

# パスウェイデータベース

京都大学化学研究所バイオインフォマティクスセンター  
五斗 進

<http://goto.kuicr.kyoto-u.ac.jp/>

AJACS京都 2016/9/1

## 講習会の流れ

- ・ パスウェイデータベースの概要
  - ・ パスウェイデータに触れてみる
- ・ データをパスウェイにマッピングする 1
  - ・ BioCyc と Reactome でのオミックス解析
- ・ データをパスウェイにマッピングする 2
  - ・ KEGG でのオミックス解析

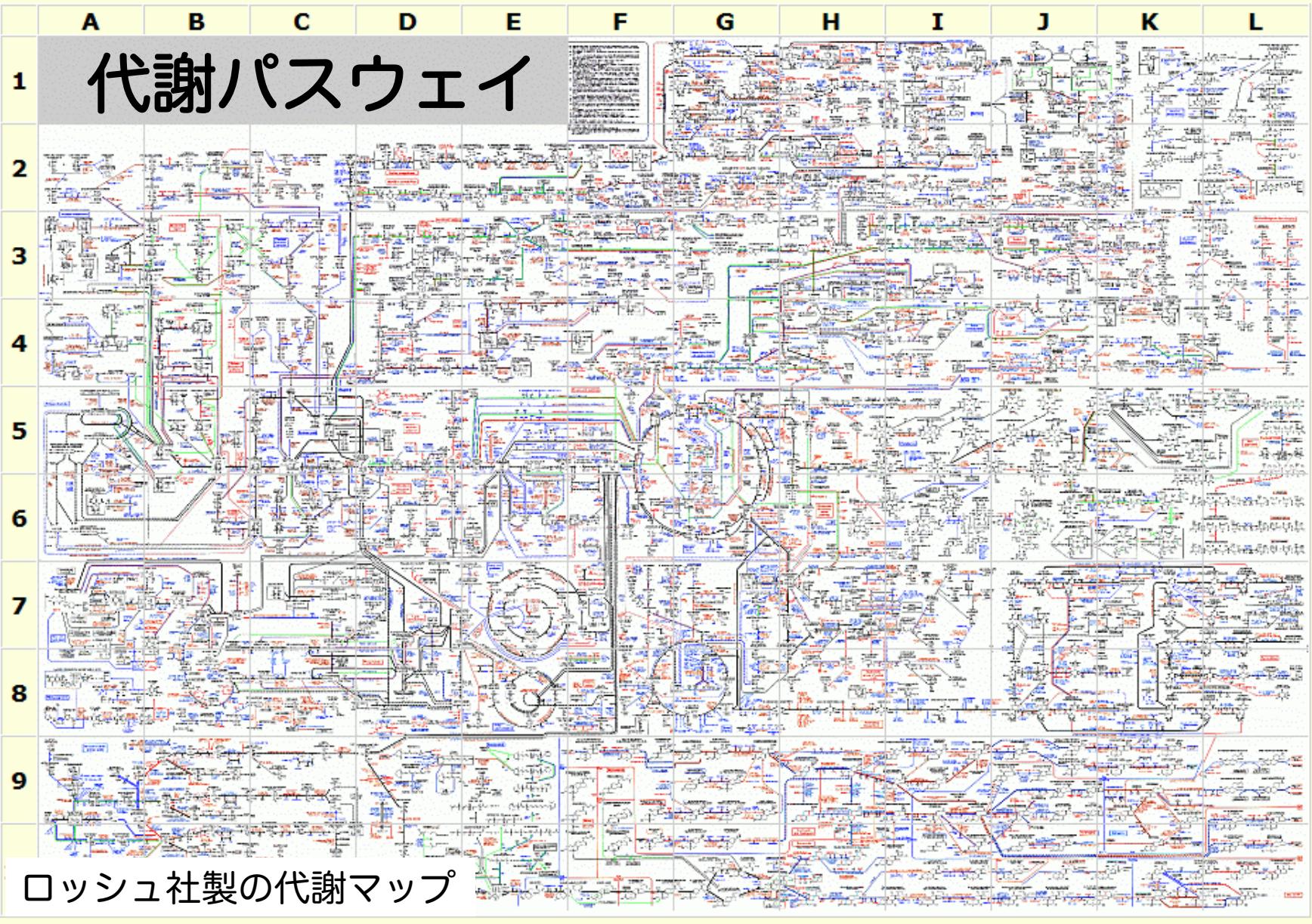
演習のサイト

<http://goto.kuicr.kyoto-u.ac.jp/lecture/pathway> 2

## パスウェイ

- ・既存の知識の統合化・可視化
  - ・生体内での遺伝子・タンパク質・化合物などの分子間相互作用を「経路」として表現したもの
  - ・生体分子ネットワークの容易な理解
- ・ダイアグラムによるマップ表現
  - ・ベーリンガー・マンハイム社 (Gerhard Michal 博士) の代謝マップ (1965年~、現ロッシュ)
  - ・初期は壁紙、後にデータベース化

# 代謝パスウェイ



Roche - Pathways

www.roche.com/sustainability/what\_we\_do/for\_communities\_and\_environment/philanthropy/science\_education/pathways.htm

Menu Search Sustainability > What we do > For communities and the environment > Biochemical Pathways DE Roche

**Science in the center**

## Roche Biochemical Pathways

The Biochemical Pathways Wall Charts have enjoyed worldwide popularity as a standard reference for nearly 50 years.

Because we care about science and scientists, the charts can now be accessed as an interactive online service with additional features.

**Explore the pathways**

Mapping the paths of life

Biochemical Pathways provide an overview of the chemical reactions of cells in various species and organs. Dr. Michal first compiled the Pathways Chart in 1965 and has been fine-tuning it ever since. Today, and with the collaboration of Roche, the two enormous posters can be found hanging in just about every research institute from Argentina to New Zealand.

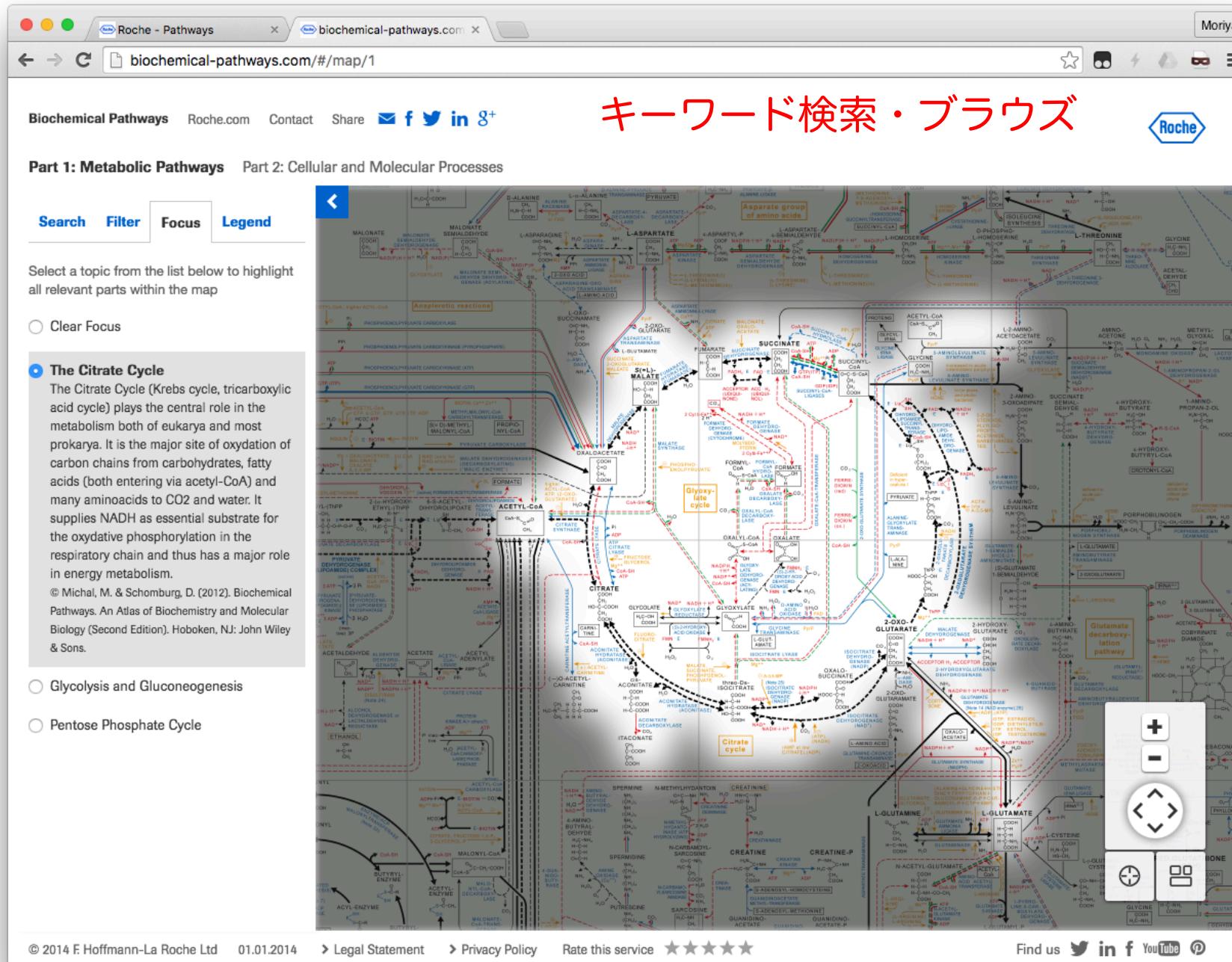
You have to be someone with tenacity and patience. And love for science.

**Dr. Gerhard Michal**  
Editor of the Roche Biochemical Pathways

By the numbers

**49 years**  
continuously fine-tuned by the editor himself.

[http://www.roche.com/sustainability/what\\_we\\_do/for\\_communities\\_and\\_environment/philanthropy/science\\_education/pathways.htm](http://www.roche.com/sustainability/what_we_do/for_communities_and_environment/philanthropy/science_education/pathways.htm)



<http://biochemical-pathways.com/#/map/1>

## パスウェイデータベース

- ・ 計算機による網羅的な解析への応用
  - ・ ゲノムの機能アノテーション。
  - ・ 生物種間比較と進化解析。
  - ・ 遺伝子発現データなどのエンリッチメント解析。
  - ・ モデリング、シミュレーション、予測

## 今日の目標

- パスウェイデータベース（KEGG、BioCyc、Reactome）とその活用法について学んで、実際のデータを適用できるようになる。
- ゲノムやメタゲノムの遺伝子機能アノテーションから、どのような代謝機能や生理機能があるかを調べることができるようになる。
- 複数のゲノム間やサンプル間での機能の比較ができるようになる。

# BioCyc

- ・ ウェブサイト : <http://biocyc.org/>
- ・ 開発 : SRIインターナショナル (Stanford Research Institute)
- ・ 対象 : 大腸菌からヒトまで、異株を含めて 7,600 種以上
- ・ 専門家が手作業で作成した文献ベースのデータ + 自動ツール
- ・ 代謝パスウェイ、制御系
- ・ 利用 : アカデミックフリー
- ・ データ形式 : BioPAX

