

University of Massachusetts Amherst

ScholarWorks@UMass Amherst

Organismic and Evolutionary Biology Graduate
Student Publication Series

Organismic and Evolutionary Biology Graduate
Program

2023

Remember your roots: Biogeographic properties of plants' native habitats can inform invasive plant risk assessments

William Pfadenhauer

University of Massachusetts Amherst, wpfadenhauer@umass.edu

Michael Nelson

University of Massachusetts Amherst

Britt Laginhas

University of Massachusetts Amherst

Bethany Bradley

University of Massachusetts Amherst

Follow this and additional works at: https://scholarworks.umass.edu/oeb_grad_pubs



Part of the [Ecology and Evolutionary Biology Commons](#), and the [Plant Sciences Commons](#)

Recommended Citation

Pfadenhauer, William; Nelson, Michael; Laginhas, Britt; and Bradley, Bethany, "Remember your roots: Biogeographic properties of plants' native habitats can inform invasive plant risk assessments" (2023). *Diversity and Distributions*. 3.
<https://doi.org/10.1111/ddi.13639>

This Article is brought to you for free and open access by the Organismic and Evolutionary Biology Graduate Program at ScholarWorks@UMass Amherst. It has been accepted for inclusion in Organismic and Evolutionary Biology Graduate Student Publication Series by an authorized administrator of ScholarWorks@UMass Amherst. For more information, please contact scholarworks@library.umass.edu.

Remember your roots: Biogeographic properties of plants' native habitats can inform invasive plant risk assessments

William G. Pfadenhauer¹  | Michael F. Nelson² | Brit B. Laginhas^{2,3} |
Bethany A. Bradley^{1,2}

¹Organismic and Evolutionary Biology, University of Massachusetts Amherst, Amherst, Massachusetts, USA

²Environmental Conservation, University of Massachusetts Amherst, Amherst, Massachusetts, USA

³Center for Geospatial Analytics, North Carolina State University, Raleigh, North Carolina, USA

Correspondence

William G. Pfadenhauer, Organismic and Evolutionary Biology, University of Massachusetts Amherst, Amherst, MA, USA.

Email: wpfadenhauer@umass.edu

Funding information

Lotta Crabtree Trust; Massachusetts Agricultural Experiment Station; National Institute of Food and Agriculture; U.S. Geological Survey, Grant/Award Number: G19AC00091; University of Massachusetts Amherst Department of Environmental Conservation, Grant/Award Number: MAS00033

Editor: Marta Carboni

Abstract

Aim: Reducing the effects of invasive plants is best accomplished by predicting which species will invade and preventing their introduction. To do this, risk assessments rely on a variety of plant traits and biogeographic properties to predict potential invasiveness. However, the relative importance of these traits and properties is unknown. Determining which biogeographic properties contribute the most to predicting invasiveness could improve the accuracy and reduce the time needed to complete future risk assessments. Here, we provide a comprehensive analysis and ranking of the biogeographic properties that best differentiate invasive and noninvasive plant species.

Location: Conterminous United States.

Methods: We compiled county-level distributions of 10,721 vascular plant species native to the conterminous United States of which 884 were established elsewhere and 131 were invasive elsewhere. For each species, we used native distribution data to calculate biogeographic properties, including range size, human modification and abiotic niche breadth. We assessed the ability of biogeographic properties to predict whether each species was invasive outside of the United States using random forest classification models.

Results: Variables that represent the breadth of a species' native range, including the ranges of soil textures, ranges of soil fertility and total geographic area, are strong predictors of plant invasiveness. Models that included these variables correctly classified 86% of invasive species and 62% of noninvasive species. Variables representing resource availability and disturbance regime were not useful for distinguishing between established and invasive species.

Main conclusions: Focusing on niche breadth properties could improve the accuracy of risk assessments and reduce the effort spent compiling information with lower predictive power. The importance of niche breadth in this analysis supports previous findings that broad physiological tolerance enables survival and reproduction in numerous environments, thereby increasing the likelihood of invasion.

This is an open access article under the terms of the [Creative Commons Attribution](https://creativecommons.org/licenses/by/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2022 The Authors. *Diversity and Distributions* published by John Wiley & Sons Ltd.

KEYWORDS

biogeographic properties, invasive alien species, machine learning, screening tools, vascular plants, weed risk assessments

1 | INTRODUCTION

Increased globalization is linked to both the intentional and unintentional introductions of nonnative plants to novel environments (Hulme, 2012; Simberloff et al., 2012). An estimated 12% of these introduced plant species could become invasive (Hulme, 2012; Simberloff et al., 2012), resulting in negative impacts in their new environments (Mack et al., 2000; Pimentel et al., 2005; Simberloff, 2001; Vilà et al., 2011). As a result, there is a pressing need to identify species at high risk of becoming invasive and prevent their introduction. Current risk assessments identify potentially problematic species using a variety of plant characteristics, including phenotypic, ecological and biogeographic properties (e.g., Koop et al., 2012; Pheloung et al., 1999). While existing research has suggested that many biogeographic properties could predict invasion risk (Fridley & Sax, 2014; Gallagher et al., 2015; Higgins & Richardson, 2014), determining the significance and relative importance of these predictors remains largely unanswered. A comprehensive analysis of biogeographic properties as predictors of invasiveness is needed to inform risk assessments.

Eradicating invasive plants after they have been introduced is rarely successful (e.g., Mack & Lonsdale, 2002; Rejmánek & Pitcairn, 2002), whereas preventing invasive species introduction has large ecological and economic benefits (Keller et al., 2007; Leung et al., 2002). Despite the greater expense and lower efficacy of management and control postinvasion, policies in many places remain overwhelmingly reactive and lack mechanisms to prevent the introduction of high-risk species (Beaury et al., 2021; Early et al., 2016). An important step towards proactive management would be to identify high-risk species and prevent their introduction (e.g., Davies & Johnson, 2011; Davies & Sheley, 2007; Mack & Lonsdale, 2002; Simberloff, 2001). Weed risk assessments (WRAs) are designed to do just that and are increasingly used by different states, regions and countries (Heikkilä, 2011; Koop et al., 2012; Pheloung et al., 1999; Roy et al., 2018) to evaluate risk and inform decisions as to whether introduction (e.g., as ornamental plants or biofuels) should be prohibited. However, risk assessments tend to be inconsistent across borders (Roy et al., 2018) and can be extremely time-consuming depending on their level of detail (Koop et al., 2012; Verbrugge et al., 2010). Thus, identifying the best predictors of invasiveness would support more effective and efficient risk assessment.

We define biogeographic properties as any biotic or abiotic habitat characteristics associated with a species' native range. Biogeographic properties are ones that could be calculated using distribution data for a species in its native range. This definition inevitably includes some properties which could be classified as

ecological rather than biogeographic (e.g., species richness in the native range), but we use the term "biogeographic properties" here for brevity (Box 1).

Biogeographic properties are often included in existing risk assessments (Heikkilä et al., 2016; Koop et al., 2012; Ou et al., 2008; Panetta, 1993; Parker et al., 2007; Pheloung et al., 1999; Reichard & Hamilton, 1997). For example, "number of USDA cold hardiness zones suitable for survival," and "broad climate suitability" are included in WRAs by Koop et al. (2012) and Pheloung et al. (1999), respectively. The use of biogeographic properties in WRAs is supported by many recent studies, which have indicated that certain biogeographic properties and native habitats are associated with invasive plants (e.g., Fridley & Sax, 2014; Gallagher et al., 2015; Hejda et al., 2009; Higgins & Richardson, 2014; Hock et al., 2020; Weber et al., 2009).

BOX 1 Glossary of terms and abbreviations used in the present article

Biogeographic property	Any biotic or abiotic habitat characteristic associated with a species' native range
CONUS	Continental United States
Established	Plant species whose entire native ranges are within the Continental United States and that are surviving and reproducing in at least one location outside the Continental United States without causing negative effects
Invasion status	The classification of each plant species as either native, established or invasive; that is, a categorical response variable with three levels
Invasive	Plant species whose entire native ranges are within the Continental United States and that are surviving and reproducing in at least one location outside the Continental United States and that are also actively spreading and causing tangible negative effects in these new location(s)
Native	Plant species only found within the Continental United States
Nonnative	Established and invasive plants together; that is, plant species whose entire native ranges are within the Continental United States and that are surviving and reproducing in at least one location outside the Continental United States
WRA	Weed Risk Assessment

While many traits and properties are hypothesized to be related to invasiveness, it is unclear which ones are useful predictors that warrant inclusion in risk assessments. For example, several biological traits are often associated with invasive plants, including high specific leaf area, high fecundity, shorter reproductive cycles and fast growth (Baker, 1965; Barker & Francis, 2021; Kleunen et al., 2010; Pyšek & Richardson, 2007). These biological traits are often included in WRAs, but previous studies have shown that many are poor predictors and do not add value to risk assessments (Caley & Kuhnert, 2006; Conser et al., 2015; Pyšek & Richardson, 2007; Weber et al., 2009). Some of these studies also question the efficacy of biogeographic properties, emphasizing the need for a comprehensive evaluation.

