

パスウェイデータベースの紹介

奥田修二郎

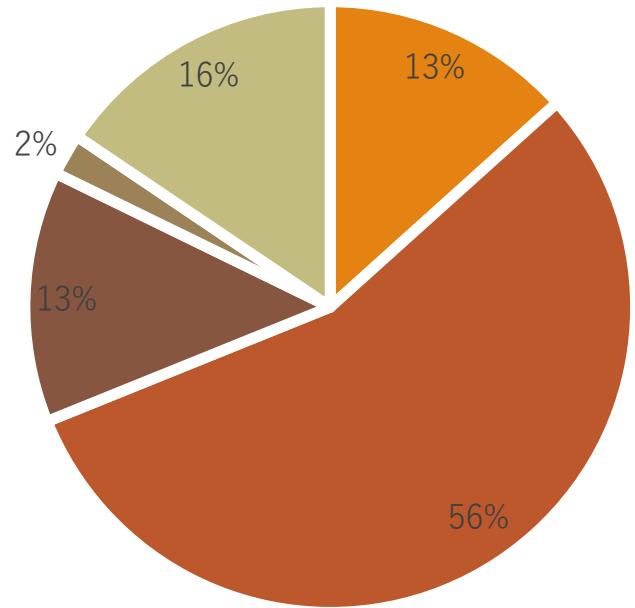
新潟大学大学院医歯学総合研究科

AJACS越後@新潟大学理学部マルチメディア教室

2018年6月27日

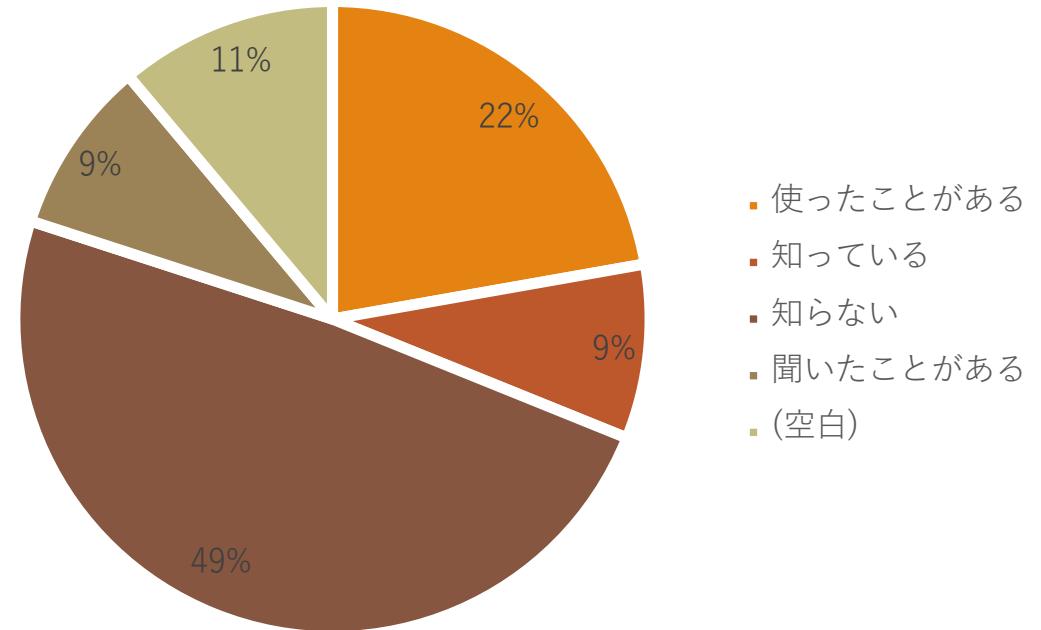
事前アンケート

【設問8-1】 パスウェイ関連のデータベースを使ったことがありますか。



- 今後使いたい
- 使う予定はすぐにはないが使い方は覚えたい
- 使ったことがある
- 日常的に使っている
- (空白)

【設問8-2】 KEGGを使ったことがありますか。



- 使ったことがある
- 知っている
- 知らない
- 聞いたことがある
- (空白)

パスウェイデータベースとは

- パスウェイとは、生体内での遺伝子やタンパク質、その他の化合物等の分子間相互作用を "経路" として表現したものです。相互作用の知識を集積して可視化、電子化したものがパスウェイデータベースと呼ばれています。歴史的には、代謝経路の表現から始まりました
 - ベーリンガーマンハイム社（現ロシュ社）の代謝マップ (1965~)
 - 最初は紙媒体での出版。のちに電子化されています
- 計算機上で表現することで、様々な可視化が行えるようになり、パスウェイ全体を俯瞰したり、一部の相互作用に注目したり、生命現象の理解が行い易くなります。また、データベース化することで網羅的に扱えるようになり、コンピューターでの利用が可能になります
 - ゲノムアノテーションや種間比較、進化解析
 - 遺伝子発現などのエンリッチメント解析
 - モデル化、シミュレーション、予測

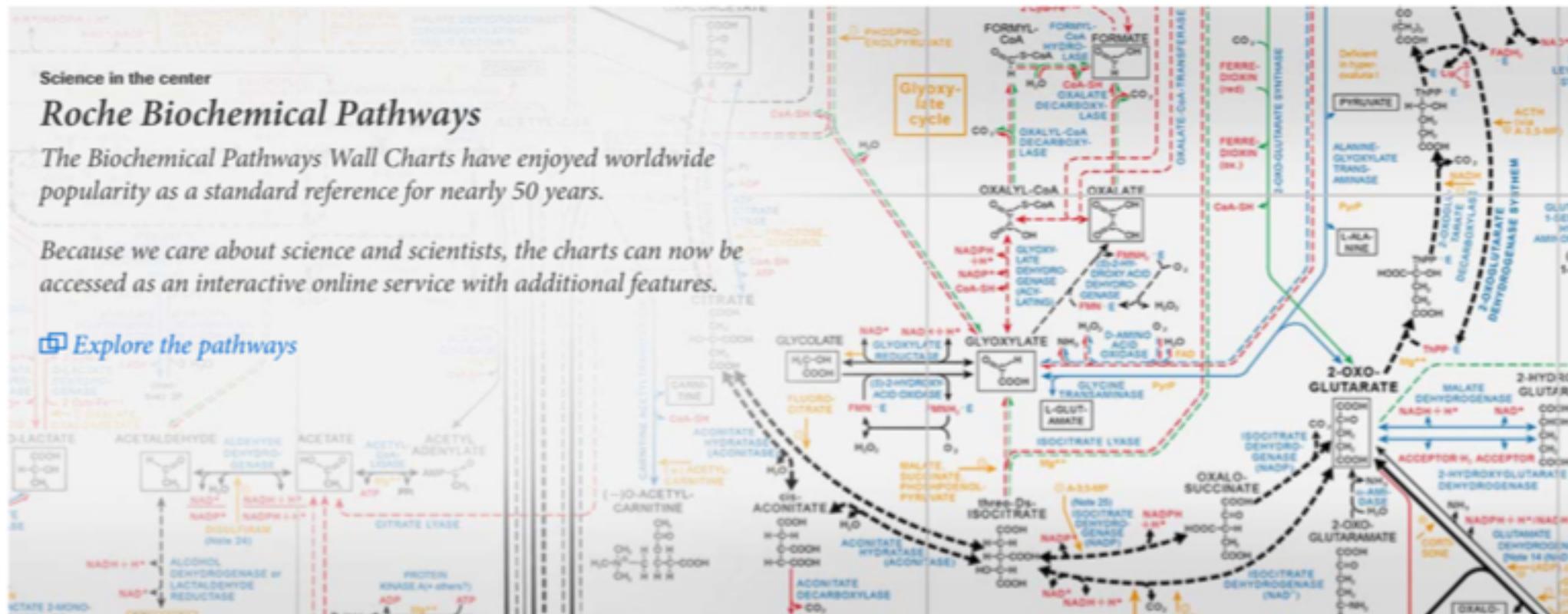
Science in the center

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


[Share](#) [Rate](#)

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.

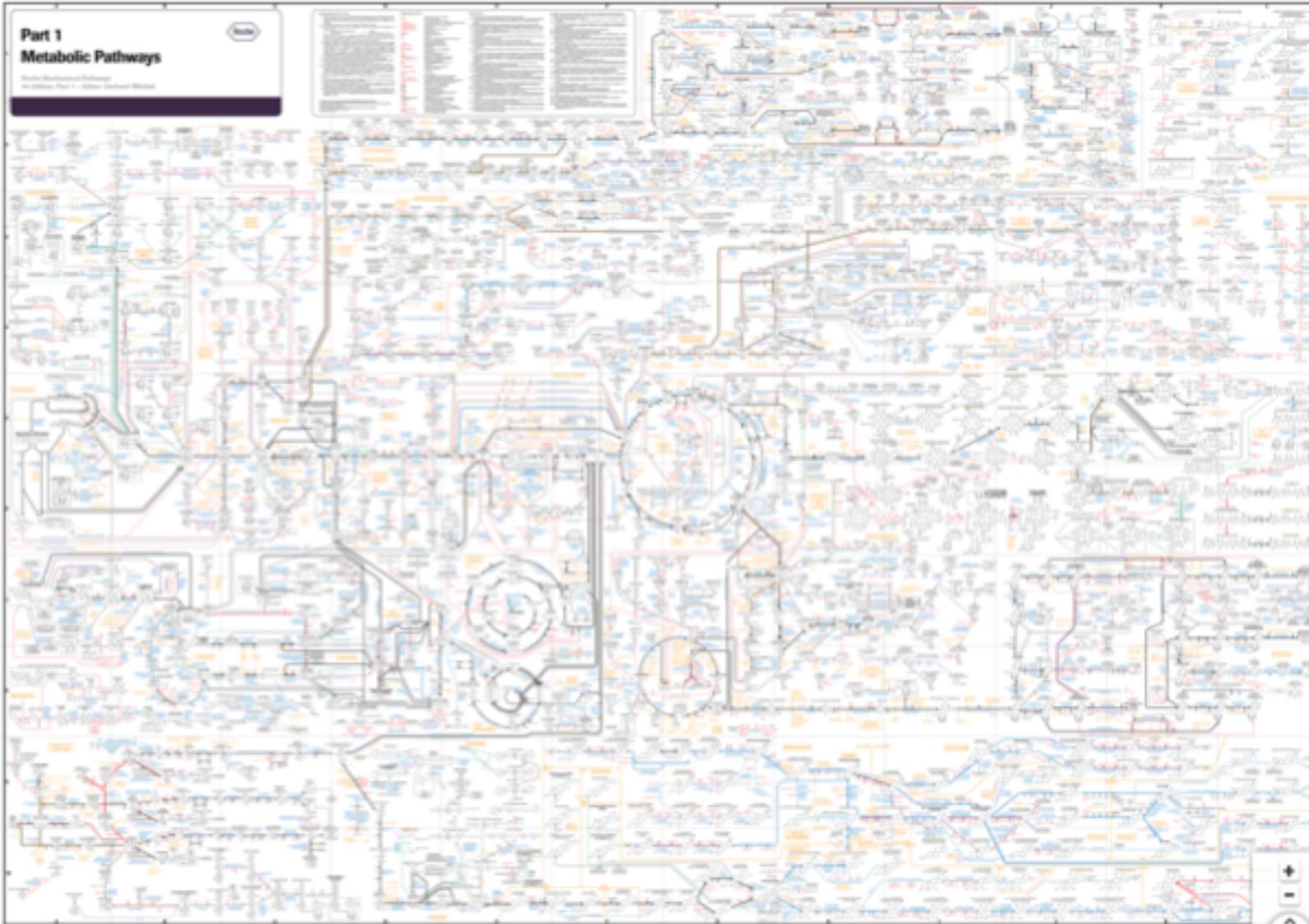
You have to be someone with tenacity and patience. And love for science

Dr Gerhard Michal
Editor of the Roche Biochemical Pathways

By the numbers

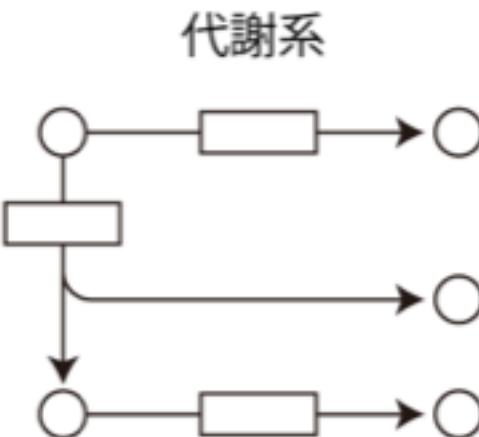
49 years
continuously fine-tuned by the editor himself.

Click here to open the viewer

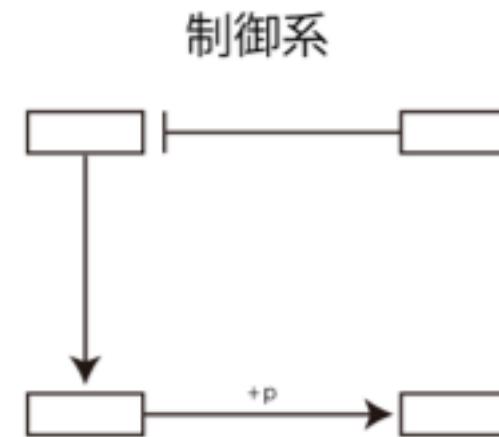


パスウェイデータベースでの表現

- 見やすくするために、ダイアグラムで表現されていることが多いっています。
- KEGG Pathway での表現例
 - 代謝パスウェイでは代謝産物をノード、酵素反応をエッジとして表現されています
 - 制御系ではタンパク質や遺伝子、その他の小分子をノード、その関係性（活性化、抑制、リン酸化など）がエッジとして表現されています



代謝産物(○)と酵素(□)の関係



タンパク質や遺伝子(□)間の関係

パスウェイデータベースのデータ形式

- 計算機での取り扱いを目的として XML (Extensible Markup Language) で記述されていることが多いっています。
 - [KGML](#) (KEGG Markup Language) は分子間の関係とダイアグラムのレイアウトを取り扱うための KEGG 独自のフォーマット
 - [SBML](#) (Systems Biology Markup Language)、[CellML](#)、[CSML](#) (Cell System Markup Language) はパスウェイのシミュレーションやモデリングを行うためのフォーマット
 - [PSI-MI](#) (Proteomics Standards Initiative Molecular Interaction XML Format) はタンパク質間相互作用を記述するためのフォーマット
 - [BioPAX](#) (Biological Pathways Exchange) は様々なパスウェイデータを統合したり、データ交換を行うことを目的として策定された標準化を目指したフォーマット
- これらのデータ形式を扱うことのできるネットワーク可視化ソフトウェアには [Cytoscape](#) や [VisANT](#) などがあります。
 - [AJACS58 : Cytoscapeを使ったデータの可視化](#)
- どのパスウェイデータベースを研究に使えば良いかは、対象生物や対象パスウェイ、目的によって異なってきます。
- 今回は BioCyc, Reactome, KEGG PATHWAY のブラウザ上での使い方を紹介します。

BioCyc

- ウェブサイト：<https://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

The screenshot shows the BioCyc Database Collection homepage. At the top right, there are links for "Log In", "My Log In", and "Create New Account". Below that is a search bar with the placeholder "Search for gene, protein, metabolite or pathway". To the right of the search bar are buttons for "Save Search" and "New Search". The main content area has a header "BioCyc Database Collection". It features a "RouteSearch" section with two diagrams: one for "Long Route" and another for "Short Route: Atoms Conserved". The "RouteSearch" section also includes a "RouteSearch: Search for Paths through the Metabolic Network" link and a "Learn More" button. On the left, there's a "Getting Started" section with a "How to BioCyc? Typical usage is:" list. The list includes instructions for selecting databases, searching for genes/pathways, and using the Quick Search. Below this is a "New User Guide" button. On the right, there's a "Tools" section with a list of BioCyc tools: Genome browser, Display of individual metabolic pathways, Full metabolic maps, Multiple analysis methods, Store groups of genes and pathways, and Comparative analysis tools. Further down are sections for "BioCyc Databases" (with a note about three tiers: Tier 1, Tier 2, and Tier 3), "Funding Sources" (mentioning NIH grant GM080746), and a "Bottom" footer.

