Biopython Project Update 2017

Sourav Singh[a], Christian Brueffer[b], Peter Cock[c],
and the Biopython Contributors

18th Bioinformatics Open Source Conference (BOSC) 2017, Prague, CZ

Website: http://biopython.org
Repository: https://github.com/biopython/biopython
License: Biopython License Agreement (BSD like, see http://www.biopython.org/DIST/LICENSE)

The Biopython Project is a long-running distributed collaborative effort, supported by the Open Bioinformatics Foundation, which develops a freely available Python library for biological computation [1].

We present here details of the latest Biopython release – version 1.69. New features include: a new parser for the ExPASy Cellosaurus cell line database, catalogue and ontology; Bio.AlignIO now supports the UCSC Multiple Alignment Format (MAF), including indexed access to large files using SQLite3; Bio.SearchIO.AbiIO can now parse FSA files; an extended Bio.Affy module supporting version 4 of the Affymetrix CEL format; updated Uniprot parsers to support the "submittedName" XML element and features with unknown locations; better handling of internal node comments in the NEXUS parser to improve compatibility with tools such as BEAST TreeAnnotator; an update to Bio.Restriction to include the REBASE February 2017 restriction enzyme list; updated Bio.SeqIO parsers for GenBank, EMBL, and IMGT that now record the molecule type from the LOCUS/ID line in the record.annotations dictionary and can cope with more format variations; Bio.PDB.PDBList now can download PDBx/mmCif (new default), PDB (old default), PDBML/XML and mmtf format protein structures; enhanced PyPy support by taking advantage of NumPy and compiling most of the Biopython C code modules; the Bio.Seq module now offers a complement function for consistency and a SeqFeature object's qualifiers attribute is now an explicitly ordered dictionary.

Biopython 1.69 included the start of a re-licensing plan, to transition away from our liberal but unique *Biopython License Agreement* to the similar but very widely used *3-Clause BSD License*. We are reviewing the code base authorship file-by-file, in order to gradually dual license the entire project.

We are currently preparing a new release – version 1.70 – that will feature an extended Bio.AlignIO module that supports Mauve's eXtended Multi-FastA (XMFA) file format.

Additionally we fixed miscellaneous bugs, enhanced our test suite and continued our efforts to abide by the PEP8 and PEP257 coding style guidelines which is now checked automatically with GitHub-integrated continous integration testing using TravisCI. Current efforts include improving the unit test coverage, which is easily viewed online at CodeCov.io.

References

[1] Cock, P.J.A., Antao, T., Chang, J.T., Chapman, B.A., Cox, C.J., Dalke, A., Friedberg, I., Hamelryck, T., Kauff, F., Wilczynski, B., de Hoon, M.J. (2009) Biopython: freely available Python tools for computational molecular biology and bioinformatics. *Bioinformatics* **25**(11) 1422-3. doi:10.1093/bioinformatics/btp163

[a]Affiliation. Email: ssouravsingh12@gmail.com
[b]Department of Clinical Sciences, Lund University, Lund, SE
[c]Information and Computational Sciences, James Hutton Institute, Invergowrie, Dundee, UK