Title
A test of Darwin's naturalization hypothesis in the thistle tribe shows that close relatives make bad neighbors

Permalink
https://escholarship.org/uc/item/6hc2d637

Journal
PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF THE UNITED STATES OF AMERICA, 110(44)

ISSN
0027-8424

Authors
Park, DS
Potter, D

Publication Date
2013-10-29

DOI
10.1073/pnas.1309948110

Peer reviewed
A test of Darwin’s naturalization hypothesis in the thistle tribe shows that close relatives make bad neighbors

Daniel S. Park1 and Daniel Potter

Department of Plant Sciences, University of California, Davis, CA 95616

Edited by Peter H. Raven, Missouri Botanical Garden, St. Louis, Missouri, and approved September 18, 2013 (received for review May 25, 2013)

Invasive species have great ecological and economic impacts and are difficult to control once established, making the ability to understand and predict invasive behavior highly desirable. Preemptive measures to prevent potential invasive species from reaching new habitats are the most economically and environmentally efficient form of management. Darwin’s naturalization hypothesis predicts that invaders less related to native flora are more likely to be successful than those that are closely related to natives. Here we test this hypothesis, using the weed-rich thistle tribe, Cardueae, in the California Floristic Province, a biodiversity hotspot, as our study system. An exhaustive molecular phylogenetic approach was used, generating and examining more than 100,000 likely phylogenies of the tribe based on nuclear and chloroplast DNA markers, representing the most in-depth reconstruction of the clade to date. Branch lengths separating invasive and noninvasive introduced taxa from native California taxa were used to represent phylogenetic distances between these groups and were compared at multiple biogeographical scales to ascertain whether invasive thistles are more or less closely related to natives than noninvasive introduced thistles are. Patterns within this highly supported clade show that not only are introduced thistles more closely related to natives more likely to be invasive, but these invasive species are also evolutionarily closer to native flora than by chance. This suggests that preadaptive traits are important in determining an invader’s success. Such rigorous molecular phylogenetic analyses may prove a fruitful means for furthering our understanding of biological invasions and developing predictive frameworks for screening potential invasive taxa.

Significance

Invasive species negatively impact both natural ecosystems and human society and are notoriously difficult to control once established. Thus, identifying potentially invasive taxa and preventing their dislocation is the most efficient management method. Darwin’s naturalization hypothesis, which predicts that the less closely related to native flora species are, the more likely they are to succeed as invaders, is tested here with an unprecedentedly thorough molecular phylogenetic approach, examining >100,000 phylogenies of the weed-rich thistle tribe Cardueae. Branch lengths between taxa were used as measures of evolutionary relatedness. Results show that invasive thistles are more closely related to natives than noninvasive introduced thistles, suggesting they share preadapative traits with the natives that make them more likely to succeed as invaders.
pathogens (40, 41). The enemy escape hypothesis also supports this view (42, 43). An alternative, opposing hypothesis is that relatedness to native taxa may convey degrees of preadaptation to the conditions of the invaded environment, rendering close relatives more likely to succeed once introduced (32, 39, 44).

Previous studies have been equivocal, finding evidence both for (30, 31, 38, 45–50), and against (24, 32, 41, 44, 51–55) Darwin’s hypothesis. However, few have used a strict phylogenetic approach based on evolutionary divergence, instead predominantly relying on taxonomic ranks (e.g., refs. 45 and 56, reviewed in refs. 57 and 58), which are highly subjective as measures of relatedness (59). In instances where phylogenetic trees were used, some have used supertrees compiled from multiple studies (46, 60, 61), with estimated branch lengths that may not accurately reflect the evolutionary distances between taxa. Other studies have been based on community phylogenetic trees (47, 60, 62).

This approach may be problematic, as communities are not necessarily monophyletic groups, but collections of co-occurring species whatever their evolutionary relationships may be (63), and sampling in such studies is unlikely to include adequate representation of all lineages present, whereas phylogenetic analyses assume monophyly of the ingroup and their accuracy is dependent on sampling that is representative of the diversity within (64, 65).

