

化合物データベース・パスウェイ データベースの紹介

統合データベース講習会 AJACSこまち
2016/8/9 9:30-11:30

国立遺伝学研究所 DDBJセンター
日本DNAデータバンク アノテーター

時松 敏明
(tokimatu@nig.ac.jp)

2016/8/9

1

本日の内容

- パスウェイデータベースの紹介
 - パスウェイデータベースの概要
 - BioCyc の紹介・使い方
 - PMN の紹介
 - KEGG PATHWAY の紹介・使い方
- 化合物データベースの紹介
 - 化合物データベースの概要
 - PubChem の紹介・使い方
 - KEGG COMPOUND/GenomeNet の紹介・使い方
 - KNApSACk の紹介・使い方

2016/8/9

2

パスウェイデータベース

2016/8/9

3

はじめに

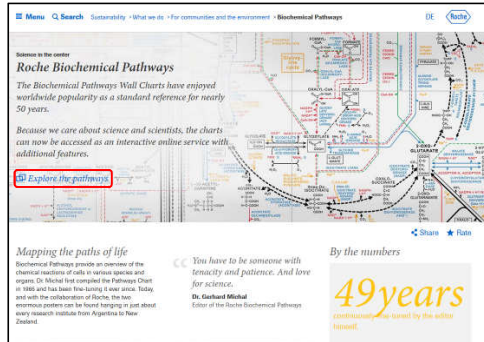
- パスウェイデータベースに関しては、過去のAJACSで
繰り返し講習がされています。過去の資料もぜひご参
照ください
 - 昨年度、本年度のパスウェイデータベースの講習
 - AJACS53:パスウェイデータベース KEGG/GenomeNetのWebサービ
スの紹介
 - AJACS54:パスウェイ情報を中心としたKEGG/GenomeNet Webサービ
スの紹介
 - AJACS55:パスウェイデータベースの紹介
 - AJACS58:パスウェイデータベースの紹介とKEGG PATHWAYの使い方
 - AJACS60:パスウェイデータベースの紹介とKEGG PATHWAYの使い方
 - 今回は化合物DBと合わせての紹介になりますので、
AJACS60(守屋講師)の内容をベースに要約して講習を行いま
す。一部簡略化していますので、AJACS60などの資料も御
参照ください。

2016/8/9

4

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、タンパク質、その他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。パスウェイを集積して可視化、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。



ペーリンガーマンハイム社(現ロシュ社)の代謝マップ(1965～)
(紙媒体で出版、のちに電子化)

2016/8/9

5

パスウェイデータベースとは



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、生命現象を理解しやすくなります。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

今回の講習は、パスウェイデータベースへのマッピングを中心に実習します。

2016/8/9

6

いろいろなパスウェイデータベース

Pathguide (<http://www.pathguide.org/>)

- パスウェイリソースのリスト約 550 (2013)
 - Availability: 有料か無料か
 - Standards: 標準データ形式(BioPAX, SBML等)に準拠しているかどうか



2016/8/9

7

いろいろなパスウェイデータベース

- 歴史的には代謝経路の表現から始まりましたが、現在ではタンパク質間相互作用、シグナル伝達系、遺伝子制御、環境シグナルなど様々な生命現象がパスウェイとして表現されています。
- Pathguide での分類
 - 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・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

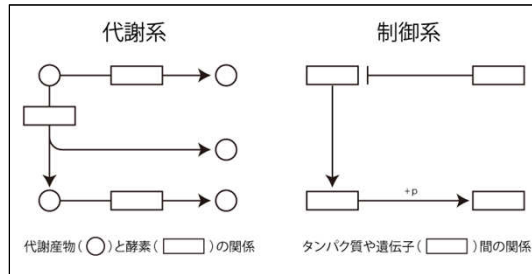
2016/8/9

8

パスウェイデータベースの表現

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くなっています

- 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



2016/8/9

9

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くなっています

- [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 [Cyroscape](#) や [VisANT](#) などがあります

– [AJACS58: Cytoscapeを使ったデータの可視化](#)

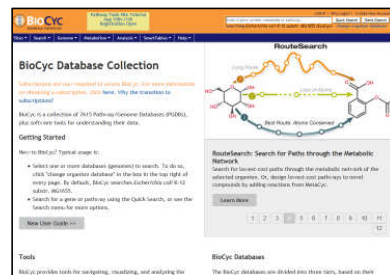
どのパスウェイデータベースを研究に使えば良いかは、対象生物や対象パスウェイ、目的によって異なってきます。今回は BioCyc, KEGG PATHWAY のブラウザ上での使い方を紹介します

2016/8/9

10

BioCyc (<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(→ 要 subscription ?)
- データ形式: BioPAX

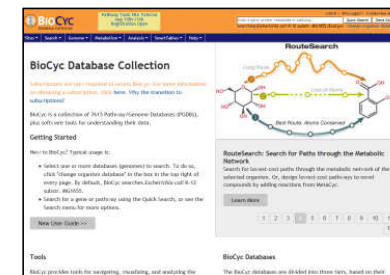


2016/8/9

11

BioCyc 実習1

- 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(→ 要 subscription ?)
- データ形式: BioPAX



