Testing Darwin's Naturalization Conundrum using phylogenetic relationships: Generalizable patterns across disparate communities?

Julienne Ng | William N. Weaver | Robert G. Laport

Abstract

Aim: Alternative hypotheses of Darwin's Naturalization Conundrum (DNC) predict that the non-native species that successfully establish within a community are those either more closely or more distantly related to the resident native species. Despite the increasing number of studies using phylogenetic data to test DNC and evaluate community assembly, it remains unknown whether phylogenetic relationships alone can be used to predict invasion susceptibility across communities differing environmentally and in disturbance history. In this study, we evaluate whether phylogenetic structure of diverse native communities predicts the occurrence of non-native species and offers insight into community assembly.

Location: Eastern United States of America.

Methods: We examine multiple communities across a north–south transect of the eastern United States to test whether non-native species richness and abundance are associated with phylogenetic diversity measures of the native community. We also test whether non-native species are consistently closely or distantly related to native species using two approaches differing in phylogenetic scale and whether this differs with ecologically successful species.

Results: Our analyses did not unambiguously resolve DNC. Non-native species richness and abundance decreased with increasing native species phylogenetic diversity. Within some communities, non-native species were significantly more closely related to native species than expected by chance, and tended to be more often closely related to a native species than that native species was to other native relatives. When considering species abundance, only one community showed that ecologically successful non-native species were closely related to resident species.

Main conclusions: Phylogenetic relationships can reveal important details about community assembly in diverse ecological settings. However, given the multifaceted nature of community assembly, phylogenetic metrics alone have limited utility as a general predictive tool for community invasion. Our study highlights a need to
1 | INTRODUCTION

The introduction, establishment and spread of non-native species can dramatically alter native community structure and the functioning of ecosystems (Chapin et al., 2000; Vitousek, Walker, Whiteaker, Mueller-Dombois, & Matson, 1987; Winter et al., 2009). Although some introductions can have a positive effect on native species biodiversity (Bartomeus, Vilà, & Santamaria, 2008; Rodríguez, 2006), more often than not, the establishment of non-native species has detrimental effects on native biodiversity, such as causing local extinctions through interspecific competition (e.g., Pile et al., 2017; Wells, de Winton, & Clayton, 1997) or by reducing pollinator visitation to native species (e.g., Brown, Mitchell, & Graham, 2002; Chittka & Schürkens, 2001). It is therefore a major goal of invasion ecology and conservation biology to identify the factors that lead to the successful establishment of non-native species. Species invasions also represent natural experiments, mimicking the natural process of community assembly, which can also serve as useful study systems to more broadly elucidate the process of community formation (Sargent, Angert, & Williams, 2017).

One key approach to predicting the success and impact of an invader is to identify the characteristics that make some communities more susceptible to harbouring a greater number of novel species than others (Rejmánek, Richardson, & Pyšek, 2013; Richardson & Pyšek, 2006; Shea & Chesson, 2002; Tilman, 2004). For example, since Darwin (1859), it has been thought that the success of an invading species can be explained by its relatedness to the resident species of a community. Darwin (1859, p. 115) cited observations made by Alphonse de Candolle and Asa Gray to suggest that non-native species are more likely to become established if closely related species are absent from the community. Often referred to as "Darwin's naturalization hypothesis" (Daehler, 2001; Rejmánek & Richardson, 1996), this prediction stemmed from the idea that non-native species that inhabit a different ecological niche from the resident species alleviate competition for available resources and are therefore more successful at establishing within a community. Darwin (1859), however, also considered the alternative hypothesis that non-native species would be more likely to establish in a community of close relatives, because the phenotypic similarity (as predicted by his theory of "descent with modification") would make the non-native species more suited to living in the same environment.

A number of studies have since tested Darwin's opposing hypotheses or "Darwin's Naturalization Conundrum" (DNC) (Diez, Sullivan, Hulme, Edwards, & Duncan, 2008). Although many early studies used taxonomy-based metrics of species relatedness following Darwin (e.g., comparing generic affinities), the increased availability of sequence data and computing power has led to a rise in the number of studies testing DNC with metrics that quantify relatedness from phylogenetic relationships (Gerhold et al., 2011; Strauss, Webb, & Salamin, 2006; Tan, Pu, Ryberg, & Jiang, 2015). These studies have revealed that the process of community assembly is more complicated than previously thought, with different studies providing support for both alternatives of DNC (reviewed in Thuiller et al., 2010; Jones, Nuismer, & Gomulkiewicz, 2013; Ma et al., 2016; Marx, Giblin, Dunwiddie, & Tank, 2016). The lack of a general pattern appears to be largely due to studies differing in temporal scale (i.e., stage of invasion) (Li et al., 2015; Ma et al., 2016), spatial scale (e.g., local versus regional) (Carboni et al., 2013; Davies, Cavender-Bares, & Deacon, 2011; Ma et al., 2016; Schaefer, Hardy, Silva, Barraclough, & Savolainen, 2011) and/or phylogenetic scale (Procheș, Wilson, Richardson, & Rejmánek, 2008; Thuiller et al., 2010). For example, a number of prior tests of DNC have focused on a single taxonomic group or clade (e.g., Park & Potter, 2015; Sandel & Tsiorogiannis, 2016; Strauss et al., 2006). Although this approach illuminates how phylogenetic relatedness influences invasion success among close relatives, it has limited utility for understanding the invasion dynamics and assembly process of a community comprising species from a wider diversity of evolutionary lineages (e.g., see Cadotte, Hamilton, & Murray, 2009; Lososová et al., 2015; Ordonez, 2014; Qian & Sandel, 2017; Schaefer et al., 2011).

