POINTS OF VIEW

Complex models of sequence evolution require accurate estimators as exemplified with the invariable site plus Gamma model

Lam-Tung Nguyen[1], Arndt von Haeseler[1,2] and Bui Quang Minh[1*]

[1] Center for Integrative Bioinformatics Vienna,
Max F. Perutz Laboratories, University of Vienna, Medical University Vienna
Vienna, Austria

[2] Bioinformatics and Computational Biology
Faculty of Computer Science
University of Vienna
Vienna, Austria

Corresponding author:
Bui Quang Minh
Center for Integrative Bioinformatics Vienna (CIBIV)
Max F. Perutz Laboratories
Campus Vienna Biocenter 5 (VBC5), Ebene 1, Room 1812.4
A-1030 Vienna, Austria
Tel: 0043 1 4277 24007
Fax: 0043 1 4277 24098
Email: minh.bui@univie.ac.at

Keywords: phylogenetic inference, maximum likelihood, rate heterogeneity among sites, invariable sites, Gamma model

Running head: Accurate estimator for I+Γ models

Abstract

The invariable site plus Γ model is widely used to model rate heterogeneity among alignment sites in maximum likelihood and Bayesian phylogenetic analyses. The proof that the invariable site plus continuous Γ model is identifiable (model parameters can be inferred correctly given enough data) has increased the creditability of its application to phylogeny reconstruction. However, most phylogenetic software implement the invariable site plus discrete Γ model, whose identifiability is likely but unproven. How well the parameters of the invariable site plus discrete Γ model are estimated is still disputed. Especially the dependency between the fraction of invariable sites and the fractions of sites with a slow evolutionary rate are discussed as problematic. We show that many phylogenetic software cannot reliably estimate the shape parameter, the proportion of invariable sites and the tree length due to the implemented optimization heuristics. We propose a new optimization heuristic to resolve this problem. Thus, the parameters of the invariable site plus discrete Γ model can be accurately estimated. Our results signify the importance of verifying and developing effective estimation methods for complex models of sequence evolution.

Introduction

In model based phylogenetic analysis, the invariable site plus Γ model (Yang 1994; Gu, et al. 1995), hereafter referred to as I+Γ, is widely used to model rate heterogeneity among sites, because it often fits the data better than the Γ model or the invariable-sites model alone (Sullivan and Swofford 1997). Thus, the I+Γ model is frequently selected by MODELTEST (Posada and Crandall 1998). The I+Γ model has two parameters: the proportion of invariable sites p_{inv} ($0 \leq p_{inv} < 1$) and the shape parameter α (>0) of the Γ distribution. A small α (<1) indicates strong rate heterogeneity, whereas a large α (>1) corresponds to weak rate heterogeneity. Under

58 certain conditions p_{inv} and α compete with each other for the same phylogenetic
59 signal. For example, $\alpha \leq 1$ already accounts for sites with low rates; that interferes
60 with p_{inv} and causes a correlation between the parameters making reliable estimation
61 of those parameters difficult (Sullivan, et al. 1999; Mayrose, et al. 2005; Yang 2006).
62 Despite this interference, it has been shown that the I+ continuous Γ model is
63 identifiable for "*all but members of the F81 family of rate matrices on any phylogeny*
64 *with more than two distinct interspecies distances*" (Rogers 2001; Allman and Rhodes
65 2008; Chai and Housworth 2011). Since the I+ continuous Γ model is identifiable,
66 reliable parameter estimation for this model should be possible for sufficiently long
67 multiple sequence alignments.
68 However, most phylogenetic software only implement the I+ discrete Γ (Yang
69 1994) model as an approximation to the continuous Γ model because of its
70 computational efficiency. The discussed competition between p_{inv} and α is based on
71 the analysis of the discrete Γ-distribution. The results have led to the suggestion to
72 discourage the use of the I+ discrete Γ model (Yang 2006; Jia, et al. 2014; Stamatakis
73 2014).
74 On the other hand, the identifiability of the I+ discrete Γ model is likely, but
75 unproven (Chai and Housworth 2011) and it is unclear how accurate popular
76 phylogenetic software estimate parameters of the I+ discrete Γ model.
77 Thus, we used simulations to assess the accuracy of the I+ discrete Γ estimators
78 implemented in three maximum likelihood (ML) phylogenetic software: RAxML
79 (Stamatakis 2014), PhyML (Guindon, et al. 2010), IQ-TREE (Nguyen, et al. 2015)
80 and one Bayesian inference program MrBayes (Ronquist, et al. 2012). More precisely,
81 we simulated 100,000-bp long alignments along three balanced trees of 6, 24 and 96
82 taxa. The lengths of the alignments ensure the recovery of the correct tree topology.

83 The three trees have uniform branch lengths of 0.1 substitutions per site except for
84 one internal branch on the 6-taxon tree whose length equals 0.2 to allow for 3 distinct
85 distances between the sequences as required for identifiability in the continuous case
86 (Chai and Housworth 2011). We assumed the K2P model (Kimura 1980) with a
87 transition/transversion ratio of 2.0 and the rate heterogeneity model I+ discrete Γ with
88 four rate categories. For each tree and each pair $(p_{inv}, \alpha) \in \{0.0,\ 0.1, \ldots, 0.9\} \times$
89 $\{0.1, 0.5, 1.0\}$ we simulated 100 alignments using Seq-Gen (Rambaut and Grassly
90 1997). We used RAxML version 8.2.2, PhyML version 20141029, IQ-TREE version
91 1.3.7 and MrBayes version 3.2.6 compiled with the BEAGLE library (Ayres, et al.
92 2012) to infer the invariable proportion, the shape parameter and the tree length from
93 the simulated alignments. For RAxML, PhyML and IQ-TREE we used the default
94 options.

