fastx-utils之截断序列：truncate

一、fastx-utils truncate介绍

功能描述：

`fastx-utils truncate`对序列进行截断操作，可指定序列长度。

命令行接口：

```
$ fastx-utils truncate

Usage: fastx-utils truncate [options] <fasta/q>

Options:
  -b      Truncate sequence from left side, default: [right]
  -d      Discard sequence short than L , defalt: [0]
  -l INT  Truncate sequence length to length 'L', defalt: [NONE]
```

可选参数：

```
-b          从左侧截断序列，默认为右侧；
-d          过滤掉序列长度低于指定值的序列，默认为0；
-l          截断序列到指定长度，默认为空；
```

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： 针对454数据分析，对序列进行截断以及过滤过短序列

示例演示：

示例文件： `sequence.fastq`

```
$ cat  sequence.fastq | head -n 6
```

```
@EJFW8:00682:05789
TAATACGGAGGGTGCAAGCGGTTGAATCGGAATAACTGGGCGTGAAAGCAGCACGCAGGCGGTTTTGTTAAGTCAGATGTGGAAATCCCCCGGGCTCAACCTGGGAACTGCATCTGATACTGGCAAGCTTGAGTCTCGTAGAGGGGGGTAGAATTCCAGGTGTAGCGGTGAAATGCGTAGAGATCTGGAGGAATACCGGTGGCGAAGGCGGCCCCCTGGACGAAGACTGACGCTCAGGTGCGAAAGCGTGGGGAGCAAACAGGATTAGATACCCTGGATACGTCCACGCCGTAAACGATGTCGACTTGGAGGTTGTGCCCTTGAGGCGTGGCTTCCGGAGCTAACGCGTTAAGTCGACCGCCTGGGGAGTACGGCCGCAAGGTTA
+
//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
TAATACGGAGGATTCAAGCGTTATCCGGATTTATTGGGTTTAAAGGGTGCGTAGGCGGTTTGATAAGTTAGAGGTGAAATTTCGGGGCTCAACCCTGAACGTGCCTCTAATACTGTTTAGCTAGAGAGTAGTTGCGGTAGGCGGAATGTATGGTGTAGCGGTGAAATGCTTAGAGATCATACAGAACACCGATTGCGAAGGCAGCTTACCAAACTATATCTGACGTTGAGGCACGAAAGCGTGGGGAGCAAACAGGATTAGATACCCGTGGTAGTCCACGCAGTAAACGATGATAACTCGTTGTCGGCGATAACACAGTCGGTGACTAAGCGAAAGCGATAAGTTATCACCTGGGAGTACGTTCGCAAGAATG
```

运行命令：

参数选项1： 设定`-l`参数, 从右侧截断序列，使得最后序列长度为`300`。

```
$ fastx-utils truncate  -l 300  sequence.fastq | head -n 6
```

```
@EJFW8:00682:05789
TAATACGGAGGGTGCAAGCGGTTGAATCGGAATAACTGGGCGTGAAAGCAGCACGCAGGCGGTTTTGTTAAGTCAGATGTGGAAATCCCCCGGGCTCAACCTGGGAACTGCATCTGATACTGGCAAGCTTGAGTCTCGTAGAGGGGGGTAGAATTCCAGGTGTAGCGGTGAAATGCGTAGAGATCTGGAGGAATACCGGTGGCGAAGGCGGCCCCCTGGACGAAGACTGACGCTCAGGTGCGAAAGCGTGGGGAGCAAACAGGATTAGATACCCTGGATACGTCCACGCCGTAAACGATG
+
//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.--'----4
@EJFW8:00704:05760
TAATACGGAGGATTCAAGCGTTATCCGGATTTATTGGGTTTAAAGGGTGCGTAGGCGGTTTGATAAGTTAGAGGTGAAATTTCGGGGCTCAACCCTGAACGTGCCTCTAATACTGTTTAGCTAGAGAGTAGTTGCGGTAGGCGGAATGTATGGTGTAGCGGTGAAATGCTTAGAGATCATACAGAACACCGATTGCGAAGGCAGCTTACCAAACTATATCTGACGTTGAGGCACGAAAGCGTGGGGAGCAAACAGGATTAGATACCCGTGGTAGTCCACGCAGTAAACGATGATAACTCG
```

参数选项2： 设置 `-b` 参数，从左侧开始截断序列到`300`。

```
$ fastx-utils truncate -b  -l 300  sequence.fna | head -n 6
```

```
@EJFW8:00682:05789
TCCCCCGGGCTCAACCTGGGAACTGCATCTGATACTGGCAAGCTTGAGTCTCGTAGAGGGGGGTAGAATTCCAGGTG
TAGCGGTGAAATGCGTAGAGATCTGGAGGAATACCGGTGGCGAAGGCGGCCCCCTGGACGAAGACTGACGCTCAGGT
GCGAAAGCGTGGGGAGCAAACAGGATTAGATACCCTGGATACGTCCACGCCGTAAACGATGTCGACTTGGAGGTTGT
GCCCTTGAGGCGTGGCTTCCGGAGCTAACGCGTTAAGTCGACCGCCTGGGGAGTACGGCCGCAAGGTTA
+
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
GTGAAATTTCGGGGCTCAACCCTGAACGTGCCTCTAATACTGTTTAGCTAGAGAGTAGTTGCGGTAGGCGGAATGTA
TGGTGTAGCGGTGAAATGCTTAGAGATCATACAGAACACCGATTGCGAAGGCAGCTTACCAAACTATATCTGACGTT
GAGGCACGAAAGCGTGGGGAGCAAACAGGATTAGATACCCGTGGTAGTCCACGCAGTAAACGATGATAACTCGTTGT
CGGCGATAACACAGTCGGTGACTAAGCGAAAGCGATAAGTTATCACCTGGGAGTACGTTCGCAAGAATG
```

参数选项3： 设置 `-d` 参数，过滤掉序列长度低于100的序列。

```
$ fastx-utils truncate -d  -l 300  sequence.fastq   | head -n 6
@EJFW8:00682:05789
TAATACGGAGGGTGCAAGCGGTTGAATCGGAATAACTGGGCGTGAAAGCAGCACGCAGGCGGTTTTGTTAAGTCAGA
TGTGGAAATCCCCCGGGCTCAACCTGGGAACTGCATCTGATACTGGCAAGCTTGAGTCTCGTAGAGGGGGGTAGAAT
TCCAGGTGTAGCGGTGAAATGCGTAGAGATCTGGAGGAATACCGGTGGCGAAGGCGGCCCCCTGGACGAAGACTGAC
GCTCAGGTGCGAAAGCGTGGGGAGCAAACAGGATTAGATACCCTGGATACGTCCACGCCGTAAACGATG
+
//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.--'----4
@EJFW8:00704:05760
TAATACGGAGGATTCAAGCGTTATCCGGATTTATTGGGTTTAAAGGGTGCGTAGGCGGTTTGATAAGTTAGAGGTGA
AATTTCGGGGCTCAACCCTGAACGTGCCTCTAATACTGTTTAGCTAGAGAGTAGTTGCGGTAGGCGGAATGTATGGT
GTAGCGGTGAAATGCTTAGAGATCATACAGAACACCGATTGCGAAGGCAGCTTACCAAACTATATCTGACGTTGAGG
CACGAAAGCGTGGGGAGCAAACAGGATTAGATACCCGTGGTAGTCCACGCAGTAAACGATGATAACTCG
```

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