

統合データベース講習会：AJACSみちのく2 2012年8月23日、24日

国内の主要なDBの使い方 (DDBJ, PDBj, KEGG)

情報・システム研究機構 (ROIS)
ライフサイエンス統合データベースセンター (DBCLS)
河野 信

1

©2012 統合データベース講習会 Licensed Under CC 表示 2.1

代表的な生命科学関係のデータベース

DNA塩基配列	GenBank/EMBL/ DDBJ
タンパク質アミノ酸配列	UniProt (Swiss-Prot + TrEMBL)
タンパク質立体構造	PDB SCOP, CATH
モチーフ	InterPro, Pfam, PROSITE, ProDom
ゲノム	Ensembl, H-InvDB, MGI, FlyBase, TAIR, SGD, KEGG GENES
化合物	PubChem, ChEBI, KEGG LIGAND
パスウェイ	KEGG PATHWAY , Reactome, BioCyc
遺伝子発現	GEO, ArrayExpress, BioGPS
文献	PubMed
その他	GO, NCBI Taxonomy, OMIM, GOLD

3

©2012 統合データベース講習会 Licensed Under CC 表示 2.1

注意点

- ◆ 参加人数が多いため、サイトにつながりにくくなることが予想されます。
- 資料を見ながら適当にタイミングをずらして実行してみてください
- 反応が無くても、何度もクリックしない
 - ますます遅くなるだけです。おおらかな気持ちで臨みましょう
- わからないことがあつたら、講習会のスタッフに気軽に聞いてください

2

©2012 統合データベース講習会 Licensed Under CC 表示 2.1

本日紹介するDB

- ◆ **DDBJ** ([DNA Data Bank of Japan](#))
 - DNA塩基配列を収集
 - 国立遺伝学研究所DDBJセンター（静岡県三島市）
- ◆ **PDBj** ([Protein Data Bank Japan](#))
 - タンパク質の立体構造を収集
 - 大阪大学蛋白質研究所
- ◆ **KEGG** ([Kyoto Encyclopedia of Genes and Genomes](#))
 - 生命情報をシステムとして表現
 - 京都大学化学研究所バイオインフォマティクスセンター

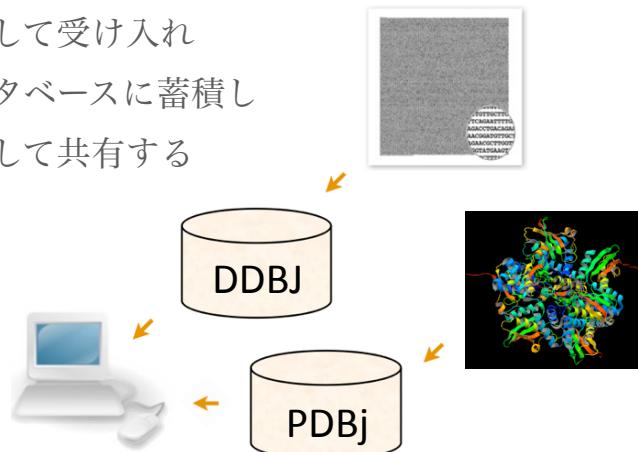
4

©2012 統合データベース講習会 Licensed Under CC 表示 2.1

データバンク

全世界で解読された塩基配列/立体構造情報を

- 査定して受け入れ
- データベースに蓄積し
- 公開して共有する



5

©2012 統合データベース講習会 Licensed Under CC 表示 2.1 © 2012 中村保一 (国立遺伝学研究所) licensed under CC表示2.1日本

DDBJ

DNA Data Bank of Japan

7

遺伝子・立体構造の論文には登録が不可欠



PLOS BIOLOGY

a peer-reviewed open-access journal published by the Public Library of Science

[Login](#) | [Create Account](#) | [Feedback](#)

Search articles...



Advanced Search

[Browse](#) [RSS](#)

[Journals](#) [Hubs](#) [PLOS.org](#)

Accession Numbers

All appropriate datasets, images, and information should be deposited in public resources. Please provide the relevant accession numbers (and version numbers, if appropriate). Accession numbers should be provided in parentheses after the entity on first use. Suggested databases include, but are not limited to:

- ▷ [ArrayExpress](#)
- ▷ [BioModels Database](#)
- ▷ [Database of Interacting Proteins](#)
- ▷ [DNA Data Bank of Japan \[DDBJ\]](#)
- ▷ [DRYAD](#)
- ▷ [EMBL Nucleotide Sequence Database](#)
- ▷ [GenBank](#)
- ▷ [Gene Expression Omnibus \[GEO\]](#)
- ▷ [Protein Data Bank](#)
- ▷ [UniProtKB/Swiss-Prot](#)
- ▷ [ClinicalTrials.gov](#)

論文投稿時の注意：論文の著者は、論文で言及した塩基配列や立体構造などのデータについて、インターネットで参照可能な公共データベースの登録番号を掲載しなければならない

6

©2012 統合データベース講習会 Licensed Under CC 表示 2.1 © 2012 PLoS Licensed Under CC Attribution 2.5

国際塩基配列データベースの一員

International Nucleotide Sequence Databank Collaboration (INSDC)

- 米国 : GenBank
- 欧州 : ENA
- 日本 : DDBJ



- ▷ (新型) DNAシーケンサーで解読されたDNA塩基配列を収集

8

©2012 統合データベース講習会 Licensed Under CC 表示 2.1 © 2012 中村保一 (国立遺伝学研究所) licensed under CC表示2.1日本

DDBJ登録ファイルの例

LOCUS AB091058 2109 bp DNA linear BCT 02-SEP-2003 CDS 1035..2096 /codon_start=1 /gene="ccp" /product="endo-beta-1,4-glucanase, cellulose complementing protein, complete" /protein_id="BA0C82441.."

DEFINITION Gluconacetobacter xylinus cmcase, ccp genes for endo-beta-1,4-glucanase, cellulose complementing protein, complete

ACCESSION AB091058

VERSION AB091058.1

KEYWORDS .

SOURCE Gluconacetobacter xylinus

ORGANISM Gluconacetobacter xylinus

Bacteria; Proteobacteria; Alphaproteobacteria; Rhodospirillales; Acetobacteraceae; Gluconacetobacter.

REFERENCE 1 (base 1 to 2109)

AUTHORS Kawano,S., Tajima,T., Uemori,Y., Yamashita,H., Erata,T., Munekata,M. and Takai,M.

TITLE Direct Submission

JOURNAL Submitted (28-AUG-2002) to the DDBJ/EMBL/GenBank databases.

Contact:Kenji Tajima
Hokkaido University, Graduate School of Engineering; N13W8,
Kitaku, Sapporo, Hokkaido 060-8628, Japan

REFERENCE 2

AUTHORS Kawano,S., Tajima,T., Uemori,Y., Yamashita,H., Erata,T., Munekata,M. and Takai,M.

TITLE Cloning of Cellulose Synthesis Related Genes from Acetobacter xylinum ATCC23769 and ATCC53582 Comparison of Cellulose Synthetic Ability Between ATCC23769 and ATCC53582

JOURNAL Unpublished (2002)

COMMENT

FEATURES source Location/qualifiers 1..2109 /db_xref="taxon:28448" /mol_type="genomic DNA" /note="Acetobacter xylinum" /organism="Gluconacetobacter xylinus" /strain="ATCC 53582"

CDS 10..1039 /codon_start=1 /gene="cmcase" /product="endo-beta-1,4-glucanase" /protein_id="BA0C82440..1" /transl_table="MSML" /translation="MSSVAAAGGAQVLSSPGAFADTAQAVQAQNAIFRASYLRLPGR VVDTONGESHSREGQGYMLPASAGDLASFOSMMQHARNLQIHTNDLFSPRLKXH QFPVFDKNNATDGLLIALLAGRAKRPRQRPYIQDAMAIYGDVNLMMKAGPVYVL MPAGAVGFTKDSVILNLSSYYVHPSLQLAQADLTADFRRNQVNMEDIGRLVSAGRFQQRLL PFDLAVNRATGALSLASWPFPLSTDAIRFLYTFNHLAPVNLADTFWNRNGA HALPWNOLTTGARSQYPAHPOYLAVAECTYGLDSMACHEPLTDHAFYISAALTLVY ARAAETIK"

BASE COUNT 343 a 661 c 661 g 444 t

ORIGIN

1 cggttcctta tgcggatcat ggccgcgcgtt ggaggccgcg agggtttttt atccacccgt
61 gttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt
121 ttccatccat ccggccggcc tttttttttt tttttttttt tttttttttt tttttttttt
181 ggccgggtt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt
241 atgttgtgtt ggccggatgtt tttttttttt tttttttttt tttttttttt tttttttttt
301 ttccatcaag ggccggatccgc cccgggtcc cccatccatay acacgttgtt ttccgtgggg
361 ttccatcaag tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt
421 ggccggatccgc cccgggtcc tttttttttt tttttttttt tttttttttt tttttttttt
481 gtccatcca tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt
541 tcattatccatc tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt
601 ctgtttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt
661 ctgtttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt
721 tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt
781 atgttgtgtt cccacgtttt tttttttttt tttttttttt tttttttttt tttttttttt
841 ggccgtcgag tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt
901 ggatgtttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt
961 ggccggatccgc cccgggtcc tttttttttt tttttttttt tttttttttt tttttttttt
1021 ggccggatccgc tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt
1081 ggccggatccgc tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt
1141 ctatcccgcc gttttttttt tttttttttt tttttttttt tttttttttt tttttttttt
1201 ggccggatccgc tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt
1261 atccatccat ccggccggcc tttttttttt tttttttttt tttttttttt tttttttttt
1321 aaacccggcc cccggccggcc tttttttttt tttttttttt tttttttttt tttttttttt
1381 cccggccggcc tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt
1441 cccggccggcc tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt
1501 cccggccggcc tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt
1561 cccggccggcc tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt
1621 agttccggccatc gttccggccatc gggggaaacc cttttttttt tttttttttt
1681 cttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt
1741 atgttgtgtt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt
1801 tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt
1861 tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt
1921 acacgtttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt
1981 ccacaaatccggcc tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt
2041 ggccggatccgc tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt
2101 statatcca

9

©2012 統合データベース講習会 Licensed Under CC 表示 2.1 //

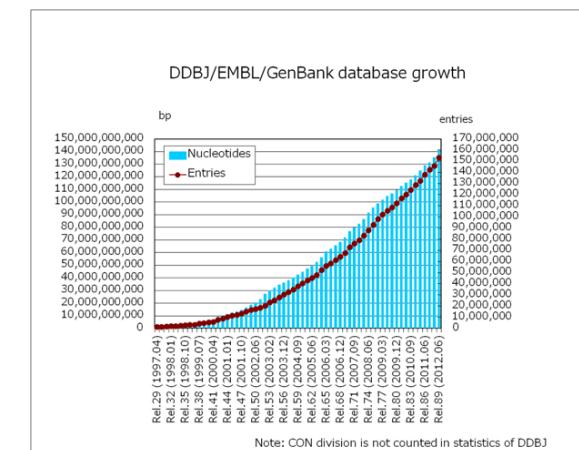
DDBJデータベースを検索してみましょう

11

現在の塩基配列データの量

塩基数：1,400億

登録数：1.5億



10

©2012 統合データベース講習会 Licensed Under CC 表示 2.1 © 2012 中村保一 (国立遺伝学研究所) licensed under CC表示2.1日本

実習1

- ◆DDBJデータベースを"ARSA"という、キーワード検索ツールで検索してみましょう
- 例として大腸菌O157の全ゲノムエントリを検索してみます

DDBJにアクセスするには「DDBJ」で検索
もしくは <http://www.ddbj.nig.ac.jp/> を直接入力

©2012 統合データベース講習会 Licensed Under CC 表示 2.1

12

DDBJトップページ

The screenshot shows the DDBJ homepage. At the top right, there are links for ENGLISH, a search bar, and a site search button. Below the header, there's a navigation menu with HOME, 基因配列の登録, 利用の手引き, 検索・解析, FTP・WebAPI, レポート・統計, and お問い合わせ. On the left, there's a sidebar with sections for DDBJの紹介, Q&A集, 基因配列の登録 (SAKURA, 大量登録システム(MSS), データの修正・更新, DDBJ Sequence Read Archive, DDBJ Trace Archive), プロジェクトの登録 (DDBJ BioProject Database), スーパーコンピュータ利用 (スノコンの利用申請, スノコンの利用方法, スノコンマニュアル), 検索 (genbank, ARSA, TXSearch, BLAST), and a license notice (CC BY). The main content area features a banner for 'DDBJ : DNA Data Bank of Japan' with a photo of Mount Fuji, followed by sections for Hot Topics (with links to DDBJ Web Magazine RSS, Papilio xuthus, Papilio polytes EST data, and Solanum lycopersicum genome data), Maintenance, Information (with a link to DDBJ Web Magazine No. 71), and download links for GenBank and FTP.

