# PhylogeneticGraph
# Program Documentation
# Version 0.1

Ward C. Wheeler
Division of Invertebrate Zoology,
American Museum of Natural History,
200 Central Park West, New York, NY, 10024, USA;
wheeler@amnh.org

August 19, 2021

# Contents

# 1 Introduction

This document contains descriptions of algorithms, procedures, data structures and other aspects of the source code documentation for the program PhylogeneticGraph (`PhyG`).

`PhyG` is a sucessor program to `POY` (Gladstein and Wheeler, 1997; Wheeler et al., 2005; Varón et al., 2008, 2010; Wheeler et al., 2013, 2015) `https://github.com/wardwheeler/POY5`, although a "complete" Haskell rewrite, optimized C (and even some assembler) was ported over from POY for pairwise alignment of small alphabet (¡8) sequences. These functions are access via the Haskell FFI.

# 2 Overview of Code base

Source code structure.

# 3 Command Parsing

# 4 Post-Order Graph Traversal

## 4.1 Trees

A decorated Graph (tree) is created for each character for each block for the graph. For exact characters, where no addition traversals are required, the specified or default outgroup sets the direction of the graph. For non-exact (e.g. sequence) characters the best traversal rooting is stored for each character in each block although the cost of the graph is recalculated based on the best traversal (over all edges in the graph), the preliminary (post-order) states are not propagated back to the decorated graph (third field of phylogenetic graph). After the pre-order pass, the final states are propagated back. Vertices are not renumbered during the rerooting process, so indices remain unchanged.

# 5 Pre-Order Graph Traversal

## 5.1 Trees

Final state assignments of root vertices are set to the preliminary, post-order state. Final states are propagated back to the decorated graph (third field of phylogenetic graph). Vertices are not renumbered during the rerooting process, so indices remain unchanged.

## 5.2 Execution in Parallel

By default the program will execute using a single process core. By specifying the options '+RTS -NX -RTS' where 'X' is the number of processors offered to the program. These are specified after the program as in (for 4 parallel threads):

PhyGraph +RTS -N4 -RTS other options...

Parallel code options are set using a parmap-type strategy throughout the code. The basic definitions of this functionality are found in ParallelUtilities.hs

## Acknowledgments

The author would like to thank DARPA SIMPLEX N66001-15-C-4039, the Robert J. Kleberg Jr. and Helen C. Kleberg foundation grant "Mechanistic Analyses of Pancreatic Cancer Evolution", and the American Museum of Natural History for financial support.

# 6 Bibliography


Cardona, G., Russelló, F., and Valiente, G. 2008. Extended newick: it is time for a standard representation of phylogenetic networks. *BMC Bioinformatics* 9. doi: 10.1186/1471-2105-9-532.

Gladstein, D. S. and Wheeler, W. C. 1997. POY version 2.0. program and documentation available at http://research.amnh.org/scicomp/projects/poy.php. American Museum of Natural History, New York.

Goloboff, P., Farris, J. S., and Nixon, K. 2008. Tnt, a free program for phylogenetic analysis. *Cladistics* 24:774–786.

Pearson, W. R. and Lipman, D. J. 1988. Improved tools for biological sequence comparison. *PNAS* 85:2444–2448.

Varón, A., Vinh, L. S., Bomash, I., and Wheeler, W. C. 2008. Poy 4.0. American Museum of Natural History. http://research.amnh.org/scicomp/projects/poy.php.

Varón, A., Vinh, L. S., and Wheeler, W. C. 2010. POY version 4: Phylogenetic analysis using dynamic homologies. *Cladistics* 26:72–85.

Wheeler, W. C. 2021. Phylogenetic Supergraphs. Cladistics, in press.

Wheeler, W. C., Gladstein, D. S., and De Laet, J. 1996-2005. POY version 3.0. program and documentation available at http://research.amnh.org/scicomp/projects/poy.php (current version 3.0.11). documentation by D. Janies and W. C. Wheeler. commandline documentation by J. De Laet and W. C. Wheeler. American Museum of Natural History, New York.

Wheeler, W. C., Lucaroni, N., Hong, L., Crowley, L. M., and Varón, A. 2013. POY version 5.0. American Museum of Natural History. http://research.amnh.org/scicomp/projects/poy.php.

Wheeler, W. C., Lucaroni, N., Hong, L., Crowley, L. M., and Varón, A. 2015. POY version 5: Phylogenetic analysis using dynamic homologies under multiple optimality criteria. *Cladistics* 31. 189-196.

Wheeler, W. C. and Washburn, A. J. 2019. Fastc: a file format for multi-character sequence data. *Cladistics* 35:573–575.