**TITLE PAGE** (Journal: Ecology Letters)

**Title:** Deconstructing Darwin’s Naturalization Conundrum in the San Juan Islands using community phylogenetics and functional traits.

**Authors:** Hannah E. Marx, David Giblin, Peter Dunwiddie, David C. Tank

**Running Title:** Darwin’s Conundrum in the San Juan Islands

**Keywords:** San Juan Islands, Pacific Northwest, Darwin’s Naturalization Conundrum, Darwin’s Naturalization Hypothesis, mega-phylogeny, community phylogenetics, invasive species, functional traits, phylogenetic diversity, phylogenetic nearest native distance, islands.

**Type of Article:** Letters

**Number of words in the abstract:** (150 max) 146

**Number of words in the main text:** (5000 max) 6403

**Number of words in each text box:** Table 1 = 76,Figure 1 = 110, Figure 2 = 217, Figure 3 = 509, Figure 4 = 135, Figure 5 = 123.

**Number of References:** (50 max) 71

**Number of figures, tables, and text boxes:** (6 max) 6 (1 table, 5 figures)

**Name and mailing address for correspondence:**

Hannah E. Marx

Institute for Bioinformatics and Evolutionary Studies, University of Idaho, 875 Perimeter Drive MS 3051, Moscow, Idaho 83844-3051 USA

Phone: (208) 885-7033

Fax: ??

Email: [marx6272@vandals.uidaho.edu](mailto:marx6272@vandals.uidaho.edu)

**STATEMENT OF AUTHORSHIP:**

PD and DG conducted the floristic survey of the San Juan Islands. HEM and DCT conceptualized the study and analyzed output. HEM performed phylogenetic and statistical analyses. HEM wrote the first draft of the manuscript, and all authors contributed to revisions. **ABSTRACT**

Darwin posed a conundrum about species invasions, postulating both the importance of functional distinctiveness from the receiving native community to avoid competition, and, at the same time, the importance of shared similarity to pass environmental filters and successfully establish. Using diversity metrics derived from both phylogenetic information and functional trait data, we assessed this conundrum in the flora of 80 uninhabited islands in the San Juan Islands of Washington State, where ~30% of the species are invasive. We discovered support for both sides of Darwin’s conundrum coin—although invasive species have phylogenetically close native relatives, functional traits differ between the two status groups, at least for maximum height and specific leaf area. We highlight the importance of publicly available datasets to explore phylogenetic and functional diversity, and the necessity of clear hypotheses and biologically relevant null expectations to disentangle ecological processes that drive patterns of invasion.

**INTRODUCTION**

Species distributions have constantly been reshuffled on the earth’s surface, implicating many species as invasive to a given community at some point. Ranges of invasive species are currently increasing with human-mediated dispersal—especially on islands {Pysek:2006dw}—and in the face of accelerating global change, it is of economic importance to predict which species are most likely to become invasive and their potential for future spread and impact {Bellard:2013bt}. Modern species introductions are an important process to study not only for economic and conservation concerns, but also, when considered as recent colonizers in an ecological community, invaders can be used to understand the basic processes of community assembly in natural systems {Tilman:2004wy}.

Darwin was among the first to realize the value of invasive species in eco-evolutionary studies for understanding community assembly processes. He postulated that the more distantly related species were to each other, the greater their difference in functional traits should be, and thus, the less competition for resources they would experience due to “diversification of structure” {Darwin:1859vq}. Therefore, invasive species that are distantly related to the native community would be more likely to possess ‘novel’ traits allowing them to successfully establish via niche differentiation and escape from competition, and this pattern is often defined as “Darwin’s Naturalization Hypothesis” {Rejmanek:1996tq}. However, Darwin also discussed the importance of ancestral similarity, and hypothesized that pre-adapted traits would be important to pass environmental filters and survive within a particular environment. Highlighting the importance of both diversification and conservation of traits for invasion success, these two theories have been united as “Darwin’s Naturalization Conundrum” (DNC) {Diez:2008is}. Ordonez (2014) recently shed light on this conundrum by separating the later prediction into a third option, where invasive species might be closely related to the native community (due to an abiotic filter), but differ in some traits in order to “fill” available niche space. Thus, we are left with three possible scenarios for invasion success: (1) that naturalized invasive species need to be completely distinct from the recipient native community, distantly related to the natives and at the edge of the spectrum of functional traits (“Darwin’s Naturalization Hypothesis” {Rejmanek:1996tq, Daehler:2001cm}), (2) that invasive species would be similar to the recipient community, closely related to the native community and share similar functional traits (“matching hypothesis” {Ordonez:2014gp}, {Proches:2008kx}), or (3) that the successful invaders would be phylogenetically similar, but differ in some functional traits to fill unoccupied niche spaces (“filling hypothesis” {Ordonez:2014gp, Proches:2008kx}).

