Parsimony via Consensus

TREVOR C. BRUEN¹ AND DAVID BRYANT²

¹Department of Mathematics, University of California, Berkeley, USA; E-mail: tbruenc@math.berkeley.edu
²Department of Mathematics, University of Auckland, New Zealand; E-mail: d.bryant@auckland.ac.nz

Abstract.—The parsimony score of a character on a tree equals the number of state changes required to fit that character onto the tree. We show that for unordered, reversible characters this score equals the number of tree rearrangements required to fit the tree onto the character. We discuss implications of this connection for the debate over the use of consensus trees or total evidence and show how it provides a link between incongruence of characters and recombination. [Compatibility; homoplasy; maximum parsimony; SPR distance.]

The parsimony length of an unordered, reversible character on a tree equals the minimum number of state changes required to fit the character onto a tree. Here we introduce an alternative way to view parsimony length, one that is dual to the original definition. We show how the parsimony length of a character equals the minimum number of changes in the tree required to fit the tree onto the character. This new perspective on parsimony has applications to consensus trees and recombination detection.

The first application we discuss is how the reformulation of parsimony provides a closer link between parsimony-based analysis and supertree methods. We demonstrate that the maximum parsimony tree can be viewed as a type of median consensus tree, where the median is computed with respect to the SPR distance (see below). As well, the result shows how to conduct a parsimony-based analysis not just on characters but on trees, without having to recode the trees as binary character matrices. This opens the way to a hybrid between the consensus approach and the total evidence approach, where the data are a mixture of characters, trees, and subtrees.

The second application of the main result is to the analysis of pairs of characters. We show that the score of the maximum parsimony tree for two characters is a simple function of the smallest number of recombinations required to explain the incongruence between the characters without homoplasy. This result provides the basis of a highly efficient test for recombination (Bruen et al., 2006).

Here and throughout the paper we assume that all phylogenetic trees are fully resolved (bifurcating) and that by “parsimony” we refer to Fitch or Wagner parsimony, where the character states are unordered and reversible. Some of the results presented here can be extended to other forms of parsimony (Bruen, 2006), and possibly to incompletely resolved trees, but these lie beyond the scope of this paper.

Note that in this paper we are dealing with unrooted SPR rearrangements, which are those used in tree searches. There is a related, but distinct, concept of rooted SPR rearrangements, where the rearrangements are restricted to obey a type of temporal constraint (Song, 2003). It is this latter class of rooted SPR rearrangements that are used to model lateral gene transfers and recombination. It would be a worthwhile, but challenging, goal to investigate whether any of the results on unrooted SPR rearrangements in this paper can be extended to rooted SPR rearrangements.

Linking Parsimony with SPR

A subtree-prune and regraft (SPR) rearrangement is an operation on phylogenetic trees whereby a subtree is removed from one part of the tree and regrafted to another part of the tree, see Figure 1 (Felsenstein, 2004; Swoford et al., 1996). These SPR rearrangements are widely used by tree searching software packages like PAUP (Swofford, 1998) and Garli (Zwickl, 2006). The SPR distance between two trees can be defined as the minimum number of SPR rearrangements required to transform one tree into the other (Hein, 1990; Allen and Steel, 2001; Goloboff, 2007). For example, the two trees $T_1$ and $T_2$ in Figure 1 can be transformed into each other using a minimum of two SPR rearrangements, via the tree $T_3$, so their SPR distance is two.

The parsimony length of a character on a tree is the minimum of steps required to fit that character on the tree, as computed by the algorithm of Fitch (1971). The length of a character $X_i$ on a tree $T$ is denoted $\ell(X_i, T)$. A character with $r_i$ states therefore has parsimony length at least $(r_i - 1)$, as every state not at the root has to arise at least once. A character is compatible with a tree if it requires at most $(r_i - 1)$ changes on that tree (Felsenstein, 2004).

Instead of fitting a character onto a tree we could just as well fit the tree onto the character. If the character and the tree are compatible, then we have a perfect fit. When there is not a perfect fit we can measure how many SPR rearrangements are required to give a tree that does make a perfect fit. It turns out that this measure gives an equivalent score to parsimony length. More formally:

**Theorem 1.** Let $X_i$ be a character with $r_i$ states and let $T$ be a fully resolved phylogenetic tree. It takes exactly $\ell(X_i, T) - (r_i - 1)$ SPR rearrangements to transform $T$ into a tree compatible with $X_i$. The result still holds if $X_i$ has some missing states.
As an example, consider the character $X_1$ mapping taxa $A, C, D, F$ to one and $B, E, G$ to zero. The length of this character on tree $T_1$ of Figure 1 is three, and the number of SPR rearrangements needed to transform $T_1$ onto some tree $T_3$ compatible with $X_1$ is two. Note that there could be other trees compatible with $X_1$ that are further than two SPR rearrangements away: the result only gives the number of rearrangements required to obtain the closest tree.

