



Fig. 1. (a) Phylogeny T and skeleton $s(T, C')$, $C' = \{3, 4\}$. Edges are labeled with characters that mutate μ and super nodes with tags t . (b) Transform to remove a degree 2 Steiner root from a super node. Note: the size of the phylogeny is unchanged.

We will assume that the above lemma holds on the input matrix for the rest of the paper. Note that such characters c', c'' are identical (after possibly relabeling one character) and are usually referred to as non-informative. It is not hard to show that this preprocessing step does not change the correctness or running time of our algorithm.

The following additional definitions are required for the description and analysis of the simple algorithm:

Definition 9: For any phylogeny T and set of characters $C' \subseteq C$:

- a *super node* is a maximal connected subtree T' of T s.t. for all edges $e \in T'$, $\mu(e) \notin C'$
- the *skeleton* of T , $s(T, C')$, is the tree that results when all super nodes are contracted to a vertex. The vertex set of $s(T, C')$ is the set of super nodes. For all edges $e \in s(T, C')$, $\mu(e) \in C'$.

Definition 10: A tag $t(u) \in \{0, 1\}^m$ of super node u in $s(T, C')$ has the property that $t(u)[c'] = v[c']$ for all $c' \in C'$, vertices $v \in u$; $t[u][i] = 0$ for all $i \notin C'$.

Throughout this paper, we will assume without loss of generality that we are working with phylogenies and skeletons that are rooted at the all zeros taxon and tag respectively. Furthermore, the skeletons used in this work themselves form a perfect phylogeny in the sense that no character mutates more than once in the skeleton. Note that in such skeletons, tag $t(u)[i] = 1$ if and only if character i mutates exactly once in the path from the root to u . Figure 1(a) shows an example of a skeleton of a phylogeny. We will use the term *sub-phylogeny* to refer to a subtree of a phylogeny.