

統合データベース講習会: AJACS  
2012年6月30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

河野 信

## 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

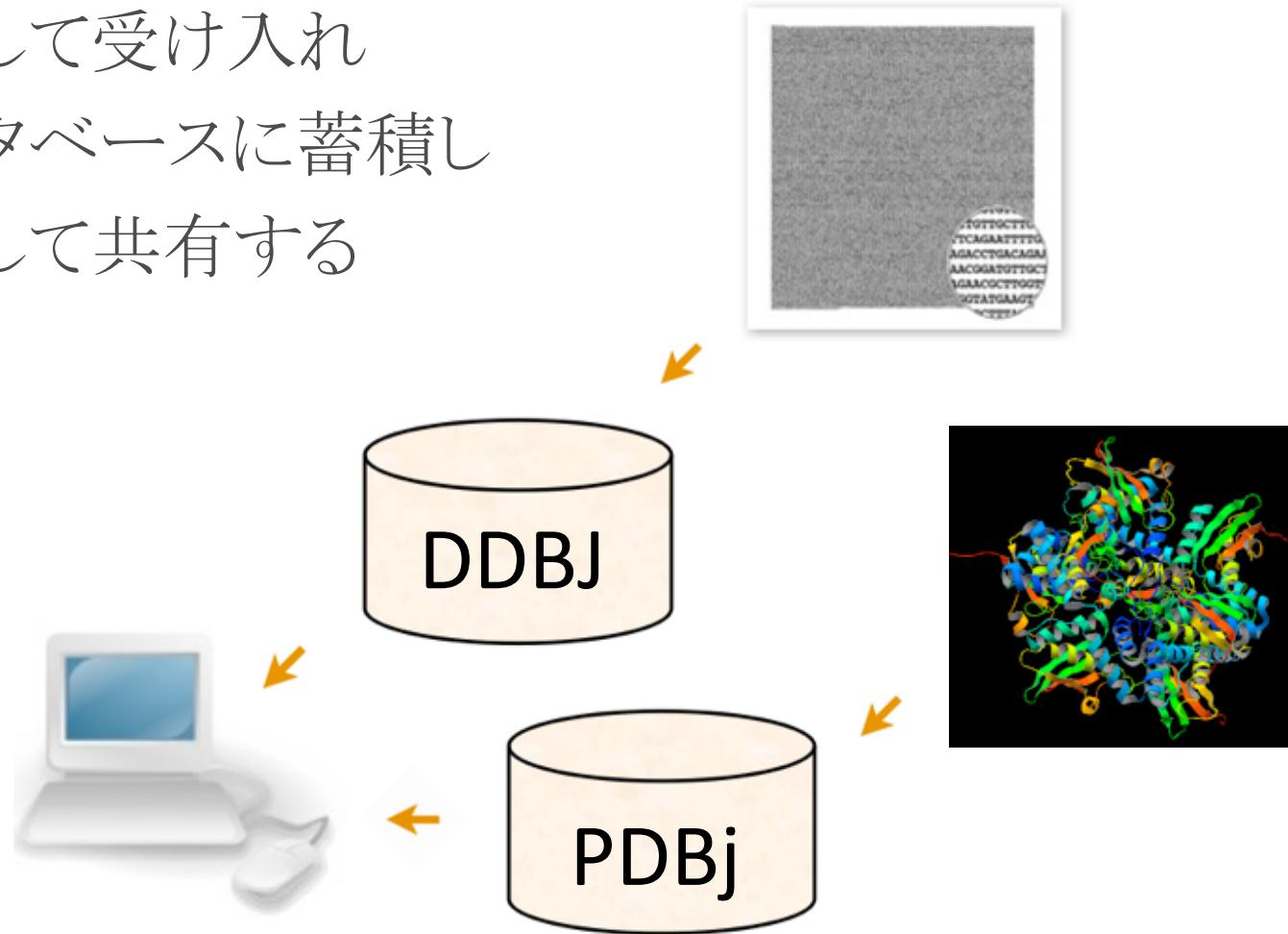
# 本日紹介するDB

---

- ◆ 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

# 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



# 遺伝子・立体構造の論文には登録が不可欠

The screenshot shows the PLOS BIOLOGY homepage. At the top right are links for 'Login', 'Create Account', and 'Feedback'. Below that is a search bar with 'Search articles...' and a 'GO' button. To the right of the search bar are links for 'Advanced Search', 'Browse', and 'RSS'. The main navigation menu at the bottom includes 'Home', 'Browse Articles', 'About', 'For Readers', 'For Authors and Reviewers', 'Journals', 'Hubs', and 'PLOS.org'. The logo for 'PLOS BIOLOGY' is on the left, described as 'a peer-reviewed open-access journal published by the Public Library of Science'.

## 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



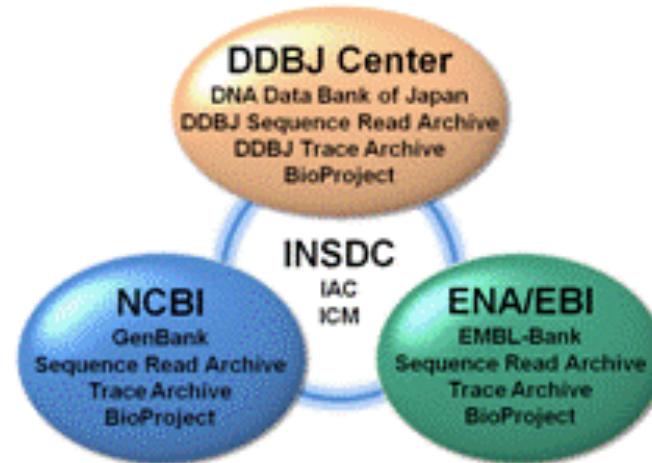
# DDBJ

DNA Data Bank of Japan

# 国際塩基配列データベースの一員

## ◆ International Nucleotide Sequence Databank Collaboration (INSDC)

- 米国: GenBank
- 歐州: 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

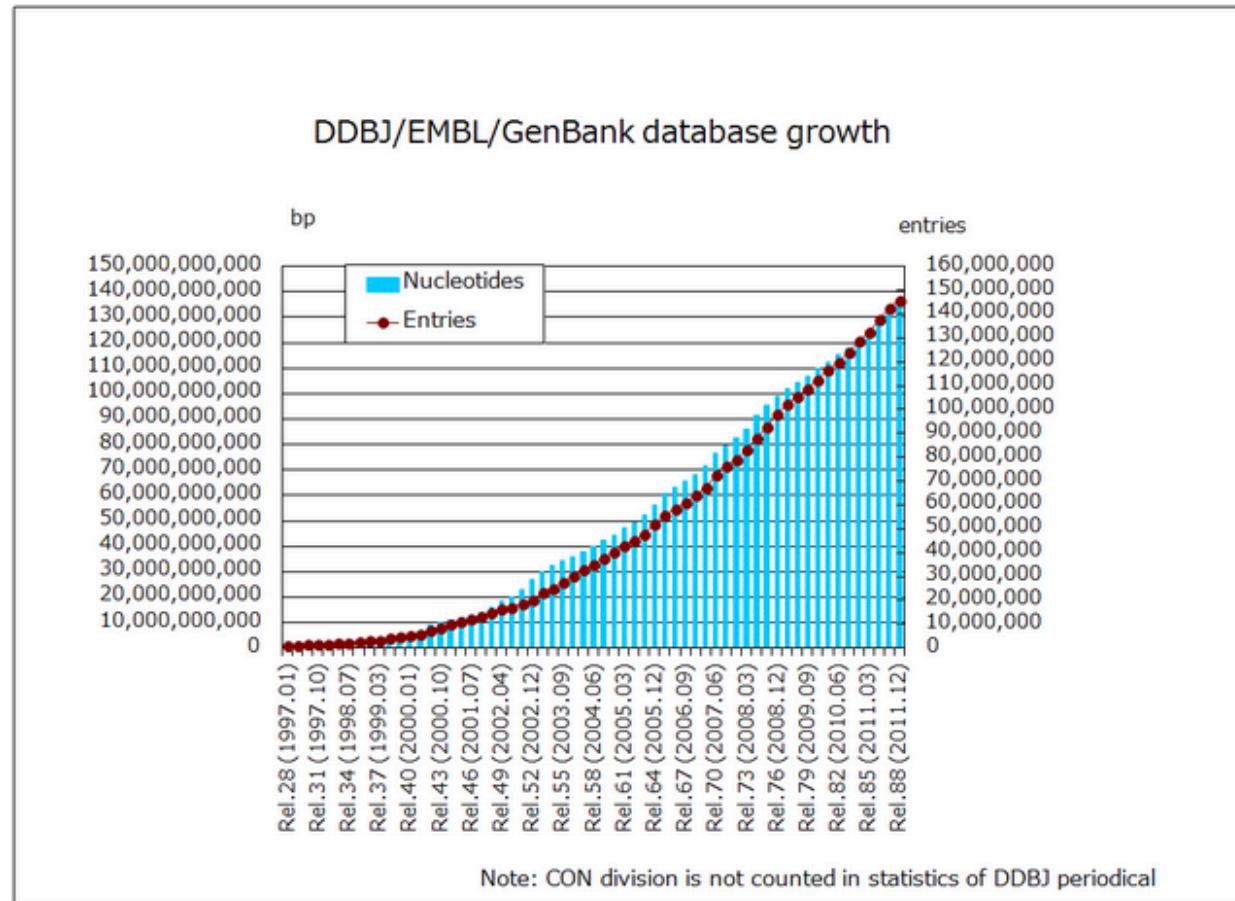
# DDBJ登録ファイルの例

LOCUS	AB091058	2109 bp	DNA	linear	BCT 02-SEP-2003	CDS	1035..2096
DEFINITION	Gluconacetobacter xylinus cmcase, ccp genes for endo-beta-1,4-glucanase, cellulose complementing protein, complete cds.					/codon_start=1 /gene="ccp" /product="cellulose complementing protein" /protein_id=" <a href="#">BAC82541.1</a> " /transl_table=11 /translation="MSASGSDEVAGGGQAGSPQDFQVRVLSRGVEGGQYSYRPFVDRSFDVTGVPEAVERHFQAEHDTAVEEQVTAPQIAVAPPVVPPDPPAIVTETAPPPP VVVSAPVTYEPAAVPAEPPVQEAPVQAAPVPPAPVPPIAEQAPPAAPDPA VAAAPVPPDPAPVTPAPQARVTGPNTRMVEFSRPQVRTQVQEGATPSRVSRSN RTSASSISERPVDRGVADEWSPVPKARLSPRERPRPGDLSFFFQGMRDTRDEKKFFPV ASTRSVRSNVSRSMTSMTDTNSSQASRPGSPVASPDSPTMAEVFMTLGGRATELLS PRPLSREALLRRRENEES"	
ACCESSION	<a href="#">AB091058</a>					343 a	661 c
VERSION	AB091058.1					661 g	444 t
KEYWORDS	.						
SOURCE	Gluconacetobacter xylinus						
ORGANISM	<a href="#">Gluconacetobacter xylinus</a>						
REFERENCE							
AUTHORS	Kawano,S., Tajima,K., 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, Kita-ku, Sapporo, Hokkaido 060-8628, Japan						
REFERENCE	2						
AUTHORS	Kawano,S., Tajima,K., 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	Location/Qualifiers						
source	1..2109 <a href="#">/db_xref="taxon:28448"</a> <a href="#">/mol_type="genomic DNA"</a> <a href="#">/note="synonym:Acetobacter xylinum"</a> <a href="#">/organism="Gluconacetobacter xylinus"</a> <a href="#">/strain="ATCC 53582"</a>						
CDS	10..1038 <a href="#">/codon_start=1</a> <a href="#">/gene="cmcase"</a> <a href="#">/product="endo-beta-1,4-glucanase"</a> <a href="#">/protein_id="BAC82540.1"</a> <a href="#">/transl_table=11</a> <a href="#">/translation="MSVMAAMGGAQVLSSSTGAFADTAPDAVAQQWAIFRAKYLRPSGR VVDTGNNGESHSEGQGYMLFAASAGDLASFQSMWWARTNLQHTNDKLFWRFLKGH QPWPVDKNNATGDLLIALALGRAGRKFQRPDYIQDAMAIYGDVLNLMTMKAGPYVVL MPGAVGFTTKDVSILNLNSYVMPSSLQAFDLTADPRWRQVMEDGIRLVSAGRFGQWRL PPDWLAVNRATGALSIASGWPPRFSYDAIRVPLYFYWAHMLAPNVLADFTRFWNNFGA NALPGWVLDLTGARSPYNAPPGLAVAEACTGLDSAGELPILDHAPDYYSAALTLLVYI ARAEEITK"</a>						
BASE COUNT							
ORIGIN	1 cgttccttta tgcgggtcat ggccggcgat ggaggggccc aggtgcatttc atccaccgg 61 gcgttcgcac accacccccc cgatcgccgc cgccgacaat cggccatctt ccgcggcca 121 tatcttcgtc ccacggcgc tgcgtggat acggcgaatg gtggcgaatc ccatatgg 181 gggcagggtct atggcatgt ctggccgc tcggcggggg accttgcgc tgcgtgg 241 atgtggatgt gggcgccgcac caacctgcag cataccaatg acaagctgtt ttccgtgg 301 ttccctaagg ggcatcagcc cccgggtcccc gacaagaaca atgccacaga tggcgcact 361 ctgatcgcgc ttgcgttgg tcgtgcgggc aagcgttttcc agcgcggccaa ttatcatcg 421 gacccatgg ccattatgg cgatgtgtc aacctgtatg cgtatgaagg gggacccgtat 481 gtgcgttca tgcccggtc tgcgtgggtt accaagaagg acagcgtgtt ctcacatcg 541 tcctattacg tcatgccttc gtcgtgcgcg gtcgtgcggcc ttacggccca cccgcgttgg 601 cgtcaggta tggaaagacgg gattgcgtt gttccgcgg gccgtttccg gcagtggcgc 661 ctggcccccgg actggctggc ggtgaatgcg gccaccggc cgctgtcgat cgcacatcg 721 tggccggcgc gcttttctta tgcgtgcgtt cgggtgcgc tttatttta ttggggccat 781 atgcgtggcgc cgaacgtgtt ggctgtatcc acccgatattt cggggccat 841 gcctccgcgcatgggttgc tgcgtggatcc aggggccgcgtt cgcggccat 901 ggatatcttc ctgttgcgcg acgcacgggg ctgttgcgtt cggggaaactt cccgcacact 961 gatcatgcgc ccgattattttt tccgcgcgcg ttgcgtgcgc tgcgttgcgt 1021 gagggactataaaggatgtgtt ctgcgtgggtt gtcgtggatggt ggctgggggg gggcaggcgt 1081 gaagtccgc ggtatccatgg cgggtcctgc gttcttttggt tgcgtggatggt gggcaggatt 1141 ctcacccggcc gttgtgttgc ctgttgcgtt atgtgcacgg cgtgtccgcgc gctgttggaaa 1201 ggcacttcgcgatgggttgc tgcgtggatcc aggggccgcgtt cgcggccat 1261 aaatcgcgcgatgggttgc cccgcggccatgg tccgcgttgc cccggccat 1321 aaacccgcgc cccgcggccatgg tccgcgttgc gtcgtggatggt ctcgttgcgtt 1381 cccgcgttgc ggcagagccatgg cccgttgcagg aagcccccgtt gcaaggccgcgc cccggccat 1441 cccgcgttgc gcccccgttgc gtcgtggatggt ctcgttgcgtt 1501 tgcgttgc gtcgtggatggt ctcgttgcgtt 1561 cccgcgttgc gtcgtggatggt ctcgttgcgtt 1621 aggtccgcgc gtcgtggatggt ctcgttgcgtt 1681 ctttcccccgc gtcgtggatggt ctcgttgcgtt 1741 atgaatggatggt ctcgttgcgtt 1801 atctgtatggt ctcgttgcgtt 1861 tggcgttgcac gtcgtggatggt ctcgttgcgtt 1921 acaccaatccgc gtcgtggatggt ctcgttgcgtt 1981 ccacaaatccgc gtcgtggatggt ctcgttgcgtt 2041 gtcgttgcgtt ctcgttgcgtt 2101 ctatattca						

# 現在の塩基配列データの量

塩基数:1,400億

登録数:1.5億



DDBJデータベースを検索してみましょう

---

# 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

# DDBJトップページ



ENGLISH



サイト内検索

HOME

塩基配列の登録

利用の手引き

検索・解析

FTP・WebAPI

レポート・統計

お問い合わせ

▶ DDBJの紹介

▶ Q&A集

■ 塩基配列の登録

▶ 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

▶ スバコンの利用申請

▶ スバコンの利用方法

▶ スバコンマニュアル

■ 検索

▶ getentry

▶ ARSA

▶ TXSearch

▶ BLAST

## DDBJ : DNA Data Bank of Japan

DDBJ（日本DNAデータバンク）は欧州と米国との対応機関

(EBIおよびNCBI)と密接に協力しながら DDBJ/EMBL/GenBank

国際塩基配列データベースを構築している三大国際DNAデータバンクのひとつです



Photo by Hidki Nagasaki

### Hot Topics

▶ 一覧へ

■ 2012.06.05 DDBJ Web Magazine のRSS配信を開始します

■ 2012.06.04 ナミアゲハ (*Papilio xuthus*) と シロオビアゲハ (*Papilio polytes*) EST データの公開

■ 2012.06.04 トマト (*Solanum lycopersicum*) 全ゲノムデータの公開

### Maintenance

▶ 一覧へ

### Information

■ DDBJ Web Magazine No.71

### 塩基配列の登録・更新

#### ■ 塩基配列の登録

塩基配列の登録手順を御案内します。

### FTP・Web API

#### ■ FTP ( <ftp://ftp.ddbj.nig.ac.jp> )

DDBJリリースなどのデータファイルをダウンロード



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

# ARSA

ARSA All-round Retrieval of Sequence and Annotation

DDBJQuick Search DDBJAdvanced Search

E. coli O157 を検索

QuickSearch E. coli O157

検索条件を複数入力する場合は、&(AND条件)、|(OR条件)、!(AND NOT条件)を指定することができます。

ARSA All-round Retrieval of Sequence and Annotation

DDBJQuick Search DDBJAdvanced Search

QuickSearch E. coli O157

検索条件を複数入力する場合は、&(AND条件)、|(OR条件)、!(AND NOT条件)を指定することができます。

FlatFile XML fasta [View](#) [Download](#)

