

統合データベース講習会：AJACS駿河

2013年1月12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)  
国立遺伝学研究所 DDBJセンター

中村 保一

# 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ったら、講習会のスタッフに気軽に聞いてください

# 代表的な生命科学関係のデータベース

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

# 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

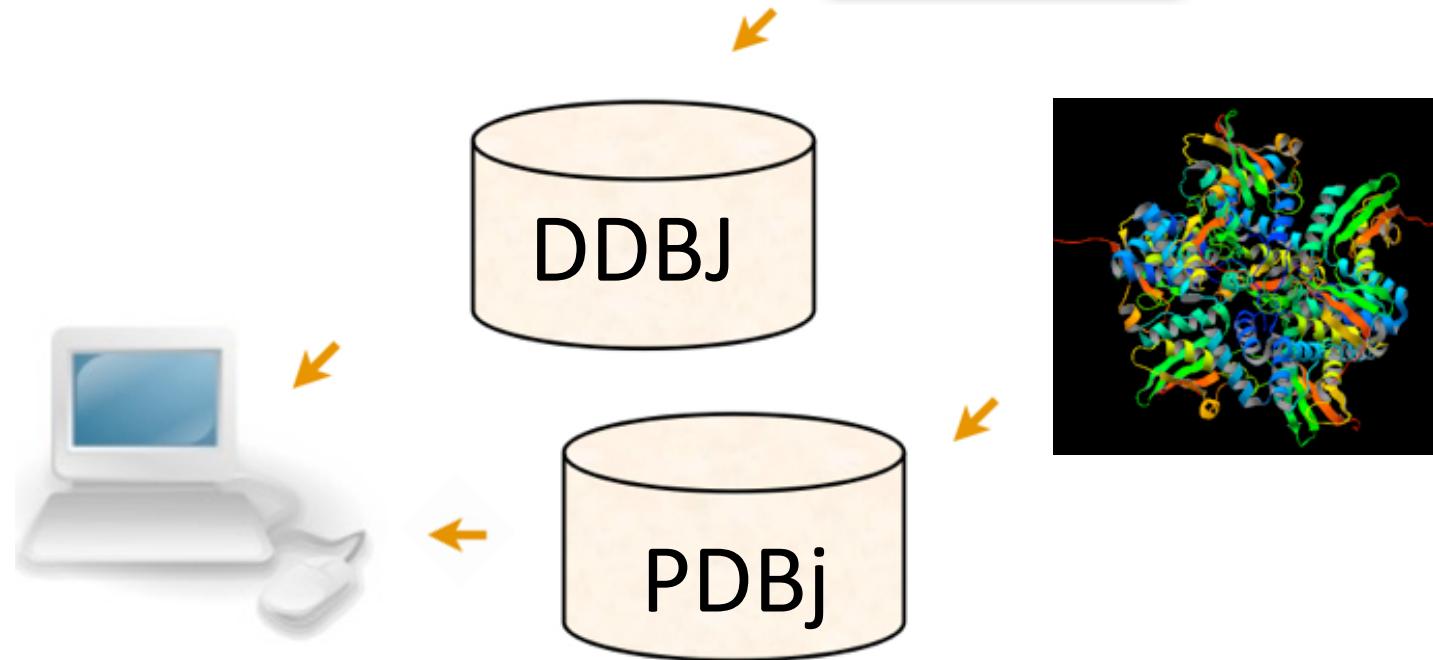
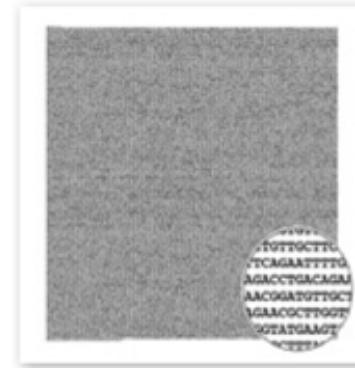
# データバンク

◆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 PLOS logo and the journal title 'PLOS BIOLOGY' are prominently displayed on the left.

## 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

# DDJ

DNA Data Bank of Japan

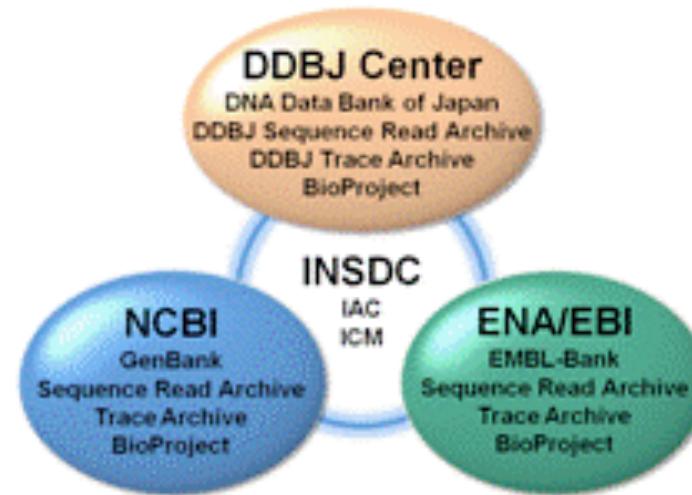
# 国際塩基配列データベースの一員

## ◆ International Nucleotide Sequence Databank Collaboration (INSDC)

○米国：GenBank

○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

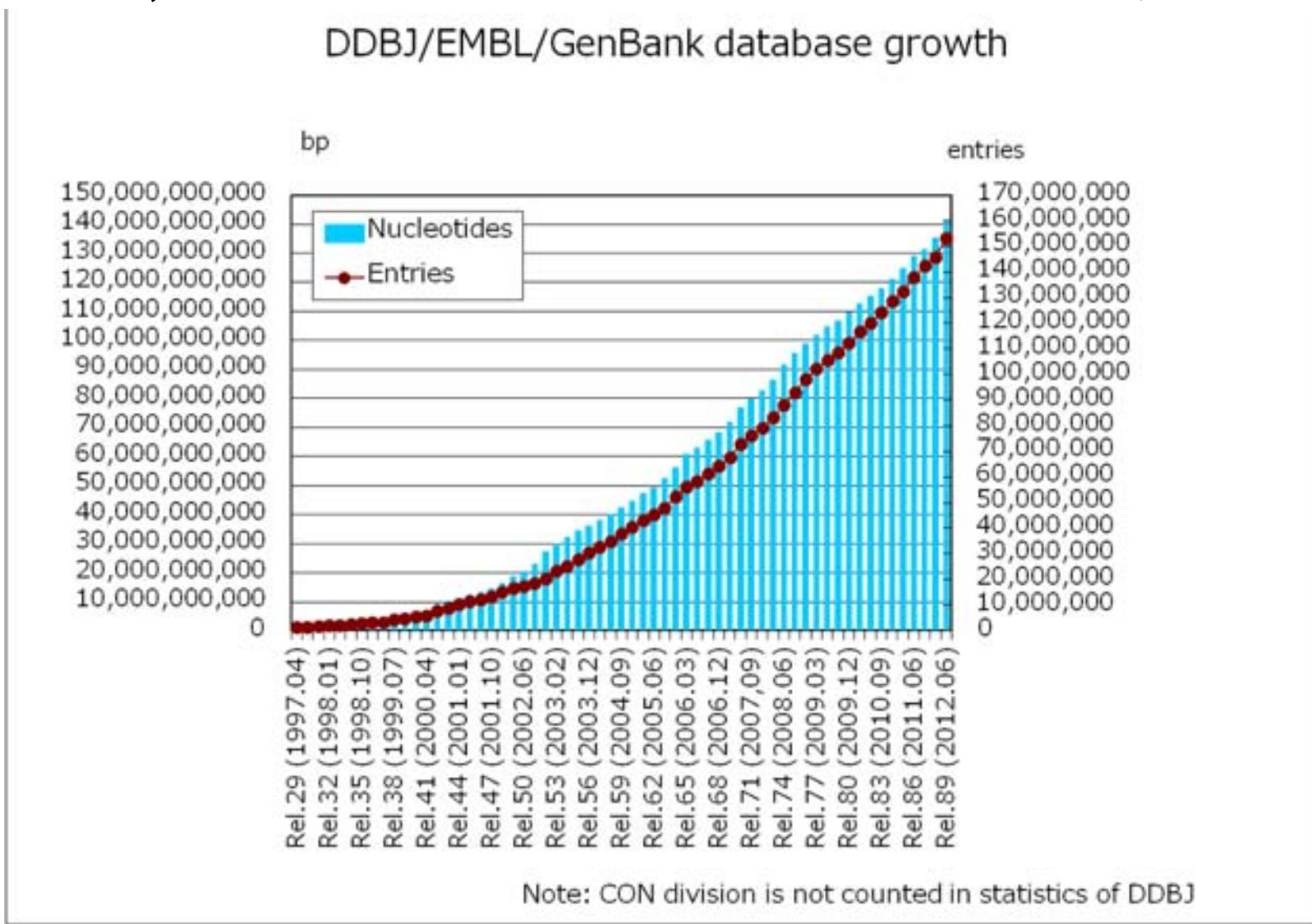
# DDBJ登録ファイルの例

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	
ACCESSION	<a href="#">AB091058</a>						/gene="ccp"	
VERSION	AB091058.1						/product="cellulose complementing protein"	
KEYWORDS	.						/protein_id=" <a href="#">BAC82541.1</a> "	
SOURCE	Gluconacetobacter xylinus						/transl_table=11	
ORGANISM	<a href="#">Gluconacetobacter xylinus</a>						/translation="MSAGSDEVAGGGQAGSPQDFQRVLRSFGVEGGQYSYRPFVDRS	
REFERENCE	Bacteria; Proteobacteria; Alphaproteobacteria; Rhodospirillales; Acetobacteraceae; Gluconacetobacter.						FDTVGVPPEAVERHFQAEHDTAVEEQVTAPQIAVAPPVVVPDPPAIVTETAPPPP	
AUTHORS	Kawano,S., Tajima,K., Uemori,Y., Yamashita,H., Erata,T., Munekata,M. and Takai,M.						VVVSAPVPEPAAVPAEPPVQEAPVQAPVPPVPIAEQAPPAAPDPASPVYAN	
TITLE	Direct Submission						VAAAPVPPDAPVTPAPQARVTGPNTMVEPFSRPQVRVQEGATPSRVSPSRMNAFP	
JOURNAL	Submitted (28-AUG-2002) to the DDBJ/EMBL/GenBank databases.						RTSASSISERPVDRGVADEWSPVPKARLSPRERRPRPGDLSFFFQGMRDTRDEKKFFPV	
COMMENT	Contact:Kenji Tajima Hokkaido University, Graduate School of Engineering; N13W8, Kita-ku, Sapporo, Hokkaido 060-8628, Japan						ASTRSVRNSVSRMTSMKTDTNSSQASRPGSPVASPDSPTMAEVFMTLGGRATELLS	
FEATURES	Location/Qualifiers						PRPSLREALRRRENEEES"	
source	1..2109 /db_xref=" <a href="#">taxon:28448</a> " /mol_type="genomic DNA" /note="synonym:Acetobacter xylinum" /organism=" <a href="#">Gluconacetobacter xylinus</a> " /strain="ATCC 53582"					BASE COUNT	343 a	661 c
CDS	10..1038 /codon_start=1 /gene="cmcase" /product="endo-beta-1,4-glucanase" /protein_id=" <a href="#">BAC82540.1</a> " /transl_table=11 /translation="MSVMAAMGGAQVLSSTGAFADTAPDAVAQQWAIFRAKYLRPSGR VVDTGNNGESHSEGQGYGMLFAASAGDLASFQSWMWARTNLQHTNDKLFWRLKGH QPPVPDFKNNTADGDLILIALGRAGKRFQRPDYIQLDAMAIYGDVLNLMTMKAGPYVVL MPGAVGFTKKDSVILNLSSYYVMPSSLQAFDLTADPFRWRQVMEDGIRLVSAGRFGQWRL PPDWLAVNRATGALSIASWGPPRFSYDAIRVPLYFYWAHMLAPNVLADFTRFWNNFGA NALPGWVWLTTGARSPYNAPPGYLAVAECTGLDSAGELPILDHAPDYYSAALTLVYI ARAEETIK"					661 g	444 t	
ORIGIN						ORIGIN		
						1 cgttccctta tgccggcat ggcggcgat ggagggcgcc aggtgcttc atccacccgt 61 gcgttcgcag acaccggccc cgatcgccgc ggcgcacat gggccatctt ccgcgcacag 121 tatcttcgtc ccagcgacg tgcgtggat acggcaatg gtggcgaatc ccatagttag 181 gggcagggtc atggcatgtc ctggccgcg tcggcggggg accttgcgtc gttccagtcg 241 atgtggatgt gggcgcgcac caacctgcgc cataccaatg acaagctgtt ttccctggcg 301 ttccctcaagg ggcataccgc cccgtgccc gacaagaaca atgccacaga tggcgacttgc 361 ctgatcgccgc ttgcgttgg tcgtcgccgc aagcgttcc acgcgcggca ttacattcg 421 gacgcattgg ccatttatgg ccatgtgtc aacctgtatg cgtatgaaagg gggaccgtat 481 gtcgtcccta tggcccggtc tgcgtggctt accaagaagg acacgcgtat cctcaacctg 541 ttccattaaatgc tcatgcgcgc gtcgtgcgcg tcgtcgacc ttacggccga cccgcgttgc 601 cgtcagggtg tggaaagacgg gattcgccgtt gtcgtccgcg ggcgttccgg gcaatggccg 661 ctgcgtccgcg actggcgtgc ggtatcgccgc cccatcgatcgcgat cgcgtatcgat 721 tggcccgccgc gttttccata tgcgtcgat cgggtgcgcg ttatatttttta ttggcgcat 781 atgcgtggccgc cgaacgtgtt ggctgatcc acccgattct ggaataattt cggggctaat 841 gcccgtcccg gatgggttga tctgacaaca gggccgcgtt cgcgtatcaa cgcgcggcct 901 ggatatcttg ctgttgcgcg atgcacggg ctgttgcgtc cggggactt cccgacactg 961 gatcatcgccgc cgcattatata ttccgcgcgc ttgcgtgcgc tcgttgcgtc cgcgcggcg 1021 gaggagacta taaagtgtatg gtcgttgcgtt ctgtatgagg ggtctggggg gggcaggctg 1081 gaatgtccgcgat ggttttccatgc cgggtcgcgc gtttttttgc tgcgtatgg gggcgttatt 1141 ctatccggcc gttttgttgc cgttgcgtt atgtatgg cgttgcgcgact ggttgttggaaa 1201 ggcacttcgcgat tcaggccggat catgacacgg cgttgcgttgc cgcgttgcgc 1261 aaatcgcgtc cgcaccgcgc cccgcgcgc tcgttgcgtc cgcgcggcc atcgatcgatgg 1321 aaaccgcgcgc cccgcgcgc tcgttgcgtc ggcgtccgcgt cacgtatgaa cccgcgcgc 1381 cgcgcgtccgc ggcagacgc cccgttgcagg aagccccgtt gggcggccgc cccgcgcgc 1441 cgcgcgcgtt gcccccgtt gccggacgg cttcccccgc ggcgcggac cccgcacccgc 1501 tgcgtatgc gaaatcgccgc gcaacccgc tttccatgc tccgcacccgc gttacgcctg 1561 cgcgcgcgc ggcgttgcgtc gggccgcgc cccgtatgtt gggccctt tccgcgcgc 1621 aggtccgcac ggttgcaggag gggccgcgc cgttgcgttgc accttcgcgt tcaatgcac 1681 ctgttgcgttgc cacaatcgatc tcgttgcgttgc gtttgcgttgc ggttgcgttgc 1741 atgaatggat ggttgcgttgc aaggccgcgc tcgcgcgcgc ggcgttgcgtc cgcgcgc 1801 atgcgttgcgttgc tttccatgc gggatcgccgc acacccgttgc tggaaaggat tttccatgc 1861 tggcgccac ggcgttgcgttgc ggcgttgcgttgc gtttgcgttgc gggccgttgc 1921 acaccaatgc tttccatgc ggcgttgcgttgc ggcgttgcgttgc gggccgttgc 1981 ccacaatgc ggcgttgcgttgc ggcgttgcgttgc gggccgttgc gggccgttgc 2041 gtcgttgcgttgc ggcgttgcgttgc ggcgttgcgttgc gggccgttgc 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  

