Title

Biogeography of the world's worst invasive species has spatially biased knowledge gaps but is predictable

Permalink

Journal

Authors

Jenkins, David G.
Bevan, Hannah R.
Chen, Wei
et al.

Publication Date

DOI

Copyright Information

Peer reviewed

Biogeography of the world's worst invasive species has spatially-biased knowledge gaps but is predictable

David G. Jenkins[1]* (iD), Hannah R. Bevan[1] (iD), Wei Chen[1] (iD),
Jacob D. Hart[1] (iD), Amanda Lindsay[1], Laura Macamo[1] (iD),
Mekail Negash[1] (iD), Leo Ohyama[2] (iD), Alessandra Pandolfi[1] (iD),
and George Zaragoza[1] (iD)

[1]Department of Biology, University of Central Florida, Orlando, FL 32816, USA;

[2]Biodiversity Institute, University of Florida, Gainesville, FL, USA.

*Correspondence: David G. Jenkins, david.jenkins@ucf.edu

Abstract

The world's "100 worst invasive species" were listed in 2000. The list is taxonomically diverse, often cited (typically for single-species studies), and its species are frequently reported in global biodiversity data bases. We acted on the principle that these notorious species should be well-reported to help answer two questions about global biogeography of invasive species (i.e., not just their invaded ranges): (1) "how are data distributed globally?" and (2) "what predicts diversity?" We collected location data for each of the 100 species from multiple data bases; 95 had sufficient data for analyses. For question (1), we mapped global species richness and cumulative occurrences since 2000 in (0.5 degree)^2 grids. For question (2) we compared alternative regression models representing non-exclusive hypotheses for geography (i.e., spatial autocorrelation), sampling effort, climate, and anthropocentric effects. Reported locations of the invasive species were spatially-biased, leaving large gaps on multiple continents. Accordingly, species richness was best explained by both anthropocentric effects not often used in biogeographic models (Government Effectiveness, Voice & Accountability, human population size) and typical natural factors (climate, geography; $R^2 = 0.87$). Cumulative occurrence was strongly related to anthropocentric effects ($R^2 = 0.62$). We extract five lessons for invasive species biogeography; foremost are the importance of anthropocentric measures for understanding invasive species diversity patterns and large lacunae in their known global distributions. Despite those knowledge gaps, advanced models here predict well the biogeography of the world's worst invasive species for much of the world.

Highlights

- Countries with more effective governments and accessible information are more likely to monitor and report invasive species.

- Despite geographically-biased data, richness of the world's worst invasive species is predicted well globally. Occurrences are also predictable but with more residual variation.

- Greater access to species identification and reporting applications on mobile devices in regions with limited current data will greatly improve understanding of invasive species distributions. Invasive species are best known where governments are most effective and information is freely available; the same may also be true for native species data. This represents an accessory benefit of government reform, given that invasive species mitigation and control depends on reported locations.

Keywords: climate, geography, Government Effectiveness index, human population, nonnative, occurrence, species richness, Voice & Accountability index

Introduction

Humans have a long history of introducing species from native ranges into new habitats, which expands biogeographic distributions of those species and can cause ecological effects in the introduced range. An introduced species may become invasive (i.e., causing ecological or socioeconomic costs), but this process differs among species and spatiotemporal contexts; varied concepts and terms have accrued (Colautti and MacIsaac 2004, Falk-Petersen et al. 2006, Pereyra 2016, Courchamp et al. 2017). Some species do not clearly become invasive because evidence is difficult to obtain or a species exhibits weak effects (e.g., Hasselman et al. 2012). Other species clearly exert ecological, evolutionary, and economic effects (e.g., Mooney and Cleland 2001, Ehrenfeld 2011, Bennett et al. 2011, Paini et al. 2016, Crystal-Ornelas and Lockwood 2020). Ways to sort invasive from non-invasive species are emerging (e.g., Kolar and Lodge 2001, Guo et al. 2022), but multiple challenges remain regarding jargon, expectations, and impacts of invasive species (e.g., Richardson and Ricciardi 2013, Ricciardi and Ryan 2018, Kourantidou et al. 2022).

Some clarity for understanding invasive species may come from studying the *most* invasive species because the extra attention applied to notorious species yields more data than for lesser-known species (e.g., Loss et al. 2013). One hundred of the world's worst invasive species were listed and described in 2000 by the Invasive Species Specialist Group at the International Union for Conservation of Nature (IUCN) to "enhance awareness of the fascinating complexity, and also the terrible consequences, of invasive alien species" (Lowe et al. 2000). One species per genus was selected if it had a "serious impact on biodiversity and/or human-related activities" and represented important aspects of biological invasion (Lowe et al. 2000). The list brought substantial attention to a diverse set of selected species, as evidenced by 5,300 citations (per Google Scholar as of November 2023), where papers typically address a single species in the list. Accordingly, the world's worst invasive species (Lowe et al. 2000) are a well-documented and diverse sample of global invasive species distributions that now represents >20 years of data. We regarded the list as a robust resource to potentially help understand general diversity patterns of invasive species, including invasive species that are less notorious. We evaluated available data for the listed species to cautiously consider that potential, and then developed models to predict diversity and abundance of the listed species using the available data.

Others have examined the biogeography of multiple invasive species but have worked at relatively coarse spatial scales (i.e., continents or countries/states), and/or focused on different questions, including species-area accrual (van Kleunen et al. 2015), economic costs (Cuthbert et al. 2022), hot spots for one clade in a region (e.g., Schneider et al. 2021), or the effect of trade (Westphal et al. 2008). Unlike prior research, we answered two questions: (1) "how are data distributed globally?" and (2) "what predicts diversity?" for the 100 worst invasive species (Lowe et al. 2000). Our answer to the first question critically evaluates available data accrued since 2000, at relatively fine spatial scale (not per country) and available in publicly available online data bases (other data may exist elsewhere). Spatial knowledge gaps are expected (e.g., Meyer et al. 2015, Nori et al. 2023), but not yet described for this list of notorious invasive species. Because we sought global patterns for the listed species, we did not separate putative native and non-native distributions. Our answer to the second question ("what predicts diversity?") evaluates statistical predictors of the available data. We tallied and mapped global records from databases at (0.5 degree)^2 resolution and then statistically modeled species richness and records at that spatial grain accrued during the 22-year interval 2000-2021. We addressed spatial autocorrelation in patterns (Bahn et al. 2006) and used generalized linear mixed models (GLMMs) to evaluate predictors of complex spatial patterns while handling skewed data distributions. Hypothesized predictors included classic climate effects (Fridley et al. 2007) and anthropocentric factors, including socioeconomic measures and governments (Table 1). Socioeconomic and governmental predictors are rarely used to model invasive species distributions but can be important to results (Brenton-Rule et al. 2016, Pouteau et al. 2021). We expected anthropocentric factors (Table 1) to reflect observer effects (e.g., more people may report more invasives) and/or information availability (e.g., via more open government and stronger economic support), separately from effects of land cover (here anthropogenic biomes, or anthromes; Ellis and Ramankutty 2013).

Our study differed from other studies of invasive species in several ways. Foremost are the goals here; rather than study potential *processes* leading to their invasive status, we more simply evaluated the prerequisite matter of global *patterns* in available data for these well-known but taxonomically-diverse species. Thus, our analyses appraise the current state of knowledge for invasive species that should be well-studied. We followed up by evaluating predictors of those collective data patterns but did not evaluate species individually. Also, this work used global distributions of the invasive species rather than focus only on invaded ranges because: (a) our questions were broader than a focus on non-native ranges; (b) native range data are often essential to model ongoing invasions; and (c) despite common expectations, native ranges are not often clearly described, especially at relative fine spatial grains used here. As a result of the approach applied here, results are organized as cautionary "lessons learned" rather than about invasion processes.

Materials & Methods

Data Acquisition and Processing

Data were acquired from multiple data bases for the 100 invasive species in February 2022 using the spocc package in R (Chamberlain 2021).

Data sources (in alphabetical order) included: the Atlas of Living Australia ('ALA'; https://www.ala.org.au); eBird (http://www.ebird.org/home; Sullivan et al. 2009); the Integrated Digitized Biocollections ('iDigBio'; https://www.idigbio.org; Matsunaga et al. 2013); the Global Biodiversity Information Facility (GBIF (https://www.gbif.org); Ocean 'Biogeographic' Information System ('OBIS'; https://portal.obis.org; Grassle and Stocks 1999); VertNet (https://vertnet.org; Constable et al. 2010); and the US Geological Survey's Biodiversity Information Serving Our Nation ('BISON'; replaced December 2021 by GBIF). Several databases set limits to 100,000 initial point records (before cleaning, described below) when accessed using spocc. As a result, data for 19 species with >100,000 point records (e.g., the European starling (*Sturnus vulgaris* Linnaeus) had >23 million point records) were obtained directly from GBIF on 23-25 February 2022, which included records already contributed to GBIF from multiple databases.

All searches were based on genus and species epithets, where taxonomic changes in the historical records required decisions. Where an epithet changed since the 100 species were listed in 2000 (Lowe et al. 2000), both former and current names were searched and concatenated. For example, Lowe et al. (2000) listed the American bullfrog as *Rana catesbeiana* Shaw, 1802 which is now *Lithobates catesbeianus* (Shaw, 1802); both were included in searches, as well as synonyms. Taxonomic synonyms were resolved by referring to the Catalogue of Life (https://www.catalogueoflife.org/) Centre for Agriculture and Bioscience International (http://www.cabi.org) and World Flora Online (http://worldfloraonline.org). Listed synonyms and new combinations were included in data, whereas undocumented synonyms (i.e., provided in a database but not resolved above) were excluded. Database entries that lacked species epithets (i.e., genus only) were excluded and all identities were at the species level. Some taxa formerly identified in (Lowe et al. 2000) as a species are now subspecies (e.g., the red-ear slider *Trachemys stricta* (Thunberg in Schoepff, 1792) is now *Trachemys stricta elegans* (Wied-Neuwied, 1838)). For those taxa, data may be more inclusive in current taxonomy than the original intent. However, our use of species-level identities includes sub-specific hybrids (e.g., Parham et al. 2020). Overall, our approach: matched the taxonomic resolution of (Lowe et al. 2000); recognized variation through time and space; and included potential hybrids among subspecies.