Reactome

- ウェブサイト：<https://www.reactome.org/>
- 開発：EMBLE-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
(<https://reactome.org/documentation/inferred-events/>)
 - 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



KEGG PATHWAY

- ウェブサイト：<https://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、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
 - [KCPA VS 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] (<http://apps.cytoscape.org/apps/keggscape>) でネットワーク可視化ソフト Cytoscape に読み込み可能

KEGGはデータベースの集合

- KEGG2をクリック
- KEGG PATHWAY を含むシステム情報データベースの他に、遺伝情報、化学情報、健康情報などのデータベースがリンクしています。

KEGG Home
Release notes
Current statistics
Plea from KEGG

KEGG Database
KEGG overview
Searching KEGG
KEGG mapping
Color codes

KEGG Objects
Pathway maps
Brite hierarchies
KEGG DB links

KEGG Software
KegTools
KEGG API
KGML

KEGG FTP
Subscription

KEGG: Kyoto Encyclopedia of Genes and Genomes

KEGG is a database resource for understanding high-level functions and utilities of the biological system, such as the cell, the organism and the ecosystem, from molecular-level information, especially large-scale molecular datasets generated by genome sequencing and other high-throughput experimental technologies. See [Release notes](#) (January 1, 2018) for new and updated features.

Announcement: KEGG NETWORK released

[Point to the KEGG web service](#)

KEGG2 KEGG Table of Contents [Update notes | Release history]

Data-oriented entry points

KEGG PATHWAY	KEGG pathway maps
KEGG BRITE	BRITE hierarchies and tables
KEGG MODULE	KEGG modules
KEGG ORTHOLOGY	KO functional orthologs [Annotation]
KEGG GENOME	Genomes [Pathogen Virus Plant]
KEGG GENES	Genes and proteins [SeqData]
KEGG COMPOUND	Small molecules
KEGG GLYCAN	Glycans

Classification

Pathway
Brite
Brite table
Module
KO (Function)
Organism
Compound
Network

Menu PATHWAY BRITE MODULE KO GENES LIGAND NETWORK DISEASE DRUG DBGET

Search KEGG for Go

Data-oriented entry points

Category	Entry Point	Content	DBGET Search
Systems information	KEGG PATHWAY KEGG BRITE KEGG MODULE	KEGG pathway maps BRITE hierarchies and tables KEGG modules	PATHWAY BRITE MODULE
Genomic information	KO (KEGG Orthology) KEGG GENOME KEGG GENES KEGG SSDB	Functional orthologs KEGG organisms (complete genomes) Genes and proteins GENES sequence similarity	ORTHOLOGY GENOME GENES
Chemical information (KEGG LIGAND)	KEGG COMPOUND KEGG GLYCAN KEGG REACTION KEGG ENZYME	Small molecules Glycans Reactions and reaction classes Enzyme nomenclature	COMPOUND GLYCAN REACTION RCLASS ENZYME
Health information	KEGG NETWORK <i>New!</i> KEGG DISEASE KEGG DRUG KEGG ENVIRON KEGG MEDICUS	Disease-related network elements Human diseases Drugs and drug groups Health related substances Japanese drug labels (JAPIC) FDA drug labels (DailyMed)	NETWORK VARIANT DISEASE DRUG DGROUP ENVIRON

対象生物種を見る

- データベースのテーブルの下、KEGG organisms をクリック

Subject-oriented entry points

KEGG Cancer	Cancer research
KEGG Pathogen	Pathogen genomes and antimicrobial resistance
KEGG Virus	Virus research
KEGG Plant	Plant research
KEGG Annotation	KO annotation of genes and proteins
KEGG SeqData	KEGG sequence data collection
KEGG RModule	Architecture of metabolic networks

Organism-specific entry points

[KEGG organisms](#) the list of currently available organisms

Select Go Clear (examples) [hsa](#) [mmu](#) [sce](#) [eco](#) [bsu](#) [syn](#)

Organism group	Pangenome		Genome
Eukaryotes	Animals Plants Fungi Protists	Genus level pangenomes	Species level pangenomes
Prokaryotes	Bacteria Archaea		KEGG organisms

See [KEGG GENOME](#) for more details.

KEGG Organisms

- KEGG では 3-4 文字の独自の生物種コードを使用している

Eukaryotes

Category	Organisms	Source
	hsa Homo sapiens (human)	RefSeq
	ptr 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
	csab Chlorocebus sabaeus (green monkey)	RefSeq
	rro Rhinopithecus roxellana (golden snub-nosed monkey)	RefSeq
	rbb Rhinopithecus bieti (black snub-nosed monkey)	RefSeq
	cjc Callithrix jacchus (white-tufted-ear marmoset)	RefSeq
	sbq Saimiri boliviensis boliviensis (Bolivian squirrel monkey)	RefSeq
	mmu Mus musculus (mouse)	RefSeq
	rno Rattus norvegicus (rat)	RefSeq
	cge Cricetus griseus (Chinese hamster)	RefSeq
	ngi Nannospalax galili (Upper Galilee mountains blind mole rat)	RefSeq
	hgl Heterocephalus glaber (naked mole rat)	RefSeq
	ccan Castor canadensis (American beaver)	RefSeq
	ocu Oryctolagus cuniculus (rabbit)	RefSeq
	tup Tupai chinensis (Chinese tree shrew)	RefSeq

KEGG Organisms

- 生物種コードのリンクをクリックすると、種の情報が表示されます
 - Annotation
 - manual : 手作業によるアノテーション（ヒト(hsa)等）
 - KOALA : SSEARCH ベースの自動ツールによるアノテーション（ゴリラ(ggo)等）
 - BlastKOALA : BLAST ベースの自動ツールによるアノテーション（ドラフトゲノム）
 - GhostKOALA : GhostX ベースの自動ツールによるアノテーション（メタゲノム）

KEGG Homo sapiens (human)

Genome info	Pathway map	Brite hierarchy	Module	Genome map	Blast	Taxonomy
Search genes: <input type="text"/>	<input type="button" value="Go"/>	<input type="button" value="Clear"/>				
Genome information						
T number	T01001					
Org code	hsa					
Aliases	HUMAN, 9606					
Full name	Homo sapiens (human)					
Definition	Homo sapiens (human)					
Annotation	manual					
Taxonomy	View details					
Lineage	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Haplorrhini; Catarrhini; Hominoidea; Homo					
Data source	RefSeq (Assembly: GCF_000001405.37) BioProject: 168					
Original DB	NCBI, OMIM, HGNC, HPRD, Ensembl, Vega, Pharos					
Statistics	Number of protein genes: 20364 Number of RNA genes: 19178					
Created	2000					

KEGG Pan paniscus (bonobo)

Genome info	Pathway map	Brite hierarchy	Module	Genome map	Blast	Taxonomy
Search genes: <input type="text"/>	<input type="button" value="Go"/>	<input type="button" value="Clear"/>				
Genome information						
T number	T02283					
Org code	ppb					
Aliases	9997					
Full name	Pan paniscus (bonobo)					
Definition	Pan paniscus (bonobo)					
Annotation	KOALA					
Taxonomy	View details					
Lineage	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Haplorrhini; Catarrhini; Hominoidea; Pan					
Data source	RefSeq (Assembly: GCF_000258655.2) BioProject: 169343					
Original DB	Max-Planck					
Statistics	Number of protein genes: 20581 Number of RNA genes: 4217					
Created	2012					

データベースリスト

- Genomes : 主に NCBI Refseq、GenBank に登録された生物種
- Species : 異株を一つにまとめたデータベース
- Genus : Genus レベルでまとめたデータベース
- Draft : Genomes に入っていない真核生物
- Meta : メタゲノム



KEGG Organisms: Complete Genomes

Eukaryotes: 429 Bacteria: 4521 Archaea: 263

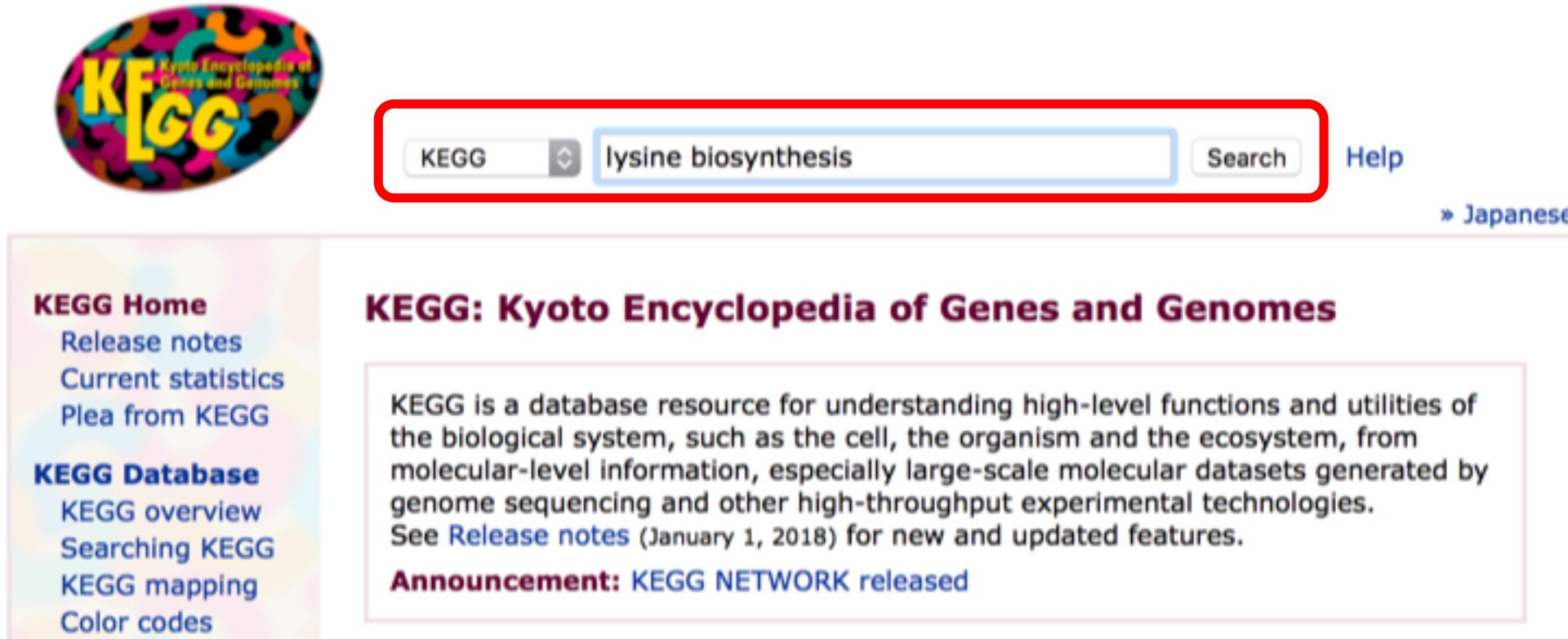
[Genomes | Species | Genus | Viruses | Meta]

Eukaryotes

Category	Organisms		Source
	hsa	Homo sapiens (human)	RefSeq
	ptr	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 leucomoanus (northern white-cheeked gibbon)	RefSeq

KEGGパスウェイマップを見る

- トップページ 上方の検索ボックスで "lysine biosynthesis" や "glycolysis" やなどの生命現象関連の単語を入力し、Search ボタンをクリック



The screenshot shows the KEGG homepage. On the left, there is a navigation sidebar with links to "KEGG Home", "Release notes", "Current statistics", "Plea from KEGG", "KEGG Database", "KEGG overview", "Searching KEGG", "KEGG mapping", and "Color codes". The main content area features the KEGG logo at the top left. Below it is a search bar with the text "lysine biosynthesis" in the input field, which is highlighted with a red rectangle. To the right of the search bar are buttons for "Search" and "Help", and a link to "» Japanese". The main heading "KEGG: Kyoto Encyclopedia of Genes and Genomes" is displayed prominently. A text box below the heading provides a brief description of KEGG's purpose and resources.

KEGG is a database resource for understanding high-level functions and utilities of the biological system, such as the cell, the organism and the ecosystem, from molecular-level information, especially large-scale molecular datasets generated by genome sequencing and other high-throughput experimental technologies. See [Release notes](#) (January 1, 2018) for new and updated features.

Announcement: KEGG NETWORK released

KEGGパスウェイマップ

- KEGG データベース全体でヒットしたエントリーが全てリストアップされ、KEGG PATHWAY にヒットがあれば、一番上に表示されます。

KEGG Search KEGG for lysine biosynthesis Go Clear

Database: KEGG - Search term: lysine biosynthesis

KEGG PATHWAY

map00300
Lysine biosynthesis

map01064
Biosynthesis of alkaloids derived from ornithine, lysine and nicotinic acid

KEGG MODULE

M00016
Lysine biosynthesis, succinyl-DAP pathway, aspartate => lysine

M00030
Lysine biosynthesis, AAA pathway, 2-oxoglutarate => 2-aminoadipate => lysine

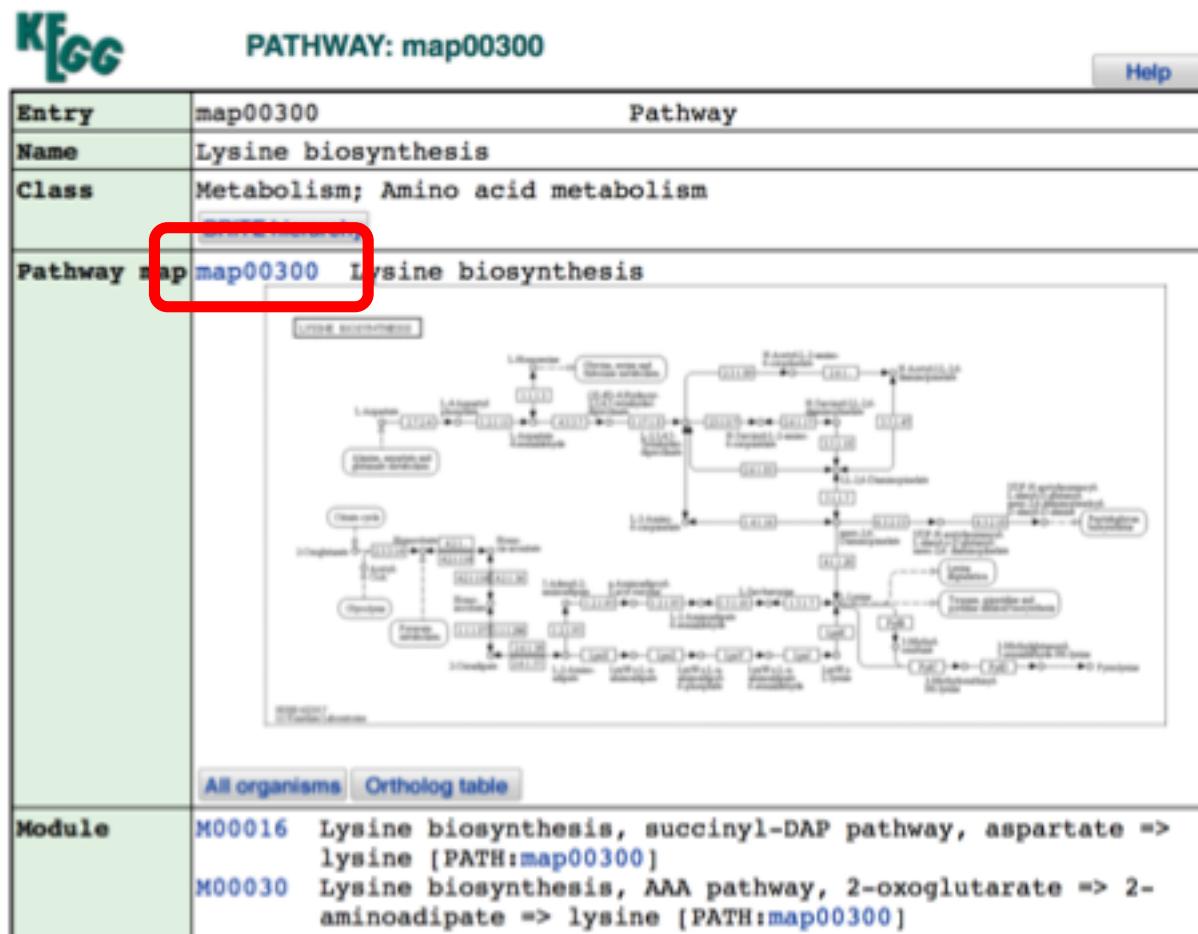
M00031
Lysine biosynthesis, mediated by LysW, 2-aminoadipate => lysine

