

PLANT–POLLINATOR COMMUNITIES

environmental gradients, trade-offs, and controllability



FERNANDO CAGUA

December 2019 – Version 1.2.0

Doctoral thesis
School of Biological Sciences
University of Canterbury

Fernando Cagua. *Plant–pollinator communities: environmental gradients, trade-offs, and controllability*; Doctoral thesis, December 2019

SUPERVISORS:

Daniel Stouffer

Jason Tylianakis

Christchurch, New Zealand

Con amor, en memoria de Betty Helena Bermudez.

1964 – 2009

CONTENTS

PUBLICATIONS

The following peer-reviewed publications have been published by the candidate during the PhD term.

Not everything is a straight line.

— **tukey_exploratory_1977**

COAUTHORSHIP DECLARATION

All chapters in this dissertation have been extracted from co-authored work.

Chapter 1 and **Chapter 2** have been published as a pre-print in *bioRxiv*. For **Chapter 1** (p. 866772; doi: [10.1101/866772](https://doi.org/10.1101/866772)) the candidate analysed all data (100%), wrote the manuscript's first draft (100%), and contributed to data collection (20%) and data cleaning (80%). For **Chapter 2** (p. 865279; doi: [10.1101/865279](https://doi.org/10.1101/865279)) the candidate analysed all data (100%) and wrote the manuscript's first draft (100%).

Chapter 3 has been published in *Journal of Ecology* (107.4, pp. 1365–2745 doi: [10.1111/1365-2745.13147](https://doi.org/10.1111/1365-2745.13147)). The candidate analysed all data (100%), wrote the the manuscript's first draft (100%), and contributed to the development of the theoretical framework (80%).

Finally, **Appendix A** has been published in *Ecology* (98.10, pp. 2640–2652; doi: [10.1002/ecy.1955](https://doi.org/10.1002/ecy.1955)). The candidate contributed to data analysis (60%), manuscript writing (40%), and data visualisation (70%).

On behalf of all co-authors, the undersigned certifys that: the above statement correctly reflects the nature and extent of the PhD candidate's contribution to this co-authored work; and in cases where the candidate was the lead author of the co-authored work he wrote the text.

Christchurch, New Zealand, December 2019



Daniel B. Stouffer

ACKNOWLEDGMENTS

Gracias a mi familia porque sin ellos no estaría aquí. En especial a mis hermanitos a los que siempre admiro y extraño. Děkuji Petrovi za to, že mi dal pevné zázemí a křídla. A Bernat por siempre estar ahí en las buenas y malas.

Thanks to Daniel for always encouraging me to give my best. To Jason for providing safe spaces where ideas can grow. To both of them for their guidance and advice, I feel lucky to have them both. I am also in debt with my co-authors, Matt, Kate, Hugo, and Audrey, as this thesis contains many of their ideas and contributions. Thanks to present and past members of the Stouffer and Tylianakis lab; it was an honour to be in the same team.

To all of those with which we smiled together over the last five years, you all make New Zealand feel home.

*We don't see things as they are
we see them as we are.*

— anonymous

PREFACE

This thesis is composed of three scientific articles. All of these articles study the processes that influence the structure of pollination networks and their implications. Each chapter is a standalone piece of research and, therefore, I only provide a brief general Introduction and Conclusion linking chapters together. In the Introduction, I focus on describing the joint context from which the research questions tackled in each chapter originate. In the Conclusion, I focus on the relationship between each chapter's results and discuss the implications of this relationship for our understanding of pollination networks.

In ?? I include another article that, although it did not end up being an integral part my thesis, it represented an important outcome provided an opportunity to learn and practice essential skills for a successful PhD and eventually led me to the topic of this dissertation.

PLANT–POLLINATOR COMMUNITIES

INTRODUCTION

When one starts to trace out the dependence of one animal upon another, one soon realises that it is necessary to study the whole community living in one habitat, since the interrelations of animals ramify so far.

