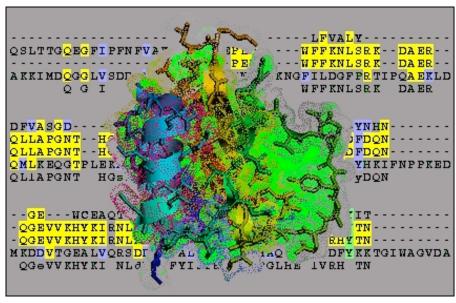
An Orientation to the *Biology Workbench*

The Biology Workbench is a web-based tool for biologists. The Workbench allows biologists to search many popular protein and nucleic acid sequence databases. Database searching is integrated with access to a wide variety of analysis and modeling tools, all within a point and click interface that eliminates file format compatibility problems.

Available online at: http://workbench.sdsc.edu



About this document		p. 2
About the Biology Workbench		p. 2
Getting started	Software / hardware requirements	p. 3
	User accounts	p. 4
An introduction to the interface	Toolsets	p. 5
	Controlling the Biology Workbench	p. 7
Examples	Example 1: Text search of a database	p. 7
	Example 2: Multiple Sequence Alignment	p. 9
Tips and tricks	Controlling a supercomputer from a browser	p. 12
	Printing	p. 12
	More information about tools and databases	p. 12
More information	Biology Student Workbench	p. 13
	BEDROCK	p. 13
	The Biology Workbench FAQ	p. 14
	References	p. 14
Appendix A - Tools		p. 15
Appendix B - Databases & Genomes		p. 19

About this document

This document is designed to provide a quick introduction to the Biology Workbench including an overview of how it is organized, an introduction to some of the resources it includes, and hints to make it easier to use. This overview is not designed to provide a technical introduction to the operation of the Biology Workbench or provide documentation for the various analysis tools and data resources collected in the Biology Workbench. For more information on the Biology Workbench project and the resources available within the Biology Workbench please see the references section, appendices and the Biology Workbench website.

Computer tools will do whatever you ask them to and it is up to the operator to infer biological meaning from the computational results. Beginning biologists can gain many insights using the Biology Workbench and I have found it to be a very rewarding way to engage students with research data and tools. Biological concepts such as the distinction between the measurement of genetic similarity and the inference of evolutionary homology can be explored using the Biology Workbench but the tools themselves do not teach these concepts. Carefully constructed curricular materials are important to make students' experience with the Biology Workbench meaningful and stimulating.

About the Biology Workbench

The Biology Workbench is a free, web accessible suite of resources for working with molecular sequence and structure data. The idea of a "workbench" is to provide a single place to go for all the stuff—both raw materials and tools—you might need to do your work. Toward that end, the Biology Workbench provides users with access to many molecular data bases (the raw materials) and analysis programs (the tools) for managing and analyzing molecular data.

Three important features of the Biology Workbench make it particularly useful to researchers and students alike.

- 1. It provides a unified interface to access a variety of tools and databases.
- 2. It is online and can be accessed with a standard web browser.
- 3. It is available at no cost to academic users.

The Biology Workbench was developed by Shankar Subramaniam and currently resides on a supercomputer at the San Diego Supercomputing Center. Thus, it provides access to both extensive computational power and huge data storage capacity. By providing a unified interface the Biology Workbench makes it easier to pull data from several sources and use a series of analysis tools without investing time in systems administration and data conversions. Because the Biology Workbench is accessed with a standard web browser it can be used with almost any personal computer that is connected to the Internet. The Biology Workbench also allows you to maintain user accounts. The user accounts make it possible to save your work online and access it later from any networked computer.

Most of the tools and databases available through the Biology Workbench are public domain resources that are available individually elsewhere. The power of the Workbench architecture is that it collects these distributed resources and provides a unified interface to access them. The Biology Workbench also facilitates communication between the tools and data sources by doing a variety of data format conversions so that they can "talk" to one another. Much of this activity

is hidden from the user behind the web interface making it much easier to focus on the research you are pursuing (Subramaniam, 1998).

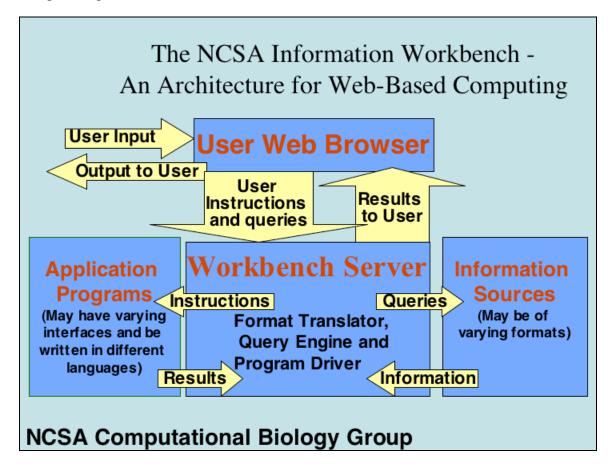


Figure 1: The architecture of the information workbench environment (developed by the NCSA computational biology group).

Getting started

Software / hardware requirements

The following information was provided at the Biology Workbench homepage in March 2003. For updates please check http://workbench.sdsc.edu.

Suggested Web Browser: we suggest that people use Netscape Communicator or Navigator, preferably the latest version that can be obtained (any 4.x version should work well). Microsoft Internet Explorer (especially older versions) can be unpredictable when loading the Biology Workbench. In addition, because we are unable to force Internet Explorer to open secondary windows with our software, showing database records and reading help pages can be a bit clumsy. Nonetheless, most Biology Workbench operations *should* work within Internet Explorer.

Some people notice browser-related problems that go away when one clears the disk cache, and turning off the disk cache altogether when using the Biology Workbench might be a good idea. Also, your memory cache should be set as high as comfortable, as some of our pages can take up quite a bit of space in your browser. We suggest a minimum value of 10 megabytes for your memory cache, if possible.