The susceptibility of a community to invasion has also been hypothesized to vary with environmental differences (reviewed in Gallien & Carboni, 2017), further explaining a lack of consistent patterns in prior investigations of DNC. For example, species in colder climates are thought to be more closely related to each other than those in warmer climates due to historical processes (Latham & Ricklefs, 1993; Qian & Sandel, 2017; Wiens & Donoghue, 2004). This may have downstream effects on the phylogenetic patterns observed between successful invaders and resident species. However, previous studies have largely focused on communities representing either geographically or environmentally restricted distributions, thus limiting their utility for determining community susceptibility to the establishment of new species from a wider diversity of habitats or selective regimes. Moreover, habitat disturbance is thought to increase the success of non-native species establishment (D’Antonio, Dudley, & Mack, 1999; Elton, 1958). Variation in site-specific disturbance history or intensity may consequently obscure consistent patterns of phylogenetic relatedness between non-native and native species making it difficult to
draw general conclusions across communities. Therefore, given the potential that site-specific ecological and evolutionary histories could lead to community differences in susceptibility to non-native species, and thus the process of community assembly, it is unclear whether generalizable patterns of phylogenetic relatedness between non-native and native species would emerge across communities differing environmentally and in disturbance history.

Here, we investigate whether phylogenetic data alone offer insight into community assembly by identifying the susceptibility of diverse plant communities to non-native species establishment. We reconstruct phylogenetic relationships among co-occurring non-native and native species and test DNC across communities along a north–south transect of the United States that spans varying eco-climatic regions and disturbance histories. We focus on non-native woody perennial species and how they have altered phylogenetic patterns within woody plant communities, as they have been a particular threat to the structure and function of forests and forest ecosystems in the eastern United States (Webster, Jenkins, & Jose, 2006). Woody species also encompass a variety of different growth forms and habits, and are likely to have interactions with one another year-round, as compared to annuals where the interactions may be ephemeral. With these communities, we investigate: (a) Can the occurrence of non-native species be predicted by the phylogenetic diversity of native species?; (b) Can the presence and success of non-native species be explained by how closely or distantly related they are to the resident native species?; and (3) Can differences in phylogenetic relationships between non-native and native species across communities be explained by environmental differences? Although DNC was based on the assumption that relatedness could be used as a proxy for niche or phenotypic similarity, we specifically avoid these assumptions as the same phylogenetic patterns could arise by different processes, such as facilitation, mutualisms or evolutionary convergence (Cavender-Bares, Kozak, Fine, & Kembel, 2009; Losos, 2011; Mayfield & Levine, 2010; Münkemüller, Boucher, Thuiller, & Lavergne, 2015). Instead, we seek to identify general phylogenetic patterns within communities for insight into how the evolutionary diversification of lineages can influence more contemporary patterns of species coexistence (Gerhold, Cahill, Winter, Bartish, & Prinzing, 2015; Sargent et al., 2017) and whether phylogenetic relatedness can be used as a predictive tool for invasiveness across habitat types.

2 | METHODS

2.1 | Plant community data collection

We obtained plant community data from surveys conducted by the National Ecological Observatory Network (NEON; www.neonscience.org). We focused on seven sites established by NEON spanning a north–south transect of the eastern United States: Harvard Forest (HARV), Jones Ecological Research Center (JERC), Oak Ridge National Laboratory (ORNL), Ordway-Swisher Biological Station (OSBS), Smithsonian Conservation Biology Institute (SCBI), Smithsonian Environmental Research Center (SERC) and Talladega National Forest (TALL) (Figure 1). These sites span approximately 13° of latitude from Massachusetts to Florida, representing five NEON-defined ecoclimatic regions of the United States, each of which is distinct in vegetation, landforms, climate and ecosystem dynamics: Northeast, Mid-Atlantic, Appalachians/Cumberland Plateau, Ozarks Complex and Southeast (Figure 1). We used the NEON data portal to obtain species occurrence and abundance data of woody perennial plants sampled from multiple replicated 400-m² plots at each site (Table 1). We classified species as woody perennials using the USDA PLANTS database (plants.usda.gov); all species that had a perennial habit and were not strictly herbaceous were considered woody perennials, regardless of whether they always produce true anatomical woody. These included trees, shrubs, subshrubs and vines. Species that had an annual and/ or biennial habit in addition to a perennial habit were also included. When NEON and USDA PLANTS relied on different taxonomic authorities for naming species, we deferred to that used by NEON. In the few cases where different NEON surveyors used synonymous species names, we followed the taxonomic convention of USDA PLANTS.

We determined the native status of all study species by following the designation assigned in the USDA PLANTS database. Non-native
species were defined as those that are likely recent additions to the contiguous continental United States. Though not necessarily invasive, many of these species are considered horticultural or agricultural nuisances.

