The Impact of Modelling Rate Heterogeneity among Sites on Phylogenetic Estimates of Intraspecific Evolutionary Rates and Timescales

Fangzhi Jia*, Nathan Lo, Simon Y. W. Ho

School of Biological Sciences, University of Sydney, Sydney, New South Wales, Australia

Abstract

Phylogenetic analyses of DNA sequence data can provide estimates of evolutionary rates and timescales. Nearly all phylogenetic methods rely on accurate models of nucleotide substitution. A key feature of molecular evolution is the heterogeneity of substitution rates among sites, which is often modelled using a discrete gamma distribution. A widely used derivative of this is the gamma-invariable mixture model, which assumes that a proportion of sites in the sequence are completely resistant to change, while substitution rates at the remaining sites are gamma-distributed. For data sampled at the intraspecific level, however, biological assumptions involved in the invariable-sites model are commonly violated. We examined the use of these models in analyses of five intraspecific data sets. We show that using 6–10 rate categories for the discrete gamma distribution of rates among sites is sufficient to provide a good approximation of the marginal likelihood. Increasing the number of gamma rate categories did not have a substantial effect on estimates of the substitution rate or coalescence time, unless rates varied strongly among sites in a non-gamma-distributed manner. The assumption of a proportion of invariable sites provided a better approximation of the asymptotic marginal likelihood when the number of gamma categories was small, but had minimal impact on estimates of rates and coalescence times. However, the estimated proportion of invariable sites was highly susceptible to changes in the number of gamma rate categories. The concurrent use of gamma and invariable-site models for intraspecific data is not biologically meaningful and has been challenged on statistical grounds; here we have found that the assumption of a proportion of invariable sites has no obvious impact on Bayesian estimates of rates and timescales from intraspecific data.

Citation: Jia F, 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(5): e95722. doi:10.1371/journal.pone.0095722

Editor: Corrie S. Moreau, Field Museum of Natural History, United States of America

Received December 9, 2013; **Accepted** March 28, 2014; **Published** May 5, 2014

Funding: FJ was supported by an Australian Postgraduate Award. NL and SYWH were supported by the Australian Research Council (DP1097265 and DP110100383). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: fangzhi.jia@sydney.edu.au

Introduction

In phylogenetic analyses of DNA sequence data, the evolutionary process is usually described using models of nucleotide substitution. These models commonly assume that substitutions occurring at each site are described by a Markov chain and that different sites evolve in a mutually independent manner. In practice, almost all models of nucleotide substitution are time-reversible. The general time-reversible (GTR) model, formally described by Tavaré [1], includes parameters that allow unequal frequencies for the four nucleotides and a distinct rate for each of the six pairwise nucleotide substitutions. By constraining one or more of these parameters, a family of time-reversible models can be generated. This gives rise to special cases of the GTR model, such as the Jukes–Cantor [2], Kimura [3], and Hasegawa–Kishino–Yano [4] models.

In their basic form, nucleotide substitution models assume that the evolutionary process is homogeneous across sites. In reality, however, rates of mutation can vary among sites because of selective pressures associated with structural and functional constraints [5–7]. Some sites, such as CpG islands in mammalian taxa, have a higher propensity to mutate [8]. Failure to take into account this rate heterogeneity among sites (RHAS) can lead to biased estimation of branch lengths, with corresponding impacts on estimates of phylogenies, substitution rates, and evolutionary timescales [9–13].

RHAS can be modelled in a number of ways, but the most popular approach is to assume that the rate at each site is a random variable drawn from a statistical distribution. The gamma distribution (+Γ) is most commonly used for this purpose [14,15], owing to its ease of interpretation and its good fit to empirical data [5]. The shape of the gamma distribution, governed by the shape parameter α, can range from bell-shaped ($\alpha > 1$) to L-shaped ($\alpha < 1$). Consequently, the gamma distribution is capable of modelling various degrees of RHAS [5].

To reduce computational burden, most methods employ a discrete gamma model in which the continuous distribution is approximated by several rate classes with equal percentiles and probabilities [14]. Within each class, all of the rates are represented by the mean or median. The higher the number of rate categories, the better is the fit of the discrete gamma distribution to the continuous gamma distribution. The number of rate categories (k) used in phylogenetic analyses of nucleotide sequences generally ranges from 2 to 32 [16–18], with most analyses employing 4–10 categories. In an analysis of two small data sets (4 sequences, 1352 sites and 5 sequences, 570 sites), Yang [14] found that 4 rate categories provided sufficiently good approximations of α and the likelihood and that there was little improvement in estimation accuracy when more than 8 categories were used. A greater number of rate categories might be preferable for large data sets [19].

RHAS can also be described using discrete rate-class models, in which the rate categories are not determined from some underlying distribution. A special case of the discrete-rates model is the invariable-sites model (+I), which assumes that there is a rate class with a rate of 0 [4,20–23]. In this model, some proportion of sites (p_{inv}) is assumed to be invariable, or completely resistant to change, because mutations at these sites are strongly deleterious or fatal. The remaining sites are assumed to evolve at non-zero rates. The idea of invariable sites was inspired by studies of protein structures and has intuitive appeal [20,24]. The invariable-sites model can be combined with the gamma model of RHAS (+Γ+I) [25,26], and the resulting gamma-invariable mixture models are widely used.