Previously, a major challenge for assessing traits as predictors of invasiveness has been the difficulty in distinguishing between established species and invasive species. Established species have persistent populations in the introduced range, whereas invasive species are established and spreading (*sensu* Richardson et al., 2000). Fortunately, the Global Naturalized Alien Flora (GloNAF; van Kleunen et al., 2019) database now enables us to identify established plants globally. Three other recent databases - Global Plant Invaders (GPI; Laginhas & Bradley, 2021), Centre for Agriculture and Bioscience International Invasive Species, Compendium (CABI ISC; CAB International, 2021) and Global Invasive Species Database (GISD; International Union for Conservation of Nature [IUCN], 2021)—also allow us to identify invasive plants globally.

We identified five main ways that biogeographic properties of the native range could affect the likelihood that a plant will establish elsewhere and/or become invasive elsewhere (Figure 1):

- Human Contact:** Since humans are the most important vectors for invasive species relocation (Hulme, 2009; Pouteau et al., 2021), any biogeographic property that represents the intensity of human presence in a given environment (e.g., the amount of developed land or the density of airports) could act

as a proxy for the likelihood of transportation outside their native range.

- Niche Breadth:** Existing literature has shown that invasive species have broad abiotic niches. A broad abiotic niche suggests that invasive species have large physiological tolerances, which in turn allows for survival and colonization in numerous other environments (Gallagher et al., 2015; Higgins & Richardson, 2014). Breadth of abiotic conditions in the native range can be described by range size as well as range of temperature, precipitation and soil textures.
- Historical Disturbance:** Disturbance regime could predict plant invasiveness because species that have evolved with high disturbance frequency could capitalize on disturbances (both natural and anthropogenic) in nonnative ranges, and invasive plants are well known to be disturbance responsive (Bhattarai & Cronin, 2014; Brooks & Pyke, 2001; Hobbs & Huenneke, 1992).
- Resource Availability:** Species originating in areas with plentiful resources (such as available water content or soil organic carbon) typically have more enemies (herbivores and pathogens) and may be more likely to experience enemy release in the nonnative range (the resource-enemy release hypothesis; Blumenthal, 2006; Blumenthal et al., 2009). Alternatively, invasive species from high-resource environments tend to be strong competitors, leading them to outcompete native species when resources are plentiful (Daehler, 2003; Davis et al., 2000).
- Native Plant Richness:** High species richness in a plant's native range could lead to "fitter" species due to a longer evolutionary history which might yield a competitive advantage elsewhere (the evolutionary imbalance hypothesis; Fridley & Sax, 2014).

Here, we assess the predictive contributions of biogeographic properties within WRAs. Specifically, our analyses address the following five hypotheses: (1) species whose native ranges include areas with extensive human contact are more likely to become established elsewhere, (2) species whose native ranges include a broad set of abiotic conditions are more likely to be invasive elsewhere, (3) species whose native ranges are prone to natural disturbance are

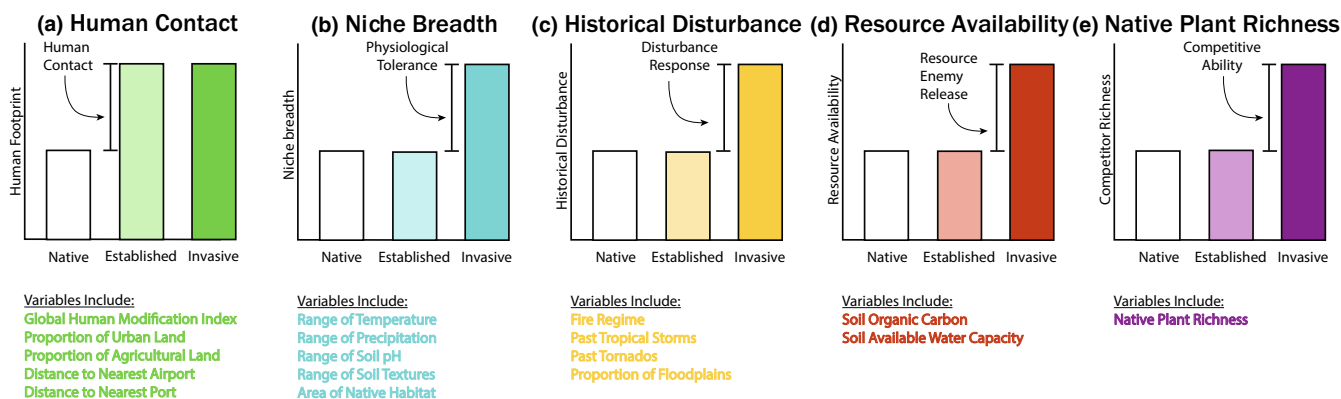


FIGURE 1 Hypotheses for the five main variable groups. Uncategorized variables (EPA Ecoregions and growth habits) are not associated with a single hypothesis.

more likely to be *invasive* elsewhere, (4) species whose native ranges are resource-rich are more likely to be *invasive* elsewhere, (5) species whose native ranges include high plant diversity are more likely to be *invasive* elsewhere (Figure 1).

2 | METHODS

2.1 | Target species and study area

For this study, we included plant species native to only the conterminous United States (CONUS). Our initial list of candidate species consisted of all documented vascular plant species (excluding subspecies, hybrids and varieties) in the USDA PLANTS database (Natural Resource Conservation Service [NRCS], 2021) classified as “L48(N),” that is, native to the CONUS. Our spatial variables were constrained by the borders of the CONUS, so any plants whose native ranges extended into other countries were excluded. For example, white spruce (*Picea glauca*) is native in many northern areas of the CONUS, but it is also native to every Canadian province. Therefore, calculating biogeographic properties within only its U.S. range could mischaracterize the species' native range.

We classified the remaining plants based on the invasion status in regions outside of CONUS (Box 1):

1. Species only found in the CONUS (hereafter “native species”; $n = 9706$)
2. Species established outside of the CONUS but not identified as *invasive* in any location (hereafter “established species”; $n = 884$), and
3. Species which were *invasive* in at least one location outside of the CONUS (hereafter “*invasive* species”; $n = 131$).

For models that evaluated differences between native and non-native species, we combined established and *invasive* species into a single category (hereafter “nonnative species,” $n = 1,015$). We used the GloNAF database to differentiate between native and established plants (van Kleunen et al., 2019) and we used GPI, CABI and GISD to differentiate between established and *invasive* plants (CAB International, 2021; IUCN, 2021; Laginhas & Bradley, 2021). Species listed as *invasive* in any location by the GPI, CABI or GISD databases were classified as *invasive*. Species not listed in any of the *invasive* databases, but listed as established in any location outside the CONUS by GloNAF were classified as established. All other species were classified as native.

We considered each species' native range to consist of all counties where the plant was listed as native in the PLANTS database (NRCS, 2021) as of June 2021. At the time of data collection, the PLANTS database contained species presence/absence data for all CONUS counties except those within Louisiana, Oglala Lakota (formerly Shannon) County in South Dakota, Cherokee County in Alabama, Broomfield County in Colorado and Juneau County in Wisconsin.

2.2 | Biogeographic variables

To test our hypotheses, we identified 18 continuous predictor variables that were representative of the five distinct biogeographic processes hypothesized above (see Table 1 for a complete list of data sources and variables). To characterize the biogeography of each species' native range, we calculated the mean, range or sum of each variable (as appropriate), aggregated over the species' native counties. For example, we calculated the mean values of urban land cover, agricultural land cover, distance to airports, distance to seaports and Global Human Modification Index (Theobald et al., 2020) as representative of human contact in each species' native range (Table 1).

To account for possible interactions between biogeographic variables and plant trait variables, we also included the presence of each species within each of the 85 EPA Level III Ecoregions (revised in 2013) and which of the six USDA-designated growth habits (tree, shrub, subshrub, graminoid, forb/herb and vine) was associated with each species. Ecoregions are areas with similar suites of environmental variables, and a species' growth habit is the result of many smaller biologic and phenotypic traits. We did not associate these additional 91 categorical variables with any one of our five hypotheses, since they likely encompass elements of multiple processes.

2.3 | Significance testing

To test for significant differences (p -value $\leq .05$) between native, established and *invasive* plants, we used the nonparametric Kruskal–Wallis and post hoc pairwise Wilcoxon rank-sum tests to compare the means of each continuous predictor variable for native, established and *invasive* plants (Table 2). We used nonparametric tests instead of ANOVA and Tukey post hoc tests because many variables had heterogeneity of variance and nonnormal residuals, which means they violated two important assumptions of ANOVAs. While Kruskal–Wallis and Wilcoxon tests can help exclude variables that are not good predictors of invasion status, significant differences between invasion statuses that result from these tests do not necessarily indicate that the variable has strong predictive capability. Therefore, we used classification models to identify the variables that made the best predictors of establishment and/or invasion.

2.4 | Classification models

We used random forests, support vector machines and logistic models to evaluate the predictive power of biogeographic properties. The three model types had similar results, so we report the methods and results of the random forest models in the main text and include details for the other models in Appendix S1.

