
诺禾致源
元基因组交付目录说明手册
(V5.0)



2018 年 2 月 25 日

目录

(注：单击即可跳转至相应文档的详细说明)

-- 02.Assembly ——	【元基因组组装结果】	3
-- total.scafSeq.stat.info.xls ——	【所有样品 Scaffold 信息表】	4
-- total.scaftigs.stat.info.xls ——	【所有样品 Scatigs 信息表】	4
-- ReadsMapping ——	【将各样品 Clean Data mapping 至组装 Scaftigs 上的结果】	5
-- Sample ——	【各样品 ReadsMapping 结果，文件夹以样品名称来命名】	5
-- Sample/NOVO_MIX ——	【各样品对应的组装结果，文件夹以样品名称来命名;NOVO_MIX 为 unmapped reads 混合组装的结果】	6
-- *.scafSeq.fa ——	【单样品 scaffold 序列，FASTA 格式】	6
-- *.scafSeq.500.ss.txt ——	【按照长度 500 进行过滤后，单样品 scaffold 序列信息统计表】	6
-- *.scaftigs.500.ss.txt ——	【单样品 Scaftigs 序列信息统计表】	7
-- *.len.{png svg} ——	【Scaftigs 长度分布图，png 或 svg 格式】	8

|-- 02.Assembly —— 【元基因组组装结果】

| |-- total.scaffigs.stat.info.xls —— 【所有样品 Scaffigs 信息表】

该文件即对应的是结题报告中的组装结果 Scaffigs 的统计表，可以用 excel 打开该文件，各列所代表的含义如下：

列数	列标题	说明
1	SampleID	样品名称
2	Total len.(bp)	组装得 到的 Scaffigs 的总长， 单位为 bp
3	Num.	组装得到的 Scaffigs 总条数
4	Average len.(bp)	Scaffigs 的平均长度
5	N50 Len.(bp)	Scaffigs 的 N50
6	N90 Len.(bp)	Scaffigs 的 N90
7	Max len.(bp)	组装得到的最长 Scaffigs 的长度值

| |-- total.scafSeq.stat.info.xls —— 【所有样品 Scaffold 信息表】

该文件为组装结果 Scaffold 的统计表，可以用 excel 打开该文件，各列所代表的含义如下：

列数	列标题	说明
1	SampleID	样品名称
2	Total len.(bp)	组装得 到的 Scaffold 的总长，单位为 bp
3	Num.	组装得到的 Scaffold 总条数
4	Average len.(bp)	Scaffold 的平均长度
5	N50 Len.(bp)	Scaffold 的 N50
6	N90 Len.(bp)	Scaffold 的 N90
7	Max len.(bp)	组装得到的最长 Scaffold 的长度值

| |-- total.scaftigs.stat.info.xls —— 【所有样品 Scaftigs 信息表】

将 scaffold 从 N 连接处打断得到 scaftigs，该文件为组装结果 Scaftigs 的统计表，可以用 excel 打开该文件，各列所代表含义同 total.scafSeq.stat.info.xls，序列信息将 scaffold 更改为 scaftigs，：

| |-- ReadsMapping —— 【将各样品 Clean Data mapping 至组装 Scaffigs 上的结果】

| | `-- Sample —— 【各样品 ReadsMapping 结果，文件夹以样品名称来命名】

| | |-- coverage_depth.{png|svg} —— 【覆盖深度分布图，png 和 svg 格式】

这两个文件为对应的样品的覆盖深度分布图，其横轴代表的是测序深度，纵轴代表的是属于该测序深度的序列数目。

| | |-- coverage.depth.table.xls —— 【各 Scaffigs 覆盖度总体情况统计,包含覆盖度，覆盖长度等信息】

该文件是对 reads mapping 后的结果进行的统计，用 excel 打开该文件后，各列所代表的含义如下：

列数	列标题	说明
1	Reference_ID	Scaffigs 的编号
2	Reference_size(bp)	Scaffigs 长度
3	Covered_length(bp)	覆盖长度
4	Coverage(%)	覆盖度
5	Depth	深度
6	Depth_single	单碱基位点深度之和

| | |-- *.unmapping.{fq1|fg2}.gz ——【各样品没有 map 上 Scaffigs 的 read1 和 read2 的 FASTQ 文件】

关于 FASTQ 文件格式介绍，请参考结题报告中的常见数据格式说明文档。

| |-- Sample/NOVO_MIX ——【各样品对应的组装结果，文件夹以样品名称来命名;NOVO_MIX 为 unmapped reads 混合组装的结果】

| | |-- *.scafSeq.fa ——【单样品 scaffold 序列，FASTA 格式】

关于 FASTA 文件格式介绍，请参考结题报告中的常见数据格式说明文档。

| | |-- *.scafSeq.500.ss.txt ——【按照长度 500 进行过滤后，单样品 scaffold 序列信息统计表】

在该文件中，储存的是相应样品组装所得到的 scaffold 的平均长度，N50，N90 等基本指标，可以用写字板或记事本打开该文件。
该文件中，各列所代表的含义如下：

行数	行标题	说明
1	Statistical level	统计下方指标时的过滤阈值，例如括号中标明了 500 的即是过滤掉 500bp 以下的序列进行的统计

2	Total number	序列数目
3	Total length of (bp)	序列总长度
4	Gap number (bp)	Gap 的碱基长度
5	Average length (bp)	平均长度
6	N50 Length (bp)	序列 N50
7	N90 Length (bp)	序列 N90
8	Maximum length (bp)	最长序列长度
9	Minimum length (bp)	最短序列长度
10	GC content is (%)	序列 GC 含量

| | |-- *.scaffigs.fa ——【单样品 Scaffigs 序列，FASTA 格式】

关于 FASTA 文件格式介绍，请参考结题报告中的常见数据格式说明文档。

| | |-- *.scaffigs.500.ss.txt ——【单样品 Scaffigs 序列信息统计表】

在该文件中，储存的是相应样品组装所得到的 Scaffigs 的平均长度，N50，N90 等基本指标，可以用写字板或记事本打开该文件。
在该文件中，各列所代表的含义如下：

行数	行标题	说明
1	Statistical level	统计下方指标时的过滤阈值，例如括号中标明了 500 的即是过滤掉 500bp 以下的序列进行的统计
2	Total number	序列数目
3	Total length of (bp)	序列总长度
4	Gap number (bp)	Gap 的碱基长度
5	Average length (bp)	平均长度
6	N50 Length (bp)	序列 N50
7	N90 Length (bp)	序列 N90
8	Maximum length (bp)	最长序列长度
9	Minimum length (bp)	最短序列长度
10	GC content is (%)	序列 GC 含量

| | `-- *.len.{png|svg} —— 【Scaftigs 长度分布图，png 或 svg 格式】

这个图片展示的是某个样品中 Scaftigs 的长度分布，横轴表示 Scaftigs 的长度，第一纵轴（Frequency(#)) 表示 Scaftigs 数目；第二纵轴（Percentage (%)）表示 Scaftigs 数目的百分比，从这个图上我们可以看出，组装后得到的 Scaftigs 的长度分布情况。

坐标轴	标题	说明
横轴	Scaftig Length(bp)	Scaftigs 的长度
第一纵轴	Frequence	Scaftigs 数目
第二纵轴	Percentage(%)	Scaftigs 数目的百分比

`-- 02.Assembly--ReadMe.pdf ——【02.Assembly 交付结果目录说明】

Novogene
诺禾致源