



Biopython Project Update

Eric Talevich*, Peter Cock†, Brad Chapman‡ *et al.*

Bioinformatics Open Source Conference (BOSC) 2012, Long Beach, California, USA

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 2011, we have made two releases. With Biopython 1.58 (August 2011), we included support for ABI chromatogram files, and also gained interoperability with PAML by merging an independent project, Brandon Invergo's pypaml. Biopython 1.59 (February 2011) added support for TogoWS, an integrated web resource for bioinformatics databases and services, and new features in GenomeDiagram [2]. Biopython 1.60 is expected to have been released by BOSC 2012 [NB: to be revised during abstract review]. All releases have seen more unit tests, more documentation, and more new contributors. Additionally, we have submitted a paper describing the recent `Bio.Phylo` module for phylogenetics.

In summer 2011 we had three Google Summer of Code (GSoC) students, all building features for protein structure analysis: Mikael Trellet (biomolecular interface analysis for `Bio.PDB`), Michele Silva (Python bridge for `Mocapy++` [3] and linking it to `Bio.PDB` to enable statistical analysis of protein structures), and Justinas Daugmaudis (Python-based extension system for `Mocapy++`). Previous GSoC students João Rodrigues and Eric Talevich are working to refactor and merge the results of these projects under a unified API covering the PDB, mmCIF and PBDML formats for 3D molecular structures.

Two additional students are expected to work with us for GSoC 2012 to implement a Biopython equivalent to BioPerl's SearchIO (covering BLAST, HMMER, FASTA etc. search results), and representation and manipulation of genomic variants (HGVS, GFF, VCF files) [NB: to be revised during abstract review].

We are now encouraging early adopters to help beta test Biopython under Python 3 and PyPy. The use of nightly unit tests via the OBF BuildBot server (<http://testing.open-bio.org/>) continue to be very helpful for cross-platform validation (covering Windows, Linux and Mac OS X for Python 2.5–2.7, Jython 2.5, plus Python 3.1–3.2 and soon PyPy) as well as catching general regressions.

Mailing list discussions continue to be active, with lots of new work and new contributors coming forward.

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] Pritchard, L., White, J.A., Birch, P.R., Toth, I. (2010) GenomeDiagram: a python package for the visualization of large-scale genomic data. *Bioinformatics* **2**(5) 616-7. doi:10.1093/bioinformatics/btk021
- [3] Paluszewski, M., and Hamelryck, T. (2010) Mocapy++—a toolkit for inference and learning in dynamic Bayesian networks. *BMC Bioinformatics* **11** 126. doi:10.1186/1471-2105-11-126

*Institute of Bioinformatics, University of Georgia, Athens, GA, USA

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

‡Bioinformatics Core Facility, Harvard School of Public Health, Harvard University, Boston, MA, USA