**Microsoft Biology Foundation**

**Beta 1.0 Release Notes**

We're happy to announce the first public release of the MBF framework – Beta 1!

This release comprised the entire set of code and binaries (including an installer) for the Microsoft Biology Foundation:

* Core MBF system library (bio.dll)
* Core MBF visualization library (MBF.Visualizations.Common.dll)
* Test harness using NUnit for validating functionality of core MBF components (bio.test.dll)
* Trident workflows for generating genomics workflows using the Trident workflow workbench (bio.workflow.dll)
* IronPython project to show how the MBF library of functional can be used with the Python scripting language (BioDemo.py)
* Microsoft Research Biology Extension for Excel, an Excel add-in, which exposes the capabilities of the MBF library in the form of Excel operations.
* Sequence Assembler, a WPF application provided as a sample/example of how to expose the MBF library features in a dedicated genomics program.
* Simulator, a small sample application which allows for splitting of large sequences into multiple shorter “reads” – intended to replicate what might come off a next-generation sequencing machine.

Additionally, we are providing a Beta 1 – Dev10 Preview release. The intention of this release is to provide a preview of where the MBF library of functionality is going. Included in this release is:

* Core MBF system library and visualization library, built against the .NET 4.0 Beta 2 release.
* **Support for .NET Parallel Extensions.**  This version of the code and binaries are utilizing the latest in technology improvements in the .NET 4.0 and Visual Studio 2010 Beta 2 release.  The inclusion of a novel multiple sequence alignment algorithm, PAMSAM, as been provided as an example of how this technology can be used to turn a commodity desktop computer into a valuable research asset.

These release notes are included in both beta releases (Beta 1 and Beta 1 – Dev10 Preview), but the releases are provided as separate downloads because of the relatively large file sizes of each. We recommend you try both, and encourage the review of the capabilities provided in the .NET 4.0 and Visual Studio 2010 Beta 2 offerings; we will continue to push the MBF library to become more and more dependent on these technologies and will likely deprecate any future MBF compatibility with previous versions of .NET and Visual Studio.

This is considered the first public production release of the framework.  Though we have had limited distribution Alphas to this point, we are excited to receive feedback from the larger community.

Please make sure to log feedback, suggestions and bugs via the connect site feedback system.  Issues input here will be logged directly against our internal database for issue tracking.  We will be paying attention to vote count to help us prioritize the issues we look at first.

Thank you in advance for helping to make the Microsoft Biology Foundation more relevant to your research!

The Microsoft Biology Framework Team

**Software requirements for the Beta1 release:**

**Mandatory:**

* **Core framework**
  + .NET 3.5 SP1

http://www.microsoft.com/downloads/details.aspx?FamilyID=ab99342f-5d1a-413d-8319-81da479ab0d7&DisplayLang=en

* + Visual Studio 2008 SP1

http://www.microsoft.com/downloads/details.aspx?FamilyId=FBEE1648-7106-44A7-9649-6D9F6D58056E

**Optional:**

* Excel Add-in
  + Microsoft Office Excel SDK Runtime for the Excel Add-In http://www.microsoft.com/downloads/details.aspx?familyid=5272E1D1-93AB-4BD4-AF18-CB6BB487E1C4&displaylang=en
  + Microsoft Visual Studio Tools for the Microsoft Office system (version 3.0 Runtime)

http://www.microsoft.com/downloads/details.aspx?familyid=54EB3A5A-0E52-40F9-A2D1-EECD7A092DCB&displaylang=en

* Silverlight 2.0 SDK for the Sequence Assembler application

http://www.microsoft.com/downloads/details.aspx?displaylang=en&FamilyID=c22d6a7b-546f-4407-8ef6-d60c8ee221ed

* NUnit 2.4.8 for the Test Framework

http://nunit.org/index.php?p=download

* IronPython 2.0.1 Runtime for the IronPython scripts

http://www.codeplex.com/IronPython

* Trident CTP for using Trident Workflows

http://research.microsoft.com/en-us/collaboration/tools/trident.aspx

**Software Requirements for the Beta1 Dev10 Preview release:**

**Mandatory**:

* Core framework
  + Visual Studio 2010 and .NET Framework 4 Beta 2

http://msdn.microsoft.com/en-us/vstudio/dd582936.aspx

**Optional**:

* NUnit 2.4.8 for the Test Framework

<http://nunit.org/index.php?p=download>

IMPORTANT NOTE:

It is recommended that the full version of Visual Studio 2008 SP1 or Visual Studio 2010 Beta 2 be used with this project to build and run properly. We recommend that you investigate obtaining a free copy of the full suite of Visual Studio products via the [DreamSpark](https://www.dreamspark.com/Products/ProductList.aspx) program at Microsoft (for students), or the [MSDN Academic Alliance](http://msdn.microsoft.com/en-us/academic/default.aspx) program (for faculty). Alternatively, you may use the provided installer to obtain runtime versions of these components for evaluation purposes.

Known Issues with Visual Studio C# Express:

* When using Visual Studio Express products to evaluate and build source code, only the following projects are supported in Visual Studio C# Express:
  + Bio
  + Bio.Test
  + Visualizations
  + Simulator
* SequenceAssembler will build in Visual Studio C# Express, but may not operate properly because of its dependence upon the Silverlight components.