When performing a phylogenetic analysis, the inclusion of a model of RHAS is usually determined using a model-selection approach. A common practice is to compare a range of substitution models using a model-selection criterion, such as the Akaike information criterion [27] or Bayesian information criterion [28]. In this respect, the Bayesian information criterion has been shown to perform well across a range of simulation scenarios [29]. The best-fitting model is then used in subsequent analyses of the data set.

Evolutionary models chosen using objective criteria are not always the most biologically pertinent [30], which raises questions about the meaning of the resulting estimates of parameters. A prominent example is the +I model, which is sometimes selected as the best-fitting RHAS model for data that have been sampled at the population level. Intraspecific data tend to display lower levels of variation than sequences that have been sampled at the species level and above. This makes the evaluation of p_{inv} highly sensitive to taxon sampling [5]. We would expect the number of constant sites to decline as sample size increases, leading to lower estimates of p_{inv}. Sites observed to be constant among sequences might not be invariable, but might simply have not experienced any mutations among the sequences that have been sampled. Additionally, deleterious mutations are much more likely to be found in population-level data [31,32], so that sites that would typically be treated as 'invariable' at the phylogenetic level might contain transient polymorphisms at the population level.

The +Γ+I mixture model, first used by Gu *et al.* [25], has been criticised on the grounds that the two parameters involved – p_{inv} and α – cannot be optimised independently of each other [15,19,33,34]. An L-shaped gamma distribution ($\alpha<1$) already accommodates a proportion of low-variability sites; as a consequence, adding a parameter to account for invariable sites creates a strong correlation between p_{inv} and α [34]. This might cause considerable problems during the parameter optimisation process, since it is impossible to obtain reliable estimates of both parameters simultaneously [33]. Combining this with the aforementioned sensitivity of p_{inv} to the size of the data set and the level of divergence, applying the +Γ+I model to intraspecific data sets appears particularly problematic, at least on theoretical grounds.

The impact of using different RHAS models in phylogenetic analyses of intraspecific data is not well understood. The impact of the choice of RHAS model is most likely to be seen in estimates of branch lengths, which can have subsequent effects on the inferred tree topology. Here we investigate how varying the RHAS model affects phylogenetic analyses based on the molecular clock. We focus on five intraspecific data sets, four of which comprise heterochronous sequences (ancient DNA and viruses). In heterochronous data set, the ages of the sequences provide internal calibrations for the molecular clock, making such data ideal for studying evolutionary rates and timescales at the intraspecific level. We test whether increasing the number of gamma categories or assuming a proportion of invariable sites affects estimates of substitution rates and coalescence times.

Materials and Methods

We assembled five intraspecific data sets: (i) complete mitogenomes from human haplogroup C1, (ii) complete mitogenomes from hominins, (iii) mitochondrial D-loop from muskox, (iv) *PB2* gene from H1N1 human influenza virus, and (v) concatenated genome fragment from HIV-1. The first data set comprised isochronous sequences from the present day, whereas the last four data sets comprised heterochronous sequences of known ages. Details of the datasets are listed in Table 1.

For each data set, sequence alignments were done using MUSCLE 3.8.31 [35] and adjusted manually. Sequences with uncertain ages were removed. For each data set, we performed preliminary Bayesian phylogenetic analyses using the best-fitting substitution models selected by the Bayesian information criterion. Alignments for all datasets used in this study are available in File S1. Bayesian phylogenetic analyses of the data sets were performed in BEAST v1.7.5 [36]. The best-fitting model of nucleotide substitution was selected for each data set using the Bayesian information criterion in Modelgenerator 0.85 [30]. We tested four different RHAS models: equal rates among sites, +Γ, +I, and +Γ+ I. For the +Γ models, we repeated the analysis using various numbers of rate categories ranging from 3–32. To test for rate heterogeneity among lineages, we initially used the uncorrelated lognormal relaxed clock [37]. For the human mitogenome data sets (i and ii), the coefficient of rate variation did not provide any evidence of rate variation among lineages. Marginal likelihoods were estimated using the harmonic-mean estimator [38].

Using the maximum-clade-credibility trees from our Bayesian analyses, we inferred the number of substitutions at each site using stochastic mutational mapping in SIMMAP [39,40]. We used the empirical nucleotide frequencies and applied the mean rate estimate from the Bayesian phylogenetic analysis of each data set to scale the branch lengths. Numbers were rounded to the nearest integer. The distribution of inferred mutational counts provided a picture of the RHAS pattern for each data set. We performed chi-squared tests to compare the goodness-of-fit of gamma and negative binomial distributions to the site-specific substitution counts. In the absence of RHAS, the distribution of site-specific substitution counts should conform to the Poisson distribution. In the presence of gamma-distributed RHAS, we expect the number of changes per site to conform to the negative binomial distribution. We simultaneously estimated the values of α by minimising χ^2. We compared the fit of these two distributions using the Akaike information criterion.

In the analyses of the heterochronous sequences (data sets ii to v), the molecular clock was calibrated using the ages of the sequences. Following previous studies of viruses [41] and ancient DNA [42], we used date-randomisation tests to check that the spread and structure of the sequence ages were sufficient for calibrating estimates of substitution rates (Figure S1). This test involves the random reassignment of the dates to the sequences. If the mean posterior rate estimated from the original data set is included in any of the 95% credibility intervals of the rates estimated from the date-randomised replicates, the sequence ages

Table 1. Details of five intraspecific data sets analysed in this study.

