

# **Bioinformatics Toolbox™**

## **Release Notes**

---

## How to Contact MathWorks



[www.mathworks.com](http://www.mathworks.com)  
[comp.soft-sys.matlab](mailto:comp.soft-sys.matlab@mathworks.com)  
[www.mathworks.com/contact\\_TS.html](http://www.mathworks.com/contact_TS.html)

Web  
Newsgroup  
Technical Support



[suggest@mathworks.com](mailto:suggest@mathworks.com)  
[bugs@mathworks.com](mailto:bugs@mathworks.com)  
[doc@mathworks.com](mailto:doc@mathworks.com)  
[service@mathworks.com](mailto:service@mathworks.com)  
[info@mathworks.com](mailto:info@mathworks.com)

Product enhancement suggestions  
Bug reports  
Documentation error reports  
Order status, license renewals, passcodes  
Sales, pricing, and general information



508-647-7000 (Phone)



508-647-7001 (Fax)



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

For contact information about worldwide offices, see the MathWorks Web site.

*Bioinformatics Toolbox™ Release Notes*

© COPYRIGHT 2003–2012 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](http://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](http://www.mathworks.com/patents) for more information.

<b>Summary by Version</b> .....	<b>1</b>
<b>Version 4.1 (R2012a) Bioinformatics Toolbox</b> Software .....	<b>5</b>
<b>Version 4.0 (R2011b) Bioinformatics Toolbox</b> Software .....	<b>10</b>
<b>Version 3.7 (R2011a) Bioinformatics Toolbox</b> Software .....	<b>16</b>
<b>Version 3.6 (R2010b) Bioinformatics Toolbox</b> Software .....	<b>21</b>
<b>Version 3.5 (R2010a) Bioinformatics Toolbox</b> Software .....	<b>28</b>
<b>Version 3.4 (R2009b) Bioinformatics Toolbox</b> Software .....	<b>32</b>
<b>Version 3.3 (R2009a) Bioinformatics Toolbox</b> Software .....	<b>41</b>
<b>Version 3.2 (R2008b) Bioinformatics Toolbox</b> Software .....	<b>44</b>
<b>Version 3.1 (R2008a) Bioinformatics Toolbox</b> Software .....	<b>51</b>
<b>Version 3.0 (R2007b) Bioinformatics Toolbox</b> Software .....	<b>58</b>
<b>Version 2.6 (R2007a+) Bioinformatics Toolbox</b> Software .....	<b>63</b>

<b>Version 2.5 (R2007a) Bioinformatics Toolbox</b>	
Software .....	66
<b>Version 2.4 (R2006b) Bioinformatics Toolbox</b>	
Software .....	74
<b>Version 2.3 (R2006a+) Bioinformatics Toolbox</b>	
Software .....	78
<b>Version 2.2.1 (R2006a) Bioinformatics Toolbox</b>	
Software .....	81
<b>Version 2.2 (R14SP3+) Bioinformatics Toolbox</b>	
Software .....	82
<b>Version 2.1.1 (R14SP3) Bioinformatics Toolbox</b>	
Software .....	84
<b>Version 2.1 (R14SP2+) Bioinformatics Toolbox</b>	
Software .....	85
<b>Version 2.0.1 (R14SP2) Bioinformatics Toolbox</b>	
Software .....	87
<b>Version 2.0 (R14SP1+) Bioinformatics Toolbox</b>	
Software .....	88
<b>Version 1.1.1 (R14SP1) Bioinformatics Toolbox</b>	
Software .....	92
<b>Version 1.1 (R14) Bioinformatics Toolbox Software ...</b>	93
<b>Version 1.0 (R13+) Bioinformatics Toolbox Software ..</b>	97
<b>Compatibility Summary for Bioinformatics Toolbox</b>	
Software .....	100

## Summary by Version

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

<b>Version (Release)</b>	<b>New Features and Changes</b>	<b>Version Compatibility Considerations</b>	<b>Fixed Bugs and Known Problems</b>
<b>Latest Version V4.1 (R2012a)</b>	Yes Details	Yes Summary	Bug Reports Includes fixes
V4.0 (R2011b)	Yes Details	Yes Summary	Bug Reports Includes fixes
V3.7 (R2011a)	Yes Details	Yes Summary	Bug Reports Includes fixes
V3.6 (R2010b)	Yes Details	Yes Summary	Bug Reports Includes fixes
V3.5 (R2010a)	Yes Details	Yes Summary	Bug Reports Includes fixes
V3.4 (R2009b)	Yes Details	Yes Summary	Bug Reports Includes fixes
V3.3 (R2009a)	Yes Details	Yes Summary	Bug Reports Includes fixes
V3.2 (R2008b)	Yes Details	Yes Summary	Bug Reports Includes fixes
V3.1 (R2008a)	Yes Details	Yes Summary	Bug Reports Includes fixes
V3.0 (R2007b)	Yes Details	Yes Summary	Bug Reports Includes fixes
V2.6 (R2007a+)	Yes Details	Yes Summary	Bug Reports Includes fixes
V2.5 (R2007a)	Yes Details	Yes Summary	Bug Reports Includes fixes
V2.4 (R2006b)	Yes Details	No	Bug Reports Includes fixes

