

# Máster Universitario en Bioinformática

Generación y mantenimiento de datos ómicos

Curso académico 2024-25



**Universidad**  
Internacional  
de Valencia

Dr. Jordi Tronchoni  
[jordi.tronchoni@professor.universidadviu.com](mailto:jordi.tronchoni@professor.universidadviu.com)

23/04/2024

De:  
 Planeta Formación y Universidades

# Tema 4

## Bases de datos y herramientas bioinformáticas

23/04/2024

# Contenidos

## Tema 1. Introducción a la bioinformática

- 1.1 Historia de la bioinformática
- 1.2 Bioética aplicada al análisis de datos

## Tema 2. Principales flujos de trabajo en bioinformática

- 2.1 Genómica
- 2.2 Metagenómica y metataxonómica
- 2.3 Transcriptómica
- 2.4 Proteómica

## Tema 3. Gestión de entornos y paquetes

- 3.1 Conda

## Tema 4. Bases de datos y herramientas bioinformáticas

- 4.1 Principales bases de datos
- 4.2 Otros recursos online

## Tema 5. Alineamiento de secuencias

- 5.1 Introducción al alineamiento de secuencias
- 5.2 Alineamientos Pairwise
- 5.3 Alineamientos Múltiples

## Tema 6. Métodos de secuenciación

- 6.1 Primera generación de secuenciadores
- 6.2 Segunda generación de secuenciadores
- 6.3 Tercera generación de secuenciadores
- 6.4 Comparación de plataformas de secuenciación

## Tema 7. Pre-procesado y calidad de secuencias

- 7.1 Calidad de secuencias
- 7.2 Pre-procesado de secuencias

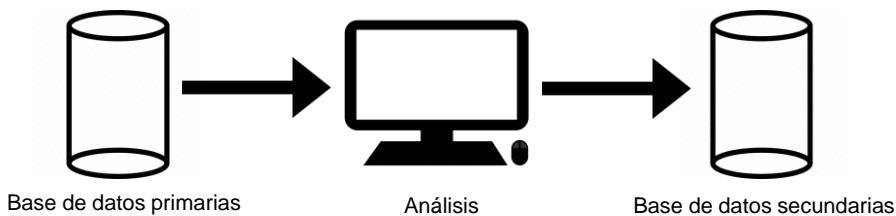
## Bases de datos y herramientas bioinformáticas

Las secuencias generadas se alojan en bases de datos dentro de cada uno de los campos de la bioinformática

Su principal función es la de almacenar toda la información generada a través de las estrategias de secuenciación y ensamblaje para poder caracterizar la nueva información que se generará en un futuro

- Bases de datos primarias
- Bases de datos secundarias
- Bases de datos especializadas

Generalmente la información es ofrecida como texto plano



23/04/2024

Toda la información generada a través de la secuenciación y análisis bioinformático se almacena en bases de datos. Generalmente cada una de las técnicas bioinformáticas tiene su conjunto de bases de datos, permitiendo no solo almacenar la nueva información generada si no la comparación y caracterización de nueva información respecto a lo almacenado. Estas bases de datos se dividen en tres grupos: bases de datos primarias, secundarias y especializadas. Las bases de datos primarias almacenan lo generado mediante las diversas técnicas, mientras que las secundarias almacenan la información procesada desde las bases de datos primarias. Por su parte, las especializadas reúnen información sobre, por ejemplo, organismos modelos concretos.

## Bases de datos y herramientas bioinformáticas

- Una de las principales herramientas en biología son las bases de datos
- Incluyen datos de secuencias de nucleótidos, proteínas, estructura de proteínas, ...etc
- Concebidas para la consulta humana, distintos programas capaces de leerlas
- Generales o específicas

Principales bases de datos se secuencias de nucleótidos:

 **Enlace 1**  
**NCBI** (EEUU)  
<http://www.ncbi.nlm.nih.gov/>

 **Enlace 2**  
**EMBL** (Europa)  
<http://www.ebi.ac.uk/embl/>

 **Enlace 3**  
**DDBJ** (Japón)  
<http://www.ddbj.nig.ac.jp>

Algunas bases de datos de genomas de organismos concretos: Principales bases de datos de proteínas:

 **Enlace 4**  
**Flybase** (Drosophila)  
<http://flybase.bio.indiana.edu/>

 **Enlace 5**  
**SGD** (Levadura)  
<http://www.yeastgenome.org/>

 **Enlace 6**  
**TAIR** (Arabidopsis)  
<http://www.arabidopsis.org/>

 **Enlace 8**  
**Uniprot**  
<http://www.pir.uniprot.org/>

 **Enlace 9**  
**Swiss-prot**  
<http://expasy.org/sprot/>

 **Enlace 10**  
**PDB**  
<http://www.rcsb.org/pdb/home/home.do>

23/04/2024

Dentro de las bases de datos primarias podemos encontrar bases de datos de nucleótidos, de proteínas y de estructuras. Las principales bases de datos primarias son GenBank, pertenecientes al grupo del NCBI; ENA, del grupo EBI; y DDBJ, "DNA Data Bank of Japan". La particularidad de estas bases de datos es que comparten información, sincronizándose entre ellas cada 24h. Por otra parte, la principal base de datos de proteínas es, a día de hoy, Uniprot. En esta base de datos tenemos información acerca de la secuencia de aminoácidos de todas las proteínas desde secuencias codificantes alojadas en las bases de datos de nucleótidos. La base de datos de estructura tridimensional de macromoléculas PDB almacena información relacionada con las coordenadas atómicas de las estructuras proteínas medidas por cristalografía de rayos X y RMN.

## Bases de datos y herramientas bioinformáticas

- Una de las principales herramientas en biología son las bases de datos
- Incluyen datos de secuencias de nucleótidos, proteínas, estructura de proteínas, ...etc
- Concebidas para la consulta humana, distintos programas capaces de leerlas
- Generales o específicas

Bibliografía:

-  **Enlace 13**  
**Pubmed**  
<http://www.NCBI.nlm.nih.gov/sites/entrez>
-  **Enlace 14**  
**Scopus**  
<https://www.Scopus.com/>
-  **Enlace 15**  
**Web of Knowledge**  
<http://www.webofknowledge.com/>

Rutas metabólicas:

-  **Enlace 16**  
**KEGG**  
<http://www.genome.jp/kegg/>

Enfermedades genéticas humanas:

-  **Enlace 17**  
**OMIM**  
<http://www.NCBI.nlm.nih.gov/sites/entrez?db=omim>

# Bases de datos y herramientas bioinformáticas

Principales bases de datos se secuencias de nucleótidos:



**Enlace 1**

**NCBI (EEUU)**

<http://www.ncbi.nlm.nih.gov/>



**Enlace 2**

**EMBL (Europa)**

<http://www.ebi.ac.uk/embl/>

- [EMBL](#)
  - [EBI](#)
- [NCBI](#)

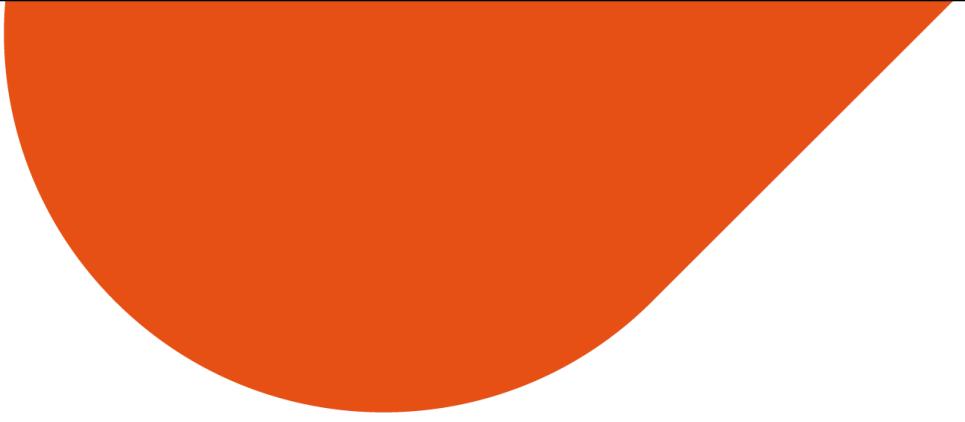


## 4.1 Principales bases de datos

EMBL

NCBI

En general podemos referirnos a ellas como Recursos de datos o más exactamente, recursos de datos y herramientas.



