Published in final edited form as:
Stat Med. 2011 September 30; 30(22): 2708–2720. doi:10.1002/sim.4287.

Bayesian methods for fitting mixture models that characterize branching tree processes: an application to development of resistant TB strains

Alane Izu[a,*], **Ted Cohen**[b], **Carole Mitnick**[c], **Megan Murray**[d], and **Victor DeGruttola**[e]
[a,e]Harvard School of Public Health, 655 Huntington Avenue, Boston, MA 02115
[b,d]Harvard School of Public Health, 641 Huntington Avenue, Boston, MA 02115
[c]Harvard Medical School, 641 Huntington Avenue, Boston, MA 02115

Abstract

For pathogens that must be treated with combinations of antibiotics and acquire resistance through genetic mutation, knowledge of the order in which drug-resistance mutations occur may be important for determining treatment policies. Diagnostic specimens collected from patients are often available; this makes it possible to determine the presence of individual drug-resistance conferring mutations and combinations of these mutations. In most cases, these specimens are only available from a patient at a single point in time; it is very rare to have access to multiple specimens from a single patient collected over time as resistance accumulates to multiple drugs. Statistical methods that use branching trees have been successfully applied to such cross-sectional data to make inference on the ordering of events that occurred prior to sampling. Here we propose a Bayesian approach to fitting branching tree models that has several advantages, including the ability to accommodate prior information regarding measurement error or cross-resistance and the natural way it permits the characterization of uncertainty. Our methods are applied to a data set for drug resistant tuberculosis in Peru; the goal of analysis is to determine the order with which patients develop resistance to the drugs commonly used for treating TB in this setting.

Keywords

Bayesian Networks; Branching; TB drug resistance; Tree Inference

1. Introduction

Resistance to antimicrobial drugs complicates the development of strategies to treatment and control infection from many different types of microbes. Because many diseases such as tuberculosis (TB), human immunodeficiency virus (HIV), hepatitis B virus (HBV), hepatitis C virus (HCV) require regimens that include multiple drugs, determining the best strategy for combining and sequencing drugs is a considerable challenge. Developing a better understanding of the order of events that lead to the loss of efficacy of individual drugs or of combination regimens can potentially help with this challenge. Such knowledge might, for example, help to preserve future drug options for patients who have failed a regimen if the probabilities of certain sequences of events can be well estimated. Here the events of importance could either be genetic events, such as individual point mutations or

[*]Correspondence to: Harvard School of Public Health, 655 Huntington Avenue, Boston, MA 02115. aizu@hsph.harvard.edu.

combinations of mutations, or changes in phenotype, such as development of in-vitro resistance to particular drugs.

The best information for reconstructing the order of events of interest arises from longitudinal studies; however, in some settings, much more cross-sectional than longitudinal data are available. This situation arises from the high cost of longitudinal studies together with the constrained resources available in some settings with the highest prevalence and incidence of disease. However, considerable work has been done to make inference about the order of events from cross-sectional data under some assumptions about the underlying process governing the onset of these events. For example, single and multiple branching trees have been used to model the sequence of copy number aberrations from comparative genome hybridization data for renal cancer and to model the order of acquisition of resistance mutations for genetic HIV data [1, 2, 3]. A branching tree is a special Bayesian network that models the joint distribution of events and imposes constraints on the dependencies between events and the order in which they can occur; it consists of a root, nodes or vertices, edges connecting the nodes and edge weights. Figure 1 gives an example of a branching tree in the context of modeling phenotypic resistance for TB. The root in this setting represents a wild type state (i.e. full sensitivity to all TB drugs); the nodes represent the acquisition of drug resistance to different drugs. The edges connecting the nodes signify that the event represented by the offspring (child) node can only occur given that the event represented by the parent node has already occurred; the edge weights are the conditional probabilities of these events. A node that has no offspring is called a leaf. In this paper, we will adopt the notation of Beerenwinkel et al where a branching tree $\mathcal{T}_k = (V, E, r, bp)$; V represents the set of events, E represents the set of edges, $r \in V$ is the root and bp is a probability mapping such that $bp: E \rightarrow [0, 1]$ [2].

Single branching trees may not be sufficient to describe the underlying processes of interest. To address this issue, Beerenwinkel et al proposed a model that includes multiple branching trees, referred to as a K-mutagenetic tree mixture model, to describe the order of acquisition of resistance mutations [2]. A K-mutagenetic tree mixture model contains not a set of trees (which were refer to as a tree structure) but also weights that correspond to the probability associated with each tree and probabilities bp that are associated with the edges. These mixture models often contain a special noise component or “star” tree, in which all nodes originate in the root. Figure 2 provides an example of a mixture model, in which the first tree is the “star” tree. We refer to a tree structure as the graph of the mixture model without the probability mapping $bp: E \rightarrow [0, 1]$, i.e. the collection of trees, $\mathcal{T}_k' = (V, E, r)$.

Desper et al derived a algorithm for making inferences from tree models based on the idea of a maximum-weight branching in a graph, and implemented these methods with data on comparative genome hybridization in cancer [1]. The approach of Beerenwinkel et al to construct mixtures of branching trees was used to estimate genetic pathways from cross-sectional data on mutations associated with HIV drug resistance [2]. Healy, DeGruttola and Hu extended this approach using bootstrap methods to estimate functions of the tree parameters that incorporate uncertainty in the tree set [3]. Their method was used to calculate functions of the tree set parameters and standard errors for these functions.

The primary aim of this paper is to propose a Bayesian approach to identify tree structures and to estimate the parameters that characterize the mixture models associated with these structures. This approach has the advantages of being easily extended to joint models or other more complex models that can incorporate measurement error; it can also accommodate prior information regarding cross-resistance. A Bayesian model can easily estimate parameters from hierarchal models, whereas extending the approach of Desper and Beerenwinkel to accommodate more complex models could be computationally intensive.

These methods are applied to cross-sectional data on phenotypic TB resistance data to provide insight regarding the order of acquisition of resistance to anti-tuberculosis antibiotics.

We apply our methods to drug-resistant TB, because it has emerged as an important public threat; several countries now face epidemics in which almost half of the reported TB cases are resistant to at least one drug and one-fifth have multidrug-resistant TB (MDR-TB) [4]. MDR-TB is defined as resistance to isoniazid and rifampicin, the two most important drugs included in standard first-line antibiotic regimens [5]. Extensively drug-resistant tuberculosis (XDR-TB) is defined as MDR-TB with additional resistance to a quinolone and to kanamycin, capreomycin, or amikacin [6]. As of January 2010, a total of 58 countries had reported at least one such case; XDR-TB is associated with lower cure rates and higher mortality than other less drug-resistant forms of TB [4, 6]. The emergence of MDR-TB and XDR-TB raises concerns of an epidemic of essentially an untreatable disease [7]. Efforts to assess the risk of continued appearance and spread of drug-resistance and to plan appropriate interventions require an understanding of how MDR and XDR resistance arises within individual hosts and subsequently spreads in populations. Knowledge of the sequence of events that lead to combination drug resistance and the conditional probabilities each of these individual events will improve our understanding of this process.