TABLE 1 Independent variables and data sources grouped by hypothesized processes

Hypothesized process	Predictor variable (data type)	Calculation	Range of values	Data source
A. Human contact	Urban land use (continuous) Includes NLCD classes 21, 22, 23, and 24	Mean	0.2%–81.5%	Homer et al. (2004)
	Agricultural land use (continuous) Includes NLCD classes 81 and 82	Mean	0%–86.6%	
	Airport proximity (continuous)	Mean	13.2–196 km	United States Geological Survey (1999)
	Port proximity (continuous)	Mean	5.12–1030 km	Waterborne Commerce Statistics Center (2013)
	Global Human Modification Index value^a (continuous)	Mean	0.008–0.836	Theobald et al. (2020)
B. Niche breadth	Precipitation (continuous)	Range	32.7–5937 mm/yr ^b	PRISM Climate Group (2015b)
	Temperature^a (continuous)	Range	10.8–43.3°C ^b	PRISM Climate Group (2015a)
	Native range size (continuous)	Sum	415–4,240,000 km ²	NRCS (2021)
	Soil pH (continuous)	Range	0.300–7.46 pH ^b	Miller and White (1998)
	Clay content (continuous)	Range	3.1%–57.7% ^b	
	Silt content^a (continuous)	Range	3.30%–76.2% ^b	
	Sand content^a (continuous)	Range	3.8%–90.3% ^b	
C. Historical disturbance	Tropical storm paths^a (continuous)	Mean	0–5.5 × 10 ⁻⁰⁸ #/km ²	Knapp et al. (2010, 2018)
	Fire regime (categorical) Either Fire Regime Group I, II, III, IV, or V	Mode	-	LANDFIRE (2016)
	Tornado touchdowns^a (continuous)	Mean	0–5.3 × 10 ⁻⁰⁸ #/km ²	National Weather Service (2009)
	Floodplains (continuous)	Mean	0.7%–99.5%	Woznicki et al. (2019)
D. Resource availability	Soil organic carbon (continuous)	Mean	1.20–10.5 kg SOC/m ²	Guevara et al. (2020)
	Soil available water capacity (continuous)	Mean	3.01%–22.2%	Miller and White (1998)
E. Native plant richness	Native plant species richness (continuous)	Mean	2.5 × 10 ⁻⁰⁸ – 1.5 × 10 ⁻⁰⁶ species/km ²	NRCS (2021)
F. Uncategorized	EPA Level III Ecoregions (categorical)	Presence	-	Omernik and Griffith (2014)
	Growth habit (categorical) Either tree, vine, shrub, subshrub, forb/herb, or graminoid	Y/N	-	NRCS (2021)

Note: Mean measurements are the mean value across each species' county distribution; range measurements are the differences between the maximum and minimum values across each species' county distribution.

^aVariables that were excluded from models due to high correlation.

^bValues are the ranges of variation within a single variable (the result of calculating the range for values that are already ranges). For example, the species with the least variation in soil pH was *Orbexilum macrophyllum*. For this species, the maximum and minimum soil pH values differed by 0.3. This species did not inhabit areas with soil pH values equal to 0.3.

2.4.1 | Random forests

Random forests are a form of supervised classification which use collections of single decision trees, which generate “branches” based on the predictor variables that have the largest variation in respect to the outcome classifications (Ho, 1995). Random forest predictions

are based on the most common, or consensus, outcomes from the individual trees, which reduces error through the aggregation of many individual predictions.

Prior to running our random forest models, we excluded all highly correlated (Pearson >0.75) predictor variables. These variables were as follows: tornado and tropical storm frequency, temperature range,

TABLE 2 *p*-Values from pairwise Wilcoxon tests for all variables included in random forest models

Predictor variable	Native-established	Established-invasive	Native-invasive
Urban land use	<2e-16*	0.031*	4.40e-8*
Agricultural land use	<2e-16*	0.088	9.00e-16*
Airport proximity	<2e-16*	0.032*	0.037*
Port proximity	4.20e-16*	0.0091*	0.4605
Precipitation range	1.10e-11*	1.80e-6*	2.30e-11*
Soil pH range	0.0024*	2.60e-8*	4.60e-10*
Clay range	<2e-16*	7.70e-10*	<2e-16*
Native range size	<2e-16*	2.80e-7*	<2e-16*
Floodplains	<2e-16*	0.33	5.20e-15*
Soil organic carbon	<2e-16*	0.089	0.058
Soil available water capacity	<2e-16*	0.14	3.60e-14*
Native plant richness	<2e-16*	0.016*	4.80e-7*

Note: Significant values (<0.05) are marked with an asterisk. These values correspond to the letters (a, b, c) in Figure 2.

TABLE 3 Classification models with resulting kappa and overall accuracy values

Classification type	Mean kappa	Model type	Mean accuracy
Native versus nonnative	0.499 ± 0.08	Random Forest	0.78 ± 0.03
Established versus invasive	0.302 ± 0.10	Random Forest	0.64 ± 0.05
Native versus established versus invasive	0.396 ± 0.05	Random Forest	0.61 ± 0.03

Note: The kappa and accuracy values are averaged across all 500 iterations of each model. Means are followed by standard deviations.

GHMI and plant presence in Ecoregions 1, 4, 11, 12, 22, 25, 27, 30, 32, 36–39, 51, 52, 55, 61, 65–67, 69, 71 and 73. Even though sand and silt content did not exceed correlations of 0.75, we excluded these variables as well because they are inherently influenced by the value of clay content. The remaining continuous predictor variables were centred and scaled using the “scale” function in R (R Core Team, 2020).

To prepare our response variable (invasion status) for the random forests, we created three different classification schemes. First, we converted invasion status into a binary response variable with two categories: native (base case) and nonnative species. Next, we kept invasion status as a binary variable, but this time we used established (base case) and invasive as the two categories. Finally, we treated invasion status as an ordinal response variable with three categories: native (base case), established and invasive.

To train and test each iteration within each random forest model, we conducted two rounds of stratified random sampling. First, to reduce the likelihood that our largest category (native species) would bias the predictions, we selected 131 species from each invasion status (native, established and invasive). This number (131) of species was equivalent to the size of our smallest category (invasive). Next, we further subsampled, without replacement, each group of 131 species, randomly selecting two-thirds of the species (86 species for each invasion status) to train each model. We then used the remaining 45 nonsampled species from each of the categories to test the accuracy of the resulting invasion status predictions. We conducted 500 iterations of this process

(sampling, then subsampling, then model training and then testing) for each of the three classification schemes (see Table 3 for model details). We fit the 1500 total random forest models in R using the package “randomForest” (R Core Team, 2020). We ran 500 decision trees in each of the random forests and tried 10 variables at each split.

To compile results, we averaged Cohen's kappa, overall accuracy and mean decrease in accuracy across the testing groups for all 500 iterations within each classification scheme. Cohen's kappa was used to quantify the consistency with which species were classified into their true invasion statuses. Overall accuracy was calculated as the percent of species whose true invasion statuses were correctly predicted by each model. Mean decrease in accuracy was used to determine variable importance rankings. We also compiled a list of established species that were misclassified as invasive more often than correctly classified as established (Appendix S2) by the model that compared native versus established versus invasive species (Table 3, row 3).

3 | RESULTS

3.1 | Dataset characteristics

Our dataset encompassed a total of 10,721 species. A majority of species (7846) were forbs/herbs, 1467 were shrubs, 945 were graminoids, 250 were subshrubs, 161 were trees, and 52 were vines

