[KEGG PATHWAY Database][https://www.genome.jp/kegg/pathway.html]

[Pathway Maps][https://www.genome.jp/kegg/pathway.html]

KEGG PATHWAY是人工审核的代表分子间相互作用, 反应和相互关系的网络图:

- Metabolism: [Global/overview](#) [Carbohydrate](#) [Energy](#) [Lipid](#) [Nucleotide](#) [Amino acid](#) [Other amino](#) [Glycan](#) [Cofactor/vitamin](#) [Terpenoid/PK](#) [Other secondary metabolite](#) [Xenobiotics](#) [Chemical structure](#)
- Genetic Information Processing
- Environmental Information Processing
- Celluar Processes
- Organismal Systems
- Human Diseases
- Drug Developement

KEGG PATHWAY是[KEGG Mapper][https://www.genome.jp/kegg/mapper.html] 用于pathway mapping的参考数据库; 其中各个PATHWAY又细分为多个项

Pathway Identifiers

Each pathway map is identified by the combination of 2-4 letter prefix code and 5 digit number (see KEGG Identifier). The prefix has the following meaning:

```
map     manually drawn reference pathway
ko      reference pathway highlighting KOs
ec      reference metabolic pathway highlighting EC numbers
rn      reference metabolic pathway highlighting reactions
<org>   organism-specific pathway generated by converting KOs to gene identifiers
```

and the numbers starting with the following:

```
011     global map (lines linked to KOs)
012     overview map (lines linked to KOs)
010     chemical structure map (no KO expansion)
07      drug structure map (no KO expansion)
other   regular map (boxes linked to KOs)
```

are used for different types of maps.

[Functional Hierarchies][https://www.genome.jp/kegg/brite.html]

KEGG BRITE是一个等级分类系统, 包含多种生物对象(biological objects)的功能性等级分类. 这些生物对象生物实体, 从分子到更高水平的被KEGG所收录的数据, 例如基因和蛋白, 小分子, 反应, 通路, 疾病和药物.

KEGG BRITE集合了多种不同类型的关系:

[1. Genes and Proteins](#) [2. Compounds and Reactions](#) [3. Drugs](#) [4. Diseases](#) [5. Organisms and Cells](#)

KEGG BRITE是[KEGG Mapper][https://www.genome.jp/kegg/mapper.html]用于brite mapping的参考数据库

Brite Identifiers

Each brite hierarchy/table is identified by the combination of 2-4 letter prefix code and 5 digit number (see KEGG Identifier). The prefix has the following meaning:

br   non-gene/protein reference hierarchy or table
jp   non-gene/protein reference hierarchy or table in Japanese
ko   gene/protein reference hiearchy for KO expansion
<org> organism-specific hierarchy generated by converting KOs to gene identifiers

[KEGG MODULE Database][https://www.genome.jp/kegg/module.html]

KEGG MODULE database目前正改变为主要针对于代谢通路.

1. 非代谢通路的模型和分子复合体都被去除. 相同数据可以在KEGG BRITE数据中得到, 同时KEGG Mapper的Reconstruct Pathway tool可用于完整性检测.
2. 通过RM numbers识别的Reaction modules被合并到MODULE flat文件中, 供DBGET搜索系统使用. 在html文件中仍可查询到反应模型的实际内容.
3. MODULE flat文件不再包含物种特异性的完整模型. 所有模型, 完整或不完整的, 和物种特异性的, 仍然可在module map界面查询到

KEGG MODULE数据目前由KEGG modules, 由M number代表, 和KEGG reaction modules, 由RM number代表, 所组成. 它们是人工审核定义的基因集和反应集的功能单元. KEGG modules进一步划分为pathway modules和signature modules.

- pathway modules: 基因集, 包括分子复合体, 在代谢通路中的功能单元
- signature modules: 表现出表型特征的基因集的功能单元
- reaction modules: 代谢通路中连续性反应步骤的功能单元

完整的KEGG modules和KEGG reaction modules可在BRITE hierarchy 文件中查看到:

KEGG modules/KEGG reaction modules

[KO][https://www.kegg.jp/kegg/ko.html]

KO(KEGG Orthology)数据库是一个分子功能数据库, 以functional othologs来表示. functional ortholog是人工定义的KEGG 分子网络内容, 也就是KEGG pathway maps, BRITE hierarchies和 KEGG modules. 例如, 当一个通路图被绘制, 每个box为给定的一个KO identifier(K number), 同时特异性的物种的实验验证的基因和蛋白被用来在其他物种中搜索orthologs.

