

---

诺禾致源  
宏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)

-- 05.FunctionAnnotation——【常用功能数据库注释及丰度分析结果】 .....	6
-- CAZy——【CAZy 数据库分析结果】 .....	6
-- CAZy_Anno——【CAZy 注释结果统计】 .....	6
-- CAZy_MAT——【CAZy 相对丰度和绝对丰度分析结果：ec 为酶，level1 为六大功能类，level2 为子功能】 .....	9
-- GeneNums——【注释到的基因数目、基因 id 统计，ec 为酶，level1 为六大功能类，level2 为子功能】 .....	10
-- GeneNums.BetweenSamples——【各样品中基因数目统计】 .....	10
-- top/top_group——【各样品（组）注释到的基因数目的柱状图】 .....	11
-- GeneNums.BetweenSamples.heatmap——【各样品注释到的基因数目的聚类图】 .....	11
-- heatmap/heatmap_group——【各样品中的酶的相对丰度聚类图以及作图数据】 .....	12
-- MetaStats——【各样品的 MetaStats 统计结果，EC 为酶，level2 为子功能】 .....	12
`-- PCA ——【各样品的 PCA 分析结果，EC 为酶，level1 为六大功能类，level2 为子功能】 .....	12
`-- PCoA ——【各样品的 PCoA 分析结果，EC 为酶，level1 为六大功能类，level2 为子功能】 .....	12
`-- NMDS——【各样品的 NMDS 分析结果，EC 为酶，level1 为六大功能类，level2 为子功能】 .....	13
`-- Anosim——【各组 Anosim 分析结果】 .....	13
-- LDA——【基于 CAZY 功能丰度的 LDA 分析结果，默认以 level2 层级丰度进行分析】 .....	14

			-- eggNOG——【eggNOG 数据库分析结果】 .....	14
			-- eggNOG_Anno——【eggNOG 注释结果统计】 .....	14
			-- eggNOG_MAT——【eggNOG 相对丰度和绝对丰度分析结果：level1 为第一层级，level2 为第二层级，og 为直系同源簇】 .....	17
			-- GeneNums——【注释到的基因数目、基因 id 统计，相对丰度和绝对丰度分析结果】 .....	18
			-- GeneNums.BetweenSamples——【各样品中基因数目统计，相对丰度和绝对丰度分析结果】 .....	19
			---top/top_group——【各样品注（组）释到的基因数目的柱状图】 .....	19
			-- GeneNums.BetweenSamples.heatmap——【各样品注释到的基因数目的聚类图，相对丰度和绝对丰度分析结果】 .....	20
			-- heatmap/heatmap_group——【各样品(组)中的酶的相对丰度聚类图以及作图数据，相对丰度和绝对丰度分析结果】 .....	21
			-- MetaStats——【各样品的 MetaStats 统计结果，level2 为第二层级，og 为直系同源簇】 .....	21
			`-- PCA ——【各样品的 PCA 分析结果】 .....	21
			`-- PCoA ——【各样品的 PCoA 分析结果】 .....	22
			`-- NMDS——【各样品的 NMDS 分析结果，og 为直系同源簇，level1 为第一层级，level2 为第二层级】 .....	22
			`-- Anosim——【各组 Anosim 分析结果，默认以 og 和 level1 层级丰度进行分析】 .....	22
			-- LDA——【基于 eggNOG 功能丰度的 LDA 分析结果，默认以 og 层级丰度进行分析】 .....	23
			`-- KEGG ——【KEGG 数据库分析结果】 .....	23
			-- GeneNums——【注释到的基因数目、基因 id 统计，ec 为酶，ko 为直系同源的功能，level1 为生物代谢通路，level2 为子功能，level3 为代谢通路图，module 为模块分析】 .....	23
			-- GeneNums.BetweenSamples——【各样品中基因数目统计】 .....	24

---

	--top/top_group——【各样品注（组）释到的基因数目的柱状图】 .....	24
	-- GeneNums.BetweenSamples.heatmap——【各样品注释到的基因数目的聚类图】 .....	25
	-- heatmap/heatmap_group——【各样品中的酶的相对丰度聚类图以及作图数据】 .....	26
	-- KEGG_Anno——【KEGG 注释结果统计】 .....	26
	-- KEGG_MAT——【KEGG 相对丰度和绝对丰度分析结果：ec 为酶，ko 为直系同源的功能】 .....	29
	-- MetaStats——【各样品的 MetaStats 统计结果】 .....	31
	-- pathwaymaps——【代谢通路比较结果】 .....	31
	`-- PCA——【各样品的 PCA 分析结果】 .....	31
	`-- NMDS——【各样品的 NMDS 分析结果】 .....	32
	`-- Anosim——【各组 Anosim 分析结果，默认以 ko、module、level1、level2、level3 层级丰度进行分析】 .....	33
	-- LDA——【基于 KEGG 功能丰度的 LDA 分析结果，默认以 KO 层级丰度进行分析】 .....	33
	-- CARD——【CARD 数据库分析结果】 .....	34
	-- Readme.pdf【CARD 数据库分析结果 readme 文档】 .....	34
	-- stat_result【对 Unigenes 中 ARO 预测结果的分析目录】 .....	34
	-- bar【柱状图结果】 .....	34
	-- box【箱形图结果】 .....	34
	-- Venn_flower【样品（组）抗性基因韦恩（花瓣）图】 .....	35
	-- heatmap/heatmap_group【top30 ARO 各样品中丰度热图结果目录】 .....	36

