Contents lists available at ScienceDirect

MethodsX

The use of an integrated framework combining eco-evolutionary data and species distribution models to predict range shifts of species under changing climates ☆

Wen-Xun Lu, Guang-Yuan Rao*

School of Life Sciences, Peking University, Beijing, China

ARTICLE INFO

ABSTRACT

Species distribution models (SDMs) are powerful tools that can predict potential distributions of species under climate change. However, traditional SDMs that rely on current species occurrences may underestimate their climatic tolerances and potential distributions. To address this limitation, we developed an integrated framework that incorporates eco-evolutionary data into SDMs. In our approach, the fundamental niches of species are constructed by their realized niches in different periods, and those fundamental niches are used to predict potential distributions of species. Our framework includes multiple phylogenetic analyses, such as niche evolution rate estimation and ancestral area reconstruction. These analyses provide deeper insights into the responses of species to climate change. We applied our approach to the *Chrysanthemum zawadskii* species complex to evaluate its efficacy through comprehensive performance evaluations and validation tests. Our framework can be applied broadly to species with available phylogenetic data and occurrence records, making it a valuable tool for understanding species adaptation in a rapidly changing world.

- Integrating the niches of species in different periods estimates more complete climatic envelopes for them.
- Combining eco-evolutionary data with SDMs predicts more comprehensive potential distributions of species under climate change.
- Our framework provides a general procedure for species with phylogenetic data and occurrence records.

Specifications Table

|	
Subject area:	Agricultural and Biological Sciences
More specific subject area:	Species distribution modeling
Name of your method:	Phylogenetically-informed species distribution model
Name and reference of original method:	Lu, W.-X., Wang, Z.-Z., Hu, X.-Y., & Rao, G.-Y. Incorporating eco-evolutionary information into species distribution models provides comprehensive predictions of species range shifts under climate change. Sci. Total Environ. 912 (2024) 169,501. https://doi.org/10.1016/j.scitotenv.2023.169501.
Resource availability:	R software

Method details

The overall workflow of our framework is depicted in Fig. 1. Our analysis process involves the following steps including data collection, data preprocessing, standard species distribution modeling (SDM), phylogenetically-informed SDM, and additional phylogenetic analysis. All analyses within this study were conducted using R *v.* 4.1.3 [1].

Data collection

Our framework requires three types of data: occurrence records of species, the dated phylogenetic tree of the target species and its closely related species, and predictor variables for modeling (e.g., climatic and topographic variables). The distribution data of species can be obtained through field surveys, museum records, related literature, or citizen science initiatives, which depend on specifics of the target species. There are multiple ways to acquire the required dated phylogenies. We can download molecular data of species from the National Center for Biotechnology Information or other databases and then use those data as inputs to reconstruct dated phylogeny with some phylogenetics software, e.g., BEAST [3], PAML [4], and others. In addition, previous studies can also provide the required phylogenetic trees. Notably, some software such as SoTree [5] and V.PhyloMaker2 [6] can generate phylogenies for vascular plants according to the species names provided. Predictor variables for modeling are usually downloaded from some public databases including WorldClim [7], ENVIREM [8], EarthEnv project [9], Oscillayers [10], and so on.

