

統合データベース講習会 : AJACS筑波2  
2012年8月6日

# 国内の主要なDBの使い方 (DDBJ, PDBj, KEGG)

---

バイオサイエンスデータベースセンター  
(NBDC)

三橋 信孝

# 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、何度もクリックすることはやめましょう。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。

# DBやツールを使い倒そう: 統合TV



- ・統合TVは、生命科学分野の有用なデータベースやウェブツールの活用法を動画で紹介するウェブサイトです
- ・2007年の開始から通算で600本以上の動画を制作公開しました
- ・統合プロジェクトの講演動画、講習会動画もここから公開されています
- ・DBCLSのリサーチアシスタントが非常に大きな戦力です

# 代表的な生命科学関係のデータベース

データベースの内容	データベース例
DNA塩基配列	GenBank, EMBL, DDBJ, UCSC
アミノ酸配列	Swiss-Prot, PIR, UniProt
タンパク質立体構造	PDB, SCOP, CATH
アミノ酸配列ドメイン	Pfam
アミノ酸配列モチーフ	PROSITE, BLOCKS
パスウェイ	KEGG, Reactome
遺伝病	OMIM
文献	MEDLINE
遺伝子発現	GEO

# 生命科学データベース統合に関する調査研究(H17~H19)

## 構築法から見たデータベースの分類

型	情報源の種類	処理方法	処理主体	データ形式	例
バンク型	測定器と登録者		不特定多数	構造化テキスト	DDBJ
プロジェクト型	測定器と実験者		特定人間	構造	Fantom, モデル生物DB
プログラム型	データベースレコード	機械的処理	マシン	構造	UniGene
キュレーション型	データベースレコード	高度情報処理	特定人間	構造	SCOP
知識モデル型	読み物	高度情報処理	特定人間	構造	KEGG
総説型	読み物	高度情報処理	特定人間	構造化テキスト	OMIM

# 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)

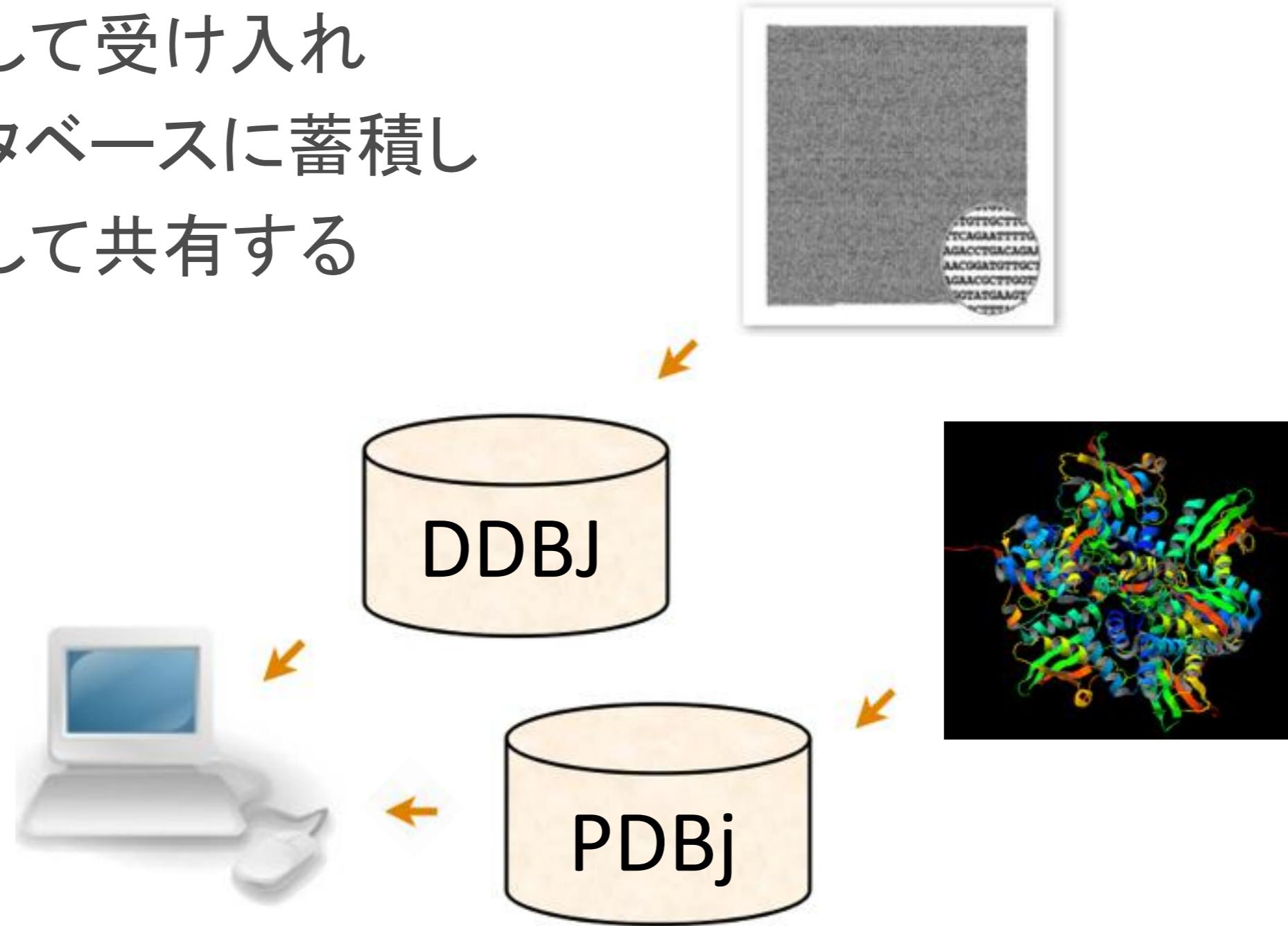
- タンパク質の立体構造を収集
- 大阪大学蛋白質研究所

## ◆ 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



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

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

Search articles...

GO

Advanced Search

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

[Home](#) [Browse Articles](#) [About](#) [For Readers](#) [For Authors and Reviewers](#)

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

## Accession Numbers

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

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

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

# 生命科学のデータベースの歴史

## アミノ酸配列データベース

- ・ 1968年 デイホフAtlas of Protein Sequence and Structure
- ・ 1980年 NBRFデータベース
- ・ 1984年 PIR (Protein Identification Resource) データベース
- ・ 2005年 UniProt (PIR+Swiss-Prot+TrEMBL)データベース

## 塩基配列データベース

- ・ 1982年 米国DNAデータベース GenBank ロスアラモス国立研究所
- ・ 1982年 欧州EMBL データベース 欧州分子生物学研究所
- ・ 1984年 日本 DDBJ(DNA Data Bank of Japan) 遺伝学研究所
- ・ 1992年 GenBank ロスアラモスから NCBI へ移管
- ・ 1994年 EMBL 本部から英国の EBI へ移転

## 立体構造データベース

- ・ 1971年 Protein Data Bank (PDB)米国ブルックヘブン国立研究所
- ・ 1999年 PDB Research Collaboratory for Structural Bioinformatics (RCSB) に移管

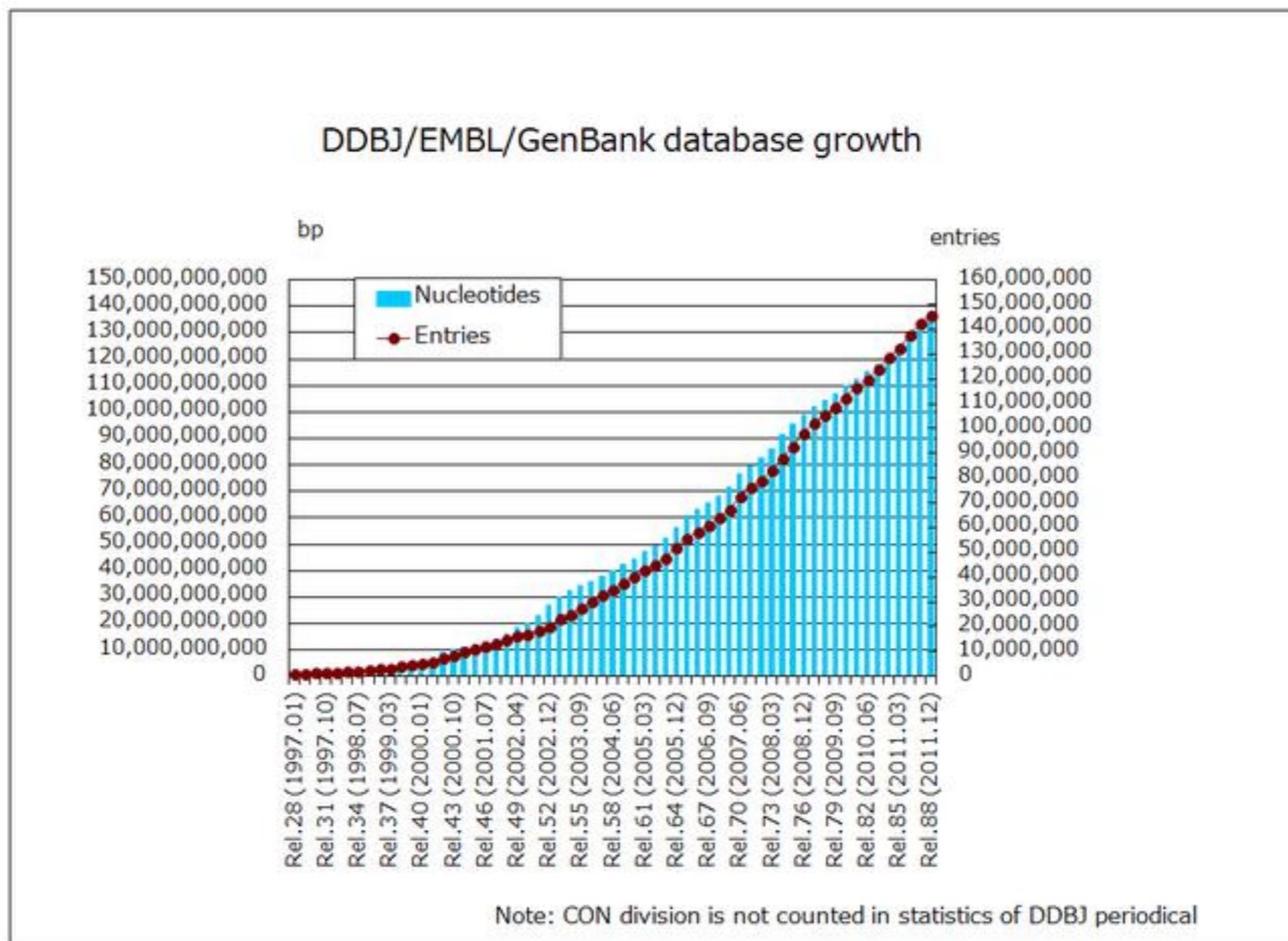
## 文献データベース

- ・ 1879年 Index Medicus創刊
- ・ 1926年 Biological Abstracts 米国フィラデルフィアの非営利組織 BIOSIS
- ・ 1907年 Chemical Abstracts 米国化学会
- ・ 1971年 MEDLINE オンラインサービス
- ・ 1996年 PubMed公開

# 現在の塩基配列データの量

塩基数 : 1,400億

登録数 : 1.5億



# 世界中で進行するゲノム・メタゲノムプロジェクト

**GOLD**  
Genomes Online Database

Last update: 2012-07-24  
Total # of genomes: **17246**

[Home](#)

Home  
Genome Map  
Genome Earth  
Search  
News  
Statistics  
Team  
Reference  
Contact

[f](#) [W](#) [Blogger](#)

**JGI**  
DOE JOINT GENOME INSTITUTE  
US DEPARTMENT OF ENERGY  
OFFICE OF SCIENCE

**Version 4.0**

**Welcome to the Genomes OnLine Database**

**GOLD:** Genomes Online Database, is a World Wide Web resource for comprehensive access to information regarding genome and metagenome sequencing projects, and their associated metadata, around the world.

**Metagenomes**

[Classification](#)  
• [Studies: 335](#)  
• [Samples: 2053](#)

**Isolate Genomes**

[Complete Projects: 3402](#)  
 [Incomplete Projects: 13796](#)  
 [Targeted Projects: 1814](#)

**Genome Distribution**

- [Project Type](#)
- [Sequencing Status](#)
- [Phylogenetic](#)

**1. Register**

Register your project information and Metadata in Genomes Online Database

[Register](#)

**2. Annotate**

Annotate your microbial genome or metagenome with IMG/ER or IMG/MER

[Annotate](#)

**3. Publish**

Publish your genome or metagenome in open access standards-supportive journal.

[Publish](#)

©2012 The Regents of the University of California  
[Disclaimer](#) | [Credits](#)

U.S. DEPARTMENT OF ENERGY | Office of Science

Genomes OnLine Database <http://www.genomesonline.org/>

# 転換期を迎える生命科学

---

- ・ 生命科学の情報爆発
- ・ 仮説検証型からデータ駆動型の科学への転換
- ・ データ, データベースを自由に使いこなす
- ・ 情報学, 統計学, 数学, 計算機科学などの融合
- ・ データの統合による知識発見



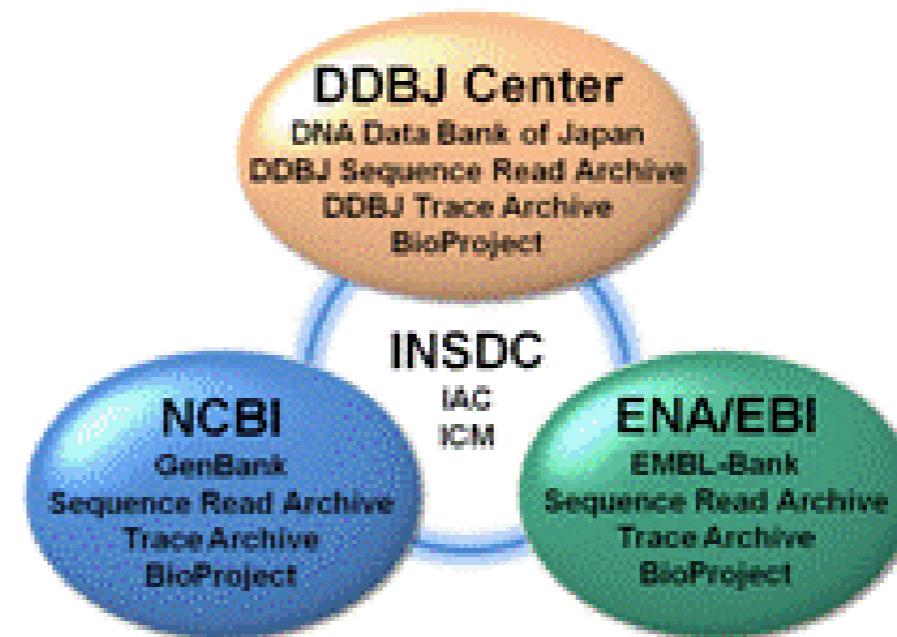
# DBJ

DNA Data Bank of Japan

# DDBJは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塩基配列も収集

# DDBJ登録ファイルの例

LOCUS AB091058 2109 bp DNA linear BCT 02-SEP-2003

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

ACCESSION [AB091058](#)

VERSION AB091058.1

KEYWORDS .

SOURCE Gluconacetobacter xylinus

ORGANISM [Gluconacetobacter xylinus](#)

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

REFERENCE 1 (bases 1 to 2109)

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  
[/db\\_xref="taxon:28448"](#)  
[/mol\\_type="genomic DNA"](#)  
[/note="synonym:Acetobacter xylinum"](#)  
[/organism="Gluconacetobacter xylinus"](#)  
[/strain="ATCC 53582"](#)

CDS 10..1038  
[/codon\\_start=1](#)  
[/gene="cmcase"](#)  
[/product="endo-beta-1,4-glucanase"](#)  
[/protein\\_id="BAC82540.1"](#)  
[/transl\\_table=11](#)  
[/translation="MSVMAAMGGAQVLSSTGAFADTAPDAVAQQWAIFRAKYLRPSGR  
VVDTGNGGESHSEGQGYGMLFAASAGDLASFQSMWMWARTNLQHTNDKLFSWRFLKGH  
QPPVPDKNNATDGDLIALALGRAGKRFQRPDIQDAMAIYGDVLNLMTMKAGPYVVL  
MPGAVGFTKKDSVILNLSYYVMPSLLQAFDLTADPRWRQVMEDGIRLVSAGRFGQWRL  
PPDWLAVNRATGALSIASGWPPRFSYDAIRVPLYFYWAHMLAPNVLADFTRFWNNFGA  
NALPGWVDLTTGARSPYNAPPGYLAVAECTGLDSAGELPTLDHAPDYYSAALTLLVYI  
ARAETIK"](#)

CDS

1035..2096

[/codon\\_start=1](#)

[/gene="ccp"](#)

[/product="cellulose complementing protein"](#)

[/protein\\_id="BAC82541.1"](#)

[/transl\\_table=11](#)

[/translation="MSASGSDEVAGGGQAGSPQDFQRVLRSFGVEGGQYSYRPFVDRS  
FDVTGVPEAVERHFDQAEHTAVEEQVTAPQIAVAPPPPVVPDPPAIVTETAPPPP  
VVVSAPVTYEPPAAAVPAEPPVQEAPVQAAPVPPAPVPPIAEQAPPAAPDPASVPYAN  
VAAAPVPPDPAPVTPAPQARVTGPNTRMVEPFSRPQVRTVQEGATPSRVSRSRMNAFP  
RTSASSISERPVDRGVADEWSPVPKARLSPRERPRPGDLSFFFQGMRDTRDEKKFFPV  
ASTRSVRNSVSRMTSMTKTDTNSSQASRPGSPVASPDGSPTMAEVFMTLGGRATELLS  
PRPLSLREALLRRRENEEES"](#)

BASE COUNT 343 a 661 c 661 g 444 t

ORIGIN

1 cgttccttta tgcgttcat ggccggcgatg ggaggggcgc aggtgcttc atccaccgg  
61 gcgttcgcag acaccggccc cgtatcggtc ggcgcacaaat gggccatctt ccgcgc  
121 tatcttcgtc ccagcggacg tgcgtggat acggcaatg gtggcgaatc ccata  
181 gggcaggct atggcatgct ctttgcgcg tcggcgaaaa accttgcgtc gttcc  
241 atgtggatgt gggcgccac caacctgcag cataccaatg acaagctgtt ttcc  
301 ttccctcaagg ggcattcagcc cccgggtgcc gacaagaaca atgc  
361 ctgatcgccg ttgcgttgg tgcgtgcggc aacgcatttc agcgc  
421 gacgcattgg ccatttatgg cgtatgtcgt aacctgatga cgtatgaaggc  
481 gtcgtcccta tgcccggtc tgcgtgcgtt accaagaagg acagcgtat  
541 tccttattacg tcatgcctc gtcgtgcag gcgttcgacc ttac  
601 cgtcaggta tggaagacgg gattgcgc  
661 ctgc  
721 tggccgcgc gctttctt  
781 atgc  
841 gccc  
901 ggat  
961 gatcat  
1021 gagg  
1081 gaagt  
1141 cctacc  
1201 ggcact  
1261 aaat  
1321 aaacc  
1381 cgc  
1441 cgc  
1501 tgc  
1561 cgc  
1621 agg  
1681 ctt  
1741 atgaat  
1801 atct  
1861 tgg  
1921 acac  
1981 ccaca  
2041 gtc  
2101 ctatattca

# DDBJ の塩基配列の登録について

The screenshot shows the DDBJ website with a navigation bar at the top. The main menu includes links for HOME, 塩基配列の登録 (highlighted in yellow), 利用の手引き, 検索・解析, FTP・WebAPI, レポート・統計, and お問い合わせ. There is also an ENGLISH link, a search bar, and a site search button.

The main content area is titled "塩基配列の登録". A blue banner at the top of this section reads "「DDBJ サービスの状況について(4/20時点での新システムへの移行後の状況)」". Below this, there is a large block of Japanese text explaining the service status, mentioning the International Nucleotide Sequence Database (INSD) and the transition to a new system.

