



Certificat International en Bioinformatique et Génomique

CIBiG 2024

9 September - 4 October 2024
WAVE center - Abidjan

<https://cibig-wave.github.io/>
bioinfo@wave-center.org



CIBiG in a few numbers

- Born from a **strong partnership of nearly 10 years**
- An **international coordination** led by **4 institutes** (3 countries)
- An international teaching team of **16 instructors**
- An international admissions committee of 6 members

Impossible without the logistical support from the WAVE center !

Coordination

Justin PITA
Fidèle TIENDREBOGO
Ezechiel Bionimian TIBIRI
Serge ANOMAN

Support : Mrs. Dago

Romaric K. NANEMA



CIBiG 2024
organization



Teaching team

Justin PITA
Fidèle TIENDREBOGO
Ezechiel Bionimian TIBIRI
Serge ANOMAN
Support : Mrs. Dago



Romaric K. NANEMA
Seydou KONSIBO
Ousmane BARRA



Palwendé Romuald BOUA



CIBiG 2024 organization

Marie COUDERC
Julie ORJUELA-BOUNIOL
François SABOT
Ndomassi TANDO
Christine TRANCHANT-DUBREUIL



Comte Aurore
Martine BANGRATZ
Sébastien CUNNAC
Dereepper Alexis
Sebastien RAVEL



+ seminar & internship



All members & 3 bioinformatics resources

Justin PITA
Fidèle TIENDREBOGO
Ezechiel Bionimian TIBIRI
Serge ANOMAN

Support : Mrs. Dago



Romaric K. NANEMA
Seydou KONSIBO
Ousmane BARRA



Palwendé Romuald BOUA



CIBiG 2024
organization

Marie COUDERC
Julie ORJUELA-BOUNIOL
François SABOT
Ndomassi TANDO
Christine TRANCHANT-DUBREUIL



Comte Aurore
Martine BANGRATZ
Sébastien CUNNAC
Dereepper Alexis
Sebastien RAVEL

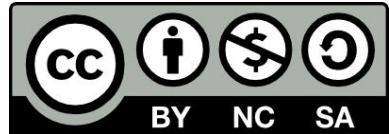


BBI platform

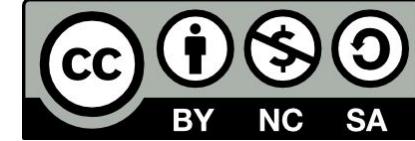


CIBiG information

- All our courses:
 - Description <https://cibig-wave.github.io/01-description.html>
 - Session 2024 : <https://cibig-wave.github.io/05-2024.html>



The educational material used for these teachings is made available under the terms of the Creative Commons Attribution - NonCommercial - ShareAlike (BY-NC-SA) 4.0 International license."
<http://creativecommons.org/licenses/by-nc-sa/4.0/>



Biological databanks and resources

CIBIG 2024

9 September - 4 October 2024
WAVE center - Abidjan

<https://cibig-wave.github.io/>
bioinfo@wave-center.org

Trainers



Ezechiel Bionimian Tibiri
Bioinformatics / Molecular
Biology PhD
INERA Ouagadougou



Alexis Dereeper
Bioinformatics engineer
PHIM - IRD



**Christine
Tranchant-Dubreuil**
Bioinformatics engineer
IR DIADE - IRD

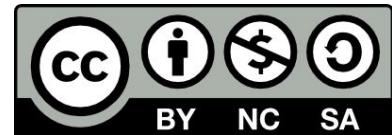
Course materials

Content licensed under CC BY-NC-SA 4.0 & created by :

E. Tibiri, A. Dereeper, C. Tranchant. Formation CIBIG 2024

Adapted from:

“Introduction aux bases de données et aux ressources” E. Tibiri 2024

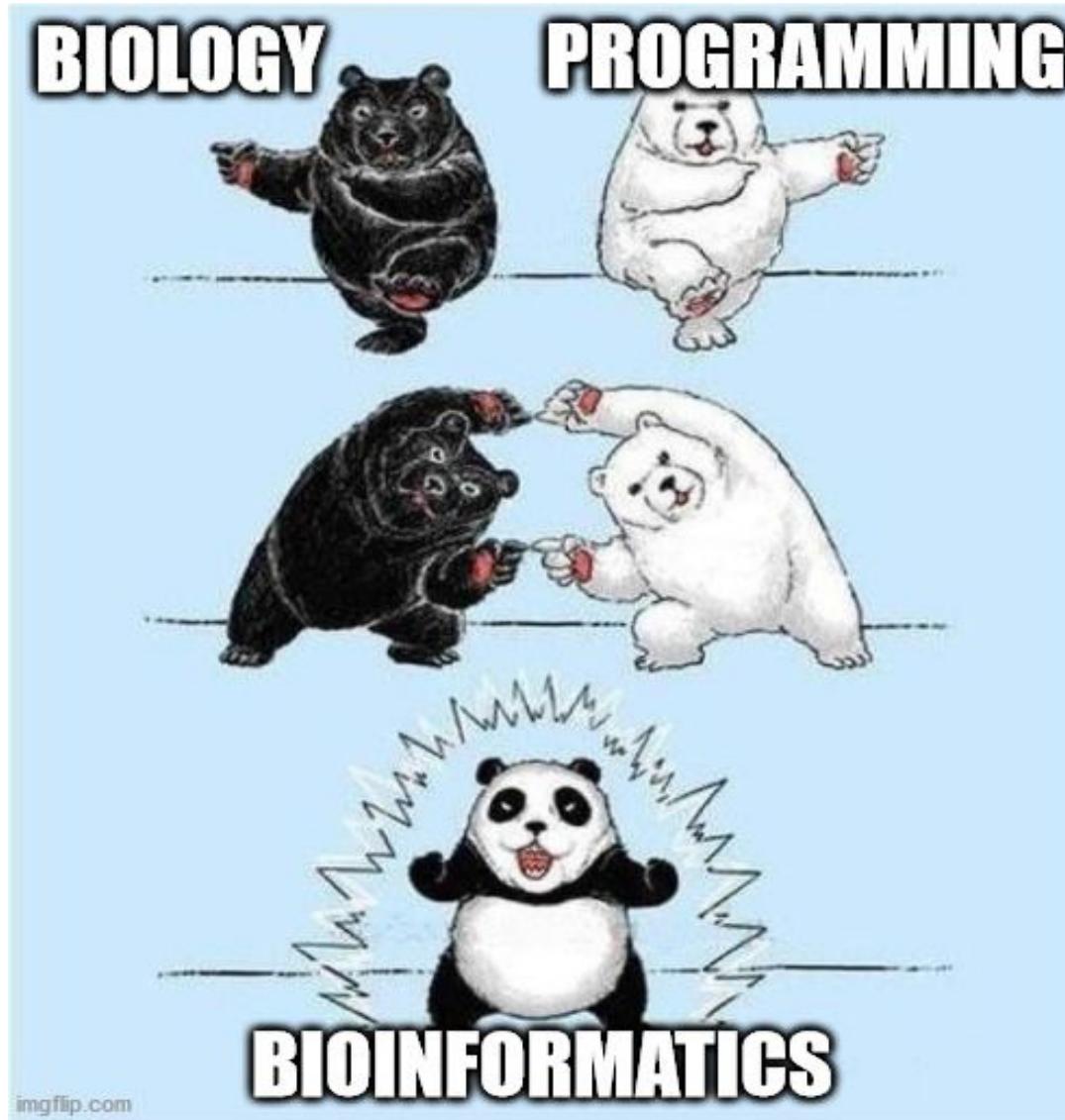


The educational material used for these teachings is made available under the terms of the Creative Commons Attribution - NonCommercial - ShareAlike (BY-NC-SA) 4.0 International license.
<http://creativecommons.org/licenses/by-nc-sa/4.0/>

Objectives of this course

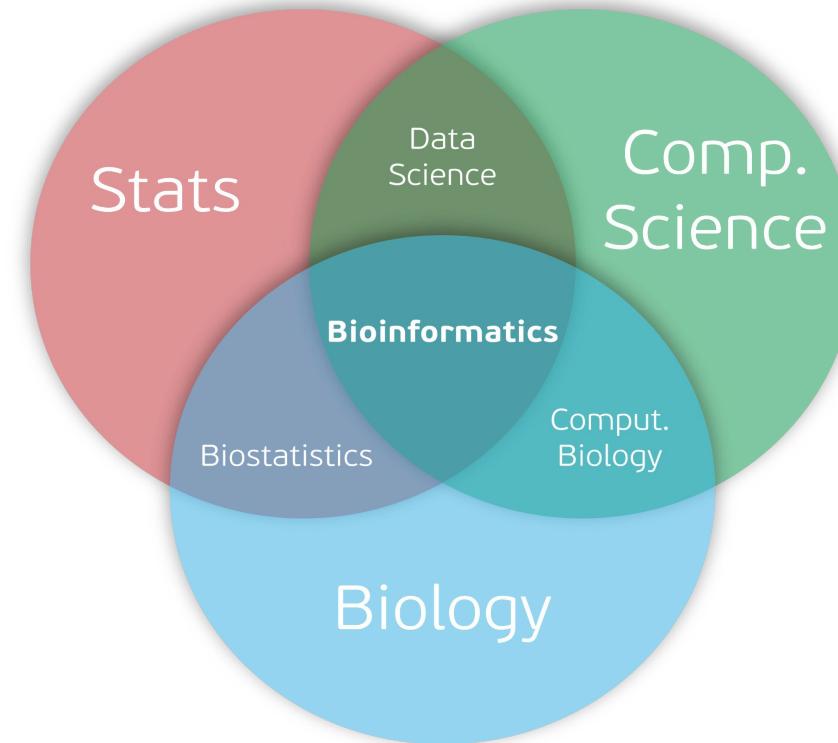
- Understand the structure and layout of data resources at NCBI and EBI
- Understand differences between databases, tools and repositories
- Search of datasets in several databases using accession numbers, gene names, etc.
- Exploit NCBI and EBI resources

What is bioinformatics



imgflip.com

A interdisciplinary science



What is bioinformatics

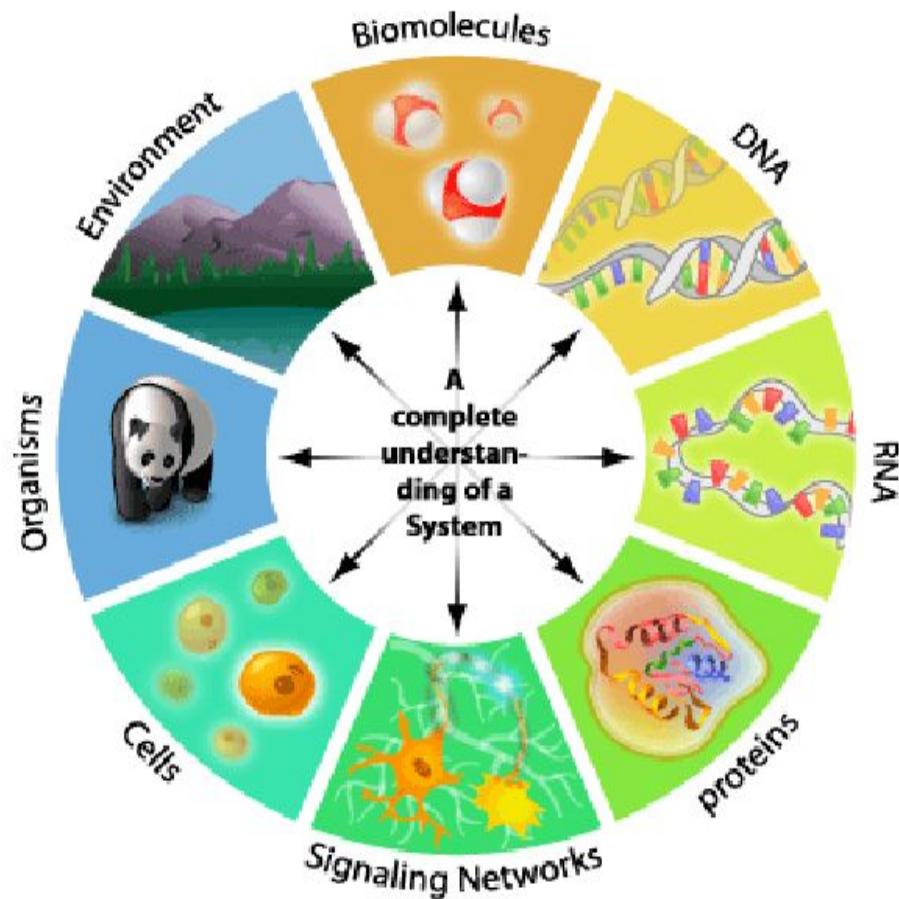
- **Definition of Bioinformatics:** It is the processing of stored biological information in the form of easily accessible and exploitable data.
- **Bioinformatics** is defined as the use of databases and computer algorithms to analyze genes, proteins and the complete collection of DNA (genome) of a living organism.

