

16S rDNA Amplicon 测序 交付文档

上海锐翌基因科技有限公司

目录

1. 分析结果文件列表.....	3
2. FASTA 格式.....	11
3. FASTQ 格式.....	11
4. 文件打开或浏览方法.....	12

1. 分析结果文件列表

RESULTS

└─04_data_statistics	[Reads]
length_distrubution.pdf	[reads 长度]
length_distrubution.png	[reads 长度]
length_distrubution.tsv	[reads 长度统计数据]
pick_otu_summary.tsv	[reads 长度]
└─alpha	
chao1.pdf	[Alpha 多样性 chao1 指数]
chao1.png	[chao1 指数]
chao1.txt	[chao1 指数的数据]
observed_species.pdf	[observed_species 指数]
observed_species.png	[observed_species 指数]
observed_species.txt	[observed_species 指数的数据]
└─05_OTU_analysis	[OTU 分析]
└─all	[所有]
otu_downsize_stat.tsv	[OTU 抽平结果]
otu_statistic.tsv	[OTU 统计]
otu_table.tsv	[OTU 丰度表]
profile_tree.tsv	[层次分类丰度表]
rep_set.fna	[OTU 代表序列]
tax_assignment.tsv	[OTU 注释文件]
└─specaccum	[物种累积曲线]
specaccum.pdf	
specaccum.png	
└─wf_taxa_summary	[OTU 注释到各个水平]
otu_table_L2.txt	[门水平]
otu_table_L3.txt	[纲水平]
otu_table_L4.txt	[目水平]
otu_table_L5.txt	[科水平]
otu_table_L6.txt	[属水平]

		└─group [group 分组一]
		otu_downsize_stat.tsv [OTU 抽平结果]
		otu_statistic.tsv [OTU 数量统计]
		otu_table.tsv [OTU 丰度表]
		profile_tree.tsv [层次分类丰度表]
		tax_assignment.tsv [OTU 注释文件]
		└─core_otu [共有 OTU]
		core_otu.pdf [共有 OTU]
		core_otu.png [共有 OTU]
		core_otu.txt [共有 OTU 数据]
		for_plot.txt [画图数据]
		└─flower [花瓣图]
		for_plot.txt [画图数据]
		flower.png
		└─otu_pca [PCA 图]
		otu_pca.pdf
		otu_pca.png
		└─venn [维恩图]
		for_plot.txt
		venn.png
		└─wf_taxa_summary [OTU 注释到各个水平]
		otu_table_all.txt [所有水平]
		otu_table_L1.txt [界水平]
		otu_table_L2.txt [门水平]
		otu_table_L3.txt [纲水平]
		otu_table_L4.txt [目水平]
		otu_table_L5.txt [科水平]
		otu_table_L6.txt [属水平]
		└─06_classification_abundance_analysis
		└─all [所有]
		krona.html [注释结果可视化]

			└─bar_plot	[物种注释柱状图]
			└─class	[纲水平]
			bar_plot.pdf	[柱状图]
			bar_plot.png	
			for_plot.txt	[画图数据]
			└─family	[科水平]
			bar_plot.pdf	
			bar_plot.png	
			for_plot.txt	
			└─genus	[属水平]
			bar_plot.pdf	
			bar_plot.png	
			for_plot.txt	
			└─order	[目水平]
			bar_plot.pdf	
			bar_plot.png	
			for_plot.txt	
			└─phylum	[门水平]
			bar_plot.pdf	
			bar_plot.png	
			for_plot.txt	
			└─rank_abundance	[rank_abundance 曲线图]
			rank_abundance.pdf	
			rank_abundance.png	
			└─tax_star	[星形图]
			for_star_plot.txt	
			tax_star.pdf	
			tax_star.png	
			└─group	[分组 1]
			krona.html	[注释结果可视化]

		└─bar_plot	[物种注释柱状图]
		└─class	[纲水平]
		bar_plot.pdf	[柱状图]
		bar_plot.png	[柱状图]
		for_plot.txt	[画图数据]
		└─family	[科水平]
		bar_plot.pdf	[柱状图]
		bar_plot.png	
		for_plot.txt	
		└─genus	[属水平]
		bar_plot.pdf	[柱状图]
		bar_plot.png	
		for_plot.txt	
		└─order	[目水平]
		bar_plot.pdf	[柱状图]
		bar_plot.png	
		for_plot.txt	
		└─phylum	[门水平]
		bar_plot.pdf	[柱状图]
		bar_plot.png	
		for_plot.txt	
		└─heatmap	[热图]
		for_plot.txt	
		heatmap.pdf	
		heatmap.png	
		└─phylo_tree	[进化树]
		tax_phylo.nwk	
		└─tax_bar_tree	[物种聚类]
		bar_tree.pdf	
		bar_tree.png	

