Global rise in emerging alien species results from increased accessibility of new source pools


Our ability to predict the identity of future invasive alien species is largely based upon knowledge of prior invasion history. Emerging alien species—those never encountered as aliens before—therefore pose a significant challenge to biosecurity interventions worldwide. Understanding their temporal trends, origins, and the drivers of their spread is pivotal to improving prevention and risk assessment tools.

Here, we use a database of 45,984 first records of 16,019 established alien species to investigate the temporal dynamics of occurrences of emerging alien species worldwide. Even after many centuries of invasions the rate of emergence of new alien species is still high: One-quarter of first records during 2000–2005 were of species that had not been previously recorded anywhere as alien, though with large variation across taxa. Model results show that the high proportion of emerging alien species cannot be solely explained by increases in well-known drivers such as the amount of imported commodities from historically important source regions. Instead, these dynamics reflect the incorporation of new regions into the pool of potential alien species, likely as a consequence of expanding trade networks and environmental change. This process compensates for the depletion of the historically important source species pool through successive invasions. We estimate that 1–16% of all species on Earth, depending on the taxonomic group, qualify as potential alien species. These results suggest that there remains a high proportion of emerging alien species we have yet to encounter, with future impacts that are difficult to predict.

Significance

Our ability to predict the identity of future invasive alien species is largely based upon knowledge of prior invasion history. Emerging alien species—those never before encountered as aliens—therefore pose a significant challenge to biosecurity interventions worldwide. Using a global database of the first records of alien species covering the years 1500–2005 we detected a surprisingly high proportion of species in recent records that have never been recorded as alien before. The high proportion of these emerging alien species mainly resulted from the increased accessibility of new source species pools in the native range. Risk assessment approaches that rely less on invasion history will need to be prioritized.
which we define as first records of so-called emerging alien species. Note that the term “emerging alien species” describes a transient status of each alien species at its first detection globally. Hence, every alien species was an emerging alien species once. The dynamics of emerging alien species accumulation provides a direct measure of ongoing invasion dynamics without the confounding effect of subsequent introductions either from the native range or from already occupied regions in the alien range.

In addition, we investigated the proportions of first records of emerging alien species among all alien species. From the proportions of emerging alien species we can infer the size of the global source pools of potential new alien species (hereafter “candidate species pool”) for different taxa and their changes over time. The candidate species pools include those native species with a high chance of becoming an alien species somewhere else at some time (this pool does not encompass all native species but is limited to those with a high potential of being introduced and establishing in a new region). A high proportion of emerging alien species indicates that the alien species originated from a source pool of candidate species that is far from being depleted. Knowledge about the proportion of emerging alien species will also be important for biosecurity, which often relies on information of known alien species (15, 16), and horizon scanning studies aiming at identifying “door-knocker” species, which are species not yet recorded but suspected to have a high risk of arrival and impacts (17).

Our study specifically addresses five questions. (i) How did first records of emerging alien species develop during recent centuries? (ii) Do we find evidence for depletion of the source pool of potential new alien species? (iii) How does variation in sampling intensity affect the observed patterns? (iv) What are the drivers of the temporal dynamics? (v) Do the spatiotemporal dynamics vary among major taxonomic groups?

Results

The distribution of the number of first records per alien species was highly skewed, with the majority of species (n = 9,984, 58%) having just a single first record in the database (Fig. 1). Eighty-six percent of all species have no more than two first records on the same continent, which indicates a narrow distribution in the alien range for most species, similar to what has been found in other studies (18, 19), and a comparatively low number of first records due to subsequent introductions to the same continent. By contrast, 26 species had more than 50 first records, with the top five being the domestic pigeon (Columbia livia, first records in 197 regions), longhorn crazy ant (Paratrechina longicornis, 134), big-headed ant (Pheidole megacephala, 92), house sparrow (Passer domesticus, 87), and common rabbit (Oryctolagus cuniculus, 82). The vascular plant with the highest number of first records is Canadian horseweed (Erigeron canadensis, 40).

Global Temporal Dynamics of Emerging Alien Species. As with all alien species, the first-record rates of emerging alien species increased distinctly over time, particularly during the 20th century (circles in Fig. 2). The proportion of emerging alien species among all alien species generally declined during recent centuries (Fig. 2). However, the proportion of emerging alien species was still high in the most recent years captured by our database (2000–2005), with the highest values found among molluscs and other invertebrates (every second first record was that of an emerging alien species), followed by crustaceans and vascular plants (every third record was an emerging alien species), fishes, mammals, and insects (every fourth). By contrast, for alien birds only every 16th first record in the period 2000–2005 was that of an emerging alien species.

A decline in the proportions of emerging alien species with time can be expected for two reasons: (i) a limited pool of potential new alien species that should deplete with ongoing establishment events, resulting in a declining number of emerging alien species, and (ii) an increase in first records due to subsequent occurrences of nonemerging alien species. The latter, however, should have a low influence on results given the comparatively low number of first records of the same species in the database. To analyze the influence of a depleting candidate species pool we analyzed the accumulation of alien species using a simple invasion model, simulating the spread of individuals from an estimated candidate species pool into a new environment (Materials and Methods). By fitting this model to observed first-record rates we were able to estimate the size of the candidate species pool X at year t. This approach is similar to those applied to estimate the total number of species on Earth derived from the rate of newly described native species (20). Assuming a constant pool of candidate species over time (Xt = X, for all t), this model already captures a large amount of the observed variability in the frequency of emerging alien species, thereby supporting the idea of a depleting candidate species pool (SI Appendix, Fig. S1).

