

Robustness of Compound Dirichlet Priors for Bayesian Inference of Branch Lengths

CHI ZHANG^{1,2,3,4}, BRUCE RANNALA^{1,4}, AND ZIHENG YANG^{1,5,*}

¹Center for Computational and Evolutionary Biology, Institute of Zoology, Chinese Academy of Sciences, Beijing 100101, China; ²State Key Laboratory of Integrated Management of Pest Insects and Rodents, Institute of Zoology, Chinese Academy of Sciences, Beijing 100101, China; ³Graduate University of Chinese Academy of Sciences, Beijing 100049, China; ⁴Genome Center and Department of Evolution & Ecology, University of California, Davis, CA 95616, USA; ⁵Department of Biology, University College London, Darwin Building, Gower Street, London WC1E 6BT, UK;

*Correspondence to be sent to: Department of Biology, University College London, Darwin Building, Gower Street, London WC1E 6BT, UK; E-mail: z.yang@ucl.ac.uk.

Received 10 September 2011; reviews returned 20 November 2011; accepted 13 January 2012

Associate Editor: Peter Foster

Abstract.—We modified the phylogenetic program MrBayes 3.1.2 to incorporate the compound Dirichlet priors for branch lengths proposed recently by Rannala, Zhu, and Yang (2012). Tail paradox, partial identifiability and influential priors in Bayesian branch length inference. *Mol. Biol. Evol.* 29:325–335.) as a solution to the problem of branch-length overestimation in Bayesian phylogenetic inference. The compound Dirichlet prior specifies a fairly diffuse prior on the tree length (the sum of branch lengths) and uses a Dirichlet distribution to partition the tree length into branch lengths. Six problematic data sets originally analyzed by Brown, Hedtke, Lemmon, and Lemmon (2010). When trees grow too long: investigating the causes of highly inaccurate Bayesian branch-length estimates. *Syst. Biol.* 59:145–161) are reanalyzed using the modified version of MrBayes to investigate properties of Bayesian branch-length estimation using the new priors. While the default exponential priors for branch lengths produced extremely long trees, the compound Dirichlet priors produced posterior estimates that are much closer to the maximum likelihood estimates. Furthermore, the posterior tree lengths were quite robust to changes in the parameter values in the compound Dirichlet priors, for example, when the prior mean of tree length changed over several orders of magnitude. Our results suggest that the compound Dirichlet priors may be useful for correcting branch-length overestimation in phylogenetic analyses of empirical data sets. [Bayesian phylogenetics; branch lengths; compound Dirichlet prior; MrBayes.]

It has recently been noted that Bayesian analyses of phylogenies using the program MrBayes (Ronquist and Huelsenbeck 2003) may produce unreasonably large trees; the tree length (the sum of branch lengths) may be orders of magnitude greater than the maximum likelihood estimate (MLE) under the same model. The phenomenon was investigated by Brown et al. (2010), Marshall (2010), and Rannala et al. (2012). Brown et al. (2010) proposed several hypotheses for possible causes of the problem, such as the existence of multiple local peaks or large nearly flat regions in the posterior that cause mixing problems for the Markov chain Monte Carlo (MCMC), or an overly informative prior favoring unreasonably large branch lengths. A detailed analysis by Rannala et al. (2012) led to the suggestion that the fundamental cause of the problem is the poor default prior on branch lengths used in MrBayes, which assigns independent and identical distributions (i.i.d.) for branch lengths on the tree. This prior places too much probability density on large tree lengths, and the degree of misspecification increases for large trees comprising many taxa because the prior mean and variance of the tree length increase linearly with the number of taxa.

To resolve the problem of extreme posterior branch lengths, Rannala et al. (2012) suggested the use of compound Dirichlet priors on branch lengths, in which a diffuse gamma or inverse gamma distribution is assigned on the tree length, and the Dirichlet distribution (conditional on tree length) is applied to partition the total tree length into branch lengths. The compound Dirichlet priors provide flexible distributions on the tree length and

branch lengths, and the prior mean of tree length does not necessarily increase with an increase of the number of taxa. Rannala et al. (2012) implemented the priors to estimate branch lengths on a star tree. Tests on simulated data suggested that the priors appeared to be diffuse enough to allow the data to dominate the posterior estimates of branch lengths and tree length. However, the priors have not been implemented in any tree inference programs.

We implemented the compound Dirichlet priors (Rannala et al. 2012), as well as the two-exponential prior (Yang and Rannala 2005; Yang 2007), by modifying the source code of MrBayes 3.1.2 (Ronquist and Huelsenbeck 2003). We then used the modified program to analyze the six (two simulated and four empirical) data sets of Brown et al. (2010). While the default i.i.d. exponential prior implemented in MrBayes generated unreasonably large tree lengths and branch lengths for those data sets, the Bayesian estimates are all reasonable under the new compound Dirichlet priors; for example, the posterior means are now close to the MLEs. The two-exponential prior, with the mean of internal branch lengths smaller than for external ones, performed better than the default i.i.d. exponential prior but not as well as the compound Dirichlet priors.