<b>Version (Release)</b>	<b>New Features and Changes</b>	<b>Version Compatibility Considerations</b>	<b>Fixed Bugs and Known Problems</b>
V2.3 (R2006a+)	Yes Details	No	Bug Reports Includes fixes
V2.2.1 (R2006a)	No	No	Bug Reports Includes fixes
V2.2 (R14SP3+)	Yes Details	No	Bug Reports Includes fixes
V2.1.1 (R14SP3)	No	No	Bug Reports Includes fixes
V2.1 (R14SP2+)	Yes Details	No	Bug Reports Includes fixes
V2.0.1 (R14SP2)	Yes Details	No	Bug Reports Includes fixes
V2.0 (R14SP1+)	Yes Details	No	No bug fixes
V1.1.1 (R14SP1)	No	No	No bug fixes
V1.1 (R14)	Yes Details	No	No bug fixes
V1.0 (R13+)	Yes Details	No	No bug fixes

## Using Release Notes

Use release notes when upgrading to a newer version to learn about:

- New features
- Changes
- Potential impact on your existing files and practices

Review the release notes for other MathWorks® products required for this product (for example, MATLAB® or Simulink®). Determine if enhancements, bugs, or compatibility considerations in other products impact you.

If you are upgrading from a software version other than the most recent one, review the current release notes and all interim versions. For example, when you upgrade from V1.0 to V1.2, review the release notes for V1.1 and V1.2.

## **What Is in the Release Notes**

### **New Features and Changes**

- New functionality
- Changes to existing functionality

### **Version Compatibility Considerations**

When a new feature or change introduces a reported incompatibility between versions, the **Compatibility Considerations** subsection explains the impact.

Compatibility issues reported after the product release appear under Bug Reports at the MathWorks Web site. Bug fixes can sometimes result in incompatibilities, so review the fixed bugs in Bug Reports for any compatibility impact.

### **Fixed Bugs and Known Problems**

MathWorks offers a user-searchable Bug Reports database so you can view Bug Reports. The development team updates this database at release time and as more information becomes available. Bug Reports include provisions for any known workarounds or file replacements. Information is available for bugs existing in or fixed in Release 14SP2 or later. Information is not available for all bugs in earlier releases.

Access Bug Reports using your MathWorks Account.

## **Documentation on the MathWorks Web Site**

Related documentation is available on [mathworks.com](http://mathworks.com) for the latest release and for previous releases:

- Latest product documentation

- Archived documentation



## Version 4.1 (R2012a) Bioinformatics Toolbox Software

This table summarizes what's new in Version 4.1 (R2012a):

New Features and Changes	Version Compatibility Considerations	Fixed Bugs and Known Problems
Yes Details below	Yes—Details labeled as <b>Compatibility Considerations</b> , below. See also Summary.	Bug Reports Includes fixes

New and updated features in this version include:

- “Update to Jmol Functions” on page 5
- “Enhancements to Objects for NGS Data” on page 5
- “Enhancements to the NGS Browser” on page 7
- “Sequence Statistics Functions” on page 7
- “Demo for DNA Methylation Analysis” on page 7
- “Functionality Being Changed or Removed” on page 8

### 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 aaccount and basecount functions no longer accept the 'Others' name-value pair. Use the 'Ambiguous' or 'Gaps' name-value pair instead.
- The aaccount and basecount functions no longer accept the 'Structure' name-value pair. Use the 'Ambiguous' name-value pair with either 'ignore' or 'warn' instead.
- The aaccount, 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 5.
'FASTQFile', <i>File</i> pair as input to the BioRead constructor	Errors	<i>File</i>	See the Compatibility Considerations subheading in “Enhancements to Objects for NGS Data” on page 5.
'SAMFile', <i>File</i> pair as input to the BioRead or BioMap constructor	Errors	<i>File</i>	See the Compatibility Considerations subheading in “Enhancements to Objects for NGS Data” on page 5.
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.

<b>Functionality</b>	<b>What Happens When You Use This Functionality?</b>	<b>Use This Instead</b>	<b>Compatibility Considerations</b>
'Others' name-value pair as input to <code>aaccount</code> and <code>basecount</code> functions	Errors	'Ambiguous' or 'Gaps' name-value pair as input to <code>aaccount</code> and <code>basecount</code> functions	Replace instances of 'Others' with 'Ambiguous' or 'Gaps'.
'Structure' name-value pair as input to <code>aaccount</code> and <code>basecount</code> functions	Errors	'Ambiguous' name-value pair with either 'ignore' or 'warn' as input to <code>aaccount</code> and <code>basecount</code> functions	Replace instances of 'Structure' with 'Ambiguous' paired with 'ignore' or 'warn'.
Others field from the output structure returned by <code>aaccount</code> , <code>basecount</code> , <code>codoncount</code> , or <code>dimercount</code> .	Errors	Ambiguous field	Replace instances of Others (as an input) with Ambiguous.