New features in this release that are considered "fully implemented" and ready for evaluation:

* MBF Core
  + Object Model
    - Support for genomic interval data as defined by the UCSB BED format via a new SequenceRange object; also uses the SequenceRangeGrouping object to hold a series of these objects.
    - Similarity matrices have been refactored to allow easier extension/modification of the provided matrices.
  + Formatters/Parsers
    - GenBank Metadata enhancements – all forms of the metadata associated with a GenBank file are now supported.
    - SNP parser/formatter for reading/writing SNP objects into a Sequence object.
    - Contig Parser/Formatter for reading/writing a collection of SNPs and a contig to which they are aligned.
    - BED file format parser/formatters for reading/writing BED genomic interval files.
  + Algorithms
    - Support for the IUPAC standard ambiguity symbols as part of a consensus sequence.
    - Large performance improvement in the way streaming is employed as part of SuffixTree generation.
    - Multiple scores and consensus outputs now stored as part of a SequenceAlignment object.
    - Merge, Intersect and Subtract methods for the SequenceRangeGrouping objects (similar in functionality to the tools provided by the Galaxy web service).
    - PAMSAM – Parallelized Multiple Sequence Alignment on Microsoft Biology Foundation. This is a novel MSA algorithm based initially on the techniques employed in MUSCLE, but incorporates significant use of the .NET Parallel Extensions resulting in much greater use of multiple core processors. NOTE: This is part of a separate download package that is dependent on the recent release of .NET 4 Beta 2 and Visual Studio 2010 Beta 2.
  + Web Services
    - The web services component of the base library has been improved to better support web service connection extensions, especially those not directly related to BLAST queries.
    - NCBI and EBI Blast services have been updated to adhere to the newly formed web service connector refactor.
    - Azure BLAST web service integration with MBF library. Exposed via the Excel Add-in and Sequence Assembler application.
  + Self-Registration
    - We’ve implemented the ability to register self-contained components (ie., MyNewAlgorithm.dll) which are dependent upon the MBF framework (ie., bio.dll).
    - The purpose of this feature is to allow extension of the framework without having to re-compile, redistribute and/or revise the MBF base library distributed with this release.
    - For more information on how to use this feature see the RegistrationTest provided with the source code part of this release.
* Applications
  + Trident
    - Workflow Integration: many of the features supported by the MBF library are now also included as native workflow activities that can be used inside a Trident Workflow.
    - You can find the Trident documentation, binaries and examples here: <http://research.microsoft.com/en-us/collaboration/tools/trident.aspx>.
  + Excel Add-in
    - To use: double-click the ExcelWorkbench.vsto file in the binary installation location or where the results of the compilation of the source code are directed. This should automatically install the addin for use in Excel.
    - Major performance improvements in the way data is presented in worksheets.
    - Status provided in bottom left corner of Excel during execution of MBF operations.
    - Ability to distinguish a worksheet between a Sequence object and a SequenceRange object.
      * Associated tools/methods are enabled/disabled as appropriate depending on which type of worksheet is currently selected.
    - Ability to select the result of an alignment (consensus) as the input to a BLAST query.
    - Color coding of values is now possible via the Configuration option on the Bioinformatics ribbon.
    - Removed 3D Molecular Viewer button until this feature gets implemented.
  + Sequence Assembler
    - Be sure to install the Silverlight 2.0 SDK as the pre-req to properly build this application. Only installing Siverlight 3.0 SDK will not allow the project to build correctly.
  + IronPython
    - The addition of 5 new script-based operations have been introduced that provide some equivalency to the EMBOSS tools currently available on other platforms:
      * Concatenate Sequence
      * Strip non-alphabetic characters
      * Remove Poly-A tail from a sequence
      * Perform logical union of two sequence files
      * Find difference between two sequences
  + Simulator
    - Inclusion of this sample application as a simple way to create “short reads” out of an already aligned sequence.
    - Also a good example of how to create a very simple UI to operate with MBF features.
    - We will likely continue to invest in this tool as a utility for testing/benchmarking algorithms.
* Installer
  + A first-draft, simple installer has been provided which will install the binaries and associated applications on the user’s machine.
  + User must agree to the MS-PL license terms
  + The list of pre-requisites is given as links to the appropriate download/installation site, but no “automatic” installation of pre-requisites is provided.

Bug Fixes:

|  |  |  |
| --- | --- | --- |
| **ID** | **Priority** | **Description** |
| 141 | 1 | Bio.WorkFlow should not reference binaries from GAC |
| 269 | 1 | ExcelAddIn : ExcelAddIn Work items for M5 |
| 287 | 1 | ExcelAddIn Assembly Error |
| 289 | 1 | Source \*requires\* VS 2008 SP1 |
| 320 | 1 | SequenceRange Data: Need to support metadata from BED file |
| 339 | 1 | Blast Web Service : Support for Synchronous calls in NCBI and EBI |
| 345 | 1 | SequenceRangeGrouping : Intersect operation output is wrong for intersect two bed files with minimal overlap value. |
| 346 | 1 | SequenceRangeGrouping : Subtract operation output is wrong when we do subtract with NonOverlapping pieces of Intervals with minimal value |
| 370 | 1 | ExcelAddIn: Importing a GenBank file hangs Excel |
| 371 | 1 | ExcelAddIn: Versioning Issue for the excel-workbench project and it needs auto increment the version similar to Bio.dll |
| 373 | 1 | Bed:Same metadata information is displaying twice even if both the metadata are same when we merge file A with file A. |
| 377 | 1 | Sequence Assembly : Performance issue : Smith Waterman and NW align algorithms taking more time to align Sequences |
| 392 | 1 | BED parser : IndexOutOfRange exception when parsing BED file. |
| 422 | 1 | GFF Parser : Automation and Unit Test cases of GFF parser are failing |
| 428 | 1 | PamSam : Unit Test cases for PamSam are failing |
| 431 | 1 | ExcelAddIn : BED merge operation output is wrong when we merge single bed sheet |
| 432 | 1 | Add copy right information for all MBF framewrok dll. |
| 115 | 1 | Blast Web Service : Implementation for Canceling a web request with the Azure Blast service is not present |
| 237 | 1 | FeatureItem: MBF does not support these attributes at the property level (except for the type and an inflexible location property). |
| 238 | 1 | GenBank: Lack of Location object |
| 239 | 1 | Muscle MSA : PAMSAMMultipleSequenceAligner now takes 1.6 hr to finish alignment for 122 Dna sequences; It was only 102 seconds in the 19th August dpk |
| 245 | 1 | ExcelAddIn : Worksheets should dynamically update the columns when the value is changed |
| 253 | 1 | ExcelAddIn : Performance issue : Rendering Sequence alignment and contig is not consistent across the application |
| 254 | 1 | ExcelAddIn WebService : Unable to use the result of an alignment (consensus) to select BLAST and BlastWorksheet should not generate if there is no data recivied |
| 262 | 1 | ExcelAddIn ConsensusView : Color coding should apply to worksheet as well |
| 263 | 1 | ExcelAddIn Color : Excel’s ‘Conditional Formatting’ should be used define to cell fill coloring schemes for each molecule type (DNA, RNA, Protein). |
| 264 | 1 | ExcelAddIn : Dialog should launch with specifics of MBF library version and MS copyright information when the ExcelAddIn Icon button is clicked |
| 265 | 1 | ExcelAddIn Worksheet : Sequence data naming improvements. |
| 272 | 1 | Aligner : Conversion of [column, row] to [row, column] in Similarity Matrix and Aligners to improve read-ability |
| 277 | 1 | PAMSAM : DynamicProgrammingProfileAlignerSerial and DynamicProgrammingProfileAlignerParallel Align(lst Sequences) method and AlignSimple(lstSequences) has no implementation and is throwing InvalidOperationException |
| 278 | 1 | MUMmer : MUMmer takes long time to align a sequences |
| 282 | 1 | PAMSAM : Properties linked with Resource file is throwing unhandled exception |
| 283 | 1 | ParseXML: sequences omitted from multi-sequence XML output |
| 285 | 1 | Sequence Assembler crash when custom alignment range selection changed |
| 291 | 1 | Excel Addin BLAST Web Service (NCBI): Object reference not set to an instance of an object |
| 297 | 1 | Utilities: Intersect Method |
| 299 | 1 | Genomic Interval Tool: Merge Method |
| 309 | 1 | Snp Parser : Warnings in Snp Parser code while building the source depot |
| 311 | 1 | Unable to load Silverlight project files because they silverlight msbuild targets at non-standard location |
| 312 | 1 | ExcelAddIn: Need to disable menu options if no valid workbook is available. |
| 314 | 1 | Excel Add-in: Need to be able to select Consensus as SequenceData for BLAST. |
| 319 | 1 | Azure Blast Web Service : AzureBlastService.cs is throwing warnings while building |
| 325 | 1 | Sequence Assembler: It's throwing casting exception when MUMmer selected as an aligner |
| 328 | 1 | SequenceRange : Intersect SequenceRangeGrouping result contains duplicate SequenceRanges. |
| 329 | 1 | ExcelAddIn : Unable to parse FastQ file |
| 347 | 1 | ExcelAddIn : Default color coding for alphabets A,C,G,T/U should be implemented as mentioned in the repro steps. |
| 349 | 1 | ExcelAddIn : ExcelAddIn application hangs when loading the HumanReferenceGene BED file |
| 351 | 1 | ExcelAddIn : The resulting sheet of a BED operation cannot be used for further BED operations. |
| 352 | 1 | SequenceRangeGrouping : Metadata is not populating in resultant BED SequenceRange after performing any BED operation. |
| 250 | 2 | Default email parameters for the EBI web service should have "msrerbio@microsoft.com" email |
| 296 | 2 | BioIronPython : The 6th Option "Find differences between two sequence " is not generating any output for few set of sequences |
| 298 | 2 | GenomicInterval Tool: Subtract Method |
| 322 | 2 | Azure WebService : Protocol Exception (The remote server returned an unexpected response: (400) Bad Request) when search requests are submitted input sequences > 50KB in size, |
