Running head: Simulation-based examination of priors in biogeographic dating

Title: Towards a methodological background for using Biological data to inform Geological models: Interaction of Biological and Paleogeographic priors in Biogeographic Dating

Authors: Luke Sparreo1, Ana Bedoya1\*

1*Center for Biodiversity & Evolution, New York Botanical Garden, 2900 Southern Boulevard, Bronx, NY, USA 10458*

\*Corresponding author: [abedoya@nybg.org](mailto:abedoya@nybg.org)

# Abstract

Biological and geographic data have long reciprocally informed one another. In this study, we simulate 10 simple molecular phylogenies with known histories and date them with historic geological priors of varying information levels and accuracy in order to test the influence of geological knowledge and molecular data on phylogenetic dating. In our simulations, models prohibiting gene flow with unknown geological histories performed just as well as models prohibiting gene flow with correctly informed geological priors. The result that accurate molecular data performs similarly under correct and unknown geographic knowledge shows promise for the utilization of the geogenomic method, which has been used to date geographic events with biological data. However, the introduction of gene flow or the incorporation of incorrect historic geological data shifted the accuracy of our phylogenetic dating. These findings show that systematists must remain cautious when conducting biogeographic dating.

Keywords: biogeography, geogenomics, dating, molecular, simulation

# Introduction

The recognition of the interplay of biological evolution and geological change laid the foundation for historical biogeography (Wallace, 1854). The discipline centers on the study of how the geographic distribution of organisms has changed over evolutionary time and how those changes are shaped by Earth’s geological history. Therefore, biogeographic thinking inherently forces the interpretation of biological data in reference to a specified geological background. However, as organismal evolution often proceeds in parallel with changes in abiotic conditions (e.g., landscape or climate shifts; Dong et al., 2024), biological data has been used to propose and inform geological models. In other words, patterns of geographic co-variation across taxa have been used in concert with genomic data, as proxies to infer the timing of formation of shared abiotic processes that may have driven those patterns (e.g., formation of a mountain as a barrier to gene flow).

Examples of the use of biological data for testing or formulating landscape change evolution scenarios include: 1) the formulation (and later rejection) of the forest refugia hypothesis to explain a mechanism for the high biodiversity observed in regions like the Amazon (Haffer, 1969). 2) The development of a model for the timing and mode of uplift of the Andean cordillera (Sanín et al., 2022). 3) The estimation of the timing of river connectivity in norther South America since the Miocene (Bedoya et al., 2021). This approach was unified in a theoretical framework (i.e., geogenomics; Baker et al., 2014), which involves the reciprocal integration of geological, climatic, ecological, paleontological, and genomic data to test landscape or climate evolution hypotheses (Baker et al., 2014, Dolby et al., 2022; Rahbeck et al., 2019; Badgley et al., 2017). This theoretical framework has also discussed the main characteristics of ideal study systems, as well the geographic and geological scales at which geological models can be developed from biological data (Baker et al., 2014, Dolby et al., 2022; Bedoya, 2024).

Using biological and geological data independently, or interpreting them in reference to each other (e.g., establishing geological hypotheses as ground truth against which biological data are interpreted), may introduce biases inherent to each biology and geology. Standard approaches for the analysis of geological and of biological data inherently carry uncertainty (e.g., the interpretation of zircon data in provenance analyses and the estimation of divergence times across lineages (Weltje & von Eynatten, 2004). Further, as the responses of organisms to a shared paleogeographical event are idiosyncratic, careful attention should be placed to the model system used to infer geological models from biological data. Organisms best serve as appropriate model systems for this purpose when their evolutionary trajectories mirror geological history. In this regard it is worth noting that even if evolutionary patterns of a given set of organisms are shaped by a landscape event, the predictability of landscape hypotheses from evolutionary patterns may be affected by common and stochastic processes (e.g., gene flow).

An example that illustrates the overall biases in the use of biological and geological data to date the timing of geomorphic events, is the study of the timing and pattern of closure of the Isthmus of Panama. The traditional model for the evolution of the Panama Isthmus suggests that its formation took place ~4.2–3.5 Ma (late Pliocene; Keigwin, 1978, 1982; Coates et al., 1992; Jackson & O’Dea, 2013) or earlier at 2.8 Ma (O’Dea et al., 2016). Geochronological data and provenance analyses have challenged this model, pushing back the timing of closure of the Central American Seaway to 13–15 Ma (middle Miocene; Montes et al., 2012a, 2012b, 2015). This more recently published model is based on the interpretation of detrital zircon provenance data (Montes et al., 2015), and has been called into question (O’Dea et al., 2016). However, the evidence used to this end has in turn been challenged (Jaramillo et al., 2017; Molnar, 2017; Jaramillo, 2018). Inferred divergence times of several taxa distributed across the Americas —including freshwater fishes, ground sloths, shrimp, and several plant taxa— are older than 3 Ma, supporting the notion of pre-Pliocene contact (Erkens et al., 2007; Hurt et al., 2009; Weir et al., 2009; Cody et al., 2010; Pinto-Sánchez et al., 2012; Bacon et al., 2013; Leigh et al., 2014; Barker et al., 2015). Multiple pulses of migration across the Isthmus prior to 3.5 Ma were inferred by Bacon et al., 2015, in opposition to the traditional model of formation of the Isthmus. While earlier pulses of organismal migration between the Americas may be evidence for an earlier onset of a land bridge connecting the Americas, earlier divergence times have sometimes been interpreted as episodes of stochastic long-distance dispersal by defendants of the traditional model of closure of the Isthmus of Panama (O’Dea et al., 2016). This underscores the importance of considering species-specific traits when interpreting bio/phylogeographic patterns used for inferring geological models.

