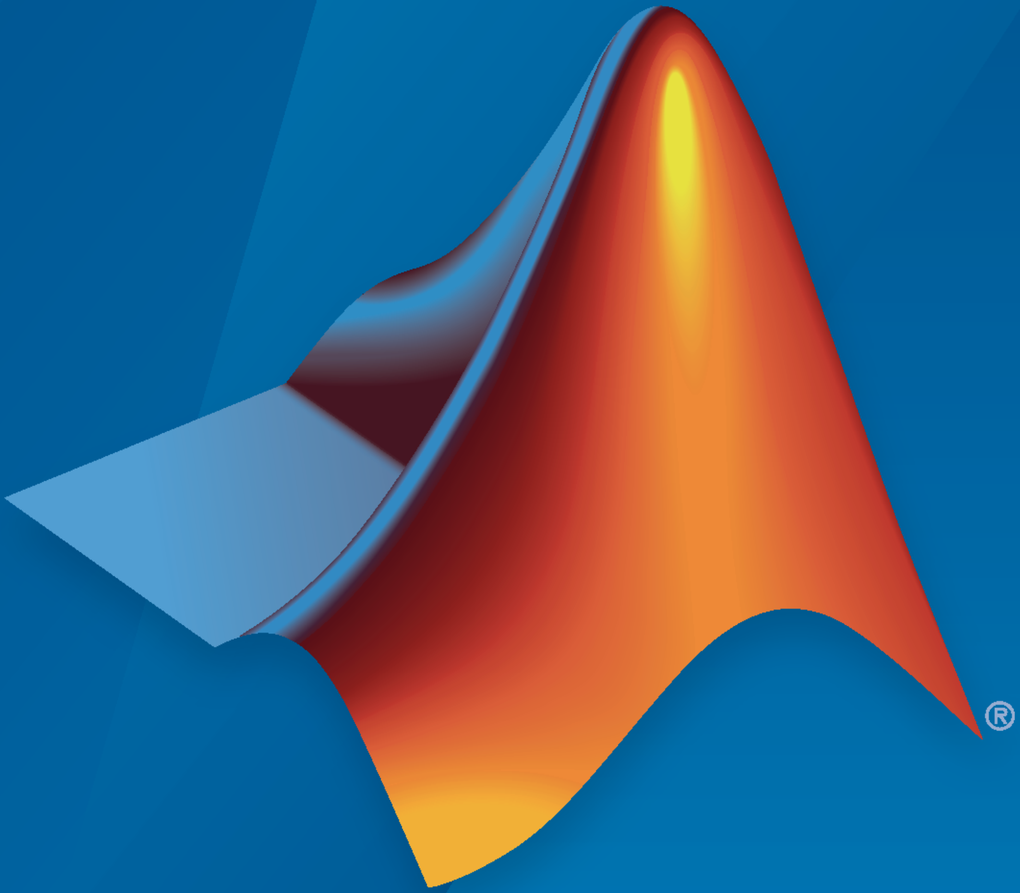
Bioinformatics Toolbox™ Release Notes



MATLAB®

MathWorks®

How to Contact MathWorks

Latest news: `www.mathworks.com`

Sales and services: `www.mathworks.com/sales_and_services`

User community: `www.mathworks.com/matlabcentral`

Technical support: `www.mathworks.com/support/contact_us`

Phone: 508-647-7000

The MathWorks, Inc.
1 Apple Hill Drive
Natick, MA 01760-2098

Bioinformatics Toolbox™ Release Notes

© COPYRIGHT 2003–2023 by The MathWorks, Inc.

The software described in this document is furnished under a license agreement. The software may be used or copied only under the terms of the license agreement. No part of this manual may be photocopied or reproduced in any form without prior written consent from The MathWorks, Inc.

FEDERAL ACQUISITION: This provision applies to all acquisitions of the Program and Documentation by, for, or through the federal government of the United States. By accepting delivery of the Program or Documentation, the government hereby agrees that this software or documentation qualifies as commercial computer software or commercial computer software documentation as such terms are used or defined in FAR 12.212, DFARS Part 227.72, and DFARS 252.227-7014. Accordingly, the terms and conditions of this Agreement and only those rights specified in this Agreement, shall pertain to and govern the use, modification, reproduction, release, performance, display, and disclosure of the Program and Documentation by the federal government (or other entity acquiring for or through the federal government) and shall supersede any conflicting contractual terms or conditions. If this License fails to meet the government's needs or is inconsistent in any respect with federal procurement law, the government agrees to return the Program and Documentation, unused, to The MathWorks, Inc.

Trademarks

MATLAB and Simulink are registered trademarks of The MathWorks, Inc. See `www.mathworks.com/trademarks` for a list of additional trademarks. Other product or brand names may be trademarks or registered trademarks of their respective holders.

Patents

MathWorks products are protected by one or more U.S. patents. Please see `www.mathworks.com/patents` for more information.

Contents

R2023a

Bioinformatics Pipeline **1-2**

Overwrite Option for seqtrim and seqfilter **1-2**

Functionality being removed or changed **1-2**
- probesetlink has been removed **1-2**
- display name-value argument of pdbsuperpose will be removed in a future release **1-2**
- pdbsuperpose no longer opens Molecule Viewer **1-2**
- molviewer and Molecule Viewer app will be removed **1-2**
- bowtie has been removed **1-2**
- bowtiebuild has been removed **1-3**

R2022b

Functionality being removed or changed **2-2**
- bowtie will be removed **2-2**
- bowtiebuild will be removed **2-2**
- graphallshortestpaths has been removed **2-2**
- graphconncomp has been removed **2-2**
- graphisdiag has been removed **2-2**
- graphisomorphism has been removed **2-2**
- graphisspantree has been removed **2-2**
- graphmaxflow has been removed **2-2**
- graphminspantree has been removed **2-2**
- graphpred2path has been removed **2-2**
- graphshortestpath has been removed **2-3**
- graphtopoorder has been removed **2-3**
- graphtraverse has been removed **2-3**
- biograph has been removed **2-3**

R2022a

Windows support for Bowtie 2, BWA, and Cufflinks support packages . . **3-2**

Functionality being removed or changed **3-2**
Bioinformatics Toolbox Interface for Bowtie Aligner support package has been replaced **3-2**
bowtie will be removed **3-2**
bowtiebuild will be removed **3-2**
graphallshortestpaths will be removed **3-2**
graphconncomp will be removed **3-2**
graphisdiag will be removed **3-2**
graphisomorphism will be removed **3-2**
graphisspantree will be removed **3-3**
graphmaxflow will be removed **3-3**
graphminspantree will be removed **3-3**
graphpred2path will be removed **3-3**
graphshortestpath will be removed **3-3**
graphtopoorder will be removed **3-3**
graphtraverse will be removed **3-3**
biograph will be removed **3-3**
bamindexread has been removed **3-4**

R2021b

rnaseqde: Perform differential expression analysis on RNA-seq count data **4-2**

Functionality being removed or changed **4-2**
graphallshortestpaths will be removed **4-2**
graphconncomp will be removed **4-2**
graphisdiag will be removed **4-2**
graphisomorphism will be removed **4-2**
graphisspantree will be removed **4-2**
graphmaxflow will be removed **4-2**
graphminspantree will be removed **4-2**
graphpred2path will be removed **4-2**
graphshortestpath will be removed **4-3**
graphtopoorder will be removed **4-3**
graphtraverse will be removed **4-3**
'Depth' option for 'DFS' method is no longer supported **4-3**
biograph will be removed **4-3**
Full matrices are supported for most graph theory functions **4-4**
bamindexread will be removed **4-5**
blastlocal has been removed **4-5**
blastformat has been removed **4-5**
blastreadlocal has been removed **4-5**

R2021a

Functionality being removed or changed **5-2**
'TimeOut' option has been added **5-2**
showalignment has been removed **5-2**
seqshowwords has been removed **5-2**
blastlocal will be removed **5-2**
blastformat will be removed **5-2**
blastreadlocal will be removed **5-3**
NGS Browser has been removed **5-3**

R2020b

Burrows-Wheeler Aligner (BWA): Perform read alignments by mapping sequence reads to a reference genome **6-2**

Functionality being removed or changed **6-2**
blastlocal will be removed **6-2**
blastformat will be removed **6-2**
blastreadlocal will be removed **6-2**
Local BLAST functions are not supported for macOS version 10.15 or later **6-2**
showalignment will be removed **6-2**
seqshowwords will be removed **6-2**
molviewer will be removed **6-2**
Molecule Viewer app will be removed **6-2**
evalrasmolscript will be removed **6-2**

R2020a

Functionality being removed or changed **7-2**
showalignment will be removed **7-2**
seqshowwords will be removed **7-2**

R2019b

Genomics Viewer: View NGS sequences and annotations **8-2**

seqqcplot: View NGS sequence and quality data from SAM or BAM files and BioRead or BioMap objects **8-2**

Sort SAM and BAM files **8-2**

Functionality being removed or changed **8-2**
NGS Browser will be removed **8-2**
BioReadQualityStatistics has been removed **8-2**
plotSummary has been removed **8-2**

R2019a

Cufflinks: Perform statistical and differential expression analysis on RNA-sequence data **9-2**

Functionality being removed or changed **9-2**
getgenbank and getgenpept no longer append data to existing file **9-2**
NGS Browser will be removed **9-2**

R2018b

Next-Generation Sequencing: Count the number of reads in BAM-formatted files using the featurecount function **10-2**

R2018a

Next-Generation Sequencing: Import BAM files that use the new Concise Idiosyncratic Gapped Alignment Report (CIGAR) format **11-2**

Next-Generation Sequencing: Map reads using the Bowtie2 software **11-2**

R2017b

NCBI Database: Download genomic data from NCBI using XML format **12-2**
blastncbi Updates **12-2**
getblast Updates **12-2**
blastread Updates **12-2**

Functionality being changed or removed **12-2**
'psiblast' BLAST program has been removed **12-2**

'Inclusion' option in blastncbi has been removed 12-3
'Descriptions' option in blastncbi has been removed 12-3
'Alignments' option in blastncbi has been removed 12-3
'GapOpen' option in blastncbi has been removed 12-3
'ExtendGap' option in blastncbi has been removed 12-3
'Pct' option in blastncbi has been removed 12-3
'Alignments' option in getblast has been removed 12-3
'Descriptions' option in getblast has been removed 12-3
'FileFormat' option in getblast has been removed 12-3
'R2012b' option in seqalignviewer has been removed 12-3
knnclassify has been removed 12-4
featuresparse has been removed 12-4
featuresmap has been removed 12-4
princomp method of DataMatrix object has been renamed 12-4
seqlinkage correctly computes the input pairwise distances 12-4

R2017a

Next-Generation Sequencing QC Plots: Generate quality control plots from FASTQ data 13-2

Functionality being changed or removed 13-2

R2016b

Next Generation Sequencing: Preprocess NGS data with seqfilter, seqtrim, seqsplit, and seqsplitpe functions 14-2

Functionality being changed or removed 14-2

R2016a

featurecount: Summarize sequence reads for large NGS datasets 15-2

Functionality being changed or removed 15-2

R2015b

Functionality being changed or removed 16-2

R2015a

Bug Fixes

R2014b

Small sample unpaired hypothesis tests for count data 18-2

Functions for navigating the Gene Transfer Format (GTF) hierarchy to assist with alternative gene splicing and isoform analyses 18-2

Attractor metagene algorithm for feature engineering using mutual information-based clustering 18-2

Functionality being changed or removed 18-2

R2014a

Functionality being changed or removed 19-2

R2013b

Functionality being changed or removed 20-2

R2013a

Saving to FASTQ, FASTA, SAM, and BAM files from a BioMap object . . . 21-2

Sorting unordered BAM files using BioMap objects 21-2

Visualize and explore quality statistics of unmapped short-read data using BioRead and BioReadQualityStatistics objects 21-2

Select and move behaviors in the Sequence Alignment app 21-2

Random access of annotation object data, for consistency with BioMap object data access 21-2

Functionality being changed 21-3

R2012b

Multiple reference sequences in BioMap objects 22-2

Mapping single and paired-end short read data to reference genomes 22-2

Increased performance when reading BAM files 22-2

Name changes for multialignviewer, phytreetool, and seqtool tools . . . 22-3

Method removed from neighbor-joining method for phylogenetic tree construction 22-3

Functionality being changed or removed 22-3

R2012a

Update to Jmol Functions 23-2

Enhancements to Objects for NGS Data 23-2

Enhancements to the NGS Browser 23-3

Sequence Statistics Functions 23-3

Demo for DNA Methylation Analysis 23-3

Functionality Being Changed or Removed 23-3

R2011b

Visualizing and Investigating Short-Read Alignments and Feature Annotations in the NGS Browser 24-2

Objects for Genomic Feature Annotations 24-2

Enhancements to BioRead and BioMap Objects 24-2
Enhancements to the saminfo and baminfo Functions 24-3
Conversion of Error and Warning Message Identifiers 24-3
Function Elements Being Removed 24-4

R2011a

Data Format and Database Functions 25-2
Sequence Statistics Functions 25-2
Updates to the BioIndexedFile Class, Properties, and Methods 25-2
Updates to BioRead and BioMap Classes and Methods 25-2
Demos for High-Throughput Sequence Analysis 25-3
Support Vector Machine (SVM) Functions 25-3
Function Elements Being Removed 25-4

R2010b

Data Format and Database Functions 26-2
Sequence Conversion Functions 26-2
Sequence Statistics Functions 26-2
Pairwise Sequence Alignment Functions 26-3
Multiple Sequence Alignment Functions 26-3
Updates to BioMap Class, Methods, and Properties 26-4
Function Elements Being Removed 26-4

R2010a

Data Format and Database Functions 27-2

Pairwise Sequence Alignment Functions 27-2

Multiple Sequence Alignment Functions 27-2

Phylogenetic Tree Tools and Methods 27-3

BioIndexedFile Function, Object, Methods, and Properties 27-3

BioRead Function, Object, Methods, and Properties 27-3

BioMap Function, Object, Methods, and Properties 27-3

Function Elements Being Removed 27-3

R2009b

Data Format and Database Functions 28-2

Protein Analysis Functions 28-2

Data Visualization Functions 28-2

Sequence Statistics Functions 28-3

Sequence Utility Functions 28-3

Sequence Visualization Functions 28-3

Pairwise Sequence Alignment Functions 28-3

Multiple Sequence Alignment Functions 28-3

Phylogenetic Tree Tools and Methods 28-4

Clustergram Window 28-4

Clustergram Methods and Properties 28-4

HeatMap Object, Methods, and Properties 28-5

DataMatrix Methods 28-6

Microarray Functions, Objects, Methods, and Properties 28-6

Mass Spectrometry Functions 28-6

Demos for Sequence Analysis 28-6

Demos for Microarray Analysis 28-7

R2009a

Data Visualization Functions 29-2

Sequence Utility Functions 29-2

Sequence Conversion Functions 29-2

Bioanalytic and Mass Spectrometry Functions 29-2

Microarray Functions 29-2

Demo for Sequence Analysis 29-3

R2008b

Data Format and Database Functions 30-2

Sequence Utility Functions 30-2

Multiple Sequence Alignment Functions 30-3

Gene Ontology Functions 30-3

Protein Analysis Functions 30-3

Mass Spectrometry Functions 30-3

Microarray File Format Functions 30-4

Microarray Functions 30-4

DataMatrix Object 30-5

DataMatrix Methods 30-5

Demo for Visualization Tools 30-5

Demo for Sequence Analysis 30-5

Demos for Microarray Data Analysis 30-5

R2008a

Data Format and Database Functions 31-2
Sequence Utility Functions 31-2
Pairwise Sequence Alignment Functions 31-3
Phylogenetic Tree Tools Function 31-3
Protein Analysis Functions 31-3
Microarray File Format Functions 31-3
Microarray Functions 31-3
Object 31-5
Clustergram Methods 31-5
Demo for Sequence Analysis 31-5
Demo for Microarray Data Analysis 31-5
Demo for Visualization Tools 31-5
Demos for Mass Spectrometry Data Analysis 31-5