		for_plot.txt	
		└─tax_tree	[物种分类树]
		tax_ass_modified.txt	
		tax_tree.nwk	
		└─07_Alpha_diversity	[alpha 多样性]
		└─group	[分组一]
		alpha_marker.tsv	[alpha 显著差异 marker]
		alpha_statistic.tsv	[alpha 多样性统计]
		chao1.pdf	[chao1 指数]
		chao1.png	
		chao1.txt	
		goods_coverage.pdf	[goods_coverage 指数]
		goods_coverage.png	
		goods_coverage.txt	
		observed_species.pdf	[observed_species 指数]
		observed_species.png	
		observed_species.txt	
		PD_whole_tree.pdf	[PD_whole_tree 指数]
		PD_whole_tree.png	
		PD_whole_tree.txt	
		shannon.pdf	[shannon 指数]
		shannon.png	
		shannon.txt	
		simpson.pdf	[simpson 指数]
		simpson.png	
		simpson.txt	
		└─box_plot	[盒型图]
		chao1.boxplot.pdf	[chao1 指数]
		chao1.boxplot.png	
		goods_coverage.boxplot.pdf	[goods_coverage 指数]
		goods_coverage.boxplot.png	
		observed_species.boxplot.pdf	[observed_species 指数]
		observed_species.boxplot.png	
		PD_whole_tree.boxplot.pdf	[PD_whole_tree 指数]
		PD_whole_tree.boxplot.png	

		shannon.boxplot.pdf	[shannon 指数]
		shannon.boxplot.png	
		simpson.boxplot.pdf	[simpson 指数]
		simpson.boxplot.png	
		total_alpha_rare	[总的 alpha 多样性]
		alpha_statistic.tsv	[alpha 多样性统计]
		chao1.pdf	[chao1 指数]
		chao1.png	
		chao1.txt	
		goods_coverage.pdf	[goods_coverage 指数]
		goods_coverage.png	
		goods_coverage.txt	
		observed_species.pdf	
		observed_species.png	[observed_species 指数]
		observed_species.txt	
		PD_whole_tree.pdf	
		PD_whole_tree.png	[PD_whole_tree 指数]
		PD_whole_tree.txt	
		shannon.pdf	
		shannon.png	[shannon 指数]
		shannon.txt	
		simpson.pdf	
		simpson.png	[simpson 指数]
		simpson.txt	
		08_Beta_diversity	[beta 多样性]
		group	[分组一]
		unweighted_unifrac_otu_table.tsv	[未加权 unifrac 距离表]
		weighted_unifrac_otu_table.tsv	[加权 unifrac 距离表]
		anosim	[anosim 结果]
		unweighted_unifrac.anosim.pdf	
		unweighted_unifrac.anosim.png	
		weighted_unifrac.anosim.pdf	
		weighted_unifrac.anosim.png	
		cluster	[聚类结果]


```

|      |      mapfile.txt
|      |      unweighted_unifrac_cluster.pdf
|      |      unweighted_unifrac_cluster.png
|      |      unweighted_unifrac_otu_table.txt
|      |      weighted_unifrac_cluster.pdf
|      |      weighted_unifrac_cluster.png
|      |      weighted_unifrac_otu_table.txt
|      |
|      |      └─heatmap                                [热图]
|      |      |      unweighted_unifrac.heatmap.pdf
|      |      |      unweighted_unifrac.heatmap.png
|      |      |      weighted_unifrac.heatmap.pdf
|      |      |      weighted_unifrac.heatmap.png
|      |
|      |      └─nmds                                  [nmds 结果]
|      |      |      unweighted_unifrac.nmds.pdf
|      |      |      unweighted_unifrac.nmds.png
|      |      |      weighted_unifrac.nmds.pdf
|      |      |      weighted_unifrac.nmds.png
|      |
|      |      └─pcoa                                  [pcoa 结果]
|      |      |      unweighted_unifrac.pcoa.pdf
|      |      |      unweighted_unifrac.pcoa.png
|      |      |      weighted_unifrac.pcoa.pdf
|      |      |      weighted_unifrac.pcoa.png
|
|      └─09_diff_analysis                             [差异分析]
|          └─group                                    [分组一]
|              └─genus_diff                           [属水平差异]
|                  |      genus_diff.marker.filt.tsv    [过滤后的差异 maker]
|                  |      genus_diff.marker.tsv         [差异 maker]
|                  |
|                  |      └─boxplot                     [盒型图]
|                  |      |      diff_boxplot.for_plot_top_20.txt
|                  |      |      diff_boxplot.pdf
|                  |      |      diff_boxplot.png
|                  |      |
|                  |      └─heatmap                     [热图]

