

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.....	4
3. FASTQ 格式.....	4

1. 分析结果文件列表

--Large project number_subproject code/	
--01_Reads/	[Reads]
--chao1.pdf/png	[chao1 指数]
--observed_species.pdf/png	[observed_species 指数]
--length_distribution.pdf/png	[reads 长度]
--pick_otu_summary.tsv	[质控统计文件]
--02_OTU	
--all/	
--bar_plot/	[各个分类等级的柱状图]
--otu_statistic.tsv	[注释结果初步统计]
--profile_tree.tsv	[注释结果树形统计]
--specaccum.pdf/png	[specaccum 图]
--tax_assignment.tsv	[OTU 注释结果]
--otu_downsize_stat.tsv	[抽平前后的 OTU 数量]
--otu_table.tsv	[OTU 丰度表]
--rank_abundance.pdf/png	[rank_abundance 图]
--rep_set.fna	[OTU 代表序列]
--tax_star.pdf/png	[饼图]
--wf_taxa_summary/	[各个水平的统计结果]
--group/ (其他文件夹)	[分组方案的结果]
--bar_plot/	[各个分类等级的柱状图]
--heatmap.pdf/png	[OTU 热图]
--otu_downsize_stat.tsv	[抽平前后的 OTU 数量]
--otu_table.tsv	[OTU 丰度表]
--tax_assignment.tsv	[OTU 注释结果]
--wf_taxa_summary/	[各个水平的统计结果]
--core_otu.pdf/png	[共有 OTU 柱状图]
--core_otu.tsv	[共有 OTU 统计]
--otu_pca.pdf/png	[OTU PCA 图]
--otu_statistic.tsv	[注释结果初步统计]
--profile_tree.tsv	[注释结果树形统计]
--venn.png	[维恩图]
--03_Alpha_diversity	[Alpha 多样性]
--total_alpha_rare/	
--alpha_statistic.tsv	[样品 Alpha 多样性统计表]
--chao1.pdf/png	[chao1 指数图]
--goods_coverage.pdf/png	[goods_coverage 图]

4

```
| | | | --taxall_diff.marker.filt.tsv [所有水平的显著差异]
| | --16S 交付文档.pdf
| | --16S 分析方法.pdf
```

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
AAATACTACCTACATTTTTTATTAGGCTTATTTTATTAATAAAGGTTGGT

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