Correlated Evolution in the Small Parsimony Framework*

Brendan Smith[1], Cristian Navarro-Martinez[1], Rebecca Buonopane[1], S. Ashley Byun[1], and Murray Patterson[2][0000−0002−4329−0234]

[1] Fairfield University, Fairfield CT 06824, USA
{brendan.smith1, rebecca.buonopane, cristian.navarro-martinez}@student.fairfield.edu
sbyun@fairfield.edu
[2] Georgia State University, Atlanta GA 30303, USA
mpatterson@cs.gsu.edu

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, *e.g.*, 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 detecting correlated evolution among pairs of discrete characters, 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. We implement this method: `parcours` (PARsimonious CO-occURrenceS) and use it to study the correlated evolution among vocalizations and morphological characters in the Felidae family, revealing some interesting results.
`parcours` is freely available at https://github.com/murraypatterson/parcours

Keywords: Parsimony · Evolution · Correlation

1 Introduction

The principle of parsimony first appeared in the context of computational evolutionary biology in [8,15]. A few years later, Sankoff (and Rousseau) generalized this to allow the association of a (different) cost to each transition between a pair

* Research supported by: Fairfield University Science Institute grants for SAB and MP; Georgia State University startup grant for MP.

of states [32]. A few years after that, Felsenstein noticed that parsimony could produce misleading results when the evolutionary rate of change on the branches of the phylogeny is high [12]. 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 [12,10]. 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 RAxML [34], PHYLIP [11] and the Bio++ library [16]. Even more recently, some Bayesian approaches have appeared, which sample the space of likely trees using the Markov-Chain Monte Carlo (MCMC) method, resulting in tools such as MrBayes [18] and Beast [7].

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 [35]. Finally, when the evolutionary rate of change of different characters is heterogeneous, maximum parsimony has even been shown to perform *substantially better* than maximum likelihood and Bayesian approaches [22]. A perfect example is cancer phylogenetics [17,33], 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 [20] and SASC [4].

The setting of the present work is an evolutionary study of vocalizations and morphological characters [3] among members of the family Felidae. Felids possess a range of intraspecific vocalizations for close, medium, and long range communication. There is no evidence that Felid vocalizations are learned: it is more likely that these calls are genetically determined [27,9]. While there are 14 major discrete and graded calls documented in Felidae, not all calls are produced by all species [29]. In [30], the authors map some of these calls to a molecular phylogeny of Felidae, to show that it was consistent with what was previously known, strengthening the argument that vocalizations are genetically determined. Here we consider a similar approach, but because (a) an obvious model of evolution is lacking in this case; (b) the possibility that vocalizations within a given group of species can evolve at considerably different rates [30]; and (c) that rates for specific characters can differ between different lineages within that group [30], it follows from [22] that parsimony is more appropriate than maximum likelihood or a Bayesian approaches.

In this work, we develop a general framework to determine correlated (or co-) evolution among pairs of characters in a phylogeny, 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, and how they might

correlate (or have co-evolved) with each other. 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* (see 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. We implement this approach in a tool called `parcours`, and use it to detect correlated evolution among 14 Felid vocalizations and 10 morphological characters, obtaining results that are consistent with the literature [30] as well as some interesting associations. While various methods for detecting correlated evolution exist, they tend to use only *phylogenetic profiles* [24], or are based on maximum likelihood [2,6], where a model of evolution is needed. Methods that determine co-evolution in the parsimony framework exist as well [26], however they are aimed at reconstructing ancestral gene *adjacencies*, given extant co-localization information. Finally, while some of the maximum likelihood software tools have a "parsimony mode", *e.g.*, [11,16], the character information must be encoded using restrictive alphabets, and there is no easy or 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* (CSV) file, inferring the alphabet from this input, and efficiently computes all small parsimonies, automatically. In summary, our contribution is a methodology for inferring correlation among pairs of characters when no model of evolution is available, and implemented into an open-source software that is so easy to use that it could also be used as a pedagogical tool.

