

パスウェイ情報を中心とした KEGG/GenomeNet Webサービスの紹介

統合データベース講習会：AJACS尾張

奈良先端科学技術大学院大学

武藤 愛 muto@bs.naist.jp

2017年1月31日（火）15:15-16:45@藤田保健衛生大学

Table of Contents

- GenomeNet/KEGGの概要説明
- ブラウザを使ったGenomeNet/KEGGの利用法
 - GenomeNet実習
 - bfind, boget, LinkDB
 - KEGG実習（KEGG PATHWAYを中心に）
 - KEGG Mapper, Blast KOALA

GenomeNet

■ ゲノム情報を基盤とした新しい生命科学研究と創薬・医療・環境保全への応用を推進するためのインターネットサービス

■ KEGGを主とするデータベース群と、それらを利用した解析を行なうための計算ツール群からなる

The screenshot shows the GenomeNet homepage with a search bar and links to various databases and tools. Key visible links include: ゲノムネット (GenomeNet), ゲノムネットとはお知らせ (About GenomeNet), 総合データベース (Comprehensive Database), 総合CBの概要 (Summary of Comprehensive Bioresource), KEGGの概要 (Summary of KEGG), リース情報 (Release Information), 薬業品データベース (Pharmaceutical Database), KEGG, varDB, 研究支援データベース (Research Support Database), 計算ツール (Computational Tools), and FTP.

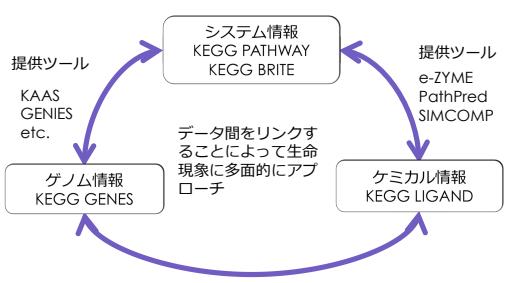
KEGG(Kyoto Encyclopedia of Genes and Genomes)

■ 分子レベルの情報から細胞、個体、エコシステムといった高次生命システムレベルの機能や有用性を理解するためのリソース

■ 遺伝子や化合物などの生体関連物質のデータベースと、それらをつなぐネットワークのデータベースからなる

The screenshot shows the KEGG homepage with a navigation menu on the left and a main content area. Key visible links include: KEGG Home, Release notes, Current statistics, Plea from KEGG, KEGG Database, KEGG Objects, KEGG Software, and KEGG Encyclopedia. The main content area features a large KEGG logo and sections for KEGG Pathway, Brite hierarchies, and Reaction Ontology.

KEGG



KEGGのシステム情報

■ PATHWAY-パスウェイマップ

■ BRITE-機能階層・オントロジー

■ MODULE-機能ユニット

The screenshot shows the KEGG - Table of Content page with a navigation menu at the top and a main content area. The menu includes: Menu, PATHWAY, BRITE, MODULE, KO, GENOME, GEN. The main content area has a search bar and a table titled 'Data-oriented entry points' with columns for Category, Entry Point, and Content. Key entries include: Systems Information (KEGG PATHWAY, KEGG BRITE, KEGG MODULE), Genomic Information (KEGG GENOME, KEGG GENES), and KO (KEGG Orthology) (Functional orthologs).

KEGG PATHWAY

生命システムに関する分子間ネットワーク図のデータベース

□ 1. 代謝系

- 2. 遺伝子制御系
- 3. 環境シグナル
- 4. 細胞プロセス
- 5. 生体システム
- 6. 疾患
- 7. 薬剤開発

KEGG PATHWAY Database
Wiring diagrams of molecular interactions, etc.

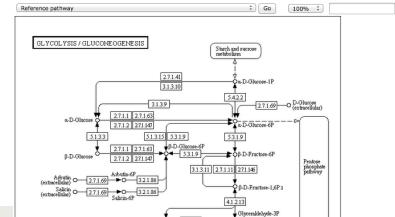
Select prefix: map
Enter keywords: Organism: []
[New pi]

Pathway Maps
KEGG PATHWAY is a collection of manually drawn pathway maps representing molecular interaction and reaction networks for:
1. Metabolism
2. Genetic Information Processing
3. Environmental Information Processing
4. Cellular Processes
5. Organismic Systems
6. Human Diseases
and also on the structure relationships (KEGG drug structure maps) in:
[Drug Development]

KEGG PATHWAY

Glycolysis / Gluconeogenesis - Reference pathway

[Pathway menu / Organism menu] [Entry] [Hide description] [Other data mapping]
The reference pathway involving pathways that are the major source of ATP energy and NADH reducing power. It is a central pathway that provides important precursor metabolites, i.e.-carbon compounds of glucose-6P, fructose-6P and three-carbon compounds of glyceraldehyde-3P, glyceraldehyde-3P, phosphoglycerate, and pyruvate (MD-H00027). When the enzymes genes of this pathway are examined in completely sequenced genomes, the reaction mechanisms of glycolysis/gluconeogenesis are found in all organisms except for some viruses and almost all organisms and which sometimes contains operon structures in bacterial genomes. Gluconeogenesis is a synthesis pathway that uses glycolytic precursors. It is essentially a reversal of glycolysis with minor variations of alternative paths (MD-H00028).