M00433
Lysine biosynthesis, 2-oxoglutarate => 2-oxoadipate

M00525
Lysine biosynthesis, acetyl-DAP pathway, aspartate => lysine

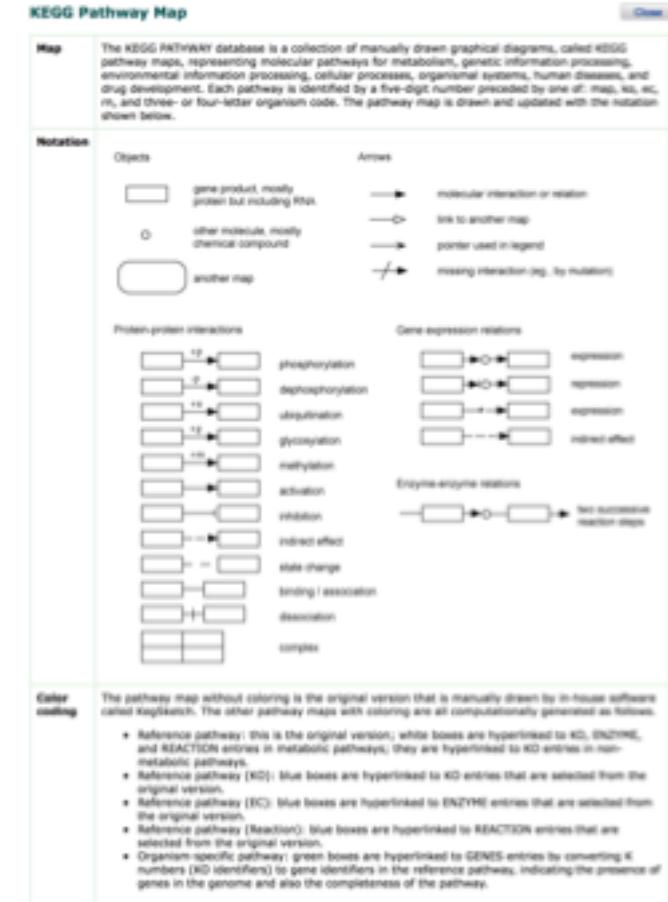
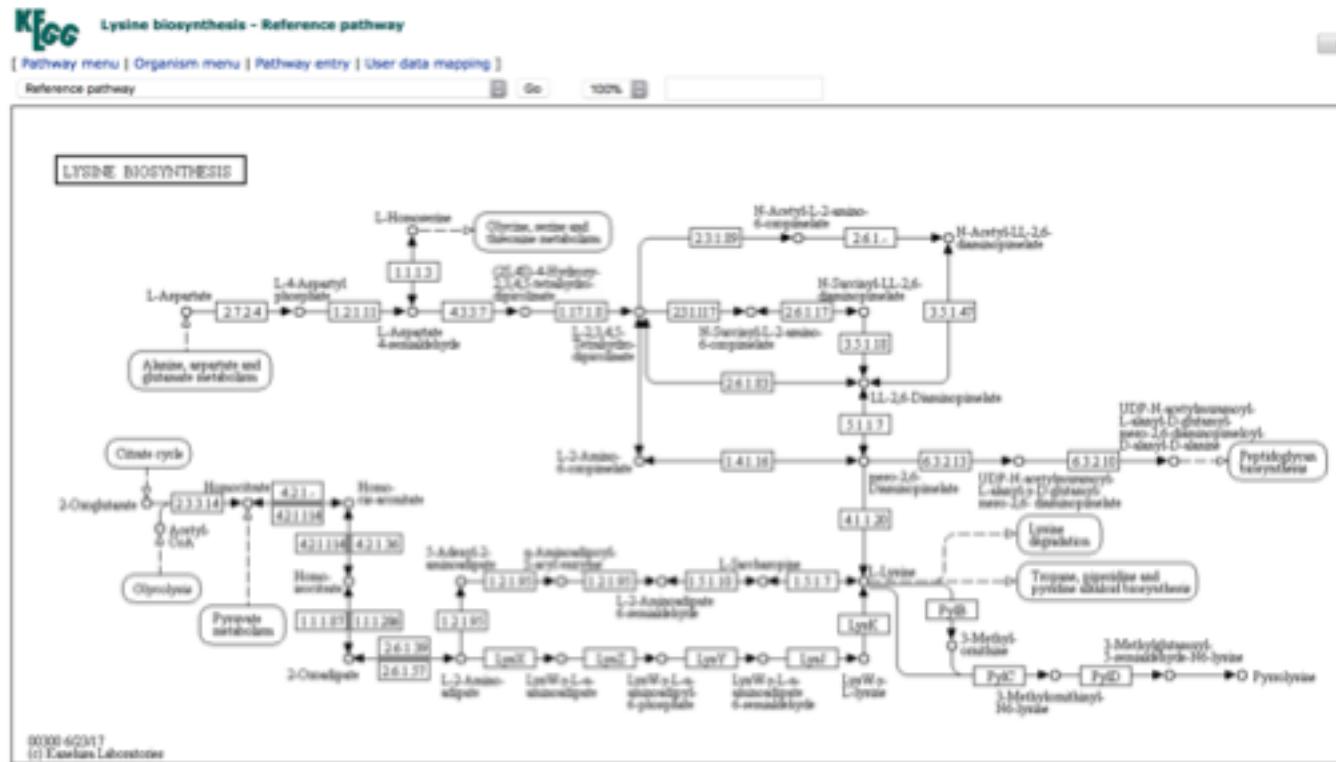
KEGGパスウェイマップ

- 各パスウェイの情報が表示されます。
 - KEGGにおけるパスウェイの最小単位で、ダイアグラム画像を“マップ”と呼んでいます。



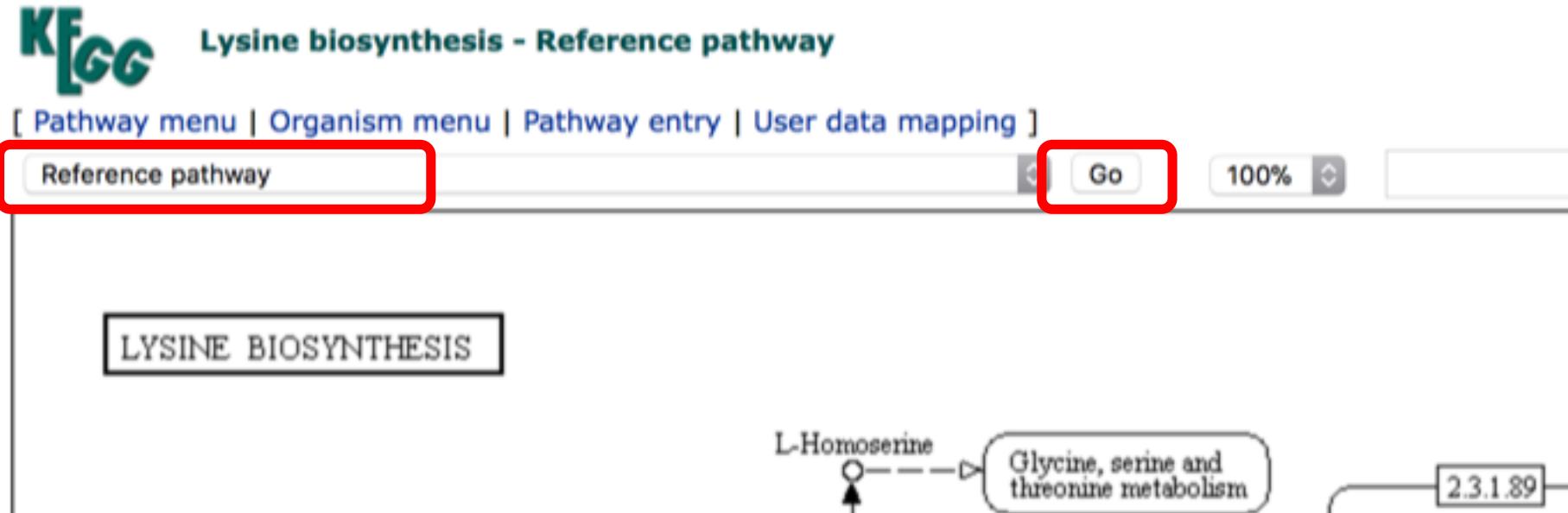
KEGGパスウェイマップ

- この色のついていない白いダイアグラムが、専門家が手作業で文献ベースから作成したリファレンスパスウェイになります
 - ボックスが遺伝子やタンパク質などの配列情報、丸が代謝産物、環境物質などの化合物
 - 各図形の説明は右上の Help から見られます



好きな生物のパスウェイを見る

- プルダウンメニューから好きな生物を選択して Go をクリック
 - リストが多すぎて選びにくいので
 - <Sort below by alphabet> を選択して Go をクリックでリストをソート
 - <Set personalized menu> を選択して Go をクリックでポップアップウィンドウからリストの絞り込み



別のアプローチ

- 種、属でまとめたパスウェイ、ドラフトゲノム、メタゲノムのパスウェイはここからは選べないので、生物種リストのページから、種のページ、パスウェイリストへ移動する必要があります。

KEGG Homo sapiens (human)

Genome info Pathway map Brite hierarchy Module Genome map Blast Taxonomy

Search genes: Go Clear

Prokaryotes

Category

Category	Organism
	eco Escherichia coli K-12 MG1655
	ecj Escherichia coli K-12 W3110
	ecd Escherichia coli K-12 DH10B
	ebw Escherichia coli BW2952
	ecok Escherichia coli K-12 MDS42
	ece Escherichia coli O157:H7 EDL933 (EHEC)
	ecs Escherichia coli O157:H7 Sakai (EHEC)
	ecf Escherichia coli O157:H7 EC4115 (EHEC)
	etw Escherichia coli O157:H7 TW14359 (EHEC)
	elx Escherichia coli O157:H7 Xuzhou21 (EHEC)
	eoj Escherichia coli O26:H11 11368 (EHEC)
	eoi Escherichia coli O111:H- 11128 (EHEC)

Pathway map (Red box)

Genome information

T number	T01001
Org code	hsa
Aliases	HUMAN, 9606
Full name	Homo sapiens (human)
Definition	Homo sapiens (human)
Annotation	manual
Taxonomy	TAX: 9606
2009 GenBank	
2012 GenBank	
2009 GenBank	
2009 GenBank	

