

# atlas-utils之标注序列文件的注释信息：label

## 一、atlas-utils label介绍

功能描述：

`atlas-utils label` 标注序列文件的注释信息。

命令行接口：

```
$ atlas-utils label

Usage: atlas-utils label [options] <db> <fasta>
Options:
  -p string  string padding to headline [default ' '].
  -r          discard squence without label.
```

可选参数：

-p 字符串	填充指定字符到标题,默认为' ';
-r	过滤掉没有标签的序列;

## 二、使用场景实例及其用法

使用场景经典案例：

示例演示

示例文件：

```
$ cat classify.txt|head -n 6
```

```
ZOTU_19
d:Bacteria,p:"Bacteroidetes",c:Sphingobacteriia,o:"Sphingobacteriales",f:Chitino
phagaceae,g:Parafilimonas,s:Parafilimonas_terrae
ZOTU_5  d:Bacteria,p:"Bacteroidetes"
ZOTU_9
d:Bacteria,p:"Bacteroidetes",c:Flavobacteriia,o:"Flavobacteriales",f:Flavobacter
iaceae,g:Flavobacterium
ZOTU_12
d:Bacteria,p:"Bacteroidetes",c:"Bacteroidia",o:"Bacteroidales",f:"Porphyromonada
ceae",g:Petrimonas,s:Petrimonas_sulfuriphila
ZOTU_7
d:Bacteria,p:"Bacteroidetes",c:"Bacteroidia",o:"Bacteroidales",f:"Porphyromonada
ceae"
ZOTU_18
d:Bacteria,p:"Bacteroidetes",c:"Bacteroidia",o:"Bacteroidales",f:"Porphyromonada
ceae",g:Macellibacteroides,s:Macellibacteroides_fermentans
```

```
$ cat zotus.fasta | head -n 6
```

```
>ZOTU_1
TGGGGAATATTGCACAATGGACGAAAGTCTGATGCAGCAACGCCGCGTGGAGGATGACACATTTTCGGTGCGTAAACTCCT
TTTATATAGGAAGATAATGACGGTACTATATGAATAAGCACCGGCTAACTCCGTGCCAGCAGCCGCGGTAATACGGAGGG
TGCAAGCGTTACTCGGAATCACTGGGCGTAAAGAGCGGTGATAGCGGGTATATAAGTCAGAAAGTAAATCCAATAGCTTAA
CTATTGAACTGCTTTTGAAACTGTATACCTAGAATGTGGGAGAGGTAGATGGAATTTCTGGTGTAGGGGTAAAATCCGTA
GAGATCAGAAGGAATACCGATTGCGAAGGCGATCTACTGGAACATTATTGACGCTGAGACGCGAAAGCGTGGGGAGCAAA
CAGG
>ZOTU_2
TGGGGAATATTGGACAATGGGGGAACCTGATCCAGCCATGCCGCGTGTGTGAAGAAGGCCCTCGGGTTGTAAAGCACT
TTCAGTGAGGAAGAACGCCTGGTGGTTAATACCCATCAGGAAAGACATCACTCACAGAAGAAGCACCGGCTAACTCCGTG
CCAGCAGCCGCGGTAATACGGAGGGTGCGAGCGTTAATCGGAATTACTGGGCGTAAAGCGCGCGTAGGTGGCTTGATAAG
CCGGTTGTGAAAGCCCCGGGCTCAACCTGGGAACGGCATCCGGAAGTGTCAAGCTAGAGTGCAGGAGAGGAAGGTAGAAT
TCCCGGTGTAGCGGTGAAATGCGTAGAGATCGGGAGGAATACCAAGTGGCGAAGGCGGCCTTCTGGACTGACACTGACACT
GAGGTGCGAAAGCGTGGGTAGCAAACAGG
>ZOTU_3
TGGGGAATATTGGACAATGGGCGCAAGCCTGATCCAGCAATGCCGCGTGTGTGAAGAAGGTTTTTCGGATCGTAAAGCACT
TTCGGCGGGGACGATGATGACGGTACCCGAGAGAAGCCCCGGCTAACTTCGTGCCAGCAGCCGCGGTAATACGAAGGG
GGCTAGCGTTGCTCGGAATTACTGGGCGTAAAGGCGCGTAGGCGGCTTGGCCAGTCAGGCGTGAAATTCCTGGGCTCAA
CCTGGGACTGCGCTTGATACTGCTGAGCTTGAGCCGAGAGAGGGTGGTGAATTCACAGTGTAGAGGTGAAATTCGTA
GATATTGGGAAGAACACCGGTGGCGAAGGCGACCACCTGGCTCGTACTGACGCTGAGGCGCGACAGCGTGGGGAGCAAA
CAGG
```

