

Title page

Title: Loss of mutualistic interactions loosens restrictions on evolutionary history in legume plant communities

Authors: Anna Simonsen¹ and Russell Dinnage¹

Affiliations:

1. Institute of Environment, Florida International University, Miami, Florida, USA.

Email: Anna Simonsen: asimonse@fiu.edu; Russell Dinnage: rdinnage@fiu.edu

Running Title: Mutualism affects community phylodiversity

Keywords: evolutionary history, phylogenetic structure, mutualism, legumes, Fabaceae, community ecology, community assembly, invasion biology

Type of Article: Letter

of Words - Abstract: 148; **Main text:** 4960

of references: 49; **# of figures:** 4; **# of tables:** 2; **# of text boxes:** 0

Corresponding Author: Anna Simonsen **email:** asimonse@fiu.edu **address:** 11200 SW 8th Street OE-148, Miami, FL 33199 **phone:** +1 786 710 6326

Statement of authorship: AS and RD collected and analyzed the data, and wrote the manuscript.

Data accessibility statement: Should the manuscript be accepted, the data supporting the results will be archived in FigShare and the data DOI will be included at the end of the article.

Abstract

Mutualistic interactions are increasingly recognized as playing important roles in community assembly. We hypothesized that mutualisms can influence the accumulation of evolutionary history within communities through indirect interactions, which we investigated by quantifying the impact of mutualism gains and losses on phylogenetic structure in the *Fabaceae* family. Analyzing global distribution data, we find that legumes lacking mutualistic interactions exhibit reduced phylogenetic clustering, resulting in higher phylogenetic diversity in regions richer in non-mutualistic legumes. Moreover, the probability of a plant species being introduced to a new range is negatively related to phylogenetic distance to its nearest native relative, but this effect is weaker for species without mutualistic interactions. These findings highlight the significant role of mutualism in restricting the local distribution of evolutionary history at a global scale. Our study advances community assembly theory and underscores the importance of considering mutualism in the conservation and restoration of phylogenetic diversity.

Introduction

Mutualistic interactions, where both species involved derive a benefit, are a fundamental aspect of biodiversity and ecosystem function (Bronstein 1994). These interactions can significantly influence the evolution and distribution of species, shaping community structure and dynamics. They can drive speciation, influence community assembly, and affect ecosystem stability and resilience (Bertness & Callaway 1994; Thompson 2005; Bascompte & Jordano 2007). However, the intricate nature of these interactions and their potential implications for the distribution of evolutionary history within communities remain poorly understood. One hypothesis we will address in this study posits that mutualism could restrict the accumulation of evolutionary history within a community, and will occur if closely related species are more likely to share mutualistic partners and that these partners are a locally limiting factor (Cavender-Bares *et al.* 2009). This theory predicts that mutualistic interactions could lead to phylogenetic clustering in communities, where closely related species are more likely to co-occur than expected by chance. However, empirical tests of this theory, particularly at a global scale, have been limited, leaving a significant gap in our understanding of the role of mutualism in shaping biodiversity patterns. Filling this gap is important because of the established significance of evolutionary history to community assembly (Webb *et al.* 2002; Kraft *et al.* 2007; Cavender-Bares *et al.* 2009; Mouquet *et al.* 2012; Davies 2021), ecosystem function (Maherali & Klironomos 2007; Cadotte *et al.* 2008, 2012, 2017; Flynn *et al.* 2011; Srivastava *et al.* 2012; Cadotte 2013), trophic cascades (Dinnage *et al.* 2012; Dinnage 2013; Ibanez *et al.* 2016; Staab *et al.* 2021), and conservation (R. I. Vane-Wright *et al.* 1991; Faith 1992; Crozier 1997; Purvis *et al.* 2005; Isaac *et al.* 2007; Lean & Maclaurin 2016; Lean 2017).

The plant family Fabaceae presents a unique opportunity to test and explore this hypothesis. This family is characterized by a diverse range of mutualistic interactions with symbiotic soil bacteria, known as rhizobia. These interactions, which have been both gained and lost throughout the evolutionary history (Werner *et al.* 2014) of the Fabaceae, enable legumes to fix atmospheric nitrogen, providing a significant advantage in nitrogen-limited soils (Sprent & James 2007). However, mutualistic relationships may also limit the ability of legumes to establish in new regions lacking suitable rhizobia (Richardson *et al.* 2000; Simonsen *et al.* 2017; Harrison *et al.* 2018). At the same time, there is evidence that many mutualistic interactions are phylogenetically structured, with closely related species more likely to share similar mutualistic partner communities (Rezende *et al.* 2007; Gómez *et al.* 2010), as is the case between legumes and rhizobia, at least

at broad sub-family levels within Fabaceae (Andrews & Andrews 2017). If legumes are more likely to share symbionts with close relatives, this could potentially influence the phylogenetic structure of mutualistic plant communities, leading to a restriction in the accumulation of evolutionary history of legume community members engaged in mutualism, more so than for non-mutualistic plant community members.

However, this expected pattern has proven difficult to test because multiple simultaneous community processes are expected to affect phylogenetic diversity. For example, it is often assumed that competition will drive increased phylogenetic diversity due to phylogenetic conservation of traits combined with competitive limiting similarity. Likewise, it is often assumed that environmental filtering will lead to restricted phylogenetic diversity because of phylogenetic conservation of traits combined with trait-based environmental tolerances (Cavender-Bares *et al.* 2009; Cadotte & Tucker 2017). If an introduced species is closely related to an established community due to shared or similar abiotic habitat requirements, then distantly related species are predicted to be able to have reduced establishment abilities in novel communities, as is classically predicted through Darwin's Naturalization Hypothesis (Kraft *et al.* 2015; Cadotte & Tucker 2017). However, because species may also share similar biotic resource requirements as well, (i.e. a shared mutualism requirements such as a pollinator species), abiotic environmental filtering and mutualism show the same predictive effects in terms of phylogenetic distance between introduced and established community members. If this is the case, environmental filtering should be difficult to distinguish from the effects of mutualisms, since the direction of their predicted effects on community phylogenetic structure is the same (Jones *et al.* 2013). Because of the ubiquity of mutualisms in natural communities – to our knowledge – no study has explicitly been able to separate or distinguish these effects during community assembly. This is normally challenging because of the difficulty of comparing mutualistic to non-mutualistic sub-communities. Typically mutualistic and non-mutualistic organisms will differ in many other ways besides their mutualism trait, making it difficult to conclude any differences are due to mutualism per se and hence isolate the role of mutualism in either restricting or increasing evolutionary history during plant community assembly. Here we take advantage of a natural evolutionary and ecological experiment, wherein mutualistic interactions with symbiotic soil bacteria have been independently gained and lost within the plant family Fabaceae (Werner *et al.* 2014) allowing us to estimate the independent effect of mutualism within a closely related group that we expect to be otherwise similar to each other in most respects. Furthermore, hundreds of Fabaceae species have been introduced to new ranges at a

global scale (Simonsen *et al.* 2017), and knowledge of introduction status and native range information of legume species combined with species-level symbiotic status enables insight into the role of rhizobial symbiosis traits in influencing plant community assembly processes in both native and introduced regions.

