ORIGINAL RESEARCH

WILEY Ecology and Evolution Open Access

Incomplete lineage sorting impacts the inference of macroevolutionary regimes from molecular phylogenies when concatenation is employed: An analysis based on Cetacea

Anieli G. Pereira | Carlos G. Schrago [iD]

Department of Genetics, Federal University of Rio de Janeiro, Rio de Janeiro, Brazil

Correspondence
Carlos G. Schrago, Instituto de Biologia, Departamento de Genética, Universidade Federal do Rio de Janeiro, CCS, A2-092, Rua Prof. Rodolpho Paulo Rocco, SN, Cidade Universitária, Rio de Janeiro, RJ, CEP: 21941-617 Brazil.
Email: carlos.schrago@gmail.com

Abstract

Interest in methods that estimate speciation and extinction rates from molecular phylogenies has increased over the last decade. The application of such methods requires reliable estimates of tree topology and node ages, which are frequently obtained using standard phylogenetic inference combining concatenated loci and molecular dating. However, this practice disregards population-level processes that generate gene tree/species tree discordance. We evaluated the impact of employing concatenation and coalescent-based phylogeny inference in recovering the correct macroevolutionary regime using simulated data based on the well-established diversification rate shift of delphinids in Cetacea. We found that under scenarios of strong incomplete lineage sorting, macroevolutionary analysis of phylogenies inferred by concatenating loci failed to recover the delphinid diversification shift, while the coalescent-based tree consistently retrieved the correct rate regime. We suggest that ignoring microevolutionary processes reduces the power of methods that estimate macroevolutionary regimes from molecular data.

KEYWORDS
concatenation, diversification, multispecies coalescent, simulation

1 | INTRODUCTION

Evolutionary studies are recently abounding with analyses that use molecular phylogenies to investigate macroevolutionary problems related to lineage diversification rates. The number of methods proposed to pursue this task has increased significantly, all relying on time-calibrated phylogenetic trees with extant species, adopting both likelihood and Bayesian frameworks to estimate the number, magnitude, and locations of diversification rate shifts (Alfaro et al., 2009; Rabosky, 2014; Rabosky & Huang, 2016; Rabosky et al., 2014). In parallel, over the last few years, the inference of molecular phylogenies itself has been transformed at its very core by the development of the multispecies coalescent theory (Edwards, 2009;

Xu & Yang, 2016), which explicitly separates the statistical process that models the species tree (speciation) from the process that models gene genealogies in such trees (multispecies coalescent). These two novel theoretical developments, however, are rarely considered jointly (Degnan & Rosenberg, 2009; Song, Liu, Edwards, & Wu, 2012). This is problematic because the estimation of macroevolutionary speciation rates should be carried out using the species tree instead of gene trees, which is the standard practice (Song et al., 2012).

Theoretically, employing gene trees rather than the species tree to estimate macroevolutionary parameters may lead to significant bias; especially in shallow divergences, where incomplete lineage sorting (ILS) has a relatively larger impact on species divergence

times (Angelis & Dos Reis, 2015; Leaché, Harris, Rannala, & Yang, 2014). This issue is not ameliorated using the tree built from concatenating genes into a supermatrix because even if the phylogeny is correctly determined, the biological meaning of branch lengths in such trees is elusive, as the age of nodes do not mirror speciation times (Angelis & Dos Reis, 2015; Edwards & Beerli, 2000). Population-level phenomena, such as ILS, gene duplication and loss, and hybridization will increase the probability of gene tree/species tree mismatch (Degnan & Rosenberg, 2009; Liu, Yu, Pearl, & Edwards, 2009; Schrago, Menezes, Furtado, Bonvicino, & Seuanez, 2014; Song et al., 2012; Tonini, Moore, Stern, Shcheglovitova, & Ortí, 2015).

Even though, the standard practice when studying lineage diversification rates is to rely on molecular phylogenies inferred from concatenated loci (Arbour & Santana, 2017; Shi & Rabosky, 2015). Concatenation ignores differences in evolutionary histories between loci, which are useful to inform on ancestral population dynamics, allowing the decouple of the processes that generate the phylogeny and gene trees (Liu, Yu, Kubatko, Pearl, & Edwards, 2009). Moreover, phylogeny inference from concatenated genomic regions without accounting for conflicting histories is a case of model misspecification, which has been shown to impact estimates of macroevolutionary parameters (Revell, Harmon, Glor, & Linder, 2005). Previous studies demonstrated that concatenating sequences may lead to biased estimates of the species phylogeny (Edwards et al., 2016; Liu, Xi, & Davis, 2015). Accounting for variation in gene genealogies in a population-level framework while estimating phylogenies is computationally demanding, and heuristic approaches were devised. However, the performance of such approaches has been questioned (Gatesy & Springer, 2013), and further investigation is required. Notwithstanding, it is desirable that biologists should account for population genetics principles, modeled by the multispecies coalescent process, while estimating species trees, and associated macroevolutionary diversification parameters, from biological sequences. To date, the consequences of adopting gene trees rather than the species tree in macroevolutionary inference have not been thoroughly examined.