Data preprocessing

In species distribution modeling, it is necessary to filter and screen the data to improve model quality [11]. Specifically, data preprocessing can be divided into two parts in our approach: reduction of spatial autocorrelation in occurrence records and selection of predictor variables.
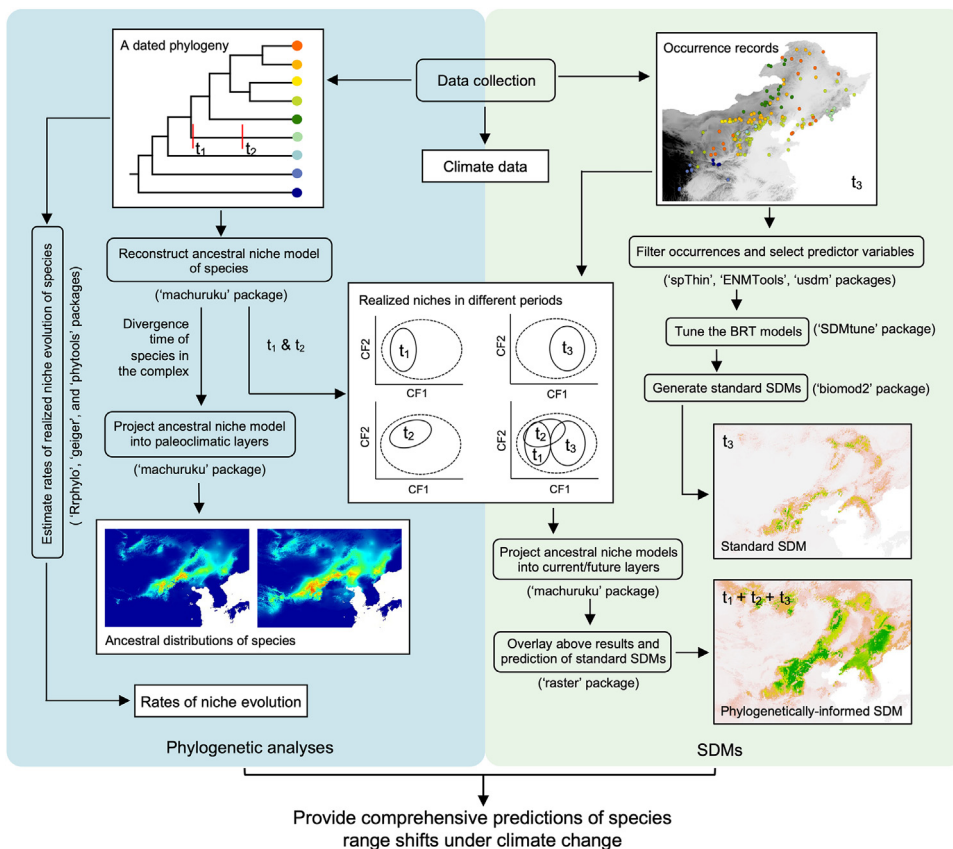
Fig. 1. The overall workflow of the integrated framework which combines eco-evolutionary data with species distribution models (SDMs) and R packages used in this study. Adopted from Lu et al. [2]. The ancestral realized niches of the target species contain two past periods, the early (t_1) and middle stages (t_2) in the course of its evolution. The realized niches of two past periods (t_1 and t_2) and the current period (t_3) are used to calibrate the "fundamental" niche of the target species. CF1 and CF2 indicate climatic factor 1 and climatic factor 2, respectively.

When performing SDMs, reducing spatial autocorrelation between occurrence records of species is crucial [11]. We here recommended two methods to minimize spatial autocorrelation in occurrence records of species. One of them is filtering the occurrence points of species to keep pairwise distances of their occurrences greater than a proper distance. Filtering of species occurrences can be performed using the 'spThin' R package [12]. Except for filtering species records, we also suggested implementing a cross-validation procedure to partition the occurrence records into training and test datasets when performing modeling analyses. This step, as proposed by previous studies [11,13], can enhance the spatial independence between training and test data, thereby improving the robustness and validity of our modeling results.

Because constructing models with highly correlated predictor variables may lead to some detrimental outcomes, such as model overfitting and debated model extrapolation [11,14], variables with high correlation should be removed from the model. A common method used to evaluate the multicollinearity of predictor variables is to calculate Pearson's correlation coefficients (r) between different variables. This assessment can be performed using the 'ENMTools' R package [15], and variables exhibiting low pairwise correlations (e.g., $|r| < 0.80$) should be selected to avoid collinearity issues.

Standard species distribution modeling

For comparisons with SDMs that integrate phylogenetic information (phylogenetically-informed SDMs), we conducted standard SDMs for the target species in our framework. Notably, previous studies indicated that splitting species which consist of populations with clear genetic and ecological differentiation into subunits could improve potential distribution predictions for them [16–18]. Therefore, we suggested determining whether models below species level are needed to predict species distributions for the target species before performing SDMs. There are many algorithms for predicting species distribution, which could be selected according to specific needs. The well-tuned boosted regression trees (BRT) model is identified as one of the best algorithms for SDM, and some studies suggested that the BRT model even outperforms the ensemble model which integrates multiple modeling algorithms [19,20]. Thus, we chose the well-tuned BRT model for standard SDMs in our framework. The subsequent modeling process is also exemplified by the BRT model.

The process of standard SDM commenced with the generation of pseudo-absences for the target species. Most algorithms such as the BRT model require both presence and absence data of species, but distribution data of species obtained from related databases are often composed of presence-only records. Barbet-Massin et al. [21] provided a guideline to generate pseudo-absences for different algorithms. For example, this guideline suggested that generating pseudo-absences with a "random" strategy can yield the most reliable distribution models when using regression techniques. After determining the strategy and number for generating pseudo-absences, we can perform this step using some R packages, e.g., ENMTools [15] and biomod2 [22].

Subsequently, the BRT models should be tuned to optimize their performance. The BRT model possesses a series of hyperparameters, including "shrinkage", "n.trees", "distribution", "bag.fraction", and "interaction.depth". While default values for these hyperparameters have been predefined in the BRT model, their optimal settings for specific taxa or species remain uncertain. Given the impact of parameter configurations on model complexity and performance, pre-defining these hyperparameters is critical [23]. In our study, we specifically adjusted three hyperparameters ("shrinkage", "n.trees", and "interaction.depth") following guidelines by Elith et al. [24]. R packages such as SDMtune [23] can be utilized for tuning the BRT model of the target species. The optimal combination of hyperparameters was determined based on the area under the receiver operator characteristic curve (AUC) metric in SDMtune. Significantly, if other modeling algorithms are selected, the models also need parameter tuning to optimize model performance.

