



# Determining Significant Correlation Between Pairs of Extant Characters in a Small Parsimony Framework

KAUSTUBH KHANDAI,<sup>1</sup> CRISTIAN NAVARRO-MARTINEZ,<sup>2</sup> BRENDAN SMITH,<sup>2</sup>  
REBECCA BUONOPANE,<sup>2</sup> SOYONG ASHLEY BYUN,<sup>2,\*</sup> and MURRAY PATTERSON<sup>1,\*</sup>

## ABSTRACT

When studying the evolutionary relationships among a set of species, the principle of parsimony states that a relationship involving the fewest number of evolutionary events is likely the correct one. Due to its simplicity, this principle was formalized in the context of computational evolutionary biology decades ago by, for example, Fitch and Sankoff. Because the parsimony framework does not require a model of evolution, unlike maximum likelihood or Bayesian approaches, it is often a good starting point when no reasonable estimate of such a model is available. In this work, we devise a method for determining if pairs of discrete characters are significantly correlated across all most parsimonious reconstructions, given a set of species on these characters, and an evolutionary tree. The first step of this method is to use Sankoff's algorithm to compute *all* most parsimonious assignments of ancestral states (of each character) to the internal nodes of the phylogeny. Correlation between a pair of evolutionary events (e.g., absent to present) for a pair of characters is then determined by the (co-) occurrence patterns between the sets of their respective ancestral assignments. The probability of obtaining a correlation this extreme (or more) under a null hypothesis where the events happen randomly on the evolutionary tree is then used to assess the significance of this correlation. We implement this method: *parcours* (PARsimonious CO-occURrenceS) and use it to identify significantly correlated evolution among vocalizations and morphological characters in the Felidae family.

**Keywords:** correlation, parsimony, phylogenetics, statistical significance.

## 1. INTRODUCTION

THE PRINCIPLE OF PARSIMONY first appeared in the context of computational evolutionary biology in Edwards and Cavalli-Sforza (1963), Camin and Sokal (1965), Farris (1970), Fitch (1971). Following this, Sankoff and Rousseau generalized this to allow the association of a (different) cost to each transition between a pair of states (Sankoff, 1975; Sankoff and Rousseau, 1975). A few years after that, Felsenstein

<sup>1</sup>Department of Computer Science, Georgia State University, Atlanta, Georgia, USA.

<sup>2</sup>Department of Biology, Fairfield University, Fairfield, Connecticut, USA.

\*Joint authors.

noticed that parsimony could produce misleading results when the evolutionary rate of change on the branches of the phylogeny is high (Felsenstein, 1978). Because parsimony does not take branch length into account, it ignores the fact that many changes on a long branch—while being far from parsimonious—may not be so unlikely in this case, resulting in a bias coined as “long branch attraction.” This paved the way for a proposed refinement to parsimony known as the maximum likelihood method (Felsenstein, 1978, 1981). Because of an explosion in molecular sequencing data, and the sophisticated understanding of the evolutionary rate of change in this setting, maximum likelihood has become the de facto framework for inferring phylogeny. Many popular software tools which implement maximum likelihood include PHYLIP (Felsenstein, 1989), RAxML (Stamatakis, 2006), and the Bio++ library (Guéguen et al., 2013). Even more recently, some Bayesian approaches have appeared, which sample the space of likely trees using the Markov-Chain Monte Carlo method, resulting in tools such as MrBayes (Huelsenbeck and Ronquist, 2001) and Beast (Drummond and Rambaut, 2007).

However, in cases where there is no model of evolution, or no reasonable estimation of the rate of evolutionary change on the branches of the phylogeny, parsimony is a good starting point. In fact, even in the presence of such a model, there are still conditions under which a maximum likelihood phylogeny is always a maximum parsimony phylogeny (Tuffley and Steel, 1997; Steel and Penny, 2004). Finally, when the evolutionary rate of change of different characters is heterogeneous, that is, there are (different) rates of change, maximum parsimony has even been shown to perform *substantially better* than maximum likelihood and Bayesian approaches (Kolaczowski and Thornton, 2004). A good example is cancer phylogenetics (Hajirasouliha and Raphael, 2014; Schwartz and Schaffer, 2017; El-Kebir, 2018; Ciccolella et al., 2020b), where little is known about the modes of evolution of cancer cells in a tumor. Due to its importance, this setting has seen a renewed interest in high-performance methods for computing maximum parsimonies in practice, such as SCITE (Jahn et al., 2016) and SASC (Ciccolella et al., 2020a).

The setting of the present work is an evolutionary study of vocalizations and morphological characters (Castelló, 2020) among members of the family Felidae. Felids possess a range of intraspecific vocalizations for close, medium, and long range communication. These vocalizations can vary from discrete calls such as the spit or hiss to more graded calls such as the mew, main call, growl, and snarl (Sunquist and Sunquist, 2002). There is no evidence that felid vocalizations are learned: it is more likely that these calls are genetically determined (Peters, 1978; Ehret, 1980; Romand and Ehret, 1984). While there are 14 major discrete and graded calls documented in Felidae, not all calls are produced by all species (Peters, 1991). In Peters and Tonkin-Leyhausen (1999), the authors map some of these calls to a molecular phylogeny of Felidae to show that it is consistent with what was previously known, strengthening the argument that vocalizations are genetically determined. In this study, we consider a similar approach, but because (1) an obvious model of evolution is lacking in this case; (2) the possibility that vocalizations within a given group of species can evolve at considerably different rates (Peters and Tonkin-Leyhausen 1999); and (3) that rates for specific characters can differ between different lineages within that group (Peters and Tonkin-Leyhausen 1999), it follows from Kolaczowski and Thornton (2004) that parsimony is more appropriate than maximum likelihood or Bayesian approaches.

In this work, we develop a general framework to determine, among pairs of characters in a phylogeny, which of them are significantly correlated (or co-occurring), when no model of evolution is present. We then use this framework to understand how these vocalizations and morphological characters may have evolved within Felidae, how they might correlate (or have coevolved) with each other, and which ones among these are significantly correlated. The first step of this approach is to infer, for each character, the set of *all* most parsimonious assignments of ancestral states (small parsimonies) in the phylogeny. Then, for each character, we construct from its set of small parsimonies, a *consensus map* (Section 3) for each (type of) evolutionary event (e.g., absent to present) along branches of the phylogeny. Correlation between a pair of evolutionary events (for a pair of characters) is then determined by how much their respective consensus maps overlap. Finally, this correlation is deemed significant if it is higher than expected in a null model where the evolutionary events happen randomly on the tree. We implement this approach in a tool called *parcours* and use it to detect significantly correlated evolution among felid vocalizations and morphological characters, obtaining results that are consistent with the literature (Peters and Tonkin-Leyhausen 1999), as well as some interesting associations.

While various methods for detecting correlated evolution exist, they tend to use only *phylogenetic profiles* (Marcotte et al., 2000), or are based on maximum likelihood (Barker et al., 2007; Cohen et al., 2013; Patterson et al., 2016), where a model of evolution is needed. Methods that determine coevolution in

the parsimony framework exist as well (Patterson et al., 2013; Duchemin et al., 2017); however, they are aimed at reconstructing ancestral gene *adjacencies*, given extant colocalization information. Finally, while some of the maximum likelihood software tools have a “parsimony mode,” for example, Felsenstein (1989), Guéguen et al. (2013), the character information must be encoded using restrictive alphabets, and there is no automatic way to compute *all* most parsimonious ancestral states for a character—something which is central to our framework. On the contrary, *parcours* takes character information as a *column-separated values* file, inferring the alphabet from this input, and efficiently computes all small parsimonies. In summary, our contribution is a methodology for inferring correlation among pairs of characters when no model of evolution is available, assessing the significance of these associations, and implementing it in an open-source software that is easy enough to be used as a pedagogical tool. The *parcours* tool is freely available at <https://github.com/murraypatterson/parcours>.

This article is structured as follows. In Section 2, we provide the background on parsimony, as well as our approach for efficiently computing *all* most parsimonious ancestral states in a phylogeny, given a set of extant states. In Section 3, we present our approach for computing correlation between pairs of characters from all such parsimonies. In Section 4, we describe how to compute the probability of obtaining a correlation as high (or higher) by chance. In Section 5, we describe an experimental analysis of the implementation, *parcours*, of our method on felid vocalizations and morphological characters and then discuss the results in Section 6. Finally, we conclude the article in Section 7 with a discussion of future directions.

## 2. SMALL PARSIMONY

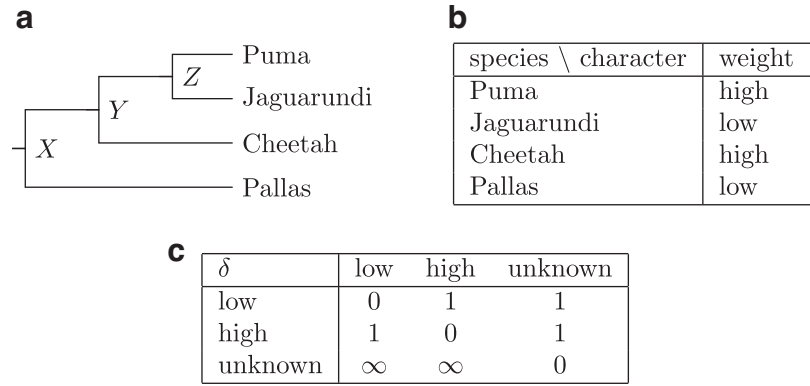
In computing a parsimony, the input is typically *character based*, involving a set of species, each over a set of *characters* (e.g., weight category), where each character can be in a number of *states* (e.g., low, medium, high, unknown, and so on). The idea is that each species is in one of these states for each character—for example, in the Puma, the weight [a morphological character (Castelló, 2020)] is “high”—and we want to understand what states the ancestors of these species could be in. Given a *phylogeny* on these species: a tree where all of these species are leaves, a *small parsimony* is an assignment of states to the internal (ancestral) nodes of this tree which minimizes the number of changes of state among the characters along branches of the tree.\* We illustrate this with the following example.

Suppose we have the four species: Puma, Jaguarundi, Cheetah (of the Puma lineage), and Pallas cat (outgroup species from the nearby Leopard cat lineage), alongside the phylogeny depicted in Figure 1a, implying the existence of the ancestors  $X$ ,  $Y$ , and  $Z$ . We are given some character, for example, weight, which can be in one of a number of states taken from *alphabet*  $\Sigma = \{\text{low, high, unknown}\}$ . If weight is high only in Puma and Cheetah as in Figure 1b, then the assignment of low to all ancestors would be a small parsimony. This parsimony has two changes of state: a change  $\text{low} \rightarrow \text{high}$  on the branches  $Z \rightarrow \text{Puma}$  and  $Y \rightarrow \text{Cheetah}$ —implying convergent increase of weight in these species. Another small parsimony is that weight is high in  $Y$  and  $Z$  (low in  $X$ )—implying that high weight was ancestral (in  $Y$ ) to the Puma, Jaguarundi, and Cheetah and later decreased in the Jaguarundi. A principled way to infer all small parsimonies is to use Sankoff’s algorithm (Sankoff, 1975), which also makes use of a *cost matrix*  $\delta$ , depicted in Figure 1c, that encodes the evolutionary cost of each state change along a branch. Since the change  $\text{low} \rightarrow \text{high}$  costs  $\delta_{\text{low, high}} = 1$ , and vice versa ( $\delta_{\text{high, low}} = 1$ ), it follows that each of the small parsimonies mentioned above have an overall cost, or *score* of 2 in this framework. A simple inspection of all possibilities shows that 2 is the minimum score of any assignment of ancestral states.

Sankoff’s algorithm (Sankoff, 1975) computes all (small) parsimonies given a phylogeny and the extant states of a character in a set of species, and a cost matrix  $\delta$ , for example, Figure 1. The algorithm has a bottom-up phase (from the leaves to the root of the phylogeny) and then a top-down phase (from the root to the leaves). The bottom-up phase is to compute  $s_i(u)$ : the minimum score of an assignment of ancestral states in the subtree of the phylogeny rooted at node  $u$  when  $u$  has state  $i \in \Sigma$ , according to the recurrence:

---

\*A *large parsimony* is then a phylogeny which results in a small parsimony with the fewest changes (and this resulting small parsimony).