---

	-- stat.ARO.absolute.xls 【各样品中 ARO 的绝对丰度信息表, ARO 按最大值排序】 .....	37
	-- stat.ARO.id.xls 【各样品中 ARO 的绝对丰度信息表, 含有抗性基因 ID, 未排序】 .....	37
	-- stat.ARO.relative.xls 【各样品中 ARO 的相对丰度信息表, ARO 按最大值排序】 .....	37
	`-- twocircle.....	37
	`-- MD_taxonomy.{png,svg} 【某分组中抗性基因与物种归属关系双圈图】 .....	37
	`-- Unigenes.protein.rgi.del.txt 【RGI 软件预测的结果文件】 .....	37
<hr/>		
-- 05.FunctionAnnotation --ReadMe.pdf ——— 【05. FunctionAnnotation 交付结果目录说明】 .....		39



---

## |-- 05.FunctionAnnotation——【常用功能数据库注释及丰度分析结果】

### | |-- CAZy——【CAZy 数据库分析结果】

#### | | |-- CAZy\_Anno——【CAZy 注释结果统计】

#### | | | |-- cazy.unigenes.num.{pdf|png}——【注释到 CAZy 第一层级的基因数目统计图，pdf 及 png 格式】

对应的是结题报告中的 CAZy 注释结果统计图，图中，横轴是数据库中各功能类型的代码，代码的解释见对应的图例说明，纵轴代表注释为相应功能类的基因数目。

#### | | | |-- cazy.unigenes.num.txt ——【注释到 CAZy 第一层级的基因数目统计结果】

储存了六大功能类对应的总体注释结果，可以用 excel 打开该文件，在该文件中，各列所代表的含义如下：

列数	列标题	说明
1	Name	六大功能类的缩写

---

2	Discription	各功能类对应的描述
3	Unique.Genes.filter.anno	该功能类注释上的基因数目

---

| | | |-- Unigenes.blast.m8.filter.anno.xls ——【过滤后的 blast 结果的注释信息】

为基于 Unique.Genes.filter 进行的 CAZy 注释结果，可以用 excel 打开。该文件共分为 4 列，各列所代表的含义如下：

列数	列标题	说明
1	Gene_id	基因的 ID 号
2	Subject_id	比对上的序列在 genebank 中的 accession number
3	CAZy_Family	比对上的序列所属的子功能类
4	Family_Description	该功能类的描述

---

---

| | | |-- Unigenes.blast.m8.filter.xls —— **【过滤后的 blast 结果文件，Blast 软件的 m8 格式】**

即为 blast 后的 m8 格式的文件，关于 m8 格式的详细解释可以参看结题报告中的常用数据格式说明。

| | | |-- Unigenes.level1.bar.{png|svg}—— **【CAZy 第一层级上的相对丰度统计图】**

对应的是结题报告中的 CAZy 基因注释结果在第一层级上的统计图，纵轴表示注释到某类型的功能的相对比例；横轴表示样品名称；各颜色区块对应的功能类别见右侧图例。

| | | |-- Unigenes.level1.bar.tree.{png|svg}—— **【CAZy 功能聚类分析结果】**

对应的是结题报告中的 CAZy 功能聚类分析图，图中心是 BC 距离聚类树结构，外圈各层是各样品在第一层级上的功能相对丰度分布，各颜色区块对应的功能类别见左上角图例。

| | | |-- Unigenes.CAZY.tax.xls—— **【CAZY 各个层级对应的物种注释信息】**

该文件为通过注释到 CAZY 不同层级的基因对应到其相关的物种注释信息，可以通过该文件筛选关注功能的物种信息。



---

| | | |-- CAZy\_MAT——【CAZy 相对丰度和绝对丰度分析结果：ec 为酶，level1 为六大功能类，level2 为子功能】

| | | |-- Absolute ——【各样品在 ec，level1 和 level2 不同水平的绝对丰度矩阵】

在该文件夹中，一共有三个文件，可以用 excel 打开这些文件，其中 Unigenes.absolute.level1.xls 是对第一层级的六大功能类在各样品中的绝对丰度进行的统计，在该文件中，第一行为样品信息，第一列为第一层级六大功能类的代号，最后一列为代号的详细说明，其余的数字则代表某个功能类在某个样品中的绝对丰度。

在 Absolute 文件夹中，存在的另外两个文件 Unigenes.absolute.level2.xls 和 Unigenes.absolute.ec.xls 的展示形式和 Unique.Genes.level1.absolute.xls 文件的展示形式是一样的，不同的是，Unigenes.absolute.level2.xls 是对子功能类在各样品中的绝对丰度进行的统计，而 Unigenes.absolute.ec.xls 则是对所有能够注释上 CAZy 数据库的基因所属的酶，在各样品中的绝对丰度进行的统计。

| | | |-- Relative ——【各样品（组）在 ec，level1 和 level2 不同水平的相对丰度矩阵】

在该文件夹中，一共有三个文件，可以用 excel 打开这些文件，其中 Unigenes.relative.level1.xls 是对第一层级的六大功能类在各样品中的相对丰度进行的统计，在该文件中，第一行为样品信息，第一列为第一层级六大功能类的代号，最后一列为代号的详细说明，其余的数字则代表某个功能类在某个样品中的相对丰度。