13

ARSA

The screenshot shows the ARSA search results for 'E. coli O157'. The search bar at the top has 'E. coli O157' highlighted. Below the search bar, there's a note about search operators. The results table has columns for PrimaryAccessionNumber, Definition, moltype, Organism, and Length. The first result is 'AB011548 Escherichia coli O157:H7 str. Sakai plasmid pOSAK1 DNA, complete sequence.' and the last result is 'AB035920 Escherichia coli O157:H7 hemG, rrsA, ileT, alaT, rrfA, mobB, mobA genes for protoporphyrin oxidase protein, 16S rRNA, isoleucine tRNA 1, alanine tRNA 1B, 23S rRNA, 5S rRNA, molybdopterin-guanine dinucleotide biosynthesis protein B, molybdopterin-guanine dinucleotide biosynthesis protein A, complete and partial cds.'.

E. coli O157 を検索

検索結果：

ヒット数が多すぎるので
絞り込みが必要

詳細検索

(Advanced Search)へ

14

いくつかの特徴で絞り込み

The screenshot shows the ARSA Advanced Search interface. It includes fields for combining searches (AND/OR), text search, accession number, primary accession number, division, sequence length, molecular type (radio buttons for sDNA, rRNA, cRNA, mRNA, rRNA, tRNA, with 'circular' and 'linear' checked), date, definition, comment, keyword, organism ('E. coli' highlighted), and taxon. There are also buttons for 'Search' and 'reset'.

15

ヒットしない…

The screenshot shows the ARSA search results for 'C. glutamicum-E. coli'. The search bar at the top has 'C. glutamicum-E. coli' highlighted. The results table has columns for PrimaryAccessionNumber, Definition, moltype, Organism, and Length. The results are as follows:

PrimaryAccessionNumber	Definition	moltype	Organism	Length
AB602479	C. glutamicum-E. coli shuttle vector pCRB12 DNA, complete sequence.	DNA	C. glutamicum-E. coli shuttle vector pCRB12	4569
AB671168	E. coli-T. thermophilus shuttle vector pTRK1T DNA, complete sequence.	DNA	E. coli-T. thermophilus shuttle vector pTRK1T	7482
AB671169	E. coli-T. thermophilus shuttle vector pTRH1T DNA, complete sequence.	DNA	E. coli-T. thermophilus shuttle vector pTRH1T	6057
HM126493	C. glutamicum-E. coli shuttle vector pCRB62, complete sequence.	DNA	C. glutamicum-E. coli shuttle vector pCRB62	5914
HM126494	C. glutamicum-E. coli shuttle vector pCRB12, complete sequence.	DNA	C. glutamicum-E. coli shuttle vector pCRB12	4569

[1]

原因：Organismの項目には正式名称しか書かれていないため

16

Escherichia coli O157 Sakai で再検索

Molecular

Type DNA RNA cRNA mRNA rRNA tRNA
Form circular linear

Date
Definition
Comment
Keyword
Organism Escherichia coli O157 Sakai
Taxon

ARSA All-round Retrieval of Sequence and Annotation

DBQuick Search DBAdvanced Search

FlatFile XML Fasta View Download

All Select

PrimaryAccessionNumber	Definition	moltype	Organism	Length
AB011548	Escherichia coli O157:H7 str. Sakai plasmid pOSAK1 DNA, complete sequence.	DNA	Escherichia coli O157:H7 str. Sakai	3306
AB011549	Escherichia coli O157:H7 str. Sakai plasmid pO157 DNA, complete sequence.	DNA	Escherichia coli O157:H7 str. Sakai	92721
BA000007	Escherichia coli O157:H7 str. Sakai DNA, complete genome.	DNA	Escherichia coli O157:H7 str. Sakai	5498450

BA000007 をクリックすると O157のゲノムエントリを表示

チェックを入れて"Download"をクリックするとエントリ、配列をダウンロード可能

[1] ©2012 統合データベース講習会 Licensed Under CC 表示 2.1

実習2

◆先ほどと同じ検索をNCBI (GenBankの提供機関) の検索システム"Entrez"で実行してみましょう

●例として大腸菌O157の全ゲノムエントリを検索してみます

NCBI Entrezにアクセスするには「Entrez」で検索もしくは <http://www.ncbi.nlm.nih.gov/Entrez/> を直接入力

18

[cc BY] ©2012 統合データベース講習会 Licensed Under CC 表示 2.1

NCBI Entrez Search

文献

PubMed All Databases Human Genome GenBank Map Viewer BLAST

Search across databases GO Clear Help

Welcome to the Entrez cross-database search page

PubMed: biomedical literature citations and abstracts	Books: online books
PubMed Central: free, full text journal articles	OMIM: online Mendelian Inheritance in Man
Site Search: NCBI web and FTP sites	

塩基配列

EST

アミノ酸配列

ゲノム配列

立体構造

Nucleotide: Core subset of nucleotide sequence records
EST: Expressed Sequence Tag records
GSS: Genome Survey Sequence records
Protein: sequence database
Genome: whole genome sequences
Structure: three-dimensional macromolecular structures
Taxonomy: organisms in GenBank
SNP: short genetic variations

dbGaP: genotype and phenotype
UniGene: gene-oriented clusters of transcript sequences
CDD: conserved protein domain database
Clone: integrated data for clone resources
UniSTS: markers and mapping data
PopSet: population study data sets
GEO Profiles: expression and molecular abundance profiles
GEO DataSets: experimental sets of GEO data

[cc BY] ©2012 統合データベース講習会 Licensed Under CC 表示 2.1

Escherichia coli O157 Sakai で検索

文献

PubMed All Databases Human Genome GenBank Map Viewer BL

Search across databases Escherichia coli o157 sakai GO Clear Help

- Result counts displayed in gray indicate one or more terms not found

105 PubMed: biomedical literature citations and abstracts	none Books: online books
233 PubMed Central: free, full text journal articles	none OMIM: online Mendelian Inheritance in Man
1 Site Search: NCBI web and FTP sites	

塩基配列

EST

アミノ酸配列

ゲノム配列

立体構造

17610 Nucleotide: Core subset of nucleotide sequence records
4 EST: Expressed Sequence Tag records
none GSS: Genome Survey Sequence records
31034 Protein: sequence database
1 Genome: whole genome sequences
3 Structure: three-dimensional macromolecular structures
none SNP: short genetic variations

2 dbGaP: genotype and phenotype
UniGene: gene-oriented clusters of transcript sequences
CDD: conserved protein domain database
Clone: integrated data for clone resources
UniSTS: markers and mapping data
PopSet: population study data sets
GEO Profiles: expression and molecular abundance profiles
GEO DataSets: experimental sets of GEO data

[cc BY] ©2012 統合データベース講習会 Licensed Under CC 表示 2.1

20

Entrez: 大腸菌ゲノムページ

NCBI Resources How To My NCBI Sign In

Genome Genome Escherichia coli o157 sakai Search Save search Limits Advanced Help

Display Settings: Overview Send to: Related information

Organism Overview: Genome Project Report; Genome Annotation Report; Plasmid Annotation Report

Escherichia coli
A well-studied enteric bacterium

Lineage: Bacteria[2817]; Proteobacteria[1179]; Gammaproteobacteria[514]; Enterobacterales[131]; Enterobacteriaceae[13]; Escherichia[1]; Escherichia coli[1]

Escherichia coli. This organism was named for its discoverer, Theodore Escherich, and is one of the premier model organisms used in the study of bacterial genetics, physiology, and biochemistry. This enteric organism is typically present in the lower intestine of humans, where it is the dominant facultative anaerobe present, but it is [More...](#)

Dendrogram (based on genomic BLAST)

Search details: [Escherichia coli O157[Organism] AND sakai[All Fields]]

Recent activity: Escherichia coli (Turn Off Clear)

Genome Sequencing Projects: Chromosomes [1], Scaffolds or contigs [0], SRA or Traces [0], No data [0]

Organism	BioProject	Assembly	Status	Chrs	Plasmids	Size (Mb)	GC%	Gene	Protein
Escherichia coli O157:H7 str. Sakai	PRJNA7781, PRJNA226	ASM886v1	♦	1	2	5.59	50.4	5,460	5,318

Genome Region: Go to nucleotide Graphics FASTA GenBank

©2012 統合データベース講習会 Licensed Under CC 表示 2.1

21

大腸菌O157 Sakai株のページ

NCBI Resources How To My NCBI Sign In

Genome Genome Search Save search Limits Advanced Help

Display Settings: Overview Send to: Related information

Organism Overview: Genome Project Report; Genome Annotation Report; Plasmid Annotation Report

Escherichia coli O157:H7 str. Sakai
Enterohemorrhagic Escherichia coli

Lineage: Bacteria[2817]; Proteobacteria[1179]; Gammaproteobacteria[514]; Enterobacterales[131]; Enterobacteriaceae[13]; Escherichia[7]; Escherichia coli[1]; Escherichia coli O157:H7[1]; Escherichia coli O157:H7 str. Sakai[0]

Escherichia coli O157:H7: This strain is associated with Hamburger disease, which is caused by the contamination of meat products by enterohemorrhagic *E. coli* (EHEC). The identifier O157:H7 refers to the serotype of EHEC, and reflects the specific antigenic markers found on the surface of the cell. EHEC attaches and effaces to cells [More...](#)

Genome Sequencing Projects

Recent activity: Turn Off Clear

Genome Region: Go to nucleotide Graphics FASTA GenBank

©2012 統合データベース講習会 Licensed Under CC 表示 2.1

22

豊富なリンクとツール群

NCBI Resources How To My NCBI Sign In

Nucleotide Nucleotide Search Limits Advanced Help

Display Settings: GenBank Send to: Change region shown

Sequence not displayed. Use 'Customize View' section for control.

Escherichia coli O157:H7 str. Sakai chromosome, complete genome
NCBI Reference Sequence: NC_002695

FASTA Graphics

Go to: BLASTによる類似性検索 プライマー設計ツール 文献へのリンク

LOCUS NC_002695 549850 bp DNA circular BCT 25-JAN-2012
DEFINITION Escherichia coli O157:H7 str. Sakai chromosome, complete genome.
ACCESSION NC_002695
VERSION NC_002695.1 GI:15829254
DBLINK Project: [57781](#)
KEYWORDS
SOURCE Escherichia coli O157:H7 str. Sakai
ORGANISM Escherichia coli O157:H7 str. Sakai
Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacterales; Enterobacteriaceae; Escherichia.
REFERENCES 1 to 5498459
AUTHORS Bergholz,T.M., Wick,L.M., Qiu,W., Riordan,J.T., Ouellette,L.M. and Whittam,T.S.
TITLE Global transcriptional response of Escherichia coli O157:H7 to growth transitions in glucose minimal medium
JOURNAL BMC Microbiol. 7, 97 (2007)
PUBMED 17967175
REMARKS Project Status: Online Only
REFERENCE 2 (sites)
AUTHORS Hayashi,T., Makino,K., Ohnishi,M., Kurokawa,K., Ishii,Y., Yokoyama,K., Han,C.G., Ohtsubo,E., Nakayama,K., Murata,T., Tanaka,M., Tobe,T., Iida,T., Takami,H., Honda,T., Sasakawa,C., Ogasawara,N., Yasunaga,T., Kuhara,S., Shiba,T., Hattori,M. and Shirane,M.
TITLE Complete genome sequence of enterohemorrhagic *Escherichia coli* O157:H7 and genomic comparison with a laboratory strain K-12
JOURNAL DNA Res. 8 (1), 11-22 (2001)
PUBMED 11258796
REMARK Erratum:[DNA Res 2001 Apr;27(8):96]