95 For MrBayes we used the default priors, i.e., uniform distribution within interval
96 [0,1] for p_{inv}, exponential distribution with mean 1.0 for α, non-clocklike uniform
97 Dirichlet distribution for branch lengths and Γ distribution with mean of 10 for tree
98 lengths (Unconstrained:GammaDir(1.0,0.1,1.0,1.0)). The sequential version of
99 MrBayes was run with four chains (one hot and three cold chains) and one million
100 MCMC generations. 1,489 (16.5%) non-convergent MrBayes runs, where the
101 effective sample sizes (ESS) on p_{inv}, α, or tree lengths are smaller than 100, were
102 repeated with five million generations. However, 52 of the extended re-runs were
103 stopped after four weeks without completing all five million generations. We note that
104 207 of the 1,489 re-runs still did not converge. MrBayes estimates are then
105 summarized as the mean of the posterior distribution with a burn-in of 25%.

106 Current phylogenetic programs do not produce accurate estimates
107 for the I+ discrete Γ model

108 Figure 1 displays the averages $\bar{\hat{\alpha}}$ of the estimated shape parameter α, the averages
109 $\overline{\widehat{p_{inv}}}$ of the estimated invariable fraction p_{inv} and the average $\bar{\hat{l}}$ of the estimated tree
110 length l produced by PhyML, RAxML, IQ-TREE and MrBayes for the 100
111 alignments simulated from each parameter combinations. A program is called
112 *accurate* if the estimated averages $\bar{\hat{\alpha}}$, $\overline{\widehat{p_{inv}}}$, $\bar{\hat{l}}$ deviate no more than 10% from the true
113 values (depicted in green in Figure 1).
114 None of the tested programs estimated all parameter combinations accurately. The
115 problem is especially pronounced for the 6-taxon alignments. For extreme rate
116 heterogeneity ($\alpha = 0.1$) MrBayes and PhyML recovered the true α, p_{inv} and l for
117 9/10 and 5/10 parameter combinations respectively, whereas the average estimates
118 from IQ-TREE and RAxML were inaccurate. For strong rate heterogeneity ($\alpha = 0.5$),
119 the degrees of inaccuracy observed among all programs differ unsystematically. On
120 the one hand, IQ-TREE and MrBayes accurately estimated the parameters in four and
121 six settings. On the other hand, RAxML and PhyML could not estimate accurately the
122 three parameters for any of the ten parameter-combinations. For medium rate
123 variation ($\alpha = 1.0$) only IQ-TREE produced the accurate estimates for all settings.
124 All other programs exhibited varying degrees of inaccuracy.
125 For the 24- and 96-taxon alignments we observed an increase in the number of
126 accurate estimates for all programs. These results corroborate a previous study
127 (Sullivan et al. 1999) showing that increased taxon sampling leads to more reliable
128 estimates. However, under extreme rate heterogeneity ($\alpha = 0.1$), only MrBayes
129 estimated all parameter sets accurately.
130 We also observed that inaccurate estimates of α and p_{inv} could sometimes lead to
131 tree lengths that substantially deviate from the simulated lengths. For instance, for the
132 96-taxon alignments simulated with $\alpha = 0.1$ and $p_{inv} = 0.8$ (expected tree length =

133 18.9) IQ-TREE estimated an average tree length of 177.0 that is nine times longer
134 than the simulated tree length. The other programs also sometimes produced tree
135 lengths that were considerably longer than the simulated ones.
136 In terms of computing times PhyML, RAxML and IQ-TREE needed for all
137 analyses 62,441, 12,563 and 7,675 CPU hours, respectively. MrBayes needed 740,681
138 CPU hours to complete one million MCMC generations, thus it is 96.5 times slower
139 than the fastest ML program. We note that this is only a lower-bound for the effective
140 time one needs to wait for MrBayes results because 16.5% of MrBayes runs did not
141 converge after one million MCMC generations. These runs were repeated with five
142 times more generations, that led to significantly more computations.

143 Multiple local optima on the likelihood surface cause inaccurate 144 estimates

145 Because the tested programs performed quite differently with respect to the accuracy
146 of parameter estimation, the number of taxa cannot be the only explanation. We
147 suspected that the optimization heuristics as implemented in the programs drive the
148 accuracy. Examining the likelihood surfaces for many simulated alignments revealed
149 a common feature that the parameter space has two distinct peaks of high log-
150 likelihoods (Figure 2): one global close to the true parameters and one suboptimal
151 peak with slightly lower log-likelihood ($\Delta LnL = -27$ in this example) separated by a
152 flat valley from the true parameters. In this particular instance MrBayes and PhyML
153 found the true parameters whereas RAxML and IQ-TREE were trapped in the local
154 maximum (not necessarily the case for other instances). In fact whether the global or
155 local optimum is detected depends on the starting values of the numerical
156 optimization routines.

157 To summarize, we compared for each simulated alignment the log-likelihoods of the
158 estimates with the log-likelihoods obtained for the true parameters. Table 1 show how
159 often the true parameter combination produced a higher likelihood than the inferred
160 parameters from MrBayes, PhyML, RAxML and IQ-TREE. These fractions are
161 particularly high for the 6-taxon tree and for the ML inference programs. Most ML
162 phylogenetic programs use general-purpose numerical methods to find substitution
163 model parameters, α, p_{inv} and branch lengths (e.g. BFGS (Fletcher 1987), Brent
164 (Brent 1973) and Newton-Raphson (Press 2007) method). These methods are
165 obviously not well adapted to the complex likelihood surface of phylogenetic
166 inferences and explain the poor overall performance of the ML programs (Figure 1).
167

Program	6-taxon tree	24-taxon tree	96-taxon tree
MrBayes	34.7 %	6.6 %	5.6 %
PhyML	60.2 %	21.9 %	45 %
RAxML	89.2 %	37.7 %	49.5 %
IQ-TREE	36 %	34.1 %	44 %

168 **Table 1.** Percentage of alignments where the true simulation parameters result in
169 higher log-likelihoods than the inferred parameters from the programs for three
170 simulation scenarios (6-, 24- and 96-taxon trees).

