Biopython Project Update

Peter Cock,* *et al.*

Bioinformatics Open Source Conference (BOSC) 2013, Berlin, Germany

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

In this talk we present the current status of the Biopython project, a long-running, distributed collaboration producing a freely available Python library for biological computation [1]. Biopython is supported by the Open Bioinformatics Foundation (OBF).

Since BOSC 2012, we will have made two releases, and a manuscript describing the `Bio.Phylo` module for phylogenetics has been published [2]. All our releases have had more unit tests, more documentation, and new contributors. In addition to nightly unit tests run on Linux, Windows and Mac OS X via an OBF hosted buildbot, we now run continuous integration tests via TravisCI. Together this covers the main operating systems (Linux, Mac OS X and Windows) and Python implementations (Jython, PyPy and both versions 2 and 3 of the primary Python implementation in C).

Biopython 1.61 (February 2013) included a number of small enhancements and additions. This release also introduced the idea of including some *beta* modules within the standard install for testing new experimental code, used to ship some of the GSoC work (see below).

Biopython has for some time run under Python 3, and we have been asking users to try this for testing purposes for some time. Biopython 1.62 (expected May 2013) is intended be our first release to officially support Python 3. [*NB: to be revised during abstract review*].

In Summer 2012 we had two Google Summer of Code (GSoC) students. Wibowo 'Bow' Arindrarto wrote a Biopython equivalent to BioPerl's SearchIO covering sequence search results from BLAST, HMMER, FASTA etc able to read and write multiple file formats with a common object model, and index large files for memory efficient random access. Lenna Peterson worked on the representation and manipulation of genomic variants (HGVS, GFF, VCF files). Both students finished the GSoC programme successfully, and have continued to be contribute to the project since. Bow's SearchIO code has already been included in recent releases, marked as experimental for wider testing and feedback. Biopython hopes to mentor more students for GSoC 2013, for example some phylogenetics project ideas have been put forward which would be mentored under the NESCent project. [*NB: to be revised during abstract review*].

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

[2] Talevich, E., Invergo, B.M., Cock, P.J.A., and Chapman, B.A. (2012) Bio.Phylo: A unified phylogenetics toolkit for Biopython. *BMC Bioinformatics* **13**:209. doi:10.1186/1471-2105-13-209

*Information and Computational Sciences, James Hutton Institute (formerly SCRI), Invergowrie, Dundee DD2 5DA, UK