KO system是一个基于网络的KOs分类, 包含6个KEGG pathway maps大类, 一个BRITE hierarchies大类和其他没包含在以上两大类的一个类别.

KO Assignment and KEGG Mapping: KEGG中的基因组注释包含两个独立的方向, ortholog注释和network重构.

Otholog annotaion(KO assignment): 储存在KO(KEGG Orthology)数据库中的分子功能包含了实验验证了的基因/蛋白的同系物; KEGG中的基因组的注释为关联基因组中基因到KO identifier(KO number).

Network reconstruction(KEGG mapping): Functional othologs定义为KEGG pathway maps和其他分子网络, 均通过K number节点构建成网络; 通过KEGG mapping, 基因组注释过程将基因组中的一个基因集转换为一个K number集, 用于自动重建KEGG pathways和其他网络, 使之能够解释高水平功能(不分物种搜索).

[KEGG Mapper][https://www.genome.jp/kegg/mapper.html]

KEGG Mapper是用于KEGG mapping的工具集, 包含最流行的KEGG pathway mapping, 还有BRITE mapping和MODULE mapping. 'Reconstruct Pathway', 'Search Pathway'和'Search&Color Patway'允许同时进行多重比多操作. 'Color Pathway'和'Join Brite'用于比对单个pathway map或单个birte hierarchy/table 文件.

General mapping tools against multiple databases

[Reconstruct Pathway][https://www.genome.jp/kegg/tool/map_pathway.html]为基本的比对工具, 支持基因组和metagenome注释. 用于处理KO annotation(K number assignment)数据, 可在内部针对 KEGG GENES, 外部使用[BlastKOALA][https://www.kegg.jp/blastkoala/]服务器, 以及其他注释服务器. **输入数据为针对单个物种的单个基因列表或者针对多重物种的多个基因列表; 每个基因列表格式 为** `gene_id K number` **; 通过K numbers将其进行pathway maps/brite hierarchies/brite tables/modules比对.**

[Search Pathway](#) 和 [Search&Color Pathway](#) 为KEGG 项目开始使用的传统工具, 但是其数据库已经得到扩展. 当对象同时存在参考数据和物种特异性通路中时, 这些工具方实现匹配. **针对给定的对象(genes, proteins, compouds, glycans, reactions, drugs, etc), 根据搜索模式搜索KEGG pathway maps/Brite hierarchies/Brite tables/KEGG modules/KEGG network variation maps/KEGG disease entries; 输入为空格或换行隔开的KEGG identifiers.**

Tool	Search mode	Target database	Query data (KEGG identifier)
Reconstruct Pathway	Reference	Pathway Brite hierarchy/table Module	K number
Search Pathway	Reference	Pathway Brite hierarchy/table Module	K/R/EC number C/G/D/E number
	Human-specific	Pathway Brite hierarchy Module Network Disease	Human gene identifier C/G/D number
	Other organism-specific	Pathway Brite hierarchy Module	Gene identifier K/R/RC/EC number C/G/D number
Search&Color Pathway	Reference	Pathway Brite hierarchy Module	K/R/EC number C/G/D/E number
	Organism-specific	Pathway Brite hierarchy Module	Gene identifier K/R/RC/EC number C/G/D number Outside gene identifier

针对不同的工具, 其搜素数据需要经过不同处理

Tool	Data format	Remark
Reconstruct Pathway	K number in the second column of each line The first column may be used for user's gene ID	Same as BlastKOALA output
Search Pathway	Free format - KEGG identifiers separated by whitespace characters (space, tab, newline)	Filter option available
Search&Color Pathway	KEGG identifier in the first column of each line Optional color specification in the second column (background color and forground color)	

[KEGG Annotation][https://www.genome.jp/kegg/annotation/]

KEGG database包含三个主要组成用于genome/metagenome注释:

- 完整基因组(KEGG organisms)的注释基因分类
- KEGG PATHWAY, BRITE和MODULE中的, 已知的高水平的功能关系, 包含分子相互作用, 反应和关系网络
- KO database中已知的分子水平相关的ortholog groups, 大部分KO entries定义为context-dependent manner, 作为KEGG 分子网络的节点

一般而言, KO entries(K numbers)以代表了序列相似性groups. 因此, KEGG GENES中 query genome的序列相似性搜索就在在搜索最恰当的K numbers, 并且得到的K numbers集合可用于重建 KEGG pathway maps, BRITE hierarchies和KEGG modules, 用于即使高水平功能相互作用 ([BlastKOALA][https://www.kegg.jp/blastkoala/]/[GhostKOALA][https://www.kegg.jp/ghostkoala/])

Otholog Table

OT(orthology table)自KEGG项目初就一直存在. 针对给定的一套K numbers, 用于展示基因在KEGG organisms中的位置:

Enter K numbers   (Example) K14579 K14580 K14578 K14581 K14582 K14583 K18242 K18243

[Go] [Filter] [Clear]

Module Table

MT(module table)是KEGG新的注释资源. 针对给定的一组M &/ K numbers, 识别包含完整modules &/ KO groups的完整模型:

Enter M/K numbers   (Example) M00595 K16952 M00596

[Go] [Filter] [Clear]

非模式生物KEGG pathway分析

1. 基因组蛋白注释

- 使用[KAAS][https://www.genome.jp/tools/kaas/](kegg automatic annotation server)对基因进行功能性注释. 通过BLAST/GHOST对序列进行人工审核的KEGG GENEs数据库注释. 结果包含 KO(KEGG Orthology)和自动生成的KEGG pathways.

Request

Complete or Draft Genome: 针对基因组中的完整的一整套基因而言, **KASS**注释效果最好. 准备氨基酸序列并使用**BBH(bi-directional best hit)**搜索:**KAAS job request (BBH method)**

Partial Genome: KASS同样可用于有限数目的基因搜索. 准备氨基酸序列并使用**SBH(single-directional best hit)**搜索同源序列: **KAAS job request (SBH method)**

Metagenomes: 当搜索序列包含来自混合物种的大量序列时, 例如来自**metagenome**测序数据, 推荐使用**GHOSTX**搜索和**SBH**搜索: **KAAS job request (SBH method for amino acid sequence query)**

- 使用[BlastKOALA][https://www.kegg.jp/blastkoala/](automatic annotaion and KEGG mapping service)自动注释和比对. KOALA(KEGG Orthology And Links Annotation)为使用SSEARCH计算用于注释K number的KEGG内部注释工具. BlastKOALA和GhostKOALA都可使用一套非冗杂的KEGG GENES, 针对提交序列分别进行BLAST和GHOSTX搜索. KofamKOALA是使用HMM profile搜索的注释工具.

BlastKOALA

Query Data Input

Automatic annotation and KEGG mapping service

BlastKOALA	**GhostKOALA**	**KofamKOALA**

KOALA job status 2020/02/18 22:47:22 (GMT+9)

	Blast	Ghost	Kofam
Number of jobs in the queue	2	0	0
Submission of last completed job	2020/02/18 22:33:21	2020/02/18 22:39:00	2020/02/18 22:17:47

其中**BlastKOALA**接受较小的查询序列文件, 适用于注释高质量基因组; **GhostKOALA**接受较大的数据集, 适用于指示**metagenomes**; **KofamKOALA**使用**KOfam(a customized HMM database of KEGG Orthologs(KOs)), 可设置阈值进行过滤.**

完成注释后, 获得蛋白名称所对应的K numbers; 根据分析所需, 例如将蛋白名称更换为locus_tag信息:

```
1 IX87_RS00300    K01934
2 IX87_RS00310    K21449
3 IX87_RS00325    K01338
4 IX87_RS00330    K00783
5 IX87_RS00335    K15777
```

2. 通路注释

- 提取其中的k nubmers, 使用[KO(KEGG ORTHOLOGY) Database][https://www.kegg.jp/kegg/ko.html]对K numbers进行KEGG pathway/BRITE/MODULE mapping(不限制物种).