R2007b

Data Format and Database Functions 32-2
Microarray File Format Functions 32-2
Microarray Functions 32-2
Sequence Conversion, Utility, and Visualization Functions 32-2
Mass Spectrometry Functions 32-3
Statistical Learning Functions 32-3
Gene Ontology Methods 32-3

Demos for Microarray Data Analysis 32-3

Demos for Sequence Analysis 32-4

Demo for Graph Theory Analysis 32-4

R2007a+

Data Formats and Databases Functions 33-2

Microarray File Formats Functions 33-2

Microarray Utility Functions 33-2

Microarray Normalization and Filtering Functions 33-3

Mass Spectrometry Functions 33-3

Demos for Mass Spectrometry Functions 33-3

R2007a

Data Formats and Database Functions 34-2

Demo for Data Formats and Database Functions 34-2

Statistical Learning Functions 34-2

Protein Analysis and Sequence Utilities Functions 34-3

Demo for Protein Analysis and Sequence Utilities Functions 34-3

Sequence Alignment Functions 34-3

Demo for Sequence Alignment Functions 34-4

Microarray File Formats Functions 34-4

Microarray Normalization and Filtering Functions 34-4

Demo for Microarray File Formats, Normalization, and Filtering Functions 34-4

Microarray Data Analysis and Visualization Functions 34-4

Demo for Microarray Data Analysis and Visualization Functions 34-5

Mass Spectrometry Functions 34-5
Phylogenetic Tree Tools Functions 34-5
Demos for Phylogenetic Tree Tools Functions 34-6
Phylogenetic Tree Methods 34-6

R2006b

Data Formats and Database Functions 35-2
Sequence Utilities Functions 35-2
Sequence Visualization Functions 35-2
Multiple Sequence Alignment Functions 35-2
Microarray File Formats 35-2
Microarray Data Analysis and Visualization Functions 35-3
Graph Theory Functions 35-3
Graph Visualization Methods 35-3
Phylogenetic Tree Methods 35-4

R2006a+

Data Formats and Databases Functions 36-2
Sequence Utilities Functions 36-2
Sequence Visualization Functions 36-2
Statistical Learning Functions 36-2
Microarray Functions 36-2
Demo for Microarray Functions 36-3

R2006a

No New Features or Changes

R14SP3+

Multiple Sequence Alignment Viewer 38-2
Microarray Functions for Agilent Software 38-2
Gene Ontology Database Functions 38-2
Demo for Gene Ontology Functions 38-2

R14SP3

No New Features or Changes

R14SP2+

Sequence Alignment Functions 40-2
Sequence Statistics Functions 40-2
Sequence Utilities Functions 40-2
Phylogenetic Tree Functions 40-2
Phylogenetic Tree Methods 40-2
Microarray Functions 40-2
Statistics Functions 40-2

Updated RBASE Table **41-2**

Expanded Bioperl Demonstration **41-2**

R2023a

Version: 4.17

New Features

Bug Fixes

Compatibility Considerations

Bioinformatics Pipeline

A bioinformatics pipeline is a set of connected algorithms (or blocks) that are executed in a predefined order to process and analyze next-generation sequencing (NGS) data locally or in cluster environments. You can build such pipelines interactively using an app or programmatically.

The toolbox provides several built-in blocks that perform specific tasks, such as filtering genomic reads or mapping reads to a reference genome. In addition, you can also convert any custom function into a block that you can use in your analysis pipeline. Using a combination of built-in blocks and custom blocks enables you to build various analysis workflows and pipelines.

For more details about the app and examples, see **Biopipeline Designer**. For command line workflows and examples, see `bioinfo.pipeline.Pipeline`.

Overwrite Option for seqtrim and seqfilter

`seqtrim` and `seqfilter` now have the `OverWrite` option to allow you to overwrite any existing files. By default, the value is `false`.

Compatibility Considerations

However, the `bioinfo.pipeline.blocks.SeqTrim` and `bioinfo.pipeline.blocks.SeqFilter` blocks always overwrite existing files.

Functionality being removed or changed

probesetlink has been removed
Errors

`probesetlink` has been removed.

display name-value argument of pdbsuperpose will be removed in a future release
Warns

The `display` name-value argument of `pdbsuperpose` will be removed in a future release.

pdbsuperpose no longer opens Molecule Viewer
Behavior change

`pdbsuperpose` no longer opens the **Molecule Viewer** app to display the superposed 3-D protein structures. In other words, the default value of the `display` name-value argument is now `false`.

molviewer and Molecule Viewer app will be removed
Warns

`molviewer` and the corresponding **Molecule Viewer** app will be removed in a future release.

bowtie has been removed
Errors

`bowtie` has been removed. Use `bowtie2` instead.

bowtiebuild has been removed

Errors

`bowtiebuild` has been removed. Use `bowtie2build` instead.

2

R2022b

Version: 4.16.1

Bug Fixes

Compatibility Considerations

Functionality being removed or changed

bowtie will be removed

Warns

`bowtie` will be removed in a future release. Use `bowtie2` instead.

bowtiebuild will be removed

Warns

`bowtiebuild` will be removed in a future release. Use `bowtie2build` instead.

graphallshortestpaths has been removed

Errors

`graphallshortestpaths` has been removed. Use `distances` instead.

graphconncomp has been removed

Errors

`graphconncomp` has been removed. Use `conncomp` instead.

graphisdiag has been removed

Errors

`graphisdag` has been removed. Use `isdag` instead.

graphisomorphism has been removed

Errors

`graphisomorphism` has been removed. Use `isomorphism` instead.

graphisspantree has been removed

Errors

`graphisspantree` has been removed. A graph is a spanning tree if and only if all nodes are reachable from an arbitrary start node and `E == N-1`, where *E* is the number of edges and *N* is the number of nodes. You can use either `bfsearch` or `dfsearch` to check if such conditions are true for a given graph.

graphmaxflow has been removed

Errors

`graphmaxflow` has been removed. Use `maxflow` instead. Note that `maxflow` returns only one solution while `graphmaxflow` returns multiple solutions if they exist.

graphminspantree has been removed

Errors

`graphminspantree` has been removed. Use `minspantree` instead.

graphpred2path has been removed

Errors

`graphpred2path` has been removed.

graphshortestpath has been removed

Errors

`graphshortestpath` has been removed. Use `shortestpath` or `shortestpathtree` instead.

graphtopoorder has been removed

Errors

`graphtopoorder` has been removed. Use `toposort` instead.

graphtraverse has been removed

Errors

`graphtraverse` has been removed. Use `bfsearch` or `dfsearch` instead. Note that `bfsearch` and `dfsearch` do not have the `'Depth'` option. An example of an alternative approach is as follows:

```
s = [1 1 2 2 2 3 4 7 8 8 8 8];
t = [3 4 7 5 6 2 6 2 9 10 11 12];
g = digraph(s,t);
v = bfsearch(g,1);
d = distances(g,1,v);
maxDepth = 2;
v2 = v(d <= maxDepth);
```

biograph has been removed

Errors

The `biograph object` and its methods have been removed. See the following table for alternative functions from `graph` and `digraph` objects.

Methods of biograph	Alternative Functions from graph and digraph Objects
`allshortestpaths (biograph)`	`distances`
`conncomp (biograph)`	`conncomp`
`dolayout (biograph)`	Not applicable
`getancestors (biograph)`	`predecessors`. Note that the function does not let you specify `NumGenerations`. You can use `distances` to find the distances from all nodes to the target node after setting the edge weight to 1, and filter the nodes by distance to the appropriate depth.
`getdescendants (biograph)`	`successors`. If you need to specify `NumGenerations`, see the preceding workaround for `getancestors`.
`getedgesbynodeid (biograph)`	`findedge`
`getmatrix (biograph)`	`adjacency`
`getnodesbyid (biograph)`	`findnode`
`getrelatives (biograph)`	`neighbors`. If you need to specify `NumGenerations`, see the preceding workaround for `getancestors`.
`getweightmatrix (biograph)`	`adjacency`

Methods of biograph	Alternative Functions from graph and digraph Objects
`isdag (biograph)`	`isdag`
`isomorphism (biograph)`	`isomorphism`
`isspantree (biograph)`	A graph is a spanning tree if and only if all nodes are reachable from an arbitrary start node, and `E == N-1`, where *E* is the number of edges and *N* is the number of nodes. You can use either `bfsearch` or `dfsearch` to check if such conditions are true for a given graph.
`maxflow (biograph)`	`maxflow`
`minspantree (biograph)`	`minspantree`
`shortestpath (biograph)`	`shortestpath`
`topoorder (biograph)`	`toposort`
`traverse (biograph)`	`bfsearch` or `dfsearch`
`view (biograph)`	`plot`. The name-value arguments of the `plot` function can be used to control the properties of the graph. The arguments can be used as replacements for many of the `biograph` object properties to change the appearance of the graph.

R2022a

Version: 4.16

New Features

Bug Fixes

Compatibility Considerations

Windows support for Bowtie 2, BWA, and Cufflinks support packages

You can now run the following Bioinformatics Toolbox Software Support Packages after you install Windows® Subsystem for Linux® (WSL) and a Linux distribution on your Windows computer.

- Bowtie 2 Support Package for Bioinformatics Toolbox™
- Cufflinks Support Package for the Bioinformatics Toolbox
- BWA Support Package for Bioinformatics Toolbox

For details on installing WSL, see here.

Compatibility Considerations

The previous support package Bioinformatics Toolbox Interface for Bowtie Aligner is only compatible with R2021b or earlier versions of MATLAB®. Starting in R2022a, install the new Bowtie 2 Support Package for Bioinformatics Toolbox (download link) instead to run `bowtie2`, `bowtie2build`, or `bowtie2inspect` in MATLAB.

Functionality being removed or changed

Bioinformatics Toolbox Interface for Bowtie Aligner support package has been replaced

The support package Bioinformatics Toolbox Interface for Bowtie Aligner has been replaced. It is only compatible for R2021b or earlier versions of MATLAB. Staring in R2022a, install the new support package Bowtie 2 Support Package for Bioinformatics Toolbox (download link) instead.

bowtie will be removed
Still runs

`bowtie` will be removed in a future release. Use `bowtie2` instead.

bowtiebuild will be removed
Still runs

`bowtiebuild` will be removed in a future release. Use `bowtie2build` instead.

graphallshortestpaths will be removed
Warns

`graphallshortestpaths` will be removed in a future release. Use `distances` instead.

graphconncomp will be removed
Warns

`graphconncomp` will be removed in a future release. Use `conncomp` instead.

graphisdiag will be removed
Warns

`graphisdag` will be removed in a future release. Use `isdag` instead.

graphisomorphism will be removed
Warns

`graphisomorphism` will be removed in a future release. Use `isomorphism` instead.

graphisspantree will be removed

Warns

`graphisspantree` will be removed in a future release. A graph is a spanning tree if and only if all nodes are reachable from an arbitrary start node and `E == N-1`, where *E* is the number of edges and *N* is the number of nodes. You can use either `bfsearch` or `dfsearch` to check if such conditions are true for a given graph.

graphmaxflow will be removed

Warns

`graphmaxflow` will be removed in a future release. Use `maxflow` instead. Note that `maxflow` returns only one solution while `graphmaxflow` returns multiple solutions if they exist.

graphminspantree will be removed

Warns

`graphminspantree` will be removed in a future release. Use `minspantree` instead.

graphpred2path will be removed

Warns

`graphpred2path` will be removed in a future release.

graphshortestpath will be removed

Warns

`graphshortestpath` will be removed in a future release. Use `shortestpath` or `shortestpathtree` instead.

graphtopoorder will be removed

Warns

`graphtopoorder` will be removed in a future release. Use `toposort` instead.

graphtraverse will be removed

Warns

`graphtraverse` will be removed in a future release. Use `bfsearch` or `dfsearch` instead. Note that `bfsearch` and `dfsearch` do not have the `'Depth'` option. An example of an alternative approach is as follows:

```
s = [1 1 2 2 2 3 4 7 8 8 8 8];
t = [3 4 7 5 6 2 6 2 9 10 11 12];
g = digraph(s,t);
v = bfsearch(g,1);
d = distances(g,1,v);
maxDepth = 2;
v2 = v(d <= maxDepth);
```

biograph will be removed

Warns

The `biograph object` and its methods will be removed in a future release. See the following table for alternative functions from `graph` and `digraph` objects.

Methods of biograph	Alternative Functions from graph and digraph Objects
`allshortestpaths (biograph)`	`distances`
`conncomp (biograph)`	`conncomp`
`dolayout (biograph)`	Not applicable
`getancestors (biograph)`	`predecessors`. Note that the function does not let you specify `NumGenerations`. You can use `distances` to find the distances from all nodes to the target node after setting the edge weight to 1, and filter the nodes by distance to the appropriate depth.
`getdescendants (biograph)`	`successors`. If you need to specify `NumGenerations`, see the preceding workaround for `getancestors`.
`getedgesbynodeid (biograph)`	`findedge`
`getmatrix (biograph)`	`adjacency`
`getnodesbyid (biograph)`	`findnode`
`getrelatives (biograph)`	`neighbors`. If you need to specify `NumGenerations`, see the preceding workaround for `getancestors`.
`getweightmatrix (biograph)`	`adjacency`
`isdag (biograph)`	`isdag`
`isomorphism (biograph)`	`isomorphism`
`isspantree (biograph)`	A graph is a spanning tree if and only if all nodes are reachable from an arbitrary start node, and `E == N-1`, where E is the number of edges and N is the number of nodes. You can use either `bfsearch` or `dfsearch` to check if such conditions are true for a given graph.
`maxflow (biograph)`	`maxflow`
`minspantree (biograph)`	`minspantree`
`shortestpath (biograph)`	`shortestpath`
`topoorder (biograph)`	`toposort`
`traverse (biograph)`	`bfsearch` or `dfsearch`
`view (biograph)`	`plot`. The name-value arguments of the `plot` function can be used to control the properties of the graph. The arguments can be used as replacements for many of the `biograph` object properties to change the appearance of the graph.

bamindexread has been removed

Errors

`bamindexread` has been removed.

R2021b

Version: 4.15.2
New Features
Bug Fixes
Compatibility Considerations

rnaseqde: Perform differential expression analysis on RNA-seq count data

Identify which features or genes are expressed differentially between RNA-seq samples using the `rnaseqde` function. The function uses an exact test to determine the differences between two groups of counts that are assumed to follow the negative binomial distribution. It provides different ways to model the variance of each feature or gene.

Functionality being removed or changed

graphallshortestpaths will be removed
Still runs

`graphallshortestpaths` will be removed in a future release. Use `distances` instead.

graphconncomp will be removed
Still runs

`graphconncomp` will be removed in a future release. Use `conncomp` instead.

graphisdiag will be removed
Still runs

`graphisdag` will be removed in a future release. Use `isdag` instead.

graphisomorphism will be removed
Still runs

`graphisomorphism` will be removed in a future release. Use `isomorphism` instead.

graphisspantree will be removed
Still runs

`graphisspantree` will be removed in a future release. A graph is a spanning tree if and only if all nodes are reachable from an arbitrary start node, and `E == N-1`, where E is the number of edges and N is the number of nodes. You can use either `bfsearch` or `dfsearch` to check if such conditions are true for a given graph.

graphmaxflow will be removed
Still runs

`graphmaxflow` will be removed in a future release. Use `maxflow` instead. Note that `maxflow` returns only one solution while `graphmaxflow` returns multiple solutions if they exist.

graphminspantree will be removed
Still runs

`graphminspantree` will be removed in a future release. Use `minspantree` instead.

graphpred2path will be removed
Still runs

`graphpred2path` will be removed in a future release.

graphshortestpath will be removed

Still runs

`graphshortestpath` will be removed in a future release. Use `shortestpath` or `shortestpathtree` instead.

graphtopoorder will be removed

Still runs

`graphtopoorder` will be removed in a future release. Use `toposort` instead.

graphtraverse will be removed

Still runs

`graphtraverse` will be removed in a future release. Use `bfsearch` or `dfsearch` instead. Note that `bfsearch` and `dfsearch` do not have the `'Depth'` option. An example of an alternative approach would be as follows:

```
s = [1 1 2 2 2 3 4 7 8 8 8 8];
t = [3 4 7 5 6 2 6 2 9 10 11 12];
g = digraph(s,t);
v = bfsearch(g,1);
d = distances(g,1,v);
maxDepth = 2;
v2 = v(d <= maxDepth);
```

'Depth' option for 'DFS' method is no longer supported

Behavior change

For `graphtraverse`, the `'Depth'` name-value argument is no longer supported for the `'DFS'` method.

biograph will be removed

Still runs

The `biograph object` and its methods will be removed in a future release. See the following table for alternative functions from `graph` and `digraph` objects.

