッピングしたいFASTQファイル名を書く。3列目(緑の下線)のpreやpostの部分は、任意の文字列でよい。

```
iu@bielinux[srp017156] pwd
/home/iu/Documents/srp017156
iu@bielinux[srp017156] ls
QC.1.trimmed.fastq.gz  SRR616268sub_1.fastq.gz
QC.2.trimmed.fastq.gz  SRR616268sub_2.fastq.gz
result2
iu@bielinux[srp017156] wget -cq http://www.iu.a.u-tokyo.ac.jp/~kadota/book/JSLAB5_4.txt
iu@bielinux[srp017156] ls [ 1:48午後 ]
JSLAB5_4.txt result2
QC.1.trimmed.fastq.gz  SRR616268sub_1.fastq.gz
QC.2.trimmed.fastq.gz  SRR616268sub_2.fastq.gz
iu@bielinux[srp017156] more JSLAB5_4.txt [ 1:48午後 ]
FileName1      FileName2      SampleName
SRR616268sub_1.fastq.gz SRR616268sub_2.fastq.gz pre
QC.1.trimmed.fastq.gz  QC.2.trimmed.fastq.gz  post
iu@bielinux[srp017156] [ 1:48午後 ]
```



②

# W14-3: Rスクリプト

①Rスクリプトファイル(JSLAB5\_5.R)のダウンロードと、②中身の表示。nkfは文字化け回避用。in\_f1がリストファイル[W14-2]。in\_f2がリファレンス配列[W13-1]。gzip圧縮リファレンス配列ファイルの解凍は後で。

```
File Edit View Search Terminal Help
iu@bielinux[srp017156] wget -cq http://www.iu.a.u-tokyo.ac.jp/~ka
dota/book/JSLAB5_5.R
iu@bielinux[srp017156] nkf JSLAB5_5.R [ 1:24午後]
in_f1 <- "JSLAB5_4.txt" #入力ファイル名を指定してi
n_f1に格納 (RNA-seqリストファイル)
in_f2 <- "/home/iu/Documents/genomes/Lactobacillus_casei_12a.GCA
000309565.2.30.dna.toplevel.fa"#入力ファイル名を指定してin_f2に格
納 (リファレンス配列)
#必要なパッケージをロード
library(QuasR) #パッケージの読み込み
#本番 (マッピング)
out <- qAlign(in_f1, in_f2) #マッピングを行うqAlign関
数を実行した結果をoutに格納
#ファイルに保存 (QCレポート用のpdfファイル作成)
out_f <- sub(".bam", "_QC.pdf", out@alignments[,1])#Quqlity Contr
olレポートのpdfファイル名を作成した結果をout_fに格納
qQCReport(out, pdfFilename=out_f) #QCレポート結果をファイル
に保存
iu@bielinux[srp017156] [ 1:24午後]
```

# W14-4: カラー表示

カラー表示。実際のコマンドはごくわずかであることがわかります。①qAlign関数部分がマッピング本番。②qQCReport関数は、PDFレポート作成用

## • カラー表示[W14-4]

Rスクリプトファイル [JSLAB5 5.R](#) の中身を表示。moreで見たものと基本的に同じです。in\_f1には、マップしたいFASTQファイルのリストをQuasRの入力形式に従って作成したファイルの名前([JSLAB5 4.txt](#))、in\_f2にはマップされる側のリファレンス配列を指定します。ここでは作業ディレクトリ上にないRelease 30の乳酸菌ゲノムファイル(解凍したファイル)を絶対パスで指定しています。

```
in_f1 <- "JSLAB5_4.txt" #入力ファイル名を指定してin_f1に格納(RNA-seqリソ  
in_f2 <- "/home/iu/Documents/genomes/Lactobacillus_casei_12a.GCA_000309565.2.30.dna  
  
#必要なパッケージをロード  
library(QuasR) #パッケージの読み込み  
  
#本番(マッピング)  
out <- qAlign(in_f1, in_f2) #マッピングを行うqAlign関数を実行した結果をoutに  
  
#ファイルに保存(QCレポート用のpdfファイル作成)  
out_f <- sub(".bam", "_QC.pdf", out@alignments[,1]) #Quqlity Controlレポートのpdfファ  
qQCReport(out, pdfFilename=out_f) #QCレポート結果をファイルに保存
```

①

②

①gzip圧縮されたリファレンスゲノム配列ファイルを解凍

# W14-5: 解凍

```
File Edit View Search Terminal Help 13:38
iu@bielinux[srp017156] cd ~/Documents/genomes [ 1:38午後 ]
iu@bielinux[genomes] ls [ 1:38午後 ]
JSLAB5_1.R
JSLAB5_3.R
Lactobacillus_casei_12a.GCA_000309565.1.22.dna.toplevel.fa
Lactobacillus_casei_12a.GCA_000309565.2.30.dna.toplevel.fa.gz
result_JSLAB1.txt
① iu@bielinux[genomes] gunzip Lactobacillus_casei_12a.GCA_000309565
.2.30.dna.toplevel.fa.gz
iu@bielinux[genomes] ls [ 1:38午後 ]
JSLAB5_1.R
JSLAB5_3.R
Lactobacillus_casei_12a.GCA_000309565.1.22.dna.toplevel.fa
Lactobacillus_casei_12a.GCA_000309565.2.30.dna.toplevel.fa
result_JSLAB1.txt
iu@bielinux[genomes] [ 1:38午後 ]
```

# W14-5: マッピング本番

QuasRでのマッピング用のRスクリプトファイルJSLAB5\_5.Rを実行。トータルで約15分かかる。

```
iu@bielinux[genomes] cd ~/Documents/srp017156 [ 1:40午後 ]
iu@bielinux[srp017156] pwd [ 1:40午後 ]
/home/iu/Documents/srp017156
iu@bielinux[srp017156] ls [ 1:40午後 ]
JSLAB5_4.txt result2
JSLAB5_5.R SRR616268sub_1.fastq.gz
QC.1.trimmed.fastq.gz SRR616268sub_2.fastq.gz
QC.2.trimmed.fastq.gz
iu@bielinux[srp017156] R --vanilla --slave < JSLAB5_5.R
```



①

# W14-5: 途中経過1

リターンキーを押して数秒後の状態。①まず最初にやっているのは、リファレンス配列のインデックス化。インデックス化(indexing)することでマッピングを高速に行うことができます。数MB程度の乳酸菌ゲノムの場合は比較的短時間(数分のオーダー)で終わりますがヒトゲノムだと数十分以上はかかるのではと思います。ただし、同じリファレンス配列を使って別のデータのマッピングを行う場合には、既にインデックス化されたものを使うのでこの部分はスキップできます。

```
iu@bielinux[~/Documents/srp017156]
int,
  rownames, sapply, setdiff, sort, table, ta
,
  unlist, unsplit

Loading required package: S4Vectors
Loading required package: stats4
Creating a generic function for 'nchar' from p
ckage 'S4Vectors'
Loading required package: IRanges
Loading required package: GenomeInfoDb
Loading required package: Rbowtie
Creating .fai file for: /home/iu/Documents/genomes/Lactobacillus_
casei_12a.GCA_000309565.2.30.dna.toplevel.fa
[fai_load] build FASTA index.
alignment files missing - need to:
  create alignment index for the genome
  create 2 genomic alignment(s)
① Creating an Rbowtie index for /home/iu/Documents/genomes/Lactobac
illus_casei_12a.GCA_000309565.2.30.dna.toplevel.fa
Finished creating index
Testing the compute nodes...OK
Loading QuasR on the compute nodes...█
```

# W14-5: 途中経過2

①マッピングがスタート。この種のプログラムは実行ログファイルを作成する場合があります。QuasRも②の絶対パスで示したファイル名にログを書き込んでいます

```
File Edit View Search Terminal Help
ckage 'S4Vectors'
Loading required package: IRanges
Loading required package: GenomeInfoDb
Loading required package: Rbowtie
Creating .fai file for: /home/iu/Documents/genomes/Lactobacillus_
casei_12a.GCA_000309565.2.30.dna.toplevel.fa
[fai_load] build FASTA index.
alignment files missing - need to:
  create alignment index for the genome
  create 2 genomic alignment(s)
Creating an Rbowtie index for /home/iu/Documents/genomes/Lactobac
illus_casei_12a.GCA_000309565.2.30.dna.toplevel.fa
Finished creating index
Testing the compute nodes...OK
Loading QuasR on the compute nodes...OK
Available cores:
nodeNames
bielinux
      1
> ① Performing genomic alignments for 2 samples. See progress in the
log file:
/home/iu/Documents/srp017156/QuasR_log_47ee3bd7d050.txt ②
```

# W14-5: 途中経過3

①samやbamと書かれているが、これは多くのマッピングプログラム(QuasRのデフォルトは内部的にBowtieプログラムを利用)の結果ファイルの形式がbam形式だから。bamはsamのバイナリ版。

```
File Edit View Search Terminal Help
Loading required package: GenomeInfoDb
Loading required package: Rbowtie
Creating .fai file for: /home/iu/Documents/genomes/Lactobacillus_
casei_12a.GCA_000309565.2.30.dna.toplevel.fa
[fai_load] build FASTA index.
alignment files missing - need to:
  create alignment index for the genome
  create 2 genomic alignment(s)
Creating an Rbowtie index for /home/iu/Documents/genomes/Lactobac
illus_casei_12a.GCA_000309565.2.30.dna.toplevel.fa
Finished creating index
Testing the compute nodes...OK
Loading QuasR on the compute nodes...OK
Available cores:
nodeNames
bielinux
      1
Performing genomic alignments for 2 samples. See progress in the
log file:
/home/iu/Documents/srp017156/QuasR_log_47ee3bd7d050.txt
[samopen] SAM header is present: 1 sequences.
[bam_sort_core] merging from 2 files...
```



# W14-5: 途中経過4

①2回目のsamやbamの記述。おそらく2つめのサンプル(リストファイルの3行目。この場合FaQCs実行後のファイルQC.\*.fastq.gz)のマッピングを行っているのだろう。

```
File Edit View Search Terminal Help
Creating .fai file for: /home/iu/Documents/genomes/Lactobacillus_
casei_12a.GCA_000309565.2.30.dna.toplevel.fa
[fai_load] build FASTA index.
alignment files missing - need to:
    create alignment index for the genome
    create 2 genomic alignment(s)
Creating an Rbowtie index for /home/iu/Documents/genomes/Lactobac
illus_casei_12a.GCA_000309565.2.30.dna.toplevel.fa
Finished creating index
Testing the compute nodes...OK
Loading QuasR on the compute nodes...OK
Available cores:
nodeNames
bielinux
    1
Performing genomic alignments for 2 samples. See progress in the
log file:
/home/iu/Documents/srp017156/QuasR_log_47ee3bd7d050.txt
[samopen] SAM header is present: 1 sequences.
[bam_sort_core] merging from 2 files...
[samopen] SAM header is present: 1 sequences.
[bam_sort_core] merging from 2 files...
```



# W14-5: 途中経過5

- ①マッピングは無事に終了したようだ。
- ②QC情報を得ようとしているのだろう。

```
alignment files missing - need to:
  create alignment index for the genome
  create 2 genomic alignment(s)
Creating an Rbowtie index for /home/iu/Documents/genomes/Lactobacillus_casei_12a.GCA_000309565.2.30.dna.toplevel.fa
Finished creating index
Testing the compute nodes...OK
Loading QuasR on the compute nodes...OK
Available cores:
nodeNames
bielinux
      1
Performing genomic alignments for 2 samples. See progress in the log file:
/home/iu/Documents/srp017156/QuasR_log_47ee3bd7d050.txt
[samopen] SAM header is present: 1 sequences.
[bam_sort_core] merging from 2 files...
[samopen] SAM header is present: 1 sequences.
[bam_sort_core] merging from 2 files...
① Genomic alignments have been created successfully
② collecting quality control data
```

# W14-5: 無事終了

```
File Edit View Search Terminal Help 13:54
create alignment index for the genome
create 2 genomic alignment(s)
Creating an Rbowtie index for /home/iu/Documents/genomes/Lactobacillus_casei_12a.GCA_000309565.2.30.dna.toplevel.fa
Finished creating index
Testing the compute nodes...OK
Loading QuasR on the compute nodes...OK
Available cores:
nodeNames
bielinux
1
Performing genomic alignments for 2 samples. See progress in the log file:
/home/iu/Documents/srp017156/QuasR_log_47ee3bd7d050.txt
[samopen] SAM header is present: 1 sequences.
[bam_sort_core] merging from 2 files...
[samopen] SAM header is present: 1 sequences.
[bam_sort_core] merging from 2 files...
Genomic alignments have been created successfully

collecting quality control data
creating QC plots
iu@bielinux[srp017156] █ [ 1:54午後]
```



# W15-1: 結果の解説

①lsした結果。②多数のファイル(計8ファイル)が生成されていることがわかる。マッピング前[W14-5]と比べてみるとよい

```
File Edit View Search Terminal Help
[samopen] SAM header is present: 1 sequences.
[bam_sort_core] merging from 2 files...
Genomic alignments have been created successfully

collecting quality control data
creating QC plots
iu@bielinux[srp017156] ls
JSLAB5_4.txt
JSLAB5_5.R
QC.1.trimmed_47ee41c6ec8b.bam
QC.1.trimmed_47ee41c6ec8b.bam.bai
QC.1.trimmed_47ee41c6ec8b.bam.txt
QC.1.trimmed.fastq.gz
QC.2.trimmed.fastq.gz
QuasR_log_47ee3bd7d050.txt
result2
SRR616268sub_1_47ee4589e65f.bam
SRR616268sub_1_47ee4589e65f.bam.bai
SRR616268sub_1_47ee4589e65f.bam.txt
SRR616268sub_1_47ee4589e65f_QC.pdf
SRR616268sub_1.fastq.gz
SRR616268sub_2.fastq.gz
iu@bielinux[srp017156]
```

[ 1:54午後 ]

[ 1:56午後 ]

# W15-1: 結果の解説

マッピング結果ファイルのメインは①と②で示した.bam。この形式のファイルを入力としてその後の解析を行うプログラムも多い。③エラーが出たりすることがなければlogファイルの中身をあまり見ることはないが、この中をよく見るとマッピング時に用いたオプション情報などを読み取ることができる。

[ 1:54午後 ]

