phyphy: Python package for facilitating the execution and parsing of HyPhy standard analyses

Stephanie J. Spielman[1]

1 Institute for Genomics and Evolutionary Medicine, Temple Univeristy, Philadelphia, PA 19122

Summary

HyPhy (Kosakovsky Pond, Frost, and Muse 2005) is a powerful bioinformatics software platform for evolutionary comparative sequence analysis and testing hypotheses of natural selection from sequence data. Combined with its accompanying webserver Datamonkey (**???**), HyPhy has garnered over 3500 citations since its introduction in 2005 and has greatly accelerated the pace of biomedical and epidemiological research. I introduce phyphy (**P**ython **HyPhy**), a Python package aimed to i) faciliate the execution of standard HyPhy analyses, and ii) extract analysis information from the JSON-formatted HyPhy output into user-friendly formats. phyphy will greatly improve the batch-users' experience by allowing users to bypass the interactive HyPhy command-line prompt and execute hundreds or thousands of analyses directly from a Python script. In addition, phyphy makes it simple to obtain key information from an executed HyPhy analysis, including fitted model parameters, annotated phylogenies with analysis output formatted for downstream visualization, and CSV files containing the most relevant output for a given method. phyphy is compatible with Hyphy version >=2.3.7 and is freely available under a FreeBSD license from https://github.com/sjspielman/phyphy.

References

Kosakovsky Pond, S. L., S. D. W. Frost, and S. V. Muse. 2005. "HyPhy: Hypothesis Testing Using Phylogenetics." *Bioinformatics* 21 (5): 676–79.