Although the model is able to reproduce the general decline in the proportions of emerging alien species, it does not capture observed deviations from the declining trends such as those for vascular plants and mammals (Fig. 2). We therefore modified the model to allow for temporal variation in the predicted pool of candidate species X and determined the temporal development of the species pool that resulted in the best description of the dynamics of emerging alien species numbers. This model extension describes the observed development of the proportion of emerging alien species with high confidence (red dots in Fig. 2), with R² values of 0.8 or higher for six taxonomic groups. Crucially, the candidate species pool predicted by the model increased over time (Fig. 2, Lower), particularly in the 20th century. At their maxima, the candidate species pools derived from simulation results were lowest for mammals (499 species) and highest for vascular plants (26,048) (Table 1).

The predicted size of the candidate species pools indicates that between 24% (insects) and 65% (birds) of the number of species in the candidate species pools have already been established somewhere outside their native ranges (Table 1). Given that first records are not available for all invasion events, the number of first records is likely larger and consequently the estimated size of the candidate species pools may also be higher. Using reported total numbers of established alien species available from the literature, the current estimated full candidate species pools (Materials and Methods) are as follows (Table 1): 425 species (crustaceans), 539 species (molluscs), 890 species (other invertebrates), 1,000 species (plants), 1,200 species (birds), and 1,800 species (mammals).
Fig. 2. Time series of observed first-record rates (black circles) and simulation results of the invasion model (red dots) for eight taxonomic groups. Observed first-record rates are shown for all alien species (Upper), emerging alien species (Upper Middle), and the proportion of emerging among all alien species (Lower Middle). In the invasion model, species were randomly selected from the predicted candidate species pool according to the first-record rate of all alien species to simulate the dynamics of emerging alien species. The temporal development of the size of the candidate species pool (Lower) is the result of the model fitting and represents the species pools, which are necessary to reproduce the dynamics of emerging alien species best. Black dots indicate the knots, which define the temporal dynamics of the candidate species pool (Materials and Methods). The congruence between observed and predicted proportions of emerging alien species is shown as \( R^2 \) values. Note that the \( y \) axes vary in scale.

A sensitivity analysis, attempting to capture the influence of sampling intensity across time on first-record rates, revealed that without the influence of sampling intensity first-record rates should have increased earlier than observed, which would have resulted in higher absolute numbers of first records before the 20th century and lower numbers after ca. 1900 (SI Appendix, Fig. S2). However, the overall patterns of the time series remained similar to the original data. Furthermore, simulation results of the invasion model fitted to the corrected time series of first records only showed minor deviations from those obtained with the original first-record rates (compare Fig. 2 and SI Appendix, Fig. S3). In addition, first-record rates may be affected by different rates of description of new species varying among taxonomic groups, which may at least partly explain the late increase in first-record rates of invertebrates, and by spatial variation in sampling intensity, which may particularly affect first-record rates of less investigated regions such as Africa, South America, or central Asia.

The emergence of a new alien species represents an initial stage of the invasion process, and thus information about the hotspots of emerging alien species indicates the typical entrance regions of new alien species and likely starting points for further spread. Many of the regions with the highest proportions of emerging alien species represent the large and emerging economies around the globe, but also many islands, particularly for invertebrates (SI Appendix, Fig. S4 and see SI Appendix, Fig. S5 for an examination of Europe). In general, the proportions were low for vascular plants, mammals, and birds and high for invertebrates. The absolute numbers of emerging alien species showed a similar global pattern (SI Appendix, Fig. S6).

Drivers of Temporal Dynamics of First Records. To analyze the drivers shaping the temporal dynamics of first-record rates we performed a regression analysis using a set of up to eight predictors (time series of temperature, relative humidity, import values, three land-use categories, number of botanical gardens, and human population size). We applied generalized additive mixed models (GAMMs) to model temporal variation in first records of emerging and nonemerging alien species (the latter representing first records of species with already known occurrences elsewhere), thereby accounting for nonlinearity and spatial and temporal autocorrelation, with continent as a random effect variable. The statistical analysis revealed that the value of imported commodities ("imports" in Fig. 3) was frequently selected as a significant predictor of temporal variation in first-record rates of emerging alien species for vascular plants, birds, fishes, insects, molluscs, and other invertebrates. Changes in land use were also a significant predictor for many taxonomic groups; however, the type of land-use change...
Table 1. Estimates of candidate species pools and their relation to reported numbers of native and alien species

