RESEARCH ARTICLE



Predicting the invasion range for a highly polyphagous and widespread forest herbivore

Laura M. Blackburn¹, Joseph S. Elkinton², Nathan P. Havill³, Hannah J. Broadley², Jeremy C. Andersen², Andrew M. Liebhold^{1,4}

I Northern Research Station, USDA Forest Service, Morgantown, West Virginia, USA 2 Department of Environmental Conservation, University of Massachusetts, Amherst, Massachusetts, USA 3 Northern Research Station, USDA Forest Service, Hamden, Connecticut, USA 4 Faculty of Forestry and Wood Sciences, Czech University of Life Sciences Prague, 165 00 Praha 6 – Suchdol, Czech Republic

Corresponding author: Andrew Liebhold (andrew.liebhold@usda.gov)

Academic editor: Ingolf Kühn | Received 23 April 2020 | Accepted 26 June 2020 | Published 28 July 2020

Citation: Blackburn LM, Elkinton JS, Havill NP, Broadley HJ, Andersen JC, Liebhold AM (2020) Predicting the invasion range for a highly polyphagous and widespread forest herbivore. NeoBiota 59: 1–20. https://doi.org/10.3897/ neobiota.59.53550

Abstract

Here we compare the environmental niche of a highly polyphagous forest Lepidoptera species, the winter moth (*Operophtera brumata*), in its native and invaded range. During the last 90 years, this European tree folivore has invaded North America in at least three regions and exhibited eruptive population behavior in both its native and invaded range. Despite its importance as both a forest and agricultural pest, neither the potential extent of this species' invaded range nor the geographic source of invading populations from its native range are known. Here we fit a climatic niche model, based on the MaxEnt algorithm, to historical records of winter moth occurrence in its native range and compare predictions of suitable distributions to records from the invaded range. We modeled this distribution using three spatial bins to overcome sampling bias for data obtained from public databases and averaged the multi-continental suitable habitat prediction. Results indicate that this species is distributed across a wide range of climates in its native range but occupies a narrower range in its invaded habitat. Furthermore, the lack of a close fit between climatic conditions in parts of its invaded range and its known native range suggests the possibility that this species has adapted to new climatic conditions during the invasion process. These models can be used to predict suitable habitats for winter moth invasions worldwide and to gain insight into possible origins of North American populations.

Keywords

bioclimatic modelling, biological invasions, climatic envelope, geographical distribution, invasive alien species, range projection, species distribution models

Copyright Laura M. Blackburn et al. This is an open access article distributed under the terms of the Creative Commons Attribution License (CC BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Introduction

With heightened awareness of the damage caused by biological invasions, biosecurity programs take on increasing importance for preventing new invasions (Hulme 2009). Risk assessment plays a key role in biosecurity programs, providing information on likely invasion success and impacts of species, and is used to prioritize preventative measures (Hayes 2003). A crucial component of the risk assessment process is mapping the potential range of a species should it become established (Venette et al. 2010). Knowledge of the potential geographical distribution of a candidate invasive species can guide implementation of both pre- and post-border biosecurity activities such as surveillance.

Understanding and quantifying the ecological niche of a species in its native range can be used to predict its potential distribution in a novel environment. Ecological niche models use occurrence data and environmental variables to predict habitat suitability (Guisan and Zimmermann 2000). These models typically use available species occurrence records to quantify the association of a species' distribution with various components of its habitat such as climate. Ecological niche models assume that a species is well adapted to present climatic conditions and this acclimation assumption is carried into future distribution forecasts. For practicality, these models typically ignore the effects of biotic interactions which may limit a species' potential distribution, and thus are expressed as a realized niche. Application of ecological niche models to predict an invading species future range assumes that successful invasion of novel environments requires similarly matched climates between the native and novel regions (Panetta and Mitchell 1991).

Unfortunately, systematic surveys for most species throughout their ranges are often impractical, though a wealth of distribution information resides in global databases and museum collections worldwide. These datasets, such as the Global Biodiversity Information Facility (GBIF, GBIF.org 2018), assemble occurrence records from many different sources; however, the intensity of sampling behind these records often contains a sampling bias where more records exist in certain areas (such as near research facilities or locations with extensive sampling by hobbyists).

A number of methods can be applied to limit this spatial bias. One method of accounting for sampling bias is to use frequencies of background occurrence records of a conspecific species or an entire genus, often referred to as target group background bias records, as proxies for sampling effort (Yates et al. 2010). Another method involves spatially filtering the occurrence records to remove records at distances greater than the resolution of cells in the environmental layers (Aiello-Lammens et al. 2015; Boria et al. 2014; Hijmans 2012; Kramer-Schadt et al. 2013; Pearson et al. 2007; Veloz 2009). Occurrence records can also be thinned based on their environmental/climatic values, discarding repetitive occurrences in similar climatic conditions (Varela et al. 2014), or records can be filtered geographically, splitting occurrence records into geographic bins (Jiménez-Valverde et al. 2011; Radosavljevic and Anderson 2014).

This study focuses the use of ecological niche models for predicting the potential distribution of the winter moth, *Operophtera brumata*. The native distribution of this

species ranges widely through most of Europe (Kozhanchikov 1950) (Fig. 1). Furthermore, larvae of this species are known to have an extremely broad host range, feeding on the foliage of a wide variety of broadleaf trees and occasionally on certain conifer species (Wint 1983). Recurrent outbreaks of this insect occur in several regions of its native range, especially in mountain birch (*Betula pubescence*) forests in Fennoscandia (Jepsen et al. 2008), but also on other hosts, such as Sitka spruce (*Picea sitchensis*) planted in the British Isles (Stoakley 1985; Watt and McFarlane 1991) and even on heather (*Calluna vulgaris*) in Scotland (Kerslake et al. 1996).

Non-native populations of the species exist in portions of North America with alien populations established in Nova Scotia, the Pacific Northwest, and New England (Fig. 1). Winter moth was first accidentally introduced to Nova Scotia in the 1930s (Hawboldt and Cuming 1950), Oregon in the 1950s (Kimberling and Miller 1988), near Vancouver, British Columbia around 1970 (Gillespie et al. 1978), and eastern Massachusetts in the 1990s (Elkinton et al. 2010). It is unknown whether these represent separate introductions from the native range or a single invasion that has spawned several secondary introductions. Following the initial establishment of each of these populations there was some initial range expansion, though in each case, spread may



Figure 1. A map showing the distribution of the native range for winter moth (Operophtera brumata), recreated from Kozhanchikov (1950), as well as native winter moth occurrence records used as training records – colored according to three geographic filters (British Isles in red, Fennoscandia in blue and Central/Southern Europe in green) and evaluation records shown in yellow **B** map of North American records reflecting the invaded range separated into three geographic regions (western Canada in brown, New England in orange and eastern Canada in violet).

have been limited by climatic conditions, introduction of biological control agents, and/or by hybridization with the native congener, Bruce spanworm, *O. bruceata* (El-kinton et al. 2010).

The objective of this study was to fit ecological niche models based on winter moth occurrence records from its native range in order to predict the potential invaded range of this species. Furthermore, as there is much uncertainty about where in Europe the non-native populations of winter moth originated from (Andersen et al. 2017; Elkinton et al. 2010; Gwiazdowski et al. 2013), we use geographic filtering to compare the climatic suitability of different potential source populations to each of the invaded North American populations. We then comment on the utility of geographic filtering to reduce sample bias in datasets based on public records.

Methods

Modelling approach

The winter moth's climatic niche was quantified using the machine learning algorithm, MaxEnt v. 3.4.1 (Phillips et al. 2006), using presence-only data for this species. MaxEnt uses both distribution information and environmental variables to predict potential areas of distribution for a species. It provides an estimated likelihood for a species range that is near uniform and still subject to environmental confines (Elith et al. 2011). MaxEnt, the most widely used species niche and distribution modelling algorithm (Fourcade et al. 2014), is efficient for complex interactions between response and predictor variables with little sensitivity to small sample sizes. However, difficulties in accuracy arise when modelling generalist or widespread species (Connor et al. 2018; Jiménez-Valverde et al. 2008; Luoto et al. 2005; Marmion et al. 2009; Segurado and Araujo 2004).

Occurrence records

Presence-only distribution data were assembled from various sources: GBIF (www. GBIF.org, taxon key = 1972449), Barcode Of Life Database (BoldData, www.barcodinglife.org), Canadian Forest Invasive Alien Species (CanFIAS, www.exoticpests. gc.ca) database, Elkinton et al. (2010), Andersen et al. (2017), and Tenow et al. (2013). These records were cleaned and separated into two subsets – all of the GBIF records were placed into a training dataset (n = 4151), while records from BoldData, CanFIAS, Elkinton, Andersen and Tenow were separated into a testing dataset (n = 518). Records from the training dataset, consisting of occurrence records for winter moth in the GBIF database, were downloaded on November 29, 2018 using the dismo package in R (Hijmans et al. 2017) and the gbif function. Both training and testing records were cleaned to remove any records where either the latitude or longitude coordinates were missing. Additionally, these records were filtered temporally, keeping only those records between 1970 and 2010, to best align with the environmental layers while maintaining a large sample size. Next, these records were visualized in ArcMap (ESRI 2017) to group locations (Fig. 1) into three regions for native range records: the British Isles (n = 2846), Fennoscandia (n = 894), and Central/ Southern Europe (n = 398); and three regions for invaded range records (n = 103): Western Canada, Eastern Canada and New England. The three native range regions were arbitrarily selected as geographically isolated areas, each with unique densities of occurrence records. The three invaded range regions were selected because they represent isolated ranges, though there is a small coastal area where the Eastern Canada and New England regions are contiguous (Fig. 1). Finally, only unique records were kept – using the delete identical tool with the advanced license of ArcMap. These cleaned records were re-projected into the World Molliweide projection, to match that of the environmental layers, and the latitude and longitude were recalculated before creating a CSV file to be used in MaxEnt.

Sampling bias

Accurate application of MaxEnt necessitates accounting for the effects of geographical sampling bias in locations of occurrence data. Use of occurrence data sets that are spatially biased can result in over-representation of certain environmental features prevalent in more intensively surveyed areas (Phillips et al. 2009). If the sampling effort is known, it can be included in the model calibration to correct for sampling bias (Anderson 2012; Phillips et al. 2009). However, this is rarely the case especially when using museum records from databases, such as GBIF.

We focused our analysis on geographic filtering or splitting of the data into bins to overcome sampling bias. Due to the winter moth's extensive geographical range and the comparatively high density of records in the United Kingdom compared to Central/Southern Europe, we selected to split location records from the native range into three geographical bins: the British Isles, Fennoscandia, and Central/Southern Europe (Fig. 1) as the simplest method for handling sampling bias.

Environmental variables

Environmental variables included in the model were selected from WorldClim 10 minute resolution variables (Fick and Hijmans 2017). WorldClim version 2 includes gridded values for 19 bioclimatic variables derived from temperature and precipitation measurements spanning 1970–2000 collected from a global network of meteorological stations. All environmental variables were resampled in R using the bilinear method to a 25 km cell size in the Mollweide projection, which maintains equal areas – an assumption that is required for MaxEnt models (Elith et al. 2011). We chose to use the following temperature variables that are likely to be of biological importance, BIO1–4,

Variable	Description				
BIO1	Annual Mean Temperature (°C)				
BIO2	Mean Diurnal Range (Mean of monthly (max temp – min temp)) (°C)				
BIO3	Isothermality ((BIO2/BIO7) * 100)				
BIO4	Temperature Seasonality (standard deviation *100)				
BIO6	Min Temperature of Coldest Month (°C)				
BIO7	Temperature Annual Range (BIO5-BIO6) (°C)				
BIO10	Mean Temperature of Warmest Quarter (°C)				
BIO11	Mean Temperature of Coldest Quarter (°C)				
BIO14	Precipitation of Driest Month (mm)				

Table 1. WorldClim v.2 bioclimatic variables included in the model, and their descriptions.

6–7, 10–11 and 14 (Table 1). Insects are ectotherms and therefore their development is sensitive to accumulated degree days as well as to extreme temperatures (Battisti and Larsson 2015). For this reason, we concluded that temperature would be more likely to influence the distribution of winter moth than precipitation, but also identified drought as possibly affecting the pupal stage which reside in the top layer of the soil throughout the summer and early autumn, hence the inclusion of BIO14. Given the polyphagous nature of winter moth larvae, we expected that suitable host plants are present in most locations of the native and invaded ranges and therefore we did not include any variable describing vegetation as predictive environmental variables.

Model parameters

We fit MaxEnt models using the following adjustments to default settings. We generated response curves and jackknife statistics to measure variable importance. Samples files consisted of training datasets for each spatial bin. Environmental layers were clipped to a 400 km buffer around each sample file. The projection layers directory consisted of environmental variables clipped to latitudes above 20°N. The algorithm created 100 replicate models for cross validation. The test sample file was the corresponding testing dataset, the maximum iterations was changed to 5000 for reaching algorithm optimization. A statistical analysis was performed on data extrapolated from each model run, using the receiver operating characteristic (ROC) plot to evaluate model performance. The area under the curve (AUC) of an ROC curve ranges in values from 0 to 1 (Fielding and Bell 1997) with a value of 0.5 indicating that the model performed as good as random and a value of 1 indicates the model has perfect discrimination.

Selection of an appropriate background extent during ecological niche modeling is often overlooked. If the considered extent is too narrow to accurately represent the potential movement of a species over time, the importance of climatic variables in demarcating a species' distribution may be underestimated (Barve et al. 2011). We chose a buffer distance of 400 km, roughly the distance between the Nova Scotia population and the Massachusetts population. However, this distance is much larger than the likely generational winter moth maximum dispersal distance; windborne first instar larvae are known to balloon 50 m (Edland 1971; Huntley et al. 1995) and adult females are incapable of flight.

Model complexity can be varied by altering the regularization parameter; this parameter reduces omission rates. After running the models with regularization values of 0.1, 1 and 3, we chose to use a regularization parameter of 3 to avoid over-fitting our distribution model.

Analyzing the climatic niche

Ridgeline plots of the distribution of environmental variables (BIO 1–4, 6–7, 10–11, 14) among samples were created to further identify differences and similarities in the abiotic niche for each spatial bin from the native versus novel locales. Next, principal components analysis was applied to the nine environmental variables for the pooled occurrence records (both native and invaded ranges) and scores for the first two principal components were plotted separately for each spatial bin (British Isles, Fennoscandia, and Central/Southern Europe, Western Canada, Eastern Canada and New England) in order to discern climatic similarities and differences among regions. ArcMap was used to create a 15 km fishnet of points for the entire study area, which extends 400 km beyond sites of winter moth occurrences in both the native and novel ranges. Next, WorldClim layers were speared to assign their values to each point location. Additionally, cells were coded based on their geographic location (spatial bin) and if within 15 km of a winter moth occurrence. These occurrence data were then exported to R and principal components analyzed with the 'prcomp' function in the base R language.

Distribution modelling

Three different MaxEnt models, one fit to occurrence data from each of the three geographical bins of the native range, were used to predict probabilities of suitable habitat for winter moth in North America. These three model predictions were then averaged to create a combined model and these probabilities were classified into three levels of habitat suitability

MaxEnt output consist of continuous probability values ranging from 0 (unsuitable habitat) to 1 (suitable habitat). MaxEnt output provides the modeler with 11 thresholds to choose from when converting the suitability map to a binary map, all of these thresholds provide a balance between commission and omission rates (Fielding and Bell 1997; Phillips et al. 2017). Model outputs were converted to classified suitability maps. To do this we employed MaxEnt's 'balance' threshold which minimizes 6.00 * training omission rate + 0.04 * cumulative threshold + 1.60 * fractional predicted area. Values below this threshold were dropped and a composite map was created, averaging across predicted probabilities from models derived from each of the three geographical bins in the native range.

Results

Predictions based on native range occurrences

All three models fit to native range records from spatial bins (Fig. 2) performed better than random (Table 2; British Isles AUC = 0.795, Fennoscandia AUC = 0.75, Central/Southern Europe AUC = 0.816). Predictions of suitable winter moth habitat in Europe and North America based on averaging probabilities from the three different models are shown in Fig. 3. The predicted distribution in Europe based on the average of the three models mostly corresponds well with the extent of the distribution map from Kozhanchikov (1950) (Fig. 1). However, the predicted area of suitable habitat extends beyond the eastern boundary of occurrence records into the Carpathian and Caucasus Mountain ranges (Fig. 3). The predicted distribution in North America correctly predicts the winter moth distribution in western Canada, but also predicts extensive areas of suitable habitat north of the invaded range, along the coast of the Cordillera region in Canada and into the Pacific Mountain System of Alaska including the Alaska Peninsula. The prediction for suitable habitat in eastern North America includes a small portion of Nova Scotia, but overall misses most of the Winter Moth's current eastern North American range and instead predicts moderately suitable habitat farther north in eastern Newfoundland. Areas of low habitat suitability in North America are also predicted in the Canadian Rockies, Northern Idaho, Western Montana and the Appalachian Mountains.

Comparison of predicted distributions based on regional bins

Predicted suitable habitat in North America varies among models fit to different geographic bins of native occurrence records (Fig. 2). Models fit to occurrences in the British Isles identified only a small region of suitable habitat in North America, in British Columbia along the North, South and West coasts of Vancouver Island, as well as in Naikoon Provincial Park on Graham Island and in Washington in the San Juan Islands, Fidalgo Island, Whidbey Island and along the northern coast of the Olympic Peninsula from Port Townsend to Neah Bay. Models fit to occurrences in Fennoscandia predicted the greatest amount of suitable habitat in North America, including the entire coast of British Columbia and the southern coast of Alaska from the Prince of Wales Island to the Alaska Peninsula in Western North America and in the East suitable habitat is found throughout Newfoundland and in spots along the southern and southwestern coast of Nova Scotia and on Cape Breton Island. The models fit to occurrences in Central/Southern Europe found moderate to low suitability along the western coast of British Columbia north to the Alaska Peninsula with a localized region of moderately high suitability in the vicinity of Vancouver in Western North America, low predicted suitability along the southeastern coast of Newfoundland, moderate to low suitability along the southern coast of Nova Scotia, as well as low suitability in the Appalachian



Figure 2. Model predictions for each spatial bin in the winter moth native range. The column on the left shows the winter moth training records (shown as black dots) used to make the predictions for suitable habitat in the native range and the column on the right shows the suitable habitat prediction for the invaded range. The prediction is shown from blue (being less suitable) to red (being most suitable).

Mountains and along the New England coast with Cape Cod and Nantucket Island displaying moderately suitable habitat. Models based on the different geographic bins highlight varying degrees of suitable habitat in the vicinity of invaded regions of Vancouver and Vancouver Island; however, the predictions of suitable habitat in Eastern North America were generally not as congruous with the invaded areas with very little overlap between models.

Table 2. Model results for each spatial bin. Percent contribution of environmental variables are in bold for those variables that showed the highest model gain in isolation; values highlighted in gray represent the most information not present in other variables, and * denotes balance threshold used for classified maps which seeks to balance training omission, predicted area and threshold value cloglog threshold.

Winter Moth	sample	threshold	AUC	% Contribution of Environmental Variables								
Region	size	values*		BIO1	BIO2	BIO3	BIO4	BIO6	BIO7	BIO10	BIO11	BIO14
British Isles	381	0.1432	0.795	1.1	0.6	9.3	42.8	14.5	0.1	29.5	0.4	1.7
Fennoscandia	379	0.1132	0.75	5.1	6	5.6	0.7	16.4	0.3	47.1	18.7	0.1
Interior Europe	224	0.0806	0.816	1.4	0.6	7	2.4	1.1	33.7	6.2	2.1	45.6
Model Averages				2.5	2.4	7.3	15.3	10.7	11.4	27.6	7.1	15.8



Figure 3. Predicted suitable habitat in native range (panel **A**) and invaded range (panel **B**) with occurrence records (shown as black dots), this reclassified map is based on the averaged prediction for three spatial bins in the native range. Probabilities falling below balance threshold values shown in Table 2 for each of the three models were classified as "not suitable". Probabilities ≤ 0.25 were classified as "low suitability", values > 0.25 but ≤ 0.5 were classified as "medium suitability" and values > 0.5 were considered "high suitability".

Analysis of environmental variables

Environmental variables contributed differently for each spatial bin (Table 2). Jackknife analyses indicated that the relative contribution of different environmental variables differed considerably among the three models. Temperature seasonality (BIO 4) is the environmental variable that contributed the most to the model fit from occurrences from the British Isles. In the model fit to Central/Southern Europe occurrences, precipitation during the driest month (BIO 14) contributed the most to this model. Mean temperature of the warmest quarter (BIO 10) contributed most to the model fit from Fennoscandia occurrences. When averaging model results for variable contribution across spatial bins, the environmental variables contributing the most influence in the native range are BIO 10, 14, 4, 7 and 6.

While there was considerable overlap in the distribution of climatic variables between the native and invaded ranges, ridgeline plots highlight the variation among populations (Fig. 4). These plots highlight the similarities between the New England



Figure 4. Distribution of bioclimatic variables among occurences within various geographic bins. Gray shading represents winter moth records in the native range. Bioclimatic variables found to contribute the most for each model area shown here, panel **A** shows BIO4 (temperature seasonality, standard deviation *100), panel **B** shows BIO6 (minimum temperature of coldest month, °*C*), panel **C** shows BIO7 (temperature annual range, °*C*), panel **D** shows BIO10 (mean temperature of warmest quarter, °*C*), panel **E** shows BIO 11 (mean temperature of coldest quarter, °*C*) and panel **F** shows BIO 14 (precipitation of driest month, *mm*).

region and Eastern Canada, with New England showing more variability in BIO 6, 7, 10 and 11. When comparing Eastern Canada to the native regions, more climatic values overlap with the Fennoscandia region (BIO 4, 6, 7 and 11) but values of only two variables (BIO 10 and 14) overlap with the Central/Southern Europe region. Values of climatic variables from the New England region overlap with the distribution of values in both Fennoscandia and Central/Southern Europe. The Western Canada region seems most similar to the British Isles across most climatic variables.

Finally, we used principal components analysis to compare the environmental niche at occurrence sites for winter moth in each geographic region (Fig. 5). We analyzed the same climatic variables used in the MaxEnt models (BIO 1–4, 6–7, 10–11 and 14). The first axis generally corresponded to values of BIO3, BIO4, BIO6, BIO7, BIO11 and BIO14 while the second axis was most strongly related to values of BIO10, but also corresponded to values of BIO2. The three native range regions and three invaded regions were generally segregated in the space defined by these two axes. However, there was considerable overlap in the distribution of Central/Southern Europe with western



Figure 5. Plot of occurrences in each geographic region based on their values for the first two principal component axes derived from values of the nine environmental variables used in the MaxEnt models (Table 1).

North America. There was also some overlap in the distribution of Fennoscandia and Central/Southern Europe. Of all invaded regions, the climate in the New England occurrence locations were generally the most different from climates within the native range.

Discussion

As expected, each of the three models predicted suitable habitat in portions of Europe from which occurrence data were located (Fig. 2). Averaging predictions across all three models from the native range (Fig. 3A) produces a predicted European range very similar to the range described in Kozhanchikov (1950) (Fig. 1), though the model fails to predict suitable habitat in Ukraine and southeastern Russia. However, the lack of predicted presence in these areas could have arisen simply from inadequate sampling in that region. Winter moth was recently reported from Tunisia (Mannai et al. 2015), and genetic analyses suggest this population is native to the region (Andersen et al. 2019b). Andersen et al. (2019b) also reported winter moth occurrence in southern Spain and Italy. Because dates of these observations fall outside of the range of dates that were used for the climate data (1970–2010), they were not used in our analysis, though their inclusion could further change the predicted distribution of winter moth to include additional Mediterranean locations. The species is also reported as an established non-native species in Iceland (Halldórsson et al. 2019).

The average of predictions from the models based on the three native range regions predicts suitable habitat in northwestern Washington, along the coast of Western Canada and northward along the coast into Alaska (Fig. 3B). Though winter moth has been established in part of this region (i.e. coastal British Columbia) for over 50 years it has not spread into most of the predicted region. It is possible that populations in this region are constrained from expanding into a climatically suitable region by a biotic factor, such as the introduced parasitoid Cyzenis albicans (Elkinton et al. 2015; Roland and Embree 1995), other natural enemies (Broadley et al. 2019) though there are only a few examples where natural enemies are known to constrain the range of an insect species (e.g., Elkinton et al. 2006, Parry 2008). Another possibility is that hybridization with the native congener, Bruce spanworm (Andersen et al. 2019a) alters the fitness of invading populations, thereby constraining their range. Predictions of the model appear to align with invaded portions of Eastern North America more poorly. The model predicts most of the invaded portions of Nova Scotia and New England as low suitability but classifies all of Newfoundland and portions of Labrador, areas that have never been invaded, as either low or medium suitability. However, it is possible that winter moth has never had the opportunity to invade these regions since the Gulf of St. Lawrence separates them from the currently invaded area. As stated earlier, this lack of agreement in portions of the Eastern North America between the model and winter moth occurrence could also be due to hybridization with Bruce spanworm (Havill et al. 2017).

Predictions of suitable habitat based on climatic niche models fit to native occurrence records sometimes do not coincide well with actual invaded regions (Broennimann et al. 2007; Jiménez-Valverde et al. 2011; Ørsted and Ørsted 2019; Roura-Pascual et al. 2006; Steiner et al. 2008), and this may result from factors such as species not being in equilibrium with the local climate, release from competitors and predators, or sampling and dispersal limitations. Fitzpatrick et al. (2007) found similar results of under-predicting the invasive potential of fire ants using native range occurrence records and over-predicting the southern boundary of the native range using invaded range records, which they concluded was due to fire ants establishing in a novel environment similar to their native environment and then expanding into novel climatic environments not available in their native range.

It is not unusual for populations of various species to become locally adapted to their climate and such local adaptation can result in variation in the potential alien range of populations originating from different portions of the native range. In such cases, models built with spatially partitioned occurrence records from the native range may provide some indication of the geographic origins of invaded populations (Steiner et al. 2008). However, without prior knowledge of the extent of local adaptation, such reconstruction of the origins of invading populations may not be possible. In our analysis, models based on records from all three native regions predict suitability in the western Canada invaded range near Vancouver and Victoria (Fig. 2). Analysis of climatic conditions (Fig. 5) indicates considerable overlap in climate between invaded areas of Western Canada and all three native regions. Thus, our analysis does not show any uniquely high similarity of the invaded portion of Western Canada with any single region of Europe. Only two models fit to European regions (Fennoscandia and the Central/Southern Europe) show suitable habitat in Nova Scotia. In the ridgeline plots (Fig. 4) and the PCA plots (Fig. 5) the records from Eastern Canada fall closest to those of Fennoscandia, suggesting this region as a possible origin for the Eastern Canada population. The model fit to the Central/Southern Europe occurrence records is the only model that predicts suitable habitat in New England, specifically near Cape Cod. When plotting climatic conditions using PCA (Fig. 5), the New England records appear to occupy a different niche mainly outside of the distribution of records from all other regions, though perhaps most similar to the Central/Southern Europe native region, supporting the hypothesis of that region being the origin of the invasion. However, we have no evidence of local adaptation to climate in the winter moth's native range, so such attribution is tentative.

Adaptation to local environments is often observed in species with large geographical ranges (Osborne and Suárez-Seoane 2002; Peterson and Vargas 1993). Differences in predictions from models based on native range occurrence records are apparent in AUC scores (Table 2), the variables of greatest contribution (Table 2) and the averaged prediction (Fig. 3). Probabilities averaged across the three models based on native range regions accurately predicted suitable North American habitat in most of the invaded regions, but predicted particularly high habitat suitability in invaded portions of the south coast of British Columbia (Fig. 3). However, it also predicted large areas of medium to high habitat suitability in uninvaded regions north of both the western and eastern invaded regions. It is suggested that distribution models may be useful in predicting regions where species are likely to invade, but may be flawed in predicting subsequent spread in novel regions (Fitzpatrick et al. 2007; Loo et al. 2007). Given that many of the large areas of predicted suitable habitat are adjacent to currently invaded areas but remain uninvaded, we suspect that there may be unknown biotic or abiotic factors (other than the climatic variables considered here) that limit North American populations into these northern uninvaded areas. However, the identity of such biotic factors remains unknown and we encourage further research into the biotic variables that may be limiting the spread of winter moth in its invaded regions.

Conclusions

We focused our study on a generalist herbivore, the winter moth, to predict areas in North America where this species is likely to invade. We applied MaxEnt, the most widely used species distribution and niche modelling algorithm, to predict the potential range of suitable habitat for winter moth. Preliminary model runs fit to large regions of winter moth occurrence highlighted a high sampling bias in the United Kingdom. We implemented a simple method of applying spatial filters based on geography to reduce sampling bias. Environmental variables were selected based on expectations of climatic factors likely to be important to the biology of this species. We chose to use environmental variables at a coarse grain (25 km) due to the widespread nature of this species and multi-continental areas of interest.

Differentiating the geographic origins for each of the North American winter moth ranges based on predictions from the various native geographic bins is possible, assuming local adaptation to climate in native populations. Based upon climatic similarity, central Europe appears to be the most likely origin of non-native populations in New England. Climatic similarity of the invaded range in Eastern Canada with Fennoscandia and Central Europe suggests those regions as likely origins. However, the Western Canada invaded range appeared equally similar to all native regions and thus there was no evidence regarding possible origins. All of these conclusions regarding origin remain speculative and would require confirmation based on genetic similarities. Combining molecular marker studies in ecological niche modelling approaches can help advance this field (Scoble and Lowe 2010) and improve future predictions of species distribution trajectories. Furthermore, evidence of hybridization with the native Bruce spanworm has been shown in New England, Oregon, British Columbia and Nova Scotia (Andersen et al. 2019a; Elkinton et al. 2010; Havill et al. 2017). If alleles from Bruce spanworm can introgress into winter moth, then this hybridization may increase winter moth's ability to adapt to a novel environment, leading to a shift in the species' fundamental niche (Holt and Gaines 1992).

Predicting the potential North American distribution of this invasive species can aid managers in proactively selecting survey locations for this destructive moth. Areas outside the current species' distribution, which are highly suitable for winter moth, may be prioritized for biosecurity measures to help prevent establishment of this species. However, it remains to be confirmed whether winter moth could establish in the vast regions predicted to be suitable north of currently invaded areas. Given that winter moth is not currently expanding its range into these areas, many of which are adjacent to currently invaded regions, there may be unknown biotic factors that limit the range of this species in ways that are not currently understood.

Acknowledgements

This research was funded by the USDA Forest Service. LB acknowledges support from University of Tennessee National Institute for Mathematical and Biological Synthesis (NIMBioS) at a tutorial session "Applications of Spatial Data: Ecological Niche Modeling". AML was supported by grant EVA4.0, No. CZ.02.1.01/0.0/0.0/16_019 /0000803 financed by OP RDE. Funding to JSE was from USDA-APHIS grant Nos AP17PPQS&T00C068. AP19PPQFO000C125

References

- Aiello-Lammens ME, Boria RA, Radosavljevic A, Vilela B, Anderson RP (2015) spThin: an R package for spatial thinning of species occurrence records for use in ecological niche models. Ecography 38: 541–545. https://doi.org/10.1111/ecog.01132
- Andersen JC, Havill NP, Broadley HJ, Boettner GH, Caccone A, Elkinton JS (2019a) Widespread hybridization among native and invasive species of *Operophtera* moths (Lepidoptera: Geometridae) in Europe and North America. Biological Invasions 21: 3383–3394. https://doi.org/10.1007/s10530-019-02054-1
- Andersen JC, Havill NP, Caccone A, Elkinton JS (2017) Postglacial recolonization shaped the genetic diversity of the winter moth (Operophtera brumata) in Europe. Ecology and evolution 7: 3312–3323. https://doi.org/10.1002/ece3.2860
- Andersen JC, Havill NP, Mannai Y, Ezzine O, Dhahri S, Ben Jamâa ML, Caccone A, Elkinton JS (2019b) Identification of winter moth (*Operophtera brumata*) refugia in North Africa and the Italian Peninsula during the last glacial maximum. Ecology and evolution 9: 13931–13941. https://doi.org/10.1002/ece3.5830
- Anderson RP (2012) Harnessing the world's biodiversity data: promise and peril in ecological niche modeling of species distributions. Annals of the New York Academy of Sciences 1260: 66–80 https://doi.org/10.1111/j.1749-6632.2011.06440.x
- Barve N, Barve V, Jiménez-Valverde A, Lira-Noriega A, Maher SP, Peterson AT, Soberón J, Villalobos F (2011) The crucial role of the accessible area in ecological niche modeling and species distribution modeling. Ecological Modelling 222: 1810–1819. https://doi. org/10.1016/j.ecolmodel.2011.02.011
- Battisti A, Larsson S (2015) Climate change and insect pest distribution range. In: Bjorkman C, Niemela P (Eds) Climate Change and Insect Pests. CABI, Wallingford, 1–15. https:// doi.org/10.1079/9781780643786.0001
- Boria RA, Olson LE, Goodman SM, Anderson RP (2014) Spatial filtering to reduce sampling bias can improve the performance of ecological niche models. Ecological Modelling 275: 73–77. https://doi.org/10.1016/j.ecolmodel.2013.12.012
- Broadley HJ, Kula RR, Boettner GH, Andersen JC, Griffin BP, Elkinton JS (2019) Recruitment of native parasitic wasps to populations of the invasive winter moth in the northeastern United States. Biological Invasions 21: 2871–2890. https://doi.org/10.1007/s10530-019-02019-4
- Broennimann O, Treier UA, Müller-Schärer H, Thuiller W, Peterson A, Guisan A (2007) Evidence of climatic niche shift during biological invasion. Ecology letters 10: 701–709. https://doi.org/10.1111/j.1461-0248.2007.01060.x
- Connor T, Hull V, Viña A, Shortridge A, Tang Y, Zhang J, Wang F, Liu J (2018) Effects of grain size and niche breadth on species distribution modeling. Ecography 41: 1270–1282. https://doi.org/10.1111/ecog.03416
- Edland T (1971) Wind dispersal of the winter moth larvae *Operophtera brumata* L.(Lep., Geometridae) and its relevance to control measures. Nor Entomol Tidsskr 18: 103–107.
- Elith J, Phillips SJ, Hastie T, Dudík M, Chee YE, Yates CJ (2011) A statistical explanation of MaxEnt for ecologists. Diversity and Distributions 17: 43–57. https://doi.org/10.1111/ j.1472-4642.2010.00725.x

- Elkinton JS, Parry D, Boettner GH (2006) Implicating an introduced generalist parasitoid in the invasive browntail moth's enigmatic demise. Ecology 87: 2664–2672. https://doi. org/10.1890/0012-9658(2006)87[2664:IAIGPI]2.0.CO;2
- Elkinton J, Boettner G, Liebhold A, Gwiazdowski R (2015) Biology, spread, and biological control of winter moth in the eastern United States. FHTET-2014-07 Morgantown, WV: US Department of Agriculture, Forest Service, Forest Health Technology Team 22 p, 1–22.
- Elkinton JS, Boettner GH, Sremac M, Gwiazdowski R, Hunkins RR, Callahan J, Scheufele SB, Donahue CP, Porter AH, Khrimian A (2010) Survey for winter moth (Lepidoptera: Geometridae) in northeastern North America with pheromone-baited traps and hybridization with the native Bruce spanworm (Lepidoptera: Geometridae). Annals of the Entomological Society of America 103: 135–145. https://doi.org/10.1603/AN09118
- ESRI (2017) ArcGIS Desktop 10.5.1. Redlands, CA.
- Fick SE, Hijmans RJ (2017) Worldclim 2: New 1-km spatial resolution climate surfaces for global land areas. International Journal of Climatology 37: 4302–4315. https://doi. org/10.1002/joc.5086
- Fielding AH, Bell JF (1997) A review of methods for the assessment of prediction errors in conservation presence/absence models. Environmental conservation 24: 38–49. https:// doi.org/10.1017/S0376892997000088
- Fitzpatrick MC, Weltzin JF, Sanders NJ, Dunn RR (2007) The biogeography of prediction error: why does the introduced range of the fire ant over-predict its native range? Global Ecology and biogeography 16: 24–33. https://doi.org/10.1111/j.1466-8238.2006.00258.x
- Fourcade Y, Engler JO, Rödder D, Secondi J (2014) Mapping species distributions with MAX-ENT using a geographically biased sample of presence data: a performance assessment of methods for correcting sampling bias. PLoS One 9: e97122. https://doi.org/10.1371/ journal.pone.0097122
- GBIF.org (2018) GBIF Home Page. Checklist dataset. https://www.gbif.org [29 November 2018]
- Gillespie D, Finlayson T, Tonks N, Ross D (1978) Occurrence of the winter moth, Operophtera brumata (Lepidoptera: Geometridae), on southern Vancouver Island, British Columbia. Canadian Entomologist. https://doi.org/10.4039/Ent110223-2
- Guisan A, Zimmermann NE (2000) Predictive habitat distribution models in ecology. Ecological Modelling 135: 147–186. https://doi.org/10.1016/S0304-3800(00)00354-9
- Gwiazdowski RA, Elkinton JS, Dewaard JR, Sremac M (2013) Phylogeographic diversity of the winter moths *Operophtera brumata* and *O. bruceata* (Lepidoptera: Geometridae) in Europe and North America. Annals of the Entomological Society of America 106: 143–151. https://doi.org/10.1603/AN12033
- Halldórsson G, Hrafnkelsdóttir B, Oddsdóttir ES (2019) Invasive forest pests in Iceland. Abstracts of NordGen Forest Conference 2019, 13–15.
- Havill N, Elkinton J, Andersen J, Hagen S, Broadley HJ, Boettner G, Caccone A (2017) Asymmetric hybridization between non-native winter moth, *Operophtera brumata* (Lepidoptera: Geometridae), and native Bruce spanworm, *Operophtera bruceata*, in the Northeastern United States, assessed with novel microsatellites and SNPs. Bulletin of Entomological Research 107: 241–250. https://doi.org/10.1017/S0007485316000857

- Hawboldt L, Cuming F (1950) Cankerworms and European Winter Moth in Nova Scotia. Bi-monthly Progress Report Forest Insect Investigations Department of Agriculture Canada 6: 1–2.
- Hayes KR (2003) Biosecurity and the role of risk assessment. In: Carlton J (Ed.) Invasive species: vectors and management strategies. Island Press, Washington, DC, 382–414.
- Hijmans RJ (2012) Cross-validation of species distribution models: removing spatial sorting bias and calibration with a null model. Ecology 93: 679–688. https://doi.org/10.1890/11-0826.1
- Hijmans RJ, Phillips S, Leathwick J, Elith J (2017) dismo: Species Distribution Modeling. R package version 1.1–4. https://CRAN.R-project.org/package=dismo
- Holt RD, Gaines MS (1992) Analysis of adaptation in heterogeneous landscapes: implications for the evolution of fundamental niches. Evolutionary Ecology 6: 433–447. https://doi.org/10.1007/BF02270702
- Hulme PE (2009) Trade, transport and trouble: managing invasive species pathways in an era of globalization. Journal of Applied Ecology 46: 10–18. https://doi.org/10.1111/j.1365-2664.2008.01600.x
- Huntley B, Berry PM, Cramer W, McDonald AP (1995) Special paper: modelling present and potential future ranges of some European higher plants using climate response surfaces. Journal of biogeography: 967–1001. https://doi.org/10.2307/2845830
- Jepsen JU, Hagen SB, Ims RA, Yoccoz NG (2008) Climate change and outbreaks of the geometrids Operophtera brumata and Epirrita autumnata in subarctic birch forest: evidence of a recent outbreak range expansion. Journal of Animal Ecology 77: 257–264. https://doi. org/10.1111/j.1365-2656.2007.01339.x
- Jiménez-Valverde A, Peterson AT, Soberón J, Overton J, Aragón P, Lobo JM (2011) Use of niche models in invasive species risk assessments. Biological Invasions 13: 2785–2797. https://doi.org/10.1007/s10530-011-9963-4
- Jiménez-Valverde A, Lobo JM, Hortal J (2008) Not as good as they seem: the importance of concepts in species distribution modelling. Diversity and Distributions 14: 885–890. https://doi.org/10.1111/j.1472-4642.2008.00496.x
- Kerslake J, Kruuk L, Hartley S, Woodin S (1996) Winter moth (Operophtera brumata (Lepidoptera: Geometridae)) outbreaks on Scottish heather moorlands: effects of host plant and parasitoids on larval survival and development. Bulletin of Entomological Research 86: 155–164. https://doi.org/10.1017/S0007485300052391
- Kimberling DN, Miller JC (1988) Effects of temperature on larval eclosion of the winter moth, Operophtera brumata. Entomologia Experimentalis et Applicata 47: 249–254. https://doi. org/10.1111/j.1570-7458.1988.tb01143.x
- Kozhanchikov IV (1950) Life cycle development and geographical distribution of frost spanworm, *Operophtera brumata* (L.), Leopid. Geometridae). Entomologuicheskoye Obozreniye (Entomological Review) 31: 178–197.
- Kramer-Schadt S, Niedballa J, Pilgrim JD, Schröder B, Lindenborn J, Reinfelder V, Stillfried M, Heckmann I, Scharf AK, Augeri DM (2013) The importance of correcting for sampling bias in MaxEnt species distribution models. Diversity and Distributions 19: 1366–1379. https://doi.org/10.1111/ddi.12096