# BioCyc

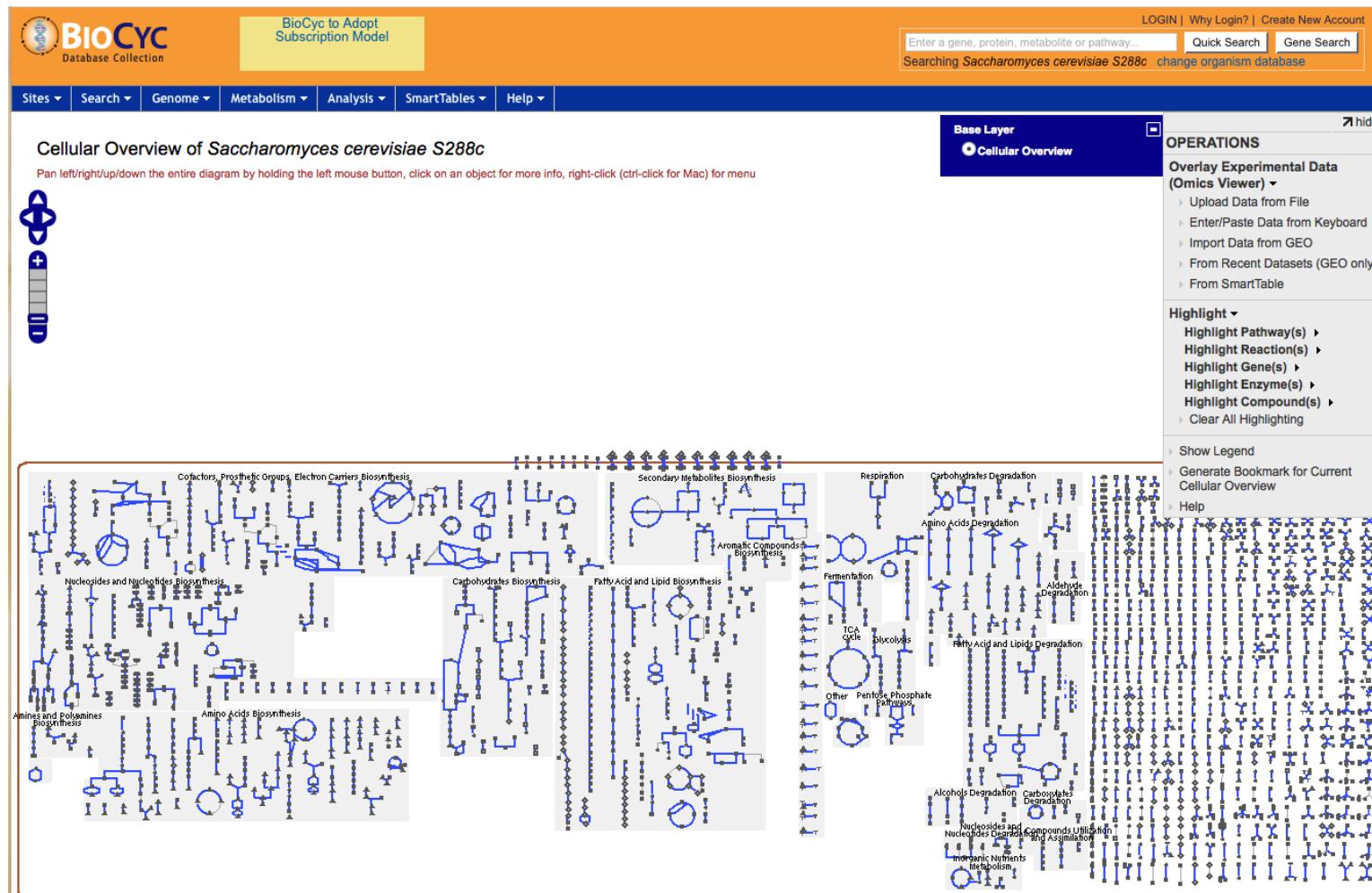
The screenshot shows the BioCyc homepage. At the top, there's a navigation bar with links for LOGIN, Why Login, Pathway Tools FBA Tutorial, Aug 10th-11th Registration Open, and a search bar. Below the navigation is a menu bar with Sites, Search, Genome, Metabolism, Analysis, SmartTables, and Help. The main content area has a title 'BioCyc Database Collection' and a sub-section 'BioCyc is a collection of 7615 Pathway/Genome Databases (PGDBs), plus software tools for understanding their data.' It also includes a 'Getting Started' section with usage instructions and a 'New User Guide >' link. On the right side, there's a large diagram titled 'RouteSearch' showing a metabolic pathway. The diagram illustrates three routes: a 'Long Route' (orange circles), a 'Loss of Atoms' route (green dashed line), and the 'Best Route: Atoms Conserved' (blue line). A chemical structure of a molecule is shown with numbered points 1, 2, and 3 indicating specific steps or atoms. Below the diagram is a text box for 'RouteSearch: Search for Paths through the Metabolic Network' and a 'Learn More' button. At the bottom of the page, there are navigation links from 1 to 10.

- 3つのレベル
- Tier 1 :
  - 専門家が手作業で作成した文献ベースのデータから構築したデータベース (7 DBs)
- Tier 2 :
  - 自動ツールで作成したデータを手作業で修正 (>40 DBs)
- Tier 3 :
  - 自動ツールで作成 (>7,500 DBs)
- MetaCyc
  - 2,000 種以上から構築したリファレンス代謝パスウェイ

<http://biocyc.org/biocyc-pgdb-list.shtml>

# 例えば、YeastCyc by BioCyc

Tier 1 → YeastCyc → Tools の metabolic map diagram



右上のサーチボックスでキーワード（例：TCA）を入れると個別マップの検索もできる

# YeastCyc の詳細

[YeastCyc Home](#) | [Search](#) | [Tools](#) | [Downloads](#) | [Help](#) | [Pathway](#)

**Saccharomyces cerevisiae Pathway: TCA cycle, aerobic respiration**

glycolysis  
pyruvate  
lysine biosynthesis  
aspartate biosynthesis  
exaloacetic acid  
malate  
fumarate  
succinate  
succinyl-CoA  
2-oxoglutarate  
glutamate biosynthesis from ammonia  
glutamate biosynthesis from glutamine

If an enzyme name is shown in bold, there is experimental evidence for this enzymatic activity.

Locations of Mapped Genes:

Synonyms: TCA, TCA cycle, tricarboxylic acid cycle, citric acid cycle

Superclasses: [Generation of Precursor Metabolites and Energy](#) → [TCA cycle](#)

Net Reaction Equation: acetate =  $2\text{CO}_2 + \text{H}_2\text{O}$

Summary:  
Eukaryotes have both NADP(+)-dependent isocitrate dehydrogenases (EC 1.1.1.42) and NAD(+)-dependent isocitrate dehydrogenases (EC 1.1.1.41). It has been shown that NAD(+)-IDH is involved in respiration while NADP(+)-IDH is involved in glutamate biosynthesis.

Citations: [[Haselbeck93](#), [Yasutake03](#), [Oyedotun97](#)]

Superpathways: [superpathway of TCA cycle and glyoxylate cycle](#)

Variants: [glyoxylate cycle](#)

Pathway Evidence Glyph:

Key to pathway glyph edge colors: ?

An enzyme catalyzing this reaction is present in this organism  
The reaction and any enzyme that catalyzes it (if one has been identified) is unique to this pathway

**SGD YeastCyc BIOCHEMICAL PATHWAYS**

Searching *Saccharomyces cerevisiae* Pathway Database

[YeastCyc Home](#) | [Search](#) | [Tools](#) | [Downloads](#) | [Help](#) | [Reaction](#)

**Saccharomyces cerevisiae Reaction: 6.2.1.5**

Species Comparison

Superclasses: [Reactions-Classified-By-Conversion-Type](#) → [Simple-Reactions](#) → [Chemical-Reactions](#) → [EC-Reactions](#) → [6 .. Ligases](#) → [6.2 .. Forming carbon-sulfur bonds](#) → [6.2.1 .. Acid--thiol ligases](#)  
[Reactions-Classified-By-Substrate](#) → [Small-Molecule-Reactions](#)

Enzymes and Genes:  
**succinyl-CoA ligase**: [LSC2](#), [LSC1](#)

In Pathway: [TCA cycle, aerobic respiration](#)

The reaction direction shown, that is, A + B  $\rightleftharpoons$  C + D versus C + D  $\rightleftharpoons$  A + B, is in accordance with the Enzyme Commission system.

Enzyme Commission Primary Name for this Reaction: Succinate--CoA ligase (ADP-forming)

Enzyme Commission Synonyms for this Reaction: Succinate thiokinase, Succinyl-CoA synthetase (ADP-forming), succinyl CoA synthesis

$\Delta G^\circ$  (kcal/mol): -0.7 KCAL/MOLE

Gene-Reaction Schematic: ?

Unification Links: [BRENDA:6.2.1.5](#), [ENZYME:6.2.1.5](#)

[Report Errors or Provide Feedback](#)  
Page generated by SRI International [Pathway Tools version 15.0](#) on Wed Oct 2, 2013.

[Search \*Saccharomyces cerevisiae\* Pathway Database](#)  Quick Search

[Return to SGD](#) [Send a Message to the SGD Curators](#)

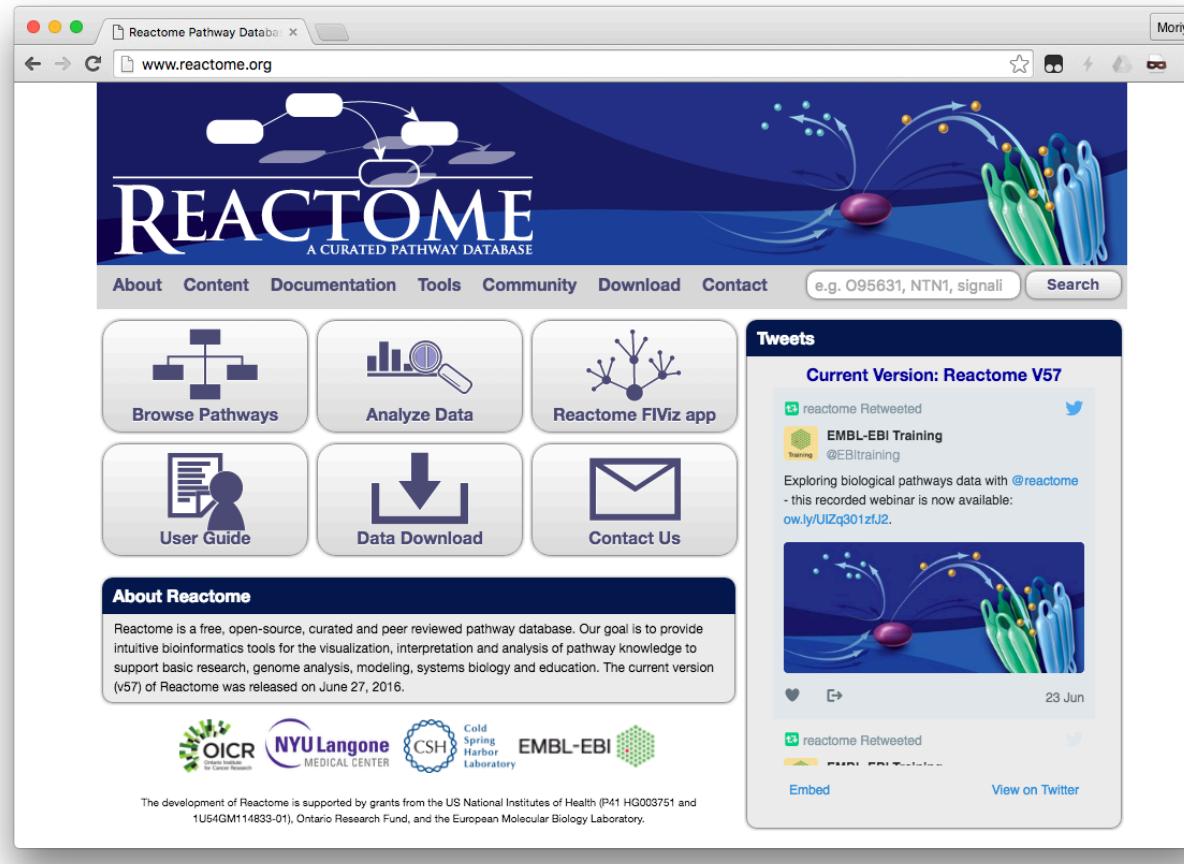
YeastCyc: 酵母のパスウェイDB

# YeastCyc に GEO のデータをマッピング

The screenshot shows the YeastCyc interface for *Saccharomyces cerevisiae* S288c. A central metabolic pathway diagram is displayed, with various nodes and connections. A floating window titled "Omics Viewer Data and Parameters" is open, showing search results for "yeast". The results list includes several GEO datasets, such as GDS5814, GDS5630, GDS3866, GDS3751, GDS3725, GDS3591, GDS3438, GDS3437, GDS3332, and GDS3245. Below the list are options for "Type of display" (Static (single timepoint)), "Use data from" (Single column), and "Select Column" (GSM1232160: wild type strain control, biological rep1). To the right of the diagram, a sidebar titled "OPERATIONS" is visible, with a red circle highlighting the "Import Data from GEO" option under the "Overlay Experimental Data (Omics Viewer)" section.