Despite the existence of a theoretical background for the integration of biological, paleontological, and geological data to test geomorphic hypotheses, a unified methodological framework is still lacking. Such a framework would allow testing of how geological priors and biological data interact. This would facilitate the exploration of the effect of incorrect, vague, and correct geological models (i.e., those that deviate from reality, incorporate great uncertainty, or accurately describe the timing of a geomorphic event), as well as the impact of using biological study systems that variably mirror geological processes. Here, we address this gap using biogeographic dating, a data-dependent and process-based method that uses biogeographic information alongside molecular data, and time-calibrated paleogeographical priors for divergence dating (Landis, 2017; Landis et al., 2018). In biogeographic dating, the observed biogeographic states of extant taxa induce a posterior distribution of dated speciation times. Specifically, observed distributional data are used to model dispersal rates between areas. The estimation of such dispersal rates results from the interaction of tip data with a specified paleogeographic prior (e.g., for the emergence or disappearance of a barrier), where the latter influences the expected probability that a particular dispersal event between areas takes place (Fig. 1a).

In this study, we aimed to test the reciprocal influence of biological and geological data and priors, leveraging biogeographic dating as a method where both are specified and analyzed simultaneously. We generated a hypothetical scenario where a temporal and spatial correspondence of biological and geological processes was enforced, such that divergence times are a proxy for the timing of a landscape change event (Fig. 1b). We specified geological priors with varying levels of accuracy and simulated the evolution of two model systems whose evolutionary histories do and do not mirror an underlying geological event. In using this simulated dataset with known geological, evolutionary histories, we test the potential for well-calibrated phylogenies to estimate paleogeographic events and the impact of accuracy and uncertainty in geologic models.

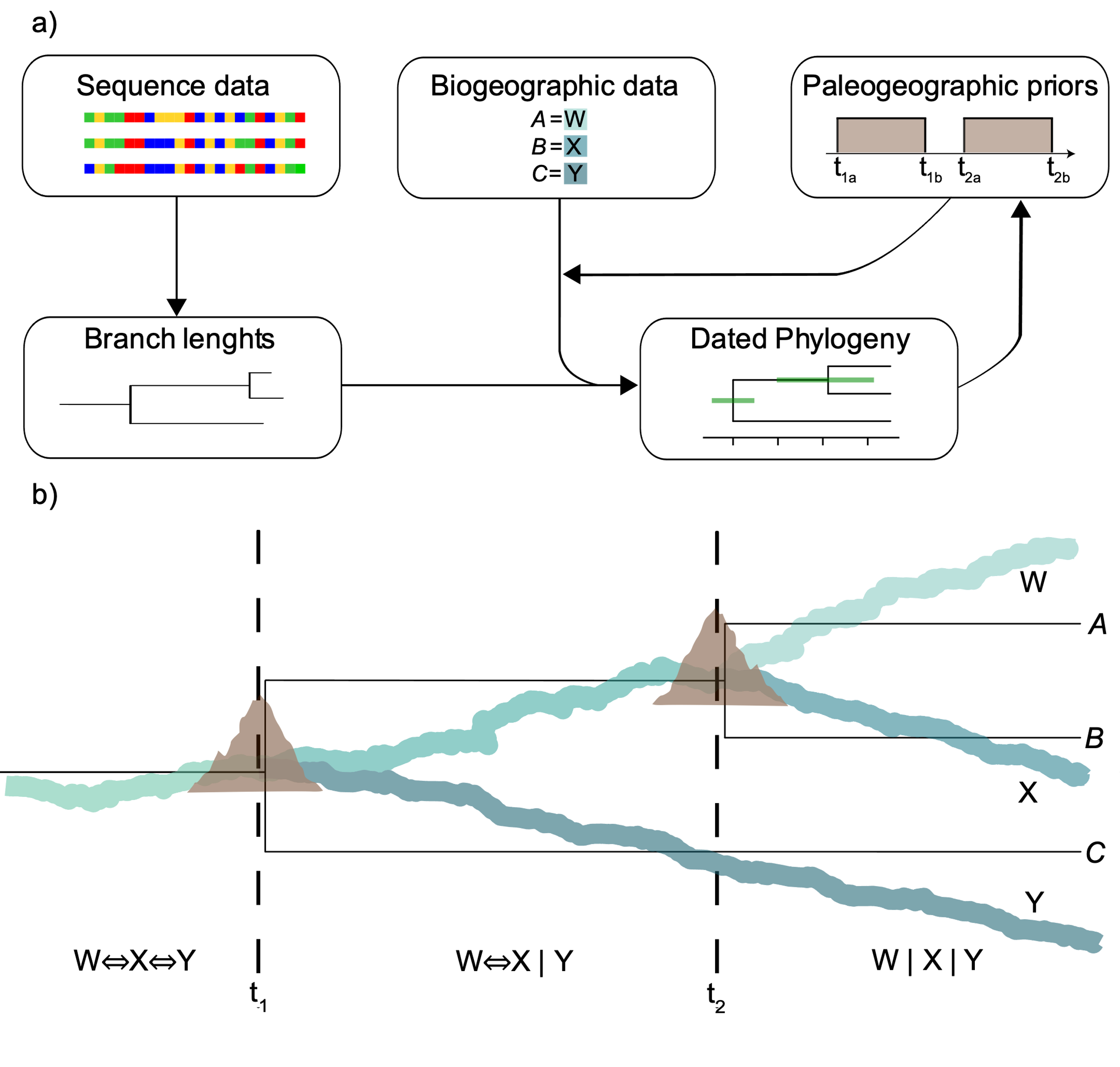
****

Figure 1: Data, priors, and their interaction in biogeographic dating for the estimation of divergence times. The framework facilitates the reciprocal influence of biological and geological data. b) Simulated scenario illustrating the splitting or rivers with surface uplift and the parallel evolution of a river plant lineage. Dotted lines represent known geological dates of uplift and river splits at 6 Ma and 2 Ma. The outgroup is not represented in the figure.

# Materials & Methods

## *Evolutionary and Geological scenarios*

We built a hypothetical scenario where a river-dwelling plant lineage occurs across rivers in a basin that experienced three historical rearrangements; a landscape change event (i.e., surface uplift) triggered the split of rivers at 10, 6, and 2 Ma. River splitting and the formation of isolated river branches (W, X, Y, and Z) result in lineage splits and divergence, and the origin of taxa A, B, C, and D respectively, whose phylogenetic relationships mirror the history of riverine connections (Fig. 1b). Initially, we assume that our study system is bounded to rivers, thus having a limited dispersal.