[All Select](#)

PrimaryAccessionNumber	Definition	moltyp	Organism	Length
<input type="checkbox"/> AB011548	Escherichia coli O157:H7 str. Sakai plasmid pOSAK1 DNA, complete sequence.	DNA	Escherichia coli O157:H7 str. Sakai	3306
<input type="checkbox"/> AB011549	Escherichia coli O157:H7 str. Sakai plasmid pO157 DNA, complete sequence.	DNA	Escherichia coli O157:H7 str. Sakai	92721
<input type="checkbox"/> AB035920	Escherichia coli O157:H7 hemG, rrsA, ileT, alaT, rrlA, 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.	DNA	Escherichia coli O157:H7	7003

検索結果：  
ヒット数が多くるので  
絞り込みが必要



詳細検索  
(Advanced Search)へ

<a href="#">DDBJ(2195)</a>
<a href="#">Patent_AA(3)</a>
<a href="#">Escherichia coli(731)</a>
<a href="#">Unknown_(561)</a>
<a href="#">Escherichia coli O157:H7(241)</a>
<a href="#">unidentified(167)</a>
<a href="#">synthetic</a>
<a href="#">construct(160)</a>
<a href="#">Escherichia coli O157:H7 str.</a>
<a href="#">EDL933(130)</a>
<a href="#">Pasteuria penetrans(15)</a>
<a href="#">Gallus gallus(10)</a>
<a href="#">Hydractinia echinata(8)</a>
<a href="#">Mycobacterium bovis</a>
<a href="#">AF2122/97(8)</a>
<a href="#">All other taxa(167)</a>



# いくつかの特徴で絞り込み

ARSA All-round Retrieval of Sequence and Annotation

DDBJQuick Search DDBJAdvanced Search

- フィールド内で検索条件を複数入力する場合は、& (AND条件)、| (OR条件)、! (AND NOT条件) を指定することができます。
- ダブルクオーテーション ("")で囲まれた文字列は、1つのキーワードとして認識されます。
- 検索方法および検索条件の入力例などを知りたい方は[こちら](#)をクリックして下さい。

Search reset

Combine Searches with &(AND)

All Text =  -

Accession Number =  -

Primary Accession Number =  -

Division  BCT  CON  ENV  HTC  HTG  HUM  INV  
 MAM  PAT  PHG  PLN  PRI  ROD  STS  
 SYN  TSA  UNA  VRL  VRT

Sequence Length =  -  =  -

---

Molecular

Type  DNA  RNA  cRNA  mRNA  rRNA  tRNA

Form  circular  linear

---

Date =  -  =  -

Definition =  -

Comment =  -

Keyword =  -

Organism =  -  E. coli

Taxon =  -

ヒットしない…

ARSA All-round Retrieval of Sequence and Annotation

DDBJQuick Search DDBJAdvanced Search

DDBJ(5) Refine Search

FlatFile XML fasta View Download

All Select

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

# 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

DDBJQuick Search DDJAdvanced Search

DDBJ(3) Refine Search

FlatFile XML fasta View Download

All Select

PrimaryAccessionNumber	Definition	moltype	Organism	Length
<input type="checkbox"/> AB011548	Escherichia coli O157:H7 str. Sakai plasmid pOSAK1 DNA, complete sequence.	DNA	Escherichia coli O157:H7 str. Sakai	3306
<input type="checkbox"/> AB011549	Escherichia coli O157:H7 str. Sakai plasmid pO157 DNA, complete sequence.	DNA	Escherichia coli O157:H7 str. Sakai	92721
<input checked="" type="checkbox"/> 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]

6

## 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

# NCBI Entrez Search

文献

塩基配列  
EST

アミノ酸配列  
ゲノム配列  
立体構造

NCBI

HOME | SEARCH | SITE MAP

PubMed All Databases Human Genome GenBank Map Viewer BLAST

Search across databases    Help

Welcome to the Entrez cross-database search page

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

 <b>Nucleotide:</b> Core subset of nucleotide sequence records	 <b>dbGaP:</b> genotype and phenotype
 <b>EST:</b> Expressed Sequence Tag records	 <b>UniGene:</b> gene-oriented clusters of transcript sequences
 <b>GSS:</b> Genome Survey Sequence records	 <b>CDD:</b> conserved protein domain database
 <b>Protein:</b> sequence database	 <b>Clone:</b> integrated data for clone resources
 <b>Genome:</b> whole genome sequences	 <b>UniSTS:</b> markers and mapping data
 <b>Structure:</b> three-dimensional macromolecular structures	 <b>PopSet:</b> population study data sets
 <b>Taxonomy:</b> organisms in GenBank	 <b>GEO Profiles:</b> expression and molecular abundance profiles
 <b>SNP:</b> short genetic variations	 <b>GEO DataSets:</b> experimental sets of GEO data



# Escherichia coli O157 Sakai で検索

文献

塩基配列  
EST

アミノ酸配列

ゲノム配列

立体構造

The screenshot shows the NCBI Entrez search results for the query "Escherichia coli o157 sakai". The search bar contains the query. Below it, a note says "- Result counts displayed in gray indicate one or more terms not found". The results are grouped into two main sections: "PubMed" and "All Databases".

**PubMed:**

- 105 PubMed: biomedical literature citations and abstracts
- 233 PubMed Central: free, full text journal articles
- 1 Site Search: NCBI web and FTP sites

**All Databases:**

- none Books: online books
- none OMIM: online Mendelian Inheritance in Man

**Other Databases:**

- 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 Taxonomy: organisms in GenBank
- none SNP: short genetic variations
- 2 dbGaP: genotype and phenotype
- none UniGene: gene-oriented clusters of transcript sequences
- 2 CDD: conserved protein domain database
- none Clone: integrated data for clone resources
- none UniSTS: markers and mapping data
- 3591 PopSet: population study data sets
- none GEO Profiles: expression and molecular abundance profiles
- 142 GEO DataSets: experimental sets of GEO data



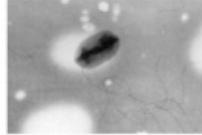
# Entrez: 大腸菌ゲノムページ

NCBI Resources How To My NCBI Sign In

Genome Genome Escherichia coli o157 sakai Save search Limits Advanced Help

Display Settings: Overview Send to:

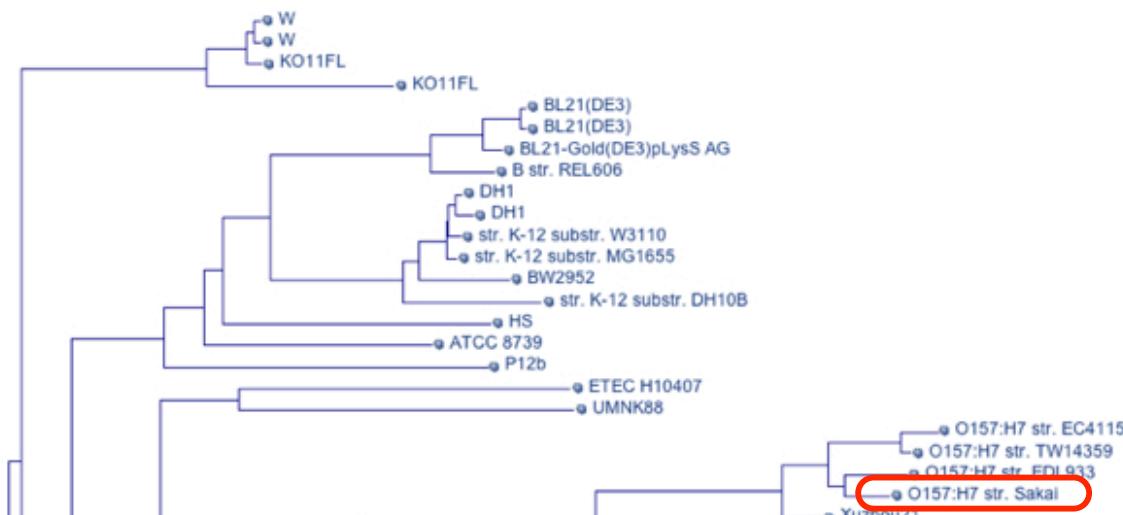
Organism Overview ; Genome Project Report ; Genome Annotation Report ; Plasmid Annotation Report

 **Escherichia coli**  
A well-studied enteric bacterium

Lineage: Bacteria[2817]; Proteobacteria[1179]; Gammaproteobacteria[514]; Enterobacteriales[131]; Enterobacteriaceae[131]; Escherichia[7]; 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)



Related information

- BioProject
- Gene
- Protein Clusters
- Components
- Protein
- PubMed
- Taxonomy

Search details

"Escherichia coli O157"[Organism] AND sakai[All Fields]

Search See more...

Recent activity

Turn Off Clear

Escherichia coli Genome

Escherichia coli o157 sakai (1) Genome

See more... 20



# 大腸菌O157 Sakai株のページ

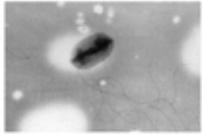
NCBI Resources How To My NCBI Sign In

Genome Genome Search Limits Advanced Help

Display Settings: Overview Send to:

Return to *Escherichia coli*

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]; Enterobacteriales[131]; Enterobacteriaceae[131]; *Escherichia*[7]; *Escherichia coli*[1]; *Escherichia coli* O157:H7[0]; *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

Organism	BioProject	Assembly	Status	Chrs	Plasmids	Size (Mb)	GC%	Gene	Protein
<i>Escherichia coli</i> O157:H7 str. Sakai	PRJNA57781, PRJNA226	ASM886v1	◆	1	2	5.59	50.4	5,460	5,318

Chromosomes [1] Scaffolds or contigs [0] SRA or Traces [0] No data [0]

Genome Region

Go to nucleotide Graphics FASTA GenBank

See more...

Related information

- BioProject
- Gene
- Protein Clusters
- Components
- Protein
- PubMed
- Taxonomy

Recent activity

- Turn Off Clear
- Escherichia coli O157:H7 str. Sakai Genome
- Escherichia coli Genome
- Escherichia coli o157 sakai (1) Genome

# 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: 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\_002695.1

FASTA Graphics

Go to:

Locus NC\_002695 5498450 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; Enterobacteriales;  
Enteracteriaceae; Escherichia.  
Reference 1 (bases 1 to 5498450)  
Authors Bergholz,T.M., Wick,L.M., Qi,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  
Remark Publication Status: Online Only  
Reference 2 (sites)  
Authors Hayashi,T., Makino,K., Ohnishi,M., Kurokawa,K., Ishii,K.,  
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  
Shinagawa,H.  
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(2):96]

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

Customize view

Basic Features  
 Default features  
 Gene, RNA, and CDS features only

Display options  
 Show sequence  
 Show reverse complement

Update View

Analyze this sequence

Run BLAST

Pick Primers

Highlight Sequence Features

LinkOut to external resources

REBASE enzyme XfaMrrP  
[REBASE - The Restriction Enzy...]  
REBASE enzyme M.Ecp933DamP  
[REBASE - The Restriction Enzy...]  
REBASE enzyme M.EphHK97DamP  
[REBASE - The Restriction Enzy...]  
REBASE enzyme M.EcoCR63FP  
[REBASE - The Restriction Enzy...]  
REBASE enzyme M.EcoVT2Dam  
[REBASE - The Restriction Enzy...]  
REBASE enzyme S.EckoKO157ORFAP  
[REBASE - The Restriction Enzy...]  
REBASE enzyme EcoKO157ORF5262P  
[REBASE - The Restriction Enzy...]

24

# 配列を取得した後の解析例

- ◆ “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を設計する
- ◆ ”clustalW”等でマルチプ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

## データを扱う際の注意点

---

# Strange things in the Refseq / nrDB

---

- ◆ protain (183) < protein
- ◆ imilar to (28) < similar to
- ◆ simila to (22) < similar to
- ◆ cromosome (4) < chromosome
- ◆ RNA olymerase < RNA polymerase
- ◆ dehydrogenas, ehydrogenase
- ◆ transposas, ransposase

現状、ヘンなアノテーションが頻出しています

# Identifier “mutation” by Excel

**BMC Bioinformatics**



Correspondence

**Open Access**

## **Mistaken Identifiers: Gene name errors can be introduced inadvertently when using Excel in bioinformatics**

Barry R Zeeberg<sup>†1</sup>, Joseph Riss<sup>†2</sup>, David W Kane<sup>3</sup>, Kimberly J Bussey<sup>1</sup>, Edward Uchio<sup>4</sup>, W Marston Linehan<sup>4</sup>, J Carl Barrett<sup>2</sup> and John N Weinstein\*<sup>1</sup>

Address: <sup>1</sup>Genomics & Bioinformatics Group, Laboratory of Molecular Pharmacology, Center for Cancer Research (CCR), National Cancer Institute (NCI), National Institutes of Health (NIH), Bldg 37 Rm 5041, NIH, 9000 Rockville Pike, Bethesda, MD 20892 USA, <sup>2</sup>Laboratory of Biosystems and Cancer, CCR, Bldg 37 Rm 5032, NIH, 9000 Rockville Pike, Bethesda, MD 20892 USA, <sup>3</sup>SRA International, 4300 Fair Lakes CT, Fairfax, VA 22033 USA and <sup>4</sup>Urologic Oncology Branch, Bldg 10 Rm 2B47, National Institutes of Health, Bethesda, MD 20892 USA

Email: Barry R Zeeberg - barry@discover.nci.nih.gov; Joseph Riss - rissj@helix.nih.gov; David W Kane - david\_kane@sra.com; Kimberly J Bussey - busseyk@mail.nih.gov; Edward Uchio - eu8v@nih.gov; W Marston Linehan - linehanm@mail.nih.gov; J Carl Barrett - barrett@mail.nih.gov; John N Weinstein\* - weinstein@dtpvx2.ncifcrf.gov

\* Corresponding author    †Equal contributors

Published: 23 June 2004

Received: 05 March 2004

BMC Bioinformatics 2004, 5:80 doi:10.1186/1471-2105-5-80

Accepted: 23 June 2004

This article is available from: <http://www.biomedcentral.com/1471-2105/5/80>

© 2004 Zeeberg et al; licensee BioMed Central Ltd. This is an Open Access article: verbatim copying and redistribution of this article are permitted in all media for any purpose, provided this notice is preserved along with the article's original URL.

# Identifier “mutation” by Excel

The screenshot shows an Excel spreadsheet with the following data:

	gene names	internal date format	default date format	gene names	internal date format	default date format	gene names	internal date format	default date format
1	APR-1	35885	1-Apr	OCT-1	36068	1-Oct	SEP2	36039	2-Sep
2	APR-2	35886	2-Apr	OCT-2	36069	2-Oct	SEP3	36040	3-Sep
3	APR-3	35887	3-Apr	OCT-3	36070	3-Oct	SEP4	36041	4-Sep
4	APR-4	35888	4-Apr	OCT-4	36071	4-Oct	SEP5	36042	5-Sep
5	APR-5	35889	5-Apr	OCT-6	36073	6-Oct	SEP6	36043	6-Sep
6	DEC-1	36129	1-Dec	OCT1	36068	1-Oct	SEPT1	36038	1-Sep
7	DEC-2	36130	2-Dec	OCT11	36078	11-Oct	SEPT2	36039	2-Sep
8	DEC1	36129	1-Dec	OCT2	36069	2-Oct	SEPT3	36040	3-Sep
9	DEC2	36130	2-Dec	OCT3	36070	3-Oct	SEPT4	36041	4-Sep
10	MAR1	35854	1-Mar	OCT4	36071	4-Oct	SEPT5	36042	5-Sep
11	MAR2	35855	2-Mar	OCT6	36073	6-Oct	SEPT6	36043	6-Sep
12	MAR3	35856	3-Mar	OCT7	36074	7-Oct	SEPT7	36044	7-Sep
13	NOV1	36099	1-Nov	SEP-1	36038	1-Sep	SEPT8	36045	8-Sep
14	NOV2	36100	2-Nov	SEP-2	36039	2-Sep	SEPT9	36046	9-Sep
15				SEP1	36038	1-Sep			

# SEPT2 → 2-Sep case in Refseq

LOCUS XM\_392412 2125 bp mRNA linear INV 12-APR-2011  
DEFINITION PREDICTED: Apis mellifera **septin-2 (2-Sep)**, mRNA.  
ACCESSION XM\_392412  
VERSION XM\_392412.4 GI:328785636  
KEYWORDS .  
SOURCE Apis mellifera (honey bee)  
ORGANISM Apis mellifera  
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
Neoptera; Endopterygota; Hymenoptera; Apocrita; Aculeata; Apoidea;  
Apidae; Apis.  
COMMENT MODEL REFSEQ: This record is predicted by automated computational analysis. This record is derived from a genomic sequence (NW\_003378075) annotated using gene prediction method: GNOMON, supported by EST evidence.  
Also see:  
Documentation of NCBI's Annotation Process

On Apr 12, 2011 this sequence version replaced gi:110757583.

