



## Biopython Project Update 2017

**Sourav Singh**, Christian Brueffer, Peter Cock,  
and the Biopython Contributors

Savitribai Phule Pune University  
India

Bioinformatics Open Source Conference 2017, Prague, CZ

# 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)

## In a Nutshell, Biopython...

... has had 11,237 commits made by 192 contributors representing 291,535 lines of code

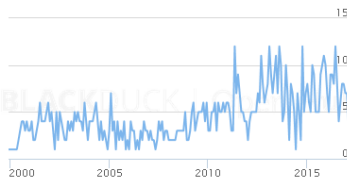
... is mostly written in Python with an average number of source code comments

... has a well established, mature codebase maintained by a very large development team with increasing Y-O-Y commits

... took an estimated 75 years of effort (COCOMO model) starting with its first commit in December, 1999 ending with its most recent commit 3 months ago

## Community

Contributors per Month



Source: <https://www.openhub.net/p/biopython>

# 80 named contributors this year, 51 newcomers with star!

- |                     |                           |                       |                             |
|---------------------|---------------------------|-----------------------|-----------------------------|
| ▶ Aaron Kitzmiller* | ▶ Chris Warth             | ▶ João Rodrigues      | ▶ Rodrigo Dorantes-Gilardi* |
| ▶ Aaron Rosenfeld   | ▶ Connor T. Skennerton    | ▶ Kai Blin            | ▶ Sacha Laurent*            |
| ▶ Adam Kurkiewicz*  | ▶ Emmanuel Noutahi        | ▶ Kristian Davidsen*  | ▶ Sebastian Bassi           |
| ▶ Adam Novak*       | ▶ Eric Rasche             | ▶ Kurt Graff*         | ▶ Sourav Singh*             |
| ▶ Adil Iqbal*       | ▶ Eric Talevich           | ▶ Lenna Peterson      | ▶ Spencer Bliven*           |
| ▶ Adrian Altenhoff* | ▶ Foen Peng*              | ▶ Leonhard Heizinger* | ▶ Stefans Mezulis           |
| ▶ Allis Tauri*      | ▶ Francesco Gastaldello*  | ▶ Marcin Magnus*      | ▶ Steve Bond                |
| ▶ Andrew Dalke      | ▶ Francisco Pina-Martins* | ▶ Markus Piotrowski   | ▶ Steve Marshall*           |
| ▶ Andrew Guy*       | ▶ Francisco               | ▶ Mateusz Korycinski* | ▶ Ted Cybulski*             |
| ▶ Andrew Szczesnak* | ▶ François Coste*         | ▶ Maximilian Greil*   | ▶ Tiago Antao               |
| ▶ Anthony Bradley*  | ▶ Frederic Sapet*         | ▶ Michał J. Gajda*    | ▶ Travis Wrightsman         |
| ▶ Ariel Aptekmann*  | ▶ Hector Martinez*        | ▶ Michiel de Hoon     | ▶ Uri Laserson              |
| ▶ Ben Fulton        | ▶ Iddo Friedberg          | ▶ Milind Luthra*      | ▶ Uwe Schmitt*              |
| ▶ Bernhard Thiel*   | ▶ Jacek Śmietański        | ▶ morrme*             | ▶ Veronika Berman*          |
| ▶ Bertrand Caron*   | ▶ Jack Twilley*           | ▶ Noam Kremen*        | ▶ Vincent Davis             |
| ▶ Bertrand Néron    | ▶ Jared Andrews*          | ▶ Olivier Morelle*    | ▶ Wibowo Arindrarto         |
| ▶ Blaise Li*        | ▶ Jeroen Van Goey*        | ▶ Oscar G. Garcia*    | ▶ Xiaoyu Zhuo*              |
| ▶ Brandon Carter*   | ▶ Jimmy O'Donnell*        | ▶ Owen Solberg        | ▶ Zheng Ruan                |
| ▶ Brandon Invergo   | ▶ John Kern*              | ▶ Patrick Kunzmann*   |                             |
| ▶ Carlos Pena       | ▶ Jordan Willis*          | ▶ Peter Cock          |                             |
| ▶ Carlos Ríos       | ▶ Joshua Meyers*          | ▶ Rasmus Fonseca*     |                             |
| ▶ Chris Rands       |                           | ▶ Richard Neher*      |                             |

Biopython 1.68 (released 2016-08-25)

## New Features in Biopython 1.68

- ▶ Final Release supporting Python 2.6
- ▶ `Bio.PDB` can now parse the new MMTF format.
- ▶ `Bio.pairwise2` rewritten to address speed and problems with local alignments.
- ▶ `SeqGui` and `xbbtools` rewritten using `tkinter`.
- ▶ Biopython now uses HTTPS instead of HTTP to connect to NCBI Entrez and QBLAST API.

