

もし「ネットから取り込んだアプリは実行できないよ」と言われた場合、

Mac - インターネットからダウンロードしたアプリの実行許可

<https://pc-karuma.net/download-app-security/>

Mac - すべてのアプリケーションの実行を許可 - PC設定のカルマ

<https://pc-karuma.net/mac-os-sierra-allow-apps-from-anywhere/>

①XXXX.appをクリックして起動する場合には、クリックの代わりにShiftを押しながらクリック確認のウィンドウが出るので、「開く」をクリック。次からはクリックだけで機能できる。

または

- ②ターミナル画面でコマンドラインにコマンドを打ち込んで起動する場合には、
- アップルメニューの「システム環境設定」⇒「セキュリティとプライバシー」⇒「一般」タブ
 - ⇒ 下段の「変更するにはカギをクリック」 (パスワード要求)
 - ⇒ 「ダウンロードした。。。」の「確認済みの。。。」をチェック
 - ⇒ 下段の「変更できないようにするにはカギをクリック」をクリック (元へ戻す)
 - ⇒ ターミナルのコマンドラインから

```
Rsudo spctl --master-disable
```

これでコマンドラインから起動できる。次からはコマンドを打ち込むだけで起動できる

- これで アプリケーションフォルダのFastQC.app をクリックして起動
 - 上部の「File」タブで、対象ファイルを指定する
 - この時は、--nogroup は指定できない (どう指定するか分からない)
- 又は、ターミナル (コマンドライン) からだと、作業ディレクトリBreseq-Seminarの中で、

```
Applications/FastQC.app/Contents/MacOS/FastQC --nogroup Anc_R1_head.fastq.gz
```

なお、Linuxでは

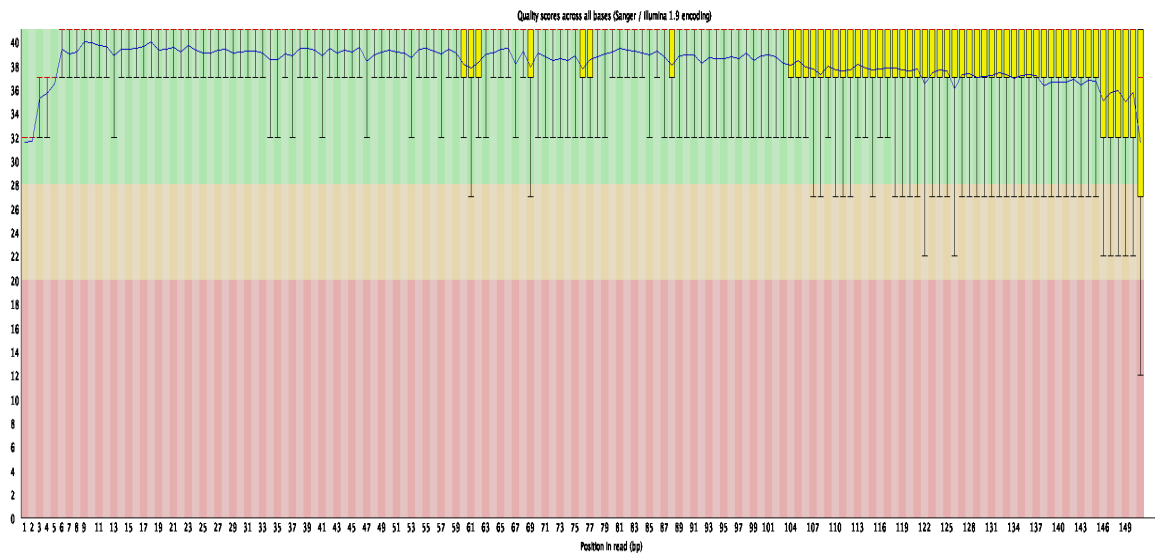
```
fastqc --nogroup Anc_R1_head.fastq.gz
```

- groupは、3'側のベースを、リードをまとめて表示する機能だが、短ければ画面に入るので、個別塩基で表示する--nogroup指定をした方が良さらしい
 - 出力は、同じディレクトリ内に Anc_R1_head_fastqc.html としてできている。
-

