fastx-utils之获取序列集合的子集或补集：subseq

一、fastx-utils subseq介绍

功能描述：

`fastx-utils 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。

命令行接口：

```
$ fastx-utils subseq

Usage: fastx-utils subseq [options] <in.fa/fq> <name.list>

Options:
  -s   supplementary set option.
```

可选参数：

```
-s      补充设置选项；
```

二、使用场景实例及其用法

示例演示：

示例演示：

示例文件：`protein.faa, list.txt`

```
$ cat protein.faa | head -n 6
```

```
>AGE80385.1
MKKIIIISATTIVIGITSFAYFGSKTPLHNEAKAVESQKHNNHKKEEIPAFPKADHNAKKIDNDFSVVTNPKSNLVL
INKHRKLPDGYIPEDLTRPNVPFISPKDKEKTLLRKDAAEALENMFKAAKKEGLDLTAVSGYRSYKRQKSLHDTYVR
RQGKAEANSVSAIPGTSEHQTGLAMDISSKSAKFQLEPIFGETAEGKWVAEHAHEFGFVIRYLEDKTDTTEYAYEPW
HLRYVGNPYATYLYKHHLTLEEAMEDKK
>AGE79558.1
MINHELVERINFLAKKAKAEGLTEEEQRERQSLREQYLKGFRQNMLNELKGIKVVNEEGTDVTPTKLKALKKQDNAK
LN
>AGE81073.1
MDKVISNEILQQFKDRMRLGDDEDANLRRILFASNKDLTRVCGNYDLNIDEVFKELVFERSRYVYNDALEYFDKNFL
SQINSLSIGKALEAIKLDGD
```

```
$ cat list.txt | head -n 6
```

```
AGE75591.1      chromosomal replication initiation protein [Bacillus thuringiensis serovar kurstaki str. HD73]
AGE75592.1      DNA polymerase III, beta subunit [Bacillus thuringiensis serovar kurstaki str. HD73]
AGE75593.1      hypothetical protein HD73_0003 [Bacillus thuringiensis serovar kurstaki str. HD73]
AGE75594.1      recombination protein F [Bacillus thuringiensis serovar kurstaki str. HD73]
AGE75595.1      DNA gyrase subunit B [Bacillus thuringiensis serovar kurstaki str. HD73]
AGE75596.1      DNA gyrase, A subunit [Bacillus thuringiensis serovar kurstaki str. HD73]
```

运行命令：

参数选项**1**： 默认参数, 求列表序列的交集。

```
$ fastx-utils subseq  protein.faa  list.txt | head -n 6
```

```
>AGE79558.1
MINHELVERINFLAKKAKAEGLTEEEQRERQSLREQYLKGFRQNMLNELKGIKVVNEEGTDVTPTKLKALKKQDNAKLN
>AGE76491.1
MKEMTAKELEEKLLRKEAVNIVDVREVEEVAEGKIPEACNIPLGLLEFRMHELDKKKEYIIVCRSGGRSARAVQFLESYGFQAINMVGGMLAWEGKVV
>AGE77971.1
MEFQLLVTCILQEGNAYFLVTKVDDVITLKVPITAGVAGLFLALGVPRCS
```

参数选项**1**： 设置`-s`, 求列表序列的补集。

```
$ fastx-utils subseq  protein.faa -s list.txt | head -n 6
```

```
>AGE80385.1
MKKIIIISATTIVIGITSFAYFGSKTPLHNEAKAVESQKHNNHKKEEIPAFPKADHNAKKIDNDFSVVTNPKSNLVLINKHRKLPDGYIPEDLTRPNVPFISPKDKEKTLLRKDAAEALENMFKAAKKEGLDLTAVSGYRSYKRQKSLHDTYVRRQGKAEANSVSAIPGTSEHQTGLAMDISSKSAKFQLEPIFGETAEGKWVAEHAHEFGFVIRYLEDKTDTTEYAYEPWHLRYVGNPYATYLYKHHLTLEEAMEDKK
>AGE81073.1
MDKVISNEILQQFKDRMRLGDDEDANLRRILFASNKDLTRVCGNYDLNIDEVFKELVFERSRYVYNDALEYFDKNFLSQINSLSIGKALEAIKLDGD
>AGE81522.1
MKKTILTTTAALTMMGTGMGINVDHIKPAEVKADTISFYDVPNNHWATKAITNLANRNIVVGYGNGQFGFGDNVTRGQVARMIYNYLKPADAGNFKNPFSDIKGHMFEKEILALAKVGIIKGYGEGKFGPDDILTREQMAQVLTNAFKFEGTKKTSFVDVDKNSWSYKAIGALEEKGVTIGTGGNMYSPTSVVTREQYSQFLFNSINVIEKETKPEEKPNTGGEVKPEEKPNTGGEVKPEEKPNTGEETKPVNIPEWLETSLATNDFTFTQAWYDGSEAINKAASTNAQQIVKNINSKYGTNLKYSEVGAIVQLVDGAREQLWLAGMNVNDFRVTFRVSNNAMIELTKELVTLVNSDLNLDQEIQEIPSAPMKIKNVEKGDYKIRISPAMADQMITIIIEKK
```


Last Update: 2020-08-10 11:56 AM