Cladistics 28 (2012) 188–189

Cladistics

10.1111/j.1096-0031.2011.00369.x

Trivial minimization of extra-steps under dynamic homology

Ward C. Wheeler*

Division of Invertebrate Zoology, American Museum of Natural History, Central Park West at 79th Street, New York, NY 10024-5192, USA

Accepted 30 June 2011

Abstract

Farris (1983) stated that the rationale of the parsimony criterion was to minimize extra (i.e. non-minimal) steps. For traditional characters, this is equivalent to minimizing total steps (i.e. length). Under dynamic homology (sensu Wheeler, 2001), this identity is broken. Here, it is shown that extra steps (but not total) can be minimized trivially (to zero) for all data sets on all trees when insertion-deletion events are considered.

In his discussion of the motivation for the parsimony criterion in phylogenetic analysis, Farris (1983) emphasized the logical importance of minimizing ad hoc hypotheses (i.e. homoplasy). The total length of a tree (its parsimony score or length) would be the sum of those transformations that change minimally and those that do not, the extra steps. Given a traditional character matrix or prealigned molecular data set, whether one minimizes total or extra steps makes no difference. They are simultaneously optimized. When insertion-deletion events are directly considered in a dynamic homology (Wheeler, 2001) framework, this co-optimization need not occur.

The direct optimization algorithm (Wheeler, 1996) embodied this idea in its procedural minimization of total length or cost. The POY program implements these ideas (Varón et al., 2010). Emphasizing this distinction, Kluge and Grant (2006) (and Grant and Kluge, 2009) based their justification of parsimony on the anti-superfluity principle, in opposition to the minimization of ad hoc hypothesis of Farris (this view was criticized by Farris, 2008).

Kluge and Grant (2006) show an example where total steps and extra steps yield conflicting results. Here, I take this further showing that in all cases, extra steps can be trivially optimized, leaving no basis for the distinction among phylogenetic hypotheses.

An example

Consider a data set with ten taxa and a single nucleotide observation for each (Fig. 1). In one scenario (top), there is an alignment with a single position, yielding a cladogram length of three (indels and substitutions equally weighted). Of these three steps, there is a single homoplasy or extra step. The second (middle scenario) has eight aligned positions, yielding a tree length of eight with no homoplastic changes. This inflation of length comes from the novel insertion of each of the ‘‘G’’ and ‘‘C’’ nucleotides and the second pair of ‘‘A’’s in taxa T8 and T9. If we are to minimize extra steps, we should choose this scenario (even with five more total steps) than the previous since it requires fewer (0 in fact) homoplastic changes.

Trivial solutions

If we take the example of Fig. 1 further, we can have each nucleotide observation derived via a unique insertion event (the third, bottom scenario). This case, like the second, has 0 extra steps, but with a total cost of ten steps. Unlike the second case, however, this scenario will have this same total length of ten with 0 homoplasies for all trees. Extra steps can be trivially minimized and offers no manner to distinguish among phylogenetic hypotheses. In the general case, there will be 0 extra steps and a total cost equal to the number of observations in all taxa, for all trees.