— elton_animal_1927

From food and freshwater production to recreation and carbon sequestration, ecosystems provide a wide range of services of considerable value to humans ([costanza_value_1997](#)). Unfortunately, global change is currently threatening the ability of ecosystems to provide these services ([schroter_ecosystem_2005](#)). Climate change and invasive alien species, in particular, are some of the most significant causes of ecosystem degradation. A necessary step to anticipate, prevent, and reverse ecosystem degradation is to understand the factors that determine their response to disturbances.

A substantial amount of research indicates that the way ecosystems respond to disturbances is strongly determined by the network of interactions formed by the species that inhabit them ([tylianakis_global_2008](#); [bascompte_asymmetric_2006](#)). This is so, because this network, which connects all organisms in an ecological community, underpins ecosystem functioning and structure, and, therefore, can modulate the resilience of ecosystem services to disturbances ([reiss_emerging_2009](#); [dobson_habitat_2006](#)). However, we still do not understand enough about the processes that shape interaction networks in ecological communities to harness them for better ecological management. The central aim of my doctoral research [is to shed light into outstanding questions about the factors that determine the structure of ecological networks and how this structure aims to better understand these processes and explores whether a mathematically rigorous network thinking](#)



can be effectively leveraged for improved management of ecosystem services.

In this thesis, I focus on the network of mutualistic interactions between plants and pollinators. These networks, which form the base of pollination systems, play a globally significant role in the maintenance of biodiversity and crop production (**bascompte_plant-animal_2007; klein_importance_2007**). Pollination systems are locally critical too; for instance, birds or insects pollinate two-thirds of New Zealand plants (**cox_pollinator_2000**), and this includes iconic native plants (like kōwhai and pōhutukawa), and economically important crops (like kiwifruit, apples and grapes). Regrettably, just like other species interactions, the relationship between plants and pollinators, ~~are-is~~ currently being disrupted by ~~multiple drivers of human-driven~~ global change at a worldwide scale (**cox_pollinator_2000**).

The number of partners species ~~hve-have~~ is a defining feature of the roles they play in its community (**cirtwill_review_2018**). At the species level it determines whether a species is a specialist or a generalist. At the community level, the distribution of the number of partners species have in the community is the main ingredient defining network structure. However, the number of partners a species has is not constant across the different ecological communities the species may inhabit (**gravel Bringing_2018**). It has been shown that the environment can influence how species interact and therefore, ~~it-the environment~~ can also be responsible for some of the differences observed across communities (**tylianakis_ecological_2017**). However, how exactly the environment may influence the number of partners a species has, particularly in plant-pollinator communities, is not well understood. In ??, I investigate precisely this and explore how the environment may affect the specialisation of species in its community. Because multiple abiotic factors (e.g. temperature, precipitation, etc.) can have contrasting effects on species, I explore how the *stresses* the environment imposes on interacting species ~~may~~-affect network structure, irrespective of the particular variable responsible for the stress. To do that, I use a global dataset of pollination networks and complement it with information about the global climate and the occurrence of species. Importantly,

because the environment can also drive changes ~~in-community composition which species might be present there in the first place~~ ([anderson_navigating_2011](#)), I look at the effect of environmental stress on the number of partners after accounting for the number of possible partners in the community.

After exploring how abiotic factors may influence the specialisation, in ?? I investigate the possible implications for ~~the pollination service~~[pollination](#). Pollination networks are deemed to be relatively generalised when compared with other types of ecological networks. That is, pollinators tend to interact with a large number of plants and vice-versa, which ~~biases the influences the distribution of the number of partners species have in the community—the~~ degree distribution. Previous theoretical work has suggested that this tendency of species to have a large degree and incidentally share a large number of partners, is responsible for the impressive biodiversity of pollination communities ([bastolla_architecture_2009](#)). Theoretically, this partner sharing increases the possible positive feedback loops between plants and pollinators, which offset the antagonistic interactions that may exist among each guild ([moeller_facilitative_2004](#)). These findings imply that coexistence of species is maximised when pollinator sharing is the highest. This implication is, in turn, based on the assumption that pollination interactions are primarily mutualistic. However, there is ample empirical evidence going back to the end of the 19th century emphasising the competitive aspects of pollination and showing that plant reproduction depends strongly on the quality of the mutualistic service ([mitchell_new_2009](#)).