| 323 | 2 | Azure WebService : Web service parameters seem case sensitive, returns: The remote server returned an unexpected response: (400) ‘Bad Request’. |
| 331 | 2 | Installer : Addin folder to be included as part of Installer |
| 333 | 2 | Installer : NUnit binaries not be be part of the installer binaries |
| 335 | 2 | ExcelAddIn :While populating consensus view; the status bar says “Alignment completed” and cursor is also not in waiting state. User is not able to know if the operation is still going on |
| 341 | 2 | Installer: PDB files are not being added as part of installer |
| 342 | 2 | SequenceRangeGrouping : Framework should not allow to create SequenceRange with values Start Index greater than End Index. |
| 343 | 2 | ExcelAddIn :Newly added blank cell is also getting coloured |
| 353 | 2 | Excel WorkBench : Currently all binaries related to excel addin is not being copied to the binaries folder |
| 354 | 2 | Build : PamSam.Test binaries are not being put as part of the build |
| 355 | 2 | ExcelAddIn: Name for pop up windows is not proper for BED |
| 356 | 2 | ExcelAddIn: Label name is missing for Gap penalty parameter. |
| 357 | 2 | Installer : wid files are not being copied as part of Installer build |
| 359 | 2 | ExcelAddIn: Click on the "Cancel" button without selecting any BLAST service button pops an alert saying "Failed to cancel the blast request. Object reference not set to an instance of an object". |
| 360 | 2 | ExcelAddIn: While importing LargeSize.gbk, it throws an error message "Could not parse LargeSize.gbk : Object reference not set to an instance of an object". |
| 361 | 2 | ExcelAddIn: The colour is not removed on Sequence Wrap Around Column |
| 362 | 2 | Installer : EULA should be present in one of the installation wizard |
| 363 | 2 | Build : Source : Files from Testdata folder in Bio.Test project is missing |
| 364 | 2 | Build : Source : Files from TestUtils folder in Bio.TestAutomation project is missing |
| 366 | 2 | Build : Source : PopUps project build is failing in Visual Studio |
| 367 | 2 | Build : Source : PamSam project build is failing in Visual Studio |
| 368 | 2 | Build : Source : ExcelAddIn project build is failing in Visual Studio |
| 369 | 2 | ExcelAddIn: Metadata for FASTA\FASTQ files missing from import into cells |
| 372 | 2 | ExcelAddIn: Another Medium size fastq file is not getting parsed after parsing one fasta and one medium size fast q file |
| 375 | 2 | ExcelAddIn: Sequence Wrap Column with align all sheets option adds extra consensus view empty sheets with colors |
| 379 | 2 | ExcelAddIn : No error caption is getting displayed for web service popup error. |
| 381 | 2 | ExcelAddIn : Excel Addin VSTO crashes after inserting a protein item in dna sequence sheet and aligning it. |
| 382 | 2 | ExcelAddIn : Blast Web Service : Parameters Window should not close when clicking on Error popup. |
| 386 | 2 | Installer : All the binaries and pdb files should be moved to one folder |
| 423 | 2 | PamSam : Algorithm name for 'MUSCLE' has a spelling mistake |
| 425 | 2 | Installer : Extra character are being added for the hyperlinks |
| 430 | 2 | ExcelAddIn : Excel Workbench application hangs when configuring column sequence with empty or invalid value. |
| 267 | 2 | ExcelAddInWorkSheets : Auto-fit rows and auto-zoom out to allow the entire 80 column sequence fit in view when the data populates the worksheet on import (70% seems about right for my laptop) |
| 268 | 2 | ExcelAddIn : Web service status bar should use native Excel progress bar to show status of service (similar to Alignments). |
| 270 | 2 | Alignment Algorithm : Offset, Aligned Sequence, Score, Consensus incorrect for more than 2 sequence passed as input for local alignment |
| 273 | 2 | Parsers : Provide Overloading Parse() function to update the ISequence "IsReadyOnly" property |
| 274 | 2 | Alignment Algorithms : Maximum Overlap score not tracked correctly |
| 275 | 2 | Alignment Algorithm : PairwiseOverlapAligner : Traceback doesn’t stop when it reaches row 0 or column 0 |
| 276 | 2 | SimpleConsensusResolver : Logic in Simple consensus resolver doesn’t deal with ambiguous DNA input correctly |
| 279 | 2 | Alignment Smith Waterman: Current implementation of SmithWaterman returns only 1 best score alignment |
| 281 | 2 | BasicDerivedSequence : Constructor returns a valid sequence for RangeStart and RangeLength set to 0 |
| 284 | 2 | PAMSAM : PairwiseDistanceMatrixGenerator class distance matrix property is null after instantiating the class |
| 292 | 2 | Mummer : Mummer is not using "UseGapExtensionCost" when alignment with SmithWaterman Pairwise aligner. |
| 313 | 2 | ExcelAddIn: "SequenceData" should be the SequenceID |
| 316 | 2 | MBF : Rename Bio.Net to MBF across entire framework |
| 324 | 2 | NCBI & EBI blast services : Clean up old implementation of NCBI & EBI Blast web service |
| 327 | 2 | Hardcoded "UseBrowserProxy = true" causes NCBI Web Blast to fail. |
| 330 | 2 | Excel Addin: Remove 3D molecular viewer button |
| 332 | 2 | PamSam: Seperate test code and dev code for PamSam |
| 338 | 2 | SequenceRangeGrouping: SequenceRangeGrouping intersect operation output is wrong when intersect two identical SequenceRangeGroupings |
| 340 | 2 | SequenceRange : There is no validation for Negative values for Range start and end index in SequenceRange constructor. |
| 280 | 3 | Sparse Sequence : Sparse sequence ctor should accept parameter size instead of user explicitly needs to set the count property |
| 288 | 3 | Segmented Sequence: Suggestion to change Segmented sequence error message. |

