

## Two sides of the same coin? Rare and pest plants native to the United States and Canada

JOHN PAUL SCHMIDT,<sup>1</sup> PATRICK R. STEPHENS, AND JOHN M. DRAKE

*Odum School of Ecology, University of Georgia, Athens, Georgia 30602 USA*

**Abstract.** Plant biodiversity is at risk, with as many as 10% of native species in the United States being threatened with extinction. Habitat loss has led a growing number of plant species to become rare or threatened, while the introduction or expansion of pest species has led some habitats to be dominated by relatively few, mostly nonindigenous, species. As humans continue to alter many landscapes and vegetation types, understanding how biological traits determine the location of species along a spectrum from vulnerability to pest status is critical to designing risk assessment protocols, setting conservation priorities, and developing monitoring programs.

We used boosted regression trees to predict rarity (based on The Nature Conservancy global rankings) and pest status (defined as legal pest status) from data on traits for the native vascular flora of the United States and Canada including Hawaii, Puerto Rico, and the Virgin Islands ( $n \approx 15\,000$ ). Categories were moderately to highly predictable ( $AUC_{\text{pest}} = 0.87$  on 25% holdout test set,  $AUC_{\text{rarity}} = 0.80$  on 25% holdout test set). Key predictors were chromosome number, ploidy, seed mass, and a suite of traits suggestive of specialist vs. generalist adaptations (e.g., facultative wetland habitat association and phenotypic variability in growth form and life history). Specifically, pests were associated with high chromosome numbers, polyploidy, and seed masses ranging from 0.1 to 100 mg, whereas rare species were associated with low chromosome numbers, low ploidy, and large ( $>1000$  mg) seed masses. In addition, pest species were disproportionately likely to be facultatively associated with wetlands, and variable in growth form and life history, whereas rare species exhibited an opposite pattern. These results suggest that rare and pest species contrast along trait axes related to dispersal and performance in disturbed or novel habitats.

**Key words:** *boosted regression trees; chromosome number; dispersal; disturbed habitat; novel habitat; pest plants; plants native to the U.S. and Canada; ploidy; rarity; seed mass.*

### INTRODUCTION

Due to human activities, many species are becoming increasingly rare and vulnerable to global extinction, while other species, whether apparently native or documented introductions, are becoming more abundant and widespread. A subset of the winners in the current human-dominated landscape, whether native or introduced, has come to dominate some communities or habitats to an economically damaging extent. When the term “invasive” is applied to economic pests, nonnative status is generally implied, although some (Valery et al. 2008, 2009) argue for broadening the definition to include natives. Because we focus exclusively on species native to North America in this study, we identify problem species as “pest,” rather than invasive, species. Identifying biological traits associated with tendencies either to vulnerability or to establishment and spread is therefore a critical step toward conserving biodiversity. A large body of literature has developed over the last

several decades establishing that plant species likely to naturalize and become pests within an introduced range can be predicted by traits (see reviews in Cadotte et al. 2006, Pyšek and Richardson 2007, van Kleunen et al. 2009). Another body of literature has compared traits of rare and common plant species (reviews in Kunin and Gaston 1993, Gaston and Kunin 1997, Murray et al. 2002) in terms of measurable properties such as range size, average local abundance, or extinction risk. Constrained by the availability of data, studies of rarity and pest status have generally compared small numbers of species (usually  $<100$ ), have contrasted closely related sets of rare and common or pest (generally nonnative) and non-pest species, and/or have focused on limited geographic regions (reviewed in Pyšek and Richardson 2007). At present, we lack robust empirical generalizations concerning which traits consistently relate to either pest status (Pyšek and Richardson 2007) or rarity (Bevill and Louda 1999, Duncan et al. 2011).

Hybridization and/or polyploidy in plants have been hypothesized to be key factors promoting invasion success (Ellstrand and Schierenbeck 2000, Lee 2002), and recent comparative studies have found evidence for this relationship among both pest (Pandit et al. 2011,

Manuscript received 22 October 2011; revised 7 February 2012; accepted 13 March 2012. Corresponding Editor: R. A. Huffbauer.

<sup>1</sup> E-mail: jps@uga.edu

Schmidt and Drake 2011a) and rare (Pandit 2006) plant species. These studies showed that, compared to overall ploidy frequencies among angiosperms, rare plants are disproportionately likely to be diploid and pest plants are disproportionately likely to be polyploid. One possible explanation for this pattern is that polyploidy provides immediate fitness advantages, particularly in novel environments such as those encountered by newly introduced species or brought about locally by environmental change. Proposed advantages of increased ploidy include higher rates of self-fertilization, greater phenotypic plasticity, and higher levels of heterozygosity that allow for multiple locally adapted types or more rapid adaptation to local conditions (Soltis et al. 2010). To date, few studies have tested for the advantages of polyploidy per se (reviewed in Soltis et al. 2010). However, there is evidence from a meta-analysis linking greater phenotypic plasticity to invasion success in plant species (Davidson et al. 2011; but see Palacio-Lopez and Gianoli 2011). Ploidy and measures of phenotypic plasticity therefore represent axes along which rare and pest species potentially may be discriminated. In addition to ploidy, if a major component of the success of pest plants is the movement of propagules and the subsequent establishment of new populations, an equally important dimension of rarity or pest status should be seed traits that promote or limit dispersal and seedling establishment.

A link between rarity and (generally exotic) pest status has been suggested by van Kleunen and Richardson (2007), with rare and pest species hypothesized to occupy opposing ends of a trait continuum (see background on this hypothesis in Jeschke and Strayer 2008). Comparative studies testing this hypothesis have yielded conflicting results. Jeschke and Strayer (2008) and Blackburn and Jeschke (2009) found little evidence to support this hypothesis in comparative studies of European fish and bird taxa. In contrast, Larson and Olden (2010) found that rare and pest crayfish in the southeastern United States strongly segregate along trait axes such as size, fecundity, and habitat specificity. In studies of plants, a large comparative study of legumes (Bradshaw et al. 2008) and a study across all angiosperms (Pandit et al. 2011) showed opposing tendencies for rare vs. invasive species with respect to chromosome number and ploidy. Other studies have shown that plants that are rare, threatened, or extinct are those with low population growth rates (Pilgrim et al. 2004), little long-distance dispersal (Kolb et al. 2006, Farnsworth 2007, Farnsworth and Ogurcak 2008), and specialized habitat requirements. In contrast, introduced plants that successfully invade are well-adapted to disturbance (Cadotte et al. 2006), tolerant of a range of habitat types, and readily able to disperse over sizeable distances (Schmidt and Drake 2011a, b). These patterns appear likely to hold for similarly successful and expanding native plants (Thompson et al. 1995, Meiners 2007, Leishman et al. 2010, Thompson and Davis 2011).

Here we investigate the role of species traits in determining whether plants are rare or successful. To control for historical factors, we compared the species within the native flora of the United States and Canada. We tested whether rare species and native species identified as economic pests within the same region can be discriminated on the basis of traits related to phenotypic plasticity, chromosome number and ploidy, seed mass, and additional traits for which data were available. We also investigated whether the same traits discriminate among species with small and large geographic ranges. Comparative studies of many species across sizable geographic regions are increasingly enabled by the current expansion of databases cataloging the occurrence, conservation or pest status, and traits of species, and by the development of a variety of analytical tools. However, to our knowledge, no studies have compared the conservation status of a similarly large set of vascular plants (>15 000 species) over a continental geographic area.

## MATERIALS AND METHODS

### *Data sources*

We tested the relationship between species status (*rare* or *pest*) and traits for the entire set of angiosperm species native to the United States and Canada including Hawaii, Puerto Rico, and the Virgin Islands. Similar to Pandit et al. (2011), we used categorical designations of species status (e.g., *critically rare* or *not critically rare*, *pest* or *not pest*), rather than more sensitive continuous measures such as global population size or rate of change, because categorical measures are available for a much larger set of species. Because defining rarity is also complicated by scale dependency (Hartley and Kunin 2003), we relied on independent global ranks assigned by NatureServe (NatureServe 2010), which summarize a set of interlinked variables: total number and condition of populations, population size, range extent and area of occupancy, short- and long-term trends in the foregoing factors, threats, and environmental specificity (NatureServe 2010). We excluded subspecies, species not ranked by NatureServe (1459 species), and species for which no trait data could be compiled (974 species). We considered as rare any species ranked by NatureServe as either “critically imperiled” (<5 occurrences, <1000 individuals, <2000 acres of habitat) or “possibly extinct” ( $n = 1258$  rare species out of 15 228 species in total). For pest species, we considered those species that are listed as “weeds,” a watch list compiled by Plants National Database (USDA NRCS 2012) from sources throughout the United States, of species that are, or are considered to have the potential to become, weedy or problematic in all or part of their U.S. range ( $n = 352$ ; data *available online*).<sup>2</sup> We considered pests to be only those species that are native to North America or the

