Predicting future invaders and future invasions

Alice Fournier,a Caterina Penoneb, Maria Grazia Pennino,c and Franck Courchampa,a,1

*Ecologie Systématicque Evolution, Univ Paris-Sud, CNRS, AgroParisTech, Université Paris-Saclay, 91400 Orsay, France; bInstitute of Plant Sciences, University of Bern, CH-3013 Bern, Switzerland; and cInstituto Español de Oceanografía, Centro Oceanográfico de Vigo, 36390 Vigo, Pontevedra, Spain

Edited by Daniel S. Simberloff, The University of Tennesseee, Knoxville, TN, and approved February 27, 2019 (received for review March 7, 2018)

Invasive alien species are a great threat to biodiversity and human livelihoods worldwide. The most effective way to limit their impacts and costs is to prevent their introduction into new areas. Identifying invaders and invasions before their occurrence would arguably be the most efficient strategy. Here, we provide a profiling method to predict which species—with which particular ecological characteristics—will invade, and where they could invade. We illustrate our approach with ants, which are among the most detrimental invasive species, as they are responsible for declines of numerous taxa, are involved in local extinctions, disturb ecosystem functioning, and impact multiple human activities. Based on statistical profiling of 1,002 ant species from an extensive trait database, we identify 13 native ant species with an ecological profile that matches that of known invasive ants. Even though they are not currently described as such, these species are likely to become the next global invaders. We couple these predictions with species distribution models to identify the regions most at risk from the invasion of these species: Florida and Central America, Brazil, Central Africa and Madagascar, Southeast Asia, Papua New Guinea Northeast Australia, and many islands worldwide. This framework, applicable to any other taxa, represents a remarkable opportunity to implement timely and specifically shaped proactive management strategies against biological invasions.

Invasive alien species (IAS) are one of the main drivers of biodiversity loss and disruptors of ecosystem functioning and services (1–4). In addition to ecological impacts, IAS also have substantial impacts on the economy and human health (5). With increasing international trade and changing climates (6, 7), increasing numbers of invasions—and their subsequent impacts—are expected worldwide (8–12).

The most cost-effective means to reduce the impacts of IAS is to prevent their importation or implement rapid response treatment programs, which rely on early detection (13–15). Consequently, there is a pressing need to find robust ways to predict invasions before their occurrence to adjust surveillance or management policies and thereby reduce the likelihood of establishment, spread, and overall cost (16). Identifying future invaders before they become so has been a longstanding and previously unreachable objective of invasion biology (17, 18).

Until now, species distribution models have forecasted the most suitable areas, and hence the areas most at risk from known IAS (7, 19, 20). Other types of models have tried to identify what characterizes IAS based on the drivers of their invasion success, their life-history traits, or their occurrences worldwide (e.g., ref. 18). Such models have been built for various taxa, including plants, ants, fish, amphibians, reptiles, birds, and mammals (21–30). For ants, studies based on various databases characterize certain ecological traits often associated with exotic or invasive species (28, 31, 32). All these pioneer studies provide a valuable starting point, but they still have fundamental limitations. First, they remain limited in terms of coverage; notably, all studies based on interception lists rely on only a few countries. Second, the studies exclusively consider the few species that have complete data for all of the studied traits, thus removing the majority of species. Removing species with missing data drastically reduces the sample size and statistical power and also introduces biases and thus improper predictions (33–36). Finally, the existing models are still mostly limited to species of already known invasive status (i.e., invasive or exotic) or to providing traits associated with invasiveness; thus, existing models do not use these traits to predict new invaders yet to move to another area. Nevertheless, this corpus of pioneer studies shows that, by using modern statistical inference tools, improved ecological knowledge on current IAS, and large databases on species ecological traits, it is now possible to reach a higher level of prediction in invasion biology and identify in advance which species could become invasive. Establishing a list of future potential IAS also allows us to predict where they are likely to invade and where special detection efforts should be deployed.