The inclusion of evolutionary information using a phylogenetic approach provides an accurate quantification of relatedness by comparing phylogenetic distances between native and invasive species {Proches:2008kx, Thuiller:2010dc}, and the use of phylogenies to investigate patterns of the naturalization process is becoming standard {Strauss:2006tt, Schaefer:2011kk, Carboni:2012wo, Park:2013dx, Ordonez:2014gp}. In addition, because related species are more likely to share ecologically relevant functional traits {Webb:2002be, CavenderBares:2009kf}, phylogenetic distance as measured on a molecular phylogeny of the species within a community is often viewed as a proxy for ecological similarity {Webb:2000dy, Webb:2002be, Cadotte:2009fh, CavenderBares:2006tj}.

However, because complex mechanisms like competition can produce contrasting community patterns {Mayfield:2010cs}, linking an ecological process to the phylogenetic patterns is controversial, especially at varying spatial scales (reviewed in {Pavoine:2010fl}). The addition of functional trait-based metrics with evolutionary distances helps to distinguish between different assembly scenarios {Cadotte:2009fh, Mouquet:2012he, Cadotte:2013hy}. Importantly, quantifiable measures of functional traits have the potential to be compared across ecosystems, and there are increasing attempts to move beyond species and describe functional diversity to search for more generalizable patterns of assembly {McGill:2006fb, Cadotte:2011iv, CavenderBares:2009kf, Cianciaruso:2011un, Roquet:2012vt}. For example, in plant systems, functional traits such as specific leaf area (SLA), height, and seed mass are assumed to be correlated with ecological strategy or primary production, and as such, are considered an appropriate proxy for selective forces that shape the community or ecosystem {Westoby:1998ug}.

With increasing availability of molecular sequence data, advances in phylogenetic methods (e.g., {Smith:2009go}) and the availability of large trait databases (e.g., {Kattge:2011fc}), more studies are integrating phylogenies and functional traits into investigations of Darwin’s Naturalization Conundrum in natural systems (e.g. {Carboni:2012wo, Schaefer:2011kk, Ordonez:2014gp}), and relatedness has been shown experimentally to be a useful predictor of invasion success {Jiang:2010ds, Cadotte:2010hc}. Because these studies have been conducted across systems ranging from coastal Mediterranean dunes {Carboni:2012wo} to global communities defined by habitat types {Ordonez:2014gp}, it is difficult to compare results due to differences in spatial and taxonomic scales (e.g. {CavenderBares:2006tj, Cadotte:2009ej}), phylogenetic estimation methods, varying metrics used to summarize evolutionary distance, the choice of functional traits, the definition of an invasive species, and importantly, the choice of null models and significance tests to assess hypotheses of invasiveness and invasibility (reviewed in {Thuiller:2010dc}, Table 1). Not surprisingly, their results show various conclusions about DNC, and the explanations for the ecological processes potentially driving these patterns of invasion are often weakly supported or conflicting.

For example, Schaefer et al. (2011) compiled a list of the angiosperm flora of the Azores, and used a species-level phylogeny and seven ecological characteristics to assess the invasive potential of species at various spatial scales. With this approach, they found invasive species were more distantly related to their nearest native relatives than non-invasive introduced species were on islands greater than one km2, yet found no significant pattern in plots less than 100m2. Increased phylogenetic distance combined with difference in life form from the nearest native angiosperm relative significantly explained the probability of invasive success. However, observed trait values for life form and seed mass explained more variation than the phylogenetic distance. They conclude that greater phylogenetic distance, yet functional similarity for most traits, suggests support for an ‘enemy release hypothesis’, the relatively low herbivore or pathogen pressure compared to the native community, in determining invasive success in Azores, but did not test these observed patterns against any random expectation {Schaefer:2011kk}. In contrast, Carboni et al. (2012) found invasive species were clustered to their nearest native and the native community at 4m2 plots on invaded coastal dunes, but not at coarser spatial scales. Because functional differences existed for seed mass, leaf thickness, and seed shape, they hypothesized their observed pattern was due to biotic resistance through competition at fine spatial resolution, and habitat filtering at coarser spatial scales {Carboni:2012wo}. More recently, a global comparison of vascular plants showed that alien taxa tend to have close native relatives in the community, yet are functionally distinct from the community of native species as well as to their nearest native relatives for specific leaf area (SLA), maximum height, and seed mass, highlighting the importance of both environmental filtering and competitive interactions in the naturalization process {Ordonez:2014gp}.

A comprehensive botanical survey of 80 uninhabited islands in the San Juan Islands archipelago off the Northwest corner of the United States revealed that despite the current lack of permanent human settlements, these native communities have all been invaded. European history is relatively recent in Northwestern North America, therefore these vascular plant communities have all been invaded within the last several hundred years. Ranging in size from 1m2 – 0.57km2, the replication of invasions across a gradient of spatial scales and the extent of invasion in these uninhabited communities provides an ideal system to investigate patterns of invasion in light of both phylogenetic and functional metrics of similarity.