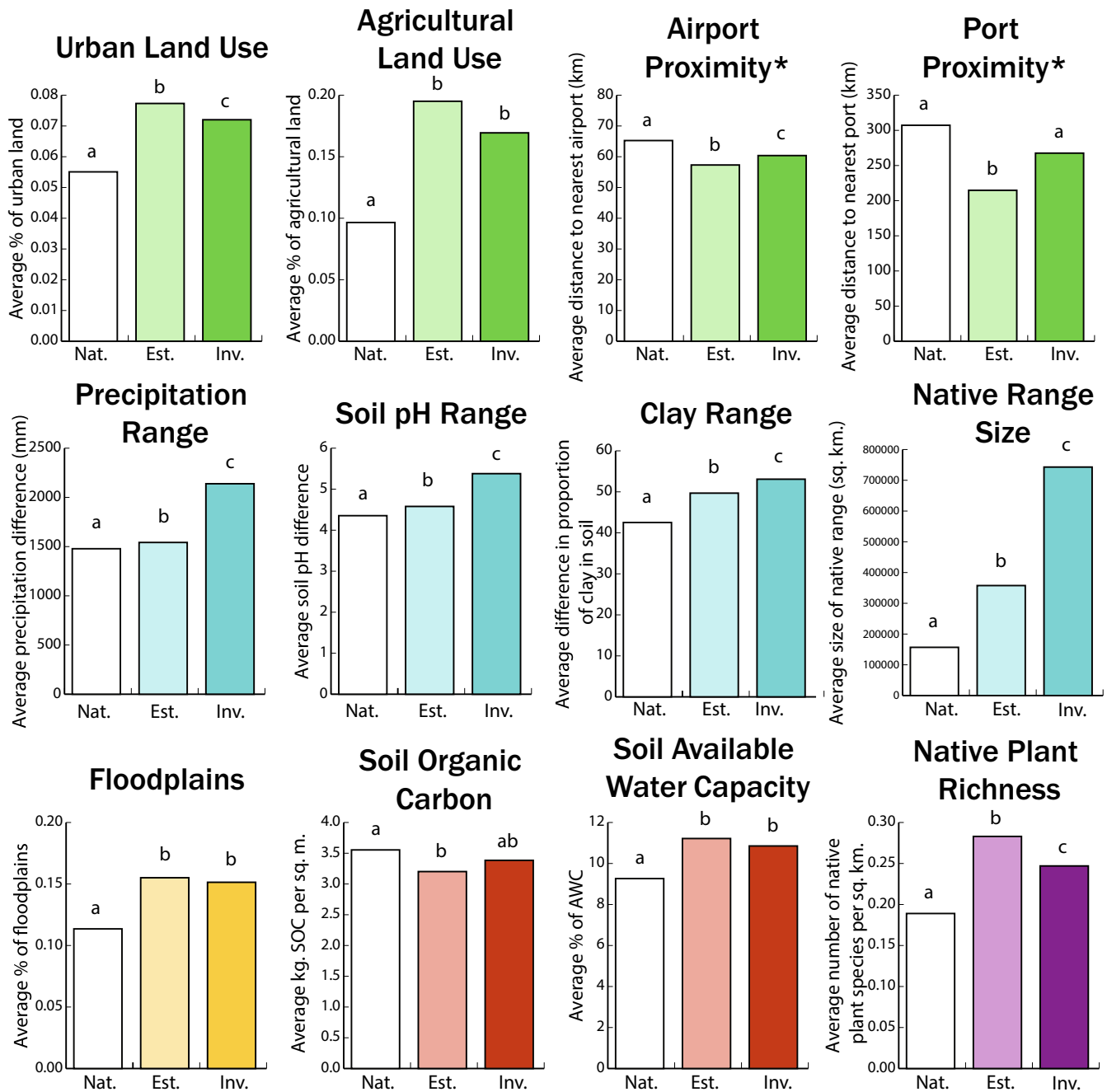


FIGURE 2 Native and established species differed significantly for all continuous predictor variables used in random forest models, but established and invasive mainly differed in niche breadth variables. Pairwise Wilcoxon rank-sum test results for continuous variables, as reported in Table 2. Letters (a, b, c) indicated significant differences between groups with a p -value $\leq .05$. *Airport and port proximity results appear inverted in comparison with the human contact hypothesis because larger distances on the y-axes represent less human contact.

(although none of the growth habits were good predictors of invasiveness). Plant presence was reported for 2981 counties, and each species was found within a mean of 46 counties, with a median of 12 counties per species. The mean county size was 2561.43 square kilometres. A total of 244 established species were misclassified as invasive more often than they were correctly classified as established in the model that compared all three invasion statuses (Appendix S2).

3.2 | How are nonnative species different from native species?

Significance testing (using Kruskal–Wallis and Wilcoxon tests) revealed differences between established and native species for all continuous predictor variables except for breadth of temperature in the native range. These results matched our hypothesis for human contact variables; we predicted that established species would be subjected

to more human contact in their native range than native species, and this appeared to be true. Results for the other predictor variable categories largely diverged from our hypotheses; native and established species differ in the niche breadth, historical disturbances, resource availability and native plant richness of their native ranges (Figure 2). Categorical predictor variables also showed some differences between native and nonnative species. For example, the largest relative proportion of nonnative species was determined to have originated in ecoregion 49 (Northern Minnesota Wetlands; 77% nonnative), while the largest relative proportion of native species was determined to have originated in ecoregion 16 (Idaho Batholith; 93% native).

3.3 | Which differences between native and nonnative species are most useful for prediction?

Niche breadth variables appeared to be the best predictors of “non-nativeness,” but variables related to all hypotheses had some importance. Random forests which classified species as either native or nonnative performed the best out of the three types of random forests (Table 3). These models had an average overall accuracy of 0.78 and average overall kappa of 0.49. Variable importance rankings from these models suggested that two niche breadth variables are the best predictors of nonnativeness: the range of the proportion of clay found within a species' native range and the geographic area of a species' native range (Figure 3). The proportion of floodplains within the native range was ranked third, followed by urban land use, available water capacity and native plant richness in the native range.

3.4 | How are invasive species different from established species?

Kruskal–Wallis and Wilcoxon tests showed that invasive and established species had significant differences for most (13/18) continuous predictor variables, including: all but one variable (agricultural land use) associated with human contact, all niche breadth variables, tornado touchdowns and native plant richness. Thus, our hypotheses of no difference between established and invasive species for human contact variables were not supported by our results. We hypothesized differences between established and invasive species for niche breadth and native plant richness, which aligned with the results of the significance testing. For categorical predictor variables, a higher proportion of invasive species originated in EPA ecoregion 49 as compared to the other ecoregions.

3.5 | Which differences between established and invasive species are most useful for prediction?

The range of clay proportions, the range of soil pH values and the geographic area of a species' native range were the three most

important variables in the random forest model that differentiated established species from invasive species. These results aligned with the other two random forest models; clay proportions, soil pH and native range size were ranked within the top eight in all models. Aside from these three niche breadth variables, the only other variable (out of the 85 total variables considered) which ranked within the top 10% for all random forest models was agricultural land use in the native range. Ecoregions 24 (Chihuahuan Desert), 84 (Atlantic Coastal Pine Barrens) and 50 (Northern Lakes and Forests) were ranked fifth, sixth and eighth, respectively.

3.6 | Which differences are most useful for prediction across native, established and invasive species?

Similarly, niche breadth variables were also the most important predictors in the model that compared all three invasion statuses (native vs. established vs. invasive). The range of clay proportions in the native range and the size of the native range once again were ranked first and second, respectively. Following native range size, there was a noticeable drop in mean decrease in accuracy before the next highest variables, which were ecoregion 59 (Northeastern Coastal Zone), available water content and proportion of floodplains (Figure 3a). Overall, niche breadth variables were consistently ranked as the best predictors of nonnativeness and invasiveness across all models.

4 | DISCUSSION

Many studies have described conditions about the native range that could increase the risk that an introduced plant will become invasive. Risk assessments would benefit from including only the predictor variables that can effectively differentiate invasives. The present study provides a comprehensive analysis of the significance and predictive power of biogeographic properties for identifying invasion risk. Variables representing niche breadth in the native range were the strongest predictors of invasiveness in our models, and spatial datasets of these variables are widely available, making them relatively easy to incorporate into future risk assessments.

4.1 | How are nonnative species different from native species?

Nearly all predictor variables showed significant differences between native species and species that were established elsewhere (Figure 2, Table 2). These results support the hypotheses that human contact, niche breadth, disturbance, resource availability and native plant richness are related to the successful establishment of plants outside of their native ranges. For example,

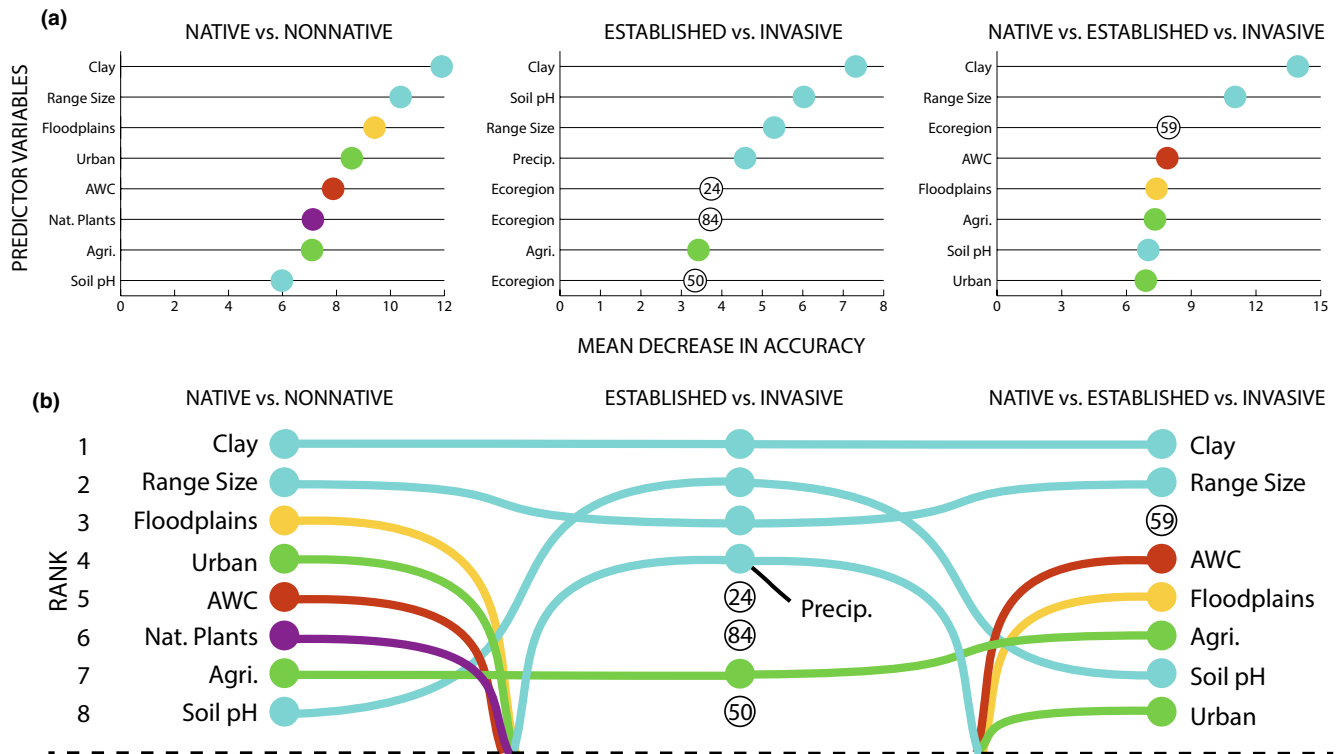


FIGURE 3 Niche breadth variables are consistently important predictors across models. (a) Variable importance as determined by mean decrease in accuracy from each random forest model. (b) The eight most important variables from each random forest model (full rankings provided in supplementary material A). Colours correspond to hypothesized underlying mechanisms from Figure 1. Numbered circles represent the corresponding EPA Level III Ecoregions. Abbreviations: AWC = soil available water capacity, Urban = urban land cover, Agri. = agricultural land cover, Nat. Plants = native plant species richness, Precip. = precipitation.

proximity to human infrastructure increases the likelihood that a plant will be discovered and introduced to a new environment, thereby increasing the chance of establishment (Hulme, 2009; Pouteau et al., 2021). Similarly, species that can survive a wide range of abiotic conditions are more likely to successfully establish in a novel environment (Gallagher et al., 2015; Higgins & Richardson, 2014). Plants with greater disturbance in their native range could be more responsive to human disturbance and therefore apt to establish in disturbed environments in a novel range. Higher resource availability and native plant richness in the native range could lead to more efficient or more competitive plants that are likely to establish in disturbed environments elsewhere (Blumenthal, 2006; Davis et al., 2000; Fridley & Sax, 2014). Our results suggest that all of these hypotheses could lead to significant differences for native and established plant species.

