

Subject: Genetics and plant breeding

CBCS Pattern Practical Examination

M.Sc (Ag.) II Semester

Course: Computers and Bioinformatics

PART II

Unit -10 Agricultural biotechnology

Unit-11 Access to literature

GenBank:

GenBank[®] is the NIH genetic sequence database, an annotated collection of all publicly available DNA sequences. GenBank is part of the International Nucleotide Sequence Database Collaboration, which comprises the DNA DataBank of Japan (DDBJ), the European Nucleotide Archive (ENA), and GenBank at NCBI.

A GenBank release occurs every two months and is available from the ftp site. The release notes for the current version of GenBank provide detailed information about the release and notifications of upcoming changes to GenBank.

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 Enters Nucleotide.

Search and align GenBank sequences to a query sequence using BLAST (Basic Local Alignment Search Tool). BLAST searches CoreNucleotide, dbEST, and dbGSS independently; see BLAST info for more information about the numerous BLAST databases.

Search, link, and download sequences programatically 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-asn1> 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, some submitters may claim patent, copyright, or other intellectual property rights in all or a portion of the data they have submitted.

NCBI is not in a position to assess the validity of such claims, and therefore cannot provide comment or unrestricted permission concerning the use, copying, or distribution of the information contained in GenBank.

Some authors are concerned that the appearance of their data in GenBank prior to publication will compromise their work.

GenBank will, upon request, withhold release of new submissions for a specified period of time. However, if the accession number or sequence data appears in print or online prior to the specified date, your sequence will be released.

In order to prevent the delay in the appearance of published sequence data, we urge authors to inform us of the appearance of the published data. As soon as it is available, please send the full publication data--all authors, title, journal, volume, pages and date--to the following address: update@ncbi.nlm.nih.gov

Privacy

If you are submitting human sequences to GenBank, do not include any data that could reveal the personal identity of the source.

GenBank assumes that the submitter has received any necessary informed consent authorizations required prior to submitting sequences.

What is dbEST

dbEST 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).

About ESTs

Expressed Sequence Tags (ESTs) are short (usually <1000 bp), single-pass sequence reads from mRNA (cDNA). Typically they are produced in large batches. They represent a snapshot of genes expressed in a given tissue and/or at a given developmental stage. They are tags (some coding, others not) of expression for a given cDNA library.

Most EST projects develop large numbers of sequences. These are commonly submitted to GenBank and dbEST as batches of dozens to thousands of entries, with a great deal of redundancy in the citation, submitter and library information. To improve the efficiency of the submission process for this type of data, we have designed a special streamlined submission process and data format

dbEST also includes sequences that are longer than the traditional ESTs, or are produced as single sequences or in small batches.

Among these sequences are products of differential display experiments and RACE experiments. The thing that these sequences have in common with traditional ESTs, regardless of length, quality, or quantity, is that there is little information that can be annotated in the record.

If a sequence is later characterized and annotated with biological features such as a coding region, 5'UTR, or 3'UTR, it should be submitted through the regular GenBank submissions procedure (via BankIt or Sequin), even if part of the sequence is already in dbEST.

dbEST is reserved for single-pass reads. Assembled sequences should not be submitted to dbEST. GenBank will accept assembled EST submissions for the TSA (Transcriptome Shotgun Assembly) division.

Please contact gb-admin@ncbi.nlm.nih.gov for more information about submitting EST assemblies. The individual reads which make up the assembly should be submitted to dbEST, the Trace archive or the Short Read Archive (SRA) prior to the submission of the assemblies. For additional information about submitting to Trace or SRA please see [Trace web site](#).

NOTE: Sequences derived from "next generation" sequencing platforms, including Roche 454, Illumina, Applied Biosystems SOLiD, and Helicos Biosciences HeliScope, should be submitted to the [Short Read Archive \(SRA\)](#) (For information contact sra@ncbi.nlm.nih.gov .)

Sequences which should not be included in EST submissions include the following:

mitochondrial sequences,

rRNA, viral sequences,

vector sequences.

- Vector and linker regions should be removed from EST sequences before submission.

TIGR Gene Indices

Expressed Sequence Tags (ESTs) have provided a first glimpse of the collection of transcribed sequences in a variety of organisms.

However, a careful analysis of this sequence data can provide significant additional functional, structural, and evolutionary information. Our analysis of the public EST sequences, available through the TIGR Gene Indices (TGI; <http://www.tigr.org/tdb/tdb.html>), is an attempt to identify the genes represented by that data and to provide additional information regarding those genes.

Gene Indices are constructed for selected organisms by first clustering, then assembling EST and annotated gene sequences from GenBank. This process produces a set of unique, high-fidelity virtual transcripts, or Tentative Consensus

(TC) sequences. The TC sequences can be used to provide putative genes with functional annotation, to link the transcripts to mapping and genomic sequence data, and to provide links between orthologous and paralogous genes.

Tentative Consensus (TC) sequences which are annotated using a variety of tools including ORF prediction, putative annotation using a controlled vocabulary, Gene Ontology and Enzyme Commission number assignments, and maps to completed and draft genomes.

The TCs are used to construct a variety of other databases, including the Eukaryotic Gene Orthologue (EGO) database and RESOURCERER, a database that annotates and cross-references microarray resources for human, mouse, and rat. At present, 57 species are represented in the Gene Index Databases, including 19 animals, 17 plants, 7 fungi and 14 protists; this includes the 24 species most highly represented in by public EST sequencing projects.

Individual databases are updated and released three times yearly, on February 1, June 1, and October 1, if the number of available ESTs for that species has increased than 10% or 25,000, whichever is fewer.

U K CropNet

UK Crop Net: Comparative genomics (ascribing function through alignment of similar nucleotide or protein sequences) has become a key area of research in plant biotechnology. This is because genomes of closely related plant species have been found to have remarkably similar genes and gene functions. As the vast amount of plant genomic data becoming available, the use of bioinformatics to improve plant varieties is also becoming vital. To make sense of all the genomic data, UK CropNet was established in 1996 with specific aims of developing software and databases that will facilitate the querying of genomic information from different crop species. Particular emphasis has been placed on developing software tools for comparative mapping. UK CropNet has used the AceDB (Durbin & Thierry-Mieg, 1992) database system to create separate databases for each of the UK CropNet Projects with individual databases for *Arabidopsis*, Barley, *Brassica*, Forage grasses, and Millet.

PlantGDB: A resource for comparative plant genomics

PlantGDB serves the plant research community by providing access to plant sequence data as well as a variety of sequence and genome analysis tools in a single online resource [(1,2); Table 1].

This update outlines recent developments at PlantGDB that have expanded its usefulness as a tool for comparative genomics. Key features include:

- expanded EST assemblies;
- new genome browsers for a larger number of species;
- Overnight annotation of emerging genome sequences;
- and novel tools for sequence retrieval and analysis,

PlantGDB: Plant sequence data and transcript assemblies

PlantGDB periodically uploads and parses all Viridiplantae sequences from GenBank and Uniprot into 70 000 individual data sets according to species or subspecies origin.

PlantGDB's sequence data are refreshed approximately every 4 months, coinciding with alternate bimonthly GenBank version releases.

GenBank and Uniprot sequences are uploaded, parsed by (sub)species, indexed for BLAST and GeneSeqer analysis, and loaded into MySQL.

For all species with *410 000 published transcripts, a non-redundant set*

of PlantGDB-generated Unique Transcripts (PUTs) is generated using a custom assembly pipeline.