In ?? I return to the longstanding view of pollination as a balance between facilitation and competition ~~among plants~~. Specifically, I explore how the sharing of partners, a common feature of generalised pollination networks, involves trade-offs ~~in the quality of pollination that may cascade into changes to network structure~~[between the quantity and purity of pollination](#). A shortcoming of previous empirical evidence was that it focused primarily on pairs of species, or, at most, small subsets of ecological communities. However, ecological communities are highly diverse, and multiple biotic, confounding factors—beyond the number of shared pollinators—can also affect the quality of the pollination

service ([flanagan_effects_2011](#)). Here, I expand the analysis of competition for pollination to ecological communities using a comprehensive dataset collected by Hugo Marrero and collaborators in the Argentinean Pampas ([marrero_agricultural_2016](#); [marrero_effect_2014](#); [marrero_exotic_2017](#)).

In ??, I move from how biotic and abiotic factors may influence the structure of ecological networks into how the structure may be used to inform ecological management. For this purpose, I build upon recent work from theoretical physics and engineering concerned with the control of complex networks ([liu_control_2016](#)). By controlling a network, I mean being able to modify the state of an ecological community to an arbitrary stable state (where the abundance of the constituent species defines the state of a community). Traditionally, it has been assumed that species central to the network of interactions, often species with a large degree, are essential to control the ecological network ([jordan_keystone_2009](#)). These species are often termed keystone species because they play a crucial role in ecosystem functioning as they can have profound impacts on the abundances of other species in the community ([mills_keystone-species_1993](#)). However, whether central species can modify the abundance of others is not based on a mechanistic understanding of how species affect each other in a network context, but merely on the assumption that species that are better connected are more influential. I use recent developments on structural controllability of complex networks to investigate whether ~~this is the case~~ [central species are more influential](#) or not.

1

EFFECT OF ENVIRONMENT ON SPECIALISATION

*Stare at the world,
not at your model.*

— kling_practitioners_2007

This Chapter An earlier version of this chapter has been published as a pre-print by cagua_environment_2019

SUMMARY

What determines whether or not a species is a generalist or a specialist? Evidence that the environment can influence species interactions is rapidly accumulating. However, but can it also influence whether a species is a generalist or a specialist? To date, a systematic link between environment and the number of partners a species interacts with the environment and specialisation has been elusive so far. Presumably, because environmental gradients appear to have, perhaps because different environmental variables produce contrasting effects on species depending on the environmental variable different species. Here, we test for a relationship use a different approach and test whether there is a link between the stresses imposed by the environment, instead of environmental gradients directly, and species specialisation using . Using a global dataset of plant-pollinator interactions. We found that the environment can play a significant effect on specialisation, we found evidence that, even when accounting for changes in community composition, likely by interacting with species' traits and evolutionary history. Species that have a large number

of interactions are more likely environmental stress can indeed influence intermediate specialisation. Under stressful environmental conditions, species that would otherwise interact with many partners tended to focus on a smaller number of fewer, presumably higher-quality, interactions under stressful environmental conditions. Interestingly, the specialists. In contrast, specialists that were present in multiple locations are more likely communities tended to broaden their niche, presumably engaging in opportunistic interactions to cope with increased environmental stress. Indeed, many apparent specialists effectively behave as facultative generalists. Overall, many of the species we analysed are not inherently generalist or specialist. Instead, species' level of specialisation should be considered on a relative scale depending on where they are found and the environmental conditions at that location, presumably by engaging in more opportunistic interactions.

