Bioinformatics Toolbox Release Notes

The Chapter 1, "Bioinformatics Toolbox 1.1 Release Notes" describe the changes introduced in the latest version of the Bioinformatics Toolbox. The following topics are discussed in these Release Notes:

- "New Features" on page 1-2
- "Major Bug Fixes" on page 1-6
- "Platform Limitations for HP and IBM" on page 1-6

If you are upgrading from a version earlier than the Bioinformatics Toolbox 1.1 (Release 14), you should also see:

- Chapter 2, "Bioinformatics Toolbox 1.0 Release Notes"

Printing the Release Notes

If you would like to print the Release Notes, you can link to a PDF version.

Contents

Bioinformatics Toolbox 1.1 Release Notes

New Features

This section summarizes the new features and enhancements introduced in the Bioinformatics Toolbox 1.1.

- "Phylogenetic Analysis Functions" on page 1-2
- "Phylogenetic Tree Object and Methods" on page 1-3
- "Hidden Markov Model (HMM) Profiles" on page 1-3
- "aminolookup, baselookup" on page 1-4
- "blastncbi, blastread, getblast" on page 1-4
- "clustergram" on page 1-4
- "imageneread" on page 1-4
- "mapcaplot" on page 1-4
- "isoelectric" on page 1-4
- "seqdisp" on page 1-4
- "seqdotplot" on page 1-4
- "seqmatch" on page 1-5
- "New Demos" on page 1-5

Phylogenetic Analysis Functions

New functions for phylogenetic tree creation and analysis.

- phytreeread — Read a Newick formatted tree file into the MATLAB workspace and return a phytree object with data from the file. Data in the file uses the Newick (New Hampshire) format for describing trees.
- phytreewrite — Copy the contents of a phytree object from the MATLAB workspace to a file.
- phytreetool — Interactive GUI that allows you to view, edit, and explore phylogenetic tree data. This GUI allows branch pruning, reordering, renaming, and distance exploring. It can also open or save Newick formatted files.
- seqlinkage — Construct a phylogenetic tree from pairwise distances.

• seqpdist — Calculate the pairwise distance between biological sequences.

Phylogenetic Tree Object and Methods

New object for manipulating phylogenetic tree data.

• phytree — Function to create a phytree object.

• phytree/get — Get property values from a phytree object

• phytree/getbyname — Get node names from a phytree object.

• phytree/pdist — Calculate the patristic distances between pairs of leaf nodes.

• phytree/plot — Draw a phylogenetic tree object in a MATLAB figure window as a phylogram, cladogram, or radial tree.

• phytree/prune — Remove nodes from a phylogenetic tree.

• phytree/select — Select branches and leaves from a phylogenetic tree using a specified criteria.

• phytree/view — Opens a phylogenetic tree in a phytreetool window.

Hidden Markov Model (HMM) Profiles

Updated Hidden Markov Model profile functions

• The model structure, that HMM functions use, now includes loop and null transition probabilities. You can read null and loop probabilities from PFAM files using pfamhmmread and from PFAM web databases using gethmmprof..

• When the function hmmprofstruct builds an HMM model, the loop and null transition probabilities default to pre-defined values. If necessary, you can later modify the probabilities using the same function.

• hmmprofalign includes two new properties to control the scoring of flanking states and null transition probabilities. In addition, a third output argument with indices pointing to the respective symbols of the query sequence was added.

aminolookup, baselookup

Updated functions (aminolookup, baselookup) to get IUB/UPAC character codes, integer codes, and names for nucleotides and amino acids.

blastncbi, blastread, getblast

New functions (blastncbi, blastread, getblast) BLAST sequences and view results from within MATLAB.

clustergram

Updated function (clustergram) to do two way bi-clustering.

imageneread

New function (imageneread) to read microarray data from an ImaGene Results file. Other functions read microarray data from Affymetrix GeneChip files (affyread) and GenePix Results (GPR) files (gprread).

mapcaplot

New function (mapcaplot) to create a Principal Component plot of expression profile data

isoelectric

New function (isoelectric) for protein analysis. Estimate the isoelectric point (the pH at which the protein has a net charge of zero) for an amino acid sequence and estimate the charge for a given pH.

seqdisp

New function (seqdisp) formats sequence output for easy viewing.

seqdotplot

Enhanced function (seqdotplot) now returns a second output (the matrix of matches as a sparse matrix).

seqmatch

New function (`seqmatch`) to find matches for every string in a library.