<sup>2</sup> <http://plants.usda.gov/>

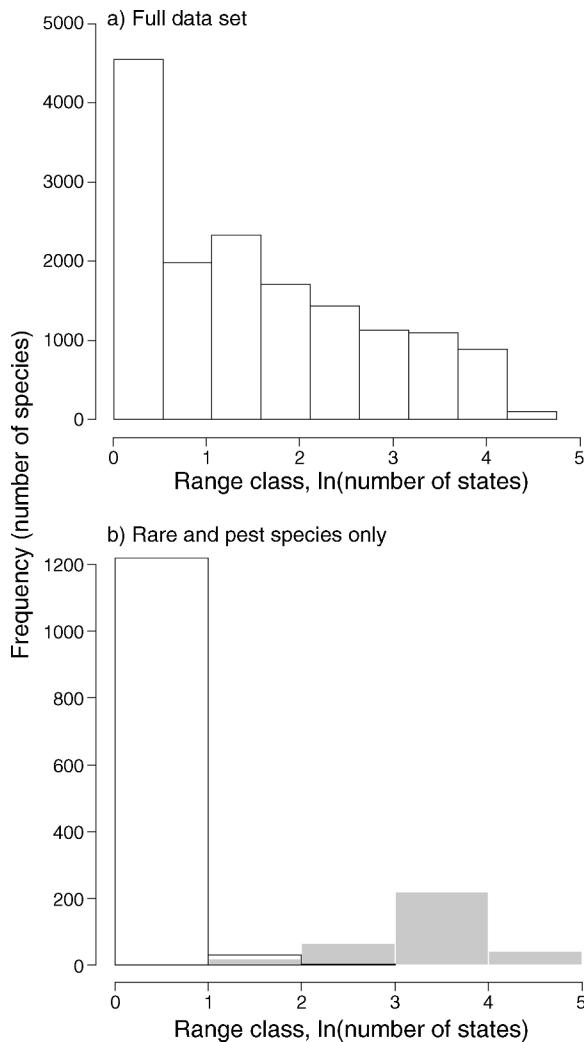


FIG. 1. Histograms of (a) the entire data set and (b) rare and pest species only, by range size class (natural log of number of states within range). Rare species are shown as open bars, and pest species as gray histogram bars in panel (b). Range class 1 includes species known from a single state or province (2084 species, including most of the rare species); nearly all pest species were among the 4548 species making up classes 4 or 5 (13–61 states/provinces).

adjacent Caribbean and that are also pests in those places. We did not include as pests species native to North America or the Caribbean, but anthropogenically introduced to Hawaii, because our objective was to investigate the relationship between traits and pest status within native rather than introduced floras.

To control for the potential economic and political biases in rare and pest designations, we also tested the relationship between traits and a more explicitly ecological variable, range size. As a measure of range size, we obtained from Plants National Database the number of states and provinces forming the range of a species. Obviously, both number of states and area of states in which a species occurs can distort actual range

sizes. Because we were primarily interested in measuring pest status, we chose number of states as less distorting, and then created classes based on the natural logarithm of the number of states/provinces in the United States and Canada. Because Plants National Database does not provide data on range by subnational units for countries beside the United States and Canada, a value of 1 was added for each additional country (e.g., Mexico, Jamaica, Cuba). Because nearly all rare species were among the 2084 species known from a single state or province, class 1 (Fig. 1), and nearly all pest species were among the 4548 species making up classes 4 or 5 (13–61 states/provinces, Fig. 1), we predicted “small range” species vs. all others and “widespread” species vs. all others as a function of traits.

As predictors, we selected traits that captured key aspects of plant biology and genetics and for which data were abundant. We classified 9761 species as biologically pollinated (i.e., insect or vertebrate pollinated) or not biologically pollinated, based on family-level floral morphology (presence or absence of animal attractants such as large, showy flowers and nectar rewards). As potential surrogates for specialist vs. generalist requirements, we classified 15 213 species according to life history (annual, perennial, biennial), 15 223 species by growth form (tree, shrub, subshrub, vine, forb, graminoid), and 15 179 species as obligate wetland, facultative wetland, or non-wetland in habitat association based on data available in Plants National Database. To quantify phenotypic plasticity, we also summed for each species the number of life histories and growth forms known, and further classified species as single or multiple for each set of phenotypes. Multiple life histories were reported for 812 species and multiple growth forms for 3222 species.

For 4189 species, we obtained values for chromosome number from the Index to Plant Chromosome Numbers (IPCN, *available online*)<sup>3</sup> and the Royal Botanic Garden Kew Database (Bennett and Leitch [2005]; updates *available online*).<sup>4</sup> For a subset of these species ( $n = 297$ ) for which multiple chromosome numbers have been reported, we recorded maximum and minimum chromosome number. We obtained values for average seed mass for 4239 species from Royal Botanic Garden Kew Seed Information Database and mean seed mass of congeneric species for 3944 species (Kew Database *available online*).<sup>5</sup> Values for mature height for 1477 species were obtained from Plants National Database. Although seed mass, chromosome number, and height data were available for a fraction of the species, we expected these variables to be highly predictive based on previous studies (Pandit et al. 2011, Schmidt and Drake 2011a).

<sup>3</sup> <http://www.tropicos.org/Project/IPCN>

<sup>4</sup> <http://data.kew.org/cvalues/>

<sup>5</sup> <http://www.kew.org/data/sid>

### *Ploidy*

For each species for which congener data were available (3546 species, of which 275 had multiple values reported), we considered five measures of ploidy in our analyses (following Pandit et al. [2011]). Because some species had multiple values, we calculated (1) highest count for each species, (2) lowest count for each species, (3) the ratio of a species' maximum reported chromosome number to the minimum chromosome number reported for that genus, (4) the ratio of a species' minimum reported chromosome number to the genus minimum, and (5) whether a species had more than one chromosome count reported. Note that highest and lowest raw chromosome numbers and the ploidy ratios based on them differed for only the 275 species for which multiple chromosome numbers were reported.

Because one of the main advantages postulated for polyploids and polyploid hybrids is greater phenotypic plasticity, we investigated potential relationships between ploidy measures and phenotypic variation for those species for which data were available. Accordingly, we classified all species for which we were able to calculate ploidy ratios as diploid if the ploidy ratio was less than or equal to 1 and as polyploid if ploidy ratio was greater than 1. We then performed two-way contingency table analysis on facultative wetland habitat association, life history, and growth form (single vs. multiple) by ploidy class (diploid vs. polyploidy), and chromosome number class (single vs. multiple).

### *Data analysis*

Conventional (linear) multivariate statistics have two weaknesses for the purposes of our study: (1) they are limited to complete data sets (i.e., complete case analysis), and (2) linear analyses cannot easily be used to explore nonlinear relationships between variables. Complete case analysis would have limited us to ~1800 species, greatly reducing statistical power, and there was no a priori reason to expect strictly linear relationships between response and predictor variable (indeed, our results showed that this was frequently not the case). Furthermore, our primary goal was classification: discriminating the critically rare or pest species from the entire pool. Therefore, we used boosted regression trees for classification of rare and pest species from the pool of species native to North America. Because the response variable in each model was binary, models were fit using a Bernoulli distribution.

Although traditional regression trees produce a single, complicated "best" model or tree, boosted regression trees improve on this method by building and merging results from multiple models using a forward, stage-wise, numerical strategy to minimize the loss in predictive performance (Ridgeway 1999, Elith et al. 2008). Boosted regression trees offer the advantage of allowing nonlinear responses and complex interactions. Results reported here were obtained using the *gbm* package in R (Ridgeway 2006). Importantly, *gbm*

handles missing data seamlessly by assigning the value "missing" to records for which no information is available and then allowing use of this value as a possible splitting value in the learned trees. Because every species in our data set is populated for some fields, including all species in the analysis (i.e., not just those with completely populated data fields) allowed for much greater coverage. Because we wished to generalize our results beyond the studied sample and to avoid the unrealistically low error rates returned by in-sample comparisons, we randomly divided the data, stratified by class, into training (75%) and test (25%) sets. We compared models according to the area under the receiver operating curve (AUC), a value ranging between 0 and 1 summarizing the probability that a randomly chosen positive case (invasive) has a higher predicted probability than a randomly chosen negative case (noninvasive). The closer AUC is to 1, the better a model is at discriminating pest from non-pest species. The prediction of withheld test data for final model evaluation was performed only after final selection of predictors and model parameters. We report classification accuracy on the holdout test set in terms of (1) the fraction of rare or pest species correctly labeled, and (2) the fraction of non-rare or non-pest species misclassified. Because accuracy may be optimized to maximize the capture rate or minimize the false positive rate, we report results for either case. Details of the modeling process are provided in the Appendix.

