

atlas-utils之链接PE双端序列：linkpairs

一、atlas-utils linkpairs介绍

功能描述：

`atlas-utils linkpairs` 链接PE双端序列，中间填充指定数量 `N`。

命令行接口：

```
$ atlas-utils linkpairs

Usage: atlas-utils linkpairs [options] <in1.fq> <in2.fq>

Options:
  -n INT    inserting padding number of 'N's, default:[8]
```

可选参数：

`-n` 整数 序列中间填充的 'N' 的数量，默认为8；

二、使用场景实例及其用法

示例演示：

示例文件： `A-1_1.fastq` , `A-1_2.fastq`

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

```
@HISEQ:483:HLJ2LBCXY:1:1101:7924:2136 1:N:0:CACCGG
CCTATGGGACGCAGCAGTGGGGAATATTGGACAATGGGCGCAAGCCTGATCCAGCCATGCCGCGTGAGTGATGAAGGCCC
TAGGGTTGTAAAGCCCTTTTCGGCGGGGAAGATAATGACGGTACCCGCAGAAGAAGCCCCGGCTAACTTCGTGCCAGCAGC
CGCGGTAATACGAAGGGGCTAGCGTTGCTCGGAATTACTGGGCGTAAAGCGCACGTAGGCGGCTTTCTAAGTCGGGGGT
GAA
+
HH1CHHHHDFHDHDH?C<DEG/CEHHIHEHCHECFHICHHHHHHDHIICHHIEHCHCEGHHIICH?EEHFHHHIECEE?HE?
1GHHHHDDHHHHHFFGHHHCHHHII?H?
```

```
$ cat A-1_2.fastq | head -n 4
```

```
@HISEQ:483:HLJ2LBCXY:1:1101:7924:2136 2:N:0:CACCGG
GGACTACTTGGGTATCTAATCCTGTTTGCTCCCCACGCTTTTCGCACCTCAGCGTCAGTACCGGGCCAGTGAGCCGCCTTC
GCCACTGGTGTTCTTTCGGAATATCTACGAGTTTCACCTCTACACTCGCAGTTCCACTCACCTCTCCCGGACTCGAGCTCT
CCAGTATCGAAGGCAGTTCTGGAGTTAAGCTCCAGGAGTTCACCCCCGACTTAGAAAGCCGCCTACGTGCGCTTTACGCC
CAGGC
+
HG1EG1F?CGEHGHCHH@GC1<@GEHE?GHHHIIHCHHHHEHE=EFHIIHIGED@GHHFH=HD?CCG1D1C?
EDCCHCG10C00DGH@CCGHHIGD/C/CECHHEHIIGHECEHHHHHHIIGHGFH@HIHEEHCGEHFHDGHHIHEH,CCHH
ECHHHIHHIGFABGC@EHICHHHFHHEHH?AG.AFBCHCAHHIHH-
8B@C-.8>=>+6CEC@G@@@EHHCDDH=CH@D+5@EEF--7CEG##
```