MATERIALS AND METHODS

The Exponential and Compound Dirichlet Priors

By default, MrBayes uses i.i.d. exponential priors on branch lengths (Ronquist and Huelsenbeck 2003).

The probability density function of each branch length t_i is

$$f(t_i|\beta) = \beta e^{-\beta t_i}, \quad i = 1, 2, \dots, (2s - 3), \quad (1)$$

where $1/\beta$ is the mean and s is the number of taxa. The prior distribution of the tree length, $T = \sum_{i=1}^{2s-3} t_i$, has the gamma distribution with shape parameter $(2s - 3)$ and rate parameter β .

$$f(T|\beta, s) = \frac{\beta^{2s-3}}{(2s-4)!} T^{2s-4} e^{-\beta T}. \quad (2)$$

When $s \gg 1$, the gamma distribution is approximately normal, with a very large mean $(2s - 3)/\beta$ and a moderate standard deviation $\sqrt{2s - 3}/\beta$. For example, if $s=100$ and $\beta = 10$, the prior mean of tree length is 19.7 and the 99% interval is (16.3, 22.9). For data sets of highly similar sequences, this will be an extremely informative prior, favoring unreasonably large branch and tree lengths.

The compound Dirichlet prior specifies a fairly diffuse prior on the tree length T and then partitions the tree length into the branch lengths according to a Dirichlet distribution with concentration parameter α (Rannala et al. 2012). If $T \sim \text{Gamma}(\alpha_T, \beta_T)$:

$$f(T) = \frac{\beta_T^{\alpha_T}}{\Gamma(\alpha_T)} T^{\alpha_T-1} e^{-\beta_T T}, \quad \alpha_T > 0, \beta_T > 0, \quad (3)$$

with mean α_T/β_T and variance α_T/β_T^2 , the joint distribution of branch lengths $\mathbf{t} = \{t_1, t_2, \dots, t_{2s-3}\}$ is

$$f(\mathbf{t}|\alpha_T, \beta_T, \alpha, c) = \frac{\beta_T^{\alpha_T}}{\Gamma(\alpha_T)} e^{-\beta_T \sum_{i=1}^{2s-3} t_i} \times \frac{1}{B(\alpha, c)} \prod_{j=1}^s t_j^{\alpha-1} \prod_{k=1}^{s-3} t_k^{\alpha c-1} \times \left(\sum_{i=1}^{2s-3} t_i \right)^{\alpha_T - \alpha s - \alpha c(s-3)} \quad (4)$$

(Rannala et al. 2012, Equation 36). Here, the concentration parameter α of the Dirichlet distribution is inversely related to the variance of the branch lengths, while c is the ratio of the prior means for the internal and external branch lengths. Rannala et al. (2012) recommended $\alpha_T = 1$ in the gamma prior, while β_T should be chosen so that the prior mean of T is reasonable for the data set being analyzed.

Alternatively if $T \sim \text{invGamma}(\alpha_T, \beta_T)$:

$$f(T) = \frac{\beta_T^{\alpha_T}}{\Gamma(\alpha_T)} T^{-\alpha_T-1} e^{-\beta_T/T}, \quad \alpha_T > 2, \beta_T > 0, \quad (5)$$

with mean $\beta_T/(\alpha_T - 1)$ and variance $\beta_T^2/((\alpha_T - 1)^2(\alpha_T - 2))$, and if the same Dirichlet distribution is used to partition the tree length into branch lengths, the joint distribution of branch lengths is

$$f(\mathbf{t}|\alpha_T, \beta_T, \alpha, c) = \frac{\beta_T^{\alpha_T}}{\Gamma(\alpha_T)} e^{-\beta_T/\sum_{i=1}^{2s-3} t_i} \times \frac{1}{B(\alpha, c)} \prod_{j=1}^s t_j^{\alpha-1} \prod_{k=1}^{s-3} t_k^{\alpha c-1} \times \left(\sum_{i=1}^{2s-3} t_i \right)^{-\alpha_T - \alpha s - \alpha c(s-3)} \quad (6)$$

(Rannala et al. 2012, Equation 39). Rannala et al. (2012) recommended $\alpha_T = 3$ in the inverse gamma prior, while β_T should be chosen so that the prior mean fits the data set.

Note that the default i.i.d. exponential prior is a special case of the compound gamma-Dirichlet prior. If we set $\alpha_T = 2s - 3$, $\beta_T = \beta$, and $\alpha = c = 1$, Equation 4 becomes

$$f(\mathbf{t}) = \beta_T^{2s-3} e^{-\beta_T \sum_{i=1}^{2s-3} t_i}. \quad (7)$$

This is the joint probability density of $2s - 3$ i.i.d. exponential random variables with common rate β_T (cf. Equation 1).

