



# KEGG 数据库信息挖掘手册 V2.0



诺禾致源微生物事业部

2017-11-01

# 目录

<b>1.KEGG 数据库概述</b>	1
1.1 KEGG 数据库分类	1
1.2 KEGG identifiers	2
1.3 KEGG ORTHOLOGY 数据库	3
1.4 KEGG Pathway 数据库结构	5
<b>2. 宏基因组 KEGG 数据库注释过程及结果展示</b>	6
2.1 宏基因组 KEGG 数据库标准注释过程	6
2.2 KEGG Pathway 数据库注释结果展示	7
<b>3. 宏基因组 KEGG 数据库注释常见问题</b>	12
3.1 KEGG 注释结果中哪个文件将物种与功能联系起来？	12
3.2 KEGG 代谢通路中，各元素代表的含义	13
3.3 有时候，图中方框变成绿色或蓝色代表什么？	14
3.4 如何对代谢通路图中某个感兴趣的酶进行标记？	16
3.5 global map , overview map 及 regular map ?	18
<b>4.参考文献</b>	19

**KEGG ( Kyoto Encyclopedia of Genes and Genomes ) 数据库为基因及基因组的“百科全书”，是一个大的综合性数据库。其主要目的是确定基因或基因组在分子水平或更高水平的功能。**

## **1. KEGG 数据库概述**

### **1.1 KEGG 数据库分类**

KEGG数据库包括四大类，16个数据库，详细数据库分类及说明如**Table 1**所示。其中，KEGG ORTHOLOGY数据库及KEGG PATHWAY数据库为KEGG最重要的两部分。另外，自2016年10月起，数据库中的KEGG RPAIR 数据库已停止使用。

**Table 1** KEGG数据库

Category & Database	Content
<b>Systems information</b>	
KEGG PATHWAY	KEGG pathway maps
KEGG BRITE	BRITE hierarchies and tables
KEGG MODULE	KEGG modules
<b>Genomic information</b>	
KEGG ORTHOLOGY(KO)	KO groups for functional orthologs
KEGG GENOME	KEGG organisms (complete genomes) and selected viruses
KEGG GENES	Gene catalogs of KEGG organisms, viruses, and addendum category
KEGG SSDB	Sequence similarity among GENES entries (computationally generated)
<b>Chemical Information</b>	
KEGG COMPOUND	Metabolites and other small molecules
KEGG GLYCAN	Glycans
KEGG REACTION	Biochemical reactions
KEGG RCLASS	Reaction class
KEGG ENZYME	Enzyme nomenclature
<b>Health Information</b>	
KEGG DISEASE	Human diseases

KEGG DRUG	Drugs
KEGG DGROUP	Drug groups
KEGG ENVIRON	Crude drugs and health-related substances

## 1.2 KEGG identifiers

KEGG object 是分子到更高水平的生物实体，反映在数据库中则被称为一个记录（entry），KEGG identifier 是每个 object 的唯一标识符和身份证，同时也是数据库每个 entry 的识别符。KEGG identifier 一般由相应数据库的前缀（database-dependent prefix）加上一个 5 位数的数字（a five-digit number）组成，各数据库 KEGG identifier 的组成如 **Table 2** 所示。

**Table 2** KEGG identifier

Database	Object	Prefix	Example
KEGG PATHWAY	Pathway map	map/ko/ec/rn/ (org)	map04930
KEGG BRITE	Functional hierarchy	br/ jp/ ko/ (org)	ko01003
KEGG MODULE	KEGG module	M/(org)_M	M00010
KEGG ORTHOLOGY	KO group	K	K04527
KEGG GENOME	KEGG organism	org code / T number	hsa/T01001
KEGG GENES	Gene / protein	org:gene/ag:protein	hsa:3643
KEGG COMPOUND	Small molecule	C	C00031
KEGG GLYCAN	Glycan	G	G00109
KEGG REACTION	Reaction	R	R00259
KEGG RCLASS	Reaction class	RC	RC00046
KEGG ENZYME	Enzyme		ec:2.7.10.1
KEGG DISEASE	Human disease	H	H00004
KEGG DRUG	Drug	D	D01441
KEGG DGROUP	Drug group	DG	DG00710
KEGG ENVIRON	Crude drug, etc.	E	E00048

从表中可以看到某些标识码由 org:gene 组成，其中 org 由三/四个字母组


成, gene 为 GeneID , 两者共同组成 KEGG GENES 数据库生物分类的标识符 ( identifier ) , 例如 hsa 代表 *Homo sapiens* ( 智人 ) ; ptr 代表 *Pan troglodytes* ( 黑猩猩 ) 。 KEGG GENES 数据库可分为 3 大类 , 除占据主要部分的生物分类外 , 还包括病毒及附录分类。病毒分类的分类码为两个字母的 'vg' , 其标识符由 vg 和 Locus\_tag 组成 , 附录分类是为功能特征蛋白手动创建的数据库 , 分类码为两个字母的 'ag' , 标识符由 ag 和 ProteinID 组成。

### 1.3 KEGG ORTHOLOGY 数据库

介绍 KEGG ORTHOLOGY 数据库之前 , 我们先了解下 Ortholog , Ortholog 可翻译为直系 / 直向 / 垂直同源 ( 基因 ) , 指位于不同的物种间的 , 在物种形成过程中源自某一共同祖先的基因 : 1 ) 从进化的角度来看 , 这类基因通常具有相同的功能 , 但并非绝对 ; 2 ) 因此 , 当我们比较一对直系同源基因时 , 可能会发现有的已经丧失了固有的功能或者进化出了新的功能。。

KEGG ORTHOLOGY 是储存分子水平功能的数据库 , 属于直系同源的基因或蛋白被归为一类 , 打上 KO(K) 的标签。因此一个 KO 标识符可对应多个基因或蛋白。

在 KEGG ORTHOLOGY 数据库页面搜索 K number ( 如 K19188 ) , 我们得到以下结果 ( **Fig 1** ) , 在 Genes 一栏我们可以看到被归于 K19188 这一类的基因或蛋白。

 <b>ORTHOLOGY: K19188</b> <a href="#">Help</a>	
<b>Entry</b>	K19188 KO
<b>Name</b>	dhponh
<b>Definition</b>	2,6-dihydroxypseudooxynicotine hydrolase [EC:3.7.1.19]
<b>Pathway</b>	ko00760 Nicotinate and nicotinamide metabolism ko01120 Microbial metabolism in diverse environments
<b>Module</b>	M00810 Nicotine degradation, pyridine pathway, nicotine => 2,6-dihydroxypyridine/succinate semialdehyde
<b>Brite</b>	KEGG Orthology (KO) [BR:ko00001] Metabolism Metabolism of cofactors and vitamins 00760 Nicotinate and nicotinamide metabolism K19188 dhponh: 2,6-dihydroxypseudooxynicotine hydrolase KEGG modules [BR:ko00002] Pathway module Secondary metabolism Aromatics degradation M00810 Nicotine degradation, pyridine pathway, nicotine => succinate semialdehyde K19188 dhponh: 2,6-dihydroxypseudooxynicotine hydrolase Enzymes [BR:ko01000] 3. Hydrolases 3.7 Acting on carbon-carbon bonds 3.7.1 In ketonic substances 3.7.1.19 2,6-dihydroxypseudooxynicotine hydrolase K19188 dhponh: 2,6-dihydroxypseudooxynicotine hydrolase <a href="#">BRITE hierarchy</a>
<b>Other DBs</b>	RN: R07515 GO: 0034948
<b>Genes</b>	ROP: ROP_27470 ROA: Pd630_LPD07472 NCA: Noca_0613 AG: CAD47941(pnh) <a href="#">Taxonomy</a> <a href="#">KOALA</a> <a href="#">UniProt</a>
<b>Reference</b>	PMID:16321959
<b>Authors</b>	Sachelaru P, Schiltz E, Igloi GL, Brandsch R.
<b>Title</b>	An alpha/beta-fold C—C bond hydrolase is involved in a central step of nicotine catabolism by <i>Arthrobacter nicotinovorans</i> .
<b>Journal</b>	J Bacteriol 187:8516-9 (2005) DOI:10.1128/JB.187.24.8516-8519.2005
<b>Sequence</b>	[ag:CAD47941]
<b>LinkDB</b>	<a href="#">All DBs</a>