---

在 Relative 文件夹中，存在的另外两个文件 Unigenes.relative.level2.xls 和 Unigenes.relative.ec.xls 的展示形式和 Unique.Genes.level1.relative.xls 文件的展示形式是一样的，不同的是，Unigenes.relative.level2.xls 是对子功能类在各样品中的相对丰度进行的统计，而 Unigenes.absolute.ec.xls 则是对所有能够注释上 CAZy 数据库的基因所属的酶，在各样品中的相对丰度进行的统计。

该文件夹中的另外三个文件为以组别为单位的数据他统计。

### | | |-- GeneNums——【注释到的基因数目、基因 id 统计，ec 为酶，level1 为六大功能类，level2 为子功能】

在该文件夹中有三个文件，文件展示形式类似。文件 Unigenes.absolute.level1.xls，共三列，第一列为第一层级六大功能类的代号，第二列为注释上的基因数，第三列为注释上该功能的所有基因 id。文件 Unigenes.absolute.level2.xls，共四列，第一列为子功能名称，第二列为子功能的详细描述，第三列为注释上的基因数，第四列为注释上该功能的所有基因 id。文件 Unigenes.absolute.ec.xls，共三列，第一列为酶的名称，第二列为注释上该酶的基因数，第三列为注释上该酶的所有基因 id。

### | | |-- GeneNums.BetweenSamples——【各样品中基因数目统计】

在该文件夹中有三个文件，文件展示形式相同。文件 Unigenes.absolute.level1.xls, Unigenes.absolute.level2.xls, Unigenes.absolute.ec.xls, 内容为不同功能水平上在各样品中注释上的基因数目统计结果。

---

| | |-- top/top\_group——【各样品（组）注释到的基因数目的柱状图】

| | | |-- figure ——【样品（组）在各水平丰度排名前 10 柱状图，svg 及 png 格式】

| | | `-- table ——【样品（组）在各水平丰度排名前 10 柱状图分析所使用的文件】

| | |-- GeneNums.BetweenSamples.heatmap——【各样品注释到的基因数目的聚类图】

| | | |-- ec——【各样品中的酶的绝对丰度聚类图】

| | | |-- level1 ——【各样品中在 CAZy 第一层级的绝对丰度聚类图】

| | | `-- level2 ——【各样品中在 CAZy 第二层级的绝对丰度聚类图】

以上 ec、level1、level2 三个文件夹中的结果，是从 GeneNums.BetweenSamples 中的表格出发，进行聚类得到的热图。

---

| | |-- heatmap/heatmap\_group——【各样品中的酶的相对丰度聚类图以及作图数据】

| | |-- MetaStats——【各样品的 MetaStats 统计结果，EC 为酶，level2 为子功能】

| | | |-- EC ——【各样品在 EC 水平的 MetaStats 统计结果】

| | | `-- level2 ——【各样品在 CAZy 第二层级的 MetaStats 统计结果】

| | `-- PCA ——【各样品的 PCA 分析结果，EC 为酶，level1 为六大功能类，level2 为子功能】

| | ||-- ec——【各样品在 EC 水平的 PCA 分析结果】

| | ||-- level1 ——【各样品在 CAZy 第一层级的 PCA 分析结果】

| | ||`-- level2 ——【各样品在 CAZy 第二层级的 PCA 分析结果】

| | `-- PCoA ——【各样品的 PCoA 分析结果，EC 为酶，level1 为六大功能类，level2 为子功能】

---

| | ||-- ec——【各样品在 EC 水平的 PCoA 分析结果】

| | ||-- level1 ——【各样品在 CAZy 第一层级的 PCoA 分析结果】

| | ||`-- level2 ——【各样品在 CAZy 第二层级的 PCoA 分析结果】

| | `-- NMDS——【各样品的 NMDS 分析结果，EC 为酶，level1 为六大功能类，level2 为子功能】

| | ||-- ec——【各样品在 EC 水平的 NMDS 分析结果】

| | ||-- level1 ——【各样品在 CAZy 第一层级的 NMDS 分析结果】

| | ||`-- level2 ——【各样品在 CAZy 第二层级的 NMDS 分析结果】

| | |`-- Anosim——【各组 Anosim 分析结果】

分别在 level、level2、ec 三个层级进行分析，每个层级结果如下：

---

| | | |--{pdf|png}——【各个分组间 Anosim 分析箱图，pdf 和 png 格式】

| | | |--stat\_anosim.txt ——【Anosim 分析的统计结果】

| | |--LDA——【基于 CAZY 功能丰度的 LDA 分析结果，默认以 level2 层级丰度进行分析】

| | | |--LDA ——【LDA 值分布柱状图结果】

| | | |--heatmap——【LDA 筛选的差异功能丰度聚类热图】

| |--eggNOG——【eggNOG 数据库分析结果】

| | |--eggNOG\_Anno——【eggNOG 注释结果统计】

| | | |--eggNOG.unigenes.num.{pdf|png}——【注释到 eggNOG 第一层级的基因数目统计图，pdf 及 png 格式】

对应的是结题报告中的 eggNOG 注释结果统计图,图中，横轴是数据库中各功能类型的代码，代码的解释见对应的图例说明，纵轴代表注释为相应功能类的基因数目。

---

| | | |-- eggNOG.unigenes.num.txt —— 【注释到 eggNOG 第一层级的基因数目统计结果】

储存了 25 大功能类对应的总体注释结果，可以用 excel 打开该文件，在该文件中，各列所代表的含义如下：

列数	列标题	说明
1	Functional_Category	25 大功能类的缩写
2	Description	各功能类对应的描述
3	Num	该功能类注释上的基因数目

| | | |-- Unigenes.blast.m8.filter.anno.xls —— 【过滤后的 blast 结果的注释信息】

储存的是基于 blast m8 文件进行的 eggNOG 注释结果，该文件可以用 excel 打开，该文件共分为 5 列，各列所代表的含义如下：

列数	列标题	说明
1	Query_id	基因的 ID 号

---

4	Subject_id	比对上的序列号
5	Ortholog_Group	该序列所属的 Orthologous id
6	Functional_Category	该 Orthologous id 所属的功能类代号
7	OG_Description	该 Orthologous id 多对应的描述