[ 1:56午後 ]

```
[samopen] SAM header is present: 1 sequences.  
[bam_sort_core] merging from 2 files...  
Genomic alignments have been created successfully
```

```
collecting quality control data  
creating QC plots
```

```
iu@bielinux[srp017156] ls  
JSLAB5_4.txt  
JSLAB5_5.R  
① QC.1.trimmed_47ee41c6ec8b.bam  
QC.1.trimmed_47ee41c6ec8b.bam.bai  
QC.1.trimmed_47ee41c6ec8b.bam.txt  
QC.1.trimmed.fastq.gz  
QC.2.trimmed.fastq.gz  
QuasR_log_47ee3bd7d050.txt ③  
result2  
② SRR616268sub_1_47ee4589e65f.bam  
SRR616268sub_1_47ee4589e65f.bam.bai  
SRR616268sub_1_47ee4589e65f.bam.txt  
SRR616268sub_1_47ee4589e65f_QC.pdf  
SRR616268sub_1.fastq.gz  
SRR616268sub_2.fastq.gz  
iu@bielinux[srp017156]
```

# W15-1: 結果の解説

```
File Edit View Search Terminal Help 13:56
[samopen] SAM header is present: 1 sequences.
[bam_sort_core] merging from 2 files...
Genomic alignments have been created successfully

collecting quality control data
creating QC plots
iu@bielinux[srp017156] ls
JSLAB5_4.txt
JSLAB5_5.R
QC.1.trimmed_47ee41c6ec8b.bam
QC.1.trimmed_47ee41c6ec8b.bam.bai
QC.1.trimmed_47ee41c6ec8b.bam.txt
QC.1.trimmed.fastq.gz
QC.2.trimmed.fastq.gz
QuasR_log_47ee3bd7d050.txt
result2
SRR616268sub_1_47ee4589e65f.bam
SRR616268sub_1_47ee4589e65f.bam.bai
SRR616268sub_1_47ee4589e65f.bam.txt
SRR616268sub_1_47ee4589e65f_QC.pdf
SRR616268sub_1.fastq.gz
SRR616268sub_2.fastq.gz
iu@bielinux[srp017156]
```

[ 1:54午後 ]

[ 1:56午後 ]

# W15-1: 結果の解説

①このPDFファイル中には、入力ファイル (paired-end RNA-seqリード)のQC情報や、どれだけマップされたかの結果などが含まれる。

```
File Edit View Search Terminal Help
[samopen] SAM header is present: 1 sequences.
[bam_sort_core] merging from 2 files...
Genomic alignments have been created successfully

collecting quality control data
creating QC plots
iu@bielinux[srp017156] ls
JSLAB5_4.txt
JSLAB5_5.R
QC.1.trimmed_47ee41c6ec8b.bam
QC.1.trimmed_47ee41c6ec8b.bam.bai
QC.1.trimmed_47ee41c6ec8b.bam.txt
QC.1.trimmed.fastq.gz
QC.2.trimmed.fastq.gz
QuasR_log_47ee3bd7d050.txt
result2
SRR616268sub_1_47ee4589e65f.bam
SRR616268sub_1_47ee4589e65f.bam.bai
SRR616268sub_1_47ee4589e65f.bam.txt
SRR616268sub_1_47ee4589e65f_QC.pdf
SRR616268sub_1.fastq.gz
SRR616268sub_2.fastq.gz
iu@bielinux[srp017156]
```

[ 1:54午後 ]

[ 1:56午後 ]





# W15-2: リファレンスのほう

①リファレンスゲノムファイルがあるディレクトリをls。②リファレンスとして指定したファイル。③W14-5の最初でリファレンスゲノムのインデックス化を行っていたが、そのときに作成されたのが赤枠の3ファイル。「~/Documents/genomes」の所有者が自分なので、これらのファイルを作成することができた。が、スパコンなどで共用のリファレンスゲノムのディレクトリを利用する際には、書き込み権限がないことに起因するエラーが起こるかもしれないので記憶に留めておこう

```
File Edit View Search Terminal Help
QC.1.trimmed_47ee41c6ec8b.bam.txt
QC.1.trimmed.fastq.gz
QC.2.trimmed.fastq.gz
QuasR_log_47ee3bd7d050.txt
result2
SRR616268sub_1_47ee4589e65f.bam
SRR616268sub_1_47ee4589e65f.bam.bai
SRR616268sub_1_47ee4589e65f.bam.txt
SRR616268sub_1_47ee4589e65f_QC.pdf
SRR616268sub_1.fastq.gz
SRR616268sub_2.fastq.gz
iu@bielinux[srp017156] pwd
/home/iu/Documents/srp017156
iu@bielinux[srp017156] ls ~/Documents/genomes [ 2:06午後 ]
JSLAB5_1.R
JSLAB5_3.R
Lactobacillus_casei_12a.GCA_000309565.1.22.dna.toplevel.fa
Lactobacillus_casei_12a.GCA_000309565.2.30.dna.toplevel.fa
Lactobacillus_casei_12a.GCA_000309565.2.30.dna.toplevel.fa.fai
Lactobacillus_casei_12a.GCA_000309565.2.30.dna.toplevel.fa.md5
Lactobacillus_casei_12a.GCA_000309565.2.30.dna.toplevel.fa.Rbowtie
result_JSLAB1.txt
iu@bielinux[srp017156] [ 2:06午後 ]
```



# W15-3: QCレポート

QuasRでマッピングしたのは、QCレポートを眺めるのが主目的。ここでは、①pdfファイルを共有フォルダ(~/Desktop/mac\_share)にコピーしてホストOS上で眺めるが...

```
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.toplevel.fa
QC.1.trimmed_fastqc.html
QC.1.trimmed_fastqc.zip
QC.2.trimmed_fastqc.html
QC.2.trimmed_fastqc.zip
SRR616268sub_1_fastqc.html
SRR616268sub_1_fastqc.zip
SRR616268sub_2_fastqc.html
SRR616268sub_2_fastqc.zip
iu@bielinux[srp017156] cp *.pdf ~/Desktop/mac_share
iu@bielinux[srp017156] █
```

[ 2:09午後 ]

[ 2:09午後 ]

[ 2:09午後 ]



# W15-3: QCレポート

①引出しアイコンをクリックして  
いてゲストOS上で眺めてもよい。

File Edit View Go Bookmarks Help

Home Documents srp017156

Places

- Recent
- Home
- Desktop
- Documents
- Downloads
- Music
- Pictures
- Videos
- Trash

Devices

- Computer

Network

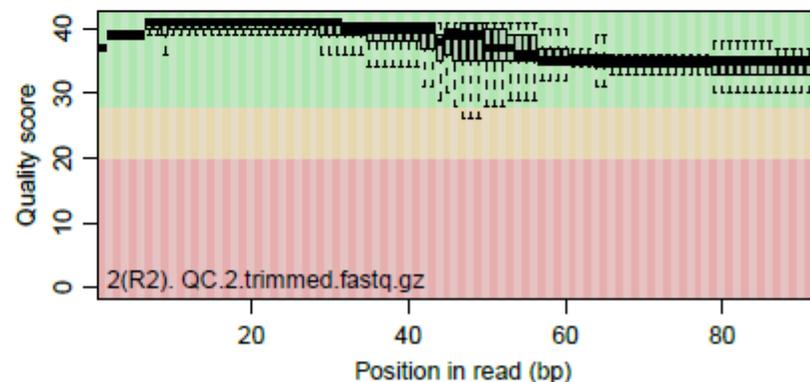
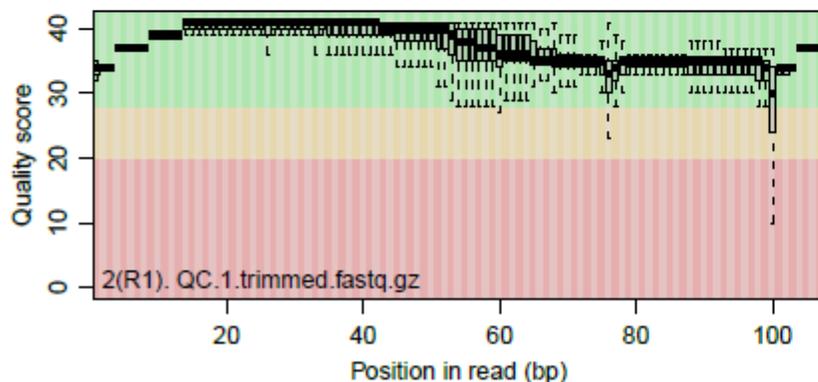
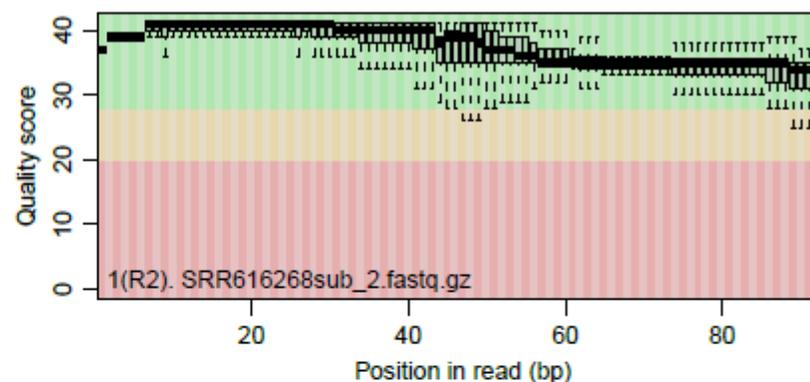
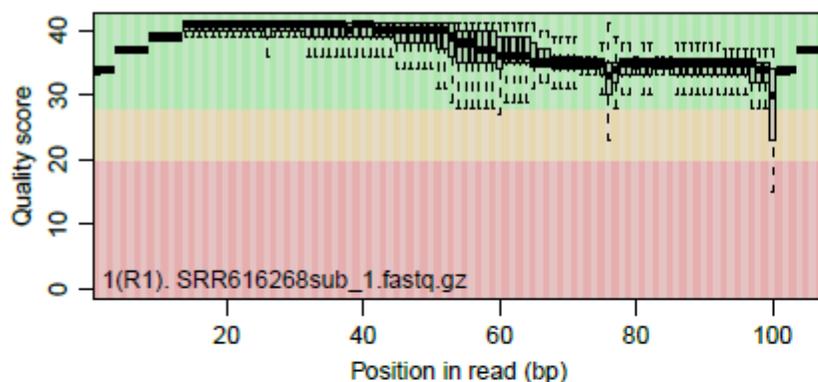
- Browse Network
- Connect to Server

Name	Size	Type	Modified
result2	11 items	Folde	11:36
SRR616268sub_1_47ee4589e65f_QC.pdf	80.4 kB	Docur	13:54
QuasR_log_47ee3bd7d050.txt	1.7 kB	Text	13:53
QC.1.trimmed_47ee41c6ec8b.bam.txt	620 bytes	Text	13:53
QC.1.trimmed_47ee41c6ec8b.bam.bai	6.2 kB	Binary	13:53
QC.1.trimmed_47ee41c6ec8b.bam	138.9 MB	Archiv	13:53
SRR616268sub_1_47ee4589e65f.bam.txt	624 bytes	Text	13:47
SRR616268sub_1_47ee4589e65f.bam.bai	6.2 kB	Binary	13:47
SRR616268sub_1_47ee4589e65f.bam	142.1 MB	Archiv	13:47
JSLAB5_5.R	698 bytes	Text	13:23
QC.2.trimmed.fastq.gz	65.9 MB	Archiv	11:37
QC.1.trimmed.fastq.gz	73.7 MB	Archiv	11:37
SRR616268sub_2.fastq.gz	68.7 MB	Archiv	12月 9
SRR616268sub_1_47ee4589e65f_QC.pdf	80.4 kB	Docur	13:54

"SRR616268sub\_1\_47ee4589e65f\_QC.pdf" selected (80.4 kB)

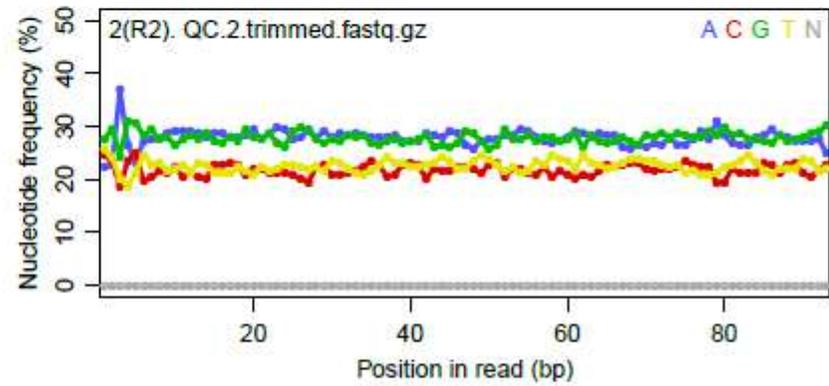
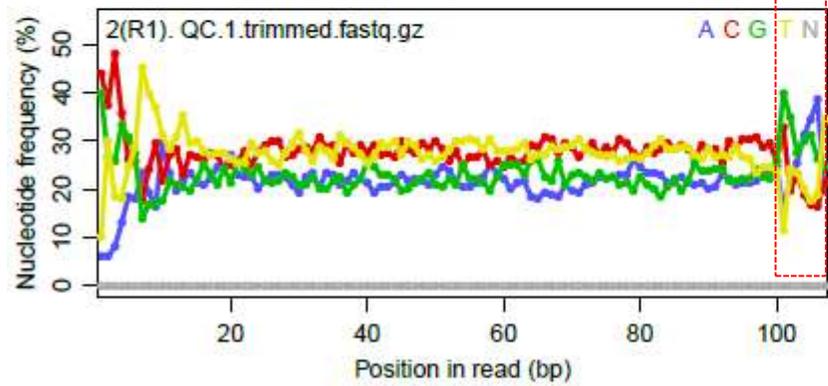
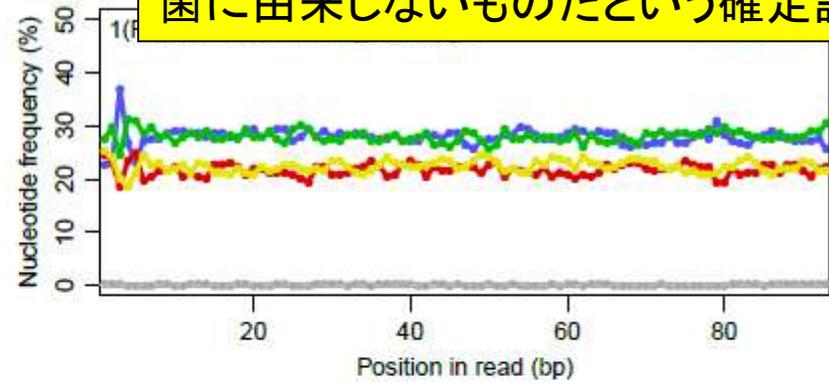
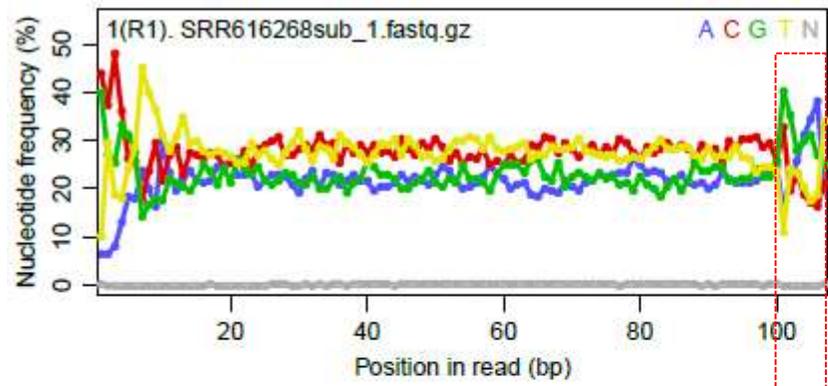
# W15-4: PDF解説