©2012 統合データベース講習会 Licensed Under CC 表示 2.1

23

配列を取得したとの解析例

- ◆ “blast”等の配列類似性検索を実行して、類似の配列を収集する
- ◆ “primer3”等で配列をクローニングするためのプライマーを設計する
- ◆ “clustalW”等でマルチブルアラインメントを作成し、配列の共通部分や進化関係を調べる
- ◆ “interproscan”等でモチーフ構造を調べる
- ◆ “swiss-model”等で立体構造を予測する

24

新型シーケンサからのデータ

25

©2012 統合データベース講習会 Licensed Under CC 表示 2.1

SRAs: Survey of Read Archives

- ◆SRA/DRAに登録されているデータを
メタデータで整理

◆<http://sra.ncbi.nlm.nih.gov/>

- 生物種
 - 解析プラットフォーム
 - キーワード
- などで検索可能

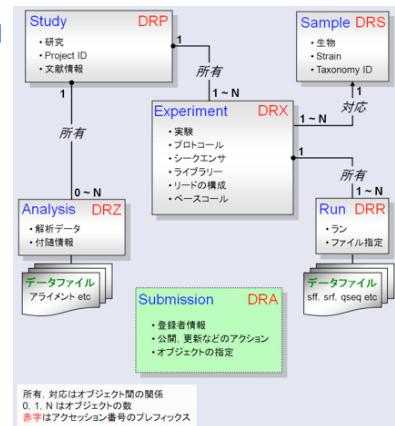
©2012 統合データベース講習会 Licensed Under CC 表示 2.1

DDBJ Sequence Read Archive (DRA)

◆新型シーケンサデータを保存・共有

登録されているデータ構造は少々複雑ですが、
DRAのページでは「日本語」での詳しい説明がある
登録されているデータはあまりうまく
整理されていない

新型シーケンサのデータ構造



26

©2012 統合データベース講習会 Licensed Under CC 表示 2.1

SRAs: Survey of Read Archives

- ◆統計値から、分類をたどってデータにアクセス
することも可能

Search by statistics - 統計値から探す

The number of projects are indicated in "Study Types" table. The totals in "Platforms" and "Species of Samples" are larger than one of "Study Types" because a project can contain some platforms and sample species.

実際のプロジェクト数は"Study Types"に書かれているものです。"Platforms"と"Species of Samples"の総数は、複数のプラットフォームで行われた実験が1つのプロジェクトでなされる場合が多くあります。ダブルカウントしているので、数字が大きくなっています。

2012-06-21 updated.

Study Types

	Whole Genome Sequencing	Transcriptome Analysis	Growth of the number of studies (including Total)
Metagenomics	6024	1657	12,000
Epigenetics	1270		
Resequencing	886		
Other	628		
RNASeq	615		
Population Genomics	585		
Gene Regulation Study	162		
Cancer Genomics	68		
Exome Sequencing	41		
Pooled Clone Sequencing	39		
Synthetic Genomics	21		
Forensic or Paleo-genomics	6		
Total	12008	(studies)	14554

* Growth of the number of studies (without Total)

©2012 統合データベース講習会 Licensed Under CC 表示 2.1

Platform

Illumina Genome Analyzer II	3629	Illumina MiSeq 2000	45470
454 GS FLX Titanium	2848	Illumina Genome Analyzer II	36940
Illumina HiSeq 2000	2328	454 GS FLX Titanium	11977
454 GS FLX	1755	Illumina Genome Analyzer	9621
Illumina Genome Analyzer	1167	454 GS FLX	8839
Illumina Genome Analyzer IX	565	Illumina Genome Analyzer IX	8760
454 GS 20	325	AB SOLID 4 System	2000
AB SOLID System 3.0	176	AB SOLID System 3.0	1080
AB SOLID 4 System	80	454 GS 20	686
Illumina MiSeq	64	Complete Genomics	401
AB SOLID System 2.0	61	unspecified	394
unspecified	60	AB SOLID System 2.0	362
Ion Torrent PGM	56	Helicos BioScope	332
AB SOLID System	37	PacBio RS	326

Species of samples (top 15)

Homo sapiens	1107	Homo sapiens	38506
unidentified	886	Streptococcus pneumoniae	8678
Mus musculus	603	Mus musculus	7155
Drosophila melanogaster	253	Drosophila melanogaster	7004
Caenorhabditis elegans	181	Caenorhabditis elegans	5523
metagenome sequence	180	human metagenome	3028
marine metagenome	160	Danio rerio	2940
Escherichia coli str. K-12 substr. MG1655	135	Plasmodium falciparum	2281
Arabidopsis thaliana	104	Staphylococcus aureus	1959
soil metagenome	115	Mycobacterium tuberculosis	1377
Saccharomyces cerevisiae	107	Saccharomyces cerevisiae	1325
Mustela putorius furo	99	Oryza sativa	1279
Human rhinovirus A	94	Sus scrofa	1268
Salmo salar	56	marine metagenome	1217
Plasmodium falciparum	50	Arabidopsis thaliana	1212
Total	14554	Total	127426
(studies)	(experiments)		

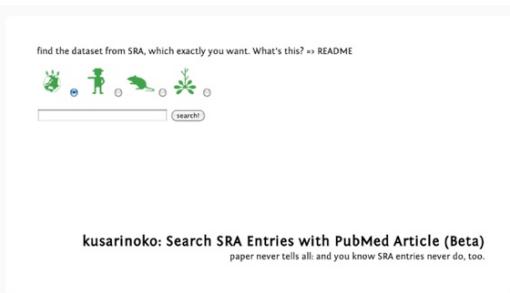
28

鎖鋸 (kusarinoko)

♦論文が出ているSRA/DRAエントリのまとめ

- 論文が出ているということは、査読を経ているので、一定のデータの質は担保されている（はず）

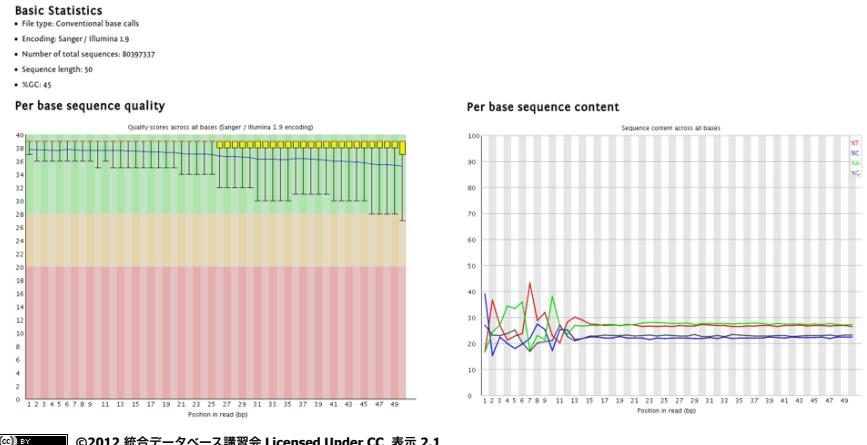
♦<http://g86.dbcls.jp/kusarinoko>



29

鎖鋸 (kusarinoko)

♦独自に”FastQC”をかけて、それぞれのデータの質を評価

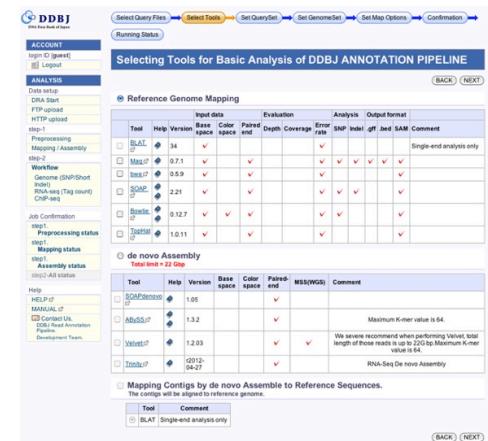


30

DDBJ Read Annotation Pipeline

♦新型シーケンサデータの解析パイプライン

♦<http://p.ddbj.nig.ac.jp/>



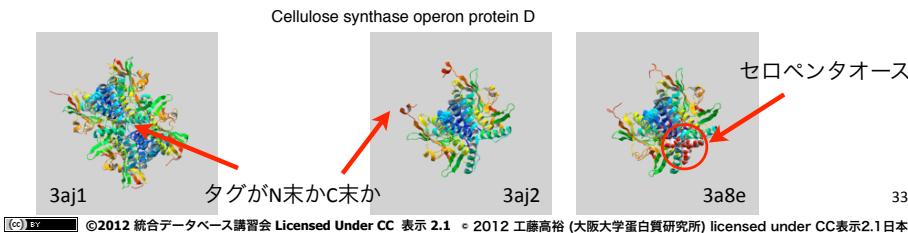
31

PDBj
Protein Data Bank Japan

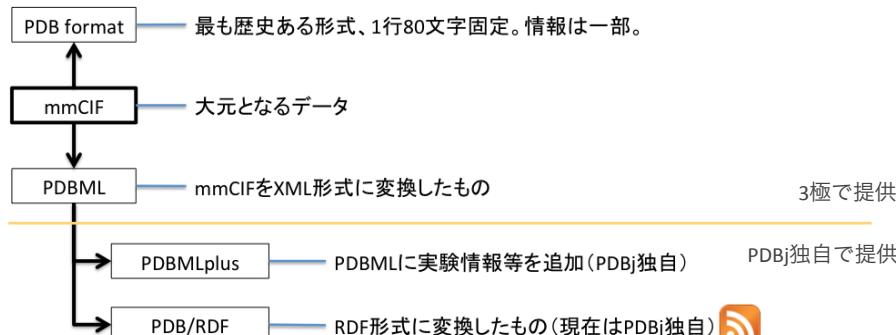
32

PDBについて

- ◆主にタンパク質、核酸の立体構造データを集めたデータバンク
- 現在のエントリ数は約84,000
- ひとつのタンパク質でも、リガンドの有無や配列の改変などの違いによって、複数のエントリが登録されていることがある



PDBのデータ形式



いずれもテキストデータ(エディタで読める)

mmCIF: macromolecular Crystallographic Information File
RDF: Resource Description Framework

35

世界蛋白質構造データバンクの一員

- ◆Worldwide Protein Data Bank (wwPDB)

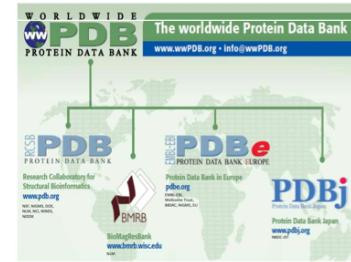
●米国 : RCSB-PDB (Research Collaboratory for Structural Bioinformatics)

●欧州 : PDBe

●日本 : PDBj

●米国 : BMRB

(Biological Magnetic Resonance Data Bank)



- ◆X線結晶解析、NMR、電子顕微鏡を使った「実験」にもとづいて決定された構造を収集

[CC BY] ©2012 統合データベース講習会 Licensed Under CC 表示 2.1 © 2012 工藤高裕 (大阪大学蛋白質研究所) licensed under CC表示2.1日本