Structure Viewing: PDB structures can be viewed for PDBF inder records that are returned from a database search. One way to do this is to use the Rasmol program. The Chime plugin is another option for viewing structures on Windows and Macintosh machines, and we may eventually provide a Java-based structure viewer. For molecules with PDB structures, we also provide links to the PDB Structure Explorer page for that particular molecule, and to the Protein Explorer display for that particular molecule.

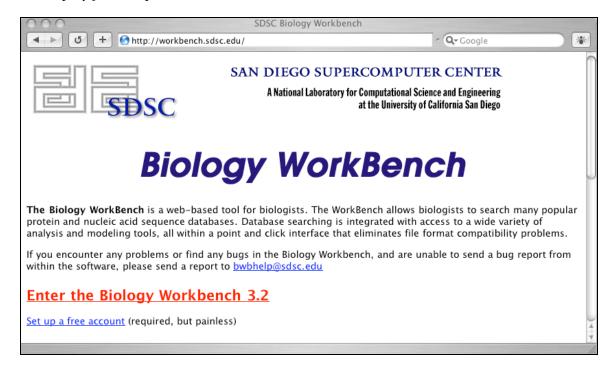


Figure 2: The top portion of the home page for the Biology Workbench.

User accounts

Access to the Biology Workbench requires a user account. While this may seem like an inconvenience to setup an account it is actually one of the most useful features. A Biology Workbench account allows you to store sequence data on a remote server meaning that you can access your account from any networked computer and continue your research without reload your data files.

Follow the "Setup a free account" link to register as a user. After registering and logging in you are ready to get started using the Biology Workbench.

Full Name:	
eMail Address:	
User ID: Single word (alphanumeric characters only)	
Password:	
Password: (again)	
Register	

Figure 3: The registration form for opening a new Biology Workbench account.

An introduction to the interface

Using the Biology Workbench is fairly simple and does not require any specialized technical skills. This introduction focuses on orienting the reader to how the interface is organized by providing a general overview of how to find and use different tools.

Warning: The simplicity of manipulating the Biology Workbench interface should not be confused with the tasks of interpreting the biological or statistical significance of results. As a collection of algorithms the parts of the Biology Workbench will behave in fairly predictable ways. Making inferences from the results that are biologically meaningful is another matter altogether.

Toolsets

Once you are logged on to the Biology Workbench you will see that the resources are organized into five main categories which I will call "tool sets." Each is a collection of functions/scripts/programs for working with a partic

ular type of data.

Note: the first screen you see when you log in has a large image and various announcements toward the top of the page. In order to get to the tools you will need to scroll down until you see the following button options.

Session Tools: tools to manage sessions—which are like folders or directories, different places to store work.

Nucleic Tools: tools for working with nucleic acid data.

Protein Tools: tools for working with amino acid data.

Alignment Tools: tools for working with sequence alignments.

Structure Tools: these are not really ready for use and will not be discussed further here.



Figure 4: The buttons that allow you to move between tool sets within the Biology Workbench. Note that you can also change the background color before you enter a tool set.

Within each of the tool sets you will see the same basic interface. From the top to the bottom there are the following features:

- a set of buttons that allows you to switch tool sets. Note that the current tool set is highlighted.
- the session label. If you do not create a new session or resume an old one the Biology Workbench will save your work into a "Default Session".
- the list of available tools. This list varies depending on the tool set you are in. For a complete list of tools see Appendix A.
- the control buttons. These allow you to initiate various actions.
- any data that you have in that toolset. Note that data storage is tool sets and session specific. In other words, it is stored in a particular tool set of a particular session, meaning that is the only place where you will have access to it.

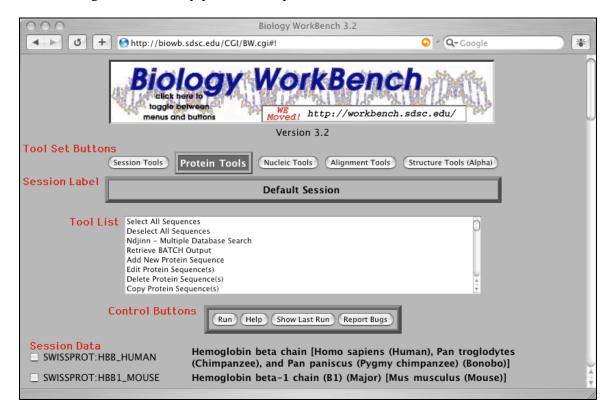


Figure 5: The Biology Workbench interface with highlighted features (red labels).

Controlling the Biology Workbench

The basic operation of any Biology Workbench tool involves the following steps:

- 1. selecting the tool it will be highlighted in the tool list
- 2. selecting data may or may not be necessary based on the tool selected
- 3. clicking the Run button

Most tools will then present a parameter page that will allow you to make some tool specific choices about how the analysis is done and how the results are displayed. Once you hit submit on the parameter page the Biology Workbench will return your results or carry out the function you requested.

Example 1: Text search of a database

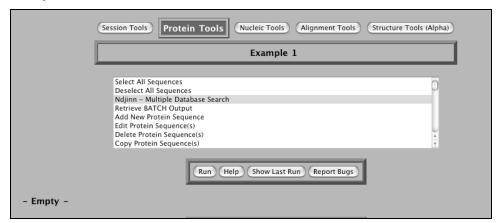
This example will walk you though the steps involved in doing a text string search of a sequence database.