### *Phylogenetic analysis*

Because we were primarily interested in prediction, we opted for the advantages of the boosted regression tree algorithm, rather than adopting a classification method that incorporates phylogeny directly, such as generalized estimating equations (GEE; Paradis 2006). Even if we desired to use the latter method, our data set is too large to accommodate GEE, nor would GEE or related approaches enable the detection of nonlinear relationships between variables. However, machine learning approaches to prediction that are also purpose-built to control for the phylogenetic nonindependence of the unit of analysis, in this case the species, are so far lacking. We therefore incorporated phylogeny at levels above genus by using, as additional explanatory variables a distance matrix created from a supertree of angiosperm families (Davies et al. 2004). In the first method, family-family phylogenetic distances were joined to the data matrix so that, for each species within the same family, values for phylogenetic distance from all other families were the same. Boosted regression tree models were then run with and without the addition of phylogenetic distances as predictors. We were unable to include relationships below the family level because to do so would have required the inclusion of more variables than can be handled (450) by the software that we used. We identified potential phylogenetic nonindependence in our analyses by (1) quantifying

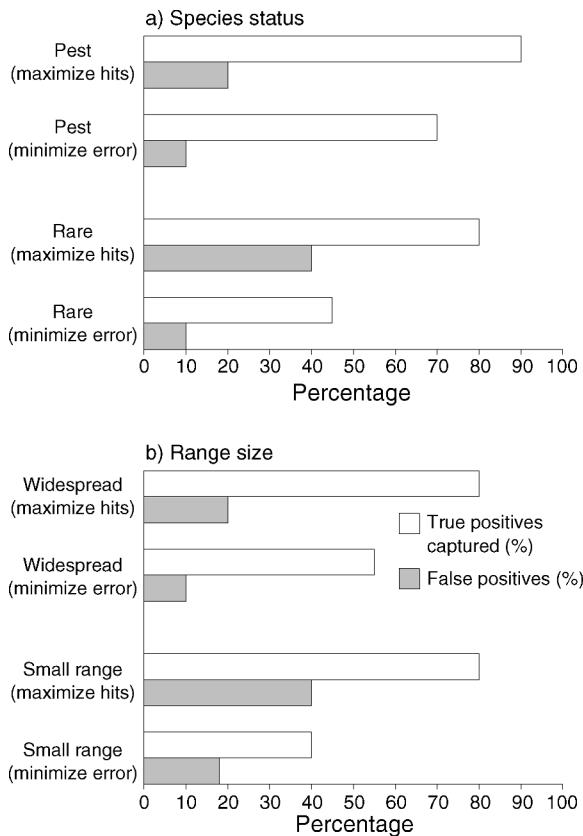


FIG. 2. Error rates for (a) rare and pest species and (b) range size models at two optima, one maximizing the true positive or hit rate, and one minimizing total error (false positives plus false negatives). Values on the  $x$ -axis are percentage of true positives (white bars) and percentage of false positives (gray bars).

improvements in model performance, and (2) determining whether the relative importance of explanatory variables shifted with the addition of phylogenetic information. We also used Blomberg's  $K$  (Blomberg et al. 2003) to directly quantify phylogenetic signal in the

proportion of rare species and the proportion of pest and rare species in families.  $K$  is a measure of phylogenetic signal that varies between 0 and  $\infty$ , where 0 indicates that a character varies independently of phylogeny, values between 0 and 1 indicate partial phylogenetic dependence, a value of 1 indicates phylogenetic dependence similar to that expected under a Brownian motion model of character evolution, and values of greater than 1 indicate greater phylogenetic dependence than expected under a Brownian motion model of character evolution. The statistical significance of observed signal was assessed using a randomization test for signal in continuous characters, as described in Blomberg and Garland (2002). Null distributions were constructed from 1000 randomizations. Estimation of  $K$  and randomization tests used the Davies et al. (2004) angiosperm supertree and were implemented in the R package picante v 1.2-0 (Kembel et al. 2010).

## RESULTS

We achieved high predictive performance in classifying pest species, capturing 82.5% of pest species with a corresponding cost in terms of false positives of only 16.5%, and moderate to high predictive performance in classifying rare species (true positive rate = 60%, false positive rate = 20%). With the goal to optimize accuracy (rather than correctly identifying as many true cases as possible while accepting a moderate level of error), we achieved a hit rate of 65% for pest species with only an 8.3% false positive rate, and a 40% true positive rate with a 6.4% false positive rate for rare species (Fig. 2).

Although some traits such as seed mass, chromosome number, and ploidy ratio were key predictors of both rarity and pest status (Table 1), the relationship of traits to pest status or rarity contrasted sharply. Pest species were most likely at seed masses in the range of 10–100 mg, and were unlikely at extremes of the seed mass distribution (<0.1 mg or >100 mg), whereas rare species were most likely at large (>1 g) seed masses (Fig. 2). The suite of predictors related to ploidy and chromosome number presents a more complex pattern (Fig. 2). Pest

TABLE 1. Comparison of importance values (IV) of continuous variables as predictors of pest vs. rare status and widespread vs. small range.

Predictor variable	IV	Predictor variable	IV
<b>Pest species</b>		<b>Rare species</b>	
$\log_{10}$ (seed mass)	26.3	$\log_{10}$ (seed mass)	29.2
$\log_{10}$ (highest chromosome number)	6.3	$\log_{10}$ (lowest chromosome number)	4.5
$\log_{10}$ (lowest chromosome number)	5.5	ploidy ratio (high)	4.4
ploidy ratio (high)	3.2	ploidy ratio (low)	4.2
ploidy ratio (low)	3.0	$\log_{10}$ (highest chromosome number)	2.5
<b>Widespread range</b>		<b>Small range</b>	
$\log_{10}$ (seed mass)	22.2	$\log_{10}$ (seed mass)	18.8
ploidy ratio (high)	5	ploidy ratio (high)	5.7
$\log_{10}$ (highest chromosome number)	2.9	ploidy ratio (low)	2.8
$\log_{10}$ (lowest chromosome number)	2.8	$\log_{10}$ (lowest chromosome number)	2.6
ploidy ratio (low)	2.4	$\log_{10}$ (highest chromosome number)	2.2

Note: Importance of covariates was calculated by randomly permuting each predictor variable at a time and computing the associated reduction in predictive performance.

TABLE 2. Comparison of categorical variables as predictors of pest vs. rare species. The relationship is given as positive (+) or negative (-).

Variable, by trait type	Relationship, by species status			Relationship, by range		
	Pest	Rare	Opposite	Widespread	Small range	Opposite
Life history						
Annual	+	+		-†	+	X
Biennial	+	+		+	-†	X
Perennial	-	-		+†	-	X
Multiple life histories	+	-	X	+	-	X
Growth form						
Forb	-	+	X	+†	-†	X
Graminoid	+	-	X	+	-	X
Subshrub	+	-	X	-†	+†	X
Shrub	-	+	X	-	+	X
Vine	+	+		+	+	
Tree	-	+	X	-	-†	
Multiple growth forms	+	-	X	+	-	X
Habitat						
Facultative wetland	+	-	X	+	-	X
Obligate wetland	+	-	X	+	-	X
Pollination						
Biotic pollination	-	+	X	-	+	X
Cytology						
Multiple counts	+	-	X	+	-	X

Notes: Traits for which the relationships are opposite for pest vs. rare status or widespread vs. small range are indicated by an "X." Dagger symbols (†) mark divergences of range size relationships from those of pest/rarity status.

species were likely at low values (<12) for highest chromosome number and especially at high chromosome numbers (20–300), whereas rare species were likeliest only at low values (<35). Pest species showed a similar relationship to lowest chromosome number as to highest, with the difference that pest status was unlikely at very low chromosome numbers (<10). Rare species, however, were most likely at high chromosome counts (>50 and >100). Echoing the pattern of raw counts, rare species were likeliest at low values of ploidy ratio (1–2.5) when calculated from the highest per species chromosome count, but at high values (>2) when based on the lowest per species value. Nearly opposite was the relationship between pest species and ploidy, with pest species being most likely at low to moderate values (1–6) when based on lowest per species chromosome count, but uniform across a range of values >1 when based on highest per species counts. Thus, rare species may be species with high chromosome counts and high ploidy ratios in the case of species with only a single chromosome number reported, but are very unlikely to have high chromosome counts or ploidy ratios associated with polyploid cytotypes. In contrast, pest species are associated with low or high chromosome counts when multiple counts are reported, with high chromosome counts otherwise, and generally with high ploidy ratios.

