Opinion
Quantifying Differences Between Native and Introduced Species
Nathan P. Lemoine,1,* Deron E. Burkepile,2 and John D. Parker3

Introduced species have historically been presumed to be evolutionarily novel and ‘different’ from native species. Recent studies question these assumptions, however, as the traits and factors promoting successful introduced and native species can be similar. We advocate a novel statistical framework utilizing quantifiable metrics of evolutionary and ecological differences among species to test whether different forces govern the success of native versus introduced species. In two case studies, we show that native and introduced species appear to follow the same ‘rules’ for becoming abundant. We propose that incorporating quantitative differences in traits and evolutionary history among species might largely account for many perceived effects of geographic origin, leading to more rigorous and general tests of the factors promoting organism success.

Does Geographic Origin Matter?
The question of whether introduced species differ from native species has formed the cornerstone of invasion biology for decades. As a result, invasion biology contains nearly 30 hypotheses attempting to determine when introduced species will have greater abundances than native species [1]. Several of the most prominent hypotheses focus on the advantage conferred by novel geographic origin. For example, introduced species often are more abundant [2], have lost natural enemies [3], possess superior intrinsic traits [4], have unique evolutionary histories [5,6], and respond differently to global change [7–9] compared with native species. As a result, many argue that novel geographic origin is a strong predictor of a species’ ability to spread aggressively [10]. Here we argue that examining the success of introduced species based on geographic origin (i.e., introduced vs native) can be a false dichotomy. Specifically, we posit that categorical tests of geographic origin are inadequate tests of the invasion process because they generally fail to test whether successful native and introduced species possess similar traits, evolutionary histories, and environmental preferences. Thus, they do not test whether similar assembly rules govern the success of both native and introduced species.

Instead, we propose that researchers statistically test whether functional similarity and evolutionary history explain variation in species’ abundances and whether patterns differ for native and introduced species. Incorporating quantitative differences in traits and evolutionary history among species might therefore account for many perceived effects of novel geographic origin. This hypothesis rests on four observations. First, geographic novelty per se often confers little advantage to introduced species [11–13] and the performance of many of the world’s most abundant introduced species can be quite similar in both the novel and native ranges [14]. Second, some native species, such as the western juniper [15], cattail [16], black vultures [17], Canada goose [18], and white-tailed deer [19], also exhibit ‘invasive’ tendencies in their native

*Correspondence: lemoine.nathan@gmail.com (N.P. Lemoine).

© 2016 Elsevier Ltd. All rights reserved.
ranges given a large disturbance or release from previous population constraints. Third, even problematic ‘invasive’ species experience the same constraints on occurrence and abundance as native species [20]. Fourth, recent proposals that novel geographic origin explains much about the potential impact of a species on the recipient community rely on the fact that introduced species often possess traits different from those of native species [10]. Thus, the perceived effect of geographic origin might stem from quantitative trait differences between native and introduced species.

Together, these observations suggest that the enhanced performance of many introduced species compared with native species arises from sampling effects [21,22]; that is, the selective study of problematic introduced species that possess a particular suite of traits allowing rapid population growth (Box 1). Although the study of these species is undoubtedly important, this bias could lead to the assumption that introduced species are inherently different from natives [10]. However, this assumption is not adequately tested by ‘ANOVA-style’ statistics that use geographic origin as a binary predictor, largely because these models do not test whether the

---

**Box 1. Are Successful Introduced Species an Idiosyncratic Sampling Effect?**

Sampling effects in studies of biodiversity arise when one species drives observed effects of biodiversity by virtue of traits inherent to that species [21]. Since introduced species are often just as abundant in their native range [13,14] and some native species possess the ability to become ‘invasive’ [51], the tendency for some introduced species to spread rapidly may arise from simple sampling effects. That is, abundant and problematic introduced species may simply be more fecund than natives and a given environment selects for particularly fecund species [4] (Figure IA). The alternative hypothesis is that geographic novelty confers an advantage to non-native species and introduced species become more abundant despite having similar fecundity to native species (Figure IB).

Numerous experiments have examined how range expansion alters the fecundity or growth of introduced species [52–54] and whether introduced species are more fecund than native species [3]. Some researchers have pointed out, however, that such comparisons may be biased in species choice, contrasting highly fecund introduced species with less fecund native species [11,55]. Only a community-level analysis can distinguish between these two hypotheses. The appropriate approach would be to regress abundance against fecundity for all (or most) species within a community allowing different slopes for native and introduced species. Lack of an interaction would indicate that introduced species perform identically to native species and succeed by virtue of trait differences (Figure A). A significant interaction might indicate that introduced species are abundant despite possessing similar traits to native species (Figure IB).