What is bioinformatics

Bioinformatics is concerned with data:

- The **genome** is the total DNA of an individual or species.
- The **transcriptome** is the set of messenger RNAs transcribed from the genome.
- The **proteome** (all biosynthesized proteins).
- The **metabolome**: all the metabolites produced by a cell (organic molecules that are neither DNA nor protein).

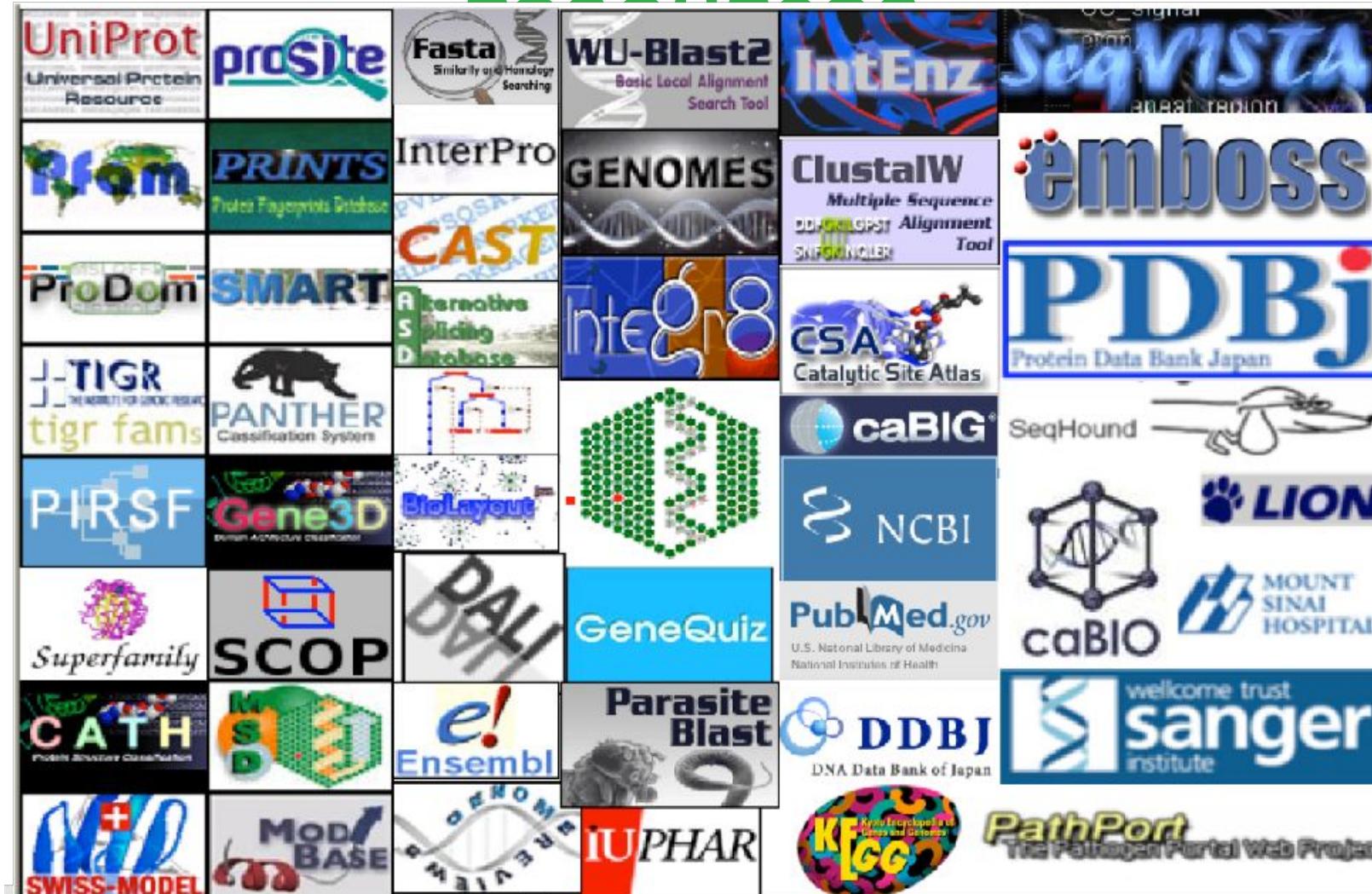
Various types of data



Introduction

- Multiple databases and resources online
- Need to know :
 - What databases and resources exist?
 - What tools are available to exploit these resources?
 - What tools are available for searching resources?

A large panel of databases and



Biological databases

- Databases are convenient system to properly store, search and retrieve any type of data
- Databases are different types based on nature of information and manner (complexity) of data storage

Biological databases

- Biological databases are:
 - Generalist or specialized
 - Primary or secondary
 - Public or private
 - Data can be: Protein, nucleotidic, protein structure, litterature, genome annotation...
 - Focused on sequence (aa or nt) or genome

Some famous biological databanks

- Sequences are nucleic acids (DNA or mRNA); **EMBL, GenBank, DDBJ**
- Sequences are amino acids (proteins); **Swiss-Prot, wwPDB**
- Bibliographic references; **PubMed**
- General informations on genes and diseases; **EntrezGene, OMIM, HGMD**
- informations on tridimensional structure of proteins or DNA; **PDB**

Primary databases

- International Nucleotide Sequence Database Collaboration (INSDC)
- Genomic sequence data stored in 3 public databases: NCBI, EBI and DDBJ
- Each database has their own tools and accession numbers



Primary databases

The GenBank database is designed to provide and encourage access within the scientific community to the most up to date and comprehensive DNA sequence information. Therefore, NCBI places no restrictions on the use or distribution of the GenBank data. However, some submitters may claim patent, copyright, or other intellectual property rights in all or a portion of the data they have submitted.

Secondary databases

- Specialized databases built from primary databases
- Result of analyses of sequences in the primary database
- Secondary database analyzed in a variety of ways and contain different information in different formats
- Provide multiple resources and different annotations
- Examples: Prosite, Prints, Blocks

Secondary Databases



BLOCKS

PRINTS



The most popular biological resources

- National Centre for Biotechnology Information (NCBI)



- European Bioinformatics Institute (EMBL-EBI)



Search into biological database: NCBI

NIH National Library of Medicine
National Center for Biotechnology Information

All Databases

[NCBI Home](#)

[Resource List \(A-Z\)](#)

[All Resources](#)

[Chemicals & Bioassays](#)

[Data & Software](#)

[DNA & RNA](#)

[Domains & Structures](#)

[Genes & Expression](#)

[Genetics & Medicine](#)

[Genomes & Maps](#)

[Homology](#)

[Literature](#)

[Proteins](#)

[Sequence Analysis](#)

[Taxonomy](#)

[Training & Tutorials](#)

[Variation](#)

Welcome to NCBI

The National Center for Biotechnology Information advances science and health by providing access to biomedical and genomic information.

[About the NCBI](#) | [Mission](#) | [Organization](#) | [NCBI News & Blog](#)

Submit Deposit data or manuscripts into NCBI databases 	Download Transfer NCBI data to your computer 	Learn Find help documents, attend a class or watch a tutorial 
Develop Use NCBI APIs and code libraries to build applications 	Analyze Identify an NCBI tool for your data analysis task 	Research Explore NCBI research and collaborative projects 

Popular Resources

[PubMed](#)

[Bookshelf](#)

[PubMed Central](#)

[BLAST](#)

[Nucleotide](#)

[Genome](#)

[SNP](#)

[Gene](#)

[Protein](#)

[PubChem](#)

NCBI News & Blog

Submitting High-Throughput Sequence Data to Gene Expression Omnibus (GEO)
04 Sep 2024
[Submit your transcriptomic and](#)

New Milestone! NCBI Pathogen Detection Reaches 2 Million Isolates
03 Sep 2024
[NCBI's Pathogen Detection resource](#)

Search into biological database: NCBI

← → ⌛ ncbi.nlm.nih.gov

An official website of the United States government [Here's how you know](#) ▾

 **National Library of Medicine**
National Center for Biotechnology Information

ezechiel.tibiri@u...

All Databases ▾

- All Databases
- Assembly
- Biocollections
- BioProject
- BioSample
- Books
- ClinVar
- Conserved Domains
- dbGaP
- dbVar
- Gene**
- Genome
- GEO DataSets
- GEO Profiles
- GTR
- Identical Protein Groups
- MedGen
- MeSH
- NLM Catalog

o NCBI

Center for Biotechnology Information advances science and health by
s to biomedical and genomic information.

[BLI](#) | [Mission](#) | [Organization](#) | [NCBI News & Blog](#)

mit

Download

Transfer NCBI data to your computer

Learn

Find help documents, attend a class or watch a tutorial

Analysis

Research

Popular Resources

- PubMed
- Bookshelf
- PubMed Central
- BLAST
- Nucleotide
- Genome
- SNP
- Gene
- Protein
- PubChem

NCBI News & Blog

Submitting High-Throughput Sequence

NCBI includes more than 30 different databases:

- **Litterature:** PubMed Central (PMC), Bookshelf and PubReader
- **Health:** ClinVar, dbGaP, dbMHC, the Genetic Testing Registry, HIV-1/Human Protein Interaction Database and MedGen
- **Genomes:** BioProject, Assembly, Genome, BioSample, dbSNP, dbVar, Nucleotide, Probe and RefSeq.
- **Genes:** Gene, Gene Expression Omnibus (GEO), HomoloGene, PopSet, Refseq and UniGene.
- **Proteins:** Protein, the Conserved Domain Database (CDD), COBALT, Conserved Domain Architecture Retrieval Tool (CDART), the Molecular Modeling Database (MMDB), Refseqp and Protein Clusters.
- **Chemical products:** Biosystems and PubChem

NCBI : Litterature

NIH National Library of Medicine
National Center for Biotechnology Information

All Databases ▾ **Search**

NCBI Home
Resource List (A-Z)
All Resources
Chemicals & Bioassays
Data & Software
DNA & RNA
Domains & Structures
Genes & Expression
Genetics & Medicine
Genomes & Maps
Homology
Literature
Proteins
Sequence Analysis
Taxonomy
Training & Tutorials
Variation

NIH National Library of Medicine
National Center for Biotechnology Information

PubMed.gov

african rice domestication population genomics × **Search** User Guide

Advanced Create alert Create RSS

Save Email Send to

Sorted by: Best match Display options

9 results Page 1 of 1

1 article found by citation matching

Transcriptome population genomics reveals severe bottleneck and domestication cost in the African rice (*Oryza glaberrima*).
Nabholz B, et al. Mol Ecol. 2014. PMID: 24684265

Transcriptome population genomics reveals severe bottleneck and domestication cost in the African rice (*Oryza glaberrima*).
Nabholz B, Sarah G, Sabot F, Ruiz M, Adam H, Nidelet S, Ghesquière A, Santoni S, David J, Glémén S. Mol Ecol. 2014 May;23(9):2210-27. doi: 10.1111/mec.12738. Epub 2014 Apr 18. PMID: 24684265
The African cultivated rice (*Oryza glaberrima*) was domesticated in West Africa 3000 years ago. ...This work represents the first genome-wide survey of the African rice genetic diversity and paves the way for further comparison between the ...