PUTs are aligned to UniProt entries using BLAST, and the best matches (if any) and UniProt-associated Gene Ontology (GO) annotations (7) are stored.

Currently, 116 species have PUT assemblies at PlantGDB, spanning

diverse taxonomic groups (Figure 2). Users can track assembly progress in PlantGDB at <http://www.plantgdb.org/prj/ESTCluster/progress.php>.

PlantGDB:

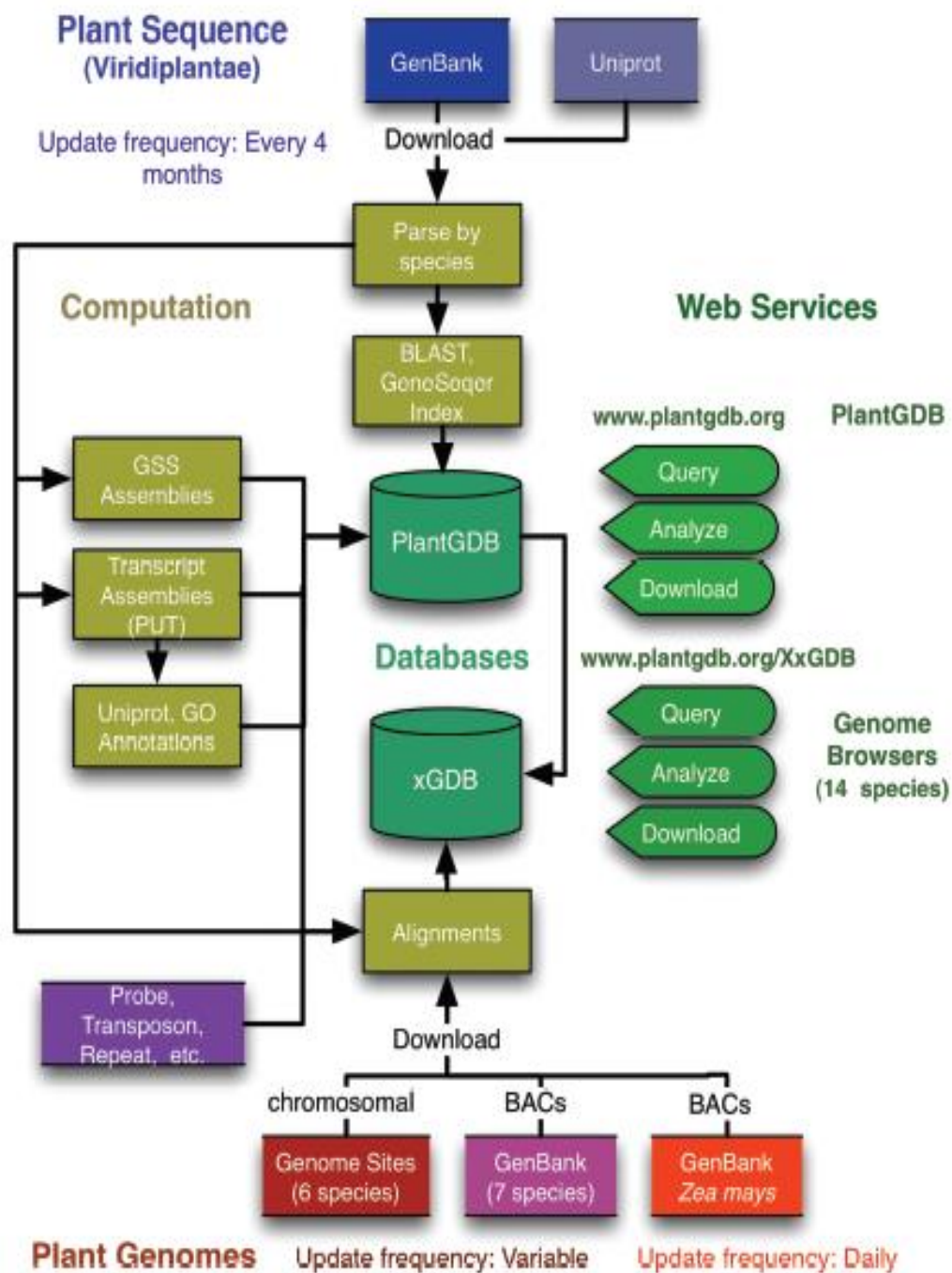


Figure 1. Database schema for PlantGDB, showing data sources, update frequency, computation and web services. PlantGDB is accessible at <http://www.plantgdb.org>, and genome browsers are accessible at <http://www.plantgdb.org/XxGDB>, where Xx is the first letter of the genus and species (e.g. A xGDB = *Arabidopsis thaliana* genome database).

PlantGDB: Genome Browsers

PlantGDB provides genome browsers (xGDB) for 14 plant species whose genomes have been completely or partially sequenced (12). AtGDB, OsGDB, MtGDB, SbGDB, VvGDB and PtGDB are chromosome-based genome browsers for *Arabidopsis thaliana* (thale cress), *Oryza sativa* (rice), *Medicago truncatula* (barrel medic), *Sorghum bicolor* (sorghum), *Vitis vinifera* (grapevine) and *Populus trichocarpa* (western balsam poplar), respectively,

while ZmGDB, GmGDB, HvGDB, SIGDB, GhGDB,

TaGDB, BrGDB and LjGDB are BAC-based browsers

for *Zea mays* (corn or maize), *Glycine max* (soybean),

Hordeum vulgare (barley), *Solanum lycopersicum*

(tomato), *Gossypium hirsutum* (cotton), *Triticum aestivum*

(bread wheat), *Brassica rapa* (field mustard) and *Lotus*

japonicus, respectively

(http://www.plantgdb.org/prj/Genome_browser.php).

GeneSeqer@PlantGDB: gene structure prediction in plant genomes

The GeneSeqer@PlantGDB Web server

(<http://www.plantgdb.org/cgi-bin/GeneSeqer.cgi>) provides a gene structure prediction tool tailored for applications to plant genomic sequences.

Predictions are based on spliced alignment with source-native ESTs and full-length cDNAs or non-native probes derived from putative homologous genes.

The tool is illustrated with applications to refinement of current gene structure annotation and de novo annotation of draft genomic sequences.

The service should facilitate expert annotation as a community effort by providing convenient access to all public plant sequences via the PlantGDB database, a simple four-step protocol for spliced alignment and visually appealing displays of the predicted gene structures in addition to detailed sequence alignments.

GrainGenes: the genome database for small-grain crops

GrainGenes, <http://www.graingenes.org>, is the international database for the wheat, barley, rye and oat genomes.

For these species it is the primary repository for information about genetic maps, mapping probes and primers, genes, alleles and QTLs.

Documentation includes such data as primer sequences, polymorphism descriptions, genotype and trait scoring data, experimental protocols used, and photographs of marker polymorphisms, disease symptoms and mutant phenotypes.

These data, curated with the help of many members of the research community, are integrated with sequence and bibliographic records selected from external databases and results of BLAST searches of the ESTs.

