Model identifiability is a key component of any proof of statistical consistency. Identifiability means that it is possible to infer all of the model’s parameters given an infinite amount of data from the model. For phylogenetic inference under the F81+Γ model, for example, the parameters are the unrooted phylogenetic tree with branch lengths, the particular F81 rate matrix (Felsenstein 1981), and a shape parameter for the gamma (Γ) distribution describing the rate heterogeneity. The F81 rate matrix is particularly simple with the rate of mutating to state i depending only on the long-run frequency of state i. This model is not identifiable using only pairwise species comparisons, that is, the joint pairwise DNA state distributions (Steel 2009).

For any F81 matrix, any two distinct gamma distributions, and any set of four or more species, Steel (2009) showed that there are distinct topologies for each gamma distribution that give the same joint pairwise DNA sequence distributions for those species. Any statistical estimator of these model parameters using only pairwise comparisons, such as distance-based methods, will be trying to estimate two separate points in parameter space, violating the definition of consistency.

For the purposes of this work, we define generic identifiability to mean that the set of parameters for which a model is not identifiable has a smaller dimension than the whole parameter space. Wu and Susko (2010) proved generic identifiability for the general time reversible (GTR) + Γ model from pairwise comparisons. The model parameters are the phylogeny with branch lengths, the 4 × 4 instantaneous rate matrix describing GTR DNA evolution along the phylogeny, and a shape parameter for the gamma distribution. Specifically, they proved that for all but the F81 family of matrices and for all phylogenies with at least two distinct interspecies distances, the rate matrix, the shape of the gamma distribution, and the phylogeny and its branch lengths are identifiable from pairwise comparisons. Because a GTR Markov matrix is parameterized using three nonzero eigenvalues but the F81 subfamily of matrices is parameterized using only one nonzero eigenvalue, the F81 model is a lower dimensional subset of the whole GTR class. Similarly, given more than two taxa, phylogenies with only one value for all of their interspecies distances make up a lower dimensional subset of the whole of tree space.

For phylogenetic inference under GTR + Γ + I, we add an additional parameter, which is the proportion of invariable (I) sites (Gu et al. 1995). Rogers (2001) argued that this popular model was generically identifiable from pairwise comparisons and that all other aspects of Wald’s proof that maximum likelihood estimators are consistent held for phylogenetic inference. His argument of identifiability, however, contained a flaw (Allman et al. 2008).

Allman et al. (2008) gave a correct proof, using three-taxon comparisons, of identifiability for the model without invariable sites for generic GTR Markov matrices. The exceptional cases when using joint three-taxon distributions only involve conditions on the GTR rate matrices, thus eliminating the exception against phylogenies with only one interspecies distance. For 4 × 4 DNA matrices, the authors go on to handle the exceptional cases, so the specific results of Allman et al. (2008) show that the model with an F81 rate matrix and gamma-distributed rate heterogeneity is always identifiable using joint three-taxon distributions. Thus, although pairwise distributions can be the same for two distinct models with F81 rate matrices, three-taxon distributions cannot be. However, the sufficiency of pairwise comparisons is of interest because distance-based methods traditionally use only pairwise comparisons and proofs of their consistency rely on identifiability based only on pairwise information.

In this paper, we complete Rogers’ proof of generic identifiability for the model with invariable sites using pairwise comparisons for all but members of the F81 family of rate matrices on any phylogeny with more than two distinct interspecies distances. Our proof also works on phylogenies with only two distinct pairwise interspecies distances if the rate matrix has three distinct nonzero eigenvalues, a condition which rules out not only the F81 family of matrices but also families such as K2P (Kimura 1980) with only two distinct nonzero eigenvalues. However, we show in the next section that deciding whether we are in the generic case or in an exceptional case as well as the nature of the exceptional case (whether it comes from not having enough distinct pairwise interspecies distances or from not having enough distinct nonzero eigenvalues) can be determined solely from the pairwise comparisons. All of these exceptional cases may still be identifiable.
using information from the joint distributions of three or more taxa in which case maximum likelihood techniques would be consistent as would appropriately generalized distance-based methods (Ranwez and Gascuel 2002; Contois and Levy 2005).