Methods of `biograph`	**Alternative Functions from `graph` and `digraph` Objects**
`allshortestpaths (biograph)`	`distances`
`conncomp (biograph)`	`conncomp`
`dolayout (biograph)`	Not applicable
`getancestors (biograph)`	`predecessors`. Note that the function does not let you specify `NumGenerations`. You can use `distances` to find the distances from all nodes to the target node after setting the edge weight to 1, and filter the nodes by distance to the appropriate depth.
`getdescendants (biograph)`	`successors`. If you need to specify `NumGenerations`, see the preceding workaround for `getancestors`.
`getedgesbynodeid (biograph)`	`findedge`

Methods of biograph	**Alternative Functions from graph and digraph Objects**
`getmatrix (biograph)`	`adjacency`
`getnodesbyid (biograph)`	`findnode`
`getrelatives (biograph)`	`neighbors`. If you need to specify `NumGenerations`, see the preceding workaround for `getancestors`.
`getweightmatrix (biograph)`	`adjacency`
`isdag (biograph)`	`isdag`
`isomorphism (biograph)`	`isomorphism`
`isspantree (biograph)`	A graph is a spanning tree if and only if all nodes are reachable from an arbitrary start node, and `E == N-1`, where *E* is the number of edges and *N* is the number of nodes. You can use either `bfsearch` or `dfsearch` to check if such conditions are true for a given graph.
`maxflow (biograph)`	`maxflow`
`minspantree (biograph)`	`minspantree`
`shortestpath (biograph)`	`shortestpath`
`topoorder (biograph)`	`toposort`
`traverse (biograph)`	`bfsearch` or `dfsearch`
`view (biograph)`	`plot`. The name-value arguments of the `plot` function can be used to control the properties of the graph. The arguments can be used as replacements for many of the `biograph` object properties to change the appearance of the graph.

Full matrices are supported for most graph theory functions

Behavior change

The following graph theory functions now support full matrices in addition to sparse matrices as input.

- `graphallshortestpaths`
- `graphconncomp`
- `graphisdag`
- `graphisomorphism`
- `graphisspantree`
- `graphmaxflow`
- `graphminspantree`
- `graphshortestpath`
- `graphtopoorder`
- `graphtraverse`

`graphpred2path` is the only function that does not support full matrices.

bamindexread will be removed

Warns

`bamindexread` will be removed in a future release.

blastlocal has been removed

Errors

`blastlocal` has been removed.

blastformat has been removed

Errors

`blastformat` has been removed.

blastreadlocal has been removed

Errors

`blastreadlocal` has been removed.

5

R2021a

Version: 4.15.1

Bug Fixes

Compatibility Considerations

Functionality being removed or changed

'TimeOut' option has been added

Behavior change

A new name-value argument `'TimeOut'` has been added to the following functions. Use this option to set the maximum connection timeout to retrieve data remotely.

- `blastncbi`
- `emblread`
- `fastainfo`
- `fastaread`
- `genbankread`
- `genpeptread`
- `getblast`
- `getembl`
- `getgenbank`
- `getgenpept`
- `gethmmalignment`
- `gethmmprof`
- `gethmmtree`
- `getpdb`
- `multialignread`
- `pdbread`
- `pfamhmmread`
- `phytreeread`
- `geoseriesread`
- `geosoftread`
- `getgeodata`

showalignment has been removed

Errors

`showalignment` has been removed.

seqshowwords has been removed

Errors

`seqshowwords` has been removed.

blastlocal will be removed

Warns

`blastlocal` will be removed in a future release.

blastformat will be removed

Warns

`blastformat` will be removed in a future release.

blastreadlocal will be removed
Warns

`blastreadlocal` will be removed in a future release.

NGS Browser has been removed
Errors

The **NGS Browser** app and `ngsbrowser` function have been removed. Use **Genomics Viewer** instead.

6

R2020b

Version: 4.15

New Features

Bug Fixes

Compatibility Considerations

Burrows-Wheeler Aligner (BWA): Perform read alignments by mapping sequence reads to a reference genome

You can map sequence reads to a reference genome using BWA. Build BWA index files using `bwaindex` and align reads to a reference sequence using `bwamem`. Both functions require the BWA Support Package for Bioinformatics Toolbox. You can download the support package from the Add-On Explorer. For details, see Bioinformatics Toolbox Software Support Packages.

Functionality being removed or changed

blastlocal will be removed

Still runs

`blastlocal` will be removed in a future release.

blastformat will be removed

Still runs

`blastformat` will be removed in a future release.

blastreadlocal will be removed

Still runs

`blastreadlocal` will be removed in a future release.

Local BLAST functions are not supported for macOS version 10.15 or later

Behavior change

`blastlocal`, `blastreadlocal`, and `blastformat` are not supported for macOS version 10.15 (Catalina) or later.

showalignment will be removed

Warns

`showalignment` will be removed in a future release.

seqshowwords will be removed

Warns

`seqshowwords` will be removed in a future release.

molviewer will be removed

Still runs

`molviewer` will be removed in a future release.

Molecule Viewer app will be removed

Still runs

The **Molecule Viewer** app will be removed in a future release.

evalrasmolscript will be removed

Still runs

`evalrasmolscript` will be removed in a future release.

7

R2020a

Version: 4.14

Bug Fixes

Compatibility Considerations

Functionality being removed or changed

showalignment will be removed

Still runs

The `showalignment` function will be removed in a future release.

seqshowwords will be removed

Still runs

The `seqshowwords` function will be removed in a future release.

R2019b

Version: 4.13

New Features

Bug Fixes

Compatibility Considerations

Genomics Viewer: View NGS sequences and annotations

The **Genomics Viewer** app lets you view and explore integrated genomic data with an embedded version of the Integrative Genomics Viewer (IGV). The genomic data include NGS read alignments, genome variants, and segmented copy number data.

seqqcplot: View NGS sequence and quality data from SAM or BAM files and BioRead or BioMap objects

`seqqcplot` now accepts the following file types (in addition to FASTQ files) and objects.

- SAM files
- BAM files
- `BioRead` object
- `BioMap` object

Sort SAM and BAM files

You can sort SAM and BAM files using the `samsort` and `bamsort` functions. The functions sort the alignment records by the reference sequence name first and then by position within the reference.

Functionality being removed or changed

NGS Browser will be removed

Warns

The **NGS Browser** app and `ngsbrowser` function will be removed in a future release. Use **Genomics Viewer** instead.

BioReadQualityStatistics has been removed

Errors

The `BioReadQualityStatistics` object has been removed. Use `seqqcplot` instead to generate quality control plots for sequence data.

plotSummary has been removed

Errors

The `plotSummary` function has been removed. Use `seqqcplot` instead to generate quality control plots for sequence data.

R2019a

Version: 4.12

New Features

Bug Fixes

Compatibility Considerations

Cufflinks: Perform statistical and differential expression analysis on RNA-sequence data

You can now assemble transcriptomes and perform differential expression analysis on RNA-Seq data using a suite of cufflinks functions, including `cufflinks`, `cuffcompare`, `cuffmerge`, `cuffquant`, `cuffdiff`, and `cuffnorm`. In addition, you can use `cuffgffread` to filter and convert GFF and GTF files. `cuffgtf2sam` lets you convert the assembled transcripts in the input GTF files to SAM files. These functions require the Cufflinks Support Package for Bioinformatics Toolbox. You can download the support package from the Add-On Explorer.

Functionality being removed or changed

getgenbank and getgenpept no longer append data to existing file

Behavior change

`getgenbank` and `getgenpept` no longer append data to an existing file. The functions now overwrite the contents of the existing file without warning.

NGS Browser will be removed

Still runs

The **NGS Browser** app and `ngsbrowser` function will be removed in a future release.

R2018b

Version: 4.11

New Features

Bug Fixes

Next-Generation Sequencing: Count the number of reads in BAM-formatted files using the featurecount function

The `featurecount` function lets you count the number of reads from BAM-formatted and SAM-formatted files.

11

R2018a

Version: 4.10

New Features

Bug Fixes

Next-Generation Sequencing: Import BAM files that use the new Concise Idiosyncratic Gapped Alignment Report (CIGAR) format

You can now import BAM- and SAM-formatted files that contain `=` and `X` characters in the CIGAR strings. Use the BioMap function to import.

Next-Generation Sequencing: Map reads using the Bowtie2 software

You can now map sequence reads to a reference sequence using the `bowtie2` function, which requires the Bioinformatics Toolbox Interface for Bowtie Aligner support package. You can download this support package from the Add-On Explorer.

12

R2017b

Version: 4.9

New Features

Bug Fixes

Compatibility Considerations

NCBI Database: Download genomic data from NCBI using XML format

Access the NCBI BLAST web services using improved `blastncbi`, `getblast`, and `blastread` functions.

blastncbi Updates

New name-value pair arguments are:

- `'MaxNumberSequences'`– Specifies the maximum number of hits to return. This argument replaces the previous `'Alignments'` and `'Descriptions'` name-value pairs.
- `'MatchScores'`– Defines the matching and mismatching scores in a nucleotide alignment (for `blastn` and `megablast` only).
- `'CompositionAdjustment'`– Specifies the type of compositional adjustment to apply to compensate for the amino acid compositions of sequences being compared.

Updated name-value pair arguments are:

- `'GapCosts'`– This argument is now available for both amino acid and nucleotide searches.
- `'Word'`– Available word sizes have been updated.
- `'Filter'`– You can now specify multiple filters at once. For example, `'Lm'` applies both the low compositional complexity filter and the mask.
- `'Database'`– Available database choices have been updated.

For the list of name-value pair arguments that have been removed from `blastncbi`, see “Functionality being changed or removed” on page 12-2.

getblast Updates

`getblast` lets you retrieve the BLAST search results and save the report in an XML-formatted file using the `'ToFile'` option. This function returns an output structure that contains the BLAST report data with the following fields: `RID`, `Algorithm`, `Database`, `QueryID`, `QueryDefinition`, `Hits`, `Parameters`, and `Statistics`.

For the list of name-value pair arguments that have been removed from `getblast` functions, see “Functionality being changed or removed” on page 12-2.

blastread Updates

`blastread` now takes only an XML-formatted BLAST report file as an input. The support for reading from URLs and other file types has been removed.

The output structure has new fields: `QueryDefinition`, `QueryID` (replaces `Query`), `Parameters`, and `HSPs.AlignmentLength`.

Functionality being changed or removed

'psiblast' BLAST program has been removed

Errors

The BLAST program `'psiblast'` has been removed from one of supported programs by `blastncbi`.

'Inclusion' option in blastncbi has been removed

Errors

The `'Inclusion'` name-value pair for `blastncbi` has been removed since it only applies to the `psiblast` program which has been also removed.

'Descriptions' option in blastncbi has been removed

Errors

The `'Descriptions'` name-value pair for `blastncbi` has been removed. Use `'MaxNumberSequences'` instead to specify the maximum number of hits to return.

'Alignments' option in blastncbi has been removed

Errors

The `'Alignments'` name-value pair for `blastncbi` has been removed. Use `'MaxNumberSequences'` instead to specify the maximum number of hits to return.

'GapOpen' option in blastncbi has been removed

Errors

The `'GapOpen'` name-value pair for `blastncbi` has been removed. Use `'GapCosts'` instead.

'ExtendGap' option in blastncbi has been removed

Errors

The `'ExtendGap'` name-value pair for `blastncbi` has been removed. Use `'GapCosts'` instead.

'Pct' option in blastncbi has been removed

Errors

The `'Pct'` name-value pair for `blastncbi` has been removed.

'Alignments' option in getblast has been removed

Errors

The `'Alignments'` name-value pair for `getblast` has been removed. The number of hits returned in the output is controlled by the number of hits in the input BLAST report.

'Descriptions' option in getblast has been removed

Errors

The `'Descriptions'` name-value pair for `getblast` has been removed. The number of hits returned in the output is controlled by the number of hits in the input BLAST report.

'FileFormat' option in getblast has been removed

Errors

The `'FileFormat'` name-value pair for `getblast` has been removed. The file is XML-formatted automatically.

'R2012b' option in seqalignviewer has been removed

Errors

The `'R2012b'` name-value pair for `seqalignviewer`has been removed. The default version of `seqalignviewer` runs more robustly than the previous version (R2012b).

knnclassify has been removed

Errors

The `knnclassify` function has been removed. Use `fitcknn` to fit a `knn` classification model and classify data using the `predict` function of `ClassificationKNN` object.

featuresparse has been removed

Errors

The `featuresparse` function has been removed. Use `featureparse` instead.

featuresmap has been removed

Errors

The `featuresmap` function has been removed. Use `featureview` instead.

princomp method of DataMatrix object has been renamed

Errors

The `princomp` method of `DataMatrix object` has been renamed. Replace instances of `princomp` with `pca`.

seqlinkage correctly computes the input pairwise distances

Behavior change

For the R2017a or earlier versions, `seqlinkage` incorrectly doubled the input pairwise distances when building a tree. This bug has been fixed in R2017b.

If you have been previously `selecting` a subset of the tree returned by `seqlinkage` with a distance threshold, consider dividing the threshold by 2.

Note that the tree topology has always been computed correctly and not affected by this bug.

13

R2017a

Version: 4.8

New Features

Bug Fixes

Compatibility Considerations

Next-Generation Sequencing QC Plots: Generate quality control plots from FASTQ data

You can access the quality of FASTQ data using the `seqqcplot` function. The function provides several quality control plots such as the boxplot of average quality score, line plot of sequence base composition, quality score distribution plot, GC distribution plot, and sequence length distribution plot. You can specify the trimming and filtering criteria and access the data quality of only those sequences that meet the criteria without modifying the original data.

Functionality being changed or removed

Functionality	Result	Use This Instead	Compatibility Considerations
`BioReadQualityStatistics` object	Warns	Use `seqqcplot` instead to generate quality control plots for sequence data.	This object will be removed in a future release.
`getCounts`	—	—	The method now returns a vector or cell array depending on whether multiple references are specified.
`'alu_repeats'` as a database to search using `blastncbi`	Errors	`'alu'`	The `'alu_repeats'` database option has been renamed to `'alu'`.

14

R2016b

Version: 4.7

New Features

Bug Fixes

Compatibility Considerations

Next Generation Sequencing: Preprocess NGS data with seqfilter, seqtrim, seqsplit, and seqsplitpe functions

You can preprocess next generation sequencing data using various functions. `seqfilter` lets you filter out sequences that do not meet specified criteria such as a maximum threshold on the number of low quality bases allowed, a minimum length, or a minimum average sequence quality. `seqtrim` lets you trim sequences at the sequence ends, or when the sequence quality goes below a given threshold. `seqsplit` splits the sequences in a FASTQ-file based on the corresponding barcodes allowing a specified number of mismatches. `seqsplitpe` splits merged paired-end sequences into two separate files.