PDBファイルの例

SHEET	6	A 7 PHE A 216	GLU A 222	1	N PHE A 218	O PHE A 221
SHEET	7	B 2 GLY A 241	LEU A 243	1	N CYS A 241	O GLN A 221
SHEET	1	B 2 GLN A 302	ARG A 307	0		
SHEET	2	B 3 GLY A 401	THR A 406	-1	N PHE A 405	O ILE A 406
SHEET	3	B 3 VAL A 462	HIS A 466	-1	N ILE A 465	O PHE A 402
SHEET	1	C 2 LEU A 411	ASN A 416	0		
SHEET	2	C 2 SER A 449	LEU A 454	-1	N LEU A 454	O LEU A 411
SHEET	1	D 2 GLY A 422	CYS A 427	0		
SHEET	2	D 2 CYS A 441	VAL A 444	-1	N VAL A 444	O GLY A 422
SHEET	1	E 2 GLY A 444	CYS A 449	0		
SHEET	1	F 2 CYS A 134	CYS A 148	0		
SHEET	2	CYS A 425	CYS A 437	0		
SHEET	3	CYS A 501	CYS I 518	0		
SHEET	4	CYS I 520	CYS I 523	0		
SHEET	5	CYS I 517	CYS I 531	0		
LINK	0	OD1 ASN A 98	CA CA A 601	1555	1555	2.44
LINK	0	OD2 ASN A 98	CA CA A 601	1555	1555	2.32
LINK	0	OD1 ASP A 155	CA CA A 601	1555	1555	2.54
LINK	0	OD2 ASP A 155	CA CA A 601	1555	1555	2.47
LINK	0	O HIS A 189	CA CA A 601	1555	1555	2.37
LINK	0	CA CA A 601	O HOH A 619	1555	1555	2.54
LINK	0	CA CA A 601	O HOH A 628	1555	1555	2.47
LINK	0	CA CA A 601	O HOH A 611	1555	1555	2.58
LINK	0	C PCA A 1	N LYS A 2	1555	1555	1.33
OSPEP	1	V ALA 123	PRO A 124	0		6.37
OSPEP	2	GLU A 119	PRO A 120	0		-5.50
OSPEP	3	ALA A 185	GLU A 222	ASP A 287		
SITE	1	AC1 7 ASN A 98	ARG A 146	ASP A 155	HIS A 189	
SITE	2	AC1 7 HOH A 611	HOH A 619	HOH A 628		
KEYWDS	INSECT ALPHA-AMYLASE INHIBITOR, AMARANTHUS HYPOCHONDRIACUS, 2 YELLOW MEAL WORM, X-RAY STRUCTURE, KNOTTIN, HYDROLASE					
X-RAY DIFFRACTION						
AUTHOR	P.J.B.PEREIRA,V.LOZANOV,A.PATTHY,R.HUBER,W.BODE,S.PONGOR,					
AUTHOR	2 S.S.TROBL					
REVDAT	3 24-FEB-09 1CLV 1 VERSN					
REVDAT	2 01-APR-03 1CLV 1 JRN					
REVDAT	1 03-MAY-00 1CLV 0					
JRNL	AUTH P.J.B.PEREIRA,V.LOZANOV,A.PATTHY,R.HUBER,W.BODE,					
JRNL	AUTH 2 S.PONGOR,S.TROBL					
JRNL	TITL SPECIFIC INHIBITION OF INSECT ALPHA-AMYLASES:					
JRNL	TITL 2 YELLOW MEAL WORM ALPHA-AMYLASE IN COMPLEX WITH THE					
JRNL	TITL 3 AMARANTH ALPHA-AMYLASE INHIBITOR AT 2.0 A					
JRNL	KEYWDS 4 RESOLUTION.					
JRNL	REF 5 STRUCTURE FOLD.DES. V. 7 1079 1999					
JRNL	REFN ISSN 9969-2126					
JRNL	PBD 10508772					
JRNL	DOI 10.1016/S0969-2126(99)80175-0					
REMARK 1						
REMARK 2						
REMARK 2 RESOLUTION_2.00 ANGSTROMS						
ATOM 9 N LYS A 2	29.983	7.994	5.735	1.00	14.51	N
ATOM 10 CA LYS A 2	30.178	8.269	4.313	1.00	13.28	C
ATOM 11 C LYS A 2	28.999	8.963	3.640	1.00	16.12	C
ATOM 12 O LYS A 2	29.027	9.224	2.435	1.00	17.54	O
ATOM 13 CB LYS A 2	30.534	6.982	3.574	1.00	13.33	C
ATOM 14 CG LYS A 2	31.829	6.365	4.059	1.00	14.70	C
ATOM 15 CD LYS A 2	32.140	5.082	3.331	1.00	17.22	C

座標情報

[CC BY] ©2012 統合データベース講習会 Licensed Under CC 表示 2.1

36

実習3

- ◆PDBjの検索サービス”PDBj Mine”で
「 α アミラーゼ」を検索してみましょう
- 同一IDのPDBデータは3極どこで見ても同じ。
- ただし、PDBjだと日本語でも検索可能

PDBjにアクセスするには「PDBj」で検索
もしくは <http://pdbj.org/> を直接入力

37

©2012 統合データベース講習会 Licensed Under CC 表示 2.1

PDBj Mine検索結果

Mine

検索結果ページ

(PDB-IDをクリックすると、詳細情報をご覧いただけます)

1 - 16 / 319

1 2 3 4 5 6 ... 20 次へ

クエリ : PDB IDまたはキーワード 表示順: 一致件数

変換クエリ : (alpha amylase) I (alfa amylase) ← 実際に検索に使われた文字列

個別エントリページへ (英語に自動変換して検索)

1clv

分子名稱 : ALPHA-AMYLASE, ALPHA-AMYLASE INHIBITOR
タイトル : YELLOW MEAL WORM ALPHA-AMYLASE IN COMPLEX WITH THE AMARANTH ALPHA-AMYLASE INHIBITOR
著者 : Pereira, P.J.B., Lozanov, V., Pathy, A., Huber, R., Bode, W., Pongor, S., Strobl, S.
実驗手法 : X-RAY DIFFRACTION
登録日 : 1999-05-04
公開日 : 2000-05-03
最終更新日 : 2009-02-24

1bvn

分子名稱 : PROTEIN (ALPHA-AMYLASE) (3.2.1.1)
タイトル : PIG PANCREATIC ALPHA-AMYLASE IN COMPLEX WITH THE PROTEINACEOUS INHIBITOR TENDAMISTAT
著者 : Machiusi, M., Wiegand, G., Epp, O., Huber, R.
実驗手法 : X-RAY DIFFRACTION
登録日 : 1998-09-16
公開日 : 1998-09-23
最終更新日 : 2009-02-24

39

©2012 統合データベース講習会 Licensed Under CC 表示 2.1

PDBj トップページ

PDBj

English simplified Chinese traditional Chinese Korean

統計情報 ヘルプ FAQ お問い合わせ

62522 entries available on 20 Jun. 2012 (00:00 UTC) 09:00(JST)

日本蛋白質データバンク (PDB: Protein Data Bank Japan) は JST-NBDC と大阪大学の支援を受け、米国 RCSB, BMRB、および欧州 PDB と協力して、生体高分子の立体構造データベースを国際的に統一化

日本蛋白質データバンク (PDB: Protein Data Bank Japan) は JST-NBDC と大阪大学の支援を受け、米国 RCSB, BMRB、および欧州 PDB と協力して、生体高分子の立体構造データベースを国際的に統一化

データ登録 >> AGDT: PDB Deposition AGDT: NMR

検索 >>

Search PDB (PDBj Mine) PDB-INFO, chem, comp/3D Latest Release Search Sequence-Navigator Structure Navigator SeqSAV Ligand Binding Sites (GIRAF) EM NMR Data Search NMR Data (BMRB) Status Search

サービスツール >> V/G: Graphic Viewer V/B (Yonodrum) Protein Globe ASH MAFFT/Flash SEALER Structure Prediction >> CRNPPRED Spanner SFAS 二次データベース >> eF-site/eF-seek/eFsurf eF-site ProMode/Elastic /ProMode/Oligomer Molecule of the Month ダウンロード >> PDB Archive/Snapshot Archive リンク集

最新情報

2012/6/5 今日の分子の記事数が150件に達しました。今月の記事
2012/5/1 2012年4月30日をもまして、xPSSSサービスおよびSC
2012/4/30 2012年4月30日（月・祝）に行われました大阪大学いち
2012/4/18 2012年6月20日（水）に、第12回蛋白質科学学会年会にて
2012/2/29 xPSSS新バージョン (V4.2) がリリースされました。eF-siteも改
2012/3/16 PDBで公開済みのエントリー数が、8万件を超える
2012/3/7 EMデータバンクがPDBアーカイブに追加されました
2012年3月7日に、EMデータバンク (EMDB) がPDBア [詳細...]

PDB検索 Mine Mine日本語ページについて

PDB IDまたはキーワード 検索

詳細条件検索 >>

NMRデータ検索 Accession number Deposition code

PDB検索 Mine Mine日本語ページについて

PDB IDまたはキーワード 検索

詳細条件検索 >>

38

©2012 統合データベース講習会 Licensed Under CC 表示 2.1

PDBj エントリページ

Mine

概要[1clv]

日本ページについて
PDBj Mineについて
更新情報

概要 [1clv] 構造情報 実験情報 離胞情報 相同蛋白質 ダウンロード/画面表示 外部データベース

PDB IDまたはキーワード

<非対称単位> =生物学的単位

1clv 配列情報 (FASTA形式) PDBファイルのダウンロード ← PDBファイルのダウンロード

分子名稱 ALPHA-AMYLASE, ALPHA-AMYLASE INHIBITOR
タイトル YELLOW MEAL WORM ALPHA-AMYLASE IN COMPLEX WITH THE AMARANTH ALPHA-AMYLASE INHIBITOR
機能のキーワード INSECT ALPHA-AMYLASE INHIBITOR, AMARANTHUS HYPOCHONDRIACUS, YELLOW MEAL WORM, KNOTTIN, HYDROLASE
由來する生物種 *Tenrebi molitor* (yellow mealworm)
由來する組織 [UNP_P80403] Seed
ポリマー鎖の合計数 2
分子量の合計数 54934.3 (詳細は構造情報のページ)
著者 Pereira, P.J.B., Lozanov, V., Pathy, A., Huber, R., Bode, W., Pongor, S., Strobl, S. (登録日: 1999-05-04, 公開日: 2000-05-03, 最終更新日: 2009-02-24)
引用文献 Pereira, P.J., Lozanov, V., Pathy, A., Huber, R., Bode, W., Pongor, S., Strobl, S. Specific inhibition of insect alpha-amylases: yellow meal worm alpha-amylase in complex with the amaranth alpha-amylase inhibitor at 2.0 Å resolution. *Biochemistry* 1999, 10(8), 1088-1095. [Published: 10500777] (DOI: 10.1016/S0886-2126(98)80175-0)
実驗手法 X-RAY DIFFRACTION (2.00Å)
他のデータベース情報 UniProt (P56634, P80403), eF-site, KEGG (EC 3.2.1.1), ProTherm, EzCatDB, PISA, PQS, PDB-INFO
NMR情報 BMRB

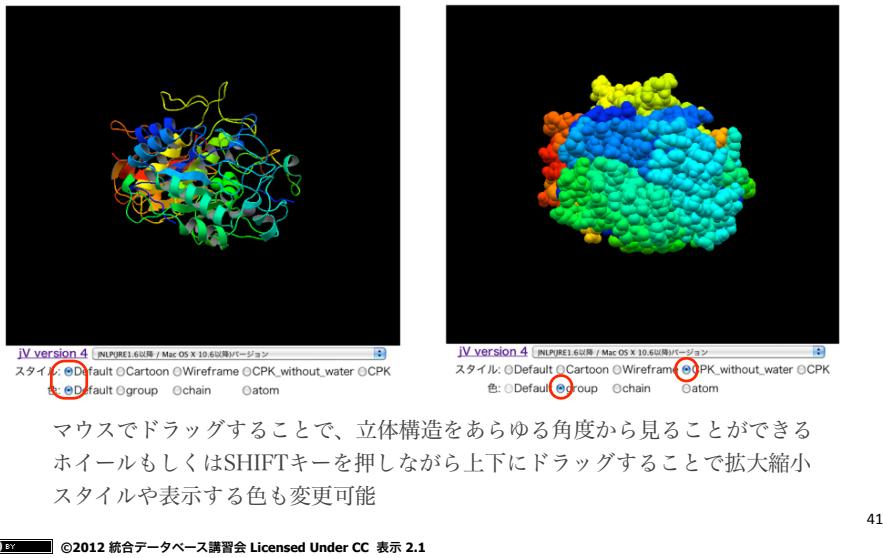
マウスで動かせる画像を表示

JAVAの拡張機能を使っているため、ビューアが起動する前に色々聞いてきます
実行もしくは許可することによって、画像を表示できるようになります
インストールされているJAVAのバージョンによってはうまく表示できない場合があります

40

©2012 統合データベース講習会 Licensed Under CC 表示 2.1

jv4での表示



構造情報ページ

日本語ページについて
PDB Mineについて
更新情報

Mine 構造情報[1clv]

検索結果 実験情報 構造情報 相同蛋白質 ダウンロード/画面表示 外部データベース PDB IDまたはキーワード 検索

エンティティ

識名	説明	種類	データベース名(アクセス番号)	分子量	分子数	由来する生物種	エンティティの一般名
A PROTEIN (ALPHA-AMYLASE)	polymer	UniProt (P5634)	51263.6	1	Tenebrio molitor (yellow mealworm)	ALPHA-1,4-GLUCAN-4-GLUCANOHYDROLASE	
I PROTEIN (ALPHA-AMYLASE INHIBITOR)	polymer	UniProt (P80403)	3595.1	1			
CALCIUM ION	non-polymer	PDBchem (CA)	40.1	1			
CHLORIDE ION	non-polymer	PDBchem (CL)	35.5	1			
water	water	PDBchem (HOH)	18.0	273			

