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

Hanno Seebens[1,2]*, Tim M. Blackburn[3,4,5], Ellie E. Dyer[3,4], Piero Genovesi[6,7], Philip E. Hulme[8], Jonathan M. Jeschke[9,10,11], Shyama Pagad[12], Petr Pyšek[13,14], Mark van Kleunen[15,16], Marten Winter[17], Michael Ansong[18], Margarita Arianoutsou[19], Sven Bacher[20], Bernd Blasius[21], Eckehard G. Brockerhoff[22], Giuseppe Brundu[23], César Capinha[24,25], Charlotte E. Causton[26], Laura Celesti-Grapow[27], Wayne Dawson[28], Stefan Dullinger[2], Evan P. Economo[29], Nicol Fuentes[30], Benoit Guénard[31], Heinke Jäger[26], John Kartesz[32], Marc Kenis[33], Ingolf Kühn[34,35,17], Bernd Lenzner[2], Andrew Liebhold[36], Alexander Mosena[37], Dietmar Moser[2], Wolfgang Nentwig[38], Misako Nishino[32], David Pearman[39], Jan Pergl[13], Wolfgang Rabitsch[40], Julissa Rojas-Sandoval[41], Alain Roques[42], Stephanie Rorke[43], Silvia Rossinelli[20], Helen E. Roy[43], Riccardo Scalera[44], Stefan Schindler[2], Kateřina Štajerová[13,14], Barbara Tokarska-Guzik[45], Kevin Walker[39], Darren F. Ward[46], Takehiko Yamanaka[47] & Franz Essl[2,6,*]

Affiliations:

[1]Senckenberg Biodiversity and Climate Research Centre (BiK-F), Senckenberganlage 25, 60325 Frankfurt am Main, Germany

[2]Department of Botany and Biodiversity Research, University of Vienna, Rennweg 14, 1030 Vienna, Austria

[3]Department of Genetics, Evolution and Environment, Centre for Biodiversity and Environment Research, University College London, Gower Street, London WC1E 6BT, UK

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

[5]Centre for Invasion Biology, Department of Botany and Zoology, Stellenbosch University, Stellenbosch, South Africa

[6]Institute for Environmental Protection and Research (ISPRA), Via Vitaliano Brancati 48, 00144 Rome, Italy

[7]Chair IUCN Species Survival Commission Invasive Species Specialist Group (ISSG), 00144 Rome, Italy

[8]Bio-Protection Research Centre, Lincoln University, Christchurch 7648, New Zealand

[9]Leibniz-Institute of Freshwater Ecology and Inland Fisheries (IGB), 12587 Berlin, Germany

[10]Institute of Biology at the Department of Biology, Chemistry, Pharmacy, Freie Universität Berlin, 14195 Berlin, Germany

[11]Berlin-Brandenburg Institute of Advanced Biodiversity Research (BBIB), 14195 Berlin, Germany

[12]IUCN Species Survival Commission Invasive Species Specialist Group (ISSG), University of Auckland, New Zealand, 1072

[13] The Czech Academy of Sciences, Institute of Botany, Department of Invasion Ecology, CZ-252 43 Průhonice, Czech Republic

[14]Department of Ecology, Faculty of Science, Charles University, Viničná 7, CZ-128 44 Prague, Czech Republic

[15]Ecology, Department of Biology, University of Konstanz, Universitätsstrasse 10, 78457 Konstanz, Germany

[16]Zhejiang Provincial Key Laboratory of Plant Evolutionary Ecology and Conservation, Taizhou University, Taizhou 318000, China

[17]German Centre for Integrative Biodiversity Research (iDiv), Halle-Jena-Leipzig, Deutscher Platz 5e, 04103Leipzig, Germany

[18]Department of Silviculture and Forest Management, Faculty of Renewable Natural Resource, KNUST, Ghana

[19]Department of Ecology and Systematics, Faculty of Biology, National and Kapodistrian University of Athens, 15784 Greece

[20]Department of Biology, University of Fribourg, Chemin du Musée 10, 1700 Fribourg, Switzerland

[21]Institute for Chemistry and Biology of the Marine Environment, University of Oldenburg, Carl-von-Ossietzky Straße 9-11, 26111 Oldenburg, Germany