## Version 4.0 (R2011b) Bioinformatics Toolbox Software

This table summarizes what's new in Version 4.0 (R2011b):

New Features and Changes	Version Compatibility Considerations	Fixed Bugs and Known Problems
Yes Details below	Yes—Details labeled as <b>Compatibility Considerations</b> , below. See also Summary.	Bug Reports Includes fixes

New and updated features in this version include:

- “Visualizing and Investigating Short-Read Alignments and Feature Annotations in the NGS Browser” on page 10
- “Objects for Genomic Feature Annotations” on page 10
- “Enhancements to BioRead and BioMap Objects” on page 11
- “Enhancements to the saminfo and baminfo Functions” on page 12
- “Conversion of Error and Warning Message Identifiers” on page 12
- “Function Elements Being Removed” on page 14

### 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 11.
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 11.
'FASTQFile', <i>File</i> pair as input to the BioRead constructor	Warns	<i>File</i>	See the Compatibility Considerations subheading in “Enhancements to BioRead and BioMap Objects” on page 11.
'SAMFile', <i>File</i> pair as input to the BioRead or BioMap constructor	Warns	<i>File</i>	See the Compatibility Considerations subheading in “Enhancements to BioRead and BioMap Objects” on page 11.

<b>Function Element Name</b>	<b>What Happens When You Use This Function Element</b>	<b>Use This Instead</b>	<b>Compatibility Considerations</b>
Indexed name-value pair argument as input to <code>getSubset</code> method of the <code>BioRead</code> or <code>BioMap</code> class	Warns	InMemory name-value pair argument	See the Compatibility Considerations subheading in “Enhancements to <code>BioRead</code> and <code>BioMap</code> Objects” on page 11.
Reference field of structure returned by <code>baminfo</code>	Errors	ScannedDictionary field	See the Compatibility Considerations subheading in “Enhancements to the <code>saminfo</code> and <code>baminfo</code> Functions” on page 12.

## Version 3.7 (R2011a) Bioinformatics Toolbox Software

This table summarizes what's new in Version 3.7 (R2011a):

New Features and Changes	Version Compatibility Considerations	Fixed Bugs and Known Problems
Yes Details below	Yes—Details labeled as <b>Compatibility Considerations</b> , below. See also Summary.	Bug Reports Includes fixes

New and updated features in this version include:

- “Data Format and Database Functions” on page 16
- “Sequence Statistics Functions” on page 17
- “Updates to the BioIndexedFile Class, Properties, and Methods” on page 17
- “Updates to BioRead and BioMap Classes and Methods” on page 18
- “Demos for High-Throughput Sequence Analysis” on page 19
- “Support Vector Machine (SVM) Functions” on page 19
- “Function Elements Being Removed” on page 20

### 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 Consideration

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

In Bioinformatics Toolbox Version 3.7, the `aaccount` 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 `IndexedByKey`.
- `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 `svmsmose` function is incorporated into the `svmtrain` and `statset` functions. Although `svmsmose` 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 <code>aaccount</code> and <code>basecount</code> functions	Errors	'Ambiguous' or 'Gaps' name-value pair as input to <code>aaccount</code> and <code>basecount</code> functions	See the Compatibility Considerations subheading in “Sequence Statistics Functions” on page 17.
'Structure' name-value pair as input to <code>aaccount</code> and <code>basecount</code> functions	Errors	'Ambiguous' name-value pair with either 'ignore' or 'warn' as input to <code>aaccount</code> and <code>basecount</code> functions	See the Compatibility Considerations subheading in “Sequence Statistics Functions” on page 17.
<code>getCoverage</code> method of <code>BioMap</code> class	Warns	<code>getBaseCoverage</code> , <code>getCounts</code> , and <code>getIndex</code> methods	See the Compatibility Considerations subheading in “Updates to BioRead and BioMap Classes and Methods” on page 18.
<code>svmsmoset</code> function	Still runs	<code>svmtrain</code> and <code>statset</code> functions	<code>svmsmoset</code> is not recommended. Use <code>svmtrain</code> and <code>statset</code> instead.



## Version 3.6 (R2010b) Bioinformatics Toolbox Software

This table summarizes what's new in Version 3.6 (R2010b):

New Features and Changes	Version Compatibility Considerations	Fixed Bugs and Known Problems
Yes Details below	Yes—Details labeled as <b>Compatibility Considerations</b> , below. See also Summary.	Bug Reports Includes fixes

New and updated features in this version include:

- “Data Format and Database Functions” on page 21
- “Sequence Conversion Functions” on page 22
- “Sequence Statistics Functions” on page 22
- “Pairwise Sequence Alignment Functions” on page 23
- “Multiple Sequence Alignment Functions” on page 23
- “Updates to BioMap Class, Methods, and Properties” on page 24
- “Function Elements Being Removed” on page 25

### 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:

- `aaccount` — 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 `aaccount` 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 `aaccount` and `basecount` functions still allow `'Others'` and `'Structure'` property name/property value pairs, but display a warning. Now the `aaccount` 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:

- `nwalignment` — 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 <code>aaccount</code> and <code>basecount</code> functions	Warns	'Ambiguous' or 'Gaps' property name/property value pair as input to <code>aaccount</code> and <code>basecount</code> functions	See the Compatibility Considerations subheading in “Sequence Statistics Functions” on page 22.
'Structure' property name/property value pair as input to <code>aaccount</code> and <code>basecount</code> functions	Warns	'Ambiguous' property name/property value pair with either 'ignore' or 'warn' as input to <code>aaccount</code> and <code>basecount</code> functions	See the Compatibility Considerations subheading in “Sequence Statistics Functions” on page 22.