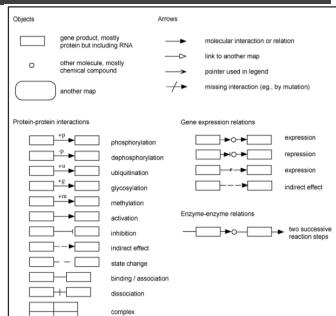
KEGG PATHWAY

□ 箱: 遺伝子 丸: 化合物

□ 矢印:

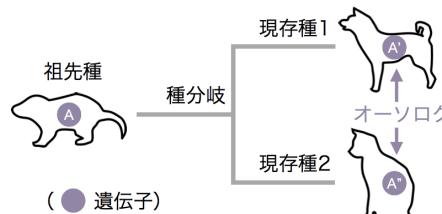
代謝系では代謝反応

それ以外の系ではリン酸化、活性化など様々なタイプがある



KEGG ORTHOLOGY (KO)

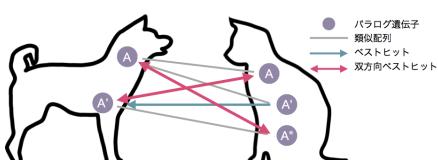
□ オーソログ遺伝子



種分岐によって生じた生物間で対応する遺伝子で、配列相同性を持ち、同一の生物学的機能を持っていると考えられる遺伝子。オーソログ遺伝子を見つけることは機能アノテーションを行う上で重要

KEGG ORTHOLOGY (KO)

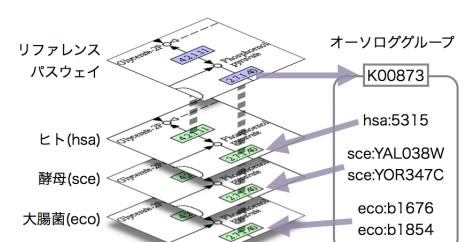
□ KOを介してシステム情報へマッピング



ただ似ているだけでなく、双方向ベストヒットである遺伝子がオーソログの関係にある可能性が高い

KEGG PATHWAY

□ リファレンスパスウェイ



KOと化合物のネットワークであるリファレンスパスウェイを基に、生物種毎のパスウェイが作られる

まずはゲノムネットにアクセスしましょう

□ <http://www.genome.jp/ja/>

GenomeNet

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

□ 生命情報データベース

ゲノムネット 総合データベース
LinkDB search SPARQL エンドポイント
KEGG: 生命システム構成統合データベース
KEGG BRITE: ジャンル構成 バスキュイ
KEGG PATHWAY: システム構成 バスキュイ
KEGG BRITEL: リスト構成 バスキュイ
KEGG MODULE: 全般構成 バスキュイ
KEGG GENES: ゲノム情報
KEGG LIGAND: クラス別構成
KEGG EXPRESSION: 表現子発現情報
KEGG MEDICUS: 薬品・医薬品情報
KEGG MGNEST: メタゲノム情報
Virus-Host DB: ワイルスの生息情報
Taxonomy: 生物分類の早期表 [new!](#)
Reaction Ontology: 反応パターンと分類
LinkDB: 関連データベース
研究者データベース
CYORF: シアノバクテリア
BSORF: バクテリア
EXPRESSION: 運送子発現プロファイル
EST (EGENES)
KEGG OG: リスト構成
KEGG MGNEST: メタゲノム情報
Virus-Host DB: ワイルスの生息情報
Taxonomy: 生物分類の早期表 [new!](#)
Reaction Ontology: 反応パターンと分類
LinkDB: 関連データベース
研究者データベース
CYORF: シアノバクテリア
BSORF: バクテリア
EXPRESSION: 運送子発現プロファイル
EST (EGENES)
KEGG OG: リスト構成

BRITE 構成階層 (日本語)
KEGG バスキュイマップ
KEGG 生物種
生物種ごとのゲノム一覧
データベース (KEGG MGNEST)
メタ (KEGG MGNEST)
ワイルス (Virus)
EST (EGENES)
オーソログラスター (OC)