2.2 | Phylogenetic reconstruction

To reconstruct the phylogenetic relationships among co-occurring species, we obtained sequence data for our study species from GenBank as well as generated new sequence data for species within the community that did not have publicly available sequence data for one or more of our target gene regions. To reduce error in node age estimates that can occur if phylogenetic inference only includes data from members within a community (Park, Worthington, & Xi, 2018), we also obtained GenBank sequences for 10% more species, selecting from the non-woody perennial species within the same study communities. We also included sequence data for all non-native woody perennial species occurring within a 160 km radius around the communities of focus, as identified using the USDA PLANTS database, to use as a regional species pool for analyses (see “Phylogenetic relationships between native and non-native species” section). *Lycium japonicum* and *L. microphyllum* were included as outgroups. We focused on obtaining two chloroplast regions, *rbcL* (ribulose-1,5-bisphosphate carboxylase/oxygenase) and *matK* (maturase K), and one nuclear region, ITS (internal transcribed spacer 1 and 2, and 5.8S ribosomal DNA). To generate our own sequences, we obtained leaf tissue from herbarium specimens housed at the Rocky Mountain Herbarium (RM) and the Missouri Botanical Gardens Herbarium (MO), preferring green tissue from recent specimens (i.e., post-1980) when available. We isolated genomic DNA from 10 to 20 mg of tissue using the QIAGEN DNAeasy Plant Kit (QIAGEN, Valencia, CA, USA) and amplified the two chloroplast genes using universal primers (rbcL: 1F, 724R, 636F, 1368R, Lledo, Crespo, Cameron, Fay, & Chase, 1998; matK: MF, TF, trnK01, Aoki & Ito, 2000; 4L, 1777R, Hu, Lavin, Wojciechowski, & Sanderson, 2000; 1350R, Clarkson et al., 2004). PCR was performed in 25 μl reactions with 2 μg genomic DNA, 0.125 μl of each primer (4.0 mM final concentration), 0.125 μl dNTP mix (0.4 mM final concentration), 2.5 μl Standard Taq Reaction Buffer (10×) and 0.125 μl Choice Taq polymerase (5 units/μl; Denville Scientific, Holliston, MA). Thermocycling proceeded as follows: 95°C for 3 min to denature the DNA followed by 10 cycles of 95°C for 45 s, 55°C for 45 s decreasing 1°C per cycle, 72°C for 45 s; 20 cycles of 95°C for 45 s, 50°C for 45 s, 72°C for 45 s. Completion of these cycles was followed by a final extension of 72°C for 5 min. We amplified ITS using universal primers (ITS2, ITS4, White, Bruns, Lee, & Taylor, 1990; Q1, Q2, Samuel, Bachmair, Jobst, & Ehrendorfer, 1998; ITS leu.1, Andreassen, Baldwin, & Bremer, 1999; ITS273F, Ng & Smith, 2016) in 25 μl PCRs with 2 μl genomic DNA, 0.125 μl of each primer (4.0 mM final concentration), 0.125 μl dNTP mix (0.4 mM final concentration), 2.5 μl 25 mM MgCl₂, 2.5 μl Standard Taq Reaction Buffer (10×) and 0.125 μl Choice Taq polymerase (5 units/μl; Denville Scientific). Thermocycling proceeded as follows: 94°C for 2 min to denature the DNA followed by 35 cycles of 94°C for 1 min, 48°C for 1 min, 72°C for 1 min. Completion of these cycles was followed by a final extension of 72°C for 10 min.

We sent PCR products to GENEWIZ (Cambridge, MA) for cleanup and sequencing reactions. High-quality sequence data were obtained for both chloroplast and nuclear regions in all sampled taxa (GenBank acccessions: KX610097–KX610103; KX610105–610108; KY427297; KY427299–427303; KY427320–427325; KYS84297–KYS84387; KY707176–KY707180) and were manually edited using *Genedius* (v6.0.5; Biomatters Ltd., Auckland, NZ). For ITS sequences, we occasionally observed secondary peaks beneath primary peaks in the DNA chromatograms. We resolved ambiguities when possible by calling the primary/dominant peak, but used ambiguity codes in cases when making base calls was too difficult.

With 1,654 sequences obtained from GenBank and 119 newly generated sequences, we combined the sequence data and aligned each gene region using Mafft (v7; Katoh & Standley, 2013) and manually checked the alignments using Mesquite (v3.10; Maddison &

| Site | # 400-m² plots | Total species richness | # represented in tree | Native species (abundance) | Non-native species (abundance) | H’ (H’ native sp) | Standardized PD (Stand. PD native sp) |
|------|----------------|-----------------------|----------------------|-----------------------------|-----------------------------|------------------|-----------------------------------|
| HARV | 31             | 83                    | 80                   | 73 (614)                    | 10 (47)                     | 4.08 (3.98)      | 0.169 (0.167)                     |
| SCBI | 31             | 61                    | 60                   | 46 (428)                    | 15 (146)                    | 3.80 (3.53)      | 0.223 (0.254)                     |
| SERC | 6              | 50                    | 49                   | 37 (139)                    | 13 (45)                     | 3.76 (3.48)      | 0.230 (0.256)                     |
| ORNL | 28             | 83                    | 83                   | 69 (673)                    | 14 (72)                     | 4.07 (3.90)      | 0.194 (0.206)                     |
| TALL | 23             | 106                   | 105                  | 102 (723)                   | 4 (21)                      | 4.35 (4.31)      | 0.163 (0.162)                     |
| JERC | 8              | 113                   | 113                  | 108 (281)                   | 5 (6)                       | 4.45 (4.41)      | 0.147 (0.149)                     |
| OSBS | 26             | 109                   | 107                  | 103 (597)                   | 6 (7)                       | 4.33 (4.30)      | 0.160 (0.157)                     |