We illustrate our methods using data resulting from about 5000 drug-susceptibility tests (DST) performed in Peru on sputum samples that were cultured for Mycobacterium tuberculosis and tested for drug resistance. Recent work by Bergval et al. raises questions about the clinical relevance of estimates of drug resistance conferring mutation rates derived from in vitro studies [8]. They found that the majority of drug-resistant strains generated in the laboratory setting do not have the same mutations as those found in clinical isolates. Accordingly, examining patterns of drug resistance from clinical isolates may better inform our understanding of how drug resistance, and multiple drug resistance in particular, arises within human hosts.

2. Methods

2.1. Tree Based Analysis

We continue to follow the notation of Beerenwinkel et al regarding mixture models that characterize mutation events[2]. Let K be the number of trees in a mixture model, defined as

$$\mathcal{M}=\sum_{k=1}^{K}\alpha_k\mathcal{T}_k \text{ with } \alpha_k \in [0,1] \text{ and } \sum_{k=1}^{K}\alpha_k=1$$

generates the random variable $Y=\sum_{k=1}^{K}\Delta_k Y_k$ where $\Delta_1, \ldots, \Delta_K \in [0, 1]$ are binary random variable with $P(\Delta_k = 1) = \alpha_k$ and $Y_1, \ldots, Y_K$ are multivariate discrete random variables that are distributed according to branching trees $\mathcal{T}_k = (V, E_k, r, bp_k)$.

Calculation of the likelihood associated with the random variable Y requires the definition of states, which are determined by the pattern of events for a particular observation. In this paper, an event is phenotypic resistance to a given drug; the number of possible states is 2^n, where n is the number of ordered drugs. Y is a multinomial random variable with probability vector **p**, where the elements of **p** represent the probability associated with each possible state. As an example, Table 1 lists the possible states for a drug regimen consisting of three drugs: drug-1, drug-2 and drug-3. S_{000} represents the state in which a patient is not resistant

to any of the drugs. S_{100} represents the state in which the patient is resistant to drug-1, but not to drug-2 or drug-3. In this example, Y is distributed as a multinomial with corresponding probabilities $p_{000}\cdots p_{111}$. Each p_{abc} is a function of the parameters of a branching tree. For example, the probability, p_{100} of being in S_{100} given the mixture model in Figure 2 is:

$$p_{100}=\alpha_1 bp_{1_1}(1-bp_{1_2})(1-bp_{1_3})+(1-\alpha_1)bp_{2_1}(1-bp_{2_2})(1-bp_{2_{3|1}})$$

The p_{abc} are potentially complicated functions of the mixture model parameters, especially as the number of nodes increases.

We use a Bayesian approach to estimate the posterior distributions of mixture model parameters, where bp_k are the branching tree parameters and α_k are the tree weights. Define θ_j as the set of tree weight parameters and branching tree parameters for a given mixture model. To calculate $P(\theta_j|Y)$, we use a likelihood based on the state of each observation, and place non-informative priors on the edge and the tree weights. Posterior distribution for the edge and tree weight parameters are obtained from an MCMC implemented using WinBUGS. Because there can exist a group of models sharing similarities that equally describe the data well, we use the deviance information criteria (DIC), which is not based on the belief of one "true" model, to choose a mixture model from the set of all possible such models. The DIC is a generalisation of the AIC and for non-hierarchical models, such as the one in this paper, the two measurements are aproximately the same. For a more complete treatment of the DIC the interested reader is referred to Spiegelhalter et al [9]. As with the DIC, the full marginal likelihood could be approximated by evaluating the likelihood at the posterior mode. Unlike the DIC, however, this measurement would not penalize for an increase in the number of parameters. We can also use $P(M_j|Y)$ to compare models, where M_j is a binary random variable that takes on the value of 1 when the model is the j^{th} mixture model. $P(M_j|Y)$ can be estimated using Bayes Theorem from $P(Y|M_j)$ and $P(M_j)$. If we assume a uniform distribution on M, then $P(M_j|Y)=P(Y|M_j)/\sum_{i=1}^{I}P(Y|M_i)$ where I represents the number of models in the candidate set. Kass & Raftery review several methods to estimate $P(Y|M_j)$; we follow the simplest method that uses samples from the posterior distribution [10].

$$P(Y|M_j)=\left[\frac{1}{T}\sum_{t=1}^{T}\frac{1}{P(Y|\theta^t,M_j)}\right]^{-1}$$

where θ^t is a draw from $P(\theta|Y,M_j)$. Our methods assume the number of trees, K, must be known. A cross validation approach, suggested by Beerenwinkel et al, is used to compute the optimal number of trees in a given mixture model [2]. The software Mtreemix is used to carry out this approach [11]. It should be noted that when using this software, the inclusion of a star tree is automatic even when the number of trees in the model is one.

To reduce the set of candidate models, we focus on reducing the set of candidate tree structures associated with those models. We propose four ways to accomplish this goal. The first is to enumerate all possible tree structures a procedure that can be very cumbersome if the number of drugs is large. The second is to consider restrictions implied by the data themselves; the third, to restrict tree structures to those that are near some referent; and the fourth, to reduce the number of drugs under consideration.

The second approach above makes use of the fact that the presence of observed states rules out tree structures in which these are not possible; and the absence of certain states, rules out structures in which they should be fairly common. For example, referring to Figure 1, if there exists one or more individuals who are resistant to $drug_3$ and not to $drug_1$, this tree structure cannot describe the data. Of course, including the star tree (Figure 2) in the tree structure makes all states possible, but such inclusion may not be desirable. When there are states that are not observed, the set of candidate tree structures can be restricted to those that represent paths to observed states. The tree in Figure 1, for example, allows the possibility that an isolate is resistant to $drug_2$, but not to $drug_1$. If this state is not observed, we may consider excluding this tree structure if the state would have fairly high probability of occurrence.

