Bioinformatics Toolbox Release Notes

Contents

Summary by Version

This table provides quick access to what's new in each version. For clarification, see "About Release Notes" on page 2.

Version (Release)	New Features and Changes	Version Compatibility Considerations	Fixed Bugs and Known Problems	Related Documentation at Web Site
Latest Version V2.6 (R2007a+)	Yes Details	Yes Summary	Bug Reports Includes fixes	Printable Release Notes: PDF Current product documentation
V2.5 (R2007a)	Yes Details	Yes Summary	Bug Reports Includes fixes	No
V2.4 (R2006b)	Yes Details	No	Bug Reports Includes fixes	No
V2.3 (R2006a+)	Yes Details	No	Bug Reports Includes fixes	No
V2.2.1 (R2006a)	No	No	Bug Reports Includes fixes	No
V2.2 (R14SP3+)	Yes Details	No	Bug Reports Includes fixes	No
V2.1.1 (R14SP3)	No	No	Bug Reports Includes fixes	No
V2.1 (R14SP2+)	Yes Details	No	Bug Reports Includes fixes	No
V2.0.1 (R14SP2)	Yes Details	No	Bug Reports Includes fixes	No
V2.0 (R14SP1+)	Yes Details	No	No bug fixes	No
V1.1.1 (R14SP1)	No	No	No bug fixes	No

Version (Release)	New Features and Changes	Version Compatibility Considerations	Fixed Bugs and Known Problems	Related Documentation at Web Site
V1.1 (R14)	Yes Details	No	No bug fixes	No
V1.0 (R13+)	Yes Details	No	No bug fixes	V1.0 product documentation

About Release Notes

Use release notes when upgrading to a newer version to learn about new features and changes, and the potential impact on your existing files and practices. Release notes are also beneficial if you use or support multiple versions.

If you are not upgrading from the most recent previous version, review release notes for all interim versions, not just for the version you are installing. For example, when upgrading from V1.0 to V1.2, review the New Features and Changes, Version Compatibility Considerations, and Bug Reports for V1.1 and V1.2.

New Features and Changes

These include

- New functionality

- Changes to existing functionality

- Changes to system requirements (complete system requirements for the current version are at the MathWorks Web site)

- Any version compatibility considerations associated with each new feature or change

Version Compatibility Considerations

When a new feature or change introduces a known incompatibility between versions, its description includes a **Compatibility Considerations** subsection that details the impact. For a list of all new features and

changes that have compatibility impact, see the "Compatibility Summary for Bioinformatics Toolbox" on page 41.

Compatibility issues that become known after the product has been released are added to Bug Reports at the MathWorks Web site. Because bug fixes can sometimes result in incompatibilities, also review fixed bugs in Bug Reports for any compatibility impact.

Fixed Bugs and Known Problems

MathWorks Bug Reports is a user-searchable database of known problems, workarounds, and fixes. The MathWorks updates the Bug Reports database as new problems and resolutions become known, so check it as needed for the latest information.

Access Bug Reports at the MathWorks Web site using your MathWorks Account. If you are not logged in to your MathWorks Account when you link to Bug Reports, you are prompted to log in or create an account. You then can view bug fixes and known problems for R14SP2 and more recent releases.

The Bug Reports database was introduced for R14SP2 and does not include information for prior releases. You can access a list of bug fixes made in prior versions via the links in the summary table.

Related Documentation at Web Site

Printable Release Notes (PDF). You can print release notes from the PDF version, located at the MathWorks Web site. The PDF version does not support links to other documents or to the Web site, such as to Bug Reports. Use the browser-based version of release notes for access to all information.

Product Documentation. At the MathWorks Web site, you can access complete product documentation for the current version and some previous versions, as noted in the summary table.

Version 2.6 (Release 2007a+) Bioinformatics Toolbox

This table summarizes what's new in Version 2.6 (Release 2007a+):

New Features and Changes	Version Compatibility Considerations	Fixed Bugs and Known Problems	Related Documentation at Web Site
Yes Details below	Yes Summary	Bug Reports Includes fixes	Printable Release Notes: PDF Current product documentation

New and updated functions in this version include:

- "Data Formats and Databases Functions" on page 4
- "Microarray File Formats Functions" on page 5
- "Microarray Utility Functions" on page 5
- "Microarray Normalization and Filtering Functions" on page 6
- "Mass Spectrometry Functions" on page 6
- "Demos for Mass Spectrometry Functions" on page 6

Data Formats and Databases Functions

The following functions are updated:

- `affyread` — Read microarray data from Affymetrix GeneChip file. Updated to read Affymetrix files from expression, genotyping, or resequencing assays on all platforms, except Solaris.

- `celintensityread` — Read probe intensities from Affymetrix CEL files. Updated to read Affymetrix CEL and CDF files from expression or genotyping assays on all platforms, except Solaris.

- `mzxmlread` — Read mzXML file into MATLAB as structure. Updated to read mzXML files that conform to the mzXML 2.1 specification or earlier specifications.