PDF1枚目。入力ファイルのQuality score分布。  
FastQC Report中の項目「Per base sequence quality」と同じ。上段がFaQCs実行前(pre)、下段が実行後(post)。左がforward側、右がreverse側。ここでの目的はFaQCs実行前後の比較ではなく、マップされなかったリードの割合や、数少ないマップされたリードの調査なので、劇的な違いはないが気にしない。



# W15-5: PDF解説

PDF2枚目。ポジションごとの塩基の出現確率。FastQC Report中の項目「Per base sequence content」と同じ(但し色は異なる)。①forward側の100-107bp付近(赤枠部分)の分布が不自然。このような結果は、FastQCをデフォルトオプションで実行すると得られない。この結果と後のほうのPDFレポートと合わせることで、これがトリムしきれていない乳酸菌に由来しないものだという確定診断が下される。



# W15-6: PDF解説

PDF4枚目。全リード(forward, reverse合わせて約200万リード)のうち、マップされたリードの割合はFaQCs実行前(pre; 上側)が0.4%、実行後(post; 下側)が0.5%。ほとんどマップされなかったことを意味する

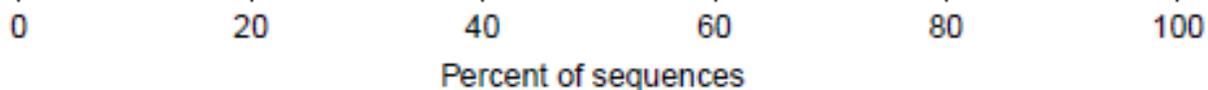
■ mapped ■ unmapped

1. SRR616268sub\_1.fastq.gz

0.4% total=2e+06 99.6%

2. QC.1.trimmed.fastq.gz

0.5% total=1.95e+06 99.5%



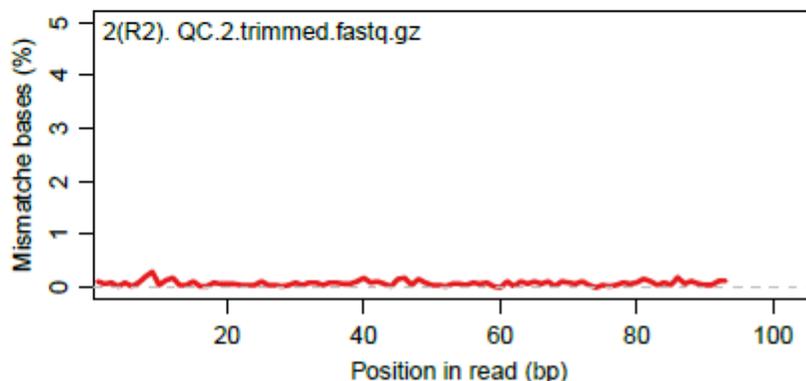
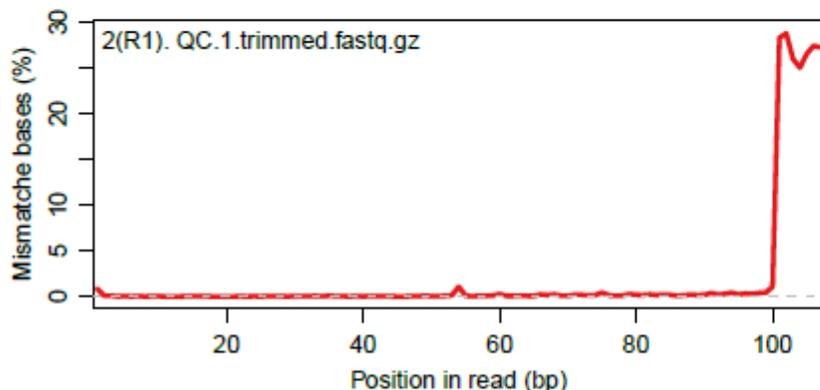
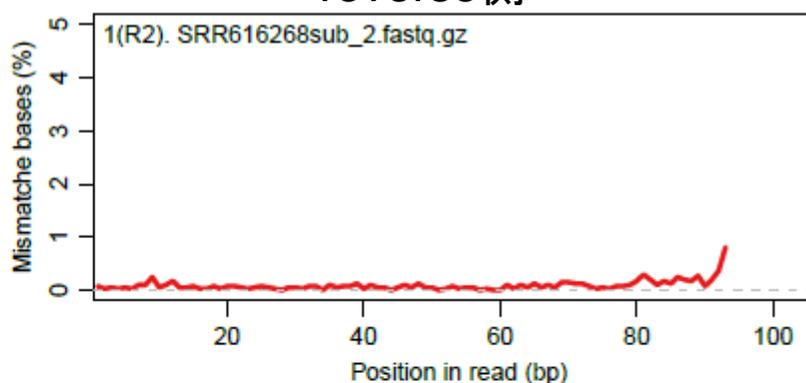
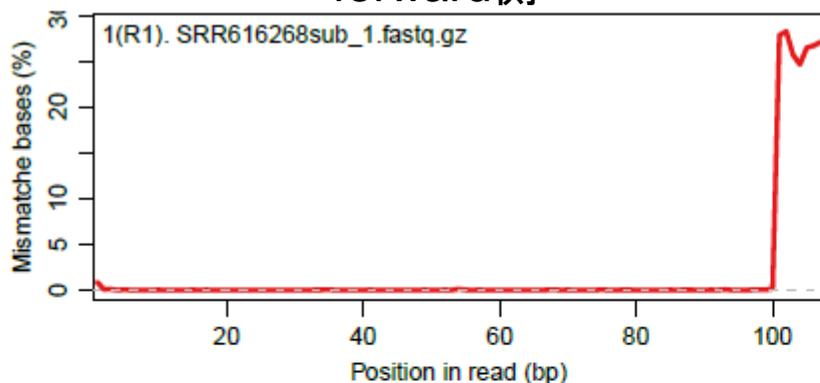
# W15-7: PDF解説

PDF6枚目。マニュアルを読んでもよくわからなかったが、おそらく縦軸がMismatch basesとなっているので、 mismatchesを許容してマップされたリードの中でどこに mismatchesがあったかを表示しているものと思われる。多少解釈が間違っていたとしても、①このプロット分布を見れば、「forward側の100-107 bp付近が犯人」という結論は不変

forward側



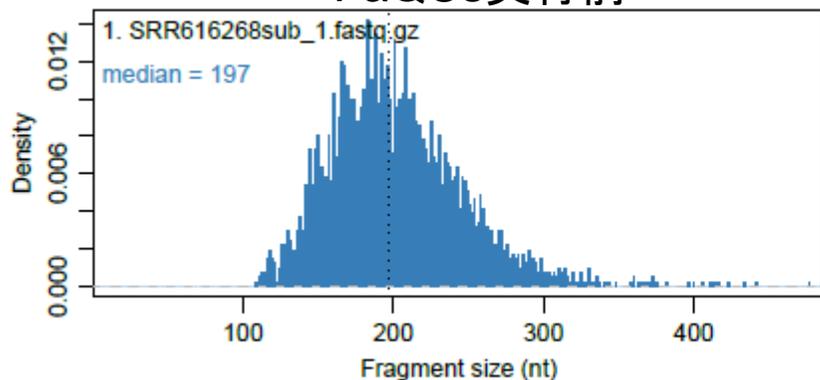
reverse側



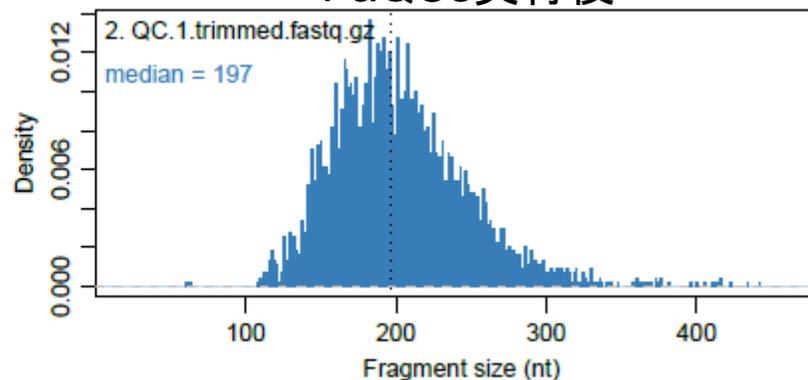
# W15-8: PDF解説

PDF8枚目。入力はpaired-endなので、おそらくforward側とreverse側両方でマップされたリードのみを取り扱っている。ゲノム配列上でのforwardとreverse間の距離分布をプロットしているものと思われる。

FaQCs実行前



FaQCs実行後







# W16-1:トリミング

①作業ディレクトリはここ。②ダウンロードしたJSLAB5\_6.Rの最初の3行分を表示。③入力ファイルは相対パスで示したSRR616268sub\_1.fastq.gz。④3'末端の8塩基を除去した結果を⑤hoge\_1.fastq.gzというファイル名で保存。⑥スクリプトファイルの実行。数秒

```
iu@bielinux[~] File Edit View Search Terminal Help
iu@bielinux[~] cd ~/Documents/
iu@bielinux[Documents] mkdir srp017156_trim1
iu@bielinux[Documents] cd srp017156_trim1
iu@bielinux[srp017156_trim1] pwd
/home/iu/Documents/srp017156_trim1
iu@bielinux[srp017156_trim1] wget -cq http://www.iu.a.u-tokyo.ac.jp/~kadota/book/JSLAB5_6.R
iu@bielinux[srp017156_trim1] nkf JSLAB5_6.R | head -n 3
in_f <- "~/Documents/srp017156/SRR616268sub_1.fastq.gz" #入力ファイル名を指定してin_fに格納
out_f <- "hoge_1.fastq.gz" #出力ファイル名を指定してout_fに格納
param_trim <- 8 #3'末端のトリムしたい塩基数を指定
iu@bielinux[srp017156_trim1] R --vanilla --slave < JSLAB5_6.R
```



# W16-1:トリミング

①reverse側のファイルとしてSRR616268sub\_2.fastq.gzを作業ディレクトリにコピー。②hoge\_1.fastq.gz (ファイルサイズ66,235,765 bytes)は、JSLAB5\_6.Rの実行結果ファイル。③ JSLAB5\_6.Rの入力ファイル(SRR616268sub\_1.fastq.gz)は76,659,501 bytes。107 bpが99 bpになった結果のファイルサイズの減少度合い的に妥当

```
File Edit View Search Terminal Help
[999996] 99 abbeeccgggggiiiihiiiiifhiihi
ccccbcbccbc]`
[999997] 99 abbeecegggggiiihghihiiiiifh
cbPT`aa^bcaca_
[999998] 99 bbbeeeefggggghiihiighiiiiini...ccccccccccccca_
ac]^accdcca^a
[999999] 99 ab_eeeeebegggghiiiiiiiiiiiiighi...ccccbcbccccccca
acc^`acaac_ac
[1000000] 99 bbbeeeegggggiiiefghiiiiigiiiihi...eedddddccccccc
cccccccccccccc
① iu@bielinux[srp017156_trim1] cp ~/Documents/srp017156/SRR616268sub
_2.fastq.gz .
iu@bielinux[srp017156_trim1] ls -l [ 2:42午後 ]
total 131764
-rw-rw-r-- 1 iu iu 66235765 12月 22 14:36 hoge_1.fastq.gz ②
-rw-rw-r-- 1 iu iu 1298 9月 14 18:22 JSLAB5_6.R
-rw-rw-r-- 1 iu iu 68682959 12月 22 14:42 SRR616268sub_2.fastq.gz
③ iu@bielinux[srp017156_trim1] ls -l ~/Documents/srp017156/SRR616268
sub_1.fastq.gz
-rw-rw-r-- 1 iu iu 76659501 12月 9 15:24 /home/iu/Documents/srp01
7156/SRR616268sub_1.fastq.gz
iu@bielinux[srp017156_trim1] [ 2:42午後 ]
```

# W16-2:トリミング

①FASTX-Toolkitのfastx\_trimmerを利用するやり方。②fastx\_trimmerはgzip圧縮ファイルの入力を受け付けないので、gunzipした結果をパイプで渡している

```
iu@bielinux[srp017156_trim1] pwd [ 2:52午後 ]
/home/iu/Documents/srp017156_trim1
iu@bielinux[srp017156_trim1] ls [ 2:52午後 ]
hoge_1.fastq.gz JSLAB5_6.R SRR616268sub_2.fastq.gz
iu@bielinux[srp017156_trim1] gunzip -c ~/Documents/srp017156/SRR61 [ 2:52午後 ]
6268sub_1.fastq.gz | fastx_trimmer -l 99 - | gzip > hoge_2.fastq.g
Z
iu@bielinux[srp017156_trim1] ls -l [ 2:52午後 ]
total 201408
-rw-rw-r-- 1 iu iu 66235765 12月 22 14:36 hoge_1.fastq.gz
-rw-rw-r-- 1 iu iu 71314779 12月 22 14:52 hoge_2.fastq.gz
-rw-rw-r-- 1 iu iu 1298 9月 14 18:22 JSLAB5_6.R
-rw-rw-r-- 1 iu iu 68682959 12月 22 14:42 SRR616268sub_2.fastq.gz
iu@bielinux[srp017156_trim1] █ [ 2:52午後 ]
```



# W16-2:トリミング

③この「- (ハイフン)」は、パイプで渡したものを入力として受け付けるという明示的な意思表示。省略することができるコマンド(or プログラム)もあるが、fastx\_trimmerは明示しないと怒られるのでつけている。

```
iu@bielinux[srp017156_trim1] pwd [ 2:52午後 ]
/home/iu/Documents/srp017156_trim1
iu@bielinux[srp017156_trim1] ls [ 2:52午後 ]
hoge_1.fastq.gz JSLAB5_6.R SRR616268sub_2.fastq.gz
iu@bielinux[srp017156_trim1] gunzip -c ~/Documents/srp017156/SRR61
6268sub_1.fastq.gz | fastx_trimmer -l 99 - | gzip > hoge_2.fastq.g
z
iu@bielinux[srp017156_trim1] ls -l [ 2:52午後 ]
total 201408
-rw-rw-r-- 1 iu iu 66235765 12月 22 14:36 hoge_1.fastq.gz
-rw-rw-r-- 1 iu iu 71314779 12月 22 14:52 hoge_2.fastq.gz
-rw-rw-r-- 1 iu iu 1298 9月 14 18:22 JSLAB5_6.R
-rw-rw-r-- 1 iu iu 68682959 12月 22 14:42 SRR616268sub_2.fastq.gz
iu@bielinux[srp017156_trim1] █ [ 2:52午後 ]
```



