

Genome information resource

Source textbook

“Introduction to Bioinformatics”

Chapter 4

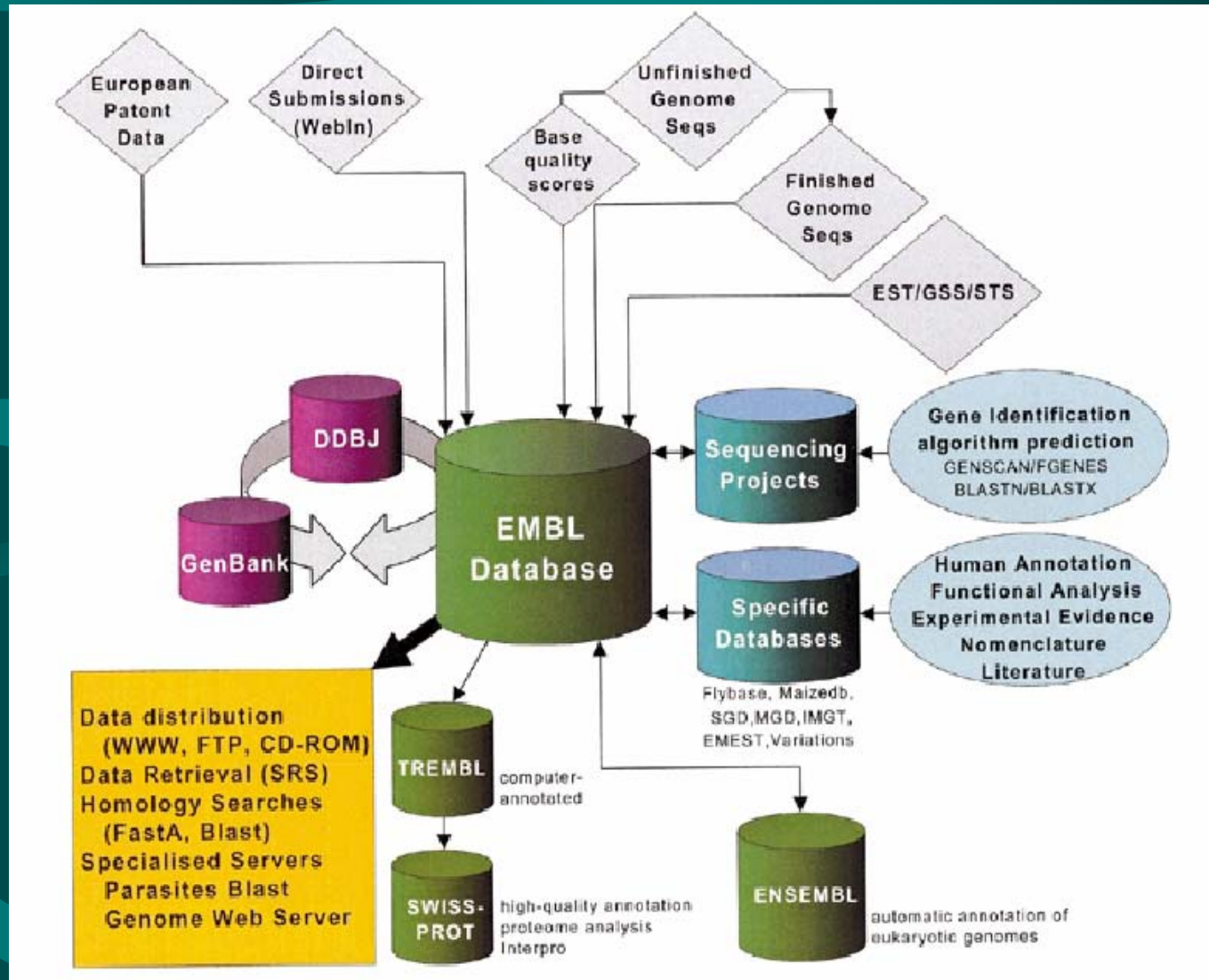
DNA Sequence database

- EMBL <http://www.ebi.ac.uk/embl/>
- DDBJ <http://www.ddbj.nig.ac.jp/>
- GenBank <http://www.ncbi.nlm.nih.gov/Genbank>
- dbEST <http://www.ncbi.nlm.nih.gov/dbEST/>
- GSDB <http://www.ncgr.org/research/>

EMBL

- ❖ From the EBI(European Bioinformatics Institute)
- ❖ The database is produced in collaboration with DDBJ and GenBank
- ❖ The rate of growth of DNA databases has been following an exponential trend. In January 1998, EMBL contained more than a million entries, representing more 15500 species.
- ❖ Can be retrieved using the SRS.
- ❖ Can be searched with query sequences via the EBI's Web interfaces to the BLAST and FastA programs.