4.2 | Which differences between native and nonnative species are most useful for prediction?

In addition to having significant differences for native and nonnative species, some properties associated with each of our categories

were also important predictors of establishment. (Figure 3). A model based on these biogeographic properties had high overall accuracy and moderate agreement based on the kappa statistic (Cohen, 1960). Thus, if our goal was to assess whether a species could establish in a nonnative range, a broad set of biogeographic properties could be both significant and important for prediction. Unfortunately, predicting establishment alone is not sufficient for risk assessment, since invasive plants make up a substantial fraction of established species (~12%; Simberloff et al., 2012) and have extensively documented negative impacts (Vilà et al., 2011).

4.3 | How are invasive species different from established species?

Established and invasive species differed in the breadth of their niches, and in the amount of human contact they were exposed to in their native ranges. While results for niche breadth variables largely matched our hypotheses for established and invasive species, the other variables did not. Unexpectedly, when compared to established species, invasive species had (on average) less human contact, equivalent disturbance, equivalent resource availability and less native plant richness in their native ranges.

4.4 | Which differences between established and invasive species are most useful for prediction?

Of our three models, the one that compared established versus invasive species is likely to be the one with the most practical significance. In order for a species to be proposed for introduction in a new location, someone (the proposer and possibly others) is likely to have an existing belief that the species can survive in the new location. In other words, it would be unusual to propose a species for introduction to a new location if there was little chance of it surviving. Thus, plants which are perceived as unlikely to survive outside of their native ranges (e.g., the “native” species in the present study) would be selectively removed before the start of the WRA evaluation process. This means that in practice, WRAs are most useful when they can distinguish between established and invasive species (regardless of the ability to classify native species) since most species that ultimately undergo evaluation will have a high likelihood of establishment.

Variables representing niche breadth were the most important predictors for the model that compared established versus invasive species (Figure 3). Thus, our analysis clearly supports the inclusion of niche breadth variables in invasive plant risk assessments. In particular, the native range area, range of soil pH and range of soil textures were all important predictors of invasiveness. These findings support Higgins and Richardson (2014), who found similar importance of abiotic niche breadth for distinguishing between established versus invasive eucalypt and acacia species. It appears that a broad environmental tolerance enables species to both establish within and invade nonnative ranges. The combination of significance and importance of niche breadth for differentiating established from invasive species suggests that native range size and range of soil characteristics should be included in WRAs.

Interestingly, our results did not find the breadth of precipitation in a species' native range to be an especially important predictor of invasiveness. Precipitation was ranked fourth in the model comparing established and invasive species, but tenth in the other two random forest models. These results are supported by Gardner et al. (2019), who found that precipitation was a relatively poor predictor of plant distributions in comparison with other variables that more accurately represented water availability. While climate matching (between possible new locations and existing ranges) is effective for predicting plants' future ranges (Franklin, 1995), the breadth of individual climate properties appears to be less useful for risk assessment. Relatedly, Willi and Van Buskirk (2022) noted that plant survival in climatic extremes is commonly associated with smaller size and lower fecundity, and Kalusová et al. (2013) found invasive species to be less likely to originate in environments with extreme climates. In these cases, plant species with especially large climatic tolerances may actually be less likely to invade since the physiological requirements for surviving such harsh conditions may lead to lower competitiveness. Ultimately, our results suggest that variables representing climatic breadth are not likely to improve WRAs.

Contrary to the evolutionary imbalance hypothesis (Fridley & Sax, 2014), we did not find evidence that native plant richness

influences invasiveness. However, our analysis used native plant species richness as a proxy for phylogenetic diversity; therefore, it is not an exact test of the mechanism proposed by Fridley and Sax (2014) and better proxies of phylogenetic diversity could prove to be important predictors. Similarly, we did not find evidence that a native range with high resource availability influences invasiveness (Blumenthal, 2006). The significance of resource availability and disturbance for predicting establishment but not invasion suggests that ruderal species are more likely to establish elsewhere (particularly in disturbed areas), but are not more likely to become invasive.

4.5 | Importance of native ecoregions

Although it is unlikely to be generalizable for risk assessment purposes, certain source ecoregions were important predictors of invasiveness. This finding supports existing studies that have suggested that certain habitats serve as sources for disproportionately large numbers of invasive species (e.g., Hejda et al., 2009). EPA ecoregion 24 (Chihuahuan Deserts) ranked highly in the model that compared established and invasive species (Figure 3). Over half (76) of the invasive species used in our analyses had some portion of their native range in the Chihuahuan Deserts. Despite being a strong predictor of invasiveness, the Chihuahuan Deserts did not contain a proportionately large number of invasive plants compared with the number of native and established-elsewhere plants also found within the ecoregion. Previous research has suggested that invasive plants are most likely to come from areas that themselves are heavily invaded (Kalusová et al., 2013), and the Chihuahuan Deserts are notable for several major invasions of both native and nonnative species (Brooks & Pyke, 2001). The importance of ecoregion variables in our models indicates that unique properties of specific native geographies increase invasion likelihood (Kalusová et al., 2013). In other words, for reasons not explained by other environmental predictors, some donor ranges produce relatively large numbers of invasive plant species (Figure 4).

4.6 | Implications for improving WRAs

It is challenging to compare the accuracy estimates from this analysis to previous accuracy assessments of weed risk assessments. First, most WRAs have been created and validated using comparatively small sample sizes. For example, Pheloung et al. (1999) used a sample of 370 plants to create their WRA, while Koop et al. (2012) used 204 species. Pheloung et al. (1999) reported that the correlation between the scores for their WRA and the invasiveness scores assigned independently by experts is 0.686. While this represents good agreement, there remains room for improvement. Our random forest model comparing established versus invasive plants had an overall accuracy of 0.64 and a Cohen's kappa of 0.3 (fair agreement). Taken together, these values indicate that abiotic niche breadth variables could be important predictors in weed risk assessments, but

