Cardador, Laura. «Incorporating phylogeographic information in alien bird distribution models increases geographic extent but not accuracy of predictions». Biological invasions, Vol. 24, issue 3 (March 2022), p. 683–695. DOI 10.1007/s10530-021-02673-7

Incorporating phylogeographic information in alien bird distribution models increases geographic extent but not accuracy of predictions

Laura Cardador[1,2], Pedro Abellán[3], Tim M. Blackburn[1,4]

[1]Centre for Biodiversity and Environment Research, Department of Genetics, Evolution, and Environment, University College London, London WC1E 6BT, United Kingdom

[2]Ecological and Forestry Applications Research Centre, Campus de Bellaterra (UAB) Edifici C, 08193 Cerdanyola del Vallès, Spain

[3]Department of Zoology, Universidad de Sevilla, Facultad de Biología, 41012 Seville, Spain

[4]Institute of Zoology, Zoological Society of London, Regent's Park, London, NW1 4RY, UK

*Corresponding author.
Email: l.cardador@creaf.uab.cat; lcardador81@gmail.com

ACKNOWLEDGEMENTS

We thank all authors and observers who contributed to the datasets on species distributions and phylogeographic patterns, without which this study could not have been possible. This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie Grant Agreement No. 752149. L.C. also received funding from the Beatriu de Pinós fellowship program (funded by the Catalan Government and EU COFUND program nº 801370). PA is funded by 'V Plan Propio de Investigación' of the Universidad de Sevilla (Spain). The authors acknowledge the use of the UCL Myriad High Performance Computing Facility (Myriad@UCL), and associated support services, in the completion of this work.

ABSTRACT

Species distribution models (SDM) have been proposed as valuable first screening tools for predicting species responses to new environmental conditions. SDMs are usually conducted at the species level, assuming that species-environment relationships are a species-specific feature that do not evolve and show no variability across a species' range. However, broad environmental tolerances at the species level can encompass narrower and different environmental tolerances for specific lineages or populations. In this study, we evaluate whether SDMs that account for within-taxon niche variation in climate and human-habitat associations provide better fits between projected distributions and real occurrence data for alien bird species than species-level SDMs. Our study focuses on eight alien bird species with established alien populations for which detailed phylogeographic information was available. Similarity in climates and human disturbance conditions occupied by different phylogenetic groups within species was low and not greater than random expectations. Accounting for intraspecific niche variation in SDMs modified the distribution and extent of suitable habitat predicted as susceptible to invasion, but did not result in more accurate model predictions in alien ranges. Until more accurate information on intraspecific variability is available, species-level models can be reasonable candidates. When phylogeographic information is available, the use of the most conservative criterion (i.e. to model both species and lineages on the basis of the actual range) is recommended.

Keywords: alien species, birds, climate, human disturbance, intraspecific niche variation, invasion risks, species distribution models

INTRODUCTION

Human activities are accelerating the rate and magnitude of changes in species geographic ranges worldwide. Climate change, land-use transformations, habitat fragmentation or environmental pollution, among others, threaten the persistence of several species in formerly suitable locations, leading to species range contractions or extinctions, or pushing species to track new suitable environmental conditions beyond former range limits (Pimm et al. 1995; Vitousek 1997; Parmesan et al. 1999). The increase and development of trade and transport infrastructures, in turn, have accelerated the dispersal, and subsequent establishment, of alien species in places far away from their native range (Blackburn et al. 2009; Hulme 2009). Such changes have notably altered the distribution of species worldwide, homogenizing species assemblages (McKinney and Lockwood 1999; Capinha et al. 2015; Sayol et al. 2021), with detrimental effects on biodiversity and ecosystems.

Although biotic interactions and dispersal are important in constraining species ranges, environmental factors can exert a primary role (Gaston 2003; Huntley et al. 2007). Species distribution models (SDMs; models that statistically relate observed species occurrences to environmental variables) have thus been proposed as valuable first screening tools for predicting species responses to new environmental conditions in new geographic areas (e.g. invasion risk assessments) or under future environmental scenarios (e.g. global climate-change) based on current species occurrence-environment relationships (Guisan and Thuiller 2005; Araújo and Peterson 2012). SDMs rely on ecological niche theory, which predicts that for relatively recent events such as biological invasions, the environmental niche is expected to be conserved (Peterson 2011). SDMs often focus on macroclimatic variables. Additionally, accounting for environmental factors other than climate, such as habitat characteristics and human disturbance, can substantially improve model predictions (Strubbe et al. 2015; Cardador and Blackburn 2020).

SDMs are usually conducted at the species level, assuming that species-environment relationships are a species-specific feature that does not evolve and shows no variability across a species' range. However, widely distributed species often encompass different taxonomic or evolutionary units, which can reflect the existence of ecotypes and locally adapted populations (Smith et al. 2019). Spatial heterogeneity in environments coupled with reduced gene flow can encourage local adaptation and functional differences, leading to divergence in niches among closely related lineages. In fact, recent work has suggested that broad environmental tolerances at the species level usually encompass narrower and different environmental tolerances for specific lineages or populations within the species (Peterson and Holt 2003; Pearman et al. 2010). Hence, modelling a species as a single undifferentiated entity may obscure the possibility that these lineages occupy distinct niches and, as a consequence, miss the idiosyncratic response of intraspecific lineages to changing environmental conditions (Pearman et al. 2010; Lecocq et al. 2019). Accounting for intraspecific niche variation in SDMs has thus been highlighted as important for forecasting species range shifts under changing environmental conditions, particularly under global climate change (Pearman et al. 2010; Peterson et al. 2019). In particular, phylogeographic structures have been proposed as a useful proxy to incorporate intraspecific differentiation in SDMs. However, the efficiency of using these proxies in SDMs remains largely unknown, in part because spatiotemporally independent data to test the accuracy of model predictions are often unavailable (Peterson et al. 2019).

Biological invasions represent unique, large-scale biogeographical experiments for evaluating model transferability (Liu et al. 2020). Large numbers of alien species have been introduced well outside their native ranges, resulting in geographically independent datasets. However, the relevance of including intraspecific niche variation in invasion risk assessments has not often been addressed (but see exceptions, Strubbe et al. 2015; Godefroid et al. 2016). Many alien species currently arrive at new areas because they are imported as trade commodities (Hulme et al. 2008; Abellán et al. 2016; Cardador et al. 2017, 2019). If individuals coming from different geographic origins belong to different lineages that have particular ecological niches (in terms of climate or human tolerances), their invasion success is also likely to differ across different recipient environments. The omission of intraspecific niche structure from niche modelling exercises may lead to some lineages having little representation in the resulting species models (Pearman et al. 2010; D'Amen et al. 2013). This can lead to underestimation of the climate tolerances of alien species and, as a consequence, their potential for establishment and spread in new environments.