This prompted us to evaluate the impact of incomplete lineage sorting on the inference of macroevolutionary regimes. To do so, we employed the widely used Cetacea phylogeny of Steeman et al. (2009), which is regarded as an exemplary case of a tree topology containing an evolutionary radiation (family Delphinidae). We investigated the performance of phylogenies obtained by both concatenation and the multispecies coalescence in recovering such a macroevolutionary regime.

2 | MATERIALS AND METHODS

2.1 | Simulating gene trees and alignments

To investigate the effects of incomplete lineage sorting on the inference of macroevolutionary parameters, we simulated gene trees, as well as their respective sequence alignments, using the time-dated

phylogeny (used here as the species tree) of cetaceans provided by Steeman et al. (2009), hereafter S09, which encompasses 98% of the extant cetacean diversity. This phylogeny was employed because it has been repeatedly used as a study case in recent analyses, where a shift of macroevolutionary regime for higher diversification rates during the early diversification of dolphins was unambiguously recovered (Rabosky, 2014). Although the S09 time-dated tree was not originally inferred using coalescent-based methods; we used it as a template to investigate macroevolutionary estimation because the occurrence of a diversification rate shift in stem Delphinidae has not been questioned so far.

Gene trees were simulated under the multispecies coalescent model using four increasing ancestral effective population sizes: $N_e = 10^4$, 10^5, 10^6, and 10^7 Wright-Fisher individuals, which resulted in four independent datasets. For each population size dataset, we composed 100 replicates containing 15 bifurcating gene trees that were evolved employing HYBRID-LAMBDA software (Zhu, Degnan, Goldstein, & Eldon, 2015). Branch lengths in simulated gene trees, which were originally in coalescent units ($2N_e$ generations), were transformed into mutation units using the genomic rate of 10^{-8} mutations/site/generation and assuming generation time = 10 years, which is close to the mean generation time of delphinids, 9.6 years (Pacifici et al., 2013). Mutation rate per generation was based on the human estimates obtained from genome-wide studies (Kong et al., 2012; Lipson et al., 2015). Gene trees in mutation units were then used to simulate gene alignments with the EVOLVER program of the PAML package (Yang, 2007) under the Jukes-Cantor substitution model. Alignment lengths of each gene (locus) were sampled from a uniform distribution from 160 to 2306 bp, which corresponds to the minimum and maximum lengths of genes used in Steeman et al. (2009). Therefore, for each N_e value, we obtained 1,500 independent alignments. A summary of the entire simulation procedure is found in Figure 1.