GrainGenes:

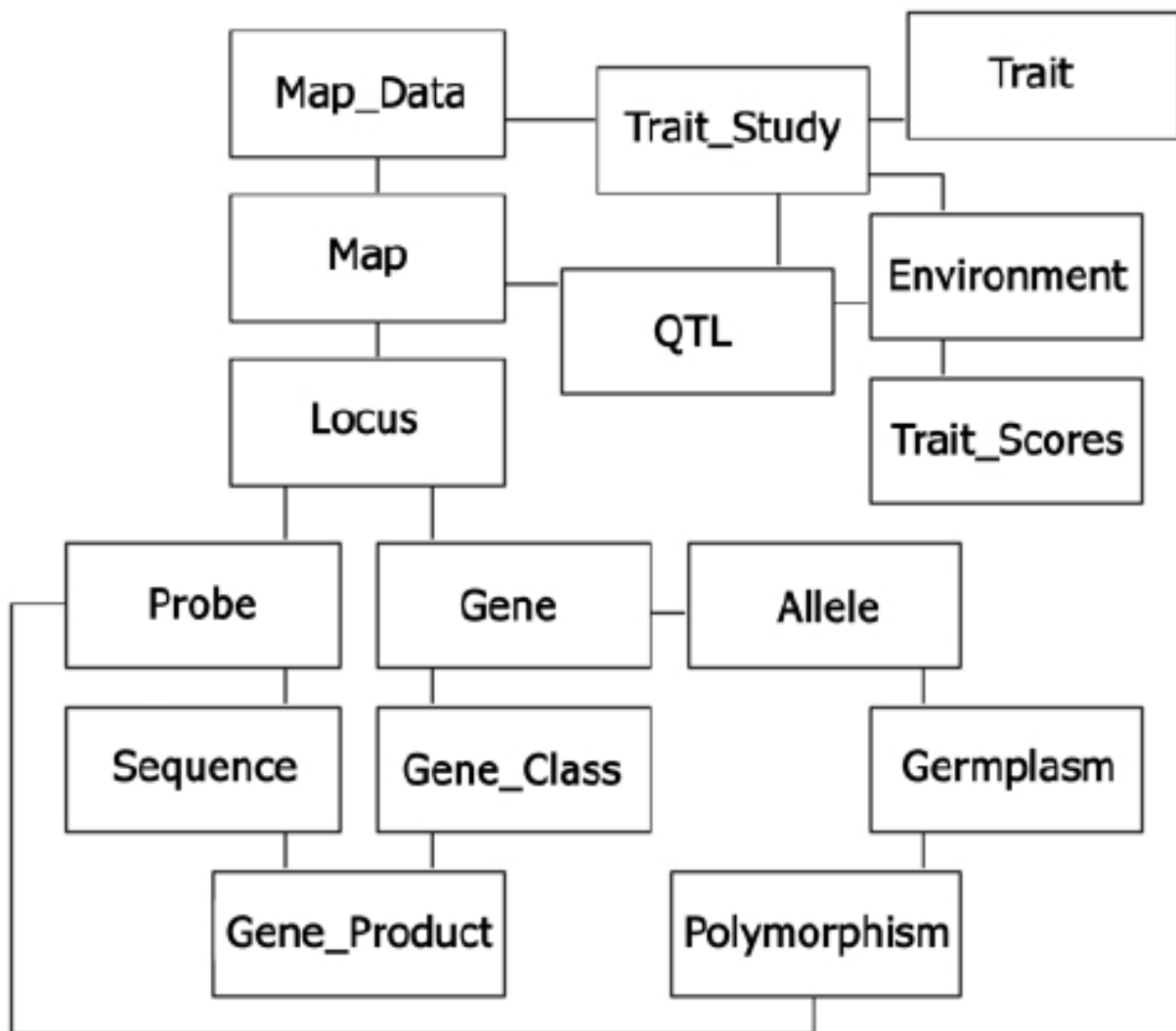


Figure 2. Recommended browse map of GrainGenes. Secondary interconnections and some additional data classes are omitted for clarity.

Gramene 2016: comparative plant genomics and pathway resources

Gramene (<http://www.gramene.org>) is an online resource for comparative functional genomics in crops and model plant species. Its two main frameworks are genomes (collaboration with Ensembl Plants) and pathways (The Plant Reactome and archival BioCyc databases). Since our last NAR update, the database website adopted a new Drupal

management platform. The genomes section features 39 fully assembled reference genomes that are integrated using ontology-based annotation and comparative analyses, and accessed through both visual and programmatic interfaces. Additional community data, such as genetic variation, expression and methylation, are also mapped for a subset of genomes.

The Plant Reactome pathway portal (<http://plantreactome.gramene.org>) provides a reference resource for analyzing plant metabolic and regulatory pathways. In addition to ~200 curated rice reference pathways, the portal hosts gene homology-based pathway projections for 33 plant species. Both the genome and pathway browsers interface with the EMBL-EBI's

KOMUGI

KOMUGI is an integrated database of wheat created by Japanese researchers in wheat sciences. "Komugi" is a Japanese word meaning wheat. This database was originally created in 1996 by a voluntary group of Japanese scientists who were concerned about the genetic stocks of wheat (the genus *Triticum*) and related species (the genus *Aegilops*) conserved in Japanese institutions. The genetic stocks include the original accessions of *Triticum* and *Aegilops* collected during expeditions led by the late Dr. H. Kihara and his colleagues to their places of origin. They also include research stocks of wheat that were produced by many scientists world-wide as shown in the table below and some accessions of rye and oats. KOMUGI also contains the Catalogue of Gene Symbols for Wheat and gene sequence data.

The Committee has been attempting to improve KOMUGI by avoiding duplicated registration and adding more information on the chromosomal constitutions and phenotypes (especially photographic

images) of the seed stocks. It has also started to create substantial databases of DNA clones and sequences.

North American Barley Genome Project

The North American Barley Genome Project (NABGP) has supported QTL analysis of malting and agronomic traits in three reference populations of doubled haploid lines derived from the following crosses: Steptoe x Morex; Harrington x TR306; and Harrington x Morex. Steptoe and Morex are six-rowed varieties. Harrington and TR306 are two-rowed varieties. Harrington and Morex, respectively, are the two-rowed and six-rowed malting quality standards for North America. Steptoe and TR306 do not have acceptable malting quality profiles, but they have desirable agronomic attributes.

Rice Genome Research Program

Rice International Rice Genome Sequencing Project (IRGSP) - The International Rice Genome Sequencing Project (IRGSP), a consortium of publicly funded laboratories, was established in 1997 to obtain a high quality, map-based sequence of the rice genome using the cultivar Nipponbare of *Oryza sativa* ssp. japonica. It is currently comprised of ten members: Japan, the United States of America, China, Taiwan, Korea, India, Thailand, France, Brazil, and the United Kingdom. The IRGSP adopts the clone-by-clone shotgun sequencing strategy so that each sequenced clone can be associated with a specific position on the genetic map and adheres to the policy of immediate release of the sequence data to the public domain. In December 2004, the IRGSP completed the sequencing of the rice genome. The high-quality and map-based sequence of the entire genome is now available in public databases.

US Wheat Genome Project

The IWGSC, with 2,400 members in 68 countries, is an international, collaborative consortium, established in 2005 by a group of wheat growers, plant scientists, and public and private breeders. The vision of the IWGSC is a high quality genome sequence of bread wheat that serves as a foundation for the accelerated development of improved varieties and that empowers all aspects of basic and applied wheat science.

The four key milestones of the IWGSC roadmap are:

- (1) use of survey sequences of the 21 bread wheat chromosomes to assign gene sequences to individual chromosomes;
- (2) develop physical maps to provide resources for sequencing;
- (3) deliver a reference sequence for each of the chromosomes; and
- (4) produce a gold standard genome sequence by integrating chromosome based genomic resources with the IWGSC whole genome assembly.

Maize GDB

First released in 1991 with the name MaizeDB, the Maize Genetics and Genomics Database.