The third approach to choosing candidate tree structures considers those that occur in bootstrap samples as well as others that are near them. To develop a distance measure, we first define a “move” as the number of differences in the sets of edges for each tree. Distance between two trees may be defined as the number of edges that have been moved. For example, to go from the tree in Figure 3(a) to that in Figure 3(b), requires moving drug-3 from below drug-1 to below drug-2. These two trees are thus one “move” away from each other. By contrast, moving from the tree in Figure 3(a) to that in 3(c) requires starting a new branch from the root with drug-3 and then moving drug-1 underneath drug-3. This maneuver requires two “moves.” Distance between a single tree structure and a set of tree structures is defined as the minimum of the distances between the single tree structure and each of tree structures in the set.

The final method to reduce the set of possible tree structures is to reduce the number of drugs under consideration. Contingency table analysis can be used to assess the dependence of resistance to a given drugs on resistance to the other drugs. Correlation analysis can provide useful guidance. If resistance to a specific drug is jointly independent of that of all other drugs, we may eliminate it from the analysis. Alternatively, a high dependence between two drugs in resistance status may allow us to eliminate one from consideration.

3. Identifiability & Simulation

3.1. Simulation: One Tree

This section explores the performance of our methods when data are simulated from single trees. We refer to the tree structure associated with the data generating model as the data generating tree structure. Our simulation study used data generated from each of the trees shown in Figure 4; these consisted of trees with five nodes and one leaf (4(a)), four nodes and two leaves (4(b)) and five nodes and three leaves (4(c)). A tree with one leaf imposes the maximum restrictions by limiting the number of possible states. For the tree in Figure 4(c), the structure and edge weights were chosen to mirror the results when modeling the data from the results of DST on Category 1 drugs; for the other trees the structures and edge weights were selected arbitrarily. For these examples, the set of all possible tree structures could be enumerated because of the small number of nodes (≤ 5). The set of candidate structures was reduced by eliminating any tree structure associated with a mixture model with zero likelihood given the simulated data, leaving only structures for which the observed states in the simulated data were possible. In each simulation, the probability of the model given the data, $P(M|Y)$, is 1 for the data-generating mixture model and 0 for all other models in the candidate set. In each simulation, the model with the lowest DIC was the data-generating model. Thus, when the original model contains a single tree, our method always returns the correct mixture model. This may be due to the large number of restrictions to the likelihood, which aids in the estimation of the conditional probabilities. Beerenwinkel’s method also identifies the correct model. For every parameter, the 95% credible interval

contains the true value. The results for Figure 4(c) are shown in Table 2; the results for the other trees show a similarly good performance. These results show that using Bayesian methods produces accurate results in these settings.

3.2. Simulation: More than one tree

Data were simulated from the mixture model containing two trees, each of which has one leaf, as shown in Figure 5. The set of candidate tree structures were the most frequent tree structures associated with the mixture models selected by the methods of Beerenwinkel et al applied to 30 bootstrap samples as well as all trees one move away from the structures chosen by the bootstrap method [2]. The latter contained a structure which is similar to the data generating tree structure except for the inclusion of a star tree. This model was selected by the DIC and is shown in Figure 6, along with the posterior modes and standard errors corresponding to each edge weight. These values and the 95% credible interval for each parameter for this model are shown in Table 3. We note that the weight on the star tree is extremely close to zero; therefore, it is reasonable to simplify this mixture model by removing the star tree, leaving us with the actual data generating tree structure. The star tree is not, in fact, necessary since all states from the simulated data can be reached by the remaining trees in the tree structure. Removing the star tree is associated with further reduction in the DIC; hence our approach does lead us to the correct tree structure. The results from fitting the model without the star tree are also shown in Table 3. Regardless of whether the star tree is included, each credible interval captures the true value of the tree weights and the edge weights, although some of these intervals are quite large. In addition, for both models, the posterior distributions are centered around the correct values.

When data were generated from a more complex mixture model containing multiple leaves and missing nodes (edges with weight of zero) as shown in Figure 7, the tree structures that resulted from the bootstrap samples were not as similar to the generating tree structures as those described above. As before, we included all structures that were one move away from the set of candidate tree structures created from applying Beerenwinkel's method to 30 bootstrap samples. The model favored by the DIC is shown in Figure 8; the posterior mode and 95% credible interval for each parameter are shown in Table 4. All of the credible intervals for the edges that do appear in the data generating tree structure contain the true value. However, the third tree of this structure does not contain any edges that appear in the data generating tree structure. This simulation illustrates that for data generating structures with a high degree of complexity, the cross-sectional data are not sufficient for accurate estimation of the mixture model. To show how well our model estimates the posterior distribution when the correct tree structure is specified, we fit the structure shown in Figure 7 and the results are found in Table 4. Our methods do center the posterior distribution around the correct value and the standard errors are small.

Identifiability cannot be assured when tree structures are complex because the trees do not provide a sufficient number of restrictions, and thereby permit different mixture models to be equally good in describing the data–a situation that is implied when there are two or more models with similar DIC. Because the tree parameters are complicated functions of the multinomial parameters described above, it is difficult to provide a general set of conditions to assure that parameters will be identifiable. In the case where the number of parameters exceeds the degrees of freedom for the multinomial model ($2^n - 1$), parameters will not be identifiable, but this condition is not necessary for non-identifiability. One approach to assessing identifiability is to inspect the solution(s) to the system of equations derived from the relationship between the multinomial parameters and the mixture model parameters. The 2^n multinomial parameters are functions of the tree weights and edge weights of the mixture model. By using the MLE of the multinomial parameters (which is just the observed proportions) these functions create a system of equations with the number of unknown

variables equal to the number of parameters in the mixture model. If there are multiple solutions to this system, the parameters will be unidentifiable; this situation leads to multiple models that give similar posterior distributions for the multinomial probabilities and hence wide credible intervals for the mixture model parameters. If there is a unique solution, the mixture model parameters will be simple functions of the MLE and no inference can be made. If there is no solution, we can use our methods to estimate the distribution of these parameters and make meaningful inferences. Beerenwinkel and Drton give a detailed theoretical discussion of identifiability for mixture models in their chapter in Algebraic Statistics for Computational Biology [12]. In one of the main results in this chapter, the authors show for $n \geq 3$, the dimension of a 2-tree mixture model, one tree of which is a star tree with equal weights, is $n + 2$. By induction, it can be shown that $2^n - 1 > n + 2$; therefore the degrees of freedom is always greater than the number of parameters. This result is consistent with our discussion of identifiability above.

