

# fastx-utils之拆分文件：split

## 一、fastx-utils split介绍

功能描述：

`fastx-utils split` 通过指定序列数目，对文件进行拆分，返回拆分的文件数字。

命令行接口：

```
1 $ fastx-utils split
2
3 Usage: fastx-utils split [options] <fasta/q> <size> <prefix>
4 Options:
5 -f fastx file suffix: default: [fasta];
```

可选参数：

```
1 -f 指定文件后缀，默认为'fasta'
```

## 二、使用场景实例及其用法

使用场景经典案例：

1. 序列相似性搜索，比如HMMER，BLAST等线程并行不理想场景，可以自行对数据集进行拆分，并使用进程提交。

示例演示：

示例文件： `protein.faa`

```
1 $ cat protein.faa | seqtk seq -l 60 | head -n 6
```

```
1 >AGE80385.1
2 MKKIIIIISATTIVIGITSFAYFGSKTPLHNEAKAVESQKHNNHKKEEIPAFPKADHNAKK
3 IDNDFSVVTPKSNLVLINKHRKLPDGYIPEDLTRPNVPFISPKDKEKTLLRKDAAEALE
4 NMFKAAKKEGLDLTAVSGYRSYKRQKSLHDTYVRRQGKAEANSVSAIPGTSEHQTGLAMD
5 ISSKSAKFQLEPIFGETAEGKWVAEHAHEFGFVIRYLEDKTDTTEYAYEPWHLRYVGNPY
6 ATYLYKHHLTLEEAMEDKK
```

运行命令：指定分割的文件序列数目 `10`

```
1 $ fastx-utils split -f fasta protein.faa 10 A1
```

```
1 2
```

```
1 $ ls
```

```
1 A1_1.fasta A1_2.fasta protein.faa
```

**注意事项：** 该程序会返回分割的文件数目。

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