Such analyses were conducted by both Meiners [11] and Leishman et al. [12]. Meiners used a 50-year time series to determine whether the 25 most frequent native and exotic species (50 species total) exhibit different successional trajectories. He found that the maximum potential growth rate and the maximum frequency of occurrence across the landscape were identical for native and introduced species across the entire community [11]. Leishman et al. recorded leaf traits for a large number of native and introduced taxa and calculated the slope between all possible pairwise combinations of traits. In 89% of cases, there were no differences in trait scaling relationships between native and introduced species, suggesting that native and introduced species have identical leaf economies and carbon capture strategies. In both of these studies, community-level analyses support our hypothesis that functional and evolutionary differences drive species performance, not geographic origin (see Figure IA in main text).

---

**Figure I. Hypothetical Demonstration of (A) a Sampling Effect that Introduced Species Are Simply More Fecund and (B) Real Differences Between Introduced and Native Species.**
same factors that promote introduced species also promote the abundance of similar native species (Box 1) [4,23–30]. Instead, we advocate the use of regression-style models that can better address whether native and introduced species exhibit quantitative differences in their relationships among organism abundance, species’ traits, evolutionary history, and abiotic drivers in a given environment.

Community assembly rules invoking species’ traits and evolutionary histories can be traced back to Darwin. In *Origin of Species*, he suggested that introduced species with close relatives in the recipient community should have difficulty establishing due to high functional similarity with resident species [23,31]. At the same time, Darwin recognized that phylogenetic similarity might pre-adapt an introduced species to local environmental conditions, thereby increasing its odds of establishment within a community. This mechanism led to the environmental filtering hypothesis; that is, the notion that abiotic conditions select for a subset of species with similar, adaptive traits [32,33]. However, invasive species are often less related to native communities than would be expected if colonization were random [5,28,29,34–37], implying that competition with native species is a strong determinant of invasion success. Despite this broad generality, these studies often do not simultaneously determine whether native taxa are similarly constrained by evolutionary similarity. We propose that a more rigorous test of community assembly mechanisms for introduced species would quantitatively compare whether evolutionary or functional dissimilarity promotes the success of introduced and native species equally (Box 1).

Here we describe new statistical tools incorporating ecological traits and evolutionary relationships that allow more sophisticated tests of community assembly mechanisms for native and introduced species. Ultimately, we argue that species’ traits and phylogenetic information provide quantifiable metrics that capture ecologically relevant differences between native and introduced species, and that quantitative analyses of these differences provide more rigorous and comprehensive tests of the factors promoting organism success than simply comparing ‘native versus introduced’.

**Traditional Tests of Community Assembly**

Historically, community assembly processes have been inferred from randomization tests of functional trait structure (e.g., [38]). For these tests, a suite of ecologically relevant traits (e.g., specific leaf area, leaf nitrogen content, fecundity, growth rate for plants) is measured on most species in a regional species pool. Each species present in a study site is assigned a measure of trait similarity to all other co-occurring species; that is, the minimum absolute difference or mean difference from all other species. Species-level metrics are then averaged across all species within a site, yielding a site-level metric of trait similarity. Random assemblages are then generated from the regional species pool to generate a null distribution. If the observed site-level metric is significantly lower than the null distribution (i.e., trait distances are smaller than expected by chance), it is assumed that environmental filtering selects for functionally similar species. If the observed site-level metric is significantly larger than the null distribution, niche partitioning is assumed to be the dominant mechanism (i.e., trait distances between species are larger than expected by chance).