MaizeGDB has transitioned from a focus on comprehensive curation of the literature, genetic maps and stocks to a paradigm that accommodates the recent release of a reference maize genome sequence, multiple diverse maize genomes and sequence-based gene expression data sets.

The MaizeGDB Team is relatively small, and relies heavily on the research community to provide data, nomenclature standards and most importantly, to recommend future directions, priorities and strategies. Key aspects of MaizeGDB's intimate interaction with the community are the co-location of curators with maize research groups in multiple locations across the USA as well as coordination with MaizeGDB's close partner, the Maize Genetics Cooperation—Stock Center.

MaizeGDB support updates to the functional and structural annotation of the B73 reference genome.

MaizeGDB :Genome annotation in the past

Over the years, about 6000 functional genes described in the literature have been curated into MaizeGDB. However, this count is dwarfed by the 32,540 gene models predicted for the B73 genome by the Maize Genome Sequencing Consortium, the 25,703 RefSeq cDNA-unigenes in GenBank , and the approximately 10,000 gene models currently thought not to exist in B73, but to be present in other inbred lines.

Genome annotation in the future

Its two top priorities, to have the maize B73 sequence assembly improved and annotated. At the request of the Maize Genetics Executive Committee, we are in the process of forming a Maize Genome Annotation Consortium. The MaizeGDB Team appreciates the importance of helping people work together toward a common goal and has recently posted guidelines on its website, <http://www.maizegdb.org/assembly.php>, to groups who plan to engage in sequence assembly and annotation.

The Arabidopsis Information Resource (TAIR)

Arabidopsis thaliana is the most widely-studied plant today. The concerted efforts of over 11 000 researchers and 4000 organizations around the world are generating a rich diversity and quantity of information and materials. This information is

made available through a comprehensive on-line resource called the Arabidopsis Information Resource (TAIR) (<http://arabidopsis.org>), which is accessible via commonly used web browsers and can be searched and downloaded in a number of ways.

Efforts have been focused on increasing data content and diversity, functionally annotating genes and gene products with controlled vocabularies, and improving data retrieval, analysis and visualization tools. New information include sequence polymorphisms including alleles, germplasms and phenotypes,

Gene Ontology annotations, gene families, protein information, metabolic pathways, gene expression data from microarray experiments and seed and DNA stocks.

New data visualization and analysis tools include SeqViewer, which interactively displays the genome from the whole chromosome down to 10 kb of nucleotide sequence and

AraCyc, a metabolic pathway database and map tool that allows overlaying expression data onto the pathway diagrams.

Finally, we have recently incorporated seed and DNA stock information from the Arabidopsis Biological Resource Center (ABRC) and implemented a shopping-cart style on-line ordering system.

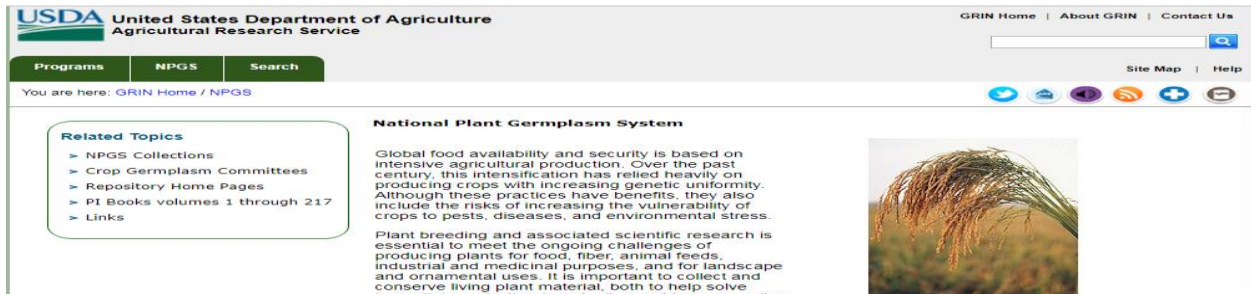
Clemson University Genomics Institute (CUGI)

The Clemson University Genomics Institute (CUGI) offers a full complement of state-of the-art tools and services to aid discovery and functional analysis of important genes and genomic regions. CUGI is organized into six groups.

- (1) The BAC/EST Resource Center (BRC) produces high quality bacterial artificial chromosome (BAC), small insert/shotgun and cDNA libraries from plants, animals and microbes. CUGI has one of the largest collections of BAC libraries in the world.
- (2) The Bioinformatics Group supports the operations of the other five laboratory-based branches of CUGI by providing computational infrastructure.
- (3) The Functional Genomics Group is equipped with a Agilent BioAnalyzer, a Agilent microarray scanner, and a Stratgene Mx-3000 real-time PCR System.
- (4) The Physical Mapping Group uses a combination of BAC fingerprinting and marker hybridization to assemble BAC clones into contigs and to anchor them to chromosomes.
- (5) The Proteomics lab suite includes a hybrid quadrupole time of flight mass spectrometer (QToF, Waters) with electrospray ionization.
- (6) The DNA Sequencing Group delivers highthroughput DNA sequencing of

any source of DNA. The center also performs sequence assembly, gap closure, and finishing.

National Plant Germplasm System (NPGS)



USDA United States Department of Agriculture
Agricultural Research Service

GRIN Home | About GRIN | Contact Us

Programs | NPGS | Search

You are here: GRIN Home / NPGS


Related Topics

- > NPGS Collections
- > Crop Germplasm Committees
- > Repository Home Pages
- > PI Books volumes 1 through 217
- > Links

National Plant Germplasm System

Global food availability and security is based on intensive agricultural production. Over the past century, this intensification has relied heavily on producing crops with increasing genetic uniformity. Although these practices have benefits, they also include the risks of increasing the vulnerability of crops to pests, diseases, and environmental stress.

Plant breeding and associated scientific research is essential to meet the ongoing challenges of producing plants for food, fiber, animal feeds, industrial and medicinal purposes, and for landscape and ornamental uses. It is important to collect and conserve living plant material, both to help solve immediate agricultural production problems as well as

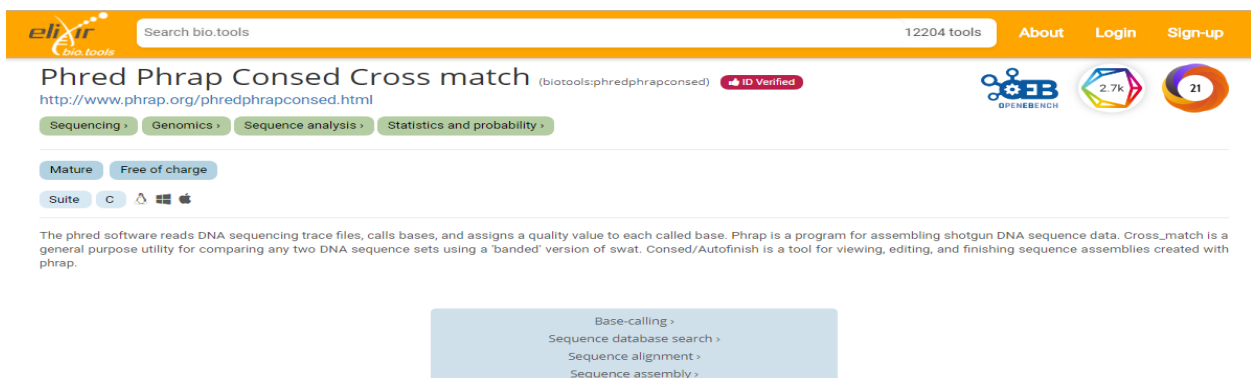


The mission of the NPGS is to support agricultural production by:

- acquiring crop germplasm
- conserving crop germplasm
- evaluating and characterizing crop germplasm
- documenting crop germplasm
- distributing crop germplasm

Through these efforts, NPGS assists in improving the quality and productivity of crops. The [GRIN](#) database is managed by the Database Management Unit ([DBMU](#)), while the acquisition of plants is managed by the Plant Exchange Office ([PEO](#)).