**Fig 1 K19188搜索结果**

另外，我们还可在 KEGG ORTHOLOG 页面的搜索框内输入任意 K 标识码（如 K19188），点击 Ortholog table 就能得到该 KO 分类在哪个物种中存在以及特定的名称（Fig 3）。

Ortholog table			
All ▼			
Grp	Genus	Organism	K19188 (dhponh)[4]
B.Act	Rhodococcus	<a href="#">rop</a>	<a href="#">ROP_27470</a>
B.Act	Rhodococcus	<a href="#">roa</a>	<a href="#">Pd630_LPD07472</a>
B.Act	Nocardioide	<a href="#">nca</a>	<a href="#">Noca_0613</a>

**Fig 3 Ortholog table ( K19188 )**

## 1.4 KEGG Pathway 数据库结构

KEGG Pathway 数据库，将代谢通路划分为 7 类，包括 Metabolism ( 新陈代谢 )、Genetic Information Processing ( 遗传信息处理 )、Environmental Information Processing ( 环境信息处理 )、Cellular Processes ( 细胞过程 )、Organismal Systems ( 生物系统 )、Human Diseases ( 人类疾病 )、Drug Development ( 药物发展 )，其中每一类又被系统分为二、三、四层。第二层目前包括 66 种子 pathway；第三层即为代谢通路图；第四层为每个代谢通路图的具体注释信息。打开 KEGG Pathway 网页，我们可以看到 7 类代谢通路及 KEGG Pathway 数据库标识符的组成。每一类代谢通路均可点击，直接跳转到该分类的子 pathway 及代谢通路图；另外 KEGG Pathway 数据库标识符，前面我们已经提到，其 prefix 包括 map/ko/ec/rn/<org>，而数字组成开始部分是根据 map 类型决定的；

<b>4. Cellular Processes</b>
<b>4.1 Transport and catabolism</b>
04144 Endocytosis
04145 Phagosome
04142 Lysosome
04146 Peroxisome
04140 Autophagy - animal
04138 Autophagy - yeast
04136 Autophagy - other
04137 Mitophagy - animal
04139 Mitophagy - yeast
<b>4.2 Cell growth and death</b>
04110 Cell cycle
04111 Cell cycle - yeast
04112 Cell cycle - Caulobacter
04113 Meiosis - yeast
04114 Oocyte meiosis
04210 Apoptosis
04214 Apoptosis - fly
04215 Apoptosis - multiple species
04216 Ferroptosis
04217 Necroptosis
04115 p53 signaling pathway
04218 Cellular senescence
<b>4.3 Cellular community - eukaryotes</b>
04510 Focal adhesion
04520 Adherens junction
04530 Tight junction
04540 Gap junction
04550 Signaling pathways regulating pluripotency of stem cells
<b>4.4 Cellular community - prokaryotes</b>
02024 Quorum sensing
05111 Biofilm formation - Vibrio cholerae
02025 Biofilm formation - Pseudomonas aeruginosa
02026 Biofilm formation - Escherichia coli
<b>4.5 Cell motility</b>
02030 Bacterial chemotaxis
02040 Flagellar assembly
04810 Regulation of actin cytoskeleton

Fig 4 Cellular Process 的第二及第三层级



我们以 Cellular Process 为例，点击该分类，页面跳转至以下页面（**Fig 4**），可直观的看到该分类包括 4.1-4.5 五个子 pathway，而每个子 pathway 下一层级为代谢通路图，每个代谢通路图前的数字为 KEGG Pathway 标识符的数字部分，如 map 04144 即可搜索到 Endocytosis 代谢通路图。

## 2. 宏基因组 KEGG 数据库注释过程及结果展示

### 2.1 宏基因组 KEGG 数据库标准注释过程

通过组装、基因预测等标准分析步骤，我们得到了 gene catalog，称为 Unigenes，其为各个基因对应的氨基酸序列信息，通过以下步骤，我们进行 KEGG 数据库标准分析：

- 1) KEGG GENES 数据库储存了各个基因的氨基酸序列信息，我们使用 DIAMOND 软件（V0.7.9）将 Unigenes 与功能数据库进行比对，参数设置：blastp，-e 1e-5；
- 2) BLAST 结果过滤：对于每一条序列的 BLAST 结果，选取 score 值最高的比对结果（one HSP $\geq$ 60 bits）进行后续分析；
- 3) KEGG GENES 数据库中，每条基因有所归属的 KEGG ORTHOLOGY 数据库中的信息（即 KO 层级）。通过 KEGG PATHWAY 数据库可获得每个 KO 参与的 pathway 信息。每个 pathway 可分为 level1（first level）、level2（second level）、level3（third level）这几个层级；

KEGG 数据库详细划分层级如下（**Table 3**）：

**Table 3** KEGG 数据库层级说明

划分层级	层级说明
level1	KEGG 代谢通路第一层级 7 大代谢通路



level2	KEGG 代谢通路第二层级 57 种子 pathway
level3	KEGG pathway id ( 例 : map00010 )
KO	KEGG ortholog group ( 例 : K00010 )
ec	KEGG EC number ( 例 EC : 3.4.1.1 )
module	KEGG Module number ( 例 : M00165 )

4) 从比对结果出发, 统计不同功能层级的相对丰度 ( 各层级的相对丰度等于注释为功能层级基因的相对丰度之和 ), 并进行统计分析, 分析结果详见网页版报告的交付文件目录列表。

## 2.2 KEGG Pathway 数据库注释结果展示

1) 诺禾致源为满足客户需求, 展现了不同分组 ( 样品 ) 在代谢通路图中的差异, 绘制了代谢通路网页版结果展示。在标准分析结题报告中, 找到代谢通路比较分析, 点击 “展示结果请点击”, 即可跳转至代谢通路网页版分析结果。

2) 点击进入后, 可发现整体网页版分析结果分为两部分:

The annotated pathways

for mPATH analysis

第一部分分析结果

Compare Samples		Strict Link
CLC_VS_CLW		> <a href="#">please click</a>
CLW_VS_CLF		> <a href="#">please click</a>
CLC_VS_CLF		> <a href="#">please click</a>

for other pathways' genenum and EC's MetaStats Analysis

第二部分分析结果

Pathway ID	Pathway Level1	Pathway Level2	Pathway Level3	Identified ECs	Sign_diff ECs
<a href="#">map04151</a>	Environmental Information Processing	Signal transduction	PI3K-Akt signaling pathway	2.7.11.1	
<a href="#">map04070</a>	Environmental Information Processing	Signal transduction	Phosphatidylinositol signaling system	3.1.4.11; 2.7.7.41; 2.7.1.107; 3.1.3.25	

第一部分为 KEGG 9 大 pathway overview 图, 图中展示了两个分组 ( 或两个样品 ) 共有及特有的代谢通路信息, 在代谢通路图中, 节点代表各种化合物,

边代表一系列的酶类反应，红色代表两个分组（或两个样品）共有的酶类反应，蓝色代表分组 A（或样品 A）独有的酶类反应，绿色代表分组 B（或样品 B）独有的酶类反应；

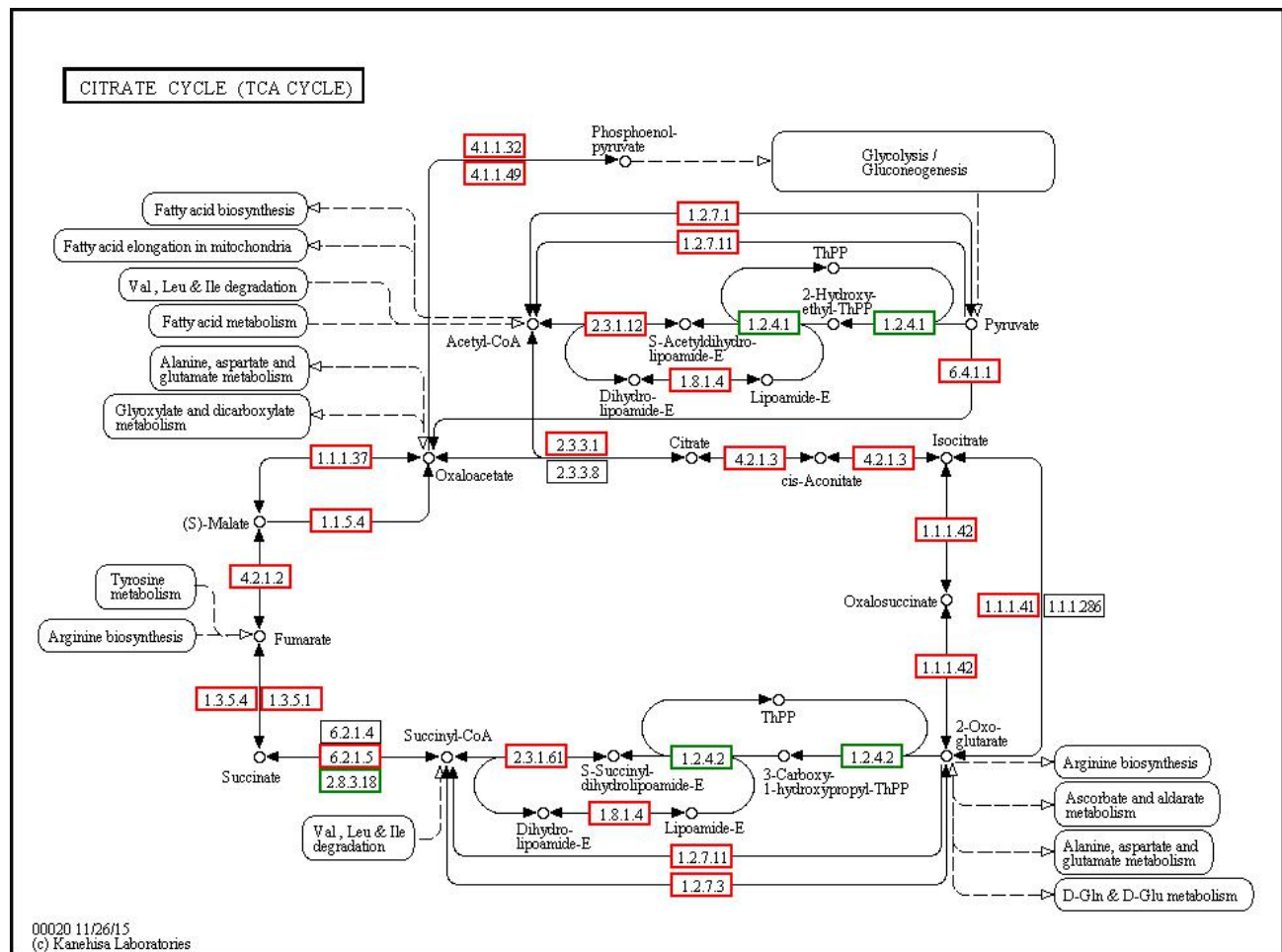


Fig 5 TCA cycle 图

通过 pathway 图，我们展示了两个分组（样品）共有及特有的代谢通路信息，比如老师量查看两分组在 *TCA cycle* 代谢上的整体差异，可在打开的网页通过 control+f 快速定位到 *TCA cycle* 代谢通路图，点击第一列的标识符，即可查看两组在 *TCA cycle* 代谢通路上的差异，蓝色或绿色代表独有的代谢反应（Fig 5），鼠标移动至蓝线（框）或绿线（框）上可显示两组独有酶类的 KO

number , EC:number 及独有的反应 R number , 点击后即可跳转至 KEGG 数据库网站 , 显示所有标注到的信息,以酶 1.2.4.1 为例 : **Fig 6** ;

KEGG

ORTHOLOGY: K00163

Help

Entry

K00163

KO

Name

aceE

Definition

pyruvate dehydrogenase E1 component [EC:1.2.4.1]

KEGG

ORTHOLOGY: K00161

Help

Entry

K00161

KO

Name

PDHA, pdhA

Definition

pyruvate dehydrogenase E1 component alpha subunit [EC:1.2.4.1]

KEGG

ORTHOLOGY: K00162

Help

Entry

K00162

KO

Name

PDHB, pdhB

Definition

pyruvate dehydrogenase E1 component beta subunit [EC:1.2.4.1]

KEGG

ENZYME: 1.2.4.1

Help

Entry

EC 1.2.4.1

Enzyme

Name

pyruvate dehydrogenase (acetyl-transferring);  
 MtPDC (mitochondrial pyruvate dehydrogenase complex);  
 pyruvate decarboxylase;  
 pyruvate dehydrogenase;  
 pyruvate dehydrogenase (lipoamide);  
 pyruvate dehydrogenase complex;  
 pyruvate:lipoamide 2-oxido-reductase (decarboxylating and acceptor-acetylating);  
 pyruvic acid dehydrogenase;  
 pyruvic dehydrogenase

KEGG

REACTION: R00014

Help

Entry

R00014

Reaction

Name

pyruvate:thiamin diphosphate acetaldehydetransferase (decarboxylating)

Definition

Pyruvate + Thiamin diphosphate <=> 2-(alpha-Hydroxyethyl)thiamine diphosphate + CO2