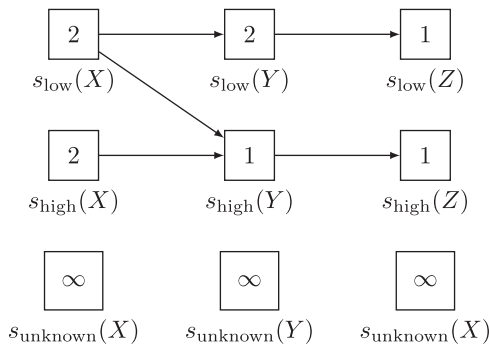
**FIG. 1.** A (a) phylogeny and (b) the extant state of character weight in four species; and (c) the cost  $\delta_{i,j}$  of the change  $i \rightarrow j$  from state  $i$  in the parent to state  $j$  in the child along a branch of the phylogeny, for example,  $\delta_{\text{high}, \text{unknown}} = 1$  ( $\delta_{\text{unknown}, \text{high}} = \infty$ ).

$$s_i(u) = \sum_{v \in C} \min_{j \in \Sigma} \{s_j(v) + \delta_{i,j}\} \quad (1)$$

where  $C$  is the set of children of node  $u$  in the phylogeny. The idea is that the score is known for any extant species (a leaf node in the phylogeny) and is coded as  $s_i(u) = 0$  if species  $u$  has state  $i$  and  $\infty$  otherwise. The score for the internal nodes is then computed according to Equation (1) in a bottom-up dynamic programming manner, starting from the leaves. The boxes in Figure 2 depict the result of this first phase on the instance of Figure 1, for example,

$$\begin{aligned} s_{\text{high}}(Z) &= \min\{s_{\text{low}}(\text{Jaguarundi}) + \delta_{\text{high}, \text{low}}, s_{\text{high}}(\text{Jaguarundi}) + \delta_{\text{high}, \text{high}}, \\ &\quad s_{\text{unknown}}(\text{Jaguarundi}) + \delta_{\text{high}, \text{unknown}}\} \\ &= \min\{s_{\text{low}}(\text{Puma}) + \delta_{\text{high}, \text{low}}, s_{\text{high}}(\text{Puma}) + \delta_{\text{high}, \text{high}}, \\ &\quad s_{\text{unknown}}(\text{Puma}) + \delta_{\text{high}, \text{unknown}}\} \\ &= \min\{0 + 1, \infty + 0, \infty + 1\} + \min\{\infty + 1, 0 + 0, \infty + 1\} \\ &= \min\{1, \infty, \infty\} + \min\{\infty, 0, \infty\} = 0 + 1 = 1 \end{aligned} \quad (2)$$

$$\begin{aligned} s_{\text{high}}(Y) &= \min\{s_{\text{low}}(\text{Cheetah}) + \delta_{\text{high}, \text{low}}, s_{\text{high}}(\text{Cheetah}) + \delta_{\text{high}, \text{high}}, \\ &\quad s_{\text{unknown}}(\text{Cheetah}) + \delta_{\text{high}, \text{unknown}}\} \\ &= \min\{s_{\text{low}}(Z) + \delta_{\text{high}, \text{low}}, s_{\text{high}}(Z) + \delta_{\text{high}, \text{high}}, s_{\text{unknown}}(Z) + \delta_{\text{high}, \text{unknown}}\} \\ &= \min\{\infty + 1, 0 + 0, \infty + 1\} + \min\{1 + 1, 1 + 0, \infty + 1\} \\ &= \min\{\infty, 0, \infty\} + \min\{2, 1, \infty\} = 0 + 1 = 1 \end{aligned} \quad (3)$$



**FIG. 2.** Graph structure resulting from Sankoff's algorithm on the instance of Figure 1. The boxes contain the values of  $s_i(u)$  computed in the bottom-up phase, for each state  $i \in \Sigma$  and internal node  $u \in \{X, Y, Z\}$  of the phylogeny. The arrows connecting these boxes are computed in the top-down phase.

$$\begin{aligned}
s_{\text{low}}(X) &= \min\{s_{\text{low}}(\text{Pallas}) + \delta_{\text{low, low}}, s_{\text{high}}(\text{Pallas}) + \delta_{\text{low, high}}, \\
&\quad s_{\text{unknown}}(\text{Pallas}) + \delta_{\text{low, unknown}}\} \\
&\quad + \min\{s_{\text{low}}(Y) + \delta_{\text{low, low}}, s_{\text{high}}(Y) + \delta_{\text{low, high}}, \\
&\quad s_{\text{unknown}}(Y) + \delta_{\text{low, unknown}}\} \\
&= \min\{0+0, \infty+1, \infty+1\} + \min\{2+0, 1+1, \infty+1\} \\
&= \min\{0, \infty, \infty\} + \min\{2, 2, \infty\} = 0+2 = 2
\end{aligned} \tag{4}$$

In general, for a given character with  $|\Sigma|=k$ , this procedure would take time  $O(nk)$ , where  $n$  is the number of species, since we compute  $nk$  values, and computing each one takes (amortized) constant time.

After the bottom-up phase, we know the minimum score of any assignment of ancestral states (it is the minimum of those  $s_i$  for  $i \in \Sigma$ , at the root), but we do not yet have an ancestral assignment of states. In this study, since  $X$  is the root in Figure 1a, we see from Figure 2 that the minimum score for this example is 2, as we saw earlier. Note that this minimum may not be unique; indeed, here  $s_{\text{low}}(X) = s_{\text{high}}(X) = 2$ , meaning that in at least one parsimony, weight is low in  $X$ , and in at least one other parsimony, weight is high in  $X$  [but never unknown in any parsimony, i.e.,  $s_{\text{unknown}}(X) = \infty$ ]. Now, to reconstruct one of these ancestral assignments of minimum score—the top-down phase of Sankoff's algorithm—we first assign to the root, one of the states  $i \in \Sigma$  for which  $s_i$  is minimum. We then determine those states in each child of the root from which  $s_i$  can be derived (these may not be unique either), and assign those states accordingly. We continue, recursively, in a top-down manner until we reach the leaves of the tree, having assigned all states at this point. For example,  $s_{\text{low}}(X) = 2$ , and can be derived in Equation (4) from  $s_{\text{high}}(Y)$  [and  $s_{\text{low}}(\text{Pallas})$ ], which is in turn derived in Equation (3) from  $s_{\text{high}}(Z)$  [and  $s_{\text{high}}(\text{Cheetah})$ ]. This corresponds to the second parsimony mentioned earlier, where high weight was ancestral in  $Y$  and later decreased in the Jaguarundi. Notice that  $s_{\text{low}}(X)$  can also be derived in Equation (4) from  $s_{\text{low}}(Y)$ , which is in turn derived from  $s_{\text{low}}(Z)$ . This is the first parsimony where weight is low in all ancestors.

One can compactly represent *all* parsimonies as a graph structure with a *box* for each  $s_i$  at each internal node of the phylogeny (e.g., boxes of Fig. 2), and an *arrow* from box  $s_i(u)$  to  $s_j(v)$  for some node  $u$  and its child  $v$  in the phylogeny, whenever  $s_i(u)$  can be derived from  $s_j(v)$  (e.g., arrows of Fig. 2). A parsimony is then some choice of exactly one box in this graph for each internal node of the phylogeny in such a way that they form an underlying directed spanning subtree in this graph, in terms of the arrows which join the boxes. This spanning subtree will have the same topology as the phylogeny, in fact, and the choice of box will correspond to the choice of ancestral state of the corresponding internal nodes of this phylogeny.

In Figure 2, for example, the spanning subtree  $s_{\text{low}}(X) \rightarrow s_{\text{low}}(Y) \rightarrow s_{\text{low}}(Z)$  corresponds to the parsimony where weight is low in all ancestors, and  $s_{\text{low}}(X) \rightarrow s_{\text{high}}(Y) \rightarrow s_{\text{high}}(Z)$  corresponds to the parsimony where high weight was ancestral in  $Y$  and decreased in the Jaguarundi. Implicitly the leaves are also included in these spanning subtrees, but since there is no choice of state for extant species, they are left out for simplicity—see Clemente et al. (2009) for another example of this graph structure, with the leaves included. Notice, from Figure 2 that there is a third solution  $s_{\text{high}}(X) \rightarrow s_{\text{high}}(Y) \rightarrow s_{\text{high}}(Z)$  corresponding to a parsimony where high weight was ancestral (in  $X$ ) to all species here and then decreased in both the Pallas cat and the Jaguarundi. Note that for internal nodes  $u$  other than the root, that a state  $i \in \Sigma$  for which  $s_i(u)$  is *not* minimized can appear in a solution, for example,  $s_{\text{low}}(Y)$ . In general, for  $|\Sigma|=k$ , the procedure for building this graph structure (e.g., in Fig. 2) would take time  $O(nk^2)$ , since at each node  $u$ , each of the  $k$  values  $s_i(u)$  is derived from  $O(k)$  values at the (amortized constant number of) children of  $u$ .

Note that there can be  $O(k^{n-1})$  parsimonies in general. For example, consider the extreme case where all costs in  $\delta$  are 0. For each internal node  $v$  and its parent  $u$  in the phylogeny, the set  $s_i(v)$  of  $k$  boxes forms a complete bipartite subgraph with the set  $s_j(u)$  of  $k$  boxes in the corresponding graph structure. Any choice of  $k$  states in each of the  $n-1$  internal nodes (along with the unique choice for each leaf) would form a spanning subtree, giving rise to  $k^{n-1}$  parsimonies. In the general (less extreme) case, we first reduce the set of states for each internal node to those in a connected component of the graph containing a minimum score root state, as only those have the potential to be in a spanning subtree. For example, in Figure 2, such (sets of) states are  $s_{\text{low}}(u)$  and  $s_{\text{high}}(u)$  for  $u \in \{X, Y, Z\}$ . Viewing the combinations of each possible choice of state for each internal node as elements of a Cartesian product (a subset of  $\{1, \dots, k\}^{n-1}$ ), a standard (e.g., based on the left-hand rule) depth-first traversal of this graph structure would visit the elements in *lexicographical order*. However, it is possible to visit the Cartesian product of a set in *Gray code* order

(Mossige, 1977), so that each consecutive element in the visit differs by exactly one co-ordinate. Visiting the graph structure in this manner allows for constant-time updates between consecutive potential combinations. That is, if a combination forms a spanning subtree, then the next combination, differing only by co-ordinate  $u$  (a node in the tree), updates only the edge to the parent and to the (amortized constant number of) children of  $u$  in this spanning subtree. It then suffices to check if this update is also a spanning subtree to verify if the current combination also forms a spanning subtree. While this graph structure (Fig. 2) is mentioned (Clemente et al., 2009), how to efficiently enumerate all its parsimonies is rarely discussed (that we know of). In contrast, since having all parsimonies is central to our approach, the above Gray code enumeration strategy is implemented into parcours.

### 3. CORRELATION

We want to determine correlated evolution among pairs of characters. Suppose we have a second character, dental profile (number of teeth), which can have states from  $\{28, 30, \text{unknown}\}$ . Here, the dental profile is “fewer” (28) only in the Pallas cat as depicted in Figure 3a, and we want to understand how this is correlated to the weight of Figure 1b.

The idea is that we first construct likely hypotheses for what ancestral states each character  $\alpha$  might have, namely the set  $P(\alpha)$  of all small parsimonies. Then we determine if there are any correlated changes in state of pairs of characters along branches of the phylogenetic tree given their sets of parsimonies. While a drawback of parsimony is potentially many solutions, without any a priori knowledge of ancestral state, choosing one solution (or a subset of solutions) is an arbitrary decision. Since such a decision could potentially bias any downstream analysis, our approach attempts to avoid this, in using the entire set of parsimonies. Let  $i \rightarrow j$  be a change in state (for some pair  $i, j$  of states) of character  $\alpha$ . Let  $p|_{i \rightarrow j}$  be the *multiset* (even though each element will have multiplicity 0 or 1) of *branches* in the phylogenetic tree where change  $i \rightarrow j$  occurs in parsimony  $p \in P(\alpha)$ . The *consensus map*  $C_\alpha(i \rightarrow j)$  for change  $i \rightarrow j$  of character  $\alpha$  is then the multiset resulting from

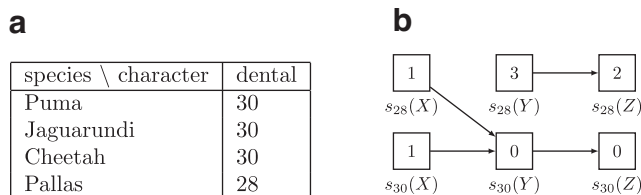
$$\bigcup_{p \in P(\alpha)} p|_{i \rightarrow j}, \quad (5)$$

in preserving the information of *all* parsimonies. The idea is that the consensus map captures the area of the tree (a set of branches), where this type of change occurs. Note that being a multiset is important, as the same change from different parsimonies may occur on the same branch of the tree, and we want to count this. The *correlation* between some change  $i \rightarrow j$  of character  $\alpha$  and some change  $i' \rightarrow j'$  of a character  $\beta$  is then the *weighted Jaccard index* (Jaccard, 1908), also known as the Ružička index (Ružička, 1958)

$$\frac{|C_\alpha(i \rightarrow j) \cap C_\beta(i' \rightarrow j')|}{|C_\alpha(i \rightarrow j) \cup C_\beta(i' \rightarrow j')|}. \quad (6)$$

The weighted Jaccard index measures the degree to which the respective consensus maps overlap, taking into account multiplicities on the branches.

For example, given the characters weight (Fig. 1b) and dental profile (Fig. 3a), we compute the correlation of the change low  $\rightarrow$  high (in weight) and  $28 \rightarrow 30$  (in dental profile) in the phylogeny of Figure 1a (according to cost of Fig. 1c) as follows. Character weight has the set  $P(\text{weight}) = \{p_1, p_2, p_3\}$  of three parsimonies, with  $p_1|_{\text{low} \rightarrow \text{high}} = \{Z \rightarrow \text{Puma}, Y \rightarrow \text{Cheetah}\}$ ,  $p_2|_{\text{low} \rightarrow \text{high}} = \{X \rightarrow Y\}$ , and  $p_3|_{\text{low} \rightarrow \text{high}} = \emptyset$ . It follows from Equation (5) that the consensus map  $C_{\text{weight}}(\text{low} \rightarrow \text{high}) = \{Z \rightarrow \text{Puma}, Y \rightarrow \text{Cheetah}, X \rightarrow Y\}$ . By inspecting Figure 3b, character dental has the set  $P(\text{dental}) = \{q_1, q_2\}$  of two parsimonies, with  $q_1|_{28 \rightarrow 30} = \{X \rightarrow Y\}$  and  $q_2|_{28 \rightarrow 30} = \emptyset$ , hence  $C_{\text{dental}}(28 \rightarrow 30) = \{X \rightarrow Y\}$ . It then follows from Equation (6) that the correlation



**FIG. 3.** The (a) extant state of character dental in four species and (b) the structure resulting from Sankoff's algorithm on these extant states along with the phylogeny of Figure 1a and cost of Figure 1c. Values  $s_2$  have been omitted for compactness, since they are all  $\infty$ , like in Figure 2.

of this co-event is  $1/3 \approx 0.33$ . The correlation of the co-event  $\text{high} \rightarrow \text{low}$  and  $30 \rightarrow 28$  in weight and dental is also  $1/3$ , while the correlation of  $\text{low} \rightarrow \text{high}$  in weight and  $30 \rightarrow 28$  in dental (and  $\text{high} \rightarrow \text{low}$  in weight and  $28 \rightarrow 30$  in dental) is 0, because there is no overlap between sets  $p|_{\text{low} \rightarrow \text{high}}$  and  $q|_{30 \rightarrow 28}$  for any pair of parsimonies  $p \in P(\text{weight})$  and  $q \in P(\text{dental})$  (and vice versa).

For completeness, the correlation of any combination of event involving an unknown state in one character with any other event in the other character is 0 because none of these events happen in any parsimony of weight or dental. We use the weighted Jaccard index because it measures the amount of event co-occurrence, normalized by the amount of independent occurrences of either event in the set of all parsimonies, taking multiplicities into account. If one wanted focus on just the events on the different branches (without multiplicity), one could use the unweighted Jaccard index, which is Equation (6) where all multisets are treated as *sets* (all nonzero multiplicities cast to 1).