4. Application

Most of the 5166 patients on whom we have information about resistance had already received several unsuccessful courses of TB therapy. Although the detailed individual treatment history is known for only a small subset of this patient group, most received drugs in the following sequence (Mitnick personal communication):

- Category I: isoniazid (H), rifampicin (R), ethambutol (E), pyrazinamide (Z)
- Category II: isoniazid, rifampicin, ethambutol, pyrazinamide, streptomycin (S)
- Retreatment: isoniazid, thiacetazone, rifampicin, ethambutol (ETA), pyrazinamide, kanamycin
- “Standard Treatment Regimen” (STR): kanamycin, pyrazinamide, ethambutol, ethionamide, ciprofloxacin

It should be noted that only a small percentage received retreatment and a significant number did not receive STR. In the majority of isolates, susceptibility testing was performed to the following 12 drugs or drug-groups: capreomycin, cycloserine, ethambutol, ethionamide, isoniazid, kanamycin or amikcacin, PAS, pyrazinamide, rifampicin, streptomycin, any 1st-generation-fluoroquinolones (ciprofloxacin, ofloxacin), and any later-generation fluoroquinolones (gatifloxacin, levofloxacin, moxifloxacin).

The focus of our application of the tree-based methods is on inference regarding the order of acquisition of resistance to drugs in the Category 2 regimen used to treat TB patients with a history of prior TB treatment. Category 2 includes the following drugs: isoniazid (H), rifampicin (R), ethambutol (E), pyrazinamide (Z) and streptomycin (S). Because most of the patients had failed previous courses of TB therapy, a high proportion are resistant to both isoniazid and rifampicin. As expected, when the model was fitted to the data, the posterior mode for the weight on the edge between rifampicin and isoniazid was approximately 0.9958, reflecting a high degree of concurrency in resistance to the two drugs in our sample. Therefore, the nodes of isoniazid and rifampicin could be exchanged; we removed the rifampicin node and examined the order of acquisition of resistance to the remaining four drugs.

Using the susceptibility test results for isoniazid (H), ethambutol (E), pyrazinamide (Z) and streptomycin (S), Beerenwinkel’s cross-validation method in the Mtreemix package suggested using two trees in the mixture model (including a star tree). The most frequent tree structures that resulted from mixture models from 30 bootstrap samples were chosen to create a set of four candidate tree structures consisting of two or three trees each. The four models were fit to the data using WinBUGS and the DIC was used to compare them; the

model with the lowest DIC is shown in Figure 9. The posterior mode, standard errors and 95% credible intervals are shown in Table 5. The trees implied that the presence of isoniazid resistance increases the probability of developing additional resistance to other drugs, especially ethambutol. We also note that except in the star tree, streptomycin and ethambutol branch off isoniazid, but resistance to either appears to arise independently of the other. The simulation studies caution us about inference in a setting of multiple trees and leaves. The tree structure here, however, provides a somewhat better basis for inference than those with multiple trees and leaves described above, because of the similarity of the two non-star trees: they differ by only one edge- that with child pyrazinamide. This similarity increases the number of restrictions implied by the structure. In addition, the finding regarding ethambutol is supported by both of the non-star trees, and hence does not depend on accurate estimation of the weights on them. Nonetheless, the credible intervals for the parameters on the middle tree are large (Table 5) possibly due to the fact that this tree is so similar to the other two trees in the structure.

We also considered trees that were one move away from the tree structure in Figure 9. The mixture model with the lowest DIC among these additional models (113.7 compared to 122.2) contained a tree structure that differed from that above only in the placement of the Z node in the tree on the right, which moved from under ethambutol node to under the wildtype node. While the parameter estimates changed somewhat (Table 5), the results were qualitatively similar for the two models.

5. Discussion

Branching trees are a powerful tool to graphically display the order in which a sequence of events occur. In this paper, we develop a Bayesian method to estimate branching trees parameters for investigation of the order of acquisition of TB drug resistance using information on the patterns of drug resistance observed in clinical isolates from treatment-experienced TB patients. From our simulations, we conclude that data-generating tree structures containing few leaves permit more reliable inference regarding mixture models. When there are many leaves in the data-generating tree structure, additional assumptions may be required. However, similarities among candidate tree structures may provide useful information about the underlying data-generating process.

We note that we could use Bayesian methods to estimate the joint posterior distribution of the state probabilities and therefore the conditional distributions of interest (e.g. probability of the occurrence of resistance to one drug given the presence of resistance to another), without having to model trees. A tree-based approach, however, provides additional information about possible pathways to multiple drug resistance. Quantifying the probability of each individual tree allows us to estimate the proportion of the population whose experience is represented by the pathway for that tree. Tree structures are essentially restrictions on the conditional probabilities of interest. In general, it may not be possible to infer these restrictions when estimating the joint distribution directly, especially when the population is split into sub-populations that follow different pathways. The tree structures provide an ordering that aids in making inference. An additional advantage is that graphical representations of the pathways provide an interpretation of joint distributions.

Our methods are applied to phenotypic data but are equally applicable to genetic data. Although inference is challenging in this setting, the clinical relevance of the ordering of resistant phenotypes may be greater than the ordering of genotypes. Studies have shown that the costs of drug-resistance conferring mutations is dependent on the host environment [13]. Because the correlation between growth rates in laboratory settings and fitness in real life is

unclear, it is advantageous to use phenotypic data which already accounts for fitness and compensation costs [14, 15].

Our results have the advantage of deriving from clinical isolates rather than strains grown in-vitro. While laboratory experiments about the acquisition of resistance provide the researcher with full control over timing of the testing and choice of drugs to which the strains are exposed, the behavior of the bacillus in-vitro may not reflect what happens in human hosts. [13, 16]. In addition, drugs are given individually in laboratory studies of resistance acquisition (either in culture or animal models), whereas patients are treated with multi-drug combinations. Data from these studies do not account for the environmental changes that occur in the presence of additional drugs.

Our application and interpretations are limited to development of drug resistance within the hosts previously treated for TB. Because most patients had failed multiple courses of therapy, the isolates in the study represent strains that had an opportunity to acquire resistance to multiple first-and second-line antibiotics. The fact that most isolates were MDR (INH & RIF resistance), precludes inference into ordering of resistance acquisition early in treatment; further studies in which isolates were collected for individuals in their first or second treatment courses for tuberculosis would be needed to generate better insight into the ordering of resistance to first-line agents.

Our methods and conclusions are also limited by the absence of information regarding drug treatment history. Knowledge of treatment regimens can enhance our inferences. For example, if streptomycin had been the child node to either pyrazinamide or ethambutol or both, knowledge that these drugs had been given concurrently would permit the inference that resistance to pyrazinamide or ethambutol predisposes resistance to streptomycin. However, if most of the patients received category 1 drugs (isoniazid, rifampicin, ethambutol and pyrazinamide) prior to those in category 2 (category 1 + streptomycin), then timing rather than predisposition might explain this ordering. Large data sets that have both treatment history and resistance phenotype would allow such investigations. The methods can also be extended to accommodate measurement error and some longitudinal data, should such information be available.

