

# tsv-utils之合并数值表: distribution

## 一、tsv-utils distribution介绍

功能描述:

`tsv-utils distribution` 根据**bin**文件以及多样本元素（elements）的分类以及数值映射表合并新的数值表，新的数值表关联bin文件的主键和elements表中的特征。。

命令行接口:

```
1 $ tsv-utils distribution
2
3 Usage: tsv-utils distribution <bins> <abundance>
```

## 二、使用场景实例及其用法

示例演示

示例文件:

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

```
1 #catalog      members
2 K14048  TRINITY_g266049_i1_1,TRINITY_g453672_i2_3,TRINITY_g474650_i1_4
3 K00360  TRINITY_g317023_i1_2,TRINITY_g83207_i1_1
4 K02305  TRINITY_g209940_i1_1,TRINITY_g212274_i1_1,TRINITY_g282567_i1_1
5 K19823  TRINITY_g120231_i1_1,TRINITY_g170185_i1_1,TRINITY_g8936_i1_1
6 K01672  TRINITY_g220886_i1_1
```

注意事项: 该输出只是演示.

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

```
1 TRINITY_g741_i1_1      Chordata      0.000000
2 TRINITY_g1252_i1_1     Chordata      0.000000
3 TRINITY_g2524_i1_1     Streptophyta  0.000000
4 TRINITY_g3978_i1_1     Chordata      0.000000
5 TRINITY_g4010_i1_1     Chordata      11.574194
6 TRINITY_g5274_i1_1     Chordata      0.000000
```

注意事项: abundance文件(比如phylum.txt) 必须包含三列, 第一列主键为 `bins` 文件的元素, 第二列为需要汇总的元素分类, 第三列为数值.

运行命令:

```
1 $ tsv-utils distribution ko.txt phylum.txt |cut -f1,2,4,5| head -n 6
```

1	#labels	Arthropoda	Cyanobacteria	Chordata
2	k00260	10.023299	0.000000	2.624316
3	k04561	0.000000	0.000000	0.000000
4	k00262	0.000000	0.000000	8.431389
5	k00362	0.000000	0.000000	0.000000
6	k00264	0.000000	0.000000	4.767519

**注意事项：** 计算结果为不同属对特征的贡献度。

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: 8/30/2020 7:03:58 PM