Domestication history and geographical adaptation inferred from a SNP map of African rice.
Meyer RS, Choi JY, Sanches M, Plessis A, Flowers JM, Amas J, Dorph K, Barreto A, Gross B, Fuller DQ, Bimpang IK, Ndjiondjop MN, Hazzouri KM, Gregorio GB, Purugganan MD. Nat Genet. 2016 Sep;48(9):1083-8. doi: 10.1038/ng.3633. Epub 2016 Aug 8. PMID: 27500524

African rice (*Oryza glaberrima* Steud.) is a cereal crop species closely related to Asian rice (*Oryza*

NCBI : Taxonomy

[Log in](#)[All Databases](#)[Search](#)

NCBI Home
Resource List (A-Z)
All Resources
Chemicals & Bioassays
Data & Software
DNA & RNA
Domains & Structures
Genes & Expression
Genetics & Medicine
Genomes & Maps
Homology
Literature
Proteins
Sequence Analysis
Taxonomy
Training & Tutorials
Variation

Taxonomy Browser

Search for as lock

Display 3 levels using filter: none

[Oryza sativa](#)

Taxonomy ID: 4530 (for references in articles please use NCBI:txid4530)

current name
Oryza sativa L., 1753

Genbank common name: **Asian cultivated rice**
NCBI BLAST name: **monocots**
Rank: **species**
Genetic code: [Translation table 1 \(Standard\)](#)
Mitochondrial genetic code: [Translation table 1 \(Standard\)](#)
Plastid genetic code: [Translation table 11 \(Bacterial, Archaeal and Plant Plastid\)](#)
Other names:
common name(s)
red rice, rice

Lineage (full)
[cellular organisms](#); [Eukaryota](#); [Viridiplantae](#); [Streptophyta](#); [Streptophytina](#); [Embryophyta](#); [Tracheophyta](#); [Euphyllophyta](#); [Spermatophyta](#); [Magnoliopsida](#); [Mesangiospermae](#); [Liliopsida](#); [Petrosaviiidae](#); [commelinids](#); [Poales](#); [Poaceae](#); [BOP clade](#); [Oryzoideae](#); [Oryzeae](#); [Oryzinae](#); [Oryza](#)

Comments and References:

 [GRIN \(Oct 18, 2016\)](#)
Name accessed on 18 October 2016 in: USDA, ARS, National Genetic Resources Program. Germplasm Resources Information Network - (GRIN) [Online Database]. National Germplasm Resources Laboratory, Beltsville, Maryland.

 [Flora of China - Poaceae](#)
Chen S-L et al. 2006. Poaceae (R. Brown) Barnhart. In Wu, Z. Y., P. H. Raven & D. Y. Hong, eds. Flora of China. Vol. 22 (Poaceae). Science Press, Beijing, and Missouri Botanical Garden Press, St. Louis. Online at Flora of China: www.efloras.org

 [The 3,000 rice genomes project](#)
The 3,000 rice genomes project. GigaScience 2014, 3:7. DOI: <http://dx.doi.org/10.1186/2047-217X-3-7>

Entrez records		
Database name	Subtree links	Direct links
Nucleotide	2,291,284	322,323
Protein	444,228	62,067
Structure	275	76
Genome	1	1
Popset	1,234	1,082
Conserved Domains	12	5
GEO Datasets	22,604	16,467
PubMed Central	34,990	34,990
Gene	95,353	149
HomoloGene	9,787	9,787
SRA Experiments	109,838	26,120
GEO Profiles	670,939	670,939
Protein Clusters	15,559	-
Identical Protein Groups	202,266	44,157
BioProject	6,895	5,349
BioSample	110,196	59,234
Assembly	105	55
PubChem BioAssay	483	449
Taxonomy	9	1

NCBI : Nucleotide



All Databases ▾

Search



Log in

Nucleotide

Nucleotide ▾ txid4530[Organism:exp]

Create alert Advanced

Search

Help

Species

Plants (2,291,254)

Bacteria (112)

Viruses (6)

Customize ...

Molecule types

genomic DNA/RNA (915,442)

mRNA (1,363,554)

rRNA (196)

Customize ...

Source databases

INSDC (GenBank) (2,236,534)

RefSeq (53,619)

Customize ...

Sequence Type

Nucleotide (391,273)

EST (1,255,251)

GSS (644,760)

Genetic compartments

Chloroplast (3,516)

Mitochondrion (208)

Plasmid (109)

Plastid (3,521)

Sequence length

Custom range...

Summary ▾ 20 per page ▾ Sort by Default order ▾

Send to: ▾

Filters: Manage Filters



TAXONOMY

[Oryza sativa](#)

Was this helpful?

Asian cultivated rice (*Oryza sativa*) is a species of monocot in the family Poaceae (grass family).

Taxonomy ID: 4530

[Genomes](#) [Genes](#) [BLAST](#)

Results by taxon

Top Organisms [Tree]

[Oryza sativa](#) (2291274)

synthetic construct (5)

[Zea mays](#) (2)

Cre expression vector pTN75 (1)

Plastid transformation vector pMSK49 (1)

All other taxa (1)

More...

Find related data

Database: Select

Find Items

Search details

txid4530[Organism:exp]

Search

See more...

Items: 1 to 20 of 2291284

<< First < Prev Page 1 of 114565 Next > Last >>

- [Oryza sativa cultivar Jinhui3 PPR830 \(PPR830\), fertility restorer \(Rf19\), hypothetical protein \(ORF2\), hypothetical protein \(ORF3\), and hypothetical protein \(ORF4\).genes, complete cds](#)
1. [37,185 bp linear DNA](#)

Accession: ON855493.1 GI: 2294270732

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

Popular Resources

[PubMed](#)

[Bookshelf](#)

[PubMed Central](#)

[BLAST](#)

[Nucleotide](#)

[Genome](#)

[SNP](#)

[Gene](#)

[Protein](#)

[PubChem](#)

NCBI News & Blog

New ClinVar graphical display

Maps clinically significant variants to gene and position! ClinVar is now a public archive of clinical variant data.

Celebrating 1 Year of NCBI Virtual Outreach Events

We launched the NCBI Virtual Event series in the fall of 2022!



NCBI : Genome resources

NIH National Library of Medicine
National Center for Biotechnology Information

Genome Genome txid4530[Organism:exp] Log in

Create alert Limits Advanced Help

Oryza sativa (Asian cultivated rice)
Reference genome: Oryza sativa Japonica Group (assembly IRGSP-1.0)
Download sequences in FASTA format for genome, transcript, protein
Download genome annotation in GFF, GenBank or tabular format
BLAST against Oryza sativa genome
All 95 genomes for species:
Browse the list
Download sequence and annotation from RefSeq or GenBank
NEW Try the NCBI Datasets Taxonomy page - a new way to access genomic data, including reference genomes

Display Settings: Overview Send to: ID: 10

Organism Overview ; Genome Assembly and Annotation report [95] ; Organelle Annotation Report [8]

Oryza sativa (Asian cultivated rice)

Oryza sativa Organism overview

Lineage: Eukaryota[10183]; Viridiplanteae[1033]; Streptophyta[942]; Embryophyta[935]; Tracheophyta[923]; Spermatophyta[909]; Magnoliopsida[889]; Liliopsida[155]; Poales[96]; Poaceae[88]; BOP clade[47]; Oryzoideae[18]; Oryzeae[18]; Oryzinae[16]; Oryza[15]; Oryza sativa[1]

Rice is one of the most important food crops in the world and feeds more people than any other crop. Rice belongs to the genus Oryza which includes approximately 24 species. They are widely distributed growing in different habitats and different soil types. They show differences in plant growth, yield, pest and disease resistance, stress tolerance [More...](#)

Summary

Sequence data: genome assemblies: 95; sequence reads: 3173 (See Genome Assembly and Annotation report)
Statistics: median total length (Mb): 388.93
median protein count: 38007
median GC%: 43.5525
NCBI Annotation Release: 102

Publications (limited to 20 most recent records)

1. Rationally Designed APOBEC3B Cytosine Base Editors with Improved Specificity. Jin S, et al. Mol Cell 2020 Sep 3
2. Multicentric origin and diversification of *atp6-orf79*-like structures reveal mitochondrial gene flows in *Oryza rufipogon* and *Oryza sativa*. He W, et al. Evol Appl 2020 Oct
3. Large-scale identification and functional analysis of *NLR* genes in blast resistance in the Tetep rice genome sequence. Wang L, et al. Proc Natl Acad Sci U S A 2019 Sep 10

[More...](#)

Representative (genome information for reference and representative genomes)

Reference genome:
Oryza sativa Japonica Group
Submitter: National Institute of Agrobiological Sciences

Loc	Type	Name	RefSeq	INSDC	Size (Mb)	GC%	Protein	rRNA	tRNA	Other RNA	Gene	Pseudogene
Chr	1	NC_029256.1	AP014957.1	43.27	43.8	5,850	-	84	1,237	4,630	158	
Chr	2	NC_029257.1	AP014958.1	35.94	43.3	4,826	2	69	1,311	3,769	117	

NCBI Resources Genome Data Viewer

Tools BLAST Genome

Related information Assembly BioProject Gene Components Protein PubMed Taxonomy

Search details txid4530[Organism:exp] Search See more...

Recent activity Turn Off Clear

Oryza sativa Genome

txid4530[Organism:exp] (1) Genome

embryophyta AND ((refseq[filter] OR swissprot[filter])) (7447863) Protein

embryophyta AND (refseq[filter]) (7408233) Protein

(oryza) AND "Oryza sativa"[orgn] (444207) Protein

See more...

Search

Popular Resources

PubMed Bookshelf PubMed Central BLAST Nucleotide Genome SNP Gene Protein PubChem

NCBI News & Blog

New ClinVar graphical display

Maps clinically significant variants across the genome and position! ClinVar is a public archive of variants reported in association with human diseases.

Celebrating 1 Year of NCBI Virtual Outreach Events

We launched the NCBI Virtual Event series in the fall of 2020!