Equation

C00022 + C00068 <=> C05125 + C00011

**Fig 6** ec:1.2.4.1 标注信息网站展示

第二部分为注释到的 pathway 代谢通路图，在代谢通路图中，节点代表各种化合物, 方框代表酶类信息（默认边框为黑色，背景为白色），不同颜色的方框代表注释为该酶类的不同 Unigenes 数目，黄色背景的酶类代表在分组间具有显著差异的酶类（若没有进行显著差异分析，则没有此部分信息），鼠标移动至该酶类，可显示差异酶类在不同分组间的丰度分布箱图。

通过第二部分的 pathway 代谢通路图，我们可以知道：

1) 在该代谢通路图中, 都有哪些酶类被注释上了? 外框有颜色的酶类, 代表被注释上的酶类;

2) 注释上该酶类的基因数目是多少? 外框不同的颜色代表注释为该酶类的基因数目高低;

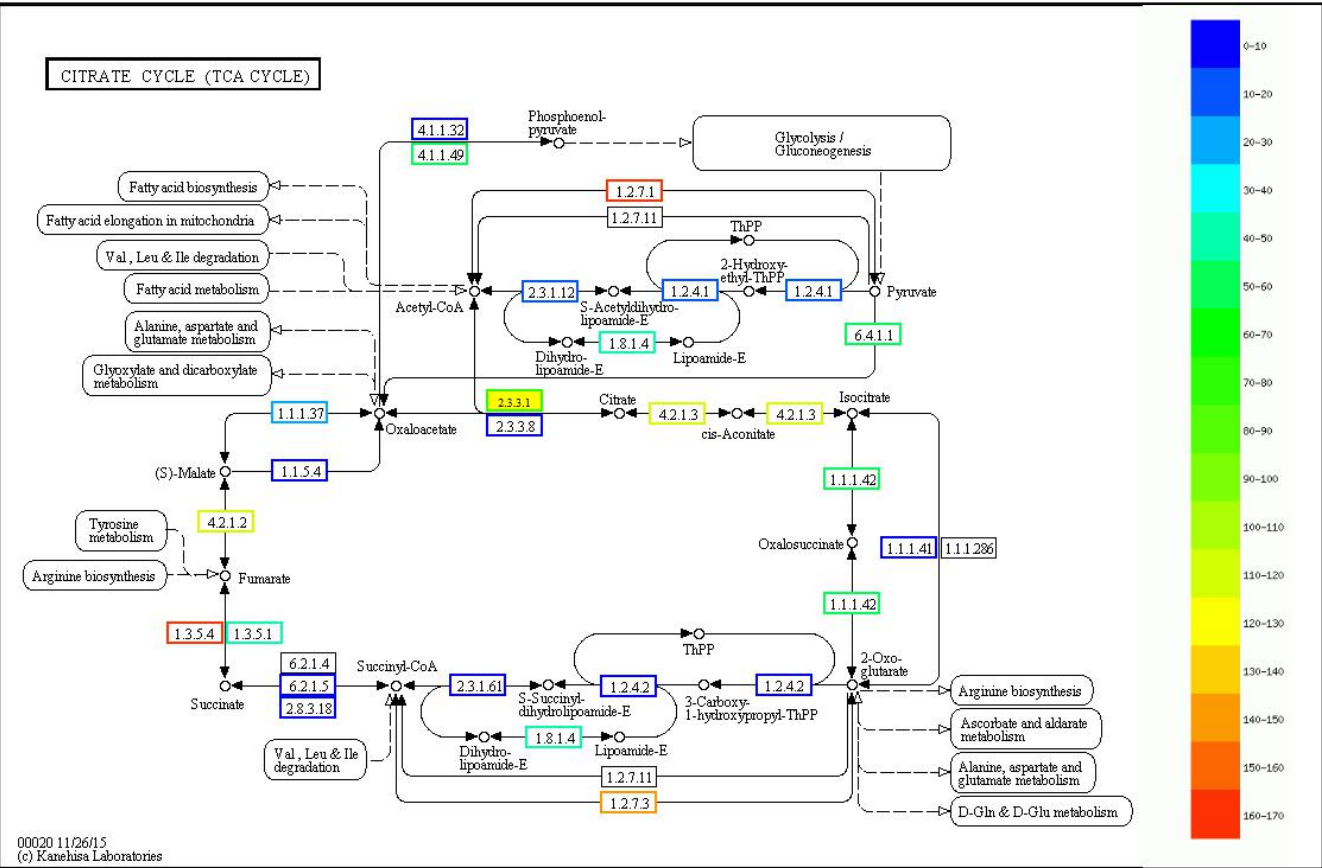
3) 哪些酶类是在组间具有显著差异的? 方框背景变为黄色表明有差异的酶类;

4) 具备差异的酶类在组间丰度分布是怎样的? 鼠标移动到差异酶类方框处, 即会弹出差异酶类在组间丰度分布箱图, 横坐标代表分组, 纵坐标代表丰度, \*代表差异显著  $q\ value < 0.05$ , \*\*代表差异极显著  $q\ value < 0.01$ 。

比如, 老师想查看各组在 *TCA cycle* 代谢通路图上的注释情况, 可通过 `control+f` 快速定位到 *TCA cycle* 代谢通路图, 点击 pathway 标识符, 即可看到各组在该代谢通路上的注释情况 (Fig 7)。测试报告显示, 能够被注释上的酶类有(显著在 Identified Ecs 一栏): 1.8.1.4; 1.2.4.2; 2.3.1.61; 6.2.1.5; 1.1.1.42; 1.1.1.41; 2.3.3.8; 4.2.1.2; 4.2.1.3; 2.3.3.1; 1.1.1.37; 6.4.1.1; 1.2.4.1; 2.3.1.12; 4.1.1.49; 4.1.1.32; 2.8.3.18; 1.3.5.1; 1.1.5.4; 1.3.5.4; 1.2.7.1; 1.2.7.3, 这些酶类会在 pathway 图中用不同颜色外框展示 (Fig 8); 注释到的酶类中具备差异的酶为 2.3.3.1, 该差异酶在组间的丰度分布情况, 可用鼠标移动至 *TCA cycle* 代谢通路图中该差异酶方框处 (黄色背景标注), 即会在网页右边显示该酶在组间丰度分布箱图 (Fig 9)。

Pathway ID	Pathway Level1	Pathway Level2	Pathway Level3	Identified ECs	Sign_diff ECs
<a href="#">map00020</a>	Metabolism	Carbohydrate metabolism	Citrate cycle (TCA cycle)	1.8.1.4; 1.2.4.2; 2.3.1.61; 6.2.1.5; 1.1.1.42; 1.1.1.41; 2.3.3.8; 4.2.1.2; 4.2.1.3; 2.3.3.1; 1.1.1.37; 6.4.1.1; 1.2.4.1; 2.3.1.12; 4.1.1.49; 4.1.1.32; 2.8.3.18; 1.3.5.1; 1.1.5.4; 1.3.5.4; 1.2.7.1; 1.2.7.3	2.3.3.1

Fig 7 TCA cycle 注释结果



注：该图中颜色图例在网页版报告中是不显示的，而是保存在结果文件中  
 result/05.FunctionAnnotation/KEGG/pathwaymaps/color\_direction.gif

