

=====文件=====

1. All_Database_annotation.xls	所有新基因功能注释汇总表(6 个 sheets)
(1). Integrated_Function.anno:	所有基因的注释信息
注: #GeneID 基因名称	
COG_class	COG 库注释的功能编码
COG_class_annotation	COG 库功能注释
GO_annotation	GO 注释
KEGG_annotation	KEGG 库注释
KOG_class	KOG 库注释的功能编码
KOG_class_annotation	KOG 库注释
Swiss-Prot_annotation	Swissprot 库注释
TrEMBL_annotation	TrEMBL 库注释
eggNOG_class	eggNOG 分类编码
eggNOG_class_annotation	eggNOG 分类注释
NR_annotation	nr 库注释
NT_annotation	nt 库注释
Pfam_annotation	Pfam 库注释
(2). Function_anno.stat:注释结果统计	
注: #Anno_Database	注释用的数据库
Annotated_Number	注释到的基因数目
300<=length<1000	注释到的基因长度在 300 到 1000 的数目
length>=1000	注释到的基因长度大于等于 1000 的基因数目
(3). GO.list:	每个基因注释到 GO 数据库上的编号
(4). GO_tree.stat:	对于各个 GO 功能分类,注释到的基因统计
注: #GO_Function	GO 功能分类
Unigene_number	注释到的基因数目
Unigene_ID	注释到的基因名称
(5). KEGG.pathway:KEGG 上注释到的基因信息	
注: #pathway pathway	名称
pathway_id	pathway ko 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 KO(或 K)标签)
(6). KEGG.ko :	每个基因的 KEGG 注释信息
注: #Gene_id	基因编号
KO e_value Database_Genes Anno	四列值,分别表示基因的 KO 编号,E 值,数据库中注释的基因以及功能注释
2.*.newGene.longest_transcript.fa.COG_class.txt	新基因在 COG 上注释的详细信息
注: #Gene name	基因名称
Protein name in COG	COG 中注释的相应蛋白质的名字
E_value	E 值
Identity	比对上的碱基数占总比对碱基长度百分比

	Score	打分值
	Organism	物种名
	COG id	COG 的功能编号
	COG class defination	COG 对功能的描述
	Function code	COG 功能分类编号
	Functional categories	COG 对功能类别的描述(如果同时注释到两个功能类别,此处为--)
	Function class defination	COG 对功能类别的定义(如果同时注释到两个功能类别,此处为--)
3. *.newGene.longest_transcript.fa.GO.anno.txt		各基因注释到 GO 数据库的信息统计
注: #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
4. *.newGene.longest_transcript.fa.KOG_class.txt		新基因在 Kog 上注释的详细信息
注: #Gene name		基因名称
	Protein name in KOG	KOG 中注释的相应蛋白质的名字
	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
	Organism	物种名
	KOG id	KOG 的功能编号
	KOG class defination	KOG 对功能的描述
	Function code	KOG 功能分类编号
	Functional categories	KOG 对功能类别的描述(如果同时注释到两个功能类别,此处为--)
	Function class defination	KOG 对功能类别的定义(如果同时注释到两个功能类别,此处为--)
5. *.newGene.longest_transcript.fa.NR.anno.txt		基因的 nr 注释结果
注: #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
6. *.newGene.longest_transcript.fa.NT.anno.txt		新基因在 nt 数据库中的注释结果
注: #NtGeneID		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
7. *.newGene.longest_transcript.fa.Swiss-Prot.anno.txt		基因在 Swiss-Prot 上的注释结果
注: #SwissprotGeneID		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	比 打分值
	Annotation	功能注释
8. *.newGene.longest_transcript.fa.Pfam.anno.txt		新基因在 Pfam 数据库中的注释结果
注: #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
9. *.newGene.longest_transcript.fa.Pfam.anno.details		新基因注释到 Pfam 数据库的结果统计
注: # target name		注释到的基因名
accession		注释到的基因在 Pfam 数据库中的检索号
query name		检索基因名(新基因)
accession		查询序列的检索号,如果没有用“-”表示序列匹配的 E 值
E-value(全长序列的比对结果)		序列匹配的 E 值
score(全长序列的比对结果)		序列匹配的打分
bias(全长序列的比对结果)		基于 null2 模型的偏差校正(高值可能意味着假阳性结果)
E-value(比对上的最好的一个 domain		的结果) 序列匹配的 E 值,如果这个值比较高,而整个序列的比对 E 值较低,说明整个结果可能是一个假阳性,较低的 E 值是由多个不够好的 hits 组合形成的,或者序列中含有重复序列。
score		序列匹配的打分
bias		null2 偏差校正
exp(domain 的数目评估)		domains 数据的期望值
reg(domain 的数目评估)		彼此独立的区域(discrete regions)的数目。一般来说和 exp 很接近,如果两者相差很大,说明这些区域互不独立。
clu(domain 的数目评估)		可能有多个 domain 的区域,通常是 0
ov(domain 的数目评估)		在由随机回溯分类中定义类别和其他的类有重合的数目
env(domain 的数目评估)		定义的所有类的数目,包括单独的类区域或者由随机回溯聚类成的一个区域中的多个类。
dom(domain 的数目评估)		定义的 domain 的数目。通常和 env 的数目一致
rep(domain 的数目评估)		满足输出阈值的 domain 的数目
inc(domain 的数目评估)		满足内置阈值的 domain 的数目
description of target		比对上的基因的描述
10. *.newGene.longest_transcript.fa.TrEMBL.anno.txt		新基因在 TrEMBL 上的注释结果
注: #TrEMBLGeneID		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

11. *.newGene.longest_transcript.fa.eggNOG_class.txt	新基因在 eggNOG 上注释的详细信息
注: #Query	基因名称
Match	匹配上的蛋白质的名称
NOG	eggNOG 库中的名称
hsp_score	比对打分值 (High-scoring Segment Pair)
Functional Category	eggNOG 的分类编号
Description	蛋白功能描述
Function class defination	功能分类的描述

注: 结果文件中没有的表或列表示没有对该数据库做注释

=====文件夹=====

1. BMK_1_PNG 新基因注释结果图及相应的统计表
2. BMK_2_KEGG_map 新基因所在的所有通路图