Focusing on categorical traits (Table 2), we found that the direction of the relationship between traits and rarity or pest status was opposite for 11 of 15 traits. Annual and biennial life history traits and vine growth

form were positively associated with either class, whereas perennial life history was negative in both cases. Consistent with our predictions, pest species were positively associated with life history and growth form plasticity, with either facultative or obligate wetland affinities, with multiple reported chromosome counts, and with abiotic pollination. In contrast, rare species tended to be phenotypically and cytologically monomorphic and tended not to be associated with wetlands or to be pollinated by animals.

Results of two-way contingency analysis revealed that life history and growth form plasticity were disproportionately associated with polyploidy, but not with multiple chromosome counts, whereas facultative wetland habitat association was disproportionately associated with both polyploidy and multiple chromosome counts (Table 3). In summary, pest species are disproportionately likely to be polyploid and to be positive for the phenotypic plasticity measures associated with polyploidy, whereas rare species are likely to be diploid and to exhibit little phenotypic plasticity.

#### Range size

Although minority classes were much larger, range size classes were less predictable than rarity/pest classes in that, for the same true positive rates, false positive rates were significantly higher (Fig. 2). With the exception of life history and growth form traits (annual, biennial, perennial, forb, subshrub, tree), range size classes showed the same relationships to categorical variables (Table 2): species with small ranges, like rare

TABLE 3. Two-way contingency analysis of species investigating the relationship between polyploidy and multiple reported cytological counts of chromosomes and three measures of phenotypic plasticity.

Predictor	Multiple life histories?				Multiple growth forms?				Facultative wetland?			
	Yes	No	$\chi^2$	<i>P</i>	Yes	No	$\chi^2$	<i>P</i>	Yes	No	$\chi^2$	<i>P</i>
Polyploidy			5	<b>0.024</b>			18	<b>0.0001</b>			1012	<b>0.0001</b>
Yes	195	1919			150	2347			776	1199		
No	62	833			36	1199			300	2347		
Multiple chromosome counts			2	0.184			0	0.661			6	<b>0.017</b>
Yes	25	185			30	180			72	138		
No	232	2310			392	2150			677	1865		

Notes: For all comparisons,  $df = 1$ . The three measures of phenotypic plasticity are: multiple vs. single known life history, multiple vs. single known growth form, and whether classed as facultative wetland in habitat association. Data on chromosome numbers are from the Missouri Gardens Index of Plant Chromosome Numbers. Phenotypic data are from the Plants National Database. Mantel test significance values  $>0.05$  within the 99% confidence interval are in boldface.

species, are unlikely to be associated with wetlands, and are likely to have multiple chromosome counts, multiple growth forms, and to be abiotically pollinated, whereas widespread species tend toward the opposite pattern. Relationships to continuous traits differed somewhat. Although seed mass was a key predictor in all cases (Table 1), the ploidy ratio of highest (species) chromosome number to genus minimum was a much more important predictor of either small or widespread ranges than of rare or pest status (see Plate 1).

Although general trends in the relationships between chromosome number and ploidy were similar, the relationships between seed mass and range size classes differed in important ways from relationships between seed mass and rare or pest status. In contrast to the relationship in the case of pest species, the likelihood that a species is widespread decreases with seed mass, such that small-seeded species ( $<0.01$ mg) are most likely to be widespread. Although it is skewed to the right, as in the case of rarity, the likelihood that a species has a small range increases with seed mass in several apparent thresholds. Extremely small-seeded species are unlikely to have small ranges, and extremely large species are unlikely to be widespread. Thus, in our data, rare species were much more likely to have large seeds than the set of species with small ranges of which they form a subset. Similarly, widespread species were more likely to have small seeds than the subset of widespread species classed as pest species.

#### Geography

Because over one-third of the rare species in our data set were Hawaiian endemics (for which patterns might plausibly be expected to differ), we analyzed the rare species data without including Hawaiian species. Predictive performance was lower (Appendix: Table A1, Fig. A1), but key predictors and trait patterns were similar. Three families in particular (Gesneriaceae, Campanulaceae, and Palmae) exhibited especially high proportions of rare species relative to other large families, reflecting the large representation of these families in the Hawaiian flora.

#### Phylogeny

The inclusion of phylogenetic and taxonomic variables at the family level and above did not affect accuracy when models were tested on the holdout test set. Major predictors and the relationships of predictors to responses remained largely unchanged. Both rarity ( $K = 0.672$ ,  $P = 0.483$ ) and pest status ( $K = 0.847$ ,  $P = 0.073$ ) showed values of  $K$  indicating partial phylogenetic dependence, although in neither case was the observed signal statistically significant.

#### DISCUSSION

Our results suggest that rarity is associated with either low or high chromosome numbers, but not with multiple chromosome numbers or low ploidy. Rare species are also species with large seed masses, are generally animal-pollinated, and are not strongly associated with wetlands. In contrast, pest species are likely to be polyploid, have multiple and high chromosome numbers, tend toward the mid-range of seed masses, tend to be abiotically pollinated, and tend to be associated with wetlands. One of the most important predictors of either rare or pest status was seed mass, which is strongly correlated with variation in seed dispersal and seedling establishment among species (Leishman et al. 2000). The importance of seed mass in predicting rarity and pest status very likely reflects the role of dispersal and establishment in the invasion process and in limiting population growth, particularly in species that have undergone habitat loss and fragmentation. Seeds larger than 100 mg are generally vertebrate-dispersed, and in some floras the largest seeds are dispersed by mammals rather than birds (Leishman et al. 2000). In our study, large-seeded ( $>1.5$  g) species were much more likely to be rare, perhaps due to dependence on large-mammal herbivores, now extirpated or extinct, for dispersal. Species at the center and to the right of the size distribution (between  $\sim 0.1$  mg and 100 mg), where seeds are small enough to be dispersed by vertebrates and yet large enough to be well-provisioned, appear to have the greatest probability of becoming pests. Species with

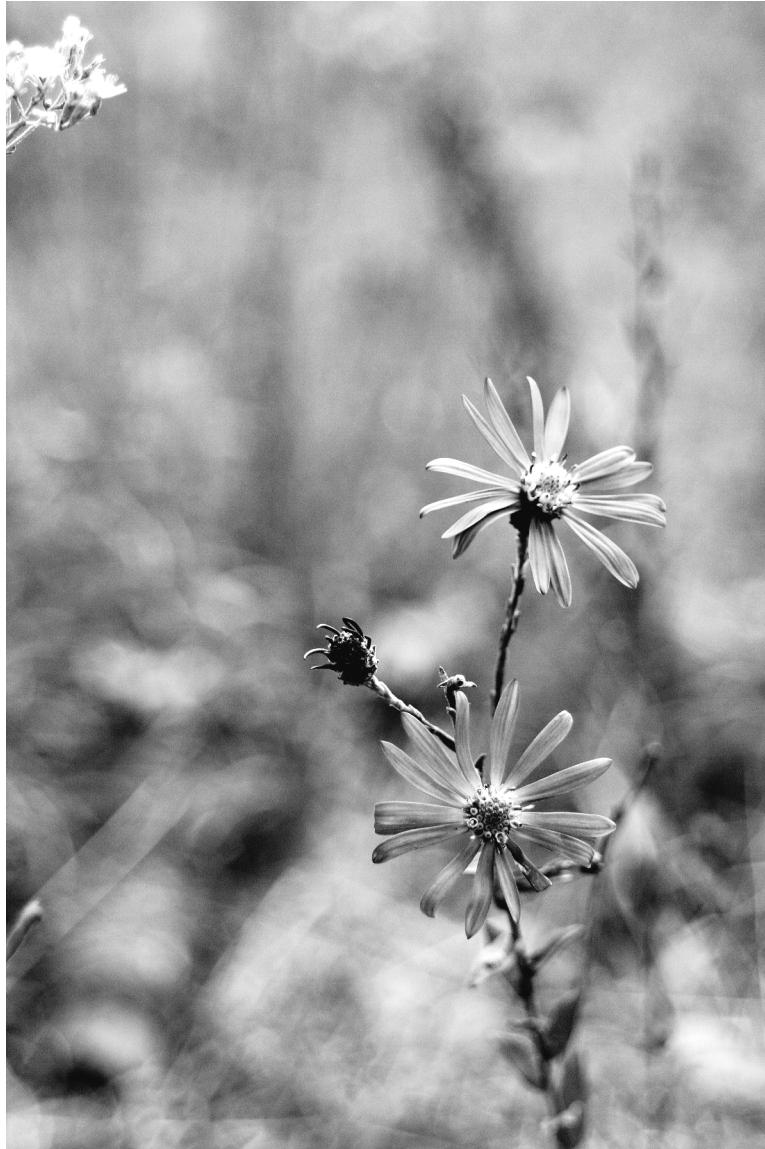


PLATE 1. Georgia aster (*Symphyotrichum georgianum*), pictured, a rare aster (ranked as G3, “vulnerable”, by NatureServe) found at only a relatively few locations in the southeastern U.S. is a decaploid ( $2n = 50$ ), whereas late purple aster (*Symphyotrichum patens*), a tetraploid, is common and ranges across the eastern and midwestern states. Photo credit: J. P. Schmidt.

seeds at the low of end of the size distribution ( $<5$  mg) are unlikely to be either rare or pests. If small-seeded species are effectively dispersed over distances separating suitable habitat, as they appear to be from results of the range size analysis, one predicts that, therefore, they would be less likely to become rare. We speculate that because small seeds are poorly provisioned, small-seeded species are unlikely to be strong competitors (especially in shaded environments) and are less likely to establish populations that will multiply to pest densities. In addition to dispersal and seedling survival, seed mass has been shown to be correlated across floras with plant growth form and height (Leishman et al. 2000). This single variable appears to be particularly information

dense, and was shown to independently predict well which species introduced to Hawaii have become noxious (Schmidt and Drake 2011a).