Table 1. Alternative, non-exclusive hypotheses evaluated here, with predictors used. Each row below completes a hypothesis statement: "[Concept] best describes invasive species distributions because [Reason] as measured by [Predictors]." Note that multiple predictors were evaluated for climate and anthropocentric effects, and a null model was also evaluated. Random effects were included in all subsequent models (explained below) but omitted here for simplicity. Cumulative records were also used as a covariate for species richness models to address sampling effects on species detection.

Concept	Reason	Predictors
Random effects	anthropocentric effects are reported by countries, and land use patterns drive suitable habitat for invasive species	countries and anthromes
Geography	spatial autocorrelation is a rule (Tobler's 1st Law of Geography) and should apply here	longitude and latitude
Climate	invasive species respond to climate, similar to native species	annual mean temperature, annual mean precipitation
Geography + Climate	spatial autocorrelation and climatic effects (above) are not mutually exclusive	longitude and latitude, annual mean temperature, annual mean precipitation
Anthropocentric Effects	human socioeconomic and political systems affect invasive species distributions (introductions, spread, monitoring) as well as via monitoring and reporting infrastructure	human population, gross domestic product per capita (GDP$_{pc}$), Government Effectiveness, Voice and Accountability
Geography + Anthropocentric Effects	spatial autocorrelation and anthropocentric effects (above) are not mutually exclusive	longitude and latitude, human population, GDP$_{pc}$, Government Effectiveness, Voice and Accountability
Geography + Climate + Anthropocentric Effects	the hypotheses (above) are not mutually exclusive. In addition, Government Effectiveness may interact with human population to cause greater effects in more populous nations	longitude and latitude, annual mean temperature, annual mean precipitation, human population, GDP$_{pc}$, Government Effectiveness, Voice and Accountability

We set a threshold for a species to be included in analyses at > 30 records because we judged distributions with fewer records to be inadequately represented. As a result, four species (notably disease agents or vectors) had too few data to be analyzed here: *Aphanomyces astaci* Schikora, 1906, *Cinara cupressi* (Buckton, 1881), *Plasmodium relictum* (Grassi & Feletti, 1891), and *Trogoderma granarium* Everts, 1898. Likewise, banana bunchy top virus was not present in databases, despite a reported global distribution (https://www.cabi.org/isc/datasheet/8161). As a potential alternative, we searched for its aphid vector (*Pentalonia nigronervosa* Coquerel, 1859) but obtained records that fully lacked Africa and Asia, despite the widespread tropical distribution of the virus. We thus treated banana bunchy top virus as an under-reported species and omitted it here. Finally, rinderpest was listed by Lowe et al. (2000) but since eradicated (Morens et al. 2011). Following Luque et al. (2014), we replaced rinderpest with *Salvinia molesta* D. S. Mitch, leaving 95 species to evaluate.

Species data were cleaned using two R packages. The scrubr package (https://github.com/ropensci/scrubr) was used with default settings to exclude records with geographic coordinates that were lacking, impossible, incomplete, imprecise, or unlikely. Data were further cleaned using the CoordinateCleaner package (Zizka et al. 2019), where records were excluded if geographic coordinates were zero (i.e., a flag for probable data error), near a country's capital and geographic centroid, or near administrative locations (e.g., museums, GBIF headquarters). Data were then restricted to unique spatio-temporal records during the years 2000-2021 to exclude duplicate entries. This step also omitted older records that tend to have greater taxonomic and geographic uncertainty (e.g., GPS selective availability was removed in 2000). Finally, resulting maps were visually examined, where oddities (e.g., a tropical species located on Baffin Island or a terrestrial species in mid-ocean) were manually excluded from data. That last step removed a few locations per species, if any. As a result of the above process, data were cleaned to be conservative for errors in geographic distribution and consistent in taxonomy with Lowe et al. (2000).

Aggregation and Mapping

We spatially aggregated point data per species in (0.5 degree)^2 grid cells, using the World Geodetic System (WGS84); the basis for the geospatial positioning system. Thus analyses below and summarized data refer to (0.5 degree)^2 grid cells as units of study. Aggregation in space simplified variable coordinate accuracy in original records while retaining substantial resolution for global analyses. For two reasons that affected interpretations, we also aggregated data in time by pooling all records obtained for years 2000-2021. First, species richness is then based on presence/absence of reported species at any time during 22 consecutive years and should be sensitive to infrequent observations or occurrences. We reasoned that species absence maintained through two decades was either: (a) likely true or (b) due to

lack of submitted records for that location, where the difference may be inferred from spatial patterns of records. Secondly, the difference between species richness (i.e., presence/absence) and cumulative occurrence was enhanced. Species richness is fully insensitive to commonality or rarity; a single record here obtains the same result as daily repeats for 22 years. In contrast, cumulative occurrences may range from 0 to thousands of (0.5 degree)^2 pixels during 22 years and could indicate commonality or rarity. Therefore, fundamental differences between species richness and occurrences were enhanced here by using data for years 2000-2021. We mapped species richness and cumulative records to address question 1 ("how are data distributed globally?").

Potential Predictors of Invasive Species

We analyzed spatial autocorrelation (using longitude and latitude of 0.5° grid cell centers) with local estimation ("loess") regression to obtain a surface representing only geographic coordinate effects. Loess regression is a robust, nonparametric approach to represent a complex spatial surface (Ferrier et al. 2002, Helsel and Ryker 2002) and is not too computationally-intensive for fine-grained global data, unlike approaches based on covariance matrices or network meshes. The spatial texture of a loess regression surface is determined by its span, where values <1 have more texture and values >1 are smoother. We modeled species richness and cumulative records using the loess command in R, with degree = 2 (i.e., 2nd order polynomial) and least-squares fitting. We iteratively adjusted span to minimize the residual standard error and maximize the correlation between predicted values and actual total records. Predicted values represented spatial autocorrelation alone.

Subsequent models using additional predictors (Table 1) included predictions from the loess model to evaluate those additional effects after already accounting for spatial autocorrelation. In addition, hierarchical structure (i.e., non-independence of grids) of grid cells within countries and anthropogenic biomes (anthromes) was handled using spatial GLMMs (Dormann et al. 2007).

All other predictors were matched to the 0.5° gridded species data using projectRaster in the R raster package. Climate conditions were obtained from WorldClim (Fick et al. 2017) per grid cell and represented by annual mean temperature (BIO1) and annual mean precipitation (BIO12). Initial scatter plots of data indicated potential curved relationships between invasive species measures (i.e., richness and total occurrence) and climate variables, consistent with expected general tolerance limits given global extremes. Accordingly, we included quadratic terms for those predictors in models (because a quadratic term is the most parsimonious means to include curvature in a linear model).

Anthropocentric effects were represented by four measures. Human population size in 2010 was used to represent the 2000-2021 interval based on the decadal census and obtained from the Socioeconomic Data and Applications Center (SEDAC; https://sedac.ciesin.columbia.edu/data/collection/gpw-v4).

We expected human population size to represent multiple potential effects, including observer effects and/or demographic effects *per se* (e.g., more people releasing invasive species, and/or causing different land use intensities). We obtained *per capita* GDP (GDP$_{pc}$) data from (Kummu et al. 2018) and used the median of 2000-2015 per country to represent economic effects.

Potential governmental effects were evaluated using two indices of governmental systems from the World Bank (Kaufmann et al. 2011). We used median scores per country for 2000-2020 to represent a country for each index. Governmental Effectiveness scores the quality of public and civil services, governmental independence from political pressures, quality of policy formulation and implementation, and credibility of governmental commitment to policies. Governmental Effectiveness is strongly and inversely correlated with the Control of Corruption score, where greater corruption in a source nation predicted more invasive species intercepted in trade to New Zealand (Brenton-Rule et al. 2016). We expected greater Governmental Effectiveness would either negatively predict invasive species richness (Brenton-Rule et al. 2016) if effective governments also control invasive species well, or positively predict invasive species richness if they instead report them more often.

We also evaluated Voice & Accountability (Kaufmann et al. 2011) as a predictor, which indicates the extent that citizens can participate in selecting their government, an uncensored media and freedoms of expression and association. Voice & Accountability is a component of the more inclusive Governmental Effectiveness, and thus potentially collinear. We examined potential collinearity while including Voice & Accountability in analyses because we expected greater Voice and Accountability would predict greater invasive species if reporting was inversely related to censorship. To be clear, all human socioeconomic and governmental predictors may represent positive (e.g., sampling and reporting) and/or negative (e.g., extirpation) effects on invasive species distributions, where resulting signs of model coefficients would help infer net effects.

Statistical Analyses

To address question 2 (what predicts diversity?), we analyzed hypothesized drivers of patterns (Table 1) alone and in combinations using GLMMs with fixed and random effects. Estimated fixed effects represented mean effects for hypothesized predictors, whereas random effects represented the spatial nesting of grid cells within larger spatial categories of countries and anthromes. Both countries and anthromes were used here as random effects because we expected each to represent different effects and because preliminary modeling showed that using both terms was more effective than either alone. Models representing random effects only and a null model (i.e., intercept-only model) were also computed and compared to models of main interest.