On the left side, there is a sidebar with several sections:

- 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
  - ▶ TPAの登録
  - ▶ WGSの登録
  - ▶ TSAの登録
  - ▶ DDBJ Sequence Read Archive
  - ▶ DDBJ Trace Archive
- プロジェクトの登録**
  - ▶ DDBJ BioProject Database
- 塩基配列登録の前に**
  - ▶ 受け入れ基準
  - ▶ データの遷移
  - ▶ データ公開原則
  - ▶ 登録データの取り扱い
  - ▶ 用語解説
  - ▶ 特許出願関連の注意

Below the sidebar, there is a section titled "塩基配列データの登録に必要な情報" which lists several links:

- 塩基配列登録の前に
- DDBJに登録可能なデータ
- 登録データの取り扱い
- 塩基配列登録の流れ
- データ登録に関するFAQ
- DDBJのデータ公開形式 (flat file) の説明
- The DDBJ/EMBL/GenBank Feature Table Definition

# アクセション番号のプレフィックスの意味



JAPANESE [Twitter icon](#) [RSS icon](#)

Site Search

HOME Submission How to use Search and Analysis FTP and WebAPI Report/Statistics Contact us

HOME > Nucleotide Sequence Submission > Prefix Letter List

▶ About DDBJ  
▶ How to Use  
▶ Q and A

■ Sequence Submission  
▶ SAKURA  
▶ Mass Submission  
▶ Data Updates  
▶ Submission for TPA  
▶ Submission for WGS  
▶ Submission for TSA  
▶ DDBJ Sequence Read Archive  
▶ DDBJ Trace Archive

■ Project Registration  
▶ DDBJ BioProject Database

■ Before Submission  
▶ Acceptable Data  
▶ Data Transition  
▶ Principle of Data Release  
▶ INSD Policies  
▶ Description of Terms

## Prefix Letter List

Status July 3, 2012

- [For General Accession Number](#)
- [For WGS Accession Number](#)
- [For Protein\\_id](#)
- [For Sequence Read Archive](#), see also [DRA metadata outline](#)
- [For BioProject](#)

### For General Accession Number

A	EPO	Patent
AA	GenBank	EST
AB	DDBJ	Direct submission
AC	GenBank	HTG
AD	GenBank	Sequences received at GSDB
AE	GenBank	Genome project data
AF	GenBank	Direct submission
AG	DDBJ	GSS
AH	GenBank	CON
AI	GenBank	Other project
AJ	EMBL-Bank	Direct submission
AK	DDBJ	FLI_cDNA
AL	EMBL-Bank	Genome project data
AM	EMBL-Bank	Direct submission
AN	EMBL-Bank	CON
AP	DDBJ	Genome project data
AQ	GenBank	GSS
AR	USPTO	Patent
AS	GenBank	Other project

# DDBJ のデータ公開形式 (flat file) の説明

The screenshot shows the DDBJ website's 'Flat File' section. The top navigation bar includes links for HOME, 塩基配列の登録 (highlighted in yellow), 利用の手引き, 検索・解析, FTP・WebAPI, レポート・統計, and お問い合わせ. The right side features ENGLISH, Twitter, and RSS icons, along with a search bar and a 'サイト内検索' button. The main content area has a sidebar with links for DDBJ紹介, Q&A集, 塩基配列の登録 (SAKURA, MSS, Data修正・更新, TPA登録, WGS登録, TSA登録, DDBJ Sequence Read Archive, DDBJ Trace Archive), プロジェクトの登録 (DDBJ BioProject Database), and 塩基配列登録の前に (受け入れ基準, データの遷移, データ公開原則, 登録データの取り扱い, 用語解説, 特許出願関連の注意). The main content box is titled 'DDBJ のデータ公開形式 (flat file) の説明'. It explains that the International Nucleotide Sequence Database (DDBJ/EMBL-Bank/GenBank) is a collection of experimental data from researchers worldwide, following specific data structure rules. It also mentions that data from the Japan Patent Office (JPO), European Patent Office (EPO), United States Patent and Trademark Office (USPTO), and Korean Intellectual Property Office (KIPO) is included. The flat file format is described as a collection of entries, each following a specific schema. A note indicates that email addresses and phone numbers are no longer displayed. Below this, a sample flat file entry is shown:

LOCUS	AB000000	450 bp	mRNA	linear	HUM 01-JUN-2009
DEFINITION	Homo sapiens GAPD mRNA for glyceraldehyde-3-phosphate dehydrogenase, partial cds.				
ACCESSION	AB000000				
VERSION	AB000000.1				
KEYWORDS	.				
SOURCE	Homo sapiens (human)				
ORGANISM	Homo sapiens				

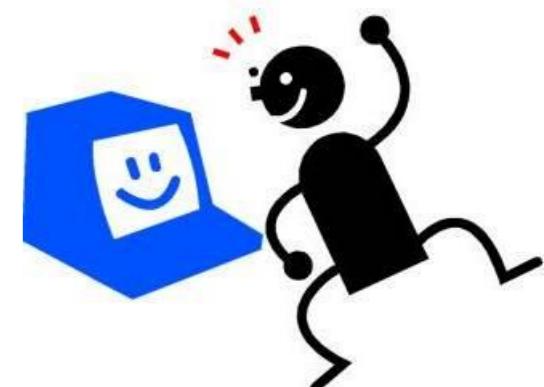
DDBJデータベースを検索してみましょう

---

# 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
  - 例としてヨーロッパブドウ(ヴィニフェラ種、Vitis Vinifera)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**

- [▶ スパコンの利用申込](#)
- [▶ スパコンの利用方法](#)
- [▶ スパコンマニュアル](#)

**■ 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 Hideki 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 \(\[ftp.ddbj.nig.ac.jp\]\(http://ftp.ddbj.nig.ac.jp\)\)](#)  
DDBJリリースなどのデータファイルをダウンロード

# getentry: アクセッショント番号等によるエントリ検索

 **DDBJ**  
DNA Data Bank of Japan

**getentry** --[アクセッショント番号等によるエントリ検索](#)--

[Help](#)  [\[English\]](#)

ID :

DNA データベース :  DDBJ / GenBank / EMBL  MGA

Protein データベース :  UniProt  PDB  DAD  Patent

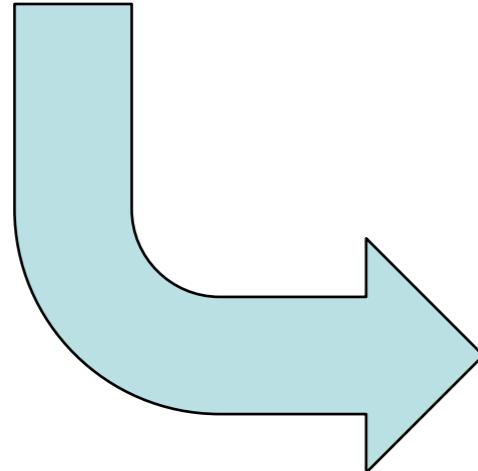
取得方法 :

出力形式 :

出力形式 :

上限 :  件

Copyright© DDBJ All rights reserved.



LOCUS	FN597015	23037639 bp	DNA	linear	CON 25-NOV-2009
DEFINITION	Vitis vinifera, whole genome shotgun sequence of line PN40024, unoriented chromosome 1, chr1.				
ACCESSION	FN597015	CU462738			
VERSION	FN597015.1				
DBLINK	Project:PRJEA18785				
KEYWORDS					
SOURCE	Vitis vinifera				
ORGANISM	<a href="#">Vitis vinifera</a>				
REFERENCE	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons; rosids; Vitales; Vitaceae; Vitis.				
AUTHORS	1 (bases 1 to 23037639) Scalabrin,S.				
JOURNAL	Submitted (30-SEP-2009) to the INSDC. Scalabrin S., Istituto di Genomica Applicata, Via J.Linussio, 51, Udine 33100 Italy, ITALY.				
REFERENCE	2 Jaillon,O., Aury,J., Noel,B., Policriti,A., Clepet,C., Casagrande,A., Choisne,N., Aubourg,S., Vitulo,N., Jubin,C., Vezzi,A., Legeai,F., Hugueney,P., Dasilva,C., Horner,D., Mica,E., Jublot,D., Poulain,J., Bruyere,C., Billault,A., Segurens,B., Gouyvenoux,M., Ugarte,E., Cattonaro,F., Anthouard,V., Vico,V., Del,Fabbro,C., Alaux,M., Di,Gaspero,G., Dumas,V., Felice,N., Paillard,S., Juman,I., Moroldo,M., Scalabrin,S., Canaguier,A., Le,Clainche,I., Malacrida,G., Durand,E., Pesole,G., Laucou,V., Chatelet,P., Merdinoglu,D., DelleDonne,M., Pezzotti,M., Lecharny,A., Scarpelli,C., Artiguenave,F., Pe,M., Valle,G., Morgante,M., Caboche,M., Adam-Blondon,A., Weissenbach,J., Quetier,F. and Wincker,P.				
AUTHORS	The French-Italian Public Consortium for Grapevine Genome Characterization				
CONSRTM					
TITLE	The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla				
JOURNAL	Nature 449(7161), 463-467(2007).				
PUBMED	17721507				
COMMENT	This is a 'working draft' sequence, generated by whole genome shotgun. This assembly was made at a 12X coverage of the genome, sequenced by Sanger method. Sequencing effort has been performed by GENOSCOPE, CRIBI (Consortium VIGNA) and IGA. More information available at <a href="http://www.genoscope.cns.fr/vitis">http://www.genoscope.cns.fr/vitis</a> and <a href="http://www.vitisgenome.it/">http://www.vitisgenome.it/</a> and <a href="http://www.appliedgenomics.org/">http://www.appliedgenomics.org/</a> Email: seqref@genoscope.cns.fr.				
FEATURES	Location/Qualifiers source 1..23037639 /organism=" <a href="#">Vitis vinifera</a> " /chromosome="1" /contig="PN40024"				

# getentry をwebAPIで利用する

- URLに検索したいIDを指定するだけ
  - `http://getentry.ddbj.nig.ac.jp/getentry/データベース名/アクセッション番号`
  - `http://getentry.ddbj.nig.ac.jp/getentry/ddbj/AB091058`
- 塩基配列をfasta形式で取得
  - <http://getentry.ddbj.nig.ac.jp/getentry/ddbj/AB091058?format=fasta>
- 複数アクセションの取得
  - <http://getentry.ddbj.nig.ac.jp/getentry/ddbj/AK388128,AK388131,AK388124,AK388126>
- たくさんのファイルを圧縮ファイルとして取得
  - `http://getentry.ddbj.nig.ac.jp/getentry/ddbj/AK377185-AK378194?filetype=gz`

# 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**

- [▶ スパコンの利用申込](#)
- [▶ スパコンの利用方法](#)
- [▶ スパコンマニュアル](#)

**■ 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 Hideki 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

# ARSA: DDBJの全文検索

ARSA All-round Retrieval of Sequence and Annotation

DDBJQuick Search DDBJAdvanced Search

QuickSearch

検索条件を複数入力する場合は、&(AND条件)、|(OR条件)、!(AND NOT条件)を指定することが可能です。

“vitis vinifera”のゲノム配列を検索で取得してみましょう

ARSA All-round Retrieval of Sequence and Annotation

DDBJQuick Search DDBJAdvanced Search

QuickSearch

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。

検索結果：  
ヒット数が多すぎるので  
絞り込みが必要  
(768931件だから…)



詳細検索  
(Advanced Search)へ

[DDBJ\(847794\)](#)  
[Patent AA\(42\)](#)

[Vitis vinifera\(768931\)](#)  
[Coffea arabica\(53560\)](#)  
[Pinus taeda\(7284\)](#)  
[Coffea canephora\(5843\)](#)  
[Coffea racemosa\(3098\)](#)  
[uncultured Glomus\(650\)](#)  
[Jatropha curcas\(472\)](#)  
[Vigna unguiculata\(404\)](#)  
[Pinus radiata\(380\)](#)  
[Grapevine leafroll-associated virus 3\(341\)](#)  
[All other taxa\(6873\)](#)

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

All Select

PrimaryAccessionNumber	Definition	moltype	Organism	Length
<input type="checkbox"/> <a href="#">A27657</a>	V.vinifera GPO1 gene.	DNA	Vitis vinifera	1990
<input type="checkbox"/> <a href="#">AB001375</a>	Vitis vinifera mRNA for soluble NSF attachment protein homologue, complete cds.	mRNA	Vitis vinifera	1167
<input type="checkbox"/> <a href="#">AB015870</a>	Vitis vinifera gVvPAL1 gene for phenylalanine ammonia-lyase, partial	DNA	Vitis vinifera	1938

# 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) を指定することが可能です。
- 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 =  **vitis vinifera**  
Taxon =

# 検索結果

ARSA All-round Retrieval of Sequence and Annotation

DDBJQuick Search DDBJAdvanced Search

DDBJ(35) Refine Search

FlatFile XML fasta View Download

All Select

PrimaryAccessionNumber	Definition	moltype	Organism	Length
<input type="checkbox"/> AH015709	Vitis vinifera cultivar Sugraone Gret1 retrotransposon 5' LTR and 3' LTR.	DNA	Vitis vinifera	1658
<input type="checkbox"/> FN597014	Vitis vinifera, whole genome shotgun sequence of line PN40024, scaffold_2838.assembly12x.	DNA	Vitis vinifera	2038
<input checked="" type="checkbox"/> FN597015	Vitis vinifera, whole genome shotgun sequence of line PN40024, unoriented <u>chromosome 1, chr1</u> .	DNA	Vitis vinifera	23037639
<input type="checkbox"/> FN597016	Vitis vinifera, whole genome shotgun sequence of line PN40024, unordered unoriented chromosome 1, chr1_random.	DNA	Vitis vinifera	568933
<input type="checkbox"/> FN597017	Vitis vinifera, whole genome shotgun sequence of line PN40024, unoriented <u>chromosome 2, chr2</u> .	DNA	Vitis vinifera	18779844
<input type="checkbox"/> FN597018	Vitis vinifera, whole genome shotgun sequence of line PN40024, unoriented <u>chromosome 3, chr3</u> .	DNA	Vitis vinifera	19341862
<input type="checkbox"/> FN597019	Vitis vinifera, whole genome shotgun sequence of line PN40024, unordered unoriented chromosome 3, chr3_random.	DNA	Vitis vinifera	1220746
	Vitis vinifera, whole genome shotgun sequence of line PN40024, unoriented <u>chromosome 4, chr4</u> .	DNA	Vitis vinifera	23867706
	Vitis vinifera, whole genome shotgun sequence of line PN40024, unordered unoriented chromosome 4, chr4_random.	DNA	Vitis vinifera	76237
	Vitis vinifera, whole genome shotgun sequence of line PN40024, unoriented <u>chromosome 5, chr5</u> .	DNA	Vitis vinifera	25021643

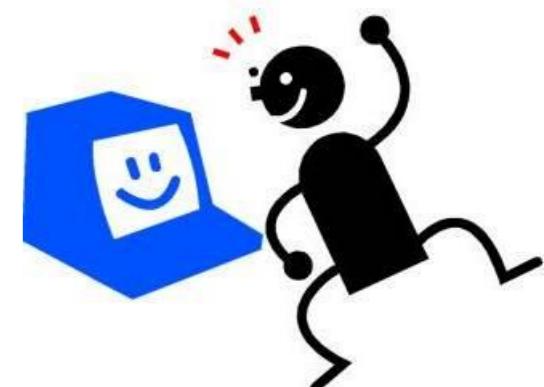
FN 597015をクリックすると  
1番染色体ゲノムエントリを表示

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

# 実習2

- ◆ 同じ検索をNCBI の検索システム ”Entrez”で実行してみましょう
  - 例として”vitis vinifera”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/> を入力



# Vitis Vinifera で検索

NCBI Entrez, The Life Sciences Search Engine

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

Search across databases **vitis vinifera** GO Clear Help

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

1376  PubMed: biomedical literature citations and abstracts	1  Books: online books
2322  PubMed Central: free, full text journal articles	1  OMIM: online Mendelian Inheritance in Man
17  Site Search: NCBI web and FTP sites	

131622  Nucleotide: Core subset of nucleotide sequence records	none  dbGaP: genotype and phenotype
511683  EST: Expressed Sequence Tag records	93397  UniGene: gene-oriented clusters of transcript sequences
229279  GSS: Genome Survey Sequence records	none  CDD: conserved protein domain database
98705  Protein: sequence database	104198  Clone: integrated data for clone resources
1  Genome: whole genome sequences	983  UniSTS: markers and mapping data
16  Structure: three-dimensional macromolecular structures	748  PopSet: population study data sets
1  Taxonomy: organisms in GenBank	none  GEO Profiles: expression and molecular abundance profiles
470835  SNP: short genetic variations	593  GEO DataSets: experimental sets of GEO data

文献

塩基配列

EST

アミノ酸配列

ゲノム配列

立体構造

# ヨーロッパブドウ(ビニフェラ種)ゲノムページ

NCBI Resources How To My NCBI Sign In

Genome Genome vitis vinifera Search Save search Limits Advanced Help

Display Settings: Overview Send to:

[Organism Overview](#) ; [Genome Project Report](#) ; [Genome Annotation Report](#) ; [Organelle Annotation Report](#)

 **Vitis vinifera**  
Vitis vinifera overview

**Lineage:** Eukaryota[1626]; Viridiplantae[474]; Streptophyta[454]; Embryophyta[448]; Tracheophyta[440]; Spermatophyta[431]; Magnoliophyta[400]; eudicotyledons[303]; core eudicotyledons[291]; rosids[182]; Vitales[5]; Vitaceae[5]; Vitis[5]; Vitis vinifera[1]

Vitis vinifera is the most cultivated and economically important grape species. Grapes are not only consumed fresh but processed in making juice and wine. Though new cultivars of grapes have been developed over last few decades, their introduction into the market has been difficult. An existing cultivar can also be manipulated for a desirable [More...](#)

**Chromosomes**

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 10\_un 11\_un  
12\_un 13\_un 16\_un 9\_un 18\_un 1\_un 3\_un 4\_un 5\_un 7\_un 17\_un

Click on chromosome name to open MapViewer

**Genome Sequencing Projects**

◆ Chromosomes [1] ◇ Scaffolds or contigs [0] ♦ SRA or Traces [0] ♦ No data [2]

Organism	BioProject	Assembly	Status	Chrs	Organelles	Size (Mb)	GC%	Gene	Protein
Vitis vinifera	PRJNA33471, PRJEA18785, PRJNA16688, PRJNA34679	12X	◆	33	2	487.2	4.6	25,584	24,623

**Related information**

BioProject  
Gene  
Protein Clusters  
Components  
Protein  
PubMed  
Taxonomy

**Search details**

"Vitis vinifera"[Organism] OR vitis vinifera[All Fields]

Search See more...