1.1 INTRODUCTION

Species interactions are known to vary widely across space and time (laliberte_cascading_2012; trojelsgaard_geographical_2015; schleuning_specialization_2012; poisot_beyond_2015). There are multiple examples of species that interact with a large number of partners in a particular community or season, but with fewer in another (olesen_temporal_2008; dupont_spatio-temporal_2009; olesen_strong_2011; rabeling_seasonal_2019; benadi_specialization_2014). Some of this variation can be attributed to environmental drivers (tylianakis_ecological_2017). However, how exactly it remains unknown how generalisable characteristics of the environment, specifically the stress it imposes on species, affects whether two species interact or not, and ultimately, and ultimately, the species' specialisation. Understanding how the environment drives the number of partners is crucial because it underpins the species' role in its community and shapes the structure of the network of interactions (cirtwill_review_2018). This structure, in turn, determines ecosystem function and stability (thebault_stability_2010).

Species interactions are determined in part by niche processes (the matching of traits) and partly by neutral processes (more abundant species are more likely to encounter each other and, thus, interact) ([vazquez_evaluating_2009](#); [vazquez_uniting_2009](#)).

The environment can influence both of these processes ([godsoe_integrating_2017](#)). It is, therefore, not surprising that, despite limitations on the spatial extent or the number of environmental gradients considered, multiple studies have ~~been able to show~~ [shown](#) how changes to interactions can be related to environmental change ([tylianakis_ecological_2017](#)). ~~For instance, some studies suggest that the strength of some trophic interactions, like predation and herbivory, can increase with temperature but might decrease with precipitation. Some other studies, however, have shown either no effect (on average) or non-linear effects of temperature or precipitation on plant-pollinator interactions.~~ Despite widespread evidence that species interactions and their networks can change along environmental gradients, the direction of change can be variable ([pires_predicted_2016](#); [baskett_latitudinal_2018](#); [devoto_patterns_2005](#)), such that until now it has been impossible to extrapolate previous findings to new kinds of gradients. Overall, while it looks clear that pairwise interactions respond to environmental drivers, there is high variability in the response ([tylianakis_global_2008](#)).

One possible explanation for the seemingly contradictory evidence is that different bioclimatic factors (like temperature or precipitation) can have contrasting effects on species and their partners, ~~such that we lack a common currency with which to derive general responses to the environment.~~ Here we attempt to ~~simplify this situation~~ [overcome these barriers to deriving general rules](#) by reducing multiple factors into a single measure of environmental stress. Previous research suggests that environmental stress may affect the number of ~~partners interaction~~ [partners a species has](#) in different ways depending on ~~its the species'~~ role in the community (for example its trophic guild) ~~or even the species itself~~ ([tylianakis_ecological_2017](#)). Specifically, we propose two alternative hypotheses of how environmental stress may affect specialisation. ~~On the one hand, it is possible that when species are~~ [First, species](#) under environmental stress, ~~they~~ might be “pressured” to focus on partners with which they are best adapted to interact. [In other words under environmental](#)



~~stress they are more specialised than otherwise.~~ For instance, hoiss_altitu~~2012~~ found increased phylogenetic clustering between plants and pollinators at higher altitudes ~~compared to lower altitudes~~; while peralta_phylogenetic_2015 found that parasitoids in plantation forest, where environmental stress was higher than in native forests, ~~were constrained to interact with hosts, they were best adapted to attack~~constrained their host use to those with which phylogenetic matching was strongest. Similarly, lavandero_genotype_2013 showed that environmental stress due to higher temperature reduced the trophic niche breadth ~~of parasitoids (range of host genotypes used)~~ of parasitoids, suggesting higher specialisation.

~~On the other hand~~Alternatively, it is also possible that when species are under environmental stress, they are forced to be more flexible in their interactions. Higher environmental stress is likely to be reflected in greater energetic or reproductive costs. Therefore, they might not be able to sustain encounter rates with their preferred partners at sufficient levels. In line with this hypothesis, hoiss_interactive_2015 found that the specialisation of plant-pollinator networks decreased both with elevation and after extreme drought events. Likewise, pellissier_spatial_2010 found a positive relationship between niche breadth and environmental stress: disk- or bowl-shaped blossoms (which allow a large number of potential pollinator species to access pollen and nectar rewards) dominated at high altitude flower communities.