Here, we assess the key assumption of distribution modelling theory – that the environmental niche remains conserved across species native ranges – as applied to the bird invasion process. For this, we focused on eight alien bird species with established alien populations, selected because of the availability of robust phylogeographic information, and for which different phylogenetic lineages (i.e., phylogroups) have been identified in previous studies. We explored niche variation in the climatic and human disturbance spaces occupied by the different lineages in the native range in order to assess whether the different phylogeographic lineages or genetic units occupy different niches. It should be noted that the niches considered here relate to the realized niche (occupied niche) and the Grinnellian niche concept: that is, the response of species to a set of non-consumable environmental variables that influence their large-scale geographical distribution (Soberón 2007). As evidence of niche conservatism was not found, we assessed how accounting for intraspecific niche variation in SDMs influences predictions about potential distributions in adventive regions for the whole species. Accounting for intraspecific variation is expected to improve the representation of different phylogroups in models (particularly the scarcest and narrowly distributed in the native range), and thus to produce better fits between projected distributions and real occurrence data for alien bird species than species-level models not considering within-taxon niche structure.

METHODS

Phylogenetic and occurrence data

Our study focuses on 8 alien bird species that have established alien populations in different regions of the world and for which detailed phylogeographic information derived from analyses of mitochondrial and/or nuclear DNA sequences was available from the literature (Table 1). We only considered studies reporting a clear definition of intra-specific phylogenetic divisions and covering substantial parts of species' native ranges. Thiessen polygons were applied to locations with available phylogenetic data to delimit geographic boundaries of different phylogroups (Figs. 1a and S1 in Supp. Info.) (Strubbe et al. 2015). Thiessen polygons define an area of influence around each sampled point (points with genetic data in our study), where every location of the study area (species ranges in our case) is nearer to this point than to all the others. Thiessen polygons derived from sampled points where a given phylogroup was identified were merged together to obtain the phylogroup geographic boundaries. We used ArcGIS 10.5 for those analyses. Note that for most species, different phylogroups were allopatric (Fig. S1), although cases of partial overlap in geographic ranges also occur. For each species, occurrence data for niche and modelling analyses were compiled from the Global Biodiversity Information Facility (GBIF, GBIF.org, 2017,Table S1). GBIF records spanned the years 1744 to 2017 (Table S2). Compiled records were classified as pertaining to the native breeding range or alien established range according to range maps provided by the BirdLife International & NatureServe (2014) and the Global Avian Invasions Atlas (Dyer et al. 2017), respectively. Note that in the case of *Copsychus saularis*, BirdLife International & NatureServe (2014) provided separate maps for *Copsychus saularis* and *Copsychus mindanensis*. Based on the phylogenetic data available, we considered these two taxa as the same species and considered their range maps jointly (Sheldon et al. 2009). Species occurrences in the native range were then assigned to different phylogroups according to geographic boundaries defined by Thiessen Polygons (Figs. 1a and S1 in Supp. Info.). Occurrence data were aggregated at 5-arcminute resolution, which corresponds approximately to 10 × 10 km. This resolution was considered to be representative of the size of the smallest cities, and thus adequate to capture the main responses of bird species to humanized environments with acceptable computing time. Samples with reported geographical issues, location uncertainty above 5 km, or with central grid coordinates of atlases of >10 km resolution, were removed from analyses. Duplicate samples at the 5-arcminute resolution were handled as single observations. Final sample sizes ranged from 3,156 to 32,052 for different species in the native range, and from 2 to 2,368 in the alien range (Table 1).

Environmental variables

We considered eight bioclimatic variables (obtained from WorldClim 1, http://www.worldclim.org/) (Hijmans et al. 2005), which are known to affect bird distributions (Strubbe et al. 2015; Cardador et al. 2016): annual mean temperature, temperature seasonality (standard deviation ×100), maximum temperature of the warmest month, minimum temperature of the coldest month, annual precipitation, precipitation of the driest month, precipitation of the wettest month and precipitation seasonality (coefficient of variation). However, annual mean temperature, temperature seasonality and annual precipitation were highly correlated with other climate variables across the world ($r \geq 0.90$) and thus removed from analyses (Cardador and Blackburn 2019). We considered two variables as descriptors of human transformed environments: i) the Global Human Influence Index, which provides a weighed composite map of anthropogenic impacts including urban extent, population density, land cover, night lights and distance to roads, railways, navigable rivers and coastlines (Sanderson et al. 2002), and ii) the percentage of urban habitats, as a more specific descriptor of urbanization. The percentage of urban habitats at the 5-arcminute resolution was derived from MODIS-based global land cover climatology data at 500m resolution (Broxton et al. 2014).

Niche analyses

We compared the climatic and human disturbance niches of different phylogroups of a given species in its native range using the framework proposed by Broennimann et al. (2012). These analyses involved four steps: (1) definition of a two-dimensional gridded environmental space, (2) calculation of the density of occurrences along the two-dimensional environmental space using kernel smoothers (3) measurement of niche overlap between occurrence densities of two given groups along the environmental space and (4) statistical tests of niche similarity using a randomization approach. For climatic niche analyses the gridded environmental space was that formed by the first two axes of a PCA on the five climatic variables considered (see above) in 5-arcminute pixels across the world. These axes explained 78% of the inertia. The first PCA axis (45%) predominantly represented temperature gradients (with higher values representing warmer climates) while the second axis (33%) represented precipitation gradients (with higher values representing drier climates with higher precipitation seasonality, Table S3). For the human space, the two axes represented the two human-related variables considered (i.e., the Global Human Influence Index and the percentage of urban environments - see above). Niche similarity was assessed using the Schoener's *D* metric, calculated from the occupancies in the environmental space depicted by the two first axes of the PCA. This metric indicates the overall match between two niches over the whole climatic or human spaces and ranges from 0 (no overlap) to 1 (complete overlap). We calculated niche similarity between each phylogroup of a species and all the other phylogroups using one-to-one comparisons.

We assessed niche conservatism by conducting niche similarity tests, whereby each obtained value of niche similarity was compared against a null distribution of 100 simulated similarity values (obtained when comparing the observed niche of one phylogroup with niches obtained by drawing occurrences at random within available habitats and vice versa) (Warren et al. 2008; Broennimann et al. 2012). Note that the niche similarity test is thus bidirectional, and two tests were conducted for each comparison between two phylogroups. As available habitat for each phylogroup in species' native ranges for analyses, we considered two alternative approaches. First we considered all ecoregions (Olson et al. 2001) occupied by each species in its native range (Figs. 1b and S2), as this might represent the complete gradient of climatic and human conditions that the study species could have reasonably encountered considering that dispersal is expected to be mainly limited by major biogeographical barriers in native ranges (Soberon and Peterson 2005). Second, we accounted for potential spatial constraints limiting the access of different phylogroups to available habitat for the whole species. For this, we repeated the niche similarity tests considering as available habitat for each phylogroup that present within the geographic range actually occupied by that particular phylogroup: i.e., that limited by the geographic boundaries derived from Thiessen polygons (Fig.S1). The results of both analyses were highly consistent (Table S4), and for simplicity we thus only provide those for the former approach in the main manuscript. All analyses were conducted using the 'ecospat' library in R software (Broenniman et al. 2014). The minimum sample size for analyses was five occurrences (Broennimann et al. 2012).