Display chain - A, J All (in new window)

Chain-A: [polypeptide(L)] ALPHA-1,4-GLUCAN-4-GLUCANOHYDROLASE

Number: 1 11 21 31 41 51 61 71 81 91 101 111 121 131 141 151

Sequence: KDNFAFSQNSVHLFEVKVNDIACERFLQGQFGVQSPNPELVACBPRWVEYQPVSYNTESCSAFTDMTEKNDAGVTVYAVNHMTGNGTGSASSADHGMYNTFVYGGDFHESFCEVNNTQADNNVINCIVLGRDNGQS

Structure: Site 44 161 171 181 191 201 211 221 231 241 251 261 271 281 291 301 311 321

331 341 351 361 371 381 391 401 411 421 431 441 451 461 471

非対称単位の内容

ポリマー	鎖の数	分子量の合計
	2	54858.7
非ポリマー	2	75.6
全て	2	54934.3

*水分子は含んでいません

二次構造、ジスルフィド結合、結合部位情報など

42

実験情報ページ

実験情報

精密化の統計情報

標準偏差[A]	119.250	119.250	64.780
標準偏差[実]	90.00	90.00	120.00
空隙割合	P 61		
分解能[A] (低 - 高)	8.00 - 2.00		
R-work	最も高い分解能シェルの値	2.03 - 2.00	
R-free	最も高い分解能シェルの値	0.2577	
R-free	最も高い分解能シェルの値	0.161	
結合角の平均二乗偏差(RMSD)[A]	0.2435		
結合角の平均二乗偏差(RMSD)[実]	0.007		
結合角の平均二乗偏差(RMSD)[R]	1.778		

回折データの統計情報

分解能[A] (低 - 高)	20.0 - 2.00		
最も高い分解能シェルの値	2.03 -		
独立反射数	33598		
観測反射数	279403		
R-measure()	0.135		
完全性[%]	最も高い分解能シェルの値	0.339	
完全性	最も高い分解能シェルの値	94.4	
穴異性	最も高い分解能シェルの値	68.3	

PDBjが文献から抽出した独自の情報
(3極共通のPDBには入っていない)

結晶ID 方法 pH pHの範囲 温度 単位

1	Vapor diffusion, hanging drop	5.4	5.0 - 6.0	23	°C
---	-------------------------------	-----	-----------	----	----

試薬の結晶化条件

ID	結晶ID	溶媒	試薬名	濃度(単位)	詳細
1	1	drop	sodium acetate	5(mM)	
2	1	drop	CaCl ₂	0.1(mM)	
3	1	drop	protein	10(mg/ml)	
4	1	reservoir	PEG1000	12(%(w/v))	
5	1	reservoir	PEG8000	12(%(w/v))	

(*) PDBjによる注釈記

©2012 統合データベース講習会 Licensed Under CC 表示 2.1

43

機能情報のページ

PDBデータベースに由来する情報 site_id: ACS

種類	サイト
残基数	3
詳細	CATALYTIC SITE

site_id: AC1

種類	サイト
残基数	7
詳細	BINDING SITE FOR RESIDUE CA A 601

PDBデータベースに由来する情報 site_id: ACS

site_id	種類	残基数	名前空間
ACS	SITE	3	CATALYTIC SITE
AC1	SITE	7	BINDING SITE FOR RESIDUE CA A 601
AC2	SITE	3	BINDING SITE FOR RESIDUE CL A 602

(UniProt情報の詳細)

SwissProtUniProtに記載されている蛋白質分子機能情報

site_id	種類	残基数	名前空間
SWS_FT_F1	enzyme active site(ACT_SITE)	1	Proton donor.
SWS_FT_F2	enzyme active site(ACT_SITE)	1	Nucleophile
SWS_FT_F3	metal binding site(METAL)	2	Calcium
SWS_FT_F4	metal binding site(METAL)	2	Calcium; via carb
SWS_FT_F5	binding site(BNDING)	3	Chloride,
SWS_FT_F6	other interesting site(SITE)	1	Transition state st

(UniProt情報の詳細)

CSAにおける酵素触媒機能の情報 site_id: AC2

site_id	種類	残基数	名前空間
CSA1	catalytic site	2	P-eBLAST alignment on 1am by CSA
CSA2	catalytic site	3	P-eBLAST alignment on 2cp by CSA
CSA3	catalytic site	3	P-eBLAST alignment on 1nok by CSA

(CSA情報の詳細)

site_id: AC1

種類	サイト
残基数	3
詳細	BINDING SITE FOR RESIDUE CL A 602

site_id: AC2

種類	サイト
----	-----

44

相同蛋白質のページ

概要 調査情報 実験情報 標本情報 **相同蛋白質** ダウンロード/画面表示 外部データベース

PDB ID または キーワード [検索]

エンティティを選択してください。

>1clv A PROTEIN (ALPHA-AMYLASE) Entity#1

EDKANFASGRNVIHLFEWKWNDIADECERFLQPGFGGVQISPNEYLVADGRPWERYQPVSYINTRSGDESAFTDMTRCNDAV
TRRCNDAGVRYIHDVAVINHMTGMNGVGTGSSADHDGMNPYAVPYGSQDHFSPCEVNYYODADVNVRNCLEVLGRDLNGQS
DYVRGLYIDYMNHHMDLGAVGFRVDAKHMSPGDNLQKLNNTDYGAFDAGARPFYIEVIDLGEAISKENYTGF
CYLEFVGFLSGLQPGGNQLKLNANWPGEWLGLQEGDAAVFPVHNDRGTTGGSQILTYKPNPKPVYKMAIAFMALAHYPTT
RIMSSDFDTONDQGPQPGNSQNLSPGNDNTNCQSYVCHRRWVYQGMVGRFNRNAVETGVENVWNSNDNQIAFSRGSQ
QFVAFATGQDQNLLNLTGPACTYQVLSGQGSGCCTQSVTCDNGSADISLGSQAEDDVGLAIHVNAKL

配列が似ているPDBエントリを検索する

>1clv B PROTEIN (ALPHA-AMYLASE) INHIBITOR Entity#2

CIPKWNQDNLQNLTPGACTYQVLSGQGSGCCTQSVTCDNGSADISLGSQAEDDVGLAIHVNAKL

タンパク質が複数あるエントリでは、
どのタンパク質を対象とするか選択する

結果 (1~50) / 75

1clvA 完全一致: 1jmaA 1lmA

Seq.Identity:100% Seq.Positions:100% E-value:0 Score:2514 Query Coverage:100% Compound:PROTEIN (ALPHA-AMYLASE) (3.2.1.1)

1jmaA 完全一致: 1tmaA 検索の重ね合わせ

1tmaA 1471 QKDANFASGRNVIHLFEWKWNDIADECERFLQPGFGGVQISPNEYLVADGRPWERYQPVSYINTRSGDESAFTDMTRCNDAV
1lmaA 1471 QKDANFASGRNVIHLFEWKWNDIADECERFLQPGFGGVQISPNEYLVADGRPWERYQPVSYINTRSGDESAFTDMTRCNDAV

1lmA 完全一致: 1jmaA 1chA

Seq.Identity:99% Seq.Positions:100% E-value:0 Score:2499 Query Coverage:100% Compound:ALPHA-AMYLASE, ALPHA-AMYLASE-INHIBITOR

1jmaA 完全一致: 1tmaA 検索の重ね合わせ

1tmaA 1471 QKDANFASGRNVIHLFEWKWNDIADECERFLQPGFGGVQISPNEYLVADGRPWERYQPVSYINTRSGDESAFTDMTRCNDAV
1tmaA 1471 QKDANFASGRNVIHLFEWKWNDIADECERFLQPGFGGVQISPNEYLVADGRPWERYQPVSYINTRSGNEASFTDMTRCNDAV

Seq.Identity:52% Seq.Positions:57% E-value:5.59566e-143 Score:1296 Query Coverage:106% Compound:alpha-amylase

1tmaA 完全一致: 1tmaA 検索の重ね合わせ

1tmaA 1471 QKDANFASGRNVIHLFEWKWNDIADECERFLQPGFGGVQISPNEYLVADGRPWERYQPVSYINTRSGDESAFTDMTRCNDAV
3vmaA 1498 QKPNPNTDRGTAIVHLEFWRAJIAEACERFLGPXFGAFGVQISPNEYHEVSPWPWWQRYQPSIYNCLCSRSGGENELRDMTRCNDAV

立体構造を重ねあわせて表示することが可能

ダウンロード/画面表示のページ

- ◆PDB形式、mmCIF形式、XML形式
それぞれのテキストファイルを表示
もしくはデータのダウンロードが可能

Mine		ダウンロード/画面表示 [1clv]		日本語ページについて PDB Mineについて 更新	
		紙面 要約 機構情報 実験情報 機能情報	相同蛋白質 ダウンロード/画面表示	外部データベース	
				POB IDまたはキーワード	検索
PDB形式	ファイル形式		ファイル名	画面表示	ダウンロード
	全ての情報 ヘッダのみ	pdb1clv.ent.gz(88k) pdb1clv.ent(364k)	圧縮 非圧縮	(表示) (表示)	(ダブルクリック) (ダブルクリック)
mmCIF	全ての情報 ヘッダのみ	pdb1clv.ent.gz(7k)	メタデータのみ (原子座標なし)	(表示)	(ダブルクリック)
		1clv.cif.gz(11k)		(表示)	(ダブルクリック)
XML	全ての情報 ヘッダのみ (no-atom) 座標情報のみ (ext-atom)	PDBML PDBMLplus	tclv.xml.gz(71k) tclv-neatom.xml.gz(27k) tclv-pdbmlplus.xml.gz(30k)	メタデータのみ メタデータのみ (PDBJ独自)	(表示) (ダブルクリック) (ダブルクリック)
	RDF		tclv-extatom.xml.gz(96k)	原子座標データのみ (メタデータなし)	(表示)
	生物学的単位 (PDB形式)		tclv-neatom.rdf.gz(18k)	RDF形式データ (PDBJ独自)	(表示)
			tclv.pdb1.gz(84k)	*author_defined_assembly_2分子 (dimeric)	(ダブルクリック)

相同蛋白質のページ

立体構造を取得した後の解析例

- ◆ “PyMol”, “Chimera”等で立体構造を描画する
 - ◆ ”DALI”, “MATRAS”等で立体構造を比較する
 - ◆ “SURFNET”等でタンパク質表面のポケット
(化合物が結合しやすい) を検出する
 - ◆ ”DelPhi”等でタンパク質の表面電荷を計算する
 - ◆ ”UCSF DOCK”, “AutoDock”等で化合物との
ドッキングシミュレーションを行う

KEGG

Kyoto Encyclopedia of Genes and Genomes

49

©2012 統合データベース講習会 Licensed Under CC 表示 2.1

KEGG トップページ <http://www.kegg.jp/>

KEGG Home
Release notes
Current statistics
Plea from KEGG
KEGG Database
KEGG の概要
Searching KEGG
KEGG mapping
Contact
KEGG Objects
Pathway maps
Brite hierarchies
KEGG Software
KegTools
KEGG API
KGML
KEGG FTP
利用申請込み
ゲノムネット
DBGET/LinkDB
Feedback
Kanehisa Labs

KEGG: Kyoto Encyclopedia of Genes and Genomes
Announcements
In the next scheduled release on July 1, 2012 a new version of the KEGG API will be made available. It is a REST-based API service, replacing the current SOAP-based API service. more ...
To accommodate the rapid increase of new genomes, five KEGG organism codes will no longer be limited to three letters. more ...
● 興味のあるエントリーポイント
KEGG2
データタイプごとのエントリーポイント
KEGG PATHWAY
KEGG BRITE
KEGG MODULE
KEGG DISEASE
KEGG DRUG
KEGG ENVIRON
KEGG MEDICUS
KEGG ORTHOLOGY
KEGG GENOME
KEGG GENES
KEGG LIGAND
● 生物種ごとのエントリーポイント
KEGG Organism
生物種コード(複数可) [Go] hsa hsa eco

解説ツール
KEGG Mapper
KEGG Atlas
KAAS
BLAST/FASTA
SIMCOMP
PathPred

Copyright 1995-2012 Kanehisa Laboratories

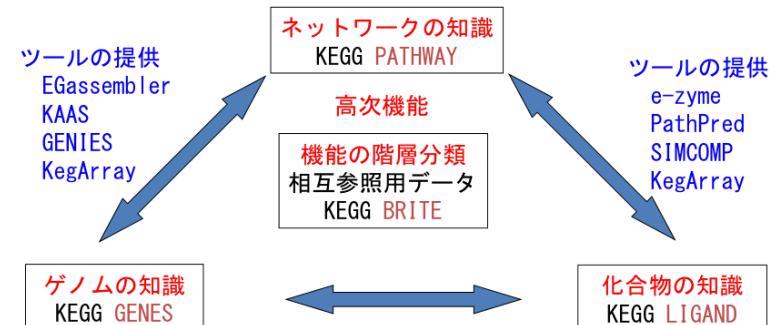
KEGG - Table of Contents
KEGG2 PATHWAY BRITE MODULE DISEASE DRUG KO GENOME GENES LIGAND DBGET
Search (KEGG) for [Go] Clear
Category Entry Point Release Info Search & Compute DBGET Search
Systems information
KEGG PATHWAY
KEGG BRITE
KEGG MODULE
KEGG Mapper
KEGG Atlas
KEGG DISEASE
KEGG DRUG
KEGG ENVIRON
KEGG MEDICUS
KEGG ORTHOLOGY
Genomic information
KEGG GENES
KEGG GENOME
KEGG Organisms
Chemical information
KEGG LIGAND
KEGG COMPOUND
KEGG REACTION
KEGG REACTION
See Kanehisa et al. (2012) for the new features of KEGG.