Note. The number of 400-m² plots sampled at each site, and species richness and abundance (in parentheses) of native and non-native species are reported. The number of species within each community represented in the phylogeny is also shown. Shannon diversity (H’), and Faith’s phylogenetic diversity (PD), standardized by species richness and depth of the community tree, are also reported for the entire community as well as native species alone (in parentheses). Site name abbreviations follow Figure 1.
Maddison, 2017). We concatenated all regions and used Bayesian inference to reconstruct a time-calibrated phylogeny for all species from our study communities as implemented in BEAST2 (v2.4.5; Bouckaert et al., 2014), with each gene region as a separate partition. We followed the age estimates of Magallón, Gómez-Acevedo, Sánchez-Reyes, and Hernández-Hernández (2015) to assign age constraints for major clades with each node calibrated with a log-normal distribution, a mean of 0.01 and a standard deviation of 1.0 (Supporting Information Appendix S1). Two runs of 120 million generations, sampling every 12,000 generations, were conducted on the CIPRES Science Gateway (www.phylo.org). We verified that both runs reached stationarity and convergence using TRACER v1.6 (Rambaut, Suchard, Xie, & Drummond, 2014) and discarded 50% of the trees from the first run and 60% from the second run as burn-in. We then combined and summarized trees as a maximum clade credibility (MCC) tree using the LOGCOMBINER and TREEANNOTATOR tools in BEAST2. To account for phylogenetic uncertainty, we subsampled 1,000 trees from the posterior distribution of trees using LOGCOMBINER and calculated all phylogenetic diversity metrics on each subsampled tree as well as the MCC tree. All trees were pruned to comprise only the community members prior to analyses.

2.3 | Community diversity and non-native species

To investigate differences in diversity across our study communities, we calculated Shannon diversity (H') for each community using the diversity function in the r package “VEGAN” (Oksanen et al., 2016) and Faith's phylogenetic diversity (PD; Faith, 1992) using the pd function in the r package “PISCANET” (Kembel et al., 2010). As PD is highly influenced by species richness, we also standardized PD by species richness and the depth of the community tree (Scheiner, Kosman, Presley, & Willig, 2017). To assess the diversity of the community without non-native species, we calculated H' and PD using only the native species of each community. We then investigated whether the relative species richness or relative abundance of non-native species could be predicted by the phylogenetic diversity of native species within the community by conducting regressions using the lm function in r.

2.4 | Phylogenetic relationships between non-native and native species

We used two approaches that differ in phylogenetic scale to evaluate whether the phylogenetic relationships between non-native and native species in communities were more consistent with either of the alternate DNC hypotheses. First, we used a broad-scale simulation approach to test whether non-native species were more closely or more distantly related to the native species of the community than expected by chance. We calculated the mean nearest ancestor distance between non-native species and their closest native relative (MNTDnn-n) for each community using a modified version of the mntd function in “PISCANET” (r script available at https://github.com/Gene-Weaver/Testing-Darwin-s-Naturalization-Conundrum). To test for significance, we generated a null distribution of MNTDnn-n values by conducting 1,000 randomizations that each involved replacing non-native community members with randomly drawn species from the site-specific regional non-native species pool (see “Phylogenetic Reconstruction” section) and recalculating MNTDnn-n. As we did not have any non-native gymnosperms within our communities (Appendix S2), we excluded drawing any gymnosperms from the non-native regional pool in our randomizations to ensure an unbiased comparison. We considered non-native species to be significantly more closely or distantly related to the co-occurring native species if the observed MNTDnn-n fell below 2.5% or above 97.5% of the null distribution, respectively.

Our second approach to testing DNC involved examining each non-native species at a finer phylogenetic scale. For each non-native species, we calculated the phylogenetic distance to the most closely related native species (NTDnn-n) and compared this value to the phylogenetic distance between the same native species and its closest relative native (NTDn-n). If NTDnn-n > NTDn-n, the non-native species is more distantly related to a native species than the native species is to its closest native relative, consistent with "Darwin's naturalization hypothesis" (one hypothesis of DNC). Alternatively, if NTDnn-n < NTDn-n, the non-native species is more closely related to a native species than the native species is to its closest native relative, consistent with the alternative hypothesis of DNC. We assessed significance by comparing our observed values to values calculated on trees with randomized non-native species (see previous paragraph) and considered the observed pattern to be significant if the observed values fell outside of 95% of the randomized distribution (r script available at https://github.com/Gene-Weaver/Testing-Darwin-s-Naturalization-Conundrum).

We also considered the ecological success of non-native species in light of DNC by integrating abundance data with phylogenetic data using the phylogenetic abundance evenness metric (PAE; Cadotte et al., 2010). Specifically, we compared PAE calculated only on the native species of a community to that of the whole community. PAE measures how evenly distributed individuals are among evolutionary lineages by first scaling the tips of the tree in proportion to each species’ abundance and then summing the total branch lengths of the community. If lineages are evenly abundant PAE = 1, while PAE < 1 when more abundant species are on shorter terminal branches, and PAE > 1 when more abundant species are on longer terminal branches. As we expect that a greater abundance of non-native species within a community should reflect their ability to successfully occupy and exploit novel habitat (e.g., Cleland et al., 2004; Li et al., 2015), PAE provides insight into whether successful non-native species are on longer or shorter terminal phylogenetic branches. We calculated PAE using an r script written by J. Hidasi-Neto (https://github.com/hidasi/rfunctions).

Last, we assessed whether differences in MNTDnn-n or PAE across communities could be explained by environmental differences. We quantified the environmental variation across our study communities using climate data from the WorldClim v.2 database at a 30 arc second spatial resolution (Fick & Hijmans, 2017). We obtained 19 bioclimatic variables and altitude for each community's
location and conducted a principal component analysis (PCA) to reduce collinearity among the variables. We used the resulting first two principal components (PC1 and PC2) to conduct regressions with MMTD_{inv} and PAE using the lm function in R.