<table>
<thead>
<tr>
<th>Metric</th>
<th>Birds</th>
<th>Crustaceans</th>
<th>Fishes</th>
<th>Insects</th>
<th>Mammals</th>
<th>Molluscs</th>
<th>Other invertebrates</th>
<th>Vascular plants</th>
</tr>
</thead>
<tbody>
<tr>
<td>Estimated candidate species pool</td>
<td>625</td>
<td>1,565</td>
<td>1,354</td>
<td>20,611</td>
<td>499</td>
<td>1,289</td>
<td>3,268</td>
<td>26,048</td>
</tr>
<tr>
<td>No. of alien species in analysis</td>
<td>406</td>
<td>430</td>
<td>478</td>
<td>4,992</td>
<td>248</td>
<td>441</td>
<td>780</td>
<td>7,380</td>
</tr>
<tr>
<td>Percentage of established alien species, %</td>
<td>65</td>
<td>27</td>
<td>35</td>
<td>24</td>
<td>50</td>
<td>34</td>
<td>28</td>
<td></td>
</tr>
<tr>
<td>Reported total no. of alien species</td>
<td>971</td>
<td>425</td>
<td>944</td>
<td>445</td>
<td>539</td>
<td>13,168</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Estimated true candidate species pool</td>
<td>1,494</td>
<td>1,574</td>
<td>2,697</td>
<td>890</td>
<td>1,585</td>
<td>47,029</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Estimated total no. of native species on Earth</td>
<td>10,000</td>
<td>150,000</td>
<td>40,000</td>
<td>5,500</td>
<td>200,000</td>
<td>368,000</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The estimated candidate species pool represents the maximum size of the species pool predicted by the model (Fig. 2, Lower), while the percentage of established alien species denotes the proportion of species from this pool that have already established an alien population according to our dataset. As first records are not available for all invasion events, the total number of alien species in nature is larger than those recorded in the first-record database. To circumvent this limitation, reported total numbers of alien species were taken from the literature instead to calculate the estimated full candidate species pool using the same proportion of established alien species. For insects and other invertebrates no reports are available.

* Dyer et al. (14).
* Pagad et al. (42) (mostly marine species).
* Froese and Pauly (39).
* Dawson et al. (41).
* Capinha et al. (2).
* van Kleunen et al. (40).
* Chapman (21).

The occurrence of a high proportion (SI Appendix, Fig. S4) of emerging alien species in a region indicates likely starting points for further spread of these newly appearing alien species. Our identified starting points generally agree with identified hotspots of predicted future increases in invasion threats being mostly located in Europe, North America, and East Asia (24) but also in emerging economies such as Brazil or Argentina (25). In addition, our analysis highlights the distinct variation in starting points among taxonomic groups and the importance of many islands as likely gateways for emerging alien species. It is unclear, however, whether these species will spread further or remain in their narrow alien ranges.

The high number of species with only a very few records in our database indicates that most species will not spread widely, but we are still far from being able to determine those species with a high potential for spread.

The statistical analysis suggests that the increases in first-record rates are associated with increasing import values and changes in land use, though the relative importance of these drivers varied among taxonomic groups (Fig. 3). Changes in temperature and relative humidity were infrequently selected as significant predictors; however, the effect of climatic drivers may have been underestimated in our approach as we could only consider changes in the recipient region, rather than environmental matches between donor and recipient regions, which may (fraction of urban areas, pasture, or cropland) varied among taxonomic groups. Variation in climactic conditions, indicated by annual mean temperature and relative humidity, was less important. The number of botanical gardens was a significant predictor for emerging alien vascular plants. The predictor “year” was frequently selected as a significant driver, which indicates that parts of the long-term trends in first-record rates could not be explained by the drivers considered here. In addition, some GAMMs (e.g., for fishes or insects) have very low predictive power, though with several significant drivers (Fig. 3), which further highlights the lack of essential variables. The GAMM analyses should therefore be considered as an initial step toward a general overview of the most important drivers of temporal variation in first records of alien species.

**Discussion**

Our analysis builds on a previous study (1) by analyzing emerging alien species separately, to remove the effect of multiple first records of the same species due to, for example, secondary spread on first-record rates and allow us to draw more direct inferences about the past and future dynamics of alien species introductions. For all taxonomic groups first records of all alien species increased distinctly during the 19th and 20th centuries, which should—based on theoretical considerations—result in an accelerated decline in the proportion of emerging alien species due to a faster depletion of the pool of potential new alien species (SI Appendix, Fig. S1). In contrast, we found comparatively low rates of decline in the proportions of emerging alien species. For vascular plants, mammals, and fishes the proportions even remained constant during the last 150 y, although the total number of alien species records increased. These dynamics cannot be explained by increasing drivers of alien species introductions such as the amount of imported commodities, which to the contrary should result in an even faster depletion of candidate species pools. The high proportion of emerging alien species in recent years is likely a consequence of an increased accessibility of candidate species pools in the native range (Fig. 2).