Based on residual distributions and iterative model selection, we used zero-inflated models, which include two parts: conditional and zero-inflation models.

The conditional model estimated linear effects of predictors after having accounted for "excess" zeroes in a logistic (i.e., binary) zero-inflation model. This detail was important to results because models used the many apparent "zeroes" (due to true absences or sampling effects) in data to generate predicted values in those unreported locations. We compared model predictions to observed data to validate model results, especially for predicting invasive species in locations lacking observations.

We used cumulative occurrences per grid cell as a covariate in species richness models to represent direct sampling effects on recorded species (akin to rarefaction), based on the expectation that more occurrences represent more reported observations. Thus, species richness results here were statistically distinct from those for cumulative records and were adjusted for sampling effects. In those same models, we then treated Government Effectiveness and Voice & Accountability as representing indirect effects due to infrastructure to report invasive species and access to those data.

Analyses of both species richness and cumulative records used the glmmTMB package in R (Brooks et al. 2017) with a zero-inflated model based on a truncated negative binomial distribution (nbinom2 with a log link). Potential alternative families (nbinom1 and Gamma) were found either less efficient or could not be solved. Importantly for interpretations, all fixed effect predictors were numeric and scaled (as Z scores) for fair, quantitative comparisons of effect sizes within a model, despite varied units. A scaled effect size with 95% confidence interval that did not include zero was interpreted as having a clear effect (Dushoff et al. 2019). Because we considered scaled terms (expressed as standard deviations of an effect) to be important to interpretations, we log-transformed human population, precipitation, and GDP$_{pc}$ to better match Z score calculations because original data ranged greatly and had strongly skewed distributions. Importantly for interpretations of results, we retained zeroes in predictors for zero-inflation models by first adding one to transformed predictors. Thus models extended predicted values to locations without reported species, rather than omitting many locations from analyses and predictions.

Alternative models were compared using the Akaike Information Criterion (AIC) with the bbmle package (Bolker and R Core Team 2022). Model comparisons emphasized AIC weight (w_i; the probability that a model is most efficient among those listed) and δAIC, where values >2 indicate clear model choice (Burnham and Anderson 2002). The use of AIC is asymptotically equivalent to leave-one-cluster-out cross validation (Fang 2011) and ensured the most efficient model (among those compared) was discussed. Model predictors were quantitative and scaled as standard deviations for fair comparison of relative effects on invasive species measures. A scaled effect size with $p < 0.05$ (i.e., its 95% confidence interval did not include zero) was interpreted as having a clear effect (Dushoff et al. 2019).

The performance package (Lüdecke et al. 2021) was used to evaluate model assumptions, potential collinearity (using variable inflation factors; VIF) and overdispersion. The fit of each final model was critiqued using Nakagawa's R^2 for fixed (i.e., marginal effects) and fixed + random (i.e., conditional, or total) effects. Code and data files are available at the link in the Data Accessibility Statement.

Results

Mapped Data

Cleaned and gridded data for all results included 56,775 gridded locations (96.5% of global locations on land) distributed in 160 countries and 52 anthromes (Fig. 1). Large areas did not have reports for any of the 95 invasive species; most reported species (Fig. 1A) and cumulative records (Fig. 1B) were from North America, Europe, and eastern Australia. In addition, human population centers within regions had many reports of invasive species.

Species Richness 2000-2021

The most inclusive species richness model (i.e., based on spatial autocorrelation, climate, anthropocentric effects and random effects of country and anthrome; Table 1) was clearly more plausible than less inclusive models, based on AIC weight =1.000 and δAIC values >> 2. The model fit data well, and fixed effect predictors were important to that model fit (marginal R^2 = 0.82; conditional R^2 = 0.87; Table 2). As expected, Voice & Accountability and Governmental Effectiveness were moderately collinear (VIFs = 4.6 and 5.8, respectively). All other model terms had low VIFs (< 3.25).

Countries were especially important in predicting zeroes in data (see Zero Inflation model in Table 2). All predictors except GDP_{pc} had clear effects on species richness after accounting for all other analyzed effects (Table 2). In preliminary models, we had expected cumulative records would be important to the zero-inflation model, but cumulative records more plausibly "explained" species richness as a fixed effect in the conditional model. Only anthropocentric predictors and geography (spatial autocorrelation) were retained in the zero-inflation model (Table 2), and Governmental Effectiveness and Voice & Accountability were especially important to the zero-inflation model.
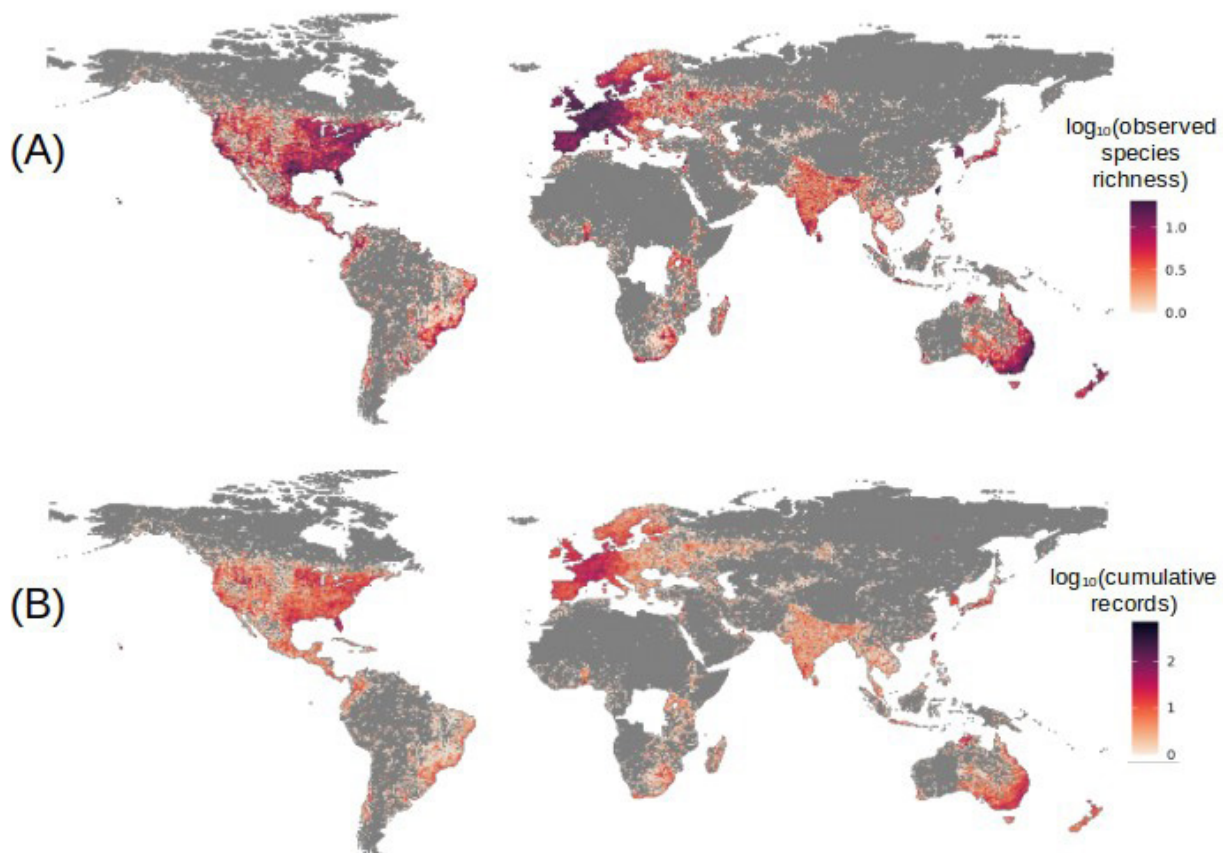
Fig 1. Global diversity of the world's worst invasive species (Lowe et al. 2000) during 2000-2021. (A) species richness and (B) cumulative records per grid cell during that interval. Spatial grain = 0.5° x 0.5°. Log scales are used to better illustrate areas with few species or no data (grey).

Random effects (especially countries; Table 2) were important to the species richness model and represented different spatial patterns (Figs. 2B & 2C). Among fixed effects, species richness was most strongly affected by mean annual temperature, human population, Government Effectiveness, and Voice & Accountability (Table 2). A strong, hump-shaped effect of temperature was supported, and weaker quadratic effects of other shapes (e.g., upward curving) were supported for effects of human population and precipitation (Fig. 3). Finally, an interaction between Government Effectiveness and quadratic human population was partially supported (Table 2).

Predicted species richness essentially smoothed observed patterns (compare Figs. 1A and 2A), and the model predicted <1 species where no species were observed and tended to conservatively predict species richness where few existed (Fig. 4A). Overall, the species richness model predicted well the collective global distributions of the world's worst invasive species and accounted for the many zeroes, spatial autocorrelation and sampling effects. Among predictors of main interest, patterns were primarily a function of climate, Governmental Effectiveness, Voice & Accountability, and human population size.