Phred Phrap Consed Cross match






bio.tools Search bio.tools 12204 tools About Login Sign-up

Phred Phrap Consed Cross match (bio.tools:phredphrapconsed) ID Verified

<http://www.phrap.org/phredphrapconsed.html>

Sequencing > Genomics > Sequence analysis > Statistics and probability >

Mature Free of charge

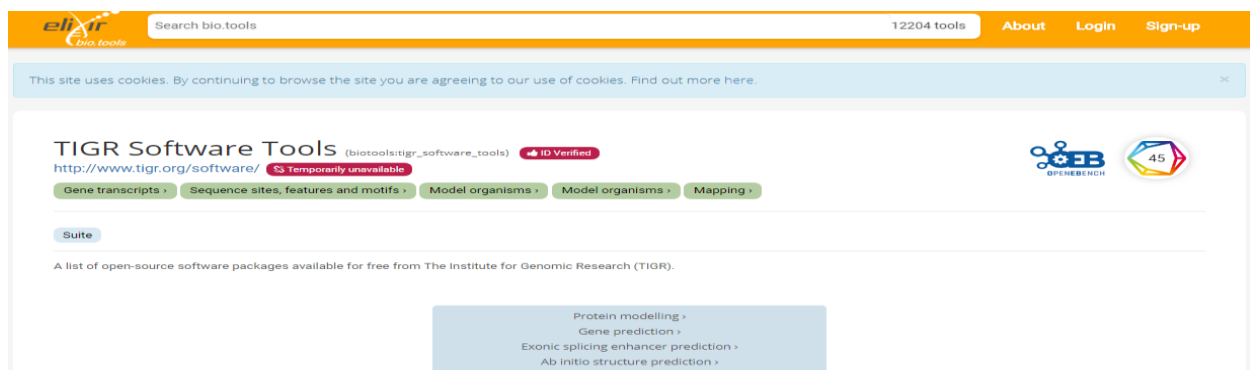
Suite C   

The phred software reads DNA sequencing trace files, calls bases, and assigns a quality value to each called base. Phrap is a program for assembling shotgun DNA sequence data. Cross_match is a general purpose utility for comparing any two DNA sequence sets using a 'banded' version of swat. Consed/Autofinish is a tool for viewing, editing, and finishing sequence assemblies created with phrap.

Base-calling >
Sequence database search >
Sequence alignment >
Sequence assembly >

The phred software reads DNA sequencing trace files, calls bases, and assigns a quality value to each called base. Phrap is a program for assembling shotgun DNA sequence data. Cross_match is a general purpose utility for comparing any two DNA sequence sets using a 'banded' version of swat. Consed/Autofinish is a tool for viewing, editing, and finishing sequence assemblies created with phrap.

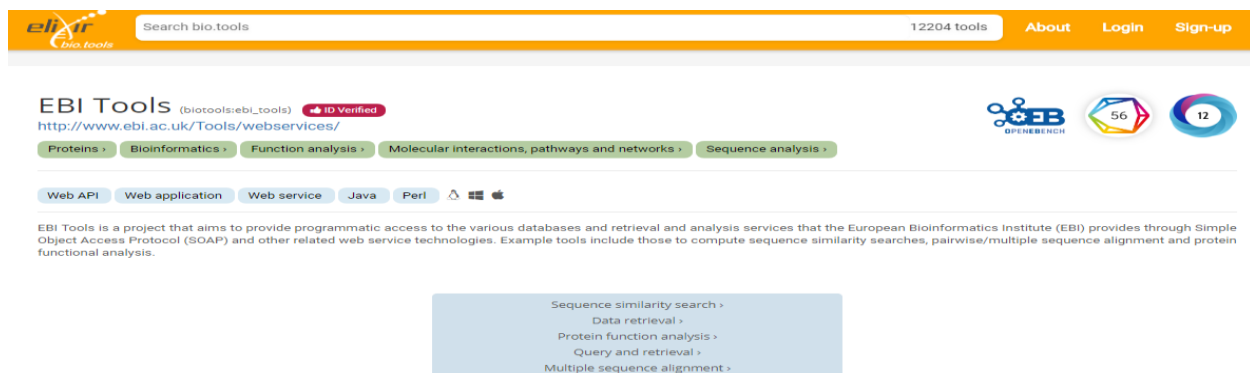
TIGR Software Tools



The screenshot shows the TIGR Software Tools page on the bio.tools website. The page features a search bar at the top with the text "Search bio.tools" and a button labeled "12204 tools". To the right of the search bar are links for "About", "Login", and "Sign-up". Below the search bar is a cookie consent banner. The main content area displays the title "TIGR Software Tools" with a sub-label "(biotools:tigr_software_tools)" and a "ID Verified" badge. The URL "http://www.tigr.org/software/" is shown, along with a "Temporarily unavailable" status. Navigation tabs include "Gene transcripts", "Sequence sites, features and motifs", "Model organisms", "Model organisms", and "Mapping". A "Suite" button is also present. Below the tabs, a description states: "A list of open-source software packages available for free from The Institute for Genomic Research (TIGR)." A list of tools is displayed in a light blue box: "Protein modelling >", "Gene prediction >", "Exonic splicing enhancer prediction >", and "Ab initio structure prediction >".

A list of open-source software packages available for free from The Institute for Genomic Research (TIGR)