Compatibility Considerations

In Bioinformatics Toolbox Version 2.6, the structure returned by `affyread` when reading a CHP file from an expression assay no longer contains a `ProbePairs` field. The `ProbePairs` field still exists in the structure returned by `affyread` when reading a CDF file.

Microarray File Formats Functions

The following functions are updated:

- `affyread` — Read microarray data from Affymetrix GeneChip file. Updated to read Affymetrix files from expression, genotyping, or resequencing assays on all platforms, except Solaris.

- `celintensityread` — Read probe intensities from Affymetrix CEL files. Updated to read Affymetrix CEL and CDF files from expression or genotyping assays on all platforms, except Solaris.

Compatibility Considerations

In Bioinformatics Toolbox Version 2.6, the structure returned by `affyread` when reading a CHP file from an expression assay no longer contains a `ProbePairs` field. The `ProbePairs` field still exists in the structure returned by `affyread` when reading a CDF file.

Microarray Utility Functions

The following function is updated:

- `probesetplot` — Plot Affymetrix probe set intensity values. Updated to accept structures created from CEL and CDF files, instead of a structure created from a CHP file.

Compatibility Considerations

In Bioinformatics Toolbox Version 2.5 and earlier, the `probesetplot` function accepted a structure created from a CHP file as input. Currently it requires two structures: one created from a CEL file and one created from a CDF library file. If you have any scripts that call the `probesetplot` function, you need to update them to provide the correct input arguments.

5

Microarray Normalization and Filtering Functions

Following is a new function:

- zonebackadj — Perform background adjustment on Affymetrix microarray probe-level data using zone-based method.

Mass Spectrometry Functions

The following function is updated:

- mzxmlread — Read mzXML file into MATLAB as structure. Updated to read mzXML files that conform to the mzXML 2.1 specification or earlier specifications.

Following is a new function you can use to calibrate and/or synchronize multidimensional mass spectrometry data:

- samplealign — Align two data sets containing sequential observations by introducing gaps.

Demos for Mass Spectrometry Functions

The following are two new mass spectrometry demos:

- Visualizing and Preprocessing Hyphenated Mass-Spectrometry Data Sets for Metabolite and Protein/Peptide Profiling

- Differential Analysis of Complex Protein and Metabolic Mixtures Using Liquid Chromatography/Mass Spectrometry (LC/MS)

Version 2.5 (Release 2007a) Bioinformatics Toolbox

This table summarizes what's new in Version 2.5 (Release 2007a):

New Features and Changes	Version Compatibility Considerations	Fixed Bugs and Known Problems	Related Documentation at Web Site
Yes Details below	Yes Summary	Bug Reports Includes fixes	No

New, updated, and deprecated functions in this version include:

- "Data Formats and Database Functions" on page 8

- "Demo for Data Formats and Database Functions" on page 8

- "Statistical Learning Functions" on page 9

- "Protein Analysis and Sequence Utilities Functions" on page 9

- "Demo for Protein Analysis and Sequence Utilities Functions" on page 10

- "Sequence Alignment Functions" on page 10

- "Demo for Sequence Alignment Functions" on page 11

- "Microarray File Formats Functions" on page 11

- "Microarray Normalization and Filtering Functions" on page 11

- "Demo for Microarray File Formats, Normalization, and Filtering Functions" on page 12

- "Microarray Data Analysis and Visualization Functions" on page 12

- "Demo for Microarray Data Analysis and Visualization Functions" on page 12

- "Mass Spectrometry Functions" on page 12

- "Phylogenetic Tree Tools Functions" on page 13

- "Demos for Phylogenetic Tree Tools Functions" on page 13

- "Phylogenetic Tree Methods" on page 14

Data Formats and Database Functions

Following are new functions for reading and creating files:

- `affyprobeseqread` — Read data file containing probe sequence information for Affymetrix® GeneChip® array.

- `pdbwrite` — Write to file using Protein Data Bank (PDB) format.

The following functions were updated:

- `celintensityread` — Read probe intensities from Affymetrix CEL files (Windows 32). Updated so that the order of columns (CEL files) in return matrices `PMIntensities` and `MMIntensities` matches the order of CEL files in the *CELFiles* input argument.

- `pdbread` — Read data from Protein Data Bank (PDB) file. Updated so that the six fields containing coordinate information (`Atom`, `AtomSD`, `AnisotropicTemp`, `AnisotropicTempSD`, `Terminal`, and `HeterogenAtom`) are now subfields within the `Model` field of the MATLAB structure. Updated to include a new property, `ModelNum`, which reads only the specified model from a PDB-formatted text file.

Compatibility Considerations

In Bioinformatics Toolbox Version 2.4 and earlier, the `celintensityread` function ordered the columns (CEL files) of return matrices `PMIntensities` and `MMIntensities` alphabetically.

