Maximum Likelihood Supertrees

Mike Steel, Allen Rodrigo

Syst. Biol. 57(2):243–250, 2008

ISSN: 1063-5157 print / 1076-836X online
DOI: 10.1080/10635150802033014

Maximum Likelihood Supertrees

MIKE STEEL[1] AND ALLEN RODRIGO[2]

[1]*Allan Wilson Centre for Molecular Ecology and Evolution, Department of Mathematics and Statistics, University of Canterbury, Christchurch, New Zealand; E-mail: m.steel@math.canterbury.ac.nz*
[2]*The Bioinformatics Institute and the Allan Wilson Centre for Molecular Ecology and Evolution, University of Auckland, New Zealand, and Laboratoire d'Informatique, de Robotique et de Microelectronique de Montpellier, France; E-mail: a.rodrigo@auckland.ac.nz*

Abstract.—We analyze a maximum likelihood approach for combining phylogenetic trees into a larger "supertree." This is based on a simple exponential model of phylogenetic error, which ensures that ML supertrees have a simple combinatorial description (as a median tree, minimizing a weighted sum of distances to the input trees). We show that this approach to ML supertree reconstruction is statistically consistent (it converges on the true species supertree as more input trees are combined), in contrast to the widely used MRP method, which we show can be statistically inconsistent under the exponential error model. We also show that this statistical consistency extends to an ML approach for constructing species supertrees from gene trees. In this setting, incomplete lineage sorting (due to coalescence rates of homologous genes being lower than speciation rates) has been shown to lead to gene trees that are frequently different from species trees, and this can confound efforts to reconstruct the species phylogeny correctly. [Gene tree; maximum likelihood; phylogenetic supertree; species tree; statistical consistency.]

Combining trees on different, overlapping sets of taxa into a parent "supertree" is now a mainstream strategy for constructing large phylogenetic trees. The literature on supertrees is growing steadily: new methods of supertree reconstruction are being developed (Cotton and Wilkinson, 2007) and supertree analyses are shedding light on fundamental evolutionary questions (Bininda-Emonds et al., 2007). Despite this surge in research activity, it is probably fair to say that biologists are still confused about what supertrees really are and what it is we do when we build a supertree. Are we, as some maintain, simply summarizing the phylogenetic information contained in a group of subtrees? Or are we trying to derive the best estimate of phylogeny given the information at hand? Nor is it clear which of these two conceptually different objectives underpin the various supertree reconstruction methods.

We take the view that what biologists really want a supertree reconstruction method to deliver is the best hypothesis of evolutionary relationships that can be inferred from the data available. Obviously, it is not the case that the supertree constructed as a summary statistic will necessarily be the best estimate of phylogeny. Nonetheless, if we are prepared to consider supertree reconstruction a problem of phylogenetic *estimation*, we have at our disposal an arsenal of phylogenetic tools and methods that have been tried and tested. Matrix representation with parsimony (MRP; Baum and Ragan, 1992), weighted MRP (Bininda-Emonds and Sanderson, 2001), matrix representation with compatibility (MRC; Rodrigo, 1996; Ross and Rodrigo, 2004), and, most recently, Bayesian supertree reconstruction (BSR; Ronquist et al., 2004) are undoubtedly inspired by standard phylogenetic methods. A gap remains, though, as there has been remarkably little development of likelihood-based methods for supertree reconstruction.

In this paper, we analyze an approach to obtain maximum likelihood (ML) estimates of supertrees, based on a probability model that permits errors in subtree topologies. The approach is of the type described by Cotton and Page (2004), and it permits supertrees to be estimated even if there is topological conflict amongst the constituent subtrees. We show that ML estimates of supertrees so obtained are statistically consistent under fairly general conditions. By contrast, we show that MRP may be inconsistent under these same conditions. We then consider a further complication that arises in the supertree setting when combining gene trees into species trees—in addition to the possibility that the input gene trees are reconstructed incorrectly (either a consequence of the reconstruction method used, or some sampling error), there is a further stochastic process that leads to the (true) gene trees differing from their underlying species tree (a consequence of incomplete lineage sorting). Although simple strategies such as gene concatenation have recently been shown to be potentially misleading (Degnan and Rosenberg, 2006), we show that an ML supertree approach for combining gene trees is also statistically consistent.

Terminology

Throughout this paper, unless stated otherwise, phylogenetic trees may be either rooted or unrooted, and we will mostly follow the notation of Semple and Steel (2003). In particular, given a (rooted or unrooted) phylogenetic tree $\mathcal{T}$ on a finite set X of taxa (which will always label the leaves of the tree), any subset Y of X induces a phylogenetic tree on taxon set Y, denoted $\mathcal{T}|Y$, which, informally, is the subtree of $\mathcal{T}$ that connects the taxa in Y only. In the *supertree* problem, we have a sequence $\mathcal{P} = (\mathcal{T}_1, \mathcal{T}_2, \ldots, \mathcal{T}_k)$ of input trees, called a *profile*, where $\mathcal{T}_i$ is a phylogenetic tree on taxon set X_i. We wish to combine these trees into a phylogenetic tree $\mathcal{T}$ on the union of the taxon sets (i.e., $X = X_1 \cup X_2 \cup \cdots \cup X_k$). We assume that the trees in $\mathcal{P}$ are either all rooted or all unrooted, and that $\mathcal{T}$ is rooted or unrooted accordingly. We will mostly assume that trees are *fully resolved* (i.e., binary trees, without polytomies); in a remark following Theorem 4, we briefly describe how this restriction can

be lifted. Furthermore, in this paper we consider just the tree topology, not the branch lengths.

A special case of the supertree problem arises when the taxon sets of the input trees are all the same ($X_1 = X_2 = \cdots = X_k$). This is the much studied *consensus tree* problem. In an early paper, McMorris (1990) described how, in this consensus setting, the majority-rule consensus tree can be given a maximum likelihood interpretation. However, this approach is quite different from the one described here (it is restricted to the consensus setting and is based on a very particular probability model).

In this paper, we will denote the underlying ("true") species tree as $\mathcal{T}_0$ (assuming that such a tree exists and that the evolution of the taxa has not involved reticulate processes such as the formation of hybrid taxa). In an ideal world, we would like $\mathcal{T}_i = \mathcal{T}_0|X_i$ for each tree $\mathcal{T}_i$ in the profile—that is, we would like each of the reconstructed trees to be identical to the subtree of the true tree for the taxa in X_0. But in practice, the trees $\mathcal{T}_1, \ldots, \mathcal{T}_k$ are unlikely to even be compatible (i.e., no phylogenetic tree $\mathcal{T}$ exists for which $\mathcal{T}_i = \mathcal{T}|X_i$ for all i).

