

# atlas-utils之：orient

## 一、atlas-utils orient介绍

功能描述：

`atlas-utils orient` 根据数据库信息,将序列调整和数据库序列方向一致

命令行接口：

```
1 $ atlas-utils orient
2
3 Usage: atlas-utils orient [options] <db> <fasta/q>
4
5 Options:
6 -r          retain sequence without strand information.
```

可选参数：

```
1 -r 保留序列去除标准化信息；
```

## 二、使用场景实例及其用法

使用场景经典案例：

扩增子荟萃分析, 根据数据库信息将序列调整方向一致

示例演示：

示例文件： `454.fastq`, `454.txt`

```
1 $ cat 454.fastq | head -n 4
```

```
1 @EJFW8:00682:05789
2 GTGCCAGCAGCCGCGGTAATACGGAGGGTGCAAGCGGTTGAATCGGAATAACTGGGCGTGAAAGCAGCACGCAGGCG
3 GTTTTGTTAAGTCAGATGTGAAATCCCCGGGCTCAACCTGGGAAGCTGCATCTGATACTGGCAAGCTTGAGTCTCG
4 TAGAGGGGGGTAGAATTCCAGGTGTAGCGGTGAAATGCGTAGAGATCTGGAGGAATACCGGTGGCGAAGGCGGCCCC
5 CTGGACGAAGACTGACGCTCAGGTGCGAAAGCGTGGGGAGCAAACAGGATTAGATACCCTGGATACGTCCACGCCGT
6 AAACGATGTCGACTTGAGGTTGTGCCCTTGAGGCGTGGCTTCCGGAGCTAACGCGTTAAGTCGACCGCCTGGGGAG
7 TACGGCCGCAAGGTTAAAACTCAAATGAATTGACGGATCGAATAACCTT
8
9 +
10 4555.5@:90/49//)//39977+//:1//849/4-3-33,33849133,3333(333337.3322222=
11 <<:7::.444&4:5:7;=<===9988.44(33333$3:1;9975:588;5;+//77::BBCBB?@@@<99+--
12 -059>>>9449::333333#39991713,3<6;?;;;7703<991=887667785.../)/:)/+/:.404;::
13 <A4;2;75;<<<,=<8777=8>>BBAAA@@@?8939@?998(7<;7777(7377?/74>69959>>7888.88(---
14 --992..2605.--'----448;88557277(-(---/(67<<=<?;;;7::38385:98;<166;
15 <<4947/////377178;;;/;/5499)/)/--(-181899+---9928;<-3-33445;.4457;.7;)0*
```

```
1 $ cat 454.txt | head -n 6
```

1	EJFW8:00702:05923	+
2	EJFW8:00674:06008	-
3	EJFW8:00698:06058	+
4	EJFW8:00675:05875	-
5	EJFW8:00683:05878	+
6	EJFW8:00689:05965	+

运行命令:

文件 454.txt 可使用 `usearch_global` 生成, 或其它软件, 比如 `minimap2`.

```

1 $ usearch \
2   -usearch_global 454.fastq \
3   -db rdp_16s_v16_sp.fasta \
4   -threads 10 \
5   -id 0.5 \
6   -maxhits 1 \
7   -strand both \
8   -userfields query+qstrand \
9   -userout 454.txt \
10  -log 454.log

```

```
1 $ atlas-utils orient -r 454.txt 454.fastq | head -n 8
```

```

1 @EJFW8:00682:05789
2 GTGCCAGCAGCCGCGTAAATACGGAGGGTGCAAGCGTTGAATCGGAATAACTGGGCGTGAAAGCAGCACGCAGGCG
  GTTTTGTTAAGTCAGATGTGGAAATCCCCCGGGCTCAACCTGGGAACTGCATCTGATACTGGCAAGCTTGAGTCTCG
  TAGAGGGGGGTAGAATTCCAGGTGTAGCGGTGAAATGCGTAGAGATCTGGAGGAATACCGGTGGCGAAGCGGCCCC
  CTGGACGAAGACTGACGCTCAGGTGCGAAAGCGTGGGGAGCAAACAGGATTAGATACCCTGGATACGTCCACGCCGT
  AAACGATGTGCACTTGGAGGTTGTGCCCTTGAGGCGTGGCTTCCGGAGCTAACGCGTTAAGTCGACCGCTGGGGAG
  TACGGCCGCAAGGTTAAACTCAAATGAATTGACGGATCGAATAACCTT
3 +
4 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*
5 @EJFW8:00704:05760
6 GTGCCAGCAGCCGCGTAAATACGGAGGATTCAAGCGTTATCCGATTTATTGGGTTTAAAGGGTGCGTAGCGGTTT
  GATAAGTTAGAGGTGAAATTTCTGGGGCTCAACCCTGAACGTGCCTCTAATACTGTTTAGCTAGAGAGTAGTTGCGGT
  AGGCGGAATGTATGGTGTAGCGGTGAAATGCTTAGAGATCATACAGAACACCGATTGCGAAGGCAGCTTACCAAAT
  ATATCTGACGTTGAGGCACGAAAGCGTGGGGAGCAAACAGGATTAGATACCCGTGGTAGTCCACGCAGTAAACGATG
  ATAACGTTGTCGCGGATAACACAGTCGGTGACTAAGCGAAAGCGATAAGTTATCACCTGGGAGTACGTTGCAAG
  AATGAACTTAAAGGAATTGACGGATCGAATAACCTT
7 +
8 56<<6B;;:::29A@=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;>>:>>>A7>>;?A7A9@9?8==>>;>BC?>;;5;:>9

```

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