## *Demographic and Sequence Simulation*

We simulated the evolution of our hypothetical river plant lineage under two scenarios with msprime (Baumdicker et al., 2022). In the first scenario, the demographic model exactly mirrored the known history of river splits (gene flow between taxa was not allowed such that geological events completely isolated taxonomic groups). We modelled the first split at t=10 Ma, leading to outgroup taxon D and taxon ABC. A second split was specified at t=6Ma, resulting in taxon C splitting from taxon AB. Finally, a third split was specified at t=2Ma, separating taxon AB into taxa A and B. We specified a generation time of one year (in line with what has been reported in real-world river-dwelling angiosperms; Philbrick & Retana, 1998). We generated a DNA matrix of 10,000 bp under the HKY substitution model, a mutation rate of 1e-8 substitutions per site per generation (Ossowski et al., 2010; as in Exposito-Alonso et al., 2022), and a recombination rate of 4.8 cM/Mb (Rizzon et al., 2006; as in Brown et al., 2016). All invariant sites were filled in with random invariant bases using a custom script (fill\_invariant\_sites.py from https://github.com/Bedoya-Research-Group/BioGeoDating\_Geogenomics/).

The second scenario we modeled the evolution of taxa whose evolutionary history do not perfectly match the evolution of riverine connections due to migration. We allowed for gene flow to take place across taxa A and B located in two adjacent rivers (W and X) at 2–1 Ma. Gene flow was simulated using the migration function in msprime (rate = 0.2). Demographic models were visualized using the package DemesDraw (Gower et al., 2022). When prohibiting gene flow, our demographic model produced divergence patterns and times that mirror the known geological event times from our hypothetical scenario (Figs. 2a and 1b). Contrastingly, the demographic model permitting gene flow produced a tree that is inconsistent with the geological history in our model (Fig. 2b). Inferred trees resulting from our simulations in msprime were used as input in downstream analyses.

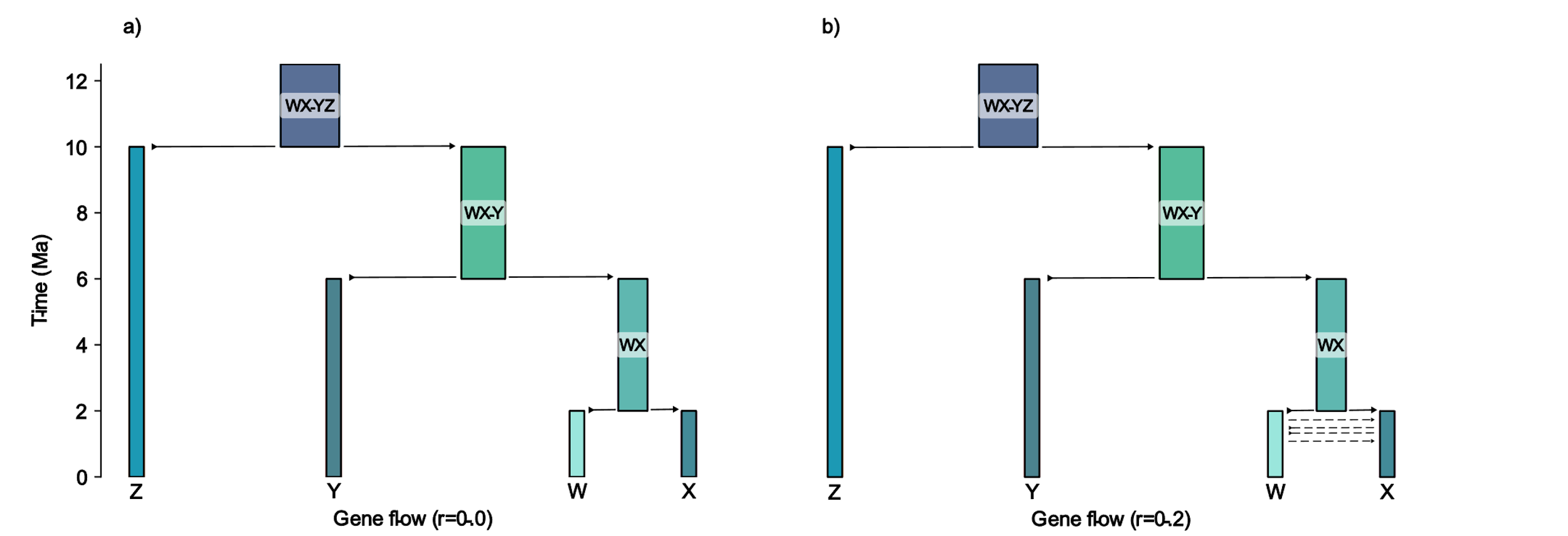
****

Figure 2: Demographic models used to generate sequence data and trees with msprime. a) Scenario without gene flow, where biological history mirrors geological events. b) Hypothetical scenario where biological and geological events do not match due to migration across landscape units (gene flow is allowed between taxa W and X from 2 to 1 Ma).

## *Biogeographic dating with differing Biological data and Geological Priors*

Under the scenario where biological and geological events coincide (no gene flow; 2a and 1b), we specified five different geological priors in five independent biogeographic-dating analyses (Fig. 3); informed with hard bounds, informed with soft bounds, incorrect with hard bounds, incorrect with soft bounds, and unknown. These five priors represent different degrees of accuracy and certainty in geological models. For instance, informed models with hard bounds illustrate the formation of the Hawaiian archipelago (Clague & Dalrymple, 1994) as used in empirical and simulation studies (Lerner et al., 2011; Clague et al., 2010). Assumptions on the timing of events can be relaxed to accommodate uncertainty (i.e., using soft bounds). In all cases, hard bounds and soft bounds were modelled using uniform and gamma distributions respectively (Fig. 3). The uniform distributions were 1Ma in length, extending 0.5Ma from either end of the time. A gamma distribution centered on the time was used to approximate a normal distribution with a SD of 0.1 due to challenges working with normal distributions (not type RealPos) in ensuing computational analyses. In informed models, the distribution bounds were centered on 6 and 2 Ma respectively, while incorrect models centered distributions on 8 and 2 Ma. We then used the five geological priors to run another set of five independent biogeographic-dating analyses, but under the second scenario where biological and geological events did not perfectly match (gene flow allowed; Fig. 2b).