Species distribution models

We fitted SDMs calibrated on occurrences of different phylogroups (single-phylogroup models) and for the species as a whole (species-level model) to generate global predictions of species potential distributions outside their native ranges. For robustness of analyses only phylogroups with more than 50 occurrence locations (exceptions: 1 phylogroup for the northern cardinal *Cardinalis cardinalis*, 1 for the oriental magpie-robin *Copsychus saularis* and 9 for the ring-necked parakeet *Psittacula krameri*) were retained for single-phylogroup models (Stockwell and Peterson 2002). Data from phylogroups with fewer than 50 occurrence locations were also disregarded from species-level models, to avoid biases in model comparisons related to differences in the data used. Predictions were derived from an ensemble model of three techniques – generalized linear models, MAXENT and random forest – using R library ‘biomod2’. Both the linear and quadratic terms of the climate and human predictors were considered, to account for positive or negative responses to intermediate values of the variables. All models were run with a single set of a maximum of 10,000 pseudo-absences randomly drawn from all ecoregions occupied by each species across its native range as with niche analyses (Figs. 1b and S2). Presences and pseudoabsences were weighted as such to ensure neutral (0.5) prevalence.

To reduce the potential effect of sampling biases in the data, a bias file was created by retrieving from GBIF occurrence data at the family level for each species (Elith et al. 2010; Cardador and Blackburn 2019). We derived a kernel density map of sampling bias at a 5-arcminute resolution using ArcMap 10.5 to be included as a fixed effect in model training. Occurrence data from species in the same taxonomic family are expected to suffer from the same detection limitations, reducing the effect of sampling biases in observed distribution patterns. To further account for potential effects of the data selection, we conducted 10 replicates for each model by using random samples (70%) of the complete datasets. Final ensemble model predictions for each species and phylogroup were generated as averaged means of all model replicates conducted (Fig. 1c-e). Sampling bias was set to its maximum value for model predictions. For each species, a composite model prediction integrating information on all single-phylogroup models was then developed (phylogroup-composite model) (Fig. 1f). For this, single-phylogroup predictions were first standardized to a maximum value of 1 to make them comparable. We then calculated the mean probability of occurrence of at least one of the related phylogroups using the multiplicative probability method described in Pearman et al. (2010). We converted continuous model predictions into binary presence–absence maps by implementing a threshold for species presence that maximized sensitivity plus specificity (Liu et al. 2005) in the training region (Fig. 1g). In the case of composite models, binary maps were obtained by assigning species presence to each cell that was predicted suitable for at least one single-phylogroup model (Fig. 1h). To reduce problems related to model extrapolation, model projections were adjusted using multivariate environmental similarity surfaces (MESSs) (Mateo et al. 2014) (Fig. S3). Environmental suitability in dissimilar areas (MESS <0) was considered to be zero. However, analyses using non-adjusted model projections were highly concordant (see results).

Model accuracy of phylogroup-composite and species-level models in predicting species occurrences in the native range were evaluated using the Boyce index (which ranges from -1 to 1, with higher values indicating higher match (Hirzel et al. 2006)) and AUC (which ranges from 0 to 1, with values up to 0.5 representing models not better than random (Phillips et al. 2006)), using the libraries ‘pROC’ and ‘ecospat’ in R. Sensitivity (i.e., the proportion of correctly classified presences) was also computed using the binary maps derived from continuous predictions. Potential differences in accuracy metrics among species-level and phylogroup-composite models in native ranges were evaluated using non-parametric paired Wilcoxon signed rank tests.

Geographic extent of SDM predictions

We calculated the extent of predicted suitable habitat in km^2 for each species at a global scale according to binary map projections of single-phylogroup, phylogroup-composite and species-level models, using the library ‘raster’ from R. Accounting for intraspecific variation is expected to improve the representation of different phylogroups in models (particularly the most scarce and narrowly distributed in the native range), and thus to increase the geographical distribution of habitat conditions predicted as suitable for alien birds: environments occupied by rarer phylogroups are expected to be marginal, and thus result in low probabilities of occurrence in those environments when distribution is modelled at the species level. Differences in the extent of phylogroup-composite and species-level model predictions in alien ranges were evaluated using one-tailed paired t-tests. We also tested the hypothesis that the niches occupied by the most narrowly distributed phylogroups in the native range are little represented in species-level models. For this, we assessed the relationship between the percentage of habitat predicted as suitable by single-phylogroup models also predicted as suitable by species-level models and phylogroup relative range size (i.e., the proportion of species native range covered by a given phylogroup). We used general linear models (GLM) for those analyses.

Accuracy of SDM predictions

SDM-based predictions were tested against real occurrence data in alien ranges using the Boyce index, AUC and sensitivity using the libraries ‘pROC’ and ‘ecospat’ in R as in native ranges. As background for AUC and Boyce calculations, we followed the framework proposed by (Strubbe et al. 2013, 2015) to estimate the area that could have been effectively accessible to introduced birds. We buffered each 5-arcminute alien locality with a distance equal to the minimum invasion speed recorded for birds (i.e. 4.59 km year, derived from (Blackburn et al. 2009)) multiplied by the number of years since introduction (Fig. S4). For localities with duplicate records the oldest year was used in analyses. When the specific year of introduction was not provided (≤ 4% of total alien localities for each species), we were conservative and only considered that locality (not a buffer around it) for background calculations. Buffers were only allowed to cover areas effectively outside species native breeding ranges. Both occurrence localities in regions where the species has effectively established (see ‘Phylogenetic and occurrence data’ section) and known introduction localities (compiled from (Redding et al. 2019)) were used. Differences in the accuracy of phylogroup-composite and species-level model predictions in alien ranges were evaluated using one-tailed paired t-tests (AUC) and one-tailed paired Wilcoxon signed rank tests (Boyce and sensitivity) according to fit of normality and homoscedasticity assumptions.

RESULTS

Intraspecific climatic and human niche variation

Phylogenetic groups within species occupied partially overlapping portions of the climate and human spaces available in the native range. However, climatic niche similarity among phylogroups of a species was low (mean ± SD, $D = 0.07 \pm 0.09$, $N = 8$) and not more similar than expected by chance for the vast majority of phylogroups within species (95% of 101 reciprocal similarity tests involving 32 phylogroups from eight species had $P > 0.05$, Table S4). Human niche similarity was higher than climatic niche similarity ($D = 0.32 \pm 0.13$), but generally not more similar (97% of tests had $P > 0.05$) than expected by chance for most comparisons.