It will involve:

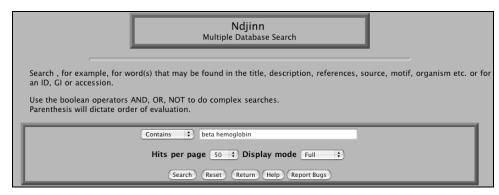
- running the Ndjinn tool
- selecting a database to search
- viewing results
- importing the results into your session

Steps:

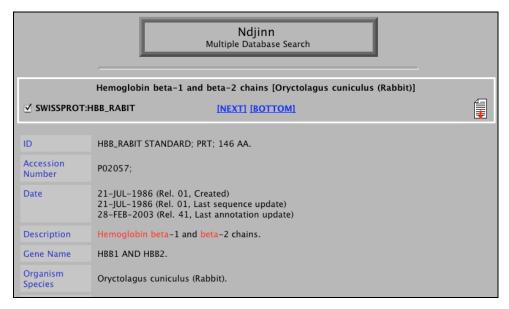
- 1. Move to the Protein tool set by clicking on the Protein Tools button.
- 2. Select the Ndjinn Multiple Database Search tool and click run there is no need to select any data with this tool.



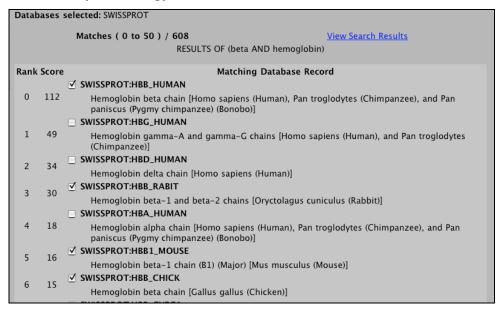
3. On the next screen enter the text string that you would like to search for and select the databases that you would like to search. Scroll down to see the databases and genomes. In this example I have searched for "beta hemoglobin" in the Swissprot database and had 50 hits per page displayed.



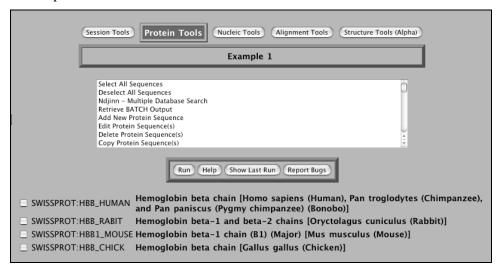
4. Select a sequence record from the results page and view the record by clicking the show record button.



5. When you have found the records you want, select them and use the import button to save them into your Biology Workbench session.



6. The data files should now appear in your session and you are ready to move on to the next example.



Example 2: Multiple Sequence Alignment

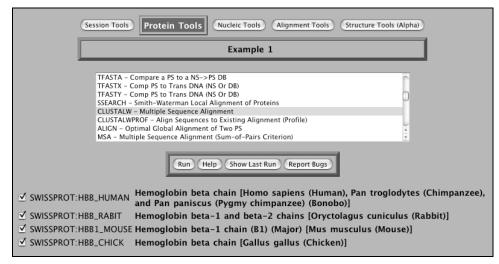
This example will step you through the steps of aligning a set of sequences. You can work with any sequences you collected from Example 1.

It will involve:

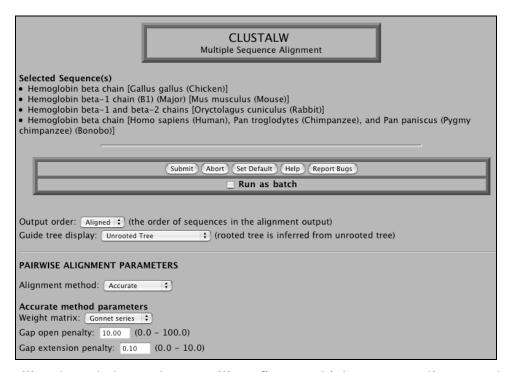
- Running the ClustalW tool
- Importing an alignment
- Viewing an alignment in Boxshade

Steps:

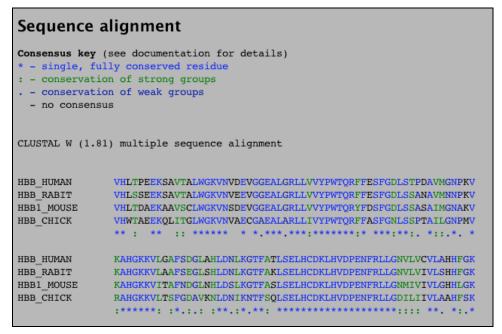
1. To start this exercise you will need some sequence data in your session. Select the ClustalW tool and several sequences to analyze.

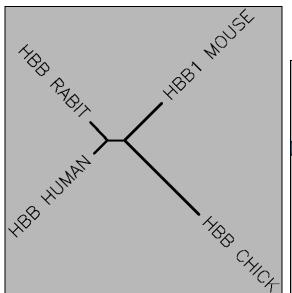


2. The ClustalW parameters page will allow you control many aspects of how the analysis is run and how the results are displayed. See the help file for more details on these parameters. In this example the analysis was run with the default settings.



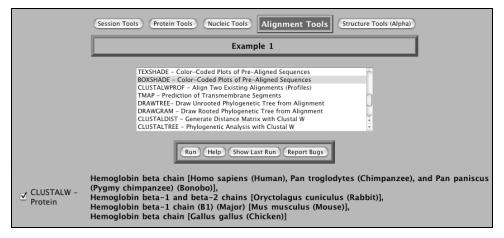
3. Scrolling through the results you will see first a multiple sequence alignment, then a distance based tree, followed by a list of the pair wise alignment scores.



