A dataset containing speciation rates, uncertainty, and presence-absence matrices for more than 34,000 vertebrate species

Juan D. Vásquez-Restrepo[1]

[1]*Laboratorio de Herpetología, Museo de Zoología "Alfonso L. Herrera", Universidad Nacional Autónoma de México, Ciudad de México, México. ORCID: 0000-0002-8306-408X.*

Corresponding author: juanda037@outlook.com

Acknowledgements

I would like to thank Julián A. Velasco and the *Instituto de Ciencias de la Atmósfera y Cambio Climático* at the *Universidad Nacional Autónoma de México*, for allowing me to use their informatic infrastructure, which was acquired supported by the DGAPA grant UNAM-PAPIIT IA201320.

Conflict of interest

The authors declare that they have no conflicts of interest.

Funding sources

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Biosketch

Juan D. Vásquez-Restrepo is a Colombian biologist interested in the study of large-scale macroecological and macroevolutionary patterns of Neotropical species, as well as the eco-evolutionary processes driving their diversity.

A dataset containing speciation rates, uncertainty, and presence-absence matrices for more than 34,000 vertebrate species

Abstract

Motivation: Today, the large amount and availability of phylogenetic data have allowed the exploration of different types of patterns at various evolutionary and spatial scales. However, the calculation of some metrics used for this purpose can be computationally demanding and complex, such as those related to diversification rates. Therefore, the existence of a database that provides such information can streamline the process. It also makes science fairer and more accessible for those without access to great computational power or advanced programming skills.

Main types of variables contained: Continuous and binary data depicting the speciation rates and presence-absence matrices, respectively, for more than 34,000 vertebrate species.

Spatial location and grain: Worldwide at a specific level.

Time period and grain: Present, but based on phylogenetic hypotheses for groups with between 300–100 Myr of evolutionary history.

Major taxa and level of measurement: Vertebrates, including the Lissamphibia subclass (amphibians), the Squamata suborder (reptiles), and the classes Aves (birds), Mammalia (mammals), and Chondrichthyes (sharks), all of them at the species level.

Software format: Various, including Excel files (.xlsx), nexus files (.nex), tree files (.tre), text files (.txt), raster files (.tiff), and RData files (.RData).

1. Introduction

It is not a secret that higher-income countries have had a predominance in biodiversity research, mainly driven by their access to technological and economic resources (Tydecks et al., 2018; Hughes et al., 2021; Rodríguez et al., 2022). On the other hand, it is also true that increasingly, more researchers and institutions are advocating for open science policies that make knowledge and science fairer and accessible (Chavan and Penev, 2011), particularly for those in countries and/or institutions with limited or no access to strong financial support or technological

infrastructure (UNESCO, 2021). This has driven science towards shifting to open-access policies not just for the publication of scientific papers but also its associated data. From a Popperian view, publishing biodiversity data allows replicability and falsifiability, facilitating and promoting the development of new methods and hypothesis, making it possible for science to move forward.

In recent years, data papers (*sensu lato*), namely, those properly classified as data papers or research articles that make available large amounts of data, have become more common in the scientific literature. These resources provide tons of valuable information, both raw and pre-processed, facilitating research that requires such types of data but is not focused on generating it, thus, reducing the time and cost associated with the publishing process. In a certain way, data papers serve as a solution to the "dog in the manger" metaphor, where a person who has no need for, or ability to use a possession that would be of use or value to others, prevents them from having it. Perhaps, the most successful example of that is the WorldClim data, which, between its first and second versions, has accumulated more than 20,000 citations in only 18 years (Hijmans et al., 2005; Fick and Hijmans, 2017).

Based on the above, I aim to make available a dataset on diversification metrics for five of the major vertebrate radiations and their respective presence-absence matrices. Given the processing of this data can be computationally demanding and requires intermedium to advanced programming skills, I believe this dataset is a valuable source of information for those interested in macroecology and macroevolution. These data were generated during an exploratory analysis motivated by genuine curiosity, but for which I do not have a practical purpose in the short term. Thus, I prefer to make it available for the community. Here, I will briefly present a description of the data, but I will not delve into the causes or consequences of the observed biodiversity patterns in detail.

2. Materials and methods

2.1. Phylogenetic sampling

I used available "fully-sampled" time-calibrated super-trees for five monophyletic groups of living vertebrates: amphibians (subclass Lissamphibia, 7238 species), reptiles (order Squamata, 9754 species), birds (class Aves, 9993 species), mammals (class Mammalia, 5911 species), and sharks (class Chondrichthyes, 1190 species). Trees correspond to those in Jetz et al. (2012), Tonini et al. (2016), Upham et al. (2016), Jetz and Pyron (2018), and Stein et al. (2018), and were downloaded from the VertLife website (http://vertlife.org). Since these phylogenies were generated using a combined method that relies on a base topology inferred from species with DNA and the imputation of those without genetic information, their consensus are highly polytomic. Therefore, to account for the phylogenetic uncertainty, I subsampled 100 random dichotomized trees for each group and ran the subsequent analyses on this set of trees (Rangel et al., 2015).

2.2. Diversification rates

In order to compare their variability and provide two different alternative metrics, I calculated diversification rates using both a deterministic (DivRate) and a probabilistic approach (BAMM), which are described below:

2.2.1. DivRate

The DivRate index aims to capture clade-wide "diversification" estimates based on the branching pattern along birth-only and homogeneous birth-death trees (Jetz et al., 2012; Belmaker and Jetz, 2015). Despite its name, DivRate does not correspond to a net diversification estimate (speciation + extinction) but only speciation (Title and Rabosky, 2018). DivRate is the inverse of Evolutionary Distinctiveness (ED), which is a measure of the evolutionary isolation of a species (Redding and Mooers, 2006; Isaac et al., 2007; Title & Rabosky, 2018). The ED can be calculated using either the "fair proportion" approach, dividing branch lengths by the number of descendant species, or the "equal splits" approach, dividing branch lengths by the number of descendant clades (Wicke and Steel, 2020). For this case, I employed both of them. To calculate the ED, I used the function *evol_distinct* in R's package *phyloregion* (Daru et al., 2020), which is computationally more efficient than other libraries.