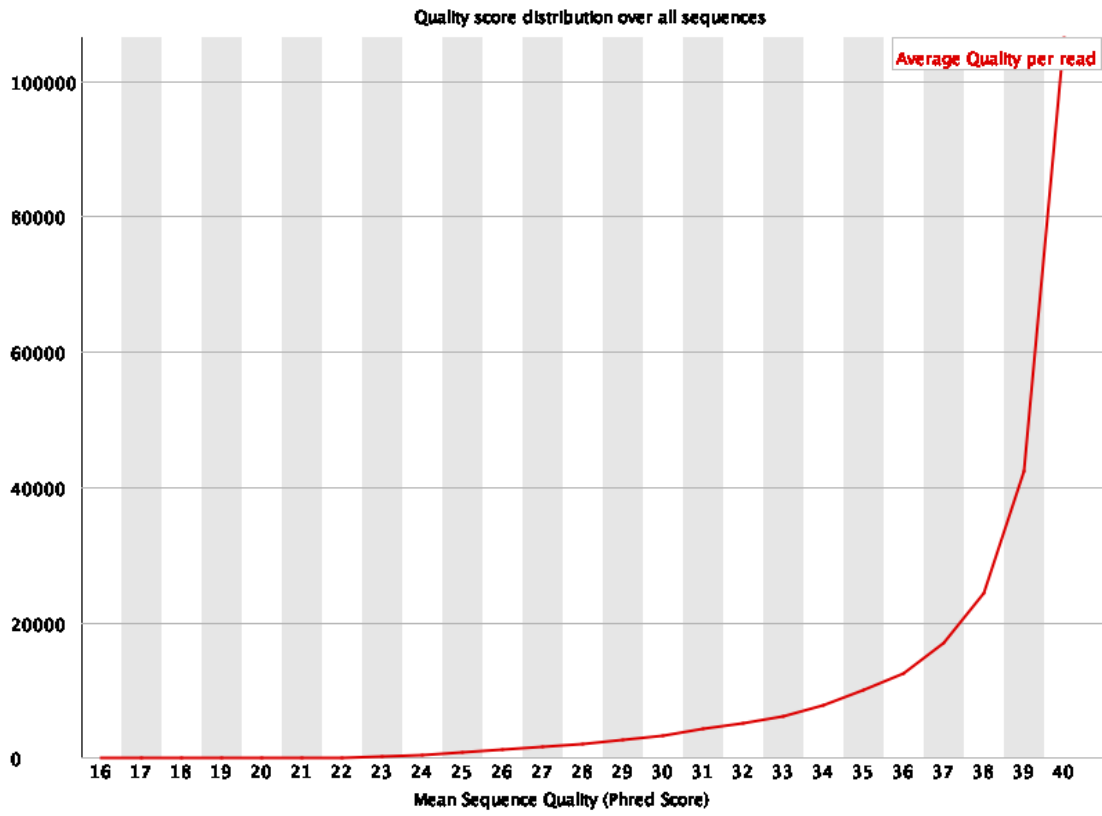
Basic Statistics

Measure	Value
Filename	Anc_R1_head.fastq.gz
File type	Conventional base calls
Encoding	Sanger / Illumina 1.9
Total Sequences	250000
Sequences flagged as poor quality	0
Sequence length	151
%GC	50