In this study, we assessed Darwin’s Naturalization Conundrum (DNC) in the San Juan Islands, using metrics and approaches that are comparable to previous studies. We use branch lengths from a time-calibrated species level mega-phylogeny to calculate the evolutionary distances between invasive and native plants across all 80 islands in the San Juan archipelago, so that this study may be compared to others in the future. Additionally, we compared values for five ecologically relevant traits between native and invasive species to evaluate functional similarity and distinguish between processes driving phylogenetic patterns in the light of the three scenarios of Darwin’s Conundrum. Specifically, we addressed (1) how invasive plants are related to their nearest native relative and to the greater native community, and (2) if invasive and native species differ in values for phenotypic traits overall, compared to their nearest native, or to the average of the native community. We highlight the importance of biological surveys and publically available datasets to describe multiple aspects of biodiversity patterns, and the use of biologically meaningful null expectations to understand macro-ecological processes of invasion across the San Juan Islands archipelago.

**MATERIALS AND METHODS**

**Floristic survey and island community matrix**

Floristic surveys were conducted across islands in the San Juan Islands archipelago from 2005-2010. Islands were included in the survey based on several criteria. A primary interest was to compile a complete inventory of the vascular flora on each island visited. Therefore, surveys focused only on smaller islands – generally <25 ha – where we were confident that all habitats could be carefully searched. Most islands of this size are in public ownership, and permission to conduct surveys and make collections was secured in advance from the appropriate agencies. In addition, several privately owned islands were also visited with permission. Islands were accessed using small launches by survey teams that included multiple experts in the local flora to ensure consistent and thorough recognition of all species. To document the flora of each island as completely as possible, multiple visits were made that were timed to facilitate detection of species throughout the growing season. Generally, visits were made in late-April, late-May, and early September.

The definition of what constitutes an invasive species is often difficult to determine due to different stages in the process of invasion {Vermeij:1996uv}. The introduction, establishment and naturalization of a species involves generating populations outside of the native range, and thus relies more heavily on abiotic interactions {Schaefer:2011kk}. In contrast, the final integration or ‘invasion’ stage implies a spread across natural habitats, and is more likely to depend on biotic interactions within the community {Richardson:2006dk, Vermeij:1996uv}. Because Darwin’s Naturalization Conundrum is specifically addressing biotic interactions such as competition, tests of DNC should focus on this final stage of the process and upon introduced species that have naturalized to become invasive {Thuiller:2010dc, Schaefer:2011kk}. Daehler (1998) distinguished ‘invasive’ introduced plants as introduced species that have self-sustaining populations in natural, undisturbed areas (in contrast to agricultural lands, urban areas, roadsides, etc.) {Daehler:1998wt}. Because only uninhabited, undisturbed islands were included in this study, all introduced species are defined as ‘invasive’. Here, invasive status was determined by multiple floras – the Flora of North America {Committee:1993wf}, the Flora of the Pacific Northwest {Hitchcock:1973vc}, the Jepson Manual {Baldwin:2012vc}, the Flora of British Columbia {EFloraBCElectron:2013va}, and the Oregon Flora Project {OregonFloraProject:2014tp}.

**Phylogeny estimation**

Recent studies testing hypotheses of community assembly have incorporated phylogenetic distances using either a Phylomatic supertree {Strauss:2006tt, Carboni:2012wo, Ordonez:2014gp} or a mega-phylogeny approach {Schaefer:2011kk, Park:2013dx}. The Phylomatic method prunes species that were absent from the community under investigation from a larger summary phylogeny, collapses intraspecific taxa into species to avoid pseudoreplication, and collapses species into polytomies at the genus level {Webb:2005cx}. In contrast, the mega-phylogeny approach builds a community phylogeny from a supermatrix of DNA sequence data {Smith:2009go} and incorporates uncertainty in topology and branch lengths using statistical phylogenetic methods. With the constant expansion of readily available sequence data, the advantages of the mega-phylogeny approach to macro-ecological studies are becoming clear {Roquet:2012vt}. To assess the relationship among vascular plants in the San Juan Islands community, we took a supermatrix approach to infer a mega-phylogeny.

Using the total list of species occurrences across all of the 80 uninhabited San Juan Islands that were sampled, GenBank release 195 was queried to retrieve publically available sequence data using the PHLAWD pipeline {Smith:2009go}. Five gene regions (*atpB*, *rbcL*, *matK*, *trnTLF* and ITS) that have been frequently used in plant phylogenetic studies were included, allowing us to compile sequence data from regions with varying rates of evolution that resolve relationships at both deep and shallow phylogenetic scales. When multiple sequences were retrieved for the same species, the longest sequence was retained.