This study sought to address theoretical and methodological issues that may have limited progress toward resolving Darwin’s naturalization conundrum (56) by using phylogenies based on molecular markers to assess the evolutionary distances between native and nonnative taxa in a strongly supported monophyletic group, the thistle tribe (Cardueae, Asteraceae), in a well-defined biogeographic area, the California Floristic Province (CAFP) (66, 67). The thistles of California offer an ideal opportunity to test Darwin’s hypothesis. The tribe, which boasts an impressive list of Mediterranean and temperate invaders, is most prolific in Mediterranean climate regions, which not only rank among the most biodiversity-rich biomes on the planet (68, 69), but among the most imperiled as well (70–74). The CAFP is one of such biodiversity hotspots, as defined by Conservation International, and is often recognized as a biogeological entity (75, 76). Also, programs and legislation regarding invasive taxa are usually state specific, making the study of invasive Cardueae in California both politically and biologically meaningful.

Phylogenies of Cardueae based on sequences from three genomic regions commonly used in phylogenetic studies of angiosperms were generated with taxon sampling representing the full lineage diversity of the tribe, including all species native and naturalized in the C AFP. Phylogenetic trees were constructed using parsimony, maximum likelihood, and Bayesian approaches. Phylogenetic distances of invasive and noninvasive introduced species from natives were compared using a comprehensive set of metrics and statistical tests to assess the utility of phylogenetic distance from natives as a predictor of invasive behavior. This study finds evidence contrary to Darwin’s hypothesis and demonstrates the robustness of such metrics, which should be more informative and meaningful than taxonomic groupings (30, 46, 77–79), especially when supported by a well-resolved molecular phylogeny.

Results

Phylogenetic trees of 202 species spanning the entire diversity of Cardueae, including the 51 species that occur in the CAFP, were constructed using Bayesian inference, maximum likelihood, and parsimony analyses based on two different combinations of three genomic regions: the internal transcribed spacer 1, 5.8S rRNA gene, and internal transcribed spacer 2 (nDNA, ITS); the trnL intron, the 3′ trnL-exon, and the intergenic spacer between trnL and trnF (cpDNA, trnL-trnF IGS); and maturase K (cpDNA, matK), per Sussanna et al. (80). Of the 73 recognized genera (81), 61 were represented. The issue of whether wider sampling of taxa is more important to accurate phylogenetic reconstructions than wider sampling of characters has been debated (82). Hence, two datasets, one with 165 taxa represented by three markers and another with 202 taxa represented by two markers, were examined. The chloroplast matK region was excluded in the two-marker dataset, as it contained the least number of variable sites.

The total number of possible topologies examined for both datasets and three tree building methods combined was 112,626. These trees represent the most thoroughly sampled phylogenetic reconstructions of the tribe Cardueae to date. The Bayesian consensus tree in Fig. 1 illustrates the phylogenetic distribution of species of the four possible biogeographic and ecological categories used in this study (invasive, noninvasive exotic, native, and not present in CAFP), and is largely congruent with previous studies, including the paraphyly of subtribe Carduinae and strong support for the monophyly of Centaureinae as traditionally circumscribed (80, 83). Nonnative (introduced) species were classified as either invasive or noninvasive based on the California Invasive Plant Council invasive plant inventory (84).

The mean phylogenetic distance between each introduced species and all native species (MPD) and that between each introduced species and its nearest native relative (MNND) were calculated separately for invasive and noninvasive introduced species, and the results were compared for all 112,626 trees. Invasive species were significantly more closely related to the
native community than were noninvasive exotic species in all topologies examined \( (P < 0.05) \). Invasive species also tended to have significantly closer nearest native relatives than noninvasive exotics in 99% of the trees tested (Table 1). Also, Cardueae species not occurring in CA were found to be significantly less related to the native CA thistle community than were invasive nonnative CA Cardueae in all cases (Table S1).