3 | RESULTS

3.1 | Plant communities

Our southern study communities comprised a higher number of species than northern communities, with communities ranging from having 50 woody perennial species at SERC in the Mid-Atlantic ecoclimatic region to 113 species at JERC in the Southeast ecoclimatic region (Table 1). This latitudinal trend appeared to be driven by native species, with more native species occurring in southern communities. This pattern was not mirrored by non-native species; non-native species ranged from 3.8% of the woody community members at TALL in the Ozarks Complex ecoclimatic region to 26% of species at SERC in the Mid-Atlantic ecoclimatic region (Table 1; Figure 1; Supporting Information Appendix S2).

3.2 | Phylogenetic reconstruction

Our final sequence alignment contained 654 species, including 319 study species, and was 11,375 base pairs in length. This included most, or all, of the species present within each community (Table 1; Figure 2). Our time-calibrated phylogenetic reconstructions returned well-resolved relationships. Nodes were well supported across the tree with a median posterior probability of 1.0 and most values being above 0.90 (TreeBase ID 22,057; Figure 2). Phylogenetic relationships were congruent with the most recent phylogeny of angiosperms (which made up the largest number of sampled taxa in our analysis) (Angiosperm Phylogeny Group, 2016).

3.3 | Community diversity and non-native species

Shannon diversity across communities was relatively consistent, ranging from $H' = 3.76–4.45$ (native species only = 3.48–4.41) (Table 1). PD ranged from 3905.0 to 5925.3 (native species only = 3278.4–5718.2) (Figure 3), while standardized PD ranged from 0.147 to 0.230 (native species only = 0.149–0.256) (Table 1).
We found a significant negative correlation between relative non-native species richness and the PD of native species at each community ($R^2 = 0.80, p = 0.004$; Figure 3a), indicating that more non-native species were significantly more closely related to co-occurring native species for three communities, SCBI, SERC and ORNL (Figure 4). For the other four communities, non-native species were neither more closely nor more distantly related to co-occurring native species than expected by chance. These results held true even when considering phylogenetic uncertainty, whereby non-native species were significantly more closely related to native species for the majority of the 1,000 posterior trees (100% SCBI trees, 95% SERC trees and 96.6% ORNL trees). JERC showed significant patterns for a small proportion of the 1,000 posterior trees (17.2% JERC trees), while none of the posterior trees for the other three communities showed significant phylogenetic relationships between co-occurring non-native and native species.

When examining the relationships between non-native and native species at a finer phylogenetic scale, we found that in all but two communities, most non-native species were more closely related to a native species than the same native species was to its own closest native relative (Figure 5). When compared to a null distribution, this pattern was only found to be significant at OSBS where $NTD_{nn-n} < NTD_{n-n}$ for all non-native species. At HARV and ORNL, there were equal proportions of non-native species that were more closely ($NTD_{nn-n} < NTD_{n-n}$) and more distantly ($NTD_{nn-n} > NTD_{n-n}$) related to a resident native species than the native species was to another native species (Figure 5).

When combining species abundance with phylogenetic data for native species alone, we found that in most communities, the most abundant native species were on relatively long branches of the phylogeny (PAE > 1; Figure 6). However, at SERC, the most abundant native species were on shorter branches of the phylogeny (PAE < 1), while at SCBI, native species were evenly abundant within the community (PAE = 1). When non-native species were included in the PAE analyses, however, we did not find a consistent pattern; PAE decreased for some communities but increased for others (Figure 6). For four communities (HARV, OSBS, JERC, SERC), the inclusion of non-native species in the calculations shifted PAE values towards 1, indicating that the non-native species made the lineages within the communities more evenly distributed in abundance across the phylogeny. At SCBI, the shift from a PAE of 1 to less than 1 when including non-native species indicates that the more abundant non-native species, compared to the other species within the community, tend to be on relatively short terminal branches of the phylogeny (Figure 6). This suggests that the ecologically successful non-native species within this community are those that have more recently diverged from the resident species (native or non-native).

We found that environmental differences among sites were poor predictors of $MNTD_{nn-n}$ and PAE across communities. The first two axes of the PCA of the environmental variables explained 90.1% of the variance. PC1 explained 66% of the variance and mainly corresponded to temperature while PC2 explained 24.1% of the variance and mainly corresponded to precipitation. We did not find any significant relationships between $MNTD_{nn-n}$, and PC1 ($R^2 = 0.08, p = 0.549$; Figure 3c) or PC2 ($R^2 = 0.07, p = 0.283$). We also did not...
process of community assembly while offering predictive tools for the native community. We found that communities comprising along a latitudinal cline exclusively by the phylogenetic structure of their native species.

Our study analysed woody perennial plant communities to test whether the occurrence of non-native species could be predicted along a latitudinal cline exclusively by the phylogenetic structure of the native community. We found that communities comprising native species with lower PD tended to have more non-native species and the non-native species were also more abundant in these communities. We also found that non-native species tended to be closely related to resident native species, but this pattern was only significant for a subset of communities. The conflicting patterns across and within communities cannot be explained by environmental differences among sites and suggests that phylogenetic relationships should not be the only metric used to broadly predict non-native species establishment, particularly across communities differing environmentally and in disturbance history. Instead, our study suggests that the lack of a generalizable phylogenetic pattern may be due to site-specific historical influences or other unmeasured ecological and phenotypic attributes, or that the unique composition of evolutionary lineages within different communities has individualized responses to invading species pressure.