Although the effects of mutualism on community assembly can be thought of as a type of environmental filtering – specifically filtering by a part of the biotic environment – nevertheless is important to distinguish the two forces because their consequences can be strikingly different due to notably different causal structures. In Figure S1 we plot proposed causal graphs showing how the effect of the presence of a native resident legume can affect the probability of another species invading, as mediated through their shared evolutionary history. In the case of abiotic environmental filtering the pathway linking the resident species to the invader species is non-causal. Correlations are induced by a shared causal relationship with the environment of a site, which is mediated through the two species environmental tolerances, which in turn, are linked by their phylogenetic similarity. On the other hand, with mutualism, the pathway linking the presence of the resident with the invasion probability of the invader is indirectly causal, because the resident species influences the presence of shared mutualistic partners, and the presence of partners then causally influences the probability of an invader establishing. This effect is mediated through phylogenetic similarity, because phylogenetic similarity influences the similarity of the species mutualistic partner compatibilities, which affects both how the resident influences the mutualistic partner community, and how the invasion probability responds to it. Figure S1 shows how the undirected structure of the two causal graphs is similar, which leads to similar predicted relationships between variables. But the direction of causal arrows makes a big difference when considering possible interventions. In the case of environmental filtering, a direct intervention on the presence of the native resident species will make no difference to the invader. The fact that the resident species is able to exist at the site provides all the information needed to predict the invader potential for establishment, whether it actually exists there makes no difference. On the other hand, with mutualism, intervening on the presence of the resident can directly affect the invasion potential of the invader. For example, if the resident species goes extinct locally, its loss can affect the presence or abundance of its mutualistic partners, which then in turn will change the probability of the invader's successful establishment. The causal structure of the mutualism mechanisms also allows the establishment of feedback loops, which is not the case for abiotic environmental filtering. So the difference

between these two factors, despite their similar patterns, is critical because it directly determines what interventions are potentially effectual.

In this study, we test how the absence of mutualistic interactions with a nutritional symbiont in legume plants affects the phylogenetic structure of legume plant communities at regional scales on a global distribution dataset. We specifically investigate two key signatures of phylogenetic structure: the degree of phylogenetic clustering of native regional legume species assemblages, and whether phylogenetic distance to a destination community can predict the probability of a legume species being introduced to a new range. We find these two independent perspectives on community assembly, normally studied in isolation, show strongly concurrent results in terms of their predicted patterns with respect to how mutualism modulates the evolutionary history of regional legume species assemblages. Our findings have implications not only for our understanding of the role of mutualism in shaping global biodiversity patterns, but also for community assembly theory and the conservation and restoration of phylogenetic diversity.

Material and Methods

Legume distribution and trait data

We used data from Simonsen et al (2017) on the introduced ranges of legumes across the world. Simonsen et al. (2017) includes details on the data collection, summarized in brief here. We collected data on the symbiotic nitrogen-fixation status and global distribution of approximately 3,500 legume species. Nitrogen-fixation status was extracted from a publicly available database compiled by Werner et al. (Werner *et al.* 2014), which categorized each species as either 'symbiotic' or not. For global distribution data, we used the International Legume Database and Information Service (ILDIS: <https://www.ildis.org>) (Roskov *et al.* 2005; Bruneau *et al.* 2019), extracting geographic distribution information for each legume species found in the nitrogen fixation database. The geographic data was then converted into usable geographic coordinate data using a shapefile containing standardized geographic regions. Additional species trait data, including plant life-form, life-history, and information on human uses, were also collected from ILDIS. Missing data for life form and life history were imputed based on taxonomic grouping. For a detailed description of the data collection methods, please refer to the original paper.

Legume Phylogeny

For the legume phylogeny, we used the most complete published phylogeny of *Fabaceae* available, a genus level phylogeny containing a total of 1,002 genera spanning all Families and covering ~80% of all described genera in the clade (Li *et al.* 2015). In this analysis we assigned all species within genera to an unresolved polytomy below the genus-level, with zero branch-lengths between them. Thus, all phylogenetic analyses should be interpreted as containing only genus-level or above phylogenetic effects. Using this phylogeny, we were able to utilize all species in our distribution dataset, minimizing the introduction of bias in response variables (see next sections). Li *et al.* has details of the phylogeny construction, but we summarize it briefly here. The phylogeny was constructed using data from 1,002 genera for three genetic markers from the plastid genome (*rbcL*, *matK*, and *trnL-F*). The phylogenetic analysis was performed using Maximum Likelihood in RAxML v7.6.6 (Stamatakis 2014). The branching events within the nitrogen-fixing clade were dated using a penalized likelihood method, with 27 fossil taxa serving as minimum calibration points. Additionally, a Bayesian relaxed clock method was used for comparison on a reduced data set of 232 taxa. For a detailed description of the phylogeny construction, please refer to the original paper (Li *et al.* 2015).

Geographic Data

Data on native and introduced ranges were based on polygon information of global legume distribution from ILDIS (see Simonsen *et al.* 2017 for details). The present or absence of species was scored for 320 geographic locales. What we call locales are based on the Taxonomic Database Working Group (TDWG) Geography Standard version 2.0 level 4 areas (ref). In addition, we use the TDWG levels 1 and 2 as higher-level hierarchical structures in our models to account for spatial non-independence in our models. We call these levels Continents, and Regions.

Statistical analysis

All statistical models were fit using the R package lme4 (Bates *et al.* 2014). Code syntax to run the models is given in the Supporting Information.

Community Phylogenetic Distance of Native Species

We modeled the mean phylogenetic diversity (MPD) of native legume species in locales using a generalized linear mixed model (GLMM). MPD of each locale was calculated in picante using mpd function (Kembel *et al.* 2010). To account for the possible effects of the structure of the legume phylogeny itself, we calculated z-scores based on the deviation of observed MPD from expected MPD under several standard null models, as implemented in picante. The null models we used (with picante null.model argument in brackets) were: 1) Richness Null Model ("richness"), 2) Tip Reshuffle Null Model ("taxa.labels"), 3) Phylogeny Pool Null Model ("phylogeny.pool"), 4) Frequency Null Model ("frequency"), and 5) Independent Swap Null Model ("independentswap"). For model 2, null model simulations were carried out within the mutualistic and non-mutualistic sub-communities separately (see below).

We modeled the Mean Phylogenetic Distance (either raw values, or z-values calculated from the null models) as a function of the proportion of the legume community that was mutualistic (model 1), or whether the sub-community was mutualistic or non-mutualistic (model 2). For model 2 we split each locale into two data points, one for the mutualistic sub-community and one for the non-mutualistic sub-community and the dataset was augmented with a factor predictor with the mutualistic status of the sub-community. We also accounted for a number of potentially confounding covariates, including the latitude of the community or sub-community's region, its area, and its native legume species richness. We also accounted for the terrestrial biome of the region as a random effect, as well as two hierarchical levels of geography (region and continent) to help account for spatial non-independence. We chose appropriate transformation for each predictor to decrease skew, which reduces high leverage points and improves alignment with model assumptions (e.g. heterogeneity of variance and linearity). See Table S1 for transformations and Model Equations in the Supplementary File.

Introduced Species

We modeled the presence or absence of an introduced species in a destination region as a function of whether it was mutualistic or not, and the phylogenetic distance from it to the native community of the destination region. Of particular interest was the interaction between the presence of mutualism and phylogenetic distance, because this tells us whether mutualism affects the strength of any effects related to phylogenetic history. We controlled for a large number of covariates including traits of the introduced species,

properties of the source and destination regions, and differences between the source and destination regions, as well as controlling for sources of non-independence at the species-level and at various geographic levels using random effects. Other factors were included as predictors, including absolute latitude, total area of source native range, destination native legume richness, human uses, and the proportion of destination species that are mutualistic (see Supplementary Table S2 for full list of model factors). As with the model above, modeling MPD as a response, terrestrial biome of the region as a random effect, as well as two hierarchical levels of geography (region and continent) to help account for spatial non-independence. We chose appropriate transformation for each predictor to decrease skew, which reduces high leverage points and improves alignment with model assumptions (e.g. heterogeneity of variance and linearity). See Supplementary Table S2 for transformations.