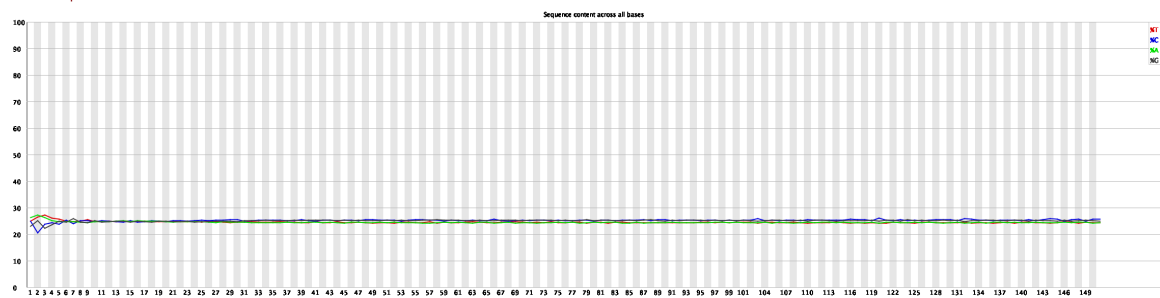
Per base sequence quality



✔ Per sequence quality scores



✔ Per base sequence content



トリミング ⇒ trimmomatic

プログラム trimmomatic

trimmomatic が広範に使われているらしい

(Usadellab, Institute of Botany and Molecular Genetics at RWTH Aachen University) <http://www.usadellab.org/cms/?page=trimmomatic>

- 品質の低い部分（先頭・末尾、スライディングウィンドウ）を切り捨てる
- 結果として短くなり過ぎたリードは、取り除く
- アダプターの除去

Javaの実行環境が必要

Javaの動作の確認

```
java -version
```

なければダウンロードインストール

「詳しい情報」 → oracle.comのページ ⇒ DOWNLOADマークでJDKをダウンロード

trimmomaticのダウンロード

ダウンロード：

<http://www.usadellab.org/cms/index.php?page=trimmomatic>

⇒ binary ⇒ zipファイルダウンロード ⇒ zipファイルを作業ディレクトリ Breseq-Seminar へ移動

⇒ 展開 ⇒ 作業ディレクトリの下にディレクトリtrimmomatic-0.39 が出る

その中に本体 trimmomatic-0.39.jar と ディレクトリadapters

パラメータ指定

- ILLUMINACLIP: Cut adapter and other illumina-specific sequences from the read.
- SLIDINGWINDOW: Perform a sliding window trimming, cutting once the average quality within the window falls below a threshold.
- LEADING: Cut bases off the start of a read, if below a threshold quality
- TRAILING: Cut bases off the end of a read, if below a threshold quality
- CROP: Cut the read to a specified length
- HEADCROP: Cut the specified number of bases from the start of the read
- MINLEN: Drop the read if it is below a specified length
- TOPHRED33: Convert quality scores to Phred-33
- TOPHRED64: Convert quality scores to Phred-64

作業ディレクトリの中で、

```
java -Xmx4g \  
-jar Trimmomatic-0.39/trimmomatic-0.39.jar\  
PE -phred33 \  
Anc_R1_head.fastq.gz Anc_R2_head.fastq.gz\  
Anc_for_paired.fq.gz Anc_for_unpaired.fq.gz\  
Anc_rev_paired.fq.gz Anc_rev_unpaired.fq.gz\  
ILLUMINACLIP:Trimmomatic-0.39/adapters/TruSeq3-PE.fa:2:30:10\  
LEADING:3 TRAILING:3 \  
SLIDINGWINDOW: 4:15 MINLEN:36
```

-Xmx4g : 仮想メモリを4G取る -jar <jarファイル> : Javaのプログラム本体のjarファイルを指定 PE : Pair-ended ⇒ Single-endedならSE -phred33 : クオリティをphred33で書く (又はphred64)

ILLUMINACLIP

ILLUMINACLIP:<fastaWithAdaptersEtc>:<seed mismatches>:<palindrome clip threshold>:<simple clip threshold>

- fastaWithAdaptersEtc: specifies the path to a fasta file containing all the adapters, PCR sequences etc. The naming of the various sequences within this file determines how they are used. See below.
- seedMismatches: specifies the maximum mismatch count which will still allow a full match to be performed
- palindromeClipThreshold: specifies how accurate the match between the two 'adapter ligated' reads must be for PE palindrome read alignment.
- simpleClipThreshold: specifies how accurate the match between any adapter etc. sequence must be against a read.

```
ILLUMINACLIP:adapters/TruSeq3-SE.fa:2:30:10 \  

```

TruSeq3-SE.fa : アダプタ配列データのファイル。 Trimmomatic-0.39/adaptersまたはイルミナのサイトから合ったものを持ってくる

```
LEADING:3 \  
TRAILING:3 \  
SLIDINGWINDOW:4:15 \  
MINLEN:36
```

先頭の3塩基、終端の3塩基は、品質が閾値より低ければ カットする

スライディングウィンドウを、窓幅4、品質閾値15で行う

- ウィンドウを先頭から1塩基ずつずらしながら、
- ウィンドウ幅4で品質の平均を取り、
- もし平均が品質閾値15より低ければカットする (ウィンドウの部分のカット?)