**Previous version release notes:**

Build 1339.0 Release Notes – Alpha 3

New features in this release that are considered "fully implemented" and ready for evaluation:

* MBF Object Model
  + GenBank Metadata handling
  + Qualitative Sequence (to support quality metrics, i.e, FastQ file format)
  + Compound Sequence Item (to support ambiguity statistics in base pairs)
  + Formatters/Parsers
    - FASTQ
    - NEWICK
    - SNP
  + Alignment Algorithms
    - PAMSAM – Parallelized Multiple Sequence Alignment on Microsoft Biology Foundation. This is a novel MSA algorithm based initially on the techniques employed in MUSCLE, but incorporates significant use of the .NET Parallel Extensions resulting in much greater use of multiple core processors.
    - Performance Optimizations on all pair-wise aligners.
    - NUCMER Pairwise Global Alignment (extension of MUMmer)
* Applications
  + Excel Add-in
    - The MBF framework and all associated capabilities have been incorporated into a specialized Bioinformatics focused menu in Excel 2007 and 2010.
    - This is a early attempt to provide some useful functionality to the genomics community via increasing the value of Excel as a scratch pad for performing analysis on next gen sequence data.
    - We are seeking feedback on how we can improve this experience and what components are missing that would make this feature more relevant to your work.
  + Sequence Assembler
    - Inclusion of all new features as menu options for execution (ie., new algorithms).
    - Numerous bug fixes to improve functionality as well as tidy up the experience.
    - We are looking for feedback on how relevant this type of application would be to your work, and how this compares to your envisioned value of the Excel Add-in (ie., where should we spend our time).

Bug Fixes:

|  |  |  |
| --- | --- | --- |
| **ID** | **Priority** | **Description** |
| 116 | 1 | Consensus: Need to expand the design and business logic around consensus generation. |
| 131 | 1 | Exell Add-in/Sequence Assembler; allow users to manuall select parsers for unknown file types |
| 146 | 1 | ISequence: Need to include method for IndexOfKnownNonGap() and LastIndexOfKnownNonGap() |
| 210 | 1 | Snp Parser : Unable to parse a simple TSV file which is in source depot data folder |
| 217 | 1 | MuscleMSA : MuscleMultpleSequenceAlignment is throwig exception "Item not being mapped" while aligning RNA and Protein Sequences |
| 218 | 1 | Muscle MSA : MuscleMultpleSequenceAlignment is not returning aligned sequences for Protein sequences with Profile Function name as SymmetrizedEntropy |
| 219 | 1 | GenBank Features.:Not able to get all genBank features count upon parsing RNA sequence. |
| 223 | 1 | Serialized form of SparseSequence is taking up a large amount of data |
| 225 | 1 | SmithWaterMan : Offset value is not proper for SW algorithm |
| 226 | 1 | NUCmer : NUCmer should provide the alignment score |
| 227 | 1 | NUCmer : Break Length parameter is currently not exposed in NUCmer |
| 229 | 1 | SequenceAlignment : NeedlemanWunsch affine gap penalty initialization should be done using gap extension penalty |
| 230 | 1 | ExcelAddIn : Build warnings while building ExcelAddIn project in Visual Studio |
| 231 | 1 | Muscle MSA : The application is not responding and running with infinite loop in stage 3 if stage 3 score is less than stage 1 score |
| 232 | 1 | Sequence : Object reference not set to an instance of an object error while trying to create a protein Sequence using Nucleotides encoding. |
| 236 | 1 | Muscle MSA : PAMSAMMultipleSequenceAligner throws unhandled exception if sequence list is passed as null |
| 242 | 1 | No way to set Nucmer Align Parameters such as BreakLength,mumLength and etc. |
| 243 | 1 | Mummer : object reference not set to an instance of an object while aligning demo data. |
| 246 | 1 | ExcelAddIn : Error "Cannot run the macro display chart " when try attempt chart |
| 252 | 1 | ExcelAddIn WorkSheet: Data block should update properly when deleting sequence item. |
| 256 | 1 | Excel Addin: Use sequence ID's instead of generic "SequenceData" name. |
| 257 | 1 | Excell Add-in: names of worksheets should be derived from files being loaded |
| 258 | 1 | Build system needs to remove old builds by policy |
| 148 | 2 | Sequence Assembler : No way to select input parameters such as consensusThreshold,mergeThreshold,gapCost and so.. to assemble sequences using different algorithm. |
| 211 | 2 | NeedlemanWunsch Alignment : Fasta Parser is throwing unhandled exception if invalid sequence is passed. |
| 212 | 2 | NeedlemanWusnch Algorithm : Align method throws unhandled exception if similarity matrix is passed as null |
| 214 | 2 | NeedlemanWusch Alignment : Fasta Parser throws unhandled Exception if empty sequence fasta file is passed. |
| 216 | 2 | Muscle MSA : Align() method of MuscleMultpleSequenceAlignment.cs is throwing exception “Index out of Range” with ProfileScoreFunctionName as "WeightedEuclideanDistance" |
| 221 | 2 | Muscle MSA : BinaryTree Class CutTree() method throws InValidCastException if any valid edgeindex is passed to cut the tree |
| 222 | 2 | NUCmer : Object Reference error while getting an aligned sequence with MinimumScore property set |
| 224 | 2 | Sequence Assembler : Consensus should not display gap characters (-) instead It should display X or N or as is the case using Ambiguity characters |
| 228 | 2 | Muscle MSA: Dna sequences alignement throws exception with profile score method name as "PearsonCorrelation" |
| 233 | 2 | Muscle MSA : "EuclideanDistance" enum value of DistanceFunctionTypes is wrongly spelled as "EuclieanDistance" |
| 234 | 2 | Muscle MSA : PAMSAMMultipleSequenceAligner throws unhandled exception with invalid similarity Matrix |
| 235 | 2 | Muscle MSA : PAMSAMMultipleSequenceAligner throws unhandled Exception if dna sequences are passed and molecule type is "RNA" |
| 241 | 2 | SimilarityMatrix needs a clearer description of its implementation |
| 247 | 2 | ExcelAddIn : Re-organize groups in ExcelWork sheet with mentioned order (left->right) |
| 248 | 2 | ExcelAddIn : Sequence Data drop down list file names should be renamed as mentioned in repro steps section. |
| 249 | 2 | ExcelAddIn :Default worksheet configure value should be 80 columns. |
| 251 | 2 | ExcelAddIn : Names of aligners to choose from should just be the name of the algorithm |
| 255 | 2 | ExcelAddIn Consensus : Consensus view display should be modified with changes mentioned in the repro steps. |
| 271 | 2 | Algorithms : The consensus sequences should set sequenceID's. |
| 145 | 3 | Sequence Assembler : Word Size text box in blast parameters dialog should not be allow user to enter characters . |
| 167 | 3 | Sequence Assembler : Periods at the end of tool tips are inconsistent across application |
| 172 | 3 | Sequence Assembler : No Exit option in the Load menu |
| 174 | 3 | Sequence Assembler : While exiting the application there is should be one prompt for exit. |
| 176 | 3 | Sequence Assembler : For custom alignment view end point should be visible |
| 178 | 3 | Sequence Assembler : Custom Alignment top view has only start and end value. Need more values |
| 206 | 3 | Sequence Assembler : Sequence Editor window should not close when clicking on "OK" popup window. |
| 213 | 3 | Sequence : It throws unhandled exception if alphabet is passed as null with invalid sequence |
| 215 | 3 | Sequence Assembler : Tool tip not found for all controls in Sequence Assembler |