Software packages such as PAUP (Swofford 1991) and PhyML (Guindon and Gascuel 2003) use a discrete gamma distribution introduced by Yang (1995) for computational efficiency and do not actually use the theoretical continuous version discussed in this paper. PHYLIP (Felsenstein 2005) uses either a discrete gamma distribution or a Bayesian method for incorporating the gamma distribution for rate heterogeneity efficiently into its software code. Identifiability of models using the discrete gamma distribution has not been analyzed except in the case where the number of discrete classes is fewer than the number of states in the transition matrix, so fewer than four for DNA (Allman and Rhodes 2006).

**Identifiability of the GRT +Γ+ I Model**

We state the formal result and briefly give Rogers’ proof up to the missing component. We sketch the idea behind the proof of the missing component in this section and give the rigorous details in the Appendix.

**Theorem 1.** Let T be an unrooted phylogenetic tree with three or more distinct interspecies distances. Let Q be any GTR instantaneous rate matrix with two or more distinct nonzero eigenvalues. Let R be a mixture of a gamma distribution with mean 1 and invariable sites. Then, sequence data produced under this model identifies this model using pairwise comparisons. That is, with infinite sequence data at the species tips, we can recover all of the model parameters in T, Q, and R used to generate the species data using only pairwise species comparisons. Additionally, if the tree has only two distinct interspecies distances but the rate matrix has three distinct nonzero eigenvalues, then the model is also identifiable.

We work only with four-state DNA rate matrices although everything herein would generalize to more states, such as for models of protein evolution. Rogers’ proof is nicely constructive in the sense that it not only says that the model can be identified but also explains how to identify the components.

**What Was Previously Proved**

Most of the components are easily identifiable.

1. The stationary nucleotide frequencies, \( \Pi = (\pi_A, \pi_C, \pi_G, \pi_T) \), can be identified simply by determining the relative proportion of A’s, C’s, G’s, and T’s in the species sequences. In general, the stationary nucleotide frequency is not enough to determine the rate matrix Q. However, if additionally we know that the rate matrix has only one nonzero eigenvalue, then we know that the rate matrix is in the F81 family and thus has the form

\[
Q = \begin{pmatrix}
A & C & G & T \\
A & -\pi_C & \pi_G & \pi_T \\
C & \pi_A & -\pi_G & \pi_T \\
G & \pi_A & \pi_C & -\pi_T \\
T & \pi_A & \pi_C & \pi_G & -
\end{pmatrix}
\]

where the dashes are understood to be what is required to make the row sum to zero.

2. The rate matrix Q is determined by its eigenvalues and its eigenvectors. The observed DNA state transition matrix for a pair of species separated by distance \( \tau \) is \( P(\tau) = E[\exp(\tau Q)] \), where \( r \) is the random rate from the mixture distribution of the gamma distribution plus invariable sites over which the expectation is taken. The observed matrix \( P(\tau) \) has the same eigenvectors as the rate matrix Q. The eigenvalues of the two matrices are also related. If the eigenvalues for Q are \((-\lambda_1, -\lambda_2, -\lambda_3, 0)\) (where \( 0 < \lambda_2 \leq \lambda_3 \leq \lambda_4 \)), then the eigenvalues for \( P(\tau) \) will be \( E[\exp(\tau \lambda_i)] \). These latter values are observable from the transition matrix, \( P(\tau) \). Let \( \mu(x) = E[\exp(-\tau)] = \pi + (1 - \pi)(1 + x/\alpha)^{-\alpha} \) be the moment generating function evaluated at \(-x\) for the rate distribution which is a mixture of invariable sites (with proportion \( \pi \), not to be confused with the stationary distribution frequencies \( \pi_A, \pi_C, \pi_G, \pi_T \), which have subscripts) plus a mean 1 gamma distribution (with proportion \((1 - \pi)\) and shape parameter \(\alpha\)). If \(\alpha\) and \(\pi\) were known, then the functional form of \(\mu\) is known and one can solve for the nonzero eigenvalues. That is, for the \( i \)th pair of species tips with distance \( \tau \) between them and for eigenvalue \( \lambda_i \), \( m_{ij} = E[\exp(-\tau \lambda_i)] = \mu(\tau \lambda_i) \) is observed and \( \tau \lambda_i = \mu^{-1}(m_{ij}) \).

3. Again, if the functional form of \(\mu\) was known, then all the interspecies distances can be determined by inverting. Knowing all pairwise distances exactly determines the phylogeny (Buneman 1971).