Trait patterns of widespread vs. narrowly distributed species were generally congruent with those of rare vs. pest species, but diverged in several interesting ways. First, ploidy and chromosome number were much less important as predictors of range-size categories, suggesting that, although large range size is associated with polyploidy, it is the subset of species with large ranges that become economic pests and are most likely to be polyploid. Second, widespread species and narrowly distributed species were probable across a much larger range of seed masses than were rare or pest species (Fig.

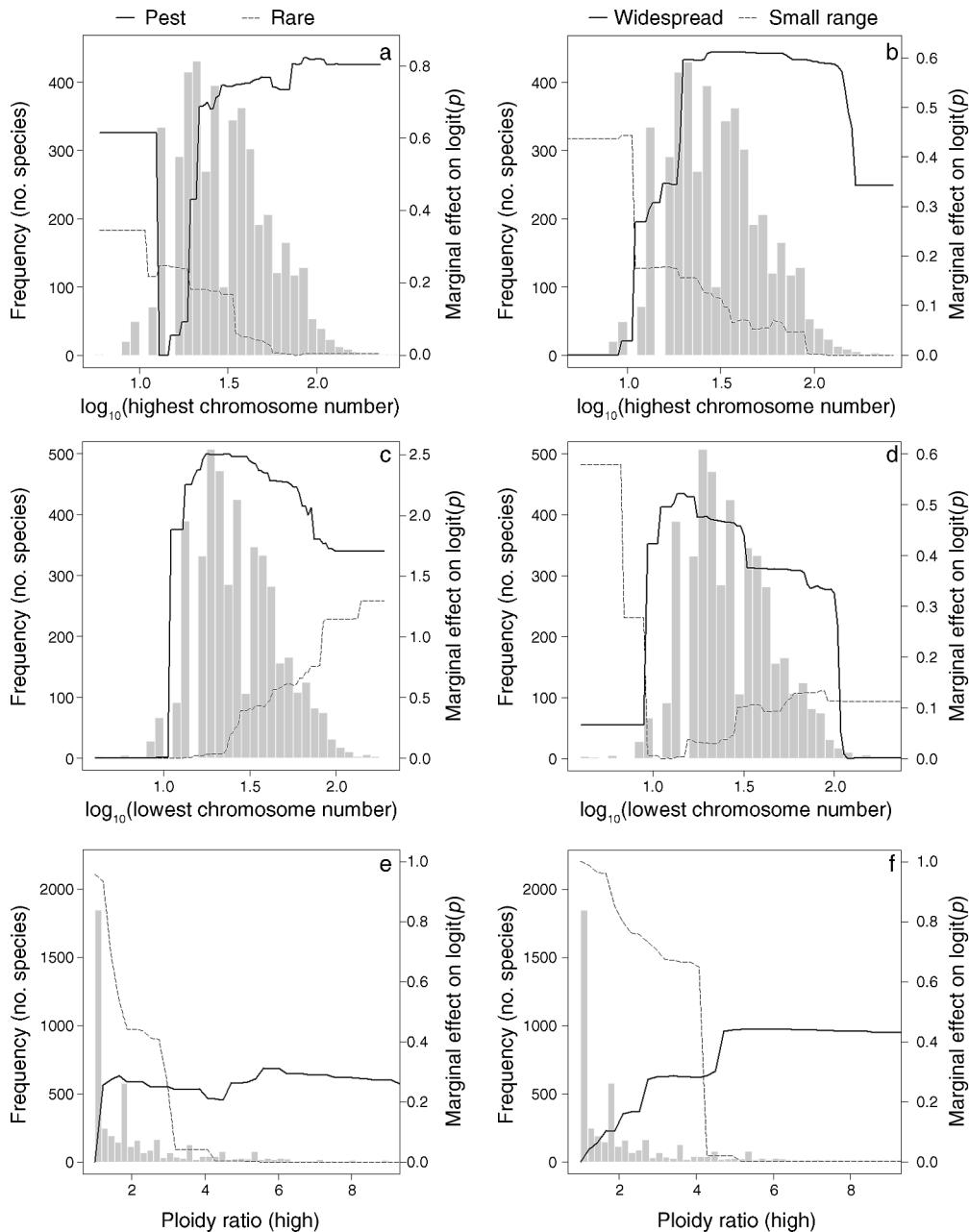


FIG. 3. Plots showing the improvement of “gbm” models (in R) as a function of a single predictor (Friedman 2001). Likelihood of rare or pest status (left column) and widespread or small range (right column) is shown as a function of (a, b) highest raw chromosome number; (c, d) lowest raw chromosome number; (e, f) high ploidy ratio (chromosome number of species/genus minimum chromosome number); (g, h) low ploidy ratio; and (i, j) effect of seed mass. Functional values (logit scale, right-hand y-axis:  $\log(p/1-p)$  where  $p$  is the probability of the status, pest vs. rare or widespread vs. small range) were standardized by shifting the lowest value to 0.

3i, j). Likelihood of a large range declined, as a general trend, with seed mass, whereas narrowly distributed species showed a bimodal relationship that included both large- and small-seeded species. Thus, with respect to seed mass (the most important predictor), rare and pest species appear to form restricted subsets of species with either small or large ranges.

### Ploidy

After seed mass, chromosome number and ploidy were together the most important predictors of either rarity or pest status. The relationship of pest status and rarity to ploidy agrees with the finding of Pandit et al. (2011) who, from global compilations, compared a set of rare and a set of invasive exotic species to congeners.

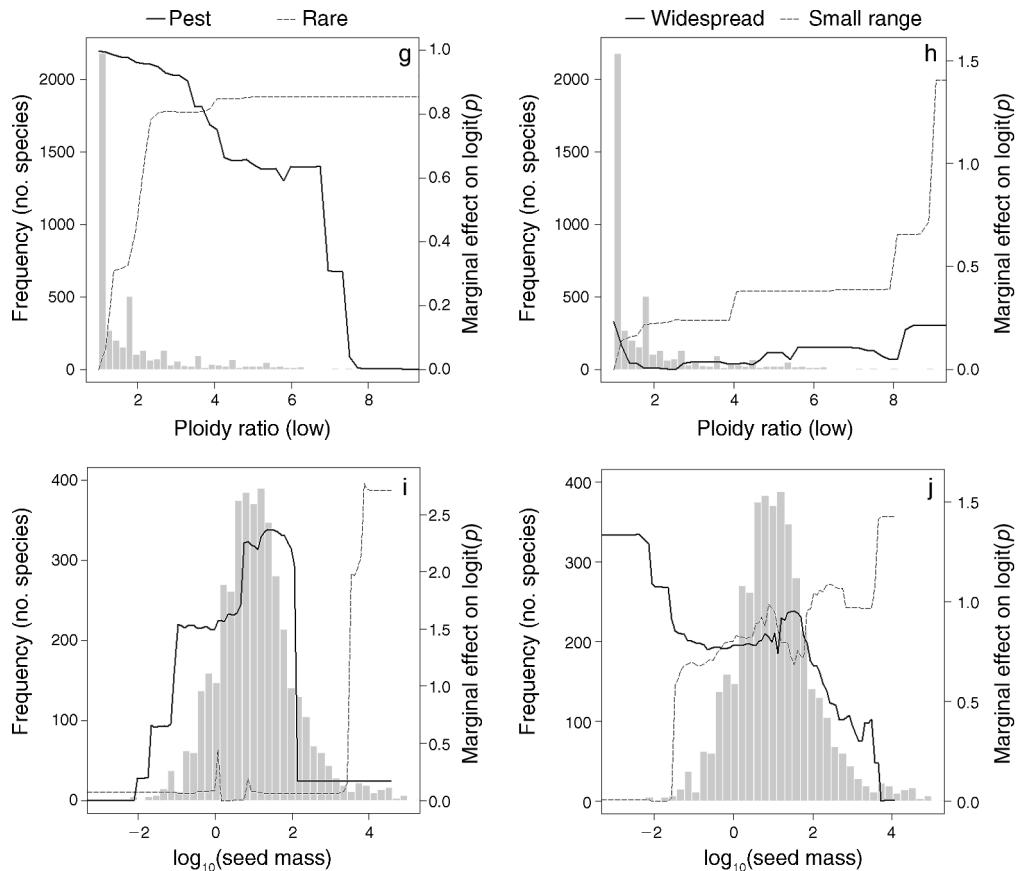


FIG. 3. Continued.