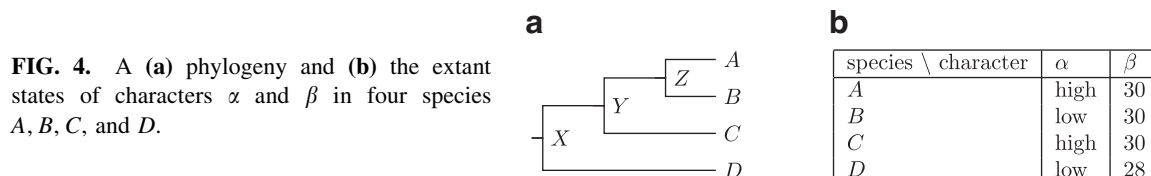
In a similar approach (Patterson et al., 2016), although in a maximum likelihood setting where there are branch lengths, the authors map the probabilities of gain and loss to each branch of the phylogeny in the most likely ancestral assignment. Correlation between such events in a pair of characters is then assessed by computing the mutual information between the associated sets of probabilities. Our approach could be viewed as a discrete analog of this.

#### 4. SIGNIFICANT CORRELATION

The correlation between two changes (in two characters), presented in Section 3, is the degree of overlap between their consensus maps. However, we want to know if this correlation is *significant*, that is, if it is higher than expected by chance. A consensus map  $C_\alpha(i \rightarrow j)$  for change  $i \rightarrow j$  in character  $\alpha$  is constructed from the set  $P(\alpha)$  of small parsimonies, each of which represents a (smallest) number of particular changes on the tree. Here, the null hypothesis posits that these particular changes happened randomly on the tree. A “random” consensus map for change  $i \rightarrow j$  of character  $\alpha$  is then obtained assuming that the changes, in each  $p \in P(\alpha)$  of the parsimonies it is constructed from, happened randomly on the tree. To determine if a correlation between some change  $i \rightarrow j$  in character  $\alpha$  and some change  $i' \rightarrow j'$  in character  $\beta$  is significant, we need to determine if the probability of obtaining this correlation is significantly higher than the correlation of a pair of random consensus maps for  $i \rightarrow j$  and  $i' \rightarrow j'$ . To do this, we need to compute, for change  $i \rightarrow j$  (and  $i' \rightarrow j'$ , respectively), the collection  $C_\alpha^{i \rightarrow j}$  (and  $C_\beta^{i' \rightarrow j'}$ , respectively) of consensus maps obtained from all possible combinations of ways in each  $p \in P(\alpha)$ , of distributing the changes in  $p$  on the tree. A (null) distribution of correlation is then obtained by computing the correlation for each possible pair of consensus maps from collections  $C_\alpha^{i \rightarrow j}$  and  $C_\beta^{i' \rightarrow j'}$ . A correlation between  $i \rightarrow j$  and  $i' \rightarrow j'$  is then deemed significant if the probability of obtaining it is sufficiently extreme in this distribution.

We continue the example of the phylogenetic tree of Figure 1a and character weight (Fig. 1b) and dental profile (Fig. 3a) to illustrate the notion of significant correlation. For compactness, we relabel Puma, Jaguarundi, Cheetah, and Pallas Cat, species  $A, B, C$ , and  $D$ , respectively, as depicted in Figure 4a. We also relabel weight and dental profile, characters  $\alpha$  and  $\beta$ , respectively (Fig. 4b). For further compactness, we refer to a branch according to its *descendant* species, and so, for example, the branch from  $Y$  to  $C$  would be referred to as  $C$ . For even further compactness, since our characters  $\alpha$  and  $\beta$  are binary, for a given branch, for example,  $Y$ , we use the upper case  $Y$  to refer to a change of state  $\text{low} \rightarrow \text{high}$  or  $28 \rightarrow 30$  (an increase) for their respective characters  $\alpha$  and  $\beta$  on this branch and the lower case  $y$  to refer to change  $\text{high} \rightarrow \text{low}$  or  $30 \rightarrow 28$  (a decrease) on this branch, respectively, when the context is clear.

For character  $\alpha(\text{weight})$ , the set  $P(\alpha)$  is then  $p_1 = \{A, C\}$ ,  $p_2 = \{b, d\}$ , and  $p_3 = \{Y, b\}$ , and for character  $\beta$  (dental profile), the set  $P(\beta)$  is then  $q_1 = \{Y\}$  and  $q_2 = \{d\}$ . It follows that the consensus maps  $C_\alpha(\text{low} \rightarrow \text{high})$  and  $C_\alpha(\text{high} \rightarrow \text{low})$ , for the changes  $\text{low} \rightarrow \text{high}$  and  $\text{high} \rightarrow \text{low}$  in character  $\alpha$ , respectively, are



**FIG. 4.** A (a) phylogeny and (b) the extant states of characters  $\alpha$  and  $\beta$  in four species  $A, B, C$ , and  $D$ .

$$C_\alpha(\text{low} \rightarrow \text{high}) = \bigcup_{p \in \{p_1, p_2, p_3\}} p|_{\text{low} \rightarrow \text{high}} = \{A, C, Y\}, \quad (7)$$

$$C_\alpha(\text{high} \rightarrow \text{low}) = \bigcup_{p \in \{p_1, p_2, p_3\}} p|_{\text{high} \rightarrow \text{low}} = \{b, b, d\}. \quad (8)$$

The consensus maps for the changes  $28 \rightarrow 30$  and  $30 \rightarrow 28$  in character  $\beta$  are then  $C_\beta(28 \rightarrow 30) = q_1 = \{Y\}$  and  $C_\beta(30 \rightarrow 28) = q_2 = \{d\}$ , respectively. The correlation of a change  $\text{low} \rightarrow \text{high}$  in character  $\alpha$  with  $28 \rightarrow 30$  in character  $\beta$  is then

$$\frac{|C_\alpha(\text{low} \rightarrow \text{high}) \cap C_\beta(28 \rightarrow 30)|}{|C_\alpha(\text{low} \rightarrow \text{high}) \cup C_\beta(28 \rightarrow 30)|} = \frac{|\{Y\}|}{|\{A, C, Y\}|} = \frac{1}{3}. \quad (9)$$

To determine if this correlation is significant, we need to compute the null distribution—the correlation between each pair from (the Cartesian product of) collections  $\mathcal{C}_\alpha^{\text{low} \rightarrow \text{high}}$  and  $\mathcal{C}_\beta^{28 \rightarrow 30}$ . Consensus map  $C_\alpha(\text{low} \rightarrow \text{high})$  comes from  $P(\alpha) = \{p_1, p_2, p_3\}$ , so we need to compute all possible combinations of ways that the events of each of  $p_1, p_2$ , and  $p_3$  can happen on the tree and then take a consensus map from each member of the (Cartesian) product of these sets of possible ways to obtain  $\mathcal{C}_\alpha^{\text{low} \rightarrow \text{high}}$ . For some alphabet  $\Sigma$ , let  $t(\Sigma) = \{i \rightarrow j | i, j \in \Sigma, j \neq i\}$ , that is, the set of possible changes (or transitions) between any pair of characters from  $\Sigma$ . For example, the alphabet of character  $\alpha$  is  $\Sigma_\alpha = \{\text{low}, \text{high}\}$ , so  $t(\Sigma_\alpha) = \{\text{low} \rightarrow \text{high}, \text{high} \rightarrow \text{low}\}$ . Let  $e^\alpha(p)$  then be the number of instances of each type of change (or event) from  $t(\Sigma_\alpha)$  in parsimony  $p \in P(\alpha)$  for character  $\alpha$ —effectively a multiset—expressed as a tuple (in some fixed ordering, e.g., lexicographic, on  $\Sigma_\alpha$ ). For example, where  $\Sigma_\alpha = \{\text{low}, \text{high}\}$ ,  $e^\alpha(p_1) = (2, 0)$ , for 2  $\text{low} \rightarrow \text{high}$  events and 0  $\text{high} \rightarrow \text{low}$  events, while  $e^\alpha(p_2) = (0, 2)$  and  $e^\alpha(p_3) = (1, 1)$ . Now, let  $W_T^\alpha(e)$  be the set of ways that some multiset  $e$  of events [from set  $t(\Sigma_\alpha)$  of possible events for character  $\alpha$ ] can happen on a tree  $T$ . For example, in phylogeny (call it  $T$ ) of Figure 4a,  $W_T^\alpha(2, 0) = \{\{A, B\}, \{C, A\}, \{C, B\}, \{Z, C\}, \{D, A\}, \{D, B\}, \{D, Z\}, \{D, C\}, \{Y, D\}\}$ —the set of ways that 2  $\text{low} \rightarrow \text{high}$  events (and 0  $\text{high} \rightarrow \text{low}$  events) in character  $\alpha$  could happen on  $T$ . Note that  $p_1, p_2$ , and  $p_3$  are each just one of the possible ways from  $W_T^\alpha(2, 0)$ ,  $W_T^\alpha(0, 2)$ , and  $W_T^\alpha(1, 1)$ , respectively. The collection  $\mathcal{C}_\alpha^{\text{low} \rightarrow \text{high}}$  is then

$$\left\{ \bigcup_{w \in \mathcal{W}} w|_{\text{low} \rightarrow \text{high}} \mid \mathcal{W} \in \times_{p \in P(\alpha)} W_T^\alpha(e^\alpha(p)) \right\}, \quad (10)$$

which is effectively the consensus map  $C_\alpha(\text{low} \rightarrow \text{high})$  [Eq. (5)] for change  $\text{low} \rightarrow \text{high}$  of character  $\alpha$  from each element [one of the possible combinations of ways that the events of each  $p \in P(\alpha)$  can happen on the tree] of the Cartesian product  $\times_{p \in P(\alpha)} W_T^\alpha(e^\alpha(p))$ . The remaining detail lies in how we compute  $W_T^\alpha(e)$  for some multiset  $e$  of events for character  $\alpha$  on a tree  $T$ .

Computing  $W_T^\alpha(e)$  for some multiset  $e$  of events in character  $\alpha$  on a tree  $T$  is done using a dynamic programming approach which combines the sets of possible ways the events can happen on subtrees of  $T$  in a bottom-up manner. This is similar to the computation of  $W_F(n, m|b)$  from Maddison (1990), hence the same  $W$  notation.<sup>†</sup> More precisely, for some node  $u$  in tree  $T$ ,  $W_u^\alpha(r_1, \dots, r_m|i)$  is the set of ways of having  $r_i$  events of type  $i$  (e.g., a  $\text{low} \rightarrow \text{high}$  event), for  $i \in \{1, \dots, m\}$ , given that the state (from alphabet  $\Sigma_\alpha$ ) of node  $u$  is  $i$ . That is, the  $r_1, \dots, r_m$  are just the elements of the tuple representing some multiset  $e$  of events from  $t(\Sigma_\alpha)$  in character  $\alpha$ , that is,  $m = |t(\Sigma_\alpha)|$ . The  $W_T^\alpha(r_1, \dots, r_m)$  from above is then

$$\bigcup_{i \in \Sigma_\alpha} W_u^\alpha(r_1, \dots, r_m|i), \quad (11)$$

where  $u$  is the root of tree  $T$ . Since this is computed using a bottom-up dynamic programming procedure, the base case for each leaf node  $\ell$  of tree  $T$  is

$$\begin{aligned} \forall i \in \Sigma_\alpha W_\ell^\alpha(0, \dots, 0|i) &= \{\emptyset\}, \\ \forall i \in \Sigma_\alpha \forall r_1, \dots, r_m > 0 W_\ell^\alpha(r_1, \dots, r_m|i) &= \emptyset, \end{aligned} \quad (12)$$

where  $\emptyset$  is the empty set. The recursive step for internal node  $u$  with the set  $V = \{v_1, \dots, v_n\}$  of children is then

<sup>†</sup>This is a generalization of the computation of  $W$  from Maddison (1990), to sets of ways a character on any number of states can happen in a general (multifurcating) tree, in fact.



$$W_u^\alpha(r_1, \dots, r_m|i) = \bigcup_{s \in \mathcal{S}} \bigcup_{p_1 \in P_V(r_{1'})} \dots \bigcup_{p_m \in P_V(r_{m'})} \bigcup_{\pi \in \Pi(s)} (\otimes_{v \in V} W_v^\alpha(p_1^v, \dots, p_m^v|\pi^v)) \otimes \pi', \quad (13)$$

where

- $\mathcal{S}$  is the number of multisets of size  $n$  taken from  $\Sigma_\alpha$ , that is, the number of ways to distribute events on the branches below  $u$  (note that  $i \rightarrow i$  is the nonevent);
- $r'_t = r_t - m_s(j)$ , if  $t$  represents transition (event)  $i \rightarrow j$  and  $m_s(j)$  is the *multiplicity* of element  $j$  in multiset  $s$  (o.w.,  $r'_t = r_t$ );
- $P_V(k) = \{ \{p^v | v \in V, p^v \in \mathbb{N}\} | \sum_{v \in V} p^v = k \}$  is the number of ways to partition  $k$  objects into the  $(n)$  bins  $V$ , that is, the number of ways to distribute the remaining events to the subtrees below  $u$ ;
- $\Pi(s)$  is the set of *unique* permutations of multiset  $s$ ;
- $\mathcal{S} \otimes \mathcal{T} = \{S \cup T | S \in \mathcal{S}, T \in \mathcal{T}\}$ , which is somewhat like a Cartesian product, but it returns a set of sets, rather than a set of tuples;
- $\pi^v$  is the  $v$ -th element of permutation  $\pi$ ; and
- $\pi' = \{ \{v : i \rightarrow j | v \in V, j = \pi^v \neq i\} \}$  is the set of changes  $v : i \rightarrow j$  from some state  $i$  to some other state  $j \neq i$  happening on the branch leading to node  $v$  from its parent (node  $u$ ).