In Bioinformatics Toolbox Version 2.4 and earlier, the `pdbread` function stored coordinate information in six fields (`Atom`, `AtomSD`, `AnisotropicTemp`, `AnisotropicTempSD`, `Terminal`, and `HeterogenAtom`) within the MATLAB structure. These six fields are now subfields within the `Model` field of the MATLAB structure.

Demo for Data Formats and Database Functions

The Accessing NCBI Entrez Databases with E-Utilities demo illustrates how to programatically search and retrieve data.

Statistical Learning Functions

Following are new functions:

- `optimalleaforder` — Determine optimal leaf ordering for hierarchical binary cluster tree.

- `svmsmoset` — Create or edit Sequential Minimal Optimization (SMO) options structure.

The following function was updated:

- `svmtrain` — Train support vector machine classifier. Updated to include a new SMO method and a new property, `SMO_Opts`, which provides options for the `SMO` method. The `BoxConstraint` property has changed, including a new default value.

Compatibility Considerations

In Bioinformatics Toolbox Version 2.4 and earlier, the `svmtrain` function used a `BoxConstraint` property with a default of $\dfrac{1}{\sqrt{eps}}$. In Bioinformatics Toolbox Version 2.5, the default is 1, which can lead to slightly different results.

Protein Analysis and Sequence Utilities Functions

Following are new functions:

- `evalrasmolscript` — Send RasMol script commands to molecule viewer.

- `molviewer` — Display and manipulate 3-D molecule structure.

- `proteinpropplot` — Plot properties of amino acid sequence.

- `seqinsertgaps` — Insert gaps into nucleotide or amino acid sequence.

The following functions were updated:

- `featuresparse` — Parse features from GenBank, GenPept, or EMBL data. Updated to include a new property, `Sequence`, which controls the extraction, when possible, of the sequences.

- `oligoprop` — Calculate sequence properties of DNA oligonucleotide. Updated to handle ambiguous `N` characters in a sequence.

The following function is obsolete:

- `pdbplot` — Plot 3-D protein structure. This function was replaced by the `molviewer` function.

Compatibility Considerations

In Bioinformatics Toolbox Version 2.5, the `pdbplot` function was replaced by the `molviewer` function. If you have any scripts that call the `pdbplot` function, you need to update them to call the `molviewer` function.

Demo for Protein Analysis and Sequence Utilities Functions

The Visualizing the Three-dimensional Structure of a Molecule demo illustrates the `molviewer` function.

Sequence Alignment Functions

The following function was updated:

- `seqpdist` — Calculate pair-wise distance between sequences. Updated to assume that all input sequences are aligned if they have the same length, regardless of the presence of gaps. If you know your input sequences are not aligned, you can align them before passing them to `seqpdist` (for example, using `multialign`), or set `PairwiseAlignment` to `true` when using `seqpdist`.

Compatibility Considerations

In Bioinformatics Toolbox Version 2.4 and earlier, the `seqpdist` function assumed all input sequences were aligned if they had the same length and at least one gap.

Demo for Sequence Alignment Functions

The Comparing Whole Genomes demo illustrates how to compare features of organisms on a genomic evolution scale.

Microarray File Formats Functions

Following is a new function:

- `affyprobeseqread` — Read data file containing probe sequence information for Affymetrix GeneChip array.

The following function was updated:

- `celintensityread` — Read probe intensities from Affymetrix CEL files (Windows 32). Updated so that the order of columns (CEL files) in return matrices `PMIntensities` and `MMIntensities` matches the order of CEL files in the *CELFiles* input argument.

Compatibility Considerations

In Bioinformatics Toolbox Version 2.4 and earlier, the `celintensityread` function ordered the columns (CEL files) of return matrices `PMIntensities` and `MMIntensities` alphabetically.

Microarray Normalization and Filtering Functions

Following are new functions:

- `affyprobeaffinities` — Compute Affymetrix probe affinities from their sequences and MM probe intensities.
- `gcrmabackadj` — Perform GC Robust Multi-array Average (GCRMA) background adjustment on Affymetrix microarray probe-level data using sequence information.
- `gcrma` — Perform GC Robust Multi-array Average (GCRMA) background adjustment, quantile normalization, and median-polish summarization on Affymetrix microarray probe-level data.

11

Demo for Microarray File Formats, Normalization, and Filtering Functions

The Preprocessing Affymetrix Microarray Data at the Probe Level demo illustrates the `affyprobeseqread`, `affyprobeaffinities`, `gcrmabackadj`, and `gcrma` functions.

Microarray Data Analysis and Visualization Functions

Following is a new function:

- `mafdr` — Estimate false discovery rate (FDR) of differentially expressed genes from two experimental conditions or phenotypes.

The following function was updated:

- `mattest` — Perform two-tailed t-test to evaluate differential expression of genes from two experimental conditions or phenotypes. Updated to include a new property, `Permute`, which controls whether permutation tests are run.

Demo for Microarray Data Analysis and Visualization Functions

The Exploring Gene Expression Data demo illustrates the `mattest` and `mafdr` functions.