2.2.2. BAMM

The Bayesian Analysis of Macroevolutionary Mixtures (BAMM) uses Markov Chain Monte Carlo (MCMC) to model complex dynamics of speciation (λ) and extinction (μ), allowing for heterogeneity in evolutionary rates (Rabosky, 2014). For each of the 100 trees in the five vertebrate groups, I ran a BAMM analysis of 2×10^7 generations, with four MCMC chains and sampling every 1000 iterations for a total of 20,000 samples, using BAMM v.2.5.0 (Rabosky, 2014). Default control file was implemented with custom priors obtained for each tree using the function *setBAMMpriors* in R's package *BAMMtools* v.2.1.10 (Rabosky et al., 2014). Due to the computationally demanding nature of this analysis, it was run on a server equipped with 40 Intel Xeon Gold 6230 2.10 GHz processors and 125 GB of RAM. Additionally, to optimize resources, each set of 100 runs was partitioned into four parallel processes of 25 replicates. The BAMM outputs were analyzed and processed locally using RStudio v.2022.7.2 (RStudio Team, 2022) with R v.4.0.5 (R Core Team, 2021). A script for running multiple BAMM analyses automatically can also be found at GitHub (https://github.com/VR-Daniel/Multiple-BAMM).

Because trees were randomly dichotomized, they differ in topology, therefore, they are unlikely to converge to the same log-likelihood value as expected in replicates of a fixed tree. Nevertheless, if each one converges independently to its most optimal solution, I expect to observe a convergence pattern rather than random variation (Fig. S1). To validate the MCMC convergence, I visually assessed it through a mean log-likelihood trace plot using the 100 tree replicates per group, discarding the first 10% of the iterations as burn-in (Fig. S2). In addition, I evaluated the Effective Sample Size (ESS) for both the number of shifts and log-likelihood parameters using the *effectiveSize* function in the *coda* library (Plummer et al., 2006). A major issue when analyzing such amount of data is that, as tree size increases, the number of generations required to adequately sample the posterior distribution also increases, significantly but unpredictably. Under the conditions in which I ran the analyses, it took more than 4300 hours (ca. 6 months) to complete. Thus, increasing the number of generations an n-number of times would be unfeasible in terms of computational cost-benefit. However, given that the idea behind the ESS is checking for sample independence, I consider running 100 independent BAMM analyses may help us to explore different peaks of the parameter landscape by generating a distribution frequency, compensating for the ESS when the expected minimum (100–200) is not met in individual runs.

2.3. Matrices of presence-absence

To facilitate the data spatialization, I generated presence-absence matrices for each vertebrate group using the distribution polygons provided by the IUCN Red List (IUCN, https://www.iucnredlist.org). For amphibians (7958 spp.), reptiles (9821 spp.), mammals (5883 spp.), and sharks (1201 spp.), I gathered them from the most recent IUCN update (2023-1), while for birds (10920 spp.), due to data accession is restricted, I employed a previously acquired dataset from a prior update (2021-3). First, I fixed the invalid geometries of the polygons using the built-in *fix geometries* algorithm in QGIS v.3.28 with the structure method, which attempts to fix the polygons' topology while maintaining their node integrity. Then, I obtained the presence-absence matrices through the function *lets.presab* in the *letsR* v.5.0 R package (Vilela and Villalobos, 2015), at a spatial resolution of one degree in geographical coordinates (WGS84). In order to spatialize the speciation rates considering the species richness by cell, I standardized it as the ratio between the mean diversification rates of the species present in each cell, over the weighted mean of all cells using species richness as the weighting variable. Thus, values above zero indicate higher speciation rates than the average, while those below, lower than the mean.

2.4 Taxonomic harmonization

Because the binomial names of the species may vary over time and datasets, I have included the GBIF's usage key as a unique numeric identifier for each taxon. The GBIF keys are invariant IDs associated with specific taxonomic ranks, allowing and facilitating the species tracking regardless of whether they have invalid names. For this purpose, I utilized the GBIF's built-in *Look Up Tool* to match the species names from phylogenetic trees and IUCN polygons with GBIF's backbone taxonomy. This tool performs a search at three levels: 1) when the provided name and the GBIF's name are the same (EXACT); 2) when the provided name and the GBIF's name slightly differ (FUZZY), for example, when the name's ending has changed to comply with grammatical gender concordance; and 3) when there is no match (HIGHERRANK). For the diversification rates datasets, 11 amphibians (0.15%), 6 reptiles (0.06%), 2 birds (0.02%), 9 mammals (0.15%), and 6 sharks (0.5%) were missing from GBIF's backbone taxonomy. On the other hand, for the presence-absence matrices, 26 amphibians (0.33%), 4 reptiles (0.04%), 1 bird (0.01%), 2 mammals (0.03%), and 2 sharks (0.17%) had no match. Posteriorly, non-matching species

between datasets were subjected to a string similarity test to reduce the number of errors due to misspelled names or ambiguities in GBIF's backbone taxonomy. This test employs the Optimal String Alignment algorithm based on the Levenshtein distance to count the number of deletions, insertions, and substitutions required to turn a B string into A, but allowing the transposition of adjacent characters. For this, I built a custom script which is available through GitHub (https://github.com/VR-Daniel/taxTools).

3. Results

For both DivRate and BAMM, tip speciation rates for all vertebrate groups tended to show relatively slow rates, exhibiting an almost generalized right-skewed distribution (Fig. 1). However, values between metrics varied significantly, with those of BAMM presenting less dispersion (Fig. 2A). Despite the high phylogenetic uncertainty for the divergence times within groups, the variability in speciation rates was consistently low for two of the three metrics (Fig. 2B). At least for the DivRate based on the fair proportion approach and BAMM, the coefficient of variation for the 75% of the data remained below the 10% threshold, except for sharks. On the other hand, the presence-absence matrices successfully recovered the well-known latitudinal patterns of diversity for the five vertebrate groups (Jenkins et al., 2013; Raz et al., 2023), exhibiting a general trend of higher species richness in tropical regions (Fig. 3), although showing specific variations in the main areas with the highest species richnes. Due to differences in data availability within each data source, diversification metrics include fewer species (34,086) than presence-absence matrices (35,783). However, there were at least 6532 amphibians, 8957 reptiles, 8373 birds, 5494 mammals, and 1011 sharks common between both datasets. A summary list of all the files included in this work is presented in Table 1.

Regarding the geographical patterns of diversification rates, I noticed that these can change significantly depending on the metric used (Fig. 3). For example, in amphibians, the New World and the Palearctic region exhibit the highest relative average speciation rates using values from BAMM, while for DivRate, these are also concentrated towards the Afrotropical region. Something similar occurs in reptiles, where the New World and the Australian and Oriental regions show the highest values according to BAMM, with also high speciation rates towards the

Afrotropical region according to DivRate. In mammals, the change is quite notable, as according to BAMM's average rates, marine mammals of the tropical belt exhibit higher speciation rates, even more than continental ones, whereas with DivRate, continental mammals of the Neo- and Afrotropics show elevated rates, as do marine mammals of temperate zones. Birds in general exhibit a pattern of low and relatively homogeneous rates according to BAMM, with some peaks in the New World, the Palearctic region, the eastern part of the Afrotropical region, and some marine areas of high latitudes, while for DivRate, rates are higher, including also the Australian and Oriental regions. Finally, for sharks, BAMM shows relatively low rates, except for some regions at higher latitudes, while according to DivRate, the highest relative speciation rates are concentrated towards the Indian and Pacific oceans.