171 Effective optimization heuristic produces accurate estimates

172 As remedy, we propose an alternative optimization heuristic which employs the
173 EM algorithm (Dempster, et al. 1977) to estimate p_{inv}. The remaining parameters are
174 estimated using general-purpose numerical methods. We assume a discrete Γ
175 distribution with k rate categories. Under the I+ discrete Γ model, the site rates follow
176 a discrete mixture model consisting of $k + 1$ categories with rates $r_0, .., r_k$, where

$r_0 = 0$ represents invariable sites and $r_i > 0$ $(i = 1, .., k)$ are the k rates determined from the shape parameter α of the discrete Γ distribution (Yang 1994). Given a tree topology, the optimization heuristic does the following:

1. Choose initial values for α and p_{inv}.
2. Optimize branch lengths by the Newton-Raphson method.
3. Optimize substitution model parameters by the BFGS method.
4. For each alignment site D_i compute its posterior probability of being invariable ($1 \leq i \leq n$, where n is the number of alignment sites):

$$P(r_0|D_i) = \frac{P(D_i|r_0)w_0}{\sum_{j=0}^{k} P(D_i|r_j)w_j}$$

   where $P(D_i|r_j)$ is the likelihood of site D_i having rate r_j and $w_0 = p_{inv}$, $w_j = \frac{1-p_{inv}}{k}$ $(1 \leq j \leq k)$.
5. Update $p_{inv} = \frac{1}{n}\sum_{i=1}^{n} P(r_0|D_i)$.
6. Optimize α by the Brent method.
7. If the log-likelihood improvement is greater than a predefined ϵ value, go back to step 2. Otherwise, stop the parameter optimization.

Step 4 and 5 correspond to the E- and M-step of the EM algorithm respectively. To avoid being stuck in local optima, we repeat this optimization procedure from ten starting values of p_{inv} evenly spaced between 0 and the fraction of constant sites observed in the alignment. The initial value of α is always set to 1.0.

We implemented the new optimization heuristic in IQ-TREE now called IQ-TREE-EM (IQ-TREE version 1.4.3) and repeated the previous simulations. Figure 3 shows that IQ-TREE-EM successfully recovered the true parameters for all but one parameter combination (6-taxon, $\alpha = 0.5$ and $p_{inv} = 0.0$) where the average estimates ($\bar{\hat{\alpha}} = 0.59$ and $\overline{\widehat{p_{inv}}} = 0.06$) slightly deviated from the true values.

Also the percentage of instances where the estimated log-likelihoods were lower than the log-likelihood for the true parameters dropped considerably (0.06% 6-taxon tree, 0.0% 24-taxon tree, 0.03% 96-taxon tree; compare also with Table 1).

This increase in accuracy comes at the cost of an increased total computing time by a factor of 1.3 compared to IQ-TREE.

Thus, we conclude that the inaccurate parameter estimation of the I+ discrete Γ shown for the tested phylogenetic programs is caused by ineffective optimization methods.

Impact on real data

TreeBase ID #seqs x #sites		PhyML	RAxML	IQ-TREE	IQ-TREE-EM	Variance
DNA M10243	$\hat{\alpha}$	0.919	1.338	0.975	0.610	**0.0893**
203 x 1771	$\widehat{p_{inv}}$	0.145	0.232	0.163	0	0.0096
	$\hat{l}$	1.709	1.7025	1.705	1.696	0
DNA M1838	$\hat{\alpha}$	0.653	0.644	0.635	0.635	0.0001
228 x 1131	$\widehat{p_{inv}}$	0.319	0.324	0.319	0.319	0
	$\hat{l}$	31.056	33.685	34.208	34.231	**2.2906**
DNA M4326	$\hat{\alpha}$	0.749	0.834	0.651	0.602	**0.0107**
227 x 4055	$\widehat{p_{inv}}$	0.516	0.528	0.487	0.470	0.0007
	$\hat{l}$	1.669	1.656	1.663	1.653	0.0001
DNA M4794	$\hat{\alpha}$	0.392	0.396	0.385	0.387	0
204 x 12113	$\widehat{p_{inv}}$	0.329	0.334	0.330	0.331	0
	$\hat{l}$	66.148	68.993	71.191	71.722	**6.4295**
DNA M7024	$\hat{\alpha}$	0.638	0.635	0.632	0.630	0
767 x 5814	$\widehat{p_{inv}}$	0.375	0.385	0.378	0.372	0
	$\hat{l}$	63.914	64.766	66.390	65.782	**1.1988**
DNA M8984	$\hat{\alpha}$	1.002	1.207	1.066	0.961	**0.0116**
201 x 3931	$\widehat{p_{inv}}$	0.245	0.285	0.263	0.231	0.0006
	$\hat{l}$	0.996	0.997	0.995	0.998	0
DNA M9915	$\hat{\alpha}$	1.106	1.083	1.070	1.073	0.0003
504 x 2757	$\widehat{p_{inv}}$	0	0.0004	0.0004	0.0004	0
	$\hat{l}$	139.283	141.211	143.153	142.477	**2.8944**
Prot M11335	$\hat{\alpha}$	1.054	1.026	1.039	1.038	0.0001
99 x 696	$\widehat{p_{inv}}$	0.049	0.041	0.05	0.050	0
	$\hat{l}$	32.307	34.615	34.544	34.606	**1.3019**
Prot M11338	$\hat{\alpha}$	1.119	1.111	1.077	1.109	0.0003
100 x 567	$\widehat{p_{inv}}$	0.018	0.010	0.018	0.019	0
	$\hat{l}$	44.454	46.138	47.428	46.755	**1.6223**