ゲノムネット計算ツール

配列解析
BLAST / FASTA - ホモジジー検索
MOTIF - モチーフ検索
MAFFT / CLUSTALW / PRIN - 復数アライメント
TREE - 株式樹形図 [new!](#)
KEGG: 生命システム構成データベース
KEGG BRITE: ジャンル構成 バスキュイ
KEGG PATHWAY: システム構成 バスキュイ
KEGG MODULE: 全般構成 バスキュイ
KEGG GENES: ゲノム情報
KEGG LIGAND: クラス別構成
KEGG EXPRESSION: 表現子発現情報
KEGG MEDICUS: 薬品・医薬品情報
KEGG MGNEST: メタゲノム情報
Virus-Host DB: ワイルスの生息情報
Taxonomy: 生物分類の早期表 [new!](#)
Reaction Ontology: 反応パターンと分類
LinkDB: 関連データベース
研究者データベース
CYORF: シアノバクテリア
BSORF: バクテリア
EXPRESSION: 運送子発現プロファイル
EST (EGENES)
KEGG OG: リスト構成
KEGG MGNEST: メタゲノム情報
Virus-Host DB: ワイルスの生息情報
Taxonomy: 生物分類の早期表 [new!](#)
Reaction Ontology: 反応パターンと分類
LinkDB: 関連データベース
研究者データベース
CYORF: シアノバクテリア
BSORF: バクテリア
EXPRESSION: 運送子発現プロファイル
EST (EGENES)
KEGG OG: リスト構成

ゲノムネット計算ツール

□ データ解析用計算ツール

配列解析

BLAST / FASTA - ホモジジー検索

MOTIF - モチーフ検索

MAFFT / CLUSTALW / PRIN - 復数アライメント

TREE - 株式樹形図 [new!](#)

(この例はご利用登録が必要です)

ゲノム構成解析

OC: オーソログラスター

REST サーバー

KEA - KEGG BRITE ナレーシングバー

MAPLE - 代謝経路の構造的計算

MapSearch - 生物種間の経路検索

PathPred - 分析・合成の経路予測

E-Prime - 合成物質の静電反応予測

ケミカル情報解析

DATA: 化合物構造検索

化合物構造検索

REST サーバー

KCM - 格納場活性検索

PathPred - 代謝経路の構造的計算

PathSearch - 部位ごとの経路検索

PathPred - 分析・合成の経路予測

E-Prime - 合成物質の静電反応予測

実習1 : bget/bfindで検索してみましょう

□ bget: GenomeNetのデータベースエントリーID検索

□ bfind: GenomeNetのデータベースに対するキーワード検索

ページ上部のテキスト検索ボックスに、好きなキーワードを入れて検索してみる。DB名:エントリーIDという形式で入力するとbgetモード(例 eco:b0002)、それ以外では、bfindモードで検索が行われる(例 Histidine kinase, Cyanamide等)。

GenomeNet

Search 総合データベース : for cytochrome Go Clear

bfindで"cytochrome"を検索した結果

□ キーワードがヒットする全てのエントリが表示される

GenomeNet

Search : 総合データベース - Search term: cytochrome Go Clear

Database: 総合データベース - Search term: cytochrome

オントロジー

データベース名

KEGG BRITE

ko00001 KO: KEGG Orthology (KO)
ko00002 Module: KEGG modules
ko00003 Photosynthesis; Photosynthesis proteins
ko00019 Cytochrome P450
ko01000 Enzyme, Enzymes
...+ display all

GO

0000277 [GO] ...

実習2 : LinkDBでID対応表を取得しましょう

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

ゲノムネット 総合データベース

LinkDB search SPARQL エンドポイント

KEGG: 生命システム構成統合データベース

KEGG BRITE: ジャンル構成 バスキュイ

KEGG PATHWAY: システム構成 バスキュイ

KEGG MODULE: 全般構成 バスキュイ

KEGG GENES: ゲノム情報

KEGG LIGAND: クラス別構成

KEGG EXPRESSION: 表現子発現情報

KEGG MEDICUS: 薬品・医薬品情報

KEGG MGNEST: メタゲノム情報

Virus-Host DB: ワイルスの生息情報

Taxonomy: 生物分類の早期表

Reaction Ontology: 反応パターンと分類

LinkDB: 関連データベース

研究者データベース

CYORF: シアノバクテリア

BSORF: バクテリア

EXPRESSION: 運送子発現プロファイル

EST (EGENES)

KEGG OG: リスト構成

KEGG MGNEST: メタゲノム情報

Virus-Host DB: ワイルスの生息情報

Taxonomy: 生物分類の早期表

Reaction Ontology: 反応パターンと分類

LinkDB: 関連データベース

研究者データベース

CYORF: シアノバクテリア

BSORF: バクテリア

EXPRESSION: 運送子発現プロファイル

EST (EGENES)

KEGG OG: リスト構成

KEGG MGNEST: メタゲノム情報

Virus-Host DB: ワイルスの生息情報

Taxonomy: 生物分類の早期表

Reaction Ontology: 反応パターンと分類

LinkDB: 関連データベース

研究者データベース

CYORF: シアノバクテリア

BSORF: バクテリア

EXPRESSION: 運送子発現プロファイル

EST (EGENES)

KEGG OG: リスト構成

KEGG MGNEST: メタゲノム情報

Virus-Host DB: ワイルスの生息情報

Taxonomy: 生物分類の早期表

Reaction Ontology: 反応パターンと分類

LinkDB: 関連データベース

研究者データベース

CYORF: シアノバクテリア

BSORF: バクテリア

EXPRESSION: 運送子発現プロファイル

EST (EGENES)