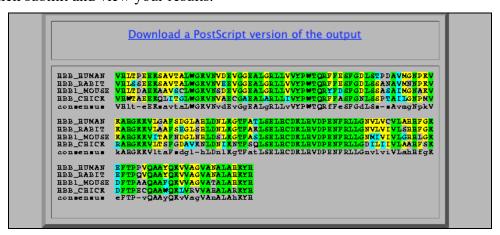


```
Sequence type explicitly set to Protein
Sequence format is Pearson
Sequence 1: HBB HUMAN
                            146 aa
Sequence 2: HBB RABIT
                           146 aa
Sequence 3: HBB1 MOUSE
                           146 aa
Sequence 4: HBB CHICK
                           146 aa
Start of Pairwise alignments
Aligning ...
Sequences (1:2) Aligned. Score:
Sequences (1:3) Aligned. Score:
Sequences (1:4) Aligned. Score: 69
Sequences (2:3) Aligned. Score: 79
Sequences (2:4) Aligned. Score: 67
Sequences (3:4) Aligned. Score: 65
Time for pairwise alignment: 0.019447
```

4. Import your alignment by clicking the import button. You will automatically be moved to the alignment tool set. Select the alignment and the Boxshade tool.



- 5. Again, the parameters page will provide you with a variety of ways to control the analysis and output.
- 6. Click submit and view your results.



Tips and tricks

Controlling a supercomputer from a browser

Navigating: It is best <u>not</u> to use the browser's navigation buttons to move around in the Biology Workbench. The browser buttons call up cached pages that are held on your local drive and may not reflect changes that have been made on the server side—where the work is actually getting done. It is possible to lose data and otherwise corrupt your session by using the browser's navigation buttons. Use the buttons provided by the Biology Workbench interface to navigate (e.g., "return" will take you back to the tool set window).

New windows: In some browsers the Biology Workbench will generate new browser windows to display analysis results, help pages and other information. Because the window is new you can not use the browser's back button to return to the previous screen and there is no link in the window to take you back. In these situations the best thing to do is either close the new window or select the old one to bring it to the front.

Waiting: Under a variety of conditions you may need to wait patiently for 15 to 30 seconds for the browser display to refresh. This generally occurs while some computations are being done on the server or data is being uploaded to your session. Usually the mouse clicks will be ignored during this time but lots of fiddling around could cause the browser to hang.

Printing

Within the browser: Because everything is presented as a web page it can easily be printed from the browser. Some files are very long so it may be best to first use the print preview function to identify the pages of the document you want to print.

Post Script files: Some figures are downloadable as Post Script files. These are handy because you can then scale the document and maintain a high resolution image. You will need a Post Script viewer to display the files. GSView (PC) and MacGSView (Mac) can be downloaded from http://www.cs.wisc.edu/~ghost/.

Screen captures: Another option is to capture part a screen or copy and past some text into a word processor.

More information about tools and databases

Help files: Many of the tools in the Biology Workbench have help files that provide detailed information about how the tool works. To get the help file highlight the tool you are interested in and then click the help button. Appendix A contains a complete list of the available tools.

Web sites: Most of the tools and databases available in the Biology Workbench are also available elsewhere on the Internet. A Google search on the tool name, database name or author usually makes it possible to find the stand-alone version of the resource and additional documentation. Published references are usually listed at the end of each help file.

Appendices: See the Appendices below for list of the tools and databases available in the Biology Workbench as of March 2003.

Switching between tool list interfaces

The tools in any tool set can be displayed in 3 different ways (buttons, scrolling list, pull-down menu). The screen shots in this document generally use the scrolling list option and this is the default when you first enter the Biology Workbench. The different tool list interfaces are shown in the figure below. The other tool list formats (buttons and pull-down menu) appear under your data on the page so that if you have a long list of sequences you will need to scroll down to find the tools. To toggle between the tool list views click the area highlighted in the figure below.

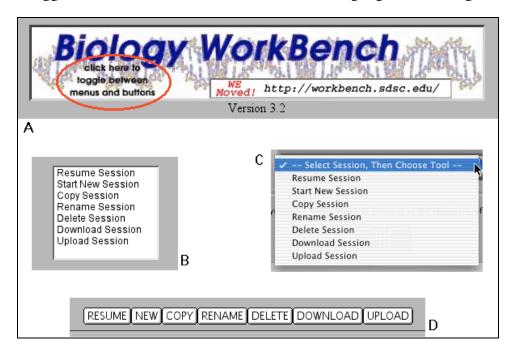


Figure 6: Toggling between tool list interfaces. (A) The screen area to click to toggle the tool list interfaces. (B) The scrolling list interface. (C) The pull down menu interface. (D) The button interface.

More information

Biology Student Workbench

There are a variety of tutorials written for using the Biology Workbench. Many of them were developed as part of the Biology Student Workbench Project and can be found at http://peptide.ncsa.uiuc.edu/. There is also a student interface to the Biology Workbench available at this site.

BEDROCK

The BEDROCK project is a NSF funded bioinformatics education project. The project focuses on faculty development and the national dissemination of resources supporting the integration of bioinformatics across the undergraduate biology curriculum. Many materials that take advantage of the Biology Workbench are available as part of the BEDROCK project http://bioquest.org/bioinforamtics.

The Biology Workbench FAQ

The Biology Workbench maintains a list of frequently asked questions at the following web site http://workbench.sdsc.edu/faq.html

References

- Subramaniam, S. (1998). The Biology Workbench A seamless database and analysis environment for the biologist. Proteins: Structure, Function, and Genetics, 32, 1-2. http://www3.interscience.wiley.com/cgi-bin/abstract/36266/START
- Unwin, R., Fenton, J., Whitsitt, M., Jamison, C., Stupar, M., Jakobsson, E., Subramaniam, S. (1998). Biology Workbench: A WWW-based virtual computing and analysis environment for the biological sciences. *In* Bioinformatics: Databases and Systems. S. Letovsky (Ed.) p. 233-244. http://bsw.ncsa.uiuc.edu/Papers/
- Cotter, Dawn (1999). About the Biology Workbench. BioQUEST Notes. Volume 9 number 3 p. 8-10. http://bioquest.org/indexbqnotes.html
- Jakobsson, Eric (2000). Portals to Knowledge: Information Technology, Research, and Training. *In* Graduate Education in the Chemical Sciences: Issues for the 21st Century: Report of a Workshop, pp. 56-72. National Academy Press http://books.nap.edu/books/0309071305/html/56.html

Appendix A - Tools

This appendix contains a list of the tools available in the Biology Workbench as of March 2003. Because the Biology Workbench is a flexible architecture tools can be added and deleted over time making this list a snapshot in time. The tool lists are organized, as they are in the Biology Workbench, into tool sets. Many of these tools have help files associated with them. You can view the help files by selecting the tool and then clicking the help button in the Biology Workbench.

