PNAS RESEARCH ARTICLE | ECOLOGY

The worldwide networks of spread of recorded alien species

César Capinha[a,b,1], Franz Essl[c], Miguel Porto[d,e,f], and Hanno Seebens[g]

Edited by Daniel Simberloff, University of Tennessee at Knoxville, Knoxville, TN; received February 2, 2022; accepted November 15, 2022

Our ability to predict the spread of alien species is largely based on knowledge of previous invasion dynamics of individual species. However, in view of the large and growing number of alien species, understanding universal spread patterns common among taxa but specific to regions would considerably improve our ability to predict future dynamics of biological invasions. Here, using a comprehensive dataset of years of first record of alien species for four major biological groups (birds, nonmarine fishes, insects, and vascular plants), we applied a network approach to uncover frequent sequential patterns of first recordings of alien species across countries worldwide. Our analysis identified a few countries as consistent early recorders of alien species, with many subsequent records reported from countries in close geographic vicinity. These findings indicate that the spread network of alien species consists of two levels, a backbone of main dispersal hubs, driving intercontinental species movement, and subsequent intracontinental radiative spread in their vicinity. Geographical proximity and climatic similarity were significant predictors of same-species recording among countries. International trade was a significant predictor of the relative timing of species recordings, with countries having higher levels of trade flows consistently recording the species earlier. Targeting the countries that have emerged as hubs for the early spread of alien species may have substantial cascading effects on the global spread network of alien species, significantly reducing biological invasions. Furthermore, using these countries as early-warning system of upcoming invasions may also boost national prevention and invasion preparedness efforts.

alien species | globalization | invasion dynamics | network | spread routes

The expansion of transportation networks and the increasing movement of people and goods around the world are causing an unprecedented influx of alien species to regions outside of their native ranges (1–3). Many of these species established populations in these regions, thereby reshaping biogeography (4–6) and causing major negative impacts on biodiversity (7–8), economic activities (9), and human health (10–11). Reducing these impacts requires a comprehensive understanding of universal patterns of spread, which would be of crucial importance to improve our ability to predict the spread of alien species.

The spread of alien species is complex, often resulting from the interaction of human-mediated introductions and species' own dispersal capabilities. Although much effort has been invested to improve our understanding of spreading dynamics, this has been mainly assessed for single or few species (e.g., refs. 12 and 13), often at limited (i.e., national or subnational) extents (e.g., ref. 14) or considering a limited set of regions of the world (e.g., refs. 15 and 16). In addition, the diversity of applied approaches renders a comparison of studies difficult. Thus, it remains unknown if the patterns of alien species spread are consistent across major taxonomic groups at large scales. Given the tight relationship between anthropogenic drivers and alien species introductions, it seems likely that some countries are frequently invaded first and serve as hubs for the spread to adjacent regions. Furthermore, spatiotemporal patterns of alien species spread may differ between taxonomic groups, resulting from differences in natural dispersal abilities or already colonized realms (e.g., fishes versus birds) or in how distinct human activities shape the patterns of species introductions (17–19). Knowing such general patterns of spread opens opportunities for targeted preventive alien species management.

Here, we use the by far most comprehensive database on the year of first recording of alien species in regions worldwide (1) to analyze the cross-taxonomic patterns of interregional spread of alien species worldwide. This database has been analyzed already before in other studies, which, however, mostly focused on temporal developments of alien species accumulations, while here, we use first records to infer spatial dynamics of spread and to identify commonalities in spread dynamics across taxa. We focus on the four most data-rich taxonomic groups (birds, nonmarine fish, insects, and vascular plants) and use sequence mining techniques (20) to identify frequent sequences of first records of alien species across countries of the world. We perform this assessment for the period from

Significance

Identifying whether alien species follow similar spatiotemporal spread patterns as they disperse across the world could help guide transnational cooperation and national efforts for the prevention of biological invasions. Whether such cross-taxonomic patterns of spread exist remained hitherto unknown. Here, we show that these patterns exist and are characterized by the early presence of alien species in a few countries, playing central roles in the global trade network, followed by radiative spread into geographically close countries with similar climates. This supports a hierarchical type of spread, where some countries act as major dispersal hubs and determine global invasion routes. These findings help to identify high-risk countries as hubs of biological invasions.

Author contributions: C.C., F.E., M.P., and H.S. designed research; C.C. performed research; C.C. analyzed data; and C.C., F.E., M.P., and H.S. wrote the paper.

The authors declare no competing interest.

This article is a PNAS Direct Submission.

[1] To whom correspondence may be addressed. Email: cesarcapinha@campus.ul.pt.

This article contains supporting information online at https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2201911120/-/DCSupplemental.

1950 onward, which covers the unprecedented acceleration of alien species spread that lasts until today (1, 21, 22). Subsequently, we identify drivers of the sequential patterns of alien species spread. We focus on two complementary components of this phenomenon: i) the proportion of the same species recorded between countries (= compositional similarity component) and ii) the relative timing of species recording in each country (= temporal component). Concerning the first component, our expectation is that countries which are geographically closer have strong trade relationships and are climatically similar will have more similar alien species assemblages, owing to easier interregional alien species flows driven by human activities and postestablishment spread (15, 23–26). As for the second component, our expectation is that the spread of alien species will occur earlier in countries with higher human population sizes and higher levels of economic activities as a result of elevated magnitudes and earlier onsets of propagule and colonization pressures (27, 28).

Results

Spatiotemporal Patterns of Spread. The analysis is based on first records of 839 alien birds, 366 alien nonmarine fishes, 3,702 alien insects, and 10,479 alien vascular plants in 53, 44, 56, and 53 countries having 10 or more first records for each group, respectively (*SI Appendix*, Fig. S1 and Table S1). We calculated a metric assessing the proportion of species recorded first in one country and next in another from 1950 to the present, for all pairs of countries, which we call the proportion of sequential records or simply "$S_{\overrightarrow{ab}}$" consisting of the number of species recorded first in country *a* and next in country *b* divided by the minimum species number of *a* and *b*. The values of this metric range from 0 to 1, where higher values indicate a higher proportion of species observed first in country *a* and afterward in country *b*. On average, the highest proportion of sequential records of alien species among country pairs is found for nonmarine fishes (average $S_{\overrightarrow{ab}}$ = 0.19 ± 0.16 SD), followed by birds (0.18 ± 0.15), insects (0.13 ± 0.12), and vascular plants (0.11 ± 0.14). That is, alien fishes tend to be more evenly distributed compared with other taxa, while plants have more alien species with narrow ranges.

A network visualization of $S_{\overrightarrow{ab}}$ values shows the relatedness of countries according to the similarity of recorded alien species (Fig. 1). In these networks, countries (nodes) sharing more species are grouped closer together, whereas those with few or no species in common are shown further apart. Directed links of varying thickness represent the individual $S_{\overrightarrow{ab}}$ values. Strong links indicate that species recorded in one country (origin) are frequently recorded later in another (destination). Note that these links may deviate from the actual physical path taken by the species (i.e., they may spread directly from the origin to the destination country or through other countries).

For birds, countries are largely grouped by their geographic location, which is indicated by countries with the same symbols clustering together in Fig. 1. This means that, in general, alien birds tend to spread over short distances to adjacent countries. However, a few countries having more alien birds in common with other countries (i.e., having a higher total strength in the network) stick out (Fig. 1*A* and *SI Appendix*, Fig. S2*A*), namely the United States (USA), Spain (ESP), Portugal (PRT), the United Kingdom (GBR), the Dominican Republic (DOM), Germany (DEU), and France (FRA), which are mostly located at the center of the network, detached from the continent-level groupings (Fig. 1*A* and *SI Appendix*, Fig. S3). These countries can be considered as being global hubs of alien bird spread, sharing many species records among each other, and forming a backbone network of strongly connected countries, which are tightly connected to the respective continental subnetworks, which were mostly determined by geographical proximity.

The analysis of sequential recordings provided additional insights into the dynamics of spread. By comparing the strength of outflowing link weights ("outstrength") and inflowing link weights ("instrength") of each country, we could identify countries recording new alien species earlier or later than other countries (Fig. 2). Countries found on the left side of the main diagonal line in Fig. 2 have a higher outstrength and record the species earlier, and those below the diagonal have a higher instrength and report them later in general. This analysis revealed that many species were recorded first in the USA and subsequently somewhere else (Fig. 2*A*). In addition, alien birds were frequently recorded first in Germany, New Zealand (NZL), the United Kingdom, and Australia (AUS). Conversely, Spain, Portugal, Cuba, and the Dominican Republic are identified as "sink" countries, i.e., species are generally recorded there later than in most other countries (Fig. 2*A*).