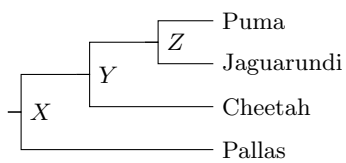
This paper is structured as follows. In Sec. 2, we provide the background on parsimony, as well as our approach for efficiently computing *all* most parsimonious ancestral states in a phylogeny, given a set of extant states. In Sec. 3, we present our approach for computing correlation between pairs of characters from all such parsimonies. In Sec. 4, we describe an experimental analysis of the implementation, `parcours`, of our method on Felid vocalizations and morphological characters. Finally, in Sec. 5, we conclude the paper with a discussion of the results obtained, and 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, etc.). The idea is that each species is in one of these states for each character — *e.g.*, in the Puma, the weight (one of the 10 morphological characters) 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 Fig. 1a, implying the existence of the ancestors X, Y and Z. We are given some character, *e.g.*, weight, which can be in one of a number of states taken from *alphabet* $\Sigma = \{\text{low, high, unknown}\}$, which we code here as $\{0, 1, 2\}$ for compactness. If weight is high only in Puma and Cheetah as in Fig. 1b, then the assignment of 0 (low) to all ancestors would be a small parsimony. This parsimony has two changes of state: a change $0 \to 1$ on the branches $Z \to$ Puma and $Y \to$ Cheetah — implying convergent increase of weight in these species. Another small parsimony is that weight is high in X and Y (low in Z) — 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 [32], which also makes use of a *cost matrix* δ, depicted in Fig. 1c, that encodes the evolutionary cost of each state change along a branch. Since the change $0 \to 1$ (low to high) costs $\delta_{0,1} = 1$, and vice versa ($\delta_{1,0} = 1$), 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 possibilities shows that 2 is the minimum score of any assignment of ancestral states.



(a)

sp. \ ch.	weight
Puma	1
Jaguarundi	0
Cheetah	1
Pallas	0

(b)

δ	0	1	2
0	0	1	1
1	1	0	1
2	∞	∞	0

(c)

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 $i \to j$ from state i in the parent to state j in the child along a branch of the phylogeny, *e.g.*, $\delta_{1,2} = 1$ ($\delta_{2,1} = \infty$).

Sankoff's algorithm [32] computes all (small) parsimonies given a phylogeny and the extant states of a character in a set of species, and a cost matrix δ, *e.g.*, Fig. 1. The algorithm has a bottom-up phase (from the leaves to the root of the phylogeny), and then a top-down phase. The first (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:

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

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 ∞ otherwise. The score for the internal nodes is then computed according to Eq. 1 in a bottom-up dynamic programming fashion, starting from the leaves. The boxes in Fig. 2 depict the result of this first phase on the instance of Fig. 1, for example:

$$\begin{aligned} s_0(X) = &\min\ \{s_0(\text{Pallas}) + \delta_{0,0},\ s_1(\text{Pallas}) + \delta_{0,1},\ s_2(\text{Pallas}) + \delta_{0,2}\} \\ &+ \min\ \{s_0(Y) + \delta_{0,0},\ s_1(Y) + \delta_{0,1},\ s_2(Y) + \delta_{0,2}\} \\ = &\min\ \{0+0,\ \infty+1,\ \infty+1\} + \min\ \{2+0,\ 1+1,\ \infty+1\} \\ = &\min\ \{0,\ \infty,\ \infty\} + \min\ \{2,\ 2,\ \infty\} = 0 + 2 = 2 \end{aligned} \tag{2}$$

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.

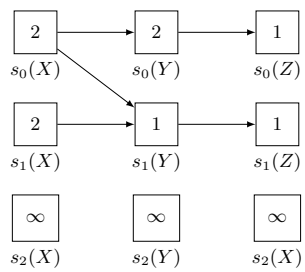


Fig. 2: Graph structure resulting from Sankoff's algorithm on the instance of Fig. 1. The *boxes* contain the values of $s_i(u)$ computed in the bottom-up phase, for each state $i \in \Sigma$ and internal node $u \in \{X, Y, Z\}$ of the phylogeny. The *arrows* connecting these boxes were computed in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

After the bottom-up phase, we know the minimum score of any assignment of ancestral states, but we do not yet have an ancestral assignment of states. Here, since X is the root in Fig. 1a, we see from Fig. 2 that the minimum score for this example is 2, as we saw earlier. Note that this minimum may not be unique; indeed, here $s_0(X) = s_1(X) = 2$, meaning that in at least one parsimony, weight is low (0) in X, and in at least one other parsimony, weight is high (1) in X (but never unknown, *i.e.*, $s_2(X) = \infty$). Now, to reconstruct one of these ancestral assignments of minimum score, 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 fashion until we reach the leaves of the tree, having assigned all states at this point. For example, $s_0(X) = 2$, and can be derived in Eq. 2 from $s_1(Y)$ (and $s_0(\text{Pallas})$), which is in turn derived from $s_1(Z)$. This corresponds to the second parsimony mentioned earlier, where high weight was ancestral in Y, and later decreased in the Jaguarundi. Notice that $s_0(X)$ can also be derived in Eq. 2 from $s_0(Y)$,