Note that this recursive step at node  $u$  is just computing all possible ways to have any  $v : i \rightarrow j$  event on any combination of each branch leading to some child  $v \in V$ , for all combinations of possible ways ( $W_v^\alpha(p_1^v, \dots, p_m^v|\pi^v)$ ) events can happen in the subtree rooted at each node  $v \in V$ . At any given step, the set, for example,  $W_u^\alpha(e|i)$  of ways some multiset  $e$  of events in character  $\alpha$  can happen on tree  $T$ , given that the state of  $\alpha$  at node  $u$  is  $i$ , is then just a set of sets of these  $v : i \rightarrow j$  events. For example, if character  $\alpha$  represents weight from above, then the set  $W_X^\alpha(2, 0|\text{low})$  of ways that 2 low  $\rightarrow$  high events and 0 high  $\rightarrow$  low events can happen on the tree  $T$  of Figure 4a, given that the state of character  $\alpha$  at node  $X$  in  $T$  is low, is

$$\begin{aligned} W_X^\alpha(2, 0|\text{low}) = & \{ \{A : \text{low} \rightarrow \text{high}, B : \text{low} \rightarrow \text{high}\}, \{C : \text{low} \rightarrow \text{high}, A : \text{low} \rightarrow \text{high}\}, \\ & \{C : \text{low} \rightarrow \text{high}, B : \text{low} \rightarrow \text{high}\}, \{Z : \text{low} \rightarrow \text{high}, C : \text{low} \rightarrow \text{high}\}, \\ & \{D : \text{low} \rightarrow \text{high}, A : \text{low} \rightarrow \text{high}\}, \{D : \text{low} \rightarrow \text{high}, B : \text{low} \rightarrow \text{high}\}, \\ & \{D : \text{low} \rightarrow \text{high}, Z : \text{low} \rightarrow \text{high}\}, \{D : \text{low} \rightarrow \text{high}, C : \text{low} \rightarrow \text{high}\}, \\ & \{Y : \text{low} \rightarrow \text{high}, D : \text{low} \rightarrow \text{high}\} \end{aligned} \quad (14)$$

or  $\{\{A, B\}, \{C, A\}, \{C, B\}, \{Z, C\}, \{D, A\}, \{D, B\}, \{D, Z\}, \{D, C\}, \{Y, D\}\}$  in our shorthand devised for binary characters [see Eq. (55) of Appendix A1].

We now detail the computation of the correlation between each pair from the Cartesian product of collections  $\mathcal{C}_\alpha^{\text{low} \rightarrow \text{high}}$  and  $\mathcal{C}_\beta^{28 \rightarrow 30}$  for the example depicted in Figure 4. To compute  $\mathcal{C}_\alpha^{\text{low} \rightarrow \text{high}}$ , based on Equation (10), we first compute  $W_T^\alpha(2, 0)$ ,  $W_T^\alpha(0, 2)$ , and  $W_T^\alpha(1, 1)$  according to Equation (11)—the ways that the events of each of the respective parsimonies  $p_1, p_2$ , and  $p_3$  for character  $\alpha$  can happen on  $T$  of Figure 4a, as follows

$$\begin{aligned} W_T^\alpha(2, 0) &= W_X(2, 0|\text{low}) \cup W_X(2, 0|\text{high}) = \\ & \{ \{A, B\}, \{C, A\}, \{C, B\}, \{Z, C\}, \{D, A\}, \{D, B\}, \{D, Z\}, \{D, C\}, \{Y, D\} \} \end{aligned} \quad (15)$$

$$\begin{aligned} W_T^\alpha(0, 2) &= W_X(0, 2|\text{low}) \cup W_X(0, 2|\text{high}) = \\ & \{ \{a, b\}, \{c, b\}, \{c, a\}, \{z, c\}, \{d, b\}, \{d, a\}, \{d, c\}, \{d, z\}, \{y, d\} \} \end{aligned} \quad (16)$$

$$\begin{aligned} W_T^\alpha(1, 1) &= W_X(1, 1|\text{low}) \cup W_X(1, 1|\text{high}) = \\ & \{ \{Z, b\}, \{Z, a\}, \{Y, b\}, \{Y, a\}, \{Y, c\}, \{Y, z\}, \\ & \{z, A\}, \{z, B\}, \{y, A\}, \{y, B\}, \{y, Z\}, \{y, C\} \} \end{aligned} \quad (17)$$

Note that the details of computing  $W_X^\alpha(2, 0|\text{low})$  [Eq. (55)],  $W_X^\alpha(2, 0|\text{high})$  [Eq. (58)],  $W_X^\alpha(0, 2|\text{low})$  [Eq. (53)],  $W_X^\alpha(0, 2|\text{high})$  [Eq. (56)],  $W_X^\alpha(1, 1|\text{low})$  [Eq. (54)], and  $W_X^\alpha(1, 1|\text{high})$  [Eq. (57)] are found in

Appendix A1. Collection  $\mathcal{C}_\alpha^{\text{low} \rightarrow \text{high}}$  is then the  $(9 \cdot 9 \cdot 12 = 972)$  sets resulting from the Cartesian product of  $W_T^\alpha(2, 0)$ ,  $W_T^\alpha(0, 2)$ , and  $W_T^\alpha(1, 1)$  [from Eqs. (15)–(17), respectively] computed according to Equation (10). The first two elements and the last element of  $\mathcal{C}_\alpha^{\text{low} \rightarrow \text{high}}$  are as follows

$$\{A, B\}|_{\text{low} \rightarrow \text{high}} \cup \{a, b\}|_{\text{low} \rightarrow \text{high}} \cup \{Z, b\}|_{\text{low} \rightarrow \text{high}} = \{A, B, Z\} \quad (18)$$

$$\{A, B\}|_{\text{low} \rightarrow \text{high}} \cup \{a, b\}|_{\text{low} \rightarrow \text{high}} \cup \{Z, a\}|_{\text{low} \rightarrow \text{high}} = \{A, B, Z\} \quad (19)$$

⋮

$$\{Y, D\}|_{\text{low} \rightarrow \text{high}} \cup \{y, d\}|_{\text{low} \rightarrow \text{high}} \cup \{y, C\}|_{\text{low} \rightarrow \text{high}} = \{Y, D, C\} \quad (20)$$

To compute  $\mathcal{C}_\beta^{28 \rightarrow 30}$ , we first compute  $W_T^\beta(1, 0)$  and  $W_T^\beta(0, 1)$  according to Equation (11)—the ways that the events of each of the respective parsimonies  $q_1$  and  $q_2$  for character  $\beta$  can happen on  $T$  of Figure 4a as follows

$$W_T^\beta(1, 0) = W_X^\beta(1, 0|28) \cup W_X^\beta(1, 0|30) = \{\{B\}, \{A\}, \{C\}, \{Z\}, \{D\}, \{Y\}\} \quad (21)$$

$$W_T^\beta(0, 1) = W_X^\beta(0, 1|28) \cup W_X^\beta(0, 1|30) = \{\{a\}, \{b\}, \{z\}, \{c\}, \{y\}, \{d\}\} \quad (22)$$

Note that the details of computing  $W_X^\beta(1, 0|28)$  [Eq. (72)],  $W_X^\beta(1, 0|30)$  [Eq. (74)],  $W_X^\beta(0, 1|28)$  [Eq. (71)], and  $W_X^\beta(0, 1|30)$  [Eq. (73)] are found in Appendix A1. Collection  $\mathcal{C}_\beta^{28 \rightarrow 30}$  is then the  $(6 \cdot 6 = 36)$  sets resulting from the Cartesian product of  $W_T^\beta(1, 0)$  and  $W_T^\beta(0, 1)$  [from Eqs. (21) and (22), respectively] computed according to Equation (10). The first two elements and the last element of  $\mathcal{C}_\beta^{28 \rightarrow 30}$  are as follows

$$\{B\}|_{28 \rightarrow 30} \cup \{a\}|_{28 \rightarrow 30} = \{B\} \quad (23)$$

$$\{B\}|_{28 \rightarrow 30} \cup \{b\}|_{28 \rightarrow 30} = \{B\} \quad (24)$$

⋮

$$\{Y\}|_{28 \rightarrow 30} \cup \{d\}|_{28 \rightarrow 30} = \{Y\} \quad (25)$$

Finally, the set of correlations (of size  $972 \cdot 36 = 34992$ ) obtained by computing the correlation [according to Eq. (6)] of each pair from the Cartesian product of the collections  $\mathcal{C}_\alpha^{\text{low} \rightarrow \text{high}}$  and  $\mathcal{C}_\beta^{28 \rightarrow 30}$  is computed. The first two correlations and last correlation of this set are as follows

$$\frac{|\{A, B, Z\} \cap \{B\}|}{|\{A, B, Z\} \cup \{B\}|} = \frac{|\{B\}|}{|\{A, B, C\}|} = \frac{1}{3} \quad (26)$$

$$\frac{|\{A, B, Z\} \cap \{B\}|}{|\{A, B, Z\} \cup \{B\}|} = \frac{|\{B\}|}{|\{A, B, C\}|} = \frac{1}{3} \quad (27)$$

⋮

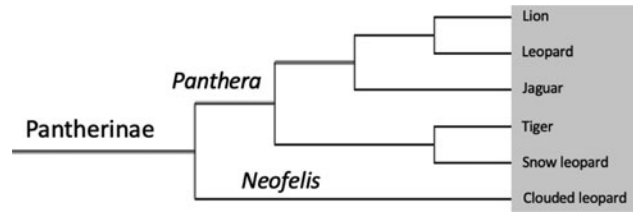
$$\frac{|\{Y, D, C\} \cap \{Y\}|}{|\{Y, D, C\} \cup \{Y\}|} = \frac{|\{Y\}|}{|\{Y, D, C\}|} = \frac{1}{3} \quad (28)$$

Of these 34,992 correlations, 16,092 of them are  $1/3$ , and 18,900 of them are 0. This distribution has a mean of  $\mu = 0.153$  and a standard deviation of  $\sigma = 0.166$ . It follows that the probability of obtaining a correlation as extreme as  $1/3$  in this (null) distribution of correlation is 0.46 ( $p$ -value), and  $1/3$  is  $1.084\sigma$  from the mean ( $z$ -score), hence this correlation is far from being significant by standard conventions ( $p < 0.05$  or  $z \geq 2$ ).

## 5. EXPERIMENTAL ANALYSIS

To test our approach, we used a set of vocalization and morphological data from the subfamily Pantherinae, commonly known as the big cats. We chose this clade in our analyses because their vocalization

**FIG. 5.** The subfamily Pantherinae consists of six species and two genera: *Panthera* (lions, leopards, jaguars, tigers, and snow leopards) and *Neofelis* (clouded leopards). Phylogeny adapted from Johnson et al. (2006), Li et al. (2016), Figueiro et al. (2017).



data are best documented among Felidae. Pantherinae consists of six species and two genera: *Panthera* (lions, leopards, jaguars, tigers, and snow leopards) and *Neofelis* (clouded leopards) (Fig. 5). The set of extant states are for 10 of the vocalizations documented in Sunquist and Sunquist (2002) and 9 morphological characteristics compiled from the various sources within Castelló (2020) (Table 1). The vocalizations had states {absent, present, unknown}, while the morphological characteristics were binned into ranks starting from 1 (largest, or longest, etc.), to 2, to 3, etc. Finally, the cost for changing to a different state is 1 and to an *unknown* state is  $\infty$  (e.g., Fig. 1c). Since unknown states are artifacts of the collection process, by assigning a high evolutionary cost to the changes from any unknown to some known state (which makes little evolutionary sense), we mitigate the propagation of any unknown state to ancestral nodes in any parsimony, whenever some known state (e.g., low or high) can explain the data. For example, in the instance of Figure 1, if the weight was instead unknown in the Jaguarundi and Pallas cat, then it would only have the unique parsimony where all ancestors X, Y, and Z have the high state. This is why we use the approach of Sankoff (1975) [instead of, e.g., (Fitch, 1971)], because we need these transition-specific costs.

Our approach and resulting implementation into the software tool *parcours* begins with the efficient computation of all parsimonies described in Section 2, followed by the computation of correlation from these parsimonies, as described in Section 3. Finally, for each of these correlations, the probability of obtaining a correlation as extreme (or more) by chance is computed according to Section 4. Since *parcours* starts with the same input as Sankoff's algorithm (e.g., Fig. 1), we input the phylogenetic tree and extant states of the 19 characters and the Pantherinae phylogeny mentioned above. The cost matrix is computed automatically by *parcours*, this being the default, when no specific cost matrix is provided. From this, *parcours* returns the correlations of all possible events between all pairs of characters according to Equation (6) and then computes the probability obtaining a correlation this extreme by chance according to Section 4.

## 6. RESULTS AND DISCUSSION

We used two measures of significance. In a normal distribution, a standard  $p$  value of 0.05 is equivalent to a  $z$  score of two or more. However, because the distributions here are computed from a (large, yet finite) number of ways some sets of events may occur on a tree, they are discrete, not normal distributions. As

TABLE 1. VOCALIZATIONS AND MORPHOLOGICAL CHARACTERISTICS USED IN THE EXPERIMENT

<i>Vocalizations</i>	<i>Morphological characteristics</i>
Spit	Gestation time
Hiss	SW
Growl	SL
Snarl	BL
Prusten	Dental profile
Puff	Weight
Mew	SL/SW ratio
Main call	BL/SL ratio
Roaring sequence	Chromosome number
Grunt	

BL, body length; SL, skull length; SW, skull width.