Similar to alien birds, the network of sequential recordings for nonmarine fishes shows two levels of clustering of countries: one largely according to their geographic and continental location and another at the center of the network formed by countries from distinct continents and forming the backbone network (Figs. 1*B* and 2*B*). Only Asian countries are spread throughout the network. The clustering according to continental regions again indicates the sharing of high proportions of same-species records among adjacent countries. A few major hubs form the central network of global fish spread including countries mainly located in East and Southeast Asia and North and South America such as Haiti (HTI), Vietnam (VNM), Japan (JPN), Philippines (PHL), and China (CHN) (*SI Appendix*, Fig. S2*B*). A few countries show a marked trend for recording species earlier, namely Japan, the USA, Philippines, and Thailand (THA), whereas countries where alien fishes are consistently recorded later include Haiti, Cambodia (KHM), and Tunisia (TUN) (Fig. 2*B*).

For insects, the network of sequential recordings is densely connected among European countries, and links are sparser to countries of other continents and between countries of other continents (Fig. 1*C*). In general, the arrangement of countries reflects their geographical proximity, which is particularly apparent for European countries (e.g., most Scandinavian, eastern European, and Mediterranean countries are shown close to each other). This indicates higher proportions of shared species records among neighbor countries. Italy (ITA) and France (FRA) emerge as central hubs, sharing many species records with several other European countries (*SI Appendix*, Fig. S2*C*) while also being the countries where alien insects are generally recorded earlier (Fig. 2*C*). Some countries are also consistently identified as sinks, namely Albania (ALB) and Ukraine (UKR) (Fig. 2*C*); however, the total strength of these countries in the network is lower than that for Italy and France (*SI Appendix*, Fig. S2*C*), indicating a lower proportion of species shared with other countries. Outside Europe, only a few strong links emerge, for example, the USA followed by Canada (CAN), China (CHN) followed by Taiwan (TWN), and Chile followed by Argentina, which again highlights the geographic structure of the network (Fig. 1*C*).

Finally, the network of sequential recordings for vascular plants is densely connected among European countries, but less so with countries on other continents and within those (Fig. 1*D*), similar to what has been observed for insects. As for previous groups, the highest sharing of species takes place between geographically close

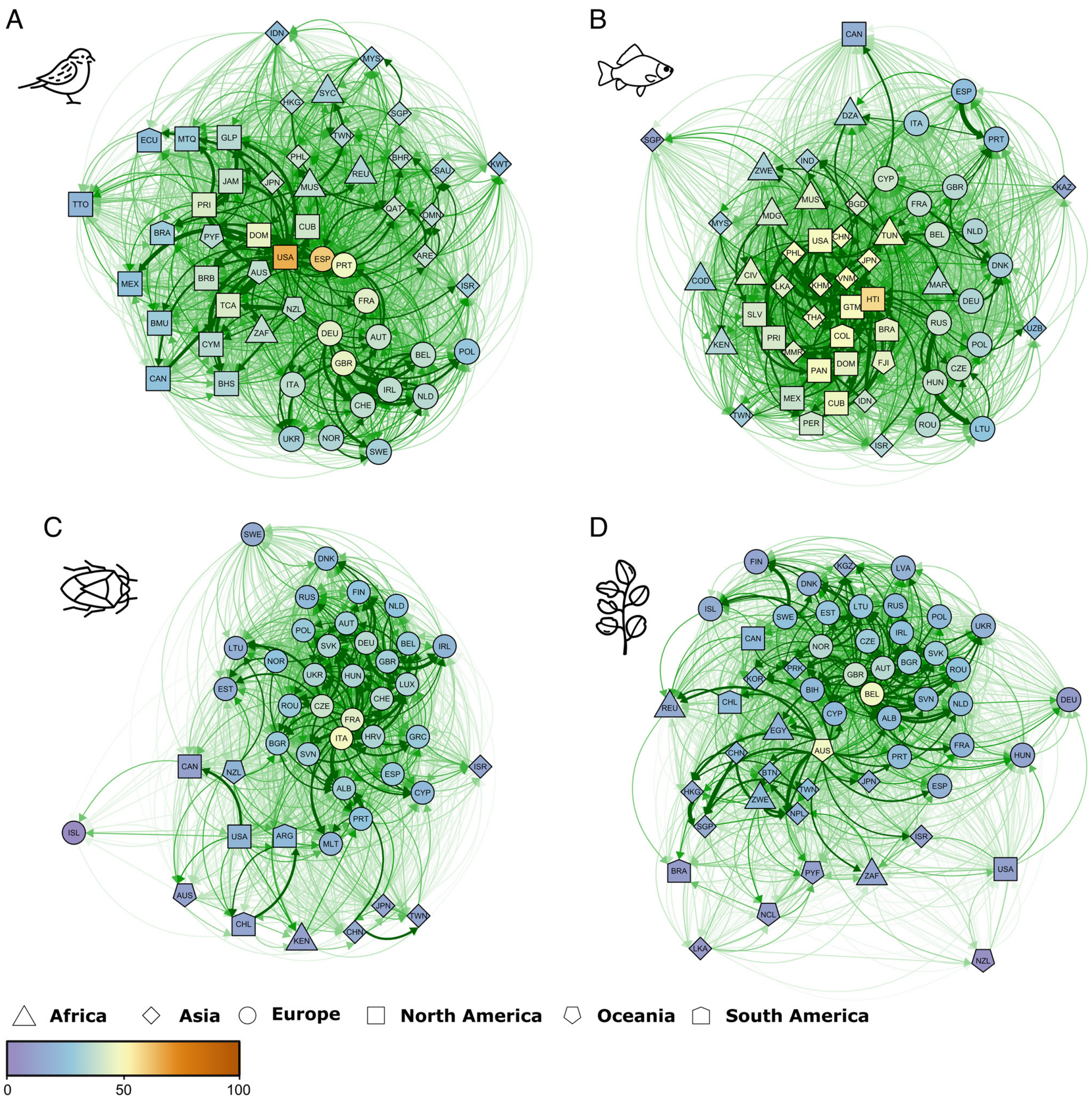

Fig. 1. Spread networks of alien (*A*) birds, (*B*) nonmarine fishes, (*C*) insects, and (*D*) vascular plants across countries of the world. Countries are represented as nodes, and the proportion of sequential records is given as directed weighted links. A thick link flowing from one country to another means that a high proportion of species recorded in the first country was later recorded in the second country regardless of the actual physical path taken by the species (e.g., directly between the two countries or through other countries). Countries with higher similarity in the identity of species recorded are grouped closer together. Node colors correspond to the sum of their link weights and correspond to node "strength" in the network, which indicates the proportion of species shared with other countries. A color at the right edge of the gradient indicates that a country has its species in common with many other countries (i.e., the maximum possible strength), while a color at the left edge indicates that the country has no species in common with other countries. Abbreviations of country names correspond to the ISO 3166-1 alpha-3 codes.

countries, leading to country groupings largely reminiscent of continents. Countries with higher proportions of alien plant species in common with other countries are Belgium (BEL), Australia, and the United Kingdom (*SI Appendix*, Fig. S2*D*), which are also prominent early recorders of the species (Fig. 2*D*). However, while the abovementioned European countries have in common many early records of the same species with other European countries, alien plant species recorded in Australia have mostly been recorded later in countries on other continents, predominantly in Asia and Africa, including Singapore (SGP), Hong Kong (HKG), Egypt (EGY), South Africa (ZAF), or Reunion (REU) (Fig. 1*D*).

Sensitivity of Spread Patterns to Number of First Records per Country, Geographical Representativity, and Surveillance Capacity. These analyses are based on a reference set of well-sampled countries, each having a minimum of 10 first records for each taxonomic group (*SI Appendix*, Table S1). To assess the impact of using other thresholds of minimum number of records per country, we calculated $S_{\overrightarrow{ab}}$ values along a gradient of thresholds, ranging from 1 to 20. We found that the values of total strength, outstrength, and instrength calculated with different thresholds were strongly positively correlated with those obtained for the reference set of countries for all taxonomic