4. Discussion

While the data presented here do not represent the first attempt to calculate diversification rates in vertebrates, it is an attempt to present a pre-processed dataset compiled, organized, and publicly available. This dataset is intended to be an asset for the study of patterns in macroecology and macroevolution, as well as systematic ecology, since calculating diversification rates can be computationally expensive and time-consuming, also requiring intermediate to advanced programming skills. Here, the process is simplified by providing a well-supported series of results for five groups of vertebrates, which can be employed to explore a wide variety of theoretical and applied questions. For example, through simple filters, it is possible to associate diversification rates and/or lineage age information with geographic or climatic data, enabling their spatialization for the search of patterns at the geographical level. However, when spatializing the data, it is important to consider their right-skewed nature and the difference in the number of species per pixel, as this can affect measures of central tendency such as the mean. Therefore, it is more appropriate to use alternatives such as the median or the geometric mean, as well as standardizing the values to relative measures. Similarly, the way data is presented can drastically influence its interpretation, because if its range of variation is small, geographical patterns may be barely visible on a continuous scale but noticeable on discrete ones (such as quantiles).

It is also important to note that diversification metrics and presence-absence matrices may differ in their taxonomy due to variations in publication dates and data sources. Therefore, users must harmonize the taxonomy of their species of interest to ensure proper data matching. The simplest option we suggest is to use a unique ID that remains constant, unlike binomial names, such as the GBIF usage key we provided. The harmonizing IDs we included seem to be useful for most cases since they match most species between datasets. However, harmonizing biodiversity data is an arduous task, and the IDs we provided were found using a heuristic method, so, some specific taxa present in both datasets may not be properly matched due to differences in GBIF's backbone taxonomy. Furthermore, when diversification rates are calculated, it is common for no measure of uncertainty to be presented. In contrast, in this dataset, the methodological approach including replicates can be useful for this purpose. It is important to notice that diversification metrics may vary depending on the number of lineages included in the tree, that is, they are sensitive to the non-random incomplete taxon sampling. This is why metrics derived from super-trees accounting for most species may be a good choice to large-scale analyses than those from pruned trees. Users should be aware of potential limitations associated with this dataset, whereby we encourage them first to get familiarized with the metrics calculation and interpretation.

Data availability statement

As a data paper, all the data used and generated outputs can be found in Figshare: http://doi.org/10.6084/m9.figshare.24894363.

References

Belmaker, J. & Jetz, W. (2015). Relative roles of ecological and energetic constraints, diversification rates and region history on global species richness gradients. *Ecology Letters*, *18*(6), 563–571. https://doi.org/10.1111/ele.12438

Chavan, V. & Penev, L. (2011). The data paper: a mechanism to incentivize data publishing in biodiversity science. *BMC Bioinformatics*, *12*(S15). https://doi.org/10.1186/1471-2105-12-s15-s2

Daru, B. H., Karunarathne, P. & Schliep, K. (2020). phyloregion: R package for biogeographical regionalization and macroecology. *Methods in Ecology and Evolution*, *11*(11), 1483–1491. https://doi.org/10.1111/2041-210x.13478

Fick, S. E. & Hijmans, R. J. (2017). WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas. *International Journal of Climatology*, 37(12), 4302–4315. https://doi.org/10.1002/joc.5086

Hijmans, R. J., Cameron, S. E., Parra, J. L., Jones, P. G. & Jarvis, A. (2005). Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology, 25*(15), 1965–1978. https://doi.org/10.1002/joc.1276

Hughes, A. C., Orr, M. C., Ma, K., Costello, M. J., Waller, J., Provoost, P., Yang, Q., Zhu, C. & Qiao, H. (2021). Sampling biases shape our view of the natural world. *Ecography, 44*(9), 1259–1269. https://doi.org/10.1111/ecog.05926

Isaac, N. J. B., Turvey, S. T., Collen, B., Waterman, C. & Baillie, J. (2007). Mammals on the EDGE: Conservation Priorities Based on Threat and phylogeny. *PLOS ONE, 2*(3), e296. https://doi.org/10.1371/journal.pone.0000296

Jenkins, C. N., Pimm, S. L. & Joppa, L. (2013). Global patterns of terrestrial vertebrate diversity and conservation. *Proceedings of the National Academy of Sciences of the United States of America, 110*(28). https://doi.org/10.1073/pnas.1302251110

Jetz, W. & Pyron, R. A. (2018). The interplay of past diversification and evolutionary isolation with present imperilment across the amphibian tree of life. *Nature Ecology and Evolution, 2*(5), 850–858. https://doi.org/10.1038/s41559-018-0515-5

Jetz, W., Thomas, G. H., Joy, J. B., Hartmann, K. & Mooers, A. Ø. (2012). The global diversity of birds in space and time. *Nature, 491*(7424), 444–448. https://doi.org/10.1038/nature11631

Plummer, M., Best, N., Cowles, K. & Vines, K. (2006). CODA: Convergence Diagnosis and Output Analysis for MCMC. *R News*, 6(1), 7–11.

Rabosky, D. L. (2014). Automatic detection of key innovations, rate shifts, and Diversity-Dependence on phylogenetic trees. *PLOS ONE, 9*(2), e89543. https://doi.org/10.1371/journal.pone.0089543

Rabosky, D. L., Gründler, M. C., Anderson, C., Title, P. O., Shi, J. J., Brown, J. W., Huang, H. & Larson, J. G. (2014). BAMMtools: an R package for the analysis of evolutionary dynamics on phylogenetic trees. *Methods in Ecology and Evolution, 5*(7), 701–707. https://doi.org/10.1111/2041-210x.12199

Rangel, T. F., Colwell, R. K., Graves, G. R., Fučíková, K., Rahbek, C. & Diniz-Filho, J. A. F. (2015). Phylogenetic uncertainty revisited: Implications for ecological analyses. *Evolution*, *69*(5), 1301–1312. https://doi.org/10.1111/evo.12644