FEATURES Location/Qualifiers

source 1..2125  
/organism="Apis mellifera"  
/mol\_type="mRNA"  
/strain="DH4"  
/db\_xref="taxon:7460"  
/linkage\_group="LG6"

gene 1..2125  
/gene="**2-Sep**"  
/note="Derived by automated computational analysis using gene prediction method: GNOMON. Supporting evidence includes similarity to: 436 ESTs, 11 Proteins"  
/db\_xref="BEEBASE:GB17411"  
/db\_xref="GenID:408882"

misc\_feature 164..166  
/gene="**2-Sep**"  
/note="upstream in-frame stop codon"

CDS 194..1444  
/gene="**2-Sep**"  
/codon\_start=1  
/product="septin-2"  
/protein\_id="XP\_392412.2"

septin-2 は  
**SEPT2**と記述されるはず、が  
Excelの自動変換機能により  
**2-Sep** と記載されてしまった

教訓1: DBのアノテーションを過信してはいけない

教訓2: Excelを使うときには自動変換に注意

新型シーケンサからのデータ

---

# DDBJ Sequence Read Archive (DRA)

## ◆新型シーケンサデータを保存・共有

1. データ公開

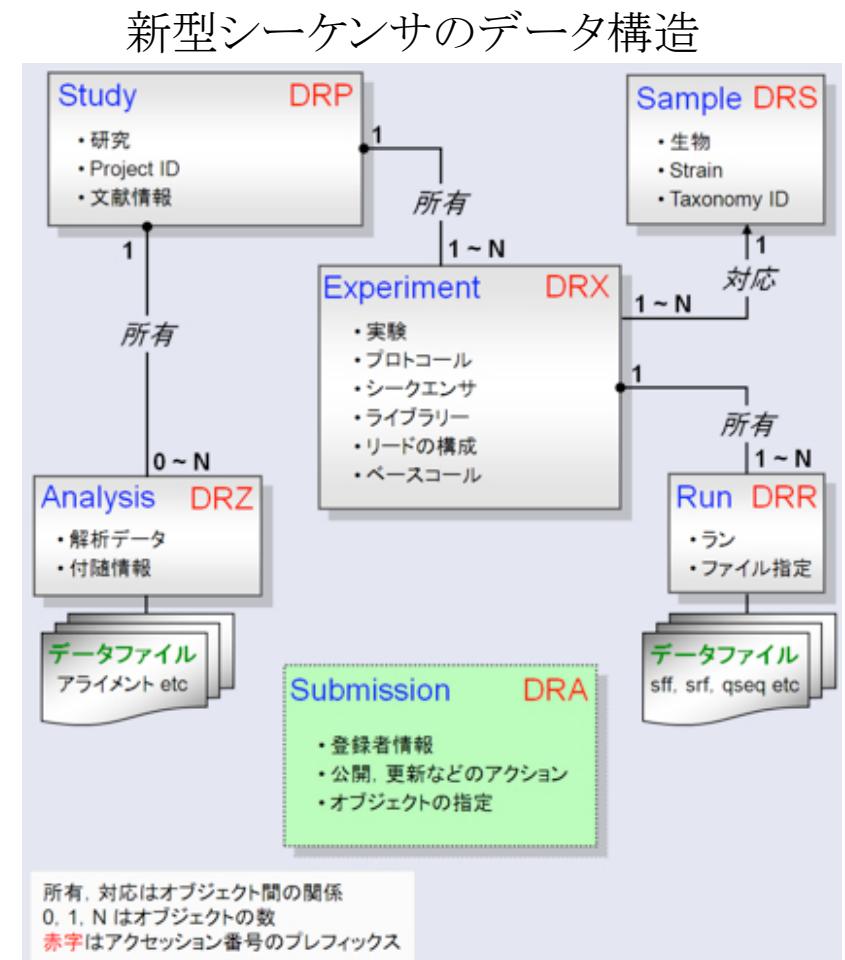
DDBJ Sequence Read Archive (DRA) では以下の場合にデータを公開しています。

- A. 登録者から公開依頼の連絡を受けた場合
- B. 登録者がアクセッション番号を公表した事を確認した場合。  
登録者以外の第三者が過失により他のアクセッション番号を論文や学会等で公表した場合は該当しません。  
公表とは、アクセッション番号を不特定多数の対象に知らせる行為（学術論文、学会、インターネット、報道機関などを媒体とした発表）を指します。
- C. 公開予定日が到来した場合
- D. DRA Run (DRR) アクセッション番号が引用されている DDBJ/EMBL-Bank/GenBank レコード (TSA, WGS, CON など) が公開された場合

B, C または D の場合は登録者の了解がなくても例外なく公開します。

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



# 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.dbcls.jp/>

- 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

The screenshot shows the homepage of the SRAs website. At the top, there is a logo featuring a cartoon submarine and the text "SRAs: Survey of Read Archives". Below the logo, there are two main sections: "What's SRAs? - SRAsって何?" and "Simple Lists - まずは見てみる". The "What's SRAs?" section contains a brief description of SRAs as an index of next-generation sequencing data and a link to the public databases (SRA, ENA, DRA). The "Simple Lists" section provides links to search by study, experiment, or run. Below these, there is a search form for studies with fields for SRA#, Study#, TAXON ID, Platform, Title keyword, and Type, along with "Search Study" and "Reset" buttons.

# 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.

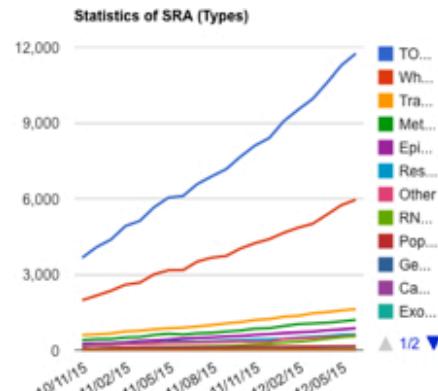
実際のプロジェクト数は"Study Types"に書かれているものです。"Platforms"や"Species of Samples"のtotalの値は、複数のプラットフォームで行われた実験が1つのプロジェクトでなされる場合が多くあり、ダブルカウントしているので、数字が大きくなっています。

2012-06-21 updated.

### Study Types

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

#### • Growth of the number of studies (including Total)



#### • Growth of the number of studies (without Total)

### Platform

Illumina Genome Analyzer II	3629	Illumina HiSeq 2000	45470
454 GS FLX Titanium	2848	Illumina Genome Analyzer II	36940
Illumina HiSeq 2000	2328	454 GS FLX Titanium	11977
454 GS FLX	1735	Illumina Genome Analyzer	9051
Illumina Genome Analyzer	1167	454 GS FLX	8939
Illumina Genome Analyzer Ibx	565	Illumina Genome Analyzer Ibx	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 HeliScope	332
AB SOLID System	37	PacBio RS	326

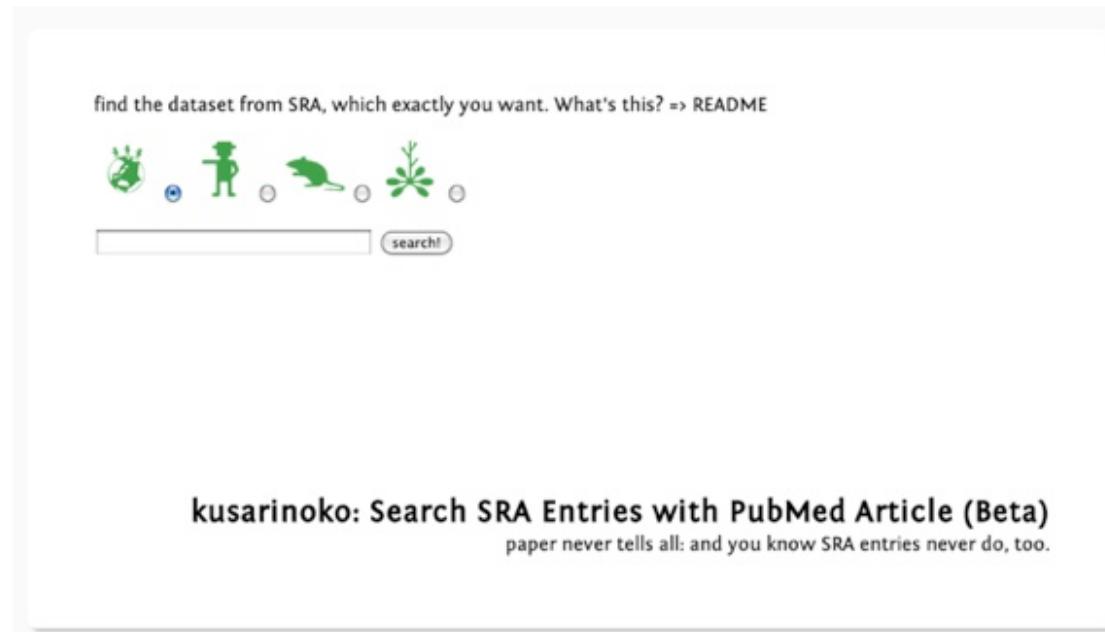
### Species of samples (top 15)

<i>Homo sapiens</i>	1107	<i>Homo sapiens</i>	38506
unidentified	886	unidentified	8678
<i>Mus musculus</i>	603	<i>Streptococcus pneumoniae</i>	7155
<i>Drosophila melanogaster</i>	253	<i>Mus musculus</i>	7004
<i>Caenorhabditis elegans</i>	181	<i>Drosophila melanogaster</i>	5523
metagenome sequence	180	human metagenome	3026
marine metagenome	160	<i>Danio rerio</i>	2940
<i>Escherichia coli</i> str. K-12 substr. MG1655	135	<i>Plasmodium falciparum</i>	2281
<i>Arabidopsis thaliana</i>	134	<i>Staphylococcus aureus</i>	1965
soil metagenome	115	<i>Mycobacterium tuberculosis</i>	1377
<i>Saccharomyces cerevisiae</i>	107	<i>Saccharomyces cerevisiae</i>	1325
<i>Mustela putorius furo</i>	99	<i>Oryza sativa</i>	1279
<i>Human rhinovirus A</i>	94	<i>Sus scrofa</i>	1268
<i>Salmo salar</i>	56	marine metagenome	1217
<i>Plasmodium falciparum</i>	50	<i>Arabidopsis thaliana</i>	1212
Total	14554	Total	127426
(studies)	(experiments)	(studies)	(experiments)



# 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>



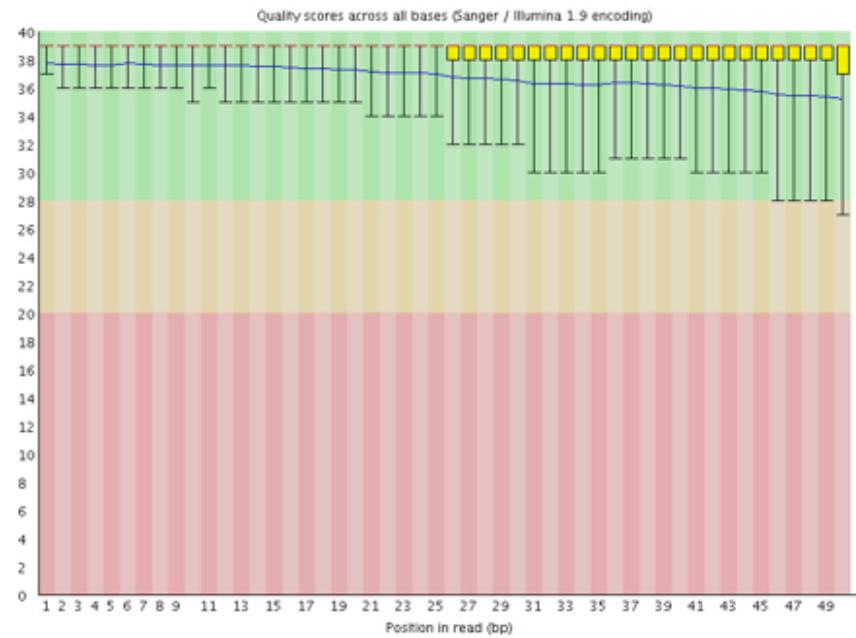
# 鎖鋸(kusarinoko)

◆ 独自に”FastQC”をかけてそれぞれのデータの質を評価

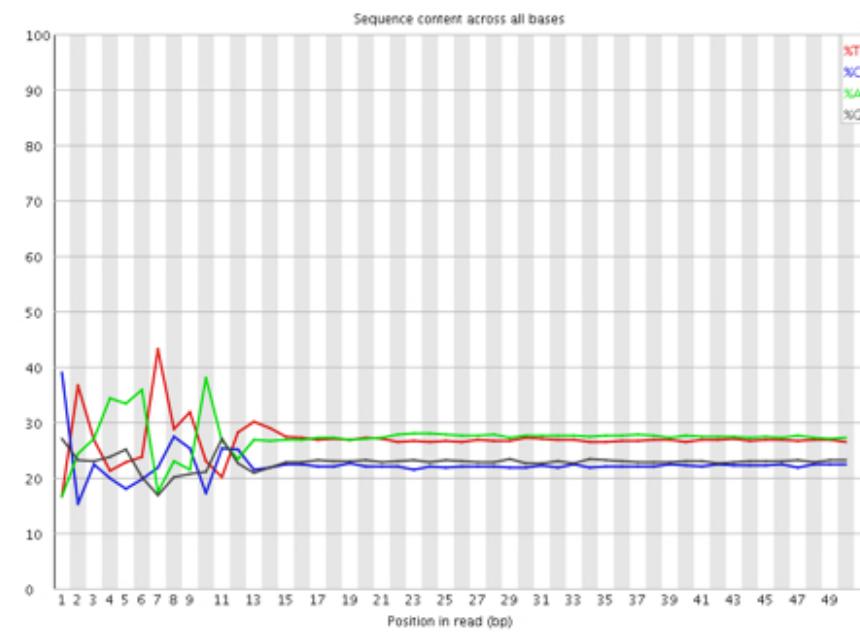
## Basic Statistics

- File type: Conventional base calls
- Encoding: Sanger / Illumina 1.9
- Number of total sequences: 80397337
- Sequence length: 50
- %GC: 45

## Per base sequence quality



## Per base sequence content



# DDBJ Read Annotation Pipeline

## ◆新型シーケンサデータの解析パイプライン

The screenshot shows the 'Select Tools' step of the DDBJ Read Annotation Pipeline. The interface includes a navigation bar at the top with steps: Select Query Files, Select Tools (highlighted in orange), Set QuerySet, Set GenomeSet, Set Map Options, and Confirmation. On the left, there's a sidebar with sections for ACCOUNT (login ID [guest], Logout), ANALYSIS (Data setup, step-1: Preprocessing, Mapping / Assembly; step-2: Workflow, Genome (SNP/Short Indel), RNA-seq (Tag count), ChIP-seq), Job Confirmation (step1: Preprocessing status, step1: Mapping status, step1: Assembly status, step2: All status), Help, HELP (link), MANUAL (link), and Contact Us (link). The main content area is titled 'Selecting Tools for Basic Analysis of DDBJ ANNOTATION PIPELINE'. It contains three tables:

- Reference Genome Mapping:** A table comparing tools based on input data (Base space, Color space, Paired-end), Evaluation (Depth, Coverage, Error rate), Analysis (SNP, Indel), and Output format (.gff, .bed, SAM). Tools listed include BLAT, Maq, bwa, SOAP, Bowtie, and TopHat.
- de novo Assembly:** A table listing tools: SOAPdenovo, ABySS, Velvet, and Trinity. It notes a total limit of 22 Gbp and a warning for Velvet regarding read length and K-mer value.
- Mapping Contigs by de novo Assemble to Reference Sequences:** A table showing a single entry for BLAT with the comment "Single-end analysis only".

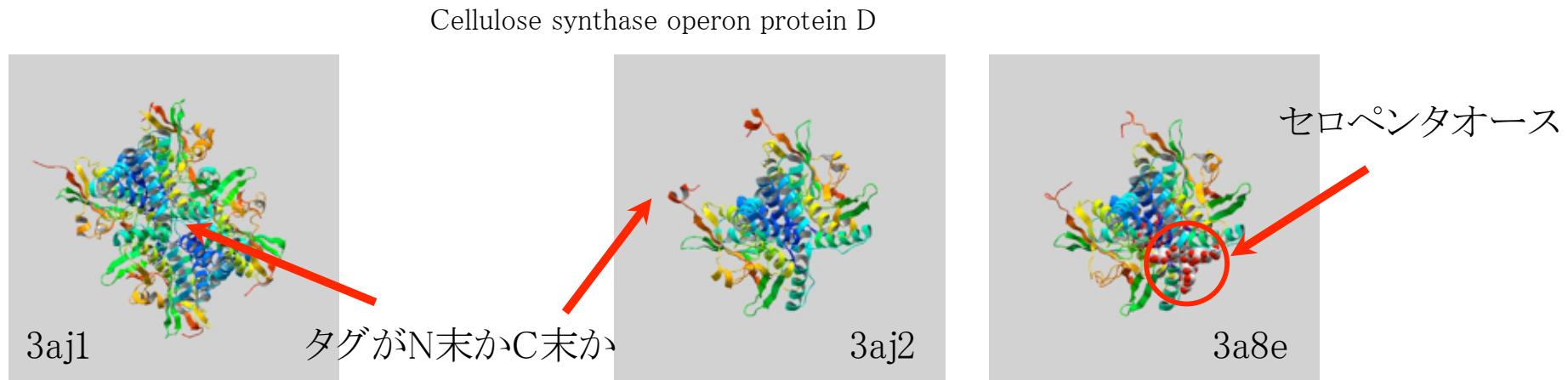
At the bottom right are BACK and NEXT buttons.