- GEO のデータをキーワード（例 : yeast）で検索してパスウェイにマッピングできる
- 複数データの比較やアニメーションもできる

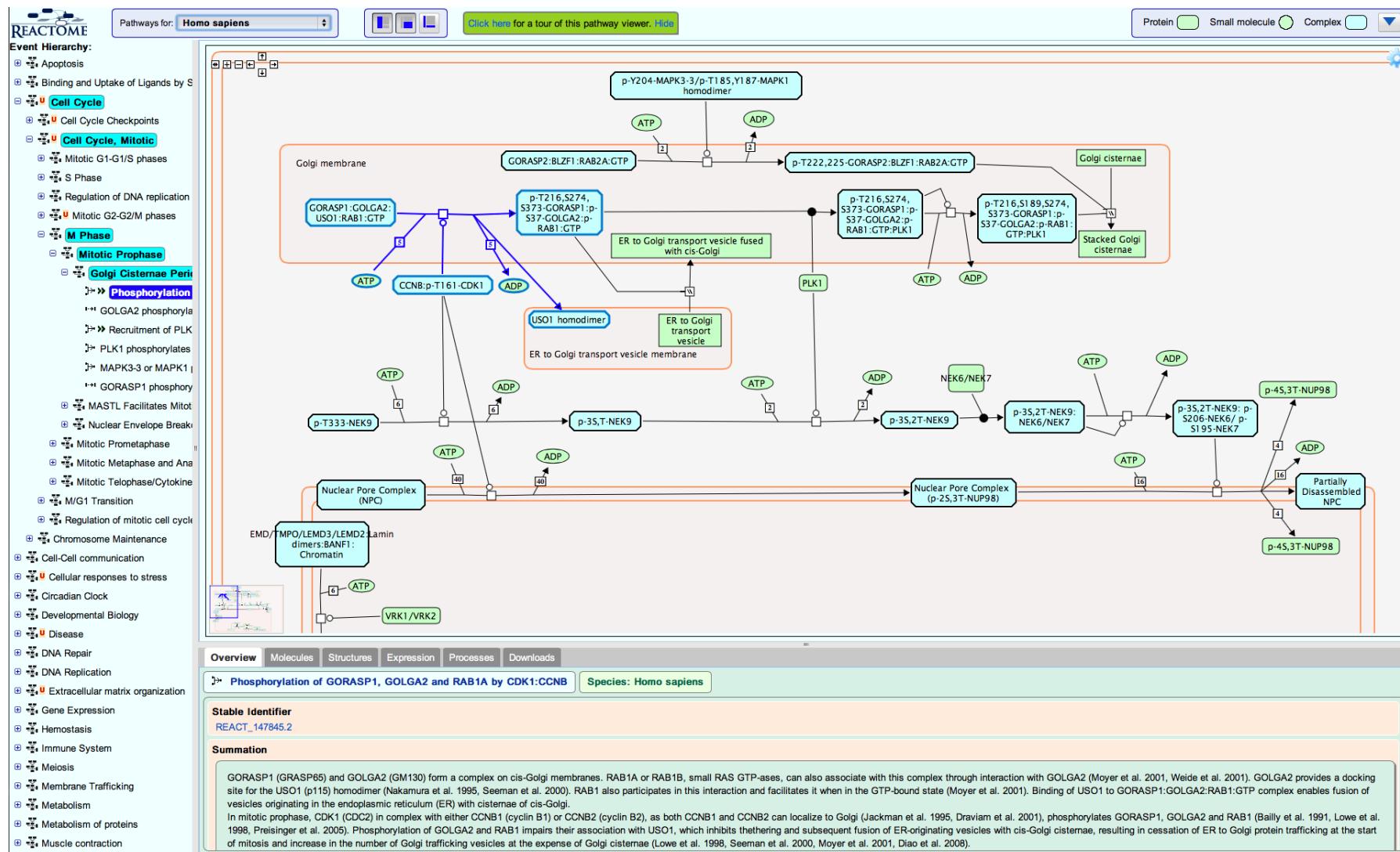
# Reactome



<http://www.reactome.org/>

- 開発：
  - EMBL-EBI 他
- 対象：
  - ヒトを中心に脊椎動物、酵母、植物、19種
- ヒト：
  - 専門家が手作業で作成した文献ベースのデータ
- その他：
  - 計算機での推定
- 代謝パスウェイ、シグナル伝達系、他
- 利用：フリー
- データ形式：
  - BioPAX, SBML

# Reactome



<http://www.reactome.org/>

# パスウェイデータベースのカタログ

## PathGuide

<http://www.pathguide.org/>

The screenshot shows the homepage of PathGuide at [www.pathguide.org](http://www.pathguide.org/). The page title is "Pathguide the pathway resource list". On the left, there is a navigation sidebar with categories like "Navigation", "Search", "Analysis", and "Contact". The main content area displays a "Complete Listing of All Pathguide Resources" with a table of 547 entries. The table includes columns for "Database Name", "Details", "Availability", and "Standards". The "Availability" column is highlighted with red circles around the "Free" and "X" icons. A news box on the right side of the page indicates a "Major new update of Pathguide August 2013".

- 約550のリスト
- Availability
  - 有料か無料か
- Standards
  - 標準データ形式に準拠しているか
- BioPAX
- SBML



- ## パスウェイ解析ツール
- Cytoscape
  - VizANT

## パスウェイデータの分類 (PathGuide)

- タンパク質間相互作用
- 代謝パスウェイ
- シグナリングパスウェイ
- パスウェイダイアグラム
- 転写因子と遺伝子制御ネットワーク
- タンパク質-化合物間相互作用
- 遺伝的相互作用ネットワーク
- アミノ酸配列解析
- その他

# KEGG

ウェブサイト : <http://www.kegg.jp/>

開発 : 京都大学

対象 :

- ゲノムの決まった全生物種（異株を含む）4,200 種以上 (>300真核生物、>3,700真正細菌、>220古細菌)
- 真核ドラフトゲノム 25種（なくなる予定）
- 環境メタゲノム 300サンプル、生体メタゲノム 700サンプル

専門家が手作業で作成した文献ベースのデータ + 自動ツール

- リファレンスパスウェイ : 専門家が手作業で文献ベースから作成
- 生物種パスウェイ
- 自動ツールでリファレンスパスウェイから作成し、手作業でキュレーション
- 自動ツールで作成（自動ツールにも段階があります）

代謝パスウェイ、シグナル伝達系、他

利用 : アカデミックフリー

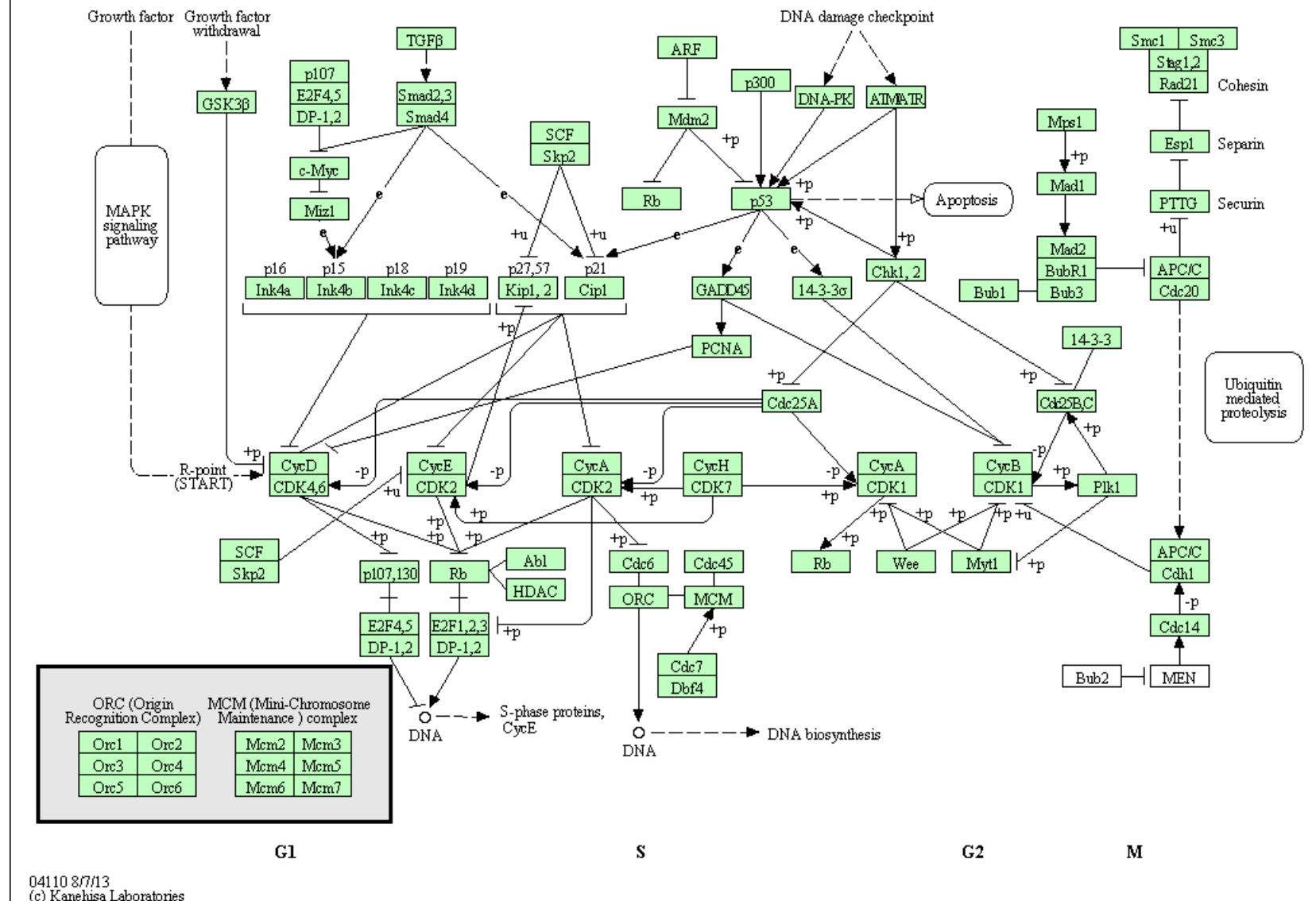
データ形式 : KGML

- KCPAVS KEGG-XML converter などで代謝パスウェイ、シグナル伝達などの多くのパスウェイを標準形式に変換可能
- KEGGscape でネットワーク可視化ソフト Cytoscape に読み込み可能

## KEGG の代謝パスウェイマップ

[http://www.genome.jp/kegg-bin/show\\_pathway?map00020](http://www.genome.jp/kegg-bin/show_pathway?map00020)

# シグナル伝達パスウェイ (KEGG の細胞周期マップの例)



# KEGG

The screenshot shows the KEGG homepage. The left sidebar contains links to KEGG Home, Database, Objects, Software, FTP, GenomeNet, DBGET/LinkDB, Feedback, and Kanehisa Labs. The main content area features the KEGG logo, a search bar, and a link to Japanese. Below this is the title "KEGG: Kyoto Encyclopedia of Genes and Genomes". A text box explains KEGG as a database for biological systems. An announcement states that KEGG RPAIR is discontinued. The main content is organized into sections: "Main entry point to the KEGG web service" (KEGG2, Table of Contents, Update notes), "Data-oriented entry points" (KEGG PATHWAY, BRITE, MODULE, ORTHOLOGY, GENOME, GENES, COMPOUND, GLYCAN, REACTION, ENZYME, DISEASE, DRUG, MEDICUS), "Organism-specific entry points" (KEGG Organisms, Analysis tools like Mapper, Atlas, BlastKOALA, GhostKOALA, BLAST/FASTA, SIMCOMP), and "Analysis tools" (Mapper, Atlas, BlastKOALA, GhostKOALA, BLAST/FASTA, SIMCOMP).

- KEGG: Kyoto Encyclopedia of Genes and Genomes
- ゲノムとパスウェイのデータベース
- 遺伝子
  - →オーソログ
  - →パスウェイ
- 遺伝子の機能分類

<http://www.kegg.jp/kegg/>

<http://www.genome.jp/kegg/>

# KEGG

KEGG GENOME Database  
Organisms and ecosystems with genome sequence information

KEGG2 PATHWAY BRITE MODULE KO GENOME GENES SSDB Organisms

Search GENOME for  Go Clear

**KEGG Organisms**

KEGG GENOME is a collection of **KEGG organisms**, which are the organisms with complete genome sequences and each of which is identified by the three- or four-letter organism code, and additional organisms with draft genome sequences. It now also contains selected viral genomes that are linked from KEGG DISEASE entries. KEGG GENOME is supplemented by MGENOME, a collection of metagenome sequences from environmental samples (ecosystems). The three categories, complete genomes, draft genomes and metagenomes, correspond to the gene catalog databases of GENES, DGENES and MGENES, respectively (see KEGG GENES).