# W16-2:トリミング

④fastx\_trimmer実行の本体部分。-lでリードの何塩基目までを残すかを指定。ここでは、(100塩基目以降をトリムしたいので)99塩基目まで残すという指定を行っている。

```
iu@bielinux[srp017156_trim1] pwd [ 2:52午後 ]
/home/iu/Documents/srp017156_trim1
iu@bielinux[srp017156_trim1] ls [ 2:52午後 ]
hoge_1.fastq.gz JSLAB5_6.R SRR616268sub_2.fastq.gz
iu@bielinux[srp017156_trim1] gunzip -c ~/Documents/srp017156/SRR61
6268sub_1.fastq.gz | fastx_trimmer -l 99 - | gzip > hoge_2.fastq.g
z
iu@bielinux[srp017156_trim1] ls -l [ 2:52午後 ]
total 201408
-rw-rw-r-- 1 iu iu 66235765 12月 22 14:36 hoge_1.fastq.gz
-rw-rw-r-- 1 iu iu 71314779 12月 22 14:52 hoge_2.fastq.gz
-rw-rw-r-- 1 iu iu 1298 9月 14 18:22 JSLAB5_6.R
-rw-rw-r-- 1 iu iu 68682959 12月 22 14:42 SRR616268sub_2.fastq.gz
iu@bielinux[srp017156_trim1] █ [ 2:52午後 ]
```



# W16-2:トリミング

⑤ここでは、トリミングした結果をパイプで流してgzip圧縮してhoge\_2.fastq.gzというファイル名で保存するという指令。「| gzip -> ...」とハイフン(-)を明示してもよい。fastx\_trimmerの-zや-oオプションを使う書き方もある。表現方法はいろいろ。

```
iu@bielinux[srp017156_trim1] pwd
/home/iu/Documents/srp017156_trim1
iu@bielinux[srp017156_trim1] ls
hoge_1.fastq.gz  JSLAB5_6.R  SRR616268sub_2.fastq.gz
iu@bielinux[srp017156_trim1] gunzip -c ~/Documents/srp017156/SRR61
6268sub_1.fastq.gz | fastx_trimmer -l 99 - | gzip > hoge_2.fastq.g
z
iu@bielinux[srp017156_trim1] ls -l
total 201408
-rw-rw-r-- 1 iu iu 66235765 12月 22 14:36 hoge_1.fastq.gz
-rw-rw-r-- 1 iu iu 71314779 12月 22 14:52 hoge_2.fastq.gz
-rw-rw-r-- 1 iu iu 1298 9月 14 18:22 JSLAB5_6.R
-rw-rw-r-- 1 iu iu 68682959 12月 22 14:42 SRR616268sub_2.fastq.gz
iu@bielinux[srp017156_trim1] █
```

[ 2:52午後 ]

[ 2:52午後 ]



# W16-3: 確認

①ls -lでファイルサイズを確認。RのBiostringsでの実行結果(hoge\_1.fastq.gz)とfastx\_trimmerでの実行結果(hoge\_2.fastq.gz)のファイルサイズが異なっている。この場合、どちらかのプログラムにバグがある可能性を考えるのが自然。

```
File Edit View Search Terminal Help
iu@bielinux[srp017156_trim1] pwd
/home/iu/Documents/srp017156_trim1
iu@bielinux[srp017156_trim1] ls [ 2:52午後 ]
hoge_1.fastq.gz JSLAB5_6.R SRR616268sub_2.fastq.gz
iu@bielinux[srp017156_trim1] gunzip -c ~/Documents/srp017156/SRR61
6268sub_1.fastq.gz | fastx_trimmer -l 99 - | gzip > hoge_2.fastq.g
z
iu@bielinux[srp017156_trim1] ls -l [ 2:52午後 ]
total 201408
-rw-rw-r-- 1 iu iu 66235765 12月 22 14:36 hoge_1.fastq.gz
-rw-rw-r-- 1 iu iu 71314779 12月 22 14:52 hoge_2.fastq.gz
-rw-rw-r-- 1 iu iu 1298 9月 14 18:22 JSLAB5_6.R
-rw-rw-r-- 1 iu iu 68682959 12月 22 14:42 SRR616268sub_2.fastq.gz
iu@bielinux[srp017156_trim1] [ 2:52午後 ]
```



-rw-rw-r--	1	iu	iu	66235765	12月	22	14:36	hoge_1.fastq.gz
-rw-rw-r--	1	iu	iu	71314779	12月	22	14:52	hoge_2.fastq.gz
-rw-rw-r--	1	iu	iu	1298	9月	14	18:22	JSLAB5_6.R
-rw-rw-r--	1	iu	iu	68682959	12月	22	14:42	SRR616268sub_2.fastq.gz

# W16-3: 確認

①RのBiostringsでの実行結果(hoge\_1.fastq.gz)と②fastx\_trimmerでの実行結果(hoge\_2.fastq.gz)の最初の4行分を表示。両者の違いは赤枠部分のdescription情報の有無だけのようにであり、バグではなさそうだ。

```
File Edit View Search Terminal Help
iu@bielinux[srp017156_trim1] ls -l
total 201408
-rw-rw-r-- 1 iu iu 66235765 12月 22 14:36 hoge_1.fastq.gz
-rw-rw-r-- 1 iu iu 71314779 12月 22 14:52 hoge_2.fastq.gz
-rw-rw-r-- 1 iu iu      1298  9月 14 18:22 JSLAB5_6.R
-rw-rw-r-- 1 iu iu 68682959 12月 22 14:42 SRR616268sub_2.fastq.gz
iu@bielinux[srp017156_trim1] gunzip -c hoge_1.fastq.gz | head -n 4
@SRR616268.7 2291:6:1101:1412:2249 length=107
AGCCCGACTTTCGTCCCTGCTCGACTTGTCAGTCTCGCAGTCAAGCTCCCTTATACCTTTACTC
TGCGAATGATTCCAACCATTCTGAGGGAACCT
+
bbbeeeeegggggiiiiiiiiiiiiiiiiighiihihihihihihihihfihiighhigggggggeeeee
dddcacccccccddccccccdbccaaacccb
iu@bielinux[srp017156_trim1] gunzip -c hoge_2.fastq.gz | head -n 4
@SRR616268.7 2291:6:1101:1412:2249 length=107
AGCCCGACTTTCGTCCCTGCTCGACTTGTCAGTCTCGCAGTCAAGCTCCCTTATACCTTTACTC
TGCGAATGATTCCAACCATTCTGAGGGAACCT
+SRR616268.7 2291:6:1101:1412:2249 length=107
bbbeeeeegggggiiiiiiiiiiiiiiiiighiihihihihihihihihfihiighhigggggggeeeee
dddcacccccccddccccccdbccaaacccb
iu@bielinux[srp017156_trim1] [ 2:57午後 ]
```

# W16-3: 確認

①RのBiostringsでの実行結果(hoge\_1.fastq.gz)と②fastx\_trimmerでの実行結果(hoge\_2.fastq.gz)の最後の4行分を表示。大丈夫そうだ。

```
File Edit View Search Terminal Help
iu@bielinux[srp017156_trim1] ls -l
total 201408
-rw-rw-r-- 1 iu iu 66235765 12月 22 14:36 hoge_1.fastq.gz
-rw-rw-r-- 1 iu iu 71314779 12月 22 14:52 hoge_2.fastq.gz
-rw-rw-r-- 1 iu iu 1298 9月 14 18:22 JSLAB5_6.R
-rw-rw-r-- 1 iu iu 68682959 12月 22 14:42 SRR616268sub_2.fastq.gz
iu@bielinux[srp017156_trim1] gunzip -c hoge_1.fastq.gz | tail -n 4
[SRR616268.1000860 2291:6:1101:11638:95311 length=107
GCCTTGTCAATCAAGGTGAGCATGTCGCCCATGCCAGAATTCGGTTGCCATGCGATCAGGATAG
AAGACATCCAGCGCATCCATCTTTTCACCTTGA
+
bbbeeeeeggggggiiiefghiiiigiiihiiiiihiiiiiiiiihifiiiiiiiiiiiiihgggggg
eeeeeddddddcccccccccccccccccccccc
iu@bielinux[srp017156_trim1] gunzip -c hoge_2.fastq.gz | tail -n 4
[SRR616268.1000860 2291:6:1101:11638:95311 length=107
GCCTTGTCAATCAAGGTGAGCATGTCGCCCATGCCAGAATTCGGTTGCCATGCGATCAGGATAG
AAGACATCCAGCGCATCCATCTTTTCACCTTGA
+SRR616268.1000860 2291:6:1101:11638:95311 length=107
bbbeeeeeggggggiiiefghiiiigiiihiiiiihiiiiiiiiihifiiiiiiiiiiiiihgggggg
eeeeeddddddcccccccccccccccccccccc
iu@bielinux[srp017156_trim1] [ 2:58午後 ]
```

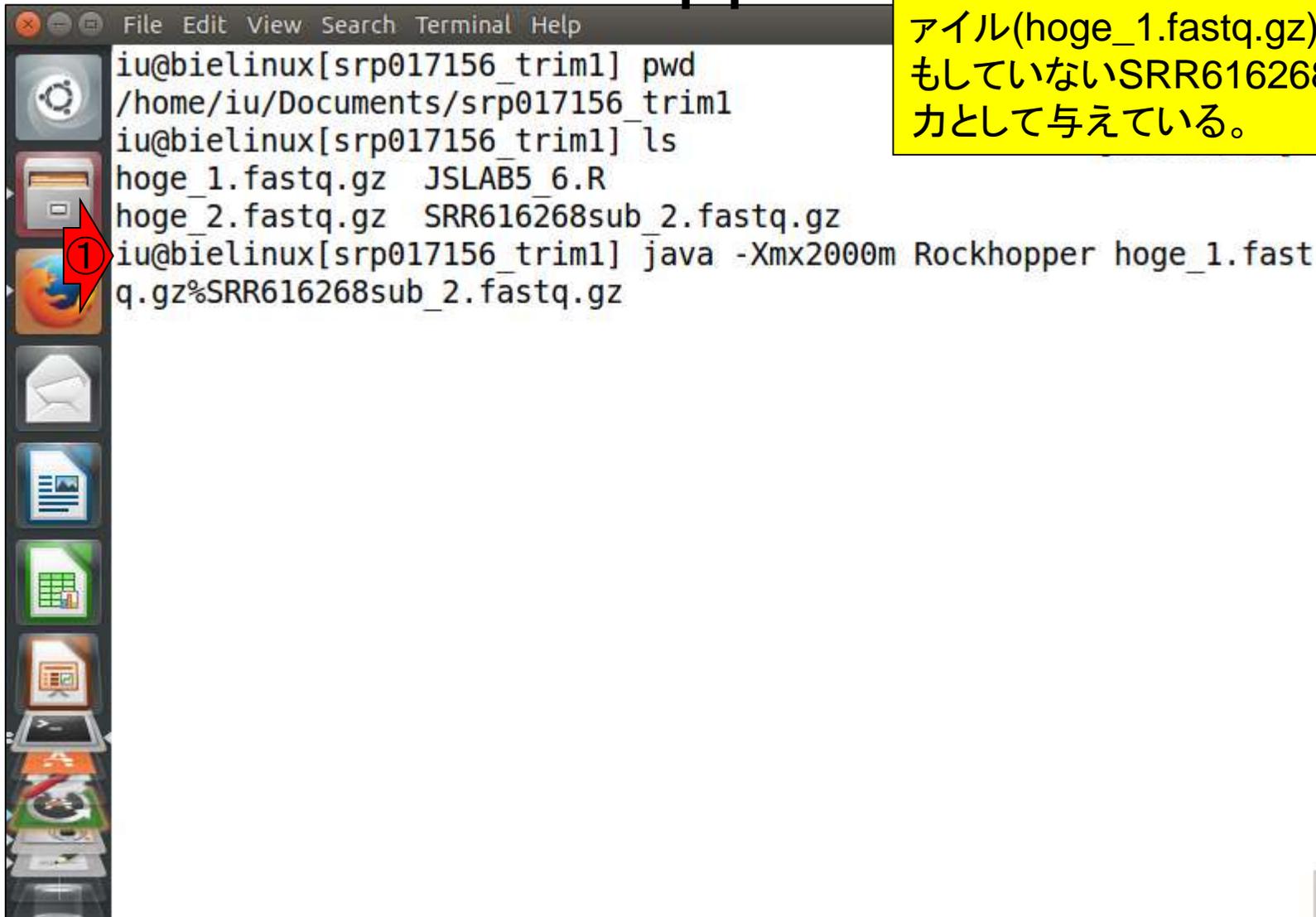
# W16-4: Tips

素朴な疑問として、よく赤下線部分の「description情報の記述が変わってないけど...」という質問が出ます。これはdescription行部分の①スペース以降の記述は任意のため、トリム用プログラムは、この赤下線部分は「ただの文字列」として取り扱います。そんなもんです。

```
iu@bielinux[srp017156_trim1] ls -l
total 201408
-rw-rw-r-- 1 iu iu 66235765 12月 22 14:36 hoge_1.fastq.gz
-rw-rw-r-- 1 iu iu 71314779 12月 22 14:52 hoge_2.fastq.gz
-rw-rw-r-- 1 iu iu      1298  9月 14 18:22 JSLAB5_6.R
-rw-rw-r-- 1 iu iu 68682959 12月 22 14:42 SRR616268sub_2.fastq.gz
iu@bielinux[srp017156_trim1] gunzip -c hoge_1.fastq.gz | tail -n 4
@SRR616268.1000860 2291:6:1101:11638:95311 length=107
GCCTTGTCAATCAAGGTGAGCATGTCGCCCATGCCAGAATTCGGTTGGCCATGCGATCAGGATAG
AAGACATCCAGCGCATCCATCTTTTCACCTTGA
+
bbbeeeeeggggggiiiefghiiiigiiihiiiihiiiiiiiiiihifhiiiiiiiiiiiiihgggggg
eeeeeddddddccccccccccccccccccccccccc
iu@bielinux[srp017156_trim1] gunzip -c hoge_2.fastq.gz | tail -n 4
@SRR616268.1000860 2291:6:1101:11638:95311 length=107
GCCTTGTCAATCAAGGTGAGCATGTCGCCCATGCCAGAATTCGGTTGGCCATGCGATCAGGATAG
AAGACATCCAGCGCATCCATCTTTTCACCTTGA
+SRR616268.1000860 2291:6:1101:11638:95311 length=107
bbbeeeeeggggggiiiefghiiiigiiihiiiihiiiiiiiiiihifhiiiiiiiiiiiiihgggggg
eeeeeddddddccccccccccccccccccccccccc
iu@bielinux[srp017156_trim1] █ [ 2:58午後 ]
```