In summary, the methods proposed are useful in suggesting longitudinal sequences of events from very limited cross-sectional data, whether the events are based on phenotype or genotype. Availability of both types of data, along with other relevant information would permit more precise inference and the flexible Bayesian methods can be easily adapted.

Acknowledgments

Contract/grant sponsor: NIAID R01 51164, 2 T32 AI 7358-21, 5 T35 ES 7293-9

References

1. Desper R, Jiang F, Kallioniemi OP, Moch H, Papadimitriou CH, Schaffer AA. Inferring tree models for oncogenesis from comparative genome hybridization data. J Comput Biol. 1999; 6(1):37–51. [PubMed: 10223663]
2. Beerenwinkel N, Rahnenfhrer J, Dumer M, Hoffmann D, Kaiser R, Selbig J, Lengauer T. Learning multiple evolutionary pathways from cross-sectional data. J Comput Biol. 2005; 12(6):584–598.10.1089/cmb.2005.12.584. [PubMed: 16108705]
3. Healy BC, DeGruttola VG, Hu C. Accommodating uncertainty in a tree set for function estimation. Stat Appl Genet Mol Biol. 2008; 7(1):Article5.10.2202/1544-6115.1324 [PubMed: 18312210]
4. WHO. Technical Report. World Health Organization; Geneva, Switzerland: February. 2008 Anti-tuberculosis drug resistance in the world: Report no. 4.

5. Mitnick CD, Shin SS, Seung KJ, Rich ML, Atwood SS, Furin JJ, Fitzmaurice GM, Viru FAA, Appleton SC, Bayona JN, et al. Comprehensive treatment of extensively drug-resistant tuberculosis. N Engl J Med. 2008; 359(6):563–574.10.1056/NEJMoa0800106 [PubMed: 18687637]
6. WHO. Technical Report. World Health Organization; Geneva, Switzerland: March. 2010 Multidrug and extensively drug-resistant tuberculosis: 2010 global report on surveillance and response.
7. Raviglione MC. Facing extensively drug-resistant tuberculosis–a hope and a challenge. N Engl J Med. 2008; 359(6):636–638.10.1056/NEJMe0804906 [PubMed: 18687645]
8. Bergval IL, Schuitema ARJ, Klatser PR, Anthony RM. Resistant mutants of mycobacterium tuberculosis selected in vitro do not reflect the in vivo mechanism of isoniazid resistance. J Antimicrob Chemother. 2009; 64(3):515–523.10.1093/jac/dkp237 [PubMed: 19578178]
9. Spiegelhalter DJ, Best NG, Carlin BP, Van Der Linde A. Bayesian measures of model complexity and fit. Journal of the Royal Statistical Society: Series B (Statistical Methodology). 2002; 64(4): 583–639.10.1111/1467-9868.00353
10. Kass R, Raftery A. Bayes factors. J Amer Statist Assoc. 1995; 90:773–395.
11. Beerenwinkel N, Rahnenführer J, Kaiser R, Hoffmann D, Selbig J, Lengauer T. Mtreemix: a software package for learning and using mixture models of mutagenetic trees. Bioinformatics. 2005; 21(9):2106–2107.10.1093/bioinformatics/bti274 [PubMed: 15657098]
12. Pachter, L.; Sturmfels, B. Algebraic statistics for computational biology. Cambridge University Press; Cambridge: 2005.
13. Bjorkman J, Nagaev I, Berg OG, Hughes D, Andersson DI. Effects of environment on compensatory mutations to ameliorate costs of antibiotic resistance. Science. 2000; 287(5457): 1479–1482.10.1126/science.287.5457.1479 [PubMed: 10688795]
14. Luciani F, Sisson SA, Jiang H, Francis AR, Tanaka MM. The epidemiological fitness cost of drug resistance in mycobacterium tuberculosis. Proc Natl Acad Sci U S A. 2009; 106(34):14 711–14 715.10.1073/pnas.0902437106
15. Cohen T, Sommers B, Murray M. The effect of drug resistance on the fitness of mycobacterium tuberculosis. Lancet Infect Dis. 2003; 3(1):13–21.10.1016/S1473-3099(03)00483-3 [PubMed: 12505028]
16. Gagneux S, Long CD, Small PM, Van T, Schoolnik GK, Bohannan BJM. The competitive cost of antibiotic resistance in mycobacterium tuberculosis. Science. 2006; 312(5782):1944–1946.10.1126/science.1124410 [PubMed: 16809538]

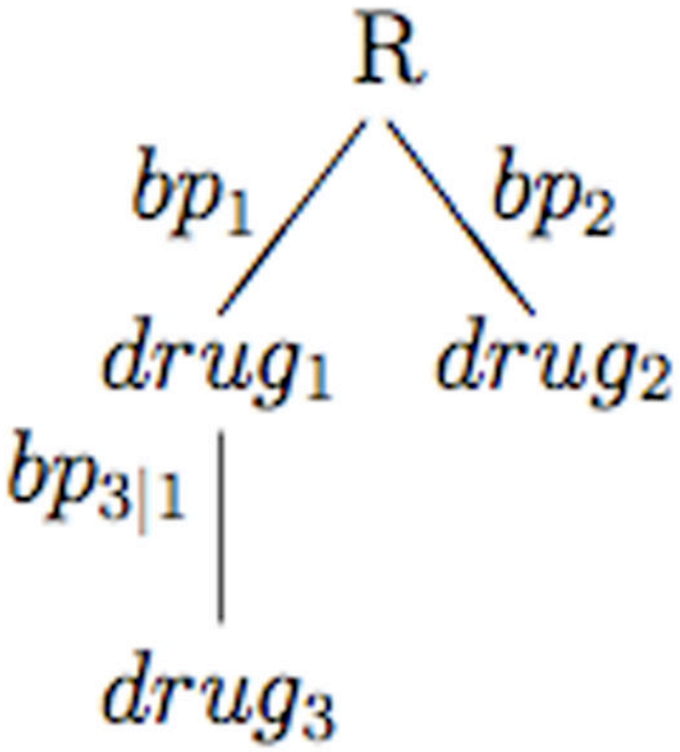


Figure 1.