Once stated, the theorem is not too difficult to prove. First show that performing an SPR rearrangement decreases the length by at most one step. Hence it takes at least $\ell(X_1, T) - (r_i - 1)$ SPR rearrangements to transform $T$ into a tree compatible with the character $X_i$. Then show that this is the minimum required. A formal proof is presented in Appendix 1. A restricted (binary character) version of this theorem was proved in Bryant (2003).

The theorem captures an issue that is central to the interpretation of incongruence: whether an observed incongruence is to be explained by positing homoplasy or by modifying the tree. Define the SPR distance from a tree $T$ to a character $X_i$ to be the SPR distance from $T$ to the closest tree $T'$ that is compatible with $X_i$. Theorem 1 then tells us that the SPR distance from $T$ to a character $X_i$ is equal to the difference between the length $\ell(X_i, T)$ of $X_i$ on $T$ and the minimum possible length of $X_i$ on any tree. That is, the amount that the length of a character exceeds the minimum possible length $(r_i - 1)$ is equal to the amount we have to modify $T$ in order to remove all incongruence between $T$ and $X_i$.

**Consensus Trees, Supertrees, and Parsimony**

In their insightful overview of supertree methods, Thorley and Wilkinson (2003) characterize a family of supertree methods that all minimize a sum of the form

$$\sum_{i=1}^{n} d(T, t_i) = d(T, t_1) + d(T, t_2) + \cdots + d(T, t_n).$$

(1)
Here \( t_1, t_2, \ldots, t_n \) are the input trees and \( d(T, t_i) \) is a measure of the distance between the input tree \( t_i \) with the supertree \( T \). There are many choices for the distance measure \( d \), and it need not be the case that the distance measure satisfies the symmetry condition \( d(T, t_i) = d(t_i, T) \). Gordon (1986) was the first to propose this description of supertrees. Many supertree methods can be described in these terms, including matrix representation with parsimony (MRP) (Baum, 1992; Ragan, 1992), minimum flip supertrees (Chen et al., 2006), the median supertree (Bryant, 1997), majority rule supertree (Cotton and Wilkinson, 2007) and the average consensus supertree (Lapointe and Cucumel, 1997).

Let \( ds(T, X_i) \) denote the SPR distance from \( T \) to the closest fully resolved tree \( T_i \) that is compatible with \( X_i \). By Theorem 1, a maximum parsimony tree for \( X_1, \ldots, X_m \) is one that minimizes the expression

\[
\sum_{i=1}^{m} ds(T, X_i) = ds(T, X_1) + ds(T, X_2) + \cdots + ds(T, X_m).
\]  

(2)

In this way, maximum parsimony is a form of median consensus. The significance of this observation does not come from the fact that we can write the the parsimony score of \( T \) in the form (2); it is from the close connection with SPR distances, and from the way we will now use this connection to combine different kinds of data in the same theoretical framework.

An SPR median tree for fully resolved trees \( t_1, \ldots, t_n \) on the same leaf set is a tree \( T \) that minimizes

\[
\sum_{i=1}^{n} ds(T, t_i) = ds(T, t_1) + ds(T, t_2) + \cdots + ds(T, t_n),
\]

where here \( d(T, t_i) \) denotes the SPR distance from \( T \) to \( t_i \) (Hill, 2007). We extend this directly to a supertree method by mimicking the situation for characters. Suppose that \( t_i \) is a phylogenetic tree, not necessarily fully resolved, on a subset of the set of leaves. We say that a fully resolved tree \( T \) on the full set of leaves is compatible with \( t_i \) (equivalently, \( T \) displays \( t_i \)) if we can obtain \( t_i \) from \( T \) by pruning off leaves and contracting edges. In this general situation, we let \( ds(T, t_i) \) denote the SPR distance from \( T \) to the closest fully resolved tree \( T_i \) that is compatible with \( t_i \). This is equivalent to pruning leaves off \( T \) and then computing the distance from this pruned tree to \( t_i \).

Now suppose that we have both characters and trees in the input. Both types of phylogenetic data can summarized by an SPR median tree \( T \), chosen to minimize the sum

\[
\sum_{i=1}^{n} ds(T, t_i) + \sum_{i=1}^{m} ds(T, X_i).
\]

We have, then, a way to bring together both the supertree/consensus methodology and the total evidence methodology. In the case that the data comprise only trees, the tree is a median supertree; in the case that the data comprise only character data, the tree is the maximum parsimony tree.