Université Félix Houphouët-Boigny Institut de Recherche pour le Développement

NCBI : Genome resources

<https://www.ncbi.nlm.nih.gov/datasets/genomes>

Datasets Taxonomy **Genome** Gene Command-line tools Documentation

Genome

BETA

Download a genome data package including genome, transcript and protein sequence, annotation and a data report

Selected taxa

Dioscorea cayenensis subsp. rotundata (Guinea yam) Enter one or more taxonomic names

Filters

Download

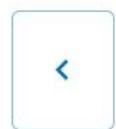
Select columns

5 genomes

Rows per page

20

1-5 of 5



<input type="checkbox"/> Assembly	Scientific name	Modifier	Annotation	Size (Mb)	Level	Year	WGS acc	Action
<input type="checkbox"/> TDr96_F1_v2_PseudoChrom... RefSeq: GCF_009730915.1 GenBank: GCA_009730915.2	Dioscorea cayenensis subsp. rotundata (Guinea yam)	TDr96_F1 cultivar	NCBI RefS...	584.2	Chromosome	2019	BLBR01	
<input type="checkbox"/> TDr96_F1_Pseudo_Chromos... GenBank: GCA_002240015.2	Dioscorea cayenensis subsp. rotundata (Guinea yam)	TDr96_F1 cultivar		456.7	Chromosome	2017	BDMI01	
<input type="checkbox"/> TDr96x99_v1.0.fasta GenBank: GCA_002260605.1	Dioscorea cayenensis subsp. rotundata (Guinea yam)	TDr96/00629 x cultivar		594.2	Scaffold	2017	BBQW01	
<input type="checkbox"/> TDr97_00777_Male_DDN GenBank: GCA_002260645.1	Dioscorea cayenensis subsp. rotundata (Guinea yam)	TDr97_00777 cultivar		683.3	Scaffold	2017	BDMK01	

NCBI Blast

National Library of Medicine
National Center for Biotechnology Information

BLAST®

Important update
The core nucleotide database (*core_nt*) is now the default nucleotide BLAST database. Learn more about *core_nt*.

Basic Local Alignment Search Tool

BLAST finds regions of similarity between biological sequences. The program compares nucleotide or protein sequences to sequence databases and calculates the statistical significance. [Learn more](#)

BLAST+ 2.16.0 is here!
Faster searches and new features.
Mon, 22 Jul 2024 [More BLAST news...](#)

Web BLAST

Nucleotide BLAST
nucleotide ▶ nucleotide

blastx
translated nucleotide ▶ protein

tblastn
protein ▶ translated nucleotide

BLAST Genomic

Enter organism common name, scientific name, or tax id **Search**

Human Mouse Microbes

Choose Search Set

Database Standard databases (nr etc.) rRNA/ITS databases Genomic + transcript databases Betacoronavirus Experimental data

Organism Core nucleotide database (*core_nt*)

Exclude Enter organism name or id - completions will be suggested exclude Add organism

Limit to Enter organism common name, binomial, or tax id. Only 20 top taxa will be shown

Entrez Query Enter an Entrez query to limit search

Program Selection

Optimize for Highly similar sequences (megablast) More dissimilar sequences (discontiguous megablast) Somewhat similar sequences (blastn)

Choose a BLAST algorithm

BLAST Search database *core_nt* using Megablast (Optimize for highly similar sequences) Show results in a new window

+ Algorithm parameters

Log in

How to read this report? [BLAST Help Videos](#) [Back to Traditional Results Page](#)

Edit Search Save Search Search Summary

Job Title U37574.1 Human BRCA1 gene, partial cds

RID DJT22NX6013 Search expires on 09-06 15:56 pm [Download All](#)

Program BLASTN Citation

Database core_nt [See details](#)

Query ID IclQuery_2281045

Description U37574.1 Human BRCA1 gene, partial cds

Molecule type dna

Query Length 3798

Other reports Distance tree of results MSA viewer

Filter Results

Organism only top 20 will appear exclude

Type common name, binomial, taxid or group name [+ Add organism](#)

Percent Identity to **E value** to **Query Coverage** to

Descriptions **Graphic Summary** **Alignments** **Taxonomy**

hover to see the title click to show alignments

100 sequences selected

Distribution of the top 81361 Blast Hits on 100 subject sequences

Query

CIBiG - 9 September - 4 October 2024

Centre National de Recherche sur les Maladies Infectieuses et le Développement des Plantes (CIRAD)
Institut de Recherche pour le Développement (IRD)
Université Félix Houphouët-Boigny (UFHB)
Institut de Recherche pour le Développement pour la Sécurité Alimentaire et la Santé (IRD-SAS)
Institut de Recherche pour le Développement pour l'Environnement et le Climat (IRD-ECO)

Tools: Primer Blast (for design PCR)

An official website of the United States government [Here's how you know](#).

NIH National Library of Medicine
National Center for Biotechnology Information

tibionez@gmail.com

Primer-BLAST

A tool for finding specific primers

Finding primers specific to your PCR template (using Primer3 and BLAST).

Primers for target on one template Primers common for a group of sequences

PCR Template

Enter accession, gi, or FASTA sequence (A refseq record is preferred) Range
Forward primer To
Reverse primer

Or, upload FASTA file aucun fichier sélectionné

Primer Parameters

Use my own forward primer (5'->3' on plus strand)
Use my own reverse primer (5'->3' on minus strand)
PCR product size Min Max
of primers to return
Primer melting temperatures (T_m) Min Opt Max Max T_m difference

Exon/intron selection

A refseq mRNA sequence as PCR template input is required for options in the section

Exon junction span No preference

Exon junction match Min 5' match Min 3' match Max 3' match
Minimal and maximal number of bases that must anneal to exons at the 5' or 3' side of the junction

Intron inclusion Primer pair must be separated by at least one intron on the corresponding genomic DNA

Intron length range Min Max

Primer Pair Specificity Checking Parameters

Specificity check Enable search for primer pairs specific to the intended PCR template

Specialized databases

- Most of sequences are also stored in GenBank/EMBL database
- May contain whole genomes sequences
- May contain specialized resources
- May contain specific tools to exploit the data

Examples of specialized databases

- Plasmodium <https://plasmodb.org/plasmo/app>
- Specialized data collections of Sanger Institute
<https://www.sanger.ac.uk>
- Eukaryotic Pathogens (EuPathD)
<https://veupathdb.org/veupathdb/app>
- Hepatitis database
https://hcv.lanl.gov/content/sequence/HCV/news/old_news.html
- Influenza research database
<https://www.fludb.org/brc/home.spq?decorator=influenza>



Specialized databases: BV-BRC

BACTERIAL AND VIRAL BIOINFORMATICS RESOURCE CENTER

Welcome to the Bacterial and Viral Bioinformatics Resource Center (BV-BRC), an information system designed to support research on bacterial and viral infectious diseases. The BV-BRC combines the data and tools from the Legacy BRC resources: PATRIC, the bacterial BRC, and IRD and ViPR, the viral BRCs. If you are transitioning from PATRIC or IRD/ViPR, please refer to the [Quick Start Guide](#) to learn how to get started with BV-BRC. Please [contact us](#) if you have any questions or issues.

SEARCH All Data Types All terms

BROWSE [BACTERIA](#) [ARCHAEA](#) [VIRUSES](#) [EUKARYOTIC HOSTS](#)

ANALYZE DATA IN BV-BRC

Upload and analyze your data in the private workspace. [Register](#) or [Login](#) to get started.

SEARCH	ANALYZE	MANAGE DATA	BATCH ACCESS
Taxa	Assembly	Access Private Workspace	Command-Line Interface
Genomes	Annotation	Save Search Results	Data API
Proteins	BLAST	Upload Data	FTP
Specialty Genes	MSA	Access Analysis Jobs	
All Searches...	All Tools & Services...	Share and Publish	

[QUICK START](#) [REFERENCE GUIDES](#) [TUTORIALS](#)

Information for PATRIC Users

The new BV-BRC system is built on the PATRIC system. Viral data and analysis tools and services from IRD/ViPR have now been integrated into the resource.

[BV-BRC Data and Functionality Overview](#)

Information for IRD/ViPR Users

The new BV-BRC system is built on the PATRIC Bacterial BRC system and incorporates data and tools from IRD and ViPR. The following materials are designed to aid IRD and ViPR users in using BV-BRC.

[IRD/ViPR to BV-BRC Mapping](#)

OUTBREAKS

Influenza H5N1 2024
SARS-CoV-2

NEWS & ANNOUNCEMENTS

BV-BRC Renewal

The BV-BRC will continue to be supported by the National Institute of Allergy and Infectious Diseases of the National Institutes of Health under Award Number U24AI183849. We will continue to support all of the current datasets and tools along with existing (and new) user workspaces. Additional enhancements to our data and tools will also be coming in future updates. ... [read more](#)

COMMUNITY NEWS

BV-BRC 286 abonnés 12 h.

Updated map now includes cattle HPAI detection in California

2024 H5N1 Cases in Cattle and Humans

Specialized databases: Nextstrain

The Nextstrain platform provides real-time tracking of pathogen evolution. It features a central dashboard with links to various analyses and tools, and a detailed phylogenetic tree for the SARS-CoV-2 outbreak in North America.

Key Features:

- About us:** An open-source project to harness the scientific and public health potential of pathogen genome data.
- Core pathogens:** Continually updated views of a range of pathogens maintained by the Nextstrain team.
- Nextclade:** In-browser phylogenetic placement, clade assignment, mutation calling and sequence quality checks.
- Featured analyses:** Mpx in the DRC, Mpox, SARS-CoV-2, H5N1 cattle outbreak, Yersinia pestis, Measles, RSV, and Ebola in the DRC.
- SARS-CoV-2:** Up-to-date analyses and a range of resources for SARS-CoV-2, the full genome analysis of the ongoing influenza A/H5N1 cattle outbreak in North America.
- Transmissions:** A map showing the geographic spread of the A/H5N1 cattle outbreak across North America, with pie charts indicating host distribution (Avian, Nonhuman Mammal, Human, Cattle) at specific locations.
- Nucleotide diversity of genome:** A bar chart showing the diversity of genomes over time.