**Recent activity**

Turn Off Clear

Vitis vinifera  
vitis vinifera (1)

# アセンブル結果

12X

Organism name: [Vitis vinifera](#)

Submitter: International Grape Genome Program

Assembly level: Chromosome

Genome representation: full

GenBank Assembly ID: GCA\_000003745.2 (latest)

RefSeq Assembly ID: GCF\_000003745.2 (latest)

RefSeq Assembly and GenBank Assembly Identical: no

WGS Project: [CAAP03](#)

[History](#) ([Show revision history](#))

Global statistics

Total sequence length	486,260,629
Total assembly gap length	15,058,008
Gaps between scaffolds	154
Number of scaffolds	2,067
Scaffold N50	3,426,264
Number of contigs	14,657
Contig N50	102,754
Total number of chromosomes and plasmids	21

See [Genome Information for Vitis vinifera](#)

Access the data

[Download the full sequence report](#)

[RefSeq FTP site](#)

Assembly Information

[Assembly Help](#)

Assembly Basics

[NCBI Assembly Data Model](#)

[Assembly Definition](#)

[Assembly Statistics](#)

Global assembly definition

Click on the table row to see sequence details in the table to the right

Assembly Unit Name
Primary Assembly
non-nuclear

Assembly Unit: Primary Assembly

[Download the full sequence report](#)

Molecule name	GenBank ID	=	RefSeq ID	Unlocalized sequences count
1	<a href="#">FN597015.1</a>	=	<a href="#">NC_012007.3</a>	0
2	<a href="#">FN597017.1</a>	=	<a href="#">NC_012008.3</a>	0
3	<a href="#">FN597018.1</a>	=	<a href="#">NC_012009.3</a>	0
4	<a href="#">FN597020.1</a>	=	<a href="#">NC_012010.3</a>	0

# 豊富なリンクとツール群

## Vitis vinifera chromosome 1, 12X, whole genome shotgun sequence

NCBI Reference Sequence: NC\_012007.3

[FASTA](#) [Graphics](#)

塩基配列

[Go to:](#) ▾

LOCUS	NC_012007	23037639 bp	DNA	linear	CON 07-DEC-2011
DEFINITION	Vitis vinifera chromosome 1, 12X, whole genome shotgun sequence.				
ACCESSION	NC_012007	GPC_000000699			
VERSION	NC_012007.3	GI:357773492			
DBLINK	Project: <a href="#">33471</a>				
KEYWORDS	WGS.				
SOURCE	Vitis vinifera				
ORGANISM	<a href="#">Vitis vinifera</a>				
	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons; rosids; Vitales; Vitaceae; Vitis.				
REFERENCE	1				
AUTHORS	Jaillon,O., Aury,J.M., Noel,B., Pollicetti,A., Clepet,C., Casagrande,A., Choisne,N., Aubourg,S., Vitulo,N., Jubin,C., Vezzi,A., Legeai,F., Hugueney,P., Dasilva,C., Horner,D., Mica,E., Jublot,D., Poulain,J., Bruyere,C., Billault,A., Segurens,B., Gouyvenoux,M., Ugarte,E., Cattonaro,F., Anthouard,V., Vico,V., Del Fabbro,C., Alaux,M., Di Gaspero,G., Dumas,V., Felice,N., Paillard,S., Juman,I., Moroldo,M., Scalabrin,S., Canaguier,A., Le Clainche,I., Malacrida,G., Durand,E., Pesole,G., Laucou,V., Chatelet,P., Merdinoglu,D., Delledonne,M., Pezzotti,M., Lecharny,A., Scarpelli,C., Artiguenave,F., Pe,M.E., Valle,G., Morgante,M., Caboche,M., Adam-Blondon,A.F., Weissenbach,J., Quetier,F. and Wincker,P.				
CONSRM	French-Italian Public Consortium for Grapevine Genome Characterization				
TITLE	The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla				
JOURNAL	Nature 449 (7161), 463-467 (2007)				
PUBMED	<a href="#">17721507</a>				

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

- Change region shown ▾
- Customize view ▾
- Analyze this sequence ▾
  - Run BLAST
  - Pick Primers
- Related information ▾
  - Assembly
  - BioProject
  - Components (Core)
  - Gene
  - Genome
  - Identical GenBank Sequence
  - Map Viewer
  - mRNA
  - Protein
  - PubMed
  - PubMed (Weighted)
  - Taxonomy
  - UniSTS
- Recent activity ▾
  - [Turn Off](#) [Clear](#)

# 配列を取得したあとの解析例

- ◆ “blast”等の配列類似性検索を実行して類似の配列を収集する
- ◆ ”primer3”等で配列をクローニングするためのプライマを設計する
- ◆ ”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

# 統合TV: ここまででの参考動画

- ・ DDBJ
  - ・ DDBJのサービス紹介
- ・ NCBI
  - ・ Entrezを使って配列を検索する ~ヒトのheat shock factor のアミノ酸配列を得る
  - ・ 遺伝子のRefSeq IDを調べる 2011
  - ・ NCBI BioProjectの使い方
- ・ EBI
  - ・ Biomart v0.8を使ってIDから遺伝子情報を取得する



## 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 a Microsoft Excel spreadsheet titled "excel.gene2date.xls". The table consists of 16 rows and 11 columns. The columns are labeled A through K. The first column contains gene names, and the subsequent columns represent different date formats. The data is as follows:

	A	B	C	D	E	F	G	H	I	J	K
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				
16											

# 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="GeneID: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

---

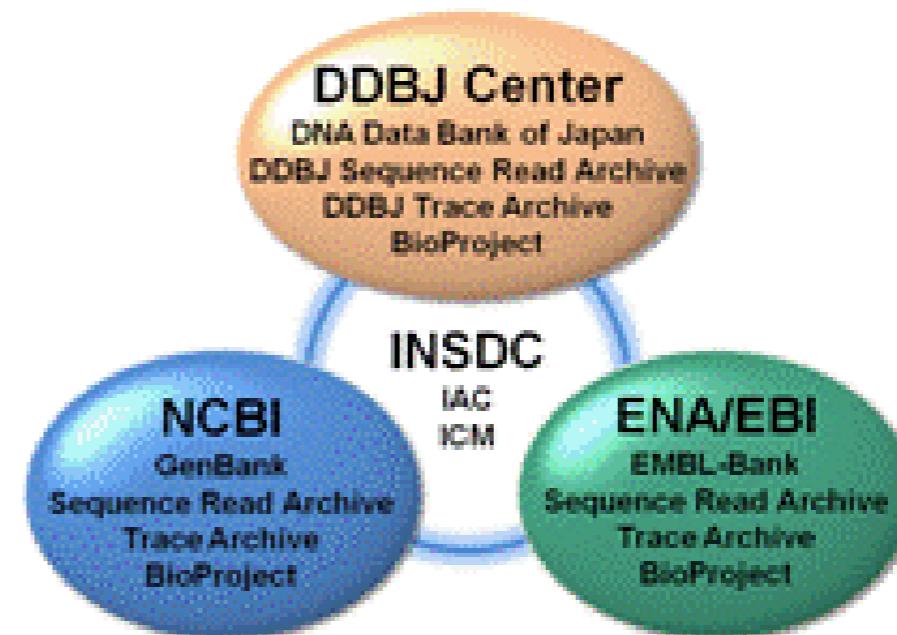
# Sequence Read Archive

- Roche 454 GS System®, Illumina Genome Analyzer®, Applied Biosystems SOLiD® Systemなどの次世代シーケンサからの出力データのためのデータベース

DDBJ:DRA

NCBI:SRA

EBI:ERA



# DRAについては次の講習で詳しく解説

DDBJパイプラインとDBCLS Galaxyの紹介:

河野信(ライフサイエンス統合データベースセンター)

## DRA Search ← → データ構造

**DRA Search**

Send Feedback | Search Home | DRA Home

Accession :   
Organism :  StudyType :   
CenterName :  Platform :   
Keyword :

Show 20 records Sort by Study Search Clear

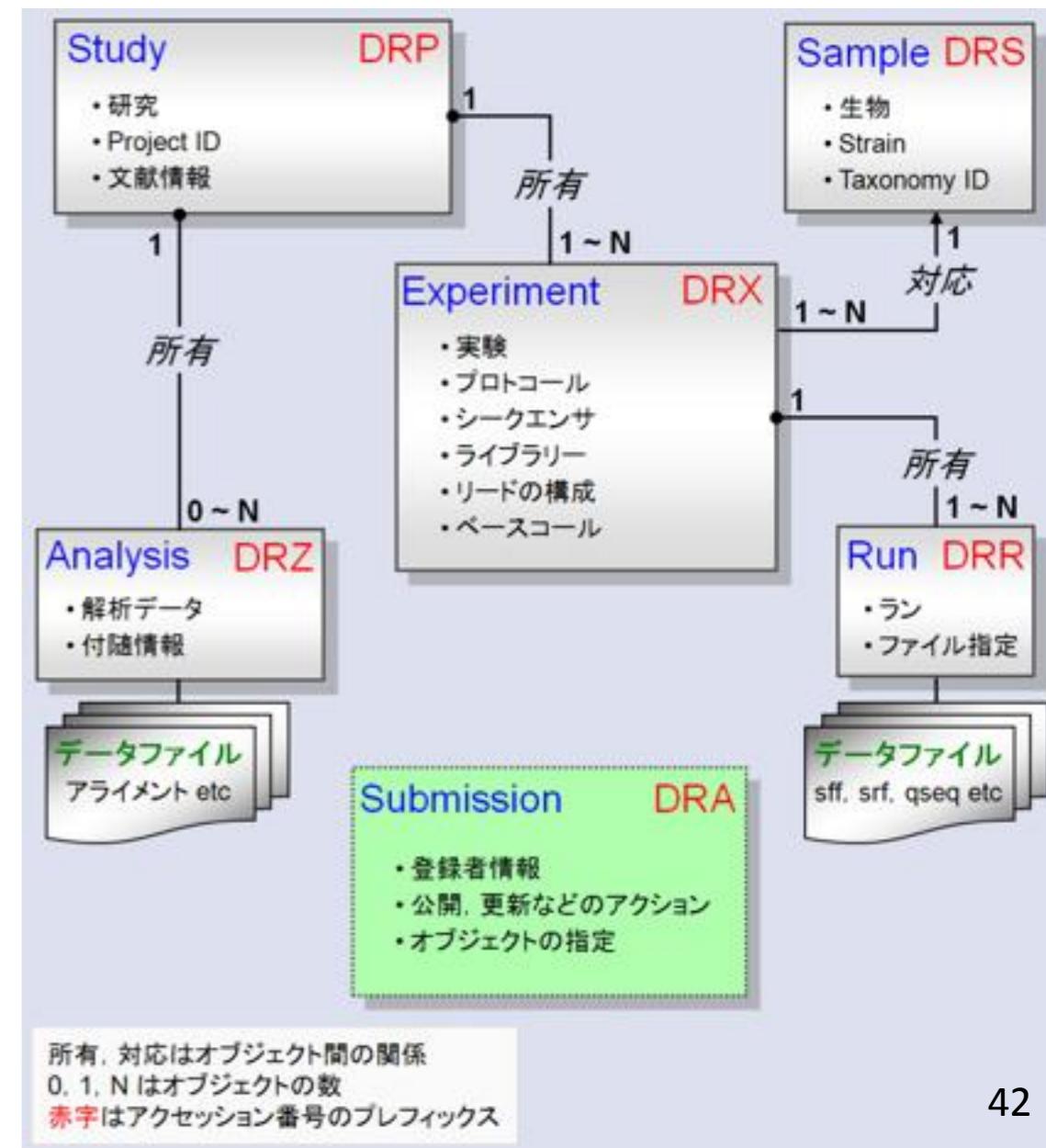
Data Last Update 2012-07-25  
WebSite Last Update 2012-04-23

**Statistics**

Type	Count
Submission	77293
Study	12729
Experiment	132252
Sample	303189
Run	228071

Organism      Study Type      Center Name [All List]

#	Organism Name	Study	#	Study Type	Study	#	Center Name	Study
1	Homo sapiens	1154	1	Whole Genome Sequencing	6264	1	JCVI	1888
2	unidentified	884	2	Transcriptome Analysis	1727	2	GEO	1831
3	Mus musculus	725	3	Metagenomics	1515	3	JGI	1768



# 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 the letters 'SRAs' in blue with a small cartoon character and a green submarine. To the right of the logo, the text 'SRAs: Survey of Read Archives' is displayed. 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 what SRAs are and links to external databases. The 'Simple Lists' section provides links to search by study, experiment, or run. Below these sections is a search form titled 'Search studies - まずは検索してみる', which includes fields for SRA#, Study#, TAXON ID, Platform, Title keyword, Type, and a dropdown menu. At the bottom of the page are '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.

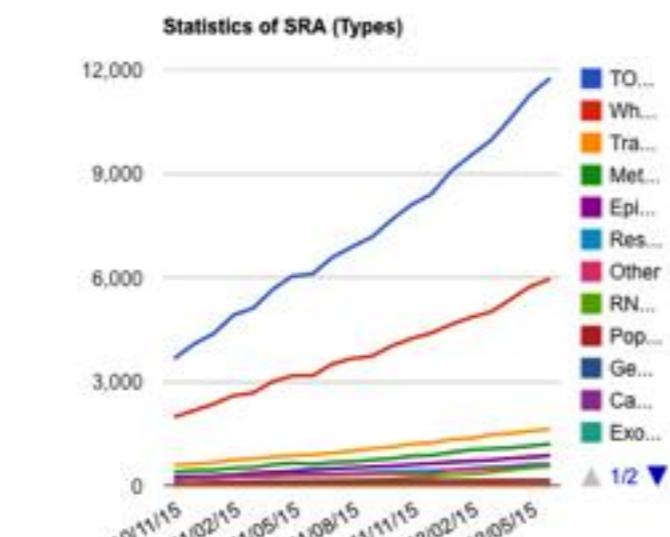
実際のプロジェクト数は"Study Types"に書かれているものです。"Platforms"や"Species of Samples"のtotalの値は、複数のプラットフォームで行われた実験が1つのプロジェクトでなされる 場合が多くあり、ダブルカウントしているので、数字が大きくなっています。

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)

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



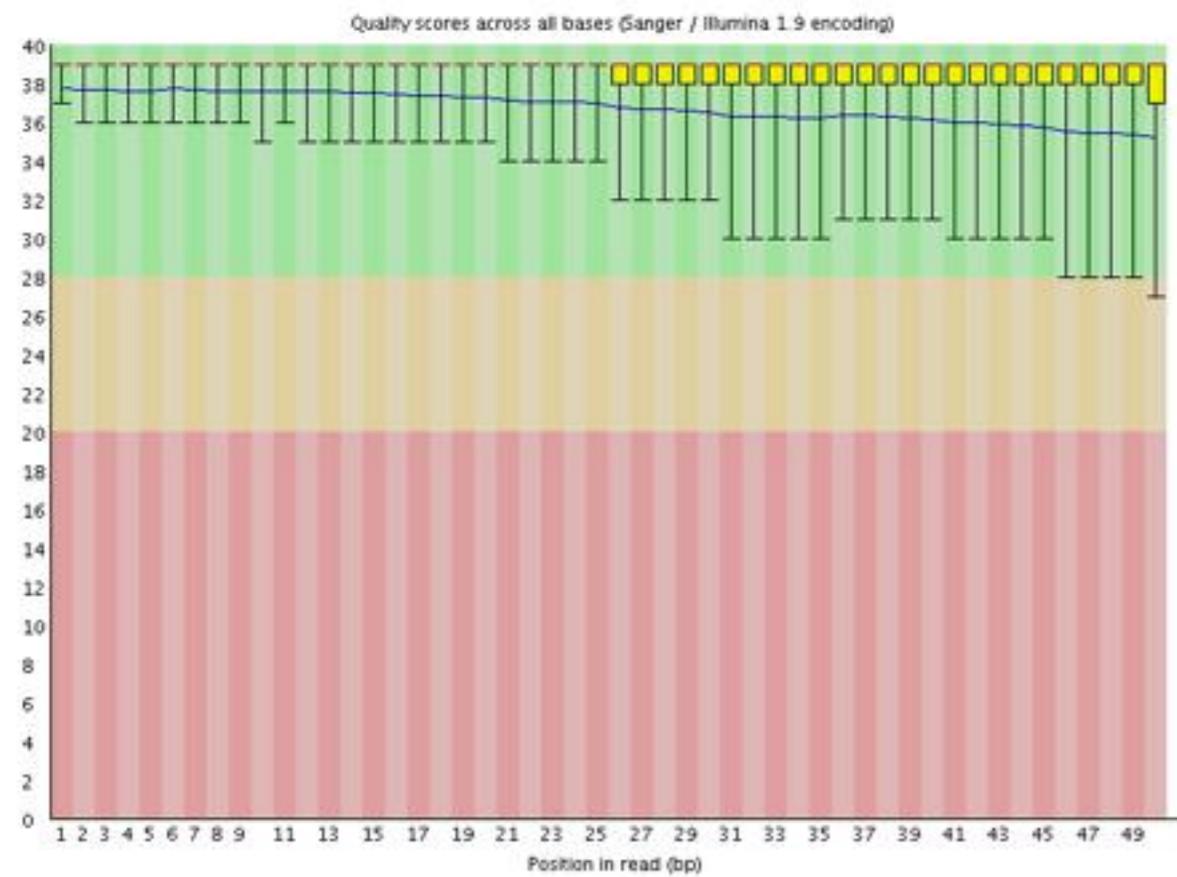
# 鎖鋸(kusarinoko)

- ◆ 独自に”FastQC”をかけてそれぞれのデータの質を評価

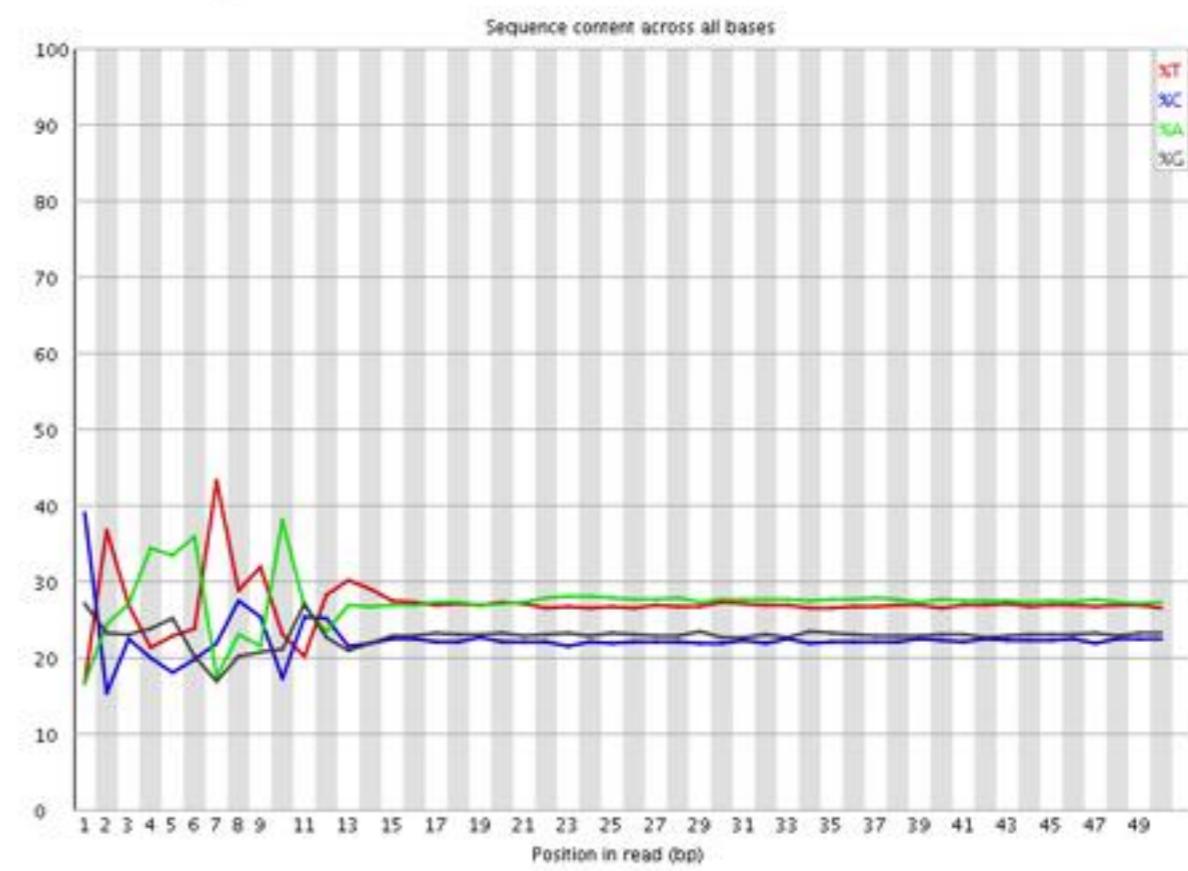
## Basic Statistics

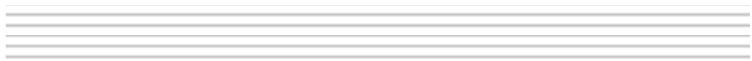
- File type: Conventional base calls
- Encoding: Sanger / Illumina 1.9
- Number of total sequences: 80397337
- Sequence length: 50
- %GC: 45

## Per base sequence quality



## Per base sequence content



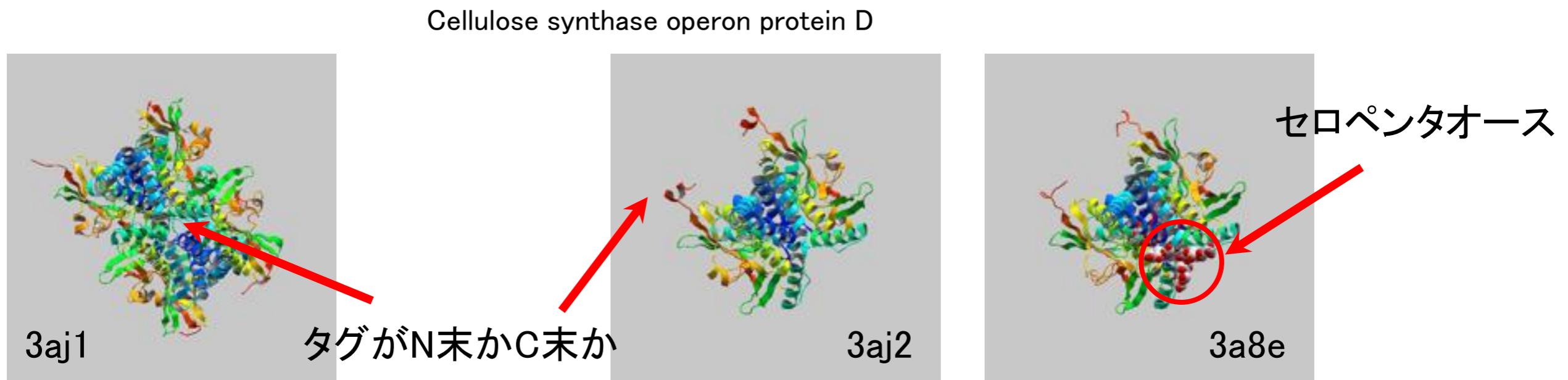


# PDBj

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83,400
  - 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



