

What is Biopython?

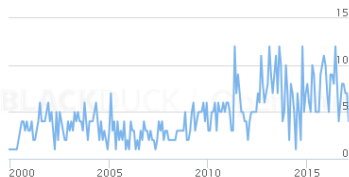
- ▶ Collection of modules for biological computation in Python
 - ▶ Sequence handling and motifs, parsers, database queries, protein structures, phylogenetics, tool wrappers and more.
- ▶ Started in 1999, first release in 2000
- ▶ Open source and freely available (Biopython license)

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>

Lots of new contributors!

- ▶ Adam Kurkiewicz
- ▶ Adam Novak
- ▶ Adrian Altenhoff
- ▶ Allis Tauri
- ▶ Andrew Guy
- ▶ Andrew Sczesnak
- ▶ Blaise Li
- ▶ Brandon Carter
- ▶ Foen Peng
- ▶ Francesco Gastaldello
- ▶ Francisco Pina-Martins
- ▶ Hector Martinez
- ▶ Jack Twilley
- ▶ Jeroen Van Goey
- ▶ Joshua Meyers
- ▶ Kurt Graff
- ▶ Leonhard Heizinger
- ▶ Marcin Magnus
- ▶ Maximilian Greil
- ▶ Michał J. Gajda
- ▶ Milind Luthra
- ▶ Oscar G. Garcia
- ▶ Richard Neher
- ▶ Sourav Singh
- ▶ Spencer Bliven
- ▶ Steve Marshall
- ▶ Veronika Berman
- ▶ Aaron Kitzmiller
- ▶ Bertrand Caron
- ▶ François Coste
- ▶ Frederic Sapet
- ▶ Jimmy O'Donnell
- ▶ John Kern
- ▶ Mateusz Korycinski
- ▶ morrme
- ▶ Noam Kremen
- ▶ Rasmus Fonseca
- ▶ Rodrigo Dorantes-Gilardi
- ▶ Sacha Laurent

Biopython 1.69 (released 2015-10-21)

Start of re-licensing plan

- ▶ 1.69 marks the start of start of our re-licensing plan, to transition away from our liberal but unique *Biopython License Agreement* to the widely used *3-Clause BSD License*.
- ▶ The code base-authorship is being reviewed file-by-file, in order to gradually dual license the entire project.

New parser for ExPASy cell line database

- ▶ Bio.ExPASy:
 - ▶ Cellosaurus cell line database, cell ontologies and cell line catalogues.
 - ▶ Now accessible through the Bio.ExPASy module.

```
from __future__ import print_function
from Bio import ExPASy, SwissProt

# 'PDOC00965', 'PDOC00022', 'PDOC50853', Chalcone
#   synthases from Orchid
ids = ['PDOC00965', 'PDOC00022', 'PDOC50853']

for id in ids:
    handle = ExPASy.get_sprot_raw(id)
    record = SwissProt.read(handle)
    print("description: %s" % record.description) # Print
        Description
    for ref in record.references:
        print("authors: %s" % ref.authors)
        print("title: %s" % ref.title)
```

Bio.AlignIO and Bio.SearchIO.AbiIO Updates

- ▶ Bio.AlignIO: Classes that deal with multiple sequence alignment files.
- ▶ Support for UCSC MAF format.
 - ▶ MAF files are now supported using the Bio.AlignIO.MafIO module
 - ▶ Also offers indexed access for large files using SQLite3.
- ▶ Extended to support the parsing of FSA files.

Other new features to Biopython

- ▶ Update to `Bio.Restriction` to include the REBASE February 2017 restriction enzyme list.
- ▶ `Bio.PDB.PDBList` now can download PDBx/mmCif (new default), PDB (old default), PDBML/XML and mmtf format protein structures.
- ▶ `Bio.Affy` module supports version 4 of the Affymetrix CEL format
- ▶ `Bio.Seq` offers complement function for consistency
- ▶ `Bio.SeqIO` parsers updated to cope with more format variations.