Functionality being changed or removed

Functionality	Result	Use This Instead	Compatibility Considerations
The `RowLabelsColor` and `ColumnLabelsColor` properties of `HeatMap` and `clustergram` objects	Warns	Set `LabelsWithMarkers` to `true` for colored markers instead of colored texts.	These properties will be removed in a future release. Set `LabelsWithMarkers` to `true` for colored markers instead of colored texts. For details, see `HeatMap object`.
The `princomp` method of `DataMatrix object`	Warns	`pca`	Replace instances of `princomp` with `pca`.
The `'Type'` name-value pair argument of `gethmmtree`.	Errors	To download the `'seed'` tree, use `gethmmtree` without any extra input arguments. To obtain the `'full'` tree, you may use `gethmmalignment` to download the `'full'` alignment and build a tree using the `seqpdist` and `seqneighjoin` functions.	This name-value pair has been removed.
The `'Mirror'` name-value pair argument of `gethmmprof` and `gethmmalignment`	Errors	The function now searches the PFAM database at http://pfam.xfam.org/.	This name-value pair has been removed.
Calling the `biograph` function without any input arguments	Warns	To create a random biograph object, specify a random connectivity matrix as an input argument.	Calling the function without any input arguments returns a random biograph with 15 nodes. This behavior will change in a future release.

R2016a

Version: 4.6

New Features

Bug Fixes

Compatibility Considerations

featurecount: Summarize sequence reads for large NGS datasets

The `featurecount` function lets you count the number of reads from the next-generation sequencing data that map to genomic features of interest using SAM and GTF file inputs. You can summarize the sequence reads at the different feature levels such as exon, transcripts, or genes. The function supports stranded sequencing protocols as well as single-end and paired-end read counting. Various choices for multiple mapping reads are available, suitable for both RNA-Seq and CHIP-Seq data analysis, as well as different methods for read (or fragment) disambiguation.

Functionality being changed or removed

Functionality	Result	Use This Instead	Compatibility Considerations
The `Color` field of a structure for the `RowLabelsColor` and `ColumnLabelsColor` properties of `HeatMap` and `clustergram` objects	Still runs	Not applicable	Independent color support for each row or column label has been removed. If there are multiple colors, the default color (black) is used as the label text color. Set `LabelsWithMarkers` to `true` for colored markers instead of colored text. For details, see `HeatMap object`.
`featuresparse`	Warns	`featureparse`	The function has been renamed to `featureparse`.
`featuresmap`	Warns	`featureview`	The function has been renamed to `featureview`.
`affyinvarsetnorm`	Still runs	Not applicable	The function has been updated to realign with the behavior of `R2012b` or earlier releases.
`getTranscripts` and `getGenes` methods of the `GTFAnnotation` class	Still runs	Not applicable	These two methods have been updated to improve the computation time.

16

R2015b

Version: 4.5.2

Bug Fixes

Compatibility Considerations

Functionality being changed or removed

Functionality	Result	Use This Instead	Compatibility Considerations
`cleave`	Still runs	Not applicable	When cleaving a sequence using trypsin, the function now applies trypsin's exception rules by default. As a result, the default output may differ from earlier releases. To prevent the use of the default exceptions for trypsin, use an empty string as the exception rule when you run `cleave`. To see the exception rules, check the table listed in `cleavelookup`.
`knnimpute`	Still runs	Not applicable	The function now errors if the number of nearest neighbors (`k`) is not a positive scalar integer.

R2015a

Version: 4.5.1

Bug Fixes

R2014b

Version: 4.5

New Features

Bug Fixes

Compatibility Considerations

Small sample unpaired hypothesis tests for count data

You can perform an unpaired hypothesis test for count data (from high-throughput sequencing assays such as RNA-Seq or ChIP-Seq) with small numbers of samples or replicates using `nbintest`. For instance, you can use this function to decide if an observed difference in read counts between two conditions is significant for a given gene. The function assumes read counts follow a negative binomial or Poisson distribution.

Functions for navigating the Gene Transfer Format (GTF) hierarchy to assist with alternative gene splicing and isoform analyses

The following functions of the `GTFAnnotation` class help you navigate the GTF information hierarchy to perform alternative gene splicing and isoform analyses:

- `getSegments` returns a table of nonoverlapping segments built by flattening the transcripts.
- `getGenes` returns a table of unique genes referenced by exons.
- `getTranscripts` returns a table of unique transcripts referenced by exons.
- `getExons` returns a table of exons.

Attractor metagene algorithm for feature engineering using mutual information-based clustering

The `metafeatures` function uses the attractor metagene algorithm, which is an unsupervised learning algorithm for feature engineering using mutual information-based learning.

Functionality being changed or removed

Functionality	What Happens When You Use This Functionality?	Use This Instead	Compatibility Considerations
`knnclassify`	Still runs	`fitcknn`	Use `fitcknn` to fit a `knn` classification model and classify data using the `predict` function of `ClassificationKNN` object.

Functionality	What Happens When You Use This Functionality?	Use This Instead	Compatibility Considerations
Default values for the `knn` classifier of `randfeatures`	Still runs	—	When you specify `'knn'` as the classifier, `randfeatures` now uses the following new defaults. • The default function is `fitcknn`. • For the `'ClassOptions'` name-value pair argument, the defaults are `{'Distance','corelation','NumNeighbors',5}`. • For the `'PerformanceThreshold'` name-value pair argument, the default is `0.7`. • For the `'ConfidenceThreshold` name-value pair argument, the default is `1`.
The `'Type'` name-value pair argument of `gethmmtree`	Warns	To download the `'seed'` tree, use `gethmmtree` without any extra input arguments. To obtain the `'full'` tree, you may use the `gethmmalignment` function to download the `'full'` alignment and build a tree using the `seqpdist` and `seqneighjoin` functions.	Setting `'Type'` to `'seed'` or `'full'` is now ignored since the PFAM database no longer provides trees for the `'full'` alignment.

R2014a

Version: 4.4

Bug Fixes

Compatibility Considerations

Functionality being changed or removed

Functionality	What Happens When You Use This Functionality?	Use This Instead	Compatibility Considerations
`'R2012b'` name-value pair input argument for the `seqalignviewer` function	Warns	The default version of `seqalignviewer` runs more robustly than the previous version (`R2012b`), and the default version is recommended to use. This name-value pair is intended only for customers who need the previous version.	See the Compatibility Considerations subheading in “Select and move behaviors in the Sequence Alignment app” on page 21-2.
`bowtieread` function	Errors	- If you have the original FASTQ (or FASTA) file, use the `bowtie` function (for UNIX® and Mac users only) to remap your files. This will create BAM files that are compatible with the toolbox. - If you have old BOWTIE files without the sequence files, you can read the files using `textscan`.	When using other BOWTIE mapper/aligner programs, set appropriate option(s) to create either a SAM or BAM output file. Then use the `Biomap` object or the `samread` or `bamread` function to access the mapped short reads.

R2013b

Version: 4.3.1

Bug Fixes

Compatibility Considerations

Functionality being changed or removed

Functionality	What Happens When You Use This Functionality?	Use This Instead	Compatibility Considerations
`Index` name-value pair argument as input to the `bamread` function	Errors	—	Remove instances of the `Index` name-value pair argument. See the Compatibility Considerations subheading in “Increased performance when reading BAM files” on page 22-2.
`'average'` as a choice for the `Method` input argument to the `seqneighjoin` function	Errors	`'equivar'`	Replace instances of `'average'` as an input to `seqneighjoin` with `'equivar'`.
Changes to tool names: • `multialignviewer` • `phytreetool` • `seqtool`	Errors	• `seqalignviewer` • `phytreeviewer` • `seqviewer`	• Replace instances of `multialignviewer` with `seqalignviewer`. • Replace instances of `phytreetool` with `phytreeviewer`. • Replace instances of `seqtool` with `seqviewer`.
`'natural'` as a choice for the *`Output`* name-value pair input argument to these functions: • `affyrma` • `affygcrma` • `rmasummary`	Errors	`'linear'`	Replace instances of `'natural'` as the value of the *`Output`* name-value pair input argument with `'linear'` for these functions: • `affyrma` • `affygcrma` • `rmasummary`

21

R2013a

Version: 4.3

New Features

Bug Fixes

Compatibility Considerations

Saving to FASTQ, FASTA, SAM, and BAM files from a BioMap object

You can write the information of any `BioRead/BioMap` object to a file using the `write` function of the object.

Sorting unordered BAM files using BioMap objects

You can pass an unordered BAM file to a `BioMap` constructor, which then creates a new ordered file.

Visualize and explore quality statistics of unmapped short-read data using BioRead and BioReadQualityStatistics objects

You can obtain quality control plots for short-read data using the `plotSummary` function of the `BioRead` object. The function creates a figure containing six plots that present summary statistics of the data stored in a FASTQ file.

In addition, you can use the `BioReadQualityStatistics` object to:

- Parse FASTQ files without creating a `BioRead` object.
- Interact with the quality data to compare different data sets or filtering options.
- Create customized plots.

Select and move behaviors in the Sequence Alignment app

You can select a block from aligned sequences and move it horizontally if gaps are available.

Compatibility Considerations

To use the previous version of `seqalignviewer`, set the name-value pair argument `'R2012b'` to `true`.

Random access of annotation object data, for consistency with BioMap object data access

You can have random access to data in `GFFAnnotation` and `GTFAnnotation` objects by using these functions:

- `getSubset`
- `getData`
- `getIndex`

Functionality being changed

Functionality	What Happens When You Use This Functionality?	Use This Instead	Compatibility Considerations
`bowtieread` function	Warns	`bowtie` function for UNIX and Mac users.	When using other BOWTIE mapper/aligner programs, set appropriate option(s) to create either a SAM or BAM output file. Then use the `BioMap` object or the `samread` or `bamread` function to access the mapped short reads.
`'natural'` as a choice for the *`OutputValue`* name-value pair input argument to these functions: • `affyrma` • `affygcrma` • `rmasummary`	Warns	`'linear'`	Replace instances of `'natural'` as the value of the *`OutputValue`* name-value pair input argument with `'linear'` for these functions: • `affyrma` • `affygcrma` • `rmasummary`
`'R2012b'` name-value pair input argument for the `seqalignviewer` function.	Still runs	—	See the Compatibility Considerations subheading in “Select and move behaviors in the Sequence Alignment app” on page 21-2.

22

R2012b

Version: 4.2

New Features

Bug Fixes

Compatibility Considerations

Multiple reference sequences in BioMap objects

You can now store information about short reads mapped to multiple references in a `BioMap` object. The new `SequenceDictionary` property contains the catalog of references available in a `BioMap` object.

Compatibility Considerations

For `BioMap` objects created using R2012b:

- The `Reference` property is now a cell string of length `obj.NSeqs`, for both `BioMap` objects with multiple references in the `SequenceDictionary` and objects with only one reference. For `BioMap` objects created before R2012b—which can only have a single reference—the `Reference` property is a string.
- `BioMap` methods that access data by genomic ranges now accept `BioMap` objects with multiple references. To use these methods, you must specify the reference or references to operate on. The affected methods are:
  - `getBaseCoverage`
  - `getCounts`
  - `getAlignment`
  - `getCompactAlignment`
  - `getIndex`

Mapping single and paired-end short read data to reference genomes

Two new functions generate an index and map short reads to a reference sequence using the Burrows-Wheeler transform.

- `bowtiebuild` builds index files using an input reference sequence.
- `bowtie` maps single and paired-end short reads to indexed reference files.

Note `bowtiebuild` and `bowtie` run on Mac and UNIX platforms only.

Increased performance when reading BAM files

The `bamread` function no longer requires the `Index` name-value pair argument to provide index information from a structure in the MATLAB workspace. Indexing happens automatically without a decrease in performance.

Compatibility Considerations

The `Index` name-value pair argument as input to the `bamread` function will be removed in a future release. There is no need to replace it, only remove it.

Name changes for multialignviewer, phytreetool, and seqtool tools

Three tools in Bioinformatics Toolbox are renamed. The old names return a warning and will be removed in a future release.

Compatibility Considerations

- `multialignviewer` is being replaced with `seqalignviewer`.
- `phytreetool` is being replaced with `phytreeviewer`.
- `seqtool` is being replaced with `seqviewer`.

Method removed from neighbor-joining method for phylogenetic tree construction

Compatibility Considerations

The choice of `'average'` for the *`Method`* input argument to the `seqneighjoin` function warns and will be removed in a future release. It is replaced with `'equivar'`.

Functionality being changed or removed

Functionality	What Happens When You Use This Functionality?	Use This Instead	Compatibility Considerations
`Index` name-value pair argument as input to the `bamread` function	Warns	—	Remove instances of the `Index` name-value pair argument. See the Compatibility Considerations subheading in "Increased performance when reading BAM files" on page 22-2.
`'average'` as a choice for the *`Method`* input argument to the `seqneighjoin` function	Warns	`'equivar'`	Replace instances of `'average'` as an input to `seqneighjoin` with `'equivar'`
Changes to tool names: • `multialignviewer` • `phytreetool` • `seqtool`	Warns	• `seqalignviewer` • `phytreeviewer` • `seqviewer`	• Replace instances of `multialignviewer` with `seqalignviewer` • Replace instances of `phytreetool` with `phytreeviewer` • Replace instances of `seqtool` with `seqviewer`

Functionality	What Happens When You Use This Functionality?	Use This Instead	Compatibility Considerations
`'natural'` as a choice for the *`Output`* name-value pair input argument to these functions: • `affyrma` • `affygcrma` • `rmasummary`	Still runs	`'linear'`	Replace instances of `'natural'` as the value of the *`Output`* name-value pair input argument with `'linear'` for these functions: • `affyrma` • `affygcrma` • `rmasummary`

`setName`, `getName`, and `getNSeqs` methods of `BioRead` and `BioMap` objects	Still runs	Dot notation	Replace instances of:
			`setName(`*`BioObj`*`, `*`name`*`)`
			with:
			`BioObj``.Name = `*`name`*
			Replace instances of:
			`getName(`*`BioObj`*`)`
			with:
			`BioObj``.Name`
			Replace instances of:
			`getNSeqs(`*`BioObj`*`)`
			with:
			`BioObj``.NSeqs`

`isequalwithequalnans` for DataMatrix object	Still runs	`isequaln`	Replace instances of `isequalwithequalnans` with `isequaln`
`princomp` for DataMatrix object	Still runs	`pca`	Replace instances of `princomp` with `pca`

R2012a

Version: 4.1

New Features

Bug Fixes

Compatibility Considerations

Update to Jmol Functions

The following functions are updated to use Version 12.0.5 of the Jmol molecule viewer:

- `evalrasmolscript` — Send RasMol script commands to Molecule Viewer window.
- `molviewer` — Display and manipulate 3-D molecule structure.

Enhancements to Objects for NGS Data

You now can construct and access information in a `BioMap` object (created from a BAM-formatted file) more efficiently. Filtering, binning, counting, and base-coverage calculation operations are now faster because source file scanning is no longer needed.

When using the `BioIndexedFile`, `BioRead`, or `BioMap` constructor to create an object from a FASTA-, FASTQ-, or SAM-formatted file, the source file no longer has a size limit of 4 GB.

Compatibility Considerations

The `BioRead` and `BioMap` constructors are changed as follows:

- When creating a `BioMap` object from a SAM- or BAM-formatted file containing multiple reference sequences, the `BioMap` constructor by default uses the first reference listed in the Dictionary of the source file.
- The following syntaxes, which take a `BioIndexedFile` object as an input, have been removed:

  `BioReadobj = BioRead(BioIFobj)`

  `BioMapobj = BioMap(BioIFobj)`

  There is no longer a need to use this syntax, as you can create an indexed object directly from the SAM- or BAM-formatted source file. See Representing Sequence and Quality Data in a BioRead Object or Representing Sequence, Quality, and Alignment/Mapping Data in a BioMap Object.
- The following syntaxes have been removed:

  `BioReadobj = BioRead('SAMFile', File)`

  `BioReadobj = BioRead('FASTQFile', File)`

  Use this simpler syntax instead:

  `BioReadobj = BioRead(File)`
- The following syntax has been removed:

  `BioMapobj = BioMap('SAMFile', File)`

  Use this simpler syntax instead:

  `BioMapobj = BioMap(File)`
- The `Indexed` name-value pair argument as input to the `getSubset` method of the `BioRead` or `BioMap` class has been removed. Use the `InMemory` name-value pair argument instead.
- The `'SubsetRef'` name-value pair argument of the `BioMap` constructor has been removed. Use the `'SelectRef'` name-value pair argument instead.