[HOME](#) [データ登録](#) [利用の手引き](#) [検索・解析](#) [FTP・WebAPI](#) [レポート・統計](#) [お問い合わせ](#) [サイト内検索](#)

**DDBJ の紹介**

- [DDBJの紹介](#)
- [Q&A集](#)

**塩基配列の登録**

- [DDBJ塩基配列登録システム](#) NEW
- [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**

- [スパコンの利用申込](#)
- [スパコンの利用方法](#)
- [スパコン紹介資料](#)

**検索**

- [getentry](#)
- [ARSA](#)
- [TXSearch](#)
- [BLAST](#)

**DDBJ : DNA Data Bank of Japan**

DDBJ（日本DNAデータバンク）は欧州と米国に対応機関（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 Hideki Nagasaki

**Hot Topics**

- 2013.01.11 [DDBJ Web Magazine No.78 配信](#)
- 2012.12.27 [シーラカンス \(\*Latimeria chalumnae\*\) GSS データの公開](#)
- 2012.12.06 [「第26回 DDBJing 講習会 in 三島」開催のお知らせ \(参加申込み受付中\)](#)

**Maintenance**

- 2012.12.05 [DDBJ 年末年始休業のお知らせ](#)
- 2012.11.26 [11/22 の落雷によるサービス停止のお詫び](#)

**Information**

- [2012.07.31] [富田勝教授の個人ゲノム配列が公開](#)

**塩基配列の登録・更新**

- [塩基配列の登録](#)

**FTP・Web API**

- [FTP \(\[ftp.ddbj.nig.ac.jp\]\(ftp://ftp.ddbj.nig.ac.jp\)\)](#)

# 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

You can use "Boolean Operators" ("AND(&)", "OR(|)", and "AND NOT(!)").

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 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
<input type="checkbox"/> AB035921	Escherichia coli O157:H7 murI, rrsB, gltT, rrlB, rrfB, murB genes for glutamate racemase, 16S rRNA, glutamate tRNA 2, 23S rRNA, 5S rRNA, UDP-N-acetyl enolpyruvyl glucosamine reductase, partial cds.	DNA	Escherichia coli O157:H7	6903

検索結果：  
ヒット数が多くすぎるので  
絞り込みが必要



詳細検索  
(Advanced Search)へ

DDBJ(2201)  
Patent\_AA(3)

Escherichia coli(731)  
Unknown\_(561)  
Escherichia coli  
O157:H7(241)  
unidentified(167)  
synthetic  
construct(160)  
Escherichia coli  
O157:H7 str.  
EDL933(131)  
Pasteuria penetrans(15)  
Gallus gallus(10)  
Hydractinia echinata(8)  
Mycobacterium bovis  
AF2122/97(8)  
All other taxa(172)

PAT(1287)  
BCT(819)  
PHG(38)  
ECT(241)

# いくつかの特徴で絞り込み

ARSA All-round Retrieval of Sequence and Annotation

DBJQuick Search DBJAdvanced 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。
- ダブルクォーテーション ("")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
<input type="checkbox"/> AB602479	C. glutamicum-E. coli shuttle vector pCRB12 DNA, complete sequence.	DNA	C. glutamicum-E. coli shuttle vector pCRB12	4569
<input type="checkbox"/> AB671168	E. coli-T. thermophilus shuttle vector pTRK1T DNA, complete sequence.	DNA	E. coli-T. thermophilus shuttle vector pTRK1T	7482
<input type="checkbox"/> AB671169	E. coli-T. thermophilus shuttle vector pTRH1T DNA, complete sequence.	DNA	E. coli-T. thermophilus shuttle vector pTRH1T	6057
<input type="checkbox"/> HM126493	C. glutamicum-E. coli shuttle vector pCRB62, complete sequence.	DNA	C. glutamicum-E. coli shuttle vector pCRB62	5914
<input type="checkbox"/> 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

# 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

DDJBQuick Search DDJAdvanced Search Refine Search

DDBJ(3) 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]

17

## 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  Entrez, The Life Sciences Search Engine.

HOME SEARCH SITE MAP PubMed All Databases Human Genome GenBank Map Viewer BLAST

Search across databases |  GO Clear 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

NCBI Entrez, The Life Sciences Search Engine.

HOME SEARCH SITE MAP 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	

17610  Nucleotide: Core subset of nucleotide sequence records	2  dbGaP: genotype and phenotype
4  EST: Expressed Sequence Tag records	none  UniGene: gene-oriented clusters of transcript sequences
none  GSS: Genome Survey Sequence records	2  CDD: conserved protein domain database
31034  Protein: sequence database	none  Clone: integrated data for clone resources
1  Genome: whole genome sequences	none  UniSTS: markers and mapping data
3  Structure: three-dimensional macromolecular structures	3591  PopSet: population study data sets
none  Taxonomy: organisms in GenBank	none  GEO Profiles: expression and molecular abundance profiles
none  SNP: short genetic variations	142  GEO DataSets: experimental sets of GEO data

# Entrez: 大腸菌ゲノムページ



Genome

Genome

Escherichia coli O157 sakai

Search

Help

Display Settings: Overview

Send to:

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



## Escherichia coli

A well-studied enteric bacterium

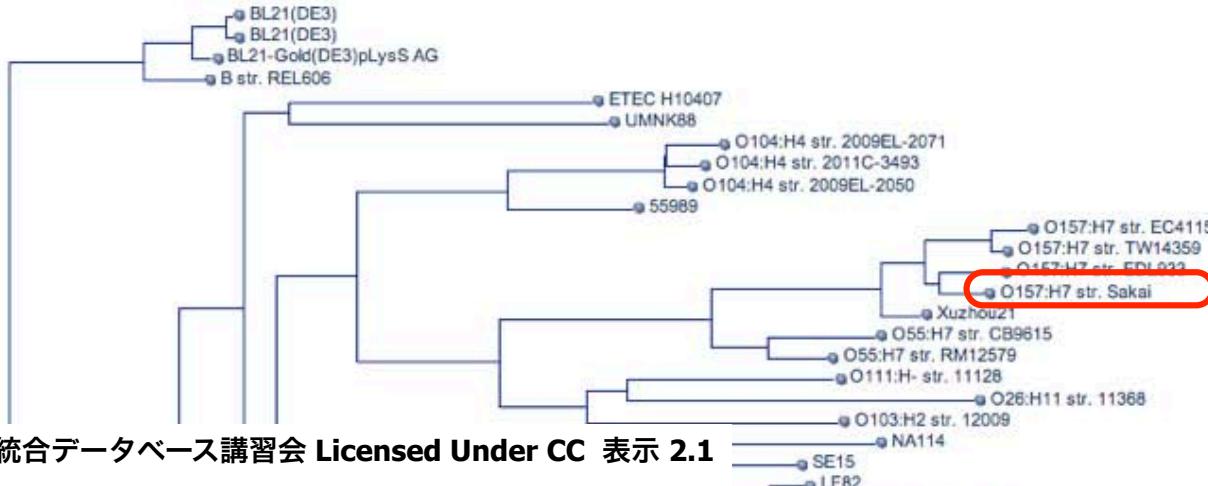
Lineage: Bacteria[3655]; Proteobacteria[1529]; Gammaproteobacteria[689]; Enterobacteriales[156]; Enterobacteriaceae[156]; Escherichia[7]; Escherichia coli[1]

**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...](#)

### Representatives

1. Reference genome, Community selected, Proteomics : Escherichia coli str. K-12 substr. MG1655
2. Reference genome, Community selected : Escherichia coli O157:H7 str. Sakai
3. Calculated : Escherichia coli SE11

### 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...

# 大腸菌O157 Sakai株のページ

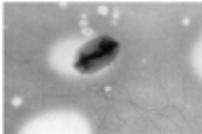
NCBI Resources How To My NCBI Sign In

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...

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 Limits Advanced Search 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;  
Enterobacteriaceae; 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.Ecop933DamP [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.EcoKO157ORFAP [REBASE - The Restriction Enzy...]  
REBASE enzyme EcoKO157ORF5262P [REBASE - The Restriction Enzy...]

# 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ーを設計する
- ◆ “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

	gene names	internal date format	default format		gene names	internal date format	default format		gene names	internal date format	default 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

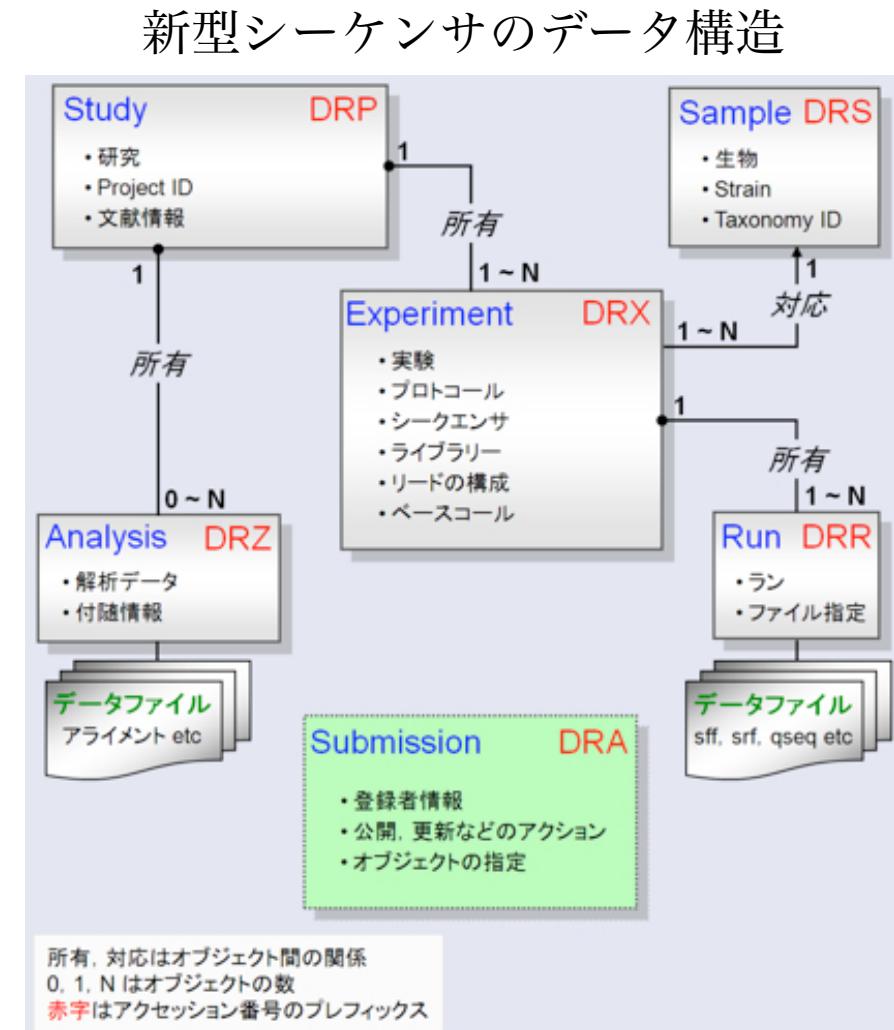
1. データ公開  
DDBJ Sequence Read Archive (DRA) では以下の場合にデータを公開しています。

- 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人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. 公開予定期日が到来した場合
- 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 blue submarine and a cartoon character, with the text "SRAs: Survey of Read Archives". Below the logo, there are three main sections: "What's SRAs? - SRAsって何?", "Simple Lists - まずは見てみる", and "Search studies - まずは検索してみる". The "What's SRAs?" section provides a brief description of the service. The "Simple Lists" section offers links to search by study, experiment, or run. The "Search studies" section contains a search form 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.