# PDBjは世界蛋白質構造データバンクの一員

## ◆ Worldwide Protein Data Bank (wwPDB)

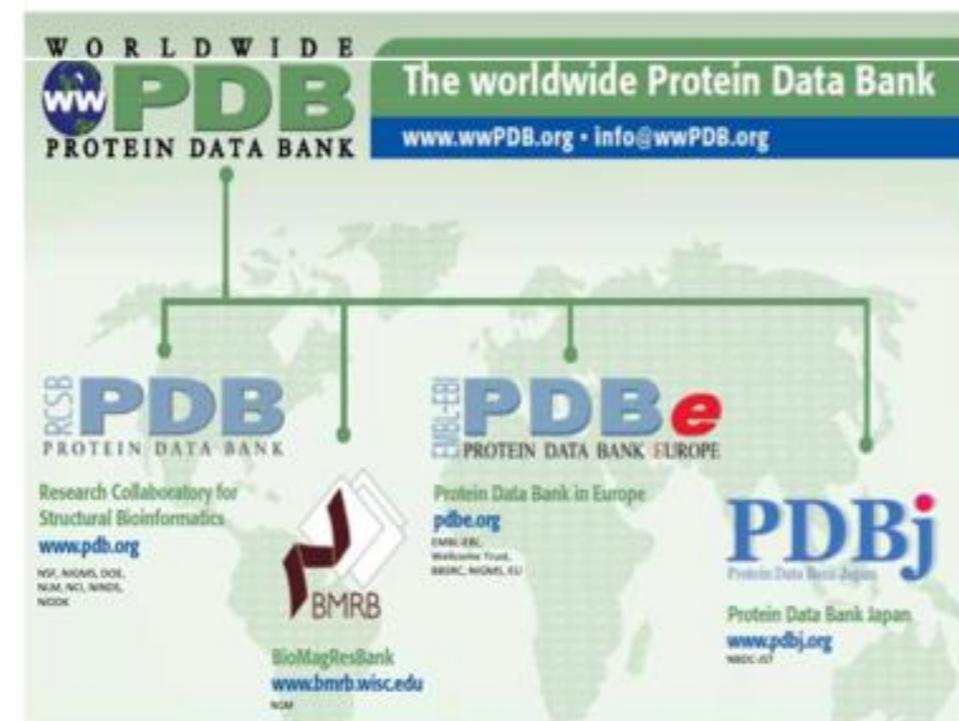
● 米国: RCSB-PDB (Research Collaboratory for Structural Bioinformatics)

● 欧州: PDBe

● 日本: PDBj

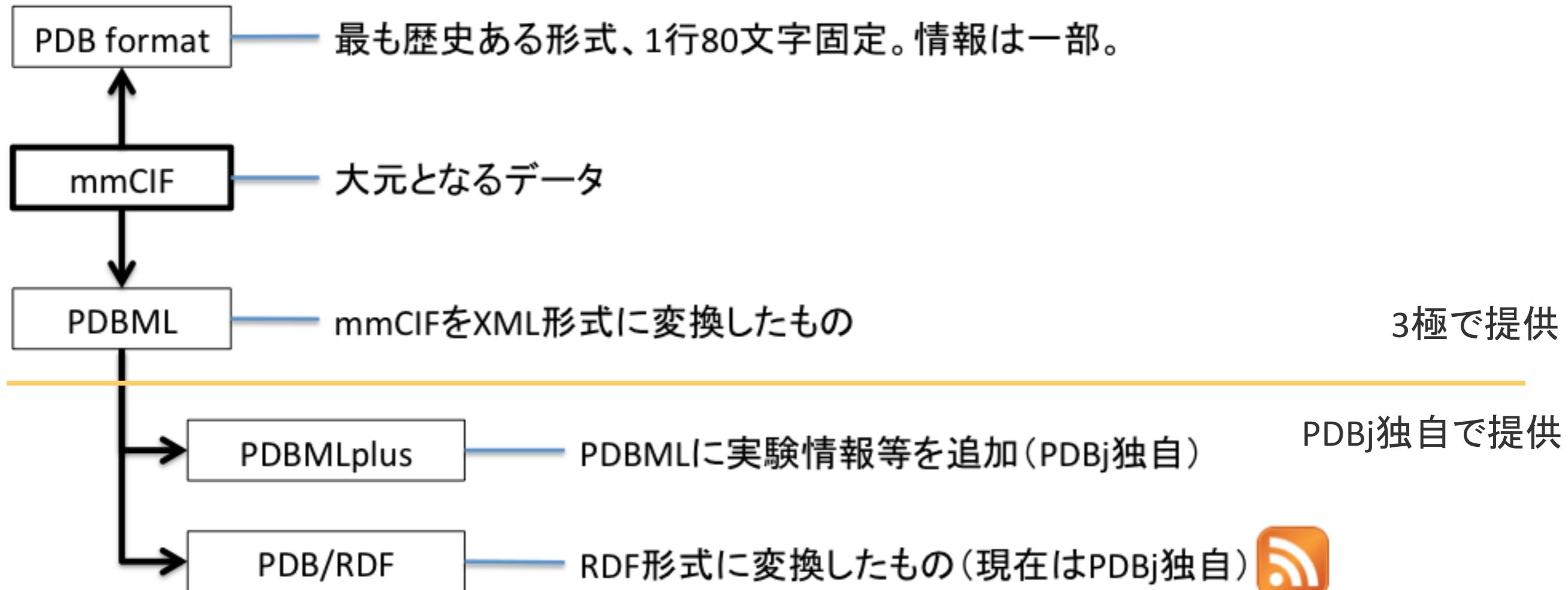
● 米国: BMRB

(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

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 CAA 601 1555 1555 2.44  
 LINK O ARG A 146 CA CAA 601 1555 1555 2.32  
 LINK OD1 ASP A 155 CA CAA 601 1555 1555 2.74  
 LINK OD2 ASP A 155 CA CAA 601 1555 1555 2.47  
 LINK O HIS A 189 CA CAA 601 1555 1555 2.37  
 LINK CA CAA 601 O HOH A 619 1555 1555 2.54  
 LINK CA CAA 601 O HOH A 628 1555 1555 2.47  
 LINK CA CAA 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

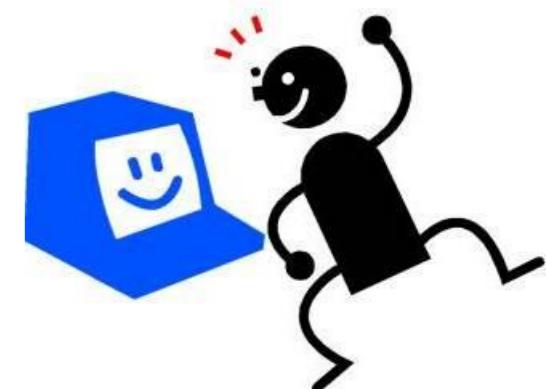
座標情報

# 実習3

◆ PDBjの検索サービス”PDBj Mine”で 「αアミラーゼ」を検索してみましょう

- 同一IDのPDBデータは3極どこで見ても同じ。
- ただし、PDBjだと日本語でも検索可能

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



# PDBj トップページ

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。

統計情報 ヘルプ FAQ お問い合わせ

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

TOP PAGE

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

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

NMRデータ登録 ADIT-NMR

NMRデータ検索

Accession number  
Deposition code

最新情報

2012/6/5 今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

2012/3/16 PDBで公開済みのエントリー数が、8万件を超える

2012/3/7 EMデータバンクがPDBアーカイブに追加されました

2012年3月7日に、EMデータバンク（EMDB）がPDB

詳細条件検索 >>

(cc) BY © 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

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)

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

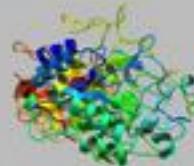
Mine

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

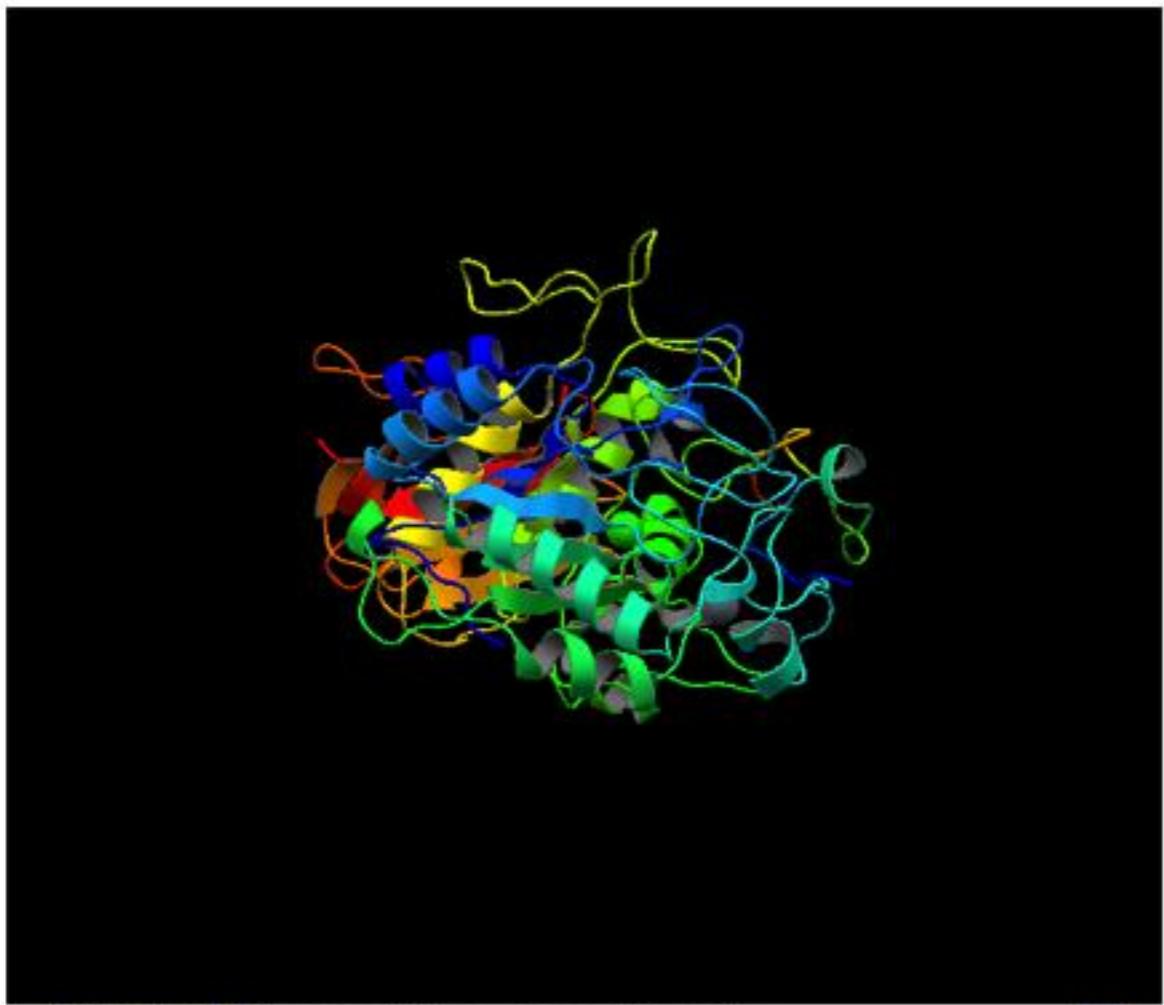
3次元構造ビューア  
(JView / Jmol) で見る \*1

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		

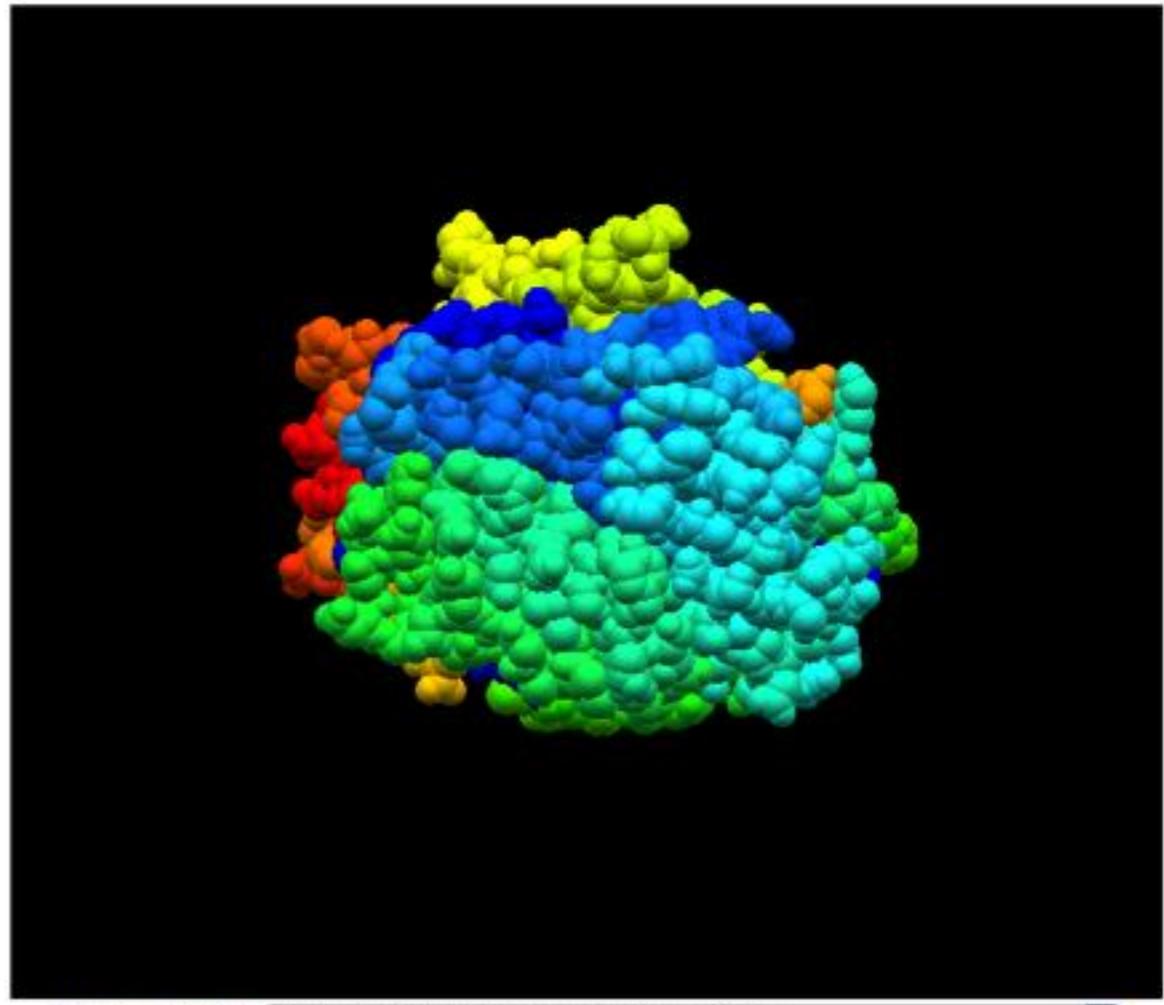
マウスで動かせる画像を表示

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

# jV4での表示



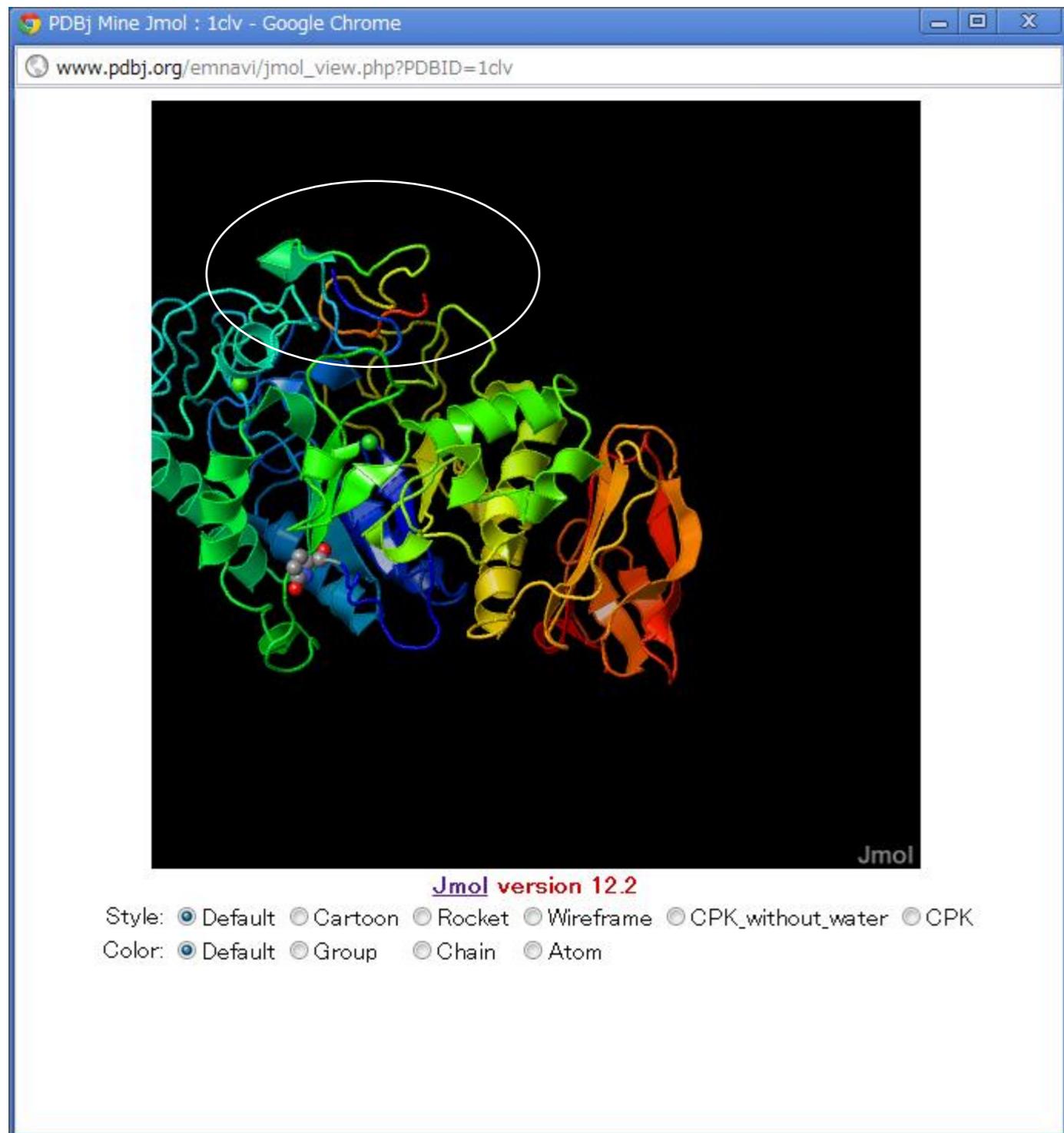
jV version 4 JNLP(JRE1.6以降 / Mac OS X 10.6以降)バージョン  
スタイル:  Default  Cartoon  Wireframe  CPK\_without\_water  CPK  
色:  Default  group  chain  atom



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

# JMolで見る



- AMARANTH ALPHA-AMYLASE INHIBITOR

# 構造情報ページ

## 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** 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

## 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( $\bar{I}$ )	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_F11	enzyme active site(ACT_SITE)	1	Proton donor.
SWS_FT_F12	enzyme active site(ACT_SITE)	1	Nucleophile.
SWS_FT_F13	metal binding site(METAL)	2	Calcium
SWS_FT_F14	metal binding site(METAL)	2	Calcium; via carboxylate
SWS_FT_F15	binding site(BINDING)	3	Chloride.
SWS_FT_F16	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