[22]Scion (New Zealand Forest Research Institute), PO Box 29237, Christchurch 8440, New Zealand

[23]Department of Agriculture, University of Sassari, Viale Italia 39, 07100 Sassari, Italy

[24]CIBIO/InBIO, Centro de Investigação em Biodiversidade e Recursos Genéticos, Cátedra Infraestruturas de Portugal-Biodiversidade, Universidade do Porto, Campus Agrário de Vairão, 4485-661 Vairão, Portugal

[25]Global Health and Tropical Medicine, GHTM, Instituto de Higiene e Medicina Tropical IHMT. Universidade Nova de Lisboa, UNL. Rua da Junqueira 100, 1349-008 Lisboa, Portugal.

[26]Charles Darwin Foundation for the Galapagos Islands, Puerto Ayora, Santa Cruz, Galapagos, Ecuador

[27]Department of Environmental Biology, Sapienza University, p. Moro, 5, 00185 Rome, Italy

[28]School of Biological and Biomedical Sciences, Durham University, South Road, Durham, DH1 3LE, United Kingdom

[29]Okinawa Institute of Science and Technology Graduate University, 1919-1 Tancha, Onna, Okinawa, 904-0495, Japan

[30]Departamento de Botánica, Facultad de Ciencias Naturales y Oceanográficas, Universidad de Concepción, Casilla 160-C, Concepción, Chile

[31]School of Biological Sciences, The University of Hong Kong, Kadoorie Biological Sciences Building, Pok Fu Lam Road, Hong Kong SAR, China

[32]Biota of North America Program (BONAP), Chapel Hill, North Carolina 27516, USA

[33]CABI, Rue des Grillons 1, 2800 Delémont, Switzerland

[34]Helmholtz Centre for Environmental Research - UFZ, Department of Community Ecology, Theodor-Lieser-Str. 4, 06120 Halle, Germany

[35]Department of Geobotany and Botanical Garden, Martin Luther University of Halle-Wittenberg, Am Kirchweg 2, 06108 Halle, Germany

[36]US Forest Service Northern Research Station, Morgantown, WV 26505, USA

[37]Department of Experimental and Systems Ecology & Center for Interamerican Studies (CIAS), Bielefeld University, Universitätsstraße 25, 33615 Bielefeld, Germany

[38]Institute of Ecology and Evolution, University of Bern, Baltzerstrasse 6, CH-3012 Bern, Switzerland

[39]Botanical Society of Britain and Ireland (BSBI) Suite 14, Bridge House, 1-2 Station Bridge, Harrogate, HG1 1SS, UK

[40]Department of Biological Diversity and Nature Conservation, Environment Agency, Spittelauer Laende 5, 1090 Vienna, Austria

[41]Department of Botany, National Museum of Natural History, MRC-166 Smithsonian Institution, P.O. Box 37012, Washington, DC 20013, USA

[42]Institut National de la Recherche Agronomique (INRA), UR 0633, Zoologie Forestière, 45075 Orléans, France

[43]Centre for Ecology & Hydrology, Wallingford, OX10 8BB, United Kingdom

[44]IUCN Species Survival Commission Invasive Species Specialist Group (ISSG), Rome, Italy

[45]Department of Botany and Nature Protection, Faculty of Biology and Environmental Protection, University of Silesia, 40-032 Katowice, Poland

[46]School of Biological Sciences, University of Auckland, PB 92019, Auckland, New Zealand Landcare Research, PB 92170, Auckland, New Zealand

[47]Institute for Agro-Environmental Sciences, NARO (NIAES), 3-1-3 Kannondai, Tsukuba, 305-8604, Japan

*Correspondence to: hanno.seebens@senckenberg.de and franz.essl@univie.ac.at

Abstract

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 statement

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 regional 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.

\body

Thousands of species have been introduced to regions outside their native ranges by humans, and many have become permanent additions to local faunas and floras. The number of these established alien species has strongly increased worldwide during the past two centuries (1), thereby redefining the classical boundaries of biogeography (2), affecting ecosystem functioning (3), human health (4), and economies (5, 6). Alien species have thus emerged as a defining feature of the Anthropocene (7).