An Exponential Model of Phylogenetic Error

Species trees that have been inferred from data may differ from the true underlying species tree for numerous reasons, including sampling effects (short and/or site-saturated sequences, or poorly defined characters), model violation, sequencing or alignment errors, and so forth. In this section, we will assume a simple model of phylogenetic error in which the probability of observing a given tree falls off exponentially with its distance from an underlying generating tree (e.g., the true species tree $\mathcal{T}_0$). This type of model has been described by Holmes (2003) in the general setting of Bayesian statistical analysis in phylogenetics. Suppose d is some metric on resolved phylogenetic trees. There are several possible choices for d, depending on the biological context and computational considerations, and we describe some of these shortly. In the exponential model, the probability, denoted $\mathbb{P}_{\mathcal{T},Y}[\mathcal{T}']$ (or more briefly $\mathbb{P}_{\mathcal{T}}[\mathcal{T}']$) of reconstructing any species tree $\mathcal{T}'$ on any taxon set Y (where $Y \subset X$), when $\mathcal{T}$ is the generating tree (on taxon set X) is proportional to an exponentially decaying function of the distance from $\mathcal{T}'$ to $\mathcal{T}|Y$. In other words,

$$\mathbb{P}_{\mathcal{T},Y}[\mathcal{T}'] = \alpha \exp[-\beta d(\mathcal{T}', \mathcal{T}|Y)]. \quad (1)$$

The constant β can vary with Y and other factors (such as the quality of the data); for example, trees constructed from long high-fidelity sequences are likely to have a larger β than trees constructed from shorter and/or noisier sequences. The constant α is simply a normalizing constant to ensure that $\sum_{\mathcal{T}'} \mathbb{P}_{\mathcal{T},Y}[\mathcal{T}'] = 1$, where the sum is over all fully resolved phylogenetic trees $\mathcal{T}'$ on taxon set Y. When we have a sequence $(X_1, X_2, \ldots)$ of subsets of X, we will reflect the dependence of α, β on X_i by writing α_i and β_i. Note that α_i is determined entirely by β_i and $|X_i|$.

Note that, implicit in (1), the probability of $\mathcal{T}'$ depends only on the subtree of $\mathcal{T}$ connecting the species in $\mathcal{T}'$ and not on the other species in $\mathcal{T}$ that are not present in $\mathcal{T}'$.

Now, suppose we observe the profile of trees $\mathcal{P} = (\mathcal{T}_1, \mathcal{T}_2, \ldots, \mathcal{T}_k)$ as above, where $\mathcal{T}_i$ has leaf set X_i. Assume that, for each i, the tree $\mathcal{T}_i$ has been independently sampled from the exponential distribution (1) with $\beta = \beta_i$. Select a phylogenetic X tree $\mathcal{T}$ that maximizes the probability ($\mathbb{P}_{\mathcal{T}}(\mathcal{T}_1, \ldots, \mathcal{T}_k)$) of generating the observed profile $\mathcal{P}$—we call this type of a tree $\mathcal{T}$ an *ML supertree for* $\mathcal{P}$.

An ML supertree has a simple combinatorial description as a (weighted) median tree, as the following result shows. Note that the use of sums of tree-to-tree distances as optimality criteria has previously been suggested by Wilkinson et al. (2001).

Proposition 1. *For any metric d on phylogenetic trees, an ML supertree for a profile $\mathcal{P}$ under the exponential model (1), is precisely a tree $\mathcal{T}$ that minimizes the weighted sum:*

$$\sum_{i=1}^{k} \beta_i d(\mathcal{T}_i, \mathcal{T}|X_i).$$

Proof. By the independence assumption,

$$\mathbb{P}_{\mathcal{T}}[(\mathcal{T}_1, \mathcal{T}_2, \ldots, \mathcal{T}_k)] = \prod_{i=1}^{k} \mathbb{P}_{\mathcal{T},X_i}[\mathcal{T}_i],$$

and, by (1), $\mathbb{P}_{\mathcal{T},X_i}[\mathcal{T}_i] = \alpha_i \exp[-\beta_i d(\mathcal{T}_i, \mathcal{T}|X_i)]$. Consequently, $\mathbb{P}_{\mathcal{T}}[(\mathcal{T}_1, \mathcal{T}_2, \ldots, \mathcal{T}_k)]$ is proportional to

$$\exp\left[-\sum_{i=1}^{k} \beta_i d(\mathcal{T}_i, \mathcal{T}|X_i)\right],$$

and this is maximised for any tree $\mathcal{T}$ that minimizes $\sum_{i=1}^{k} \beta_i d(\mathcal{T}_i, \mathcal{T}|X_i)$. This completes the proof.

Note that there may exist more than one ML supertree. This would be more likely if one has a small number of input trees and the distance metric is "coarse" (e.g., the symmetric difference metric) rather than "fine" (e.g., SPR: subtree-prune-and-regraft distance) and if the β_i values are all equal.

In the special case where d is the nearest-neighbor interchange (NNI) metric, and the β_i values are all equal, this ML supertree was described by Cotton and Wilkinson (2007). Moreover, in the consensus tree setting, and where d is the symmetric difference (Robinson-Foulds) metric, the consensus of the ML supertrees is the same as the usual majority-rule consensus tree. This follows from earlier results by Barthélemy and McMorris (1986; see also Cotton and Wilkinson, 2007).

The choice of metric d should be guided by the biological context and computational expediency—-for example, the SPR metric appropriately models horizontal gene transfer events, whereas the NNI metric may be appropriate for trees that lack local resolution; however,

the symmetric difference metric may also be useful as it is fast to compute.

Note also that the β_i values are not regarded as variables to be optimized in the ML procedure (to do so would lead to all trees being optimal solutions because $\beta_i = 0$ would be the optimal selection for all i). Rather, they allow other external factors (e.g., how well supported each input tree is by the data) to be reflected in the model. As a default option (in the absence of such knowledge), one might set all the β_i values equal.

The use of the exponential error model requires some explanation. In standard, character-based, or sequence-based phylogenetic analysis, the probability of obtaining the data is derived using a model of character or sequence substitution. This "process-based" approach to calculating the likelihood differs from the "error-based" approach we have adopted here. With an error-based approach, we model the probability distribution of outcomes, in this case the distribution of trees one may obtain from subsets of data. The value of this approach is that it is independent of process. The choice of an exponential distribution to describe the fall-off in probability of a tree as one moves further from the true tree is somewhat arbitrary but is due to simplicity and computational efficiency (Proposition 1). Regardless of the reasons why subtrees differ from the true supertree (i.e., regardless of the process), if the distribution of these differences can be modeled using an exponential distribution, our results hold.

