Additional File 3 — A sketch of a proof for the formula of limiting UA on Hennigian comb-shaped trees

For a Hennigan comb shaped tree, let $Z$ be an internal node, under which the number of leaves is large. Let $X$ and $Y$ be the left and right children of $Z$, thus $X$ is a leaf and $Y$ is a root of a smaller comb-shaped tree. Suppose that the branch length of $ZX$ tends towards infinity, then by the Jukes-Cantor model the true state at $X$ is randomized, that is, $\Pr_X[X = i] = \frac{1}{N}$ for $i \in S$, where $t_X$ denotes the true state at $X$.

Let $i$ be the cardinality of the set of states that the Fitch algorithm reconstructs at $Y$. Then the reconstructed set at $Z$ contains either $i + 1$ or 1 state(s), with probability $\frac{N-i}{N}$ and $\frac{1}{N}$, respectively. As a result, the cardinalities of the reconstructed state set by the Fitch algorithm from leaves to the root can be formulated by a Markov process, in which the state set is $\{1, 2, \ldots, N\}$ and the transition matrix is

$$T = \begin{bmatrix}
\frac{1}{N} & \frac{N-1}{N} & 0 & \cdots & 0 \\
\frac{1}{N} & \frac{N-2}{N} & \frac{N-1}{N} & \cdots & 0 \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
\frac{1}{N} & 0 & 0 & \cdots & \frac{1}{N} \\
1 & 0 & 0 & 0 & 0
\end{bmatrix} \tag{1}$$

For every pair of states $i$ and $j$, there is a walk $i, 1, 1, \ldots, 1, 2, 3, \ldots, j - 1, j$ from $i$ to $j$ of length $N$ with non-zero probability. Thus, the transition matrix $T$ is primitive. By the Perron-Frobenius theorem [1], there exists an equilibrium state vector $\vec{v} = (v_1, v_2, \ldots, v_N)$, such that $\vec{v} = \vec{v}T$. Since the state at $X$ is randomized, the probability that the reconstructed cardinality 1 set contains the true state is $\frac{1}{N}$. Thus, the limiting UA is $\frac{1}{N} v_1$, which is $\sum_{i=1}^{N} \frac{N-i}{N} \frac{N-i-1}{(N-1)!}$.

References