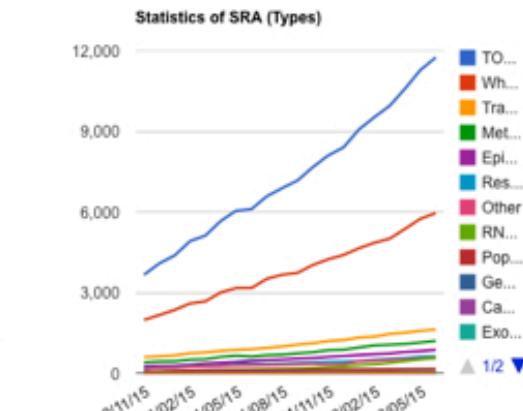
実際のプロジェクト数は"Study Types"に書かれているものです。"Platforms"や"Species of Samples"のtotalの値は、複数のプラットフォームで行われた実験が1つのプロジェクトでなされる場合が多くあり、ダブルカウントしているので、数字が大きくなっています。

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

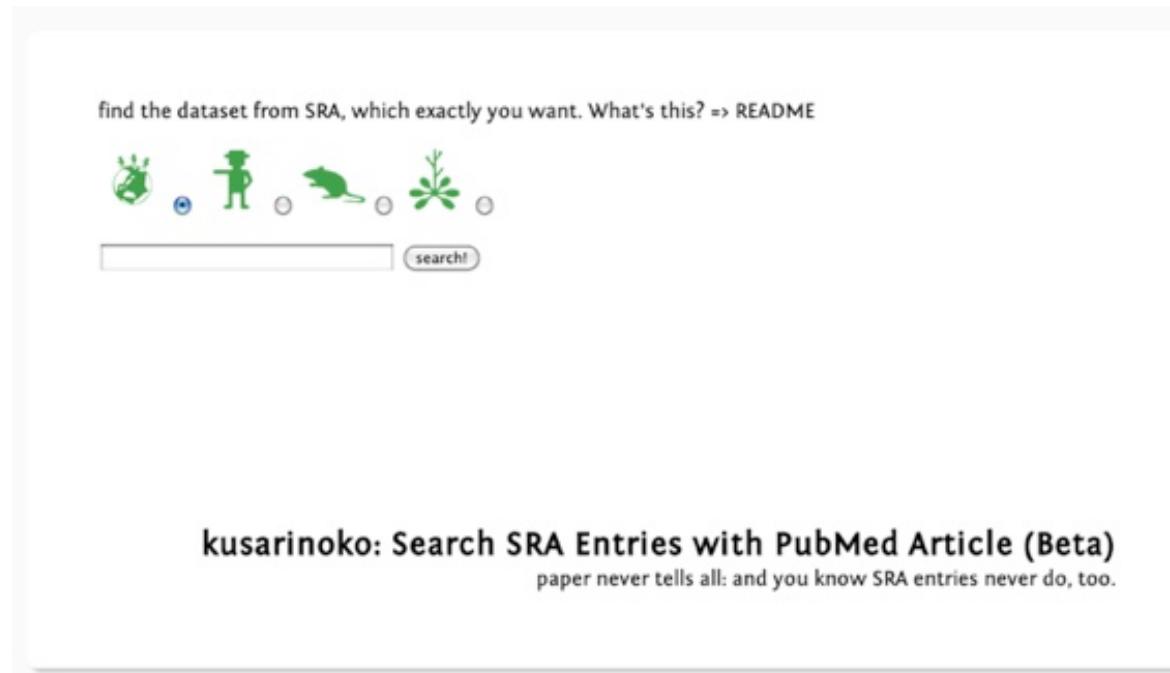
Illumina Genome Analyzer II	3629	Illumina HiSeq 2000	45470
454 GS FLX Titanium	2848	Illumina Genome Analyzer II	36940
Illumina HiSeq 2000	2328	454 GS FLX Titanium	11977
454 GS FLX	1735	Illumina Genome Analyzer	9051
Illumina Genome Analyzer	1167	454 GS FLX	8939
Illumina Genome Analyzer Iix	565	Illumina Genome Analyzer Iix	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 (studies)	Total	127426 (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>



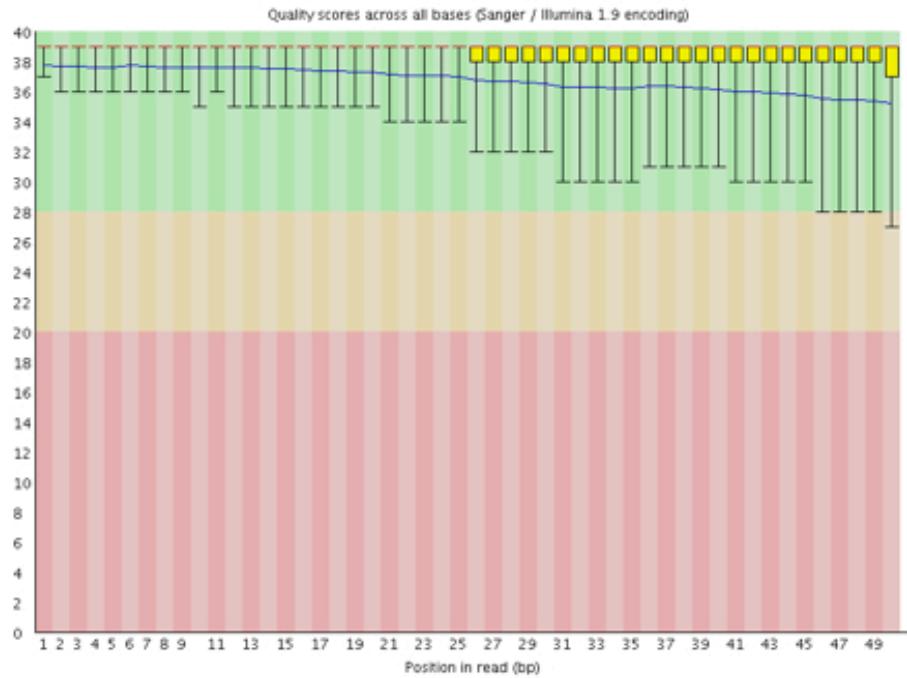
# 鎖鋸 (kusarinoko)

◆ 独自に”FastQC”をかけて、  
それぞれのデータの質を評価

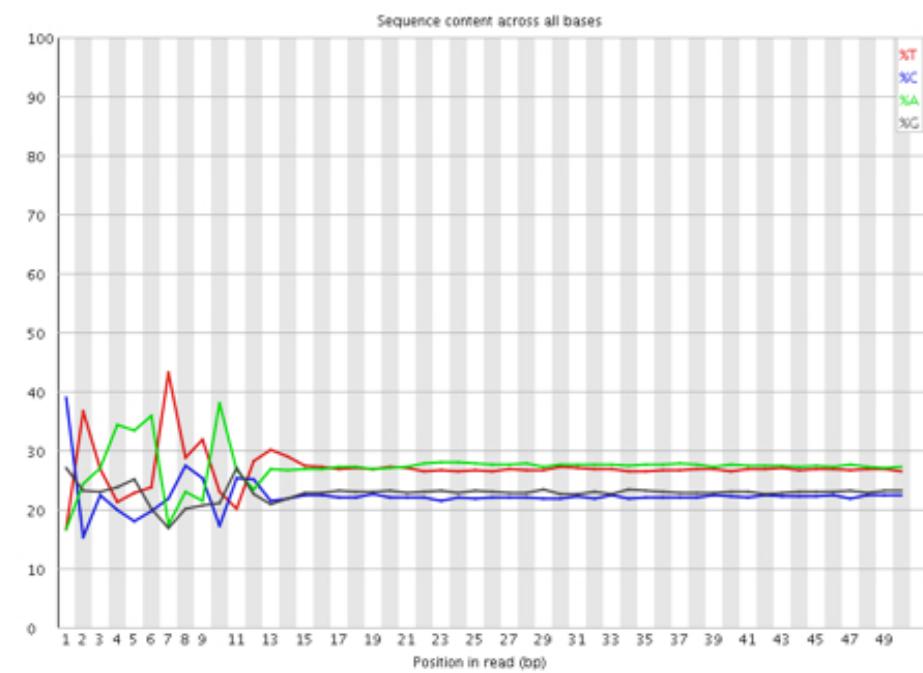
## Basic Statistics

- 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

○ <http://p.ddbj.nig.ac.jp/>

The screenshot shows the 'Select Tools' step of the pipeline. The top navigation bar includes 'Select Query Files', 'Select Tools' (which is highlighted in yellow), 'Set QuerySet', 'Set GenomeSet', 'Set Map Options', and 'Confirmation'. Below this is a 'Running Status' section. The main content area is titled 'Selecting Tools for Basic Analysis of DDBJ ANNOTATION PIPELINE'. It contains three sections: 'Reference Genome Mapping', 'de novo Assembly', and 'Mapping Contigs by de novo Assemble to Reference Sequences'. Each section has a table with columns for Tool, Help, Version, Base space, Color space, Paired-end, Depth, Coverage, Error rate, SNP, Indel, gff, bed, SAM, and Comment.

**Reference Genome Mapping**

Tool	Help	Version	Base space	Color space	Paired-end	Depth	Coverage	Error rate	SNP	Indel	gff	bed	SAM	Comment
BLAT	<a href="#">?</a>	34	✓					✓						Single-end analysis only
Maq	<a href="#">?</a>	0.7.1	✓		✓			✓	✓	✓	✓	✓	✓	
bwa	<a href="#">?</a>	0.5.9	✓		✓			✓					✓	
SOAP	<a href="#">?</a>	2.21	✓		✓			✓	✓	✓			✓	
Bowtie	<a href="#">?</a>	0.12.7	✓	✓	✓			✓	✓				✓	
TopHat	<a href="#">?</a>	1.0.11	✓		✓			✓					✓	

**de novo Assembly**  
Total limit = 22 Gbp

Tool	Help	Version	Base space	Color space	Paired-end	MSS(WGS)	Comment
SOAPdenovo	<a href="#">?</a>	1.05			✓		
ABySS	<a href="#">?</a>	1.3.2			✓		Maximum K-mer value is 64.
Velvet	<a href="#">?</a>	1.2.03			✓	✓	We severly recommend when performing Velvet, total length of those reads is up to 22G bp. Maximum K-mer value is 64.
Trinity	<a href="#">?</a>	r2012-04-27			✓		RNA-Seq De novo Assembly

**Mapping Contigs by de novo Assemble to Reference Sequences.**  
The contigs will be aligned to reference genome.

Tool	Comment
BLAT	Single-end analysis only

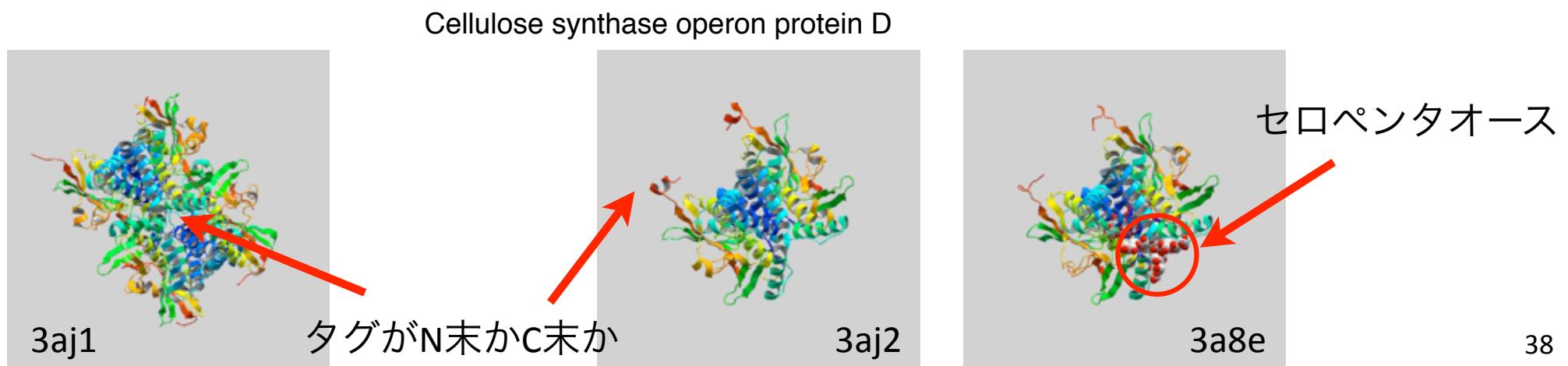
BACK NEXT

# 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86,000 ([www.wwpdb.org](http://www.wwpdb.org))
- 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)

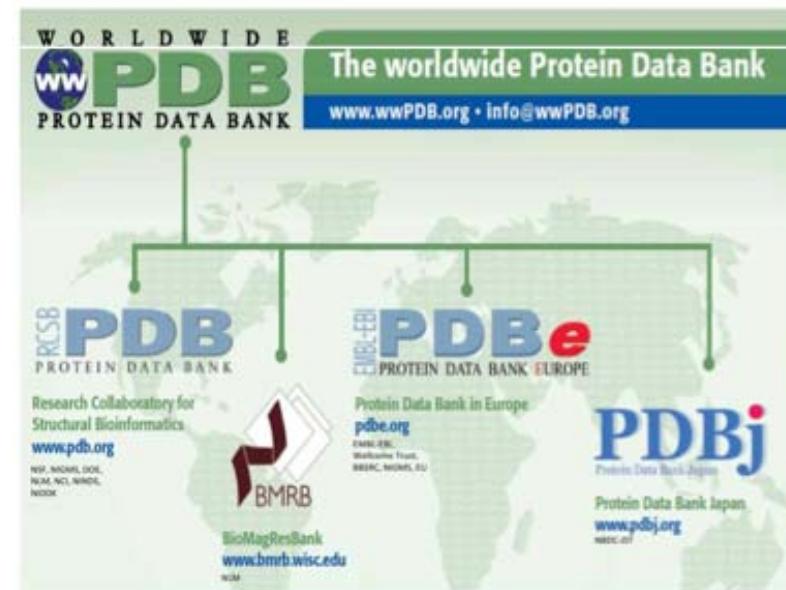
○米国：RCSB-PDB (Research Collaboratory for Structural Bioinformatics)