Prot M4539	$\hat{\alpha}$	0.834	0.615	0.831	0.832	**0.0118**
59 x 12428	$\widehat{p_{inv}}$	0.135	0.0001	0.135	0.135	0.0046
	$\hat{l}$	14.780	14.917	14.853	14.847	0.0032
Prot M4884	$\hat{\alpha}$	0.742	0.493	0.740	0.742	**0.0155**
50 x 11827	$\widehat{p_{inv}}$	0.183	0.0001	0.182	0.183	0.0084
	$\hat{l}$	6.770	6.829	6.783	6.784	0.0007
Prot M5379	$\hat{\alpha}$	0.922	0.6334	0.922	0.922	**0.0209**
60 x 7776	$\widehat{p_{inv}}$	0.154	0.0001	0.154	0.154	0.0059
	$\hat{l}$	11.215	11.376	11.248	11.249	0.0051
Prot M7078	$\hat{\alpha}$	0.807	0.508	0.807	0.807	**0.0223**
77 x 12457	$\widehat{p_{inv}}$	0.212	0.0001	0.211	0.211	0.0112
	$\hat{l}$	9.522	9.812	9.516	9.516	0.0217
Prot M7729	$\hat{\alpha}$	0.751	0.534	0.751	0.752	**0.0118**
62 x 2973	$\widehat{p_{inv}}$	0.189	0.021	0.190	0.191	0.0071
	$\hat{l}$	17.168	18.964	17.425	17.437	0.6716
Prot M8175	$\hat{\alpha}$	1.098	0.830	1.091	1.093	**0.0174**
194 x 665	$\widehat{p_{inv}}$	0.077	0.0001	0.077	0.078	0.0015
	$\hat{l}$	11.839	11.806	11.945	11.947	0.0052

Table 2. TreeBase alignments that showed high variance (highlighted in bold-face in the last column) in the estimated α (>0.01) or tree lengths (>1.0) across four ML programs.

To investigate the impact of accuracy on real data for ML estimates, we analyzed 70 DNA and 45 protein TreeBase alignments (see Nguyen, et al. 2015 for alignment selection criteria). We applied the GTR+I+Γ4 and LG+I+Γ4 models for DNA and protein data, respectively. Among 115 alignments, we detected 15 alignments where the estimated model parameters differ widely as indicated by the high variance of the estimated parameters for the four ML programs (Table 2). For alignment M10243 all programs estimated different α and p_{inv}. Moreover, for protein alignments RAxML tends to underestimate α and p_{inv} (sometimes close to zero) relative to PhyML, IQ-TREE and IQ-TREE-EM. Since we do not know the truth in such instances, it is advisable to compare the estimates from many different ML programs, before one comes to conclusions about the evolution of the organism of interest.

Discussion

225 Our simulations revealed a major issue for parameter estimation of the I+ discrete
226 Γ model as implemented in phylogenetic software. Despite using very long
227 alignments, none of the tested programs recovered the true α, p_{inv} and tree length for
228 all parameter combinations. Often, the estimates deviated heavily from the true values
229 and different programs estimated different values for the same evolutionary
230 parameters, although all programs inferred the true tree. Our further analysis of 115
231 TreeBase alignments identified 15 alignments where PhyML, RAxML, IQ-TREE and
232 IQ-TREE-EM also showed certain discrepancies. Thus, this problem occurred in 10%
233 of the real data. Because we cannot predict from the alignments, when the estimates
234 may differ, the only way to be confident to get accurate estimate is to run many
235 different programs and compare the results.
236 We showed that the estimation heuristics implemented in popular phylogenetic
237 programs causes such inaccurate estimates and not the correlation between a small α
238 and the proportion of invariable sites. The relatively good performance of MrBayes is
239 likely attributed to the Bayesian sampling of the parameter space but comes at the
240 cost of excessive computing time.
241 With IQ-TREE-EM we provided an alternative optimization heuristic for ML
242 methods that allows accurate estimation of the parameters for the I+ discrete Γ model.
243 IQ-TREE-EM combines two optimization techniques: the multiple starting point
244 strategy and the EM algorithm. While the former allows to escape local optima, the
245 latter helps to speedup the optimization using analytical formula for p_{inv}. This new
246 approach effectively infers the true evolutionary parameters for long alignments.
247 Thus, it is tempting to speculate that the GTR+I+ discrete Γ model is also identifiable
248 as shown for the GTR+I+ continuous Γ model (Chai and Housworth 2011).

Our observations show that as models of sequence evolution get more and more complex (e.g., Dirichlet rate and other mixture models) more sophisticated numerical optimization are necessary to achieve accurate estimates of evolutionary parameters. It is not enough to recover the true tree, if one wants to understand how evolutionary forces shaped contemporary genomes. The effect of wrong parameter estimates for the substitution model on the total tree length is sometimes dramatic (see Figure 1). This may in turn bias downstream analysis such as divergence time dating, inference of site-specific evolutionary rates and ancestral sequence reconstruction, which are sensitive to the parameter estimates. Thus, one should critically scrutinize the heuristics implemented in popular programs. A more thorough evaluation of phylogenetic inference programs allowing for very complicated models of sequence evolution is necessary, but beyond the scope of this article.

Finally, we would like to point out that we only addressed the accurate computation of p_{inv} and α for the widely-used I+ discrete Γ model. We do not discuss the biological interpretation of p_{inv}. The estimate of p_{inv} depends very much on the multiple sequence alignment at hand. p_{inv} may change if we enlarge the alignment. Thus, drawing an absolute conclusion from p_{inv} is in any case questionable.

Acknowledgements

The authors would like to thank Heiko Schmidt for fruitful discussions, two anonymous reviewers, Fredrik Ronquist and Edward Susko for constructive and helpful comments on a earlier version of the manuscript. The computational results presented have been achieved using the Vienna Scientific Cluster 3 (VSC-3). This work was supported by the Austrian Science Fund – FWF (grant no. I-2805-B29 and I-1824-B22).

