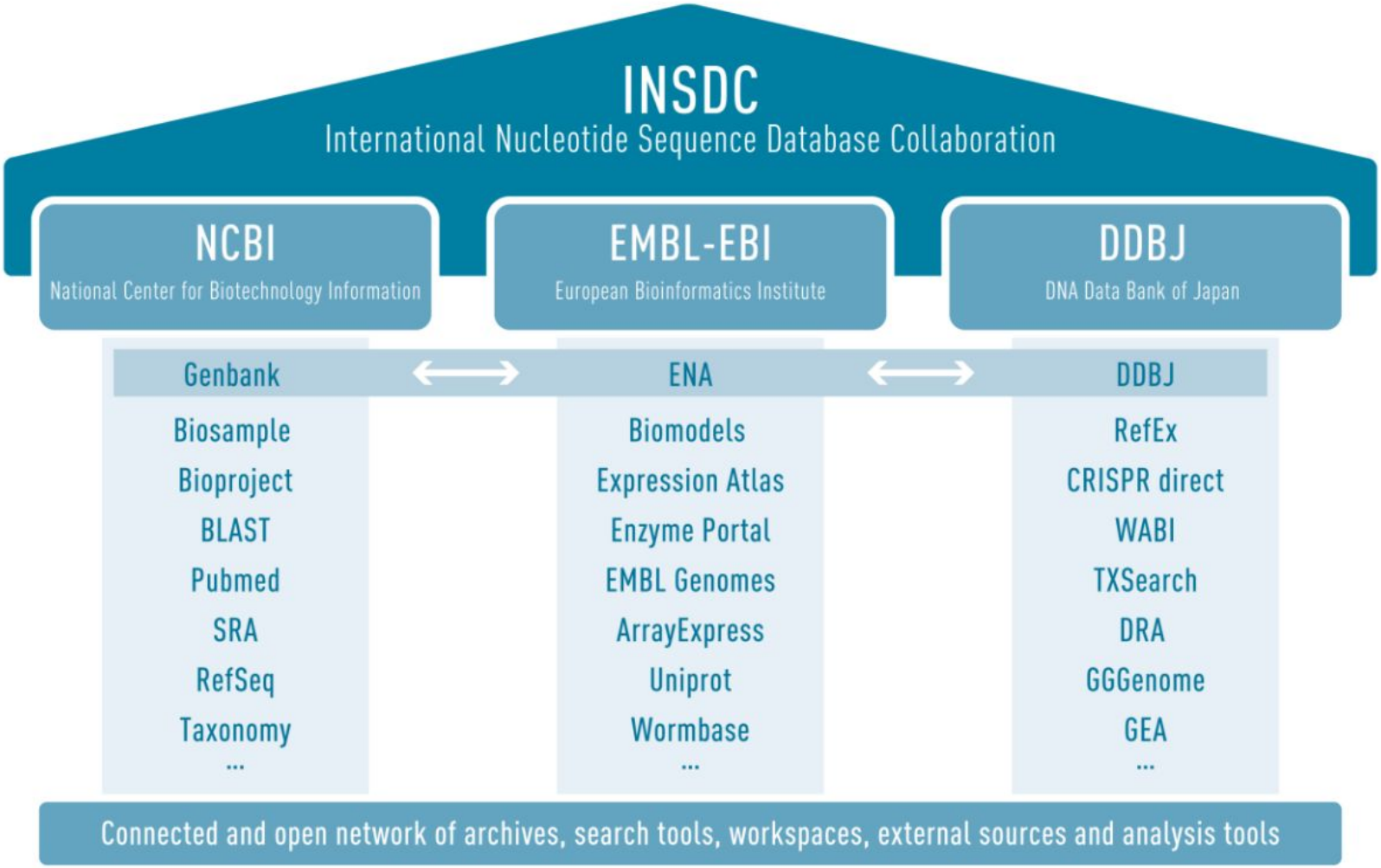
# Bases de Datos
# y
# Navegadores para plantas

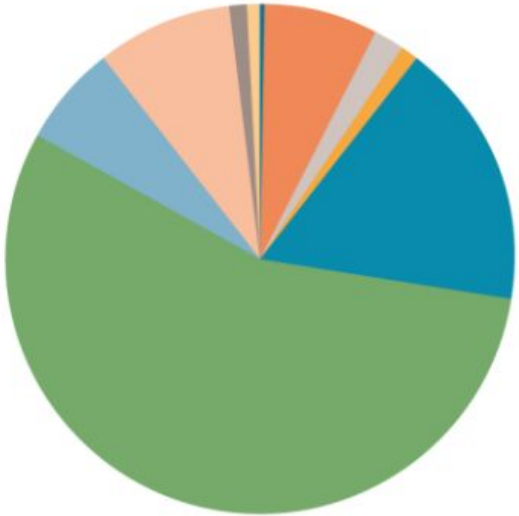

INSDC
International Nucleotide Sequence Database Collaboration
NCBI
National Center for Biotechnology Information
EMBL-EBI
European Bioinformatics Institute
DDBJ
DNA Data Bank of Japan
Genbank
ENA
DDBJ
Biosample
Bioproject
BLAST
Pubmed
SRA
RefSeq
Taxonomy
...
Biomodels
Expression Atlas
Enzyme Portal
EMBL Genomes
ArrayExpress
Uniprot
Wormbase
...
RefEx
CRISPR direct
WABI
TXSearch
DRA
GGGenome
GEA
...
Connected and open network of archives, search tools, workspaces, external sources and analysis tools