Fig 8 TCA cycle 注释结果



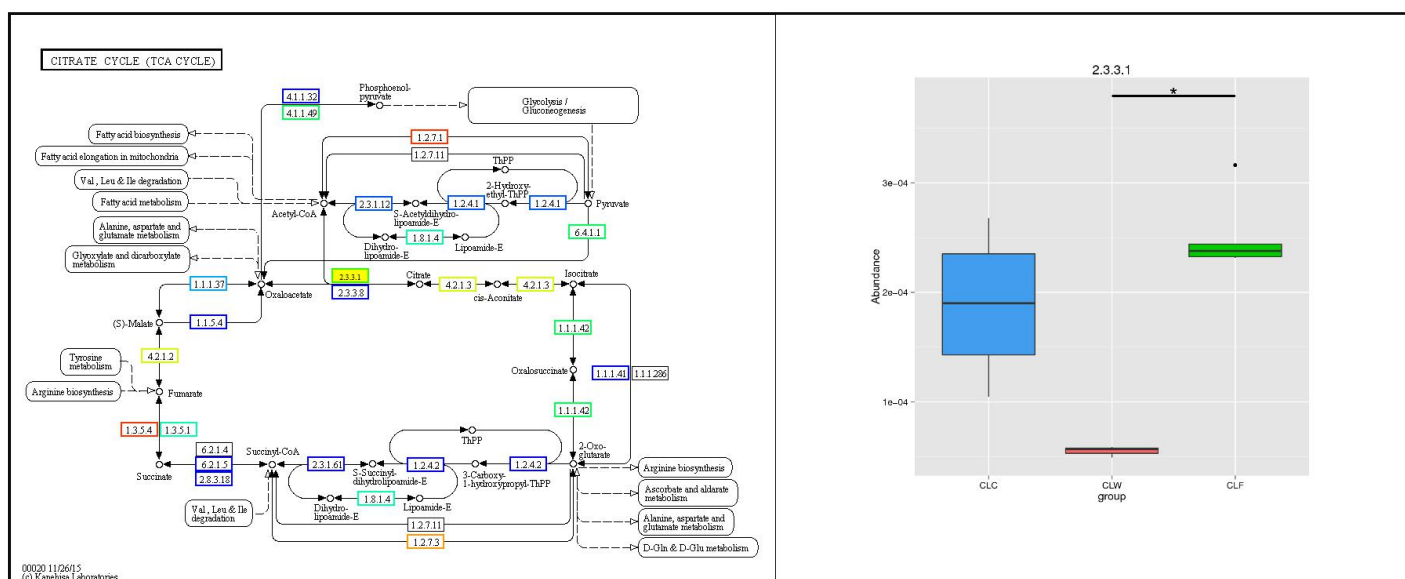


Fig 9 差异酶及其在各组间丰度分布箱图

### 3. 宏基因组 KEGG 数据库注释常见问题

#### 3.1 KEGG 注释结果中哪个文件将物种与功能联系起来？

将物种注释结果与 KEGG 功能注释结果联系起来的文件为 Unigenes.KEGG.tax.xls，位于结果文件的：/TJPROJ1/MICRO/gaojianpeng/meta/NHM140053-N01\_12\_wolffeces\_meta\_20161021/result/05.FunctionAnnotation/KEGG/KEGG\_Anno/Unigenes.KEGG.tax.xls；

从 KEGG Pathway 分析结果，我们可以找到某些分组独有的酶或反应，而我们想知道是那些物种导致了这种差异，是通过查看 Unigenes.KEGG.tax.xls

( Fig 10 )

Map_ID	KO_ID	EC_ID	Gene_ID	Taxonomy
map00010	K15633	5. 4. 2. 12	NOVO_MIX	k_Bacteria;p_Proteobacteria;c
map00010	K01810	5. 3. 1. 9	NOVO_MIX	k_Bacteria;p_Proteobacteria;c
map00010	K00121	1. 1. 1. 28	NOVO_MIX	k_Bacteria;p_Proteobacteria;c
map00010	K00128	1. 2. 1. 3	NOVO_MIX	k_Bacteria;p_Proteobacteria;c
map00010	K00162	1. 2. 4. 1	NOVO_MIX	k_Bacteria;p_Proteobacteria;c
map00010	K03841	3. 1. 3. 11	NOVO_MIX	k_Bacteria;p_Proteobacteria;c
map00010	K01895	6. 2. 1. 1	NOVO_MIX	k_Bacteria;p_Proteobacteria;c