After model tuning, we performed the well-tuned BRT model with ten cross-validation replicates using the optimal combinations of hyperparameters in the 'biomod2' R package [22] for the target species. Each BRT model replicate was trained on 80% of the occurrence data, with the remaining 20% data reserved for testing. To evaluate model performance, we utilized the true skill statistic (TSS) metric and the AUC metric. Given that the value of TSS metric > 0.7 indicates high model performance, we only used models with TSS values greater than 0.7 to build the ensemble BRT model of the target species for further projections. These ensemble models were created using both the probability weighted mean method and the committee averaging method. Their performances were also assessed based on the AUC and TSS metrics. Then, ensemble models generated by two methods were projected into the current climatic layers to predict the potential distribution of species in the current period.

In the final phase, we converted the continuous suitability predictions of those ensemble models into binary outputs. A previous study proposed that the "maximizing the sum of sensitivity and specificity" (maxSSS) statistic is a well-performing threshold selection method for transforming continuous predictions into binary maps [25]. Thus, we calculated the maxSSS values for the ensemble models of species, and the habitat suitability value of each raster cell below the corresponding maxSSS value was set to 0.

Phylogenetically-informed SDMs

To forecast more comprehensive potential distributions of species, we employed phylogenetically-informed SDMs to model their distribution. This method integrates phylogenetic information with traditional SDM to enhance model performance. Standard SDMs characterize environmental envelopes of species based on the relationships between environmental variables and observed occurrences of species [26,27]. However, these models typically reflect the realized niches of species, potentially underestimating their climatic tolerances and full potential distributions [26]. The multi-temporal calibration approach, proposed as a solution to this limitation, operates under the premise that while the fundamental niche of a species remains relatively stable over time, its realized niche can shift [28,29]. Studies have shown that the multi-temporal calibrated SDMs, which incorporate geographic locations of

species from different periods (e.g., fossil records), can enhance model transferability and provide estimation closer to the fundamental niche of species [30–33]. Yet, the scarcity of fossil records for many species impedes the verification of their past distributions and characterization of their realized niches in the past.

Our solution is to integrate eco-evolutionary information into SDMs. Considering the fundamental niche is often regarded as a lineage-specific trait that is conserved within and even across lineages [34,35], the fundamental niche of the target species can be informed by the niches of its related species in a phylogenetic framework [36]. Hence, we drew on the conceptual frame of the multi-temporal calibration and proposed our phylogenetically-informed SDM. The process involves several key steps:

We first reconstructed ancestral climatic envelopes for the target species and its closely related species, utilizing current climate data and the dated phylogeny. This was done with the 'machuruku' R package [37]. The reconstruction of ancestral niches involves two stages: a) constructing Bioclim models for all species using their observed occurrence records and selected predictor variables. The current climatic envelopes of these species are described by some parameters, e.g., the mean, standard deviation, and skewness; b) utilizing the dated phylogeny and the derived parameters of the climatic envelope as inputs to estimate ancestral character states at specific past time points under the Brownian Motion (BM) model. We here focused on two past time points for the target species, i.e., the early and middle stages respectively in the course of their evolution, with the timing dependent on the divergence times of species. These ancestral characters were then converted to ancestral niche models of species with the 'machuruku' package [37]. Significantly, the BM model assumes that niche evolution of species is gradual, which may oversimplify the process of niche evolution. This is an omission needs to be addressed, and some progress has been made [38], which is also the direction for us to optimize our framework in the future.

Subsequently, those ancestral niche models were projected into current climatic layers, transforming the ancestral climatic envelopes into binary suitability maps using the 'machuruku' package [37]. As potential distributions predicted by modeling essentially represent the spatial manifestation of the realized niche of species, we overlaid these potential distributions, generated from ancestral niches, with those from standard SDMs. This overlay, conducted using the 'raster' R package [39] allowed us to integrate realized niches of the target species from different periods to predict a more comprehensive potential distribution for species.

Future projections of species distribution

After constructing SDMs with different methods, the next step is to predict the potential distributions for the target species across various future periods to explore species-specific responses to changing climates. For standard SDMs, we employed the BRT ensemble model, previously generated for the target species, and projected it into future climatic layers to predict future potential distributions of species. These projections were then transformed into binary suitability maps for subsequent analyses. For phylogenetically-informed SDMs, the ancestral niche models, developed in our earlier analyses, were similarly projected into climatic layers of future periods. This projection process allowed us to convert these models into spatial representations. We then overlaid the "future" potential distributions forecasted by both the ancestral niche models and the standard SDM to provide a potential distribution of phylogenetically-informed SDM for the target species.