Raz, T., Allison, A., Ávila, L. J., Bauer, A. M., Böhm, M., De Oliveira Caetano, G. H., Colli, G. R., Doan, T. M., Doughty, P., Grismer, L., Itescu, Y., Kraus, F., Martins, M. R., Morando, M., Murali, G., Nagy, Z. T., Campos Nogueira, C., Novosolov, M., Oliver, P. M., Passos, P., Pincheira-Donoso, D., Sindaco, R., Slavenko, A., Torres-Carvajal, O., Uetz, P., Wagner, P., Zimin, A., Roll, U. & Meiri, S. (2023). Diversity gradients of terrestrial vertebrates – substantial variations about a common theme. *Journal of Zoology, Early View*. https://doi.org/10.1111/jzo.13130

R Core Team. (2021). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org

Redding, D. W. & Mooers, A. Ø. (2006). Incorporating Evolutionary Measures into Conservation Prioritization. *Conservation Biology*, *20*(6), 1670–1678. https://doi.org/10.1111/j.1523-1739.2006.00555.x

Rodríguez, J. P., Sucre, B., Mileham, K., Sánchez-Mercado, A., de Andrade, N., Bezeng, S. B., Croukamp, C., Falcato, J., Borboroglu, P. G., González, S., González-Ciccia, P., González-Maya, J. F., Kemp, L. V., Kusrini, M. D., López-Gallego, C., Luz, S., Menon, V., Moehlman, P. D., Raimondo, D., Sillero-Zubiri, C., Street, W., Vincent, A. & Xie, Y. (2022). Addressing the Biodiversity Paradox: Mismatch between the Co-Occurrence of Biological Diversity and the Human, Financial and Institutional Resources to Address Its Decline. *Diversity*, *14*(9), 708. https://doi.org/10.3390/d14090708

RStudio Team. (2022). RStudio: Integrated Development Environment for R. PBC, Boston, United States. https://posit.co

Stein, R. W., Mull, C. G., Kuhn, T. S., Aschliman, N. C., Davidson, L. N. K., Joy, J. B., Smith, G. J., Dulvy, N. K. & Mooers, A. Ø. (2018). Global priorities for conserving the evolutionary history of sharks, rays and chimaeras. *Nature Ecology and Evolution*, *2*(2), 288–298. https://doi.org/10.1038/s41559-017-0448-4

Title, P. O. & Rabosky, D. L. (2019). Tip rates, phylogenies and diversification: What are we estimating, and how good are the estimates? *Methods in Ecology and Evolution*, *10*(6), 821–834. https://doi.org/10.1111/2041-210x.13153

Tonini, J. F. R., Beard, K. H., Ferreira, R. B., Jetz, W. & Pyron, R. A. (2016). Fully-sampled phylogenies of squamates reveal evolutionary patterns in threat status. Biological *Conservation*, *204*, 23–31. https://doi.org/10.1016/j.biocon.2016.03.039

Tydecks, L., Jeschke, J. M., Wolf, M., Singer, G. & Tockner, K. (2018). Spatial and topical imbalances in biodiversity research. *PLOS ONE*, *13*(7), e0199327. https://doi.org/10.1371/journal.pone.0199327

UNESCO. (2021). *UNESCO Recommendation on Open Science*. United Nations Educational, Scientific and Cultural Organization. https://doi.org/10.54677/mnmh8546

Upham, N. S., Esselstyn, J. A. & Jetz, W. (2019). Inferring the mammal tree: Species-level sets of phylogenies for questions in ecology, evolution, and conservation. *PLOS Biology*, *17*(12), e3000494. https://doi.org/10.1371/journal.pbio.3000494

Vilela, B. & Villalobos, F. (2015). letsR: a new R package for data handling and analysis in macroecology. *Methods in Ecology and Evolution*, *6*(10), 1229–1234. https://doi.org/10.1111/2041-210x.12401

Wicke, K. & Steel, M. (2019). Combinatorial properties of phylogenetic diversity indices. *Journal of Mathematical Biology*, *80*(3), 687–715. https://doi.org/10.1007/s00285-019-01438-0

Table 1. Files' structure, names, formats, and content included in this data paper.

File	Format	Description
GROUP	.zip	Diversification rates dataset main folder by GROUP
Branch lengths		Subfolder
GROUP_BRANCH LENGTHS	.xlsx	Branch lengths file for the 100 trees and summary statistics by group (in Species/Myr)
Speciation rates[1]		Subfolder
GROUP_BAMM	.xlsx	BAMM estimates file for the 100 trees and summary statistics by group (in Species/Myr)
GROUP_DivRate_ES	.xlsx	DivRate "Equal Splits" estimates file for the 100 trees and summary statistics by group (in Species/Myr)
GROUP_DivRate_FP	.xlsx	DivRate "Fair Proportion" estimates file for the 100 trees and summary statistics by group (in Species/Myr)
GROUP_BAMM FILES	.zip	Raw BAMM outputs, including the runs info, mcmc outputs, event data, chain swap, and the 100 species trees
GROUP_SAMPLED TREES	.nex	Species trees
Presence-Absence	.zip	Presence-Absence dataset main folder
paMa_GROUP[2]	.RData	Presence-Absence matrices by GROUP
Rasters	.zip	Diversification rates and richness raster files
Speciation rates		Subfolder
GROUP_BAMM	.tiff	Raster file for the standardized BAMM speciation rates (at 1°)
GROUP_DivRate_FP	.tiff	Raster file for the standardized DivRate "Fair Proportion" speciation rates (at 1°)
Richness		Subfolder
GROUP_Richness	.tiff	Raster file for the species richness (at 1°)
Summaries		Subfolder
paMa_summ_GROUP_METRIC	.RData	Summarized Presence-Absence matrices by pixel, including mean speciation rates, species richness, and standardized values

Phylo_Taxonomy	.xlsx	Harmonized taxonomy for the diversification rates dataset
paMA_Taxonomy	.xlsx	Harmonized taxonomy for the Presence-Absence dataset

[1]Herein, we focus on speciation rather than extinction rates, because when based only on living-species phylogenetic data, the latter can be biased despite being correlated with the true rates (Rabosky, 2014). However, since we included all the BAMM output files, if interested, users can also easily extract additional information such as the extinction rates or the a posteriori maximum probability shift configuration for each run, using the command *getEventData* in the *BAMMtools* package (see the package's help files for further details).

[2]Presence-Absence matrices for each group are provided as .RData files in order to reduce the file sizes. These data can be directly imported into R using the command *load* and easily exported to comma separated values (.csv) using *write.csv*.

