Wheeler (1993) attempts to show that procedures wherein character-transformation matrices have or permit violations of the triangle inequality are generally to be avoided, as illogical and inconsistent. “Only through [the triangle inequality’s] application can logical and consistent cost scenarios be created.” This derives from the assumption that the transformation values are (Euclidean) distances and that they therefore must follow the four conditions of three-point metric space—namely, $d_{aa} = 0; d_{ab} \geq 0; d_{ab} = d_{ba};$ and $d_{ab} \leq d_{a6} + d_{6b}$. There are other metric spaces in which the fourth condition is replaced. For example, the four-point or additive metric requires $d_{ab} + d_{6g} \leq d_{a6} + d_{6y} = d_{a6} + d_{by}$. I shall argue that, although what Wheeler says is quite helpful, it is overly general because it focuses on metricity rather than on the goal of accurate reconstruction of phylogeny and ancestral character states.

Atchley and I (Fitch and Atchley 1985a, 1985b), have endeavored to reconstruct the known history of the old strains of inbred mice. We have recently developed a new method of phylogeny inference for these data, which we call “loss parsimony” (unpublished data). All the strains are inbred to the point where all loci examined are homozygous, although different strains are homozygous for different alleles. Moreover, all alleles in the original analysis were present in wild populations. It therefore appeared that a model close to what actually happened was that a female Mus mus domesticus mouse from central England was crossed with a M. m. musculus or M. mollosinus male from eastern Europe or farther east and that the process of inbreeding caused the loss of the heterozygosity produced in the original cross (Fitch and Atchley 1985b). We now know that there was more than one male involved, but the model is realistic for some subset of the data, and that is all that needs possibly to be true to make my point.

The method requires every allele at every locus to be thought of as a separate character. Every strain has one and only one allele for each locus, so that a majority of the characters are absent for every strain. The outgroup (the presumed hybrid resulting from the hybrid cross) has all characters. By reversing the character-state assignment ($0 =$ present; and $1 =$ absent), one can convert PHYLIP’s Camin-Sokal algorithm (Camin and Sokal 1965) from one that creates a new state but never returns to an old state into a method that loses an old state as often as is necessary but never regains it. It precisely matches the case of crossing two animals (or even a small, closely related population of them) and deriving several inbred lines from their offspring and from descendants thereof.

This method violates Wheeler’s strictures because the “distance” from present to absent is very much less than the (infinite) distance from the absent to present. This is not really a violation of the triangle inequality so much as it is a violation of the second assumption of metricity, $d_{ab} = d_{ba}$ (see Wheeler’s section on symmetry and his fig. 1c, left). The method properly reflects that the underlying process of inbreeding allows only loss (if new mutants arise, they are, in practice, removed from that lineage). The phylogeny would have been different if gains had the same value as losses.

The fact that the underlying genetic process is metric (indeed, it is the even more restrictive four-point metric!) does not mean that methods must eschew what are the molecular equivalent of cladistic polarity statements. Loss parsimony is but an extreme example. A restriction-enzyme site for a 4-cutter has 81 alternative possible nucleotide substitutions available to it, as does a second potential site that is only one nucleotide
different from such a four-cutter site. In the first case, all 81 substitutions destroy the site; in the second case, only 1 of the 81 substitutions produces a new site. Thus a cladogram that had a restriction site arise once and be lost once rather than arise independently twice would seem to be preferable. Any time that the triangle inequality is violated, we should look carefully to see that we have not fallen into some subtle trap, but biology is more important than metricity.

WALTER M. FITCH
Editor in Chief

LITERATURE CITED