KEGG OG: リスト構成

KEGG MGNEST: メタゲノム情報

Virus-Host DB: ワイルスの生息情報

Taxonomy: 生物分類の早期表

Reaction Ontology: 反応パターンと分類

LinkDB: 関連データベース

研究者データベース

CYORF: シアノバクテリア

BSORF: バクテリア

EXPRESSION: 運送子発現プロファイル

EST (EGENES)

KEGG OG: リスト構成

KEGG MGNEST: メタゲノム情報

Virus-Host DB: ワイルスの生息情報

Taxonomy: 生物分類の早期表

Reaction Ontology: 反応パターンと分類

LinkDB: 関連データベース

研究者データベース

CYORF: シアノバクテリア

BSORF: バクテリア

EXPRESSION: 運送子発現プロファイル

EST (EGENES)

KEGG OG: リスト構成

KEGG MGNEST: メタゲノム情報

Virus-Host DB: ワイルスの生息情報

Taxonomy: 生物分類の早期表

Reaction Ontology: 反応パターンと分類

LinkDB: 関連データベース

研究者データベース

CYORF: シアノバクテリア

BSORF: バクテリア

EXPRESSION: 運送子発現プロファイル

EST (EGENES)

KEGG OG: リスト構成

KEGG MGNEST: メタゲノム情報

Virus-Host DB: ワイルスの生息情報

Taxonomy: 生物分類の早期表

Reaction Ontology: 反応パターンと分類

LinkDB: 関連データベース

研究者データベース

CYORF: シアノバクテリア

BSORF: バクテリア

EXPRESSION: 運送子発現プロファイル

EST (EGENES)

KEGG OG: リスト構成

KEGG MGNEST: メタゲノム情報

Virus-Host DB: ワイルスの生息情報

Taxonomy: 生物分類の早期表

Reaction Ontology: 反応パターンと分類

LinkDB: 関連データベース

研究者データベース

CYORF: シアノバクテリア

BSORF: バクテリア

EXPRESSION: 運送子発現プロファイル

EST (EGENES)

KEGG OG: リスト構成

KEGG MGNEST: メタゲノム情報

Virus-Host DB: ワイルスの生息情報

Taxonomy: 生物分類の早期表

Reaction Ontology: 反応パターンと分類

LinkDB: 関連データベース

研究者データベース

CYORF: シアノバクテリア

BSORF: バクテリア

EXPRESSION: 運送子発現プロファイル

EST (EGENES)

KEGG OG: リスト構成

KEGG MGNEST: メタゲノム情報

Virus-Host DB: ワイルスの生息情報

Taxonomy: 生物分類の早期表

Reaction Ontology: 反応パターンと分類

LinkDB: 関連データベース

研究者データベース

CYORF: シアノバクテリア

BSORF: バクテリア

EXPRESSION: 運送子発現プロファイル

EST (EGENES)

KEGG OG: リスト構成

KEGG MGNEST: メタゲノム情報

Virus-Host DB: ワイルスの生息情報

Taxonomy: 生物分類の早期表

Reaction Ontology: 反応パターンと分類

LinkDB: 関連データベース

研究者データベース

CYORF: シアノバクテリア

BSORF: バクテリア

EXPRESSION: 運送子発現プロファイル

EST (EGENES)

KEGG OG: リスト構成

KEGG MGNEST: メタゲノム情報

Virus-Host DB: ワイルスの生息情報

Taxonomy: 生物分類の早期表

Reaction Ontology: 反応パターンと分類

LinkDB: 関連データベース

研究者データベース

CYORF: シアノバクテリア

BSORF: バクテリア

EXPRESSION: 運送子発現プロファイル

EST (EGENES)

KEGG OG: リスト構成

KEGG MGNEST: メタゲノム情報

Virus-Host DB: ワイルスの生息情報

Taxonomy: 生物分類の早期表

Reaction Ontology: 反応パターンと分類

LinkDB: 関連データベース

研究者データベース

CYORF: シアノバクテリア

BSORF: バクテリア

EXPRESSION: 運送子発現プロファイル

EST (EGENES)

KEGG OG: リスト構成

KEGG MGNEST: メタゲノム情報

Virus-Host DB: ワイルスの生息情報

Taxonomy: 生物分類の早期表

Reaction Ontology: 反応パターンと分類

LinkDB: 関連データベース

研究者データベース

CYORF: シアノバクテリア

BSORF: バクテリア

EXPRESSION: 運送子発現プロファイル

EST (EGENES)

実習2： LinkDBでID対応表を取得しましょう

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

The screenshot shows the LinkDB homepage with the KEGG database selected. A green box highlights the 'Genes' link under the KEGG section. Below the main navigation, there are download links for 'Organism' and 'Pfam' in 'text' or 'format' (CSV) and a 'Download' button.

実習2： LinkDBでID対応表を取得しましょう

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

The screenshot shows the 'Find organism' search interface. A red box highlights the 'Selected' button in the search results. Below the search bar, there are download links for 'Organism' and 'Pfam' in 'text' or 'format' (CSV) and a 'Download' button.