EMBL



DDBJ

- ❖ DNA Data Bank of Japan, which began in 1986 as a collaboration with EMBL and GenBank.
- ❖ May be submitted to it from all corners of the world by means of a Web-based data-submission tool.
- ❖ Also used to provide standard search tools, such as FastA and BLAST.

DDBJ

Search

GO

To Japanese Page

- DDBJ Top Page
- What's NEW
- Welcome to DDBJ
- DDBJing
- SAKURA
- Mass Data
- Update/Correction
- getentry
- SRS
- Homology Search
- CIB Homepage

SITE MAP

DNA Data Bank of Japan

[MEXT, Japan](#) [National Institute of Genetics](#)

[Center for Information Biology and DNA Data Bank of Japan](#)

- Jump -

WHAT'S NEW

- Hot Topics: DDBJ rel. 45 Completed, including 11,434,113 entries, 12,207,092,905 bases
 - supernig periodic maintenance: 9 a.m. May 7th-9 a.m. May 10th
 - [DDBJ will stop accepting submissions by 'Authorin' and 'Submission Form'](#)

Welcome (to DDBJ)

- [about DDBJ/CIB](#)
- [DDBJing](#) (Japanese only)
- [DDBJ Statistics](#)
- [DDBJ Publications Online](#)
- [Addresses Related to DDBJ Activities](#)

Data Submission

- [Submission via SAKURA](#)

Database Search

- [getentry](#)
- [SRS](#)
- [Homology Search](#) (FASTA/BLAST/SSEARCH)
- [TXSearch](#)

Data Analysis

- [ClustalW](#) (multiple alignment and tree making)
- [Use of Super Computer minerva\(VPP5000\)](#)

GenBank

- ❖ From the National Center for Biotechnology Information (NCBI).
- ❖ The increasing size of the database have made it convenient to split GenBank into smaller , discrete divisions (17 to date) *See next Page
- ❖ Also used to provide standard search tools, such as FastA and BLAST.

GenBank

Division	Sequence subset	Division	Sequence subset
PRI	Primate	PHG	Bacteriophage 抗菌素
ROD 嚙齒動物	Rodent	SYN	Synthetic 人造的
MAM 哺乳類	Other mammalian	UNA	Unannotated 未註解的
VRT 脊椎動物	Other vertebrate	EST	Expressed Sequence Tags
INV	Invertebrate	PAT	Patent
PLN 黴菌 藻類	Plant,fungal,algal	STS	Sequence Tagged Sites
BCT 細菌	Bacterial	GSS	Genome Survey Sequences
RNA	Structural RNA	HTG	High Throughput Genomic Sequence
VRL 濾過性病毒	Viral		

The three-letter codes for each of the 17 divisions of GenBank.

網址(D) http://www.ncbi.nlm.nih.gov/Genbank/

 **NCBI Submit to GenBank®**

PubMed Entrez BLAST OMIM Taxonomy Structure

Search for

GenBank
Genomes
LocusLink
PubMed
OMIM
Proteins
Structures

NCBI
SITE MAP

Accession numbers
to cite in your manuscript

BankIt
Introductory information

Sequin
Introductory information

Special submissions
genomes, batch sequences, alignments

Sending data to GenBank

Submitting Sequence Data to GenBank

The most important source of new data for GenBank is direct submissions from scientists. GenBank depends on its contributors to help keep the database as comprehensive, current, and accurate as possible. NCBI provides timely and accurate processing and biological review of new entries and updates to existing entries, and is ready to assist authors who have new data to submit.