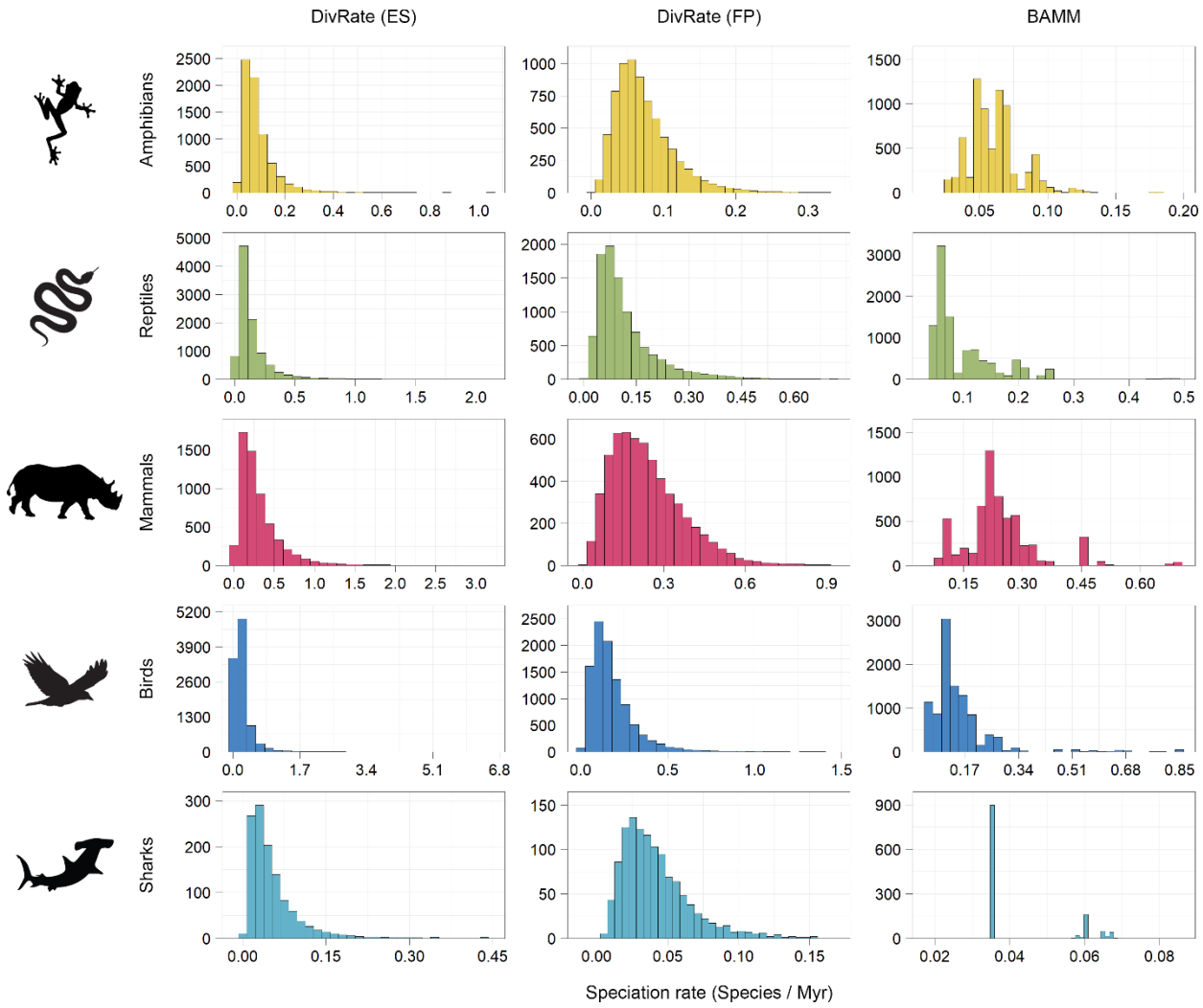


Figure 1. Histograms showing the distribution of the speciation rates for the five vertebrate groups according to the metric used.

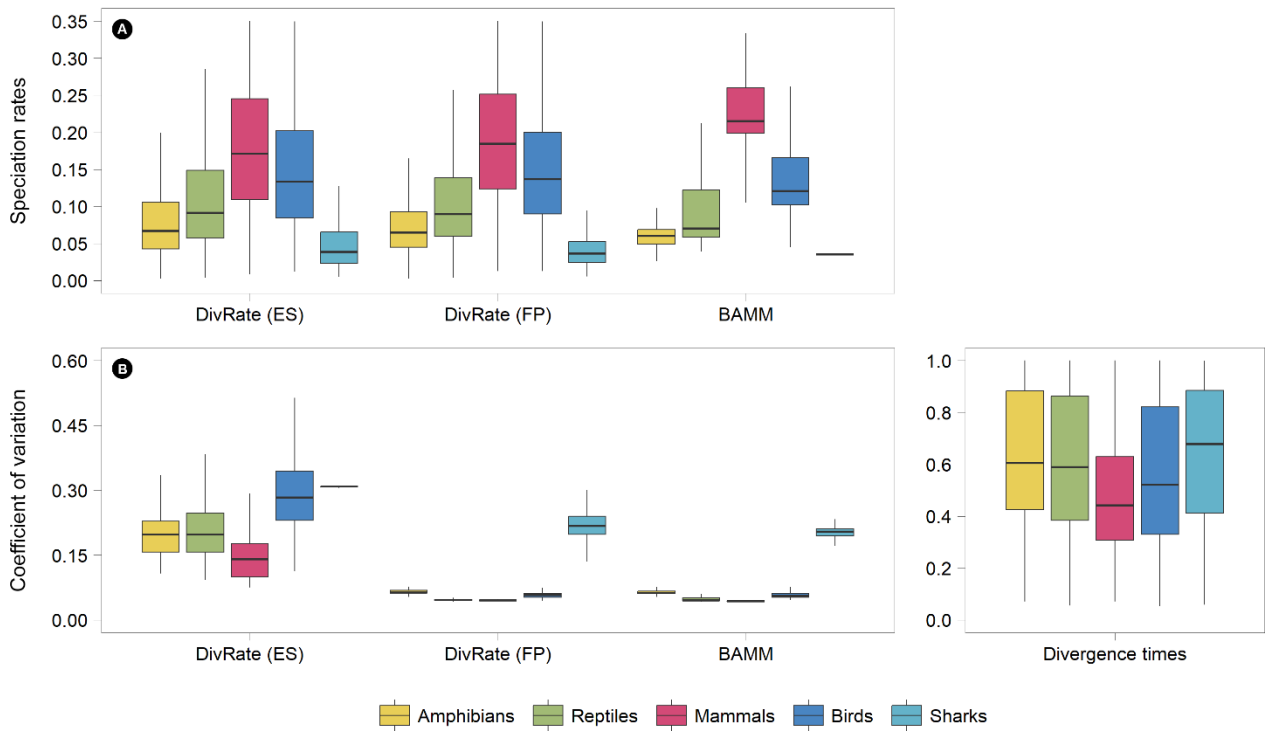


Figure 2. Boxplot showing the variation in the speciation rates for the five vertebrate groups according to the metric used (A), and coefficient of variation for the 100 replicates per group, metric, and divergence times (B).