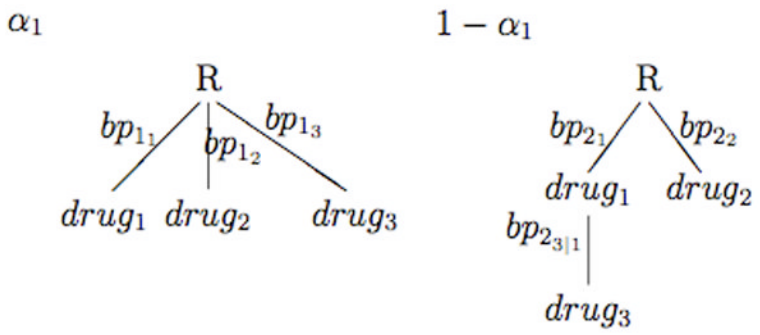


Figure 2.

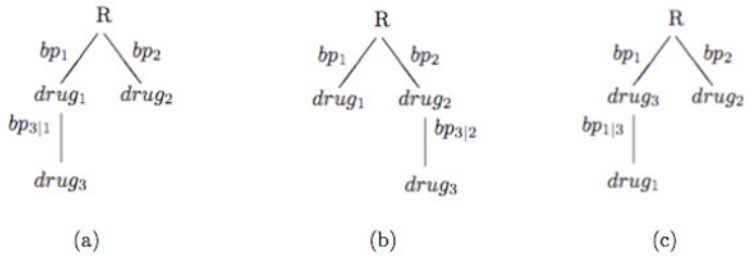
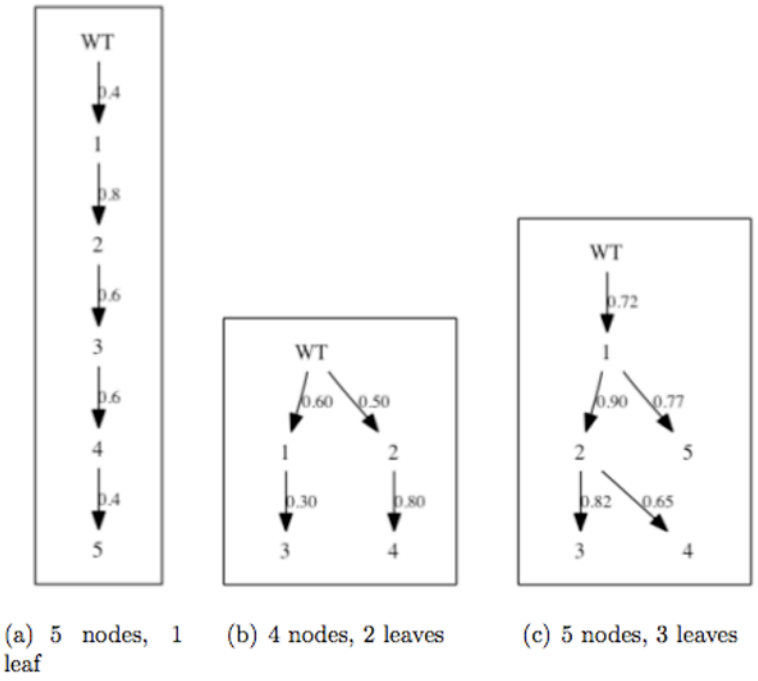


Figure 3.



(a) 5 nodes, 1 leaf (b) 4 nodes, 2 leaves (c) 5 nodes, 3 leaves

Figure 4.

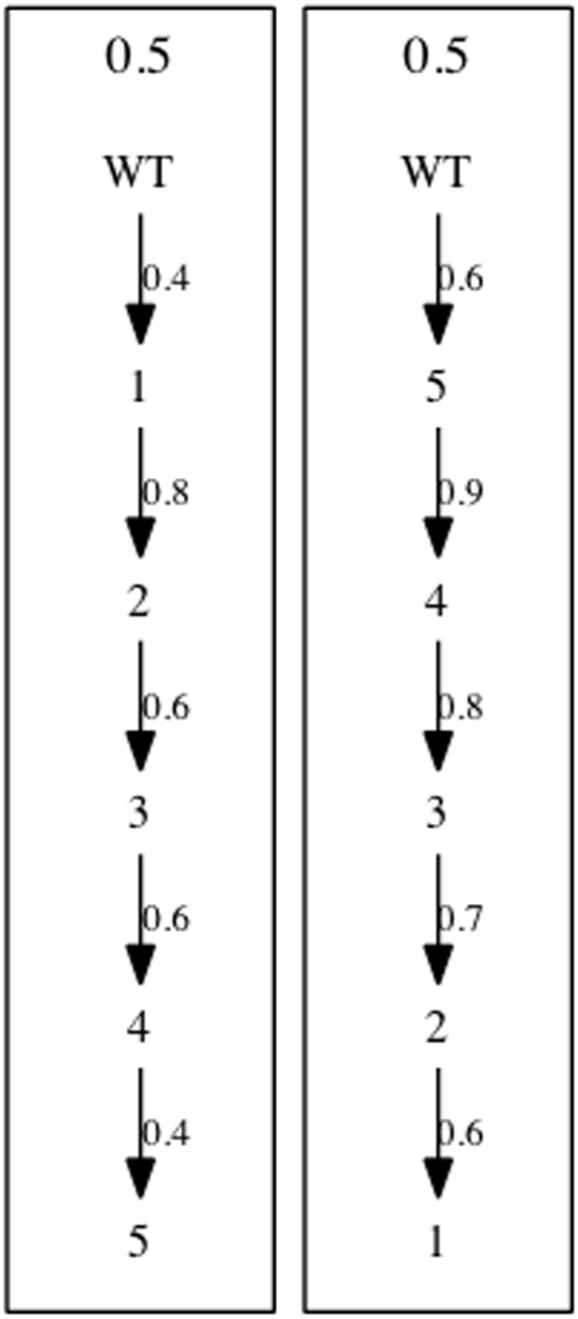


Figure 5.