Here, we investigate whether and how environmental stress can systematically affect specialisation. Our main aim is to test the two hypotheses mentioned above that relate environmental stress ~~and with a~~ species' number of partners and investigate whether this changes ~~systematically~~ across species or between trophic guilds. We propose that specialist species can become "facultative" generalists to reduce their vulnerability to the absence of preferred partners (benadi_specialization_2014). In other words, we expect that, as environmental stress increases, specialists should be more likely to engage with more partners. Species with ~~a large number of partners, on the other hand many partners, conversely~~, should have a larger pool of available partners and might, therefore, be more likely to specialise ~~under~~

~~environmental stress and focus~~ on the most beneficial partners under environmental stress. Importantly, when testing these hypotheses, we control for the potential effects of the environment ~~in-on~~ community composition (`gravel_bringing_2018`). We test these hypotheses using data on plant-pollinator interactions. These interactions provide a particularly interesting system to test these hypotheses~~because, due to the multiple trade-offs involved in the pollination service, there are multiple intuitive ways in which we could imagine species respond to environmental stress given the available partners.~~ Plant-pollinator interactions are loose enough ~~detect a signal of flexibility and partner preference but strong~~  ~~enough to drive coevolutionary processes~~ (`hutchinson_cophylogenetic_2017`). We estimate the stress species might experience in ~~their community by calculating the bioclimatic suitability of their communities given a given community based on~~ the species' patterns of global occurrence.

1.2 METHODS

We retrieved plant-pollinator networks from the Web of Life database (`fortuna_web_2014`). This database contains datasets originating from 57 studies published in the primary literature between 1923 and 2016. Calculating the environmental stress of species in their community and their potential partners required us to reduce both the taxonomic and distributional/locational uncertainty. A critical step towards reducing this uncertainty is to ensure that the names used to identify species are valid and unambiguous, which in turn allow us to obtain further information from biological databases and accurately match species across studies. Therefore, our first step was to ensure consistent spelling and standardisation of species names synonyms (see **Supplementary Methods**). The cleaning process resulted on a total of 2,555 plants and 8,406 pollinator species distributed across 73 locations around the globe (Figure ?? and **S1**).

After matching species across studies as accurate as possible, we carried on two more steps. First, we calculated the environmental

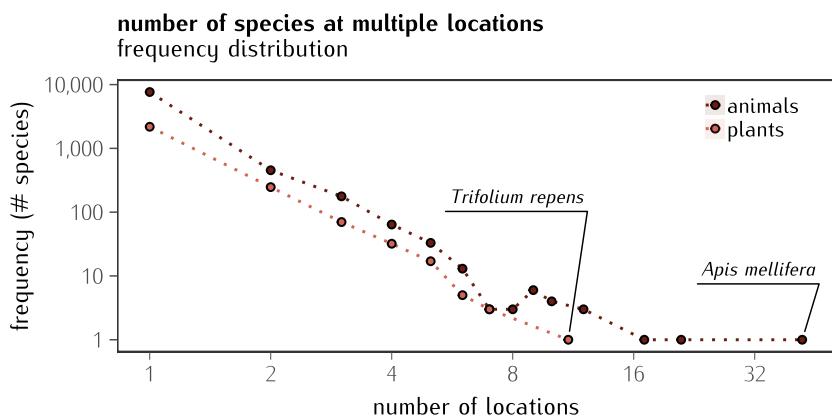


Figure 1.1: Frequency distribution of the number of locations in which a species is present. The most common pollinator species was *Apis mellifera*, which was sampled on 42 locations, while the most common plant species was *Trifolium repens*, which was sampled on 11 locations

stress of species in their communities. Second, we relate the species stress in their community with the number of partner species it has as a metric of their level of specialisation.