# W17-1: Rockhopper

②Rockhopper2によるde novo transcriptome assemblyをトリム後のデータで再実行。  
forward側はRのBiostringsを用いて得られたファイル(hoge\_1.fastq.gz)、reverse側は特に何もしていないSRR616268sub\_2.fastq.gzを入力として与えている。

A terminal window with a dark background and a light-colored text. The window title bar shows 'File Edit View Search Terminal Help'. The terminal content shows a user at a prompt 'iu@bielinux[srp017156\_trim1]' performing several commands: 'pwd' (returns '/home/iu/Documents/srp017156\_trim1'), 'ls' (lists 'hoge\_1.fastq.gz', 'JSLAB5\_6.R', 'hoge\_2.fastq.gz', 'SRR616268sub\_2.fastq.gz'), and 'java -Xmx2000m Rockhopper hoge\_1.fastq.gz %SRR616268sub\_2.fastq.gz'. A red arrow with the number '1' points to the 'java' command line. The terminal window is overlaid on a desktop environment with a vertical dock on the left containing icons for a terminal, file manager, Firefox, mail, documents, spreadsheet, presentation, and a stack of folders.

```
iu@bielinux[srp017156_trim1] pwd
/home/iu/Documents/srp017156_trim1
iu@bielinux[srp017156_trim1] ls
hoge_1.fastq.gz  JSLAB5_6.R
hoge_2.fastq.gz  SRR616268sub_2.fastq.gz
iu@bielinux[srp017156_trim1] java -Xmx2000m Rockhopper hoge_1.fast
q.gz%SRR616268sub_2.fastq.gz
```

# W17-1: Rockhopper

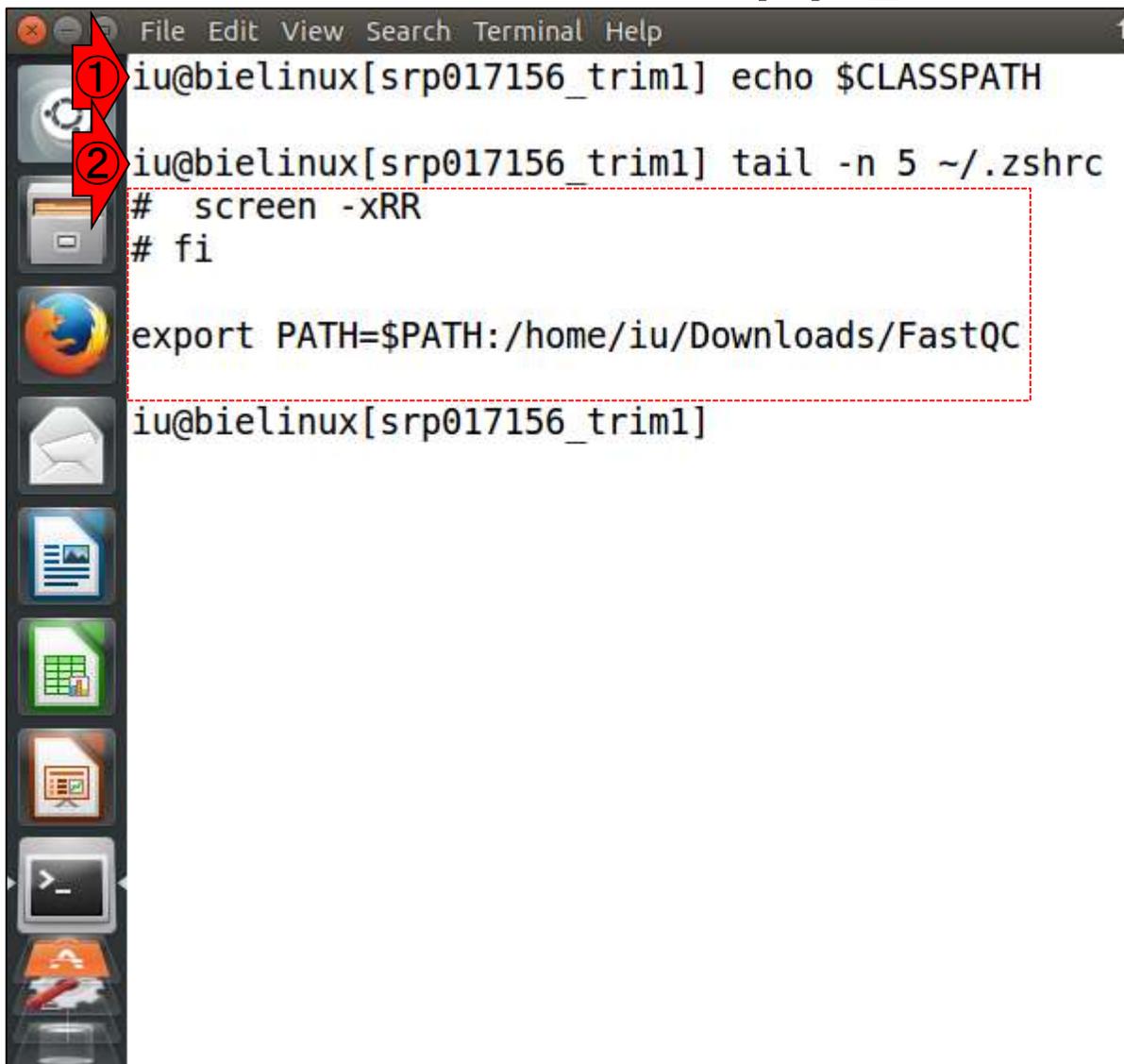
実行結果。エラーが出ていることがわかる。これは、今実行したターミナルは、クラスパスを設定したターミナル[W4]とは異なるものだから。もし同じターミナルだったら、エラーは出ない



```
iu@bielinux[srp017156_trim1] pwd [12:12午後]
/home/iu/Documents/srp017156_trim1
iu@bielinux[srp017156_trim1] ls [12:12午後]
hoge_1.fastq.gz JSLAB5_6.R
hoge_2.fastq.gz SRR616268sub_2.fastq.gz
iu@bielinux[srp017156_trim1] java -Xmx2000m Rockhopper hoge_1.fastq.gz%SRR616268sub_2.fastq.gz
Error: Could not find or load main class Rockhopper
iu@bielinux[srp017156_trim1] [12:16午後]
```

# W17-2: echoで書き込み

- ①W4で設定したクラスパスが、このターミナルでは無効になっていることを確認。環境設定の永続化は、第4回のW10-3で行った、~/.zshrcファイルへの書き込み。
- ②~/.zshrcファイルの最後の5行分を表示。これがクラスパス書き込み前の状態



```
File Edit View Search Terminal Help
iu@bielinux[serp017156_trim1] echo $CLASSPATH
iu@bielinux[serp017156_trim1] tail -n 5 ~/.zshrc
# screen -xRR
# fi
export PATH=$PATH:/home/iu/Downloads/FastQC
iu@bielinux[serp017156_trim1]
```

[ 3:16午後 ]

# W17-2: echoで書き込み

「gedit ~/.zshrc」で.zshrcファイルを編集してもよいが、せつくないので「echoで表示させた文字列をファイルに追加書き込みする」やり方を伝授。①や②で示すように、シングルクォーテーション(')で囲まれた文字列を画面右上に出力するのがecho

```
iu@bielinux[srp017156_trim1] echo $CLASSPATH
iu@bielinux[srp017156_trim1] tail -n 5 ~/.zshrc
# screen -xRR
# fi
export PATH=$PATH:/home/iu/Downloads/FastQC
① iu@bielinux[srp017156_trim1] echo 'hoge' [ 3:16午後 ]
hoge
② iu@bielinux[srp017156_trim1] echo 'export CLASSPATH=/home/iu/Downl
oads/Rockhopper.jar'
export CLASSPATH=/home/iu/Downloads/Rockhopper.jar
iu@bielinux[srp017156_trim1] [ 3:18午後 ]
```

# W17-2: >>で追加書き込み

①echoで表示させた、~/.zshrcファイルの最後に書き込みたい内容を「>>」で追加書き込み。「>」では追加ではなく上書きになってしまうので注意!「cp ~/.zshrc ~/.zshrc\_org」などとしてバックアップファイルを作成しておくほうがいいかもしれない。②追加書き込み後にtailコマンドで最後の5行分を再表示。追加書き込みが正常終了。

```
iu@bielinux[srp017156_trim1] echo $CLASSPATH
iu@bielinux[srp017156_trim1] tail -n 5 ~/.zshrc
# screen -xRR
# fi
export PATH=$PATH:/home/iu/Downloads/FastQC
iu@bielinux[srp017156_trim1] echo 'hoge'
hoge
iu@bielinux[srp017156_trim1] echo 'export CLASSPATH=/home/iu/Downl
oads/Rockhopper.jar'
export CLASSPATH=/home/iu/Downloads/Rockhopper.jar
① iu@bielinux[srp017156_trim1] echo 'export CLASSPATH=/home/iu/Downl
oads/Rockhopper.jar' >> ~/.zshrc
② iu@bielinux[srp017156_trim1] tail -n 5 ~/.zshrc
# fi
export PATH=$PATH:/home/iu/Downloads/FastQC
export CLASSPATH=/home/iu/Downloads/Rockhopper.jar
iu@bielinux[srp017156_trim1] █
```

[ 3:16午後 ]  
[ 3:19午後 ]  
[ 3:19午後 ]

# W17-3: sourceして確認

```
File Edit View Search Terminal Help
① iu@bielinux[srp017156_trim1] echo $CLASSPATH
② iu@bielinux[srp017156_trim1] source ~/.zshrc
iu@bielinux[srp017156_trim1] echo $CLASSPATH
/home/iu/Downloads/Rockhopper.jar
iu@bielinux[srp017156_trim1]
```

ただの復習(第4回のW10-4)。~/.zshrcにきちんと書き込みできていたとしても②source関数を実行して環境設定ファイル(~/.zshrc)のリロードを行わなければいけない。①リロード前と③リロード後で「echo \$CLASSPATH」実行結果が異なっていることがわかる。

[ 3:26午後 ]

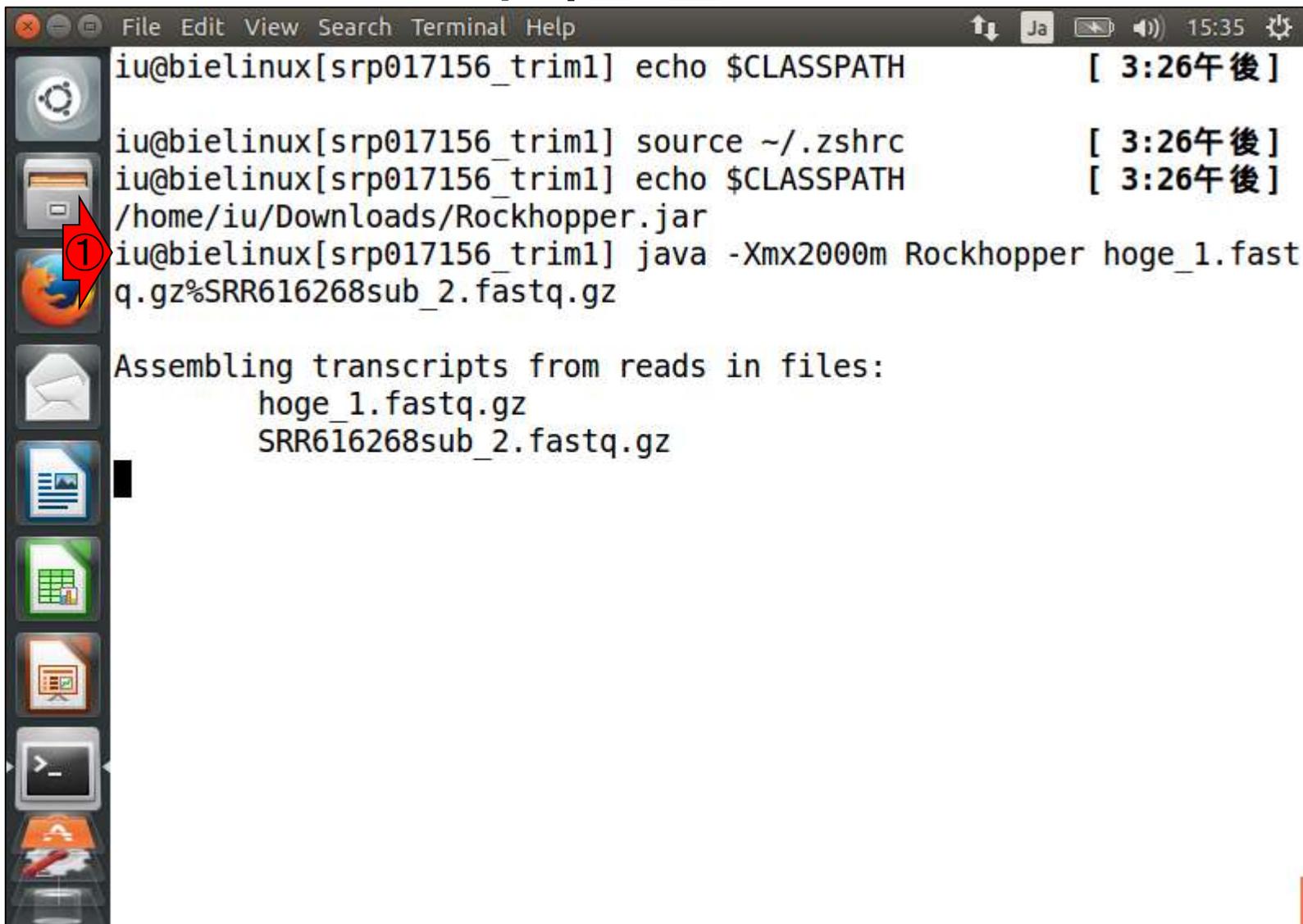
# W17-4: Rockhopper

Rockhopper2を再々トライ。RのBiostrings  
を利用したファイルhoge\_1.fastq.gzを  
forward側として入力する場合。約2分

```
iu@bielinux[srp017156_trim1] echo $CLASSPATH [ 3:26午後 ]
iu@bielinux[srp017156_trim1] source ~/.zshrc [ 3:26午後 ]
iu@bielinux[srp017156_trim1] echo $CLASSPATH [ 3:26午後 ]
/home/iu/Downloads/Rockhopper.jar
iu@bielinux[srp017156_trim1] java -Xmx2000m Rockhopper hoge_1.fast
q.gz%SRR616268sub_2.fastq.gz
```