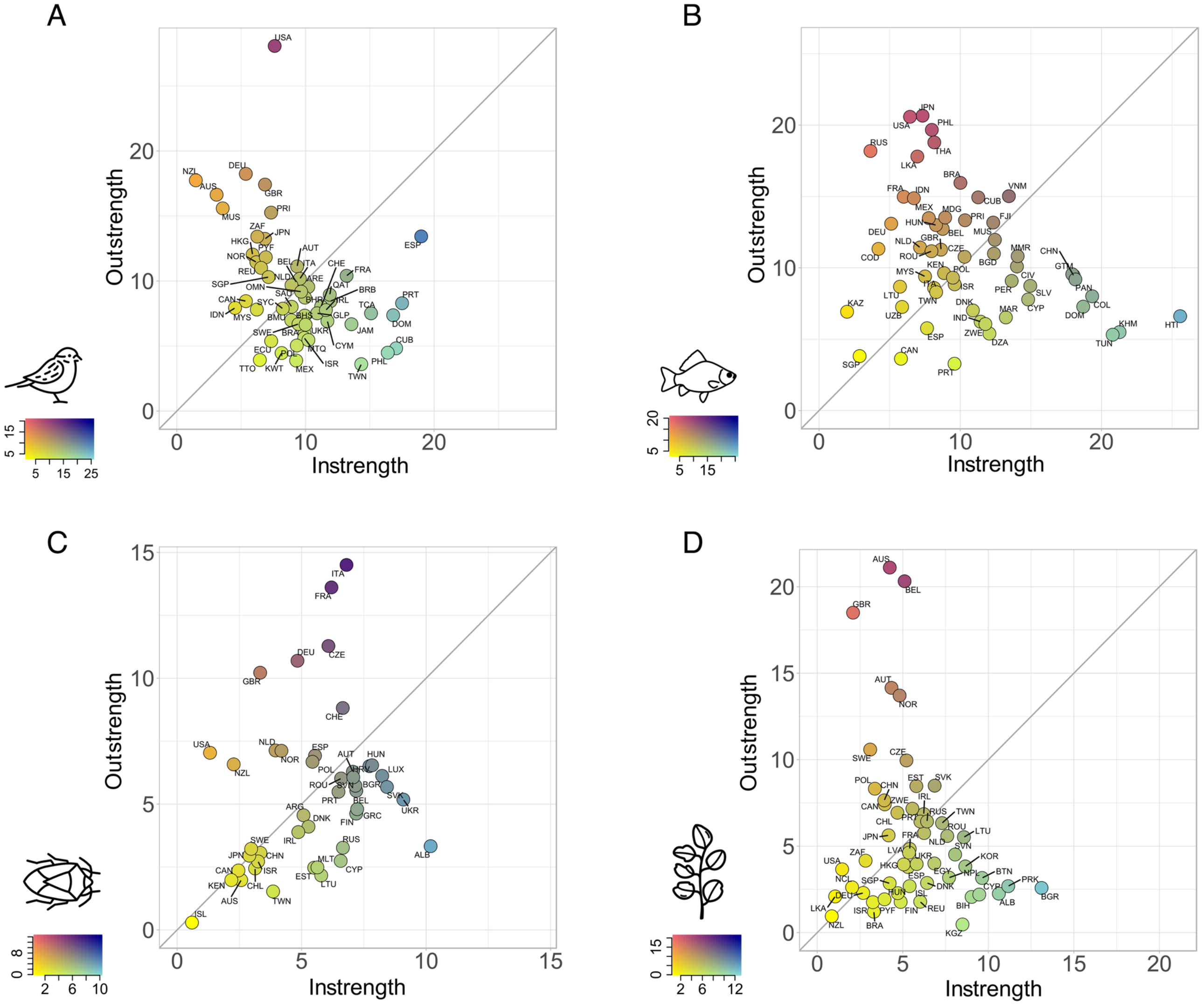

Fig. 2. Relative timing of species recording for alien (*A*) birds, (*B*) nonmarine fishes, (*C*) insects, and (*D*) vascular plants. Scatterplots represent the relationship between the sum of countries' outflowing link weights (outstrength), which indicates earlier recordings, and sum of inflowing link weights (instrength), denoting later recordings. Countries with outstrength exceeding instrength are found on the left side of the main diagonal line and record, on average, the species in earlier stages of spread. Conversely, countries with higher instrength values record, on average, the species later than other countries. Country colors correspond to the bivariate combination of *x* and *y* axes. Abbreviations of country names correspond to the ISO 3166-1 alpha-3 codes.

groups (mean Pearson correlation coefficient ≥0.88; *SI Appendix*, Table S2), suggesting an overall robustness of identified patterns to variations in the minimum number of records per country. These patterns may also be sensitive to differences in the number of countries representing each continent. Therefore, we calculated $S_{\overrightarrow{ab}}$ values for 100 stratified random samples of countries, with each sample having a maximum of five countries per continent (*Materials and Methods*). For birds, nonmarine fishes, and vascular plants, the average total strength of countries in the replicated networks (*SI Appendix*, Fig. S4), as well as the average relative timings of species recording (*SI Appendix*, Fig. S5), showed only a few country-level deviations from those using data for all countries. This suggests the overall robustness of results based on all countries to geographical bias. For insects, a few more relevant changes occur, mainly affecting non-European countries. These include a substantial increase in species sharing for the USA and Argentina (ARG) (*SI Appendix*, Fig. S4*C*), as well as an increase in the earliness of relative timings of species recording for the USA (*SI Appendix*, Fig. S5*C*). The increase in species sharing for non-European countries is unsurprising considering the dominance of European countries in the whole dataset and suggests caution in interpreting patterns for alien insects beyond this continent.

We also assessed potential effects of different levels of species recording effort in shaping the obtained patterns. This consisted of calculating $S_{\overrightarrow{ab}}$ values using only the subset of 25% best-known species, which can plausibly be considered easier to identify and thus less reliant on specialized resources to be recorded. To identify this set of species, we extracted the frequency of each species' scientific name in a large corpus of multilingual books published between 1950 and 2019 (*Materials and Methods*). For all four species groups, the patterns for the best-known species show noticeable differences for a few countries (for example, Albania reduced its position as a sink country for insects) (*SI Appendix*, Figs. S6–S8). One difference consistently verified across countries and taxa concerns the occurrence of higher compositional similarities, as evidenced by warmer node colors in the networks (*SI Appendix*, Fig. S6) and corresponding higher total strength values (*SI Appendix*, Fig. S7). This increase likely means that the best-known species are generally more widespread across countries or are better recorded, as hypothesized. However, the overall structures of geographical clustering and relative timing of species recording among countries are largely similar to the ones obtained for all species. This similarity is supported by a strong correlation between the values of total strength, outstrength, and instrength of countries when using all species and the best-known species only (Pearson correlation coefficient >0.76; *SI Appendix*, Table S3), suggesting the robustness of identified patterns to differences in levels of surveillance capacity of countries.

Potential Human and Environmental Drivers. To test for significant associations between sequential recordings and geographical, political, and socioeconomic relationships of countries (*SI Appendix*, Table S4), we performed regression analyses accounting for the dyadic nature of the variables (*Materials and Methods*). Concerning the compositional similarity component of these sequences—which is given by the sum of link weights for each pair of countries—we found a significant (α = 0.05) negative relationship for geographical distance and for absolute difference in mean temperature and a positive relationship for the sum of trade flows in the 1990s and 2000s for all taxonomic groups (Fig. 3). For birds and nonmarine fishes, significant negative relationships were also identified for absolute differences in mean annual precipitation. Recent colonial relationships (i.e., if countries have had a common colonizer or colonial relationship after 1945) were nonsignificant across taxa. All models explained a substantial amount of variation in $S_{\overline{ab}}$ values ($R^2 \geq 0.53$) and had a reasonable level of out-of-sample predictive accuracy (fivefold relative absolute error (RAE) ≤ 0.68).

The significant negative relationship between geographic distance and compositional similarity is robust to variations of the threshold of minimum number of first records per country (*SI Appendix*, Fig. S9 and Dataset S1), being identified in all models across a gradient of thresholds and for the four taxonomic groups, regardless of the period of the trade data used. The negative relationship identified for temperature differences is also verified in all or the majority of replicate models, with the only exception occurring for insects when using trade data from the 1980s. The positive relationship for the sum of trade flows in the 1990s and 2000s has also been frequently identified, although with a lower prevalence than the two previous predictors. The significant negative relationship for absolute differences in mean annual precipitation has been recurrently identified for birds, but less so for