For each gene region, the alignment from the PHLAWD output was cleaned using phyutility {Smith:2008gb} to remove sites that were missing from >50% of the taxa. Gene trees were inferred under maximum likelihood (ML) as implemented in RAxML version 7.4.2 {Stamatakis:2006de}, using the GTR-CAT model of nucleotide evolution, and the simultaneous rapid bootstrap and ML search algorithm with 1000 bootstrap replicates. Rogue and outlier taxa were identified using the program RogueNaRok {Aberer:2011vh} and by visual comparison of topology to accepted relationships (e.g. {Soltis:2011di}). Sequences identified as GenBank submission errors (misidentified, mislabeled, and/or messy) were removed, alignments for each gene region were cleaned again, and then concatenated into a final alignment using phyutility {Smith:2008gb}. A final by-gene partitioned ML estimate of species relationships for all vascular plants in the San Juan Island species pool was inferred using the concatenated alignment, implemented in RAxML version 7.4.2 {Stamatakis:2006de}.

Using a tree with branch lengths proportional to time is the standard for conducting phylogenetic comparative analyses. However, because generating a timetree at this scale is a non-trivial task, other large-scale comparative analyses have used trees with relativized branch lengths that do not incorporate real time estimates {Ackerly:1999wk, GarlandJr:2000by, Webb:2000dy}. This is less than ideal, and makes it difficult to compare metrics across studies. In contrast, time-scaled branch lengths place communities on the same axis making such comparisons possible. Because the focus (i.e., community level) and scale (i.e., vascular plants) of this phylogeny is inappropriate for direct divergence time estimation using fossil calibrations, the mega-phylogeny estimate was scaled to time using the Congruification approach {Eastman:2013bh} with the function ‘congruify’ in the R-package Gieger v. 2.0 {Pennell:2014jf}. This method resolves topological inconsistencies between two trees with the aim of mapping dates from a timetree to concordant nodes in an unscaled tree. We used the detailed divergence time analysis of the Soltis et al. (2011) {Soltis:2011di} angiosperm phylogeny from Zanne et al. (2014) {Zanne:2014kg, DatafromThreekey:wf} as the reference timetree to provide dates for mapping to the nodes in the San Juan Islands community phylogeny. Dates from concordant nodes provided the maximum and minimum age constraints that were applied to each calibration point, and penalized-likelihood rate smoothing {Sanderson:1997tk}, as implemented in treePL {Smith:2012dn}, was used to generate a distribution of time-scaled estimates of the community mega-phylogeny with a smoothing parameter set to 10 after optimization on the maximum likelihood tree.

**Functional traits**

We compiled data for five quantitative traits—specific leaf area (fresh area/dry mass) {Ordonez:2010km, Royer:2012cs}, maximum height {Ordonez:2010km}, leaf size {Royer:2012cs}, seed mass {Ordonez:2010km}, and leaf nitrogen content {Royer:2012cs} (DRYAD\_1). Published data sets were supplemented with additional records for height gathered from the Electronic Atlas of the Flora of British Columbia {EFloraBCElectron:2013va} by taking the harmonic means of all values for maximum height that were found for each species.

In addition to being well sampled across land plants, these five traits have gained acceptance as measurable proxies to link variation in plant phenotypes to ecological strategies {Westoby:2002ft, Reich:2003jy, Cornwell:2014dx}. Specific leaf area (SLA), maximum height, and seed mass comprise the L-H-S scheme, widely used to relate interspecific trait variation to tradeoffs in ecological strategies proposed by Grime {Westoby:1998ug, Grime:1974ur}. For example, shorter plants with thinner, narrower leaves (low SLA) fall toward the “stress-tolerant” pole of ecological strategies. Seed mass describes a tradeoff between the dispersal capability and longevity of small seeds, and the establishment success of large seeds, balanced by reproductive effort {Westoby:2002ft}. SLA and nitrogen content together comprise the “leaf economic spectrum” in which species near the high end with large, broad leaves and high nutrient content tend to have a high growth rate and are good light competitors, but more susceptible to herbivory {Westoby:2002ft, Wright:2004jb}. Values for each trait vary greatly across the Tree of Life, and they have been shown to be correlated with major functional shifts across clades {Cornwell:2014dx}. Therefore, the ecological relevance of these five traits implicate them as being important for community assembly {Kraft:2007dv}, and they have been used to explain ecological similarity of the native community and functional uniqueness of invasives in previous ecophylogenetic tests of DNC {Schaefer:2011kk, Carboni:2012wo, Ordonez:2014gp}.

**Statistical analyses and test of Darwin’s Naturalization Conundrum**

Various distance-based metrics have been widely used to identify non-random phylogenetic structure in community composition and to summarize phylogenetic patterns {Hardy:2008cf, Hardy:2008bz, Webb:2002be}. Previous tests of DNC have used a variety of different metrics to summarize phylogenetic and functional diversity (Table 1), although most have calculated some measure of a ‘nearest taxon distance’, capturing interactions between close relatives (‘tip effects’) and a ‘mean pairwise distance’, describing community-wide interactions (reviewed in {Thuiller:2010dc}). We tested Darwin’s Naturalization Conundrum by measuring both nearest-taxon and mean diversity for phylogenetic and functional traits.

