Biopython Project Update 2019

Peter Cock[a],
and the Biopython Contributors[b]

20th Bioinformatics Open Source Conference (BOSC) 2019, Basel, Switzerland

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]. This talk will look ahead to the year to come, and give a summary of the project news since the 1.72 release in June 2018, and the talk at GCCBOSC 2018.

While there were no major new modules introduced in Biopython 1.73 (December 2018) or Biopython 1.74 (expected May/June 2019), there have been lots of incremental improvements. In terms of lines of code changed, a substantial proportion have been in-line documentation (Python docstrings), used to generate human readable API documentation. While we are still using `epydoc` for this, our continuous integration system has been generating more modern HTML output using `sphinx`, which we hope to host on our domain, or at Read The Docs, making this work much more visible to the world. We have been using the tool `flake8` with various plugins for this (as well as checking coding style), showing a steady improvement in best practice compliance - every public API should be documented this year.

In 2017 we started 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, to gradually dual license the entire project. All new contributions are dual licensed, and currently half the Python files in the main library have been dual licensed.

Another important going effort is improving the unit test coverage. Sadly This is currently fairly static at about 85% (excluding online tests), but can be viewed online at CodeCov.io.

We are using GitHub-integrated continuous integration testing on Linux (using TravisCI) and Windows (using AppVeyor), including enforcing the Python PEP8 and PEP257 coding style guidelines. We hope to be able to recommend a simple `git pre-commit` hook for our contributors shortly, and have discussed the idea of adopting the new yet popular Python code formatting style tool `black` to reduce the human time costs in writing compliant code.

Looking further ahead, in 2020, in line with most major scientific Python libraries, we will be dropping support for Python 2. See https://python3statement.org/

Finally, since our last update talk in June 2018, Biopython has had 32 named contributors including 14 newcomers. This reflects our policy of trying to encourage even small contributions. This brings our total named contributor count to 248 since the project began, and looks likely to break 250 by our 20th Birthday in August 2019.

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]Information and Computational Sciences, James Hutton Institute, Invergowrie, Dundee, UK
[b]See contributor listing on GitHub.