The introduction of alien species has a centuries-long history and one might expect that the proportion of emerging alien species would have declined to low levels. Surprisingly, the proportion was still high in 2000–2005, with on average every fourth new first record being of an emerging alien species. Alien birds seemed to be the only exception to this trend, as the proportion of emerging alien species in this group has distinctly declined recently. This suggests that many alien bird species, which have already established around the world, are currently expanding their alien range either through natural or human-assisted dispersal. This can partly be attributed to the intensified trade of a specific pool of bird species used for cultural practices such as prayer releases in Asia, and the concomitant increased likelihood of accidental releases (22). However, this may change in the future as there are many birds not yet established outside their native ranges, which may be attractive for the Asian market, and thus more emerging alien species are likely to appear in the future (14). Indeed, there are already signs that neotropical bird species are increasing in the Asian bird markets (23).
have been important as well (26). Interaction terms between environmental variables never improved the model fits. Import value already proved to be a good predictor of alien species richness and is a commonly used proxy for introduction rates of alien species (9–11). Land use is also known to affect biological invasions as land degradation increases the chance of establishment of alien species, though the significance of this driver has been mostly reported for alien vascular plants (12), while it was often found to be not significant in cross-taxonomic analyses (11, 27). Our statistical analysis showed that both changes in introduction rates, as indicated by the significant effect of imports, and establishment rates, as indicated by the significant effect of land use, were likely important drivers of the accumulation of alien species in general. However, in some cases the GAMMs could only explain a very low amount of the observed variation, which is indicated by an adjusted $R^2$ of zero (Fig. 3). This shows that although many predictors have significant effects on the time series of first-record rates, important predictor variables are seemingly still lacking. In addition, changes in the candidate species pools could not be considered in the statistical analysis, which may at least partly explain the low predictive power of the statistical models.

For alien vascular plants, the number of botanical gardens was the distinct increase in the first-record rates, may indicate (i) an increase in the rate of establishment or (ii) an increase in the size of the global candidate species pool. An increase in the rate of establishment is supported by the statistical analysis, which shows a significant influence of changes in land use on first-record rates (Fig. 3). However, the effect of land use on first-record rates was not consistent among taxonomic groups and cannot explain the temporal development of emerging alien species for all taxonomic groups. Simulation results show an increase in the candidate species pools for all taxonomic groups (Fig. 2), which can explain the flattening of the proportions of emerging alien species, which supports previous findings of the role of botanical gardens for the introduction of alien vascular plants (28). The absolute number of botanical gardens may not be the most appropriate predictor as it ignores the number of planted species and species origins, which were not available. Likewise, data on other drivers such as introductions by acclimatization societies (29), European explorers or settlers (30), and plant hunters (31) are largely lacking, which highlights the need to improve the availability of historical data for more detailed analyses of spatiotemporal invasion dynamics.

The still high and in some cases even constant proportion of emerging alien species among first records, in combination with the distinct increase in the first-record rates, may indicate (i) an increase in the rate of establishment or (ii) an increase in the size of the global candidate species pool. An increase in the rate of establishment is supported by the statistical analysis, which shows a significant influence of changes in land use on first-record rates (Fig. 3). However, the effect of land use on first-record rates was not consistent among taxonomic groups and cannot explain the temporal development of emerging alien species for all taxonomic groups. Simulation results show an increase in the candidate species pools for all taxonomic groups (Fig. 2), which can explain the flattening of the proportions of emerging alien species.
very well. An increase in candidate species pools seems likely to be a consequence of an increasing accessibility of these species pools, which may be due to increasing access to and integration of new source regions into the global exchange network (new routes of invasion) (25) and the emergence of new introduction pathways (e.g., fashion trends and the pet trade) (32). In addition, other factors such as changes in environmental conditions or land use may have enabled other alien species to establish, which should also result in an increasing size of candidate species pools (24). Our results indicate that the increases in alien species numbers in general and that of emerging alien species in particular can be explained by the interplay of increases in candidate species pools in the native range, increases in introduction rates due to, for example, greater volume of imports, and probably rising establishment rates as a consequence of land degradation in the recipient regions. The predicted rise in the size of candidate species pools likely compensated for the effect of their depletion due to elevating introduction and establishment rates.

The candidate species pools are predicted to encompass 1–16% of all species on Earth for the various taxonomic groups, with particularly high values observed for vertebrates and vascular plants (Table 1). Those are groups with the most comprehensive data and, thus, the low numbers for invertebrates may also be affected by sampling biases. These numbers are, however, fraught with uncertainties as both the true size of the candidate species pools and the true number of species on Earth are poorly known, and thus these results should be interpreted as rough estimates. Given the highly uneven spatial distributions of origins of alien species (14) we can expect that some regions (33) or habitats (34) provide a considerably larger number of potential alien species, which need to be identified to improve our predictions of alien species dynamics.

In conclusion, our study reveals that global invasion dynamics are still prominently driven by the introductions of emerging alien species and thus by primary introductions, while widespread alien distributions are comparatively rare. This, however, also depends on the resolution of the analysis, and the relationship between emerging and nonemerging alien species will certainly change at finer spatial resolution. So far, the proportions of emerging alien species have declined only recently and moderately in most taxonomic groups. This shows that the introduction of new alien species is still ongoing at high rates, and that we can expect many more invasions in the future with large and emerging economies being likely starting points of future spread (SI Appendix, Fig. S4). Furthermore, which aliens are the primary reason of establishment and mitigation of further spread, often relies on warning lists based on information of species that are alien elsewhere (15, 16). This is particularly relevant for horizon scanning studies that aim to identify door-knocker species, which are those not yet recorded but suspected to have a high risk of arrival and impacts (17). Consequently, emerging alien species pose a particular challenge to biosecurity as they have no invasion history elsewhere, and their identities and potential impacts are difficult to predict. These species therefore may have higher chances to slip through border controls and elude early response management.