**EMBL → EBI → *data resources***

En general podemos referirnos a ellas como Recursos de datos o más exactamente, recursos de datos y herramientas.

## Bases de datos y herramientas bioinformáticas

- Laboratorio Europeo de Biología Molecular (EMBL)

**Institución de investigación en biología molecular financiada por 18 países europeos** que fue creada en 1974. Tiene laboratorios en Heidelberg (Alemania), Hamburgo (Alemania), Grenoble (Francia) y Hinxton (Reino Unido); un programa de investigación en Monterotondo (Italia), y **una subsede en Barcelona** (España), desde el 23 de diciembre de 2016.

### Funciones del EMBL:

- ✓ la realización de investigación básica en biología molecular;
- ✓ la formación y entrenamiento de científicos, estudiantes y visitantes;
- ✓ el patrocinio de programas activos de ciencia en la sociedad.

## EMBL – Heidelberg - GER



23/04/2024

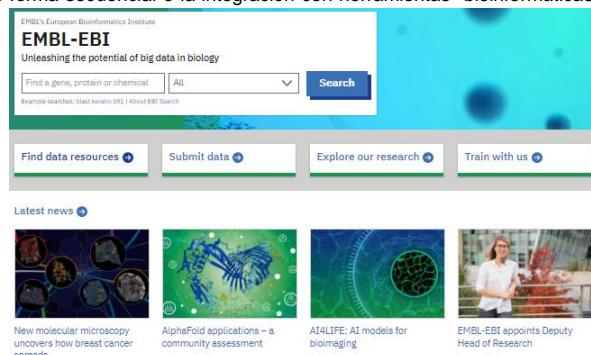
## EMBL – Heidelberg - GER



## Bases de datos y herramientas bioinformáticas

- Instituto Europeo de Bioinformática (EBI) (EMBL's European Bioinformatics Institute)
  - ✓ Líder a escala mundial en desarrollo y mantenimiento de recursos bioinformáticos.
  - ✓ Gama más amplia del mundo de recursos de datos moleculares disponibles de forma gratuita y actualizada.
  - ✓ Servicios web que permiten consultar sus bases de datos biológicos de manera programada, y posibilitan el desarrollo de análisis de forma secuencial o la integración con herramientas bioinformáticas de análisis propios y de otras instituciones.

<http://www.ebi.ac.uk/>



23/04/2024

## EBI - Hinxton - UK



EMBL-EBI



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

## EMBL-EBI data resources and tools

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

Find a data resource or a tool

Search

23/04/2024

EBI



EMBL's European Bioinformatics Institute

# EMBL-EBI

Unleashing the potential of big data in biology

Find a gene, protein or chemical

All

Science search

- Genomes & metagenomes
- Nucleotide sequences
- Protein sequences
- Small molecules
- Gene expression
- Gene-Disease Associations
- Diseases
- Molecular interactions
- Reactions & pathways
- Protein families
- Literature
- Samples & ontologies

Search web content

- EMBL-EBI People
- FMRI -FRT web

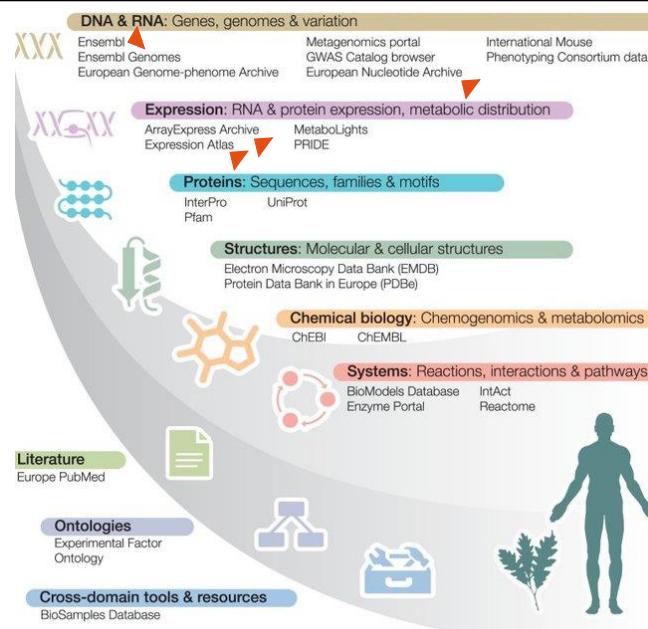
Find data resources →

Latest news →

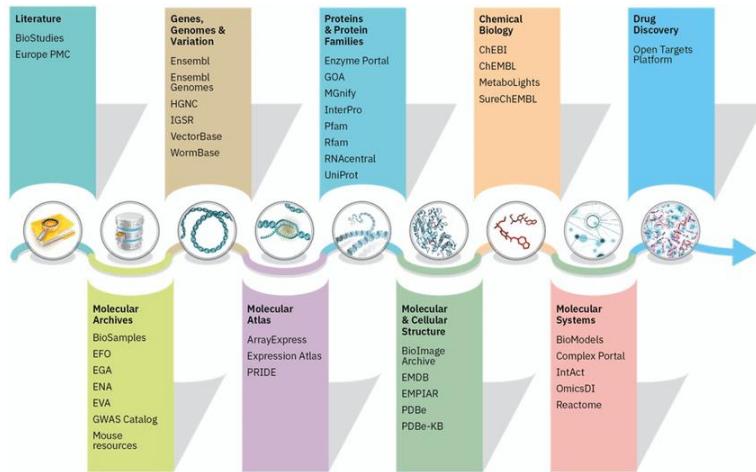
Explore our r

23/04/2024

## Data resources

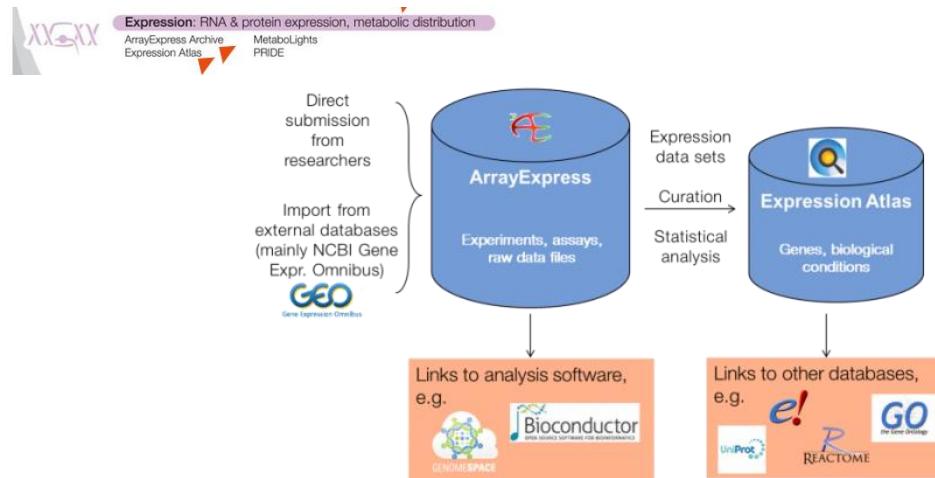


## Data resources



23/04/2024

## Bases de datos y herramientas bioinformáticas



23/04/2024

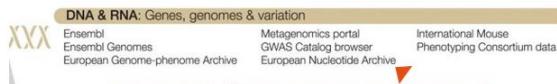
En realidad tendemos a llamar de forma generalista a todo bases de datos, y esto genera cierta confusión. Además las bases de datos tienen numerosos enlaces y a veces es difícil saber donde estas o porque un link te ha llevado donde te ha llevado. Aumenta la confusión que los datos estén duplicados en numerosas bases de datos...

Por ejemplo, ArrayExpress o Expression Atlas, son dos bases de datos. Dos bases de datos independientes, pero altamente conectadas. Mientras que ArrayExpress contiene Experimentos, o archivos en bruto, Expression Atlas contiene los resultados de análisis de ArrayExpress. En ambas bases de datos encontraremos múltiples links que nos llevaran fuera de la base de datos a herramientas de análisis de los datasets donde nos encontramos o a otras bases de datos ligadas al dataset que estábamos analizando.

ArrayExpress y Expression Atlas son dos bases de datos que se encuentran entre los recursos de datos del EBI, el Instituto Europeo de Bioinformática, que pertenece (cuelga de) al EMBL, el Laboratorio Europeo de Biología Molecular.