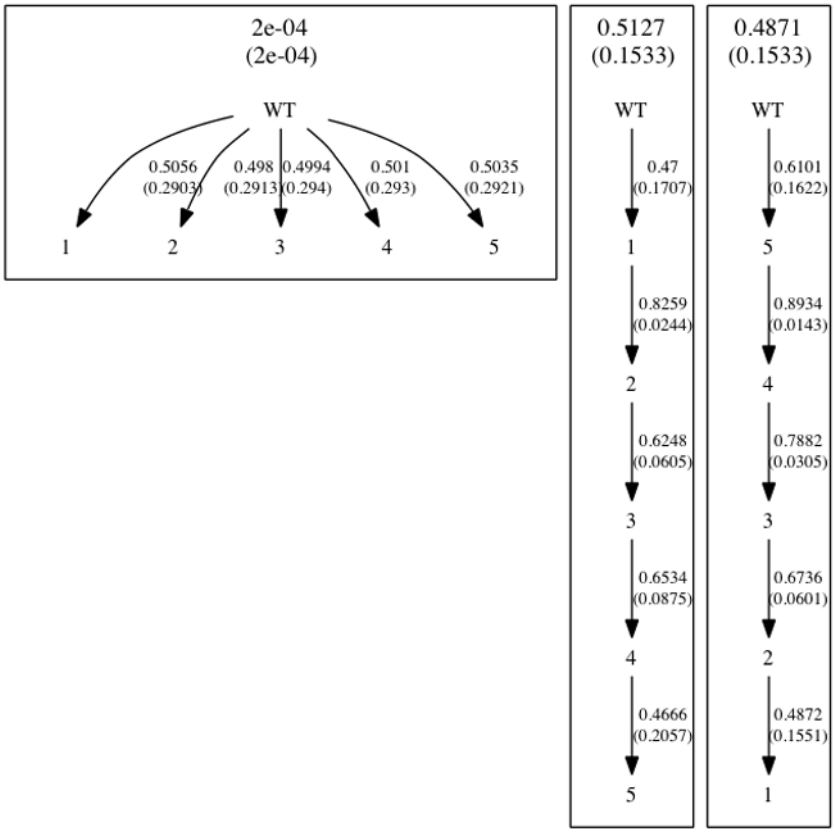


Figure 6.

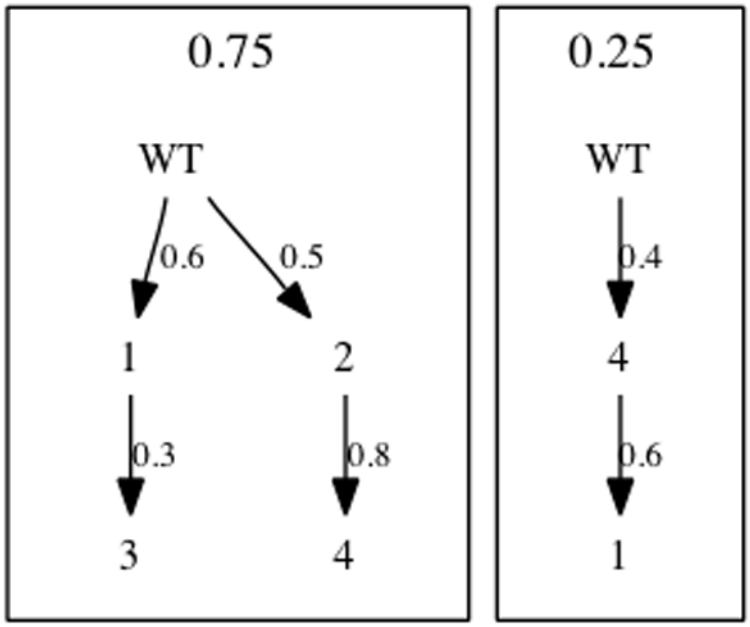


Figure 7.

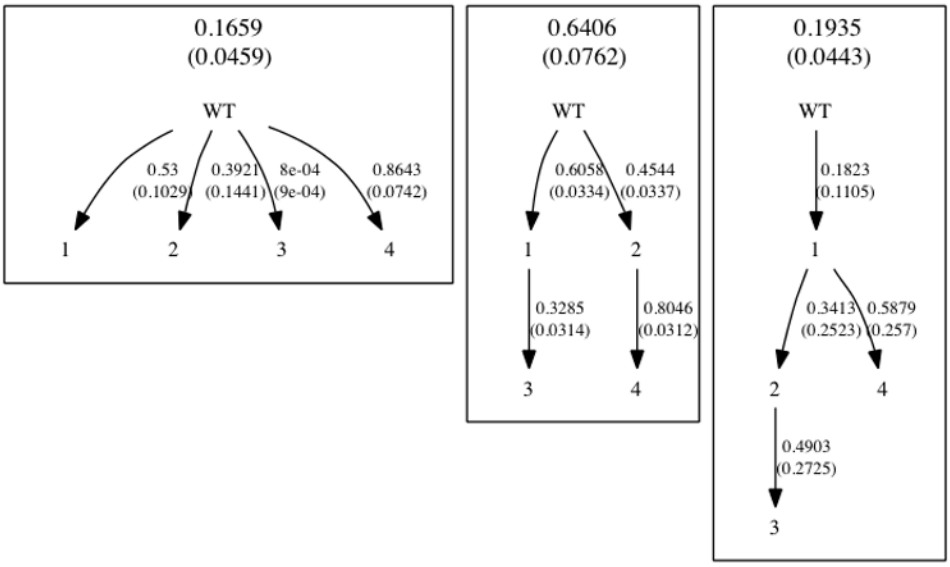


Figure 8.

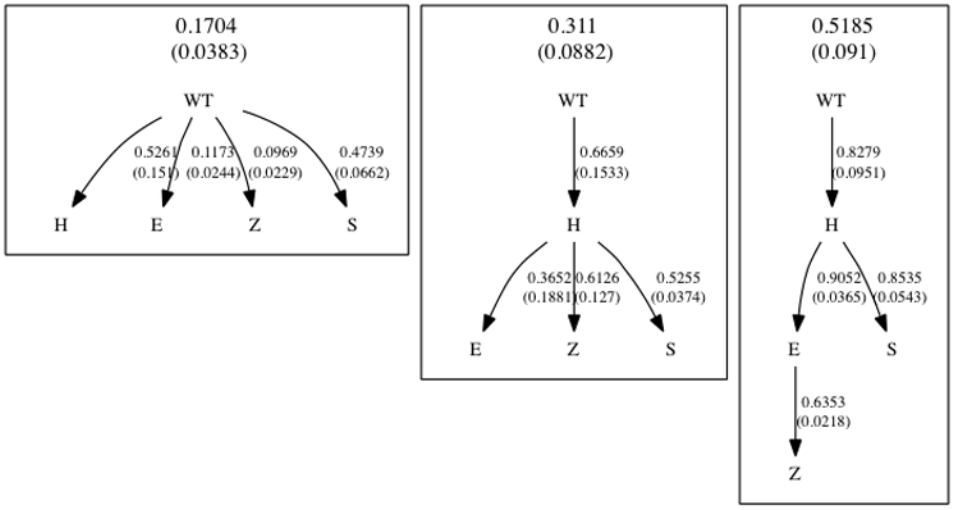


Figure 9.

Table 1

The eight possible states regarding drug resistance from a treatment consisting of three drugs.