Species	Marker	Length (bp)	Sequences (ancient/modern)	Temporal span (years)	Best-fitting model (BIC)	Molecular clock	Primary source
Human (hg C1)	Mitogenome	15,031	0/184	–	TrN+I+Γ	Strict	–
Hominin[a]	Mitogenome	16,607	61/56	65000	TrN+I+Γ	Strict	[50–52]
Muskox (*Ovibos moschatus*)	D-loop	682	104/16	45740	TVM+I+Γ	Uncorrelated lognormal	[53]
Human influenza A H1N1 virus	*PB2*	2345	115/16	82	TVM+Γ	Uncorrelated lognormal	–
HIV-1	Concatenated genome fragments	1063	157/4	46	GTR+I+Γ	Uncorrelated lognormal	[54]

[a] The hominin data set includes sequences from 56 extant and 55 ancient modern humans (*Homo sapiens*), 5 Neanderthals (*Homo neanderthalensis*), and the Denisovan hominin.
doi:10.1371/journal.pone.0095722.t001

are considered to have insufficient structure and spread for calibrating the molecular clock. For the mitogenomes from human haplogroup C1 (data set i), we assumed an age of 21,700 years (standard deviation 2700 years) for the root of the tree, following the study by Kumar *et al.* [43].

A constant-size coalescent prior was used for each analysis, based on Bayes factors calculated from the marginal likelihoods of each +Γ model. Posterior distributions of parameters, including the tree, were estimated by Markov chain Monte Carlo (MCMC) sampling. MCMC chains were run for at least 2×10^7 steps, with samples drawn every 10^3 steps. After discarding an appropriate proportion of the samples as burn-in, we checked for acceptable sampling, mixing, and convergence to the stationary distribution in each case.

Results

Analysis of rate variation among sites

All data sets showed a negative relationship between the number of sites in a rate category and the number of inferred substitutions per site. The category of zero inferred changes had the highest count for all data sets and, with the exception of the HIV data set, most sites were observed to be constant for each data set (Figure 1). For all data sets, the Poisson distribution provided a poor fit to the site-specific substitution counts ($P \ll 0.001$), with the negative binomial distribution providing a significantly better fit (Table 2). Despite this improvement, neither model provided a good approximation of the pattern of RHAS in the two human mitogenome data sets (i and ii), which were the largest and least variable. The HIV data set (v) features a large proportion of sites inferred to be non-constant, and shows considerable deviation from both the Poisson and negative binomial distributions. There was a particularly large proportion of sites with high inferred numbers of substitutions.

Effect of RHAS model on estimates of likelihoods and parameters

For four out of the five data sets, the marginal likelihood increased with the number of gamma rate categories (Figure 2). In general, 6–10 rate categories provided a good approximation of the asymptotic likelihood value for +Γ models, whereas there was minimal improvement in likelihood when using greater than 10 gamma rate categories. It is noticeable that marginal likelihood is generally higher with +Γ+I models than with +Γ models, and that there is more rapid convergence towards the asymptotic value in +Γ+I models. In the analyses of the four heterochronous data sets (ii–v), neither varying the number of gamma rate categories nor allowing a proportion of invariable sites had any obvious impact on estimates of the coalescence time (root age) or the substitution rate (Figure 3). In almost all cases, the 95% credibility intervals overlapped substantially and the variance in means was <6% of the average 95% CI width (except for the HIV data set). The only exception was the rate estimate for the HIV data set (v), which slightly decreased with an increasing number of gamma rate categories. In the analysis of the mitogenome sequences from human haplogroups C1 (data set i), neither varying the number of gamma categories nor allowing a proportion of invariable sites had any noticeable effect on the estimate of the substitution rate (Figure S2).

The relationships between the α shape parameter or p_{inv} and the number of gamma rate categories were less clear (Figure 2). The general pattern seemed to be a negative relationship between both parameters and the number of gamma rate categories, which can best seen in the declining posterior means and the non-