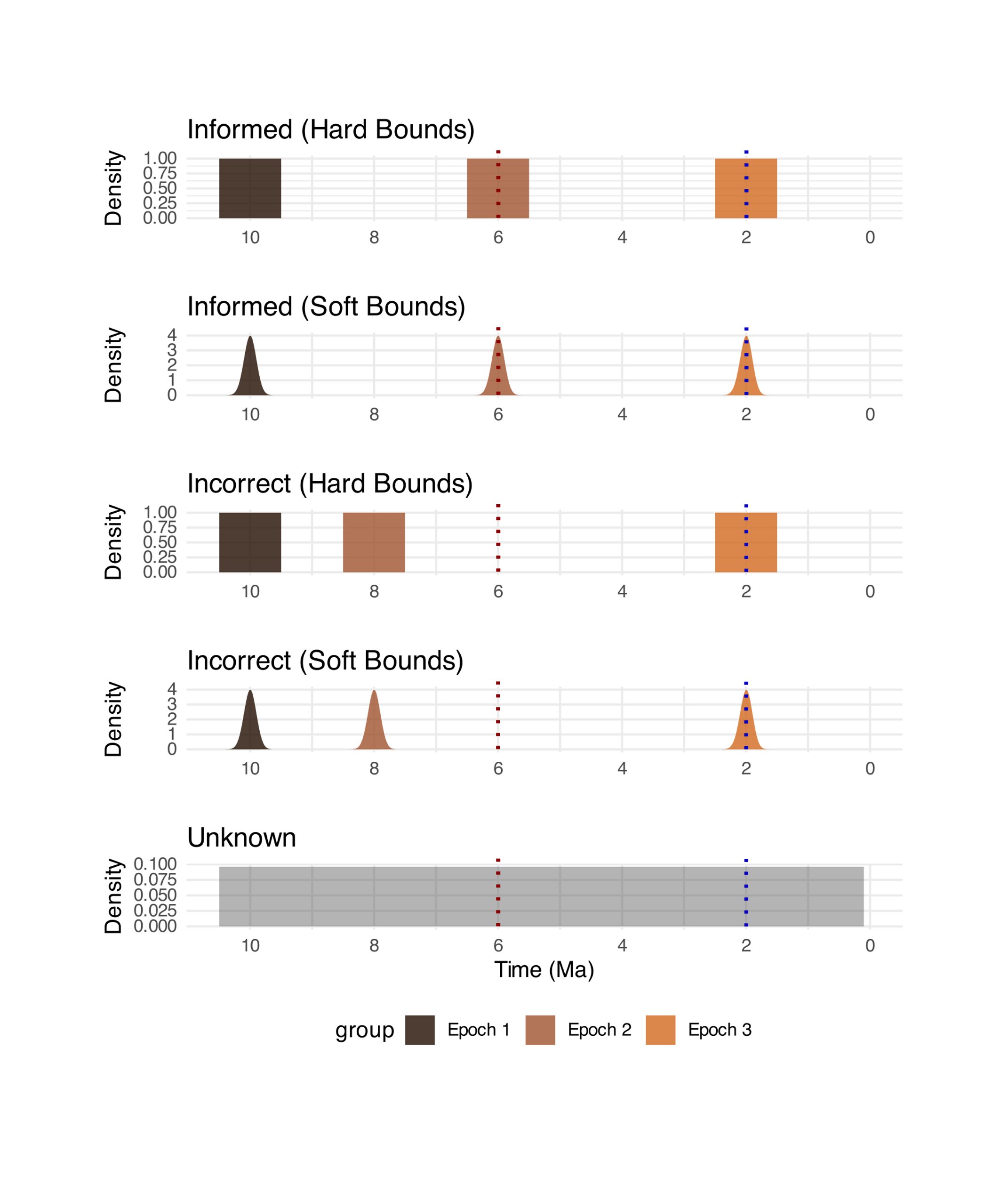
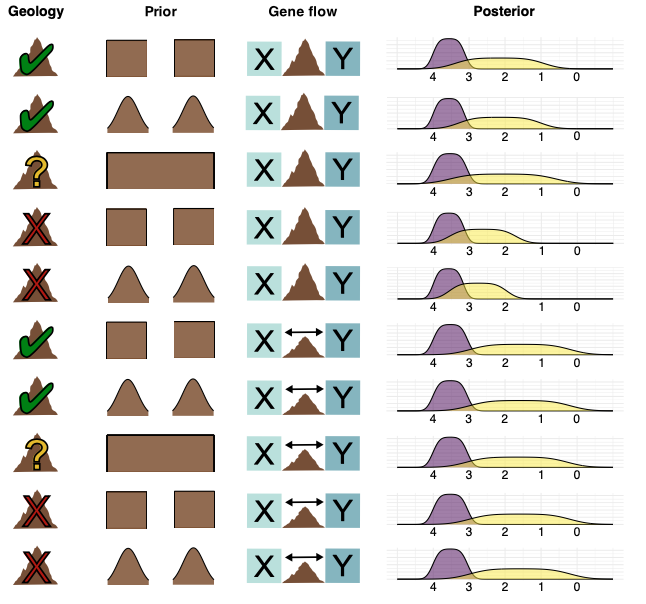
****

Figure 3. Geological priors specified for the two biological scenarios modeled. Red dotted line represents the true time for the node where taxon C splits from taxa A and B, and blue dotted line represents the true time for the node where taxon A splits from taxon B.

Biogeographic dating analyses for the model data were performed in RevBayes (Höhna et al., 2016) under the Dispersal-Extinction-Cladogenesis (DEC) model (Landis, 2017; Landis et al., 2018). We created new scripts adapted from Landis, 2017 for biogeographic dating in RevBayes available at https://github.com/Bedoya-Research-Group/BioGeoDating\_Geogenomics.git. Molecular sequence data, a phylogeny, present ranges, distances between areas, geological event times, and dispersal probabilities across the latter, are defined as inputs to jointly estimate divergence times, and range evolution. Phylogenetic relationships can be specified a priori or estimated in parallel. Two MCMC chains were run with 10 million generations, saving a tree every 100 generations, for each of the ten analyses. In all analyses we assigned a uniform distribution from 0 to 10.5 Ma for the root age of the tree. The extirpation rate was set to zero. Chain convergence was assessed with Tracer v1.7.2 (Rambaut et al., 2018) and with the R package Convenience (Fabreti & Höhna, 2022). The two generated tree files for each analysis were then combined after removing 0.25 burn-in fraction on each file. The final generated trees and posterior distributions were visualized in R using RevGadgets (Tribble et al., 2021).