However, our results differed in that we also found raw chromosome number to be an important predictor. In our study, interestingly, pest and widespread species shared a tendency to low chromosome counts and high ploidy (or the capacity for high ploidy), whereas rare species and species with restricted ranges tended toward low or high chromosome counts but low ploidy. If this pattern proves to be a general one, it suggests that patterns of plant species abundance and distribution may well be causally related to genome multiplication.

However, whether polyploidy is a cause or an effect of rarity or pest status, and whether the more important factor in the success of pest species is the ability of a taxonomic species to form multiple cytotypes, or the polyploid cytotypes themselves, remain open questions. In the case of pest species, polyploidy may be a response to selection in novel or harsh environments (Ramsey and Schemske 1998), or, as Pandit et al. (2011) suggest, the large number of reproductive events presumably characteristic of pest species invasions increases the likelihood that genome duplication events will occur.

The existence of many species known to occur as both diploids and polyploids in their native ranges, but most commonly as polyploids in introduced ranges, e.g., *Solidago gigantea* (Schlaepfer et al. 2010a), *Centaurea*

*stoebe* (Henery et al. 2010), *Centaurea maculosa* (Trier et al. 2009), *Hypericum perforatum* (Maron et al. 2004), *Senecio inaequidens* (Lafuma et al. 2003), and *Lythrum salicaria* (Kubátová et al. 2008), supports the commonly held belief that adaptive advantages are present prior to introduction (but see Mandák et al. 2003). Moreover, the number of ploidy levels in the native range was a predictor of invasion success among Central European plants introduced to other regions (Pyšek et al. 2009). However, the few studies that have compared diploid and polyploid cytotypes for traits that might affect fitness in the species' native ranges have yielded mixed results. Treier et al. (2009), for *C. maculosa*; Schlaepfer et al. (2010b), for 14 congeneric pairs of herbaceous species; and Thebault et al. (2011), for *C. stoebe* and *S. inaequidens*, found support for preadaptation by polyploids in the form of tolerance of drier conditions, poly- rather than monocarpic life history, increased clonal growth, and higher leaf specific area and stem height, whereas Hull-Sanders et al. (2009) for *Solidago gigantea* found little difference between cytotypes in physiology, performance, or secondary chemistry, and Kubátová et al. (2008) found little difference in growth characteristics.

Results from studies comparing the performance of pest and non-pest congeners (van Kleunen et al. 2007, Schlaepfer et al. 2010b) without considering ploidy level provide further support for preadaptation. In addition, many well-known pest plant taxa are polyploids resulting from hybridization events, e.g., *Tragopogon*, *Spartina*, *Senecio*. Indeed, in some cases taxonomic groups once considered to be (auto)polyploid subspecies have been shown subsequently to be allopolyploid hybrids (e.g., Vilatersana et al. 2007). Both allopolyploids and autopolyploids frequently exhibit larger range sizes than closely related diploid taxa, e.g., in the genus *Clarkia* (Lowry and Lester 2006), and often thrive where parental species cannot, in harsh conditions at higher altitudes and latitudes and in low-nutrient or drought-prone environments (Levin 2003). Accordingly, we suggest that polyploidy may generally be a causal factor rather than the result of invasion success, and that taxonomic complexes that generate polyploid taxa are more likely to give rise to species successful enough to become pests or dominants within native or introduced ranges. In this view, we find support from the conclusions of a comprehensive review by te Beest et al. (2012), who also conclude that polyploidy can be an important factor in plant invasions through an increased ability to adapt to novel conditions made possible by greater genetic diversity, thereby assisting the “evolution of invasiveness.” Alternatively, polyploidy may foster invasiveness either by restoring sexual reproduction following hybridization or by promoting asexual reproduction in the absence of suitable mates.

As an additional factor related to invasiveness, we showed that polyploid taxa were disproportionately likely to exhibit phenotypic plasticity in the form of multiple growth forms, multiple life histories, or broad ecological tolerances. Although it is unclear whether these findings actually reflect trait plasticity within individuals or rather the evolutionary potential of polyploids that leads to multiple genotypes adapted to multiple conditions, invasion success, particularly in disturbed environments, appears to be strongly linked to trait variability, and thus, at least indirectly, to polyploidy.

#### *Phylogenetic signal*

The inclusion of phylogenetic and taxonomic variables along with traits as predictors did not improve prediction of rarity or pest status, and we found little evidence of phylogenetic trends in pest status and rarity at the family level. Because we lacked phylogenetic information within families, our analyses cannot rule out the influence of phylogeny at lower taxonomic levels. Indeed, previous work suggests that within some families particular genera disproportionately contain pest species (Pandit et al. 2011, Schmidt and Drake 2011b) and rare species (Pandit et al. 2011). However, although phylogenetic factors within genera, families,

and higher taxonomic groupings may constrain seed mass, wetland adaptation, chromosome counts, and the degree of polyploidization, it is not clear that this would limit the generality of our findings, given that neither rarity nor pest status shows statistically significant phylogenetic signal.

#### *Predictability of rare/pest and small/broad ranges*

Overall, the level of accuracy in predicting pest species was high, but was only moderate for rare species. Bias in data availability may have significantly reduced our ability to predict rarity. For example, chromosome count data were available for only 162 and seed mass data for only 30 of 1258 rare species, but for 286 and 198, respectively, of 352 pest species. The difference in accuracy may also, of course, indicate that rarity is actually related to traits unknown to us and is not correlated with candidate predictors included in our analysis. Additionally, pest species designations may be a more reliable and, therefore, predictable category because there is less ambiguity in whether a species is a nuisance or not. By contrast, assigning all species within the North American flora to a rarity class necessarily required judgments that frequently must be made using meager information on the status and extent of populations. Furthermore, rarity ranks are a composite of many factors and they group potentially distinct classes of rare plants that may have different relationships to traits. Although this is also true of pest status ranks, we suggest that the ecological, geographic, or evolutionary causes of rarity may be more disparate than the reasons for considering a species to be a pest. For instance, consider the effect of commonness and rarity on a species distribution. All economic pests will exhibit high densities and rates of spread at some localities or habitats. Yet, among rare species, narrow endemics may be either locally abundant but sparsely distributed in a landscape, or relatively wide-ranging but never reaching high densities locally (Rabinowitz 1981), whereas still other species are highly dependent on disturbance regimes such as periodic fire or flooding, and are therefore less frequent or absent in the current landscape than under historical conditions.

Reflective of these patterns among rare and pest classes, classification performance was also better for widespread than for narrowly distributed species. Data limitations may again be implicated in the disparity. In addition, large rather than small ranges may be better resolved by our state-based distributions: ranges that cover at most of a single large state (e.g., California, Texas) should not count as small. Overall, predictability of either range size class was lower than that of either rare or pest classes. This reduced predictability may imply that rare and pest classes are more cohesive labels than those based on range size alone, or may simply reflect the coarse resolution of range sizes and the cutoffs that we used to define large and small ranges.

*Historical and geographic factors affecting rarity*

An additional issue is that rarity may be more strongly affected by historical factors that we were not able to include as predictors. The degree to which human alteration of landscapes coincides with the range or habitat of a species is likely to be a key determinant of rarity status (Kunin and Gaston 1993, Duncan and Young 2000, Murray et al. 2002). Regions with high floristic diversity or endemism (e.g., Hawaii, California) and high human population densities account for a larger share of rare species. Of the rare species included in this study, 84.9% are restricted to a single state or province. Hawaii, unsurprisingly, accounts for 35.9% of the total number of rare species in the data set, and 48.7% of native Hawaiian species are rare. Although 16.7% of all rare species are restricted to California (area = 423 970 km<sup>2</sup>, human population in 2010 = 37 million), which has both high plant diversity and endemism and high human population density, only 15.2% of California endemics are classed as rare. In Florida (area = 170 304 km<sup>2</sup>, population in 2010 = 18.5 million), however, 25.3% of species endemic to the state are rare, but only 9.2% of angiosperms species endemic to Texas (area = 696 241 km<sup>2</sup>, population in 2010 = 25 million) are rare. Of these four states, Texas has the least number of endemics, the largest area, and the lowest population density. In the same way, habitats hosting endemic species or ecosystems with high diversity that have been particularly subject to destruction (e.g., granite outcrops, bogs, and longleaf pine savannas in the Southeast) can be expected to generate more rare species. The importance of historical factors (e.g., time since introduction, propagule pressure) in explaining the establishment and naturalization of introduced species has been demonstrated (see reviews in Cadotte et al. 2006, Pyšek and Richardson 2007, Pemberton and Liu 2009). Historical factors may be similarly important in predicting which native species have become most threatened and which have become pests.