<b>Function Element Name</b>	<b>What Happens When You Use This Function Element</b>	<b>Use This Instead</b>	<b>Compatibility Considerations</b>
'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 23.
'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 23.
The following properties of a clustergram object: <ul style="list-style-type: none"> <li>• ColumnMarker</li> <li>• Impute</li> <li>• Ratio</li> <li>• RowMarker</li> <li>• SymmetricRange</li> </ul>	Errors	New properties of a clustergram object: <ul style="list-style-type: none"> <li>• ColumnGroupMarker</li> <li>• ImputeFun</li> <li>• DisplayRatio</li> <li>• RowGroupMarker</li> <li>• Symmetric</li> </ul>	See “Clustergram Methods and Properties” on page 37.
'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 54.

<b>Function Element Name</b>	<b>What Happens When You Use This Function Element</b>	<b>Use This Instead</b>	<b>Compatibility Considerations</b>
'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 54.
pdbplot function	Errors	molviewer function	See the Compatibility Considerations subheading in “Protein Analysis and Sequence Utilities Functions” on page 68.
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 78.
mamadnorm and mameannorm functions	Errors	manorm function	See “Microarray Analysis” on page 90.

## Version 3.5 (R2010a) Bioinformatics Toolbox Software

This table summarizes what's new in Version 3.5 (R2010a):

New Features and Changes	Version Compatibility Considerations	Fixed Bugs and Known Problems
Yes Details below	Yes—Details labeled as <b>Compatibility Considerations</b> , below. See also Summary.	Bug Reports Includes fixes

New and updated features in this version include:

- “Data Format and Database Functions” on page 28
- “Pairwise Sequence Alignment Functions” on page 29
- “Multiple Sequence Alignment Functions” on page 29
- “Phylogenetic Tree Tools and Methods” on page 30
- “BioIndexedFile Function, Object, Methods, and Properties” on page 30
- “BioRead Function, Object, Methods, and Properties” on page 30
- “BioMap Function, Object, Methods, and Properties” on page 30
- “Function Elements Being Removed” on page 31

### 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 `'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 29
'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 29.

## Version 3.4 (R2009b) Bioinformatics Toolbox Software

This table summarizes what's new in Version 3.4 (R2009b):

New Features and Changes	Version Compatibility Considerations	Fixed Bugs and Known Problems
Yes Details below	Yes—Details labeled as <b>Compatibility Considerations</b> , below. See also Summary.	Bug Reports Includes fixes

New and updated features in this version include:

- “Data Format and Database Functions” on page 33
- “Protein Analysis Functions” on page 33
- “Data Visualization Functions” on page 33
- “Sequence Statistics Functions” on page 34
- “Sequence Utility Functions” on page 34
- “Sequence Visualization Functions” on page 35
- “Pairwise Sequence Alignment Functions” on page 35
- “Multiple Sequence Alignment Functions” on page 35
- “Phylogenetic Tree Tools and Methods” on page 36
- “Clustergram Window” on page 36
- “Clustergram Methods and Properties” on page 37
- “HeatMap Object, Methods, and Properties” on page 38
- “DataMatrix Methods” on page 39
- “Microarray Functions, Objects, Methods, and Properties” on page 39
- “Mass Spectrometry Functions” on page 39
- “Demos for Sequence Analysis” on page 40
- “Demos for Microarray Analysis” on page 40

## 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<sup>®</sup>, 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.
- `tgspread` — 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

## Version 3.3 (R2009a) Bioinformatics Toolbox Software

This table summarizes what's new in Version 3.3 (R2009a):

New Features and Changes	Version Compatibility Considerations	Fixed Bugs and Known Problems
Yes Details below	Yes—Details labeled as <b>Compatibility Considerations</b> , below. See also Summary.	Bug Reports Includes fixes

New and updated features in this version include:

- “Data Visualization Functions” on page 41
- “Sequence Utility Functions” on page 41
- “Sequence Conversion Functions” on page 42
- “Bioanalytic and Mass Spectrometry Functions” on page 42
- “Microarray Functions” on page 42
- “Demo for Sequence Analysis” on page 43

### 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.
- `illumbslookup` — Look up Illumina® BeadStudio™ target (probe) sequence and annotation information. Updated to read Illumina microRNA array annotation files.
- `illumbsread` — 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 `matstest` 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 `matstest` function used equal variance for the test. In Bioinformatics Toolbox Version 3.2, the `matstest` 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

## Version 3.2 (R2008b) Bioinformatics Toolbox Software

This table summarizes what's new in Version 3.2 (R2008b):

New Features and Changes	Version Compatibility Considerations	Fixed Bugs and Known Problems
Yes Details below	Yes—Details labeled as <b>Compatibility Considerations</b> , below. See also Summary.	Bug Reports Includes fixes

New and updated features in this version include:

- “Data Format and Database Functions” on page 44
- “Sequence Utility Functions” on page 46
- “Multiple Sequence Alignment Functions” on page 46
- “Gene Ontology Functions” on page 47
- “Protein Analysis Functions” on page 47
- “Mass Spectrometry Functions” on page 47
- “Microarray File Format Functions” on page 48
- “Microarray Functions” on page 49
- “DataMatrix Object” on page 49
- “DataMatrix Methods” on page 50
- “Demo for Visualization Tools” on page 50
- “Demo for Sequence Analysis” on page 50
- “Demos for Microarray Data Analysis” on page 50

### Data Format and Database Functions

Following are new functions:

- `affygcma` — 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:

- `affygcma` — 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.
- `affysnannotread` — 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.
- `affygcrrma` — 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 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 `affysnpintensitiesplit`.