State	Resistant to $drug_1$	Resistant to $drug_2$	Resistant to $drug_3$
S_{000}	No	No	No
S_{100}	Yes	No	No
S_{010}	No	Yes	No
S_{001}	No	No	Yes
S_{110}	Yes	Yes	No
S_{101}	Yes	No	Yes
S_{011}	No	Yes	Yes
S_{111}	Yes	Yes	Yes

Table 2

Data were simulated from a single tree shown in Figure 4(c). The posterior mode and 95% credible interval from the model with the lowest DIC is shown here.

Parameter	True Value	Posterior Mode	95% Credible Interval
WT→1	0.72	0.7197	(0.711, 0.7285)
1→2	0.90	0.9033	(0.8964, 0.9104)
2→3	0.82	0.8235	(0.8137, 0.833)
3→4	0.65	0.6588	(0.6469, 0.6705)
4→5	0.77	0.7739	(0.7641, 0.7833)

Table 3

Simulation results from the model with the lowest DIC shown in Figure 6. The true value, posterior mode and 95% credible interval are displayed. The DIC when fitting the model in Figure 6 is 93.84. When the star tree was removed the DIC decreased to 89.49, favoring the removal of the star tree.

			Figure 6 w/ star tree	Figure 6 w/o star tree
Parameter	**Tree**	**Original**	**Post. Mode (95% CI)**	**Post. Mode (95% CI)**
Tree Weight	1	-	1.8e-4 (4.1e-6,6.8e-4)	-
Tree Weight	2	0.5	0.51 (0.22,0.72)	0.57 (0.3,0.76)
Tree Weight	3	0.5	0.49 (0.28,0.78)	0.43 (0.24,0.7)
WT→1	1	-	0.51 (0.026,0.98)	-
WT→2	1	-	0.50 (0.028,0.97)	-
WT→3	1	-	0.50 (0.025,0.98)	-
WT→4	1	-	0.50 (0.023,0.97)	-
WT→5	1	-	0.50 (0.023,0.97)	-
WT→1	2	0.4	0.47 (0.30,0.87)	0.40 (0.27,0.64)
1→2	2	0.8	0.83 (0.78,0.87)	0.82 (0.78,0.87)
2→3	2	0.6	0.62 (0.52,0.74)	0.62 (0.52,0.73)
3→4	2	0.6	0.65 (0.49,0.80)	0.65 (0.49,0.79)
4→5	2	0.4	0.47 (0.044,0.76)	0.45 (0.038,0.75)
2→1	3	0.6	0.49 (0.046,0.64)	0.49 (0.07,0.64)
3→2	3	0.7	0.67 (0.52,0.74)	0.68 (0.53,0.74)
4→3	3	0.8	0.79 (0.71,0.83)	0.79 (0.72,0.83)
5→4	3	0.9	0.89 (0.86,0.91)	0.89 (0.86,0.91)
WT→5	3	0.6	0.61 (0.36,0.93)	0.67 (0.42,0.98)

Table 4

Simulation results from the model with the lowest DIC shown in Figure 8. The true value, posterior mode and 95% credible interval are displayed. In the third tree of the structure in Figure 8, it should be noted that there are edges that do not appear in the data generating tree structure. The DIC for this model is 98.79. This table also shows how well our model estimates the posterior distribution when the correct model is specified (found in the column labelled “Original Model”). The DIC for this model is 102.70.

			Figure 8	Original Model
Parameter	**Tree**	**Original**	**Post. Mode (95% CI)**	**Post. Mode (95% CI)**
Tree Weight	1	-	0.17 (0.081,0.27)	-
Tree Weight	2	0.75	0.64 (0.42,0.74)	0.76 (0.74,0.77)
Tree Weight	3	0.25	0.19 (0.14,0.32)	0.24 (0.23,0.26)
WT→1	1	-	0.53 (0.25,0.63)	-
WT→2	1	-	0.39 (0.099,0.67)	-
WT→3	1	-	8.5e-4 (2.1e-5,3.3e-3)	-
WT→4	1	-	0.86 (0.72,0.99)	-
WT→1	2	0.6	0.61 (0.55,0.70)	0.60 (0.58,0.62)
WT→2	2	0.5	0.45 (0.37,0.50)	0.50 (0.48,0.51)
1→3	2	0.3	0.33 (0.29,0.42)	0.29 (0.28,0.30)
2→4	2	0.8	0.80 (0.77,0.90)	0.81 (0.79,0.82)
4→1	3	0.6	-	0.62 (0.59,0.65)
WT→4	3	0.4	-	0.40 (0.37,0.43)
WT→1	3	-	0.18 (8.8e-3,0.42)	-
1→2	3	-	0.34 (0.013,0.92)	-
2→3	3	-	0.49 (0.037,0.97)	-
1→4	3	-	0.59 (0.053,0.97)	-

Table 5

Column 1 shows the posterior mode and 95% credible interval from the model with the lowest DIC of the candidate set from the most frequent structures from the bootstrap samples (Figure 9) and column 2 displays the results from the model with the lowest DIC from the models that are one move away from the candidate set created from the bootstrap samples.

		Column 1 (Figure 9)	Column 2
Parameter	**Tree**	**Post. Mode (95% CI)**	**Post. Mode (95% CI)**
Tree Weight	1	0.17 (0.086,0.24)	0.20 (0.11,0.26)
Tree Weight	2	0.31 (0.14,0.47)	0.36 (0.26,0.48)
Tree Weight	3	0.52 (0.35,0.69)	0.44 (0.32,0.53)
WT→H	1	0.53 (0.13,0.73)	0.71 (0.54,0.78)
WT→E	1	0.12 (0.073,0.17)	0.16 (0.1,0.22)
WT→Z	1	0.097 (0.056,0.15)	0.057 (0.024,0.11)
WT→S	1	0.47 (0.34,0.6)	0.63 (0.55,0.75)
WT→H	2	0.67 (0.37,0.97)	0.39 (0.19,0.56)
H→E	2	0.37 (0.024,0.66)	0.33 (0.025,0.63)
H→Z	2	0.61 (0.4,0.92)	0.61 (0.37,0.93)
H→S	2	0.53 (0.45,0.6)	0.36 (0.035,0.55)
WT→H	3	0.83 (0.66,0.99)	0.98 (0.97,0.99)
H→E	3	0.91 (0.84,0.98)	0.93 (0.86,1.0)
H→Z	3	0.64 (0.59,0.68)	-
WT→Z	3	-	0.65 (0.62,0.68)
H→S	3	0.85 (0.78,0.98)	0.85 (0.78,0.98)