Species distribution models in native ranges

Species-level and phylogroup-composite model predictions showed a good agreement with species occurrences in the native ranges (mean ± SD, species-level: Boyce = 0.88 ± 0.24, AUC = 0.86 ± 0.11, sensitivity = 0.78 ± 0.19; phylogroup-composite: Boyce = 0.91 ± 0.12, AUC = 0.84 ± 0.09, sensitivity = 0.89 ± 0.10). No significant differences in model performance between species-level and phylogroup-composite models were observed (one-tailed paired Wilcoxon rank test: Boyce, $V = 21.5$, $P = 0.71$; AUC, $V = 28$, $P = 0.93$; sensitivity, $V = 6$, $P = 0.054$).

Geographic extent and accuracy of model projections in alien ranges

Species-level and phylogroup-composite models provided similar but not equal predictions of habitat susceptible to invasion (Figs. 1e-h, S5 and S6). Notably, phylogroup-composite models projected significantly larger habitat suitability at a global scale than species-level models (Fig. 2a, t = -2.0, df = 7, $P = 0.04$). On average, only 30 ± 16% of total pixels predicted as suitable by binary maps derived from both species-level and phylogroup-composite models were coincident between both types of models, while 52 ± 30% of pixels were predicted as suitable only by phylogroup-composite models and 18 ± 19% only by species-level models. The capacity of species-level models to predict the occurrence of individual phylogroups was significantly related to phylogroup range size (estimate: 0.68 ± 0.13, $P < 0.001$, $R^2 = 0.55$; Fig. 3).

When used to predict occurrences in alien ranges, model accuracy was on average good according to different metrics considered (Figs. 2b-d), but high variability was observed among species (Figs. 2b-d; see also Fig S7 for comparisons of models not adjusted by MESS analyses). Accounting for intraspecific niche variation did not improve model accuracy (Boyce, $V = 21$, $P = 0.69$; AUC, $t = -0.09$, $df = 7$, $P= 0.47$; sensitivity, $V = 9$, $P= 0.22$). These result hold when omitting *Platycercus elegans* from analyses (Boyce, $V = 14$, $P = 0.53$; AUC, $t = -0.29$, $df = 6$, $P= 0.39$; sensitivity, $V = 9$, $P= 0.22$), for which sample size in the alien range was very low (Table 1).

DISCUSSION

There is a limited number of studies considering intraspecific niche variation in models assessing environmental susceptibility to the colonization by alien birds, and they have rarely compared modelled predictions against independent sets of occurrence data (Peterson et al., 2019). Our results agree with recent evidence suggesting that conspecific phylogenetic lineages of a species can differ in the climates and human disturbance conditions they experience in native ranges (Peterson and Holt 2003; Pearman et al. 2010; D'Amen et al. 2013): we find little climatic niche similarity among phylogroups of the 8 species in our study. Accounting for intraspecific variation in SDMs modified the distribution and the extent of potential suitable habitat for the whole species. However, contrary to our expectations, accounting for intraspecific niche variation did not result in more accurate model predictions, according to current distributions of established alien species.

The increased extent of phylogroup-composite model predictions, when compared to classical species-level models, suggests that the omission of intraspecific niche structure from species distribution models underestimates intraspecific realized niche variation, and thus species-level prediction of habitat susceptible to invasion (Peterson and Holt 2003; D'Amen et al. 2013; Godefroid et al. 2016). The extents of geographic areas occupied by different phylogroups of a species in our study were not equivalent (Fig. S1), and this may lead to some phylogroups (particularly the most scarce and narrowly distributed in the native range) having little representation in model predictions obtained by classical species level models (Fig. 3).

However, contrary to our expectations, accounting for intraspecific niche variation did not improve model accuracy when predicting occurrences in alien ranges. These results could be related to different, not mutually exclusive processes. First, it should be noted that phylogroup separation – in this study and in general – relies on neutral molecular markers. Strong structuring in these markers across populations indicates reduced dispersal and thus increased potential for local genetic adaptation to emerge (Lenormand 2002). Yet, we do not know whether genetic structuring in neutral markers really reflects local genetic adaptation to climatic and human environmental conditions experienced (Holderegger et al. 2006), and thus in these species tolerances, or just reflects differences in the realized (i.e., occupied) niche of different phylogroups (Guisan et al. 2014; Peterson et al. 2019). In this sense, since most phylogroups within species considered have allopatric distributions, observed realized niche divergence may have been driven by different environmental conditions in the range of each phylogroup, rather than by adaptation to different conditions within a shared spatial distribution (Maia-Carvalho et al. 2018). If so, dividing a species' range into several groups might have yielded different climate–occupancy relationships for each group even in the absence of local adaptation. In our study, this could explain the larger species range predictions of models considering phylogroup information, but the quite similar prediction performances. For some species it is also possible that local adaptation occurs at higher or lower taxonomic levels (Peterson et al. 2019; Smith et al. 2019) or geographic scales (Cardador et al. 2016).

Second, lineages little represented in classical species-level models are also those less likely to be introduced and subsequently established in new areas, given the positive effects of abundance and geographic range size on introduction and establishment success in alien species (Blackburn and Duncan 2001b, a). If this is the case, accounting for intraspecific niche variation would result in small differences in model prediction for the most commonly translocated taxa. At the same time, accounting for intraspecific niche variation might overestimate the potential alien range size, as the environmental tolerances of alien individuals would actually be narrower than that of the species as a whole. Under this hypothesis (i.e., the more common phylogroups in native areas are more often introduced and established in alien ranges), higher model accuracy for the more common phylogroups should be expected. However, post-hoc analyses assessing the relationship between accuracy of single-phylogroup model predictions and phylogroup relative range size (i.e., the proportion of species native range covered by a given phylogroup) offer little support for this hypothesis (Pearson correlation coefficients between accuracy metrics and phylogroup range size, AUC: r = -0.20, P =0.33; Boyce: r = -0.01, P = 0.97, N = 25).

Third, as most bird introductions are relatively recent, species might not occupy all of the potential suitable environments available in the invaded range, due to dispersal limitations (Blackburn et al. 2009; Ascensão et al. 2020). For some species, current alien distributions may thus reflect the characteristics of the new introduction localities rather than optimum environmental conditions, which might be more likely to be moved into during spread (Abellán et al. 2017). Furthermore, while climate and human variables appeared to be major factors shaping alien species distributions (Cardador and Blackburn 2019), omission of other important drivers of bird distributions, such as interspecific interactions (Blackburn et al. 2009; Redding et al. 2019), might also produce an overestimate of the projected suitable area for species. Some species might even undergo niche shifts in alien ranges, although this seems to be less frequent (Broennimann et al. 2007; Strubbe et al. 2013; Cardador and Blackburn 2020). All of these issues may have contributed to reduce model accuracy for both species-level and phylogroup-composite model predictions in alien ranges, masking potential differences.