Results

Effect of mutualism on MPD of native regional legume species assemblages

Here, we modeled the Mean Phylogenetic Distance (either raw values, or z-values calculated from the null models) as a function of the proportion of the legume community that was mutualistic (for model 1), or whether the sub-community was mutualistic or non-mutualistic (for model 2). We define sub-communities by dividing the total legume community into two subcomponents, one composed only of the mutualistic legumes, and the other composed only of the non-mutualistic legumes. We then asked whether one of these sub-communities has higher phylogenetic diversity than the other, using both raw MPD, and after accounting for expected MPD using several different null models. We also accounted for a number of potentially confounding covariates in our models, including the latitude of the community or sub-community's region, its area, and its native legume species richness. We also accounted for the terrestrial biome of the region as a random effect, as well as two hierarchical levels of geography (region and continent) to help account for spatial non-independence. All factors and geographic regions are listed in Table S1 and S3 respectively.

For "Model 1", we found that the proportion of mutualistic legumes in a community had a significant negative effect on the mean phylogenetic distance of legume communities (Figure 1; Table 1). This result was robust to the inclusion of a number of covariates that may be confounded with the proportion of mutualistic

legumes or their phylogenetic diversity. Mutualism accounts for as much as 45% of the variation explained by the model, according to hierarchical partitioning. This result was similar for MPD z values, which were calculated to control for the results of 5 different null models. All results were consistent across all null models, and were robust to the inclusion of several covariates of known importance to community assembly (Table 1).

For “Model 2” we found that mutualistic legume sub-communities have significantly lower mean phylogenetic distance compared with non-mutualistic sub-communities from the same regions (Figure 2; Table 1). This result was robust to the inclusion of a number of covariates that may be confounded with the proportion of mutualistic legumes or their phylogenetic diversity. Mutualism accounts for as much as 50% of the variation explained by the model, according to hierarchical partitioning. This result was similar for MPD z values, which were calculated to control for the results of 5 different null models. All results were consistent across all null models. Non-mutualistic species are more common near the equator, and communities near the equator tend to have higher MPD (Figure S2; Table 1), however, even after accounting for the absolute latitude of the region, there is still a significant effect of mutualism.

Analysis of Introduced Species

We found that mutualistic legumes had a significantly lower probability of being introduced, and that the phylogenetic distance between an introduced legume species and the destination legume community had a strong negative effect on the probability of being introduced (Table 2; Figure 3). This suggests that being closely related to the destination community makes it much more likely that a legume species will become introduced in a region. Importantly, there was also a significant negative interaction between mutualism and the phylogenetic distance between an introduced species and the destination community. This means that the negative effect of phylogenetic distance on introduction success is weaker for non-mutualistic legume species, i.e. non-mutualistic legumes show less restrictions with respect to their phylogenetic history during introduced species establishment (Figure 3). The results are consistent whether using the nearest neighbor distance or median phylogenetic distance. Generally the nearest neighbor on the phylogeny for successfully introduced non-mutualistic legumes came from anywhere across the whole phylogeny, whereas mutualistic legumes nearest neighbor was more restricted (Figure 4). The mutualism effect was robust to the inclusion of a large number of covariates (Table 2).

Discussion

The intricate interplay of ecological and evolutionary processes that shape community assembly is a central theme in ecology. Though mutualisms were often overlooked in early work on community assembly, relative to other interactions, especially competition and predation, more recent work has made significant progress in integrating mutualistic and facilitative interactions into important assembly-related ecological theory, such as niche theory (Koffel *et al.* 2021). In the context of how mutualisms influence the representation of evolutionary history, mutualistic interactions can generate patterns of phylogenetic diversity that are difficult to distinguish from environmental filtering, leading potentially to an underdetermination of theory by data. Therefore, without experimental manipulations, it is challenging to disentangle and demonstrate the contribution of mutualism to patterns of evolutionary history in communities. These patterns hold import for understanding how deep-time evolution and ecology interact, and are important for conservation and restoration research, given that evolutionary diversity has been shown to contribute to ecosystem function (Cadotte *et al.* 2008, 2012, 2017; Dinnage *et al.* 2012; Srivastava *et al.* 2012; Cadotte 2013), as well as being of significant conservation interest in and of itself (Faith 1992; Crozier 1997; Purvis *et al.* 2005; Isaac *et al.* 2007). However, in this study, we have leveraged a 'natural experiment' with the presence or absence of mutualism in comparable species, providing a unique opportunity to explore this issue.

Our results demonstrate that mutualism has a strong, measurable impact on the evolutionary structure of regionally sized communities at a global scale. We found that native phylogenetic diversity increases when non-mutualistic legumes are present in the community, and non-mutualistic sub-communities have higher phylogenetic diversity. This pattern is not merely a result of non-mutualists being placed on the phylogeny in a way that leads to high PD. Communities with non-mutualistic legumes also have higher than expected PD under several different null models that control for the structure of the phylogeny and species richness (Webb *et al.*, 2002). This suggests that the amount of evolutionary history within a community is being restricted by the presence of mutualistic partners. Mutualistic partners likely act as a filter, shaping community composition and structure. There is evidence that many mutualistic interactions are phylogenetically structured, with closely

related species more likely to share similar mutualistic partner communities (Bascompte and Jordano, 2007). Previous studies suggest that legumes indeed show patterns of broad specialization with different rhizobia clades and that these are phylogenetically conserved to some degree (Sprent, 2007). In this study, the relative explanatory power of the presence of mutualism is even comparably strong to other important factors such as latitude, area, and biome, underscoring the importance of considering mutualistic interactions when studying community assembly and biodiversity patterns.

In terms of introduction patterns, we found that legumes that have been introduced to non-native regions by humans have a higher probability of becoming established if they are closely related to the existing legume community. However, for non-mutualistic legumes, this relationship is considerably weaker. This suggests again that phylogenetic conservatism of mutualistic partners is reducing the ability of distantly related species to establish in an existing community, leading ultimately to a restriction on the ability of a local community to accumulate evolutionary history. Importantly, because of the relatively recent occurrence of species introductions, this analysis is not confounded by biogeography or other long-term processes which could have contributed to the pattern found in the analysis on native legume assemblages.

In both our analysis of native legume communities and our analysis of introduced legume species, we found that the overall effect of evolutionary history, regardless of mutualistic status, was one of restriction, that is, it is consistent with theoretical effects of a phylogenetically conserved latent probability of existing or establishing across locales. This leads to a lower mean phylogenetic distance in native communities relative to that expected under a set of null models, and a lower chance of establishing an introduced species when the native community is distantly related. However, since we had comparable mutualistic and non-mutualistic species in our dataset, we could estimate how much this overall negative effect was changed by mutualism. We found the overall negative effect was less negative in non-mutualists. Arguably this estimate represents the amount of the evolutionary history restriction effect that is due to the effect of mutualism specifically (e.g. the contribution of phylogenetic conservation in mutualistic interaction to the overall effect). For the native community analysis, fully mutualistic communities of legumes had mean phylogenetic distance ~ 1.28 standard deviation less than expected (under independent swap null), which means the combined effect of mutualism and all other effects (e.g. environmental filtering) on phylogenetic restriction was -1.28 . On the other hand, based on model 1, a full non-mutualistic community would have a mean phylogenetic distance ~ 1.02 less than expected. This means

the approximate contribution of mutualism to restricting evolutionary history is $(-1.28 - -1.02) / -1.28 = \sim 0.20$, about 20%. In the introduced species analysis, the mean effect of median phylogenetic distance to the nearest relative in the native community was ~ -0.51 (every million years of distance led to a 0.51 reduction in establishment probability on the logit scale). Again this effect combines the effects of mutualism and all other effects. Without mutualism, the mean effect was ~ -0.43 according to the model. So the approximate effect of mutualism was $(-0.51 - -0.43) / -0.51 = \sim 0.16$, about 16%, a surprisingly similar proportion to native species. In general, most of the model variants we tried estimated an effect between 10% to 25% due to mutualistic interactions, which represents a fairly substantial proportion.