Given that integrating distance constraints can enhance the accuracy of SDMs [40], we suggested generating an accessible area for the target species. This was achieved by first establishing a minimum convex polygon (MCP) for each species, followed by the creation of an appropriate buffer zone surrounding each MCP. The size of buffer zone should be tailored to the dispersal ability of species being studied. After that, the predictions of different modeling strategies were processed based on the corresponding accessible area. Following this, we employed the 'biomod2' package [22] in R to compare the changes in potential distributions of the target species across different future periods.

Additional phylogenetic analysis

Even if multi-temporal calibrated SDMs can provide estimation closer to the fundamental niche of species, the projections they generate should be approached with caution due to potential overly optimistic expectations in responses of species to climate change. Furthermore, some studies have underscored the significance of evolutionary history of a species in forecasting its future distribution under changing climatic conditions [41,42]. To enhance our understanding of the responses of species to climate shifts, we undertook comprehensive phylogenetic analyses in our framework. These analyses include estimating the rate of niche evolution and reconstructing ancestral areas.

To gain a deeper insight into how different species have adapted to climate change, we quantified their mean rates of realized niche evolution. In this study, we adopted a phylogenetic ridge regression-based method to calculate niche evolution rates [43,44]. This approach enables the estimation of evolutionary rates across all branches of the phylogenetic tree. Specifically, we accounted for the ancestral characters of main nodes as reference values when computing rates of niche evolution. Initially, we identified the optimal model of niche evolution for the target species using the 'geiger' R package [45] based on the climatic data of their occurrences and the dated phylogeny. We evaluated four evolution models: White Noise, Brownian Motion, Ornstein-Uhlenbeck, and Early Burst. Their performance was assessed using Akaike's Information Corrected Criterion (AICc) values. Subsequently, the ancestral characters of those climatic variables were estimated with the 'phytools' package [46], based on the optimal evolution model to provide reference values for the estimation of niche evolution rates. Finally, we used the dated phylogeny, climatic data, and the reference values of main nodes as inputs to compute rates of niche evolution for the target species, employing the 'RRphylo' R package [44].

To reconstruct the historical range dynamics of the target species, we estimated its ancestral distributions. This analysis commenced with the acquisition of paleoclimate data. In recent years, high-quality paleoclimatic data has become available, which contributes

to the development of ancestral range estimation [10,47]. We can download paleo-climatic layers with high spatial resolution from some databases (e.g., Oscillayers [10] and PaleoClim [47]) according to the requirements. Subsequently, using the processes detailed in the 'Phylogenetically-informed SDMs' section, we generated ancestral niche models for the target species utilizing the 'machuruku' R package [37]. These models were then projected into the corresponding paleo-climatic layers to deduce potential ancestral distributions for the target species. These ancestral distributions were further transformed into binary suitability maps through the 'machuruku' R package, providing a clearer picture of the historical range dynamics of species.

An empirical examination of the *Chrysanthemum zawadskii* species complex

To assess the effectiveness of our approach, we here focused on the *Chrysanthemum zawadskii* species complex within the genus *Chrysanthemum* L. (Anthemidae, Asteraceae). This complex contains seven closely related species in China, i.e., *C. zawadskii, C. maximowiczii, C. naktongense, C. chanetii, C. mongolicum, C. hypargyrum,* and *C. oreastrum*, which exhibit diverse habitats and different patterns of distribution [48,49]. Previous studies reconstructed a well-resolved phylogeny for this complex and traced its evolutionary history, serving as a foundation for our analysis and modeling in this study [50,51]. Notably, two narrowly distributed species, i.e., *C. oreastrum* and *C. hypargyrum*, consist of distinct geographical lineages with clear genetic and ecological differentiation [50,51]. Therefore, we implemented SDMs at lineage levels for these two species. We predicted the potential distribution of different species and lineages within this complex, calculated their niche evolution rates, and reconstructed their ancestral distributions following the previously described analysis process. Data collection and preprocessing of the empirical examination are detailed in supplementary material.

We then conducted a comparative analysis between phylogenetically-informed SDMs and standard SDMs to evaluate the contributions of ancestral realized niches to the total niche of the target species. Specifically, the values of two niche breadth metrics, Levins' $B1$ and $B2$ (as proposed by Levins [52]) were measured for each species and lineage based on their potential distributions. These metrics range from 0 to 1, with values nearing 1 denoting a wider niche breadth, and those closer to 0 indicating a narrower one. We used predictions of species and lineage which added distance constraints as inputs to calculate these niche breadth metrics using the 'ENMtools' R package [15]. Subsequently, we compared the current potential distribution of each species and lineage based on different modeling strategies with the 'biomod2' R package [22]. Our niche comparison results revealed that the niche breadths provided by phylogenetically-informed SDMs are consistently broader than those predicted by standard SDMs for all species and lineages, indicating that our approach estimates broader niches for all species and lineages (Table S1). Similarly, phylogenetically-informed SDMs also provided broader potential distributions for all species and lineages than those generated by standard SDMs (Table S1).