Specialized databases: Phytozome

Welcome to

Phytozome ▾

Overview

Release Notes

News

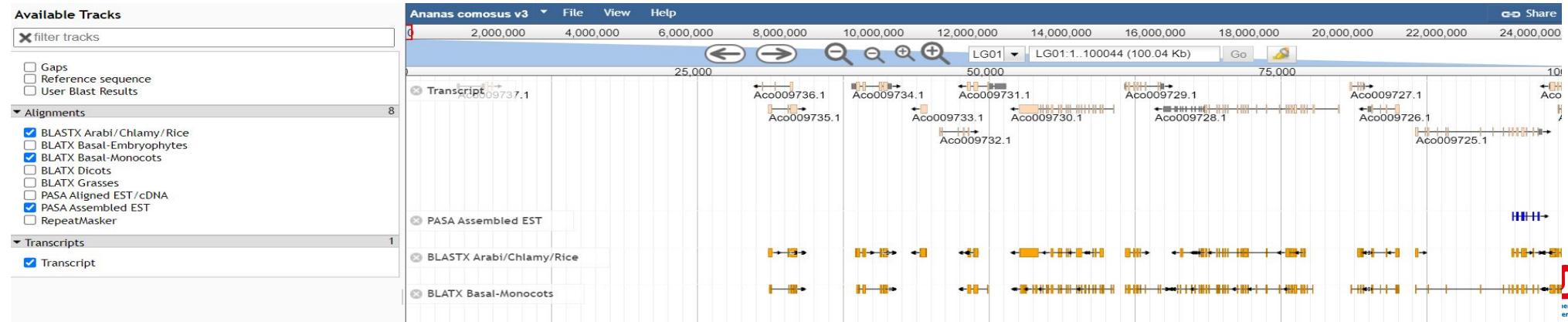
Recent Genome Releases

Genome	Common name	Release Date
<i>Gossypium hirsutum</i> v3.1	upland cotton	Aug 16, 2022
<i>Lens culinaris</i> v1	lentil	Aug 16, 2022
<i>Lens eruvirens</i> v1	wild lentil	Aug 16, 2022
<i>Glycine max</i> Wm82 ISU-01 v2.1	soybean	Aug 16, 2022
<i>Chlamydomonas reinhardtii</i> CC-4532 v6.1	green algae	Jun 17, 2022
<i>Kalanchoe laxiflora</i> FTBG2000359A v3.1		Mar 1, 2022
<i>Gossypium hirsutum</i> CSX8308 v1.1	upland cotton	Mar 1, 2022

Phytozome, the Plant Comparative Genomics portal of the Department of Energy's Joint Genome Institute, provides JGI users and the broader plant science community a hub for accessing, visualizing and analyzing JGI-sequenced plant

1. Choose genomes by selecting from tree or type genus/species/common name 0 genomes selected ▾

2. find genes by keyword search by BLAST get standard data files build custom data sets



Specialized databases: TAIR (Arabidopsis)

<https://www.arabidopsis.org/>

The screenshot shows the homepage of the TAIR website. At the top, there is a navigation bar with links to Home, Help, Contact, About Us, Subscribe, Login, Register, and an institution section for IRD Montpellier (subscribed). Below the navigation bar is a search bar with a placeholder "Enter search text" and a dropdown menu set to "Gene". The main content area features two large boxes: "Bulk gene search and download" and "Bulk protein search and download", both highlighted with red boxes. To the right of these boxes is a "Breaking News" sidebar with several news items and social media links.

The Arabidopsis Information Resource

Bulk downloads have moved!

Bulk gene search and download

Bulk protein search and download

About TAIR

The Arabidopsis Information Resource (TAIR) maintains a database of genetic and molecular biology data for the model higher plant *Arabidopsis thaliana*. Data available from TAIR includes the complete genome sequence along with gene structure, gene product information, gene expression, DNA and seed stocks, genome maps, genetic and physical markers, publications, and information about the Arabidopsis research community. Gene product function data is updated every week from the latest published research literature and community data submissions. TAIR also provides extensive linkouts from our data pages to other Arabidopsis resources.

TAIR is located at [Phoenix Bioinformatics](#) and funded by [subscriptions](#).

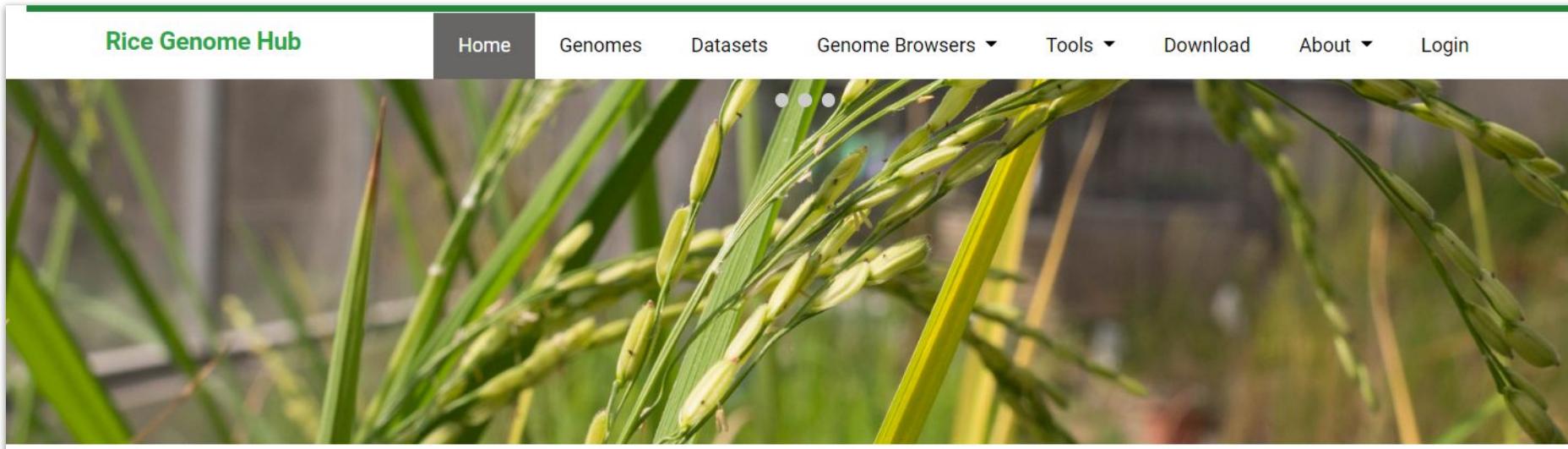
Full access to TAIR requires a subscription. Please see our [subscription page](#) for further details.

General comments or questions: curator@arabidopsis.org

Breaking News

- News RSS has been updated [Jul 31, 2024]
- Our news RSS feed is active again. Subscribe using: <https://www.arabidopsis.org/api/news/breakingnews/rss>
- Bulk downloads have moved!** [Jul 24, 2024]
- Wondering where the bulk downloads went? They moved to the Gene and Protein search pages. Upload your list of AGI locus IDs and download from the search results. See our [Help | YouTube](#).
- Arabidopsis thaliana UniProt data updated on TAIR [Jul 10, 2024]

Specialized databases: Rice Genome Hub



Available Tools

The Rice Genome Hub provides a serie of tools to browse, visualize and search among all data sets available.



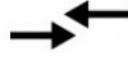
DIANE

Tool for RNA-seq data analyses, from raw count to gene regulatory network. Allow the user to...



Gene Search

Search for a gene by name, location, functional annotation keywords...



Primer Designer

Primer Designer allows users to design new target-specific primers in one step as well as to...



Primer Blaster

Check PCR primer specificity on any Rice Genome



SRA (Sequence Reads Archive)(NCBI) ENA (European Nucleotide Archive) (EBI)

An official website of the United States government [Here's how you know](#)

National Library of Medicine
National Center for Biotechnology Information

SRA SRA Advanced Search Log in Help

SRA - Now available on the cloud

Sequence Read Archive (SRA) data, available through multiple cloud providers and NCBI servers, is the largest publicly available repository of high throughput sequencing data. The archive accepts data from all branches of life as well as metagenomic and environmental surveys. SRA stores raw sequencing data and alignment information to enhance reproducibility and facilitate new discoveries through data analysis.

National Library of Medicine
National Center for Biotechnology Information

SRA SRA cassava Advanced Search Log in

Access Public (11,292)
Source DNA (7,422)
RNA (3,759)
Type genome (1,174)
Library Layout paired (0,074)
single (5,218)
Platform ABI SOLID (12)
BGISEQ (174)
Capillary (1)
Illumina (10,705)
Ion Torrent (7)
LS454 (37)
Oxford Nanopore (94)
PacBio SMRT (251)
Strategy Epigenomics (71)
Exome (258)

Summary 20 per page View results as an expanded interactive table using the RunSelector. Send results to Run selector
Send to: Filters: Manage Filters

Results by taxon
Top Organisms [Tree]
Manihot esculenta (322)
Brassica napus (392)
soil metagenome (342)
Bemisia tabaci (232)
plant metagenome (220)
All other taxa (1784)
More...

Search results
Items: 1 to 20 of 11292 << First < Prev Page 1 of 585 Next > Last >>
Illumina NovaSeq 6000 paired end sequencing
1. 1 ILLUMINA (Illumina NovaSeq 6000) run: 28.9M spots, 8.7G bases, 2.6Gb downloads
Accession: ERX12972232
Illumina NovaSeq 6000 paired end sequencing
2. 1 ILLUMINA (Illumina NovaSeq 6000) run: 34M spots, 10.3G bases, 3.1Gb downloads
Accession: ERX12972231
Illumina NovaSeq 6000 paired end sequencing
3. 1 ILLUMINA (Illumina NovaSeq 6000) run: 70.4M spots, 21.3G bases, 6.4Gb downloads
Accession: ERX12972230
Illumina NovaSeq 6000 paired end sequencing
4. 1 ILLUMINA (Illumina NovaSeq 6000) run: 32.6M spots, 9.8G bases, 3Gb downloads
Accession: ERX12958352

Search in related databases

Database	Access
	public controlled
BioSample	8,645
BioProject	275
dbGaP	1
GEO Datasets	388

Find related data

ENA European Nucleotide Archive

Home Submit Search Rulespace About Support

We recommend that you subscribe to the ENA-announce mailing list for updates on services.

For SARS-CoV-2 data submissions, users should contact us in advance of submission at virus-dataflow@ebi.ac.uk for specific advice on options and to access the highest levels of support. We have also launched a Drag-and-Drop Data Submission Service (currently in Beta) suitable for certain SARS-CoV-2 submissions. We are inviting submitters to try this out. Please contact us at the email above for details.

European Nucleotide Archive

The European Nucleotide Archive (ENA) provides a comprehensive record of the world's nucleotide sequencing information, covering raw sequencing data, sequence assembly information and functional annotation. [More about ENA](#).

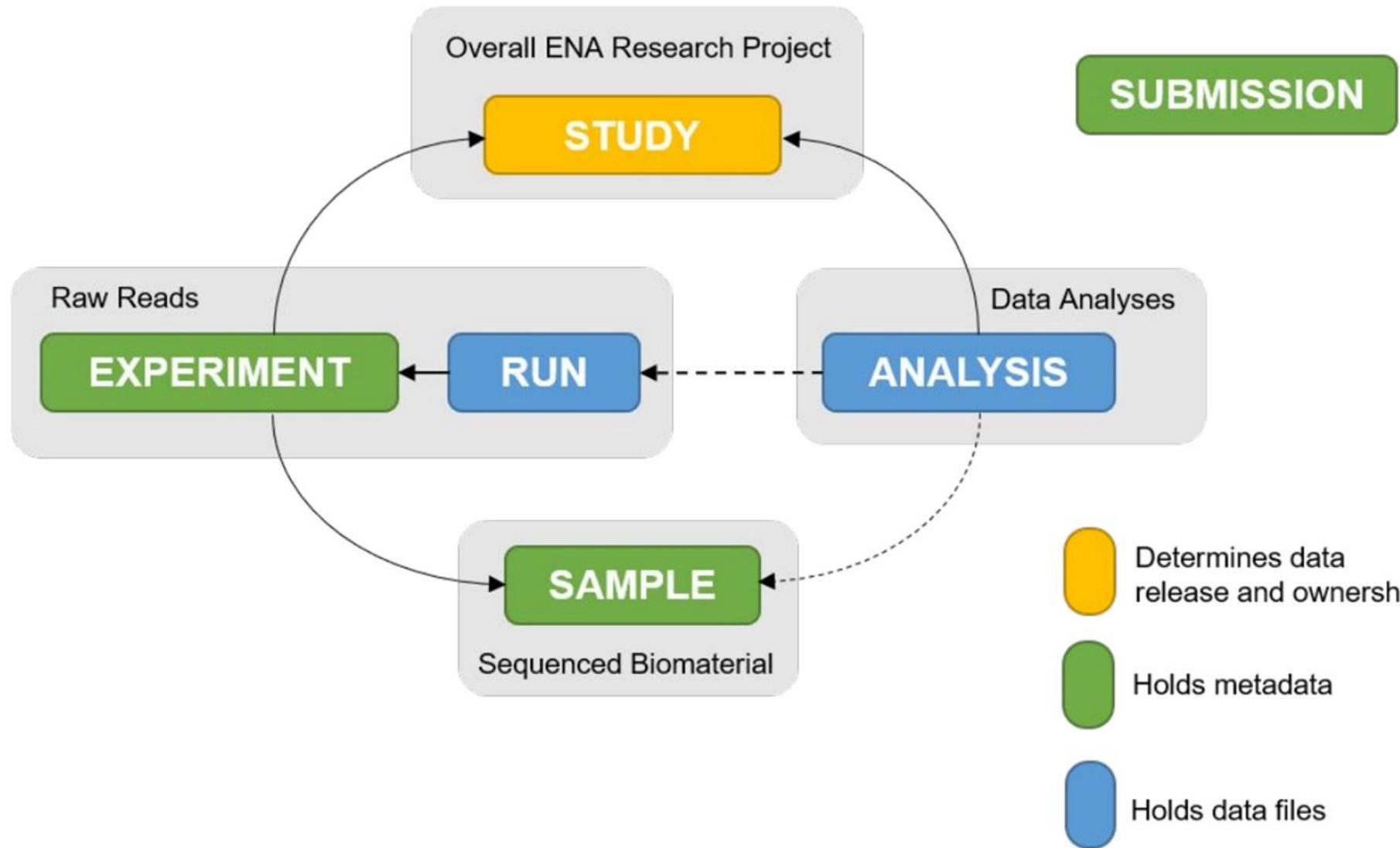
Access to ENA data is provided through the browser, through search tools, through large scale file download and through the API.