Materials and Methods

First Record Database. This analysis is based on a global dataset of first records of alien species that have become established in one or more mainland or island regions (1). The regions largely correspond to countries, while large islands belonging politically to a mainland country but located in biogeographically different areas or with extensive independent samples such as Hawaii, Galapagos, Azores, or Puerto Rico are considered as different regions. The delineations of the regions were obtained from the Global Administrative Areas database and we supplemented this database with information about islands from ref. 35. A first record in a region in our database is either for a species that had already established an alien population elsewhere or for a species that was never before recorded as alien anywhere in the world. We define the latter as emerging alien species. Note that all alien species count as “emerging” once (for their earliest record in our dataset). Compared with a previous study (1) the first-record database was updated and revised, now including in total 48,611 first records (+6%) from 17,130 established alien species (+1%) in 276 regions.

The first records were compiled from >100 different sources including online databases, published articles and books, and personal collections, which are listed in SI Appendix, Table S1. We adopted the categorization of the invasion status of alien species (casual/established) if provided in the original data source. If the invasion status was not provided, we considered the first record to be from an established alien species sensu ref. 36 as this is the most common status reported. We admit that this approach may lead to an overestimation in established alien species; however, the main findings of this analysis are robust to changes in the number of first records considered as shown by the sensitivity analysis.

First records from the original data sources were assigned to specific regions in the first record database. This was not possible for one dataset of alien insects encompassing first records for the combined regions of the United States and Canada. Comparing lists of alien insects in the United States (37) and Canada (38) revealed that roughly one-third (32%) of all alien insects reported for the combined region were found only in Canada and two-thirds (68%) only in the United States. We therefore randomly assigned two-thirds of the first records (n = 1,905) to the United States and one-third (n = 953) to Canada. While this may result in misspecified alien insects for the United States and Canada, this does not affect the continental analyses. The final first-record database was compiled to analyze large-scale temporal trends of alien species accumulation. For detailed information more specific databases and publications should be consulted such as refs. 14 and 39–42 or those listed in SI Appendix, Table S1.

The analysis was restricted to first records from eight major taxonomic groups with a sufficient number of first records and only included records up to 2005 to account for delays in reporting alien species records into databases. This resulted in 45,984 first records of 16,019 established alien species across 270 regions worldwide (SI Appendix, Fig. S7). In this dataset, most first records are for vascular plants (53% of all records), followed by insects (26%), birds (6%), fishes (4%), other invertebrates (3%), mammals (3%), molluscs (2%), and crustaceans (2%). The geographic distribution of first records is biased toward Europe (39% of all first records), followed by North America (24% + 6% Australasia + Pacific Islands; 20%), Asia (9%), and South America (including Central America, 5%), and Antarctica (0.3%). As most regions considered here refer to countries, the distribution of first records is affected by the distribution of country sizes worldwide. Using only the first record of a species on a continent, thereby removing multiple records of a species on that continent, revealed slightly different proportions [Europe: 27% (± 12 percentage points); North America: 26% (+6); Australasia + Pacific Islands: 24% (+4); Asia: 9% (+0); Southern America: 8% (+3); Africa: 6% (+0); and Antarctica: 0.3% (+0)].

Model to Estimate Source Pools. To investigate the influence of the native species pools on the proportions of emerging alien species we established a simple model of invasion dynamics. The model simulates the spread of individuals from a candidate species pool of unknown size X to a new region, thereby estimating the size of the candidate alien species pool as a function of emerging among all alien species. This approach is similar to those applied to estimate the total number of species on Earth, using the rate of description of new native species (20). In the candidate species pool we assume that species abundances are log-normally distributed [log (mean) = –2 and log (SD) = 1], which is a common way to describe the distribution of species in natural communities, and that each individual has the same probability of being introduced and establishing in the alien range. In a first step, the size of this candidate species pool remained constant within the simulation time. At each time step (here the year t), we randomly selected S species with probability according to the log-normal distribution from the candidate species pool X and placed them into a new range where they were alien. S, corresponds to the number of observed first records at year t, which was obtained from the time series of first-record rates of all alien species (upper panel for each taxon in Fig. 2). Each introduced species is considered to be able to establish an alien population in a region. Thus, the new range, where the species is alien, is large and suitable enough to allow the establishment of all introduced alien species. A species may be selected multiple times from the candidate species pool, which reflects the ongoing process of invasion into different regions. Emerging alien species were determined as the first occurrence of that species in the alien range. The numbers of first records of all alien species and those of emerging alien species were recorded. To obtain the size of X, the simulation was repeated 100 times and the resulting predicted average time series of emerging alien species was fitted to the observed time
series of emerging alien species. The deviation between predicted and observed values was measured as the root-mean-squared error (RMSE) between both time series. Fitting was done using the Nelder–Mead optimization algorithm implemented in the optim function of the base R language (43), which tries to find a parameter set (here only X) minimizing RMSE. The optimization was performed several times for each taxonomic group with different initial parameter settings to ensure not being trapped in a local minimum in the fit landscape. In the first approach the only fitted parameter was the size of the candidate species pool X. In a second step the same model was applied in the same way, but now the candidate species pool X was allowed to vary with time t. As we had no prior knowledge about the functional form of Xt, we used a very flexible function, thereby only defining four knots at certain times (two at the years 1000 and 2005, respectively, and another two in-between), which can be of any positive value. The knot at year 1000 was included to allow species with a known first record before 1500 to establish. The knots at 1000 and 2005 were fixed in time and represent the boundaries of the simulation period. Thus, six parameters have to be fitted in total: the size of the candidate species pools at the four knots and the timing of the two intermediate knots within the boundaries of 1000 and 2005. Between these knots, Xt was linearly interpolated to obtain a continuous function for the full time period. We applied this approach to yield a flexible function, which at the same time resulted in the convergence of the optimization algorithm to a meaningful solution. We also tested other functional forms such as piecewise linear regression functions or functions with fewer or more knots, which, however, did not improve the fits. Fitting was done in the same way as described above.