*Corresponding author:
E-mail address: wheeler@amnh.org

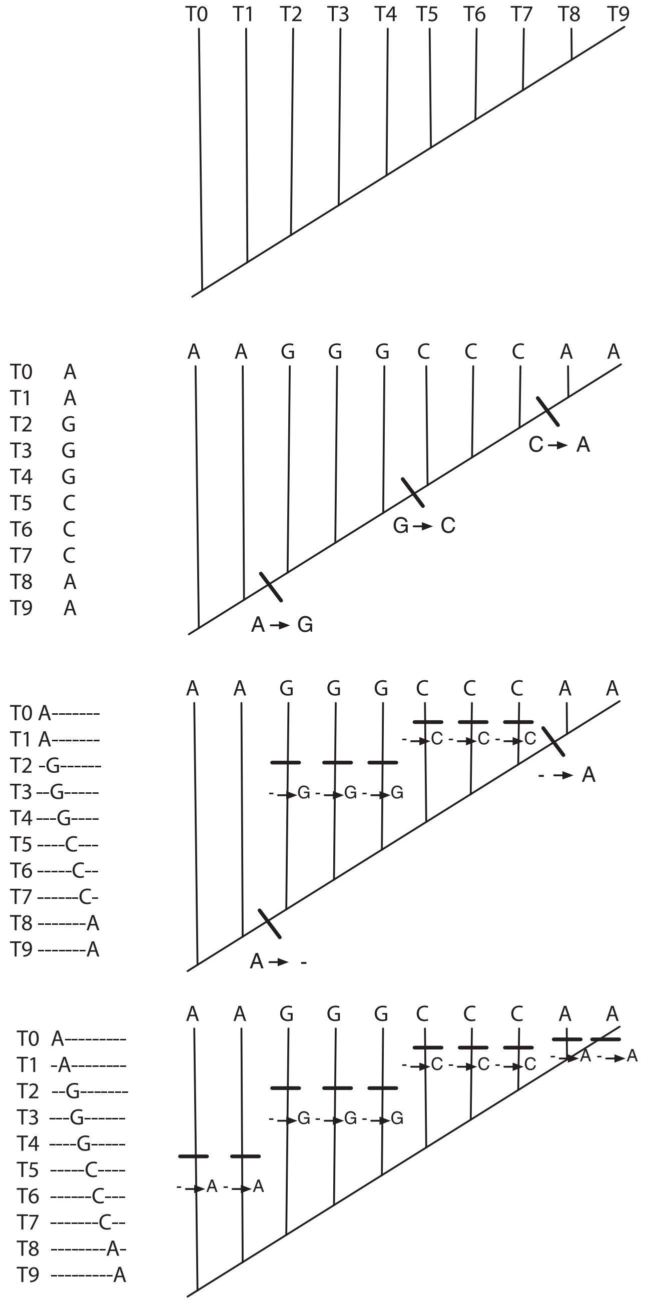

Fig. 1. A tree for ten taxa (top) as a basis for optimizing the alignments (upper, middle, and lower). The upper alignment requires three total steps one of which is "extra", the middle alignment requires eight steps in total with 0 extra steps, and the lower alignment ten steps in total, again with 0 extra steps. All transformations (indels included) costing 1, hence all steps with equal weight.

Clearly, such a situation of trivial optimization is of little use in the search for parsimonious solutions. With dynamic homology, parsimony must signify the minimization of total cost.

Acknowledgements

I would like to thank James Carpenter for discussion of these ideas and comments on this manuscript. Research was supported by the US NSF-ITR grant "Building the tree of life: A national resource for phyloinformatics and computational phylogenetics" [NSF EF 03-31495], "An Integrated approach to the origin and diversification of protostomes" [NSF DEB 05-31677], and "Assembling the tree of life: phylogeny of spiders" [NSF EAR 02-28699].

References

Farris, J.S., 1983. The logical basis of phylogenetic analysis. In: Platnick, N.I., Funk, V.A. (Eds.), Advances in Cladistics, Proceedings of the Second Meeting of the Willi Hennig Society. Columbia University Press, New York, NY, Vol. 2, pp. 7–36.

Farris, J.S., 2008. Parsimony and explanatory power. Cladistics 24, 825–847.

Grant, T., Kluge, A.G., 2009. Parsimony, explanatory power, and dynamic homology testing. System. Biodivers. 7, 357–363.

Kluge, A.G., Grant, T., 2006. From conviction to anti-superuity: old and new justifications for parsimony in phylogenetic inference. Cladistics 22, 276–288.

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

Wheeler, W.C., 1996. Optimization alignment: the end of multiple sequence alignment in phylogenetics? Cladistics 12, 1–9.

Wheeler, W.C., 2001. Homology and the optimization of DNA sequence data. Cladistics 17, S3–S11.