We implemented the compound Dirichlet priors (Rannala et al. 2012) and the two-exponential prior (Yang and Rannala 2005; Yang 2007) by modifying the source code of MrBayes 3.1.2 (Ronquist and Huelsenbeck 2003). We validated our modification to the program by making use of the fact that the posterior should become the prior when the likelihood is set to a constant. We thus compared the posterior of branch and tree lengths sampled using the MCMC when no sequence data are used (a constant likelihood) with the theoretical expectations from the priors (see supplementary material for our test results: DOI:10.5061/dryad.j1kn5tq6). The test indicates that our modifications are correct.

Reanalysis of the Six Data Sets of Brown et al. (2010)

We applied the modified program to analyze the six data sets of Brown et al. (2010). Two of the six data sets are simulated (Brown and Lemmon 2007), while the other four are for lizards (Leache and Mulcahy 2007), frogs (Gamble et al. 2008), clams (Hedtke et al. 2008), and froglets (Symula et al. 2008). See Brown et al. (2010) for more details of the data sets.

We set the prior mean on the tree length to be 0.01, 0.1, 0.2, 0.5, 1, 2, 10, and 100, respectively. In the compound Dirichlet priors, we used the recommended value $\alpha_T = 1$, so that $\beta_T = 100, 10, 5, 2, 1, 0.5, 0.1$, and 0.01, respectively, for the gamma prior on tree length. For the inverse gamma prior, we used the recommended value $\alpha_T = 3$, so that $\beta_T = 0.02, 0.2, 0.4, 1, 2, 4, 20$, and 200, respectively. Parameters $\alpha = c = 1$ are fixed, which specify a uniform Dirichlet distribution of branch lengths given the tree length. For empirical data analysis, we recommend that the β_T parameter be chosen to reflect the sequence divergence level in the data set being analyzed. Here, we used many different values to assess the impact of the prior on the posterior estimates of branch lengths and tree length.

For comparison, we also used the default exponential priors for branch lengths implemented in MrBayes. We analyzed the data sets using both the default prior mean of 0.1 and using several different prior means for the tree length. For the i.i.d. exponential prior, the rate is set to $\beta = (2s - 3)/\bar{T}$. For the two-exponential prior, the internal branch branches are set to one-tenth of the external ones; that is, the internal rate is set to $\beta_I = (10s + s - 3)/\bar{T}$, and the external rate is set to $\beta_E = (10s + s - 3)/(10\bar{T})$. We

note that the two-exponential priors were proposed to reduce the spuriously high posterior probabilities for trees or clades and not intended for branch-length estimation.

We initially used the substitution model GTR + I + Γ_4 (Yang 1994a, 1994b) in the Bayesian analysis because this was the model used by Brown et al. (2010). However, as pointed out by several authors (e.g., Minin et al. 2003; Ren et al. 2005), the widely used I + Γ_4 model is pathological because both the invariable sites (“I”) and the gamma distribution (“ Γ ”) describe the same phenomenon of variable rates among sites, and the use of both in the same model leads to a strong correlation in the estimates of the proportion of the invariable sites p_0 and the gamma shape parameter α . The correlation may in turn impact estimation of branch lengths and tree length, especially in the maximum likelihood (ML) analysis. Thus, we also conducted the Bayesian analysis under GTR + Γ_4 without the I component. For each parameter setting, we ran four MCMC chains (one cold and three hot), 10 million iterations, and took samples every 1000 iterations. The last 8000 samples were used to calculate the posterior distributions for the parameters.

RESULTS AND DISCUSSION

The results of analyzing the six data sets of Brown et al. (2010) under the GTR + I + Γ_4 model and a variety of branch-length priors are shown in Figure 1 and Table 1. For each data set, the posterior means and medians of tree lengths are very similar, so only the means are shown.

The Exponential Priors

The default exponential prior for branch lengths in MrBayes causes the posterior tree lengths to be dominated by the prior (Table 1: OneExp). The posterior tree length is either too small, or too large, by comparison with the MLE depending on whether the prior mean is smaller, or larger, than the MLE. The default prior mean 0.1 ($\beta = 10$) gives posterior tree length 5.25 (3.91, 6.72) for simulated data set A, 5.25 (3.86, 6.83) for simulated data set B, 0.96 (0.83, 1.11) for the frogs data set, 14.0 (10.86, 17.59) for clams, 4.54 (3.81, 5.47) for the lizards data set, and 2.49 (1.62, 3.34) for the froglets data set (Brown et al. 2010). All these estimates are too large by comparison with the MLEs. If one uses the MLE of tree length to calculate the mean branch length and uses it as the prior mean (instead of the default prior mean 0.1) for the independent exponential prior, the posterior will be close to the MLE. However, the procedure is non-Bayesian, and the resulting exponential prior will be too informative (and thus will produce credible intervals for the branch lengths that are too narrow). The two-exponential prior performs slightly better than the default independent exponentials but it also has too much influence on the posterior (Table 1: twoExp).