- Loo SE, Nally RM, Lake PS (2007) Forecasting New Zealand mudsnail invasion range: model comparisons using native and invaded ranges. Ecological applications 17: 181–189. https://doi.org/10.1890/1051-0761(2007)017[0181:FNZMIR]2.0.CO;2
- Luoto M, Pöyry J, Heikkinen R, Saarinen K (2005) Uncertainty of bioclimate envelope models based on the geographical distribution of species. Global Ecology and biogeography 14: 575–584. https://doi.org/10.1111/j.1466-822X.2005.00186.x
- Mannai Y, Ezzine O, Nouira S, Ben J (2015) First report of the winter moth Operophtera brumata on Quercus canariensis and Q. afares in North West of Tunisia. Tunisian Journal of Plant Protection 10: 69–73.
- Marmion M, Luoto M, Heikkinen RK, Thuiller W (2009) The performance of state-of-the-art modelling techniques depends on geographical distribution of species. Ecological Modelling 220: 3512–3520. https://doi.org/10.1016/j.ecolmodel.2008.10.019
- Ørsted IV, Ørsted M (2019) Species distribution models of the Spotted Wing Drosophila (*Drosophila suzukii*, Diptera: Drosophilidae) in its native and invasive range reveal an ecological niche shift. Journal of Applied Ecology 56: 423–435. https://doi.org/10.1111/1365-2664.13285
- Osborne PE, Suárez-Seoane S (2002) Should data be partitioned spatially before building largescale distribution models? Ecological Modelling 157: 249–259. https://doi.org/10.1016/ S0304-3800(02)00198-9
- Panetta F, Mitchell N (1991) Homoclime analysis and the prediction of weediness. Weed Research 31: 273–284. https://doi.org/10.1111/j.1365-3180.1991.tb01767.x
- Parry D (2008) Beyond Pandora's box: quantitatively evaluating nontarget effects of parasitoids in classical biological control. Biological Invasions 11: 47–58. https://doi.org/10.1007/ s10530-008-9319-x
- Pearson RG, Raxworthy CJ, Nakamura M, Townsend Peterson A (2007) Predicting species distributions from small numbers of occurrence records: a test case using cryptic geckos in Madagascar. Journal of biogeography 34: 102–117. https://doi.org/10.1111/j.1365-2699.2006.01594.x
- Peterson AT, Vargas N (1993) Ecological diversity in scrub jays, Aphelocoma coerulescens. In: Ramamoorthy TP, Bye R, Lot A, Fa J (Eds) Biological diversity of Mexico: origins and distribution. Oxford University Press, New York, 309–317.
- Phillips SJ, Anderson RP, Dudík M, Schapire RE, Blair ME (2017) Opening the black box: an open-source release of Maxent. Ecography 40: 887–893. https://doi.org/10.1111/ ecog.03049
- Phillips SJ, Anderson RP, Schapire RE (2006) Maximum entropy modeling of species geographic distributions. Ecological Modelling 190: 231–259. https://doi.org/10.1016/j.ecolmodel.2005.03.026
- Phillips SJ, Dudík M, Elith J, Graham CH, Lehmann A, Leathwick J, Ferrier S (2009) Sample selection bias and presence-only distribution models: implications for background and pseudo-absence data. Ecological applications 19: 181–197. https://doi.org/10.1890/07-2153.1
- Radosavljevic A, Anderson RP (2014) Making better Maxent models of species distributions: complexity, overfitting and evaluation. Journal of biogeography 41: 629–643. https://doi. org/10.1111/jbi.12227

- Roland J, Embree DG (1995) Biological control of the winter moth. Annual review of entomology 40: 475–492. https://doi.org/10.1146/annurev.en.40.010195.002355
- Roura-Pascual N, Suarez AV, McNyset K, Gómez C, Pons P, Touyama Y, Wild AL, Gascon F, Peterson AT (2006) Niche differentiation and fine-scale projections for argentine ants based on remotely sensed data. Ecological applications 16: 1832–1841. https://doi.org/10.1890/1051-0761(2006)016[1832:NDAFPF]2.0.CO;2
- Scoble J, Lowe AJ (2010) A case for incorporating phylogeography and landscape genetics into species distribution modelling approaches to improve climate adaptation and conservation planning. Diversity and Distributions 16: 343–353. https://doi.org/10.1111/j.1472-4642.2010.00658.x
- Segurado P, Araujo MB (2004) An evaluation of methods for modelling species distributions. Journal of biogeography 31: 1555–1568. https://doi.org/10.1111/j.1365-2699.2004.01076.x
- Steiner FM, Schlick-Steiner BC, VanDerWal J, Reuther KD, Christian E, Stauffer C, Suarez AV, Williams SE, Crozier RH (2008) Combined modelling of distribution and niche in invasion biology: a case study of two invasive Tetramorium ant species. Diversity and Distributions 14: 538–545. https://doi.org/10.1111/j.1472-4642.2008.00472.x
- Stoakley J (1985) Outbreaks of Winter moth, Operophthera brumata L. (Lep., Geometridae) in young plantations of Sitka spruce in Scotland 1: Insecticidal control and population assessment using the sex attractant pheromone. Zeitschrift für angewandte Entomologie 99: 153–160. https://doi.org/10.1111/j.1439-0418.1985.tb01973.x
- Tenow O, Nilssen AC, Bylund H, Pettersson R, Battisti A, Bohn U, Caroulle F, Ciornei C, Csóka G, Delb H (2013) Geometrid outbreak waves travel across Europe. Journal of Animal Ecology 82: 84–95. https://doi.org/10.1111/j.1365-2656.2012.02023.x
- Varela S, Anderson RP, García-Valdés R, Fernández-González F (2014) Environmental filters reduce the effects of sampling bias and improve predictions of ecological niche models. Ecography 37: 1084–1091. https://doi.org/10.1111/j.1600-0587.2013.00441.x
- Veloz SD (2009) Spatially autocorrelated sampling falsely inflates measures of accuracy for presence-only niche models. Journal of biogeography 36: 2290–2299. https://doi. org/10.1111/j.1365-2699.2009.02174.x
- Venette RC, Kriticos DJ, Magarey RD, Koch FH, Baker RH, Worner SP, Gómez Raboteaux NN, McKenney DW, Dobesberger EJ, Yemshanov D (2010) Pest risk maps for invasive alien species: a roadmap for improvement. BioScience 60: 349–362. https://doi. org/10.1525/bio.2010.60.5.5
- Watt A, McFarlane A (1991) Winter moth on Sitka spruce: synchrony of egg hatch and budburst, and its effect on larval survival. Ecological Entomology 16: 387–390. https://doi. org/10.1111/j.1365-2311.1991.tb00231.x
- Wint W (1983) The role of alternative host-plant species in the life of a polyphagous moth, Operophtera brumata (Lepidoptera: Geometridae). The Journal of Animal Ecology: 439–450. https://doi.org/10.2307/4564
- Yates CJ, McNeill A, Elith J, Midgley GF (2010) Assessing the impacts of climate change and land transformation on Banksia in the South West Australian Floristic Region. Diversity and Distributions 16: 187–201. https://doi.org/10.1111/j.1472-4642.2009.00623.x

RESEARCH ARTICLE



Pet or pest? Stable isotope methods for determining the provenance of an invasive alien species

Katherine G. W. Hill¹, Kristine E. Nielson², Jonathan J. Tyler², Francesca A. McInerney², Zoe A. Doubleday³, Greta J. Frankham⁴, Rebecca N. Johnson^{4,5}, Bronwyn M. Gillanders¹, Steven Delean¹, Phillip Cassey¹

I School of Biological Sciences and the Environment Institute, The University of Adelaide, Adelaide, South Australia, Australia 2 Department of Earth Sciences and Sprigg Geobiology Centre, University of Adelaide, Adelaide, South Australia 3 Future Industries Institute, University of South Australia, Adelaide, South Australia, Australia 4 Australian Centre for Wildlife Genomics, Australian Museum Research Institute, Sydney, New South Wales, Australia 5 Smithsonian National Museum of Natural History, Washington, D.C., USA

Corresponding author: Katherine Hill (katherine.hill@adelaide.edu.au)

Academic editor: E. García-Berthou | Received 27 April 2020 | Accepted 29 May 2020 | Published 28 July 2020

Citation: Hill KGW, Nielson KE, Tyler JJ, McInerney FA, Doubleday ZA, Frankham GJ, Johnson RN, Gillanders BM, Delean S, Cassey P (2020) Pet or pest? Stable isotope methods for determining the provenance of an invasive alien species. NeoBiota 59: 21–37. https://doi.org/10.3897/neobiota.59.53671

Abstract

The illegal pet trade facilitates the global dispersal of invasive alien species (IAS), providing opportunities for new pests to establish in novel recipient environments. Despite the increasing threat of IAS to the environment and economy, biosecurity efforts often lack suitable, scientifically-based methods to make effective management decisions, such as identifying an established IAS population from a single incursion event. We present a proof-of-concept for a new application of a stable isotope technique to identify wild and captive histories of an invasive pet species. Twelve red-eared slider turtles (Trachemys scripta elegans) from historic Australian incursions with putative wild, captive and unknown origins were analysed to: (1) present best-practice methods for stable isotope sampling of T. s. elegans incursions; (2) effectively discriminate between wild and captive groups using stable isotope ratios; and (3) present a framework to expand the methodology for use on other IAS species. A sampling method was developed to obtain carbon (δ^{13} C) and nitrogen ($\delta^{15}N$) stable isotope ratios from the keratin layer of the carapace (shells), which are predominantly influenced by dietary material and trophic level respectively. Both δ^{13} C and δ^{15} N exhibited the potential to distinguish between the wild and captive origins of the samples. Power simulations demonstrated that isotope ratios were consistent across the carapace and a minimum of eight individuals were required to effectively discriminate wild and captive groups, reducing overall sampling costs. Statistical classification effectively separated captive and wild groups by $\delta^{15}N$ (captive: $\delta^{15}N\% \ge 9.7\%$, minimum of 96% accuracy). This study outlines a practical and accessible method for detecting IAS incursions, to potentially provide biosecurity staff and decision-makers with the tools to quickly identify and manage future IAS incursions.

Copyright Katherine G.W. Hill et al. This is an open access article distributed under the terms of the Creative Commons Attribution License (CC BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Keywords

biosecurity, invasive species, pet trade, provenancing, stable isotopes, Trachemys scripta elegans, wildlife trade

Introduction

Wildlife trade, in particular the legal and illegal pet trade, facilitates the worldwide movements of invasive alien species (IAS), providing novel introduction pathways into new environments (Russello et al. 2008, Lockwood et al. 2019). Accidental escapes or intentional release of alien pets provide numerous opportunities for these species to establish, particularly where the propagule pressure is high from repeated or mass releases (Vall-Ilosera and Cassey 2017). If new populations are not detected rapidly, complete eradication is unlikely and often extremely costly and resource-intensive (Mack et al. 2000). IAS are a key threatening process to global biodiversity loss, thus the prevention of further establishment is critical (Lodge et al. 2006, Baillie et al. 2010). Current methods for discriminating between a recently escaped or released captive individual and an individual from an established population are ineffective and rarely identify the threat early enough for effective eradication (Schmidt et al. 2017). Here, we explore the potential for a novel application of carbon and nitrogen biogeochemistry for determining the provenance of a vertebrate IAS incursion.

The relative abundance of stable isotopes within a material is a function of its synthesis and environmental history, which, in the case of vertebrate animals, predominantly relates to their diet (Camin et al. 2016). In vertebrate animal tissue, stable carbon isotope ratios (a measure of the relative abundance of ${}^{13}C/{}^{12}C$, reported as $\delta^{13}C$) are linked to the $\delta^{13}C$ of the animal's diet, which, in turn, is strongly influenced by the relative proportion of the C3 and C4 plants an animal directly or indirectly consumes. Nitrogen stable isotope ratios (a measure of the relative abundance of ${}^{15}N/{}^{14}N$, reported as $\delta^{15}N$) are also influenced by the animal's diet; specifically, they indicate the trophic position of the animal. Given that a difference in the diet between wild and captive animals is extremely likely, stable isotope ratios can potentially utilise these differences to provide information on the origin of the animal (Ziegler et al. 2018).

Stable isotopes are a well-established forensic technique and are a strong candidate for identifying the origin of IAS incursions (Cerling et al. 2016). Environmental research using stable isotopes include tracing vertebrate movements, including tracking migratory animals where satellite trackers cannot be used (e.g. MacKenzie et al. 2011, Madigan et al. 2017) and identifying diets and niche positions (e.g. Haubrock et al. 2020, Pearson et al. 2013). Previous studies have explored stable isotope tools to distinguish between wild and captive animals with success, including, but not limited to: short-beaked echidnas (*Tachyglossus aculeatus*) (Brandis et al. 2018); wolves (*Canis lupis*) (Kays and Feranec 2011); African grey parrots (*Psittacus erithacus*) (Alexander et al. 2019, Symes et al. 2017); reticulated pythons (*Python reticulatus*) (Natusch et al. 2017); and crocodile lizards (*Shinisaurus crocodilurus*) (van Schingen et al. 2016, Ziegler et al. 2018). In all of these cases, stable isotopes have proved efficient at identifying environmental histories and diets. Yet, with the exception of insects (Holder et al. 2014, Hood-Nowotny et al. 2016), the use of stable isotopes for determining the provenance of IAS, early in the incursion process, is relatively unexplored.

Trachemys scripta elegans (red-eared slider turtles) were selected as a case study to test the efficacy of δ^{13} C and δ^{15} N for biosecurity applications. As one of the world's top 100 most invasive species, T. s. elegans have the potential to establish and spread in urban and semi-rural areas worldwide (IUCN 2000, Rodder et al. 2009, Banha et al. 2017). T. s. elegans are the most-traded turtle species in the world, where more than 50 million individuals were exported from USA between 1989 to 1997 to supply the global pet trade (Telecky 2001) and the trade continues illegally despite being restricted in most regions (Kitowski and Pachol 2009; García-Díaz et al. 2015). These animals grow large quickly, resulting in being intentionally released into waterways when they become undesirable as pets. As a consequence, T. s. elegans have established nearly 200 identified breeding populations worldwide (Kikillus et al. 2010). They are a significant threat to biodiversity, as they compete with native turtles for food and shelter (Pearson et al. 2013, Balzani et al. 2016) and carry exotic diseases including Ranavirus and Chlamydia spp. (Johnson et al. 2007, Mitura et al. 2017). As T. s. elegans are omnivorous, the proportions of meat and plant material that may vary between wild and captive diets are likely to drive differences in δ^{13} C and δ^{15} N, making them good candidates for our case study.

New methodologies are urgently needed to provide early identification of incursions as distinct from established populations, to allow for quick and effective eradication (Lodge et al. 2006). Here, we present a new application of δ^{13} C and δ^{15} N using historical Australian incursion samples of *T. s. elegans* of putative wild and captive origins to: (i) determine best-practice methods for sampling *T. s. elegans* incursions; (ii) evaluate the use of δ^{13} C and δ^{15} N to discriminate between wild and captive individuals; and (iii) provide a framework to expand the methodology for use on other IAS species.

Methods

Sample collection

T. s. elegans post-mortem specimens were loaned from the Queensland Museum, the Department of Primary Industries and Regions, South Australia and the Australian Museum Research Institute Herpetology Collection. These animals were collected by state wildlife compliance agencies under their powers to seize animals being kept in contradiction to legislation or found at-large in wild environments. All animals were euthanised as per state and territory biosecurity protocols and stored frozen. The national collection contains seized *T. s. elegans* incursions from various locations across Australia. Due to the nature of the limited sample collection and the value of biosecurity material, twelve animals from various Australian locations (Fig. 1) with sufficient details to determine the accuracy of their environmental histories were selected, which were: (i) seized from illegal captive holding or commercial sale (n = 4); (ii) surrendered



Figure 1. Locations of Australian *T. s. elegans* incursion samples used in this study. Identified established populations exist in Sydney, New South Wales (Burgin 2006, Mo 2018) and multiple locations in Queensland (O'Keeffe 2005).

by members of the public (n = 3); or (iii) found at-large in wild-states (n = 5). Sex and age were determined by dissection and secondary characteristics, according to Gradela et al. (2017). Due to the lack of confirmed established populations, widespread sampling of different environments and, thus, potential environmental variation in δ^{13} C and δ^{15} N, was not possible. However, it was assumed these variances would be captured in the between-individual variation.

Based on the assumed environmental history of the individual turtles, we assigned the variable "status" and classified individuals as "wild" or "captive". While the majority of animals used in this study had relatively high confidence of their origin, there remains uncertainty in the status of individuals being correctly assigned by authorities. Therefore, we created an index to determine the percent confidence of correct classification, based on how many secondary characteristics matched the original assessment by authorities, including: (i) proximity to a known established population; (ii) presence of algae or wild features on the carapace; (iii) seized by authorities or surrendered by a member of the public. This provided a confidence scale for selecting the individuals used for a decision model (Suppl. material 1). Four individuals seized from captivity had 100% certainty in status, while the remaining individuals contained varying degrees of uncertainty. Individuals with less than 50% certainty were classified as "unknown" (n = 2).

Isotopic analysis

Measuring stable isotope ratios from a slow-turnover and inert tissue provides a longterm record of an animal's environmental history (Dalerum and Angerbjörn 2005). The keratin covering a turtles' carapace in scale-like sections (scutes) was selected due to its slow, annual growth and ease of sampling (Schneider et al. 2015). Scutes grow by adding new layers to the base, whilst widening in each layer from the perimeter to account for growth. Old scutes are then shed from the perimeter of the shell (Clinical Anatomy and Physiology of Exotic Species 2005). Assuming *T. s. elegans* scute growth is similar to *Chrysemys picta* from the same sub-family (Emydidae), a new layer of scute grows in warmer months of spring to summer, while the previous year's growth is shed; providing temporal comparisons between years (Alibardi 2005). Shed scute was available for three turtles (C1, U1 and U2), with primary scutes retained on the carapace and secondary scute peeled off after freeze-drying. On one individual, a tertiary layer was available as a second layer of peeling scute.

Carapaces were washed, removed from the body and freeze-dried to separate partially shed scutes, to exclude water contamination and to ensure only one layer of scute was sampled at a time. Samples on shed scutes were cut using sterile dissecting scissors, while shavings were collected on attached scutes using sterile scalpels. Scute samples were weighed and placed in tin capsules for continuous-flow isotope ratio mass spectrometry (CF-IRMS) using an Elementar elemental analyser coupled to a Nu Horizon mass spectrometer at the University of Adelaide. Standards of glycine, glutamic acid and USGS41 (L-glutamic acid; Reston Stable Isotope Laboratory 2011) were run periodically to correct for mass effects and instrumental drift during and between runs. Isotope ratios are reported in per mil (‰), where δ^{13} C is reported relative to the Vienna Pee Dee Belemnite (VPDB) standard and δ^{15} N is reported relative to AIR. The values of δ^{13} C and δ^{15} N were measured during the same analysis.

Sampling size and design

To improve the detectable δ^{13} C and δ^{15} N separation between captive and wild groups or to increase effect size differences, the variance of each hierarchal level of sampling (indi-

vidual > scute > sample) needed to be minimised without oversampling (Nicholson and Holmes 2017). A pilot analysis was performed to provide estimates of the variances of each hierarchal level of sampling. Calculated variances, along with proposed measures of detectable differences in isotope ratios, were used to compare different sampling designs that were generated through simulation (Green and MacLeod 2016) with the aim to determine a suitable sample size of individual animals and number of scutes and samples within scutes per individual. The most practical sampling method was determined as that with power > 80% with minimum sampling (Suppl. material 2) to detect the minimum observed difference of δ^{15} N between animals from different status groups.

Analysis of turtle provenance

Differences in mean isotope ratios amongst individual turtle specimens were evaluated using linear mixed effects models. The values of δ^{13} C and δ^{15} N were fitted independently as response variables, with individual turtles as a fixed effect and scute as a random effect to allow for variation between repeated measurements within a scute. The models explicitly allowed for differences in variation between individuals, because heterogeneity within individuals violated the constant variance assumption of the linear mixed effect models. The effects of sex on δ^{13} C and δ^{15} N were investigated using linear mixed effect models with and without sex as a term and examined significance of dropping different independent variables using a Pearson's chi-squared test. We were unable to investigate other variables of interest, such as location and climate, due to the broad variety of the small number of representative samples. Instead, these contribute to the between-individual variation.

The overall objective was to evaluate if a decision rule could be developed that allowed wild and captive individuals to be identified, based on their δ^{13} C and δ^{15} N values. To assess this, a classification tree approach was adopted by introducing the status as a response variable. As the data consisted of multiple observations from the same individual turtles, a structured cross-validation approach was used to evaluate the prediction error, with all observations from the same individual included in the 'hold-out' set for prediction; and to avoid over-fitting. Individuals with an unknown status were omitted, as well as juvenile turtle W1 due to potential differences in diet between juvenile and adult turtles (Reed and Krysko 2014). The optimal classification tree was selected using the classification parameter; where misclassification errors were not significantly improved with the penalty of adding further nodes. A classification tree was constructed using this process, including all individual observations across all turtles to allow for variation within individuals. To analyse the temporal variability across the shed and retained scutes, separate linear models for each isotope were fitted, to assess changes in δ^{13} C and δ^{15} N

All analyses were conducted in the R software environment for statistical and graphical computing (V 3.5.3; R Core Team 2019). Linear mixed effects models were

fitted with the R package "lme4" (Bates et al. 2014), simulations were performed in "simr" (Green and MacLeod 2016) and classification trees with "rpart" (Trevor et al. 2009, Therneau et al. 2018).

Results

Sampling size and design

Power simulations indicated nine samples across two separate scutes on the carapace were sufficient to capture individual variation, while retaining a detectable difference between wild and captive individuals. Variation between scutes of the same layer (primary and secondary) was minimal when compared to variation between individuals. Sampling four individuals per status group ("captive" and "wild" groups, eight individual turtles in total) provided the greatest power at a minimum of 96%. The position of the samples within the scute had no significant effect on $\delta^{15}N$ ($\chi_6^2 = 1.76$, p > 0.05) nor $\delta^{13}C$ ($\chi_6^2 = 0.840$, p > 0.05).

Analysis of turtle provenance

Status (wild versus captive) was the main factor underlying differences in isotope values. There was evidence for an effect of status on isotopic ratios (χ_1^2 = 4.02, *p* = 0.0451), but no clear differences between the sexes (χ_2^2 = 3.66, *p* = 0.160).

Individual turtles had their own unique δ^{13} C and δ^{15} N values and within-individual variation was generally less than between-individual variation (Table 1). Wild turtles exhibited lower δ^{15} N values and a greater spread in δ^{13} C values compared to captive turtles (Fig. 2).

Turtle	δ ¹⁵ N mean	δ^{15} N SE	δ ¹³ C mean	$\delta^{13}CSE$	n
C1	13.08	0.40	-21.41	0.22	25
C2	10.51	0.40	-22.24	0.27	26
C3	10.13	0.42	-21.62	0.66	18
C4	10.90	0.40	-19.67	0.28	16
C5	12.58	0.45	-18.33	0.26	15
C6	13.62	0.43	-19.09	0.27	18
U1	8.03	0.40	-22.53	0.27	18
U2	12.39	0.40	-18.70	0.27	18
W1	7.42	0.42	-20.09	0.52	23
W2	6.44	0.40	-27.26	0.28	24
W3	8.71	0.42	-25.34	0.32	26
W4	9.22	0.41	-22.63	0.29	16

Table 1. δ^{15} N and δ^{13} C means, standard error (SE) and sample sizes (*n*) for individual turtles.



Figure 2. A Confidence in original status assignment based on select characteristics (Suppl. material 1); **B** plot of δ^{13} C and δ^{15} N means and 95% confidence intervals for captive and wild *T. s. elegans* individuals, coloured according to their assignment confidence described in Figure 2A.

The classification tree showed clear differences between captive and wild groups associated with $\delta^{15}N$ (Fig. 3). Only $\delta^{15}N$ was required to separate captive and wild groups, with captive animals identified by $\delta^{15}N\% \ge 9.7\%$, with a success rate of at least 96%. If $\delta^{15}N$ information was not included, groups were poorly separated with captive turtles identified by $\delta^{13}C\% \ge -22\%$ with success of 81%. Using the generated classification tree, the two unknown turtles "U1" and "U2" were classified as wild and captive, respectively (Fig. 3).

Of the three turtles with shed scutes available, all revealed significant differences between layers in δ^{13} C (C1: $F_{1,31} = 100.4$, p < 0.0001; U1: $F_{3,51} = 0.0006$, p < 0.0001; U2: $F_{2,34} < 0.0001$, p < 0.0001) and δ^{15} N for turtles C1 and U1 ($F_{1,31} = 100.4$, p < 0.0001; $F_{2,50} = 82.7$, p < 0.0001), but not U2 ($F_{1,34} = 1.771$, p = 0.192). However, there was no consistency in the direction of change between layers. Furthermore, the δ^{15} N values for each scute layer remained within the classification range of their assigned status; "wild" and "captive".



Figure 3. $\delta 13C$ and $\delta 15N$ means and 95% confidence intervals for the shed and retained scute on three individuals: captive C1 and unknowns U1 and U2. Layers are labelled as primary (newest growth; retained on carapace) and secondary (older growth; shed scute). An older scute layer was available on turtle U1, named tertiary (oldest growth; retained scute).

Discussion

Captive and wild *T. s. elegans* are effectively differentiated by their δ^{15} N. Sampling scute proved to be a simple method; no specialist equipment was required for collection and samples could be taken anywhere on the scute and across multiple scutes with minimal variation within the individual. Although individuals were dissected for this study, the use of scute shavings is potentially a non-invasive method. This makes the technology accessible for non-specialist practitioners, such as biosecurity or veterinary staff and for samples to be collected and sent to a laboratory for analysis and determination of their origins. Furthermore, the power simulations demonstrated that minimal sampling per individual is required, reducing the overall sampling costs in time, effort and welfare, as well as monetary cost.

As material of high biosecurity risk is inherently difficult to obtain, the availability of *T. s. elegans* and other reptile IAS is limited, while information surrounding an animal's history is not always accessible. Wild *T. s. elegans* specimens are rare in Australia, as at-large populations have only been confirmed in Sydney (Burgin 2006, Robey et al. 2011; Mo 2019) and Queensland (O'Keeffe 2005). Furthermore, *T. s. elegans* are

illegal to import or keep in Australia without a licence, limiting available samples to those confiscated from illegal keeping (Department of Agriculture and Water Resources 2017). Here, we used a selection of samples for which we were initially confident in the assignment of status to an individual. Furthermore, the use of a power analysis demonstrated that a minimum of eight ($n \ge 8$) individuals of known origin was sufficient to obtain a detectable effect size to effectively separate wild and captive groups.

Separation of wild and captive groups used a simple classification tree model, which effectively differentiated wild and captive individuals with minimal misclassification error. As samples with relatively high status confidence were used, this classification tree can be adopted as a set of best-practice methods and model to determine the origins of *T. s. elegans* individuals found in wild-states. However, further refinement of the model is required, such as including a wider range of locations of samples to improve the discrimination power.

Differences in the δ^{13} C and δ^{15} N composition of scutes from different status groups are likely primarily influenced by different proportions and sources of plant and animal material within a turtle's diet, as well as varied sources of these food groups (Balzani et al. 2016). The trophic level of captive turtles, as inferred from δ^{15} N, is consistently higher than for wild populations, despite potential δ^{15} N enrichment from agricultural fertilisers in wild environments (Hofmeister et al. 2013). This is likely influenced by a higher consumption of meat-based products by captive turtles, including commercial turtle food (Mazumder et al. 2018). Commercial turtle food often contains marine origin proteins, which may increase δ^{13} C and δ^{15} N (Schoeninger and DeNiro 1984). A similar result has been identified in studies on other reptiles, such as crocodile lizards (*Shinisaurus crocodilurus*) and monitor lizards (*Varanus* spp.), where captive animals possessed enriched δ^{15} N (van Schingen et al. 2016, Natusch et al. 2017, Ziegler et al. 2018). Therefore, δ^{15} N is a strong candidate for the expansion of this forensic technique into other reptile groups, but likely requires species-specific validation.

The δ^{13} C exhibited little power for separating wild and captive groups. As with δ^{15} N, δ^{13} C is influenced by a variety of environmental factors. However, δ^{13} C was identified as the most significant separator for wild and captive juvenile *T.s.scripta* by Aresco and James (2005). Juvenile turtle W1 showed enrichment in δ^{13} C, which may be due to an ontogenetic shift in diet by adult *T. s. elegans* (Reed and Krysko 2014). While δ^{13} C had some separating power and may be applied for a more detailed analysis of *T. s. elegans* dietary behaviour in wild and captive states, we did not find it informative as a biosecurity tool for separating adult incursions.

For each turtle, where shed scute was available, the $\delta^{15}N$ and $\delta^{13}C$ exhibited significant differences between successive active seasons. However, as there was no consistent direction of change in the isotope data, it is unlikely the changes are due to tissue degradation and instead likely reflected temporal variability in the turtle's diet. The variance in $\delta^{15}N$ was sufficiently small to ensure that the specimen remained within the same status group, based on the $\delta^{15}N \approx 9.7\%$ discrimination value.

It is important to note that the status assignment refers to the confidence that the turtle was wild or captive for the entirety of the scute growth period. The natal origin (birthplace) of the turtle cannot be determined using scute growth alone as scutes are shed yearly (Alibardi 2005). Natal origin determination requires sampling of a tissue which has remained inert since hatching, such as bone, as explored by Holder et al. (2014). Determining a tissue which has remained inert since hatching in *T. s. elegans* and a comparison to scute material will be extremely useful for future biosecurity efforts.

The exploration of additional biogeochemical tracers may be useful to create a more diverse set of methods and potentially obtain greater evidence of environmental origin. Stable isotopes relating to the animal's water source such as hydrogen (${}^{2}H/{}^{1}H$) and oxygen (${}^{17}O/{}^{16}O$) or ${}^{18}O/{}^{16}O$) may provide useful information on the animal's geographical origin and have been used in other animal tracking applications (Bowen et al. 2005, Hobson and Wassenaar 2018). Furthermore, controlled experiments to determine the scute-diet fractionation factor in *T. s. elegans* may provide greater information on the animals' diets.

Conclusions

The values of δ^{13} C and δ^{15} N in scute keratin are effective at filling the requirement for the urgent need for effective forensic techniques to quickly identify the origin of *T. s. elegans* (red-eared slider turtle) incursions and has promising potential for applications on other high-risk IAS species (Lodge et al. 2006, McFadden et al. 2017). Stable isotope ratios in the scutes of *T. s. elegans* provide long-term information on individual environmental histories and, thus, provide an effective forensic method for identifying the origins of individuals found in wild-states. This study provides a set of best-practice, relatively accessible methods for sampling IAS incursions and subsequent analysis, including a classification tree, to determine the risk of future incursions. These approaches, using an emerging and effective forensic technique, contribute to the continuing development of various forensic techniques that are crucial for effective biosecurity efforts.

Acknowledgements

Turtle carcasses were kindly provided through scientific loans by the Queensland Museum (Patrick Couper), Department of Primary Industries and Regions, South Australia (Lindell Andrews) and Australian Museum's Herpetology Collection (Jodi Rowley and Stephen Mahoney). For sample collection and seizure information, we thank the Primary Industries and Regions SA (Lindell Andrews), the Department of Primary Industry, NSW (Alyssa Trotter, Nathan Cutter), the Elizabeth Macarthur Agricultural Institute (Brendon O'Rourke, Naomi Porter) and Department of Jobs, Precincts and Regions/Agriculture Victoria (Jesse Miller). We are extremely grateful to Mark Rollog for assistance with CF-IRMS data collection, Jennifer Pistevos for her work on pilot analyses and Talia Wittmann for research assistance. This research was supported by the University of Adelaide Environment Institute, by ARC FT110100793 to F.A. McInerney and by Invasive Animals CRC (Project 1L4) and Centre for Invasive Species Solutions (Project PO1-I-002) funding to P. Cassey. An ARC LIEF grant (LE120100054) funded the IRMS used for analyses.

References

- Alexander J, Downs CT, Butler M, Woodborne S, Symes CT (2019) Stable isotope analyses as a forensic tool to monitor illegally traded African grey parrots. Animal Conservation 22: 134–143. https://doi.org/10.1111/acv.12445
- Alibardi L (2005) Proliferation in the epidermis of chelonians and growth of the horny scutes. Journal of Morphology 265: 52–69. https://doi.org/10.1002/jmor.10337
- Aresco MJ, James FC (2005) Ecological relationships of turtles in northern Florida lakes: a study of omnivory and the structure of a lake food web. Florida State University Tallahassee, Florida, USA.
- Baillie J, Griffiths J, Turvey S, Loh J, Collen B (2010) Evolution lost: status and trends of the world's vertebrates. Zoological Society of London, London, UK.
- Balzani P, Vizzini S, Santini G, Masoni A, Ciofi C, Ricevuto E, Chelazzi G (2016) Stable isotope analysis of trophic niche in two co-occurring native and invasive terrapins, *Emys* orbicularis and *Trachemys scripta elegans*. Biological Invasions 18: 3611–3621. https://doi. org/10.1007/s10530-016-1251-x
- Banha F, Gama M, Anastacio PM (2017) The effect of reproductive occurrences and human descriptors on invasive pet distribution modelling: *Trachemys scripta elegans* in the Iberian Peninsula. Ecological Modelling 360: 45–52. https://doi.org/10.1016/j.ecolmodel.2017.06.026
- Bates D, Mächler M, Bolker B, Walker S (2014) Fitting linear mixed-effects models using lme4. Journal of Statistical Software 67: 1–48. https://doi.org/10.18637/jss.v067.i01
- Bowen GJ, Wassenaar LI, Hobson KA (2005) Global application of stable hydrogen and oxygen isotopes to wildlife forensics. Oecologia 143: 337–348. https://doi.org/10.1007/s00442-004-1813-y
- Brandis KJ, Meagher PJB, Tong LJ, Shaw M, Mazumder D, Gadd P, Ramp D (2018) Novel detection of provenance in the illegal wildlife trade using elemental data. Scientific Reports 8: 8. https://doi.org/10.1038/s41598-018-33786-0
- Burgin S (2006) Confirmation of an established population of exotic turtles in urban Sydney. Australian Zoologist 33: 379–384. https://doi.org/10.7882/AZ.2006.011
- Burleigh R, Brothwell D (1978) Studies on Amerindian dogs, 1: Carbon isotopes in relation to maize in the diet of domestic dogs from early Peru and Ecuador. Journal of Archaeological Science 5: 355–362. https://doi.org/10.1016/0305-4403(78)90054-7
- Camin F, Bontempo L, Perini M, Piasentier E (2016) Stable isotope ratio analysis for assessing the authenticity of food of animal origin. Comprehensive Reviews in Food Science and Food Safety 15: 868–877. https://doi.org/10.1111/1541-4337.12219
- Cerling TE, Barnette JE, Bowen GJ, Chesson LA, Ehleringer JR, Remien CH, Shea P, Tipple BJ, West JB (2016) Forensic stable isotope biogeochemistry. Annual Review of Earth and Planetary Sciences 44: 175–206. https://doi.org/10.1146/annurev-earth-060115-012303
- Clinical Anatomy and Physiology of Exotic Species (2005) Chapter 3 Tortoises and turtles. In: O'Malley B (Ed.) Clinical Anatomy and Physiology of Exotic Species. W.B. Saunders Ltd., Edinburgh, 41–56. https://doi.org/10.1016/B978-070202782-6.50006-5

- Dalerum F, Angerbjörn A (2005) Resolving temporal variation in vertebrate diets using naturally occurring stable isotopes. Oecologia 144: 647–658. https://doi.org/10.1007/s00442-005-0118-0
- Department of Agriculture and Water Resources (2017) Biosecurity Regulation 2016. In: DoAaW (Ed.) Resources. Canberra, Australia.
- García-Díaz P, Ross JV, Ayres C, Cassey P (2015) Understanding the biological invasion risk posed by the global wildlife trade: propagule pressure drives the introduction and establishment of Nearctic turtles. Global Change Biology 21: 1078–1091. https://doi.org/10.1111/ gcb.12790
- Gradela A, Santiago TOC, Pires IC, Silva, AdCS, de Souza LC, de Faria MD, Neto JP, Milanelo L (2017) Sexual dimorphism in red-eared sliders (*Trachemys scripta elegans*) from the Wild Animal Triage Center of the Tiete Ecological Park, São Paulo, Brazil. Acta Scientiae Veterinariae 45: 1–10. https://doi.org/10.22456/1679-9216.80442
- Green P, MacLeod CJ (2016) SIMR: an R package for power analysis of generalized linear mixed models by simulation. Methods in Ecology and Evolution 7: 493–498. https://doi. org/10.1111/2041-210X.12504
- Haubrock JF, et al. (2020) When alien catfish meet resource overlap between the North American *Ictalurus punctatus* and immature European *Silurus glanis* in the Arno River (Italy). Ecology of Freshwater Fish, 29: 4–17. https://doi.org/10.1111/eff.12481
- Hobson KA, Wassenaar LI (2018) Tracking animal migration with stable isotopes. Academic Press. https://doi.org/10.1016/B978-0-12-814723-8.00001-5
- Hofmeister NR, Welk M, Freedberg S (2013) Elevated levels of δ15N in riverine Painted Turtles (Chrysemys picta): trophic enrichment or anthropogenic input? Canadian Journal of Zoology 91: 899–905. https://doi.org/10.1139/cjz-2013-0121
- Holder PW, Armstrong K, Van Hale R, Millet MA, Frew R, Clough TJ, Baker JA (2014) Isotopes and trace elements as natal origin markers of *Helicoverpa armigera* - an experimental model for biosecurity pests. PLOS ONE 9: e92384. https://doi.org/10.1371/journal. pone.0092384
- Hood-Nowotny R, Mayr L, Saad N, Seth RK, Davidowitz G, Simmons G (2016) Towards incorporating insect isotope analysis using cavity ring-down spectroscopy into areawide insect pest management programs. Florida Entomologist 99: 177–184. https://doi. org/10.1653/024.099.sp121
- Hu Y, Luan F, Wang S, Wang C, Richards MP (2009) Preliminary attempt to distinguish the domesticated pigs from wild boars by the methods of carbon and nitrogen stable isotope analysis. Science in China Series D: Earth Sciences 52: 85–92. https://doi.org/10.1007/ s11430-008-0151-z
- IUCN (2000) IUCN guidelines for the prevention of biodiversity loss caused by alien invasive species. In: International Union for the Conservation of Nature. Gland, Switzerland.
- Johnson AJ, Pessier AP, Jacobson ER (2007) Experimental transmission and induction of ranaviral disease in western ornate box turtles (*Terrapene ornata ornata*) and red-eared sliders (*Trachemys scripta elegans*). Veterinary Pathology 44: 285–297. https://doi.org/10.1354/ vp.44-3-285

- Kays R, Feranec R (2011) Using stable carbon isotopes to distinguish wild from captive wolves. Northeastern Naturalist 18: 253–264. https://doi.org/10.1656/045.018.0301
- Kikillus KH, Hare KM, Hartley S (2010) Minimizing false-negatives when predicting the potential distribution of an invasive species: a bioclimatic envelope for the red-eared slider at global and regional scales. Animal Conservation 13: 5–15. https://doi.org/10.1111/ j.1469-1795.2008.00299.x
- Kitowski I, Pachol D (2009) Monitoring the trade turnover of red-eared terrapins (*Trachemys scripta elegans*) in pet shops of the Lublin region, east Poland. North-Western Journal of Zoology 5: 34–39.
- Lockwood JL, Welbourne DJ, Romagosa CM, Cassey P, Mandrak NE, Strecker A, Leung B, Stringham OC, Udell B, Episcopio-Sturgeon DJ, Tlusty MF, Sinclair J, Springborn MR, Pienaar EF, Rhyne AL, Keller R (2019) When pets become pests: the role of the exotic pet trade in producing invasive vertebrate animals. Frontiers in Ecology and the Environment 17: 323–330. https://doi.org/10.1002/fee.2059
- Lodge DM, Williams S, MacIsaac HJ, Hayes KR, Leung B, Reichard S, Mack RN, Moyle PB, Smith M, Andow DA (2006) Biological invasions: recommendations for US policy and management. Ecological Applications 16: 2035–2054. https://doi.org/10.1890/1051-0761(2006)016[2035:BIRFUP]2.0.CO;2
- Mack RN, Simberloff D, Lonsdale WM, Evans H, Clout M, Bazzaz FA (2000) Biotic invasions: causes, epidemiology, global consequences, and control. Ecological Applications 10: 689–710. https://doi.org/10.1890/1051-0761(2000)010[0689:BICEGC]2.0.CO;2
- MacKenzie KM, Palmer MR, Moore A, Ibbotson AT, Beaumont WRC, Poulter DJS, Trueman CN (2011) Locations of marine animals revealed by carbon isotopes. Scientific Reports 1: 21. https://doi.org/10.1038/srep00021
- Madigan D, Baumann Z, Carlisle A, Snodgrass O, Dewar H, Fisher N (2017) Isotopic insights into migration patterns of Pacific bluefin tuna in the eastern Pacific Ocean. Canadian Journal of Fisheries and Aquatic Sciences 75: 260–270. https://doi.org/10.1139/ cjfas-2016-0504
- Mazumder D, Johansen MP, Fry B, Davis E (2018) Muscle and carapace tissue–diet isotope discrimination factors for the freshwater crayfish *Cherax destructor*. Marine and Freshwater Research 69: 56–65. https://doi.org/10.1071/MF16360
- McFadden MS, Topham P, Harlow PS (2017) A ticking time bomb: is the illegal pet trade a pathway for the establishment of corn snake (*Elaphe guttata*) populations in Australia? Australian Zoologist 38: 499–504. https://doi.org/10.7882/AZ.2017.006
- Mitura A, Niemczuk K, Zareba K, Zajac M, Laroucau K, Szymanska-Czerwinska M (2017) Free-living and captive turtles and tortoises as carriers of new *Chlamydia* spp. PLOS ONE 12: 15. https://doi.org/10.1371/journal.pone.0185407
- Mo M (2019) Red-eared Sliders *Trachemys scripta elegans* in southern Sydney, including new incursions. Australian Zoologist 40: 314–325. https://doi.org/10.7882/AZ.2018.022
- Natusch DJD, Carter JF, Aust PW, Tri NV, Tinggi U, Mumpuni, Riyanto A, Lyons JA (2017) Serpent's source: determining the source and geographic origin of traded python skins using isotopic and elemental markers. Biological Conservation 209: 406–414. https://doi. org/10.1016/j.biocon.2017.02.042

- Nicholson G, Holmes C (2017) A note on statistical repeatability and study design for highthroughput assays. Statistics in Medicine 36: 790–798. https://doi.org/10.1002/sim.7175
- O'Keeffe S (2005) Investigating in conjecture: eradicating the red-eared slider in Queensland. In: 13th Australasian Vertebrate Pest Conference. Te papa, Wellington, New Zealand, 169–176 pp.
- Pearson SH, Avery HW, Kilham SS, Velinsky DJ, Spotila JR (2013) Stable isotopes of C and N reveal habitat dependent dietary overlap between native and introduced turtles *Pseudemys rubriventris* and *Trachemys scripta*. PLOS ONE 8: e62891. https://doi.org/10.1371/ journal.pone.0062891
- R Core Team (2019) R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.
- Reed RN, Krysko, KL (2014) Chapter 28 Invasive and introduced reptiles and amphibians. In: Mader DR, Divers SJ (Eds) Current Therapy in Reptile Medicine and Surger. W.B. Saunders, St. Louis, 304–309. https://doi.org/10.1016/B978-1-4557-0893-2.00028-4
- Reston Stable Isotope Laboratory (2011) Report of stable isotopic composition. Reference material USGS41. United States Geological Survey, Reston, Virginia.
- Robey J, Burgin S, Hitchen DJ, Ross G (2011) Status of an urban feral red-eared slider (*Trachemys scripta elegans*) population in Sydney a decade on. Australian Zoologist 35: 822–825. https://doi.org/10.7882/AZ.2011.033
- Rodder D, Schmidtlein S, Veith M, Lotters S (2009) Alien invasive slider turtle in unpredicted habitat: a matter of niche shift or of predictors studied? PLOS ONE 4: e7843. https://doi.org/10.1371/journal.pone.0007843
- Russello MA, Avery ML, Wright TF (2008) Genetic evidence links invasive monk parakeet populations in the United States to the international pet trade. Bmc Evolutionary Biology 8: 11. https://doi.org/10.1186/1471-2148-8-217
- Schmidt B, Stevenson K, Bloink C (2017) Surveys for the smooth newt (*Lissotriton vulgaris*) in south-east Melbourne. Ecology Australia Pty Ltd, Melbourne.
- Schneider L, Eggins S, Maher W, Vogt RC, Krikowa F, Kinsley L, Eggins SM, Da Silveira R (2015) An evaluation of the use of reptile dermal scutes as a non-invasive method to monitor mercury concentrations in the environment. Chemosphere 119: 163–170. https://doi. org/10.1016/j.chemosphere.2014.05.065
- Schoeninger MJ, DeNiro MJ (1984) Nitrogen and carbon isotopic composition of bone collagen from marine and terrestrial animals. Geochimica et Cosmochimica Acta 48: 625–639. https://doi.org/10.1016/0016-7037(84)90091-7
- Symes C, Skhosana F, Butler M, Gardner B, Woodborne S (2017) Isotope (delta C-13, delta N-15, delta H-2) diet-tissue discrimination in African grey parrot *Psittacus erithacus*: implications for forensic studies. Isotopes in Environmental and Health Studies 53: 580–596. https://doi.org/10.1080/10256016.2017.1319832
- Telecky TM (2001) United States import and export of live turtles and tortoises, 8–13.
- Therneau T, Atkinson B, Ripley B, Ripley MB (2018) rpart: Recursive Partitioning and Regression Trees.
- Trevor H, Robert T, JH F (2009) The elements of statistical learning: data mining, inference, and prediction. New York, NY: Springer.