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

## Version 3.1 (R2008a) Bioinformatics Toolbox Software

This table summarizes what's new in Version 3.1 (R2008a):

New Features and Changes	Version Compatibility Considerations	Fixed Bugs and Known Problems
Yes Details below	Yes—Details labeled as <b>Compatibility Considerations</b> , below. See also Summary.	Bug Reports Includes fixes

New and updated features in this version include:

- “Data Format and Database Functions” on page 51
- “Sequence Utility Functions” on page 52
- “Pairwise Sequence Alignment Functions” on page 53
- “Phylogenetic Tree Tools Function” on page 53
- “Protein Analysis Functions” on page 53
- “Microarray File Format Functions” on page 54
- “Microarray Functions” on page 54
- “Object” on page 56
- “Clustergram Methods” on page 56
- “Demo for Sequence Analysis” on page 57
- “Demo for Microarray Data Analysis” on page 57
- “Demo for Visualization Tools” on page 57
- “Demos for Mass Spectrometry Data Analysis” on page 57

### 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.

## Version 3.0 (R2007b) Bioinformatics Toolbox Software

This table summarizes what's new in Version 3.0 (R2007b):

New Features and Changes	Version Compatibility Considerations	Fixed Bugs and Known Problems
Yes Details below	Yes—Details labeled as <b>Compatibility Considerations</b> , below. See also Summary.	Bug Reports Includes fixes

New and updated features in this version include:

- “Data Format and Database Functions” on page 58
- “Microarray File Format Functions” on page 59
- “Microarray Functions” on page 59
- “Sequence Conversion, Utility, and Visualization Functions” on page 59
- “Mass Spectrometry Functions” on page 60
- “Statistical Learning Functions” on page 60
- “Gene Ontology Methods” on page 60
- “Demos for Microarray Data Analysis” on page 61
- “Demos for Sequence Analysis” on page 61
- “Demo for Graph Theory Analysis” on page 62

### 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:

- `mssalign` — 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

## Version 2.6 (R2007a+) Bioinformatics Toolbox Software

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

New Features and Changes	Version Compatibility Considerations	Fixed Bugs and Known Problems
Yes Details below	Yes—Details labeled as <b>Compatibility Considerations</b> , below. See also Summary.	Bug Reports Includes fixes

New and updated functions in this version include:

- “Data Formats and Databases Functions” on page 63
- “Microarray File Formats Functions” on page 64
- “Microarray Utility Functions” on page 64
- “Microarray Normalization and Filtering Functions” on page 65
- “Mass Spectrometry Functions” on page 65
- “Demos for Mass Spectrometry Functions” on page 65

### 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 Metabolic Mixtures Using Liquid Chromatography/Mass Spectrometry (LC/MS)

## Version 2.5 (R2007a) Bioinformatics Toolbox Software

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

New Features and Changes	Version Compatibility Considerations	Fixed Bugs and Known Problems
Yes Details below	Yes—Details labeled as <b>Compatibility Considerations</b> , below. See also Summary.	Bug Reports Includes fixes

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

- “Data Formats and Database Functions” on page 67
- “Demo for Data Formats and Database Functions” on page 67
- “Statistical Learning Functions” on page 68
- “Protein Analysis and Sequence Utilities Functions” on page 68
- “Demo for Protein Analysis and Sequence Utilities Functions” on page 69
- “Sequence Alignment Functions” on page 69
- “Demo for Sequence Alignment Functions” on page 70
- “Microarray File Formats Functions” on page 70
- “Microarray Normalization and Filtering Functions” on page 70
- “Demo for Microarray File Formats, Normalization, and Filtering Functions” on page 71
- “Microarray Data Analysis and Visualization Functions” on page 71
- “Demo for Microarray Data Analysis and Visualization Functions” on page 71
- “Mass Spectrometry Functions” on page 71
- “Phylogenetic Tree Tools Functions” on page 72
- “Demos for Phylogenetic Tree Tools Functions” on page 72
- “Phylogenetic Tree Methods” on page 73

## 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 programmatically 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  $\frac{1}{\sqrt{\epsilon}}$ . 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 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.
- `mssalign` — 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.
- `dndsm1` — 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.

## Version 2.4 (R2006b) Bioinformatics Toolbox Software

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

New Features and Changes	Version Compatibility Considerations	Fixed Bugs and Known Problems
Yes Details below	Yes Summary	Bug Reports Includes fixes

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

- “Data Formats and Database Functions” on page 74
- “Sequence Utilities Functions” on page 75
- “Sequence Visualization Functions” on page 75
- “Multiple Sequence Alignment Functions” on page 75
- “Microarray File Formats” on page 75
- “Microarray Data Analysis and Visualization Functions” on page 76
- “Graph Theory Functions” on page 76
- “Graph Visualization Methods” on page 77
- “Phylogenetic Tree Methods” on page 77

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

## Graph Theory Functions

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

- `graphallshortestpaths` — Find all shortest paths in graph.
- `graphconncomp` — Find strongly or weakly connected components in graph.
- `graphisdag` — Test for cycles in directed graph.
- `graphisomorphism` — Find isomorphism between two graphs.
- `graphisspantree` — Determine if tree is spanning tree.
- `graphmaxflow` — Calculate maximum flow and minimum cut in directed graph.
- `graphminspantree` — Find minimal spanning tree in graph.
- `graphpred2path` — Convert predecessor indices to paths.