4 | DISCUSSION

The introduction and establishment of non-native species has been a major threat to native biodiversity on a variety of scales (Blackburn, Cassey, Duncan, Evans, & Gaston, 2004; Burbidge & Manly, 2002; Mooney & Cleland, 2001). Yet, species invasions also represent natural experiments that can offer insight into the process of community assembly while offering predictive tools for identifying communities most susceptible to non-native species. Our study analysed woody perennial plant communities to test whether the occurrence of non-native species could be predicted along a latitudinal cline exclusively by the phylogenetic structure of the native community. We found that communities comprising native species with lower PD tended to have more non-native species and the non-native species were also more abundant in these communities. We also found that non-native species tended to be closely related to resident native species, but this pattern was only significant for a subset of communities. The conflicting patterns across and within communities cannot be explained by environmental differences among sites and suggests that phylogenetic relationships should not be the only metric used to broadly predict non-native species establishment, particularly across communities differing environmentally and in disturbance history. Instead, our study suggests that the lack of a generalizable phylogenetic pattern may be due to site-specific historical influences or other unmeasured ecological and phenotypic attributes, or that the unique composition of evolutionary lineages within different communities has individualized responses to invading species pressure.

4.1 | Correlates of non-native species occurrence

Across communities, the PD of native species was a good predictor of non-native species richness and abundance. The negative association between native PD and non-native species richness and abundance is consistent with Elton’s classic hypothesis that more diverse communities are more resistant to species invasions (Elton, 1958; Levine & D’Antonio, 1999), which has also been supported by empirical studies (e.g., Kennedy et al., 2002; Naeem et al., 2000; but see Stohlgren, Barnett, & Kartesz, 2003). From a broader ecological perspective, the number of native species in our study communities also largely corresponds with the well-documented latitudinal biodiversity gradient observed for plants and animals (Hillebrand, 2004; Pianka, 1966), as we found that native species richness is higher at lower latitudes (Table 1: Figure 1). Communities with higher biodiversity at lower latitudes may therefore be more resistant to non-native species unless there are species within the community that

FIGURE 4 Mean nearest taxon distance between non-native species and the phylogenetically nearest native species (MNTD_nn-n) within each community calculated from the pruned MCC tree for each community (diamonds) compared to a random expectation (coloured dots). The simulated MNTD_nn-n expectations were generated by conducting 1,000 randomizations that involved randomly replacing non-natives from a regional pool of non-native species and recalculating MNTD_nn-n for each new random community. The observed values for three communities (SCBI, SERC and ORNL, denoted by an asterisk) fall below 95% of the null distribution, suggesting that non-native species at these sites are significantly more closely related to co-occurring native species than expected by chance. The observed values for all other communities are within 95% of the null distribution, indicating that there is no significant pattern (i.e., non-native species are neither more closely nor more distantly related to co-occurring native species than expected by chance).

FIGURE 5 Proportion of non-native species that are more closely related to a resident native species than the same native species is to its closest native relative (NTD_nn-n < NTD_n-n) versus the proportion of non-native species that are more distantly related to a native species than the same native species is to its closest native relative (NTD_nn-n > NTD_n-n). The asterisk beside OSBS indicates that the proportion of non-native species more closely related to a resident native species than they are to each other is greater than expected by chance.
facilitate the establishment of novel species, or the invaders possess novel ecological traits to exploit unoccupied ecological niches and can immediately out-compete resident species upon introduction.

Our finding that more non-native species occur in communities at higher latitudes may also be the result of differing disturbance histories. The northern latitudes of eastern North America were glaciated repeatedly throughout much of the Pleistocene and as recently as ~18,000 years ago (Davis, 1983; Pielou, 1991). This glaciation would have driven many of the species currently found at northern latitudes into southern refugia and would also have opened up newly available habitat upon glacial retreat (Davis, 1983; Jaramillo-Correa, Beaulieu, Khasa, & Bousquet, 2009; Sewell, Parks, & Chase, 1996). It may be that because of the history of repeated glacial–interglacial cycles over the last several thousand years, the northern communities in our study have not yet reached equilibrium with respect to species richness, allowing immigrant/non-native species to readily colonize these areas despite the resident species assemblages. In contrast, the southern communities likely remained entirely unglaciated during the Pleistocene, resulting in longer spans of time for speciation, adaptation and opportunities for the resident southern species to form associations with other refugial species. A history of species diversification and interactions over time may have led to a greater filling of ecological niches in southern communities, leading to contemporary immigrant/non-native species being largely precluded from establishing in these communities at present.

4.2 | Darwin’s Naturalization Conundrum persists

Our assessments of phylogenetic relatedness between co-occurring non-native and native species revealed mixed patterns across and within communities. Within three communities, we found that non-native species are more closely related to native species than expected by chance. At two of these communities (SCBI and SERC), most of the non-native species were more closely related to a resident native species than the same native species was to another native species (ND_Dn = ND_Dn). Indeed, most other communities also tended towards this same pattern, suggesting non-native species often co-occur with a phylogenetically close relative. However, OSBS was the only community to show a significant pattern at this finer phylogenetic scale, and all non-native species within OSBS were more closely related to a resident native species than that native species was to another native species (ND_Dn < ND_Dn). Despite this finding, non-native species at OSBS did not show significantly closer phylogenetic relationships to native species at the community level (Figure 4).