The observed growth in alien species numbers have been largely attributed to increases in drivers of alien species introductions, such as import volumes and human mobility, and rising establishment rates due to land-degradation (8–12). However, alien species numbers in a region may also be affected by changes in the accessibility of source pools of species in the native range (13). For example, it has been shown that 'historical' alien bird introductions (1500 – 1903AD) were largely driven by European colonial expansion and thus mostly drawn from birds originating in Europe and European colonies, whereas 'modern' bird introductions (1983 – 2000) primarily relate to introductions via the pet bird trade, and concern species native to regions close to key trade hubs (14). These new source pools provide many new potential alien species when old source pools start to deplete (the depletion of source pools refers to the proportion of new alien species in that pool, which declines with every newly selected species), thereby maintaining the rate of alien species establishments in new regions. Disentangling the factors underpinning the accumulation of alien species will improve our understanding of past invasion dynamics and result in better-informed predictions of future trajectories of alien species accumulation.

To dissect the drivers of emerging alien species, we analysed a global database of 45,984 regional first records of 16,019 established alien species from most major taxonomic groups (vascular plants, mammals, birds, fishes, insects, crustaceans, molluscs, and other invertebrates) (1) during recent centuries (1500 – 2005AD). A first record constitutes the year of first detection of an alien species that later has become established in a region (usually a country or an island). From this database, we determined the first records of species appearing for the first time as alien worldwide, 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,

9

which are species not yet recorded but suspected to have a high risk of arrival and impacts (17).

Our study specifically addresses the following 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 spatio-temporal 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 (*Columba 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 non-emerging 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 analyse the influence of a depleting candidate species pool, we analysed 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 (see 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 ($X_t = X$, for all t), this model already captures a large amount of

11

the observed variability in the frequency of emerging alien species, thereby supporting the idea of a depleting candidate species pool (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_t, 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 R^2 values of 0.8 or higher for six taxonomic groups. Crucially, the candidate species pool predicted by the model increased over time (lower panels in Fig. 2), 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 (see Methods) are (Table 1): 425 species (crustaceans), 539 species (molluscs), 890 species (mammals), 1,494 species (birds), 2,697 species (fishes), and 47,029 (vascular plants). These estimated full candidate species pools represent between 1% (molluscs) to 16% (mammals) of the estimated total number of species in each taxonomic

group on Earth (21). As simulations revealed a consistent increase in candidate species pools during the 20th century, these pools can be expected to be still larger yet.

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 prior to the 20th century and lower numbers after ca. 1900 (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 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 (Fig. S4, see 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 (Fig. S6).

Drivers of temporal dynamics of first records

To analyse 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 (GAMM) to model temporal variation in first records of emerging and non-emerging alien species (the latter representing first records of species with already known occurrences elsewhere), thereby accounting for non-linearity, 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 (fraction of urban areas, pasture or cropland) varied among taxonomic groups. Variation in climatic 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 towards a general overview of the most important drivers of temporal variation in first records of alien species.

Discussion

Our analysis reveals novel insights into emerging alien species dynamics. It builds on a previous study (1) by analysing emerging alien species separately, to remove the effect of multiple first records of the same species due to e.g. 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 century, 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 (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 years, 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).

The occurrence of a high proportion (Fig. S4) and number (Fig. S6) 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 the hotspots of predicted future increases in invasion threats being mostly located in Europe, North America and East Asia (24). 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 have been important as well (25). 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, 26). 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 significantly related to first-record rates of emerging alien species, which supports previous findings of the role of botanical gardens for the introduction of alien vascular plants (27). 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 acclimatisation societies (28), European explorers or settlers (29), and plant hunters (30) are largely lacking, which highlights the need to improve the availability of historical data for more detailed analyses of spatio-temporal 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) (31), and the emergence of new introduction pathways (e.g., fashion trends, 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 e.g. 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 numbers 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 wide-spread alien distributions are comparatively rare. This, however, also depends on the resolution of the analysis, and the relationship between emerging and non-emerging 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 (Fig. S4). Biosecurity, which aims at the prevention 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 data set 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 (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 data set). Compared to 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 Table S2. We adopted the categorisation 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 (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 data set of alien insects encompassing first records for the combined regions of USA and Canada. Comparing lists of alien insects in the USA (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 USA. We therefore randomly assigned two third of the first records (n=1905) to the USA and one third (n=953) to Canada. While this may result in misspecified alien insects for the USA and Canada, this did not affect the continental and global analyses. Note that the first record database was compiled to analyse large-scale temporal trends of alien species accumulation. For detailed information more specific databases and publications should be consulted such as (14, 39–42) or those listed in Table S2.