- The `getCoverage` method of the `BioMap` class has been removed. Use the `getBaseCoverage`, `getCounts`, or `getIndex` method instead.

Enhancements to the NGS Browser

When you import short-read alignment data from a SAM- or BAM-formatted file into the NGS Browser:

- SAM-formatted files no longer have a size limit of 4 GB. Now, the size of both SAM- and BAM-formatted files is limited only by your operating system and available memory.
- The SAM- or BAM-formatted file can contain alignment data for multiple references. When importing short reads, you can select one reference sequence from those listed in the file header, or scan the file to see a list of the actual reference sequences and the aligned read count for each reference sequence.

Sequence Statistics Functions

Compatibility Considerations

- The `aacount` and `basecount` functions no longer accept the `'Others'` name-value pair. Use the `'Ambiguous'` or `'Gaps'` name-value pair instead.
- The `aacount` and `basecount` functions no longer accept the `'Structure'` name-value pair. Use the `'Ambiguous'` name-value pair with either `'ignore'` or `'warn'` instead.
- The `aacount`, `basecount`, `codoncount`, and `dimercount` functions no longer include an `Others` field in the output structure. Use the `Ambiguous` field instead.

Demo for DNA Methylation Analysis

The following demo describes how to identify and compare potential cancer-related methylations at the base-pair level:
“Exploring Genome-Wide Differences in DNA Methylation Profiles”

Functionality Being Changed or Removed

Functionality	What Happens When You Use This Functionality?	Use This Instead	Compatibility Considerations
`SubsetRef` name-value pair argument as input to the `BioMap` constructor function	Errors	`'SelectRef'` name-value pair argument	Replace instances of `SubsetRef` with `SelectRef`.
`BioIndexedFile` object as input to the `BioRead` or `BioMap` constructor function	Errors	A FASTQ-, SAM-, or BAM-formatted file	See the Compatibility Considerations subheading in “Enhancements to Objects for NGS Data” on page 23-2.

Functionality	**What Happens When You Use This Functionality?**	**Use This Instead**	**Compatibility Considerations**
`'FASTQFile', File` pair as input to the `BioRead` constructor	Errors	*`File`*	See the Compatibility Considerations subheading in "Enhancements to Objects for NGS Data" on page 23-2.
`'SAMFile', File` pair as input to the `BioRead` or `BioMap` constructor	Errors	*`File`*	See the Compatibility Considerations subheading in "Enhancements to Objects for NGS Data" on page 23-2.
`Indexed` name-value pair argument as input to the `getSubset` method of the `BioRead` or `BioMap` class	Errors	`InMemory` name-value pair argument	Replace instances of `'Indexed', false` pair with `'InMemory', true` pair.
`getCoverage` method of the `BioMap` class	Errors	`getBaseCoverage`, `getCounts`, or `getIndex` method	Replace all instances of `getCoverage` with `getBaseCoverage`, `getCounts`, or `getIndex`.
`'Others'` name-value pair as input to `aacount` and `basecount` functions	Errors	`'Ambiguous'` or `'Gaps'` name-value pair as input to `aacount` and `basecount` functions	Replace instances of `'Others'` with `'Ambiguous'` or `'Gaps'`.
`'Structure'` name-value pair as input to `aacount` and `basecount` functions	Errors	`'Ambiguous'` name-value pair with either `'ignore'` or `'warn'` as input to `aacount` and `basecount` functions	Replace instances of `'Structure'` with `'Ambiguous'` paired with `'ignore'` or `'warn'`.
`Others` field from the output structure returned by `aacount`, `basecount`, `codoncount`, or `dimercount`.	Errors	`Ambiguous` field	Replace instances of `Others` (as an input) with `Ambiguous`.

R2011b

Version: 4.0

New Features

Bug Fixes

Compatibility Considerations

Visualizing and Investigating Short-Read Alignments and Feature Annotations in the NGS Browser

The NGS Browser lets you visually verify and investigate the alignment of short-read sequences to a reference sequence. For more information, see Visualizing and Investigating Short-Read Alignments and `ngsbrowser`.

Objects for Genomic Feature Annotations

Following are new classes for objects that contain genomic feature annotations for nucleotide sequences:

- `GFFAnnotation` — Contain data from GFF file.
- `GTFAnnotation` — Contain data from GTF file.

These classes have properties and methods that you can use to explore, access, filter, and manipulate all or a subset of the feature annotation data. For more information, see Storing and Managing Feature Annotations in Objects.

Enhancements to BioRead and BioMap Objects

You can now construct a `BioMap` object from a BAM-formatted file.

When constructing these objects from source files, by default the data is indexed, which is more efficient for construction and data access. The `BioRead` and `BioMap` constructors now include an `IndexDir` name-value pair argument, which lets you specify the location of the index file.

You can still construct these objects with the data in memory, which lets you modify all the properties of the objects. The `BioRead` and `BioMap` constructors now include an `InMemory` name-value pair argument, which lets you construct the objects with the data in memory.

For details on the previous enhancements, see Storing and Managing Short-Read Sequence Data in Objects.

Compatibility Considerations

The `BioRead` and `BioMap` constructors are changed as follows:

- The following syntaxes that take a `BioIndexedFile` object as an input will be removed in a future release:

  `BioReadobj = BioRead(BioIFobj)`

  `BioMapobj = BioMap(BioIFobj)`

  There is no longer a need to use this syntax, as you can create an indexed object directly from the SAM- or BAM-formatted source file. See Representing Sequence and Quality Data in a BioRead Object or Representing Sequence, Quality, and Alignment/Mapping Data in a BioMap Object.
- The following syntaxes will be removed in a future release:

  `BioReadobj = BioRead('SAMFile', File)`

```
BioReadobj = BioRead('FASTQFile', File)
```

Use this syntax instead:

```
BioReadobj = BioRead(File)
```

- The following syntax will be removed in a future release:

  ```
  BioMapobj = BioMap('SAMFile', File)
  ```

  Use this syntax instead:

  ```
  BioMapobj = BioMap(File)
  ```
- The `Indexed` name-value pair argument as input to the `getSubset` method of the `BioRead` or `BioMap` class will be removed in a future release. Use the `InMemory` name-value pair argument instead.
- The `'SubsetRef'` name-value pair argument of the `BioMap` constructor will be removed in a future release. Use the `'SelectRef'` name-value pair argument instead.
- If you use the `getSubset` method of a `BioRead` or `BioMap` object, and specify the same element more than once, the method errors, even if the object is in memory.

Enhancements to the saminfo and baminfo Functions

The `saminfo` and `baminfo` functions now include a `ScanDictionary` name-value pair argument, which controls the return of the reference names and the number of reads aligned to each reference from a SAM- or BAM-formatted file in new fields, `ScannedDictionary` and `ScannedDictionaryCount`. This information is needed when constructing a `BioMap` object from a file with multiple reference sequences. For more information, see Constructing a BioMap Object from a SAM- or BAM-Formatted File.

Compatibility Considerations

The `Reference` field is no longer returned in the output structure for `baminfo`. The `ScannedDictionary` field now includes names of the reference sequences.

Conversion of Error and Warning Message Identifiers

For R2011b, some error and warning message identifiers have changed in Bioinformatics Toolbox.

Compatibility Considerations

If you have scripts or functions that use message identifiers that changed, you must update the code to use the new identifiers. Typically, message identifiers are used to turn off specific warning messages, or in code that uses a try/catch statement and performs an action based on a specific error identifier.

For example, the `Bioinfo:nwalign:InvalidScoringMatrix` identifier has changed to `bioinfo:nwalign:InvalidScoringMatrix`. If your code checks for `Bioinfo:nwalign:InvalidScoringMatrix`, you must update it to check for `bioinfo:nwalign:InvalidScoringMatrix` instead.

To determine the identifier for a warning, run the following command just after you see the warning:

```
[MSG,MSGID] = lastwarn;
```

The preceding command saves the message identifier to the variable MSGID.

To determine the identifier for an error, run the following command just after you see the error:

```
exception = MException.last;
MSGID = exception.identifier;
```

Note Warning messages indicate a potential issue with your code. While you can turn off a warning, a suggested alternative is to change your code so it runs warning-free.

Function Elements Being Removed

Function Element Name	What Happens When You Use This Function Element	Use This Instead	Compatibility Considerations
`'SubsetRef'` name-value pair argument as input to `BioMap` constructor function	Warns	`'SelectRef'` name-value pair argument	See the Compatibility Considerations subheading in "Enhancements to BioRead and BioMap Objects" on page 24-2.
`BioIndexedFile` object as input to the `BioRead` or `BioMap` constructor function	Warns	A FASTQ-, SAM-, or BAM-formatted file	See the Compatibility Considerations subheading in "Enhancements to BioRead and BioMap Objects" on page 24-2.
`'FASTQFile',` *`File`* pair as input to the `BioRead` constructor	Warns	*`File`*	See the Compatibility Considerations subheading in "Enhancements to BioRead and BioMap Objects" on page 24-2
`'SAMFile',` *`File`* pair as input to the `BioRead` or `BioMap` constructor	Warns	*`File`*	See the Compatibility Considerations subheading in "Enhancements to BioRead and BioMap Objects" on page 24-2.
`Indexed` name-value pair argument as input to `getSubset` method of the `BioRead` or `BioMap` class	Warns	`InMemory` name-value pair argument	See the Compatibility Considerations subheading in "Enhancements to BioRead and BioMap Objects" on page 24-2.
`Reference` field of structure returned by `baminfo`	Errors	`ScannedDictionary` field	See the Compatibility Considerations subheading in "Enhancements to the saminfo and baminfo Functions" on page 24-3.

25

R2011a

Version: 3.7

New Features

Bug Fixes

Compatibility Considerations

Data Format and Database Functions

The following functions have a new field, `FilePath`, in their output structure:

- `fastainfo` — Return information about FASTA file.
- `fastqinfo` — Return information about FASTQ file.
- `saminfo` — Return information about Sequence Alignment/Map (SAM) file.

The `fastainfo` function has two additional fields in its output structure: `Header` and `Length`.

Sequence Statistics Functions

Compatibility Considerations

In Bioinformatics Toolbox Version 3.6, the `aacount` and `basecount` functions still allowed `'Others'` and `'Structure'` name-value pairs, but displayed a warning.

In Bioinformatics Toolbox Version 3.7, the `aacount` and `basecount` functions do not allow `'Others'` and `'Structure'` name-value pairs, and return an error if you use them. Now you must use the `'Ambiguous'` and `'Gaps'` name-value pairs, which specify whether to count or ignore ambiguous characters and gaps, as well as specify how to count ambiguous characters, and whether to display a warning.

Updates to the BioIndexedFile Class, Properties, and Methods

The following name-value pairs of the `BioIndexedFile` constructor function are renamed:

- `MapKeys` is now `IndexedByKeys`.
- `MemMapIndex` is now `MemoryMappedIndex`.

Note The former name-value pairs are still valid for Bioinformatics Toolbox Version 3.7 (R2011a).

The `MemoryMappedIndex` property of the `BioIndexedFile` class is now editable, which lets you load and unload file indices in memory.

The `BioIndexedFile` class includes the following new methods:

- `getDictionary` — Retrieve reference sequence names from SAM-formatted source file associated with `BioIndexedFile` object.
- `getSubset` — Create object containing subset of elements from `BioIndexedFile` object.

Updates to BioRead and BioMap Classes and Methods

The `BioMap` constructor includes a new name-value pair, `SubsetRef`, which lets you specify one reference sequence in the input argument (`BioIndexedFile` object, SAM-formatted file, or structure) when constructing the `BioMap` object.

The following method of the `BioRead` and `BioMap` classes is updated:

`getSubset` — Create object containing subset of elements from object. Updated with addition of the `Indexed` name-value pair, which lets you use the `BioIndexedFile` object when creating a new object, thus saving memory. This name-value pair is ignored if your `BioRead` or `BioMap` object was not created from a `BioIndexedFile` object.

Following are new methods of the `BioMap` class:

- `getBaseCoverage` — Return base-by-base alignment coverage of reference sequence in `BioMap` object.
- `getCounts` — Return count of read sequences aligned to reference sequence in `BioMap` object.
- `getIndex` — Return indices of read sequences aligned to reference sequence in `BioMap` object.

The `getCoverage` method of the `BioMap` class is being removed in a future release. Use the `getBaseCoverage`, `getCounts`, and `getIndex` methods instead.

Compatibility Considerations

In Bioinformatics Toolbox Version 3.6 and earlier, the `BioMap` class included a `getCoverage` method, which computes read coverage in a `BioMap` object.

In Bioinformatics Toolbox Version 3.7, the `getCoverage` method still runs, but displays a warning. Now use the `getBaseCoverage`, `getCounts`, and `getIndex` methods of the `BioMap` class.

Demos for High-Throughput Sequence Analysis

Following are two new high-throughput sequence analysis demos:

- “Exploring Protein-DNA Binding Sites from Paired-End ChIP-Seq Data”
- “Identifying Differentially Expressed Genes from RNA-Seq Data”

Support Vector Machine (SVM) Functions

The functionality of the `svmsmoset` function is incorporated into the `svmtrain` and `statset` functions. Although `svmsmoset` is still valid, it is no longer documented.

The `svmtrain` function has been updated:

- The function can now handle `NaN` values in the training matrix input and performs more checks of parameters you supply.
- The function now includes Sequential Minimal Optimization (SMO) functionality plus four new name-value pairs: `kernelcachelimit`, `kktviolationlevel`, `options`, and `tolkkt`.
- The default training method is `SMO`, even if you have Optimization Toolbox™ installed.
- The `QuadProg_Opts` and `SMO_Opts` name-value pairs have been replaced by the `options` name-value pair. Although the former name-value pairs are still valid, the recommended ways to perform quadratic programming (QP) training and SMO training are summarized in the following bullets.
- The recommended way to include `QP` options for `svmtrain` is to use the `QP` training method and use the new `options` name-value pair. For the `options` value, use a structure you create with `optimset`.
- The recommended way to include `SMO` options for `svmtrain` is to use the default `SMO` training method and use the new `kernelcachelimit`, `kktviolationlevel`, `options`, and `tolkkt`

name-value pairs. For the `options` value, use a structure you create with the `statset` function and its `Display` and `MaxIter` name-value pairs.

Compatibility Considerations

In Bioinformatics Toolbox Version 3.6 and earlier, if you had Optimization Toolbox installed, `QP` was the default training method for the `svmtrain` function. Now the default training method is `SMO`.

Function Elements Being Removed

Function Element Name	What Happens When You Use This Function Element	Use This Instead	Compatibility Considerations
`'Others'` name-value pair as input to `aacount` and `basecount` functions	Errors	`'Ambiguous'` or `'Gaps'` name-value pair as input to `aacount` and `basecount` functions	See the Compatibility Considerations subheading in “Sequence Statistics Functions” on page 25-2.
`'Structure'` name-value pair as input to `aacount` and `basecount` functions	Errors	`'Ambiguous'` name-value pair with either `'ignore'` or `'warn'` as input to `aacount` and `basecount` functions	See the Compatibility Considerations subheading in “Sequence Statistics Functions” on page 25-2.
`getCoverage` method of `BioMap` class	Warns	`getBaseCoverage`, `getCounts`, and `getIndex` methods	See the Compatibility Considerations subheading in “Updates to BioRead and BioMap Classes and Methods” on page 25-2.
`svmsmoset` function	Still runs	`svmtrain` and `statset` functions	`svmsmoset` is not recommended. Use `svmtrain` and `statset` instead.