- Vall-Ilosera M, Cassey P (2017) Leaky doors: private captivity as a prominent source of bird introductions in Australia. PLOS ONE 12: e0172851. https://doi.org/10.1371/journal. pone.0172851
- van Schingen M, Ziegler T, Boner M, Streit B, Nguyen TQ, Crook V, Ziegler S (2016) Can isotope markers differentiate between wild and captive reptile populations? A case study based on crocodile lizards (*Shinisaurus crocodilurus*) from Vietnam. Global Ecology and Conservation 6: 232–241. https://doi.org/10.1016/j.gecco.2016.03.004
- Ziegler S, Giesen K, van Schingen M, Rauhaus A, Ziegler T (2018) Testing the applicability of 15 N isotopic marker in skin tissue to distinguish between captive and wild monitor lizards, 73–83.

Supplementary material I

Tables S1.1, S1.2. A detailed description of indexes used for calculating confidence of status of *Trachemys scripta elegans individuals*

Authors: Katherine G. W. Hill, Kristine E. Nielson, Jonathan J. Tyler, Francesca A. McInerney, Zoe A. Doubleday, Greta J. Frankham, Rebecca N. Johnson, Bronwyn M. Gillanders, Steven Delean, Phill Cassey

Data type: species data

Copyright notice: This dataset is made available under the Open Database License (http://opendatacommons.org/licenses/odbl/1.0/). The Open Database License (ODbL) is a license agreement intended to allow users to freely share, modify, and use this Dataset while maintaining this same freedom for others, provided that the original source and author(s) are credited.

Link: https://doi.org/10.3897/neobiota.59.53671.suppl1

Supplementary material 2

Table S2.1; Figure S1. Explanation of methods for determining the optimal sampling size and design, using a power analysis on pilot data

Authors: Katherine G. W. Hill, Kristine E. Nielson, Jonathan J. Tyler, Francesca A. McInerney, Zoe A. Doubleday, Greta J. Frankham, Rebecca N. Johnson, Bronwyn M. Gillanders, Steven Delean, Phill Cassey

Data type: statistical data

Copyright notice: This dataset is made available under the Open Database License (http://opendatacommons.org/licenses/odbl/1.0/). The Open Database License (ODbL) is a license agreement intended to allow users to freely share, modify, and use this Dataset while maintaining this same freedom for others, provided that the original source and author(s) are credited.

Link: https://doi.org/10.3897/neobiota.59.53671.suppl2
Supplementary material 3

Figure S1. Results from a carbon decision tree

Authors: Katherine G. W. Hill, Kristine E. Nielson, Jonathan J. Tyler, Francesca A. McInerney, Zoe A. Doubleday, Greta J. Frankham, Rebecca N. Johnson, Bronwyn M. Gillanders, Steven Delean, Phill Cassey

Data type: statistical data

Copyright notice: This dataset is made available under the Open Database License (http://opendatacommons.org/licenses/odbl/1.0/). The Open Database License (ODbL) is a license agreement intended to allow users to freely share, modify, and use this Dataset while maintaining this same freedom for others, provided that the original source and author(s) are credited.

Link: https://doi.org/10.3897/neobiota.59.53671.suppl3

Supplementary material 4

Table S4.1. Determining confidence in status assignment

Authors: Katherine G. W. Hill, Kristine E. Nielson, Jonathan J. Tyler, Francesca A. McInerney, Zoe A. Doubleday, Greta J. Frankham, Rebecca N. Johnson, Bronwyn M. Gillanders, Steven Delean, Phill Cassey

Data type: statistical data

Copyright notice: This dataset is made available under the Open Database License (http://opendatacommons.org/licenses/odbl/1.0/). The Open Database License (ODbL) is a license agreement intended to allow users to freely share, modify, and use this Dataset while maintaining this same freedom for others, provided that the original source and author(s) are credited.

Link: https://doi.org/10.3897/neobiota.59.53671.suppl4



A workflow for standardising and integrating alien species distribution data

Hanno Seebens¹, David A. Clarke², Quentin Groom³, John R. U. Wilson^{4,5}, Emili García-Berthou⁶, Ingolf Kühn^{7,8,9}, Mariona Roigé¹⁰, Shyama Pagad¹¹, Franz Essl¹², Joana Vicente¹³, Marten Winter⁹, Melodie McGeoch²

I Senckenberg Biodiversity and Climate Research Centre, Senckenberganlage 25, 60325 Frankfurt, Germany 2 School of Biological Sciences, Monash University, Clayton 3800, VIC, Australia 3 Meise Botanic Garden, Meise, Belgium 4 Centre for Invasion Biology, Department of Botany and Zoology, Stellenbosch University, South Africa 5 South African National Biodiversity Institute, Kirstenbosch Research Centre, Cape Town, South Africa 6 GRECO, Institute of Aquatic Ecology, University of Girona, 17003 Girona, Spain 7 Helmholtz Centre for Environmental Research – UFZ, Department of Community Ecology, Theodor-Lieser-Str. 4, 06120 Halle, Germany 8 Martin Luther University Halle-Wittenberg, Geobotany and Botanical Garden, Am Kirchweg 2, 06108 Halle, Germany 9 German Centre for Integrative Biodiversity Research (iDiv) Halle-Jena-Leipzig, Deutscher Platz 5e, 04103 Leipzig, Germany 10 AgResearch, Biocontrol and Biosecurity, Private Bag 4749, Christchurch 8140, New Zealand 11 IUCN Species Survival Commission Invasive Species Specialist Group (ISSG), University of Auckland, Auckland 1072, New Zealand 12 Department of Botany and Biodiversity and Genetic Resources (CIBIO) / InBIO Research Network in Biodiversity and Evolutionary Biology, Campus Agrário de Vairão, Rua Padre Armando Quintas nº 7, 4485-641, Vairão, Vila do Conde, Portugal

Corresponding author: Hanno Seebens (hanno.seebens@senckenberg.de)

Academic editor: Maud Bernard-Verdier | Received 23 April 2020 | Accepted 3 July 2020 | Published 28 July 2020

Citation: Seebens H, Clarke DA, Groom Q, Wilson JRU, García-Berthou E, Kühn I, Roigé M, Pagad S, Essl F, Vicente J, Winter M, McGeoch M (2020) A workflow for standardising and integrating alien species distribution data. NeoBiota 59: 39–59. https://doi.org/10.3897/neobiota.59.53578

Abstract

Biodiversity data are being collected at unprecedented rates. Such data often have significant value for purposes beyond the initial reason for which they were collected, particularly when they are combined and collated with other data sources. In the field of invasion ecology, however, integrating data represents a major challenge due to the notorious lack of standardisation of terminologies and categorisations, and the application of deviating concepts of biological invasions. Here, we introduce the SINAS workflow, short for Standardising and Integrating Alien Species data. The SINAS workflow standardises terminologies following Darwin Core, location names using a proposed translation table, taxon names based on the GBIF backbone taxonomy, and dates of first records based on a set of predefined rules. The output of the SINAS

Copyright Hanno Seebens et al. This is an open access article distributed under the terms of the Creative Commons Attribution License (CC BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

workflow provides various entry points that can be used both to improve coherence among the databases and to check and correct the original data. The workflow is flexible and can be easily adapted and extended to the needs of different users. We illustrate the workflow using a case-study integrating five widely used global databases of information on biological invasions. The comparison of the standardised databases revealed a surprisingly low degree of overlap, which indicates that the amount of data may currently not be fully exploited in the original databases. We highly recommend the use and development of publicly available workflows to ensure that the integration of databases is reproducible and transparent. Workflows, such as SInAS, ultimately increase trust in data, study results, and conclusions.

Keywords

databases, Darwin Core, GBIF, invasive alien species, R software environment, reproducibility, standardisation, taxonomy, workflow

Introduction

In recent years, we have observed a tremendous rise in the availability of data in all fields of biodiversity research (La Salle et al. 2016), including invasion ecology. In particular, initiatives have emerged to map the occurrence of specific taxa with alien populations – called 'alien taxa' in the following – for major groups such as plants, birds, amphibians and reptiles (van Kleunen et al. 2015; Dyer et al. 2017a; Capinha et al. 2017); to assess the extent of invasions in particular geographical regions (e.g., Europe, DAISIE 2009) and habitats (e.g., marine, Ahyong et al. 2019); to document particular events (e.g., dates of record, Seebens et al. 2017); or to identify and record the presence of alien species that have negative impacts (e.g., Pagad et al. 2018). Although analyses of these data sources have led to valuable insights on the historic and current spatial and temporal patterns and processes of biological invasions (Dyer et al. 2017a; Dawson et al. 2017; Pyšek et al. 2017; Bertelsmeier et al. 2017; Seebens et al. 2018), these new aggregations of alien species data differ in various respects and are not interoperable.

Biodiversity data sources are often not standardised or directly comparable (Guralnick et al. 2018), which limits their value for conservation and research (Bayraktarov et al. 2019). In invasion ecology, new databases have recently been produced for a range of different purposes, although they have, to date, been produced largely in isolation. To remedy this, individual workflows have been created to harmonise and integrate the information in order to meet particular project goals. These workflows have used different taxonomic and geographical standards and practices, but such standardisations are not always clearly documented. As a result, databases are often not comparable and cannot be readily linked, which hampers progress towards improving the taxonomic and geographic coverage of alien species data and potential insights for research and management that might be derived as a consequence (McGeoch et al. 2012). The widespread lack of standardisation across key data sources on alien species also hinders clear communication with managers and policy makers (Gatto et al. 2013; McGeoch and Jetz 2019).

Progress in biodiversity research has been facilitated by the development of data standards (Guralnick and Hill 2009), powerful analytical tools and coherent work-

flows to, for instance, develop and calculate Essential Biodiversity Variables (EBVs, Kissling et al. 2018; Jetz et al. 2019) or to clean biodiversity data (Mathew et al. 2014; Jin and Yang 2020). Recently, using three exemplar alien species, a workflow was constructed and tested to integrate data from multiple sources for alien species (Hardisty et al. 2019). For most comprehensive databases in invasion ecology, the publication of such workflows and detailed descriptions of database generation remains rare (but see Dyer et al. 2017b; Pagad et al. 2018). Thus, data management in invasion ecology does not often meet open science principles, and the databases produced do not qualify as FAIR, i.e. Findable, Accessible, Interoperable, and Reusable (Wilkinson et al. 2016). Although the procedures for collating data are often described, the descriptions and associated metadata are generally insufficient for the workflow to be reproduced. Computer scripts and guidance documents are often not publicly available, which further impedes reproducibility. Using a standardised, publicly available workflow would enable alien species databases to be combined in a transparent and repeatable way, and improve the format, contents, and interoperability of databases (Mathew et al. 2014). Such annotated workflows would also guide future data collation efforts such that they achieve both their own goals and contribute to community-wide efforts to enhance the quality and quantity of data on alien and invasive species (Hobern et al. 2019). In particular, any integration of species databases requires a well-documented, repeatable, coherent, and standardised workflow to match nomenclature and taxonomy based on a standard concept (e.g., Boyle et al. 2013; Murray et al. 2017), or even to map different taxonomic concepts to each other (Berendsohn 1995). The availability of large online infrastructures for biodiversity research, such as the Global Biodiversity Information Facility (GBIF), enables taxonomic standardisation in a reproducible and standardised way, but the potential is still not fully exploited in studies addressing biological invasions.

Here, we introduce the SInAS (Standardising and Integrating Alien Species data) workflow that was developed within the course of the synthesis working group "Theory and Workflows for Alien and Invasive Species Tracking" (sTWIST) at sDiv, Leipzig, Germany. Following Hardisty and Roberts (2013), we use the term "workflow" as a description of a series of processes of data manipulation and integration, including the codes allowing a largely automated approach (see also van der Aalst and van Hee 2002, who use the term "workflow" for a series of standardised processes). The SInAS workflow serves to integrate databases of regional checklists including information on spatial and temporal dynamics of alien species using a standardised protocol to merge taxon and location names. The SInAS workflow combines public taxonomic infrastructures with procedures, resolutions, and concepts commonly used in biodiversity research in general and invasion ecology in particular. In the following, we provide a detailed description of the SInAS workflow and its implementation in R. We demonstrate its functionality using an example of merging five of the most comprehensive open access alien species databases currently available. Although the workflow was developed for merging databases of alien species occurrences, it can be readily adapted to other databases, including those with associated spatial information.

The SInAS workflow

The SInAS workflow was created to integrate databases organised as individual spreadsheet tables, which is the most common format for alien species occurrence information. In contrast to databases of native species, alien species occurrences are often associated with a date of first introduction or first date of report for a region as an alien or naturalised species. Here, we adopt a common use of these "first records", which represent the first record of a taxon in a particular region. Following Darwin Core terminology (Darwin Core Task Group 2009), first records are called "event dates" in the following.

Three major steps, organised in sequence, form the primary components of the workflow: 1) initial check and preparation of the original databases; 2) standardisation of the databases; and 3) merging of the standardised databases (Fig. 1). Standardisation (step 2) is the most complex step and can be subdivided into specific tasks that each involves the standardisation of one of eight variables: taxon names, location names, event dates, occurrence status, establishment means, degree of establishment, pathway, and habitat. An overview of all variables used in this workflow together with definitions and explanations are given in Suppl. material 2: Tables S1–S4. Each specific task requires a reference against which data will be standardised (e.g., a list of location names in a particular format or a list of accepted taxon names and their synonyms). Each task produces intermediate output tables to report where there was standardisation (e.g., replacements of original names) and where standardisation was not possible (e.g., missing names and unresolved names). As input files, each step of the workflow requires the output of the previous step as input except for step one, where the original database and its metadata have to be provided (currently implemented as *.xlsx files). In the following section, a comprehensive overview of the SInAS workflow is provided, while the detailed description can be found in the Suppl. material 1. The full workflow implemented in R together with all required input files, examples databases, and a manual are provided as the SInAS workflow package (see section 'Data and code availability' below).

Step 1: Preparation of databases

The first step includes a check of the availability of variables in the original databases. Variables are categorised into three classes: i) required variables, which must be provided (i.e., taxon and location names); ii) optional variables, which are associated to the taxon occurrence (e.g., occurrence status or pathway) or represent entries potentially useful for data standardisation (e.g., extra taxonomic information); and iii) additional variables, which are not used within the workflow, but are retained as presented in the original databases throughout standardisation (e.g., traits). An overview of variables and definitions is provided in Suppl. material 2: Table S1. The column names of the required and optional variables in the input databases are harmonised.



Figure 1. Overview of the Standardising and Integrating Alien Species data (SInAS) workflow that can be used to merge alien species databases. The workflow consists of three consecutive steps: 1. preparation of databases, 2. standardisation, and 3. merging. The standardisation step is subdivided into the standardisation of: 2a. terminology, 2b. location names, 2c. taxon names, and 2d. event dates (i.e., first records). The user can modify the workflow by adjusting the reference tables under 'user-defined input'. At each step of standardisation, changes and missing entries are exported as intermediate output that can be used to check the workflow, the reference tables, or the input data.

Step 2: Standardisation

2a: Terminology

Records of alien species are often associated with information about their occurrence status, the degree of establishment, and their pathway(s) of introduction. Such information is standardised in this step using translation tables (Suppl. material 1). Translation tables provide information about the entries in the original databases and the corresponding terms that are to be used in the merged database. These are part of the workflow package (see section 'Data and code availability' below), and follow the recommendations by Groom et al. (2019) in standardising the Darwin Core terms 'establishmentMeans', 'occurrenceStatus' and 'pathway', and adopting their suggestion to include a new term 'degreeOfEstablishment', describing the status of the taxon at a particular location (Suppl. material 2: Table S1). Strictly speaking, this status is not associated to a taxon, but a specific population. This means, as Colautti & MacIsaac (2004) already pointed out, that alien or nonindigenous species are misnomers and these attributes, frequently referred to simply as "status", are associated at population level (i.e., intersecting taxon name with locality). In databases covering large regions, such attributes must properly be assigned at the right level. However, to be comparable with the wealth of invasion literature that does not properly attribute "status", and for reasons of linguistic simplicity, we still refer to alien species rather than using the correct alien populations. Although the proposal by Groom et al. (2019) has not yet been ratified by the Biodiversity Information Standards organisation, we used it in the workflow as the proposed terminology covers dimensions critical to invasion biology, policy, and management (McGeoch and Jetz 2019), and thus will provide helpful information irrespective of its official incorporation into Darwin Core. The Darwin Core term 'habitat' is also standardised within the workflow; however, as a categorisation of different habitats is not provided by Darwin Core, we provide one in the respective translation table (Suppl. material 1) based on the distinction between terrestrial, freshwater, marine, and brackish habitats. The translation tables can be adjusted by the user in any way, but we highly recommend adhering to the proposed Darwin Core terminology to avoid having incomparable entries. Nonmatching terms are exported so they can be manually checked.

2b: Location names

Location names are standardised using a user-defined translation table (Suppl. material 1), which includes the master location names and the corresponding alternative formats, languages, and spellings. Locations represent administrative units such as countries, states or islands. The majority of location names (89%) conform to the 2-digit ISO code (ISO 3166-1 alpha-2) classification. For the remaining locations, countries were split into sub-national units which are geographically separated from each other (be they islands, states or mainland areas). For instance, Alaska, Hawaii, and US Minor Outlying Islands were separated from mainland United States; the Azores were distinguished from Portugal; and Tasmania from Australia. The full list of location names can be found in the input file "AllLocations.xlsx" as part of the workflow package. Altogether, we used a set of 262 non-overlapping locations covering the terrestrial surface of the world. Similar resolutions are used in many studies of biological invasions (Seebens et al. 2017; Capinha et al. 2017; Dyer et al. 2017b). The location categorisation can be easily adjusted to any spatial delineation in a user-friendly way by modifying the input file. Additional information for the location such as two- and three-digit ISO codes of countries, continents or the World Geographical Scheme for Recording Plant Distributions regions (WGSRPD, Brummitt 2001) are also provided. Non-matching location names are exported for reference. A shapefile is provided, which relates the location to georeferenced polygons for mapping.

2c: Taxon names

Taxonomic standardisation is one of the most important and challenging tasks in biodiversity data integration (Rees and Cranston 2017) as taxon names are often considered the fundamental unit to which other information types are linked (Patterson et al. 2010; Koch et al. 2018). This, however, necessitates the use of a taxonomic backbone against which all species names are assessed during the standardisation process. In the absence of a single authoritative nomenclature across all taxa (Bánki et al. 2018), we used the GBIF taxonomic backbone, which is itself primarily based on the Catalogue of Life (Bánki et al. 2018) (43 % overlap of GBIF backbone taxonomy and Catalogue of Life at the time of access) and complemented with 50+ other sources of taxonomic information. The details of these taxonomic sources can be found at the GBIF Secretariat (2019) and the full taxonomy is available for download (http://rs.gbif.org/datasets/backbone/). If the taxon name could be found in GBIF either as an exact match, a synonym or a fuzzy match with a high confidence (see Suppl. material 1), the obtained 'accepted taxon name' according to GBIF, as well as its given synonym and further taxonomic information, are returned and stored. Taxon names identified as synonyms according to GBIF are replaced with the accepted name obtained from GBIF. To avoid mismatches due to spelling errors, GBIF performs fuzzy matching of the full taxon names. This involves a calculation of similarity between the provided taxon names and the record provided by GBIF. GBIF returns the result of fuzzy matching by the summary metric "confidence", which involves cross-checks of taxon names, authorities and taxonomic information with different weightings (see http://www.gbif.org/developer/ species#searching for more details). In addition to the taxon names, the taxonomic tree (species, genus, family, order, class, phylum, and kingdom) is obtained from GBIF. In the SInAS workflow, all taxon names that could not be resolved are exported as a list of missing taxon names for further reference. A complete list of all taxon names (including the original names provided in the individual databases, taxonomic information, taxonomic status of the name, and search results) is exported as a separate list of taxon names (Suppl. material 1). The user can provide a list of species names and synonyms to resolve conflicts and errors in GBIF entries.

2d: Event dates

In the SInAS workflow presented here, event dates represent the time of the first documented occurrence of a species in a region outside its native range, which is also called 'first record' (Seebens et al. 2017). Ideally, event dates for the first record of an alien species are provided as a single year, which is then retained in the workflow. But often other time ranges are provided. To enable merging and cross-checking of first records among databases and further analysis, it is necessary to translate these different time ranges into single years. Such an adjustment of first records requires a set of rules (e.g., Seebens et al. 2017; Dyer et al. 2017b), which define how a time range should be treated to obtain a single year. In the simplest case, the start and the end years of the time range are provided, and their arithmetic mean is used as the new single event date. In other cases, time ranges are described in alternative ways such as "1920ies" or "<1920". In translating this information, we followed primarily the rules defined in table 3 of Dyer et al. (2017b). The rules are currently provided as a textual description and the user has to "translate" non-standard event dates into a single year format according to the guidelines and examples provided in the file 'Guidelines_eventDate.xlsx' as part of the workflow package. The user has the opportunity to modify the rules, but we recommend sticking to the proposed ones as a standard in biological invasions. Cases of entries that could not be adjusted are exported from the workflow for cross-checking.

Step 3: Merging

In the final step of the workflow, the standardised databases are merged into a single master database. Merging is based on the entries of taxon and location names. That is, all entries with exactly the same taxon and location name will be merged to obtain a single entry for each existing combination of taxon and location. This is achieved by first merging columns of the standardised databases to concatenate their contents and, second, by merging rows of the final database to remove duplicate entries. Conflicts of multiple event dates for the same event are resolved by adopting the earlier of the first records. In cases where conflicts cannot be resolved, the respective entries of all databases are combined to a single entry of the master database. For instance, if a taxon X in location Y is classified as 'introduced' in one database and 'uncertain' in another, the entry in the final master database for X in Y will be 'introduced; uncertain'. The user will be informed that conflicts still exist, which might be solved by adjusting the translation tables or by checking the original data.

In principle, the SInAS workflow is fully automated once metadata are provided at step 1. This, however, requires accepting all defaults such as location names and taxonomic classification by GBIF and, more importantly, keeping all unresolved conflicts that might include unmatched location names or misspellings in the original data. We therefore recommend running the workflow in an iterative process of running the workflow, checking warnings and intermediate output tables, resolving conflicts and errors, and re-running the workflow. Such an iterative process should increase the match between databases, and therefore the coverage of the final merged database.

A case study

We applied and tested the workflow using five global databases of spatio-temporal alien species occurrences (Table 1): three with a taxonomic focus, one each on alien birds (GAVIA, Dyer et al. 2017b), vascular plants (GloNAF, van Kleunen et al. 2019), and amphibians and reptiles (AmphRep, Capinha et al. 2017); one multi-taxon database with a focus on temporal dynamics (FirstRecords version 1.2, Seebens et al. 2017); and one with a focus on alien species with negative environmental or socio-economic impacts, i.e. "invasive alien species" (GRIIS, Pagad et al. 2018; accessed 10th September 2019). These databases are currently among the most up-to-date and comprehensive global data sources for alien species distributions, dynamics, and impacts. All databases are publicly available. The lack of accessibility impeded the incorporation of other global databases such as the World Register of Marine Introduced Species (WRiMS) or the CABI Invasive Species Compendium. The databases used here are of varying size, ranging from 1,118 (AmphRep) to 232,042 (GloNAF) records and including 277 (AmphRep) to 33,687 (GRIIS) taxa. The databases have different spatial resolutions and follow different taxonomic standards. Variables from the different databases were mapped onto the variables provided in the SInAS workflow as outlined in Suppl. material 2: Tables S1–S4. As location names were

Table 1. The taxonomic coverage and size of the original databases on the occurrence of alien taxa before and after standardisation and merging using the Standardising and Integrating Alien Species data (SInAS) workflow (see Figure 1). Records were counted multiple times when they were obtained from different databases. Reductions in total record number were mostly a result of aggregation from the finer spatial resolution of the original databases to the higher spatial resolution used in the SInAS workflow.

Database	Reference	Focus of database	Total records		Number of taxa	
			(original)	(merged)	(original)	(merged)
GloNAF	van Kleunen et al. (2019)	Vascular plants	232,042	71,468	14,053	13,545
AmphRep	Capinha et al. (2017)	Amphibians, reptiles	1,118	854	277	276
GAVIA	Dyer et al. (2017b)	Birds	27,723	4,494	971	968
GRIIS	Pagad et al. (2018)	Invasive species	107,302	96,655	33,687	27,128
FirstRecords	ords Seebens et al. (2017) First records		45,402	45,060	15,231	14,990

provided in different columns in GloNAF and GAVIA, these were merged manually to obtain a better match with the classification of locations used in the SInAS workflow.

Merging of the five databases resulted in a new database (the sTWIST database) consisting of two interlinked tables containing records of alien species per location and a full list of taxa including further taxonomic information (Suppl. material 3). Depending on the success of the integration of the specific databases, several additional files will be created during the workflow providing missing taxa and location names, unresolved terms (e.g., of occurrence status and pathways), translated location names and event dates, and unresolved event dates. In our cases, 17 of these tables were exported from the workflow for further cross-checking (Suppl. material 5) together with 25 tables, which include the output of each intermediate step and database (Suppl. material 4). The sTWIST database contains 156,900 records of 35,150 taxa in 257 locations. The resulting alien species numbers globally are in line with the reported hotspots of biological invasions being the USA (excluding Hawaii and Alaska), the United Kingdom, New Zealand, Hawaii, and Australia (fig. 2, Dawson et al. 2017). One consequence of the workflow was that, after cleaning and standardisation, the number of records dropped (Table 1). For example, the merged sTWIST database contained only ~30% of the original GloNAF database. This was mostly due to the GloNAF database having a finer spatial resolution than the sTWIST database (1,029 vs. 257 regions). Consequently, many regions were combined and records merged.

Altogether, 53,546 taxon names were obtained from all five databases, including synonyms and multiple entries of individual taxa due to different spellings. A small proportion (5 %) of these taxon names could not be found in GBIF for different reasons such as misspellings, missing information or unresolved taxonomies. This often involved subspecies, varieties or hybrids and can be checked in the output files "Missing_Taxa_*" for the individual databases. Most of these unresolved taxon names were obtained from GRIIS (1,610; 6 % of GRIIS taxa) followed by FirstRecords (802; 5%), AmphRep (10; 4%), GloNAF (261; 2%) and GAVIA (8; <1%). Unresolved taxon names "Taxa_FullList.csv". Standardisation during the SInAS workflow identified 7,174 syn-

Table 2. Overlap (in %) of locations, taxa, and taxa by location record between taxonomic and crosstaxon databases. An overlap between two databases is defined as the number of entries in the taxon-specific database shared with the cross-taxon database divided by the total number of entries from the taxonspecific database. It therefore shows how many records of the taxon-specific databases are found in the cross-taxon ones.

	GRIIS	FirstRecords
Locations		
GloNAF	76	97
GAVIA	76	98
AmphRep	74	98
Taxa		
GloNAF	69	45
GAVIA	54	86
AmphRep	61	63
Taxa by location		
GloNAF	44	20
GAVIA	26	78
AmphRep	29	41



Figure 2. The number of alien taxa per region as presented in the final sTWIST database. Smaller island regions are depicted by circles, with the size of the circles proportional to the numbers of taxa. Region delineations are based on Global Administrative Areas (GADM).

onyms (13%), which were replaced by the accepted names provided by GBIF. This finally reduced the number of taxa to 35,150 distinct taxon names.

After standardisation of taxon and location names, the overlap of taxon-specific databases with the cross-taxon ones was surprisingly low (Table 2). Most regions were represented in all databases; however, the overlaps for taxa and taxon by location combinations were often far below 50%. For instance, only 26% of all records in GAVIA can also be found in GRIIS, while 20% of the GloNAF records were also included in FirstRecords. The comparatively low overlap of locations in GRIIS with taxon-specific databases stems from a few locations only considered separately in GRIIS.

The SInAS workflow is, to the best of our knowledge, the most comprehensive workflow to standardise and integrate alien species occurrence databases to date. It is also in full compliance with the FAIR data principles (Wilkinson et al. 2016). The workflow provides a foundation to develop and apply standards for the harmonisation of taxon names, geographic resolutions, and event dates. It achieves this using translation tables and rules that are transparent and linked to existing international schemes such as accepted taxonomic backbones that can be easily updated as needed. The SInAS workflow also offers the opportunity to adapt individual steps to the respective user's needs, and enables the user to conveniently report on deviations from the suggested workflow. Reporting of such adjustments is essential for reproducibility, particularly in the field of invasion ecology, which is rich in competing concepts and terminologies (Falk-Petersen et al. 2006). Thus, the SInAS workflow will help to differentiate and integrate the various approaches, and finally will increase trust not only in data but also in study results and conclusions communicated to the decision makers and the general public (Franz and Sterner 2018). The potential to customise and extend the workflow increases the range of possible applications such as the calculation of indicators (e.g., Wilson et al. 2018), the ability to conduct global and regional assessments of invasive alien species and their control, and the global collaboration being proposed as essential for dealing with priority invaders (Blackburn et al. 2020).

We introduced the SInAS workflow as a tool to integrate databases, but it can also assist with standardisation within a database to ensure that region or taxon names are consistent, and that terminologies of individual checklists are reported in a more standardised way. Although the flexibility built into the SInAS workflow makes it more broadly useful, providing flexibility in a workflow does bear the risk that databases remain incompatible. For instance, users of the workflow can define their own categorisation of locations, which might result in even more heterogeneous databases in addition to those that already exist. It is essential, therefore, that modifications of the workflow are clearly communicated. As best practice, we recommend that modifications of the input files such as translation tables, taxon names or any modification of the workflow itself are clearly reported and published together with the final database. For instance, a change in the list of geographic regions can be easily attached as a table to the respective publication together with the link to our workflow. In this way, modifications can be traced back to their origin and databases remain comparable despite adaptations to individual project goals. We believe that our proposed workflow will smooth this process and make it easier for individual researchers to publish not only scientific results in a more consistent way, but also the underlying workflows to enhance the transparency and reproducibility of the science.

The comparison of the individual databases that resulted from the integration work done here highlighted an unexpectedly low degree of overlap between them. This re-emphasizes, in spite of significant recent advances in alien species data collation, the importance of: 1) joint collaborative work, 2) freely available data, and 3) shared workflows to improve the taxonomic, geographic, and temporal coverage and resolution of alien species data (Hardisty et al. 2019). The low degree of overlap was obviously related to the scope of the individual databases – the taxon-specific databases focussed on a high level of spatial and taxonomic coverage, while cross-taxonomic databases harvest information on a specific topic such as event dates or impact. Moreover, the databases drew original data records from different sources, and so each database was constructed using different workflows with divergent assumptions and supporting concepts. This clearly shows that not only does the merging of individual databases have to be standardised as proposed here, but the integration of primary data from the original sources needs to be done in a more reproducible and transparent way as well (Vanderhoeven et al. 2017; Pagad et al. 2018). Our case study also highlights that the SInAS workflow and the associated scripts could be used to assess the reliability of different databases and their components (e.g., Cano-Barbacil et al. 2020) and to identify potential areas of improvement for the respective databases.

Our workflow was developed to integrate taxon lists for individual regions, so-called checklists. Checklists represent by far the most common representation of spatial information on alien species occurrences (Pyšek et al. 2012; Brundu and Camarda 2013). This is somewhat different to other fields of biodiversity research, where occurrence data are often provided as range maps, grids, plot based lists or point coordinates. In contrast to populations of native taxa, alien taxa populations are categorised as being alien only for a particular region and timeframe. The importance of decision-making in an applied science, such as invasion ecology, means that policies are commonly made for the administrative units (such as countries or states/provinces) responsible for control efforts, and the spatial resolution of presence-absence data is low resolution to accommodate both uncertainty and the precautionary principle when data are intended to inform policy and management. As a consequence, the decision of what is considered as being alien is often taken for administrative regions. This is somewhat different for aquatic alien species, which are categorised depending on marine regions or water sheds, but these spatial units can be easily incorporated as additional entries in the table of geographic regions. In its current form, the SInAS workflow is not capable of handling coordinate-based occurrences. While including point-wise occurrences might be possible in future versions of the workflow, a practical solution would be to assign the coordinate-based location to a region and add the region to the workflow. For example, point-wise occurrence data for the Western Mediterranean Sea could be attributed to this region and added to the workflow.

The pervasive challenge in the integration of alien species data from multiple sources is the variability in the use of terminology (McGeoch et al. 2012). For example, the term 'invasive species' has at least three working definitions: alien populations that are self-sustaining and have naturally spread; alien populations that negatively impact native species, ecosystems, the economy or human health; or populations (be they native or alien) that have recently increased in abundance or extent (Richardson et al. 2000; Blackburn et al. 2011; Carey et al. 2012). As a consequence, merging databases that use different definitions of alien and invasive alien species could result in a misleading collation of taxa. Currently, terminologies are not consistently used across databases, although standard concepts have been published (Blackburn et al. 2011). In the SInAS workflow, we provide a translation of terms following common standards (Darwin Core Task Group 2009; Groom et al. 2019), but the definitions of these terms may vary among primary sources and projects, which often cannot be standardised ret-

rospectively. It is therefore essential to stick to common definitions and transparent workflows already in the primary literature, to clearly specify which definition is used.

A further difficulty in combining species data lies in the application of different taxonomic concepts (Berendsohn 1995) by the data recorders. This is a general problem in biodiversity and taxonomic research and is not solved within the SInAS workflow: it requires collaborative solutions from the relevant research community. While resolving such taxonomic conflicts would mean the SInAS workflow is more useful, one should keep in mind that a complete taxonomic resolution is not necessarily required to provide useful information (Gerwing et al. 2020). Unless this workflow is used by experienced taxonomists for taxonomic resolution, we recommend sticking to standards offered by other authorities such as GBIF and report deviations from these standards. Our workflow eases this reporting process by providing the opportunity to submit information of modifications together with the databases.

While advancements have been made in other fields of biodiversity research, with online platforms such as GBIF including a full and citable version control, many databases on biological invasions are still curated by individuals or research groups and might not be publicly available at all. Changing this situation will require there being: 1) an incentive for researchers to publish their data online, ideally with a digital object identifier (DOI) and versioning as provided by online platforms such as GBIF or long-term archives such as Zenodo (https://zenodo.org/) or Dryad (https://datadryad.org), and following the FAIR principles of data management; 2) professional training and technical support for data management; and 3) clear guidelines and standards to ease such data publications (Groom et al. 2019). For some of these aspects, support is already available but still not widely adopted such as the "Guide to Data Management in Ecology and Evolution" published by the British Ecological Society (2014). For other aspects, financial and personnel support is required as individual researchers often do not have the capacity to ensure long-term maintenance and support, which can only be achieved from institutions. The importance of adopting the FAIR data principles has been increasingly recognised by international institutions such as the Intergovernmental Science-Policy Platform of Biodiversity and Ecosystem Services [IPBES, currently conducting a thematic assessment on invasive alien species and their control (https://ipbes.net/invasivealien-species-assessment) that depends on the integration of data sources as we have discussed here] and the European Commission, which provide incentives to scientists to make their data comparable and available. We believe the workflow presented here addresses these challenges by providing an example of how to achieve standardisation across databases and to facilitate the kind of standardisation chosen by the researchers.

The modular structure of the SInAS workflow means that it can form the basis for the development of future data integration workflows. We foresee several opportunities for extensions. Translation tables of additional variables such as taxon traits and variables related to regions and relevant for understanding drivers of biological invasions (environmental, socio-economic, historic) would add another level of value for both research and application. The workflow could also be extended to allow for coordinatebased occurrence records by integrating information of region delineations using Geographic Information System (GIS) tools. Thus, the SInAS workflow, focussed as it is on essential variables for tracking biological invasions (distribution, time, and impact, Latombe et al. 2017), can be considered the core of an integrated comprehensive work-flow of data on biological invasions. Global collaborative efforts, supported by readily accessible, globally representative evidence, are key to stemming the invasion tide.

Data and code availability

The full SInAS workflow including all required R scripts, input files, example databases and a manual is made freely available at a repository at Zenodo (https://doi.org/10.5281/ zenodo.3944432) together with the coordinate-based delineations of regions. The releases at Zenodo are linked to a GitHub repository, which ensures full version control of the code. New releases will be provided under the same DOI. All additional files related to the case study are attached to this publication as supplementary materials.

Acknowledgements

This paper is a joint effort of the sTWIST working group (Theory and Workflows for Invasive Species Tracking) supported by sDiv, the Synthesis Centre of iDiv (DFG FZT 118 – 202548816). It is a contribution to the Species Populations Working Group of the Group on Earth Observations Biodiversity Observation Network (GEO BON; https://geobon. org/ebvs/workinggroups/species-populations). We thank Wolfgang Traylor for advice on structuring the R code, Carlos Eduardo Arlé Ribeiro de Souza for providing the shapefile and Gabriele Rada for support on graphic design. Support from the following funding agencies is acknowledged: HS – Belmont Forum-BiodivERsA project AlienScenarios through the national funders German Federal Ministry of Education and Research (BMBF; grant 01LC1807A). MAM – Australian Research Council (DP200101680). FE – BiodivERsA-Belmont Forum Project AlienScenarios (FWF project no I 4011-B32). JRUW - South African Department of Forestry, Fisheries and the Environment (DFFtE) for funding noting that this publication does not necessarily represent the views or opinions of DFFtE or its employees. DAC – Australian Government Research Training Program (RTP) scholarship. EGB - Spanish Ministry of Science and Innovation (projects CGL2016-80820-R, PCIN-2016-168 and RED2018-102571-T) and the Government of Catalonia (ref. 2017 SGR 548). QG – Belgian Science Policies Brain program (BR/165/A1/TrIAS).

