TWO CONJECTURES CONCERNING MAXIMUM PARSIMONY TREES

MP1

There exists a constant $c > 0$ such that, for any fully-resolved phylogenetic tree $\mathcal{T}$, there exists a sequence of at most $\lfloor c \cdot \log(n(\mathcal{T}))\rfloor$ binary characters on X that has $\mathcal{T}$ as its unique maximum parsimony tree, where $n(\mathcal{T})$ is the number of leaves of $\mathcal{T}$.

MP2

Suppose we generate a 2-state character ('site pattern') on the leaves of a binary phylogenetic tree by the symmetric 2-state Markov model (also known as the CFN, or Neyman 2-state model). Suppose that each edge of the tree has the same substitution probability p. Let $\mu_{\mathcal{T}}(\mathcal{T}')$ denote the expected parsimony score on $\mathcal{T}'$ of a character evolved under the 2-state Markov model on $\mathcal{T}$. Then if p is sufficiently small (as a function of n, the number of leaves of $\mathcal{T}$) it is relatively easy to show that:

$$\mu_{\mathcal{T}}(\mathcal{T}') > \mu_{\mathcal{T}}(\mathcal{T}), \tag{1}$$

for every binary phylogenetic trees $\mathcal{T}'$ that has the same leaf set as $\mathcal{T}$ but which is different (in topology) from $\mathcal{T}$. However perhaps there exists a constant p_0 that is independent of n for which (1) holds for *all* binary phylogenetic trees $\mathcal{T}$ (and $\mathcal{T}' \neq \mathcal{T}$ with the same leaf set as $\mathcal{T}$) for all $p \in (0, p_0)$ [Note: This is equivalent to asserting the statistical consistency of maximum parsimony on all binary trees for this restriction on the substitution probabilities, and is a special case of a more general conjecture by Vic. Albert in his book 'Parsimony, Phylogeny and Genomics', Oxford UP, 2005 (Conjecture 1.3.1, p.8)]