TABLE 2. OF THE 61 PAIRS OF EVENTS WITH NONZERO CORRELATION, THE 30 WHICH ARE SIGNIFICANTLY CORRELATED, AS JUDGED BY A  $P$ -VALUE  $<0.05$  OR A  $Z$ -SCORE  $\geq 2$ 

	$c_1$	$c_2$	$t_1$	$t_2$	Correlation	$\mu$	$\sigma$	$p$	$z$ -score
1	Puff	Prusten	Absent $\rightarrow$ present	Present $\rightarrow$ absent	1.000	0.100	0.300	0.100	3.000
2	Grunt	Roaring-seq	Absent $\rightarrow$ present	Absent $\rightarrow$ present	1.000	0.100	0.300	0.100	3.000
3	BL	Main-call	2 $\rightarrow$ 1	Present $\rightarrow$ absent	0.500	0.095	0.196	0.189	2.069
4	Dental	BL	1 $\rightarrow$ 2	2 $\rightarrow$ 3	1.000	0.100	0.300	0.100	3.000
5	Dental	BL	2 $\rightarrow$ 1	3 $\rightarrow$ 2	1.000	0.100	0.300	0.100	3.000
6	Weight	SL	1 $\rightarrow$ 2	1 $\rightarrow$ 2	0.667	0.253	0.145	0.031*	2.846
7	Weight	BL	1 $\rightarrow$ 3	2 $\rightarrow$ 3	1.000	0.117	0.322	0.117	2.744
8	Weight	Dental	1 $\rightarrow$ 3	1 $\rightarrow$ 2	1.000	0.100	0.300	0.100	3.000
9	Weight	Dental	3 $\rightarrow$ 1	2 $\rightarrow$ 1	1.000	0.100	0.300	0.100	3.000
10	SL-SW	Main-call	4 $\rightarrow$ 1	Present $\rightarrow$ absent	0.500	0.095	0.196	0.189	2.069
11	SL-SW	BL	4 $\rightarrow$ 1	2 $\rightarrow$ 1	1.000	0.147	0.210	0.025*	4.055
12	SL-SW	BL	4 $\rightarrow$ 3	2 $\rightarrow$ 3	1.000	0.112	0.316	0.112	2.811
13	SL-SW	Dental	3 $\rightarrow$ 4	2 $\rightarrow$ 1	1.000	0.100	0.300	0.100	3.000
14	SL-SW	Dental	4 $\rightarrow$ 3	1 $\rightarrow$ 2	1.000	0.100	0.300	0.100	3.000
15	SL-SW	Weight	4 $\rightarrow$ 3	1 $\rightarrow$ 3	1.000	0.117	0.322	0.117	2.744
16	BL-SL	Roaring-seq	3 $\rightarrow$ 1	Absent $\rightarrow$ present	0.500	0.086	0.189	0.172	2.193
17	BL-SL	Roaring-seq	4 $\rightarrow$ 1	Absent $\rightarrow$ present	0.500	0.090	0.192	0.179	2.138
18	BL-SL	Grunt	3 $\rightarrow$ 1	Absent $\rightarrow$ present	0.500	0.086	0.189	0.172	2.193
19	BL-SL	Grunt	4 $\rightarrow$ 1	Absent $\rightarrow$ present	0.500	0.090	0.192	0.179	2.138
20	BL-SL	SL	4 $\rightarrow$ 1	2 $\rightarrow$ 1	0.333	0.116	0.106	0.127	2.056
21	BL-SL	BL	1 $\rightarrow$ 4	2 $\rightarrow$ 3	1.000	0.117	0.322	0.117	2.744
22	BL-SL	BL	3 $\rightarrow$ 4	2 $\rightarrow$ 3	1.000	0.125	0.331	0.125	2.643
23	BL-SL	Dental	1 $\rightarrow$ 4	1 $\rightarrow$ 2	1.000	0.100	0.300	0.100	3.000
24	BL-SL	Dental	3 $\rightarrow$ 4	1 $\rightarrow$ 2	1.000	0.100	0.300	0.100	3.000
25	BL-SL	Dental	4 $\rightarrow$ 1	2 $\rightarrow$ 1	0.500	0.090	0.192	0.179	2.138
26	BL-SL	Dental	4 $\rightarrow$ 3	2 $\rightarrow$ 1	0.500	0.100	0.200	0.199	2.005
27	BL-SL	Weight	1 $\rightarrow$ 4	1 $\rightarrow$ 3	1.000	0.126	0.332	0.126	2.633
28	BL-SL	Weight	3 $\rightarrow$ 4	1 $\rightarrow$ 3	1.000	0.133	0.340	0.133	2.552
29	BL-SL	SL-SW	1 $\rightarrow$ 4	4 $\rightarrow$ 3	1.000	0.117	0.322	0.117	2.744
30	BL-SL	SL-SW	3 $\rightarrow$ 4	4 $\rightarrow$ 3	1.000	0.125	0.331	0.125	2.643

Each row is correlation between a pair  $t_1$  and  $t_2$  of events in characters  $c_1$  and  $c_2$ , respectively. The  $\mu$  and  $\sigma$  are the mean and standard deviation of the corresponding null distribution, respectively. The  $p$ -value is the probability of obtaining a correlation this extreme in the null distribution, while the  $z$ -score is the number of standard deviations ( $\sigma$ ) this probability ( $p$ ) is from the mean ( $\mu$ ).

such, we defined significance using either  $p < 0.05$  and/or  $z$  score of two or more. Table 2 reports all correlations found by parcours where either  $p < 0.05$  or  $z \geq 2$ . We examined all such significant co-occurrences to assess if they were consistent with known biological data as a way to evaluate parcours.

Out of the 30 significant correlations detected by parcours, 12 were between traits associated with overall body size such as weight and body length (BL). This is not surprising given that many features like body weight and BL have allometric relationships (Calder, 1984). Significant co-occurrences within Pantherinae include a decrease in both weight and skull length (SL) ( $p=0.03$ ) and an increase in both the SL/skull width (SL/SW) ratio and BL ( $p=0.025$ ). For example, these significant co-occurrences were detected in the lineage leading to the African leopard and African lion, respectively. These findings are consistent with the observation that (1) African leopards are smaller in both weight and SL than either of its sister taxa, the lions and jaguars, and (2) lions have the largest BL and SL/SW ratio of any member of Pantherinae.

Prusten and puff are two types of friendly close-range vocalizations. Felid species have only one type of friendly call in their repertoire as the prusten and puff are functionally equivalent (Sunquist and Sunquist, 2002). Prusten is found in all members of Pantherinae except for lions and leopards which instead have a puff (Peters and Tonkin-Leyhausen, 1999). A significant co-occurrence of a loss of prusten and a gain of puff ( $z=3.0$ ) along the lineage leading to lions and leopards is consistent with the observation that prusten is most likely the ancestral friendly call type in Pantherinae which was replaced by a puff in lions and leopards.

The dental formula (DF) is defined as the total number of each type of tooth in both the upper and lower jaw. Within Pantherinae, all species have a DF of 30 with the exception of the Indochinese Clouded Leopard (*Neofelis nebulosa*), which has a DF of 28–30. Clouded leopards are the smallest member of Pantherinae in terms of weight and overall BL. Weighing an average of 17 kg and 88 cm in length, the clouded leopards are mixed prey feeders, adapted to taking both large and small prey. All other members of Pantherinae are large prey feeders with relatively wide muzzles (Meachen-Samuels and Valkenburgh, 2009). This is consistent with the observation that a decrease in SL/SW ratio co-occurs with an increase in DF and vice versa ( $z=3.0$ ). The direct co-occurrence of BL with DF and weight with DF ( $z=3.0$ ) is also consistent with the fact that the clouded leopard is the smallest felid in Pantherinae and the only felid with a DF <30.

A roaring sequence is characterized by one to two soft moans followed by full throated roars and a series of grunts of diminishing amplitude (Stander and Stander, 1988). There was a significant co-occurrence between an increase in BL/SL ratio from a medium sized to a large bodied *Panthera* species and the gain of this call along the lineage leading to lions, leopards, and jaguars ( $z=2.2$ ). This is consistent with the fact that the roaring sequence is known only in those felid species. While grunts are most often used as part of a roaring sequence, grunts can also be used alone in certain situations such as females calling for cubs (Sunquist and Sunquist, 2002). Not surprisingly, the gain of a grunt also significantly co-occurs with an increase in BL/SL ratio along the same lineage ( $z=2.2$ ). Finally, there is significant association between the gain of a grunt and a gain in the roaring sequence ( $z=3.0$ ), also along this same lineage. This is consistent with the observation that the grunt is an integral part of the roaring sequence which supports their non-random co-occurrence.

The main call is defined as a high intensity form of a mew. Mews are the predominant vocalization of felids but have low to medium intensities. Main calls are found in all felids except for lions (Sunquist and Sunquist, 2002). We detected a significant co-occurrence in the increase in both SL/SW ratio and BL with the loss of a main call ( $z=2.1$ ) along the lineage leading to lions. This is consistent with main calls being absent in lions and African lions being the largest felid within Pantherinae in terms of BL and having the highest SL/SW ratio of all felids in this subfamily.

## 7. CONCLUSION

Understanding the evolution of complex traits such as vocalization and morphological characters remains a key challenge in evolutionary studies. In this work, we propose a new approach to detecting significantly correlated evolution among discrete characters in a phylogenetic tree by leveraging the information from *all* parsimonies—its implementation, parcours, efficiently automating this process in an easy-to-use and customizable way. We demonstrate its use on a set of vocalizations and morphological characteristics in the Pantherinae subfamily.

While parcours already discovers trends in these data which are supported in the literature, possible refinements to the ranking of characteristics (such as dental profile), or the cost matrix (e.g., making  $4 \rightarrow 1$  more expensive than  $2 \rightarrow 1$ ), could even further improve these results. While we used the weighted Jaccard index as correlation here, a full exploration of similarity indexes (Warrens, 2016) is the subject of future work. Here, we consider the set of all parsimonies in this framework. However, another idea is to allow the user to place constraints on this set, based on a priori knowledge, such as certain ancestral states, or the least common ancestor of a given set of states. To assess significant correlation, we compute all possible ways some sets of events can happen on a tree. However, this can result in runtime issues, something which was encountered also in the method (Maddison, 1990) which inspired the design of our approach.

A potential solution could be to subsample such set of possible ways, allowing for a faster approximation of significant correlation. Finally, a future experimental direction is to reconstruct even more detailed information about felid vocalizations (e.g., mean frequency, harmonics) from audio recordings. We have already devised a machine learning framework for automatically extracting many such acoustic features from sound recordings and have some interesting preliminary results on the Pantherinae subfamily. Using parcours to then automate the search for significant correlations among the many features extracted, along with morphological characters (e.g., mean frequency and body size), will provide valuable insight into the evolution of their vocalizations in the context of morphology and perhaps even habitat.

## AUTHORS' CONTRIBUTIONS

S.A.B., M.P.: Conceptualization. M.P.: Methodology. K.K., C.N.M., M.P.: Software. K.K., S.A.B., M.P.: Validation. M.P.: Formal Analysis. S.A.B., M.P.: Investigation. C.N.M., B.S., R.B., S.A.B.: Resources. C.N.M., B.S., R.B., S.A.B.: Data curation. All: Writing—Original Draft. All: Writing—Review and Editing. K.K., C.N.M.: Visualization. S.A.B., M.P.: Supervision. S.A.B., M.P.: Project Administration. S.A.B., M.P.: Funding Acquisition.

## ACKNOWLEDGMENTS

The authors thank Steven Patterson for insightful discussions about the statistical aspects of this work.

## AUTHOR DISCLOSURE STATEMENT

The authors declare they have no conflicting financial interests.

## FUNDING INFORMATION

Research supported by Fairfield University Science Institute and Fredrickson Family Innovation Lab grants for S.A.B. and M.P.; Georgia State University startup grant for M.P.

## REFERENCES

- Barker, D., Meade, A., and Pagel, M. 2007. Constrained models of evolution lead to improved prediction of functional linkage from correlated gain and loss of genes. *Bioinformatics* 23, 14–20.
- Calder, W. 1984. *Size, Function and Life History*. Harvard University Press, Cambridge Massachusetts.
- Camin, J., and Sokal, R. 1965. A method for deducing branching sequences in phylogeny. *Evolution* 19, 311–326.
- Castelló, J. 2020. *Felids and Hyenas of the World*. Princeton University Press, Princeton, New Jersey, USA.
- Ciccolella, S., Ricketts, C., Soto Gomez, M., et al. 2020a. Inferring cancer progression from Single-Cell Sequencing while allowing mutation losses. *Bioinformatics*. [Epub ahead of print]; DOI: 10.1093/bioinformatics/btaa722.
- Ciccolella, S., Soto Gomez, M., Patterson, M.D., et al. 2020b. Gpps: An ILP-based approach for inferring cancer progression with mutation losses from single cell data. *BMC Bioinformatics* 21, 413.
- Clemente, J., Ikeo, K., Valiente, G., et al. 2009. Optimized ancestral state reconstruction using Sankoff parsimony. *BMC Bioinformatics* 10. [Epub ahead of print]; DOI: 10.1186/1471-2105-10-51.
- Cohen, O., Ashkenazy, H., Karin, E.L., et al. 2013. CoPAP: Coevolution of presence-absence patterns. *Nucleic Acids Res.* 41, 232–237.
- Drummond, A., and Rambaut A. 2007. BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evol. Biol.* 7, 1–8.
- Duchemin, W., Anselmetti, Y., Patterson, M., et al. 2017. DeCoSTAR: Reconstructing the ancestral organization of genes or genomes using reconciled phylogenies. *Genome Biol. Evol.* 9, 1312–1319.
- Edwards, A., and Cavalli-Sforza, L. 1963. The reconstruction of evolution. *Ann. Hum. Genet.* 27, 105–106.
- Ehret, G. 1980. Development of sound communication in mammals (179–225). In Rosenblatt, S., Hindle, R.A., Beer, C., and Busnel, M.-C. (eds): *Advances in the Study of Animal Behavior*. Academic Press, Cambridge, Massachusetts, USA.
- El-Kebir, M. 2018. SPhyR: Tumor phylogeny estimation from single-cell sequencing data under loss and error. *Bioinformatics* 34, i671–i679.
- Farris, J.S. 1970. Methods for computing Wagner trees. *Syst. Biol.* 19, 83–92.
- Felsenstein, J. 1978. Cases in which parsimony or compatibility methods will be positively misleading. *Syst. Zool.* 27, 401–410.
- Felsenstein, J. 1981. Evolutionary trees from DNA sequences: A maximum likelihood approach. *J. Mol. Evol.* 17, 368–376.
- Felsenstein, J. 1989. PHYLIP—Phylogeny Inference Package (Version 3.2). *Cladistics* 5, 164–166.
- Figueiro, H.V., Li, G., Trindade, F.J., et al. 2017. Genome-wide signatures of complex introgression and adaptive evolution in the big cats. *Sci. Adv.* 3.
- Fitch, W.M. 1971. Toward defining the course of evolution: minimum change for a specific tree topology. *Syst. Zool.* 20, 406–416.
- Guéguen, L., Gaillard, S., Gouy, B.B.M., et al. 2013. Bio++: Efficient extensible libraries and tools for computational molecular evolution. *Mol. Biol. Evol.* 30, 745–1750.