Similar null model methods are used to determine the effect of evolutionary history on the invasion process, typically by replacing functional trait distances with pairwise branch distances derived from phylogenetic trees [5,37–39]. Importantly, these phylogenetic distances are often meant to serve as proxies for functional similarity based on the assumption that ecologically relevant traits are phylogenetically conserved [40]. Although this assumption can be violated [41], phylogenetic distances remain useful as they may describe aspects of functional relatedness not incorporated in trait measurements.
Null model approaches have greatly advanced our understanding of community assembly mechanisms. Nevertheless, null models suffer from several limitations that limit their generality. First, they condense species-level information into a single, site-level number, discarding data on species that do not occur at a given site. Accordingly, this method cannot consider how similarity affects absences from a community (i.e., failed introductions). Also, similarity metrics are often unweighted by species abundances, such that very common and very rare species are given equal weights in null models [28,29,37,39]. Weighting, however, can dramatically alter the outcome and conclusions of such tests [20], reflecting the crucial underpinning that common species interact more strongly than rare species [42]. Finally, quantitative assessment of how environmental drivers impact assembly mechanisms [38] or differences between native and introduced species [39] cannot be conducted with the null model approach, which instead relies on qualitative comparisons of results from different sites [43].

**A Novel Statistical Framework for Testing Community Assembly**

Regression techniques circumvent many of the above problems by allowing researchers to: (i) use both presences and absences (i.e., failed invasions); (ii) incorporate species abundances; (iii) include functional and/or phylogenetic similarity as a predictor of species’ abundances; and (iv) assess the influence of environmental factors on assembly mechanisms. Unfortunately, the structure of community composition data often does not conform to easy analysis via traditional regression methods for several reasons. First, most species from the regional species pool do not occur in a given site, such that zeros dominate abundance data at a given site. Second, species’ abundances within a site often follow a log-normal or gamma distribution because most species are rare. For these reasons community composition data are often collapsed into presence/absence data and analyzed using permutational tests, which have several drawbacks as described above.

Lemoine et al. [20] described a new statistical model for describing the relationship between traits, phylogenetic distances, and species’ abundances. Assuming that species’ absences from a site are ‘true’ absences and not missed due to observation error, the data can be analyzed using a zero-altered distribution [44]. Zero-altered models comprise a statistical distribution constrained to positive values, such as the gamma or log-normal distribution, augmented to allow zero values with a piecewise likelihood function. The likelihood \[ L(\tilde{y}) \] of species \( i \) occurring in site \( j \) then becomes:

\[
L(\tilde{y}_i) = \begin{cases} 
1 - \pi_i & \text{if } y_i = 0 \\
\pi_i \times \text{Gamma}(\alpha, \beta_i) & \text{if } y_i > 0
\end{cases}
\]

where \( \pi_i \) is the probability of the \( i \)th species occurring within a site \( j \). The shape and rate parameters of the gamma distribution are given by \( \alpha \) and \( \beta_i \), respectively. Allowing rate to vary by observation \( \beta_i \) allows the variance to increase as the square of the mean, which is often necessary for relative abundance data. This assumption can be relaxed, however, fixing both shape and rate or allowing shape to vary by observation instead, depending on the distribution of the data.

Both the probability of occurrence \( (\pi_i) \) and the relative abundance \( \tilde{y}_i \) can be made linear functions of functional similarity, phylogenetic relatedness, and geographic origin within a site:

\[
\pi_i = \gamma_0 + \gamma_1 \text{FuncDist}_i + \gamma_2 \text{PhyloDist}_i + \gamma_3 \text{Origin}_i + \gamma_4 \text{Origin}_i \times \text{FuncDist}_i + \gamma_5 \text{Origin}_i \times \text{PhyloDist}_i
\]

\[
\log(\tilde{y}_i) = \theta_0 + \theta_1 \text{FuncDist}_i + \theta_2 \text{PhyloDist}_i + \theta_3 \text{Origin}_i + \theta_4 \text{Origin}_i \times \text{FuncDist}_i + \theta_5 \text{Origin}_i \times \text{PhyloDist}_i
\]

where \( \tilde{y}_i \) is the predicted relative abundance of species \( i \) in site \( j \), \( \text{FuncDist}_i \) is a metric of functional similarity of species \( i \) to all other species in site \( j \), \( \text{PhyloDist}_i \) is a metric of phylogenetic
relatedness of species $i$ to all other species in site $j$, and \textit{Origin}_i denotes whether the $i$th species is native (0) or introduced (1). For gamma distributions, predicted relative abundance is converted to the shape parameter in the likelihood function as $\beta_{ij} = \alpha / \gamma_{ij}$. Importantly, including a variable for introduced status and all relevant interactions allows researchers to statistically determine whether traits and evolutionary history similarly constrain the abundance of native and introduced species [20].

