Biopython Project Update 2018
Standing on each other’s shoulders

Peter Cock (@pjacock on Twitter),
The Biopython Contributors (@biopython on Twitter)
What is Biopython?

- Collection of modules for biological computation in Python
  - Sequence handling and motifs, parsers, database queries, protein structures, phylogenetics, tool wrappers and more.
- Started in 1999, first release in 2000
- Open source and freely available (Biopython license)
- https://biopython.org and @Biopython on Twitter
43 named contributors in last year, 18 newcomers with star!

- Adhemar Zerlotini
- Ariel Aptekmann
- Benjamin Vaisvil*
- Blaise Li
- Chad Parmet
- Chris Rands
- Christian Brueffer
- Connor T. Skennerton
- Erik Cederstrand*
- Fei Qi*
- Francesco Gastaldello
- James Jeffryes*
- Jerven Bolleman*
- Joe Greener*
- Joerg Schaarschmidt*
- Joo Rodrigues
- Jeroen Van Goey
- Jun Aruga*
- Kai Blin
- Kozo Nishida
- Lewis A. Marshall*
- Markus Piotrowski
- Michiel de Hoon
- Nicolas Fontrodonia*
- Pamela Russell*
- Peter Cock
- Philip Bergstrom*
- rht*
- Saket Choudhary
- Shuichiro Makigaki*
- Shyam Saladi*
- Siong Kong
- Spencer Bliven
- Stefans Mezulis
- Steve Bond
- Yasar L. Ahmed*
- Zachary Sailer*
- Zaid Ur-Rehman*
- Wibowo 'Bow' Arindrarto
Biopython 1.71 (April 2018)

- Python 3-compatible rich comparison for all Bio.PDB objects
- New options and fixes for parsing and writing files, including PIR format for sequences, mmCIF files, and Gene files.
- New codon tables 27-31 from NCBI (NCBI genetic code table version 4.2) were added.
  - Now supports ambiguous codons for either stop or amino acid
Biopython 1.72 (June 2018)

- NumPy no longer needed at compile time
- Bio.Entrez now supports NCBI API key
- New alternative pairwise alignment module
- Bio.SearchIO is no longer considered experimental
Python Versions

- Currently support Python 2.7, 3.4, 3.5 and 3.6
- Clear end of life for Python 2 support in 2020, pledged on http://python3statement.org/
- Also support and test on PyPy
- Support for Jython now deprecated
- Next release will officially support Python 3.7
PpPI, pip and wheels

- Recommend “pip install biopython“
- This fetches from Python Package Index (PyPI)
- Will use pre-compiled wheel files if available
- We build the wheels on AppVeyor & TravisCI
  - Following NumPy community in using the multibuild system, developed by Matthew Brett and the MacPython project. https://github.com/matthew-brett/multibuild
  - https://github.com/biopython/biopython-wheels
Biopython’s Open Source License

- The Open Source Initiative https://opensource.org/ maintains a list of approved open source licenses.
- This is the *de facto* gold standard.
- Biopython’s license is not (quite) on that list.
Permission to use, copy, modify and distribute this software and its documentation for any purpose and without fee is hereby granted, provided that the above copyright notice appear in all copies[,] [and] that both [that] [the] copyright notice and this permission notice appear in supporting documentation[, and that the name [of] <copyright holder> [or <related entities>] not be used in advertising or publicity pertaining to distribution of the software without specific, written prior permission]. [<copyright holder> makes no representations about the suitability of this software for any purpose. It is provided “as is” without express or implied warranty.]

https://opensource.org/licenses/HPND
Biopython License Agreement

*Italics highlighting ony substantial difference to the HPND:*

Permission to use, copy, modify, and distribute this software and its documentation *with or without modifications* and for any purpose and without fee is hereby granted, provided that any copyright notices appear in all copies and that both those copyright notices and this permission notice appear in supporting documentation, and that the names of the contributors or copyright holders not be used in advertising or publicity pertaining to distribution of the software without specific prior permission.
Biopython’s Open Source License

• We could ask the Open Source Initiative (OSI) to approve our license (as an HPND variant?)
• We’ve agreed to gradually dual-license the code under our old license and the 3-clause BSD license
• Requires checking each file to confirm all contributors agree to dual licensing
• As of Biopython 1.72, about 20% of code files done
• Only one minor contributor has refused to date
On going work

- Moving from epydoc to Sphinx for API docs (and from \LaTeX{} to Sphinx for Tutorial?)
- Improving compliance with PEP8 and PEP257 style guidelines
- Improving code test coverage
  https://codecov.io/github/biopython/biopython/
- Planning what to do with legacy Alphabet objects
- Simplifying our release process
- Other new contributor driven efforts
Changes to hep Community Building

- GitHub Issue templates (to help with bug reporting)
- Pull request templates (to help with expectations)
- CODEOWNERS file to help assign code reviews
- Using *Easy Fix* tag on some issues, intended for new contributors
- Already have a CONTRIBUTING text file
- Do we need a CODE_OF_CONDUCT too?
- **What else should we be doing?**
Acknowledgements

Thank you to:

• All our contributors to date
• Google Summer of Code (supporting past students)
• My travel costs paid for by:
  • James Hutton Institute
  • Open Bioinformatics Foundation
    (as a BOSC committee member)