New Demos

- **Bicluster demo** — Demonstrates some of the options of the clustergram function.

- **Bioperl demo** — Illustrates the interoperability between MATLAB and Bioperl - passing arguments from MATLAB to Perl scripts and pulling BLAST search data back to MATLAB

- **Phytree demo for Hominidae species** — A phylogenetic tree is constructed from mtDNA sequences for the Hominidae taxa (also known as pongidae). This family embraces the gorillas, chimpanzees, orangutans and the humans.

- **Phytree demo for HIV/SIV** — Analyzes the reconstruction of phylogenetic trees from infected HIV/SIV organisms.

Major Bug Fixes

The Bioinformatics Toolbox 1.1 includes several bug fixes made after the version 1.0 release.

Corrected Emission Probability

For Hidden Markov Model profiles, the emission probability for the first symbol in the profile is now correct. Previously the emission probability did not include null transition probabilities. Ignoring the null transition probabilities caused a small difference in the answer.

Platform Limitations for HP and IBM

HP-UX Help Browser Limitations

When opening the MATLAB Help browser or a function that users the browser, to display results, a number of exceptions are written to the MATLAB Command Window. These exculpations do not affect viewing information in the browser window.

However, in the left pane, the text labels for the tabs are missing. These tabs are, from left t right, **Contents**, **Index**, **Search**, **Demos**, and **Favorites**.

To view documentation for the Bioinformatics Toolbox, open a Web browser (for example, IE or Netscape), and enter the address.

```
http://www.mathworks.com/access/helpdesk/help/toolbox/bioinfo/
```

Bioinformatics Toolbox 1.0 Release Notes

Introduction

The Bioinformatics Toolbox extends MATLAB with basic sequence analysis and gene expression analysis functions. The Bioinformatics Toolbox is a collection of tools built on the MATLAB numeric computing environment. The toolbox supports a wide range of common sequence analysis and expression analysis tasks, from accessing web-based databases, to sequence alignment, to microarray normalization and visualization.

The Bioinformatics Toolbox is dependent upon many functions from the Statistics Toolbox including some functions only available in the latest version of the Statistics Toolbox, 4.1. We recommend that you install the latest version of the Statistics Toolbox before running the Bioinformatics Toolbox.

Features

This section introduces the features for the Bioinformatics Toolbox 1.0. The Bioinformatics Toolbox has more than 100 functions implemented using M-files. For a complete list of functions, in the **MATLAB Command Window**, type

```
help bioinfo
```

Data I/O
The toolbox provides functions to directly access many standard Web-based databases such as GenBank, EMBL, PIR, and PDB. There are also functions to read many standard file formats, including FASTA and PDB. For microarray data, there are functions to read Affymetrix, GenePix, SPOT format data, and a function to access data directly from the NCBI Gene Expression Omnibus Web site.

Sequence Alignment
The toolbox has functions for pairwise sequence alignment and for hidden Markov model-based sequence profile alignment, including efficient MATLAB implementations of the Needleman-Wunsch and Smith-Waterman algorithms. In addition to the alignment functions there are several tools for visualizing

sequence alignments. The toolbox provides many standard scoring matrices, including the PAM and BLOSUM families.

Sequence Utilities and Statistics

The toolbox contains many functions for working with sequences. There are functions for converting DNA sequences to RNA or amino acid sequences; there are functions that report various statistics about sequences, and functions to search for patterns within the sequence; there are functions for creating random sequences, and there are functions to perform in-silico digestion of sequences with restriction enzymes and proteases.

Microarray Normalization and Visualization

The toolbox contains a number of functions for normalizing microarray data including lowess normalization, global mean normalization, and MAD normalization. The toolbox provides several functions for visualizing microarray data, including spatial heat maps, box plots, loglog, and I-R plots. The toolbox also uses functions from the Statistics Toolbox to perform cluster analysis and to visualize the results.

Protein Structure Analysis

In addition to standard sequence analysis functions, there is also a graphical user interface (GUI), `proteinplot`, for visualizing properties of protein sequences.

Tutorial Demonstrations

There are also several tutorial examples that demonstrate how to use the functions in the toolbox. These tutorials would be a good place to start using the toolbox.