

fastx-utils之修改序列的名字: rename

一、fastx-utils rename介绍

功能描述:

`fastx-utils rename` 修改序列的名字, 统一使用前缀, 比如 **"ZOTU_"**。

命令行接口:

```
1 $ fastx-utils rename
2
3 Usage: fastx-utils rename [options] <fasta/q> <prefix>
4
5 Options:
6   -r retain original sequence identifier.
7   -m print the mapping info to stderr streaming.
```

可选参数:

```
1 -l 保留原始序列标识符;
2 -m 输出映射信息到标准错误;
```

二、使用场景实例及其用法

使用场景经典案例:

1. 16S扩增子数据分析, 构建的OTU名字可以自己定义, 比如 **ZOTU**, 并重新进行索引, `_1`, `_2`

示例演示:

示例文件: `sequence.fastq`

```
1 $ cat sequence.fastq | head -n8
```

[illegible]

运行命令：

参数选项1: 默认参数, 修改序列名字为 A1:

```
1 | $ fastx-utils rename sequence.fastq A1 | head -n8
```

```

1 | @A1_1
2 | CCTATGGGATGCACCAAGTGGGGAATATTGGACAATGGGCGAAAGCCTGATCCAGCCATGCCGCGTGAGTGATGAAGG
   | CCTTAGGGTTGTAAAGCTCTTTCGGCGGGGAAGATAATGACGGTACCCGCGAGAAGAAGCCCCGGCTAACTTCGTGCC
   | AGCAGCCGCGTAATACGAAGGGGGCTAGCGTTGTTTCGGAACCACTGGGCGTAAAGCGCGTGAGGCGGATTGTTAA
   | GTCGGGGGTGAAATCCTGGGGCTCAACCTCAGAACTGCCTTCGATACTGGCGATCTTGAGTCCGGGAGAGGTGAGTG
   | GTATTCCTAGTGTAGAGGTGAAATTCGTAGATATTAGGAAGAACACCAAGTGGCGAAGGCGGCTCACTGGCCCGGTAC
   | TGACGCTGAAACGCGAAAGCGTGGGGAGCAAACAGGATTAGATACCCGTGTAGTCC
3 | +
4 | IIIIIIIHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHH
   | IIIIIIIHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHH
   | IIGHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHH
   | IIGHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHH
   | BFEHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHH
   | F?
   | FB.FHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHH
   | IHHHGHGHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHH
   | IIIIHEHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHH
   | IIIIHEHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHH
5 | @A1_2
6 | CCTACGGGAGGCACCAAGTGGGGAATATTGGACAATGGGCGAAAGCCTGATCCAGCCATGCCGCGTGAGTGATGAAGG
   | CCCTAGGGTTGTAAAGCTCTTTCACCGGTGAAGATAATGACGGTAACCGGAGAAGAAGCACCGGCTAACTTCGTGCC
   | AGCAGCCGCGTAATACCAAGGGGGCTAGCGTTGTTTCGGAATTACTGGGCGTAAAGCGCACGTAGGCGGATATTTAA
   | GTCAGGGGTGAAATCCCGGGGCTCAACCCCGGAACTGCCTTTGATACTGGGTATCTAGAGTATGGTAGAGGTGAGTG
   | GAATCCGAGTGTAGAGGTGAAATTCGTAGATATTCGAGGAACACCAAGTGGCGAAGGCGGCTCACTGGACCATTAC
   | TGACGCTGAGGTGCGAAAGCGTGGGGAGCAAACAGGATTAGATACCCATGTAGTCC
7 | +
8 | IIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIII
   | IIIIIIIHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHH
   | IIIIIIIHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHH
   | IIIIIIIHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHH
   | KKKKKKKKKKKKKHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHH
   | GHE<<<IIIIHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHH
   | IIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIII

```

参数选项2： 设置 `-r` 参数，保留原始序列标识符。

```
1 | $ fastx-utils rename -r sequence.fastq A1 | head -n8
```

[illegible]

参数选项3: 设置 `-m` 参数, 将映射关系输出标准错误。

```
1 $ fastx-utils rename -r -m sequence.fastq A1 1>rename.fastq 2>map.txt
2 $ cat map.txt | head -n 6
```

1	A1_1	HISEQ:483:HLJ2LBCXY:1:1111:14104:10730
2	A1_2	HISEQ:483:HLJ2LBCXY:1:2112:7631:97645
3	A1_3	HISEQ:483:HLJ2LBCXY:1:1204:10672:14348
4	A1_4	HISEQ:483:HLJ2LBCXY:1:1208:10369:65205
5	A1_5	HISEQ:483:HLJ2LBCXY:1:1110:8032:54809
6	A1_6	HISEQ:483:HLJ2LBCXY:1:2108:16285:9884

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: 2020-08-10 11:56 AM

