fastx-utils之拆分文件：shred

一、fastx-utils shred介绍

功能描述：

`fastx-utils shred` 将长序列拆分成具有 `L` 交叠长度的序列集合。

命令行接口：

```
1  $ fastx-utils shred
2
3  Usage: fastx-utils shred <fasta> <length> <overlap>
```

二、使用场景实例及其用法

使用场景经典案例：

1. 454 Newber基因组组装软件，只支持长度小于2000bp序列，多基因组组装可将Contigs序列打断成有一定长度，比如100bp 交叠的序列群。

示例演示：

示例文件： `CP004075.fasta`

```
1  $ cat CP004075.fasta | seqtk seq -l  80  | head -n 8
```

```
1  >CP004075.1 Bacillus thuringiensis serovar kurstaki str. HD73 plasmid pHT8_2,
   complete sequence
2  ATTTAAAGGGGCGTAAAGCGTGCTACAAGGGACACAAAACGTGCCACAAAAGTGCTTTAAACGTACACAAAACGGGC
   TAC
3  AAAGCCCAAAAAGGCATCCCGTAATTCACAAAATAGAAACCCAATAGGCTCAGCCTATTTCATCACTTTTTTTATAA
   AAT
4  GGGATTCGGTTTCCTTATGCTCTTCTTTAAAAAAAGGTTTATATGGCTTTTTAAGCTGGTAAATATTTTGTATTAAA
   AAA
5  AGAAAATGCAGGATAAGGAGAGGGACACAGTTGTGTCTACGTGTATCAGTCAAAGCCTGATAGAATAATTTTTCATT
   GTG
6  TACAGAGTTGTGACCACGATTGAAAAAGGGAATGCGTTTTGTGGTCACAAAACATTTATAAGGTCTATTTTACTAAT
   GTT
7  TTGATGTAGCCTTACTAATAGATCCGACAGATTAAATAATTAATGTTTGTGGCAATTAGCGAGGAATGGCGAAGATT
   AGG
8  GAAGACATTTTTTGAAGTATATTGGGGCGAAGTATTAAGATTTTTATCCAAGTAATCGTTATAAAAAAATCAAAATG
   AAA
```

运行命令： 模拟 `100bp` 长度, `20bp` 交叠

```
1  $ fastx-utils  shred CP004075.fasta  100  20 | head -n8
2  >CP004075.1_1
3  ATTTAAAGGGGCGTAAAGCGTGCTACAAGGGACACAAAACGTGCCACAAAAGTGCTTTAAACGTACACAAAACGGGC
   TACAAAGCCCAAAAAGGCATCCC
4  >CP004075.1_2
5  AAAGCCCAAAAAGGCATCCCGTAATTCACAAAATAGAAACCCAATAGGCTCAGCCTATTTCATCACTTTTTTTATAA
   AATGGGATTCGGTTTCCTTATGC
6  >CP004075.1_3
7  GGGATTCGGTTTCCTTATGCTCTTCTTTAAAAAAAGGTTTATATGGCTTTTTAAGCTGGTAAATATTTTGTATTAAA
   AAAAGAAAATGCAGGATAAGGAG
8  >CP004075.1_4
9  AGAAAATGCAGGATAAGGAGAGGGACACAGTTGTGTCTACGTGTATCAGTCAAAGCCTGATAGAATAATTTTTCATT
   GTGTACAGAGTTGTGACCACGAT
```

本文材料为 **BASE** (**B**iostack **A**pplied bioinformatic **SE**ies) 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