which is in turn derived from $s_0(Z)$. This is the first parsimony where weight is low (0) 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 Fig. 2, for example, the spanning subtree $s_0(X) \to s_0(Y) \to s_0(Z)$ corresponds to the parsimony where weight is low in all ancestors, and $s_0(X) \to s_1(Y) \to s_1(Z)$ 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 [5] for another example of this graph structure, with the leaves included. Notice, from Fig. 2 that there is a third solution $s_1(X) \to s_1(Y) \to s_1(Z)$ corresponding to a parsimony where high weight was ancestral (in X) to all species here, and then decreased in both the Pallas 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, *e.g.*, $s_0(Y)$. 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 $O(k^n)$ parsimonies in general. For example, consider the extreme case where all costs in δ are 0. For each internal node v and its parent u in the phylogeny, the set $s_i(v)$ of k boxes form 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 Fig. 2, such (sets of) states are $s_0(u)$ and $s_1(u)$ 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 $\{1, \ldots, k\}^{n-1}$), a standard (*e.g.*, based on the left-hand rule) depth-first traversal (DFT) 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 [25], 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

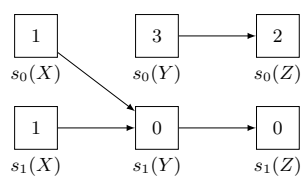
tree) updates only the edge to the parent and to the (amortized constant number of) children of u in this spanning subtree. It then suffices to check if this update is also a spanning subtree in order to verify if the current combination also forms a spanning subtree. This shaves a linear $O(n)$ factor off of the runtime of the naive DFT approach, which can be substantial in the case of dense (potential) solution spaces. While this graph structure (Fig. 2) is mentioned [5], how to efficiently enumerate all its parsimonies is rarely discussed (that we know of). On the other hand, 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 $\{28, 30, \text{unknown}\}$, which we again code as $\{0, 1, 2\}$ for compactness. Here, the dental profile is "fewer" (0) only in the Pallas cat as depicted in Fig. 3a, and we want to understand how this is correlated to the weight of Fig. 1b.

sp. \ ch.	dental
Puma	1
Jaguarundi	1
Cheetah	1
Pallas	0

(a)



(b)

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 Fig. 1a and cost of Fig. 1c. Values s_2 have been omitted for compactness, since they are all ∞, like in Fig. 2.

The idea is that we first construct likely hypotheses for what ancestral states each character c might have, namely the set $P(c)$ 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 $a = \alpha \to \beta$ be a change in state (for some pair, α, β of states) of character c. Let $B_p(a)$ be the *multiset* (even though each element will have multiplicity 0 or 1) of *branches* in the phylogenetic tree where change a occurs in parsimony $p \in P(c)$. The *consensus map* $C_c(a)$ for change a of character c is

then the multiset resulting from

$$\bigcup_{p \in P(c)} B_p(a) \ , \tag{3}$$

in preserving the information of *all* parsimonies. Finally, the *correlation* between some change a of character c and some change b of a character d is then the *weighted Jaccard index* [19] (also known as the Ružička index [31])

$$\frac{|C_c(a) \cap C_d(b)|}{|C_c(a) \cup C_d(b)|} \ . \tag{4}$$

For example, given the characters weight (Fig. 1b) and dental profile (Fig. 3a), we compute the correlation of the change $0 \to 1$ (low to high in weight, and 28 to 30 in dental) of both of these characters in the phylogeny of Fig. 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 $B_{p_1}(0 \to 1) = \{Z \to \text{Puma}, Y \to \text{Cheetah}\}$, $B_{p_2}(0 \to 1) = \{X \to Y\}$, and $B_{p_3}(0 \to 1) = \emptyset$. It follows from Eq. 3 that the consensus map $C_{0\to 1}(\text{weight}) = \{Z \to \text{Puma}, Y \to \text{Cheetah}, X \to Y\}$. By inspecting Fig. 3b, character dental has the set $P(\text{dental}) = \{q_1, q_2\}$ of two parsimonies, with $B_{q_1}(0 \to 1) = \{X \to Y\}$ and $B_{q_2}(0 \to 1) = \emptyset$, hence $C_{0\to 1}(\text{dental}) = \{X \to Y\}$. It then follows from Eq. 4 that the correlation of this co-event is $1/3 \approx 0.33$. The correlation of the co-event $1 \to 0$ in both weight and dental is also $1/3$, while the correlation of $0 \to 1$ in weight and $1 \to 0$ in dental (and vice versa) are 0, because there is no overlap between sets $B_p(0 \to 1)$ and $B_q(1 \to 0)$ 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 $0 \to 2$ or $2 \to 0$ with any other event in either weight or dental 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 Eq. 4 where all multisets are treated as *sets* (all non-zero multiplicities cast to 1).