# Nucleotide sequence Data

## What is the biological scope of the NSD available in Genbank?

ENTRIES

BASE PAIRS

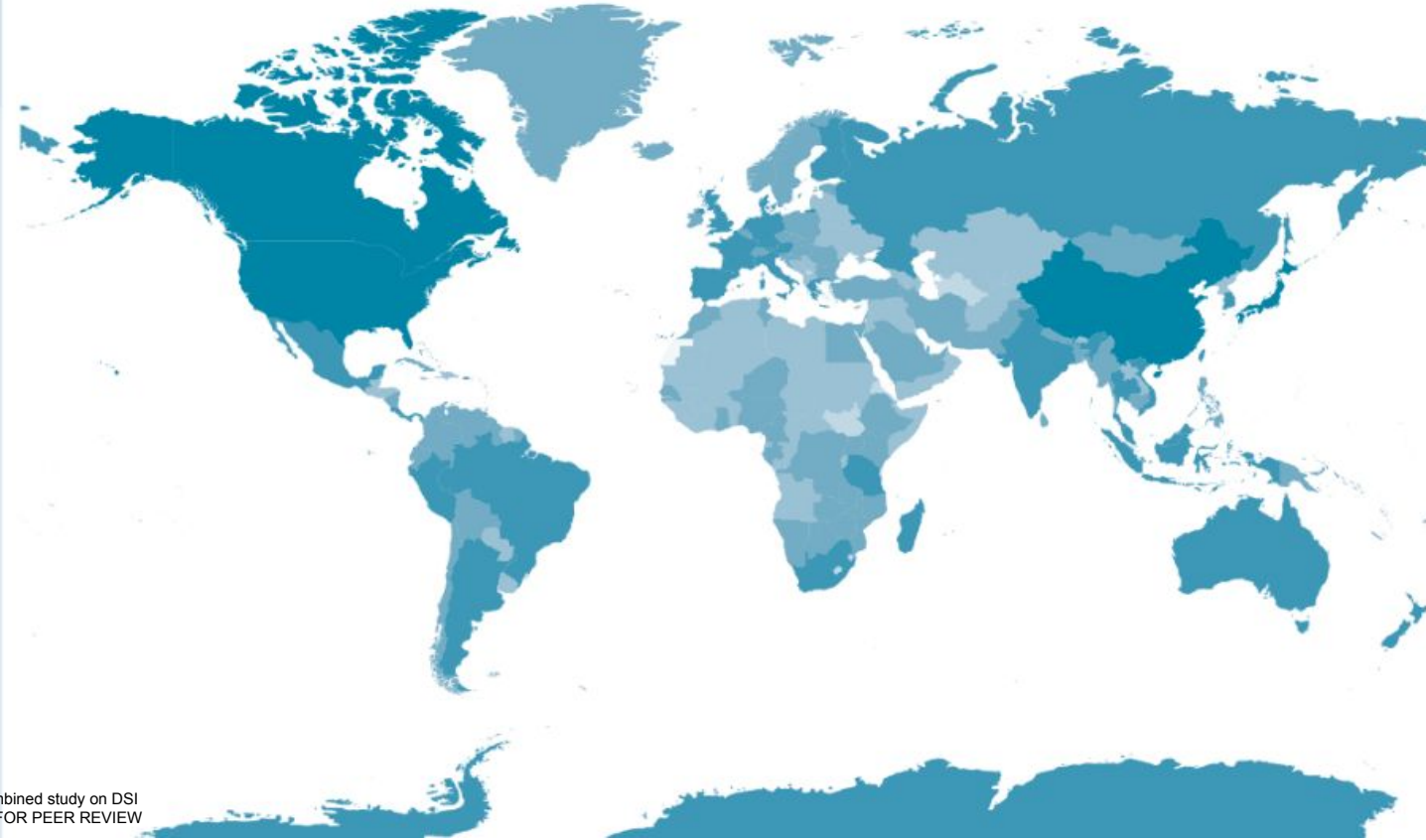


| TAXONOMIC GROUP | ENTRIES | BASE PAIRS |
|---|---|---|
| Viruses | 1.21% | 0.32% |
| Bacteria/Archaea | 5.60% | 7.37% |
| Fungi | 3.29% | 1.95% |
| Single-celled Eukaryotes | 2.56% | 0.91& |
| Plants & Algae | 21.61% | 17.12% |
| Animals | 28.45% | 55.62% |
| Human | 11.52% | 6.25% |
| Model Organisms | 11.93% | 8.67% |
| Other/synthetic (microorganisms) | 8.03% | 1.16% |
| Unidentified (no tax. ID) | 5.82% | 0.61% |

Rohden, F., Huang, S., Dröge, G., & Scholz, A. H. (2023). Combined study on DSI in public and private databases and DSI traceability. (DRAFT FOR PEER REVIEW )

## What is the country of origin for non-human NSD?

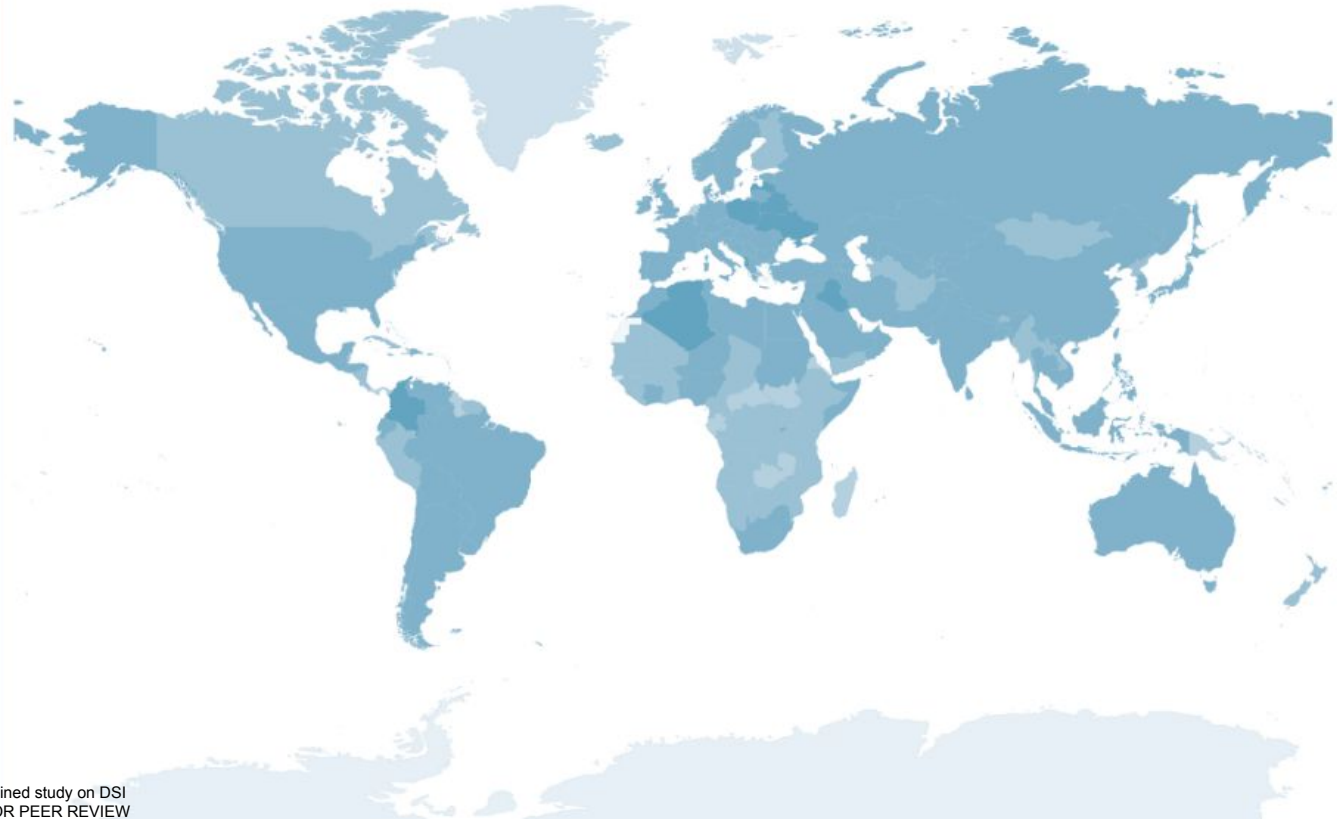
| | | |
|---|---|---|
| 1. | China | 18.23 % |
| 2. | United States | 17.39 % |
| 3. | Canada | 9.10 % |
| 4. | Japan | 7.24 % |
| 5. | India | 3.46 % |
| 6. | Australia | 2.66 % |
| 7. | Mexico | 2.54 % |
| 8. | Brazil | 2.30 % |
| 9. | Germany | 1.83 % |
| 10. | Spain | 1.58 % |