Mass Spectrometry Functions

Following are new functions:

- `msdotplot` — Plot set of peak lists from LC/MS or GC/MS data set.
- `mspalign` — Align mass spectra from multiple peak lists from LC/MS or GC/MS data set.
- `mspeaks` — Convert raw mass spectrometry data to peak list (centroided data).
- `msppresample` — Resample mass spectrometry signal while preserving peaks.
- `mzxml2peaks` — Convert mzXML structure to peak list.

The following function was updated:

- `msheatmap` — Create pseudocolor image of set of mass spectra. Updated to handle LC/MS and GC/MS data.

Phylogenetic Tree Tools Functions

Following is a new function:

- `seqinsertgaps` — Insert gaps into nucleotide or amino acid sequence.

The following functions were updated:

- `dnds` — Estimate synonymous and nonsynonymous substitution rates. Updated to include two new properties, `Verbose`, which controls the display of the codons considered in the computations and their amino acid translations, and `Window`, which performs the calculations over a sliding window.
- `dndsml` — Estimate synonymous and nonsynonymous substitution rates using maximum likelihood method. Updated to include a new property, `Verbose`, which controls the display of the codons considered in the computations and their amino acid translations.
- `seqpdist` — Calculate pair-wise distance between sequences. Updated to assume that all input sequences are aligned if they have the same length, regardless of the presence of gaps. If you know your input sequences are not aligned, you can align them before passing them to `seqpdist` (for example, using `multialign`), or set `PairwiseAlignment` to `true` when using `seqpdist`.

Compatibility Considerations

In Bioinformatics Toolbox Version 2.4 and earlier, the `seqpdist` function assumed all input sequences were aligned if they had the same length and at least one gap.

Demos for Phylogenetic Tree Tools Functions

The following demos illustrate the `nwalign`, `seqinsertgaps`, `dnds`, and `multialign` functions:

- Analyzing Synonymous and Nonsynonymous Substitution Rates

- Investigating the Bird Flu Virus

The Reconstructing the Origin and the Diffusion of the SARS Epidemic demo presents an analysis of the SARS epidemic.

Phylogenetic Tree Methods

Following is a new method of a phytree object:

- reorder — Reorder leaves of phylogenetic tree.

Version 2.4 (Release 2006b) Bioinformatics Toolbox

This table summarizes what's new in Version 2.4 (Release 2006b):

New Features and Changes	Version Compatibility Considerations	Fixed Bugs and Known Problems	Related Documentation at Web Site
Yes Details below	Yes Summary	Bug Reports Includes fixes	No

New functions, obsoleted functions, and changes introduced in this version are

- "Data Formats and Database Functions" on page 15
- "Sequence Utilities Functions" on page 16
- "Sequence Visualization Functions" on page 16
- "Multiple Sequence Alignment Functions" on page 16
- "Microarray File Formats" on page 16
- "Microarray Data Analysis and Visualization Functions" on page 17
- "Graph Theory Functions" on page 17
- "Graph Visualization Methods" on page 18
- "Phylogenetic Tree Methods" on page 18

Data Formats and Database Functions

Following is a new function for getting data into MATLAB:

- mzxmlread — Read mzXML file into MATLAB as structure.

The following functions were updated:

- celintensityread — Read probe intensities from Affymetrix CEL files (Windows 32). Updated to include a new property, Verbose, which controls the display of a progress report showing the name of each CEL file as it is read.

15

- `fastaread` — Read data from FASTA file. Updated to include a new property, `Blockread`, which controls reading a single entry or block of entries from a file.

- `geosoftread` — Read Gene Expression Omnibus (GEO) SOFT format data. Updated to read Data Set (GDS) files as well as Sample (GSM) files.

- `getblast` — BLAST report from NCBI Web site. Updated to include a new property, `WaitTilReady`, which pauses MATLAB and waits a specified time (minutes) for a report from the NCBI Web site.

- `scfread` — Read trace data from SCF file. Updated to include more output options.

Sequence Utilities Functions

Following is a new function for parsing sequence data:

- `featuresparse` — Parse features from GenBank, GenPept, or EMBL data.

Sequence Visualization Functions

The following function was updated:

- `seqtool` — Open tool to interactively explore biological sequences. Updated to download sequences from the EMBL database, interactively move the viewing frame in the Sequence Viewer by pressing and holding **Ctrl** while click-dragging, and export an amino acid translation as a FASTA file or to the MATLAB Workspace.

Multiple Sequence Alignment Functions

The following function was updated:

- `multialignviewer` — Open viewer for multiple sequence alignments. Updated to export consensus sequences.

Microarray File Formats

The following function was updated:

- `celintensityread` — Read probe intensities from Affymetrix CEL files (Windows 32). Updated to include a new property, `Verbose`, which controls

the display of a progress report showing the name of each CEL file as it is read.

Microarray Data Analysis and Visualization Functions

The following functions were updated:

- `clustergram` — Create dendrogram and heat map. Updated to include a new property, `OptimalLeafOrder`, which enables or disables the optimal leaf ordering calculation, which determines the leaf order that maximizes the similarity between neighboring leaves.