# W17-4: 途中経過



```
iu@bielinux[srp017156_trim1] echo $CLASSPATH [ 3:26午後 ]
iu@bielinux[srp017156_trim1] source ~/.zshrc [ 3:26午後 ]
iu@bielinux[srp017156_trim1] echo $CLASSPATH [ 3:26午後 ]
/home/iu/Downloads/Rockhopper.jar
iu@bielinux[srp017156_trim1] java -Xmx2000m Rockhopper hoge_1.fast
q.gz%SRR616268sub_2.fastq.gz
Assembling transcripts from reads in files:
    hoge_1.fastq.gz
    SRR616268sub_2.fastq.gz
```

# W17-4: 実行結果

トリム前の無残な結果[W5-2]やreverse側のsingle-endのみの結果[W6-4]と比べても、①転写物数(794 transcripts)や②総塩基数(449,115 bases)の点で劇的にアセンブルが改善されたことがわかる!

```
iu@bielinux[~/Documents/srp017156_trim1]
hoge_1.fastq.gz
SRR616268sub_2.fastq.gz

Aligning reads to assembled transcripts using files:
hoge_1.fastq.gz
SRR616268sub_2.fastq.gz

Total reads in files:          987886
Perfectly aligned reads:      579770  59%

Total number of assembled transcripts:  794 ①
Average transcript length:             565
Median transcript length:              306
Total number of assembled bases:      449115 ②

Summary of results written to file:      Rockhopper
_Results/summary.txt
Details of assembled transcripts written to file:  Rockhopper
_Results/transcripts.txt

FINISHED.

iu@bielinux[srp017156_trim1] [ 3:36午後 ]
```

# W17-5: Rockhopper

①fastx\_trimmerでの実行結果ファイル(hoge\_2.fastq.gz)を入力として、念のため実行。②

Rockhopper\_Resultsディレクトリ中の以前の実行結果ファイルは上書きされてなくなるので注意！  
様々なオプションや入力ファイルの結果を保存したい場合は「summary.txtやtranscripts.txt」のファイル名をその都度変更しておく。ここは同じ結果になることを確認するだけなので気にしない

```
File Edit View Search Terminal Help
Total reads in files:          987886
Perfectly aligned reads:      579770  59%

Total number of assembled transcripts: 794
Average transcript length:      565
Median transcript length:       306
Total number of assembled bases: 449115

Summary of results written to file:
_Rockhopper_Results/summary.txt
Details of assembled transcripts written to file:
_Rockhopper_Results/transcripts.txt

FINISHED.

iu@bielinux[srp017156_trim1] ls [ 3:36午後 ]
hoge_1.fastq.gz  JSLAB5_6.R          SRR616268sub_2.fastq.gz
hoge_2.fastq.gz  Rockhopper_Results

iu@bielinux[srp017156_trim1] ls Rockhopper_Results [ 3:42午後 ]
genomeBrowserFiles intermediary summary.txt transcripts.txt

iu@bielinux[srp017156_trim1] java -Xmx2000m Rockhopper hoge_2.fast
q.gz%SRR616268sub_2.fastq.gz
```

②

①

# W17-5: Rockhopper

確かに同じ結果になった！2つのトリミングプログラムともに正しく動作していることも、ポジティブなアセンブル結果から証明されたといえる。

```
File Edit View Search Terminal Help
hoge_2.fastq.gz
SRR616268sub_2.fastq.gz

Aligning reads to assembled transcripts using files:
hoge_2.fastq.gz
SRR616268sub_2.fastq.gz

Total reads in files:          987886
Perfectly aligned reads:      579770  59%

Total number of assembled transcripts:  794
Average transcript length:             565
Median transcript length:              306
Total number of assembled bases:      449115

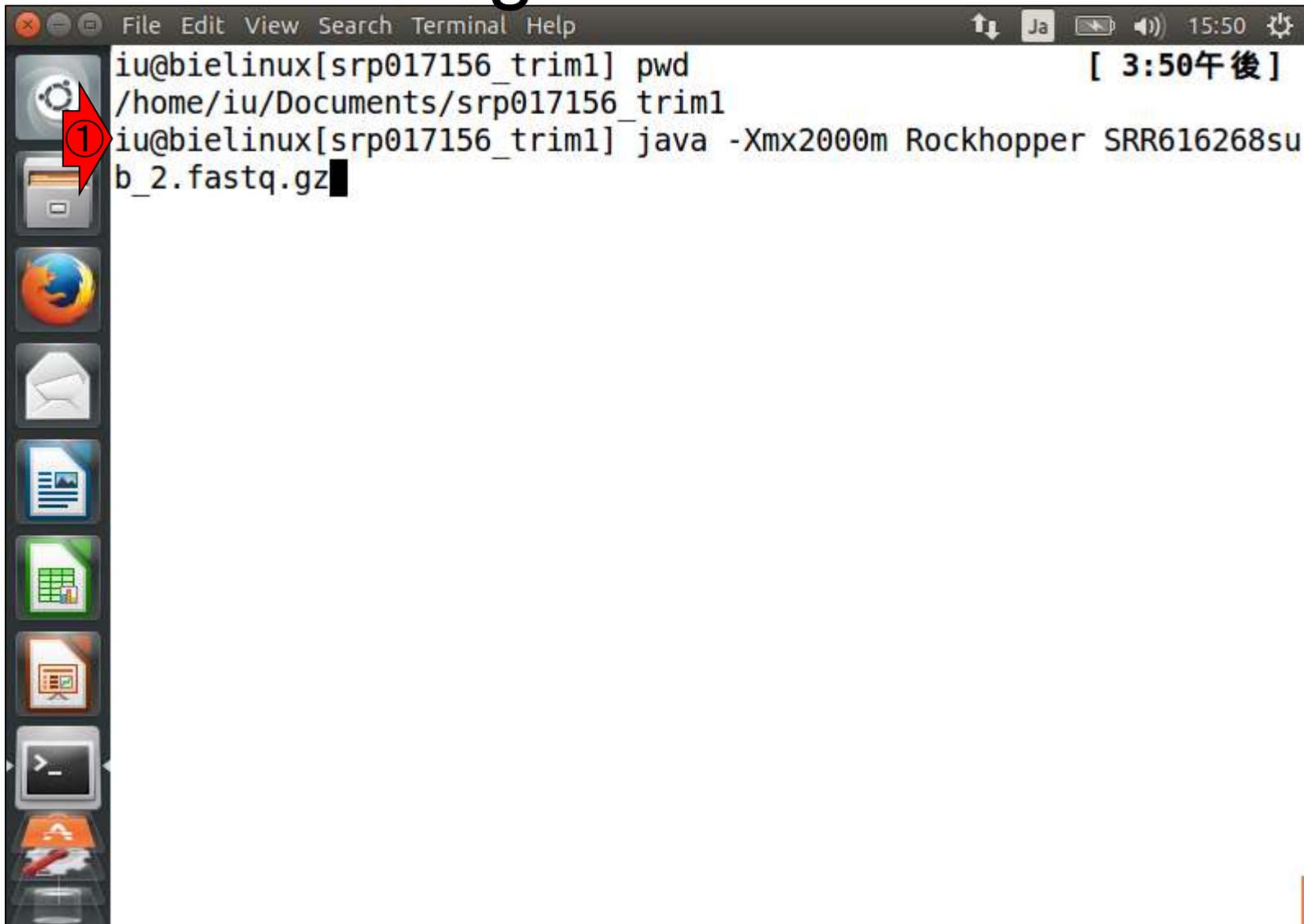
Summary of results written to file:      Rockhopper
_Results/summary.txt
Details of assembled transcripts written to file:  Rockhopper
_Results/transcripts.txt

FINISHED.

iu@bielinux[srp017156_trim1] [ 3:44午後 ]
```

# W17-6: single-end

比較用に①何もしていないreverse側のsingle-endのみ(SRR616268sub\_2.fastq.gz)で実行



A terminal window titled "Terminal" with a menu bar (File, Edit, View, Search, Terminal, Help) and a system tray (Ja, 15:50). The terminal shows the following commands and output:

```
iu@bielinux[srp017156_trim1] pwd [ 3:50午後 ]
/home/iu/Documents/srp017156_trim1
iu@bielinux[srp017156_trim1] java -Xmx2000m Rockhopper SRR616268sub_2.fastq.gz
```

A red arrow with the number "1" points to the second command line.

# W17-6: single-end

①アセンブルされた転写物数は424個、②総塩基数は185,233。③入力リード数983,854個のうち、72% (710,393個)がマップされていることがわかる。FaQCs実行後のファイルを入力とした結果(W6-4)よりもわずかによい結果といえるかもしれないが、事実上誤差範囲。

```
File Edit View Search Terminal Help
Assembling transcripts from reads in file:
ub_2.fastq.gz
Aligning reads to assembled transcripts using file:      SRR616268s
ub_2.fastq.gz
Total reads in file:          983854
Perfectly aligned reads:      710393  72%
Total number of assembled transcripts: 424
Average transcript length:    436
Median transcript length:     228
Total number of assembled bases: 185233
Summary of results written to file:      Rockhopper
_Results/summary.txt
Details of assembled transcripts written to file:  Rockhopper
_Results/transcripts.txt
FINISHED.
iu@bielinux[srp017156_trim1] [ 3:51午後 ]
```



# W18-1: QuasR

①トリム後のデータでマッピングを再実行すべく、リストファイル(JSLAB5\_7.txt)をダウンロード。forward側はhoge\_1.fastq.gz、reverse側は特に何もしていないSRR616268sub\_2.fastq.gz。②moreで中身を表示。③比較のため、特に何もしていないforward側のファイルでのマッピングも行っている(pre\_7bp\_trimの行に相当)

```
iu@bielinux[srp017156_trim1] pwd
/home/iu/Documents/srp017156_trim1
iu@bielinux[srp017156_trim1] ls
hoge_1.fastq.gz  JSLAB5_6.R          SRR616268sub_2.fastq.gz
hoge_2.fastq.gz  Rockhopper_Results
iu@bielinux[srp017156_trim1] wget -cq http://www.iu.a.u-tokyo.ac.jp/~kadota/book/JSLAB5_7.txt
iu@bielinux[srp017156_trim1] more JSLAB5_7.txt  [ 4:00午後 ]
FileName1      FileName2      SampleName
SRR616268sub_1.fastq.gz SRR616268sub_2.fastq.gz pre_7bp_trim
hoge_1.fastq.gz SRR616268sub_2.fastq.gz post_7bp_trim
iu@bielinux[srp017156_trim1] [ 4:00午後 ]
```



①Rスクリプトファイル(JSLAB5\_8.R)をダウンロードし、②最初の2行分を表示

# W18-2: QuasR

```
iu@bielinux[srp017156_trim1] pwd [ 4:00午後 ]
/home/iu/Documents/srp017156_trim1
iu@bielinux[srp017156_trim1] ls [ 4:00午後 ]
hoge_1.fastq.gz JSLAB5_6.R SRR616268sub_2.fastq.gz
hoge_2.fastq.gz Rockhopper_Results
iu@bielinux[srp017156_trim1] wget -cq http://www.iu.a.u-tokyo.ac.jp/~kadota/book/JSLAB5_7.txt
iu@bielinux[srp017156_trim1] more JSLAB5_7.txt [ 4:00午後 ]
FileName1 FileName2 SampleName
SRR616268sub_1.fastq.gz SRR616268sub_2.fastq.gz pre_7bp_trim
hoge_1.fastq.gz SRR616268sub_2.fastq.gz post_7bp_trim
① iu@bielinux[srp017156_trim1] wget -cq http://www.iu.a.u-tokyo.ac.jp/~kadota/book/JSLAB5_8.R
② iu@bielinux[srp017156_trim1] nkf JSLAB5_8.R | head -n 2 -
in_f1 <- "JSLAB5_7.txt" #入力ファイル名を指定してin_f1に格納(RNA-seqリストファイル)
in_f2 <- "/home/iu/Documents/genomes/Lactobacillus_casei_12a.GCA_00309565.2.30.dna.toplevel.fa"#入力ファイル名を指定してin_f2に格納(リファレンス配列)
iu@bielinux[srp017156_trim1] [ 4:09午後 ]
```

①比較用入力ファイル(SRR616268sub\_1.fastq.gz)のコピーと②確認

# W18-3: QuasR

```
iu@bielinux[~/Documents/srp017156_trim1]
iu@bielinux[srp017156_trim1] pwd [ 4:14午後 ]
/home/iu/Documents/srp017156_trim1
iu@bielinux[srp017156_trim1] ls [ 4:14午後 ]
hoge_1.fastq.gz JSLAB5_7.txt SRR616268sub_2.fastq.gz
hoge_2.fastq.gz JSLAB5_8.R
JSLAB5_6.R Rockhopper_Results
① iu@bielinux[srp017156_trim1] cp ~/Documents/srp017156/SRR616268sub
_1.fastq.gz .
② iu@bielinux[srp017156_trim1] ls [ 4:14午後 ]
hoge_1.fastq.gz JSLAB5_7.txt SRR616268sub_1.fastq.gz
hoge_2.fastq.gz JSLAB5_8.R SRR616268sub_2.fastq.gz
JSLAB5_6.R Rockhopper_Results
iu@bielinux[srp017156_trim1] [ 4:14午後 ]
```

# W18-4: QuasR

①Rスクリプトファイル(JSLAB5\_8.R)の  
実行。約13分。

```
iu@bielinux[srp017156_trim1] pwd [ 4:14午後 ]
/home/iu/Documents/srp017156_trim1
iu@bielinux[srp017156_trim1] ls [ 4:14午後 ]
hoge_1.fastq.gz JSLAB5_7.txt SRR616268sub_2.fastq.gz
hoge_2.fastq.gz JSLAB5_8.R
JSLAB5_6.R Rockhopper_Results
iu@bielinux[srp017156_trim1] cp ~/Documents/srp017156/SRR616268sub
_1.fastq.gz .
iu@bielinux[srp017156_trim1] ls [ 4:14午後 ]
hoge_1.fastq.gz JSLAB5_7.txt SRR616268sub_1.fastq.gz
hoge_2.fastq.gz JSLAB5_8.R SRR616268sub_2.fastq.gz
JSLAB5_6.R Rockhopper_Results
① iu@bielinux[srp017156_trim1] R --vanilla --slave < JSLAB5_8.R
```