Rohden, F., Huang, S., Dröge, G., & Scholz, A. H. (2023). Combined study on DSI in public and private databases and DSI traceability. (DRAFT FOR PEER REVIEW )

## How does database usage compare to provided sequences?

| | | |
|---|---|---|
| 1. | Ukraine | 4.91 |
| 2. | Bahrain | 2.97 |
| 3. | Lebanon | 1.55 |
| 4. | Iraq | 1.48 |
| 5. | Albania | 1.45 |
| 6. | Colombia | 1.31 |
| 7. | Algeria | 1.23 |
| 8. | Singapore | 1.20 |
| 9. | Belarus | 1.18 |
| 10. | Poland | 1.13 |



Rohden, F., Huang, S., Dröge, G., & Scholz, A. H. (2023). Combined study on DSI in public and private databases and DSI traceability. (DRAFT FOR PEER REVIEW )

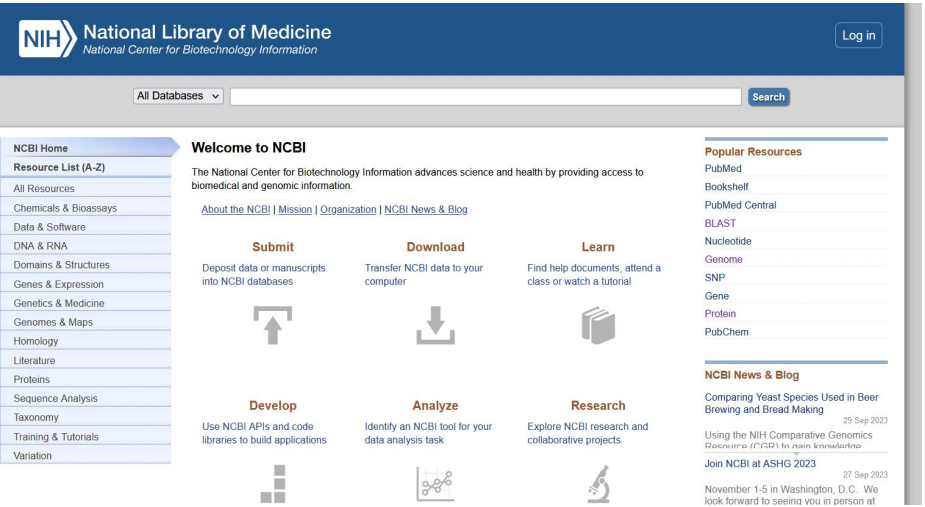
# NCBI

https://www.ncbi.nlm.nih.gov/

National Center for Biotechnology Information

NCBI

Se estableció el 4 de Nov de 1988




Objetivo del NCBI Desarrollar bases de datos públicas, herramientas de software, difundir información biomédica

Funciones del NCBI

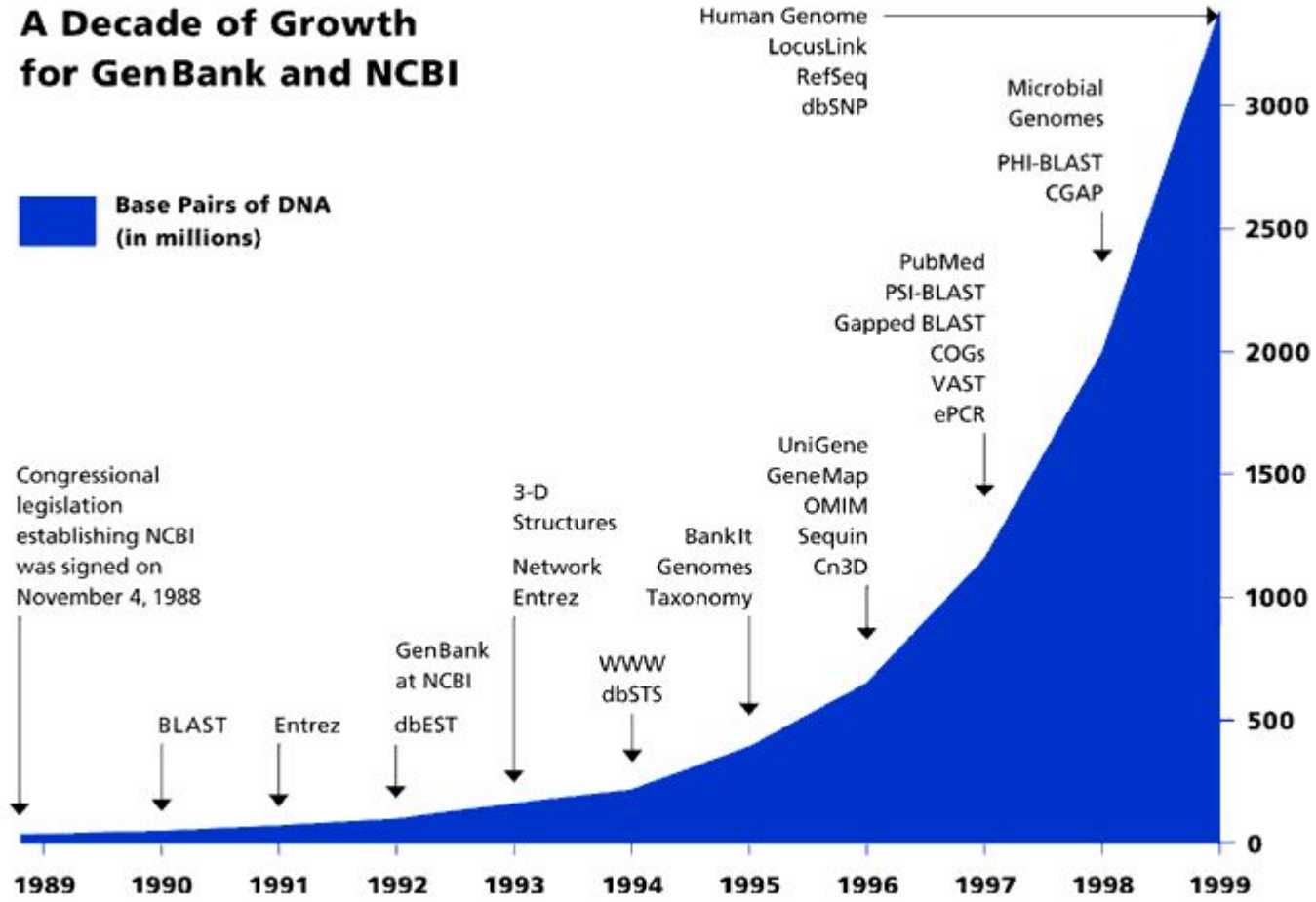
1.Mantenimiento de Bases de Datos Biológicas.

