LETTER TO THE EDITOR

To the Editor:

The relative merits of likelihood and parsimony usually are discussed in terms of philosophy (Siddall and Kluge, 1997) or of ability to reflect a modeled process (Huelsenbeck, 1995; Siddall, 1998). Little regard has been given to how the two methods view evidence and character independence. Siddall and Kluge (1997: 322) suggested that one cannot claim "'that nucleotide sites are treated independently and simultaneously 'correct for' multiple substitutions." Here we give sample data sets to demonstrate the very different perspectives that maximum likelihood and parsimony take on the nature of evidence and on whether or not characters are treated independently (all analyses were conducted with PAUP* ver. 4.0b2 for Macintosh OS).

For a four-taxon statement, with taxa "one," "two," "three," and "four," there are three possible hypotheses of relationships, namely ((one four) (two three)), ((one three) (two four)), and ((one two) (three four)).

Consider matrix "I":

one	AGGACTTC
two	GAAGTCCT
three	GGAATTCC
four	GGAATTCC

Parsimony analysis finds all trees equally optimal. So too does maximum likelihood using a Jukes–Cantor model.

Consider matrix "II":

one	AGGACTTC C
two	GAAGTCCT T
three	GGAATTCC C
four	GGAATTCC T

Parsimony and likelihood using a Jukes–Cantor model each prefer tree ((one three) (two four)), as is evidenced by the shared states for the ninth character.

Consider matrix "III":

one	AGGACTTC A
two	GAAGTCCT A
three	GGAATTCC G
four	GGAATTCC G

Parsimony obviously prefers tree ((one two) (three four)) in light of the evidence in the ninth character, but likelihood using a Jukes–Cantor model does not distinguish among any of the possible topologies.

Consider matrix "IV":

one	AGGACTTC A C
two	GAAGTCCT A T
three	GGAATTCC G C
four	GGAATTCC G T

Parsimony considers the ninth and tenth characters to be equally compelling evidence for ((one two) (three four)) or ((one three) (two four)). Likelihood using a Jukes–Cantor model prefers only ((one three) (two four)).

Matrix "V"

one	AGGACTTC AT CA
two	GAAGTCCT AT TG
three	GGAATTCC GC CG
four	GGAATTCC GC TA,

which has two characters with identical distributions and two with other distributions, is considered by parsimony to provide unambiguous evidence for ((one

(AP)

two) (three four)). Likelihood using a Jukes–Cantor model finds the other two trees ((one four) (two three)) and ((one three) (two four)) more compelling.

Parsimony analysis sees no evidence in matrix "VI"

one	ACGT
two	CGTA
three	GTAC
four	TACG

though likelihood (using a Kimura two-parameter model, estimating kappa from the data) is compelled to choose ((one three) (two four)).

With matrix "VII"

one	AAAAAAAA
two	GGGGGGGG
three	TTTTTTTT
four	CCCCCCCC

likelihood (using HKY85) chooses ((one two) (three four)). Matters change for matrix "VIII"

one	AAAAAAAA GGGGGGGGAAAAAAAA
two	GGGGGGGG GGGGGGGGAAAAAAAA
three	TTTTTTTT GGGGGGGGAAAAAAAA
four	CCCCCCCC GGGGGGGGAAAAAAAA

The same approach now is obliged to choose the other two trees. However, the effect goes away in matrix "IX"

one	AAAAAAAA GGGGGGGGAAAAAAAA
two	GGGGGGGG GGGGGGGGAAAAAAAA
three	TTTTTTTT GGGGGGGGAAAAAAAA
four	C-C-C-C- GGGGGGGGAAAAAAAA
five	-C-C-C-C GGGGGGGGAAAAAAAA

The point of these simple examples is only to underscore the fact that likelihood methods do not evaluate characters independently and that they see evidence where we believe most would not. The differences of opinion between likelihood and parsimony in evaluating matrices III, IV, and V stem from likelihood expectations of homogeneity, i.e., what applies to one character must be appropriate for all, and independent changes for the ninth character on the branches leading to one and two are deemed probable in light of the autapomorphies in characters 1 through 8. Similarly, in matrices VI and VII, even if none is specified, the method estimates a substitution bias and chooses a tree to minimize the number of transversions. Those same transversions in matrix VII become much more "probable" to likelihood when faced with the additional 16 constant purine characters in matrix VIII and so a different tree is chosen.

If, as it would seem to us, this problem of homogenous expectations across sites and lineages is unreasonable, a solution would be to estimate models for each site and each lineage separately. That is, unlike modeled data (e.g., Huelsenbeck, 1995) where there is a common mechanism creating the data, expect no common mechanism. This approach to rendering maximum likelihood reasonable will yield "precisely the maximum parsimony trees" (Tuffley and Steel, 1997: 599), that is, parsimony and maximum likelihood with no common mechanism will yield identical results.

REFERENCES

Huelsenbeck, J. P. (1995). Performance of phylogenetic methods in simulation. *Syst. Biol.* **44:** 17–48.

Siddall, M. E. (1998). Success of parsimony in the four-taxon case: Long-branch repulsion by likelihood in the Farris zone. *Cladistics* **14:** 209–220.

Siddall, M. E., and Kluge, A. G. (1997). Probabilism and phylogenetic inference. *Cladistics* **13:** 313–336.

Tuffley, C., and Steel, M. (1997). Links between maximum likelihood and maximum parsimony under a simple model of site substitution. *Bull. Math. Biol.* **59:** 581–607.

Mark E. Siddall
Division of Invertebrate Zoology
American Museum of Natural History
New York, New York 10024
Arnold G. Kluge
Museum of Zoology
University of Michigan
Ann Arbor, Michigan 48109