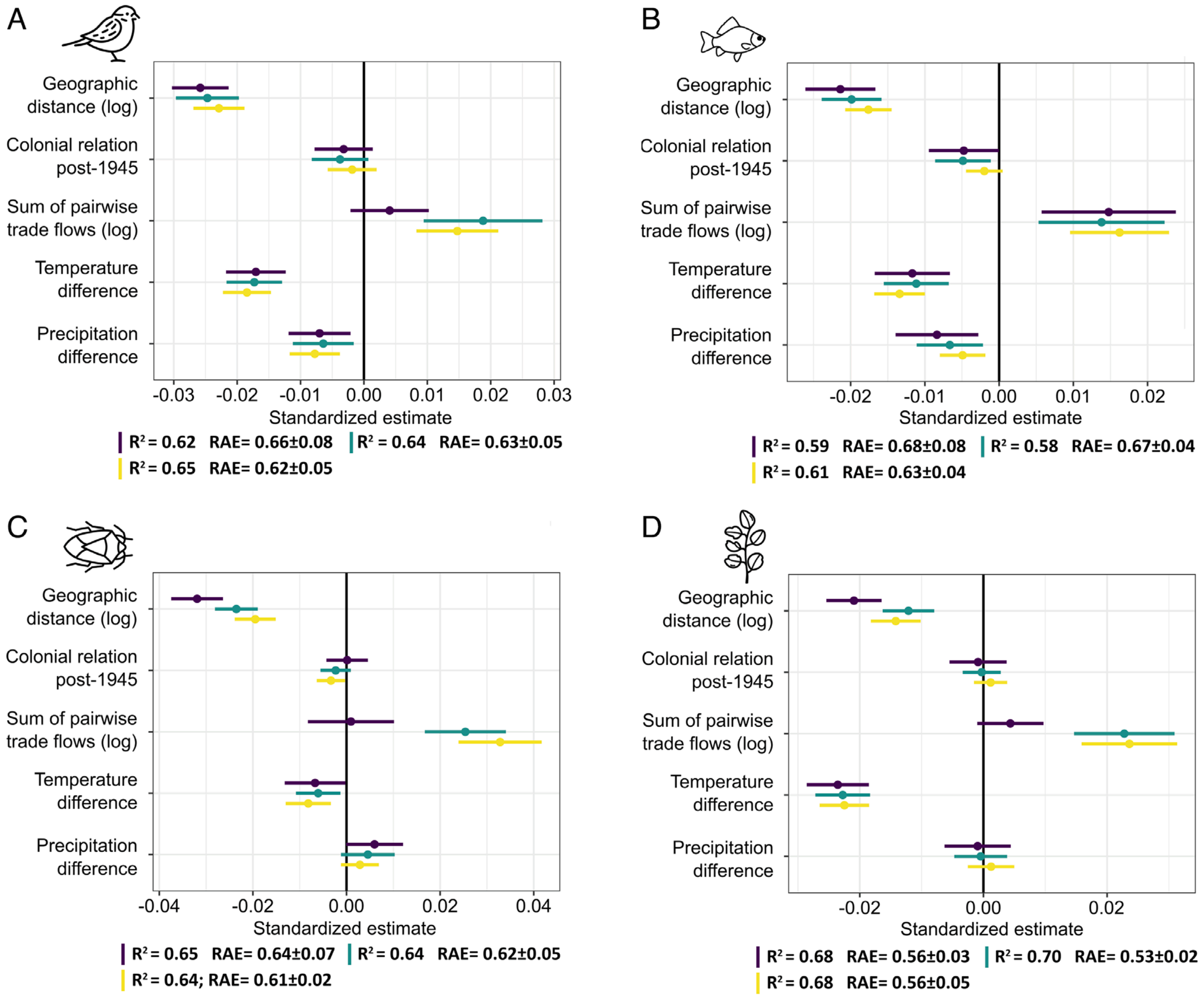

Fig. 3. Results of regression models testing for predictors of compositional similarity (i.e., the sum of $S_{\overline{ab}}$ values) between pairs of countries, for alien (A) birds, (B) nonmarine fishes, (C) insects, and (D) vascular plants. The dependent variable is logit transformed, and five variables describing pairwise relationships between countries were used as predictors. The coefficients and SE for each predictor are given for three models, each representing the sum of pairwise trade flows in a different period: 1981 to 1990 (purple), 1991 to 2000 (green), and 2001 to 2010 (yellow). Goodness of fit (R^2) and out-of-sample predictive accuracy (relative absolute error; RAE) are given for each model.

nonmarine fish, particularly in models using trade data from the 1980s. Recent colonial relationships show the highest prevalence of nonsignificant relationships and of variations in the type of significant relationships, yet for birds a significant negative relationship has been identified for this predictor in more than half of the models, regardless of the period of the trade data used. Considering only the best-known species, the relationships for geographical distance, absolute temperature differences, and sum of trade flows during the 2000s are fully maintained, whereas the relationship between sum of trade flows in the 1990s and compositional similarity for nonmarine fishes becomes nonsignificant (*SI Appendix*, Fig. S10). Absolute differences in precipitation also become nonsignificant for nonmarine fishes and vascular plants in models using trade data for the 1990s and 2000s and significant in all models for birds and insects. Replicate regression analyses for randomly sampled subsets of countries show variation in the significance of variables and in the type of relationship. For birds, nonmarine fishes, and vascular plants, significant negative relationships between compositional similarity and absolute differences in temperature and geographical distance were the most robust, occurring in all or in the majority of replicate models (*SI Appendix*, Fig. S11 and Dataset S2). On the other hand, a negative relationship between geographical distance and compositional similarity was the only relationship consistently found in replicate models for insects. The significance of relationships for the sum of trade flows had a low consistency across taxa. Altogether, these results support our findings from the network analyses that geographically close countries and countries positioned in the same climatic zones (as expressed by their mean annual temperature) have a higher propensity to record the same species.

As for the temporal component, all four taxonomic groups showed a significant positive relationship with bilateral trade values for the three time periods represented by the predictors (i.e., 1980s, 1990s, and 2000s; Fig. 4), implying that alien species are consistently recorded earlier in countries having higher trade flows (a depiction of these flows are provided in *SI Appendix*, Fig. S12). In addition, we found that alien birds, nonmarine fishes, and vascular plants are recorded significantly earlier in countries with a high GDP per capita, although depending on the period represented. No significant relationships were identified for the variable

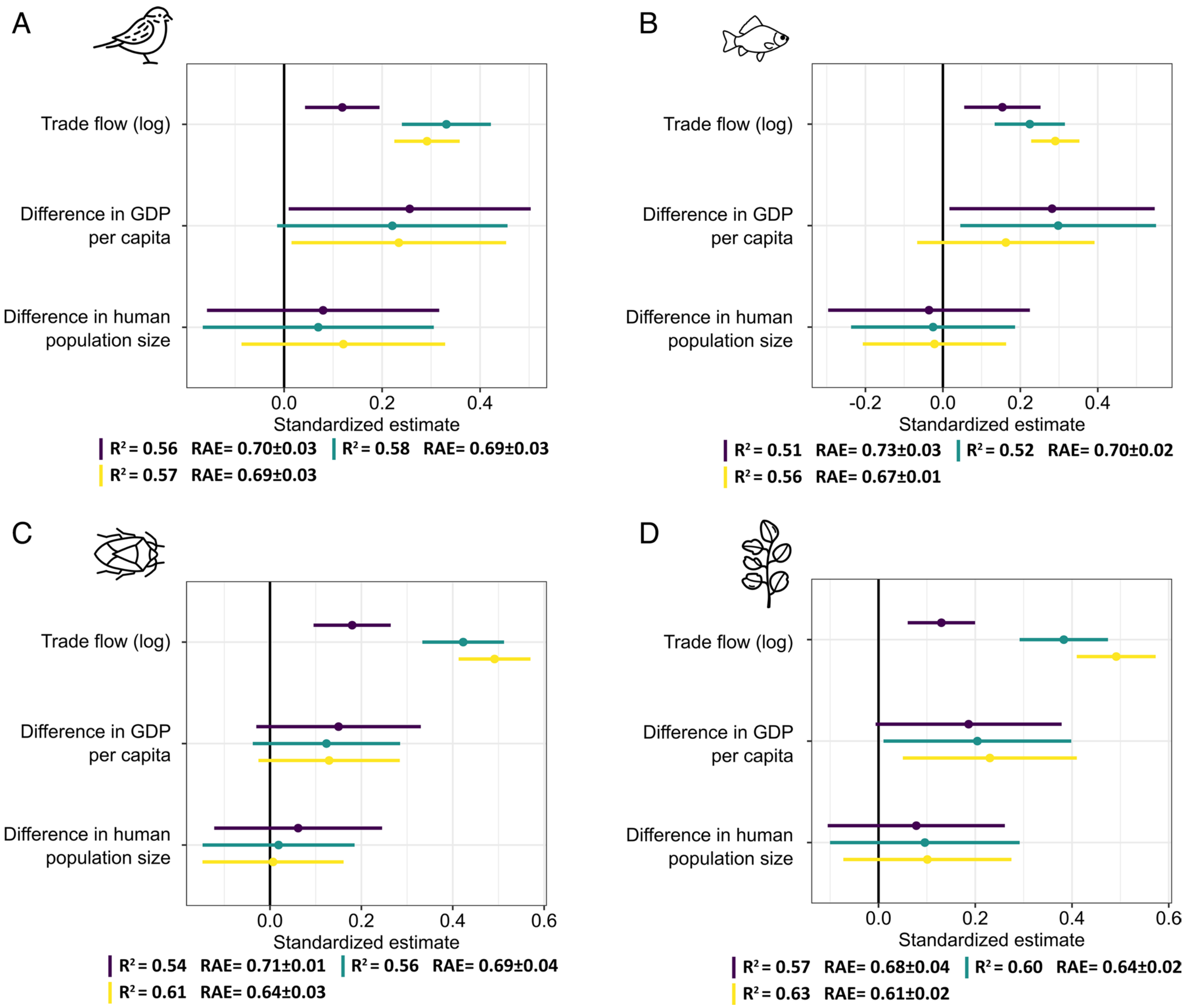

Fig. 4. Results of regression models testing for predictors of differences in the relative timing of species recording among countries (i.e. the $S_{\overrightarrow{ab}}$ values), for alien (*A*) birds, (*B*) nonmarine fishes, (*C*) insects, and (*D*) vascular plants. The dependent variable was logit transformed, and three variables describing socioeconomic relationships between pairs of countries were used from 1981 to 1990 (purple), 1991 to 2000 (green), and 2001 to 2010 (yellow). Goodness of fit (R^2) and out-of-sample predictive accuracy (relative absolute error; RAE) are given for each model.

representing differences in human population size. All models explained a substantial amount of variation in $S_{\overrightarrow{ab}}$ values ($R^2 \geq 0.51$) and had reasonable levels of out-of-sample predictive accuracy (fivefold relative absolute error ≤ 0.73).