R2010b

Version: 3.6

New Features

Bug Fixes

Compatibility Considerations

Data Format and Database Functions

The following new functions allow indexed file access to BAM-formatted files:

- `bamindexread` — Read Binary Sequence Alignment/Map Index (BAI) file.
- `baminfo` — Return information about Binary Sequence Alignment/Map (BAM) file.
- `bamread` — Read data from Binary Sequence Alignment/Map (BAM) file.

The following new functions let you read Bowtie- and SOAP-formatted files:

- `bowtieread` — Read data from Bowtie file.
- `soapread` — Read data from Short Oligonucleotide Analysis Package (SOAP) file.

Sequence Conversion Functions

The following new functions support CIGAR strings for sequence mapping and alignment:

- `align2cigar` — Convert aligned sequences to corresponding Compact Idiosyncratic Gapped Alignment Report (CIGAR) format strings.
- `cigar2align` — Convert unaligned sequences to aligned sequences using Compact Idiosyncratic Gapped Alignment Report (CIGAR) format strings

Sequence Statistics Functions

The following functions are updated:

- `aacount` — Count amino acids in sequence. Updated by adding the `Ambiguous` property, which lets you specify how to count ambiguous amino acid characters. Updated by adding the `Gaps` property, which lets you specify to count or ignore gaps. The `Others` and `Structure` properties still work, but display a warning, indicating that they will be invalid in future versions of Bioinformatics Toolbox. The `Others` field in the output structure is replaced by the `Ambiguous` field.
- `basecount` — Count nucleotides in sequence. Updated by adding the `Ambiguous` property, which lets you specify how to count ambiguous nucleotide characters. Updated by adding the `Gaps` property, which lets you specify to count or ignore gaps. The `Others` and `Structure` properties still work, but display a warning, indicating that they will be invalid in future versions of Bioinformatics Toolbox. The `Others` field in the output structure is replaced by the `Ambiguous` field.
- `codonbias` — Calculate codon frequency for each amino acid coded for in nucleotide sequence. Updated by adding the `Ambiguous` property, which lets you specify how to count codons containing ambiguous nucleotide characters.
- `codoncount` — Count codons in nucleotide sequence. Updated by adding the `Ambiguous` property, which lets you specify how to count codons containing ambiguous nucleotide characters. Updated by adding the `GeneticCode` property, which lets you overlay a grid that groups the synonymous codons on the heat map of the codon counts. The `Others` field in the output structure is replaced by the `Ambiguous` field.
- `dimercount` — Count dimers in nucleotide sequence. Updated by adding the `Ambiguous` property, which lets you specify how to count dimers containing ambiguous nucleotide characters. The `Others` field in the output structure is replaced by the `Ambiguous` field.

Compatibility Considerations

In Bioinformatics Toolbox Version 3.5 and earlier, the `aacount` and `basecount` functions included `'Others'` and `'Structure'` property name/property value pairs, which let you specify how to count ambiguous characters and gaps, and whether to display a warning. These functions also returned a structure with an `Others` field.

In Bioinformatics Toolbox Version 3.6, the `aacount` and `basecount` functions still allow `'Others'` and `'Structure'` property name/property value pairs, but display a warning. Now the `aacount` and `basecount` functions include the `'Ambiguous'` and `'Gaps'` property name/property value pairs, which specify whether to count or ignore ambiguous characters and gaps, as well as specify how to count ambiguous characters, and whether to display a warning. These functions now return a structure with an `Ambiguous` field, which replaces the `Others` field.

In Bioinformatics Toolbox Version 3.6, the `codoncount` and `dimercount` functions return a structure with an optional `Ambiguous` field, which replaces the `Others` field.

Pairwise Sequence Alignment Functions

The following function is updated:

- `nwalign` — Globally align two sequences using Needleman-Wunsch algorithm. Updated to support semiglobal or “glocal” alignments by addition of `Glocal` property.

Multiple Sequence Alignment Functions

The following new functions support CIGAR strings for sequence mapping and alignment:

- `align2cigar` — Convert aligned sequences to corresponding Compact Idiosyncratic Gapped Alignment Report (CIGAR) format strings.
- `cigar2align` — Convert unaligned sequences to aligned sequences using Compact Idiosyncratic Gapped Alignment Report (CIGAR) format strings

The following functions are updated:

- `multialign` — Align multiple sequences using progressive method. Updated to include a new property, `'UseParallel'`, which lets you use `parfor`-loops and compute in parallel mode.
- `seqpdist` — Calculate pairwise distance between sequences. Updated to include a new property, `'UseParallel'`, which lets you use `parfor`-loops and compute in parallel mode.

Compatibility Considerations

In Bioinformatics Toolbox Version 3.4 and earlier, the `multialign` and `seqpdist` functions included `'JobManager'` and `'WaitInQueue'` property name/property value pairs, which let you process in parallel, including support for the MATLAB scheduler for clusters.

In Bioinformatics Toolbox Version 3.5, the `multialign` and `seqpdist` functions allowed the `'JobManager'` and `'WaitInQueue'` property name/property value pairs, but displayed a warning.

In Bioinformatics Toolbox Version 3.6, the `multialign` and `seqpdist` functions error if you use the `'JobManager'` or `'WaitInQueue'` property name/property value pair. Instead they include the

`'UseParallel'` property name/property value pair, which lets you process in parallel, including support for:

- Local workers for multicore machines
- The MATLAB scheduler for clusters
- Third-party schedulers for clusters

Updates to BioMap Class, Methods, and Properties

You can now create a `BioMap` object from a MATLAB structure containing sequence and alignment information, returned by the `bamread` function.

The following method of the `BioMap` class is updated:
`getCoverage` — Compute read coverage in BioMap object. Updated to return the coverage of multiple regions of the reference sequence.

The `BioMap` class includes the following new methods:

- `getCompactAlignment` — Construct compact alignment represented in `BioMap` object.
- `getMatePosition` — Retrieve mate positions of read sequences from `BioMap` object.
- `setMatePosition` — Set mate positions of read sequences in `BioMap` object.

The `BioMap` class includes the following new property:

- `MatePosition` — Positions of the mates for all read sequences represented in the `BioMap` object.

Function Elements Being Removed

Function Element Name	What Happens When You Use This Function Element	Use This Instead	Compatibility Considerations
`'Others'` property name/property value pair as input to `aacount` and `basecount` functions	Warns	`'Ambiguous'` or `'Gaps'` property name/property value pair as input to `aacount` and `basecount` functions	See the Compatibility Considerations subheading in “Sequence Statistics Functions” on page 26-2.
`'Structure'` property name/property value pair as input to `aacount` and `basecount` functions	Warns	`'Ambiguous'` property name/property value pair with either `'ignore'` or `'warn'` as input to `aacount` and `basecount` functions	See the Compatibility Considerations subheading in “Sequence Statistics Functions” on page 26-2.
`'JobManager'` property name/property value pair as input to `multialign` and `seqpdist` functions	Errors	`'UseParallel'` property name/property value pair as input to `multialign` and `seqpdist` functions	See the Compatibility Considerations subheading in “Multiple Sequence Alignment Functions” on page 26-3.

Function Element Name	**What Happens When You Use This Function Element**	**Use This Instead**	**Compatibility Considerations**
`'WaitInQueue'` property name/property value pair as input to `multialign` and `seqpdist` functions	Errors	`'UseParallel'` property name/property value pair as input to `multialign` and `seqpdist` functions	See the Compatibility Considerations subheading in “Multiple Sequence Alignment Functions” on page 26-3.
The following properties of a clustergram object: • `ColumnMarker` • `Impute` • `Ratio` • `RowMarker` • `SymmetricRange`	Errors	New properties of a clustergram object: • `ColumnGroupMarker` • `ImputeFun` • `DisplayRatio` • `RowGroupMarker` • `Symmetric`	See “Clustergram Methods and Properties” on page 28-4.
`'Dimension'` property name/property value pair as input to `clustergram` function	Errors	`'Cluster'` property name/property value pair as input to `clustergram` function	See the Compatibility Considerations subheading in “Microarray Functions” on page 31-3.
`'Pdist'` property name/property value pair as input to `clustergram` function	Errors	Either `'RowPdist'` or `'ColumnPdist'` property name/property value pair as input to `clustergram` function	See the Compatibility Considerations subheading in “Microarray Functions” on page 31-3.
`pdbplot` function	Errors	`molviewer` function	See the Compatibility Considerations subheading in “Protein Analysis and Sequence Utilities Functions” on page 34-3.
`getpir` and `pirread` functions	Errors	Use `getembl`, `getgenpept`, and `getpdb` to retrieve protein sequences from Web databases. Use `emblread`, `genpeptread`, and `pdbread` to read protein sequence data.	See “Data Formats and Databases Functions” on page 36-2.
`mamadnorm` and `mameannorm` functions	Errors	`manorm` function	

R2010a

Version: 3.5

New Features

Bug Fixes

Compatibility Considerations

Data Format and Database Functions

The following functions are new:

- `saminfo` — Return information about Sequence Alignment/Map (SAM) file.
- `samread` — Read data from Sequence Alignment/Map (SAM) file.

The following functions are updated:

- `fastaread` — Read data from FASTA file. Updated to allow trimming of the headers in the output structure by addition of `TrimHeaders` property.
- `fastqread` — Read data from FASTQ file. Updated to allow trimming of the headers in the output structure by addition of `TrimHeaders` property.
- `phytreeread` — Read phylogenetic tree file. Updated to return a second output containing bootstrap values for tree nodes.

Pairwise Sequence Alignment Functions

The following function is updated:

- `fastaread` — Read data from FASTA file. Updated to allow trimming of the headers in the output structure by addition of `TrimHeaders` property.

Multiple Sequence Alignment Functions

The following functions are updated:

- `fastaread` — Read data from FASTA file. Updated to allow trimming of the headers in the output structure by addition of `TrimHeaders` property.
- `multialign` — Align multiple sequences using progressive method. Updated to include a new property, `'UseParallel'`, which lets you use `parfor`-loops and compute in parallel mode.
- `seqpdist` — Calculate pairwise distance between sequences. Updated to include a new property, `'UseParallel'`, which lets you use `parfor`-loops and compute in parallel mode.

Compatibility Considerations

In Bioinformatics Toolbox Version 3.4 and earlier, the `multialign` and `seqpdist` functions included `'JobManager'` and `'WaitInQueue'` property name/property value pairs, which let you process in parallel, including support for the MATLAB scheduler for clusters.

In Bioinformatics Toolbox Version 3.5, the `multialign` and `seqpdist` functions do not include the include the `'JobManager'` and `'WaitInQueue'` property name/property value pairs. Instead they include the `'UseParallel'` property name/property value pair, which lets you process in parallel, including support for:

- Local workers for multicore machines
- The MATLAB scheduler for clusters
- Third-party schedulers for clusters

Phylogenetic Tree Tools and Methods

The following functions are updated:

- `phytreeread` — Read phylogenetic tree file. Updated to return a second output containing bootstrap values for tree nodes.
- `seqpdist` — Calculate pairwise distance between sequences. Updated to include a new property, `'UseParallel'`, which lets you use `parfor`-loops and compute in parallel mode.

BioIndexedFile Function, Object, Methods, and Properties

Following is a new class for an object that lets you extract information from large multi-entry text files.

- `BioIndexedFile` — Allow quick and efficient access to large text file with nonuniform-size entries.

This class has properties and methods that are useful for accessing, reading, and parsing data from a large source file.

BioRead Function, Object, Methods, and Properties

Following is a new class for an object that contains data from short-read sequences, including sequence headers, nucleotide sequences, and the quality scores for the sequences.

- `BioRead` — Contain sequence and quality data.

This class has properties and methods that you can use to explore, access, filter, and manipulate all or a subset of the data, before doing subsequent analyses or sequence alignment and mapping.

BioMap Function, Object, Methods, and Properties

Following is a new class for an object that contains data from short-read sequences, including sequence headers, read sequences, quality scores for the sequences, and data about alignment and mapping to a single reference sequence.

- `BioMap` — Contain sequence, quality, alignment, and mapping data.

This class has properties and methods that you can use to explore, access, filter, and manipulate all or a subset of the data, before doing subsequent analyses or viewing the data.

Function Elements Being Removed

Function Element Name	What Happens When You Use This Function Element	Use This Instead	Compatibility Considerations
`'JobManager'` property name/property value pair as input to `multialign` and `seqpdist` functions	Warns	`'UseParallel'` property name/property value pair as input to `multialign` and `seqpdist` functions	See the Compatibility Considerations subheading in “Multiple Sequence Alignment Functions” on page 27-2
`'WaitInQueue'` property name/property value pair as input to `multialign` and `seqpdist` functions	Warns	`'UseParallel'` property name/property value pair as input to `multialign` and `seqpdist` functions	See the Compatibility Considerations subheading in “Multiple Sequence Alignment Functions” on page 27-2.

R2009b

Version: 3.4

New Features

Bug Fixes

Compatibility Considerations

Data Format and Database Functions

Following are new functions:

- `fastainfo` — Return information about FASTA file.
- `fastqinfo` — Return information about FASTQ file.
- `fastqread` — Read data from FASTQ file.
- `fastqwrite` — Write to file using FASTQ format.
- `sffinfo` — Return information about SFF file.
- `sffread` — Read data from SFF file.
- `tgspcinfo` — Return information about SPC file.
- `tgspcread` — Read data from SPC file.

The following functions are updated:

- `affyread` — Read microarray data from Affymetrix® GeneChip® file. Updated to read cell layout files (CLF) and background probe (BGP) files.
- `multialignwrite` — Write multiple alignment to file. Updated to write a file in either ClustalW ALN format (default) or MSF format.

Protein Analysis Functions

Following is a new function:

- `isotopicdist` — Calculate high-resolution isotope mass distribution and density function.

The following function is updated:

- `cleave` — Cleave amino acid sequence with enzyme. Updated to let you specify an exception to the enzyme's cleavage rule and to let you specify a maximum number of missed cleavage sites. Also updated to return the number of missed cleavage sites per peptide fragment.

Data Visualization Functions

The following functions are updated:

- `microplateplot` — Display visualization of microtiter plate. Display updated so that first row of input matrix appears at the top and is labeled row A. Updated to return the handle to the axes of the plot, which lets you reverse the order or the rows or columns in the display. Updated to include a new property, `'TextFontSize'`, which lets you control the font size of text labels.
- `multialignviewer` — Display and interactively adjust multiple sequence alignment. Updated to accept a list of names to label the sequences in the Multiple Sequence Alignment Viewer window.
- `showalignment` — Display color-coded sequence alignment. Updated to control the inclusion or exclusion of terminal gaps from the count of matches and similar residues when displaying a pairwise alignment.

Compatibility Considerations

In Bioinformatics Toolbox Version 3.3, the default layout for the plot returned by `microplateplot` displayed the first row of the input matrix at the bottom.

In Bioinformatics Toolbox Version 3.4, the plot displays the first row of the input matrix at the top.

Sequence Statistics Functions

The following function is updated:

- `seqshowwords` — Graphically display words in sequence. Updated to search for multiple words in a sequence.

Sequence Utility Functions

The following functions are updated:

- `cleave` — Cleave amino acid sequence with enzyme. Updated to let you specify an exception to the enzyme's cleavage rule and to let you specify a maximum number of missed cleavage sites. Also updated to return the number of missed cleavage sites per peptide fragment.
- `rebasecuts` — Find restriction enzymes that cut nucleotide sequence. Updated to use Version 904 of REBASE®, the Restriction Enzyme Database.
- `restrict` — Split nucleotide sequence at restriction site. Updated to use Version 904 of REBASE, the Restriction Enzyme Database.

