

fastx-utils之抽取序列文件对应的信息：view

一、fastx-utils view介绍

功能描述：

`fastx-utils view` 抽取序列文件对应的信息，以表格形式展示。

命令行接口：

```
1 $ fastx-utils view
2
3 Usage: fastx-utils view [options] <fasta/q>
4
5 Options:
6   -a STR    print additional fileds
7   -c        print comment fileds
8   -l        print length fileds
9   -s        print sequence fileds
```

可选参数：

```
1  -a        输出附加信息；
2  -c        输出注释信息；
3  -l        输出长度信息；
4  -s        输出序列信息；
```

二、使用场景实例及其用法

示例演示：

示例文件： `protein.faa`

```
1 $ cat protein.faa | head -n 6
```

```
1 >AGE80385.1 D-alanyl-D-alanine carboxypeptidase [Bacillus thuringiensis
  serovar kurstaki str. HD73]
2 MKKIIIIISATTIVIGITSFAYFGSKTPLHNEAKAVESQKHNNHKKEEIPAFPKADHNAKKIDNDFS
  VVTNPKSNLVL
  INKHKRLPDGYIPEDLTRPNVPFISPKDKEKTLRKDAEAELENMFKA AKKEGLDLTAVSGYRSYKR
  QKSLHDTYVR
  RQGKAEANSVSAIPGTSEHQTGLAMDISSKSAKFQLEPIFGETAEGKWAEHAHEFGFVIRYLEDKT
  DTTTEYAYEPW
  HLRYVGNPYATYLYKHHLTLEEAMEDKK
3 >AGE79558.1 hypothetical protein HD73_3980 [Bacillus thuringiensis sero
  var kurstaki str. HD73]
4 MINHELVERINFLAKKAKAEGLTEEEQRERQSLREQYLKGRQNMNLKGIKVVNEEGTDVTPTKLK
  ALKKQDNAK
  LN
5 >AGE81073.1 hypothetical protein HD73_5496 [Bacillus thuringiensis sero
  var kurstaki str. HD73]
6 MDKVISNEILQQFKDRMRLGDDEDANLRRILFASNKDLTRVCGNYDLNIDEVFKELVFERSRYVY
  NDALEYFDKNFL
  SQINLSIGKALEAIKLDGD
```

运行命令：显示 `protein.faa` 文件信息。

```
$ fastx-utils view protein.faa | head -n6
```

```
1 AGE80385.1
2 AGE79558.1
3 AGE81073.1
4 AGE81522.1
5 AGE80137.1
6 AGE76491.1
```

参数选项1： 设置 `-c` 参数，输出注释信息。

```
1 $ fastx-utils view -c protein.faa | head -n6
```

```
1 AGE80385.1      D-alanyl-D-alanine carboxypeptidase [Bacillus thuringiensis
serovar kurstaki str. HD73]
2 AGE79558.1      hypothetical protein HD73_3980 [Bacillus thuringiensis
serovar kurstaki str. HD73]
3 AGE81073.1      hypothetical protein HD73_5496 [Bacillus thuringiensis
serovar kurstaki str. HD73]
4 AGE81522.1      S-layer domain-containing protein (plasmid) [Bacillus
thuringiensis serovar kurstaki str. HD73]
5 AGE80137.1      Methylenetetrahydrofolate reductase [Bacillus thuringiensis
serovar kurstaki str. HD73]
6 AGE76491.1      Rhodanese domain protein [Bacillus thuringiensis serovar
kurstaki str. HD73]
```

参数选项2： 设置 `-l` 参数，输出长度信息。

```
1 $ fastx-utils view -l protein.faa | head -n6
```

```
1 AGE80385.1      259
2 AGE79558.1      79
3 AGE81073.1      97
4 AGE81522.1      402
5 AGE80137.1      610
6 AGE76491.1      98
```