2.2 | Phylogenies and divergence times

To mimic the procedure carried out by researchers when inferring macroevolutionary regimes, for each population size dataset, phylogenies and divergence times were estimated from simulated alignments. When inferring phylogenies, we used two competing approaches commonly implemented in recent literature: concatenation and species tree inference based the multispecies coalescent (MSC) theory. For both approaches, maximum likelihood trees were estimated in PhyML 3 (Guindon et al., 2010) under the Jukes-Cantor model. Because the coalescent program STEM requires ultrametric trees, and as sequence data were simulated under a single substitution rate, branch lengths in PhyML trees were re-estimated with the program BASEML of the package PAML 4, enforcing the molecular clock (Yang, 1997, 2007). Divergence time estimation was conducted in treePL (Smith & O'Meara, 2012). The algorithm was run without cross-validation and setting the smooth parameter = 0.1. Five calibrations were used according to the fossils and ages employed in S09, plus a maximum limit of
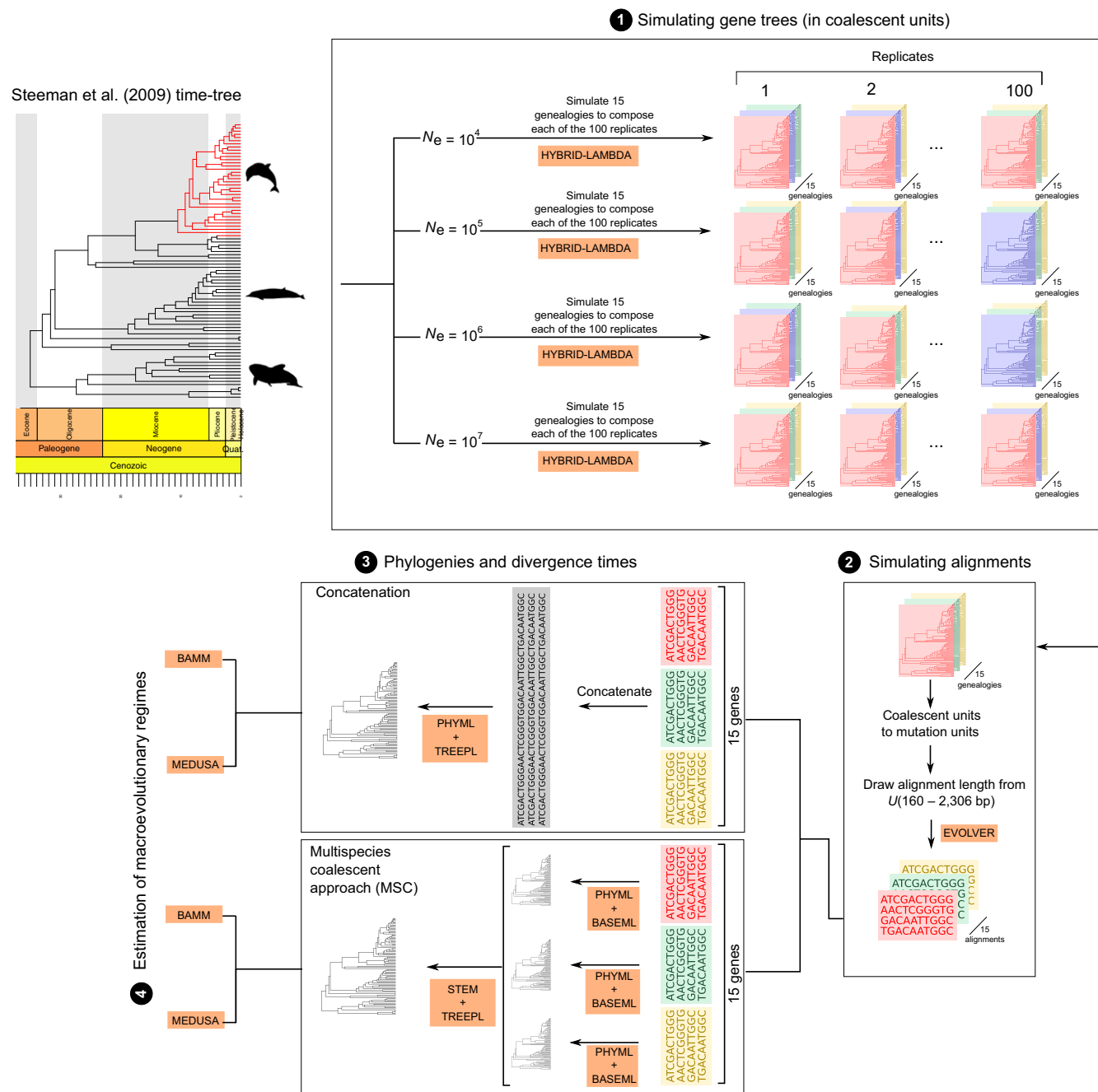
FIGURE 1 Summary of the simulation steps used to generate data for evaluating the impact of population processes on the estimation of diversification rates and shifts. Steeman et al. (2009) time-tree, which contains a diversification shift near the diversification of delphinids, was used as a template to generate gene genealogies under four different population size scenarios. A total of 100 replicates, with 15 gene genealogies each, was produced (1). Alignments were simulated after transforming branch lengths of gene trees from coalescent units to mutation units and by sampling an alignment length from an empirically-driven uniform distribution (2). Phylogenies and divergence times were then estimated using simulated alignments by employing both concatenation and a MSC approach (3). The inferred dated phylogenies were finally used to estimate the shift of macroevolutionary regime (4)

40 Mya to the age of crown Cetacea based on Chen, Xu, Zhou, and Yang (2011) (Table 1). Species tree estimation under the MSC model was implemented in STEM (Kubatko, Carstens, & Knowles, 2009) by fixing θ values from 0.0004 ($N_e = 10^4$) to 0.4 ($N_e = 10^7$), respectively. As means of evaluating the accuracy and precision of inferred trees, for each replicate, we calculated the topological distance from the estimated trees to the S09 phylogeny using

the Penny and Hendy (1985) metric implemented in the R package 'ape' (Paradis, Claude, & Strimmer, 2004) (Figure 1).

2.3 | Estimation of macroevolutionary regimes

The dynamics of species diversification was inferred for all dated phylogenies obtained from concatenation and the MSC method

using the widely employed software in the literature, namely BAMM (Rabosky, 2014) and MEDUSA (Alfaro et al., 2009; Pennell et al., 2014). BAMM accounts for the variation in diversification rates through time and among lineages using transdimensional (reversible-jump) Markov chain Monte Carlo (rjMCMC) (Rabosky, 2014). In BAMM, Markov chains were sampled every 1,000th generation until 7,500 trees were collected after a burn-in of 25%. Prior distributions were set according to the *setBAMMPriors* function from the BAMMtools R package (Rabosky et al., 2014). MEDUSA assumes constant diversification rates through time to detect rate variations

in lineages based on a maximum likelihood approach. We used the birth-death (BD) model to run MEDUSA analyses.

When evaluating the performance of trees estimated using either concatenation or the MSC-based method in recovering the same macroevolutionary regime inferred for the S09 tree (Steeman et al., 2009), we measured (a) the number of shifts inferred on the tree (BAMM and MEDUSA); (b) the marginal posterior probability for 0, 1 and 2-diversification rate shifts regimes in BAMM; and (c) the age of the inferred diversification shift (BAMM and MEDUSA). Item (iii) was evaluated because, in addition to correctly inferring the number of diversification shifts, we expect the age in which the change of macroevolutionary regime took place to be correctly recovered.

