

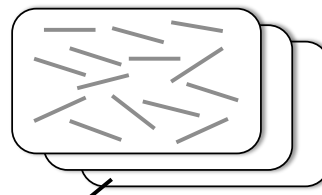
ショートリードのマッピング

ゲノム配列
(リファレンス reference 配列)

形式(配列)

```
>chr
AGCTTTTCATTCTGACTGCAACGGGCAATATGTCT
CTGTGTGGATTAAAAAGAGTGTCTGATAGCAGC
TTCTGAACGGTTACCTGCGGTGAGTAAATAAAA
TTTATTGACCTAGCTCACTAAATCTTTAACCAA
TATAGGCATAGGCGACAGACAGATAAAATTACAG
ACTACACAACATCCATGAAACGATTAGCACCACC
ATTACCACCACCATCACCATTACCACAGGTAACGG
```

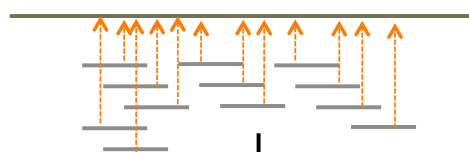
サンプル(ゲノムDNA/RNA)
(リード read 配列)



形式
(配列+クオリティ値)

```
@SRR1515276.1 HWI-ST808:151:D2D13ACXX:2:1207:
ATCCGGCTGGCGCACCACCTATGTCCGGGCAATACAAAGCTGG
+SRR1515276.1 HWI-ST808:151:D2D13ACXX:2:1207:
@@@AD>DDFF7DC?FFEBF@DFII<DF@AAA6AEFDBDCA?>A?
@SRR1515276.2 HWI-ST808:151:D2D13ACXX:2:1207:
CACCGTGTAGTACCAGCATCTCGGTACAATCAGCAATCCAGTC
+SRR1515276.2 HWI-ST808:151:D2D13ACXX:2:1207:
CCCFDFDFHDFHIIIEGHIJJJJGPHGGHGGGIIJDGIJHH
@SRR1515276.3 HWI-ST808:151:D2D13ACXX:2:1207:
CAGGACATCGCCTTTGATCGGTTTCAGACTTCGACCAACCTGCAI
+SRR1515276.3 HWI-ST808:151:D2D13ACXX:2:1207:
CCCFDFDFAFHFIJGHIJJJJJGHEHIJGHIJFEHIIA@FIF
```

リファレンス配列へのマッピング



クオリティチェック
アダプター除去

形式(マッピング結果)

```
@HD VN:1.0 SO:unsorted
@SQ SN:chr LN:4639675
@PG ID:bowtie2 PN:bowtie2 VN:2.2.4 CL:"/bio/bin/bowtie2-align
SRR1515276.40 0 chr 4423609 42 51M * 0 0 GGAATTCCTCACTGCCA
SRR1515276.158 16 chr 501700 42 51M * 0 0 ACGCACCGAGTGCAAAG
SRR1515276.212 4 * 0 0 * * 0 0 GGCCGCTTTCAGCGTGT
SRR1515276.319 0 chr 2922768 42 51M * 0 0 GCTTAAGTTGATTAAGG
SRR1515276.367 16 chr 2753873 42 51M * 0 0 GCGGTCCGTCGCCAGC
SRR1515276.411 0 chr 3440721 42 51M * 0 0 ACGGCATAAATTTCTTGA
```

復習: cutadaptによるアダプターの除去

入力

- リード配列 (FASTQ 形式; paired-end)

etec_1.fq

etec_2.fq

- アダプター配列 (それぞれを3'端から除去)

Adapter1: GATCGGAAGAGCGGTT

Adapter2: GATCGGAAGAGCGTCG

- ◆ アダプター配列除去の実行

除去後のデータ (FASTQ形式) は etec_1.cut.fq, etec_2.cut.fq とする)

\$ cutadapt

復習: **bowtie2** 用インデックスの作成

入力

- ゲノムデータ (FASTA形式)
eco_o139.fa 腸管毒素原性大腸菌(ETEC) O139:H28のゲノム配列

◆ bowtie2用インデックスの作成(インデックス名は etec)

\$ bowtie2-build

復習: **bowtie2**の実行 (paired-end)

入力

- リード配列 (FASTQ 形式; paired-end; アダプターを除去したもの)
etec_1.cut.fq
etec_2.cut.fq
- リファレンス配列のインデックス名
etec (先ほど作ったもの)

◆ bowtie2によるマッピングの実行 (出力:etec_bowtie2.sam)

\$ bowtie2

復習: **SAM**から**BAM**への変換

入力

- SAMファイル（さきほどbowtie2によって作成されたもの）
etec_bowtie2.sam

- ◆ SAMからBAMへ変換する（出力ファイル名:etec_bowtie2.bam）

\$ samtools

- ◆ 作成したBAMファイルをヘッダ付きでSAMに変換してlessで表示する

\$ samtools

復習: **BAM**ファイルのインデックスづけ

入力

- BAMファイル（さきほどSAMからの変換によって作成されたもの）
etec_bowtie2.bam

- ◆ リファレンス配列上の位置の順にソートする
（出力ファイル:etec_bowtie2_sorted.bam）

\$ samtools

- ◆ ソートされたBAMファイルに対してインデックスを作成する

\$ samtools

- ◆ インデックスを使って、リファレンスの染色体配列（染色体名:ETEC_chr）の
10000-12000 の範囲にマッピングされた結果のみを表示する

% samtools