Because we were specifically interested in the relationship of invasives to the nearest native species, we calculated the distance to the nearest native species (MNNPD) following the metric described by Thuiller et al. (2010), and similar to those used in Schaefer et al. (2011) and Ordonez (2014). Calculated for each invasive (MNNPDi) and native species (MNNPDn), MNNPD is the minimum phylogenetic distance (difference in branch length) to the nearest native relative. As such, MNNPD captures tip effects, or interactions between close relatives, and directly tests DNC. Further, we examined the phylogenetic distinctiveness of the invasive species to the average native community. We calculated the phylogenetic distance from each species to the mean of the phylogenetic distance of all the native species in the community for each invasive (MPDi) and native (MPDn) species. This allows assessment of the general importance of phylogenetic similarity to the native community for invasion success.

To test if invasive plants are functionally unique from the native community in which they have naturalized, we first compared observed trait values for each ecologically relevant functional trait between invasive and native status groups on each island.Similar to phylogenetic distinctiveness, we assessed the functional distinctiveness of each species to its nearest native relative as the nearest native functional difference (MNNFD), or difference in trait values between each species and its nearest native relative (identified by the MNNPD). When there were multiple equally distant nearest native relatives, the median trait value was used to calculate the MNNFD. Finally, we calculated the mean functional difference (MFD), the difference in trait values between each species in both status groups and the mean trait value of all co-occurring natives in each community.

For each island community, the dated ML phylogeny of the vascular flora of the San Juan Islands was pruned to include just species co-occurring on the focal island (i.e. community). Metrics to describe phylogenetic (MNNPDi / MNNPDn ; MPDi/ MPDn) and functional distinctiveness (MNNFDi / MNNFDn ; MFDi/ MFDn) were calculated for each species. For each metric, differences in observed means between each status group were compared using an unpaired t-test across each island. Then to assess DNC, we compared the direction of the means for each significant relationship. For example, if the mean metric for invasive species (e.g. MNNPDi) was greater than the mean metric for native species (e.g. MNNPDn), we would conclude distinctiveness of the invasives from the native(s) to be important for the successful naturalization of invasive species in that community.

Finally, we tested if the observed patterns between invasive and native species were significantly different from that given any other random invasive species that could possibly colonize each community. We randomly replaced invasive species in each community with others from the pool of all invasive species occurring in the San Juan Islands without replacement, while preserving the observed number of invasives on each island. This null model maintains the evolutionary structure of the native community and the richness of the invasive species within the community, but randomizes the identity and the evolutionary distances of the invasive species. For each simulated community the mean of each metric was recalculated, and the randomization was repeated 1000 times to obtain a null distribution of means. The observed means were compared to each null distribution by measuring the standardized effect size (SES, similar to that described in the R package picante {Kembel:2010ft}, and equivalent to -1 times NRI or NTI described in {Webb:2002be}), which is the difference in the observed and expected mean under the null model, divided by the standard deviation of the distribution of the null. Positive values of the SES*metric* indicate that observed invasive species are more distinct (phylogenetically or functionally) from the nearest native or native community that any other random invasive species in the species pool (phylogenetic overdispersion), and negative values indicate that invasives are more similar to the nearest native and/or native community. Statistical significance was determined for each SES*metric* by calculating a *P* values, the proportion of simulated means that were as or more extreme than the observed mean {Kembel:2010ft}. All statistical analyses were conducted in R {RAlanguageanden:wf}. Custom R scripts for all analyses are available on Dryad and Github (DOI XXXXX) (DRYAD\_0).

**RESULTS**

**Island communities**

Voucher specimens were deposited in the Burke Museum of Natural History and Culture Herbarium at the University of Washington (WTU), databased via pnwherbaria (http://www.pnwherbaria.org), and comprehensive lists were compiled of all species occurring on 80 uninhabited islands ranging from 1m2 – 0.57km2 (2.725784 km2 total). Where possible, nomenclature follows Flora of North America {Committee:1993wf}. A total of 442 vascular plants were identified and are included in the Floristic Atlas of the San Juan Islands, Washington, which is publically available online

(<http://biology.burke.washington.edu/herbarium/resources/sanjuanatlas.php>). Intraspecific taxa were collapsed to species to avoid pseudoreplication, resulting in a total of 415 vascular plants across all islands used in this study (DRYAD\_1), 150 (30%) of which were identified as invasive (DRYAD\_2). Islands where no sequence data was retrieved from GenBank (Little Oak 2, Swirl Rock East), in which there was only one native (Smallest unnamed island by Long Island), or no invasive species (East Sucia 5 Island, Shag Reef, Smaller Island near Charles, Swirl Rock West) were removed, resulting in a total of 73 islands used in the following analyses (DRYAD\_3).