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 (Fig. S7). In this data set, 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 towards Europe (39% of all first records), followed by North America (20%), Australasia + Pacific Islands (20%), Asia (9%), Africa (6%), Southern 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); 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 using the ratio 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 (standard deviation) = 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_t 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_t 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 (top 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 optimisation algorithm implemented in the 'optim' function of the base R language (43), which tries to find a parameter set (here only X) minimising RMSE. The optimisation 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_t was allowed to vary with time t. As we had no prior knowledge about the functional form of X_t, 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. The locations of the other two knots were determined by the optimisation algorithm, thereby restricting them to lie within 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, X_t 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 optimisation algorithm to a meaningful solution. We also tested other functional forms such as piece-wise 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 pools 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 data sets 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 import values of trade, human population sizes, and the number of botanical gardens. All variables were extracted for each region and time period of five or two years, respectively, depending on the analysis (see below). Note that not all data were available for all regions and times.

The land-use data were obtained from the data set 'Harmonized Global Land Use for Years 1500–2100, V1' (http://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 five years, the mean proportion of the respective variable was calculated. Historic environmental data were taken from the Twentieth Century Reanalysis project provided by the NOAA/OAR/ESRL PSD, Boulder, Colorado, USA (http://www.esrl.noaa.gov/psd/). These data sets contain monthly averages of environmental variables from 1871 to 2012 at a spatial resolution of two degrees 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 (http://themasites.pbl.nl/tridion/en/themasites/hyde/). Human population densities were mostly available for decades from 1800 to 2005 at a spatial resolution of 5 minutes, 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 five or two year 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 data set and the predictive power of the variable. From 'Botanic Gardens Conservation International' (www.bgci.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 e.g. 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 a standard deviation of one prior to 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 analysed 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 1870–2005 due to the availability of explanatory variables for emerging alien species and compared to non-emerging alien species, which represents 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 non-linear as indicated in previous studies (1, 47), (ii) first-record rates are not normally distributed, (iii) many time series of first records have gaps, (iv) the statistical design is imbalanced as some regions were sampled more intensively than others, (v) the first-record rates are spatially, and (vi) temporally autocorrelated and (vii) distinct variation in e.g. sampling intensity, number of regions, but also in the temporal development of first-record rates exists among continents.

(i) To account for non-linearity, we applied Generalized Additive Mixed Models (GAMM) (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 non-linearity is estimated by generalized cross-validation, which always prefers 'smoother' relationships over more complex ones. That is, linear relationships are preferred over non-linear 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 to the widely applied Generalized Linear Mixed Models, interpretation of the results is less straightforward and mostly 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 'shrinkage smoothers', so that strong enough penalization will shrink the coefficients of the smoothers to near zero. This effectively removes less suitable explanatory variable 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 two-dimensional 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 model fit by applying a Likelihood Ratio Test on the full and the nested model.

(ii) First records represent count data, which are most appropriately modelled 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 achieve more complete time series, 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 five years resulted in a sample size of >100 for most taxonomic groups. This was found to produce robust estimations of coefficients. For some time series (non-emerging alien crustaceans, molluscs, and other invertebrates), the sample size was below 100. We therefore repeated the analysis for these groups with a time interval of two years, which increased the sample size. As the results did not differ distinctly, we only present the results for the 5-year intervals.

(iv) The number of first records varied distinctly among continents and taxonomic groups, which may result in an imbalanced design of the regression analysis with some combinations of continent and taxonomic group having very low numbers of first records compared to others. To avoid the analysis of these very incomplete time series, only continental time series of first-record rates that had at least 15 samples out of 28 possible samples for one time series (i.e., 28 possible five-year intervals during 1870–2005) were included in the analysis.