Debido a ello, la denominación, recursos de datos y herramientas bioinformáticas es más acertado.

## Bases de datos y herramientas bioinformáticas



- **European Nucleotide Archive (ENA)**, del Laboratorio Europeo de Biología Molecular (EMBL) e Instituto Europeo de Bioinformática (EBI).  
<https://www.ebi.ac.uk/ena/browser/home>

**ENA** forma parte -actualmente es su mayor contribuyente: **>925 millones de secuencias de ADN registradas**- de la **International Nucleotide Sequence Database Collaboration (INSDC)**, que también incluye el **GenBank del NCBI** -con >761 millones de secuencias- y el **DNA Data Bank of Japan (DDBJ)** -que contiene ~76 millones de secuencias.

23/04/2024

La base de datos ENA es la base de nucleótidos europea. Presenta una interfaz diferente a la del NCBI pero su contenido es similar. Presenta una particularidad, y es que almacena secuencias EST procedentes de transcriptomas

# Bases de datos y herramientas bioinformáticas

## European Nucleotide Archive (ENA)

The screenshot shows the ENA homepage with a teal header featuring the ENA logo and "European Nucleotide Archive". The header includes links for EMBL-EBI, Services, Research, Training, About us, and Search. Below the header is a search bar with fields for "Enter text search terms" and "Enter accession", both containing placeholder text like "Example: Taxon:9608, SRA00098, PRJEB40". A "Search" button is also present. The main content area has a light blue background with a message about subscribing to the ENA-announce mailing list. It also contains a note about SARS-CoV-2 data submissions and a Drag-and-Drop Data Submission Service. At the bottom are four teal buttons labeled "Submit", "Search", "Rulespace", and "Support". To the right of the main content is a sidebar showing a Twitter feed from "@ENASequencing" with the first tweet visible.

- Secuencias de ADN y ARN anotadas.
- Información complementaria:
  - ✓ detalles del ensamblaje de secuencias
  - ✓ Procedimientos experimentales
  - ✓ Otros metadatos relacionados con los proyectos de secuenciación (muestra, configuración experimental, configuración de la máquina, mapeo, ensamblaje, etc.)

23/04/2024

### European Nucleotide Archive (ENA)

#### Enlace de interés

<https://www.ebi.ac.uk/ena/submit/data-formats>

En esta página web se pueden ver los datos de ENA organizados en una serie de clases distintas divididas en tres niveles (**reads**, **assembly**, **annotation**), cada una de las cuales tiene sus propios formatos y estándares.

- How to Explore an ENA Project

## Organización de los datos

Data Domain	Description	Example Records
Projects/Studies	Contains information on a biological research project. This holds all the data generated as part of this research	PRJEB1787 (ERP001736)
Samples	Represents biological samples collected and sequenced in real life	SAMEA2620084 (ERS488919)
Reads (Runs/Experiments)	Hold raw read files and sequencing methods	ERR1701760 ERX1772048
Analyses	Hold results files of analyses performed on sequencing data and analysis methods	ERZ1195979
Contig set	Hold contig sets generated as part of a genome or transcriptome assembly.	CABHOY01000000.1
Assemblies	Represents an entire genome assembly and holds any contig sets or sequence records generated as part of the assembly	GCA_000001405.28
Assembled/Annotated Sequences (*)	Any sequence records from coding or non-coding regions to full assembled chromosomes	CM000667.2
Taxon	The sequenced organism or metagenome of a sample	Taxon:9604
Sample Checklist	The checklist of metadata that the sample was registered with	ERC000013

\* Assembled and annotated sequence records fall into different data classes. Read more about the different classes of sequences [here](#).

23/04/2024

## Bases de datos y herramientas bioinformáticas



- **Ensembl**, del Laboratorio Europeo de Biología Molecular (EMBL) e Instituto Europeo de Bioinformática (EBI).

<http://www.ensembl.org/index.html>

**Base de datos de genomas completos de organismos vertebrados** a través de la cual es posible realizar investigación en **genómica comparativa, evolución, variación de secuencia y regulación transcripcional**. Ensembl anota genes, permite efectuar alineaciones múltiples, predecir la función reguladora de un gen y recopilar datos de enfermedades.

23/04/2024

# Bases de datos y herramientas bioinformáticas

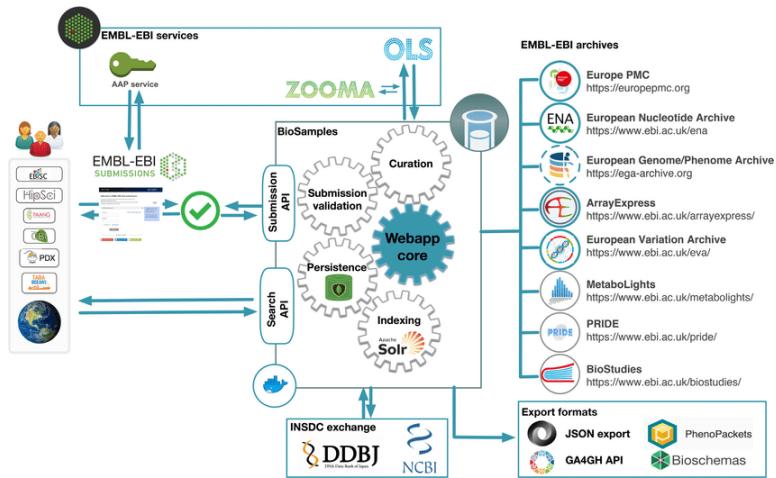
## Ensembl

The screenshot shows the Ensembl genome browser interface. At the top, there's a navigation bar with links for 'BLAST/BLAT', 'VEP', 'Tools', 'BioMart', 'Downloads', and 'More'. Below the navigation is a search bar with placeholder text 'Search all species' and a dropdown menu set to 'All species'. To the right of the search bar is a button with a magnifying glass icon. The main content area has several sections: 'Tools' (with links to 'BioMart', 'BLAST/BLAT', and 'Variant Effect Predictor'), 'Search' (with a text input field containing 'e.g. BRCA2 or rat 5:62797383..63627669 or rs699 or coronary heart disease'), 'All genomes' (with a dropdown menu 'Select a species' showing 'All species'), and 'Favourite genomes' (listing 'Human (GRCh38.p12)', 'Donkey (GRCh32?)', and 'Mouse (GRCm38.p6)'). On the right side, there's a sidebar with information about Ensembl (a genome browser for vertebrate genomes), the 'Ensembl Release 95 (January 2019)' (listing updates like new regulatory build for human, mouse GRCh38.p12, and updated genomes for various species), and other news from the blog.

- Incluye programas bioinformáticos, como, por ejemplo, **BLAST/BLAT**, **BioMart** y **Variant Effect Predictor (VEP)**, cuyas descripciones y usos aparecen detallados en los enlaces correspondientes.

23/04/2024

## Bases de datos y herramientas bioinformáticas



23/04/2024

El proceso de subida de datos a las bases de datos es complejo para intentar mantener un orden y una clasificación adecuada. Son muchas las solicitudes de subida de datos que se producen ya que para poder publicar un trabajo científico, un "paper" los datos deben de encontrarse en un repositorio público.

# Data submission

[EMBL-EBI home](#) > [Services](#) > [Data submission](#)

## Data submission

Use this data submission wizard to find the right archive for your data in a few simple steps.

- 1 What type of data do you have?



### Why submit data to an archive?

- Submission of primary data and derived information to public data repositories is an essential step in the scientific process.
- Through submission, the scientific community is fed the raw materials for the building and maintenance of the complete and up-to-date data sets that support searches and analysis on the latest sequences, structures and molecular profiles of living systems.
- Serving as a complement to the literature publication process and supporting early data sharing, the EMBL-EBI offers a number of submission services appropriate for different types and scales of data.

### Need help?

If you need help with your data submission, please [contact support](#).

### All EMBL-EBI data repositories

Array Express: functional genomics data  
BioImage Archive: bioimaging data  
BioModels: computational models  
BioSamples: reference sample data  
BioStudies: biological research data  
ChEBI: chemical entities  
DGVa: structural genetic variation data  
EFO: experimental variables  
EGA: human data that requires controlled access  
EMPIAR: raw image data  
ENA: nucleotide sequence data  
EVA: genetic variation data  
GO: Gene ontology annotations  
GWAS Catalog: Genome-wide association study data  
IntAct: molecular interactions  
IntEnz: enzyme nomenclature  
Metabolights: metabolomics data  
Metagenomics: raw sequence data & associated meta-data

23/04/2024

Para ello el EBI cuenta con una herramienta para “facilitar” la subida de los datos de forma correcta, esto es bien etiquetados (metadatos) al repositorio correcto.