From the model results the maximum size of the candidate species pool was determined and the proportion of already established alien species in the first record database was calculated (Table 1). As the first record database did not include first records for all invasion events the size of the candidate species pool is underestimated. We therefore collated total alien species numbers from the literature and online databases if available and calculated the full candidate species pool assuming the same relationship between candidate species pool and established alien species as observed for the species considered in this study. We compare these results with estimates of the total number of native species on Earth for the various taxonomic groups (21).

Data on Drivers of First-Record Rates. The explanatory variables needed to be regional time series with a global coverage spanning at least one century to ensure a sufficient number samples per continent and taxonomic group. Only a few datasets of potential drivers fulfilled these requirements, and thus the analysis of drivers was inevitably limited by data availability. We considered eight drivers in total: three measures of temporal change in land use (proportion of urban area, pasture, and cropland), two of climatic conditions (annual mean temperature and precipitation), total imports of trade, human population sizes, and the number of botanical gardens. All variables were extracted for each region and time period of 5 or 2 y, respectively, depending on the availability (discussed below). Note that not all data were available for all regions and times.

The land-use data were obtained from the dataset Harmonized Global Land Use for Years 1500–2100, V1 (daac.ornl.gov/cgi-bin/dsviewer.pl?ds_id=1248) (44), which is provided by NASA’s Earthdata service (https://earthdata.nasa.gov). These data represent proportions of land cover annually for 1500–2100 at 0.5° spatial resolution. For each region and time period of 5 y the mean proportion of the respective variable was calculated. Historic environmental data were taken from the Twentieth Century Reanalysis project provided by the National Oceanic and Atmospheric Administration/Oceanic and Atmospheric Research/Earth System Research Laboratory Physical Sciences Division, Boulder, CO (www.esrl.noaa.gov/psd/). These datasets contain monthly averages of environmental variables from 1871 to 2012 at a spatial resolution of 2° latitude and longitude. We extracted the near-surface air temperature and relative humidity as indicators for climatic conditions and calculated averages for each region and time period. Import values were obtained from the Correlates of War project (45), providing bilateral trade values exchanged between countries during 1870–2009. The number of countries with available trade data increased over the course of time and consequently more trade data are available in recent times. The consideration of trade in the analysis reduced the total number of regions, and thus the sample size of the analysis. Human population densities were obtained from the HYDE 3.1 (46) database (thematics.pbl.nl/tridion/en/themasites/hyde). Human population densities were mostly available for decades from 1800 to 2005 at a spatial resolution of 5 min, which were summed to get human population sizes for each region. To get a common temporal resolution for all explanatory variables we linearly interpolated the time series of population sizes to a 5- or 2-y resolution, respectively. This arbitrarily increased the sample size of human population size, which may confound the statistical analysis. However, the original data were a very smooth time series and thus the interpolation should not affect the variability of the dataset and the predictive power of the variable. From Botanic Gardens Conservation International (www.bcgi.org) we obtained the year of foundation of 1,571 botanical gardens during 1800–2005 worldwide, which were attributed to the respective regions and time period. The number of botanical gardens can only be a very rough proxy for the influence on plant invasions, because, for example, sizes of botanical gardens or planted species are not available. Human population sizes and import values were log-transformed and all data were rescaled to a mean of zero and an SD of one before the regression analysis.

Statistical Analysis of Drivers of First-Record Rates. In a previous analysis (1) we detected distinct variation in the time series of first-record rates among continents. Thus, we analyzed temporal dynamics in first records by continent, such that a species could now be an emerging alien species multiple times, once on each continent where it is not native. The analysis was carried out on time series of first records from 1870 to 2005 due to the availability of explanatory variables for emerging alien species and compared with nonemerging alien species, which represent first records of alien species already known from other sites. An analysis of the first-record rates is not straightforward as several statistical challenges have to be overcome: (i) the relationships of first-record rates to explanatory variables can be expected to be nonlinear as indicated in previous studies (1, 47), (ii) first-record rates are not normally distributed, (iii) many time series of first records are nonstationary, (iv) the statistics of the remaining few regions where some areas were sampled more intensively than others, (v) the first-record rates are spatially and (vi) temporally autocorrelated, and (viii) distinct variation in, for example, sampling intensity and number of regions, but also in the temporal development of first-record rates, exists among continents.