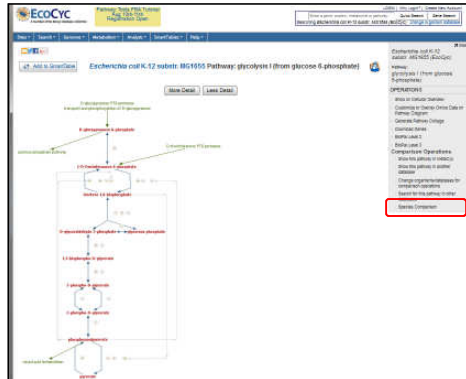
2016/8/9

12

BioCyc 実習3

種間比較をしてみよう

- 右の Option メニューの [Species Comparison](#) をクリック

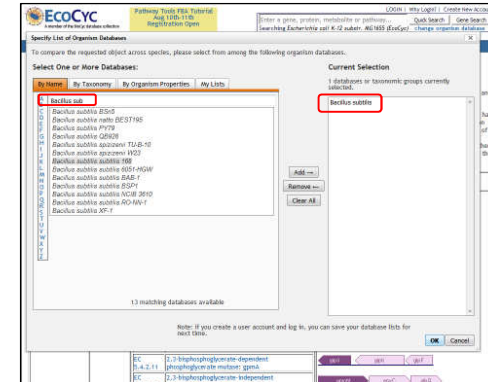


2016/8/9

17

BioCyc 実習3

- 比較する種の選択画面が表示されるので、比較したい好きな種を入力し、OK をクリック (例: *Bacillus subtilis*)
- パスウェイや遺伝子、オペロン構造などが比較できます



2016/8/9

18

BioCyc 実習4

好きな生物の Overview パスウェイを見てみよう

- 上のメニューの Metabolism > [Cellular Overview](#) をクリック

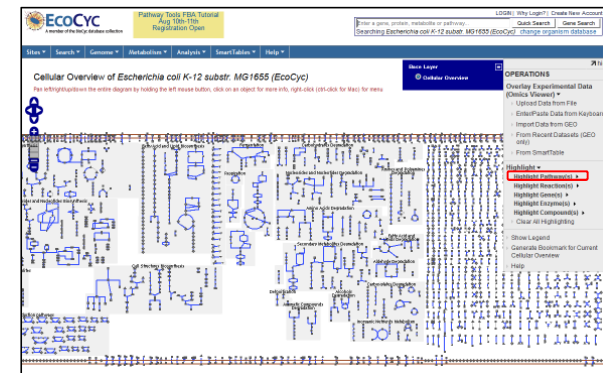
The screenshot shows the BioCyc interface with the 'Cellular Overview' option highlighted in the top navigation bar. The main content area displays a complex metabolic pathway diagram for 'Escherichia coli K-12 substr. MG1655 (EcoCyc)'.

2016/8/9

19

BioCyc 実習4

パスウェイや反応、遺伝子、酵素、化合物の名前や ID で Overview パスウェイをハイライト (例: TCA cycle)

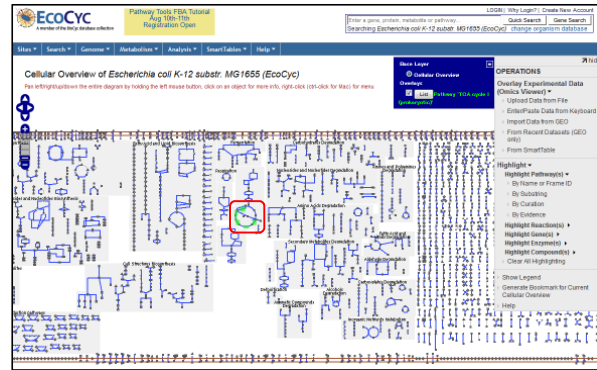


2016/8/9

20

BioCyc 実習4

パスウェイや反応、遺伝子、酵素、化合物の名前やID で Overview パスウェイをハイライト (例: TCA cycle)



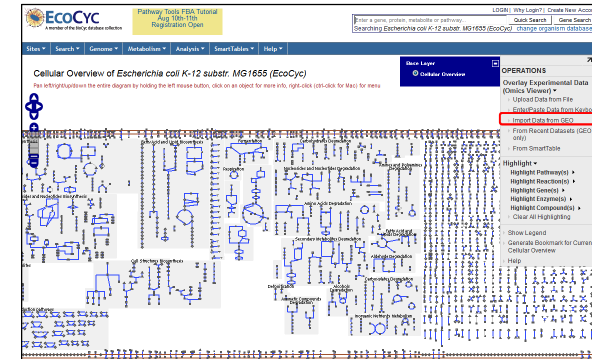
2016/8/9

21

BioCyc 実習5

遺伝子発現データをマッピングしてみよう

- NCBI GEO (Gene Expression Omnibus)のデータを直接マッピングできます
- 右のメニューの Import Data from GEO をクリック

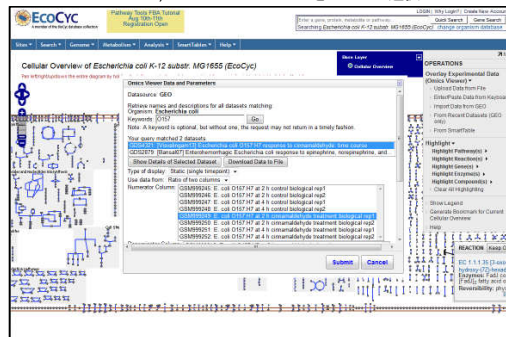


2016/8/9

22

BioCyc 実習5

- キーワード検索で GEO のデータを検索し、データ一つ選択 (例: O157)
- マッピングオプション例
 - Type of display: Static (single timepoint)を選択
 - Use data form: Ratio of two columnsを選択
 - Numerator Column, Denominator Column を一つずつ選択

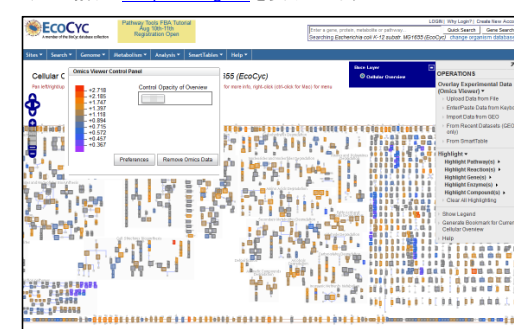


2016/8/9

23

BioCyc 実習5

- 例) 病原性大腸菌O157のシンナムアルデヒド応答、2時間目と4時間目の遺伝子発現の割合
 - この例の場合、オレンジ-赤で示された遺伝子は発現が増え、青-紫で示された遺伝子は発現が減っています
- 余裕のある人は、いろいろなオプション(アニメーション、テーブル出力など)を試してみましょう
- もっと詳しく知りたい場合は [BioCyc User's guide](#) を参照しましょう