# Results

****

**Figure 4:** Posterior distributions produced from 10 biogeographic-dating analyses conducted using different demographic models and geological priors. Purple: Yellow:

When using geological priors that conform to our hypothetical landscape change scenario (Fig. 4 and Figs. S1, S2, S6, S7), the 95% HPD of the estimated divergence time for the splitting event between taxa A and B (rivers X and Y) spanned the true date (i.e., 1 Ma; Table S1). However, in all cases the mean estimated age for the node was older than simulated. When gene flow was modeled, divergence times for this node were more recent (~ 0.7 Ma younger in models permitting gene flow). For correctly informed models, the choice between uniform and normal priors had little influence on absolute node age regardless of gene flow. However, priors with a normal distribution resulted in slightly older age distribution for the node in question in the absence of gene flow.

In the geology unknown models (Fig. 4 and Figs. S3, S8), analyses with or without gene flow resulted in , the models found an absolute age for both nodes similar (around 0.1 million years) to the correctly informed models. Like the correctly informed model, the introduction of a short period of gene flow in the demographic history of geology unknown models had a large influence on the predicted absolute age of node two. The absolute age of node two was found to be around 0.5 million years younger when gene flow was included when compared to the unknown model prohibiting gene flow. Additionally, the age distributions for node two were similar to the correctly informed models when gene flow was permitted. However, in the unknown model prohibiting gene flow, the age distribution for node two was dated more broadly than the correctly informed models that prohibited gene flow.

When incorrect geological priors were used in the model (Figs. S4, S5, S9, S10), results for absolute node age and the age distribution for node two shifted in models prohibiting gene flow. The absolute age for node two was about 0.5 years older in both incorrect models prohibiting gene flow, and the age distribution for node two was dated as older as well. Interestingly, in the models that permitted gene flow, the incorrect, correctly informed, and unknown models all produced similar (around 0.1 million years) results. Predictably, the estimated absolute age for node one remained roughly unchanged, as it was not manipulated in the epoch time estimation nor the demographic model. There were no notable differences in the age estimations between using the uniform or normal distributions in the incorrect models.

# Discussion

## *Geologically unknown prior data dates phylogenies comparably to accurate geological priors*

Our results found that models with a flat, unknown geological prior performed nearly as well at estimating divergence times as models with a uniform or normal geological prior centered around correct times. Correct, well-informed simulated data successfully being able to predict the timing of the emergence of biogeographic barriers without prior historic geographic information suggests exciting opportunities to test new biogeographic hypotheses. These results confirm the critical idea in geogenomics that biology and geography may reciprocally inform each other.

While the strong performance of the molecular data is exciting, it should be noted that in our simulations, our molecular data was “perfect”: a model scenario in which genetic sequences had no missing data. Additionally, we assume our taxon sampling was complete; there were no unknown extant taxa nor extinct taxa in our group that lacked molecular data. Yet, has been thought that missing data at the molecular (Roure et al., 2013) and taxonomic levels (Zwickl & Hillis, 2002) may negatively influence the accuracy of phylogenies, which could in turn create issues downstream in dating methods. Additionally, our model dataset lacked common informatic errors like inaccurate molecular alignment (Ogden & Rosenberg, 2006) or erroneous taxonomic identification (de Almeida et al., 2023) both of which may lead to inaccurate phylogenies and biogeographic dating. In realistic datasets, imperfect molecular data may interfere with the ability of molecular data to accurately test geographic hypotheses.

Furthering genomic data acquisition and empirical analyses in model systems such as montane or riparian species has been proposed (Bedoya, 2024) as the next steps in furthering the empirical testing of reciprocal geographic-biological dating methods. Model systems such as these that lack known demographic histories of gene flow, have strongly-supported biogeographic histories, and already have accepted dated phylogenies through adequate fossil calibrations (such as Landis et al., 2021) will be necessary to explore the possibilities of dating geographic events through biological data.

## *Incorrect geographic historical data alters dating accuracy*

With the introduction of incorrect prior information, however, models performed significantly worse in predicting geological history. This supports previous studies that have indicated incorrect assumptions within the DEC (Magalhaes et al., 2021) and fossilized birth-death (Barido-Sottani et al., 2020) models have large impacts on inference in biogeographic estimations. Considering debate continues over the timing of influential geological events (White et al., 2013; Zheng, 2015), the selection of accurately resolved paleogeographic priors is critical. Further work should also explore the impact of these historic geographic assumptions in other biogeographic models like DEC+J (Matzke, 2014), which have been debated regarding ability to accurately parameterize the rates of speciation (Ree & Sanmartin, 2018), as the accuracy of model assumptions likely plays a large role in the accuracy of phylogenetic dating results.

Researchers should be conservative and cautious when creating strict priors in biogeographic dating methods, especially considering the fact that with our accurate simulated molecular data, a flat geological prior performed nearly as well as correctly-informed priors.

## *Demographic models that allow gene flow alter dating accuracy*

When demographic models allowed for gene flow, the predicted cladogenesis times were found to be significantly younger, regardless of the historic geographic prior used. The fact that all models allowing gene flow, regardless of the geological information present, produced similar results suggests that the existence of gene flow in molecular data creates a significant barrier to the accurate implementation in using geological data to date phylogenies. Current work continues to reveal histories of introgression (Stull et al., 2023), including from extinct lineages (Tricou et al., 2022), which did not exist in our model system. Researchers should take caution in using geological priors to date phylogenies with known or suspected gene flow.