参数选项3： 设置 `-s` 参数，输出序列信息。

```
1 $ fastx-utils view -s protein.faa | head -n6
```

```

1 AGE80385.1
  MKKIIIIISATTIVIGITSFAYFGSKTPLHNEAKAVESQKHNNHKKEEIPAFPKADHNAKKIDNDFS SVTNPKSNLVL
  INKHRKLPDGYIPEDLTRPNVPFISPKDKEKTL LRKDAAEALENMFKA AKKEGLDLTAVSGYRSYKRQKSLHDTYVR
  RQGKAEANSVSAIPGTSEHQTGLAMDISSKSAKFQLEPIFGETAEGKWVAEHAHEFGFVIRYLEDKTDTTEYAYEPW
  HLRYVGNPYATYLYKHHLTLEEAMEDKK
2 AGE79558.1
  MINHELVERINFLAKKAKAEGLTEEEQRERQSLREQYLKGFRQNMLNELKGIKVVNEEGDVTPTKLKALKKQDNAK
  LN
3 AGE81073.1
  MDKVISNEILQQFKDRMLGDDDEDANLRRILFASNKDLTRVCGNYDLNIDEVFKELVFERSRYVYNDALEYFDKNFL
  SQINLSIGKALEAIKLDGD
4 AGE81522.1
  MKKTILTTAALTMMGTGMGINVDHIKPAEVKADTISFYDVPNNHWATKAITNLANRNIVVGYGNGQFGFGDNVTRG
  QVARMYNYLKPADAGNFKNPFS DIKGHMFKEILALAKVGIKGYGEGKFGPDDILTREQMAQVLTNAFKFEGTKK
  TSFVDVDKNSWSYKAIGALEEKGVITGTGGNMYSPTSVVTREQYSQFLFNSINVIEKETKPEEKPNTGGEVKPEEK
  NTGGEVKPEEKPNTGEETKPVNIPEWLETSLATNDFTFTQAWYDGEAINKAASTNAQQIVKNINSKYGTNLKYSEV
  GAIVQLVDGAREQLWLAGMNVNDFRVTFRVSNAMI E LTKELVTLVNSDLNLDQEIQEIP SAPMKIKNVEKG DYKIR
  ISPAMADQMITIIIEKK
5 AGE80137.1
  MKLLDLLSKGIVIGDGAVGTLLHSHGLQSSFEELNVSDPDLIISIHKQYVAAGADVIQTNTYGANEA KL RMYGLENQ
  VTKINRAAVKLAKASVTDKNAILGTIGGMKHIGAVTTTDMEREFMLLEQAGALLEQVDGLLLET FYDEFELLHAVK
  VLRKQTNIPIVAQLALHEAGTTQNGNDVNEILKQFIDYGANVVGLNCQLGPLHMT EAFKMISIPQNGYLSAYPNAGL
  PNYVEGRYVYEGSPAYFEAMTPNFIEQGIRLLGGCCGTTPEHIQSMKRAVANITPVIEKETIQRPKV VHTHEKRSKA
  HVTLAEKAKKQTTVVVELDPPKTLDTQRFFEGARALKRAGADAITLADNSLASPRVSNMAMGALLTKHDIPVLTHLT
  CRDHNVI GLQSHLLGLSALGMEEVLALTGD PARVGD FPGATSVYDLSSIELIKMIKEMNDGRSILGKSIGPATRFSV
  GGAFNPHVRHLKAAVKRMERKIDAGAEYFLTQPIYDIKLIEEVYEATKHLEQPIFIGIMPLVSKRNADFLHFEVPGI
  TLPEEIRERMDGHETKEAAIEEGIRISQELVDAAMKYFNGIYLITPFLKYEITEHLVKYVREKQEVKEGIN
6 AGE76491.1
  MKEMTAKELEEKLLRKEAVNIVDVREVEEVAEGKIPEACNIPLGLLEFRMHELD KKKKEYIIVCRSGGRSARAVQFLE
  SYGFQAINMVGGMLAWEGKV

```

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