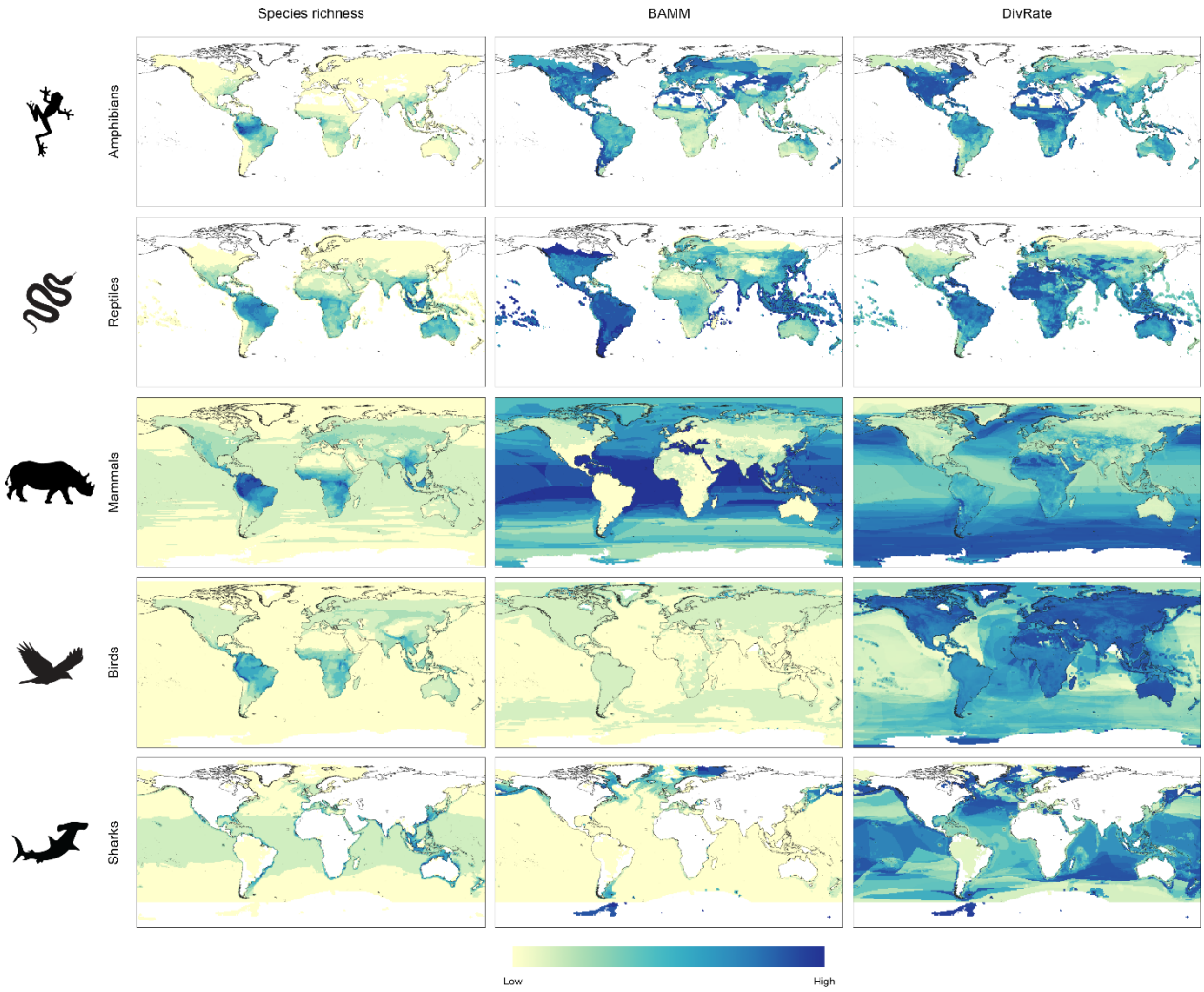
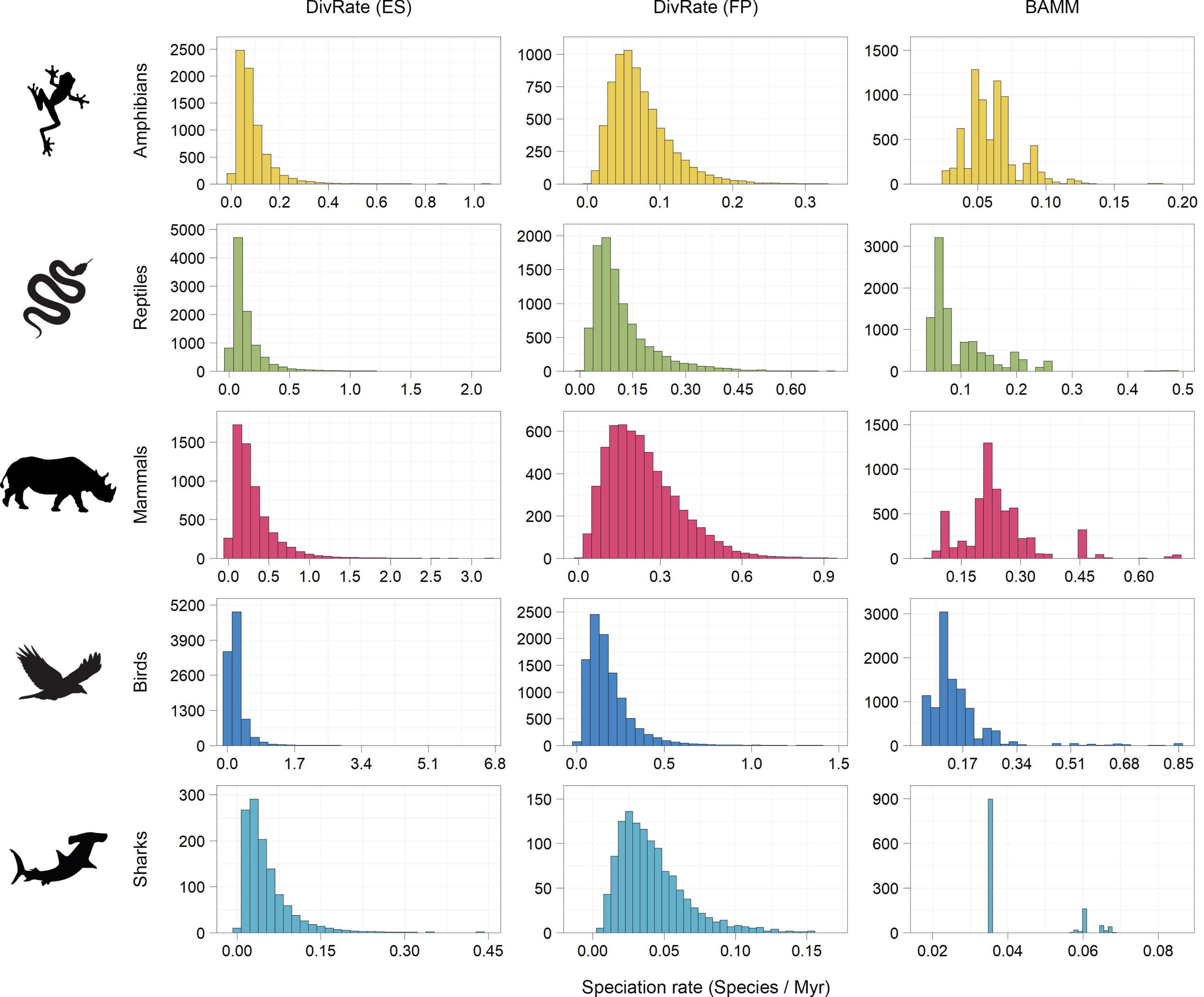
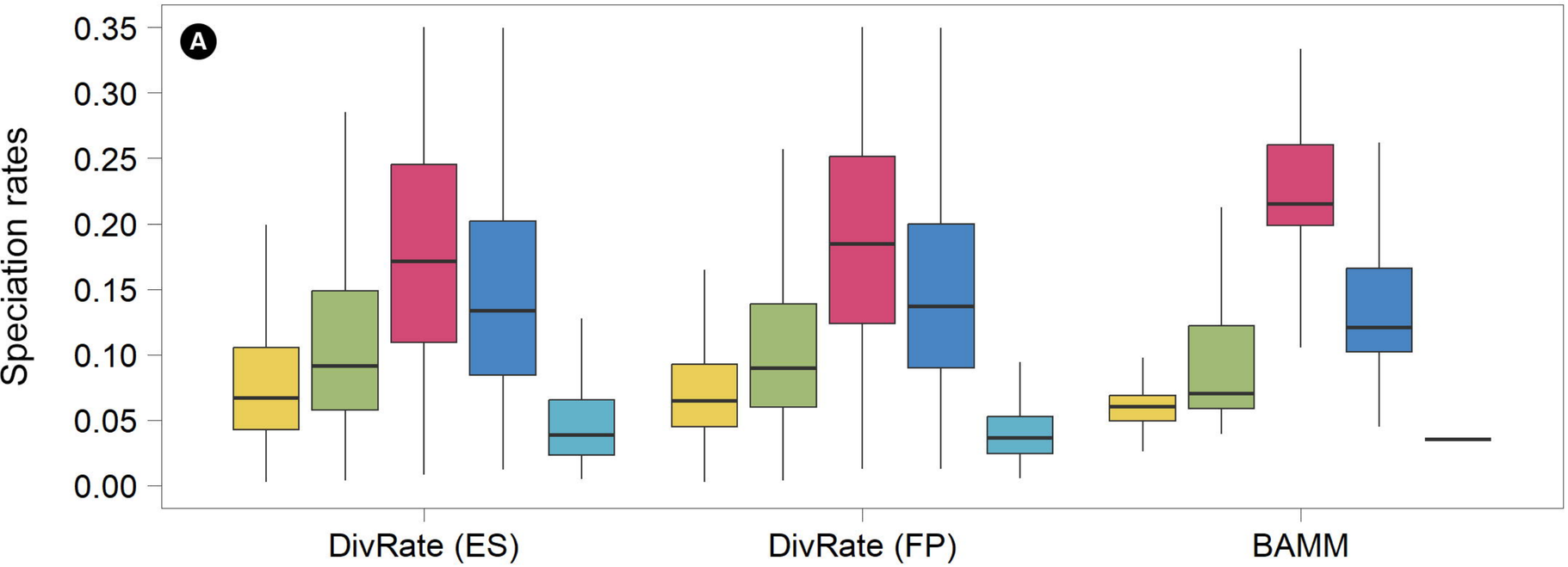