*Comparisons to other tests of the "two-sides-of-the-same-coin" hypothesis across taxa*

Our results rather strongly support the notion that rare and pest species can be predicted by the same set of traits and that their relationship to a set of categorical traits contrasts diametrically. Although rare and pest species do not appear to differ greatly in life history traits (annual, biennial, perennial), rare species are more likely to be plastic in regard to life history, growth form, and habitat affinity. With respect to ploidy and seed mass, patterns contrast sharply, suggesting that (1) rare plants are likely to be dispersal-limited and pest plants typically possess readily dispersible, but well-provisioned seeds, and (2) rare plants tend toward low or, to a much lesser degree, very high ploidy, whereas pest plants tend to be polyploid, although not in the extreme. Further, polyploidy appears to be related to plasticity in traits. Broadly, then, our results are congruent with two

of the few studies explicitly testing the "two-sides-of-the-same-coin" hypothesis. Comparisons in these studies were slightly different than ours in that the contrast was rare vs. pest anywhere, rather than pest within, or adjacent to, putative native range. Yet, rare species of legumes (Bradshaw et al. 2008) and crayfish (Larson and Olden 2010) tend to be habitat specialists, whereas pest species tend to be habitat generalists. At odds with these results are those of Blackburn and Jeschke (2009) and Jeschke and Strayer (2008), who tested the "two-sides-of-the-same-coin" hypothesis using data on invasion success, extinction risk, and traits relating to life history and association with humans across all European and North American species of birds and freshwater fish. Of the predictors that those authors used, none, aside from native range size, was an important predictor of both threat and invasive status. Contrary to their predictions, invasive fish species had smaller native ranges than noninvasive introduced species, and diet breadth was not predictive of threat status among bird species. Indeed, although large, slow-to-reproduce bird species were more likely to be imperiled, and number of clutches per year was negatively related to threat status in fish, other reproductive variables were not significant predictors of invasion success. Instead, key predictors of invasion success appeared to be related to human affiliation and propagule pressure rather than to life history traits. Thus, a specialist-generalist opposition was apparent in three studies, but absent in two others.

What accounts for these contrasting results? We suggest that a key difference between studies finding support for the two-sides-of-the-coin hypothesis is the degree to which human decisions and actions determine which species were introduced between continents and with how much effort. Establishing introduced populations of birds and fish appears (not surprisingly) to require substantial human intervention. The propagation effort necessary to establish introduced populations of invertebrates or plants may be much lower. Further, invasion success in Jeschke and Strayer (2008) was better predicted than threat status, leading the authors to conclude, as we did, that habitat destruction and degree of human disturbance, factors that they were unable to include in the analysis, were more important determinants of threat than pest status. Thus, controlling for anthropogenic factors would improve prediction and potentially resolve apparent contradictions between findings across taxa.

*Conclusions*

We assembled and analyzed data for the entire native flora of the continental United States and Canada to develop predictive models of rarity and pest status as a function of traits. Our success, despite data limitations, indicates that better prediction of rare and pest status than has been presumed possible (Lodge 1993), particularly for rarity, is achievable. Thus, our results have important implications for conservation planning and

biosecurity. We have also identified patterns related to seed mass, chromosome number, and ploidy that are congruent with previous results (Pandit et al. 2011, Schmidt and Drake 2011a, b), and support the idea that polyploidy and hybridization (or their absence) are closely linked to functional traits that can promote either vulnerability or pest status.

## ACKNOWLEDGMENTS

We thank Kevin Samples and Krisztian Magori for technical assistance, and Rich Shefferson and two anonymous reviewers for many helpful comments. This work was supported by a cooperative agreement with the USDA-ERS Program of Research on the Economics of Invasive Species (Cooperative Agreement No. 58-7000-8-0111).

## LITERATURE CITED

- Bennett, M. D., and I. J. Leitch. 2005. Plant DNA C-values database (release 4.0). Royal Botanic Gardens, Kew, Richmond, Surrey, UK. <http://www.kew.org.uk/cvalues/>
- Bevill, R. L., and S. M. Louda. 1999. Comparisons of related rare and common species in the study of plant rarity. *Conservation Biology* 13:493–498.
- Blackburn, T. M., and J. M. Jeschke. 2009. Invasion success and threat status: two sides of a different coin? *Ecography* 32:83–88.
- Blomberg, S. P., and T. Garland, Jr. 2002. Tempo and mode in evolution: phylogenetic inertia, adaptation and comparative methods. *Journal of Evolutionary Biology* 15:899–910.
- Blomberg, S. P., T. Garland, Jr., and A. R. Ives. 2003. Testing for phylogenetic signal in comparative data: behavioral traits are more labile. *Evolution* 57:717–745.
- Bradshaw, C. J. A., X. L. Giam, H. T. W. Tan, B. W. Brook, and N. S. Sodhi. 2008. Threat or invasive status in legumes is related to opposite extremes of the same ecological and life-history attributes. *Journal of Ecology* 96:869–883.
- Cadotte, M. W., B. R. Murray, and J. Lovett-Doust. 2006. Ecological patterns and biological invasions, using regional species inventories in macroecology. *Biological Invasions* 8:809–821.
- Davidson, A. M., M. Jennions, and A. B. Nicotra. 2011. Do invasive species show higher phenotypic plasticity than native species and, if so, is it adaptive? A meta-analysis. *Ecology Letters* 14:981–993.
- Davies, T. J., T. G. Barraclough, M. W. Chase, P. S. Soltis, D. E. Soltis, and V. Savolainen. 2004. Darwin's abominable mystery: Insights from a supertree of the angiosperms. *Proceedings of the National Academy of Sciences USA* 101:1904–1909.
- Duncan, R. P., and J. R. Young. 2000. Determinants of plant extinction and rarity 145 years after European settlement of Auckland, New Zealand. *Ecology* 81:3048–3061.
- Elith, J., J. R. Leathwick, and T. Hastie. 2008. A working guide to boosted regression trees. *Journal of Animal Ecology* 77:802–813.
- Ellstrand, N. C., and K. A. Schierenbeck. 2000. Hybridization as a stimulus for the evolution of invasiveness in plants? *Proceedings of the National Academy of Sciences USA* 97:7043–7050.
- Farnsworth, E. J. 2007. Plant life history traits of rare versus frequent plant taxa of sandplains: implications for research and management trials. *Biological Conservation* 136:44–52.
- Farnsworth, E. J., and E. J. Ogurcak. 2008. Functional groups of rare plants differ in levels of imperilment. *American Journal of Botany* 95:943–953.
- Friedman, J. H. 2001. Greedy function approximation: a gradient boosting machine. *Annals of Statistics* 29:1189–1232.
- Gaston, K. F., and W. E. Kunin. 1997. Rare–common differences, an overview. Pages 12–29 in W. E. Kunin and K. J. Gaston, editors. *The biology of rarity: causes and consequences of rare–common differences*. Chapman and Hall, London, UK.
- Hartley, S., and W. E. Kunin. 2003. Scale dependency of rarity, extinction risk, and conservation priority. *Conservation Biology* 17:1559–1570.
- Henery, M. L., G. Bowman, P. Mráz, U. A. Treier, E. Gex-Fabry, U. Schafner, and H. Müller-Schärer. 2010. Evidence for a combination of pre-adapted traits and rapid adaptive change in the invasive plant *Centaurea stoebe*. *Journal of Ecology* 98:800–813.
- Hull-Sanders, H. M., R. H. Johnson, H. A. Owen, and G. A. Meyer. 2009. Effects of polyploidy on secondary chemistry, physiology, and performance on native and invasive genotypes of *Solidago gigantea* (Asteraceae). *American Journal of Botany* 96:762–770.
- Jeschke, J. M., and D. L. Strayer. 2008. Are threat status and invasion success two sides of the same coin? *Ecography* 31:124–130.
- Kembel, S. W., D. D. Ackerly, S. P. Blomberg, W. K. Cornwell, P. D. Cowan, M. R. Helmus, H. Morlon, and C. O. Webb. 2010. Picante: R tools for integrating phylogenies and ecology v 1.2-0. *Bioinformatics* 26:1463–1464. <http://cran.r-project.org/web/packages/picante/>
- Kolb, A., F. Barsch, and M. Diekmann. 2006. Determinants of local abundance and range size in forest vascular plants. *Global Ecology and Biogeography* 15:237–247.
- Kubátová, B., P. Trávníček, D. Bastlová, V. Čurn, V. Jarolímová, and J. Suda. 2008. DNA ploidy-level variation in native and invasive populations of *Lythrum salicaria* at a large geographic scale. *Journal of Biogeography* 35:167–176.
- Kunin, W. E., and K. J. Gaston. 1993. The biology of rarity: patterns, causes, and consequences. *Trends in Ecology and Evolution* 8:298–301.
- Lafuma, L., K. Balkwill, E. Imbert, R. Verlaque, and S. Maurice. 2003. Ploidy level and origin of the European invasive weed *Senecio inaequidens* (Asteraceae). *Plant Systematics and Evolution* 243:59–72.
- Larson, E. R., and J. D. Olden. 2010. Latent extinction and invasion risk of crayfishes in the southeastern United States. *Conservation Biology* 24:1099–1110.
- Lee, C. E. 2002. Evolutionary genetics of invasive species. *Trends in Ecology and Evolution* 17:386–391.
- Leishman, M. R., V. P. Thomson, and J. Cooke. 2010. Native and exotic plants have fundamentally similar carbon capture strategies. *Journal of Ecology* 98:28–42.
- Leishman, M. R., I. J. Wright, A. T. Moles, and M. Westoby. 2000. The evolutionary ecology of seed size. Pages 31–57 in M. Fenner, editor. *Seeds: the ecology of regeneration in plant communities*. Second edition. CAB [Center for Agricultural Bioscience] International, Wallingford, UK.
- Levin, D. A. 2003. The ecological transition in speciation. *New Phytologist* 161:91–96.
- Lodge, D. M. 1993. Biological invasions: lessons for ecology. *Trends in Ecology and Evolution* 8:133–137.
- Lowry, E., and S. E. Lester. 2006. The biogeography of plant reproduction: potential determinants of species' range sizes. *Journal of Biogeography* 33:1975–1982.
- Mandák, B., P. Pyšek, M. Lysák, J. Suda, A. Krahulcová, and K. Bímová. 2003. Variation in DNA-ploidy levels of *Reynoutria* taxa in the Czech Republic. *Annals of Botany* 92:265–272.
- Maron, J. L., M. Vila, R. Bommarco, S. Elmendorf, and P. Beardsley. 2004. Rapid evolution of an invasive plant. *Ecological Monographs* 74:261–280.
- Meiners, S. J. 2007. Native and exotic plant species exhibit similar population dynamics during succession. *Ecology* 88:1098–1104.