These results are largely robust to variations in the threshold of minimum number of first records per country. The significant positive relationship for bilateral trade values is found in the majority of models along a gradient of thresholds for birds, insects, and vascular plants, regardless of the period assessed (*SI Appendix*, Fig. S13 and Dataset S1). For nonmarine fish, this relationship is also identified, although with high consistency only in models for the 2000s. On the other hand, the significant relationship between early species recording and high GDP is consistent across thresholds only for vascular plants in the 1990s and 2000s. Differences in human population were dominantly nonsignificant for all taxa and time periods. The results obtained for the best-known species are also similar to the ones for all species, with a significant positive relationship with bilateral trade values being found in all models except the one representing this variable in the 1980s for nonmarine fishes (*SI Appendix*, Fig. S14). The significance of relationships for GDP also changed slightly, with birds and vascular plants now having significant positive relationships for the three time periods. Replicate regression analyses for randomly sampled subsets of countries further support the importance of trade as being the most robust predictor for sequential recordings, occurring in most replicate models for all species groups (*SI Appendix*, Fig. S15 and Dataset S2). Relationships with other variables received very little support, including GDP per capita. Additional regression models were run including a predictor representing the capacity of countries to record alien species in their early stages of invasion, which identified a significant positive relationship with trade flows for all species groups and time periods, and nonsignificant relationships for all other variables except GDP per capita in the 1990s and 2000s for nonmarine fish (*SI Appendix*, Fig. S16). Altogether, these results support that alien species are consistently recorded earlier in countries having higher trade flows.

Discussion

Despite the publication of various studies on the global distribution of alien species (e.g., refs. 29 and 30), the dynamics of their spread remain surprisingly poorly known, although an understanding of general spread patterns is essential to predict the distribution of alien species and prevent their impacts (31). Using data for four major taxonomic groups, we here show that the initial stages of global spread of alien species tend to be concentrated in a recurrent set of countries and that subsequent spread mainly takes place in their geographical proximity. These patterns suggest a hub-and-spoke model (32), where species' spread is driven by a few countries, facilitating spread to major hubs and to adjacent countries in geographic close proximity on the same continent. This interpretation is backed by the large number of documented cases of hub-and-spoke type of spread for individual species, including birds (13), insects (12), nonmarine fishes (33), and vascular plants (34), and by the recent recognition that secondary species introductions (those originating from regions where species are nonnative) are very common in the global spread routes of alien species, the so-called bridgehead effect (16). Here, we show that this does not only apply to individual species, but indeed, such hub-and-spoke network structures can also be identified for major taxonomic groups using information on the spread for hundreds of alien species. Statistical analyses support our conclusions that the observed patterns are largely driven by geographic distance at continental scale and anthropogenic factors such as trade imports. Our analyses also indicate that species sharing is generally higher among countries having similar macroclimatic conditions, supporting our initial expectation of recurrent postestablishment spread where environmental conditions are suitable to the species requirements. This finding adds to the body of work that identifies climatic similarity between regions as an important indicator of invasion risk (6, 15, 23–25).

The degree of distinction between a backbone network of major hubs and continental subnetworks differs between taxa and is most apparent for birds and fishes. For insects, indirect support for this type of spread is available as Italy and France have been identified previously as being countries of frequent first recordings of alien insects and subsequent spread within Europe (35). However, for alien insects and plants, the networks of sequential recordings are dominated by European countries, which is likely a consequence of data availability. Although both groups represent those with the highest numbers of first records in the database, a high proportion of first records is from Europe (57% for insects and 61% for plants), resulting in high proportions of species sharing and strong links in the networks of sequential recordings for European countries. Even if many records exist for non-European countries, the unequal distribution hinders the identification of backbone networks as it was possible for birds and fishes. The influence of unequal sampling is also apparent at the continental level, where countries with particularly comprehensive data such as Belgium are also found in the center of well-connected countries and early recordings. Furthermore, the presented results are based on recordings, which are often delayed compared with actual introduction and incomplete, and do not necessarily reflect actual spread routes. Still, the generality of the results across analyses supported by sensitivity analyses indicates that the observed findings are robust.

Our results show how countries relate in terms of the identity of alien species and the relative timing of their recording. Thus, our findings offer guidance on the prioritization of efforts aiming to minimize invasion risks at transnational scales. Assuming a dominant manifestation of hub-and-spoke species spread, efforts seeking to minimize invasion risks at the global and continental scales should prioritize spread prevention from countries identified to be at the top of the spread hierarchy – i.e., the spread hubs (e.g., Australia for vascular plants, the USA for birds, Japan and the USA for nonmarine fish, and Italy and France for insects). The identification of the measures required to enable this minimization is a complex task (31), which falls outside the scope of our work; however, it is apparent that coordinated global efforts reducing the probability of spread through major hubs can have substantial knock-on effects on alien species spread worldwide (36). The spread patterns may also indicate the invasion patterns that will prevail during the coming years, reflecting the unfolding of invasions that have already been set in motion (37). For example, in the absence of better information, or of known changes in the invasion dynamics, there is no reason to expect that alien vascular plants previously recorded in Australia will stop dispersing along the Indo-Pacific region or that alien insects recorded in Italy or France will stop spreading to other European countries.

Materials and Methods

Data on First Observations of Alien Species. Data on the year of first report of alien species in regions worldwide were retrieved from the Alien Species First Record Database v2 (1, 38), which integrates previously scattered data on the year of first records of alien species in regions (e.g., countries and subnational units such as provinces or islands). This database is by far the most up-to-date and largest of its kind, holding over 60,000 first records of more than 22,000 species from a total of 280 regions (1, 38).

This database includes records of established, nonestablished, eradicated, or extinct alien species in countries and subnational regions. As our aim is to analyze general patterns of species spread, we considered all types of records, i.e., regardless of the species' invasion status. However, it is important to note that all records refer to the recording of species in the wild (i.e., individuals under human care or cultivation are not considered). In addition, recording rates for established alien species are expected to be more prevalent in the data as they represent more than 55% of all records for which establishment information is known.

We only used records for countries and autonomous subnational regions (e.g., Puerto Rico, Bermuda) (hereafter referred to as "countries" for simplicity) and excluded marine species to comply with the terrestrial-based delineation of the study units. Marine and nonmarine organisms were distinguished using the habitat information provided by the World Register of Marine Species (WoRMS; 39). Species listed as "marine" in WoRMS were removed from our analysis. As information about the realm is missing for many species, we conducted a complementary assessment for groups with a high proportion of aquatic species (fish, crustaceans, mollusks, bryozoa, and algae). Species with no records on WoRMS and identified in the literature as marine were also excluded. Finally, to reduce the sensitivity of the analysis to biases in recording dates, we reduced the temporal resolution of records from annual to quinquennial (i.e., 5-y periods). To increase the robustness of results, we only analyzed groups of species represented by a minimum of 20 countries, each country having at least 10 first records from 1950 onward. Four taxonomic groups satisfied these criteria: birds, insects, nonmarine fishes, and vascular plants, totaling 2,763, 7,175, 1,443, and 27,761 first records, respectively.