**Phylogeny estimation**

After removal of GenBank sequences identified as incorrect, genetic data for 366 / 415 (88%) species were retained, with a total combined aligned length of 5745 bp after cleaning to remove sites with <50% coverage (Appendix S1). GenBank accession numbers for sequences that were used for phylogeny estimation are available from the sequence alignments on Dryad (DRYAD\_4). Because there were no major conflicts among the separate gene trees (Appendix S2), the concatenated mega-phylogeny was used for all downstream analyses (Fig. 1, see Appendix S0 for labeled tips). Highly supported and ‘congruified’ nodes were distributed evenly across the tree, resolving all deep relationships with high support and very few inconsistencies at the genus level. For example, within Asteridae (BS = 96%), the Ericales are sister to the subclade Lamiidae + Campanulidae (BS = 98%), consistent with established angiosperm relationships (e.g. {Soltis:2011di}, Appendix S3).

**Functional traits**

Visual inspection of trait values that were recovered for each species mapped across the tips on the phylogeny showed good coverage for most traits across the tree (Fig. 1). Trait values were not normally distributed, and were therefore log transformed so extreme values did not drive results {Westoby:1998ug}. In cases where there were two trait values for a species (e.g. for intraspecific taxa), the harmonic mean value was used. Maximum height, seed mass, and SLA have data for more than 55% of species overall and more than 60% of the species within most islands, while leaf size and leaf nitrogen content have less complete coverage (Appendix S4).

**Phylogenetic relatedness patterns**

Results from our phylogenetic analyses indicate that invasive species have closer native relatives than the native species in each island community (Fig. 2a; MNNPDi < MNNPDn), and this difference is significant for 40 (54.05%) of the islands (Figure 4a, Appendix S6). This pattern switches at the scale of the whole San Juan Island archipelago, and we found that invasive species are phylogentically distinct, however this difference was not significant (*P* = 0.7569). The observed mean MNNPDi was less than expected under our null model, indicating that the invasive species occurring on a particular island are significantly more clustered to their nearest native relative than any other random invasive from the species pool on 22 (30.56%) of the islands (Fig.4b, Appendix S6). Thus, we found statistical support that invasive species are phylogenetically more closely related to their nearest native relative than the native species are to theirs. Further, comparison to the null model shows that invasives occurring on the islands are more closely related to their nearest native relative than a random invasive from the species pool, and are never more distantly related than a random invasive.

The observed mean MPDi was significantly different than the mean MPDn for 21 (28.38%) islands (Fig.2b, Fig.4a), and was observed mostly on the larger islands (20.27%). Compared to the null of the invasive species pool, 22 (30.56%) islands were more significantly evenly distributed within the overall native community than random (Fig. 4b; observed mean MPDi > null mean MPDi). On a few islands (*n* = 5, 6.94%), invasive species were clustered within the native community (observed mean MPDi < null mean MPDi), but never on larger islands (Fig. 4b). Given these results, we found that successful invasive species were evenly dispersed across the native communities of the San Juan Islands, with the exception of a few small islands were invasives where clustered in the native community as a whole.

**Functional trait patterns**

Each trait showed slightly different trends (Fig. 3a, Appendix S7), however, invasive species were determined to have different trait values when compared to native species for maximum height (61.81%), SLA (58.11%), and leaf nitrogen (41.89%). Although invasives were found to have smaller seeds and leaves than the natives across all islands, this difference was not often very significant (Fig. 5a). In general, invasive species were found to be shorter, with a higher SLA, and more leaf nitrogen content than the native species, but with similar leaf and seed sizes. Functionally, this translates to invasive plants in this system tending to be easily dispersed, short annuals with broad, thin, nitrogen rich leaves. In addition, they fall towards the ‘fast-return’ end of the leaf economic spectrum (high SLA, high nitrogen content; {Wright:2004jb}), which may indicate higher growth and photosynthetic rates compared to the native species. These patterns are similar to other global comparisons of functional traits between native and invasive plants that have also found seed mass, SLA, and maximum height to differ {Ordonez:2010km, vanKleunen:2010gz}, as well as meta-analyses that have found all traits except for leaf size to frequently differ between invasives and natives {Pysek:2007vf}.

Comparing the difference in functional trait values for each species to its nearest native relative, we found little significant difference between invasives (MNNFDi) and natives (MNNFDn) (Fig. 3b, Fig.5a). The means of MNNFDi and MNNFDn were most noticeably different for maximum height (28.17% of the islands) and SLA (16.39% of the islands). However, only seed mass and maximum height indicated significant functional similarity (i.e. clustering) for invasives to their nearest native relatives when compared to any random invasive in the greater San Juan Island species pool (Fig.5b; 18.01% and 9.86% of the islands, respectively). For the other traits (e.g. SLA, leaf size, leaf nitrogen content), differences in trait values for invasives and their nearest native relative were not significant between means or compared to the null expectation on most islands (Fig. 5b), which may be due to the amount of missing data for these traits (Appendix S4, Appendix S5).

