

# Biological Databases

- Different types of databases
- Different levels of biological databases

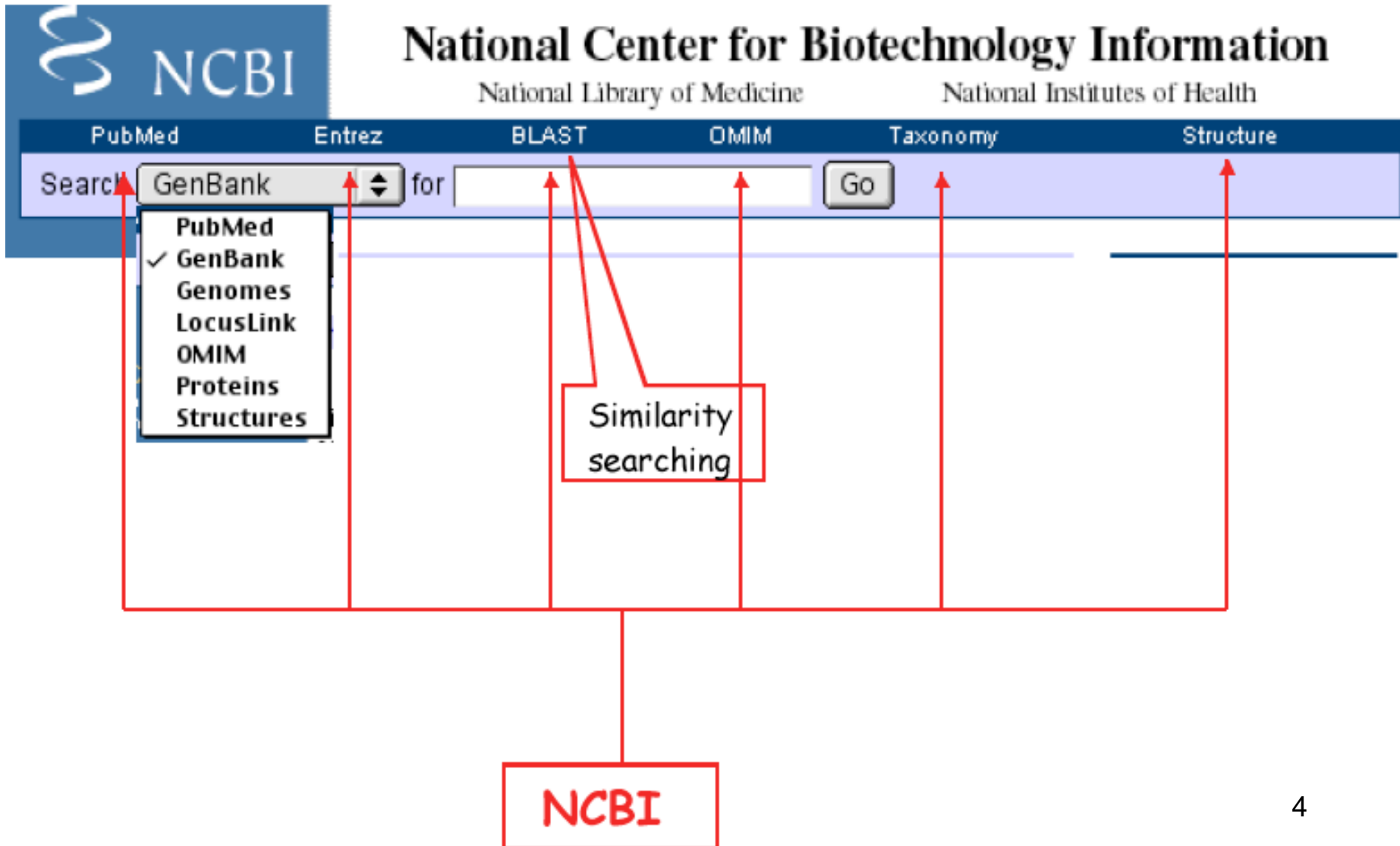
# Types of Databases

- Hierarchical
- Object-oriented
- Relational

# Levels of Biological Databases

- Primary
  - Raw data: EMBL, GenBank, DDBJ
- Secondary
  - Human-curated (knowledge base):  
SwissProt, UniProt, RefSeq, SEED, etc...
- Tertiary/ Specialized
  - NMPDR, FLYBase, WormBase, etc...