There is a small redundancy that all of this is known up to an arbitrary scale factor because the distances and rates always appear multiplied together. We can set the distances and the nonzero eigenvalues \( \lambda_i \) by fixing a pair of
species with pairwise distance $\tau_i$, we can determine how many distinct nonzero eigenvalues there are. If there is only one distinct nonzero eigenvalue, the matrix comes from the F81 family and one can determine the exact F81 matrix, as noted above. The observed moments also, for a fixed nonzero eigenvalue, determine how many distinct interspecies distances are in the phylogenetic tree. Thus, we can tell if we are in the exceptional case where there is only one or two distinct interspecies distances.

**What Was Incorrectly Proved**

The difficulty in the argument is determining the functional form for $\mu_i$; that is, in determining $\pi$ and $\alpha$. At the point in Rogers’ argument where he is trying to show that the form of the moment generating function is determined by the pairwise joint sequence distributions, the tree and the rate matrix describing DNA evolution generally have not yet been determined. Thus, the question is whether there can be two different distributions, the tree and the rate matrix describing DNA evolution, for two distinct interspecies distances with a rate matrix, as noted above. The observed moments also, for at least three distinct eigenvalues $\lambda_l$ and $\mu$ is another moment generating function for a rate distribution associated with interspecies distances $I_l$ and non-zero eigenvalues $I_i$, and $\mu$ is another moment generating function for a rate distribution associated with interspecies distances $\tau_i$ and nonzero eigenvalues $\lambda_i$. Rogers noted that if there were two distinct interspecies distances ($\tau_1 \neq \tau_2$), this could only happen if $\gamma^{-1}(\mu(\tau_1 \lambda'_i))/\gamma^{-1}(\mu(\tau_2 \lambda'_i)) = I_1/I_2$ for three distinct eigenvalues $\lambda_i$ of a generic rate matrix. The argument Rogers gave that this implies the two moment generating functions are identical was flawed (Rogers 2001, Allman et al. 2008).

We note the symmetry here played by the distances and the nonzero eigenvalues. If there are just two distinct nonzero eigenvalues ($\lambda_2 \neq \lambda_3$) but many values for the distances separating pairs of tips, then we have $\gamma^{-1}(\mu(\tau_1 \lambda'_2))/\gamma^{-1}(\mu(\tau_3 \lambda'_3)) = I_2/I_3$ for at least three distinct interspecies distances $\tau_i$ on a generic phylogeny. Our main result, Proposition 1, proved in the Appendix, states that $\gamma^{-1}(\mu(\tau_1 \lambda'_2))/\gamma^{-1}(\mu(\tau_3 \lambda'_3))$ equals a constant for at most two distinct values of $\tau$ unless the two moment generating functions $\gamma$ and $\mu$ are identical. A mathematically identical argument works for two distinct interspecies distances with a rate matrix with three distinct nonzero eigenvalues by showing that Rogers’ original ratio, $\gamma^{-1}(\mu(\tau_1 \lambda_i))/\gamma^{-1}(\mu(\tau_2 \lambda_i))$, can equal a given constant at most twice in the eigenvalue variable $\lambda$. We omit the statement and proof of this symmetrical result from the Appendix.

**The Ideas Behind Our Proof**

We make use of Rogers’ original and useful trick of considering $\gamma^{-1}(\mu(x))$, which changes the nonlinear multivariable problem of determining how many solutions are possible to the equations

$$\mu(\tau_i \lambda_i) = \pi + (1 - \tau)(1 + \tau \lambda_i/\alpha)^{-\alpha} = m_{ij}$$

in terms of the scale of the gamma distribution $\alpha$ and the proportion of invariable sites $\pi$ into a single variable calculus problem of determining how many times the function $\gamma^{-1}(\mu(\tau_i \lambda_i))/\gamma^{-1}(\mu(\tau_j \lambda_i))$ of the single variable $\tau$ crosses a horizontal line. A continuous, smooth function crosses a horizontal line at most once if it is monotone (increasing or decreasing); it crosses at most twice if its derivative is zero at most once; and it crosses at most three times if its derivative is zero at most twice, which happens if its second derivative is zero at most once. Those observations and some algebraic manipulations form the basis of the proof of this missing component in Rogers’ argument. The rigorous proof is provided in the Appendix.