The Compound Dirichlet Priors

The compound Dirichlet priors have much less influence on the posterior estimates. With the default shape

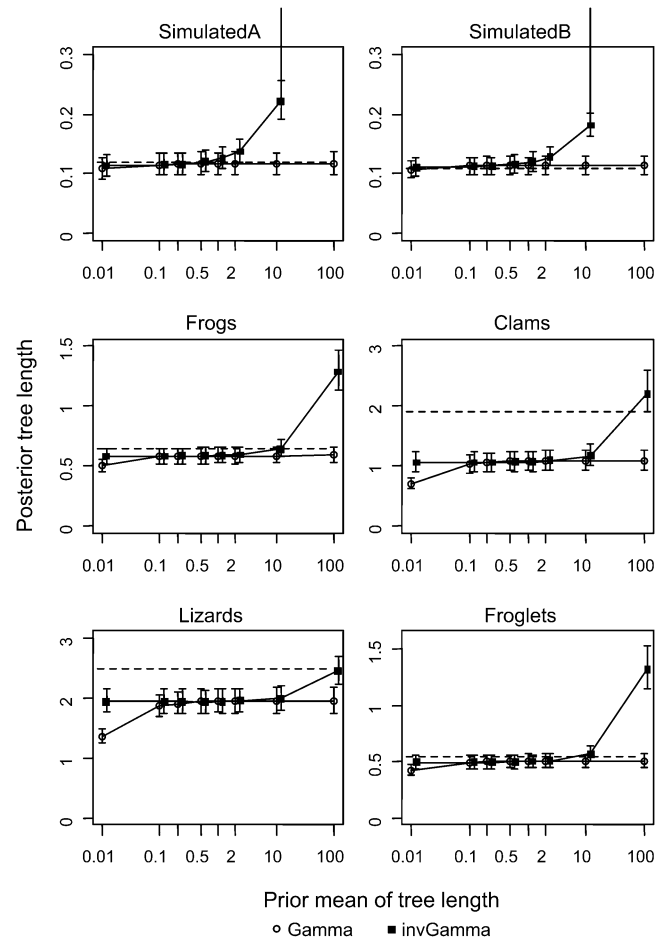


FIGURE 1. The posterior means and 95% CIs obtained in reanalysis of the six data sets of Brown et al. (2010) using the compound Dirichlet priors of Rannala et al. (2012). For the gamma prior on tree length, the shape parameter is fixed at $\alpha_T = 1$, while the prior mean of tree length is fixed at 0.01, 0.1, 0.2, 0.5, 1, 2, 10, 100, so that $\beta_T = 100, 10, 5, 2, 1, 0.5, 0.1, 0.01$. For the inverse gamma prior, $\alpha_T = 3$, so that $\beta_T = 0.02, 0.2, 0.4, 1, 2, 4, 20, 200$. The x -axis is on the logarithmic scale. The concentration parameter of the Dirichlet distribution is fixed at $\alpha = 1$. We used $c = 1$, so that the prior means of internal and external branch lengths are the same. The MLEs are indicated by the dashed lines.

parameter $\alpha_T = 1$ for the gamma Dirichlet prior, the posterior tree lengths are quite stable for all six data sets when the prior mean (or β_T) varied over four orders of magnitude (Fig. 1 and Table 1: GammaDir). Even though the inverse gamma–Dirichlet prior (with the default shape parameter $\alpha_T = 3$) shows more sensitivity, the posterior tree lengths are stable for the six data sets when the prior mean (or β_T) varied over two or three orders of magnitude (Fig. 1 and Table 1: invGammaDir). In real data analysis, one will usually have prior information about the level of sequence divergence that provides an order-of-magnitude estimate of the tree length, so that a more informative prior mean can be specified in the inverse gamma–Dirichlet prior.

For the two simulated data sets, the posterior means are close to the MLEs and the 95% credibility intervals (CIs) cover the MLEs. For the frogs and froglets data set,