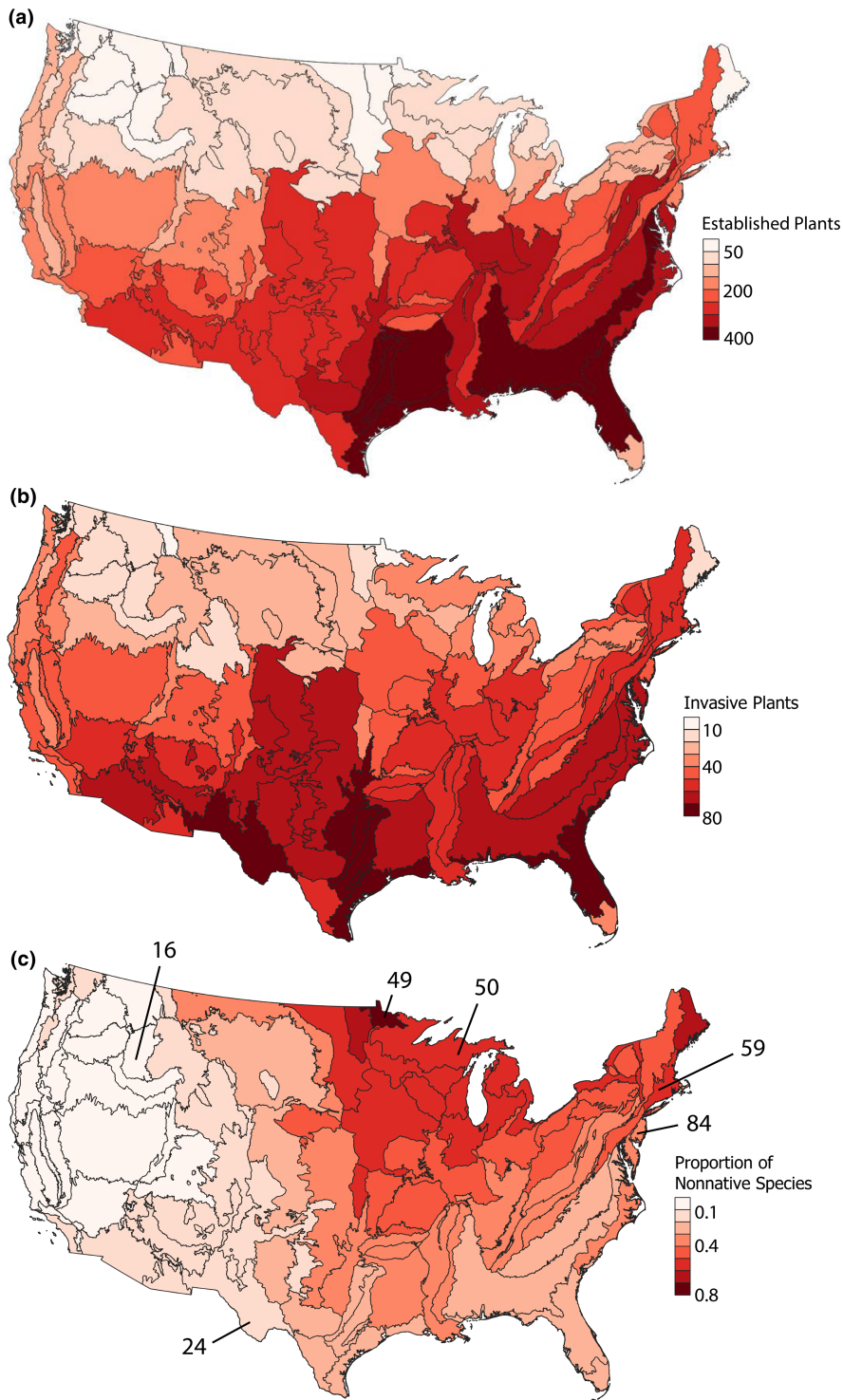


FIGURE 4 Proportions of native, established and invasive species vary across ecoregions, although none are consistently useful predictors of invasiveness. (a) The absolute number of plant species that are native to each of the EPA Level III Ecoregions and established in at least one location elsewhere in the world. (b) The absolute number of plant species that are native to each of the ecoregions and invasive elsewhere in the world. Ecoregion 24 (Chihuahuan Deserts) was ranked fifth in the model that compared established and invasive species. (c) The proportions of the total species native to each ecoregion that have become either established or invasive elsewhere in the world.

are likely not sufficient on their own for identifying invasive plants. We are unaware of a comprehensive analysis of biological traits for established versus invasive species (although Sutherland, 2004 compared life history traits). Therefore, although we cannot determine the relative importance of biogeographic properties compared with biological traits, it is clear that niche breadth properties should be included.

Another reason for including biogeographic variables in WRAs is that these variables are easy to find for many parts of the world,

especially for common donor regions like North America and Europe (van Kleunen et al., 2015). For example, species distribution data (for calculating range size; Chytrý et al., 2016) and soil predictor data (Fabian et al., 2014 for soil pH; Ballabio et al., 2016 for soil texture) are readily available in Europe. Whereas determining the biological traits of plants requires individual measurements for each species (which are lacking for understudied species; Kattge et al., 2011), biogeographic properties can be determined once for a given source habitat and subsequently applied to all the species that are native to

the habitat. This is particularly useful since a lack of detailed information about biological traits has previously affected which species scientists assess, which traits are considered for inclusion in WRAs, and the predictive performance of WRAs (van Kleunen et al., 2010; Pheloung et al., 1999). Thus, the inclusion of biogeographic properties might make it possible for WRAs to assess more species before they become invasive.

4.7 | Additional considerations

Limitations of temporal and spatial scale could affect our results. First, we assumed that established species were not invasive, but lag phases in plant invasion can last for decades (Aikio et al., 2010) so it is likely that some species we classified as established will become invasive. Appendix S2 includes a list of potentially high-risk established species that were consistently predicted by our models as invasive and therefore may currently be in lag phases. Second, the U.S. county scale is a fairly coarse resolution that could mask some important finer scale patterns. These data are the most consistent currently available for all flora, but future analyses would optimally use higher resolution distribution data.

While we see consistent biogeographic predictors for U.S. species (Figure 3), it is worth noting that native range biogeography in the U.S. has distinct differences from other parts of the world. For example, our study area mainly consisted of temperate or semi-arid ecosystems. Thus, our results may not hold for species originating in the tropics. North American plants also have a shorter evolutionary history alongside humans, which might lead to fewer human-associated plants than in Africa or Eurasia. In both cases, these could lead to different predictive power of native range properties for plants originating from other geographies.

5 | CONCLUSION

Our results demonstrate that biogeographic properties which quantify the breadth of plant species' native ranges are strong predictors of plant invasiveness. These trends indicate that species capable of surviving under a range of habitat conditions are more likely to both establish and invade in new locations. This information has practical significance for streamlining risk assessments and improving the accuracy of future invasive plant predictions.

ACKNOWLEDGEMENTS

This research was supported by DOI Northeast Climate Adaptation Science Center graduate fellowships awarded to BBL and WGP (Grant No. G19AC00091 from the U.S. Geological Survey), by the National Institute of Food and Agriculture, U.S. Department of Agriculture, the Massachusetts Agricultural Experiment Station and the Department of Environmental Conservation under project no. MAS00033 and by a fellowship from the Lotta Crabtree Trust awarded to WGP. The authors wish to thank the members of the

UMass Spatial Ecology Lab as well as Matthew Fertakos, Meghan Graham MacLean and Katrina Zarella Smith for feedback on analyses, figures and phrasing.

CONFLICT OF INTEREST

We have no conflict of interest to declare.

DATA AVAILABILITY STATEMENT

The data and code that support the findings of this study are openly available on GitHub at <https://github.com/wpfadenhauer/Remember-Your-Roots>.

PEER REVIEW

The peer review history for this article is available at <https://publons.com/publon/10.1111/ddi.13639>.

ORCID

William G. Pfadenhauer  <https://orcid.org/0000-0001-7519-1014>

REFERENCES

- Aikio, S., Duncan, R. P., & Hulme, P. E. (2010). Lag-phases in alien plant invasions: Separating the facts from the artefacts. *Oikos*, 119(2), 370–378. <https://doi.org/10.1111/j.1600-0706.2009.17963.x>
- Baker, H. G. (1965). Characteristics and modes of origin of weeds. In *Characteristics and modes of origin of weeds* (pp. 147–172). Academic Press Inc.
- Ballabio, C., Panagos, P., & Monatanarella, L. (2016). Mapping topsoil physical properties at European scale using the LUCAS database. *Geoderma*, 261, 110–123. <https://doi.org/10.1016/j.geoderma.2015.07.006>
- Barker, K., & Francis, R. A. (2021). *Routledge handbook of biosecurity and invasive species*. Routledge.
- Beaury, E. M., Fusco, E. J., Allen, J. M., & Bradley, B. A. (2021). Plant regulatory lists in the United States are reactive and inconsistent. *Journal of Applied Ecology*, 58(9), 1957–1966. <https://doi.org/10.1111/1365-2664.13934>
- Bhattarai, G. P., & Cronin, J. T. (2014). Hurricane activity and the large-scale pattern of spread of an invasive plant species. *PLoS One*, 9(5), e98478. <https://doi.org/10.1371/journal.pone.0098478>
- Blumenthal, D. (2006). Interactions between resource availability and enemy release in plant invasion. *Ecology Letters*, 9(7), 887–895. <https://doi.org/10.1111/j.1461-0248.2006.00934.x>
- Blumenthal, D., Mitchell, C. E., Pysek, P., & Jarosik, V. (2009). Synergy between pathogen release and resource availability in plant invasion. *PNAS*, 106(19), 7899–7904.
- Brooks, M. L., & Pyke, D. A. (2001). Invasive plants and fire in the deserts of North America. In K. Galley & T. Wilson (Eds.), *Proceedings of the invasive plant workshop* (pp. 1–14). Tall Timbers Research Station. https://www.firescience.gov/projects/00-1-2-04/project/00-1-2-04_00_1_2_04_Deliverable_04.pdf
- CAB International. (2021, May 11). *Invasive species compendium*. CABI. <https://www.cabi.org/isc>
- Caley, P., & Kuhnert, P. M. (2006). Application and evaluation of classification trees for screening unwanted plants. *Austral Ecology*, 31(5), 647–655. <https://doi.org/10.1111/j.1442-9993.2006.01617.x>
- Chytrý, M., Hennekens, S. M., Jiménez-Alfaro, B., Knollová, I., Dengler, J., Jansen, F., Landucci, F., Schaminée, J. H. J., Acic, S., Agrillo, E., Ambarlı, D., Angelini, P., Apostolova, I., Attorre, F., Berg, C., Bergmeier, E., Biurrún, I., Botta-Dukát, Z., Brisse, H., ... Yamalov, S.