This is the one place where Rogers’ proof is not constructive in that the identification of $\alpha$ and $\pi$ is not explicit at this point. In practice, for $m_{ij} = \mu(\tau_i \lambda_i) = \pi + (1 - \pi)(1 + \tau_i \lambda_i/\alpha)^{-\alpha}$, we would use three distinct interspecies distances $\tau_1, \tau_2$, and $\tau_3$, two distinct nonzero eigenvalues $\lambda_2$ and $\lambda_3$, and solve the equations

$$\mu^{-1}(m_{12})/\mu^{-1}(m_{13}) = \mu^{-1}(m_{22})/\mu^{-1}(m_{23}) = \mu^{-1}(m_{32})/\mu^{-1}(m_{33})$$

for $\alpha$ and $\pi$ using numerical software.

These equations can be solved exactly in this context because we have an infinite amount of data and thus have the observed moments $m_{ij}$ exactly. With estimates of these moments from a finite amount of pairwise data, a least squares approach derived from these theoretical equations for estimating the common quotient, $\alpha$, and $\pi$ could be developed. This approach is unnecessary for maximum likelihood methods that have natural ways for estimating these parameters but might be useful for distance-based methods that do not currently provide estimates for $\alpha$ and $\pi$.

**ACKNOWLEDGMENTS**

We thank John Rhodes, Elizabeth Allman, the three attentive reviewers—Cecile Ané, Mike Steel, and one who remained anonymous—and Associate Editor Olivier Gascuel for numerous corrections and helpful comments, which substantially improved the manuscript.

**REFERENCES**


**APPENDIX**

**Proof of Rogers’ Claim**

**Proposition 1 Assume**

\[ \mu(x) = \pi + (1-\pi)(1+x/\alpha)^{-\alpha} \]

\[ \nu(y) = p + (1-p)(1+y/a)^{-a} \]

are the parameterizations of the two moment generating functions (evaluated at \(-x\)) for the rate distribution which is a mixture of a gamma distribution with mean 1 and a point mass at zero for invariable sites. Also assume that \(\pi \geq p\). (If this is not true, reverse the roles of \(\nu\) and \(\mu\).) Furthermore, assume that the rate matrix \(Q\) in the model for DNA evolution associated with \(\mu\) has at least two distinct nonzero eigenvalues \((-\lambda_3\) and \(-\lambda_3\)) so that \(\lambda_2 < \lambda_3\). Let the corresponding eigenvalues associated with \(\nu\) be \(-l_2\) and \(-l_3\), which need not be distinct. Then the equation

\[ \frac{\nu^{-1}(\mu(x))}{\nu^{-1}(\mu(x))} = \frac{l_2}{l_3} = A \]

has at most two solutions in the variable \(x > 0\) whenever \(\mu \neq \nu\).

**Proof.** Note that

\[ \nu^{-1}(\mu(x)) = a \left\{ \frac{\pi - p + (1-\pi)(1+x/\alpha)^{-\alpha}}{1-p} \right\}^{-1/a} - 1 \]

and that showing (1) has at most two solutions for \(x > 0\) is equivalent to showing that the function

\[ f(x) = \left[ (\pi - p + (1-\pi)(1+x/\alpha)^{-\alpha})^{-1/a} - 1 \right] \]

satisfies

\[ f(x) - Af(Bx) = (1 - A)\left[ 1 - p \right]^{-1/a} \]

for at most two \(x > 0\). In this equation, \(A = l_2/l_3 \leq 1\) because the eigenvalues are ordered and \(B = \lambda_3/\lambda_2 > 1\). The inequality for \(B\) is strict by the hypothesis that the rate matrix of the model associated with \(\mu\) has at least two distinct nonzero eigenvalues. We further note that the equality holds at \(x = 0\), and we will bound the number of times that the derivative of \(f(x) - Af(Bx)\) is zero. This is equivalent to bounding the number of times \(f'(x)/f''(Bx) = AB\).