TABLE 1. Posterior means and 95% CIs of the tree length

Data set	Prior mean \bar{T}	Tree length (mean, 95% CI)			
		GammaDir(1, β_1 , 1, 1)	invGamDir(3, β_2 , 1, 1)	oneExp(β)	twoExp(β_I , β_E)
Simulated A: $s = 29$, MLE = 0.12 (0.12, $\alpha =$ 0.02) (Brown and Lemmon 2007)	0.01	0.109 (0.092, 0.126)	0.114 (0.097, 0.131)	0.034 (0.030, 0.038)	0.028 (0.024, 0.032)
	0.1	0.115 (0.099, 0.135)	0.115 (0.098, 0.134)	0.111 (0.097, 0.126)	0.087 (0.075, 0.100)
	1	0.116 (0.099, 0.135)	0.126 (0.108, 0.145)	0.155 (0.133, 0.182)	0.143 (0.123, 0.165)
	10	0.116 (0.099, 0.135)	0.221 (0.190, 0.256)	9.746 (7.334, 12.67)	3.809 (2.722, 5.151)
	100	0.116 (0.099, 0.136)	78.73 (28.78, 203.4)	98.37 (73.72, 127.7)	41.02 (30.12, 54.13)
Simulated B: $s = 29$, MLE = 0.11 (0.11, $\alpha =$ 0.14) (Brown and Lemmon 2007)	0.01	0.107 (0.094, 0.122)	0.111 (0.096, 0.126)	0.038 (0.034, 0.042)	0.032 (0.028, 0.037)
	0.1	0.113 (0.098, 0.128)	0.112 (0.098, 0.128)	0.110 (0.097, 0.123)	0.090 (0.079, 0.101)
	1	0.113 (0.098, 0.128)	0.120 (0.105, 0.136)	0.139 (0.123, 0.158)	0.133 (0.117, 0.150)
	10	0.113 (0.098, 0.129)	0.181 (0.163, 0.202)	9.686 (7.262, 12.54)	3.839 (2.726, 5.113)
	100	0.113 (0.099, 0.129)	88.90 (26.47, 244.9)	98.86 (74.16, 127.3)	41.58 (29.88, 55.02)
Frogs: $s = 66$, MLE = 0.64 (0.61, $\alpha = 0.20$) (Gamble et al. 2008)	0.01	0.499 (0.450, 0.552)	0.580 (0.516, 0.648)	0.043 (0.040, 0.047)	0.038 (0.032, 0.043)
	0.1	0.574 (0.510, 0.642)	0.579 (0.517, 0.647)	0.272 (0.251, 0.295)	0.194 (0.173, 0.216)
	1	0.584 (0.522, 0.653)	0.586 (0.523, 0.654)	0.673 (0.608, 0.743)	0.503 (0.457, 0.551)
	10	0.585 (0.523, 0.655)	0.641 (0.574, 0.715)	0.941 (0.815, 1.089)	0.836 (0.738, 0.950)
	100	0.586 (0.523, 0.655)	1.285 (1.129, 1.468)	69.94 (57.34, 83.99)	19.48 (12.05, 29.96)
Clams: $s = 93$, MLE = 1.96 (1.75, $\alpha = 0.25$) (Hedtke et al. 2008)	0.01	0.707 (0.620, 0.802)	1.055 (0.908, 1.231)	0.022 (0.020, 0.025)	0.021 (0.019, 0.024)
	0.1	1.019 (0.871, 1.184)	1.055 (0.901, 1.227)	0.202 (0.183, 0.222)	0.188 (0.167, 0.210)
	1	1.073 (0.919, 1.246)	1.066 (0.914, 1.239)	1.033 (0.928, 1.148)	0.963 (0.850, 1.085)
	10	1.079 (0.930, 1.249)	1.170 (1.008, 1.351)	5.730 (4.033, 7.610)	3.318 (2.616, 4.132)
	100	1.080 (0.923, 1.256)	2.204 (1.894, 2.600)	99.21 (85.67, 114.6)	66.14 (54.95, 78.47)
Lizards: $s = 123$, MLE = 2.48 (2.11, $\alpha = 0.19$) (Leache and Mulcahy 2007)	0.01	1.361 (1.243, 1.484)	1.939 (1.755, 2.147)	0.034 (0.032, 0.037)	0.027 (0.025, 0.029)
	0.1	1.858 (1.677, 2.044)	1.938 (1.747, 2.155)	0.303 (0.281, 0.325)	0.235 (0.216, 0.256)
	1	1.948 (1.758, 2.147)	1.935 (1.744, 2.140)	1.457 (1.351, 1.565)	1.190 (1.096, 1.289)
	10	1.947 (1.751, 2.164)	1.986 (1.789, 2.189)	3.379 (3.006, 3.800)	2.873 (2.584, 3.204)
	100	1.956 (1.748, 2.176)	2.448 (2.221, 2.699)	57.38 (49.82, 74.43)	24.34 (18.71, 29.74)
Froglets: $s = 92$, MLE = 0.55 (0.55, $\alpha = 0.12$) (Symula et al. 2008)	0.01	0.426 (0.380, 0.477)	0.494 (0.437, 0.555)	0.032 (0.029, 0.034)	0.016 (0.015, 0.018)
	0.1	0.493 (0.437, 0.556)	0.495 (0.438, 0.557)	0.214 (0.197, 0.232)	0.115 (0.103, 0.127)
	1	0.503 (0.443, 0.568)	0.503 (0.444, 0.565)	0.648 (0.581, 0.720)	0.389 (0.354, 0.426)
	10	0.501 (0.446, 0.564)	0.568 (0.507, 0.637)	1.609 (1.208, 2.170)	0.868 (0.755, 0.996)
	100	0.502 (0.444, 0.568)	1.318 (1.146, 1.528)	98.96 (84.45, 113.5)	4.854 (2.094, 11.44)