We also did not find a general pattern when incorporating species abundance into our phylogenetic analyses. Within four communities (SERC, OSBS, HARV, JERC), individuals were roughly equally distributed among non-native and native lineages (Figure 6) suggesting that non-native species within these communities were not always more ecologically successful than native species and also were not tightly associated with either long or short terminal branches. We found that in one community (SCBI), the more ecologically successful non-native species (i.e., those that are most abundant) were those that had more recently diverged from either a non-native or native resident species, and indeed, the more abundant non-native species often had another representative of the same genus inhabiting the same community (Supporting Information Appendix S2). This may reflect that within this community, the non-native species that co-occur with phylogenetically close relatives are able to progress further through the different stages of establishment and spread (Richardson & Pyšek, 2012) while non-native species that are more distantly related to the resident species are inhibited. Future studies in which the timing of non-native species introductions is known would help disentangle these alternatives for novel species establishment. For example, while the plant community data we obtained were among the initial surveys NEON released, plant surveys are planned to continue over 30 years (Field et al., 2006). Using these temporal data for the same seven communities analysed in this study could provide detailed insight into the dynamics between non-native and native species and allow further investigation into how different stages of non-native species establishment influence phylogenetic community structure.

The major findings of our study are consistent with one proposition of Darwin’s Naturalization Conundrum: Non-native species tend to be closely related to co-occurring native species. However,
in our broad survey across communities from multiple ecoclimatic regions, we did not find this to be a consistent pattern within every community. This suggests that the likelihood of a non-native species establishing within a community may largely be community-specific. As we did not find any significant associations between environmental variables and variation in MNTD and PAE among communities, the lack of generalizable patterns of relatedness are unlikely to be due solely to community-specific environmental differences. Instead, disturbance history or other unmeasured ecological factors could influence whether a non-native species is likely to invade a community. Furthermore, given the diversity of evolutionary lineages within our study communities ("woody perennials," loosely defined), the specific phenotypes of the community members, as well as the different processes that can allow species to co-occur with one another (e.g., adaptation, mutualism, facilitation; Cadotte, Davies, & Peres-Neto, 2017; Cavender-Bares et al., 2009; Gerhold et al., 2015; Mayfield & Levine, 2010), it perhaps is not surprising that we do not find a clear overarching pattern to help resolve DNC. That communities comprise collections of many evolutionary lineages may mean that some lineages are more receptive to co-occurring with closely related non-native species while other lineages may only be receptive to more distantly related non-native species. Moreover, whether a non-native species is ecologically successful may be community- or lineage-specific. In one community, only the non-native species closely related to the resident native species occurred at high abundance (Figure 6), while in other communities, there was equal abundance across both non-native and native lineages. Interestingly, our finding that fewer non-native species occur in phylogenetically diverse communities, and at lower abundance (Figure 3), suggests that communities comprising a diversity of lineages may always be resistant to new species, perhaps due to having a mix of species that both resist closely related and distantly related non-natives. Therefore, overall, our study shows that while patterns of phylogenetic relatedness among non-native and native species within a community can offer insight into a community’s susceptibility to invasion, phylogenetic relatedness alone may be an insufficient predictive tool for understanding the rules of community assembly. Rather, future efforts integrating phylogenetic diversity with other measures, such as phenotypic traits (e.g., Gaynor, Ng, & Laport, 2018; Marx et al., 2016; Schaefer et al., 2011), and abiotic factors (e.g., Lim, Crawley, De Vere, Rich, & Savolainen, 2014; Qian & Sandel, 2017), will help elucidate whether phylogenetic relatedness together with other factors can be used to predict a community’s susceptibility to the establishment of new species.

5 | CONCLUSIONS

Our analyses of seven communities of non-native and native woody perennial species across different ecoclimatic regions revealed striking differences in phylogenetic community structure across the eastern United States. We found that communities of phylogenetically diverse native species had fewer non-native species and that non-native species were often more closely related to native species than native species were to each other, as predicted by one proposition of DNC. Patterns of phylogenetic relationships between non-native and native species, however, were not consistent across all seven communities, suggesting that measures of phylogenetic relationships cannot solely be used as a broad predictive tool for community invasion. Although phylogenetic assessments can reveal important details about community assembly in diverse ecological settings, site-specific differences in biogeography, ecology, geology, and the evolutionary lineages present likely help explain the lack of a generalizable phylogenetic pattern across communities. Our study highlights the need to incorporate additional types of data into a phylogenetic framework, such as phenotypic information, to better understand the process of community assembly and why some communities are more susceptible to the establishment of non-native species.

ACKNOWLEDGEMENTS

We thank Vivianna Sanchez, Michelle Gaynor, Jessica Persinger and Nathaniel Pinson for their help generating sequence data, as well as Stacey Smith for generously providing research space. We are also grateful for the gracious assistance of the staff at the Rocky Mountain Herbarium (RM) and the Missouri Botanical Gardens Herbarium (MO) in providing access to sample specimens. This research was supported by the University of Colorado’s Biological Sciences Initiative BURST, UROP and BSI Scholar research grants to W.N.W., and a National Science Foundation grant and REU supplement to J.N. and R.G.L. (NSF-EF 1550813).

DATA ACCESSIBILITY

All DNA sequences used in this study have been deposited in GenBank (accession numbers: KX610097–KX610103; KX610105–610108; KY427297; KY427299–427303; KY427320–427325; KYSB4297–KYSB4387; KY707176–KY707180). Sequence alignments and the MCC phylogenetic tree can be found in TreeBASE (TreeBASE ID: 22057). R scripts used for analyses are available at https://github.com/Gene-Weaver/Testing-Darwin-s-Naturalization-Conundrum.