It is important to note the difference between this approach and the MRP method (Baum, 1992; Ragan, 1992), which could also be used to combine trees and characters. In MRP, the trees are broken down into multiple independent characters. This is a problem, because the characters encoding a tree are nowhere near independent. In contrast, the SPR median tree approach treats a tree as a single indivisible unit of information.

There is one critical issue that has been side-stepped: computation time. At present, computational limitations make the construction of SPR median trees infeasible for all but the smallest data sets. Computing the SPR distance between two trees is an NP-hard problem (Hickey et al., 2007) and exact special case algorithms (Hickey et al., 2006) that could be applied to the problem. Below we describe a lower bound method for the SPR distance that should also aid construction of these SPR median trees.

**PARSIMONY ON PAIRS OF CHARACTERS**

Another valuable application of Theorem 1 follows when we consider parsimony analysis of just two unordered and reversible characters. The concept of pairwise character compatibility was introduced by Le Quesne (1969); see also Felsenstein (2004). Two binary undirected characters with states 0 and 1 are incompatible if and only if all four combinations of 00, 01, 10, and 11 are present as combination of states for the two characters. In a standard setting, character incompatibility is interpreted as implying that at least one of the characters has undergone convergent or recurrent mutation. In other words, for every possible phylogeny describing the history of the two characters, at least one homoplasy is posited for one of the characters (assuming the characters were correctly scored). Another interpretation of incompatibility of two characters is that characters evolved without homoplasy on two different phylogenies, where the phylogenies differ by one or more SPR rearrangement (Sneath et al., 1975; Hudson and Kaplan, 1985).

Define the total incongruence score \( i(X_1, X_2) \) for two multistate unordered characters \( X_1 \) and \( X_2 \) (with \( r_1 \) and \( r_2 \) states respectively) as

\[
i(X_1, X_2) = \min_{\ell} [\ell(X_1, T) + \ell(X_2, T)] - (r_1 - 1) - (r_2 - 1).
\]

(3)

This is the maximum parsimony score of the two characters \( X_1, X_2 \) minus the minimum number of changes required for each character. Equation (3) generalizes the incompatibility notion for two binary characters. It is also equivalent to the incongruence length difference statistic applied to only two characters (Farris et al., 1995). Importantly, the total incongruence score can be computed rapidly (Bruen and Bryant, 2008). The following
consequence of Theorem 1 strengthens the connection between incongruence and SPR rearrangements.

**Theorem 2.** The total incongruence score \( i(X_1, X_2) \) for two characters equals the minimum SPR distance between a tree \( T_1 \) and a tree \( T_2 \) such that \( T_1 \) is compatible with \( X_1 \) and \( T_2 \) is compatible with \( X_2 \).

Although the notion of total incongruence for two characters has been considered before in the context of character selection and weighting (Penny and Hendy, 1986), it has not been considered in the context of genealogical similarity. Essentially, Theorem 2 shows that the total incongruence score equals the minimum possible number of SPR rearrangements that could have occurred between the phylogenetic histories for both characters, assuming that the characters have different histories with which they are each perfectly compatible.

Indeed, Theorem 2 suggests a natural way to interpret genealogical similarity between two characters, which we have used to develop a powerful test for recombination (Bruen et al., 2006). Choosing two characters from two different genes (which have possibly different histories) gives a simple approach to identify the distinctiveness of the histories of the genes.

We can also apply Theorem 2 to obtain a lower bound on the SPR distance between two trees. Suppose that we have two trees \( T_1 \) and \( T_2 \) and we wish to obtain a lower bound on the SPR distance \( d(T_1, T_2) \) between the two trees. If we choose any character \( X_1 \) compatible with \( T_1 \) and any character \( X_2 \) compatible with \( T_2 \) then, by Theorem 2, we have that \( i(X_1, X_2) \leq d(T_1, T_2) \). By carefully choosing \( X_1 \) and \( X_2 \) we can obtain tighter bounds.

**DISCUSSION AND EXTENSIONS**

We have presented a reformulation of parsimony that is complementary or dual to the standard definitions. Instead of measuring how well a character fits onto a tree, we look at how well the tree fits onto the character. A consequence of this new perspective is that we can combine trees and character data using one general SPR framework, and we also obtain new results connecting incongruence measures and recombination. Nevertheless, it is not immediately clear how the new reformulation can be interpreted in itself.

