

fastx-utils之根据index数据拆分: demultiplex

一、fastx-utils demultiplex

功能描述:

`fastx-utils demultiplex` 将PE双端序列转成交叠的序列文件。

命令行接口:

```
1 $ fastx-utils demultiplex
2
3 Usage: fastx-utils demultiplex [options]
4
5 Options:
6   -d output directory for demultiplex reads. default: [./]
7   -i index sheet [sample\tindex].
8   -1 forward reads.
9   -2 reverse reads.
```

参数接口:

- d 数据输出目录
- i 信息表 [样本\t索引]
- 1 正向测序数据.
- 2 反向测序数据.

二、使用场景实例及其用法

示例演示:

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

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

```
1 @A01050:39:HJ5FKDRXX:2:2135:18141:18004 1:N:0:TCCGCGAA+GGCTCTGA
2 GCCCTGTAGGACTACAAGGGTTTCTAATCCTGTTTCGATACCCGCACTTTCGAGCTTCAGCGTCAGTTGCGCTCCAGT
3 GAGCTGCCTTCGCAATCGGAGTTCTTCGTGATATCTAAGCATTTACCGCTACACCACGAATTCGCCCCACTTTGTG
4 CGTACTCAAGGAAACCAAGTTTCGCGCTGCAGTTCAGATGTTGAGCATCTACATTTACAAACACGCTTAATCTCCGGCC
5 TACGCTCCCTTTAAACCA
6 +
7 F:FF,::FFFFFF:FFFFFFFFFFFFFFF:,FFFF,FF:,FFFFFFFF:FFFF:FFFF:FFFF,FFFFFFFF:
8 FFFFFFF:FFF:FFFF,FFFFFFFF:,FFFF:FF:,FF:,FFFFFF:FFFF,F:FFFF,FFFFFF:FFFF:FF
9 FFFFFFFFFFFFF:F:FFFF,FFFFFFFFFFFFFFFFFFFFFFFFFFFFF:,FFFFFFFFFFFF,F,.,FF:,FF
10 :FFFF:FFFF,F:FFFF
11 @A01050:39:HJ5FKDRXX:2:2116:24451:15969 1:N:0:TCCGCGAA+GGCTCTGA
12 CTGCTACAGTGCCAGCCGCGGTAATACGGAGGATGCGAGCGTTCTCCGATTTATTGGGTTTAAAGGGTGCCTA
13 GGCGGAAAGTTAAGTCAGCGGTAAAAATCGAGAGGCTCAACCTCTTTCAGCCGTTGAAACTGGTTTTCTAGAGTGAGT
14 GAGAAGTACGCGGAATGCGAGGTGTAGCGGTGAAATGCATAGATATCACGCAGAACTCCAATTGCGAAGGCAGCGTA
15 CCGGCACTCAACTGACGCT
```

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

```
1 @A01050:39:HJ5FKDRXX:2:2135:18141:18004 2:N:0:TCCGCGAA+GGCTCTGA
2 TACCGGAAGTGCCAGCAGCCGCGTAATACGGAAGGTCCGGGCGTTATCCGGATTTATTGGGTTTAAAGGGAGCGTA
  GGCCGGAGATTAAGCGTGTTGTGACATGTAGATGCTCAACATCTGAAGTGCAGCGCGAACTGGTTTCCTTGAGTACG
  CACAAAGTGGGCGGAATTCGTGGTGTAGCGGTGAAATGCTTAGATATCACGAAGAACTCCGATTGCGAAGGCAGCTC
  ACTGGAGCGCACCTGACGC
3 +
4 FFFFFFFFFFFFFFFFFFFFFFFFFF,FFFFFFFFFFFF:FFFF:FFFFF:,FFFFFFFFFFFFFFFF,FFFFFF
  FFFFFFFFFFFFFFFFFF::FFFF,FFFFF:F:FFF:FFFFFFFFFFFFFFFFFFFF, :FF:FFFFFFFFFFFFFFFF
  FFFFFF::FFFFFFFFFFFFFFFF,FFFFFFFFFFFF:FFFFFFFFFFFF:FFFFFFFF:FFFFFF:FFFFFFFF
  FFF,FFF,FFF:FFFFFFF
5 @A01050:39:HJ5FKDRXX:2:2116:24451:15969 2:N:0:TCCGCGAA+GGCTCTGA
6 TAAATGCGGGACTACAGGGGTATCTAATCCTGTTCGATCCCCACGCTTTCGTGCTTCAGCGTCAGTTGAGTGCCGGT
  ACGCTGCCTTCGCAATTGGAGTTCTGCGTGATATCTATGCATTTACCGCTACACCTCGCATTCCGCGTACTTCTCA
  CTCCTCTAGAAAACGAGTTTCAACGGCTGAAAGAGGTTGAGCCTCTCGATTTTACCGCTGACTTAACTTTCCGCCT
  ACGCACCTTTAAACCCAA
```

```
1 $ cat index.txt
```

```
1 D001      TCCGCGAA+GGCTCTGA
```

运行命令：将双端序列转成交叠的序列文件。

```
1 $ fastx-utils demultiplex -d ./ -i index.txt -1 A1.1.fastq.gz -2
  A1.2.fastq.gz
```

```
1 $ ls
2 A1.1.fastq.gz A1.2.fastq.gz D001_R1.fastq D001_R2.fastq index.txt
```

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: 9/2/2020 12:37:53 AM