References

Allman ES, Rhodes JA. 2008. Identifying evolutionary trees and substitution parameters for the general Markov model with invariable sites. Math. Biosci. 211:18-33.
Ayres DL, Darling A, Zwickl DJ, Beerli P, Holder MT, Lewis PO, Huelsenbeck JP, Ronquist F, Swofford DL, Cummings MP, et al. 2012. BEAGLE: an application programming interface and high-performance computing library for statistical phylogenetics. Syst. Biol. 61:170-3.
Brent RP. 1973. Algorithms for Minimization Without Derivatives: Courier Corporation.
Chai J, Housworth EA. 2011. On Rogers' proof of identifiability for the GTR + Gamma + I model. Syst. Biol. 60:713-8.
Dempster AP, Laird NM, Rubin DB. 1977. Maximum likelihood from incomplete data via the EM algorithm. J. Roy. Stat. Soc. B:1-38.
Fletcher R. 1987. Practical methods of optimization. New York: John Wiley and Sons.
Gu X, Fu YX, Li WH. 1995. Maximum-Likelihood-Estimation of the Heterogeneity of Substitution Rate among Nucleotide Sites. Mol. Biol. Evol. 12:546-57.
Guindon S, Dufayard JF, Lefort V, Anisimova M, Hordijk W, Gascuel O. 2010. New algorithms and methods to estimate maximum-likelihood phylogenies: Assessing the performance of PhyML 3.0. Syst. Biol. 59:307-21.
Jia FZ, Lo N, Ho SYW. 2014. The Impact of Modelling Rate Heterogeneity among Sites on Phylogenetic Estimates of Intraspecific Evolutionary Rates and Timescales. Plos One 9.
Kimura M. 1980. A Simple Method for Estimating Evolutionary Rates of Base Substitutions through Comparative Studies of Nucleotide-Sequences. J. Mol. Evol. 16:111-20.
Mayrose I, Friedman N, Pupko T. 2005. A Gamma mixture model better accounts for among site rate heterogeneity. Bioinformatics 21:ii151-ii8.
Nguyen LT, Schmidt HA, von Haeseler A, Minh BQ. 2015. IQ-TREE: A Fast and Effective Stochastic Algorithm for Estimating Maximum-Likelihood Phylogenies. Mol. Biol. Evol. 32:268-74.
Posada D, Crandall KA. 1998. MODELTEST: testing the model of DNA substitution. Bioinformatics 14:817-8.
Press WH. 2007. Numerical recipes 3rd edition: The art of scientific computing: Cambridge University Press.
Rambaut A, Grassly NC. 1997. Seq-Gen: an application for the Monte Carlo simulation of DNA sequence evolution along phylogenetic trees. Comput. Appl. Biosci. 13:235-8.
Rogers JS. 2001. Maximum likelihood estimation of phylogenetic trees is consistent when substitution rates vary according to the invariable sites plus gamma distribution. Syst. Biol. 50:713-22.
Ronquist F, Teslenko M, van der Mark P, Ayres DL, Darling A, Hohna S, Larget B, Liu L, Suchard MA, Huelsenbeck JP. 2012. MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. Syst. Biol. 61:539-42.
Stamatakis A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 30:1312-3.
Sullivan J, Swofford DL. 1997. Are guinea pigs rodents? The importance of adequate models in molecular phylogenetics. J. Mammal. Evol. 4:77-86.

323 Sullivan J, Swofford DL, Naylor GJP. 1999. The effect of taxon sampling on
324 estimating rate heterogeneity parameters of maximum-likelihood models. Mol. Biol.
325 Evol. 16:1347-56.
326 Yang Z. 2006. Computational Molecular Evolution. New York: Oxford University
327 Press.
328 Yang Z. 1994. Maximum likelihood phylogenetic estimation from DNA sequences
329 with variable rates over sites: approximate methods. J. Mol. Evol. 39:306-14.
330

Figure legends

Figure 1. The averages $\bar{\hat{\alpha}}$ of the estimated shape parameter α, the averages $\overline{\widehat{p_{inv}}}$ of the estimated invariable fraction p_{inv} and the average $\bar{\hat{l}}$ of the estimated tree length l produced by PhyML, RAxML, IQ-TREE and MrBayes for the 100 alignments simulated from each parameter combinations. The averages are colored according to their differences from the true values: red (more than 25% deviation), yellow (10% to 25% deviation) and green (less than 10% deviation). For $p_{inv} = 0.0$ the estimated $\widehat{p_{inv}}$ is green if $0 \leq \widehat{p_{inv}} \leq 0.01$, yellow if $0.01 \leq \widehat{p_{inv}} \leq 0.05$ and red if $0.05 < \widehat{p_{inv}}$.

Figure 2. The likelihood surface for one simulated alignment as a function of α and p_{inv}.

Figure 3. The averages $\bar{\hat{\alpha}}$ of the estimated shape parameter α, the averages $\overline{\widehat{p_{inv}}}$ of the estimated invariable fraction p_{inv} and the average $\bar{\hat{l}}$ of the estimated tree length l produced by IQ-TREE-EM for the 100 alignments simulated from each parameter combinations. The color code is explained in Figure 1.

Figure 1

PhyML

α \ p_{inv}	0.00	0.10	0.20	0.30	0.40	0.50	0.60	0.70	0.80	0.90
	6-taxon tree (tree length = 1.0)									
0.10	0.00	0.10	0.20	0.25	0.39	0.50	0.61	0.89	0.90	0.90
	0.10	0.10	0.10	0.06	0.09	0.10	0.11	0.56	0.19	0.11
	1.00	1.00	1.00	0.92	0.98	0.99	1.17	1.71	3.68	0.76
0.50	0.19	0.18	0.17	0.17	0.25	0.36	0.43	0.53	0.68	0.69
	0.84	0.62	0.47	0.35	0.33	0.32	0.26	0.23	0.22	0.04
	1.00	1.00	1.00	0.99	0.98	0.98	0.95	0.91	0.91	18.58
1.00	0.11	0.20	0.20	0.19	0.17	0.15	0.17	0.29	0.45	0.77
	1.50	1.47	0.99	0.67	0.45	0.30	0.21	0.17	0.13	0.52
	1.00	1.00	1.00	1.00	1.00	0.98	0.94	0.90	0.80	4.54
	24-taxon tree (tree length = 4.5)									
0.10	0.41	0.28	0.35	0.30	0.43	0.51	0.69	0.70	0.80	0.93
	0.55	0.30	0.26	0.10	0.13	0.11	0.18	0.10	0.10	0.17
	4.93	4.93	5.14	4.48	4.79	5.93	26.92	4.43	4.53	10.76
0.50	0.01	0.10	0.20	0.30	0.40	0.50	0.60	0.70	0.80	0.90
	0.51	0.50	0.51	0.50	0.50	0.50	0.50	0.50	0.50	0.50
	4.50	4.50	4.50	4.50	4.49	4.51	4.50	4.52	4.49	4.45
1.00	0.00	0.10	0.20	0.30	0.40	0.50	0.60	0.70	0.80	0.89
	1.01	1.01	1.00	1.00	0.99	0.99	1.00	0.99	1.00	0.87
	4.50	4.50	4.49	4.50	4.50	4.51	4.50	4.52	4.54	136.95
	96-taxon tree (tree length = 18.9)									
0.10	0.00	0.10	0.49	0.55	0.61	0.50	0.62	0.70	0.90	0.92
	0.10	0.10	0.28	0.28	0.28	0.10	0.13	0.10	0.45	0.17
	18.83	18.91	13.90	14.00	14.06	18.69	17.95	18.27	35.95	38.45
0.50	0.00	0.10	0.20	0.30	0.40	0.50	0.60	0.70	0.80	0.90
	0.50	0.50	0.50	0.50	0.50	0.50	0.50	0.50	0.50	0.50
	18.88	18.88	18.88	18.82	18.90	18.88	18.94	18.77	18.89	18.29
1.00	0.00	0.10	0.20	0.30	0.40	0.50	0.60	0.70	0.80	0.90
	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.03
	18.90	18.89	18.89	18.87	18.94	18.93	18.99	18.77	18.92	18.49