Statistical Consistency of ML Supertrees under the Exponential Model

Is the ML procedure statistically consistent as the number (k) of trees in the profile grows? More precisely, under what conditions is the method guaranteed to converge on the underlying generating tree $\mathcal{T}_0$ as we add more trees to the analysis? The problem is slightly different from other settings (such as the consistency of ML for tree reconstruction from aligned sequence data) where one has a sequence of i.i.d. random variables. In the supertree setting, it is perhaps unrealistic to expect that the data sets are generated according to an identical process, since the sequence of subsets $X_1, X_2, \ldots$, of X is generally deliberately selected.

To formalize the statistical consistency question in this setting, let $X_1, X_2, \ldots$, be a sequence of subsets of finite set X. It is clear that the X_is must cover X in some reasonable way in order for the ML supertree method to be consistent—for example, if some taxon is not present in any X_i, or is present in only a small number of input trees, then we cannot expect the location of this taxon in any supertree to be strongly supported.

Thus, we will assume that the sequence of subsets of X satisfies the following *covering property*: For each subset Y of taxa from X of size m (where $m = 3$ for rooted trees or $m = 4$ for unrooted trees), the proportion of subsets X_i that contain Y does not decay to 0 as the sequence length (of subsets) increases. More formally, for each such subset Y of X we assume there is some $\epsilon > 0$ and some K sufficiently large for which:

$$\frac{1}{k}|\{i \leq k : Y \subseteq X_i\}| \geq \epsilon \text{ for all } k \geq K. \tag{2}$$

If a subset of taxa, Y, is only found in one or a few trees and is never seen again in trees that are subsequently added, this property will not hold.

We now establish statistical consistency of ML supertrees under a much more general class of models than the exponential model. Consider any model M for generating phylogenetic trees (without branch lengths) on subsets of X that has a generating phylogenetic X tree $\mathcal{T}$ as its sole underlying parameter. Such a model will typically derive from a more complex model containing other parameters (such as branch lengths, population sizes, and so forth), but we will assume that these have a prior distribution and that they have been integrated out, so our model has just one parameter—the tree topology.

Given a sequence $(X_1, X_2, \ldots, X_k \ldots)$ of subsets of X we say that M satisfies the property of *centrality* if, for some $\eta > 0$, and all $i \geq 1$, we have:

$$\mathbb{P}_{\mathcal{T}, X_i}[\mathcal{T}|X_i] - \mathbb{P}_{\mathcal{T}, X_i}[\mathcal{T}'] \geq \eta \tag{3}$$

for all trees $\mathcal{T}'$ on leaf set X_i that are different from $\mathcal{T}|X_i$. This condition says that, if $\mathcal{T}$ is the generating tree, then amongst all phylogenetic trees on leaf set X_i, the one that $\mathcal{T}$ induces on X_i is always (strictly) more probable than any other tree.

As a related condition, we say that M satisfies the property of *basal centrality* if, for some $\eta > 0$, and all $i \geq 1$, we have:

$$\mathbb{P}_{\mathcal{T}, X_i}[\mathcal{T}|Y] - \mathbb{P}_{\mathcal{T}, X_i}[\mathcal{T}'] \geq \eta \tag{4}$$

for all trees $\mathcal{T}'$ on leaf set Y that are different from $\mathcal{T}|Y$, and all $Y \subseteq X_i$, of size m (= 3 for rooted trees and = 4 for unrooted trees). This condition says that, if $\mathcal{T}$ is the generating tree, then amongst all the small (m element) subsets Y of X_i the tree that $\mathcal{T}$ induces on Y is (strictly) more probable that any other tree.

Notice that the exponential model (with the β_i values bounded away from 0) satisfies centrality (take $\eta = \min_{i \geq 1}[\alpha_i(1 - e^{-\beta_i})]$). We will see later that basal centrality is a useful concept for another setting (lineage sorting). Note also that basal centrality (for a sequence X_i) is not a special case of centrality, and, conversely, centrality is not a special case of basal centrality, though of course the two notions coincide in the special case where $|X_i| = m$ for all i.

We now show that the selection of an ML supertree is statistically consistent for any model M that satisfies either centrality or basal centrality when applied to a sequence of subsets of X that obeys the covering property. Its proof (along with the proofs of all following propositions and theorems) is given in the Appendix.

Theorem 2. *Given a sequence $X_1, X_2, \ldots,$ which satisfies the covering property (2), consider a profile $\mathcal{P}_k = (\mathcal{T}_1, \ldots, \mathcal{T}_k)$,*

where $\mathcal{T}_i$ is generated independently, from a tree $\mathcal{T}_0$ with taxon set X, according to a model M that satisfies either the centrality or basal centrality property for this sequence. Then the probability that $\mathcal{P}_k$ has a unique ML supertree and that this tree is $\mathcal{T}_0$ tends to 1 as $k \to \infty$.

Remarks

- We can easily modify the ML process if some of the input trees are not fully resolved (due to "soft" polytomies). For a general phylogenetic tree t_i (possibly with polytomies) on taxon set $X_i \subseteq X$, and a generating fully resolved phylogenetic tree $\mathcal{T}$ on taxon set X, let $\phi(t_i|\mathcal{T})$ be the probability of the event that the tree $\mathcal{T}_i$ that $\mathcal{T}$ generates under the exponential model is a refinement of t_i. More precisely,

$$\phi(t_i|\mathcal{T}) = \sum_{\mathcal{T}_i \geq t_i} \alpha_i \exp[-\beta_i d(\mathcal{T}_i, \mathcal{T}|X_i)]$$

  where $\mathcal{T}_i \geq t_i$ indicates that the (fully resolved) tree $\mathcal{T}_i$ contains all the splits present in t_i and has the same leaf set (X_i). Notice that $\phi(t_i|\mathcal{T})$ is not a probability distribution on phylogenetic trees with the leaf set X_i (its sum is > 1). Nevertheless, given a profile $\mathcal{P} = (t_1, \ldots, t_k)$ of phylogenetic trees (some or all of which may have polytomies), one can perform ML to select the tree $\mathcal{T}$ that maximizes the joint probability $\prod_{i=1}^{k} \phi(t_i|\mathcal{T})$ of the events $\mathcal{T}_i \geq t_i$ for $i = 1, \ldots, k$.
- We point out an alternative way of viewing this ML procedure applied to a profile $\mathcal{P} = (\mathcal{T}_1, \ldots, \mathcal{T}_k)$ when d is one of two well-known metrics on trees (SPR and TBR). Suppose that we were to extend each tree $\mathcal{T}_i$ in $\mathcal{P}$ to a tree $\mathcal{T}_i'$ on the full set of taxa (X). We could regard the placement of those taxa that are missing in $\mathcal{T}_i$ (namely the taxa in $X - X_i$) to form a tree $\mathcal{T}_i'$ on the full leaf set X to be "nuisance parameters" in a maximum likelihood framework (under the exponential model), and thereby seek to find the tree $\mathcal{T}$ and extensions $(\mathcal{T}_1', \ldots, \mathcal{T}_k')$ to maximize the joint probability:

$$\mathbb{P}_{\mathcal{T}}[(\mathcal{T}_1', \mathcal{T}_2', \ldots, \mathcal{T}_k')] \text{ subject to } \mathcal{T}_i = \mathcal{T}_i'|X_i \text{ for all } i.$$