(v) The spatial autocorrelation of first-record rates among regions was addressed by aggregating the first records to the level of continents: spatial autocorrelation among continents can be assumed to be much lower compared to inter-regional dynamics. For this continental analysis, an emerging alien species was considered to be the first record of that alien species on that continent. Consequently, a species can be an emerging alien species multiple times, once on each continent where it has an alien population. Furthermore, continent was included in the GAMMs as a random effect variable, which addresses spatial autocorrelation among continents.

(vi) Time series of first-record rates are temporally autocorrelated, which may bias the calculation of regression coefficients. As 'gamm4' did not allow for the implementation of an autoregressive correlation structure, we included 'year' as an additional predictor in the GAMMs. This effectively removed temporal autocorrelation in the residuals, which was checked using the autocorrelation function 'acf' in basic R (43) applied to normalized residuals.

(vii) The distinct variations in the time series of first-record rates among continents was addressed by incorporating continent as a random effect variable in the GAMMs. This also

accounted for a potential systematic influence of the number of regions per continent or the total area of all regions of the same continent. In addition, we also tested to include these variables as an offset variable in the 'gamm4' function, which should remove this influence. However, this did not improve the model fit and thus an offset variable was not included in the final model.

The aforementioned methodology is complex and thus includes some opportunities to perform the same analysis in a slightly different way. However, our approach is appropriate for an analysis of data of this kind and the results are robust allowing general insights in the temporal development of first-record rates. Nevertheless, interpretation of the results should be done carefully, also because of the lack of data for some drivers of alien species introductions.

Sensitivity analysis on sampling intensity

Sensitivity analysis

It is likely that the sampling intensity of alien species has increased during recent centuries, with a particularly 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 herbaria sampling intensity (51) or expert judgements (52) 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 high sampling

intensity results (i) in an earlier detection 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 years in 1500 to five years in 2005 (Fig. S8A). 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 Fig. S8A, and subtracted it from the respective first record. For example, a maximum of 100 years was subtracted from a first record from 1500 and five years from a first record of 2005. This resulted in much earlier first records centuries ago compared to 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 standard errors of the means of first-record rates were calculated (red lines in 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 Fig. S8A 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 data set. We again assumed an exponentially increasing sampling intensity with

time, and thus the proportion of removed first records increased likewise from zero in 1500 to 50% in 2005 (Fig. S8B). This analysis was repeated 100 times, and the means and standard errors of the means of first-record rates were determined (blue lines in 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 misspecification of first records of up to 100 years. 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.

References

1. Seebens H, et al. (2017) No saturation in the accumulation of alien species worldwide. *Nat Commun* 8:14435.

2. Capinha C, Essl F, Seebens H, Moser D, Pereira HM (2015) The dispersal of alien species redefines biogeography in the Anthropocene. *Science* 348(6240):1248–1251.

3. Simberloff D, et al. (2013) Impacts of biological invasions: what's what and the way forward. *Trends Ecol Evol* 28(1):58–66.

4. Pyšek P, Richardson DM (2010) Invasive Species, Environmental Change and Management, and Health. *Annu Rev Environ Resour* 35(1):25–55.

5. Kettunen M, et al. (2009) *Technical support to EU strategy on invasive species (IAS) — Assessment of the impacts of IAS in Europe and the EU: Final report for the European Commission, Institute for European Environmental Policy (IEEP)* (Brussel, Belgium).

6. Vilà M, et al. (2010) How well do we understand the impacts of alien species on ecosystem services? A pan-European, cross-taxa assessment. *Front Ecol Environ* 8(3):135–144.

7. Lewis SL, Maslin MA (2015) Defining the Anthropocene. *Nature* 519(7542):171–180.

8. Hulme PE (2009) Trade, transport and trouble: managing invasive species pathways in an era of globalization. *J Appl Ecol* 46(1):10–18.

9. Levine JM, D'Antonio CM (2003) Forecasting biological invasions with increasing international trade. *Conserv Biol* 17(1):322–326.

10. Seebens H, Essl F, Blasius B (2017) The intermediate distance hypothesis of biotic invasions. *Ecol Lett* 20:158–165.