```

```

| | heatmap.pdf
| | heatmap.png
| |
| └─pca [pca 图]
|     diff_pca.pdf
|     diff_pca.png
|
| └─LEfSe [LEfSe 结果]
|     LDA.cladogram.pdf
|     LDA.cladogram.png
|     LDA.pdf
|     LDA.png
|     otu_table_for_lefse.txt
|
| └─otu_diff [OTU 水平差异]
|     | otu_diff.marker.filt.tsv
|     | otu_diff.marker.tsv
|     |
|     └─boxplot [盒型图]
|         | diff_boxplot.for_plot_top_20.txt
|         | diff_boxplot.pdf
|         | diff_boxplot.png
|         |
|         └─heatmap [热图]
|             | heatmap.pdf
|             | heatmap.png
|             |
|             └─pca [pca 图]
|                 | diff_pca.pdf
|                 | diff_pca.png
|                 |
|                 └─taxall_diff [所有水平的差异]
|                     | taxall_diff.marker.filt.tsv
|                     | taxall_diff.marker.tsv
|                     |
|                     └─boxplot [盒型图]
|                         | diff_boxplot.for_plot_top_20.txt
|                         | diff_boxplot.pdf

```

```

|      diff_boxplot.png
|
|—heatmap      [热图]
|      heatmap.pdf
|      heatmap.png
|—pca          [pca 图]
      diff_pca.pdf
      diff_pca.png
  
```

2. FASTA 格式

FASTA 格式（又称为 Pearson 格式），是一种基于文本用于表示核苷酸序列或氨基酸序列的格式。在这种格式中碱基对或氨基酸用单

个字母来编码，且允许在序列前添加序列名及注释。序列文件的第一行是由大于号">"或分号";"打头的任意文字说明（习惯常用">"作为起

始），用于序列标记。从第二行开始为序列本身，只允许使用既定的核苷酸或氨基酸编码符号。通常核苷酸符号大小写均可，而氨基酸常

用大写字母。如：

```

>scaffold1 35.9

AACTCCAAATGTTTTACATCCTTTTTTTATCCATAATATATAATCAACTGATATACAAAATGAAA
AAATACTACCTACATTTTTTATTAGGCTTATTTTATTAAAATAAGGTTGGT

GTTGTGGAAATAGCCATTCT.....
  
```

3. FASTQ 格式

Single-end 测序，每个样品只有 1 个 FASTQ 文件；Paired-end 测序，*_1.fq 和*_2.fq 分别对应一个 FASTQ 文件。

FASTQ 文件是用户得到的最原始文件，其中每 4 行表示一条 read，即：

```

@FC4290FAAXX:4:1:3:84#CAGATC/1

CCAACTATGATAGCCAANAAGGGAAAGCCATAGAG...

+

abb_aab_aa`a^aba^D[`a_`aaaa`_a_`__a...
  
```

每个序列共有 4 行，第 1 行是@序列 ID，包括 index 序列及 read1 或 read2

标志，由测序仪产生；第 2 行是碱基序列，大写“ACTGN”；

第三行是“+”，省略了序列 ID；第 4 行是序列对应的测序质量值序列，每个字母对应第 2 行每个碱基，第四行每个字母对应的 ASCII 值

减去 33，即为该碱基的测序质量值，比如 I 对应的 ASCII 十进制值为 73，那么其对应的碱基质量值是 40。Solexa 碱基质量值范围为 0 到

41。下表为 Solexa 测序错误率与测序质量值简明对应关系，Solexa 碱基质量值用 sQ 表示，则有以下表关系：

表 2-1 Solexa 测序错误率与测序质量值简明对应关系

测序错误率测序质量值对应字符

5.00%	20	5
0.10%	30	?
0.01%	40	I

4. 文件打开或浏览方法

所有提供的文件均为 Linux 系统下的文件，压缩包使用“tar -zcvf ”命令压缩，以下为不同系统用户解压缩的方法：

Unix/Linux/Mac 用户：使用 tar -zcvf *.tar.gz 命令

Windows 用户：使用 WinRAR 软件解压缩

如在本附录中无特殊说明，所有提供的文件均为 Linux 系统下文本文件。

Unix/Linux 用户可以使用 more 或 less 命令查看该文本文件内容；Windows 用户，一般文本文件可以使用写字板打开。推荐使用开源文本编辑器 gedit for win32 版本(<http://projects.gnome.org/gedit/>)。当文件比较大时，打开文件可能导致 Windows 系统死机，建议使用性能较好的计算机或者使用更适合处理大量数据的 Unix/Linux 系统打开。

数据中可能包含部分图像文件，一般图像文件后缀名为.png、.pdf、tiff 等，对于图像文件，Windows 用户可以使用图片浏览器打开，Linux/Unix 用户使用 display 命令打开。

Linux 下的表格均为制表符 (Tab) 分割的文本, 为了便于阅读, 建议使用 excel 或 openoffice 等办公软件用表格形式打开, 打开时请用 “Tab 分割” 方式。