Overall, our results show variability in model predictions linked to taxonomic level considered. However, comparisons of model predictions with current available distribution data in alien ranges do not provide evidence of an improvement in prediction accuracy for models accounting for intraspecific niche variation. This result is relevant given the lack of information about the presence and geographical distribution of phylogeographic lineages for many species: according to our results, until more accurate information on intraspecific variability is available, species-level models can be reasonable candidates. However, when phylogeographic information is available, the use of the most conservative criterion (i.e. to model both species and lineages on the basis of the actual range, e.g. Mori et al. 2019) is recommended, given that the ultimate purpose of such modelling exercises is to reduce invasion risks and their consequences on biodiversity conservation.

DECLARATIONS

Funding

This project has received funding from the European Union’s Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie Grant Agreement No. 752149. L.C. also received funding from the Beatriu de Pinós fellowship program (funded by the Catalan Government and EU COFUND program nº 801370). PA is funded by ‘V Plan Propio de Investigación’ of the Universidad de Sevilla (Spain).

Conflicts of interest/Competing interests

The authors have no conflict of interest to disclose.

Availability of data and material

All data used in this manuscript have already been published or archived and can be accessible via Figshare *http://dx.doi.org/10.6084/m9.figshare.4234850* (Dyer et al. 2016), GBIF.org (2018 - for a complete list of GBIF occurrence downloads see Table S1), WorldClim database (Hijmans et al. 2005) available for download from *http://www.worldclim.org,* BirdLife International & NatureServe, (2014) available for download from *http://datazone.birdlife.org/species/requestdis and terrestrial ecoregions of the world* (Olson et al. 2001) *available for download from http://www.worldwildlife.org/publications/terrestrial-ecoregions-of-the-world*. Data on geographic extent and accuracy of model projections will be accessible via Figshare should the manuscript be accepted.

Code availability

Code associated with this manuscript will be accessible via Figshare should the manuscript be accepted.

Authors' contributions

'Not applicable'

Ethics approval

'Not applicable'

Consent to participate

'Not applicable'

Consent for publication 544

'Not applicable'

REFERENCES

Abellán P, Carrete M, Anadón JD, et al (2016) Non-random patterns and temporal trends (1912-2012) in the transport, introduction and establishment of exotic birds in Spain and Portugal. Divers Distrib 22:263–273. https://doi.org/10.1111/ddi.12403

Abellán P, Tella JL, Carrete M, et al (2017) Climate matching drives spread rate but not establishment success in recent unintentional bird introductions. Proc Natl Acad Sci 114:9385–9390. https://doi.org/10.1073/pnas.1704815114

Araújo MB, Peterson AT (2012) Uses and misuses of bioclimatic envelope modeling. Ecology 93:1527–1539. https://doi.org/10.1890/11-1930.1

Ascensão F, Latombe G, Anadón JD, et al (2020) Drivers of compositional dissimilarity for native and alien birds: the relative roles of human activity and environmental suitability. Biol Invasions. https://doi.org/10.1007/s10530-020-02196-7

BirdLife International, NatureServe (2014) Bird Species Distribution Maps of the World

Blackburn TM, Duncan RP (2001a) Determinants of establishment success in introduced birds. Nature 414:195–197. https://doi.org/10.1038/35102557

Blackburn TM, Duncan RP (2001b) Establishment patterns of exotic birds are constrained by non-random patterns in introduction. 927–939

Blackburn TM, Lockwood JL, Cassey P (2009) Avian invasions. The ecology and evolution of exotic birds. Oxford University Press, Oxford

Broenniman O, Petitpierre B, Randin C, et al (2014) Package ' ecospat '

Broennimann O, Fitzpatrick MC, Pearman PB, et al (2012) Measuring ecological niche overlap from occurrence and spatial environmental data. Glob Ecol Biogeogr 21:481–497. https://doi.org/10.1111/j.1466-8238.2011.00698.x

Broennimann O, Treier UA, Müller-Schärer H, et al (2007) Evidence of climatic niche shift during biological invasion. Ecol Lett 10:701–709. https://doi.org/10.1111/j.1461-0248.2007.01060.x

Broxton PD, Zeng X, Sulla-Menashe D, Troch PA (2014) A global land cover climatology using MODIS data. J Appl Meteorol Climatol 53:1593–1605. https://doi.org/10.1175/JAMC-D-13-0270.1

Capinha C, Essl F, Seebens H, et al (2015) The dispersal of alien species redefines biogeography in the Anthropocene. Science (80-) 348:1248–1251. https://doi.org/10.1126/science.aaa8913

Cardador L, Blackburn TM (2020) A global assessment of human influence on niche shifts and risk predictions of bird invasions. Glob Ecol Biogeogr

Cardador L, Blackburn TM (2019) Human-habitat associations in the native distributions of alien bird species. J Appl Ecol 56:1189–1199.

https://doi.org/10.1111/1365-2664.13351

Cardador L, Carrete M, Gallardo B, Tella JL (2016) Combining trade data and niche modelling improves predictions of the origin and distribution of non-native European populations of a globally invasive species. J Biogeogr 43:967–978. https://doi.org/10.1111/jbi.12694

Cardador L, Lattuada M, Strubbe D, et al (2017) Regional Bans on Wild-Bird Trade Modify Invasion Risks at a Global Scale. Conserv Lett 10:717–725. https://doi.org/10.1111/conl.12361

Cardador L, Tella J, Anadón J, et al (2019) The European trade ban on wild birds reduced invasion risks. Conserv Lett 12:e12631

D'Amen M, Zimmermann NE, Pearman PB (2013) Conservation of phylogeographic lineages under climate change. Glob Ecol Biogeogr 22:93–104. https://doi.org/10.1111/j.1466-8238.2012.00774.x

Dyer EE, Redding DW, Blackburn TM (2016) Data from: The global avian invasions atlas, a database of alien bird distributions worldwide. Figshare, https://doi.org/106084/m9.figshare4234850

Dyer EE, Redding DW, Blackburn TM (2017) The global avian invasions atlas, a database of alien bird distributions worldwide. Sci Data 4:170041. https://doi.org/10.1038/sdata.2017.41

Elith J, Kearney M, Phillips S (2010) The art of modelling range-shifting species. Methods Ecol Evol 1:330–342. https://doi.org/10.1111/j.2041-210X.2010.00036.x

Gaston KJ (2003) The structure and dynamics of geographic ranges. Oxford University Press, New York