2. Proporcionar un sistema de recuperación de datos. p.ej. Entrez

3. Proporcionar recursos computacionales para el análisis de datos de GenBank.

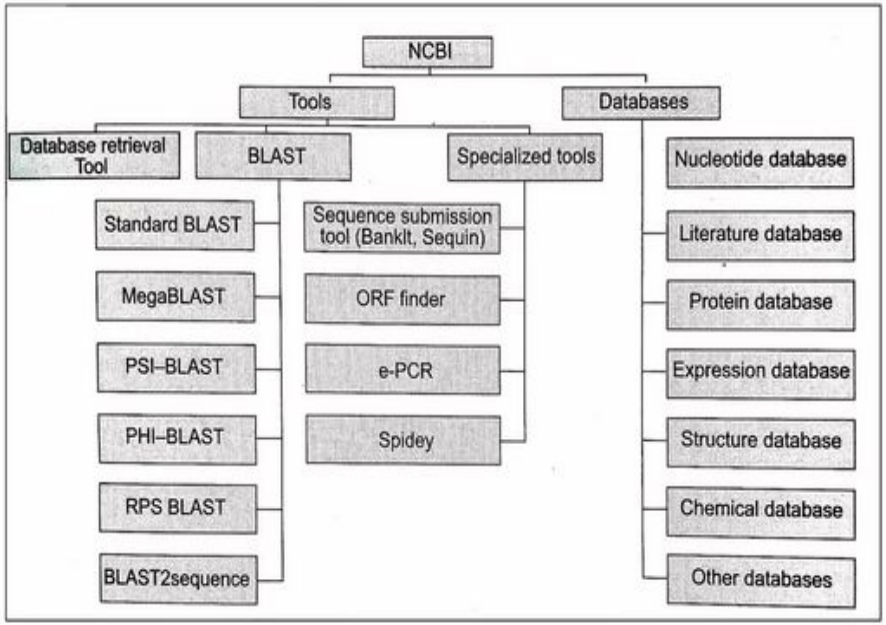
“ NCBI ofrece además algunas herramientas bioinformáticas para el análisis de secuencias de ADN, ARN y proteínas, siendo BLAST una de las más usadas”

Cañedo, 2009)




https://www.ncbi.nlm.nih.gov/Web/Newsltr/Summer99/decade.html

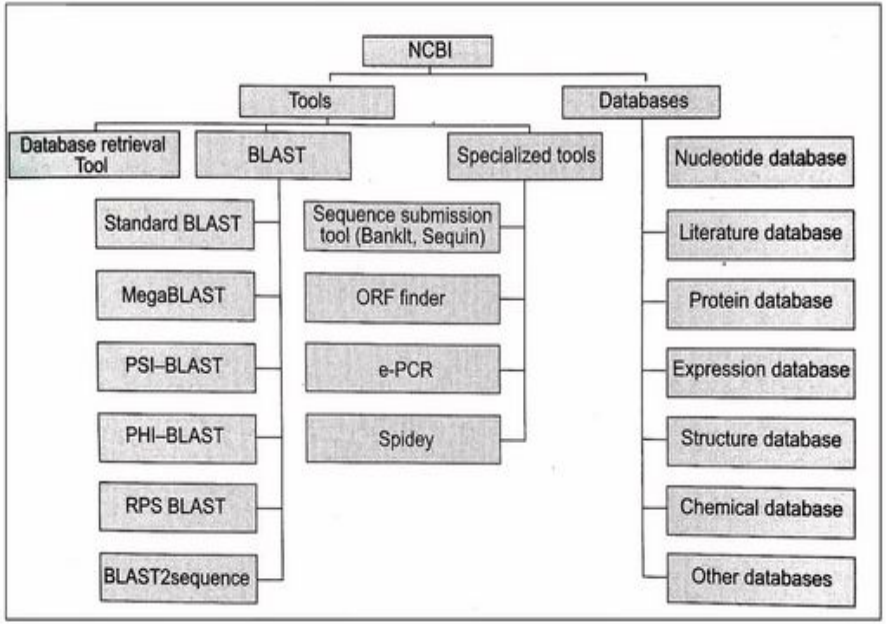
# Herramientas de NCBI




Basic Local Alignment Search Tool (BLAST)
Batch Entrez
BLAST Microbial Genomes
BLAST RefSeqGene
BLAST Tutorials and Guides
CDTree
Cn3D
COBALT
Coffee Break
Comparative Genome Viewer (CGV)
Conserved Domain Architecture Retrieval Tool (CDART)
Conserved Domain Search Service (CD Search)
Ebot
E-Utilities
Gene Expression Omnibus (GEO) BLAST
Genetic Codes
Genome BLAST
Genome Data Viewer (GDV)
Genome Remapping Service
Genome Workbench
LinkOut
Map Viewer
Multiple Sequence Alignment Viewer
NCBI News
NCBI Toolbox
Open Reading Frame Finder (ORF Finder)
OSIRIS
Performs a BLAST search of the genomic sequences in the RefSeqGene/LRG set. The default display provides ready navigation to review alignments in the Graphics display.
Phenotype-Genotype Integrator (PheGenI)
Primer-BLAST
ProSplign
Provides a configurable graphical display of a nucleotide or protein sequence and features that have been annotated on that sequence. In addition to use on NCBI sequence database pages, this viewer is available as an embeddable webpage component. Detailed documentation including an API Reference guide is available for developers wishing to embed the viewer in their own pages.
PSSM Viewer
PubChem Power User Gateway (PUG)
PubChem Standardization Service
PubChem Structure Search
PubMed Clinical Queries
PubMed Tutorials
Related Structures
Sequence Viewer
SNP Database Specialized Search Tools
Splign
Taxonomy Browser
Taxonomy Common Tree
Taxonomy Statistics
Taxonomy Status Reports
Tree Viewer
Variation Viewer
VecScreen
Vector Alignment Search Tool (VAST)
Viral Genotyping Tool

