

# Acceso a BDs Biológicas: Entrez System

Luis E. Garreta U

[luis.garreta@javerianacali.edu.co](mailto:luis.garreta@javerianacali.edu.co)

## **Curso de Bioinformática**

Pontificia Universidad Javeriana – Cali

Facultad de Ingeniería - Carrera de Biología

13 de agosto de 2018

# Acceso a las BDs Biológicas

- ▶ Búsqueda y recuperación textual
- ▶ Tres grandes herramientas integradas para búsquedas y recuperación:
  - ▶ **Entrez**: desarrollado en el NCBI
  - ▶ **SRS**: desarrollada en el EBI
  - ▶ **DBGET**: desarrollada como parte del GenomeNet de Japón
- ▶ Además de las búsquedas, se ofrece enlaces a información relacionada
- ▶ Las tres herramientas difieren ligeramente:
  - ▶ En las BD que consultan,
  - ▶ En los enlaces a la información relacionada
- ▶ Nos enfocaremos en unos de los más usados: **ENTREZ**

# GenBank

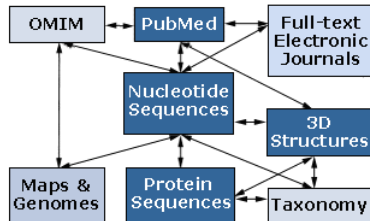
- ▶ Colección anotada de secuencias del NCBI
- ▶ Las secuencias pueden ser de diversos tipos y alcances:
  - ▶ Secuencia de ADN, ARN, aminoácidos
  - ▶ Secuencia de transcrito, gen, cromosoma, genoma
  - ▶ Secuencia de mutación (SNP)
  - ▶ ... hasta 40 BDs distintas
- ▶ **PubMed**: complementa a GenBank con una colección anotada de artículos científicos
- ▶ **Entrez** es la herramienta del NCBI para facilitar las búsquedas

# Entrez (<https://www.ncbi.nlm.nih.gov/>)

Brinda una herramienta integrada para acceder a las BDs del NCBI

- ▶ Secuencias de nucleótidos
- ▶ Secuencias de proteínas
- ▶ Estructuras macromoleculares
- ▶ Genomas y mapas genéticos
- ▶ Literatura: MEDLINE
- ▶ Mas de 38 DBs

...



# Ejemplo: *Búsqueda estructura 3D de la Villin*

NCBI Resources How To

## Search NCBI databases

Results found in 31 databases for "villin"

Literature			Genes		
<b>Books</b>	29	books and reports	<b>EST</b>	438	expressed sequence tag sequences
<b>MeSH</b>	7	ontology used for PubMed indexing	<b>Gene</b>	4,380	collected information about gene products
<b>NLM Catalog</b>	1	books, journals and more in the NLM Collections	<b>GEO DataSets</b>	398	functional genomics studies
<b>PubMed</b>	1,267	scientific & medical abstracts/citations	<b>GEO Profiles</b>	30,110	gene expression and molecular analysis
<b>PubMed Central</b>	4,090	full-text journal articles	<b>HomoloGene</b>	12	homologous gene sets for selected species
<b>Health</b>			<b>PopSet</b>	13	sequence sets from phylogenetic studies
<b>ClinVar</b>	22	human variations of clinical significance	<b>UniGene</b>	138	clusters of expressed transcripts
<b>dbGaP</b>	0	genotype/phenotype interaction studies	<b>Proteins</b>		
<b>GTR</b>	2	genetic testing registry	<b>Conserved Domains</b>	17	conserved protein domains
<b>MedGen</b>	0	medical genetics literature and links	<b>Protein</b>	17,014	protein sequences
<b>OMIM</b>	14	online mendelian inheritance in man	<b>Protein Clusters</b>	16	sequence similarity-based protein clusters
<b>PubMed Health</b>	0	clinical effectiveness, disease and drug reports	<b>Structure</b>	116	experimentally-determined biomolecular structures
<b>Genomes</b>			<b>Chemicals</b>		
<b>Assembly</b>	0	genome assembly information	<b>BioSystems</b>	866	molecular pathways with links to chemical reactions
<b>BioCollections</b>	0	museum, herbaria, and other biorepository collections			
<b>BioProject</b>	27	biological projects providing data to NCBI			

# Ejemplo: *Lista de Resultados por estructura*


NCBI Resources How To


Structure  [Create alert](#) [Advanced](#)


Summary 20 per page Sort by Default order Send to:

**Search results**

Items: 1 to 20 of 116 << First < Prev Page 1 of 6 Next > Last >>

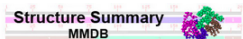
- ☐ 

[Crystal structure of calcium-free human gelsolin\[Actin binding Protein\]](#)  
 Taxonomy: Homo sapiens  
 Proteins: 1 modified: 2017-11-20  
 MMDB ID: 77149 PDB ID: 3FFN  
[View in ICn3D](#) [Similar Structures](#) [PubMed](#) [Proteins](#) [Conserved Domains](#)
- ☐ 

[ATP bound gelsolin\[CONTRACTILE PROTEIN, STRUCTURAL PROTEIN\]](#)  
 Taxonomy: Equus caballus  
 Proteins: 2 Chemicals: 2 modified: 2018-01-08  
 MMDB ID: 38727 PDB ID: 2FGH  
[View in ICn3D](#) [Similar Structures](#) [PubMed](#) [Proteins](#) [Conserved Domains](#) [PubChem Compound](#)
- ☐ 

[The Crystal Structure Of Calcium-Free Equine Plasma Gelsolin\[Contractile Protein\]](#)  
 Taxonomy: Equus caballus  
 Proteins: 2 modified: 2012-11-27  
 MMDB ID: 11164 PDB ID: 1D0N  
[View in ICn3D](#) [Similar Structures](#) [PubMed](#) [Proteins](#) [Conserved Domains](#)

# Ejemplo: Selección de un registro particular


 Enter PDB ID or MMDB ID 

**5VNT: Solution NMR Structure of the C-terminal Headpiece Domain of Villin 4 From *A.thallana*, the First Non-vertebrate Headpiece Structure**

Citation: ⓘ

No citation available for this structure.

**PDB ID:** 5VNT  ⓘ

**MMDB ID:** 150755 ⓘ

**PDB Deposition Date:** 2017/5/1 ⓘ

**Updated in MMDB:** 2017/05 ⓘ

**Experimental Method:** solution nmr ⓘ

**Source Organism:** *Arabidopsis t...* ⓘ

**Similar Structures:**  ⓘ

ⓘ

Biological Unit for 5VNT: monomeric; determined by author ⓘ

Molecular Graphic ?

Interactions ?

☐ Protein ☐ Nucleotide ☐ Chemical

Download Structure Data ?

Format: ASN.1 (Cn3D) v

Data Set: Single 3D structure Cn3D

# Formas de Realizar las Búsquedas

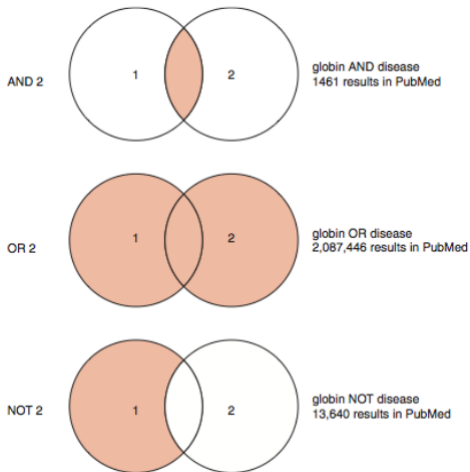
Las búsquedas en las BDs pueden ser hechas de varias formas:

- ▶ Por tópico:
  - ▶ Ingresar palabra o "frase
  - ▶ Utilizar operadores booleanos: AND, OR, NOT
  - ▶ Utilizar comodines: cyanobact\* (cyanobacterium and cyanobacterial)
- ▶ Por identificadores únicos:
  - ▶ Secuencias de ADN y proteínas se les asigna un único número de acceso cuando se ingresan a la BD (e.g
  - ▶ CAA50375.1 (Villin, partial [Homo sapiens])
- ▶ Por autor:
  - ▶ Apellido + Iniciales
  - ▶ e.g: mann nh



# Búsqueda avanzada en Entrez

- ▶ Uso de manuales
  - ▶ Fundamental para cualquier herramienta bioinformática
- ▶ Operadores booleanos: AND, OR, NOT
  - ▶ horse OR horses
- ▶ Filtros: corchetes tras el nombre:
  - ▶ horse[Organism]
  - ▶ BRCA1[Gene Name]
- ▶ Comillas: para coincidencia exacta si hay más de una palabra
  - ▶ "Equus caballus"



J. Pevsner, *Bioinformatics and Functional Genomics*. 2009