i) To account for nonlinearity, we applied GAMMs (48), which represent a powerful and flexible way of regression analysis. A GAMM is a generalized linear mixed model using smoothed splines fitted to the explanatory variables rather than the original values of the variables. The degree of nonlinearity is estimated by generalized cross-validation, which always prefers “smoother” relationships over more complex ones. That is, linear relationships are preferred over nonlinear ones. We used the function gamm4 from the package gamm4 (49) in the R language (43) to perform the analysis. In gamm4, parameters are estimated using maximum likelihood and thus model fits of nested models can be compared using the Akaike information criterion (AIC). In comparison with the widely applied generalized linear mixed models, in- terpretation of the resulting smooths is based on visual inspection of diagnostic plots. Following recommendations (50), we analyzed diagnostic plots (not shown) of fitted values, residuals, and the fitted splines for each term of each GAMM to assess the goodness of fits. gamm4 allows the application of “shrinking smoothers,” so that strong-enough penalization will shrink the coefficients of the smoothers to near zero. This automatically removes less plausible information available from the model without the need to refit a nested version of the full model. Hence, model selection is done in one step with all explanatory variables included in the model. Here, we used the shrinkage version of cubic regression splines implemented in gamm4 for each single predictor variable. To test for potential interactions of temperature and relative humidity we added a tensor product smooth, which represents a 2D spline fit, of both variables to the GAMMs. As the tensor product smooth never improved the model fit we do not show this result here. The importance of one explanatory variable for the fit is expressed by a leave-one-out cross-validation approach, thereby comparing the full model with a nested model without the predictor under consideration using AIC. We tested for significant improvements of the models by applying a likelihood ratio test on the full and the nested model.

ii) First records represent count data, which are most appropriately modeled using a Poisson-distributed GAMM with a canonical log link function.

iii) For some taxonomic groups and continents the number of first records is rather low either because of lower sampling intensity or lower number of alien species. This may result in incomplete and noisy time series. To avoid the bias due to the low sample size we aggregated the number of first records to time intervals. This reduces the sample size, which can also affect the estimation of regression coefficients. We therefore attempted to find a balance between the width of the time interval...
and the sample size. An aggregation of first records to a time interval of 5 y resulted in a sample size of >100 for most taxonomic groups. This waning process is reflected in the estimated number of first records. For example, the total number of first records for 5 y from 1500 and 5 y from a first record of 2005. This resulted in much earlier first records centuries ago compared with recent first records and should remove the recording lag due to variation in sampling intensity. This procedure was repeated 100 times, and the means and SEMs of first-record rates were calculated (red lines in SI Appendix, Fig. S2). As mentioned above, the time series of recording lags is purely arbitrary, but we believe that it is a reasonable assumption. Furthermore, the exact parameter choice for the rate of decline, the maximum time lags, or the exact shape of the function shown in Si Appendix, Fig. SBA had only minor effects on the results. For example, using a linear instead of an exponential function reduces the observed differences between first records before and after 1900, but the overall patterns of the time series remain similar.

ii) Second, to account for the potentially increasing number of alien species detected in recent times due to intensified sampling we randomly removed a proportion of first records from the dataset. We again assumed an exponentially increasing sampling intensity with time, and thus the proportion of removed first records increased likewise from zero in 1500–50% in 2005 (SI Appendix, Fig. SBA). This analysis was repeated 100 times, and the means and SEMs of recorded first-record rates were determined (blue lines in SI Appendix, Fig. S2).

The modifications of first records in the sensitivity analysis were substantial, with up to 50% of all first records being removed and a misclassification of first records up to 100 y. Such distinct modifications helped clarify how long-term trends of first-record rates may be affected by temporal variation in sampling intensity. However, the resulting time series of modified first-record rates should not be considered to represent the actual rates as we do not know the true changes in sampling intensities and the consequences for first-record rates. The results of this sensitivity analysis only allow us to draw general conclusions about the robustness of the results.

