

Bio.Phylo: A unified phylogenetics toolkit for Biopython

Eric W Talevich*¹ and Brad A Chapman²

¹Institute of Bioinformatics, University of Georgia, 120 Green Street, Athens, GA 30602

²Massachusetts General Hospital, 55 Fruit Street, Boston, MA 02114

*Presenting author e-mail: etal@uga.edu

Project web site: <http://biopython.org>

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

License: Biopython Public License

Background: The number and range of tools developed for phylogenetic analysis has expanded dramatically, creating new opportunities to combine results with other sources of information to obtain an enhanced evolutionary perspective. This leads to the additional challenge of integrating data found in a wide variety of formats. PhyloXML, an XML-based file format for richly annotated phylogenetic data, is one effort to address this problem. However, for researchers to benefit from standardization, there must also be a complementary software ecosystem that can read, manipulate and transform this information into the various forms required to build computational pipelines.

Results: We developed a Biopython software library capable of parsing common file formats for phylogenetic trees, performing basic transformations and manipulations, attaching rich annotations, and visualizing trees. Parsing and serialization code is separated from the internal tree object representation; this allows us to support a set of common operations on trees independent of the source format, as well as convert between formats. The complete phyloXML specification is implemented, providing full interoperability with popular tools such as Archaeopteryx. Nexus and Newick support was obtained through a refactoring of the Bio.Nexus module by Cymon J. Cox and Frank Kauff. Several mechanisms for displaying trees were also implemented, taking advantage of existing libraries for visualization (matplotlib) and graph manipulation (NetworkX and Graphviz).

Conclusions: Bio.Phylo meets a growing need in bioinformatics for working with heterogeneous types of evolutionary data. By supporting interoperability with multiple file formats and leveraging existing Biopython features, this library simplifies the construction of phylogenetic workflows. We also acknowledge the role of Google Summer of Code and the NESCent Phyloinformatics program in sponsoring and initiating this project.

Availability: Bio.Phylo is included with Biopython version 1.54.