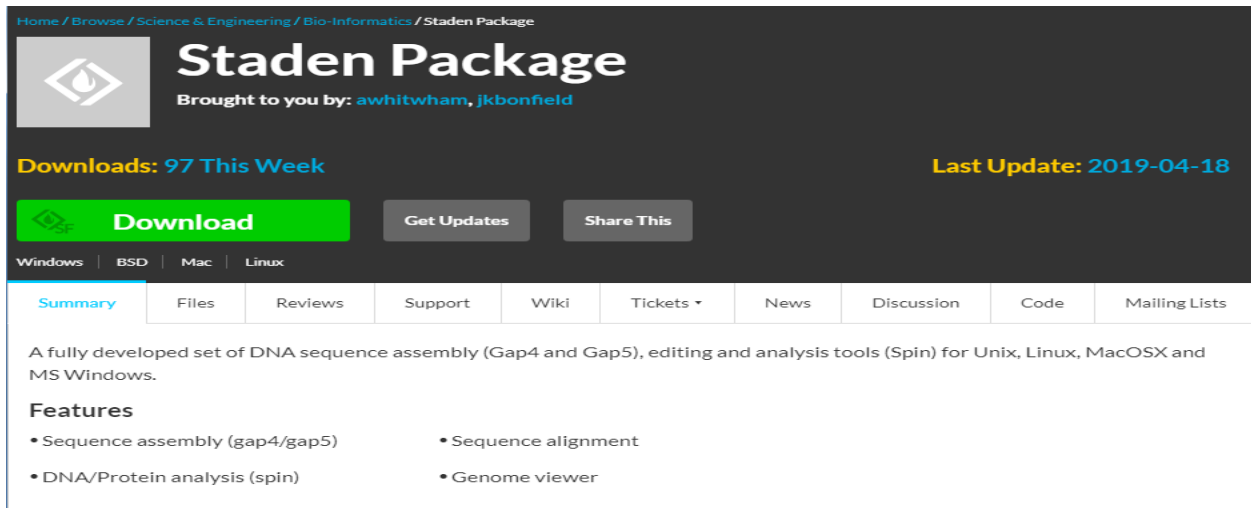
EBI Tools



The screenshot shows the EBI Tools page on the bio.tools website. The page features a search bar at the top with the text "Search bio.tools" and a button labeled "12204 tools". To the right of the search bar are links for "About", "Login", and "Sign-up". Below the search bar is a cookie consent banner. The main content area displays the title "EBI Tools" with a sub-label "(biotools:ebi_tools)" and a "ID Verified" badge. The URL "http://www.ebi.ac.uk/Tools/webservices/" is shown. Navigation tabs include "Proteins", "Bioinformatics", "Function analysis", "Molecular interactions, pathways and networks", and "Sequence analysis". Below the tabs, a list of tool categories is shown: "Web API", "Web application", "Web service", "Java", "Perl", and icons for Linux, Windows, and Mac. A description states: "EBI Tools is a project that aims to provide programmatic access to the various databases and retrieval and analysis services that the European Bioinformatics Institute (EBI) provides through Simple Object Access Protocol (SOAP) and other related web service technologies. Example tools include those to compute sequence similarity searches, pairwise/multiple sequence alignment and protein functional analysis." A list of tools is displayed in a light blue box: "Sequence similarity search >", "Data retrieval >", "Protein function analysis >", "Query and retrieval >", and "Multiple sequence alignment >".

EBI Tools is a project that aims to provide programmatic access to the various databases and retrieval and analysis services that the European Bioinformatics Institute (EBI) provides through Simple Object Access Protocol (SOAP) and other related web service technologies. Example tools include those to compute sequence similarity searches, pairwise/multiple sequence alignment and protein functional analysis.

Staden Package



The screenshot shows the project page for the Staden Package. At the top, there is a breadcrumb trail: Home / Browse / Science & Engineering / Bio-Informatics / Staden Package. The main header features the Staden Package logo (a stylized flame or drop) and the text "Staden Package Brought to you by: [awhitwham](#), [jkbonfield](#)". Below the header, it displays "Downloads: 97 This Week" and "Last Update: 2019-04-18". There are three buttons: a prominent green "Download" button, a grey "Get Updates" button, and a grey "Share This" button. Below these buttons, there are tabs for operating systems: Windows, BSD, Mac, and Linux. A navigation menu includes links for Summary, Files, Reviews, Support, Wiki, Tickets, News, Discussion, Code, and Mailing Lists. The main content area starts with a description: "A fully developed set of DNA sequence assembly (Gap4 and Gap5), editing and analysis tools (Spin) for Unix, Linux, MacOSX and MS Windows." followed by a "Features" section with a bulleted list:

- Sequence assembly (gap4/gap5)
- Sequence alignment
- DNA/Protein analysis (spin)
- Genome viewer

A fully developed set of DNA sequence assembly (Gap4 and Gap5), editing and analysis tools (Spin) for Unix, Linux, MacOSX and MS Windows.

Access to literature

Literature (Resources):

- Books (printed or e-books)
- Journals (printed or e-books)
- Research reports (both)
- Institutional publications(both)
- Govt. publications (both)
- Various NGO's/ INGO's publications
- Internet (Online resources)
- Intranet (Offline resources)

Sources of Information (Types) :

By nature/content of information-

1. Primary literature

2. Secondary literature

3. Tertiary literature

By format/media/channel

a. Hard copy (Print)

b. Soft copy(Electronic)

Primary literature

- Primary sources are the authentic publication of an expert's new evidence, conclusions and proposals (case reports, clinical trials, etc).

- It usually published in a peer-reviewed journal.
- Preliminary reports, congress papers and preprints also constitute primary literature

Secondary literature

Secondary sources are systematic review articles or meta-analyses where material derived from primary source literature are inferred and evaluated.

Tertiary literature

It consists of collections that compile information from primary or secondary literature (eg., reference books).

How can we search? :

A. Printed

B. Electronic

(a) Offline (Intranet) e.g. DVD, CD, cassette

(b) Online (Internet)

- **Internet searching** for the purpose of academic, business and others.
- **Pubmed (Databases)** searching exclusively for medical literature

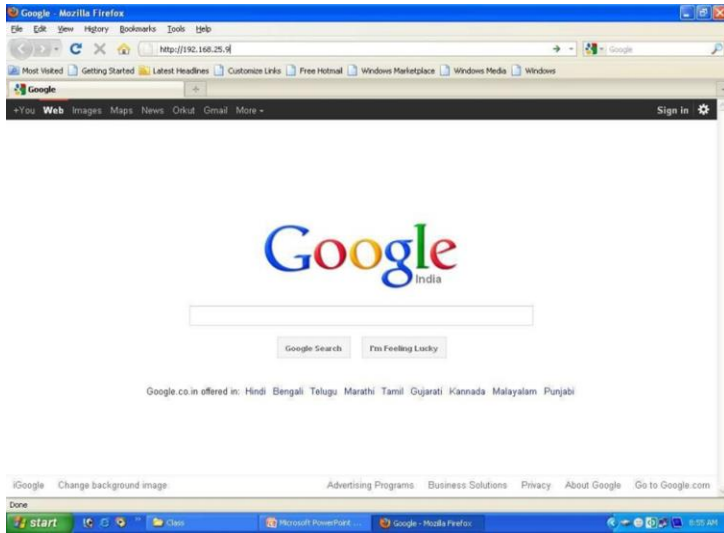
Types A and B(a) can be access in Library

Searching.....:

We should know the techniques-

- to access to relevant information,
- to get the recall (unwanted) less and precise (relevant) information more, and

- familiar with various techniques that the Search engine such as Google, Yahoo, Altavista, pubmed etc. used.



Resource	Web address
Search engines	
Google	http://www.google.com
Google Scholar	http://www.scholar.google.com
Yahoo	http://www.yahoo.com
Electronic source of database	
PubMed	https://www.nlm.nih.gov/pubmed
MeSH	http://www.ncbi.nlm.nih.gov/mesh
Medline (Medical Literature Analysis and Retrieval System Online)	https://www.nlm.nih.gov
CINAHL (The Cumulative Index to Nursing and Allied Health)	https://www.cinahl.com
Embase (Excerpta Medica Database)	https://store.elsevier.com/embase
SCOPUS	https://www.scopus.com/
Ind Med: Indian Database	https://www.medind.nic.in
ERIC	https://www.eric.ed.gov
ProQuest	http://proquest.com

Resource	Web address
The Cochrane Database of Systematic Reviews	http://www.cochranelibrary.com/
The ACP Journal Club	http://search.ebscohost.com/
Dartmouth EBM Database	http://www.dartmouth.edu/~library/biomed/resources/ejournals.html
Evidence updates	http://plus.mcmaster.ca/evidenceupdates/
e Medicine	http://emedicine.medscape.com/
National Guideline Clearinghouse	http://www.guideline.gov/
Ovid Medline	http://www.dartmouth.edu/~library/biomed/resources/ovid.html
PubMed	http://www.ncbi.nlm.nih.gov/pubmed/
TRIP (Turning Research into Practice) Database Plus	https://www.tripdatabase.com/
UpToDate	http://www.uptodate.com/online

- Acta Horticulture
- Agricola
- GeoBase
- Informit
- OECD iLibrary
- Scopus
- SciFinder Web
- Sci-hub
- Web of Science
- AGRIS (AGRIcultural Sciences)
- CAB abstract
- CARIS (current agricultural research information system)

SHODHGANGA

The Shodhganga@INFLIBNET Centre provides a platform for research students to deposit their Ph.D. theses and make it available to the entire scholarly community in open access at <http://shodhganga.inflibnet.ac.in>

Other Tips....