- `mairplot` — Create intensity versus ratio scatter plot for microarray signals. Updated to include a new property, `Type`, which creates either an IR plot or MA plot, changing the plot axes to log scale, and adding plot interactive features such as displaying gene labels, changing factor lines, normalizing data, and exporting data.

- `mapcaplot` — Create Principal Component plot of expression profile data. Updated by adding an export feature.

- `redgreencmap` — Create red and green colormap. Updated to include a new property, `Interpolation`, which sets the method for color interpolation.

Graph Theory Functions

Following are new functions for applying basic graph theory algorithms to sparse matrices:

- `graphallshortestpaths` — Find all shortest paths in graph.

- `graphconncomp` — Find strongly or weakly connected components in graph.

- `graphisdag` — Test for cycles in directed graph.

- `graphisomorphism` — Find isomorphism between two graphs.

- `graphisspantree` — Determine if tree is spanning tree.

- `graphmaxflow` — Calculate maximum flow and minimum cut in directed graph.

- `graphminspantree` — Find minimal spanning tree in graph.

- `graphpred2path` — Convert predecessor indices to paths.

- `graphshortestpath` — Solve shortest path problem in graph.
- `graphtopoorder` — Perform topological sort of directed acyclic graph.
- `graphtraverse` — Traverse graph by following adjacent nodes.

Graph Visualization Methods

Following are new methods for applying basic graph theory algorithms to a `biograph object`:

- `allshortestpaths` — Find all shortest paths in biograph object.
- `conncomp` — Find strongly or weakly connected components in biograph object.
- `getmatrix` — Get connection matrix from biograph object.
- `isdag` — Test for cycles in biograph object.
- `isomorphism` — Find isomorphism between two biograph objects.
- `isspantree` — Determine if tree created from biograph object is spanning tree.
- `maxflow` — Calculate maximum flow and minimum cut in biograph object.
- `minspantree` — Find minimal spanning tree in biograph object.
- `shortestpath` — Solve shortest path problem in biograph object.
- `topoorder` — Perform topological sort of directed acyclic graph extracted from biograph object.
- `traverse` — Traverse biograph object by following adjacent nodes.

Phylogenetic Tree Methods

Following is a new method for the `phytree` object:

- `getmatrix` — Convert phytree object into a relationship matrix.

Version 2.3 (Release 2006a+) Bioinformatics Toolbox

This table summarizes what's new in Version 2.3 (Release 2006a+):

New Features and Changes	Version Compatibility Considerations	Fixed Bugs and Known Problems	Related Documentation at Web Site
Yes Details below	No	Bug Reports at Web site	No

New functions, obsoleted functions, and changes introduced in this version are

- "Data Formats and Databases Functions" on page 19

- "Sequence Utilities Functions" on page 20

- "Sequence Visualization Functions" on page 20

- "Statistical Learning Functions" on page 20

- "Microarray Functions" on page 20

- "Demo for Microarray Functions" on page 21

Data Formats and Databases Functions

The following functions are obsolete:

- getpir — Sequence data from PIR-PSD database. This function retrieved data from the PIR-PSD database. This database has been discontinued and this function no longer retrieves data. See `http://pir.georgetown.edu/pirwww/dbinfo/nref.shtml` for more details.

- pirread — Read data from Protein Information Resource (PIR) file. This function supported the data format of the PIR-PSD database. This database has been discontinued. See `http://pir.georgetown.edu/pirwww/dbinfo/nref.shtml` for more details.

Sequence Utilities Functions

The following function was updated to include five new databases, including refseq_rna, refseq_genomic, env_nt, refseq_protein, and env_nr:

- blastncbi — Generate remote BLAST request.

Sequence Visualization Functions

Following is a new function for visualizing sequence data:

- featuresmap — Draw linear or circular map of features from GenBank structure.

Statistical Learning Functions

The following function was updated to include three new properties, including RBF_Sigma, BoxConstraint, and Autoscale:

- svmtrain — Train support vector machine classifier.

Microarray Functions

The following function is supported on the Windows 32 platform only:

- affyread — Read microarray data from Affymetrix GeneChip file (Windows 32).

Following are new functions for preprocessing Affymetrix probe-level microarray data:

- celintensityread — Read probe intensities from Affymetrix CEL files (Windows 32).
- rmabackadj — Perform background adjustment on Affymetrix microarray probe-level data using Robust Multi-array Average (RMA) procedure.
- rmasummary — Calculate gene (probe set) expression values from Affymetrix microarray probe-level data using Robust Multi-array Average (RMA) procedure.
- affyinvarsetnorm — Perform rank invariant set normalization on probe intensities from multiple Affymetrix CEL or DAT files.

Following is a new function for two-color microarray normalization:

- `mainvarsetnorm` — Perform rank invariant set normalization on gene expression values from two experimental conditions or phenotypes.