実習2： LinkDBでID対応表を取得しましょう

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

The screenshot shows the LinkDB homepage with the KEGG database selected. A green box highlights the 'Genes' link under the KEGG section. Below the main navigation, there are download links for 'Organism' and 'Pfam' in 'text' or 'format' (CSV) and a 'Download' button.

実習2： LinkDBでID対応表を取得しましょう

■ 遺伝子IDとPfam IDの
対応表が取れる

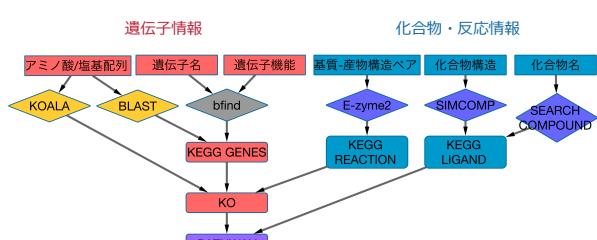
eco_pfam-1.list — 編集済み		
eco:b0001	pf:PF08254	original
eco:b0002	pf:PF00696	original
eco:b0002	pf:PF00742	original
eco:b0002	pf:PF01842	original
eco:b0002	pf:PF03447	original
eco:b0002	pf:PF13840	original
eco:b0003	pf:PF00288	original
eco:b0003	pf:PF08544	original
eco:b0004	pf:PF0291	original
eco:b0004	pf:PF14821	original
eco:b0005	pf:PF10697	original
eco:b0006	pf:PF03883	original
eco:b0007	pf:PF01235	original
eco:b0008	pf:PF00923	original
eco:b0009	pf:PF00994	original
eco:b0010	pf:PF01184	original
eco:b0011	pf:PF03937	original
eco:b0011	pf:PF03981	original
eco:b0011	nf:PF13099	original

次はKEGGにアクセスしましょう

■ <http://www.kegg.jp/ja/>

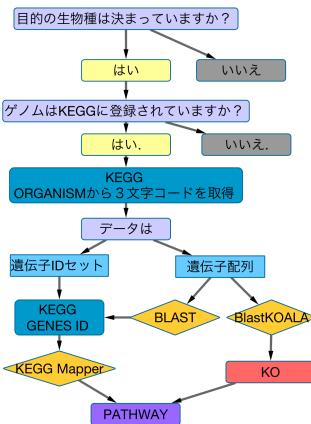
解析の流れ

■ あなたの持っているデータはどのデータですか？



解析の流れ

- ゲノム配列が解読されている生物種の場合



KEGG ORGANISMS

- KEGGにゲノムが登録されている生物種には3文字の生物種コードが与えられている

www.genome.jp/kegg/catalog/org_list.html

KEGG Organisms: Complete Genomes

Eukaryotes: 304 Bacteria: 3393 Archaea: 207

[Genomes | Species | Genus | Draft | Meta]

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

実習3: KEGG Mapperを用いたパスウェイ再構築

www.kegg.jp/kegg/mapper.html

KEGG Mapper

About KEGG Mapper

Search Pathway
Search&Color Pathway
Search&Color Brite
Join Brite
Search Module
Search&Color Module
Annotate Sequence
BlastKOALA
Reconstruct Pathway
Reconstruct Brite
Reconstruct Module
Map Taxonomy

Pathway mapping tool

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.
- Example: hsa,ko,ec,org,hsadd,or org

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)
- Example: hsa,ko,ec,org,hsadd,or org
- NCBI and UniProt identifiers may be used

www.genome.jp/kegg/tool/map_pathway2.html

KEGG Mapper – Search&Color Pathway

生物種 3 文字コード

Search against: hsa Enter: map, ko, ec, rn, hsadd, or org Primary ID: KEGG identifiers ID の種類を選択 Enter objects one per line followed by bgcolor, fgcolor:

Examples: 7167 red,blue C00118 pink

遺伝子や化合物のID 塗りつぶし色,線の色

Alternatively, enter the file name containing the data: ファイルが選択されています。

If necessary, change default bgcolor: pink

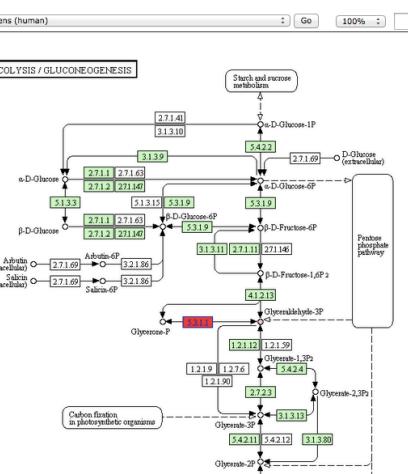
Include aliases Use uncolored diagrams Display objects not found in the search Search pathways containing all the objects (AND search)