  We call such a tree $\mathcal{T}$ an *extended ML tree* for the profile $\mathcal{P}$; it turns out to be just the same as the ML trees we have defined above, as the next result shows.

  Proposition 3. *For $d = SPR$ or $d = TBR$, and any profile $\mathcal{P}$ of fully resolved, unrooted phylogenetic trees, the extended ML tree(s) for $\mathcal{P}$ coincides precisely with the ML tree(s) for $\mathcal{P}$.*

Relation to MRP and Its Statistical Inconsistency

In the MRP (matrix representation with parsimony) supertree method, the input trees are coded as characters—one for each interior edge of each tree—with character states 0, 1, ? depending, respectively, on whether the taxon in question is on one side, or the other, of the edge, or is absent from the tree. A most parsimonious tree (or trees) is then reconstructed from this data matrix. As shown recently by Bruen and Bryant (2007), there is a close analogy between MRP and consensus tree methods, which seek a median tree computed using the SPR (subtree prune and regraft) or TBR (tree bisection and reconnection) metric d (recall that a median tree for a profile $\mathcal{P} = (\mathcal{T}_1, \ldots \mathcal{T}_k)$ of trees that all have same leaf set X, is a tree $\mathcal{T}$ that minimizes the sum $\sum_{i=1}^{k} d(\mathcal{T}, \mathcal{T}_i)$; cf. Proposition 1). However, the result from Bruen and Bryant (2007) does not guarantee that MRP produces an ML supertree even when $\beta_i = 1$ for all i, as their approach constructs a median on a space of trees that are defined by splits rather than particular trees.

We turn now to the question of the statistical consistency of MRP under the exponential model (1). It can be shown that MRP will be statistically consistent under the covering property (2) in some special cases. Two such cases that can be formally established (details omitted) are (i) when all the subsets X_i are of size m ($= 3$ for rooted trees and $= 4$ for unrooted trees); or (ii) when β_i is sufficiently large (in relation to $|X|$). However, in general, we have the following result.

Theorem 4. *A $\beta > 0$ exists for which MRP is statistically inconsistent even in the special (consensus) case where, for all i, X_i is the same set of six taxa and $\beta_i = \beta$. More precisely, for this value of β and with unrooted fully-resolved phylogenetic trees on these (equal) taxon sets, the probability that $\mathcal{T}_0$ is an MRP tree (for a profile of trees generated under (1)) converges to 0 as k tends to infinity.*

Remarks

- The formal proof of Theorem 4 is given in the Appendix. However, the intuition behind the proof is that if β is sufficiently small (but still positive), then trees generated by the exponential model are close to a uniform random distribution on trees. For such input there is a slight bias of MRP towards unbalanced trees. This shape bias property of MRP has been explored (in related settings) by Wilkinson et al. (2005).
- The exact condition (on the β_is) at which inconsistency of MRP occurs is not clear, though some general comments can be made. For large values of β_i, the decay of the exponential function is rapid, and the probability of producing an incorrect subtree some distance away from the true tree is small. In these cases, MRP is likely to work well. However, when β_i is small, the probability of producing an incorrect subtree is relatively high, even when this subtree is some distance away from the true tree. An interesting theoretical question is whether a value $s \in (0, 1)$ exists for which MRP is statistically consistent (for arbitrarily large taxon sets) under the conditions of Theorem 4, whenever $\beta_i = \beta \geq s$.

Statistical Consistency of ML Species Supertrees from Multiple Gene Trees

A current problem in phylogenetics is how best to infer species trees from gene trees (Degnan and Rosenberg,

2006; Gadagkar et al., 2005; Liu and Pearl, 2007). Even in the consensus setting (i.e., when the set of taxa for each gene tree is the complete set of taxa under study), Degnan and Rosenberg (2006) have demonstrated how incomplete lineage sorting on gene trees can mean that the most likely topology for a gene tree can differ from the underlying species tree (for certain rooted phylogenetic trees on four taxa and for all rooted phylogenetic trees on five or more taxa). This surprising result implies that simplistic majority-rule approaches to finding a consensus species tree can be problematic.

The phenomenon described by Degnan and Rosenberg (2006) is based on the coalescent model for studying lineage sorting in evolving populations. The surprising behavior arises only when the effective population sizes and the branch lengths of the species tree are in appropriate ranges. Moreover, for rooted three-taxon trees, the most probable gene tree topology always agrees with the species tree topology. Nevertheless, the fact that larger gene trees can favor an incorrect species tree might easily complicate some standard statistical approaches.

In this section, we show how, despite the phenomena described above (from Degnan and Rosenberg, 2006), and even in the more general supertree setting (where some gene trees may have some missing taxa), a maximum likelihood approach to supertree construction of a species tree from gene trees is statistically consistent.

Consider first a somewhat idealist situation where we have a sequence of rooted gene trees, each correctly inferred (but possibly differing from the species tree due to lineage sorting) for a sequence of (arbitrary size) subsets of X that satisfy the covering property. The statistical consistency of ML under this scenario follows immediately from Theorem 2 (using basal centrality, not centrality) because lineage sorting (regarded as an error model M for deriving gene trees from species trees under the coalescent model) satisfies basal centrality. This is because, under the coalescent model, $\mathbb{P}^c_{\mathcal{T},Y}[\mathcal{T}|Y] = 1 - \frac{2}{3}e^{-\lambda}$, whereas $\mathbb{P}^c_{\mathcal{T},Y}[\mathcal{T}''] = \frac{1}{3}e^{-\lambda}$ for the two other choices of $\mathcal{T}'' \neq \mathcal{T}|Y$, where λ is (up to a factor of 2) the ratio of time (measured in generations) to the effective population size (see, e.g., Rosenberg, 2002; Tajima, 1983). Consequently,

$$\mathbb{P}_{\mathcal{T},Y}[\mathcal{T}|Y] - \mathbb{P}_{\mathcal{T},Y}[\mathcal{T}'] \geq 1 - e^{-\lambda} > 0,$$

for all subsets Y of X_i of size 3. The limitation of this consistency result is that it invokes the unrealistic assumption that the gene trees have all been correctly inferred. We thus consider a more realistic scenario; however, in order to prove a consistency result, we must restrict the input trees to have just 3 leaves each. We thus consider the following situation:

Triplet-based supertree with dual error: a sequence of rooted gene trees inferred, possibly with error, on subset of X of size 3 that satisfy the covering property, provided the error in the reconstructed gene tree (differing from the true gene tree) is described by the exponential model.