Table 2. Summary of statistical model results for cumulative species richness (2000-2021) of 95 of the world's worst invasive species. Data include 56,775 grid cells (0.5° x 0.5°) in 160 countries and 52 anthromes. The model was clearly most efficient (AIC weight = 1.000, all δAIC >>2) and includes a conditional model for quantitative records after taking into account the binomial zero-inflation model (which predicts absence (0) of invasive species). Effect sizes are scaled as Z scores (+ 95% confidence intervals, or CI) for fair comparisons and expressed in addition to the intercept term. Geographic effect was estimated as a loess surface model (richness ~ longitude * latitude) with span = 0.007 (i.e., 410 neighboring grid cells). Marginal R^2 (fixed effects) = 0.825; conditional R^2 (fixed + random effects) = 0.871.

| Predictors	Coefficient	p value
Conditional Model		
Random Effects: Standard Deviation		
anthromes	0.038	- -
countries	0.335	- -
Fixed Effects: Scaled Effect Size (95% CI)		
(Intercept)	-0.756 (0.116)	<0.001
geography	0.100 (0.010)	<0.001
cumulative records	0.523 (0.014)	<0.001
annual mean temperature	0.651 (0.091)	<0.001
(annual mean temperature)^2	-0.387 (0.059)	<0.001
annual mean precipitation	-0.098 (0.174)	0.27
(annual mean precipitation)^2	0.144 (0.147)	0.05
Voice & Accountability	0.258 (0.123)	<0.001
Governmental Effectiveness (GE)	0.377 (0.129)	<0.001
human population	0.460 (0.260)	<0.001
(human population)^2	0.150 (0.158)	0.06
GDP_{pc}	0.019 (0.059)	0.54
$(GDP_{pc})^2$	-0.005 (0.051)	0.86
GE: human population	-0.018 (0.179)	0.84
GE: (human population)^2	-0.164 (0.109)	0.003
Zero-inflation Model (predicts zeroes)		
Random Effects: Standard Deviation		
anthromes	0.273	- -
countries	1.228	- -
Fixed Effects: Scaled effect sizes as odds ratios (upper & lower CIs)		
(Intercept)	2.042 (1.522, 2.739)	<0.001
geography	0.078 (0.069, 0.088)	<0.001
Governmental Effectiveness	0.342 (0.225, 0.519)	<0.001
Voice & Accountability	0.620 (0.405, 0.949)	0.03
human population	0.183 (0.170, 0.198)	<0.001

Cumulative Records

Whereas species richness represented the reported presence of invasive species any time during the years 2000-2021, cumulative records represented reported abundances in that period. Cumulative records were modeled well, though not as well as species richness; fixed effects R^2 = 0.51; total effects R^2 = 0.68 (Table 3), which is also indicated by comparing model validation plots (Fig. 4). Random effects were again important to models, where countries represented more variation than anthromes in both zero-inflation (which predicts zeroes) and conditional models (Table 3). All predictors here had low collinearity (VIF < 3.1).

Interestingly, fewer predictors clearly affected cumulative records than for species richness: only temperature and the interaction of Governmental Effectiveness and human population had a clear effect, after accounting for geography and zeroes (Table 3), where the relationships looked very similar to those for species richness (Fig. 3). Geography and anthropocentric predictors were again important to the zero-inflation model, where Voice & Accountability and Governmental Effectiveness again had the strongest effects.

As was the case for species richness, predictions from the cumulative records model generally smoothed and captured the global invasive species distributions to represent regional or continental patterns (compare Figs. 1B and 5A). Places with numerous reported occurrences (e.g., Florida in the US, Western Europe, Eastern Australia) remained "hot" in the mapped model predictions. However, neighboring areas were also predicted to have numerous occurrences that exceeded observed patterns. Whereas model validation for species richness (Fig. 4A) provided some assurance of accuracy, the cumulative occurrences model was less accurate in its predicted values, especially where fewer occurrences had been reported (Fig. 4B).
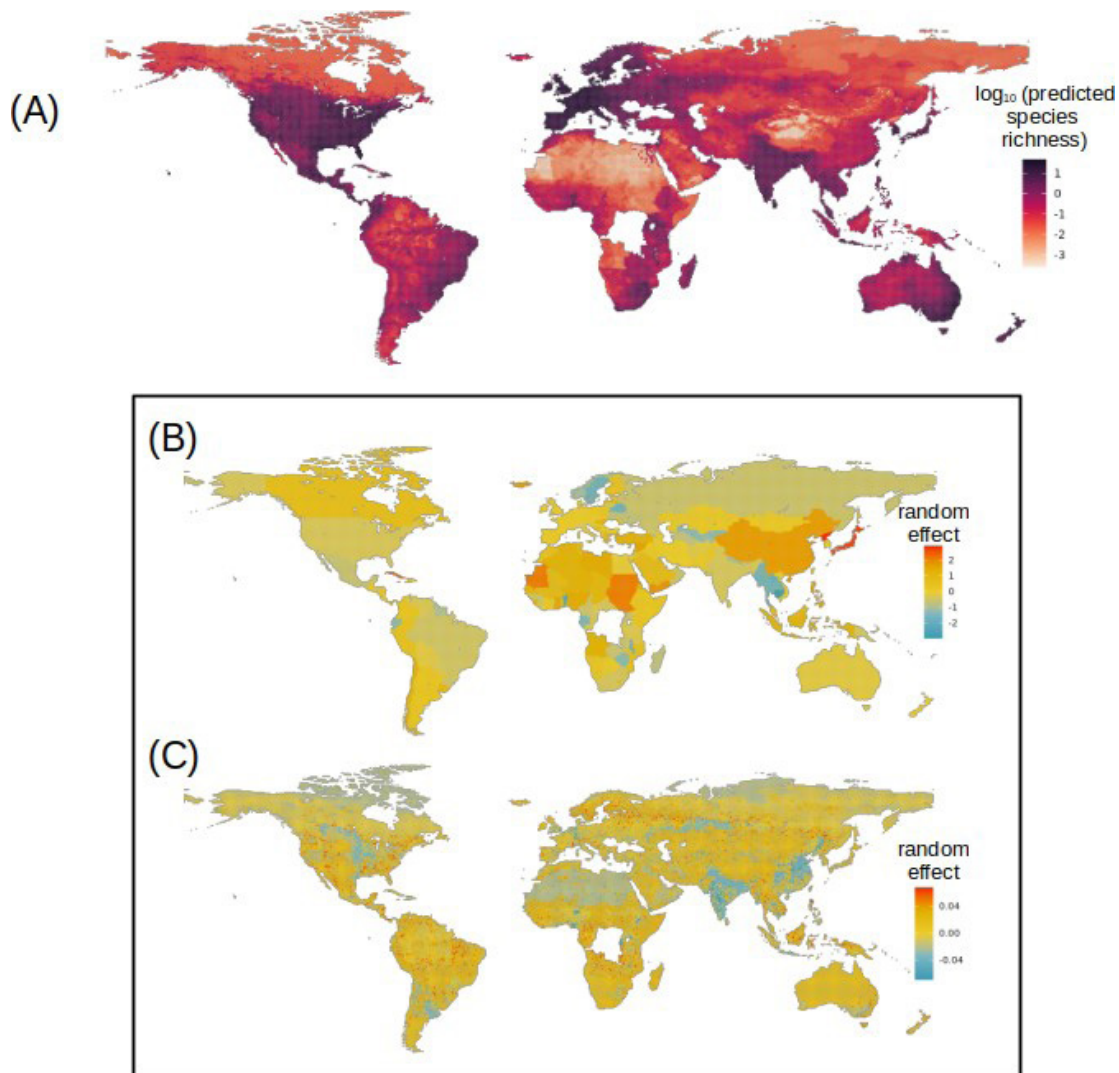
Fig 2. (A) Predicted species richness for 95 of the 100 world's worst invasive species (Lowe et al. 2000), 2000-2021. Compare to Fig. 1A; the model predicted very low species richness where no species were reported. Note predicted values are mapped as $\log_{10}$-transformed values to make low values more apparent. Predictions come from the model detailed in Table 2. Countries (B) and (C) anthromes varied as random effects in the mixed effects model.
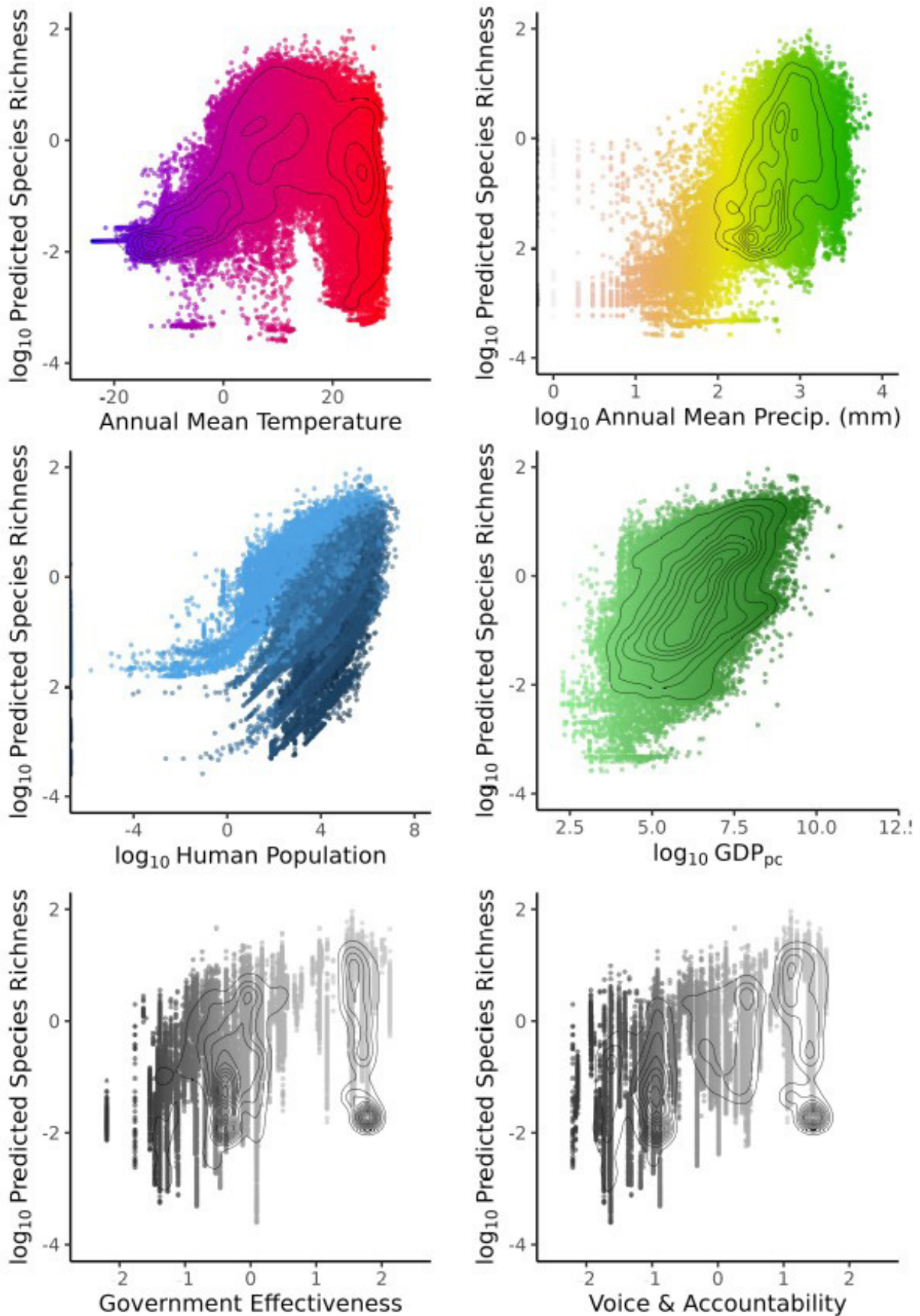
Fig 3. Predicted species richness as a function of model predictors. Note that axes change for predictors and species richness is log-scaled to help visualize patterns. The human population plot is colored to represent the interaction with Government Effectiveness (GE), where lighter blue indicates greater GE.