To further examine the nature of the nonnative community of thistles in CA, the net relatedness index (NRI), a standardized measure that quantifies the degree of clustering of taxa over the entire phylogeny, and nearest taxon index (NTI), a measure of nearest distance that quantifies the clustering of the terminal nodes (85, 86), were calculated as follows:

\[
NRI = - \left( \frac{MPD_{obs} - MPD_{rndm}}{SdsMPD_{rndm}} \right)
\]

\[
NTI = - \left( \frac{MNND_{obs} - MNND_{rndm}}{SdsMNND_{rndm}} \right)
\]

where MPD_{obs} and MNND_{obs} are the mean phylogenetic distance and mean nearest neighbor distance between introduced species and native species of the observed original data and MPD_{rndm} and MNND_{rndm} are those of 10,000 randomly generated communities. SdsMPD_{rndm} and SdsMNND_{rndm} refer to the SDs of the MPDs and MNNDs of the random assemblages. \( P \) values for the calculated NRI/NTIs were obtained by dividing the number of random assemblages with means greater than or equal to observed by the number of randomizations + 1 (85). Thus, positive NRI/NTI values indicate phylogenetic clustering, and negative values indicate overdispersion, relative to the native community. Positive values were retrieved in all instances for invasive taxa, indicating that invasive taxa were more closely related to their respective nearest native relatives, as well as to the native thistle community as a whole, than would be expected by chance. Results were statistically significant \( (P < 0.05) \) in nearly all of the topologies investigated (Table 1). In contrast, noninvasive exotic taxa were not significantly closer to native species than by chance.

To examine Darwin’s naturalization hypothesis at a local and possibly more ecologically meaningful scale (57), the aforementioned analyses were replicated at the level of individual bioregions, using the phytogeographic boundaries as defined in The Jepson Manual (Table 1) (87). Again, the majority of the investigated evolutionary topologies resolved invasive species as more closely related to the natives than are noninvasive exotic taxa. Despite smaller sample sizes, in most bioregions, invasive species were significantly more closely related to native species than by random chance as well. However, these patterns were not as well supported in the Great Central Valley as in the other bioregions, which generally represent coastal and mountainous areas of California.

These results suggest that, within the thistle tribe, introduced taxa are more closely related to the native community and/or have closer native relatives are more likely to become invasive than those that are more distantly related to natives. The robustness of this pattern is demonstrated by the fact that it was upheld at multiple geographic scales, across various tree-building approaches.

**Discussion**

In the midst of a plethora of conflicting results and methodologies, some have argued that predicting which species will be invasive is...
nigh impossible (11); however, there may exist an optimal scale and scope of analysis at which the prediction of invasiveness is feasible, focusing on meaningfully circumscribed taxonomic and biogeographic entities (88). The phylogenetic scope of this study, encompassing a highly supported clade of 202 species within 64 genera, allows for circumvention of some of the problems that might occur in similar studies aimed at testing Darwin’s naturalization hypothesis at lesser or greater taxonomic scales. Examining the relationships between introduced species and native species where the community is composed of taxa classified in multiple families (47) and possibly separated by over a hundred million years of evolution (89), may not be ecologically meaningful, as it becomes difficult to assume ecological similarities between such taxa are the result of shared evolutionary history. Also, at such a large scale, it is difficult to construct well-sampled, robust phylogenies of an all-encompassing monophyletic group.

At the other extreme, examining the relationships among species within a smaller clade, as might be represented by a single genus, would likely provide results whose applicability to other situations would be quite limited. The scale used here provides a reasonable middle ground where taxa maintain a relatively high level of genetic and structural uniformity, yet are diverse enough to display a wide range of ecological adaptations.

The reliability of such comparative analyses depends on whether the phylogenetic scale used accurately reflects the true evolutionary history of the taxa involved. However, phylogenetic reconstructions are prone to error and uncertainty, and most studies currently use one or few working phylogenies (90). Here, by examining a large number of likely evolutionary scenarios generated by a range of methods, it was possible to deal, to an extent, with the problem of phylogenetic uncertainty.

