

祖先形質の最節約復元順序集合について

– On MPR-posets in phylogeny –

東海大・理・情報数理 成嶋 弘 (Hiroshi Narushima)

A mathematical theory for the subject on ancestral character-state reconstructions under the maximum parsimony in phylogeny has been developing ([2]–[10]).

We use the notations in [2] and [5]. Let Ω denote the set that may be either the set $\mathbf{R}$ of real numbers or the set $\mathbf{N}$ of nonnegative integers. Note that Ω expresses the linearly ordered character-states. Let $T = (V = V_O \cup V_H, E, \sigma)$ be any undirected tree with the endnodes evaluated by a weight function $\sigma : V_O \rightarrow \Omega$, where V is the set of nodes, V_O is the set of endnodes which are nodes of degree one, V_H is the set of internal nodes, and E is the set of branches. We call this tree an *el-tree*. For an el-tree T , we define an assignment $\lambda : V \rightarrow \Omega$ such that $\lambda|_{V_O}$ (the restriction of λ to V_O) = σ , where $\lambda(u)$ is called a *state* of u under λ . This assignment is called a *reconstruction* on an el-tree T . For each branch e in E of an el-tree T with a reconstruction λ , we define the *length* $l(e)$ of branch $e = \{u, v\}$ by $|\lambda(u) - \lambda(v)|$. Then the *length* $L(T|\lambda)$ of an el-tree T under the reconstruction λ is the sum of the lengths of the branches. That is, $L(T|\lambda) = \sum_{e \in E} l(e)$. Furthermore we define the minimum length $L^*(T)$ of T by

$$L^*(T) = \min\{L(T|\lambda) \mid \lambda \text{ is a reconstruction on } T\}.$$

Note that $L^*(T)$ is well-defined. A *Most-Parsimonious Reconstruction* denoted by MPR on an el-tree T is a reconstruction λ such that $L(T|\lambda) = L^*(T)$. Generally an el-tree T has more than one MPR. The set $\{\lambda(u) \mid \lambda \text{ is an MPR on } T\}$ of states is called the *MPR-set* of a node u and written as S_u .

Let $T = (V, E)$ be a rooted (directed) tree, where V is the set of nodes and $E(\subseteq V \times V)$ is the set of branches. For each u and v in V , we write $u \rightarrow v$ or $u = p(v)$ when $(u, v) \in E$, i.e., u is a *parent* of v (or v is a *child* of u). For each u and v in V , u is called an *ancestor* of v , written $u \xrightarrow{*} v$, if there is a sequence of nodes $u = u_1, u_2, \dots, u_n = v$ in V such that $u_i \rightarrow u_{i+1}$ ($i \in [n-1]$). In a rooted tree, there is only one node without a parent, which is called the *root*, and a node without a child is called a *leaf*. For each u in V , we denote a *subtree* of T induced from a subset $\{u\} \cup \{v \in V \mid u \xrightarrow{*} v\}$ of V by $T_u = (V_u, E_u)$. Note that u is the root of T_u .

For a given el-tree $T = (V_O \cup V_H, E, \sigma)$, we define a *rooted el-tree* $T^{(r)}$ rooted at any element r in $V = V_O \cup V_H$. The rooted el-tree $T^{(r)}$ is simply written T if it is understood. In addition, if r is an endnode, i.e., $r \in V_O$ and s is its unique child, we denote the rooted tree $T^{(r)}$ by (T_s, r) to visualize the structure. In this case, the subtree T_s is called the *body* of the tree $T^{(r)}$; otherwise, i.e., if $r \in V_H$, the body of $T^{(r)}$ is $T^{(r)}$ itself.

Let $I_i = [a_i, b_i]$ ($i \in [m]$) be any family of closed intervals in Ω . Let all the endpoints a_i and b_i of I_i ($i \in [m]$) be sorted in ascending order and then be arranged as follows:

$$x_1 \leq x_2 \leq \cdots \leq x_m \leq x_{m+1} \leq \cdots \leq x_{2m}.$$

Then we call the closed interval $[x_m, x_{m+1}]$ in Ω the *median interval* of the closed intervals I_i ($i \in [m]$), which is the key concept in a series of our papers, and denote it by $\text{med}\langle I_1, I_2, \dots, I_m \rangle$ or $\text{med}\langle I_i : i \in [m] \rangle$.