The structure tools are not listed here and are not recommended.

Protein Tools

Select All Sequences

Deselect All Sequences

Ndjinn - Multiple Database Search

Retrieve BATCH Output

Add New Protein Sequence

Edit Protein Sequence(s)

Delete Protein Sequence(s)

Copy Protein Sequence(s)

View Protein Sequence(s)

Download Protein Sequence(s)

View Database Records of Imported Sequences

View Available Scoring Matrices

BL2SEQ - Compare proteins to each other with BLAST

BL2SEQX - Compare proteins to nucleotide sequences with BLAST

BLASTP - Compare a PS to a PS DB

TBLASTN - Compare a PS to a translated DB

PSIBLAST - Position Specific Iterative BLAST

FASTA - Heuristic Sequence Similarity Search (PS Or DB)

TFASTA - Compare a PS to a NS -> PS DB

TFASTX - Comp PS to Trans DNA (NS Or DB)

TFASTY - Comp PS to Trans DNA (NS Or DB)

SSEARCH - Smith-Waterman Local Alignment of Proteins

CLUSTALW - Multiple Sequence Alignment

CLUSTALPROF - Align Sequences to Existing Alignment (Profile)

Orientation to the Biology Workbench

ALIGN - Optimal Global Alignment of Two PS

MSA - Multiple Sequence Alignment (Sum-of-Pairs Criterion)

LALIGN - Calculate N-Best Local PS Alignments

LFASTA - Local Alignment of Two PS

ROBUST - Global alignment of Two PS (Show Robust Pairs)

SIM - N-Best Local Similarities Using Affine Weights

BESTSCOR - Calculate the Best Self-Comparison Score

CTREE - Align protein sequences with confidence estimates

PRSS - Compare a PS to a Shuffled PS

SAPS - Statistical Analysis of PS

AASTATS - Statistics Based on Amino Acid Abundance, including weight and specific volume

GREASE - Kyte-Doolittle Hydropathy Profile

RPBLAST - Compare a PS to a Conserved Domain DB

FingerPRINTScan - PRINTS fingerprint identification

PROSEARCH - Search Prosite DB for Patterns in a PS

PPSEARCH - Search Prosite DB for Patterns in a PS

PFSCAN - Sequence Search Against a Set of Profiles (PROSITE and PFAM)

HMMPFAM - Search against Pfam HMM database

BLIMPS - Sequence Search Against a Set of Profiles (BLOCKS)

PATTERNMATCHDB - Search for Regular Expressions (Patterns) in a protein sequence DB

PATTERNMATCH - Search for Regular Expressions (Patterns) in a protein sequence

GOR4 - Predict Secondary Structure of PS

RANDSEQ - Randomize a Sequence

CHOFAS - Predict Secondary Structure of PS(s) (Chou-Fasman)

HTH - Predict HTH Motifs in Protein Chains

PELE - Protein Structure Prediction

DSSP - Secondary Structure/Solvent Exposure of PDB Proteins

TMAP - Predication of Transmembrane Segments

TMHHMM - Predict location of transmembrane helices and location of intervening loop regions*

EXTCOEF - Extinction coefficient calculation

PI - Isoelectric point determination

Nucleic Tools

Select All Sequences

Deselect All Sequences

Ndjinn - Multiple Database Search

Retrieve BATCH Output

Add New Nucleic Sequence

Edit Nucleic Sequence(s)

Delete Nucleic Sequence(s)

Copy Nucleic Sequence(s)

View Nucleic Sequence(s)

Download Nucleic Sequence(s)

View Database Records of Imported Sequences

BL2SEQ - Compare nucleotides to a reference nucleotide with BLAST

BL2SEQX - Compare nucleotides protein sequences with BLAST

BLASTN- Compare a NS to a NS DB

BLASTX - Compare a PS-Derived-from-NS to a PS DB

TBLASTX - Compare a translated NS to a translated DB

FASTA - Nucleic Acid Sequence Comparisons (NS or DB)

FASTX - Compare Translated NS to PS DB

FASTY - Compare Translated NS to PS DB

SSEARCH - Smith-Waterman Local Alignment

CLUSTALW - Multiple Sequence Alignment

CLUSTALPROF - Align Sequences to Existing Alignment (Profile)

ALIGN - Optimal Global Sequence Alignment

LALIGN - Calculate Optimal Local Sequence Alignments

LFASTA - Calculate Local Sequence Alignments (Heuristic)

PATTERNMATCHDB - Search for Regular Expressions (Patterns) in a nucleic sequence DB

PATTERNMATCH - Search for Regular Expressions (Patterns) in a nucleic sequence

TACG - Analyze a NS for Restriction Enzyme Sites

PRIMER3 - Design Primer Pairs and Probes

NASTATS - Nucleic Acid Statistics

BESTSCOR - Calculate the Best Self-Comparison Score

PFSCAN - Sequence Search Against a Set of Profiles (PROSITE)

PRIMERCHECK - Calculates melting point, length, %GC for a primer sequence

PRIMERTM - Designs end primers based on a minimum Tm

SIXFRAME - Generate & Import 6 Frame Translations on a NS

REVCOMP - Generate Reverse Complement of NS

RANDSEQ - Randomize a Sequence

Alignment Tools

Select All Alignments

Deselect All Alignments

Ndjinn - Multiple Database Search (pronounced "engine")

Add New Aligned Sequence

Retrieve BATCH Output

Edit Aligned Sequence(s)