4 Experimental Analysis

We used the following data to validate our approach. The phylogeny is that of the family Felidae from Johnson et al. [21], where Leopard and Jaguar are "swapped"[3], due to more recent findings [13]. The set of extant states are for the 14 vocalizations documented in Sunquist and Sunquist [36], and the 10 morphological characters compiled from the various sources within [3]. Finally, the cost for changing to a different state is 1, and to an *unknown* state is ∞ (*e.g.*, Fig. 1c).

[3] see example at https://github.com/murraypatterson/parcours/tree/main/felidae

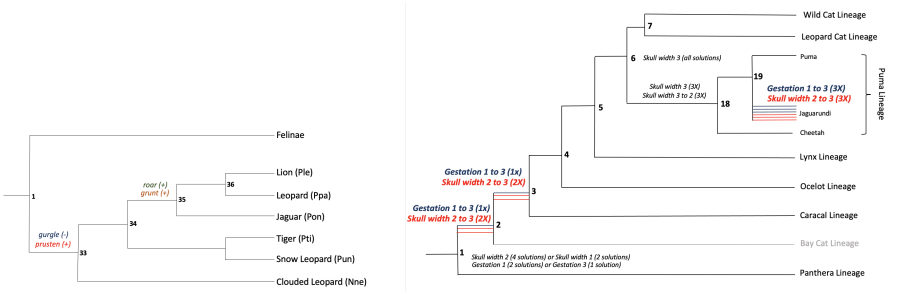
Since unknown states are artifacts of the collection process, by assigning a high evolutionary cost to the changes of any unknown to some known state (which makes little evolutionary sense) we mitigate the propagation of any unknown state to ancestral nodes in any parsimony, whenever some known state (*e.g.*, 0 or 1) can explain the data. For example, in the instance of Fig. 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 state 1 (high). This is why we use the approach of Sankoff [32] (instead of, *e.g.*, Fitch [15]), 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 Sec. 2, followed by the computation of correlation from these parsimonies, as described in Sec. 3. 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 24 characters and the Felidae phylogeny mentioned above. The cost matrix mentioned above 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 Eq. 4. Since each correlation is, by the definition in Eq. 4, a value between 0 and 1, we report in Table 1 all events for all pairs of characters that had a correlation of 0.5 or greater. Fig. 4 show some consensus maps on the tree for selected characters of Table 1.

Table 1: Pairs of correlated events[a]

	c_1	c_2	t_1	t_2	Correlation	Jaccard
1	gurgle	prusten	$1 \to 0$	$1 \to 0$	1.0	1.0
2	gurgle	gestation	$1 \to 0$	$3 \to 1$	1.0	1.0
3	prusten	gestation	$0 \to 1$	$3 \to 1$	1.0	1.0
4	grunt	roar	$0 \to 1$	$0 \to 1$	1.0	1.0
5	gestation	weight	$4 \to 1$	$3 \to 2$	0.5	0.5
6	weight	skull L.	$4 \to 3$	$3 \to 2$	0.57	0.5
7	weight	skull L.	$1 \to 2$	$1 \to 2$	0.64	1.0
8	skull W.	gestation	$2 \to 3$	$1 \to 3$	0.71	1.0
9	sk.L./sk.W.	dental	$4 \to 3$	$1 \to 2$	0.67	0.67

[a] Pairs t_1, t_2 of events (columns 3–4) in pairs c_1, c_2 of characters (columns 1–2) with correlation (Eq. 4) of 0.5 or greater (column 5). Jaccard index (see Sec. 3) is also included for reference (column 6). Vocalizations (gurgle, prusten, grunt, roar) are all binary (0 or 1), while the remaining morphological characters have categorical ranks.



(a) (b)

Fig. 4: The consensus maps (Eq. 3) corresponding to the pairs of events in: (a) rows 1 and 4 of Table 1, where $(+)$ is $0 \to 1$ and $(-)$ is $1 \to 0$; and (b) row 8 of Table 1, where the number of lines on a branch indicates multiplicity. Note here, that the Bay cat lineage (grayed out) was excluded (automatically by `parcours`) from analyses due to lack of data on species of that group.