Lysine biosynthesis マップ

- 一部のボックスが緑色で塗られる、その生物（またはサンプル）の持つ遺伝子を示しています。

00240 Pyrimidine metabolism

Amino acid metabolism

00250 Alanine, aspartate and glutamate metabolism

00260 Glycine, serine and threonine metabolism

00270 Cysteine and methionine metabolism

00280 Valine, leucine and isoleucine degradation

00290 Valine, leucine and isoleucine biosynthesis

Lysine biosynthesis

00310 Lysine degradation

00220 Arginine biosynthesis

00330 Arginine and proline metabolism

00340 Histidine metabolism

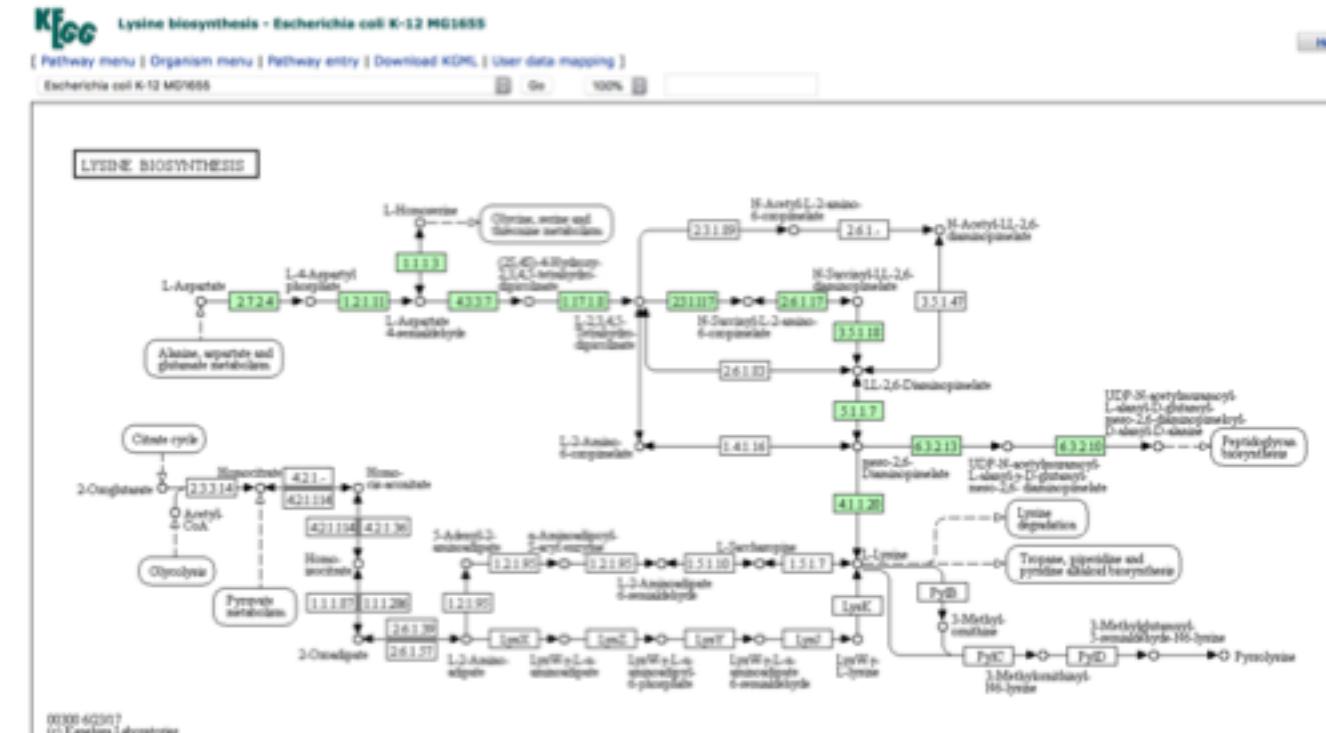
00350 Tyrosine metabolism

00360 Phenylalanine metabolism

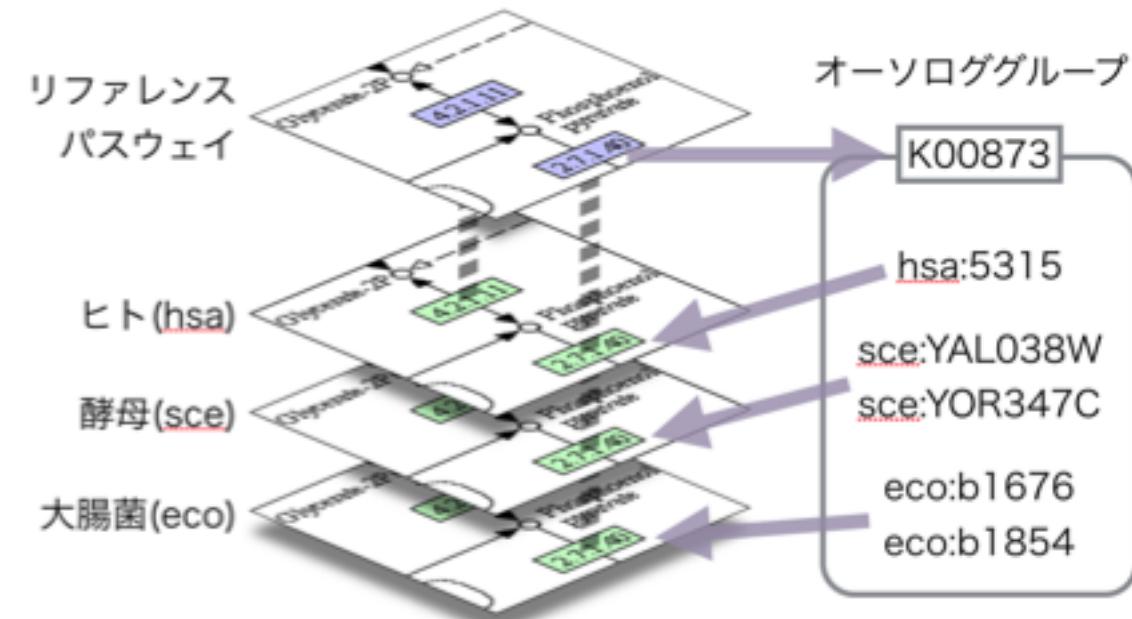
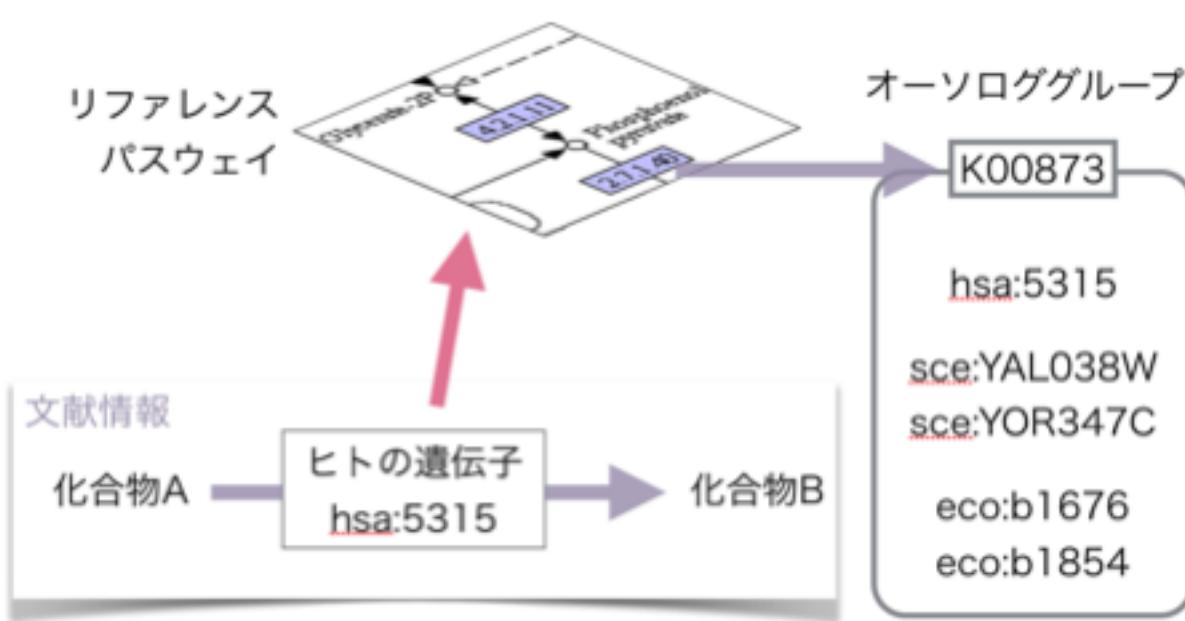
00380 Tryptophan metabolism

00400 Phenylalanine, tyrosine and tryptophan biosynthesis

Metabolism of other amino acids



リファレンスパスウェイと種毎のパスウェイの関係



Overviewマップを見る

- <https://www.kegg.jp/kegg/pathway.html>
- 1.0 Global and overview maps の Metabolic pathways をクリック
 - 右の [KEGG Atlas] は Java で動くビューウィーで、自由度が少し高い分、動作が重たい
- 左にモジュールのリスト（KEGG におけるパスウェイの小さい機能単位）、右にマップが表示
- 機能単位毎にパスウェイを強調表示できる

1. Metabolism

1.0 Global and overview maps

01100 Metabolic pathways	[KEGG Atlas]
01110 Biosynthesis of secondary metabolites	[KEGG Atlas]
01120 Microbial metabolism in diverse environments	[KEGG Atlas]
01130 Biosynthesis of antibiotics	[KEGG Atlas]
01200 Carbon metabolism	[KEGG Atlas]
01210 2-Oxocarboxylic acid metabolism	[KEGG Atlas]
01212 Fatty acid metabolism	[KEGG Atlas]
01230 Biosynthesis of amino acids	[KEGG Atlas]
01220 Degradation of aromatic compounds	[KEGG Atlas]
1.1 Carbohydrate metabolism	
00010 Glycolysis / Gluconeogenesis	
00020 Citrate cycle (TCA cycle)	

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

Reference pathway

Go

35%

KEGG module

Energy metabolism

 Carbon fixation

- 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 cycle
- M00171 C4-dicarboxylic acid cycle
- M00170 C4-dicarboxylic acid cycle
- M00173 Reductive citrate cycle
- M00579 Phosphate acetyltransferase

 Nitrogen metabolism

- M00175 Nitrogen fixation

 Methane metabolism

- M00567 Methanogenesis
- M00174 Methane oxidation

 Sulfur metabolism

- M00176 Assimilatory sulfur metabolism
- M00596 Dissimilatory sulfur metabolism
- M00595 Thiosulfate oxidation

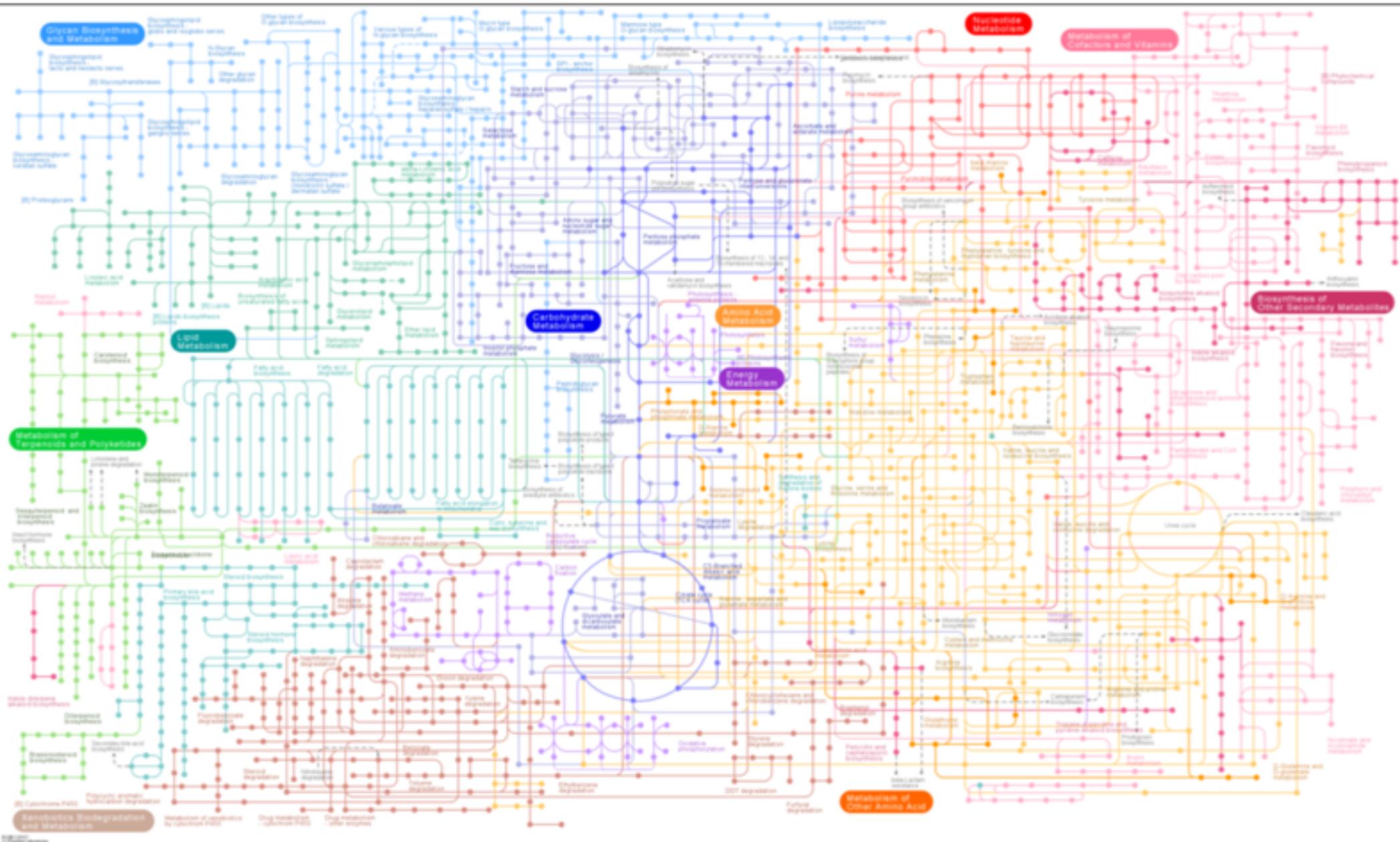
Carbohydrate and lipid metabolism

 Central carbohydrate metabolism

- M00001 Glycolysis (Embden-Meyerhof-Parnet pathway)
- M00002 Glycolysis, core
- M00003 Gluconeogenesis
- M00307 Pyruvate oxidation
- M00009 Citrate cycle (TCA cycle)
- M00010 Citrate cycle, first step
- M00011 Citrate cycle, second step
- M00004 Pentose phosphate pathway
- M00006 Pentose phosphate pathway
- M00007 Pentose phosphate pathway
- M00580 Pentose phosphate pathway
- M00005 PRPP biosynthesis
- M00008 Entner-Doudoroff pathway
- M00308 Semi-phosphoryl transferase
- M00633 Semi-phosphoryl transferase
- M00309 Non-phosphoryl transferase

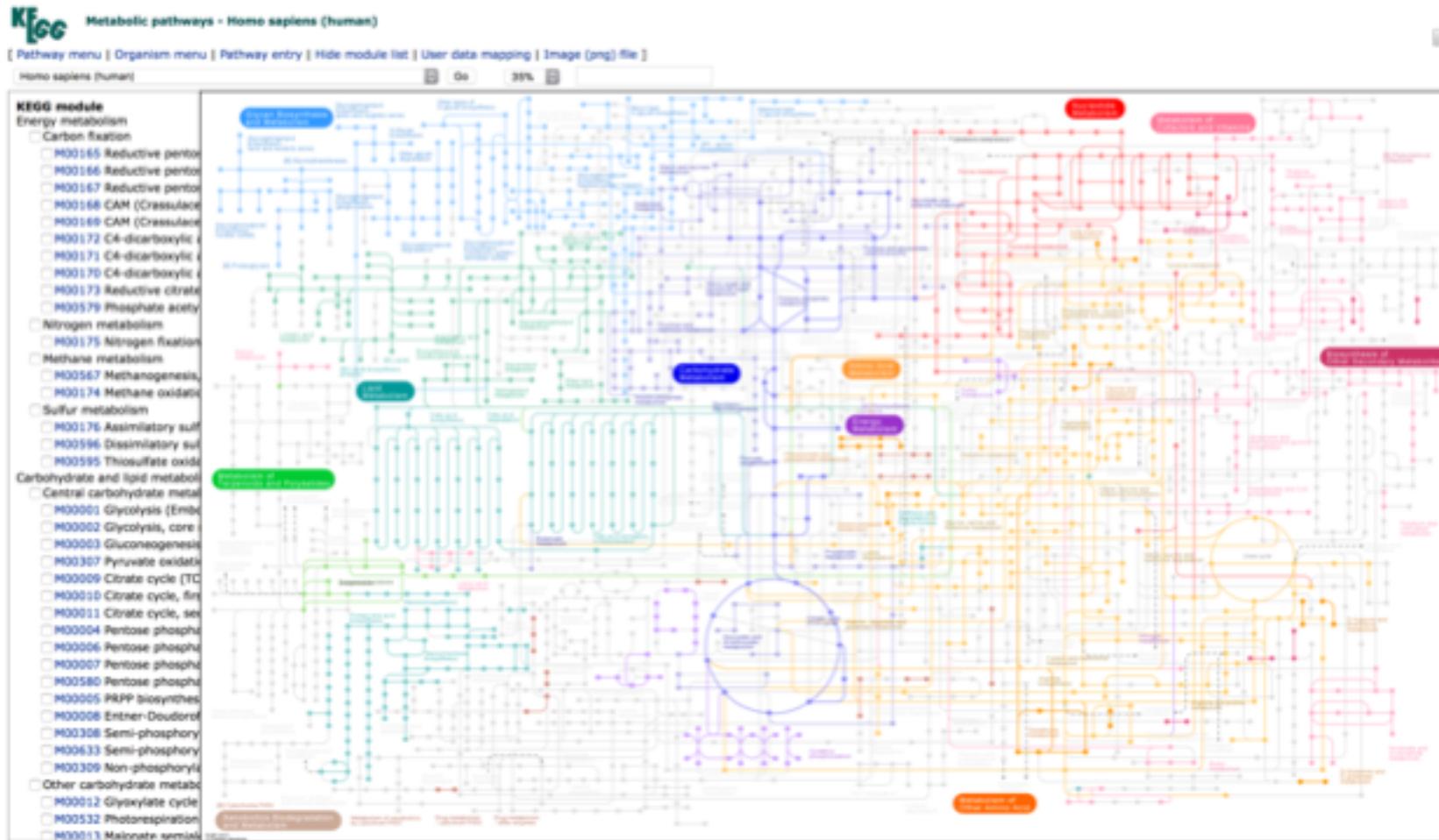
 Other carbohydrate metabolism

- M00012 Glyoxylate cycle
- M00532 Photorespiration
- M00013 Malonate semialdehyde cycle