We will use the following equivalent formulas for the derivative:

\[ f'(x) = \frac{1 - \pi}{a} \left[ (\pi - p) + (1 - \pi)(1 + x/\alpha)^{-\alpha} \right]^{-\alpha} \]

\[ \left[ (1 - p)(1 + x/\alpha)^{-\alpha} \right]^{-\alpha} \]

\[ + (1 - \pi)\left[ (1 + x/\alpha)^{-\alpha} + \left(1 - \pi\right) \right]^{-\alpha} \]

\[ = \frac{1 - \pi}{a} \left[ (\pi - p)(1 + x/\alpha)^{-\alpha} \right]^{-\alpha} \]

\[ + (1 - \pi)\left[ (1 + x/\alpha)^{-\alpha} \right]^{-\alpha} \]

\[ + \left[ (1 - \pi)\left(1 + x/\alpha\right)^{-\alpha} \right]^{-\alpha} \]

If \(\alpha \leq a\), then using the second expression for the derivative we see that

\[ f'(x)/f''(Bx) = \frac{(1 + Bx/\alpha)}{1 + x/a} \left[ \left(1 - \pi\right)(1 + Bx/\alpha)^{-\alpha} + \left(1 - \pi\right) \right]^{-\alpha+1} \]

\[ + \left[ (1 - \pi)(1 + x/\alpha)^{-\alpha} \right]^{-\alpha} \]

Because both parts on the right hand side are increasing, \(f'(x)/f''(Bx)\) crosses \(AB\) at most once. Thus the equation \(f(x) - Af(Bx) = (1 - A)(1 - p)^{-1/a}\) has at most one additional solution for \(x > 0\). That the second term is increasing in \(x\) is not completely obvious. With some algebraic manipulation its derivative can be written as:

\[ \frac{d}{dx} \left[ \left(\pi - p\right)(1 + \frac{Bx}{\alpha})^{-\alpha} + \left(1 - \pi\right) \right] = \frac{\left(\pi - p\right)}{\left(\pi - p\right)^{\alpha+1} \left(1 + \frac{Bx}{\alpha}\right)^{\alpha+1}} \]

\[ \left(1 - \pi\right)^{-\alpha} \left(1 + \frac{Bx}{\alpha}\right)^{\alpha+1} \left(1 + \frac{x}{\alpha}\right)^{-\alpha} \]

\[ + \frac{\left(\pi - p\right)}{\left(1 + \frac{Bx}{\alpha}\right)^{\alpha+1}} \left(1 + \frac{Bx}{\alpha}\right)^{-\alpha+1} \]

In the case \(\alpha \geq 1\) the term \(B\left(1 + \frac{Bx}{\alpha}\right)^{-\alpha+1}\) is increasing with a minimum of \(B\) at \(x = 0\). In the case \(\alpha < 1\), it is decreasing with a minimum of \(B^{\alpha}\) as \(x \to \infty\). In either
case, it is greater than 1. So the derivative is positive and the function is increasing as claimed.

If \( \alpha > 1 \), we will use the third expression for the derivative. For simplicity of notation and to facilitate our calculations, we will change the parameterization. Bounding the number of times \( f'(x)/f''(Bx) = AB \) is equivalent to bounding the number of times the function

\[
g(x) = \frac{q(1 + Bx)^s}{q(1 + x)^s} = c
\]

where \( c \) is a constant, \( q = (\pi - p)/(1 - \pi) < 1 \), \( B > 1 \), \( 0 < r = a(\alpha + 1)/(\alpha + 1) \), \( 0 < s = (\alpha - a)/(\alpha + 1) \), and \( s + r = \alpha \). To do this, we will show that \( g'(x) = 0 \) at most once. Note that \( g'(x) = 0 \) if and only if

\[
0 = B[qr(1 + Bx)^{-r-1} - s(1 + Bx)^{-s-1}][q(1 + x)^r + (1 + x)^{-r}] - [qr(1 + x)^r + (1 + x)^{-r}]q[r(1 + x)^{r-1} - s(1 + x)^{-s-1}]
\]

\[
= qr(Bx)^{-r-1}(B - 1) + qr(Bx)^{-r-1}(1 + x)^{-s} - (1 + Bx)^{-s} (1 + x)^{r-1)}
\]

\[
= q(Bx)^{-r-1}(1 + x)^{r-1} - (1 + Bx)^{-s} (1 + x)^{r-1})
\]

\[
= s(1 + Bx)^{-s-1}(1 + x)^{r-1} - (1 + Bx)^{-s} (1 + x)^{r-1})
\]

Placing the terms with a multiple of \( s \) on the other side and multiplying through by \( (1 + x)^{-r-1}(1 + Bx)^{s+1} \), we see that the last equation is zero if and only if

\[
\frac{s(B - 1)}{(1 + x)^{s+1}} + q(1 + x) \left( B - \frac{1 + Bx}{1 + x} \right)^{r+1} = 0 \tag{4}
\]

\[
q^2r(B - 1)(1 + Bx)^{r+s} + qr(1 + Bx) \left( B - \frac{1 + Bx}{1 + x} \right)^{r+s} - 1). \tag{5}
\]