2016/8/9

24

PMN (<http://www.plantcyc.org/>)

- 開発: S. Rhee, [Department of Plant Biology, Carnegie Institution, USA](#)
- 対象: 350以上の植物情報からなる植物リファレンスパスウェイ+シロイヌナズナなど22種+外部6種
 - SRI International のPathway Tools Software を使用
 - 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



2016/8/9

25

KEGG PATHWAY (<http://www.kegg.jp/>)

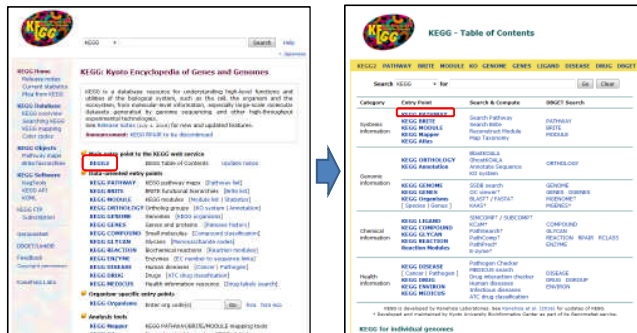
- 開発: 京都大学
- 対象: ゲノムの決まった全生物種(異株を含む) 4,200 種以上 (>300真核生物、>3,700真正細菌、>220古細菌)、真核ドラフトゲノム 25種、環境メタゲノム 300サンプル、生体メタゲノム 700サンプル
 - 専門家が手作業で作成した文献ベースのデータ+自動ツール
 - リファレンスパスウェイ: 専門家が手作業で文献ベースから作成
 - 生物種パスウェイ
 - 自動ツールでリファレンスパスウェイから作成し、手作業でキュレーション
 - 自動ツールで作成(自動ツールにも段階があります)
 - 代謝パスウェイ、シグナル伝達系、他
- 利用: アカデミックフリー
- 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

2016/8/9

26

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
- 今回は PATHWAY, COMPOUND(後述) 以外の詳細は省くので、それ以外の詳細は過去の AJACS 資料を参照 (AJACS50, AJACS51 (付録資料), AJACS53, AJACS54)



2016/8/9

27

KEGG PATHWAY 実習1

対象生物種を見てみよう

- データベースのテーブルの下、[KEGG organisms](#) をクリック



- KEGG では 3-4 文字の独自の生物種コードを使用している



2016/8/9

28

KEGG PATHWAY 実習1

- 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)



2016/8/9

29

KEGG PATHWAY データベースリスト

- 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: 342 Bacteria: 3778 Archaea: 229		
Genomes Species Genus Draft 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 abelli (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

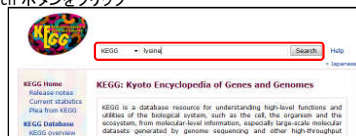
2016/8/9

30

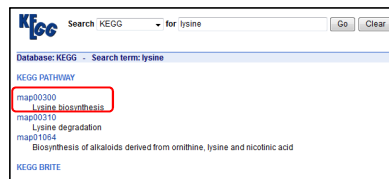
KEGG PATHWAY 実習2

パスウェイマップを見てみよう

- トップページ** 上方の検索ボックスで "lysine biosynthesis" や "glycolysis" 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



- 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



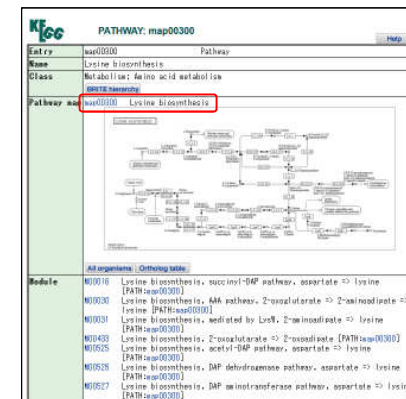
2016/8/9

31

KEGG PATHWAY 実習2

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

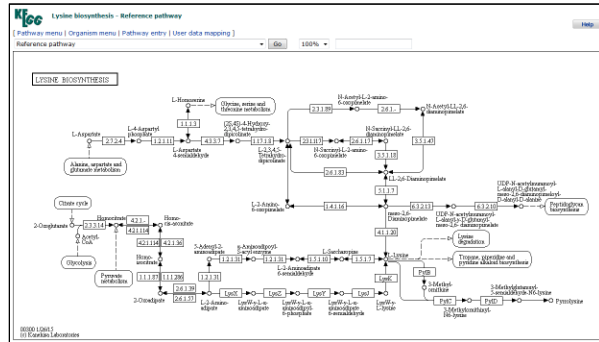


2016/8/9

32

KEGG PATHWAY 実習2

- 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

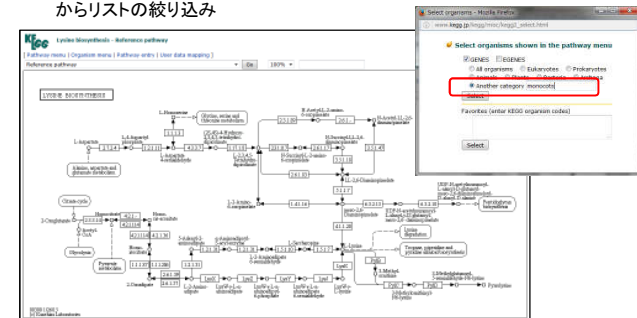


2016/8/9

33

KEGG PATHWAY 実習3

- 好きな生物のパスウェイを見てみよう
- 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



2016/8/9

34

KEGG PATHWAY 実習3

- 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 Draft Genomes		
Eukaryotes		
Category	Organisms	Source
Animal	110012 (Homo sapiens)	Genome Project
Animal	110013 (Homo sapiens)	Genome Project
Plant	110014 (Arabidopsis thaliana)	Genome Project
Plant	110015 (Arabidopsis thaliana)	Genome Project
Algae	110016 (Chlamydomonas reinhardtii)	Genome Project
Fungi	110017 (Saccharomyces cerevisiae)	Genome Project
Fungi	110018 (Saccharomyces cerevisiae)	Genome Project
Basidiomycetes	110019 (Phanerochaete chrysosporium)	Genome Project