- Hajirasouliha, I., and Raphael, B.J. 2014. Reconstructing mutational history in multiply sampled tumors using perfect phylogeny mixtures (354–367). In Brown, D., and Morgenstern, B. (eds): *Algorithms in Bioinformatics*. Springer-Verlag, Berlin-Heidelberg, Germany.
- Huelsenbeck, J.P., and Ronquist, F. 2001. MRBAYES: Bayesian inference of phylogeny. *Bioinformatics* 17, 754–755.
- Jaccard, P. 1908. Nouvelles Recherches Sur La Distribution Florale. *Bulletin de la Société Vaudoise des Sciences Naturelles*. (New Research on the Floral Distribution. *Bulletin of the Vaudoise Society of the Natural Sciences*.) 44, 223–270.
- Jahn, K., Kuipers, J., and Beerenwinkel, N. 2016. Tree inference for single-cell data. *Genome Biol.* 17, 86.
- Johnson, W.E., Eizirik, E., Pecon-Slaterry, J., et al. 2006. The late Miocene radiation of modern Felidae: A genetic assessment. *Science* 311, 73–77.
- Kolaczowski, B., and Thornton, J. 2004. Performance of maximum parsimony and likelihood phylogenetics when evolution is heterogeneous. *Nature* 431, 980–984.
- Li, G., Davis BW, Eizirik E, et al. 2016. Phylogenomic evidence for ancient hybridization in the genomes of living cats (Felidae). *Genome Res.* 26, 1–11.
- Maddison, W.P. 1990. A method for testing the correlated evolution of two binary characters. *Evolution* 44, 539–557.
- Marcotte, E.M., Xenarios, I., Blik, A.M. van der, et al. 2000. Localizing proteins in the cell from their phylogenetic profiles. *Proc Natl Acad Sci USA* 97, 12115–12120.
- Meachen-Samuels, J., and Valkenburgh, B.V. 2009. Craniodental indicators of prey size preference in the Felidae. *Biol. J. Linn. Soc.* 96, 784–799.
- Mossige, S. 1977. An algorithm for gray codes. *Computing* 18, 89–92.
- Patterson, M., Bernard, T., and Kahn, D. 2016. Correlated evolution of metabolic functions over the tree of life. *bioRxiv* 093591. [Epub ahead of print]; DOI: 10.1101/093591.
- Patterson, M., Szöllősi, G., Daubin, V., et al. 2013. Lateral gene transfer, rearrangement, reconciliation. *BMC Bioinformatics* 14. [Epub ahead of print]; DOI: 10.1186/1471-2105-14-S15-S4.
- Peters, G. 1978. Vergleichende Untersuchung Zur Lautgebung Einiger Feliden (Mammalia, Felidae). (Comparative Study of the Vocalization of Some Felids.) *Spixiana* 1, 1–283.
- Peters, G. 1991. Vocal communication in cats (76–77). Seidensticker, J., and Lumpkin, S. (eds): *Great Cats—Majestic Creatures of the Wild*. Rodale Press, Emmaus, Pennsylvania, USA.
- Peters, G., and Tonkin-Leyhausen, B.A. 1999. Evolution of acoustic communication signals of mammals: Friendly close-range vocalizations in Felidae (Carnivora). *J. Mammal. Evol.* 6, 129–159.
- Romand, R., and Ehret, G. 1984. Development of sound production in normal, isolated and deafened kittens during the first postnatal months. *Dev. Psychobiol.* 17, 629–649.
- Ružička, M. 1958. Anwendung Mathematisch-Statistischer Methoden in Der Geobotanik (Synthetische Bearbeitung von Aufnahmen). [Application of Mathematical-Statistical Methods in Geobotany (Synthetic Processing of Recordings).] *Biología* 13, 647–661.
- Sankoff, D. 1975. Minimal mutation trees of sequences. *SIAM J. Appl. Math.* 28, 35–42.
- Sankoff, D., and Rousseau, P. 1975. Locating the vertices of a Steiner tree in an arbitrary metric space. *Math. Programm.* 9, 240–246.
- Schwartz, R., and Schäffer, A. 2017. The evolution of tumour phylogenetics: Principles and practice. *Nat. Rev. Genet.* 18, 213–229.
- Stamatakis, A. 2006. RAxML-VI-HPC: Maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* 22, 2688–2690.
- Stander, P., and Stander, J. 1988. Characteristics of lion roars in Etosha National Park. *Madoqua* 15, 315–318.
- Steel, M., and Penny, D. 2004. Two further links between MP and ML under the Poisson model. *Appl. Math. Lett.* 17, 785–790.
- Sunquist, M., and Sunquist, F. 2002. *Wild Cats of the World*. University of Chicago Press, Chicago and London (452 pgs).
- Tuffley, C., and Steel, M. 1997. Links between maximum likelihood and maximum Parsimony under a simple model of site substitution. *Bull. Math. Biol.* 59, 581–607.
- Warrens, M.J. 2016. Inequalities between similarities for numerical data. *J. Classif.* 33, 141–148.

Address correspondence to:

Dr. Murray Patterson  
Department of Computer Science  
Georgia State University  
Atlanta, GA 30303  
USA

E-mail: mpatterson30@gsu.edu

(Appendix follows →)

# Appendix

## APPENDIX A1

*Ways that two events in character  $\alpha$  can happen in a tree*

Given character  $\alpha$  from Figure 4b ( $\Sigma_\alpha = \{\text{low}, \text{high}\}$ ), we detail how to compute all possible ways that two events, from  $t(\Sigma_\alpha) = \{\text{low} \rightarrow \text{high}, \text{high} \rightarrow \text{low}\}$ , can happen on tree  $T$  of Figure 4a, given either state low or high at the root ( $X$ ) of the tree  $T$ , that is,  $W_X^\alpha(0, 2|\text{low})$  [Eq. (53)],  $W_X^\alpha(1, 1|\text{low})$  [Eq. (54)],  $W_X^\alpha(2, 0|\text{low})$  [Eq. (55)],  $W_X^\alpha(0, 2|\text{high})$  [Eq. (56)],  $W_X^\alpha(1, 1|\text{high})$  [Eq. (57)], and  $W_X^\alpha(2, 0|\text{high})$  [Eq. (58)]. The base case(s) for the leaf nodes  $A, B, C$ , and  $D$  of tree  $T$  is set according to Equation (12) as follows

$$\begin{aligned} W_A^\alpha(0, 0|\text{low}) &= W_A^\alpha(0, 0|\text{high}) = \{\emptyset\} \\ W_B^\alpha(0, 0|\text{low}) &= W_B^\alpha(0, 0|\text{high}) = \{\emptyset\} \\ W_C^\alpha(0, 0|\text{low}) &= W_C^\alpha(0, 0|\text{high}) = \{\emptyset\} \\ W_D^\alpha(0, 0|\text{low}) &= W_D^\alpha(0, 0|\text{high}) = \{\emptyset\} \\ \forall r_1, r_2 > 0 \quad W_A^\alpha(r_1, r_2|\text{low}) &= W_A^\alpha(r_1, r_2|\text{high}) = \emptyset \\ \forall r_1, r_2 > 0 \quad W_B^\alpha(r_1, r_2|\text{low}) &= W_B^\alpha(r_1, r_2|\text{high}) = \emptyset \\ \forall r_1, r_2 > 0 \quad W_C^\alpha(r_1, r_2|\text{low}) &= W_C^\alpha(r_1, r_2|\text{high}) = \emptyset \\ \forall r_1, r_2 > 0 \quad W_D^\alpha(r_1, r_2|\text{low}) &= W_D^\alpha(r_1, r_2|\text{high}) = \emptyset \end{aligned}$$

The recursive step(s) for each internal node  $Z, Y$  and finally  $X$  of tree  $T$  is then computed according to Equation (13) as follows

$$\begin{aligned} W_Z^\alpha(0, 0|\text{low}) &= \\ &W_A^\alpha(0, 0|\text{low}) \otimes W_B^\alpha(0, 0|\text{low}) \otimes \{\emptyset\} \\ &= \{\emptyset\} \end{aligned} \tag{29}$$

$$\begin{aligned} W_Z^\alpha(0, 1|\text{low}) &= \\ &W_A^\alpha(0, 0|\text{low}) \otimes W_B^\alpha(0, 1|\text{low}) \otimes \{\emptyset\} \\ &\cup W_A^\alpha(0, 1|\text{low}) \otimes W_B^\alpha(0, 0|\text{low}) \otimes \{\emptyset\} \\ &= \emptyset \end{aligned} \tag{30}$$

$$\begin{aligned} W_Z^\alpha(0, 2|\text{low}) &= \\ &W_A^\alpha(0, 0|\text{low}) \otimes W_B^\alpha(0, 2|\text{low}) \otimes \{\emptyset\} \\ &\cup W_A^\alpha(0, 1|\text{low}) \otimes W_B^\alpha(0, 1|\text{low}) \otimes \{\emptyset\} \\ &\cup W_A^\alpha(0, 2|\text{low}) \otimes W_B^\alpha(0, 0|\text{low}) \otimes \{\emptyset\} \\ &= \emptyset \end{aligned} \tag{31}$$

$$\begin{aligned} W_Z^\alpha(1, 0|\text{low}) &= \\ &W_A^\alpha(0, 0|\text{low}) \otimes W_B^\alpha(1, 0|\text{low}) \otimes \{\emptyset\} \\ &\cup W_A^\alpha(1, 0|\text{low}) \otimes W_B^\alpha(0, 0|\text{low}) \otimes \{\emptyset\} \\ &\cup W_A^\alpha(0, 0|\text{high}) \otimes W_B^\alpha(0, 0|\text{low}) \otimes \{\{A\}\} \\ &\cup W_A^\alpha(0, 0|\text{low}) \otimes W_B^\alpha(0, 0|\text{high}) \otimes \{\{B\}\} \\ &= \{\{A\}\}, \{\{B\}\} \end{aligned} \tag{32}$$



$$\begin{aligned}
W_Z^\alpha(1, 1|\text{low}) = & \\
& W_A^\alpha(0, 0|\text{low}) \otimes W_B^\alpha(1, 1|\text{low}) \otimes \{\emptyset\} \\
& \cup W_A^\alpha(0, 1|\text{low}) \otimes W_B^\alpha(1, 0|\text{low}) \otimes \{\emptyset\} \\
& \cup W_A^\alpha(1, 0|\text{low}) \otimes W_B^\alpha(0, 1|\text{low}) \otimes \{\emptyset\} \\
& \cup W_A^\alpha(1, 1|\text{low}) \otimes W_B^\alpha(0, 0|\text{low}) \otimes \{\emptyset\} \\
& \cup W_A^\alpha(0, 0|\text{high}) \otimes W_B^\alpha(0, 1|\text{low}) \otimes \{\{A\}\} \\
& \cup W_A^\alpha(0, 0|\text{low}) \otimes W_B^\alpha(0, 1|\text{high}) \otimes \{\{B\}\} \\
& \cup W_A^\alpha(0, 1|\text{high}) \otimes W_B^\alpha(0, 0|\text{low}) \otimes \{\{A\}\} \\
& \cup W_A^\alpha(0, 1|\text{low}) \otimes W_B^\alpha(0, 0|\text{high}) \otimes \{\{B\}\} \\
& = \emptyset
\end{aligned} \tag{33}$$

$$\begin{aligned}
W_Z^\alpha(2, 0|\text{low}) = & \\
& W_A^\alpha(0, 0|\text{low}) \otimes W_B^\alpha(2, 0|\text{low}) \otimes \{\emptyset\} \\
& \cup W_A^\alpha(1, 0|\text{low}) \otimes W_B^\alpha(1, 0|\text{low}) \otimes \{\emptyset\} \\
& \cup W_A^\alpha(2, 0|\text{low}) \otimes W_B^\alpha(0, 0|\text{low}) \otimes \{\emptyset\} \\
& \cup W_A^\alpha(0, 0|\text{high}) \otimes W_B^\alpha(1, 0|\text{low}) \otimes \{\{A\}\} \\
& \cup W_A^\alpha(0, 0|\text{low}) \otimes W_B^\alpha(1, 0|\text{high}) \otimes \{\{B\}\} \\
& \cup W_A^\alpha(1, 0|\text{high}) \otimes W_B^\alpha(0, 0|\text{low}) \otimes \{\{A\}\} \\
& \cup W_A^\alpha(1, 0|\text{low}) \otimes W_B^\alpha(0, 0|\text{high}) \otimes \{\{B\}\} \\
& \cup W_A^\alpha(0, 0|\text{high}) \otimes W_B^\alpha(0, 0|\text{high}) \otimes \{\{A, B\}\} \\
& = \{\{A, B\}\}
\end{aligned} \tag{34}$$

$$\begin{aligned}
W_Z^\alpha(0, 0|\text{high}) = & \\
& W_A^\alpha(0, 0|\text{high}) \otimes W_B^\alpha(0, 0|\text{high}) \otimes \{\emptyset\} \\
& = \{\emptyset\}
\end{aligned} \tag{35}$$

$$\begin{aligned}
W_Z^\alpha(0, 1|\text{high}) = & \\
& W_A^\alpha(0, 0|\text{high}) \otimes W_B^\alpha(0, 1|\text{high}) \otimes \{\emptyset\} \\
& \cup W_A^\alpha(0, 1|\text{high}) \otimes W_B^\alpha(0, 0|\text{high}) \otimes \{\emptyset\} \\
& \cup W_A^\alpha(0, 0|\text{high}) \otimes W_B^\alpha(0, 0|\text{low}) \otimes \{\{b\}\} \\
& \cup W_A^\alpha(0, 0|\text{low}) \otimes W_B^\alpha(0, 0|\text{high}) \otimes \{\{a\}\} \\
& = \{\{b\}\}, \{\{a\}\}
\end{aligned} \tag{36}$$

$$\begin{aligned}
W_Z^\alpha(0, 2|\text{high}) = & \\
& W_A^\alpha(0, 0|\text{high}) \otimes W_B^\alpha(0, 2|\text{high}) \otimes \{\emptyset\} \\
& \cup W_A^\alpha(0, 1|\text{high}) \otimes W_B^\alpha(0, 1|\text{high}) \otimes \{\emptyset\} \\
& \cup W_A^\alpha(0, 2|\text{high}) \otimes W_B^\alpha(0, 0|\text{high}) \otimes \{\emptyset\} \\
& \cup W_A^\alpha(0, 0|\text{high}) \otimes W_B^\alpha(0, 1|\text{low}) \otimes \{\{b\}\} \\
& \cup W_A^\alpha(0, 0|\text{low}) \otimes W_B^\alpha(0, 1|\text{high}) \otimes \{\{a\}\} \\
& \cup W_A^\alpha(0, 1|\text{high}) \otimes W_B^\alpha(0, 0|\text{low}) \otimes \{\{b\}\} \\
& \cup W_A^\alpha(0, 1|\text{low}) \otimes W_B^\alpha(0, 0|\text{high}) \otimes \{\{a\}\} \\
& \cup W_A^\alpha(0, 0|\text{low}) \otimes W_B^\alpha(0, 0|\text{low}) \otimes \{\{a, b\}\} \\
& = \{\{a, b\}\}
\end{aligned} \tag{37}$$