1.2.1 Environmental stress

We calculated the environmental stress of species in their communities. We assume that stress a species experiences in a particular location is inversely related to the suitability of the average environmental conditions in that place. As we aim to compare specialisation levels for different levels of environmental stress, we only calculate bioclimatic suitability for species that were present in at least two communities. To calculate the bioclimatic suitability of a species in a particular location, we used a niche-factor analysis ([hirzel_ecological-niche_2002](#); [broennimann_measuring_2012](#)). This approach is based on the probability density function of species distribution in an environmental variable space. Habitats are characterised by a collection of environmental variables. In a nutshell, those habitats in which the species occurs more often are deemed to be more suitable for the species than habitats in which the species has never been observed. As bioclimatic suitability is calculated in a scale from zero to one following the niche-factor analysis, for simplicity, we define environmental stress as one minus suitability.

The niche factor analysis requires two critical pieces of information. First, it requires information about the occurrences of the species of interest. Second, the method requires information about the environmental conditions for all the locations in which the species occurs. We retrieved 38.1 million occurrences from the Global Biodiversity Information Facility (GBIF; <https://www.gbif.org>). Issues with data quality are a central issue hampering the use of publicly available species occurrence GBIF data in ecology and biogeography ([jetz_essential_2019](#)). We, therefore, followed a series of filters and geographic heuristics to correct or remove erroneous and imprecise referencing records ([zizka_coordinatecleaner_2019](#)) which allowed us to identify and remove 7.5 million potentially problematic oc-

currences from further analysis. We integrated the occurrences from our plant-pollinator communities to the cleaned occurrences retrieved from GBIF.

We retrieved environmental data from WorldClim V2.0, which includes 19 bioclimatic variables commonly used in species distribution modelling ([fick_worldclim_2017](#)). We then complemented data obtained from WorldClim with data from Envirem ([title_envirem_2017](#)), which includes 16 extra bioclimatic and two topographic variables. The additional set of variables from Envirem are relevant to ecological or physiological processes and thus have the potential to improve our suitability estimation ([title_envirem_2018](#)). We obtained all environmental data as rasters composed by cells of 2.5 arc-minutes. We chose this resolution because it provides a reasonable match to the locational accuracy of the species occurrences found in GBIF, particularly those that originate from preserved specimens in museum collections.

After obtaining information about species occurrence and the environment, we then merged these two datasets such that a vector with details of our 37 bioclimatic and topographic variables characterised the location of each occurrence. Sets of occurrence data tend to be spatially aggregated due to sample bias (tendency to collect close to cities, certain countries). Moreover, spatial autocorrelation arises in ecological data because geographically clumped records tend to be more similar in physical characteristics and/or species abundances than do pairs of locations that are farther apart. To account for such spatial dependency in occurrence data, we only included one occurrence record if a species had more than one within a cell of the bioclimatic raster. We did this to avoid giving more weight to areas with a high number of occurrences, a common scenario in occurrence records collected opportunistically as the ones we use here. In this step we removed 85.4% of the occurrences which resulted in a total of 4.5 million occurrences used in our niche analysis.

A common issue of terrestrial bioclimatic datasets is that the boundaries of the cells with information do not precisely match the landmass boundaries. The result of this mismatch is that not all environmental variables were available for 3,273 of the

raster cells with occurrences (0.8% of the total). As expected, the vast majority of these problematic cells were close to the shore. To address this issue, we calculated the average value of environmental variables within a 5km buffer of the centre of the cell where the variable was missing and used it to approximate the value of the variable in that cell. Using this procedure, we were able to fill environmental variables for 89.3% of the cells where they were missing. To fill the remaining 350 cells, we repeated the aforementioned procedure but instead using a 10km buffer. We removed from further analysis occurrences located within the 135 cells for which we were unable to fill environmental variables (0.03% of the total).