Exec Clear

実習3: パスウェイ

結果

指定した色でオブジェクトが塗られている



実習3: KEGG Mapperを用いたパスウェイ再構築

例題

ある研究チームは、大腸菌K-12株の遺伝子「dapA」高発現下で細胞のL-リジン生産量が上がることを発見した。

この結果を説明する代謝経路をKEGG mapperを用いて検索せよ。

自動アノテーションシステム



input: 大規模シーケンスによって得られた生物種のアミノ酸配列や塩基配列

output: 配列名とKOの対応表、PATHWAYマップ



KAASと同様にアミノ酸配列や塩基配列にKOをアサイン

SEARCHを利用。KEGG内部でKOのアサインにはKOALAを利用

実習4: BlastKOALAを用いた自動アノテーション及pathway再構築

www.kegg.jp/kegg/mapper.html

KEGG Mapper

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

About KEGG Mapper

KEGG Mapper is a collection of tools for KEGG mapping. KEGG pathway mapping tools: Search, 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 tools have search and mapping features. 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.)
- Mutual search objects may be specified in one line together with color attributes (background color and foreground color).
- Multiple objects can be specified in one line.

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).
- Multiple objects can be specified in one line.

NCBI and UniProt identifiers may be used

www.kegg.jp/kegg/tool/annotate_sequence.html

KEGG Mapper – Annotate Sequence by BlastKOALA

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

Upload query amino acid sequences in FASTA format

Enter FASTA sequences

Or upload file: [選択...] ファイルが選択されていません。

Select GENES family/genus dataset to be searched

Example: Query data: sequence.txt
Select Buchnera (32199)

リンク先を全選択してコピー

Select Family/Genus:

Exec Clear

Annotate Sequence is an interface to the BlastKOALA server for assigning KO identifiers (K numbers) to a given sequence data set for subsequent analyses with Reconstruct Pathway, Reconstruct Brite, and Reconstruct Module tools. The BlastKOALA computation is performed in an interactive mode using an appropriate subset of KEGG GENES (corresponding to family/genus of your organism). This tool is useful for converting your sequence IDs to K numbers.

www.kegg.jp/kegg/tool/annotate_sequence.html

KEGG Mapper – Annotate Sequence by BlastKOALA

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

Upload query amino acid sequences in FASTA format

Enter FASTA sequences
>[c]NZ_ACNKO1000001_1.gntr_WP_009873963_1 [Gene=C55_RS0100010] [protein_id=WP_009873963_1] [location=202..208] MTNLNRDVIVVVGAGHIGTEAMAMASSRGMGCKTLLTQKISDGLASCNPAGIGGKSHLVKE IDALGGHHM AMVAVVAGGQDFPILJLSSSKGPVRTRAQOKLYIETYVKKILKKQNNLLEAEVKDLFLKYN5 VVGVLT QNEINYSSRSVVLAGTFLGKIHIGLKSYSACRGDKQSAISDLSVRLRELSLVNLKLTGTPP RIDINTV NPNLNLNSDTPVPVFSPGMNVSHHPKOPCPYLTHTNEKTHEIRKWLDKSPVNTGPKLGKLG Or upload file: [選択...] ファイルが選択されていません。

Select GENES family/genus dataset to be searched

Example: Query data: sequence.txt
Select Buchnera (32199)

Exec Clear

Annotate Sequence is an interface to the BlastKOALA server for assigning KO identifiers (K numbers) to a given sequence data set for subsequent analyses with Reconstruct Pathway, Reconstruct Brite, and Reconstruct Module tools. The BlastKOALA computation is performed in an interactive mode using an appropriate subset of KEGG GENES (corresponding to family/genus of your organism). This tool is useful for converting your sequence IDs to K numbers.

www.kegg.jp/kegg-bin/list_taxonomy/mode=select

今 Buchnera

			Yersinia rohdei	29485
			Yersinia ruckeri	29486
			Shigella flexneri	623
			Shigella sonnei	624
			Shigella dysenteriae	625
			Pectobacterium atrosepticum	29471
			Pectobacterium carotovorum	554
			Pectobacterium wasabiae	55208
			Pectobacterium sp. SCGI193	1166016
			Erwinia tasmaniensis	338565
			Erwinia pytorfana	799907
			Erwinia herbicola	553
			Erwinia bilogiae	182337
			Erwinia sp. Etp217	215669
			Photobacter 29487	29488
			Photobacter luminescens	291112
			Photobacter symbiotica	291112
			Buchnera aphidicola	9
			Wiggeworthia glossinifolia	31229
			Sodalis sp. 851	63912
			Sodalis sp. H51	1239307
			Candidatus Sodalis peramtonius	1486991
			Enterobacter sp. 638	399742
			Enterobacter cloacae	950
			Enterobacter laevulans	1134193
			Chromobacter sakazakii	28141
			Esco bacter	1649298
			Esco bacter turicensis	4135902
			Klebsiella pneumoniae	573
			Klebsiella variicola	244366
			Klebsiella oxytoca	571
			Klebsiella michiganensis	1134687
			Chlorobacter koren	945

今 Buchnera

すべて強調表示 大文字/小文字を強調 件中 1 件目

実習4: 自動アノテーション

今少し待っていると画面が自動で切り替わる

www.kegg.jp/kegg-bin/blastkoala_result?id=63ad1534d877b373e512a155

Annotate Sequence Result

Your Annotate Sequence job

Query dataset: 575 entries
KEGG database searched: 32199.pep
Job submitted: Tue Jun 16 08:00:44 JST 2015
Job completed: Tue Jun 16 08:01:26 JST 2015