Following are new functions for microarray differential expression analysis:

- `mattest` — Perform two-sample, two-tailed t-test to evaluate differential expression of genes from two experimental conditions or phenotypes.

- `mavolcanoplot` — Create significance versus gene expression ratio (fold change) scatter plot of microarray data.

Demo for Microarray Functions

New demo of the new microarray functions (Analyzing Affymetrix Microarray Gene Expression Data).

Version 2.2.1 (Release 2006a) Bioinformatics Toolbox

This table summarizes what's new in Version 2.2.1 (Release 2006a):

New Features and Changes	Version Compatibility Considerations	Fixed Bugs and Known Problems	Related Documentation at Web Site
No	No	Bug Reports at Web site	No

Version 2.2 (Release 14SP3+) Bioinformatics Toolbox

This table summarizes what's new in Version 2.2 (Release 14SP3+):

New Features and Changes	Version Compatibility Considerations	Fixed Bugs and Known Problems	Related Documentation at Web Site
Yes Details below	No	Bug Reports at Web site	No

New features and changes introduced in this version are

- "Multiple Sequence Alignment Viewer" on page 23
- "Microarray Functions for Agilent Software" on page 23
- "Gene Ontology Database Functions" on page 23
- "Demo for Gene Ontology Functions" on page 24

Multiple Sequence Alignment Viewer

- `multialignviewer` — Interactively view, explore alignments, and make manual modifications.

Microarray Functions for Agilent Software

- `agferead` — Read an Agilent Feature Extraction Software file.
- `magetfield` — Utility function to extract data from a microarray.

Gene Ontology Database Functions

- `geneont` — Import the Gene Ontology database from the Web.
- `getancestors`, `getdescendants`, `getrelatives` — Get a subset of the ontology.
- `goannotread` — Parse Gene Ontology Annotated files.
- `num2goid` — Convert numbers to Gene Ontology IDs.

Demo for Gene Ontology Functions

New demo for the new Gene Ontology functions (`geneontologydemo`) and working with whole genomes (`biomemorymapdemo`).

Version 2.1.1 (Release 14SP3) Bioinformatics Toolbox

This table summarizes what's new in Version 2.1.1 (Release 14SP3):

New Features and Changes	Version Compatibility Considerations	Fixed Bugs and Known Problems	Related Documentation at Web Site
No	No	Bug Reports at Web site	No

Version 2.1 (Release 14SP2+) Bioinformatics Toolbox

This table summarizes what's new in Version 2.1 (Release 14SP2+)

New Features and Changes	Version Compatibility Considerations	Fixed Bugs and Known Problems	Related Documentation at Web Site
Yes Details below	No	Bug Reports at Web site	No

New features and changes introduced in this version are:

- "Sequence Alignment Functions" on page 26
- "Sequence Statistics Functions" on page 27
- "Sequence Utilities Functions" on page 27
- "Phylogenetic Tree Functions" on page 27
- "Phylogenetic Tree Methods" on page 27
- "Microarray Functions" on page 27
- "Statistics Functions" on page 27

Sequence Alignment Functions

- multialign — Align multiple sequences using a progressive method with Distributed Computing Toolbox support.
- multialignread — Read multiple sequence alignment file.
- profalign — Align two profiles using Needleman-Wunsch global alignment.
- showalignment — Updated to show multiply aligned sequences.
- seqpdist — Updated to calculate pair-wise distances between observations with Distributed Computing Toolbox support.

Sequence Statistics Functions

- codonbias — Calculate codon frequency for each amino acid in a DNA sequence.

- cpgisland — Locate CpG islands in a DNA sequence.

Sequence Utilities Functions

- rebasecuts — Find restriction enzymes that cut a protein sequence.

- seqtool — Graphical User Interface (GUI) for single sequence analysis.

Phylogenetic Tree Functions

- dnds, dndsml — Estimate synonymous and nonsynonymous substitutions rates.

- seqneighjoin — Reconstruct a phylogenetic tree with a Neighbor-joining method.

Phylogenetic Tree Methods

- getcanonical — Calculate the canonical form of a phylogenetic tree.

- getnewwickstr — Create a Newick formatted string.

- reroot — Change the root of a phylogenetic tree.

- subtree — Extract a subtree.

- weights — Calculate weights for a phylogenetic tree.

Microarray Functions

probesetplot — Plot values for an Affymetrix CHP file probe set.

Statistics Functions

rankfeatures — Renamed function. The previous name was sqtlfeatures.

Version 2.0.1 (Release 14SP2) Bioinformatics Toolbox

This table summarizes what's new in Version 2.0.1 (Release 14SP2):

New Features and Changes	Version Compatibility Considerations	Fixed Bugs and Known Problems	Related Documentation at Web Site
Yes Details below	No	Bug Reports at Web site	No

New features and changes introduced in this version are

- "Updated RBASE Table" on page 28
- "Expanded Bioperl Demonstration" on page 28

Updated RBASE Table

RBASE is the enzyme table that the function `restrict` uses to locate sequence patterns.