This model allows the effects of traits, phylogeny, geographic origin, and their interactions to vary among multiple sites by modeling each parameter as a linear function of environmental variables measured at each site. To illustrate, $\theta_{ij}$ (the effect of functional distance on relative abundance at site $j$) is randomly drawn from a normal distribution with a predicted value that depends on, for example, mean annual temperature (MAT) and precipitation (MAP):

$$\theta_{ij} \sim N(\bar{\theta}_{ij}, \sigma^2_{\theta_i})$$  \(4\)
$$\bar{\theta}_{ij} = \mu_0 + \mu_1 \text{MAT}_j + \mu_2 \text{MAP}_j$$  \(5\)

This statistical framework assesses the influence of functional similarity and phylogenetic relationships on both species occurrence and abundance simultaneously (Box 2). Negative slopes within a site indicate environmental filtering, wherein the probability of occurrence or abundance declines with increasing functional or phylogenetic distance. Conversely, positive slopes indicate niche partitioning, wherein the probability of occurrence or abundance increases with increasing functional or phylogenetic distance. The inclusion of site-level predictors can determine whether the abiotic template of a site affects the community assembly process or alters the differences between native and introduced species. Thus, rather than numerous randomization tests that both discard large amounts of data and inflate the possibility of Type I errors, this model tests all relevant questions at once using all available data.

**Case Study 1: Temperate Forest Understories**

Forests along the Mid-Atlantic coast of the USA contain a large number of introduced species. Lemoine et al. [20] hypothesized that introduced species experience stronger environmental filtering than native species, as demonstrated previously [39]. To test this hypothesis, they combined relative abundance data, leaf and seed traits, and a molecular phylogeny for 71 species in 25 different forest stands. They also quantified multiple environmental variables describing the abiotic context of each forest stand (litter depth, light availability, and soil moisture content). They used the model described here to determine how mean phylogenetic distance (MPD), mean trait distance (MTD), and introduced status (native/introduced) interactively determine both the occurrence and abundance of species in forest understories and whether environmental context influenced the observed patterns.

Introduced species were less likely to occur than native species in older forests and were more abundant than native species in young forests (Figure 1C,H). Species with highly dissimilar traits were less likely to occur within a forest stand and, importantly, this effect was identical for both native and introduced species (Figure 1B,E). Species that were closely related and possessed leaf characteristics [specific leaf area (SLA), toughness] similar to those of other species in the community were more abundant. Again, these patterns were identical for native and introduced species except for a few of the younger forests (Figure 1F–J). Additionally, the model was able to describe how environmental context affected these patterns. Increasing litter depth, for example, decreased the probability of introduced species occurring in a forest stand, and introduced species were both more abundant and suffered stronger environmental filtering in younger forests with higher soil moisture content [45].
Overall, environmental filtering selected for closely related species possessing similar leaf economies in all forest stands. Geographic origin explained very little variation in community assembly processes beyond evolutionary history and trait differences. However, the few differences between native and introduced species shed light on the invasion process. Introduced...
Case Study 2: California Grasslands

In contrast to forests, grasslands exhibit intense competition among grasses and forbs at small spatial scales [46,47], and competition can be most intense between closely related grass species [47] (but see [48]). Grasslands may therefore undergo different community assembly processes than forests with potentially stronger niche partitioning. As a result, niche partitioning of introduced species may be more important in grassland ecosystems than in forests.
Cadotte et al. [39] tested whether native and introduced species experience similar community assembly mechanisms at four California grassland sites using traditional null model approaches. They found that introduced species demonstrated clustered phylogenetic patterns at low species richness while native species demonstrated overdispersed patterns at high species richness. These results imply that evolutionary history differentially affects the occurrence of native and introduced species and that these effects vary along a species-richness gradient.

We reanalyzed their data using the statistical framework outlined here. First, we averaged the relative abundance of each species across all plots within a site, such that the response variable was average relative cover of grass and forb species in 1 m² plots at each of the four sites, and then calculated weighted MPD for each species within each site. Site-level predictors included total species richness, precipitation, elevation, and average biomass per plot to determine how environmental factors influenced community assembly.

In our model there were no fundamental differences in community assembly mechanisms between native and introduced species (Figure 2). Introduced species were more likely than native species to occur at sites with low species richness, but MPD did not affect the probability of occurrence at any site. Introduced species were more likely than natives to occur at the sites with 34 species, but were also unaffected by MPD. Native species underwent significant environmental filtering at the most speciose sites where introduced species were relatively more abundant. The lack of significant interactions indicates that MPD affects native and introduced species similarly at all sites.