GBIF.org (2017) GBIF Home Page

Godefroid M, Rasplus J-Y, Rossi J-P (2016) Is phylogeography helpful for invasive species risk assessment? The case study of the bark beetle genus Dendroctonus. Ecography (Cop) 39:1197–1209. https://doi.org/10.1111/ecog.01474

Guisan A, Petitpierre B, Broennimann O, et al (2014) Unifying niche shift studies: Insights from biological invasions. Trends Ecol Evol 29:260–269. https://doi.org/10.1016/j.tree.2014.02.009

Guisan A, Thuiller W (2005) Predicting species distributions: offering more than simple habitat models. Ecol Lett 8:993–1009

Haring E, Gamauf A, Kryukov A (2007) Phylogeographic patterns in widespread corvid birds. Mol Phylogenet Evol 45:840–862. https://doi.org/10.1016/j.ympev.2007.06.016

Hijmans RJ, Cameron SE, Parra JL, et al (2005) Very high resolution interpolated climate surfaces for global land areas. Int J Climatol 25:1965–1978

Hirzel AH, Gwenaëlle LL, Helfer V, et al (2006) Evaluating the ability of habitat suitability models to predict species presences. Ecol Modell 1999:142–152. https://doi.org/https://doi.org/10.1016/j.ecolmodel.2006.05.017

Holderegger R, Kamm U, Gugerli F (2006) Adaptive vs. neutral genetic diversity: implications for landscape genetics. Landsc Ecol 21:797–807. https://doi.org/10.1007/s10980-005-5245-9

Hulme PE (2009) Trade, transport and trouble: Managing invasive species pathways in an era of globalization. J Appl Ecol 46:10–18. https://doi.org/10.1111/j.1365-2664.2008.01600.x

Hulme PE, Bacher S, Kenis M, et al (2008) Grasping at the routes of biological invasions: A framework for integrating pathways into policy. J Appl Ecol 45:403–414. https://doi.org/10.1111/j.1365-2664.2007.01442.x

Huntley B, Green RE, Collingham YC, Willis SG (2007) A climatic atlas of European breeding birds. Durham University, The RSPB and Lynx Edicions, Barcelona

Joseph L, Dolman G, Donnellan S, et al (2008) Where and when does a ring start and end? Testing the ring-species hypothesis in a species complex of Australian parrots. Proc Biol Sci 275:2431–40. https://doi.org/10.1098/rspb.2008.0765

Lecocq T, Harpke A, Rasmont P, Schweiger O (2019) Integrating intraspecific differentiation in species distribution models: Consequences on projections of current and future climatically suitable areas of species. Divers Distrib 25:1088–1100. https://doi.org/10.1111/ddi.12916

Lenormand T (2002) Gene flow and the limits to natural selection. Trends Ecol Evol 17:183–189. https://doi.org/10.1016/S0169-5347(02)02497-7

Liu C, Berry PM, Dawson TP, Pearson RG (2005) Selecting thresholds of occurrence in the prediction of species distributions. Ecography (Cop) 3:385–393

Liu C, Wolter C, Xian W, Jeschke JM (2020) Species distribution models have limited spatial transferability for invasive species. Ecol Lett ele.13577. https://doi.org/10.1111/ele.13577

Liukkonen-Anttila T, Uimaniemi L, Orell M, Lumme J (2002) Mitochondrial DNA variation and the phylogeography of the grey partridge (Perdix perdix) in Europe: From pleistocene history to present day populations. J Evol Biol 15:971–982. https://doi.org/10.1046/j.1420-9101.2002.00460.x

Maia-Carvalho B, Vale CG, Sequeira F, et al (2018) The roles of allopatric fragmentation and niche divergence in intraspecific lineage diversification in the common midwife toad (Alytes obstetricans). J Biogeogr 45:2146–2158. https://doi.org/10.1111/jbi.13405

Mateo RG, Broennimann O, Petitpierre B, et al (2014) What is the potential of spread in invasive bryophytes? Ecography (Cop) n/a-n/a. https://doi.org/10.1111/ecog.01014

McKinney M, Lockwood J (1999) Biotic homogenization: a few winners replacing many losers in the next mass extinction. Trends Ecol Evol 14:450–453

Mori E, Menchetti M, Zozzoli R, Milanesi P (2019) The importance of taxonomy in species distribution models at a global scale: the case of an overlooked alien squirrel facing taxonomic revision. J Zool 307:43–52. https://doi.org/10.1111/jzo.12616

Olson DM, Dinerstein E, Wikramanayake ED, et al (2001) Terrestrial ecoregions of the world: a new map of life on Earth. Bioscience 51:933–938

Parmesan C, Ryrholm N, Stefanescu C, et al (1999) Poleward shifts in geographical ranges of butterfly species associated with regional warming. Nature 399:579–583. https://doi.org/doi.org/10.1038/21181

Pearman PB, Amen MD, Graham CH, et al (2010) Within-taxon niche structure : niche conservatism , divergence and predicted effects of climate change. https://doi.org/10.1111/j.1600-0587.2010.06443.x

Peterson AT (2011) Ecological niche conservatism: a time-structured review of evidence. J Biogeogr 38:817–827. https://doi.org/10.1111/j.1365-2699.2010.02456.x

Peterson AT, Holt RD (2003) Niche differentiation in Mexican birds: Using point occurrences to detect ecological innovation. Ecol Lett 6:774–782. https://doi.org/10.1046/j.1461-0248.2003.00502.x

Peterson ML, Doak DF, Morris WF (2019) Incorporating local adaptation into forecasts of species ’ distribution and abundance under climate change. Glob Chang Biol 775–793. https://doi.org/10.1111/gcb.14562

Phillips SJ, Anderson RP, Schapire RE (2006) Maximum entropy modeling of species geographic distributions. Ecol Modell 190:231–259.

https://doi.org/10.1016/j.ecolmodel.2005.03.026

Pimm SL, Russell GJ, Gittleman JL, Brooks TM (1995) The future of biodiversity. Science (80-) 269:347–350

Redding DW, Pigot AL, Dyer EE, et al (2019) Location-level processes drive the establishment of alien bird populations worldwide. Nature 571:103–106. https://doi.org/10.1038/s41586-019-1292-2

Sanderson EW, Malanding J, Levy MA, et al (2002) The human footprint and the last of the wild. Bioscience 52:891–904. https://doi.org//10.1641/0006-3568(2002)052[0891:THFATL]2.0.CO;2

Sayol F, Cooke RSC, Pigot AL, Blackburn TM, Tobias JA, Steinbauer MJ, Antonelli A, Faurby, S. (2021). Loss of functional diversity through anthropogenic extinctions of island birds is not offset by biotic invasions. Science Advances, in press.