RAxML

α \ p_{inv}	0.00	0.10	0.20	0.30	0.40	0.50	0.60	0.70	0.80	0.90	
	6-taxon tree (tree length = 1.0)										
0.10	0.71	0.74	0.77	0.79	0.82	0.84	0.86	0.89	0.90	0.92	$\widehat{p_{inv}}$
	4.63	3.85	3.08	2.36	1.69	1.15	0.67	0.56	0.24	0.22	$\hat{\alpha}$
	1.03	1.04	1.07	1.10	1.17	1.33	1.78	1.74	4.39	1.38	$\hat{l}$
0.50	0.30	0.34	0.40	0.45	0.51	0.57	0.64	0.72	0.81	0.91	$\widehat{p_{inv}}$
	1.31	1.17	1.05	0.95	0.83	0.75	0.68	0.63	0.63	0.73	$\hat{\alpha}$
	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.99	$\hat{l}$
1.00	0.16	0.22	0.29	0.37	0.45	0.53	0.61	0.71	0.80	0.90	$\widehat{p_{inv}}$
	1.89	1.74	1.59	1.45	1.35	1.25	1.17	1.14	1.10	1.05	$\hat{\alpha}$
	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.99	$\hat{l}$
	24-taxon tree (tree length = 4.5)										
0.10	0.64	0.67	0.70	0.73	0.76	0.73	0.77	0.82	0.88	0.94	$\widehat{p_{inv}}$
	0.82	0.72	0.64	0.58	0.53	0.20	0.17	0.18	0.24	0.29	$\hat{\alpha}$
	5.18	5.87	6.74	7.69	8.91	16.64	20.97	28.88	8.72	3.38	$\hat{l}$
0.50	0.01	0.11	0.21	0.30	0.40	0.50	0.60	0.70	0.80	0.90	$\widehat{p_{inv}}$
	0.51	0.51	0.51	0.51	0.51	0.51	0.50	0.50	0.50	0.51	$\hat{\alpha}$
	4.50	4.50	4.50	4.50	4.50	4.50	4.50	4.50	4.50	4.51	$\hat{l}$
1.00	0.01	0.10	0.20	0.30	0.40	0.50	0.60	0.70	0.80	0.90	$\widehat{p_{inv}}$
	1.02	1.02	1.01	1.01	1.00	1.00	1.00	1.00	1.00	1.00	$\hat{\alpha}$
	4.50	4.50	4.50	4.50	4.50	4.50	4.50	4.51	4.50	4.52	$\hat{l}$
	96-taxon tree (tree length = 18.9)										
0.10	0.37	0.43	0.49	0.65	0.67	0.68	0.74	0.81	0.87	0.94	$\widehat{p_{inv}}$
	0.27	0.27	0.28	0.62	0.49	0.28	0.26	0.27	0.27	0.28	$\hat{\alpha}$
	13.85	13.88	13.94	13.43	13.79	14.08	14.55	14.32	14.20	11.90	$\hat{l}$
0.50	0.00	0.10	0.20	0.30	0.40	0.50	0.60	0.70	0.80	0.90	$\widehat{p_{inv}}$
	0.50	0.50	0.50	0.50	0.50	0.50	0.50	0.50	0.50	0.50	$\hat{\alpha}$
	18.89	18.90	18.91	18.89	18.89	18.89	18.89	18.92	18.91	18.88	$\hat{l}$
1.00	0.00	0.10	0.20	0.30	0.40	0.50	0.60	0.70	0.80	0.90	$\widehat{p_{inv}}$
	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	$\hat{\alpha}$
	18.90	18.89	18.90	18.89	18.90	18.90	18.90	18.89	18.90	18.92	$\hat{l}$