Country-Level Descriptors. We calculated the closest distance between each pair of countries (in km) in R (40) using the country delineations in Natural Earth's "Admin 0" shapefile (version 5.1.1; https://www.naturalearthdata.com). Countrywide averages of temperature and total precipitation were calculated in GIS using the raster files "annual mean temperature" and "annual precipitation" supplied by WorldClim (https://www.worldclim.org/; version 2.1; 41) for the period 1970 to 2000. The variables on the socioeconomic characteristics of countries and their historical relations were extracted from the CEPII Gravity database v.202102 (http://www.cepii.fr/; 42). These variables were a) whether the countries have ever established a colonial relationship or were part of the same colonial empire since 1945 (coded as 1 if so or 0 if otherwise); b) the difference between the GDP per capita of the first and the second country in the pair (in US$); and c) the difference between the human population size of the first and the second country in the pair. Data on countries' bilateral trade flows were extracted (in GBP and converted to US$) from the CEPII TradeHist database v.4 (http://www.cepii.fr/; 43). To account for changes in international trade patterns and socioeconomic status of countries over time, the average differences in GDP per capita and in human population size and average bilateral trade flows were calculated for three time periods: 1981 to 1990, 1991 to 2000, and 2001 to 2010. Data for previous decades were too incomplete to use.

Sequential Pattern Mining. To identify spatiotemporal patterns in first observations of alien species, we used sequential pattern mining (20). This technique takes the temporal order of events into account and has been applied to analyze sequences of frequently purchased products (44), sequences of medical prescriptions (45), or spatiotemporal patterns of pollution levels (46). In sequential pattern mining, the interest of each sequence is provided by a "support measure." This measure can reflect different criteria (e.g., sequence length and "profit"; 20), but it generally refers to the frequency of occurrence of the sequence, i.e., a high support value implies that a specific sequence appears more frequently. Following this approach, we here assessed the support of each sequence formed by an ordered pair of countries sharing the same alien species occurrence. The support was measured in terms of the proportion of recordings of the sequence relative to the number of sequential events possible, as follows:

$$S_{\overrightarrow{ab}} = \frac{\sum\left(\overrightarrow{ab}\right)}{min\left(n(a), n(b)\right)},$$

where $\overrightarrow{ab}$ is a sequence formed by an alien species recorded first in country *a* and afterward in country *b*, *n(a)* is the number of alien species recorded in country *a*, and *n(b)* is the number of alien species recorded in country *b*. We call $S_{\overrightarrow{ab}}$ the proportion of sequential recordings. Importantly, and because our interest concerns only the sequences ending in 1950 or afterward (see above), for the calculation of $\overline{ab}$, the alien species in country *a* may have been recorded prior to 1950, but in country *b*, they must have been recorded in this year or afterward. In the same manner, parameter *n(a)* refers to all alien species recorded in country *a*, but parameter *n(b)* refers to those recorded in 1950 or afterward. Thus, the support value $S_{\overrightarrow{ab}}$ corresponds to the number of species recorded first in country a and next in country b divided by the potential number of these sequences, which is expressed by the minimum species number of *a* and *b*. A valuable property resulting from the standardization of the number of spread events by the minimum size of the species pool is that the proportion obtained is robust to differences in species richness between countries, similar to the results of the Simpson metric in the calculation of compositional similarities (47). *SI Appendix*, Fig. S17 provides an illustrated example of the calculation of $S_{\overrightarrow{ab}}$ values for a pair of countries.

The values of $S_{\overrightarrow{ab}}$ range from 0, corresponding to a situation where no species in country *a* was later found in country *b*, to 1, when all species in country *a*–irrespective of their number–were found later in *b*–irrespective of how many other species were found in *b*. The calculation of this metric for all ordered pairs of countries produces a nonsymmetrical square matrix, with the matrix rows representing the value of $S_{\overrightarrow{ab}}$ when countries are first in the sequence (i.e., country *a* on the $\overrightarrow{ab}$ sequence) and the matrix columns representing the value of $S_{\overrightarrow{ab}}$ when countries are second in the sequence (i.e., country *b* on the $\overrightarrow{ab}$ sequence).

We visualized the $S_{\overrightarrow{ab}}$ matrix with each country being represented as a node of a network and the values of $S_{\overrightarrow{ab}}$ being represented as weighted directed links between country pairs. We used the Fruchterman–Reingold algorithm to find a two-dimensional layout of the networks (48). This is an iterative algorithm, which scales the width of links as a function of their weight and adjusts node positions so that countries from pairs having higher support values are grouped closer together, whereas those with lower support are shown further apart. These techniques were implemented using the "qgraph" R package (49).

We also examined the relative timing of alien species recording by country. This was performed by calculating the outstrength and instrength metrics (50) for each country in the network of $S_{\overrightarrow{ab}}$ values. The outstrength of a particular node (country) in the network corresponds to the sum of its outflowing link weights, while its instrength corresponds to the sum of the inflowing link weights. A high outstrength value means that alien species recorded in that country are frequently recorded subsequently in other countries. Conversely, a high instrength value means that a country generally behaves as a receptor of alien species from all the other countries. Accordingly, we plotted the outstrength and instrength values of countries by means of a scatterplot to assess their relative timings of alien species recording.

Testing for Drivers of Sequential Patterns of First Records. Finally, we tested for significant associations between the values of $S_{\overrightarrow{ab}}$ and country-level descriptors. We used the social relations regression model (51), which allows considering the nonindependence of dyadic-type observations (i.e., observations pertaining to a pair of subjects), as in the case of $S_{\overrightarrow{ab}}$ values.

For each species group, we performed two sets of analyses, one testing for predictors of compositional similarity among countries–given by the sum of $S_{\overrightarrow{ab}}$ for each pair of countries–and the other testing for predictors of relative timing of species recording–given by the $S_{\overrightarrow{ab}}$ values directly. In the first set, we used the five variables describing geographic and political contexts as predictors, namely geographical distance (log transformed), colonial relations after 1945, pairwise sums of trade flows, and difference in annual mean temperature and total precipitation. Three models were built, each representing the sum of trade flows in a distinct decade (1981 to 1990, 1991 to 2000, and 2001 to 2010). These variables are symmetrical, meaning that their value is the same, irrespective of the order in which the two countries are compared. To conform with this property, as a dependent variable, pairwise sums of $S_{\overrightarrow{ab}}$ values were used. These pairwise sums correspond to the link strength between a pair of countries in a network and quantify their degree of compositional similarity. These values were logit transformed to improve the robustness of model estimates (52). We hypothesized that pairs of countries showing higher relatedness (i.e., being geographically close, sharing similar climates, and sharing recent colonial relationships) will also have in common more alien species (15, 23–25).

In the second set of models, we used the socioeconomic variables GDP per capita, human population, and trade flow (log transformed) as predictors.

Three models were built for each species group, each representing values of GDP per capita, human population, and trade flows for a distinct decade (1981 to 1990, 1991 to 2000, and 2001 to 2010). Similar to $S_{\overrightarrow{ab}}$ values, these variables are directed, meaning that their values reflect the order in which the two countries are compared. Logit-transformed $S_{\overrightarrow{ab}}$ values were used as the dependent variable. We expect that countries having higher values of GDP, trade flows, and human population will generally record alien species earlier due to the combined effects of increased and earlier onset of propagule and colonization pressures (27, 28).

For both sets of models, we measured their predictive accuracy using a fivefold cross-validation procedure. In this procedure, 80% of the data are used to fit the model and 20% are left out to compare with predictions. This assessment is repeated 5 times, each time using a distinct data fold for model evaluation. Predictive performance was measured with the RAE, for which a value of zero represents a perfect match between predicted and observed values, while a value of 1 corresponds to the level of error of a "naive" predictor, consisting simply the average of observed values used for evaluation (53). Models were fit using the "amen" R package (51, 54). To rule out issues of multicollinearity among the predictor variables entering the models, we measured their variance inflation factor (VIF) using the "vif" function of "car" package (55). No relevant levels of multicollinearity were identified (i.e., all VIF values <2; *SI Appendix*, Tables S5 and S6).

Sensitivity of Results to Threshold of Minimum Number of Records per Country. Countries with a low number of first records are likely to be less representative of the general patterns of species spread as they are more sensitive to randomness in the driving processes and to potential data errors (caused, for example, by errors in species identification). On the other hand, restricting the analysis to countries with a high number of first records reduces the number of those included in the analysis and the geographical representativeness of results. We use a reference threshold of 10 or more first records per country, which we believe to provide a good trade-off between these two criteria. However, we also assess the sensitivity of results to variation in the value of this threshold. For that purpose, we calculated the $S_{\overrightarrow{ab}}$ values for thresholds ranging from 1 to 20. We then measured the Pearson correlation coefficient between countries' total strength, outstrength, and instrength for each threshold and the values of these metrics using the reference threshold (i.e., 10 or more first records). In addition, for each threshold, we reran the regression analyses testing for significant predictors of compositional similarity and sequential recordings of species among countries.