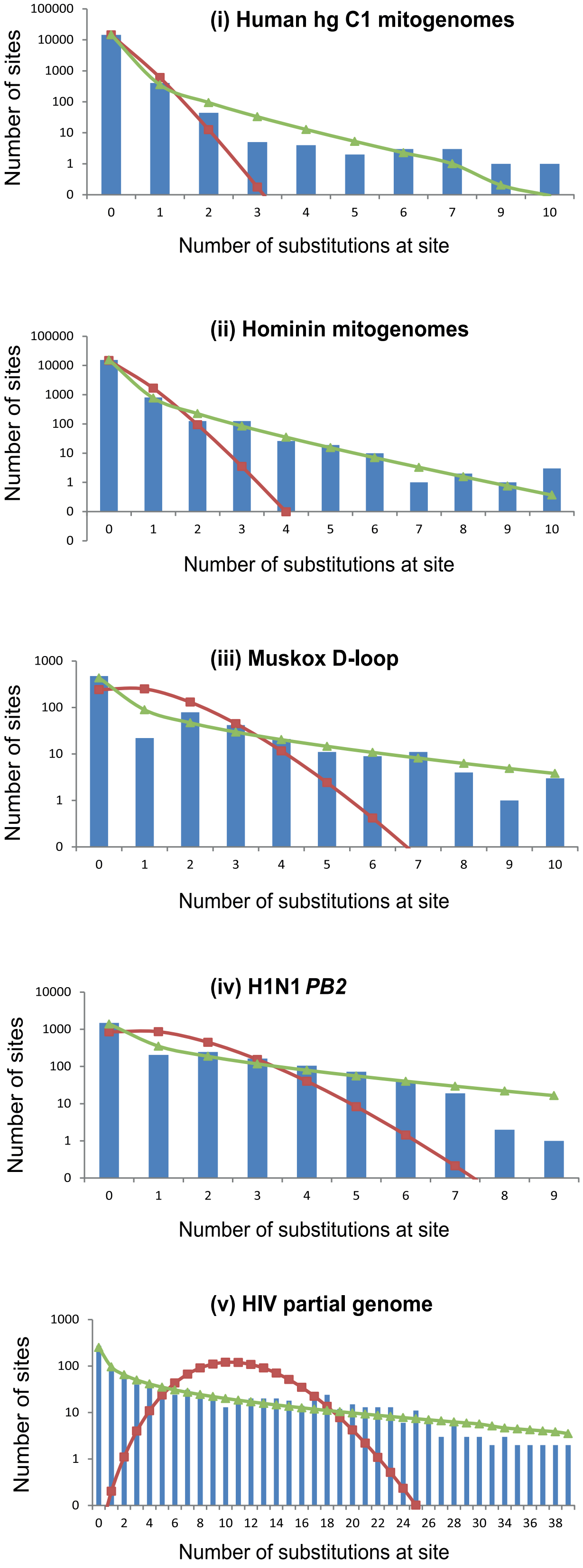

Figure 1. Semi-logarithmic plots of substitutions per site for five intraspecific DNA data sets. Number of substitutions at each site were inferred using parsimony on the Bayesian estimates of the tree topologies. Columns indicate the number of sites against the number of substitutions occurring at each site. Red and green lines indicate the best-fitting Poisson and negative binomial distributions, respectively.

doi:10.1371/journal.pone.0095722.g001

overlapping 95% CIs of estimates based on small and large numbers of gamma categories. However, exceptions to both of these rules were observed in the sequences from human influenza A H1N1 virus (data set iv).

Increasing the number of gamma rate categories led to a linear increase in computation time (Figure 4). The computational costs associated with an increase in the number of gamma categories were similar between +Γ and +Γ+I models. With regard to the topologies of the maximum-clade-credibility trees, we observed that the major clades were unaffected by changes in the number of gamma categories or the inclusion of invariable sites (results not shown). Since this study is limited to data at the population level, which are described by the coalescent process, we did not investigate the impacts of different models on detailed phylogenetic relationships.

Discussion

Our phylogenetic analyses of five intraspecific data sets have provided a number of insights into the performance of RHAS models. Varying the RHAS model, including the number of gamma rate categories or a proportion of invariable sites (p_{inv}), had negligible impacts on our estimates of root age and substitution rate from heterochronous data sets. We found evidence of a complex interplay between α, the number of gamma rate categories, and p_{inv}. When a proportion of the sites were assumed to be invariable, increasing the number of gamma rate categories generally caused a decrease in both α and p_{inv}. This is because the sites that are changing rapidly (mutational hotspots) are preferentially accommodated over less variable sites when there are few rate categories. When there is a limited number of rate categories, this results in a gamma distribution with a higher α. The presence of the invariable-sites parameter in this situation mitigates this bias, but results in an overestimation of p_{inv}.

Our results highlight a trade-off between computational cost and improved accuracy when an increasing number of gamma rate categories are used to model RHAS. The marginal likelihood, which describes the average fit of a model to the data, reaches a plateau as the number of gamma rate categories increases, a result that echoes those of Yang [14]. Using a large number of categories, however, incurs a significant computational cost, increasing both the RAM and time requirements for the likelihood calculations [44]. Our results suggest that using 6–10 rate categories provides a good approximation of the plateau likelihood value when not using invariable-sites models, and that increasing the number of rate categories incurs greater computational cost with minimal benefit. This contradicts suggestions that using a relatively small number of rate categories is insufficient to capture the complexities of the molecular evolutionary process [17].

For population-level analyses that aim to estimate substitution rates or coalescence times, our results suggest that greatly increasing the number of gamma categories typically does not lead to substantial changes in parameter estimates. An exception is when the evolutionary rate is highly variable among sites and deviates strongly from a gamma distribution, as in the case of the HIV data set examined here. In such cases, a higher number of rate categories (8–10) might lead to a modest improvement in estimation accuracy. Surprisingly, the H1N1 influenza virus data did not show a positive relationship between marginal likelihood

Table 2. Fit of Poisson and negative binomial distributions to the site-specific substitution counts in Figureô 1, estimates of α, the ratio of the P values for both distributions, the proportion of constant sites for five data sets and the substitution rate estimate for each data set.

	Poisson	Negative binomial	α_{NB}[a]	ΔAIC	Number of constant sites[b]	Mean substitution rate (site^{-1} year^{-1})
(i) Human hg C1 mitogenomes	4.22×10^{-8}	$\chi^2_5 = 58.63$	0.048	809.7	96.8%	4.22×10^{-8}
	$P = 3.73 \times 10^{-62}$	$P = 2.33 \times 10^{-11}$				
(ii) Hominin mitogenomes	$\chi^2_3 = 481.21$	$\chi^2_6 = 76.44$	0.092	2617.6	93.2%	2.04×10^{-8}
	$P = 5.62 \times 10^{-104}$	$P = 1.94 \times 10^{-14}$				
(iii) Muskox	$\chi^2_5 = 626.76$	$\chi^2_6 = 84.22$	0.24	829.0	69.5%	9.51×10^{-7}
	$P = 3.3 \times 10^{-133}$	$P = 1.89 \times 10^{-15}$				
(iv) H1N1	$\chi^2_5 = 1793.11$	$\chi^2_{10} = 153.95$	0.31	1622.0	63.6%	1.75×10^{-3}
	$P \ll 10^{-300}$	$P = 5.72 \times 10^{-28}$				
(v) HIV	$\chi^2_{21} = 3860.99$	$\chi^2_{38} = 475.95$	0.39	13845.0	27.5%	3.05×10^{-3}
	$P \ll 10^{-300}$	$P = 4.50 \times 10^{-77}$				

