

COMPOSITIONAL BIAS, CHARACTER-STATE BIAS, AND CHARACTER-STATE RECONSTRUCTION USING PARSIMONY

TIMOTHY M. COLLINS,¹ PETER H. WIMBERGER,^{1,2} AND GAVIN J. P. NAYLOR¹

¹Department of Biology, University of Michigan, Ann Arbor, Michigan 48109-1048, USA

Abstract.—Compositional bias, the occurrence of the four bases in unequal proportions, is a common feature of nucleotide sequences. We analyzed patterns of character-state reconstruction using maximum parsimony in two empirical data sets exhibiting compositional bias. For each case in which the inferred reciprocal numbers of changes for a pair of bases differed substantially, the two bases also differed markedly in relative abundance and the asymmetry of reconstructed transformations favored changes from the common state to the rare state. In addition, the compositional biases of the inferred ancestral sequences were more extreme than that seen in the terminal taxa, having an excess of common states relative to the terminals. Both of these features suggested that patterns of character-state reconstruction might be systematically distorted when compositional bias results in unequal representation of character states within a character, a condition we term character-state bias. Character-state bias is essentially compositional bias within a character. Simulation studies showed that highly asymmetric patterns of character-state reconstruction can be produced in the face of an underlying symmetry of character-state transformations in the presence of compositional bias. Rates of change are also important. The asymmetry of transformations produced in the simulations matched the pattern found in empirical data sets, with transformations from the common state to the rare state being more abundant. Rare states tend to be autapomorphic, typically requiring a change to the rare state on a tree. Conversely, changes from rare states to common states are systematically lost. These results are significant for methods that rely on character-state reconstruction using maximum parsimony, for example, to develop weighting schemes for phylogenetic analysis or to study patterns of correlated character evolution. Techniques that rely on character-state reconstruction may often be compromised by the distorting influence of character-state bias. [Compositional bias; character-state bias; character-state reconstruction; parsimony.]

One of the primary advantages of maximum parsimony over other methods of phylogenetic inference is that it allows for the reconstruction of ancestral character states at internal nodes (Donoghue, 1989; Maddison and Maddison, 1992). Knowledge of ancestral character states is a prerequisite to the study of character evolution. For example, hypotheses regarding the origin of adaptations and correlated character evolution require reconstruction of ancestral character states (Coddington, 1988; Maddison, 1990). The efficacy of many methods currently used in evolutionary biology hinges on the assumption that ancestral character states can be correctly re-

constructed. In this paper, we show that maximum parsimony can result in systematically biased ancestral character-state reconstructions when the different states within a character are in unequal proportions. The purpose of this communication is to promote caution and judicious interpretation of ancestral character-state reconstructions when character states differ markedly in their frequencies for a class of characters. We present two empirical case studies using DNA sequence data in which maximum parsimony implies ancestral character states that are likely incorrect and demonstrate the effect with computer simulations.

BACKGROUND

DNA sequences lend themselves well to studies investigating evolutionary dynamics

² Present address: Department of Biology, University of Puget Sound, Tacoma, Washington 98416, USA.