○欧州：PDBe

○日本：PDBj

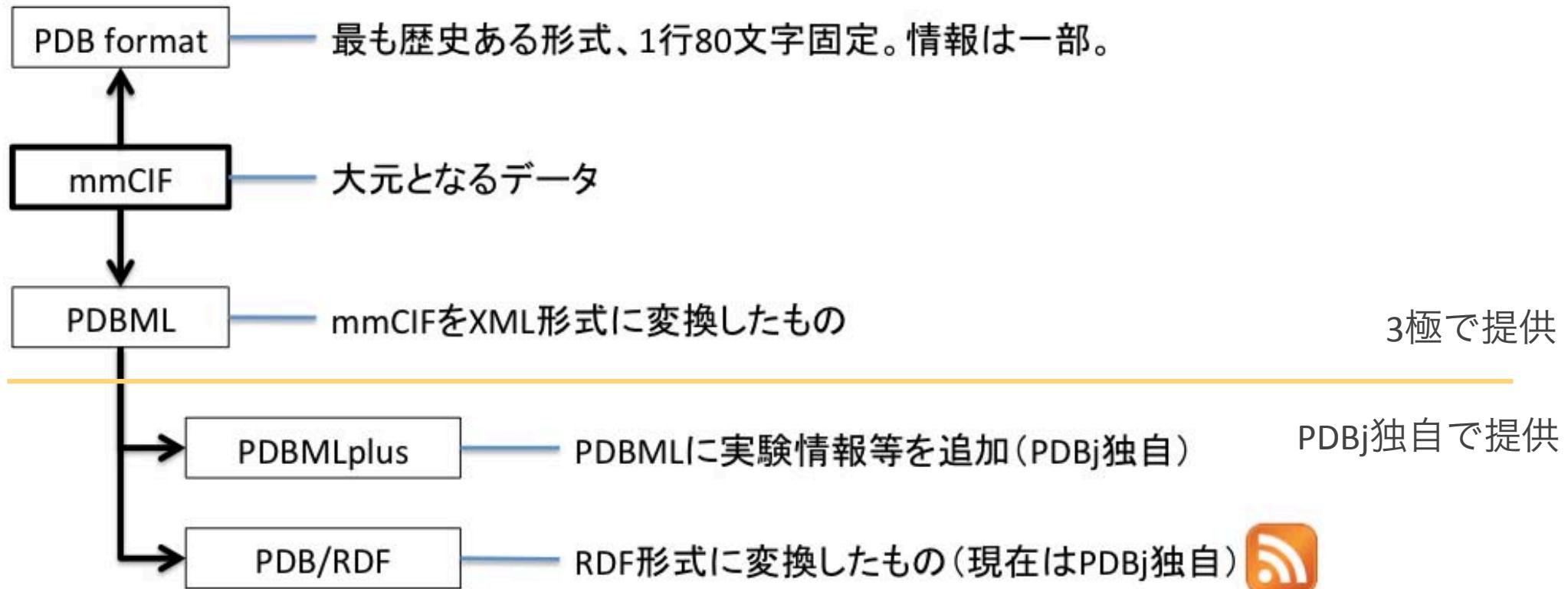
○米国：BMRB

(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 Å  
 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 3 RESOLUTION 2.00 ANGSTROMS

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 GLN A 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 GLUI 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://pdjb.org/> を直接入力

# PDBj トップページ

The screenshot shows the PDBj homepage with a red box highlighting the "Japanese" link in the top navigation bar. A red arrow points from the search bar area on the left to a larger search interface on the right.

**Top Navigation:** PDBj, English, Japanese (highlighted), Simplified Chinese, traditional Chinese, Korean, 統計情報, ヘルプ, FAQ, お問い合わせ

**Left Sidebar:**

- トップページ
- データ登録 >>
  - ADIT: PDB Deposition
  - ADIT-NMR
- 検索 >>
  - Search PDB (PDBj 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
  - SEALa
  - 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

**Center Content:**

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。

**Search Bar (Left):** 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 検索  
αアミラーゼ

**Search Bar (Right):** 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 検索  
αアミラーゼ

**Right Sidebar:**

- 82522 entries available on 20 Jun., 2012 00:00(UTC) / 09:00(JST)
- WORLDWIDE PROTEIN DATA BANK
- eProtS Encyclopedia of Protein Structures
- Protein Globe
- DBCLS
- KU-OPT
- C Protein Center

# 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

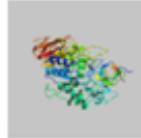
クエリ : αアミラーゼ PDB ID または キーワード 表示順: 一致件数

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

	<b>1clv</b>	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
		公開日 : 2000-05-03
		最終更新日 : 2009-02-24

	<b>1bvn</b>	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
		公開日 : 1998-09-23
		最終更新日 : 2009-02-24

# PDBjエントリページ

日本語ページについて  
PDBj Mineについて  
更新情報

概要 [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

<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  
(jV4 / Jmol) で見る \*1

配列の表示

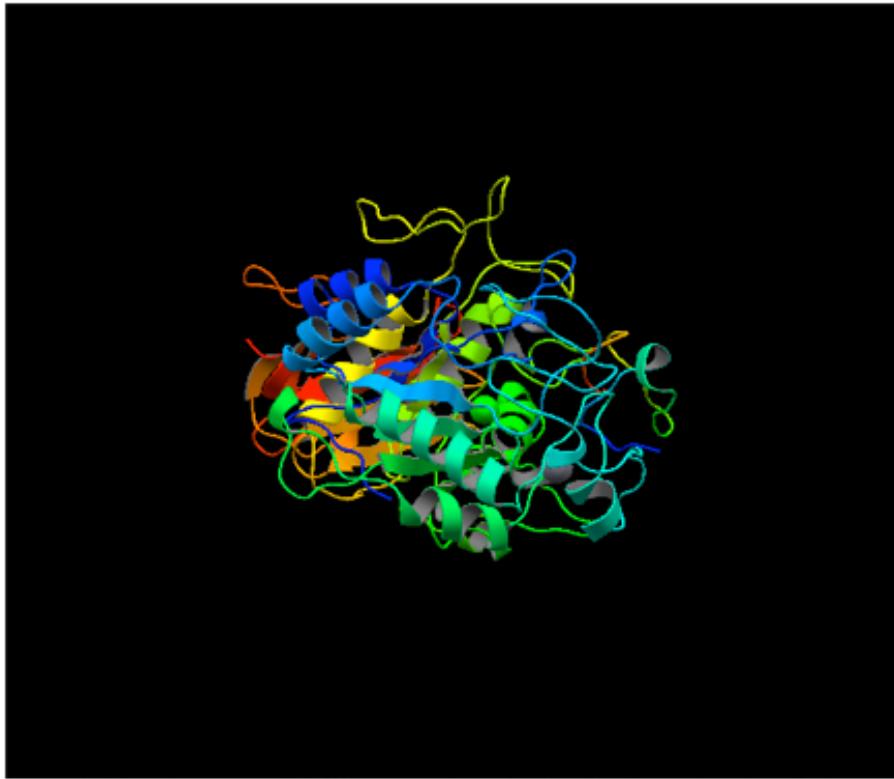
PDBファイルのダウンロード

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	万見(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

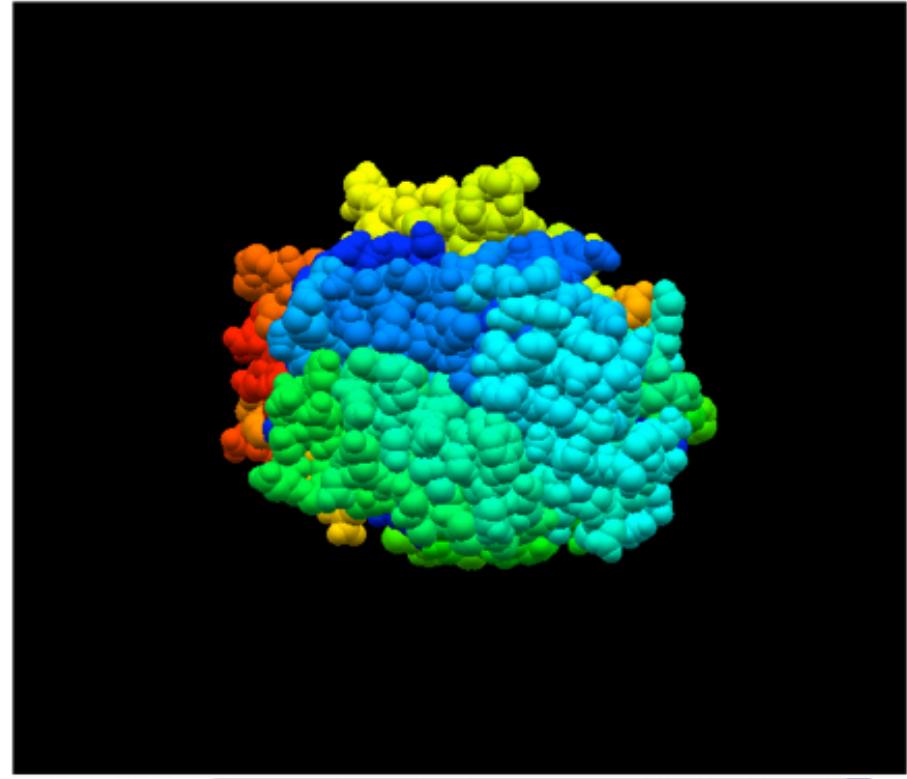
マウスで動かせる画像を表示

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<sub>45</sub>

# 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**  
**PDB 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

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, 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: KDAFNASGRNSIVHLFEWKWNDADECERFLQPQGFGGVQISPPNEYLVADGRPWWERYQPVSYIINTRSGDESFTDMTRRCNDAGVRIYVDavinHMTGMNGVGTGSSADHDGMNYPAVPYGSGDFHSPCEVNYYQDADNRNCELVGLRDLNQGS  
Structure:  
Site:

161 171 181 191 201 211 221 231 241 251 261 271 281 291 301 311 321  
DYVRGVVIDYMNHMIDLGVAGFRVDAAKHMSPGDLSVIESGLKNLNNTDYGFADGARPFYQEVIDLGGEAIKNEYTGFGCVLEFQEGVSLGNNAFGQGGNQLKLNLANWGPENGLEGLDAVVFDNHNDNQRTGGSQLITYKNPKPYKMAIAFMLAHPYGTTTRIMSSFDTD

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** 適能情報 相同蛋白質 ダウンロード/画面表示 外部データベース  
検索条件 PDF 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	CaCl2	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 carbon
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



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#:

EKDANFASGRNSIVHLFEWKWNDIADECERFLQPQGFGGVQISPPNEYLVADGRPWWERYQPVSYIINTRSGDESAFTDM  
TRRCNDAGVRIYDAVINHMTGMNGVGTSGSSADHDGMNYPAVPYGSGDFHSPCEVNYYQDADNVRNCELVGLRDLNQGS  
DYVRGVLDYMNHMIDLGVAGFRVDAAKHMSPGDLSVIFSGKLKNLNTDYGFADGARPFYQEVIDLGGEAISKNEYTGFG  
CVLEFQFGVSLGNNAFQGGNNQLKLNLANWGPWEWLLEGGLDAVVFDNHDNQRTGGSQLTYKNPKPYKMAIAFMLAHPYGTT  
RIMSSFDFTDNDQGPPQDGSGNLISPGINDDNTCSNGYVCEHRWRQVYGMVGFRNAVEGTQVENWWSNDDNQIAFSRGSQ  
GFVAFTNGGDLNQNLTGLPAGTYCDVISGELSGSCTGKSCTVGDNGSADISLSAEDDGVLAIHVNAKL

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

CIPKWNRCPKMDGVPCCPYTCTSDYYGNCS

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 1471QKDANFASGRNSIVHLFEWKWNDIADECERFLQPQGFGGVQISPPNEYLVADGRPWWERYQPVSYIINTRSGDESAFTDMTRRCNDAGV  
1tmqA 1471QKDANFASGRNSIVHLFEWKWNDIADECERFLQPQGFGGVQISPPNEYLVADGRPWWERYQPVSYIINTRSGDESAFTDMTRRCNDAGV

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 \[1viwA\]](#) 構造の重ね合わせ

1clvA 1471QKDANFASGRNSIVHLFEWKWNDIADECERFLQPQGFGGVQISPPNEYLVADGRPWWERYQPVSYIINTRSGDESAFTDMTRRCNDAGV  
1viwA 1471QKDANFASGRNSIVHLFEWKWNDIADECERFLQPQGFGGVQISPPNEYLVADGRPWWERYQPVSYIINTRSGNESAFDMTRRCNDAGV

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 1498QKDANFASGRNSIVHLFEWKWNDIADECERFLQPQGFGGVQISPPNEYLVADG--RPWWERYQPVSYIINTRSGDESAFTDMTRRCNDAGV  
3vm5A 1498QHNPNTRDGRTAIVHLFEWRWADIAECERFLGPKGFAGVQISPPNEHILVSSPWRPWQRYPISYNLCSRSGGENELRDMITRCNNV

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

Mine PDB

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 外部リンク  
PDB ID または キー:

## 1clvA (クエリ) と 1viwA (テンプレート) の重ね合わせ

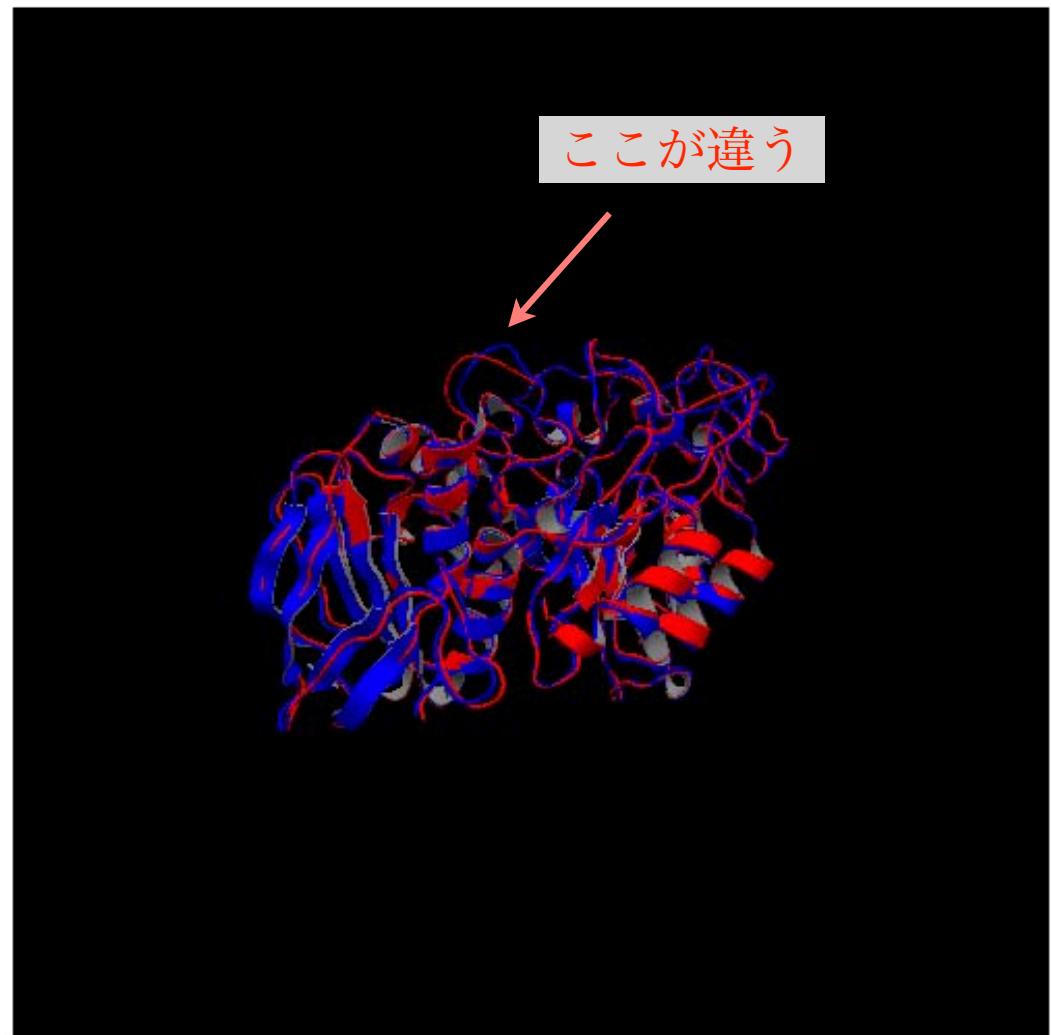
整列した残基数 = 470

### RMSD 重ね合わせ:

CA RMSD = 0.86Å [PDBファイル JV4 [旧Java Plug-in / 次世代Java Plug-in ル]

1clvA	KDANFASGRNSIVHLFEWKWNDADECERFLQPQGFGGVQISPPNEYLVADGRPWE
1clvA	CCCCCCCCCCCEEECCCCHHHHHHHHCCCCCCCCCEEECCCCCCCCCCCCCCCCHHHHHI
1viwA	KDANFASGRNSIVHLFEWKWNDADECERFLQPQGFGGVQISPPNEYLVADGRPWE
1viwA	CCCCCCCCCCCCCCCCCCCCCCHHHHHHHHCCHHCCCEEECCCCCCCCCCCCCCCCHHHH
Ner類似度	998999
1clvA	PVSYIINTRSGDESAFTDMTRRCNDAGVRIVDAVINHMTGMNGVGTSGSSADHDGMN
1clvA	CCCCCCCCCCCHHHHHHHHHHHHHCCCEEEEEECCCCCCCCCCCCCCCCCCCCCCCCCCCC
1viwA	PVSYIINTRSGNEASAFTDMTRRCNDAGVRIVDAVINHMTGMNGVGTSGSSADHDGMN
1viwA	CCCCCCCCCCCHHHHHHHHHHHHHCCCEEEEEECCCCCCCCCCCCCCCCCCCCCCCCCCCC
Ner類似度	999

## 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

Mine

日本語ページについて  
PDBj Mineについて  
更新情報

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)	圧縮	<a href="#">表示</a>	<a href="#">ダウンロード</a>	
	ヘッダのみ	pdb1clv.ent(364k)	非圧縮	<a href="#">表示</a>	<a href="#">ダウンロード</a>	
mmCIF	全ての情報	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>	
		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)	メタデータのみ	<a href="#">表示</a>	<a href="#">ダウンロード</a>
	座標情報のみ (ext-atom)	PDBMLplus	1clv-pdbmlplus.xml.gz(30k)	メタデータのみ (PDBj独自)	<a href="#">表示</a>	<a href="#">ダウンロード</a>
	RDF		1clv-extatom.xml.gz(96k)	原子座標データのみ (メタデータなし)	<a href="#">表示</a>	<a href="#">ダウンロード</a>
生物学的単位 (PDB形式)		1clv-noatom.rdf.gz(18k)	RDF形式データ (PDBj独自)	<a href="#">表示</a>	<a href="#">ダウンロード</a>	
		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あとの解析例

---

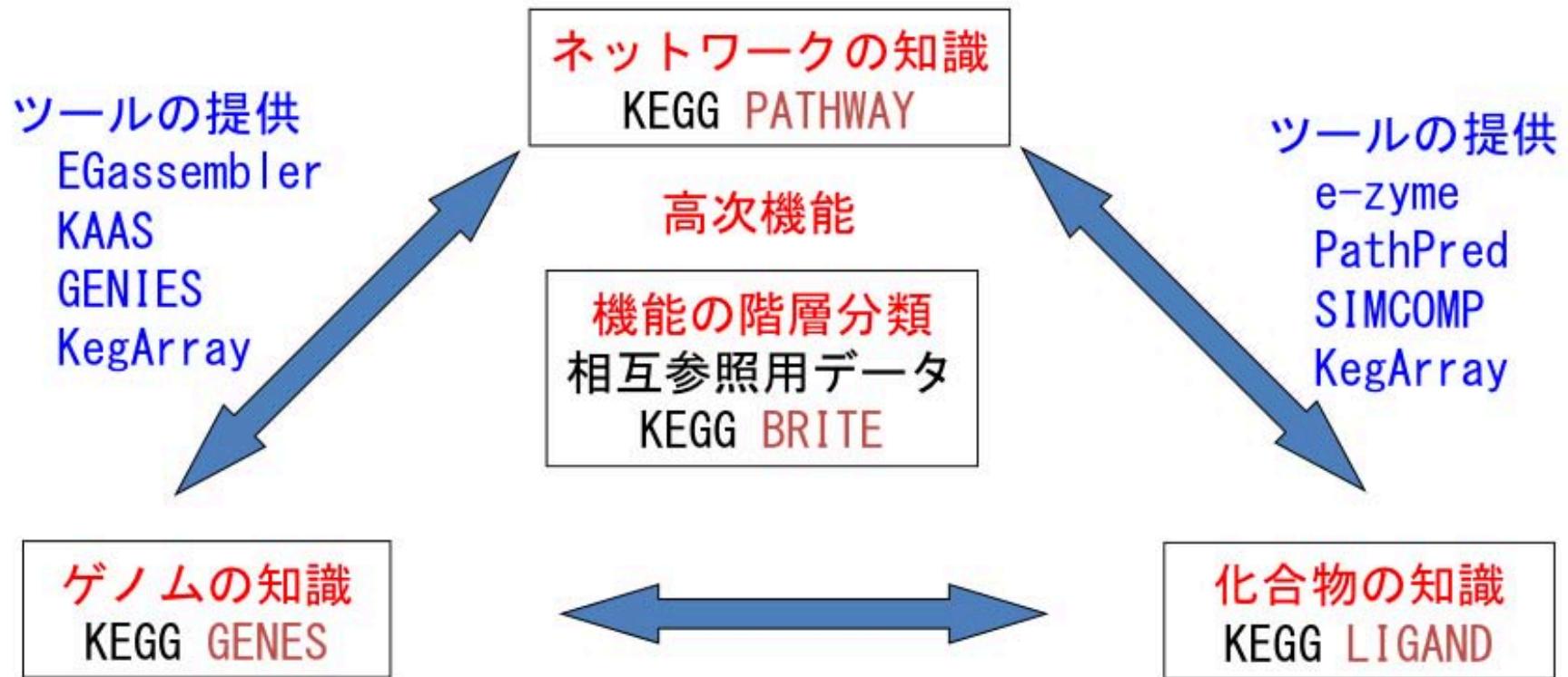
- ◆ “PyMol”, “Chimera”等で立体構造を描画する
- ◆ ”DALI”, “MATRAS”等で立体構造を比較する
- ◆ ”SURFNET”等でタンパク質表面のポケット  
(化合物が結合しやすい) を検出する
- ◆ ”DelPhi”等でタンパク質の表面電荷を計算する
- ◆ ”UCSF DOCK”, “AutoDock”等で化合物とのドッキングシミュレーションを行う

# 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/>

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** 製能階層・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** オーソログアナテーション
  - 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

## KEGG - Table of Contents

Search KEGG  for [Go](#) [Clear](#)

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

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

---

◆ “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

# 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	EST datasets	Draft genomes with automatic (KAAS) annotation
EGENES	Metagenomes	EST contigs with automatic (KAAS) annotation
MGENES	Viruses	Metagenomes with automatic (KAAS) annotation
VGENES		No annotation; available only in <a href="#">DBGET</a>

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: 178 Bacteria: 2104 Archaea: 145

[ Genomes | Draft | ESTs | Meta | Pan ]

#### Eukaryotes

Category	Species	Source
Mammals	hsa Homo sapiens (human)	RefSeq
	ptr Pan troglodytes (chimpanzee)	RefSeq
	pps Pan paniscus (bonobo)	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
	fca Felis catus (domestic cat)	RefSeq
Vertebrates	bta Bos taurus (cow)	RefSeq
	ssc Sus scrofa (pig)	RefSeq
	ecb Equus caballus (horse)	RefSeq
	mdo Monodelphis domestica (opossum)	RefSeq
	shr Sarcophilus harrisii (Tasmanian devil)	RefSeq
	oaa Ornithorhynchus anatinus (platypus)	RefSeq
	gga Gallus gallus (chicken)	RefSeq
	mgp Meleagris gallopavo (turkey)	RefSeq
	tgu Taeniopygia guttata (zebra finch)	RefSeq
	acs Anolis carolinensis (green anole)	RefSeq
Reptiles	xla Xenopus laevis (African clawed frog)	RefSeq
	xtr Xenopus tropicalis (western clawed frog)	RefSeq
	dre Danio rerio (zebrafish)	RefSeq
	tru Takifugu rubripes (torafugu)	RefSeq
Fishes	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
Lancelets	dpo Drosophila pseudoobscura pseudoobscura	RefSeq
	dan Drosophila ananassae	RefSeq
	--	--
Animals		

## 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:280  
AMY2B, AMY2; amylase, alpha 2B (pancreatic) (EC:3.2.1.1); K01176 alpha-amylase [EC:3.2.1.1]  
hsa:9654  
TTLL4; tubulin tyrosine ligase-like family, member 4; K16601 tubulin polyglutamylase TTLL4 [EC:6.---]  
hsa:23093  
TTLL5, KIAA0998, STAMP; tubulin tyrosine ligase-like family, member 5; K16602 tubulin polyglutamylase TTLL5 [EC:6.---]  
hsa:25941  
TPGS2, C18orf10, HsT3006, L17, PGs2; tubulin polyglutamylase complex subunit 2; K16605 tubulin polyglutamylase complex subunit 2  
hsa:25809  
TTLL1, C22orf7, HS323M22B; tubulin tyrosine ligase-like family, member 1; K16599 tubulin polyglutamylase TTLL1 [EC:6.---]  
hsa:91978  
TPGS1, C19orf20, GTRGEO22, PGs1; tubulin polyglutamylase complex subunit 1; K16581 tubulin polyglutamylase complex subunit 1  
hsa:79739  
TTLL7, RP5-836J3.2; tubulin tyrosine ligase-like family, member 7; K16583 tubulin polyglutamylase TTLL7 [EC:6.---]  
hsa:93432  
maltase-glucoamylase (alpha-glucosidase) pseudogene  
hsa:83887  
TTLL2, C6orf104, NYD-TSPG, DJ366N23.3; tubulin tyrosine ligase-like family, member 2; K16600 tubulin polyglutamylase TTLL2 [EC:6.---]  
hsa:164395  
TTLL9, C20orf125, DJ310O13.1; tubulin tyrosine ligase-like family, member 9; K16603 tubulin polyglutamylase TTLL9 [EC:6.---]  
hsa:284076  
TTLL6, TTL.6; tubulin tyrosine ligase-like family, member 6; K16582 tubulin polyglutamylase TTLL6/13 [EC:6.---]  
hsa:440307  
TTLL13; tubulin tyrosine ligase-like family, member 13; K16582 tubulin polyglutamylase TTLL6/13 [EC:6.---]  
hsa:277  
AMY1B, AMY1; amylase, alpha 1B (salivary) (EC:3.2.1.1); K01176 alpha-amylase [EC:3.2.1.1]  
hsa:278  
AMY1C, AMY1; amylase, alpha 1C (salivary) (EC:3.2.1.1); K01176 alpha-amylase [EC:3.2.1.1]  
hsa:276  
AMY1A, AMY1; amylase, alpha 1A (salivary) (EC:3.2.1.1); K01176 alpha-amylase [EC:3.2.1.1]  
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]

