
诺禾致源
元基因组交付目录说明手册
(V3.1)



2015 年 09 月 30 日

目录

(注: 单击即可跳转至相应文档的详细说明)

---05.FUNCTIONANNOTATION --- 【功能注释分析结果】	4
-- CAZY --- 【CAZY 数据库分析结果】	4
---CAZY_ANNO --- 【CAZY 注释结果统计】	4
---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】	7
---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】	8
---GENENUMS.BETWEENSAMPLES --- 【各样品中基因数目统计】	8
---GENENUMS.BETWEENSAMPLES.HEATMAP --- 【各样品注释到的基因数目的聚类热图, 若层级名称>70 个字符, 展示前70 字符】	9
---HEATMAP --- 【不同层级在各样品中的相对丰度聚类热图以及作图数据, 若层级名称>70 个字符, 展示前70 字符】	9
---HEATMAP_group --- 【不同层级在各组中的相对丰度聚类热图以及作图数据, 若层级名称>70 个字符, 展示前70 字符】	9
---METASTATS --- 【不同层级的 METASTATS 统计结果, EC 为酶, LEVEL1 为六大功能类, LEVEL2 为子功能】	9
-- EGGNOG --- 【EGGNOG 数据库分析结果】	9
---EGGNOG_ANNO --- 【EGGNOG 注释结果统计】	9
---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】	11
---GENENUMS --- 【各层级注释到的基因数目、基因 ID 统计】	12
---GENENUMS.BETWEENSAMPLES --- 【各样品中基因数目统计结果】	13
---GENENUMS.BETWEENSAMPLES.HEATMAP --- 【各样品注释到的基因数目的聚类热图, 若层级名称>70 个字符, 展示前70 字符】	14

---HEATMAP ---	【各样品中的相对丰度聚类热图以及作图数据，若层级名称>70 个字符，展示前70 字符】	14
---HEATMAP_group ---	【各组中的相对丰度聚类热图以及作图数据，若层级名称>70 个字符，展示前70 字符】	14
---NOG.TAX ---	【ORTHOLOG GROUP 物种归属分析结果】	14
---METASTATS ---	【各样品的 METASTATS 统计结果，LEVEL1 为第一层级，LEVEL2 为第二层级，OG 为直系同源簇】	14
`---PCA ---	【各样品的 PCA 分析结果】	14
`-- KEGG ---	【KEGG 数据库分析结果】	14
---GENENUMS ---	【注释到的基因数目、基因 ID 统计，EC 为酶，KO 为直系同源，LEVEL1 为六大代谢通路，LEVEL2 为子代谢通路，LEVEL3 为代谢通路图】	14
---GENENUMS.BETWEENSAMPLES ---	【各样品中基因数目统计】	15
---GENENUMS.BETWEENSAMPLES.HEATMAP ---	【各样品注释到的基因数目的聚类热图，若层级名称>70 个字符，展示前70 字符】	15
---HEATMAP ---	【各样品中的酶的相对丰度聚类图以及作图数据，若层级名称>70 个字符，展示前70 字符】	15
---HEATMAP_group ---	【各组中的酶的相对丰度聚类图以及作图数据，若层级名称>70 个字符，展示前70 字符】	15
---KEGG_ANNO ---	【KEGG 注释结果统计】	15
---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】	18
---METASTATS ---	【各样品的 METASTATS 统计结果】	19
---PATHWAYMAPS ---	【代谢通路比较结果】	19
`---PCA ---	【各样品的 PCA 分析结果】	19
`---05.FUNCTIONANNOTATION---README.PDF ---	【05.FUNCTIONANNOTATION 交付结果目录说明】	19

|-- 05.FunctionAnnotation ——【功能注释分析结果】

| |-- 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	CAZy_Class	六大功能类的缩写

2	Discription	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】

为基于 Unigenes.blast.m8.filter.xls 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 GI 号
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。

| | | |-- 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 不同水平的均一化后的绝对丰度矩阵】 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 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