KEGG Draft Genomes		
Eukaryotes		
Category	Organisms	Source
Animal	110012 (Homo sapiens)	Genome Project
Animal	110013 (Homo sapiens)	Genome Project
Plant	110014 (Arabidopsis thaliana)	Genome Project
Plant	110015 (Arabidopsis thaliana)	Genome Project
Algae	110016 (Chlamydomonas reinhardtii)	Genome Project
Fungi	110017 (Saccharomyces cerevisiae)	Genome Project
Fungi	110018 (Saccharomyces cerevisiae)	Genome Project
Basidiomycetes	110019 (Phanerochaete chrysosporium)	Genome Project

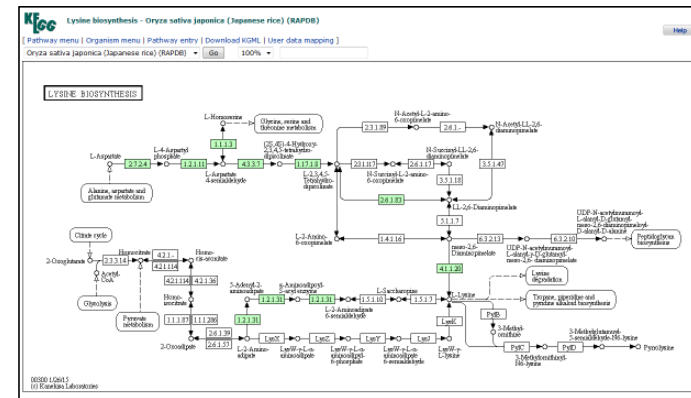
KEGG 110013	
Genome info	Pathway map
Search genes:	
KEGG pathway maps	
Metabolism	
Global and overview maps:	
01100 Metabolic pathways	
01110 Biosynthesis of secondary metabolites	
01200 Carbon metabolism	
01210 2-Oxocarboxylic acid metabolism	
01212 Fatty acid metabolism	
01220 Biosynthesis of amino 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	
00050 Amino sugar and nucleoside sugar metabolism	
00060 Inositol metabolism	
00060 Glyoxylate and dicarboxylate metabolism	
00060 Propanoate metabolism	
00060 Butyrate metabolism	
00060 CS-branching amino acid metabolism	
00062 Insulin-like growth factor metabolism	

2016/8/9

35

KEGG PATHWAY 実習3

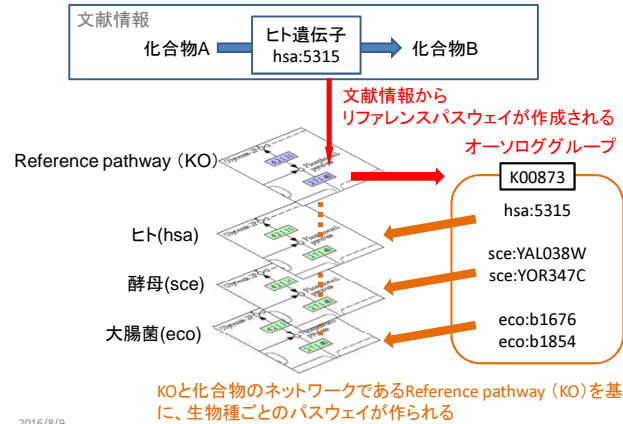
- 一部のボックスが緑色で塗られる、その生物(またはサンプル)の持つ遺伝子を示しています



2016/8/9

36

リファレンスパスウェイと種毎のパスウェイの関係



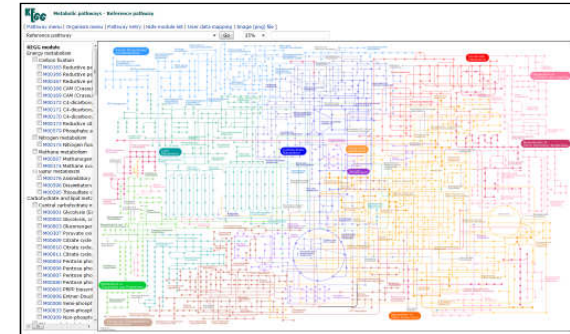
2016/8/9

37

KEGG PATHWAY 実習4

Overview マップを見てみよう

- PATHWAY top page の 1.0 Global and overview maps の [Metabolic pathways](#) をクリック
 - 右の [\[KEGG Atlas\]](#) は Java で動くビューワーで、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

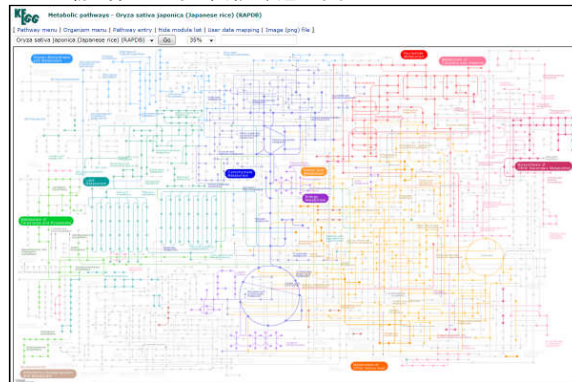


2016/8/9

38

KEGG PATHWAY 実習4

- 生物種毎の Overview マップを見てみよう
 - 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