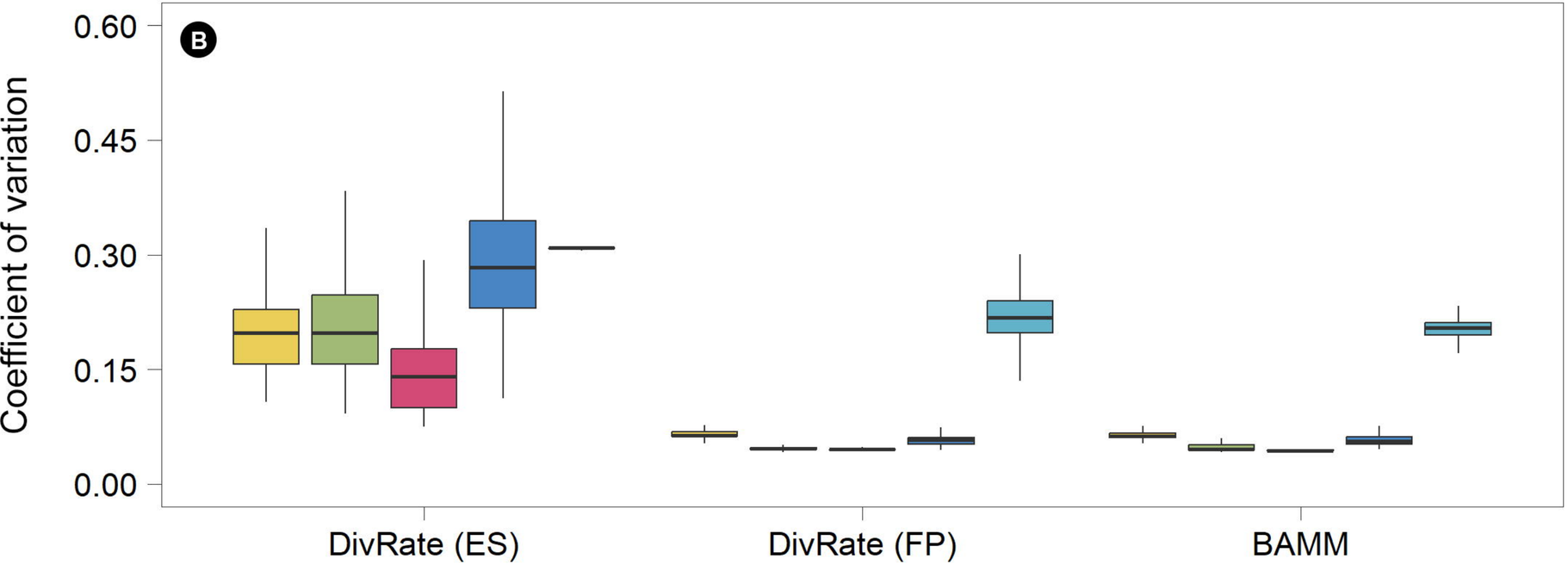
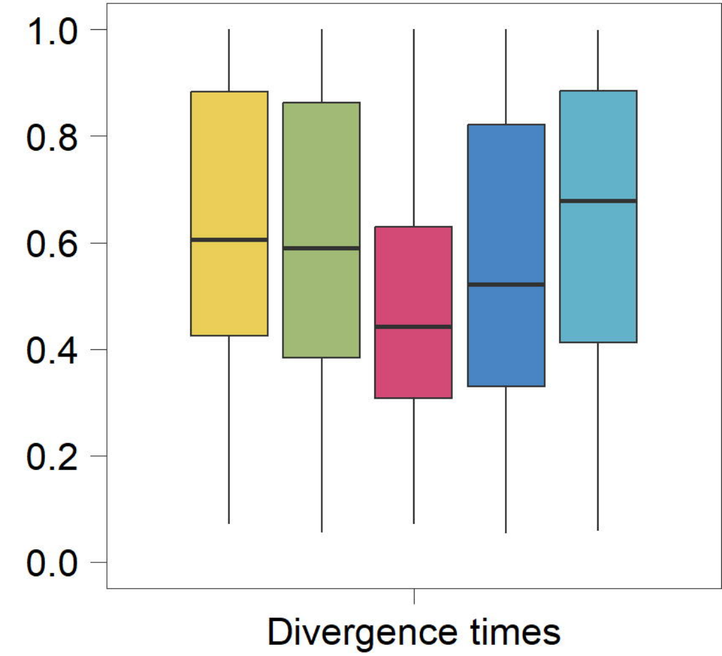


