

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 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.txt 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