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="#">BRITIE hierarchy</a>		
SSDB	<a href="#">Ortholog</a> <a href="#">Paralog</a> <a href="#">GFIT</a>		
Motif	Pfam: Alpha-amylase Alpha-amylase_C DUF3078 <a href="#">Motif</a>		
Other DBs	NCBI-GI: 56549662 NCBI-GenID: 277 OMIM: 104701 HGNC: 475 HPRD: 18510 Ensembl: ENSG00000174876 Vega: OTTHUMG00000011021 UniProt: P04745		
Structure	PDB: 3DHP 3BLP 1Z32 1SMD 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>	MKLFWLLFTIGFCWAQYSSNTQQGRTSIVHLFEWRWVDIALECERYLAPKGFGGVQVSPP NENVAHNPFPRPWERYQPVSYKLCTRSQNEDEFRNMTVRCCNNVGVRIVYDAVINHMCGN AVSAGTSSTCGSYFNPGSRDFPAVPYSQGWDFNDGKCKTGSGDIENYNDATQVRDCRLSGL LDLALGKDYVRSKIAEYMNHLIDIGVAGFRIDASKHMWPGDIKAILDKLHNLNLSNWFPPEG SKFIYIQLVIDLGEPISKSDYFGNGRVTEFKYGAKLGTWIRKWNNGEKMSYLNWNGEWG FMPSDRALVFVDNHDNQRGHGAGGASILTFWDARLYKMAVGFMLAHPYGFTRVMSSYRWP RYFENGKDVNDWVGPNDNGVTKEVTPDTTCGNDWVCEHRWRQIRNMVNFRNVDDQGP FTNWYDNGSNQVAFGRGNRGFIVFNNDWTFSLTLQTGLPAGTYCDVISGDKINGNCTGI KIYVSSDGKAHSISNSAEDPFIAIHAESKL	
NT seq	1536 nt <a href="#">NT seq</a>	atgaagctctttgggtgccttcaccattgggtctgctgggctcagttcccaa acacaacaaggacgaaactcttgcacatcgatggcgatgggttatattgc cttgaatgtggcgatatttagtcctccaaggatggaggggttcaggctctccacca	

# KEGG ORTHOLOGYのエントリ



## ORTHOLOGY: K01176

[Help](#)

<b>Entry</b>	K01176	KO
<b>Name</b>	E3.2.1.1, amyA, malS	
<b>Definition</b>	alpha-amylase [EC:3.2.1.1]	
<b>Pathway</b>	ko00500 Starch and sucrose metabolism ko04973 Carbohydrate digestion and absorption	
<b>Brite</b>	KEGG Orthology (KO) [BR:ko00001] Metabolism Carbohydrate metabolism 00500 Starch and sucrose metabolism K01176 E3.2.1.1, amyA, malS; alpha-amylase Organismal Systems Digestive system 04973 Carbohydrate digestion and absorption K01176 E3.2.1.1, amyA, malS; alpha-amylase Enzymes [BR:ko01000] 3. Hydrolases 3.2 Glycosylases 3.2.1 Glycosidases, i.e. enzymes that hydrolyse O- and S-glycosyl 3.2.1.1 alpha-amylase K01176 E3.2.1.1, amyA, malS; alpha-amylase <a href="#">BRITE hierarchy</a>	
<b>Other DBs</b>	RN: R02108 COG: COG0366 GO: 0004556	
<b>Genes</b>	HSA: 276(AMY1A) 277(AMY1B) 278(AMY1C) 279(AMY2A) 280(AMY2B) PTR: 457067(AMY2B) 469397(AMY1A) 736558(AMY2A) PPS: 100980682 100983088(AMY2B) 100983778(AMY2A) PON: 100446394(AMY1A) 100447008 MCC: 712893(AMY2B) 713027(AMY2A) MMU: 100043684(Amy2a4) 100043686(Amy2a3) 100043688(Amy2a2) 109959(Amy2a5) 11722(Amy1) RNO: 24203(Amy1a) 497039(Amy2) CFA: 480825 607460	

### All links

- Ontology (8)
  - KEGG BRITE (6)
  - GO (1)
  - COG (1)
- Pathway (6)
  - KEGG PATHWAY (6)
- Drug (1)
  - KEGG DRUG (1)
- Chemical reaction (2)
  - KEGG ENZYME (1)
  - KEGG REACTION (1)
- Gene (3346)
  - KEGG GENES (994)
  - KEGG DGENES (23)
  - KEGG EGENES (319)
  - KEGG MGENES (2010)
- Protein sequence (847)
  - UniProt (847)
- Literature (2)
  - PubMed (2)
- All databases (4212)

[Download RDF](#)

# KAAS: <http://www.genome.jp/tools/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.

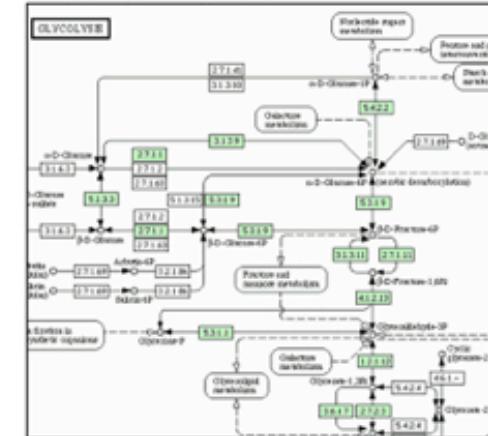
- [KAAS job request \(BBH method\)](#)
- [KAAS job request \(SBH method\)](#)