For each node u in the body of a rooted el-tree T , we assign a closed interval $I(u)$ of Ω recursively as follows:

$$I(u) = \begin{cases} [\sigma(u), \sigma(u)] & \text{if } u \text{ is a leaf,} \\ \text{med}\langle I(v) : u \rightarrow v \rangle & \text{otherwise.} \end{cases}$$

We call $I(u)$ the *characteristic interval* of a node u and so I is called the *characteristic interval map* on T .

We now restate the results in the previous paper [2], which are used in this paper. Let T be a rooted el-tree (T_s, r) and I be the characteristic interval map on T . Let $\lambda_{<u>}$ denote the restriction $\lambda|_{V_u}$ of a reconstruction λ on T to a subtree T_u of T . Then a set $\mathbf{Rmp2}(r, s)$ of reconstructions on T is defined recursively as follows:

$$\lambda_{<s>} \in \mathbf{Rmp2}(r, s) \iff \begin{cases} \lambda(s) \in \text{med}\langle [\lambda(r), \lambda(r)], I(t) : s \rightarrow t \rangle, \\ \text{and } \forall t(s \rightarrow t) (\lambda_{<t>} \in \mathbf{Rmp2}(s, t)). \end{cases}$$

Note that $\lambda_{<s>}$ (with $\lambda(r) = \sigma(r)$) can be considered a reconstruction on T . The following are Theorem 1 (Theorem 3 (ii)) and Corollary 5 in [2].

Theorem A. *For any endnode r of an el-tree T , $\mathbf{Rmp2}(r, s)$ is the set of all MPRs on T*

Noting that generally a phylogenetic tree has more than one MPR, Swofford and Maddison [9] have defined more explicitly the ACCTRAN reconstruction originated with Farris [1], and the DELTRAN reconstruction, which are considered to be more meaningful and useful MPRs in phylogeny. Then Minaka [3] has introduced the usual partial ordering on the set of all possible MPRs on a phylogenetic tree, in order to investigate the relationships among the ACCTRAN, the DELTRAN, and other MPRs.

For any λ and μ in $\mathbf{Rmp}(T)$, the partial ordering $\lambda \leq \mu$ is defined by $\lambda(u) \leq \mu(u)$ for all u in V . The partially ordered set $(\mathbf{Rmp}(T), \leq)$ is called the *MPR-poset* or Minaka poset. From a lattice-theoretic point of view, we first have a question whether there exists the greatest element (or the least element) in the MPR-poset or not.

The following is Proposition 5 in [7], which answers to the above question.

Proposition B. *Let T be an el-tree. Let $\lambda_{\max}(\lambda_{\min})$ denote a reconstruction λ on T such that $\lambda(u) = \max(S_u)$ ($\min(S_u)$) for any internal node u . Then the reconstruction $\lambda_{\max}(\lambda_{\min})$ on T is the greatest (least) element of the MPR-poset $(\mathbf{Rmp}(T), \leq)$.*

In Narushima and Misheva [6, 7], and Narushima [8], the two remarkable properties of ACCTTRAN reconstructions have been shown, and also some conditions for an ACCTTRAN reconstruction to be the greatest element or the least element in the MPR-poset have been given.

In order to investigate ACCTTRAN and DELTRAN reconstructions from another point of view, Minaka [4] has implicitly defined another partial ordering “ a is ancestral to b ” on a polarized transformation series, and then has introduced a partial ordering called “MPR partial order” on $\mathbf{Rmp}(T)$. We now give a mathematically explicit definition for the MPR partial order.

We first define a binary relation $\leq_{\sigma(r)}$ on Ω as follows. Let T be a rooted el-tree (T_s, r) . For a and b in Ω , $a \leq_{\sigma(r)} b$ if and only if $\sigma(r) \leq a \leq b$ or $\sigma(r) \geq a \geq b$. Then, it is easily shown that the relation $\leq_{\sigma(r)}$ is a partial-ordering on Ω .

We next define a binary relation $\leq_{\sigma(r)}$ on $\mathbf{Rmp}(T)$ as follows. Let T be a rooted el-tree (T_s, r) . For λ and μ in $\mathbf{Rmp}(T)$, $\lambda \leq_{\sigma(r)} \mu$ if and only if $\lambda(u) \leq_{\sigma(r)} \mu(u)$ for all u in V_H . Clearly, the binary relation $\leq_{\sigma(r)}$ on $\mathbf{Rmp}(T)$ is a partial-ordering, and then the partially ordered set $(\mathbf{Rmp}(T), \leq_{\sigma(r)})$ is called a $\sigma(r)$ -version MPR-poset.