Sensitivity of Results to Number of Countries per Continental Region. To assess the sensitivity of results to differences in the number of countries representing each continent, we randomly resampled, without replacement, up to five countries per continental region. A total of 100 random samples were made, and for each sample, we computed 1) the outstrength and instrength of each country, and 2) the regression analyses for compositional similarity and sequential species recordings among countries. Outstrength and instrength values were averaged across the 100 replicate samples and visualized through a scatterplot. Results of regression analyses were summarized by summing the number of significant positive, significant negative, and nonsignificant relationships ($\alpha = 0.05$) for each predictor across the 100 samples.

Robustness of Results to Variation in Species Sampling Effort among Countries. We also assessed the robustness of the results to varying levels of species recording effort between countries. To perform this, we focused only on a subset of best-known species in each taxonomic group. Species better known to experts and nonexperts tend to be easier to identify and record (56), plausibly minimizing the role of available taxonomic expertise and surveillance efforts of countries as drivers of species recording patterns. To quantify human familiarity with each species, we used the Google Ngram (57), an algorithm that returns relative frequencies of words, or word sequences, appearing in several million digitized books published between 1500 and 2019. By this, we make the reasonable assumption that species appearing most frequently in this corpus are most familiar to experts and nonexperts. Using the ngramr package (58) for R, we obtained annual relative frequencies of the scientific name of each species for the time period covered in our analysis (1950 to 2019). These values were obtained from seven separate book corpora, each comprising books in one of seven languages (Chinese, English, French, German, Italian, Russian, and Spanish). We averaged the values for each species across years and across the seven corpora, so that species that appear most frequently across years and across a wider range of languages receive higher scores. The three most familiar species of each group, as indicated by this procedure, were *Columba livia*, *Passer domesticus*, and *Gallus gallus* for birds, *Drosophila melanogaster*, *Aedes aegypti,* and *Musca domestica* for insects, *Cyprinus carpio*, *Carassius auratus*, and *Esox lucius* for nonmarine fishes, and *Zea mays*, *Phaseolus vulgaris*, and *Vicia faba* for vascular plants.

Using the 25% highest ranking species in each taxonomic group, we computed 1) the network of first recording patterns, 2) scatterplots of countries' outstrength and instrength values, and 3) the regression analyses testing for predictors of temporal recording patterns and compositional similarity.

Robustness of Results of Temporal Models to Variation in the Capacity of Countries to Record Species in Early Stages of Invasion. We also assessed the robustness of results from temporal models to variation in the capacity of countries to record alien species in early stages of invasion. This capacity differs from "simple" species recording effort (considered above) as it is mainly determined by countries' proactiveness toward the identification of invasions (e.g., surveillance initiatives), while recording efforts are also determined by reactive efforts (e.g., after impacts being apparent). To perform this assessment, we collected the categorization of countries according to their proactive capacity toward biological invasions performed by Early et al. (59; https://www.fabiogeography.com/data). These authors ranked countries from 0 (lowest capacity) to 3 (highest capacity), according to their capacity to prevent the introduction of IAS and identify and control newly emerging invasions. This assessment was based on reports to the Convention on Biological Diversity and considered, among other things, existing legislation on cargo inspection procedures, public awareness programs, and resources for species identification and reporting of emerging or spreading invasions. Nine countries are at the top of the proactiveness ranking: Australia, Austria, Belgium, Canada, Switzerland, Germany, the United Kingdom, New Zealand, and the USA. We used the countrywide differences in this ranking as an additional predictor in the temporal models and compared the results obtained with those of models not including this predictor.

Data, Materials, and Software Availability. Previously published data were used for this work (All data are freely available online. The major database of first records underlying the analysis is available at https://doi.org/10.5281/zenodo.4632335. Sources of the predictor variables are indicated when the variable is first mentioned in the materials and methods section. R code and post-processed data sets are available at: https://doi.org/10.5281/zenodo.6683720) (60).

ACKNOWLEDGMENTS. We thank Dr. Peter Hoff for advice on the calculation of coefficients of determination for amen models and Dr. Regan Early for making the data on countries' proactive capacity toward biological invasions available. We appreciate the helpful comments of two anonymous reviewers on a previous version of the manuscript. C.C. was supported by Portuguese national funds through FCT–Fundação para a Ciência e a Tecnologia, I.P.–under research grants CPCA/A1/2790/2020CPCA/A1/394984/2021, and PTDC/GES-OUT/30210/2017, support to the CEG/IGOT Research Unit (UIDB/00295/2020 and UIDP/00295/2020), and the program of "Stimulus of Scientific Employment–Individual Support" within the contract "CEECIND/02037/2017." M.P. was supported by Portuguese national funds through FCT in the scope of Norma Transitória–DL57/2016/CP1440/CT0017. F.E. and H.S. received funding from the 2017 to 2018 Belmont Forum and BiodivERsA joint call for research proposals, under the BiodivScen ERA-Net COFUND program, and with the funding organization Austrian Science Foundation FWF (grant I 4011-B32) and BMBF (grant number 16LC1807A). We are grateful to the following individuals for creating and making available, under a creative commons license, the species icons used in this study: Alex Muravev (fish), Olena Panasovska (plant), and Yu Luck (bird and insect).

Author affiliations: [a]Centre of Geographical Studies, Institute of Geography and Spatial Planning, University of Lisbon, 1600-276 Lisboa, Portugal; [b]Associated Laboratory Terra, Tapada da Ajuda, 1349-017 Lisboa, Portugal; [c]Bioinvasions, Global Change, Macroecology Group-Department of Botany and Biodiversity Research, University of Vienna, 1030 Vienna, Austria; [d]Centro de Investigação em Biodiversidade e Recursos Genéticos, Rede de Investigação em Biodiversidade e Biologia Evolutiva, Laboratório Associado, Campus de Vairão, Universidade do Porto, 4485-661 Vairão 4485-661, Portugal; [e]Centro de Investigação em Biodiversidade e Recursos Genéticos, Rede de Investigação em Biodiversidade e Biologia Evolutiva, Laboratório Associado, Instituto Superior de Agronomia, Universidade de Lisboa, 1349-017 Lisboa, Portugal; [f]BIOPOLIS Program in Genomics, Biodiversity and Land Planning, Centro de Investigação em Biodiversidade e Recursos Genéticos,Campus de Vairão, 4485-661 Vairão, Portugal; and [g]Senckenberg Biodiversity and Climate Research Centre, 60325 Frankfurt, Germany