The mean functional difference showed similar trends as the nearest native functional difference. For seed mass and maximum height, the means of MFDi and MFDn were significantly different (on 15.07% and 23.94% of the islands, respectively, Fig. 5a). These differences were maintained when compared to the random invasive species pool, where observed invasives were more significantly similar to the average value for seed mass across the native community than any random invasive on 23.61% of the islands, and were similar to the native average for maximum height on 16.90% of the islands. Because leaf size and nitrogen content have data for less than 60% of the species across most of the islands, comparisons between invasives and nearest natives for these traits are likely less meaningful.

**DISCUSSION**

Contrary to many other phylogenetic tests of DNC in plant systems {Strauss:2006gr, Schaefer:2011kk, Carboni:2012wo, Park:2013dx}(Table 1) our analysis of the flora of the San Juan Islands shows that invasive plant species have close native relatives in these island communities (i.e., non-random phylogenetic clustering)—rejecting “Darwin’s Naturalization Hypothesis.” Differences in mean MNNPDi and MNNPDn were significant across the majority of the islands, and when the observed mean MMNNPDi was compared to the null distribution randomizing invasives across the islands, only support for phylogenetic similarity to the nearest native was found. Less of the islands showed a significant difference in means between MPDi and MPDn, and we observed the opposite relationship comparing invasives to the null hypothesis—observed invasives are mostly evenly distributed within the native community as a whole, when compared to a random expectation of the possible invaders. Thus, the invasive flora of the San Juan Islands exhibits a nested structure of relatedness, where successful invaders are closely allied with natives at the tips, yet are spread evenly throughout the evolutionary history of the native flora, rather then clustered within one lineage.

Given the phylogenetic similarity between invasive plants and their close native relatives, we hypothesized habitat filtering to be the driving force in community assembly, and expected that invasive species should also be functionally similar to the natives because of the importance of pre-adapted traits for colonization of the new habitat. Despite this simple prediction, we see complex patterns emerging with the inclusion of functional traits. Invasive species often differ from natives in functional traits (e.g. {Pysek:2007vf, vanKleunen:2010gz}), and both experimental {Pearson:2011ee} and observational studies {Rejmanek:1996vw, Rejmanek:1996tq, Pysek:2006dw, Ordonez:2010km} have shown that many traits are associated with invasive plants. Consistent with other studies investigating trait differences between invasive species and the native flora they invade (Table 1), observed mean trait values were significantly different between the native and invasive species on many islands, at least for a few ecologically relevant traits (maximum height, SLA, and leaf nitrogen). Yet, when accounting for phylogenetic history in functional trait comparisons, we see multiple patterns emerging.

First, the absence of significant differences between observed mean trait values does not necessarily negate the importance of shared similarity to close native relatives or the native community given the invasive species that were observed to occur. Invasive and native species were shown to have similar sized seeds overall, similar seed sizes as their nearest native relative (mean MNNFDi = mean MNNFDn), and to the mean seed size of the native community (mean MFDi = mean MFDn). But when compared to the randomized invasive community, more islands show significant similarity (clustering) in seed size of the occurring invasives to both the nearest native and the native community, highlighting the importance of comparing observed relationships to biologically informed null expectations (see Thuiller et al. 2010 for a discussion of null models relevant to DNC). Thus, although there was no significant difference in means, the invasive species that occur within each island have more similar seed mass to the natives than random. Together with overall phylogenetic similarity to the native community, this suggests support for the “matching hypotheses” for seed mass.

Second, when invasive and native species differed in their observed functional traits, invasive species were not necessarily functionally distinct from their nearest native relative or the native community. This is especially apparent for SLA, where we found a significant difference in the observed mean SLA for 43 out of 74 islands, but only 10 islands showed a significant difference in SLA compared to the nearest native, which was not significantly different from any other random invader across the San Juan Islands (Fig. 4). A similar pattern was observed for maximum height. Therefore, our results suggest that for maximum height and SLA, only the disparity in trait values matters for successful invasion, and the identity of the invader relative to the native community is obsolete. This pattern suggests support for the “filling hypothesis”—invasives are closely related to the native community overall to pass an environmental filter, but certain traits differ to fill available niche spaces, escape competition, and persist over time.