## Data submission: <https://www.ebi.ac.uk/submit/>

[EMBL-EBI home](#) > [Services](#) > [Data submission](#)

### Data submission

Use this data submission wizard to find the right archive for your data in a few simple steps.

1 What type of data do you have?



#### Why submit data to an archive?

- Submission of primary data and derived information to public data repositories is an essential step in the scientific process.
- Through submission, the scientific community is fed the raw materials for the building and maintenance of the complete and up-to-date data sets that support searches and analysis on the latest sequences, structures and molecular profiles of living systems.
- Serving as a complement to the literature publication process and supporting early data sharing, the EMBL-EBI offers a number of submission services appropriate for different types and scales of data.

#### Need help?

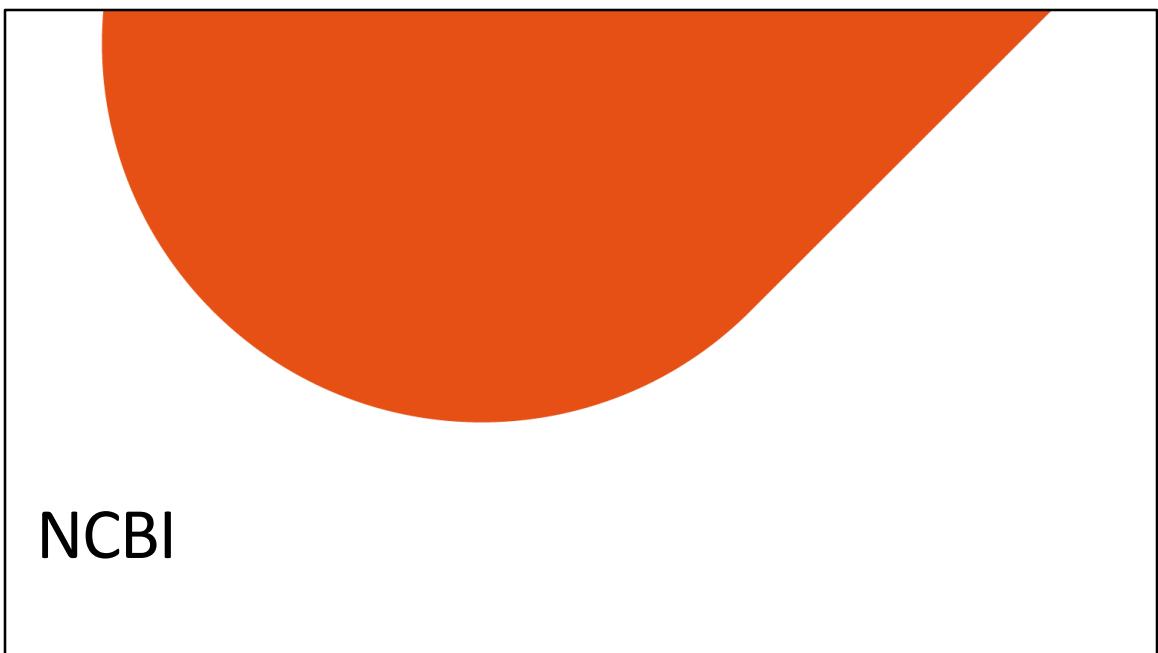
If you need help with your data submission, please [contact support](#).

#### All EMBL-EBI data repositories

Array Express: functional genomics data  
BioImage Archive: bioimaging data  
BioModels: computational models  
BioSamples: reference sample data  
BioStudies: biological research data  
ChEBI: chemical entities  
DGVa: structural genetic variation data  
EFO: experimental variables  
EGA: human data that requires controlled access  
EMPIAR: raw image data  
ENA: nucleotide sequence data  
EVA: genetic variation data  
GO: Gene ontology annotations  
GWAS Catalog: Genome-wide association study data  
IntAct: molecular interactions  
IntEnz: enzyme nomenclature  
Metabolights: metabolomics data  
Metagenomics: raw sequence data & associated meta-data

23/04/2024

A pesar de lo que indican, la realidad es otra, es complejo y bastante tedioso, razón por la que mayoritariamente, no encontramos los datos subidos correctamente o etiquetados correctamente. Es muy recomendable no dejar este paso para el final, hacerlo correctamente lleva tiempo. Es interesante revisar toda la documentación antes de abordar el depósito de los datos.



## Bases de datos y herramientas bioinformáticas

- Centro Nacional para la Información Biotecnológica de los Estados Unidos (NCBI).

Es parte de la Biblioteca Nacional de Medicina de los Estados Unidos (National Library of Medicine), una rama de los Institutos Nacionales de Salud (National Institutes of Health, NIH). Está localizado en Bethesda, Maryland, y fue fundado el 4 de noviembre de 1988 con **la misión de ser una importante fuente de información y desarrollo en biología molecular.**

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

## Bases de datos y herramientas bioinformáticas

- Centro Nacional para la Información Biotecnológica de los Estados Unidos (NCBI).

The screenshot shows the main page of the NCBI website. At the top, there's a navigation bar with links for 'NCBI', 'Resources', 'How To', 'All Databases', and a search bar. On the left, there's a sidebar with a 'Resource List (A-Z)' menu containing items like 'NCBI Home', 'Chemicals & Bioassays', 'Data & Software', 'DNA & RNA', 'Genes & Expression', 'Genetics & Medicine', 'Genomes & Maps', 'Homology', 'Literature', 'Proteins', 'Sequence Analysis', 'Taxonomy', 'Training & Tutorials', and 'Variation'. The main content area has several sections: 'Welcome to NCBI' with a brief introduction, 'Popular Resources' (PubMed, Bookshelf, PubMed Central, BLAST, Nucleotide, Genome, SNP, Gene, Protein, PubChem), 'NCBI News & Blog' (with a recent entry about the novel coronavirus genome), and four main functional sections: '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), and 'Research' (Explore NCBI research and collaborative projects). Each section includes a small icon and a brief description.

23/04/2024

## Bases de datos y herramientas bioinformáticas

- Centro Nacional para la Información Biotecnológica de los Estados Unidos (NCBI).

El **NCBI** ofrece un número variado de recursos científicos, p. ej.:

**GenBank** (almacena y constantemente actualiza información referente a secuencias genómicas)

**PubMed** (contiene resúmenes y citas de artículos científicos referentes a biomedicina, biotecnología, bioquímica, genética y genómica)

**OMIM** (contiene una recopilación de enfermedades genéticas humanas)

**dbSNP** (base de datos de polimorfismos de nucleótidos simples)

**RefSeq** (incluye la anotación precisa de genes conocidos y permite la predicción de nuevos genes basados en la evidencia de transcripción disponible)

**BLAST** (alineamiento múltiple de secuencias). <https://blast.ncbi.nlm.nih.gov/Blast.cgi>

## Bases de datos y herramientas bioinformáticas

- Centro Nacional para la Información Biotecnológica de los Estados Unidos (NCBI).

Todas las bases de datos del NCBI están disponibles en [línea de manera gratuita](#) y son accesibles usando el buscador [Entrez](#) (barra de búsqueda localizada en la parte superior central de cada una de las bases de datos).

## Bases de datos y herramientas bioinformáticas

### Genbank

- Creada en 1982 por Walter Goad y el Laboratorio Nacional de Los Álamos
- >761 millones de secuencias de ADN de más de 300.000 especies registradas
- es parte también de la *International Nucleotide Sequence Database Collaboration (INSDC)*.
- crece a un ritmo exponencial (se duplica aprox cada 18 meses).
- obtiene sus datos principalmente a través de los **registros de laboratorios individuales** y de **lotes de proyectos de secuenciación a gran escala** que incluyen proyectos de WGS y proyectos de muestreo ambiental.
- la mayoría de los registros se realizan utilizando los programas **BankIt**, **Tbl2asn**, **Submission Portal** y **Sequin**, y el personal de GenBank asigna números de acceso a los datos registrados.