Delete Aligned Sequence(s)

Copy Aligned Sequence(s)

View Aligned Sequence(s)

Split Alignment Into Component Sequences

Split Alignment Into Component Sequences and Remove Gap Characters (-)

Download Aligned Sequence(s)

TEXTSHADE - Color-Coded Plots of Pre-Aligned Sequences

BOXSHADE - Color-Coded Plots of Pre-Aligned Sequences

CLUSTALPROF - Align Two Existing Alignments (Profiles)

TMAP - Prediction of Transmembrane Segments

DRAWTREE- Draw Unrooted Phylogenetic Tree from Alignment

DRAWGRAM- Draw Rooted Phylogenetic Tree from Alignment

CLUSTALDIST- Generate Distance Matrix with Clustal W

CLUSTALTREE- Phylogenetic Analysis with Clustal W

DNADIST - Compute Evolutionary Distance Matrix from NS Alignment

PROTDIST - Compute Evolutionary Distance Matrix from PS Alignment

DNAPARS - Infer an Unrooted Phylogeny from NS Alignment

PROTPARS - Infer an Unrooted Phylogeny from PS Alignment

MVIEW - Multiple Alignment Display

Appendix B - Database and Genomes

This appendix contains lists of the databases and genomes available in the Biology Workbench as of March 2003. Because the Biology Workbench is a flexible architecture databases and genomes can be added and deleted over time making these lists snapshots in time. The 72 databases are listed first followed by the list of 159 genomes.

Databases

AAINDEX_1	Amino Acid index database Amino Acid Indices
AAINDEX_2	Amino Acid index database Amino Acid Mutation Matrices
AFCS	Alliance for Cellular Signaling Database
BLOCKS	Multiple alignments of conserved regions in protein families
DBCAT	DBCat Public Catalog of Databases
ECDC	E.Coli Database Collection
ENZYME	Repository of information relative to the nomenclature of enzymes
EPD	Eukaryotic Promoter Database
FLYBASE	A comprehensive database for information on the genetics and molecular biology of Drosophila
GBBCT	Genbank Bacterial Sequences
GBEST1	GenBank Expressed Sequence Tags Part 1
GBEST2	GenBank Expressed Sequence Tags Part 2
GBEST3	GenBank Expressed Sequence Tags Part 3
GBEST4	GenBank Expressed Sequence Tags Part 4
GBEST5	GenBank Expressed Sequence Tags Part 5
GBEST6	GenBank Expressed Sequence Tags Part 6
GBEST7	GenBank Expressed Sequence Tags Part 7
GBEST8	GenBank Expressed Sequence Tags Part 8
GBEST9	GenBank Expressed Sequence Tags Part 9
GBGSS1	GenBank Genome Survey Sequence entries Part 1
GBGSS2	GenBank Genome Survey Sequence entries Part 2
GBGSS3	GenBank Genome Survey Sequence entries Part 3
GBHTC	GenBank High Throughput cDNA Sequencing Entries

GBHTG1	GenBank High Throughput Genomic Sequencing Entries Part 1
GBHTG2	GenBank High Throughput Genomic Sequencing Entries Part 2
GBINV	GenBank Invertebrate Sequences
GBMAM	GenBank Mamalian Sequences
GBNEW	GenBank Updates
GBPAT	GenBank Patent Sequences
GBPHG	GenBank Phage Sequences
GBPLN	GenBank Plant Sequences incl. fungi and algae
GBPRI	GenBank Primate Sequences
GBREFSEQ	Genbank Refseq
GBROD	GenBank Rodent Sequences
GBSTS	GenBank Sequence Tagged Site Entries
GBSYN	GenBank Synthetic and Chimeric Sequences
GBUNA	GenBank Unannotated Sequence Entries
GBVRL	GenBank Viral Sequences
GBVRT	GenBank Misc. Vertebrate Sequences
GENPEPT	GenBank Gene Products Last Full Release
GENPEPTNEW	GenBank Gene Products Updates
KEGG_GENES	KEGG Genes Database
KEGG_ORTHOLOG	KEGG Ortholog Database
KEGG_PATHWAY	KEGG Pathway Database
LIGAND_COMPOUND	Ligand Compound Database
LIGAND_ENZYME	Ligand Enzyme Database
МНСРЕР	A database of MHC binding peptides
NRL3D	Sequence-structure database derived from the 3-dimensional structure of proteins deposited with the Protein Data Bank
OMIM	Online Mendelian Inheritance in Man
PDBFINDER	Sequences Derived from the PDB, DSSP, and HSSP databases
PDBSEQRES	Sequences derived from PDB and PDBFINDER
PFAM	PFAM A database of protein domain family alignments and
·	

	HMMs
PFAM_B	PFAM-B A database of protein domain family alignments and HMMs, unannotated
PIR	Protein Information Resource, All Entries - National Biomedical Research Foundation
PRINTS	Protein Motif Fingerprint Database
PRODOM	The Protein Domain Database
PROSITE	Prosite Data
PROSITE_DOC	Prosite Documentation
PROSITE_QDOC	Prosite Prerelease Documentation
REBASE	The Restriction Enzyme Database
REBASE_ABS	The Restriction Enzyme Database Abstracts
REBASE_SRC	Restriction Enzymes And Methylases Commercially Available
SDSCNR	Non-Redundant sequence database made from all protein databases within the Biology Workbench
SEQANALREF	Sequence Analysis Bibliographic Reference Data Bank
SWISSPROT	Swissprot Database
TRANSFAC_CELL	Transcription Factor Database CELL component: Explanations of Cellular Sources that Interact with Sites in Transfac SITE
TRANSFAC_CLASS	Transcription Factor Database CLASS component: Explanations of DNA Binding Domains of Transcription Factor Classes
TRANSFAC_FACTOR	Transcription Factor Database FACTOR Component: Eukaryotic cis-acting regulatory DNA elements and transacting factors
TRANSFAC_GENE	Transcription Factor Database GENE component
TRANSFAC_SITE	Transcription Factor Database SITE component: Putative regulatory protein binding sites
TrEMBL	Computer-annotated protein sequence database supplementing the SWISS-PROT Protein Sequence Data Bank
TrEMBLnew	Computer-annotated protein sequence database supplementing the SWISS-PROT Protein Sequence Data Bank Updates