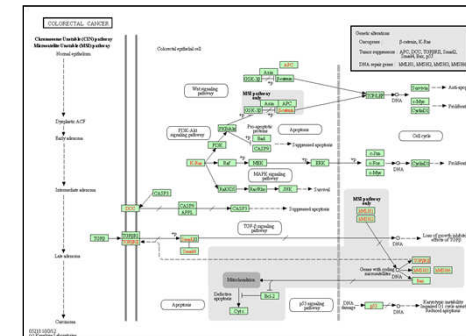
2016/8/9

39

KEGG PATHWAY 実習5

ヒトの疾患パスウェイを見てみよう

- 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



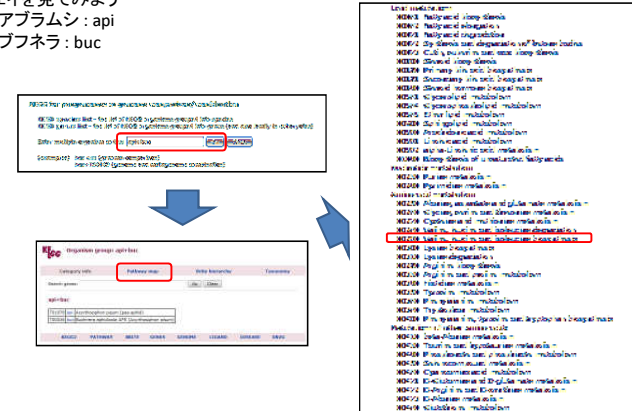
2016/8/9

40

KEGG PATHWAY 実習6

- アラムシとブフネラで、00290 : Valine, leucine and isoleucine biosynthesis の主観比較パスウェイを見てみよう
- アラムシ: api
 - ブフネラ: buc
- LEU1 isoform 1a:

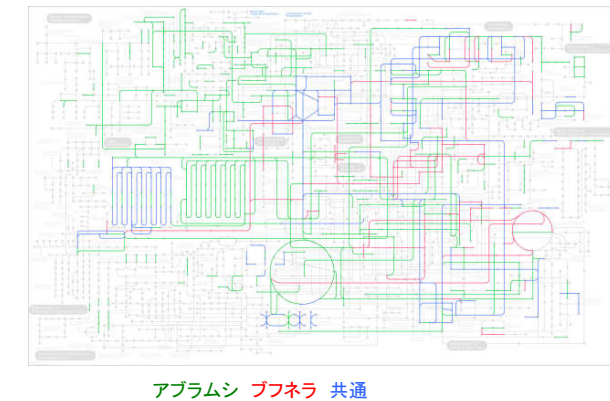
 - MDH3: N-acetylserine transaminase
 - MDH4: N-acetylserine lyase
 - MDH5: N-acetylserine aminotransferase
 - MDH6: Glyoxylate synthetase
 - MDH7: Glutamate decarboxylase



KEGG PATHWAY 実習6

Overviewマップで、共生生物間の補完関係を見ることができる

- アブラムシ : api(緑)、ブフネラ : buc(ピンク)
- 3生物以上でも比較可能



KEGG には遺伝子・化合物リストからパスウェイにマッピングするツールが組み込まれています

- [KEGG Mapper](#)
- 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-GeneID, 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-GeneID, 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 - SearchColor Pathway

About KEGG Mapper

Search Pathway

 Color Pathway (SwissL)

Search Brn

Search Hmbr

Annotation Sequence

Approximate Pathway

Search against: Brn: map, ks, wu, rs, hmd, or

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

Enter objects one per line followed by **logcolor**, **logcolor**:

Examples
 Select

Alternatively, enter the file name containing the data:

45

KEGG PATHWAY 実習7

- 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)

[illegible]

46

KEGG PATHWAY 実習7

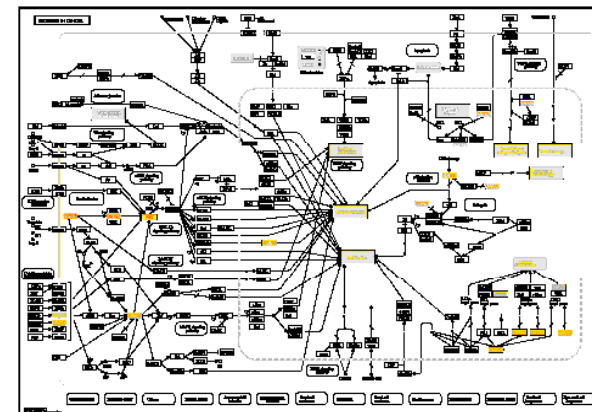
- **数値データをマッピングしてみよう**
- [Color Pathway](#) をクリック
- 右のサンプル CML-COSMIC をダウンロード
 - 中身は配列 ID と数値の対応リスト
- Select KEGG pathway map: でパスウェイを指定 (hsa05200)
- Enter file name containing the data: でダウンロードしたサンプルファイルを指定
- File type: で Numerical value を選択
- Exec ボタンをクリック

[illegible]

47

KEGG PATHWAY 実習7

数値がグラデーションになってマッピングされる



48

化合物データベース

2016/8/9

49

化合物データベース ～Introduction～

- 歴史は長い、有償の文化
 - Chemical Abstracts (1907～)[<http://www.cas-japan.jp/>]
 - ACS(米国化学会)の下部組織、CASによる発行(有償)
 - 化学関連文献・特許、化学構造・反応など総合的な情報を含む
 - CAS登録番号(CAS RN)は化合物IDとして事実上の標準
 - 現在、SciFinderによるオンライン検索が主体
 - 冊子体、CD、オンライン検索がある
 - The Combined Chemical Dictionary (CRC Press)
 - 下記4種の辞典を収載
 - Dictionary of Natural Products (DNP)
 - Dictionary of Organic Compounds
 - Dictionary of Inorganic/Organometallic Compounds
 - Dictionary of Drugs
 - 冊子体、CD(以上DNPのみ)、オンライン検索がある
 - CHEMnetBASEによるオンライン検索(<http://www.chemnetbase.com/>)
 - 日本では、いずれも化学情報協会(JAICI)の取扱い
- 遺伝子データベースの発展とともに、化合物データベースにおいても様々な無償のデータベースが作られる