The statistical consistency of ML supertrees under this scenario (Triplet-based supertree with dual error) is due to the following result (where M_1 is the lineage sorting model that derives a gene tree from a species tree, whereas M_2 models the error in reconstructing the gene tree correctly).

Proposition 5. *Given a sequence $X_1, X_2, \ldots,$ of subsets of X of size 3 that satisfies the covering property (2), consider a profile $\mathcal{P}_k = (\mathcal{T}_1, \ldots, \mathcal{T}_k)$, where tree $\mathcal{T}_i$ is generated independently, from a tree $\mathcal{T}_0$ with taxon set X, by first generating a tree on taxon set X_i under some central model, M_1, and then, independently, using this tree to generate $\mathcal{T}_i$ according to an exponential model, M_2. Then the probability that $\mathcal{P}_k$ has a unique ML supertree and that this tree is $\mathcal{T}_0$ tends to 1 as $k \to \infty$.*

Discussion

To develop a likelihood-based supertree reconstruction method, it is necessary to define a model that delivers the probability of obtaining a series of subtree topologies, given a hypothesized supertree. We have chosen a very simple yet intuitive probability function whereby the probability of observing a wrong subtree (i.e., one where the topology differs from that of a pruned supertree) decreases exponentially as its topology becomes increasingly distant from that of the hypothesised supertree. Consequently, the ML supertree can be estimated even when the constituent subtrees have conflicting topological signals.

Our approach is model based, but one may reasonably ask whether the model described here is a biologically realistic one. We suggest that it is. For one thing, we expect, for a variety of reasons, to see conflicts between the topologies of subtrees and the reconstructed supertree. With gene sequences obtained from different species, for instance, incomplete lineage sorting and ancestral heterozygosity frequently lead to differences between gene trees and species trees. Convergent and parallel evolution can confound phylogenetic reconstruction, as can long-branch attraction. We have chosen to use the exponential distribution to describe this steady decrease in probabilities as the distances between subtrees and supertrees increase. The value of using the exponential distribution lies in the ease with which it can be manipulated when we compute log-likelihoods. Additionally, we have noted that our model is an error-based model; i.e., it describes the distribution of subtree-to-supertree distances without regard for the underlying processes that cause topological conflicts. However, we suggest that one fruitful research project may be to explore other possible probability distributions, other tree-to-tree distance metrics, as well as process-based models of topological conflict. In this respect, coalescent-based models of incomplete lineage sorting (Carstens and Knowles, 2007) may hold some promise. As with the original phylogenetic likelihood methods designed for character and sequence data, we hope to see new models emerge as other researchers explore the use of ML supertrees. Even if one is reluctant to use the exponential distribution, the statistical consistency of ML supertrees is nonetheless guaranteed, provided the probability distribution of subtree-to-supertree distances assigns the highest probability to a distance of zero.

The likelihood framework provides an additional benefit: a rich body of statistical and phylogenetic methods already use likelihood. Moreover, statistical consistency holds for maximum likelihood supertrees under weak conditions, in contrast to MRP, which can be inconsistent in some cases. We also show that the ML supertree approach developed here provides a statistically consistent strategy for combining gene trees even when there is the possibility that these trees may be different from the true species tree. An obvious application of ML supertrees will be their use in statistical tests of topological hypotheses, and we already know how to do this with standard ML phylogenies (Goldman et al., 2000).

We also recognize that our particular likelihood implementation is closely related to the Majority-Rule(−) Supertree construction proposed by Cotton and Wilkinson (2007). More precisely, when the tree metric is the symmetric difference (Robinson-Foulds) metric, then the Marjority-Rule(−) Supertree is, in effect, the strict consensus of our ML supertrees. However, the approach in Cotton and Wilkinson (2007) is quite different: they show how to extend majority rule from the consensus to the supertree setting. Nonetheless, they converge on the same optimality criterion that we use; i.e., a supertree that minimizes the sum of distances to a set of trees. One should not be surprised that the same optimality criterion can emerge from different conceptual bases. With standard phylogenetic reconstruction, choosing the tree that minimizes the number of evolutionary changes can be justified philosophically (with the principle of maximum parsimony) as a consensus method (Bruen and Bryant, 2007) or by using an explicitly statistical approach such as likelihood (Steel and Penny, 2000).

We have not discussed algorithms to search for ML supertrees. As with classical phylogenetic likelihood methods, maximum likelihood supertrees will be obtained using a variety of approaches, including heuristic methods, genetic algorithms, simulated annealing, and so on. We also direct readers to the discussion in Cotton and Wilkinson (2007), because the criterion we use is similar to theirs.

ACKNOWLEDGEMENTS

We thank the Allan Wilson Centre for Molecular Ecology and Evolution for supporting this work. Allen Rodrigo began this project while he was working with Olivier Gascuel at the Laboratoire d'Informatique, de Robotique et de Microelectronique de Montpellier.

REFERENCES

Barthélemy, J. P., and F. R. McMorris. 1986. The median procedure for n-trees. J. Classif. 3:329–334.

Baum, B. R., and M. A. Ragan. 1992. Combining trees as a way of combining data sets for phylogenetic inference, and the desirability of combining gene trees. Taxon 41:3–10.

Bininda-Emonds, O. R. P., M. Cardillo, K. E. Jones, R. D. E. MacPhee, R. M. D. Beck, R. Grenyer, S. A. Price, R. A. Vos, J. L. Gittleman, and A. Purvis. 2007. The delayed rise of present-day mammals. Nature 446:507–512.

Bininda-Emonds, O., and M. J. Sanderson. 2001. Assessment of the accuracy of matrix representation with parsimony analysis supertree construction. Syst. Biol. 50:565–579.

Bruen, T., and D. Bryant. 2007. Parsimony as consensus. Syst. Biol. In press.