IQ-TREE

α \ p_{inv}	0.00	0.10	0.20	0.30	0.40	0.50	0.60	0.70	0.80	0.90
	6-taxon tree (tree length = 1.0)									
0.10	0.71	0.74	0.77	0.79	0.82	0.84	0.86	0.89	0.90	0.92
	4.80	4.02	3.27	2.51	1.80	1.15	0.64	0.55	0.28	0.17
	1.03	1.04	1.06	1.10	1.17	1.32	1.76	1.73	3.47	1.16
0.50	0.18	0.23	0.30	0.36	0.42	0.51	0.61	0.70	0.80	0.91
	0.82	0.74	0.68	0.62	0.56	0.53	0.52	0.52	0.53	0.68
	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.99	0.87
1.00	0.01	0.10	0.20	0.30	0.40	0.50	0.60	0.70	0.80	0.90
	1.02	1.01	1.01	1.00	1.01	1.00	1.02	1.04	1.06	1.03
	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.99	0.99	0.98
	24-taxon tree (tree length = 4.5)									
0.10	0.64	0.67	0.70	0.73	0.76	0.79	0.76	0.81	0.87	0.95
	0.82	0.72	0.64	0.58	0.52	0.29	0.14	0.14	0.14	0.30
	5.17	5.86	6.73	7.68	8.89	62.72	41.99	40.81	41.04	43.68
0.50	0.00	0.10	0.20	0.30	0.40	0.50	0.60	0.70	0.80	0.90
	0.50	0.50	0.50	0.50	0.50	0.50	0.50	0.50	0.50	0.50
	4.50	4.50	4.50	4.50	4.50	4.50	4.50	4.50	4.51	4.53
1.00	0.00	0.10	0.20	0.30	0.40	0.50	0.60	0.70	0.80	0.90
	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	4.50	4.50	4.50	4.50	4.50	4.50	4.50	4.51	4.51	4.59
	96-taxon tree (tree length = 18.9)									
0.10	0.36	0.43	0.59	0.65	0.70	0.75	0.79	0.80	0.87	0.93
	0.27	0.27	0.57	0.62	0.63	0.64	0.63	0.28	0.17	0.16
	13.83	13.87	13.26	13.48	13.69	14.00	14.61	19.03	177.99	74.08
0.50	0.00	0.10	0.20	0.30	0.40	0.50	0.60	0.70	0.80	0.90
	0.50	0.50	0.50	0.50	0.50	0.50	0.50	0.50	0.50	0.50
	18.89	18.90	18.90	18.88	18.92	18.95	18.99	19.09	19.23	19.57
1.00	0.00	0.10	0.20	0.30	0.40	0.50	0.60	0.70	0.80	0.90
	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.99
	18.90	18.89	18.90	18.90	18.92	18.94	18.96	18.99	19.13	19.56

MrBayes

α \ p_{inv}	0.00	0.10	0.20	0.30	0.40	0.50	0.60	0.70	0.80	0.90	
	6-taxon tree (tree length = 1.0)										
0.10	0.01	0.10	0.20	0.30	0.40	0.50	0.60	0.70	0.80	0.90	$\widehat{p_{inv}}$
	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.07	$\hat{\alpha}$
	1.00	1.00	1.00	1.00	1.00	1.00	1.01	1.01	1.02	7.00	$\hat{l}$
0.50	0.10	0.14	0.20	0.28	0.38	0.49	0.60	0.70	0.66	0.74	$\widehat{p_{inv}}$
	0.65	0.57	0.52	0.50	0.48	0.49	0.50	0.50	0.26	0.11	$\hat{\alpha}$
	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.84	0.51	$\hat{l}$
1.00	0.04	0.09	0.18	0.28	0.27	0.29	0.24	0.32	0.46	0.68	$\widehat{p_{inv}}$
	1.14	1.01	0.96	0.96	0.75	0.62	0.32	0.21	0.13	0.10	$\hat{\alpha}$
	1.00	1.00	1.00	1.00	1.00	0.99	0.96	0.91	0.80	0.63	$\hat{l}$
	24-taxon tree (tree length = 4.5)										
0.10	0.00	0.10	0.20	0.30	0.40	0.50	0.60	0.70	0.80	0.90	$\widehat{p_{inv}}$
	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.10	$\hat{\alpha}$
	4.48	4.50	4.50	4.51	4.51	4.51	4.51	4.52	4.56	4.70	$\hat{l}$
0.50	0.00	0.10	0.20	0.30	0.40	0.50	0.60	0.70	0.80	0.87	$\widehat{p_{inv}}$
	0.51	0.50	0.50	0.50	0.50	0.50	0.50	0.50	0.50	0.25	$\hat{\alpha}$
	4.50	4.50	4.50	4.50	4.50	4.50	4.50	4.50	4.51	3.69	$\hat{l}$
1.00	0.00	0.10	0.20	0.30	0.40	0.50	0.60	0.70	0.80	0.66	$\widehat{p_{inv}}$
	1.01	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.10	$\hat{\alpha}$
	4.50	4.50	4.50	4.50	4.50	4.50	4.51	4.51	4.51	2.45	$\hat{l}$
	96-taxon tree (tree length = 18.9)										
0.10	0.00	0.10	0.20	0.30	0.40	0.50	0.60	0.70	0.80	0.90	$\widehat{p_{inv}}$
	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.10	$\hat{\alpha}$
	18.85	18.91	18.92	18.92	18.90	18.93	18.99	19.01	18.97	18.93	$\hat{l}$
0.50	0.00	0.10	0.20	0.30	0.40	0.50	0.60	0.70	0.76	0.88	$\widehat{p_{inv}}$
	0.50	0.50	0.50	0.50	0.50	0.50	0.50	0.50	0.36	0.37	$\hat{\alpha}$
	18.88	18.90	18.91	18.90	18.90	18.90	18.90	18.93	11.91	13.89	$\hat{l}$
1.00	0.00	0.10	0.20	0.30	0.40	0.50	0.60	0.70	0.80	0.89	$\widehat{p_{inv}}$
	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.83	0.62	$\hat{\alpha}$
	18.90	18.89	18.90	18.89	18.91	18.91	18.91	18.90	16.95	12.44	$\hat{l}$