References

- Ahyong S, Costello MJ, Galil BS, et al. (2019) World Register of Introduced Marine Species (WRiMS).
- Bánki O, Döring M, Holleman A, Addink W (2018) Catalogue of Life Plus: innovating the CoL systems as a foundation for a clearinghouse for names and taxonomy. Biodiversity Information Science and Standards 2: e26922. https://doi.org/10.3897/biss.2.26922

- Bayraktarov E, Ehmke G, O'Connor J, et al. (2019) Do big unstructured biodiversity data mean more knowledge? Frontiers in Ecology and Evolution 7: 1–5. https://doi.org/10.3389/ fevo.2019.00319
- Berendsohn WG (1995) The concept of "potential taxa" in databases. Taxon 44: 207–212. https://doi.org/10.2307/1222443
- Bertelsmeier C, Ollier S, Liebhold A, Keller L (2017) Recent human history governs global ant invasion dynamics. Nature Ecology & Evolution 1: 0184. https://doi.org/10.1038/ s41559-017-0184
- Blackburn GS, Bilodeau P, Cooke T, Cui M, Cusson M, Hamelin RC, Keena MA, Picq S, Roe AD, Shi J, Wu Y, Porth I (2020) An Applied Empirical Framework for Invasion Science: Confronting Biological Invasion Through Collaborative Research Aimed at Tool Production. Annals of the Entomological Society of America. https://doi.org/10.1093/aesa/saz072
- Blackburn TM, Pyšek P, Bacher S, Carlton JT, Duncan RP, Jarošík V, Wilson JRU, Richardson DM (2011) A proposed unified framework for biological invasions. Trends in Ecology & Evolution 26: 333–339. https://doi.org/10.1016/j.tree.2011.03.023
- Boyle B, Hopkins N, Lu Z, Garay JAR, Mozzherin D, Rees T, Matasci N, Narro ML, Piel WH, Mckay SJ, Lowry S, Freeland C, Peet RK, Enquist BJ (2013) The taxonomic name resolution service: an online tool for automated standardization of plant names. BMC Bioinformatics 14: 1–16. https://doi.org/10.1186/1471-2105-14-16
- British Ecological Society (2014) A Guide to Data Management in Ecology and Evolution. https:// www.britishecologicalsociety.org/wp-content/uploads/Publ_Data-Management-Booklet.pdf
- Brummitt RK (2001) World Geographical Scheme for Recording Plant Distributions. Hunt Institute for Botanical Documentation, Carnegie Mellon University, Pittsburgh.
- Brundu G, Camarda I (2013) The Flora of Chad: a checklist and brief analysis. PhytoKeys 23: 1–18. https://doi.org/10.3897/phytokeys.23.4752
- Cano-Barbacil C, Radinger J, García-Berthou E (2020) Reliability analysis of fish traits reveals discrepancies among databases. Freshwater Biology 65: 863–877. https://doi.org/10.1111/ fwb.13469
- Capinha C, Seebens H, Cassey P, García-Díaz P, Lenzner B, Mang T, Moser D, Pyšek P, Rödder D, Scalera R, Winter M, Dullinger S, Essl F (2017) Diversity, biogeography and the global flows of alien amphibians and reptiles. Diversity and Distributions 23: 1313–1322. https://doi.org/10.1111/ddi.12617
- Carey MP, Sanderson BL, Barnas KA, Olden JD (2012) Native invaders challenges for science, management, policy, and society. Frontiers in Ecology and the Environment 10: 373–381. https://doi.org/10.1890/110060
- Colautti RI, MacIsaac HJ (2004) A neutral terminology to define 'invasive' species. Diversity and Distributions 10: 135–141. https://doi.org/10.1111/j.1366-9516.2004.00061.x
- DAISIE (2009) Handbook of Alien Species in Europe. Springer, Dordrecht.
- Darwin Core Task Group (2009) Darwin Core (Kampmeier G, review manager) Biodiversity Information Standards (TDWG). http://www.tdwg.org/standards/450
- Dawson W, Moser D, van Kleunen M, Kreft H, Pergl J, Pyšek P, Weigelt P, Winter M, Lenzner B, Blackburn TM, Dyer EE, Cassey P, Scrivens SL, Economo EP, Guénard B, Capinha C, Seebens H, García-Díaz P, Nentwig W, García-Berthou E, Casal C, Mandrak NE, Fuller P, Meyer C, Essl F (2017) Global hotspots and correlates of alien species richness

across taxonomic groups. Nature Ecology & Evolution 1: 0186. https://doi.org/10.1038/ s41559-017-0186

- Dyer EE, Cassey P, Redding DW, Collen B, Franks V, Gaston KJ, Jones KE, Kark S, Orme CDL, Blackburn TM (2017a) The Global Distribution and Drivers of Alien Bird Species Richness. PLoS Biology 15: e2000942. https://doi.org/10.1371/journal.pbio.2000942
- Dyer EE, Redding DW, Blackburn TM (2017b) The global avian invasions atlas, a database of alien bird distributions worldwide. Scientific Data 4: 170041. https://doi.org/10.1038/sdata.2017.41
- Falk-Petersen J, Bøhn T, Sandlund OT (2006) On the Numerous Concepts in Invasion Biology. Biological Invasions 8: 1409–1424. https://doi.org/10.1007/s10530-005-0710-6
- Franz NM, Sterner BW (2018) To increase trust, change the social design behind aggregated biodiversity data. Database 2018: 1–12. https://doi.org/10.1093/database/bax100
- Gatto F, Katsanevakis S, Vandekerkhove J, Zenetos A, Cardoso AC (2013) Evaluation of Online Information Sources on Alien Species in Europe: The Need of Harmonization and Integration. Environmental Management 51: 1137–1146. https://doi.org/10.1007/s00267-013-0042-8
- GBIF Secretariat (2019) GBIF Backbone Taxonomy.
- Gerwing TG, Cox K, Allen Gerwing AM, Campbell L, Macdonald T, Dudas SE, Juanes F (2020) Varying intertidal invertebrate taxonomic resolution does not influence ecological findings. Estuarine, Coastal and Shelf Science 232: 106516. https://doi.org/10.1016/j. ecss.2019.106516
- Groom Q, Desmet P, Reyserhove L, Adriaens T, Oldoni D, Vanderhoeven S, Baskauf SJ, Chapman A, McGeoch M, Walls R, Wieczorek J, Wilson JRU, Zermoglio PFF, Simpson A (2019) Improving Darwin Core for research and management of alien species. Biodiversity Information Science and Standards 3: e38084. https://doi.org/10.3897/biss.3.38084
- Guralnick R, Hill A (2009) Biodiversity informatics: automated approaches for documenting global biodiversity patterns and processes. Bioinformatics 25: 421–428. https://doi. org/10.1093/bioinformatics/btn659
- Guralnick R, Walls R, Jetz W (2018) Humboldt Core toward a standardized capture of biological inventories for biodiversity monitoring, modeling and assessment. Ecography 41: 713–725. https://doi.org/10.1111/ecog.02942
- Hardisty A, Roberts D (2013) A decadal view of biodiversity informatics: challenges and priorities. BMC Ecology 13: 1–16. https://doi.org/10.1186/1472-6785-13-16
- Hardisty AR, Belbin L, Hobern D, McGeoch MA, Pirzl R, Williams KJ, Kissling WD (2019) Research infrastructure challenges in preparing essential biodiversity variables data products for alien invasive species. Environmental Research Letters 14: 025005. https://doi. org/10.1088/1748-9326/aaf5db
- Hobern D, Baptiste B, Copas K, Guralnick R, Hahn A, van Huis E, Kim E-S, McGeoch M, Naicker I, Navarro L, Noesgaard D, Price M, Rodrigues A, Schigel D, Sheffield CA, Wieczorek J (2019) Connecting data and expertise: a new alliance for biodiversity knowledge. Biodiversity Data Journal 7: e33679. https://doi.org/10.3897/BDJ.7.e33679
- Jetz W, McGeoch MA, Guralnick R, Ferrier S, Beck J, Costello MJ, Fernandez M, Geller GA, Keil P, Merow C, Meyer C, Muller-Karger FE, Pereira HM, Regan EC, Schmeller DS, Turak E (2019) Essential biodiversity variables for mapping and monitoring species populations. Nature Ecology and Evolution 3: 539–551. https://doi.org/10.1038/s41559-019-0826-1

- Jin J, Yang J (2020) BDcleaner: A workflow for cleaning taxonomic and geographic errors in occurrence data archived in biodiversity databases. Global Ecology and Conservation 21: e00852. https://doi.org/10.1016/j.gecco.2019.e00852
- Kissling WD, Ahumada JA, Bowser A, Fernandez M, Fernández N, García EA, Guralnick RP, Isaac NJB, Kelling S, Wouter L, McRae L, Mihoub J-B, Obst M, Santamaria M, Skidmore AK, Williams KJ, Donat A, Amariles D, Arvanitidis C, Bastin L, De Leo F, Willi E, Elith J, Hobern D, Martin D, Pereira HM, Pesole G, Peterseil J, Saarenmaa H, Schigel D, Schmeller DS, Segata N, Turak E, Uhlir PF, Wee B, Hardisty AR (2018) Building essential biodiversity variables (EBVs) of species distribution and abundance at a global scale. Biological Reviews 93: 600–625. https://doi.org/10.1111/brv.12359
- Koch MA, German DA, Kiefer M, Franzke A (2018) Database Taxonomics as Key to Modern Plant Biology. Trends in Plant Science 23: 4–6. https://doi.org/10.1016/j. tplants.2017.10.005
- La Salle J, Williams KJ, Moritz C (2016) Biodiversity analysis in the digital era. Philosophical Transactions of the Royal Society. https://doi.org/10.1098/rstb.2015.0337
- Latombe G, Pyšek P, Jeschke JM, Blackburn TM, Bacher S, Capinha C, Costello MJ, Fernández M, Gregory RD, Hobern D, Hui C, Jetz W, Kumschick S, McGrannachan C, Pergl J, Roy HE, Scalera R, Squires ZE, Wilson JRU, Winter M, Genovesi P, McGeoch MA (2017) A vision for global monitoring of biological invasions. Biological Conservation 213: 295–308. https://doi.org/10.1016/j.biocon.2016.06.013
- Mathew C, Güntsch A, Obst M, Vicario S, Haines R, Williams AR, de Jong Y, Goble C (2014) A semi-automated workflow for biodiversity data retrieval, cleaning, and quality control. Biodiversity Data Journal 2: e4221. https://doi.org/10.3897/BDJ.2.e4221
- McGeoch M, Jetz W (2019) Measure and Reduce the Harm Caused by Biological Invasions. One Earth 1: 171–174. https://doi.org/10.1016/j.oneear.2019.10.003
- McGeoch MA, Spear D, Kleynhans EJ, Marais E (2012) Uncertainty in invasive alien species listing. Ecological Applications 22: 959–971. https://doi.org/10.1890/11-1252.1
- Murray BR, Martin LJ, Phillips ML, Pyšek P (2017) Taxonomic perils and pitfalls of dataset assembly in ecology: a case study of the naturalized Asteraceae in Australia. NeoBiota 34: 1–20. https://doi.org/10.3897/neobiota.34.11139
- Pagad S, Genovesi P, Carnevali L, Schigel D, McGeoch MA(2018) Introducing the Global Register of Introduced and Invasive Species. Scientific Data 5: 170202. https://doi. org/10.1038/sdata.2017.202
- Patterson DJ, Cooper J, Kirk PM, Pyle RL, Remsen DP (2010) Names are key to the big new biology. Trends Ecol Evol 25: 686–691. https://doi.org/10.1016/j.tree.2010.09.004
- Pyšek P, Danihelka J, Sádlo J, Jr JC (2012) Catalogue of alien plants of the Czech Republic: checklist update, taxonomic diversity and invasion patterns. Preslia 84: 155–255.
- Pyšek P, Pergl J, Essl F, Lenzner B, Dawson W, Kreft H, Weigelt P, Winter M, Kartesz J, Nishino M, Antonova LA, Barcelona JF, Cabezas FJ, Cárdenas D, Cárdenas-Toro J, Castaño N, Chacón E, Chatelain C, Dullinger S, Ebel AL, Figueiredo E, Fuentes N, Genovesi P, Groom QJ, Henderson L, Inderjit, Kupriyanov A, Masciadri S, Maurel N, Meerman J, Morozova O, Moser D, Nickrent D, Nowak PM, Pagad S, Patzelt A, Pelser PB, Seebens H, Shu W, Thomas J, Velayos M, Weber E, Wieringa JJ, Baptiste M, van Kleunen M (2017) Naturalized alien flora of the world: species diversity, taxonomic and phylogenetic patterns,

geographic distribution and global hotspots of plant invasion. Preslia 89: 203–274. https://doi.org/10.23855/preslia.2017.203

- Rees J, Cranston K (2017) Automated assembly of a reference taxonomy for phylogenetic data synthesis. Biodiversity Data Journal 5: e12581. https://doi.org/10.3897/BDJ.5.e12581
- Richardson DM, Pyšek P, Rejmanek M, Barbour MG, Panetta FD, West CJ (2000) Naturalization and invasion of alien plants: concepts and definitions. Diversity and Distributions 6: 93–107. https://doi.org/10.1046/j.1472-4642.2000.00083.x
- Seebens H, Blackburn TM, Dyer EE, Genovesi P, Hulme PE, Jeschke JM, Pagad S, Pyšek P, Winter M, Arianoutsou M, Bacher S, Blasius B, Brundu G, Capinha C, Celesti-Grapow L, Dawson W, Dullinger S, Fuentes N, Jäger H, Kartesz J, Kenis M, Kreft H, Kühn I, Lenzner B, Liebhold A, Mosena A, Moser D, Nishino M, Pearman D, Pergl J, Rabitsch W, Rojas-Sandoval J, Roques A, Rorke S, Rossinelli S, Roy HE, Scalera R, Schindler S, Štajerová K, Tokarska-Guzik B, van Kleunen M, Walker K, Weigelt P, Yamanaka T, Essl F (2017) No saturation in the accumulation of alien species worldwide. Nature Communications 8: 14435. https://doi.org/10.1038/ncomms14435
- Seebens H, Blackburn TM, Dyer EE, Genovesi P, Hulme PE, Jeschke JM, Pagad S, Pyšek P, van Kleunen M, Winter M, Ansong M, Arianoutsou M, Bacher S, Blasius B, Brockerhoff EG, Brundu G, Capinha C, Causton CE, Celesti-Grapow L, Dawson W, Dullinger S, Economo EP, Fuentes N, Guénard B, Jäger H, Kartesz J, Kenis M, Kühn I, Lenzner B, Liebhold AM, Mosena A, Moser D, Nentwig W, Nishino M, Pearman D, Pergl J, Rabitsch W, Rojas-Sandoval J, Roques A, Rorke S, Rossinelli S, Roy HE, Scalera R, Schindler S, Štajerová K, Tokarska-Guzik B, Walker K, Ward DF, Yamanaka T, Essl F (2018) Global rise in emerging alien species results from increased accessibility of new source pools. Proceedings of the National Academy of Sciences 115: E2264–E2273. https://doi.org/10.1073/pnas.1719429115
- van der Aalst W, van Hee KM (2002) Workflow Management: Models, Methods, and Systems. MIT Press Cambridge, London. https://doi.org/10.7551/mitpress/7301.001.0001
- van Kleunen M, Dawson W, Essl F, Pergl J, Winter M, Weber E, Kreft H, Weigelt P, Kartesz J, Nishino M, Antonova LA, Barcelona JF, Cabezas FJ, Cárdenas D, Cárdenas-Toro J, Castaño N, Chacón E, Chatelain C, Ebel AL, Figueiredo E, Fuentes N, Groom QJ, Henderson L, Inderjit, Kupriyanov A, Masciadri S, Meerman J, Morozova O, Moser D, Nickrent DL, Patzelt A, Pelser PB, Baptiste MP, Poopath M, Schulze M, Seebens H, Shu W-S, Thomas J, Velayos M, Wieringa JJ, Pyšek P (2015) Global exchange and accumulation of non-native plants. Nature 525: 100–103. https://doi.org/10.1038/nature14910
- van Kleunen M, Pyšek P, Dawson W, et al. (2019) The Global Naturalized Alien Flora (GloN-AF) database. Ecology 100: e02542.
- Vanderhoeven S, Adriaens T, Desmet P, Strubbe D, Backeljau T, Barbier Y, Brosens D, Cigar J, Coupremanne M, De Troch R, Eggermont H, Heughebaert A, Hostens K, Huybrechts P, Jacquemart A-L, Lens L, Monty A, Paquet J-Y, Prévot C, Robertson T, Termonia P, Van De Kerchove R, Van Hoey G, Van Schaeybroeck B, Vercayie D, Verleye TJ, Welby S, Groom QJ (2017) Tracking Invasive Alien Species (TrIAS): Building a datadriven framework to inform policy. Research Ideas and Outcomes 3: e13414. https://doi.org/10.3897/rio.3.e13414
- Wilkinson MD, Dumontier M, Aalbersberg IJ, Appleton G, Axton M, Baak A, Blomberg N, Boiten J-W, da Silva Santos LB, Bourne PE, Bouwman J, Brookes AJ, Clark T, Crosas M,

Dillo I, Dumon O, Edmunds S, Evelo CT, Finkers R, Gonzalez-Beltran A, Gray AJG, Groth P, Goble C, Grethe JS, Heringa J, Hoen TAC, Hooft R, Kuhn T, Kok R, Kok J, Lusher SJ, Martone ME, Mons A, Packer AL, Persson B, Rocca-Serra P, Roos M, van Schaik R, Sansone S-A, Schultes E, Sengstag T, Slater T, Strawn G, Swertz MA, Thompson M, van der Lei J, van Mulligen E, Velterop J, Waagmeester A, Wittenburg P, Wolstencroft K, Zhao J, Mons B (2016) The FAIR Guiding Principles for scientific data management and stewardship. Scientific Data 3: 160018. https://doi.org/10.1038/sdata.2016.18

Wilson JRU, Faulkner KT, Rahlao SJ, Richardson DM, Zengeya TA, van Wilgen BW (2018) Indicators for monitoring biological invasions at a national level. Journal of Applied Ecololy 55: 2612–2620. https://doi.org/10.1111/1365-2664.13251

Supplementary material I

Technical description and manual of the SInAS workflow implementation in R

Authors: Hanno Seebens, David A. Clarke, Quentin Groom, John R. U. Wilson, Emili García-Berthou, Ingolf Kühn, Mariona Roigé, Shyama Pagad, Franz Essl, Joana Vicente, Marten Winter, Melodie McGeoch

Data type: text

- Explanation note: This document contains a detailed description of the implementation of the SInAS workflow in R and its application.
- Copyright notice: This dataset is made available under the Open Database License (http://opendatacommons.org/licenses/odbl/1.0/). The Open Database License (ODbL) is a license agreement intended to allow users to freely share, modify, and use this Dataset while maintaining this same freedom for others, provided that the original source and author(s) are credited.

Link: https://doi.org/10.3897/neobiota.59.53578.suppl1

Supplementary material 2

Supplementary Tables S1–S4

Authors: Hanno Seebens, David A. Clarke, Quentin Groom, John R. U. Wilson, Emili García-Berthou, Ingolf Kühn, Mariona Roigé, Shyama Pagad, Franz Essl, Joana Vicente, Marten Winter, Melodie McGeoch

Data type: tables

- Explanation note: Tables S1–S4 provide descriptions of variables used in the SInAS workflow and how these were mapped on the databases of the case study.
- Copyright notice: This dataset is made available under the Open Database License (http://opendatacommons.org/licenses/odbl/1.0/). The Open Database License (ODbL) is a license agreement intended to allow users to freely share, modify, and use this Dataset while maintaining this same freedom for others, provided that the original source and author(s) are credited.

Link: https://doi.org/10.3897/neobiota.59.53578.suppl2

Supplementary material 3

Final output files of the case study applying the SInAS workflow

Authors: Hanno Seebens, David A. Clarke, Quentin Groom, John R. U. Wilson, Emili García-Berthou, Ingolf Kühn, Mariona Roigé, Shyama Pagad, Franz Essl, Joana Vicente, Marten Winter, Melodie McGeoch

Data type: tables

- Explanation note: The zip contains the final output files of the application of the SInAS workflow in the case study. It includes the merged database, a full list of taxon names and the translated location names and event dates (first records).
- Copyright notice: This dataset is made available under the Open Database License (http://opendatacommons.org/licenses/odbl/1.0/). The Open Database License (ODbL) is a license agreement intended to allow users to freely share, modify, and use this Dataset while maintaining this same freedom for others, provided that the original source and author(s) are credited.

Link: https://doi.org/10.3897/neobiota.59.53578.suppl3

Supplementary material 4

Intermediate output files of the case study applying the SInAS workflow

Authors: Hanno Seebens, David A. Clarke, Quentin Groom, John R. U. Wilson, Emili García-Berthou, Ingolf Kühn, Mariona Roigé, Shyama Pagad, Franz Essl, Joana Vicente, Marten Winter, Melodie McGeoch

Data type: tables

Explanation note: The zip file contains all intermediate output files, which represent the output of each individual step of the SInAS workflow applied in the case study.

Copyright notice: This dataset is made available under the Open Database License (http://opendatacommons.org/licenses/odbl/1.0/). The Open Database License (ODbL) is a license agreement intended to allow users to freely share, modify, and use this Dataset while maintaining this same freedom for others, provided that the original source and author(s) are credited.

Link: https://doi.org/10.3897/neobiota.59.53578.suppl4

Supplementary material 5

Unresolved entries of the case study applying the SInAS workflow

Authors: Hanno Seebens, David A. Clarke, Quentin Groom, John R. U. Wilson, Emili García-Berthou, Ingolf Kühn, Mariona Roigé, Shyama Pagad, Franz Essl, Joana Vicente, Marten Winter, Melodie McGeoch

Data type: tables

- Explanation note: The zip file contains all files with unresolved records such as unmatched taxon names, missing location names, unresolved event dates or missing terms from the application of the workflow in the case study. These files could be used for cross-checking and correction errors and mis-matches to improve the final output.
- Copyright notice: This dataset is made available under the Open Database License (http://opendatacommons.org/licenses/odbl/1.0/). The Open Database License (ODbL) is a license agreement intended to allow users to freely share, modify, and use this Dataset while maintaining this same freedom for others, provided that the original source and author(s) are credited.

Link: https://doi.org/10.3897/neobiota.59.53578.suppl5

RESEARCH ARTICLE



Origin of climatic data can determine the transferability of species distribution models

Arunava Datta^{1,2,3}, Oliver Schweiger¹, Ingolf Kühn^{1,4,5}

I Department of Community Ecology, Helmholtz Centre for Environmental Research–UFZ, Theodor-Lieser-Straße 4, D-06120 Halle, Germany 2 Centre for Invasion Biology, Department of Botany and Zoology, Stellenbosch University, Matieland, South Africa 3 South African National Biodiversity Institute, Kirstenbosch National Botanical Gardens, Claremont, South Africa 4 Institute of Biology/Geobotany and Botanical Garden, Martin Luther University Halle-Wittenberg, Am Kirchtor 1, D-06108 Halle, Germany 5 German Centre for Integrative Biodiversity Research (iDiv) Halle-Jena-Leipzig, Deutscher Platz 5e, 04103 Leipzig, Germany

Corresponding author: Arunava Datta (arunava.datta@ufz.de)

Academic editor: M. Rejmanek	Received 19 May 2019	Accepted 31 May 2020	Published 28 July 2020
------------------------------	----------------------	----------------------	------------------------

Citation: Datta A, Schweiger O, Kühn I (2020) Origin of climatic data can determine the transferability of species distribution models. NeoBiota 59: 61–76. https://doi.org/10.3897/neobiota.59.36299

Abstract

Methodological research on species distribution modelling (SDM) has so far largely focused on the choice of appropriate modelling algorithms and variable selection approaches, but the consequences of choosing amongst different sources of environmental data has scarcely been investigated. Bioclimatic variables are commonly used as predictors in SDMs. Currently, several online databases offer the same sets of bioclimatic variables, but they differ in underlying source of raw data and method of data processing (extrapolation and downscaling). In this paper, we asked whether predictive performance and spatial transferability of SDMs are affected by the choice of two different bioclimatic databases viz. WorldClim 2 and Chelsa 1.2. We used presence-absence data of the invasive plant Ageratina adenophora from the Western Himalaya for training SDMs and a set of independently-collected presence-only datasets from the Central and Eastern Himalaya to evaluate the transferability of the SDMs beyond the training range. We found that the performance of SDMs was, to a large degree, affected by the choice of the climatic dataset. Models calibrated on Chelsa 1.2 outperformed WorldClim 2 in terms of internal evaluation on the calibration dataset. However, when the model was transferred beyond the calibration range to the Central and Eastern Himalaya, models based on WorldClim 2 performed substantially better. We recommend that, in addition to the choice of predictor variables, the choice of predictor datasets with these variables should not be based merely on subjective decision whenever several options are available. Instead, such decisions should be based on robust evaluation of the most appropriate dataset for a given geographic region and species being modelled. Moreover, decisions could also depend on the objective of the study, i.e. projecting within the calibration range or beyond. Therefore, a quantitative evaluation of predictor datasets from alternative sources should be routinely performed as an integral part of the modelling procedure.

Copyright Arunava Datta et al. This is an open access article distributed under the terms of the Creative Commons Attribution License (CC BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Keywords

Ageratina adenophora, climatic database, invasive species, model transfer, species distribution modelling

Introduction

Correlative species distribution models (SDMs, also referred to as ecological niche models or habitat suitability models) are used to estimate the potential geographic distribution of species by modelling the relationship between known occurrences of a species with its environmental conditions (Guisan and Zimmermann 2000; Pearson and Dawson 2003; Elith and Leathwick 2009). These models directly relate the occurrence of a species to its realised multi-dimensional niche (Hutchinson 1957; Pearson and Dawson 2003) in the environmental space (Soberón and Nakamura 2009; Peterson et al. 2011) that is provided by the chosen predictor variables. Climatic conditions are crucial in determining the large-scale distribution patterns of organisms (Woodward 1987; Woodward et al. 2004) and are hence widely used for modelling species distributions (Pearson and Dawson 2003).

SDMs are frequently applied in invasion biology, conservation biology, evolutionary biology and agriculture due to their versatility (Elith and Leathwick 2009; Peterson et al. 2011). SDMs of invasive species are often used to make temporal and spatial predictions of climatically-suitable regions that could potentially be invaded (Thuiller et al. 2005; Ervin and Holly 2011; Jaryan et al. 2013) and thus aid in early detection, control and eradication of the invasive species (Thuiller et al. 2005; Peterson et al. 2011). The distribution of invasive plants will most likely change due to climate change and therefore future projections of invasion from SDMs will further help in taking long-term management decisions (Thuiller et al. 2005; Peterson et al. 2011).

To avoid misleading recommendations for such management decisions, SDMs and the resulting predictions or future projections of suitable environmental conditions and corresponding invasion risks need to be highly reliable. Much of past research has focused on the development of modelling algorithms and model (i.e. variable) selection to increase the performance of SDMs (Guisan and Zimmermann 2000; Elith and Leathwick 2009). Ample studies are available on different methodological aspects, such as the choice of different modelling algorithms, sample size, sample density, variable selection and spatial resolution of environmental layers on model accuracy and transferability (Randin et al. 2006; Peterson and Nakazawa 2008; Heikkinen et al. 2012; Wenger and Olden 2012).

Model transferability, either in space or time (Randin et al. 2006; Elith and Leathwick 2009), is of particular importance for invasive species to reliably assess their response to climate change or to predict their invasive potential in novel areas and for corresponding management decisions (Clark et al. 2001; Yates et al. 2018). Therefore, it is essential to assess the predictive accuracy of an SDM, not only within the region in which it was fitted (i.e. internal validation within the calibration range), but also in a geographic region different from the calibration range (i.e. external validation on an independent dataset) (Heikkinen et al. 2012; Wenger and Olden 2012; Fernández and Hamilton 2015). The model transfer may often involve extrapolation if the ranges of the predictors are beyond the calibration range of the model. Model transferability is a particularly challenging issue in species distribution modelling (Araújo and Guisan 2006; Elith and Leathwick 2009; Peterson et al. 2011; Wenger and Olden 2012). A recent review on challenges in transferability of ecological models has flagged many pertinent issues, such as the choice of response variables, sampling bias, choice of modelling algorithm and non-stationarity etc. (Yates et al. 2018).

It has also been shown that the choice of predictor variables can impact model accuracy and transferability (Bobrowski et al. 2017; Karger et al. 2017; Petitpierre et al. 2017), but studies, focusing exclusively on the consequences of choosing different sources providing the same set of predictor variables, are very scarce (Peterson et al. 2011). Consequently, researchers often rely on their subjective decisions for choosing one source of predictor datasets over others, even if the same set of (potential predictor) variables are available from different sources.

SDMs have increasingly benefitted from the availability of climatic predictors at very high resolutions in the form of rasterised GIS layers available from different sources (Soberón and Nakamura 2009; Peterson et al. 2011). Despite offering the same variables, such different climatic databases could differ in their actual values since they rely on different source data and use different interpolation or downscaling algorithms (Bobrowski and Schickhoff 2017; Karger et al. 2017). Such differences could be particularly relevant in regions of high orographic heterogeneity, which have been shown to be highly sensitive to prediction errors for multiple plant species (Hanspach et al. 2011).

The most widely-used variables for SDMs are the set of 19 bioclimatic variables (Peterson and Nakazawa 2008; O'Donnell and Ignizio 2012) that do not only include annual averages, but also climatic extremes limiting the physiological performance of biological organisms (O'Donnell and Ignizio 2012). Currently, several databases offer free access to these bioclimatic variables. WorldClim was one of the first and most frequently used high resolution (30 arc seconds) global bioclimatic dataset derived from ground weather stations across the globe and interpolated by using latitude, longitude and elevation as independent variables (Hijmans et al. 2005). In the recent version of WorldClim (Version 2; Fick and Hijmans 2017), hereafter referred to as WorldClim 2, satellite-derived covariates, such as land surface temperature and cloud cover, have also been used in the interpolation process to improve the data quality in areas where ground observations are scarce. Chelsa (Version 1.2; Karger et al. 2017), hereafter referred to as Chelsa 1.2, is another bioclimatic database that accounts for orographic patterns of precipitation in mountainous terrains, i.e. it accounts for factors, such as aspect and valley exposition by including wind effects (see Karger et al. 2017). Therefore, it can be assumed that, due to the methodological differences in generating the raster layers, these databases are not equivalent and hence their use in SDMs could result in differences in predictive accuracy and, moreover, in transferability.

In this paper, we asked, whether models calibrated on Chelsa 1.2 and WorldClim 2, respectively, differ in terms of internal and external predictive performance. To this end, we used the invasive plant species *Ageratina adenophora* (Spreng.) R.M.King &

H.Rob. in the Himalaya as our study system. Using presence-absence data of *A. ad-enophora* from the Western Himalaya as the response, we calibrated generalised linear models on Chelsa1.2 and WorldClim2 data. Transferability of models calibrated on these two datasets was evaluated using an independent set of presence-only data from Central and Eastern parts of the Himalaya.

Methods

Target species

Ageratina adenophora (Crofton weed, Asteraceae) is a plant species native to Mexico and invasive (or even noxious) in more than 30 countries in subtropical regions across the globe (Auld and Martin 1975; Qiang 1998; Tian et al. 2007; Muniappan et al. 2009; Poudel et al. 2019). It is a multi-stemmed, perennial herb or undershrub that grows up to 2 metres and flowers profusely in spring (Tripathi et al. 2012). It was introduced as an ornamental plant to England in the 19th century (Auld and Martin 1975) and was later introduced in different parts of the world (Muniappan et al. 2009), such as the Himalaya (Dehradun, India) in the early 20th century (Datta et al. 2017). In South Asia, it has expanded its distribution almost throughout the subtropical and sub-temperate belts of the Himalaya, ranging from Arunachal Pradesh in the east to Himachal Pradesh in the west (Raizada 1976; Tripathi et al. 2012) and also flourishes in mountains of peninsular India (Muniappan and Viraktamath 1993; Muniappan et al. 2009).

Study area and distribution survey

Our study was carried out in a region of the Western Himalaya (Singh and Singh 1987) between 29.96N and 32.55N latitudes and 75.77E and 78.43E longitudes. Our study area covered five provinces in north-western India and stretched from Dhauladhar range (Himachal Pradesh province) in the west to the mountains of Gharwal region (Uttrakhand province) in the east. We also covered a considerable part of low-lying foothills of the Himalaya (Siwalik range).

We haphazardly surveyed 389 locations and recorded the presence or absence of *A. adenophora* in the subtropical and temperate zones of the Western Himalaya between 300 m to 3000 m elevation (Fig.1). We targeted this elevational belt based on prior knowledge about the distribution of the plant from previous reconnaissance surveys and existing literature on its distribution (Datta et al. 2017). The surveys were conducted in the vegetation periods of 2014 and 2015. Most of the surveys were carried out along road- and riversides as these are conduits for dispersal of propagules and are also initial establishment sites of *A. adenophora* (Z. Lu and Ma 2006; Wang et al. 2011). However, many high elevational areas beyond 2500 m were not accessible by road and, hence, we used trekking trails for surveying such remote locations. Alpine



Figure 1. Survey locations of *Ageratina adenophora*. The region marked by the blue rectangle **a** shows the survey area in the Western Himalaya from which 192 presences (red circles) and 197 genuine absences (blue circles) were used to train the model. The region marked by the green rectangle **b** shows the Central and Eastern Himalaya from where an additional set of 85 presence only locations (green circles) were obtained for evaluating the transferability of the species distribution models trained in the Western Himalaya. The relief map of the region is depicted in brown. The relief map was made with layer obtained from Natural Earth and the international borders were digitized from political map of India (9th edition) published by survey of Inida..

and subalpine regions (> 3500 m) were not surveyed since the plant is known to be entirely absent from these regions due to extremely low temperatures (Datta et al. 2017). At the scale of the used climatic variables (30 arc seconds or 1 km²), microclimatic variations due to trails, roads and water conduits are not a limiting factor for the distribution of the species. Hence, this potential bias in data acquisition should not influence the model outcome and general conclusions. To assess model transferability, we used an independent set of presence-only records (N = 85) that were collected by experts (see acknowledgements) from Central and Eastern Himalaya (Fig. 1).

Climatic data and variable selection

Due to collinearity amongst the 19 bioclimatic variables, we used a cluster analysis to select variables seperately for WorldClim 2 and Chelsa 1.2 (Dormann et al. 2013). All the 19 bioclimatic variables were scaled to zero and unit standard deviation prior to

cluster analysis. The dendrogram was constructed, based on Spearman's rank correlation (ρ) using UPGMA (unweighted pair-group method with arithmetic averages) agglomeration method. A threshold value of of $\rho = |0.7|$ (Dormann et al. 2013) was used to prune the dendrogram and select variables that were not highly collinear. This procedure resulted in five clusters for WorldClim1.2 and seven clusters for Chelsa 2 (see Suppl. material 1). Selection of a variable within a cluster was primarily based on its ecological relevance to the study species. For example, the plant is known to be limited by low temperatures in higher elevations (Datta et al. 2017), therefore the minimum temperature of the coldest month was selected (bio 6). Similarly, germination of seeds is known to be limited by moisture in the lower elevations, hence precipitation of the driest month (bio 14) was preferred over other variables (Datta et al. 2017). In order to make the models based on WorldClim 2 and Chelsa 1.2 comparable, we ensured that the set of selected variables was common. However, due to inherent differences in the correlation sructure, the variable selection procedure yielded slightly different sets of variables for the two datasets. Finally, five variables were selected for WorldClim 2, while two additional variables were selected in the case of Chelsa 1.2 (see Table 1).

In addition to the two models based on WorldClim 2 and Chelsa 1.2 data, we calibrated a third model based on Chelsa 1.2 data, but using the same set of five variables that were selected specifically for WorldClim 2 (Table 1). This allowed us to make direct and unbiased comparison between the predictive performance of WorldClim 2 and Chelsa 1.2 and to assess whether our conclusions are potentially confounded by differences in model performance caused by the initial variable selection procedure.

Modelling procedure

We used a multi-model inference approach (Burnham and Anderson 2002) to arrive at the final model to be used for prediction (Grueber et al. 2011; Symonds and Moussalli 2011; Burnham 2015). The following steps were taken: (1) We fitted generalised linear models with binomial error distribution and a logit link function to the presence-absence data of A. adenophora in the Western Himalaya using previously selected climatic variables (Table 1). All predictor variables were scaled to zero mean and unit standard deviation. (2) We then obtained models with all possible variable combinations using the "dredge" function in the "MuMIn" package (Barton 2015) of R (R Core Team 2017). (3) A subset of best models that differed by 2 or less in AIC from the best model was considered for a model averaging process (hereafter referred to as "best subset") (Grueber et al. 2011). (4) We then averaged model coefficients, weighted by the corresponding Akaike weights across all models in the best subset. We used the default "full average" method for calculating the averaged coefficients (if a variable is absent from one of the component models, a parameter estimate of "zero" is substituted in the averaging process (Symonds and Moussalli 2011). This method results in shrinkage of parameter estimates for those variables which are less important (Grueber et al. 2011) and has been suggested when prediction from an averaged model is intended (Symonds and Moussalli 2011).

Table 1. Variable selection for Chelsa 1.2 and WorldClim 2 datasets using UPGMA cluster analysis to reduce collinearity amongst the variables. Highly correlated variables were removed from each dataset (using threshold of Spearman's $\rho = 0.7$, see text for details). The selected variables from Chesla 1.2 and WorldClim 2 are represented by tick mark (\checkmark) against the respective variable.

Climatic variable	Abbreviatio n	WorldClim2	Chelsa1.2
Isothermality	bio3	\checkmark	\checkmark
Temperature Seasonality	bio4		\checkmark
Min Temperature of Coldest Month	bio6	\checkmark	\checkmark
Temperature Annual Range	bio7		\checkmark
Annual Precipitation	bio12	\checkmark	\checkmark
Precipitation of Driest Month	bio14	\checkmark	\checkmark
Precipitation Seasonality	bio15	\checkmark	\checkmark

Model evaluation

To obtain binary predictions (i.e. presence or absence output) from continuous probability values, a threshold was selected by maximising the true skill statistic (TSS), which accounts for both omission and commission errors and is known to be independent of prevalence (Allouche et al. 2006). The value of TSS can range from -1 to +1. A value close to +1 indicates good agreement, while a value close to 0 indicates that the model does not perform better than a random model (Allouche et al. 2006). A value close to -1 suggests that a completely inverse model would be better. AUC is a common traditionally-used metric for evaluating the performance of SDMs; however, its efficiency has been questioned (Jiménez-Valverde et al. 2008) and, therefore, we do not report AUC values.

To assess the transferability (i.e. predictive performance of the model beyond the calibration area in the Western Himalaya), we used the independent set of presenceonly data from the Central and Eastern Himalaya (Nepal, Sikkim, Darjeeling and Bhutan; see acknowledgements for contributors). Since we did not have true absence data from these regions, we could not use ordinary model evaluation metrics such as TSS. Therefore, we used the Boyce Index for assessing transferability (Boyce et al. 2002; Hirzel et al. 2006). The Boyce Index compares the ratio of predicted frequency and expected frequency of evaluation points across the prediction gradient using a moving window approach (Hirzel et al. 2006; Petitpierre et al. 2012). It is a thresholdindependent metric ranging between -1 and +1. Positive values close to 1 indicate very good agreement of observed presences with the model prediction, while values very close to zero indicate that the predictions are not better than random. Negative values of the Boyce Index show that the model is worse than a random model and makes predictions in areas that are not suitable for the species (Hirzel et al. 2006). For this purpose, the region of evaluation was defined by drawing a convex hull around the presence-only evaluation points. The convex hull (polygon) was used to crop the prediction layer (raster) from the model. Subsequently, the predicted occurrence probabilities were used as a measure of "habitat suitability" (x-axis) and were correlated

(Spearman's correlation) with the "predicted to expected ratio" (y-axis) calculated from the presence-only evaluation points across the prediction gradient using the moving window approach (Hirzel et al. 2006).

The Boyce Index was calculated using the "ecospat.boyce" function of the "ecospat" package (Cola et al. 2017). The Boyce Index was also calculated for internal evaluation (i.e. training range) to facilitate direct comparison between Western and Central and Eastern Himalaya using presence only data.

Further, SDMs were projected to a much larger geographic area (entire South Asia) compared to the training area to allow for a general qualitative assessment (i.e. visual agreement), based on a priori knowledge about the distribution of *A. adenophora* from existing literature. R codes for the entire analysis can be found in Suppl. material 3.

Results

Here, we report the predictive performance of the three averaged models using the multimodel inference approach. The first model ("WorldClim data – WorldClim variable selection") had two component models (i.e. best subset of models that differed by 2 or less in AIC), the second model ("Chelsa data – Chelsa variable selection") had six component models, while the third model ("Chelsa data – WorldClim variable selection") had four component models. The average value of the coefficients for the bioclimatic variables also differed between the models (Suppl. material 2).

Internal evaluation of the models based on TSS, using presence-absence data, showed that Chelsa performed marginally better than WorldClim (Table 2). The "Chelsa data – Chelsa variable selection" had the highest TSS value amongst all models, while the "Chelsa data – WorldClim variable selection" performed similar to "World-Clim data – WorldClim variable selection" (Table 2). In contrast, internal evaluation using the Boyce Index (based on presence-only data) revealed that the performance of

Table 2. Model evaluation metrics for different models using Chelsa 1.2 and WorldClim 2 datasets. Database refers to the climatic database used for modelling (calibration). Variable selection refers to the specific set of variables selected using cluster analysis for Chelsa 1.2 and WorldClim 2 datasets (see Table 1 and method section for further details). Sensitivity is the rate of true positives while specificity is the rate of true negatives. Boyce internal refers to the Boyce Index calculated for the training area and Boyce external refers to the Boyce Index calculated for Central and Eastern Himalaya where the model was transferred to. Chelsa 1.2 and WorldClim 2 are written as Chelsa and WorldClim in the table.

		Internal evaluation				External evaluation			
Modelling	Variable	Thr	PCC	Sen	Spe	TSS	MSE	Boyce	Boyce index
database	selection							index	
WorldClim	WorldClim	0.69	0.76	0.6	0.92	0.52	0.24	0.61	0.64
Chelsa	Chelsa	0.46	0.81	0.76	0.86	0.62	0.19	0.59	-0.14
Chelsa	WorldClim	0.54	0.75	0.73	0.77	0.51	0.25	0.91	0.37

Thr: Threshold to translate continuous occurrence probabilities into presence/absence data; Sen: Sensitivity; Spe: Specificity; PCC: Percent correctly classified; TSS: True skill statistic; MSE: Mean square error.



Figure 2. Model projection in South Asia showing the continuous probabilities (left) and binarised prediction (right) from the models. Panel **a** and **b**: WorldClim 2 data and variables selected for WorldClim 2; panel **c** and **d**: Chelsa 1.2 data and variables selected for Chelsa 1.2; panel **e** and **f**: WorldClim 2 data but variables selected for Chelsa 1.2.

the Chelsa models was marginally lower than the WorldClim models, while "Chelsa data – WorldClim variable selection" had the highest Boyce Index (Table 2).

In contrast to internal model evaluation, transferability of the model beyond the calibration range in the Central and Eastern Himalaya was entirely based on the Boyce Index because we had only presence data from these regions. The Boyce Index was highest for the "WorldClim data – WorldClim variable selection" and was slightly negative for "Chelsa data – Chelsa variable selection". Negative value of Boyce's Index indicated that the model predicted high probability of occurrence even for regions that were almost unsuitable for the species.