23/04/2024

# Bases de datos y herramientas bioinformáticas

## Genbank

<https://www.ncbi.nlm.nih.gov/genbank/>

Sign in to NCBI

NCBI Resources How To

GenBank Nucleotide

GenBank Submit Genomes WGS Metagenomes TPA TSA INSDC Other

Search

**GenBank Overview**

**What is GenBank?**

GenBank® is the NIH genetic sequence database, an annotated collection of all publicly available DNA sequences (*Nucleic Acids Research*, 2010 Jan;31(1):139-42). GenBank is part of the International Nucleotide Sequence Database Collaboration, which comprises the DNA Database of Japan (DDBJ), the European Nucleotide Archive (ENA), and GenBank at NCBI. These three organizations exchange data on a daily basis.

A GenBank release occurs every two months and is available from the [its site](#). The release notes for the current version of GenBank provide detailed information about the release and notifications of upcoming changes to GenBank. Release notes for previous GenBank releases are also available. GenBank growth statistics for both the traditional GenBank divisions and the WGS division are available from each release. GenBank growth [statistics](#) for both the traditional GenBank divisions and the WGS division are available from each release.

An annotated sample [GenBank record](#) for a *Saccharomyces cerevisiae* gene demonstrates many of the features of the GenBank flat file format.

**Access to GenBank**

There are several ways to search and retrieve data from GenBank.

- Search GenBank for sequence identifiers and annotations with [Eutils-Nucleotide](#).
- Search and align GenBank sequences to a query sequence using [BLAST](#) (Basic Local Alignment Search Tool). BLAST searches CoreNucleotide, dbEST, and dbSSR independently; see [BLAST.info](#) for more information about the numerous BLAST databases.
- Search, link, and download sequences programmatically using [NCBI e-utilities](#).
- The ASN.1 and flatfile formats are available at NCBI's anonymous FTP server: <ftp://ftp.ncbi.nlm.nih.gov/ncbi-ssn1> and <ftp://ftp.ncbi.nlm.nih.gov/genbank>.

**GenBank Data Usage**

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,

23/04/2024

# Bases de datos y herramientas bioinformáticas

## Genbank

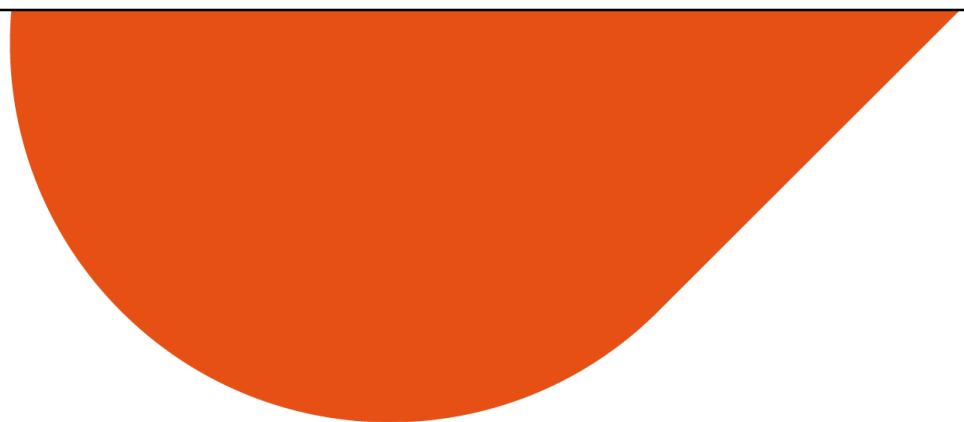
Ejemplo de un registro de datos en GenBank ([U49845](#), relativo al gen **TCP1-beta** de *Saccharomyces cerevisiae*) con el formato *GenBank Flat Format* (GFF):

<https://www.ncbi.nlm.nih.gov/genbank/samplerecord/>

The screenshot shows the NCBI GenBank sample record page for accession number U49845. The page title is "Sample GenBank Record". A note at the top states: "This page presents an annotated sample GenBank record (accession number [U49845](#)) in its *GenBank Flat File* format. You can see the corresponding [live record](#) for U49845, and see [examples of other records](#) that show a range of biological features." Below this is a table containing the GFF data for the TCP1-beta gene.

LOCUS	SCU49845	5028 bp	DNA	PLN	21-JUN-1999
DEFINITION	Saccharomyces cerevisiae TCP1-beta gene, partial cds, and Ax12p (AXL12) and Rev1p (REV7) genes, complete cds.				
ACCESSION	U49845				
VERSION	U49845.1	<a href="#">GX:1293613</a>			
KEYWORDS	.				
SOURCE	Saccharomyces cerevisiae (baker's yeast)				
ORGANISM	Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes; Saccharomycetales; Saccharomyctaceae; Saccharomyces.				
REFERENCE	1 (bases 1 to 5028)				
AUTHORS	Torpey,L.E., Gibbs,P.E., Nelson,J. and Lawrence,C.W.				
TITLE	Cloning and sequence of REV7. A gene whose function is required for				

23/04/2024



Uniprot; PDB; KEGG; SGD...

## Bases de datos biológicas

### Bases de datos primarias

**UniProt (Universal Protein Resource)** - EBI + Swiss Institute of Bioinformatics (SIB) + Georgetown University.  
Aúna Swiss-Prot + TrEMBL

- Swiss-Prot : Proteínas curadas manualmente
- TrEMBL: Automáticamente

UniProt Knowledgebase (UniProtKB) – Secuencias proteínas y sus anotaciones

UniProt Reference Clusters Database (UniRef)

23/04/2024

UniProt es la principal base de datos de proteínas la cual aúna información de otras bases de datos. Es una combinación de distintos institutos, como el EBI, el instituto suizo y la universidad de Georgetown. Esta base de datos aúna información de Swiss-Prot y TrEMBL. Estas dos bases de datos se diferencian principalmente porque Swiss-Prot revisa manualmente toda la información anotada en la misma, mientras que en TrEMBL se analiza de forma automática.

# Bases de datos biológicas

## Bases de datos primarias

The screenshot shows the UniProt homepage with a banner announcing "The new UniProt website is here! Take me to UniProt BETA". Below the banner, there are sections for UniProtKB (Swiss-Prot and TrEMBL), UniRef, UniParc, Proteomes, Supporting data (Literature citations, Taxonomy, Subcellular locations), and News. A sidebar on the right provides information about the SARS-CoV-2 portal and recent releases.

The mission of UniProt is to provide the scientific community with a comprehensive, high-quality and freely accessible resource of protein sequence and functional information.

**UniProtKB**  
UniProt Knowledgebase  
Swiss-Prot (565,254)  
Manually annotated and reviewed.  
Records with information extracted from literature and curator-evaluated computational analysis.

**UniRef**  
The UniProt Reference Clusters (UniRef) provide clustered sets of sequences from the UniProt Knowledgebase (including isoforms) and selected UniParc records.

**UniParc**  
UniParc is a comprehensive and non-redundant database that contains most of the publicly available protein sequences in the world.

**Proteomes**  
A proteome is the set of proteins thought to be expressed by an organism. UniProt provides proteomes for species with completely sequenced genomes.

**Supporting data**  
Literature citations, Taxonomy, Subcellular locations

**News**  
Forthcoming changes  
There are currently no changes planned

UniProt release 2021\_03  
The Importance of being disordered | MobiDB-lite predictions for intrinsically disordered regions | UniProtKB via AWS Open Data and Amazon...

UniProt release 2021\_02  
With a little help from my friend | SwissBioPics subcellular

We'd like to inform you that we have updated our [Privacy Notice](#) to comply with Europe's new General Data Protection Regulation (GDPR) that applies since 25 May 2018.

Do not show this banner again

23/04/2024

Esta es la imagen que presenta su página de inicio.

# Bases de datos biológicas

## Bases de datos primarias

The new UniProt website is here! Take me to UniProt BETA

UniProtKB - I3WFC7 (I3WFC7\_9BETA)

Display Help video BLAST Align Format Add to basket History Add a publication Feedback

Entry Protein Capsid scaffolding protein  
Gene N/A  
Organism Macacine betaherpesvirus 3 (Rhesus cytomegalovirus)  
Status Unreviewed - Annotation score: 0 - Protein inferred from homology

Function

Names & Taxonomy

Subcellular location

PTM / Processing

Interaction

Structure

23/04/2024

Podemos buscar el nombre de una proteína y acceder a su ficha.