Sequence Visualization Functions

The following functions are updated:

- `multialignviewer` — Display and interactively adjust multiple sequence alignment. Updated to accept a list of names to label the sequences in the Multiple Sequence Alignment Viewer window.
- `showalignment` — Display color-coded sequence alignment. Updated to control the inclusion or exclusion of terminal gaps from the count of matches and similar residues when displaying a pairwise alignment.

Pairwise Sequence Alignment Functions

Following is a new function:

- `localalign` — Return local optimal and suboptimal alignments between two sequences.

The following functions are updated:

- `multialignviewer` — Display and interactively adjust multiple sequence alignment. Updated to accept a list of names to label the sequences in the Multiple Sequence Alignment Viewer window.
- `showalignment` — Display color-coded sequence alignment. Updated to control the inclusion or exclusion of terminal gaps from the count of matches and similar residues when displaying a pairwise alignment.

Multiple Sequence Alignment Functions

The following functions are updated:

- `multialignviewer` — Display and interactively adjust multiple sequence alignment. Updated to accept a list of names to label the sequences in the Multiple Sequence Alignment Viewer window.

- `multialignwrite` — Write multiple alignment to file. Updated to write a file in either ClustalW ALN format (default) or MSF format.
- `showalignment` — Display color-coded sequence alignment. Updated to control the inclusion or exclusion of terminal gaps from the count of matches and similar residues when displaying a pairwise alignment.

Phylogenetic Tree Tools and Methods

The Phylogenetic Tree Tool includes the following updates:

- Includes two new circular print renderings: equal angle and equal daylight
- Updates to **Tools** menu, including commands to select specific branch and leaf nodes based on different criteria, such as distance, common ancestors, leaves only, and descendants.

Following is a new method:

- `cluster` — Validate clusters in phylogenetic tree.

The following method is updated:

- `plot` — Draw phylogenetic tree. Updated to include two new algorithms for circular layouts: equal angle and equal daylight. Updated to let you rotate circular trees from 0 through 360 degrees and to rotate leaf labels of circular trees so that the text is aligned to the root node. Updated the `'LeafLabels'` property so that it defaults to `true` for circular layouts and to `false` for square and angular layouts.

Compatibility Considerations

In Bioinformatics Toolbox Version 3.3, the `'LeafLabels'` property defaulted to `true` when the `'Type'` property was `'square'` or `'angular'`, and to `false` when the `'Type'` property was `'radial'`.

In Bioinformatics Toolbox Version 3.4, the `'LeafLabels'` property defaults to `false` when the `'Type'` property is `'square'` or `'angular'`, and to `true` when the `'Type'` property is `'radial'`.

Clustergram Window

The Clustergram window has two new toolbar buttons:

- **Annotate** button — Shows and hides intensity values for each area of the heat map.
- **Show Dendrogram** button — Shows and hides the dendrograms.

Clustergram Methods and Properties

The following are new methods of a clustergram object:

- `addTitle` — Add title to clustergram.
- `addXLabel` — Label *x*-axis of clustergram.
- `addYLabel` — Label *y*-axis of clustergram.
- `clusterGroup` — Select cluster group.

The following properties of a clustergram object are renamed:

- `ColumnMarker` is now `ColumnGroupMarker`.
- `Impute` is now `ImputeFun`.
- `Ratio` is now `DisplayRatio`.
- `RowMarker` is now `RowGroupMarker`.
- `SymmetricRange` is now `Symmetric`.

Note The former property names are still valid for Bioinformatics Toolbox version 3.4 (R2009b).

Following is a new property related to the display of dendrogram tree diagrams in a clustergram object:

- `ShowDendrogram`

The following are new properties related to the display of row and column labels of a clustergram object:

- `RowLabels`
- `ColumnLabels`
- `RowLabelsLocation`
- `ColumnLabelsLocation`
- `RowLabelsColor`
- `ColumnLabelsColor`
- `LabelsWithMarkers`
- `RowLabelsRotate`
- `ColumnLabelsRotate`

The following are new properties related to annotating data in a clustergram object:

- `Annotate`
- `AnnotColor`
- `AnnotPrecision`

When using clustergram properties with the `get` and `set` methods, the property names are now case sensitive.

Compatibility Considerations

In Bioinformatics Toolbox Version 3.3, the property names of a clustergram object were not case sensitive when used with the `get` and `set` methods.

In Bioinformatics Toolbox Version 3.4, property names of a clustergram object are case sensitive.

HeatMap Object, Methods, and Properties

Following is a new object:

- HeatMap object — Object containing matrix and heat map display properties.

The following are methods of a HeatMap object:

- `addTitle` — Add title to heat map.
- `addXLabel` — Label *x*-axis of heat map.
- `addYLabel` — Label *y*-axis of heat map.
- `plot` — Render heat map for object.
- `view` — Render heat map for object.

A HeatMap object includes many properties that control the creation of the heat map, row and column labels, axes labels, title, and data annotation.

DataMatrix Methods

Following is a new method of a DataMatrix object:

- `dmwrite` — Write DataMatrix object to text file.

Microarray Functions, Objects, Methods, and Properties

Following are new functions to create objects containing data from a microarray gene expression experiment:

- `bioma.ExpressionSet` — Contain data from microarray gene expression experiment.
- `bioma.data.ExptData` — Contain expression data from microarray gene expression experiment.
- `bioma.data.MetaData` — Contain sample or feature metadata from microarray gene expression experiment.
- `bioma.data.MIAME` — Contain experiment information from microarray gene expression experiment.

These objects have properties and methods that are useful for viewing and analyzing the data or a subset of the data.

Mass Spectrometry Functions

Following are new functions:

- `isotopicdist` — Calculate high-resolution isotope mass distribution and density function.
- `tgspcinfo` — Return information about SPC file.
- `tgspcread` — Read data from SPC file.

The following function is updated:

- `mspeaks` — Convert raw peak data to peak list (centroided data). Updated to include a new property, `'Style'`, which lets you specify the style for marking the peaks in the plot.

Demos for Sequence Analysis

Following are two new sequence analysis demos:

- `Working with SFF Files from the 454 Genome Sequencer FLX System`
- "Working with Illumina/Solexa Next-Generation Sequencing Data"

Demos for Microarray Analysis

Following are two new microarray analysis demos:

- "Working with Objects for Microarray Experiment Data"
- "Analyzing Illumina Bead Summary Gene Expression Data"

R2009a

Version: 3.3

New Features

Bug Fixes

Compatibility Considerations

Data Visualization Functions

Following is a new function:

- `microplateplot` — Display visualization of microtiter plate.

Sequence Utility Functions

The following functions are updated:

- `rebasecuts` — Find restriction enzymes that cut nucleotide sequence. Updated to use Version 811 of REBASE, the Restriction Enzyme Database.
- `restrict` — Split nucleotide sequence at restriction site. Updated to use Version 811 of REBASE, the Restriction Enzyme Database.

Sequence Conversion Functions

The following function is updated:

- `nt2aa` — Convert nucleotide sequence to amino acid sequence. Updated to include a new property, `'ACGTOnly'`, to support ambiguous and unknown nucleotide characters.

Bioanalytic and Mass Spectrometry Functions

The following functions are updated to use with data from any separation technique, including mass spectrometry:

- `msalign` — Align peaks in signal to reference peaks.
- `msbackadj` — Correct baseline of signal with peaks.
- `mslowess` — Smooth signal with peaks using nonparametric method.
- `msnorm` — Normalize set of signals with peaks.
- `mspeaks` — Convert raw peak data to peak list (centroided data).
- `msppresample` — Resample signal with peaks while preserving peaks.
- `msresample` — Resample signal with peaks.
- `mssgolay` — Smooth signal with peaks using least-squares polynomial.

Microarray Functions

The following functions are updated:

- `cghcbs` — Perform circular binary segmentation (CBS) on array-based comparative genomic hybridization (aCGH) data. Updated to include an optional heuristic stopping rule to improve performance.
- `ilmnbslookup` — Look up Illumina® BeadStudio™ target (probe) sequence and annotation information. Updated to read Illumina microRNA array annotation files.
- `ilmnbsread` — Read gene expression data exported from Illumina BeadStudio software. Updated to read Illumina microRNA array data files.

- `mattest` — Perform two-sample t-test to evaluate differential expression of genes from two experimental conditions or phenotypes. Updated with new property, `'VarType'`, which lets you specify equal or unequal (default) variance for the test.

Compatibility Considerations

A compatibility consideration related to the `mattest` function was introduced in Bioinformatics Toolbox Version 3.2, but not reported in the Release Notes for Version 3.2 (R2008b). Specifically, in Bioinformatics Toolbox Version 3.1 and earlier, the `mattest` function used equal variance for the test. In Bioinformatics Toolbox Version 3.2, the `mattest` function starting using unequal variance for the test.

Demo for Sequence Analysis

The following is a new sequence analysis demo:
“Predicting Protein Secondary Structure Using a Neural Network”

R2008b

Version: 3.2

New Features

Bug Fixes

Compatibility Considerations

Data Format and Database Functions

Following are new functions:

- `affygcrma` — Perform GC Robust Multi-array Average (GCRMA) procedure on Affymetrix microarray probe-level data.
- `affyrma` — Perform Robust Multi-array Average (RMA) procedure on Affymetrix microarray probe-level data.
- `affysnpannotread` — Read Affymetrix Mapping DNA array data from CSV-formatted annotation file.
- `geoseriesread` — Read Gene Expression Omnibus (GEO) Series (GSE) format data.
- `multialignwrite` — Write multiple-alignment to file using ClustalW ALN format.
- `mzcdfread` — Read mass spectrometry data from netCDF file.

The following functions are updated:

- `affyread` — Read microarray data from Affymetrix GeneChip file. Updated so that `Probes` field in the return structure is now a `single`, which reduces memory usage.
- `celintensityread` — Read probe intensities from Affymetrix CEL files. Updated so that `PMIntensities` and `MMIntensities` fields in the return structure are now `singles`, which reduces memory usage.
- `geosoftread` — Read Gene Expression Omnibus (GEO) SOFT format data. Updated to support Platform (GPL) records.
- `getgeodata` — Retrieve Gene Expression Omnibus (GEO) format data. Updated to support Platform (GPL) and Series (GSE) records.
- `goannotread` — Read annotations from Gene Ontology annotated file. Updated to include two new properties, `'Fields'` and `'Aspect'`, which let you read a subset of the data in the annotated file.
- `multialignread` — Read multiple sequence alignment file. Updated to support PHYLIP (Phylogeny Inference Package) multiple-sequence alignment files.
- `mzxmlread` — Read data from mzXML file. Improved to read larger files, faster and without running out of memory. Updated with three new properties, `'Levels'`, `'TimeRange'`, and `'ScanIndices'`, which let you filter and read a subset of the data. Updated with a `'Verbose'` property to control the progress display while reading the file.

Compatibility Considerations

In Bioinformatics Toolbox Version 3.1 and earlier, the `Probes` field, in the structure returned by `affyread`, and the `PMIntensities` and `MMIntensities` fields, in the structure returned by `celintensityread`, were `doubles`. In Bioinformatics Toolbox Version 3.2, these fields are `singles`.

Sequence Utility Functions

Following is a new function:

- `cleavelookup` — Find cleavage rule for enzyme or compound.

The following functions are updated:

- `blastncbi` — Create remote NCBI BLAST report request ID or link to NCBI BLAST report. Updated to include a `'GapCosts'` property, which lets you specify penalties for both opening and extending gaps, and an `'Entrez'` property, which lets you limit searches using Entrez query syntax.
- `cleave` — Cleave amino acid sequence with enzyme. Includes a new input argument that specifies the name of an enzyme or compound for which a cleavage rule is specified in the literature.
- `rebasecuts` — Find restriction enzymes that cut nucleotide sequence. Updated to use Version 806 of REBASE, the Restriction Enzyme Database.
- `restrict` — Split nucleotide sequence at restriction site. Updated to use Version 806 of REBASE, the Restriction Enzyme Database.
- `seqlogo` — Display sequence logo for nucleotide or amino acid sequences. Updated to return a figure handle to the sequence logo.

Multiple Sequence Alignment Functions

Following is a new function:

- `multialignwrite` — Write multiple alignment to file using ClustalW ALN format.

The following function is updated:

- `multialignread` — Read multiple sequence alignment file. Updated to support PHYLIP (Phylogeny Inference Package) multiple sequence alignment files.

Gene Ontology Functions

The following function is updated:

- `goannotread` — Read annotations from Gene Ontology annotated file. Updated to include two new properties, `'Fields'` and `'Aspect'`, which let you read a subset of the data in the annotated file.

Protein Analysis Functions

Following are new functions:

- `cleavelookup` — Find cleavage rule for enzyme or compound.
- `pdbsuperpose` — Superpose 3-D structures of two proteins.
- `pdbtransform` — Apply linear transformation to 3-D structure of molecule.

The following function is updated:

- `cleave` — Cleave amino acid sequence with enzyme. Includes a new input argument that specifies the name of an enzyme or compound for which a cleavage rule is specified in the literature.

Mass Spectrometry Functions

Following are new functions:

- `mzcdf2peaks` — Convert mzCDF structure to peak list.

- `mzcdfinfo` — Return information about netCDF file containing mass spectrometry data.
- `mzcdfread` — Read mass spectrometry data from netCDF file.
- `mzxmlinfo` — Return information about mzXML file.

The following function is updated:

- `mzxmlread` — Read data from mzXML file. Improved to read larger files, faster and without running out of memory. Updated with three new properties, `'Levels'`, `'TimeRange'`, and `'ScanIndices'`, which let you filter and read a subset of the data. Updated with a `'Verbose'` property to control the progress display while reading the file.

Microarray File Format Functions

Following are new functions:

- `affygcrma` — Perform GC Robust Multi-array Average (GCRMA) procedure on Affymetrix microarray probe-level data.
- `affyrma` — Perform Robust Multi-array Average (RMA) procedure on Affymetrix microarray probe-level data.
- `affysnpannotread` — Read Affymetrix Mapping DNA array data from CSV-formatted annotation file.
- `geoseriesread` — Read Gene Expression Omnibus (GEO) Series (GSE) format data.

The following functions are updated:

- `affyread` — Read microarray data from Affymetrix GeneChip file. Updated so that `Probes` field in the return structure is now a `single`, which reduces memory usage.
- `celintensityread` — Read probe intensities from Affymetrix CEL files. Updated so that `PMIntensities` and `MMIntensities` fields in the return structure are now `singles`, which reduces memory usage.
- `geosoftread` — Read Gene Expression Omnibus (GEO) SOFT format data. Updated to support Platform (GPL) records.
- `getgeodata` — Retrieve Gene Expression Omnibus (GEO) format data. Updated to support Platform (GPL) and Series (GSE) records.

Compatibility Considerations

In Bioinformatics Toolbox Version 3.1 and earlier, the `Probes` field, in the structure returned by `affyread`, and the `PMIntensities` and `MMIntensities` fields, in the structure returned by `celintensityread`, were `doubles`. In Bioinformatics Toolbox Version 3.2, these fields are `singles`.

Microarray Functions

Following are new functions:

- `affysnpintensitysplit` — Split Affymetrix SNP probe intensity information for alleles A and B.
- `affygcrma` — Perform GC Robust Multi-array Average (GCRMA) procedure on Affymetrix microarray probe-level data.

- `affyrma` — Perform Robust Multi-array Average (RMA) procedure on Affymetrix microarray probe-level data.
- `DataMatrix` — Create DataMatrix object.

The following functions are updated:

- `ilmnbslookup` — Look up Illumina BeadStudio target (probe) sequence and annotation information. Updated to support BGX and TXT annotation files.
- `mattest` — Perform two-sample t-test to evaluate differential expression of genes from two experimental conditions or phenotypes. Updated to use unequal variance instead of equal variance for the test.
- `probesetlookup` — Look up information for Affymetrix probe set. Updated to accept multiple probe set IDs/names or gene IDs.