Figure 2

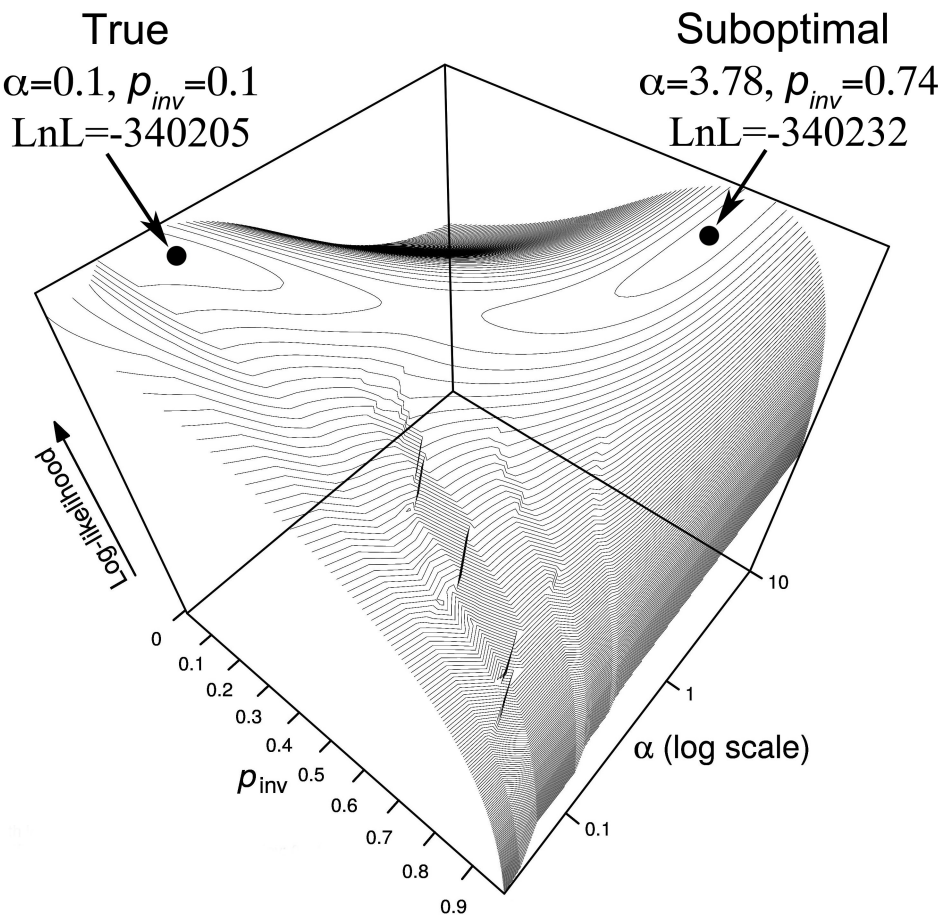
True
α=0.1, p_{inv}=0.1
LnL=-340205
Suboptimal
α=3.78, p_{inv}=0.74
LnL=-340232
Log-likelihood
0
0.1
0.2
0.3
0.4
p_{inv}
0.5
0.6
0.7
0.8
0.9
10
1
0.1
α (log scale)

Figure 3

IQ-TREE-EM											
α \ p_{inv}	0.00	0.10	0.20	0.30	0.40	0.50	0.60	0.70	0.80	0.90	
	6-taxon tree (tree length = 1.0)										
0.10	0.00	0.10	0.20	0.30	0.40	0.50	0.60	0.70	0.80	0.90	$\overline{\widehat{p_{inv}}}$
	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.10	$\overline{\hat{\alpha}}$
	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.01	0.93	$\overline{\hat{l}}$
0.50	0.06	0.10	0.19	0.27	0.38	0.48	0.59	0.70	0.80	0.90	$\overline{\widehat{p_{inv}}}$
	0.59	0.52	0.50	0.48	0.49	0.48	0.48	0.52	0.52	0.54	$\overline{\hat{\alpha}}$
	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.99	0.98	$\overline{\hat{l}}$
1.00	0.01	0.09	0.19	0.29	0.40	0.50	0.60	0.70	0.80	0.90	$\overline{\widehat{p_{inv}}}$
	1.05	1.00	0.98	0.98	1.02	1.03	1.03	1.04	1.00	1.02	$\overline{\hat{\alpha}}$
	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.99	1.00	0.99	$\overline{\hat{l}}$
24-taxon tree (tree length = 4.5)											
0.10	0.00	0.10	0.20	0.30	0.40	0.50	0.60	0.70	0.80	0.90	$\overline{\widehat{p_{inv}}}$
	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.10	$\overline{\hat{\alpha}}$
	4.50	4.49	4.49	4.50	4.51	4.51	4.51	4.52	4.52	4.19	$\overline{\hat{l}}$
0.50	0.00	0.10	0.20	0.30	0.40	0.50	0.60	0.70	0.80	0.90	$\overline{\widehat{p_{inv}}}$
	0.50	0.50	0.50	0.50	0.50	0.50	0.50	0.50	0.50	0.50	$\overline{\hat{\alpha}}$
	4.50	4.50	4.50	4.50	4.50	4.50	4.50	4.50	4.50	4.50	$\overline{\hat{l}}$
1.00	0.00	0.10	0.20	0.30	0.40	0.50	0.60	0.70	0.80	0.90	$\overline{\widehat{p_{inv}}}$
	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	$\overline{\hat{\alpha}}$
	4.50	4.50	4.50	4.50	4.50	4.50	4.50	4.51	4.50	4.50	$\overline{\hat{l}}$
96-taxon tree (tree length = 18.9)											
0.10	0.00	0.10	0.20	0.30	0.40	0.50	0.60	0.70	0.80	0.90	$\overline{\widehat{p_{inv}}}$
	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.10	$\overline{\hat{\alpha}}$
	18.90	18.89	18.90	18.91	18.89	18.92	18.97	18.98	19.02	18.80	$\overline{\hat{l}}$
0.50	0.00	0.10	0.20	0.30	0.40	0.50	0.60	0.70	0.80	0.90	$\overline{\widehat{p_{inv}}}$
	0.50	0.50	0.50	0.50	0.50	0.50	0.50	0.50	0.50	0.50	$\overline{\hat{\alpha}}$
	18.90	18.90	18.91	18.89	18.90	18.90	18.90	18.92	18.91	18.86	$\overline{\hat{l}}$
1.00	0.00	0.10	0.20	0.30	0.40	0.50	0.60	0.70	0.80	0.90	$\overline{\widehat{p_{inv}}}$
	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	$\overline{\hat{\alpha}}$
	18.90	18.89	18.90	18.89	18.90	18.90	18.90	18.89	18.90	18.87	$\overline{\hat{l}}$