We here show an example for the MPR-poset $(\mathbf{Rmp}(T), \leq)$ and an example for the $\sigma(r)$ -version MPR-poset $(\mathbf{Rmp}(T), \leq_{\sigma(r)})$. An el-tree $T = (V_O \cup V_H, E, \sigma)$ is shown in Fig.1.

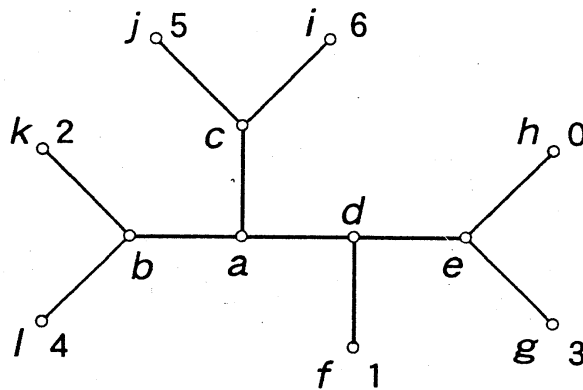
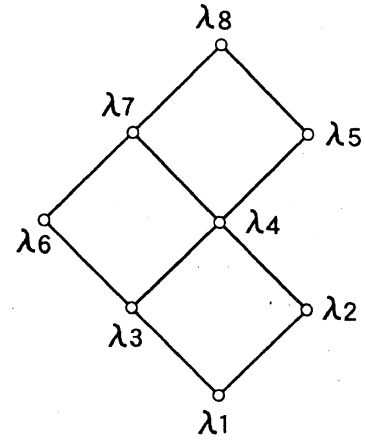


Figure 1: An el-tree T

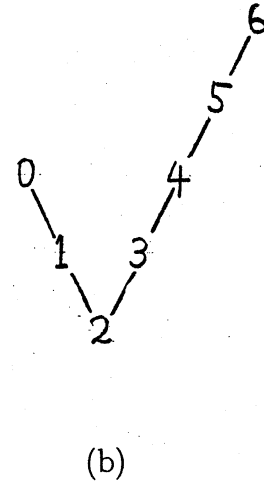
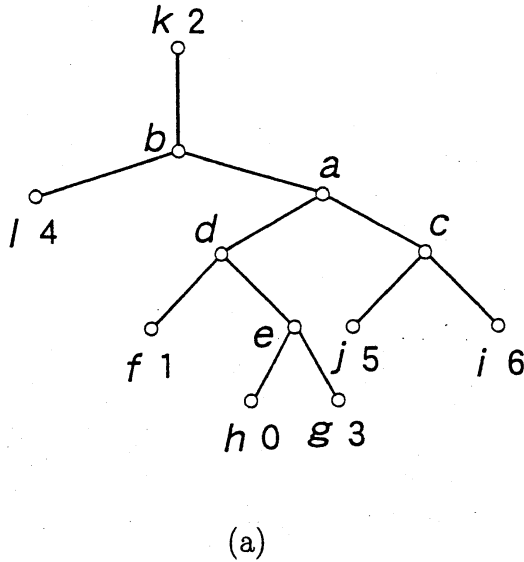
All MPRs on T are recursively generated by Hanazawa-Narushima algorithm and shown in Table 1. Then we have the MPR-poset $(\mathbf{Rmp}(T), \leq)$ shown in Fig.2.

Table 1: The set $\mathbf{Rmp}(T)$ of all MPRs

$\lambda \backslash u$	a	b	c	d	e	f	g	h	i	j	k	l
λ_1	2	2	5	1	1	1	3	0	6	5	2	4
λ_2	2	2	5	2	2	1	3	0	6	5	2	4
λ_3	3	3	5	1	1	1	3	0	6	5	2	4
λ_4	3	3	5	2	2	1	3	0	6	5	2	4
λ_5	3	3	5	3	3	1	3	0	6	5	2	4
λ_6	4	4	5	1	1	1	3	0	6	5	2	4
λ_7	4	4	5	2	2	1	3	0	6	5	2	4
λ_8	4	4	5	3	3	1	3	0	6	5	2	4

Figure 2: The MPR-poset $(\mathbf{Rmp}(T), \leq)$

Let the el-tree T in Fig.1 be rooted at k . Then we have a rooted el-tree $T^{(k)} = (T_b, k)$ shown in Fig.3 (a). Noting $\sigma(k) = 2$, we have the partial-ordering $\leq_{\sigma(k)} = \leq_2$ on Ω , of which Hasse diagram is shown in Fig.3 (b). As a result, we have the 2-version MPR-poset $(\mathbf{Rmp}(T), \leq_2)$ shown Fig.4.

