

ITS rDNA Amplicon 测序 交付文档

上海锐翌基因科技有限公司



ITS rDNA Amplicon 测序交付文档

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/				
	0	1_Reads/	[Reads]	
		chao1.pdf	[chao1 指数]	
		observed_species.pdf	[observed_species 指数]	
		length_distrubution.pdf	[reads 长度]	
		pick_otu_summary.tsv	[质控统计文件]	
	0	2_OTU		
		all/		
		otu_downsize_stat.tsv	[抽平前后的 OTU 数量]	
		otu_table.txt	[OTU 丰度表]	
		bar_plot/	[各个分类等级的柱状图]	
		rank_abundance.pdf	[rank_abundance 图]	
		specaccum.pdf	[specaccum 图]	
		tax_star.pdf	[饼图]	
		rep_set.fna	[OTU 代表序列]	
		tax_assignment.txt	[OTU 注释结果]	
		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]	
		core_otu.txt	[共有 OTU 统计]	
		core_otu.pdf	[共有 OTU 柱状图]	
		heatmap.pdf	[OTU 热图]	
		otu_pca.pdf	[OTU PCA 图]	
		venn.png	[维恩图]	
		tax_assignment.txt	[OTU 注释结果]	
		wf_taxa_summary/	[各个水平的统计结果]	
	0	3_Alpha_diversity	[Alpha 多样性]	
		total_alpha_rare/		
		alpha_statistic.tsv	[样品 Alpha 多样性统计表]	
		chao1.pdf	[chao1 指数图]	
		goods_coverage.pdf	[goods_coverage 图]	
		observed_species.pdf	[observed_species 图]	
		PD_whole_tree.pdf	[PD_whole_tree 图]	
		shannon.pdf	[shannon 图]	
		simpson.pdf	[simpson 图]	
		group(其他文件夹)	[分组方案的结果]	
		alpha_marker.tsv	[Alpha diversity 指数差异]	
		alpha_statistic.tsv	[样品 Alpha 多样性统计表]	

锐翌基因 REALGENE

ITS rDNA Amplicon 测序交付文档

	box_plot/	[Alpha 多样性的盒形图]	
04_B	Beta_diversity	[Beta 多样性]	
group			
	unweighted_unifrac_otu_table.txt	[基于 unweighted_unifrac 统计]	
	weighted_unifrac_otu_table.txt	[基于 weighted_unifrac 统计]	
	unweighted_unifrac.heatmap.pdf	[热图]	
	unweighted_unifrac.pcoa.pdf	[pcoa 图]	
	weighted_unifrac.heatmap.pdf	[热图]	
	weighted_unifrac.pcoa.pdf	[pcoa 图]	
05_diff_analysis			
group			
	LEfSe/	[LEfSe 分析结果]	
	genus_diff.marker.txt	[属水平的差异结果]	
	genus_diff/	[属水平的差异图]	
	otu_diff.marker.txt	[OTU 水平的差异结果]	
	otu_diff/	[OTU 水平的差异图]	
	taxall_diff.marker.txt	[所有水平的差异结果]	
	taxall_diff/	[所有水平的差异图]	
	phylum_diff.marker.txt	[门水平的的差异结果]	
	phylum_diff/	[门水平的差异图]	

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

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。

FASTO 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