



Biopython Project Update

Peter Cock*

Bioinformatics Open Source Conference (BOSC) 2009, Stockholm, Sweden

In this talk we present the current status of the Biopython project (www.biopython.org), focusing on features developed in the last year, and future plans. The Oxford University Press journal *Bioinformatics* has recently published an application note describing Biopython (Cock *et al.*, 2009).

Since BOSC 2008, Biopython has seen two releases. Biopython 1.49 (November 2008) was an important milestone in bringing support for Python 2.6, and in terms of our dependence on Numerical Python as we made the transition from the obsolete Numeric library to NumPy. Biopython 1.49 also added more biological methods to our core sequence object. April 2009 saw the release of Biopython 1.50, new features include:

- `GenomeDiagram` by Leighton Pritchard (Pritchard *et al.*, 2006) has been integrated into Biopython as the `Bio.Graphics.GenomeDiagram` module.
- A new module `Bio.Motif` has been added, which is intended to replace the existing `Bio.AlignAce` and `Bio.MEME` modules.
- `Bio.SeqIO` can now read and write FASTQ and QUAL files used in second generation sequencing work.

Biopython will celebrate its 10th Birthday later this year, and has now been cited or referred to in over one hundred scientific publications (a list is included on our website). We will present a brief history of the project and current work. This includes the evaluation of git (and github) as a possible distributed version control system (DVCS) to replace our existing very stable CVS server hosted by the Open Bioinformatics Foundation, which we hope will encourage more participation in the project. Also, we currently have two *Google Summer of Code* project students working on phylogenetics code for Biopython in conjunction with the National Evolutionary Synthesis Center (NESCent).

Biopython is free open source software available from www.biopython.org under the Biopython License Agreement (an MIT style license, <http://www.biopython.org/DIST/LICENSE>).

References

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* doi:10.1093/bioinformatics/btp163

Pritchard, L. *et al.* (2006) `GenomeDiagram`: A Python Package for the Visualisation of Large-Scale Genomic Data, *Bioinformatics* **22**(5), 616-617. doi:10.1093/bioinformatics/btk021

*Peter.Cock@scri.ac.uk – Plant Pathology, SCRI, Invergowrie, Dundee DD2 5DA, UK