## Bases de datos biológicas

### Bases de datos secundarias

ExPASy

#### Prosite (expasy)

Una base de datos proteicas clasificada por motivos proteicos, relacionados con función

Un motivo deriva de alineamientos múltiples y se guarda la consenso

The screenshot shows the PROSITE homepage. At the top, there's a navigation bar with links to Home, ScanProsite, ProRule, Documents, Downloads, Links, and Funding. Below the navigation is the PROSITE logo and the text "Database of protein domains, families and functional sites". A section titled "SARS-CoV-2 relevant PROSITE motifs" is visible. The main area has two main sections: "Search" (with a search bar and dropdown menu) and "Browse" (with a list of filtering options). Below these is a "Quick Scan mode of ScanProsite" section with a text input field for protein sequences. To the right, there's a "Other tools" section with a link to "PRATT". The bottom right corner of the page displays the date "23/04/2024".

Entre las bases de datos secundarias encontramos aquellas que derivan información desde la secuencia de aminoácidos. Prosite es una de ellas, y guarda información de motivos proteicos, los cuales se relacionan con la función de las proteínas

## Bases de datos biológicas



PROSITE documentation PDOC00022

### Fibronectin type-II collagen-binding domain signature and profile

Description Technical section References Copyright Miscellaneous

#### Description

Fibronectin is a plasma protein that binds cell surfaces and various compounds including collagen, fibrin, heparin, DNA, and actin. The major part of the sequence of fibronectin consists of the repetition of three types of domains, which are called type I, II, and III [1]. Type II domain (FN2) is approximately 40 residues long, contains four conserved cysteines involved in disulfide bonds and is part of the collagen-binding region of fibronectin [2]. In fibronectin the minimal collagen binding region is formed by one FN1 and two FN2 domains. This suggests that the collagen-binding sites spans multiple modules.

A schematic representation of the position of the invariant residues and the topology of the disulfide bonds in FN2 domain is shown below.



'C': conserved cysteine involved in a disulfide bond.

'#': large hydrophobic residue.

The 3D-structure of the FN2 domain has been determined (see <PDB:2FN2>) [3]. The structure consists of two double-stranded anti-parallel  $\beta$ -sheets oriented approximately perpendicular to each other and two interdigitating loops connecting the two  $\beta$ -sheets and the

23/04/2024

Aquí podemos observar un dominio de la proteína fibronectina tipo 2. Podemos leer información sobre la proteína y una representación del motivo. Este motivo está descrito y hace referencia a funciones concretas donde la proteína está involucrada.

## Bases de datos biológicas

### Bases de datos secundarias

KEGG



KEGG PATHWAY Database

Wiring diagrams of molecular interactions, reactions and relations

KEGG2 PATHWAY BRITÉ MODULE KO GENES COMPOUND DISEASE DRUG

Select prefix:

map

Organism:

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 regulation 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. Pathways in Disease
7. Drug Development

KEGG PATHWAY is the reference database for pathway mapping in KEGG Mapper.

#### 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 reaction 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)

23/04/2024

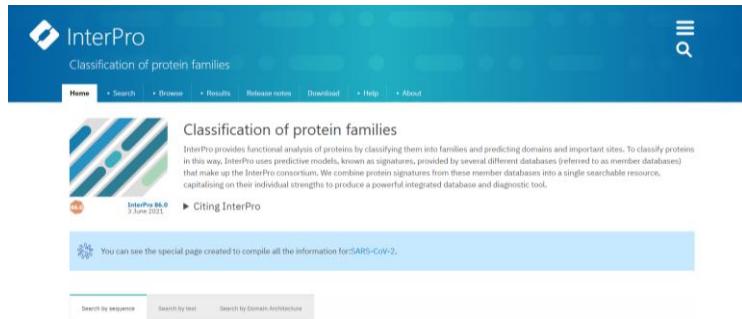
La base de datos KEGG contiene información sobre rutas metabólicas y las macromoléculas que están involucradas. De esta forma podemos predecir, en función de la información existente en la base de datos, cuales son las funciones relacionadas de nuestras proteínas.

## Bases de datos biológicas

### Bases de datos secundarias

#### Interpro

Metabase de datos que aúna otras bases de datos como Swiss-Prot, Pfam o TIGRFAMs



The screenshot shows the InterPro homepage with a blue header containing the logo and the text "Classification of protein families". Below the header is a navigation bar with links for Home, Search, Browse, Results, Release notes, Download, Help, and About. The main content area features a graphic of blue and green diagonal stripes and the text "Classification of protein families". It explains that InterPro provides functional analysis of proteins by classifying them into families and predicting domains and important sites. It mentions that InterPro uses predictive models known as signatures, provided by several different databases (referred to as member databases) that make up the InterPro consortium. The page also highlights the release of InterPro 84.0 on 3 June 2021 and a link to "Citing InterPro". A blue banner at the bottom states, "You can see the special page created to compile all the information for SARS-CoV-2." At the bottom of the page are three search options: "Search by sequence", "Search by text", and "Search by Domain Architecture".

23/04/2024

Interpro por su parte aparece como una asociación de bases de datos proteicas y bases de datos funcionales, como Swiss-Prot, Pfam y TIGRFAM, es un ejemplo de metabase de datos.

## Bases de datos biológicas

### Bases de datos secundarias

#### Interpro

The screenshot shows the InterPro database interface. At the top, there's a navigation bar with links for Home, Search, Browse, Results, Release notes, Download, Help, and About. Below the navigation is a breadcrumb trail: Home / By Entry / InterPro / IPR0000003 / Overview. The main content area displays the following information:

- Short name:** Retinoid-X\_rept/HNF4
- Entry ID:** IPR0000003
- Contributing Member Database Entry:** PRINTER (PR00545)
- Family relationships:**
  - Nuclear hormone receptor (IPR001722)
  - Retinoid X receptor/HNF4 (IPR0000003)
- Description:** Steroid or nuclear hormone receptors (NRs) constitute an important superfamily of transcription regulators that are involved in widely diverse physiological functions, including control of embryonic development, cell differentiation and homeostasis. Members of the superfamily include the steroid hormone receptors and receptors for thyroid hormone, retinoids & retinoid-like molecules (e.g. all-trans-retinoic acid and a variety of other retinoids). These molecules function as ligands for the receptor.

23/04/2024

Aquí tenemos un ejemplo del receptor X retinoide.

## Bases de datos biológicas

### Bases de datos secundarias

#### **PhenomicDB**

Base de datos que contiene datos de humano y organismos modelos

The screenshot shows the Database Commons interface. At the top, there is a search bar with placeholder text "Search by name, category, country, data type, etc." and a "Search" button. Below the search bar, there are navigation links: Home, Search, Browse, Metrics, Curations, Help, Disclaimer, and Submit. A "Sign In" link is also present. The main content area is titled "Database Profile" for "PhenomicDB". It contains two main sections: "General information" and "Ranking".

**General information:**

- URL: <http://phenomicdb.info/>
- Full name: A multi-species genotype-phenotype database for comparative phenomics.
- Description: PhenomicDB is a multi-species genotype-phenotype database, which shows phenotypes associated with their corresponding genes and grouped by gene orthologies across a variety of species.
- Year founded: 2005
- Last update: 2010-08-01
- Version: v1.0

**Ranking:**

Category	Value
All databases	1804
Genotype phenotype and variation	69
CITATIONS	4,312

23/04/2024

Existen las bases de datos con información sobre el fenotipo. En este ejemplo tenemos la base de datos PhenomicDB, con información de humano y organismos modelos

## Bases de datos biológicas

### Bases de datos secundarias

Estructuras moleculares

#### PDB (Protein data bank)

Determinación experimental de la cristalografía de macromoléculas. Generalmente hay proteínas, pero también hay ADN, ARN y complejos híbridos

The screenshot shows the homepage of the RCSB Protein Data Bank (PDB). At the top, there is a search bar with the placeholder "Enter search terms or PDB ID(s)". Below the search bar, there are links for "Advanced Search" and "Browse Annotations". A banner at the top right celebrates "Celebrating 50 YEARS OF Protein Data Bank" with icons for social media and a link to "Visit RCSB.org". On the left side, there is a sidebar with links: "Welcome", "Deposit", "Search", "Visualize", "Analyze", "Download", and "Learn". The main content area features a "Structural View of Biology" section with a brief description of the resource's purpose, mentioning its role in biomedicine and agriculture. It also highlights the "COVID-19 CORONAVIRUS Resources" and "Celebrating PROTEIN DATA BANK 50" sections. On the right, there is a "October Molecule of the Month" section showing a 3D molecular model of a protein complex, and a "Fifty Years of Open Access to PDB Structures" section.