KEGGはいろいろなDBの集合体
システムの知識
ゲノムの知識
化合物の知識

51

KEGGとは？

◆様々な種類のデータを「生命現象の総体」として再構築



研究者の知識をゲノムレベルのデータと結びつける

50

©2012 統合データベース講習会 Licensed Under CC 表示 2.1 © 2012 時松敏明 (京都大学化学研究所) licensed under CC表示2.1日本

ゲノムの知識データベース・ツール

52

©2012 統合データベース講習会 Licensed Under CC 表示 2.1

ゲノムの知識データベース・ツール

◆KEGG GENES

- ゲノムが解読された生物種の配列に対して、KEGG独自のアノテーションを付けたDB

◆KEGG ORTHOLOGY (KO)

- 各生物のオーソログ（異なる生物に存在する相同な機能を持った遺伝子群）を集めたDB

◆KAAS (KEGG automatic annotation server)

- 新規ゲノムに対して、自動的にKOをアノテーション

53

©2012 統合データベース講習会 Licensed Under CC 表示 2.1

KEGG GENES



KEGG GENES Database
Molecular building blocks of life in the genomic space

KEGG2	PATHWAY	BRITE	MODULE	KO	GENOME	GENES	SSDB	Organisms																	
Enter org:gene (Example: syn:ssr2451)	<input type="button" value="Entry"/>	<input type="button" value="Gene cluster"/>	<input type="button" value="Ortholog"/>	<input type="button" value="Paralog"/>	<input type="button" value="Motif"/>	<input type="button" value="Clear"/>																			
Gene Catalogs																									
KEGG GENES is a collection of gene catalogs for all complete genomes (see release history) generated from publicly available resources, mostly NCBI RefSeq. They are subject to SSB compilation and KO assignment (Gene annotation) by KOALA tool. KEGG GENES for draft genome of some eukaryotes and KEGG GENES for EST dataset of mostly plants are subject to automatic gene catalogs, which are given automatically by KAAS while GENES uses as a reference data set. There is now a fourth type of gene catalog, MGENES for metagenomes (see also KEGG GENOME) with automatic annotation. The viral gene catalog, VGENES, is not yet fully integrated in the KEGG system.																									
<table border="1"><thead><tr><th>Gen catalog</th><th>Category</th><th>Remark</th></tr></thead><tbody><tr><td>GENES</td><td>Complete genomes</td><td>High-quality genomes with KOALA and manual annotations</td></tr><tr><td>DGENES</td><td></td><td>Draft genomes with automatic (KAAS) annotation</td></tr><tr><td>EGENES</td><td>EST datasets</td><td>EST contigs with automatic (KAAS) annotation</td></tr><tr><td>MGENES</td><td>Metagenomes</td><td>Metagenomes with automatic (KAAS) annotation</td></tr><tr><td>VGENES</td><td>Viruses</td><td>No annotation; available only in DBGET</td></tr></tbody></table>	Gen catalog	Category	Remark	GENES	Complete genomes	High-quality genomes with KOALA and manual annotations	DGENES		Draft genomes with automatic (KAAS) annotation	EGENES	EST datasets	EST contigs with automatic (KAAS) annotation	MGENES	Metagenomes	Metagenomes with automatic (KAAS) annotation	VGENES	Viruses	No annotation; available only in DBGET							
Gen catalog	Category	Remark																							
GENES	Complete genomes	High-quality genomes with KOALA and manual annotations																							
DGENES		Draft genomes with automatic (KAAS) annotation																							
EGENES	EST datasets	EST contigs with automatic (KAAS) annotation																							
MGENES	Metagenomes	Metagenomes with automatic (KAAS) annotation																							
VGENES	Viruses	No annotation; available only in DBGET																							

Search [GENES] for amylase Go Clear

bfind mode bgte mode

Search [Organism] hsa for Go Clear

bfind mode bgte mode

2種類の検索モード

bfind: キーワード検索

bgte: KEGG GENES IDがわかっているとき

55

©2012 統合データベース講習会 Licensed Under CC 表示 2.1

演習4

◆”amylase”でKEGG GENESを検索してみましょう

- PDBjとは異なり、日本語を自動的に翻訳してくれないので、英語で入力します

KEGGにアクセスするには「KEGG」で検索
もしくは <http://www.kegg.jp/> を直接入力
そこから KEGG GENES をクリック

54

©2012 統合データベース講習会 Licensed Under CC 表示 2.1

KEGG GENES 検索結果

KEGG Search [GENES] for amylase Go Clear

Database: GENES - Search term: amylase (1 - 1000) (Next)

hsa9432	MGC138178, MGC138180; maltase-glucosidase (alpha-glucosidase) pseudogene
hsa100124692	hsa100124692
hsa110100	Starch and sucrose metabolism
hsa04970	Salivary secretion
hsa04973	Carbohydrate digestion and absorption
hsa04974	[PATH]hsa04974
hsa04975	Starch and sucrose metabolism
hsa04976	[PATH]hsa04976
hsa04977	Organismal Systems: Digestive System; Salivary secretion [PATH]hsa04977
hsa04978	Organismal Systems: Digestive System; Carbohydrate digestion and absorption [PATH]hsa04978
hsa100124691	hsa100124691
hsa100124690	hsa100124690
hsa278	[PATH]hsa04978
hsa91978	TPGS1, C19orf20, GTRGE20, PGs1; tubulin polyglutamylase complex subunit 1
hsa279	[PATH]hsa04978
hsa276	TPS2, C16orf10, DKFZp586M1523, HsT3006, L17, PGs2; tubulin polyglutamylase complex subunit 2
hsa277	[PATH]hsa04978
hsa278	AMY1, AMY1A, AMY1B, AMY1C; amylase, alpha 1C (salivary) (EC:3.2.1.1); K01176 alpha-amylase [EC:3.2.1.1]
hsa279	[PATH]hsa04978
hsa280	AMY1, AMY1A, AMY1B, AMY1C; amylase, alpha 1C (salivary) (EC:3.2.1.1); K01176 alpha-amylase [EC:3.2.1.1]
hsa281	[PATH]hsa04978
hsa282	AMY1, AMY1A, AMY1B, AMY1C; amylase, alpha 1A (salivary) (EC:3.2.1.1); K01176 alpha-amylase [EC:3.2.1.1]
hsa283	[PATH]hsa04978
hsa284	AMY2, AMY2A, AMY2B, PA; amylase, alpha 2A (pancreatic) (EC:3.2.1.1); K01176 alpha-amylase [EC:3.2.1.1]
hsa285	[PATH]hsa04978
hsa286	AMY2B, AMY2; amylase, alpha 2B (pancreatic) (EC:3.2.1.1); K01176 alpha-amylase [EC:3.2.1.1]
hsa287	[PATH]hsa04978
hsa288	AMY2A, AMY2, AMY2B, PA; amylase, alpha 2A (pancreatic) (EC:3.2.1.1); K01176 alpha-amylase [EC:3.2.1.1]
hsa289	[PATH]hsa04978
hsa290	MGAM, MG, MGA; maltase-glucosidase (alpha-glucosidase) (EC:3.2.1.20 3.2.1.3); K12047 maltase-glucosidase [EC:3.2.1.20 3.2.1.3]
hsa291	[PATH]hsa04978
hsa292	CAD: carbamoyl-phosphate synthetase 2, aspartate transcarbamoylase, and dihydroorotase (EC:2.1.3.2 3.5.2.3 3.5.5); K11540 carbamoyl-phosphate synthetase / aspartate carbamoyltransferase / dihydroorotase [EC:6.3.5.2.1 3.5.2.3]
hsa293	[PATH]hsa04978
hsa294	CAD: carbamoyl-phosphate synthetase 2, aspartate transcarbamoylase, and dihydroorotase; K11540 carbamoyl-phosphate synthase / aspartate carbamoyltransferase / dihydroorotase [EC:6.3.5.2.1 3.5.2.3]
hsa295	[PATH]hsa04978
hsa296	p735587
hsa297	tubulin polyglutamylase complex subunit 1-like
hsa298	[PATH]hsa04978
hsa299	tubulin polyglutamylase complex subunit 2-like
hsa300	[PATH]hsa04978
hsa301	tubulin polyglutamylase complex subunit 1-like
hsa302	[PATH]hsa04978
hsa303	MGAM; maltase-glucosidase, intestinal-like; K12047 maltase-glucosidase [EC:3.2.1.20 3.2.1.3]

Homo sapiens (human): 277	
Entry	277 CDS T01001
Gene name	AMY1B, AMY1, AMY1A, AMY1C, MGCI77995
Definition	amylase, alpha 1B (salivary) [EC:3.2.1.1]
Orthology	[KO]1176 phα-amylase [EC:3.2.1.1]
Organism	hsa Homo sapiens
Pathway	[hsa04974] starch and sucrose metabolism
Class	Metabolic pathways
Subsystems	Carbohydrate metabolism
Organismal Systems	Digestive System; Salivary secretion [PATH]hsa04977
Organismal Systems	Digestive System; Carbohydrate digestion and absorption [PATH]hsa04978
SSDB	[Gene] [Paralog] [GFT]
Motif	Alpha-amylase Alpha-amylase_C DUF3078
Other DBs	NCBi-G2: 3654962 NCBi-GeneID: 277 OMIM: 104701 HGNC: 1075 HPI: 18510 Ensembl: ENSG00000174876 Vega: OTTHUMG00000001021 Uniprot: P08276
Structure	PDB: 1QHJ 3BLJ 1Z32 1SM0 1JXK 1JXK 1MFU [Thumbnails] [Jmol]
Position	1q21
AA seq	511 aa [AAseq] [GFF]
NT seq	1536 nt [NTseq]

56

©2012 統合データベース講習会 Licensed Under CC 表示 2.1

KEGG ORTHOLOGYのエントリ

Ko ORTHOLOGY: K01176

Entry	K01176	KO
Name	E3.2.1.1, amyA, malS	
Definition	alpha-amylase [EC:3.2.1.1]	
Pathway	ko00500 Starch and sucrose metabolism ko04973 Carbohydrate digestion and absorption	
Class	Metabolism; Carbohydrate Metabolism; Starch and sucrose metabolism [PATH:ko00500] Organismal Systems; Digestive System; Carbohydrate digestion and absorption [PATH:ko04973] [BRITE hierarchy]	
Other DBs	RN: R02108 COG: COGO366 GO: 0004556	
Genes	HSA: 276(AMY1A) 277(AMY1B) 278(AMY1C) 279(AMY2A) 280(AMY2B) PTR: 457067(AMY2B) 469397(AMY1A) 736558(AMY2A) PON: 100446394(AMY1A) 100447008 MCC: 712893(AMY2B) 713027(AMY2A) MMU: 100043684(Amy2a) 11722(Amy1) RNO: 24203(Amy1a) 497039(Amy2) CFA: 480825 AML: 100465041 BTA: 505049(AMY2B) 539383(AMY2A) SSC: 100152255 397397(AMY2) ECB: 100049851 100051073 MDO: 100032988(AMY1A) OAA: 100078561(AMY2A) GGA: 414139(AMY1A) 414140(AMY2A)	

©2012 統合データベース講習会 Licensed Under CC 表示 2.1

57

KAAS

KAAS - KEGG Automatic Annotation Server
for ortholog assignment and pathway mapping

Request

About KAAS

KAAS (KEGG Automatic Annotation Server) provides functional annotation of genes by BLAST comparisons against the manually curated KEGG GENES database. The result contains KO (KEGG Orthology) assignments and automatically generated KEGG pathways.