5 Discussion

Understanding the evolution of complex traits such as vocalization and morphological characters remains a key challenge in evolutionary studies. We discuss the results of our experimental analysis on these characters, and future directions below. The gurgle and prusten are two of the major friendly, close range vocalizations in Felidae. These vocalizations are short, largely atonal calls produced during friendly approaches and close contact situations. Each felid species uses only one of these homologous, close-range calls [28]. The gurgle is the most widely distributed call, found in all felid species except for members of Pantherinae, and are considered to be ancestral. Along the lineage leading to Pantherinae, the prusten appears to have replaced the gurgle about 8–3 myr BP (node 1 [21] to 33) [30]. The complete correlation of a loss of gurgle and gain of prusten (row 1 of Table 1, and Fig 4a) is consistent with the idea that these calls are functionally equivalent vocalizations, one being replaced as another is gained successively over time [30]. Also, `parcours` linked an increase of gestation period [3] to the loss of the gurgle, and gain of the prusten, respectively (rows 2–3 of Table 1). Finally, the concomitant gain of grunt and roar (row 4 of Table 1 and Fig 4a) is consistent with the observation that the grunt typically occurs in a roaring sequence [36].

The `parcours` tool detected a correlation of 0.5 (row 5 of Table 1) with large increases in gestation period (rank 4 to 1) and a moderate increase in weight (rank 3 to 2). This is largely consistent with observations that body weight and gestation period in mammals are loosely correlated [1]. Felids exhibit considerable variability in body weight (> 200 kg to < 5 kg) and skull length (> 30 cm to < 10 cm). Strong allometric interspecific relationships between body size (weight) and skull length has been demonstrated in a range of non-human mammals [14]. Consistent with this, `parcours` detected a correlation of 0.57

(row 6 of Table 1) between an increase in both weight ($4 \to 3$) and skull length ($3 \to 2$). A correlation of 0.64 was also detected between a decrease in weight ($1 \to 2$) and skull length ($1 \to 2$) in Pantherinae (row 7 of Table 1). Interestingly, all co-occurrences between these two characters are on the branches going into Leopard and Snow Leopard. Not surprisingly, `parcours` picked up a pairwise correlation of 0.71 (row 8 of Table 1) between a decrease in skull width ($2 \to 3$) and gestation period ($1 \to 3$). As we see in Fig. 4b, this was due to a recurrent decrease in both traits near the common ancestor of Felidae, and again inside the Puma lineage, due to the much smaller size of the Jaguarundi, compared to the other taxa (Puma and Cheetah) of this lineage. Finally, `parcours` detected a correlation of 0.67 (row 9 of Table 1) of an increase ($4 \to 3$) in the ratio of skull length and skull width and dental profile. Since the dental formula of 28–30 teeth (rank 2) is somewhat nebulous, this particular result may be an artifact of this ranking and needs further assessment.

In conclusion, we propose a new approach to exploring correlated evolution among discrete characters in an 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 dataset of felid vocalizations and morphological characteristics. While `parcours` already discovers trends in this 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 $4 \to 1$ more expensive than $2 \to 1$), could even further improve these results. More generally, by taking into account multiplicities of events on the branches (from different parsimonies) in computing weighted Jaccard index (Eq. 4), this can be larger (row 6 of Table 1) or smaller (row 7) than Jaccard index. Since each index has its advantages and disadvantages, a full exploration of similarity indexes [37] is the subject of future work. While we consider the set of all parsimonies in this framework, another idea is to allow the user to place constraints on this set, based on a priori knowledge that a certain state is ancestral, for example, or based on the least common ancestor of a given set of states. Finally, while the threshold of 0.5 used in this experiment is somewhat arbitrary, an interesting future work is to devise a principled approach to determining *significant* correlation between pairs of traits, possibly by using ideas from [23].