Fig 10 Unigenes.KEGG.tax.xls

### 3.2 KEGG 代谢通路中，各元素代表的含义

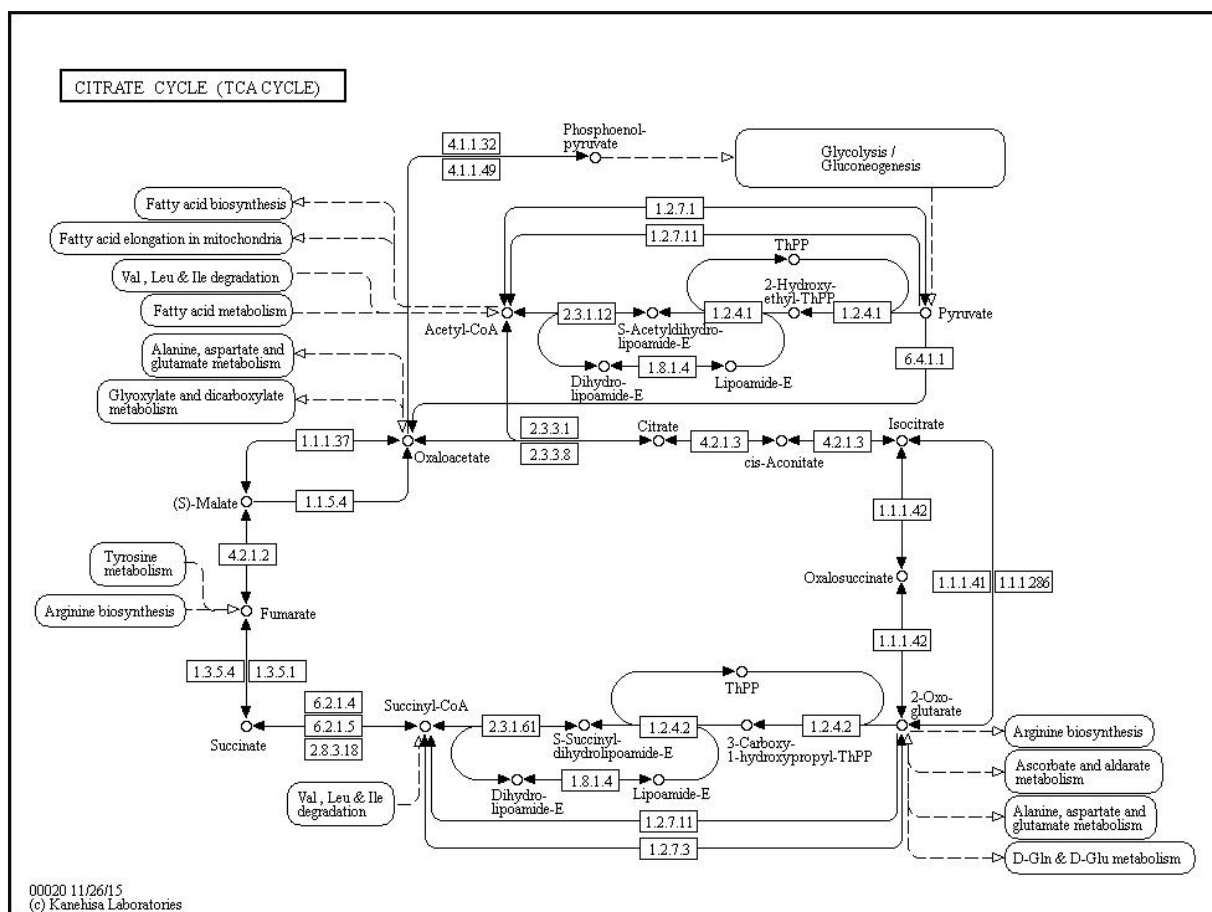


Fig 11 TCA cycle of reference pathway

以 *TCA cycle* 代谢通路图为例（Fig 11），图中各形状代表的含义如下：

- 1) 方框一般代表行使某一功能的酶类，方框内的数字为酶类的编号，鼠标移动至某一方框，会显示行使该功能的全部基因的 KO 标识符。因此，有时候方框内没有酶的编号，仅由一个名称填充，代表行使该功能的 KEGG ORTHOLOGY 并没有酶的标识符信息，只有名称信息；
- 2) 小圆圈代表代谢物，打开代谢通路图后，将鼠标移动至任一圆圈会出现 KEGG COMPOUND 数据库的标识符，如 C00068，C 代表 compound，数字代表该化合物在 KEGG 中的编号。
- 3) 实线的箭头代表反应过程，比如图中 Oxaloacetate 经过 1.1.1.37 该酶的



作用变成了 ( S ) -Malate。

- 4) 大的圆角矩形代表另外一个代谢通路图，该代谢通路图和其他代谢图相连的反应过程是由虚线的箭头来表示的；

更多详细信息，可参考如下说明 ( Fig12 )：

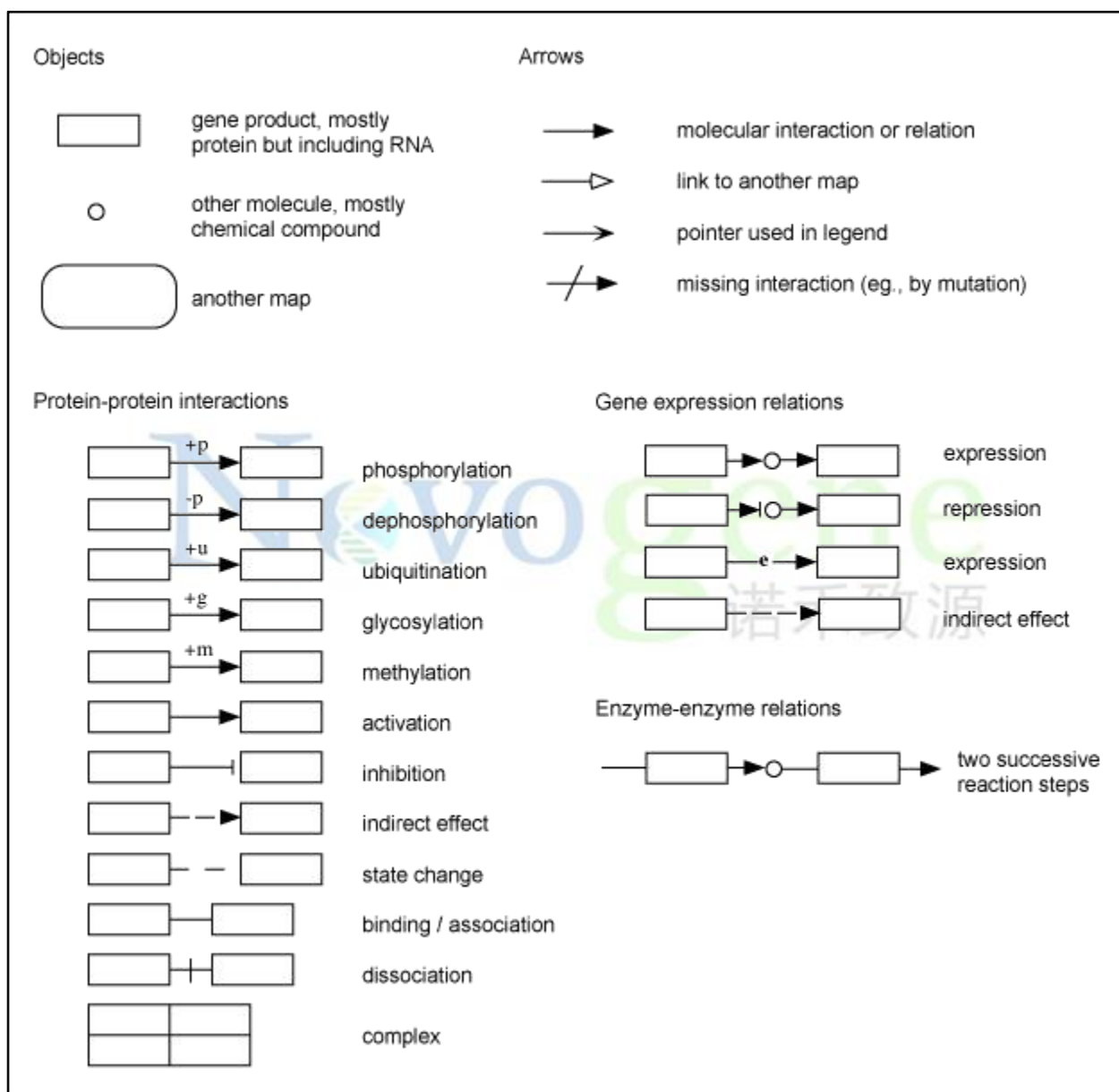


Fig 12 KEGG pathway 符号说明

### 3.3 有时候，图中方框变成绿色或蓝色代表什么？

在 KEGG 官网打开 TCA cycle 代谢通路图

( [http://www.kegg.jp/kegg-bin/highlight\\_pathway?scale=1.0&map=map00020&keyword=00020](http://www.kegg.jp/kegg-bin/highlight_pathway?scale=1.0&map=map00020&keyword=00020) ) , 可以看到该网页顶部有一个下拉菜单 , 里面包括 Reference pathway 在内的多个选项。

