

# atlas-utils之根据barcode对序列文件进行拆分：split

## 一、atlas-utils split介绍

功能描述：

`atlas-utils split` 根据 `barcode` 序列信息序列文件进行拆分，不删除 `barcode` 序列。

命令行接口：

```
1  Usage: atlas-utils split [options]
2
3  Options:
4    -1 STR  forward reads file;
5    -2 STR  reverse reads file;
6    -b STR  barcode mapping file;
7    -d STR  output filename directory;
8    -l INT  minimum barcode length, default: 5;
9    -h      display usage;
10
11 Example:
12 atlas-utils split -b barcodes.txt -1 I365.R1.fastq.gz -2 I365.R2.fastq.gz
   -l 5 -d I365
```

可选参数：

```
1  -1 字符串 正向序列文件；
2  -2 字符串 反向序列文件；
3  -b 字符串 barcode 映射文件；
4  -d 字符串 输出文件夹名称；
5  -l 字符串 最小barcode长度，默认为5；
6  -h      显示用法；
```

## 二、使用场景实例及其用法

示例演示：

示例文件：`barcodes.txt`, `A1.1.fastq.gz` `A1.2.fastq.gz`

```
1 $ zcat barcodes.txt
```

```
1 D001      TACCGGAA      GCCCTGTA
2 D002      CTGCTACA      TAAATGCG
```

```
1 $ zcat A1.1.fastq.gz | head -n 6
```

```

1 @A01050:39:HJ5FKDRXX:2:2135:18141:18004 1:N:0:TCCGCGAA+GGCTCTGA
2 GCCCTGTAGGACTACAAGGGTTTCTAATCCTGTTCGATACCCGCACTTTCGAGCTTCAGCGTCAGTTGCGCTCCAGT
  GAGCTGCCTTCGCAATCGGAGTTCTTCGTGATATCTAAGCATTTACCGCTACACCACGAATTCGCCCCACTTTGTG
  CGTACTCAAGGAAACAGTTTCGCGCTGCAGTTCAGATGTTGAGCATCTACATTTACAACACGCTTAATCTCCGGCC
  TACGCTCCCTTTAAACCCA
3 +
4 F:FF,::FFFFFF:FFFFFFFFFFFFFFF:,FFFF,FF:,FFFFFFFF:FFFF:FFFF:FFFF,FFFFFFFF:
  FFFFFFF:FFF:FFFF,FFFFFFFF:,FFFF:FF:,FF:,FFFFFF:FFF,F:FFF,FFFFFF:FFFF:FF
  FFFFFFFFFFFFF:F:FFFF,FFFFFFFFFFFFFFFFFFFFFFFFFFFFF:,FFFFFFFFFFFF,F,.,FF:,FF
  :FFFF:FFFF,F:FFFF
5 @A01050:39:HJ5FKDRXX:2:2116:24451:15969 1:N:0:TCCGCGAA+GGCTCTGA
6 CTGCTACAGTGCCAGCCGCGCGTAATACGGAGGATGCGAGCGTTCTCCGATTTATTGGGTTTAAAGGGTGCCTA
  GGCGGAAAGTTAAGTCAGCGGTAAATCGAGAGGCTCAACCTCTTCAGCCGTTGAACTGGTTCCTAGAGTGAGT
  GAGAAGTACGCGGAATGCGAGGTGTAGCGGTGAAATGCATAGATATCACGCAGAACTCCAATTGCGAAGGCAGCGTA
  CCGGCACTCAACTGACGCT

```

```
1 $ zcat A1.2.fastq.gz | head -n 6
```

```

1 @A01050:39:HJ5FKDRXX:2:2135:18141:18004 2:N:0:TCCGCGAA+GGCTCTGA
2 TACCGGAAGTGCCAGCAGCCGCGTAATACGGAAGGTCCGGCGTTATCCGGATTTATTGGGTTTAAAGGGAGCGTA
  GGCCGGAGATTAAGCGTGTGTGACATGTAGATGCTCAACATCTGAACTGCAGCGCGAACTGGTTCCTTGAGTACG
  CACAAAGTGCGCGGAATTCGTGGTGTAGCGGTGAAATGCTTAGATATCACGAAGAACTCCGATTGCGAAGGCAGCTC
  ACTGGAGCGCACCTGACGC
3 +
4 FFFFFFFFFFFFFFFFFFFFFFFFFFFFFF,FFFFFFFFFFFF:FFFF:FFFFF:,FFFFFFFFFFFFFFFF,FFFFF
  FFFFFFFFFFFFFFFFFF:FFFF,FFFF:F:FFF:FFFFFFFFFFFFFFFFFFFFF:,FF:FFFFFFFFFFFFFFFF
  FFFFFF:FFFFFFFFFFFFFFFF,FFFFFFFFFFFF:FFFFFFFFFFFF:FFFFFFFF:FFFFF:FFFFFFFF
  FFF,FFF,FFF:FFFFFFF
5 @A01050:39:HJ5FKDRXX:2:2116:24451:15969 2:N:0:TCCGCGAA+GGCTCTGA
6 TAAATGCGGGACTACAGGGTATCTAATCCTGTTCGATCCCCACGCTTTCGTGCTTCAGCGTCAGTTGAGTGCCGGT
  ACGCTGCCTTCGCAATTGGAGTTCTGCGTGATATCTATGCATTTACCGCTACACCTCGCATTCCGCGTACTTCTCA
  CTCACTCTAGAAAACAGTTTCAACGGCTGAAAGAGGTTGAGCCTCTCGATTTTACCGCTGACTTAACCTTCCGCCT
  ACGCACCTTTAAACCAA

```

运行命令：

根据 `barcode` 序列信息序列文件进行拆分，不去除 `barcode` 序列

```

1 $ atlas-utils split -b barcodes.txt -1 A1.1.fastq.gz -2 A1.2.fastq.gz -l 5 -d
  A1
2 $ ls A1

```

```
1 D001_1.fastq D001_2.fastq D002_1.fastq D002_2.fastq
```

本文材料为 **BASE (Biostack Applied bioinformatic SEies)** 课程 **Linux Command Line Tools for Life Scientists** 材料，版权归上海逻捷信息科技有限公司 所有。

Last Update: 2020-09-11 11:56 AM