The visual inspection of the prediction maps also showed that the "Chelsa data – Chelsa variable selection" model produced extremely unrealistic over-predictions (Fig. 2c). For instance, the model showed most parts of South Asia to be highly suitable for *A. adenophora*, including warm tropical regions of peninsular India. However, in reality, the species is known to be restricted to moist subtropical and temperate regions found at higher elevations (Muniappan and Viraktamath 1993).

To identify whether this over-prediction was simply caused by the selection of variables based on the Chelsa dataset, we also assessed the performance of the "Chelsa data – WorldClim variable selection" model. This increased model performance, measured with the Boyce Index, but stayed considerably below that of the "WorldClim data – WorldClim variable selection" model (Table 2). Further, transferability was slightly improved, although many potentially unsuitable regions in central and southern India were still being predicted as climatically suitable for *A. adenophora* (Fig. 2d).

Discussion

Using two openly-available bioclimatic datasets, we found that the choice of the climatic dataset had a substantial effect on transferability of SDMs in mountainous regions such as the Himalaya. It is interesting to note that, although the same set of five variables was used in the multimodel inference approach for "WorldClim data – WorldClim variable selection" and "Chelsa data – WorldClim variable selection" models, the number of component models in the "best subset" for "Chelsa data – WorldClim variable selection" was twice the number of models in "WorldClim data – WorldClim variable selection". The contribution of the variables in these two models also differed considerably. For example, in the "WorldClim data – WorldClim variables" model, bio15 was the most important variable, but in the case of "Chelsa data – WorldClim variables", bio12 was the most important variable. This suggests that the difference in predictive power between the two databases is most likely due to the underlying differences in the variables and not due to the modelling approach used by us.

We initially expected that the Chelsea 1.2 dataset would perform very well in mountainous areas because it corrects for orographic patterns of precipitation. Earlier studies, based in the Himalaya and the Swiss Alps, showed that the performance of Chelsa was superior to WorldClim. For example, it has been reported that Chelsa 1 outperformed WorldClim 1.4 in predicting the distribution of tree line forming Himalayan birch in the Himalaya (Bobrowski and Schickhoff 2017). Karger et al. (2017) also found a marginally superior performance of Chelsa 1 over WorldClim 1.4 in predicting the distribution of the distribution of 67 plant species from Switzerland. However, unlike our study, none of the previous studies verified the transferability of the models in space using an independent occurrence dataset from a different geographic region.

Our study yielded contrasting results, especially in terms of reliability when models are transferred to other regions. This difference could partly be due to the following reasons: i) earlier studies used older versions of the two climatic databases. WorldClim has considerably updated their data in the latest version (WorldClim 2) by incorporating remotely-sensed variables, such as land surface temperature and cloud cover. This update might have significantly improved the quality of the data in contrast to previous versions. ii) since Chelsa 1.2 has made several corrections to account for orographic patterns, especially in precipitation (Karger et al. 2017), these corrections might have changed the spatial pattern of the correlation structure amongst the variables at a local scale (Mesgaran et al. 2016). Therefore, the transferability of the model might be compromised when the models are projected to a new region characterised by a different correlation structure amongst the variables.

It is worth noting that the values of TSS were not very high for any of the models, indicating that climatic variables alone are not sufficient in explaining the distribution pattern of *A. adenophora*. For example, empirical studies have shown that the species has a narrow pH range from slightly acidic to neutral soils (pH 5 to 7) and cannot tolerate highly saline conditions (Lu et al. 2006). Moreover, biotic interactions and dispersal limitations are also crucial in determining plant distributions (Soberón and Nakamura 2009; Peterson et al. 2011). Therefore, including such variables could probably help in improving the general model performance and transferability for this species.

Although we found WorldClim 2 to perform better in terms of model transferability, it is premature to give generalised recommendations for preferring one dataset over the other, based on this case study alone. The species being studied and the geographic area of the study may be equally important (Hanspach et al. 2011). Providing a general overview, on how pertinent the problem is or under which conditions it applies for which type of species is beyond the scope of this study. We rather want to highlight the potential problem. We therefore recommend that the evaluation of climatic datasets should be performed routinely as an integral part of a modelling exercise and the database with best predictive performance should be chosen. For application of SDMs within the training and calibration region, internal validation is reliable, although performing an out-of-area cross-validation procedure is preferable when sample size is sufficient (Wenger and Olden 2012). However, if model transfer to a different geographic region is desired, validation against an independent occurrence dataset is highly recommended for choosing the most appropriate source of environmental data for the given study system. Therefore, a quantitative evaluation of predictor datasets from alternative sources should be routinely performed as an integral part of the modelling procedure.

Data availability

The occurrence data can be found here: https://zenodo.org/record/3875679#.Xt-g6IzozZRZ [https://doi.org/10.5281/zenodo.3875679]

Acknowledgements

We carried out this work with financial support from German Academic Exchange Service (DAAD) and institutional support from CSIR-Institute of Himalayan Bioresource Technology, Palampur and Helmholtz Centre for Environmental Research-UFZ. For contributing occurrence data, we would like to specifically thank Dr. Rajendra Yonzone from Darjeeling (India), Choki Gyeltshen from Bhutan, Bharat Pradhan from Sikkim (India), Dr. Dinesh Thakur from Jammu (India), Om Prakash from Palampur, and Dr. Bharat Shrestha from Nepal. Finally we would like to thank Dr. R.D Singh (deceased) for his motivation to carry out the field work in Himalayas.

References

- Allouche O, Tsoar A, Kadmon R (2006) Assessing the accuracy of species distribution models: Prevalence, kappa and the true skill statistic (TSS). Journal of Applied Ecology 43: 1223–1232. https://doi.org/10.1111/j.1365-2664.2006.01214.x
- Araújo MB, Guisan A (2006) Five (or so) challenges for species distribution modelling. Journal of Biogeography 33: 1677–1688. https://doi.org/10.1111/j.1365-2699.2006.01584.x
- Auld B, Martin PM (1975) The autecology of *Eupatorium adenophorum* Spreng. in Australia. Weed Research 15: 27–31. https://doi.org/10.1111/j.1365-3180.1975.tb01092.x
- Barton K (2015) MuMIn: Multi-model inference. R package version 1.9.13. Version 1: 18.
- Bobrowski M, Gerlitz L, Schickhoff U (2017) Modelling the potential distribution of *Bet-ula utilis* in the Himalaya. Global Ecology and Conservation 11: 69–83. https://doi.org/10.1016/j.gecco.2017.04.003
- Bobrowski M, Schickhoff U (2017) Why input matters: Selection of climate data sets for modelling the potential distribution of a treeline species in the Himalayan region. Ecological Modelling 359: 92–102. https://doi.org/10.1016/j.ecolmodel.2017.05.021
- Boyce MS, Vernier PR, Nielsen SE, Schmiegelow FKA (2002) Evaluating resource selection functions. Ecological Modelling 157: 281–300. https://doi.org/10.1016/S0304-3800(02)00200-4
- Burnham KP (2015) Multimodel Inference: Understanding AIC relative variable importance values.
- Burnham KP, Anderson DR (2002) Springer Model Selection and Multimodel Inference: a Practical Information-theoretic Approach. 488 pp.
- Clark JS, Carpenter SR, Barber M, Collins S, Dobson A, Foley JA, Lodge DM, Pascual M, Pielke Jr R, Pizer W, Pringle C, Reid WV, Rose KA, Sala O, Schlesinger WH, Wall DH, Wear D (2001) Ecological Forecasts: An Emerging Imperative. Science 293: 657–661. https://doi.org/10.1126/science.293.5530.657
- Cola V Di, Broennimann O, Petitpierre B, Breiner FT, Amen MD, Randin C, Engler R, Pottier J, Pio D, Dubuis A, Pellissier L, Mateo G, Hordijk W, Salamin N, Guisan A (2017) ecospat: an R package to support spatial analyses and modeling of species niches and distributions. Ecography 40(6): 774–787. https://doi.org/10.1111/ecog.02671
- Datta A, Kühn I, Ahmad M, Michalski S, Auge H (2017) Processes affecting altitudinal distribution of invasive Ageratina adenophora in western Himalaya: The role of local adaptation and the importance of different life-cycle stages. PloS ONE 12(11): e0187708. https://doi.org/10.1371/journal.pone.0187708
- Dormann CF, Elith J, Bacher S, Buchmann C, Carl G, Carré G, Marquéz JRG, Gruber B, Lafourcade B, Leitão PJ, Münkemüller T, Mcclean C, Osborne PE, Reineking B, Schröder B, Skidmore AK, Zurell D, Lautenbach S (2013) Collinearity: A review of methods to deal with it and a simulation study evaluating their performance. Ecography 36: 027–046. https://doi.org/10.1111/j.1600-0587.2012.07348.x
- Elith J, Leathwick JR (2009) Species Distribution Models: ecological explanation and prediction across space and time. Annual Review of Ecology, Evolution, and Systematics 40: 677–697. https://doi.org/10.1146/annurev.ecolsys.110308.120159
- Ervin GN, Holly DC (2011) Examining local transferability of redictive species distribution models for invasive plants: an example with Cogongrass (*Imperata cylindrica*). Invasive Plant Science and Management 4: 390–401. https://doi.org/10.1614/IPSM-D-10-00077.1
- Fernández M, Hamilton H (2015) Ecological niche transferability using invasive species as a case study. PLoS ONE 10(3): e0119891. https://doi.org/10.1371/journal.pone.0119891
- Fick SE, Hijmans RJ (2017) WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas. International Journal of Climatology 37: 4302–4315. https://doi. org/10.1002/joc.5086
- Grueber CE, Nakagawa S, Laws RJ, Jamieson IG (2011) Multimodel inference in ecology and evolution: challenges and solutions. Journal of Evolutionary Biology 24: 699–711. https://doi.org/10.1111/j.1420-9101.2010.02210.x
- Guisan A, Zimmermann NE (2000) Predictive habitat distribution models in ecology. Ecological Modelling 135: 147–186. https://doi.org/10.1016/S0304-3800(00)00354-9
- Hanspach J, Kühn I, Schweiger O, Pompe S, Klotz S (2011) Geographical patterns in prediction errors of species distribution models. Global Ecology and Biogeography 20: 779–788. https://doi.org/10.1111/j.1466-8238.2011.00649.x
- Heikkinen RK, Marmion M, Luoto M (2012) Does the interpolation accuracy of species distribution models come at the expense of transferability? Ecography 35: 276–288. https://doi.org/10.1111/j.1600-0587.2011.06999.x
- Hijmans RJ, Cameron SE, Parra JL, Jones PG, Jarvis A (2005) Very high resolution interpolated climate surfaces for global land areas. International Journal of Climatology 25: 1965–1978. https://doi.org/10.1002/joc.1276
- Hirzel AH, Le Lay G, Helfer V, Randin C, Guisan A (2006) Evaluating the ability of habitat suitability models to predict species presences. Ecological Modelling 199: 142–152. https://doi.org/10.1016/j.ecolmodel.2006.05.017
- Hutchinson GE (1957) Concluding remarks. Cold Spring Harbor Symposia on Quantitative Biology 22: 415–427. https://doi.org/10.1101/SQB.1957.022.01.039
- Jaryan V, Datta A, Uniyal SK, Kumar A, Gupta RC, Singh RD (2013) Modelling potential distribution of *Sapium sebiferum* an invasive tree species in western Himalaya. Current Science 105: 1282–1287.
- Jiménez-Valverde A, Lobo JM, Hortal J (2008) Not as good as they seem: the importance of concepts in species distribution modelling. Diversity and Distributions 14: 885–890. https://doi.org/10.1111/j.1472-4642.2008.00496.x
- Karger DN, Conrad O, Böhner J, Kawohl T, Kreft H, Soria-Auza RW, Zimmermann NE, Linder HP, Kessler M (2017) Climatologies at high resolution for the earth's land surface areas. Scientific Data 4: 170122. https://doi.org/10.1038/sdata.2017.122
- Lu P, Sang W, Ma K (2006) Effects of environmental factors on germination and emergence of Crofton weed (*Eupatorium adenophorum*). Weed Science 54: 452–457. https://doi. org/10.1614/WS-05-174R1.1
- Lu Z, Ma K (2006) Spread of the exotic croftonweed (*Eupatorium adenophorum*) across southwest China along roads and streams. Weed Science 54: 1068–1072. https://doi.org/10.1614/WS-06-040R1.1

- Mesgaran MB, Lewis MA, Ades PK, Donohue K, Ohadi S, Li C (2016) Hybridization can facilitate species invasions, even without enhancing local adaptation. Proceedings of the National Academy of Sciences 113: 10210–10214. https://doi.org/10.1073/pnas.1605626113
- Muniappan R, Raman A, Reddy GVP (2009) Ageratina adenophora (Sprengel) King and Robinson (Asteraceae). Biological Control of Tropical Weeds using Arthropods. Cambridge University Press, 63–73. https://doi.org/10.1017/CBO9780511576348.004
- Muniappan R, Viraktamath CA (1993) Invasive alien weeds in the Western Ghats. Current Science 64: 555–558.
- O'Donnell MS, Ignizio DA (2012) Bioclimatic predictors for supporting ecological applications in the conterminous United States. U.S Geological Survey Data Series 691. https://doi.org/10.3133/ds691
- Pearson RG, Dawson TP (2003) Predicting the impacts of climate change on the distribution of species: are bioclimate envelope models useful? Global Ecology and Biogeography 12: 361–371. https://doi.org/10.1046/j.1466-822X.2003.00042.x
- Peterson AT, Nakazawa Y (2008) Environmental datasets matter in ecological niche modelling: An example with Solenopsis invicta and Solenopsis richteri. Global Ecology and Biogeography 17: 135–144. https://doi.org/10.1111/j.1466-8238.2007.00347.x
- Peterson AT, Soberón J, Pearson RG, Anderson RP, Martínez-Meye E, Nakamura M, Araújo MB (2011) Ecological Niches and Geographic Distributions (MPB-49). Princeton University Press. https://doi.org/10.23943/princeton/9780691136868.001.0001
- Petitpierre B, Broennimann O, Kueffer C, Daehler C, Guisan A (2017) Selecting predictors to maximize the transferability of species distribution models: lessons from cross-continental plant invasions. Global Ecology and Biogeography 26: 275–287. https://doi.org/10.1111/geb.12530
- Petitpierre B, Kueffer C, Broennimann O, Randin C, Daehler C, Guisan A (2012) Climatic niche shifts are rare among terrestrial plant invaders. Science (New York, N.Y.) 335: 1344– 1348. https://doi.org/10.1126/science.1215933
- Poudel AS, Jha PK, Shrestha BB, Muniappan R (2019) Biology and management of the invasive weed Ageratina adenophora (Asteraceae): current state of knowledge and future research needs. Weed Research 59(2): 79–92. https://doi.org/10.1111/wre.12351
- Qiang S (1998) The history and status of the study on croftonweed (*Eupatorium adenophorum* Spreng.) A worst worldwide weed. Journal of Wuhan Botanical Research 16: 366–372.
- Raizada MB (1976) Supplement to Duthie's flora of the ppper Gangetic plain and of adjacent Siwalik and sub- Himalayan tracts. Bishen Singh Mahendra Pal Singh, Dehradun, 358 pp.
- Randin CF, Dirnböck T, Dullinger S, Zimmermann NE, Zappa M, Guisan A (2006) Are niche-based species distribution models transferable in space? Journal of Biogeography 33: 1689–1703. https://doi.org/10.1111/j.1365-2699.2006.01466.x
- Singh JS, Singh SP (1987) Forest vegetation of the Himalaya. The Botanical Review 53: 80–192. https://doi.org/10.1007/BF02858183
- Soberón J, Nakamura M (2009) Niches and distributional areas: Concepts, methods, and assumptions. Proceedings of the National Academy of Sciences 106: 19644–19650. https://doi.org/10.1073/pnas.0901637106
- Symonds MRE, Moussalli A (2011) A brief guide to model selection, multimodel inference and model averaging in behavioural ecology using Akaike's information criterion. Behavioral Ecology and Sociobiology 65: 13–21. https://doi.org/10.1007/s00265-010-1037-6

- Thuiller W, Richardson DM, Pyšek P, Midgley GF, Hughes GO, Rouget M (2005) Niche-based modelling as a tool for predicting the risk of alien plant invasions at a global scale. Global Change Biology 11: 2234–2250. https://doi.org/10.1111/j.1365-2486.2005.001018.x
- Tian Y, Feng Y, Liu C (2007) Addition of activated charcoal to soil after clearing *Ageratina adenophora* stimulates growth of forbs and grasses in China 41: 285–291.
- Tripathi RS, Yadav AS, Kushwaha SPS (2012) Biology of Chromolaena odorata, Ageratina adenophora and Ageratina riparia: a review. Invasive alien plants: an ecological appraisal for the Indian subcontinent 32: 43–56. https://doi.org/10.1079/9781845939076.0043
- Wang R, Wang J-F, Qiu Z-J, Meng B, Wan F-H, Wang Y-Z (2011) Multiple mechanisms underlie rapid expansion of an invasive alien plant. New Phytologist 191: 828–839. https://doi. org/10.1111/j.1469-8137.2011.03720.x
- Wenger SJ, Olden JD (2012) Assessing transferability of ecological models: An underappreciated aspect of statistical validation. Methods in Ecology and Evolution 3: 260–267. https://doi. org/10.1111/j.2041-210X.2011.00170.x
- Woodward FI (1987) Climate and plant distribution. Cambridge University Press, 188 pp.
- Woodward FI, Lomas M, Kelly CK (2004) Global climate and the distribution of plant biomes. Philosophical transactions of the Royal Society of London. Series B, Biological sciences 359: 1465–1476. https://doi.org/10.1098/rstb.2004.1525
- Yates KL, Bouchet PJ, Caley MJ, Mengersen K, Randin CF, Parnell S, Fielding AH, Bamford AJ, Ban S, Barbosa AM, Dormann CF, Elith J, Embling CB, Ervin GN, Fisher R, Gould S, Graf RF, Gregr EJ, Halpin PN, Heikkinen RK, Heinänen S, Jones AR, Krishnakumar PK, Lauria V, Lozano-montes H, Mannocci L, Mellin C, Mesgaran MB, Moreno-amat E, Mormede S, Novaczek E, Oppel S, Crespo GO, Peterson AT, Rapacciuolo G, Roberts JJ, Ross RE, Scales KL, Schoeman D, Snelgrove P, Sundblad G, Thuiller W, Torres LG, Verbruggen H, Wang L, Wenger S, Whittingham MJ, Zharikov Y, Zurell D, Sequeira AMM (2018) Outstanding challenges in the transferability of ecological models. Trends in Ecology & Evolution 33: 790–802. https://doi.org/10.1016/j. tree.2018.08.001

Supplementary material I

Variable selection using cluster analyss based on Spearman's rank corellation and UPGMA method for agglomeration

Authors: Arunava Datta, Oliver Schweiger, Ingolf Kühn

Data type: statistical data

Copyright notice: This dataset is made available under the Open Database License (http://opendatacommons.org/licenses/odbl/1.0/). The Open Database License (ODbL) is a license agreement intended to allow users to freely share, modify, and use this Dataset while maintaining this same freedom for others, provided that the original source and author(s) are credited.

Link: https://doi.org/10.3897/neobiota.59.36299.suppl1

Supplementary material 2

Multimodel inference table

Authors: Arunava Datta, Oliver Schweiger, Ingolf Kühn

Data type: statistical data

- Explanation note: Tables depicting all the component models of the best subset (i.e. models that differed by 2 or less in AIC).
- Copyright notice: This dataset is made available under the Open Database License (http://opendatacommons.org/licenses/odbl/1.0/). The Open Database License (ODbL) is a license agreement intended to allow users to freely share, modify, and use this Dataset while maintaining this same freedom for others, provided that the original source and author(s) are credited.

Link: https://doi.org/10.3897/neobiota.59.36299.suppl2

Supplementary material 3

R codes

Authors: Arunava Datta, Oliver Schweiger, Ingolf Kühn

Data type: R code (text)

Explanation note: R codes used in the paper.

Copyright notice: This dataset is made available under the Open Database License (http://opendatacommons.org/licenses/odbl/1.0/). The Open Database License (ODbL) is a license agreement intended to allow users to freely share, modify, and use this Dataset while maintaining this same freedom for others, provided that the original source and author(s) are credited.

Link: https://doi.org/10.3897/neobiota.59.36299.suppl3

RESEARCH ARTICLE



Xylella fastidiosa invasion of new countries in Europe, the Middle East and North Africa: Ranking the potential exposure scenarios

Michel Frem^{1,3}, Daniel Chapman², Vincenzo Fucilli¹, Elia Choueiri³, Maroun El Moujabber⁴, Pierfederico La Notte⁵, Franco Nigro¹

I Università degli Studi di Bari - Aldo Moro, Dipartimento di Scienze del Suolo, della Pianta e degli Alimenti (Di.S.S.P.A.), Via Amendola 165/A, 70126, Bari, Italia 2 University of Stirling, Department of Biological and Environmental Sciences, Stirling, FK9 4LA, UK 3 Lebanese Agricultural Research Institute, Department of Plant Protection, Tal Amara, P.O. Box 287, Zahlé, Lebanon 4 CIHEAM Bari, Istituto Agronomico Mediterraneo, Via Ceglie 9, 70010 Valenzano, Bari, Italia 5 Consiglio Nazionale delle Ricerche, Istituto per la Protezione Sostenibile delle Piante, Via Amendola 122/I, 70126, Bari, Italia

Corresponding author: Michel Frem (mefrem@lari.gov.lb)

Academic editor: Uwe Starfinger	Received 15 April 2020	Accepted 7 July 2020	Published 30 July	2020
Academic editor. Owe starmiger	Received 19 April 2020	$\pi \alpha \alpha \beta \alpha \alpha \beta \alpha $	1 ublished Jo July	2020

Citation: Frem M, Chapman D, Fucilli V, Choueiri E, Moujabber ME, Notte PL, Nigro F (2020) *Xylella fastidiosa* invasion of new countries in Europe, the Middle East and North Africa: Ranking the potential exposure scenarios. NeoBiota 59: 77–97. https://doi.org/10.3897/neobiota.59.53208

Abstract

After the recent high-impact European outbreaks of *Xylella fastidiosa* (*Xff*), a xylem-limited plant pathogenic bacterium native to the Americas, this research aims to rank the risks of potential entry, establishment and spread of *Xf* in new countries across Europe, the Middle East and North Africa. A novel risk-ranking technique is developed, based on combining entry risk drivers (imported plants, direct flights and ferry connections) with risk factors related to establishment and spread (presence of potential insect vectors, vulnerable economic crops, alternative hosts and climate suitability) of this pathogen. This reveals that western European countries have the highest risk for entry, but that the Mediterranean basin runs the highest risk for establishment and spread of *Xf*. Lebanon in particular has the highest level of risk for *Xf* dispersal within its suitable territory. Countries without current outbreaks combining high risks of *Xf* arrival and establishment are mainly in the Mediterranean basin: Turkey is at the highest level of risk, followed by Greece, Morocco and Tunisia, which are ranked at the high level. The ranking model also confirms the vulnerability, in terms of invasion by *Xf*, of southern European countries (Italy, Portugal and Spain) in which the pathogen has already been reported. High summer temperatures in these southern countries are

Copyright Michel Frem et al. This is an open access article distributed under the terms of the Creative Commons Attribution License (CC BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

likely to be the significant determinant for the overall invasion process, while northern European countries have a high level risk for the arrival of the pathogen, but relatively low summer temperatures may limit establishment and spread of major outbreaks. In general, our study provides a useful approach for mapping and comparing risks of invasive non-native species and emerging pathogens between countries, which could be useful for regional horizon scanning and phytosanitary and biosecurity management.

Keywords

alien species, biological invasion, entry, dispersal, plant bacterium, risk drivers

Introduction

Global trade networks are implicated in increasing rates of global spread of plant diseases through transport of live plants and other plant-related products (Chapman et al. 2017). The plant pathogenic bacterium *Xylella fastidiosa* (hereafter *Xf*) is an important example of this process. Originating in the Americas, this plant pathogen has recently been detected for the first time in several European and Near Eastern countries: Italy in 2013 (Saponari et al. 2013), France in 2015 (Denancé et al. 2017), Switzerland in 2015 (European and Mediterranean Plant Protection Organization 2015), Spain in 2016 (Olmo et al. 2017), Germany in 2016, Portugal in 2019, and more recently, Israel in June 2019 (European and Mediterranean Plant Protection Organization 2016, 2019a, b). This dangerous non-native pathogen causes several devastating plant diseases that have huge socio-economic and ecological impacts, motivating risk assessment for the entry and spread of *Xf* into new countries (European Food Safety Authority 2019).

Following detection of Xf outbreaks in Europe, the EU and some Middle East and North African (MENA) countries have implemented several risk reduction options to combat this plant disease and prevent its entry and spread. Despite these actions, there remains a risk that Xf will continue to spread to new countries and regions through the movement of infected host plants (asymptomatic or unknown hosts) or via unintentional transport of insect vectors through equipment as a commodity contaminant or vehicle "hitchhiker". This transport risk is likely to be highest in countries with welldeveloped transport links to current outbreak areas (Chapman et al. 2017). However, transport connectivity will only present a serious risk of damaging outbreaks if there is also a suitable climate and land use to support the establishment and spread (Early et al. 2018). Therefore, there is a need for an overall country-level risk ranking for Xf simultaneously across the EU and MENA region which would allow the selection of priority countries for more detailed country-specific pest risk analyses.

As such, we aimed to rank the likelihood of potential invasion by Xf of new countries in these regions, and to provide an overall risk rating by combining rankings for entry, establishment and spread in order to assess each country's overall vulnerability to Xf invasion. Over the past 20 years, there have been several model frameworks and studies of alien species risk assessment at an individual-country scale. These include invasive alien species risk assessment in Great Britain (Mumford et al. 2010), risk assessment simultaneously prioritizing the control of invasive plant species and the conservation of rare plant species in the USA (Miller et al. 2010), risk assessment models for invasive species and uncertainty in rankings from multi-criteria analysis in Australia (Benke et al. 2010), a weed screening tool for the US (Koop et al. 2011), generic ecological impact assessments of alien species in Norway (Sandvik et al. 2013), risk screening tools for potentially invasive plants, animals and their pathogens in Belgium (D'hondt et al. 2015), and development of a plant risk evaluation tool to assess the invasive potential of ornamental plants (Conser et al. 2015). In contrast to these country-specific assessment, only a few studies assess risk across multiple countries, for example the EPPO prioritization process for invasive alien plants (Brunel et al. 2010), the new protocols for assessment of the environmental impact of pests in the EPPO decision support scheme for pest risk analysis (Kenis et al. 2012), guidance on environmental risk assessment of plant pests (European Food Safety Authority 2010), and the generic impact scoring system (Nentwig et al. 2016) in Europe. Moreover, a few protocols regarding alien taxa have been developed at the global level, such as the framework and guidelines for implementing the proposed IUCN (International Union for Conservation of Nature) environmental impact classification for alien taxa (Hawkins et al. 2015).

Concerning Xf, the few published studies that have performed risk assessment for this biological invader are limited to one country (Afechtal et al. 2018), one region (European Food Safety Authority, 2015, 2018, 2019; Godefroid et al. 2019;) or a subset of territories (Bosso et al. 2016) mainly based on climate suitability indicators. Thus, the novelty of the current research is to integrate multiple Xf risk drivers (i.e. plant trade, human movements, insect-vectors, host-plants, temperature suitability) and simultaneously ranks all countries in Europe and MENA region in a way that is useful for decision makers. In addition, these selected ranking variables (Table 1) are quantitative (except for vectors), objective (i.e. based on statistics and not on expert opinions), transparent, consistent, sufficiently specific, and available simultaneously to all the countries concerned. By quantifying these risk drivers with best available scientific evidence, the proposed approach is simple to apply and may be used as a conceptual framework in rating-based risk assessment of other pests to support policy for appropriate and efficient biosecurity management of Xf and other invading pests.

Methods

Countries included in the risk ranking model

Fifty six countries were selected for this study as shown in Figure 1. Thirty eight in Europe (Albania/ALB, Austria/AUT Belgium/BEL, Bosnia Herzegovina/BIH, Bulgaria/BGR, Croatia/HRV, Cyprus/CYP, Czech Republic/CZE, Denmark/DNK, Estonia/EST, Finland/FIN, France/FRA, Germany/DEU, Greece/GRC, Hungary/HUN, Iceland/ISL, Ireland/IRL, Italy/ITA, Latvia/LVA, Lithuania/LTU, Luxembourg/LUX,

Macedonia/MKD, Malta/MLT, Moldova/MDA, Montenegro/MNE, Netherlands/ NLD, Norway/NOR, Poland/POL, Portugal/PRT, Romania/ROU, Serbia/SRB, Slovakia/SVK, Slovenia/SVN, Spain/ESP, Sweden/SWE, Switzerland/CHE, Turkey/TUR and United Kingdom/GBR) and eighteen in MENA (Algeria/DZA, Bahrain/BHR, Egypt/EGY, Iraq/IRQ, Israel/ISR, Jordan/JOR, Kuwait/KWT, Lebanon/LBN, Libya/ LBY, Morocco/MAR, Oman/OMN, Qatar/QAT, Saudi Arabia/, State of Palestine/ PSE, Syria/SYR, Tunisia/YUN, United Arab Emirates/UAE and Yemen/YEM). Their selection was based on: (i) their connectivity through the trade of potentially infected plants for planting and the movement of people as passive vectors, (ii) their abundance of *Xf* host-plants, (iii) their relative summer climate similarity to countries where the disease is already present, and (iv) their inclusion in the EU and MENA regions.

Data on risk drivers for rankings

The vulnerability of a target country to the entry of Xf was assessed via six key risk drivers related to entry (importation of plants for planting, direct air flights and ferry traffic from countries source of Xf) and four risk drivers related to the establishment and spread of the bacterium (potential Xf vectors, vulnerable crops, alternative hosts from the forestry and ornamental sectors, and the suitability of summer temperatures) as summarized in Table 1. The use of these indicators was justified by: (i) the consideration of trade of plants and movement of passengers as common pathways of Xf interception mainly in Europe (European Food Safety Authority 2015), (ii) the biology and ecology cycle of the pathogen as stated by Chatterjee et al. (2008) (the presence of vectors is essential to transmit the bacterium from infected to healthy plants), (iii) the economic importance of Xf host-crops in southern European countries and MENA



Figure 1. Countries including in the risk ranking model, labelled with their international codes (ISO3).

Category	Code and description of the risk driver					
Entry (plant trade)	<i>ENT</i> , Live plant imports from three EU countries with major outbreaks (ESP, FRA, ITA).					
	ENT, Live plant imports from non-EU countries in which Xf is present but from which th					
	bacterium has not been intercepted in Europe (ARG, CAN, IRN, PRY, TWN, and VEN).					
	ENT, Live plant imports from non-EU countries in which Xf is present and from which the					
	bacterium has been intercepted in Europe (BRA, CRI, ECU, HND, MEX and USA).					
Entry (human	ENT_4 Number of direct airline routes from or near outbreak regions in the EU (ESP, FRA and					
movement)	ITA).					
	ENT_5 Number of direct airline routes from non-EU countries in which Xf is present (ARC					
	BRA, CAN, CRI, ECU, HND, IRN, PRY, MEX, TWN, USA and VEN).					
	ENT ₆ Number of annual ferry sailings from ports in or near outbreak regions in the EU (E					
	FRA and ITA) and from non-EU countries in which Xf is present (ARG, BRA, CAN, CRI,					
	ECU, HND, IRN, PRY, MEX, TWN, USA and VEN).					
Establishment and	EST ₁ Presence of at least one known or potential Xf insect-vector.					
spread (vectors)						
Establishment and	<i>EST</i> ₂ Proportion of agricultural area growing susceptible crops.					
spread (host plants)	<i>EST</i> ₃ Proportion covers of forest.					
Establishment and	EST_4 Mean relative Xf growth potential in vulnerable habitats, based on summer mean					
spread (climate)	temperature.					

Table 1. Risk drivers used to assess overall Xylella fastidiosa invasion risk and their weights.

region, (iv) the abundance of forest as Xf alternative hosts mainly in northern Europe countries (European Food Safety Authority 2018), (v) the summer optimal temperature (around 28 °C) for Xf growth as studied by Feil and Purcell (2001) and, (vi) their data availability to all selected countries.

As an indicator of the entry risk from plant trade we obtained data on import volumes of Plants for Planting from potential Xf source countries between 2000 and 2016 from the Resource Trade Earth database (https://resourcetrade.earth/data) and from the Taiwanese Bureau of Foreign Trade (https://cus93.trade.gov.tw) (Suppl. material 2: Table S1). Xf is known to infect over 563 cultivated and ornamental herbaceous, shrubby and woody species present in cultivated fields, gardens, parks, woods and forests could be infected by Xf through the effective transmission of insect vectors (European Food Safety Authority 2018). Many host-plants have long asymptomatic periods or remain fully asymptomatic, causing potentially high transport risk through trade. For each country, mean annual imports (metric tons) of plant commodity groups considered capable of carrying viable bacteria were obtained (i.e. bulbs, tubers, tuberous roots, corms and rhizomes in dormant and in growth; unrooted cuttings and slips; edible fruit trees, shrubs and bushes and live plants). Import volumes were calculated separately for three sets of countries in which Xf is present but potentially differ in import risk: (i) three EU countries with outbreaks (ESP, FRA, ITA), (ii) six non-EU countries from which the bacterium has not been intercepted coming into Europe (ARG, CAN, IRN, PRY, TW, VEN), and (iii) six non-EU countries from which Xf has been intercepted entering Europe (BRA, CRI, ECU, HND, MEX, USA) (European Food Safety Authority 2018 Annex D).

Intentional or unintentional human movement of contaminated plant material or insect-vectors may also lead to new introductions of *Xf*. To quantify the entry risk from human movement through international air travel, we calculated the number of annual

direct airline flights from the airports in or near to outbreaks areas in Europe (specifically demarcated zones in ESP, FRA and ITA) and from non-EU countries in which Xf is present (ARG, BRA, CAN, CRI, ECU, HND, IRN, PRY, MEX, TW, USA, VEN). Flight data were extracted from the OpenFlights database (https://www.openflights.com) which covers all flights in 2014. Entry risk from ferry traffic was evaluated using the number of annual passenger ferry sailings from ports in or near the outbreak areas in Europe and from the same infected non-EU countries. Ferry data were extracted from scheduled sailings in 2018 listed by Ferry Lines (https://www.ferrylines.com). We initially considered including road network connectivity as a risk factor, but decided against it because the relatively small outbreak areas in Europe are far by road from other uninfected countries.

Since Xf is entirely insect-transmitted, presence of potential vectors was considered an indicator of risk of establishment and spread. Disease transmission occurs by xylem-feeding insect vectors, mainly via spittlebugs in Europe. In the Apulia region of southern Italy, the spittlebug Philaenus spumarius (L.: Superfamily Cercopoidea, Family Aphrophoridae) is considered to play the major role in transmitting Xf subspecies pauca (Saponari et al. 2014). Additionally, other spittlebug species such as Neophilaenus campestris and Philaenus italosignus (Hemiptera: Aphrophoridae) are also able to transmit Xf (European Food Safety Authority 2015, 2018, 2019) suggesting that other unknown vectors could facilitate outbreaks if Xf were introduced to other countries. As such, for all countries, we searched for occurrence of the following potential xylemfeeding vector species: Philaenus spumarius, Neophilaenus campestris, Aphrophora alni, Aphrophora salicina, Cercopis vulnerata, Cercopis sanguinolenta, Cicada orni, Cicadatra atra, Cicadivetta tibialis, Cicadella viridis, Lyristes plebejus and Tibicina haematodes. Data were obtained from Global Biodiversity Information Facility (www.gbif.org), HemBases (https://hemiptera-databases.org/), 3I Interactive Keys and Taxonomic Databases (http://dmitriev.speciesfile.org/), Fauna Europea (https://fauna-eu.org) and a previous literature review (European Food Safety Authority 2015).

Establishment and spread also requires presence of Xf host plants. This risk indicator was estimated from the cultivated areas of its main vulnerable economic hosts (i.e. almonds with shell, apricots, blueberries, cherries, sour cherries, green coffee, citrus fruit nes, stone fruit stone nes, grapefruits including pomelos, grapes, olives, oranges, peaches and nectarines, pears, plums and sloes). Production data for 2000–2015 were obtained from the FAOSTAT database (http://www.fao.org/faostat/en/) and converted into the proportion of the total agricultural area of each country containing vulnerable crops. In addition, we obtained the proportion of the total area of each target country covered by forest, from the same source, as an indicator of alternative host plants for Xf, which is capable of infecting tree species from genera including Quercus, Acer, and Ulmus.

Risk from climate suitability was assessed based on summer land surface temperatures, obtained from two regional gridded layers deriving the mean temperature of the warmest quarter (Bio10) from MODIS satellite data. Europe was covered by the EuroLST layer at 250 m resolution from MOD11A1 V005 daily temperatures, reprojected to a 0.05 degree long-lat grid (Metz et al. 2014). Non-European countries were covered by the Tropical LST layer at 0.05 degree resolution from MOD11C3v5 monthly temperatures (Deblauwe et al. 2016). Before merging the layers, they were harmonized through a linear regression fitted to their overlapping areas (EuroLST = $1.923 + 0.930 \times \text{Tropical LST}$, $R^2 = 0.983$). Summer land surface temperatures were converted into a relative Xf growth potential using published data on *in vitro* bacterial colony growth rates of Xf at different temperatures (Feil and Purcell 2001). From these data, a simple growth curve was fitted, with an optimum temperature of 28 °C and lower and upper growth temperatures of 8 and 35 °C (Suppl. material 1: Figure S1). This growth curve was rescaled between 0 and 1 and used to map the relative suitability of the summer land surface temperatures. Finally, mean suitable values within potentially invadable habitat (croplands and forests) were calculated for each country.

Risk ranking

We developed a structured system that ranks nations according to their risk of *Xf* invasion, combining the above risk drivers that constitute the components of the biological invasion process (Fig. 2). Biological invasion is a multistage process, but it can be simplified into two major stages: entry, involving the transportation and introduction of the species, and the subsequent establishment and spread of the alien species. Each stage involves a set of barriers or blocks that the species must overcome to successfully invade a new territory (Rogg et al. 2003; Blackburn et al. 2011; Seebens et al. 2015; Early et al. 2016). The overall invasion ranking from the entry and establishment and spread indicators mentioned above is based on their relative values among all countries (Early et al. 2016). An established matrix to combine both types of risk is used (Fig. 3).



Figure 2. Overview of the invasion process of *Xylella fastidiosa* into a new country, based on the general framework of Blackburn et al. (2011). The diagram highlights specific risk indicators for each stage of invasion that were considered in our risk-ranking model.



Figure 3. Martix for combining entry risk with establishment and spread risk to form an overall assessment of vulnerability to *Xf* invasion, based on Early et al. (2016). The diagram assembles five relative rank categories at each stage of invasion considered in our risk ranking model.

Variables in the analysis (risk driver data) were weighted based on their relative loadings or importance for the first two axes of a factor analysis (Table 2). Factor analysis has long been used as a multivariate method to combine correlated variables into smaller numbers of common factors (Venette 2015). In this study, it combined the multiple risk drivers into two dimensions of risk, which we interpreted as indicators of entry risk and establishment and spread risk. Prior to the factor analysis, data on all risk drivers were normalized by Box-Cox power transformation, improve their conformity to normal distributions. Bartlett's test of sphericity was used to test for significant multivariate correlations between variables. In addition, the Kaiser-Meyer-Olkin statistic was used to describe the proportion of variance captured by the factor analysis. Inspections of the loadings showed that factors 1 and 2 correspond to increasing gradients of risk for entry, and establishment and spread, respectively.

Results

The risk indicators used in the ranking are shown in Figure 4. Factor analysis on these transformed indicators produced an adequate description of their multivariate pattern. Bartlett's test of sphericity showed significant correlation among risk drivers ($\chi^2 = 305.2$, df = 45, *P* < 0.001), while the Kaiser-Meyer-Olkin statistic had a value of 0.695, which is above the 0.5 threshold that is usually accepted for adequate explained variation.

The loadings table (Table 2) shows high positive loadings for entry risk variables on the first factor, except for direct ferry lines (ENT_{o}) , while none of the entry factors loaded strongly on the second factor. By contrast, establishment and spread risk drivers

Risk driver code	Factor 1 loading	Factor 2 loading	Entry risk weight	Establishment and
				spread risk weight
ENT,	0.863	0.236	20%	
ENT,	0.825	-0.258	19%	
ENT	0.815	-0.114	19%	
ENT	0.778	-0.074	18%	
ENT	0.750	-0.442	17%	
ENT	0.308	0.195	7%	
EST	0.366	0.703		25%
EST,	-0.026	0.825		29%
EST ₃	0.351	0.462		16%
EST	0.039	0.842		30%

Table 2. Rotated loadings of the risk drivers on two factor analysis axes, showing the contribution of each risk factor. See Table 1 for explanations of the risk driver codes. Relative loadings of entry drivers on factor 1 and of establishment and spread drivers on factor 2 were used as weights to estimate their combined risks.

loaded strongly on the second factor, except forest cover, but not on the first. Given this structure, the first factor was interpreted as a latent variable associated with entry, and the second factor interpreted as a latent variable indicating risk of establishment and spread.

Entry risk

Figure 5 displays the ranking of Europe and MENA countries according to the entry risk drivers. Most western European countries have high values of most entry risk indicators (Fig. 4) and as such have the highest overall entry risk. Most MENA countries appear less exposed to entry risk factors, except for Tunisia, Morocco, Jordan, Kuwait, Saudi Arabia and the UAE, where the entry risk level is high.

Establishment and spread risk

Figure 6 shows the ranking of Europe and MENA countries according to the four establishment and spread risk factors relating to insect vectors, host plants and climate suitability (see also Fig. 4). Among the establishment and spread risk indicators, suitable crops and summer temperatures were concentrated around the Mediterranean, leading to most Middle East and Mediterranean countries being classified at high to highest risk levels.