EKDANFASGRNSIVHLFEWKWNDIADECERFLQPQGF GG VQISPPNEYLVADGRPWWERYQPVSYIINTRSGDESAFTDMTRRCNDAGVRIYVDAVINHMTGMNGVTGSSADHDGMNYPAVPYGSGDFHSPCEVNYYQDADNVRNCELVGLRDLNQGSDYVRGVVLIDYMNHMIDLGVAGFRVDAAKHMSPGDL SVIFSGLKNLNTDYGFADGARPFYQEVIDLGGEAISKNEYTGFGCVLEFQFGVSLGNNAFQGGNQLKLN LANWGPEWGLLEG LDAAVVFDNHDNQRTGGSQLTYKNPKPYKMAIAFMLAHPYGTTRIMSSFDFTDNDQGPPQDGSGNLISPGINDDNTCSNGYVCEHRWRQVYGMVGFRNAVEGTQVENWWSNDDNQIAFSRGSGFVAF TNGGDLNQNLTGLPAGTYCDVISGELSGGSCTGKSVTVDNGSADISLGSAE DDGVLAHVNAKL

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

CIPKWNRCGPKMDGVPCCEPYTCTSDYYGNCS

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 1tmgA

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 [1tmgA] 構造の重ね合わせ

1clvA 1471 QKDANFASGRNSIVHLFEWKWNDIADECERFLQPQGF GG VQISPPNEYLVADGRPWWERYQPVSYIINTRSGDESAFTDMTRRCNDAGV

1tmgA 1471 QKDANFASGRNSIVHLFEWKWNDIADECERFLQPQGF GG VQISPPNEYLVADGRPWWERYQPVSYIINTRSGDESAFTDMTRRCNDAGV

1tmgA 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 1471 QKDANFASGRNSIVHLFEWKWNDIADECERFLQPQGF GG VQISPPNEYLVADGRPWWERYQPVSYIINTRSGDESAFTDMTRRCNDAGV

1viwA 1471 QKDANFASGRNSIVHLFEWKWNDIADECERFLQPQGF GG VQISPPNEYLVADGRPWWERYQPVSYIINTRSGNESAFDMTRRCNDAGV

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 1498 QKDANFASGRNSIVHLFEWKWNDIADECERFLQPQGF GG VQISPPNEYLVADG--RPWWERYQPVSYIINTRSGDESAFTDMTRRCNDAGV

3vm5A 1498 QHNPNTRDGRTAIVHLFEWRWADIAE CERFLGPKG FAGVQVQISPPNEHILVSSPWRPWQRYQPISYNLCSRSGGENELRDMITRCNNV

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<sup>PDB</sup>

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  
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

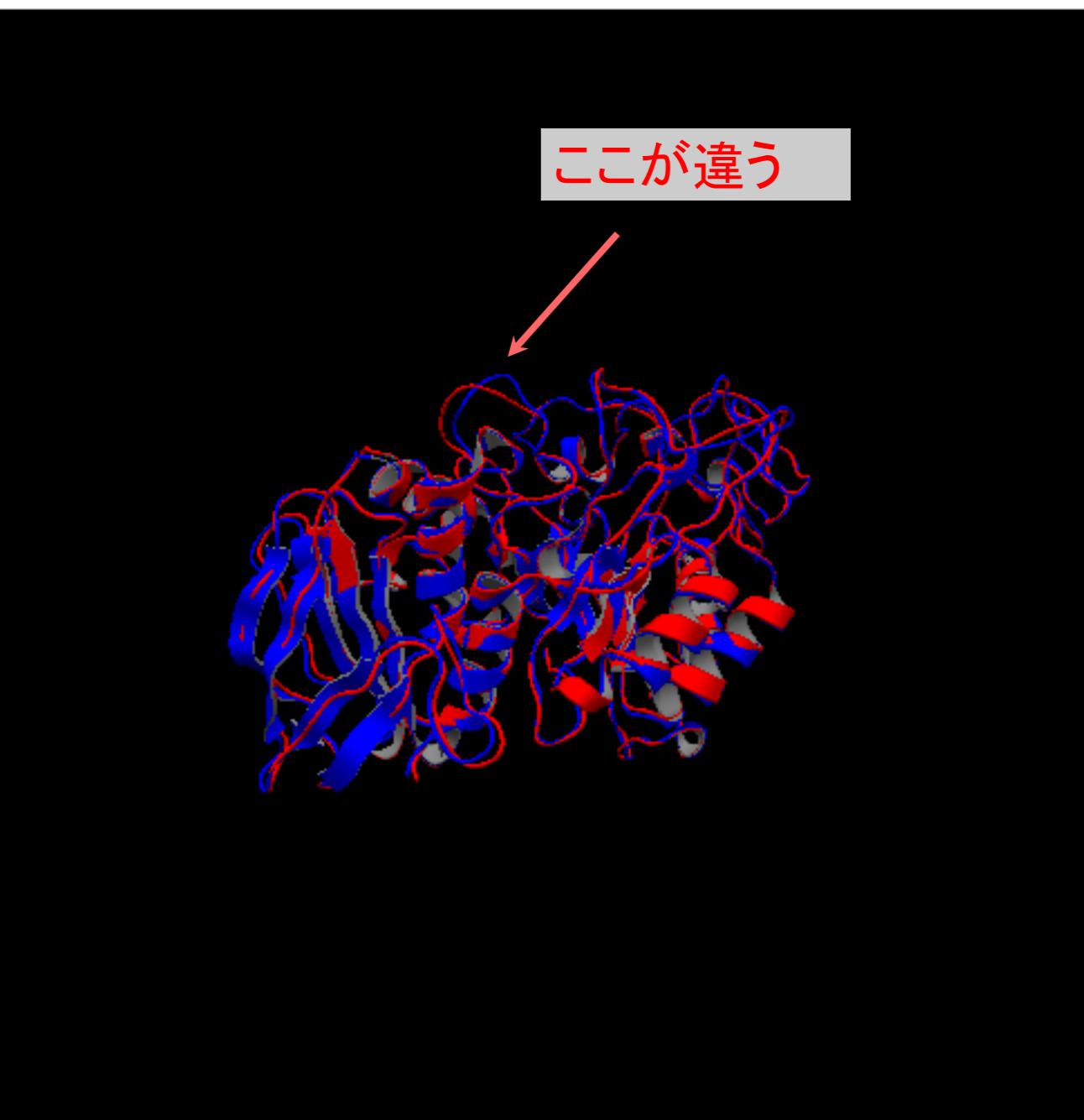
整列した残基数 = 470

### 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	KDANFASGRNSIVHLFEWKWNNDIADECERFLQPQGFGGVQISPPNEYLVADGRPWE
1clvA	CCCCCCCCCCCEEECCCCHHHHHHCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC
1viwA	KDANFASGRNSIVHLFEWKWNNDIADECERFLQPQGFGGVQISPPNEYLVADGRPWE
1viwA	CC
Ner類似度	99899
1clvA	PVSYIINTRSGDESAFTDMTRRCNDAGVRIYVDAVINHMTGMINGVTSGSSADHDGMN
1clvA	CCCCCCCCCCCHHHHHHHHHHHCCCCEEEECCCCCCCCCCCCCCCCCCCCCCCC
1viwA	PVSYIINTRSGNESAFDMTRRCNDAGVRIYVDAVINHMTGMINGVTSGSSADHDGMN
1viwA	CCCCCCCCCCCHHHHHHHHHHHCCCCEEEECCCCCCCCCCCCCCCCCCCC
Ner類似度	9997899999999

## 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** PDBj

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	<input type="button" value="表示"/>	<input type="button" value="ダウンロード"/>
	ヘッダのみ	pdb1clv.ent(364k)	非圧縮	<input type="button" value="表示"/>	<input type="button" value="ダウンロード"/>
mmCIF	全ての情報	pdb1clv.ent.gz(7k)	メタデータのみ(原子座標なし)	<input type="button" value="表示"/>	<input type="button" value="ダウンロード"/>
	ヘッダのみ (no-atom)	1clv.cif.gz(111k)		<input type="button" value="表示"/>	<input type="button" value="ダウンロード"/>
	PDBML	1clv-noatom.xml.gz(171k)		<input type="button" value="表示"/>	<input type="button" value="ダウンロード"/>
	PDBMLplus	1clv-pdbmlplus.xml.gz(27k)	メタデータのみ	<input type="button" value="表示"/>	<input type="button" value="ダウンロード"/>
XML	座標情報のみ (ext-atom)	1clv-extatom.xml.gz(96k)	メタデータのみ(PDBj独自) 原子座標データのみ(メタデータなし)	<input type="button" value="表示"/>	<input type="button" value="ダウンロード"/>
	RDF	1clv-noatom.rdf.gz(18k)	RDF形式データ(PDBj独自)	<input type="button" value="表示"/>	<input type="button" value="ダウンロード"/>
	生物学的単位 (PDB形式)	1clv.pdb1.gz(84k) *author_defined_assembly, 2 分子 (dimeric)		<input type="button" value="表示"/>	<input type="button" value="ダウンロード"/>

# 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

# 統合TV: ここまででの参考動画

- PDBj

- PDBj Mineを使ってタンパク質を検索する
- 万見(Yorodumi)の使い方～基本と連携～
- RCSB PDBを使ってタンパク質の立体構造を調べる
- CueMol2でタンパク質の立体構造を見る



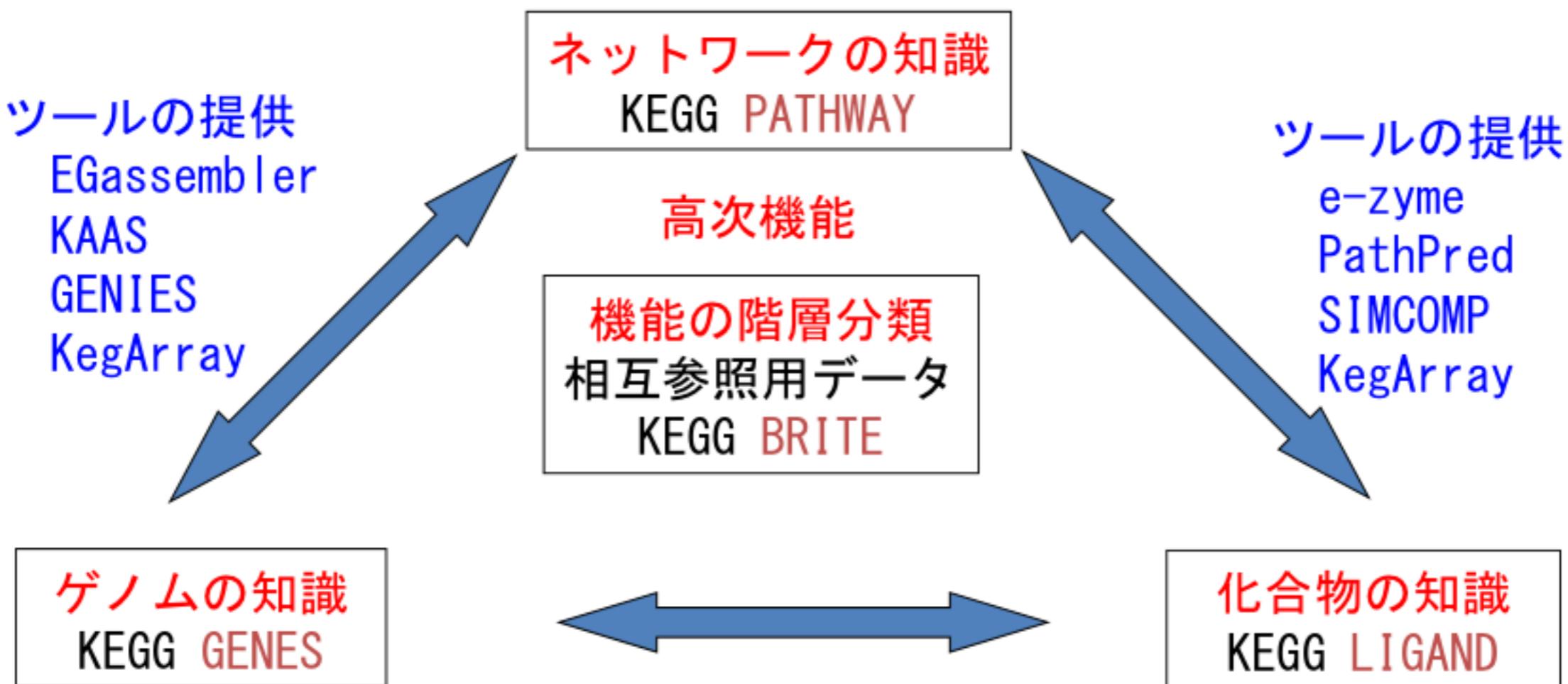


# KEGG

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

KEGG: Kyoto Encyclopedia of Genes and Genomes

Announcements  
In the next scheduled release on July 1, 2012 a new version of the KEGG API will be made available. It is a REST-based API service, replacing the current SOAP-based API service. [more ...](#)  
To accommodate the rapid increase of new genomes, 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) [お薦手帳]  
  
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

Copyright 1995-2012 Kanehisa Laboratories

KEGG - Table of Contents

KEGG2 PATHWAY BRITE MODULE DISEASE DRUG KO GENOME GENES LIGAND DBGET

Search **KEGG** for  Go Clear

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

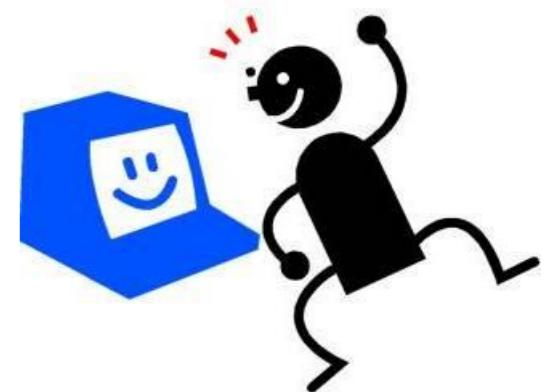
## ◆ KAAS (KEGG automatic annotation server)

- 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

- ◆ ”anthocyanidin”でKEGG GENESを検索してみましょう
  - PDBjとは異なり日本語-英語翻訳機能はありません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

## 生物種リストの表示(植物)

Entry Gene cluster Ortholog Paralog Motif Clear

**Announcement: Four-letter organism codes**

To accommodate the rapid increase of new genomes, the KEGG organism code is no longer limited to three letters. The current KEGG relational database schema allows up to five letters, but for now we still use either a three- or four-letter code for an organism in the GENES (high-quality complete genomes) and EGENES datasets. Note that the four letter codes starting with "d" for DGENES and starting with "e" for EGENES had been used for some time. Organisms in these categories are identified by the T numbers only.

Posted on June 4, 2012  
Revised on July 1, 2012

**Gene Catalogs**

KEGG GENES is a collection of gene catalogs for all complete genomes (see [release history](#)) and draft genomes from publicly available resources, mostly NCBI RefSeq. They are subject to SSDB computation and automatic annotation by KOALA tool. KEGG DGENES for draft genomes of some eukaryotes and MGENES for metagenomes are supplementary gene catalogs, which are given automatic KO assignments with GENES used as a reference data set. There is now a fourth type of gene catalogs, VGENES (see also [KEGG GENOME](#)) with automatic annotation. The viral gene catalog, VGENES, is not yet included 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 anthocyanidin Go Clear

bfind mode bget mode

Search Organism vvi for anthocyanidin Go Clear

bfind mode bget mode

Plants		Mustard family	ath	Arabidopsis thaliana (thale cress)	RefSeq
			aly	Arabidopsis lyrata (lyrate rockcress)	RefSeq
Eudicots	Pea family	gmx	Glycine max (soybean)	RefSeq	
	Willow family	mtr	Medicago truncatula (barrel medic)	RefSeq	
	Spurge family	pop	Populus trichocarpa (black cottonwood)	RefSeq	
	Grape family	rcl	Ricinus communis (castor bean)	RefSeq	
Monocots	Grass family	vvi	Vitis vinifera (wine grape)	RefSeq	
	osa	Oryza sativa japonica (Japanese rice)	RefSeq		
	dosa	Oryza sativa japonica (Japanese rice)	RAP-DB		
	bdi	Brachypodium distachyon	RefSeq		
	sbi	Sorghum bicolor (sorghum)	RefSeq		
	zma	Zea mays (maize)	RefSeq		
Ferns	smo	Selaginella moellendorffii	RefSeq		
	Mosses	ppp	Physcomitrella patens subsp. patens	RefSeq	
		cre	Chlamydomonas reinhardtii	RefSeq	
		vcn	Volvox carteri f. nagariensis	RefSeq	
Green algae		olu	Ostreococcus lucimarinus	RefSeq	
	ota	Ostreococcus tauri	RefSeq		
	cme	Cyanidioschyzon merolae	U Tokyo		
	sce	Saccharomyces cerevisiae (budding yeast)	RefSeq		
Red algae					

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

**GenomeNet**

Search T01084 for anthocyanidin Go Clear

Database: T01084 - Search term: anthocyanidin (Total 36 hits)

- 100232982 LAR1; leucoanthocyanidin reductase 1; K13081 leucoanthocyanidin reductase [EC:1.17.1.3]
- 100233142 LDOX, ANS; leucoanthocyanidin dioxygenase; K05277 leucoanthocyanidin dioxygenase [EC:1.14.11.19]
- 100266832 anthocyanidin 3-O-glucosyltransferase 5-like
- 100265092 anthocyanidin 3-O-glucosyltransferase 5-like; K12356 coniferyl-alcohol glucosyltransferase [EC:2.4.1.111]
- 100249635 anthocyanidin 3-O-glucosyltransferase 2-like
- 100254753 anthocyanidin 3-O-glucosyltransferase 2-like
- 100249544 anthocyanidin 3-O-glucosyltransferase 2-like
- 100247914 anthocyanidin 3-O-glucosyltransferase 2-like
- 100258170 anthocyanidin 3-O-glucosyltransferase 2-like
- 100258988 anthocyanidin 3-O-glucosyltransferase 6-like
- 100245030 leucoanthocyanidin dioxygenase-like
- 100241356 anthocyanidin reductase-like
- 100265454 anthocyanidin reductase-like
- 100233099 UFGT, AIUGT1, AIUGT2, FIUGT1, FIUGT2, ITUGT1, ITUGT2, RUUGT1, RUUGT2, VVGT1; UDP glucose:flavonoid 3-O-glucosyltransferase; K12930 anthocyanidin 3-O-glucosyltransferase [EC:2.4.1.115]
- 100257289 anthocyanidin 5,3-O-glucosyltransferase-like
- 100254886 anthocyanidin 3-O-glucosyltransferase 5-like
- 100247997 anthocyanidin 3-O-glucosyltransferase 5-like
- 100268157