We suggest that the two analyses produce consistent results because the patterns they estimate are the product of similar processes. For introduced species, establishment in a native community is the product of the opportunity to disperse to a site and its ability to establish once there. If we assume that phylogenetic distance of an introduced species to a native community is unlikely to strongly influence the probability of dispersal and is primarily mediated by human movement, then the primary mechanism for our result is most likely the ability of a species to establish in the site, where this ability is mediated by the increased likelihood of the presence of effective mutualistic partners when close relatives are already present. The same mechanisms could lead to the patterns we saw in the native legume communities, since one of the processes shaping the assemblage of a native community is the immigration of new species followed by subsequent establishment. However, in the case of an observed native community we expect this process to have occurred over a much longer time-scale, and to have not involved the intervention of humans (and to have also involved biogeographic processes). That is, long-term processes of dispersal, often from nearby locales, but occasionally from more distant locales, which may also involve the evolution of new species in allopatric locations followed by dispersal into the focal locale (biogeography), combined with subsequent phylogenetically biased establishment would lead to more phylogenetic clustering in mutualistic communities relative to non-mutualistic communities, mimicking to some extent the introduction process that humans have greatly accelerated in the last century.

Our ability to separate to some extent the effects of environmental filtering and mutualistic interactions, which produce similar expected patterns on evolutionary history accumulation is significant because of the very different causal implications of the two mechanisms. Because mutualism creates a causal relationship between a resident species and a potential invader by directly affecting the presence or abundance of their shared

mutualistic partners, this means that manipulating the resident can affect the probability of subsequent invasion, whereas with environmental filtering, this is not possible, because the relationship between resident and invader is non-causal (Figure S1). The causal nature of the interaction through mutualism allows for the establishment of feedback loops that would not occur with abiotic environmental filtering. For example, the establishment of an invasive mutualistic legume will then affect the probability of subsequent invasion by other mutualistic legumes. A hypothetical example is as follows: 1) an immigrating legume (legume B) shares some portion of its compatible rhizobia community with a resident legume (legume A), and the resident legume has helped to maintain these rhizobia at high abundance; 2) the immigrating legume B is able to establish thanks in part to the abundance of these compatible rhizobia; 3) the now established invader legume B helps to increase the abundance of rhizobia that it did not share with the original resident legume A; and 4) the newly resident legume B increases the probability of establishment of a third potential immigrant legume C, which has an overlapping compatible rhizobia community with legume B (but not with legume A). This sort of process could lead to a runaway positive feedback loop if it is not interrupted by other processes (such as human intervention or competition). The causal nature of the effects also suggest that removal of the introduced species would result in a decreased probability of subsequent invasion.

A large body of theory and recent empirical work has shown that the importance of facilitation relative to other interactions such as competition increases with increasing environmental stress (Bertness & Callaway 1994; He *et al.* 2013; Wang *et al.* 2013; Silliman *et al.* 2015). This suggests that the mechanisms mediating invasive species could shift more towards those driven by mutualism, like those revealed in this study, as climate change and anthropogenic land use change drive increasing levels of stress. Indeed, a recent study showed that the effect of phylogenetic distance for a single potential invader species could shift from negative to positive, depending on whether the environment it was being introduced into was drought-stressed or not (Wang *et al.* 2023). They were also able to show that the likely mechanism of the negative effect in drought conditions, similar to what we observed, was arbuscular mycorrhizal fungi (AMF) colonization of the invaders roots which was higher when close relatives were present and whose effects only overcame stronger competition amongst close relatives when drought was severe. With increasing environmental stress globally, mutualisms are likely to take on an increased importance in ecological communities, making it all the more imperative to understand their full role in community assembly, evolutionary change and ecosystem function.

Acknowledgements

We would like to thank Marcel Cardillo and Marc Cadotte for their manuscript comments. We also want to thank the Australian Research Council for providing funding that made this work possible.

References

- Andrews, M. & Andrews, M.E. (2017). Specificity in Legume-Rhizobia Symbioses. *Int. J. Mol. Sci.*, 18.
- Bascompte, J. & Jordano, P. (2007). Plant-Animal Mutualistic Networks: The Architecture of Biodiversity.
- Bates, D., Mächler, M., Bolker, B. & Walker, S. (2014). Fitting Linear Mixed-Effects Models using lme4. *arXiv [stat.CO]*.
- Bertness, M.D. & Callaway, R. (1994). Positive interactions in communities. *Trends Ecol. Evol.*, 9, 191–193.
- Bronstein, J.L. (1994). Our Current Understanding of Mutualism. *Q. Rev. Biol.*, 69, 31–51.
- Bruneau, A., Borges, L.M., Allkin, R., Egan, A.N., de la Estrella, M., Javadi, F., *et al.* (2019). Towards a new online species-information system for legumes. *Aust. Syst. Bot.*, 32, 495–518.
- Cadotte, M., Dinnage, R. & Tilman, G.D. (2012). Phylogenetic diversity promotes ecosystem stability. *Ecology*, 93, 223–233.
- Cadotte, M.W. (2013). Experimental evidence that evolutionarily diverse assemblages result in higher productivity. *Proc. Natl. Acad. Sci. U. S. A.*, 110, 8996–9000.
- Cadotte, M.W., Cardinale, B.J. & Oakley, T.H. (2008). Evolutionary history and the effect of biodiversity on plant productivity. *Proc. Natl. Acad. Sci. U. S. A.*, 105, 17012–17017.
- Cadotte, M.W., Livingstone, S.W., Yasui, S.-L.E., Dinnage, R., Li, J.-T., Marushia, R., *et al.* (2017). Explaining ecosystem multifunction with evolutionary models. *Ecology*, 98, 3175–3187.
- Cadotte, M.W. & Tucker, C.M. (2017). Should Environmental Filtering be Abandoned? *Trends Ecol. Evol.*, 32, 429–437.
- Cavender-Bares, J., Kozak, K.H., Fine, P.V.A. & Kembel, S.W. (2009). The merging of community ecology and phylogenetic biology. *Ecol. Lett.*, 12, 693–715.
- Crozier, R.H. (1997). Preserving the information content of species: genetic diversity, phylogeny, and conservation worth. *Annu. Rev. Ecol. Syst.*
- Davies, T.J. (2021). Ecophylogenetics redux. *Ecol. Lett.*, 24, 1073–1088.
- Dinnage, R. (2013). Phylogenetic diversity of plants alters the effect of species richness on invertebrate herbivory. *PeerJ*, 1, e93.
- Dinnage, R., Cadotte, M.W., Haddad, N.M., Crutsinger, G.M. & Tilman, D. (2012). Diversity of plant evolutionary lineages promotes arthropod diversity. *Ecol. Lett.*, 15, 1308–1317.
- Faith, D.P. (1992). Conservation evaluation and phylogenetic diversity. *Biol. Conserv.*
- Flynn, D.F.B., Mirotnick, N., Jain, M., Palmer, M.I. & Naeem, S. (2011). Functional and phylogenetic diversity as predictors of biodiversity--ecosystem-function relationships. *Ecology*, 92, 1573–1581.
- Gómez, J.M., Verdú, M. & Perfectti, F. (2010). Ecological interactions are evolutionarily conserved across the entire tree of life. *Nature*, 465.
- Harrison, T.L., Simonsen, A.K., Stinchcombe, J.R. & Frederickson, M.E. (2018). More partners, more ranges: generalist legumes spread more easily around the globe. *Biol. Lett.*, 14.
- He, Q., Bertness, M.D. & Altieri, A.H. (2013). Global shifts towards positive species interactions with increasing environmental stress. *Ecol. Lett.*, 16, 695–706.
- Ibanez, S., Arène, F. & Lavergne, S. (2016). How phylogeny shapes the taxonomic and functional structure of plant-insect networks. *Oecologia*, 180, 989–1000.
- Isaac, N.J.B., Turvey, S.T., Collen, B., Waterman, C. & Baillie, J.E.M. (2007). Mammals on the EDGE: conservation priorities based on threat and phylogeny. *PLoS One*, 2, e296.
- Jones, E.I., Nuismer, S.L. & Gomulkiewicz, R. (2013). Revisiting Darwin's conundrum reveals a twist on the relationship between phylogenetic distance and invasibility. *Proc. Natl. Acad. Sci. U. S. A.*, 110, 20627–20632.
- Kembel, S.W., Cowan, P.D., Helmus, M.R., Cornwell, W.K., Morlon, H., Ackerly, D.D., *et al.* (2010). Picante: R tools for integrating phylogenies and ecology. *Bioinformatics*, 26, 1463–1464.
- Koffel, T., Daufresne, T. & Klausmeier, C.A. (2021). From competition to facilitation and mutualism: a general theory of the niche. *Ecol. Monogr.*, 91.
- Kraft, N.J.B., Adler, P.B., Godoy, O., James, E.C., Fuller, S. & Levine, J.M. (2015). Community assembly, coexistence and the environmental filtering metaphor. *Funct. Ecol.*, 29, 592–599.