Next, we calculated the probability density function of the species distribution in environmental space. To determine the environmental space, we used the first two components from a principal component analysis of the 37 bioclimatic variables associated with the species occurrences. Specifically we used the `dudi.pca` function from the R package `ade4 1.7.13` ([dray_ade4_2007](#)) and center and scale all bioclimatic variables to have a mean of zero and a unit variance. We then determined the position of species occurrences in the environmental space and estimate their bivariate probability density function. We used a kernel method to estimate this density and normalised it such that it ranges between zero and one. We used the kernel density method in the niche-factor analysis ([broennimann_measuring_2012](#)) rather than the distance from the mode ([hirzel_ecological-niche_2002](#)) (as it has been proposed earlier) because it has been shown to reduce the procedure's sensitivity to sampling effort and the resolution of the environmental space. Specifically, to calculate the probability density function we used `ecospat.grid.clim.dyn` from the R package `ecospat 3.0` ([broennimann_ecospat_2018](#)) with a grid resolution of 200. We then determined the location in the environmental space of the plant-pollinator communities using the function `suprow` from `ade4`. The normalised density at that particular location ([hijmans_raster_2019](#)) corresponds the bioclimatic suitability. The result of all these steps is the environmental stress which corresponds to one minus the bioclimatic suitability for a species of a particular location.

We used a sensitivity analysis to determine the minimum number of occurrences that are necessary to have robust environmental stress estimations. For that we used the species with most occurrences available, *Archilochus colubris*, and calculated the mean absolute error of the bioclimatic suitability values obtained with one thousand subsamples from the 74,791 occurrences available from GBIF.

1.2.2 Data analysis

We then used a set of Bayesian multilevel models to evaluate the impact of environmental stress on species specialisation. Specifically, we use the normalised degree of species as our response variable; that is, the number of species it interacts with given the number of species in the opposite guild ([martin_gonzalez_centrality_2010](#)). ~~The~~
We use degree mainly because it is a simple approximation of specialisation from the "eltonian niche" perspective. Importantly quantifying specialisation using degree does not require quantitative information about the interactions. This crucial feature allowed us to include many more communities and, ultimately minimise the possible over-representation of common and cosmopolitan species in our model. In our model, the normalised degree was modelled using a logit link function and a binomial distribution in which the number of partner species a focal [species](#) interacts with is the number of successes, and the number of species in the opposite guild is the number of trials. We are aware that whether species interact or not is not a Bernoulli process as species interactions are not strictly independent from each other. However, the use of a binomial distribution allows us to account for the differences in species richness across communities indirectly. Importantly, results are qualitatively similar when we model species degree directly using a Poisson distribution and a logarithmic link function.

We evaluated four models to assess the relative importance of suitability. A first model, our baseline model, included five variables. The predictors in the baseline model were the environmental

stress, ~~its of the community for each focal species, the species'~~ number of known possible partners in the community, and both the species guild (plant or a pollinator) and its interaction with environmental stress. We included the number of known possible partners as a predictor in our models ~~as it allows us to control for the effects of the environment on community composition, effectively accounting for species co-occurrence.~~ We calculated this metric by determining: ~~We included this predictor as we would expect an increase in the number of partners with which the species is known to interact in any other community. Controlling for actual partners merely due to an increase in~~ the number of potential partners makes our model a particularly stringent test of our environmental stress hypotheses because this variable could explain a large proportion of variance. Often, the potential and the actual number of partners is the same or very close to each other, especially for rare species present only in a few communities ~~available partners. From a community ecology perspective, this is justified because our response variable can be interpreted as the "realised" portion of the niche, and the number of potential partners is a rough approximation to the "fundamental" niche.~~ From a biogeographic perspective, including the number of potential partners is an attempt to control for the effects of the environment on community composition and species co-occurrence. We determined the number of possible partners in a particular community by counting the number of species in the opposite guild with which the species interacted in any other community in our interaction dataset.

We allowed the intercept and slope of the stress-specialisation relationship to vary among species. This approach allowed us to investigate two questions. First, it allows us to inspect the extent to which environmental stress affects species ~~similarly in a similar way~~. Second, by investigating the correlation between the intercept and the slope as a model parameter, it allowed us to inspect the extent by which species with a small or large number of partner species respond to increasing levels of environmental stress. To account for unmeasured differences between communities, like sampling effort, sampling method, or diversity, we also allowed the model intercept to be different for each community in our study. To facilitate model interpretation and

convergence, we scaled all continuous variables to have a mean of zero and a unit variance.