Considering the entire archipelago as one community, the phylogenetic pattern changes at the tips, and invasive species are more distantly related to their nearest native, although not significant (Fig. 2a). Further, we found that invasives observed to occur across the islands were more significantly phylogenetically clustered (similar) to their nearest native on smaller islands, and more evenly dispersed within the native community on medium sized islands compared to a random invasive (Fig. 3b). The effect of spatial scale on phylogenetic patterns of invasiveness has been previously noted in the literature {Carboni:2012wo}. Given that different ecological and evolutionary processes will likely vary in strength across different scales this pattern is understandable. At small spatial scales (i.e. small islands), it is possible that more extreme abiotic pressures, such as tidal flooding, or nutrient limitation, are creating a stronger effect of environmental or habitat filtering. The higher proportion of islands that show significant phylogenetic similarity to the nearest native and the native community on small islands may signify that conservation of ancestral traits is important to cope with these extreme conditions. As island size increases, the landscape is more likely to become heterogeneous, increasing the complexity of both biotic and abiotic interactions. In this study, we treat small and large islands as comparable communities. In actuality, larger islands are composed of many different communities—e.g. mixed conifer forest, meadow, beaches. It may be that if communities were defined by environmental characteristics such as habitat type (e.g. grassland, rock outcrop) instead of geographic boundaries, we would see more homogenous patterns emerge. Future directions include exploring biogeographic and environmental factors that might explain the invasiveness of each species or the invasibility of each island in the San Juan floristic community, and disentangle the ecological processes influencing invasion dynamics in this system.

A phylogeny is often considered an appropriate proxy for ecological similarity in community phylogenetic studies. We have shown here that the phylogeny reflects functional patterns for certain traits (e.g. maximum height and SLA), or for specific comparisons (e.g. to the nearest native), but often contrasting patterns emerge (e.g. seed mass, mean community comparisons) that can only be distinguished with the combination of functional traits and evolutionary history. Although invasive and native species differ in observed functional traits, invasives share similar traits with their nearest native relative (seed mass and SLA) and with the overall native community (seed mass and maximum height). Even if overall functional distinctiveness of the invasive community suggests the importance of diversification of structure—as Darwin suggested—phylogenetic similarity to the native community along with shared similarity to their nearest native relative of a few functional traits imply that pre-adapted traits are also important for successful invasion in this system. Future work to understand the relative importance of shared vs. novel traits would be an important addition to understanding the invasion process, and community assembly more broadly.

In the era of “big data” and the growing availably of publically available molecular data, it is tempting to use phylogenetic approaches to explain patterns of community assembly and predict future invasions in the absence of additional trait data for each species. However, we show that with the growing body of trait databases and ecological surveys, function and phylogeny can be easily united to test eco-evolutionary hypothesis of species invasion, and their complementarity can aid in deconstructing the processes of invasion in natural systems. In particular, clearly defined null models that evaluate the phylogenetic and functional structure of communities enable identification of patterns that could not be generated randomly, and may indicate interesting ecological processes driving species invasions. Incorporating the growing body of diversity metrics with meaningful null hypotheses and unified methodologies into comparable frameworks has the potential to further deconstruct the patterns of assembly we observe, and predict future ecological change.

**ACKNOWLEDGEMENTS**

We would like to thank multiple contributors of the trait data used in this study: Angela Moles, Dana Royer, Alejandro Ordonez, Ian Wright, Peter Reich, Amy Zanne, and Will Cornwell. Also, thanks to Luke Harmon and Matthew Pennell for advice on statistics and programming, and Cody Hinchliff for help with the PHLAWD pipeline. H.E.M was supported by a NSF Graduate Research Fellowship Grant No. DGE-1144254 and a Bioinformatics and Computational Biology Graduate Fellowship from the University of Idaho.

**SUPPORTING INFORMATION**

Additional Supporting Information may be found in the online version of this article:

**Appendix S0** Community phylogeny of the vascular flora of the San Juan Islands with the tips labeled as taxa.

**Appendix S1** Table withnumbers of native and invasive species for each gene region, final sequence length, and percent of each gene partition in the final concatenated nucleotide matrix post cleaning.

**Appendix S2** Best maximum likelihood phylogram for each gene tree produced with RAxML HPC-PTHREADS version 7.4.2 {Stamatakis:2006de}.

**Appendix S3** Best maximum likelihood phylogram for the concatenated dataset produced with RAxML HPC-PTHREADS version 7.4.2 {Stamatakis:2006de}.

**Appendix S4** TableSummarizing of the number of species with data for each trait, and the mean, median, minimum, and maximum trait value across all islands.

**Appendix S5** Plots of species richness and percent of species with data for each functional trait per island.

**Appendix S6** Phylogenetic diversity measures (MNNPD, MPD, SES MNNPD, SES MNNPD MPD) for all islands.

**Appendix S7** Functional diversity measures (trait measurements, MNNFD, MFD, SES MNNFD, SES MFD) for all islands.

**DRYAD**

**DRYAD\_0** R scripts used in the analyses.

**DRYAD\_1** Species list with status, and trait values for each of the five traits (ADD TRAIT SOURCE)

**DRYAD\_2** Metadata for each island (total area, size category) and summary of species diversity observed and the percent of species with sequence data that were recovered from GenBank.

**DRYAD\_3** Community matrix with presence absence of each species across the 80 islands in the San Juan archipelago.

**DRYAD\_4** Alignments for each gene region and concatenated withGenBank accession numbers for each species.

**DRYAD\_5** Treefile(s): each gene region and concatenated

**REFERNECES**

{papers2\_bibliography}