[a] This estimate of the shape parameter for gamma-distributed rates among sites was obtained by minimising χ^2.
[b] Sites having <0.5 mutations as inferred by stochastic mutational mapping in SIMMAP.
doi:10.1371/journal.pone.0095722.t002

and the number of rate categories, although the reasons for this are unclear.

The results of our analyses using the invariable-sites model are more complex. The +Γ+I model is intuitively appealing and is, according to the Bayesian information criterion, the preferred model for four of the five data sets analysed (Table 1). However, estimates of parameters in the +Γ+I model are highly sensitive to taxon sampling [5,34] and there is a strong correlation between the proportion of invariable sites and the gamma shape parameter [34]. Our analyses reveal that estimates of the proportion of invariable sites are also highly susceptible to changes in the number of rate categories. Our results corroborate the notion that estimates of both parameters are inevitably biased when the +Γ+I model is used [34].

The interdependence of p_{inv} and α in the +Γ+I model has led some researchers to warn against its use [33,45]. For example, in the justification of their exclusion of +Γ+I models in their analyses, Ren *et al.* [45] contended that one should also consider the

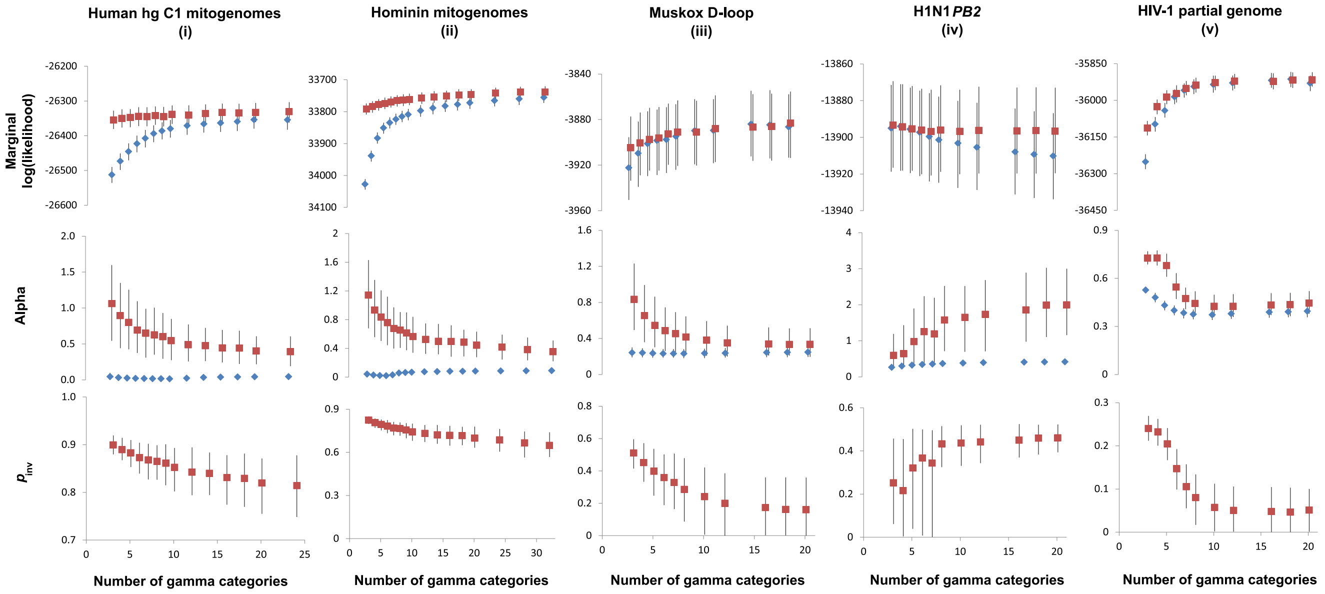

Figure 2. Bayesian phylogenetic estimates of various parameters against number of gamma rate categories for five intraspecific DNA data sets. From top to bottom, rows show estimates of marginal likelihood, the gamma shape parameter (α), and the proportion of invariable sites (p_{inv}). Filled blue and empty red markers represent parameter estimates using +Γ and +Γ+I models, respectively.
doi:10.1371/journal.pone.0095722.g002