Figure 3: (a) A rooted el-tree (T_b, k) (b) The partial-ordering $\leq_{\sigma(k)} = \leq_2$

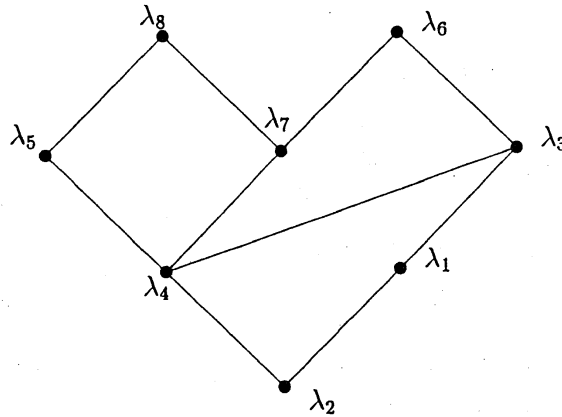


Figure 4: The MPR-poset $(\mathbf{Rmp}(T), \leq_2)$

Note that the usual MPR-poset is uniquely defined for an el-tree, but the $\sigma(r)$ -version MPR-poset depends on the root's character-state of a rooted el-tree $T = (T_s, r)$.

We here describe some lattice-theoretic problems on $\sigma(r)$ -version MPR-posets.

Some lattice-theoretic problems on $\sigma(r)$ -version MPR-posets.

1. Whether there exists the greatest element (or the least element) in each $\sigma(r)$ -version MPR-poset or not ?
2. If there is not the greatest element (or the least element), then what conditions for the existence do we have ?
3. How many maximal (or minimal) elements do we have ?
4. Does any $\sigma(r)$ -version MPR-poset form a lower-semilattice ?

References

- [1] J. M. Farris, Methods for computing Wagner trees, *Systematic Zoology* 19 (1970) 83-92.
- [2] M. Hanazawa, H. Narushima and N. Minaka, Generating most parsimonious reconstructions on a tree: a generalization of the Farris-Swofford-Maddison method, *Discrete Applied Mathematics* 56 (1995) 245-265.

- [3] N. Minaka, Parsimony, phylogeny and discrete mathematics: combinatorial problems in phylogenetic systematics (in Japanese: with English summary), Natural History Research, Chiba Prefectural Museum and Institute, Vol.2 No.2 (1993) 83 - 98.
- [4] N. Minaka, Algebraic Properties of the Most Parsimonious Reconstructions of the Hypothetical Ancestors on a Given Tree, Forma 8 (1993) 277-296.
- [5] H. Narushima and M. Hanazawa, A more efficient algorithm for MPR problems in phylogeny, Discrete Applied Mathematics 80 2-3 (1997) 227-234.
- [6] H. Narushima and N. Misheva, Characteristics of the ACCTTRAN reconstruction in the MPR-poset, Proc. The International Symposium on Combinatorics and Applications (eds:W.Y.C.Chen, D.Z.Du, D.F.Hsu, H.Y.Hap) (June 28 - 30, 1996 : Nankai Institute of Mathematics, Nankai University, Tianjin, P.R.China) 306-313.
- [7] H. Narushima and N. Misheva, On the characteristics of the ancestral character-state reconstruction under the accelerated transformation optimization, preprint.
- [8] H. Narushima, On most-parsimonious reconstruction in phylogeny and extremal properties of ACCTTRAN reconstructions, preprint.
- [9] D. L. Swofford and W. P. Maddison, Reconstructing ancestral character states under Wagner parsimony, Mathematical Biosciences 87 (1987) 199-229.
- [10] 成嶋 弘, 系統樹最節約復元問題の大域的最適解について (On globally optimal reconstructions of phylogenetic trees), 京大数解研講究録 992『計算理論とその応用』(1997年5月), 5 -11.