

fastx-utils之序列重排: reorder

一、fastx-utils reorder介绍

功能描述:

`fastx-utils reorder` 根据提供的序列索引, 对序列文件进行排序。

命令行接口:

```
1 $ fastx-utils reorder
2
3 Usage: fastx-utils reorder <fasta/q> <list>
```

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: 需要根据给定的顺序, 重新排列 `ZOTU` 代表序列

示例演示:

示例文件: `protein.faa list`

```
1 $ cat protein.faa | head -n 8
```

```
1 >AGE80385.1
2 MKKIIIIISATTIVIGITSFAYFGSKTPLHNEAKAVESQKHNNHKKEEIPAFPKADHNAKKIDNDFS SVVTNP KSNLVL
  INKHRKLPDGYIPEDLTRPNVPFISPKDKEKTLRKDAAEALENMFKA AKKEGLDLTAVSGYRSYKRQKSLHDTYVR
  RQGKAEANSVSAIPGTSEHQTGLAMDISSKSAKFQLEPIFGETAEGKWVAEHAHEFGFVIRYLEDKTDTTEYAYEPW
  HLRYVGNPYATYLYKHHLTLEEAMEDKK
3 >AGE79558.1
4 MINHELVERINFLAKKAKAEGLTEEEQ RERQSLREQYLKGF RQNMLNELKG I KVVNEEGTDVTP TKLKALKKQDNAK
  LN
5 >AGE81073.1
6 MDKVISNEILQQFKDRMRLGDDEDANLRRILFASNKDLTRVCGNYDLNIDEVFKELVFERSRYVYNDALEYFDKNFL
  SQINLSIGKALEAIKLDGD
7 >AGE81522.1
8 MKKTILTTTAALTMMGTGMGINVDHIKPAEVKADTISFYDVPNNHWATKAITNLANRNIVVGNGQGFGGDNVTRG
  QVARMIIYNLKPADAGNFKNPFSDIKGHMFEKEILALAKVGIKGYGEGKFGPDDILTREQMAQVLTNAFKFEGTKK
  TSFVDVDKNSWSYKAIGALEEKGV TIGTGGNMYSPTSVVTREQYSQFLFNSINVIEKETKPEEK PNTGGEVKPEEK P
  NTGGEVKPEEK PNTGEETKPVNIPEWLETSLATNDFTFTQAWYDGSEAINKAASTNAQQIVKNINSKYGTNLKYSEV
  GAIVQLVDGAREQLWLAGMNVNDFRVTF RVSNNAMIELTKELVTLVNSDLNLDQEIQEIPSAPMKIKNVEKG DYKIR
  ISPAMADQM ITIIIEKK
```

```
1 $ cat list | head -n 6
```

```
1 AGE77971.1
2 AGE78522.1
3 AGE76372.1
4 AGE80385.1
5 AGE79558.1
6 AGE81073.1
```

运行命令：默认删除不能匹配的序列。

```
1 $ fastx-utils reorder protein.faa list | head -n 6
2
3 >AGE77971.1
4 MEFQLLVTCILQEGNAYFLVTKVDDVITLKVPIAGVAGLFLALGVPRCS
5 >AGE78522.1
6 MNQQLVYTPLAEPFVMGMSVAVILSVFSSNLLFLSLVFLYVILIGAMHVYIRKVSERVEYSQGSSNVFIGETN
  MCKMKISNKSIFPILNIVFRFKCENKLTWNHDEINKNSTGSNYYMNFNLKGRESVSFDLQAVALKRGIKWEEVE
  IVITDPFGFITNHITYKQVGTPSYLVLPAPVKMQVPELQEWSRGFRKAMSSPLYDETKVMGVKSYENEDFRSIHWS
  ATAKTGTITAKKYERTQSDKYAIYLNQKSGISLRNDTEELIETAGICKQLLMQNCSFEWINSVKDNGLLHIK
  NGHNRKHLQNVLKVLAISDQDTPVSSTYFYTAGFRRKELDAVPLILGTSPKKYSRTNKWVVIKE
7 >AGE76372.1
8 MKLTLIIKEEKNFHLPEFIPARLIRQAPELADIPNNPGPEDMDKMVQYVVKVYGEQFTLDQYWDGVDARKFLSTT
  SDVINAIINETVGAAGGTPGTGEETNPNA
9 [biostack@biostack example]$ fastx-utils reorder protein.faa list | head
  -n 8
10 >AGE77971.1
11 MEFQLLVTCILQEGNAYFLVTKVDDVITLKVPIAGVAGLFLALGVPRCS
12 >AGE78522.1
13 MNQQLVYTPLAEPFVMGMSVAVILSVFSSNLLFLSLVFLYVILIGAMHVYIRKVSERVEYSQGSSNVFIGETN
  MCKMKISNKSIFPILNIVFRFKCENKLTWNHDEINKNSTGSNYYMNFNLKGRESVSFDLQAVALKRGIKWEEVE
  IVITDPFGFITNHITYKQVGTPSYLVLPAPVKMQVPELQEWSRGFRKAMSSPLYDETKVMGVKSYENEDFRSIHWS
  ATAKTGTITAKKYERTQSDKYAIYLNQKSGISLRNDTEELIETAGICKQLLMQNCSFEWINSVKDNGLLHIK
  NGHNRKHLQNVLKVLAISDQDTPVSSTYFYTAGFRRKELDAVPLILGTSPKKYSRTNKWVVIKE
14 >AGE76372.1
15 MKLTLIIKEEKNFHLPEFIPARLIRQAPELADIPNNPGPEDMDKMVQYVVKVYGEQFTLDQYWDGVDARKFLSTT
  SDVINAIINETVGAAGGTPGTGEETNPNA
16 >AGE80385.1
17 MKKIIIIISATTIVIGITSFAYFGSKTPLHNEAKAVESQKHNNHKKEEIPAFPKADHNAKKIDNDFSVVTNPKNLV
  LINKHRKLPDGYIPEDLTRPNVPFISPKDKEKTLRKDAEALENMFKAAKKEGLDLTAVSGYRSYKRQKSLHDTY
  VRRQGKAANSVSAIPGTSEHQTGLAMDISSSAKFQLEPIFGETAEGKWAEHAHEFGFVIRYLEDKTDTEYAY
  EPWHLRYVGNPYATYLYKHHLTLEEAMEDKK
```

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