Assignment of population parameter values to nodes in network or MUL-tree

Graham Jones

2012-04-30

The algorithm for assigning population parameter values from an array to points in the MUL-tree is complex.

First, the nodes are sorted using a criterion which places nodes having the same species clade together. The groups of nodes belonging to different clades are sorted by size so that the tip nodes with clades of size one come first. (Of less importance, but for the record: clades of the same size are sorted lexicographically using the order in which species appear in the XML element alloppSpecies. Nodes within groups with the same clade are sorted using sequence information.)

There may be one, two, or three nodes having the same species clade. The other complications are that tips are different from internal nodes, and that nodes which roots of tetratrees or just below, as well as the root are special cases. It is possible to distinguish eleven cases in total.

	Type	Tip?	#nodes	#values
1	A node in the diploid tree, not the root	No	1	1
2	A node in the diploid tree, not the root	Yes	1	2
3	The root node in the diploid tree	No	1	0
4	A foot node, from a scenario 'A' or 'C' tetratree	No	1	1
5	Two feet nodes, from a scenario 'B' tetratree	No	2	2
6	Node pair, not a tetratree root	No	2	1
7	Node pair, a tetratree root	No	2	1
8	Node pair, not a tetratree root	Yes	2	2
9	Node pair, a tetratree root	Yes	2	2
10	Node pair plus leg-join, root of scenario 'C' tetratree	No	3	1
11	Node pair plus leg-join, root of scenario 'C' tetratree	Yes	3	2

'#nodes' means the number of nodes in the MUL-tree sharing the same species clade. '#values' means the number of population parameter values which are used by this case. (Some values are copied from node to node so the number here depends on exact algorithm.) A 'tetratree' is a tetraploid subtree.

The cases are handled in sets 1-3, 4, 5, 6-9, 10-11. Nodes can have up to three values: tip, rootward, hybridization. Only tip nodes can have tip values; only tetraroots can have hybridization values. The groups of nodes are visited in the sorted order, so that tips are visited first, the root last, and children always visited before parents.

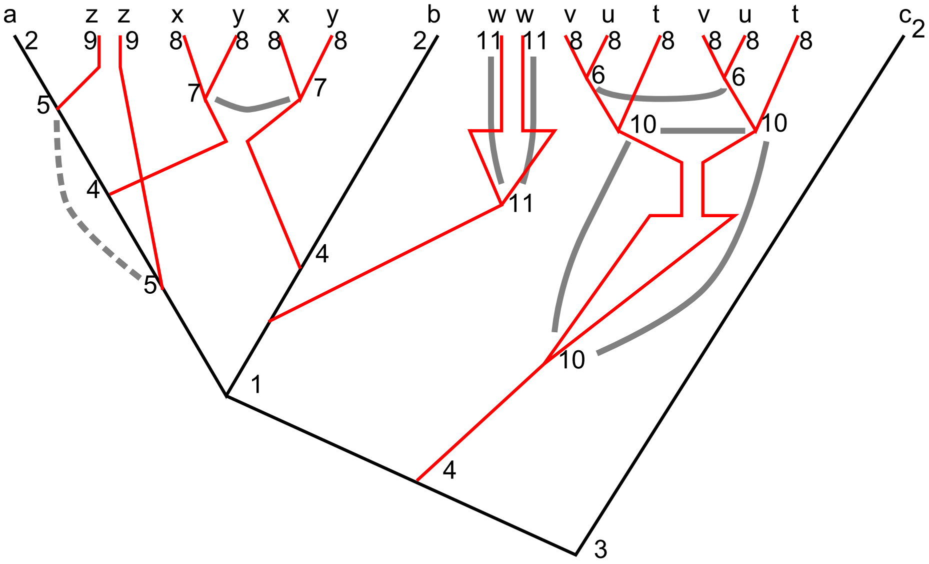

Figure 1: MUL-tree for 3 diploids a,b,c and 7 tetraploids z,y,x,w,v,u,t in 4 tetraploid subtrees. Numbers at nodes indicate cases. Grey lines join groups of nodes dealt with together (but pairs of tips are not indicated).

Cases 1-3. If the root do nothing, otherwise do the following. If a tip assign a tipward value. Assign a rootward value.

Case 4. Assign a rootward value. Copy rootward value from non-tetraroot child to tetraroot child.

Case 5. Assign a rootward value to each foot (uses two values). For each foot, copy rootward value from non-tetraroot child to tetraroot child.

Cases 6-9. If a tip assign a tipward value to both nodes. If it is a tetraroot, assign a hybridization value to both nodes, else assign a rootward value to both nodes.

Cases 10-11. If a tip assign a tipward value to all three nodes. Assign a hybridization value to all three nodes. Assign a rootward value to all three nodes. (Some assignments are redundant.)

In cases 4 and 5, which are feet nodes, the phrase ‘copy rootward value from non-tetraroot child to tetraroot child’ means that a value previously assigned is also used as the rootward value for the tetraroot. Note in cases 6-9 a rootward value is not assigned to a tetraroot.

The total number of values is

$$3n_d - 2 + \sum_{1}^{r}(3n_i + 1) = 3n - 2 + r$$

where there are n_d diploid species, r tetraploid subtrees, n_i tetraploid species in the i'th tetraploid subtree, and a total of n species.

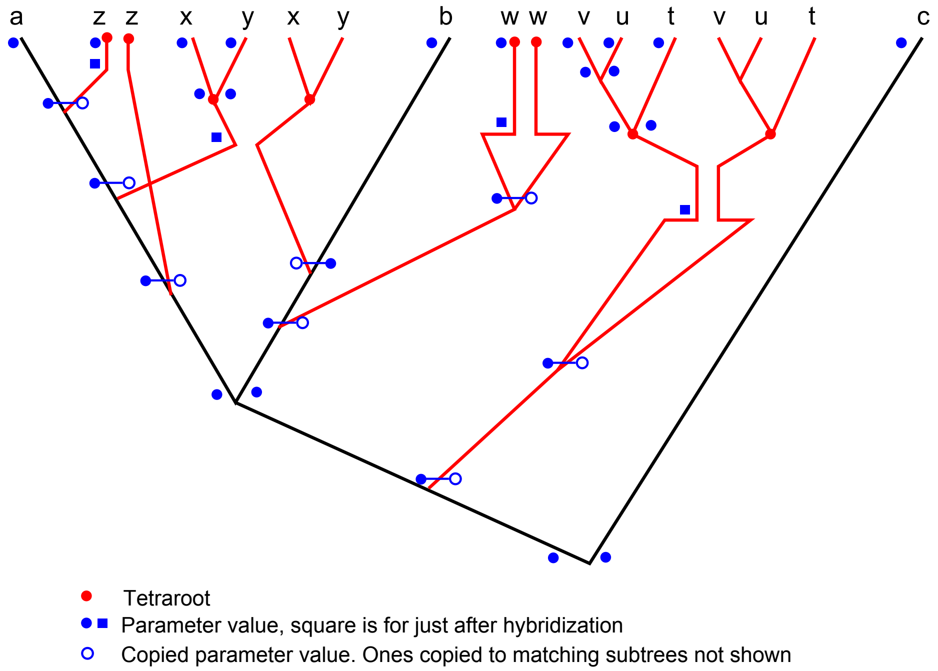

Figure 2: Same MUL-tree but now showing the positions of the 32 population parameter values.