23/04/2024

También existen las bases de datos de estructuras, siendo la principal PDB (Protein data bank). Esta base de datos reúne la información relacionada con estructura de proteínas, ADN o ARN mediante cristalografía de rayos X o RMN.

## Bases de datos biológicas

### Bases de datos secundarias

Estructuras moleculares

#### **PubChem**

Compuestos químicos e información sobre su actividad

The screenshot shows the PubChem homepage with a dark blue background featuring a hexagonal geometric pattern. At the top, there's a banner for the National Library of Medicine. Below it, the PubChem logo is displayed. A search bar contains the text "Explore Chemistry". Below the search bar, there's a message: "Quickly find chemical information from authoritative sources". The search bar also includes a magnifying glass icon and several input fields with placeholder text like "Try covid-19", "aspirin", "EGFR", etc. At the bottom of the page, there are several icons: a pencil, an upward arrow, a magnifying glass, and a grid. The date "23/04/2024" is visible in the bottom right corner.

También tenemos la base de datos PubChem, con información de compuestos químicos.

## Bases de datos biológicas

### Bases de datos especializadas

Base de datos de The Saccharomyces Genome Database (SGD)

SGD Saccharomyces GENOME DATABASE

Analysis Sequence Function Literature Community

1 of 26

Htr1p targeted genes spread evanescent into核 after heat shock (nuclear pore complex, red). Image courtesy of J. Chouillery and A. Sastri, Grana (A), CIB Health Sciences Center

About SGD

The Saccharomyces Genome Database (SGD) provides comprehensive integrated biological information for the budding yeast *Saccharomyces cerevisiae* along with search and analysis tools to explore these data, enabling the discovery of functional relationships between sequence and gene products in fungi and higher organisms.

Explore SGD

Meetings

Fungal Pathogen Genomics  
May 07 to May 12, 2023  
Wellcome Genome Campus, Hinxton, Cambridge, UK

CSHL Yeast Genetics & Genomics course  
July 26 to August 15, 2023 -  
Cold Spring Harbor Laboratory, New York

TAGC2024 The Allied Genetics Conference  
March 05 to March 10, 2024 -  
National Harbor | Washington DC Metro Area

New & Noteworthy

Regulation of yeast inducible promoters by SAGA and TFIID - November 15, 2022

Gene transcription – the elaborate process that our cells use to read genetic information stored in DNA – was long thought to be turned on only when certain regulatory factors traveled to specific DNA sequences. In a new study published in Genes & Development, Mittal et al., 2022, discovered that a subset of genes has their transcription [...] Read More

2D RNA structures from RNACentral - October 20, 2022

SGD has updated our RNA pages to add secondary structures provided by RNACentral and generated by RDOT. Thumbnails and links to RNACentral

Tweets from @yeastgenome

13 SGD Project Retweeted  
D. Allan Drummond Art @dadrummondart · Nov 15  
Inside a cell ↗ My cutaway model of a dividing cell with all its delicate machinery, mostly printed in white PLA (I used to print in Spilt the mitochondria...) Available at <https://www.etsy.com/listing> if you'd like your own!

23/04/2024

Existen repositorios o bases de datos especializadas en organismos modelo. Estas bases de datos reúnen información de organismos concretos como por ejemplo *Saccharomyces*. La base de datos SGD guarda toda la información relevante sobre este organismo, como genomas, genes, proteínas, fenotipos...

## Bases de datos biológicas

### Principales problemas de las bases de datos

- Secuencias con errores o con secuencias de vectores de clonación, adaptadores, etc...
- Información mal anotada, campos incompletos
- Sesgos sobre microorganismos muy estudiados, redundancia

23/04/2024

Las bases de datos por lo general tienen una serie de errores o problemas sistemáticos. Estos problemas derivan en su mayoría de la falta de criterio unificado a la hora de añadir nueva información a las mismas y el tipo de estudios que se realizan. Podemos encontrar secuencias con errores, los cuales deberán curarse manualmente al detectarse, o secuencias mal anotadas como vectores de secuenciación anotados como plásmidos de secuencias bacterianas. También podemos tener problema en la información que se muestran en las fichas para cada organismo o proteína, ya que a pesar de tener la información más relevante, por ser obligatoria, podemos estar perdiendo la información de, por ejemplo, la fuente desde donde se aisló la muestra. Por último podemos observar problemas en la sobrerrepresentación de diversos organismos, como los organismos modelos o bacterias (como E.coli), lo que puede generar un problema en la caracterización de nuevas secuencias en las bases de datos.

## Bases de datos y herramientas bioinformáticas

Google search results for "metschnikowia pulcherrima genome". The results include:

- Artículos académicos para metschnikowia pulcherrima genome**
- [yeast strain from the \*Metschnikowia pulcherrima\* ... - Venkatesh - Citado por 13](https://www.ncbi.nlm.nih.gov/pmc/)
- [... of the bimolar yeast \*Metschnikowia pulcherrima\* - Gore-Lloyd - Citado por 48](https://www.ncbi.nlm.nih.gov/pmc/)
- [... of a chitinase gene in \*Metschnikowia pulcherrima\* ... - Saravananumar - Citado por 69](https://www.ncbi.nlm.nih.gov/pmc/)
- [Aproximadamente 19.600 resultados \(0,37 segundos\)](https://www.ncbi.nlm.nih.gov/pmc/)
- Draft Genome Sequence of a Highly Heterozygous Yeast ...**
- [de A Venkatesh 2018 Citado por 13 - \*Metschnikowia\* strain UCD127 was isolated from soil in Ireland and sequenced. It is a highly heterozygous diploid strain with 385.000 single ...](https://www.ncbi.nlm.nih.gov/pmc/)
- [https://www.ncbi.nlm.nih.gov/worm/ ... Traducir esta página](https://www.ncbi.nlm.nih.gov/worm/)
- Taxonomy browser (*Metschnikowia pulcherrima*) - NCBI**
- [Entrar records. Database name: Direct links, Links from type: Nucleotide, 437 - 17. Protein, 15. - Genom ... 1, ~ Popset, 60, ~ GEO Datasets, 7, ~](https://www.ncbi.nlm.nih.gov/taxonomy/)
- [https://mycocosm.jgi.doe.gov/M... Traducir esta página](https://mycocosm.jgi.doe.gov/M...)
- Info - *Metschnikowia pulcherrima* NRRL Y-7111 v1.0**
- [Status : Genome Assembly size \(Mbp\), 21.18 , Sequencing read coverage depth, 643.11x; # of contigs, 140 ; # of scaffolds, 140.](https://mycocosm.jgi.doe.gov/M...)
- [Genome Assembly size \(Mbp\): 21.18](https://mycocosm.jgi.doe.gov/M...)
- [Metschnikowia pulcherrima NRRL Y-7111 v1.1 - Home](https://mycocosm.jgi.doe.gov/M...)
- [Metschnikowia pulcherrima](https://mycocosm.jgi.doe.gov/M...) is best known for being the first yeast described as ... the first known secondary metabolic gene cluster in budding yeasts?

23/04/2024

Es importante conocer como se estructura una base de datos y donde esta la información, pero a la hora de buscar *algo* sigue siendo más sencillo pasar por un buscador.

Por ejemplo, digamos que me estoy embarcando en un nuevo proyecto que va a implicar la secuenciación de una cepa de levadura de *Metschnikowia pulcherrima*. Ciertamente podría ir a la base de datos de genomas del EBI o del NCBI y buscar por organismo, y probablemente sería muy recomendable hacerlo así... pero en realidad, solemos empezar por escribir en el buscador:

*Metschnikowia pulcherrima genome*

La razón es que muy probablemente vaya a localizar esos genomas depositados en las principales bases de datos pero... puede que exista algún repositorio más pequeño, alguna universidad o proyecto que tenga los datos públicos o disponibles.

En este caso, mycocosm, e indagando vemos que esta actualizado con un genoma con PacBio en el 2020.