# PDBj

Protein Data Bank Japan

# PDBについて

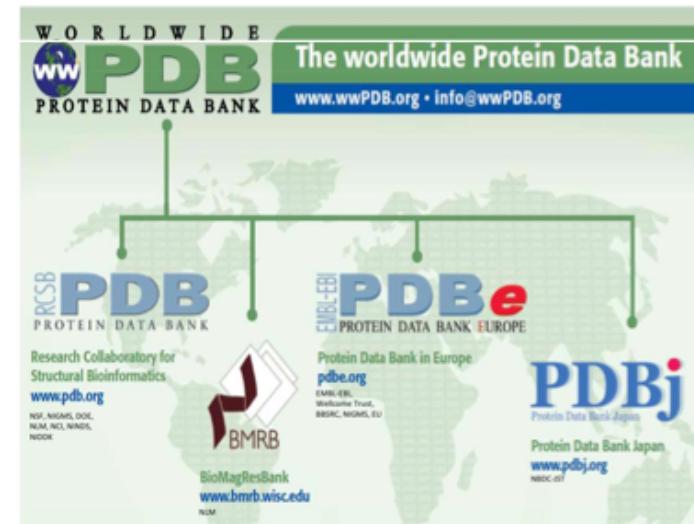
- ◆ 主にタンパク質、核酸の立体構造データを集めたデータバンク
  - 現在のエントリ数は約82,500
  - ひとつのタンパク質でも、リガンドの有無や配列の改変などの違いによって、複数のエントリが登録されていることがある



# 世界蛋白質構造データバンクの一員

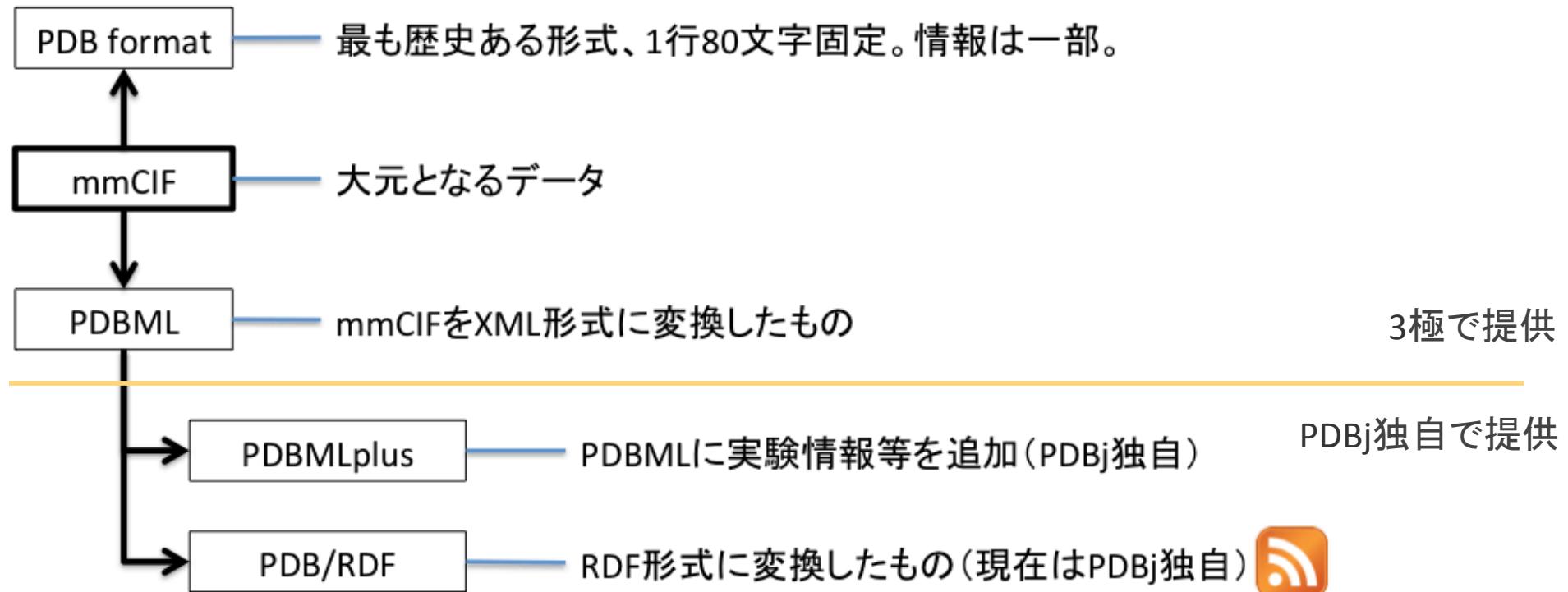
## ◆ Worldwide Protein Data Bank (wwPDB)

- 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

# 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](#)

# PDBファイルの例

HEADER HYDROLASE 04-MAY-99 1CLV  
 TITLE YELLOW MEAL WORM ALPHA-AMYLASE IN COMPLEX WITH THE AMARANTH  
 TITLE 2 ALPHA-AMYLASE INHIBITOR  
 COMPND MOL\_ID: 1;  
 COMPND 2 MOLECULE: PROTEIN (ALPHA-AMYLASE);  
 COMPND 3 CHAIN: A;  
 COMPND 4 SYNONYM: ALPHA-1,4-GLUCAN-4-GLUCANOHYDROLASE;  
 COMPND 5 EC: 3.2.1.1;  
 COMPND 6 MOL\_ID: 2;  
 COMPND 7 MOLECULE: PROTEIN (ALPHA-AMYLASE INHIBITOR);  
 COMPND 8 CHAIN: I;  
 COMPND 9 ENGINEERED: YES  
 SOURCE MOL\_ID: 1;  
 SOURCE 2 ORGANISM\_SCIENTIFIC: TENEBRIOS MOLITOR;  
 SOURCE 3 ORGANISM\_COMMON: YELLOW MEALWORM;  
 SOURCE 4 ORGANISM\_TAXID: 7067;  
 SOURCE 5 MOL\_ID: 2;  
 SOURCE 6 SYNTHETIC: YES;  
 SOURCE 7 OTHER\_DETAILS: THE PROTEIN WAS CHEMICALLY SYNTHESIZED. THE  
 SOURCE 8 SEQUENCE OF THIS PROTEIN IS NATURALLY FOUND IN AMARANTHUS  
 SOURCE 9 HYPOCHONDRIACUS (PRINCE'S FEATHER).  
 KEYWDS INSECT ALPHA-AMYLASE INHIBITOR, AMARANTHUS HYPOCHONDRIACUS,  
 KEYWDS 2 YELLOW MEAL WORM, X-RAY STRUCTURE, KNOTTIN, HYDROLASE  
 EXPDTA X-RAY DIFFRACTION  
 AUTHOR P.J.B.PEREIRA,V.LOZANOV,A.PATTHY,R.HUBER,W.BODE,S.PONGOR,  
 AUTHOR 2 S.STROBL  
 REVDAT 3 24-FEB-09 1CLV 1 VERSN  
 REVDAT 2 01-APR-03 1CLV 1 JRNL  
 REVDAT 1 03-MAY-00 1CLV 0  
 JRNL AUTH P.J.PEREIRA,V.LOZANOV,A.PATTHY,R.HUBER,W.BODE,  
 JRNL AUTH 2 S.PONGOR,S.STROBL  
 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 TITL 4 RESOLUTION.  
 JRNL REF STRUCTURE FOLD.DES. V. 7 1079 1999  
 JRNL REFN ISSN 0969-2126  
 JRNL PMID 10508777  
 JRNL DOI 10.1016/S0969-2126(99)80175-0  
 REMARK 1  
 REMARK 2  
 REMARK 2 RESOLUTION. 2.00 ANGSTROMS.  
 REMARK 3  
 REMARK 3 DEPIEMENT

メタデータ

SHEET 6 A 7 PHE A 218 GLU A 222 1 N PHE A 218 O PHE A 182  
 SHEET 7 A 7 CYS A 241 LEU A 243 1 N CYS A 241 O GLNA 221  
 SHEET 1 B 3 GLN A 392 ARG A 397 0  
 SHEET 2 B 3 GLY A 401 THR A 406 -1 N PHE A 405 O ILE A 393  
 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 425 0  
 SHEET 2 D 2 SER A 441 VAL A 444 -1 N VAL A 444 O GLY A 422  
 SSBOND 1 CYS A 28 CYS A 84 1555 1555 2.03  
 SSBOND 2 CYS A 134 CYS A 148 1555 1555 2.04  
 SSBOND 3 CYS A 354 CYS A 360 1555 1555 2.03  
 SSBOND 4 CYS A 425 CYS A 437 1555 1555 2.03  
 SSBOND 5 CYS I 501 CYS I 518 1555 1555 2.03  
 SSBOND 6 CYS I 508 CYS I 523 1555 1555 2.03  
 SSBOND 7 CYS I 517 CYS I 531 1555 1555 2.03  
 LINK OD1 ASN A 98 CA CA A 601 1555 1555 2.44  
 LINK O ARG A 146 CA CA A 601 1555 1555 2.32  
 LINK OD1 ASP A 155 CA CA A 601 1555 1555 2.74  
 LINK OD2 ASP A 155 CA CA A 601 1555 1555 2.47  
 LINK O HIS A 189 CA CA A 601 1555 1555 2.37  
 LINK CA CA A 601 O HOH A 619 1555 1555 2.54  
 LINK CA CA A 601 O HOH A 628 1555 1555 2.47  
 LINK CA CA A 601 O HOH A 611 1555 1555 2.58  
 LINK C PCA A 1 N LYS A 2 1555 1555 1.33  
 CISPEP 1 VAL A 123 PRO A 124 0 -6.37  
 CISPEP 2 GLU I 519 PRO I 520 0 -5.50  
 SITE 1 ACS 3 ASP 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  
 SITE 1 AC2 3 ARG A 183 LEU A 243 ARG A 321  
 CRYST1 119.250 119.250 64.780 90.00 90.00 120.00 P 61 6  
 ORIGX1 1.000000 0.000000 0.000000 0.000000  
 ORIGX2 0.000000 1.000000 0.000000 0.000000  
 ORIGX3 0.000000 0.000000 1.000000 0.000000  
 SCALE1 0.008386 0.004841 0.000000 0.000000  
 SCALE2 0.000000 0.009683 0.000000 0.000000  
 SCALE3 0.000000 0.000000 0.015437 0.000000  
 HETATM 1 N PCA A 1 29.020 7.713 8.323 1.00 17.69 N  
 HETATM 2 CA PCA A 1 30.380 8.263 8.128 1.00 16.55 C  
 HETATM 3 CB PCA A 1 31.390 7.193 8.612 1.00 16.19 C  
 HETATM 4 CG PCA A 1 30.495 5.943 8.987 1.00 16.93 C  
 HETATM 5 CD PCA A 1 29.101 6.476 8.787 1.00 19.39 C  
 HETATM 6 OE PCA A 1 28.089 5.796 9.037 1.00 22.92 O  
 HETATM 7 C PCA A 1 30.667 8.643 6.676 1.00 13.70 C  
 HETATM 8 O PCA A 1 31.514 9.493 6.417 1.00 14.12 O  
 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

## 実習3

---

- ◆ PDBjの検索サービス”PDBj Mine”で「 $\alpha$ 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

# PDBj トップページ

The screenshot shows the PDBj homepage with a search bar highlighted by a red box. A red arrow points from this box to a detailed search result page below.

**PDBj** (Logo: A 3D molecular model composed of spheres)

English Japanese simplified Chinese traditional Chinese Korean

統計情報 ヘルプ FAQ お問い合わせ

**トップページ**

データ登録 >>

- ADIT: PDB Deposition
- ADIT-NMR

検索 >>

- Search PDB (PDB| Mine)
- PDB/RDF, chem\_comp/RDF
- Latest Release Search
- Sequence-Navigator
- Structure-Navigator
- SeSAW
- Ligand Binding Sites (GIRAF)
- EM Navigator
- Search NMR Data (BMRB)
- Status Search

サービス&ソフトウェア >>

- jV: Graphic Viewer
- 万見 (Yorodumi)
- Protein Globe
- ASH
- MAFFTash
- SELA

Structure Prediction >>

- CRNPRED
- Spanner
- SFAS

二次データベース >>

- eF-site/eF-seek/eF-surf
- eProtS
- ProMode / ProMode Elastic / ProMode Oligomer
- Molecule of the Month

ダウンロード >>

- PDB Archive/Snapshot Archive

リンク集

日本蛋白質構造データバンク (PDBj: 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PDBe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されたPDBアーカイブとして運営するとともに、様々な解析ツールを提供しております。

データ登録

PDB登録 ADIT PDB Deposition

NMRデータ登録 ADIT-NMR

検索

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: αアミラーゼ 検索

詳細条件検索 >>

NMRデータ検索 BMRB

Accession number Deposition code

最新情報

2012/6/5 月の分子の記事数が150件に達しました。今月の記事

2012/5/1 2012年4月30日をもちまして、xPSSSサービスおよび

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回蛋白質科学会年会にて

2012/3/26 jVの最新版（jV4.2）がリリースされました。eF-siteで

ました。詳細はリリースノートをご覧ください。

2012/3/16 PDBで公開済みのエントリー数が、8万件を超ました

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

（詳細...）

検索

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: αアミラーゼ 検索

詳細条件検索 >>

82522 entries available on 20 Jun., 2012  
00:00(UTC) / 09:00(JST)

WORLDWIDE PROTEIN DATA BANK

蛋白質構造百科事典 eProtS Encyclopedia of Protein Structures

Protein Globe

DBCLS Database Center for Life Science

OC Protein Center

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

# PDBj Mine検索結果

Mine PDBj

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)

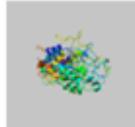
1 - 16/ 319

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:    
変換クエリ : (alpha amylase) | (alfa amylase)

実際に検索に使われた文字列  
(英語に自動変換して検索)

リセット 検索 個別エントリページへ

<a href="#">1clv</a>	分子名称	: 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., Patthy, A., Huber, R., Bode, W., Pongor, S., Strobl, S.
	実験手法	: X-RAY DIFFRACTION
	登録日	: 1999-05-04
<input type="button" value="ダウンロード"/>	公開日	: 2000-05-03
	最終更新日	: 2009-02-24

<a href="#">1bvn</a>	分子名称	: PROTEIN (ALPHA-AMYLASE) (3.2.1.1)
	タイトル	: PIG PANCREATIC ALPHA-AMYLASE IN COMPLEX WITH THE PROTEINACEOUS INHIBITOR TENDAMISTAT
	著者	: Machius, M., Wiegand, G., Epp, O., Huber, R.
	実験手法	: X-RAY DIFFRACTION
	登録日	: 1998-09-16
<input type="button" value="ダウンロード"/>	公開日	: 1998-09-23
	最終更新日	: 2009-02-24

# PDBj エントリページ

Mine PDBj

概要 [1clv]

日本語ページについて  
PDBj Mineについて  
更新情報

概要 構造情報 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  
PDB ID または キーワード 検索

<非対称単位>  
= <生物学的単位>

他の静止画像を見る

3次元構造ビューア  
(JView / Jmol) で見る \*1

配列の表示

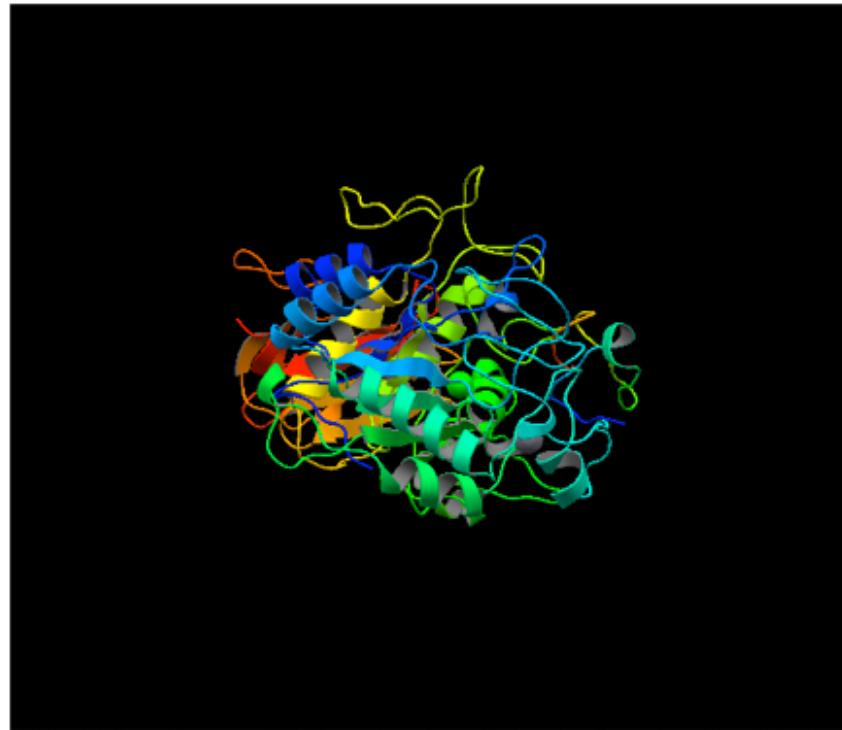
PDBファイルのダウンロード

エントリーID (PDB ID)	1clv	配列情報 (FASTA形式)	PDBファイルのダウンロード
分子名称	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		
由来する生物種	Tenebrio molitor (yellow mealworm)		
由来する組織	[UNP - P80403] Seed		
ポリマー鎖の合計数	2		
分子量の合計	54934.3 (詳細は 構造情報のページ)		
著者	Pereira, P.J.B., Lozanov, V., Patthy, 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., Patthy, 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. <i>Structure Fold Des.</i> , 7:1079 - 1088, 1999. (PubMed : 10508777) (DOI: 10.1016/S0969-2126(99)80175-0)		
実験手法	X-RAY DIFFRACTION (2.00[Å])		
他のデータベース情報	万覧(Yorodumi), CATH, CE, FSSP, SCOP, VAST, UniProt (P56634, P80403), eF-site, KEGG (EC 3.2.1.1), ProTherm, EzCatDB, PISA, PQS, PDB/RDF		
NMR情報	BMRB		