- Murray, B. R., P. H. Thrall, A. M. Gill, and A. B. Nicotra. 2002. How plant life-history and ecological traits relate to species rarity and commonness at varying spatial scales. *Austral Ecology* 27:291–310.
- NatureServe. 2010. NatureServe's central databases. NatureServe, Arlington, Virginia, USA.
- Palacio-Lopez, K., and E. Gianoli. 2011. Invasive plants do not display greater phenotypic plasticity than their native or non-invasive counterparts: a meta-analysis. *Oikos* 120:1393–1401.
- Pandit, M. K. 2006. Continuing the search for pattern among rare plants: Are diploid species likely to be rare? *Evolutionary Ecology Research* 8:543–552.
- Pandit, M. K., M. J. O. Poccock, and W. E. Kunin. 2011. Ploidy influences rarity and invasiveness in plants. *Journal of Ecology* 99:1108–1115.
- Paradis, E. 2006. Analysis of phylogenetics and evolution with R. Springer Science+Business Media, Dordrecht, The Netherlands.
- Pemberton, R. W., and H. Liu. 2009. Marketing time predicts naturalization of horticultural plants. *Ecology* 90:69–80.
- Pilgrim, E. S., M. Crawley, and K. Dolphin. 2004. Patterns of rarity in the native British flora. *Biological Conservation* 120:161–170.
- Pyšek, P., V. Jarošík, J. Pergl, R. Randall, M. Chytrý, I. Kühn, L. Tichý, J. Danihelka, J. Chrtek, Jr., and J. Sádlo. 2009. The global invasion of Central European plants is related to distribution characteristics in their native range and species traits. *Diversity and Distributions* 15:891–903.
- Pyšek, P., and D. M. Richardson. 2007. Traits associated with invasiveness in alien plants: where do we stand? *Biological Invasions* 193:97–125.
- Rabinowitz, D. 1981. Seven forms of rarity. Pages 205–217 in H. Synge and J. Chichester, editors. *The biological aspects of rare plant conservation*. John Wiley, New York, New York, USA.
- Ramsey, J., and D. W. Schemske. 1998. Pathways, mechanisms and rates of polyploid formation in flowering plants. *Annual Review of Ecology and Systematics* 29:467–501.
- Ridgeway, G. 1999. The state of boosting. *Computer Science and Statistics* 31:172–181.
- Ridgeway, G. 2006. Generalized boosted regression models. Documentation on the R Package “gbm,” version 1.5-7. <http://cran.r-project.org/web/packages/gbm/>
- Schlaepfer, D. R., P. J. Edwards, and R. Billeter. 2010. Why only tetraploid *Solidago gigantea* (Asteraceae) became invasive: a common garden comparison of ploidy levels. *Oecologia* 163:661–673.
- Schlaepfer, D. R., M. Glaetli, M. Fischer, and M. van Kleunen. 2010. A multi-species experiment in their native range indicates pre-adaptation of invasive alien plant species. *New Phytologist* 185:1087–1099.
- Schmidt, J. P., and J. M. Drake. 2011a. Time since introduction, seed mass, and genome size predict successful invaders among the cultivated vascular plants of Hawaii. *PLoS ONE*:e17391.
- Schmidt, J. P., and J. M. Drake. 2011b. Why are some plant genera more invasive than others? *PLoS ONE* 6(4):e18654.
- Soltis, D. E., J. A. Buggs, J. J. Doyle, and S. P. Soltis. 2010. What we still don't know about polyploidy. *Taxon* 59:1387–1403.
- te Beest, M., J. J. Le Roux, D. M. Richardson, A. K. Brysting, J. Suda, M. Kubešová, and P. Pyšek. 2012. The more the better? The role of polyploidy in facilitating plant invasions. *Annals of Botany* 109:19–45.
- Thebault, A., F. Gillet, H. Muller-Scharer, and A. Buttler. 2011. Polyploidy and invasion success: trait trade-offs in native and introduced cytotypes of two Asteraceae. *Plant Ecology* 2:315–325.
- Thompson, K., and M. A. Davis. 2011. Why research on traits of invasive plants tells us very little. *Trends in Ecology and Evolution* 26:155–156.
- Thompson, K., J. G. Hodgson, and T. C. G. Rich. 1995. Native and alien invasive plants: more of the same? *Ecography* 18:390–402.
- Treier, U. A., O. Broennimann, S. Normand, A. Guisan, U. Schaffner, T. Steinger, and H. Muller-Scharer. 2009. Shift in cytotype frequency and niche space in the invasive plant *Centaurea maculosa*. *Ecology* 5:1366–1377.
- USDA NRCS [Natural Resources Conservation Service]. 2012. The PLANTS Database. National Plant Data Team, Greensboro, North Carolina, USA. <http://plants.usda.gov>
- Valery, L., H. Fritz, J. C. Lefeuvre, and D. Simberloff. 2008. In search of a real definition of the biological invasion phenomenon itself. *Biological Invasions* 10:1345–1351.
- Valery, L., F. Herve, J. C. Lefeuvre, and D. Simberloff. 2009. Invasive species can also be native. *Trends in Ecology and Evolution* 24:585–585.
- van Kleunen, M., S. D. Johnson, and M. Fischer. 2007. Predicting naturalization of southern African Iridaceae in other regions. *Journal of Applied Ecology* 44:594–603.
- van Kleunen, M., and D. M. Richardson. 2007. Invasion biology and conservation biology: time to join forces to explore the links between species traits and extinction risk and invasiveness. *Progress in Physical Geography* 31:447–450.
- van Kleunen, M., E. Weber, and M. Fisher. 2009. A meta-analysis of trait differences between invasive and non-invasive plant species. *Ecology Letters* 13:235–245.
- Vilatersana, R., A. K. Brysting, and C. Brochmann. 2007. Molecular evidence for hybrid origins of the invasive polyploids *Carthamus creticus* and *C. turkestanicus* (Caryophyllales, Asteraceae). *Molecular Phylogenetics and Evolution* 44:610–621.

## SUPPLEMENTAL MATERIAL

### Appendix

Background details of boosted regression tree models, and accuracy of model predictions for models with and without (1) phylogenetic data and (2) the inclusion of Hawaiian species (*Ecological Archives* A022-079-A1).