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

## Graph Visualization Methods

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

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

## Phylogenetic Tree Methods

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

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

## Version 2.3 (R2006a+) Bioinformatics Toolbox Software

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

New Features and Changes	Version Compatibility Considerations	Fixed Bugs and Known Problems
Yes Details below	No	Bug Reports

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

- “Data Formats and Databases Functions” on page 78
- “Sequence Utilities Functions” on page 78
- “Sequence Visualization Functions” on page 79
- “Statistical Learning Functions” on page 79
- “Microarray Functions” on page 79
- “Demo for Microarray Functions” on page 80

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

### Sequence Utilities Functions

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

- `blastncbi` — Generate remote BLAST request.

## Sequence Visualization Functions

Following is a new function for visualizing sequence data:

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

## Statistical Learning Functions

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

- `svmtrain` — Train support vector machine classifier.

## Microarray Functions

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

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

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

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

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

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

Following are new functions for microarray differential expression analysis:

- `mattest` — Perform two-sample, two-tailed t-test to evaluate differential expression of genes from two experimental conditions or phenotypes.
- `mavolcanoplot` — Create significance versus gene expression ratio (fold change) scatter plot of microarray data.

### **Demo for Microarray Functions**

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

## Version 2.2.1 (R2006a) Bioinformatics Toolbox Software

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

<b>New Features and Changes</b>	<b>Version Compatibility Considerations</b>	<b>Fixed Bugs and Known Problems</b>
No	No	Bug Reports

## Version 2.2 (R14SP3+) Bioinformatics Toolbox Software

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

New Features and Changes	Version Compatibility Considerations	Fixed Bugs and Known Problems
Yes Details below	No	Bug Reports

New features and changes introduced in this version are

- “Multiple Sequence Alignment Viewer” on page 82
- “Microarray Functions for Agilent Software” on page 82
- “Gene Ontology Database Functions” on page 82
- “Demo for Gene Ontology Functions” on page 83

### 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.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`).

## Version 2.1.1 (R14SP3) Bioinformatics Toolbox Software

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

<b>New Features and Changes</b>	<b>Version Compatibility Considerations</b>	<b>Fixed Bugs and Known Problems</b>
No	No	Bug Reports



## Version 2.1 (R14SP2+) Bioinformatics Toolbox Software

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

New Features and Changes	Version Compatibility Considerations	Fixed Bugs and Known Problems
Yes Details below	No	Bug Reports

New features and changes introduced in this version are:

- “Sequence Alignment Functions” on page 85
- “Sequence Statistics Functions” on page 86
- “Sequence Utilities Functions” on page 86
- “Phylogenetic Tree Functions” on page 86
- “Phylogenetic Tree Methods” on page 86
- “Microarray Functions” on page 86
- “Statistics Functions” on page 86

### 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`, `dndsm1` — 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.
- `getnewickstr` — 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 `sqlfeatures`.

## Version 2.0.1 (R14SP2) Bioinformatics Toolbox Software

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

New Features and Changes	Version Compatibility Considerations	Fixed Bugs and Known Problems
Yes Details below	No	Bug Reports

New features and changes introduced in this version are

- “Updated RBASE Table” on page 87
- “Expanded Bioperl Demonstration” on page 87

### 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`.

## Version 2.0 (R14SP1+) Bioinformatics Toolbox Software

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

New Features and Changes	Version Compatibility Considerations	Fixed Bugs and Known Problems
Yes Details below	No	No bug fixes

New features and changes introduced in this version are

- “Mass Spectrometry Data Analysis” on page 88
- “Graph Visualization Object and Methods” on page 89
- “Statistical Learning” on page 89
- “Sequence Analysis” on page 89
- “Protein Analysis” on page 90
- “Microarray Analysis” on page 90
- “Updated Web Connectivity Function” on page 91

### Mass Spectrometry Data Analysis

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

- `msresample` — Resample with antialias filtering.
- `msbackadj` — Correct a baseline by estimation.
- `msalign` — Align a spectrum to a set of candidate peaks.
- `msheatmap` — Draw a heat map image for a set of spectra and check alignments.
- `msnorm` — Normalize a set of spectra.
- `mslowess` — Nonparametric smoothing using the Lowess method.
- `mssgolay` — Least-squares polynomial smoothing.

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

## Graph Visualization Object and Methods

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

- `biograph` — Function to create a biograph object.
- `dolayout` — Calculate node and edge positions.
- `getnodesbyid` — Get handles to nodes.
- `getedgesbynodeid` — Get handles to edges.
- `view` — Render a graph in its viewer.
- `getancestors` — Find ancestors.
- `getdescendants` — Find descendants.
- `getrelatives` — Find neighbors.