Genomes

A_thaliana	Genbank Arabidopsis thaliana genome
A_thaliana_gene	Genbank Arabidopsis thaliana CDS and proteins
Aaeo	Genbank Aquifex aeolicus genome
Aaeo_gene	Genbank Aquifex aeolicus CDS and proteins
Aper	Genbank Aeropyrum pernix genome
Aper_gene	Genbank Aeropyrum pernix CDS and proteins
AtumC58	Genbank Agrobacterium tumefaciens C58 genome
AtumC58UWash	Genbank Agrobacterium tumefaciens C58 UWash genome
AtumC58UWash_gene	Genbank Agrobacterium tumefaciens C58 UWash CDS and proteins
AtumC58_gene	Genbank Agrobacterium tumefaciens C58 CDS and proteins
Bbur	Genbank Borrelia burgdorferi genome
Bbur_gene	Genbank Borrelia burgdorferi CDS and proteins
Bhal	Genbank Bacillus halodurans genome
Bhal_gene	Genbank Bacillus halodurans CDS and proteins
Bmel	Genbank Brucella melitensis genome
Bmel_gene	Genbank Brucella melitensis CDS and proteins
Bsub	Genbank Bacillus subtilis genome
Bsub_gene	Genbank Bacillus subtilis CDS and proteins
Buch	Genbank Buchnera sp genome
Buch_gene	Genbank Buchnera sp CDS and proteins
C_elegans	Genbank C elegans genome
C_elegans_gene	Genbank C elegans CDS and proteins
Cace	Genbank Clostridium acetobutylicum genome
Cace_gene	Genbank Clostridium acetobutylicum CDS and proteins
Ccre	Genbank Caulobacter crescentus genome
Ccre_gene	Genbank Caulobacter crescentus CDS and proteins
Cglu	Genbank Corynebacterium glutamicum genome

Orientation to the Biology Workbench

Cglu_gene	Genbank Corynebacterium glutamicum CDS and proteins
Cjej	Genbank Campylobacter jejuni genome
Cjej_gene	Genbank Campylobacter jejuni CDS and proteins
Cmur	Genbank Chlamydia muridarum genome
Cmur_gene	Genbank Chlamydia muridarum CDS and proteins
Cper	Genbank Clostridium perfringens genome
Cper_gene	Genbank Clostridium perfringens CDS and proteins
CpneAR39	Genbank Chlamydophila pneumoniae AR39 genome
CpneAR39_gene	Genbank Chlamydophila pneumoniae AR39 CDS and proteins
CpneCWL029	Genbank Chlamydophila pneumoniae CWL029 genome
CpneCWL029_gene	Genbank Chlamydophila pneumoniae CWL029 CDS and proteins
CpneJ138	Genbank Chlamydophila pneumoniae J138 genome
CpneJ138_gene	Genbank Chlamydophila pneumoniae J138 CDS and proteins
Ctra	Genbank Chlamydia trachomatis genome
Ctra_gene	Genbank Chlamydia trachomatis CDS and proteins
DICTY	SDSC Dictyostelium discoideum CDS and proteins [as of now, only searchable by orf ID]
Drad	Genbank Deinococcus radiodurans genome
Drad_gene	Genbank Deinococcus radiodurans CDS and proteins
E_cuniculi	Genbank Encephalitozoon cuniculi genome
E_cuniculi_gene	Genbank Encephalitozoon cuniculi CDS and proteins
EcoliK12	Genbank Escherichia coli K12 genome
EcoliK12_gene	Genbank Escherichia coli K12 CDS and proteins
EcoliO157H7	Genbank Escherichia coli O157H7 genome
EcoliO157H7EDL933	Genbank Escherichia coli O157H7 EDL933 genome
EcoliO157H7EDL933_gene	Genbank Escherichia coli O157H7 EDL933 CDS and proteins
EcoliO157H7_gene	Genbank Escherichia coli O157H7 CDS and proteins
Fnuc	Genbank Fusobacterium nucleatum genome
Fnuc_gene	Genbank Fusobacterium nucleatum CDS and proteins
H_sapiens	Genbank H sapiens genome

H_sapiens_mRNA	Genbank H sapiens CDS and proteins
Hals	Genbank Halobacterium sp genome
Hals_gene	Genbank Halobacterium sp CDS and proteins
Hinf	Genbank Haemophilus influenzae genome
Hinf_gene	Genbank Haemophilus influenzae CDS and proteins
Hpyl26695	Genbank Helicobacter pylori 26695 genome
Hpyl26695_gene	Genbank Helicobacter pylori 26695 CDS and proteins
HpylJ99	Genbank Helicobacter pylori J99 genome
HpylJ99_gene	Genbank Helicobacter pylori J99 CDS and proteins
Linn	Genbank Listeria innocua genome
Linn_gene	Genbank Listeria innocua CDS and proteins
Llac	Genbank Lactococcus lactis genome
Llac_gene	Genbank Lactococcus lactis CDS and proteins
Lmon	Genbank Listeria monocytogenes genome
Lmon_gene	Genbank Listeria monocytogenes CDS and proteins
M_musculus	Genbank M musculus genome
M_musculus_mRNA	Genbank M musculus CDS and proteins model reference sequences only
Mgen	Genbank Mycoplasma genitalium genome
Mgen_gene	Genbank Mycoplasma genitalium CDS and proteins
Mjan	Genbank Methanococcus jannaschii genome
Mjan_gene	Genbank Methanococcus jannaschii CDS and proteins
Mlep	Genbank Mycobacterium leprae genome
Mlep_gene	Genbank Mycobacterium leprae CDS and proteins
Mlot	Genbank Mesorhizobium loti genome
Mlot_gene	Genbank Mesorhizobium loti CDS and proteins
Mpne	Genbank Mycoplasma pneumoniae genome
Mpne_gene	Genbank Mycoplasma pneumoniae CDS and proteins
Mpul	Genbank Mycoplasma pulmonis genome
Mpul_gene	Genbank Mycoplasma pulmonis CDS and proteins
Mthe	Genbank Methanobacterium thermoautotrophicum genome
Mthe_gene	Genbank Methanobacterium thermoautotrophicum CDS and
t	· · · · · · · · · · · · · · · · · · ·