In the MENA region, Lebanon is the only country at the highest level of risk, followed by Morocco, Tunisia, Egypt, Israel and State of Palestine which are classified at high risk level. Algeria, Jordan and Syria are ranked at the medium level risk, while the remaining MENA countries are at least risk rank for *Xf* establishment and spread. In Europe, Albania, Cyprus, Greece, Italy, Macedonia, Malta, Portugal, Serbia, Spain and Turkey are classified at highest risk level, while Iceland, Ireland and Norway are at least risk level.



Figure 4. Maps of the risk indicators used for ranking potential for *Xf* entry (*ENT*) and establishment and spread (*EST*) as described in Table 1.



Figure 5. Rank categorization of Europe and MENA countries according to the six entry risk drivers of *Xylella fastidiosa* in relation to the importation of plants for planting, direct air flights and ferry traffic.



Figure 6. Rank categorization of Europe and MENA countries according to the four establishments and spread risk drivers of *Xylella fastidiosa* in relation to vectors, vulnerable economical crops, alternative hosts and climate suitability.



Figure 7. Overall rank categorization of Europe and MENA region according to exposure to invasion by *Xf* with respect to combined all entry, establishment and spread risk drivers and using the matrix of Early et al. (2016).

Overall risk of exposure to invasion by Xf

When we combined the risk rankings for entry (Fig. 5) and establishment and spread (Fig. 6) using the matrix in Figure 2 the countries ranked as most vulnerable were Italy, Portugal, Spain and Turkey (Fig. 7). Additionally Greece, Morocco and Tunisia were ranked at a high risk level. Only three countries (Bahrain, Libya and Yemen) yield the least level of combined risk, although the risk was generally low in northern Europe and the non-Mediterranean Middle East.

Discussion

The threat of intentional or unintentional species movements leading to the entry and spread of invasive alien organisms is increased by international trade and travel (Tatem 2009; Early et al 2016; Chapman et al 2017). For *Xf*, recent outbreaks and interceptions of infected imported plants for planting in Europe show that *Xf* spreads to new countries via this pathway (European Food Safety Authority 2015). Most countries in Europe and the MENA region are increasingly connected with each other and with global sources of *Xf* in the Americas and Asia through the trade in imported plants and human travel. Given weaknesses in current phytosanitary regulations and airport inspection structures in parts of the region, especially in some MENA countries, it is possible that passengers could carry with them potentially infected planting material, or

that transport infrastructure could accidentally carry Xf insect vectors. Consequently, there is a real risk that Xf may continue to overcome geographical barriers and infect new countries, potentially leading to major new disease outbreaks. Our assessment of the joint risks of Xf entry and establishment and spread therefore provides a useful approach for regional risk assessment by ranking the relative risk across 56 countries.

In particular our analysis identified a contrast between entry risk and the risk of establishment and spread. This was clearly seen in the factor analysis loadings, in which entry risk drivers loaded strongly on axis 1, while establishment and spread risk drivers loaded strongly on axis 2. Based on this, countries in western Europe and also Turkey tended to score highly for entry risk, principally because they import greater volumes of plants from infected countries in Europe and globally and had greater numbers of direct flight connections originating in infected regions. By contrast, risk of establishment and spread was ranked mostly based on the degree to which crops grown in a country are known to be susceptible to Xf and whether the summer temperature was apparently well suited to Xf colony growth. Presence of vectors was also weighted strongly in the factor analysis, but did not have a big influence on the results since nearly all countries had at least one potential vector species recorded as present (Fig. 4g). Countries with vulnerable crop types and summer mean land surface temperatures close to the optimum of 28 °C (Feil and Purcell 2001) were mainly located around the Mediterranean basin, where many countries were assessed to have low entry risk. The known world distribution of Xf is shown in the suppl. material 5: Table S4.

Consequently, few countries were ranked very highly for both entry risk and establishment. Of those that were, three countries already have major Xf outbreaks, namely Italy, Spain and Portugal, whose outbreak was discovered after this study was conducted. Greece and Turkey were also ranked in the highest risk group but so far remain disease free. Güldür et al. (2005) reported the detection of Xf on seven almond trees uprooted in Turkey. However, following this detection, the Turkish National Plant Protection Organization declared that Xf did not occur in the country. Outside of the top-ranked countries, Morocco and Tunisia were ranked as high risk countries, on the basis that they scored fairly highly for both entry and establishment and spread. So far, Xf has not been detected in those countries. However, a pest risk analysis conducted for Morocco concluded that the overall likelihood of Xf spread in the country is very likely, providing support for our assessment (Afechtal et al. 2018). Interestingly, in our study France was ranked as medium overall risk, despite it having outbreaks on its south coast and on the Mediterranean island of Corsica. The primary reason for this is that France also includes more northerly regions with less suitable summer temperatures and less vulnerable crop types. Therefore, despite France clearly having areas at high risk of the disease, the national-scale resolution of the study reduced its overall risk ranking.

Our findings are broadly consistent with other risk assessment studies for Xf in Europe and MENA region. For example, data-driven species distribution modelling studies using a wider range of climate variables than we assessed generally confirm our simple mapping of temperature risk (Bosso et al. 2016; European Food Safety Authority 2019). For example, Godefroid et al. (2019) predicted the potential distribution of Xf subsp. *fastidiosa* in Spain, France, Italy, Croatia, Greece and Turkey, and the coastal

regions of North Africa, which is consistent with our simple forward projections based only on an *in vitro* growth curve (see Suppl. material 1: Fig. S1). In addition there have been recent detections of Xf in some of the countries that we rated as high risk for establishment based on temperature and crop types. This includes Lebanon, in which Temsah et al. (2015) reported Xf on oleander, although Habib et al. (2016) disagreed with this finding. It also includes Israel, in which Xf is under containment after a recent detection on almonds in the Hulla Valley (European and Mediterranean Plant Protection Organization 2019b). By contrast, countries such as Germany, Switzerland and Netherlands were ranked as having low or medium risk for establishment and spread. In those countries Xf has been intercepted in trade or infected plants within protected indoor locations (consistent with their highest ranking for entry), but has not been able to spread to the wider environment (European Food Safety Authority 2019).

Limitations of the study

Important limitations of this study include its country-level resolution, reliance on data of differing quality, missing risk factors with insufficient data to include and uncertainty about how to combine risk factors into overall risks. As discussed above, the country-level resolution of the analysis affected results for countries that appear largely unsuitable for *Xf*, but have small areas that are at high risk, such as France. Future approaches could use high resolution gridded data on the risk drivers to try to map risk at a higher resolution, addressing this problem. One reason that we were limited to a country-level analysis was that some datasets were only available at that resolutions are too poorly mapped to allow regional breakdowns. As plant trade is the major pathways of *Xf* introduction, there seems little prospect of mapping entry risk at higher resolutions. However, availability of gridded climate and land use data (see Suppl. material 1: Fig. S1) could allow higher resolution mapping of those components of establishment and spread risk.

Due to lack of adequate data across Europe and the MENA region we did not feel able to include some other potentially relevant risk drivers. Individual host plant species (other than major crops) and insect vector distributions were not mapped well enough to consider. We also did not consider variation in risk for different subspecies of *Xf*, of which at least three are present in Europe and the MENA (*Xf* subsp. *pauca*, *multiplex* and *fastidiosa*) and all differ in host plant range and temperature-growth responses (European Food Safety Authority 2019). In addition, lack of information meant our the analysis did not account for variation in management regimes in different countries, including farmers' cultural management (i.e. crop genetics, use of resist-ance/tolerant cultivars, presence of transgenic plants, vegetation, vector control etc.), surveillance and monitoring programs, or phytosanitary regulations at the import stage or testing capacity. However, with more complete data the present ranking model could be extended to include additional risk indicators.

The results of this type of study can also be sensitive to how risk indicators are combined. We suggest that our use of factor analysis to weight additive risk combination ensured that our individual risk rankings for entry and establishment and spread followed the major gradients in the assessed drivers of those risks. In addition, we explored alternative schemes, including multiplicative risk combinations, and found these produced qualitatively similar results. In addition, we used an established matrix to combine both types of risk (Fig. 3; Early et al 2016), though it is likely that our results are somewhat sensitive to this choice of matrix. Furthermore, relative risk rankings such as those produced here are sensitive to the set of countries included in the analysis, which is why we endeavored to include the widest range of countries in the geographical regions that may be threatened by Xfs arrival in Europe. Ideally, we would convert our qualitative rankings into quantitative probabilities of entry and quantitative measures of spread potential. However, until large numbers of entry and outbreak events are observed, allowing us to link particular risk drivers to actual probabilities of invasion it will be difficult to improve upon qualitative ranking.

Conclusion

Regional risk assessment for high-impact invasive alien species such as Xf requires approaches that incorporate multiple risk drivers to simultaneously rank countries for multiple stages of invasion, such as the approach developed here. The world is increasingly connected by international plant trade and human travel, which are potential drivers of Xf entry into new areas where the presence of insect vectors, the abundance of host-plants as well as the climate suitability play an important role for its dispersal. As such our approach could be useful for both individual countries to understand their risk of Xf relative to other countries, and if applied across many different pests it could be useful to identify priority species. It is also useful for supra-national organizations interested in Plant Health (i.e. EPPO, EFSA, and EU) who can use country-level risk rankings to prioritize phytosanitary resources among countries. In this context, the strength of this study is that it creates a tool for mapping, ranking and combining multiple sources of invasion risk at country-level.

Overall, we identified the most vulnerable new countries to Xf invasion are mainly located in the Mediterranean basin, particularly Turkey, Greece, Morocco and Tunisia. As such, this research provides important information in terms of potential exposure by Xf, for policy makers or stakeholders in high risk countries where Xf has not yet been reported. We suggest that these countries and other ranked with relatively high risk should conduct detailed individual risk analysis, take preventive measures, and if necessary, improve their surveillance systems for early Xf detection in plants and insect-vectors, and raise awareness to prevent socioeconomic and ecological impacts on their ecosystems. In addition, our approach could be adapted to assess the specific risks for other important invasive alien species, irrespective of their origin, potential area of invasion and whether or not they have already invaded parts of the risk assessment region. As such, it provides a useful addition to tools and methods more commonly applied in regional-scale risk assessment for invasive alien species.

Acknowledgements

Michel Frem thanks the UK Centre for Ecology and Hydrology, CIHEAM Bari (Italy) and UNIBA Aldo Moro Bari (Italy) for their hospitality, bibliography, data collection and treatment. Thanks also to Sarah Jane Christopher of UNIBA Aldo Moro Bari (Italy) for her careful revision of the Manuscript. This research was supported by CURE-*Xf*, an EU-funded project, coordinated by CIHEAM Bari (H2020-Marie Sklodowska-Curie Actions – Research and Innovation Staff Exchange. Reference number: 634353).

References

- Afechtal M, Vicent A, Saponari M, D'Onghia AM (2018) Pest Risk Analysis on *Xylella fastidi*osa in Morocco. Journal of Plant Protection Research 58(3): 215–219.
- Benke K, Steel J, Weiss J (2010) Risk assessment models for invasive species: uncertainty in rankings from multi-criteria analysis. Biological Invasions 13: 239–253. https://doi.org/10.1007/s10530-010-9804-x
- Blackburn T, Pyšek P, Bacher S, Carlton J, Duncan R, Jarošík V, Wilson J, Richardson D (2011) A proposed unified framework for biological invasions. Trends in Ecology & Evolution 26: 333–339. https://doi.org/10.1016/j.tree.2011.03.023
- Bosso L, Di Febbraro M, Cristinzio G, Zoina A, Russo D (2016) Shedding light on the effects of climate change on the potential distribution of *Xylella fastidiosa* in the Mediterranean basin. Biological Invasions 18(6): 1759–1768. https://doi.org/10.1007/s10530-016-1118-1
- Brunel S, Branquart E, Fried G (2010) The EPPO prioritization process for invasive alien plants. EPPO Bulletin, 40: 407–422. https://doi.org/10.1111/j.1365-2338.2010.02423.x
- Chapman D, Purse B, Roy H, Bullock J (2017) Global trade networks determine the distribution of invasive non-native species. Global Ecology and Biogeography 26. https://doi. org/10.1111/geb.12599
- Chatterjee S, Almeida R, Lindow S (2008) Living in two worlds: the plant and insect lifestyles of *Xylella fastidiosa*. Annual Review of Phytopathology 46: 243–271. https://doi. org/10.1146/annurev.phyto.45.062806.094342
- Conser C, Seebacher L, Fujino DW, Reichard S, DiTomaso JM (2015) The development of a plant risk evaluation (PRE) tool for assessing the invasive potential of ornamental plants. PLOS ONE 10(3): 16. https://doi.org/10.1371/journal.pone.0121053
- Deblauwe V, Droissart V, Bose R, Sonké B, Blach-Overgaard A, Svenning JC, Wieringa J, Ramesh B, Stévart T, Couvreur T (2016) Remotely sensed temperature and precipitation data improve species distribution modelling in the tropics. Global Ecology and Biogeography 25: 443–454. https://doi.org/10.1111/geb.12426
- Denancé N, Legendre B, Briand M, Olivier V, Boisseson C, Poliakoff F, Jacques MA (2017) Several subspecies and sequence types are associated with the emergence of *Xylella fastidiosa* in natural settings in France. Plant Pathology 66: 1054–1064. https://doi.org/10.1111/ ppa.12695

- D'hondt B, Vanderhoeven S, Roelandt S (2015) Harmonia+ and Pandora+: risk screening tools for potentially invasive plants, animals and their pathogens. Biological Invasions 17: 1869–1883. https://doi.org/10.1007/s10530-015-0843-1
- Early R, Bradley BA, Dukes JS, Lawler JJ, Olden JD, Blumenthal DM, Gonzalez P, Grosholz ED, Ibañez I, Miller LP, Sorte CJB, Tatem AJ (2016) Global threats from invasive alien species in the twenty-first century and national response capacities. Nature Communications 7: 12485. https://doi.org/10.1038/ncomms12485
- Early R, González-Moreno P, Murphy ST, Day R (2018) Forecasting the global extent of invasion of the cereal pest *Spodoptera frugiperda*, the fall armyworm. NeoBiota 40: 25–50. https://doi.org/10.3897/neobiota.40.28165
- European Food Safety Authority Panel on Plant Health (2010) Guidance on a harmonized framework for pest risk assessment and the identification and evaluation of pest risk management options by EFSA. EFSA Journal 8(2): 1495, 66. https://doi.org/10.2903/j. efsa.2010.1495
- European Food Safety Authority Panel on Plant Health (2015) Scientific opinion on the risks to plant health posed by *Xylella fastidiosa* in the EU territory, with the identification and evaluation of risk reduction options. EFSA Journal 13(1): 3989, 266. https://doi.org/10.2903/j.efsa.2015.3989
- European Food Safety Authority Panel on Plant Health (2018) Scientific report on the update of the *Xylella* spp. host plant database. EFSA Journal 16(9): 5408, 87. https://doi. org/10.2903/j.efsa.2018.5408
- European Food Safety Authority Panel on Plant Health (2019) Update of the scientific opinion on the risks to plant health posed by *Xylella fastidiosa* in the EU territory. EFSA Journal 17(5): 5665, 200.
- European and Mediterranean Plant Protection Organization (2015) *Xylella fastidiosa* detected in *Coffea* spp. plants imported into Switzerland. EPPO Reporting service 10/2015. Num. article 181/2015. https://gd.eppo.int/reporting/article-5128
- European and Mediterranean Plant Protection Organization (2016) First report of *Xylella fas-tidiosa* subsp. *fastidiosa* on Nerium oleander in Germany. EPPO Reporting service 7/2016. Num. article 133/2016. https://gd.eppo.int/reporting/article-5878
- European and Mediterranean Plant Protection Organization (2019a) First report of *Xylella fastidiosa* subsp. *multiplex* in Portugal. EPPO Reporting service 1/2019. Num. article 017/2019. https://gd.eppo.int/reporting/article-6447
- European and Mediterranean Plant Protection Organization (2019b) First report of *Xylella fastidiosa* in Israel. EPPO Reporting service 6/2019. Num. article 121/2019. https://gd.eppo. int/reporting/article-6551
- Feil H, Purcell A (2001) Temperature-dependent growth and survival of *Xylella fastidiosa in vitro* and in potted grapevines. Plant Disease 85: 1230–1234. https://doi.org/10.1094/PDIS.2001.85.12.1230
- Godefroid M, Cruaud A, Streito JC, Rasplus JY, Rossi JP (2019) *Xylella fastidiosa*: climate suitability of European continent. Scientific Reports, 9(1): 8844. https://doi.org/10.1038/ s41598-019-45365-y

- Güldür ME, Çağlar BK, Castellano MA, Ünlü L, Güran S, Yılmaz MA, Martelli GP (2005) First report of almond leaf scorch in Turkey. Journal of Plant Pathology 87(3): 246.
- Habib W, Nigro F, Gerges E, Jreijiri F, Al Masri Y, El Riachy M, Choueiri E (2016) *Xylella fastidiosa* does not occur in Lebanon. Journal of Phytopathology 164(6): 395–408. https://doi.org/10.1111/jph.12467
- Hawkins C, Bacher S, Essl F, Hulme P, Jeschke J, Kühn I, Kumschick S, Nentwig W, Pergl J, Pyšek P, Rabitsch W, Richardson D, Vilà M, Wilson J, Genovesi P, Blackburn T (2015) Framework and guidelines for implementing the proposed IUCN Environmental Impact Classification for Alien Taxa (EICAT). Diversity and Distributions 21. https://doi. org/10.1111/ddi.12379
- Kenis M, Bacher S, Baker RH (2012) New protocols to assess the environmental impact of pests in the EPPO decision-support scheme for pest risk analysis. EPPO Bulletin 42(1): 21–7. https://doi.org/10.1111/j.1365-2338.2012.02527.x
- Koop AL, Fowler L, Newton LP, Caton BP (2011) Development and validation of a weed screening tool for the United States. Biological Invasions 14: 273–294. https://doi. org/10.1007/s10530-011-0061-4
- Metz M, Rocchini D, Neteler M (2014) Surface temperatures at the continental scale: tracking changes with remote sensing at unprecedented detail. Remote Sensing 6: 3822–3840. https://doi.org/10.3390/rs6053822
- Miller TK, Allen, CR, Landis WG, Merchant, JW (2010) Risk assessment: simultaneously prioritizing the control of invasive plant species and the conservation of rare plant species. Biological Conservation, 143: 2070–2079. https://doi.org/10.1016/j.biocon.2010.05.015
- Mumford JD, Booy O, Baker R, Rees M, Copp G, Black K, Holt J, Leach A, Hartley M (2010) Invasive species risk assessment in Great Britain. Aspects of Applied Biology 104: 49–54.
- Nentwig W, Bacher S, Pyšek P, Vilà M, Kumschick S (2016) The generic impact scoring system (GISS): a standardized tool to quantify the impacts of alien species. Environmental Monitoring and Assessment, 188:315. https://doi.org/10.1007/s10661-016-5321-4
- Olmo D, Nieto A, Adrover F, Urbano A, Beidas O, Juan A, Marco-Noales E, López M, Navarro I, Monterde A, Montes-Borrego M, Navas Cortés J, Landa B (2017) First detection of *Xy-lella fastidiosa* on cherry (*Prunus avium*) and *Polygala myrtifolia* plants, in Mallorca Island, Spain. Plant Disease 101. https://doi.org/10.1094/PDIS-04-17-0590-PDN
- Rogg H, Buddenhagen C, Causton C (2003) Experiences and limitations with pest risk analysis in the Galapagos Islands. https://www.fao.org/3/y5968e/y5968e0m.htm
- Sandvik H, Sæther BE, Holmern T, Tufto J, Engen S, Roy HE (2013) Generic ecological impact assessments of alien species in Norway: a semi-quantitative set of criteria. Biodiversity and Conservation 22: 37–62. https://doi.org/10.1007/s10531-012-0394-z
- Saponari M, Boscia D, Nigro F, Martelli GP (2013) Identification of DNA sequences related to *Xylella fastidiosa* in oleander, almond and olive trees exhibiting leaf scorch symptoms in Apulia (southern Italy). Journal of Plant Pathology 95(3): 659–668.
- Saponari M, Loconsole G, Cornara D, Yokomi RK, De Stradis A, Boscia D, Bosco D, Martelli GP, Krugner R, Porcelli F (2014) Infectivity and Transmission of *Xylella fastidiosa* by *Philaenus spumarius* (Hemiptera: Aphrophoridae) in Apulia, Italy. Journal of Economic Entomology 107(4): 1316–1319. https://doi.org/10.1603/EC14142

- Seebens H, Essl F, Dawson W, Fuentes N, Moser D, Pergl J, Pyšek P, Van Kleunen M, Weber E, Winter M, Blasius B (2015) Global trade will accelerate plant invasions in emerging economies under climate change. Global Change Biology 21(11): 4128–4140. https://doi. org/10.1111/gcb.13021
- Tatem A (2009) The worldwide airline network and the dispersal of exotic species: 2007–2010. Ecography 32(1): 94–102. https://doi.org/10.1111/j.1600-0587.2008.05588.x
- Temsah M, Hanna I, Saad A (2015) First report of *Xylella fastidiosa* associated with oleander leaf scorch in Lebanon. Journal of Crop Protection 4(1): 131–137.
- Venette S (2015) Assessing the quality of pest risk models. In: Venette RC (Ed.) Pest risk modelling and mapping for invasive alien species. CAB International, Boston, 223–233. https://doi.org/10.1079/9781780643946.0223

Supplementary material I

Figures S1, S2

Authors: Michel Frem, Daniel Chapman, Vincenzo Fucilli, Elia Choueiri, Maroun El Moujabber, Pierfederico La Notte, Franco Nigro

Data type: measurement

- Explanation note: To map gradients of *Xylella fastidiosa* relative climate suitability in Europe and Middle East and North Africa countries.
- Copyright notice: This dataset is made available under the Open Database License (http://opendatacommons.org/licenses/odbl/1.0/). The Open Database License (ODbL) is a license agreement intended to allow users to freely share, modify, and use this Dataset while maintaining this same freedom for others, provided that the original source and author(s) are credited.

Link: https://doi.org/10.3897/neobiota.59.53208.suppl1

Supplementary material 2

Table S1

Authors: Michel Frem, Daniel Chapman, Vincenzo Fucilli, Elia Choueiri, Maroun El Moujabber, Pierfederico La Notte, Franco Nigro

Data type: measurement

- Explanation note: Importation of plants for planting from countries, source of *Xylella fastidiosa*.
- Copyright notice: This dataset is made available under the Open Database License (http://opendatacommons.org/licenses/odbl/1.0/). The Open Database License (ODbL) is a license agreement intended to allow users to freely share, modify, and use this Dataset while maintaining this same freedom for others, provided that the original source and author(s) are credited.

Link: https://doi.org/10.3897/neobiota.59.53208.suppl2

Supplementary material 3

Table S2

Authors: Michel Frem, Daniel Chapman, Vincenzo Fucilli, Elia Choueiri, Maroun El Moujabber, Pierfederico La Notte, Franco Nigro

Data type: measurement

Explanation note: Correlation matrix of the ranking model.

Copyright notice: This dataset is made available under the Open Database License (http://opendatacommons.org/licenses/odbl/1.0/). The Open Database License (ODbL) is a license agreement intended to allow users to freely share, modify, and use this Dataset while maintaining this same freedom for others, provided that the original source and author(s) are credited.

Link: https://doi.org/10.3897/neobiota.59.53208.suppl3

Supplementary material 4

Table S3

Authors: Michel Frem, Daniel Chapman, Vincenzo Fucilli, Elia Choueiri, Maroun El Moujabber, Pierfederico La Notte, Franco Nigro

Data type: measurement

- Explanation note: Rank categorization of European and Middle East and North Africa countries according to exposure to invasion by *Xylella fastidiosa* and classified from the highest to the least overall risk rank.
- Copyright notice: This dataset is made available under the Open Database License (http://opendatacommons.org/licenses/odbl/1.0/). The Open Database License (ODbL) is a license agreement intended to allow users to freely share, modify, and use this Dataset while maintaining this same freedom for others, provided that the original source and author(s) are credited.

Link: https://doi.org/10.3897/neobiota.59.53208.suppl4

Table S4

Authors: Michel Frem, Daniel Chapman, Vincenzo Fucilli, Elia Choueiri, Maroun El Moujabber, Pierfederico La Notte, Franco Nigro

Data type: occurrence

Explanation note: The known world distribution of Xylella fastidiosa.

Copyright notice: This dataset is made available under the Open Database License (http://opendatacommons.org/licenses/odbl/1.0/). The Open Database License (ODbL) is a license agreement intended to allow users to freely share, modify, and use this Dataset while maintaining this same freedom for others, provided that the original source and author(s) are credited.

Link: https://doi.org/10.3897/neobiota.59.53208.suppl5

RESEARCH ARTICLE



Effects of a recalcitrant understory fern layer in an enclosed tropical restoration experiment

Luis C. Beltrán¹, Karla María Aguilar-Dorantes², Henry F. Howe¹

 I Department of Biological Sciences (m/c 066), University of Illinois at Chicago, Chicago IL 60607, USA
2 Centro de Investigación en Biodiversidad y Conservación, Universidad Autónoma del Estado de Morelos, Avenida Universidad 1001, Col. Chamilpa, Cuernavaca, Morelos, C.P. 62209, México

Corresponding author: Luis C. Beltrán (lbeltr6@uic.edu)

Academic editor: Elizabeth Wandrag | Received 11 March 2020 | Accepted 9 July 2020 | Published 30 July 2020

Citation: Beltrán LC, Aguilar-Dorantes KM, Howe HF (2020) Effects of a recalcitrant understory fern layer in an enclosed tropical restoration experiment. NeoBiota 59: 99–118. https://doi.org/10.3897/neobiota.59.51906

Abstract

Establishing mixed-species tree plantings and fencing them to protect seedlings from herbivory is a valuable strategy for reconnecting forest fragments separated by agropastoral lands. However, fencing may provide exotic plants with the escape from herbivory required to invade the understory of planted communities. Here we take advantage of such a situation to ask how the identity of planted species and the resulting canopy cover influenced invasion success by the Asian swordtail fern (Nephrolepis brownii Desv. Nephrolepidaceae) in a 13-year-old tropical restoration experiment. Through a seed addition experiment, we also evaluated the effects the ferns had on recruiting seedlings. We found that the invasion was most acute in the unplanted control plots where canopy cover was consistently scarce. Frond density correlated negatively with canopy cover, though most of the variance in the model is explained by the design of our experiment ($r^2m = .161$, $r^2c = .460$). Between planting treatments that differed in the dispersal mode of the planted trees, the wind-dispersed treatment had higher fern density and longer fronds than the animaldispersed treatment. The animal-dispersed treatment had the highest recruiting species richness, which was negatively correlated with fern density ($r^2 = .748$). The seed addition experiment confirmed that mortality rates increased where frond density was higher ($F_{1,41} = 7.159$, p = .011) and germination rates were lowered for the smaller-seeded species ($F_{1,42}$ = 13.2, p = .002). To prevent recalcitrant understory layers from establishing in plantings in the future, we recommend: (1) establishing larger plantings or expanding existing ones to minimize edge effects (particularly light filtration), (2) supplementing young plantings with additional seedlings to prevent canopy gaps from forming, and (3) planting an assemblage of species that cover the full forest strata and have consistently full tree-canopies.

Keywords

Edge effects, invasive species, Los Tuxtlas Biosphere Reserve, *Nephrolepis brownii*, tropical restoration plantings, Veracruz

Copyright Luis C. Beltrán et al. This is an open access article distributed under the terms of the Creative Commons Attribution License (CC BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Introduction

Mixed-species plantings are established to enhance the natural succession of forests in degraded lands (Parrotta and Knowles 2001; Lamb et al. 2005; de la Peña-Domene et al. 2013). These plantings can be used to reconnect forest fragments as a stepping-stone biological corridor, a design that provides greater flexibility of land use for local land-owners than continuous biological corridors do (Rey Benayas et al. 2008). For example, the space in between plantings can be used as pasture for cattle, so long as the plantings are fenced to prevent cattle from devouring recruiting seedlings (Holl and Quiros-Nietzen 1999; Aide et al. 2000). However, fences will protect plants indiscriminately, including unwanted exotics. Under favorable light conditions, release from herbivory can enable exotic plant species to act as aggressive invaders, altering the structure of the understory and potentially the fate of the restored patches (Keane and Crawley 2002).

Invasive plant species can arrest succession by dominating the plant understory through resource competition and allelopathy, resisting displacement by native species for an extended period of time (Young and Peffer 2010). A dense understory can also benefit small seed-eating mammals such as rodents, elevating the rate of seed predation under the invaders' canopy (den Ouden 2000). An understory characterized by such pervasive disruptors of succession is known as a 'recalcitrant understory layer' (Royo and Carson 2006). Species capable of forming this layer typically have three key characteristics: (1) rapid vegetative growth (e.g., often spreading through rhizomes), (2) reduced susceptibility to elimination, and (3) have long-living genets (Young and Peffer 2010). Eliminating a recalcitrant understory layer is possible when a dense tree canopy cover has already been established, but this is time-consuming and labor-intensive (Hill and Silander 2001; Douterlungne et al. 2010).

Ferns provide multiple examples of recalcitrant understory layers (Horsley 1993; George and Bazzaz 1999a, b; Marrs et al. 2008). Their ability to suppress seed germination is supplemented by slow decomposition of fern litter that tends to accumulate under fern thickets (George and Bazzaz 1999b; Allison and Vitousek 2004). One fern species with potential for establishing recalcitrant understory layers is the Asian sword-tail (*Nephrolepis brownii* Desv.) (Hovenkamp and Miyamoto 2005; Sharpe and Shiels 2014). It is naturalized in the tropical Americas where it is considered an invasive species with important economic and ecological implications (Proctor 1989; Robinson et al. 2010; CABI 2019). Where established, it is common from sea level to 1700 m in forests, although it is most common in open vegetation (e.g., roadsides, riverbanks, open thickets). This species' ability to arrest succession has not been evaluated, despite its continued presence in some tropical restoration plantings in Veracruz, Mexico.

Here we use the presence of *N. brownii* in our 13-year-old restoration planting experiment in tropical Mexico to assess (1) the invasibility of these plantings and (2) the effect that *N. brownii* has on recruiting seedlings. For the former, we evaluate the extent of the invasion using frond density (m^{-2}) by restoration treatment and in response to canopy cover. The restoration treatments include two planting treatments that differ in the dispersal mode of the planted trees (wind-dispersal vs. animal-dispersal) and an unplanted control simulating natural succession. Canopy cover differs between plots

but is especially lacking in the control treatment. Given how species capable of forming recalcitrant understory layers typically benefit from lacking tree cover (Royo and Carson 2006), we predict that the ferns would be more numerous in the unplanted controls and in plantings with lower tree-canopy cover. As an additional assessment of invasive success, we use frond length measurements, predicting that they correlate with frond density. This hypothesis is based on previous studies that have shown reduced size-density tradeoffs for invasive plants, allowing them to dominate the understory (Jakobs et al. 2004; Royo and Carson 2006). We also predict that other disturbanceadapted species capable of dominating understory, such as exotic grasses and hemiepiphytic lianas, will be more common in the unplanted control treatment.

To assess the effect *N. brownii* has on plant recruitment we compared recruited woody species richness in response to frond density. We predicted that higher frond density would result in lower species richness. To verify that this pattern was not purely correlative, we established a seed addition experiment under varying frond densities to evaluate seedling performance. We predicted that small-seeded tree species would be most affected by frond density, with lower germination, higher mortality, and lower biomass than larger-seeded species. Examining the invasibility of these plantings revealed design improvements that could prevent future plantings from developing a recalcitrant understory of *N. brownii*, and possibly of other invasive species.

Methods

Study area

The study was conducted in the Los Tuxtlas Biosphere Reserve (~3,300 km²), in the state of Veracruz, Mexico. Mean annual rainfall in the region is 4,900 mm and temperature 27 °C. Dry seasons typically extend from March through May and rainy seasons from June to February (González-Soriano et al. 1997). Deforestation driven by the demand for cattle pasture has resulted in highly fragmented forests in the region (Mendoza et al. 2005). These fragments are home to +350 tree species, the majority of which are dispersed by animals (Ibarra-Manríquez et al. 2015).

Restoration Plantings Experiment

Restoration plantings were established in June 2006 to facilitate the movement of plants and animals in a permanent agricultural landscape. The rationale was to determine if planted animal-dispersed tree species accelerated tropical forest succession by attracting more dispersal agents that bring forest seeds with them than planted wind-dispersed stands or unplanted controls. The experiment consisted of 24 30 × 30 m fenced plots set on an 8×3 grid, each separated by 35 m of actively grazed pasture (central GPS point $18^{\circ}35'43.64"$ N, $95^{\circ}06'06.29"$ W). The plantings included three treatments: Eight plots were planted with seedlings of 12 native animal-dispersed spe-

cies (animal), eight with seedlings of 12 native wind-dispersed species (wind), and eight were left unplanted to simulate natural succession (control) (Fig. 1). Each plot is divided into four subplots of 13×13 m separated by a 4 m wide open corridor. Some planted species in the animal-dispersed treatment like *Ficus yoponensis* (Desv.), *Inga sinacae* (M. Sousa & Ibarra-Manr.), and *Stemmandenia donnel-smithii* (Rose) have been fruiting consistently since 2010. However, seed rain between planted treatments has not differed significantly (Popoca-Ortega 2016).

To date, seedlings of 78 species of trees that were not planted have recruited in experimental plots (unpublished data from 2019 census). For details on the experimental design, the study area, and early-recruitment patterns, see de la Peña-Domene et al. (2013).

Study species

Nephrolepis brownii is a terrestrial, sometimes epiphytic fern, native to Southeast Asia. This species is more drought-resistant than most ferns due in great part to its erect rhizomes (Hovenkamp and Miyamoto 2005). The capacity of *N. brownii* to spread through underground runners allows it to quickly exploit gaps in the canopy (Sharpe and Shiels 2014), a characteristic seen in other fern species capable of forming a recalci-



Figure 1. Experimental design of the 24 restoration plots belonging to three treatments: (A) plots planted with animal-dispersed trees, (W) plots planted with wind-dispersed trees, and (C) unplanted control plots. Adapted from de la Peña-Domene et al. (2013).

trant understory layer (Young and Peffer 2010). Aboveground, this species typically has 5 or 6 fronds per ramet, though the ramets themselves can be difficult to distinguish in the field. In this study, we use frond counts to estimate fern density. A full description of *N. brownii* is provided by Hovenkamp and Miyamoto (2005). Specimens collected from this study are deposited in the herbarium HUMO of the Center for Biodiversity and Conservation Research of the Autonomous University of the State of Morelos and the herbarium of the Biological Station of Los Tuxtlas.

Census of N. brownii and Recruiting Seedlings

To evaluate the extent of the invasion, we randomly established 4 1 m² quadrats (total of 384) within each subplot of every plot. To select quadrats, we used a random number generating code in RStudio to yield four numbers from 1 to 169, each number representing a possible quadrat in the 13 m² subplots. We counted the number of fronds within each quadrat. We also recorded the presence of grass or hemiepiphytic lianas in the quadrats as percent cover where present. This was done to account for other types of understory layers that might suppress tree recruitment. Quadrats were treated as "grass- or liana-dominated" when percent cover exceeded 50%. To obtain frond length measurements, we haphazardly selected five unfurled fronds per quadrat to measure. The recruiting plant community (> 10 cm tall woody plants) within the 1 m^2 guadrats and the trees creating canopy cover over the quadrat (recruits over 1.3 m tall) were recorded and identified to species. The latter was done to determine the frequency with which planted trees provided canopy cover over the recruiting seedlings and/or ferns. To estimate canopy cover, a photograph of the canopy covering each quadrat was taken from the middle of the quadrat at a height of ~ 1.3 m and analyzed using the environmental application for iOS devices, %Cover. The raw results of this study are available in Suppl. materials 4, 5.

Seed Addition Experiment

To determine whether fern cover affected the germination, growth, and survivorship of tree seedlings, we established a seed addition experiment within the restoration plots. Two late-successional, animal-dispersed tree species characteristic of the forest of Los Tuxtlas, *Brosimum alicastrum* (Sw.) and *Guarea grandifolia* (DC.), were selected for this study. Previous censuses showed that recruitment of these two species in the plantings were rare, which allowed us to more easily keep track of the fate of added seeds.

Seeds of animal-dispersed *B. alicastrum* $(8-13 \times 13-18 \times 13-18 \text{ mm})$ and *G. grandifolia* $(13-25 \times 10-15 \times 8-15 \text{ mm})$ were collected from 20 and 4 fruiting adults respectively in early-to-mid June of 2018 (seed dimensions from Ibarra-Manríquez et al. 2015). Seeds were cleared of remaining aril and wiped with 10% Ethanol before being dispersed to a randomly selected subplot within each plot in late June. Each selected subplot received 20 *B. alicastrum* and 12 *G. grandifolia* seeds spread out around four equidistant points close to the center of the subplot to minimize competition between them. The specific location of the added seeds was marked with stake flags and seedlings were tagged with flagging tape as they germinated. Monthly censuses were carried out for a year to record germination, mortality, and seedling height. Seedlings were harvested in late June of 2019 to assess performance through biomass allocation: (1) total leaf count, (2) leaf dimensions (length and width), (3) leaf mass, (4) root mass, (5) shoot mass, (6) root-to-shoot ratio, and (7) total seedling mass. Mass measurements were taken after the seedlings were dried for 48 hours in a desiccating chamber at 40 °C. The raw results of this experiment are available in Suppl. material 6.

Statistical Analysis

Differences in average frond density were analyzed using a generalized mixed-effects linear model with a negative-binomial distribution where the nested design (quadrats within subplots within plots) was included as a random effect. Differences in frond length between treatments were analyzed using a mixed-effects linear model where each frond is the experimental unit and the random nested effect accounts for this additional level (fronds within quadrats within subplots within plots). To evaluate the relationship between frond density and frond length however, we used average frond length per quadrat to avoid pseudo-replication. Here too, we used a mixed-effects model with the nested design of the census at a quadrat scale included as a random effect. The relationship between frond density with canopy cover was also analyzed using a mixed-effects linear model with the same random nested effect as the previous model. An ANCOVA was used to determine if the relationship between frond density and canopy cover differed by treatment. Differences in recruiting species richness between treatments were analyzed using a generalized linear model with a Poisson distribution log link. Linear regressions were used to evaluate the effect of frond density on species richness. An ANCOVA was used to determine if the relationship between frond density and recruiting species richness varied by treatment. For the species richness models, data was pooled by plot in light of how smaller scales artificially limit the variable's range of values. Quadrats that were grass- or liana- dominated (> 50% cover) were excluded from models that used frond density or length as a variable.

To determine if *B. alicastrum* and *G. grandifolia* differed in their response to frond density, we compared the relationship between their biomass features and frond density using ANCOVA. Frond density in these cases corresponded to the averaged frond density of the subplot where the seeds had been sown. The analyses were carried out at a subplot scale to avoid pseudo-replication of frond density values and because quadrat measurements from the earlier census did not correspond to where seeds were sown. When the interaction was significant, we evaluated the effects of frond density using linear regressions for each species.

In January 2019, trespassers cut some of the germinated seedlings. Most of the cut seedlings survived, but all were excluded from the mortality and biomass features

ANCOVAs. All analyses were carried in RStudio ("R version 3.6.1 (2019-07-05)") using the lme4 package to fit linear and generalized mixed-effects models (Bates et al. 2015), the MuMin package to estimate marginal (r^2m , considers only variance of fixed effects) and conditional correlation coefficients (r^2c , takes both fixed and random effects into consideration) (Barton 2015), and the lmerTest package to obtain p-values and approximations for denominator degrees of freedom for the mixed-effects models (Kuznetsova et al. 2017). The sjPlot package (Lüdecke 2019) was used to create mixed-effects model summaries (Suppl. material 7).

Results

Extent of Invasion of N. brownii

Of the 384 1 m² quadrats established for the census, 306 (80%) were recorded with N. brownii. The degree of invasion follows a progression by treatment; the control treatment has the highest degree of fern infestation (94.53% of quadrats have ferns), followed by the wind-dispersed treatment (82.81%), and then the animal-dispersed treatment (71.88%) (Fig. 2). Pairwise comparisons reveal that the control quadrats have significantly higher frond density than the animal-dispersed treatment (z = 3.417, p > .001) and marginally higher frond density than the wind-dispersed treatment (z = 1.729, p = .083). Differences between planting treatments were marginally significant (z = 1.743, p = .081; Fig. 3A). The controls also had longer fronds than the animal- (ANOVA, t = 7.056, p < .001) and wind-dispersed stands (t = 5.49, p < .001). Differences between planted treatments were marginally significant (t = 1.829, p = .068; Fig. 3B). Frond density is positively correlated with average frond length ($F_{1,252.76}$ = 128.8, p > .001, $r^2m = .327$, r^2c = .444, Fig. 4A) and negatively correlated with canopy cover. In the case of the latter, more of the variance in the model is accounted for when the nested design of the census is considered ($F_{1,311,29} = 33.665, p > .001, r^2m = .161, r^2c = .460$, Fig. 4B). The relationship between frond density and canopy cover does not differ between treatments (ANCOVA, $F_{2,351.65}$ = 1.126, p = .325). Control plots had more quadrats dominated by grass or lianas (18/128) than did animal- (3/128) and wind-dispersed treatments (1/128).

Effect of Invasion on Recruitment

Within the quadrats established for the census, a total of 54 woody plant species recruited, the majority of them animal-dispersed and all of them native except for *Citrus* sp. Recruited species richness was lower in controls (average = 6.8) than either planting treatment, though only marginally to the wind-dispersed treatment (vs. animal: average = 10.5, z = 2.533, p = .011, vs. wind: average = 9.38, z = 1.841, p = .066). Differ-



Figure 2. Ground cover of ferns, grasses, or lianas for the 384 1 m² quadrats excluding recruiting and planted forest plants. Each colored square represents a 1 m² quadrat in the (A) animal-dispersed, (W) wind-dispersed, or (C) unplanted control plots. Grass and liana quadrats were characterized as such when more than 50% of the ground cover was composed of grasses or lianas.