Carstens, B., and L. Knowles. 2007. Estimating phylogeny from gene tree probabilities in melanoplus grasshoppers despite incomplete lineage sorting. Syst. Biol. 56:400–411.

Cotton, J. A., and M. Wilkinson. 2007. Majority-rule supertrees. Syst. Biol. 56:445–452.

Degnan, J. H., and N. A. Rosenberg. 2006. Discordance of species trees with their most likely gene trees. PLoS Genet. 2(5):e68.

Gadagkar, S. R., M. S. Rosenberg, and S. Kumar. 2005. Inferring species phylogenies from multiple genes: Concatenated sequence tree versus consensus gene tree. J. Exp. Zool. 304B:64–74.

Goldman, N., J. P. Anderson, and A. G. Rodrigo. 2000. Likelihood-based tests of topologies in phylogenetics. Syst. Biol. 49:652–670.

Holmes, S. 2003. Statistics for phylogenetic trees. Theor. Pop. Biol. 63:17–32.

Liu, L., and D. K. Pearl. 2007. Species trees from gene trees: Reconstructing Bayesian posterior distributions of a species phylogeny using estimated gene tree distributions. Syst. Biol. 56:504–514.

McMorris, F. R. 1990. The median procedure for n-trees as a maximum likelihood method. J. Classif. 7:77–80.

Rodrigo, A. G. 1996. On combining cladograms. Taxon 45:267–274.

Ronquist, F., J. J. Huelsenbeck, and T. Britton. 2004. Bayesian supertrees. Chapter 9 *in* Phylogenetic supertrees (O. R. P. Bininda-Emonds, ed.). Kluwer Academic Publishers, Dordrecht, The Netherlands.

Rosenberg, N. A. 2002. The probability of topological concordance of gene trees and species trees. Theor. Pop. Biol. 61:225–247.

Ross, H. A., and A. G. Rodrigo. 2004. An assessment of matrix representation with compatibility in supertree reconstruction. Chapter 2 *in* Phylogenetic supertrees (O. R. P. Bininda-Emonds, ed.). Kluwer Academic Publishers, Dordrecht, The Netherlands.

Semple, C., and M. Steel. 2003. Phylogenetics. Oxford University Press.

Steel, M., and D. Penny. 2000. Parsimony, likelihood and the role of models in molecular phylogenetics. Mol. Biol. Evol. 17:839–850.

Steel, M., and L. A. Szekely. 2002. Inverting random functions (ii): Explicit bounds for discrete maximum likelihood estimation, with applications. SIAM J. Discr. Math. 15:562–575.

Steel, M. A., M. D. Hendy, and D. Penny. 1992. Significance of the length of the shortest tree. J. Classif. 9:71–90.

Tajima, F. 1983. Evolutionary relationships of DNA sequences in finite populations. Genetics 105:437–460.

Wilkinson, M., J. A. Cotton, C. Creevey, O. Eulenstein, S. R. Harris, F.-J. Lapointe, C. Levasseur, J. O. Mcinerney, D. Pisani, and J. L. Thorley. 2005. The shape of supertrees to come: Tree shape related properties of fourteen supertree methods. Syst. Biol 54:419–431.

Wilkinson, M., J. L. Thorley, D. T. L. Littlewood, and R. A. Bray. 2001. Towards a phylogenetic supertree for Platyhelminthes? Pages 292-301 *in* Interrelationships of the Platyhelminthes (D. T L. Littlewood and R. A. Bray, eds.). Taylor and Francis, London, UK.

First submitted 15 August 2007; reviews returned 19 November 2007; final acceptance 15 January 2008
Associate Editor: Olivier Gascuel

APPENDIX

PROOFS OF MAIN RESULTS

In this section we collect together the proofs of the main results. We start by stating a general sufficient condition for the consistency of ML in non-i.i.d. settings.

Consistency of ML for General (non-i.i.d.) Sequences

Here we describe a convenient way to establish the statistical consistency of maximum likelihood when we have a sequence of observations that may not be independent or identically distributed. We frame this discussion generally, as the result may be useful for other problems. In particular, in this result, we do not need to assume the sequence samples are independent (though in our applications, they are), nor identically distributed (in our applications, they are not). Suppose we have a sequence of random variables $Y_1, Y_2, \ldots$, that takes values in

some finite set W and are generated by some process that depends on an underlying discrete parameter a that can take values in some finite set A. In our setting, the Y_is are trees constructed from different data sets (e.g., gene trees), whereas a is the generating species tree topology. We assume that the model specifies the probability distribution of $(Y_1, \ldots, Y_k)$ given (just) a—for example, in our tree setting this would mean specifying prior distributions on the branch lengths and other parameters of interest (e.g., ancestral population sizes) and integrating with respect to these priors.

Given an actual sequence $(y_1, \ldots, y_k)$ of observations, the *maximum likelihood (ML) estimate* of the discrete parameter is the value a that maximizes the joint probability

$$\mathbb{P}_a[Y_1 = y_1, \ldots, Y_k = y_k]$$

(i.e., the probability that the process with parameter a generates $(y_1, \ldots, y_k)$). Now suppose that the sequence $Y_1, \ldots Y_k, \ldots$, is generated by a_0. We would like the probability that the ML estimate is equal to a_0 to converge to 1 as k increases. If this holds for all choice of $a_0 \in A$, then ML is *statistically consistent*. The following result provides a convenient way to establish this; indeed, it characterises the statistical consistency of ML.

Proposition 6. *In the general setup described above, ML is statistically consistent if and only if the following condition holds: for any two distinct elements $a, b \in A$, we can construct a sequence of events $E_1, E_2, \ldots$, where E_k is dependent on $(Y_1, \ldots Y_k)$, for which, as $k \to \infty$:*

(i) the probability of E_k under the model with parameter a converges to 1.
(ii) the probability of E_k under the model with parameter b converges to 0.

Proof. The "only if" direction is easy: Suppose ML is statistically consistent and $a, b \in A$ are distinct. Let E_k be the event that a is the unique maximum likelihood estimate obtained from $(Y_1, \ldots, Y_k)$. Then E_k satisfies conditions (i) and (ii).