	proteins
MtubCDC1551	Genbank Mycobacterium tuberculosis CDC1551 genome
MtubCDC1551_gene	Genbank Mycobacterium tuberculosis CDC1551 CDS and proteins
MtubH37Rv	Genbank Mycobacterium tuberculosis H37Rv genome
MtubH37Rv_gene	Genbank Mycobacterium tuberculosis H37Rv CDS and proteins
NmenMC58	Genbank Neisseria meningitidis MC58 genome
NmenMC58_gene	Genbank Neisseria meningitidis MC58 CDS and proteins
NmenZ2491	Genbank Neisseria meningitidis Z2491 genome
NmenZ2491_gene	Genbank Neisseria meningitidis Z2491 CDS and proteins
Nost	Genbank Nostoc sp genome
Nost_gene	Genbank Nostoc sp CDS and proteins
P_falciparum	Genbank P falciparum genome
P_falciparum_gene	Genbank P falciparum CDS and proteins
Paby	Genbank Pyrococcus abyssi genome
Paby_gene	Genbank Pyrococcus abyssi CDS and proteins
Paer	Genbank Pseudomonas aeruginosa genome
Paer_gene	Genbank Pseudomonas aeruginosa CDS and proteins
Pfur	Genbank Pyrococcus furiosus genome
Pfur_gene	Genbank Pyrococcus furiosus CDS and proteins
Phor	Genbank Pyrococcus horikoshii genome
Phor_gene	Genbank Pyrococcus horikoshii CDS and proteins
Pmul	Genbank Pasteurella multocida genome
Pmul_gene	Genbank Pasteurella multocida CDS and proteins
Pyae	Genbank Pyrobaculum aerophilum genome
Pyae_gene	Genbank Pyrobaculum aerophilum CDS and proteins
Rcon	Genbank Rickettsia conorii genome
Rcon_gene	Genbank Rickettsia conorii CDS and proteins
Rpro	Genbank Rickettsia prowazekii genome
Rpro_gene	Genbank Rickettsia prowazekii CDS and proteins
Rsol	Genbank Ralstonia solanacearum genome

Rsol_gene	Genbank Ralstonia solanacearum CDS and proteins
S_cerevisiae	Genbank S cerevisiae genome
S_cerevisiae_gene	Genbank S cerevisiae CDS and proteins
S_pombe	Genbank S pombe genome
S_pombe_gene	Genbank S pombe CDS and proteins
Salt	Genbank Salmonella typhi genome
SaltLT2	Genbank Salmonella typhimurium LT2 genome
SaltLT2_gene	Genbank Salmonella typhimurium LT2 CDS and proteins
Salt_gene	Genbank Salmonella typhi CDS and proteins
Smel	Genbank Sinorhizobium meliloti genome
Smel_gene	Genbank Sinorhizobium meliloti CDS and proteins
Ssol	Genbank Sulfolobus solfataricus genome
Ssol_gene	Genbank Sulfolobus solfataricus CDS and proteins
StaurMu50	Genbank Staphylococcus aureus Mu50 genome
StaurMu50_gene	Genbank Staphylococcus aureus Mu50 CDS and proteins
StaurN315	Genbank Staphylococcus aureus N315 genome
StaurN315_gene	Genbank Staphylococcus aureus N315 CDS and proteins
Stok	Genbank Sulfolobus tokodaii genome
Stok_gene	Genbank Sulfolobus tokodaii CDS and proteins
StpneR6	Genbank Streptococcus pneumoniae R6 genome
StpneR6_gene	Genbank Streptococcus pneumoniae R6 CDS and proteins
StpneTIGR4	Genbank Streptococcus pneumoniae TIGR4 genome
StpneTIGR4_gene	Genbank Streptococcus pneumoniae TIGR4 CDS and proteins
Stpyo	Genbank Streptococcus pyogenes genome
Stpyo_gene	Genbank Streptococcus pyogenes CDS and proteins
SynP	Genbank Synechocystis PCC6803 genome
SynP_gene	Genbank Synechocystis PCC6803 CDS and proteins
Taci	Genbank Thermoplasma acidophilum genome
Taci_gene	Genbank Thermoplasma acidophilum CDS and proteins
Tmar	Genbank Thermotoga maritima genome
Tmar_gene	Genbank Thermotoga maritima CDS and proteins

Orientation to the Biology Workbench

Tpal	Genbank Treponema pallidum genome
Tpal_gene	Genbank Treponema pallidum CDS and proteins
Tvol	Genbank Thermoplasma volcanium genome
Tvol_gene	Genbank Thermoplasma volcanium CDS and proteins
Urea	Genbank Ureaplasma urealyticum genome
Urea_gene	Genbank Ureaplasma urealyticum CDS and proteins
Vchol	Genbank Vibrio cholerae genome
Vchol_gene	Genbank Vibrio cholerae CDS and proteins
Xfas	Genbank Xylella fastidiosa genome
Xfas_gene	Genbank Xylella fastidiosa CDS and proteins
Ypes	Genbank Yersinia pestis genome
Ypes_gene	Genbank Yersinia pestis CDS and proteins