We concluded that the invasive species cumulative records model was useful to compare among predictors, but we discounted results of the cumulative occurrences model in interpretations of details (e.g., maps).

Overall, models predicted well the distributions of the world's worst invasive species since 2000, within regions where those species are reported in publicly accessible global databases. Species richness was especially predictable; cumulative records tended to overestimate counts values where they were low. In both cases, the modeled distributions represent smoothed surfaces based on reports of these worst invasive species.

Discussion

The list of the world's 100 worst invasive species (Lowe et al. 2000) is often cited on a per-species basis. To our knowledge, predictors of biogeographic diversity patterns for the listed species - as a group - have not been examined. We expected notoriety of the listed species to help accrue substantial data during the 22-year span since 2000 that would result in a well-documented set of invasive species. We were not disappointed in this regard; 95 species had sufficient data to analyze. We also expected the wide taxonomic breadth of the list to indicate invasive species patterns in general. Species richness and cumulative records (2000-2021) were each patchily distributed around the world, and within those distributions were functions of spatial autocorrelation, climate, and socioeconomic or governmental terms. We note that results here are based on data aggregated to 0.5° grids, which may afford more clear signals than country-level analyses.

Discussion is organized below as cautionary "lessons learned" about existing data to make it clear that results here evaluated patterns in available information available via multiple publicly-available data bases (including but not limited to GBIF), rather than processes that may cause those ultimate patterns. Some lessons can be described as *streetlight effects*, which apply to science in general (Freedman 2010) and derive from an old joke about someone searching under a streetlight at night for a lost wallet, regardless of where the wallet was actually dropped. For invasive species, a streetlight effect occurs when we seek to generalize diversity patterns, regardless of where data actually exist. In total, five lessons were learned from data and analyses that suggest directions for future invasive species research and work (Table 4).

Lesson #1: The world's worst invasive species vary substantially in available data.

An apparent *phylogenetic streetlight effect* exists in publicly-available online data bases for invasive species research is that some taxa are most often considered as invasive. This bias is consistent with the general Linnean shortfall in biology (Hortal et al. 2015), where larger-bodied organisms and anthropocentric diseases are most likely to be formally described and gain most attention (Diniz-Filho et al. 2010, Vicente 2010, Murray et al. 2015, Bartels et al. 2016, Parsons et al. 2022).

For example, two species omitted here due to lack of data are major disease agents in U.S forests; chestnut blight (*Cryphonectria parasitica* (Murill)) and Dutch elm disease (*Ophiostoma ulmi* (Buisman)). Data are remarkably sparse for the fungi causing those well-known tree diseases. If data here represent invasive species in general, then more occurrence data for currently under-represented invasive species need to be collected and reported to gain meaningful understanding of distributions. This need seems especially urgent for relatively small-bodied organisms (including disease agents) which can cause substantial effects themselves and compound climate change effects (Hendrix et al. 2008, Singh et al. 2013, Hoberg and Brooks 2015).
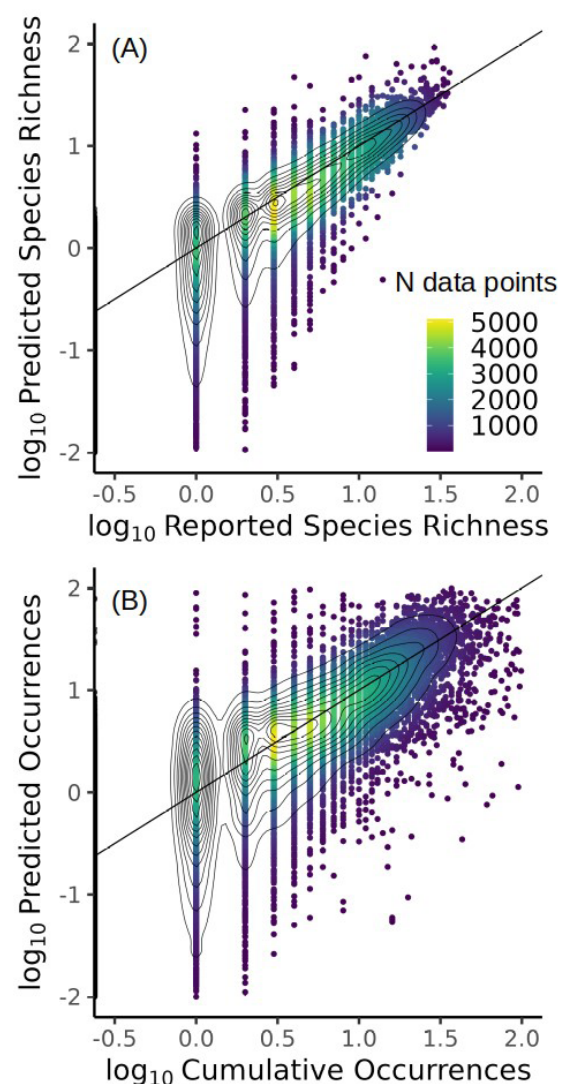
Fig. 4. Model validation for (A) species richness and (B) cumulative occurrences of the "world's worst invasive species." Species richness was conservatively predicted in most cases, whereas cumulative occurrences were not; thus we discounted detailed results of cumulative occurrences (e.g., maps) in interpretations. The line represents the 1:1 relationship in each plot.

Table 3. Summary of statistical model results for cumulative records (2000-2021) of 95 of the world's worst invasive species. The model was clearly most efficient (AIC weight = 1.000, all δAIC >2) among alternative models. It includes a conditional model for quantitative records after taking into account the binomial zero-inflation model, which predicts absence (0) of invasive species., where an odds ratio >1 indicates the increase in the odds of 0 with an increase of 1 SD in a scaled predictor. Marginal R^2 (fixed effects) = 0.40; conditional R^2 (fixed + random effects) = 0.62. All else as in Table 2.

Predictors	Coefficient	P value
Conditional Model		
Random Effects: Standard Deviation		
anthromes	0.102	- -
countries	0.594	- -
Fixed Effects: Scaled Effect Size (95% CI)		
Intercept	0.186 (0.155)	0.018
geography	0.393 (0.018)	<0.001
annual mean temperature	0.387 (0.136)	<0.001
(annual mean temperature)^2	-0.241 (0.088)	<0.001
annual mean precipitation	-0.05 (0.246)	0.708
(annual mean precipitation)^2	0.081 (0.213)	0.456
Voice & Accountability	0.163 (0.208)	0.122
Governmental Effectiveness (GE)	0.317 (0.205)	0.002
human population	-0.719 (0.322)	<0.001
(human population)^2	1.109 (0.212)	<0.001
GDP_{pc}	-0.029 (0.095)	0.550
$(GDP_{pc})^2$	-0.021 (0.091)	0.645
GE: human population	0.648 (0.223)	<0.001
GE: (human population)^2	-0.537 (0.149)	<0.001
Zero-inflation Model (predicts zeroes)		
Random Effects: Standard Deviation		
anthromes	0.322	- -
countries	1.326	- -
Fixed Effects: Scaled effect sizes as odds ratios (upper & lower CIs)		
Intercept	1.937 (1.414, 2.652)	<0.001
geography	0.098 (0.086, 0.111)	<0.001
Governmental Effectiveness	0.319 (0.204, 0.499)	<0.001
Voice & Accountability	0.518 (0.329, 0.816)	0.005
human population	0.163 (0.150, 0.176)	<0.001

Table 4. Five lessons learned from examination of "the world's worst invasive species" in publicly-available data bases, including the basis for each lesson.

|	Lesson	Basis
1	The world's worst invasive species vary substantially in available data; phylogenetic and terrestrial streetlight effects	Comparison of data available among species
2	Evidence to understand (and act on) invasive species is biased to parts of the world that appear most able to report biodiversity; a geographic streetlight effect	Maps and statistical analyses: countries and gross domestic product per capita (GDP_{pc}) predicted invasive species richness and occurrences
3	More invasive species are reported by more effective and open governments; a governmental streetlight effect	Statistical analyses, where indices for effective and open governments strongly predicted invasive species.
4	Invasive species richness follows general climatic clines, consistent with environmental tolerance limits but weighted to temperate regions	Statistical analyses, where ∩-shaped patterns were detected (i.e., reduced values at extremes and peak values at intermediate values).
5	Invasive species richness and occurrences depend on human population	Statistical analyses. This may reflect multiple drivers, including observer effects and introduction frequency (landscape conditions were addressed by anthromes).