## 运行命令:

```
$ atlas-utils label ./label/example/classify.txt ./label/example/zotus.fasta |
head -n 4
```

```
>ZOTU_1
d:Bacteria,p:"Proteobacteria",c:Epsilonproteobacteria,o:Campylobacterales,f:Campylobacteraceae,g:Arcobacter
TGGGGAATATTGCACAATGGACGAAAGTCTGATGCAGCAACGCCGCGTGGAGGATGACACATTTTCGGTGCGTAAACTCCT
TTTATATAGGAAGATAATGACGGTACTATATGAATAAGCACCGGCTAACTCCGTGCCAGCAGCCGCGGTAATACGGAGGG
TGCAAGCGTTACTCGGAATCACTGGGCGTAAAGAGCGGTGATAGCGGGTATATAAGTCAGAAAGTAAATCCAATAGCTTAA
CTATTGAACTGCTTTTGAAACTGTATACCTAGAATGTGGGAGAGGTAGATGGAATTTCTGGTGTAGGGGTAAAATCCGTA
GAGATCAGAAGGAATACCGATTGCGAAGGCGATCTACTGGAACATTATTGACGCTGAGACGCGAAAGCGTGGGGAGCAAA
CAGG
>ZOTU_2
d:Bacteria,p:"Proteobacteria",c:Gammaproteobacteria,o:Oceanospirillales,f:Halomonadaceae,g:Halomonas
TGGGGAATATTGGACAATGGGGGAACCTGATCCAGCCATGCCGCGTGTGTGAAGAAGGCCCTCGGGTTGTAAAGCACT
TTCAGTGAGGAAGAACGCCTGGTGGTTAATACCCATCAGGAAAGACATCACTCACAGAAGAAGCACCGGCTAACTCCGTG
CCAGCAGCCGCGGTAATACGGAGGGTGCGAGCGTTAATCGGAATTACTGGGCGTAAAGCGCGCGTAGGTGGCTTGATAAG
CCGGTTGTGAAAGCCCCGGGCTCAACCTGGGAACGGCATCCGGAAGTGTCAAGCTAGAGTGCAGGAGAGGAAGGTAGAAT
TCCCGGTGTAGCGGTGAAATGCGTAGAGATCGGGAGGAATACCAAGTGGCGAAGGCGGCCTTCTGGACTGACACTGACACT
GAGGTGCGAAAGCGTGGGTAGCAAACAGG
```

**参数选项1:** 使用参数 `-p`, 填充 `NNNNNNNN` 到标题

```
$ atlas-utils label -p 'NNNNNNNN' ./label/example/classify.txt
./label/example/zotus.fasta |head -n 5
>ZOTU_1NNNNNNNNNd:Bacteria,p:"Proteobacteria",c:Epsilonproteobacteria,o:Campylobacteriales,f:Campylobacteraceae,g:Arcobacter
TGGGGAATATTGCACAATGGACGAAAGTCTGATGCAGCAACGCCGCGTGGAGGATGACACATTTCCGGTTCGTAACCTCCT
TTTATATAGGAAGATAATGACGGTACTATATGAATAAGCACCGGCTAACTCCGTGCCAGCAGCCGCGGTAATACGGAGGG
TGCAAGCGTTACTCGGAATCACTGGGCGTAAAGAGCGTGTAGCGGGTATATAAGTCAGAAGTGAAATCCAATAGCTTAA
CTATTGAACTGCTTTTAAAAGTGTATACCTAGAATGTGGGAGAGGTAGATGGAATTTCTGGTGTAGGGGTAAAATCCGTA
GAGATCAGAAGGAATACCGATTGCGAAGGCGATCTACTGGAACATTATTGACGCTGAGACGCGAAAGCGTGGGGAGCAAA
CAGG
>ZOTU_2NNNNNNNNNd:Bacteria,p:"Proteobacteria",c:Gammaproteobacteria,o:Oceanospirillales,f:Halomonadaceae,g:Halomonas
TGGGGAATATTGGACAATGGGGGAACCTGATCCAGCCATGCCGCGTGTGTGAAGAAGGCCCTCGGGTTGTAAAGCACT
TTCAGTGAGGAAGAACGCCTGGTGGTTAATACCCATCAGGAAAGACATCACTCACAGAAGAAGCACCGGCTAACTCCGTG
CCAGCAGCCGCGTAATACGGAGGGTGCGAGCGTTAATCGGAATTACTGGGCGTAAAGCGCGCGTAGGTGGCTTGATAAG
CCGGTTGTGAAAGCCCCGGGCTCAACCTGGGAACGGCATCCGGAAGTGTCAAGCTAGAGTGCAGGAGAGGAAGGTAGAAT
TCCCGGTGTAGCGGTGAAATGCGTAGAGATCGGGAGGAATACCAGTGCGAAGGCGGCCTTCTGGACTGACACTGACACT
GAGGTGCGAAAGCGTGGGTAGCAAACAGG
>ZOTU_3NNNNNNNNNd:Bacteria,p:"Proteobacteria",c:Alphaproteobacteria,o:Rhodospirillales,f:Acetobacteraceae,g:Roseomonas
```

本文材料为 **BASE** (**Bioinformatics Applied bioinformatic Series**) 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: 2020-09-11 11:56 AM