---

Nevo gene  
诺木致源

| | | |-- Unigenes.blast.m8.filter.xls ——【过滤后的 blast 结果文件，Blast 软件的 m8 格式】

即为 blast 后的 m8 格式的文件，关于 m8 格式的详细解释可以参看结题报告中的常用数据格式说明。

| | | |-- Unigenes.level1.bar.{png/svg} ——【eggNOG 第一层级上的相对丰度统计图】

对应的是结题报告中的 eggNOG 基因注释结果在第一层级上的统计图，纵轴表示注释到某类型的功能的相对比例；横轴表示样品名称；各颜色区块对应的功能类别见右侧图例。



---

| | | |-- Unigenes.level1.bar.tree.{png|svg}——【eggNOG 功能聚类分析结果，pdf 与 png 格式】

对应的是结题报告中的 eggNOG 功能聚类分析图，图中心是欧氏距离聚类树结构，外圈各层是各样品在第一层级上的功能相对丰度分布，各颜色区块对应的功能类别见左上角图例。

| | | |-- Unigenes.eggNOG.tax.xls——【eggNOG 各个层级对应的物种注释信息】

该文件为通过注释到 eggNOG 不同层级的基因对应到其相关的物种注释信息，可以通过该文件筛选关注功能的物种信息。

| | | |-- eggNOG\_MAT——【eggNOG 相对丰度和绝对丰度分析结果：level1 为第一层级，level2 为第二层级，og 为直系同源簇】

| | | |-- Absolute ——【各样在 level1，level2 和 og 水平注释到的绝对丰度矩阵】

在该文件夹中，一共有三个文件，可以用 excel 打开这些文件，其中 Unigenes.absolute.level1.xls 是对第一层级的 25 大功能类在各样品中的绝对丰度进行的统计，在该文件中，第一行为样品信息，第一列为第一层级 25 大功能类的代号，最后一列为代号的详细说明，其余的数字则代表某个功能类在某个样品中的绝对丰度。

---

在 Absolute 文件夹中，存在的另外两个文件 Unigenes.absolute.level2.xls 和 Unigenes.absolute.og.xls 的展示形式和 Unique.Genes.level1.absolute.xls 文件的展示形式是一样的，不同的是，Unigenes.absolute.level2.xls 是对子功能类在各样品中的绝对丰度进行的统计，而 Unigenes.absolute.og.xls 则是对所有能够注释上 eggNOG 数据库的基因所属的 Orthologous id，在各样品中的绝对丰度进行的统计。

#### | | | `-- Relative ——【各样在 level1, level2 和 og 水平注释到的相对丰度矩阵】

在该文件夹中，一共有三个文件，可以用 excel 打开这些文件，其中 Unigenes.relative.level1.xls 是对第一层级的 25 大功能类在各样品中的绝对丰度进行的统计，在该文件中，第一行为样品信息，第一列为第一层级 25 大功能类的代号，最后一列为代号的详细说明，其余的数字则代表某个功能类在某个样品中的绝对丰度。

在 Relative 文件夹中，存在的另外两个文件 Unigenes.relative.level2.xls 和 Unigenes.relative.og.xls 的展示形式和 Unique.Genes.level1.relative.xls 文件的展示形式是一样的，不同的是，Unigenes.absolute.level2.xls 是对子功能类在各样品中的绝对丰度进行的统计，而 Unigenes.relative.og.xls 则是对所有能够注释上 eggNOG 数据库的基因所属的 Orthologous id，在各样品中的绝对丰度进行的统计。

该文件夹包含以组为单位分析结果。

#### | | | |-- GeneNums——【注释到的基因数目、基因 id 统计，相对丰度和绝对丰度分析结果】

在该文件夹中有三个文件，文件展示形式类似。文件 Unigenes.absolute.level1.xls，共三列，第一列为第一层级 25 大功能类的代号，

---

第二列为注释上的基因数，第三列为注释上该功能的所有基因 id。文件 Unigenes.absolute.level2.xls，共四列，第一列为子功能名称，第二列为子功能的详细描述，第三列为注释上的基因数，第四列为注释上该功能的所有基因 id。文件 Unigenes.absolute.og.xls，共三列，第一列为酶的名称，第二列为注释上该 Orthologous id 的基因数，第三列为注释上该 Orthologous id 的所有基因 id。

| | |-- GeneNums.BetweenSamples——【各样品中基因数目统计，相对丰度和绝对丰度分析结果】

在该文件夹中有三个文件，文件展示形式相同。文件 Unigenes.absolute.level1.xls，Unigenes.absolute.level2.xls，Unigenes.absolute.og.xls，内容为不同功能水平上在各样品中注释上的基因数目统计结果。