Expanded Bioperl Demonstration

Example of calling MATLAB from Perl scripts now includes several examples of passing various types of data (both directly and by variant variable) back and forth between Perl and a MATLAB Automation Server. To view the demo, type `bioperldemo`.

Version 2.0 (Release 14SP1+) Bioinformatics Toolbox

This table summarizes what's new in Version 2.0 (Release 14SP1+):

New Features and Changes	Version Compatibility Considerations	Fixed Bugs and Known Problems	Related Documentation at Web Site
Yes Details below	No	Bug Reports at Web site	No

New features and changes introduced in this version are

- "Mass Spectrometry Data Analysis" on page 29

- "Graph Visualization Object and Methods" on page 30

- "Statistical Learning" on page 30

- "Sequence Analysis" on page 30

- "Protein Analysis" on page 31

- "Microarray Analysis" on page 31

- "Updated Web Connectivity Function" on page 32

Mass Spectrometry Data Analysis

Following are new functions designed for preprocessing and classification of raw mass spectrometry data from SELDI-TOF and MALDI-TOF spectrometers.

- msresample — Resample with antialias filtering.

- msbackadj — Correct a baseline by estimation.

- msalign — Align a spectrum to a set of candidate peaks.

- msheatmap — Draw a heat map image for a set of spectra and check alignments.

- msnorm — Normalize a set of spectra.

- mslowess — Nonparametric smoothing using the Lowess method.

- `mssgolay` — Least-squares polynomial smoothing.

- `msviewer` — Plot a spectrum or a set of spectra.

Graph Visualization Object and Methods

New object and set of methods to view relationships between data with interactive maps.

- `biograph` — Function to create a biograph object.

- `dolayout` — Calculate node and edge positions.

- `getnodesbyid` — Get handles to nodes.

- `getedgesbynodeid` — Get handles to edges.

- `view` — Render a graph in its viewer.

- `getancestors` — Find ancestors.

- `getdescendants` — Find descendants.

- `getrelatives` — Find neighbors.

Statistical Learning

New set of functions to classify data and identify features in the data.

- `classperf` — Evaluate the performance of a classifier.

- `crossvalind` — Cross-validation index generation.

- `knnclassify` — K-Nearest neighbor classifier.

- `knnimpute` — Impute missing data using the nearest neighbor method.

- `randfeatures` — Randomized subset feature selection.

- `sqtlfeatures` — Sequential forward feature selection. This function will be renamed to `rankfeatures` in Version 2.1.

- `svmclassify` — Classify using a support vector machine classifier.

- `svmtrain` — Train a support vector machine classifier.

Sequence Analysis

New functions for analysis and visualization of multiple sequences.

- `seqconsensus` — Computes the consensus sequence for a set of sequences.
- `seqlogo` — Displays sequence logos for DNA and protein sequences.
- `seqprofile` — Computes the sequence profile of a multiple alignment.

Updated functions.

- `palindromes` — Updated to allow for gaps in the palindrome.
- `seqshoworfs`, `seqshowwords`, `showalignment` — Updated to display the results in a Figure window. (This may cause problems on the Mac.)

In Bioinformatics Toolbox 2.0 the functions `seqlogo`, `seqshowwords`, `seqshoworfs`, and `showalignment` use Java based figures. Currently on the Macintosh, Java figures are not enabled by default. If you use these functions on a Macintosh, you should start MATLAB with

```
matlab -useJavaFigures
```

Protein Analysis

- `pdbplot` — Plots 3-D backbone structure of proteins in a PDB file.

Microarray Analysis

- `quantilenorm` — Quantile normalization.

New set of functions for working with Affymetrix GeneChip data sets.

- `probelibraryinfo` — Get library information for a probe.
- `probesetlink` — Show probe set information from NetAffx.
- `probesetlookup` — Get gene information for a probe set.
- `probesetplot` — Plot probe set values.
- `probesetvalues` — Get probe set values from CEL and CDF information.
- `manorm` — Normalization by scaling and centering replaces the functions `mamadnorm` and `mameannorm`.

- `affyread` — Updated with output structures that have changed slightly. Some redundant fields have been removed from CDF and CHP structure. GIN database files are now supported. Version 4 of the Affymetrix GDAC File Access Runtime Libraries is provided.

Note If you use `mamadnorm` or `mameannorm` in any of your personal M-files, please update your files with the new function `manorm`. These functions are now obsolete and may be removed from future releases of Bioinformatics Toolbox.

- `geosoftread` — Updated with supports Gene Expression Omnibus Database records (GDS files).
- `maimage` — Updated with supports Affymetrix CEL data.
- `maboxplot` — Now supports Affymetrix CHP data.

Affymetrix and Gene Chip are registered trademarks of Affymetrix Inc.

Updated Web Connectivity Function

`getgenbank` — Now returns CDS information for a gene in a structure allowing direct access to the transcribed sequence.