Results of this study suggest that introduced species with highly negative ecological impacts are phylogenetically closer to their native counterparts than are largely benign, noninvasive introduced species. Together with the finding that these invasive species are more closely related to native taxa than to other species, this suggests that preadaptive advantages, evolved over shared history and shared through common ancestry, may outweigh the importance of enemy escape or competitive exclusion, at least in certain stages of biological invasions. This trend mirrors those found in studies using taxonomic ranks and supertrees at large continental scales (44, 55, 91, 92) but has rarely been observed at local community scales (61).

Although invasive species tended to be closer to both the native chrysanthemum. Interestingly, although further support was found when examining the species compositions of natural reserves in the University of California Natural Reserve System for which species lists are publically available (93). Of the 14 reserves in which both native and invasive thistle species co-occur, 12 had at least one invasive species in the same genus with a native, supporting the patterns found at larger scales. No noninvasive introduced thistles were found in the reserve system. However, the Great Central Valley of California seems to be an exception; although this region is far from being the smallest, or most thistle-free in region, little support was found for phylogenetic relatedness determining the success of introduced taxa. This result is likely due to the fact that the Central Valley is predominantly farmland and grazing land for livestock and is almost entirely in private ownership (94). Very little land is left unmanaged, and remaining natural habitats are mostly small and fragmented. It may be impossible to observe ecological patterns underlying biological invasions in such areas.

Only a small fraction of the many exotics that have been introduced to alien habitats over the years have become highly invasive (93). Clearly, the net outcome of a plethora of opportunities and obstacles introduced plants are faced with will determine whether their populations ultimately thrive against, or are suppressed by, the native community (2). As demonstrated here, a phylogenetic framework can represent such interactions by addressing the shared evolutionary history of species. Thus, the evolutionary relatedness of exotic taxa to natives can provide information for identifying threats to native communities, as well as extending our understanding of why certain introduced species prove to be more invasive than others. This study suggests that monitoring and regulating exotic species that are closely related to native taxa but not yet introduced or escaped should be a priority.

Methods

Study System. The Asteraceae, the largest eudicot family and possibly the largest angiosperm family, represents 8% of all flowering plant species and, although most prominent in drier, Mediterranean climates, can be found on all continents except Antarctica (95). While comprising many important crop and horticultural species, the family also includes a disproportionately high global representation of invasive species (96, 97). Of specific interest is the thistle tribe, Cardueae, one of the largest tribes of Asteraceae, with ~2,500 species. Previous molecular studies have unanimously confirmed Cardueae as monophyletic (80, 95, 98–101). The tribe comprises roughly half of the invasive Asteraceae species in California, as well as many natives and endemics (102). Native and nonnative introduced taxa were delimited according to the second edition of The Jepson Manual (87).

Collecting Materials. Sampling was based on previous systematics studies (80, 83, 101) to represent most of the genera and major clades of the tribe. The tribe comprises roughly half of the invasive species in California, as well as many natives and endemics (102). Native and nonnative introduced taxa were delimited according to the second edition of The Jepson Manual (87).

DNA Extraction, Amplification, and Sequencing. DNA extractions were performed using Qiagen miniprep kits. The ITS region was amplified and sequenced using primers ITS5 and ITS4 (103) and separately using combinations of ITS2 and ITS5, and ITS3 and ITS4, respectively (104). The trnL intron, the 3′ trnL exon, and the intergenic spacer between trnL and trnF were amplified and sequenced together. The universal primers trnL-D and trnL-E were used for amplifying the trnL-trnF IGS region. In some cases, combinations of trnL-D and trnL-E were used together as well (105). The first 1,000 bp of the 5′ end of the gene maturase K (matK), which account for most of the variability found in the gene (106), were amplified with the primers trnK-710 F (107) and AST-1R (101). In some cases, combinations of the primers matK1F, matK1R, matK2F, and matK2R were used as well (99). PCR products were run on a 1% agarose gel. Identifiable bands were cut out and purified with Qiagen Gel Extraction Kits. Sanger sequencing of the purified PCR products was performed on ABI 3730 Capillary Electrophoresis Genetic Analyzers at the University of California–Davis Center for Plant Diversity. Relevant sequences were also downloaded from GenBank. Species were chosen to represent the entire diversity of the tribe and achieve sufficient density of taxon sampling within that monophyletic group. To minimize inaccurate placements of taxa, sampling of hybrids was avoided. The origins of the samples and their GenBank accession numbers are listed in Table S2.