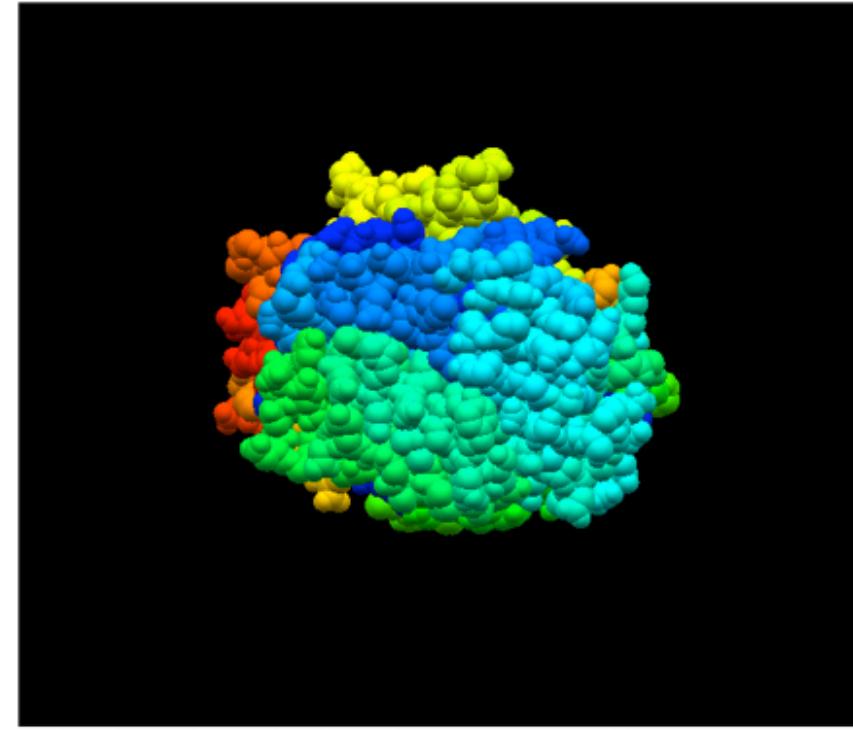
マウスで動かせる画像を表示

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

# jv4での表示



jV version 4 JNLP(JRE1.6以降 / Mac OS X 10.6以降)バージョン  
スタイル: Default Cartoon Wireframe CPK\_without\_water CPK  
色: Default group chain atom



jV version 4 JNLP(JRE1.6以降 / Mac OS X 10.6以降)バージョン  
スタイル: Default Cartoon Wireframe CPK\_without\_water CPK  
色: Default group chain atom

マウスでドラッグすることで、立体構造をあらゆる角度から見ることができる  
ホイールもしくはSHIFTキーを押しながら上下にドラッグすることで拡大縮小  
スタイルや表示する色も変更可能

# 構造情報ページ

**Mine** 構造情報[1clv] 日本語ページについて  
PDBj Mineについて 更新情報

概要 構造情報 実験情報 機動情報 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

タンパク質2分子  
非タンパク質2分子  
水273分子が登録されている

エンティティ

鎖名	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) 配列情報 (FASTA形式)	polymer	UniProt (P56634)	51263.6	1	Tenebrio molitor (yellow mealworm)	• ALPHA-1,4-GLUCAN-4-GLUCANOHYDROLASE
I	PROTEIN (ALPHA-AMYLASE INHIBITOR) 配列情報 (FASTA形式)	polymer	UniProt (P80403)	3595.1	1		
	CALCIUM ION	non-polymer	PDBeChem (CA) Chem.Comp.(CA)	40.1	1		
	CHLORIDE ION	non-polymer	PDBeChem (CL) Chem.Comp.(CL)	35.5	1		
	water	water	PDBeChem (HOH) Chem.Comp.(HOH)	18.0	273		

Display chain - A, [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: EDANFASGRNSIVHLFEWKWNIDIADECERFLQPQGFGGVQISPNEYLVADGRPWWEERYQPVSYIINTRSGDESAFTDMTRRCNDAGVRIYVDAVINHMTGMNGVGTGSSADHDGMNYPAVPYGSGDFHSPCEVNYYQDADNVRNCELVGLRDLNQGS  
Structure:  
Site: 11 161 171 181 191 201 211 221 231 241 251 261 271 281 291 301 311 321  
  
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	2
	分子量の合計	54858.7
非ポリマー	分子数	2
	分子量の合計	75.6
全て	分子量の合計	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

# 実験情報ページ

概要 構造情報 **実験情報** 検索情報 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

精密化の統計情報

格子定数 [Å]	119.250	119.250	64.780
格子定数 [度]	90.00	90.00	120.00
空間群	P 61		
分解能 [Å] (低 - 高)	8.00 - 2.00		
最も高い分解能シェルの値	2.03 - 2.00		
R-work	0.161		
最も高い分解能シェルの値	0.2577		
R-free	0.19		
最も高い分解能シェルの値	0.2435		
結合長の平均二乗偏差(RMSD) [Å]	0.007		
結合角の平均二乗偏差(RMSD) [度]	1.778		

回折データの統計情報

分解能 [Å] (低 - 高)	20.0 - 2.00
最も高い分解能シェルの値	2.05 -
独立反射数	33598
観測反射数	279403
R-merge( <i>I</i> )	0.135
最も高い分解能シェルの値	0.339
完全性 [%]	94.4
最も高い分解能シェルの値	68.3
冗長性	8.3

結晶化条件

結晶ID	方法	pH	pHの範囲	温度	単位
1	Vapor diffusion, hanging drop	5.4		23	°C

文献の結晶化試薬

ID	結晶ID	溶液	試薬名	濃度(単位)	詳細
1	1	drop	sodium acetate	5(mM)	
2	1	drop	CaCl <sub>2</sub>	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による注釈付記

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)



# 機能情報のページ

概要 構造情報 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

PDB ID または キーワード 検索

機能情報のページについて

GO (ジーン・オントロジー) に由来する情報

鎖名	GOid	名前空間	
A	0004556	molecular_function	alpha-amylase activity
I	0004866	molecular_function	endopeptidase inhibitor activity
A	0005975	biological_process	carbohydrate metabolic process
I	0015066	molecular_function	alpha-amylase inhibitor activity
A	0046872	molecular_function	metal ion binding

(GO情報の詳細)

PDBデータベースに由来する情報

site_id	種類	残基数	
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

(PDB情報の詳細) →

SwissProt/UniProtに記載されている蛋白質分子機能情報

site_id	種類	残基数	
SWS_FT_FI1	enzyme active site(ACT_SITE)	1	Proton donor.
SWS_FT_FI2	enzyme active site(ACT_SITE)	1	Nucleophile.
SWS_FT_FI3	metal binding site(METAL)	2	Calcium
SWS_FT_FI4	metal binding site(METAL)	2	Calcium; via carbonyl
SWS_FT_FI5	binding site(BINDING)	3	Chloride.
SWS_FT_FI6	other interesting site(SITE)	1	Transition state stabilizer.

(UniProt情報の詳細)

CSAにおける酵素触媒機能の情報

site_id	種類	残基数	
CSA1	catalytic site	2	PsiBLAST alignment on 1amy by CSA
CSA2	catalytic site	3	PsiBLAST alignment on 2cpu by CSA
CSA3	catalytic site	3	PsiBLAST alignment on 1uok by CSA

(CSA情報の詳細)

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

鎖名>	残基>
A	ASP185
A	GLU222
A	ASP287

site\_id: AC1

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

鎖名>	残基>
A	ASN98
A	ARG146
A	ASP155
A	HIS189
A	HOH611
A	HOH619
A	HOH628

site\_id: AC2

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

鎖名>	残基>
A	ARG183
A	LEU243
A	ARG321

# 相同蛋白質のページ

概要 構造情報 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  
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。

配列が似ているPDBエントリを検索する

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

EKDANFASGRNSIVHLFEWKWNDIADECERFLQPQGFQGVQISPPNEYLVADGRPWWERYQPVSYIINTRSGDESAFTDMTRRCNDAGV  
TRRCNDAGVRIYVDAVINHMTGMNGVGTVTSGSSADHDGMNYPAVPYGSGDFHSPCEVNYYQDADNVRNCELVGLRDLNQGS  
DYVRGVVLIDYMNHMIDLGVAGFRVDAAKHMSPGDLSVIFSGKLKNLNTDYGFADGARPFYQEVIDLGGEAISKNEYTGFG  
CVLEFQFGVSLGNNAFQGGNQLKLNLANWGPEWGLLEGDAVVFDNHDNQRTGGSQILTYKNPKPYKMAIAFMLAHPYGTT  
RIMSSFDFTNDQGPPQDGSGNLISPGINDDNTCSNGYVCEHRWRQVYGMVGFRAVEGTQVENWWNSNDNQIAFSRGSQ  
GFVAFTNGGDLNQNLTGLPAGTYCDVISGELSGSCTGKSVTVDNGSADISLGSAEEDDGVLAIHVNAKL

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

CIPKWNRCGPMDGVPCCEPYTCTSDYYGNCS

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 次へ

1clvA 完全一致: 1jaeA 1tmqA

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

新規検索 [1tmqA] 構造の重ね合わせ

1clvA 1471QKDANFASGRNSIVHLFEWKWNDIADECERFLQPQGFQGVQISPPNEYLVADGRPWWERYQPVSYIINTRSGDESAFTDMTRRCNDAGV

1tmqA 1471QKDANFASGRNSIVHLFEWKWNDIADECERFLQPQGFQGVQISPPNEYLVADGRPWWERYQPVSYIINTRSGDESAFTDMTRRCNDAGV

1tmqA 完全一致: 1jaeA 1clvA

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

新規検索 [1viwiA] 構造の重ね合わせ

1clvA 1471QKDANFASGRNSIVHLFEWKWNDIADECERFLQPQGFQGVQISPPNEYLVADGRPWWERYQPVSYIINTRSGDESAFTDMTRRCNDAGV

1viwiA 1471QKDANFASGRNSIVHLFEWKWNDIADECERFLQPQGFQGVQISPPNEYLVADGRPWWERYQPVSYIINTRSGNESAFDMTRRCNDAGV

Seq.Identity:52% Seq Positives:67% E-value:5.59566e-143 Score:1296 Query Coverage:106% Compound:alpha-amylase

新規検索 [3vm5A] 構造の重ね合わせ

1clvA 1498QKDANFASGRNSIVHLFEWKWNDIADECERFLQPQGFQGVQISPPNEYLVADG--RPWWERYQPVSYIINTRSGDESAFTDMTRRCNDAGV

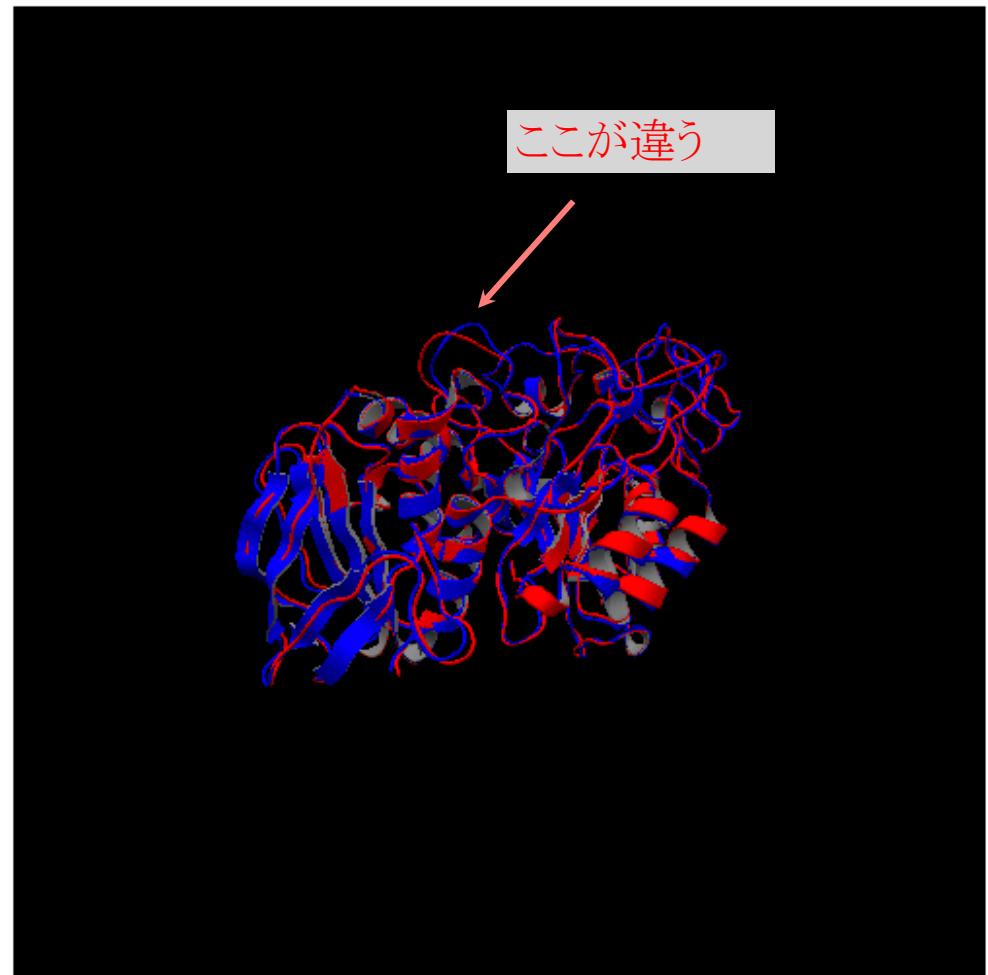
3vm5A 1498QHNPNTDRGRTAIVHLFEWRWADIAECERFLGPKGAGVQISPPNEHILVSSPWRPWQRYQPISYNLCSRSGGENELRDMITRCNNV

立体構造を重ねあわせて  
表示することが可能



# 相同蛋白質のページ

### Superposition of 1clvA (blue) and 1viwA (red)



Style:  Cartoon  Backbone  Wireframe

Color: group chain atom

# 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

日本語ページについて  
PDBj Mineについて  
更新情報

**Mine**

ダウンロード/画面表示 [1clv]

概要 構造情報 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

PDB ID または キーワード 検索

ファイル形式	ファイル名	画面表示	ダウンロード	
PDB形式	全ての情報	pdb1clv.ent.gz(88k) pdb1clv.ent(364k)	圧縮 非圧縮	<a href="#">表示</a> <a href="#">ダウンロード</a> <a href="#">ダウンロード</a>
	ヘッダのみ	pdb1clv.ent.gz(7k)	メタデータのみ(原子座標なし)	<a href="#">表示</a> <a href="#">ダウンロード</a> <a href="#">ダウンロード</a>
mmCIF		1clv.cif.gz(111k)	<a href="#">表示</a> <a href="#">ダウンロード</a>	
XML	全ての情報	1clv.xml.gz(171k)	<a href="#">表示</a> <a href="#">ダウンロード</a>	
	ヘッダのみ (no-atom)	PDBML	1clv-noatom.xml.gz(27k)	<a href="#">表示</a> <a href="#">ダウンロード</a>
		PDBMLplus	1clv-pdbmlplus.xml.gz(30k)	<a href="#">表示</a> <a href="#">ダウンロード</a>
	座標情報のみ (ext-atom)		1clv-extatom.xml.gz(96k)	<a href="#">表示</a> <a href="#">ダウンロード</a>
RDF		1clv-noatom.rdf.gz(18k)	<a href="#">表示</a> <a href="#">ダウンロード</a>	
生物学的単位 (PDB形式)		1clv.pdb1.gz(84k) *author_defined_assembly, 2 分子 (dimeric)	<a href="#">表示</a> <a href="#">ダウンロード</a>	

# 立体構造を取得した後の解析例

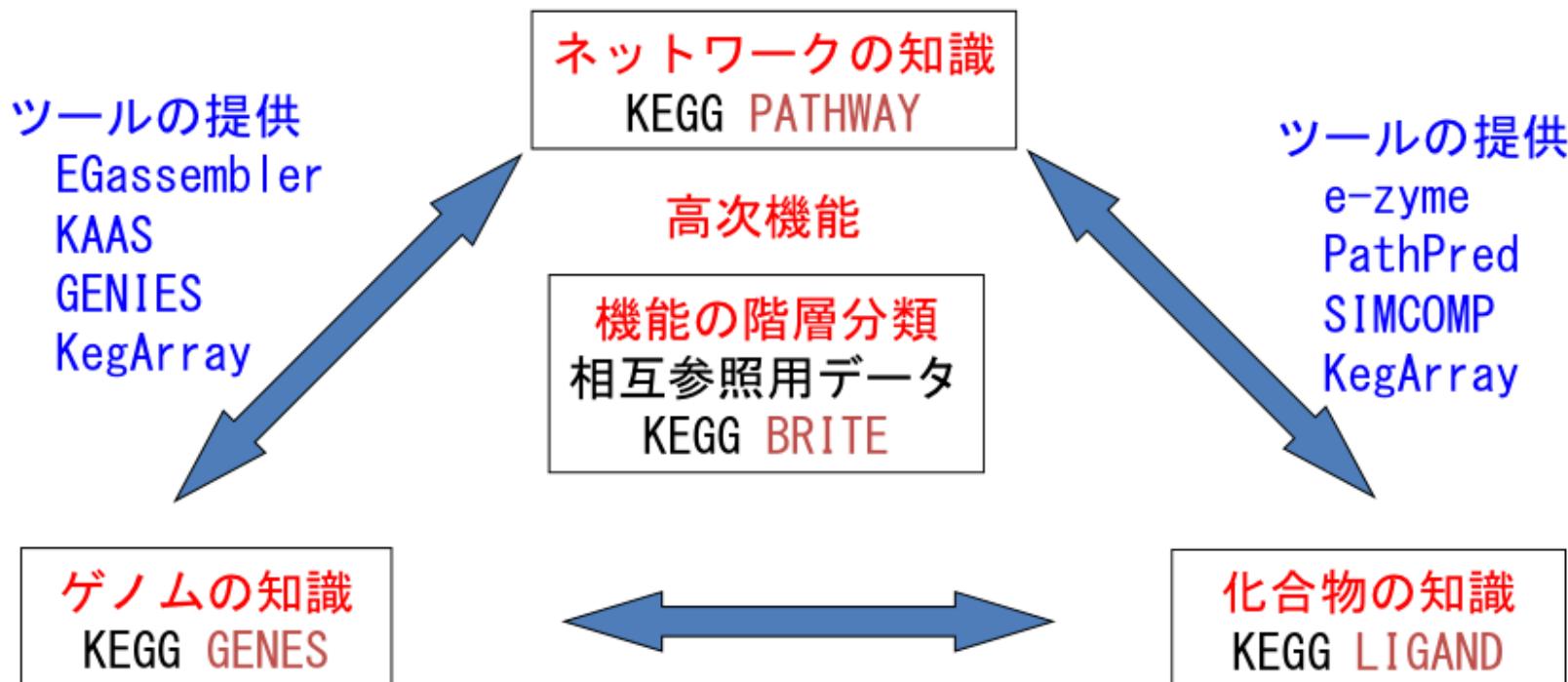
- ◆ “PyMol”, “Chimera”等で立体構造を描画する
- ◆ ”DALI”, “MATRAS”等で立体構造を比較する
- ◆ “SURFNET”等でタンパク質表面のポケット  
(化合物が結合しやすい)を検出する
- ◆ ”DelPhi”等でタンパク質の表面電荷を計算する
- ◆ ”UCSF DOCK”, “AutoDock”等で化合物との  
ドッキングシミュレーションを行う

# KEGG

Kyoto Encyclopedia of Genes and Genomes

# 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

# KEGGトップページ

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

The screenshot shows the KEGG homepage with a sidebar containing links to KEGG Home, Database, Objects, Software, FTP, Genome Network, DBGET/LinkDB, Feedback, and Kanehisa Labs. The main content area features an announcement about a new API version and a section titled "KEGG の主要エントリーポイント" (Main Entry Points) which includes a link to "KEGG2". A red arrow points from the "KEGG2" link to the corresponding entry in the Table of Contents on the right.