Category	DBGET (genome)	Identifiers	DBGET (genes)	Annotation
Complete genomes (KEGG Organisms)	GENOME	T0 numbers / three- or four-letter organism codes	GENES	KOALA/manual
Viral genomes <small>New!</small>	GENOMES	T4 numbers	DGENES	BlastKOALA
Draft genomes		T1 numbers		
Metagenomes		T3 numbers		

Enter KEGG organism code or use Organism button to select  
Organism  Go Clear (Examples) hsa mmu dme cel ath sce eco bsu syn mja

- Complete genomes
  - 4,000以上のゲノム
- Metagenomes
  - 1,000以上のサンプル
- 遺伝子の機能アノテーションとパスウェイ情報

# KEGG

 KEGG  
Kyoto Encyclopedia of  
Genes and Genomes

**KEGG Organisms: Complete Genomes**

Eukaryotes: 333 Bacteria: 3750 Archaea: 229  
[ Genomes | Species | Genus | Draft | Meta ]

**Eukaryotes**

Category	Organisms	Source
Mammals	hsa Homo sapiens (human)	RefSeq
	pu Pan troglodytes (chimpanzee)	RefSeq
	pps Pan paniscus (bonobo)	RefSeq
	ggo Gorilla gorilla gorilla (western lowland gorilla)	RefSeq
	pon Pongo abelii (Sumatran orangutan)	RefSeq
	nle Nomascus leucogenys (northern white-cheeked gibbon)	RefSeq
	mcc Macaca mulatta (rhesus monkey)	RefSeq
	mcf Macaca fascicularis (crab-eating macaque)	RefSeq
	rro Rhinopithecus roxellana (golden snub-nosed monkey)	RefSeq
	cjc Callithrix jacchus (white-tufted-ear marmoset)	RefSeq
	mmu Mus musculus (mouse)	RefSeq
	rno Rattus norvegicus (rat)	RefSeq
	cge Cricetulus griseus (Chinese hamster)	RefSeq
	ngi Nannospalax galili (Upper Galilee mountains blind mole rat)	RefSeq
	hgl Heterocephalus glaber (naked mole rat)	RefSeq
	ocu Oryctolagus cuniculus (rabbit)	RefSeq
	tup Tupaia chinensis (Chinese tree shrew)	RefSeq
	cfa Canis familiaris (dog)	RefSeq
	aml Ailuropoda melanoleuca (giant panda)	RefSeq
	umr Ursus maritimus (polar bear)	RefSeq
	fca Felis catus (domestic cat)	RefSeq
	ptg Panthera tigris altaica (Amur tiger)	RefSeq
	bta Bos taurus (cow)	RefSeq
	bomr Bos mutus (wild yak)	RefSeq
	phd Pantholops hodgsonii (chiru)	RefSeq
	chx Capra hircus (goat)	RefSeq
	oas Ovis aries (sheep)	RefSeq
	ssc Sus scrofa (pig)	RefSeq
	cfr Camelus ferus (Wild Bactrian camel)	RefSeq
	bacu Balaenoptera acutorostrata scammoni (minke whale)	RefSeq
	lve Lipotes vexillifer (Yangtze River dolphin)	RefSeq
	ecb Equus caballus (horse)	RefSeq
	myb Myotis brandtii (Brandt's bat)	RefSeq
	myd Myotis davidii	RefSeq
	nale Pteropus alecto (black flying fox)	RefSeq

Vertebrates

- Complete genomes
  - 4,000以上のゲノム
- 各生物種の名前とコードがリストされている。
- コードをクリックすると生物種の詳細情報に飛ぶ。
- 詳細情報画面からPathway mapへのリンクをたどるとパスウェイのリストが得られる。

# KEGG



**Genome info** **Pathway map** **Brite hierarchy** **Module** **Genome map** **Blast** **Taxonomy**

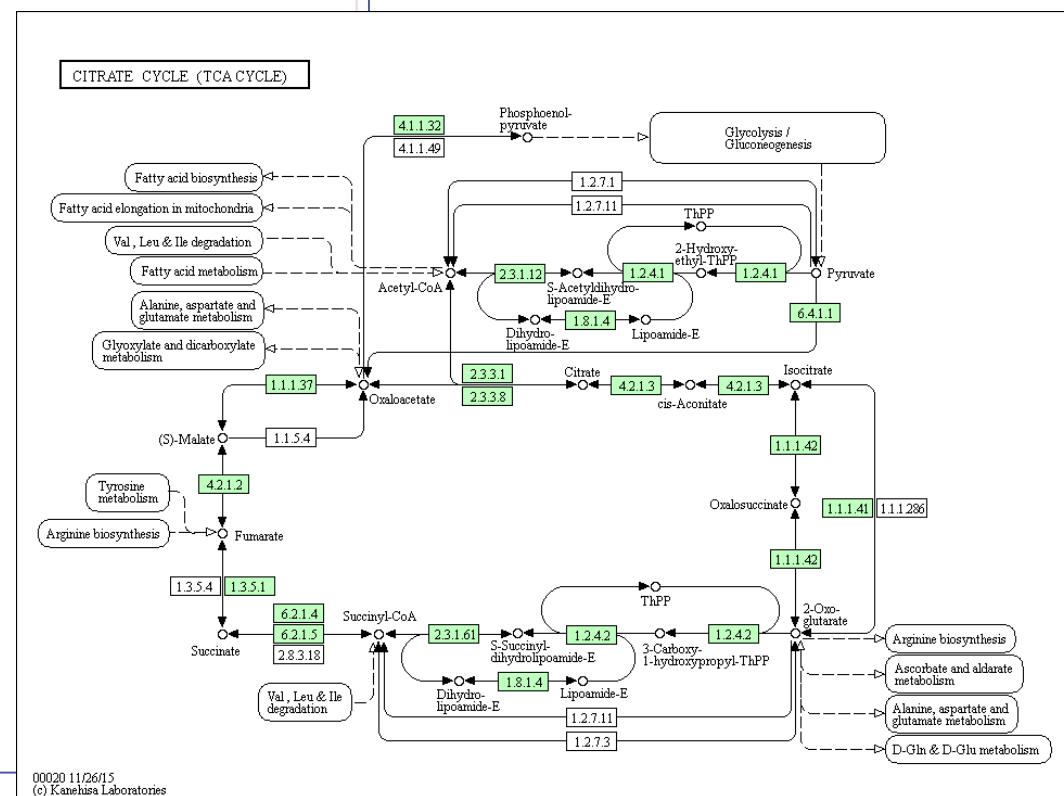
Search genes:  Go Clear KEGG Mapper: Search Pathway Search&Color Pathway

**KEGG pathway maps**

**Metabolism**

- Global and overview maps
  - 01100 Metabolic pathways
  - 01200 Carbon metabolism
  - 01210 2-Oxocarboxylic acid metabolism
  - 01212 Fatty acid metabolism
  - 01230 Biosynthesis of amino acids
- Carbohydrate metabolism
  - 00010 Glycolysis / Gluconeogenesis
  - 00020 Citrate cycle (TCA cycle)
  - 00030 Pentose phosphate pathway
  - 00040 Pentose and glucuronate interconversions
  - 00051 Fructose and mannose metabolism
  - 00052 Galactose metabolism
  - 00053 Ascorbate and aldarate metabolism
  - 00050 Starch and sucrose metabolism
  - 000520 Amino sugar and nucleotide sugar metabolism
  - 000620 Pyruvate metabolism
  - 000630 Glyoxylate and dicarboxylate metabolism
  - 000640 Propanoate metabolism
  - 000650 Butanoate metabolism
  - 000562 Inositol phosphate metabolism
- Energy metabolism
  - 00190 Oxidative phosphorylation
  - 00910 Nitrogen metabolism
  - 00920 Sulfur metabolism
- Lipid metabolism
  - 00061 Fatty acid biosynthesis
  - 00062 Fatty acid elongation
  - 00071 Fatty acid degradation
  - 00072 Synthesis and degradation of ketone bodies
  - 00100 Steroid biosynthesis
  - 00120 Primary bile acid biosynthesis
  - 00140 Steroid hormone biosynthesis
  - 00561 Glycerolipid metabolism
  - 00564 Glycerophospholipid metabolism
  - 00565 Ether lipid metabolism
  - 00600 Sphingolipid metabolism
  - 00590 Arachidonic acid metabolism
  - 00591 Linoleic acid metabolism
  - 00592 alpha-Linolenic acid metabolism
  - 01040 Biosynthesis of unsaturated fatty acids
- Nucleotide metabolism
  - 00230 Purine metabolism
  - 00240 Pyrimidine metabolism
- Amino acid metabolism
  - 00250 Alanine, aspartate and glutamate metabolism
  - 00260 Glycine, serine and threonine metabolism
  - 00270 Cysteine and methionine metabolism

- Pathway maps
  - 約200のマップ
  - リファレンスに対して色付けすることで生物種毎に再構築。



## アノテーションの比較

- ◆ KEGG GENOME database から
  - <http://www.genome.jp/kegg/genome.html>
  - KEGG Mapping for Genome Comparison and Combination で KEGG でアノテーションされたゲノムを選ぶ。
- ◆ サンプルは KEGG Metagenomes で調べることができる。
  - [http://www.genome.jp/kegg/catalog/org\\_list3.html](http://www.genome.jp/kegg/catalog/org_list3.html)
  - MGENES データベースの Metagenomes のリンクをたどっても行けます。
  - 各サンプルから得られた遺伝子配列に対して機能アノテーションをした結果を見ることができる。

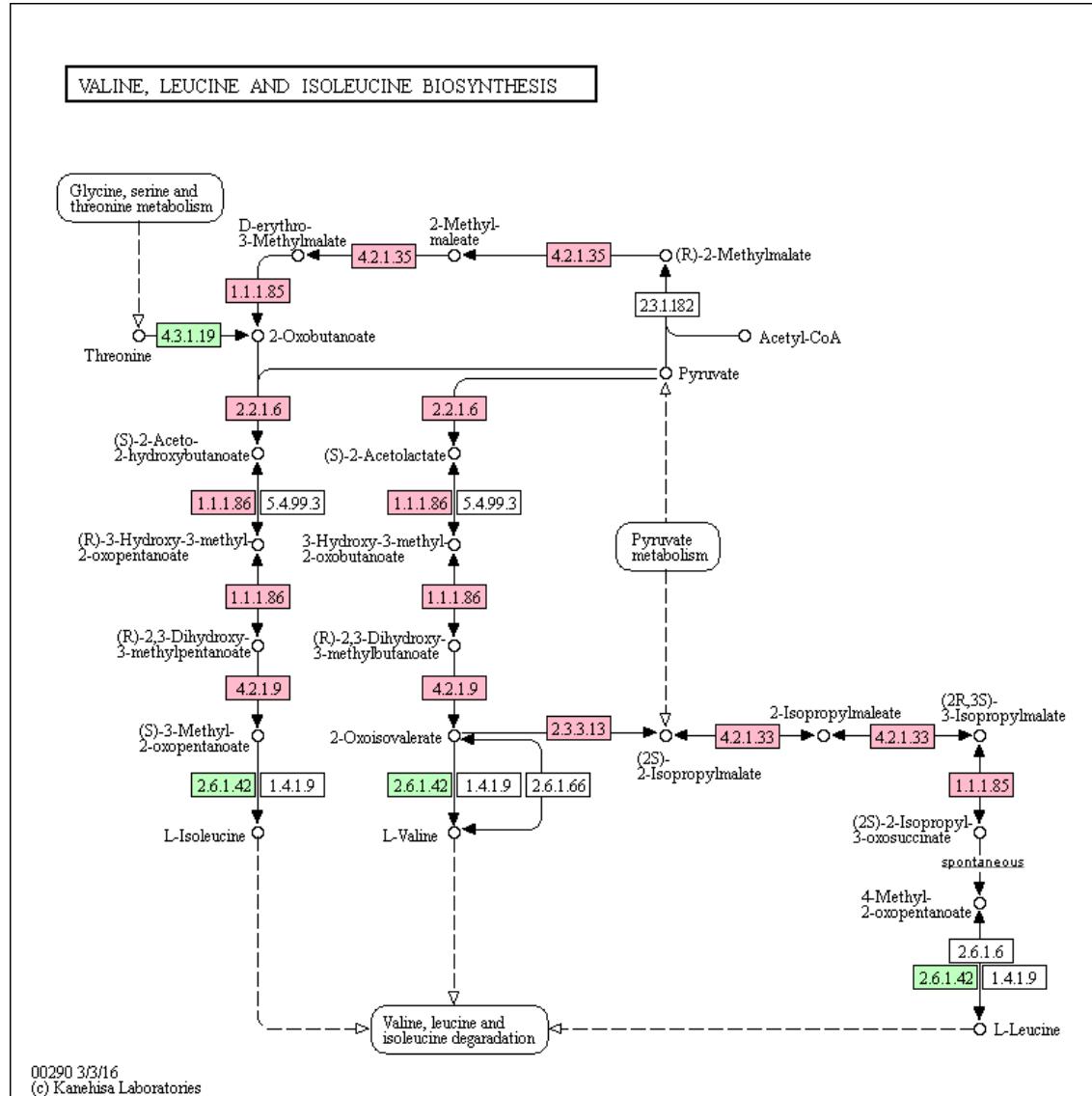
# アノテーションの比較

The screenshot shows the KEGG GENOME Database homepage. At the top, there is a logo for "KEGG GENOME Database" and a subtext "Organisms and ecosystems with genome sequence information". Below the header is a navigation bar with links: KEGG2, PATHWAY, BRITE, MODULE, KO, GENOME, GENES, SSDB, and Organisms. A search bar is present with the placeholder "Search GENOME for" and buttons for "Go" and "Clear". The main content area is titled "KEGG Organisms" and contains a brief description of what KEGG GENOME is. Below this is a section titled "KEGG Mapping for Genome Comparison and Combination". It explains that an organism group can be defined to compare or combine KEGG organisms, other organisms, and environmental samples. It includes a form to "Define organism group (enter organism codes or T numbers)" with a "Go" button. A list of predefined genome comparisons follows:

- (Genome comparison)  
hsa eco Homo sapiens (human) vs. Escherichia coli K-12 strain  
such as [Terpenoid backbone biosynthesis](#)  
where green for human, red for E.coli, and split for both
- (Host-microbiome relationship)  
hsa+T30003 Homo sapiens (human) + gut microbiome  
such as [Global metabolism map](#)  
where green for human, red for gut microbiome, and blue for both
- (Host-symbiont relationship)  
T10018+mlo Lotus japonicus (lotus) + Mesorhizobium loti
- api+buc Acyrthosiphon pisum (pea aphid) + Buchnera aphidicola
- bmy+wbm Brugia malayi (filaria) + Wolbachia
- (Host-vector-pathogen relationship)  
hsa+aga+pfa Homo sapiens (human) + Anopheles gambiae (mosquito) + Plasmodium falciparum

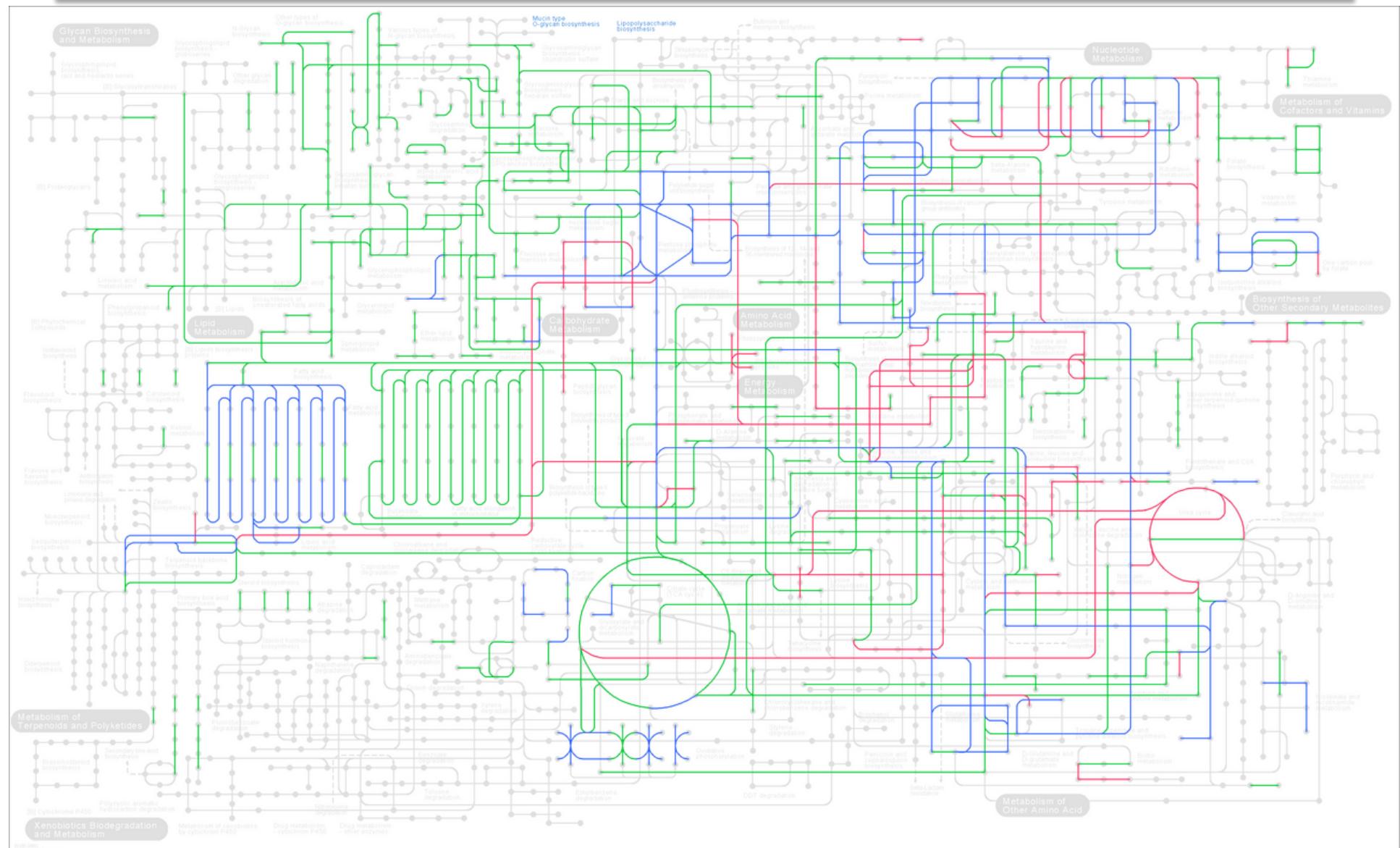
- 複数の生物種を指定することによって比較することができる。
- 例を使って見てみよう。
  - api+buc のアミノ酸合成系
- やり方は p.33~40 も参照

# 複数生物種の情報を PATHWAY にマッピング



- 共生関係
- 寄生関係
- ヒトと腸内細菌叢など
- アブラムシ  
– 昆虫
- ブフネラ  
– アブラムシの共生細菌

# 複数生物種の情報を PATHWAY にマッピング



アブラムシ ブフネラ 共通

# MGENES

**KEGG MGENES Database**  
A collection of genes from large scale metagenomics studies

**KEGG GENES GENOME Blast**

Search MGENES Go Clear  
Browse annotated metagenome samples in KEGG Metagenomes. Go  
Compare samples (enter T numbers, e.g. T30839 T30841 T30844). Go Clear

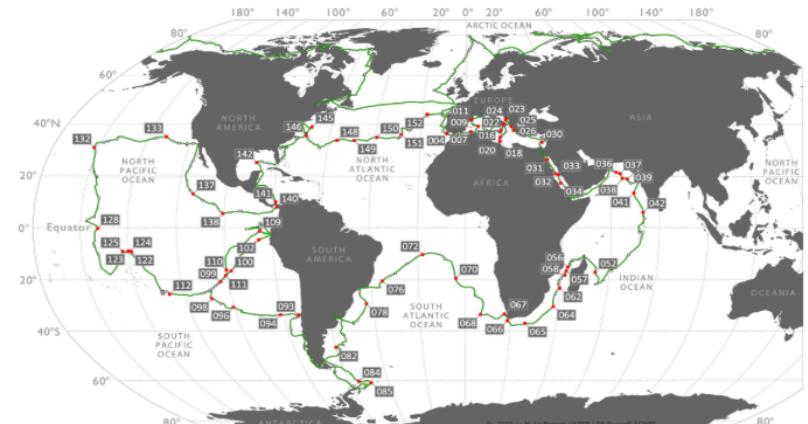
**MGENES**

KEGG MGENES is a collection of gene catalogs from large scale environmental sequencing studies. Sequence and annotation data were generated from publicly available resources with the use of recently developed GhostKOALA. MGENES links sequences, functions and environmental parameters based on the KEGG reference pathway database. MGENES also provides publicly available reference gene catalogs (e.g., OM-RGC, IGC) generated from metagenomes as well as complete genomes with the use of sequence clustering methods.

- **Metagenomes:** Annotated environmental gene catalog
- **RefGene:** Reference gene catalogs from OM-RGC and IGC
- **BLAST:** Blast search to MGENES and RefGene
- **Download sequence data:**  
The annotated sequence data are accessible from the GenomeNet FTP server [README] [FTP].

**Tara Oceans Project**

The GenomeNet and the KEGG teams are actively involved in the integration of the sequence data and metadata generated by the international-interdisciplinary *Tara Oceans* Consortium.



The route of the Tara Oceans Expedition 2009-2013.  
(c) N. Le Bescot / EPEP / SB Roscoff / CNRS

- メタゲノムのコレクション
- サンプルごとにアノテーションとパスウェイ再構築

<http://www.genome.jp/mgenes/>

# MGENES

KEGG GENOME Database  
Organisms and ecosystems with genome sequence information

KEGG2 PATHWAY BRITE MODULE KO GENOME GENES SSDB Organisms

Search GENOME for  Go Clear

**KEGG Organisms**

KEGG GENOME is a collection of **KEGG organisms**, which are the organisms with complete genome sequences and each of which is identified by the three- or four-letter organism code, and additional organisms with draft genome sequences. It now also contains selected viral genomes that are linked from KEGG DISEASE entries. KEGG GENOME is supplemented by MGENOME, a collection of metagenome sequences from environmental samples (ecosystems). The three categories, complete genomes, draft genomes and metagenomes, correspond to the gene catalog databases of GENES, DGENES and MGENES, respectively (see KEGG GENES).

Category	DBGET (genome)	Identifiers	DBGET (genes)	Annotation
Complete genomes (KEGG Organisms)	GENOME	T0 numbers / three- or four-letter organism codes	GENES	KOALA/manual
Viral genomes <i>New!</i>		T4 numbers		
Draft genomes	MGENOME	T1 numbers	DGENES	BlastKOALA
Metagenomes		T3 numbers		

Enter KEGG organism code or use Organism button to select  
Organism  Go Clear (Examples) hsa mmu dme cel ath sce eco bsu syn mja

- Complete genomes
  - 4,000以上のゲノム
- Metagenomes
  - 1,000以上のサンプル
- 遺伝子の機能アノテーションとパスウェイ情報



KEGG Metagenomes

**Environmental: 327** (Tara: 243, GOS: 80, Others: 4)

**Organismal: 712** (Oral: 374, Airway: 30, Gut: 276, Urogenital: 7, Skin: 25)

[ Genomes | Species | Draft | ESTs | Meta ]

## **Environmental samples**

Category	Project			Source
Tara	T30798	4_DCM_0d2-1d6	Tara Oceans sample - 4_DCM_0d2-1d6	Tara Oceans
	T30799	4_SRF_0d2-1d6	Tara Oceans sample - 4_SRF_0d2-1d6	Tara Oceans
		SampleSiteNo_SampleSiteDepth_FilterSize		Tara Oceans
	T30801	1_SRF_0d2-1d6	Tara Oceans sample - 1_SRF_0d2-1d6	Tara Oceans
	T30802	9_DCM_0d2-1d6	Tara Oceans sample - 9_DCM_0d2-1d6	Tara Oceans
	T30803	9	SampleSiteNo: 4~156, 69 サイト	
	T30804	11	SampleSiteDepth: SRF < DCM < MES	
	T30805	11		
	T30806	11	FilterSize: フィルター孔のサイズ	
	T30807	11	0~0d2: 0~0.22μm, ウィルス(ファージ)	
	T30808	2		
	T30809	2	0d2~1d6: 0.22~1.6μm, バクテリア、巨大ウィルス	
	T30810	2		
	T30811	2	0d8~3: 0.8~3μm, バクテリア、単細胞真核	
	T30812	25_DCM_0.2-1d2	Tara Oceans sample - 25_DCM_0.2-1d2	Tara Oceans