# NCBI Databases/ GenBank



Search  for

**NCBI Homepage**

## SITE MAP

### About NCBI

general and contact information

### GenBank

sequence submission support and software

### Molecular databases

sequences, structures and taxonomy

### Literature databases

PubMed, OMIM and PubMed

Central **NEW**

### Genomic biology

the human genome, whole genomes and related resources

### Tools

for data mining

### Research at NCBI

## What does NCBI do?

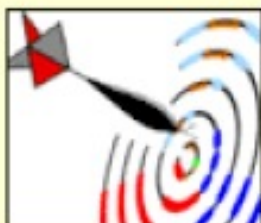
Established in 1988 as a national resource for molecular biology information, NCBI creates public databases, conducts research in computational biology, develops software tools for analyzing genome data, and disseminates biomedical information - all for the better understanding of molecular processes affecting human health and disease.

## Draft Human Genome

Explore [human genome resources](#) or browse the human genome sequence using the [Map Viewer](#).

## DART: A new tool

[Archives](#)



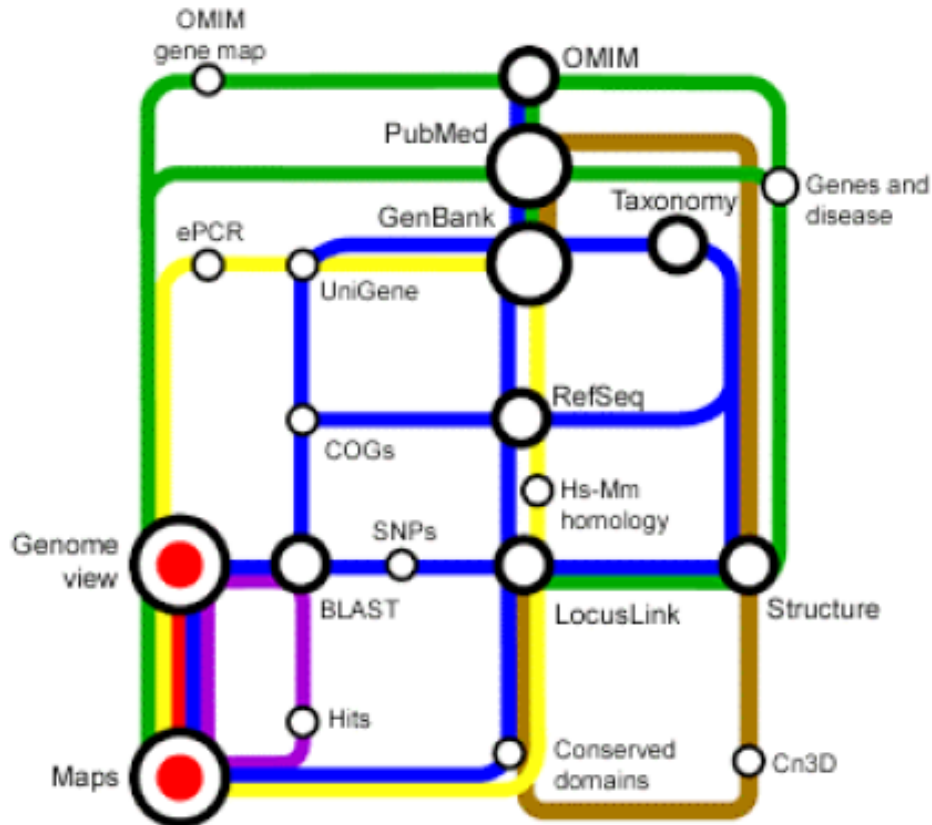
Want to locate protein neighbors by domain architecture? Learn about NCBI's new Domain Architecture Retrieval Tool...

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

## Hot Spots

- ▶ Cancer genome anatomy project
- ▶ Clusters of orthologous groups
- ▶ Coffee Break
- ▶ Electronic PCR
- ▶ Gene expression omnibus
- ▶ Genes and disease
- ▶ Human genome resources
- ▶ Human map viewer
- ▶ Human/mouse homology maps