References

1. Atanasov, A.: Allometric relationship between the length of pregnancy and body weight in mammals. Bulgarian Journal of Veterinary Medicine **8**(1), 13–22 (2005)
2. Barker, D., et al.: Constrained models of evolution lead to improved prediction of functional linkage from correlated gain and loss of genes. Bioinformatics **23** (2007)
3. Castelló, J.: Felids and Hyenas of the World. Princeton University Press (2020)
4. Ciccolella, S., et al.: Inferring cancer progression from Single-Cell Sequencing while allowing mutation losses. Bioinformatics (2020)
5. Clemente, J., Ikeo, K., Valiente, G., Gojobori, T.: Optimized ancestral state reconstruction using sankoff parsimony. BMC Bioinformatics **10**(1) (2009)

6. Cohen, O., et al.: CoPAP:. Nucleic Acids Research **41**, 232–237 (2013)
7. Drummond, et al.: BEAST:. BMC Evolutionary Biology **7**(214) (2007)
8. Edwards, et al.: The reconstruction of evolution. Ann. Hum. Gen. **27** (1963)
9. Ehret: Development of sound communication in mammals. In: Anim. Behav. (1980)
10. Felsenstein: Evolutionary trees from DNA sequences:. Mol. Evol. **17** (1981)
11. Felsenstein: PHYLIP v.3.2. Cladistics **5**, 164–166 (1989)
12. Felsenstein, J.: Cases in which parsimony or compatibility methods will be positively misleading. Systematic Zoology **27**(4), 401–410 (1978)
13. Figueiró, H.V., et al.: Genome-wide signatures of complex introgression and adaptive evolution in the big cats. Science Advances **3**(e1700299) (2017)
14. Fitch, W.T.: Skull dimensions in relation to body size in non-human mammals: The causal bases for acoustic allometry. Zoology **103**, 40–58 (2000)
15. Fitch, W.M.: Toward defining the course of evolution: Minimum change for a specific tree topology. Systematic Zoology **20**(4), 406–416 (1971)
16. Guéguen, et al.: Bio++:. Molecular Biology and Evolution **30**(8), 1745–1750 (2013)
17. Hajirasouliha, I., Raphael, B.J.: Reconstructing mutational history in multiply sampled tumors using perfect phylogeny mixtures. In: WABI (2014)
18. Huelsenbeck, J.P., Ronquist, F.: MRBAYES:. Bioinformatics **17** (2001)
19. Jaccard: Nouvelles recherches sur la distribution florale. Soc. Vaud. **44** (1908)
20. Jahn, et al.: Tree inference for single-cell data. Genome Biology **17** (2016)
21. Johnson, et al.: The late miocene radiation of modern felidae: A genetic assessment. Science **311**(5757), 73–77 (2006)
22. Kolaczkowski, B., Thornton, J.: Performance of maximum parsimony and likelihood phylogenetics when evolution is heterogeneous. Nature **431** (2004)
23. Maddison, W.P.: A method for testing the correlated evolution of two binary characters. Evolution **44**(3), 539–557 (1990)
24. Marcotte, E.M., et al.: Localizing proteins in the cell from their phylogenetic profiles. Proc. National Academy of Sciences **97** (2000)
25. Mossige, S.: An algorithm for Gray codes. Computing pp. 89–92 (1977)
26. Patterson, M., Szöllősi, G., Daubin, V., Tannier, E.: Lateral gene transfer, rearrangement, reconciliation. BMC Bioinformatics **14**(S4) (2013)
27. Peters, G.: Vergleichende Untersuchung zur Lautgebung einiger Feliden (Mammalia, Felidae). Spixiana **1**, 1–283 (1978)
28. Peters, G.: On the structure of friendly close range vocalizations in terrestrial carnivores. Zeitschrift für Säugetierkunde **49**, 157–182 (1984)
29. Peters, G.: Vocal communication in cats. In: Great Cats–. pp. 76–77 (1991)
30. Peters, G., Tonkin-Leyhausen, B.A.: Evolution of acoustic communication signals of mammals:. Journal of Mammalian Evolution **6**(2), 129–159 (1999)
31. Ružička, M.: Anwendung mathematisch-statistischer methoden in der geobotanik (synthetische bearbeitung von aufnahmen). Biológia **13**, 647–661 (1958)
32. Sankoff, D.: Minimal mutation trees of sequences. SIAM J. Appl. Math. **28** (1975)
33. Schwartz, R., Schäffer, A.: The evolution of tumour phylogenetics: principles and practice. Nature Reviews Genetics **18**(4), 213–229 (2017)
34. Stamatakis, A.: RAxML-VI-HPC:. Bioinformatics **22**(21), 2688–2690 (2006)
35. Steel, M., Penny, D.: Two further links between MP and ML under the poisson model. Applied Mathematics Letters **17**(7), 785–790 (2004)
36. Sunquist, M., Sunquist, F.: Wild Cats of the World (2002)
37. Warrens, M.J.: Inequalities between similarities for numerical data. Journal of Classification **33**, 141–148 (2016)