- (2016). European vegetation archive (EVA): An integrated database of European vegetation plots. *Applied Vegetation Science*, 19(1), 173–180. <https://doi.org/10.1111/avsc.12191>
- Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, 20(1), 37–46. <https://doi.org/10.1177/001316446002000104>
- Conser, C., Seebacher, L., Fujino, D. W., Reichard, S., & DiTomaso, J. M. (2015). The development of a plant risk evaluation (PRE) tool for assessing the invasive potential of ornamental plants. *PLoS One*, 10(3), e0121053. <https://doi.org/10.1371/journal.pone.0121053>
- Daehler, C. (2003). Performance comparisons of co-occurring native and alien invasive plants: Implications for conservation and restoration. *Annual Review of Ecology, Evolution, and Systematics*, 34, 183–211.
- Davies, K. W., & Johnson, D. D. (2011). Are we “missing the boat” on preventing the spread of invasive plants in rangelands? *Invasive Plant Science and Management*, 4(1), 166–171. <https://doi.org/10.1614/IPSM-D-10-00030.1>
- Davies, K. W., & Sheley, R. L. (2007). A conceptual framework for preventing the spatial dispersal of invasive plants. *Weed Science*, 55(2), 178–184. <https://doi.org/10.1614/WS06-161>
- Davis, M. A., Grime, J. P., & Thompson, K. (2000). Fluctuating resources in plant communities: A general theory of invasibility. *Journal of Ecology*, 88, 528–534. <https://doi.org/10.1046/j.1365-2745.2000.00473.x>
- Early, R., Bradley, B. A., Dukes, J. S., Lawler, J. J., Olden, J. D., Blumenthal, D. M., Gonzalez, P., Grosholz, E. D., Ibañez, I., Miller, L. P., Sorte, C. J. B., & Tatem, A. J. (2016). Global threats from invasive alien species in the twenty-first century and national response capacities. *Nature Communications*, 7(1), 12485. <https://doi.org/10.1038/ncomms12485>
- Fabian, C., Reimann, C., Fabian, K., Birke, M., Baritz, R., & Haslinger, E. (2014). GEMAS: Spatial distribution of the pH of European agricultural and grazing land soil. *Applied Geochemistry*, 48, 207–216. <https://doi.org/10.1016/j.apgeochem.2014.07.017>
- Franklin, J. (1995). Predictive vegetation mapping: Geographic modeling of biospatial patterns in relation to environmental gradients. *Progress in Physical Geography*, 19(4), 474–499.
- Fridley, J. D., & Sax, D. F. (2014). The imbalance of nature: Revisiting a Darwinian framework for invasion biology. *Global Ecology and Biogeography*, 23(11), 1157–1166. <https://doi.org/10.1111/geb.12221>
- Gallagher, R. V., Randall, R. P., & Leishman, M. R. (2015). Trait differences between naturalized and invasive plant species independent of residence time and phylogeny. *Conservation Biology*, 29(2), 360–369. <https://doi.org/10.1111/cobi.12399>
- Gardner, A. S., Maclean, I. M. D., & Gaston, K. J. (2019). Climatic predictors of species distributions neglect biophysically meaningful variables. *Diversity and Distributions*, 25(8), 1318–1333. <https://doi.org/10.1111/ddi.12939>
- Guevara, M., Arroyo-cruz, C. E., Brunsell, N., Cruz-gaistardo, C. O., Domke, G. M., Equihua, J., Etchevers, J., Hayes, D. J., Hengl, T., Ibellas, A., Johnson, K., de Jong, B., Libohova, Z., Llamas, R., Nave, L., Ornelas, J. L., Paz, F., Ressler, R., Schwartz, A., ... Vargas, R. (2020). *Soil organic carbon estimates for 30-cm depth, Mexico and Conterminous USA, 1991–2011*. ORNL DAAC. <https://doi.org/10.3334/ORNLD AAC/1737>
- Heikkilä, J. (2011). A review of risk prioritisation schemes of pathogens, pests and weeds: Principles and practices. *Agricultural and Food Science*, 20, 15–28.
- Heikkilä, J., Tuomola, J., Pouta, E., & Hannunen, S. (2016). FinnPRIO: A model for ranking invasive plant pests based on risk. *Biological Invasions*, 18(7), 1827–1842. <https://doi.org/10.1007/s10530-016-1123-4>
- Hejda, M., Pyšek, P., Pergl, J., Sádlo, J., Chytrý, M., & Jarošík, V. (2009). Invasion success of alien plants: Do habitat affinities in the native distribution range matter? *Global Ecology and Biogeography*, 18(3), 372–382. <https://doi.org/10.1111/j.1466-8238.2009.00445.x>
- Higgins, S. I., & Richardson, D. M. (2014). Invasive plants have broader physiological niches. *Proceedings of the National Academy of Sciences*, 111(29), 10610–10614. <https://doi.org/10.1073/pnas.1406075111>
- Ho, T. K. (1995). Random decision forests. In *Proceedings of the 3rd international conference on document analysis and recognition* (pp. 278–282). Institute of Electrical and Electronics Engineers (IEEE). <https://web.archive.org/web/20160417030218/http://ect.bell-labs.com/who/tkh/publications/papers/odt.pdf>
- Hobbs, R. J., & Huenneke, L. F. (1992). Disturbance, diversity, and invasion: Implications for conservation. *Conservation Biology*, 6(3), 324–337. <https://doi.org/10.1046/j.15231739.1992.06030324.x>
- Hock, M., Hofmann, R., Essl, F., Pyšek, P., Bruehlheide, H., & Erfmeier, A. (2020). Native distribution characteristics rather than functional traits explain preadaptation of invasive species to high-UV-B environments. *Diversity and Distributions*, 26(10), 1421–1438. <https://doi.org/10.1111/ddi.13113>
- Homer, C., Huang, C., Yang, L., Wylie, B., & Coan, M. (2004). Development of a 2001 National Land-Cover Database for the United States. *Photogrammetric Engineering & Remote Sensing*, 70(7), 829–840. <https://doi.org/10.14358/PERS.70.7.829>
- Hulme, P. E. (2009). Trade, transport and trouble: Managing invasive species pathways in an era of globalization. *Journal of Applied Ecology*, 46(1), 10–18. <https://doi.org/10.1111/j.1365-2664.2008.01600.x>
- Hulme, P. E. (2012). Weed risk assessment: A way forward or a waste of time? *Journal of Applied Ecology*, 49(1), 10–19. <https://doi.org/10.1111/j.1365-2664.2011.02069.x>
- International Union for Conservation of Nature. (2021, May 11). *Global invasive species database*. Invasive Species Specialist Group. <http://www.iucngisd.org/gisd/>
- Kalusová, V., Chytrý, M., Kartesz, J. T., Nishino, M., & Pyšek, P. (2013). Where do they come from and where do they go? European natural habitats as donors of invasive alien plants globally. *Diversity and Distributions*, 19(2), 199–214. <https://doi.org/10.1111/ddi.12008>
- Kattge, J., Díaz, S., Lavorel, S., Prentice, I. C., Leadley, P., Bönlisch, G., Garnier, E., Westoby, M., Reich, P. B., Wright, I. J., Cornelissen, J. H. C., Violle, C., Harrison, S. P., Bodegom, P. M. V., Reichstein, M., Enquist, B. J., Soudzilovskaia, N. A., Ackerly, D. D., Anand, M., ... Wirth, C. (2011). TRY – A global database of plant traits. *Global Change Biology*, 17(9), 2905–2935. <https://doi.org/10.1111/j.1365-2486.2011.02451.x>
- Keller, R. P., Lodge, D. M., & Finnoff, D. C. (2007). Risk assessment for invasive species produces net bioeconomic benefits. *Proceedings of the National Academy of Sciences*, 104(1), 203–207. <https://doi.org/10.1073/pnas.0605787104>
- Kleunen, M. V., Weber, E., & Fischer, M. (2010). A meta-analysis of trait differences between invasive and non-invasive plant species. *Ecology Letters*, 13(2), 235–245. <https://doi.org/10.1111/j.1461-0248.2009.01418.x>
- Knapp, K. R., Diamond, H. J., Kossin, J. P., Kruk, M. C., & Schreck, C. J. (2018). *International best track archive for climate stewardship (IBTrACS) project, version 4*. NOAA National Centers for Environmental Information. <https://doi.org/10.25921/82ty-9e16>
- Knapp, K. R., Kruk, M. C., Levinson, D. H., Diamond, H. J., & Neumann, C. J. (2010). The international best track archive for climate stewardship (IBTrACS): Unifying tropical cyclone best track data. *Bulletin of the American Meteorological Society*, 91, 363–376.
- Koop, A. L., Fowler, L., Newton, L. P., & Caton, B. P. (2012). Development and validation of a weed screening tool for the United States. *Biological Invasions*, 14(2), 273–294. <https://doi.org/10.1007/s10530-011-0061-4>
- Laginhas, B. B., & Bradley, B. A. (2021). Global plant invaders: A compendium of invasive plant taxa documented by the peer-reviewed