| | |--top/top\_group——【各样品注（组）释到的基因数目的柱状图】

| | | |-- figure ——【样品（组）在各水平（包括 level1、level2、og）丰度排名前 10 柱状图，svg 及 png 格式】

| | | `-- table ——【样品（组）在各水平（包括 level1、level2、og）丰度排名前 10 柱状图分析所使用的文件】

---

| | | -- GeneNums.BetweenSamples.heatmap——【各样品注释到的基因数目的聚类图，相对丰度和绝对丰度分析结果】

| | | -- og——【各样品中在 eggNOG 第一层级的绝对丰度聚类图】

| | | --level1——【各样品中在 eggNOG 第一层级的绝对丰度聚类图】

以上 og、level1 两个文件夹中的结果，是从 GeneNums.BetweenSamples 中的表格出发，进行聚类得到的热图。

Novogene  
诺禾致源

---

| | |-- heatmap/heatmap\_group——【各样品(组)中的酶的相对丰度聚类图以及作图数据, 相对丰度和绝对丰度分析结果】

| | |-- MetaStats——【各样品的 MetaStats 统计结果, level2 为第二层级, og 为直系同源簇】

| | | |-- level2 ——【各样品在 eggNOG 第二层级的 MetaStats 统计结果】

| | | `-- og——【各样品在 og 水平的 MetaStats 统计结果】

| | `-- PCA ——【各样品的 PCA 分析结果】

| | | |-- level1 ——【各样品在 eggNOG 第一层级的 PCA 分析结果】

| | | |-- level2 ——【各样品在 eggNOG 第二层级的 PCA 分析结果】

| | | `-- og——【各样品在 og 水平的 PCA 分析结果】

---

| | `-- PCoA ——【各样品的 PCoA 分析结果】

| | |-- level1 ——【各样品在 eggNOG 第一层级的 PCoA 分析结果】

| | |-- level2 ——【各样品在 eggNOG 第二层级的 PCoA 分析结果】

| | `-- og——【各样品在 og 水平的 PCoA 分析结果】

| | `-- NMDS——【各样品的 NMDS 分析结果，og 为直系同源簇，level1 为第一层级，level2 为第二层级】

| | ||-- og——【各样品在 og 水平的 NMDS 分析结果】

| | ||-- level1 ——【各样品在 eggNOG 第一层级的 NMDS 分析结果】

| | ||`-- level2 ——【各样品在 eggNOG 第二层级的 NMDS 分析结果】

| | |`-- Anosim——【各组 Anosim 分析结果，默认以 og 和 level1 层级丰度进行分析】

---

| | | |-- \*.{pdf|png}——【各个分组间 Anosim 分析箱图, pdf 和 png 格式】

| | | |-- stat\_anosim.txt ——【Anosim 分析的统计结果】

| | |-- LDA——【基于 eggNOG 功能丰度的 LDA 分析结果, 默认以 og 层级丰度进行分析】

| | | |-- LDA ——【LDA 值分布柱状图结果】

| | | |-- heatmap——【LDA 筛选的差异功能丰度聚类热图】

| |-- KEGG ——【KEGG 数据库分析结果】

| |-- GeneNums——【注释到的基因数目、基因 id 统计, ec 为酶, ko 为直系同源的功能, level1 为生物代谢通路, level2 为子功能, level3 为代谢通路图, module 为模块分析】

在该文件夹中, 一共有六个文件, 可以用 excel 打开这些文件, 其中 Unigenes.absolute.level1.xls, 共三列, 第一列为第一层级六大功能类的代号, 第二列为注释上的基因数, 第三列为注释上该功能的所有基因 id。Unigenes.absolute.level2.xls、Unigenes.absolute.level3.xls、Unigenes.absolute.ko.xls、Unigenes.absolute.ec.xls 和 Unigenes.absolute.module.xls 也是三列, 第一列分别为 level2、level3、Orthologous groups 的 id、酶和 module 的名称, 第二列为注释上的基因数, 第三列为注释上该功能的所有基因 id。

---

## |       |-- GeneNums.BetweenSamples——【各样品中基因数目统计】

在该文件夹中有六个文件，文件展示形式相同。文件 Unigenes.absolute.level1.xls，Unigenes.absolute.level2.xls，Unigenes.absolute.level3.xls，Unigenes.absolute.ec.xls，Unigenes.absolute.ko.xls、Unigenes.absolute.module.xls 内容为不同功能水平上在各样品中注释上的基因数目统计结果。

## |       |--top/top\_group——【各样品注（组）释到的基因数目的柱状图】

|   |   |-- figure ——【样品（组）在各水平（包括 level1、level2、level3、ko、ec、module）丰度排名前 10 柱状图，svg 及 png 格式】

|   |   `-- table ——【样品（组）在各水平（包括 level1、level2、level3、ko、ec、module）丰度排名前 10 柱状图分析所使用的文件】



---

|       |-- GeneNums.BetweenSamples.heatmap——【各样品注释到的基因数目的聚类图】

|       |   |-- ec——【各样品中在酶水平的绝对丰度聚类图】

|       |   |-- ko——【各样品中在 KO 水平的绝对丰度聚类图】

|       |   |-- module——【各样品中在 moduoe 水平的绝对丰度聚类图】

|       |   |-- level1——【各样品中在 KEGG 第一层级的绝对丰度聚类图】

|       |   |-- level2——【各样品中在 KEGG 第二层级的绝对丰度聚类图】

|       |   `-- level3——【各样品中在 KEGG 第三层级的绝对丰度聚类图】

以上 ec、ko、module、level1、level2 和 level3 六个文件夹中的结果，是从 GeneNums.BetweenSamples 中的表格出发，进行聚类得到的热图。