Biopython 1.69 (released 2017-04-06)

## Start of re-licensing plan

- ▶ Biopython 1.69 marked start of our re-licensing plan.
- ▶ transitioning away from liberal *Biopython License Agreement*, to the widely used *3-Clause BSD License*.
- ▶ Code base authorship being reviewed file-by-file,
- ▶ will gradually dual license the entire project.

## New features in Biopython 1.69

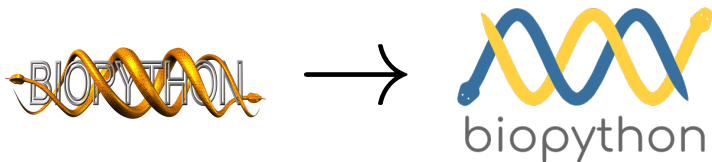
- ▶ Improved PyPy support, all NumPy and most C code works.
- ▶ Cellosaurus cell line database, cell ontologies and catalogues accessible through the `Bio.ExPASy` module.
- ▶ MAF files supported in `Bio.AlignIO`, including indexing.
- ▶ `Bio.Restriction` updated to REBASE Feb 2017 enzyme list
- ▶ `Bio.PDB` can download PDBx/mmCif (new default), PDB (old default), PDBML/XML, and MMTF protein structures.
- ▶ `Bio.Affy` supports version 4 of Affymetrix CEL format.



Biopython 1.70 (released 2017-07-10)

# New logo for Biopython

- ▶ Version 1.70 marks the transition towards a new logo for Biopython.



- ▶ The new logo was created by Patrick Kunzmann in 2017 and had support from the Python Software Foundation and the Biopython developers community.

# New features in Biopython 1.70

- ▶ Module `Bio.AlignIO`
  - ▶ Support for XMFA file format using `Bio.AlignIO.MauveIO`
- ▶ Module `Bio.SearchIO`
  - ▶ Support for new arguments to read and write blast-xml files.
- ▶ Deprecate `Bio.GA` and `Bio.NeuralNetwork` modules.
- ▶ Deprecate support for Jython.
- ▶ Dropped support for Python 3.3

## Currently supported Python versions

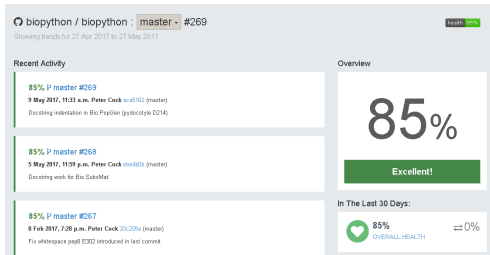
- ▶ Python 2.7
- ▶ Python 3.4
- ▶ Python 3.5
- ▶ Python 3.6
- ▶ PyPy 5.7
- ▶ PyPy3 5.8 beta
- ▶ Jython 2.7 (support is deprecated)

## On going work

- ▶ Miscellaneous bug fixes
- ▶ Improving built-in docstrings (Python API documentation)
- ▶ Test suite enhancements and coverage improvements
- ▶ Better [PEP8](#) and [PEP257](#) coding style adherence
- ▶ Automating build process for wheel packages  
<http://github.com/biopython/biopython-wheels>

## Continuous Integration

- ▶ TravisCI
  - ▶ Linux testing with coverage
  - ▶ Style checks with flake8
  - ▶ Experimenting with Build Stages
- ▶ Appveyor (Windows testing with coverage)
- ▶ Codecov.io (tracking unit test coverage)
- ▶ Quantified Code (code metrics etc, sadly shutting down)
- ▶ Landscape.io (“health score”)



# Conclusion

- ▶ Lots of new contributors
- ▶ Turning new contributors into recurring ones
- ▶ Lots of new stuff
- ▶ Lots more automation with Continuous Integration services
- ▶ Biopython 1.71 in the near future

# Acknowledgements

- ▶ Peter Cock
- ▶ Biopython Community



O|B|F



# Resources!

## Website:

- ▶ <http://biopython.org>

## Repositories:

- ▶ Main: <http://github.com/biopython/biopython>
- ▶ Website:  
<https://github.com/biopython/biopython.github.io>

## Mailing lists:

- ▶ General list: [biopython@biopython.org](mailto:biopython@biopython.org)
- ▶ Developers list: [biopython-dev@biopython.org](mailto:biopython-dev@biopython.org)

## Biostars:

- ▶ <https://www.biostars.org/t/biopython/> (“biopython” category)