运行命令： 链接 [PE](#) 双端序列。

```
$ atlas-utils linkpairs A-1_1.fastq A-1_2.fastq | head -n 4
```

```
@HISEQ:483:HLJ2LBCXY:1:1101:7924:2136
CCTATGGGACGCAGCAGTGGGGAATATTGGACAATGGGCGCAAGCCTGATCCAGCCATGCCGCGTGAGTGATGAAGGCC
TAGGGTTGTAAAGCCCTTTCGGCGGGGAAGATAATGACGGTACCCGAGAAGAAGCCCCGGCTAACTTCGTGCCAGCAGC
CGCGGTAATACGAAGGGGCTAGCGTTGCTCGGAATTACTGGGCGTAAAGCGCACGTAGGCGGCTTTCTAAGTCGGGGGT
GAANNNNNNNNNGCCTGGGCGTAAAGCGCACGTAGGCGGCTTTCTAAGTCGGGGGTGAACTCCTGGAGCTTAAGTCCAGAA
CTGCCTTCGATACTGGAGAGCTCGAGTCCGGGAGAGGTGAGTGGAAGTGCAGTGTAGAGGTGAACTCGTAGATATTCG
CAAGAACACCAGTGGCGAAGGCGGCTACTGGCCCGTACTGACGCTGAGGTGCGAAAGCGTGGGGAGCAACAGGATTA
GATACCAAGTAGTCC
+
HH1CHHHHDFHDH?C<DEG/CEHHIHEHCHECFHICHHHHHHDHICHHIEHCCEGHHIICH?EEHFHHHIECEE?HE?
1GHHHHHHHHHFFGHHHCHHHII?H?HHEHCHF1CGHC0DEHEHHD<<FEHHHGHIC??
HHCCHHHHHIHH@FFEEHIDHHCHHHIHE@GH-G??EDEGEHIIICHHC?EHH-6@8-B-
=C,5@@FHHHHHHHCEHH#####IIIIIIII##GEC7--
FEE@5+D@@EHC=HDCHHHE@@G@CEC6+>=>8.-C@B8-HHHHACHCBFA.GA?
HHEHHFHHHCIE@CGBAFGIHHHCEHHCC,HEHIIHGDFHFHEGCHEEHIH@HFGHGIHHHHHECEHGIIEHHCE
C/C/DGIHHGCC@HGD00C01GCHCCDE?C1D1GCC?DH=HFHHG@CEDEGIHIFE=EHEHHHHCHHHHHG?
EHEG@<1CG@HHCHGHEGC?F1GE1GH
```

参数选项1： 设置 `-n` 参数，修改中间填充的 N 的数量。

```
$ atlas-utils linkpairs -n 20 A-1_1.fastq A-1_2.fastq | head -n 4
```

```
@HISEQ:483:HLJ2LBCXY:1:1101:7924:2136
CCTATGGGACGCAGCAGTGGGGAATATTGGACAATGGGCGCAAGCCTGATCCAGCCATGCCGCGTGAGTGATGAAGGCC
TAGGGTTGTAAAGCCCTTTCGGCGGGGAAGATAATGACGGTACCCGAGAAGAAGCCCCGGCTAACTTCGTGCCAGCAGC
CGCGGTAATACGAAGGGGCTAGCGTTGCTCGGAATTACTGGGCGTAAAGCGCACGTAGGCGGCTTTCTAAGTCGGGGGT
GAANNNNNNNNNNNNNNNNNNNNNGCCTGGGCGTAAAGCGCACGTAGGCGGCTTTCTAAGTCGGGGGTGAACTCCTGGAGC
TTAACTCCAGAAGTGCCTTCGATACTGGAGAGCTCGAGTCCGGGAGAGGTGAGTGGAAGTGCAGTGTAGAGGTGAACT
CGTAGATATTCGAAGAACACCAGTGGCGAAGGCGGCTACTGGCCCGTACTGACGCTGAGGTGCGAAAGCGTGGGGAG
CAAACAGGATTAGATACCAAGTAGTCC
+
HH1CHHHHDFHDH?C<DEG/CEHHIHEHCHECFHICHHHHHHDHICHHIEHCCEGHHIICH?EEHFHHHIECEE?HE?
1GHHHHHHHHHFFGHHHCHHHII?H?HHEHCHF1CGHC0DEHEHHD<<FEHHHGHIC??
HHCCHHHHHIHH@FFEEHIDHHCHHHIHE@GH-G??EDEGEHIIICHHC?EHH-6@8-B-
=C,5@@FHHHHHHHCEHH#####IIIIIIIIIIIIIIIIIIII##GEC7--
FEE@5+D@@EHC=HDCHHHE@@G@CEC6+>=>8.-C@B8-HHHHACHCBFA.GA?
HHEHHFHHHCIE@CGBAFGIHHHCEHHCC,HEHIIHGDFHFHEGCHEEHIH@HFGHGIHHHHHECEHGIIEHHCE
C/C/DGIHHGCC@HGD00C01GCHCCDE?C1D1GCC?DH=HFHHG@CEDEGIHIFE=EHEHHHHCHHHHHG?
EHEG@<1CG@HHCHGHEGC?F1GE1GH
```

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

