Biopython Project Update 2017

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

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 Biopython releases since BOSC 2016, namely Biopython 1.68, 1.69 and 1.70. Together these had 82 named contributors including 51 newcomers which reflects our policy of trying to encourage even small contributions.

Biopython 1.68 (August 2016) was a relatively small release, with the main new feature being support for RSSB's new binary Macromolecular Transmission Format (MMTF) for structural data.

Biopython 1.69 (April 2017) represents the start of our 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.

Major new features include: a new parser for the ExPASy Cellosaurus cell line database, catalogue and ontology; support for the UCSC Multiple Alignment Format (MAF), FSA sequencing files, version 4 of the Affymetrix CEL format; updates to the REBASE February 2017 restriction enzyme list; Bio.PDB.PDBList now can download more formats including MMTF; enhanced PyPy support by taking advantage of NumPy and compiling most of the Biopython C code modules.

Biopython 1.70 (July 2017) has internal changes to better support the now standard `pip` tool for Python package installation. Major new features include: support for Mauve's eXtended Multi-FastA (XMFA) file format, updates to our BLAST XML and MEME parsers, ExPASy support, and phylogenetic distance matrices. This release is noteworthy for our new logo (below), contributed by Patrick Kunzmann. This draws on our original double helix logo (above), and the blue and yellow colors of the current Python logo:

All releases fixed miscellaneous bugs, enhanced the test suite, and continued efforts to follow the PEP8 and PEP257 coding style guidelines which is now checked automatically with GitHub-integrated continuous integration testing using TravisCI. We now also use AppVeyor for continuous integration testing under Windows. 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]Dept. of Computer Engineering, VIIT, Pune, India. Email: ssouravsingh12@gmail.com
[b]Department of Clinical Sciences, Lund University, Lund, SE
[c]Information and Computational Sciences, James Hutton Institute, Invergowrie, Dundee, UK
[d]See contributor listing on GitHub.