# NCBI Databases



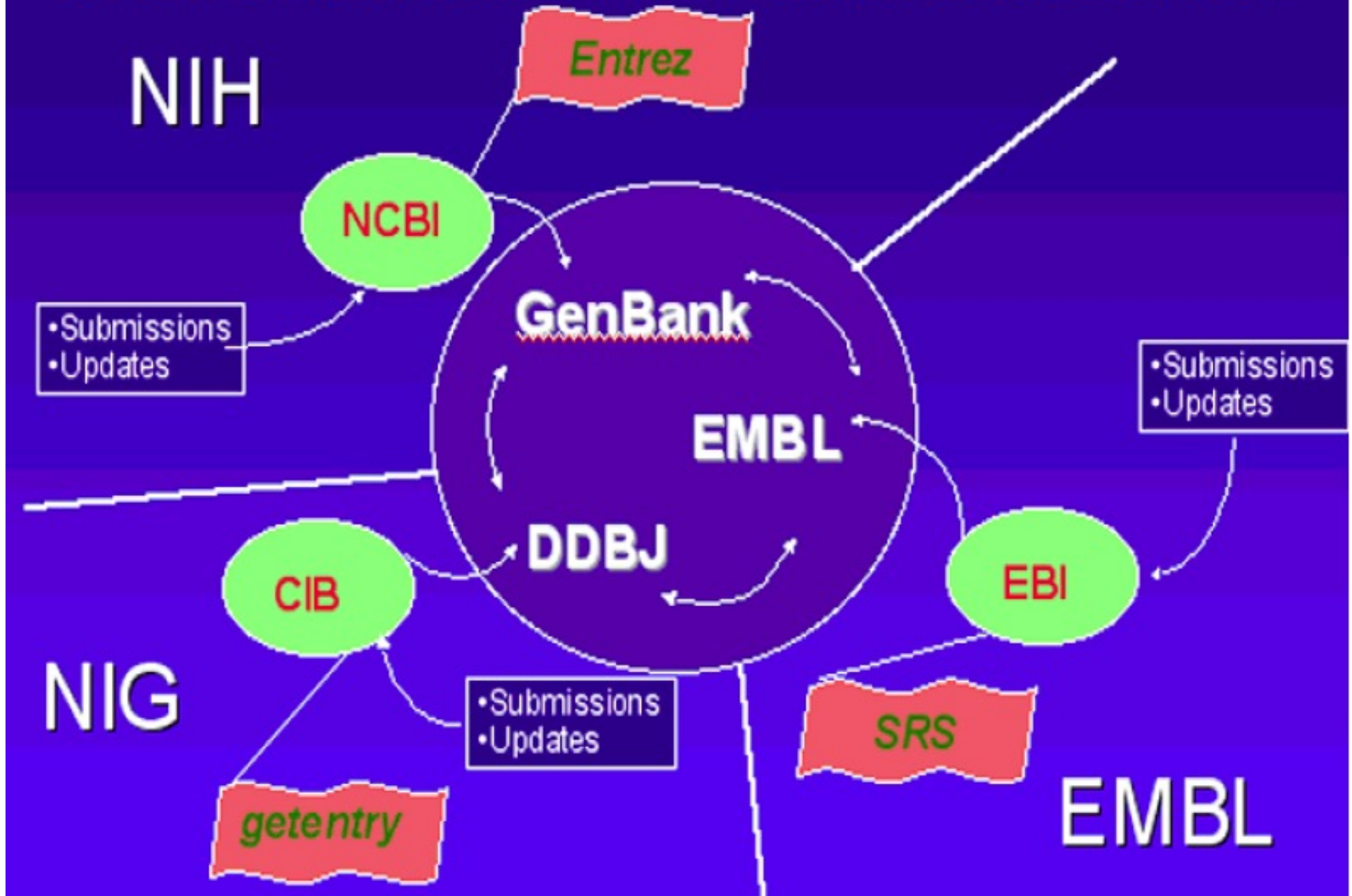
Key	
<span style="color: green;">■</span> Literature	<span style="color: purple;">■</span> BLAST
<span style="color: blue;">■</span> Sequence	<span style="color: red;">■</span> Genome View
<span style="color: yellow;">■</span> Mapping	<span style="color: brown;">■</span> Structure

[Back to showcase](#)

<http://www.ncbi.nlm.nih.gov/Tour/tour.html>

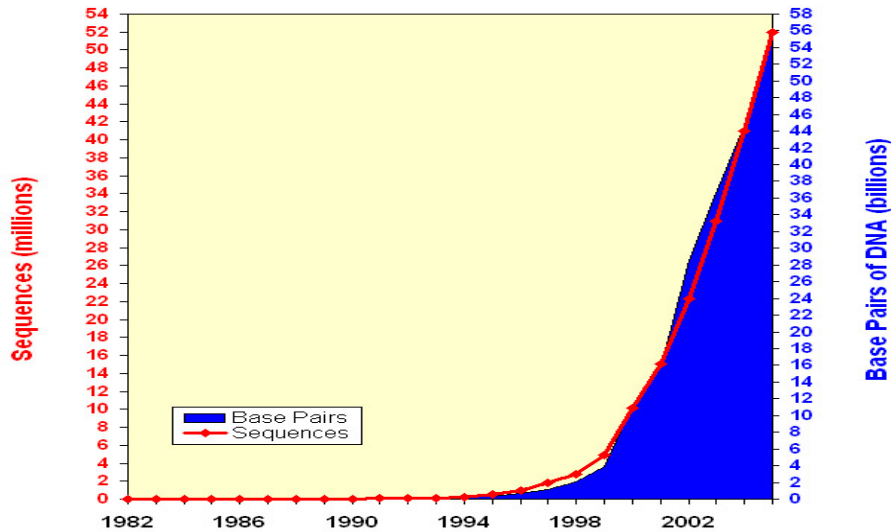
# International Databases

GenBank is part of international collaboration of NCBI, DDBJ, EMBL.