Figure 3. Diversity patterns based on the presence-absence matrices for each vertebrate group. For the speciation rates, maps are colored quantile-based using 10 bins.

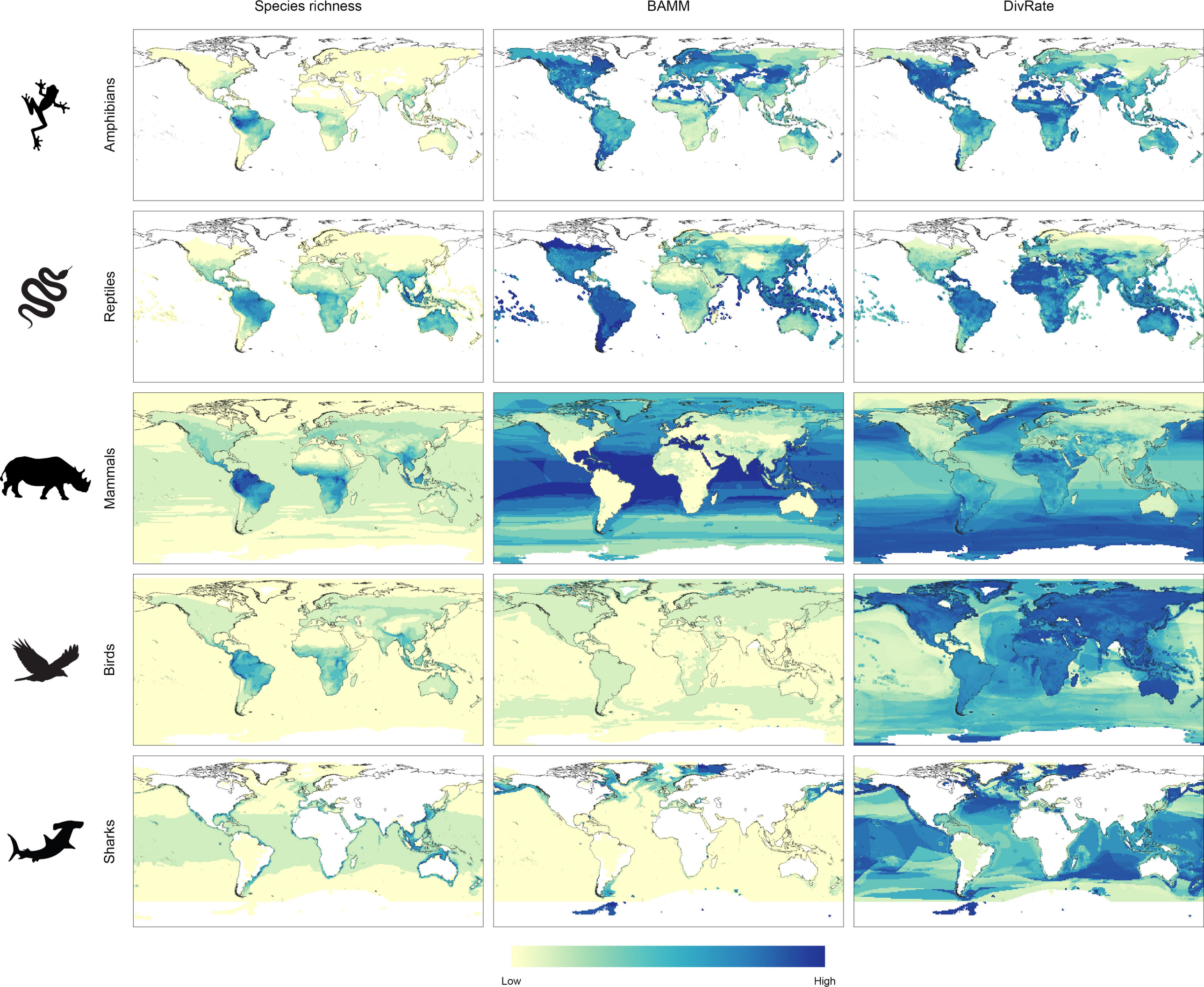

DivRate (ES)
DivRate (FP)
BAMM
Amphibians
Reptiles
Mammals
Birds
Sharks
Speciation rate (Species / Myr)







Amphibians Reptiles Mammals Birds Sharks


Species richness
BAMM
DivRate
Amphibians
Reptiles
Mammals
Birds
Sharks
Low
High