atlas-utils之切除引物序列: primer_strip

一、atlas-utils primer_strip介绍

功能描述：

`atlas-utils primer_strip` 可以根据`USEARCH`的比对结果删除引物序列(包含匹配的引物位置之前/后的序列)，可以识别单端或者双端引物序列信息。

命令行接口：

```
  $ atlas-utils primer_strip

  Usage: atlas-utils primer_strip <oligo> <fastq>

  Note: usearch search oligodb parameters:
query+qstrand+target+qlo+qhi+ql+diffs+tlo+thi

  Options:
    -w INT max shifted width, default, [20]
```

可选参数：

```
    -w  整型 最大匹配宽度，默认为20;
```

二、使用场景实例及其用法

使用场景经典案例：

扩增子数据分析,比如 `454`测序或者 `Ion Torrent` 测序数据, 切除引物序列信息。

示例演示：

示例文件：`primers.fa`，`454.fastq`

```
$ cat primers.fa
```

```
>+
GTGCCAGCMGCCGCGG
>-
CCGTCAATTCMTTTRAGTTT
```

```
$ cat 454.fastq | head -n 4
```

```
@EJFW8:00682:05789
GTGCCAGCAGCCGCGGTAATACGGAGGGTGCAAGCGGTTGAATCGGAATAACTGGGCGTGAAAGCAGCACGCAGGCG
GTTTTGTTAAGTCAGATGTGGAAATCCCCCGGGCTCAACCTGGGAACTGCATCTGATACTGGCAAGCTTGAGTCTCG
TAGAGGGGGGTAGAATTCCAGGTGTAGCGGTGAAATGCGTAGAGATCTGGAGGAATACCGGTGGCGAAGGCGGCCCC
CTGGACGAAGACTGACGCTCAGGTGCGAAAGCGTGGGGAGCAAACAGGATTAGATACCCTGGATACGTCCACGCCGT
AAACGATGTCGACTTGGAGGTTGTGCCCTTGAGGCGTGGCTTCCGGAGCTAACGCGTTAAGTCGACCGCCTGGGGAG
TACGGCCGCAAGGTTAAAACTCAAATGAATTGACGGATCGAATAACCTT
+
4555.5@:90/49//)//39977+//:1///849/4-3-33,33849133,3333(333337.3322222=
<<:7::.444&4:5:7;=<===9988.44(33333$3:1;9975:588;5;+//77::BBCBB?@@@<99+--
-059>>>9449::333333#39991713,3<6;?;;;7703<991=887667785.../)/:)/+/=:.404;:;
<A4;2;75;<<<,<=8777=8>>BBAAA@@@?8939@?998(7<;7777(7377?/74>69959>>7888.88(---
--992..2605.--'----448;88557277(-(---//(67<<=<?;;;7::38385:98;<166;
<<4947/////377178;;/;/5499)/)/--(-181899+---9928;<-3-33445;:4457;:7;)0*
```

运行命令：

```
$ usearch -search_oligodb 454.fastq -db primers.fa          \
  -userout 454.txt -strand both -maxdiffs 2                 \
  -userfields query+qstrand+target+qlo+qhi+ql+diffs+tlo+thi  \
  -log 454.log ;
```

根据`USEARCH`的比对结果删除引物序列，设置参数`-w`，指定最大匹配宽度为20

```
$ atlas-utils  primer_strip -w 20  454.txt  454.fastq | head -n 8
```

```
@EJFW8:00682:05789
TAATACGGAGGGTGCAAGCGGTTGAATCGGAATAACTGGGCGTGAAAGCAGCACGCAGGCGGTTTTGTTAAGTCAGA
TGTGGAAATCCCCCGGGCTCAACCTGGGAACTGCATCTGATACTGGCAAGCTTGAGTCTCGTAGAGGGGGGTAGAAT
TCCAGGTGTAGCGGTGAAATGCGTAGAGATCTGGAGGAATACCGGTGGCGAAGGCGGCCCCCTGGACGAAGACTGAC
GCTCAGGTGCGAAAGCGTGGGGAGCAAACAGGATTAGATACCCTGGATACGTCCACGCCGTAAACGATGTCGACTTG
GAGGTTGTGCCCTTGAGGCGTGGCTTCCGGAGCTAACGCGTTAAGTCGACCGCCTGGGGAGTACGGCCGCAAGGTTA
+
//39977+//:1///849/4-3-33,33849133,3333(333337.3322222=<<:7::.444&4:5:7;=
<===9988.44(33333$3:1;9975:588;5;+//77::BBCBB?@@@<99+--
-059>>>9449::333333#39991713,3<6;?;;;7703<991=887667785.../)/:)/+/=:.404;:;
<A4;2;75;<<<,<=8777=8>>BBAAA@@@?8939@?998(7<;7777(7377?/74>69959>>7888.88(---
--992..2605.--'----448;88557277(-(---//(67<<=<?;;;7::38385:98;<166;
<<4947/////377178;;/;/5499)/)/--(-1818
@EJFW8:00704:05760
TAATACGGAGGATTCAAGCGTTATCCGGATTTATTGGGTTTAAAGGGTGCGTAGGCGGTTTGATAAGTTAGAGGTGA
AATTTCGGGGCTCAACCCTGAACGTGCCTCTAATACTGTTTAGCTAGAGAGTAGTTGCGGTAGGCGGAATGTATGGT
GTAGCGGTGAAATGCTTAGAGATCATACAGAACACCGATTGCGAAGGCAGCTTACCAAACTATATCTGACGTTGAGG
CACGAAAGCGTGGGGAGCAAACAGGATTAGATACCCGTGGTAGTCCACGCAGTAAACGATGATAACTCGTTGTCGGC
GATAACACAGTCGGTGACTAAGCGAAAGCGATAAGTTATCACCTGGGAGTACGTTCGCAAGAATG
+
AB=@A;<79B;@A?AA@DCCF?BBB=B=@@@;@B@DD>CC>CG?ED=BBCEGCC@AA=@B<BBBC@CC?
CB;:4:@BA:AE?CCCC<CCCB>BB=ABA=@@B;;6;CBC@BBBCCDC>A:9:????@AABBB@BDC@BBB?BB?
B@@?@B>>8:<868;=8====8>ABB=BBB;;;>>@@@@??
5::=8=:616;77)/)/986606=7AA9@887>=@??===E=A>=:====CA4878::::/:
<888*8<>8==:>A<<=@@///=9766;;5;====>@@E;?>>>;;;A=AB??@767<=67----
(-665799:5<;;;@D<A>>=>7>?888B:==8:///6/556./5555;;6?888;>>:>>
```

Last Update: 2020-09-11 11:56 AM