Notes: The first column shows summary information for each data set: taxonomic group, number of taxa, MLE of the tree length, and reference. The MLEs outside the brackets are calculated using PAUP* 4b10 (Swofford 2003) under the GTR + I + Γ_4 model, as in Brown et al. (2010); those inside the brackets are calculated using PhyML (Guindon et al. 2010) under the GTR + Γ_4 model (subtree pruning and regrafting (SPR) moves), with inferred gamma shape parameters. The data sets are then reanalyzed using the modified MrBayes under the GTR + I + Γ_4 model and four priors: the gamma-Dirichlet prior (column 3), the inverse gamma-Dirichlet prior (column 4), the i.i.d. exponential prior (column 5, the default in the original MrBayes), and the two-exponential prior (column 6). For the gamma-Dirichlet prior, $\alpha_T = 1$, $\beta_T = 100, 10, 1, 0.1, 0.01$; for the inverse gamma-Dirichlet prior, $\alpha_T = 3$, $\beta_T = 0.02, 0.2, 2, 20, 200$; for the i.i.d. exponential prior, $\beta = (2s - 3)/\bar{T}$; for the two-exponential prior, $\beta_I = (10s + s - 3)/\bar{T}$ and $\beta_E = (10s + s - 3)/(10\bar{T})$. The posterior mean and 95% CI of the tree length are shown for each analysis.

the posterior means/medians are very near the MLEs but are slightly smaller. For the clams and lizards data set, most of the 95% CIs do not cover the MLEs, and the Bayesian estimates of tree lengths are smaller than the MLEs. This appears to be a defect of the MLEs, rather than the Bayesian estimates (see below). In summary, the posterior mean tree lengths (branch lengths) under the compound Dirichlet priors are much closer to the MLEs than those under the exponential priors.

Model Complexity and Partial Identifiability

We investigated the possible reasons for the greater differences observed between the Bayesian and ML estimates of tree length in the clams and lizards data sets. We conducted both ML and Bayesian analyses (using GammaDir(1, 1, 1, 1)) under several additional substitution models: JC69, JC69 + Γ_4 , JC69 + I + Γ_4 , GTR, GTR + Γ_4 , and GTR + I + Γ_4 . The results are summarized in Table 2. For the clams data set, the Bayesian posterior estimates of tree length were smaller than the MLEs

under both GTR + Γ_4 and GTR + I + Γ_4 , mainly because the MLEs of the branch lengths to two outgroup species

TABLE 2. MLEs and posterior means of the tree length for the clams and lizards data sets

Model	Tree length	
	PAUP*	MrBayes
Clams data set		
JC69	0.89	0.88
JC69 + Γ_4	1.15	1.04
JC69 + I + Γ_4	1.17	1.04
GTR	0.93	0.90
GTR + Γ_4	1.75	1.08
GTR + I + Γ_4	1.96	1.07
Lizards data set		
JC69	1.50	1.42
JC69 + Γ_4	1.77	1.62
JC69 + I + Γ_4	1.77	1.60
GTR	1.53	1.43
GTR + Γ_4	2.11	1.83
GTR + I + Γ_4	2.48	1.95

Note: The branch-length prior used in the MrBayes analysis is GammaDir(1, 1, 1, 1).

TABLE 3. Posterior means and 95% CIs of the tree length under the GTR + Γ_4 model