We claim that the function \( f(x) \) on the left hand side of this equation is increasing and the function \( g(x) \) on the right hand side of this equation is decreasing, so that the intersection occurs at most once. Thus \( g'(x) = 0 \) at most once and \( f'(x)/f''(Bx) = AB \) at most twice for \( x > 0 \). Thus, \( f(x) - Af'(Bx) = (1 - A)(1 - p)^{-1/\alpha} \) for at most two \( x > 0 \) because that equation holds for \( x = 0 \).

To prove the claims, let

\[
h_1(x) = (1 + x) \left( B - \frac{1 + Bx}{1 + x} \right)^{r+1}
\]

so that

\[
h'_1(x) = B - \left( \frac{1 + Bx}{1 + x} \right)^{r+s} \left( \frac{1}{1 + x} \right)(1 + Bx + (r + s + 1)(B - 1)).
\]

Note that \( h'_1(0) = -(B - 1)(r + s) < 0 \) and \( h'_1(x) \) is decreasing because both \( \frac{1 + Bx}{1 + x} \) and \( \frac{1 + (r+s+1)(B-1)+Bx}{1+x} \) are increasing functions. Thus, \( h_1(x) \) and, consequently, function \( (4) \) are decreasing functions.

If \( \alpha = r + s \geq 1 \), then the right hand side function \( (5) \) is clearly increasing. If \( \alpha = r + s < 1 \), then let

\[
h_2(x) = (1 + Bx) \left( B - \frac{1 + Bx}{1 + x} \right)^{-s-1} \tag{6}
\]

so that

\[
h'_2(x) = B \left( -1 + \frac{1 + x}{1 + Bx} \right) \left( B - \frac{1 + (r + s)(B - 1)}{1 + x} \right)
\]

Because \( \frac{1 + x}{1 + Bx} \) is between 0 and 1 and because \( 1 - (r + s) \) is also between 0 and 1, \( (1 + x)^{1/(r + s)} \geq \frac{1 + x}{1 + Bx} \). Thus,

\[
h'_2(x) \geq B \left( -1 + \frac{1 + x}{1 + Bx} \right) \left( B - \frac{1 + (r + s)(B - 1)}{1 + x} \right) \geq 0.
\]

Thus, \( h_2(x) \) and, consequently, \( (5) \) are increasing functions.

\[\square\]

The F81 Exceptional Case

We will show that the F81 family of rate matrices are the only ones with only one nonzero eigenvalue because this was not pointed out in Wu and Susko (2010). We can set the one nonzero eigenvalue to any negative number and the argument will be the same, so assume that the nonzero eigenvalue is \(-1\). Let \( V \) be the matrix of eigenvectors with the last column being the eigenvector associated with the eigenvalue zero so that it is the transpose of \((1,1,1,1)^\top\). Then,

\[
Q = V \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}
\]

\[
V^{-1} = -I + \begin{pmatrix} 0 & 0 & 0 & 1/2 \\ 0 & 0 & 0 & 1/2 \\ 0 & 0 & 0 & 1/2 \\ 0 & 0 & 0 & 1/2 \end{pmatrix} V^{-1}.
\]

Thus, \( Q \) has the identical entries in a given column except for the diagonal entry, making the rate matrix a member of the F81 family. The Jukes–Cantor matrix is a special member of the F81 family and can be distinguished by considering the eigenvectors because, for Jukes–Cantor, the eigenspace of the nonzero eigenvalue is orthogonal to \((1/2, 1/2, 1/2, 1/2)^\top\). Other families can overlap with the F81 family and, for the specific cases in which they do so, they fall into the exceptional (non-generic) case.

Clearly, it is not possible to determine the branch lengths, shape parameter, and proportion of invariable sites for an F81 matrix using one three-taxon tree as done in Allman et al. (2008) for the model without invariable sites because there are at most four pieces of information (the changes between pairs of species and for the
entire three-taxon tree) and five parameters to solve for. Thus, at least four taxa are necessary to identify the rate parameters and branch lengths under the F81 rate matrix. For a four-taxon tree, we will have four three-taxon subtrees and thus four equations similar to Equation (8) (more specifically, the equation just above equation (8)) in Allman et al. (2008) with two unknown parameters, \( \alpha \) and \( \pi \) to determine.