TABLE 1 Fossil calibrations used in divergence time estimation

Divergence	Age (Ma)
B. mysticeti and *D. delphis*	33.3–40 Ma
B. physalus and *M. novaeanglia*	>7.3 Ma
L. vexillifer and *D. delphis*	>23.5 Ma
I. geoffrensis and *P. blainvillei*	>12 Ma
D. leucas and *P. phocoena*	>10 Ma

3 | RESULTS AND DISCUSSION

We found that under smaller effective population sizes (10^4 and 10^5), most macroevolutionary analyses of BAMM and MEDUSA correctly recovered one diversification rate shift, independent of the tree estimation method. As population sizes increased, however,
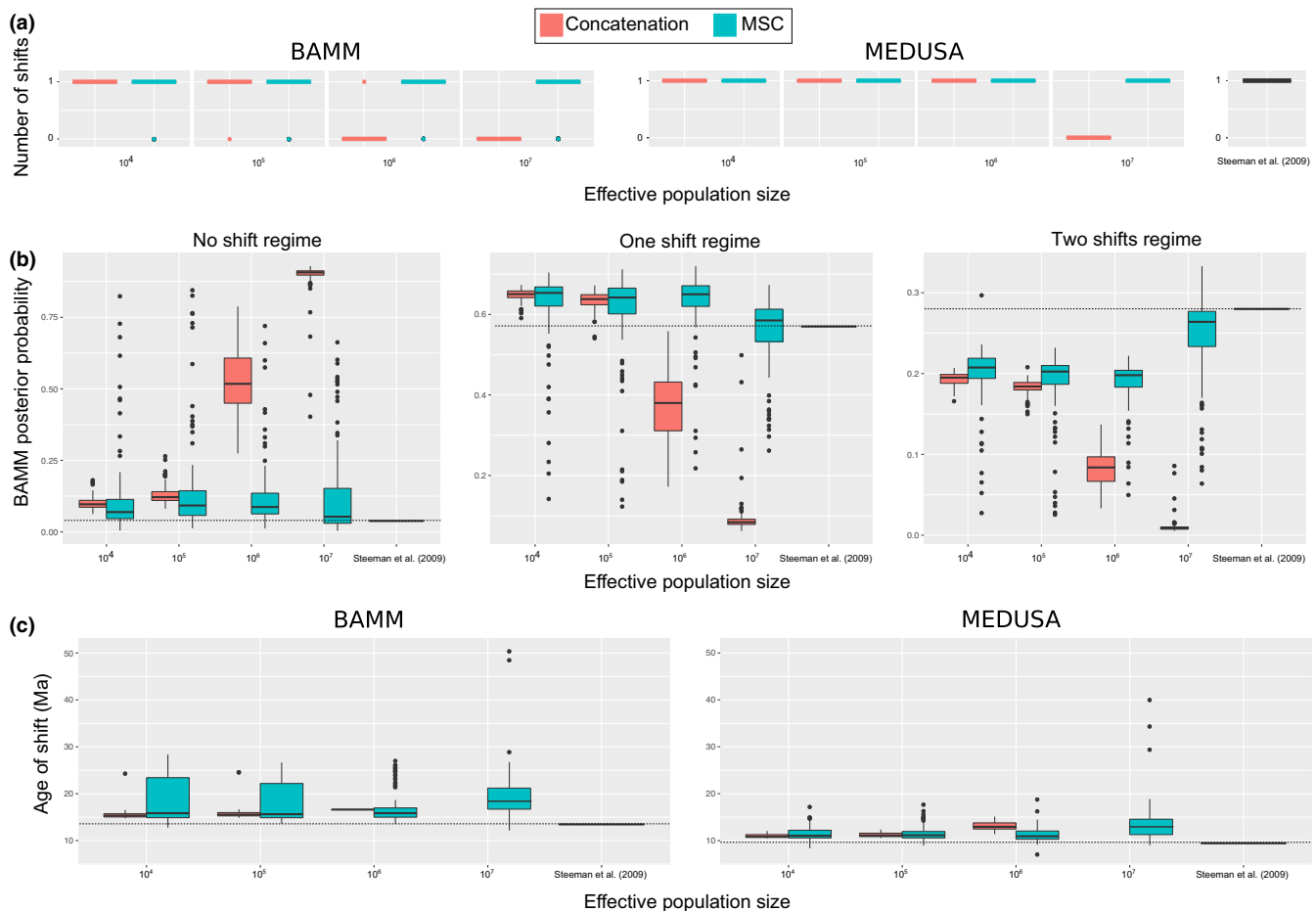

FIGURE 2 Impact of the concatenation and MSC methods in the inference of macroevolutionary rates. (a) The number of shifts inferred; (b) the posterior probability for 0-, 1- and 2-shift configurations in simulated datasets estimated in BAMM; and (c) ages of the inferred diversification rate shifts. In all panels, results obtained using the original Steeman et al. (2009) tree is also displayed for the sake of comparison

only analyses conducted using the MSC-based trees recovered a rate shift in BAMM and MEDUSA. For instance, when $N_e = 10^7$, phylogenies estimated by concatenation were unable to inform on the presence of a shift of macroevolutionary regime (0%) (Figure 2a). In BAMM, we draw the marginal posterior densities for macroevolutionary regimes with a varying number of shifts. The posterior probability of no diversification shift was low under smaller N_e (10^4 and 10^5), which correctly captured the empirical scenario of a 1-shift (Figure 2b). Under large N_e, however, only datasets in which the MSC-based trees were used rendered a low probability of a 0-shift. Trees from concatenated alignments consistently yielded >70% probability for the incorrect 0-shift configuration. Posterior densities for the regime with one diversification rate shift indicated that concatenation failed to recover the cetacean macroevolutionary regime under $N_e = 10^6$ and 10^7, whereas posteriors for the 1-shift configuration from the MSC-based trees rendered a higher probability even under large effective population sizes (Figure 2b).