ACKNOWLEDGMENTS. We thank the Comisión Nacional para el Conocimiento y Uso de la Biodiversidad team, David Baxter, Yves Frenot, Echekhaj Jäger, Wojciech Solarz, Agnes Dellinger, Stephan Gallisch, Marc Lebovici, and Zuzana Sixtova, for providing data of first records or contacts to data providers; Johan Liakka for assistance in climate data acquisition; and the DFG Grant FZT 118 (to M.W.), the Federal Ministry of Education and Research, Germany, and the Leverhulme Trust (RF/2/RFG/2010/0016) and a postdoc 3120125 (N.F.), for their help in archiving data; Bob O’Hara for statistical advice; and two anonymous reviewers for their helpful comments. First records of plants in Greece were derived from unpublished data of M. Arianoutsou, Ioannis Bazos, Pinelopi Delipetrou, Yannis Kokkoris, and Andreas Zikos. The study was supported by Deutsche Forschungsgemeinschaft (DFG) Grants SE 1891/2-1 (to H.S.) and KL 18669-1 (to M.V.K.), Invasion Dynamics Network DFG Grant JE 288B-1, Austrian Science Fund Grant I20BE-816 (to H.S., F.E., and S.D.), COST Action C1209 “Allen Challenge” (S.B., F.E., R.S., H.E.R., M. Arianoutsou, A.R., M.V.K., S.S.S., W.N., and I.K.), the Galapagos Conservancy (C.E.C. and H.J.H., contribution 2160 of the Charles Darwin Foundation for the Galapagos Islands), The Czech Academy of Sciences (research development project RVO 67960236 and Praemium Academiae award support to P.P.), and Czech Science Foundation Projects 14-36079G, Centre of Excellence PLADIAS and P505/11/1112 (P.P., J.P., and K.S.), New Zealand Ministry of Business, Innovation and Employment (MBIE) core funding (CO4X1104) to Scon and the “Better Border Biosecurity” collaboration (E.G.B.), the Polish National Science Centre (B.T.G.), the Portuguese Foundation for Science and Technology and Programa Operacional Potencial Humano/Fundo Social Europeu Grants SFRH/BPD/84422/2010 and GHTD – UID/Multi/04413/2013 (to C.C.), Fondecyt postdoc 3120125 (N.F.), the Leverhulme Trust (Rt/FZ/RGF/2010/0016) and a King Saud University Distinguished Scientist Research Fellowship (both to T.M.B.), the Federal Ministry of Education and Research, Germany project “The Americas as Space of Entanglements” (A.M.), Defra in compiling first records through the GB Non-Native Species Information Portal (H.E.R.), the German Centre for Integrative Biodiversity Research (iDiv) Halle-Jena-Leipzig Grant DFG FZT 118 (to M.W.), subsidy funding to Okinawa Institute of Science and Technology Graduate University (E.P.E.), MBIE to Landcare Research within the “Managing Invasives” Portfolio (D.W.S.). The support for the Twentieth Century Reanalysis Project dataset is provided by the US Department of Energy, Office of Science Innovative and Novel Computational Impact on Theory and Experiment program and Office of Biological and Environmental Research and by the National Oceanic and Atmospheric Administration Climate Program Office.

Sensitivity Analysis on Sampling Intensity. It is likely that the sampling intensity of alien species has increased during recent centuries, with a peculiarly high intensity in most recent decades. This may affect the analyses of long-term trends of first records, which we addressed in a sensitivity analysis. A direct measure of temporal changes in sampling intensity is not available. Known approaches like the consideration of herbivaria sampling intensity (S1) or expert judgments (S2) as a proxy for sampling intensity are useful to address variation in recent sampling intensity, but this approach is not applicable over several centuries. We therefore performed a sensitivity analysis by modifying the first records arbitrarily purely based on theoretical considerations. We identified two major consequences of and increased sampling intensity: A first record moving forward in time and an earlier discovery of a new alien species and thus in earlier first records and (ii) in more alien species recorded. We therefore performed two sensitivity analyses.

i) In nearly all cases of first records there is a lag period between the actual introduction of an alien species and the record of its first occurrence. Assuming that an intensification of sampling will result in an earlier detection of a new alien species the recording lag should decrease with time. We therefore assumed that the maximum recording lag decreased exponentially from 100 y in 1500 to 5 y in 2005 (S1 Appendix, Fig. SBA). For each first record, we randomly selected a recording lag between zero and the maximum recording lag at that time according to the relationship shown in S1 Appendix, Fig. SBA and subtracted it from the respective first-record year. For example, a first-record year of 1500 was subtracted from a recording lag of 50 y from 1500 and 5 y from a first record of 2005. This resulted in much earlier first records centuries ago compared with recent first records and should remove the recording lag due to variation in sampling intensity. This procedure was repeated 100 times, and the means and SEMs of first-record rates were calculated (red lines in S1 Appendix, Fig. S2). As mentioned above, the time series of recording lags is purely arbitrary, but we believe that it is a reasonable assumption. Furthermore, the exact parameter choice for the rate of decline, the maximum time lags, or the exact shape of the function shown in S1 Appendix, Fig. SBA had only minor effects on the results. For example, using a linear instead of an exponential function reduces the observed differences between first records before and after 1900, but the overall patterns of the time series remain similar.

ii) Second, to account for the potentially increasing number of alien species detected in recent times due to intensified sampling we randomly removed a proportion of first records from the dataset. We again assumed an exponentially increasing sampling intensity with time, and thus the proportion of removed first records increased likewise from zero in 1500–50% in 2005 (S1 Appendix, Fig. SBA). This analysis was repeated 100 times, and the means and SEMs of first-record rates were determined (blue lines in S1 Appendix, Fig. S2).

The modifications of first records in the sensitivity analysis were substantial, with up to 50% of all first records being removed and a misclassification of first records up to 100 y. Such distinct modifications helped clarify how long-term trends of first-record rates may be affected by temporal variation in sampling intensity. However, the resulting time series of modified first-record rates should not be considered to represent the actual rates as we do not know the true changes in sampling intensities and the consequences for first-record rates. The results of this sensitivity analysis only allow us to draw general conclusions about the robustness of the results.