---

|       |-- heatmap/heatmap\_group——【各样品中的酶的相对丰度聚类图以及作图数据】

|       |-- KEGG\_Anno——【KEGG 注释结果统计】

|       |   |-- kegg.unigenes.num.{pdf|png}——【注释到 KEGG 第一层级的基因数目统计图】

对应的是结题报告中的 KEGG 注释结果统计图,图中，左侧为代谢通路的描述，图中条形图上的数字为注释为该通路的基因数目。

|       |   |-- kegg.unigenes.num.txt——【注释到 KEGG 第一层级的基因数目统计图】

该文件一共分为三列，各列所代表的含义如下：

列数	列标题	说明
1	First Level	第一层六大代谢通路的描述
2	Second Level	注释出来的属于该代谢通路的第二层子通路的描述
3	Gene of Unique.Genes.KEGG.catalog	该子通路注释上的基因数目

---

| | | -- Unigenes.blast.m8.filter.anno.xls——【过滤后的 blast 结果的注释信息】

储存的是基于 blast m8 文件进行的 KEGG 注释结果，该文件也可以用 excel 打开，共分为 7 列，在最后一列中，不同的 pathway 之间用相应的 pathway 编号隔开，在每一个 pathway 中，都分为三个层级，第一层级为六大代谢通路，第二层级为 43 种子通路，第三层级则为具体的通路信息。在该文件中，各列所代表的含义如下：

列数	列标题	说明
1	Query_id	基因的 ID 号
2	Kegg_geneID	比对上的序列 ID 号也即 KEGG 中的 geneID 号
3	KO_ID	该序列所对应的，在 KEGG 数据库中的 Orthology ID 号
4	KO_NAME	该 Orthology ID 所对应的名称
5	KO_DEFINITION	该 Orthology ID 所对应的描述
6	KO_EC	该 Orthology ID 所对应的 EC 编号

---

7	Module	该 Orthologous id 所对应的 Module 信息
8	KO_PATHWAY	该 Orthologous id 所对应的 pathway 信息

---

| | |-- Unigenes.blast.m8.filter.xls——【过滤后的 blast 结果文件，Blast 软件的 m8 格式】

即为 blast 后的 m8 格式的文件，关于 m8 格式的详细解释可以参看结题报告中的常用数据格式说明。

| | |-- Unigenes.level1.bar.{png|svg}——【KEGG 第一层级上的相对丰度统计图】

对应的是结题报告中的 KEGG 基因注释结果在第一层级上的统计图，纵轴表示注释到某类型的代谢通路的相对比例；横轴表示样品名称；各颜色区块对应的代谢通路类别见右侧图例。

| | |-- Unigenes.level1.bar.tree.{png|svg}——【KEGG 功能聚类分析结果】

对应的是结题报告中的 KEGG 功能聚类分析图，图中心是欧氏距离聚类树结构，外圈各层是各样品在第一层级上的代谢通路相对丰度分布，各颜色区块对应的代谢通路见左上角图例。

---

| | | |-- Unigenes.KEGG.tax.xls——【KEGG 各个层级对应的物种注释信息】

该文件为通过注释到 KEGG 不同层级的基因对应到其相关的物种注释信息，可以通过该文件筛选关注功能的物种信息。

| |-- KEGG\_MAT——【KEGG 相对丰度和绝对丰度分析结果：ec 为酶，ko 为直系同源的功能】

| | |-- Absolute——【各样品在 ec, ko、module、level1, level2 和 level3 绝对丰度矩阵】

在该文件夹中，一共有六个文件，可以用 excel 打开这些文件，其中 Unigenes.absolute.level1.xls 是对第一层级的六大生物代谢通路在各样品中的绝对丰度进行的统计，在该文件中，第一行为样品信息，第一列为第一层级六大生物代谢通路的描述，其余的数字则代表某个代谢通路在某个样品中的绝对丰度。

在 Absolute 文件夹中，存在的另外五个文件 Unigenes.absolute.level2.xls、Unigenes.absolute.level3.xls、Unigenes.absolute.ko.xls、Unigenes.absolute.ec.xls 和 Unigenes.absolute.module.xls 的展示形式和 Unigenes.absolute.level1.xls 文件的展示形式是一样的，不同的是，Unigenes.absolute.level2.xls 是对子通路在各样品中的绝对丰度进行的统计，Unigenes.absolute.level3.xls 是对代谢通路图在各样品中的绝对丰度进行的统计，Unigenes.absolute.ko.xls 是对 Orthologous groups 在各样品中的绝对丰度进行的统计，而 Unigenes.absolute.ec.xls 则是对所有能够注释上 KEGG 数据库的基因所属的酶，Unigenes.absolute.module.xls 是 module 在个样本种的绝对丰度进行的统计，在各样品中的绝对丰度进行的统计。

---

## | | `-- Relative——【各样品在 ec, ko、module、level1, level2 和 level3 相对丰度矩阵】

在该文件夹中，一共有六个文件，可以用 excel 打开这些文件，其中 Unigenes. relative.level1.xls 是对第一层级的六大生物代谢通路在各样品中的相对丰度进行的统计，在该文件中，第一行为样品信息，第一列为第一层级六大生物代谢通路的描述，其余的数字则代表某个代谢通路在某个样品中的相对丰度。