- literature. *Ecology*, 103(2), e03569. <https://doi.org/10.1002/ecy.3569>
- LANDFIRE. (2016). *Biophysical settings*. United States Geological Survey. <https://www.landfire.gov>
- Leung, B., Lodge, D. M., Finnoff, D., Shogren, J. F., Lewis, M. A., & Lamberti, G. (2002). An ounce of prevention or a pound of cure: Bioeconomic risk analysis of invasive species. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, 269(1508), 2407–2413. <https://doi.org/10.1098/rspb.2002.2179>
- Mack, R. N., & Lonsdale, W. M. (2002). Eradicating plants: Hard-won lessons for islands. In E. C. R. Veitch & M. N. Clout (Eds.), *Turning the tide: The eradication of invasive species* (pp. 164–172). IUCN SSC Invasive Species Specialist Group.
- Mack, R. N., Simberloff, D., Lonsdale, W. M., Evans, H., Clout, M., & Bazzaz, F. A. (2000). Biotic invasions: Causes, epidemiology, global consequences, and control. *Ecological Applications*, 10(3), 689–710. [https://doi.org/10.1890/1051-0761\(2000\)010\[0689:BICEG\]2.0.CO;2](https://doi.org/10.1890/1051-0761(2000)010[0689:BICEG]2.0.CO;2)
- Miller, D. A., & White, R. A. (1998). A conterminous United States multi-layer soil characteristics data set for regional climate and hydrology modeling. *Earth Interactions*, 2, 711–724. <http://soilinfo.psu.edu>
- National Weather Service. (2009). *Tornado tracks*. National Atlas of the United States. <https://www.sciencebase.gov/catalog/item/4f4e4ac5e4b07f02db679d0c>
- Natural Resource Conservation Service. (2021). *The PLANTS database*. United States Department of Agriculture. <https://plants.usda.gov>
- Omernik, J. M., & Griffith, G. E. (2014). Ecoregions of the conterminous United States: Evolution of a hierarchical spatial framework. *Environmental Management*, 54, 1249–1266.
- Ou, J., Lu, C., & O'toole, D. K. (2008). A risk assessment system for alien plant bio-invasion in Xiamen. *China. Journal of Environmental Sciences*, 20(8), 989–997. [https://doi.org/10.1016/S1001-0742\(08\)62198-1](https://doi.org/10.1016/S1001-0742(08)62198-1)
- Panetta, F. D. (1993). A system of assessing proposed plant introductions for weed potential. *Plant Protection Quarterly*, 8(1), 10–14.
- Parker, C., Caton, B. P., & Fowler, L. (2007). Ranking nonindigenous weed species by their potential to invade the United States. *Weed Science*, 55(4), 386–397.
- Pheloung, P. C., Williams, P. A., & Halloy, S. R. (1999). A weed risk assessment model for use as a biosecurity tool evaluating plant introductions. *Journal of Environmental Management*, 57(4), 239–251. <https://doi.org/10.1006/jema.1999.0297>
- Pimentel, D., Zuniga, R., & Morrison, D. (2005). Update on the environmental and economic costs associated with alien-invasive species in the United States. *Ecological Economics*, 52(3), 273–288. <https://doi.org/10.1016/j.ecolecon.2004.10.002>
- Pouteau, R., Thuiller, W., Hobohm, C., Brunel, C., Conn, B. J., Dawson, W., Dechoum, M. de S., Ebel, A. L., Essl, F., Fragman-Sapir, O., Fristoe, T., Jogan, N., Kreft, H., Lenzner, B., Meyer, C., Pergl, J., Pyšek, P., Verkhovzina, A., Weigelt, P., ... Kleunen, M. (2021). Climate and socio-economic factors explain differences between observed and expected naturalization patterns of European plants around the world. *Global Ecology and Biogeography*, 30(7), 1514–1531. <https://doi.org/10.1111/geb.13316>
- PRISM Climate Group. (2015a). *Thirty year Normal mean temperature: Annual dataset*. Oregon State University. <http://prism.oregonstate.edu>
- PRISM Climate Group. (2015b). *Thirty Year Normal Precipitation: Annual dataset*. Oregon State University. <http://prism.oregonstate.edu>
- Pyšek, P., & Richardson, D. M. (2007). Traits associated with invasiveness in alien plants: Where do we stand? In W. Nentwig (Ed.), *Biological Invasions* (pp. 97–125). Springer. https://doi.org/10.1007/978-3-540-36920-2_7
- R Core Team. (2020). *R: A language for environment and statistical computing*. R Foundation for Statistical Computing. <http://www.R-project.org/>
- Reichard, S. H., & Hamilton, C. W. (1997). Predicting invasions of Woody plants introduced into North America. *Conservation Biology*, 11(1), 193–203. <https://doi.org/10.1046/j.1523-1739.1997.95473.x>
- Rejmánek, M., & Pyšek, P. (2002). When is eradication of exotic pest plants a realistic goal? In *Turning the tide: The eradication of invasive species: Proceedings of the International Conference on Eradication of Island Invasives* (pp. 249–253). IUCN.
- Richardson, D. M., Pyšek, P., Rejmánek, M., Barbour, M. G., Panetta, F. D., & West, C. J. (2000). Naturalization and invasion of alien plants: Concepts and definitions. *Diversity and Distributions*, 6(2), 93–107. <https://doi.org/10.1046/j.1472-4642.2000.00083.x>
- Roy, H. E., Rabitsch, W., Scaler, R., Stewart, A., Gallardo, B., Genovesi, P., Essl, F., Adriaens, T., Bacher, S., Booy, O., Branquart, E., Brunel, S., Copp, G. H., Dean, H., D'hondt, B., Josefsson, M., Kenis, M., Kettunen, M., Linnamagi, M., ... Zenetos, A. (2018). Developing a framework of minimum standards for the risk assessment of alien species. *Journal of Applied Ecology*, 55(2), 526–538. <https://doi.org/10.1111/1365-2664.13025>
- Simberloff, D. (2001). Biological invasions—How are they affecting us, and what can we do about them? *Western North American Naturalist*, 61(3), 308–315.
- Simberloff, D., Souza, L., Nuñez, M. A., Barrios-Garcia, M. N., & Bunn, W. (2012). The natives are restless, but not often and mostly when disturbed. *Ecology*, 93(3), 598–607. <https://doi.org/10.1890/11-1232.1>
- Sutherland, S. (2004). What makes a weed a weed: Life history traits of native and exotic plants in the USA. *Oecologia*, 141(1), 24–39. <https://doi.org/10.1007/s00442-004-1628-x>
- Theobald, D. M., Kennedy, C., Chen, B., Oakleaf, J., Baruch-Mordo, S., & Kiesecker, J. (2020). Earth transformed: Detailed mapping of global human modification from 1990 to 2017. *Earth System Science Data*, 12(3), 1953–1972. <https://doi.org/10.5194/essd-12-19532020>
- United States Geological Survey. (1999). *Airports of the United States, Puerto Rico and the U.S. Virgin Islands*. National Atlas of the United States.
- van Kleunen, M., Dawson, W., Essl, F., Pergl, J., Winter, M., Weber, E., Kreft, H., Weigelt, P., Kartesz, J., Nishino, M., Antonova, L. A., Barcelona, J. F., Cabezas, F. J., Cardenas, D., Cardenas-Toro, J., Castano, N., Chacon, E., Chatelain, C., Ebel, A. L., ... Pyšek, P. (2015). Global exchange and accumulation of non-native plants. *Nature*, 525, 100–103.
- van Kleunen, M., Pyšek, P., Dawson, W., Essl, F., Kreft, H., Pergl, J., Weigelt, P., Stein, A., Dullinger, S., König, C., Lenzner, B., Maurel, N., Moser, D., Seebens, H., Kartesz, J., Nishino, M., Aleksanyan, A., Ansong, M., Antonova, L. A., ... Winter, M. (2019). The global naturalized alien Flora (GloNAF) database. *Ecology*, 100(1), e02542. <https://doi.org/10.1002/ecy.2542>
- Verbrugge, L. N. H., Leuven, R. S. E. W., & van der Velde, G. (2010). *Evaluation of international risk assessment protocols for exotic species* (Series of Reports on Environmental Science, p. 59). Institute for Water and Wetland Research.
- Vilà, M., Espinar, J. L., Hejda, M., Hulme, P. E., Jarošík, V., Maron, J. L., Pergl, J., Schaffner, U., Sun, Y., & Pyšek, P. (2011). Ecological impacts of invasive alien plants: A metaanalysis of their effects on species, communities and ecosystems. *Ecology Letters*, 14(7), 702–708. <https://doi.org/10.1111/j.1461-0248.2011.01628.x>
- Waterborne Commerce Statistics Center. (2013). *Major ports dataset*. United States Army Corps of Engineers. <https://catalog.data.gov/dataset/major-ports-national>
- Weber, J., Dane Panetta, F., Virtue, J., & Pheloung, P. (2009). An analysis of assessment outcomes from eight years' operation of the Australian border weed risk assessment system. *Journal of Environmental Management*, 90(2), 798–807. <https://doi.org/10.1016/j.jenvman.2008.01.012>
- Willi, Y., & Van Buskirk, J. (2022). A review on trade-offs at the warm and cold ends of geographical distributions. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 377(1848), 20210022. <https://doi.org/10.1098/rstb.2021.0022>

Woznicki, S. A., Baynes, J., Panlasigui, S., Mehaffey, M., & Neale, A. (2019). Development of a spatially complete floodplain map of the conterminous United States using random forest. *Science of the Total Environment*, 647, 942–953. <https://doi.org/10.1016/j.scitotenv.2018.07.353>

BIOSKETCHES

William Pfadenhauer is interested in using spatial and statistical technologies to better understand plant invasions at a global scale, with an emphasis on evaluating and improving management and regulations.

Michael Nelson is interested in how spatially aware tools such as spatial statistics and agent-based modelling can help us understand ecological phenomena that are inherently spatial, such as plant invasions and the spread of forest pests and pathogens.

Brit Laginhas is interested in using geospatial methods to produce dynamic forecasts characterizing invasive species spread across landscapes.

Bethany Bradley is interested in how the geographical locations of invasive plants across landscapes and regions can inform ecological understanding of species distributions, vulnerability assessments and adaptation planning.

Author Contribution: B.B.L. and B.A.B. devised the project and conceptual framework. W.G.P. and B.B.L. collected the data. W.G.P. and M.F.N. ran the models and analyzed the results. W.G.P. and B.A.B. wrote the manuscript, with contributions from all authors.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Pfadenhauer, W. G., Nelson, M. F., Laginhas, B. B., & Bradley, B. A. (2023). Remember your roots: Biogeographic properties of plants' native habitats can inform invasive plant risk assessments. *Diversity and Distributions*, 29, 4–18. <https://doi.org/10.1111/ddi.13639>