生物種毎のOverviewマップを見る

- プルダウンメニューから生物を選択し、Go をクリック
- 生物の持っていない経路は灰色になる



ヒトの疾患パスウェイを見る

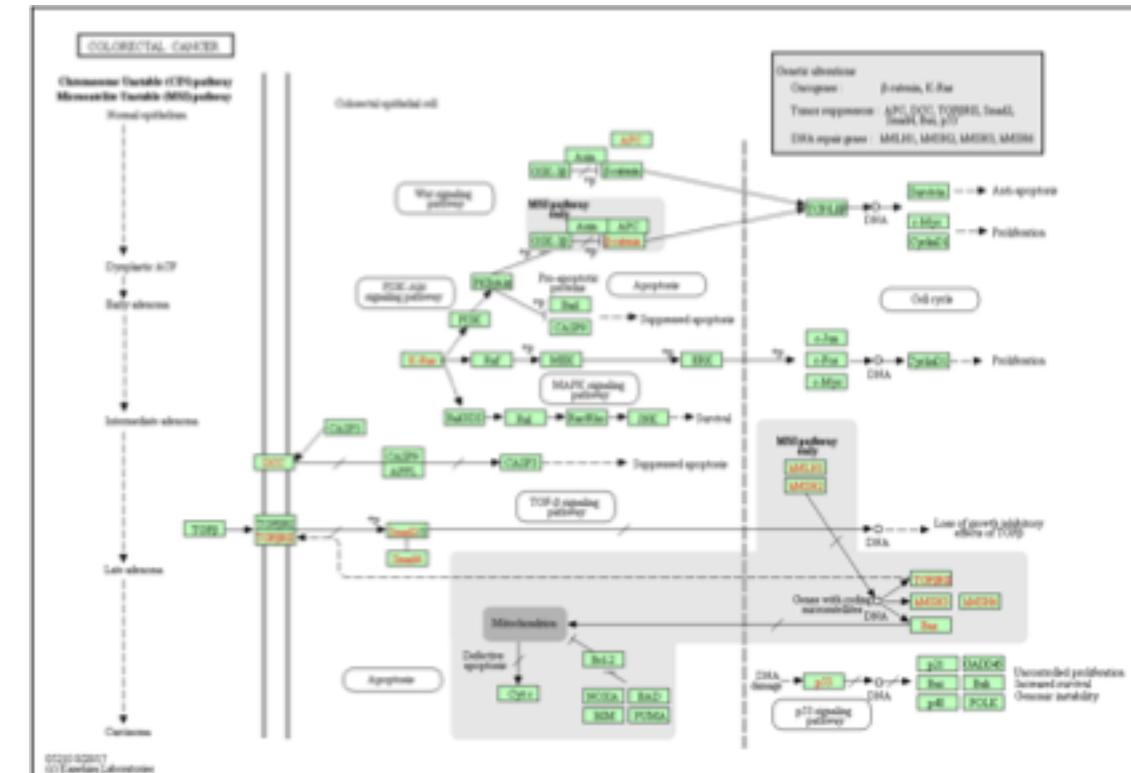
- <https://www.genome.jp/kegg/pathway.html#disease>
- がん、免疫系疾患、神経変性疾患など多因子性の疾患
- 好きな疾患パスウェイをクリック（例：大腸がん）
 - 赤字の遺伝子が疾患の病因遺伝子を示しています。

6.1 Cancers: Overview

- 05200 Pathways in cancer
- 05230 Central carbon metabolism in cancer
- 05231 Choline metabolism in cancer
- 05202 Transcriptional misregulation in cancer
- 05206 MicroRNAs in cancer
- 05205 Proteoglycans in cancer
- 05204 Chemical carcinogenesis
- 05203 Viral carcinogenesis

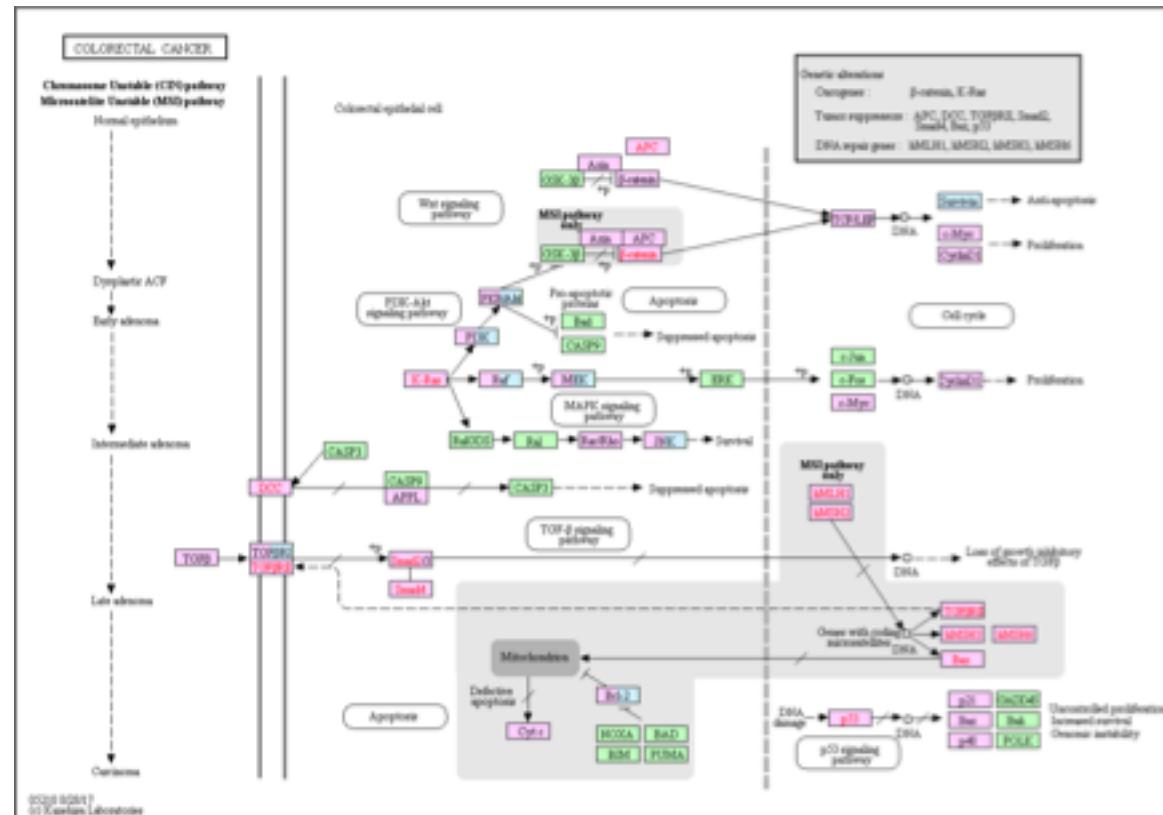
6.2 Cancers: Specific types

- 05210 Colorectal cancer
- 05212 Pancreatic cancer
- 05225 Hepatocellular carcinoma *New!*
- 05226 Gastric cancer *New!*



ヒトの疾患パスウェイ

- プルダウンメニューから Homo sapiens (human) + Disease/drug を選択
 - ピンクのボックスは何らかの疾患で病因遺伝子となっている遺伝子を示しています
 - ライトブルーのボックスは何らかの疾患で医薬品のターゲットとなっている遺伝子を示しています
- このように、KEGG では正常な状態のパスウェイの他に、病原因子や医薬品、そのターゲットなどの情報もパスウェイとして表現されています



種間比較をする

- <https://www.genome.jp/kegg/kegg2.html>
- KEGG GENOMEのリンクをクリック
- KEGG Mapping for Genome Comparison and Combination のテキストボックスに "eco ecs" と入力して Go をクリック
 - eco : 非病原性大腸菌 k-12
 - ecs : 病原性大腸菌 O157

Data-oriented entry points			
Category	Entry Point	Content	DBGET Search
Systems information	KEGG PATHWAY	KEGG pathway maps	PATHWAY
	KEGG BRITE	BRITE hierarchies and tables	BRITE
	KEGG MODULE	KEGG modules	MODULE
Genomic information	KO (KEGG Orthology)	Functional orthologs	ORTHOLOGY
	KEGG GENOME	KEGG organisms (complete genomes) Genes and proteins GENES sequence similarity	GENOME GENES
	KEGG SSDB		
Chemical information (KEGG LIGAND)	KEGG COMPOUND	Small molecules	COMPOUND
	KEGG GLYCAN	Glycans	GLYCAN
	KEGG REACTION	Reactions and reaction classes	REACTION
	KEGG ENZYME	Enzyme nomenclature	ENZYME
Health information	KEGG NETWORK <small>New!</small>	Disease-related network elements	NETWORK
	KEGG DISEASE	Human diseases	VARIANT
	KEGG DRUG	Drugs and drug groups	DISEASE
	KEGG ENVIRON	Health related substances	DRUG
	KEGG MEDICUS	Japanese drug labels (JAPIC) FDA drug labels (DailyMed)	DRUG GROUP ENVIRON

KEGG Mapping for Genome Comparison and Combination

An organism group may be defined to compare or combine KEGG organisms, other organisms and environmental samples, enabling the analysis of combined pathway maps for the group.

Define organism group (enter organism codes or T numbers):

eco ecs

Go

(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

種間比較

- 上のメニューの Pathway map をクリック
- 好きなパスウェイをクリック
 - 前者が持っている遺伝子が緑、後者が持っている遺伝子がピンクで表示
 - Overview パスウェイの場合、両者が持っている遺伝子はライトブルーで表示

KEGG Organism group: eco ecs

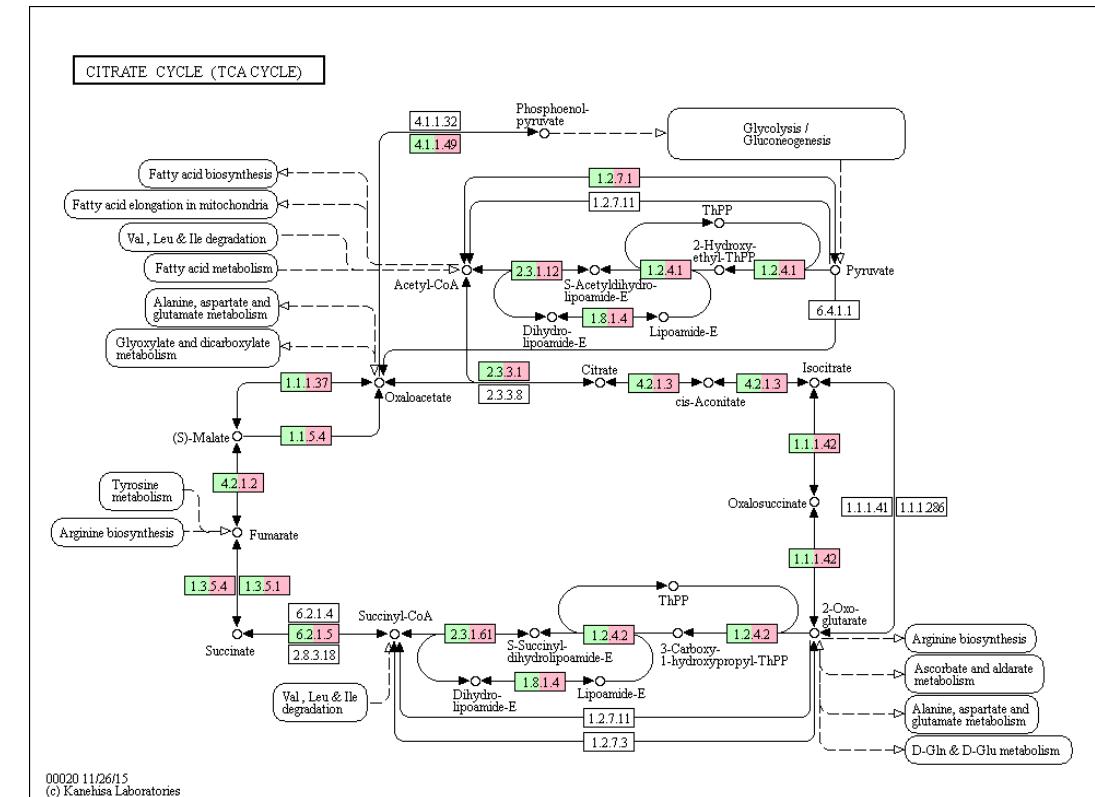
Category info Pathway map Brite hierarchy Taxonomy

Search genes: Go Clear

eco ecs

T00007 eco Escherichia coli K-12 MG1655
T00048 ecs Escherichia coli O157:H7 Sakai (EHEC)

KEGG2 PATHWAY BRITE GENES GENOME LIGAND DISEASE DRUG



Bacterial secretion system マップ

- 03070 : **Bacterial secretion system** パスウェイでは病原性に関する III型、VI型分泌装置が 0157 側だけが持っているのがわかる

Environmental Information Processing

Membrane transport

02010 ABC transporters

02060 Phosphotransferase system (PTS)

03070 Bacterial secretion system

Signal transduction

02020 Two-component system

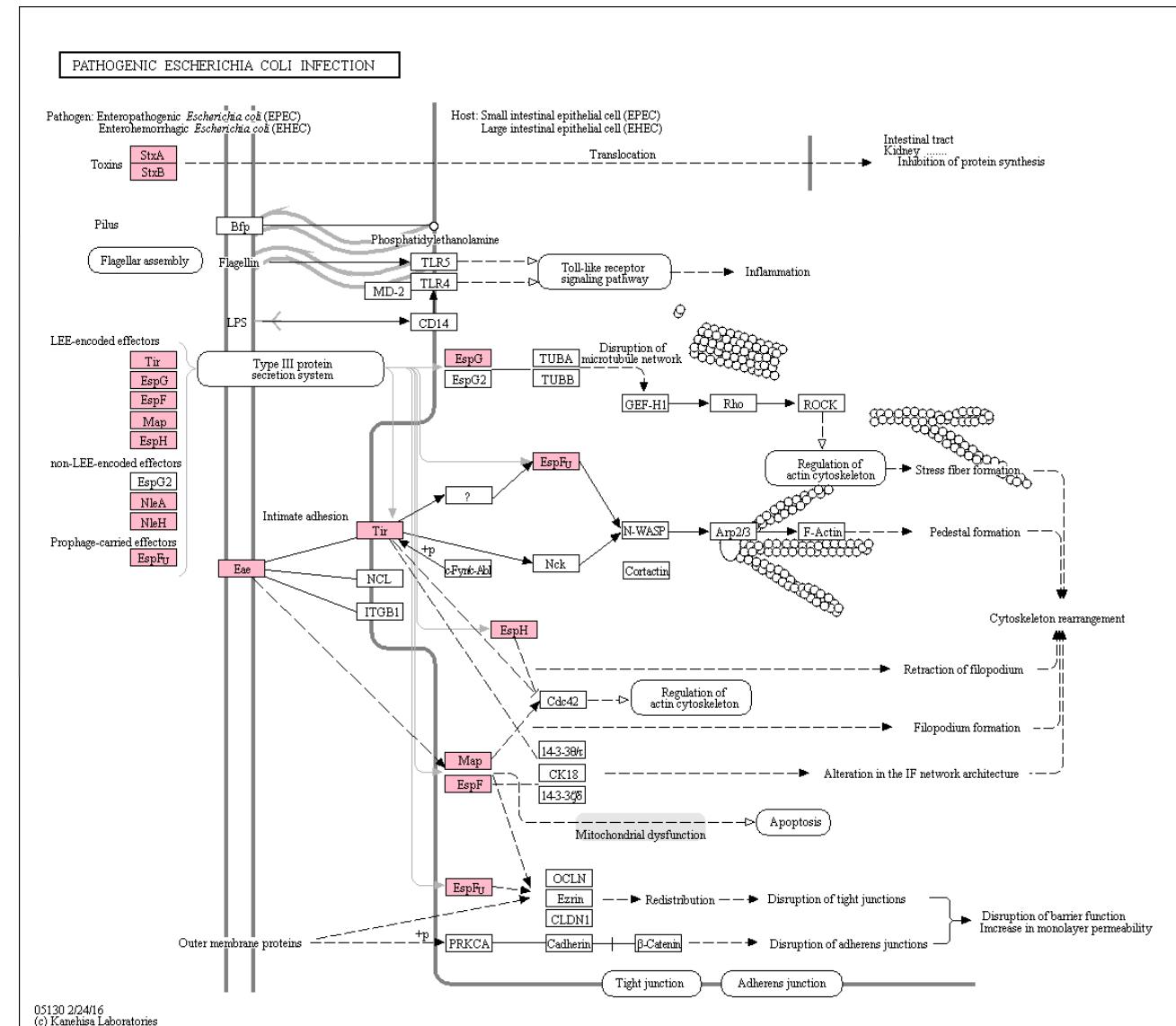
04014 ~~Two-component system~~

	Type I	Type III	Type II	Sec-SRP
Outer membrane protein (OMP)	TolC	YscF	GspD	SecD/F
Membrane fusion protein	HlyD	YscO YscP YscX	GspS	SecE
ABC transporter	HlyB	Secretin YscC	GspC GspF GspG	SecG
OMP		YscW	GspH GspI GspJ	
Inner membrane protein (IMP)		YscJ YscR YscS	GspK GspL GspM	
		YscT YscU YscV	GspE	
ATPase		YscN	Leaderpeptidase GspO	
ATPase-associated protein		YscQ YscL		
IMP			SecD/F SecE SecG	
ATPase			SecY YajC YidC	
Secretion monitor			SecA	
SRP receptor			SecM	
Targeting protein			FtsY	
			SecB ffh	