Build 1286.3 Release Notes - Alpha 2

New features in this release that are considered "fully implemented" and ready for evaluation:

* MBF Object Model
  + Virtual, Sparse, Segmented and Derived Sequence Interface and Classes
  + Compressed Entity Encoding and Decoding methods
  + Formatters/Parsers
    - GFF
  + Algorithms
    - MUMmer 3
  + Web Services
    - EBI Wu-Blast
* Prototype applications
  + Sequence Assembler: demonstrates the different functionalities and tries to illustrate via code, the usage of bio.net framework built using the Bio.NET framework. It now integrates with the web-services and also makes use of SilverMap to display the results.
  + IronPython: allows working with the bio.net framework from with-in the IronPython environment and demonstrates via code, the usage of bio.net framework.
  + SilverMap visual control for viewing Blast results (courtesy Queensland University of Technology).
* Test cases
  + We have added hundreds of test cases on the NUnit framework that we are making available for validating code changes by the community.  Be advised that you must download the nUnit framework separately from <http://sourceforge.net/projects/nunit/>.

Bug Fixes:

|  |  |  |
| --- | --- | --- |
| **ID** | **Priority** | **Description** |
| 11 | 1 | Copyright notice missing from many project code files |
| 20 | 1 | Alignment algorithm should accept two gap penalties but accepting only one |
| 95 | 1 | Clean up bio csproj file |
| 93 | 1 | Comments: should we include a reference about where our source test data came from in the comments of our test code? |
| 100 | 1 | Dynamic pairwise alignments do not create a Consensus |
| 106 | 1 | SparseSequence : Harded coded values found in SparseSequence.cs file |
| 108 | 1 | MUMmer: Need to iimplement affine gap version of Align method |
| 117 | 1 | BlastWebTest: Need to update email address |
| 104 | 1 | MUMmer : Suffix Tree: For a given sequence, the length of the Unique match is not correct |
| 127 | 1 | Clean up Exception handling to be more consistent |
| 136 | 1 | MUMmer : MUMmer is currently calculating the Overlap score twice which is not correct |
| 147 | 1 | Source Data: Need to finalize source data to use in demo |
| 151 | 1 | SilverMap: Including Silverlight projects result in a warning dialog in Visual Studio |
| 130 | 1 | Make use of "using' more consistent. |
| 132 | 1 | Nucleotide : ISequenceItem should enforce Nucleotide and Aminoacid classes to expose whether the symbol is gap or ambiguity or termination. |
| 203 | 1 | Exception thrown when serializing and deserializing empty SparseSequence |
| 134 | 2 | MUMmer : Start Node, End Node, Start Index and End Index of Suffix tree builder is not as expected |
| 169 | 2 | Sequence Assembler : Save & Save as buttons should be disabled on load |
| 173 | 2 | Sequence Assembler : Application is hanging when trying to cancel with spacebar input |
| 175 | 2 | Sequence Assembler : Drag both the slider till end and it removes the chart from the UI. |
| 179 | 2 | Sequence Assembler : Moving range selector when minimum range is selected. |
| 180 | 2 | Sequence Assembler : Application is not opening in Maximized state |
| 183 | 2 | Slider bar: Need to move events to background thread |
| 187 | 2 | MUMmer : MUMmer with Align() method is not returning correct score |
| 188 | 2 | MUMmer : Setting of Default property values |
| 198 | 2 | Sequence Assembler : sequences are not getting dragged from Sequence Tree View pane to Selected Sequences pane in second run |
| 113 | 2 | MUMmer : The minimum length of MUM should support 1 and not less than 2 |
| 208 | 2 | Remove the hardcoded file name for Application log |
| 128 | 2 | Contigs: Proptery or method |
| 129 | 2 | SequenceSplitter: Math error |