- KAAS Help

Complete or Draft Genome

KAAS works best with a complete set of genes in a genome is known. Prepare query amino acid sequences and use the BBH (bi-directional best hit) method to assign orthologs.

- KAAS job request (BBH method)

Partial Genome

KAAS can also be used for a limited number of genes. Prepare query amino acid sequences and use the SBH (single-directional best hit) method to assign orthologs.

- KAAS job request (SBH method)
- KAAS interactive

ESTs

When ESTs are comprehensive enough, a set of consensus contigs can be generated by the EGassembler server and used as a gene set for KAAS with the BBH method. Otherwise, use ESTs as they are with the SBH method.

- KAAS job request (BBH method)
- KAAS job request (SBH method)

Example of Results

KO assignment

KEGG pathway mapping

©2012 統合データベース講習会 Licensed Under CC 表示 2.1

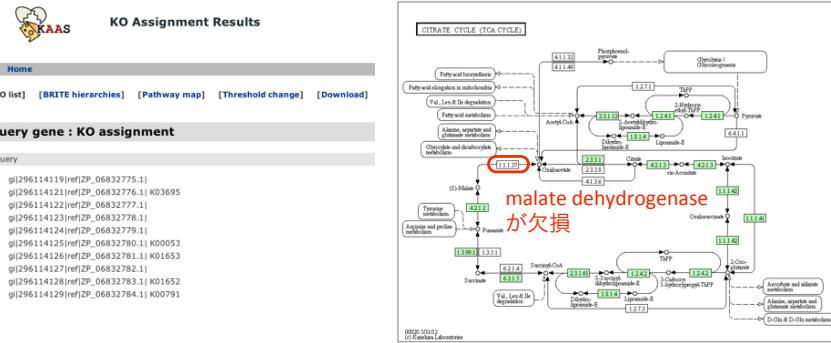
58

KAAS入力と結果

マルチFASTAファイルを入力

```
>gi|296114119|ref|ZP_06832775.1| hypothetical protein GXY_00005 [Glucanacetobacter hansenii ATCC 23769]
MRWGLVMIQACMIVL GKIEHLGSRDAPAFGVW
>gi|296114211|ref|ZP_06832776.1| chaperone clipB [Glucanacetobacter hansenii ATCC 23769]
MNIEKFTERSRGFLQAQTIAMREYNQLTPEHLKALLDDOGAASALIRAAAGQPPAIAAAVDLAK
LPKVQGGGAGGOPSATPDVLRLDAEQAQAKAGDEYVAQDRLAIAASETPAQOALRAGGATPQALDKA
IATIRKGRTVSENAEASFALKYARDTEIALQGKDPVIGRDEEIRRAIQVLARRSKN
>gi|296114122|ref|ZP_06832777.1| hypothetical protein GXY_00019 [Glucanacetobacter hansenii ATCC 23769]
MGAAALIMITLILAGYALCEWPGEEVTPALMQK
```

TCAサイクル



59

化合物の知識データベース・ツール

©2012 統合データベース講習会 Licensed Under CC 表示 2.1

60

化合物の知識データベース・ツール

◆KEGG COMPOUND

- 主に代謝化合物を収録したDB

◆KEGG REACTION

- 生体内反応のDB、これらをつなぎ合わせると
パスウェイになる

◆KEGG LIGAND

- 化合物関係のDBを総称してLIGANDと呼んでいる

61

©2012 統合データベース講習会 Licensed Under CC 表示 2.1

KEGG LIGAND検索結果

KEGG Search [COMPOUND] for phenylalanine [Go] [Clear]

Database: LIGAND - Search term: phenylalanine

KEGG COMPOUND

- C00079 KegDraw Jmol
L-phenylalanine; (S)-alpha-Amino-beta-phenylpropionic acid
C00355 KegDraw Jmol
3,4-Dihydroxy-L-phenylalanine; L-Dopa; 3-Hydroxy-L-tyrosine; L-beta-(3,4-Dihydroxyphenyl)alanine;
Levodopa; Dihydroxy-L-phenylalanine
C02057 KegDraw Jmol
Phenylalanine; DL-Phenylalanine; alpha-Amino-beta-phenylpropionic acid
C0235 KegDraw Jmol
D-Phenylalanine; D-alpha-Amino-beta-phenylpropionic acid
C03519 KegDraw Jmol
N-acetyl-phenylalanine
... display all

KEGG REACTION

- R00031 1,2-Benzenediol:oxygen oxidoreductase; Oxygen + 2-L-Tyrosine <>> 2,3,4-Dihydroxy-L-phenylalanine
R00045 3,4-dihydroxy-L-phenylalanine:oxygen oxidoreductase; Oxygen + 2,3,4-Dihydroxy-L-phenylalanine <>> 2-D-quinolone + 2 H₂O
R00066 L-Phenylalanine racemase (ATP-hydrolyzing); ATP + L-Phenylalanine + H₂O <>> AMP + Diphosphate + D-Phenylalanine
R00097 ATP:L-phenylalanine adenylyltransferase; ATP + L-Phenylalanine <>> Diphosphate + N-Adenylyl-L-phenylalanine
R00688 L-phenylalanine:NAD+ oxidoreductase (deaminating); L-Phenylalanine + H₂O + NAD+ <>> Phenylpyruvate + NH₃ + NADH + H⁺
... display all

KEGG ENZYME

- 1.4.1.20 L-phenylalanine dehydrogenase; L-phenylalanine dehydrogenase; PHD
1.13.12.9 phenylalanine 2-monooxygenase; L-phenylalanine oxiidase (deaminating and decarboxylating); phenylalanine (deaminating, decarboxylating) oxidase
1.13.12.15 3,4-dihydroxyphenylalanine oxidative deaminase; 3,4-dihydroxy-L-phenylalanine: oxidative deaminase;
oxidative deaminase; DOPA oxidative deaminase; DOPAODA
1.14.13.124 phenylalanine N-monoxygenase; phenylalanine N-hydroxylase; CYP79A2

KEGG Search [COMPOUND] for phenylalanine [Go] [Clear]

Database: COMPOUND - Search term: phenylalanine (Total 20 hits)

- C00079 KegDraw Jmol
(S)-alpha-Amino-beta-phenylpropionic acid
C00355 KegDraw Jmol
3,4-Dihydroxy-L-phenylalanine; L-Dopa; 3-Hydroxy-L-tyrosine; L-beta-(3,4-Dihydroxyphenyl)alanine;
Levodopa; Dihydroxy-L-phenylalanine
C02057 KegDraw Jmol
Phenylalanine; DL-Phenylalanine; alpha-Amino-beta-phenylpropionic acid
C0235 KegDraw Jmol
D-Phenylalanine; D-alpha-Amino-beta-phenylpropionic acid
C03519 KegDraw Jmol
N-acetyl-phenylalanine
C00079 KegDraw Jmol
N-Adenyl-L-phenylalanine
C05620 KegDraw Jmol
N-acetyl-L-phenylalanine
C08273 KegDraw Jmol
2,5-Dihydrophenylalanine; L-2,5-Dihydrophenylalanine
C1127 KegDraw Jmol
D-Phenylalanine
C1221 KegDraw Jmol
Formyl-L-leucyl-phenylalanine methyl ester; FMPL OMe
C1136 KegDraw Jmol
N-Acetylphenylalanine beta-naphthyl ester; N-Acetyl-DL-phenylalanine beta-naphthyl ester
C1159 KegDraw Jmol
Phenylalanine beta-naphthyl-L-leucyl-L-phenylalanine
C11615 KegDraw Jmol
Pristimycin IA; Vemanycin B alpha; Mikamycin B; 4-(4-(Dimethylamino)-N-methyl-L-phenylalanine)N-virginiamycin S1; Streptogramin B
C11616 KegDraw Jmol
Pristimycin IB; Vemanycin B beta; Elipristin; 4-[N-Methyl-4-(dimethylamino)-L-phenylalanine]N-virginiamycin S1
C11617 KegDraw Jmol
Pristimycin IC; Vemanycin B gamma; 2-D-Alanine-4-[4-(dimethylamino)-N-methyl-L-phenylalanine]N-virginiamycin S1
C12020 KegDraw Jmol
4-Aminophenylalanine; p-Aminophenylalanine
C17235 KegDraw Jmol
L-Homophenylalanine
C18615 KegDraw Jmol
4-Dimethylamino-L-phenylalanine
C19712 KegDraw Jmol
N-(4-hydroxyphenyl)amine; 2-(Hydroxymethyl)-3-phenylpropanoate
C19715 KegDraw Jmol
N,N-Dihydroxy-L-phenylalanine

ヒット数が多いと検索結果が省略される

display allをクリックすると全部表示される

63

演習5

◆KEGG LIGANDで”phenylalanine”を検索してみましょう



KEGG LIGAND Database

Molecular building blocks of life in the chemical space

KEGG2 PATHWAY BRITE MODULE LIGAND COMPOUND GLYCAN REACTION

Enter C/G/R/RP/RC numbers (Example) C00389 C05903 C06562 C08650 C09727 C09762

Filter Pathway mapping Brute 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, RP, RC, and EC numbers, respectively. ENZYME is derived from the IUBMB/IUPAC Enzyme Nomenclature, but the others are internally developed and maintained.

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

Search [LIGAND] for phenylalanine [Go] [Clear]
blind mode bg mode

62

©2012 統合データベース講習会 Licensed Under CC 表示 2.1

KEGG COMPOUNDのエントリ

KEGG COMPOUND: C00079

Entry	C00079 Compound
Name	L-Phenylalanine; (S)-alpha-Amino-beta-phenylpropionic acid
Formula	C9H11NO2
Exact mass	165.079
Mol weight	165.1891
Structure	
	C00079 Mol file KCF file DB search Jmol KegDraw
Remark	Same as: D00021 BRTE hierarchy
Reaction	R00686 R00687 R00688 R00689 R00690 R00691 R00692 R00693 R00694 R00695 R00697 R00698 R00699 R01375 R01376 R01795 R0360 R06744 R07211 R08435 R08463 R08652 R08690 R09535 R09578 R09579
Pathway	C00360 Phenylalanine metabolism ko00400 Phenylalanine, tyrosine and tryptophan biosynthesis ko00940 Phenylpropanoid biosynthesis ko00960 Tropane, piperidine and pyridine alkaloid biosynthesis ko00966 Glucosinolate biosynthesis ko00970 Aminocycl-RNA biosynthesis map01060 Biosynthesis of plant secondary metabolites map01061 Biosynthesis of phenylpropanoids map01063 Biosynthesis of alkaloids derived from shikimate pathway map01064 Biosynthesis of alkaloids derived from ornithine, lysine and nicotinic acid map01070 Biosynthesis of plant hormones ko01100 Metabolic pathways ko01110 Biosynthesis of secondary metabolites ko02010 ABC transporters ko04974 Protein digestion and absorption ko04978 Mineral absorption

©2012 統合データベース講習会 Licensed Under CC 表示 2.1

64

KEGG REACTIONのエントリ

KEGG REACTION: R00692

Entry	R00692	Reaction
Name	L-phenylalanine:pyruvate aminotransferase	
Definition	L-Phenylalanine + Pyruvate <=> Phenylpyruvate + L-Alanine	
Equation	$\text{C00079} + \text{C00022} \rightleftharpoons \text{C00166} + \text{C00041}$	
RPair	RP00021 C00022_C00041 main RP00057 C00079_C00166 main RP06279 C00022_C00166 trans RP06280 C00041_C00079 trans	
Enzyme	2.6.1.58	
Pathway	rn00360 Phenylalanine metabolism rn00400 Phenylalanine, tyrosine and tryptophan biosynthesis rn00960 Tropane, piperidine and pyridine alkaloid biosynthesis rn01100 Metabolic pathways rn01110 Biosynthesis of secondary metabolites	
Orthology	K05821 aromatic amino acid aminotransferase II [EC:2.6.1.58 2.6.1.28]	