Kraft, N.J.B., Cornwell, W.K., Webb, C.O. & Ackerly, D.D. (2007). Trait evolution, community assembly, and the phylogenetic structure of ecological communities. *Am. Nat.*, 170, 271–283.

Lean, C.H. (2017). Biodiversity Realism: Preserving the tree of life. *Biol. Philos.*, 32, 1083–1103.

Lean, C. & Maclaurin, J. (2016). The Value of Phylogenetic Diversity. In: *Biodiversity Conservation and Phylogenetic Systematics: Preserving our evolutionary heritage in an extinction crisis* (eds. Pellens, R. & Grandcolas, P.). Springer International Publishing, Cham, pp. 19–37.

Li, H.-L., Wang, W., Mortimer, P.E., Li, R.-Q., Li, D.-Z., Hyde, K.D., *et al.* (2015). Large-scale phylogenetic analyses reveal multiple gains of actinorhizal nitrogen-fixing symbioses in angiosperms associated with climate change. *Sci. Rep.*, 5, 14023.

Maherali, H. & Klironomos, J.N. (2007). Influence of phylogeny on fungal community assembly and ecosystem functioning. *Science*, 316, 1746–1748.

Mouquet, N., Devictor, V., Meynard, C.N., Munoz, F., Bersier, L.-F., Chave, J., *et al.* (2012). Ecophylogenetics: advances and perspectives. *Biol. Rev. Camb. Philos. Soc.*, 87, 769–785.

Purvis, A., Gittleman, J.L. & Brooks, T. (2005). *Phylogeny and Conservation*. Cambridge University Press.

Rezende, E.L., Lavabre, J.E., Guimarães, P.R., Jordano, P. & Bascompte, J. (2007). Non-random coextinctions in phylogenetically structured mutualistic networks. *Nature*, 448, 925–928.

Richardson, D.M., Allsopp, N., D'Antonio, C.M., Milton, S.J. & Rejmánek, M. (2000). Plant invasions--the role of mutualisms. *Biol. Rev. Camb. Philos. Soc.*, 75, 65–93.

R. I. Vane-Wright, Christopher J. Humphries, C. J. Humphries, Paul H. Williams & Paul Williams. (1991). What to protect?—Systematics and the agony of choice. *Biol. Conserv.*

Roskov, Y.R., Bisby, F.A., Zarucchi, J.L., Schrire, B.D. & White, R.J. (2005). ILLDIS World Database of Legumes.[Online] Available at: <http://www.illdis.org>.

Silliman, B.R., Schrack, E., He, Q., Cope, R., Santoni, A., van der Heide, T., *et al.* (2015). Facilitation shifts paradigms and can amplify coastal restoration efforts. *Proc. Natl. Acad. Sci. U. S. A.*, 112, 14295–14300.

Simonsen, A.K., Dinnage, R., Barrett, L.G., Prober, S.M. & Thrall, P.H. (2017). Symbiosis limits establishment of legumes outside their native range at a global scale. *Nat. Commun.*, 8, 14790.

Sprent, J.I. & James, E.K. (2007). Legume evolution: where do nodules and mycorrhizas fit in? *Plant Physiol.*, 144, 575–581.

Srivastava, D.S., Cadotte, M.W., MacDonald, A.A.M., Marushia, R.G. & Mirotnick, N. (2012). Phylogenetic diversity and the functioning of ecosystems. *Ecol. Lett.*, 15, 637–648.

Staab, M., Liu, X., Assmann, T., Bruehlheide, H., Buscot, F., Durka, W., *et al.* (2021). Tree phylogenetic diversity structures multitrophic communities. *Funct. Ecol.*, 35, 521–534.

Stamatakis, A. (2014). RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*, 30, 1312–1313.

Thompson, J.N. (2005). *The Geographic Mosaic of Coevolution*. University of Chicago Press.

Wang, J., Li, S.-P., Ge, Y., Wang, X.-Y., Gao, S., Chen, T., *et al.* (2023). Darwin's naturalization conundrum reconciled by changes of species interactions. *Ecology*, 104, e3850.

Wang, J., Zhang, C.B., Chen, T. & Li, W.H. (2013). From selection to complementarity: the shift along the abiotic stress gradient in a controlled biodiversity experiment. *Oecologia*, 171, 227–235.

Webb, C.O., Ackerly, D.D., McPeck, M.A. & Donoghue, M.J. (2002). Phylogenies and community ecology. *Annu. Rev. Ecol. Syst.*, 33, 475–505.

Werner, G.D.A., Cornwell, W.K., Sprent, J.I., Kattge, J. & Kiers, E.T. (2014). A single evolutionary innovation drives the deep evolution of symbiotic N₂-fixation in angiosperms. *Nat. Commun.*, 5, 4087.

Figures

Figure 1. The relationship between Mean Phylogenetic Distance (MPD) in legume communities and the proportion of the legume community composed of mutualistic legumes. The relationship is non-linear which was accounted for by a squared transformation of the predictor variable (which also fixed heteroskedasticity issues). Most of the change in MPD with respect to the proportion of mutualistic legumes occurs above 0.90, so this section has been highlighted and zoomed to better see what is happening there.