Ants represent an excellent opportunity to take the existing predictive tools a step further by including more traits, more species, and worldwide coverage. Invasive ants are among the worst IAS (37). They have enormous ecological impacts on other ant species as well as on other taxa (38, 39), ranging from decreases in the abundance and richness of native species (37) to the disruption of species interactions (40, 41), invasion meltdown (41, 42), local species extinctions (42), and whole ecosystem functioning alteration (39). Invasive ants also have tremendous negative impacts on human activities, notably agriculture (43), human health (44), and infrastructure (39). The costs of their damage reach billions of dollars annually; for example, the economic costs of the red imported fire ant alone have been estimated to reach $211 million/y in Hawaii and exceed $1.65 billion/y in Australia (16). Finally, invasive ants are extremely difficult to control (45); more than 200 species are established outside their native range (46), of which 19 are recorded as invasive (47). There are no fewer than five ant species in the International Union for the
Conservation of Nature (IUCN) list of the “100 of the world’s worst invasive alien species,” a feat equaled by no other taxonomic family (48). In this study, we will follow the definition of invasive species adopted by the Invasive Species Specialist Group of the IUCN, excluding species that are native but expanding, as well as exotic species with no impact.

Here, we have built a profiling tool that identifies which species have the ecological profile that makes them highly likely of being the next invading species. Our approach is to first define an ecological profile of the most invasive ant species and then compare it with the ecological profile of other species based on an extensive dataset of ant species traits, imputed through finely tuned and up-to-date machine-learning algorithms. Our quantitative risk assessment tool responds to two of the most crucial questions of invasion biology today: which species are likely to become invasive and where will they become invasive? This methodology is transposable to other taxa for which there is adequate ecological information on current invaders.

Results

Future Invaders. Our pool of ant species retained for the ecological profiling of invasive species was kept high on account of the setup imputation procedure, which enabled us to estimate missing values in the traits and thus have all trait values, either observed or estimated, for 1,002 species. This was done while keeping the imputation error very low (whole dataset imputation error: 5.97%), thus ensuring robust model estimates and predictions. We used this imputed dataset of four ecological traits to calculate the probability of species being invasive and detect ant species with a profile highly similar to known ant invaders. Our models enabled us to efficiently discriminate between likely and unlikely invasive species [Area Under the receiver operating characteristic Curve (AUC) = 0.95], and it indicated relatively strong relationships between species traits and invasion potential (pseudoR2 = 0.53). Our analyses suggest that the four most important traits for predicting ant invasiveness probability are the following: independent colony foundation mode, supercoloniality, generalist nesting type, and a frequent association with disturbances (SI Appendix, Table S3).

Our method also allowed us to distinguish two invasiveness profiles. Species with a probability > 0.70 (both known and predicted invasive species) are hereafter called “superinvasive.” From the 19 known invasive species, 14 have a superinvasive profile, with an invasiveness probability between 0.889 and 0.700 (mean: 0.837, Fig. 1). Four known invasive species have probabilities between 0.132 and 0.115 (mean: 0.157) and are hereafter called “invasive.” This bimodal pattern is likely linked to the colony structure of the species, as all superinvasive species are also supercolonial while all invasive species are not. Finally, one species recorded by the IUCN as invasive, Aenonymex octospinosus, has a significantly distinct profile, with an extremely low invasiveness probability of 0.155 × 10−3.

Our model also provided us with a list of 13 species that share the trait profile of current invasive species. Among the 13 species, 2 have a superinvasive profile (invasiveness probability between 0.828 and 0.838; mean: 0.830), while 11 have an invasive profile [probabilities from 0.111 to 0.267; mean: 0.163 (Table 1)]. Note that a leave-one-out sensitivity analysis [recoding known invasive species as noninvasive (SI Appendix)] classified all known invasive species as invasive in 100% of our models, thus showing that our models were reliable for predicting invasiveness (SI Appendix, Tables S4 and S5). The remainder of the 970 species that were not identified as invasive (according to the lower fifth centile of the invasive species probability distribution) in more than 90 models of 100 were not considered to be potential future invaders. Targeted studies on some of these species may, however, be relevant to confirm that they are at risk for becoming invaders in the future.