NCBI

# Bases de DATOS

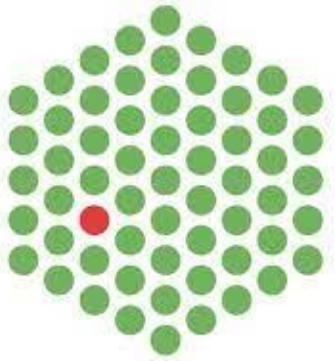



https://www.ncbi.nlm.nih.gov/guide/all/

Assembly
BioCollections
BioProject (formerly Genome Project)
BioSample
Bookshelf
ClinicalTrials.gov
ClinVar
Computational Resources from NCBI's Structure Group
Consensus CDS (CCDS)
Conserved Domain Database (CDD)
Database of Genomic Structural Variation (dbVar)
Database of Genotypes and Phenotypes (dbGaP)
Database of Short Genetic Variations (dbSNP)
GenBank
Gene
Gene Expression Omnibus (GEO) Database
Gene Expression Omnibus (GEO) Datasets
Gene Expression Omnibus (GEO) Profiles
GeneReviews
Genes and Disease
Genetic Testing Registry (GTR)
Genome
Genome Reference Consortium (GRC)
Glycans
HIV-1, Human Protein Interaction Database
Identical Protein Groups
Influenza Virus
Journals in NCBI Databases
MedGen
MeSH Database
National Library of Medicine (NLM) Catalog
NCBI C++ Toolkit Manual

NCBI Education Page
NCBI Glossary
NCBI Handbook
NCBI Help Manual
NCBI Pathogen Detection Project
Nucleotide Database
Online Mendelian Inheritance in Man (OMIM)
PopSet
Protein Clusters
Protein Database
Protein Family Models
PubChem BioAssay
PubChem Compound
PubChem Substance
PubMed
PubMed Central (PMC)
Reference Sequence (RefSeq)
RefSeqGene
Retrovirus Resources
SARS CoV
Sequence Read Archive (SRA)
Structure (Molecular Modeling Database)
Taxonomy
Third Party Annotation (TPA) Database
Viral Genomes
Virus Variation

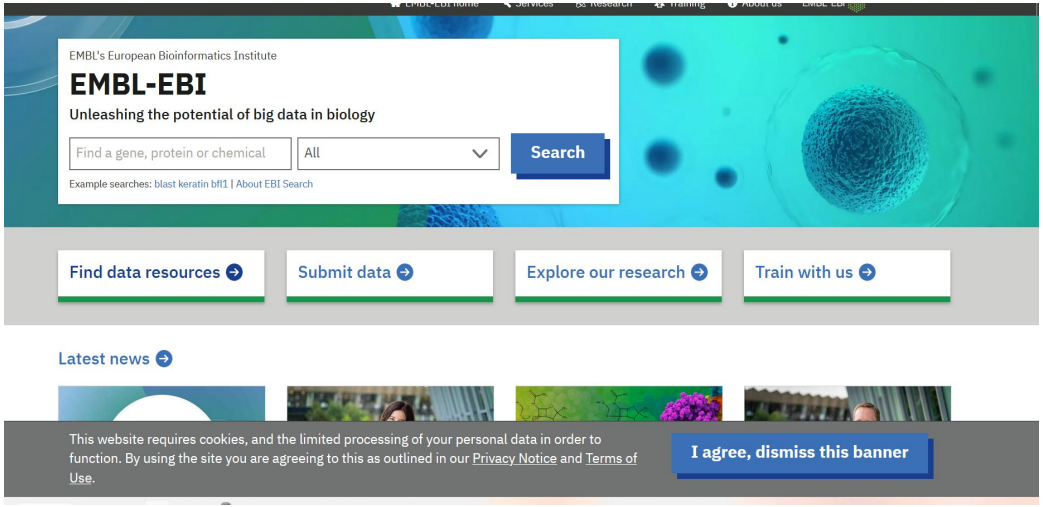
# EMBL-EBI



https://www.ebi.ac.uk/

# EMBL-EBI

European Molecular Biology Laboratory (EMBL)




En 1980, se creó en el EMBL Heidelberg la biblioteca de datos de secuencias de nucleótidos del EMBL (ahora parte del Archivo Europeo de Nucleótidos), con el objetivo de crear una base de datos central de secuencias de ADN.

# Herramientas de  EMBL-EBI

## Featured data resources

### AlphaFold DB

Database for protein structure predictions for numerous species

CC-BY

### Complex Portal

A manually curated, encyclopaedic resource of macromolecular complexes from model organisms.

CC0

### ArrayExpress

A database of functional genomics experiments, including microarray and RNAseq expression data typically related to publications.

Web API

## Featured tools

### Clustal Omega

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

Web API

### HMMER

Fast sensitive protein homology searches using profile hidden Markov models (HMMs) for querying against both sequence and HMM target databases.

Web API

### Annotation Platform

Consolidating text-mined and curated annotations

Web API

https://www.ebi.ac.uk/services/data-resources-and-tools

# DDBJ

https://www.ddbj.nig.ac.jp/

# DDBJ

Bioinformation and DDBJ Center provides sharing and analysis services for data from life science researches and advances science.