11.   Pyšek P, et al. (2010) Disentangling the role of environmental and human pressures on biological invasions across Europe. *Proc Natl Acad Sci USA* 107(27):12157–12162.

12.   Pauchard A, Alaback PB (2004) Influence of Elevation, Land Use, and Landscape Context on Patterns of Alien Plant Invasions along Roadsides in Protected Areas of South-Central Chile. *Conserv Biol* 18(1):238–248.

13.   Liebhold AM, Brockerhoff EG, Kimberley M (2017) Depletion of heterogeneous source species pools predicts future invasion rates. *J Appl Ecol* Early view. doi:10.1111/1365-2664.12895.

14.   Dyer EE, et al. (2017) The global distribution and drivers of alien bird species richness. *PLOS Biol* 15(1):e2000942.

15.   Leung B, et al. (2012) TEASIng apart alien species risk assessments: a framework for best practices. *Ecol Lett* 15(12):1475–93.

16.   European Plant Protection Organization (2015) *A1 and A2 Lists of Pests Recommended for Regulation as Quarantine Pests* Available at: http://archives.eppo.int/EPPOStandards/PM1_GENERAL/pm1-02(24)_A1A2_2015.pdf.

17.   Roy HE, et al. (2015) *Invasive Alien Species - Prioritising prevention efforts through horizon scanning* (ENV.B.2/ETU/2014/0016. European Commission).

18.   Pyšek P, et al. (2017) Naturalized alien flora of the world. *Preslia* 89(3):203–274.

19.   Dyer EE, et al. (2016) A global analysis of the determinants of alien geographical range size in birds. *Glob Ecol Biogeogr* 25(11):1346–1355.

20.   Joppa LN, Roberts DL, Pimm SL (2011) How many species of flowering plants are

there? *Proc R Soc B Biol Sci* 278(1705):554–559.

21. Chapman AD (2009) *Numbers of Living Species in Australia and the World* (Department of the Environment, Water, Heritage and the Arts, Canberra, Australia). 2nd Ed. Available at: http://www.environment.gov.au/biodiversity/abrs/publications/other/species-numbers/2009/06-references.html.

22. Blackburn TM, Dyer EE, Su S, Cassey P (2015) Long after the event, or four things we (should) know about bird invasions. *J Ornithol* 156(1):15–25.

23. Eaton JA, Leupen BTC, Krishnasamy K (2017) *Songsters of Singapore: An Overview of the Bird Species in Singapore Pet Shops* (Petaling Jaya, Selangor, Malaysia).

24. Early R, et al. (2016) Global threats from invasive alien species in the twenty-first century and national response capacities. *Nat Commun* 7:12485.

25. Petitpierre B, et al. (2012) Climatic Niche Shifts Are Rare Among Terrestrial Plant Invaders. *Science* 335(6074):1344–1348.

26. Bellard C, Leroy B, Thuiller W, Rysman JF, Courchamp F (2016) Major drivers of invasion risks throughout the world. *Ecosphere* 7(3):1–14.

27. Hulme PE (2011) Addressing the threat to biodiversity from botanic gardens. *Trends Ecol Evol* 26(4):168–174.

28. Pears N (1982) Familiar aliens: The acclimatisation societies' role in New Zealand's biogeography. *Scott Geogr Mag* 98(1):23–34.

29. di Castri F (1989) History of biological invasions with special emphasis on the Old World. *Biological Invasions: A Global Perspective*, eds Drake JA, et al. (John Wiley and

Sons, Chichester), pp 1–30.

30. Fry C (2013) *The Plant Hunters: The Adventures of the World's Greatest Botanical Explorers* (University of Chicago Press, Chicago).

31. Seebens H, et al. (2015) Global trade will accelerate plant invasions in emerging economies under climate change. *Glob Chang Biol* 21(11):4128–4140.

32. Essl F, et al. (2015) Crossing Frontiers in Tackling Pathways of Biological Invasions. *Bioscience* 65(8):769–782.

33. Casties I, Seebens H, Briski E (2016) Importance of geographic origin for invasion success: A case study of the North and Baltic Seas versus the Great Lakes–St. Lawrence River region. *Ecol Evol* 6(22):8318–8329.

