atlas-utils之过滤分类结果中线粒体和叶绿体来源的序列：filter

一、atlas-utils filter介绍

功能描述：

`atlas-utils filter` 过滤 USEARCH SINTAX/NBC 分类结果中线粒体和叶绿体来源的序列，可以指定匹配的分类学名称。

命令行接口：

```
$ atlas-utils filter

Usage: atlas-utils filter [options] <sintax>

Options:
    -t STR filter d:Bacteria or d:Archaea, or Others Default:[NONE]
           for multi-targer: d:Archaea,p:Firmicutes
    -r     print OTU IDs after filter.
```

可选参数：

```
    -t 字符串    过滤分类学中d:Bacteria或者d:Archaea或者其他物种，默认为空；
                多个分类学目标可设置如：d:Archaea,p:Firmicutes
    -r          输出过滤后的OTU ID
```

二、使用场景实例及其用法

示例演示：

示例文件：`classify.txt`

```
$ cat classify.txt |head -n 5
```

```
1  ZOTU_19 d:Bacteria(1.0000),p:"Bacteroidetes"
   (1.0000),c:Sphingobacteriia(1.0000),o:"Sphingobacteriales"
   (1.0000),f:Chitinophagaceae(1.0000),g:Parafilimonas(0.9400),s:Parafilimonas_t
   errae(0.8836)        +
   d:Bacteria,p:"Bacteroidetes",c:Sphingobacteriia,o:"Sphingobacteriales",f:Chit
   inophagaceae,g:Parafilimonas,s:Parafilimonas_terrae
2  ZOTU_5  d:Bacteria(1.0000),p:"Bacteroidetes"(1.0000),c:"Bacteroidia"
   (0.7800),o:"Bacteroidales"
   (0.6084),f:Prolixibacteraceae(0.3650),g:Mariniphaga(0.0621),s:Mariniphaga_ana
   erophila(0.0105)    +        d:Bacteria,p:"Bacteroidetes"
3  ZOTU_9  d:Bacteria(1.0000),p:"Bacteroidetes"
   (1.0000),c:Flavobacteriia(1.0000),o:"Flavobacteriales"
   (1.0000),f:Flavobacteriaceae(1.0000),g:Flavobacterium(1.0000)        +
   d:Bacteria,p:"Bacteroidetes",c:Flavobacteriia,o:"Flavobacteriales",f:Flavobac
   teriaceae,g:Flavobacterium
4  ZOTU_12 d:Bacteria(1.0000),p:"Bacteroidetes"(1.0000),c:"Bacteroidia"
   (1.0000),o:"Bacteroidales"(1.0000),f:"Porphyromonadaceae"
   (1.0000),g:Petrimonas(1.0000),s:Petrimonas_sulfuriphila(1.0000)  +
   d:Bacteria,p:"Bacteroidetes",c:"Bacteroidia",o:"Bacteroidales",f:"Porphyromon
   adaceae",g:Petrimonas,s:Petrimonas_sulfuriphila
5  ZOTU_7  d:Bacteria(1.0000),p:"Bacteroidetes"(1.0000),c:"Bacteroidia"
   (1.0000),o:"Bacteroidales"(1.0000),f:"Porphyromonadaceae"
   (1.0000),g:Macellibacteroides(0.5700),s:Macellibacteroides_fermentans(0.3249)
       +
   d:Bacteria,p:"Bacteroidetes",c:"Bacteroidia",o:"Bacteroidales",f:"Porphyromon
   adaceae"
```

运行命令：

过滤 `USEARCH SINTAX/NBC` 分类结果中为 `d:Bacteria` 的序列。

```
1  $ atlas-utils filter -t "d:Archaea" classify.txt | head -n2
```

```
1  ZOTU_19 d:Bacteria(1.0000),p:"Bacteroidetes"
   (1.0000),c:Sphingobacteriia(1.0000),o:"Sphingobacteriales"
   (1.0000),f:Chitinophagaceae(1.0000),g:Parafilimonas(0.9400),s:Parafilimonas_t
   errae(0.8836)        +
   d:Bacteria,p:"Bacteroidetes",c:Sphingobacteriia,o:"Sphingobacteriales",f:Chit
   inophagaceae,g:Parafilimonas,s:Parafilimonas_terrae
2  ZOTU_5  d:Bacteria(1.0000),p:"Bacteroidetes"(1.0000),c:"Bacteroidia"
   (0.7800),o:"Bacteroidales"
   (0.6084),f:Prolixibacteraceae(0.3650),g:Mariniphaga(0.0621),s:Mariniphaga_ana
   erophila(0.0105)    +        d:Bacteria,p:"Bacteroidetes"
```

参数选项**1**： 设置 `-r` 参数，输出过滤后的OTU ID。

```
1  $ atlas-utils filter -t d:Archaea  -r classify.txt  2> identifiers.txt  |
   head -n 2
```

```
1  ZOTU_19 d:Bacteria(1.0000),p:"Bacteroidetes"
   (1.0000),c:Sphingobacteriia(1.0000),o:"Sphingobacteriales"
   (1.0000),f:Chitinophagaceae(1.0000),g:Parafilimonas(0.9400),s:Parafilimonas_t
   errae(0.8836)          +
   d:Bacteria,p:"Bacteroidetes",c:Sphingobacteriia,o:"Sphingobacteriales",f:Chit
   inophagaceae,g:Parafilimonas,s:Parafilimonas_terrae
2  ZOTU_5  d:Bacteria(1.0000),p:"Bacteroidetes"(1.0000),c:"Bacteroidia"
   (0.7800),o:"Bacteroidales"
   (0.6084),f:Prolixibacteraceae(0.3650),g:Mariniphaga(0.0621),s:Mariniphaga_ana
   erophila(0.0105)    +        d:Bacteria,p:"Bacteroidetes"
3  [biostack@biostack example]$ atlas-utils filter -t d:Archaea  -r classify.txt
   2> identifiers.txt  | head -n 2
4  ZOTU_19 d:Bacteria(1.0000),p:"Bacteroidetes"
   (1.0000),c:Sphingobacteriia(1.0000),o:"Sphingobacteriales"
   (1.0000),f:Chitinophagaceae(1.0000),g:Parafilimonas(0.9400),s:Parafilimonas_t
   errae(0.8836)          +
   d:Bacteria,p:"Bacteroidetes",c:Sphingobacteriia,o:"Sphingobacteriales",f:Chit
   inophagaceae,g:Parafilimonas,s:Parafilimonas_terrae
5  ZOTU_5  d:Bacteria(1.0000),p:"Bacteroidetes"(1.0000),c:"Bacteroidia"
   (0.7800),o:"Bacteroidales"
   (0.6084),f:Prolixibacteraceae(0.3650),g:Mariniphaga(0.0621),s:Mariniphaga_ana
   erophila(0.0105)    +        d:Bacteria,p:"Bacteroidetes"
```

本文材料为 **BASE** (**B**iostack **A**pplied bioinformatic **SE**ies) 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。