2016/8/9

50

化合物データベース

KEGG LIGAND	www.genome.jp/kegg/ligand.html	京大・金久研。代謝産物を中心とした化合物、代謝反応など。
KNAPSAcK	kanaya.naist.jp/KNAPSAcK/	奈良先端大・金谷研。天然物・生物情報の対応が軸。植物中心
J-GLOBAL	jglobal.jst.go.jp/advancedsearch/#t=4	JST。日化辞(有機化合物DB)のWeb検索が可能(旧・日化辞Web)。
DNP	dnp.chemnetbase.com/	CRC出版。天然物DB。詳細閲覧は有償。
UNPD	pxxxjpk.edu.cn/UNPD/	北京大。天然物DB。件数多いが、出典不明なものも。
ChEBI	www.ebi.ac.uk/chebi/	EBI。化合物分類DB。分類体系は、現状化学構造寄り。
PubChem	pubchem.ncbi.nlm.nih.gov/	米国・NIH。化合物など。他の化合物DBのレポジトリ的性質。件数膨大。
ChemSpider	www.chemspider.com/	400以上の化合物DBのIDを串刺し検索できるサイト。データ更新遅め。
HMDB	www.hmdb.ca/	カナダ・アルバート大。ヒトメタボロームDB。
ECHDB	www.ecmdb.ca/	同上。大腸菌メタボロームDB。
YMDB	www.ymdb.ca/	同上。酵母メタボロームDB。
DrugBank	www.drugbank.ca/	同上。薬物のDB。
KEGG DRUG	www.kegg.jp/kegg/drug/	京大・金久研。医薬品DB。構造分類、添付文書とのリンクなども。
LIPIDMAPS	www.lipidmaps.org/	米国脂質学会。脂質DB。ポリフェノールなども含む。
Flavonoid Database	metabolomics.jp/wiki/Category:FL	metabolomics.jp(遺伝研・有田研)内にあるWikiベースのフラボノイドDB。Flavonoid ViewerのデータをベースにWiki化。

2016/8/9

(AJACS安芸 メタボローム解析(2016/7/6 櫻井講師)の資料を基に一部改変)

51

化合物構造を表記する書式

- 行列表記
 - mol (MDL/molfile)
 - 結合を結合表としてあらわす。二次元構造式の標準形式
 - sdf (sdf file)
 - mol形式を内包し1ファイルに複数化合物を記述可能にしたもの
 - 詳しくは CTfile format (最新版は BIOVIAから) を参照
- 線形表記
 - InChI、InChIKey
 - IUPACとNISTが開発、現在はIUPAC管理、IUPACとInChI Trustで維持
 - standard InChIでは、構造式から一意なInChI文字列生成可能
 - ほとんどの情報がヒト可読
 - InChIKey(hashed InChI)はWeb検索用、人には読めず、非常に稀に重複可能性あり
 - SMILES
 - 分子の二次元構造を文字列として記述
 - 情報をコンパクトに保存
 - 原子座標の羅列と違い、ユーザーにも理解しやすい
 - canonical SMILESで一つの構造に対して唯一となるSMILES表記を定めることができる

2016/8/9

52

PubChem (<https://pubchem.ncbi.nlm.nih.gov/>)

NCBIが維持管理する低分子化合物データベース

- PubChem Substance (他のDB、PubChem自身の化合物データエントリー、ID:SID)
- PubChem Compound (Substance のデータを化学構造によりまとめたもの、ID:CID)
- PubChem BioAssay (Substance に含まれる生理活性情報を集約したもの)

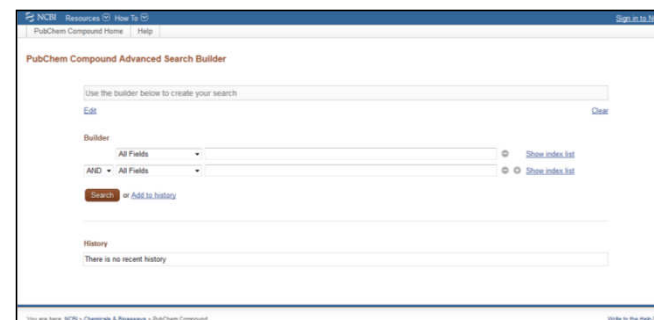


2016/8/9

53

PubChem Advanced Search

他の NCBI のデータベースと同様のインターフェースの Advanced Search Builder で項目と検索クエリーを組み合わせて検索を行うことができる



