

atlas-utils之获取双端序列的barcode序列：barcode

一、atlas-utils barcode介绍

功能描述：

`atlas-utils barcode` 从双端序列中获取barcode序列, 可以用来检查序列中存在哪些barcode序列, 检查数据拆分的情况。

命令行接口：

```
1 $ atlas-utils barcode
2
3 Usage: atlas-utils barcode <1.fq> <2.fq>
```

二、使用场景实例及其用法

使用场景经典案例：

检查数据拆分率低，或者检查barcode是否有使用错了。

示例演示：

示例文件： `A-1_1.fastq`， `A-1_2.fastq`

```
1 $ cat A-1_1.fastq | head -n 6
```

```
1 $ cat barcode/example/A-1_1.fastq | head -n 6
2 @HISEQ:483:HLJ2LBCXY:1:1101:7924:2136 1:N:0:CACCGG
3 CCTATGGGACGCAGCAGTGGGGAATATTGGACAATGGGCGCAAGCCTGATCCAGCCATGCCGCGTGAGTGATGAAGG
  CCTAGGGTTGTAAAGCCCTTTTCGGCGGGGAAGATAATGACGGTACCCGCGAGAAGAAGCCCCGGCTAACTTCGTGCC
  AGCAGCCGCGGTAATACGAAGGGGCTAGCGTTGCTCGGAATTACTGGGCGTAAAGCGCACGTAGGCGGCTTTCTAA
  GTCGGGGGTGAA
4 +
5 HH1CHHHHDFHDHDH?C<DEG/CEHHIHEHCHECFHICHHHHDHIICHHIEHCHCEGHHIICH?
  EEHFHHHIECEE?HE?1GHHHHDHHHHHFFGHHHCHHHII?H?HHEHCHF1CGHC0DEHEHHD<<FEHHHIGHIC??
  HHCGHHHHIHH@FFEEHHDHCHHHIHE@GH-G??EDEGEHIIICHHC?EHH-6@8-B-
  =C,5@FHHHHHHHCEHH#####
6 @HISEQ:483:HLJ2LBCXY:1:1101:16730:2160 1:N:0:CACCGG
7 CCTACGGGAGGCACCAGCAAGGAATTTTCGCAATGGGCGAAAGCCTGACGGAGCAACGCCGCGTGGGGGATGAAGG
  CCTTCGGGTTGTAAACCCCTTTTCGAGGGAAGAAGATCTGACGGTACCTCGCGAATAAGCCACGGCTAACTACGTG
  CCAGCAGCCGCGGTAATACGTAGGTGGCAAGCGTTGTCCGGATTACTGGGCGTAAAGCGCGCGCAGGCGGACTGGT
  AAGTCTGGGGCG
```

```
1 $ cat A-1_2.fastq | head -n 6
```

```

1 @HISEQ:483:HLJ2LBCXY:1:1101:7924:2136 2:N:0:CACCGG
2 GGACTACTTGGGTATCTAATCCTGTTTGTCTCCCCACGCTTTCGCACCTCAGCGTCAGTACCGGGCCAGTGAGCCGCC
  TTCGCCACTGGTGTCTTGCGAATATCTACGAGTTTCACCTCTACACTCGCAGTTCCACTCACCTCTCCCGGACTCG
  AGCTCTCCAGTATCGAAGGCAGTTCTGGAGTTAAGCTCCAGGAGTTACCCCCGACTTAGAAAGCCGCCTACGTGCG
  CTTTACGCCCAGGC
3 +
4 HG1EG1F?CGEHGHCHH@GC1<@GEHE?GHHHHHHCHHHHEHE=EFHHIIGEDDEC@GHHFH=HD?CCG1D1C?
  EDCCHCG10C00DGH@CCGHHIGD/C/CECHHEHIIGHECEHHHHHHIIGHGFH@HIHEEHCGEHFHDGHHIHEH,C
  CHHECHHHIIGFABGC@EHICHHHFHHEHH?AG.AFBCHCAHHIHH-
  8B@C-.8>=>+6CEC@G@@EHHHCDH=CHE@@D+5@EEF--7CEG##
5 @HISEQ:483:HLJ2LBCXY:1:1101:16730:2160 2:N:0:CACCGG
6 GGACTACACGGGTATCTAATCCGTTTGTCTCCCCACGCTTTCGCGCCTCAGCGTCAGTTGCCCCCAGAGAGTCGCT
  TTCGCCACTGGTGTCTCTCCCGATATCTACGCATTCCACCACTACACCGGGAATTCCACTCTCCTCTGAGCCCCTCA
  AGACCGGCAGTATCCGGGGACCGCTCACAGTTGAGCCGTGAGATTTACCCCAGACTTACCAGCCCCGCCTGCGCGCG
  CTTTACGCCCAGTA

```

运行命令：

获取 A-1 双端序列的 `barcode` 序列, 默认8个碱基;

```
1 $ atlas-utils barcode -l 8 A-1_1.fastq A-1_2.fastq | head
```

1	HISEQ:483:HLJ2LBCXY:1:1101:7924:2136	CCTATGGG	GGACTACT
2	HISEQ:483:HLJ2LBCXY:1:1101:16730:2160	CCTACGGG	GGACTACA
3	HISEQ:483:HLJ2LBCXY:1:1101:19412:2440	CCTATGGG	GGACTACT
4	HISEQ:483:HLJ2LBCXY:1:1101:10790:2619	CCTACGGG	GGACTACT
5	HISEQ:483:HLJ2LBCXY:1:1101:10004:2933	CCTATGGG	GGACTACA
6	HISEQ:483:HLJ2LBCXY:1:1101:16103:2752	CCTATGGG	GGACTACG
7	HISEQ:483:HLJ2LBCXY:1:1101:17844:2827	CCTATGGG	GGTCTACA
8	HISEQ:483:HLJ2LBCXY:1:1101:5069:3192	CCTATGGG	GGACTACT
9	HISEQ:483:HLJ2LBCXY:1:1101:11184:3127	CCTATGGG	GGACTACT
10	HISEQ:483:HLJ2LBCXY:1:1101:11518:3202	CCTATGGG	GGACTACT

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: 2020-09-9 11:56 AM