最終的に長さがMINLEN 36より短くなったリードは捨てる

ドキュメントには手順がもっと細かく書いてある。 http://www.usadellab.org/cms/uploads/supplementary/Trimmomatic/TrimmomaticManual_V0.32.pdf

マッピング・集計 (Breseq)

リードをリファレンス (参照配列) にマップ (貼り付ける) して、集計する

Breseq

Breseqを使うと、マッピング・集計・統計的評価をまとめて実行する。(その代わり時間がかかる)

インストール

<https://barricklab.org/twiki/pub/Lab/ToolsBacterialGenomeResequencing/documentation/installation.html#method-1-binary-download>

前提ソフトとして、Bowtie2とRが必要なので、先にインストールする。

Bowtie2のインストール

<https://sourceforge.net/projects/bowtie-bio/files/bowtie2/2.3.5.1/>

から bowtie2-2.3.5.1-macos-x86_64.zip をダウンロードフォルダにダウンロードし、展開する。

この中で、bowtie2***というファイル (15個ある) をBreseqから使えるようにしなければならない。

ここでは /usr/local/bin ディレクトリの中にコピーを作って置くことにする。

;/usr/local/ディレクトリは普通にはfinderで開けない (見えない) が、finder内でshift+command+G とすると

検索窓が開くのでそこに /usr/local/bin と入力すると、開くことができる。そこへコピーを置く。

コピーしたファイルはネットからもってきた実行ファイルなのでセキュリティチェックにひっかかるが、

先述の「コマンドラインから」の方法で回避できているはずである。

Rのインストール

Rのミラーサイト <https://cran.ism.ac.jp> からDownload R for (Mac) OS Xを選んで R-3.6.2.pkg をダウンロードディレクトリにダウンロードし、

R-3.6.2.pkgをクリックして起動、指示に従う。インストーラがほとんど自動的にインストールしてくれる。

Breseqのダウンロード・インストール

最後に Breseq 本体をダウンロードする。

<https://github.com/barricklab/breseq/releases>

から、breseq-0.35.0-MacOSX-10.9+.tar.gz をダウンロードフォルダにダウンロードし、クリックして展開する。

それによって、ディレクトリ breseq-0.35.0-MacOSX-10.9+ ができる。

このフォルダ全体を、作業ディレクトリの中にコピーし、更に、ディレクトリの名前を

breseq-0.35.0-MacOSX-10.9+ から breseq に変更する。ディレクトリ名の変更は finder の中で出来るが、

コマンドラインから変更する場合には mv 旧ファイル名 新ファイル名 とする。

ディレクトリ名もファイル名も同じに扱ってよい。(mv 旧ディレクトリ名 新ディレクトリ名)

AP012030.gbの準備

リファレンスゲノムが必要だが、最初にダウンロードしておいたはずである。

もし無ければ、NCBIのサイトからGenBank形式のファイルをダウンロードし、名前を AP012030.gb に変更しておく。

Breseqの処理

作業ディレクトリの中から、コマンドラインでbreseqを起動する。breseqの本体はディレクトリbreseqの中のbin/breseqにあるので、

コマンド名はbreseq/bin/breseqになる。出力を置く場所は-Oパラメタで指定でき、ここではOutputとしている。

入力は作業ディレクトリ内にあるAnc_for_paired.fq.gzとする。trimmomaticで順逆ファイルを1つにまとめているので、そのうちpairedの方を使う。ターミナルへの出力を一部、ファイルbreseq.outに転送しておく。

```
breseq/bin/breseq -o Output -r AP012030.gb Anc_for_paired.fq.gz Anc_rev_paired.fq.gz > breseq.out
```

で処理できる。