A *terrestrial streetlight effect* also appears in the data, because few marine or coastal species were listed by Lowe et al. (2000) and analyzed here. This effect may reflect, in part, a Wallacean Shortfall (Hortal et al. 2015), where bipedal air-breathers more often report terrestrial occurrences to databases than mid-ocean observations. Perhaps the list should be updated and expanded to better represent macro- and micro-organisms and major habitats (e.g., the 100 worst invasive species in each of terrestrial, freshwater, coastal, or marine habitats). After all, absence from the original list did "not imply that a species poses a lesser threat" (Lowe et al. 2000).

A more comprehensive list of the world's worst invasive species may help redirect efforts on those understudied species and invaded habitats in the same way that species on the original list were widely cited.

Lesson #2: Most evidence to understand (and act on) invasive species is biased to parts of the world that appear most able to document biodiversity.

This lesson may seem obvious given similar findings exist for other biodiversity analyses (e.g., Meyer et al. 2015, Nori et al. 2023). To our knowledge, it has not been reported globally among many, invasive species; perhaps related to the common focus on single species .
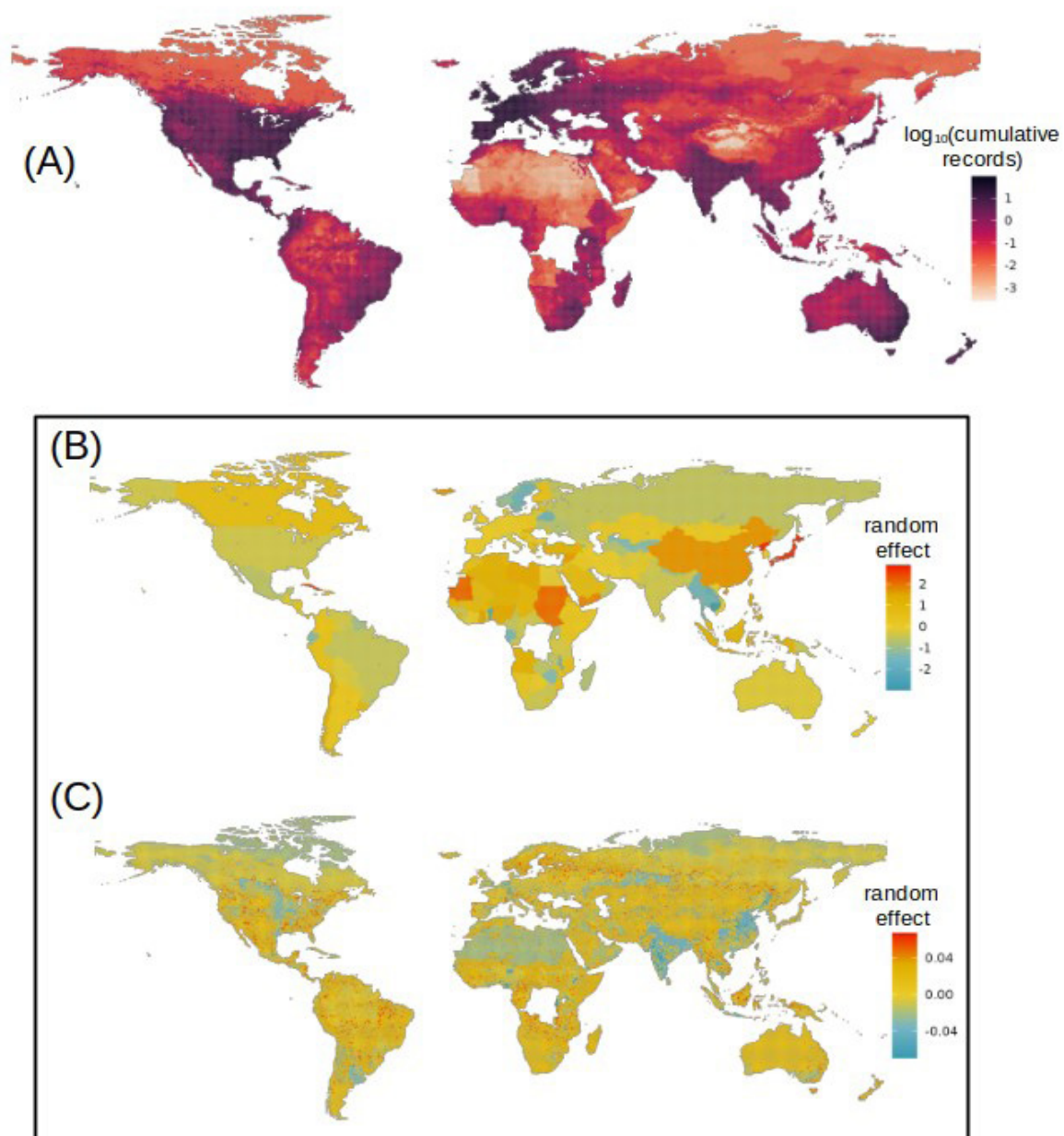
Fig 5. (A) Predicted cumulative records for 95 of the 100 world's worst invasive species (Lowe et al. 2000), 2000-2021. Predictions come from the model detailed in Table 3. Compare with mapped data in Fig. 1B. (B) Countries and (C) anthromes varied as random effects in the mixed effects model.

Most data in internet-available data bases for invasive species occur in North America and Europe but many countries elsewhere lack reports of the 95 invasive species and data are not distributed evenly or randomly across the globe. This spatial bias – despite the diversity and number of species analyzed – supported the hypothesis for geographic effects and random effects (countries and anthromes) to explain patterns and represents a *geographic streetlight effect* because we can better evaluate patterns in North America and Europe than in other continents with less data (e.g., large parts of Africa, Asia, and South America). This outcome also reflects a different piece of the Wallacean shortfall (Hortal et al. 2015), probably due to uneven distributions of people, wealth, and governmental systems around the world (also see Lesson #3 below), as reflected in statistical results here. Similar to literature searches (Zenni et al. 2023), greater multi-lingual access to enter data into online biodiversity data bases could expand known spatial distributions. However, data gaps revealed in maps also appear to extend beyond nations to regions, perhaps reflecting historical legacies of colonialism and long-term geopolitical conflicts as much as current national effects.

Based on the above, we regarded three more emerging lessons with a global view but local or regional caution. This approach was warranted because fixed effects in models represented global mean effects revealed amidst the patchy data. Also, model predictors were represented by wide ranges of potential values, and analyses addressed potentially confounding effects of spatial autocorrelation. Models appropriately estimated low species richness and fewer cumulative records where no data existed, but models merely fitted existing (globally patchy) data; the model results are contingent on reported data and are likely to change if more data were more widely available.

Lesson #3: More invasive species are reported by governments that are more effective and have more open information access.

This effect may be considered a *governmental streetlight effect* and contributes to the geographic streetlight effect. We expect this pattern is related to support for science-based policy, science infrastructure and information access (including both governmental data reporting and freedom of expression). For example, governmental support to digitize museum records and other traditional data may be needed before data may be available online (e.g., in GBIF). Also, more effective governments were less often sources of invasive species intercepted in trade entering New Zealand (Brenton-Rule et al. 2016). It appears that a more effective government exports fewer invasives abroad *and* is more transparent about invasive species within its borders. We note that both may exist in an effective government because reducing or removing invasive species may exceed governmental authority (e.g., on private lands), whereas reporting invasive species is a matter of information gathering and access.

This lesson suggests another reason for more effective government is to better combat the substantial ecological and economic costs associated with invasive species (Pimentel et al. 2005, Cuthbert et al. 2022). On the other hand, recent global trends in the mix of democracies and autocracies (Brunkert et al. 2019, Lührmann and Lindberg 2019) do not bode well for invasive species research, interception, and mitigation.

Lesson #4: Invasive species richness is clinal but unlike general expectations.

Diversity is generally expected to peak in warmest regions and decline toward the poles, related to harsh climate and well-known environmental tolerance limits. However, diversity of the "world's worst invasive species" peaked in temperate regions (see Lesson #2 above) and was truncated at temperatures less than those expected in equatorial regions. We note that this pattern was obtained after accounting for potentially-confusing effects of other predictors (e.g., spatial autocorrelation, sampling and anthropocentric effects). This effect appears most related to geographic biases in reported data, where temperate regions tend to have more data. Thus, clinal pattern here differs in detail from typical latitudinal clines and appears an artifact of sample distribution in these data. However, results here remain relevant to consequences of climate change (Hellmann et al. 2008) for the "world's worst invasive species" because they serve as a baseline for potentially shifting future spatial patterns, after accounting for potential changes in publicly-available data. Toward that aim, the modeling approach used here may help future clinal modeling to account for diverse covariates, (including spatial autocorrelation and numerous knowledge gaps) and help predict potential distribution shifts with climate change.

Lesson #5: Invasive species richness and occurrences increase with human population.