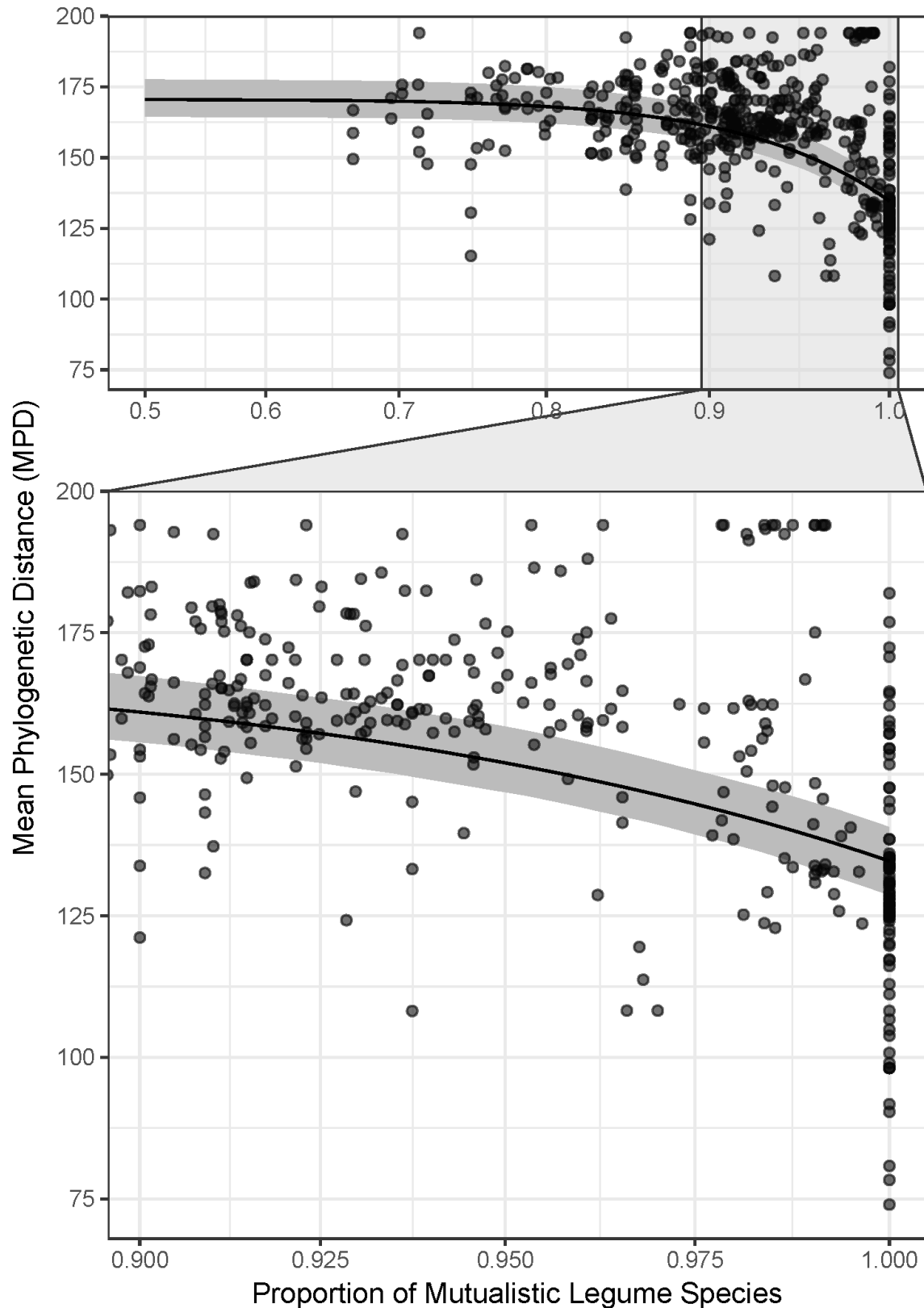


Figure 2. Plots showing the difference in Mean Phylogenetic Distance (MPD) between mutualistic and non-mutualistic sub-communities. “Z - null model name” refers to z-values calculated by subtracting the expected MPD of a null model from the observed MPD and dividing by the standard deviation of the null model replicates. Null models were run 1000 times each. Dots represent estimates from a generalized linear mixed model, after setting all covariates to zero. Error bars represent the 95% confidence interval from the model, again, setting all other covariates to zero, calculated using parametric bootstrapping. Raw MPD is always positive. Positive values of Z-values represent MPD values higher than expected based on the null model; negative values mean MPD is lower than expected.

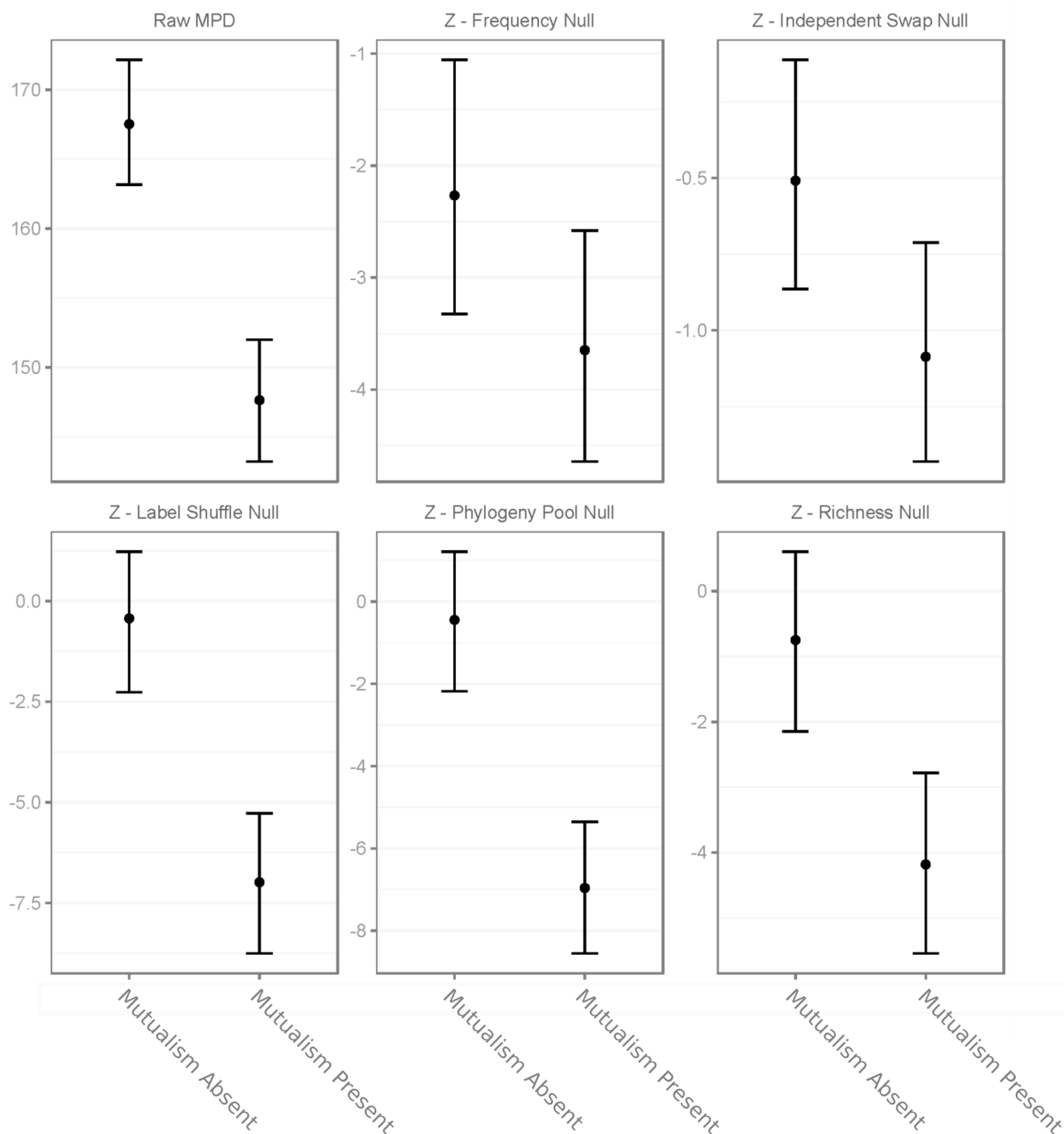


Figure 3. Relationship between phylogenetic distance of introduced species to destination community (nearest neighbour distance) and the expected number of regions it has successfully established in (e.g. probability of establishment times the number of potential destination regions), for mutualistic and non-mutualistic legume species. Line represent predictions from a generalized linear mixed model, after setting all covariates to zero. Error bands represent the 95% confidence interval from the model, again, setting all other covariates to zero, calculated using parametric bootstrapping