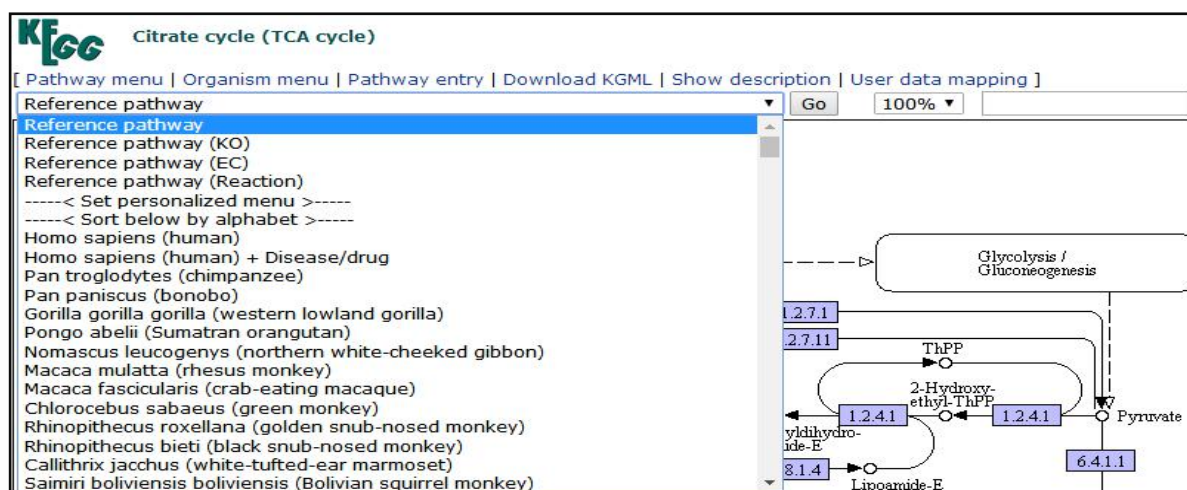


Fig 13 TCA cycle pathway 页面

在 KEGG PATHWAY 数据库中，一般有三种种代谢图（ Fig 13 ）：

1. 第一种，为参考代谢通路 reference pathway，是根据已有知识绘制的概括性的、详尽的具有一般参考意义的代谢图，这种图上都是纯黑白色的，所有的框都可以鼠标点击查看更详细的信息；

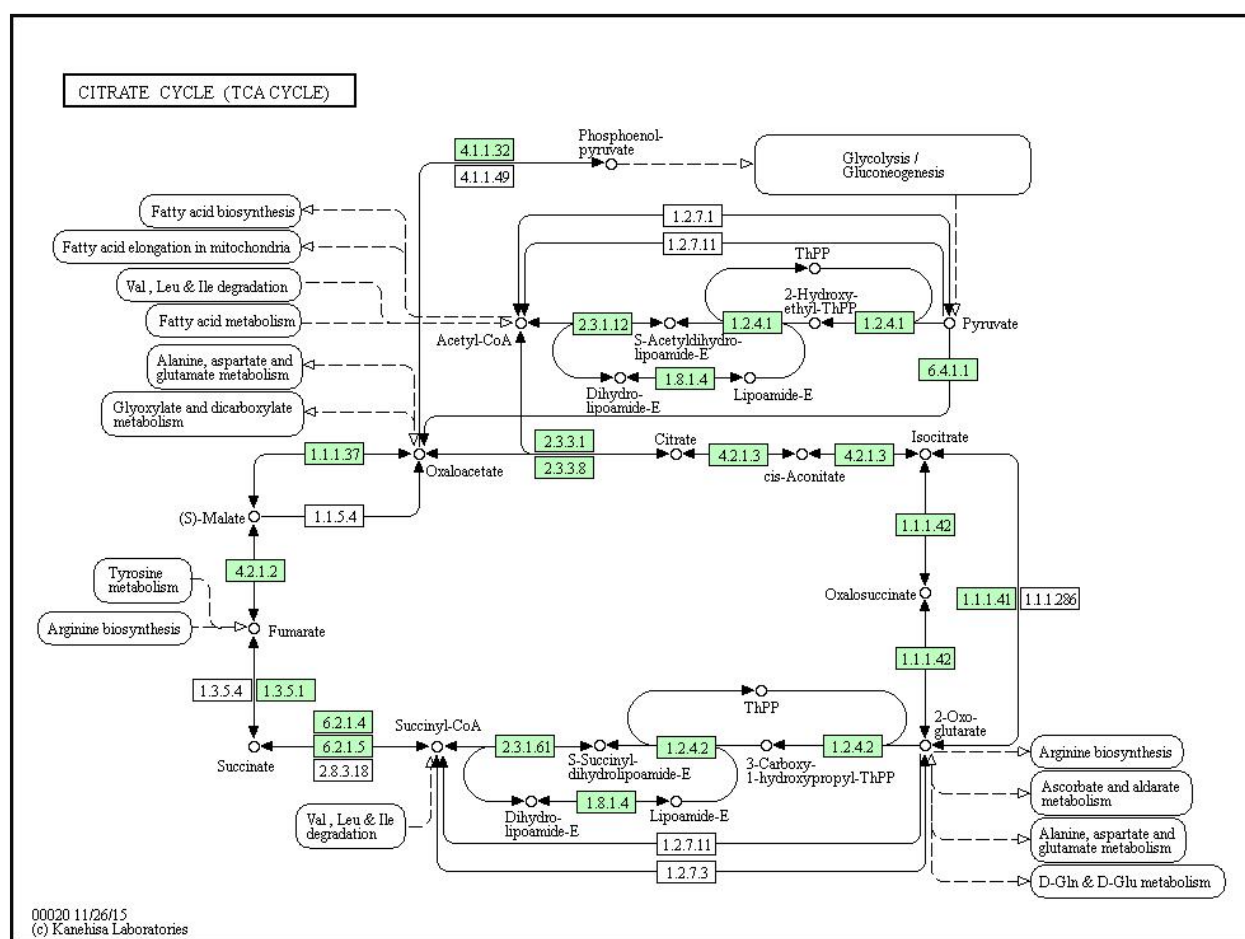
2. 第二种则是选择方框显示的信息，对应选项为 Reference pathway( KO )，Reference pathway ( EC ) 及 Reference pathway ( Reaction )，在这一类图中，方框会变成蓝色，移动到方框仅会显示 K number，EC number 及 R/RC number，方便我们仅关注一类时进行查找；

3. 第三种则是属于特定的某个物种的代谢图，图中，有些方框变为绿色，表示该物种特有的基因或酶。在这张图中仅有显示为绿色的方框可以被点击并给出详细的注释信息。另外，reference pathway 在 KEGG 中的名字是以 map 开头的，比如 map00020（即 TCA cycle 参考代谢通路图），而特定物种的代谢通

路图,开头的几个字符不是 map 而是物种英文单词的缩写,比如 Homo sapiens 的 TCA 循环代谢通路图,就是 hsa00020,物种名称缩写:

[http://www.kegg.jp/kegg/catalog/org\\_list.html](http://www.kegg.jp/kegg/catalog/org_list.html)

选择 Homo sapiens(human)后,属于人的 TCA cycle 代谢通路图则变成下面的图片 ( **Fig 14** ), 酶 4.1.1.32 为人的 TCA 循环特有的,而酶 4.1.1.49 为其他物种参与该步反应所需的,并且只有标记为绿色的酶才会显示信息。