## Statistical Learning

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

- `classperf` — Evaluate the performance of a classifier.
- `crossvalind` — Cross-validation index generation.
- `knnclassify` — K-Nearest neighbor classifier.
- `knnimpute` — Impute missing data using the nearest neighbor method.
- `randfeatures` — Randomized subset feature selection.
- `sqt1features` — Sequential forward feature selection. This function will be renamed to `rankfeatures` in Version 2.1.
- `svmclassify` — Classify using a support vector machine classifier.
- `svmtrain` — Train a support vector machine classifier.

## Sequence Analysis

New functions for analysis and visualization of multiple sequences.

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

Updated functions.

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

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

```
matlab -useJavaFigures
```

## Protein Analysis

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

## Microarray Analysis

- `quantilenorm` — Quantile normalization.

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

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

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

---

**Note** If you use `mamadnorm` or `mameannorm` in any of your personal M-files, please update your files with the new function `manorm`. These functions are not recommended and may be removed from future releases of the Bioinformatics Toolbox software.

---

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

### **Updated Web Connectivity Function**

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

## Version 1.1.1 (R14SP1) Bioinformatics Toolbox Software

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

<b>New Features and Changes</b>	<b>Version Compatibility Considerations</b>	<b>Fixed Bugs and Known Problems</b>
No	No	No bug fixes



## Version 1.1 (R14) Bioinformatics Toolbox Software

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

New Features and Changes	Version Compatibility Considerations	Fixed Bugs and Known Problems
Yes Details below	No	No bug fixes

New features and changes introduced in this version are

- “Phylogenetic Analysis Functions” on page 93
- “Phylogenetic Tree Object and Methods” on page 94
- “Hidden Markov Model (HMM) Profiles” on page 94
- “BLAST Functions” on page 94
- “Microarray Functions” on page 95
- “Protein Analysis Function” on page 95
- “Sequence Alignment Functions” on page 95
- “New Demos” on page 95

### Phylogenetic Analysis Functions

New functions for phylogenetic tree creation and analysis.

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

- `seqpdist` — Calculate the pairwise distance between biological sequences.

## Phylogenetic Tree Object and Methods

New object for manipulating phylogenetic tree data.

- `phytree` — Function to create a `phytree` object.
- `get` — Get property values from a `phytree` object
- `getbyname` — Get node names from a `phytree` object.
- `pdist` — Calculate the patristic distances between pairs of leaf nodes.
- `plot` — Draw a phylogenetic tree object in a MATLAB Figure window as a phylogram, cladogram, or radial tree.
- `prune` — Remove nodes from a phylogenetic tree.
- `select` — Select branches and leaves from a phylogenetic tree using a specified criteria.
- `view` — Open a phylogenetic tree in a `phytreetool` window.

## Hidden Markov Model (HMM) Profiles

Updated Hidden Markov Model profile functions.

- The model structure that HMM functions use now includes loop and null transition probabilities. You can read null and loop probabilities from PFAM files using `pfamhmmread`, and, from PFAM Web databases, using `gethmmprof`.
- When the function `hmmprofstruct` builds an HMM model, the loop and null transition probabilities default to predefined values. If necessary, you can later modify the probabilities using the same function.
- `hmmprofalign` includes two new properties to control the scoring of flanking states and null transition probabilities. In addition, a third output argument with indices pointing to the respective symbols of the query sequence was added.

## BLAST Functions

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

## Microarray Functions

- `imageneread` — Read microarray data from an ImaGene® Results file.
- `affyread` — Read microarray data from Affymetrix GeneChip files.
- `gprread` — Read microarray data from GenePix® Results (GPR) files.
- `mapcaplot` — Create a Principal Component plot of expression profile data.
- `clustergram` — Updated function to do two way bi-clustering.

## Protein Analysis Function

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

## Sequence Alignment Functions

- `seqdisp` — Formats sequence output for easy viewing.
- `seqmatch` — Find matches for every string in a library.
- `seqdotplot` — Updated function now returns a second output (the matrix of matches as a sparse matrix).
- `aminolookup` , `baselookup` — Updated functions to get IUB/UPAC character codes, integer codes, and names for nucleotides and amino acids.

## New Demos

- **Bicluster demo** — Demonstrates some of the options of the `clustergram` function.
- **Bioperl demo** — Illustrates the interoperability between the MATLAB software and Bioperl, passing arguments from the MATLAB software to Perl scripts and pulling BLAST search data back to the MATLAB software.
- **Phytree demo for Hominidae species**— A phylogenetic tree is constructed from mtDNA sequences for the Hominidae taxa (also known as pongidae). This family embraces the gorillas, chimpanzees, orangutans and the humans.

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

## Version 1.0 (R13+) Bioinformatics Toolbox Software

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

New Features and Changes	Version Compatibility Considerations	Fixed Bugs and Known Problems
Yes Details below	No	No bug fixes

New features and changes introduced in this version are

- “Introduction to Bioinformatics Toolbox” on page 97
- “Databases and Data Formats” on page 98
- “Sequence Alignment” on page 98
- “Sequence Utilities and Statistics” on page 98
- “Microarray Normalization and Visualization” on page 98
- “Protein Structure Analysis” on page 99

### Introduction to Bioinformatics Toolbox

Bioinformatics Toolbox Version 1.0 (Web Release R13 SP1+) extends the MATLAB software with basic sequence analysis and gene expression analysis functions. Bioinformatics Toolbox is a collection of tools built on the MATLAB numeric computing environment. The toolbox supports a wide range of common sequence analysis and expression analysis tasks, from accessing Web-based databases, to sequence alignment, to microarray normalization and visualization.