Consider the information a single character, or tree, represents. Given a single character, we can define the set or cloud of trees around the character comprising exactly those trees compatible with the character (Fig. 2). If the character evolved without homoplasy, then the true evolutionary tree must be contained somewhere within the cloud. Now suppose we have multiple characters, each with its own cloud. There may not be a single tree contained in the intersection of all of these clouds.

![Figure 2](https://academic.oup.com/sysbio/article-abstract/57/2/251/1623283)  
**Figure 2.** Cartoon representation of parsimony in terms of tree rearrangements. Each character \( X_i \) gives a “cloud” of trees containing those trees compatible with \( X_i \). The distance from \( T \) to a cloud is the SPR distance \( d(T, t_i) \) to the tree \( t_i \) in the cloud that is closest to \( T \). The maximum parsimony tree is then the tree closest to these clouds under the SPR distance.
Instead, we search for a tree that is as close as possible to all of the clouds. The distance from $T$ to the cloud associated to character $X_i$ is exactly $d(T, X_i)$, so by Theorem 1 a tree closest to all of the clouds is a maximum parsimony tree.

We note that several of the results in this article can be extended; for details, see Bruen (2006). Both Theorems 1 and 2 remain valid if we replace the SPR distance with the tree bisection and reconnection (TBR) distance. In a TBR rearrangement, a subtree is removed from the tree and then reattached elsewhere with the tree, the difference with SPR being that with SPR we can only reattach the subtree at the node where the branch was broken, whereas with TBR we can reattach from any node in the subtree (Allen and Steel, 2001; Felsenstein, 2004). The TBR distance between two trees is the minimum number of TBR rearrangements required to transform one tree into the other.

That Theorems 1 and 2 hold for the TBR distance may appear surprising because the TBR distance between two trees is always less than, or equal to, the SPR distance between the trees. However, the extension follows by a small change to the proof of Theorem 1, noting that a TBR move can only reduce the parsimony score of a character by at most one.

We have also explored extensions of the result to other distances between trees, notably the Robinson-Foulds distance and the nearest neighbor interchange distance. See Bruen (2006) for details.

ACKNOWLEDGEMENTS

We would like to thank Mike Steel, Sebastien Böcker, Olaf Bininda-Emonds, Pablo Goloboff, Melanie Pierson, Mark Wilkinson, and an anonymous referee for their valuable suggestions. This research was partially supported by the New Zealand Marsden Fund.

REFERENCES

Allen, B., and M. Steel. 2001. Subtree transfer operations and their induced metrics on evolutionary trees. Ann. Comb. 5:1–13.

Baum, B. 1992. Combining trees as a way of combining datasets for phylogenetic inference, and the desirability of combining gene trees. Taxon 41:3–10.

Bruen, T. 2006. Discrete and statistical approaches to genetics. PhD thesis, McGill University School of Computer Science, Montreal, Canada.

Bruen, T., and D. Bryant. 2008. A subdivision approach to maximum parsimony. Ann. Comb. 12:45–51.

Bruen, T., H. Philippe, and D. Bryant. 2006. A simple and robust statistical test to detect the presence of recombination. Genetics 172:1–17.

Bryant, D. 1997. Building trees, hunting for trees and comparing trees. Ann. Comb. 1:45–51.

Bryant, D. 2003. A classification of consensus methods for phylogenetic trees. Mol. Biol. Evol. 3:403–417.

Ragan, M. A. 1992. Phylogenetic inference based on matrix representations of trees. Mol. Phylog. Evol. 1:53–58.

Sneath, P., M. Sackin, and R. Ambler. 1975. Detecting evolutionary incompatibilities from protein sequences. Systematic Zoology 24:311–332.

Song, Y. S. 2003. On the combinatorics of rooted binary phylogenetic trees. Ann. Comb. 7:365–379.

Swofford, D., G. Olsen, P. Waddell, and D. Hillis. 1996. Molecular systematics, Pages 407–514 in Phylogenetic inference. Sinauer Associates, Sunderland, Massachusetts.

Swofford, D. L. 1998. PAUP*: Phylogenetic analysis using parsimony (*and other methods). Sinauer Associates, Sunderland, Massachusetts.

Thorley, J. L., and M. Wilkinson. 2003. A view of supertree methods. Pages 185–194 in Bioconsensus (M. F. Janowitz, F.-J. Lapointe, F. R. McMorris, B. Mirkin, and F. S. Roberts, eds.), volume 61 of DIMACS, Series in Discrete Mathematics and Theoretical Computer Science, The American Mathematical Society, New York.

