

# ショートリードのマッピング

ゲノム配列

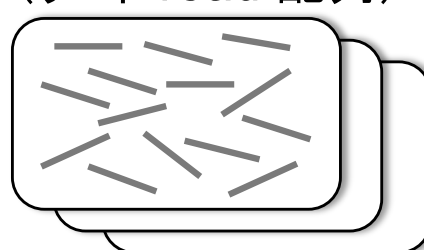
(リファレンス reference 配列)

形式 (配列)

```
>chr
AGCTTTTCATTCTGACTGCAACGGGCAATATGTCT
CTGTGTGGATTAAAAAAGAGTGTCTGATAGCAGC
TTCTGAAGTGGTTACCTGCCGTGAGTAAATAAAA
TTTTATTGACTTAGGTCACTAAATACTTTAACCAA
TATAGGCATAGCGCACAGACAGATAAAAAATTACAG
AGTACACAACATCCATGAAACGCATTAGCACCACC
ATTACCACCACCATCACCATTACCACAGGTAACGG
```

サンプル(ゲノムDNA/RNA)

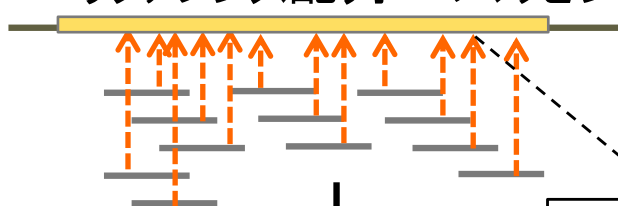
(リード read 配列)



形式  
(配列 + クオリティ値)

```
@SRR1515276.1 HWI-ST808:151:D2D13ACXX:2:1207:3625:88631 length=51
ATCCGGCTGGCGCACCCGACCTATGTTCCGGGCGAATACAAGCTGGGTGAAG
+SRR1515276.1 HWI-ST808:151:D2D13ACXX:2:1207:3625:88631 length=51
@@@AD>DDFF7DC?FFEBF@DFII<DF@AAA6AEFDBDCA?>A?B=>B::
@SRR1515276.2 HWI-ST808:151:D2D13ACXX:2:1207:3871:88513 length=51
CACCGTGTAGTACCAGCATCCTGCGTACAATCAGCAATCCCAGTCCCTCCCC
+SRR1515276.2 HWI-ST808:151:D2D13ACXX:2:1207:3871:88513 length=51
CCCFDFDFHDFHFIIEGIHJJJJGFGHGGHGGIJDGIJHHGGGHHI
@SRR1515276.3 HWI-ST808:151:D2D13ACXX:2:1207:3950:88530 length=51
CAGGACATCGCCTTTGATCGGTTTCAGACTTCGGACCAACCTGCATTTTCAG
+SRR1515276.3 HWI-ST808:151:D2D13ACXX:2:1207:3950:88530 length=51
CCCFDFDFAFHFHFIJGHIJJJJJHEHIIJGHIFEHIIA@FIFHGGIIGI
```

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 (遺伝子アノテーション)

chr	RefSeq	start_codon	190	192	1.000	+	.	gene_id	"b0001"; transcript_id	"b0001";
chr	RefSeq	CDS	190	252	1.000	+	0	gene_id	"b0001"; transcript_id	"b0001";
chr	RefSeq	stop_codon	253	255	1.000	+	.	gene_id	"b0001"; transcript_id	"b0001";
chr	RefSeq	exon	190	255	1.000	+	.	gene_id	"b0001"; transcript_id	"b0001";

形式 (マッピング結果)

```
@HD VN:1.0 SO:unsorted
@SQ LN:4639675
@PG ID:bowtie2 PN:bowtie2 VN:2.2.4 CL:"/bio/bin/bowtie2-alig
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 GCGTGTCCTGCCGAGC
SRR1515276.411 0 chr 3440721 42 51M * 0 0 ACGGCATAATTTCTTGA
SRR1515276.434 0 chr 4198737 42 51M * 0 0 GCGCGGTACGCATCTGG
```

# 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: AGATCGGAAGAGCGGTT

Adapter2: AGATCGGAAGAGCGTCG

## ◆ 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 ☐ AGATCGGAAGAGCGGTT ☐ AGATCGGAAGAGCGTCG  
           ☐ etec_1.cut.fq ☐ etec_2.cut.fq  
           etec_1.fq   etec_2.fq
```

# 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    (出力インデックス名: `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 [ ] etec [ ] etec_1.cut.fq [ ] etec_2.cut.fq  
          [ ] etec_bowtie2.sam
```

# 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   etec_bowtie2.sam  etec_bowtie2.bam
```

- ◆ 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   etec_bowtie2.bam  less
```

# 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  etec_bowtie2.bam  etec_bowtie2_sorted.bam
```

- ◆ 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  etec_bowtie2_sorted.bam
```

- ◆ 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  etec_bowtie2_sorted.bam 
```