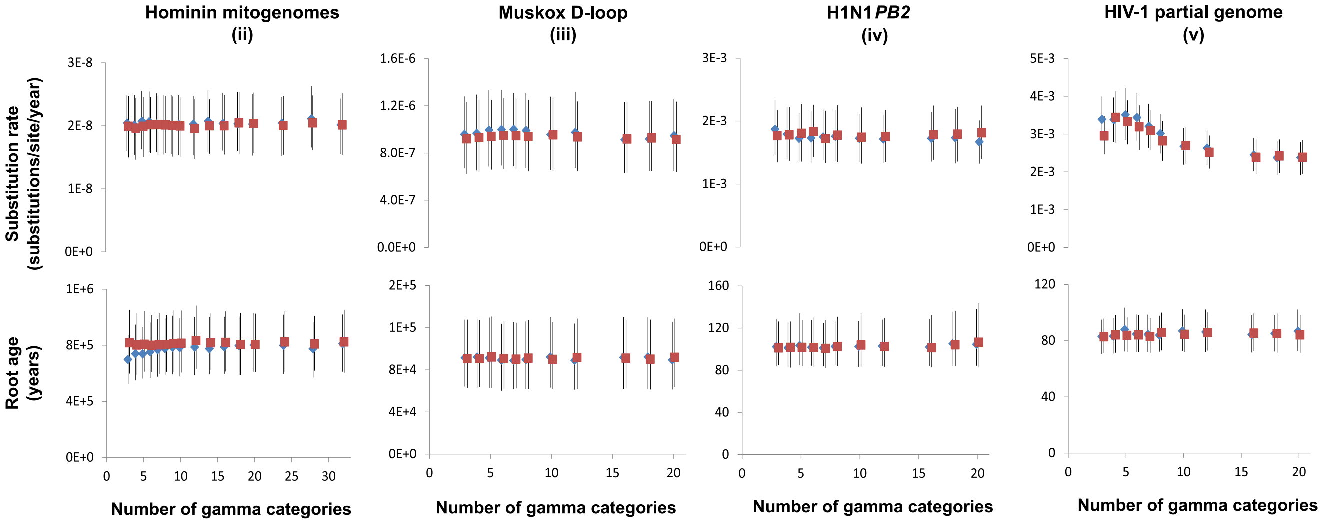

Figure 3. Bayesian phylogenetic estimates of substitution rate and root age against number of gamma rate categories for four intraspecific DNA data sets. Filled blue and empty red markers represent parameter estimates using +Γ and +Γ+I models, respectively.
doi:10.1371/journal.pone.0095722.g003

"biological interpretations of the models and the robustness of analysis to model assumptions" (p. 815), not just the fit of the model to the data. We concur with this viewpoint, considering that the invariable-sites assumption is particularly troublesome for intraspecific data. Here we would expect a +Γ+I model to perform well for data sets that show a bimodal distribution in site-specific rates (one peak at 0 and one peak at >0). However, when the aim is to estimate the substitution rate or coalescence times for a population-level data set, we found that allowing a proportion of invariable sites did not alter the results substantially. There was, however, a small computational benefit associated with +Γ+I models, since they generally outperformed +Γ models in marginal likelihood at small numbers of gamma categories, reducing the need for higher numbers of categories in +Γ models. In some instances (data set i and ii), the asymptotic marginal likelihood for +Γ+I models was slightly higher than the marginal likelihood for +Γ models.

Using a gamma distribution to model RHAS has deservedly been popular, owing to its good fit and mathematical simplicity [46]. There is, however, no reason to believe that the distribution of rates among sites actually follows the gamma distribution. Indeed, we observe that the gamma model still does not provide an accurate picture of RHAS, especially for the least variable data sets that we examined (i and ii). Some studies have explored the possibility of using alternative approaches to model RHAS. Notably, the discrete-rates CAT model, implemented in RAxML [47] and FastTree2 [44], has been shown to be computationally more efficient than the traditional +Γ model and yields tree topologies with improved likelihood values [48]. Recently, Wu *et al.* [49] proposed a Bayesian method of automatic model selection that simultaneously estimates the substitution model and rate at each site. The performance of these parameter-rich models, with regard to phylogenetic analyses of intraspecific sequence data, warrants further study.

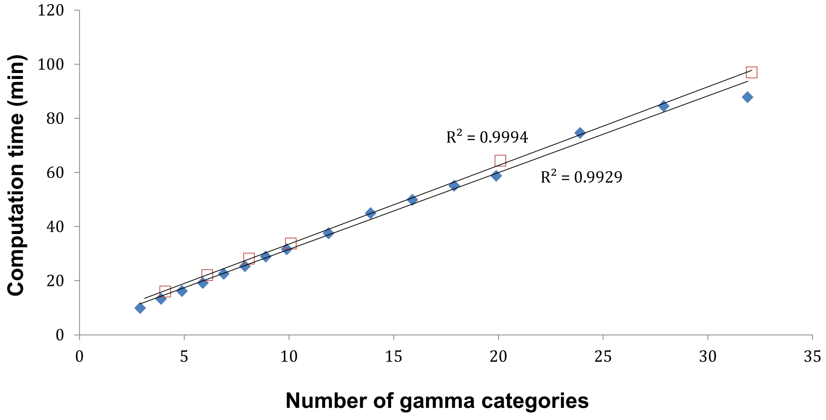

Figure 4. Computation time (min) as a function of the number of gamma rate categories, for the hominin mitogenome data set (ii). Filled blue and empty red markers represent computation time using TrN+Γ and TrN+I+Γ models, respectively. Each Markov chain was run for 10^6 steps on a six-core processor (Intel Xeon W3690).
doi:10.1371/journal.pone.0095722.g004

Conclusions

The inclusion of a parameter for the proportion of invariable sites is a legacy of studies conducted at the interspecific level. At the intraspecific level, estimating the proportion of invariable sites is primarily a statistical measure that does not have much biological meaning. Here we suggest that most intraspecific studies of substitution rates or coalescence times can use fewer than 10 gamma rate categories, in order to achieve a balance between model complexity, computational efficiency, and parameter estimation. The results of our study apply to population-level data, but are probably relevant to data sets containing sequences from closely related species. Further studies of rate variation in interspecific data sets will provide additional insights into the performance of RHAS models.