Fig. 1. Predicted invasiveness probability distributions for (1) ant species recognized as invasive by IUCN and (2) ant species for which no information on invasive status exists. Dark gray bars on the right correspond to superinvasive profiles, medium gray bars to invasive, and light gray bars to noninvasive. The asterisk (•) points to A. octospinosus, which the model shows to be wrongly classified as invasive by the IUCN. For the sake of clarity, the Bottom graph has two y axes: the light-gray axis corresponds to the light-gray bar of noninvasive species only.

Areas at Risk. Species distribution models allowed us to identify regions of climatic suitability for each of the 11 likely future invaders for which we had sufficient data (SI Appendix, Fig. S11). These predictions were of high quality: True Skill Statistics (TSS) values were between 0.88 and 0.97 and AUC values were between 0.96 and 0.99, indicating good model performances. The potential invasions are more likely in tropical regions, as illustrated in the cumulated map (Fig. 2). When overlapping all potential distributions, invasion hotspots (i.e., areas with more than eight invasive species) were predicted to occur in Florida and Central America, Brazil, Central Africa and Madagascar, Southeast Asia, Papua New Guinea, Northeast Australia, and many islands worldwide. Areas characterized by temperate and extreme climatic conditions (i.e., ice, hot desert, tundra), especially in the Paleartic and Nearctic (SI Appendix, Fig. S1) seemed less at risk from these novel ant invasive species. This was also the pattern observed for the two identified potential superinvasive species (Fig. 2), although Lepisiota canescens could also invade almost all of Africa and South East Asia, and Technomyrmex difficilis had a much more restricted potential distribution.

Discussion

Using a large database on ant species traits and an up-to-date phylogeny of ant genera, we developed a robust methodology to quantify the relative importance of ecological traits in shaping their invasiveness probability and used it to both characterize the ecological profile of current invasive ants and identify 13 noninvasive species that share this profile. We distinguished between invasive and highly invasive species based on their probability of invasiveness. We complemented this prediction by identifying regions most at risk for invasion from these species: tropical regions, both mainland and islands. Our predictive approach directly points to species that have not yet invaded and areas that are not yet invaded, but are likely to do so in the future.

Pioneer risk-assessment models are all based on datasets that only comprise information on exotic ant species, i.e., species that have already been moved and detected outside of their native zone (24, 29), or pest species, i.e., species that are already known to have concerning impacts (49, 50). As a result, these models cannot point to novel species of potential risk. In contrast, our screening tool is based on a large and collaborative dataset that contains information for 25 ecological traits for the 19 ant species recognized as invasive by the IUCN (47) as well as over 2,000 native species, of which 1,002 had all information (observed or
Invasive profiles

Species | P ± | %
---|---|---

Superinvasive profiles

*Anoplolepis gracilipes* | 0.86 ± 0.02 | 100
*Linepithema humile* | 0.86 ± 0.02 | 100
*Monomorium pharaonis* | 0.86 ± 0.02 | 100
*Myrmica rubra* | 0.86 ± 0.02 | 100
*Nylanderia pubens* | 0.86 ± 0.02 | 100
*Paratrechina longicornis* | 0.86 ± 0.02 | 100
*Tapinoma melanecephalum* | 0.86 ± 0.02 | 100
*Wasmannia auropunctata* | 0.86 ± 0.02 | 100
*Lasius neglectus* | 0.83 ± 0.02 | 100
*Solenopsis geminata* | 0.83 ± 0.02 | 100
*Solenopsis invicta* | 0.83 ± 0.02 | 100
*Technomyrmex albipes* | 0.83 ± 0.02 | 100
*Trichomyrmex destructor* | 0.83 ± 0.02 | 100
*L. canescens* | 0.83 ± 0.12 | 100
*T. difficilis* | 0.83 ± 0.02 | 100
*Pheidole megacephala* | 0.70 ± 0.05 | 100