Annotation data: View Download

Summary: 551 entries (95.8%) annotated

Functional category: Unclassified 159, Genetic Information Processing 32, Amino Acid metabolism 31, Carbohydrate metabolism 40, Nucleotide metabolism 39, Energy metabolism 47, Metabolism of cofactors and vitamins 120, Cellular Processes 3, Environmental Information Processing 1, Others 1

KEGG Mapper Reconstruct Pathway Reconstruct Brite Reconstruct Module

[Annotate Sequence | KEGG | Kanehisa Labs]

実習4: 自動アノテーション

Annotation query IDs Assigned KOs

```

[{"id": "KEGG", "query": "ACKP01000001_1_prot_WP_00873983_1_1_(620)", "ko": "K0369 gyA; RNA uridyl 5'-trisopropylbenzimidothio modification enzyme"}, {"id": "KEGG", "query": "ACKP01000001_1_prot_WP_00873994_1_2_(274)", "ko": "K0210 ATPC; F-type H+-transporting ATPase subunit a"}, {"id": "KEGG", "query": "ACKP01000001_1_prot_WP_00873995_1_3_(79)", "ko": "K0210 ATPC; F-type H+-transporting ATPase subunit c"}, {"id": "KEGG", "query": "ACKP01000001_1_prot_WP_00873996_1_4_(161)", "ko": "K0210 ATPC; F-type H+-transporting ATPase subunit b"}, {"id": "KEGG", "query": "ACKP01000001_1_prot_WP_00873997_1_5_(177)", "ko": "K0211 ATPD; F-type H+-transporting ATPase subunit delta"}, {"id": "KEGG", "query": "ACKP01000001_1_prot_WP_00873998_1_6_(512)", "ko": "K0211 ATPD; F-type H+-transporting ATPase subunit alpha"}, {"id": "KEGG", "query": "ACKP01000001_1_prot_WP_00873999_1_7_(109)", "ko": "K0211 ATPD; F-type H+-transporting ATPase subunit epsilon"}, {"id": "KEGG", "query": "ACKP01000001_1_prot_WP_00873970_1_8_(445)", "ko": "K0211 ATPD; F-type H+-transporting ATPase subunit beta [EC:3.6.3.14]"}, {"id": "KEGG", "query": "ACKP01000001_1_prot_WP_02585314_1_9_(140)", "ko": "K0211 ATPD; F-type H+-transporting ATPase subunit epsilon"}, {"id": "KEGG", "query": "ACKP01000001_1_prot_WP_00873971_1_10_(803)", "ko": "K02470 gyB; DNA gyrase subunit B [EC:5.9.1.3]"}, {"id": "KEGG", "query": "ACKP01000001_1_prot_WP_00873971_1_11_(364)", "ko": "K02338 DPO3; DNA polymerase III subunit beta [EC:2.7.7.7]"}, {"id": "KEGG", "query": "ACKP01000001_1_prot_WP_00873971_1_12_(454)", "ko": "K02338 DPO3; DNA polymerase III subunit epsilon; chaperone/restriction initiator protein"}, {"id": "KEGG", "query": "ACKP01000001_1_prot_WP_00873971_1_13_(454)", "ko": "K02338 DPO3; DNA polymerase III subunit epsilon; chaperone/restriction initiator protein L34"}, {"id": "KEGG", "query": "ACKP01000001_1_prot_WP_00873977_1_14_(532)", "ko": "K03211 ygcC; YgcC (Oxidative family membrane protein), chaperone GroES"}, {"id": "KEGG", "query": "ACKP01000001_1_prot_WP_00873977_1_15_(453)", "ko": "K03650 mmE; tRNA modification GTPase [EC:3.6.-.-]"}, {"id": "KEGG", "query": "ACKP01000001_1_prot_WP_00873979_1_16_(96)", "ko": "K04079 gnsE; chaperonin GroEL"}, {"id": "KEGG", "query": "ACKP01000001_1_prot_WP_00873980_1_17_(546)", "ko": "K04079 gnsE; chaperonin GroEL"}, {"id": "KEGG", "query": "ACKP01000001_1_prot_WP_00873980_1_18_(569)", "ko": "K04079 gnsE; chaperonin GroEL"}, {"id": "KEGG", "query": "ACKP01000001_1_prot_WP_00873982_1_19_(246)", "ko": "K02311 dnaC; DNA replication protein DnaC"}, {"id": "KEGG", "query": "ACKP01000001_1_prot_WP_00873983_1_20_(164)", "ko": "K02311 dnaT; DNA replication protein DnaT"}, {"id": "KEGG", "query": "ACKP01000001_1_prot_WP_00873984_1_21_(191)", "ko": "K03164 rsmD; 16S rRNA (guanine-N2)-methyltransferase [EC:2.1.1.171]"}, {"id": "KEGG", "query": "ACKP01000001_1_prot_WP_00873984_1_22_(378)", "ko": "K03110 ftsY; fused signal recognition particle receptor"}, {"id": "KEGG", "query": "ACKP01000001_1_prot_WP_00873984_1_23_(454)", "ko": "K02933 S023.3; RNA polymerase sigma-32 factor"}, {"id": "KEGG", "query": "ACKP01000001_1_prot_WP_00873984_1_24_(609)", "ko": "K02082 gntR; glucosamine-fructose-6-phosphate amino transferase (eximerizing) [EC:2.6.1.]"}, {"id": "KEGG", "query": "ACKP01000001_1_prot_WP_00873988_1_25_(459)", "ko": "K00404 gntR; bifunctional UDP-N-acetylglucosamine pyrophosphorylase / Glucosamine-1-glyc", {"id": "KEGG", "query": "ACKP01000001_1_prot_WP_00873988_1_26_(272)", "ko": "K07024"}, {"id": "KEGG", "query": "ACKP01000001_1_prot_WP_00873990_1_27_(75)", "ko": "K00549 metT; 5-methyltetrahydropteroylglutamate-homocysteine methyltransferase [EC:2.1.1.17]"}, {"id": "KEGG", "query": "ACKP01000001_1_prot_WP_00873991_1_28_(52)", "ko": "K00602 purN; phosphoribosylaminoimidazolecarboxamide formyltransferase / IMP cyclhydratase"}, {"id": "KEGG", "query": "ACKP01000001_1_prot_WP_00873991_1_29_(169)", "ko": "K02879 nupA; DNA-binding protein Nup-alpha"}, {"id": "KEGG", "query": "ACKP01000001_1_prot_WP_00873993_1_30_(81)", "ko": "K02311 dntA; DNA-directed RNA polymerase subunit beta' [EC:2.7.7.6]"}, {"id": "KEGG", "query": "ACKP01000001_1_prot_WP_00873994_1_31_(1407)", "ko": "K03040 rpsC; DNA-directed RNA polymerase subunit beta' [EC:2.7.7.6]"}, {"id": "KEGG", "query": "ACKP01000001_1_prot_WP_00873995_1_32_(1342)", "ko": "K03040 rpsB; DNA-directed RNA polymerase subunit beta' [EC:2.7.7.6]"}, {"id": "KEGG", "query": "ACKP01000001_1_prot_WP_00873996_1_33_(122)", "ko": "K02931 RP-L7; large subunit ribosomal protein L7/L12"}, {"id": "KEGG", "query": "ACKP01000001_1_prot_WP_00873997_1_34_(165)", "ko": "K02864 RP-L10; large subunit ribosomal protein L10"}, {"id": "KEGG", "query": "ACKP01000001_1_prot_WP_00873998_1_35_(231)", "ko": "K02863 RP-L11; large subunit ribosomal protein L11"}]