65

システムの知識データベース・ツール

©2012 統合データベース講習会 Licensed Under CC 表示 2.1

66

システムの知識データベース・ツール

◆KEGG PATHWAY

- 代謝系、制御系、疾患などのネットワークを収集したDB

◆Global Map

- 代謝系ネットワークを俯瞰する

◆KEGG MAPPER

- それぞれの酵素や化合物の色を指定して表示

67

演習6

◆KEGG PATHWAYで様々な生物種のパスウェイ、Global Mapを表示して、使われているパスウェイの違いを見てみましょう

0. Global Map

0.1 Metabolism

- Metabolic pathways [zoom out]
Biosynthesis of secondary metabolites [zoom out]
Microbial metabolism in diverse environments [zoom out]

Launch KEGG Atlas
Launch KEGG Atlas
Launch KEGG Atlas

1. Metabolism

1.1 Carbohydrate Metabolism

- Glycolysis / Gluconeogenesis
Citrate cycle (TCA cycle)
Pentose phosphate pathway
Pentose and glucuronate interconversions
Fatty acid metabolism
Galactose metabolism
Ascorbate and aldarate metabolism
Starch and sucrose metabolism
Amino sugar and nucleotide sugar metabolism
Pyruvate metabolism
Glyoxylate and dicarboxylate metabolism
Propanoate metabolism
Butanoate metabolism
C5-Branched dibasic acid metabolism
Inositol phosphate metabolism

Enzymes
Compounds with biological roles

1.2 Energy Metabolism

- Coupling phosphorylation
Photophosphorylation
Photosynthesis
Photosynthesis - antenna proteins

Photosynthesis proteins

©2012 統合データベース講習会 Licensed Under CC 表示 2.1

68

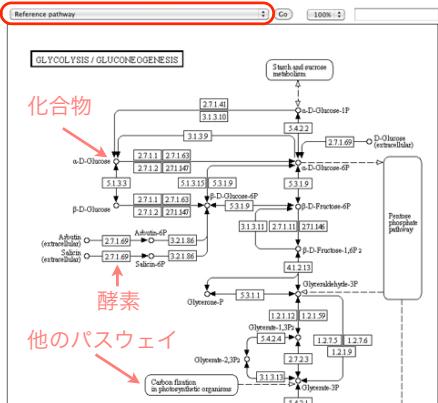
©2012 統合データベース講習会 Licensed Under CC 表示 2.1

KEGG PATHWAY MAP

Glycolysis / Gluconeogenesis - Reference pathway

[Pathway menu | Organism menu | Pathway entry | Hide description | User data mapping]

Glycolysis is the process of converting glucose into pyruvate and generating small amounts of ATP (energy) and NADH (redox power). It is a central pathway that produces important precursor metabolites; six-carbon compounds of glucose-6P and fructose-6P and three-carbon compounds of glyceraldehyde-3P, glycinate-3P, phosphoenolpyruvate, and pyruvate. Glycolysis is a highly conserved pathway found in all eukaryotes and prokaryotes. It is a core module of the metabolic network [MD:M00307]. When the enzyme genes of this pathway are examined in completely sequenced genomes, the reaction steps of three-carbon compounds from glyceraldehyde-3P to pyruvate form a conserved core module [MD:M00021], which is found in almost all organisms. This core module sometimes contains open structures in bacterial genomes. Gluconeogenesis is a synthesis pathway [MD:M00031].



©2012 統合データベース講習会 Licensed Under CC 表示 2.1

69

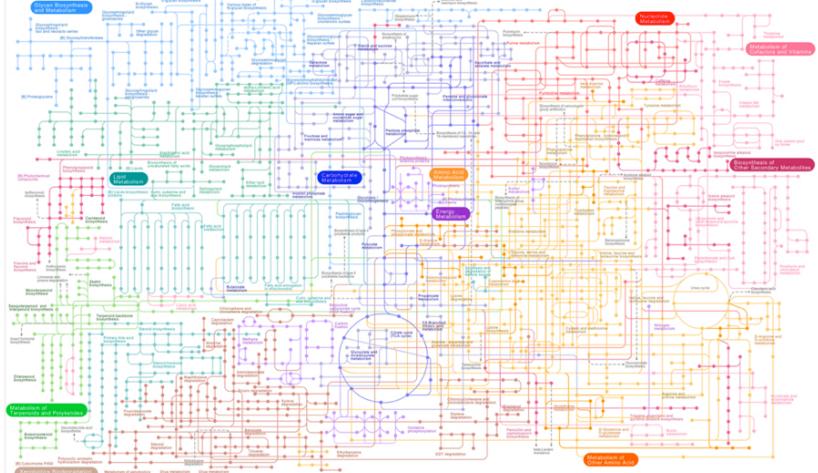
Global Map

Metabolic pathways - Reference pathway

[Pathway menu | Organism menu | Pathway entry | Download KML | User data mapping]

Reference pathway

[Go] 35% [?]



70

©2012 統合データベース講習会 Licensed Under CC 表示 2.1

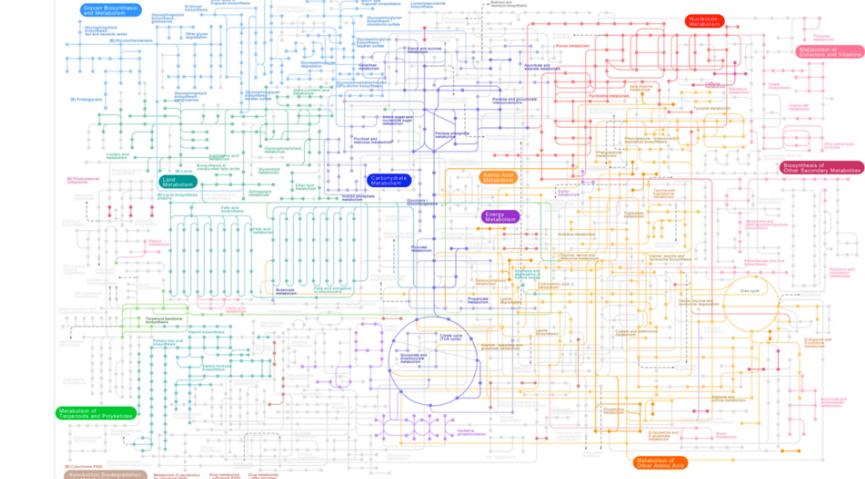
Global Map (Homo sapiens)

Metabolic pathways - Homo sapiens (human)

[Pathway menu | Organism menu | Pathway entry | Download KML | User data mapping]

Homo sapiens (human)

[Go] 35% [?]



71

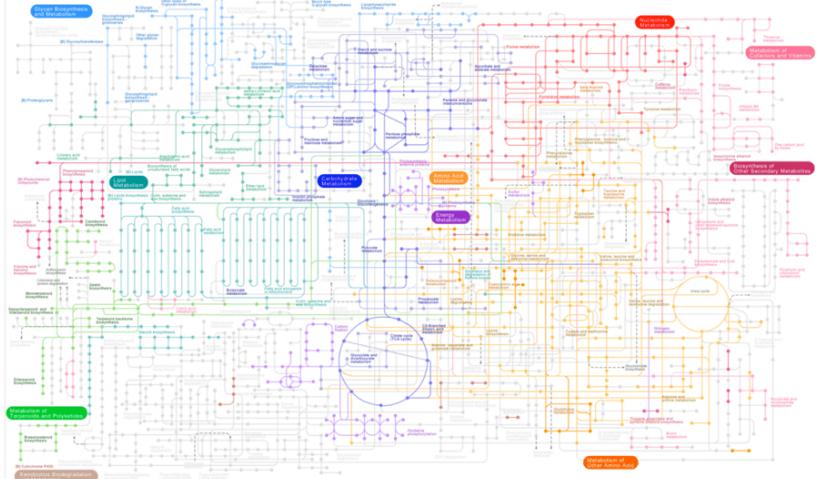
Global Map (Arabidopsis thaliana)

Metabolic pathways - Arabidopsis thaliana (thale cress)

[Pathway menu | Organism menu | Pathway entry | Download KML | User data mapping]

Arabidopsis thaliana (thale cress)

[Go] 35% [?]



72

©2012 統合データベース講習会 Licensed Under CC 表示 2.1

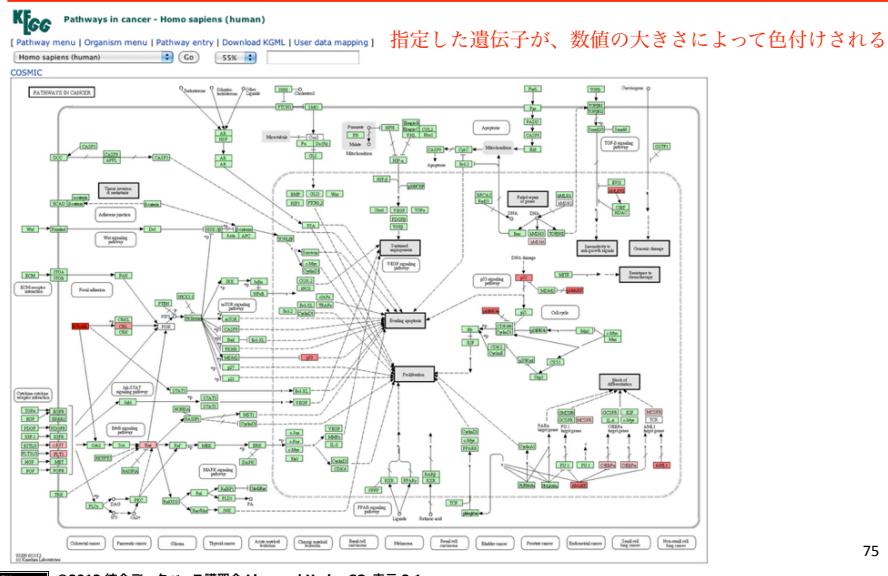
演習7

◆KEGG Mapperを使って、パスウェイの色を変更してみましょう

73

©2012 統合データベース講習会 Licensed Under CC 表示 2.1

KEGG Mapperの結果



KEGG Mapper

Pathway Mapping

KEGG PATHWAY mapping is the process to map molecular datasets, especially large-scale datasets in genomics, transcriptomics, proteomics, and metabolomics, to the KEGG pathway maps for biological interpretation of higher-level systemic functions.

- Search Pathway - basic pathway mapping tool
- Search&Color Pathway - advanced pathway mapping tool
- Color Pathway - selected pathway map coloring tool

KEGG Mapper – Color Pathway

KEGG Mapper	Search Pathway	Search&Color Pathway	Color Pathway	Reconstruct Pathway
KEGG Atlas	Search Brite	Search&Color Brite	Color Pathway 3D	Reconstruct Brite
KEGG	Search Module	Search&Color Module	Join Brite	Reconstruct Module

2. 表示パスウェイを指定

Select KEGG pathway map: **hsa05200**

Color specification example: Cancer stage
hsa05200 + CML-stage
hsa05220 + CML-stage

Enter file name containing the data:

(ファイルを選択) **hsa_CML-COSMIC.txt**

Numerical value example: Somatic mutations
hsa05200 + CML-COSMIC

File type: Color specification

Numerical value (Convert to log scale)
Gradation: from #ffffffff to #ff0000

Use uncolored diagram

3. 保存したファイルを選択

4. File typeをNumerical value, log scaleに変更

©2012 統合データベース講習会 Licensed Under CC 表示 2.1

#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
hsa:3294 94
hsa:3956 2
hsa:3717 40
hsa:3718 2
hsa:3815 5
hsa:3845 6
hsa:4436 1
hsa:4869 1
hsa:4893 15
hsa:5290 1
hsa:7157 23
hsa:7490 4
hsa:7490 2
hsa:10320 3
hsa:54790 5
hsa:55294 1
hsa:171023 6

74

おわりに

◆DDBJ、PDBj、KEGG共に、時間の都合で紹介できなかった機能がたくさんありますので、ぜひ使い倒してみてください

76

©2012 統合データベース講習会 Licensed Under CC 表示 2.1