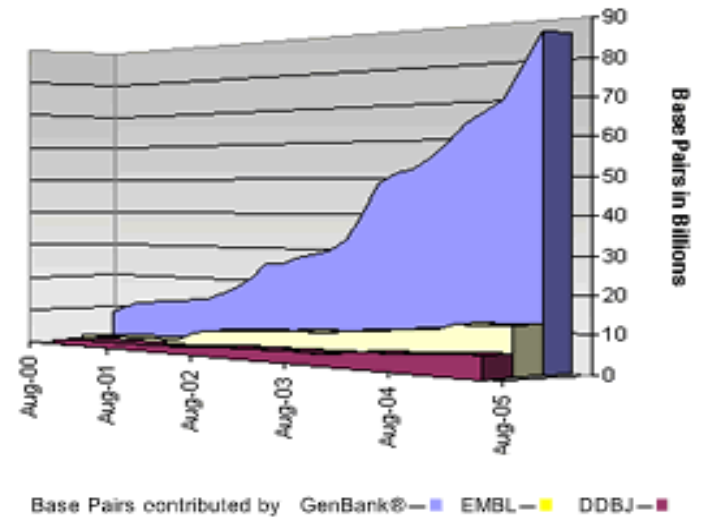


# The Data Deluge: Explosion of DBs

**Growth of GenBank  
(1982 - 2005)**



**Growth of the  
International Nucleotide Sequence Database Collaboration**

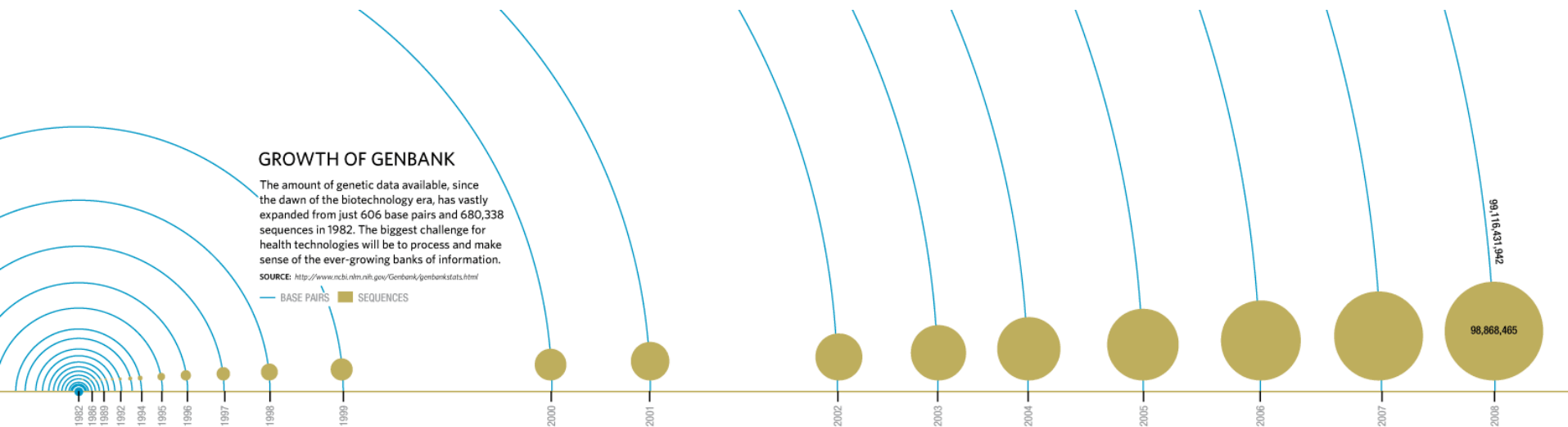


**GROWTH OF GENBANK**

The amount of genetic data available, since the dawn of the biotechnology era, has vastly expanded from just 606 base pairs and 680,338 sequences in 1982. The biggest challenge for health technologies will be to process and make sense of the ever-growing banks of information.

SOURCE: <http://www.ncbi.nlm.nih.gov/Genbank/genbankstats.html>

— BASE PAIRS ■ SEQUENCES





# A Final Word (One More Time)

On using computational tool to deal with the biological data, we have to always remember that the data are more important than the tools because these data are all about **life**: ours and our neighbors on this planet!

But we also have to remember that whatever we find in databases is as good as the laboratories producing it!  
They are not infallible!