KO (KEGG ORTHOLOGY) Database

Linking genomes to pathways by ortholog annotation

| Menu | PATHWAY | BRITE | MODULE | KO | Annotation | ENZYME | RModule | BlastKOALA |

Search [KO ⌄] for [_____] [Go]

The following interface allows some of the KEGG mapping functions (see also KEGG Annotation).

Enter K numbers   (Example) K00161 K00162 K00163 K00627 K00382

[_____]

[Filter] [Ortholog table] [Map pathway] [Map brite] [Map module] [Get title] [Get entry] [Clear]
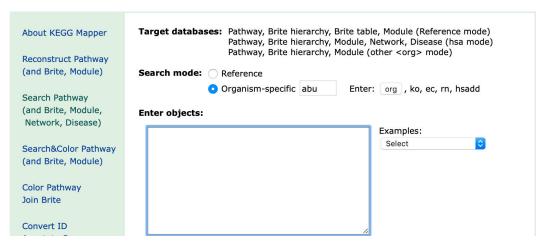
选择Map pathway, 完成注释后跳转至结果页面:

KEGG Mapper Search Result

| Pathway (186) | Brite (35) | Module (131) |

Sort by the pathway list

Show matched objects

ko01100 Metabolic pathways (280)

ko01110 Biosynthesis of secondary metabolites (121)

ko01120 Microbial metabolism in diverse environments (99)

ko01130 Biosynthesis of antibiotics (82)

- 或使用其中的k nubmers, 使用KEGG Mapper Search Pathway注释(可选物种, 选择reference时结果接近KO map?):

KEGG Mapper – Search Pathway

About KEGG Mapper

Reconstruct Pathway
(and Brite, Module)

Search Pathway
(and Brite, Module,
 Network, Disease)

Search&Color Pathway
(and Brite, Module)

Color Pathway
Join Brite

Convert ID

Target databases: Pathway, Brite hierarchy, Brite table, Module (Reference mode)
Pathway, Brite hierarchy, Module, Network, Disease (hsa mode)
Pathway, Brite hierarchy, Module (other <org> mode)

Search mode: ○ Reference
　　　　　　　　 ● Organism-specific [abu]   Enter: [org] , ko, ec, rn, hsadd

Enter objects:

[_____]   Examples:
                              [Select ⌄]

选择对应物种缩写, 完成注释后跳转至结果页面:

KEGG Mapper Search Result

| Pathway (90) | Brite (0) | Module (114) |

Sort by the pathway list
Show matched objects

abu01100 Metabolic pathways - Arcobacter butzleri RM4018 (317)

abu01110 Biosynthesis of secondary metabolites - Arcobacter butzleri RM4018 (169)

abu01130 Biosynthesis of antibiotics - Arcobacter butzleri RM4018 (112)

abu01230 Biosynthesis of amino acids - Arcobacter butzleri RM4018 (88)

3. 下载对应网页内容使用RCurl或浏览器resource下载并解析

这里选择的是BlastKOALA注释得到的K numbers, 然后使用KEGG Mapper Search Pathway比对, 选择物种abau(鲍曼):

Pathway(abau):

KEGG Mapper Search Result

| Pathway (116) | Brite (0) | Module (167) |

Sort by the pathway list
Show matched objects

abau01100 Metabolic pathways (676)

abau01110 Biosynthesis of secondary metabolites (274)

abau01120 Microbial metabolism in diverse environments (218)

abau01130 Biosynthesis of antibiotics (193)

Pathway(reference):

KEGG Mapper Search Result

| Pathway (212) | Brite (38) | Brite Table(7) | Module (169) |

Sort by the pathway list
Show matched objects

map01100 Metabolic pathways (580)

map01110 Biosynthesis of secondary metabolites (232)

map01120 Microbial metabolism in diverse environments (184)

map01130 Biosynthesis of antibiotics (153)

KO:

KEGG Mapper Search Result

| Pathway (212) | Brite (38) | Module (169) |

Sort by the pathway list
Show matched objects

ko01100 Metabolic pathways (580)

ko01110 Biosynthesis of secondary metabolites (232)

ko01120 Microbial metabolism in diverse environments (184)

ko01130 Biosynthesis of antibiotics (153)

ko01230 Biosynthesis of amino acids (96)