**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, the KEGG organism codes will no longer be limited to three letters. [more ...](#)

**KEGG の主要エントリーポイント**

- KEGG2** KEGG 全体の目次のページ [Update notes](#)
- データタイプごとのエントリーポイント**
  - KEGG PATHWAY** パスウェイマップ [パスウェイ一覧]
  - KEGG BRITE** 機能階層・オントロジー [日本語版 Brite 一覧]
  - KEGG MODULE** モジュールマップ
  - KEGG DISEASE** 病気 (日本語) [がん | 感染症]
  - KEGG DRUG** 医薬品 (日本語) [薬効分類 ATC 分類]
  - KEGG ORTHOLOGY** オーソログアノテーション
  - KEGG GENOME** ゲノム [KEGG 生物種一覧]
  - KEGG GENES** 遺伝子・タンパク質
  - KEGG LIGAND** 化合物・化学反応・糖鎖 [化合物分類]
- 一般向けのエントリーポイント**
  - KEGG MEDICUS** ゲノムと疾患・医薬品の統合リソース (日本語) [お薬手帳]
  - 生物種ごとのエントリーポイント**
    - KEGG Organisms** 生物種コード(複数可)  [Go](#) hsa hsa eco
  - 解析ツール**
    - KEGG Mapper** KEGG PATHWAY/BRITE/MODULE へのマッピングツール群
    - KEGG Atlas** KEGG グローバルマップビュア
    - KAAS** ゲノムの自動アノテーションサーバー
    - BLAST/FASTA** 類似配列検索
    - SIMCOMP** 類似化学構造
    - PathPred** 分解・合成経路予測

The screenshot shows the KEGG Table of Contents with a sidebar containing links to KEGG2, PATHWAY, BRITE, MODULE, DISEASE, DRUG, KO, GENOME, GENES, LIGAND, and DBGET. The main content area is a table with columns for Category, Entry Point, Release Info, Search & Compute, and DBGET Search. The table is divided into sections: Systems information, Genomic information, and Chemical information. The "KEGG ORTHOLOGY" section is highlighted with a red border, and the "KEGG GENOME" section is highlighted with a green border. A blue border highlights the "KEGG LIGAND" section under Chemical information.

Category	Entry Point	Release Info	Search & Compute	DBGET Search
Systems information	<b>KEGG PATHWAY</b>	New maps	Search Pathway	PATHWAY
	<b>KEGG BRITE</b>	Update history	Search Brite	
	<b>KEGG MODULE</b>	New hierarchies	Search Module	
	<b>KEGG Mapper</b>	Update history	KEGG pathway maps	
	<b>KEGG Atlas</b>		BRITE functional hierarchies	
Genomic information	<b>KEGG DISEASE</b>	New drug maps	Human diseases	DISEASE
	<b>KEGG DRUG</b>	Update history	Infectious diseases	
	<b>KEGG ENVIRON</b>		ATC drug classification	
	<b>KEGG MEDICUS</b>			
Chemical information	<b>KEGG ORTHOLOGY</b>		KEGG Orthology (KO)	ORTHOLOGY
	<b>KEGG GENOME</b>	New organisms	SSDB search	GENES
		Update history	BLAST / FASTA search	
			KAAS automatic annotation	
<b>KEGG LIGAND</b>	Map organisms to taxonomy	Generate taxonomy tree	MGENOME	
		KEGG organisms		
<b>KEGG COMPOUND</b>	SIMCOMP / SUBCOMP search	COMPOUND		
	KCaM search	GLYCAN		
	E-zyme reaction prediction	REACTION		
	PathPred pathway prediction	RPAIR		
<b>KEGG GLYCAN</b>	PathComp path computation	RCLASS		
		ENZYME		
<b>KEGG REACTION</b>	PathSearch reaction search			

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



## ゲノムの知識データベース・ツール

# 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

## 演習4

---

- ◆ “amylose”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

# 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:ssr3451

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 SSDB computation and KO assignment (gene annotation) by KOALA tool. KEGG DGENES for draft genomes of some eukaryotes and KEGG EGENES for EST datasets of mostly plants are supplementary gene catalogs, which are given automatic KO assignment by KAAS with GENES used as a reference data set. There is now a fourth type of gene catalogs, MGENES for metagenomes (see also [KEGG GENOME](#)) with automatic annotation. The viral gene catalog, VGENES, is not yet fully integrated in the KEGG system.

Gene 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

bfind mode  bget mode

Search Organism hsa for

bfind mode  bget mode

生物種リストの表示

KEGG Organisms: Complete Genomes

Eukaryotes: 155 Bacteria: 1859 Archaea: 129

[ Genomes | Draft | ESTs | Meta | Pan ]

Eukaryotes

Category	Species	Source
Mammals	hsa Homo sapiens (human)	RefSeq
	ptr Pan troglodytes (chimpanzee)	RefSeq
	pon Pongo abelii (Sumatran orangutan)	RefSeq
	mcc Macaca mulatta (rhesus monkey)	RefSeq
	mmu Mus musculus (mouse)	RefSeq
	rno Rattus norvegicus (rat)	RefSeq
	cfa Canis familiaris (dog)	RefSeq
	aml Ailuropoda melanoleuca (giant panda)	RefSeq
	bta Bos taurus (cow)	RefSeq
	ssc Sus scrofa (pig)	RefSeq
Vertebrates	ecb Equus caballus (horse)	RefSeq
	mdo Monodelphis domestica (opossum)	RefSeq
	oaa Ornithorhynchus anatinus (platypus)	RefSeq
	gga Gallus gallus (chicken)	RefSeq
	mfp Meleagris gallopavo (turkey)	RefSeq
	tgu Taenopygia guttata (zebra finch)	RefSeq
	acs Anolis carolinensis (green anole)	RefSeq
	xla Xenopus laevis (African clawed frog)	RefSeq
	xtr Xenopus tropicalis (western clawed frog)	RefSeq
	dre Danio rerio (zebrafish)	RefSeq
Birds	bfo Branchiostoma floridae (Florida lancelet)	RefSeq
	cin Ciona intestinalis (sea squirt)	RefSeq
	spu Strongylocentrotus purpuratus (purple sea urchin)	RefSeq
	dme Drosophila melanogaster (fruit fly)	RefSeq
Reptiles		
Amphibians		
Fishes		
Lancelets		
Ascidians		
Echinoderms		

2種類の検索モード  
 bfind: キーワード検索  
 bget: KEGG GENES IDがわかっているとき



# KEGG GENES 検索結果

**KEGG** Search GENES for amylase Go Clear

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

- hsa:93432 MGC138178, MGC138180; maltase-glucoamylase (alpha-glucosidase) pseudogene
- hsa:100124692 E1\_116351; maltase-glucoamylase (alpha-glucosidase) pseudogene
- hsa:277** AMY1B, AMY1, AMY1A, AMY1C, MGC177995; amylase, alpha 1B (salivary) (EC:3.2.1.1); K01176 alpha-amylase [EC:3.2.1.1]
- hsa:278 AMY1C, AMY1, AMY1A, AMY1B; amylase, alpha 1C (salivary) (EC:3.2.1.1); K01176 alpha-amylase [EC:3.2.1.1]
- hsa:91978 TPGS1, C19orf20, GTRGEO22, PGs1; tubulin polyglutamylase complex subunit 1
- hsa:25941 TPGS2, C18orf10, DKFZp586M1523, HsT3006, L17, PGs2; tubulin polyglutamylase complex subunit 2
- hsa:276 AMY1A, AMY1, AMY1B, AMY1C; amylase, alpha 1A (salivary) (EC:3.2.1.1); K01176 alpha-amylase [EC:3.2.1.1]
- hsa:280 AMY2B, AMY2; amylase, alpha 2B (pancreatic) (EC:3.2.1.1); K01176 alpha-amylase [EC:3.2.1.1]
- hsa:279 AMY2A, AMY2, AMY2B, PA; amylase, alpha 2A (pancreatic) (EC:3.2.1.1); K01176 alpha-amylase [EC:3.2.1.1]
- hsa:8972 MGAM, MG, MGA; maltase-glucoamylase (alpha-glucosidase) (EC:3.2.1.20 3.2.1.3); K12047 maltase-glucoamylase [EC:3.2.1.20 3.2.1.3]
- hsa:790 CAD; carbamoyl-phosphate synthetase 2, aspartate transcarbamylase, and dihydroorotase (EC:2.1.3.2 3.5.2.3 6.3.5.5); K11540 carbamoyl-phosphate synthase / aspartate carbamoyltransferase / dihydroorotase [EC:6.3.5.5 2.1.3.2 3.5.2.3]
- ptr:470338 CAD; carbamoyl-phosphate synthetase 2, aspartate transcarbamylase, and dihydroorotase; K11540 carbamoyl-phosphate synthase / aspartate carbamoyltransferase / dihydroorotase [EC:6.3.5.5 2.1.3.2 3.5.2.3]
- ptr:735587 tubulin polyglutamylase TTLL11-like
- ptr:455380 tubulin polyglutamylase complex subunit 2-like
- ptr:749850 tubulin polyglutamylase complex subunit 1-like
- ptr:463789 MGAM; maltase-glucoamylase, intestinal-like; K12047 maltase-glucoamylase [EC:3.2.1.20 3.2.1.3]

**KEGG Homo sapiens (human): 277**

Help

Entry	277	CDS	T01001
Gene name	AMY1B, AMY1, AMY1A, AMY1C, MGC177995		
Definition	amylase, alpha 1B (salivary) (EC:3.2.1.1)		
Orthology	K01176 alpha-amylase [EC:3.2.1.1]		
Organism	hsa Homo sapiens (human)		
Pathway	hsa00500 Starch and sucrose metabolism hsa01100 Metabolic pathways hsa04970 Salivary secretion hsa04973 Carbohydrate digestion and absorption		
Class	Metabolism; Carbohydrate Metabolism; Starch and sucrose metabolism [PATH:hsa00500] Organismal Systems; Digestive System; Salivary secretion [PATH:hsa04970] Organismal Systems; Digestive System; Carbohydrate digestion and absorption [PATH:hsa04973] <a href="#">BRITe hierarchy</a>		
SSDB	<a href="#">Ortholog</a> <a href="#">Paralog</a> <a href="#">GIFT</a>		
Motif	Pfam: Alpha-amylase Alpha-amylase_C DUF3078 <a href="#">Motif</a>		
Other DBs	NCBI-GI: 56549662 NCBI-GenelD: 277 OMIM: 104701 HGNC: 475 HPRD: 18510 Ensembl: ENSG00000174876 Vega: OTTHUMG00000011021 UniProt: P04745		
Structure	PDB: 3DHP 3BLP 1Z32 1SM0 1JXK 1JXJ 3BLK 1MFU 1MFV 1Q4N 1NM9 1C8Q 1XV8 <a href="#">Thumbnails</a>	 <a href="#">Jmol</a>	
Position	1p21		
AA seq	511 aa <a href="#">AA seq</a> <a href="#">DB search</a> MKLFWLLFTIGFCWAQYSSNTQQGRTSIVHLFEWRWVDIALECEERYLAPKGFGGVQVSPP NENVAIHNPFRPWWEERYQPVSYKLCTRSGNEDFRNMVTRCNNVGIRYVDAVINHMCGN AVSAGTSTTCGSYFPNGPSRDFPAVPYSQWDFNDGKCKTGSGDNEYNDATQVRDCRLSGL LDLALGKDYYRSKIAEYMNHLIDIGVAGFRIDASKHMWPDKIALDKLHNLNLSNWPEG SKPFIYQEVIDLGGEPIKSSDYFGNGRVTEFKYGAALKGTVKRWNGEKMSYLKNWGEGWG FMPSDRALVFVDNHDNQRGHAGGASILTWFWDARLYKMAVGFMALAHPYGFTRVMSYYRWP RYFENGKDNDVWGPNDNGVTKETVNPDPTTCGNDWVCEHRWRQIRNMVNFRNVVDGQP FTNWYDNGSNQVAFGRGNRGFIVFNNDDWTSLTQTLPGTYCDVISGDKINGNTGI KIYVSDDGKAHFSSISNAEDPFIAHESKL		
NT seq	1536 nt <a href="#">NT seq</a> atgaagcttttgttgccttaccatgggttcgtggctcagtattcccaaat acacaacaaggacgaaatctattgtcatctgttgaatggcgatgggttatgtct cttgaatgtgagcgatatttagctcccaaggattgggggttcaggctctccacca		



# KEGG ORTHOLOGYのエントリ



ORTHOLOGY: K01176

Help

Entry	K01176 KO
Name	E3.2.1.1, amyA, malS
Definition	alpha-amylase [EC:3.2.1.1]
Pathway	<a href="#">ko00500</a> Starch and sucrose metabolism <a href="#">ko04973</a> Carbohydrate digestion and absorption
Class	Metabolism; Carbohydrate Metabolism; Starch and sucrose metabolism [PATH: <a href="#">ko00500</a> ] Organismal Systems; Digestive System; Carbohydrate digestion and absorption [PATH: <a href="#">ko04973</a> ] <a href="#">BRITE hierarchy</a>
Other DBs	RN: <a href="#">R02108</a> COG: <a href="#">COG0366</a> GO: <a href="#">0004556</a>
Genes	HSA: <a href="#">276</a> (AMY1A) <a href="#">277</a> (AMY1B) <a href="#">278</a> (AMY1C) <a href="#">279</a> (AMY2A) <a href="#">280</a> (AMY2B) PTR: <a href="#">457067</a> (AMY2B) <a href="#">469397</a> (AMY1A) <a href="#">736558</a> (AMY2A) PON: <a href="#">100446394</a> (AMY1A) <a href="#">100447008</a> MCC: <a href="#">712893</a> (AMY2B) <a href="#">713027</a> (AMY2A) MMU: <a href="#">100043684</a> (Amy2a4) <a href="#">11722</a> (Amy1) RNO: <a href="#">24203</a> (Amy1a) <a href="#">497039</a> (Amy2) CFA: <a href="#">480825</a> AML: <a href="#">100465041</a> BTA: <a href="#">505049</a> (AMY2B) <a href="#">539383</a> (AMY2A) SSC: <a href="#">100152255</a> <a href="#">397397</a> (AMY2) ECB: <a href="#">100049851</a> <a href="#">100051073</a> MDO: <a href="#">100032988</a> (AMY1A) OAA: <a href="#">100078561</a> (AMY2A) GGA: <a href="#">414139</a> (AMY1A) <a href="#">414140</a> (AMY2A)

# 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 when 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 \(SBH method\)](#)

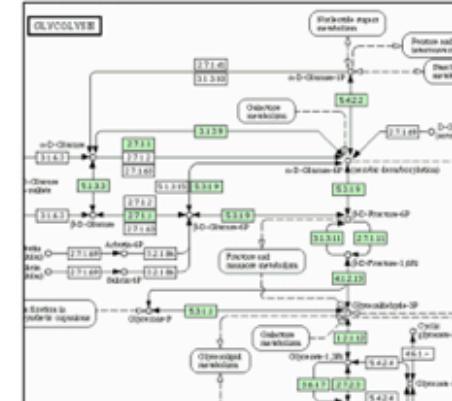
### Example of Results

#### KO assignment

The screenshot shows a table titled "KO Assignment Results" with the following data:

Query gene	KO assignment
query_0001	K00003
query_0002	K00073
query_0003	K01733
query_0005	
query_0006	
query_0007	K03310
query_0008	K05916
query_0009	K03913
query_0010	K07234
query_0011	
query_0012	
query_0013	
query_0014	K04043
query_0015	K03905
query_0016	