$$\begin{aligned}
W_Z^\alpha(1, 0|\text{high}) = & \\
& W_A^\alpha(0, 0|\text{high}) \otimes W_B^\alpha(1, 0|\text{high}) \otimes \{\emptyset\} \\
& \cup W_A^\alpha(1, 0|\text{high}) \otimes W_B^\alpha(0, 0|\text{high}) \otimes \{\emptyset\} \\
& = \emptyset
\end{aligned} \tag{38}$$

$$\begin{aligned}
W_Z^z(1, 1|\text{high}) = & \\
& W_A^z(0, 0|\text{high}) \otimes W_B^z(1, 1|\text{high}) \otimes \{\emptyset\} \\
& \cup W_A^z(0, 1|\text{high}) \otimes W_B^z(1, 0|\text{high}) \otimes \{\emptyset\} \\
& \cup W_A^z(1, 0|\text{high}) \otimes W_B^z(0, 1|\text{high}) \otimes \{\emptyset\} \\
& \cup W_A^z(1, 1|\text{high}) \otimes W_B^z(0, 0|\text{high}) \otimes \{\emptyset\} \\
& \cup W_A^z(0, 0|\text{high}) \otimes W_B^z(1, 0|\text{low}) \otimes \{\{b\}\} \\
& \cup W_A^z(0, 0|\text{low}) \otimes W_B^z(1, 0|\text{high}) \otimes \{\{a\}\} \\
& \cup W_A^z(1, 0|\text{high}) \otimes W_B^z(0, 0|\text{low}) \otimes \{\{b\}\} \\
& \cup W_A^z(1, 0|\text{low}) \otimes W_B^z(0, 0|\text{high}) \otimes \{\{a\}\} \\
& = \emptyset
\end{aligned} \tag{39}$$

$$\begin{aligned}
W_Z^z(2, 0|\text{high}) = & \\
& W_A^z(0, 0|\text{high}) \otimes W_B^z(2, 0|\text{high}) \otimes \{\emptyset\} \\
& \cup W_A^z(1, 0|\text{high}) \otimes W_B^z(1, 0|\text{high}) \otimes \{\emptyset\} \\
& \cup W_A^z(2, 0|\text{high}) \otimes W_B^z(0, 0|\text{high}) \otimes \{\emptyset\} \\
& = \emptyset
\end{aligned} \tag{40}$$

$$\begin{aligned}
W_Y^z(0, 0|\text{low}) = & \\
& W_Z^z(0, 0|\text{low}) \otimes W_C^z(0, 0|\text{low}) \otimes \{\emptyset\} \\
& = \{\emptyset\}
\end{aligned} \tag{41}$$

$$\begin{aligned}
W_Y^z(0, 1|\text{low}) = & \\
& W_Z^z(0, 0|\text{low}) \otimes W_C^z(0, 1|\text{low}) \otimes \{\emptyset\} \\
& \cup W_Z^z(0, 1|\text{low}) \otimes W_C^z(0, 0|\text{low}) \otimes \{\emptyset\} \\
& = \emptyset
\end{aligned} \tag{42}$$

$$\begin{aligned}
W_Y^z(0, 2|\text{low}) = & \\
& W_Z^z(0, 0|\text{low}) \otimes W_C^z(0, 2|\text{low}) \otimes \{\emptyset\} \\
& \cup W_Z^z(0, 1|\text{low}) \otimes W_C^z(0, 1|\text{low}) \otimes \{\emptyset\} \\
& \cup W_Z^z(0, 2|\text{low}) \otimes W_C^z(0, 0|\text{low}) \otimes \{\emptyset\} \\
& = \emptyset
\end{aligned} \tag{43}$$

$$\begin{aligned}
W_Y^z(1, 0|\text{low}) = & \\
& W_Z^z(0, 0|\text{low}) \otimes W_C^z(1, 0|\text{low}) \otimes \{\emptyset\} \\
& \cup W_Z^z(1, 0|\text{low}) \otimes W_C^z(0, 0|\text{low}) \otimes \{\emptyset\} \\
& \cup W_Z^z(0, 0|\text{high}) \otimes W_C^z(0, 0|\text{low}) \otimes \{\{Z\}\} \\
& \cup W_Z^z(0, 0|\text{low}) \otimes W_C^z(0, 0|\text{high}) \otimes \{\{C\}\} \\
& = \{\{A\}, \{B\}, \{Z\}, \{C\}\}
\end{aligned} \tag{44}$$

$$\begin{aligned}
W_Y^z(1, 1|\text{low}) = & \\
& W_Z^z(0, 0|\text{low}) \otimes W_C^z(1, 1|\text{low}) \otimes \{\emptyset\} \\
& \cup W_Z^z(0, 1|\text{low}) \otimes W_C^z(1, 0|\text{low}) \otimes \{\emptyset\} \\
& \cup W_Z^z(1, 0|\text{low}) \otimes W_C^z(0, 1|\text{low}) \otimes \{\emptyset\} \\
& \cup W_Z^z(1, 1|\text{low}) \otimes W_C^z(0, 0|\text{low}) \otimes \{\emptyset\} \\
& \cup W_Z^z(0, 0|\text{high}) \otimes W_C^z(0, 1|\text{low}) \otimes \{\{Z\}\} \\
& \cup W_Z^z(0, 0|\text{low}) \otimes W_C^z(0, 1|\text{high}) \otimes \{\{C\}\} \\
& \cup W_Z^z(0, 1|\text{high}) \otimes W_C^z(0, 0|\text{low}) \otimes \{\{Z\}\} \\
& \cup W_Z^z(0, 1|\text{low}) \otimes W_C^z(0, 0|\text{high}) \otimes \{\{C\}\} \\
& = \{\{Z, b\}, \{Z, a\}\}
\end{aligned} \tag{45}$$

$$\begin{aligned}
W_Y^z(2, 0|\text{low}) = & \\
& W_Z^z(0, 0|\text{low}) \otimes W_C^z(2, 0|\text{low}) \otimes \{\emptyset\} \\
& \cup W_Z^z(1, 0|\text{low}) \otimes W_C^z(1, 0|\text{low}) \otimes \{\emptyset\} \\
& \cup W_Z^z(2, 0|\text{low}) \otimes W_C^z(0, 0|\text{low}) \otimes \{\emptyset\} \\
& \cup W_Z^z(0, 0|\text{high}) \otimes W_C^z(1, 0|\text{low}) \otimes \{\{Z\}\} \\
& \cup W_Z^z(0, 0|\text{low}) \otimes W_C^z(1, 0|\text{high}) \otimes \{\{C\}\} \\
& \cup W_Z^z(1, 0|\text{high}) \otimes W_C^z(0, 0|\text{low}) \otimes \{\{Z\}\} \\
& \cup W_Z^z(1, 0|\text{low}) \otimes W_C^z(0, 0|\text{high}) \otimes \{\{C\}\} \\
& \cup W_Z^z(0, 0|\text{high}) \otimes W_C^z(0, 0|\text{high}) \otimes \{\{Z, C\}\} \\
& = \{\{A, B\}, \{C, A\}, \{C, B\}, \{Z, C\}\}
\end{aligned} \tag{46}$$

$$\begin{aligned}
W_Y^z(0, 0|\text{high}) = & \\
& W_Z^z(0, 0|\text{high}) \otimes W_C^z(0, 0|\text{high}) \otimes \{\emptyset\} \\
& = \{\emptyset\}
\end{aligned} \tag{47}$$

$$\begin{aligned}
W_Y^z(0, 1|\text{high}) = & \\
& W_Z^z(0, 0|\text{high}) \otimes W_C^z(0, 1|\text{high}) \otimes \{\emptyset\} \\
& \cup W_Z^z(0, 1|\text{high}) \otimes W_C^z(0, 0|\text{high}) \otimes \{\emptyset\} \\
& \cup W_Z^z(0, 0|\text{high}) \otimes W_C^z(0, 0|\text{low}) \otimes \{\{c\}\} \\
& \cup W_Z^z(0, 0|\text{low}) \otimes W_C^z(0, 0|\text{high}) \otimes \{\{z\}\} \\
& = \{\{b\}, \{a\}, \{c\}, \{z\}\}
\end{aligned} \tag{48}$$

$$\begin{aligned}
W_Y^z(0, 2|\text{high}) = & \\
& W_Z^z(0, 0|\text{high}) \otimes W_C^z(0, 2|\text{high}) \otimes \{\emptyset\} \\
& \cup W_Z^z(0, 1|\text{high}) \otimes W_C^z(0, 1|\text{high}) \otimes \{\emptyset\} \\
& \cup W_Z^z(0, 2|\text{high}) \otimes W_C^z(0, 0|\text{high}) \otimes \{\emptyset\} \\
& \cup W_Z^z(0, 0|\text{high}) \otimes W_C^z(0, 1|\text{low}) \otimes \{\{c\}\} \\
& \cup W_Z^z(0, 0|\text{low}) \otimes W_C^z(0, 1|\text{high}) \otimes \{\{z\}\} \\
& \cup W_Z^z(0, 1|\text{high}) \otimes W_C^z(0, 0|\text{low}) \otimes \{\{c\}\} \\
& \cup W_Z^z(0, 1|\text{low}) \otimes W_C^z(0, 0|\text{high}) \otimes \{\{z\}\} \\
& \cup W_Z^z(0, 0|\text{low}) \otimes W_C^z(0, 0|\text{low}) \otimes \{\{z, c\}\} \\
& = \{\{a, b\}, \{c, b\}, \{c, a\}, \{z, c\}\}
\end{aligned} \tag{49}$$

$$\begin{aligned}
W_Y^z(1, 0|\text{high}) = & \\
& W_Z^z(0, 0|\text{high}) \otimes W_C^z(1, 0|\text{high}) \otimes \{\emptyset\} \\
& \cup W_Z^z(1, 0|\text{high}) \otimes W_C^z(0, 0|\text{high}) \otimes \{\emptyset\} \\
& = \emptyset
\end{aligned} \tag{50}$$

$$\begin{aligned}
W_Y^z(1, 1|\text{high}) = & \\
& W_Z^z(0, 0|\text{high}) \otimes W_C^z(1, 1|\text{high}) \otimes \{\emptyset\} \\
& \cup W_Z^z(0, 1|\text{high}) \otimes W_C^z(1, 0|\text{high}) \otimes \{\emptyset\} \\
& \cup W_Z^z(1, 0|\text{high}) \otimes W_C^z(0, 1|\text{high}) \otimes \{\emptyset\} \\
& \cup W_Z^z(1, 1|\text{high}) \otimes W_C^z(0, 0|\text{high}) \otimes \{\emptyset\} \\
& \cup W_Z^z(0, 0|\text{high}) \otimes W_C^z(1, 0|\text{low}) \otimes \{\{c\}\} \\
& \cup W_Z^z(0, 0|\text{low}) \otimes W_C^z(1, 0|\text{high}) \otimes \{\{z\}\} \\
& \cup W_Z^z(1, 0|\text{high}) \otimes W_C^z(0, 0|\text{low}) \otimes \{\{c\}\} \\
& \cup W_Z^z(1, 0|\text{low}) \otimes W_C^z(0, 0|\text{high}) \otimes \{\{z\}\} \\
& = \{\{z, A\}, \{z, B\}\}
\end{aligned} \tag{51}$$

$$\begin{aligned}
W_Y^z(2, 0|\text{high}) = & \\
& W_Z^z(0, 0|\text{high}) \otimes W_C^z(2, 0|\text{high}) \otimes \{\emptyset\} \\
& \cup W_Z^z(1, 0|\text{high}) \otimes W_C^z(1, 0|\text{high}) \otimes \{\emptyset\} \\
& \cup W_Z^z(2, 0|\text{high}) \otimes W_C^z(0, 0|\text{high}) \otimes \{\emptyset\} \\
& = \emptyset
\end{aligned} \tag{52}$$

$$\begin{aligned}
W_X^z(0, 2|\text{low}) = & \\
& W_Y^z(0, 0|\text{low}) \otimes W_D^z(0, 2|\text{low}) \otimes \{\emptyset\} \\
& \cup W_Y^z(0, 1|\text{low}) \otimes W_D^z(0, 1|\text{low}) \otimes \{\emptyset\} \\
& \cup W_Y^z(0, 2|\text{low}) \otimes W_D^z(0, 0|\text{low}) \otimes \{\emptyset\} \\
& = \emptyset
\end{aligned} \tag{53}$$

$$\begin{aligned}
W_X^z(1, 1|\text{low}) = & \\
& W_Y^z(0, 0|\text{low}) \otimes W_D^z(1, 1|\text{low}) \otimes \{\emptyset\} \\
& \cup W_Y^z(0, 1|\text{low}) \otimes W_D^z(1, 0|\text{low}) \otimes \{\emptyset\} \\
& \cup W_Y^z(1, 0|\text{low}) \otimes W_D^z(0, 1|\text{low}) \otimes \{\emptyset\} \\
& \cup W_Y^z(1, 1|\text{low}) \otimes W_D^z(0, 0|\text{low}) \otimes \{\emptyset\} \\
& \cup W_Y^z(0, 0|\text{high}) \otimes W_D^z(0, 1|\text{low}) \otimes \{\{Y\}\} \\
& \cup W_Y^z(0, 0|\text{low}) \otimes W_D^z(0, 1|\text{high}) \otimes \{\{D\}\} \\
& \cup W_Y^z(0, 1|\text{high}) \otimes W_D^z(0, 0|\text{low}) \otimes \{\{Y\}\} \\
& \cup W_Y^z(0, 1|\text{low}) \otimes W_D^z(0, 0|\text{high}) \otimes \{\{D\}\} \\
& = \{\{Z, b\}, \{Z, a\}, \{Y, b\}, \{Y, a\}, \{Y, c\}, \{Y, z\}\}
\end{aligned} \tag{54}$$