Under $N_e = 10^4$ and 10^5, the age of the Cetacea diversification rate shift was correctly inferred independent of the tree building approach (Figure 2c). When MSC trees were employed, however, a greater variance of the age was retrieved. Theory predicts that the mean coalescent time, that is, genetic divergence between two lineages, will increase with increasing population size because the number of substitutions accumulated in the ancestral species equals the scaled population parameter $\theta = 4N_e\mu$. MSC methods thus differentiate between the mean coalescent time (genetic divergence) and the speciation time *per se* (reproductive isolation) of lineages. Therefore, as expected, when $N_e = 10^6$, the ages the delphinid macroevolutionary regime shift inferred from trees estimated using concatenation were older when compared to MSC-based trees. When $N_e = 10^7$, the comparison could not be performed, as only analyses using the MSC-based trees recovered a shift. Differences between the ages of shifts inferred by BAMM and MEDUSA might be explained by the fact that MEDUSA assigns the location of diversification rate shifts to nodes rather than along branches (Alfaro et al., 2009; Rabosky, 2014).

Ultimately, the inference of macroevolutionary regimes from molecular phylogenies evidently depends on the correct estimation of both topology and time-transformed branch lengths. In our study, as the effective population size increased, topological distances between the phylogenies inferred assuming the MSC trees and S09 tree were smaller than those calculated using concatenation (Figure 3). This corroborates the findings that the MSC-based methods are more accurate in recovering the true phylogeny under strong ILS caused by large effective population sizes (Edwards, Liu, & Pearl, 2007; Kubatko & Degnan, 2007; Kubatko et al., 2009).

Studies have highlighted the difficulty of estimating macroevolutionary parameters from molecular phylogenies (Quental & Marshall, 2010). Moreover, recent works suggested that the methodological approaches implemented in both BAMM and MEDUSA may be flawed (May & Moore, 2016; Moore, Höhna, May, Rannala, & Huelsenbeck, 2016). Although methodological improvements were already implemented (Rabosky, Mitchell, & Chang, 2017), it is
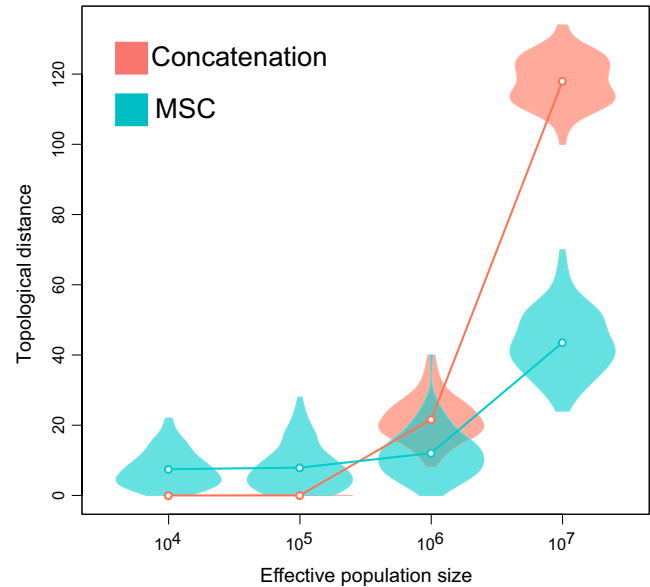

FIGURE 3 Association between topological distances of the inferred tree using concatenation or MSC and the effective population size. Violin plots depict the distribution of the topological distances of the phylogenies estimated from each tree-building method across the 100 replicates. Circles indicate the mean topological distance for replicates in each population size scenario

conceivable that dated molecular phylogenies alone are insufficient to estimate the shifts of macroevolutionary regimes reliably. In this sense, our study is an additional alert that the methodological framework used to infer the molecular phylogeny itself should account for population-level processes, such as incomplete lineage sorting. Future investigation of the impact of molecular phylogenies on estimating lineage diversification rates should be ideally conducted under the multispecies coalescent model.