Sheldon FH, Lohman DJ, Lim HC, et al (2009) Phylogeography of the magpie-robin species complex (Aves: Turdidae: Copsychus) reveals a Philippine species, an interesting isolating barrier and unusual dispersal patterns in the Indian Ocean and Southeast Asia. J Biogeogr 36:1070–1083. https://doi.org/10.1111/j.1365-2699.2009.02087.x

Smith AB, Godsoe W, Rodríguez-Sánchez F, et al (2019) Niche Estimation Above and Below the Species Level. Trends Ecol Evol 34:260–273. https://doi.org/10.1016/j.tree.2018.10.012

Smith BT, Escalante P, Hernández Bãos BE, et al (2011) The role of historical and contemporary processes on phylogeographic structure and genetic diversity in the Northern Cardinal, Cardinalis cardinalis. BMC Evol Biol 11:. https://doi.org/10.1186/1471-2148-11-136

Soberón J (2007) Grinnellian and Eltonian niches and geographic distributions of species. Ecol Lett 10:1115–1123

Soberon J, Peterson AT (2005) Interpretation of Models of Fundamental Ecological Niches and Species' Distributional Areas. Biodivers Informatics 2:1–10. https://doi.org/10.17161/bi.v2i0.4

Stockwell DR. B, Peterson AT (2002) Effects of sample size on accuracy of species distribution models. Ecol Modell 148:1–13. https://doi.org/10.1016/S0304-3800(01)00388-X

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

Strubbe D, Jackson H, Groombridge J, Matthysen E (2015) Invasion success of a global avian invader is explained by within-taxon niche structure and association with humans in the native range. Divers Distrib 21:675–685. https://doi.org/10.1111/ddi.12325

Vitousek PM (1997) Human Domination of Earth's Ecosystems. Science (80-) 277:494–499. https://doi.org/10.1126/science.277.5325.494

Warren DL, Glor RE, Turelli M (2008) Environmental niche equivalency versus conservatism: quantitative approaches to niche evolution. Evolution 62:2868–83. https://doi.org/10.1111/j.1558-5646.2008.00482.x

Zink RM, Pavlova A, Drovetski S, Rohwer S (2008) Mitochondrial phylogeographies of five widespread Eurasian bird species. J Ornithol 149:399–413. https://doi.org/10.1007/s10336-008-0276-z

Table 1. Bird species selected on the basis of the existing literature on phylogeographic structure. Species name, number of phylogroups per species, source for phylogenetic data and the number of occurrences available in the native and alien ranges according to GBIF (www.gbif.org) are provided. For native occurrences, the minimum and maximum occurrences available for different phylogroups are shown in parentheses.

Species	Phylogroups	References	Native occurrences	Alien occurrences
Alauda arvensis	2	Zink et al. 2008	19789 (232- 19557)	2378
Cardinalis cardinalis	6	Smith et al. 2011	41372 (9- 40079)	200
Copsychus saularis	3	Sheldon et al. 2009	3893 (20 - 3726)	36
Corvus frugilegus	2	Haring et al. 2007	7445 (52- 7393)	53
Perdix perdix	2	Liukkonen-Anttila et al. 2002	6444 (1804 - 6291)	2316
Pica pica	2	Haring et al. 2007	32052 (782 - 31270)	115
Platycercus elegans	2	Joseph et al. 2008	5046 (79- 4967)	2
Psittacula krameri	17	Strubbe et al. 2015	3156 (1-1834)	701

Figure Legends

Figure 1. Phylogroup distribution, native occurrences and model predictions for the Eurasian skylark *Alauda arvensis*. (a) Distribution of two different phylogroups (red and orange) as derived from Thiessen polygons and locations with genetic data (points); (b) native occurrence data (black dots) and background (i.e., all ecoregions occupied by each species in its native range, grey) used for species distribution modelling; (c-d) continuous predictions of single phylogroup models, polygons show the distribution of the phylogroup used in models; (e) species-level model continuous predictions; (f) phylogroup-composite model continuous predictions; (g) species-level model binary predictions; (h) phylogroup-composite model binary predictions. In (g) and (h), green represents habitat predicted as suitable and grey as non-suitable.

Figure 2. Comparisons of (a) extent of suitable habitat and (b – d) model accuracy in alien ranges between species distribution models accounting or not for phylogroup niche variation. Extent of suitable habitat (a) is derived from binary maps using the maximum sensitivity plus specificity threshold. Model accuracy in alien ranges is assessed by the AUC (b), TSS (c) and sensitivity (d). N = 8.

Figure 3. Similarity between single-phylogroup and species-level model predictions in relation to phylogroup relative range size. Similarity refers to the percentage of pixels predicted as suitable by a single-phylogroup model also predicted as suitable by species-level models at a global scale.

Figure 1.