## 4.2 Otros recursos online

## Herramientas online

**Software para ensamblaje y edición de fragmentos de ADN**

**gatk**

User Guide   Tool Index   Blog   Forum   Events   Download GATK4   Sign in

### Genome Analysis Toolkit

Variant Discovery in High-Throughput Sequencing Data

Developed in the Data Sciences Platform at the Broad Institute, the toolkit offers a wide variety of tools with a primary focus on variant discovery and genotyping. Its powerful processing engine and high-performance computing features make it capable of taking on projects of any size. [Learn more](#).

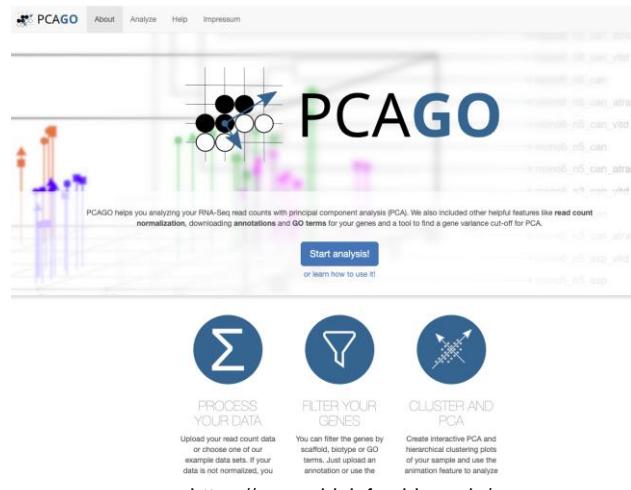
Find answers to your questions. Stay up to date on the latest topics. Ask questions and help others.

<b>Getting Started</b> Best practices, tutorials, and other info to get you started	<b>Technical Documentation</b> Algorithms, glossary, and other detailed resources	<b>Announcements</b> Blog and events
<b>Tool Index</b> Purpose, usage and options for each tool	<b>Forum</b> Ask our team for help and report issues	<b>GATK Showcase on Terra</b> Check out these fully configured workspaces
<b>Download latest version of GATK</b> The GATK package download includes all released GATK tools. Please see Getting Started for a quick start primer on using GATK, and see the Version History for release notes and highlights of differences between versions.		
		<b>Run on Cloud</b>
		<b>Run on HPC</b>

23/04/2024

## Herramientas online

### PCA transcriptómica



<https://pcago.bioinf.uni-jena.de/>

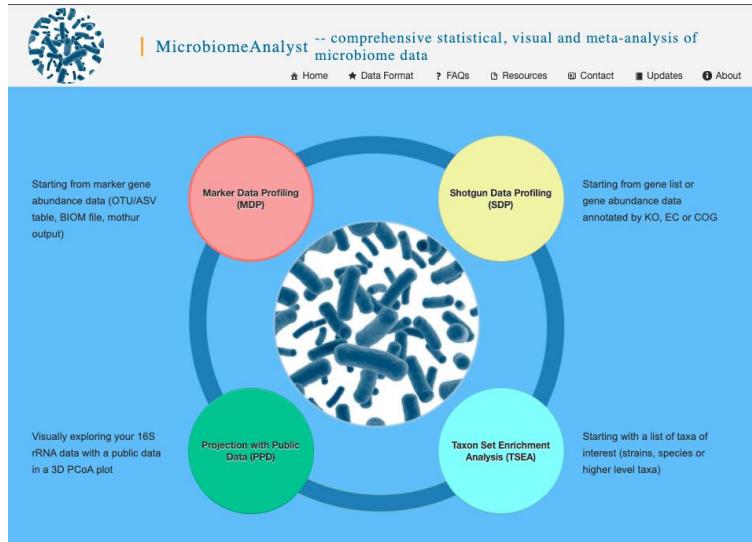
23/04/2024

## Herramientas online

The screenshot shows the NCBI BLAST homepage. At the top, there are links for NIH, U.S. National Library of Medicine, NCBI National Center for Biotechnology Information, and a sign-in area for tronchoni, My NCBI, and Sign Out. Below this, the main title "BLAST®" is displayed, followed by "Basic Local Alignment Search Tool". A brief description of BLAST is provided, mentioning it finds regions of similarity between biological sequences by comparing nucleotide or protein sequences to sequence databases and calculating statistical significance. There is a "Learn more" link and a "NEWS" sidebar with a message about a new version of IgBLAST. The date is listed as Thu, 14 Jan 2021 12:00:00 EST, and there is a link to "More BLAST news...". The "Web BLAST" section features three search tools: "Nucleotide BLAST" (nucleotide > nucleotide), "blastx" (translated nucleotide > protein), and "tblastn" (protein > translated nucleotide). The "Protein BLAST" tool is also shown, which compares protein to protein. Below these, the "BLAST Genomes" section includes a search bar for organism names and buttons for Human, Mouse, Rat, and Microbes, with a "Search" button. The footer of the page shows the date 23/04/2024.

## Herramientas online

### Software para metataxonómica



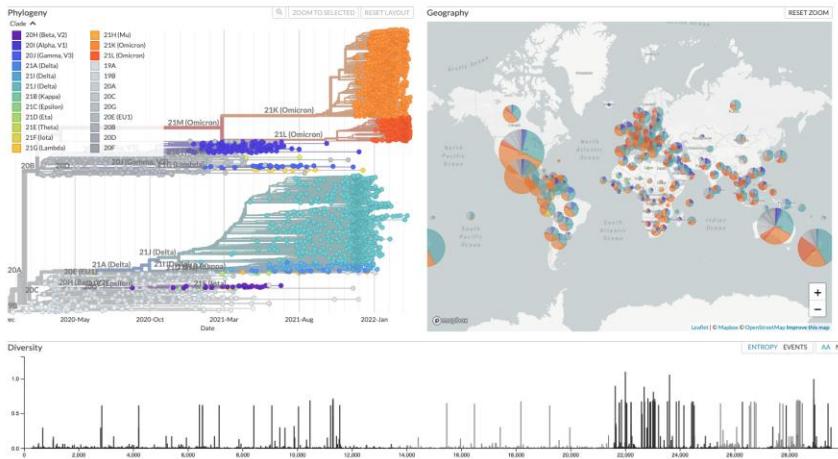
## Herramientas online

### Epidemiología Genómica de SARS-CoV-2 NextStrain

Genomic epidemiology of SARS-CoV-2 with global subsampling

Built with nextstrain.now. Maintained by the Nextstrain team. Enabled by data from GISAID.

Showing 3002 of 3002 genomes sampled between Dec 2019 and Mar 2022.



23/04/2024

## Existen otro tipo de recursos online

<a href="https://www.embl.org/">https://www.embl.org/</a>	EMBL
<a href="http://www.ebi.ac.uk/">http://www.ebi.ac.uk/</a>	EBI
<a href="https://www.ncbi.nlm.nih.gov/">https://www.ncbi.nlm.nih.gov/</a>	
<a href="https://www.ddbj.nig.ac.jp/index-e.html">https://www.ddbj.nig.ac.jp/index-e.html</a>	DDBJ
<a href="https://www.ebi.ac.uk/training/online/course-list">https://www.ebi.ac.uk/training/online/course-list</a>	ebi course
<a href="https://www.ncbi.nlm.nih.gov/books/NBK21101/">https://www.ncbi.nlm.nih.gov/books/NBK21101/</a>	ncbi books
<a href="https://www.iscb.org/cms-addon/online_courses/index.php">https://www.iscb.org/cms-addon/online_courses/index.php</a>	Iscb course
<a href="http://biochem218.stanford.edu/index.html">http://biochem218.stanford.edu/index.html</a>	Standford course
<a href="https://www.huber.embl.de/msmb/introduction.html">https://www.huber.embl.de/msmb/introduction.html</a>	Modern statistics book
<a href="http://seqanswers.com/">http://seqanswers.com/</a>	Bioinformatics forums
<a href="https://www.biostars.org/">https://www.biostars.org/</a>	& comunites

23/04/2024

Es importante conocer los dos grandes foros de bioinformática, seqanswers y biostars. Generalmente cuando tenemos un problema de instalación o hay alguna herramienta que no funciona adecuadamente, ya le ha ocurrido a alguien y es aconsejable revisar estos foros para poder encontrar la solución.

También podemos encontrar interesantes discusiones sobre la utilización de los distintos parámetros de distintas herramientas y conocer nuevas formas de realizar distintos procesos.

# ¡Gracias!



**Universidad**  
Internacional  
de Valencia

[universidadviu.com](http://universidadviu.com)

De:  
 Planeta Formación y Universidades