We expected a generally positive effect, related to observer effects and/or anthropophilic distributions weighted to points of introduction, which are often population centers (e.g., port cities). The detected upward trend was nonlinear; it was accentuated with greater human population and by an interaction with more effective government. We infer that the relationship is complex among diverse species, but is consistent with observer effects, anthropophilic distributions of invasive species, and better information access via governmental programs to report invasive species. The interaction between population and Governmental Effectiveness shows that more invasive species are reported than expected where there are fewer people if a government is more effective. We note that land use categories (that correlate with human population) were accounted for separately by anthromes as random effects only revealed after accounting for other effects, and likely a proxy for underlying processes or other factors, such as land cover effects (e.g., wild lands have relatively few observations but urban centers offer limited habitat diversity).

We suggest future research on invasive species should more often include anthropocentric predictors (e.g., human population, governmental indices) and land cover (e.g. anthromes) to better model invasive species distributions while also accounting for other covariates (e.g., climate). This approach does not appear to be common, and differs from analyses of native species in natural lands.

Conclusions

Overall, analyses here for 95 of the 100 world's worst invasive species (Lowe et al. 2000, Luque et al. 2014) transcended common views per species or within only non-native ranges to attempt a general biogeography for this well-known list of species. The species analyzed here represent a quite varied and robustly sampled subset that may enable inferences about invasive species biogeography in general. To the extent that is true, then conclusions here evaluate the general state of invasive biogeography by revealing five lessons (Table 4), including four related streetlight effects about general patterns in data for invasive species distributions and diversity.

We think the most striking are Lessons 1-3, which in sum indicate that large gaps exist in global distributions (i.e., native and non-native ranges) for even the "world's worst invasive species," despite millions of records in global data bases. Those gaps can be filled by concerted, global efforts to improve the breadth of biodiversity records and address the many burdens placed on natural and socioeconomic systems by invasive species (e.g., Pimentel et al. 2005, Ehrenfeld 2010, Paini et al. 2016, Loss et al. 2016, Cuthbert et al. 2022, Ryan et al. 2021). Until then, models such as ours use known distributions to predict invasive species distributions after addressing spatial autocorrelation, climatic conditions, and anthropocentric measures (population, socioeconomic, and governmental factors). Our model of species richness seemed particularly (but cautiously) predictive for global lands.

Global data gaps compound the challenge to estimate potential ranges of invasive species. Some species distribution models (SDMs) use pseudo-absences with corrections for sampling biases (Hertzog et al. 2014), but we think two problems emerge here. Foremost, all SDMs assume range equilibrium, which is a fundamental problem for invasive species not yet at equilibrium (Václavík and Meentemeyer 2009). Beyond customary climate and habitat variables, results here suggest that anthropocentric variables, including Government Effectiveness, Voice & Accountability and human population, may improve global invasive SDMs (e.g., Lyons et al. 2020). If so, then models may be more predictive before these problematic species reach "equilibrium," which is too late for any effective intervention. Also, data gaps shown here represent large-scale absences of data, not simpler quantitative gradients of sampling biases. Perhaps smoothed predictions similar to ours and based on spatial autocorrelation may help approach quantitative gradients of sampling biases needed in SDMs.

Data Accessibility Statement

Data and R code are available in Dryad & Zenodo (respectively) at https://doi.org/10.5061/dryad.zw3r228bh.

References

Bahn V.J., O'Connor, R.B. & Krohn, W. (2006) Importance of spatial autocorrelation in modeling bird distributions at a continental scale. Ecography, 29, 835-844. https://doi.org/10.1111/j.2006.0906-7590.04621.x

Bartels, P.J., Apodaca, J.J., Mora, C. & Nelson, D.R. (2016) A global biodiversity estimate of a poorly known taxon: phylum Tardigrada. Zoological Journal of the Linnean Society, 178, 730-736. https://doi.org/10.1111/zoj.12441

Bennett, A.E., Thomsen, M. & Strauss, S.Y. (2011) Multiple mechanisms enable invasive species to suppress native species. American Journal of Botany, 98, 1086-1094. https://doi.org/10.3732/ajb.1000177

Bolker, B., & R Core Team. (2022) bbmle: Tools for general maximum likelihood estimation. R package version 1.0.25. R Foundation for Statistical Computing. Available at https://CRAN.R-project.org/package=bbmle.

Brenton-Rule, E.C., Barbieri, R.F. & Lester PJ. (2016) Corruption, development and governance indicators predict invasive species risk from trade. Proceedings of the Royal Society B: Biological Sciences, 283, 20160901. https://doi.org/10.1098/rspb.2016.0901

Brooks, M.E., Kristensen, K., Van Benthem, K.J., et al. (2017) glmmTMB balances speed and flexibility among packages for zero-inflated generalized linear mixed modeling. The R Journal, 9, 378-400.

Brunkert, L., Kruse, S. & Welzel, C. (2019) A tale of culture-bound regime evolution: the centennial democratic trend and its recent reversal. Democratization, 3, 422-443. https://doi.org/10.1080/13510347.2018.1542430

Burnham, K. & Anderson, D. (2002) Model selection and multimodel inference. A practical information-theoretic approach. 2nd ed. Springer-Verlag, New York.

Chamberlain, S. (2021) spocc: interface to species occurrence data sources. R package version 1.2.0. https://CRAN.R-project.org/package=spocc.

Colautti, R.I. & MacIsaac, H.J. (2004) A neutral terminology to define 'invasive' species. Diversity and Distributions, 10, 135-141. https://doi.org/10.1111/j.1366-9516.2004.00061.x

Constable, H., Guralnick, R., Wieczorek, J., et al. (2010) VertNet: a new model for biodiversity data sharing. PLOS Biology, 8, e1000309. https://doi.org/10.1371/journal.pbio.1000309

Courchamp, F., Fournier, A., Bellard, C., et al. (2017) Invasion biology: specific problems and possible solutions. Trends in Ecology and Evolution, 32, 13-22. https://doi.org/10.1016/j.tree.2016.11.001

Crystal-Ornelas, R. & Lockwood, J.L. (2020) The 'known unknowns' of invasive species impact measurement. Biological Invasions, 22, 1513-1525.

Cuthbert, R.N., Diagne, C., Haubrock, P.J., Turbelin, A.J. & Courchamp, F. (2022) Are the "100 of the world's worst" invasive species also the costliest? Biological Invasions, 24, 1895-1904. https://doi.org/10.1007/s10530-020-02200-0

Diniz-Filho, J.A.F., De Marco, Jr., P.A. & Hawkins, B.A. (2010) Defying the curse of ignorance: perspectives in insect macroecology and conservation biogeography. Insect Conservation and Diversity, 3, 172-179. https://doi.org/10.1111/j.1752-4598.2010.00091.x

Dormann, C.F., McPherson, J.M., Araújo, M.B., et al. (2007) Methods to account for spatial autocorrelation in the analysis of species distributional data: a review. Ecography, 1, 609-628. https://doi.org/10.1111/j.2007.0906-7590.05171.x

Dushoff, J., Kain, M.P. & Bolker, B.M. (2019) I can see clearly now: reinterpreting statistical significance. Methods in Ecology and Evolution, 10, 756-759. https://doi.org/10.1111/2041-210X.13159

Ellis, E.C. & Ramankutty, N. (2013) Anthropogenic biomes of the World, Version 2, 2000. Palisades, NY: NASA Socioeconomic Data and Applications Center (SEDAC). Available at http://sedac.ciesin.columbia.edu/data/set/anthromes-anthropogenic-biomes-world-v2-2000.

Ehrenfeld, J.G. (2010) Ecosystem consequences of biological invasions. Annual Review of Ecology, Evolution, and Systematics, 41, 59-80. https://doi.org/10.1146/annurev-ecolsys-102209-144650

Falk-Petersen, J., Bøhn, T. & Sandlund, O.T. (2006) On the numerous concepts in invasion biology. Biological invasions, 8, 1409-1424. https://doi.org/10.1007/s10530-005-0710-6

Fang, Y. (2011) Asymptotic equivalence between cross-validations and Akaike information criteria in mixed-effects models. Journal of Data Science, 9, 15-21. https://doi.org/10.6339/JDS.201101_09(1).0002

Ferrier, S., Watson, G., Pearce, J. & Drielsma, M. (2002) Extended statistical approaches to modelling spatial pattern in biodiversity in northeast New South Wales. I. Species-level modelling. Biodiversity & Conservation, 11, 2275-307. https://doi.org/10.1023/A:1021302930424

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, 4302-15. https://doi.org/10.1002/joc.5086 Available at https://www.worldclim.org/data/bioclim.html.

Freedman, D.H. (2010) Why scientific studies are so often wrong: The streetlight effect. Discover Magazine, 26, 1-4.