Figure 3. A boxplots of frond density (m^2) and **B** frond length (cm) by restoration treatment. Lower and upper box boundaries represent 25^{th} and 75^{th} percentiles, respectively. Lower and upper error lines represent Q1 - 1.5 * IQR and Q3 + 1.5 * IQR, respectively. Middle band represents median values.

ences in richness between plantings were not statistically significant (z = .713, p = .476; Fig. 5A). The relationship between frond density and species richness varied marginally between treatments (ANCOVA, $F_{2,18} = 2.823$, p = .086). In the animal-dispersed treatment there was a strong, negative correlation between species richness and frond density ($F_{1.6} = 17.84$, p = .006, $r^2 = .748$). No significant relationships were observed between these variables in the wind-dispersed ($F_{1.6} = 2.224$, p = .186, $r^2 = .271$) and control treatments ($F_{1.6} = .929$, p = .372, $r^2 = .134$, Fig. 5B).



Figure 4. Relationship between frond density with **A** frond length and **B** percent canopy cover. Frond length values represent an average from five randomly selected fronds per quadrat. Band shades represent 95% confidence intervals.

Seed addition experiment

Frond density differentially influenced seedling performance by species (Fig. 6). The relationship between frond density and germination rates differed between *B. alicas-trum* and *G. grandifolia* (ANCOVA, $F_{1,42} = 9.88$, p = .003). For *B. alicastrum*, germination rates were negatively correlated with frond density (linear regression, $F_{1,21} = 13.2$, p = .002, $r^2 = .386$) while *G. grandifolia* showed no trend (linear regression, $F_{1,21} = .484$, p = .494, $r^2 = .023$). Species also differed marginally in the seedling height response to frond density (ANCOVA, $F_{1,36} = 3.83$, p = .058). *Brosimum alicastrum* showed no trend (linear regression, $F_{1,17} = .862$, p = .366, $r^2 = .05$) while *G. grandifolia* showed taller seedlings as frond density increased (linear regression, $F_{1,21} = 5.424$, p = .030, $r^2 = .205$).

Frond density had marginally significant, negative effects on total leaf count ($F_{3,38} = 3.103$, p = .086, $r^2 = .196$). This pattern was similar for the two species (AN-COVA, $F_{3,38} = .731$, p = .398). Leaf length, leaf width, leaf mass, total seedling mass, shoot mass, and roots mass did not show trends in response to frond density (p > .1). However, root-to-shoot ratios did show a negative correlation with frond density for *G. grandifolia* (linear regression, $F_{1,21} = 5.06$, p = .035, $r^2 = .194$) that was not observed with *B. alicastrum* (linear regression, $F_{1,17} = .289$, p = .548, $r^2 = .017$). An ANCOVA confirmed that the species differ marginally in how frond density influenced their root-to-shoot ratios (ANCOVA, $F_{3,38} = 2.88$, p = .098).



Figure 5. Species richness by **A** treatment and in response to **B** frond density at a plot scale. Error bars represent standard error. Band shade represents 95% confidence intervals. Dashed lines indicate non-significant correlations (p > .1). Solid lines indicate significant correlations (p < .05).

Mortality rates increased with increasing frond density ($F_{3,41} = 3.38, p < .001$). This pattern was similar for both species (ANCOVA, $F_{3,41} = .245, p = .624$), although *B. alicastrum* showed higher mortality rates than *G. grandifolia* ($F_{3,38} = 7.159, p = .011$). *Brosimum alicastrum* seedlings were also unable to survive past beyond 53 fronds per m² while *G. grandifolia* seedlings still survived in areas with over 80 fronds per m².

Discussion

Our results show that restoration treatments influenced the development of recalcitrant understory layers of *N. brownii*. Unplanted control plots meant to simulate natural succession were the most affected; this treatment had both the highest frond density and longest frond length. This clearly shows that low development of tree canopies enables invasion by exotic species released from herbivory. Effects of tree cover are also supported by the negative relationship between frond density and tree-canopy cover across treatments, which our model confirmed is largely explained by the design of our experiment. The importance of early development of a tree canopy layer is consistent with practices of the indigenous Lacandon Maya people of Chiapas, Mexico, who plant rapidly growing balsa (*Ochroma pyramidale* Cav. ex Lam. Urb.) to create a dense, wide canopy capable of suppressing the growth of another invasive fern species


Figure 6. Seedling biomass responses of *Brosimum alicastrum* and *Guarea grandifolia* to frond density (m²). Frond density values are the averages of each subplot's quadrat measurements where seeds were sown. Band shade represents 95% confidence intervals. Dashed lines indicate non-significant correlations (p > .1). Two-dashed lines indicate marginally significant correlations (.05). Solid lines indicate significant correlations (<math>p < .05).

(*Pteridium aquilium*, Douterlungne et al. 2010; 2013). Balsa was one of the species we planted in the wind-dispersed treatment, but its low representation in the stands (1/12 of planted trees) likely limited its fern-suppressing attributes.

Differences between planting treatments provide additional insights. The animaldispersed treatment had marginally lower frond density (average: $18.44/m^2$) than the wind-dispersed treatment (average: $25.42/m^2$). Though not as large a difference as that between either planting treatment or the control (average: $38.02/m^2$), it suggests that the assemblage of planted species influences the invasibility of the plantings. One likely reason why the wind-dispersed treatment is more vulnerable could be the higher incidence of canopy gaps (Suppl. material 1). Considering this treatment has $1.3\times$ more surviving planted trees, this is surprising (Suppl. material 2). However, when we determined the tree species that provided canopy cover over our quadrats, we found that relative to the animal-dispersed treatment, it was $1.2\times$ less likely for planted trees to be part of the canopy cover in the wind-dispersed treatment (Suppl. material 1). In other words, the boon the animal-dispersed treatment provides in curtailing the fern invasion is likely due to the growth patterns of the planted species that yield wider, more permanent canopies with multiple layers than the species we planted in wind-dispersed plots.

Differences between planted treatments are likely a byproduct of our experimental design; the species planted were chosen by dispersal mode and successional status (4 pioneers and 8 late-successionals) without focusing on the strata they filled in the forest or their phenology. As a consequence of this, the wind-dispersed treatment had one small tree species, three subcanopy species, and eight canopy tree species. In contrast, the animal-dispersed treatment is more balanced with three small, three subcanopy, and six canopy tree species (Suppl. material 3). Additionally, wind-dispersed species usually drop their leaves to maximize seed dispersal during the dry season, thereby enhancing light penetration for at least four months of a year (reviewed by van Schaik et al. 1993). Leaf drop is consistent with the increase in leaf-litter mass we detected in wind-dispersed plots (Valencia-Esquivel 2012).

Our results are consistent with the generalization that species selection influences the development of forest structure and the return of biodiversity (Parrotta and Knowles 2001; Sansevero et al. 2011; Longworth and Williamson 2018). We consider it important that future planting designs consider the strata that the assemblage of planted species will fill as they mature as well as their deciduous phenology and dispersal mode. The development of multiple, permanent or semi-permanent canopy layers will restrict the growth of recalcitrant understory layers. Furthermore, we suggest that practitioners prevent the development of canopy gaps by supplementing plantings with additional planted seedlings where necessary.

It is noteworthy that in our plantings, even a full canopy cover with multiple layers failed to prevent the invasion of *N. brownii*. This is likely due to the small size of the plots. At 30×30 m, these plantings experience adverse microclimatic conditions from edge effects known to alter forest structure (Magnago et al. 2015). The clearest way to minimize edge effects, including the light regimes that benefit invasives, is to maximize the size of the plantings (Laurance 2008).

Our survey of the plantings also confirmed that recruiting species richness is lower where frond density is higher, though this relationship was only statistically significant in the animal-dispersed treatment. The controls showed lower species richness than the planting treatments, likely because the unplanted plots are also subject to greater seed dispersal limitations that can explain the lower recruitment rates (Popoca-Ortega 2016). Nonetheless, results from the seed addition experiment strongly suggest that the relationship between frond density and species richness is a consequence of the adverse effects of fern cover on recruiting seedlings. The two sown seed species were both suppressed by fern invasion but performed differently in response to frond density. *Brosimum alicastrum* showed lower germination rates, and higher mortality rates in response to frond density while *G. grandifolia* showed greater shoot length, lower root-to-shoot ratio, and higher mortality rates. The difference in germination rates can be explained by differences in seed size between species. *Guarea grandifolia* produces seeds up to 3× larger than *B. alicastrum* (calculated from seed dimensions provided by Ibarra-Manríquez et al. 2015). Greater seed mass provides germinating seedlings with more resources with which to grow in closed canopy conditions (Foster and Janson 1985). For *G. grandifolia*, resources are evidently invested in shoot length when the seedlings are covered by the understory canopy of *N. brownii*. This is reflected in the root-to-shoot ratios too; for *G. grandifolia*, the ratio increases with increasing frond density while *B. alicastrum* shows no pattern.

The differences these species exhibit in germination suggest that the tree species most likely to grow past the understory fern layer are those with greater seed mass. This is problematic in the context of restoration plantings because large-seeded species are less likely to be dispersed across inhospitable matrices (e.g., pastures) than small-seeded species (reviewed by Wunderle 1997; Beltran and Howe 2019). Even if these large seeds are dispersed to fern-infested plots, mortality rates are still higher where fronds are denser. Ultimately this means that invasion by *N. brownii* limits tree seedling recruitment and shapes the seedling community by favoring seeds large enough to germinate and grow past the understory but small enough to be dispersed to the plantings.

We do not observe *N. brownii* growing outside the plots as well as it does inside of them. This is likely because of the cows that graze the surrounding pasture; we have observed them eating the fronds that manage to grow past the outline of the fences. However, eliminating the fences in an attempt to suppress the ferns would terminate restoration; livestock eat tree seedlings as well as *N. brownii*. We have also not found *N. brownii* growing within the Los Tuxtlas primary forest, which is likely because of the deep shade cast by multiple layers of trees (some taller than 30 m) and/or endemic grazers that browse understory plants (e.g., *Tayasu pecari*, *Sylvilagus* spp.). One solution to restoration plantings facing this problem may be direct seeding and/or planting of large-seeded tree species capable of growing past the ferns and shading them out. However, further research with additional tree species is required to ascertain how an understory layer of *N. brownii* affects species of different seed size and life history.

Conclusion

Nephrolepis brownii suppressed tropical forest succession in our tropical restoration plantings experiment. The fern invasion was most successful in unplanted control plots, highlighting the importance of tree canopy cover. In the animal-dispersed treatment where recruiting species richness was highest, there was also a negative correlation between frond density and tree-seedling recruitment. Our seed addition experiment confirmed that this inverse relationship was not simply correlative but caused by the ferns themselves. Future studies should evaluate how restoration plantings can be designed to make them less likely to develop recalcitrant understory layers. While fences cannot be removed to allow ungulate grazers in, we suggest testing the effects of (1) planting an assemblage of species that cover multiple forest strata and have more permanent tree canopies, (2) supplementing canopy gaps with additional seedlings, and/or (3) expanding the dimensions of the plantings.

Acknowledgements

We thank the staff at the Los Tuxtlas Biological Station for their support and Dario Velasco, Jorge Velasco, and Ivonne Popoca for their field assistance. We are grateful to Santiago Sinaca Colín and Álvaro Campos-Villanueva for assistance identifying tree seedlings. We are also grateful to the members of the Wise Lab at the University of Illinois, Chicago (UIC) for comments on the manuscript and the Golden Generation of Los Tuxtlas for support during the field seasons. We are grateful to Ann Marie Gawel, Dalva M. Silva Matos, Cristina Martinez-Garza, and Elizabeth Wandrag for constructive comments on the manuscript and Aaron Goodman for advice on statistical analyses and approaches. This study was supported by two Elmer Hadley Graduate Research Grants from the Department of Biological Sciences at UIC. The restoration plantings were established and maintained with funds from the United States National Science Foundation (DEB 0516259), the University of Illinois, and the National Geographic Society (9302-13). The authors declare no competing interests.

References

- Aide TM, Zimmerman JK, Pascarella JB, Rivera L, Marcano-Vega H (2000) Forest regeneration in a chronosequence of tropical abandoned pastures: amplications for restoration acology. Restoration Ecology 8: 328–338. https://doi.org/10.1046/j.1526-100x.2000.80048.x
- Allison SD, Vitousek PM (2004) Extracellular Enzyme Activities and Carbon Chemistry as Drivers of Tropical Plant Litter Decomposition1. Biotropica 36(3): 285–296. https://doi. org/10.1646/03180
- Barton MK (2015) MuMIn: MuMIn: Multi-Model Inference. R package version 1.43.17. http://cran.r-project.org/package=MuMIn
- Bates D, Machler MM, Bolker B, Walker S (2015) Fitting Linear Mixed-Effects Models Using {lme4}. Journal of statistical software 67: 1–48. https://doi.org/10.1007/978-94-009-4812-9_8
- Beltrán LC, Howe HF (2019) The frailty of tropical restoration plantings. Restoration Ecology 28: 16–21. https://doi.org/10.1111/rec.13066
- CABI (2019) Invasive Species Compendium. http://www.cabi.org/isc
- de la Peña-Domene M, Martínez-Garza C, Howe HF (2013) Early recruitment dynamics in tropical restoration. Ecological Applications 23: 1124–1134. https://doi.org/10.1890/12-1728.1
- de la Peña-Domene M, Minor ES, Howe HF (2016) Restored connectivity facilitates recruitment by an endemic large-seeded tree in a fragmented tropical landscape. Ecology 97: 2511–2517. https://doi.org/10.1002/ecy.1459
- den Ouden J (2000) The Role of Bracken (*Pteridium aquilinum*) in forest dynamics. Wageningen, Netherlands: Wageningen University.
- Douterlungne D, Levy-Tacher SI, Golicher DJ, Dañobeytia FR (2010) Applying Indigenous Knowledge to the Restoration of Degraded Tropical Rain Forest Clearings Dominated by Bracken Fern. Restoration Ecology 18: 322–329. https://doi.org/10.1111/j.1526-100X.2008.00459.x

- Douterlungne D, Thomas E, Levy-Tacher SI, Paynter Q (2013) Fast-growing pioneer tree stands as a rapid and effective strategy for bracken elimination in the Neotropics. Journal of Applied Ecology: 1257–1265. https://doi.org/10.1111/1365-2664.12077
- Foster S, Janson CH (1985) The Relationship between Seed Size and Establishment Conditions in Tropical Woody Plants. Ecology 66: 773–780. https://doi.org/10.2307/1940538
- George LO, Bazzaz FA (1999a) The Fern Understory as an Ecological Filter: Emergence and Establishment of Canopy-Tree Seedlings. Ecology 80: 833–845. https://doi. org/10.1890/0012-9658(1999)080[0833:TFUAAE]2.0.CO;2
- George LO, Bazzaz FA (1999b) The Fern Understory as an Ecological Filter: Growth and Survival of Canopy-Tree Seedlings. Ecology 80: 846–856. https://doi.org/10.1890/0012-9658(1999)080[0846:TFUAAE]2.0.CO;2
- González-Soriano E, Dirzo R, Vogt RC (1997) Historia natural de los Tuxtlas. Universidad Autonoma de Mexico (UNAM), Instituto de Biologia, Instituto de Ecología, Comisión Nacional para el Conocimiento y Uso de la Biodiversidad (CONABIO), Mexico.
- Hill JD, Silander JA (2001) Distribution and dynamics of two ferns: *Dennstaedtia puncti-lobula* (Dennstaedtiaceae) and *Thelypteris noveboracensis* (Thelypteridaceae) in a Northeast mixed hardwoods-hemlock forest. American Journal of Botany 88: 894–902. https://doi.org/10.2307/2657041
- Holl KD, Quiros-Nietzen E (1999) The effect of rabbit herbivory on reforestation of abandoned pasture in southern Costa Rica. Biological Conservation 87: 391–395. https://doi. org/10.1016/S0006-3207(98)00051-2
- Horsley SB (1993) Mechanisms of interference between hay-scented fern and black cherry. Canadian Journal of Forest Research 23: 2059–2069. https://doi.org/10.1139/x93-257
- Hovenkamp PH, Miyamoto F (2005) A conspectus of the Native and Naturalized Species of *Nephrolepis* (Nephrolepidaceae) in the World. Blumea – Biodiversity, Evolution and Biogeography of Plants 50: 279–322. https://doi.org/10.3767/000651905X623003
- Ibarra-Manríquez G, Martínez-Morales M, Cornejo-Tenorio G (2015) Frutos y semillas del bosque tropical perennifolio: región de Los Tuxtlas, Veracruz. Conabio, México, 348 pp.
- Ibarra-Manríquez G, Sinaca Colín S (1995) Lista florística comentada de la Estación de Biología Tropical "Los Tuxtlas", Veracruz, México. Revista de Biología Tropical: 75–115.
- Jakobs G, Weber E, Edwards PJ (2004) Introduced plants of the invasive *Solidago gigantea* (Asteraceae) are larger and grow denser than conspecifics in the native range. Diversity and Distributions 10: 11–19. https://doi.org/10.1111/j.1472-4642.2004.00052.x
- Kuznetsova A, Brockhoff PB, Christensen RH (2017) ImerTest package: tests in linear mixed effects models. Journal of statistical software 82: 1–26. https://doi.org/10.18637/jss.v082.i13
- Keane R, Crawley MJ (2002) Exotic plant invasions and the enemy release hypothesis. Trends in Ecology & Evolution 17: 164–170. https://doi.org/10.1016/S0169-5347(02)02499-0
- Lamb D, Erskine PD, Parrotta JA (2005) Restoration of degraded tropical forest landscapes. Science 310: 1628–1632. https://doi.org/10.1126/science.1111773
- Laurance WF (2008) Theory meets reality: How habitat fragmentation research has transcended island biogeographic theory. Biological Conservation 141: 1731–1744. https://doi. org/10.1016/j.biocon.2008.05.011

- Longworth JB, Williamson GB (2018) Composition and Diversity of Woody Plants in Tree Plantations Versus Secondary Forests in Costa Rican Lowlands. Tropical Conservation Science 11: 1–13. https://doi.org/10.1177/1940082918773298
- Lüdecke D (2019) Sjstats: Collection of Convenient Functions for Common Statistical Computations. R package version 0.18.0. https://CRAN.R-project.org/package=sjstats
- Magnago LFS, Rocha MF, Meyer L, Martins SV, Meira-Neto JAA (2015) Microclimatic conditions at forest edges have significant impacts on vegetation structure in large Atlantic forest fragments. Biodiversity and Conservation 24: 2305–2318. https://doi.org/10.1007/ s10531-015-0961-1
- Marrs RH, Johnson SW, Duc MGL (2008) Control of bracken and restoration of heathland. VIII. The regeneration of the heathland community after 18 years of continued bracken control or 6 years of control followed by recovery. Journal of Applied Ecology 35: 857– 870. https://doi.org/10.1111/j.1365-2664.1998.tb00004.x
- Mendoza E, Fay J, Dirzo R (2005) A quantitative analysis of forest fragmentation in Los Tuxtlas, southeast Mexico: patterns and implications for conservation. Revista chilena de historia natural 78(3): 451–467. https://doi.org/10.4067/S0716-078X2005000300008
- Parrotta JA, Knowles OH (2001) Restoring tropical forests on lands mined for bauxite: Examples from the Brazilian Amazon. Ecological Engineering 17: 219–239. https://doi.org/10.1016/S0925-8574(00)00141-5
- Popoca-Ortega L (2016) Lluvia de semillas en parcelas de restauración ecológica en la selva tropical de Los Tuxtlas, Veracruz, México. Tesis de Licenciatura, Facultad de Ciencias Biológicas, Universidad Autónoma del Estado de Morelos. Cuernavaca, Morelos, México.
- Proctor GR (1989) Ferns of Puerto Rico and the Virgin Islands. Memoirs of the New York Botanical Garden 53: 1–389. https://doi.org/10.2307/1547180
- Rey Benayas JM, Bullock JM, Newton AC (2008) Creating woodland islets to reconcile ecological restoration, conservation, and agricultural land use. Frontiers in Ecology and the Environment 6: 329–336. https://doi.org/10.1890/070057
- Robinson RC, Sheffield E, Sharpe JM, Mehltreter K, Walker LR, Sharpe JM (2010) Problem ferns: their impact and management. Fern Ecology: 255–322. https://doi.org/10.1017/ CBO9780511844898.009
- Royo AA, Carson WP (2006) On the formation of dense understory layers in forests worldwide: consequences and implications for forest dynamics, biodiversity, and succession. Canadian Journal of Forest Research 36: 1345–1362. https://doi.org/10.1139/x06-025
- Sansevero JBB, Prieto PV, de Moraes LFD, Rodrigues PJP (2011) Natural Regeneration in Plantations of Native Trees in Lowland Brazilian Atlantic Forest: Community Structure, Diversity, and Dispersal Syndromes. Restoration Ecology 19: 379–389. https://doi. org/10.1111/j.1526-100X.2009.00556.x
- Sharpe JM, Shiels AB (2014) Understory fern community structure, growth and spore production responses to a large-scale hurricane experiment in a Puerto Rico rainforest. Forest Ecology and Management 332: 75–86. https://doi.org/10.1016/j. foreco.2014.01.023
- van Schaik CP, Terbourgh JW, Wright SJ (1993) The Phenology of Tropical Forests: Adaptive Significance and Consequences for Primary Consumers. Annual Review of Ecology and Systematics 24: 353–377. https://doi.org/10.1146/annurev.es.24.110193.002033

- Valencia-Esquivel I (2012) Efecto de plantaciones de restauración ecológica experimental en la caída de hojarasca en un pastizal tropical lluvioso en Los Tuxtlas, Veracruz. PHD Thesis. Universidad Autónoma del Estado de Morelos.
- Wunderle JM (1997) The role of animal seed dispersal in accelerating native forest regeneration on degraded tropical lands. Forest Ecology and Management 99: 223–235. https://doi. org/10.1016/S0378-1127(97)00208-9
- Young TP, Peffer E (2010) "Recalcitrant understory layers" revisited: arrested succession and the long life-spans of clonal mid-successional species. Canadian Journal of Forest Research 40: 1184–1188. https://doi.org/10.1139/X10-066

Supplementary material I

Model output for canopy cover characteristics by restoration treatment

Authors: Luis C. Beltrán, Karla María Aguilar-Dorantes, Henry F. Howe

Data type: Statistics table

- Explanation note: Mixed-effect model output for canopy openness (100 Canopy Cover %) and frequency of planted trees providing canopy cover by restoration treatment. Canopy openness model was built using a negative binomial distribution.
- Copyright notice: This dataset is made available under the Open Database License (http://opendatacommons.org/licenses/odbl/1.0/). The Open Database License (ODbL) is a license agreement intended to allow users to freely share, modify, and use this Dataset while maintaining this same freedom for others, provided that the original source and author(s) are credited.

Link: https://doi.org/10.3897/neobiota.59.51906.suppl1

Supplementary material 2

Trees by origin and restoration treatment

Authors: Luis C. Beltrán

Data type: Count

- Explanation note: Total number of trees (>2m) per treatment, including both, recruited and planted trees. This figure comes from a separate unpublished study carried out by Luis C. Beltrán.
- Copyright notice: This dataset is made available under the Open Database License (http://opendatacommons.org/licenses/odbl/1.0/). The Open Database License (ODbL) is a license agreement intended to allow users to freely share, modify, and use this Dataset while maintaining this same freedom for others, provided that the original source and author(s) are credited.

Supplementary material 3

Information on planted tree species

Authors: Luis C. Beltrán, Karla María Aguilar-Dorantes, Henry F. Howe

Data type: Descriptive

- Explanation note: Family, life history, dispersal mode, height range, and forest strata occupied by the planted species in the experimental restoration plantings. Height ranges from Ibarra-Manríquez et al. (2015) and Ibarra-Manríquez and Sinaca Colín (1995). Strata classified from published height ranges.
- Copyright notice: This dataset is made available under the Open Database License (http://opendatacommons.org/licenses/odbl/1.0/). The Open Database License (ODbL) is a license agreement intended to allow users to freely share, modify, and use this Dataset while maintaining this same freedom for others, provided that the original source and author(s) are credited.

Link: https://doi.org/10.3897/neobiota.59.51906.suppl3

Supplementary material 4

Fern data from quadrat study

Authors: Luis C. Beltrán, Karla María Aguilar-Dorantes, Henry F. Howe

Data type: count, density, percent, measurement, and species composition

- Explanation note: This table contains the data from the 384 quadrats where we determined: frond density, frond length, canopy cover, the identity of species covering the quadrat with their canopy, and the percent understory cover of grass or lianas.
- Copyright notice: This dataset is made available under the Open Database License (http://opendatacommons.org/licenses/odbl/1.0/). The Open Database License (ODbL) is a license agreement intended to allow users to freely share, modify, and use this Dataset while maintaining this same freedom for others, provided that the original source and author(s) are credited.

Supplementary material 5

Recruit data from quadrat study

Authors: Luis C. Beltrán, Karla María Aguilar-Dorantes, Henry F. Howe

Data type: occurrence

- Explanation note: This table includes the occurrence of each plant (>10 cm tall) that we found in our study's quadrats. The life history and dispersal mode of each species is presented.
- Copyright notice: This dataset is made available under the Open Database License (http://opendatacommons.org/licenses/odbl/1.0/). The Open Database License (ODbL) is a license agreement intended to allow users to freely share, modify, and use this Dataset while maintaining this same freedom for others, provided that the original source and author(s) are credited.

Link: https://doi.org/10.3897/neobiota.59.51906.suppl5

Supplementary material 6

Seed addition experiment data

Authors: Luis C. Beltrán, Karla María Aguilar-Dorantes, Henry F. Howe Data type: Count and measurements

- Explanation note: This table contains the results from the seed addition experiment. The shadehouse treatment corresponds to seeds sown in a shadehouse to confirm seed viability (not included in analysis). The columns for each month correspond to height measurements taken monthly. Plant biomass measurements are of dry mass (g).
- Copyright notice: This dataset is made available under the Open Database License (http://opendatacommons.org/licenses/odbl/1.0/). The Open Database License (ODbL) is a license agreement intended to allow users to freely share, modify, and use this Dataset while maintaining this same freedom for others, provided that the original source and author(s) are credited.

Supplementary material 7

Output from mixed-effects models on frond density

Authors: Luis C. Beltrán, Karla María Aguilar-Dorantes, Henry F. Howe Data type: Statistics table

- Explanation note: Summary from the mixed-effects models used in this study including: (1) Frond Density x Treatment, (2) Frond Length x Treatment, (3) Frond Density x Frond Length, and (4) Frond Density x Canopy Cover.
- Copyright notice: This dataset is made available under the Open Database License (http://opendatacommons.org/licenses/odbl/1.0/). The Open Database License (ODbL) is a license agreement intended to allow users to freely share, modify, and use this Dataset while maintaining this same freedom for others, provided that the original source and author(s) are credited.



CONTAIN: Optimising the long-term management of invasive alien species using adaptive management

Xavier Lambin¹, David Burslem¹, Paul Caplat², Thomas Cornulier¹, Gabriella Damasceno⁸, Laura Fasola¹⁰, Alessandra Fidelis⁸, Pablo García-Díaz¹, Bárbara Langdon⁵, Eirini Linardaki^{1,4,9}, Lía Montti^{3,6}, Jaime Moyano⁷, Martín A. Nuñez⁷, Stephen C.F. Palmer¹, Aníbal Pauchard⁵, Euan Phimister¹, José Cristóbal Pizarro¹², Priscila Powell³, Eduardo Raffo⁹, Ignacio A. Rodríguez-Jorquera⁴, Ignacio Roesler¹¹, Jorge A. Tomasevic⁴, Justin M.J. Travis¹, Claudio Verdugo⁴

I University of Aberdeen, Aberdeen AB24 2TZ, UK 2 Queens University Belfast, UK 3 Instituto de Ecología Regional (UNT-CONICET) Tucumán, Argentina 4 Centro de Humedales Río Cruces (CEHUM), Universidad Austral de Chile, Valdivia, Chile 5 Laboratorio de Invasiones Biológicas (LIB), Facultad de Ciencias Forestales, Universidad de Concepción, Chile 6 Instituto de Investigaciones Marinas y Costeras (IIMyC-CO-NICET), Instituto de Geología de Costas-CIC, Universidad Nacional de Mar del Plata, Argentina 7 Grupo de Ecología de Invasiones, INIBIOMA-UNComa, CONICET, Bariloche, Argentina 8 Lab of Vegetation Ecology, Instituto de Biociências, Universidade Estadual Paulista (UNESP), Rio Claro, Brazil 9 Servicio Agrícola y Ganadero (SAG), Los Ríos region, Chile 10 Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET)-Dirección Regional Patagonia Norte, Administración de Parques Nacionales. Aves Argentinas-Asociación Ornitológica del Plata, Argentina 11 Departamento de Ecología, Genética y Evolución & Instituto de Ecología, Genética y Evolución de Buenos Aires (IEGEBA-CONICET), Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires, Argentinas-Asociación Ornitológica del Plata, Argentina 12 Laboratorio de Estudios del Antropoceno (LEA), Facultad de Ciencias Forestales, Universidad de Concepción, Chile

Corresponding author: Xavier Lambin (x.lambin@abdn.ac.uk)

Academic editor: Ingo Kowarik | Received 17 March 2020 | Accepted 17 July 2020 | Published 5 August 2020

Citation: Lambin X, Burslem D, Caplat P, Cornulier T, Damasceno G, Fasola L, Fidelis A, García-Díaz P, Langdon B, Linardaki E, Moyano J, Montti L, Nuñez MA, Palmer SCF, Pauchard A, Phimister E, Pizarro JC, Powell P, Raffo E, Rodriguez-Jorquera IA, Roesler I, Tomasevic JA, Travis JMJ, Verdugo C (2020) CONTAIN: Optimising the long-term management of invasive alien species using adaptive management. NeoBiota 59: 119–138. https://doi.org/10.3897/neobiota.59.52022

Abstract

Invasive Alien Species (IAS) threaten biodiversity, ecosystem functions and services, modify landscapes and impose costs to national economies. Management efforts are underway globally to reduce these impacts, but little attention has been paid to optimising the use of the scarce available resources when IAS

Copyright Xavier Lambin et al. This is an open access article distributed under the terms of the Creative Commons Attribution License (CC BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

are impossible to eradicate, and therefore population reduction and containment of their advance are the only feasible solutions.

CONTAIN, a three-year multinational project involving partners from Argentina, Brazil, Chile and the UK, started in 2019. It develops and tests, via case study examples, a decision-making toolbox for managing different problematic IAS over large spatial extents. Given that vast areas are invaded, spatial prioritisation of management is necessary, often based on sparse data. In turn, these characteristics imply the need to make the best decisions possible under likely heavy uncertainty.

Our decision-support toolbox will integrate the following components:

- the relevant environmental, social, cultural, and economic impacts, including their spatial distribution;
- (ii) the spatio-temporal dynamics of the target IAS (focusing on dispersal and population recovery);
- (iii) the relationship between the abundance of the IAS and its impacts;
- (iv) economic methods to estimate both benefits and costs to inform the spatial prioritisation of costeffective interventions.

To ensure that our approach is relevant for different contexts in Latin America, we are working with model species having contrasting modes of dispersal, which have large environmental and/or economic impacts, and for which data already exist (invasive pines, privet, wasps, and American mink). We will also model plausible scenarios for data-poor pine and grass species, which impact local people in Argentina, Brazil and Chile. We seek the most effective strategic management actions supported by empirical data on the species' population dynamics and dispersal that underpin reinvasion, and on intervention costs in a spatial context. Our toolbox serves to identify key uncertainties driving the systems, and especially to highlight gaps where new data would most effectively reduce uncertainty on the best course of action. The problems we are tackling are complex, and we are embedding them in a process of co-operative adaptive management, so that both researchers and managers continually improve their effectiveness by confronting different models to data. Our project is also building research capacity in Latin America by sharing knowledge/ information between countries and disciplines (i.e., biological, social and economic), by training earlycareer researchers through research visits, through our continuous collaboration with other researchers and by training and engaging stakeholders via workshops. Finally, all these activities will establish an international network of researchers, managers and decision-makers. We expect that our lessons learned will be of use in other regions of the world where complex and inherently context-specific realities shape how societies deal with IAS.

Keywords

abundance impact relationship, adaptive management, biological invasions, dispersal, *Ligustrum lucidum*, models, *Neovison vison, Pinus contorta, Pinus radiata, Urochloa* spp, *Vespula germanica*

Introduction

Globally, invasive alien species (IAS) threaten biodiversity, ecosystem functions and services, modify landscapes and traditional livelihoods, and impose costs to national economies. The impacts of IAS are increasingly being documented worldwide (e.g. Simberloff et al. 2013, Haider et al. 2018, Taylor et al. 2019). Eradication efforts have been successful in island ecosystems or during the early stages of invasions, but

their results are very limited for most of the worst invaders in continental situations (Jones et al 2016). In contrast, population reduction and containment of existing populations are critical to reducing the cumulative impacts of those IAS that cannot be eradicated (Bomford and O'Brien 1995). This is because the prevention and detection have failed for numerous IAS; eradication is often not possible for IAS with cultural or economic importance, or those that are already widespread (Iriarte et al. 2005; Novillo and Ojeda 2008). A variety of frameworks have accordingly been developed to help manage these IAS, but the lack of available management tools remains a constraint on the ability of resource managers to develop long-term management plans (Larson et al. 2011). Responding to the challenge of established IAS requires the urgent development and implementation of evidence-based policy and decision-making systems for evaluating cost-effective strategies that limit their populations and reduce their impacts without eradication.

While there have been many advances in control techniques and management plans for species- and context-specific cases, multiple opportunities remain to advance in more holistic cross-taxa approaches. This is particularly pressing in developing regions, such as Latin America, where resources for management are extremely scarce, knowledge and data are limited, and the threat of IAS is large and increasing (Nuñez and Pauchard 2010). In these circumstances, managers must often shift their focus to a combination of slowing down or containing the invasion complemented with the long-term management of established populations in subsets of invaded areas where this is both practically feasible and cost-effective. For these strategies to be effective over their long operational horizons, there is a need to minimise reinvasion from existing populations, which in turn leads to recurring costs as the pressure from these populations remains constant over time. Local communities and management authorities bear those costs, and management must be optimised to minimise expenditure while retaining crucial ecosystem services and allowing the sustainable economic development of local communities.

The long-term management of IAS is plagued with uncertainties and complexities. While much of the invasion science literature contributes to the understanding of the mechanisms underpinning biological invasions, the exchange of resources and information on how to deal appropriately with established IAS in different social, economic, and ecological systems is substantially less advanced. Adaptive management or 'learning by doing' provides a suitable framework to create collaborative and interdisciplinary approaches focused on resolving the problems associated with IAS (Allen 2000, Johnson et al. 2015). Adaptive management involves specifying dynamic models of the case-study system, parameterized with empirical data and recursively updated by monitoring the response of the system to management interventions. This approach helps contend with uncertainties, incorporates existing knowledge in a quantitative fashion, considers the different dimensions of IAS management, and evaluates the adequacy of the interventions.

Our aim in this paper is to introduce an applied three-year multinational project and a group of more than 20 researchers and practitioners from Argentina, Brazil, Chile and the UK, which seeks to improve the strategic, long-term management of harmful IAS that cannot be eradicated and must be managed. Our work centers around four priority, data rich, and six secondary, relatively data-poor, problematic exemplar species established in Latin America. Our priority species are the glossy privet (*Ligustrum lucidum* – privet hereafter) in Argentina, pines in Chile and Argentina (*Pinus contorta* and *P. radiata* - pines hereafter), and the American mink (*Neovison vison* – mink hereafter) in Chile and Argentina. Secondary, relatively data-poor, species are the yellowjacket wasp in Chile (*Vespula germanica* – yellowjacket hereafter), invasive African grasses (*Urochloa decumbens, U. brizantha*), and the pines *P. elliottii* and *P. taeda* in Brazil and Argentina. The name of this project (CONTAIN) refers to the impossibility of IAS eradication and the need for containment of their population growth and impacts in the face of reinvasion.

Latin America, our study system

A colonial history and a heavy reliance on horticulture, industrialised agriculture, forestry, and aquaculture, resulted in the introduction of numerous alien species in Latin America. For example, 18 alien mammal species are present in Latin America (20%) of world mammalian species introduced), creating a hotspot of alien mammals in the southern temperate ecoregion of South America (Iriarte et al. 2005; Novillo and Ojeda 2008). Some of these alien species became damaging IAS, and now pose substantial problems. Some ecosystems are being fundamentally transformed by these IAS, as they dominate landscapes and drive major ecological processes. These include the novel communities of muskrat (Ondatra zibethicus), Canadian beaver (Castor canadensis), and mink now found in parts of Patagonia (Fasola and Valenzuela 2014). Notably, one recent study estimated that should Canadian beaver occupy all its suitable habitat in Tierra Del Fuego islands, it would result in >1 million tonnes of carbon being released to the atmosphere as a result of dam-building (Papier et al. 2019). Many plants with known invasive potential have also been introduced to support horticulture and forestry, as well as for ornamental purposes and pastures. Some of these species became invasive while others are naturalized and only now are beginning to spread across both semi-natural and human-disturbed ecosystems. These IAS include several Pinus species (specifically, P. radiata, P. contorta, P. elliottii), which have been described as representing a potential time-bomb, or invasion debt, based on their impacts elsewhere (Taylor et al. 2019). IAS from other taxa are also raising concerns such as the glossy privet, an evergreen tree dispersed by native birds. These species disrupt successional processes in forests in many parts of the world, threatening biodiversity and ecosystem services (Richardson and Rejmánek 2011). Besides impacting vegetation dynamics (Damasceno et al. 2018), African invasive grasses are changing fire regimes due to the increase in fuel load, leading to more intense and severe fires in tropical savannahs such as those covering large tracts of Brazil (Gorgone-Barbosa et al. 2015). Introduced carnivorous Vespula wasps notoriously restructure communities and alter resource flows, having a

detrimental impact on pollination services and the apicultural industry in many parts of the world, including Latin America (Lester and Beggs 2019).

Project approach

Following the tenets of adaptive management, we will develop and trial on the ground a decision-support toolbox to allocate management interventions in space and time effectively, based on conceptual and practical advances from IAS management practices in e.g. New Zealand, Australia, and the United States (e.g. Baker 2017). The key elements of this decision-support toolbox include:

- (i) the relevant environmental, social, cultural and economic impacts of IAS, including their spatial and temporal distribution;
- the spatio-temporal dynamics of the target species, with a focus on understanding and forecasting how dispersal and population recovery after management shape reinvasion and spread;
- (iii) the relationship between the abundance of the focal IAS and its relevant impacts in the focal areas;
- (iv) economic methods to estimate both the benefits and costs of interventions to spatially develop and rank prioritisation of cost-effective actions to manage interventions associated with IAS in space and time.

We seek to integrate the components described above in a mechanistic and streamlined fashion adapted to the idiosyncrasies and local contexts of our case studies in Latin America. To do so, we need to identify rules for selecting management strategies based on species' life histories, environmental goals, and socio-economic objectives. Indeed, planning durable IAS management requires determining the extent to which abundance of IAS should be reduced. Specifying what residual density is tolerable is a socio-ecological question involving consideration of the resilience of native species to IAS, the economic costs of IAS damage and the management costs required to achieve the residual density. Such costs typically rise exponentially as density decreases (Holmes et al. 2015). Furthermore, the ability of native species and ecosystem functions to be maintained in the presence of IAS is highly variable (Bradley et al. 2019). Different species and economic activities have different density-impact functions on native biota (Norbury et al. 2015). Some highly vulnerable species are devastated by even occasional incursions (e.g. predator-naïve flightless birds, Blackburn et al. 2004), while others can persist under low to moderate IAS density. The propensity of a small number of IAS individuals (e.g. small tree stands) to fuel the further spread of IAS also varies according to species traits, impinging on what residual density is manageable (Yokomizo et al. 2009). Thus, specifying management objectives in a spatial and temporal context is non-trivial, and indeed, it is often the case that such objectives are lacking or only vaguely articulated.

We identify the dispersal dynamics of IAS as critical to the success of management strategies, representing both a challenge and an opportunity. Whether mainly active (in animals) or passive (in plants), dispersal is notoriously subject to complex patterns of density and resource dependence. Propagule pressure after the naturalization stage may depend upon the age and stage structure of the source populations (Travis et al. 2011) and may be constrained by the permeability of the environment (Schurr et al. 2008). Yet, despite rapid advances in the understanding of dispersal biology, a knowledge gap exists regarding how to use this understanding to reduce the negative impacts of IAS on native species. A better understanding of IAS dispersal strategies may expose their vulnerabilities and incorporating such knowledge could improve management efficiency. For instance, depending on the target species, control may be more effective if performed in areas that act as sources (Baker 2017), or where active dispersers may be intercepted before they reach vulnerable areas (Caplat et al. 2014). Other species, such as the mink, may be best controlled if habitat selection by dispersers makes them settle reliably in high quality sites that are turned into ecological traps through targeted culling (Melero et al. 2018).

A final crucial issue to consider is that there is scant guidance for practitioners on how to allocate limited effort spatially, given that IAS spread through active or seed dispersal. This is necessary, as it has been recognized that there is spatial heterogeneity in IAS impacts within invaded landscapes (Latzka et al. 2016), and it is also known that impacts often increase exponentially with IAS density (Norbury et al. 2015; Bradley et al. 2019). Considering this explicitly is a key novelty of CONTAIN. Different management actions can target different stages of IAS spread, and the effects of management depend largely on the spatial configuration of the targeted area and on spatial aspects of spread. Source areas may produce propagules that spread through the landscape, fuelling reinvasion and further spread of the IAS. Their success depends on the behavior of dispersers, and the spatial and temporal variation in establishment success. The latter often co-varies with gradients in habitat quality and conspecific density, including those created by management. Thus, the redistribution of IAS in space in response to management actions may create 'halos' of decreased density spanning larger areas and delivering collateral benefits to local communities using natural resources in the vicinity of management action (Glen et al. 2013). Conversely, compensatory reproduction and dispersal (or increased establishment rate) may negate the impact of interventions according to the prevailing flux of dispersers or variation in the effects of land-use on establishment. Exploiting the potential predictability in the patterns of dispersal-driven reinvasion is particularly valuable in areas where access is limited or difficult and agencies would require substantial resources to tackle IAS. Therefore, this predictability can be harnessed to optimize management operations in challenging and uncertain circumstances.