#### KEGG pathway mapping



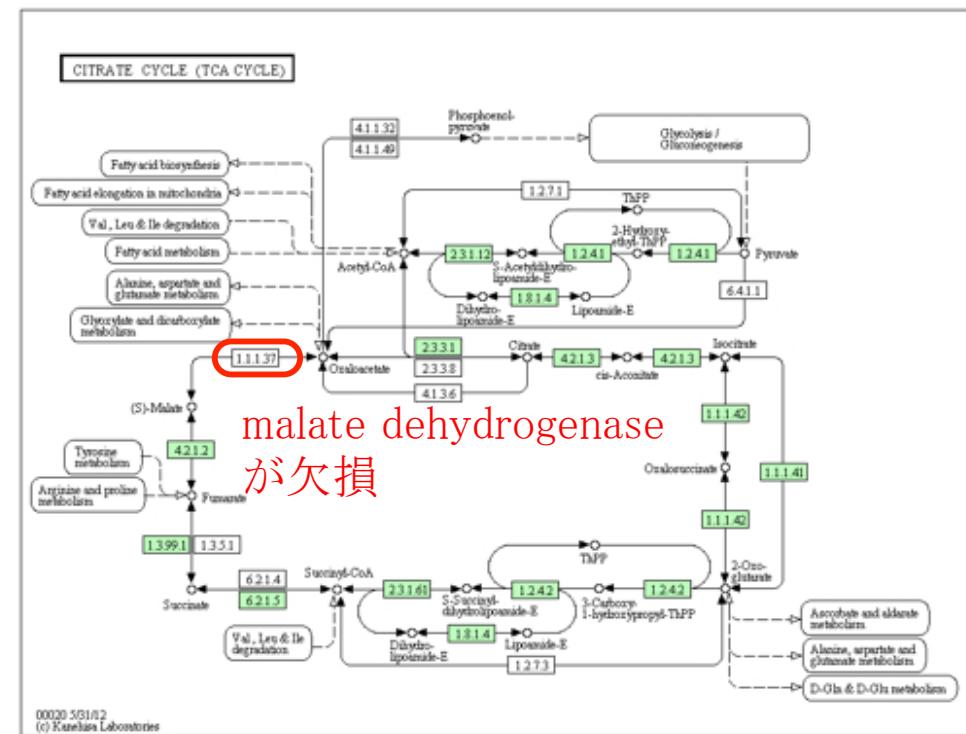
## KAAS入力と結果

マルチFASTAファイルを入力

```
>gi|296114119|ref|ZP_06832775.1| hypothetical protein GXY_00005  
[Gluconacetobacter hansenii ATCC  
23769]MRWGLVMIQACMIVVLGKEIGHLSGRDAAPAFGVW>gi|296114121|ref|ZP_06832776.1|  
chaperon clpb [Gluconacetobacter hansenii ATCC  
23769]MNIEKFTERSRGFLQAAQTIAMREYNQQLTPEHLLKALLDDQGAASALIRAAGGQPPAIAAVDTALAKLP  
KVQGGGAGQPSATPDILVRLLDAAEQAAQKAGDEYVAQDRLLAAIAASETPAGQALRAGGATPQALDKAIATIRKGRTV  
TSENAEASFDAKKYARDVTEIALQGKLDPVIGRDEEIRRQVLARRSKN>gi|296114122|ref|  
ZP_06832777.1| hypothetical protein GXY_00019 [Gluconacetobacter hansenii ATCC  
23769]MGAALIMITTLIAGYALCEWPPGEEVTPALMOIK
```

## TCAサイクル

KO Assignment Results	
<a href="#">Home</a>	
<a href="#">[KO list]</a>	<a href="#">[BRITE hierarchies]</a>
<a href="#">[Pathway map]</a>	<a href="#">[Threshold change]</a>
<a href="#">[Download]</a>	
Query gene : KO assignment	
query	
gi 296114119 ref ZP_06832775.1	
gi 296114121 ref ZP_06832776.1	K03695
gi 296114122 ref ZP_06832777.1	
gi 296114123 ref ZP_06832778.1	
gi 296114124 ref ZP_06832779.1	
gi 296114125 ref ZP_06832780.1	K00053
gi 296114126 ref ZP_06832781.1	K01653
gi 296114127 ref ZP_06832782.1	
gi 296114128 ref ZP_06832783.1	K01652
gi 296114129 ref ZP_06832784.1	K00791



# 化合物の知識データベース・ツール

# 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

# 演習5

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

The screenshot shows the KEGG LIGAND Database homepage. At the top is a logo with the letters 'K' and 'L' in yellow on a blue background, with the text 'KEGG LIGAND Database' and 'Molecular building blocks of life in the chemical space'. Below the logo is a navigation bar with links: KEGG2, PATHWAY, BRITE, MODULE, LIGAND, COMPOUND, GLYCAN, and REACTION. A search input field says 'Enter C/G/R/RP/RC numbers (Example) C00389 C05903 C06562 C08650 C09727 C09762'. Below the search field are buttons for Filter, Pathway mapping, Brite mapping, Get title, Get entry, and Clear. A section titled 'Chemical Substances and Reactions' contains a paragraph about the database's scope and a table mapping identifiers to databases. At the bottom is a search bar with 'Search LIGAND for phenylalanine' and 'Go' and 'Clear' buttons, along with 'bfind mode' and 'bget mode' radio buttons.

Database	Identifier	Content	Specialized entry point	
LIGAND	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	
	RCLASS	RC number	Reaction class	
	ENZYME	EC number	Enzyme nomenclature	



# 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  
C02265 KegDraw Jmol  
D-Phenylalanine; D-alpha-Amino-beta-phenylpropionic acid  
C03519 KegDraw Jmol  
N-Acetyl-L-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 Dopaquinone + 2 H<sub>2</sub>O  
R00686 L-Phenylalanine racemase (ATP-hydrolysing); ATP + L-Phenylalanine + H<sub>2</sub>O <=> AMP + Diphosphate + D-Phenylalanine  
R00687 ATP:L-phenylalanine adenylyltransferase; ATP + L-Phenylalanine <=> Diphosphate + N-Adenylyl-L-phenylalanine  
R00688 L-phenylalanine:NAD<sup>+</sup> oxidoreductase (deaminating); L-Phenylalanine + H<sub>2</sub>O + NAD<sup>+</sup> <=> Phenylpyruvate + NH<sub>3</sub> + NADH + H<sup>+</sup>  
... » display all

## KEGG ENZYME

- 1.4.1.20 phenylalanine dehydrogenase; L-phenylalanine dehydrogenase; PHD  
1.13.12.9 phenylalanine 2-monoxygenase; L-phenylalanine oxidase (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  
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  
C02265 KegDraw Jmol  
D-Phenylalanine; D-alpha-Amino-beta-phenylpropionic acid  
C03519 KegDraw Jmol  
N-Acetyl-L-phenylalanine  
C03709 KegDraw Jmol  
N-Adenylyl-L-phenylalanine  
C05620 KegDraw Jmol  
N-Acetyl-D-phenylalanine  
C08273 KegDraw Jmol  
2,5-Dihydrophenylalanine; L-2,5-Dihydrophenylalanine  
C11217 KegDraw Jmol  
p-Fluorophenylalanine  
C11221 KegDraw Jmol  
Formylmethionyl-leucyl-phenylalanine methyl ester; FMLP OMe  
C11341 KegDraw Jmol  
N-Acetylphenylalanine beta-naphthyl ester; N-Acetyl-DL-phenylalanine beta-naphthyl ester  
C11596 KegDraw Jmol  
FMLP; N-Formyl-L-methionyl-L-leucyl-L-phenylalanine  
C11615 KegDraw Jmol  
Pristinamycin IA; Vermamycin B alpha; Mikamycin B; 4-[4-(Dimethylamino)-N-methyl-L-phenylalanine]virginiamycin S1; Streptogramin B  
C11616 KegDraw Jmol  
Pristinamycin IB; Vermamycin B beta; Efepristin; 4-[N-Methyl-4-(methylamino)-L-phenylalanine]virginiamycin S1  
C11617 KegDraw Jmol  
Pristinamycin IC; Vermamycin B gamma; 2-D-Alanine-4-[4-(dimethylamino)-N-methyl-L-phenylalanine]virginiamycin S1  
C12033 KegDraw Jmol  
4-Aminophenylalanine; p-Aminophenylalanine  
C17235 KegDraw Jmol  
L-Homophenylalanine  
C18619 KegDraw Jmol  
4-Dimethylamino-L-phenylalanine  
C19712 KegDraw Jmol  
N-Hydroxy-L-phenylalanine; 2-(Hydroxyamino)-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

# KEGG COMPOUNDのエントリ

**Kegg** COMPOUND: C00079 [Help](#)

Entry	C00079	Compound
Name	L-Phenylalanine; (S)-alpha-Amino-beta-phenylpropionic acid	
Formula	C9H11NO2	
Exact mass	165.079	
Mol weight	165.1891	
Structure	 C00079 <a href="#">Mol file</a> <a href="#">KCF file</a> <a href="#">DB search</a> <a href="#">Jmol</a> <a href="#">KegDraw</a>	
Remark	Same as: D00021 <a href="#">BRITE hierarchy</a>	
Reaction	<a href="#">R00686</a> <a href="#">R00687</a> <a href="#">R00688</a> <a href="#">R00689</a> <a href="#">R00690</a> <a href="#">R00691</a> <a href="#">R00692</a> <a href="#">R00693</a> <a href="#">R00694</a> <a href="#">R00695</a> <a href="#">R00697</a> <a href="#">R00698</a> <a href="#">R00699</a> <a href="#">R01375</a> <a href="#">R01376</a> <a href="#">R01795</a> <a href="#">R03660</a> <a href="#">R06744</a> <a href="#">R07211</a> <a href="#">R08435</a> <a href="#">R08463</a> <a href="#">R08652</a> <a href="#">R08690</a> <a href="#">R09535</a> <a href="#">R09578</a> <a href="#">R09579</a>	
Pathway	<a href="#">ko00360</a> Phenylalanine metabolism <a href="#">ko00400</a> Phenylalanine, tyrosine and tryptophan biosynthesis <a href="#">ko00940</a> Phenylpropanoid biosynthesis <a href="#">ko00960</a> Tropane, piperidine and pyridine alkaloid biosynthesis <a href="#">ko00966</a> Glucosinolate biosynthesis <a href="#">ko00970</a> Aminoacyl-tRNA biosynthesis <a href="#">map01060</a> Biosynthesis of plant secondary metabolites <a href="#">map01061</a> Biosynthesis of phenylpropanoids <a href="#">map01063</a> Biosynthesis of alkaloids derived from shikimate pathway <a href="#">map01064</a> Biosynthesis of alkaloids derived from ornithine, lysine and nicotinic acid <a href="#">map01070</a> Biosynthesis of plant hormones <a href="#">ko01100</a> Metabolic pathways <a href="#">ko01110</a> Biosynthesis of secondary metabolites <a href="#">ko02010</a> ABC transporters <a href="#">ko4974</a> Protein digestion and absorption <a href="#">ko4978</a> Mineral absorption	

**Enzyme**

1.4.1.20	1.4.3.2	1.11.1.7	1.11.1.21
1.13.12.9	1.14.13.124	1.14.16.1	2.3.1.53
2.6.1.1	2.6.1.5	2.6.1.9	2.6.1.28
2.6.1.57	2.6.1.58	2.6.1.64	2.6.1.70
2.7.7.54	3.4.17.23	4.1.1.28	4.1.1.53
4.2.1.51	4.2.1.91	4.3.1.24	4.3.1.25
5.1.1.11	6.1.1.20		

**Other DBs**

CAS: 63-91-2  
 PubChem: [3379](#)  
 ChEBI: [17295](#)  
 KNApSAck: [C00001386](#)  
 PDB-CCD: [PHE](#)  
 3DMET: [B01151](#)  
 NIKKAJi: [J9.175H](#)

**KCF data**

[Show](#)

# KEGG REACTIONのエントリ



REACTION: R00692

Help

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

## システムの知識データベース・ツール

---

# 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

# 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

The screenshot shows the KEGG Global Map interface. At the top left, there's a section titled "0. Global Map" with a sub-section "0.1 Metabolism". Below it is a list of metabolic pathways: "Metabolic pathways [zoom out]" (highlighted with a red box), "Biosynthesis or secondary metabolites [zoom out]", and "Microbial metabolism in diverse environments [zoom out]". To the right of these are three "Launch KEGG Atlas" buttons. The main content area is titled "1. Metabolism" and contains a sub-section "1.1 Carbohydrate Metabolism". This section is also highlighted with a red box and lists various carbohydrate metabolism pathways: Glycolysis / Gluconeogenesis, Citrate cycle (TCA cycle), Pentose phosphate pathway, Pentose and glucuronate interconversions, Fructose and mannose 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, and Inositol phosphate metabolism. To the right of this list are two categories: "Enzymes" and "Compounds with biological roles". Below "1.1 Carbohydrate Metabolism" is another sub-section "1.2 Energy Metabolism" which lists Oxidative phosphorylation, Photosynthesis, and Photosynthesis - antenna proteins. To the right of "1.2 Energy Metabolism" is a category "Photosynthesis proteins".

**0. Global Map**

**0.1 Metabolism**

Metabolic pathways [zoom out]  
Biosynthesis or 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  
Fructose and mannose 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

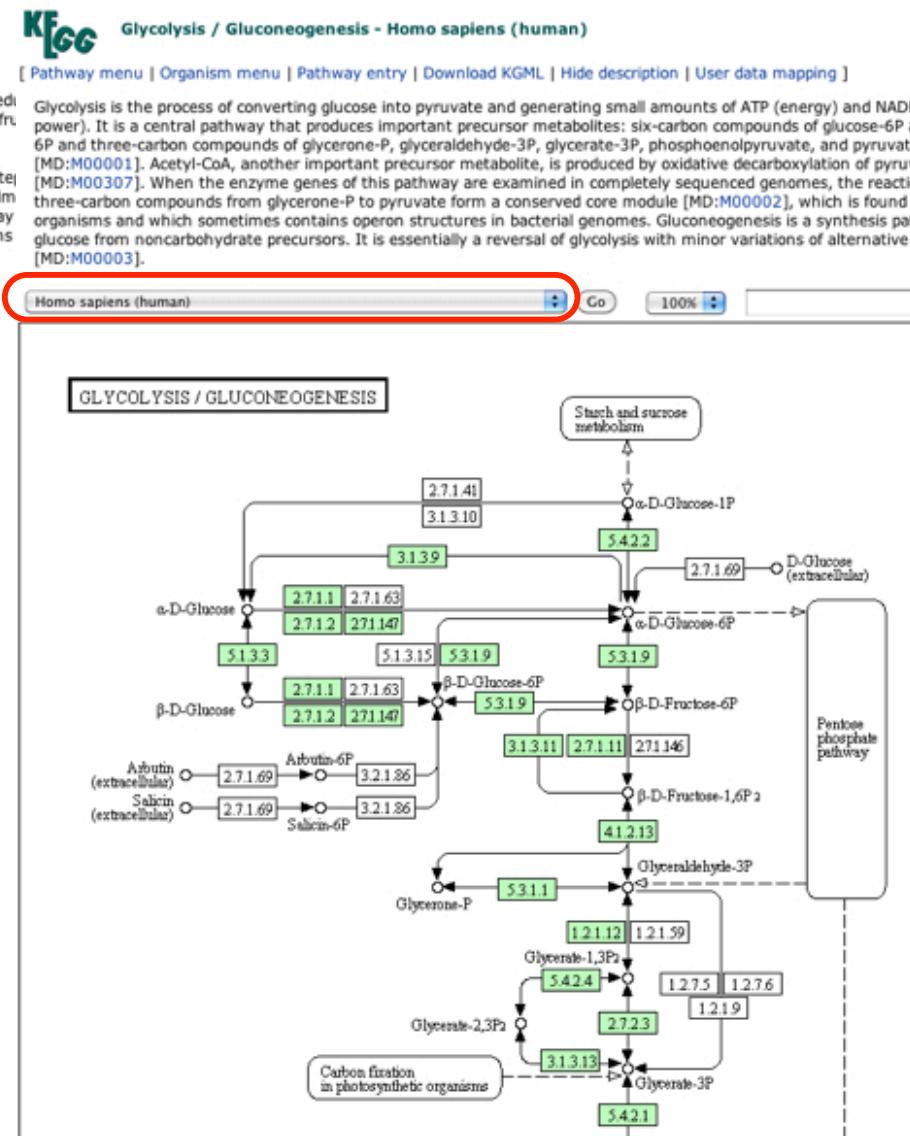
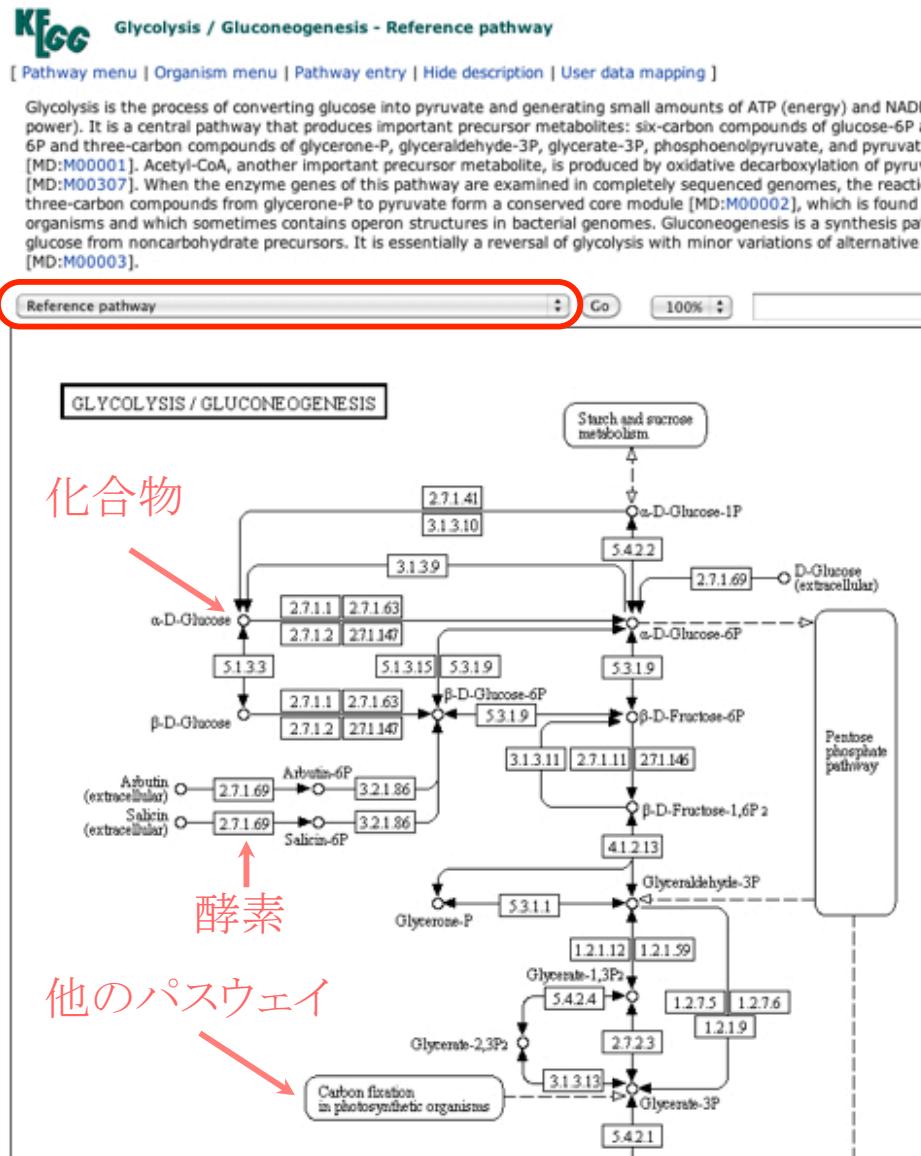
Enzymes  
Compounds with biological roles

