

atlas-utils之计算非冗余序列: uniques

一、atlas-utils uniques介绍

功能描述:

`atlas-utils uniques` 计算非冗余序列, 兼容**USEARCH**的注释文件。

命令行接口:

```
1 $ atlas-utils uniques
2
3 Usage: div_utils uniques [options] <in.fa/q>
4
5 Options:
6   -a          size annotation
7   -l STR      sequence label, default: [Uniq]
8   -w INT      number of residues per line; 0 for 2^32-1 default:[0]
```

可选参数:

```
1   -a          注释大小;
2   -l 字符串   序列标签, 默认为'Uniq';
3   -w 整数     每行的序列长度, 默认为0, 不换行;
```

二、使用场景实例及其用法

示例演示:

示例文件: `A-1_1.fastq`

```
1 $ cat A-1_1.fastq | head -n 4
```

```
1 @HISEQ:483:HLJ2LBCXY:1:1101:7924:2136 1:N:0:CACCGG
2 CCTATGGGACGCAGCAGTGGGGAATATTGGACAATGGGCGCAAGCCTGATCCAGCCATGCCGCGTGAGTGATGAAGG
3 CCCTAGGGTTGTAAAGCCCTTTTCGGCGGGGAAGATAATGACGGTACCCGCAGAAGAAGCCCCGGCTAACTTCGTGCC
4 AGCAGCCGCGGTAATACGAAGGGGGCTAGCGTTGCTCGGAATTACTGGGCGTAAAGCGCACGTAGGCGGCTTTCTAA
5 GTCGGGGGTGAA
6
7 +
8 HH1CHHHHDFHDH?C<DEG/CEHHIHEHCHECFHICHHHHDHIICHHIEHCHCEGHHIICH?
9 EEHFHHHIECEE?HE?1GHHHHDHHHHHFFGHHHCHHHII?H?HHEHCHF1CGHC0DEHEHHD<<FEHHHIGHIC??
10 HHCGHHHHIHH@FFEEHIHDHCHHHIHIHE@GH-G??EDEGEHIHIICHHC?EHH-6@8-B-
11 =C,5@@FHHHHHHHCEHH#####
```

运行命令:

获得非冗余序列。

```
1 $ atlas-utils uniques A-1_1.fastq | head -n 4
```

```

1 >Uniq1
2 CCTATGGGATGCAGCAGTAAGGAATATTGGTCAATGGACGCAAGTCTGAACCAGCCATGCCGCGTGAAGGATGAATG
  CCCTCAGGGTTGTAAACTTCTTTTATCTGGGAAGAAATCCTCGATTTCCATCGGGGTTGACGGTACCAGAGGAATAA
  GCACCGGCTAACTCCGTGCCAGCAGCCGCGGTAATACGGAGGGTGCAAGCGTTATCCGGATTTACTGGGTTTAAAGG
  GTGCGTAGGCGG
3 >Uniq2
4 CCTATGGGGTGCAGCAGTAAGGAATATTGGTCAATGGACGCAAGTCTGAACCAGCCATGCCGCGTGAAGGATGAATG
  CCCTCAGGGTTGTAAACTTCTTTTATCTGGGAAGAAATCCTCGATTTCCATCGGGGTTGACGGTACCAGAGGAATAA
  GCACCGGCTAACTCCGTGCCAGCAGCCGCGGTAATACGGAGGGTGCAAGCGTTATCCGGATTTACTGGGTTTAAAGG
  GTGCGTAGGCGG

```

参数选项1： 设置 `-a` 参数，添加大小注释。

```

1 $ atlas-utils uniques -a A-1_1.fastq | head -n 4

```

```

1 >Uniq1;size=93;
2 CCTATGGGATGCAGCAGTAAGGAATATTGGTCAATGGACGCAAGTCTGAACCAGCCATGCCGCGTGAAGGATGAATG
  CCCTCAGGGTTGTAAACTTCTTTTATCTGGGAAGAAATCCTCGATTTCCATCGGGGTTGACGGTACCAGAGGAATAA
  GCACCGGCTAACTCCGTGCCAGCAGCCGCGGTAATACGGAGGGTGCAAGCGTTATCCGGATTTACTGGGTTTAAAGG
  GTGCGTAGGCGG
3 >Uniq2;size=76;
4 CCTATGGGGTGCAGCAGTAAGGAATATTGGTCAATGGACGCAAGTCTGAACCAGCCATGCCGCGTGAAGGATGAATG
  CCCTCAGGGTTGTAAACTTCTTTTATCTGGGAAGAAATCCTCGATTTCCATCGGGGTTGACGGTACCAGAGGAATAA
  GCACCGGCTAACTCCGTGCCAGCAGCCGCGGTAATACGGAGGGTGCAAGCGTTATCCGGATTTACTGGGTTTAAAGG
  GTGCGTAGGCGG

```

参数选项2： 设置 `-l` 参数，修改标签名称为 `sequence`。

```

1 $ atlas-utils uniques -l A1 -a A-1_1.fastq | head -n 4

```

```

1 >A11;size=93;
2 CCTATGGGATGCAGCAGTAAGGAATATTGGTCAATGGACGCAAGTCTGAACCAGCCATGCCGCGTGAAGGATGAATG
  CCCTCAGGGTTGTAAACTTCTTTTATCTGGGAAGAAATCCTCGATTTCCATCGGGGTTGACGGTACCAGAGGAATAA
  GCACCGGCTAACTCCGTGCCAGCAGCCGCGGTAATACGGAGGGTGCAAGCGTTATCCGGATTTACTGGGTTTAAAGG
  GTGCGTAGGCGG
3 >A12;size=76;
4 CCTATGGGGTGCAGCAGTAAGGAATATTGGTCAATGGACGCAAGTCTGAACCAGCCATGCCGCGTGAAGGATGAATG
  CCCTCAGGGTTGTAAACTTCTTTTATCTGGGAAGAAATCCTCGATTTCCATCGGGGTTGACGGTACCAGAGGAATAA
  GCACCGGCTAACTCCGTGCCAGCAGCCGCGGTAATACGGAGGGTGCAAGCGTTATCCGGATTTACTGGGTTTAAAGG
  GTGCGTAGGCGG

```

参数选项3： 设置 `-w` 参数，指定每行序列的长度为70。

```

1 $ atlas-utils uniques -w 70 A-1_1.fastq | head

```

```
1 >Uniq1
2 CCTATGGGATGCAGCAGTAAGGAATATTGGTCAATGGACGCAAGTCTGAACCAGCCATGCCGCGTGAAGG
3 ATGAATGCCCTCAGGGTTGTAACTTCTTTTATCTGGAAGAAATCCTCGATTTCCATCGGGGTTGACGG
4 TACCAGAGGAATAAGCACCGGCTAACTCCGTGCCAGCAGCCGCGTAATACGGAGGGTGCAAGCGTTATC
5 CGGATTTACTGGGTTTAAAGGGTGCGTAGGCGG
6 >Uniq2
7 CCTATGGGGTGCAGCAGTAAGGAATATTGGTCAATGGACGCAAGTCTGAACCAGCCATGCCGCGTGAAGG
8 ATGAATGCCCTCAGGGTTGTAACTTCTTTTATCTGGAAGAAATCCTCGATTTCCATCGGGGTTGACGG
9 TACCAGAGGAATAAGCACCGGCTAACTCCGTGCCAGCAGCCGCGTAATACGGAGGGTGCAAGCGTTATC
10 CGGATTTACTGGGTTTAAAGGGTGCGTAGGCGG
```

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: 2020-09-09 11:56 AM