Submit **Search** **Rulespace** **Support**

Tweets from @ENASequence

European Nucleotide Archive (ENA) Retweeted

The ENA Metadata Model



- Maintain most comprehensive range of freely accessible and up-to-date molecular databases worldwide
- Offer online and live training for the use of their resources.

<https://www.ebi.ac.uk/training>

EMBL - EBI

The EMBL-EBI website has been redesigned. Please send us feedback about this page.

EMBL's European Bioinformatics Institute

EMBL-EBI

Unleashing the potential of big data in biology

Find a gene, protein or chemical

All

Search

Example searches: [blast keratin bfl1](#) | [About EBI Search](#)

[Find data resources](#) →

[Submit data](#) →

[Explore our research](#) →

[Train with us](#) →

[Latest news](#) →



Organisations should embrace open science faster – interview with Prof. Dame Janet Thornton

17 May 2022



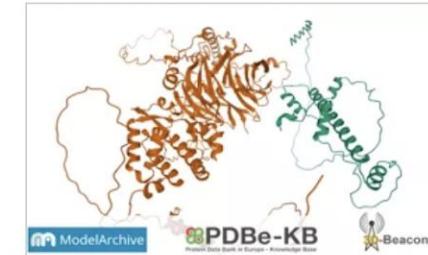
Europe PMC: Harnessing the power of text mining to accelerate life sciences research

12 May 2022



2.4 billion sequences now available in the latest MGnify protein database release

11 May 2022



[Predicted complexes from ModelArchive now on PDBe-KB pages](#)

6 May 2022

Services

[Overview](#) | [A to Z](#) | [Data submission](#) | [Research infrastructure development programme](#) | [Support](#)

The European Bioinformatics Institute (EMBL-EBI) maintains the world's most comprehensive range of freely available and up-to-date molecular data resources.

Developed in collaboration with our colleagues worldwide, our services let you share data, perform complex queries and analyse the results in different ways. You can work locally by downloading our data and software, or use our web services to access our resources programmatically.

— You can read more about our services in the journal [Nucleic Acids Research](#).

Tools & Data Resources

Tools

Clustal Omega



Multiple sequence alignment of DNA or protein sequences. Clustal Omega replaces the older ClustalW alignment tools.

[Web API](#) | [Multiple sequence alignment](#)

InterProScan



InterProScan searches sequences against InterPro's predictive protein signatures.

[Web API](#) | [Protein feature detection](#)

[Sequence motif recognition](#)

BLAST [protein]



Fast local similarity search tool for protein sequence databases.

[Web API](#) | [Sequence similarity search](#)

BLAST [nucleotide]



Fast local similarity search tool for nucleotide sequence databases.

[Web API](#) | [Sequence similarity search](#)

Data resources

Ensembl



Genome browser, API and database, providing access to reference genome annotation

[Web API](#) | [EMBL-EBI Terms of use](#)

UniProt



A comprehensive resource for protein sequence and functional annotation.

[Web API](#) | [CC-BY](#)

PDBe



The European resource for the collection, organisation and dissemination of 3D structural data (from PDB and EMDB) on biological macromolecules and their complexes.

[Web API](#) | [CC0](#)

Europe PMC



A database to search the worldwide life sciences literature

[Web API](#) | [EMBL-EBI Terms of use](#)

Browse by type

XXX	Gene Expression	Proteins
Structures	Systems	Chemical biology
Ontologies	Literature	Cross domain

Programmatic access

EMBL-EBI web services allow you to query our large biological data resources programmatically, so that you can develop data analysis pipelines or integrate public data with your own applications. The Web Services technology we use are built on open standards to ensure client and server software from various sources will work well together.

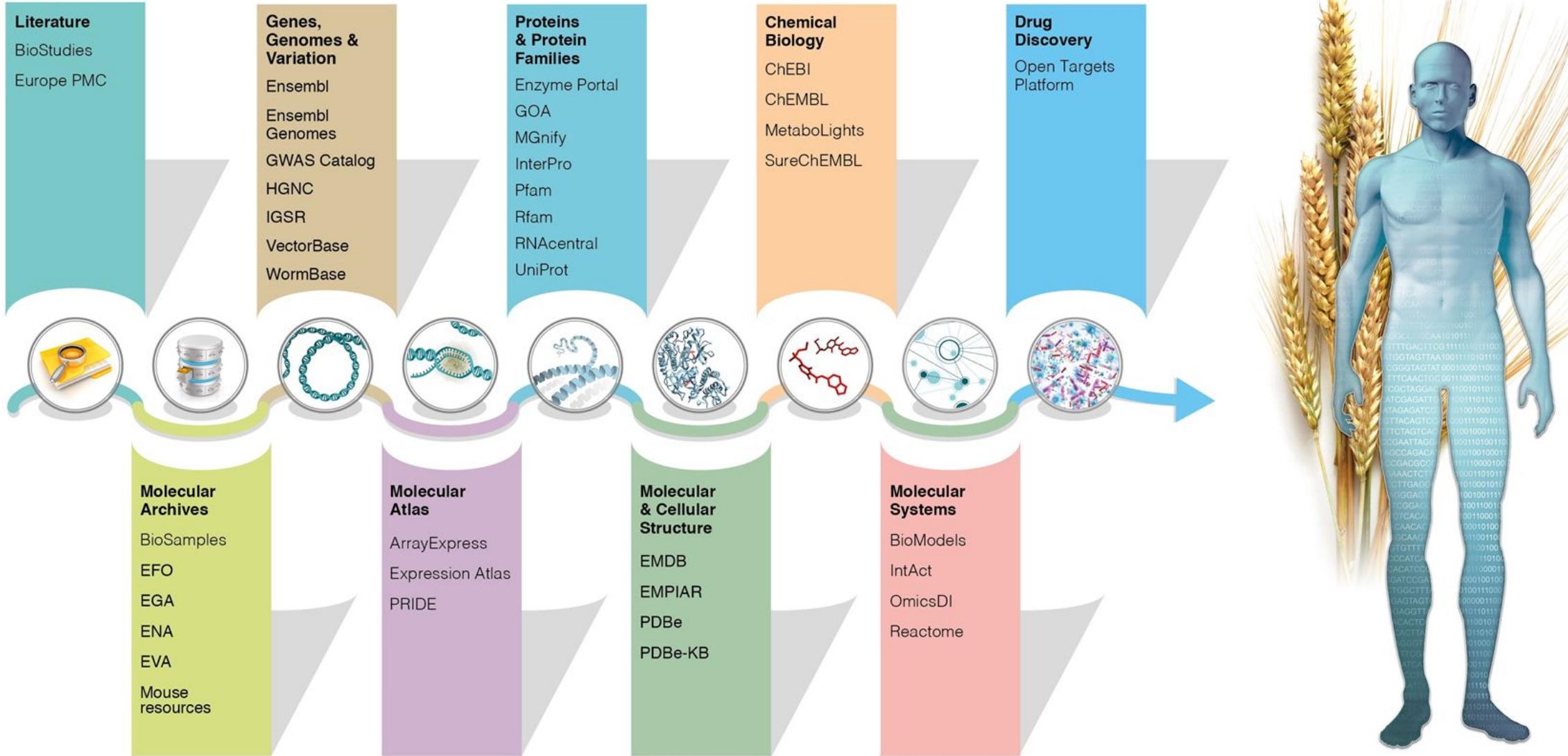
[Browse EMBL-EBI web services](#)

Principles of service provision

Open

Our data and tools are freely available, without restriction. The only exception is potentially identifiable human genetic information, for which access depends

EMBL - EBI services



EMBL - EBI databases

The screenshot shows the UniProt BLAST search interface. At the top, there's a navigation bar with links for BLAST, Align, Retrieve/ID mapping, Peptide search, Help, and Contact. A message at the top states: "From June 20, 2018 all traffic will be automatically redirected to HTTPS. More information or view this page using https". Below the navigation is a color scale from red (100) to green (0) labeled "Identity %". The main area has a large "UniProt" logo with a circular graphic. On the left, a sidebar titled "Filter by" lists categories like Reviewed (4), Unreviewed (246), 3D structure (2), Proteomes (156), and Popular organisms (Human, Mouse, Zebrafish, Rat, Bovine). The "Align" section is currently selected. The main content area displays a search result for entry P00439, comparing PH4H_HUMAN against itself. It includes an alignment overview, a sequence alignment table, and detailed statistics like E-value, Score, and Identity.

P00439

Examples: IPPO20405, Uniref, P51987, PF02932, GO:0007185

Home Search Release notes Download About InterPro Help Contact

Export TSV Export FASTA

Phenylalanine-4-hydroxylase (P00439)

Accession: P00439 (PH4H_HUMAN)
Species: Homo sapiens (Human)
Length: 452 amino acids (complete)

Entry type:

- Homolog
- Family
- Domains
- Repeats
- Site

Status:

- Unintegrated

Per-residue features:

- Residue annotation

Domains and repeats

Detailed signature matches



The screenshot shows the Ensembl Genomes homepage. The top navigation bar includes links for About us, Download, Data types, Data access, FAQs, Bacteria, Protists, Fungi, Plants, Metazoa, and Vertebrates. A sidebar on the left lists categories like All species, Bacteria, Eukaryotes, Metazoa, Plants, and Protists, along with links for genomic variation data, peptide comparative data, pan-taxonomic coverage, whole genome alignments, and sequence alignment. The main content area features the Ensembl Genomes logo and a search interface. The search interface includes fields for Species (with dropdowns for Division, Pan Comparisons, Variation, Peptide Alignments, Genome Alignments, and Other Alignments), a search term input (Match part of a name: elegans Wildcard: * may be used; esch.*.coll), and an Apply button. Below the search is a table with columns for Species, Division, Taxonomy, Assembly, Genebuild, Variation, Pan, Genome, Peptides, and Other Alignments. The table contains several rows of data, each with a red checkmark icon.