# W18-4: QuasR

無事終了。①lsで確認。bamファイルや②QCレポートファイルが作成されていることがわかる

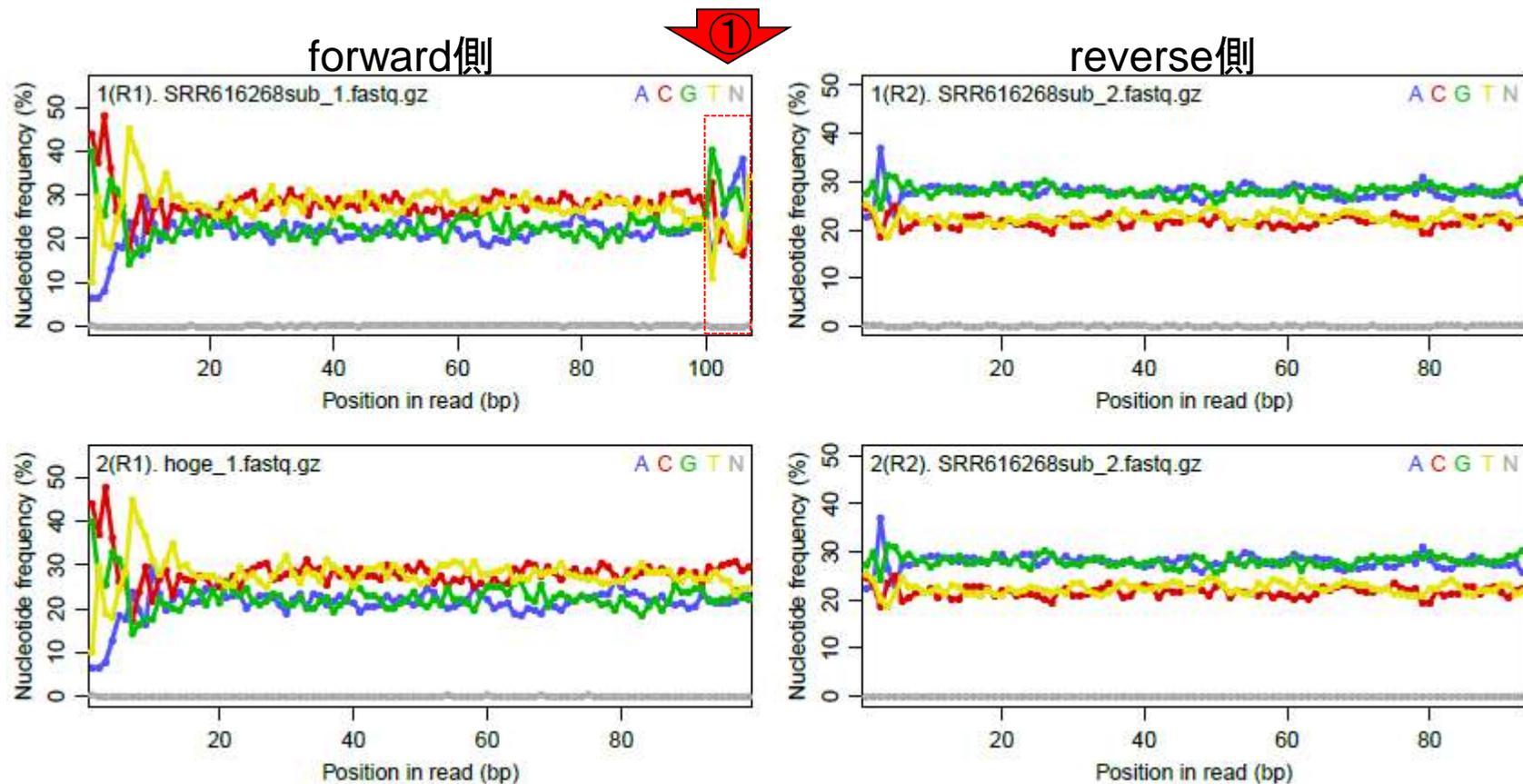
```
bielinux
1
Performing genomic alignments for 2 samples. See progress in the log file:
/home/iu/Documents/srp017156_trim1/QuasR_log_663c661fd224.txt
[samopen] SAM header is present: 1 sequences.
[bam_sort_core] merging from 2 files...
[samopen] SAM header is present: 1 sequences.
[bam_sort_core] merging from 2 files...
Genomic alignments have been created successfully

collecting quality control data
creating QC plots
iu@bielinux[srp017156_trim1] ls [ 4:30午後 ]
hoge_1_663c66aaf313.bam QuasR_log_663c661fd224.txt
hoge_1_663c66aaf313.bam.bai Rockhopper_Results
hoge_1_663c66aaf313.bam.txt SRR616268sub_1_663c2ebdd882.bam
hoge_1.fastq.gz SRR616268sub_1_663c2ebdd882.bam.bai
hoge_2.fastq.gz SRR616268sub_1_663c2ebdd882.bam.txt
JSLAB5_6.R SRR616268sub_1_663c2ebdd882_QC.pdf
JSLAB5_7.txt SRR616268sub_1.fastq.gz
JSLAB5_8.R SRR616268sub_2.fastq.gz
iu@bielinux[srp017156_trim1] cp *.pdf ~/Desktop/mac_share
```



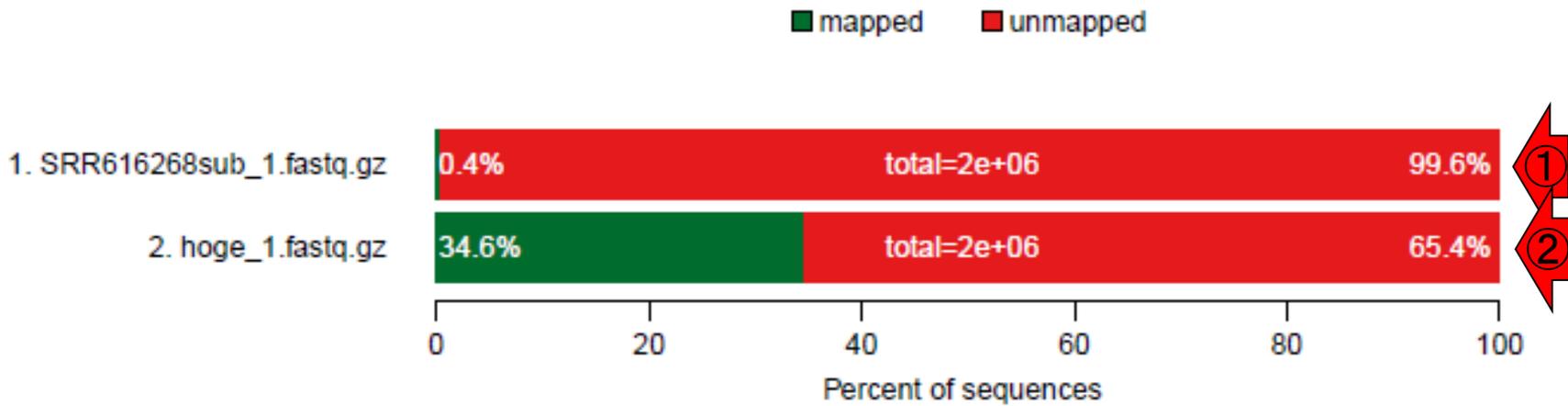
# W18-5: PDF解説

PDF2枚目。ポジションごとの塩基の出現確率。FastQC Report中の項目「Per base sequence content」と同じ(但し色は異なる)。赤枠部分がトリムされたおかげでアセンブルやマッピングが劇的に改善したことになる。



# W18-6: PDF解説

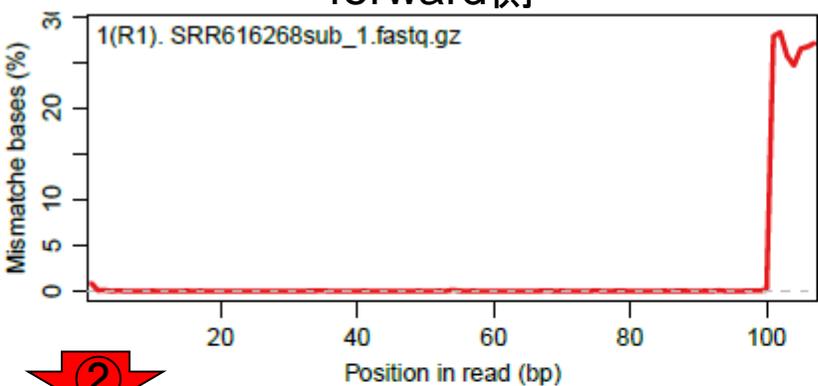
PDF4枚目。全リード(forward, reverse合わせて200万リード)のうち、マップされたリードの割合は①トリム実行前が0.4%、②実行後が34.6%。トリム後のマップ率が劇的に向上。



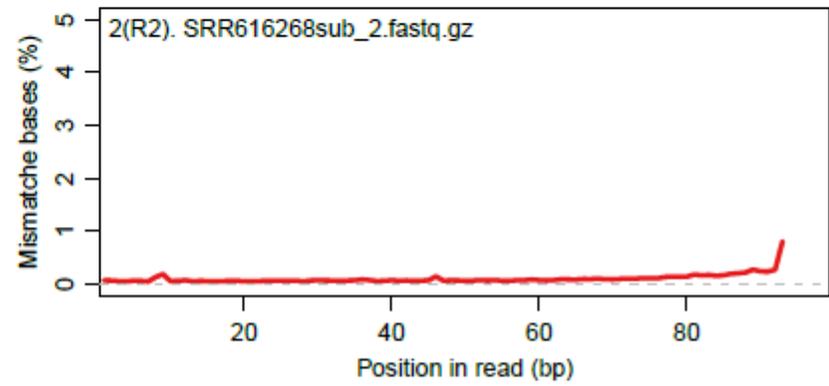
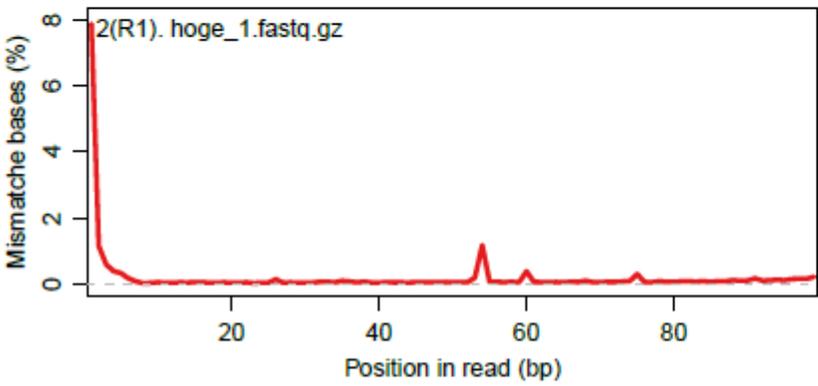
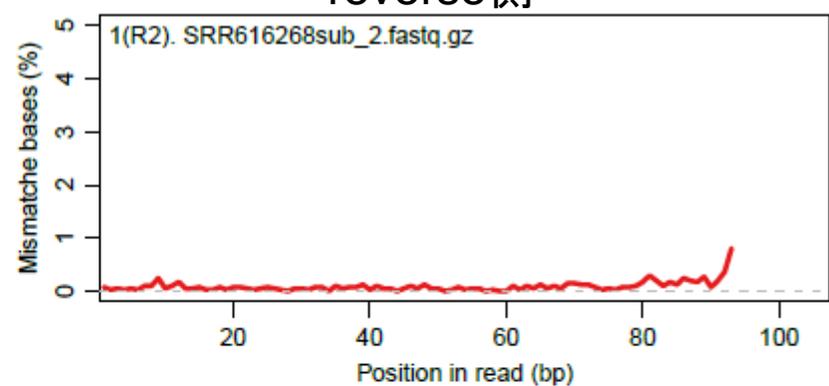
# W18-7: PDF解説

PDF6枚目。Forward側の100-107 bpをトリムしたおかげで①のミスマッチ塩基の割合が劇的に低下していることがわかる。そのおかげで、相対的なインパクトが弱かった② forward側の1塩基目あたりもミスマッチ率が高かったことがわかる。

forward側



reverse側



# W19-1: FastQC

- ① --nogroupオプションをつけてFastQCを実行。
- ②赤枠が出力ファイル。③htmlレポートのファイル名をSRR616268sub\_1\_nogroup.htmlに変更。

```
iu@bielinux[srp017156_trim1] pwd [ 4:49午後 ]
/home/iu/Documents/srp017156_trim1
iu@bielinux[srp017156_trim1] ls SRR616268sub_1* [ 4:50午後 ]
SRR616268sub_1_663c2ebdd882.bam
SRR616268sub_1_663c2ebdd882.bam.bai
SRR616268sub_1_663c2ebdd882.bam.txt
SRR616268sub_1_663c2ebdd882_QC.pdf
SRR616268sub_1.fastq.gz
iu@bielinux[srp017156_trim1] fastqc2 -v [ 4:50午後 ]
FastQC v0.11.4
① iu@bielinux[srp017156_trim1] fastqc2 -q --nogroup SRR616268sub_1.f
astq.gz
iu@bielinux[srp017156_trim1] ls SRR616268sub_1* [ 4:50午後 ]
SRR616268sub_1_663c2ebdd882.bam SRR616268sub_1_fastqc.html ②
SRR616268sub_1_663c2ebdd882.bam.bai SRR616268sub_1_fastqc.zip
SRR616268sub_1_663c2ebdd882.bam.txt SRR616268sub_1.fastq.gz
SRR616268sub_1_663c2ebdd882_QC.pdf
③ iu@bielinux[srp017156_trim1] mv SRR616268sub_1_fastqc.html SRR6162
68sub_1_nogroup.html
iu@bielinux[srp017156_trim1] cp SRR616268sub_1_nogroup.html ~/Desk
top/mac_share
iu@bielinux[srp017156_trim1] [ 4:55午後 ]
```

# W19-2: FastQC

①SRR616268sub\_1\_nogroup.htmlのKmer Content項目を表示。1-59塩基目には極端に多いk-merの上位6個は存在しないことがわかる