Supporting Information

Figure S1 Results of the date-randomisation test for temporal signal in heterochronous data. The plot shows the rate estimates and their 95% credibility intervals for the unrandomised data (empty markers)_and date-randomised replicates (filled markers). The test was conducted on the four heterochronous data sets, each with 20 replicates.
(EPS)

Figure S2 Estimates of substitution rates as a function of the number of gamma rate categories, for the isochronous human hg C1 mitogenome data set (i). Filled blue and empty red markers represent rate estimates using +Γ and +Γ+I models, respectively.
(EPS)

File S1 Alignments for all datasets used in this study.
(ZIP)

Acknowledgments

We would like to thank three anonymous reviewers for their helpful comments.

Author Contributions

Conceived and designed the experiments: FJ SYWH. Performed the experiments: FJ. Analyzed the data: FJ. Wrote the paper: FJ NL SYWH.

References

1. Tavaré S (1986) Some probabilistic and statistical problems in the analysis of DNA sequences. Lectures Math Life Sci (Amer Math Soc) 17: 57–86.
2. Jukes TH, Cantor CR (1969) Evolution of protein molecules. In: Munro HN, editor. Mammalian protein metabolism. New York: Academic Press. pp. 21–123.
3. Kimura M (1980) A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. J Mol Evol 16: 111–120.
4. Hasegawa M, Kishino H, Yano T (1985) Dating of the human-ape splitting by a molecular clock of mitochondrial DNA. J Mol Evol 22: 160–174.
5. Yang Z (1996) Among-site rate variation and its impact on phylogenetic analyses. Trends Ecol Evol 11: 367–372.
6. Simon C, Nigro L, Sullivan J, Holsinger K, Martin A, et al. (1996) Large differences in substitutional pattern and evolutionary rate of 12S ribosomal RNA genes. Mol Biol Evol 13: 923–932.
7. Simon C, Buckley TR, Frati F, Stewart JB, Beckenbach AT (2006) Incorporating molecular evolution into phylogenetic analysis, and a new compilation of conserved polymerase chain reaction primers for animal mitochondrial DNA. Annu Rev Ecol Evol Syst 37: 545–579.
8. Hodgkinson A, Eyre-Walker A (2011) Variation in the mutation rate across mammalian genomes. Nat Rev Genet 12: 756–766.
9. Wakeley J (1993) Substitution rate variation among sites in hypervariable region-1 of human mitochondrial-DNA. J Mol Evol 37: 613–623.
10. Tateno Y, Takezaki N, Nei M (1994) Relative efficiencies of the maximum-likelihood, neighbor-joining, and maximum-parsimony methods when substitution rate varies with site. Mol Biol Evol 11: 261–277.
11. Buckley TR, Simon C, Chambers GK (2001) Exploring among-site rate variation models in a maximum likelihood framework using empirical data: Effects of model assumptions on estimates of topology, branch lengths, and bootstrap support. Syst Biol 50: 67–86.
12. Sullivan J, Swofford DL (2001) Should we use model-based methods for phylogenetic inference when we know that assumptions about among-site rate variation and nucleotide substitution pattern are violated? Syst Biol 50: 723–729.
13. Soubrier J, Steel M, Lee MSY, Sarkissian CD, Guindon S, et al. (2012) The influence of rate heterogeneity among sites on the time dependence of molecular rates. Mol Biol Evol 29: 3345–3358.
14. Yang Z (1994) Maximum likelihood phylogenetic estimation from DNA sequences with variable rates over sites: approximate methods. J Mol Evol 39: 306–314.
15. Yang Z (1993) Maximum-likelihood estimation of phylogeny from DNA sequences when substitution rates differ over sites. Mol Biol Evol 10: 1396–1401.
16. Jarman SN, Elliott NG (2000) DNA evidence for morphological and cryptic Cenozoic speciations in the Anaspididae, 'living fossils' from the Triassic. J Evol Biol 13: 624–633.
17. Soares P, Ermini L, Thomson N, Mormina M, Rito T, et al. (2009) Correcting for purifying selection: an improved human mitochondrial molecular clock. Am J Hum Genet 84: 740–759.
18. de St Pierre M, Gandini F, Perego UA, Bodner M, Gomez-Carballa A, et al. (2012) Arrival of Paleo-Indians to the southern cone of South America: new clues from mitogenomes. PLOS ONE 7.
19. Mayrose I, Friedman N, Pupko T (2005) A Gamma mixture model better accounts for among site rate heterogeneity. Bioinformatics 21: 151-158.
20. Fitch WM, Margoliash E (1967) A method for estimating the number of invariant amino acid coding positions in a gene using cytochrome C as a model case. Biochem Genet 1: 65–71.
21. Fitch WM (1986) An estimation of the number of invariable sites is necessary for the accurate estimation of the number of nucleotide substitutions since a common ancestor. In: Gershowitz H, Rucknagel DL, Tashian RE, editors. Evolutionary perspectives and the new genetics. New York: Alan R. Liss, Inc. pp. 149–159.
22. Palumbi SR (1989) Rates of molecular evolution and the fraction of nucleotide positions free to vary. J Mol Evol 29: 180–187.
23. Shoemaker JS, Fitch WM (1989) Evidence from nuclear sequences that invariable sites should be considered when sequence divergence is calculated. Mol Biol Evol 6: 270–289.
24. Tourasse NJ, Gouy M (1997) Evolutionary distances between nucleotide sequences based on the distribution of substitution rates among sites as estimated by parsimony. Mol Biol Evol 14: 287–298.
25. Gu X, Fu YX, Li WH (1995) Maximum likelihood estimation of the heterogeneity of substitution rate among nucleotide sites. Mol Biol Evol 12: 546–557.
26. Waddell PJ, Steel MA (1997) General time-reversible distances with unequal rates across sites: mixing gamma and inverse Gaussian distributions with invariant sites. Mol Phylogenet Evol 8: 398–414.
27. Akaike H (1981) A new look at the statistical model identification. IEEE Trans Autom Control: 716–723.
28. Schwartz G (1978) Estimating the dimension of a model. Ann Stat 6: 461–464.
29. Luo A, Qiao HJ, Zhang YZ, Shi WF, Ho SYW, et al. (2010) Performance of criteria for selecting evolutionary models in phylogenetics: a comprehensive study based on simulated datasets. BMC Evol Biol 10.
30. Keane TM, Creevey CJ, Pentony MM, Naughton TJ, Mclnerney JO (2006) Assessment of methods for amino acid matrix selection and their use on empirical data shows that *ad hoc* assumptions for choice of matrix are not justified. BMC Evol Biol 6: 29.
31. Haag-Liautard C, Dorris M, Maside X, Macaskill S, Halligan DL, et al. (2007) Direct estimation of per nucleotide and genomic deleterious mutation rates in *Drosophila*. Nature 445: 82–85.
32. Agrawal AF, Whitlock MC (2012) Mutation load: the fitness of individuals in populations where deleterious alleles are abundant. Annu Rev Ecol Evol Syst 43: 115–135.
33. Yang Z (2006) Computational Molecular Evolution. New York: Oxford University Press.
34. Sullivan J, Swofford DL, Naylor GJP (1999) The effect of taxon sampling on estimating rate heterogeneity parameters of maximum-likelihood models. Mol Biol Evol 16: 1347–1356.
35. Edgar RC (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res 32: 1792–1797.
36. Drummond AJ, Suchard MA, Xie D, Rambaut A (2012) Bayesian phylogenetics with BEAUti and the BEAST 1.7. Mol Biol Evol 29: 1969–1973.
37. Drummond AJ, Ho SYW, Phillips MJ, Rambaut A (2006) Relaxed phylogenetics and dating with confidence. PLOS Biol 4: 699–710.
38. Suchard MA, Weiss RE, Sinsheimer JS (2001) Bayesian selection of continuous-time Markov chain evolutionary models. Mol Biol Evol 18: 1001–1013.
39. Nielsen R (2002) Mapping mutations on phylogenies. Syst Biol 51: 729–739.