## Search

Retrieve the data from the database

## Submission

Navigation for how to submit your data

## Services

Services available in DDBJ Center

## Super Computer

NIG Supercomputer

## Statistics

Statistics of DDBJ Center services

## Activities

Training sessions and achievements of DDBJ Center

## NEWS

**Updated tools related to Mass Submission System (MSS)**
2023/09/27 Announcement DDBJ DDBJ Center

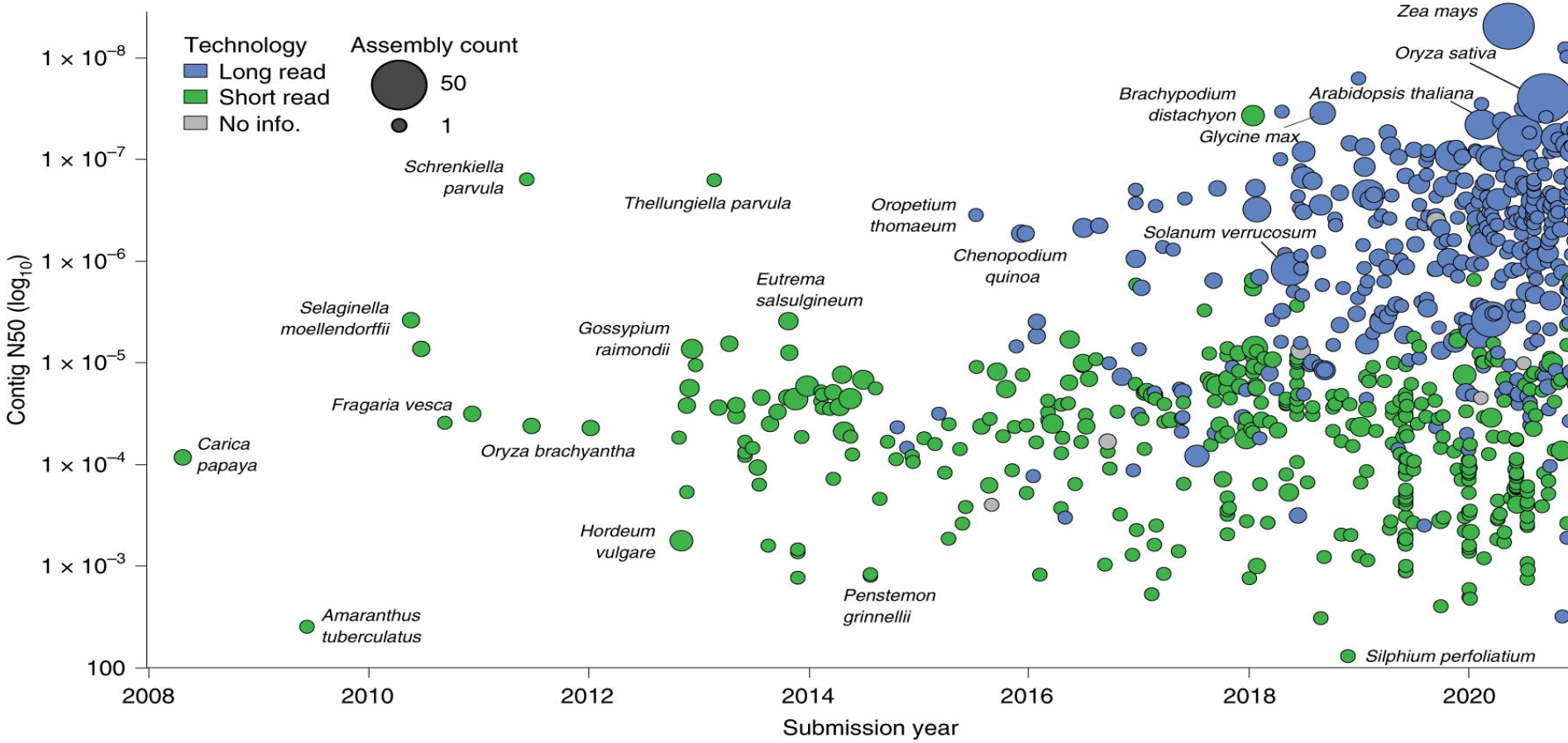
**[Resumed](September 27th) Suspension of DDBJ services due to NIG supercomputer maintenance**
2023/09/22 Maintenance DDBJ BioProject BioSample DRA GEA JGA AGD DDBJ Center

**Planned policy revisions**
2023/09/06 Announcement DDBJ BioProject BioSample MetaboBank DRA GEA JGA AGD DDBJ Center

more

# Plantas

# Genomas secuenciados de Plantas




https://www.nature.com/articles/s41477-021-01031-8

Assembly count
50
100
150
Domestication status
Domesticated
Cultivated
Feral
Natural commodity
Wild relative
Wild
Canada
USA
Mexico
Colombia
Brazil
Argentina
Chile
Sweden
Finland
United Kingdom
Denmark
Germany
Poland
Netherlands
Czech Republic
Belgium
Austria
France
Switzerland
Portugal
Spain
Italy
Turkey
Russia
Saudi Arabia
Oman
Nigeria
China
South Korea
Japan
Bangladesh
Taiwan
India
Thailand
Philippines
Malaysia
Singapore
Australia
New Zealand

# Navegadores enfocados en Plantas

National Center for Biotechnology Information
NCBI

EggNOG 5.0.0

http://eggnog5.embl.de/#/app/home

EnsemblPlants

https://plants.ensembl.org/index.html

KEGG Kyoto Encyclopedia of Genes and Genomes

https://www.genome.jp/kegg/genome/plant.html

JGI Phytozome 13
THE PLANT GENOMICS RESOURCE

https://phytozome-next.jgi.doe.gov/

Plaza

https://bioinformatics.psb.ugent.be/plaza/

# Bases de Datos en Plantas

**Table 4.5** List of plant species of relevant agronomical interest, sorted by the year of publication of their genomes. The annotation version available for these species in different reference databases is also reported