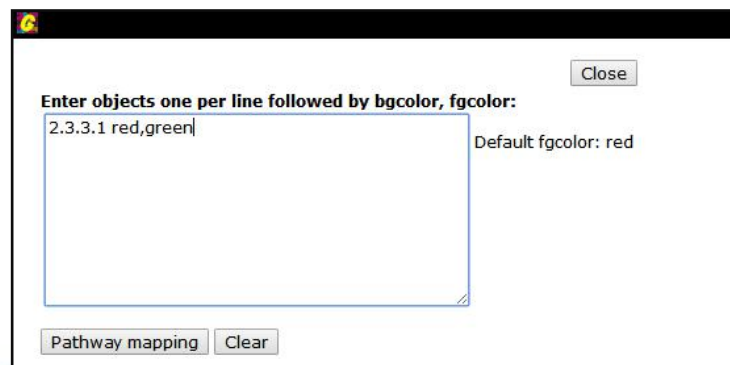
**Fig 14** TCA cycle of Homo sapiens

### 3.4 如何对代谢通路图中某个感兴趣的酶进行标记？

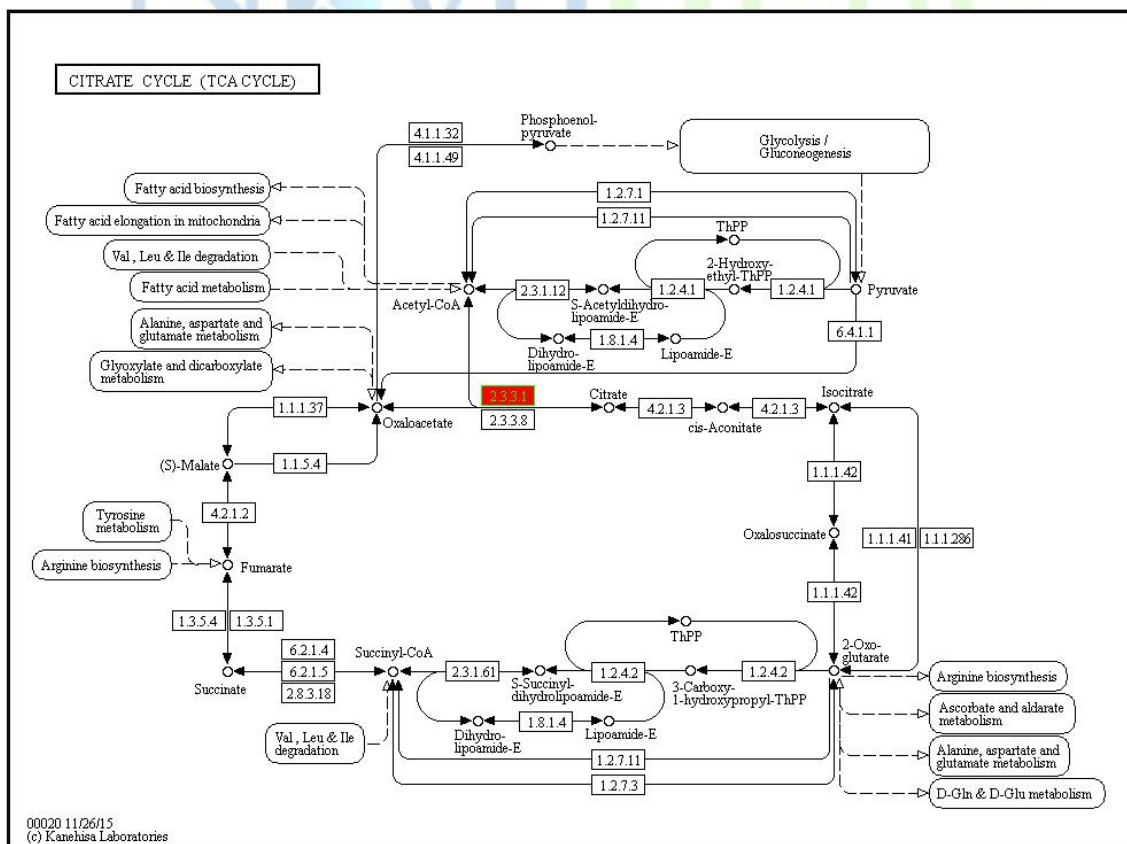
仍然以 TCA cycle 代谢通路图为例,在 KEGG 官网打开该 mapping, 可看到 **User data mapping** 选项。



打开后的对话框如下图（**Fig 15**），写入想要标记的特殊酶类，一行一条信息，其中 bgcolor 为背景色，fgcolor 代表前置色，输入完成后点击 pathway, 2.3.3.1 就会变成特殊标记了（**Fig16**）。



**Fig 15** 颜色设置页面



**Fig 16** 标记酶 2.3.3.1

### 3.5 global map , overview map 及 regular map ?

TCA cycle 代谢通路图这类 map 被称为 regular map ,通常以方框标注酶类 , 并且会展示 KO 号( k number )。而 global map 和 overview map 包含在 KEGG Pathway 数据库 Metabolism 分类下 , 是新陈代谢通路图的总览图 , 这两种 map 图中是用箭头来展示 KO 号 ,大的圆角矩形代表该代谢通路涉及到的子代谢通路图 ( map number ) , 另外这两种 map 标识符的数字部分也有区别。具体查看 **Table 4**。

**Table 4** map number of different type

Type	map number	remarks
Global map	011XX	lines linked to KOs
Overview map	012XX	lines linked to KOs
chemical structure map	010XX	no KO expansion
drug structure map	07	no KO expansion
regular map	other	boxes linked to KOs

global map 和 overview map 分别包含 4 张和 5 张图 , 具体查看 **Table 5**。

**Table 5**

category	maps	detail
global map	Metabolic pathways	总代谢通路总图
	Biosynthesis of secondary metabolites	次级代谢通路总图
	Microbial metabolism in diverse environments	不同环境中代谢通路总图
	Biosynthesis of antibiotics	抗生素合成总图
overview map	Carbon metabolism	碳代谢概览图
	2-Oxocarboxylic acid metabolism	$\alpha$ -酮酸代谢概览图 , 包含了
		TCA cycle , arginine bios-

	ynthesis 等代谢通路图
Fatty acid metabolism	脂肪酸代谢概览图 ,包含了 TCA cycle 代谢通路图
Biosynthesis of amino acids	20 种氨基酸合成概览图
Degradation of aromatic compounds	芳香族化合物降解概览图

## 4.参考文献

KEGG 官网介绍	Kanehisa M, Furumichi M, Mao T, et al. KEGG: new perspectives on genomes, pathways, diseases and drugs[J]. Nucleic Acids Research, 2017, 45(D1):D353.
Ortholog	Li W H, Al E. Expression divergence between duplicate genes.[J]. Trends in Genetics TIG, 2005, 21(11):602. Fitch W M. Distinguishing Homologous from Analogous Proteins[J]. Systematic Zoology, 1970, 19(2):99.
KEGG 宏基因组标准注释过程	Bäckhed F, Roswall J, Peng Y, et al. Dynamics and Stabilization of the Human Gut Microbiome during the First Year of Life[J]. Cell host &microbe, 2015, 17(5): 690-703. Qin J, Li Y, Cai Z, et al. A metagenome-wide association study of gut microbiota in type 2 diabetes[J]. Nature, 2012, 490(7418): 55-60. Li J, Jia H, Cai X, et al. An integrated catalog of reference genes in the human gut microbiome[J]. Nature biotechnology, 2014, 32(8): 834-841. Feng Q, Liang S, Jia H, et al. Gut microbiome development along the colorectal adenoma–carcinoma sequence[J]. Nature communications,2015, 6.