While gene flow may interfere with inferring the correct biogeographic history, further research may develop methods that allow systematists to properly date phylogenies of systems that include rampant gene flow. In fact, if properly diagnosed, the detection of gene flow itself may be a tool to understand the history of past biogeographic connections, as gene flow is more likely to occur between species that have geographic overlap (Liu et al., 2024) and in areas with ephemeral geological barriers (Araya-Donoso et al., 2022).

# Funding Acknowledgement

X

# Acknowledgements

Code used to run our model simulations was adapted from existing code produced in Landis, 2017 and Landis et al., 2018.

# Data Availability Statement

All code for the generation data and analyses included here is available at the following GitHub repository https://github.com/Bedoya-Research-Group/River\_phylogeography

# References

de Almeida, R. F., Pellegrini, M. O. O., de Morais, I. L., Simão-Bianchini, R., Rattanakrajang, P., Cheek, M., & Simões, A. R. G. (2023). Barking up the wrong tree: the dangers of taxonomic misidentification in molecular phylogenetic studies. *Plant ecology and evolution*, *156*(2), 146–159.

Araya‐Donoso, R., Baty, S. M., Alonso‐Alonso, P., Sanín, M. J., Wilder, B. T., Munguia‐Vega, A., & Dolby, G. A. (2022). Implications of barrier ephemerality in geogenomic research. *Journal of biogeography*.

Bacon, C. D., Mora, A., Wagner, W. L., & Jaramillo, C. A. (2013). Testing geological models of evolution of the Isthmus of Panama in a phylogenetic framework. *Botanical journal of the Linnean Society. Linnean Society of London*, *171*(1), 287–300.

Bacon, C. D., Silvestro, D., Jaramillo, C., Smith, B. T., Chakrabarty, P., & Antonelli, A. (2015). Biological evidence supports an early and complex emergence of the Isthmus of Panama. *Proceedings of the National Academy of Sciences of the United States of America*, *112*(19), 6110–6115.

Badgley, C., Smiley, T. M., Terry, R., Davis, E. B., DeSantis, L. R. G., Fox, D. L., Hopkins, S. S. B., et al. (2017). Biodiversity and topographic complexity: modern and geohistorical perspectives. *Trends in Ecology & Evolution*, *32*(3), 211–226.

Barido-Sottani, J., van Tiel, N. M. A., Hopkins, M. J., Wright, D. F., Stadler, T., & Warnock, R. C. M. (2020). Ignoring fossil age uncertainty leads to inaccurate topology and divergence time estimates in time calibrated tree inference. *Frontiers in ecology and evolution*, *8*.

Barker, F. K., Burns, K. J., Klicka, J., Lanyon, S. M., & Lovette, I. J. (2015). New insights into New World biogeography: An integrated view from the phylogeny of blackbirds, cardinals, sparrows, tanagers, warblers, and allies. *The Auk*, *132*(2), 333–348.

Baumdicker, F., Bisschop, G., Goldstein, D., Gower, G., Ragsdale, A. P., Tsambos, G., Zhu, S., et al. (2022). Efficient ancestry and mutation simulation with msprime 1.0. *Genetics*, *220*(3).

Bedoya, A. M. (2024). Botany and geogenomics: Constraining geological hypotheses in the neotropics with large-scale genetic data derived from plants. *American Journal of Botany*, *111*(4), e16306.

Bedoya, A. M., Leaché, A. D., & Olmstead, R. G. (2021). Andean uplift, drainage basin formation, and the evolution of plants living in fast-flowing aquatic ecosystems in northern South America. *The New Phytologist*, *232*(5), 2175–2190.

Coates, A. G., Jackson, J. B. C., Collins, L. S., Cronin, T. M., Dowsett, H. J., Bybell, L. M., Jung, P., et al. (1992). Closure of the Isthmus of Panama: The near-shore marine record of Costa Rica and western Panama. *GSA Bulletin*.

Cody, S., Richardson, J. E., Rull, V., Ellis, C., & Pennington, R. T. (2010). The Great American Biotic Interchange revisited. *Ecography*, no-no.

Dolby, G. A., Bennett, S. E. K., Dorsey, R. J., Stokes, M. F., Riddle, B. R., Lira-Noriega, A., Munguia-Vega, A., et al. (2022). Integrating Earth-life systems: a geogenomic approach. *Trends in Ecology & Evolution*, *37*(4), 371–384.

Erkens, R. H. J., Chatrou, L. W., Maas, J. W., van der Niet, T., & Savolainen, V. (2007). A rapid diversification of rainforest trees (Guatteria; Annonaceae) following dispersal from Central into South America. *Molecular Phylogenetics and Evolution*, *44*(1), 399–411.

Fabreti, L. G., & Höhna, S. (2021). Convergence Assessment for Bayesian Phylogenetic Analysis using MCMC simulation. *Methods in Ecology and Evolution*.

Gower, G., Ragsdale, A. P., Bisschop, G., Gutenkunst, R. N., Hartfield, M., Noskova, E., Schiffels, S., et al. (2022). Demes: a standard format for demographic models. *Genetics*, *222*(3).

Haffer, J. (1969). Speciation in amazonian forest birds. *Science*, *165*(3889), 131–137.

Harris, C. R., Millman, K. J., van der Walt, S. J., Gommers, R., Virtanen, P., Cournapeau, D., Wieser, E., et al. (2020). Array programming with NumPy. *Nature*, *585*(7825), 357– 362.

Harris, L. W., & Davies, T. J. (2016). A Complete Fossil-Calibrated Phylogeny of Seed Plant Families as a Tool for Comparative Analyses: Testing the “Time for Speciation” Hypothesis. *Plos One*, *11*(10), e0162907.

Hill, E. C., Gao, D. F., Polhemus, D. A., Fraser, C. J., Iova, B., Allison, A., & Butler, M. A. (2023). Testing Geology with Biology: Plate Tectonics and the Diversification of Microhylid Frogs in the Papuan Region. *Integrative organismal biology (Oxford, England)*, *5*(1), obad028.

Höhna, S., Landis, M. J., Heath, T. A., Boussau, B., Lartillot, N., Moore, B. R., Huelsenbeck, J. P., et al. (2016). RevBayes: Bayesian phylogenetic inference using graphical models and an interactive model-specification language. *Systematic Biology*, *65*(4), 726–736.