写R脚本解析获得:

```
1 ko      ko_ann  k       k_ann
2 abu0110  Metabolic pathways - Arcobacter butzleri RM4018        K00003  hom; homoserine dehydrogenase [EC:1.1.1.3]
3 abu0110  Metabolic pathways - Arcobacter butzleri RM4018        K00012  UGDH; UDPglucose 6-dehydrogenase [EC:1.1.1.22]
4 abu0110  Metabolic pathways - Arcobacter butzleri RM4018        K00013  hisD; histidinol dehydrogenase [EC:1.1.1.23]
5 abu0110  Metabolic pathways - Arcobacter butzleri RM4018        K00014  aroE; shikimate dehydrogenase [EC:1.1.1.25]
6 abu0110  Metabolic pathways - Arcobacter butzleri RM4018        K00024  mdh; malate dehydrogenase [EC:1.1.1.37]
7 abu0110  Metabolic pathways - Arcobacter butzleri RM4018        K00031  IDH1; isocitrate dehydrogenase [EC:1.1.1.42]
8 abu0110  Metabolic pathways - Arcobacter butzleri RM4018        K00052  leuB; 3-isopropylmalate dehydrogenase [EC:1.1.1.85]
9 abu0110  Metabolic pathways - Arcobacter butzleri RM4018        K00053  ilvC; ketol-acid reductoisomerase [EC:1.1.1.86]
10 abu0110  Metabolic pathways - Arcobacter butzleri RM4018       K00058  serA; D-3-phosphoglycerate dehydrogenase / 2-oxoglutarate reductase [EC:1.1.1.95 1.1.1.399]
```

Module:

KEGG Mapper Search Result

Pathway (90)	Brite (0)	Module (114)

Sort by the module list

Show matched objects

　abu_M00144 NADH:quinone oxidoreductase, prokaryotes (14)

　abu_M00026 Histidine biosynthesis, PRPP => histidine (12)

　abu_M00121 Heme biosynthesis, plants and bacteria, glutamate => heme (10)
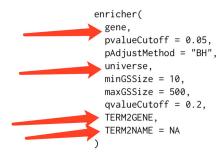
解析可得:

```
1 ko       ko_ann   k       k_ann
2 abu_M00144        NADH:quinone oxidoreductase, prokaryotes        K00330  nuoA; NADH-quinone oxidoreductase subunit A [EC:7.1.1.2]
3 abu_M00144        NADH:quinone oxidoreductase, prokaryotes        K00331  nuoB; NADH-quinone oxidoreductase subunit B [EC:7.1.1.2]
4 abu_M00144        NADH:quinone oxidoreductase, prokaryotes        K00334  nuoE; NADH-quinone oxidoreductase subunit E [EC:7.1.1.2]
5 abu_M00144        NADH:quinone oxidoreductase, prokaryotes        K00335  nuoF; NADH-quinone oxidoreductase subunit F [EC:7.1.1.2]
6 abu_M00144        NADH:quinone oxidoreductase, prokaryotes        K00336  nuoG; NADH-quinone oxidoreductase subunit G [EC:7.1.1.2]
7 abu_M00144        NADH:quinone oxidoreductase, prokaryotes        K00337  nuoH; NADH-quinone oxidoreductase subunit H [EC:7.1.1.2]
8 abu_M00144        NADH:quinone oxidoreductase, prokaryotes        K00338  nuoI; NADH-quinone oxidoreductase subunit I [EC:7.1.1.2]
```

4. 调用clusterProfiler包, 使用enricher函数富集

[enricher]

[https://bioconductor.org/packages/release/bioc/manuals/clusterProfiler/man/clusterProfiler.pdf]
是一个通用的富集分析工具, 分析需要准备差异基因(gene), gene和注释term(ko/M)之间的对应关系文件
(TERM2GENE), 还可以提供可选的背景基因(universe)和term对应的名称文件(TERM2NAME).

```
enricher(
    gene,
    pvalueCutoff = 0.05,
    pAdjustMethod = "BH",
    universe,
    minGSSize = 10,
    maxGSSize = 500,
    qvalueCutoff = 0.2,
    TERM2GENE,
    TERM2NAME = NA
)
```

5. 绘图

略!