Fridley, J.D., Stachowicz, J.J., Naeem, S., et al. (2007) The invasion paradox: reconciling pattern and process in species invasions. Ecology, 88, 3-17. https://doi.org/10.1890/0012-9658(2007)88[3:TIPRPA]2.0.CO;2

Grassle, J.F. & Stocks, K.I. (1999) A global ocean biogeographic information system (OBIS) for the census of marine life. Oceanography, 12, 12-14. https://www.jstor.org/stable/43924894

Guo, K., Pyšek, P., Chytrý, M., et al. (2022) Ruderals naturalize, competitors invade: varying roles of plant adaptive strategies along the invasion continuum. Functional Ecology, 36, 2469-2479. https://doi.org/10.1111/1365-2435.14145

Hasselman, D.J., Hinrichsen, R.A., Shields, B.A. & Ebbesmeyer, C.C. (2012) American shad of the Pacific Coast: a harmful invasive species or benign introduction? Fisheries, 37, 115-22. https://doi.org/10.1080/03632415.2012.659941

Hellmann, J.J., Byers, J.E., Bierwagen, B.G. & Dukes, J.S. (2008) Five potential consequences of climate change for invasive species. Conservation Biology, 22, 534-543. https://doi.org/10.1111/j.1523-1739.2008.00951.x

Helsel, D.R. & Ryker, S.J. (2002) Defining surfaces for skewed, highly variable data. Environmetrics: The Official Journal of the International Environmetrics Society, 13, 445-452. https://doi.org/10.1002/env.531

Hendrix, P.F., Callaham, Jr., M.A., Drake, J.M., et al. (2008) Pandora's box contained bait: the global problem of introduced earthworms. Annual Review of Ecology, Evolution, and Systematics, 39, 593-613. https://doi.org/10.1146/annurev.ecolsys.39.110707.173426

Hertzog, L.R., Besnard, A. & Jay-Robert, P. (2014) Field validation shows bias-corrected pseudo-absence selection is the best method for predictive species-distribution modelling. Diversity and Distributions, 20, 1403-1413. https://doi.org/10.1111/ddi.12249

Hoberg, E.P. & Brooks, D.R. (2015) Evolution in action: climate change, biodiversity dynamics and emerging infectious disease. Philosophical Transactions of the Royal Society B: Biological Sciences, 370, 20130553. https://doi.org/10.1098/rstb.2013.0553

Hortal, J., de Bello, F., Diniz-Filho, J.A., et al. (2015) Seven shortfalls that beset large-scale knowledge of biodiversity. Annual Review of Ecology, Evolution, and Systematics, 4, 46:523-549. https://doi.org/10.1146/annurev-ecolsys-112414-054400

Kaufmann, D., Kraay, A. & Mastruzzi, M. (2011) The worldwide governance indicators: methodology and analytical issues 1. Hague Journal on the Rule of Law, 3, 220-246. https://doi.org/10.1017/S1876404511200046 Available at https://info.worldbank.org/governance/wgi/

Kolar, C.S. & Lodge, D.M. (2001) Progress in invasion biology: predicting invaders. Trends in Ecology and Evolution, 16, 199-204. https://doi.org/10.1016/S0169-5347(01)02101-2

Kourantidou, M., Haubrock, P.J., Cuthbert, R.N., et al. (2022) Invasive alien species as simultaneous benefits and burdens: trends, stakeholder perceptions and management. Biological Invasions, 24, 1905-1926. https://doi.org/10.1007/s10530-021-02727-w

Kummu, M., Taka, M. & Guillaume, J.H. (2018) Gridded global datasets for gross domestic product and Human Development Index over 1990–2015. Scientific Data, 5, 1-5. https://doi.org/10.1038/sdata.2018.4

Loss, S.R., Will, T. & Marra, P.P. (2013) The impact of free-ranging domestic cats on wildlife of the United States. Nature Communications, 4, 1-8. https://doi.org/10.1038/ncomms2380

Lowe, S., Browne, M., Boudjelas, S. & De Poorter, M. (2000) 100 of the world's worst invasive alien species: a selection from the global invasive species database. Auckland: Invasive Species Specialist Group. www.issg.org/booklet.pdf

Lüdecke, D., Ben-Shachar, M.S., Patil, I., Waggoner, P. & Makowski, D. (2021) performance: an R package for assessment, comparison and testing of statistical models. Journal of Open Source Software, 6, 3139. https://doi.org/10.21105/joss.03139

Lührmann, A. & Lindberg, S.I. (2019) A third wave of autocratization is here: what is new about it? Democratization, 26, 1095-1113. https://doi.org/10.1080/13510347.2019.1582029

Luque, G.M., Bellard, C., Bertelsmeier, C., et al. (2014) The 100th of the world's worst invasive alien species. Biological Invasions, 16, 981-985. https://doi.org/10.1007/s10530-013-0561-5

Lyons, D.A., Lowen, J.B., Therriault, T.W., et al. (2020) Identifying marine invasion hotspots using stacked species distribution models. Biological Invasions, 22, 3403-23. https://doi.org/10.1007/s10530-020-02332-3

Matsunaga, A., Thompson, A., Figueiredo, R.J., et al. (2013) A computational and storage cloud for integration of biodiversity collections. IEEE 9th International Conference on e-Science, 78-87. https://doi.org/10.1109/eScience.2013.48.

Meyer, C., Kreft, H., Guralnick, R. & Jetz, W. (2015) Global priorities for an effective information basis of biodiversity distributions. Nature Communications 6, 8221.

Mooney, H.A. & Cleland, E.E. (2001) The evolutionary impact of invasive species. Proceedings of the National Academy of Sciences USA, 98, 5446-51. https://doi.org/10.1038/ncomms9221

Morens, D.M., Holmes, E.C., Davis, A.S. & Taubenberger, J.K. (2011) Global rinderpest eradication: lessons learned and why humans should celebrate too. Journal of Infectious Diseases, 204, 502-505. https://doi.org/10.1093/infdis/jir327

Murray, K.A., Preston, N., Allen, T., et al. (2015) Global biogeography of human infectious diseases. Proceedings of the National Academy of Sciences USA, 112, 12746-12751. https://doi.org/10.1073/pnas.1507442112

Nori, J., Cordier, J. M., Osorio-Olivera, L., & Hortal, J. (2023). Global knowledge gaps of herptile responses to land transformation. Frontiers in Ecology and the Environment, 21, 411-417. https://doi.org/10.1002/fee.2625

Paini, D.R., Sheppard, A.W., Cook, D.C., et al. (2016) Global threat to agriculture from invasive species. Proceedings of the National Academy of Sciences USA, 113, 7575-7579. https://doi.org/10.1073/pnas.1602205113

Parham, J.F., Papenfuss, T.J., Sellas, A.B., Stuart, B.L. & Simison, W.B. (2020) Genetic variation and admixture of red-eared sliders (*Trachemys scripta elegans*) in the USA. Molecular Phylogenetics and Evolution. 145, 106722. https://doi.org/10.1016/j.ympev.2019.106722

Parsons, D.J., Pelletier, T.A., Wieringa, J.G., Duckett, D.J. & Carstens, B.C. (2022) Analysis of biodiversity data suggests that mammal species are hidden in predictable places. Proceedings of the National Academy of Sciences USA, 119:e2103400119. https://doi.org/10.1073/pnas.2103400119

Pereyra, P.J. (2016) Revisiting the use of the invasive species concept: an empirical approach. Austral Ecology, 41, 519-528. https://doi.org/10.1111/aec.12340

Pereyra PJ. (2020) Rethinking the native range concept. Conservation Biology, 34, 373-377. https://doi.org/10.1111/cobi.13406

Pimentel, D., Zuniga, R. & Morrison, D. (2005) Update on the environmental and economic costs associated with alien-invasive species in the United States. Ecological Economics, 52, 273-288. https://doi.org/10.1016/j.ecolecon.2004.10.002

Pouteau, R., Thuiller, W., Hobohm, C., et al. (2021) Climate and socio-economic factors explain differences between observed and expected naturalization patterns of European plants around the world. Global Ecology and Biogeography, 30, 1514-1531. https://doi.org/10.1111/geb.13316

Ricciardi, A. & Ryan, R. (2018) The exponential growth of invasive species denialism. Biological Invasions, 20, 549-553. https://doi.org/10.1007/s10530-017-1561-7

Richardson, D.M. & Ricciardi, A. (2013) Misleading criticisms of invasion science: a field guide. Diversity and Distributions, 19, 1461-1467. https://doi.org/10.1111/ddi.12150

Ryan, S.J., Carlson, C.J., Tesla B., et al. (2021) Warming temperatures could expose more than 1.3 billion new people to Zika virus risk by 2050. Global Change Biology, 27, 84-93. https://doi.org/10.1111/gcb.15384

Schneider, K., Makowski, D., & van Der Werf, W. (2021). Predicting hotspots for invasive species introduction in Europe. Environmental Research Letters, 16, 114026. https://doi.org/10.1088/1748-9326/ac2f19

Singh, S.K., Hodda, M., Ash, G.J. & Banks, N.C. (2013) Plant-parasitic nematodes as invasive species: characteristics, uncertainty and biosecurity implications. Annals of Applied Biology, 163, 323-50. https://doi.org/10.1111/aab.12065

Sullivan, B.L., Wood, C.L., Iliff, M.J., et al. (2009) eBird: a citizen-based bird observation network in the biological sciences. Biological Conservation, 142, 2282-2292. https://doi.org/10.1016/j.biocon.2009.05.006

Václavík, T., & Meentemeyer, R. K. (2009). Invasive species distribution modeling (iSDM): are absence data and dispersal constraints needed to predict actual distributions? Ecological Modelling, 220, 3248-3258. https://doi.org/10.1016/j.ecolmodel.2009.08.013

van Kleunen, M., Dawson, W., Essl, F., et al. (2015) Global exchange and accumulation of non-native plants. Nature, 525, 100-103. https://doi.org/10.1038/nature14910

Vicente, F. (2010) Micro-invertebrates conservation: forgotten biodiversity. Biodiversity and Conservation, 19, 3629-34. https://doi.org/10.1007/s10531-010-9898-6

Westphal, M.I., Browne, M., MacKinnon, K. & Noble, I. (2008) The link between international trade and the global distribution of invasive alien species. Biological Invasions, 10, 391-8. https://doi.org/10.1007/s10530-007-9138-5

Zenni, R.D., Barlow, J., Pettorelli, N., Stephens, P., Rader, R., Siqueira, T., Gordon, R., Pinfield, T. & Nuñez, M.A. (2023) Multi-lingual literature searches are needed to unveil global knowledge. Journal of Applied Ecology 60, 380-383. https://doi.org/10.1111%2F1365-2664.14370

Zizka, A., Silvestro, D., Andermann, T., et al. (2019) CoordinateCleaner: standardized cleaning of occurrence records from biological collection databases. Methods in Ecology and Evolution. 10, 744-51. https://doi.org/10.1111/2041-210X.13152