Zwickl, D. 2006. Genetic algorithm approaches for the phylogenetic analysis of large biological sequence datasets under the maximum likelihood criterion. PhD thesis, University of Texas at Austin.

First submitted 3 April 2006; reviews returned 26 August 2006; final acceptance 6 June 2007

APPENDIX

The first observation is that a TBR rearrangement of a tree increases the length of a character by at most one. As SPR rearrangements are a special case of TBR rearrangements, the same result holds for SPR.

Lemma 1. Let $T$ be a fully resolved phylogenetic tree and $X$, an unordered reversible character. Let $T'$ be a phylogenetic tree that differs from $T$ by a single TBR rearrangement. Then $\ell(X, T') - \ell(X, T) \leq 1$.

Proof. The proof of Lemma 5.1 in Bryant (2004) for binary characters applies directly to the multistate case.

Let $d_{STR}(T, T')$ denote the unrooted SPR distance between two phylogenetic trees $T$ and $T'$.

Theorem 1 Let $X_t$ be a character with $\tau$ states and let $T$ be a fully resolved phylogenetic tree. It takes exactly $(\ell(X, T) - (\tau - 1))$ SPR rearrangements to transform $T$ into a tree compatible with $X_t$. The result still holds if $X$ has some missing states.
Proof. Let $T'$ be a fully resolved phylogenetic tree compatible with $X$, for which $d_{SPR}(T', T)$ is minimized and let $m = d_{SPR}(T', T)$. Then there exists a sequence of trees $T = T_0, \ldots, T_m = T$ such that every adjacent pair of trees in the sequence differs by exactly one SPR re-arrangement. By Lemma 1 the existence of this sequence implies that $\ell(T, X) - \ell(T', X) \leq d_{SPR}(T, T')$ and because $X$ is compatible with $X$, we have $\ell(T', X) = r_i - 1$, giving

$$\ell(T, X) - (r_i - 1) \leq d_{SPR}(T, T').$$

For the other direction, we show that we can construct a sequence of $\ell(T, X) - (r_i - 1)$ SPR rearrangements that transform $T$ into a tree $T'$ compatible with $X$. Firstly, if $\ell(T, X) - (r_i - 1) = 0$, then $T$ is compatible with $X$, so the proof is finished. Otherwise, let $\tilde{X}$ be an assignment of states to internal nodes that minimizes the number of state changes (that is, a minimum extension of $X$). Then because $X$ is not compatible with $T$ there exist three vertices $u, v$, and $w$, where $(u, v) \in E(T)$, $v$ lies on the path from $u$ to $w$ and $\tilde{X}(u) = \tilde{X}(w) \neq \tilde{X}(v)$. Perform an SPR rearrangement by removing edge $(u, v)$, suppressing the $v$ vertex and creating a new edge $(u, l)$ where $l$ is a new vertex on an edge adjacent to $u$. Furthermore, set $\tilde{X}(l) = \tilde{X}(w)$. Then the number of edges on which a change has occurred has decreased by 1, thereby decreasing the parsimony length by 1. This procedure can be repeated until the parsimony length equals $r_i - 1$, constructing the desired sequence of trees and completing the proof.

**Theorem 2.** The total incongruence score $i(X_1, X_2)$ for two characters equals the minimum SPR distance between a tree $T_1$ and $T_2$ such that $X_1$ is compatible with $T_1$ and $X_2$ is compatible with $T_2$.

Proof. Let $T_1$ and $T_2$ be any two trees compatible with $X_1$ and $X_2$, respectively. Then $\ell(X_1, T_1) = r_i - 1$ and by Theorem 1, $\ell(X_i, T_1) - (r_i - 1) \leq d_{SPR}(T_1, T_2)$. We have then that

$$i(X_1, X_2) \leq \ell(X_1, T_1) + \ell(X_2, T_1) - (r_i - 1) - (r_j - 1) \leq d_{SPR}(T_1, T_2)$$

and so $i(X_1, X_2)$ is a lower bound for $d_{SPR}(T_1, T_2)$.

We show that this bound can be achieved. Let $T$ be a maximum parsimony tree for the pair of characters $X_1, X_2$. By Theorem 1 there exist two trees $T_1$ and $T_2$ such that $T_1$ is compatible with $X_1$, $T_2$ is compatible with $X_2$, and

$$d_{SPR}(T_1, T) + d_{SPR}(T_2, T) = i(X_1, X_2),$$

implying that $d_{SPR}(T_1, T_2) \leq d_{SPR}(T_1, T) + d_{SPR}(T_2, T) = i(X_1, X_2)$ and hence

$$d_{SPR}(T_1, T_2) = i(X_1, X_2).$$