## FastQC Report

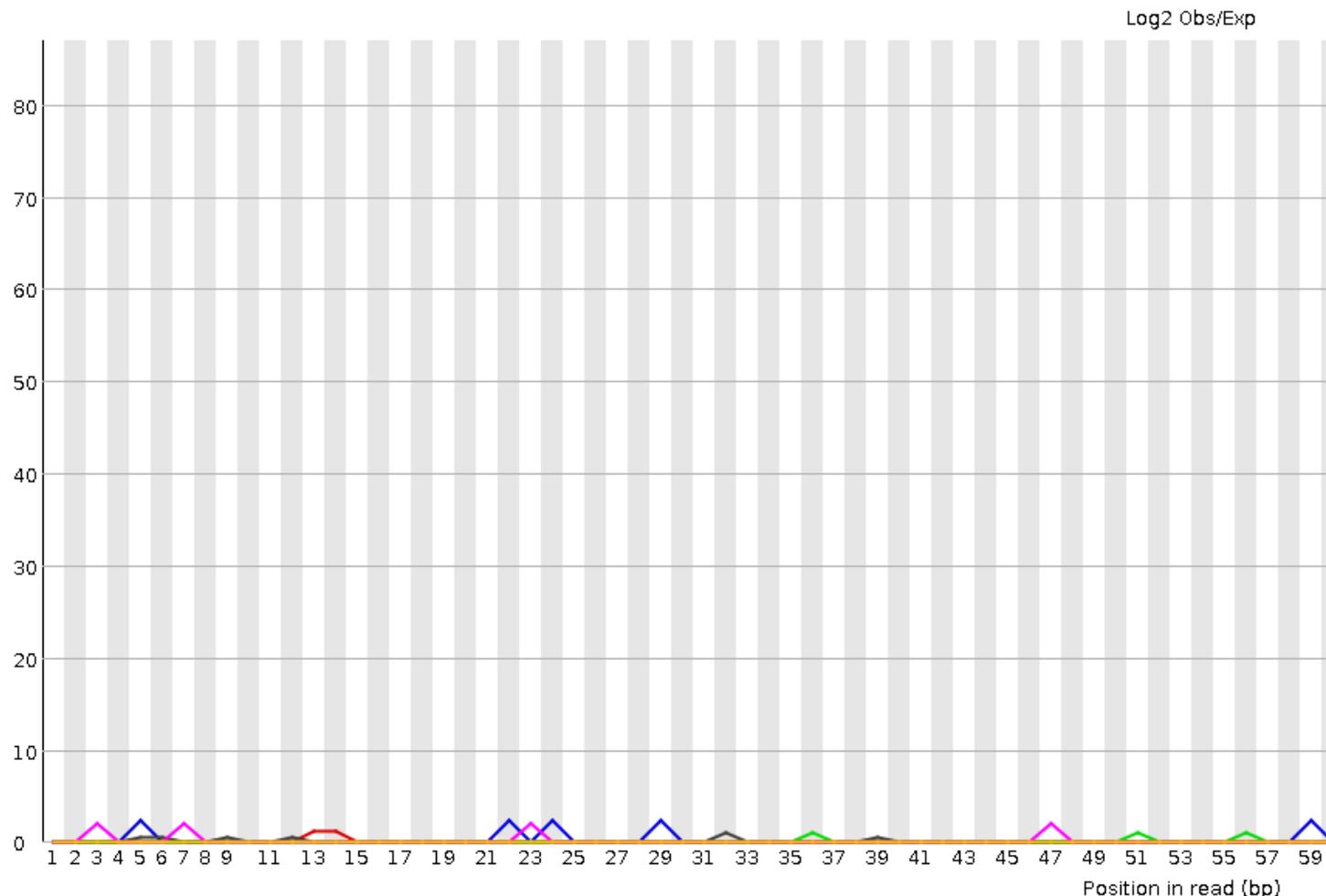
Wed 16 Sep 2015  
SRR616268sub\_1.fastq.gz

### 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

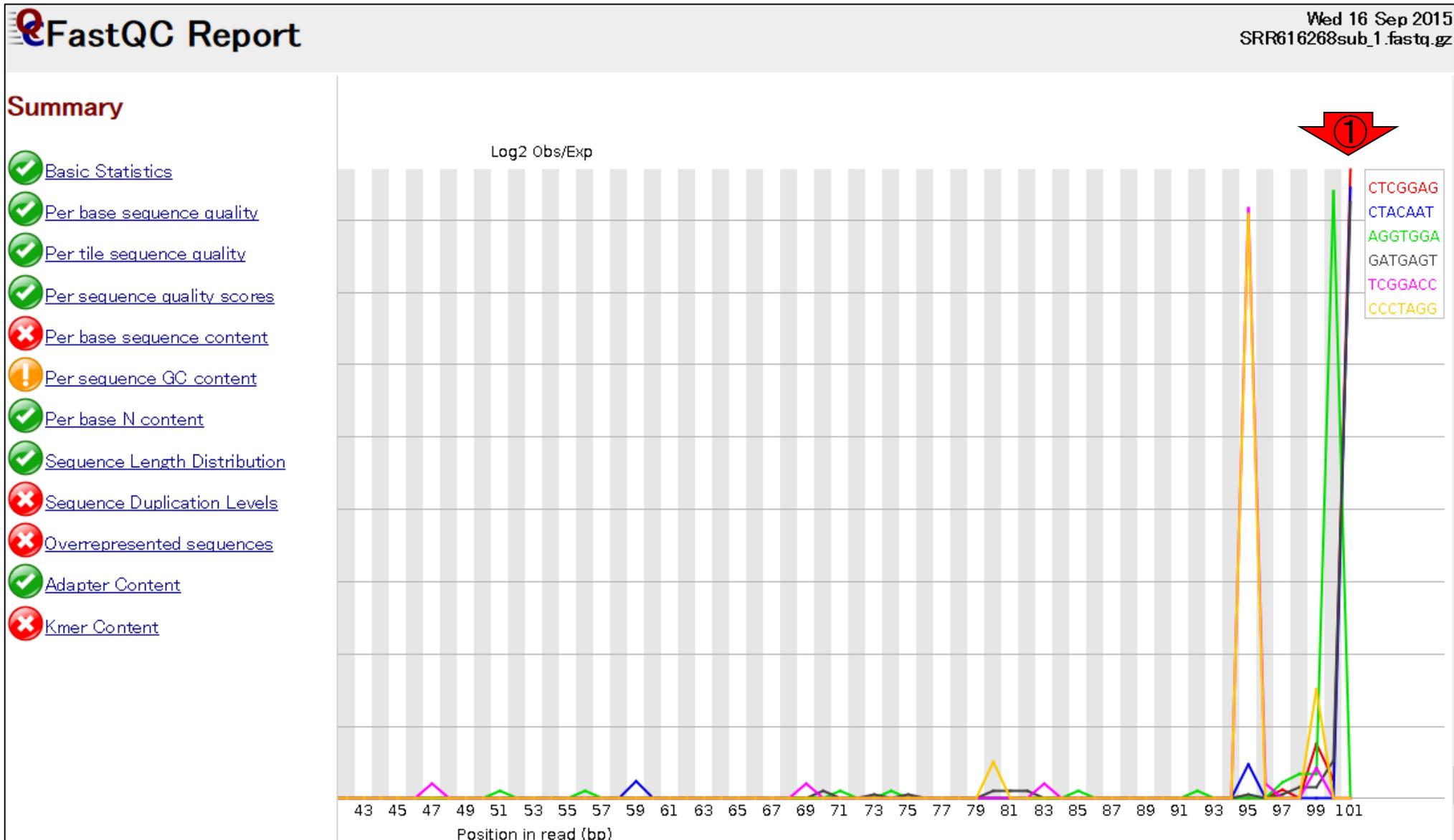


### ✗ Kmer Content



# W19-2: FastQC

①リードの右側(3'側)を表示。極端に多いk-merの上位6個が右側に偏って存在することがわかる。



# W19-2: FastQC

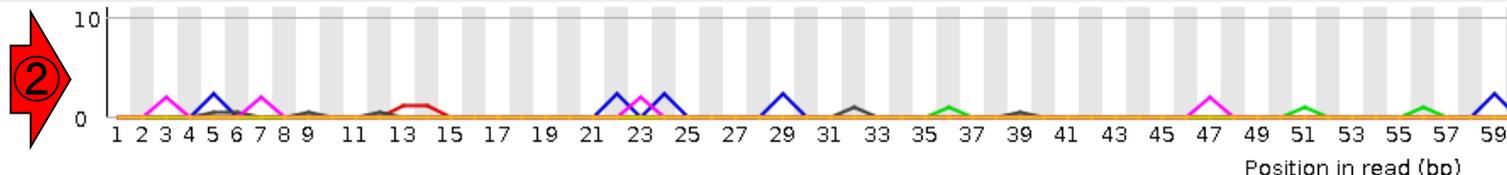
① Kmer Content項目のちょっと下のほうを表示。  
 ②上の折れ線グラフは、③赤枠で示す観測値/期待値が大きい上位6個をプロットしたもの

## FastQC Report

Wed 16 Sep 2015  
 SRR616268sub\_1.fastq.gz

### 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	Count	PValue	Obs/Exp Max	Max Obs/Exp Position
CTCGGAG	395	0.0	86.890045	101
CTACAAT	215	0.0	84.51276	101
AGGTGGA	445	0.0	83.98286	100
GATGAGT	960	0.0	82.54421	101
TCGGACC	235	0.0	81.66475	95
CCCTAGG	100	0.0	80.80512	95
CCTAGGT	45	1.4897523E-9	78.56054	96
GTTGTGG	965	0.0	75.840096	101
GGCCCTG	180	0.0	75.75481	100
TACTACA	20	0.0016262057	75.75481	96
TAGGTGG	225	0.0	74.07136	99
GTTCTCT	75	0.0	74.02692	101
CGGGCCT	235	0.0	71.00822	1
TCCCTCG	50	3.430614E-9	70.70448	98
CCCACAC	440	0.0	70.0158	95
TACTTAC	405	0.0	69.5848	95



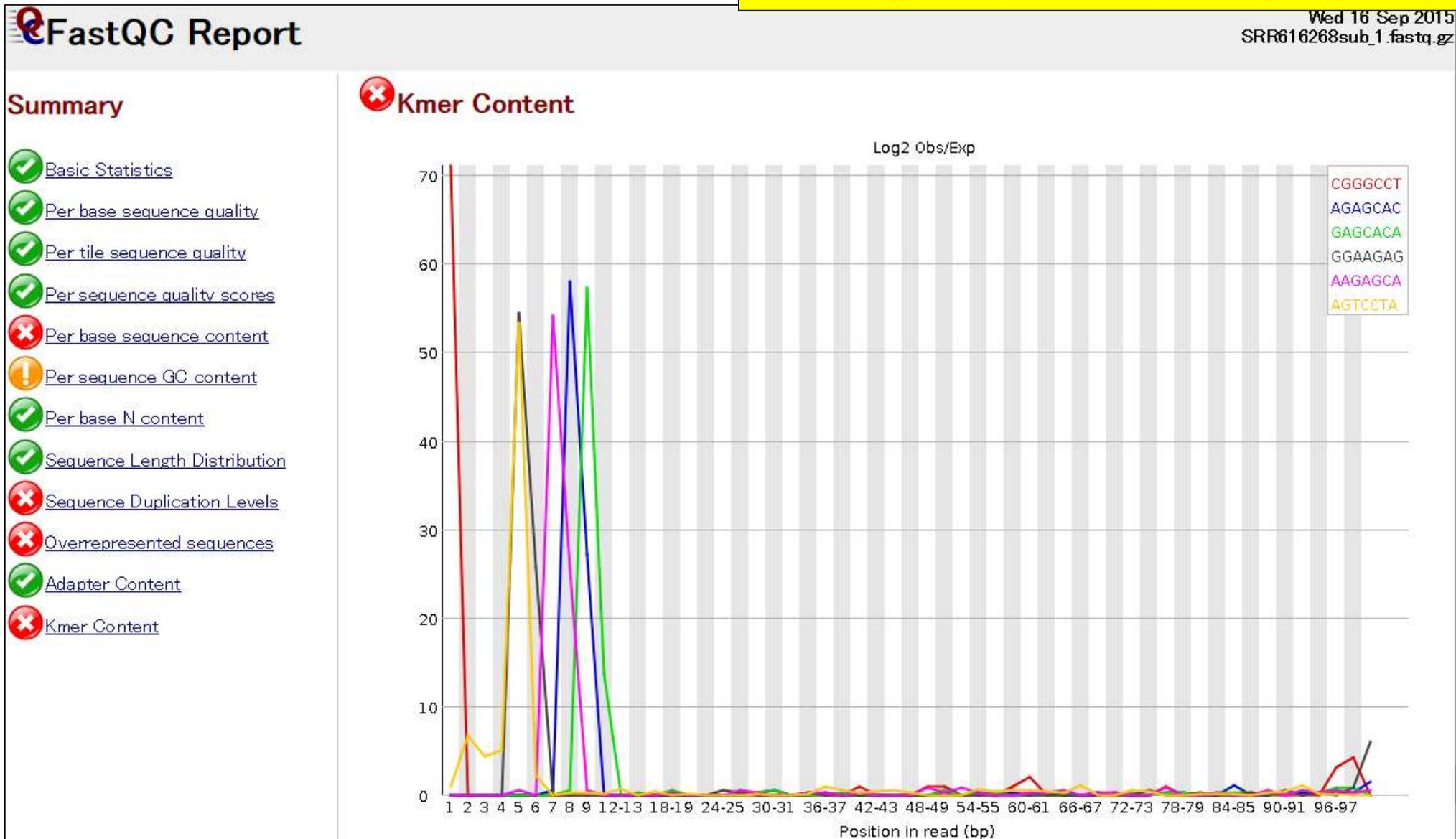
# W19-3: FastQC

①デフォルトでFastQCを実行。②htmlレポートのファイル名をSRR616268sub\_1\_default.htmlに変更

```
iu@bielinux[srp017156_trim1] pwd [ 5:09午後 ]
/home/iu/Documents/srp017156_trim1
iu@bielinux[srp017156_trim1] ls SRR616268sub_1* [ 5:09午後 ]
SRR616268sub_1_663c2ebdd882.bam SRR616268sub_1_fastqc.zip
SRR616268sub_1_663c2ebdd882.bam.bai SRR616268sub_1_fastq.gz
SRR616268sub_1_663c2ebdd882.bam.txt SRR616268sub_1_nogroup.html
SRR616268sub_1_663c2ebdd882_QC.pdf
① iu@bielinux[srp017156_trim1] fastqc2 -q SRR616268sub_1.fastq.gz
iu@bielinux[srp017156_trim1] mv SRR616268sub_1_fastqc.html SRR616268sub_1_default.html ②
iu@bielinux[srp017156_trim1] ls SRR616268sub_1* [ 5:09午後 ]
SRR616268sub_1_663c2ebdd882.bam SRR616268sub_1_default.html
SRR616268sub_1_663c2ebdd882.bam.bai SRR616268sub_1_fastqc.zip
SRR616268sub_1_663c2ebdd882.bam.txt SRR616268sub_1_fastq.gz
SRR616268sub_1_663c2ebdd882_QC.pdf SRR616268sub_1_nogroup.html
iu@bielinux[srp017156_trim1] cp SRR616268sub_1_default.html ~/Desktop/mac_share
iu@bielinux[srp017156_trim1] [ 5:09午後 ]
```

# W19-4: FastQC

①SRR616268sub\_1\_default.htmlのKmer Content項目を表示。極端に多いk-merの上位6個が左側(5'側)に偏って存在していることがわかる。



# W19-4: FastQC

① Kmer Content項目のちょっと下のほうを表示。  
 ②上の折れ線グラフは、③赤枠で示す観測値/期待値が大きい上位6個をプロットしたもの