Pathogenic Escherichia coli infection

Infectious diseases: Bacterial

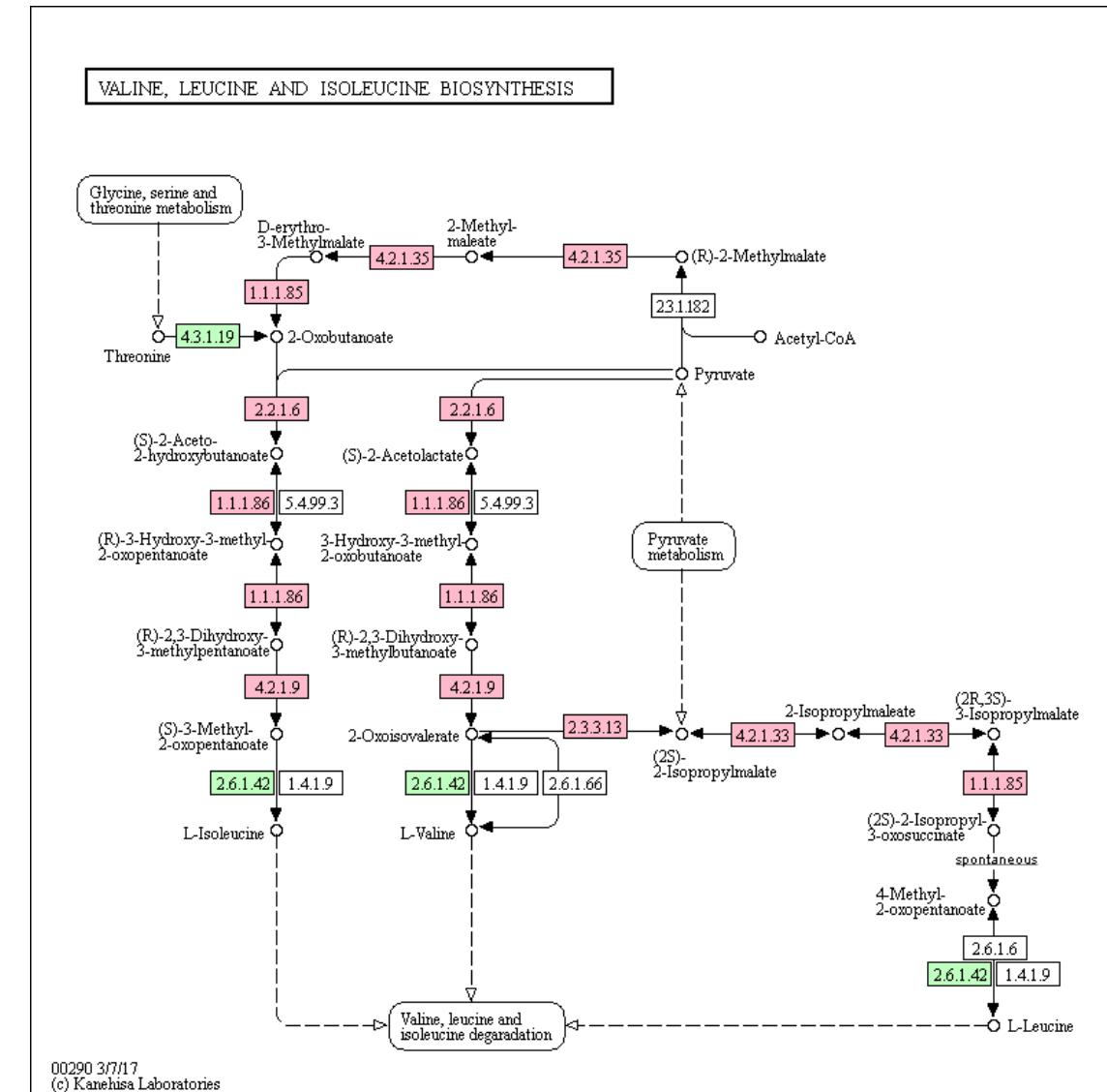
- 05110 Vibrio cholerae infection
- 05120 Epithelial cell signaling in Helicobacter pylori infection
- 05130 Pathogenic Escherichia coli infection**
- 05132 Salmonella infection
- 05131 Shigellosis
- 05133 Pertussis
- 05134 Legionellosis
- 05150 Staphylococcus aureus infection
- 05152 Tuberculosis
- 05100 Bacterial invasion of epithelial cells



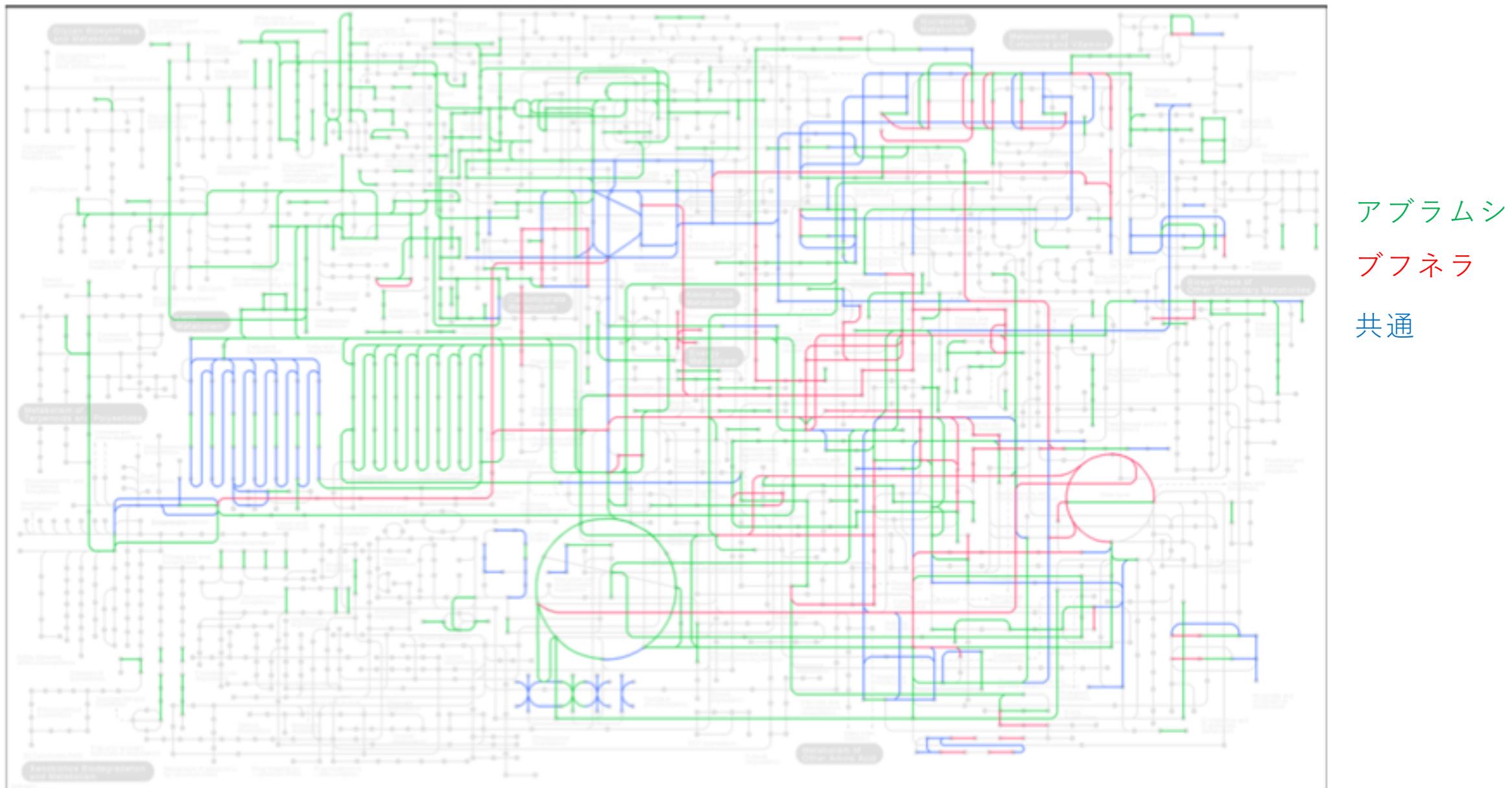
アブラムシとブフネラで種間比較

- 00290 : Valine, leucine and isoleucine biosynthesis マップ
 - アブラムシ : api
 - ブフネラ : buc
- 共生生物間のパスウェイ補完によってアミノ酸合成が可能になっていることがわかる

00240 Pyrimidine metabolism
Amino acid metabolism
00250 Alanine, aspartate and glutamate metabolism
00260 Glycine, serine and threonine metabolism
00270 Cysteine and methionine metabolism
00280 Valine, leucine and isoleucine degradation
00290 Valine, leucine and isoleucine biosynthesis
lysine biosynthesis
00300 Lysine degradation
00310 Lysine degradation
00220 Arginine biosynthesis
00330 Arginine and proline metabolism
00340 Histidine metabolism
00350 Tyrosine metabolism
00360 Phenylalanine metabolism
00380 Tryptophan metabolism



Overviewマップの場合



サンプルデータのマッピング

- KEGGには遺伝子リストからパスウェイをマッピングするツールが組み込まれています。

- KEGG Mapper
(<https://www.kegg.jp/kegg/mapper.html>)
- Pathway mapping tool の2番目のSearch&Color Pathwayをクリック
 - Search against : データベースコード
 - Primary ID : ID 種類 (KEGG ID, NCBI-GenelD, NCBI-ProteinID, UniProt)
 - テキストエリア : 要素のリスト (遺伝子、タンパク質、化合物)
 - [配列 ID or 代謝産物 ID] 塗りつぶし色[,線の色]
 - 配列 ID は KEGG gene ID, NCBI-GenelD, NCBI-ProteinID, UniProt ID
 - 代謝産物 ID は KEGG Compound ID (C番号)のみ
 - 線の色はオプション
 - 色は16進数表記か基本的なカラーネームで記述



KEGG Mapper

A suite of KEGG mapping tools

About KEGG Mapper

Search Pathway
Search&Color Pathway
Color Pathway

Search Brite
Search&Color Brite
Join Brite
Join Brite Table

Search Module
Search&Color Module

Search Disease

Reconstruct Pathway
Reconstruct Brite
Reconstruct Module
Map Taxonomy

Convert ID
Draw Phylogram

Annotate Sequence
BlastKOALA

KEGG Atlas
KEGG

About KEGG Mapper

KEGG Mapper is a collection of tools for KEGG mapping: KEGG pathway mapping, BRITE mapping, and MODULE mapping. Two pathway mapping tools, "Search Pathway" and "Color Pathway", were made available from the beginning of the KEGG project. The naming of these tools was somewhat misleading since both involved searching and coloring procedures. The latter is now called "Search&Color Pathway" and a new "Color Pathway" is introduced for coloring of a selected pathway map (without searching).

Pathway mapping tools

Search Pathway - basic pathway mapping tool

- Search against all pathway maps in one category (ko, org, etc.)
- Multiple search objects may be specified in one line
- Mapped objects are marked in red

Search&Color Pathway - pathway mapping tool with coloring options

- Search against all pathway maps in one category (ko, org, etc.)
- Each search object is specified in one line together with color attributes (background color and foreground color)
- Mapped objects can be marked in any color
- NCBI and UniProt identifiers may be used

Color Pathway - selected pathway map coloring tool

- One selected pathway map is colored in multiple ways
- Each object is specified in one line together with color attributes, i.e., multiple background and foreground colors in multiple columns
- Results shown as multiple snapshots of image maps
- When numerical values are given as attributes, they are displayed by color gradation or 3D bar graph

KEGG Mapper

- テキストエリア右の Example を選択して Exec ボタンをクリックすると、ヒットしたパスウェイのリストが表示されます（カッコの中はヒットした要素の数）

KEGG Mapper – Search&Color Pathway

Search against: hsa Enter: map, ko, ec, rn, hsadd, or org

Primary ID: KEGG identifiers (Outside IDs for organism-specific pathways only)

Enter objects one per line followed by bgcolor, fgcolor:

Examples: Homo sapience pathway

7167 red,blue
C00118 pink

Alternatively, enter the file name containing the data:

ファイルを選択 ファイル未選択

If necessary, change default bgcolor: pink

Include aliases

The screenshot shows the KEGG Mapper interface for "Search&Color Pathway". On the left, there's a sidebar with links like "About KEGG Mapper", "Search Pathway" (with "Search&Color Pathway" and "Color Pathway" options), "Search Brite" (with "Search&Color Brite", "Join Brite", and "Join Brite Table" options), "Search Module" (with "Search&Color Module"), "Search Disease", "Reconstruct Pathway" (with "Reconstruct Brite", "Reconstruct Module", and "Map Taxonomy" options), and "Convert ID". The main area has a search bar for "Search against: hsa" and a dropdown for "Primary ID: KEGG identifiers". Below that is a text input field for entering objects with their colors, showing examples "7167 red,blue" and "C00118 pink". There are also sections for entering a file name and changing the default background color, with a checkbox for including aliases.