Table S2
Phylogenetic Analyses. The ITS region, as well as the chloroplast matK gene and the trnL-F region, was chosen not only because of their widespread use in phylogenetics, but also because the combination of relatively fast-evolving regions and slower ones was necessary to provide the required deep and shallow level resolution for performing phylogenetic analyses at the species level. The combination of these markers has been used in previous studies of the group (80, 83), and Shimodaira-Hasegawa tests (108) were performed in PAUP* (109) to confirm lack of significant conflict in phylogenetic signal between nuclear and plastid regions (Table S3). Sequences of each region were aligned independently with ClustalX (110), edited further by hand using MEGAS (111), and combined into a single matrix. Multiple phylogenies of the tribe Cardueae were estimated based on two combinations of the three aforementioned markers: ITS and trnL-F (n = 202) and ITS, trnL-F, and matK (n = 165). To minimize error and bias in taxon placement and branch length calculations, all taxa in the phylogenies were represented by all markers, with 15% missing data. In addition, to investigate a full breadth of possible evolutionary scenarios, three different tree building methods were used as described below.

Garli 2.0 (112) was run on the Cipres Science Gateway server to generate maximum likelihood (ML) phylogenies. The general time reversible (GTR + I + F) (113) model was used with default settings. Four parallel runs were performed to ensure that the resulting tree was not lodged on a local optimum. Maximum parsimony phylogenies were conducted with PAUP* (114) through the Cipres Science Gateway. Analyses were run with default settings with TBR branch swapping for 43,000 Ratchet repetitions and 200 Ratchet iterations for the ITS + trnL-F dataset and 50,000 repetitions and 200 iterations for the ITS + trnL-F + matK dataset, yielding 19,092 and 13,528 most parsimonious trees, respectively. Maximum likelihood trees and bootstrap values are presented in Figs. S1 and S2. Bayesian inference analyses were carried out using MrBayes 3.2.1 (115). Models of molecular evolution were evaluated with JMODELTEST (116); the best fit models were GTR + I + F for ITS and GTR + F for plastid regions, based on both the modified Akaike information criterion and Bayesian information criterion. Analyses were run under default settings for 13 million generations, sampling two Markov chain Monte Carlo (MCMC) chains every 500 generations. A total of 40,002 trees were saved for each dataset after discarding 23% as burn-in.

Investigating Darwin’s Naturalization Hypothesis. To assess whether exotic invaders tend to be more closely related to their native relatives, the MPDs from each nonnative taxon to all native taxa, as well as the distance from each nonnative taxon to its nearest native relative, i.e., the MNNDs, were calculated using the ICOMDIST function in phylocm (85). These two metrics were compared using a t-test in R (117). To determine the degree to which variance in tree topologies affects these results, this process was repeated for all of the phylogenies generated by the aforementioned methods using a custom Perl script, for a total of >100,000 trees. Identical analyses were also conducted at the bioregion level as defined by The Jepson Manual (87), to seek a more focused view of species interactions.

Custom R scripts were used to calculate NRI and NTI across all phylogenies. In accordance to the independent swap null model, nonnative taxa were randomly sampled to generate random exotic assemblages of equal species richness as observed communities (118). A total of 10,000 random assemblages were generated per tree.

ACKNOWLEDGMENTS. We thank E. Dean and J. Shepard of the University of California–Davis Center for Plant Diversity for assistance with sampling herbarium material for DNA barcoding. V. G. Celeti, D. V. Baladandar, and S. Pecenini for invaluable help collecting in Italy; I. Korf and K. Battenberg for instructive discussions on Peri programming; and D.S.P.’s dissertation committee members for insightful comments on the project and manuscript. This work was supported by funding from Davis Botanical Society and the National Science Foundation Grant DEB1210526.

2. Colautti RI, MacIsaac HJ (2004) A neutral terminology to de