Compatibility Considerations

In Bioinformatics Toolbox Version 3.1 and earlier, the `mattest` function used equal variance for the test. In Bioinformatics Toolbox Version 3.2, the `mattest` function uses unequal variance for the test.

DataMatrix Object

Following is a new object:

- DataMatrix object — Data structure encapsulating data and metadata from microarray experiment so that it can be indexed by gene or probe identifiers and by sample identifiers.

DataMatrix Methods

There are many methods that let you create, index into, modify, create subsets, sort, perform operations on, analyze, and plot a DataMatrix object.

Demo for Visualization Tools

The “Visualizing the Three-Dimensional Structure of a Molecule” demo is updated to use the new `pdbsuperpose` function.

Demo for Sequence Analysis

The following is a new sequence analysis demo:

- Analyzing the Human Distal Gut Microbiome

Demos for Microarray Data Analysis

Following is a new microarray data analysis demo:

- “Working with GEO Series Data”

The Exploring Microarray Gene Expression Data demo is updated to use the new DataMatrix object.

The “Analyzing Affymetrix SNP Arrays for DNA Copy Number Variants” demo is updated to use two new functions: `affysnpannotread` and `affysnpintensitysplit`.

The “Preprocessing Affymetrix Microarray Data at the Probe Level” demo is updated to use two new functions: `affygcrma` and `affyrma`.

R2008a

Version: 3.1

New Features

Bug Fixes

Compatibility Considerations

Data Format and Database Functions

Following is a new function:

- `ilmnbsread` — Read microarray data exported from Illumina BeadStudio software.

The following functions are updated:

- `celintensityread` — Read probe intensities from Affymetrix CEL files. Updated output structure to include a new field, `GroupNumbers`, which contains group numbers of probes.
- `fastawrite` — Write to file using FASTA format. Updated such that if you specify an existing file, new data is appended to the file instead of overwriting it.
- `getgenbank` — Retrieve sequence information from GenBank® database. Updated such that if you use the `'ToFile'` property and specify an existing file, new data is appended to the file instead of overwriting it. Updated to allow you to access a partial sequence by adding new property `'PartialSeq'`.
- `getgenpept` — Retrieve sequence information from GenPept database. Updated such that if you use the `'ToFile'` property and specify an existing file, new data is appended to the file instead of overwriting it. Updated to allow you to access a partial sequence by adding new property `'PartialSeq'`.
- `getgeodata` — Retrieve Gene Expression Omnibus (GEO) SOFT format data. Updated to retrieve both Sample (GSM) and Data Set (GDS) data.

Compatibility Considerations

In Bioinformatics Toolbox Version 3.0 and earlier, when writing to files using the `fastawrite` function or the `getgenbank` or `getgenpept` functions with the `'ToFile'` property, if you specified an existing file, the file was overwritten. In Bioinformatics Toolbox Version 3.1, if you specify an existing file, new data is appended to the file instead of overwriting it.

Sequence Utility Functions

The following functions are updated:

- `evalrasmolscript` — Send RasMol script commands to Molecule Viewer window. Updated to use Version 11.4 of the Jmol molecule viewer.
- `molviewer` — Display and manipulate 3-D molecule structure. Updated to use Version 11.4 of the Jmol molecule viewer.
- `ramachandran` — Draw Ramachandran plot for Protein Data Bank (PDB) data. Updated to handle PDB files with multiple chains and models by adding three properties: `'Chain'`, `'Plot'`, and `'Model'`. Updated Ramachandran plot to mark glycine residues and display reference regions by adding three properties: `'Glycine'`, `'Regions'`, and `'RegionDef'`. Updated Ramachandran plot to display amino acid information in ToolTip. Updated to easily determine the names and sequence positions of amino acids corresponding to torsion angles by creating an output structure.
- `rebasecuts` — Find restriction enzymes that cut nucleotide sequence. Updated to use Version 710 of REBASE, the Restriction Enzyme Database.
- `restrict` — Split nucleotide sequence at restriction site. Updated to use Version 710 of REBASE, the Restriction Enzyme Database.

Pairwise Sequence Alignment Functions

The following functions are updated:

- `nwalign` — Globally align two sequences using Needleman-Wunsch algorithm. Updated to improve pairwise sequence performance.
- `swalign` — Locally align two sequences using Smith-Waterman algorithm. Updated to improve pairwise sequence performance.

Phylogenetic Tree Tools Function

The following function is updated:

- `dnds` — Estimate synonymous and nonsynonymous substitution rates. Updated by adding `'AdjustStops'` property to control whether stop codons are excluded from calculations.

Protein Analysis Functions

The following functions are updated:

- `evalrasmolscript` — Send RasMol script commands to Molecule Viewer window. Updated to use Version 11.4 of the Jmol molecule viewer.
- `molviewer` — Display and manipulate 3-D molecule structure. Updated to use Version 11.4 of the Jmol molecule viewer.
- `ramachandran` — Draw Ramachandran plot for Protein Data Bank (PDB) data. Updated to handle PDB files with multiple chains and models by adding three properties: `'Chain'`, `'Plot'`, and `'Model'`. Updated Ramachandran plot to mark glycine residues and display reference regions by adding three properties: `'Glycine'`, `'Regions'`, and `'RegionDef'`. Updated Ramachandran plot to display amino acid information in ToolTip. Updated to easily determine the names and sequence positions of amino acids by creating an output structure.

Microarray File Format Functions

Following is a new function:

- `ilmnbsread` — Read microarray data exported from Illumina BeadStudio software.

The following functions are updated:

- `celintensityread` — Read probe intensities from Affymetrix CEL files. Updated output structure to include a new field, `GroupNumbers`, which contains group numbers of probes.
- `getgeodata` — Retrieve Gene Expression Omnibus (GEO) SOFT format data. Updated to retrieve both Sample (GSM) and Data Set (GDS) data.

Microarray Functions

Following are new functions:

- `affysnpquartets` — Create table of SNP probe quartet results for Affymetrix probe set.
- `cghfreqplot` — Display frequency of DNA copy number alterations across multiple samples.

- `ilmnbslookup` — Look up Illumina BeadStudio target (probe) sequence and annotation information.
- `redbluecmap` — Create red and blue color map.

The following functions are updated:

- `clustergram` — Compute hierarchical clustering, display dendrogram and heat map, and create clustergram object.

  Updated properties include:

  - `'Linkage'` — Can specify linkage method separately for rows and columns.
  - `'Dendrogram'` — Can specify color threshold separately for rows and columns.

  Replaced properties include:

  - `'Dimension'` — Replaced by the `'Cluster'` property, which lets you cluster along the columns, rows, or both.
  - `'Pdist'` — Replaced by `'RowPdist'` and `'ColumnPdist'` properties.

  New properties include:

  - `'Standardize'` — Specifies the dimension for standardizing the data.
  - `'DisplayRange'` — Specifies the display range of standardized values.
  - `'LogTrans'` — Controls the $\log_2$ transform of the data.
  - `'Impute'` — Specifies a function and properties to impute missing data.
  - `'RowMarker'` — Adds color and text marker to a group of rows.
  - `'ColumnMarker'` — Adds color and text marker to a group of columns.

  The interactivity of the clustergram figure is enhanced with the following features:

  - Select a group of rows or columns and display the group number and genes or samples within.
  - Create a new clustergram of only a group of the data.
  - Export data as a clustergram object or structure in the MATLAB Workspace.
- `maboxplot` — Create box plot for microarray data. Updated by adding `'BoxPlot'` property, which lets you specify arguments to pass to the `boxplot` function, which creates the box plot.
- `mairplot` — Create intensity versus ratio scatter plot of microarray data. Updated by adding `'PlotOnly'` property, which lets you display the scatter plot without user interface components.
- `mattest` — Perform two-sample t-test to evaluate differential expression of genes from two experimental conditions or phenotypes. Updated by adding `'Bootstrap'` property to run bootstrap tests.
- `mavolcanoplot` — Create significance versus gene expression ratio (fold change) scatter plot of microarray data. Updated by adding `'PlotOnly'` property, which lets you display the volcano plot without user interface components.
- `probesetvalues` — Create table of Affymetrix probe set intensity values. Updated by adding `'Background'` property to control the background correction.
- `zonebackadj` — Perform background adjustment on Affymetrix microarray probe-level data using zone-based method. Updated to return a third output containing the estimated background values for each probe.

Compatibility Considerations

In Bioinformatics Toolbox Version 3.0 and earlier, the `clustergram` function included `'Dimension'` and `'Pdist'` properties. In Bioinformatics Toolbox Version 3.1, the `'Dimension'` property is replaced by the `'Cluster'` property, and the `'Pdist'` property is replaced by the `'RowPdist'` and `'ColumnPdist'` properties.

Object

Following is a new object:

- clustergram object — Object containing hierarchical clustering analysis data.

Clustergram Methods

The following are new methods of a clustergram object:

- `get` — Retrieve information about clustergram object.
- `plot` — Render clustergram heat map and dendrograms for clustergram object.
- `set` — Set property of clustergram object.
- `view` — View clustergram heat map and dendrograms for clustergram object.

Demo for Sequence Analysis

The following is a new sequence analysis demo:

- Performing a Metagenomic Analysis of a Sargasso Sea Sample

Demo for Microarray Data Analysis

The following is a new microarray data analysis demo:

- "Analyzing Affymetrix SNP Arrays for DNA Copy Number Variants"

Demo for Visualization Tools

The following is a new visualization tool demo:

- "Working with the Clustergram Function"

Demos for Mass Spectrometry Data Analysis

- The Batch Processing of Spectra Using Distributed Computing demo is updated to use the latest features of the Parallel Computing Toolbox™ version 3.3, and is now called "Batch Processing of Spectra Using Sequential and Parallel Computing"
- The "Preprocessing Raw Mass Spectrometry Data" demo is updated with state-of-the-art examples for peak detection using wavelets denoising, binning by hierarchical clustering, and binning by dynamic programming.

32

R2007b

Version: 3.0

New Features

Bug Fixes

Compatibility Considerations

Data Format and Database Functions

Following are new functions:

- `blastformat` — Create local BLAST database.
- `blastreadlocal` — Read data from local BLAST report.
- `cytobandread` — Read cytogenetic banding information.

The following function was updated:

- `affyread` — Read microarray data from Affymetrix GeneChip file. Updated the structure returned when reading a CDF library file. The structure contains three new subfields: `GroupNumber`, `Direction`, and `GroupName`.

Microarray File Format Functions

Following is a new function:

- `cytobandread` — Read cytogenetic banding information.

The following function was updated:

- `affyread` — Read microarray data from Affymetrix GeneChip file. Updated the structure returned when reading a CDF library file. The structure contains three new subfields: `GroupNumber`, `Direction`, and `GroupName`.

Microarray Functions

Following are new functions:

- `chromosomeplot` — Plot chromosome ideogram with G-banding pattern.
- `cghcbs` — Perform circular binary segmentation (CBS) on array-based comparative genomic hybridization (aCGH) data.

The following function is updated:

- `probesetvalues` — Create table of Affymetrix probe set intensity values. Updated return matrix, which contains intensity values for probe-level data, to include two new fields: `GroupNumber` and `Direction`. Updated to return a second output containing the column names for the return matrix, which contains intensity values for probe-level data.

Sequence Conversion, Utility, and Visualization Functions

Following are new functions:

- `blastlocal` — Perform search on local BLAST database to create BLAST report.
- `rnaconvert` — Convert secondary structure of RNA sequence between bracket and matrix notations.
- `rnafold` — Predict minimum free-energy secondary structure of RNA sequence.
- `rnaplot` — Draw secondary structure of RNA sequence.

Mass Spectrometry Functions

The following function is updated:

- `mspalign` — Align mass spectra from multiple peak lists from LC/MS or GC/MS data set. Updated to include a new property, `'ShowEstimation'`, which controls the display of an assessment plot relative to the estimation method and the vector of common mass/charge (m/z) values.

Statistical Learning Functions

The following function is updated:

- `svmsmoset` — Create or edit Sequential Minimal Optimization (SMO) options structure. Updated default values for the `'MaxIter'` and `'KernelCacheLimit'` properties. Changed the `'Display'` property so that when set to `'iter'`, a report displays every 500 iterations instead of 10.

Compatibility Considerations

In Bioinformatics Toolbox Version 2.6 and earlier, the `svmsmoset` function used a `'MaxIter'` property with a default of `1500` and a `'KernelCacheLimit'` property with a default of `7500`. In Bioinformatics Toolbox Version 3.0, the defaults are `15000` and `5000`, respectively. Also, when you set the `'Display'` property to `'iter'`, a report displays every 500 iterations instead of 10.

Gene Ontology Methods

The following methods of a gene ontology object are updated:

- `geneont.getancestors` — Find terms that are ancestors of specified Gene Ontology term. Updated to also return the number of times each ancestor is found. Updated to include two new properties, `'Relationtype'`, which specifies a relationship type to search for in the gene ontology, and `'Exclude'`, which controls excluding the original queried term(s) from the output, unless the term was reached while searching the gene ontology.
- `geneont.getdescendants` — Find terms that are descendants of specified Gene Ontology term. Updated to also return the number of times each descendant is found. Updated to include two new properties, `'Relationtype'`, which specifies a relationship type to search for in the gene ontology, and `'Exclude'`, which controls excluding the original queried term(s) from the output, unless the term was reached while searching the gene ontology.
- `geneont.getrelatives` — Find terms that are relatives of specified Gene Ontology term. Updated to also return the number of times each relative is found. Updated to include three new properties, `'Levels'`, which specifies the number of levels up and down to search in the gene ontology, `'Relationtype'`, which specifies a relationship type to search for in the gene ontology, and `'Exclude'`, which controls excluding the original queried term(s) from the output, unless the term was reached while searching the gene ontology.

Demos for Microarray Data Analysis

The following are two new microarray data analysis demos:

- “Detecting DNA Copy Number Alteration in Array-Based CGH Data”

- “Analyzing Array-Based CGH Data Using Bayesian Hidden Markov Modeling”

Demos for Sequence Analysis

The following are two new sequence analysis demos:

- “Predicting and Visualizing the Secondary Structure of RNA Sequences”
- “Identifying Over-Represented Regulatory Motifs”

The “Investigating the Bird Flu Virus” demo was updated to demonstrate how to write KML-formatted files, which can be used by Google Earth™ to display geospatial data.

Demo for Graph Theory Analysis

The following is a new graph theory demo:

- Working with Graph Theory Functions

33

R2007a+

Version: 2.6

New Features

Bug Fixes

Compatibility Considerations

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.

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 Metabolite Mixtures using Liquid Chromatography/Mass Spectrometry (LC/MS)"

34

R2007a

Version: 2.5

New Features

Bug Fixes

Compatibility Considerations

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 $\frac{1}{\sqrt{eps}}$. 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 removed:

- `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 pairwise 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.

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 Microarray 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 pairwise 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.

35

R2006b

Version: 2.4

New Features

Bug Fixes

Data Formats and Database Functions

Following is a new function for getting data into the MATLAB environment:

- `mzxmlread` — Read mzXML file into the MATLAB software 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.
- `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 the MATLAB software 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.

R2006a+

Version: 2.3

New Features

Data Formats and Databases Functions

The following functions are removed:

- `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 `https://pir.georgetown.edu/pirwww/dbinfo/pir_psd.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 `https://pir.georgetown.edu/pirwww/dbinfo/pir_psd.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).

R2006a

Version: 2.2.1

No New Features or Changes

R14SP3+

Version: 2.2

New Features

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.
- `geneont.getancestors`, `geneont.getdescendants`, `geneont.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`).

R14SP3

Version: 2.1.1

No New Features or Changes

R14SP2+

Version: 2.1

New Features

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 pairwise 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`.

R14SP2

Version: 2.0.1

New Features

Updated RBASE Table

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

Expanded Bioperl Demonstration

Example of calling the MATLAB software 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`.