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:

\$ fastx-utils rename

Usage: fastx-utils rename [options] <fasta/q> <prefix>

Options:

- -r retain original sequence identifier.
- -m print the mapping info to stderr streaming.

可选参数:

- -r 保留原始序列标识符;
- -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

\$ cat sequence.fastq | head -n8

@HISEQ:483:HLJ2LBCXY:1:1111:14104:10730

CCTATGGGATGCACCAGTGGGGAATATTGGACAATGGGCGAAAGCCTGATCCAGCCATGCCGCTGAGTGATGAAGGCCT
TAGGGTTGTAAAGCTCTTTCGGCGGGGAAGATAATGACGGTACCCGCAGAAGAAGCCCCGGCTAACTTCGTGCCAGCAGC
CGCGGTAATACGAAGGGGGCTAGCGTTGTTCGGAACCACTGGGCGTAAAGCGCGTGTAGGCGGATTGTTAAGTCGGGGGT
GAAATCCTGGGGCTCAACCTCAGAACTGCCTTCGATACTGGCGATCTTGAGTCCGGGAGAGGTGAGTGGTATTCCTAGTG
TAGAGGTGAAATTCGTAGATATTAGGAAGAACACCAGTGGCGAAGGCGGCTCACTGGCCCGGTACTGACGCTGAAACGCG
AAAGCGTGGGGAGCAAACAGGATTAGATACCCGTGTAGTCC

+

@HISEQ:483:HLJ2LBCXY:1:2112:7631:97645

CCTACGGGAGGCACCAGTGGGGAATATTGGACAATGGGCGAAAGCCTGATCCAGCCATGCCGCTGAGTGATGAAGGCCC
TAGGGTTGTAAAGCTCTTTCACCGGTGAAGATAATGACGGTAACCGGAGAAGAAGCACCGGCTAACTTCGTGCCAGCAGC
CGCGGTAATACCAAGGGGGCTAGCGTTGTTCGGAATTACTGGGCGTAAAGCGCACGTAGGCGGATATTTAAGTCAGGGGT
GAAATCCCGGGGCTCAACCCCGGAACTGCCTTTGATACTGGGTATCTAGAGTATGGTAGAGGTGAAGTTACTGACGCTGAGGTGCG
AAAGCGTGGGGAGCAAACAGGATTAGATACCCATGTAGTCC

+

运行命令:

参数选项1: 默认参数, 修改序列名字为 A1:

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

@A1_1

CCTATGGGATGCACCAGTGGGGAATATTGGACAATGGGCGAAAGCCTGATCCAGCCATGCCGCTGAGTGATGAAGGCCT
TAGGGTTGTAAAGCTCTTTCGGCGGGGAAGATAATGACGGTACCCGCAGAAGAAGCCCCGGCTAACTTCGTGCCAGCAGC
CGCGGTAATACGAAGGGGGCTAGCGTTGTTCGGAACCACTGGGCGTAAAGCGCGTGTAGGCGGATTGTTAAGTCGGGGGT
GAAATCCTGGGGCTCAACCTCAGAACTGCCTTCGATACTGGCGATCTTGAGTCCGGGAGAGGTGAGTGGTATTCCTAGTG
TAGAGGTGAAATTCGTAGATATTAGGAAGAACACCAGTGGCGAAGGCGGCTCACTGGCCCGGTACTGACGCTGAAACGCG
AAAGCGTGGGGAGCAAACAGGATTAGATACCCGTGTAGTCC

+

@A1_2

CCTACGGGAGGCACCAGTGGGGAATATTGGACAATGGGCGAAAGCCTGATCCAGCCATGCCGCTGAGTGATGAAGGCCC
TAGGGTTGTAAAGCTCTTTCACCGGTGAAGATAATGACGGTAACCGGAGAAGAAGCACCGGCTAACTTCGTGCCAGCAGC
CGCGGTAATACCAAGGGGGCTAGCGTTGTTCGGAATTACTGGGCGTAAAGCGCACGTAGGCGGATATTTAAGTCAGGGGT
GAAATCCCGGGGCTCAACCCCGGAACTGCCTTTGATACTGGGTATCTAGAGTATGGTAGAGGTGAAGTTACTGACGCTGAGGTGCG
AAAGCGTGGGGAGCAAACAGGATTAGATACCCATGTAGTCC

+

参数选项2: 设置 -r 参数, 保留原始序列标识符。

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

@A1_1 HISEQ:483:HLJ2LBCXY:1:1111:14104:10730

CCTATGGGATGCACCAGTGGGGAATATTGGACAATGGGCGAAAGCCTGATCCAGCCATGCCGCTGAGTGATGAAGGCCT
TAGGGTTGTAAAGCTCTTTCGGCGGGGAAGATAATGACGGTACCCGCAGAAGAAGCCCCGGCTAACTTCGTGCCAGCAGC
CGCGGTAATACGAAGGGGGCTAGCGTTGTTCGGAACCACTGGGCGTAAAGCGCGTGTAGGCGGATTGTTAAGTCGGGGGT
GAAATCCTGGGGCTCAACCTCAGAACTGCCTTCGATACTGGCGATCTTGAGTCCGGGAGAGGTGAGTGGTATTCCTAGTG
TAGAGGTGAAATTCGTAGATATTAGGAAGAACACCAGTGGCGAAGGCGGCTCACTGGCCCGGTACTGACGCTGAAACGCG
AAAGCGTGGGGAGCAAACAGGATTAGATACCCGTGTAGTCC

+

@A1_2 HISEQ:483:HLJ2LBCXY:1:2112:7631:97645

CCTACGGGAGGCACCAGTGGGGAATATTGGACAATGGGCGAAAGCCTGATCCAGCCATGCCGCTGAGTGATGAAGGCCC
TAGGGTTGTAAAGCTCTTTCACCGGTGAAGATAATGACGGTAACCGGAGAAGAAGCACCGGCTAACTTCGTGCCAGCAGC
CGCGGTAATACCAAGGGGGCTAGCGTTGTTCGGAATTACTGGGCGTAAAGCGCACGTAGGCGGATATTTAAGTCAGGGGT
GAAATCCCGGGGCTCAACCCCGGAACTGCCTTTGATACTGGGTATCTAGAGTATGGTAGAGGTGAAGTTACTGACGCTGAGGTGCG
AAAGCGTGGGGAGCAAACAGGATTAGATACCCATGTAGTCC

+

参数选项3: 设置 -m 参数, 将映射关系输出标准错误。

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

```
A1_1 HISEQ:483:HLJ2LBCXY:1:1111:14104:10730
A1_2 HISEQ:483:HLJ2LBCXY:1:2112:7631:97645
A1_3 HISEQ:483:HLJ2LBCXY:1:1204:10672:14348
A1_4 HISEQ:483:HLJ2LBCXY:1:1208:10369:65205
A1_5 HISEQ:483:HLJ2LBCXY:1:1110:8032:54809
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