Version 1.1.1 (Release 14SP1) Bioinformatics Toolbox

This table summarizes what's new in Version 1.1.1 (Release 14SP1):

New Features and Changes	Version Compatibility Considerations	Fixed Bugs and Known Problems	Related Documentation at Web Site
No	No	No bug fixes	No

Version 1.1 (Release 14) Bioinformatics Toolbox

This table summarizes what's new in Version 1.1 (Release 14)

New Features and Changes	Version Compatibility Considerations	Fixed Bugs and Known Problems	Related Documentation at Web Site
Yes Details below	No	No bug fixes	No

New features and changes introduced in this version are

- "Phylogenetic Analysis Functions" on page 34
- "Phylogenetic Tree Object and Methods" on page 35
- "Hidden Markov Model (HMM) Profiles" on page 35
- "BLAST Functions" on page 36
- "Microarray Functions" on page 36
- "Protein Analysis Function" on page 36
- "Sequence Alignment Functions" on page 36
- "New Demos" on page 36

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 pair-wise distances.

- `seqpdist` — Calculate the pair-wise distance between biological sequences.

Phylogenetic Tree Object and Methods

New object for manipulating phylogenetic tree data.

- `phytree` — Function to create a phytree object.

- `get` — Get property values from a phytree object

- `getbyname` — Get node names from a phytree object.

- `pdist` — Calculate the patristic distances between pairs of leaf nodes.

- `plot` — Draw a phylogenetic tree object in a MATLAB figure window as a phylogram, cladogram, or radial tree.

- `prune` — Remove nodes from a phylogenetic tree.

- `select` — Select branches and leaves from a phylogenetic tree using a specified criteria.

- `view` — Open 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 predefined 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.

BLAST Functions

`blastncbi`, `blastread`, `getblast` — BLAST sequences and view results from within MATLAB.

Microarray Functions

- `imageneread` — Read microarray data from an ImaGene Results file.

- `affyread` — Read microarray data from Affymetrix GeneChip files.

- `gprread` — Read microarray data from GenePix Results (GPR) files.

- `mapcaplot` — Create a Principal Component plot of expression profile data.

- `clustergram` — Updated function to do two way bi-clustering.

Protein Analysis Function

`isoelectric` — 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.

Sequence Alignment Functions

- `seqdisp` — Formats sequence output for easy viewing.

- `seqmatch` — Find matches for every string in a library.

- `seqdotplot` — Updated function now returns a second output (the matrix of matches as a sparse matrix).

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

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.

Version 1.0 (Release 13+) Bioinformatics Toolbox

This table summarizes what's new in Version 1.0 (Release 13+):

New Features and Changes	Version Compatibility Considerations	Fixed Bugs and Known Problems	Related Documentation at Web Site
Yes Details below	No	No bug fixes	V1.0 product documentation

New features and changes introduced in this version are

- "Introduction to Bioinformatics Toolbox" on page 38

- "Databases and Data Formats" on page 39

- "Sequence Alignment" on page 39

- "Sequence Utilities and Statistics" on page 39

- "Microarray Normalization and Visualization" on page 39

- "Protein Structure Analysis" on page 40

Introduction to Bioinformatics Toolbox

Bioinformatics Toolbox Version 1.0 (Web Release R13 SP1+) extends MATLAB with basic sequence analysis and gene expression analysis functions. 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.

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

Bioinformatics Toolbox 1.0 has more than 100 functions implemented using M-files. For a complete list of functions, in the MATLAB Command Window, type

```
help bioinfo
```

Databases and Data Formats

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 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 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.

Compatibility Summary for Bioinformatics Toolbox

This table summarizes new features and changes that might cause incompatibilities when you upgrade from an earlier version, or when you use files on multiple versions. Details are provided in the description of the new feature or change.

Version (Release)	New Features and Changes with Version Compatibility Impact
Latest Version V2.6 (R2007a+)	See the **Compatibility Considerations** subheadings for these changes: • "Data Formats and Databases Functions" on page 4 • "Microarray File Formats Functions" on page 5 • "Microarray Utility Functions" on page 5
V2.5 (R2007a)	See the **Compatibility Considerations** subheadings for these changes: • "Data Formats and Database Functions" on page 8 • "Statistical Learning Functions" on page 9 • "Protein Analysis and Sequence Utilities Functions" on page 9 • "Sequence Alignment Functions" on page 10 • "Microarray File Formats Functions" on page 11 • "Phylogenetic Tree Tools Functions" on page 13
V2.4 (R2006b)	None
V2.3 (R2006a+)	None
V2.2.1 (R2006a)	None
V2.2 (R14SP3+)	None
V2.1.1 (R14SP3)	None
V2.1 (R14SP2+)	None

Version (Release)	New Features and Changes with Version Compatibility Impact
V2.0.1 (R14SP2)	None
V2.0 (R14SP1+)	None
V1.1.1 (R14SP1)	None
V1.1 (R14)	None
V1.0 (R13+)	None