**Vitis vinifera (wine grape): 100233099**

<b>Entry</b>	100233099	CDS	T01084
<b>Gene name</b>	UFGT, AIUGT1, AIUGT2, FIUGT1, FIUGT2, ITUGT1, ITUGT2, RUUGT1, RUUGT2, VVGT1		
<b>Definition</b>	UDP:glucose:flavonoid 3-O-glucosyltransferase		
<b>Ortholog</b>	K12930 anthocyanidin 3-O-glucosyltransferase [EC:2.4.1.115]		
<b>Organism</b>	vvi Vitis vinifera (wine grape)		
<b>Pathway</b>	vvi00942 Anthocyanin biosynthesis vvi01100 Metabolic pathways vvi01110 Biosynthesis of secondary metabolites		
<b>Class</b>	Metabolism; Biosynthesis of Other Secondary Metabolites; Anthocyanin biosynthesis [PATH:vvi00942] <a href="#">BRITE hierarchy</a>		
<b>SSDB</b>	<a href="#">Ortholog</a> <a href="#">Paralog</a> <a href="#">GIFT</a>		
<b>Motif</b>	Pfam: UDPGT Glyco_trans_1_3 Glyco_tran_28_C <a href="#">Motif</a>		
<b>Other DBs</b>	NCBI-GI: 225454815 NCBI-GenelD: 100233099 UniProt: D7T7R5		
<b>Structure</b>	PDB: 2C9Z 2C1X 2C1Z <a href="#">Thumbnails</a>  <a href="#">Jmol</a>		
<b>Position</b>	16		
<b>AA seq</b>	456 aa <a href="#">AA seq</a> <a href="#">CB search</a> MSQTTPNPHAVLAFPFSTHAAPLLAVRRLLAAAAPHAVFSFFSTSQSNSAFHDSDMHTM QCNIKSYDVSDGVREGYVFAGRQREDIELFMRRAAPESFRQGMMAVAETGRPVSLVADA FIWFAADDMAREMGVWLPFWTAGPNSLSTHVYTDEIREKIIGVSGIQGREDELLNFIPGMS KVRFRDLQEGIVFGMNLNSLFSRMLHRMGOVLPKATAVFINSFEELDDSLNDLKSKLKY LNI GPFNLITPPPVIPNTTGCLQWLKERKPTSVVYISFGTVTTPPPAAELVALAEALEASR VPFIWSLRDKARVHLPEGFLEKTRGYGMVPMAPQAEVLAHEAVGAFVTHCGIMSLWESV AGGVPLICRPFYGDQRRLNGRMVEDALEIGVRI EGGVFTESGLMSCFDQILSQEKGKKLRE NLGALRETADRAGPKGSSTENFKTLVDLVSKPKDV		
<b>NT seq</b>	1371 nt <a href="#">NT seq</a> atgtctaaaccacccaccaacccccatgtggcggctctggccctccatgcaggcttc gcagccccccctccgtggcgctgtggccgcctgtggccctcatgcaggcttc tccctttcataccatgcaggccaaatccaaacgcctccatcttccacgcactccatgcataccatx caatgttaatatcaagtccatgtgtggcggactgtggccctgtggccatgtttggcc xxxxxxcccaggatatttgatgtttccatgatggccctggccatgtttggcc xxxxatgttgcgttgcgtggccgagacagxxccatgtggccatgtttggcc ttcatttttttcgtccatgtatggccatgtggccatgtttttggccatgtttggcc actgcaggccatctaactcactccatccatgtttacactgtatggaaaatcaggaaaaatgg xxatgttgcggcattcaaggccatgtggccatgtttggccatgtttggccatgtttggcc aaatggatgtttggccatgtggccatgtttggccatgtttggccatgtttggccatgtttggcc tcacgcattgtccatgtggccatgtttggccatgtttggccatgtttggccatgtttggcc tccctggccatgtggccatgtttggccatgtttggccatgtttggccatgtttggccatgtttggcc ctcaaatatgtccatgtggccatgtttggccatgtttggccatgtttggccatgtttggcc tgcctccatgtggccatgtttggccatgtttggccatgtttggccatgtttggccatgtttggcc xtcagccacccaccccccagccatgtttggccatgtttggccatgtttggccatgtttggcc xtacogtttatgttgcctcaaggacaaggccatgtttggccatgtttggccatgtttggcc		

# KEGG ORTHOLOGYのエントリ

**KEGG** ORTHOLOGY: K12930 [Help](#)

<b>Entry</b>	K12930	K0
<b>Name</b>	BZ1	
<b>Definition</b>	anthocyanidin 3-O- $\beta$ -glucosyltransferase [EC:2.4.1.115]	
<b>Pathway</b>	ko00942 Anthocyanin biosynthesis	
<b>Class</b>	Metabolism; Glycan Biosynthesis and Metabolism; Glycosyltransferases [BR:ko01003] Metabolism; Biosynthesis of Other Secondary Metabolites; Anthocyanin biosynthesis [PATH:ko00942] <a href="#">BRTE hierarchy</a>	
<b>Other DBs</b>	RN: R06534 R06535 R06536 GO: 0047213 CAZy: GT1	
<b>Genes</b>	POP: POPTR_287596(FGT2) RCU: RCOM_1324500 WI: 100233099(UFGT) <a href="#">Taxonomy</a> <a href="#">KOALA</a> <a href="#">UniProt</a>	
<b>Reference</b> <b>Authors</b> <b>Title</b>	PMID:9535914 Ford CM, Boss PK, Hoj PB Cloning and characterization of <i>Vitis vinifera</i> UDP- $\beta$ -glucosylflavonoid 3-O- $\beta$ -glucosyltransferase, a homologue of the enzyme encoded by the maize Bronze-1 locus that may primarily serve to glucosylate anthocyanidins in vivo.	
<b>Journal</b>	J Biol Chem 273:9224-33 (1998)	
<b>Reference</b> <b>Authors</b> <b>Title</b>	PMID:16482224 Offen W, Martinez-Fleites C, Yang M, Kiat-Lim E, Davis BG, Turling CA, Ford CM, Bowles DJ, Davies GJ Structure of a flavonoid glucosyltransferase reveals the basis for plant natural product modification.	
<b>Journal</b>	EMBO J 25:1396-405 (2006)	

# KAAS: 自動アノテーションツール



## 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.

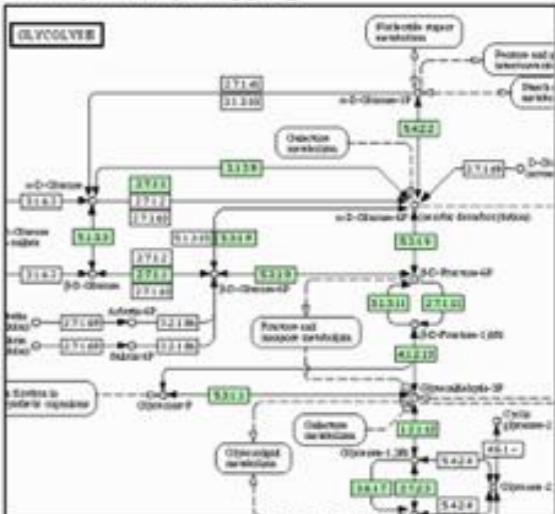
- KAAS job request (BBH method)
- KAAS job request (SBH method)