Invasive profiles

*F. yessensis* | 0.27 ± 0.03 | 100
*T. bicaudatum* | 0.23 ± 0.21 | 100
*A. spinosa* | 0.16 ± 0.02 | 100
*Cardiocondyla emeryi* | 0.16 ± 0.02 | 100
*Cardiocondyla minutor* | 0.16 ± 0.02 | 100
*Cardiocondyla wroghtoni* | 0.16 ± 0.02 | 100
*Neivamyrmex pilosus* | 0.16 ± 0.02 | 100
*Tetramorium simillimum* | 0.16 ± 0.02 | 100
*Monomorium floricola* | 0.16 ± 0.02 | 100
*Solenopsis paniculata* | 0.16 ± 0.02 | 100
*Brachyponera chinensis* | 0.13 ± 0.04 | 100
*Solenopsis richteri* | 0.13 ± 0.04 | 100
*A. trigona* | 0.13 ± 0.04 | 99
*T. laevigatum* | 0.13 ± 0.03 | 93
*L. lujae* | 0.12 ± 0.02 | 90
*A. octospinosus* | 0.00 ± 0.00 | 100

P*: invasiveness probability (mean: 100 models); “±”: invasiveness variability (SD: 100 models); “%”: percentage of models. Note that the values of *A. octospinosus* are P = 0.0001 ± 2.60671E-05.
of biosecurity in these biodiversity-rich ecosystems. It is worth mentioning here that, although occurrence data used for Species Distribution Models (SDMs) were thoroughly checked for issues typical of such data (SI Appendix, Supplementary Material and Methods), some spatiotemporal bias due to the lack of knowledge of the sampling effort remains unavoidable (61). Therefore, the resulting distribution maps should, as usual, be considered with caution.

Our list of predicted invaders crossed with climatic niche modeling offers a worldwide risk assessment of the areas that are the most at risk and informs the relevant countries about the necessity to pay attention to the introduction of ants to prevent invasions. This should be viewed together with the corresponding distribution studies of the already established invasive ant species (e.g., ref. 7). Although climatic suitability is a prerequisite for a successful invasion, the biotic and abiotic characteristics of the receiving community also condition the species establishment and spread and may modulate the local environmental suitability of these species (23). Climate and land use changes may also slightly alter the predictions and create new opportunities for these species to spread in temperate regions and at higher latitudes, as observed for many invasive species (19, 62). A promising future step would be to predict the potential impacts of species predicted to be likely future invaders (63, 64).

Our results are based on the definition of ecological traits. For example, one could argue that using the trait “supercolonial” in our model may hinder its ability to detect new potential invaders, since this behavior has mostly been observed in ant colonies after they have become invasive. However, supercoloniality has already been outlined as an important characteristic of invasive species (37, 65). It might be worth, in this context, completing the AntProfiler database (that currently contains about one-sixth of the known ant species), as having more data is the first step to obtaining more (and more accurate) identifications. A complementary and relevant study would be to analyze which traits condition each step of the invasion process (introduction, establishment, spread, and impact), thus providing some information on which invasion stage is the most critical in order to design a targeted control strategy for each species according to its specific combination of traits (30).

Invasive ant populations, once established in a climatically suitable area, are extremely difficult to control or eradicate (45). A range of control measures exists (e.g., use of chemical insecticides incorporated into baits, biocontrol), but the majority have low efficiency, high negative impacts on native taxa and ecosystems (39), or high costs (13). Overall, our tool can greatly reduce border detection efforts and costs by precisely informing management strategies and preventing some of the potential impacts of invasive ants through the implementation of proactive ant detection programs. Preventing biological invasions is generally believed to be more cost-effective and efficient than the eradication of an established population, long-term control, and repairing the damage caused by invasions (3). This study provides a methodology to predict potential invaders and likely regions of invasions, and we recommend extending and transposing it to assess the invasion potential and areas at risk for other taxa with existing trait databases such as plants, fish, and terrestrial vertebrates.