40. Bollback JP (2006) SIMMAP: stochastic character mapping of discrete traits on phylogenies. BMC Bioinformatics 7: 88.
41. Ramsden C, Holmes EC, Charleston MA (2009) Hantavirus evolution in relation to its rodent and insectivore hosts: no evidence for codivergence. Mol Biol Evol 26: 143–153.
42. Ho SYW, Lanfear R, Phillips MJ, Barnes I, Thomas JA, et al. (2011) Bayesian estimation of substitution rates from ancient DNA sequences with low information content. Syst Biol 60: 366–374.
43. Kumar S, Bellis C, Zlojutro M, Melton PE, Blangero J, et al. (2011) Large scale mitochondrial sequencing in Mexican Americans suggests a reappraisal of Native American origins. BMC Evol Biol 11: 293.
44. Price MN, Dehal PS, Arkin AP (2010) FastTree 2 – approximately maximum-likelihood trees for large alignments. PLOS ONE 5: e9490.
45. 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.
46. Felsenstein J (2001) Taking variation of evolutionary rates between sites into account in inferring phylogenies. J Mol Evol 53: 447–455.
47. Stamatakis A, Ludwig T, Meier H (2005) RAxML-III: a fast program for maximum likelihood-based inference of large phylogenetic trees. Bioinformatics 21: 456–463.
48. Stamatakis A (2006) Phylogenetic models of rate heterogeneity: a high performance computing perspective. Proceedings of the 20th IEEE International Parallel & Distributed Processing Symposium (IPDPS2006). Washington: IEEE Computer Society Press. pp. 278–286.
49. Wu CH, Suchard MA, Drummond AJ (2013) Bayesian selection of nucleotide substitution models and their site assignments. Mol Biol Evol 30: 669–688.
50. Ingman M, Kaessmann H, Paabo S, Gyllensten U (2000) Mitochondrial genome variation and the origin of modern humans. Nature 408: 708–713.
51. Brotherton P, Haak W, Templeton J, Brandt G, Soubrier J, et al. (2013) Neolithic mitochondrial haplogroup H genomes and the genetic origins of Europeans. Nat Commun 4: 1764.
52. Fu QM, Mittnik A, Johnson PLF, Bos K, Lari M, et al. (2013) A revised timescale for human evolution based on ancient mitochondrial genomes. Curr Biol 23: 553–559.
53. De MacPhee R, Tikhonov AN, Mol D, Greenwood AD (2005) Late Quaternary loss of genetic diversity in muskox (*Ovibos*). BMC Evol Biol 5: 49.
54. Worobey M, Gemmel M, Teuwen DE, Haselkorn T, Kunstman K, et al. (2008) Direct evidence of extensive diversity of HIV-1 in Kinshasa by 1960. Nature 455: 661–664.