### Example of Results

#### KO assignment

KO Assignment Results	
Home	
[KO list] [KEGG Inquiries] [Pathway map] [Threshold change] [Download]	
Query gene : KO assignment	
query_0001	ko00001
query_0002	ko00003
query_0003	ko00873
query_0004	ko1733
query_0005	
query_0006	
query_0007	ko03010
query_0008	ko00616
query_0009	ko03033
query_0010	ko07934
query_0011	
query_0012	
query_0013	
query_0014	ko04043
query_0015	ko30986
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| chaperone clpB [Gluconacetobacter hansenii ATCC 23769]  
MNIEKFTERSRGFLQAAQTIAMREYNQQLTPEHLLKALLDDDQGAASALIRAAGGQPPAIAAVDTALAK  
LPKVQGGGAGQPSATPDLVRLDAAEQAAQKAGDEYVAQDRLLAAIAASETPAGQALRAGGATPQALDKA  
IATIRKGRTVTSENAAEASFDALKKYARDVTEIALQGKLDPVIGRDEEIRRAIQVLARRSKN  
>gi|296114122|ref|ZP_06832777.1| hypothetical protein GXY_00019 [Gluconacetobacter hansenii ATCC 23769]  
MGAALIMITTLIAGYALCEWPPGEVTPALMQIK
```

## TCAサイクル

 KAAS

**KO Assignment Results**

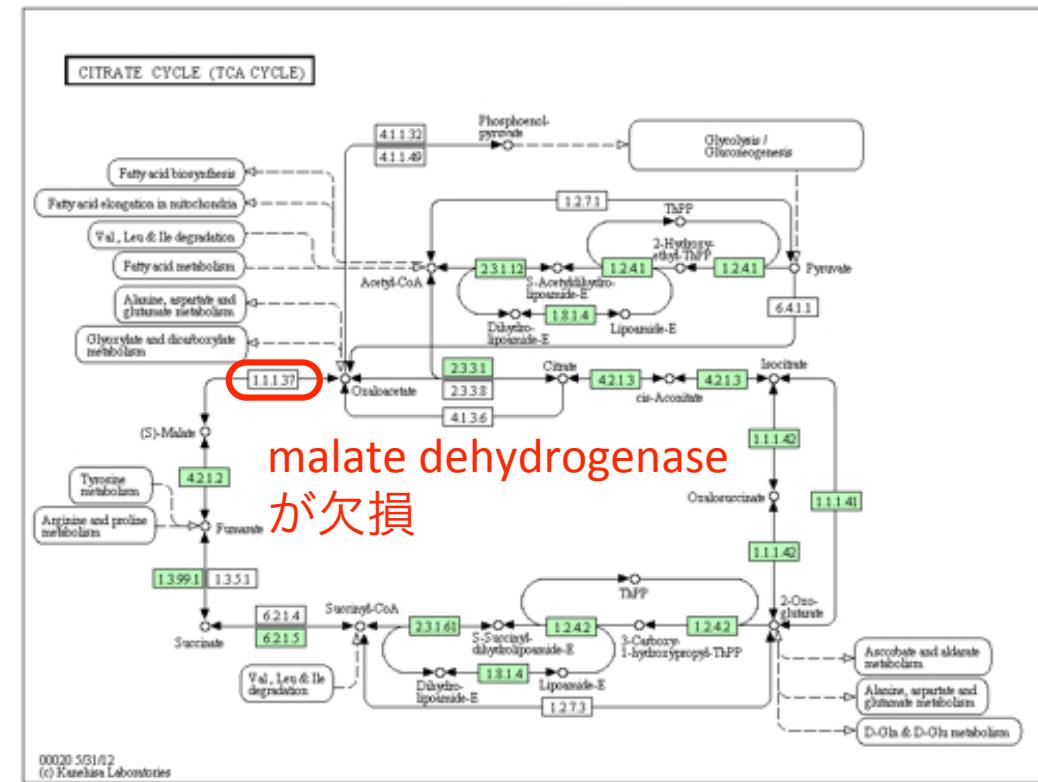
**Home**

[KO list] [BRITE hierarchies] [Pathway map] [Threshold change] [Download]

**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

 **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

**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
LIGAND	COMPOUND	C number	Chemical compound structures <a href="#">KEGG COMPOUND</a>
	GLYCAN	G number	Glycan structures <a href="#">KEGG GLYCAN</a>
	REACTION	R number	Biochemical reactions
	RPAIR	RP number	Reactant pair alignments
	RCLASS	RC number	Reaction class
	ENZYME	EC number	Enzyme nomenclature <a href="#">KEGG REACTION</a>

Search  for     
 bfind mode  bget mode

# 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-Adenyl-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-Adenyl-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; Vernamycin B alpha; Mikamycin B; 4-[4-(Dimethylamino)-N-methyl-L-phenylalanine]virginiamycin S1; Streptogramin B

C11616 KegDraw Jmol  
Pristinamycin IB; Vernamycin B beta; Efepristin; 4-[N-Methyl-4-(methylamino)-L-phenylalanine]virginiamycin S1

C11617 KegDraw Jmol  
Pristinamycin IC; Vernamycin 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="#">ko04974</a> Protein digestion and absorption <a href="#">ko04978</a> Mineral absorption	<b>Enzyme</b> <table border="1"> <tr> <td><a href="#">1.4.1.20</a></td> <td><a href="#">1.4.3.2</a></td> <td><a href="#">1.11.1.7</a></td> <td><a href="#">1.11.1.21</a></td> </tr> <tr> <td><a href="#">1.13.12.9</a></td> <td><a href="#">1.14.13.124</a></td> <td><a href="#">1.14.16.1</a></td> <td><a href="#">2.3.1.53</a></td> </tr> <tr> <td><a href="#">2.6.1.1</a></td> <td><a href="#">2.6.1.5</a></td> <td><a href="#">2.6.1.9</a></td> <td><a href="#">2.6.1.28</a></td> </tr> <tr> <td><a href="#">2.6.1.57</a></td> <td><a href="#">2.6.1.58</a></td> <td><a href="#">2.6.1.64</a></td> <td><a href="#">2.6.1.70</a></td> </tr> <tr> <td><a href="#">2.7.7.54</a></td> <td><a href="#">3.4.17.23</a></td> <td><a href="#">4.1.1.28</a></td> <td><a href="#">4.1.1.53</a></td> </tr> <tr> <td><a href="#">4.2.1.51</a></td> <td><a href="#">4.2.1.91</a></td> <td><a href="#">4.3.1.24</a></td> <td><a href="#">4.3.1.25</a></td> </tr> <tr> <td><a href="#">5.1.1.11</a></td> <td><a href="#">6.1.1.20</a></td> <td></td> <td></td> </tr> </table> <b>Other DBs</b> <table border="1"> <tr> <td>CAS: 63-91-2</td> </tr> <tr> <td>PubChem: <a href="#">3379</a></td> </tr> <tr> <td>ChEBI: <a href="#">17295</a></td> </tr> <tr> <td>KNAPSAck: <a href="#">C00001386</a></td> </tr> <tr> <td>PDB-CCD: <a href="#">PHE</a></td> </tr> <tr> <td>3DMET: <a href="#">B01151</a></td> </tr> <tr> <td>NIKKAJI: <a href="#">J9.175H</a></td> </tr> </table> <b>KCF data</b> <table border="1"> <tr> <td><a href="#">Show</a></td> </tr> </table>	<a href="#">1.4.1.20</a>	<a href="#">1.4.3.2</a>	<a href="#">1.11.1.7</a>	<a href="#">1.11.1.21</a>	<a href="#">1.13.12.9</a>	<a href="#">1.14.13.124</a>	<a href="#">1.14.16.1</a>	<a href="#">2.3.1.53</a>	<a href="#">2.6.1.1</a>	<a href="#">2.6.1.5</a>	<a href="#">2.6.1.9</a>	<a href="#">2.6.1.28</a>	<a href="#">2.6.1.57</a>	<a href="#">2.6.1.58</a>	<a href="#">2.6.1.64</a>	<a href="#">2.6.1.70</a>	<a href="#">2.7.7.54</a>	<a href="#">3.4.17.23</a>	<a href="#">4.1.1.28</a>	<a href="#">4.1.1.53</a>	<a href="#">4.2.1.51</a>	<a href="#">4.2.1.91</a>	<a href="#">4.3.1.24</a>	<a href="#">4.3.1.25</a>	<a href="#">5.1.1.11</a>	<a href="#">6.1.1.20</a>			CAS: 63-91-2	PubChem: <a href="#">3379</a>	ChEBI: <a href="#">17295</a>	KNAPSAck: <a href="#">C00001386</a>	PDB-CCD: <a href="#">PHE</a>	3DMET: <a href="#">B01151</a>	NIKKAJI: <a href="#">J9.175H</a>	<a href="#">Show</a>
<a href="#">1.4.1.20</a>	<a href="#">1.4.3.2</a>	<a href="#">1.11.1.7</a>	<a href="#">1.11.1.21</a>																																			
<a href="#">1.13.12.9</a>	<a href="#">1.14.13.124</a>	<a href="#">1.14.16.1</a>	<a href="#">2.3.1.53</a>																																			
<a href="#">2.6.1.1</a>	<a href="#">2.6.1.5</a>	<a href="#">2.6.1.9</a>	<a href="#">2.6.1.28</a>																																			
<a href="#">2.6.1.57</a>	<a href="#">2.6.1.58</a>	<a href="#">2.6.1.64</a>	<a href="#">2.6.1.70</a>																																			
<a href="#">2.7.7.54</a>	<a href="#">3.4.17.23</a>	<a href="#">4.1.1.28</a>	<a href="#">4.1.1.53</a>																																			
<a href="#">4.2.1.51</a>	<a href="#">4.2.1.91</a>	<a href="#">4.3.1.24</a>	<a href="#">4.3.1.25</a>																																			
<a href="#">5.1.1.11</a>	<a href="#">6.1.1.20</a>																																					
CAS: 63-91-2																																						
PubChem: <a href="#">3379</a>																																						
ChEBI: <a href="#">17295</a>																																						