Unigenes.relative.level1.xls 文件的展示形式是一样的，不同的是，Unigenes.relative.level2.xls 是对子功能类在各样品中的相对丰度进行的统计，而 Unigenes.relative.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。

| | |-- 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。

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

| | | -- heatmap —— 【不同层级在各样品中的相对丰度聚类热图以及作图数据】

| | | -- heatmap_group —— 【不同层级在各组中的相对丰度聚类热图以及作图数据】

| | | -- MetaStats —— 【不同层级的 MetaStats 统计结果，EC 为酶，level1 为六大功能类，level2 为子功能】

| | -- 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】

结果与 CAZy 数据库的结果类似。

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

结果与 CAZy 数据库的结果类似。

| | | |-- 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 Ortholog Group ID
6	Functional_Category	该 Ortholog Group ID 所属的功能类代号
7	OG_Description	该 Ortholog Group ID 所对应的描述

| | | |-- 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} —— 【eggNOG 第一层级上的相对丰度统计图】

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

| | |-- 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 —— 【各样品在 level1，level2 和 og 水平注释到的均一化后的绝对丰度矩阵】 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 Unigenes.absolute.level1.xls 文件的展示形式是一样的，不同的是，Unigenes.absolute.level2.xls 是对子功能类的描述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 Ortholog Group ID，在各样品中的绝对丰度进行的统计。

| | | `-- Relative ——【各样品在 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 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.og.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 是对子功能类的描述在各样品中的相对丰度进行的统计，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 Ortholog Group ID，在各样品中的相对丰度进行的统计。

| | | |-- GeneNums ——【各层级注释到的基因数目、基因 id 统计】

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，第一列为子功能描述，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 Ortholog Group ID，第二列为注释上该 Ortholog Group ID 的基因数，第三列为注释上该 Ortholog Group ID 的所有基因 id。

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

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。



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

| | | -- heatmap —— 【各样品中的相对丰度聚类热图以及作图数据】

| | | -- heatmap_group —— 【各组中的相对丰度聚类热图以及作图数据】

| | | -- NOG.Tax —— 【 ortholog group 物种归属分析结果】

| | | -- MetaStats —— 【各样品的 MetaStats 统计结果，level1 为第一层级，level2 为第二层级，og 为直系同源簇】

| | `-- PCA —— 【各样品的 PCA 分析结果】

| `-- 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，level1 为六大代谢通路，level2 为子代谢通路，level3 为代谢通路图】

在该文件夹中，一共有五个文件，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

代谢通路的名称，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。Unigenes.absolute.level2.xls、Unigenes.absolute.level3.xls、Unigenes.absolute.ko.xls 和 Unigenes.absolute.ec.xls 也是三列，第一列分别为 level2、level3、Orthologous groups 的 id 和酶的名称，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】

在该文件夹中有三个文件，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 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。

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

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

| |-- heatmap_group —— 【各组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	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】

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

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

| |-- 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】

| | |-- Absolute —— 【各样品在 ec，ko,level1，level2 和 level3 均一化后的绝对丰度矩阵】 在该文件夹中，一共有五

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，存在的另外三个文件 Unigenes.absolute.level2.xls、Unigenes.absolute.level3.xls、Unigenes.absolute.ko.xls 和 Unigenes.absolute.ec.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，在各样品中的绝对丰度进行的统计。

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

在该文件夹中，一共有五个文件，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 Relative 文件夹中，存在的另外三个文件 Unigenes.relative.level2.xls、Unigenes.relative.level3.xls、Unigenes.relative.ko.xls 和 Unigenes.relative.ec.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，而 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。

`-- MetaStats` —— **【各样品的 MetaStats 统计结果】**

`-- pathwaymaps` —— **【代谢通路比较结果】**

`-- PCA` —— **【各样品的 PCA 分析结果】**

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