**新基因注释结果**

使用BLAST软件将发掘的新基因与NR、Swiss-Prot、GO、COG、KEGG数据库进行序列比对，获得新基因的注释信息。

**====================================文件======================================**

**|-- All\_Database\_annotation.xls** **新基因整合各个数据库注释结果**

内有多个sheet工作表，分别是新基因与各个数据库比对后的结果

**|-- Function\_Annotation.stat.xls** **新基因各个数据库的统计结果**

|-- Anno\_Database 注释的数据库

|-- Annotated\_Number 注释到该数据库的新基因的数目

|-- 300<=length<1000 注释到该数据库的长度范围为300-1000的新基因数目

|-- length>=1000 注释到该数据库的长度范围为1000以上的新基因数目

**|-- Integrated\_Function.annotation.xls 新基因在各个数据库里的注释结果整合**

|-- ID 新基因ID

|-- 之后各列分别为注释到各个数据库的信息，若没有则用”--”代替

**===================================文件夹１====================================**

**|-- BMK\_1\_Annotation** **NewGene功能注释分析结果目录**

**|-- \*.Cog/Kog\_class.xls** **COG/KOG数据库注释结果文件**

|-- Gene name：基因名称；

|-- Portein name in COG：COG中注释的相应蛋白质的名字；

|-- E\_value：E值；

|-- Identify：比对上的碱基数占总比对碱基长度的百分比；

|-- Score：打分值；

|-- Organism：物种名；

|-- COG(KOG) id：COG (KOG)编号；

|-- COG(KOG) class defination：COG(KOG)分类名称；

|-- Function code：COG(KOG)功能分类编号；

|-- Functional categories：COG(KOG)功能类别名称；

|-- Function class defination：COG(KOG)功能类别定义。

**|-- \*.eggNOG\_class.xls** **eggNOG功能注释文件**

|-- Query：基因名称；

|-- Match：基因比对上的 eggNOG 数据库中的蛋白序列；

|-- eggNOG：eggNOG 数据库的ID编号；

|-- score：基因和序列比对的打分；

|-- Functional Category：eggNOG数据库对应的功能分类编号；

|-- Description：比对上的序列功能描述；

|-- Function class defination：eggNOG数据库分类类别的功能描述。

**|-- \*.GO.anno.xls** **各基因注释到GO数据库的信息统计**

|-- Gene：基因名称；

|-- Number：某基因注释到的GO类别数目；

|-- GO\_Anno：注释到的GO类别。

**|-- \*.GO.list.xls** **每个基因注释到GO数据库上的编号**

|-- 基因ID；

|-- 相应基因注释到的GO编号。

**|-- \*.Kegg.ko.xls** **每个基因的kegg注释信息**

|-- Gene\_id：基因编号；

|-- KO|e\_value|Database\_Genes|Anno：四列值,分别表示基因的注释到的K编号，blastx比对上基因 E 值,比对上的基因ID以及它的功能注释。

**|-- \*.Kegg.pathway.xls** **KEGG上注释到的基因信息**

|-- pathway：pathway名称；

|-- pathway\_id：pathway编号(一个 ko编号表示一个通路)；

|-- Gene\_number：注释到某pathway上的基因数目；

|-- Gene\_id：注释到某pathway上的基因名称；

|-- KOs：KO 分类信息(一个代谢通路上有哪些KO分类,它是蛋白质(酶)的一个分类体系,序列高度相似，并且在同一条通路上有相似功能的蛋白质被归为一组，然后打上K标签)。

**|-- \*.Pfam.anno.xls**  **所有的基因在Pfam数据库中的注释结果**

|-- Gene\_ID：基因名称；

|-- Pfam\_IDs ：比对上的Pfam数据库的编号；

|-- Pfam\_Description：比对上的Pfam数据库的描述。

**|-- \*.NR/Swissprot.anno.xls** **基因在NR/swissprot上的注释结果**

|-- SwissprotGeneID(NrGeneID)：基因名称；

|-- Database\_ID：注释到数据库的基因编号；

|-- E\_value：E值；

|-- IdeNTity：比对上的碱基数占总比对碱基长度的百分比；

|-- Score：打分值；

|-- Annotation：功能注释。

**===================================文件夹２====================================**

**|-- BMK\_2\_Statistic 统计文件+统计图**

**|-- \*.eggNOG.cluster.png** **eggNOG注释分类统计图**

横坐标为eggNOG各分类内容，纵坐标为基因数目。在不同的功能类中，基因所占多少反映对应时期和环境下代谢或者生理偏向等内容，可以结合研究对象在各个功能类的分布做出科学的解释。

**|-- \*.eggNOG.cluster.stat.xls** **eggNOG注释分类统计表**

|-- #ID class ID

|-- Class\_Name class Name

|-- Numbers 基因数目

**|-- \*.GO.png** **GO二级节点注释统计图（png格式）**

所有基因GO 分类统计结果图,此图展示的是在全部基因背景下GO各二级功能的基因比例和个数。横坐标：GO分类；纵坐标：左边为基因数目所占百分比，右边为基因数目。

**|-- \*.nr.lib.png NR注释中比对上不同物种分布统计图**

展示了注释到某物种上的基因数目以及占总检测到的基因的比例。

**|-- \*.nr.lib.stat.xls NR注释中比对上不同物种分布统计表**

|-- #Species\_Name 物种名，只展示最多的前9个物种+Other；

|-- Homologous\_Number 数目；

|-- Ratio 比例。

**===================================文件夹３====================================**

**|-- BMK\_3\_KEGG\_map KEGG通路图**

**|-- ko\*.png**  **KEGG通路注释图**