# SampleSiteNo\_SampleSiteDepth\_FilterSize

100004 | 7\_071\_002-100 1010 00000 sample = 7\_071\_002-100

T30802 9 DCM\_0d2-1d6\_Tara\_Oceans sample - 9 DCM\_0d2-1d6

SEARCHED INDEXED SERIALIZED FILED 10-15-66 10 AM 1966

T30803 9. SampleSiteNo: 4~156\_69 サ1

SampleSize: 1 150, 000

SampleSiteDenth: SRF < DCM

T30805 1) SampleSiteDepth: SKI < DEM

FilterSize: フィルタ=孔のサイズ

# FilterSize. フィルタ 孔のサイズ

T30807 | 0.042; 0.022 μm 白 /

T30808 2

0.12 1.16 0.22 1.6

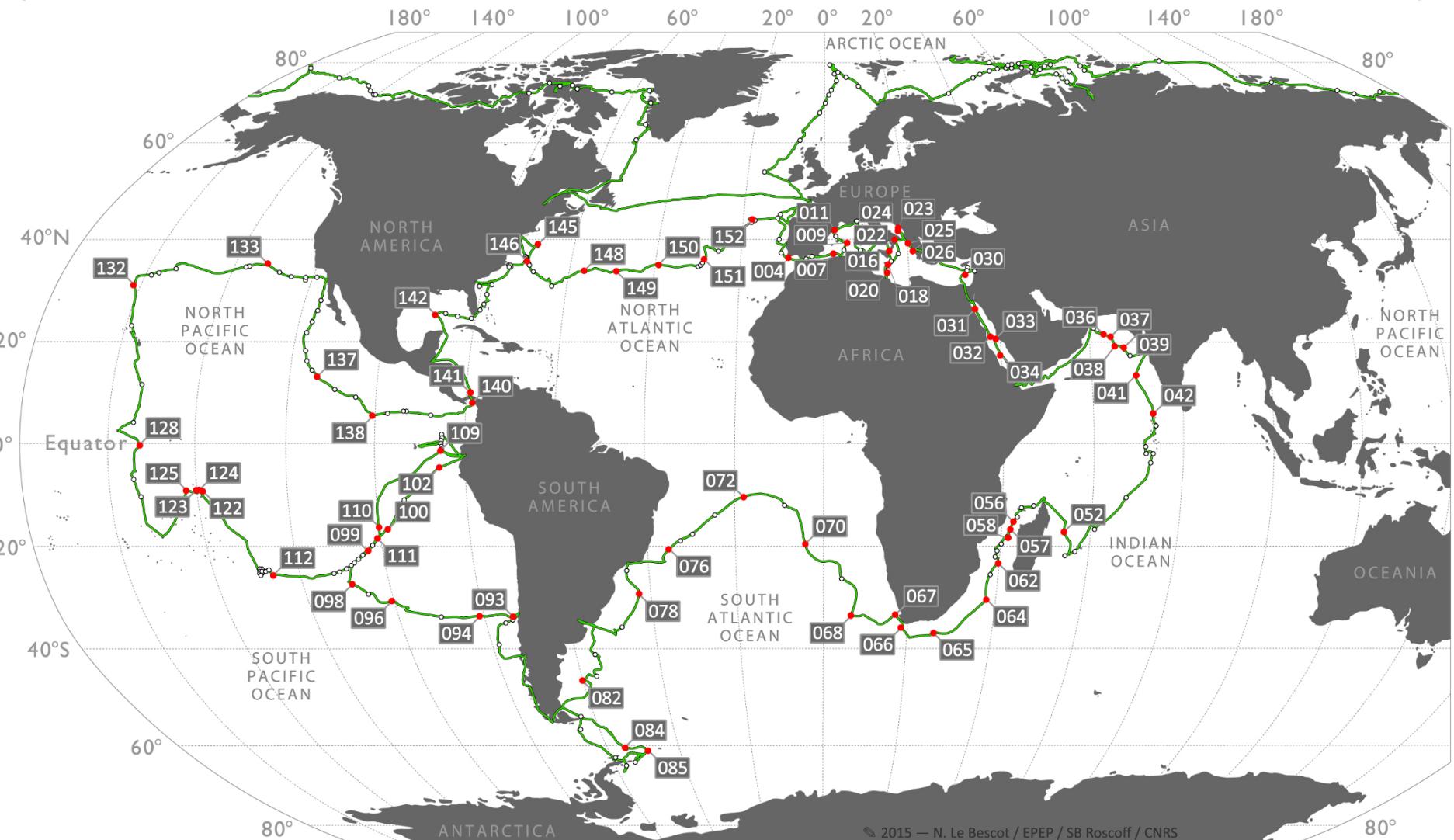
T30809 | 2 0d2~1d6: 0.22~1.6μm,

0d8~3: 0.8~3μm バクテ

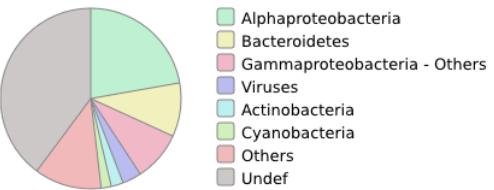
T30811 2

Digitized by srujanika@gmail.com

# サンプル地点



The route of the Tara Oceans Expedition 2009-2013.  
(c) N. Le Bescot / EPEP / SB Roscoff / CNRS

T number	T30798																		
Org code	T30798																		
Aliases	4_DCM_0d2-1d6																		
Full name	T30798																		
Definition	Tara Oceans sample - 4_DCM_0d2-1d6																		
Annotation	GhostKOALA																		
Comment	<pre> "Station"="4_DCM" "Station.number"="4" "Depth.label"="DCM" "Filter.size"="0d2-1d6" "Region"="NAO" "Mean.Date"="9/15/09 16:45" "Mean.Lat"="36.573" "Mean.Long"="-6.538667" "Mean.Pressure_Digiquartz [db]"="45" "Mean.Depth [salt water, m]"="38.689029" "Mean.Temperature [ITS-90, deg C]"="16.246331" "Mean.Salinity [PSU]"="36.556443" "Mean.Oxygen, adjusted [umol/Kg]"="NA" "Mean.Nitrates, adjusted [umol/L]"="NA" "Mean.Chloro.HPLC adjusted [mg Chl/m3]"="0.880307" "Mean.Angular scattering coef, at 117, 470 nm, adj in situ dark [m-1 Sr-1]"="NA" "Mean.bbp470, adj in situ dark [m-1]"="NA" "Mean.Fcdom, factory [ppb QSE]"="NA" "Mean.beam attenuation coef, 660 nm, adjusted deep zero [m-1]"="NA" "Mean.Flux_150m"="NA" "Mean Depth MLD Sigma (m)"="3.5" "Mean Depth Max Fluo (m)"="38" "Mean Depth Max N2 (m)"="17.5" "Mean Depth Max O2 (m)"="NA" "Mean Depth Min O2 (m)"="NA" "Mean Depth Nitracline (m)"="NA" "FLAG SEASON 1-4 (Spring-Summer-Autumn-Winter)"="2" "SEASON PART 1-3 (early middle late)"="3" "NO2 [umol/L]"="NA" "PO4 [umol/L]"="NA" "NO2NO3 [umol/L]"="NA" "SI [umol/L]"="NA" "Okubo-Weiss"="0.007355" "Lyapunov_exp."="0.023383" "grad_SST_adv"="NA" "retention"="0" "AMODIS:CDOM"="1.15753" "AMODIS:PAR8d,Einstains/m-2/d-1"="37.643" "AMODIS:PARm,Einstains/m-2/d-1"="35.829" "NPP_8d_(VGPM)_[mgC/m2/day]"="-4406.835" "NPP_month_(VGPM)_[mgC/m2/day]"="-4330.89" "total.hetero"="1226212.235" "total.auto"="56383.4666" "total.bacteria"="1282595.701" "total.pico.euk"="6889.938377" </pre>																		
Data source	Tara																		
Original DB	Tara																		
Other DBs	EBI-ENA: ERR598950 ERR599095 PANGAEA: TARA_X00000368																		
Statistics	<p>Number of nucleotides: 400933560</p> <p>Number of protein genes: 725424</p> <p>Phylum distribution:</p>  <table border="1"> <thead> <tr> <th>Phylum</th> <th>Percentage</th> </tr> </thead> <tbody> <tr> <td>Alpha proteobacteria</td> <td>~15%</td> </tr> <tr> <td>Bacteroidetes</td> <td>~10%</td> </tr> <tr> <td>Gammaproteobacteria - Others</td> <td>~8%</td> </tr> <tr> <td>Viruses</td> <td>~5%</td> </tr> <tr> <td>Actinobacteria</td> <td>~3%</td> </tr> <tr> <td>Cyanobacteria</td> <td>~2%</td> </tr> <tr> <td>Others</td> <td>~1%</td> </tr> <tr> <td>Undef</td> <td>~45%</td> </tr> </tbody> </table>	Phylum	Percentage	Alpha proteobacteria	~15%	Bacteroidetes	~10%	Gammaproteobacteria - Others	~8%	Viruses	~5%	Actinobacteria	~3%	Cyanobacteria	~2%	Others	~1%	Undef	~45%
Phylum	Percentage																		
Alpha proteobacteria	~15%																		
Bacteroidetes	~10%																		
Gammaproteobacteria - Others	~8%																		
Viruses	~5%																		
Actinobacteria	~3%																		
Cyanobacteria	~2%																		
Others	~1%																		
Undef	~45%																		

- メタゲノムのコレクションで T 番号をクリックした結果
- サンプルごとのアノテーションとパスウェイ再構築
  - 環境因子の情報
  - サンプルから得られた遺伝子（タンパク質をコードする）
  - 遺伝子のアノテーションに基づく生物種の分布

## 演習

- 海洋サンプル (*Tara Oceans Project*) から得られたメタゲノムの機能アノテーション結果を使って、サンプル間にどのような違いがあるかを調べてみよう。
  - 2つ以上の海洋サンプルデータ（ゲノムでもよい）を選び、KEGG パスウェイの階層分類を比較することによって、異なる機能をリストアップする。
  - サンプルには地点の違いや深さの違いなど多くの環境因子が付随しています。例えば深さであれば、
    - SRF (Surface) : 表層
    - DCM (Deep Chlorophyll Maximum) : 数10m
    - MES (Mesopelagic) : 数100m

# アノテーションの比較

The screenshot shows the KEGG GENOME Database interface. At the top, there's a logo and the title "KEGG GENOME Database". Below it, a search bar says "Organisms and ecosystems with genome sequence information". A navigation bar includes links for KEGG2, PATHWAY, BRITE, MODULE, KO, GENOME, GENES, SSDB, and Organisms. A search bar below the navigation bar has fields for "Search" (set to "GENOME"), "for", "Go", and "Clear". The main content area is titled "KEGG Mapping for Genome Comparison and Combination". It explains that an organism group can be defined to compare or combine KEGG organisms, other organisms, and environmental samples. Below this, a form asks "Define organism group (enter organism codes or T numbers):" with a text input containing "T30841 T30844" and a "Go" button. A large blue arrow points from the text input down to a list of pre-defined groups. The list includes:

- (Genome comparison)  
hsa eco Homo sapiens (human) vs. Escherichia coli K-12 strain such as Terpenoid backbone biosynthesis where green for human, red for E.coli, and split for both
- (Host-microbiome relationship)  
hsa+T30003 Homo sapiens (human) + gut microbiome such as Global metabolism map where green for human, red for gut microbiome, and blue for both
- (Host-symbiont relationship)  
T10018+mlo Lotus japonicus (lotus) + Mesorhizobium loti
- api+buc Acyrthosiphon pisum (pea aphid) + Buchnera aphidicola
- bmy+wbm Brugia malayi (filaria) + Wolbachia pipientis
- (Host-vector-pathogen relationship)  
hsa+aga+pfa Homo sapiens (human) + Anopheles gambiae falciparum

A blue callout box at the bottom right contains the Japanese text: "ここに比較したいゲノムやサンプルのT番号を入れて Go!".