For the converse direction, recall that the *variation distance* between two probability distributions p, q on any finite set W is

$$\max_{E \subset W} |\mathbb{P}_p(E) - \mathbb{P}_q(E)|$$

where $\mathbb{P}_p(E) = \sum_{w \in E} p(w)$ is the probability of event E under distribution p (similarly for $\mathbb{P}_q(E)$). This variation distance can also be written as $\frac{1}{2}\|p - q\|_1$, where $\|p - q\|_1 = \sum_{w \in W} |p(w) - q(w)|$ is the l_1 distance between p and q. Thus, if we let $d^{(k)}(a, b)$ denote the l_1 distance between the probability distribution on $(Y_1, \ldots, Y_k)$ induced by a and by b, then conditions (i) and (ii) imply that

$$\lim_{k \to \infty} \frac{1}{2} d^{(k)}(a, b) = 1. \tag{5}$$

Now, by the first part (equation 3.1) of Theorem 3.2 of Steel and Szekely (2002), the probability that the ML estimate is the value of A that generates the sequence $(Y_1, \ldots, Y_k)$ is at least $1 - \sum_{b \neq a}[1 - \frac{1}{2}d^{(k)}(a, b)]$ and so, by (5), this probability converges to 1 as $k \to \infty$.

Proof of Theorem 2. To establish the theorem, using Proposition 6 (stated above) it is enough to specify for each choice of distinct resolved phylogenetic X-trees $\mathcal{T}_0$ and $\mathcal{T}$, a sequence of events E_k (dependent on $\mathcal{P}_k$) for which, as k grows, E_k has a probability that tends to 1 under the distribution obtained from $\mathcal{T}_0$ and tends to 0 under the distribution obtained from $\mathcal{T}$. Since $\mathcal{T}$ differs from $\mathcal{T}_0$ a subset Y exists of size m (= 3 for rooted trees and = 4 for unrooted) for which $\mathcal{T}|Y \neq \mathcal{T}_0|Y$. Notice that the covering property (2) implies that

$$\frac{1}{k}|\{i \leq k : \mathcal{T}|X_i \neq \mathcal{T}_0|X_i\}| \geq \epsilon \text{ for all } k \geq K. \tag{6}$$

Let E_k be the event that among all those $i \in \{1, \ldots, k\}$ for which $\mathcal{T}|X_i \neq \mathcal{T}_0|X_i$, we have $\mathcal{T}_i = \mathcal{T}_0|X_i$ more often than $\mathcal{T}_i = \mathcal{T}|X_i$. Now, for a profile generated by $\mathcal{T}_0$ according to the model satisfying the centrality property, we have, for each i for which $\mathcal{T}|X_i \neq \mathcal{T}_0|X_i$,

$$\mathbb{P}_{\mathcal{T}_0, X_i}[\mathcal{T}_i = (\mathcal{T}_0|X_i)] \geq \mathbb{P}_{\mathcal{T}_0, X_i}[\mathcal{T}_i = (\mathcal{T}|X_i)] + \eta. \tag{7}$$

Similarly, for a profile generated by $\mathcal{T}$ according to a model satisfying the centrality property, and for each i for which $\mathcal{T}|X_i \neq \mathcal{T}_0|X_i$, we have

$$\mathbb{P}_{\mathcal{T}, X_i}[\mathcal{T}_i = (\mathcal{T}|X_i)] \geq \mathbb{P}_{\mathcal{T}, X_i}[\mathcal{T}_i = (\mathcal{T}_0|X_i)] + \eta. \tag{8}$$

By condition (6), there is a positive limiting proportion ($\epsilon > 0$) of i for which $\mathcal{T}|X_i \neq \mathcal{T}_0|X_i$. By independence (and the law of large numbers) it follows from inequality (7) that event E_k has a probability that tends to 1 as $k \to \infty$ for a profile generated by $\mathcal{T}_0$. Similarly, by (8), event E_k has a probability tending to 1 as $k \to \infty$ for a profile generated by $\mathcal{T}$. Statistical consistency of ML now follows by Proposition 6.

For a model M satisfying basal centrality the same argument applies if we select Y as above and modify event E_k to be the event that among all those $i \in \{1, \ldots, k\}$ for which $Y \subseteq X_i$ we have $\mathcal{T}_i|Y = \mathcal{T}_0|Y$ more often than $\mathcal{T}_i = \mathcal{T}|Y$.

Proof of Theorem 4. For two unrooted fully-resolved phylogenetic X-trees $\mathcal{T}, \mathcal{T}'$, let $L(\mathcal{T}, \mathcal{T}')$ denote the total parsimony score on $\mathcal{T}'$ of the set of splits of $\mathcal{T}$. That is,

$$L(\mathcal{T}, \mathcal{T}') = \sum_{\sigma \in \Sigma(\mathcal{T})} l(\sigma, \mathcal{T}'), \tag{9}$$

where $\Sigma(\mathcal{T})$ is the set of splits of $\mathcal{T}$ and $l(\sigma, \mathcal{T}')$ is the parsimony score of the split σ on $\mathcal{T}'$ (treating σ as a binary character; Semple and Steel, 2003). For any fully resolved phylogenetic X tree $\mathcal{T}'$, let $e(\mathcal{T}'; \mathcal{T}_0)$ be the expected total parsimony score on $\mathcal{T}'$ of the set of splits of a tree $\mathcal{T}$ randomly generated by $\mathcal{T}_0$ according to the exponential model (1). Then,

$$e(\mathcal{T}'; \mathcal{T}_0) = \sum_{\mathcal{T}} \alpha \exp[-\beta d(\mathcal{T}, \mathcal{T}_0)] \cdot L(\mathcal{T}, \mathcal{T}'). \tag{10}$$

To establish Theorem 4, it is enough to show, for some $\beta > 0$ and for two unrooted fully resolved trees $\mathcal{T}_0, \mathcal{T}_1$ on $X = \{1, \ldots, 6\}$, that $e(\mathcal{T}_0; \mathcal{T}_0) - e(\mathcal{T}_1; \mathcal{T}_0) > 0$, because if $\mathcal{T}_0$ is the generating tree, then $\mathcal{T}_1$ will be favored over $\mathcal{T}_0$ by MRP. We first show that this can occur when $\beta = 0$. In that case, $\alpha \exp[-\beta d(\mathcal{T}, \mathcal{T}_0)] = 1/105$ for all $\mathcal{T}$ (there are 105 unrooted fully-resolved phylogenetic trees on X) and so, by (10), we have

$$e(\mathcal{T}_0; \mathcal{T}_0) - e(\mathcal{T}_1; \mathcal{T}_0) = \frac{1}{105} \sum_{\mathcal{T}} [L(\mathcal{T}, \mathcal{T}_0) - L(\mathcal{T}, \mathcal{T}_1)].$$

Applying (9) and interchanging the order of summation gives:

$$e(\mathcal{T}_0; \mathcal{T}_0) - e(\mathcal{T}_1; \mathcal{T}_0) = \frac{1}{105} \sum_{\sigma} n(\sigma) \cdot [l(\sigma, \mathcal{T}_0) - l(\sigma, \mathcal{T}_1)], \tag{11}$$

where $n(\sigma)$ is the number of unrooted fully resolved phylogenetic X trees containing split σ and the summation is over all the splits of $X = \{1, \ldots, 6\}$. Moreover, if any difference $l(\sigma, \mathcal{T}_0) - l(\sigma, \mathcal{T}_1)$ is non-zero in (11), then σ is necessarily a split that partitions the taxa into sets of sizes either 2, 4 or 3, 3, and for such a split σ we have $n(\sigma) = 15$ and 9, respectively.