To further validate the effectiveness of phylogenetically-informed SDM, we focused on *C. chanetii*, given its relatively abundant occurrence records within the complex. We created partial presence datasets by randomly selecting a subset of occurrence records (50% and 75%) for each species in the complex. This random sampling was replicated five times for each subset percentage, resulting in a total of 10 partial presence datasets. These partial presence datasets were utilized to reconstruct the climatic envelopes of *C. chanetii* and predict its current potential distributions using our phylogenetically-informed SDMs, as previously described. We then compared these results with those obtained from standard SDMs to ascertain whether our approach estimates a broader niche and predicts a more extensive distribution for *C. chanetii*. The outcomes of this validation test, detailed in Table 1, underscore the superior capacity of our method in providing a broader climatic envelope for the target species compared to standard SDMs.

In practice, we found that the phylogenetically-informed SDMs also predicted broader potential distributions than those of the standard SDMs across various future periods. Besides, the results of two modeling strategies display different tendencies of range change under climate change (Table S2). In this case, integrating the results of SDMs with niche evolution rates (Table S3) and ancestral distribution of the target species (Figure S1) can help us understand their fate in a rapidly changing world.

Table 1
Validation test results for phylogenetically-informed SDMs in *Chrysanthemum chanetii*.

Method	Presence datasets	Niche breadth		Range change (%)
		Levins' $B1$	Levins' $B2$	
Standard SDM	all occurrences	0.827	0.123	/
Phylogenetically-	all occurrences	0.886	0.252	104.1
informed	50% occurrences repeat-1	0.883	0.242	96.3
SDM	50% occurrences repeat-2	0.858	0.180	45.7
	50% occurrences repeat-3	0.860	0.184	49.0
	50% occurrences repeat-4	0.880	0.233	88.8
	50% occurrences repeat-5	0.869	0.203	64.9
	75% occurrences repeat-1	0.894	0.276	124.0
	75% occurrences repeat-2	0.890	0.263	113.3
	75% occurrences repeat-3	0.875	0.219	77.9
	75% occurrences repeat-4	0.895	0.280	127.1
	75% occurrences repeat-5	0.859	0.181	47.0

Note: Validation test is based on the potential distributions of *C. chanetii* in the current period. Levins' $B1$ and $B2$ niche breadth metrics are as suggested by Levins [52].

Summary

This study introduces a comprehensive framework that combines eco-evolutionary data with species distribution models to forecast species range shifts in response to changing climates. Our approach is universally applicable to species for which phylogenetic data and occurrence records are available. Validation tests have demonstrated that incorporating the realized niches of target species from various time periods results in a more complete delineation of their climatic envelopes. Consequently, these phylogenetically-informed SDMs are capable of predicting a more extensive potential distribution for the target species. However, it is important to acknowledge that such phylogenetically-informed SDMs might yield overly optimistic predictions regarding responses of species to climate change. To address this, we advocate for incorporating additional data, such as the rate of niche evolution and ancestral distribution of species, and recommend a cautious interpretation of the results. Overall, integrating eco-evolutionary information is crucial for accurately predicting the future potential distributions of species, thereby enriching our understanding of how species respond to climate change.

Ethics statements

None

CRediT author statement

Lu Wen-Xun: Conceptualization, Methodology, Validity tests, Data curation, Visualization, Writing- Original draft preparation. **Rao Guang-Yuan**: Supervision, Writing- Reviewing and Editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The data that support the findings of this study are included in the article/Supplementary Material.

Acknowledgments

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.mex.2024.102608.

References

[1] R Core TeamR: A Language and Environment for Statistical Computing, R Foundation for Statistical Computing, Vienna, Austria, 2022 URL https://www.R-project.org/ .

[2] W.-X. Lu, Z.-Z. Wang, X.-Y. Hu, G.-Y. Rao, Incorporating eco-evolutionary information into species distribution models provides comprehensive predictions of species range shifts under climate change, Sci. Total Environ. 912 (2024) 169501, doi:10.1016/j.scitotenv.2023.169501.

[3] M.A. Suchard, P. Lemey, G. Baele, D.L. Ayres, A.J. Drummond, A. Rambaut, Bayesian phylogenetic and phylodynamic data integration using BEAST 1.10, Virus Evol. 4 (1) (2018) vey016, doi:10.1093/ve/vey016.

[4] Z. Yang, PAML 4: phylogenetic analysis by maximum likelihood, Mol. Biol. Evol. 24 (8) (2007) 1586–1591, doi:10.1093/molbev/msm088.