| Species | Year of publication | Annotation version | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | | NCBI | Ensembl Plants | PlantGDB | Phytozome | Plaza | Eggnog | KEGG |
| *Arabidopsis* | 2000 | TAIR 10 | TAIR 10 | TAIR 10 | TAIR 10 | TAIR 10 | TAIR 10 | TAIR 10 |
| Rice | 2002 | RGAP 7 | RGAP 7 | RGAP 7 | RGAP 7 | RGAP 7 | RGAP 7 | RGAP 7 |
| Grapevine | 2007 | GCF_000003745.3 | V1 Cribi | V2 Genoscope | V2 Genoscope | V2 Genoscope | V1 Cribi | GCF_000003745.3 |
| Maize | 2009 | B73_RefGen_v3 | B73_RefGen_v4 | B73_RefGen_v2 | B73_RefGen_v3 | B73_RefGen_v4 | B73_RefGen_v4 | B73_RefGen_v3 |
| Potato | 2011 | GCF_000226075.1 | GCF_000226075.1 | PGSC v.3 2.1.10 | PGSC v. 3.4 | iTAG v. 1 | GCF_000226075.1 | GCF_000226075.1 |
| Tomato | 2012 | GCF_000188115.3 | iTAG v. 2.4 | NA | iTAG v. 2.3 | iTAG v. 2.3 | iTAG v. 2.3 | GCF_000188115.3 |
| Sweet orange | 2013 | GCF_000317415.1 | NA | NA | JGI v1 | JGI v1 | NA | GCF_000317415.1 |
| Pepper | 2014 | GCF_000710875.1 | NA | NA | NA | NA | NA | NA |
| Amborella | 2013 | GCF_000471905.1 | GCF_000471905.1 | NA | AmTr_v_0.10 | AmTr_v_0.10 | NA | GCF_000471905.1 |
| Soybean | 2006 | GCF_000004515.4 | GCF_000004515.1 | Wm82.a2.v1 | Wm82.a2.v1 | Wm82.a2.v1 | GCF_000004515.1 | GCF_000004515.4 |
| Apple | 2010 | GCF_000148765.1 | NA | NA | M. Domestica v1.0 | M. Domestica v1.0 | NA | GCF_000148765.1 |
| Sorghum | 2009 | GCF_000003195.2 | GCF_000003195.2 | Sbi1.4 | Sbi3.1 | NA | GCF_000003195.2 | GCF_000003195.2 |
| Barrel medic | 2011 | GCF_000219495.2 | Mt4.0 | Mt3.5 | Mt4.0 | Mt4.0 | NA | GCF_000219495.2 |
| Banana | 2012 | GCF_000313855.1 | GCF_000313855.1 | NA | M. acuminata v 1 | NA | GCF_000313855.1 | GCF_000313855.1 |
| Cocoa | 2011 | GCF_000208745.1 | GCF_000208745.1 | NA | C. Matina v1.1 | C. Matina v1.1 | NA | GCF_000208745.1 |

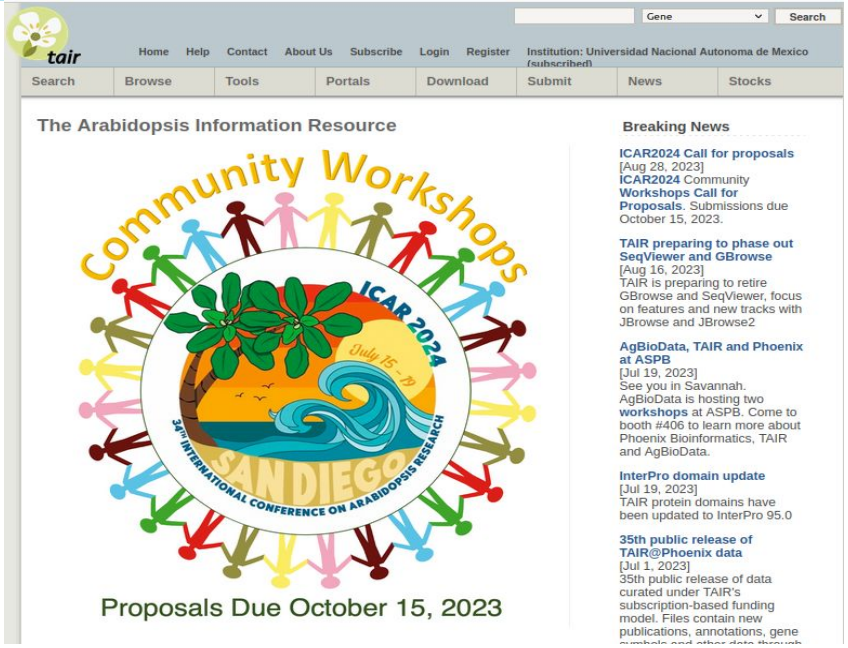
Ambrosino, L., Colantuono, C., Monticolo, F., & Chiusano, M. L. (2018). Bioinformatics resources for plant genomics: opportunities and bottlenecks in the-omics era. *Current Issues in Molecular Biology*, *27*(1), 71-88.

# TAIR

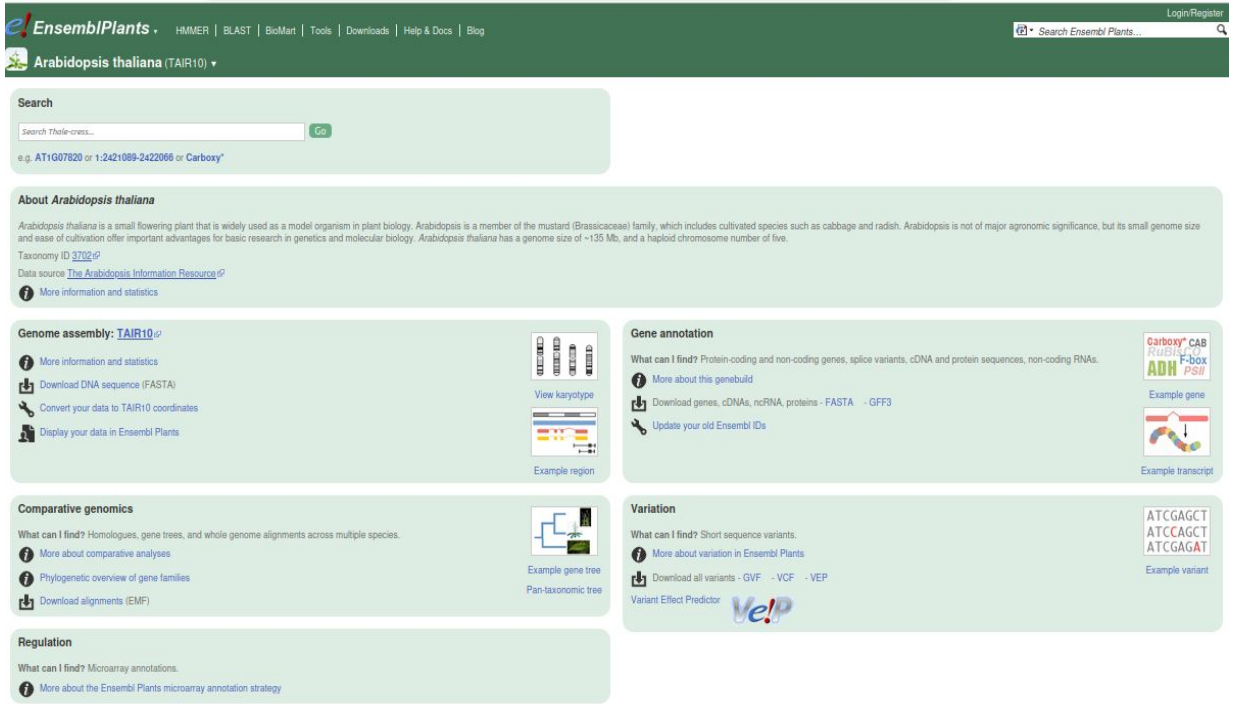
*The Arabidopsis Information Resource* (*TAIR*)