**1.2 Energy Metabolism**

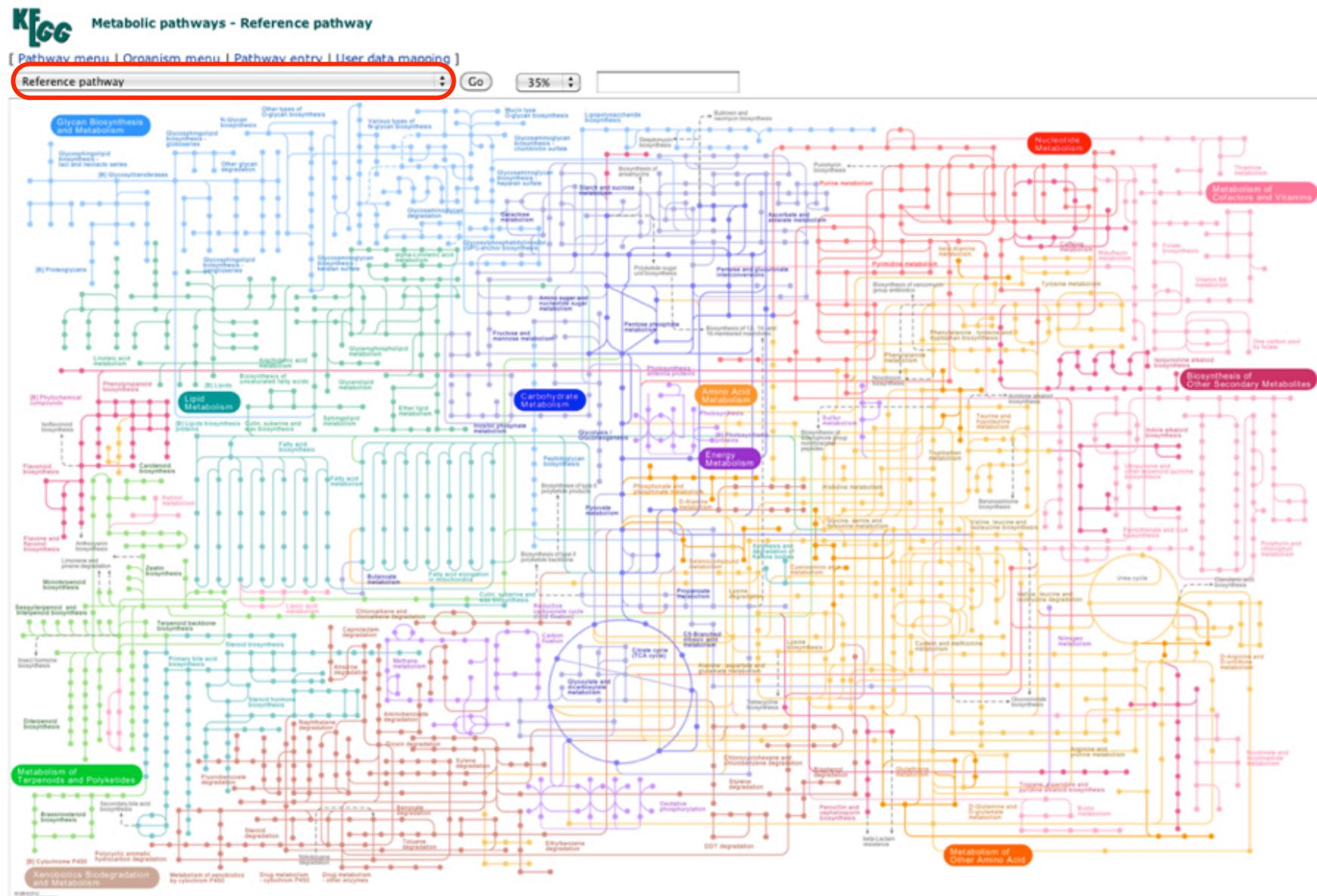
Oxidative phosphorylation  
Photosynthesis  
Photosynthesis - antenna proteins

Photosynthesis proteins

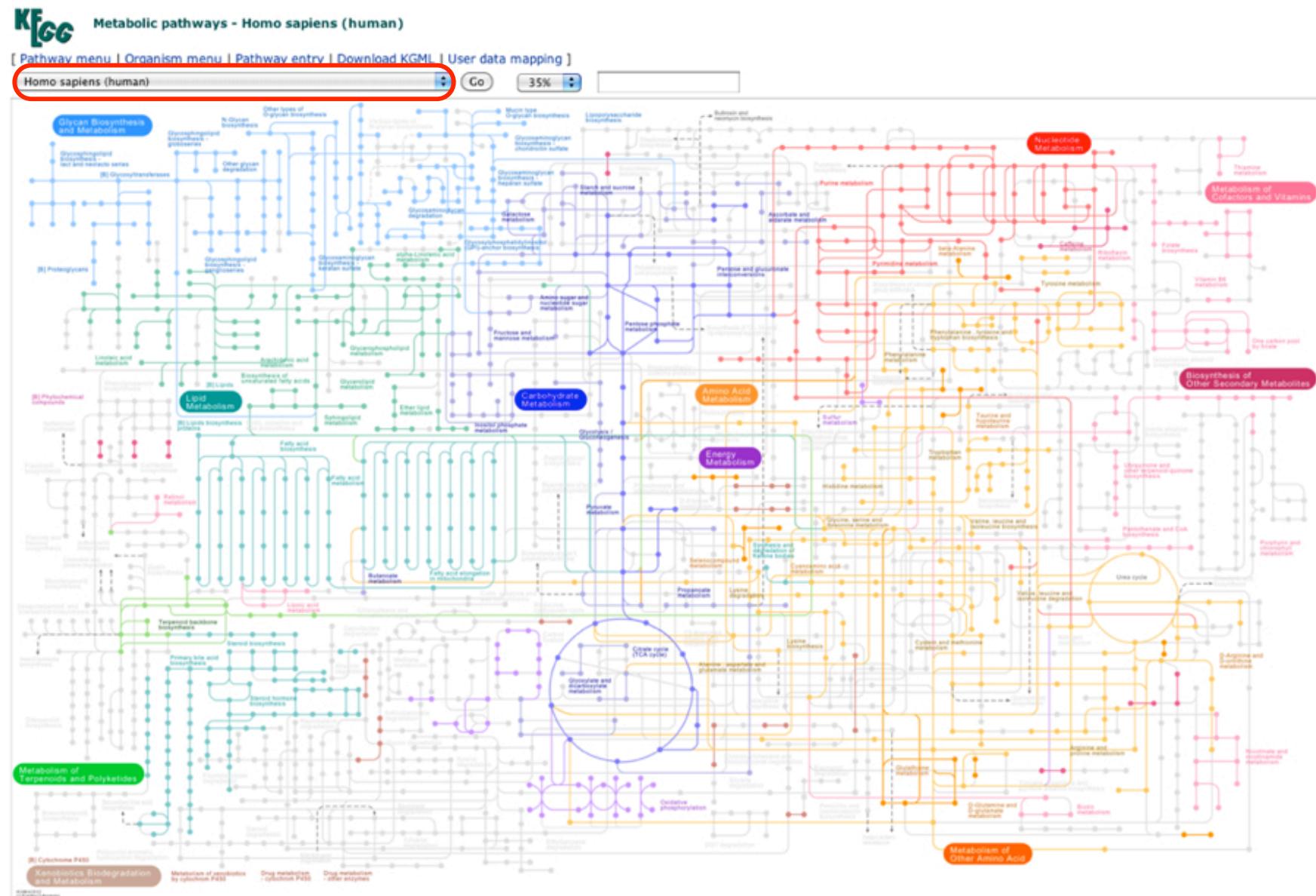
# KEGG PATHWAY MAP



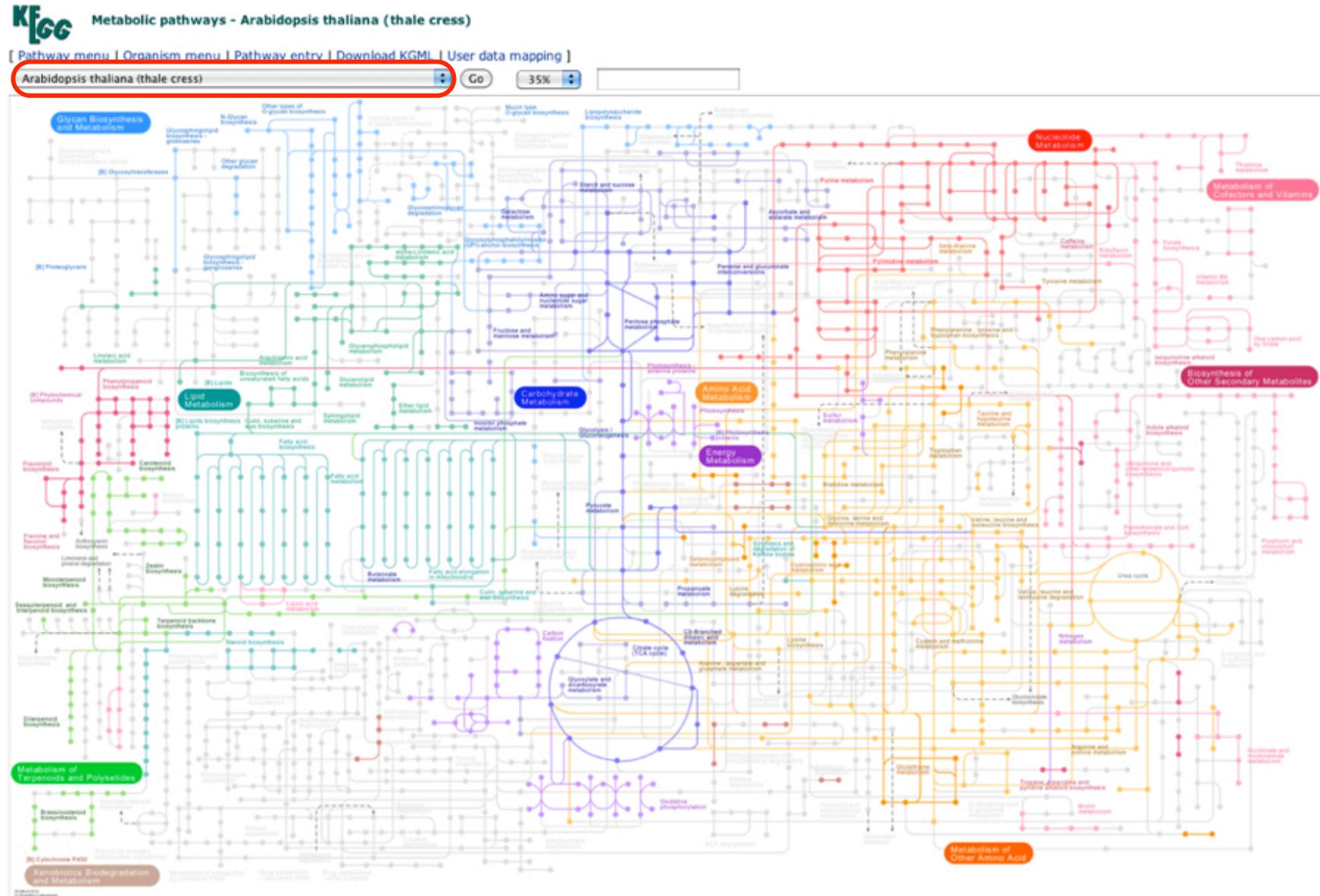
# Global Map



# Global Map (*Homo sapiens*)



# Global Map (*Arabidopsis thaliana*)



# 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

# 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	<a href="#">Search Pathway</a>	<a href="#">Search&amp;Color Pathway</a>	<a href="#">Color Pathway</a>	<a href="#">Reconstruct Pathway</a>
KEGG Atlas	<a href="#">Search Brite</a>	<a href="#">Search&amp;Color Brite</a>	<a href="#">Color Pathway 3D</a>	<a href="#">Reconstruct Brite</a>
KEGG	<a href="#">Search Module</a>	<a href="#">Search&amp;Color Module</a>	<a href="#">Join Brite</a>	<a href="#">Reconstruct Module</a>

### 2. 表示パスウェイを指定

Select KEGG pathway map:

Color specification example: Cancer stage

hsa05200 + CML-stage

hsa05220 + CML-stage

Enter file name containing the data:

hsa\_CML-COSMIC.txt

Numerical value example: Somatic mutations

hsa05200 + CML-COSMIC

File type:  Color specification

Numerical value ( Convert to log scale)

Gradation: from  to

### 1. 右クリックで保存

Use uncolored diagram

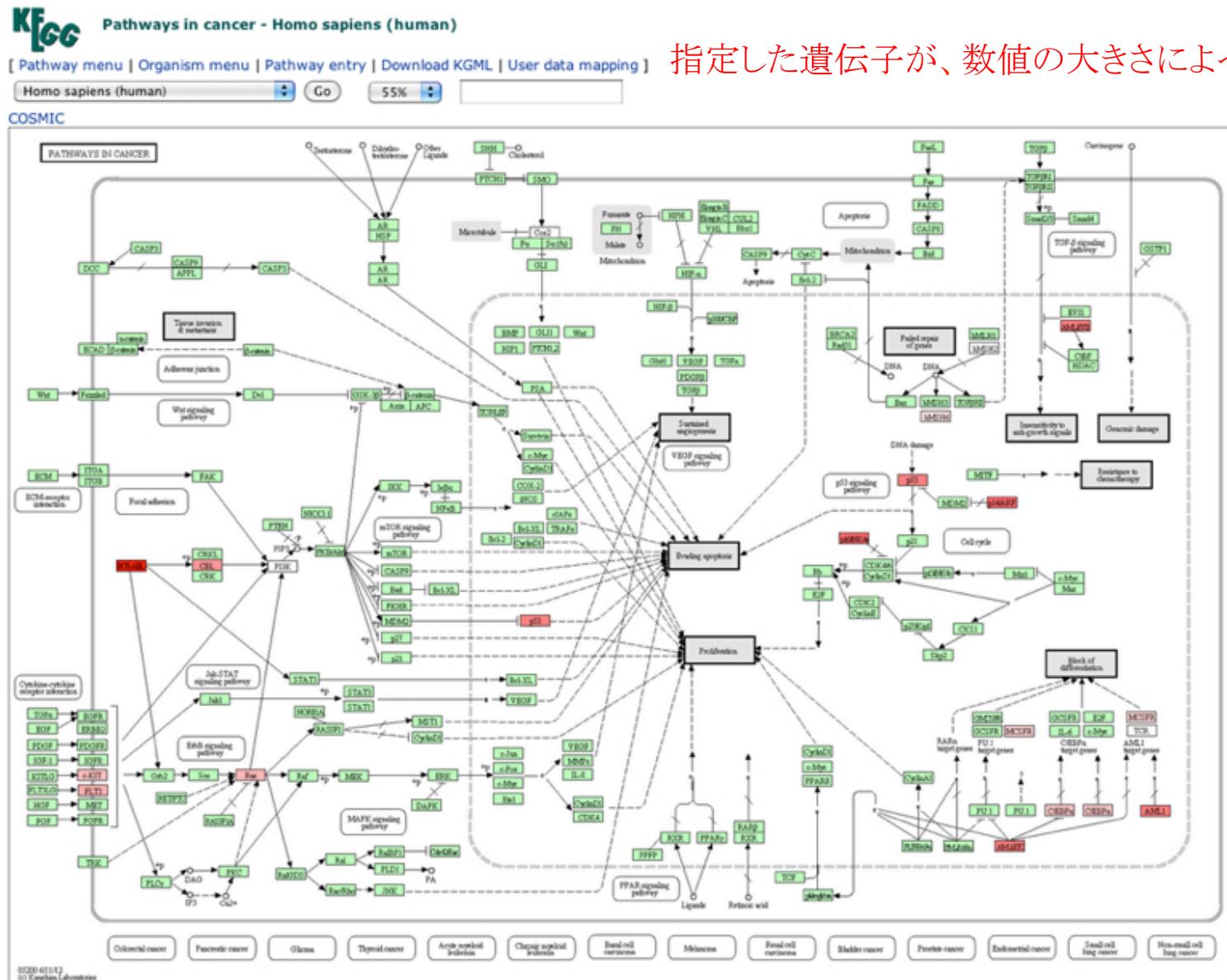
### 3. 保存したファイルを選択

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

#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:2624	9
hsa:2956	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:7403	4
hsa:7490	2
hsa:10320	3
hsa:54790	5
hsa:55294	1
hsa:171023	6



## KEGG Mapperの結果



指定した遺伝子が、数値の大きさによって色付けされる

# 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