Hurt, C., Anker, A., & Knowlton, N. (2009). A multilocus test of simultaneous divergence across the Isthmus of Panama using snapping shrimp in the genus Alpheus. *Evolution*, *63*(2), 514–530.

Jaramillo, C. A. (2018). *Evolution of the Isthmus of Panama: Biological, Paleoceanographic and Paleoclimatological Implications*. John Wiley & Sons.

Jaramillo, C., Montes, C., Cardona, A., Silvestro, D., Antonelli, A., & Bacon, C. D. (2017). Comment (1) on “formation of the isthmus of panama” by o’dea et al. *Science Advances*, *3*(6), e1602321.

Keigwin, L. (1982). Isotopic paleoceanography of the Caribbean and East pacific: role of panama uplift in late neogene time. *Science*, *217*(4557), 350–353.

Keigwin, L. D. (1978). Pliocene closing of the Isthmus of Panama, based on biostratigraphic evidence from nearby Pacific Ocean and Caribbean Sea cores. *Geology*.

Koch, M. A., Haubold, B., & Mitchell-Olds, T. (2000). Comparative evolutionary analysis of chalcone synthase and alcohol dehydrogenase loci in Arabidopsis, Arabis, and related genera (Brassicaceae). *Molecular Biology and Evolution*, *17*(10), 1483–1498.

Landis, M. J. (2017). Biogeographic dating of speciation times using paleogeographically informed processes. *Systematic Biology*, *66*(2), 128–144.

Landis, M. J., Eaton, D. A. R., Clement, W. L., Park, B., Spriggs, E. L., Sweeney, P. W., Edwards, E. J., et al. (2021). Joint Phylogenetic Estimation of Geographic Movements and Biome Shifts during the Global Diversification of Viburnum. *Systematic Biology*, *70*(1), 67–85.

Landis, M. J., Freyman, W. A., & Baldwin, B. G. (2018). Retracing the Hawaiian silversword radiation despite phylogenetic, biogeographic, and paleogeographic uncertainty. *Evolution*, *72*(11), 2343–2359.

Leigh, E. G., O’Dea, A., & Vermeij, G. J. (2014). Historical biogeography of the Isthmus of Panama. *Biological Reviews of the Cambridge Philosophical Society*, *89*(1), 148–172.

Liu, S.-Y., Yang, Y.-Y., Tian, Q., Yang, Z.-Y., Li, S.-F., Valdes, P. J., Farnsworth, A., et al. (2025). An integrative framework reveals widespread gene flow during the early radiation of oaks and relatives in Quercoideae (Fagaceae). *Journal of Integrative Plant Biology*, *67*(4), 1119–1141.

Magalhaes, I. L. F., Santos, A. J., & Ramírez, M. J. (2021). Incorporating Topological and Age Uncertainty into Event-Based Biogeography of Sand Spiders Supports Paleo-Islands in Galapagos and Ancient Connections among Neotropical Dry Forests. *Diversity*, *13*(9), 418.

Matzke, N. J. (2014). Model selection in historical biogeography reveals that founder-event speciation is a crucial process in island clades. *Systematic Biology*, *63*(6), 951–970.

Molnar, P. (2017). Comment (2) on “formation of the isthmus of panama” by o’dea et al. *Science Advances*, *3*(6), e1602320.

Montes, C, Cardona, A., Jaramillo, C., Pardo, A., Silva, J. C., Valencia, V., Ayala, C., et al. (2015). Middle Miocene closure of the Central American Seaway. *Science*, *348*(6231), 226–229.

Montes, C, Cardona, A., McFadden, R., Moron, S. E., Silva, C. A., Restrepo-Moreno, S., Ramirez, D. A., et al. (2012). Evidence for middle Eocene and younger land emergence in central Panama: Implications for Isthmus closure. *Geological Society of America Bulletin*, *124*(5–6), 780–799.

Montes, Camilo, Bayona, G., Cardona, A., Buchs, D. M., Silva, C. A., Morón, S., Hoyos, N., et al. (2012). Arc-continent collision and orocline formation: Closing of the Central American seaway. *Journal of Geophysical Research*, *117*(B4).

O’Dea, A., Lessios, H. A., Coates, A. G., Eytan, R. I., Restrepo-Moreno, S. A., Cione, A. L., Collins, L. S., et al. (2016). Formation of the Isthmus of Panama. *Science Advances*, *2*(8), e1600883.

Ogden, T. H., & Rosenberg, M. S. (2006). Multiple sequence alignment accuracy and phylogenetic inference. *Systematic Biology*, *55*(2), 314–328.

Parham, J. F., Donoghue, P. C. J., Bell, C. J., Calway, T. D., Head, J. J., Holroyd, P. A., Inoue, J. G., et al. (2012). Best practices for justifying fossil calibrations. *Systematic Biology*, *61*(2), 346–359.

Philbrick, C. T., & Novelo Retana, A. (1998). Flowering phenology, pollen flow, and seed production in Marathrum rubrum (Podostemaceae). *Aquatic Botany*, *62*(3), 199–206.

Pinto-Sánchez, N. R., Ibáñez, R., Madriñán, S., Sanjur, O. I., Bermingham, E., & Crawford, A. J. (2012). The Great American Biotic Interchange in frogs: multiple and early colonization of Central America by the South American genus Pristimantis (Anura: Craugastoridae). *Molecular Phylogenetics and Evolution*, *62*(3), 954–972.

Rahbek, C., Borregaard, M. K., Antonelli, A., Colwell, R. K., Holt, B. G., Nogues-Bravo, D., Rasmussen, C. M. Ø., et al. (2019). Building mountain biodiversity: Geological and evolutionary processes. *Science*, *365*(6458), 1114–1119.

Rambaut, A., Drummond, A. J., Xie, D., Baele, G., & Suchard, M. A. (2018). Posterior summarization in Bayesian phylogenetics using Tracer 1.7. *Systematic Biology*, *67*(5), 901–904.