### Example of Results

#### KO assignment

KO Assignment Results	
Query gene : KO assignment	
query_0001	
query_0002	K00003
query_0003	K00072
query_0004	K01733
query_0005	
query_0006	
query_0007	K03010
query_0008	K06016
query_0009	K08931
query_0010	K07534
query_0011	
query_0012	
query_0013	
query_0014	K04043
query_0015	K02098
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]  
MNIEKFTERSRGFLQAAQTIAMREYNQQLTPEHLLKALLDDDQGAASALIRAAGGQPPAIAAAVDTALAKLPKVQGGGAGQPSATPDLVRLLDAAEQAAQKAG  
DEYVAQDRLLAAIAASETPAGQALRAGGATPQALDKAIATIRKGRTVTSENAEASF DALKKYARDVTEIALQGKLDPVIGRDEEIRRAIQVLARRSKN  
>gi|296114122|ref|ZP_06832777.1| hypothetical protein GXY_00019 [Gluconacetobacter hansenii ATCC 23769]  
MGAALIMITTLIAGYALCEWPPGEETPALMQIK  
....  
....  
....
```

## TCAサイクル

 KAAS

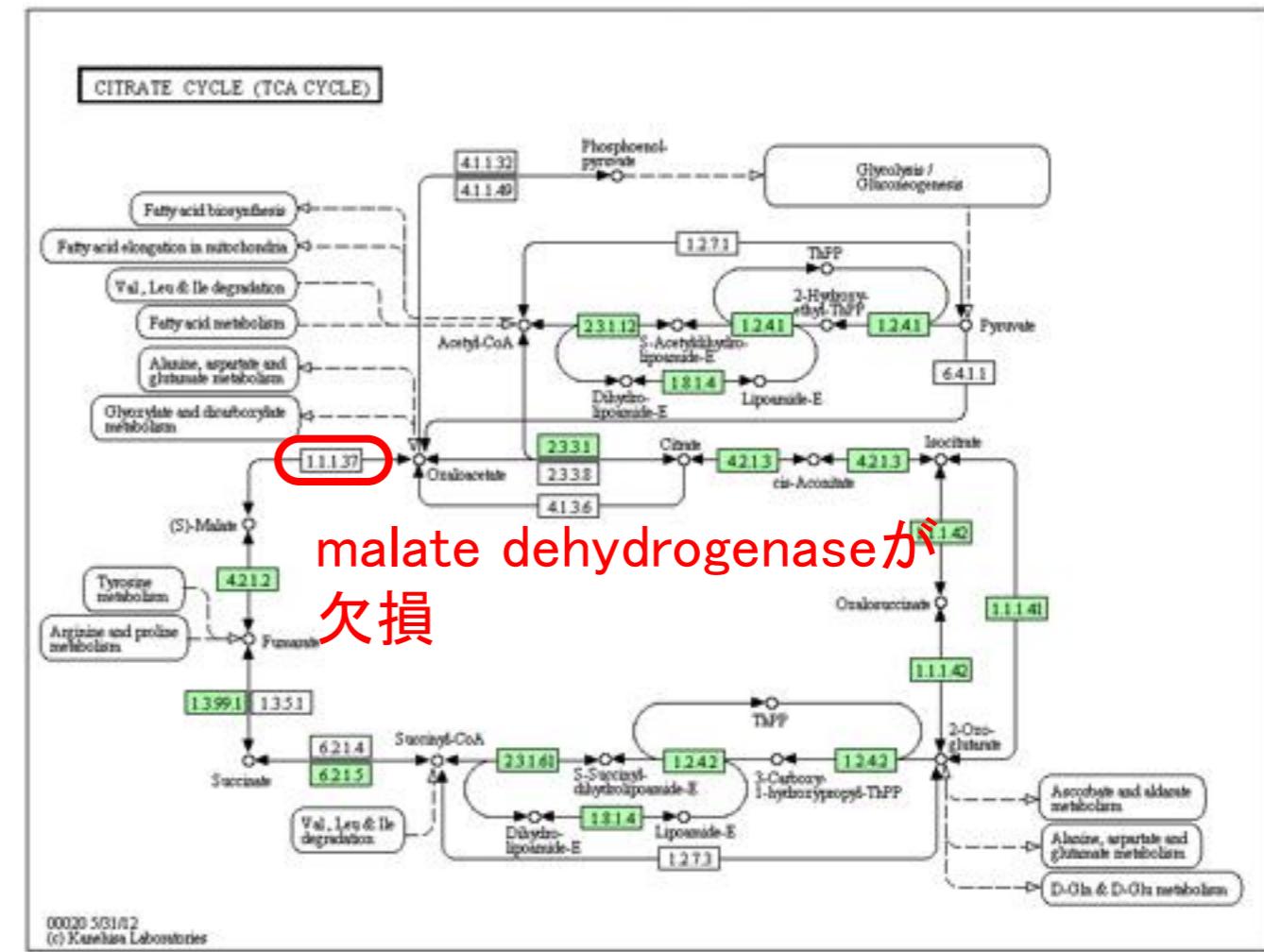
### KO Assignment Results

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

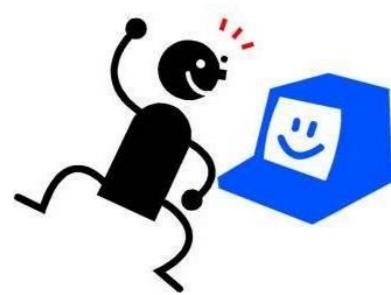
- 生体内反応のDB. これらをつなぎ合わせるとパスウェイになる

## ◆ 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”anthocyanidin”を検索してみましょう



**KEGG LIGAND Database**  
Molecular building blocks of life in the chemical space

KEGG2 PATHWAY BRITE MODULE LIGAND COMPOUND GLYCAN REACTION

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

Filter Pathway mapping Brite mapping Get title Get entry Clear

**Chemical Substances and Reactions**

KEGG LIGAND contains our knowledge on the universe of chemical substances and reactions that are relevant to life. It is a composite database consisting of COMPOUND, GLYCAN, REACTION, RPAIR, RCLASS, and ENZYME databases, whose entries are identified by C, G, R, 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	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	

Search LIGAND for anthocyanidin Go Clear  
bfnd mode bget mode

# KEGG LIGAND検索結果

KEGG  
Search LIGAND for anthocyanidin Go Clear

Database: LIGAND - Search term: anthocyanidin

KEGG COMPOUND

- C02003 KegDraw Jmol  
Anthocyanidin
- C03940 KegDraw Jmol  
Anthocyanidin-3-O-beta-D-glucoside; Anthocyanidin-3-O-D-glucoside
- C06361 KegDraw Jmol  
Anthocyanidin-3,5-diglucoside
- C06363 KegDraw Jmol  
Anthocyanidin 3-glucoside-5-hydroxycinnamoylglucoside
- C10237 KegDraw Jmol  
Proanthocyanidin A2; Epicatechin-(2beta->7,4beta->8)-epicatechin

... » display all

KEGG REACTION

- R03797 UDPglucose:anthocyanidin 3-O-beta-D-glucosyltransferase; UDP-glucose + Anthocyanidin <=> UDP + Anthocyanidin-3-O-beta-D-glucoside
- R05163 Hydroxycinnamoyl-CoA:anthocyanidin 3,5-diglucoside 5-O-glucoside-6'''-O-hydroxycinnamoyltransferase; Hydroxycinnamoyl-CoA + Anthocyanidin-3,5-diglucoside <=> CoA + Anthocyanidin 3-glucoside-5-hydroxycinnamoylglucoside
- R06534 UDP-glucose:anthocyanidin 3-O-D-glucosyltransferase; Pelargonidin + UDP-glucose <=> Pelargonidin 3-O-glucoside + UDP
- R06535 UDP-glucose:anthocyanidin 3-O-D-glucosyltransferase; Cyanidin + UDP-glucose <=> Chrysanthemin + UDP
- R06536 UDP-glucose:anthocyanidin 3-O-D-glucosyltransferase; Delphinidin + UDP-glucose <=> Delphinidin 3-O-glucoside + UDP

... » display all

KEGG ENZYME

- 1.3.1.77 anthocyanidin reductase; AtANR; MtANR
- 1.14.11.19 leucocyanidin oxygenase; anthocyanidin synthase



Search COMPOUND for anthocyanidin

Database: COMPOUND - Search term: anthocyanidin (Total 8 hits)

- C02003 KegDraw Jmol  
Anthocyanidin
- C03940 KegDraw Jmol  
Anthocyanidin-3-O-beta-D-glucoside; Anthocyanidin-3-O-D-glucoside
- C06361 KegDraw Jmol  
Anthocyanidin-3,5-diglucoside
- C06363 KegDraw Jmol  
Anthocyanidin 3-glucoside-5-hydroxycinnamoylglucoside
- C10237 KegDraw Jmol  
Proanthocyanidin A2; Epicatechin-(2beta->7,4beta->8)-epicatechin
- C15539 KegDraw Jmol  
Anthocyanidin 3-O-beta-D-glucoside
- C15540 KegDraw Jmol  
Anthocyanidin 3-O-(6-O-malonyl-beta-D-glucoside)
- C17624 KegDraw Jmol  
Cinnamtannin A1; Proanthocyanidin C1; Procyanidin C1

DBGET integrated database retrieval system

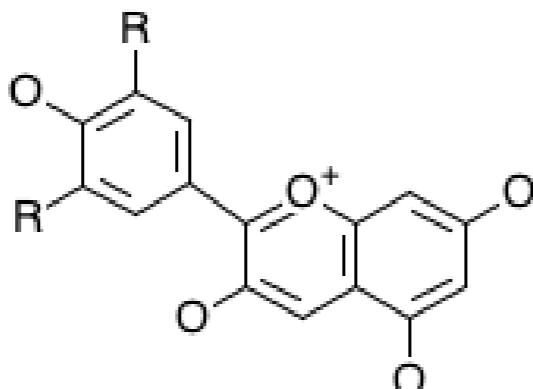
ヒット数が多いと検索結果が省略される  
display allをクリックすると全部表示される

# KEGG COMPOUNDのエントリ



COMPOUND: C02003

Help

Entry	C02003	Compound
Name	Anthocyanidin	
Formula	C15H9O5R2	
Structure	 C02003	
	<a href="#">Mol file</a> <a href="#">KCF file</a> <a href="#">DB search</a> <a href="#">Jmol</a> <a href="#">KegDraw</a>	
Comment	Type: generic compound in reaction hierarchy	
Reaction	<a href="#">R03797</a> <a href="#">R07351</a> <a href="#">R07352</a>	
Enzyme	<a href="#">1.3.1.77</a> <a href="#">2.4.1.115</a>	
Other DBs	PubChem: 5099 ChEBI: 16366	
KCF data	<a href="#">Show</a>	

» Japanese version

# KEGG REACTIONのエントリ



## REACTION: R03797

[Help](#)

Entry	R03797	Reaction
Name	UDPglucose:anthocyanidin 3-O-beta-D-glucosyltransferase	
Definition	UDP-glucose + Anthocyanidin $\rightleftharpoons$ UDP + Anthocyanidin-3-O-beta-D-glucoside	
Equation	$C00029 + C02003 \rightleftharpoons [C00015] + C03940$	
Comment	general reaction (see R06534, R06535 and R06536)	
RPair	RP00012 C00015_C00029 main RP03419 C02003_C03940 main RP07728 C00029_C03940 trans	
Enzyme	<a href="#">2.4.1.115</a>	

# KEGG ENZYMEのエントリ



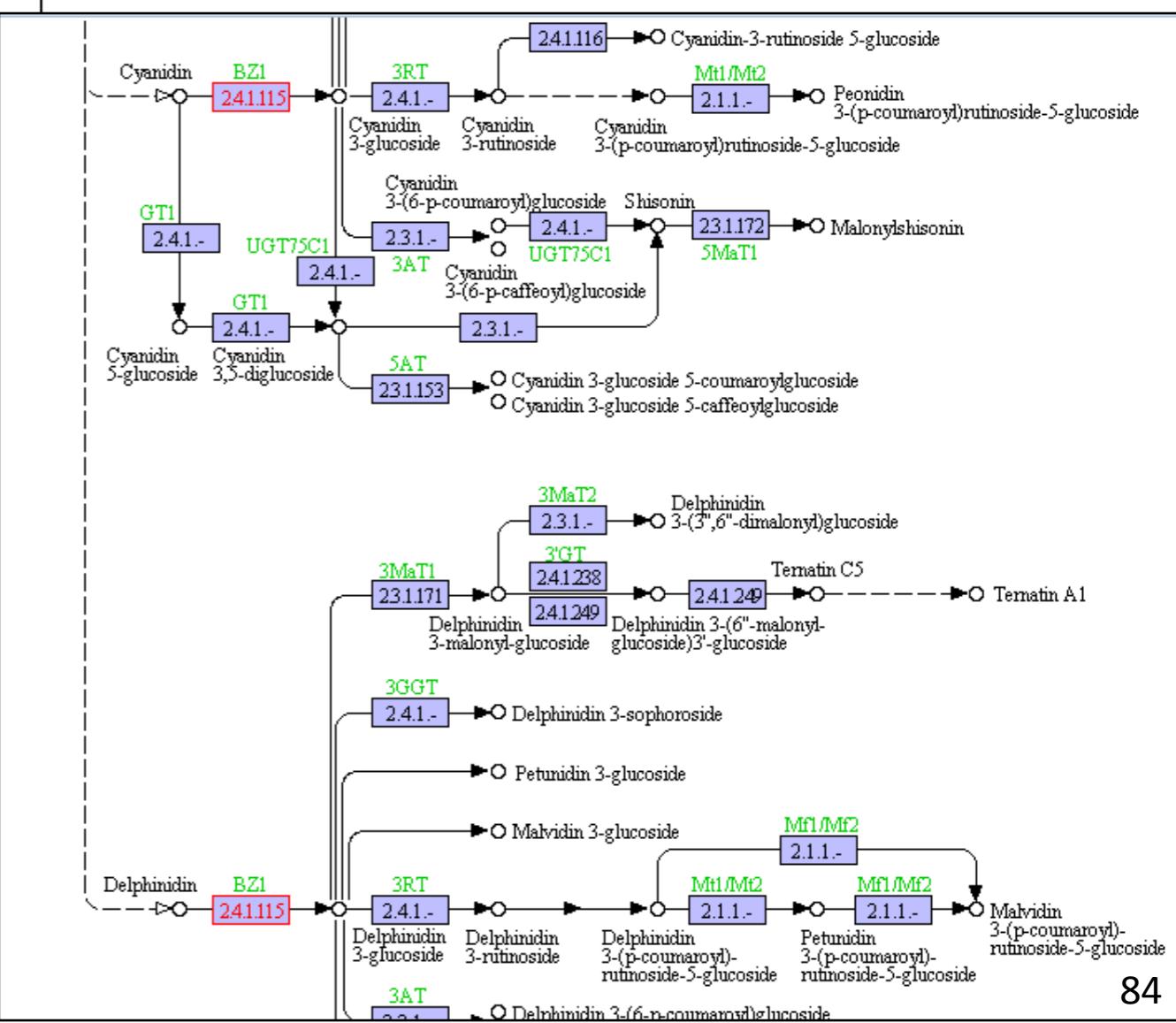
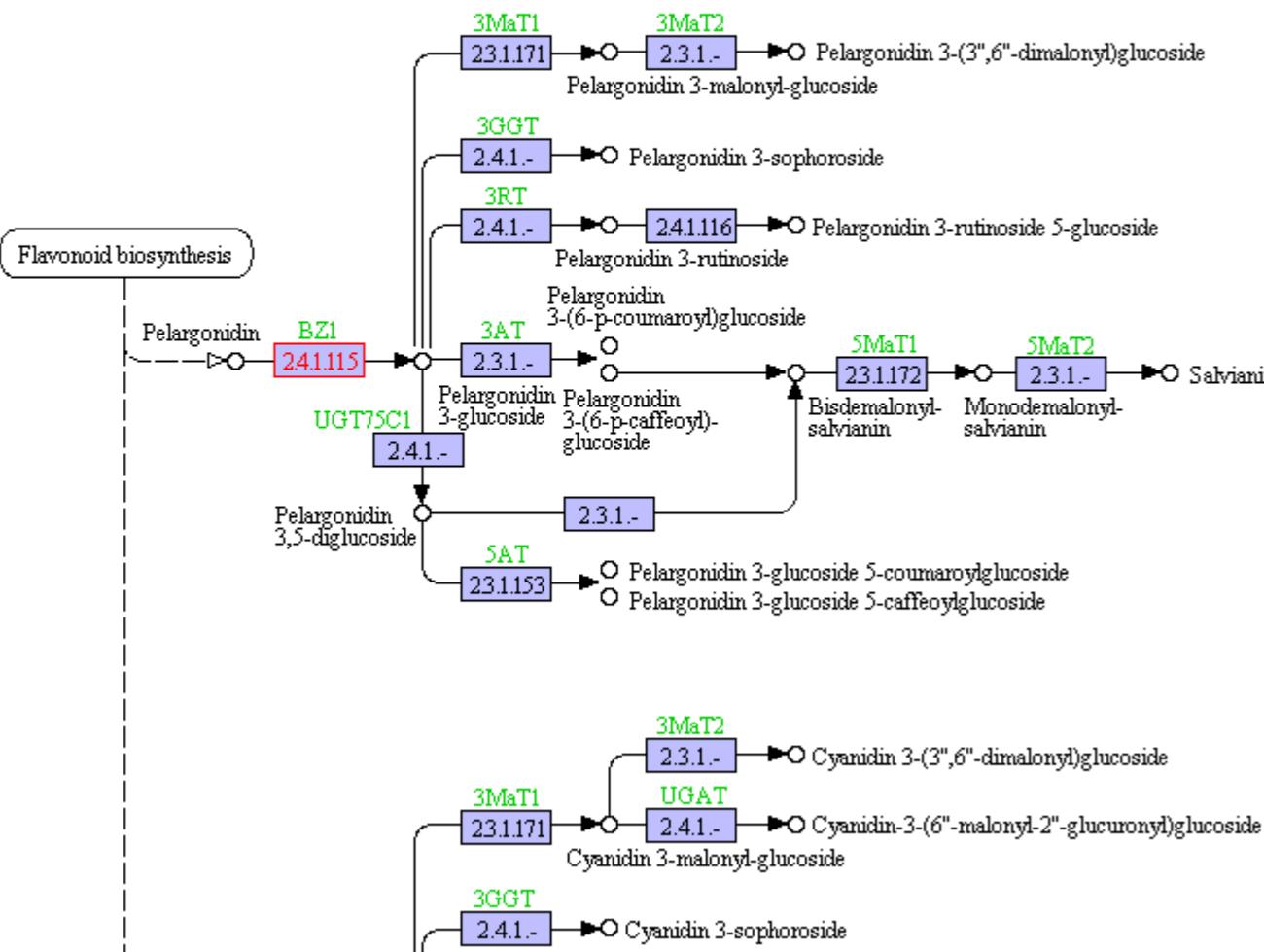
ENZYME: 2.4.1.115

Help

Entry	EC 2.4.1.115 Enzyme
Name	anthocyanidin 3-O- $\alpha$ -glucosyltransferase; uridine diphosphoglucose-anthocyanidin 3-O- $\alpha$ -glucosyltransferase; UDP- $\alpha$ -glucose:anthocyanidin/flavonol 3-O- $\alpha$ -glucosyltransferase; UDP- $\alpha$ -glucose:cyanidin-3-O- $\alpha$ -glucosyltransferase; UDP- $\alpha$ -glucose:anthocyanidin 3-O-D- $\alpha$ -glucosyltransferase; 3-GT
Class	Transferases; Glycosyltransferases; Hexosyltransferases <a href="#">BRITE hierarchy</a>
Sysname	UDP-D- $\alpha$ -glucose:anthocyanidin 3-O-beta-D- $\alpha$ -glucosyltransferase
Reaction(IUBMB)	UDP-D- $\alpha$ -glucose + an anthocyanidin = UDP + an anthocyanidin-3-O-beta-D- $\alpha$ -glucoside [RN:R03797]
Reaction(KEGG)	R03797 > R06534 R06535 R06536 <a href="#">Show all</a>
Substrate	UDP-D- $\alpha$ -glucose [CPD:C00029]; anthocyanidin [CPD:C02003]
Product	UDP [CPD:C00015]; anthocyanidin-3-O-beta-D- $\alpha$ -glucoside [CPD:C03940]
Comment	The anthocyanidin compounds cyanidin, delphinidin, peonidin and to a lesser extent pelargonidin can act as substrates. The enzyme does not catalyse glucosylation of the 5-position of cyanidin and does not act on flavanols such as quercetin and kaempferol (cf. EC 2.4.1.91 flavonol 3-O- $\alpha$ -glucosyltransferase). In conjunction with EC 1.14.11.19, leucocyanidin oxygenase, it is involved in the conversion of leucoanthocyanidin into anthocyanidin 3- $\alpha$ -glucoside. It may act on the pseudobase precursor of the anthocyanidin rather than on the anthocyanidin itself. [2]