- 複数のメタゲノムサンプルを指定するとサンプル間の比較ができる。

# アノテーションの比較

**KEGG** Organism group: T30841 T30844

Category info      **Pathway map**      Brite hierarchy      Taxonomy

Search genes:  Go Clear

**T30841 T30844**

T30841	T30841	38_MES_0d2-1d6; Tara Oceans sample - 38_MES_0d2-1d6
T30844	T30844	38_SRF_0d2-1d6; Tara Oceans sample - 38_SRF_0d2-1d6

KEGG2    PATHWAY    BRITE    GENES    GENOME    LIGAND    DISEASE    DRUG

# Pathway map



Organism group: T30841 T30844

Category info

Pathway map

Brite hierarchy

Taxonomy

Search genes:  Go Clear

## KEGG pathway maps

### Metabolism

#### Global and overview maps

- 01100 Metabolic pathways
- 01110 Biosynthesis of secondary metabolites
- 01120 Microbial metabolism in diverse environments
- 01130 Biosynthesis of antibiotics
- 01200 Carbon metabolism
- 01210 2-Oxocarboxylic acid metabolism
- 01212 Fatty acid metabolism
- 01230 Biosynthesis of amino acids
- 01220 Degradation of aromatic compounds

#### Carbohydrate metabolism

- 00010 Glycolysis / Gluconeogenesis
  - 00020 Citrate cycle (TCA cycle)
  - 00030 Pentose phosphate pathway
  - 00040 Pentose and glucuronate interconversions
  - 00051 Fructose and mannose metabolism
  - 00052 Galactose metabolism
  - 00053 Ascorbate and aldarate metabolism
  - 00500 Starch and sucrose metabolism
  - 00520 Amino sugar and nucleotide sugar metabolism
  - 00620 Pyruvate metabolism
  - 00630 Glyoxylate and dicarboxylate metabolism
  - 00640 Propanoate metabolism
  - 00650 Butanoate metabolism
  - 00660 C5-Branched dibasic acid metabolism
  - 00562 Inositol phosphate metabolism
- Energy metabolism
- 00190 Oxidative phosphorylation
  - 00195 Photosynthesis
  - 00196 Photosynthesis - antenna proteins
  - 00710 Carbon fixation in photosynthetic organisms
  - 00720 Carbon fixation pathways in prokaryotes
  - 00680 Methane metabolism
  - 00910 Nitrogen metabolism
  - 00920 Sulfur metabolism
- Lipid metabolism

グローバルマップ

生物種1：緑、生物種2：赤、共通：青  
\* 3生物種以上は未サポート

個別マップ

生物種1：薄緑、生物種2：薄赤、、、、  
\* 10生物種までサポート

ここでは生物種とサンプルはどちらも  
遺伝子の集合と考えて下さい。

# Pathway map



Organism group: T30841 T30844

Category info      Pathway map      Brite hierarchy      Taxonomy

Search genes:  Go Clear

**KEGG pathway maps**

**Metabolism**

Global and overview maps

- 01100 Metabolic pathways
- 01110 Biosynthesis of secondary metabolites
- 01120 Microbial metabolism in diverse environments
- 01130 Biosynthesis of antibiotics
- 01200 Carbon metabolism
- 01210 2-Oxocarboxylic acid metabolism
- 01212 Fatty acid metabolism
- 01230 Biosynthesis of amino acids
- 01220 Degradation of aromatic compounds

Carbohydrate metabolism

- 00010 Glycolysis / Gluconeogenesis
- 00020 Citrate cycle (TCA cycle)
- 00030 Pentose phosphate pathway
- 00040 Pentose and glucuronate interconversions
- 00051 Fructose and mannose metabolism
- 00052 Galactose metabolism
- 00053 Ascorbate and aldarate metabolism
- 00500 Starch and sucrose metabolism
- 00520 Amino sugar and nucleotide sugar metabolism
- 00620 Pyruvate metabolism
- 00630 Glyoxylate and dicarboxylate metabolism
- 00640 Propanoate metabolism
- 00650 Butanoate metabolism
- 00660 C5-Branched dibasic acid metabolism
- 00562 Inositol phosphate metabolism

Energy metabolism

- 00190 Oxidative phosphorylation
- 00195 Photosynthesis
- 00196 Photosynthesis - antenna proteins
- 00710 Carbon fixation in photosynthetic organisms
- 00720 Carbon fixation pathways in prokaryotes
- 00680 Methane metabolism
- 00910 Nitrogen metabolism
- 00920 Sulfur metabolism
- Linid metabolism

**KEGG module**

Energy metabolism

- M00165 Reductive pentose cycle
- M00166 Reductive pentose cycle
- M00167 Reductive pentose cycle
- M00168 CAM (Crassulacean Acid Metabolism)
- M00169 CAM (Crassulacean Acid Metabolism)
- M00172 C4-dicarboxylic acid metabolism
- M00171 C4-dicarboxylic acid metabolism
- M00173 Reductive citrate cycle
- M00579 Phosphate acetyltransferase
- M00175 Nitrogen fixation
- M00567 Methanogenesis
- M00174 Methane oxidation
- M00176 Assimilatory sulfate reduction
- M00596 Dissimilatory sulfate reduction
- M00595 Thiosulfate oxidation

Carbohydrate and lipid metabolism

- Central carbohydrate metabolism
- M00001 Glycolysis (Embden-Meyerhof-Parnet pathway)
- M00002 Glycolysis, core
- M00003 Gluconeogenesis
- M00307 Pyruvate oxidation
- M00009 Citrate cycle (TCA cycle)
- M00110 Citrate cycle, first step
- M00011 Citrate cycle, second step
- M00004 Pentose phosphate cycle
- M00005 Pentose phosphate cycle
- M00006 Pentose phosphate cycle
- M00007 Pentose phosphate cycle
- M00580 Pentose phosphate cycle
- M00008 PRPP biosynthesis
- M00308 Entner-Doudoroff pathway
- M00633 Semi-phosphoryl transferases
- M00309 Non-phosphoryl transferases
- Other carbohydrate metabolism
- M00012 Glyoxylate cycle
- M00013 Malonate semialdehyde dehydrogenase
- M00632 Galactose dehydrogenase

Metabolic pathways - Reference pathway + T30841 T30844

[ Pathway menu | Organism menu | Pathway entry | Hide module list | User data mapping | Image (png) file ]

Reference pathway Go 35%

Help

# Pathway map

# アノテーションの比較

**KEGG** Organism group: T30841 T30844

Category info	Pathway map	Brite hierarchy	Taxonomy
Search genes: <input type="text"/> Go Clear			
<b>T30841 T30844</b>			
T30841 T30841 38_MES_0d2-1d6; Tara Oceans sample - 38_MES_0d2-1d6			
T30844 T30844 38_SRF_0d2-1d6; Tara Oceans sample - 38_SRF_0d2-1d6			

KEGG2 PATHWAY BRITE GENES GENOME LIGAND DISEASE DRUG

- ◆ Pathway map は K 番号のアノテーション（対応する遺伝子の有無）マップを一つずつ開いてみないと分からない。
- ◆ Brite hierarchy では機能分類全体を階層的に表示してくれるので、全体を一通りチェックするために使う。

# Brite hierarchy

**KEGG** Organism group: T30841 T30844

Category info   Pathway map   Brite hierarchy   Taxonomy

Search genes:  Go Clear

**BRITE functional hierarchies**

- 00001 KEGG Orthology (KO)
- 00003 KEGG reaction modules
- 01000 Enzymes
- 01001 Protein kinases
- 01009 Protein phosphatases and associated proteins
- 01002 Peptidases
- 01003 Glycosyltransferases
- 01005 Lipopolysaccharide biosynthesis proteins
- 01004 Lipid biosynthesis proteins
- 01008 Polyketide biosynthesis proteins
- 01006 Prenyltransferases
- 01007 Amino acid related enzymes
- 00199 Cytochrome P450
- 00194 Photosynthesis proteins
- 03000 Transcription factors
- 03021 Transcription machinery
- 03019 Messenger RNA biogenesis
- 03041 Spliceosome
- 03011 Ribosome
- 03009 Ribosome biogenesis
- 03016 Transfer RNA biogenesis
- 03012 Translation factors
- 03110 Chaperones and folding catalysts
- 04131 SNAREs
- 04121 Ubiquitin system
- 03051 Proteasome
- 03032 DNA replication proteins
- 03036 Chromosome
- 03400 DNA repair and recombination proteins
- 03029 Mitochondrial biogenesis
- 02000 Transporters
- 02044 Secretion system
- 02042 Bacterial toxins
- 02022 Two-component system
- 02035 Bacterial motility proteins
- 04812 Cytoskeleton proteins
- 04147 Exosome
- 02048 Prokaryotic defense system
- 04030 G protein-coupled receptors
- 01020 Enzyme-linked receptors
- 04050 Cytokine receptors
- 03310 Nuclear receptors
- 04040 Ion channels
- 04031 GTP-binding proteins
- 04052 Cytokines
- 04516 Cell adhesion molecules and their ligands
- 04090 Cellular antigens
- 00535 Proteoglycans
- 00536 Glycosaminoglycan binding proteins
- 04091 Lectins

KEGG Orthology ID (K 番号) のうちパスウェイにマッピングされているものの階層分類

KEGG Orthology ID (K 番号) のうちパスウェイにマッピングされていないものも含めたタンパク質ファミリーや複合体の階層分類

# Brite hierarchy



Organism group: T30841 T30844

Category info    Pathway map    Brite hierarchy    Taxonomy

Search genes:  Go Clear

**BRITE functional hierarchies**

00001 KEGG Orthology (KO)  
00003 KEGG reaction modules

01000 Enzymes  
01001 Protein kinases  
01009 Protein phosphatases and associated proteins  
01002 Peptidases  
01003 Glycosyltransferases  
01005 Lipopolysaccharide biosynthesis proteins  
01004 Lipid biosynthesis proteins  
01008 Polyketide biosynthesis proteins  
01006 Prenyltransferases  
01007 Amino acid related enzymes  
00199 Cytochrome P450  
00194 Photosynthesis proteins  
03000 Transcription factors  
03021 Transcription machinery  
03019 Messenger RNA biogenesis  
03041 Spliceosome  
03011 Ribosome  
03009 Ribosome biogenesis  
03016 Transfer RNA biogenesis  
03012 Translation factors  
03110 Chaperones and folding catalysts  
04131 SNAREs  
04121 Ubiquitin system  
03051 Proteasome  
03032 DNA replication proteins  
03036 Chromosome  
03400 DNA repair and recombination proteins  
03029 Mitochondrial biogenesis  
02000 Transporters  
02044 Secretion system  
02042 Bacterial toxins  
02022 Two-component system  
02035 Bacterial motility proteins  
04812 Cytoskeleton proteins  
04147 Exosome  
02048 Prokaryotic defense system  
04030 G protein-coupled receptors  
01020 Enzyme-linked receptors  
04050 Cytokine receptors  
03310 Nuclear receptors  
04040 Ion channels  
04031 GTP-binding proteins  
04052 Cytokines  
04516 Cell adhesion molecules and their ligands  
04090 Cellular antigens  
00535 Proteoglycans  
00536 Glycosaminoglycan binding proteins  
04091 Lectins



KEGG Orthology (KO) + T30841 T30844

[ Brite menu | Organism menu | Download htext ]

Reference hierarchy (KO)

Go

▼ ▼ ▼ ▼    One-click mode

- ▶ Metabolism
- ▶ Genetic Information Processing
- ▶ Environmental Information Processing
- ▶ Cellular Processes
- ▶ Organismal Systems
- ▶ Human Diseases

[ BRITE | KEGG2 | KEGG ]

Last updated: July 4, 2015

- ▼ 左から大分類→小分類でその階層をすべて開く。
- ▶ その階層の一つ下の分類を開く。One-click mode をチェックすると、下の層全てを開く。

# Brite hierarchy



Organism group: T30841 T30844

Category info    Pathway map    Brite hierarchy    Taxonomy

Search genes:  Go Clear

**BRITE functional hierarchies**

- 00001 KEGG Orthology (KO)**
- 00003 KEGG reaction modules**
- 01000 Enzymes
- 01001 Protein kinases
- 01009 Protein phosphatases and associated proteins
- 01002 Peptidases
- 01003 Glycosyltransferases
- 01005 Lipopolysaccharide biosynthesis proteins
- 01004 Lipid biosynthesis proteins
- 01008 Polyketide biosynthesis proteins
- 01006 Prenyltransferases
- 01007 Amino acid related enzymes
- 00199 Cytochrome P450
- 00194 Photosynthesis proteins
- 03000 Transcription factors
- 03021 Transcription machinery
- 03019 Messenger RNA biogenesis
- 03041 Spliceosome
- 03011 Ribosome
- 03009 Ribosome biogenesis
- 03016 Transfer RNA biogenesis
- 03012 Translation factors
- 03110 Chaperones and folding catalysts
- 04311 SNAREs
- 04121 Ubiquitin system
- 03051 Proteasome
- 03032 DNA replication proteins
- 03036 Chromosome
- 03400 DNA repair and recombination proteins
- 03029 Mitochondrial biogenesis
- 02000 Transporters
- 02044 Secretion system
- 02042 Bacterial toxins
- 02022 Two-component system
- 02035 Bacterial motility proteins
- 04812 Cytoskeleton proteins
- 04147 Exosome
- 02048 Prokaryotic defense system
- 04030 G protein-coupled receptors
- 01020 Enzyme-linked receptors
- 04050 Cytokine receptors
- 03310 Nuclear receptors
- 04040 Ion channels
- 04031 GTP-binding proteins
- 04052 Cytokines
- 04516 Cell adhesion molecules and their ligands
- 04090 Cellular antigens
- 00535 Proteoglycans
- 00536 Glycosaminoglycan binding proteins
- 04091 Lectins

**KEGG Orthology (KO) + T30841 T30844**

[ Brite menu | Organism menu | Download httext ]