Now suppose $\mathcal{T}_0$ has a symmetric shape (i.e., an unrooted fully resolved tree of six leaves with three cherries) and $\mathcal{T}_1$ has a pectinate shape (i.e., an unrooted fully resolved tree of six leaves with two cherries). Then, by using earlier results (Steel et al., 1992, table 3) concerning the number of splits that partition the taxa into sets of size 2, 4 and 3, 3 and have parsimony score 1, 2, 3 in these two trees, it can be shown from (11) that

$$e(\mathcal{T}_0; \mathcal{T}_0) - e(\mathcal{T}_1; \mathcal{T}_0) > 0$$

So far, we have assumed that $\beta = 0$; however, $e(\mathcal{T}_0; \mathcal{T}_0) - e(\mathcal{T}_1; \mathcal{T}_0)$ is a continuous function of β, so a strictly positive value of β exists for which

$$e(\mathcal{T}_0; \mathcal{T}_0) - e(\mathcal{T}_1; \mathcal{T}_0) > 0.$$

This completes the proof.

Proof of Proposition 3. For $d = SPR$ or $d = TBR$, and for any resolved unrooted phylogenetic trees $\mathcal{T}$ on taxon set X, and $\mathcal{T}_Y$ on taxon set $Y \subseteq X$ we claim that:

$$\min\{d(\mathcal{T}', \mathcal{T}) : \mathcal{T}'|Y = \mathcal{T}_Y\} = d(\mathcal{T}_Y, \mathcal{T}|Y). \tag{12}$$

where $\mathcal{T}'$ ranges over the set of all unrooted resolved phylogenetic tree on taxon set X that induce $\mathcal{T}_Y$ when restricted to Y. To establish this claim, note firstly that, for any $\mathcal{T}'$ with taxon set X, and $Y \subseteq X$, we have $d(\mathcal{T}'|Y, \mathcal{T}|Y) \leq d(\mathcal{T}', \mathcal{T})$ and so the $\geq$ inequality holds in (12). To establish equality induction on k (starting with the base case $k = 1$) shows that if $\mathcal{T}_Y$ and $\mathcal{T}|Y$ are k SPR (or TBR) moves apart, then there exists an unrooted resolved phylogenetic $\mathcal{T}'$ on taxon set X that induces $\mathcal{T}_Y$ when restricted to Y, and which is at most k SPR (or TBR, respectively) moves apart from $\mathcal{T}$. Having established (12), Proposition 3 now follows by Proposition 1.

Note that Equation (12) does not necessarily hold for other tree metrics such as the NNI (nearest-neighbor interchange) or the partition (Robinson-Foulds) metric.

Proof of Proposition 5. By Proposition 2 it suffices to show that the composite model satisfies centrality. Given a rooted phylogenetic tree $\mathcal{T}$ on taxon set X, and a subset Y of X of size 3, label the three rooted binary trees on Y, $\mathcal{T}_1$, $\mathcal{T}_2$, $\mathcal{T}_3$, so that $\mathcal{T}_1 = \mathcal{T}|Y$. We need to show that, for some $\eta' > 0$, $\mathbb{P}_{\mathcal{T},Y}(\mathcal{T}_1) \geq \mathbb{P}_{\mathcal{T},Y}(\mathcal{T}_j) + \eta'$ for $j = 2, 3$. By the independence assumption,

$$\mathbb{P}_{\mathcal{T},Y}(\mathcal{T}_k) = \sum_{j=1}^{3} \mathbb{P}^{(1)}_{\mathcal{T},Y}(\mathcal{T}_j)\mathbb{P}^{(2)}_{\mathcal{T}_j,Y}(\mathcal{T}_k),$$

where $\mathbb{P}^{(1)}_{\mathcal{T},Y}(\mathcal{T}_j)$ is the probability of generating $\mathcal{T}_j$ under model M_1 from generating tree $\mathcal{T}$, and $\mathbb{P}^{(2)}_{\mathcal{T}_j,Y}(\mathcal{T}_k)$ is the probability of generating $\mathcal{T}_k$ under model M_2 with generating tree $\mathcal{T}_j$. Now, $\mathbb{P}^{(2)}_{\mathcal{T}_j,Y}(\mathcal{T}_k)$ takes the value α for $j = k$, and the value $\alpha e^{-\beta}$ for $j \neq k$. Thus, $\mathbb{P}_{\mathcal{T},Y}(\mathcal{T}_1) = \alpha\mathbb{P}^{(1)}_{\mathcal{T},Y}(\mathcal{T}_1) + \alpha e^{-\beta}[\mathbb{P}^{(1)}_{\mathcal{T},Y}(\mathcal{T}_2) + \mathbb{P}^{(1)}_{\mathcal{T},Y}(\mathcal{T}_3)]$, whereas $\mathbb{P}_{\mathcal{T},Y}(\mathcal{T}_2) = \alpha\mathbb{P}^{(1)}_{\mathcal{T},Y}(\mathcal{T}_2) + \alpha e^{-\beta}[\mathbb{P}^{(1)}_{\mathcal{T},Y}(\mathcal{T}_1) + \mathbb{P}^{(1)}_{\mathcal{T},Y}(\mathcal{T}_3)]$. Consequently,

$$\mathbb{P}_{\mathcal{T},Y}(\mathcal{T}_1) - \mathbb{P}_{\mathcal{T},Y}(\mathcal{T}_2) = \alpha(1 - e^{-\beta})\left[\mathbb{P}^{(1)}_{\mathcal{T},Y}(\mathcal{T}_1) - \mathbb{P}^{(1)}_{\mathcal{T},Y}(\mathcal{T}_2)\right].$$

A similar expression holds for other difference $\mathbb{P}_{\mathcal{T},Y}(\mathcal{T}_1) - \mathbb{P}_{\mathcal{T},Y}(\mathcal{T}_3)$ and because M_1 is central, we can take η' to be the minimal value of $\alpha(1 - e^{-\beta})[\mathbb{P}_{\mathcal{T},Y}(\mathcal{T}_1) - \mathbb{P}_{\mathcal{T},Y}(\mathcal{T}_k)]$ over $k = 2, 3$.