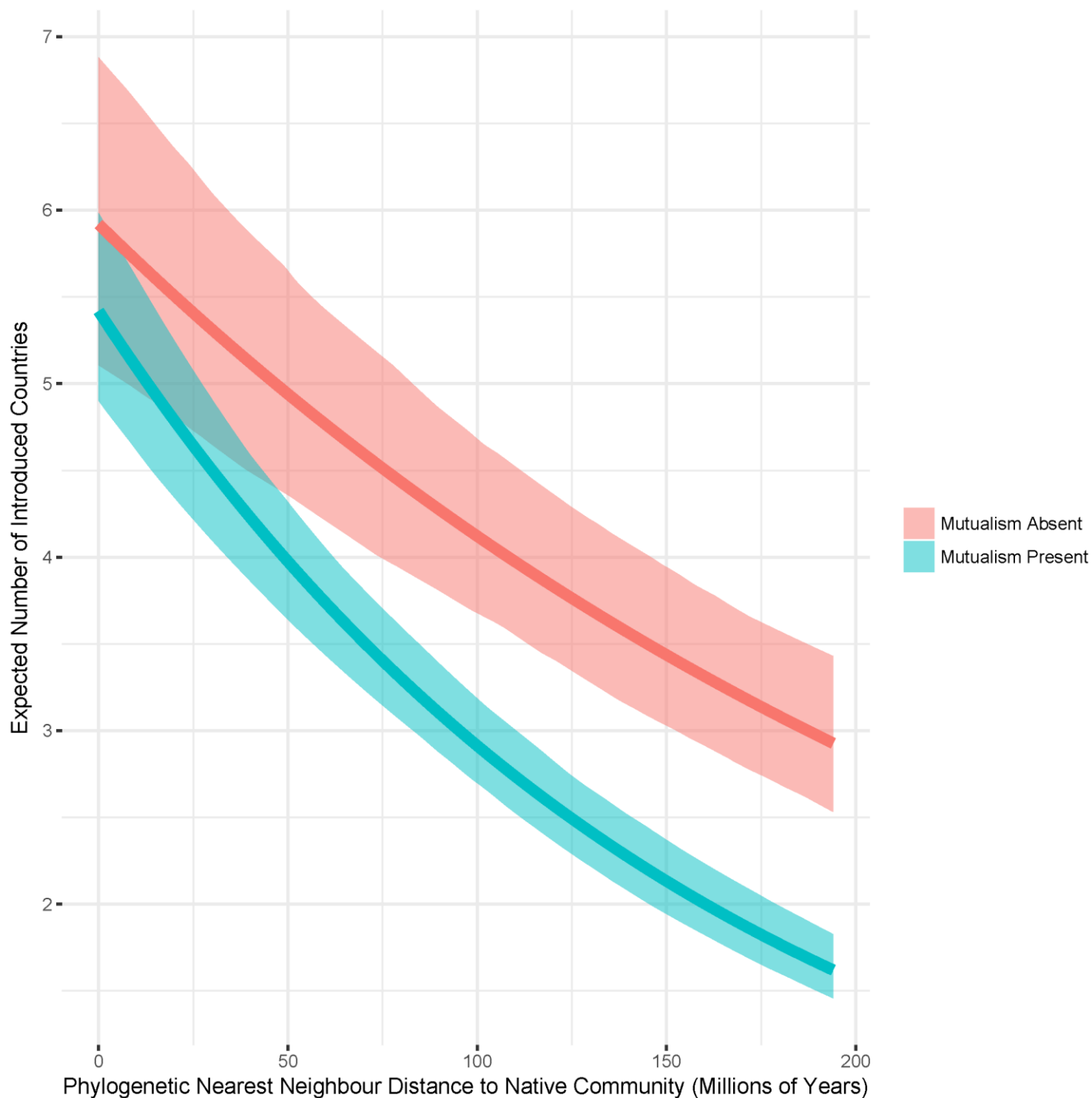
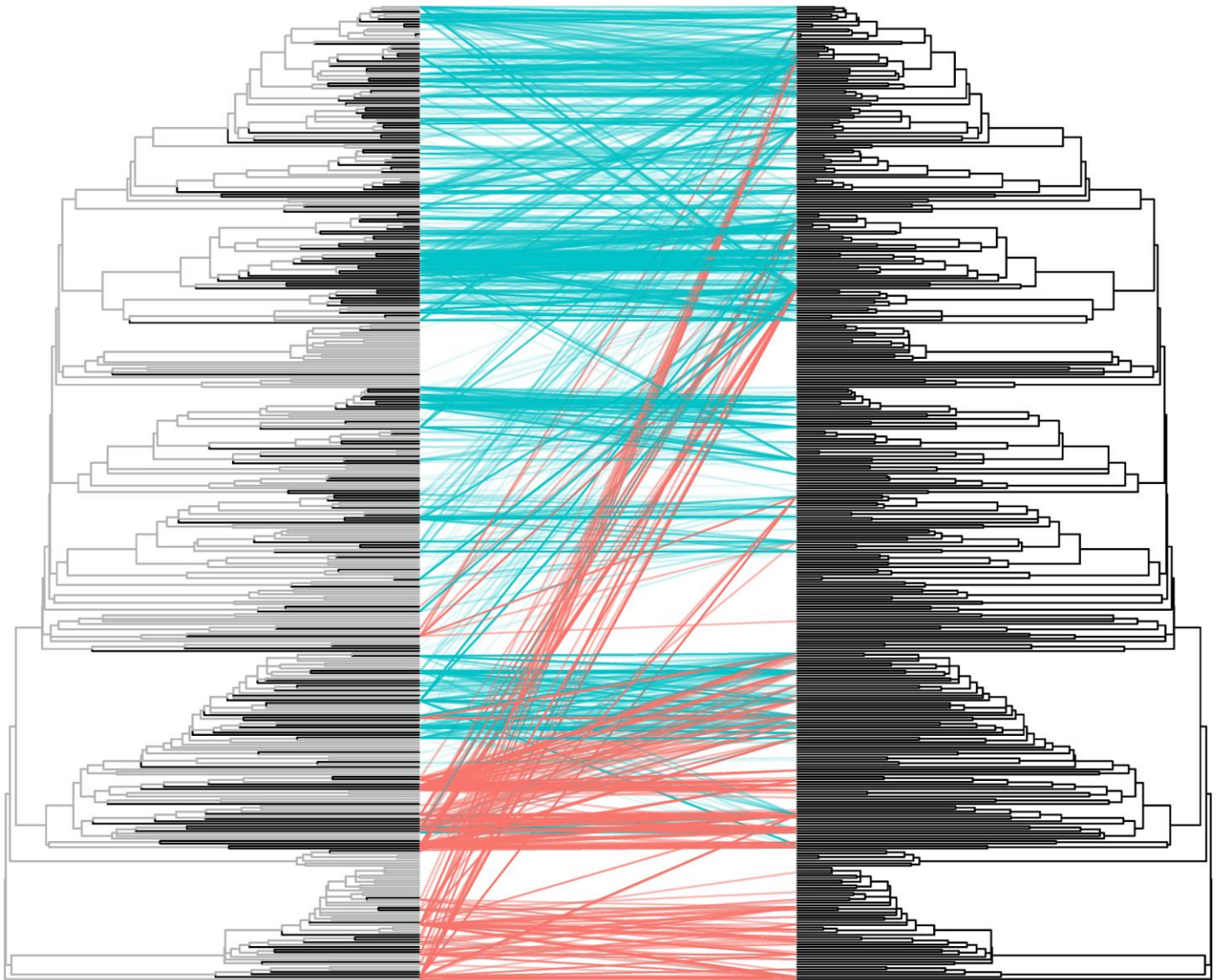