Data set	Prior mean \bar{T}	Tree length (mean, 95% CI)			
		GammaDir(1, β_1 , 1, 1)	invGamDir(3, β_2 , 1, 1)	oneExp(β)	twoExp(β_I , β_E)
Simulated A	0.01	0.118 (0.103, 0.134)	0.123 (0.107, 0.141)	0.041 (0.037, 0.046)	0.037 (0.032, 0.042)
	0.1	0.125 (0.108, 0.143)	0.124 (0.108, 0.142)	0.118 (0.104, 0.133)	0.096 (0.084, 0.108)
	1	0.125 (0.109, 0.143)	0.133 (0.115, 0.152)	0.158 (0.138, 0.181)	0.148 (0.129, 0.169)
	10	0.125 (0.109, 0.144)	0.205 (0.182, 0.230)	0.165 (0.143, 0.189)	0.164 (0.142, 0.187)
	100	0.125 (0.108, 0.144)	1.368 (1.159, 1.605)	0.166 (0.143, 0.190)	0.165 (0.143, 0.190)
Simulated B	0.01	0.108 (0.094, 0.123)	0.111 (0.097, 0.127)	0.040 (0.036, 0.045)	0.036 (0.032, 0.041)
	0.1	0.113 (0.098, 0.129)	0.112 (0.098, 0.127)	0.110 (0.097, 0.123)	0.090 (0.080, 0.102)
	1	0.113 (0.098, 0.129)	0.120 (0.106, 0.136)	0.140 (0.123, 0.158)	0.133 (0.117, 0.150)
	10	0.113 (0.098, 0.129)	0.182 (0.163, 0.202)	0.144 (0.126, 0.163)	0.143 (0.125, 0.163)
	100	0.113 (0.099, 0.129)	3.558 (2.579, 4.887)	0.145 (0.127, 0.165)	0.144 (0.127, 0.164)
Frogs	0.01	0.507 (0.459, 0.557)	0.580 (0.521, 0.642)	0.061 (0.057, 0.065)	0.057 (0.052, 0.069)
	0.1	0.577 (0.518, 0.639)	0.580 (0.520, 0.645)	0.286 (0.265, 0.308)	0.225 (0.205, 0.246)
	1	0.585 (0.526, 0.648)	0.585 (0.526, 0.649)	0.665 (0.601, 0.732)	0.509 (0.465, 0.557)
	10	0.586 (0.527, 0.650)	0.633 (0.571, 0.701)	0.866 (0.768, 0.977)	0.791 (0.704, 0.884)
	100	0.587 (0.528, 0.650)	1.138 (1.023, 1.267)	98.17 (81.52, 115.1)	0.888 (0.780, 1.004)
Clams	0.01	0.728 (0.643, 0.820)	1.062 (0.917, 1.226)	0.034 (0.031, 0.038)	0.032 (0.030, 0.035)
	0.1	1.023 (0.885, 1.179)	1.065 (0.919, 1.230)	0.242 (0.220, 0.267)	0.234 (0.215, 0.254)
	1	1.081 (0.930, 1.253)	1.075 (0.927, 1.235)	1.036 (0.936, 1.146)	0.964 (0.856, 1.084)
	10	1.084 (0.933, 1.259)	1.174 (1.015, 1.355)	4.548 (3.513, 6.073)	2.897 (2.267, 3.620)
	100	1.092 (0.940, 1.264)	2.112 (1.833, 2.445)	98.62 (84.42, 113.7)	65.64 (54.27, 78.15)
Lizards	0.01	1.498 (1.404, 1.596)	1.817 (1.690, 1.958)	0.066 (0.062, 0.070)	0.064 (0.061, 0.066)
	0.1	1.785 (1.661, 1.920)	1.821 (1.692, 1.963)	0.479 (0.457, 0.501)	0.472 (0.449, 0.497)
	1	1.825 (1.693, 1.970)	1.822 (1.701, 1.963)	1.535 (1.452, 1.623)	1.383 (1.299, 1.473)
	10	1.828 (1.697, 1.968)	1.847 (1.717, 1.989)	2.549 (2.336, 2.786)	2.328 (2.137, 2.539)
	100	1.831 (1.701, 1.972)	2.108 (1.957, 2.274)	68.03 (57.65, 80.30)	37.07 (31.89, 43.27)
Froglets	0.01	0.460 (0.414, 0.510)	0.525 (0.470, 0.584)	0.049 (0.046, 0.052)	0.040 (0.037, 0.043)
	0.1	0.523 (0.470, 0.580)	0.525 (0.471, 0.584)	0.253 (0.236, 0.270)	0.173 (0.159, 0.187)
	1	0.529 (0.474, 0.590)	0.532 (0.478, 0.591)	0.647 (0.587, 0.711)	0.418 (0.384, 0.456)
	10	0.530 (0.476, 0.590)	0.582 (0.523, 0.645)	0.947 (0.834, 1.075)	0.797 (0.711, 0.888)
	100	0.530 (0.476, 0.593)	1.054 (0.956, 1.166)	100.0 (85.71, 115.4)	0.988 (0.865, 1.130)

Note: See legend to Table 1.

(Neocorbicula and Polymesoda) were very large, while the Bayesian estimates are less extreme due to the influence of the prior. In such cases of extreme branch lengths, the Bayesian method is expected to outperform ML, even if judged by Frequentist properties (see, e.g., Berger et al. 1999). Note that the MLE of branch lengths and tree length can be infinite so that neither the mean nor the variance of the MLE exists. We also used another prior, GammaDir(1, 1, 0.5, 0.2), which favors variable branch lengths and longer external branches. The posterior tree length is 1.42 under GTR + Γ_4 , closer to the MLE under this model.

For the lizards data set, there is no clear pattern of long branches as in the clams data set. Nevertheless, the Bayesian analysis showed a moderating or shrinking effect compared with the MLEs, as in the clams data set. The ML analysis, in particular, is heavily impacted by the use of the GTR + I + Γ_4 model (Table 2). With the prior GammaDir(1, 1, 0.5, 0.2), the posterior tree length under GTR + Γ_4 is 2.00, quite similar to the MLE under this model.