Miscellaneous

- ▶ Miscellaneous bug fixes
- ▶ Test suite enhancements
- ▶ Better PEP8 and PEP257 coding style adherence
- ▶ Enhanced PyPy support by taking advantage of NumPy and compiling most of the Biopython C code modules

Biopython 1.69 Contributors

- ▶ Aaron Rosenfeld
- ▶ Adam Kurkiewicz (*)
- ▶ Adam Novak (*)
- ▶ Adrian Altenhoff (*)
- ▶ Allis Tauri (*)
- ▶ Andrew Dalke
- ▶ Andrew Guy (*)
- ▶ Andrew Szczesnak (*)
- ▶ Ben Fulton
- ▶ Bernhard Thiel (*)
- ▶ Bertrand Néron
- ▶ Blaise Li (*)
- ▶ Brandon Carter (*)
- ▶ Brandon Invergo
- ▶ Carlos Pena
- ▶ Carlos Ríos
- ▶ Chris Warth
- ▶ Emmanuel Noutahi
- ▶ Foen Peng (*)
- ▶ Francesco Gastaldello (*)
- ▶ Francisco Pina-Martins (*)
- ▶ Hector Martinez (*)
- ▶ Jacek Śmietański
- ▶ Jack Twilley (*)
- ▶ Jeroen Van Goey (*)
- ▶ Joshua Meyers (*)
- ▶ Kurt Graff (*)
- ▶ Lenna Peterson
- ▶ Leonhard Heizinger (*)
- ▶ Marcin Magnus (*)
- ▶ Markus Piotrowski
- ▶ Maximilian Greil (*)
- ▶ Michał J. Gajda (*)
- ▶ Michiel de Hoon
- ▶ Milind Luthra (*)
- ▶ Oscar G. Garcia (*)
- ▶ Owen Solberg
- ▶ Peter Cock
- ▶ Richard Neher (*)
- ▶ Sebastian Bassi
- ▶ Sourav Singh (*)
- ▶ Spencer Bliven (*)
- ▶ Stefans Mezulis
- ▶ Steve Bond
- ▶ Steve Marshall (*)
- ▶ Uri Laserson
- ▶ Veronika Berman (*)
- ▶ Vincent Davis
- ▶ Wibowo 'Bow' Arindrarto

What's cooking for Biopython 1.70?

Biopython 1.70-dev

- ▶ Module `Bio.AlignIO`
 - ▶ Support for XMFA file format.
 - ▶ using `Bio.AlignIO.MauveIO`

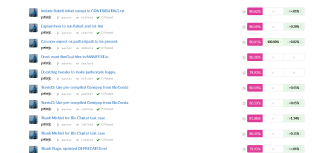
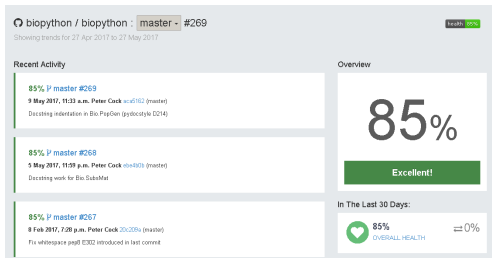
- ▶ Module `Bio.SearchIO`
 - ▶ Support for new arguments to read and write blast-xml files.

Supported Python Versions

- ▶ Python 2.7
- ▶ Python 3.3
- ▶ Python 3.4
- ▶ Python 3.5
- ▶ PyPy 5.0
- ▶ PyPy3 2.4
- ▶ Jython 2.7

Continuous Integration

- ▶ TravisCI
- ▶ Experimentation with Build Stages for CI testing.
 - ▶ Codecov.io (test coverage)
 - ▶ Quantified Code (metrics and automatic pull requests)
 - ▶ Landscape.io (“health score”)
- ▶ Currently enabled by default: Codecov.io, Quantified Code



Conclusion

- ▶ Lots of new contributors
- ▶ Lots of new stuff
- ▶ Turn new contributors into recurring ones
- ▶ Biopython 1.70 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
- ▶ Developers list: biopython-dev@biopython.org

Biostars:

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