

atlas-utils: 微生物组数据操作通用程序

一、程序介绍

1 | `atlas-utils` 为一组微生物组数据分析的工具包，主要使用`klib`进行开发，包括`khash`,
| kvect, kstring`等。

二、主程序接口

当前释放版本: `version: 0.0.1-r2`

```
1 | $ atlas-utils
2 |
3 | Usage: atlas-utils <command> <arguments>
4 | Version: 0.0.1 r2
5 |
6 | Command:
7 |     -- Fastq manipulation.
8 |         demultiplex    FASTAQ demultiplex with barcode sequence.
9 |         split          FASTAQ demultiplex with barcode without
10 |                        trimming barcode sequence.
11 |         uniques        Find unique sequences and abundances.
12 |         linkpairs      Link paired-reads with padding 'N'.
13 |         fqchk          Fastq QC (base/quality summary).
14 |         label          Add label to fasta headline.
15 |         barcode        Get barcode seq from paired-reads.
16 |         orient         Orient seq use strand +/- information.
17 |         primer_strip   Retrive sequence region matches to primer pair.
18 |                        support single end match.
19 |
20 |     -- (Z)OTU manipulation.
21 |         unique_table   Construct uniques table using exact match.
22 |         annotation     Annotate OTU table using OTU assignment from
23 |                        Sintax/RDP.
24 |         filter         1. Filter sequence from Mitochondria/Chloroplast,
25 |                        2. optional filter
26 |                        specified taxon level .
27 |         level          Abundance table for specify taxonomy level. ie.
28 |                        genus.
29 |         abundance      Abundance table for specify taxonomy level and
30 |                        sample. ie. genus.
31 |         summary        Calcuate OTU number/Tag number per sample.
32 |         rowsum          Calcuate row sum.
33 |         rank           Rank/merge for numeric table.
34 |         krona           Convert annotated OTU table to krona text format.
35 |         quantile        Calcuate nth quantile sample size for otu_table
36 |                        normalization.
37 |         mean_size      Calcuate 'Mean' sample size for otu_table
38 |                        normalization.
39 |         min_size       Calcuate 'Min' sample size for otu_table
40 |                        normalization.
```

34	max_size	Calculate 'Max' sample size for otu_table
	normalization.	
35	counts2freqs	Convert counts table from counts to frequencies.
36	trim	Delete low-abundance (Z)OTUs.
37	binary	Convert (Z)OTUS table to presence/absence values.
38	pairwise	Convert (Z)OTUS table to pairwise format for
	phylocom.	
39	rarity	rarity summary.
40	rare	Rarefy OTU table to specified number of reads.
41	group_by	group_by operation for OTU table, support average,
	sum.	
42		
43	-- PICRUST.	
44	normalization	Normalize OTU table with 16S copy Number, et..
45	melt	Calculate feature abundance. ie. KEGG, COG, with
	Greengene 13.5.	
46	kann	Calculate feature abundance using feature maps, KO
	-> Module.	
47		
48	-- auxiliary utils.	
49	view	view text file, ignor comments and blank lines.
50	getline	Get target line with headline.
51	subsamples	Get target columns with headline match.
52	partition	split OTU table into specify number file.

Licenced:

(c) 2019-2020 - LEI ZHANG

Logic Informatics Co.,Ltd.

zhanglei@logicinformatics.com

三、主要子命令功能介绍

主要子命令功能介绍:

1. [demultiplex](#) 使用双端 **barcode** 拆分数据, 支持精确匹配模式, 去除 **barcode** 序列;
2. [split](#) 使用双端 **barcode** 拆分数据, 支持精确匹配模式, 不删除 **barcode** 序列;
3. [uniques](#) 计算非冗余序列, 兼容USEARCH的注释文件;
4. [linkpairs](#) 链接PE双端序列, 中间填充指定数量'N';
5. [fqchk](#) 统计序列质量;
6. [label](#) 标注序列文件的注释信息;
7. [barcode](#) 获取PE序列的 5'->3' 前指定数量碱基作为barcode, 适用于检测序列未拆分信息;
8. [orient](#) 调整序列的方向, 根据数据库的 +/- 链特异性;
9. [primer_strip](#) 切除引物序列;
10. [unique_table](#) 使用USEARCH unique子命令构建Unique序列丰度表;
11. [annotation](#) 使用SINTAX/RDP结果注释(Z)OTU表;
12. [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;
13. [level](#) 计算给定(Z)OTU表统计指定分类学等级的丰度;
14. [abundance](#) 给定(Z)OTU表以及样本标识符, 计算对应样本的丰度;
15. [summary](#) 计算给定(Z)OTU表的(Z)OTU数目以及Tags数目;
16. [rowsum](#) 计算ZOTU表中的每个(Z)OTU 的Tags数;
17. [rank](#) 对给定(Z)OTU表进行排序, 并输出指定数据的 (Z)OTU 或者物种分类, 其余可归并到Others分类;
18. [krona](#) 给定(Z)OTU 表以及样本标识符, 生成Krona兼容的文件格式;
19. [quantile](#) 计算样本Tags数指定百分比大小数值;

20. [mean_size](#) 给定(Z)OTU表中所有样本的Tags总数的mean值;
21. [min_size](#) 给定(Z)OTU表中所有样本的Tags总数的min值;
22. [max_size](#) 给定(Z)OTU表中所有样本的Tags总数的max值;
23. [counts2freqs](#) 将(Z)OTU表的counts 转换成相对丰度表;
24. [trim](#) 使用 (Z)OTU 相对丰度的和对(Z)OTU表进行过滤。
25. [binary](#) 转换 (Z)OTU 表为presence/absence (0/1) 模式;
26. [pairwise](#) 转换(Z)OTU 表为三列 phylocom 格式;
27. [rarity](#) 稀有度分析;
28. [rare](#) (Z)OTU表抽平;
29. [group_by](#) 根据分组信息汇总(Z)OTU表
30. [melt](#) 根据Closed_OTU表预测功能多样性
31. [normalization](#) 根据标记基因拷贝数对扩增子数据进行标准化;
32. [kann](#) 用丰度表以及特征映射表, 将数值表转换成更高级分类的数值表; 比如KO表转换成Pathway表;
33. [view](#) 简单处理表格行操作, 比如添加注释符号(#), 忽略注释符号以及空行;
34. [getline](#) 根据指定的行标识符(与第一列的名称匹配), 获取对应的行;
35. [subsamples](#) 根据指定的列标识符(第一行中指定的名称), 获取对应的列或者多列信息, 强制包含第一列信息, 比如从 (Z)OTU 表中抽取一部分样本的数值, 形成新的 (Z)OTU 表;
36. [partition](#) 根据指定样本数拆分(Z)OTU表;

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: 9/2/2020 12:19:31 AM