在 Absolute 文件夹中，存在的另外五个文件 Unigenes. relative.level2.xls、Unigenes. relative.level3.xls、Unigenes. relative.ko.xls、Unigenes. relative.ec.xls 和 Unigenes.relative.module.xls 的展示形式和 Unigenes.relative.level1.xls 文件的展示形式是一样的，不同的是，Unigenes. relative.level2.xls 是对子通路在各样品中的相对丰度进行的统计，Unigenes. relative.level3.xls 是对代谢通路图在各样品中的相对丰度进行的统计，Unigenes. relative.ko.xls 是对 Orthologous groups 在各样品中的相对丰度进行的统计，Unigenes. relative.ec.xls 则是对所有能够注释上 KEGG 数据库的基因所属的酶，在各样品中的相对丰度进行的统计，Unigenes.relative.module.xls 是对 module 在各样本的相对丰度进行的统计。

诺禾致源

---

|       |-- MetaStats——【各样品的 MetaStats 统计结果】

|       |   |-- ec——【各样品在酶水平的 MetaStats 统计结果】

|       |   |-- ko——【各样品在 KO 水平的 MetaStats 统计结果】

|       |   |-- module——【各样品在 KO 水平的 MetaStats 统计结果】

|       |   |-- level2——【各样品在 module 水平的 MetaStats 统计结果】

|       |   `-- level3——【各样品在 KEGG 第三层级的 MetaStats 统计结果】

|       |-- pathwaymaps——【代谢通路比较结果】

|       `-- PCA——【各样品的 PCA 分析结果】

|           |-- ec——【各样品在酶水平的 PCA 分析结果】



---

|           |-- ko——【各样品在 KO 水平的 PCA 分析结果】

|           |-- module——【各样品在 module 水平的 PCA 分析结果】

|           |-- level1——【各样品在 KEGG 第一层级的 PCA 分析结果】

|           |-- level2——【各样品在 KEGG 第二层级的 PCA 分析结果】

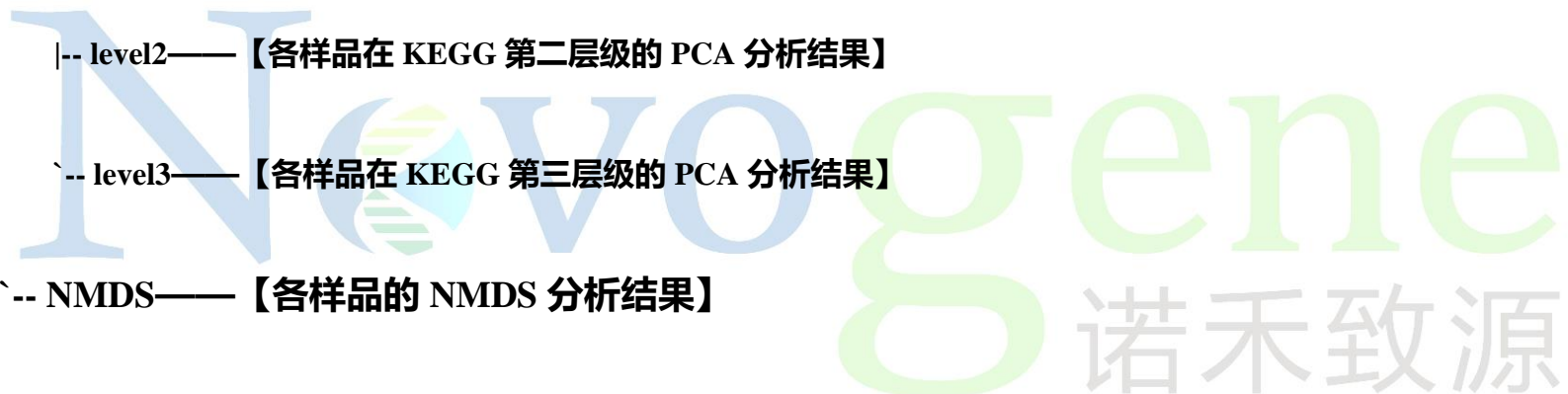
|           `-- level3——【各样品在 KEGG 第三层级的 PCA 分析结果】