Protein databanks

- **TrEMBL**: an automatic translation of DNA sequences from the EMBL ENA databank containing completely automated annotations based on homology
- **SwissProt** : a hand-curated resource where cross-references are made to the literature checking information from the automated annotations and adding more information manually.
- **PIR** (Protein Information Resource)
- **Uniprot = SwissProt + TrEMBL + PIR**



=> Relies on ExPASy server from SIB (Swiss Institute of Bioinformatics) and EBI



UniProt BLAST Align Peptide search ID mapping SPARQL UniProtKB ▾ atpase human Advanced | List Search

Status

- Reviewed (Swiss-Prot) (2,510)
- Unreviewed (TrEMBL) (232,571)

Popular organisms

- Human (5,469)
- Zebrafish (1,367)
- Mouse (149)
- S. cerevisiae (46)
- Rat (44)

Taxonomy

Group by

- Taxonomy
- Keywords
- Gene Ontology
- Enzyme Class

Proteins with

- 3D structure (846)
- Active site (6,790)
- Activity regulation (973)
- Allergen (13)
- Alternative products (isoforms) (671)
- More items

UniProtKB 235,081 results

Tools ▾ Download (235k) Add View: Cards Table Customize columns Share ▾

Entry	Entry Name	Protein Names	Gene Names	Organism
P46063	RECQL_HUMAN	ATP-dependent DNA helicase Q1 [...]	RECQL, RECQ1, RECQL1	Homo sapiens (Human)
Q6NVH7	SWAP1_HUMAN	ATPase SWSAP1 [...]		(Human)
Q9Y5K8	VATD_HUMAN	V-type proton ATPase subunit D [...]	ATP6V1D, ATP6M, VATD	Homo sapiens (Human)
Q14802	FXYD3_HUMAN	FXYD domain-containing ion transport regulator 3 [...]	FXYD3, MAT8, PLML	Homo sapiens (Human)
O14657	TOR1B_HUMAN	Torsin-1B [...]	TOR1B, DQ1, FKSG18	Homo sapiens (Human)
Q00839	HNRPU_HUMAN	Heterogeneous nuclear ribonucleoprotein U [...]	HNRNPU, C1orf199, HNRPU, SAFA, U21.1	Homo sapiens (Human)
P20648	ATP4A_HUMAN	Potassium-transporting ATPase alpha chain 1 [...]	ATP4A	Homo sapiens (Human)
O43681	GET3_HUMAN	ATPase GET3 [...]	GET3, ARSA, ASNA1, TRC40	Homo sapiens (Human)
P55072	TERA_HUMAN	Transitional endoplasmic reticulum ATPase [...]	VCP, HEL-220, HEL-S-70	Homo sapiens (Human) 806 AA
P18074	ERCC2_HUMAN	General transcription and DNA repair factor IIH helicase subunit XPD [...]	ERCC2, XPD, XPDC	Homo sapiens (Human) 760 AA
P19447	ERCC3_HUMAN	General transcription and DNA repair factor IIH helicase/translocase subunit XPB [...]	ERCC3, XPB, XPBC	Homo sapiens (Human) 782 AA
P38606	VATA_HUMAN	V-type proton ATPase catalytic subunit A [...]	ATP6V1A, ATP6A1, ATP6V1A1, VPP2	Homo sapiens (Human) 617 AA

Q6NVH7 · SWAP1_HUMAN

Protein¹ | ATPase SWSAP1
 Gene¹ | SWAP1
 Status¹ | UniProtKB reviewed (Swiss-Prot)
 Organism¹ | Homo sapiens (Human)

Amino acids | 229 (go to sequence)
 Protein existence¹ | Evidence at protein level
 Annotation score¹ | 65

Entry Variant viewer Feature viewer Genomic coordinates Publications External links History

Tools ▾ Download Add Add a publication Entry feedback

Function¹
 ATPase which is preferentially stimulated by single-stranded DNA and is involved in homologous recombination repair (HRR). Has a DNA-binding activity which is independent of its ATPase activity.

GO annotations¹
 Access the complete set of GO annotations on QuickGO

Slimming set: generic

Cell color indicative of number of GO terms

ASPECT TERM

- Cellular Component: nucleus [Source:UniprotKB-SubCell] 1 Publication
- Cellular Component: Shu complex [Source:UniprotKB] 1 Publication
- Molecular Function: ATP hydrolysis activity [Source:UniprotKB] 1 Publication
- Molecular Function: single-stranded DNA binding [Source:UniprotKB] 1 Publication
- Biological Process: double-strand break repair via homologous recombination [Source:UniprotKB] 1 Publication
- Biological Process: protein stabilization [Source:UniprotKB] 1 Publication

Protein domains and families

proSite

Home ScanProsite Browse ProRule Documentation Downloads About

Search PROSITE Search

Database of protein domains, families and functional sites

SARS-CoV-2 relevant PROSITE motifs

PROSITE consists of documentation entries describing protein domains, families and functional sites as well as associated patterns and profiles to identify them [More... / References / Commercial users].

PROSITE is complemented by ProRule, a collection of rules based on profiles and patterns, which increases the discriminatory power of profiles and patterns by providing additional information about functionally and/or structurally critical amino acids [More..].

Release 2024_04 of 24-Jul-2024 contains 1943 documentation entries, 1311 patterns, 1389 profiles and 1406 ProRule.

Search PROSITE e.g. PDOC00022, PS50089, SH3, zinc finger Search add wildcard **

Quick Scan mode of ScanProsite Quickly find matches of your protein sequences to PROSITE signatures (max. 10 sequences). [?] Examples

For UniProtKB/TrEMBL accessions/identifiers, only those of entries belonging to reference proteomes are accepted.
Scan Clear Exclude motifs with a high probability of occurrence from the scan

For more scanning options go to ScanProsite

Browse PROSITE by documentation entry by ProRule description by taxonomic scope by number of positive hits

Other tools PRATT allows to interactively generate conserved patterns from a series of unaligned proteins. MyDomains - Image Creator allows to generate custom domain figures.



Pfam

HOME | SEARCH | BROWSE | FTP | HELP | ABOUT EMBL-EBI

Pfam data and new releases are available through [InterPro](#)
The Pfam website now serves as a static page with no data updates. All links below redirect to the closest alternative page in the InterPro website.

Pfam 37.0 (21 979 entries, 709 clans)

The Pfam database is a large collection of protein families, each represented by *multiple sequence alignments* and *hidden Markov models (HMMs)*. [More...](#)

QUICK LINKS [VIEW PFAM ANNOTATION AND ALIGNMENTS](#)
[SEQUENCE SEARCH](#)
[VIEW A PFAM ENTRY](#) Enter an accession (e.g. PF02171) to see all data for that entry.
[VIEW A CLAN](#)
[VIEW A SEQUENCE](#)
[VIEW A STRUCTURE](#)
[KEYWORD SEARCH](#)
[JUMP TO](#)

InterPro

Classification of protein families

Home Search Browse Results Release notes Download Help About Contact us

InterPro 101.0 25 July 2024

Classification of protein families

InterPro provides functional analysis of proteins by classifying them into families and predicting domains and important sites. To classify proteins in this way, InterPro uses predictive models, known as signatures, provided by several different databases (referred to as member databases) that make up the InterPro consortium. We combine protein signatures from these member databases into a single searchable resource, capitalising on their individual strengths to produce a powerful integrated database and diagnostic tool.

► Citing InterPro
If you find InterPro useful, please cite the reference that describes this work:
Paysan-Lafosse T, Blum M, Chuguransky S, Gregg T, Pinto BL, Salazar GA, Bileschi ML, Bork P, Bridge A, Colwell L, Gough J, Haft DH, Letunic I, Marchler-Bauer A, Mi H, Natale DA, Orengo CA, Pandurangan AP, Rivoire C, Sigrist CJ, Silitoe I, Thanki N, Thomas PD, Tosatto SCE, Wu CH, Bateman A. *InterPro* in 2022. *Nucleic Acids Research*, Nov 2022, (doi: 10.1093/nar/gkac993)

Search by sequence Search by text Search by Domain Architecture

Scan your sequences
Enter your sequence
Choose file Example protein sequence Advanced options
Search Clear

Powered by InterProScan

Gene Ontology (GO)

 GENEONTOLOGY
Unifying Biology

About Ontology Annotations Downloads Help

Current release 2024-06-17: 42 093 GO terms | 7 624 863 annotations
1 537 645 gene products | 5 409 species (see statistics)

THE GENE ONTOLOGY RESOURCE

The mission of the GO Consortium is to develop a comprehensive, **computational model of biological systems**, ranging from the molecular to the organism level, across the multiplicity of species in the tree of life.

The Gene Ontology (GO) knowledgebase is the world's largest source of information on the functions of genes. This knowledge is both human-readable and machine-readable, and is a foundation for computational analysis of large-scale molecular biology and genetics experiments in biomedical research.

Search GO term or Gene Product in AmiGO ... 

Any Ontology Gene Product

GO Enrichment Analysis ?

Powered by PANTHER

Your gene IDs here...

biological process

Homo sapiens Examples Launch 

Hint: can use UniProt ID/AC, Gene Name, Gene Symbols, MOD IDs

 AmiGO 2 Home Search Browse Tools & Resources Help Feedback About

Quick search 

SRP-dependent cotranslational protein targeting to membrane, signal sequence recognition

Term Information ?

Accession: GO:0006617
Name: SRP-dependent cotranslational protein targeting to membrane, signal sequence recognition
Ontology: biological_process
Synonyms: SRP-dependent cotranslational membrane targeting, signal sequence recognition, SRP-dependent cot... [more](#)

Alternate IDs: None
Definition: The process in which SRP binds to the signal peptide in a nascent protein, causing protein elongation to pause, during cotranslational membrane targeting. Source: ISBN:0815316194
Comment: None
History: See term history for GO:0006617 at QuickGO
Chem. react: None
Subset: None
Related: [Link](#) to all genes and gene products annotated to SRP-dependent cotranslational protein targeting to membrane, signal sequence recognition (excluding "regulates").
[Link](#) to all direct and indirect annotations to SRP-dependent cotranslational protein targeting to membrane, signal sequence recognition (excluding "regulates").
[Link](#) to all direct and indirect annotations download (limited to first 10,000) for SRP-dependent cotranslational protein targeting to membrane, signal sequence recognition (excluding "regulates").

Include "regulates"

For more information, please see the ontology relation documentation.