[5] Z.-D. Chen, T. Yang, L. Lin, L.-M. Lu, H.-L. Li, M. Sun, B. Liu, M. Chen, Y.-T. Niu, J.-F. Ye, Z.-Y. Cao, H.-M. Liu, X.-M. Wang, W. Wang, J.-B. Zhang, Z. Meng, W. Cao, J.-H. Li, S.-D. Wu, H.-L. Zhao, Z.-J. Liu, Z.-Y. Du, Q.-F. Wang, J. Guo, X.-X. Tan, J.-X. Su, L.-J. Zhang, L.-L. Yang, Y.-Y. Liao, M.-H. Li, G.-Q. Zhang, S.-W. Chung, J. Zhang, K.-L. Xiang, R.-Q. Li, D.E. Soltis, P.S. Soltis, S.-L. Zhou, J.-H. Ran, X.-Q. Wang, X.-H. Jin, Y.-S. Chen, T.-G. Gao, J.-H. Li, S.-Z. Zhang, A.-M. Lu, China Phylogeny Consortium, Tree of life for the genera of Chinese vascular plants, J. Syst. Evol. 54 (4) (2016) 277–306, doi:10.1111/jse.12219.

[6] Y. Jin, H.V. Qian, PhyloMaker2: an updated and enlarged R package that can generate very large phylogenies for vascular plants, Plant Diversity 44 (4) (2022) 335–339, doi:10.1016/j.pld.2022.05.005.

[7] R.J. Hijmans, S.E. Cameron, J.L. Parra, P.G. Jones, A. Jarvis, Very high resolution interpolated climate surfaces for global land areas, Int. J. Climatol. 25 (15) (2005) 1965–1978, doi:10.1002/joc.1276.

[8] P.O. Title, J.B. Bemmels, ENVIREM: an expanded set of bioclimatic and topographic variables increases flexibility and improves performance of ecological niche modeling, Ecography 41 (2) (2018) 291–307, doi:10.1111/ecog.02880.

[9] G. Amatulli, S. Domisch, M.-N. Tuanmu, B. Parmentier, A. Ranipeta, J. Malczyk, W. Jetz, A suite of global, cross-scale topographic variables for environmental and biodiversity modeling, Sci. Data 5 (1) (2018) 180040, doi:10.1038/sdata.2018.40.

[10] A. Gamisch, Oscillayers: a dataset for the study of climatic oscillations over Plio-Pleistocene time-scales at high spatial-temporal resolution, Global Ecol. Biogeogr. 28 (11) (2019) 1552–1560, doi:10.1111/geb.12979.

[11] N. Sillero, A.M. Barbosa, Common mistakes in ecological niche models, Int. J. Geogr. Inf. Sci. 35 (2) (2021) 213–226, doi:10.1080/13658816.2020.1798968.

[12] M.E. Aiello-Lammens, R.A. Boria, A. Radosavljevic, B. Vilela, R.P. Anderson, spThin: an R package for spatial thinning of species occurrence records for use in ecological niche models, Ecography 38 (5) (2015) 541–545, doi:10.1111/ecog.01132.

[13] D.R. Roberts, V. Bahn, S. Ciuti, M.S. Boyce, J. Elith, G. Guillera-Arroita, S. Hauenstein, J.J. Lahoz-Monfort, B. Schröder, W. Thuiller, D.I. Warton, B.A. Wintle, F. Hartig, C.F. Dormann, Cross-validation strategies for data with temporal, spatial, hierarchical, or phylogenetic structure, Ecography 40 (8) (2017) 913–929, doi:10.1111/ecog.02881.

[14] L. Zhang, F. Huettmann, X.-D. Zhang, S.-R. Liu, P.-S. Sun, Z. Yu, C.-R. Mi, The use of classification and regression algorithms using the random forests method with presence-only data to model species' distribution, MethodsX 6 (2019) 2281–2292, doi:10.1016/j.mex.2019.09.035.

[15] D.L. Warren, N.J. Matzke, M. Cardillo, J.B. Baumgartner, L.J. Beaumont, M. Turelli, R.E. Glor, N.A. Huron, M. Simões, T.L. Iglesias, J.C. Piquet, R. Dinnage, ENMTools 1.0: an R package for comparative ecological biogeography, Ecography 44 (4) (2021) 504–511, doi:10.1111/ecog.05485.

[16] P.B. Pearman, M. D'Amen, C.H. Graham, W. Thuiller, N.E. Zimmermann, Within-taxon niche structure: niche conservatism, divergence and predicted effects of climate change, Ecography 33 (6) (2010) 990–1003, doi:10.1111/j.1600-0587.2010.06443.x.

[17] D.H. Ikeda, T.L. Max, G.J. Allan, M.K. Lau, S.M. Shuster, T.G. Whitham, Genetically informed ecological niche models improve climate change predictions, Glob. Change Biol. 23 (1) (2017) 164–176, doi:10.1111/gcb.13470.

[18] Z. Zhang, J.M. Kass, S. Mammola, I. Koizumi, X. Li, K. Tanaka, K. Ikeda, T. Suzuki, M. Yokota, N. Usio, Lineage-level distribution models lead to more realistic climate change predictions for a threatened crayfish, Divers. Distrib. 27 (4) (2021) 684–695, doi:10.1111/ddi.13225.