NOTE: The 'Authorin' submission tool and the E-mail submission form were phased out on December 31, 1998, and submissions made with those tools are no longer accepted as of that date. Instead, please use the improved submission tools. [BankIt](#) and [Sequin](#).

► **Submit now!!**

[Sequin](#)
Stand-alone sequence submission tool

[BankIt](#)
For quick and simple submissions

[VecScreen](#)
Vector contamination screening tool

► **GenBank**

[GenBank](#)
overview of the database

[Search](#)

dbEST

- Maintains its own format and identification number system.
- Contained more than 1.6 million ESTs, including more than 1 million sequences from Homo sapiens, and 300000 from Mus musculus and Mus domesticus.

EST:(Expressed Sequence Tag) A partial sequence of a clone, randomly selected from a cDNA library and used to identify genes expressed in a particular tissue. ESTs are used extensively in projects to map the human genome.



Expressed Sequence Tags database

PubMed

Entrez

BLAST

OMIM

Taxonomy

Structure

Search **EST** for

Go

Clear

modified during the last 10 Years

NCBI

SITE MAP

Human
Genome
Resources

UniGene

LocusLink

NCI_CGAP

NEW 07/15/2000 EST search method switched from IRX to Entrez. Use search box above instead of old search page.

► What is EST?

dbEST ([Nature Genetics 4:332-3;1993](#)) is a division of [GenBank](#) that contains sequence data and other information on "single-pass" cDNA sequences, or [Expressed Sequence Tags](#), from a number of organisms. A brief account of the history of human ESTs in GenBank is available ([Trends Biochem. Sci. 20:295-6;1995](#)). Also, consult the special "Genome Directory" issue of Nature (vol. 377, issue 6547S, 28 September 1995).

► Other ways to access dbEST

GSDB

- GSDB(Genome Sequence DataBase)
- Produced by the National Center for Genome Resources at Santa Fe, New Mexico.
- The format of GSDB entries is consistent with GenBANK.
- The principal difference between the two formats is the inclusion of the **GSDBID** keyword, which implements the tracking mechanism for GSDB accessions.

NATIONAL CENTER FOR GENOME RESOURCES

What's New

as of April 23,
2001

Dayan B.
Goodenowe of
Phenomenome
Discoveries will
present
"Comprehensive
Metabolome
Expression
Analysis" at
noon Monday,
May 7. Learn
about NCGR
[events.](#)

[News Releases](#)

[Site Disclaimer](#)



Better nutrition,
cleaner environment,
improved health
through bioinformatics



About Bioinformatics

Site Map
Contact
Search

ABOUT NCGR

[Programs](#)
[Leadership](#)
[Staff](#)
[Collaborations](#)
[Funding](#)
[News](#)
[Events](#)

SOFTWARE

[PathDB™](#)
[ISYS™](#)
[MAR- Finder](#)
[GeneX](#)
[PGI](#)
[MGI](#)
[PGC](#)

RESEARCH

[TAIR](#)
[Medicago](#)
[Phytophthora](#)
[Gene Expression](#)
[Pathways](#)
[Integration](#)

Specialized genomic resources

- The purpose of the specialized resources is to bring a focus
 - To species-specific genomics
 - To particular sequencing techniques.
- SGD(Saccharomyces Genome Database)
- UniGene
- TDB(TIGR database)

Summary

- The principal nucleic acid sequence databases are GenBank, EMBL and DDBJ, which each collect a portion of the total sequence data reported world-wide, and exchange new and updated entries on a daily basis.

Summary(Cont.)

- During 1992-1997, the level of EST and STS data within GenBank grew 10-fold. Nevertheless, the overall sequence information contributed by such partial data was still less than that of the higher quality sequences in the other major division.

Summary(Cont.)

- GenBank, which is produced at the NCBI, is split into smaller, discrete division. This facilitates fast, specific searches by restricting queries to particular database subset.
- In addition to the comprehensive DNA sequence databases, there is a variety of more specialised genomic resources. These so-called boutique databases bring a focus to species-specific genomics and to particular sequencing techniques.

Summary(Cont.)

- The scope of genomic resources available on the Internet is immense and has had an enormous impact on the ability of scientists to present and disseminate research results.