-Bases de Datos y Browser

https://www.arabidopsis.org/

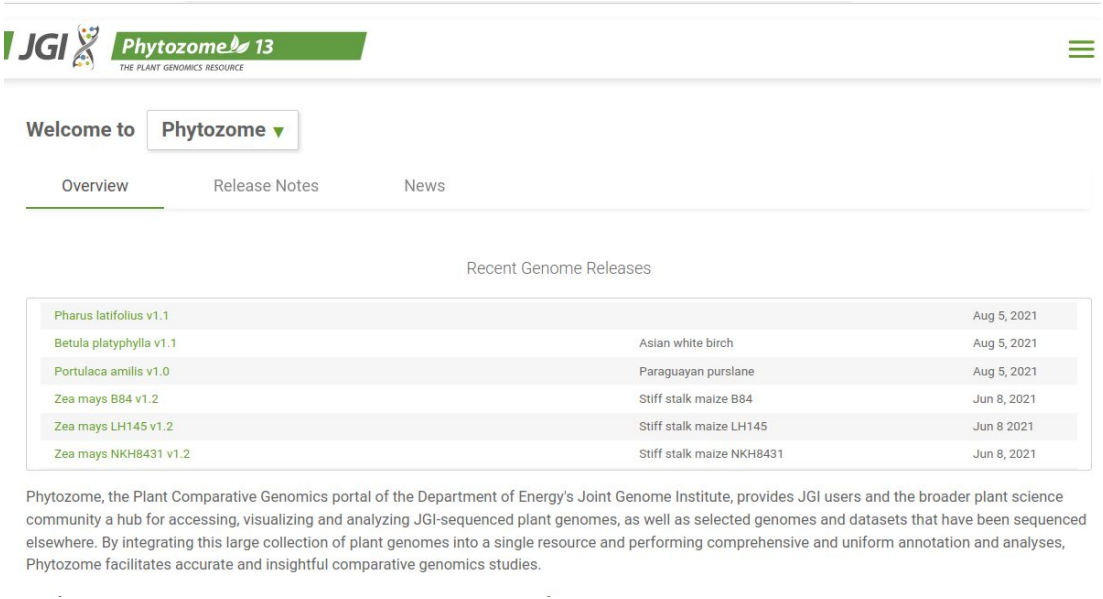


https://plants.ensembl.org/index.html



# JGI Phytozome 12

THE PLANT GENOMICS RESOURCE



# InterMINE - Phytozome

Manage Columns | Manage Filters | Manage Relationships | Generate Python code | Export

**Showing 1 to 9 of 9 rows**

| Gene Symbol | Gene Secondary Identifier | GO Annotation Ontology Term . Identifier | GO Annotation Ontology Term . Name | Code Code | Ontology Term Namespace |
|---|---|---|---|---|---|
| zen | CG1046 | GO:0000978 | RNA polymerase II cis-regulatory region sequence-specific DNA binding | IBA | molecular_function |
| zen | CG1046 | GO:0000981 | DNA-binding transcription factor activity, RNA polymerase II-specific | IBA | molecular_function |
| zen | CG1046 | GO:0000981 | DNA-binding transcription factor activity, RNA polymerase II-specific | IDA | molecular_function |
| zen | CG1046 | GO:0005634 | nucleus | IBA | cellular_component |

LIS - Legume Information System
Information about legume traits for crop improvement

**LegumeMine** v5.1.0.2 Genomic and expression data from LIS

LIS - Legume Information System
Information about legume traits for crop improvement

**MedicagoMine** v5.1.0.3 Medicago data from the Legume Information System

LegumeMine || ArachisMine | CicerMine | GlycineMine | LensMine | LupinusMine | PhaseolusMine | VignaMine | MedicagoMine

**MaizeMine** v1.5 An integrated data warehouse for MaizeGDB. Click here to go to MaizeMine v1.4 (previous release)

# AgBioData

Toward enhanced genomics, genetics, and breeding research outcomes through standardization of practices and protocols across agricultural databases

Agbase
https://agbase.arizona.edu/
AgroPortal
http://agroportal.lirmm.fr/
Alfalfa Breeders Toolbox
https://medicago.legumeinfo.org/
Animal QTLdb
http://www.animalgenome.org/cgi-bin/QTLdb/index
Bovine Genome Database
http://bovinegenome.org/
CassavaBase
https://www.cassavabase.org/
Citrus Genome Database
https://www.citrusgenomedb.org/
Citrus Greening
https://www.citrusgreening.org/
CottonGen
https://www.cottongen.org/
Crop Ontology
https://cropontology.org/
Cucurbit Genomics
http://cucurbitgenomics.org/
CyVerse
http://www.cyverse.org
EBS (Enterprise Breeding System)
https://ebs.excellenceinbreeding.org/
Ensembl
https://useast.ensembl.org/index.html
FAANG
https://data.faang.org/home
Genome Database for Rosaceae
https://www.rosaceae.org/
Genome Database for Vaccinium
https://www.vaccinium.org/
GOBii
https://excellenceinbreeding.org/toolbox/tools/gobii-genomic-data-management-system-gobii-gdm
GrainGenes
https://wheat.pw.usda.gov/GG3/
Gramene
http://www.gramene.org/
GRIN
https://www.ars-grin.gov
Hardwood Genomics
http://www.hardwoodgenomics.org/
Hymenoptera Genome Database
http://hymenopteragenome.org/
i5k Workspace@NAL
https://i5k.nal.usda.gov/
KitBase
http://kitbase.ucdavis.edu/
Legume Information System
https://legumeinfo.org/
MaizeGDB
https://maizegdb.org/
MusaBase
https://www.musabase.org/
National Animal Disease Center
https://www.ars.usda.gov/midwest-area/ames/nadc/
PeanutBase
https://peanutbase.org
Planteome
https://planteome.org/
Plant Metabolic Pathway Databases
https://plantcyc.org/
Pulse Crop Database
https://www.pulsedb.org/
Sol Genomics Network
https://solgenomics.net/
SorghumBase
https://www.sorghumbase.org
SoyBase
https://www.soybase.org
SweetPotatoBase
https://www.sweetpotatobase.org/
TAIR
http://www.arabidopsis.org
The Bio-Analytic Resource for Plant Biology (BAR)
https://bar.utoronto.ca/
The European Variation Archive
https://www.ebi.ac.uk/eva/
TreeGenes
https://treegenesdb.org
Triticeae toolbox, T3
https://triticeaetoolbox.org/wheat/
WheatIS
http://wheatis.org/
YamBase
https://www.yambase.org/

https://www.agbiodata.org/databases