ORCID

Julienne Ng https://orcid.org/0000-0002-2994-6233
Robert G. Laport https://orcid.org/0000-0001-5672-0929

REFERENCES

Andreassen, K., Baldwin, B. G., & Bremer, B. (1999). Phylogenetic utility of the nuclear rDNA ITS region in subfamily Ixoroideae (Rubiaceae): Comparisons with cpDNA rbcL sequence data. Plant Systematics and Evolution, 217, 119–135. https://doi.org/10.1007/BF00984925
Angiosperm Phylogeny Group (2016). An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG IV. Botanical Journal of the Linnean Society, 181, 1–20.
Scheiner, S. M., Kosman, E., Presley, S. J., & Willig, M. R. (2017). The components of biodiversity, with a particular focus on phylogenetic information. Ecology and Evolution, 7, 6444–6454. https://doi.org/10.1002/ece3.3199

Sewell, M. M., Parks, C. R., & Chase, M. W. (1996). Intraspecific chloroplast DNA variation and biogeography of North American Liriodendron L. (Magnoliaceae), Evolution, 50, 1147–1154. https://doi.org/10.1111/j.1558-5646.1996.tb02355.x

Shea, K., & Chesson, P. (2002). Community ecology theory as a framework for biological invasions. Trends in Ecology & Evolution, 17, 170–176. https://doi.org/10.1016/S0169-5347(02)02495-3

Stohlgren, T. J., Barnett, D. T., & Kartesz, J. T. (2003). The rich get richer: Patterns of plant invasions in the United States. Frontiers in Ecology and the Environment, 1, 11–14. https://doi.org/10.1890/1540-9295(2003)001[0001:TRGRPO]2.0.CO;2

Strauss, S. Y., Webb, C. O., & Salamin, N. (2006). Exotic taxa less related to native species are more invasive. Proceedings of the National Academy of Sciences, 103, 5841–5845. https://doi.org/10.1073/pnas.0508073103

Tan, J., Pu, Z., Ryberg, W. A., & Jiang, L. (2015). Resident-invader phylogenetic relatedness, not resident phylogenetic diversity, controls community invasibility. The American Naturalist, 186, 59–71. https://doi.org/10.1086/681584

Thuiller, W., Gallien, L., Boulangeat, I., De Bello, F., Münkmüller, T., Roquet, C., & Laveigne, S. (2010). Resolving Darwin’s Naturalization Conundrum: A quest for evidence. Diversity and Distributions, 16, 461–475. https://doi.org/10.1111/j.1472-4642.2010.00645.x

Tilman, D. (2004). Niche tradeoffs, neutrality, and community structure: A stochastic theory of resource competition, invasion, and community assembly. Proceedings of the National Academy of Sciences of the United States of America, 101, 10854–10861. https://doi.org/10.1073/pnas.0403458101

Vitousek, P. M., Walker, L. R., Whiteaker, L. D., Mueller-Dombois, D., & Matson, P. A. (1987). Biological invasion by Myrica faya alters ecosystem development in Hawaii. Science, 238, 802–804. https://doi.org/10.1126/science.238.4828.802

Webster, C. R., Jenkins, M. A., & Jose, S. (2006). Woody invaders and the challenges they pose to forest ecosystems in the Eastern United States. Journal of Forestry, 104, 366–374.

Wells, R. D. S., de Winton, M. D., & Clayton, J. S. (1997). Successive macrophyte invasions within the submerged flora of Lake Tarawera, Central North Island, New Zealand. New Zealand Journal of Marine and Freshwater Research, 31, 449–459. https://doi.org/10.1080/00288330.1997.9516778

White, T. J., Bruns, T., Lee, S., & Taylor, J. (1990). Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics. In M. Innis, D. Gelfand, J. Sninsky, & T. J. White (Eds.), PCR protocols: A guide to methods and applications (pp. 315–322). San Diego, California: Academic Press.

Wiens, J. J., & Donoghue, M. J. (2004). Historical biogeography, ecology and species richness. Trends in Ecology & Evolution, 19, 639–644. https://doi.org/10.1016/j.tree.2004.09.011

Winter, M., Schweiger, O., Klotz, S., Nentwig, W., Andriopoulos, P., Arianaoutsiou, M., ... Kühn, I. (2009). Plant extinctions and introductions lead to phylogenetic and taxonomic homogenization of the European flora. Proceedings of the National Academy of Sciences, 106, 21721–21725. https://doi.org/10.1073/pnas.0907088106

**BIOSKETCH**

Julienne Ng is an evolutionary ecologist whose research focuses on identifying the drivers of biodiversity patterns at both micro- and macroevolutionary scales. Further details about her research interests can be found at https://julienneng.weebly.com.

William N. Weaver research interests include machine learning, big data and developing tools to increase the efficiency of biological research. Further information can be found at https://williamweaver.xyz.

Robert G. Laport is a plant evolutionary ecologist interested in the processes generating patterns of biodiversity. His research focuses on the population-level interplay between genome duplication (polyploidy), ecological processes and speciation. Additional details can be found at https://robertlaport.com.

Author contributions: J.N. and R.G.L. conceived the research; all authors collected and analysed the data; J.N. and R.G.L. drafted the manuscript with input from W.N.W.

**SUPPORTING INFORMATION**

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

How to cite this article: Ng J, Weaver WN, Laport RG. Testing Darwin’s Naturalization Conundrum using phylogenetic relationships: Generalizable patterns across disparate communities? Divers Distrib. 2019;25:361–373. https://doi.org/10.1111/ddi.12861