Reference hierarchy (KO)    Go

▼ ▼ ▼    One-click mode

▼ Metabolism

► Overview

▼ Carbohydrate metabolism

- 00010 Glycolysis / Gluconeogenesis [PATH:00010]
  - K00844 HK; hexokinase [EC:2.7.1.1]
  - K12407 GCK; glucokinase [EC:2.7.1.2]
  - K00845 glk; glucokinase [EC:2.7.1.2]
  - K01810 GPI, pgk; glucose-6-phosphate isomerase [EC:5.3.1.9]
  - K06859 pgil; glucose-6-phosphate isomerase, archaeal [EC:5.3.1.9]
  - K13810 tal-pgi; transaldolase / glucose-6-phosphate isomerase [EC:2.2.1.2 5.3.1.9]
  - K15916 pgm-pni; glucose/mannose-6-phosphate isomerase [EC:5.3.1.9 5.3.1.8]
  - K00850 pfkA, PFK; 6-phosphofructokinase 1 [EC:2.7.1.11]
  - K16370 pfkB; 6-phosphofructokinase 2 [EC:2.7.1.11]
  - K03841 FBP, fbp; fructose-1,6-bisphosphatase I [EC:3.1.3.11]
  - K02446 glpX; fructose-1,6-bisphosphatase II [EC:3.1.3.11]
  - K1532 glpX-SEBP; fructose-1,6-bisphosphatase II / sedoheptulose-1,7-bisphosphatase [EC:3.1.3.11 3.1.3.37]
  - K01086 fbp-SEBP; fructose-1,6-bisphosphatase I / sedoheptulose-1,7-bisphosphatase [EC:3.1.3.11 3.1.3.37]
  - K04041 fbpI; fructose-1,6-bisphosphatase III [EC:3.1.3.11]
  - K01623 ALDO; fructose-bisphosphate aldolase, class I [EC:4.1.2.13]
  - K11645 fbaB; fructose-bisphosphate aldolase, class I [EC:4.1.2.13]
  - K01624 fbaA; fructose-bisphosphate aldolase, class II [EC:4.1.2.13]
  - K01622 fructose 1,6-bisphosphate aldolase/phosphatase [EC:4.1.2.13 3.1.3.11]
  - K16305 K16305; fructose-bisphosphate aldolase / 6-deoxy-5-ketofructose 1-phosphate synthase [EC:4.1.2.13 2.2.1.11]
  - K16306 K16306; fructose-bisphosphate aldolase / 2-amino-3,7-dideoxy-D-threo-hept-6-ulosonate synthase [EC:4.1.2.13 2.2.1.10]
  - K01803 TP1, tpiA; triosephosphate isomerase (TIM) [EC:5.3.1.1]
  - K00134 GAPDH, gapA; glyceraldehyde 3-phosphate dehydrogenase [EC:1.2.1.12]
  - K0705 GAPDHS; glyceraldehyde-3-phosphate dehydrogenase, spermaticogenic [EC:1.2.1.12]
  - K00150 gap2; glyceraldehyde-3-phosphate dehydrogenase (NAD(P)) [EC:1.2.1.59]
  - K00927 PGK, pgk; phosphoglycerate kinase [EC:2.7.2.3]
  - K01834 PGAM, gpmA; 2,3-bisphosphoglycerate-dependent phosphoglycerate mutase [EC:5.4.2.11]
  - K15633 gpmI; 2,3-bisphosphoglycerate-independent phosphoglycerate mutase [EC:5.4.2.12]
  - K15634 gpmA; probable phosphoglycerate mutase [EC:5.4.2.12]
  - K15635 apgM; 2,3-bisphosphoglycerate-independent phosphoglycerate mutase [EC:5.4.2.12]
  - K01689 ENO, eno; enolase [EC:4.2.1.11]
  - K00873 FK, pyk; pyruvate kinase [EC:2.7.1.40]
  - K12406 PKLR; pyruvate kinase isozymes R/L [EC:2.7.1.40]
  - K00163 aceF; pyruvate dehydrogenase E1 component [EC:1.2.4.1]
  - K00161 FDHA, pdha; pyruvate dehydrogenase E1 component alpha subunit [EC:1.2.4.1]
  - K00162 FDHB, pdhb; pyruvate dehydrogenase E1 component beta subunit [EC:1.2.4.1]
  - K00627 DLAT, aceF, pdhc; pyruvate dehydrogenase E2 component (dihydrolipoamide acetyltransferase) [EC:2.3.1.12]
  - K00382 DLD, lpd, pdhd; dihydrolipoamide dehydrogenase [EC:1.8.1.4]
  - K13997 PDHX; dihydrolipoamide dehydrogenase-binding protein of pyruvate dehydrogenase complex
  - K00169 porA; pyruvate ferredoxin oxidoreductase delta subunit [EC:1.2.7.1]
  - K00170 porB; pyruvate ferredoxin oxidoreductase beta subunit [EC:1.2.7.1]
  - K00171 porD; pyruvate ferredoxin oxidoreductase delta subunit [EC:1.2.7.1]
  - K00172 porC; pyruvate ferredoxin oxidoreductase gamma subunit [EC:1.2.7.1]
  - K00016 Ldh, ldh; L-lactate dehydrogenase [EC:1.1.1.27]
  - K01568 E4.1.1.1, pdc; pyruvate decarboxylase [EC:4.1.1.1]
  - K13951 ADH1\_7; alcohol dehydrogenase 1/7 [EC:1.1.1.1]
  - K13980 ADH4; alcohol dehydrogenase 4 [EC:1.1.1.1]
  - K00121 frmA, ADHS, adhc; S-(hydroxymethyl)glutathione dehydrogenase / alcohol dehydrogenase [EC:1.1.1.284 1.1.1.1]
  - K13952 ADH6; alcohol dehydrogenase 6 [EC:1.1.1.1] ...

色付けはパスウェイと同じ

## 演習

- 海洋サンプル (*Tara Oceans Project*) から得られたメタゲノムの機能アノテーション結果を使って、サンプル間にどのような違いがあるかを調べてみよう。
  - 2つ以上の海洋サンプルデータ（ゲノムでもよい）を選び、KEGG パスウェイの階層分類を比較することによって、異なる機能をリストアップする。
  - 例：表層と深海、光合成、窒素固定、エネルギー代謝、脂質代謝
- KEGG Mapper を使えば、同じことが任意の遺伝子名とK番号のリストからできます。
  - [http://www.genome.jp/kegg/tool/map\\_pathway.html](http://www.genome.jp/kegg/tool/map_pathway.html)

## 演習

- 海洋サンプル (*Tara Oceans Project*) から得られたメタゲノムの機能アノテーション結果を使って、サンプル間にどのような違いがあるかを調べてみよう。
- MGENES (T 番号) の比較では
  - 遺伝子の有無の情報のみ
  - それらがどの程度の量あるかは分からない
- 元のデータには各遺伝子にマッピングできたリード数の情報がある
  - 遺伝子の量を考えると差があるかもしれない。
  - <http://goto.kuicr.kyoto-u.ac.jp/lecture/data/> にサンプルごとの遺伝子頻度情報があります。

# KEGG Mapper – Color Pathway

**KEGG Mapper – Color Pathway**

About KEGG Mapper  
Search Pathway  
Search&Color Pathway  
Color Pathway  
Color Pathway WebGL  
  
Search Brite  
Search&Color Brite  
Join Brite  
  
Search Module  
Search&Color Module  
  
Annotate Sequence  
BlastKOALA  
  
Reconstruct Pathway  
Reconstruct Brite  
Reconstruct Module  
Map Taxonomy  
  
KEGG Atlas  
KEGG

Select KEGG pathway map: ko00195  
Enter file name containing the data: T30844.txt  
File type:  Numerical value ( Convert to log scale)  
Gradation:  minimum-maximum  negative-zero-positive  
Minimum : #ffff00 Maixmum : #ff0000 Check  
  
 Use uncolored diagram  
Exec Clear

**Color Pathway** is an extension of the Search&Color Pathway tool, allowing multiple coloring of a selected pathway map. Multiple coloring of bgcolor (or bgcolor,fgcolor), which is specified in tab-delimited columns, may represent time-dependent or site-dependent changes of gene expressions, pathways active in different cancer stages, etc., and is shown as multiple snapshots of image maps. Alternatively, numerical values may be given in multiple columns and they are displayed by gradation and optionally 3D representation as well. The objects in different types of pathway maps are specified by the following KEGG identifiers.

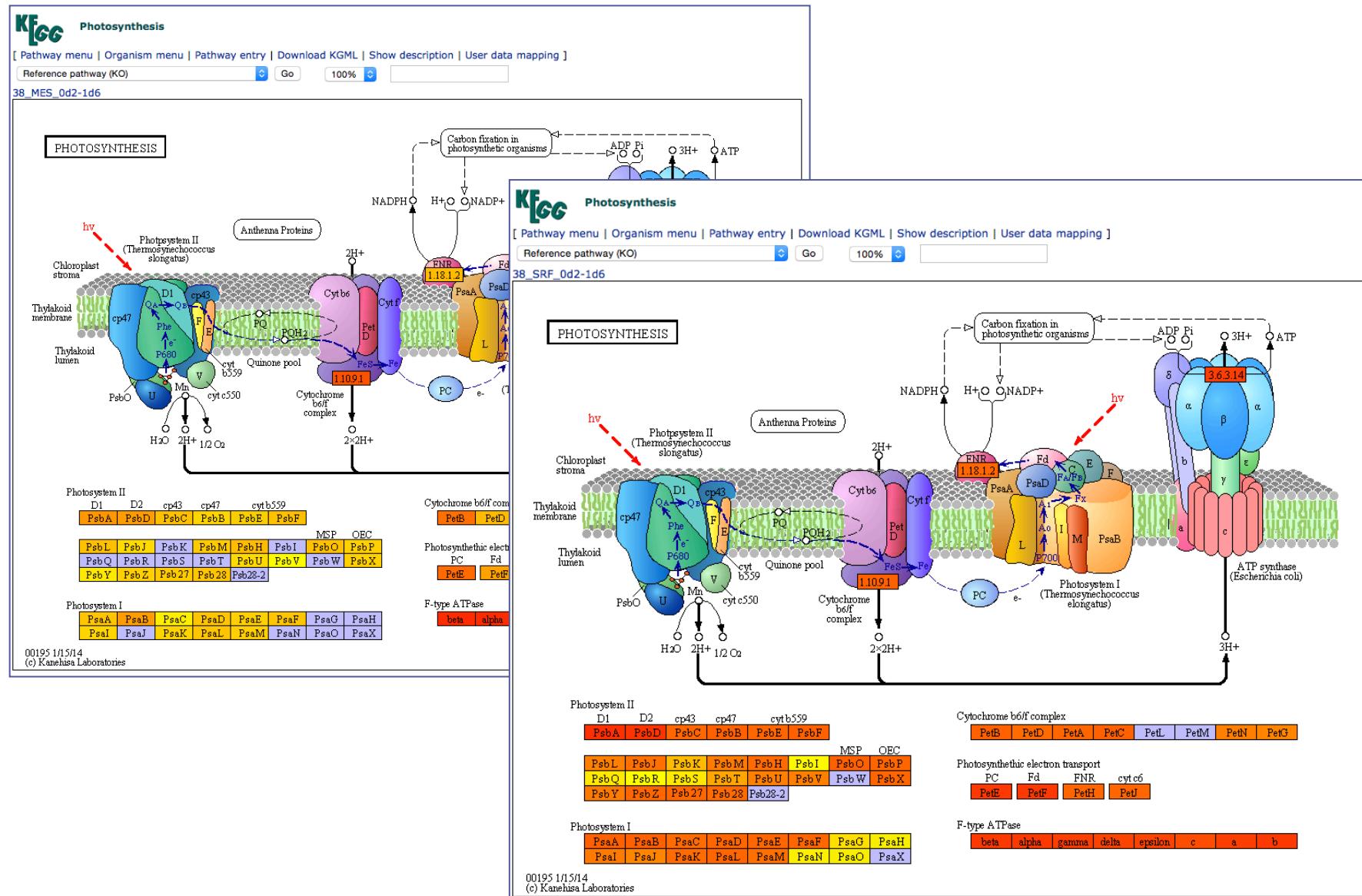
Prefix	Type	KEGG identifier
map	Reference pathway	K number
ko	Reference pathway (KO)	K number
org	Organism-specific pathway	gene identifier

Last updated: July 24, 2014

[http://www.genome.jp/kegg/tool/map\\_pathway3.html](http://www.genome.jp/kegg/tool/map_pathway3.html)  
指定したパスウェイの色付け

遺伝子または KO とその発現量とマップ番号を入力し、指定した色のグラデーションで色付けする

# KEGG Mapper – Color Pathway



# まとめ

- ・ パスウェイデータベースの概要
  - ・ パスウェイデータに触れてみた
- ・ データのパスウェイへのマッピング1
  - ・ BioCyc と Reactome でのオミックス解析
- ・ データのパスウェイへのマッピング2
  - ・ KEGG でのオミックス解析