| |    `-- NMDS——【各样品的 NMDS 分析结果】

| | |   |-- ec——【各样品酶水平的 NMDS 分析结果】

| | |   |-- ko——【各样品 KO 水平的 NMDS 分析结果】

| | |   |-- module——【各样品 module 水平的 NMDS 分析结果】





---

| | |-- level1 ——【各样品在 KEGG 第一层级的 NMDS 分析结果】

| | |-- level2 ——【各样品在 KEGG 第二层级的 NMDS 分析结果】

| | |-- level3——【各样品在 KEGG 第三层级的 NMDS 分析结果】

| | |-- Anosim——【各组 Anosim 分析结果，默认以 ko、module、level1、level2、level3 层级丰度进行分析】

| | | |-- \*.{pdf|png}——【各个分组间 Anosim 分析箱图，pdf 和 png 格式】

| | | |-- stat\_anosim.txt ——【Anosim 分析的统计结果】

| | |-- LDA——【基于 KEGG 功能丰度的 LDA 分析结果，默认以 KO 层级丰度进行分析】

| | | |-- LDA ——【LDA 值分布柱状图结果】

| | | |-- heatmap——【LDA 筛选的差异功能丰度聚类热图】

---

|-- CARD——【CARD 数据库分析结果】

| |-- Readme.pdf【CARD 数据库分析结果 readme 文档】

| |-- stat\_result【对 Unigenes 中 ARO 预测结果的分析目录】

| | |-- bar【柱状图结果】

| | | |-- stat.ARO.ppm.{png,svg}【top20 的 ARO 在各样品的相对丰度柱状图】

| | | |-- stat.ARO.RelativePercent.{png,svg}【top20 的 ARO 在各样品的相对百分含量柱状图】

| | | |-- stat.ARO.relative.per.xls【top20 的 ARO 在各样品的相对百分含量表格】

| | | `-- stat.ARO.relative.ppm.xls【top20 的 ARO 在各样品的相对丰度表格】

| | |-- box【箱形图结果】

| | | |-- arobox【ARO 数目箱型图】

| | | | |-- aro.xls【各样品 ARO 数目表格】

| | | | `-- group.ARObox.{png,pdf}【组间 ARO 数目箱型图】

---

```
| | | `-- genebox 【基因数目箱型图】
```

```
| | |      |-- gene.xls 【各样品基因数目表格】
```

```
| | |      `-- group.genebox.{png,pdf} 【组间各样品抗性基因数目箱型图】
```

```
| | |-- circos 【各样品中 ARO 丰度圈图结果目录】
```

```
| | | |-- circos.overview.{png,svg} 【各样品中 top10 ARO 丰度圈图】
```

```
| | | `-- stat.ARO.relative.circos.xls 【top10 的 ARO 在各样品中相对丰度表，单位为 ppm】
```

圈图分为两个部分，右侧为样品信息，左侧为 ARO。内圈不同颜色表示不同的样品和 ARO，刻度为相对丰度，单位为 ppm，左侧为样品中 ARO 的相对丰度之和，右侧为各 ARO 在样本中的相对丰度之和；外圈左侧为各个 ARO 中各样本所占的相对百分含量，外圈右侧为各样本中各个 ARO 的相对百分含量。

```
| | |-- getARO.xls 【各样品中抗性基因的丰度及 ARO 归属信息表】
```

```
| | |-- Venn_flower 【样品（组）抗性基因韦恩（花瓣）图】
```

```
| | | |-- venn_flow{pdf,png} 【样品韦恩（花瓣）图】
```

---

```
| | | |-- venn_flowe_G{pdf,png} 【组韦恩（花瓣）图】
```

当样本（组）数小于 5 时，展示韦恩图，当样本（组）数超过 5 个时，展示花瓣图；图中，每个圈代表一个样品；圈和圈重叠部分的数字代表样品之间共有的抗性基因个数；没有重叠部分的数字代表样品的特有抗性基因个数。

```
| | |-- heatmap/heatmap_group 【top30 ARO 各样品中丰度热图结果目录】
```

```
| | | |-- aro_bw
```

```
| | | | |-- bw.{pdf,png} 【各样品中 top30 ARO 黑白热图】
```

```
| | | | `-- stat.ARO.relative.bw.xls 【各样品中 top30 ARO 丰度表】
```

横轴为样品名称，右侧纵轴为 ARO 名称，上方颜色为样本所在分组的信息；黑白热图中黑色表示样品中含有 ARO，白色表示样品中没有该 ARO。（如果该 top30 ARO 在所有样本中均含有则不展示该图片）

```
| | | `-- aro_heat
```

```
| | | | |-- heat.{pdf,png} 【各样品中 top30 ARO 丰度聚类热图】
```

```
| | | | `-- stat.ARO.relative.heat.xls 【各样品中 top30 ARO 丰度表】
```

横轴为样品名称，右侧纵轴为 ARO 名称，上方颜色为样本所在分组的信息，中间热图对应的值为每一行 ARO 相对丰度经过标准化处理后得到的 Z 值。

---

```
| | |-- stat.ARO.absolute.xls 【各样品中 ARO 的绝对丰度信息表, ARO 按最大值排序】
| | |-- stat.ARO.id.xls 【各样品中 ARO 的绝对丰度信息表, 含有抗性基因 ID, 未排序】
| | |-- stat.ARO.relative.xls 【各样品中 ARO 的相对丰度信息表, ARO 按最大值排序】
| | `-- twocircle
| |
| | |-- card.gene.absolute.total.tax.xls 【各样本中抗性基因的丰度及物种归属信息表】
| | `-- MD_taxonomy.{png,svg} 【某分组中抗性基因与物种归属关系双圈图】
| `-- Unigenes.protein.rgi.del.txt 【RGI 软件预测的结果文件】
```

---

列标题	说明
ORF_ID	基因 ID
CUT_OFF	比对模式 (默认严格筛选, strict) 。

---

---

PASS_EVALUE	该 AROblast 比对时筛选 e-value 阈值。
-------------	------------------------------

Best_Hit_evalue	比对实际得到的 e-value 值
-----------------	-------------------

Best_Hit_ARO	比对得到的最佳 ARO
--------------	-------------

Best_Identities	序列与 ARO 序列相似度得分
-----------------	-----------------

ARO	该基因 ID 可以比对出的全部 ARO 编号
-----	------------------------

ARO_name	该基因 ID 可以比对出的全部 ARO 名称
----------	------------------------

Model_type	RGI 软件分析 ARO 时采用的计算模型
------------	-----------------------

SNP	SNP 模型
-----	--------

ARO_category	ARO 种类描述信息
--------------	------------

bit_score	ARO 的打分值
-----------	----------

Predicted_Protein	某基因中预测出的蛋白序列
-------------------	--------------

CARD_Protein_Sequence	CARD 数据库中 ARO 的蛋白序列。
-----------------------	----------------------

LABEL	基因 ID
-------	-------

---

ID

ARO 的 ID 信息

---

`-- 05.FunctionAnnotation --ReadMe.pdf      —— 【05. FunctionAnnotation 交付结果目录说明】

Novogene  
诺禾致源