Pathway	<a href="#">ec00942 Anthocyanin biosynthesis</a> <a href="#">ec01100 Metabolic pathways</a> <a href="#">ec01110 Biosynthesis of secondary metabolites</a>
Orthology	K12930 anthocyanidin 3-O- $\alpha$ -glucosyltransferase
Genes	POP: P0PTP_287596(FGT2) RCU: RCOM_1324500 WV: 100233099(UFGT)
	<a href="#">Taxonomy</a>
Reference	1 [PMID:751640]

# KEGG PATHWAYのエントリ

## ANTHOCYANIN BIOSYNTHESIS



# SIMCOMP/SUBCOMP



COMPOUND: C15549

Entry C15549 Compound  
Name Anthocyanin  
Formula C15H11O  
Exact mass 207.081  
Mol weight 207.2472

Structure

Mol file KCF file DB search Jmol KegDraw

Comment Type: generic compound in reaction hierarchy  
Reaction R07266  
Enzyme 2.4.1.238  
Other DBs CAS: 11029-12-2 PubChem: 17396541 NIKKAI: J101.736E

KCF data Show

SIMCOMP Search

SUBCOMP SIMCOMP SIMCOMP2 KEGG2

Compute Return

Query C02003

Structure

Database  COMPOUND  DRUG  KNAPSAcK  REACTION  
Options  Global search  Local search  Customized search  
Feedback KEGG GenomeNet Kyoto University Bioinformatics Center

類似化合物の検索ツール

SIMCOMP Search Result

Database : KEGG COMPOUND Number of entries in a page 20 Hide structure

Page : 1 Go of 46 Items : 1 - 20 of 920 Top Previous Next Bottom

Top 10 Clear Select operation Exec

No	Entry	Structure	Name
1	C02003		Anthocyanidin
2	C15598		2,3-cis-2R,3R-Flavan-3-ol Flavan-3-ol

# 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

KEGG PATHWAY Database  
Wiring diagrams of molecular interactions, reactions, and relations

KEGG2 PATHWAY BRITE MODULE DISEASE DRUG KO GENOME GENES LIGAND DBGET

Select prefix  Enter keywords   Help

**Pathway Maps**

KEGG PATHWAY is a collection of manually drawn pathway maps (see [new maps](#) and [update history](#)) representing our knowledge on the molecular interaction and reaction networks for:

**0. Global Map** (highlighted with a red box)

1. Metabolism Carbohydrate Energy Lipid Nucleotide Amino acid Other amino acid Glycan Cofactor/vitamin Terpenoid/PK Other secondary metabolite Xenobiotics Overview

2. Genetic Information Processing

3. Environmental Information Processing

4. Cellular Processes

5. Organismal Systems

6. Human Diseases

and also on the structure relationships (KEGG drug structure maps) in:

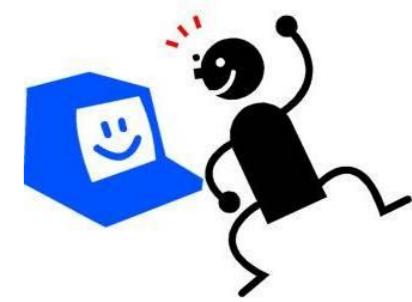
7. Drug Development

**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

**0. Global Map**

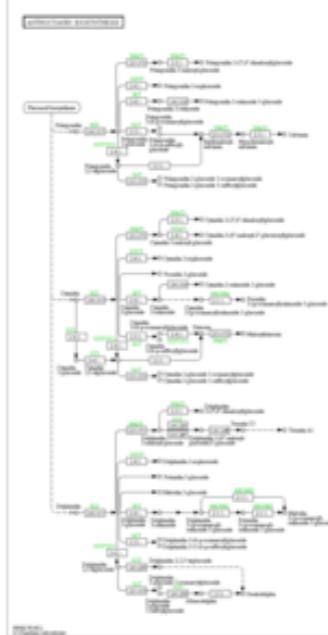
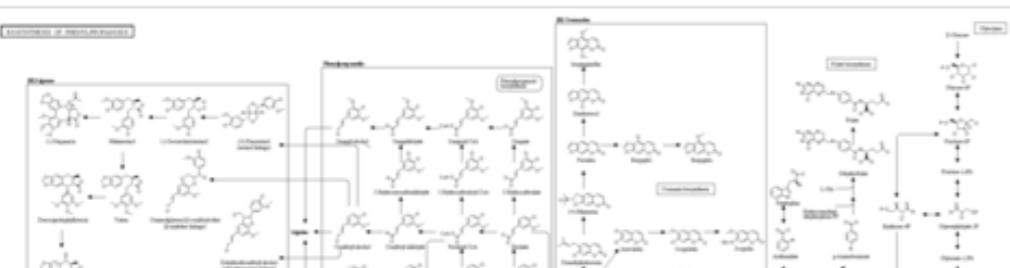


# キーワード”anthocyanidin”で検索

## Pathway Text Search

Number of entries in a page 20 ▾ Hide thumbnail

Items : 1 - 3 of 3

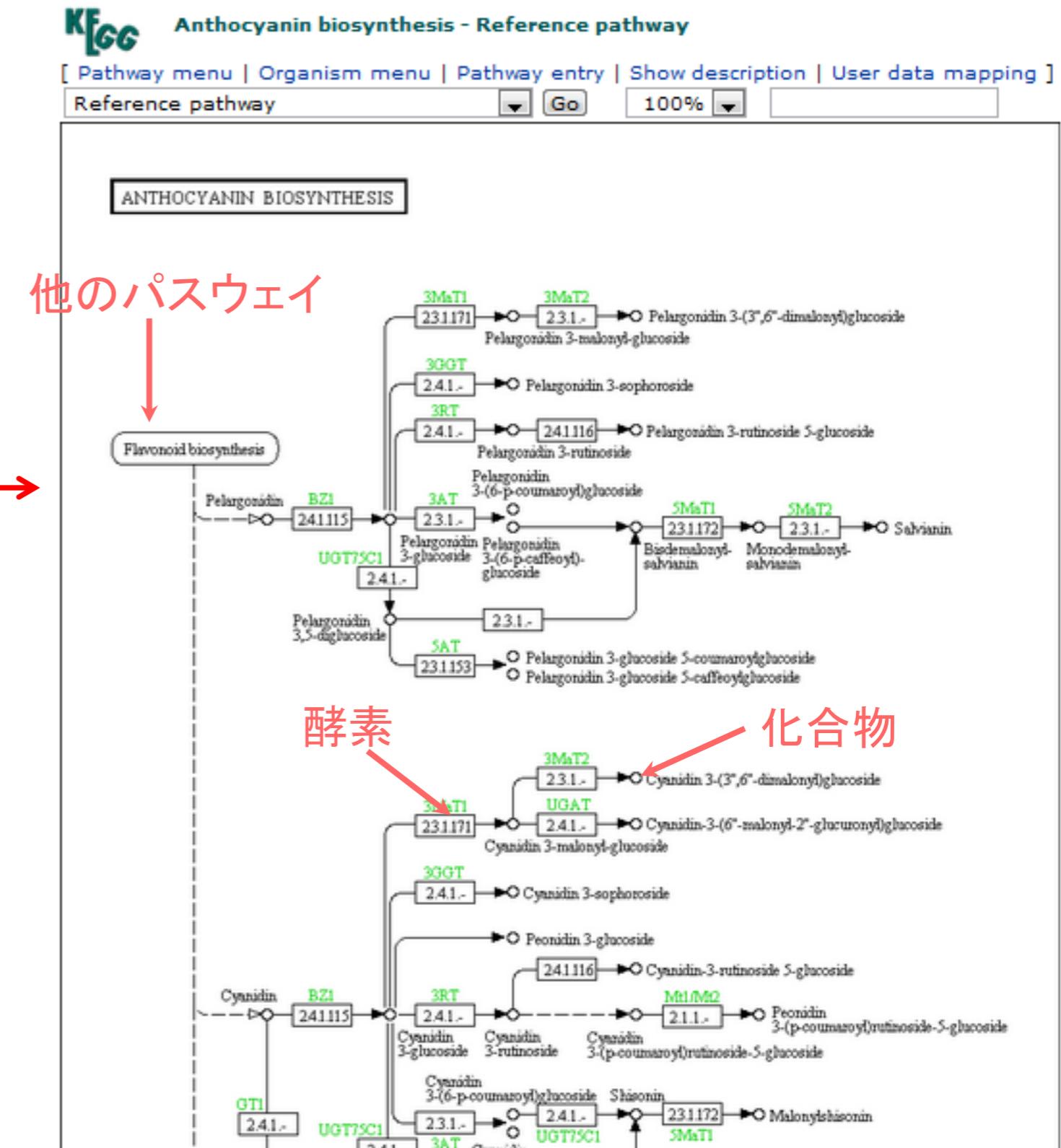
Entry	Thumbnail Image	Name	Description	Object	Legend
map00942		Anthocyanin biosynthesis	<p>Anthocyanidins (aglycones) and anthocyanins (glycosides) are common plant pigments and belong to a s...</p> <p>C16370 (Delphinidin 3-(6-p-coumaroyl)glucoside) C16368 (Pelargonidin 3-(6-p-coumaroyl)glucoside) C16...</p>	C16370 (Delphinidin 3-(6-p-coumaroyl)glucoside) C16368 (Pelargonidin 3-(6-p-coumaroyl)glucoside) Pelargonidin 3-glucos...	Delphinidin 3-(6-p-coumaroyl)glucoside Pelargonidin 3-(6-p-coumaroyl)glucoside Pelargonidin 3-glucos...
				C18083	linkage1 /resinol

# KEGG PATHWAY MAP



PATHWAY: map00942

<b>Entry</b>	map00942	Pathway
<b>Name</b>	Anthocyanin biosynthesis	
<b>Description</b>	Anthocyanidins (aglycones) and anthocyanins (glycosides) are common plant pigments and belong to a structural subclass of flavonoids characterized by a 2-phenylbenzopyrylium unit. They are derived along the flavonoid modification pathways and further separated into three types, pelargonidin, cyanidin, and delphinidin, due to the different number of hydroxyl groups in the phenyl group.	
<b>Class</b>	Metabolism; Biosynthesis of Other Secondary Metabolites <small>(BRITE hierarchy)</small>	
<b>Pathway map</b>	map00942 Anthocyanin biosynthesis	
<a href="#">All organisms</a> <a href="#">Ortholog table</a>		
<b>Other DBs</b>	GO: 0009718	
<b>Reference</b>	PMID: 17383962	
<b>Authors</b>	Unno H, Ichimaiwa F, Suzuki H, Takahashi S, Tanaka Y, Saito A, Nishino T, Kusunoki M, Nakayama T.	
<b>Title</b>	Structural and mutational studies of anthocyanin malonyltransferases establish the features of BAHD enzyme catalysis.	
<b>Journal</b>	J Biol Chem 282:15812-22 (2007)	
<b>Reference</b>	PMID: 16887235	
<b>Authors</b>	Kogawa K, Kazuma K, Kato N, Noda N, Suzuki M.	
<b>Title</b>	Biosynthesis of malonylated flavonoid glycosides on the basis of malonyltransferase activity in the petals of <i>Clitoria ternatea</i> .	
<b>Journal</b>	J Plant Physiol 164:886-94 (2007)	
<b>Reference</b>	PMID: 16945088	



# KEGG PATHWAY MAP



## Anthocyanin biosynthesis - Reference pathway

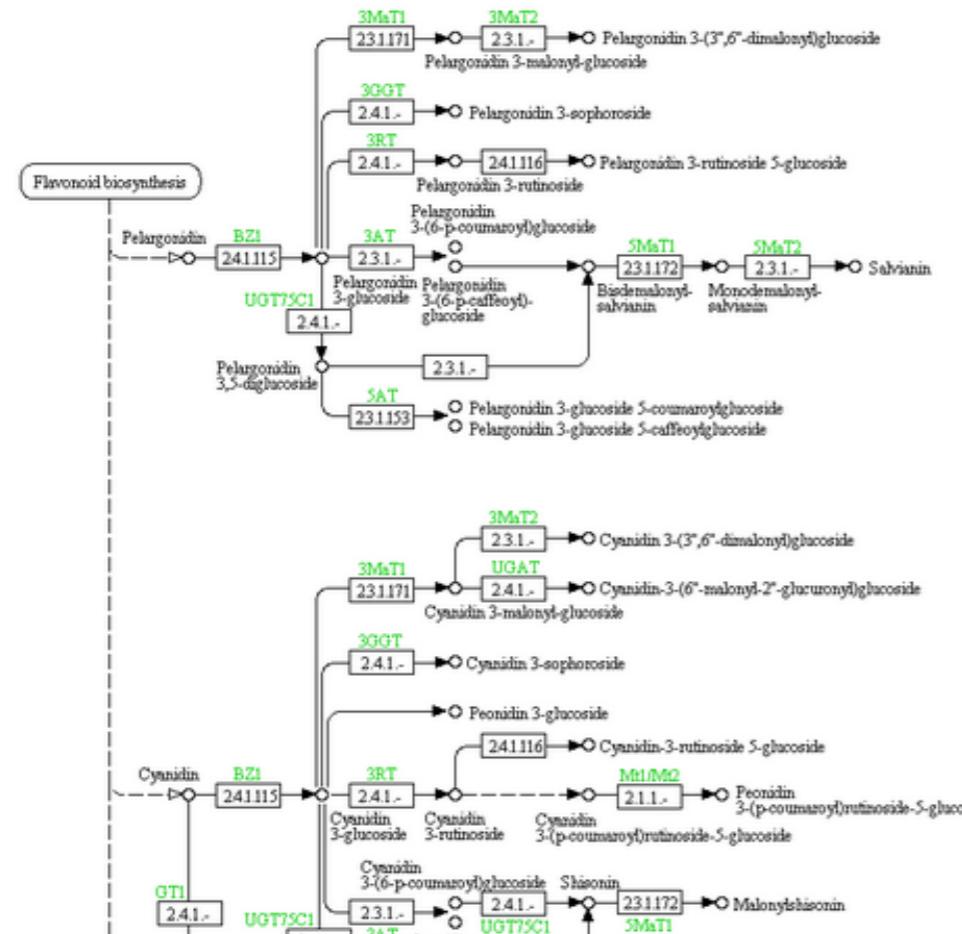
[ Pathway menu | Organism menu | Pathway entry | Show description | User data mapping ]

Reference pathway

Go

100%

### ANTHOCYANIN BIOSYNTHESIS



## Anthocyanin biosynthesis - *Vitis vinifera* (wine grape)

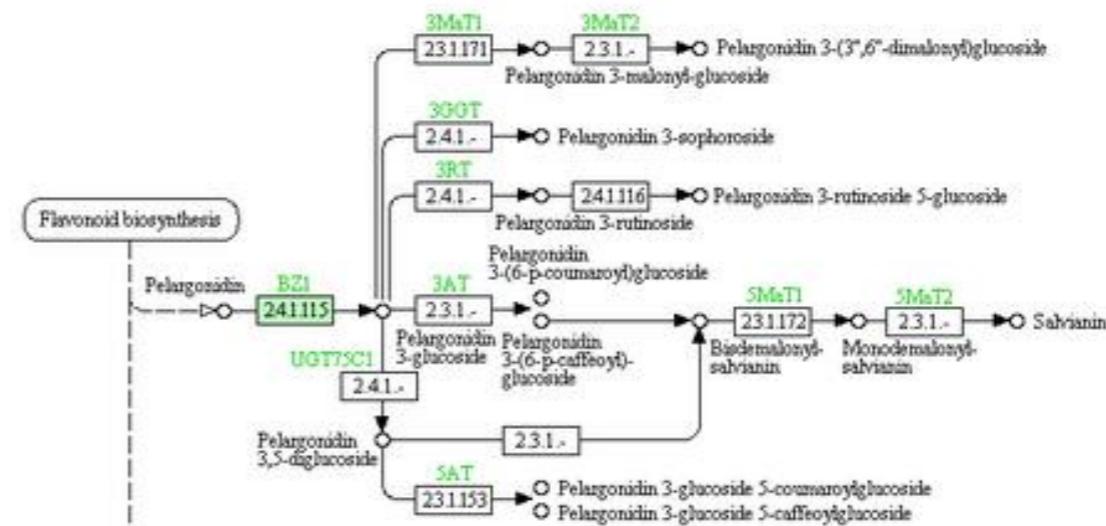
[ Pathway menu | Organism menu | Pathway entry | Download KGML | Show description | User data ]

*Vitis vinifera* (wine grape)

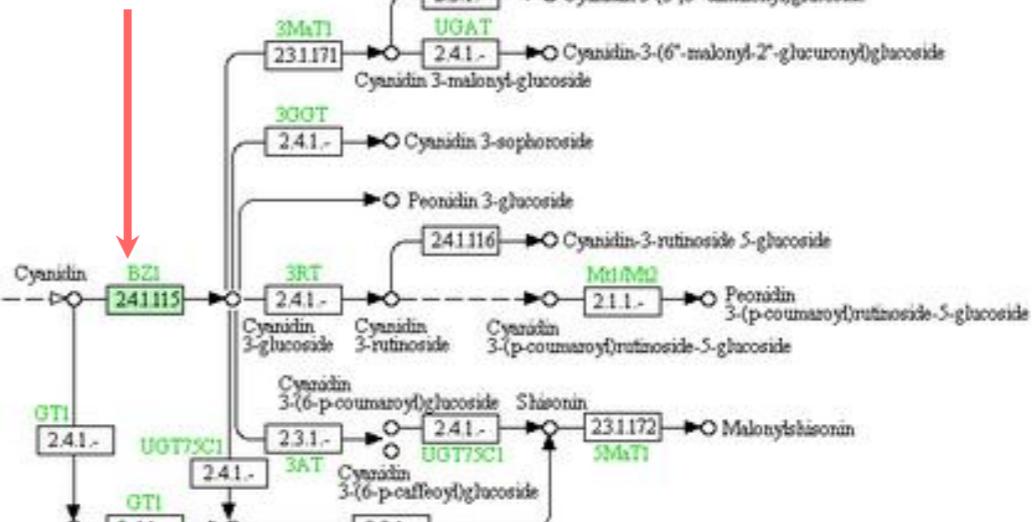
Go

100%

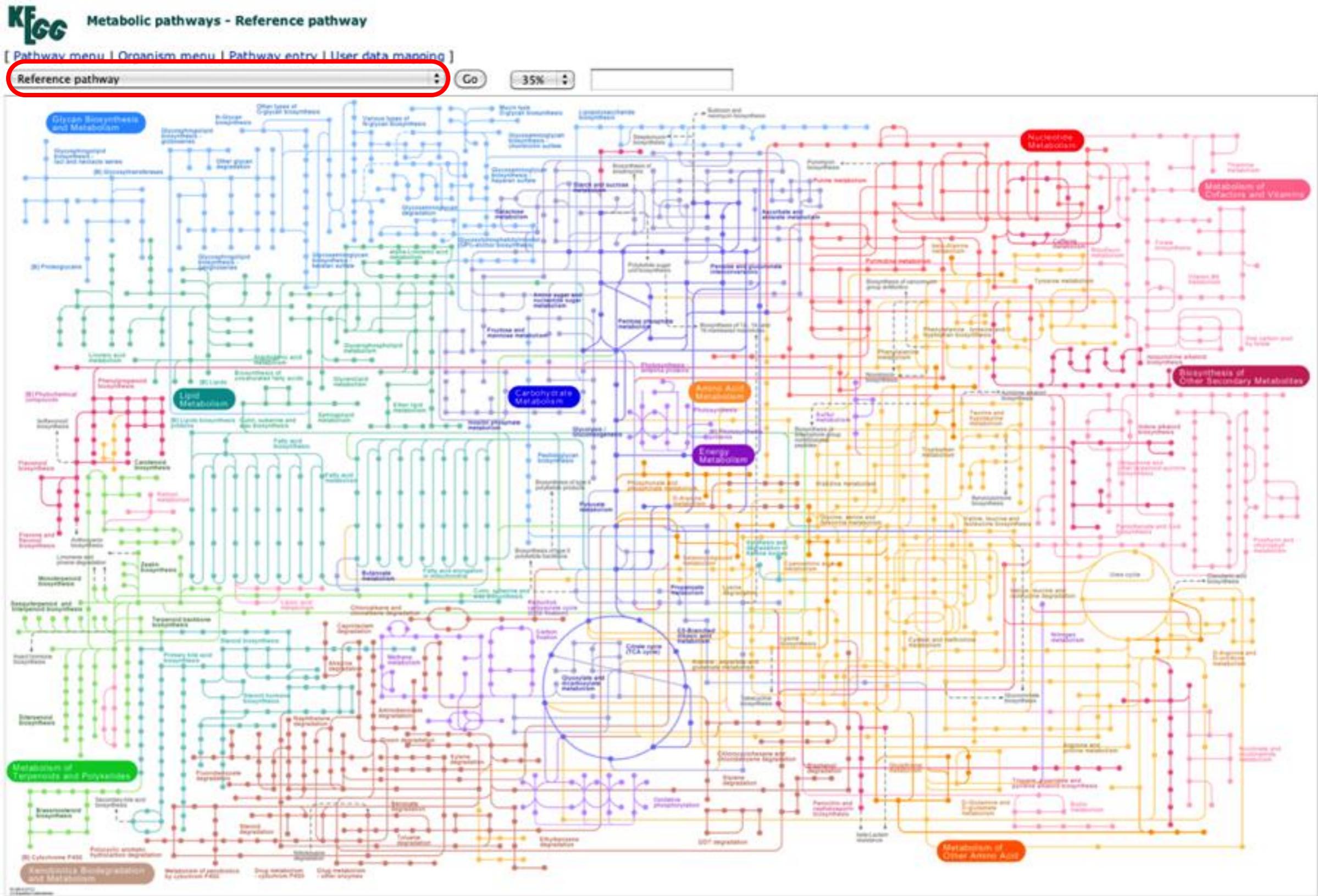
### ANTHOCYANIN BIOSYNTHESIS



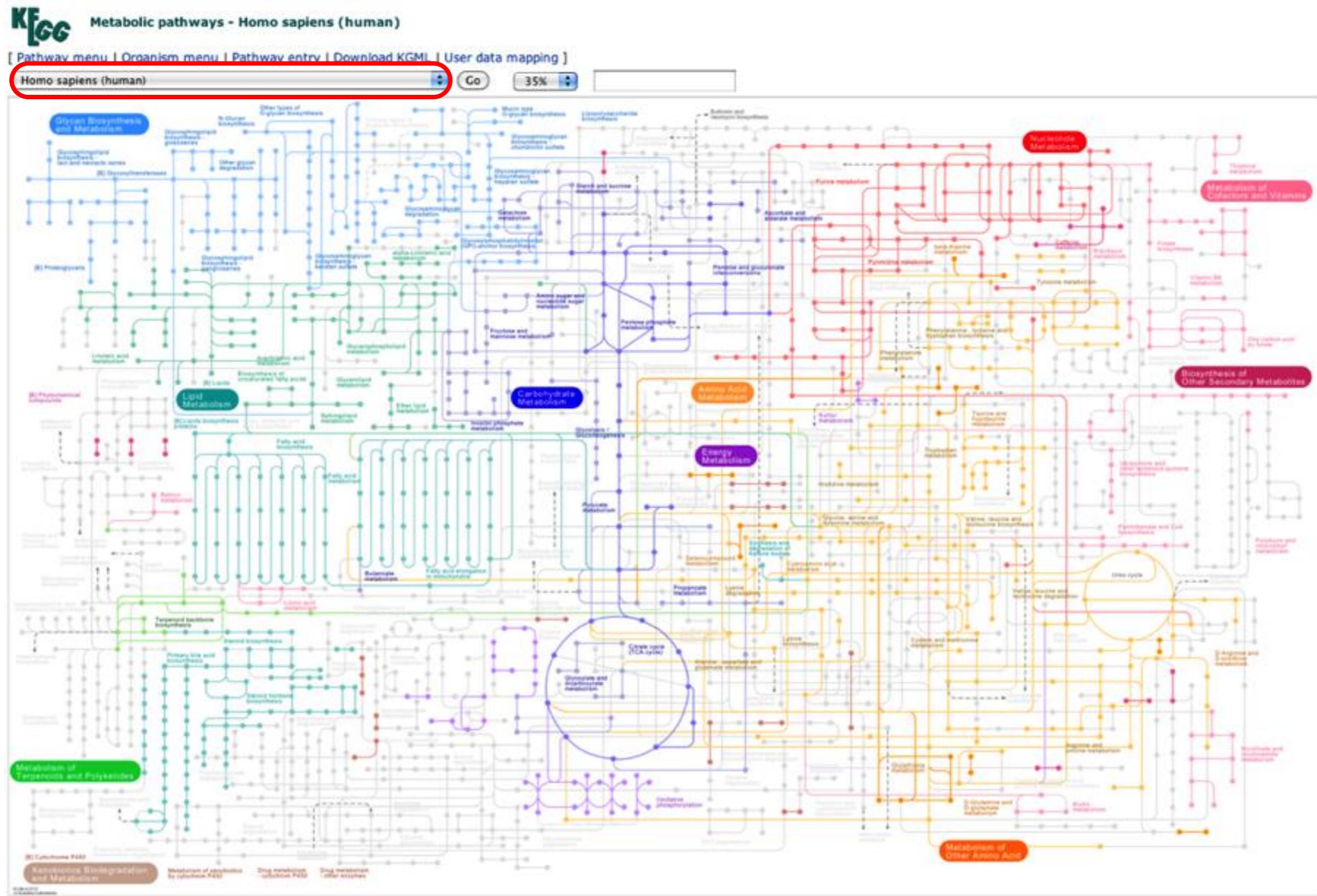
この生物がもつ酵素



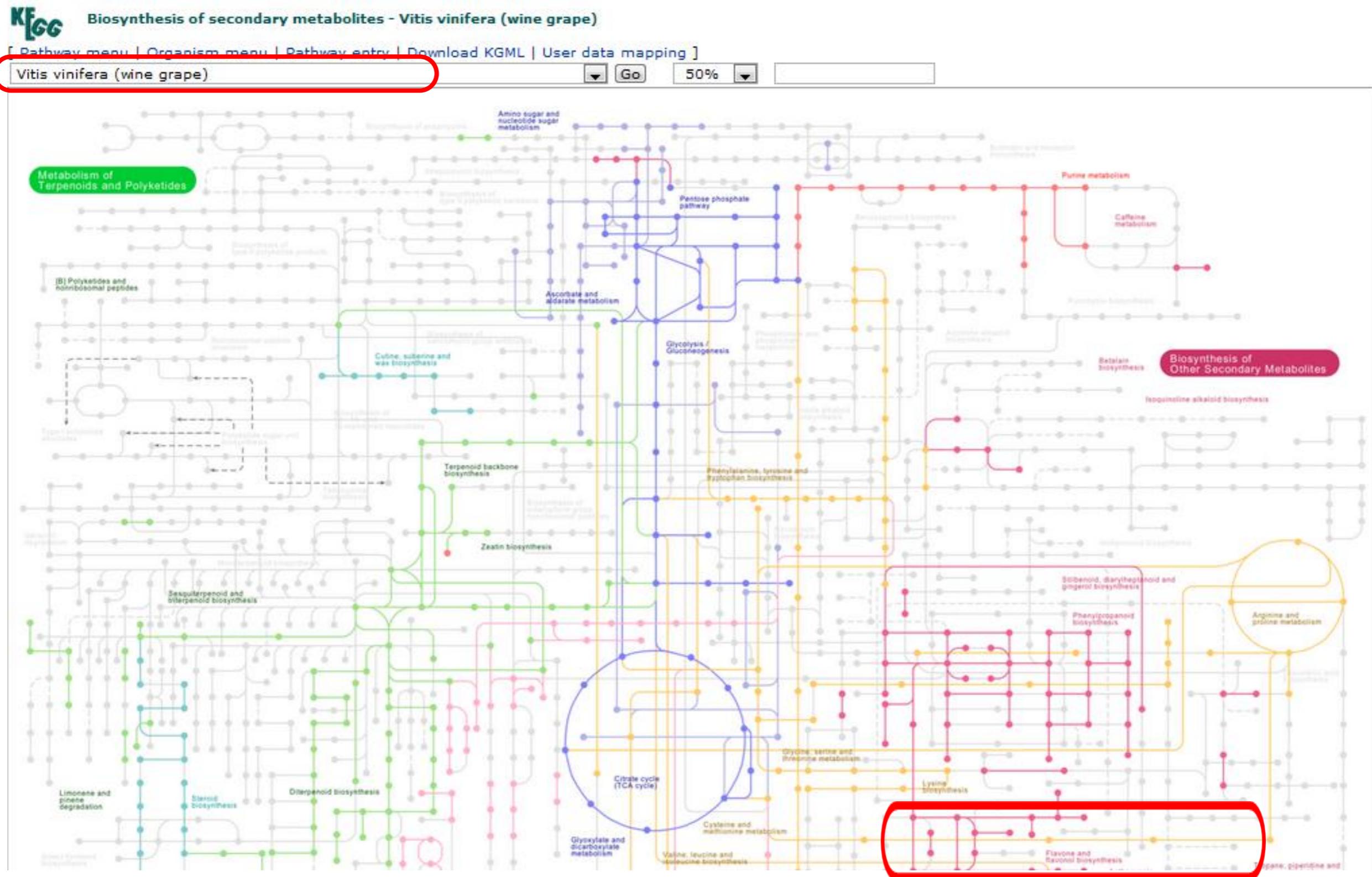
# Global Map



# Global Map (Homo sapiens)



# Global Map (Vitis Vinifera)



この辺りにアントシアニン合成系がある

# 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

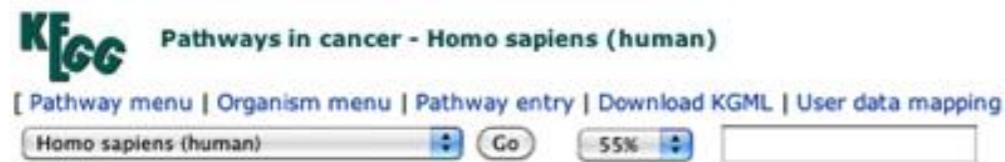


#hsa	COSMIC
hsa:25	678
hsa:861	26
hsa:867	13
hsa:1029	44
hsa:1050	4
hsa:1436	3
hsa:2146	9
hsa:2322	7
hsa:2623	2
hsa:2624	9
hsa:2956	2
hsa:3717	40
hsa:3718	2
hsa:3815	5
hsa:3845	6
hsa:4436	1
hsa:4869	1
hsa:4893	15
hsa:5290	1
hsa:7157	23
hsa:7403	4
hsa:7490	2
hsa:10320	3
hsa:54790	5
hsa:55294	1
hsa:171023	6

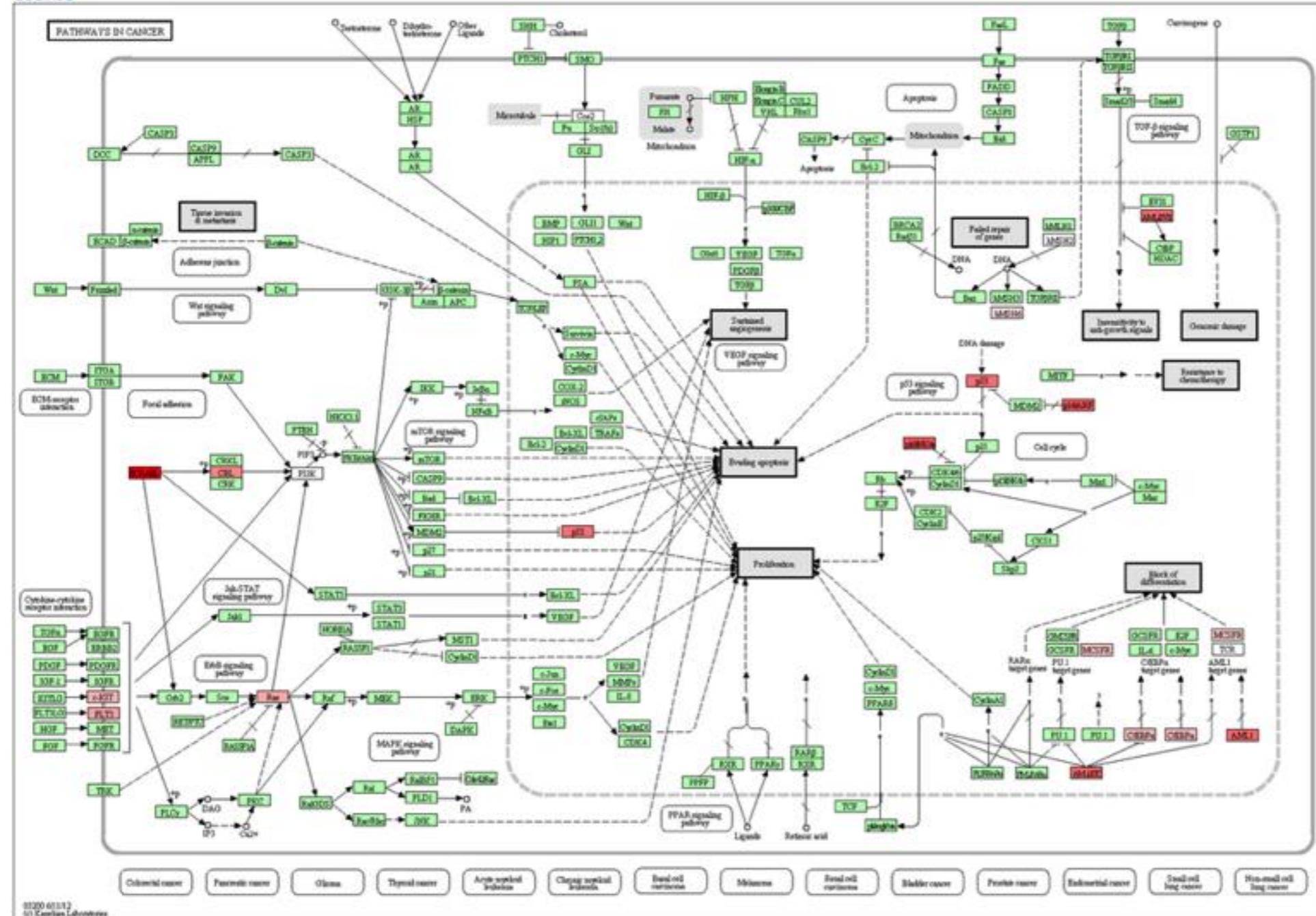
### 3. 保存したファイルを選択

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

# KEGG Mapperの結果



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

# 統合TV: ここまででの参考動画

- ・ KEGG
  - ・ KEGG、GenomeNetのサービス・利用法
  - ・ KEGG Atlas と KAAS アノテーション ~ KEGG における最近の開発から





# おわりに

---

- ◆ DDBJ, PDBj, KEGG共に、今回は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ってみてください