Preserving and enhancing the livelihoods and biodiversity affected by the most damaging IAS in Latin America is likely to require recurrent management interventions extending in perpetuity. This challenge is ideally suited for adaptive management. Despite its success for achieving good outcomes, implementations of formal adaptive management approaches are scarce, owing to a lack of suitably trained staff able to operate in an interdisciplinary context at the interface between quantitative research and management (Williams et al. 2009). This deficit of human capacity in Latin America is critical, and even though there are increasing efforts to manage IAS, they do not necessarily contribute to an applied body of knowledge, and they sometimes lack a solid scientific foundation, which increases costs and reduces effectiveness. We present a program of research and reciprocal knowledge transfer representing a genuine, multi-country partnership using an adaptive management approach for tackling the challenges of securing biodiversity for sustainable livelihoods and economy and for maintaining and restoring natural capital in the face of IAS.

Methodology

Our project is organized around five work packages (WPs, Fig. 1). The sequence of the WPs fits the adaptive management and forecasting cycle, which presupposes:

- (i) development of an initial model of the system using whatever data are available,
- (ii) use of the model to identify key parameters and uncertainties,
- (iii) design and collection of data to address those uncertainties,
- (iv) specifying and exploring management scenarios (with associated costs and benefits).
- (v) WP5 crosses over other WPs and is designed to build social capital and capabilities of managers, NGOs, government agencies, and scientists to achieve a sustainable and positive change in the way IAS are managed in Latin America.

WP1. Specify generic models of IAS range reinvasion in response to management using RangeShifter

The goal of this WP is to develop and test a modelling platform to be applied as a decision tool for informing management efforts targeted at controlling IAS. We will incorporate key ecological mechanisms as well as costs of management, in order to test the effectiveness of alternative management options.

We will build upon the strong foundations provided by the RangeShifter software developed at the University of Aberdeen (Bocedi et al. 2014). RangeShifter is a flexible platform for modelling species' ecological and evolutionary dynamics across spatially complex landscapes. It applies an individual-based modelling approach with considerable flexibility for adapting to the biology of a user's focal species. RangeShifter has already been used for a broad range of applications, which consider how demography, landscape structure and dispersal behavior influence spread rates of plants and animals, including the case of mink in Scotland (Fraser et al. 2015). The software already contains two process-based models for simulating the transfer phase of dispersal for actively dispersing animals and two dispersal kernels for simulating plant seed disper-



Figure 1. Diagram describing links between different Work packages (WPs) of CONTAIN. IAS refer to Invasive Alien Species.

sal. We will add new modules to model 1) wind-dispersal of seeds, 2) animal-mediated dispersal of seeds, 3) land use/cover changes and 4) IAS control in space and time.

We will model the spatial dynamics of wind-dispersed invasive plants using the WALD model (e.g. Caplat et al. 2012), a simplified mechanistic dispersal model which retains the essential physics while being mathematically tractable (Katul et al. 2005). WALD parameters can be calibrated from the field or estimated from literature values, and the model has been shown to provide a close fit to empirical measurements of wind dispersal (see Caplat et al. 2012).

We will also model seed dispersal by animals using the 2Dt dispersal kernel (Clark et al. 1999), because this approach allows us to fit short and long dispersal distances, which is suitable for zoochorous seed dispersal (Herrera et al. 2011). The parameters for the 2Dt kernel for privet are already available for the Yungas ecoregion in Argentina (Powell and Aráoz 2018), which is one of the study systems for the CONTAIN project, including also environmental variability effects on dispersal (e.g. tree density).

We will develop a population management module in RangeShifter to simulate the removal of varying numbers of the focal IAS across space and time. It will provide a range of management strategies that differentially target specific ages/stages/sexes and spatial locations in an approach where management options can be compared according to their *effectiveness* at reducing the impacts of IAS (Caplat et al. 2014) and limiting the damage they cause. We will further expand this approach by incorporating the costs of management into the model, enabling management options to be ranked by *cost effectiveness*, which is highly relevant for managers.

WP 2. Specify models of range reinvasion after management for exemplar species.

We will apply our generic IAS model (WP1) to the exemplar species mentioned earlier and for which sufficient knowledge exists such that we can develop management models. We will parameterize the models with data from the literature, results from ongoing management interventions and new data when necessary. We will test model predictions against data collected in the field using an integrated approach.

To parameterize the models and kickstart an adaptive management programme, we will consider three exemplar plant species (privet, *P. contorta, and P. radiata*, Box 1), and we will derive stage-structured estimates for survival and fecundity as well as rate of seedling/sapling establishment. For privet, appropriate data exist on bird-mediated seed dispersal distances, gut passage time (needed for mechanistic dispersal model, Powell and Aráoz 2018), canopy height, land cover (Montti et al. 2017), and their potential impacts (Fernandez et al. 2017). To develop spatially realistic management scenarios, we will additionally use the outputs from WP1 to include spatially varying costs for realistic management options.

To start the adaptive management of privet, we will conduct experimental management of privet invasion sources in the subtropical montane forests (Yungas) in Northwestern Argentina. We will test different types of interventions (cutting privet individuals by mechanical and chemical methods against no actions) in the invasion front of privet adjoining the native forest. Additionally, we will measure the effects of active restoration of native vegetation, which includes planting saplings of native species. We will evaluate the effectiveness of these treatments by measuring privet individual and population recovery (re-sprout, survival, seed arrival and reestablishment) and natural regeneration (seed arrival, establishment, survival and growth of native trees) to find the most effective method to manage the invasion and restore native plant diversity in invaded forests. During two fruiting seasons, we will determine how the distance to privet seed sources and seedbank suppression affects seed arrival, germination, and sapling growth of privets as well as native trees.

For pines, our mechanistic models of effective dispersal will be based on the characteristics of the source population (propagule pressure, canopy height, distance, wind speed, and direction), and habitat characteristics for seedling recruitment (canopy cover, ground cover, and microclimate). We will build upon our extensive data on *P. contorta* invasion and management, and potentially re-survey permanent plots set up in Chile and Argentina to understand invasion trajectories, legacy effects, and the reinvasion after management (Pauchard et al. 2016). We will also estimate individual seed production for pines across different sites, for both plantations and stands arising from invasion. In addition, we will evaluate the impact of pine invasions on the productivity of native grasslands along a gradient of pine invasion density. Furthermore, we will conduct experimental management of *P. radiata*, the most widely planted pine species in Chile (c. 1.8 million ha). Realistic management (i.e. following current practices in the region) will be applied (i.e. mechanical removal or girdling) to plots stratified by habitat (native forests, grasslands, and shrublands), landscape context (distance to pine plantations, adjacent patches) and invasion stage (early or advanced). We will measure reinvasion from the seed bank and adjacent stands to estimate dispersal parameters. We will assess the economic costs and logistic constraints for all these treatments in order to parameterize the models.

The estimation of model parameters for mink will be based on studies in its native and European invaded ranges, including previous removal interventions (Melero et al. 2015, 2018; Oliver et al. 2016; Fasola and Roesler 2018). Our models will account for settlement probability relative to distance from natal site, habitat metrics and conspecific density (Melero et al. 2018), and the relationships between reproductive parameters and density (Melero et al. 2015). Our models will be tested against data from empirical work on mink focussing on ongoing management in highly contrasting regions of Chile and Argentina. We will estimate the effective dispersal of reinvading mink culled in Los Ríos, Chile, and in Buenos Aires Plateau, Argentina, using georeferenced archived tissue samples. Following genotyping, we will reconstruct likely pedigrees (Oliver et al. 2016), estimating dispersal as the distance between the location of capture and likely natal site. The outcome will be a quantification of the probability of successful dispersal from one area to another given variation in mink density caused by management activity and habitat productivity. Metrics of success will include the viability and population dynamics of affected endemic bird species as well as indicators of farmers' livelihoods. We will evaluate the former through field surveys and detailed monitoring data on aquatic bird abundance and breeding success that we will retrospectively link to model predictions and observations on residual mink density. Previous surveys of farmers, collected as part of a participatory management program in Los Rios region (SAG 2017), will be extended to assess the relationship between region-wide mink density and losses of poultry to mink predation.

WP3. Explore long-term management success criteria

Providing model-informed advice on management effectiveness requires clear criteria for success to have been determined. Such criteria are lacking. Here, we take an interdisciplinary participatory approach to determine the economic impacts of IAS and the costs associated with their removal at different spatial and organisational scales.

The species chosen as case studies (Boxes 1, 2) differ in the nature of the economic costs produced, their spatial distribution, and their impact across different societal

Box I. Exemplar plant species: **a** *Pinus contorta* invasion from a commercial plantation in the Patagonian steppe in Coyhaique Alto, Aysén Region, Chile. **b** *Turdus rufiventris* (a native bird in Yungas), eating *Ligustrum lucidum* fruits from an invaded forest in north western Argentina. **c** errado (tropical savannah) invaded by *Urochloa brizantha* evidencing the dominance of the invasive species.



Pinus contorta and *P. radiata* are our priority exemplar species of wind-dispersed IAS. Some pines are fast growing conifers highly suitable for forestry, which is why so many pine species have been widely introduced in Southern America, where there are no native pines. Pines are very successful at increasing their range due to a number of adaptations: a simple breeding system that allows selfing and cross-pollination by wind; high seed production from an early age; mechanisms for long-distance seed dispersal by wind; high seedling establishment in disturbed areas; and the ability to grow in a wide range of abiotic conditions (Richardson 1998). The invasion of pines generally originates from forestry plantations, which disperse most of their seeds to a distance of 100 m, but occasional long-distance dispersal events also occur (Caplat et al. 2012). This creates an invasion front that slowly advances from the plantation edge following the prevailing wind direction, but also many invasion islands far from the plantation that contribute to a faster range expansion. In general, very few abiotic or biotic conditions stop or even slow down a pine invasion, the lack of ectomycorrhizal fungi being one of the most important constrains to their spread (Nuñez et al. 2009). Because some pines are adapted to fire, this disturbance accelerates the invasion process and creates suitable habitats for seedling establishment in the absence of competition with native plants (Singh et al. 2018).

Pine invasions in Argentina, Chile and Brazil mainly affect treeless ecosystems, such as grasslands and shrublands (Richardson et al. 1994, Simberloff et al. 2010), with substantial impacts on the hydrological and nutrient cycles and modifications to the habitats of native species (Taylor et al. 2017, 2019). As a result, pastures invaded by pines in Argentina show a reduction in productivity affecting cattle grazing (Nuñez et al. 2017). In Chile, pine invasions increase the frequency and intensity of fires, with great risks for human settlements. The process of biotic homogenization, caused by replacement of native species during pine invasions, negatively affects the touristic scenery of Patagonia, which is one of the major sources of income in the region. In Argentina, the limited management efforts against pine invasions are restricted to protected areas carried out by the National Parks Administration, with no attention to most of the invaded areas in the country (Nuñez et al. 2017). In Chile, management efforts are focused on ecosystems that are critical for biodiversity conservation and largely driven by forestry certification standards (from the Forestry Stewardship Council) which foresters need to fulfil (Nuñez et al. 2017). Successful management experiences from New Zealand and South Africa show that pine invasions can be controlled using mechanical or chemical methods. Such interventions can be costly and do not guarantee the recovery of the native ecosystems, which may need to be actively restored (Nuñez et al. 2017).

African grasses are a threat to open ecosystems of South America, such as the Cerrado and the Llanos (Milton 2004). These species were intentionally planted for cattle pastures (Brossard and Barcelos 2005), but due to their physiological characteristics, such as high biomass and seed production, they can easily spread to protected areas (Pivello et al. 1999). The main species found in protected areas in Brazil are from the genus Urochloa (U. brizantha, U. decumbens, U. humidicola). Moreover, Melinis minutiflora, Megathyrsus maximum, Hyparrenia rufa and Andropogon gayanus are other major IAS commonly found in protected areas of Cerrado. Their presence in the natural systems leads to a decrease in the abundance and diversity of native species, mostly grasses (Damasceno et al. 2018), as well as an increase in dead biomass, leading and manual removal (Assis 2017; Damasceno & Fidelis submitted). Therefore, adaptive management can be an important tool at least to reduce the damage that these species impose on invaded areas (Damasceno et al. 2018).

The privet is a bird-dispersed Asian tree species invading ecosystems globally (Aragón and Groom 2003; Aslan 2011). In Argentina, this species was introduced and initially spread by people, and is used primarily for urban shade, amenity, living fences and windbreaks. After its introduction, the first steps of the expansion process involve a rapid and massive colonization of disturbed habitats near the introduction points, which are generally human settlements (Hoyos et al. 2010, Montti et al. 2017). After this first stage, seed sources become more abundant and widespread. Finally, once the species is well established, privet may form mono-specific stands that will dominate the entire tree community of the invaded habitat. The effects of privet on native forest are evident, modifying community species composition (Ayup et al. 2014) and ecosystem functions (such as nutrient turnover), and producing shifts in environmental conditions such as soil moisture and light availability (Zamora Nasca et al. 2014). Additionally, privet can slowly invade the native forests without human intervention (Malizia et al. 2017), because the species is dispersed not only by humans but also by native fruit-eating birds (Aragón and Groom 2003). In 2015, the National Strategy on Invasive Exotic Species from the Argentinian Government included privet as one of their eight most relevant IAS to test and promote management initiatives (Ministerio de Ambiente y Desarrollo Sustentable, 2016).

groups. The plant species have impacts on both direct and indirect economic values, including through sustainability certification schemes of exotic crops. On the other hand, the animal species have potential differential and more significant effects on small-holders' livelihoods and nature-based tourism activities. Our exemplar species also differ in terms of the speed and nature of their dispersal, which affect the risks and benefits associated with immediate action relative to a responsive approach (Epanchin-Niell 2017). This WP comprises four main activities.

Identify the main management and damage costs associated with each case study species, using secondary data and benefit transfer approaches (estimating economic values for ecosystem services by transferring available information from studies elsewhere, see e.g. https://sciencebase.usgs.gov/benefit-transfer), plus focused primary data collection where necessary.

Evaluate the cost-effectiveness of management measures and policies. The range of measures to be considered will be determined through discussions with relevant stake-holders e.g. policy makers, from each national context. The impact of timing on policy effectiveness will also be considered (Sims et al. 2016).

Evaluate the potential effectiveness of proposed management measures accounting for how people's individuals' behavior may affect IAS management, e.g. cropping and management decisions made by farmers.

Gain an understanding of the degree to which individuals and businesses are likely to account for the impact of their management activity on other activities and establish how government interventions may effect changes in behavior related to managing IAS to the levels that lead to improved societal benefits.

WP4. Co-develop adaptive management solutions

Having built capacity in WP3 and show-cased the approach with the work on our focal study species, we will explore strategic options for a wider set of IAS and seek crosstaxonomic generalities. To this effect, we will co-design management strategies with stakeholders and other researchers for emerging and potential future IAS for which data are currently sparse, but for which there are high societal demands for effective management strategies.

Our partners, researchers and practitioners from Latin America, will take a leading role in this work package. For instance, while we selected our exemplar IAS because of their importance in the partner countries, other species such as *Pinus elliottii* and *P. taeda* are transforming grasslands in northern Argentina and southern Brazil (Zenni and Simberloff 2013; Brandes et al. 2019, Durigan et al. 2007; Box 1). We will carry out workshops with foresters to explore plausible management scenarios that might be applicable for data-poor species, as well as future scenarios involving synergies between IAS. Some IAS, like the yellowjackets, affect local communities wherever they occur by damaging beekeeping, horticulture, and tourism (Magunacelaya et al. 1985, Box 2), yet range-wide containment would be prohibitively expensive. Participatory management as performed by poultry farmers suffering from mink predation in Los Ríos **Box 2.** Exemplar animal species: **a** Critically endangered Hooded grebes killed by a single American mink in Austral Patagonian highland plateau **b** Mink trapped in a raft deployed along rivers draining the plateau where Hooded grebe breed **c** Yellowjackets (*Vespula germanica*) feeding on a piece of meat.



Mink are our exemplar of a highly mobile and damaging mammalian predatory IAS. In Chile alone, the losses to invasive mink are estimated at US\$9.5 million per year, with US\$8.1millions corresponding to biodiversity losses and US\$1.4 millions allocated to control operations (UNDP 2017). Mink have become extremely abundant in southern Chile, fully invading Los Ríos, Los Lagos, Aysén and Magallanes regions, which comprise part of the Valdivian rainforest ecoregion, a recognized global biodiversity hotspot. The livelihood of autochthonous communities inhabiting coastal areas is severely affected due to loss of intertidal bio-resources (Ruiz et al. 1996). Since 2015, Ministry of Agriculture staff (partners in CONTAIN) have initiated a large participatory pilot mink control project in Los Ríos region focused on reducing the hardship to smallholder poultry farmers caused by mink predation and, as a by-product, preserving the wildlife resources on which ecotourism relies. More than a thousand farmers have removed >5,000 mink in the first 5 years, resulting in marked declines in predation of poultry. Surveys of participants demonstrate very high levels (>99%) of support for the continuation of the project (SAG 2017), but the long-term goals of management and over what scale they can be achieved sustainably have not yet been articulated. In Austral Patagonia, Argentina, mink threaten endemic upland bird species, including the critically endangered hooded grebe Podiceps gallardoi, with extinction (Roesler et al 2012; Fasola and Roesler 2018). Here, Aves Argentina (partners in CONTAIN) is leading a determined, high investment, mink trapping effort to avert extinction of the grebe, focused on the Buenos Aires Plateau (Fasola and Roesler 2016), an area of few rivers and a steep productivity gradient. Dispersing mink invade the upland nesting lakes of grebes seasonally, and because grebes are naïve to predation, single dispersers devastate local populations by surplus killing. A broad strategic goal for management is to push mink back to the lowlands, near the newly established Patagonia National Park, where control can be sustained at lower costs and from where dispersers would not threaten endangered endemic species.

The German wasp or German yellowjacket is native to Eurasia and northern Africa and has invaded e.g. New Zealand, Australia, South Africa, Chile, Argentina, United States of America and Canada (D'Adamo and Lozada 2009). Since its introduction in Chile in the 1970s, it has expanded its distribution range and negatively affected several economic activities, including agriculture and tourism (Estay et al. 2008). The aggressive nature of this IAS and its impact on rural economies has moved local communities to organize around the management of this species, even though eradication is no longer possible.

region (Chile) is one solution not yet widely used in Latin America. We will kick-start participatory adaptive management with farmers in rural communities already part of the participatory program Comunidad Humedal in Chile. This will serve to evaluate the effectiveness of fipronil meat-baiting by local residents (Sackmann et al. 2001) to control yellowjackets while at the same time gathering information on the scale of subsequent recolonization by yellowjackets. This will directly feed into the pioneering local IAS management plans recently launched by the municipality of Valdivia in Los Ríos region and now emulated by others.

WP5. Train in process-based modelling, contemporary ecological statistics, and economic valuation

We implement adaptive management approaches to counteract problematic IAS. Our work involves different researchers, policy and decision-makers, and stakeholders associated with diverse species and ecosystems problems. To facilitate this international cooperation, we have created an inclusive training program open to external researchers and practitioners from other Latin America organizations. It runs in parallel with research to foster a common approach based on the modern population-modelling tools underpinning adaptive management. This was augmented by a training workshop on economic valuation which is also central to developing effective adaptive management. Our training seeks to empower researchers to represent IAS within Bayesian integrated population models (IPMs) to estimate demographic and dispersal parameters for the RangeShifter decision tool. Where observational data are scant, as is the case with some problematic IAS, prior information may be used from related systems. The ability of IPMs to propagate sources of uncertainty arising from observation error, parameter uncertainty and process stochasticity will feed into the co-design of monitoring programs that are crucial to continuously improve IAS management effectiveness through adaptive management beyond the lifespan of the funded project.

Outlook

CONTAIN brings together researchers with not only diverse taxonomic focus (mammals, insects, trees, and grasses) but also contrasting research traditions, even when it comes to IAS. This reflects the prevailing research cultures in the participating countries and a dominance of diagnostic-focussed research over management. Latin American researchers include plant scientists working on dispersal, experts in IAS biology and researchers and practitioners already involved in the management of the mink as part of mixed academic or government initiatives. The UK team contributes experience in large-scale participative adaptive management of mink, but also plant ecology, agent-based modelling, statistical methods and rural economy expertise. The researchers appointed by the project deliver crucial expertise, including from New Zealand IAS management and from the forestry industry. None of the participating researchers has individually yet attained CONTAIN's aim to combine ecological, economic and sociological knowledge in a decision-support toolbox of broad applicability to the management of IAS that cannot be eradicated. The first few months of CONTAIN have initiated a common journey towards this aim, facilitated by meetings and exchanges, involving joint learning, a blending of research cultures and a common understanding of the benefits of transnational cross-taxa approach, considering the idiosyncratic aspects of the diverse socio-ecological contexts in which long term management ought to take place. A shared and contagious vision of the importance of evidence and understanding to guide management will no doubt be one output of CONTAIN.

A major challenge CONTAIN faces is that adaptive management requires a long period of time to show results, while both the required research and implementation funding to achieve that goal are short-term, typically three years (Mill et al. 2020). There are grounds for optimism, however. We found that awareness of the societal issues caused by IAS is high and the appetite for trial solutions is higher still. Indeed, following the launch workshop of CONTAIN in Chile, where the achievements of the Service for Agriculture and Livestock (SAG)'s evidence-led participatory management program for mink in the Los Ríos region and the international interest this elicited were highlighted to regional and national authorities, substantial progress has been achieved with funding and trans-regional (in Chile) and trans-national (between Chile and Argentina) management of mink. Thus, the CONTAIN project emerges as a real opportunity to not only improve our understanding of management of IAS, but also to improve the way government and communities deal with this important problem.

Acknowledgements

Project CONTAIN is funded under the Latin American Biodiversity Programme as part of the Newton Fund (NE/S011641/1), with contributions from NERC, the Argentine National Scientific & Technical Research Council (CONICET,-2019-74-APN-DIR#CONICET), the Brazilian São Paulo Research Foundation (FAPESP 2018/14995-8), the Chilean National Commission for Scientific & Technological Research (CONICYT). AP is supported by CONICYT PIA AFB170008. AF receives grant from CNPq (303988/2018-5), GD receives grant from FAPESP (2018/09054-0). The mink control program "Control Comunitario del Vison" is funded by the regional FNDR Funds, BIP 30484635-0, with the support of the regional government council. Yellowjacket wasp control receive support from municipality of Valdivia.

References

- Allen WJ (2000) Working together for environmental management: the role of information sharing and collaborative learning. PhD thesis. Massey University New Zealand.
- Aragón R, Groom M (2003) Invasion by *Ligustrum lucidum* (Oleaceae) in NW Argentina: early stage characteristics in different habitat types. Revista de Biología Tropical 51(1): 59–70.
- Ayup MM, Montti L, Aragón R, Grau HR (2014) Invasion of *Ligustrum lucidum* (Oleaceae) in the southern Yungas: Changes in habitat properties and decline in bird diversity. Acta oecologica 54: 72–81. https://doi.org/10.1016/j.actao.2013.03.006
- Aslan CE (2011) Implications of newly-formed seed-dispersal mutualisms between birds and introduced plants in northern California, USA. Biological Invasions 13: 2829–2845. https:// doi.org/10.1007/s10530-011-9966-1
- Baker CM (2017) Target the Source: Optimal Spatiotemporal Resource Allocation for Invasive Species Control. Conservation Letters 10: 41–48. https://doi.org/10.1111/conl.12236
- Blackburn TM, Cassey P, Duncan RP, Evans KL, Gaston KJ (2004) Avian Extinction and Mammalian Introductions on Oceanic Islands. Science 305: 1955–1958. https://doi. org/10.1126/science.1101617
- Bocedi G, Palmer SCF, Pe'er G, Heikkinen RK, Matsinos YG, Watts K, Travis JMJ (2014) RangeShifter: a platform for modelling spatial eco-evolutionary dynamics and species' responses to environmental changes. Methods in Ecology and Evolution 5: 388–396. https:// doi.org/10.1111/2041-210X.12162
- Bomford M, O'Brien P (1995) Eradication or control for vertebrate pests. Wildlife Society Bulletin 23: 249–255. https://www.jstor.org/stable/3782799

- Bradley BA, Laginhas BB, Whitlock R, Allen JM, Bates AE, Bernatchez G, Diez JM, Early R, Lenoir J, Vilà M, Sorte CJB (2019) Disentangling the abundance-impact relationship for invasive species. Proceedings of the National Academy of Sciences of the United States of America 116: 9919–9924. https://doi.org/10.1073/pnas.1818081116
- Brandes AFdN, Albuquerque RP, Domingues GdAF, Barros CF, Durigan G, Abreu RCR (2019) Dendroecology of *Pinus elliottii* Engelm. reveals waves of invasion in a neotropical savanna. Biological Invasions 22: 403–419. https://doi.org/10.1007/s10530-019-02099-2
- Brossard M, Barcellos ADO (2005) Conversão do Cerrado em pastagens cultivadas e funcionamento de latossolobos. Cadernos de Ciência e Tecnologia 22: 153–168.
- Caplat P, Nathan R, Buckley YM (2012) Seed terminal velocity, wind turbulence, and demography drive the spread of an invasive tree in an analytical model. Ecology 93: 368–377. https://doi.org/10.1890/11-0820.1
- Caplat P, Hui C, Maxwell BD, Peltzer DA (2014) Cross-scale management strategies for optimal control of trees invading from source plantations. Biological Invasions 16: 677–690. https://doi.org/10.1007/s10530-013-0608-7
- Clark, JS, Silman M, Kern R, Macklin E, HilleRisLambers J (1999) Seed dispersal near and far: patterns across temperate and tropical forests. Ecology 80(5): 1475–1494. https://doi. org/10.1890/0012-9658(1999)080[1475:SDNAFP]2.0.CO;2
- D'Adamo P, Lozada M (2009) Flexible foraging behavior in the invasive social wasp Vespula germanica (Hymenoptera: Vespidae). Annals Entomological Society of America 102 (6): 1109–1115. https://doi.org/10.1603/008.102.0620
- Damasceno, G; Fidelis, A. Submitted. Management of invasive grass species in a tropical savanna is dependent on regime disturbance and climatic conditions. Journal of Environmental Management.
- Damasceno G, Souza L, Pivello VR, Gorgone-Barbosa E, Giroldo PZ, Fidelis A (2018) Impact of invasive grasses on Cerrado under natural regeneration. Biological Invasions 20: 3621–3629. https://doi.org/10.1007/s10530-018-1800-6
- Durigan G, Siqueira MF de, Franco GADC (2007) Threats to the Cerrado remnants of the state of São Paulo, Brazil. Scientia Agricola 64: 355–363. https://doi.org/10.1590/S0103-90162007000400006
- Epanchin-Niell RS (2017) Economics of invasive species policy and management. Biological Invasions 19: 3333–3354. https://doi.org/10.1007/s10530-017-1406-4
- Estay P, Ripa R, Gerding M, Araya J, Curkovic T (2008) Manejo integrado de la avispa chaqueta amarilla, *Vespula germanica* Fabricius (Hymenoptera: Vespidae). Instituto de Investigaciones Agropecuarias, Boletín INIA 174, 74 pp.
- Fasola L, Valenzuela A (2014). Invasive Carnivores in Patagonia: defining priorities for their management using the American mink (*Neovison vison*) as study case. Ecología Austral, 24: 183–192.
- Fasola L, Roesler I (2016) Invasive predator control program in Austral Patagonia for endangered bird conservation. European Journal of Wildlife Research 62: 601–608. https://doi. org/10.1007/s10344-016-1032-y
- Fasola L, Roesler I (2018) A familiar face with a novel behaviour raises challenges for conservation: American mink in arid Patagonia and a critically endangered bird. Biological Conservation 218: 217–222. https://doi.org/10.1016/j.biocon.2017.12.031

- Fernandez RD, Bulacio N, Álvarez A, Pajot H, Aragón R (2017) Fungal decomposers of leaf litter from an invaded and native mountain forest of NW Argentina. Antonie van Leeuwenhoek 110(9): 1207–1218. https://doi.org/10.1007/s10482-017-0893-8
- Fraser EJ, Lambin X, Travis JMJ, Harrington L a., Palmer SCF, Bocedi G, Macdonald DW (2015) Range expansion of an invasive species through a heterogeneous landscape--the case of American mink in Scotland. Diversity and Distributions 21: 888–900. https://doi. org/10.1111/ddi.12303
- Glen AS, Pech RP, Byrom AE (2013) Connectivity and invasive species management: towards an integrated landscape approach. Biological Invasions 15: 2127–2138. https://doi. org/10.1007/s10530-013-0439-6
- Gorgone-Barbosa E, Pivello VR, Bautista S, Zupo T, Rissi MN, Fidelis A (2015) How can an invasive grass affect fire behavior in a tropical savanna? A community and individual plant level approach. Biological Invasions 17: 423–431. https://doi.org/10.1007/s10530-014-0740-z
- Haider S, Kueffer C, Bruelheide H, Seipel T, Alexander JM, Rew LJ, Arévalo JR, Cavieres LA, McDougall KL, Milbau A, Naylor BJ, Speziale K, Pauchard A (2018) Mountain roads and non-native species modify elevational patterns of plant diversity. Global Ecology and Biogeography 27: 667–678. https://doi.org/10.1111/geb.12727
- Herrera JM, Morales JM, García D (2011) Differential effects of fruit availability and habitat cover for frugivore-mediated seed dispersal in a heterogeneous landscape. Journal of Ecology 99(5): 1100–1107. https://doi.org/10.1111/j.1365-2745.2011.01861.x
- Holmes ND, Campbell KJ, Keitt BS, Griffiths R, Beek J, Donlan CJ, Broome KG (2015) Reporting costs for invasive vertebrate eradications. Biological Invasions 17: 2913–2925. https://doi.org/10.1007/s10530-015-0920-5
- Hoyos LE, Gavier-Pizarro GI, Kuemmerle T, Bucher EH, Radeloff VC, Tecco PA (2010) Invasion of glossy privet (*Ligustrum lucidum*) and native forest loss in the Sierras Chicas of Córdoba, Argentina. Biological invasions 12(9): 3261–3275. https://doi.org/10.1007/ s10530-010-9720-0
- Iriarte A, Lobos GA, Jaksic FM (2005) Invasive vertebrate species in Chile and their control and monitoring by governmental agencies. Revista Chilena de Historia Natural 78: 143– 151. https://doi.org/10.4067/S0716-078X2005000100010
- Johnson FA, Boomer GS, Williams BK, Nichols JD, Case DJ (2015) Multilevel Learning in the Adaptive Management of Waterfowl Harvests: 20 Years and Counting. Wildlife Society Bulletin 39: 9–19. https://doi.org/10.1002/wsb.518
- Jones HP, Holmes ND, Butchart SHM, Tershy BR, Kappes PJ, Corkery I, Aguirre-Muñoz A, Armstrong DP, Bonnaud E, Burbidge AA, Campbell K, Courchamp F, Cowan PE, Cuthbert RJ, Ebbert S, Genovesi P, Howald GR, Keitt BS, Kress SW, Miskelly CM, Oppel S, Poncet S, Rauzon MJ, Rocamora G, Russell JC, Samaniego-Herrera A, Seddon PJ, Spatz DR, Towns DR, Croll DA (2016) Invasive mammal eradication on islands results in substantial conservation gains. Proceedings of the National Academy of Sciences 113: 4033–4038. https://doi.org/10.1073/pnas.1521179113
- Katul GG, Porporato A, Nathan R, Siqueira M, Soons MB, Poggi D, Horn HS, Levin SA (2005) Mechanistic analytical models for long-distance seed dispersal by wind. American Naturalist 166: 368–381. https://doi.org/10.1086/432589

- Langdon B, Pauchard A, Aguayo M (2010) *Pinus contorta* invasion in the Chilean Patagonia: local patterns in a global context. Biological Invasions 12: 3961–3971. https://doi. org/10.1007/s10530-010-9817-5
- Larson DL, Phillips-Mao L, Quiram G, Sharpe L, Stark R, Sugita S, Weiler A (2011) A framework for sustainable invasive species management: Environmental, social, and economic objectives. Journal of Environmental Management 92: 14–22. https://doi.org/10.1016/j. jenvman.2010.08.025
- Latzka AW, Hansen GJA, Kornis M, Vander Zanden MJ (2016) Spatial heterogeneity in invasive species impacts at the landscape scale. Ecosphere 7: e01311. https://doi.org/10.1002/ecs2.1311
- Lester PJ, Beggs JR (2019) Invasion success and management strategies for social *Vespula* wasps. Annual Review of Entomology 64: 51–71. https://doi.org/10.1146/annurev-ento-011118-111812
- Magunacelaya JC, Chiappa E, Ojeda P (1985) Biología, problemas y control de la avispa Chaqueta amarilla. Documento técnico 2. Chile Forestal, Corporación Nacional Forestal, Ministerio de Agricultura, Chile.
- Malizia A, Osinaga-Acosta O, Powell PA, Aragón R (2017) Invasion of Ligustrum lucidum (Oleaceae) in subtropical secondary forests of NW Argentina: declining growth rates of abundant native tree species. Journal of Vegetation Science 28: 1240–1249. https://doi. org/10.1111/jvs.12572
- Melero Y, Robinson E, Lambin X (2015) Density-and age-dependent reproduction partially compensates culling efforts of invasive non-native American mink. Biological Invasions 17: 1–13. https://doi.org/10.1007/s10530-015-0902-7
- Melero Y, Cornulier T, Oliver MK, Lambin X (2018) Ecological traps for large-scale invasive species control: Predicting settling rules by recolonising American mink post-culling. Journal of Applied Ecology 55: 1769–1779. https://doi.org/10.1111/1365-2664.13115
- Mill AC, Crowley SL, Lambin X, McKinney C, Maggs G, Robertson P, Robinson NJ, Ward AI, Marzano M (2020) The challenges of long-term invasive mammal management: lessons from the UK. Mammal Review https://doi.org/10.1111/mam.12186
- Milton SJ (2004) Grasses as invasive alien plants in South Africa. South African Journal of Science 100: 69–75.
- Ministerio de Ambiente y Desarrollo Sustentable de Argentina (2016) Especies exóticas invasoras: Newsletter septiembre 2016. Available on: https://www.argentina.gob.ar/ambiente/ biodiversidad/especiesinvasoras/proyecto
- Montti L, Carrillo VP, Gutiérrez-Angonese J, Gasparri NI, Aragón R, Grau HR (2017) The role of bioclimatic features, landscape configuration and historical land use in the invasion of an Asian tree in subtropical Argentina. Landscape Ecology 32: 2167–2185. https://doi. org/10.1007/s10980-017-0563-2
- Norbury GL, Pech RP, Byrom AE, Innes J (2015) Density-impact functions for terrestrial vertebrate pests and indigenous biota: Guidelines for conservation managers. Biological Conservation 191: 409–420. https://doi.org/10.1016/j.biocon.2015.07.031
- Novillo A, Ojeda RA (2008) The exotic mammals of Argentina. Biological Invasions 10: 1333– 1344. https://doi.org/10.1007/s10530-007-9208-8
- Nuñez MA, Horton TR, Simberloff D (2009) Lack of belowground mutualisms hinders Pinaceae invasions. Ecology 90: 2352–2359. https://doi.org/10.1890/08-2139.1

- Nuñez MA, Pauchard A (2010) Biological invasions in developing and developed countries: does one model fit all? Biological Invasions 12: 707–714. https://doi.org/10.1007/s10530-009-9517-1
- Nuñez MA, Chiuffo MC, Torres A, Paul T, Dimarco RD, Raal P, Policelli N, Moyano J, García RA, van Wilgen BW, Pauchard A, Richardson DM (2017) Ecology and management of invasive Pinaceae around the world: progress and challenges. Biological Invasions 19: 3099–3120. https://doi.org/10.1007/s10530-017-1483-4
- Oliver MK, Piertney SB, Zalewski A, Melero Y, Lambin X (2016) The compensatory potential of increased immigration following intensive American mink population control is diluted by male-biased dispersal. Biological Invasions 18: 3047–3061. https://doi.org/10.1007/s10530-016-1199-x
- Papier CM, Poulos HM, Kusch A (2019) Invasive species and carbon flux: the case of invasive beavers (*Castor canadensis*) in riparian *Nothofagus* forests of Tierra del Fuego, Chile. Climatic Change 153: 219–234. https://doi.org/10.1007/s10584-019-02377-x
- Pauchard A, Escudero A, Garcia RA, de la Cruz M, Langdon B, Cavieres LA, Esquivel J (2016) Pine invasions in treeless environments: dispersal overruns microsite heterogeneity. Ecology and Evolution 6: 447–459. https://doi.org/10.1002/ece3.1877
- Pivello VR, Shida CN, Meirelles ST (1999) Alien grasses in Brazilian savannas: A threat to the biodiversity. Biodiversity and Conservation 8: 1281–1294. https://doi. org/10.1023/A:1008933305857
- Powell PA, Aráoz E (2018) Biological and environmental effects on fine-scale seed dispersal of an invasive tree in a secondary subtropical forest. Biological Invasions 20: 461–473. https://doi.org/10.1007/s10530-017-1548-4
- Richardson DM (1998) Ecology and Biogeography of Pinus. University Press, Cambridge, 470 pp.
- Richardson DM, Williams PA, Hobbs RJ (1994) Pine Invasions in the Southern Hemisphere: Determinants of Spread and Invadability. Journal of Biogeography 21: 511–527. https:// doi.org/10.2307/2845655
- Richardson DM and M Rejmanek (2011) Trees and shrubs as invasive alien species -a global review. Diversity and Distributions 17(5): 409–420. https://doi.org/10.1111/j.1472-4642.2011.00782.x
- Roesler I, Imberti S, Casanas H, Volpe N (2012) A new threat for the globally Endangered Hooded Grebe *Podiceps gallardoi*: The American mink *Neovison vison*. Bird Conservation International, 22(4): 383–388. https://doi.org/10.1017/S0959270912000019
- Ruiz J, Schlatter R, Bücher D (1996) Estudio de la situación del visón (*Mustela vison*, Schreber 1777) y su impacto sobre las comunidades autóctonas de la X Región, como aporte a la protección y recuperación de Áreas Silvestres Protegidas del Estado. Puerto Montt, Chile
- Sackmann P, Rabinovich M, Corley JC (2001) Successful Removal of German Yellowjackets (Hymenoptera: Vespidae) by Toxic Baiting. Journal of Economic Entomology 94: 811– 816. https://doi.org/10.1603/0022-0493-94.4.811
- SAG (2017) Programa Control Comunitario del Visón (Neovison vison). Valdivia, Chile.
- Schurr FM, Steinitz O, Nathan R (2008) Plant fecundity and seed dispersal in spatially heterogeneous environments: models, mechanisms and estimation. Journal of Ecology 96(4): 628–641. https://doi.org/10.1111/j.1365-2745.2008.01371.x

- Simberloff D, Nuñez MA, Ledgard NJ, Pauchard A, Richardson DM, Sarasola M, Van Wilgen BW, Zalba SM, Zenni RD, Bustamante R, Pena E, Ziller SR (2010) Spread and impact of introduced conifers in South America: Lessons from other southern hemisphere regions. Austral Ecology 35: 489–504. https://doi.org/10.1111/j.1442-9993.2009.02058.x
- Simberloff D, Martin JL, Genovesi P, Maris V, Wardle DA, Aronson J, Courchamp F, Galil B, García- Berthou E, Pascal M, Pyšek P, Sousa R, Tabacchi E, Vilà M (2013) Impacts of biological invasions: what's what and the way forward. Trends in Ecology and Evolution 28(1): 58–66. https://doi.org/10.1016/j.tree.2012.07.013
- Sims C, Finnoff D, Shogren JF (2016) Bioeconomics of invasive species: using real options theory to integrate ecology, economics, and risk management. Food Security 8: 61–70. https://doi.org/10.1007/s12571-015-0530-1
- Singh SP, Inderjit, Singh JS, Majumdar S, Moyano J, Nuñez MA, Richardson DM (2018) Insights on the persistence of pines (*Pinus* species) in the Late Cretaceous and their increasing dominance in the Anthropocene. Ecology and Evolution 8: 10345–10359. https://doi. org/10.1002/ece3.4499
- Taylor KT, Maxwell BD, McWethy DB, Pauchard A, Nuñez MA, Whitlock C (2017) Pinus contorta invasions increase wildfire fuel loads and may create a positive feedback with fire. Ecology 98: 678–687. https://doi.org/10.1002/ecy.1673
- Taylor KT, Callaway RM, Fajardo A, Pauchard A, Nuñez MA, Brooker RW, Maxwell BD, Dimarco RD, Peltzer DA, Mason B, Routsalainen S, McIntosh ACS, Pakeman RJ, Laney Smith A and M Gundale (2019) Severity of impacts of an introduced species corresponds with regional eco-evolutionary experience. Ecography 42: 12–22. https://doi.org/10.1111/ ecog.04014
- Travis JMJ, Harris CM, Park KJ, Bullock JM (2011) Improving prediction and management of range expansions by combining analytical and individual-based modelling approaches. Methods in Ecology and Evolution 2: 477–488. https://doi.org/10.1111/j.2041-210X.2011.00104.x
- UNDP (2017) Valoración económica del impacto de siete especies exóticas invasoras sobre los sectores productivos y la biodiversidad en Chile. Santiago de Chile, United Nations Development Programme.
- Williams BK, Szaro RC, Shapiro CD (2009) Adaptive management: The U.S. Department of the Interior technical guide. Adaptive Management Working Group, US Department of the Interior, Washington, D.C.
- Yokomizo H, Possingham HP, Thomas MB, Buckley YM (2009) Managing the impact of invasive species: the value of knowing the density–impact curve. Ecological Applications 19: 376–386. https://doi.org/10.1890/08-0442.1
- Zamora Nasca L, Montti L, Grau R, Paolini L (2014) Efectos de la invasión del ligustro, *Ligus-trum lucidum*, en la dinámica hídrica de las Yungas del noroeste Argentino. Bosque 35(2): 195–205. https://doi.org/10.4067/S0717-92002014000200007
- Zenni RD, Simberloff D (2013) Number of source populations as a potential driver of pine invasions in Brazil. Biological Invasions 15: 1623–1639. https://doi.org/10.1007/s10530-012-0397-4