2016/8/9

54

PubChem 実習 1

PubChem Compound から化合物を検索してみましょう

- 画面中央の検索窓に化合物名を入力してGoをクリック
 - 特に検索対象の化合物を思い浮かべない人は「Rutin(フラボノイドの一種)」を入力してください



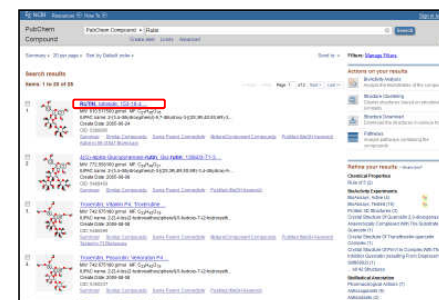
2016/8/9

55

PubChem 実習 1

検索結果から自分が検索したい化合物に一番近い化合物の名前をクリックしてください

- Rutin を入力した人は 1. の Rutin... をクリックしてください



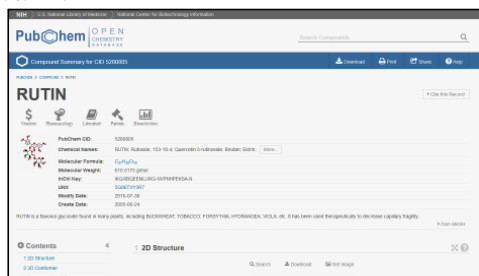
2016/8/9

56

PubChem 実習 1

検索結果の PubChem Compound エントリーが表示されます

- Rutin であれば CID 5280805
- 基本的な化学情報だけでなく、文献、特許、生理活性など多様な情報が表示



2016/8/9

57

PubChem 実習 2-1

5. Related record の項目では関連の項目を見ることができます

- 5.3 Substances の項目を見てみましょう
 - 5.3.1 Related Substances の Same の項目をクリックしてみましょう

2016/8/9

58

PubChem 実習 2-1

5. Related record の項目では関連の項目を見ることができます

- 5.3 Substances の項目を見てみましょう
 - 当該のCIDに対応する PubChem Substances のリストが出てきます

2016/8/9

59

PubChem 実習 2-2

5. Related record の項目では関連の項目を見ることができます

- 5.3 Substances の項目を見てみましょう
 - 5.3.2 Substances by Category では、Submitter のCategory 別に Substances を見ることができます。試みにツリーを開いてみましょう

2016/8/9

60

PubChem 実習 2-3

5. Related record の項目では関連の項目を見ることができます

- 5.2 Related Compounds の項目を見てみましょう
 - Same Isotope の項目をクリックしてみましょう
 - Same Isotope とは、Same Connectivity (結合が同じ) かつ Isotope も同じ、すなわち PubChem に立体異性体として登録されている化合物がリストされる

Corporate Summary for 100 Selected

Contents

- 1.2D Rotation
- 2.3D Conformer
- 3.Isomers and Isotopes
- 4.Chemical and Physical Properties
- 5.Modeled Records
- 6.Chemical Vendors
- 7.Pharmacokinetics and Toxicology
- 8.Literature
- 9.Patents
- 10.Environmental Information and Pathways
- 11.Biological Test Results
- 12.Clinical Trials
- 13.Information Sources

5.1 Related Compounds with Annotation

Isotopes (110) Isotachomers (274) Pseudo (232)

5.2 Related Compounds

Same Connectivity	49 records
Same Isotope	40 records
Same Topical Isotopes	100 records
Same Pseudo, Exact	36 records
Mixtures, Components, and Associated Purities	32 records
Isomer Compounds	10887 records
Similar Customers	20 records

2016/8/9

62

PubChem 実習 2-3

5. Related record の項目では関連の項目を見つけることができません

- リストの化学構造(InChI形式)をダウンロードしてみましょう
 - Structure Download をクリックします
 - format でInChI, compression type で none を選び, download をクリックします
 - formatではSDFなども選ぶことができます



PubChem • Download Service



[PubMed](#) | [Entrez](#) | [Structure](#) | [PubChem](#) | [Help](#)

PubChem Download Service

Download

Download the records in the format selected below: 

3DCHI

Choose a format: 

None

Choose a compression type: 

☐ Retrieve 3D records/images? 

1

Number of 3D conformers per CID: 

Save Job

Save this job in XML format (e.g. for PEX) 

[Write to Helpdesk](#) | [Disclaimer](#) | [Privacy statement](#) | [Accession policy](#) | [Data Citation Guidelines](#)

2016/8/9

64

PubChem 実習 2-3

5. Related record の項目では関連の項目を見ることができます

- 5.2 Related Compounds の項目を見てみましょう

– InChI のリストがダウンロードされています

- 立体異性を示す部分 (赤字の部分) が少しずつ違うことがわかります

- 5280805 InChI=1S/C27H30O16/c1-8-17(32)20(35)22(37)26(40-8)39-7-15-18(33)21(36)23(38)27(42-15)43-25-19(34)16-13(31)5-10(28)6-14(16)41-24(25)9-2-3-11(29)12(30)4-9/h2-6,8,15,17-18,20-23,26-33,35-38H,7H2,1H3/t8-,15+,17-,18+,20+,21-,22+,23-,26+,27-/m0/s1
- 98362683 InChI=1S/C27H30O16/c1-8-17(32)20(35)22(37)26(40-8)39-7-15-18(33)21(36)23(38)27(42-15)43-25-19(34)16-13(31)5-10(28)6-14(16)41-24(25)9-2-3-11(29)12(30)4-9/h2-6,8,15,17-18,20-23,26-33,35-38H,7H2,1H3/t8-,15+,17-,18+,20+,21-,22+,23-,26+,27-/m0/s1
- 98362682 InChI=1S/C27H30O16/c1-8-17(32)20(35)22(37)26(40-8)39-7-15-18(33)21(36)23(38)27(42-15)43-25-19(34)16-13(31)5-10(28)6-14(16)41-24(25)9-2-3-11(29)12(30)4-9/h2-6,8,15,17-18,20-23,26-33,35-38H,7H2,1H3/t8-,15-,17+,18-,20+,21-,22-,23+,26-,27+/m1/s1
- 98362681 InChI=1S/C27H30O16/c1-8-17(32)20(35)22(37)26(40-8)39-7-15-18(33)21(36)23(38)27(42-15)43-25-19(34)16-13(31)5-10(28)6-14(16)41-24(25)9-2-3-11(29)12(30)4-9/h2-6,8,15,17-18,20-23,26-33,35-38H,7H2,1H3/t8-,15-,17+,18-,20+,21-,22-,23+,26-,27+/m1/s1
- 98136788 InChI=1S/C27H30O16/c1-8-17(32)20(35)22(37)26(40-8)39-7-15-18(33)21(36)23(38)27(42-15)43-25-19(34)16-13(31)5-10(28)6-14(16)41-24(25)9-2-3-11(29)12(30)4-9/h2-6,8,15,17-18,20-23,26-33,35-38H,7H2,1H3/t8-,15+,17-,18-,20-,21-,22-,23-,26+,27+/m0/s1