[19] T. Hao, J. Elith, J.J. Lahoz-Monfort, G. Guillera-Arroita, Testing whether ensemble modelling is advantageous for maximising predictive performance of species distribution models, Ecography 43 (4) (2020) 549–558, doi:10.1111/ecog.04890.

[20] R. Valavi, G. Guillera-Arroita, J.J. Lahoz-Monfort, J. Elith, Predictive performance of presence-only species distribution models: a benchmark study with repro-ducible code, Ecol. Monogr. 92 (1) (2022) e01486, doi:10.1002/ecm.1486.

[21] M. Barbet-Massin, F. Jiguet, C.H. Albert, W. Thuiller, Selecting pseudo-absences for species distribution models: how, where and how many? Methods Ecol. Evol. 3 (2) (2012) 327–338, doi:10.1111/j.2041-210X.2011.00172.x.

[22] Thuiller, W., Georges, D., Engler, R., Breiner, F., Biomod2: ensemble platform for species distribution modeling. R package version 3.3-7.1. 2019. https://CRAN.R-project.org/package=biomod2.

[23] S. Vignali, A.G. Barras, R. Arlettaz, V. Braunisch, SDMtune: an R package to tune and evaluate species distribution models, Ecol. Evol. 10 (20) (2020) 11488–11506, doi:10.1002/ece3.6786.

[24] J. Elith, J.R. Leathwick, T. Hastie, A working guide to boosted regression trees, J. Anim. Ecol. 77 (4) (2008) 802–813, doi:10.1111/j.1365-2656.2008.01390.x.

[25] C. Liu, M. White, G. Newell, Selecting thresholds for the prediction of species occurrence with presence-only data, J. Biogeogr. 40 (4) (2013) 778–789, doi:10.1111/jbi.12058.

[26] M.B. Araújo, A. Guisan, Five (or so) challenges for species distribution modelling, J. Biogeogr. 33 (10) (2006) 1677–1688, doi:10.1111/j.1365-2699.2006.01584.x.

[27] J. Elith, J.R. Leathwick, Species distribution models: ecological explanation and prediction across space and time, Annu. Rev. Ecol. Evol. Syst. 40 (1) (2009) 677–697, doi:10.1146/annurev.ecolsys.110308.120159.

[28] O. Broennimann, A. Guisan, Predicting current and future biological invasions: both native and invaded ranges matter, Biol. Lett. 4 (5) (2008) 585–589, doi:10.1098/rsbl.2008.0254.

[29] I. Morales-Castilla, T.J. Davies, W.D. Pearse, P. Peres-Neto, N. Swenson, Combining phylogeny and co-occurrence to improve single species distribution models, Global Ecol. Biogeogr. 26 (6) (2017) 740–752, doi:10.1111/geb.12580.

[30] D. Nogués-Bravo, Predicting the past distribution of species climatic niches, Global Ecol. Biogeogr. 18 (5) (2009) 521–531, doi:10.1111/j.1466-8238.2009.00476.x.

[31] L. Maiorano, R. Cheddadi, N.E. Zimmermann, L. Pellissier, B. Petitpierre, J. Pottier, H. Laborde, B.I. Hurdu, P.B. Pearman, A. Psomas, J.S. Singarayer, O. Broen-nimann, P. Vittoz, A. Dubuis, M.E. Edwards, H.A. Binney, A. Guisan, Building the niche through time: using 13,000 years of data to predict the effects of climate change on three tree species in Europe, Global Ecol. Biogeogr. 22 (3) (2013) 302–317, doi:10.1111/j.1466-8238.2012.00767.x.

[32] F. Yu, Y. Sun, T. Wang, A.K. Skidmore, C. Ding, X. Ye, Linking the past and present to predict the distribution of Asian crested ibis (*Nipponia nippon*) under global changes, Integr. Zool. 17 (6) (2022) 1095–1105, doi:10.1111/1749-4877.12581.

[33] A. Mondanaro, M. Di Febbraro, S. Castiglione, M. Melchionna, C. Serio, G. Girardi, A.M. Belfiore, P. Raia, ENphylo: a new method to model the distribution of extremely rare species, Methods Ecol. Evol. 14 (3) (2023) 911–922, doi:10.1111/2041-210X.14066.

[34] D. Strubbe, O. Broennimann, F. Chiron, E. Matthysen, Niche conservatism in non-native birds in Europe: niche unfilling rather than niche expansion, Global Ecol. Biogeogr. 22 (8) (2013) 962–970, doi:10.1111/geb.12050.

[35] H. Qiao, A.T. Peterson, L. Ji, J. Hu, Using data from related species to overcome spatial sampling bias and associated limitations in ecological niche modelling, Methods Ecol. Evol. 8 (12) (2017) 1804–1812, doi:10.1111/2041-210X.12832.