The results for the GTR + Γ_4 model are summarized in Table 3. Similar patterns are observed as were discussed above with respect to the sensitivity of the posterior to the exponential priors and the

robustness of the compound Dirichlet priors. However, it is notable that the independent exponential prior did not produce extremely long branch lengths in the two simulated data sets, and estimates were stable over a wider range of prior mean (Table 3). The use of the complex GTR + I + Γ_4 model thus appears to have exacerbated the problem of extremely long branch lengths (Rannala et al. 2012).

In summary, the compound Dirichlet priors appear to provide more robust and reasonable results and greatly reduce the discrepancy between posterior mean branch lengths and MLEs. For the clams and lizards data sets we analyzed, in which the MLEs are much larger than the posterior estimates, we suggest that the MLEs may be worse estimators, at least partly due to the use of an overparameterized substitution model (GTR + I + Γ_4) with strong correlations between parameter estimates. We suggest that the compound Dirichlet priors generally perform better (producing more reasonable branch length estimates) than the independent exponential priors implemented in the current version of MrBayes and therefore represent a better choice of default prior.

Program availability.—The modified MrBayes 3.1.2 source code is available from <http://abacus.gene.ucl.ac.uk/software/> or <http://www.rannala.org>.

SUPPLEMENTARY MATERIAL

Supplementary material, including data files and/or online-only appendices, can be found in the Dryad data repository (DOI:10.5061/dryad.j1kn5tq6).

FUNDING

C.Z. is supported by China Scholarship Council (File No. 2010491133).

ACKNOWLEDGMENTS

We thank Paul Lewis, Jeremy Brown, Peter Foster, and Ron Deby for comments, and De-Xing Zhang for discussions. Z.Y. is a Royal Society/Wolfson Merit Award Holder.

REFERENCES

- Berger J.O., Liseo B., Wolpert R.L. 1999. Integrated likelihood methods for eliminating nuisance parameters. *Stat. Sci.* 14:1–28.
- Brown J.M., Hedtke S.M., Lemmon A.R., Lemmon E.M. 2010. When trees grow too long: investigating the causes of highly inaccurate Bayesian branch-length estimates. *Syst. Biol.* 59:145–161.
- Brown J.M., Lemmon A.R. 2007. The importance of data partitioning and the utility of Bayes factors in Bayesian phylogenetics. *Syst. Biol.* 56:643–655.
- Gamble T., Berendzen P.B., Bradley Shaffer H., Starkey D.E., Simons A.M. 2008. Species limits and phylogeography of North American cricket frogs (*Acris*: Hylidae). *Mol. Phylogenet. Evol.* 48:112–125.
- Guindon S., Dufayard J.F., 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.* 59:307–321.
- Hedtke S.M., Stanger-Hall K., Baker R.J., Hillis D.M. 2008. All-male asexuality: origin and maintenance of androgenesis in the Asian clam *Corbicula*. *Evolution*. 62:1119–1136.
- Leache A.D., Mulcahy D.G. 2007. Phylogeny, divergence times and species limits of spiny lizards (*Sceloporus magister* species group) in western North American deserts and Baja California. *Mol. Ecol.* 16:5216–5233.
- Marshall D.C. 2010. Cryptic failure of partitioned Bayesian phylogenetic analyses: lost in the land of long trees. *Syst. Biol.* 59:108–117.
- Minin V., Abdo Z., Joyce P., Sullivan J. 2003. Performance-based selection of likelihood models for phylogeny estimation. *Syst. Biol.* 52:674–683.
- Rannala B., Zhu T., Yang Z. 2012. Tail paradox, partial identifiability and influential priors in Bayesian branch length inference. *Mol. Biol. Evol.* 29:325–335.
- Ren F., Tanaka H., Yang Z. 2005. An empirical examination of the utility of codon-substitution models in phylogeny reconstruction. *Syst. Biol.* 54:808–818.
- Ronquist F., Huelsenbeck J.P. 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics*. 19:1572–1574.
- Swofford D.L. 2003. PAUP*: phylogenetic analysis using parsimony (*and other methods). Version 4. Sinauer Associates, Sunderland, Massachusetts.
- Symula R., Keogh J.S., Cannatella D.C. 2008. Ancient phylogeographic divergence in southeastern Australia among populations of the widespread common froglet, *Crinia signifera*. *Mol. Phylogenet. Evol.* 47:569–580.
- Yang Z. 1994a. Estimating the pattern of nucleotide substitution. *J. Mol. Evol.* 39:105–111.
- Yang Z. 1994b. Maximum likelihood phylogenetic estimation from DNA sequences with variable rates over sites: approximate methods. *J. Mol. Evol.* 39:306–314.
- Yang Z. 2007. Fair-balance paradox, star-tree paradox, and Bayesian phylogenetics. *Mol. Biol. Evol.* 24:1639–1655.
- Yang Z., Rannala B. 2005. Branch-length prior influences Bayesian posterior probability of phylogeny. *Syst. Biol.* 54:455–470.