Using the cetacean diversification as a template, our simulations showed that when the effective population size reached 10^6, phylogenies inferred from concatenation started to differ from the template cetacean tree, whereas coalescent-based approaches performed better in recovering the template topology. Thus, it is not surprising that under this N_e value, the inference of the diversification rate regime from coalescent-based trees outperformed concatenation in most simulations. Evidently, we should evaluate the biological relevance of such findings by contrasting our results with empirically estimated effective population sizes. Oliver (2013) compiled a list of effective population size estimates to show that microevolutionary processes may affect deep-time phylogenetic relationships. Effective population sizes as large as 10^6 have been frequently recovered for invertebrates (Papadopoulos, Peijnenburg, & Luttikhuizen, 2005), plants (McDaniel et al., 2010), and even vertebrates (Carneiro, Ferrand, & Nachman, 2009). Although rarer, N_e values greater than 10^7 were indeed reported for mollusks (7.2×10^7) (Etter et al., 2011), insects (~2.0×10^7) (Schoville, Stuckey, & Roderick, 2011) and birds (1.6×10^7) (Maley & Winker, 2010). Our results demonstrated that diversification

rate inference in such lineages must rely on coalescent-based tree inference. In such cases, the use of trees inferred from concatenated genes will increase Type II error because the probability of identifying a diversification rate shift decreases significantly. In the case of cetaceans, N_e's varied from values near 10^4 (e.g., 1,218 for bottlenose dolphins, (Viaud-Martinez, Brownell, Komnenou, & Bohonak, 2008) to 214,629 for harbor porpoises (Viaud-Martínez et al., 2007). Therefore, the range of N_e values used in our simulations included the values measured for cetaceans. Our simulations showed that, within this range, phylogenies inferred from both concatenation and coalescent-based approaches retain the information required to correctly estimate macroevolutionary regime shifts.

We suggest that further developments in macroevolutionary rate estimation should not overlook the effects of microevolutionary processes on phylogenies. Although our simulations tried to capture the intricacies of empirical data manipulation, they were still based on simplistic assumptions, such as a constant effective population size along the phylogeny. Moreover, it would be meaningful to quantify the effects of tree topology shapes that comprise differing mixtures of deep and shallow divergences. If MSC is disregarded, topologies containing both intra- and inter-species diversity are difficult to model, as they contain both the coalescent and speciation processes.

ACKNOWLEDGMENTS

AUTHORS CONTRIBUTION

AGP and CGS conceived and designed the study, performed all analyses, interpreted the results and wrote the manuscript.

DATA ACCESSIBILITY STATEMENT

All phylogenies used in this study were previously published and original sources were referenced in the text. Data generated by simulation are archived in the Dryad repository: https://doi.org/10.5061/dryad.np7332q

ORCID

Carlos G. Schrago [ID] http://orcid.org/0000-0001-6257-8906

REFERENCES

Alfaro, M. E., Santini, F., Brock, C., Alamillo, H., Dornburg, A., Rabosky, D. L., … Harmon, L. J. (2009). Nine exceptional radiations plus high turnover explain species diversity in jawed vertebrates. *Proceedings of the National Academy of Sciences*, 106, 13410–13414. https://doi.org/10.1073/pnas.0811087106

Angelis, K., & Dos Reis, M. (2015). The impact of ancestral population size and incomplete lineage sorting on Bayesian estimation of species divergence times. *Current Zoology*, 61, 874–885. https://doi.org/10.1093/czoolo/61.5.874

Arbour, J. H., & Santana, S. E. (2017). A major shift in diversification rate helps explain macroevolutionary patterns in primate species diversity. *Evolution*, 71, 1600–1613. https://doi.org/10.1111/evo.13237

Carneiro, M., Ferrand, N., & Nachman, M. W. (2009). Recombination and speciation: Loci near centromeres are more differentiated than loci near telomeres between subspecies of the European rabbit (Oryctolagus cuniculus). *Genetics*, 181, 593–606.

Chen, Z., Xu, S., Zhou, K., & Yang, G. (2011). Whale phylogeny and rapid radiation events revealed using novel retroposed elements and their flanking sequences. *BMC Evolutionary Biology*, 11, 314. https://doi.org/10.1186/1471-2148-11-314

Degnan, J. H., & Rosenberg, N. A. (2009). Gene tree discordance, phylogenetic inference and the multispecies coalescent. *Trends in Ecology & Evolution*, 24, 332–340. https://doi.org/10.1016/j.tree.2009.01.009

Edwards, S. V. (2009). Is a new and general theory of molecular systematics emerging? *Evolution*, 63, 1–19. https://doi.org/10.1111/j.1558-5646.2008.00549.x

Edwards, S. V., & Beerli, P. (2000). Perspective: Gene Divergence, population divergence, and the variance incoalescence time in phylogeographic studies. *Evolution*, 54, 1839–1854.

Edwards, S. V., Liu, L., & Pearl, D. K. (2007). High-resolution species trees without concatenation. *Proceedings of the National Academy of Sciences*, 104, 5936–5941. https://doi.org/10.1073/pnas.0607004104

Edwards, S. V., Xi, Z., Janke, A., Faircloth, B. C., McCormack, J. E., Glenn, T. C., … Davis, C. C. (2016). Implementing and testing the multispecies coalescent model: A valuable paradigm for phylogenomics. *Molecular Phylogenetics and Evolution*, 94, 447–462. Elsevier Inc. https://doi.org/10.1016/j.ympev.2015.10.027

Etter, R. J., Boyle, E. E., Glazier, A., Jennings, R. M., Dutra, E., & Chase, M. R. (2011). Phylogeography of a pan-Atlantic abyssal protobranch bivalve: Implications for evolution in the Deep Atlantic. *Molecular Ecology*, 20, 829–843. https://doi.org/10.1111/j.1365-294X.2010.04978.x