KNAPSAck: <a href="#">C00001386</a>																																						
PDB-CCD: <a href="#">PHE</a>																																						
3DMET: <a href="#">B01151</a>																																						
NIKKAJI: <a href="#">J9.175H</a>																																						
<a href="#">Show</a>																																						

# KEGG REACTIONのエントリ



REACTION: R00692

Help

Entry	R00692	Reaction
Name	L-phenylalanine:pyruvate aminotransferase	
Definition	L-Phenylalanine + Pyruvate <=> Phenylpyruvate + L-Alanine	
Equation	$C00079 + C00022 \rightleftharpoons C00166 + C00041$ <p>The diagram illustrates the reversible reaction catalyzed by L-phenylalanine:pyruvate aminotransferase. It shows four chemical structures: L-phenylalanine (C00079), pyruvate (C00022), Phenylpyruvate (C00166), and L-alanine (C00041). A double-headed red arrow connects L-phenylalanine and Phenylpyruvate. Below these, another double-headed red arrow connects pyruvate and L-alanine. The structures are arranged with L-phenylalanine and Phenylpyruvate on top, and pyruvate and L-alanine below them.</p>	
RPair	<a href="#">RP00021</a> C00022_C00041 main <a href="#">RP00057</a> C00079_C00166 main <a href="#">RP06279</a> C00022_C00166 trans <a href="#">RP06280</a> C00041_C00079 trans	
Enzyme	<a href="#">2.6.1.58</a>	
Pathway	<a href="#">rn00360</a> Phenylalanine metabolism <a href="#">rn00400</a> Phenylalanine, tyrosine and tryptophan biosynthesis <a href="#">rn00960</a> Tropane, piperidine and pyridine alkaloid biosynthesis <a href="#">rn01100</a> Metabolic pathways <a href="#">rn01110</a> Biosynthesis of secondary metabolites	
Orthology	<a href="#">K05821</a> 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. On the left, there's a navigation tree. The first level includes '0. Global Map' and '1. Metabolism'. Under '1. Metabolism', '1.1 Carbohydrate Metabolism' is expanded, showing a list of 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. This list is highlighted with a red rectangle. Below it is '1.2 Energy Metabolism' with entries: Oxidative phosphorylation, Photosynthesis, and Photosynthesis - antenna proteins. To the right, there are three columns of links: 'Metabolic pathways [zoom out]', 'Biosynthesis of secondary metabolites [zoom out]', and 'Microbial metabolism in diverse environments [zoom out]'. Each has a 'Launch KEGG Atlas' link next to it. Further down, there are links for 'Enzymes' and 'Compounds with biological roles', and at the bottom, 'Photosynthesis proteins'.

- 0. Global Map**
- 0.1 Metabolism**
  - [Metabolic pathways \[zoom out\]](#) [Launch KEGG Atlas](#)
  - [Biosynthesis of secondary metabolites \[zoom out\]](#) [Launch KEGG Atlas](#)
  - [Microbial metabolism in diverse environments \[zoom out\]](#) [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
  - 1.2 Energy Metabolism**
    - Oxidative phosphorylation
    - Photosynthesis
    - Photosynthesis - antenna proteins

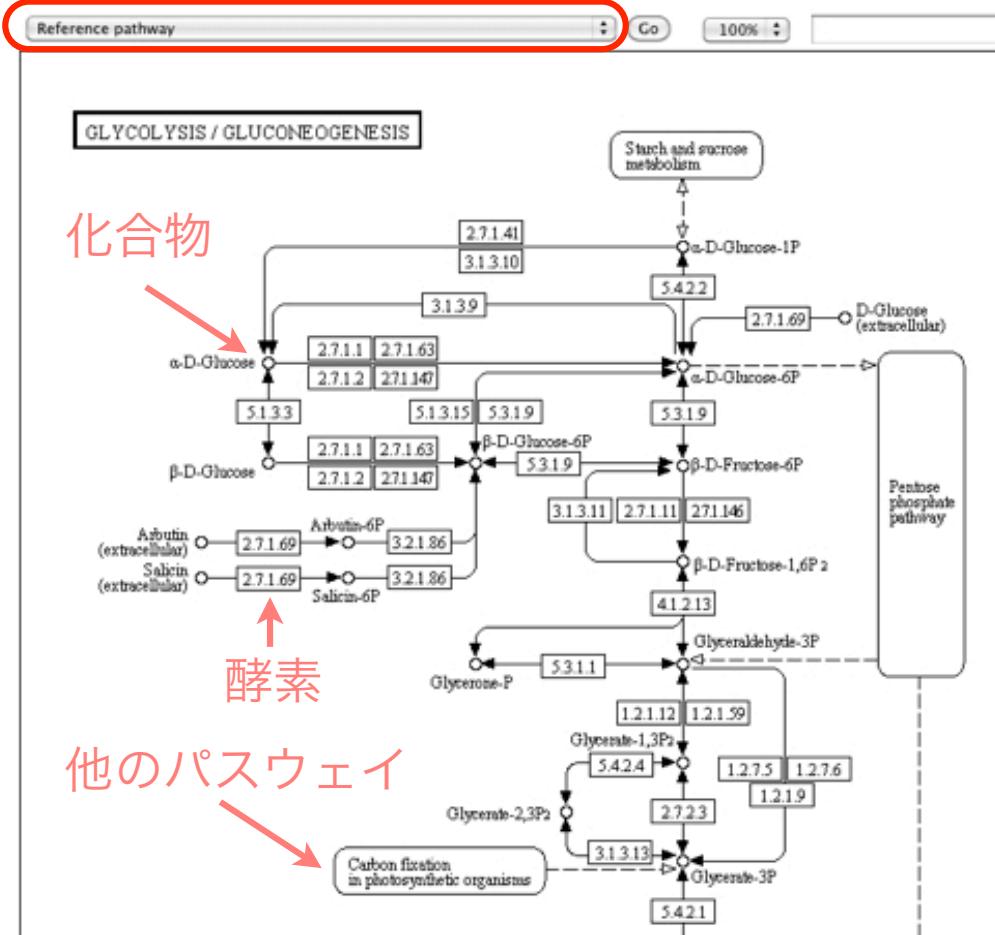
# KEGG PATHWAY MAP



Glycolysis / Gluconeogenesis - Reference pathway

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

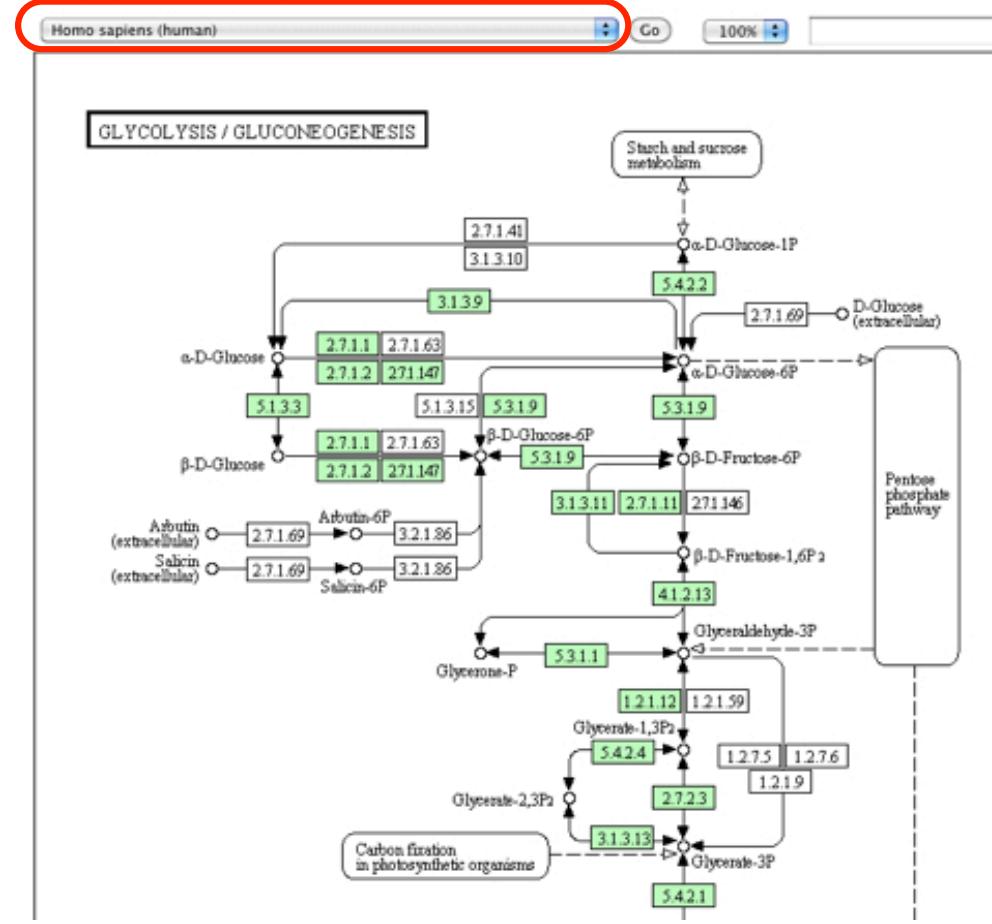
Glycolysis is the process of converting glucose into pyruvate and generating small amounts of ATP (energy) and NADH (reducing power). It is a central pathway that produces important precursor metabolites: six-carbon compounds of glucose-6P and fructose-6P and three-carbon compounds of glyceraldehyde-3P, glycerate-3P, phosphoenolpyruvate, and pyruvate [MD:M00001]. Acetyl-CoA, another important precursor metabolite, is produced by oxidative decarboxylation of pyruvate [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:M00002], which is found in almost all organisms and which sometimes contains operon structures in bacterial genomes. Gluconeogenesis is a synthesis pathway glucose from noncarbohydrate precursors. It is essentially a reversal of glycolysis with minor variations of alternative paths [MD:M00003].



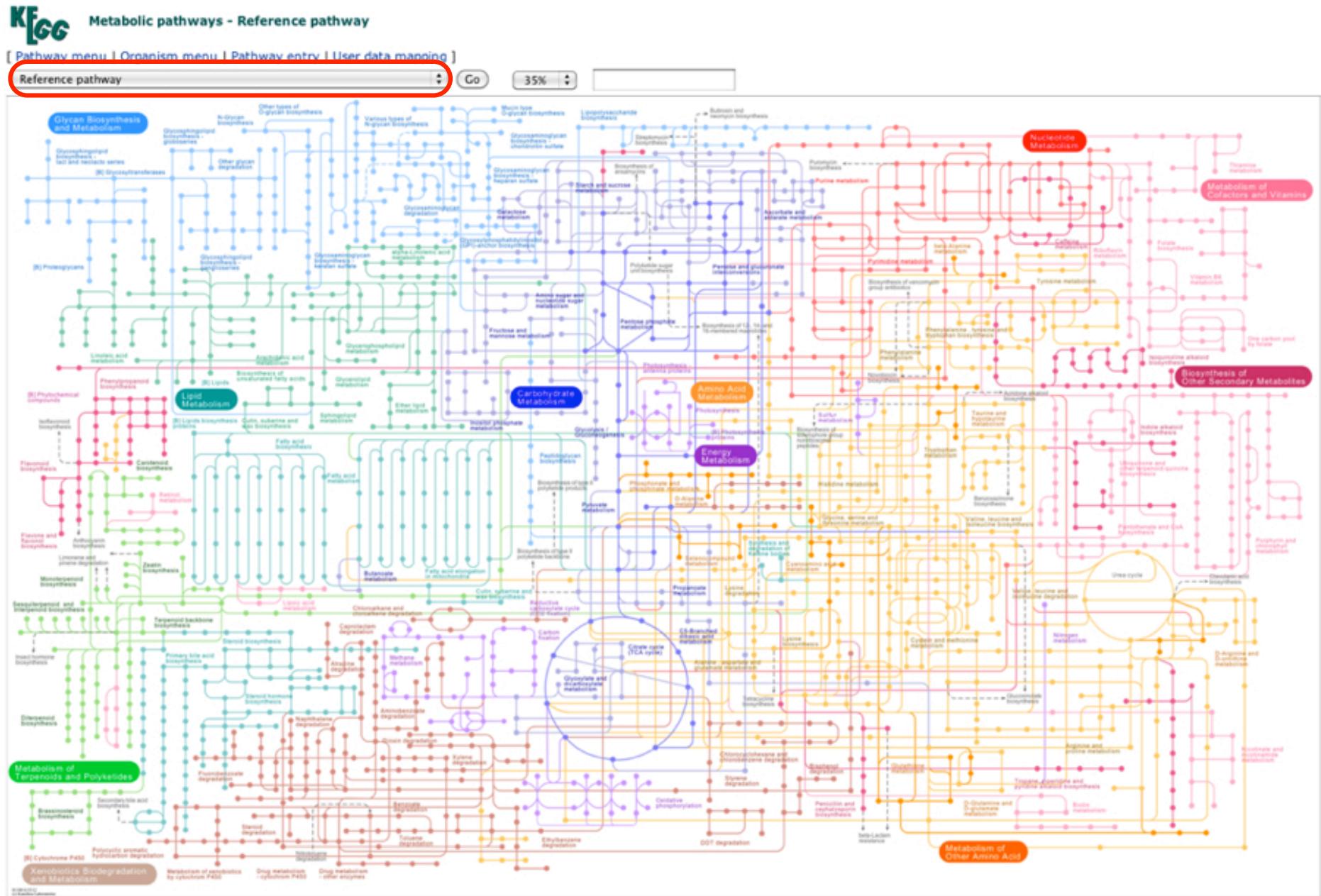
Glycolysis / Gluconeogenesis - Homo sapiens (human)

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

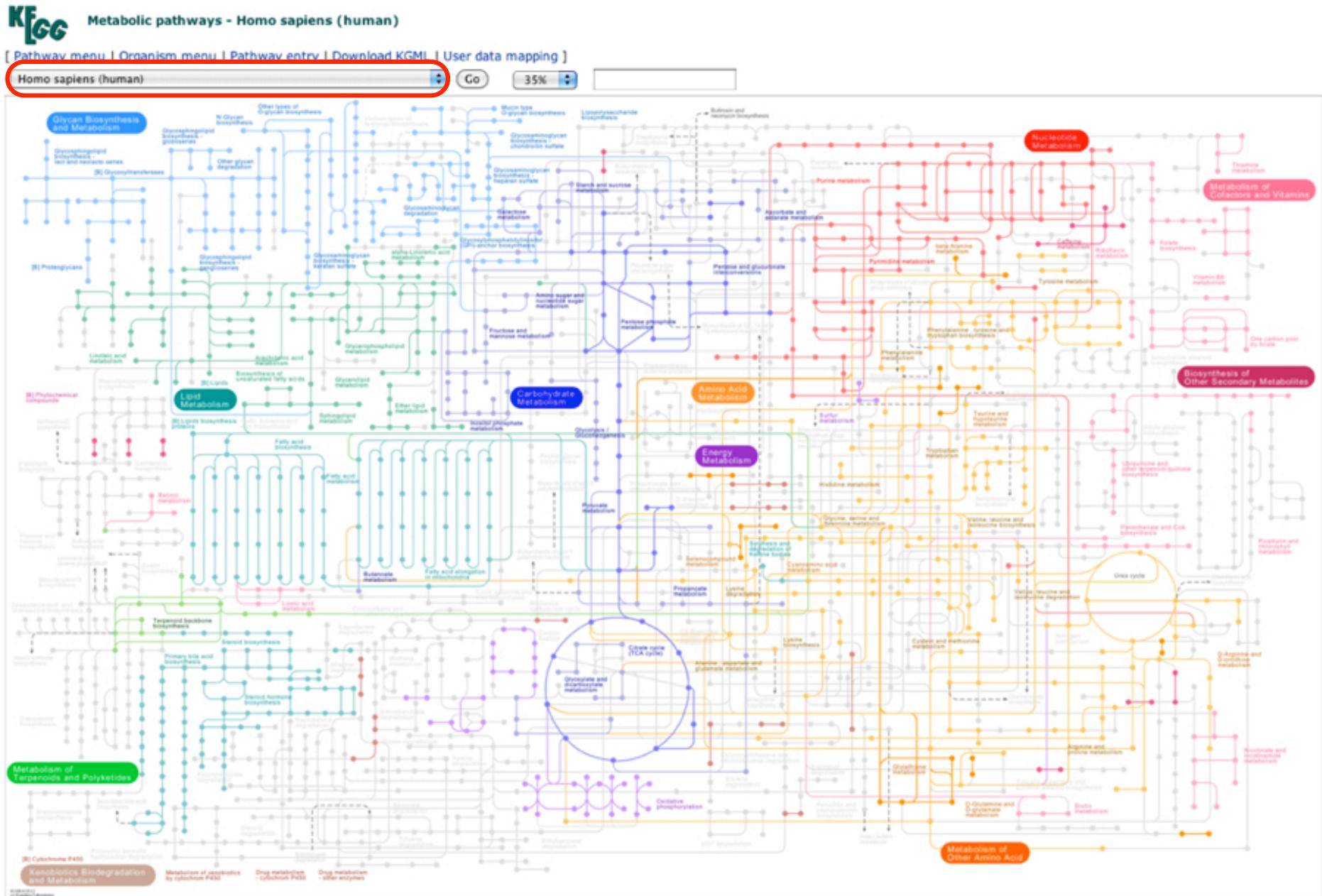
Glycolysis is the process of converting glucose into pyruvate and generating small amounts of ATP (energy) and NADH (reducing power). It is a central pathway that produces important precursor metabolites: six-carbon compounds of glucose-6P and fructose-6P and three-carbon compounds of glyceraldehyde-3P, glycerate-3P, phosphoenolpyruvate, and pyruvate [MD:M00001]. Acetyl-CoA, another important precursor metabolite, is produced by oxidative decarboxylation of pyruvate [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:M00002], which is found in almost all organisms and which sometimes contains operon structures in bacterial genomes. Gluconeogenesis is a synthesis pathway glucose from noncarbohydrate precursors. It is essentially a reversal of glycolysis with minor variations of alternative paths [MD:M00003].



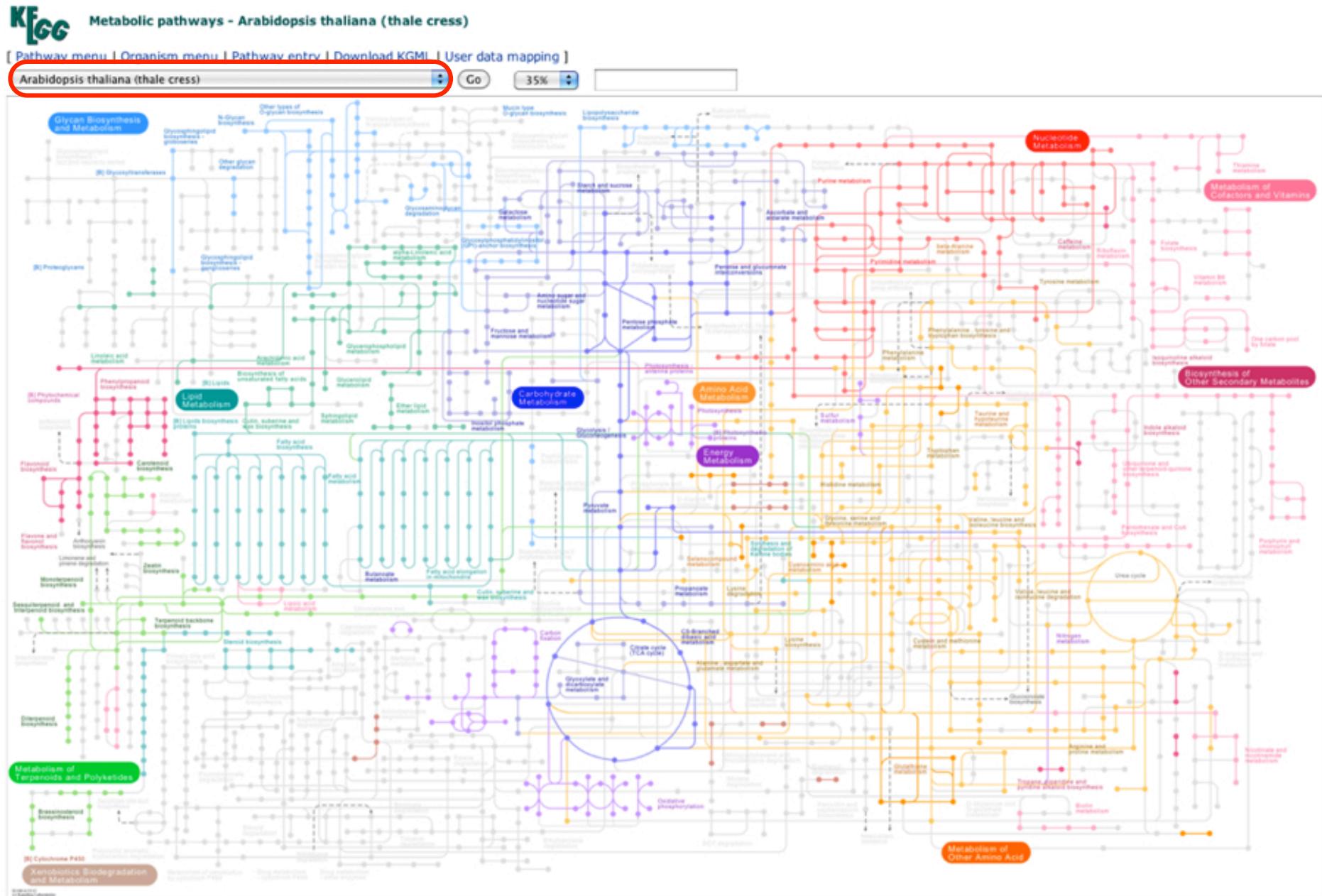
# Global Map



# 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    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:  
 **hsa\_CML-COSMIC.txt**

Numerical value example: Somatic mutations  
**hsa05200 + CML-COSMIC**

**1. 右クリックで保存**

File type:  Color specification  
 Numerical value ( Convert to log scale)

Gradation: from **#ffffff** 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**

#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

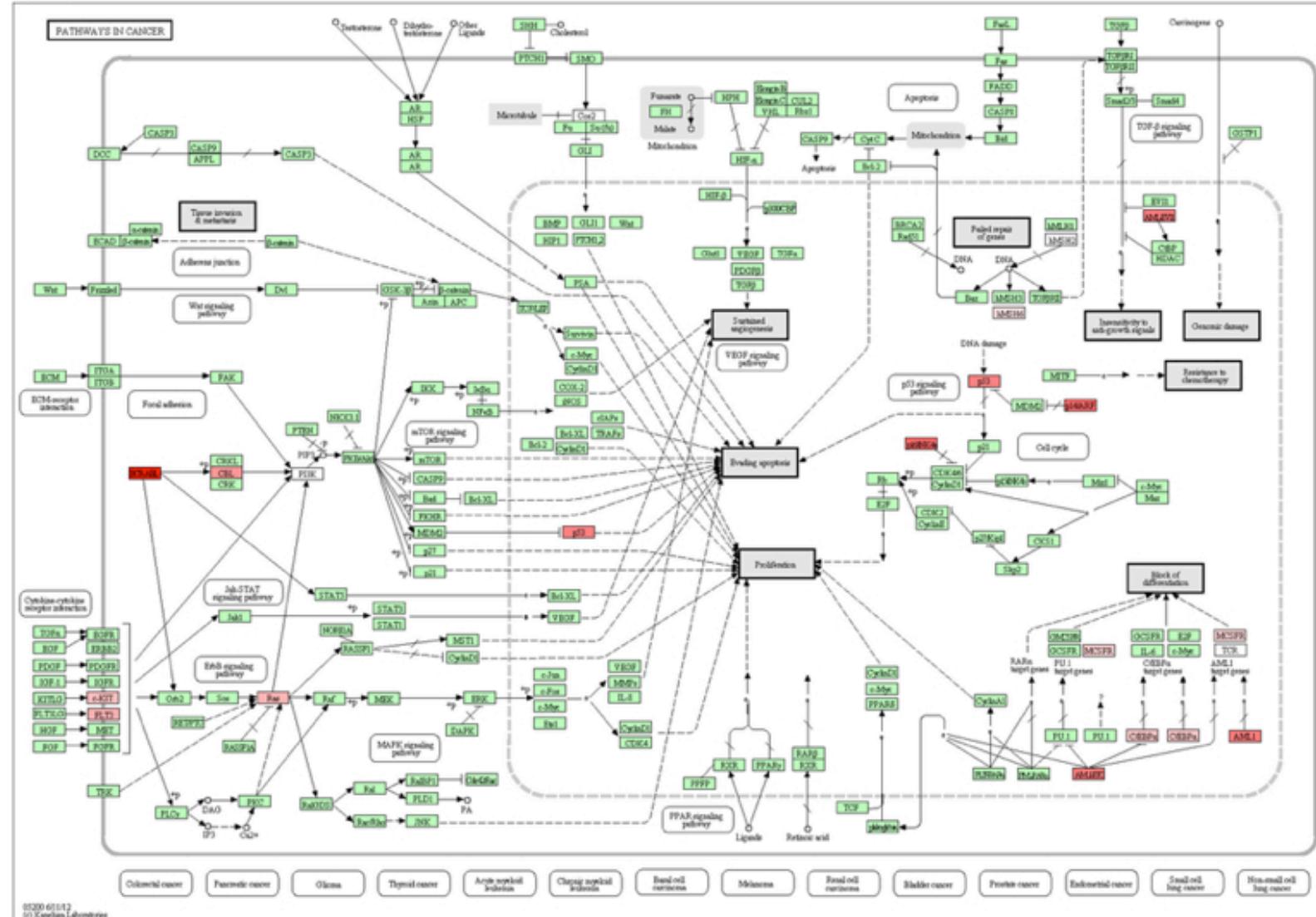
# KEGG Mapperの結果

**KEGG** Pathways in cancer - Homo sapiens (human)

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

Homo sapiens (human) Go 55% [ ]

COSMIC



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