Specify your search terms

- > Use professional terms
- > Use the keywords from standard tools- SLSH, LCSH, MeSH and others
- > Use lowercase letters
- > Ignore preposition, article and other stop words
- > Read the searching guideline of the particular databases

PubMed :



PubMed is a free resource developed and maintained by the National Center for Biotechnology Information (NCBI) at the National Library of Medicine® (NLM)

PubMed has been available since 1996. Its more than 28 million references include the MEDLINE database.

The screenshot shows a web browser window with the address bar displaying "www.ncbi.nlm.nih.gov/pubmed?term='Substance-Related Disorders/nursing'[Mesh]". The page header includes the NCBI logo and search options. The main content area shows search results for "Substance-Related Disorders/nursing"[Mesh]. The results are displayed in a list format, with the first two results visible:

- [Development of guidelines for nurse-led discharge of children presenting with toxic ingestion.](#)
Lawton L.
Nurs Child Young People. 2012 Nov;24(9):21-3.
PMID: 23311262 [PubMed - indexed for MEDLINE]
[Related citations](#)
- [The effects of cannabis on mental health.](#)
Shrubbs R.

The page also features a "Results by year" bar chart, a "14 free full-text articles in PubMed Central" section, and various filter options on the left side.

PubMed provides free access to

- ✓ MEDLINE, NLM's database of citations and abstracts in the fields of medicine, nursing, dentistry, veterinary medicine, health care systems, and preclinical sciences.
- ✓ Citations that provide a record for an article before it is indexed with MeSH® and added to MEDLINE or converted to out-of-scope status.
- ✓ Citations that precede the date that a journal was selected for MEDLINE indexing.
- ✓ Citations to articles that are out-of-scope (e.g., covering plate tectonics or astrophysics) from certain MEDLINE journals, primarily general science and general chemistry journals, for which the life sciences articles are indexed with MeSH for MEDLINE.

- ✓ Citations to some additional life science journals that submit full-text articles to PubMed Central® and receive a qualitative review by NLM.
- ✓ Citations for the majority of books and book chapters available on the NCBI Bookshelf

PubMed also

- Links to full-text articles found in PubMed Central or at publisher web sites, and other related resources.
- Provides Advanced search, Clinical Queries search filters, and Special Queries pages.
- Links to related articles and provides discovery tools for other data that may be of interest.
- Includes automatic e-mailing of search updates, the ability to save records, and filters for search results using "My NCBI".
- Includes a spell checker feature.
- Links to NCBI molecular biology resources.

MEDLINE®: Description of the Database

MEDLINE is the U.S. National Library of Medicine® (NLM) premier bibliographic database that contains more than 25 million references to journal articles in life sciences with a concentration on biomedicine. A distinctive feature of MEDLINE is that the records are indexed with NLM Medical Subject Headings (MeSH®). MEDLINE is the online counterpart to MEDLARS® (MEDical Literature Analysis and Retrieval System) that originated in 1964.

Access to MEDLINE data are also available via services and products developed by organizations that download the database from NLM. Access to various MEDLINE services is often available from medical libraries, many public libraries, and commercial sources.

MedlinePlus®, another service offered by the NLM, provides consumer-oriented health information. Health consumers are encouraged to discuss search results with their health care provider.

ISI Citation Database....

The Institute for Scientific Information (ISI) was an academic publishing service, founded by Eugene Garfield in Philadelphia in 1960. ISI offered bibliographic database services. Its specialty was citation indexing and analysis, a field pioneered by Garfield

ISI maintained citation databases covering thousands of academic journals, including a continuation of its longtime print-based indexing service the Science Citation Index (SCI), as well as the Social Sciences Citation Index (SSCI) and the Arts and Humanities Citation Index (AHCI). All of these were available via ISI's Web of Knowledge database service.

ISI was acquired by Thomson Scientific & Healthcare in 1992, and became known as Thomson ISI. It was a part of the Intellectual Property & Science business of Thomson Reuters until 2016, when the IP & Science business was sold, becoming Clarivate Analytics. In February 2018, Clarivate announced it will re-establish ISI as part of its Scientific and Academic Research group. It exists as a group within Clarivate as of November 2018.

Current contents®

Current Contents is a rapid alerting service database from Clarivate Analytics, formerly the Institute for Scientific Information and Thomson Reuters. It is published online and in several different printed subject sections.

Current Contents was first published in paper format, in a single edition devoted only to biology and medicine. Other subject editions were added later. Initially, it consisted simply of a reproduction of the title pages from several hundred major peer-reviewed scientific journals, and was published weekly, with the issues containing title pages from journal issues only a few weeks previously, a shorter time lag than any service then available. There was an author index and a crude keyword subject index only. Author addresses were provided so readers could send reprint requests for copies of the actual articles..

BIOSIS Previews....

BIOSIS Previews is an English-language, bibliographic database service, with abstracts and citation indexing. It is part of *Clarivate Analytics Web of Science* suite. BIOSIS Previews indexes data from 1926 to the present.

BIOSIS Previews is part of the *Life Sciences in Web of Science*. Its coverage encompasses the life sciences and biomedical sciences literature, with deep global coverage on a wide range of related subject areas. This is accomplished with access to indexed journal content from *Biological Abstracts*, and supplemental indexed non-journal content from *Biological Abstracts/Reports*, *Reviews*, *Meetings (BA/RRM or Biological Abstracts/RRM)* and the major publications of BIOSIS

The database is also provided by EBSCO Information Services through a partnership with Clarivate Analytics. *Biological Abstracts* consists of 350,000 references for almost 5,000 primary journal and monograph titles. *Biological Abstracts/RRM* additionally includes more than 200,000 *non-journal citations*.

Biological Abstracts/RRM is the former *BioResearch Index*.

Pascal (Database)

PASCAL is a scientific bibliographic database, which is maintained by INIST (CNRS). PASCAL covers the core scientific literature in science, technology and medicine with special emphasis on European literature.

As of 2012, PASCAL maintains a database of more than 17 million records, 90% of these are author abstracts. Its coverage is from 1973 to present. Its source documents are composed of journal articles at 88% (3,085 international titles), proceedings at 9%, and dissertations, books, patents, and reports account combined for 3%

EMBASE.....

Embase (often styled **EMBASE** for **Excerpta Medica dataBASE**) is a biomedical and pharmacological bibliographic database of published literature designed to support information managers and pharmacovigilance in complying with the regulatory requirements of a licensed drug. Embase, produced by Elsevier, contains over 32 million records from over 8,500 currently published journals from 1947^[1] to the present. Through its international coverage, daily updates, and drug indexing with Emtree, Embase enables tracking and retrieval of drug information in the published literature. Each record is fully indexed and Articles in Press are available for some records and In Process are available for all records, ahead of full indexing. Embase's international coverage expands across biomedical journals from 95 countries and is available through a number of database vendors

Cochrane Reviews Cochrane Library

The Cochrane Library (named after Archie Cochrane) is a collection of databases in medicine and other healthcare specialties provided by Cochrane and other organizations. At its core is the collection of Cochrane Reviews, a database of systematic reviews and meta-analyses which summarize and interpret the results of medical research. The Cochrane Library aims to make the results of well-conducted controlled trials readily available and is a key resource in evidence-based medicine.