$$\begin{aligned}
W_X^z(2, 0|\text{low}) = & \\
& W_Y^z(0, 0|\text{low}) \otimes W_D^z(2, 0|\text{low}) \otimes \{\emptyset\} \\
& \cup W_Y^z(1, 0|\text{low}) \otimes W_D^z(1, 0|\text{low}) \otimes \{\emptyset\} \\
& \cup W_Y^z(2, 0|\text{low}) \otimes W_D^z(0, 0|\text{low}) \otimes \{\emptyset\} \\
& \cup W_Y^z(0, 0|\text{high}) \otimes W_D^z(1, 0|\text{low}) \otimes \{\{Y\}\} \\
& \cup W_Y^z(0, 0|\text{low}) \otimes W_D^z(1, 0|\text{high}) \otimes \{\{D\}\} \\
& \cup W_Y^z(1, 0|\text{high}) \otimes W_D^z(0, 0|\text{low}) \otimes \{\{Y\}\} \\
& \cup W_Y^z(1, 0|\text{low}) \otimes W_D^z(0, 0|\text{high}) \otimes \{\{D\}\} \\
& \cup W_Y^z(0, 0|\text{high}) \otimes W_D^z(0, 0|\text{high}) \otimes \{\{Y, D\}\} \\
& = \{\{A, B\}, \{C, A\}, \{C, B\}, \{Z, C\}, \{D, A\}, \{D, B\}, \{D, Z\}, \{D, C\}, \{Y, D\}\}
\end{aligned} \tag{55}$$

$$\begin{aligned}
W_X^z(0, 2|\text{high}) = & \\
& W_Y^z(0, 0|\text{high}) \otimes W_D^z(0, 2|\text{high}) \otimes \{\emptyset\} \\
& \cup W_Y^z(0, 1|\text{high}) \otimes W_D^z(0, 1|\text{high}) \otimes \{\emptyset\} \\
& \cup W_Y^z(0, 2|\text{high}) \otimes W_D^z(0, 0|\text{high}) \otimes \{\emptyset\} \\
& \cup W_Y^z(0, 0|\text{high}) \otimes W_D^z(0, 1|\text{low}) \otimes \{\{d\}\} \\
& \cup W_Y^z(0, 0|\text{low}) \otimes W_D^z(0, 1|\text{high}) \otimes \{\{y\}\} \\
& \cup W_Y^z(0, 1|\text{high}) \otimes W_D^z(0, 0|\text{low}) \otimes \{\{d\}\} \\
& \cup W_Y^z(0, 1|\text{low}) \otimes W_D^z(0, 0|\text{high}) \otimes \{\{y\}\} \\
& \cup W_Y^z(0, 0|\text{low}) \otimes W_D^z(0, 0|\text{low}) \otimes \{\{y, d\}\} \\
& = \{\{a, b\}, \{c, b\}, \{c, a\}, \{z, c\}, \{d, b\}, \{d, a\}, \{d, c\}, \{d, z\}, \{y, d\}\}
\end{aligned} \tag{56}$$

$$\begin{aligned}
W_X^z(1, 1|\text{high}) = & W_Y^z(0, 0|\text{high}) \otimes W_D^z(1, 1|\text{high}) \otimes \{\emptyset\} \\
& \cup W_Y^z(0, 1|\text{high}) \otimes W_D^z(1, 0|\text{high}) \otimes \{\emptyset\} \\
& \cup W_Y^z(1, 0|\text{high}) \otimes W_D^z(0, 1|\text{high}) \otimes \{\emptyset\} \\
& \cup W_Y^z(1, 1|\text{high}) \otimes W_D^z(0, 0|\text{high}) \otimes \{\emptyset\} \\
& \cup W_Y^z(0, 0|\text{high}) \otimes W_D^z(1, 0|\text{low}) \otimes \{\{d\}\} \\
& \cup W_Y^z(0, 0|\text{low}) \otimes W_D^z(1, 0|\text{high}) \otimes \{\{y\}\} \\
& \cup W_Y^z(1, 0|\text{high}) \otimes W_D^z(0, 0|\text{low}) \otimes \{\{d\}\} \\
& \cup W_Y^z(1, 0|\text{low}) \otimes W_D^z(0, 0|\text{high}) \otimes \{\{y\}\} \\
& = \{\{z, A\}, \{z, B\}, \{y, A\}, \{y, B\}, \{y, Z\}, \{y, C\}\}
\end{aligned} \tag{57}$$

$$\begin{aligned}
W_X^z(2, 0|\text{high}) = & W_Y^z(0, 0|\text{high}) \otimes W_D^z(2, 0|\text{high}) \otimes \{\emptyset\} \\
& \cup W_Y^z(1, 0|\text{high}) \otimes W_D^z(1, 0|\text{high}) \otimes \{\emptyset\} \\
& \cup W_Y^z(2, 0|\text{high}) \otimes W_D^z(0, 0|\text{high}) \otimes \{\emptyset\} \\
& = \emptyset
\end{aligned} \tag{58}$$

*Ways that one event in character  $\beta$  can happen in a tree*

Given character  $\beta$  from Figure 4b ( $\Sigma_\beta = \{28, 30\}$ ), we detail how to compute all possible ways that one event, from  $t(\Sigma_\beta) = \{28 \rightarrow 30, 30 \rightarrow 28\}$  can happen on tree  $T$  of Figure 4a, given either state 28 or 30 at the root ( $X$ ) of the tree  $T$ , that is,  $W_X^\beta(0, 1|28)$  [Eq. (71)],  $W_X^\beta(1, 0|28)$  [Eq. (72)],  $W_X^\beta(0, 1|30)$  [Eq. (73)], and  $W_X^\beta(1, 0|30)$  [Eq. (74)]. The base case(s) for the leaf nodes  $A, B, C$ , and  $D$  of tree  $T$  is set according to Equation (12) as follows

$$\begin{aligned}
W_A^\beta(0, 0|28) &= W_A^\beta(0, 0|30) = \{\emptyset\} \\
W_B^\beta(0, 0|28) &= W_B^\beta(0, 0|30) = \{\emptyset\} \\
W_C^\beta(0, 0|28) &= W_C^\beta(0, 0|30) = \{\emptyset\} \\
W_D^\beta(0, 0|28) &= W_D^\beta(0, 0|30) = \{\emptyset\} \\
\forall r_1, r_2 > 0 & W_A^\beta(r_1, r_2|28) = W_A^\beta(r_1, r_2|30) = \emptyset \\
\forall r_1, r_2 > 0 & W_B^\beta(r_1, r_2|28) = W_B^\beta(r_1, r_2|30) = \emptyset \\
\forall r_1, r_2 > 0 & W_C^\beta(r_1, r_2|28) = W_C^\beta(r_1, r_2|30) = \emptyset \\
\forall r_1, r_2 > 0 & W_D^\beta(r_1, r_2|28) = W_D^\beta(r_1, r_2|30) = \emptyset
\end{aligned}$$

The recursive step(s) for each internal node  $Z, Y$  and finally  $X$  of tree  $T$  is then computed according to Equation (13) as follows

$$\begin{aligned}
W_Z^\beta(0, 0|28) = & W_A^\beta(0, 0|28) \otimes W_B^\beta(0, 0|28) \otimes \{\emptyset\} \\
& = \{\emptyset\}
\end{aligned} \tag{59}$$

$$\begin{aligned}
W_Z^\beta(0, 1|28) = & W_A^\beta(0, 0|28) \otimes W_B^\beta(0, 1|28) \otimes \{\emptyset\} \\
& \cup W_A^\beta(0, 1|28) \otimes W_B^\beta(0, 0|28) \otimes \{\emptyset\} \\
& = \emptyset
\end{aligned} \tag{60}$$

$$\begin{aligned}
W_Z^\beta(1, 0|28) = & \\
& W_A^\beta(0, 0|28) \otimes W_B^\beta(1, 0|28) \otimes \{\emptyset\} \\
& \cup W_A^\beta(1, 0|28) \otimes W_B^\beta(0, 0|28) \otimes \{\emptyset\} \\
& \cup W_A^\beta(0, 0|28) \otimes W_B^\beta(0, 0|30) \otimes \{\{B\}\} \\
& \cup W_A^\beta(0, 0|30) \otimes W_B^\beta(0, 0|28) \otimes \{\{A\}\} \\
& = \{\{B\}, \{A\}\}
\end{aligned} \tag{61}$$

$$\begin{aligned}
W_Z^\beta(0, 0|30) = & \\
& W_A^\beta(0, 0|30) \otimes W_B^\beta(0, 0|30) \otimes \{\emptyset\} \\
& = \{\emptyset\}
\end{aligned} \tag{62}$$

$$\begin{aligned}
W_Z^\beta(0, 1|30) = & \\
& W_A^\beta(0, 0|30) \otimes W_B^\beta(0, 1|30) \otimes \{\emptyset\} \\
& \cup W_A^\beta(0, 1|30) \otimes W_B^\beta(0, 0|30) \otimes \{\emptyset\} \\
& \cup W_A^\beta(0, 0|28) \otimes W_B^\beta(0, 0|30) \otimes \{\{a\}\} \\
& \cup W_A^\beta(0, 0|30) \otimes W_B^\beta(0, 0|28) \otimes \{\{b\}\} \\
& = \{\{a\}, \{b\}\}
\end{aligned} \tag{63}$$

$$\begin{aligned}
W_Z^\beta(1, 0|30) = & \\
& W_A^\beta(0, 0|30) \otimes W_B^\beta(1, 0|30) \otimes \{\emptyset\} \\
& \cup W_A^\beta(1, 0|30) \otimes W_B^\beta(0, 0|30) \otimes \{\emptyset\} \\
& = \emptyset
\end{aligned} \tag{64}$$

$$\begin{aligned}
W_Y^\beta(0, 0|28) = & \\
& W_Z^\beta(0, 0|28) \otimes W_C^\beta(0, 0|28) \otimes \{\emptyset\} \\
& = \{\emptyset\}
\end{aligned} \tag{65}$$

$$\begin{aligned}
W_Y^\beta(0, 1|28) = & \\
& W_Z^\beta(0, 0|28) \otimes W_C^\beta(0, 1|28) \otimes \{\emptyset\} \\
& \cup W_Z^\beta(0, 1|28) \otimes W_C^\beta(0, 0|28) \otimes \{\emptyset\} \\
& = \emptyset
\end{aligned} \tag{66}$$

$$\begin{aligned}
W_Y^\beta(1, 0|28) = & \\
& W_Z^\beta(0, 0|28) \otimes W_C^\beta(1, 0|28) \otimes \{\emptyset\} \\
& \cup W_Z^\beta(1, 0|28) \otimes W_C^\beta(0, 0|28) \otimes \{\emptyset\} \\
& \cup W_Z^\beta(0, 0|28) \otimes W_C^\beta(0, 0|30) \otimes \{\{C\}\} \\
& \cup W_Z^\beta(0, 0|30) \otimes W_C^\beta(0, 0|28) \otimes \{\{Z\}\} \\
& = \{\{B\}, \{A\}, \{C\}, \{Z\}\}
\end{aligned} \tag{67}$$

$$\begin{aligned}
W_Y^\beta(0, 0|30) = & \\
& W_Z^\beta(0, 0|30) \otimes W_C^\beta(0, 0|30) \otimes \{\emptyset\} \\
& = \{\emptyset\}
\end{aligned} \tag{68}$$

$$\begin{aligned}
W_Y^\beta(0, 1|30) = & \\
& W_Z^\beta(0, 0|30) \otimes W_C^\beta(0, 1|30) \otimes \{\emptyset\} \\
& \cup W_Z^\beta(0, 1|30) \otimes W_C^\beta(0, 0|30) \otimes \{\emptyset\} \\
& \cup W_Z^\beta(0, 0|28) \otimes W_C^\beta(0, 0|30) \otimes \{\{z\}\} \\
& \cup W_Z^\beta(0, 0|30) \otimes W_C^\beta(0, 0|28) \otimes \{\{c\}\} \\
& = \{\{a\}, \{b\}, \{z\}, \{c\}\}
\end{aligned} \tag{69}$$

$$\begin{aligned}
W_Y^\beta(1, 0|30) = & \\
& W_Z^\beta(0, 0|30) \otimes W_C^\beta(1, 0|30) \otimes \{\emptyset\} \\
& \cup W_Z^\beta(1, 0|30) \otimes W_C^\beta(0, 0|30) \otimes \{\emptyset\} \\
& = \emptyset
\end{aligned} \tag{70}$$

$$\begin{aligned}
W_X^\beta(0, 1|28) = & \\
& W_Y^\beta(0, 0|28) \otimes W_D^\beta(0, 1|28) \otimes \{\emptyset\} \\
& \cup W_Y^\beta(0, 1|28) \otimes W_D^\beta(0, 0|28) \otimes \{\emptyset\} \\
& = \emptyset
\end{aligned} \tag{71}$$

$$\begin{aligned}
W_X^\beta(1, 0|28) = & \\
& W_Y^\beta(0, 0|28) \otimes W_D^\beta(1, 0|28) \otimes \{\emptyset\} \\
& \cup W_Y^\beta(1, 0|28) \otimes W_D^\beta(0, 0|28) \otimes \{\emptyset\} \\
& \cup W_Y^\beta(0, 0|28) \otimes W_D^\beta(0, 0|30) \otimes \{\{D\}\} \\
& \cup W_Y^\beta(0, 0|30) \otimes W_D^\beta(0, 0|28) \otimes \{\{Y\}\} \\
& = \{\{B\}, \{A\}, \{C\}, \{Z\}, \{D\}, \{Y\}\}
\end{aligned} \tag{72}$$

$$\begin{aligned}
W_X^\beta(0, 1|30) = & \\
& W_Y^\beta(0, 0|30) \otimes W_D^\beta(0, 1|30) \otimes \{\emptyset\} \\
& \cup W_Y^\beta(0, 1|30) \otimes W_D^\beta(0, 0|30) \otimes \{\emptyset\} \\
& \cup W_Y^\beta(0, 0|28) \otimes W_D^\beta(0, 0|30) \otimes \{\{y\}\} \\
& \cup W_Y^\beta(0, 0|30) \otimes W_D^\beta(0, 0|28) \otimes \{\{d\}\} \\
& = \{\{a\}, \{b\}, \{z\}, \{c\}, \{y\}, \{d\}\}
\end{aligned} \tag{73}$$

$$\begin{aligned}
W_X^\beta(1, 0|30) = & \\
& W_Y^\beta(0, 0|30) \otimes W_D^\beta(1, 0|30) \otimes \{\emptyset\} \\
& \cup W_Y^\beta(1, 0|30) \otimes W_D^\beta(0, 0|30) \otimes \{\emptyset\} \\
& = \emptyset
\end{aligned} \tag{74}$$