fastx-utils: fasta/q操作通用程序

一、程序介绍

```
`fastx-utils` 为一组处理fasta/q文件的程序集合，主要使用`klib` 进行开发，包括`khash, kvect, kstring`等。
```

二、主程序接口

当前释放版本：`version: 0.0.1-r2`

```
$ fastx-utils
```

```
Usage:   fastx-utils <command> <arguments>
Version: 0.0.1-r2

Command:
      -- FASTA/Q Summary.
         view          extract sequence information.
         fqchk         fastq QC (base/quality summary).
         convert       common transformation of FASTA/Q.
         info          calculate reads counts/base info.
         counts        calculate sequence counts.
         length        extract sequence length.

      -- FASTA/Q retrieve.
         filter        filter PE reads with Ns and Qs.
         dedup         deduplication with sequence label.
         fetch         fetch by specified identifier in command line.
         reorder       reorder sequence by name.
         subseq        extract sequence with specified identifier set.
         uniques       Find uniques sequences.

      -- FASTA/Q Edit.
         fake          convert fasta to fake fastq with qual.
         interleave    interleave two PE FASTA/Q files.
         deinterleave  deinterleave interleaved FASTA/Q file.
         rename        rename sequence identifier with specified prefix
                       and then, _1, _2, _3..., return identifier map.
         label         add a label before/after name to relabel sequence.
         truncate      truncate sequence to specified length L.
         comment       add comment information.
         replace       replace sequence name.
         strip_stop_codon strip stop codon in AA sequence.

      -- FASTA/Q segmentation.
         partition     partition fasta/q file to N files.
         shred         shred long sequence with overlapped sequence.
         split         split fasta/q file with specified size.
         pseudo        make pseudo-sequence from a fasta file.
         demultiplex   fastq demultiplex using index.
```

```
39
40      -- auxiliary utils.
41         revcomp       revcomp DNA sequence.
42         rna2dna       convert RNA to DNA sequence.
```

```
1  Licenced:
2  (c) 2016-2020 - LEI ZHANG
3  Logic Informatics Co.,Ltd.
4  zhanglei@logicinformatics.com
```

三、主要子命令功能介绍

主要子命令功能介绍：

统计操作

1. view:抽取序列文件对应的信息，以表格形式展示；
2. fqchk：统计序列文件的碱基组成以及质量值；
3. convert：转换fastq为fasta文件；
4. info：统计序列数，碱基数，最小长度，最大长度， GCh含量等信息；
5. counts：返回序列数，输出流可以添加注释信息；
6. length：根据长度区域进行过滤，或者显示每条序列的长度信息；

查询操作

7. filter：质量控制，根据'N'的个数和质量值对序列进行过滤；
8. dedup：根据序列的名字去除重复序列；
9. fetch：根据提供的序列ID，获取其序列；
10. reorder:根据提供的序列索引，对序列文件进行排序；
11. subseq：根据提供的序列ID，获取序列集合的交集序列或者补集序列；
12. uniques：根据序列去除重复；

编辑操作

12. fake：使用指定质量值构建fastq文件；
13. interleave：将PE双端序列转成交叠的序列文件；
14. deinterleave：交叠的序列文件转换成PE双端序列；
15. rename:修改序列的名字，统一使用前缀，比如 "ZOTU_"；
16. label：修饰序列的名字，比如添加样本信息， ";sample=X1"；
17. truncate：对序列进行截断操作，可指定序列长度；
18. comment：根据Key/value 键值对，对序列进行注释；
19. replace：根据Key/value 键值对，修改序列的名字；
20. strip_stop_codon： 去除序列末端的终止密码子；

分割操作

21. partition: 指定分割文件数，对序列文件进行拆分，返回拆分的文件数字；
22. shred： 将长序列拆分成具有'L'交叠长度的序列集合；
23. split： 通过制定序列数目，对文件进行拆分，返回拆分的文件数字；
24. pseudo： 对序列集合进行合并操作，合并成一条假的染色体；
25. demultiplex： 根据index数据拆分样本信息；

辅助操作

26. revcomp： 根据提供的序列进行反向互补；

27. rna2dna： 将RNA序列文件转换成；

Last Update: 9/2/2020 12:19:31 AM