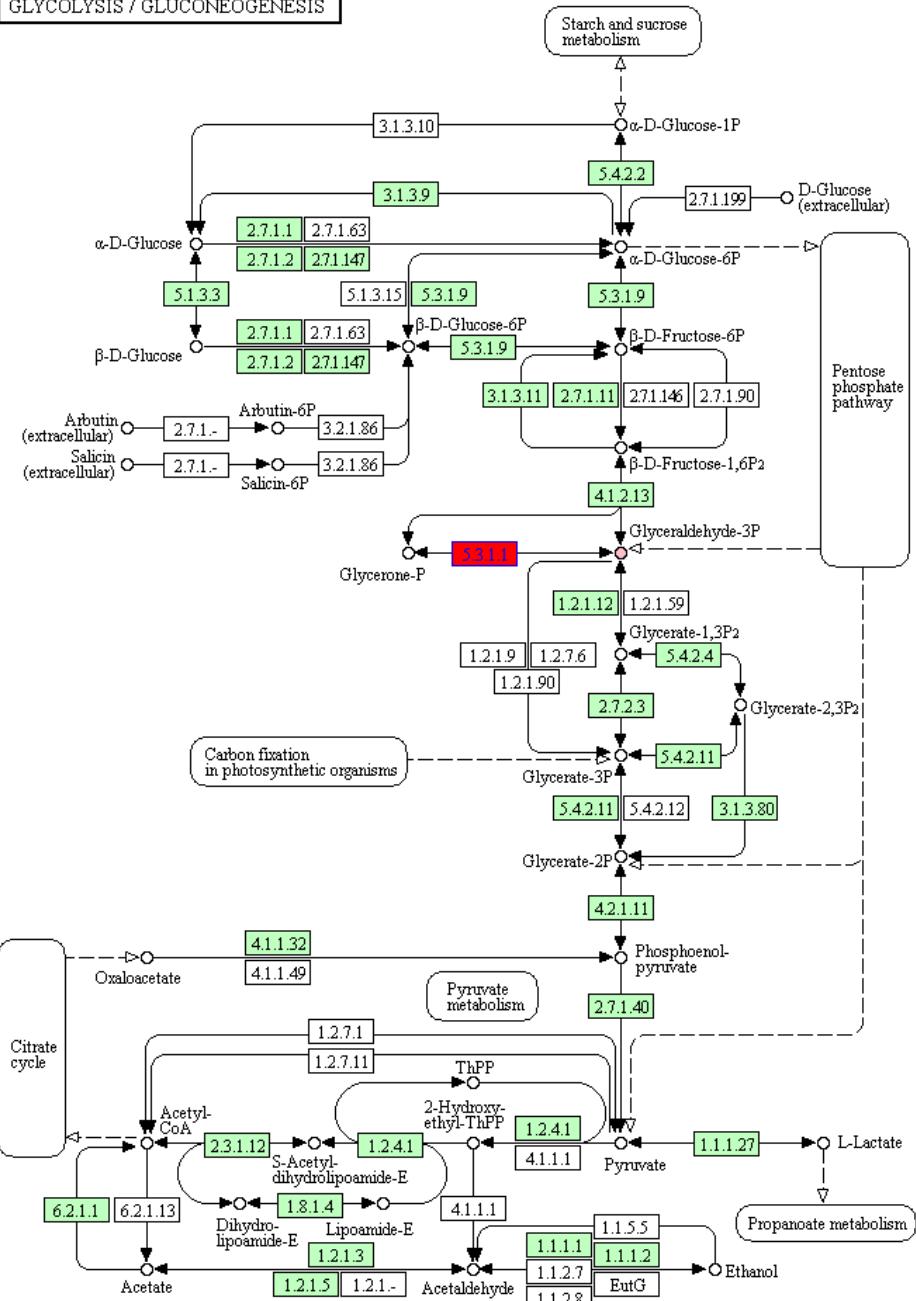
Pathway Search Result

Sort by the pathway list

Show all objects

- hsa01230 Biosynthesis of amino acids - Homo sapiens (human) (2)
- hsa01100 Metabolic pathways - Homo sapiens (human) (2)
- hsa00010 Glycolysis / Gluconeogenesis - Homo sapiens (human) (2)
- hsa00562 Inositol phosphate metabolism - Homo sapiens (human) (2)
- hsa00051 Fructose and mannose metabolism - Homo sapiens (human) (2)
- hsa01200 Carbon metabolism - Homo sapiens (human) (2)
- hsa00900 Terpenoid backbone biosynthesis - Homo sapiens (human) (1)
- hsa00750 Vitamin B6 metabolism - Homo sapiens (human) (1)
- hsa00052 Galactose metabolism - Homo sapiens (human) (1)
- hsa00030 Pentose phosphate pathway - Homo sapiens (human) (1)
- hsa00730 Thiamine metabolism - Homo sapiens (human) (1)

GLYCOLYSIS / GLUCONEOGENESIS

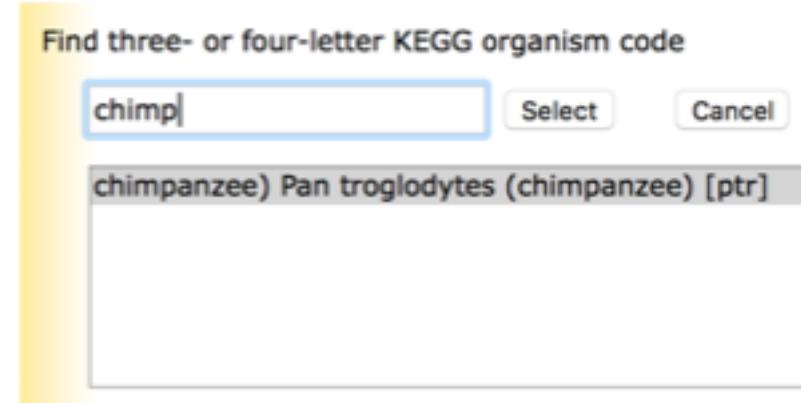


チンパンジーの遺伝子をマッピング

- テキストボックスに「例」のIDを記入
- Search against: にチンパンジーの生物種コードを入れる
 - コードがわからないので、org ボタンをクリック
 - ポップアップウィンドウでに 種名を入力すると、下のボックスに候補が出るので、選択したあと Select をクリック
 - チンパンジーのコード "ptr" が入力されていることを確認

例

453039 red
104003784 coral
453645 gray,red
453565 blue,yellow
450453 #fbfb88
463861 #88ffbb



Search against: ptr Enter: map, ko, ec, rn, hsadd, or org

Primary ID: KEGG identifiers (Outside IDs for organism-specific p

Enter objects one per line followed by bgcolor, fgcolor:

```
453039 red
104003784 coral
453645 gray,red
453565 blue,yellow
450453 #fbfb88
463861 #88ffbb
```

Examples: Select

Alternatively, enter the file name containing the data:

ファイルを選択 ファイル未選択

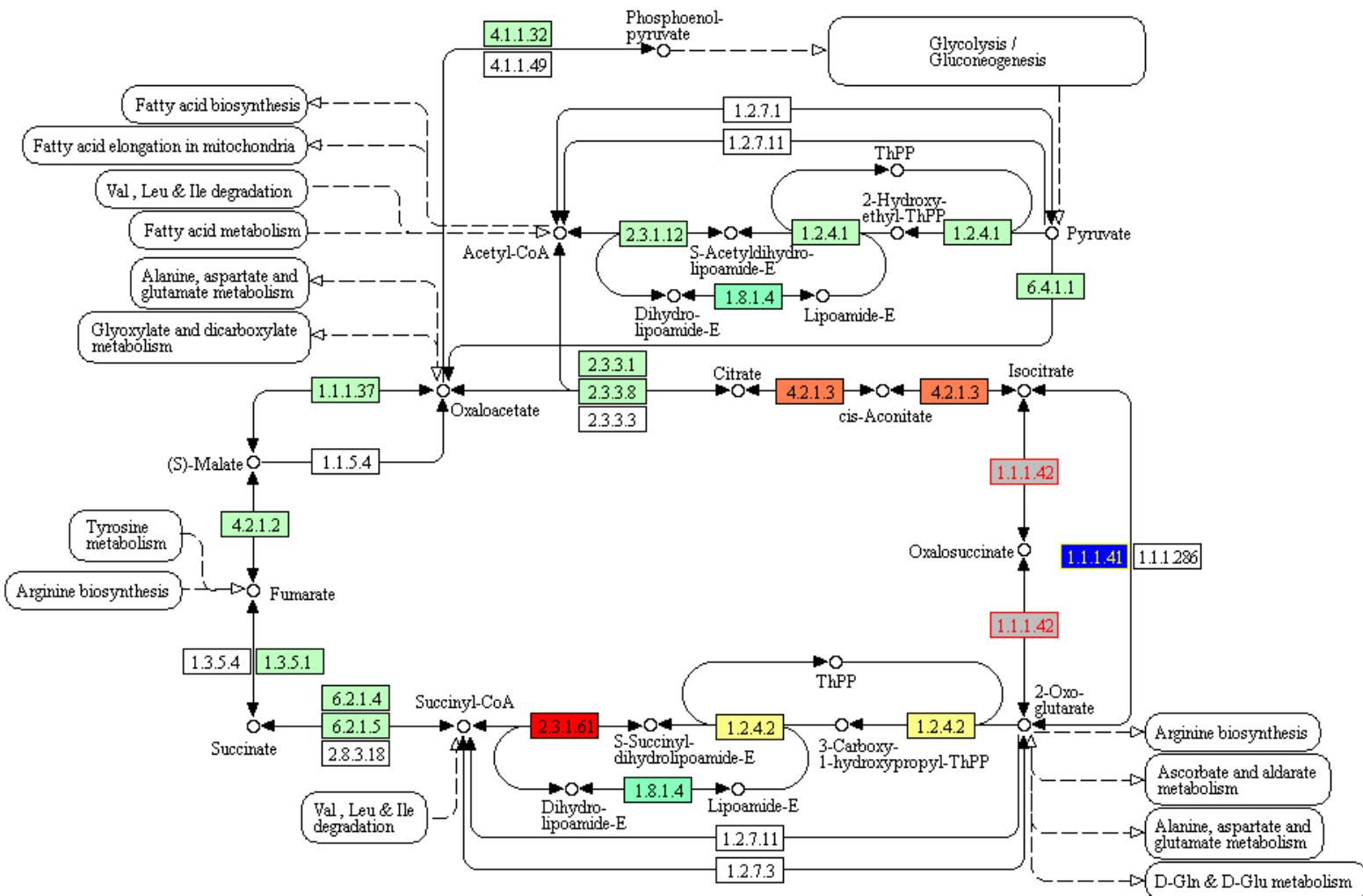
Pathway Search Result

Sort by the pathway list

Show all objects

- ptr00020 Citrate cycle (TCA cycle) - Pan troglodytes (chimpanzee) (6)**
- ptr01200 Carbon metabolism - Pan troglodytes (chimpanzee) (6)
- ptr01100 Metabolic pathways - Pan troglodytes (chimpanzee) (6)
- ptr01210 1-Oxocarboxylic acid metabolism - Pan troglodytes (chimpanzee) (3)
- ptr01230 Biosynthesis of amino acids - Pan troglodytes (chimpanzee) (3)
- ptr00630 Glyoxylate and dicarboxylate metabolism - Pan troglodytes (chimpanzee) (2)
- ptr00310 Lysine degradation - Pan troglodytes (chimpanzee) (2)
- ptr00380 Tryptophan metabolism - Pan troglodytes (chimpanzee) (1)
- ptr00280 Valine, leucine and isoleucine degradation - Pan troglodytes (chimpanzee) (1)
- ptr00640 Propanoate metabolism - Pan troglodytes (chimpanzee) (1)
- ptr00480 Glutathione metabolism - Pan troglodytes (chimpanzee) (1)
- ptr00260 Glycine, serine and threonine metabolism - Pan troglodytes (chimpanzee) (1)
- ptr00420 Pyruvate metabolism - Pan troglodytes (chimpanzee) (1)
- ptr00010 Glycolysis / Gluconeogenesis - Pan troglodytes (chimpanzee) (1)
- ptr04146 Peroxisome - Pan troglodytes (chimpanzee) (1)

CITRATE CYCLE (TCA CYCLE)



数値データをマッピング

- [Color Pathway](https://www.kegg.jp/kegg/tool/map_pathway3.html) (https://www.kegg.jp/kegg/tool/map_pathway3.html) をクリック
- 右のサンプル Numerical values to colorを選択
 - 中身は配列 ID と数値の対応リスト
- Select KEGG pathway map: でパスウェイを指定 (hsa05200)
- Option: で Numerical values converted to color gradationを選択
- Exec ボタンをクリック

Select KEGG pathway map: hsa05200

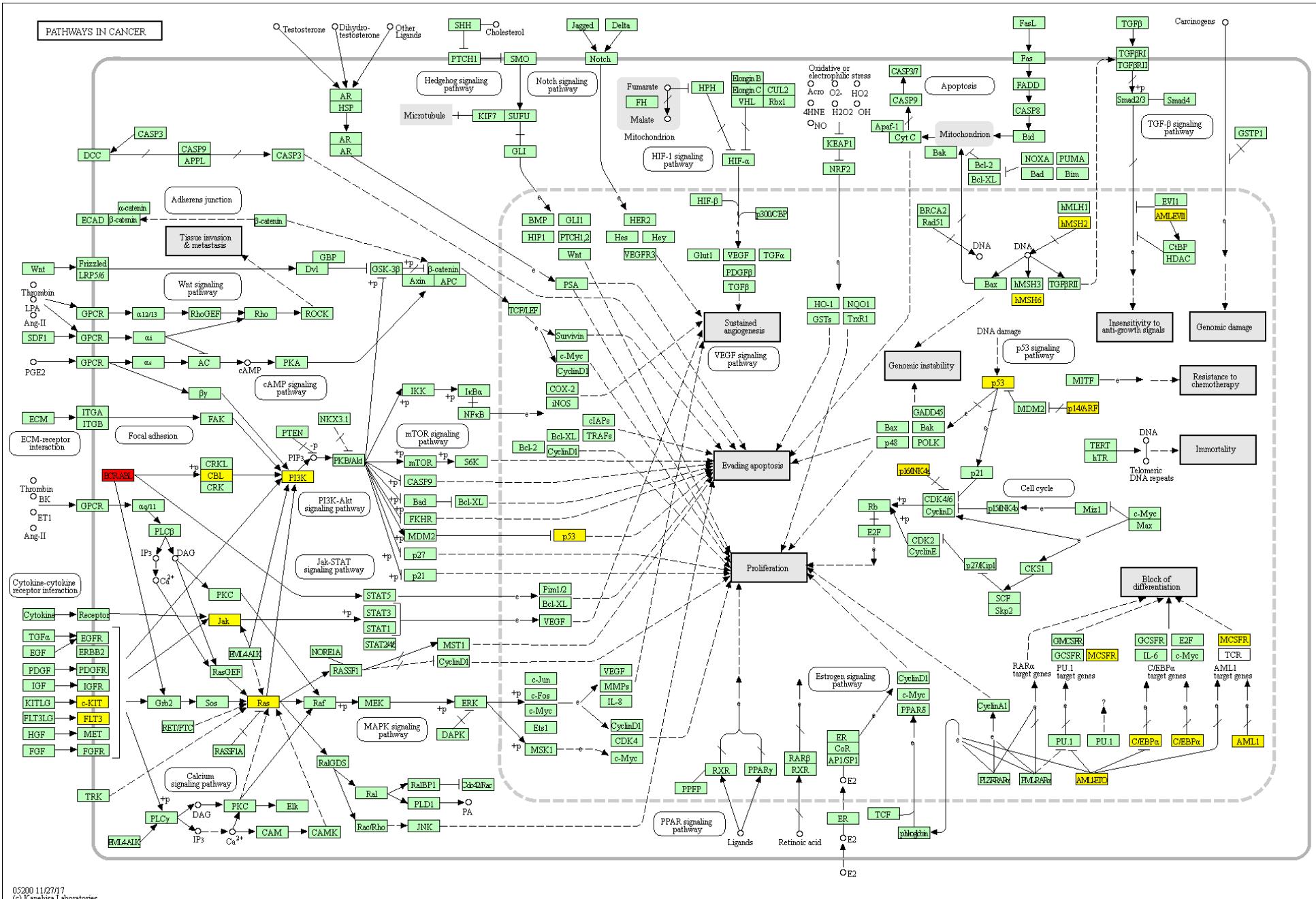
Enter data:

```
#hsa COSMIC
hsa:25    678
hsa:861   26
hsa:867   13
hsa:1029  44
hsa:1050  4
hsa:1436  3
hsa:2146  9
hsa:2322  7
hsa:2623  2
```

Examples:
Numerical values to color

Alternatively, enter the file name containing the data:

ファイルを選択 ファイル未選択



3Dグラフマッピング

- 右のサンプル Numerical values to colorを選択する

Select KEGG pathway map: hsa05200

Enter data:

```
#hsa:COSMIC
hsa:25 678
hsa:861 26
hsa:867 13
hsa:1029 44
hsa:1050 4
hsa:1436 3
hsa:2146 9
hsa:2322 7
hsa:2623 2
```

Numerical values to graph

Alternatively, enter the file name containing the data:

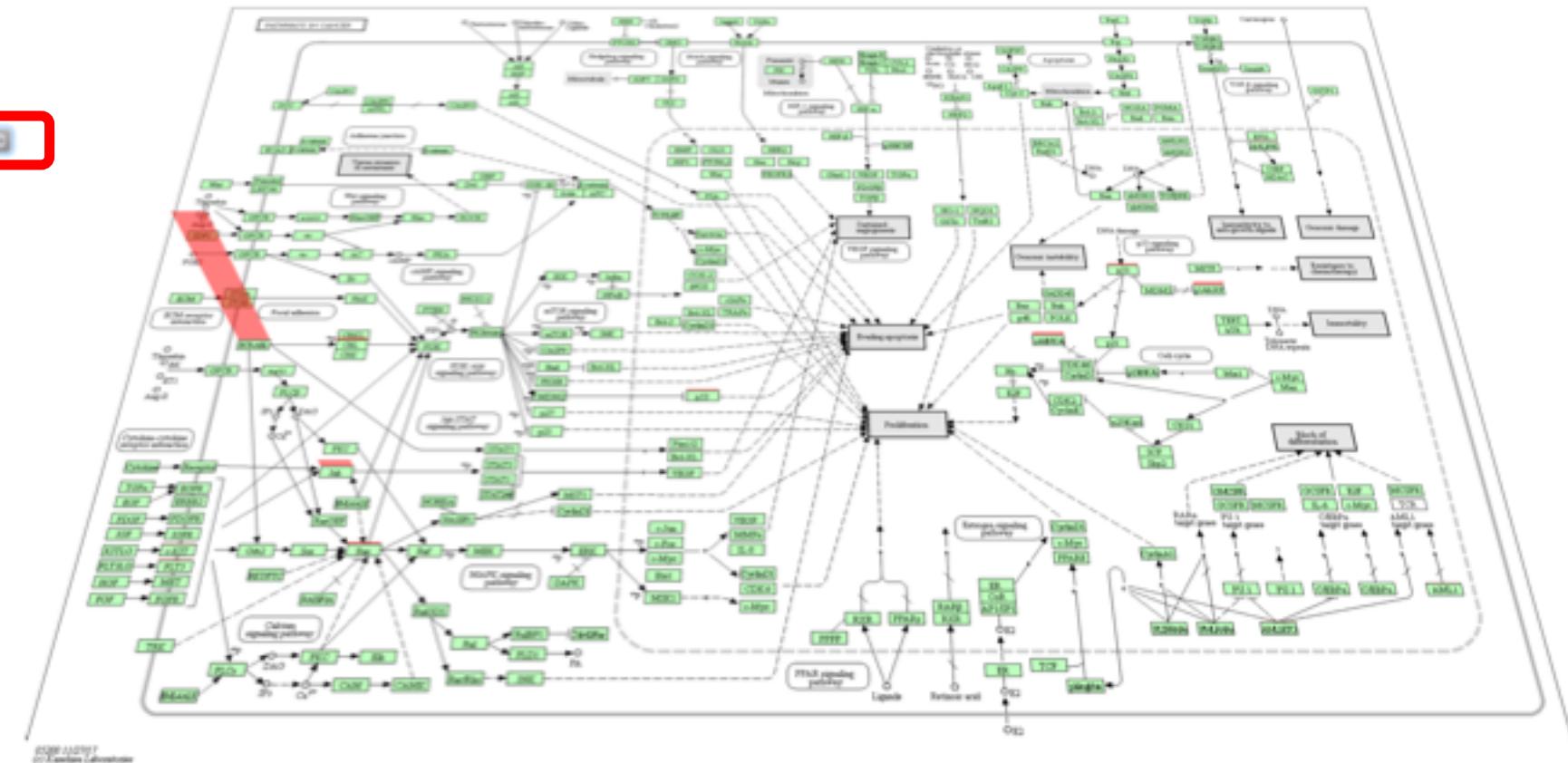
ファイルを選択 ファイル未選択

Option:

- Color specification
- Numerical values converted to color gradation
Range: minimum-maximum negative-zero-positive
minimum #ffff00 maximum #ff0000 Check
- Numerical values converted to 3D bar graph

Use uncolored diagram

Exec Clear



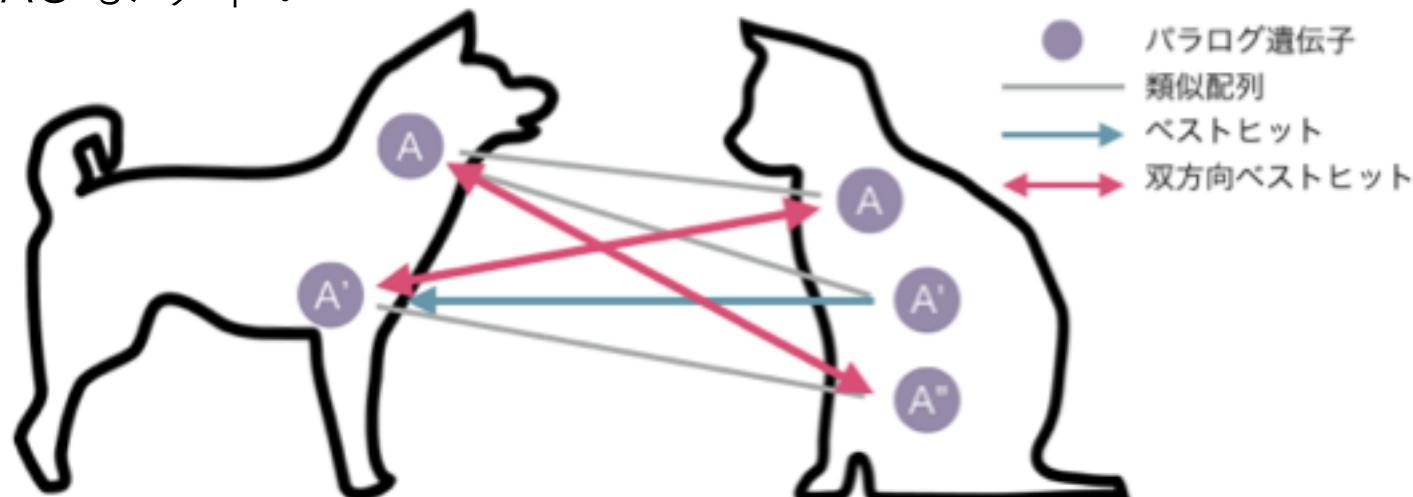
KEGGデータベースにはない遺伝子のマッピング

- KEGGに登録されている配列データと類似性を計算し、自動で遺伝子機能を推定、パスウェイへのマッピングを行う。
- KAAS (<https://www.genome.jp/tools/kaas/>)
- BlastKOALA, GhostKOALA

The screenshot shows the KAAS web interface. The top header reads "KAAS - KEGG Automatic Annotation Server for ortholog assignment and pathway mapping". The left sidebar has a "Request" tab selected, containing sections for "About KAAS", "Complete or Draft Genome", "Partial Genome", and "Metagenomes", each with links to help pages. The right sidebar shows "Example of Results" for "KO assignment" and "KEGG pathway mapping". The "KO assignment" section displays a table of query genes and their assigned KOs, while the "KEGG pathway mapping" section shows a complex network diagram of biological pathways.

配列相同意検索

- KAAS (<https://www.genome.jp/tools/kaas/>)
 - 配列類似性の計算は BLAST, GhostX, GhostZ ベースの 3 つ
 - GhostX は BLAST より精度は劣るが 100 倍早い
 - GhostZ は GhostX より精度は劣るが 2 倍早い
 - 種間で両方向ベストヒットを利用して遺伝子機能を推定（片方向も可能）
- BlastKOALA, GhostKOALA
 - 配列類似性の計算は BLAST, GhostX ベース
 - クエリーからデータベースへの片方向の計算なのと、データベースを圧縮しているぶん KAAS より早い



BlastKOALAを使ったパスウェイマッピング

- [Annotate Sequence by BlastKOALA](#)
(https://www.kegg.jp/kegg/tool/annotate_sequence.html)
- Example: の sequence.txt をコピー & ペースト、もしくはダウンロードしてファイルを選択
- Family/Genus ボタンをクリック
 - サンプルが Buchnera の仲間なので、KEGG の Buchnera データを使う
 - 新たに開いたウィンドウで、Buchnera を探し、Taxonomy番号をクリック

KEGG Mapper – Annotate Sequence by BlastKOALA

About KEGG Mapper
Search Pathway
Search&Color Pathway
Color Pathway

Search Brite
Search&Color Brite
Join Brite
Join Brite Table

Search Module
Search&Color Module

Search Disease

Reconstruct Pathway
Reconstruct Brite
Reconstruct Module
Map Taxonomy

Upload query amino acid sequences in FASTA format

Enter FASTA sequences

```
YLINFFYKTLKIKGTPIQIQFKDNENPYVKNNK  
>cl|N2_ACFK01000001.1_prot_WP_009874555.1_575 [gene=CSS_RS0103170]  
[protein=/inference=EXISTENCE: similar to AA sequence:SwissProt:P57663.1]  
[protein_id=WP_009874555.1] [location=complement(641210..641791)]  
MLNISKKNIIFFILFFLISLIFLNWKYFSLVNKENLESLKYEKIKKINKKKSKNLYEVENFIVQ  
NTSI  
YGTLTALSIAKKYVECNNLDKALLQLNNSLKYTKNEENLKNLLKINIAKIQIQQKNEENKAMNIL  
ETIQNHN  
WKNIIEHMKGDIFININNKKEAIKSWKSLFIEDSNASKEIINMKLNELKEQN
```

Or upload file:

Select GENES family/genus dataset to be searched

Select Family/Genus 32199 Buchnera

Example:
Query data: [sequence.txt](#)
[Select Buchnera \(32199\)](#)

Exec Clear

Erwinia	551	Erwinia amylovora
		Erwinia billingiae
		Erwinia sp. Ejp617
		Erwinia gerundensis
Buchnera	32199	Buchnera aphidicola
Wigglesworthia	51228	Wigglesworthia glossinidia
		Pantoea ananatis
		Pantoea vagans
		Pantoea en At-Qh

計算中

- Exec ボタンをクリック
- 数分待つ



BlastKOALA Job Request

Request accepted

Your job ID is 2fed9b298f4efd7e9a9eebed200f16a1fa687edb

Started at Mon Jan 8 15:09:48 JST 2018

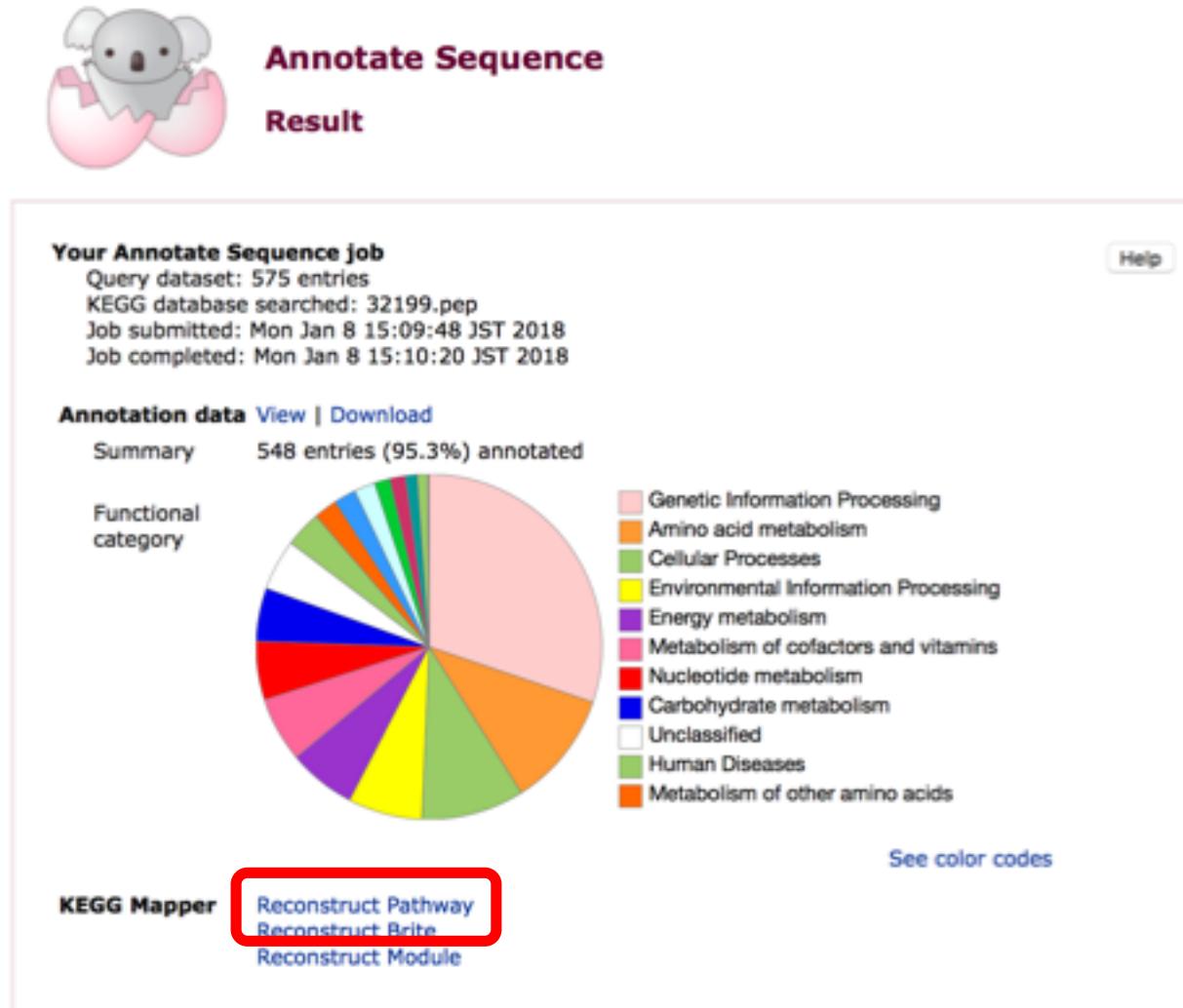
This page reloads automatically every 5 seconds

Your result will be displayed in the following page

[http://www.kegg.jp/kegg-bin/blastkoala_result?
id=2fed9b298f4efd7e9a9eebed200f16a1fa687edb&passwd=mTYL46&mode=mapper](http://www.kegg.jp/kegg-bin/blastkoala_result?id=2fed9b298f4efd7e9a9eebed200f16a1fa687edb&passwd=mTYL46&mode=mapper)

計算結果

- Reconstruct Pathway から遺伝子がマッピングされたパスウェイを見ることができる



Overviewマップの再構築結果

Pathway Reconstruction Result

Show all objects

Metabolism

Global and overview maps

- 0 100 Metabolic pathways (205)
- 0 110 Biosynthesis of secondary metabolites (97)
- 01120 Microbial metabolism in diverse environments (52)

- 01130 Biosynthesis of antibiotics (76)

- 01200 Carbon metabolism (30)

- 01210 2-Oxocarboxylic acid metabolism (10)

- 01212 Fatty acid metabolism (5)

- 01230 Biosynthesis of amino acids (62)

Carbohydrate metabolism

- 00010 Glycolysis / Gluconeogenesis (15)

- 00020 Citrate cycle (TCA cycle) (5)

- 00030 Pentose phosphate pathway (12)

- 00040 Pentose and glucuronate interconversions (1)

- 00051 Fructose and mannose metabolism (7)

- 00052 Galactose metabolism (1)

- 00500 Starch and sucrose metabolism (2)

- 00520 Amino sugar and nucleotide sugar metabolism (9)

- 00620 Pyruvate metabolism (6)

- 00630 Glyoxylate and dicarboxylate metabolism (2)

- 00640 Propanoate metabolism (3)

- 00650 Butanoate metabolism (2)

- 00660 C5-Branched dibasic acid metabolism (2)

- 00562 Inositol phosphate metabolism (2)

Energy metabolism

- 00190 Oxidative phosphorylation (27)

- 00195 Photosynthesis (8)

- 00710 Carbon fixation in photosynthetic organisms (7)

- 00720 Carbon fixation pathways in prokaryotes (4)

- 00680 Methane metabolism (8)

- 00920 Sulfur metabolism (9)

Lipid metabolism

- 00061 Fatty acid biosynthesis (5)

- 00564 Glycerophospholipid metabolism (1)

- 01040 Biosynthesis of unsaturated fatty acids (2)

Nucleotide metabolism