Gatesy, J., & Springer, M. S. (2013). Concatenation versus coalescence versus "concatalescence". *Proceedings of the National Academy of Sciences*, 110, E1179–E1179. https://doi.org/10.1073/pnas.1221121110

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. *Systematic Biology*, 59, 307–321. https://doi.org/10.1093/sysbio/syq010

Kong, A., Frigge, M. L., Masson, G., Besenbacher, S., Sulem, P., Magnusson, G., … Stefansson, K. (2012). Rate of de novo mutations and the importance of father's age to disease risk. *Nature*, 488, 471–475. Nature Publishing Group. https://doi.org/10.1038/nature11396

Kubatko, L. S., Carstens, B. C., & Knowles, L. L. (2009). STEM: Species tree estimation using maximum likelihood for gene trees under coalescence. *Bioinformatics*, 25, 971–973. https://doi.org/10.1093/bioinformatics/btp079

Kubatko, L. S., & Degnan, J. H. (2007). Inconsistency of phylogenetic estimates from concatenated data under coalescence. *Systematic Biology*, 56, 17–24. https://doi.org/10.1080/10635150601146041

Leaché, A. D., Harris, R. B., Rannala, B., & Yang, Z. (2014). The influence of gene flow on species tree estimation: A simulation study. *Systematic Biology*, 63, 17–30. https://doi.org/10.1093/sysbio/syt049

Lipson, M., Loh, P. R., Sankararaman, S., Patterson, N., Berger, B., & Reich, D. (2015). Calibrating the human mutation rate via ancestral recombination density in diploid genomes. *PLoS Genetics*, 11, 1–25.

Liu, L., Xi, Z., & Davis, C. C. (2015). Coalescent methods are robust to the simultaneous effects of long branches and incomplete lineage

sorting. *Molecular Biology and Evolution*, 32, 791–805. https://doi.org/10.1093/molbev/msu331

Liu, L., Yu, L., Kubatko, L., Pearl, D. K., & Edwards, S. V. (2009). Coalescent methods for estimating phylogenetic trees. *Molecular Phylogenetics and Evolution*, 53, 320–328. Elsevier Inc. https://doi.org/10.1016/j.ympev.2009.05.033

Liu, L., Yu, L., Pearl, D. K., & Edwards, S. V. (2009). Estimating species phylogenies using coalescence times among sequences. *Systematic Biology*, 58, 468–477. https://doi.org/10.1093/sysbio/syp031

Maley, J. M., & Winker, K. (2010). Diversification at high latitudes: Speciation of buntings in the genus Plectrophenax inferred from mitochondrial and nuclear markers. *Molecular Ecology*, 19, 785–797. https://doi.org/10.1111/j.1365-294X.2009.04513.x

May, M. R., & Moore, B. R. (2016). How well can we detect lineage-specific diversification-rate shifts? a simulation study of sequential aic methods. *Systematic Biology*, 65, 1076–1084. https://doi.org/10.1093/sysbio/syw026

McDaniel, S. F., Von Stackelberg, M., Richardt, S., Quatrano, R. S., Reski, R., & Rensing, S. A. (2010). The speciation history of the physcomitrium - Physcomitrella species complex. *Evolution*, 64, 217–231. https://doi.org/10.1111/j.1558-5646.2009.00797.x

Moore, B. R., Höhna, S., May, M. R., Rannala, B., & Huelsenbeck, J. P. (2016). Critically evaluating the theory and performance of Bayesian analysis of macroevolutionary mixtures. *Proceedings of the National Academy of Sciences*, 113, 9569–9574. https://doi.org/10.1073/pnas.1518659113

Oliver, J. C. (2013). Microevolutionary processes generate phylogenomic discordance at ancient divergences. *Evolution*, 67, 1823–1830. https://doi.org/10.1111/evo.12047

Pacifici, M., Santini, L., Di Marco, M., Baisero, D., Francucci, L., Grottolo Marasini, G., ... Rondinini, C. (2013). Generation length for mammals. *Nature Conservation*, 5, 89–94.

Papadopoulos, L. N., Peijnenburg, K. T. C. A., & Luttikhuizen, P. C. (2005). Phylogeography of the calanoid copepods *Calanus helgolandicus* and *C. euxinus* suggests Pleistocene divergences between Atlantic, Mediterranean, and Black Sea populations. *Marine Biology*, 147, 1353–1365. https://doi.org/10.1007/s00227-005-0038-x

Paradis, E., Claude, J., & Strimmer, K. (2004). APE: Analyses of phylogenetics and evolution in R language. *Bioinformatics*, 20, 289–290. https://doi.org/10.1093/bioinformatics/btg412

Pennell, M. W., Eastman, J. M., Slater, G. J., Brown, J. W., Uyeda, J. C., Fitzjohn, R. G., ... Harmon, L. J. (2014). Geiger v2.0: An expanded suite of methods for fitting macroevolutionary models to phylogenetic trees. *Bioinformatics*, 30, 2216–2218. https://doi.org/10.1093/bioinformatics/btu181

Penny, D., & Hendy, M. D. (1985). The use of tree comparison metrics. *Systematic Biology*, 34, 75–82. https://doi.org/10.1093/sysbio/34.1.75

Quental, T. B., & Marshall, C. R. (2010). Diversity dynamics: Molecular phylogenies need the fossil record. *Trends in Ecology & Evolution*, 25, 435–441.