AGRICOLA.....

AGRICOLA (AGRICultural OnLine Access) is an online database created and maintained by the United States National Agricultural Library of the United States Department of Agriculture.

The database serves as the catalog and index for the collections of the United States National Agricultural Library, but it also provides public access to information on agriculture and allied fields

Agripedia....

Agripedia is an Internet Accessible Interactive Multimedia Instructional Resource, developed by the University of Kentucky's College of Agriculture created with a USDA Higher Education Program Grant.

Agripedia has been created to allow students to use resources across the country in an integrated, interactive learning resource on the Internet.

As a "encyclopedia of agriculture", Agripedia presents facts, figures, demonstrations, examples, graphs and more regarding the concepts, practices, and vocabulary of agriculture in a multimedia format using audio clips, graphics, text and animation. AGRIPEDIA not only provides ready access to instructional sites.

Agripedia is a service to the African community to support farmers, students, and the industry. We want to provide advance trading, economic, financial, educational and social services to our clients.

Agropedia.....

Agropedia is an online knowledge repository for information related to agriculture in India.

It includes universal meta models and localized content for a variety of users with appropriate interfaces built in collaborative mode in multiple languages.

It is backed by Government of India and sponsored by the World Bank through the National Agricultural Innovation Project of the Indian Council of Agricultural Research (ICAR), the open access online resource project was launched on 12 January 2009. Eventually it aims to use weekly alerts from scientists on different crops, to send text messages to farmers across India.

Partners of Agropedia:

- Information and Communication Technology (ICT)
- Indian Institute of Technology (IIT Kanpur)
- Indian Institute of Technology (IIT Bombay)
- Indian Institute of Information Technology and Management, Thiruvananthapuram (IIITM- K)
- National Academy of Agricultural Research Management (NAARM) (Indian Council of Agricultural Research -ICAR)
- Govind Ballabh Pant University of Agriculture & Technology, Pantnagar
- University of Agricultural Sciences, Dharwad (UASD)
- International Crops Research Institute for the Semi-Arid Tropics (ICRISAT)

Create a Formatted Bibliography :

The screenshot shows the EndNote X2 interface with a bibliography table and the 'Output Styles' menu open. The table lists various references with columns for Author, Year, Title, Journal, Ref Type, and URL. The 'Output Styles' menu is currently set to 'Annotated'.

Author	Year	Title	Journal	Ref Type	URL
Winter	2007	3LGM(2)-Modeling to support manag...		Conference...	<Go to ISI>
Ash	2005	Adding insight: A qualitative cross-sit...	Internationa...	Journal Arti...	<Go to ISI>
Kaplan	1997	Addressing organizational issues int...	Journal of t...	Journal Arti...	<Go to ISI>
Simon	2007	Adoption of order entry with decision ...	Journal of t...	Journal Arti...	<Go to ISI>
Hu	2002	Adoption of telemedicine technology ...	Journal of ...	Journal Arti...	<Go to ISI>
Blobel	2006	Advanced EHR architectures - Promi...		Conference...	<Go to ISI>
Blobel	2000	Advanced tool kits for EPR security		Conference...	<Go to ISI>
Haux	1997	Aims and tasks of medical informatics		Conference...	<Go to ISI>
Tang	2002	AMIA advocates national health infor...	Journal of t...	Journal Arti...	<Go to ISI>
	99	Analysis and design of an ontology fo...	Methods of ...	Journal Arti...	<Go to ISI>
	05	Analysis of information and communi...	lee Transa...	Journal Arti...	<Go to ISI>
	07	Analysis of the geographical accessi...	Internationa...	Journal Arti...	<Go to ISI>
	99	An application server approach for in...	Journal of t...	Journal Arti...	<Go to ISI>
	04	Approach to mobile information and ...		Conference...	<Go to ISI>
	00	ARCHIMED: A Network of Integrated...	Methods of ...	Journal Arti...	<Go to ISI>
Ciccarese	2005	Architectures and tools for innovative...	Internationa...	Journal Arti...	<Go to ISI>
Joubert	1998	ARIANE: Integration of information d...	Internationa...	Journal Arti...	<Go to ISI>
Gillies	2000	Assessing and improving the quality ...	Methods of ...	Journal Arti...	<Go to ISI>
Weir	2000	Assessing the implementation process		Conference...	<Go to ISI>
Lee	2001	Assessment of HIV/AIDS-related hea...	Information ...	Journal Arti...	<Go to ISI>
Luck	2006	An automated scoring algorithm for c...	Internationa...	Journal Arti...	<Go to ISI>
Giorgino	2005	Automated spoken dialogue system f...	Internationa...	Journal Arti...	<Go to ISI>
Boudioni	2003	Availability and use of information tou...	Aslib Proce...	Journal Arti...	<Go to ISI>
Ammenwerth	2005	Bad health informatics can kill - Is ev...	Methods of ...	Journal Arti...	<Go to ISI>
Vancamp	1989	A BASIC GUIDE TO ONLINE INFOR...	Online	Journal Arti...	<Go to ISI>
Horslen	1992	BASIC HEALTH-CARE FUNCTION...	Medical Inf...	Journal Arti...	<Go to ISI>
Westberg	1999	The basis for using the Internet to sup...	Journal of t...	Journal Arti...	<Go to ISI>
Weaver	2005	Bedside, classroom and bench: Coll...	Internationa...	Journal Arti...	<Go to ISI>
Singh	1997	Bhorugram (India): Revisited - A 4 ye...	Internationa...	Journal Arti...	<Go to ISI>
Rodriguez	2007	A BPMN extension for the modeling ...	leice Trans...	Journal Arti...	<Go to ISI>
Johnson	2004	Brainstorming about next-generation ...	Internationa...	Journal Arti...	<Go to ISI>

The 'Output Styles' menu is open, showing the following options:

- New Style...
- Edit "Annotated"
- Open Style Manager...
- Annotated (selected)
- Author-Date
- Numbered

Annotated Bibliography

Blobel, B. (2000). Advanced tool kits for EPR security.

Responding to the challenge for efficient and high quality health care, the shared care paradigm must be established in health. In that context, information systems such as electronic patient records (EPR) have to meet this paradigm supporting communication and interoperation between the health care establishments (HCE) and health professionals (HP) involved. Due to the sensitivity of personal medical information, this co-operation must be provided in a trustworthy way.

Blobel, B. (2006). Advanced EHR architectures - Promises or reality.

Objectives. Forming the informational reflection of the patients and their care, the Electronic Health Record (EHR) is the core application of any complex health information system or health network. Such an ideally lifelong history file must be reliable, flexible, adaptable to new concepts and technologies, and robust, to allow for sharing knowledge over its lifetime. A sophisticated architecture must be chosen for meeting this challenge.

de Keizer, N. F., A. Abu-Hanna, et al. (1999). "Analysis and design of an ontology for intensive care diagnoses." Methods of Information in Medicine 38(2): 102-112.

Information about the patient's health status and about medical problems in general, play an important role in stratifying a patient population for quality assurance of intensive care. A terminological system which supports both the description of health problems for daily care practice and the aggregation of