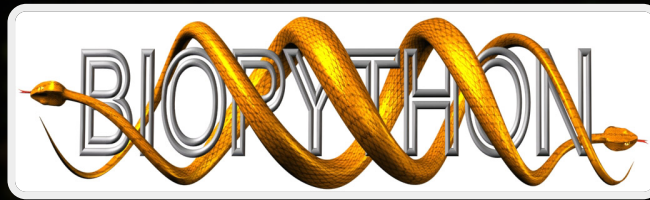


Biopython Project Update

Bioinformatics Open Source Conference
Boston, July 10th 2014



Wibowo Arindrarto (Bow), Peter Cock, Eric Talevich,
and the Biopython contributors

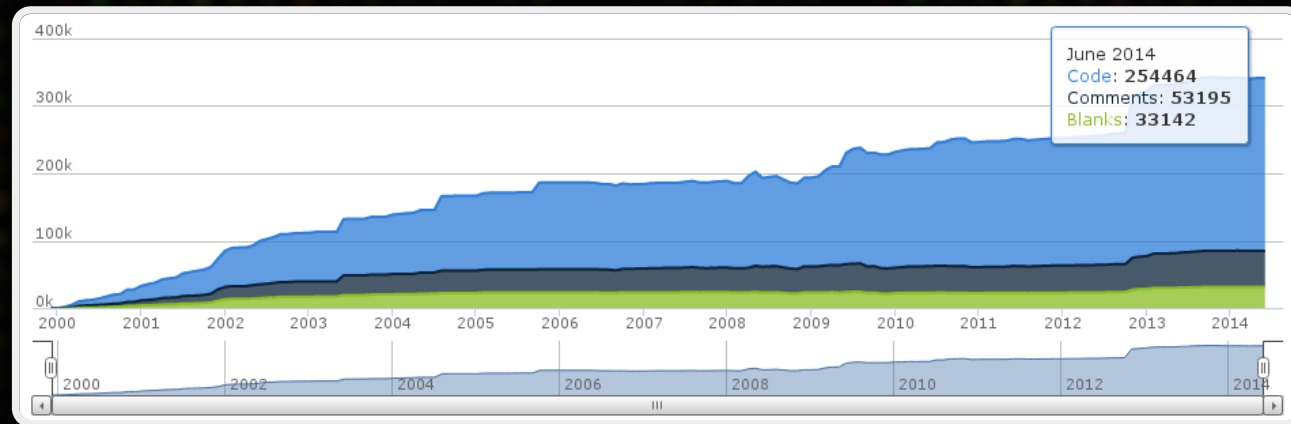
@_bow_ · #BOSC2014 · #ISMB2014



The Introduction

- Collection of modules for dealing with Biological data in Python
- Open source + freely available
- Multiple OSes + multiple Python implementations & versions
- First commit in 1999

show_stats(Bio)



via Ohloh (June 2014)

- >9,000 commits by >100 contributors
- Since BOSC 2013: >800 commits & 17 new contributors

New contributors since BOSC 2013

- Bertrand Néron
- Chris Mitchell
- Chunlei Wu
- Edward Liaw
- Gokcen Eraslan
- Josha Inglis
- Konstantin Tretyakov
- Manlio Calvi
- Markus Piotrowski
- Matsuyuki Shirota
- Melissa Gymrek
- Nigel Delaney
- Sergei Lebedev
- Vincent Davis
- Wayne Decatur
- Yanbo Ye
- Zheng Ruan

Last Time in BOSCO

- Peter Cock presented in Berlin
- Biopython 1.62 beta released (1.62 released shortly after)
- Two Google Summer of Code (GSoC) students (with NESCent)

Bio.CodonAlignment

Biopython 1.64

- GSoC 2013 project, now an experimental module
- Codon alignment analysis: objects & analysis methods
- *Mentors*: Eric Talevich & Peter Cock

Bio.Phylo enhancements

Biopython 1.64

- GSoC 2013 project
- Improved phylogenetic analysis with Bio.Phylo: tree consensus, tree comparisons, etc.
- *Mentors*: Mark Holder, Jeet Sukumaran, & Eric Talevich

Other additions

Biopython 1.62-1.65+

- Bio.Applications: `fastsimcoal`, `bwa`, and `samtools` support
- Bio.Entrez DTD caching
- Bio.Restriction enzyme list
- Bio.SearchIO, Bio.Phylo, and Bio.motifs format support
- More from yesterday's `#codefest`: Brad's update tomorrow

Streamlined Codebase

Biopython 1.63

- Single codebase supporting Python 2, Python 3, PyPy, and Jython 2.7
- Previously: 2to3.py
- Had to drop Python 2.5 (consequently Jython 2.5) support
- Documentation, tutorial, cookbook updated accordingly

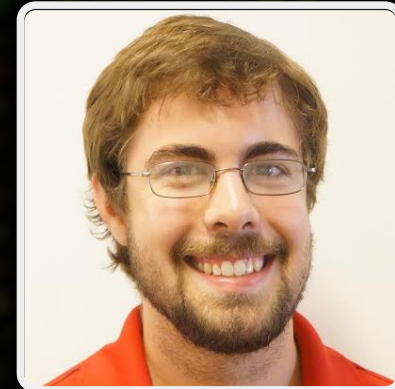
Wider Availability

- In the Galaxy ToolShed as a tool dependency
- Version 1.61 onwards

Dev repo: http://github.com/biopython/galaxy_packages

GSoC 2014

- Evan Parker, Chemistry PhD candidate
- Addition of lazy-parsing support for `Bio.SeqIO`
- Faster parsing for simple formats (e.g. FASTA) and complex, annotated formats (e.g. GenBank)
- *Mentors*: Peter Cock & Wibowo Arindrarto
- Blog: <http://blog.evanaparker.com>
- Dev repo: <http://github.com/eparker05/biopython>



Acknowledgement

The Community

contributors, users, testers, bug reporters, everybody involved

Infrastructure & Other Support



I Thank ...



Biopython lives in our:

- In-house scripts and pipelines
- Released Python packages (available in the PyPI):
 - **kMer**: k-mer analysis toolkit
 - **tssv**: short structural variation analysis
Anvar et al. (2014) DOI:10.1093/bioinformatics/btu068
- Mutalyzer web service: <http://mutalyzer.nl>
an HGVS variant nomenclature checker

Thank you!

Check us out: <http://biopython.org>

Code repository:

<http://github.com/biopython/biopython>

Mailing list:

- General list: biopython@biopython.org
- Developers list: biopython-dev@biopython.org

This presentation: <http://speakerdeck.com/bow/2014-bosc-biopython>