```

実習4: 自動アノテーション

Reconstruct Pathway をクリック

実習4: BlastKOALA自動アノテーション

Pathway Reconstruction Result

Show all objects

Metabolism

- Global and overview maps
- 01100 Metabolic pathways (206)
- 01110 Biosynthesis of secondary metabolites (98)
- 01120 Biosynthesis of antibiotics (10)
- 01130 Biosynthesis of antibiotics (69)
- 01200 Carbon metabolism (30)
- 01210 2-Oxocarboxylic acid metabolism (10)
- 01220 2-Oxocarboxylic acid metabolism (19)
- 01230 Biosynthesis of amino acids (62)
- Carbohydrate metabolism
- 00000 Carbohydrate metabolism (15)
- 00020 Citrate cycle (TCA cycle) (5)
- 00030 Pentose phosphate pathway (1)
- 00040 Pyruvate metabolism (1)
- 00051 Fructose and mannose metabolism (7)
- 00052 Galactose metabolism (1)
- 00059 Starch and sucrose metabolism (2)
- 00060 Glycolysis/gluconeogenesis (9)
- 00062 Pyruvate metabolism (5)
- 00063 Glycolysis and dicarboxylate metabolism (2)
- 00064 Pyruvate metabolism (2)
- 00065 Butanoate metabolism (2)
- 00066 C3-branched dicarboxylic acid metabolism (2)
- 00062 Inositol phosphate metabolism (2)
- Energy metabolism
- 00190 Oxidative phosphorylation (27)
- 00193 Photosynthesis (8)
- 00201 Carbon fixation by photosynthetic organisms (7)
- 00723 Carbon fixation pathways in prokaryotes (4)
- 00668 Methane metabolism (8)
- 00020 Sulfur metabolism (9)
- 00061 Fatty acid biosynthesis (5)
- 00564 Glycerophospholipid metabolism (1)
- 00230 Biosynthesis of unsaturated fatty acids (2)
- Nucleotide metabolism
- 00230 Purine metabolism (29)
- 00240 Pyrimidine metabolism (30)

実習4: BlastKOALA自動アノテーション

AssignされたKOに相当する遺伝子産物のノードが色付けされている