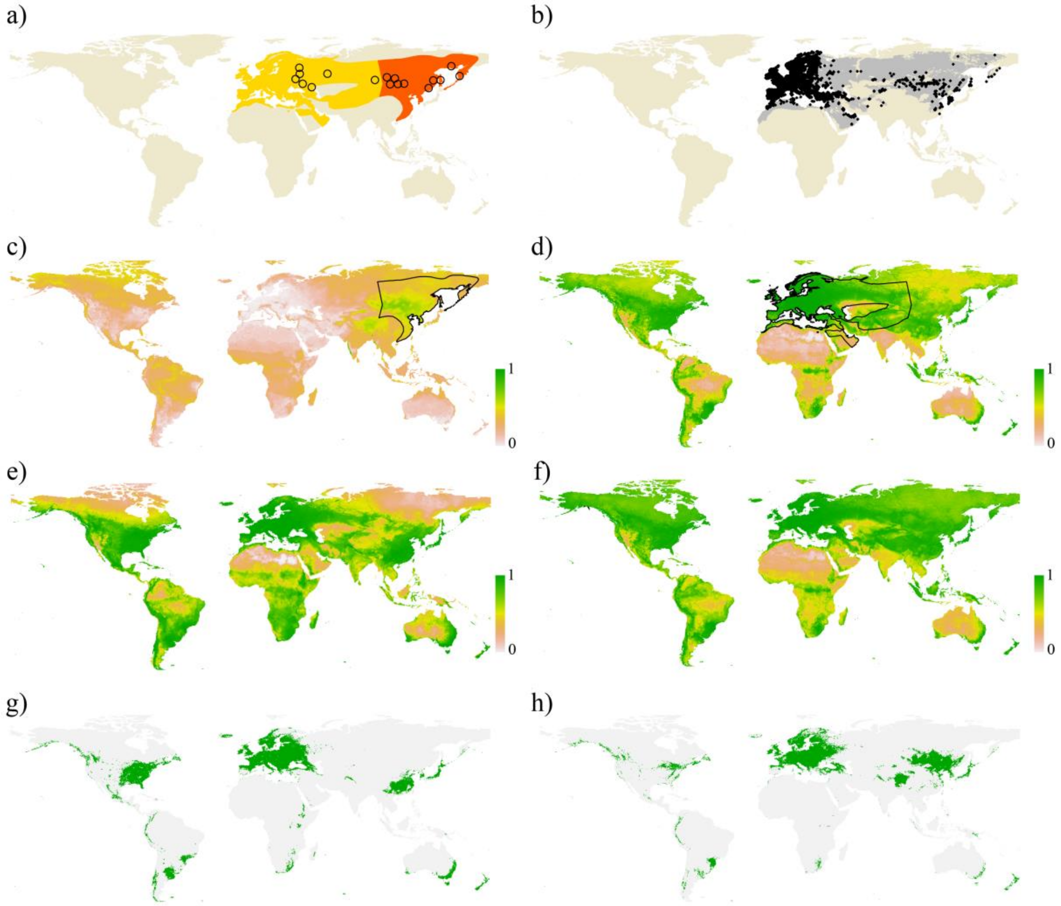
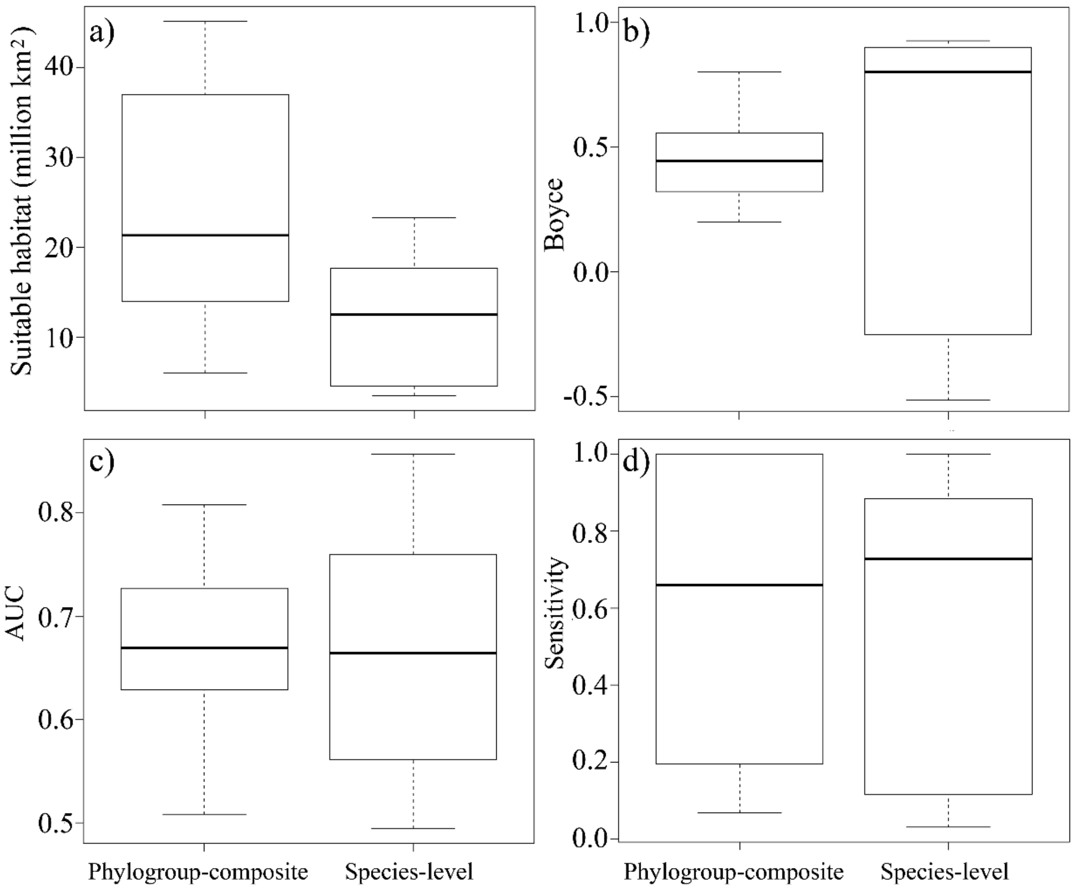
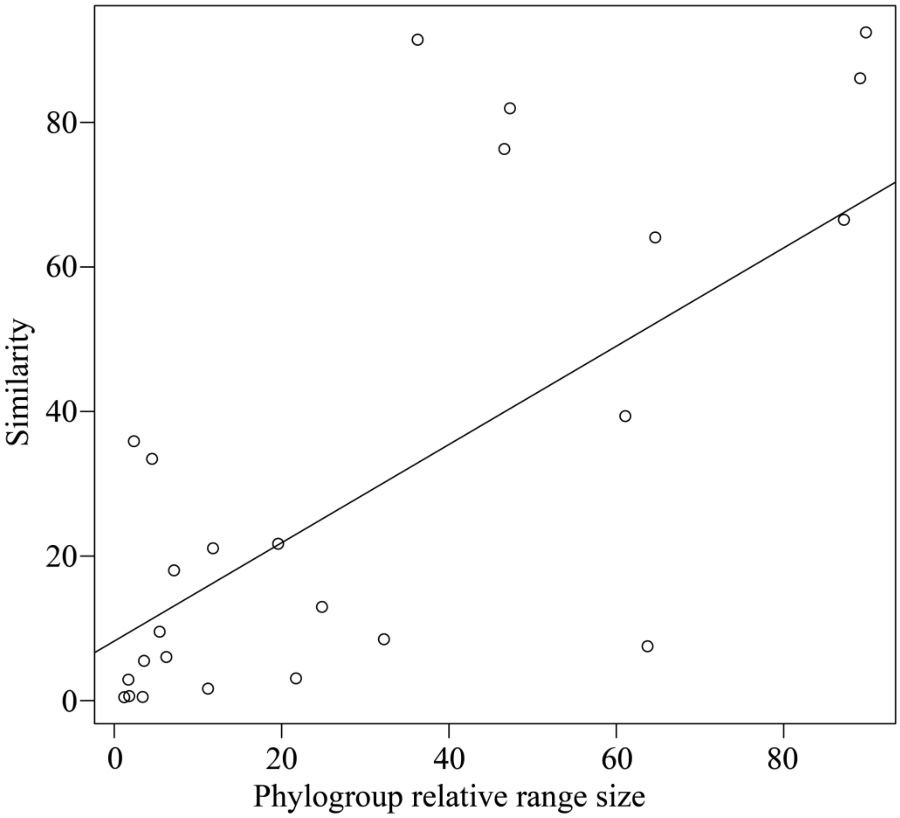


Figure 2.



Model comparisons: (a) t = -2.3, df = 7, P = 0.03; (b) V = 21, P = 0.69; (c) t = -0.09, df = 7, P = 0.47; (d) V = 9, P = 0.22

Figure 3.


Similarity
80
60
40
20
0
0
20
40
60
80
Phylogroup relative range size