Annotations Graph Views Inferred Tree View Neighborhood Mappings

- GO:0009987 cellular process
- GO:0051179 localization
- GO:0008150 biological_process
- GO:0051641 cellular localization
- GO:0033036 macromolecule localization
- GO:0070727 cellular macromolecule localization
- GO:0051668 localization within membrane
- GO:0008104 protein localization
- GO:0051234 establishment of localization
- GO:0045184 establishment of protein localization
- GO:0072657 protein localization to membrane
- GO:0033365 protein localization to organelle
- GO:0071840 cellular component organization or biogenesis
- GO:0090150 establishment of protein localization to membrane
- GO:0072594 establishment of protein localization to organelle
- GO:0070972 protein localization to endoplasmic reticulum
- GO:0006605 protein targeting
- GO:0044085 cellular component biogenesis
- GO:0016043 cellular component organization
- GO:0072599 establishment of protein localization to endoplasmic reticulum
- GO:0006512 protein targeting to membrane
- GO:0022607 cellular component assembly
- GO:0006613 cotranslational protein targeting to membrane
- GO:0045047 protein targeting to ER
- GO:0043933 protein-containing complex organization
- GO:0006503 protein-containing complex assembly
- GO:0006614 SRP-dependent cotranslational protein targeting to membrane
- ▼ GO:0006617 SRP-dependent cotranslational protein targeting to membrane, signal sequence recognition

Metabolic pathways: KEGG

KEGG Databases Tools Auto annotation Kanehisa Lab

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

KEGG2 PATHWAY BRITE MODULE KO GENES COMPOUND NETWORK DISEASE DRUG

Select prefix map Enter keywords Go Help

[New pathway maps | Update history]

Pathway Maps

KEGG PATHWAY is a collection of manually drawn pathway maps representing our knowledge of the molecular interaction, reaction and relation networks for:

1. Metabolism
 - Global/overview Carbohydrate Energy Lipid Nucleotide Amino acid Other amino Glycan Cofactor/vitamin Terpenoid/PK Other secondary metabolite Xenobiotics Chemical structure
2. Genetic Information Processing
3. Environmental Information Processing
4. Cellular Processes
5. Organismal Systems
6. Human Diseases
7. Drug Development

The pathway map viewer linked from this page contains features of KEGG mapping, especially for coloring map objects as described here.

Pathway Identifiers

Each pathway map is identified by the combination of 2-4 letter prefix code and 5 digit number (see KEGG Identifier). The prefix has the following meaning:

- map manually drawn reference pathway
- ko reference pathway highlighting KOs
- ec reference metabolic pathway highlighting EC numbers
- rn reference metabolic pathway highlighting reactions
- <org> organism-specific pathway generated by converting KOs to gene identifiers

and the numbers starting with the following:

- 011 global map (lines linked to KOs)
- 012 overview map (lines linked to KOs)
- 010 chemical structure map (no KO expansion)
- 07 drug structure map (no KO expansion)
- other regular map (boxes linked to KOs)

are used for different types of maps.

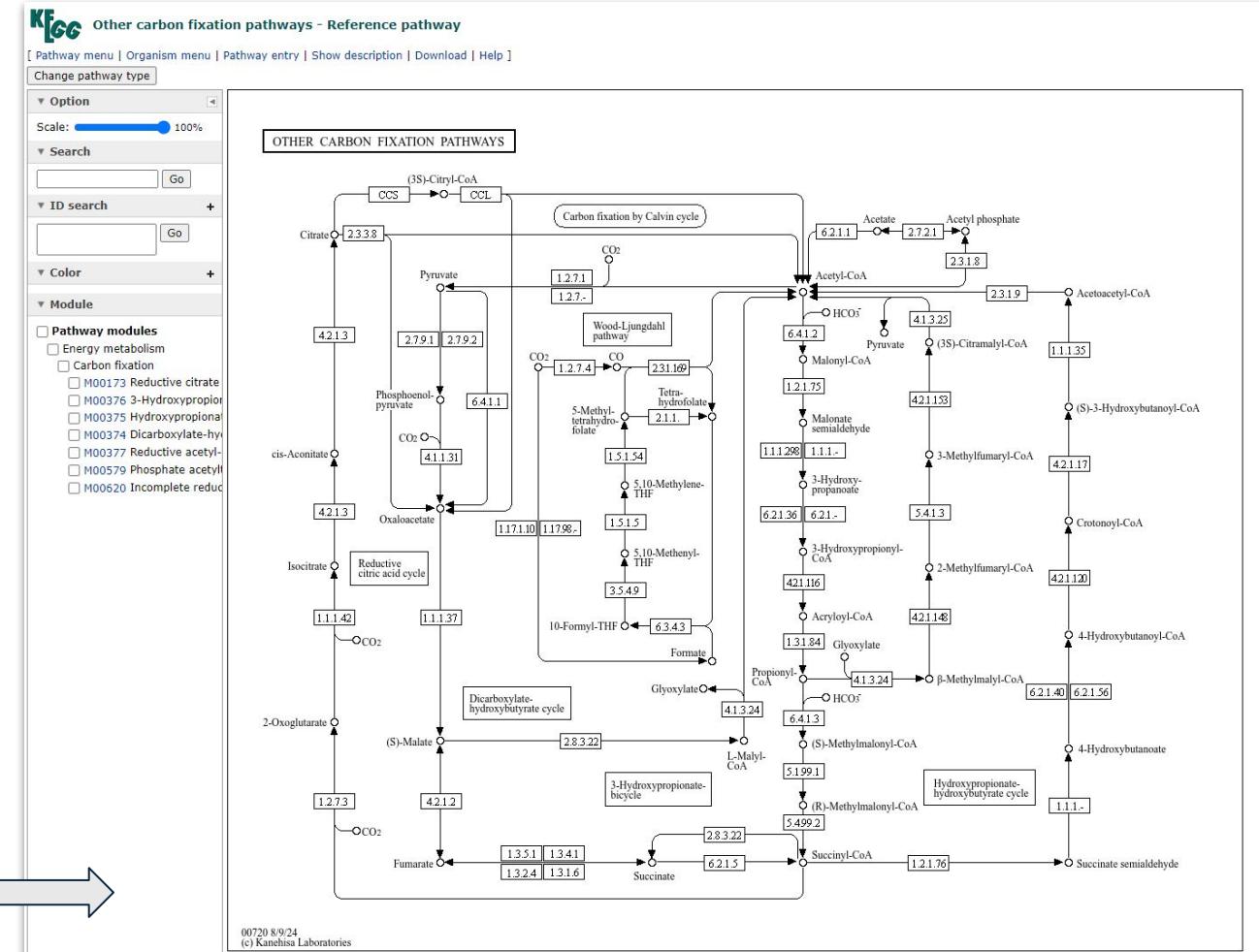
KEGG PATHWAY is integrated with MODULE and NETWORK databases as indicated below.

M - module
R - reaction module
N - network

1. Metabolism

1.0 Global and overview maps

01100 M Metabolic pathways
01110 M Biosynthesis of secondary metabolites
01120 M Microbial metabolism in diverse environments
01200 M R Carbon metabolism
01210 M R 2-Oxocarboxylic acid metabolism
01212 M R Fatty acid metabolism
01230 M R Biosynthesis of amino acids
01232 M R Nucleotide metabolism
01250 M R Biosynthesis of nucleotide sugars
01240 M R Biosynthesis of cofactors
01220 M R Degradation of aromatic compounds
01310 M Nitrogen cycle <i>New!</i>



Protein Structure database: PDB

- Protein Data Bank (PDB)
- Crystallographic database for the three-dimensional structural data of large biological molecules, such as proteins and nucleic acids.
- Data obtained by X-ray crystallography, NMR spectroscopy

Protein Structure database: PDB

<https://www.rcsb.org/>

RCSB PDB Deposit Search Visualize Analyze Download Learn About Documentation Careers COVID-19 MyPDB Contact us

PDB PROTEIN DATA BANK 224,572 Structures from the PDB 1,068,577 Computed Structure Models (CSM)

3D Structures oryza sweet Include CSM Advanced Search | Browse Annotations Help

PDB-101 PDB EMDDataResource NAKB wwPDB Foundation PDB-Dev

Welcome Deposit Search Visualize Analyze Download Learn

Access Computed Structure Models (CSMs) of available model organisms Learn more

September Molecule of the Month

Carbon Capture Mechanisms

Explore NEW Features PDB-101 Training Resources

Latest Entries As of Tue Sep 03 2024

9FHD hHKK-C in complex with BI-9787

Features & Highlights

- Register for the Sept 18 Virtual Office Hour on Molecular Animations
- Join us to learn more about RCSB PDB's molecular animations
- Register for the Sept 12 Virtual Office Hour: Streamlining PDB Deposition
- Join our Biocurators to learn about preparing files and supplementary information, starting a deposition session, navigating OneDep, and more.
- Register for the Sept 9 Mol* Virtual Office Hour
- Join RCSB PDB for quick tips on how to use Mol* to view Sequence Annotations In 3D
- PDB Archive Serves Structures Determined by Integrative and Hybrid Methods

News Publications

Poster Prize Awarded at ISMB 2024

Congratulations to Yehlin Cho for Enhancing Protein Design Robustness through Noise-Informed Sequence Design

Biocurator Milestone: >10,000 Depositions Processed

Congratulations to wwPDB's Irina Persikova on processing more than 10,000 PDB structures

Paper Published: Exploring protein 3D similarities via comprehensive structural alignments

The RCSB.org alignment tool facilitates 3D superposition of multiple protein

Search Summary This query matches 984 Structures.

Refinements Structure Determination Methodology: experimental (984)

Scientific Name of Source Organism: Oryza sativa Japonica Group (227), Homo sapiens (204), Oryza sativa (98), Thaumatomoccus daniellii (63), Dioscoreophyllum cumminsii (38), Xanthomonas oryzae pv. oryzae KACC 10331 (37), Arabidopsis thaliana (26), Mus musculus (26), Oryza sativa Indica Group (22), Prunus dulcis (21), More...

Taxonomy: Eukaryota (848), Bacteria (108)

1 to 25 of 984 Structures Page 1 of 40 Sort by Score

5CTG The 3.1 Å resolution structure of a eukaryotic SWEET transporter

Tao, Y., Perry, K., Feng, L. (2015) Nature 527: 259-263

Released 2015-10-28

Method X-RAY DIFFRACTION 3.103 Å

Organisms Oryza sativa Japonica Group

Macromolecule Bidirectional sugar transporter SWEET2b (protein)

Unique Ligands BNG, PE5, TRS

5CTH The 3.7 Å resolution structure of a eukaryotic SWEET transporter

Sequence of 5CTG | The 3.1 Å resolution structure of a eukaryotic SWEET transporter

Structure

5CTG | The 3.1 Å resolution structure of a eukaryotic SWEET transporter

Type Assembly

Asm Id 1: Author And Software ...

Dynamic Bonds Off

Measurements

Components

Density

Quality Assessment

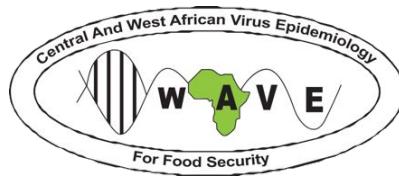
Assembly Symmetry

Export Models

Export Animation

Export Geometry

Carbon Capture Mechanisms



CIBiG 2024

Thank you for your attention!



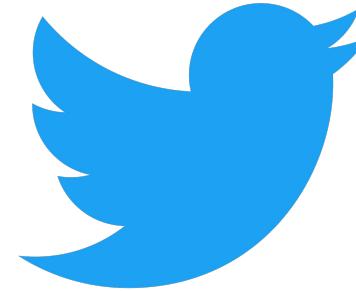
The teaching material used in these lessons is made available under the terms of the Creative Commons Attribution - Non-Commercial - Share Alike (BY-NC-SA) 4.0 International licence:

<http://creativecommons.org/licenses/by-nc-sa/4.0/>



CIBiG

FOLLOW US ON TWITTER!



FOLLOW US ON FACEBOOK !



FOLLOW US ON GITHUB !



CIBiG 2024

Don't forget to cite us!

How do you cite CIBiG?

"The authors acknowledge the International Certificate in Bioinformatics and Genomics (CIBiG) for providing capacity building that have contributed to the research results reported within this paper. URL: <https://cibig-wave.github.io>"