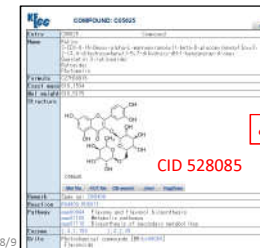
2016/8/9

65

PubChem CID で注意すべき点

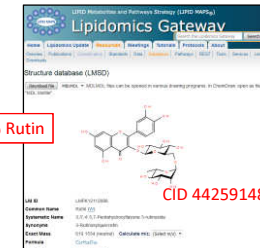
立体異性を持つ化合物では、Source側のDBの構造的書き方、mol形式からInChI形式に変換する際の立体異性変換の問題などにより、同じ化合物が同じCIDにならないことがある(特に、糖を含む化合物に顕著)

- 5280805 InChI=1S/C27H30O16/c1-8-17(32)20(35)22(37)26(40-8)39-7-15-18(33)21(36)23(38)27(42-15)43-25-19(34)16-13(31)5-10(28)6-14(16)41-24(25)9-2-3-11(29)12(30)4-9/h2-6,8,15,17-18,20-23,26-33,35-38H,7H2,1H3/t8-,15+,17-,18+,20+,21-,22+,23-,26+,27-/m0/s1
- 44259148 InChI=1S/C27H30O16/c1-8-17(32)20(35)22(37)26(40-8)39-7-15-18(33)21(36)23(38)27(42-15)43-25-19(34)16-13(31)5-10(28)6-14(16)41-24(25)9-2-3-11(29)12(30)4-9/h2-6,8,15,17-18,20-23,26-33,35-38H,7H2,1H3/t8?,15?,17-,18+,20-,21?,22?,23?,26+,27-/m0/s1



CID 528085

どちらも Rutin



CID 44259148

2016/8/9

66

KEGG LIGAND データベース

(<http://www.genome.jp/kegg/ligand.html>)

KEGG LIGAND データベース

Enter C, G, R, P, R, C, and EC numbers (Example: C00080 C00090 C00092 C00093 C00094 C00095 C00096 C00097)

Filter Pathway mapping Brite mapping Get title Get entry Clear

Chemical Substances and Reactions

KEGG LIGAND contains our knowledge on the universe of chemical substances and reactions that are relevant to life. It is a composite database consisting of COMPOUND, GLYCAN, REACTION, RPAIR, RCLASS, and ENZYME databases, whose entries are identified by C, G, R, P, R, C, and EC numbers, respectively. ENZYME is derived from the IUBMB/NCAC Enzyme Nomenclature, but the others are internally developed and maintained.

Database	Identifier	Content	Specialized entry page
COMPOUND	C number	Chemical compound structures	KEGG COMPOUND
GLYCAN	G number	Glycan structures	KEGG GLYCAN
REACTION	R number	Biochemical reactions	KEGG REACTION
RPAIR	RP number	Reactant pair alignments	KEGG REACTION
RCLASS	RC number	Reaction class	KEGG REACTION
ENZYME	EC number	Enzyme nomenclature	KEGG REACTION

Search LIGAND by: Go Close

blind mode target mode

KEGG LIGANDのサブカテゴリ

化合物、糖鎖、反応に特化した入口

ID番号のIdentifierと各DBに含まれるデータの説明

KEGGにおける「Chem(o)-」の部分、すなわち生化学的な情報全般を扱う今回は、COMPOUNDを紹介します。

2016/8/9

67

KEGG COMPOUND 実習

- 例: phenylalanine -

KEGG COMPOUND データベース

Search for: phenylalanine

Results: 1 hit

Compound	Identifier	Content
L-phenylalanine	C00031	Chemical compound structure

Chemical structure of L-phenylalanine is shown.

2016/8/9

68

KEGG COMPOUND 実習 ～ KEGG COMPOUND Entry ～

構造表示、
構造情報 (mol, KCF)、
構造検索、構造表示 (&編集)、

各種データベースへのリンク
REACTION, PATHWAY, ENZYME

外部データベースへのリンク
PubChem, ChEBI, etc.

LinkDB情報
GenomeNet内部でリンクされているDBや
対応関係のとれる外部DBを表示

69

ゲノムネットケミカル情報解析ツール

2016/8/9

<http://www.genome.jp/> (英語)
<http://www.genome.jp/ja/> (日本語)

70

GenomeNet 実習 GenomeNetからの化合物データベース一括検索

2016/8/9

英語: <http://www.genome.jp/>
日本語: <http://www.genome.jp/ja/>

71

KNAPsACK Family (http://kanaya.naist.jp/KNAPsACK_Family/)

奈良先端大・金谷研による代謝産物・生物種を軸とする代謝産物データベース

2016/8/9

72

KNaPSaCk 実習

KNaPSaCk Core System でイネ(*Oryza sativa*) が生成する代謝産物を検索してみよう

- KNaPSaCk Core をクリック

2016/8/9 73

KNaPSaCk 実習

KNaPSaCk Core System でイネ(*Oryza sativa*) が生成する代謝産物を検索してみよう

- Organism を選択し、*Oryza sativa* を入力し List をクリック

2016/8/9 74

KNaPSaCk 実習

KNaPSaCk Core System でイネ(*Oryza sativa*) が生成する代謝産物を検索してみよう

- リストから化合物IDをクリック(例: C00000091 trans-Zeatin)

2016/8/9 75

KNaPSaCk 実習

KNaPSaCk Core System でイネ(*Oryza sativa*) が生成する代謝産物を検索してみよう

- C00000091 trans-Zeatin エントリーが表示されます

2016/8/9 76

KNaPSaCk Familyの詳細な
使い方は、過去の講習会
(AJACS駿河、2013/1/12)の
資料、講習などがNBDCの講
習会サイトおよび統合TVで公
開されています
統合TVで20140227版の利
用法紹介映像もあります