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

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

help bioinfo

## **Databases and Data Formats**

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

## **Sequence Alignment**

The toolbox has functions for pairwise sequence alignment and for hidden Markov model sequence profile alignment, including efficient MATLAB implementations of the Needleman-Wunsch and Smith-Waterman algorithms. In addition to the alignment functions there are several tools for visualizing sequence alignments. The toolbox provides many standard scoring matrices, including the PAM and BLOSUM families.

## **Sequence Utilities and Statistics**

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

## **Microarray Normalization and Visualization**

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

## **Protein Structure Analysis**

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

## Compatibility Summary for Bioinformatics Toolbox Software

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

<b>Version (Release)</b>	<b>New Features and Changes with Version Compatibility Impact</b>
<b>Latest Version V4.1 (R2012a)</b>	See the <b>Compatibility Considerations</b> subheadings for: <ul style="list-style-type: none"> <li>• “Enhancements to Objects for NGS Data” on page 5</li> <li>• “Sequence Statistics Functions” on page 7</li> </ul> See “Functionality Being Changed or Removed” on page 8.
V4.0 (R2011b)	See the <b>Compatibility Considerations</b> subheadings for: <ul style="list-style-type: none"> <li>• “Enhancements to BioRead and BioMap Objects” on page 11</li> <li>• “Enhancements to the saminfo and baminfo Functions” on page 12</li> </ul> See “Function Elements Being Removed” on page 14.



<b>Version (Release)</b>	<b>New Features and Changes with Version Compatibility Impact</b>
V3.7 (R2011a)	<p>See the <b>Compatibility Considerations</b> subheadings for:</p> <ul style="list-style-type: none"> <li>• “Sequence Statistics Functions” on page 17</li> <li>• “Updates to BioRead and BioMap Classes and Methods” on page 18</li> <li>• “Support Vector Machine (SVM) Functions” on page 19</li> </ul> <p>See “Function Elements Being Removed” on page 20.</p>
V3.6 (R2010b)	<p>See the <b>Compatibility Considerations</b> subheadings for:</p> <ul style="list-style-type: none"> <li>• “Sequence Statistics Functions” on page 22</li> <li>• “Multiple Sequence Alignment Functions” on page 23</li> </ul> <p>See “Function Elements Being Removed” on page 25.</p>
V3.5 (R2010a)	See “Function Elements Being Removed” on page 31.
V3.4 (R2009b)	<p>See the <b>Compatibility Considerations</b> subheadings for:</p> <ul style="list-style-type: none"> <li>• “Data Visualization Functions” on page 33</li> <li>• “Phylogenetic Tree Tools and Methods” on page 36</li> <li>• “Clustergram Methods and Properties” on page 37</li> </ul>
V3.3 (R2009a)	<p>See the <b>Compatibility Considerations</b> subheading for: “Microarray Functions” on page 42</p>

<b>Version (Release)</b>	<b>New Features and Changes with Version Compatibility Impact</b>
V3.2 (R2008b)	<p>See the <b>Compatibility Considerations</b> subheadings for:</p> <ul style="list-style-type: none"> <li>• “Data Format and Database Functions” on page 44</li> <li>• “Microarray File Format Functions” on page 48</li> <li>• “Microarray Functions” on page 49</li> </ul>
V3.1 (R2008a)	<p>See the <b>Compatibility Considerations</b> subheadings for these changes:</p> <ul style="list-style-type: none"> <li>• “Data Format and Database Functions” on page 51</li> <li>• “Microarray Functions” on page 54</li> </ul>
V3.0 (R2007b)	<p>See the <b>Compatibility Considerations</b> subheading for:</p> <ul style="list-style-type: none"> <li>• “Statistical Learning Functions” on page 60</li> </ul>
V2.6 (R2007a+)	<p>See the <b>Compatibility Considerations</b> subheadings for:</p> <ul style="list-style-type: none"> <li>• “Data Formats and Databases Functions” on page 63</li> <li>• “Microarray File Formats Functions” on page 64</li> <li>• “Microarray Utility Functions” on page 64</li> </ul>
V2.5 (R2007a)	<p>See the <b>Compatibility Considerations</b> subheadings for these changes:</p> <ul style="list-style-type: none"> <li>• “Data Formats and Database Functions” on page 67</li> <li>• “Statistical Learning Functions” on page 68</li> <li>• “Protein Analysis and Sequence Utilities Functions” on page 68</li> <li>• “Sequence Alignment Functions” on page 69</li> <li>• “Microarray File Formats Functions” on page 70</li> <li>• “Phylogenetic Tree Tools Functions” on page 72</li> </ul>

<b>Version (Release)</b>	<b>New Features and Changes with Version Compatibility Impact</b>
V2.4 (R2006b)	None
V2.3 (R2006a+)	None
V2.2.1 (R2006a)	None
V2.2 (R14SP3+)	None
V2.1.1 (R14SP3)	None
V2.1 (R14SP2+)	None
V2.0.1 (R14SP2)	None
V2.0 (R14SP1+)	None
V1.1.1 (R14SP1)	None
V1.1 (R14)	None
V1.0 (R13+)	None