Figure 4. Network diagram showing the relationship between successfully introduced legume species and their nearest phylogenetic neighbour in the destination legume community. The left and right show the phylogeny of the legume genera in the study. On the left, species that have been successfully introduced into non-native regions are highlighted in black (non-introduced species are grey). Lines link an introduced species on the left to their nearest phylogenetic neighbours in their introduced destination regions on the right. Red lines are non-mutualistic species, whereas blue lines are mutualistic species. Where a species on the left has multiple lines radiating from it, this represents multiple introductions into different locales. It can be seen that there is a greater spread across the phylogeny in the nearest neighbour links for non-mutualistic legumes. Note what this does not show is the links for unsuccessful introductions (e.g. unobserved introductions), which is critical information for modelling the difference in introduction success rate. Nevertheless, this network diagram seems to capture the story that the statistical model has detected.



Tables

Table 1. Results of Model 1 and Model 2 on Native MPD. Estimate and Standard Error (in brackets) for fixed effect factors explaining the Mean Phylogenetic Distance of legume communities across the globe, followed by random effects factors and general information about the model. Null Model Deviations are results from models using z values calculated by subtracting expected MPD from a null model from the observed MPD and dividing by the standard deviation of null model replicates, instead of raw MPD. ***p < 0.001, **p < 0.01, *p < 0.05

	Raw MPD	Null Model Deviations			
		Richness Null	Phylogenetic Pool Null	Tip Reshuffle Null	Independent Swap Null
Model 1: Proportion of mutualistic legumes					
Intercept	152.66*** (2.80)	-5.56*** (1.05)	-5.75*** (1.06)	-5.72*** (1.05)	-1.02*** (0.21)
Mutualist	-7.21*** (0.81)	-1.02*** (0.12)	-1.04*** (0.12)	-1.04*** (0.12)	-0.26*** (0.03)
Absolute Native Latitude	-7.64*** (1.79)	-1.55*** (0.30)	-1.59*** (0.31)	-1.55*** (0.31)	-0.31*** (0.08)
Area of Region (log)	-0.76 (1.09)	0.11 (0.17)	0.12 (0.17)	0.11 (0.17)	-0.01 (0.04)
Native Legume Richness	5.24*** (1.25)	-0.93*** (0.19)	-0.99*** (0.20)	-0.99*** (0.20)	-0.10* (0.05)
Variance: Region	60.42	2.89	3.01	2.86	0.18
Variance: Biome	8.10	0.34	0.33	0.33	0.02
Variance: Continent	38.57	7.78	8.03	7.78	0.29
Variance: Residual	107.99	2.27	2.36	2.31	0.14
Model 2: Mutualistic vs. non-mutualistic sub-communities					
Intercept	167.62*** (2.43)	-0.75 (0.68)	-0.47 (0.86)	-0.46 (0.86)	-0.52** (0.20)
Mutualist	-19.99*** (1.29)	-3.43*** (0.18)	-6.52*** (0.23)	-6.54*** (0.23)	-0.58*** (0.09)
Absolute Native Latitude	-7.15*** (1.56)	-1.12*** (0.24)	-1.29*** (0.31)	-1.30*** (0.31)	-0.30** (0.10)
Area of Region (log)	-0.16 (1.05)	0.14 (0.15)	0.16 (0.19)	0.17 (0.19)	-0.02 (0.07)
Native Legume Richness	4.88*** (1.21)	-0.49** (0.17)	-1.30*** (0.22)	-1.30*** (0.22)	-0.09 (0.08)
Variance: Region	53.35	1.31	2.01	1.96	0.11
Variance: Biome	0.89	0.03	0.03	0.04	0.00
Variance: Continent	24.92	3.14	5.08	5.04	0.23
Variance: Residual	158.96	2.93	4.96	4.98	0.72

Table 2. Results of model of introduced legume species across the globe. Response was the presence or absence of an introduced species in a destination region (modelled using binomial error). Estimate and Standard Error (in brackets) of each fixed effect factor are below, followed by random effects factors and general information about the model. Four versions of the model include nearest neighbour distance and median phylogenetic distance as the response, and with or without a random slopes effect, which modelled a separate but pooled phylogenetic distance effect for each species (see methods for details). Results of most interest are bolded (other results are for covariates). ***p < 0.001, **p < 0.01, *p < 0.05.

	Nearest Neighbour Distance	Nearest Neighbour Distance (random slopes)	Median Phylogenetic Distance	Median Phylogenetic Distance (random slopes)
Intercept	-4.63*** (0.31)	-4.78*** (0.31)	-4.62*** (0.30)	-4.88*** (0.31)
Mutualist	-0.34* (0.15)	-0.34* (0.15)	-0.05 (0.25)	-0.24 (0.27)
Phylogenetic Distance to Destination Community	-0.43*** (0.02)	-0.45*** (0.04)	-0.25*** (0.04)	-0.41*** (0.05)
Mutualist by Phylogenetic Distance Interaction	-0.08*** (0.01)	-0.07*** (0.01)	-0.04** (0.01)	-0.02** (0.01)
Total Source Area (sqrt)	0.00 (0.05)	0.02 (0.05)	-0.01 (0.05)	-0.02 (0.06)
Absolute Native Latitude	0.18*** (0.05)	0.20*** (0.05)	0.15** (0.05)	0.14** (0.05)
Native Species Richness (sqrt)	0.98*** (0.10)	1.00*** (0.11)	1.15*** (0.10)	1.17*** (0.11)
Destination Area (log)	-0.07 (0.09)	-0.08 (0.09)	-0.05 (0.09)	-0.06 (0.09)
Proportion of Destination Species that are Mutualistic	0.16* (0.08)	0.15* (0.08)	0.15 (0.08)	0.14 (0.08)
Destination Region a Different Biome?	-0.80*** (0.03)	-0.80*** (0.03)	-0.88*** (0.03)	-0.88*** (0.03)
Difference in Latitude between Source and Destination Region	-0.44*** (0.02)	-0.47*** (0.02)	-0.51*** (0.02)	-0.55*** (0.02)
Distance between Source and Destination Region	0.31*** (0.02)	0.29*** (0.02)	0.27*** (0.02)	0.26*** (0.02)

Nitrogen Fixation by Proportion of Native Fixers Interaction	-0.20*** (0.04)	-0.20* (0.09)	-0.35* (0.16)	-0.43* (0.19)
Squared Difference in Latitude between Source and Destination Region	0.23*** (0.05)	0.22*** (0.05)	0.21*** (0.04)	0.20*** (0.04)
Variance: Species	1.36	1.38	1.39	1.68
Variance: Country	0.61	0.64	0.60	0.64
Variance: Region	0.92	0.95	0.93	1.00
Variance: Continent	0.50	0.51	0.46	0.46
Variance: Species by Phylogenetic Distance (random slope)		0.34		0.67
Covariance: Species Intercept and Species Slope		0.09		0.52