34. Kalusová V, et al. (2017) Naturalization of European plants on other continents: The role of donor habitats. *Proc Natl Acad Sci* (December):1–6.

35. Weigelt P, Jetz W, Kreft H (2013) Bioclimatic and physical characterization of the world's islands. *Proc Natl Acad Sci USA* 110(38):15307–12.

36. Blackburn TM, et al. (2011) A proposed unified framework for biological invasions. *Trends Ecol Evol* 26(7):333–339.

37. Aukema JE, et al. (2010) Historical Accumulation of Nonindigenous Forest Pests in the Continental United States. *Bioscience* 60(11):886–897.

38. Nealis VG, et al. (2016) Historical occurrence of alien arthropods and pathogens on trees in Canada. *Canada J For Res* 180(November 2015):1–41.

39. Froese R, Pauly D (Eds) (2015) FishBase. *http://www.fishbase.de/ Accessed at 2017-03-03*.

40. van Kleunen M, et al. (2015) Global exchange and accumulation of non-native plants. *Nature* 525(7567):100–103.

41. Dawson W, et al. (2017) Global hotspots and correlates of alien species richness across taxonomic groups. *Nat Ecol Evol* 1(186):186.

42. Pagad S, Hayes K, Katsanevakis S, Costello MJ (2017) World Register of Introduced Marine Species (WRIMS). *http://www.marinespecies.org/ Accessed at 2017-12-07*.

43. R Development Core Team (2015) *R: A language and Environment for Statistical Computing* (https://www.r-project.org/, R Foundation for Statistical Computing, Vienna, Austria) Available at: https://www.r-project.org/.

44. Chini LP, Hurtt GC, Frolking S (2014) Harmonized Global Land Use for Years 1500 - 2100, V1. doi:10.3334/ORNLDAAC/1248.

45. Barbieri K, Keshk OMG (2012) Correlates of War Project Trade Data Set Codebook, Version 3.0. Available at: http://correlatesofwar.org.

46. Klein Goldewijk K, Beusen A, Van Drecht G, De Vos M (2011) The HYDE 3.1 spatially explicit database of human-induced global land-use change over the past 12,000 years. *Glob Ecol Biogeogr* 20(1):73–86.

47. Costello C, Springborn M, McAusland C, Solow A (2007) Unintended biological invasions: Does risk vary by trading partner? *J Environ Econ Manage* 54(3):262–276.

48. Wood SN (2006) *Generalized Additive Models: An Introduction with R* (Chapman & Hall/CRC, Boca Raton, Florida).

49. Wood SN, Scheipl F (2016) *gamm4: Generalized Additive Mixed Models using "mgcv" and "lme4"* (R package version 0.2-4) Available at: https://cran.r-

project.org/package=gamm4.

50.   Zuur AF, Ieno EN, Walker N, Saveliev AA, Smith GM (2009) *Mixed effects models and extensions in ecology with R* (Springer, New York) doi:10.1007/978-0-387-87458-6.

51.   Delisle F, Lavoie C, Jean M, Lachance D (2003) Reconstructing the spread of invasive plants: taking into account biases associated with herbarium specimens. *J Biogeogr* 30(7):1033–1042.

52.   Meyer C, Kreft H, Guralnick R, Jetz W (2015) Global priorities for an effective information basis of biodiversity distributions. *Nat Commun* 6:8221.

Figure legends

Fig. 1 Frequency distribution of the number of first records per alien species.

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 (top panels), emerging alien species (second panels), and the proportion of emerging among all alien species (third panels). 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 panels) 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 (see 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.

Fig. 3 Results of the statistical analysis of time series of first-record rates of emerging and non-emerging alien species using GAMMs for eight taxonomic groups. Non-emerging alien species denote first records of those alien species, which have already established somewhere else in the world, and are shown for comparison. The importance of a single predictor (x-axes) is indicated by the size of the dots corresponding to the difference in AIC

(ΔAIC) between the full model and the one without the parameter under consideration (see

Methods). Significant improvements of the model fit tested by a likelihood ratio test are

highlighted by black outer circles. The sample size is given in the first column and the

adjusted R^2 in the last column of each panel. A missing dot denotes that the respective

predictor did not enter the GAMM.