

fastx-utils之合并序列：pseudo

一、fastx-utils pseudo介绍

功能描述：

`fastx-utils pseudo` 对序列集合进行合并操作，合并成一条假的染色体。

命令行接口：

```
1 $ fastx-utils pseudo
```

```
1 Usage: fastx-utils pseudo [options] <fasta>
2 Options:
3   -n padding number of 'N's: default: [150];
4   -l pseudo-sequence label: default: [pseudo];
```

可选参数：

```
1 -n 填充 'N'，默认数量为150；
2 -l 设置序列标签，默认为'pseudo'；
```

二、使用场景实例及其用法

使用场景经典案例：

1. 细菌基因组组装结果，可以通过Mauve通过参考基因组调整Contig的顺序和方向，并合并成一条序列，中间使用 指定 `N` 进行填充。

示例演示：

示例文件： `sequence.fastq`

[illegible]

运行命令: 默认参数: 合并序列集合, 合并成一条假序列, 默认填充150N.

```
1 $ fastx-utils pseudo sequence.fastq | seqtk seq -l 80 | head -n16
```

```

1 >pseudo
2 CCTATGGGATGCACCAGTGGGAATATTGGACAATGGGCGAAAGCCTGATCCAGCCATGCCGCGTGAGTGATGAAG
3 GCCT
4 TAGGGTTGTAAAGCTCTTTCGGCGGGGAAGATAATGACGGTACCCGCAGAAGAAGCCCCGGCTAACTTCGTGCCAG
5 CAGC
6 CGCGGTAATACGAAGGGGCTAGCGTTGTTCGGAACCACTGGGCGTAAAGCGCGTGTAGGCGGATTGTTAAGTCGG
7 GGGT
8 GAAATCCTGGGGCTCAACCTCAGAACTGCCTTCGATACTGGCGATCTTGAGTCCGGGAGAGGTGAGTGGTATTCTT
9 AGTG
10 TAGAGGTGAAATTCGTAGATATTAGGAAGAACACCAGTGGCGAAGGCGGCTCACTGGCCCCGTACTGACGCTGAAA
11 CGCG
12 AAAGCGTGGGGAGCAAACAGGATTAGATACCCGTGTAGTCCNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
13 NNNN
14 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
15 NNNN
16 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
17 CTGA
18 TCCAGCCATGCCGCGTGAGTGATGAAGGCCCTAGGGTTGTAAAGCTCTTTCACCGGTGAAGATAATGACGGTAACC
19 GGAG
20 AAGAAGCACCGGCTAACTTCGTGCCAGCAGCCGCGGTAATACCAAGGGGGCTAGCGTTGTTTCGGAATTACTGGGCG
21 TAAA
22 GCGCACGTAGGCGGATATTTAAGTCAGGGGTGAAATCCCGGGGCTCAACCCCGGAACTGCCTTTGATACTGGGTAT
23 CTAG
24 AGTATGGTAGAGGTGAGTGGAATTCCGAGTGTAGAGGTGAAATTCGTAGATATTCGGAGGAACACCAGTGGCGAAG
25 GCGG
26 CTCACTGGACCATTACTGACGCTGAGGTGCGAAAGCGTGGGGAGCAAACAGGATTAGATACCCATGTAGTCCNNNN
27 NNNN

```

参数选项： 设置 `-l` 参数，修改序列标签为'reads', 设置 `-n` 参数，设置填充'N'的数量为5。

```
1 >reads
2 CCTATGGGATGCACCACTGGGGAATATTGGACAATGGGCGAAAGCCTGATCCAGCCATGCCGCGTGAGTGATGAAG
  GCCT
3 TAGGGTTGTAAAGCTCTTTCGGCGGGGAAGATAATGACGGTACCCGCAGAAGAAGCCCCGGCTAACTTCGTGCCAG
  CAGC
4 CGCGGTAATACGAAGGGGGCTAGCGTTGTTCGGAACCACTGGGCGTAAAGCGCGTGATAGCGGATTGTTAAGTCGG
  GGGT
5 GAAATCCTGGGGCTCAACCTCAGAACTGCCTTCGATACTGGCGATCTTGAGTCCGGGAGAGGTGAGTGGTATTCCT
  AGTG
6 TAGAGGTGAAATTCGTAGATATTAGGAAGAACACCACTGGCGAAGGCGGCTCACTGGCCCCGGTACTGACGCTGAAA
  CGCG
7 AAAGCGTGGGGAGCAAACAGGATTAGATACCCGTGTAGTCCNNNNNCCTACGGGAGGCACCACTGGGGAATATTGG
  ACAA
8 TGGGCGAAAGCCTGATCCAGCCATGCCGCGTGAGTGATGAAGGCCCTAGGGTTGTAAAGCTCTTTCACCGGTGAAG
  ATAA
9 TGACGGTAACCGGAGAAGAAGCACCGGCTAACTTCGTGCCAGCAGCCGCGGTAATACCAAGGGGGCTAGCGTTGTT
  CGGA
10 ATTACTGGGCGTAAAGCGCACGTAGGCGGATATTTAAGTCAGGGGTGAAATCCCGGGGCTCAACCCCGGAACTGCC
  TTTG
11 ATACTGGGTATCTAGAGTATGGTAGAGGTGAGTGGAATTCCGAGTGTAGAGGTGAAATTCGTAGATATTCGGAGGA
  ACAC
12 CAGTGGCGAAGGCGGCTCACTGGACCATTACTGACGCTGAGGTGCGAAAGCGTGGGGAGCAAACAGGATTAGATAC
  CCAT
13 GTAGTCCNNNNNCCTACGGGGGGCACCACTAGGGAATACTGGACAATGGGCGAAAGCCTGATCCAGCCATGCCGCG
  TGAG
14 TGATGAAGGCCCTAGGGTTGTAAAGCTCTTTCAGCGGGGAAGATAATGACGGTACCCGCAGAAGAAGCCCCGGCCA
  ACTT
15 CGTGCCAGCAGCCGCGGTAATACGAAGGGGGCTAGCGTTGTTTCGGAATTACTGGGCGTAAAGCGCGCGCAGGCGGA
  CTGT
16 TAAGTCAGGGGTGAAATCCCAGGGCTCAACCTCGGAATTGCCTTTGATACTGGCAGTCTCGAGTCCGGGAGAGGTG
  AGTG
```

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