

fastx-utils之显示序列的长度信息：length

一、fastx-utils length介绍

功能描述：

`fastx-utils length` 根据长度区域进行过滤，或者显示每条序列的长度信息。

命令行接口：

```
1 $ fastx-utils length
2
3 Usage: fastx-utils length [options] <fasta/q>
4
5 Options:
6   -s INT   Discard sequence short than S , default: [0]
7   -l INT   Discard sequence longer than L, default: [0]
```

可选参数：

```
1   -s  整数  过滤掉序列长度低于指定数值的序列，默认为0；
2   -l  整数  过滤掉序列长度高于指定数值的序列，默认为0；
```

二、使用场景实例及其用法

示例文件：`sequence.fastq`

```
1 $ cat sequence.fastq | head -n 6
```

```
1 @EJFW8:00682:05789
2 TAATACGGAGGGTGCAAGCGTTGAATCGGAATAACTGGGCGTGAAAGCAGCACGCAGGCGGTTTTGTAAAGTCAGA
TGTGGAATCCCCGGGCTCAACCTGGGAAGTGCATCTGATACTGGCAAGCTTGAGTCTCGTAGAGGGGGTAGAAT
TCCAGGTGTAGCGGTGAAATGCGTAGAGATCTGGAGGAATACCGGTGGCGAAGGCGGCCCTTGACGAAGACTGAC
GCTCAGGTGCGAAAGCGTGGGAGCAAACAGGATTAGATACCTGGATACGTCCACGCCGTAAACGATGTCGACTTG
GAGGTTGTGCCCTTGAGGCGTGGCTTCCGGAGCTAACGCGTTAAGTCGACCGCCTGGGGAGTACGGCCGCAAGGTTA
3 +
4 //39977+//:1///849/4-3-33,33849133,3333(333337.332222=<=: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
5 @EJFW8:00704:05760
6 TAATACGGAGGATTCAAGCGTTATCCGATTTATTGGGTTTAAAGGGTGCGTAGGCGGTTTGATAAGTTAGAGGTGA
AATTTCCGGGGCTCAACCCTGAACGTGCCTCTAATACTGTTTAGCTAGAGAGTAGTTGCGGTAGGCGGAATGTATGGT
GTAGCGGTGAAATGCTTAGAGATCATACAGAACACCGATTGCGAAGGCAGCTTACCAAATATATCTGACGTTGAGG
CACGAAAGCGTGGGAGCAAACAGGATTAGATACCGTGGTAGTCCACGCAGTAAACGATGATAACTCGTTGTCCGC
GATAACACAGTCGGTGAATAAGCGAAAGCGATAAGTTATCACCTGGGAGTACGTTGCAAGAATG
```

运行命令：显示 `sequence.fastq` 文件的序列长度信息。

```
1 $ fastx-utils length sequence.fastq | head -n 6
```

```
1 EFW8:00682:05789      385
2 EFW8:00704:05760      373
3 EFW8:00713:05834      378
4 EFW8:00675:05875      137
5 EFW8:00683:05878      303
6 EFW8:00665:05958      376
```

参数选项： 设置 `-l` 和 `-s` 参数，寻找序列长度在一个区间的序列集合。

```
1 $ fastx-utils length -s 300 -l 380 sequence.fastq | head -n6
```

```
1 @EFW8:00704:05760
2 TAATACGGAGGATTCAACGTTATCCGATTTATTGGGTTTAAAGGGTGCCTAGGCGGTTTGATAAGTTAGAGGTGA
AATTTTCGGGGCTCAACCCTGAACGTCCTCTAATACTGTTTAGCTAGAGAGTAGTTGCGGTAGGCGGAATGTATGGT
GTAGCGGTGAAATGCTTAGAGATCATACAGAACACCGATTGCGAAGGCAGCTTACCAAACTATATCTGACGTTGAGG
CACGAAAGCGTGGGAGCAAACAGGATTAGATACCCGTGGTAGTCCACGCAGTAAACGATGATAACTCGTTGTCCGC
GATAACACAGTCGGTGACTAAGCGAAAGCGATAAGTTATCACCTGGGAGTACGTTCGCAAGAATG
3 +
4 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;>>:>>
5 @EFW8:00713:05834
6 TAAACGTAGGTACACAAGCGTTGTCGGAATTACTGGGTGTAAAGGGAGCGCAGGCGGGAAGACAAGTTGGAAGTG
AAATCCATGGGCTCAACCCATGAACTGCTTTCAAACCTGTTTTCTTGAGTAGTGCAGAGGTAGGCGGAATCCCGG
TGTAGCGGTGGAATGCGTAGATATCGGGAGGAACACAGTGGCGAAGGCGGCCTACTGGGCACCAACTGACGCTGAG
GCTCGAAAGTGTGGGTAGCAAACAGGATTAGATACCCTGGTAGTCCACACCGTAAACGATGATTACTAGGTGTTGGA
GGATCGACCCCTTCAGTGCCGCAGTTAACACAATAAGTAATCCACCTGGGGAGTACGACCGAAGGTTGA
```

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