We have no proof to date that there cannot be multiple solutions to these four equations for all parameter choices, although we hypothesize that there will not be, at least for generic branch lengths. Using Rogers’ trick of assuming that two distinct models give the same data, we do have a simple proof that two models that lie in one part of bivariate parameter space can be distinguished. Let \( \mu \) be associated with tree \( T’ = ((AB)(CD)) \) and rate distribution mixture parameters \( \alpha \) and \( \pi \). Let \( \nu \) be associated with an unknown tree \( T \) and parameters \( a \) and \( p \). Assume \( p \leq \pi \) and \( \alpha \leq a \). The first restriction, \( p \leq \pi \) simply places an ordering on the models. The line \( \alpha = a \) divides the ordered bivariate parameter space into two parts. If the models fall on one side of this line, the side where \( \alpha \leq a \), then we show below that they can be distinguished.

Let \( t_i \) be the interior branch length in \( T’ \) and \( t_A, t_B, t_C, t_D \) and \( t_D \) be the branch lengths to the respective tips in \( T’ \). Then applying Rogers’ trick to the four equations corresponding to Equation (8) or, more specifically, the equation appearing just above equation (8), in Allman et al. (2008), we have, for any F81 matrix with eigenvalue set to \(-1\) (and for the moment generating functions evaluated at the negative value):

\[
\n^{-1}(\mu(t_A + t_B)) + n^{-1}(\mu(t_A + t_C + t_s)) + n^{-1}(\mu(t_B + t_C + t_s)) = 2n^{-1}(\mu(t_A + t_B + t_C + t_s)) \\
\n^{-1}(\mu(t_A + t_B)) + n^{-1}(\mu(t_A + t_D + t_s)) + n^{-1}(\mu(t_B + t_D + t_s)) = 2n^{-1}(\mu(t_A + t_B + t_D + t_s)) \\
\n^{-1}(\mu(t_C + t_D)) + n^{-1}(\mu(t_A + t_C + t_s)) + n^{-1}(\mu(t_A + t_D + t_s)) = 2n^{-1}(\mu(t_A + t_C + t_D + t_s)) \\
\n^{-1}(\mu(t_C + t_D)) + n^{-1}(\mu(t_A + t_C + t_s)) + n^{-1}(\mu(t_A + t_D + t_s)) = 2n^{-1}(\mu(t_A + t_C + t_D + t_s))
\]

As before, writing these equations in terms of the function \( f(x) = [(\pi - p) + (1 - \pi)(1 + x/\alpha)^{-\alpha}]^{-1/\alpha} \) and letting \( s = t_C + t_s \) or \( s = t_D + t_s \), we see that the first two equations have the form:

\[
2f(t_A + t_B + s) - f(t_A + s) - f(t_B + s) = f(t_A + t_B) - f(0).
\]

When \( \left( \frac{1 - \pi}{\pi - p} \right) \left( \frac{\alpha - q}{\alpha + q} \right) \leq 1 \), \( f \) is concave. In this case, one can show that there is no solution \( s \) unless \( a = \alpha \) and \( \pi = p \). When \( \left( \frac{1 - \pi}{\pi - p} \right) \left( \frac{\alpha - q}{\alpha + q} \right) > 1 \), \( f \) has one inflection point and there appears to be at most one solution \( s \) implying that the parameters are identifiable for generic trees. However, we have been at a loss to prove this last point formally.

**Discretized Gamma Distribution**

Phylogenetic software often uses a discrete gamma distribution where the gamma distribution is broken into \( k \) intervals of equal probability and the mean of each interval is used to represent the interval. This changes the functional form of the moment generating function. Let \( -\lambda \) be a nonzero eigenvalue for the instantaneous rate matrix \( Q \) and \( \tau \) be the distance between a pair of species. The moment generating function evaluated at \(-\lambda \tau\) is then the function

\[
\mu(\lambda \tau) = \pi + \frac{1 - \pi}{k} \left( \sum_{j=1}^{k} e^{-r_j(\alpha)\lambda \tau} \right),
\]

where \( r_j(\alpha) \) is the mean of the gamma distribution, with shape parameter \( \alpha \) and mean 1, over the \( j \)th equally probable interval (out of \( k \) intervals in total). It is an open mathematical problem to determine whether or not the model with this discrete rate distribution is generically identifiable.