Rabosky, D. L. (2014). Automatic detection of key innovations, rate shifts, and diversity-dependence on phylogenetic trees. *PLoS ONE*, 9, e89543. https://doi.org/10.1371/journal.pone.0089543

Rabosky, D. L., Grundler, M., Anderson, C., Title, P., Shi, J. J., Brown, J. W., ... Larson, J. G. (2014). BAMMtools: An R package for the analysis of evolutionary dynamics on phylogenetic trees. *Methods in Ecology and Evolution*, 5, 701–707. https://doi.org/10.1111/2041-210X.12199

Rabosky, D. L., & Huang, H. (2016). A robust semi-parametric test for detecting trait-dependent diversification. *Systematic Biology*, 65, 181–193. https://doi.org/10.1093/sysbio/syv066

Rabosky, D. L., Mitchell, J. S., & Chang J. (2017). Is BAMM flawed? theoretical and practical concerns in the analysis of multi-rate diversification models. *Systematic Biology*, 66, 477–498. https://doi.org/10.1093/sysbio/syx037

Revell, L. J., Harmon, L. J., Glor, R. E., & Linder, P. (2005). Under-parameterized model of sequence evolution leads to bias in the estimation of diversification rates from molecular phylogenies. *Systematic Biology*, 54, 973–983. https://doi.org/10.1080/10635150500354647

Schoville, S. D., Stuckey, M., & Roderick, G. K. (2011). Pleistocene origin and population history of a neoendemic alpine butterfly. *Molecular Ecology*, 20, 1233–1247. https://doi.org/10.1111/j.1365-294X.2011.05003.x

Schrago, C. G., Menezes, A. N., Furtado, C., Bonvicino, C. R., & Seuanez, H. N. (2014). Multispecies coalescent analysis of the early diversification of neotropical primates: Phylogenetic inference under strong gene trees/species tree conflict. *Genome Biology and Evolution*, 6, 3105–3114. https://doi.org/10.1093/gbe/evu244

Shi, J. J., & Rabosky, D. L. (2015). Speciation dynamics during the global radiation of extant bats. *Evolution*, 69, 1528–1545. https://doi.org/10.1111/evo.12681

Smith, S. A., & O'Meara, B. C. (2012). treePL: Divergence time estimation using penalized likelihood for large phylogenies supplementary analyses. *Bioinformatics*, 28, 2689–2690. https://doi.org/10.1093/bioinformatics/bts492

Song, S., Liu, L., Edwards, S. V., & Wu, S. (2012). Resolving conflict in eutherian mammal phylogeny using phylogenomics and the multispecies coalescent model. *Proceedings of the National Academy of Sciences*, 109, 14942–14947. https://doi.org/10.1073/pnas.1211733109

Steeman, M. E., Hebsgaard, M. B., Fordyce, R. E., Ho, S. Y. W., Rabosky, D. L., Nielsen, R., ... Willerslev, E. (2009). Radiation of extant cetaceans driven by restructuring of the oceans. *Systematic Biology*, 58, 573–585. https://doi.org/10.1093/sysbio/syp060

Tonini, J., Moore, A., Stern, D., Shcheglovitova, M., & Ortí, G. (2015). Concatenation and Species Tree Methods Exhibit Statistically Indistinguishable Accuracy under a Range of Simulated Conditions. *PLOS Currents Tree of Life.* https://doi.org/10.1371/currents.tol.34260cc27551a527b124ec5f6334b6be.

Viaud-Martinez, K. A., Brownell, R. L., Komnenou, A., & Bohonak, A. J. (2008). Genetic isolation and morphological divergence of Black Sea bottlenose dolphins. *Biological Conservation*, 141, 1600–1611. https://doi.org/10.1016/j.biocon.2008.04.004

Viaud-Martínez, K. A., Vergara, M. M., Gol'din, P. E., Ridoux, V., Öztürk, A. A., Öztürk, B., ... Bohonak, A. J. (2007). Morphological and genetic differentiation of the Black Sea harbour porpoise *Phocoena phocoena*. *Marine Ecology Progress Series*, 338, 281–294. https://doi.org/10.3354/meps338281

Xu, B., & Yang, Z. (2016). Challenges in species tree estimation under the multispecies coalescent model. *Genetics*, 204, 1353–1368. https://doi.org/10.1534/genetics.116.190173

Yang, Z. (1997). PAML : A program package for phylogenetic analysis by maximum likelihood. *Cabios Applications Note*, 13, 555–556. https://doi.org/10.1093/bioinformatics/13.5.555

Yang, Z. (2007). PAML 4: Phylogenetic analysis by maximum likelihood. *Molecular Biology and Evolution*, 24, 1586–1591. https://doi.org/10.1093/molbev/msm088

Zhu, S., Degnan, J. H., Goldstein, S. J., & Eldon, B. (2015). Hybrid-Lambda: Simulation of multiple merger and Kingman gene genealogies in species networks and species trees. *BMC Bioinformatics*, 16, 292. https://doi.org/10.1186/s12859-015-0721-y