| 105 | 2 | MUMmer : Suffix Tree: For any of the Reference & Query sequence the FirstSequenceMumOrder is not getting updated and the value is always 0. |
| 118 | 2 | SparseSequence: Sequence item at zeroth position is not getting removed from Sparse Sequence instance with Remove() method |
| 119 | 2 | MUMmer : Mummer is currently not accepting RNA sequences and accepting only DNA sequences |
| 120 | 2 | EbiWuBlast : Passing "Format-Type" parameter with invalid values should through an error message |
| 121 | 2 | EBIWUBlastt : Passing program parameter with value "BLASTP" in upper case letters is throwing an exception. |
| 124 | 2 | EbiWuBlast : "Argument OutOfRangeException" for small sequence which does not have any significant hits found with input parameter. |
| 109 | 2 | NcbiBlastParameters : "Command" parameter has a invalid command "string". |
| 110 | 2 | Sequence : ByteArray : For \_valuesPerByte=1 the \_encodedValues byte[] object is getting updated |
| 101 | 2 | NcbiQBlast : Name and Description properties of NcbiQBlast is not getting set |
| 103 | 2 | NcbiBlastParameters : GENETIC\_CODE parameters has a invalid/incorrect range validation |
| 94 | 2 | WebAccessor : Exceptions are not thrown, instead return only false |
| 96 | 2 | NcbiQBlast : NCBI client should throw proper error message if the IE proxy setting was pointing to a server which was not resolving or was not a real proxy server |
| 97 | 2 | MUMmer : Alignment : Alignment score is not being calculated or currently the score of aligned sequence is 0 |
| 98 | 2 | Consensus view pending tasks |
| 99 | 2 | Code review: Consensus custom view |
| 88 | 2 | BlastXmlParser : Few properties which was got as part of response stream is not being saved |
| 89 | 2 | BlastWebTest : Infinite loop found in the unit test case for Web services |
| 90 | 2 | BlastXmlParser : Parsing empty xml file should throw an proper error message. |
| 57 | 2 | Sequence Assembler : Application is not throwing an exception even if we load empty fasta files. |
| 77 | 2 | Sequence Assembler : Molecule type inconsistency |
| 79 | 2 | Object Module : Incosistencies in casing in Alphabets and MoleculeType enums. |
| 85 | 2 | WebAccessor : \_proxy variable should be exposed for the user to update |
| 86 | 3 | Web Services : Not all files has the CopyRight information |
| 67 | 3 | FastA Parser : ParseOne(filename) should not thorw any exception if fastA file contains more than one sequence. |
| 92 | 3 | NcbiQBlast : Submitting Search request without search query should throw an proper error message. |
| 111 | 3 | Ncbi4NaEncoding : Thymine Nucleotide object name among the list of nucleotides is not matching with expected nucleotide object name |
| 112 | 3 | Ncbi4NaEncoding,IupacNaEncoding: The class summary comment contains wrong nucleotide information |
| 202 | 3 | Sequence Assembler : Change the color scheme on click of canecl button |
| 171 | 3 | Sequence Assembler : Overlapping characters when we decrease app window size |
| 205 | 3 | Sequence Assembler : Hot key for Help Menu is not present. |

Build 1256.2 Release Notes Alpha 1

Features in this release that are considered "fully implemented" and ready for evaluation:

* MBF Object Model
  + Sequence Interface and Classes
  + Encoding methods
    - Alphabets
    - Transcription and Translation
  + Formatters/Parsers
    - FASTA
    - GenBank
  + Algorithms
    - Needleman-Wunsch
    - Smith-Waterman
  + Similarity Matrices
  + Web Services
    - NCBI QBLAST

Other components implemented but not complete.  These should be evaluated for design and structure, but not as complete features:

* Bio.NET Object Model
  + Sparse Sequences
  + Derived Sequences
  + Metadata
  + Formatters/Parsers
    - GFF
    - ClustalW
  + Algorithms
    - Mummer
  + Suffix Tree

Bug Fixes:

|  |  |  |
| --- | --- | --- |
| **ID** | **Priority** | **Description** |
| 19 | 1 | Blocking: Disposing of object created is not being done in BasicSequenceFormatter class. |
| 21 | 1 | Hard coded values found in Similarity Matrix Constructor |
| 22 | 1 | Out of memory exception when the sequence file size goes beyond 9 KB |
| 23 | 1 | Aligned sequences are padded with A in the end and dropping characters at beginning for sequence of 1KB |
| 24 | 1 | Sequence Object Model : \_statistics is not getting initialized from all constructors |
| 10 | 1 | GenBankFormatter not checking to Flush() items written to Writer |
| 12 | 1 | Blocking: Not able to complement the newly created DNA sequence using Complement(dnaSource) method |
| 13 | 1 | Blocking: Not able to transcribe the newly created sequence using Transcribet(dnaSource) method |
| 16 | 1 | Can't add a U symbol to an RNA sequence. |
| 17 | 1 | Transcription class incorrect for G, C |
| 38 | 1 | Encoding : IupacNAEncoding encoding LookUpBySymbol(string) method throwing an exception,if we pass string as a parameter. |
| 39 | 1 | BasicDerivedSequence : CopyTo mehod in BasicDerivedSequence class should start copying from passed sequence array index instead of 0th index. |
| 54 | 1 | Need to be able to select All Files (\*.\*) as type in Load dialog for Sequence Assembler |
| 72 | 1 | SequenceAssembler: No way to remove sequences from Sequence Tree View |
| 73 | 1 | Seqeunce Assembler : Application should not get crash upon editing Assembled and UnAssembled sequences with invalid sequence. |
| 66 | 1 | Build: Need to propogate source code to build share for each build |
| 81 | 1 | NcbiQBlast : No framework abstraction code for Blast Web services. |
| 82 | 1 | NcbiQBlast : SubmitSearchRequest(ISequence sequence) in NcbiQBlast is taking input sequence and doing nothing with it. |
| 83 | 1 | NcbiQBlast : Possible values for NCBI Parameters are not exposed. |
| 71 | 2 | Sequence Assembler : Opening same invalid Fasta file 2nd time, gives different error message |
| 55 | 2 | SequenceAssembler: Sequence items style and Sequence file name styles should be constant. |
| 56 | 2 | Seqeunce Assembler :Sequence Items should not add to selected sequence pane upon double clicking on right mouse button. |
| 58 | 2 | FastA Parser : Parse method should throw an exception or warning message, if we parse empty fasta file. |
| 59 | 2 | FastA Parser: Exception in parsing FastA file which has first line empty and sequences start with second line in the file. |
| 60 | 2 | Seqeunce Assembler : Unassembled Sequences are saving as empty files rather than saving unassembled sequences in the saved file. |
| 61 | 2 | SequenceAssembler : Buttons styles should be constant across the application |
| 78 | 2 | Sequence Assembler : Empty seqeunce should not be saved. |
| 40 | 2 | Sequences : MoleculeType property in Sequence class should get and set value to BasicSequenceInfo. |
| 41 | 2 | MetadataListItem class should implement ISerializable. |
| 18 | 2 | Alignment algorithms do not set default properties |
| 30 | 2 | Sequence Assembler application should have icon |
| 31 | 2 | Minimum Height of the panes in the application should not be fixed. |
| 33 | 2 | In correct ghost text in Sequence assembly |
| 34 | 2 | Application shows the warning message in a incorrect scenario. |
| 25 | 2 | Sequence count is not getting updated in the selected sequence pane |
| 26 | 2 | Upon doing second assemble,Dialog prompt with Yes/No has reverse logic. |
| 27 | 2 | Total length in Assembly report is not getting displayed when a contig is not generated |
| 2 | 2 | Missing GenBank file in GenBankTest |
| 3 | 2 | Product Studio for Bio.Net does not support proper build numbers |
| 4 | 2 | Product Studio for Bio.Net does not have any Path information set |
| 5 | 2 | Product Studio for Bio.Net needs vendor registration |
| 6 | 2 | No differentiation for Sequence Formatters that support writing multiple sequences |
| 7 | 2 | No file option for formatting in ISequenceFormatter |
| 8 | 2 | Fasta formatter depends on holding the entire sequence in string format before writing to file |
| 9 | 2 | Unnecessary folder in project structure PairwiseAlignerImplementations |
| 28 | 3 | Total Processing Time = (End-Start) is not generating in Assembly report |
| 29 | 3 | Name of the files should be displayed in sequence tree view instead of file1/file2.. |
| 42 | 3 | SequenceAssembler: Title color needs to be enhanced |
| 44 | 3 | SequenceAssembler: All pop-ups come up as another item in the task bar |
| 45 | 3 | SequenceAssembler: File names are not readable in file open dialog |
| 46 | 3 | SequenceAssembler: Custom dialog with "OK" pop-up has too much space at the bottom. |
| 47 | 3 | SequenceAssembler: Scroll bars should not be visible by default. |
| 48 | 3 | SequenceAssembler: Tabbing to a control doesnt have a hover effect. |
| 49 | 3 | SequenceAssembler: No 'Esc' key support for Pop-ups |
| 50 | 3 | SequenceAssembler: No short cuts for Load menu |
| 51 | 3 | SequenceAssembler: Algo drop-down is not alphabetically sorted |
| 52 | 3 | SequenceAssembler: Rename consenus items to consensus tree view |
| 53 | 3 | SequenceAssembler: Cannot tab to 'x' button of the sequence on selected sequence. |
| 62 | 3 | Sequence Assembler :Label Molecule Type should have colon next to it. |
| 63 | 3 | SequenceAssembler : On Tab, the focus style need to applied to all Buttons and need to remove the default focus style with dotted lines. |
| 64 | 3 | GeneBank Parser : Exception or message should be thrown upon parsing empty GeneBank file using GeneBank parser. |
| 74 | 3 | Sequence Assembler : Popup with only one button(Say "OK" button) should work upon clickin on both Escape as wll as clicking on ENTER button. |
| 75 | 3 | Sequence Assembler : There is no Tool tip for the buttons in the application. |
| 76 | 3 | Sequence Assembler : There is too much space at the bottom of the Sequence File Browser dialogue. |