Ree, R. H., & Sanmartín, I. (2018). Conceptual and statistical problems with the DEC+J model of founder-event speciation and its comparison with DEC via model selection. *Journal of biogeography*, *45*(4), 741–749.

dos Reis, M., Inoue, J., Hasegawa, M., Asher, R. J., Donoghue, P. C. J., & Yang, Z. (2012). Phylogenomic datasets provide both precision and accuracy in estimating the timescale of placental mammal phylogeny. *Proceedings. Biological Sciences / the Royal Society*, *279*(1742), 3491–3500.

Rieux, A., & Balloux, F. (2016). Inferences from tip-calibrated phylogenies: a review and a practical guide. *Molecular Ecology*, *25*(9), 1911–1924.

Roure, B., Baurain, D., & Philippe, H. (2013). Impact of missing data on phylogenies inferred from empirical phylogenomic data sets. *Molecular Biology and Evolution*, *30*(1), 197– 214.

Sanín, M. J., Mejía‐Franco, F. G., Paris, M., Valencia‐Montoya, W. A., Salamin, N., Kessler, M., Olivares, I., et al. (2022). Geogenomics of montane palms points toMiocene–Pliocene Andean segmentation related to strike‐slip tectonics. *Journal of biogeography*, *49*(9), 1711–1725.

Stull, G. W., Pham, K. K., Soltis, P. S., & Soltis, D. E. (2023). Deep reticulation: The long legacy of hybridization in vascular plant evolution. *The Plant Journal*, *114*(4), 743–766.

Tiley, G. P., & Burleigh, J. G. (2015). The relationship of recombination rate, genome structure, and patterns of molecular evolution across angiosperms. *BMC Evolutionary Biology*, *15*, 194.

Jackson, J.B.C. & O’Dea, A. (2013). Timing of the oceanographic and biological isolation of the Caribbean Sea from the tropical eastern Pacific Ocean. *Bulletin of Marine Science*, *89*, 4, 779-800.

Tribble, C. M., Freyman, W. A., Landis, M. J., Lim, J. Y., Barido‐Sottani, J., Kopperud, B. T., Höhna, S., et al. (2021). RevGadgets: an R Package for visualizing Bayesian phylogenetic analyses from RevBayes. *Methods in Ecology and Evolution*.

Tricou, T., Tannier, E., & de Vienne, D. M. (2022). Ghost lineages highly influence the interpretation of introgression tests. *Systematic Biology*, *71*(5), 1147–1158.

Wallace, A. R. (1854). On the monkeys of the amazon. *Annals and Magazine of Natural History*, *14*(84), 451–454.

Wallace, A. R. (1862). On the physical geography of the malay archipelago. *Proceedings of the Royal Geographical Society of London*, *7*(5), 206.

Weir, J. T., Bermingham, E., & Schluter, D. (2009). The Great American Biotic Interchange in birds. *Proceedings of the National Academy of Sciences of the United States of America*, *106*(51), 21737–21742.

Weltje, G. J., & von Eynatten, H. (2004). Quantitative provenance analysis of sediments: review and outlook. *Sedimentary geology*, *171*(1–4), 1–11.

White, L. T., Gibson, G. M., & Lister, G. S. (2013). A reassessment of paleogeographic reconstructions of eastern Gondwana: Bringing geology back into the equation. *Gondwana Research*, *24*(3–4), 984–998.

Whittaker, R. J., Fernández-Palacios, J. M., Matthews, T. J., Borregaard, M. K., & Triantis, K. A. (2017). Island biogeography: Taking the long view of nature’s laboratories. *Science*, *357*(6354).

Zheng, H. (2015). Birth of the Yangtze River: age and tectonic-geomorphic implications. *National science review*, *2*(4), 438–453.

Zwickl, D. J., & Hillis, D. M. (2002). Increased taxon sampling greatly reduces phylogenetic error. *Systematic Biology*, *51*(4), 588–598.

# Figures

**Figure 1.** a) Data, priors, and their interaction in biogeographic dating for the estimation of divergence times. The framework facilitates the reciprocal influence of biological and geological data. b) Simulated scenario illustrating the splitting or rivers with surface uplift and the parallel evolution of a river plant lineage. Dotted lines represent known geological dates of uplift and river splits at 3 Ma and 1 Ma.

**Figure 2:** Demographic models used to generate sequence data and trees with *msprime*. a) Scenario without gene flow, where biological history mirrors geological events. b) Hypothetical scenario where biological and geological events do not match due to migration across landscape units (gene flow is allowed).

Figure 3. Geological priors specified for the two biological scenarios modeled. Red dotted line represents the true time for the node where taxon C splits from taxa A and B, and green dotted line represents the true time for the node where taxon A splits from taxon B.

**Figure 4:** Posterior distributions produced from 10 biogeographic-dating analyses conducted using different demographic models and geological priors. Purple: Yellow:

# Supplemental Information

**Table S1**: The absolute ages predicted for node 1 (taxa C and stem group AB), node 2 (taxa A and B), and the distribution of age ranges for the demographically manipulated node 2 (taxa A and B)

|  |  |  |  |
| --- | --- | --- | --- |
| **Model** | **Node 1 Absolute Age** | **Node 2 Absolute Age** | **Node 2 Age Distribution** |
| Informed uniform, no gene flow | 3.6051 | 2.1222 | 3.3801, 0.9284 |
| Informed normal, no gene flow | 3.6012 | 2.1295 | 3.2818, 1.0217 |
| Unknown, no gene flow | 3.5304 | 2.0042 | 3.5133, 0.5983 |
| Incorrect uniform, no gene flow | 3.633 | 2.6246 | 3.5475, 1.7175 |
| Incorrect normal, no gene flow | 3.6297 | 2.6844 | 3.5546, 1.9654 |
| Informed uniform, gene flow | 3.5067 | 1.3893 | 2.9633, 0.2038 |
| Informed normal, gene flow | 3.5074 | 1.403 | 2.9705, 0.2054 |
| Unknown, gene flow | 3.5198 | 1.4953 | 3.0648, 0.2794 |
| Incorrect uniform, gene flow | 3.5086 | 1.4076 | 2.9846, 0.2103 |
| Incorrect normal, gene flow | 3.5084 | 1.4034 | 2.9679, 0.2042 |

**Figures S1-S10:** Maximum clade credibility trees constructed for the ten biogeographic models tested in this study.