1. H. Seebens *et al.*, No saturation in the accumulation of alien species worldwide. *Nat. Commun.* **8**, 1–9 (2017).
2. P. Pyšek *et al.*, Scientists' warning on invasive alien species. *Biol. Rev.* **95**, 1511–1534 (2020).
3. P. E. Hulme, Unwelcome exchange: International trade as a direct and indirect driver of biological invasions worldwide. *One Earth* **4**, 666–679 (2021).
4. C. Capinha, F. Essl, H. Seebens, D. Moser, H. M. Pereira, The dispersal of alien species redefines biogeography in the Anthropocene. *Science* **348**, 1248–1251 (2015).
5. E. K. Longman, K. Rosenblad, D. F. Sax, Extreme homogenization: The past, present and future of mammal assemblages on islands. *Glob. Ecol. Biogeogr.* **27**, 77–95 (2018).
6. C. Capinha, F. Marcolin, L. Reino, Human-induced globalization of insular herpetofaunas. *Glob. Ecol. Biogeogr.* **29**, 1328–1349 (2020).
7. T. M. Blackburn, C. Bellard, A. Ricciardi, Alien versus native species as drivers of recent extinctions. *Front Ecol. Environ.* **17**, 203–207 (2019).
8. C. Bellard, C. Bernery, C. Leclerc, Looming extinctions due to invasive species: Irreversible loss of ecological strategy and evolutionary history. *Glob. Chang. Biol.* **27**, 4967–4979 (2021).
9. C. Diagne *et al.*, High and rising economic costs of biological invasions worldwide. *Nature* **592**, 571–576 (2021).
10. P. E. Hulme, Invasive species challenge the global response to emerging diseases. *Trends Parasitol.* **30**, 267–270 (2014).
11. M. Vilá *et al.*, Viewing emerging human infectious epidemics through the lens of invasion biology. *BioScience* **71**, 722–740 (2021).
12. Q. Tang, T. Bourguignon, L. Willenmse, E. De Coninck, T. Evans, Global spread of the German cockroach, Blattella germanica. *Biol. Invasions* **21**, 693–707 (2019).
13. A. Peña-Peniche, C. Mota-Vargas, M. García-Arroyo, I. MacGregor-Fors, On the North American Invasion of the house sparrow and its absence in the Yucatan Peninsula. *Avian Conserv. Ecol.* **16**, 18 (2021).
14. N. W. Siegert, D. G. McCullough, A. M. Liebhold, F. W. Telewski, Dendrochronological reconstruction of the epicentre and early spread of emerald ash borer in North America. *Divers. Distrib.* **20**, 847–858 (2014).
15. R. Kannan, C. M. Shackleton, R. Uma Shaanker, Reconstructing the history of introduction and spread of the invasive species, Lantana, at three spatial scales in India. *Biol. Invasions.* **15**, 1287–1302 (2013).
16. C. Bertelsmeier *et al.*, Recurrent bridgehead effects accelerate global alien ant spread. *Proc. Natl. Acad. Sci. U.S.A.* **115**, 5486–5491 (2018).
17. J. R. U. Wilson, E. E. Dormontt, P. J. Prentis, A. J. Lowe, D. M. Richardson, Something in the way you move: Dispersal pathways affect invasion success. *Trends Ecol. Evol.* **24**, 136–144 (2009).
18. H. Seebens, F. Essl, B. Blasius, The intermediate distance hypothesis of biological invasions. *Ecol. Lett.* **20**, 158–165 (2017).
19. J. M. Gippet, C. Bertelsmeier, Invasiveness is linked to greater commercial success in the global pet trade. *Proc. Natl. Acad. Sci. U.S.A.* **118**, e2016337118 (2021).
20. P. Fournier-Viger, J.C.-W. Lin, R. U. Kiran, Y. S. Koh, R. Thomas, A survey of sequential pattern mining. *Data Sci. Pattern Recognit.* **1**, 54–77 (2017).
21. P. E. Hulme, Trade, transport and trouble: Managing invasive species pathways in an era of globalization. *J. Appl. Ecol.* **46**, 10–18 (2009).
22. C. Bertelsmeier, S. Ollier, A. Liebhold, L. Keller, Recent human history governs global ant invasion dynamics. *Nat. Ecol. Evol.* **1**, 0184 (2017).
23. M. Clavero, E. García-Berthou, Homogenization dynamics and introduction routes of invasive freshwater fish in the Iberian Peninsula. *Ecol. Appl.* **16**, 2313–2324 (2006).
24. D. P. Bebber, T. Holmes, S. J. Gurr, The global spread of crop pests and pathogens. *Glob. Ecol. Biogeogr.* **23**, 1398–1407 (2014).
25. F. Ascensão *et al.*, Drivers of compositional dissimilarity for native and alien birds: The relative roles of human activity and environmental suitability. *Biol. Invasions.* **22**, 1447–1460 (2020).
26. Q. Yang *et al.*, The global loss of floristic uniqueness. *Nat. Commun.* **12**, 1–10 (2021).
27. J. L. Lockwood, P. Cassey, T. M. Blackburn, The more you introduce the more you get: The role of colonization pressure and propagule pressure in invasion ecology. *Divers. Distrib.* **15**, 904–910 (2009).
28. W. Dawson *et al.*, Global hotspots and correlates of alien species richness across taxonomic groups. *Nat. Ecol. Evol.* **1**, 1–7 (2017).
29. M. Van Kleunen *et al.*, Global exchange and accumulation of non-native plants. *Nature* **525**, 100–103 (2015).
30. E. E. Dyer, D. W. Redding, T. M. Blackburn, The global avian invasions atlas, a database of alien bird distributions worldwide. *Sci. Data* **4**, 1–12 (2017).
31. A. Ricciardi *et al.*, Four priority areas to advance invasion science in the face of rapid environmental change. *Environ. Rev.* **29**, 119–141 (2021).
32. J. T. Carlton, Pattern, process, and prediction in marine invasion ecology. *Biol. Conserv.* **78**, 97–106 (1996).
33. A. Simon *et al.*, Invasive cyprinid fish in Europe originate from the single introduction of an admixed source population followed by a complex pattern of spread. *PLoS One* **6**, e18560 (2011).
34. P. Pyšek, V. Jarošík, J. Müllerová, J. Pergl, J. Wild, Comparing the rate of invasion by *Heracleum mantegazzianum* at continental, regional, and local scales. *Divers. Distrib.* **14**, 355–363 (2008).
35. A. Roques *et al.*, Are invasive patterns of non-native insects related to woody plants differing between Europe and China? *Front. For. Glob. Change* **2**, 91 (2020).
36. H. Seebens, M. Gastner, B. Blasius, F. Courchamp, The risk of marine bioinvasion caused by global shipping. *Ecol. Lett.* **16**, 782–790 (2013).
37. F. Essl *et al.*, Socioeconomic legacy yields an invasion debt. *Proc. Natl. Acad. Sci. U.S.A.* **108**, 203–207 (2011).
38. H. Seebens *et al.*, Global rise in emerging alien species results from increased accessibility of new source pools. *Proc. Natl. Acad. Sci. U.S.A.* **115**, E2264–E2273 (2018).
39. WoRMS Editorial Board, World register of marine species (2019). https://www.marinespecies.org/ (Accessed 12 July 2021).
40. R Core Team, *A Language and Environment for Statistical Computing* (R Foundation for Statistical Computing, Vienna, Austria, 2012). https://www.R-project.org.
41. S. E. Fick, R. J. Hijmans, WorldClim 2: New 1-km spatial resolution climate surfaces for global land areas. *Int. J. Climatol.* **37**, 4302–4315 (2017).
42. T. Mayer, S. Zignago, Notes on CEPII's Distance Measures: The GeoDist database, (Centre d'Études Prospectives et d'Informations Internationales, Working Papers 2011-25 (2011). Retrieved 15 September 2021, from http://www.cepii.fr/PDF_PUB/wp/2011/wp2011-25.pdf.
43. M. Fouquin, J. Hugot, Two centuries of bilateral trade and gravity data: 1827–2014, (Centre d'Études Prospectives et d'Informations Internationales, Working Papers 2016-14 (2016). Retrieved 16 September 2021, from http://www.cepii.fr/PDF_PUB/wp/2016/wp2016-14.pdf.
44. R. Bhatta, C. Ezeife, M. N. Butt, "Mining sequential patterns of historical purchases for e-commerce recommendation" in *International Conference on Big Data Analytics and Knowledge Discovery* (Springer, 2019), pp. 57–72.
45. A. P. Wright, A. T. Wright, A. B. McCoy, D. F. Sittig, The use of sequential pattern mining to predict next prescribed medications. *J. Biomed. Inform.* **53**, 73–80 (2015).
46. L. Zhang, G. Yang, X. Li, Mining sequential patterns of PM2. 5 pollution between 338 cities in China. *J. Environ. Manage.* **262**, 110341 (2020).
47. A. Baselga, Partitioning the turnover and nestedness components of beta diversity. *Glob. Ecol. Biogeogr.* **19**, 134–143 (2010).
48. T. M. Fruchterman, E. M. Reingold, Graph drawing by force-directed placement. *Softw. Pract. Exp.* **21**, 1129–1164 (1991).
49. S. Epskamp, A. O. Cramer, L. J. Waldorp, V. D. Schmittmann, D. Borsboom, qgraph: Network visualizations of relationships in psychometric data. *J. Stat. Softw.* **48**, 1–18 (2012).
50. T. Opsahl, F. Agneessens, J. Skvoretz, Node centrality in weighted networks: Generalizing degree and shortest paths. *Soc. Networks* **32**, 245–251 (2010).
51. P. Hoff, Additive and multiplicative effects network models. *Stat. Sci.* **36**, 34–50 (2021).
52. D. I. Warton, F. K. Hui, The arcsine is asinine: The analysis of proportions in ecology. *Ecology* **92**, 3–10 (2011).
53. I. H. Witten, E. Frank, M. A. Hall, *Data Mining: Practical Machine Learning Tools and Techniques* (Morgan Kaufmann, ed. 3, 2011).
54. P. D. Hoff, Dyadic data analysis with amen. arXiv [Preprint] (2015). https://arxiv.org/abs/1506.08237 (Accessed 20 January 2022).
55. J. Fox, S. Weisberg, *An R Companion to Applied Regression* (Sage publications, ed. 3, 2018).
56. E. H. Boakes *et al.*, Distorted views of biodiversity: Spatial and temporal bias in species occurrence data. *PLoS Biol.* **8**, e1000385 (2010).
57. Y. Lin, "Syntactic annotations for the google books ngram corpus" in *Proceedings of the ACL 2012 System Demonstrations* (Association for Computational Linguistics, 2012), pp. 169–174.
58. S. Carmody, Ngramr: Retrieve and Plot Google N-Gram Data (R package, Version 1.7.2, 2020). https://cran.r-project.org/web/packages/ngramr/ngramr.pdf. Accessed 10 October 2021.
59. R. Early *et al.*, Global threats from invasive alien species in the twenty-first century and national response capacities. *Nat. Commun.* **7**, 1–9 (2016).
60. C. Capinha, F. Essl, M. Porto, H. Seebens, R scripts and complementary data for manuscript "The worldwide networks of spread of recorded alien species". Zenodo. https://doi.org/10.5281/zenodo.6683720. Deposited 22 June 2022.