Figure 2. Coefficient Estimates for the Effects of Mean Phylogenetic Distance (MPD), Introduced Status (Introduced), and Their Interactions on the Occurrence and Abundance of Species in Four Grassland Sites. Numbers to the right of each confidence interval indicate the probability that the coefficient is greater than 0, such that either extremely high (>0.95) or extremely low (<0.05) numbers indicate significant effects. (A) Species were less likely to occur at sites with low species richness, but (B) MPD did not affect the probability of occurrence at any site. (C) Introduced species were more likely than natives to occur at the site with 34 species, but (D) were also unaffected by MPD. (F) Native species underwent significant environmental filtering at the most speciose sites where (G) introduced species were relatively more abundant. (F) The lack of significant interactions indicates that MPD affects native and introduced species similarly at all sites.
natives to occur at only a single site (Figure 2C) but MPD had no influence on the occurrence of either native (Figure 2B) or introduced (Figure 2D) species. At the most speciose sites, however, native species underwent significant environmental filtering (Figure 2F). Introduced species were more likely to occur at these same sites (Figure 2G) but did not differ from natives in their response to MPD (Figure 2H). No site-level characteristic had a significant effect on community assembly, indicating that site-level variation in community assembly was explained by some unmeasured factor.

These results are qualitatively similar to those reported from temperate forest stands. The most common species were phylogenetically similar regardless of their native or introduced status, suggesting the widespread importance of environmental filtering. Yet, introduced species were more likely to occur and were more abundant than natives at some sites. These discrepancies between native and introduced species can therefore lead to more in-depth questions and research on the invasion process.

Concluding Remarks and Future Directions

The flexibility of our statistical model enables researchers to adapt it to answer numerous different and important questions (see Outstanding Questions). For example, both of the case studies presented here span a relatively small environmental gradient. Invasion biologists might therefore ask how community assembly mechanisms of native and introduced species vary across large, continental-scale variations in temperature, elevation, precipitation, species richness, and community composition. Our model can also incorporate temporal changes in community assembly by treating year as the grouping variable instead of site and incorporating interannual environmental variables as predictors. Finally, our model can be expanded to account for multiple hierarchical levels, allowing researchers to examine how environmental variation at multiple spatial scales influences community assembly. Relevant predictors can be included at each level, which would allow simultaneous testing of whether microhabitat variation in species richness, soil nitrogen, etc. and site-level variation in species richness, precipitation, temperature, etc. affect community assembly.

Although decades of theory posit that introduced species require unique ecological hypotheses to explain their spread and success, recent evidence suggests that native and introduced species have similar succession trajectories [11], leaf carbon capture traits [8,12], chemical defenses [49], enemy release [45], soil allelopathy [50], and community assembly mechanisms (see [20] and Case Study 2). These results suggest that introduced and native species might thrive for very similar reasons. Here we present a novel quantitative framework that allows researchers to better test whether similar assembly rules govern the presence/absence and the abundance of native and introduced species. This method provides a more rigorous and quantitative test of the invasion process than traditional introduced versus native tests and should help address the larger question of whether geographic novelty conveys any inherent advantage above and beyond quantifiable differences among species, including their traits, evolutionary similarities, and responses to key environmental drivers.

Acknowledgments

The authors thank Dr Paul Craze and three anonymous reviewers for their helpful suggestions on the manuscript. Dr Mark Cadotte provided all data for the case study in California grasslands. This work was supported by a Dissertation Year Fellowship to N.P.L. from Florida International University.

References


Outstanding Questions

We focus here on testing how traits and phylogenetic relationships determine species’ occurrences and abundances within a forest stand and whether geographic origin provides unique information once these differences among species have been statistically incorporated into the model. However, there are many more processes that can be analyzed with simpler multiple regressions that incorporate traits and phylogenetic distances. For example, herbivore feeding preferences could be modeled as a function of plant abundance, leaf toughness (or trichomes), plant relatedness to other plants, and geographic origin. A similar model could be constructed for carbon storage, primary production, drought tolerance, range expansion, or any other number of important variables. In summary, we suggest that only a whole-community approach that explicitly tests for the effects of traits, phylogenetic relatedness, and geographic origin on the process of interest provides an adequate test of invasion hypotheses.