[36] R.A. Catullo, S. Ferrier, A.A. Hoffmann, Extending spatial modelling of climate change responses beyond the realized niche: estimating, and accommodating, physiological limits and adaptive evolution, Global Ecol. Biogeogr. 24 (10) (2015) 1192–1202, doi:10.1111/geb.12344.

[37] W.X. Guillory, J.L. Brown, A new method for integrating ecological niche modeling with phylogenetics to estimate ancestral distributions, Syst. Biol. 70 (5) (2021) 1033–1045, doi:10.1093/sysbio/syab016.

[38] S. Castiglione, A. Mondanaro, M. Di Febbraro, M. Melchionna, C. Serio, G. Girardi, A.M. Belfiore, P. Raia, Testing for changes in rate of evolution and position of the climatic niche of clades, Mammal Rev 52 (4) (2022) 573–583, doi:10.1111/mam.12303.

[39] Hijmans, R.J., Raster: geographic data analysis and modeling. R package version 3.5-15. https://cran.r-project.org/web/packages/raster/index.html.

[40] P. Mendes, S.J.E. Velazco, A.F.A. de Andrade, P.D.M. Júnior, Dealing with overprediction in species distribution models: how adding distance constraints can improve model accuracy, Ecol. Model. 431 (2020) 109180, doi:10.1016/j.ecolmodel.2020.109180.

[41] M. Cordellier, M. Pfenninger, Inferring the past to predict the future: climate modelling predictions and phylogeography for the freshwater gastropod *Radix balthica* (Pulmonata, Basommatophora), Mol. Ecol. 18 (3) (2009) 534–544, doi:10.1111/j.1365-294X.2008.04042.x.

[42] L. Gallien, B. Saladin, F.C. Boucher, D.M. Richardson, N.E. Zimmermann, Does the legacy of historical biogeography shape current invasiveness in pines? New Phytol 209 (3) (2016) 1096–1105, doi:10.1111/nph.13700.

[43] C. Kratsch, A.C. McHardy, RidgeRace: ridge regression for continuous ancestral character estimation on phylogenetic trees, Bioinformatics 30 (17) (2014) i527–i533, doi:10.1093/bioinformatics/btu477.

[44] S. Castiglione, G. Tesone, M. Piccolo, M. Melchionna, A. Mondanaro, C. Serio, M. Di Febbraro, P. Raia, A new method for testing evolutionary rate variation and shifts in phenotypic evolution, Methods Ecol. Evol. 9 (4) (2018) 974–983, doi:10.1111/2041-210X.12954.

[45] L.J. Harmon, J.T. Weir, C.D. Brock, R.E. Glor, W. Challenger, GEIGER: investigating evolutionary radiations, Bioinformatics 24 (1) (2008) 129–131, doi:10.1093/bioinformatics/btm538.

[46] L.J. Revell, phytools: an R package for phylogenetic comparative biology (and other things), Methods Ecol. Evol. 3 (2) (2012) 217–223, doi:10.1111/j.2041-210X.2011.00169.x.

[47] J.L. Brown, D.J. Hill, A.M. Dolan, A.C. Carnaval, A.M. Haywood, PaleoClim, high spatial resolution paleoclimate surfaces for global land areas, Sci. Data. 5 (1) (2018) 180254, doi:10.1038/sdata.2018.254.

[48] C. Shih, G.-X. Fu, *Dendranthema* (DC.) Des Moul, in: Z.Y. Wu, P.H. Raven, D.Y. Hong (Eds.), Flora Republicae Popularis Sinicae Volume 76: Angiospermae, Dicotyledoneae, Compositae (3), Science Press, Beijing, China, 1983, pp. 28–49.

[49] P.-L. Liu, Q. Wang, Y.-P. Guo, J. Yang, G.-Y. Rao, Phylogeny of the genus *Chrysanthemum* L.: evidence from single-copy nuclear gene and chloroplast DNA sequences, PLoS ONE 7 (11) (2012) e48970, doi:10.1371/journal.pone.0048970.

[50] W.-X. Lu, X.-Y. Hu, Z.-Z. Wang, G.-Y. Rao, Hyb-Seq provides new insights into the phylogeny and evolution of the *Chrysanthemum zawadskii* species complex in China, Cladistics 38 (6) (2022) 663–683, doi:10.1111/cla.12514.

[51] W.-X. Lu, Z.-Z. Wang, X.-Y. Hu, G.-Y. Rao, Echoes of the past: niche evolution, range dynamics, and their coupling shape the distribution of species in the *Chrysanthemum zawadskii* species complex, Front. Ecol. Evol. 11 (2023) 1250491, doi:10.3389/fevo.2023.1250491.

[52] R. Levins, Evolution in Changing Environments, Princeton University Press, Princeton, USA, 1968.