Materials and Methods

To simultaneously identify the most relevant traits and keep the highest number of species, we proceeded in two steps. First, we identified and selected traits that performed well at predicting invasiveness. Second, we optimized our usable dataset by removing species with too many missing values and by estimating the values of the selected traits for the species with few missing values (using all available traits). We then used the dataset with the selected traits, either observed or estimated, to predict invasiveness probability with Generalized Linear Models. Finally, we predicted areas at risk for invasion of species identified as potential invaders, using Species Distribution Models. Our step-by-step process is illustrated Fig. 3 and is presented below, with further details in SI Appendix.

Trait Data Selection and Imputation. We searched for trait combinations that are associated with invasive species. We extracted ant species traits from our AntProfiler online database (https://antprofiler.ese.u-psud.fr). This database includes information on 2,176 ant species (134 genera) and ecological traits related to their occurrence, morphology, behavior, and invasive status (66). From this database, we selected traits noncorrelated to each other but correlated to the invasive status. The full description of the trait selection process is given in SI Appendix, Supplementary Material and Methods and illustrated in SI Appendix, Fig. S3.

Despite our data collection efforts, the database contained many missing values. We therefore used an imputation technique (missForest) to estimate the missing values (53). Before imputation, we removed the species with too many missing values to keep a dataset with a maximum of ~60% of missing information (following ref. 52). We then explored how missing data could be input into our database to minimize bias and maximize prediction accuracy. To improve the estimation of missing values, we used phylogenetic information and all traits initially downloaded from AntProfiler. We used out-of-bag error estimations to calibrate the models as these were shown to accurately estimate imputation error, and we repeated the imputation 100 times (Fig. 3 and SI Appendix, Supplementary Material and Methods).

We selected a final subset of four traits: colony foundation, colony structure, nesting type, and association with disturbed areas and 1,002 species (of 2,176).

Predictive Model Building.

Potential future invasive species. We modeled the invasive profile using generalized linear models with our four selected traits as predictors and a binomial distribution. We ran 100 models for each of the previously imputed datasets and used model predictions to determine future invasive species. For each model, we identified potentially new invasive ant species as those with a...
predicted invasiveness probability above the lower fifth centile of the 19 known invasive species probability distribution. Each identified species was considered to be potentially invasive if it was selected in at least 90 of the 100 models.

Areas at risk. SDMs are based on correlations between environmental variables and geolocalized species records and can be used to delineate potential species distributions (67). We built SDMs for the ant species found to have similar profiles to those of known invasive species to identify the areas that present suitable environmental conditions for these species and thus areas at risk for invasion. A full description of the SDM process is given in SI Appendix, Supplementary Material and Methods. Two metrics were used to evaluate SDM accuracy: the TSS and the AUC.

Ensemble models—i.e., averages of all models used, weighted by the TSS—were run for each of the predicted invaders with sufficient occurrence points to produce individual climatic suitability distribution maps (SI Appendix, Fig. S11). It is noteworthy that three species (Formica yessensis, Aphaenogaster spinosa, and Azteca trigona) had very few occurrence points (24, 33, and 42, respectively), meaning that the resulting distribution maps should be taken with additional caution. For two other species (Technomyrmex laurenti and Technomyrmex lujae), the number of occurrences was too low (<20 observations) to ensure reliable results; they were subsequently omitted from the analysis. Finally, we also combined these individual predictions by summing each species potential distribution as binary maps to obtain a cumulative invasion risk map from these future invaders. Binary transformation was based on the threshold that maximized the TSS for each species.

Data Availability. All data and scripts used in this study are available at https://github.com/caterinap/Antprofiler.

ACKNOWLEDGMENTS. We thank Cleo Bertelsmeier and Sebastien Ollier with whom we worked on different statistical approaches for a preliminary study and who provided useful comments on an earlier version of the manuscript; five anonymous reviewers who provided useful comments that led to improved analyses; Guillaume Guénard and Daniel J. Stekhoven for their advice, which helped improve our imputation methodology; Benoît Guénard and James Wetterer for sharing their ant occurrence data; and Corrie Moreau and Philip Ward who provided information and data on ant phylogeny.