We compared this baseline model with three alternative models in which we removed one predictor at a time. To quantify the difference between models, in terms of their expected out-of-sample performance, we use the Wanatabe-Akaike information criterion (WAIC). All models were fitted under a Bayesian framework using the R package `brms` 2.8.0 (**burkner_brms_2017; burkner_advanced_2018**) as an interface for Stan (**carpenter_stan_2017**). For each model, we used four Markov chains of 4,000 iterations each; we used half of the iterations for warmup. We used weakly informative priors for all model parameters. Specifically we used normal priors of mean zero and standard deviation ten for the population-level effects and the intercepts, a half-Cauchy prior with a location of zero and a scale of two for the standard deviations, and, when applicable, an LKJ-correlation prior with parameter $\zeta = 1$ for the correlation matrix between group-level parameters.

1.3 RESULTS

After performing our sensitivity analysis, we found that, for a species, we need roughly 18 independent occurrences for each community for which we aim to estimate the environmental stress. This is the number of occurrences necessary to maintain the mean absolute error of bioclimatic suitability below 0.1 (Figure ??). We therefore removed from further analyses 283 species for which we did not have enough occurrences to obtain robust estimates.

Our models performed relatively well. The Bayesian R^2 for our baseline model was 0.89, which indicates our models were able to capture a large proportion of the variability on the data. Overall, we found that environmental stress does not have a consistent effect across species. Indeed, when looking at the fixed effects, stress has virtually no relationship with the normalised degree—our metric of specialisation (Figure ??a). However,

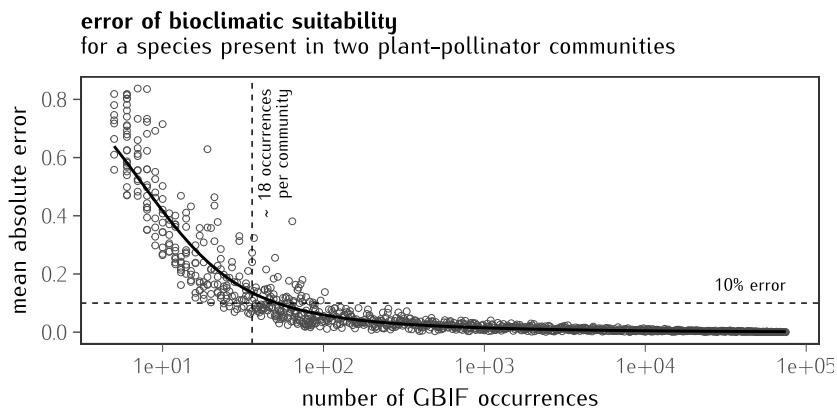


Figure 1.2: Sensitivity analysis of environmental stress error. The number of independent occurrences retrieved from GBIF is inversely related to the error of bioclimatic suitability for our plant-pollinator networks. The sensitivity analysis was performed by subsampling occurrences of *Archilochus colubris*, the species in our dataset with the largest number of occurrences in GBIF, which was recorded in two of our communities.

Table 1.1: Comparison in out of sample predictive power of the baseline model (**bold**) and their alternatives. We rank models by their expected log predictive density based on their Wanatabe-Akaike information criterion (WAIC).

predictors	WAIC	SE
stress x guild + # possible partners	6,592	170
stress + # possible partners	6,595	166
guild + # possible partners	7,081	202
stress x guild	8,041	290

~~environmental stress was still~~ Environmental stress was an important predictor in our model. The difference in WAIC between our baseline model and the model that did not include environmental stress was 489 ± 94 (Table ??). This apparent discrepancy can be explained by the variability of the specialisation-stress relationship across species.

Conditional effects of predictors in our baseline model. The shown values are based on predictions for a hypothetical community with 76 and 33 pollinators. These values correspond to the median number of species in each guild across communities. In each panel, we condition on the mean value of the other predictor in the model. We indicate mean values for each predictor with a vertical dashed line. For model fitting, we scaled all predictors to have a mean of zero and unit variance; however, here we show the unscaled predictors to facilitate interpretation. To illustrate the uncertainty around the fitted estimates, we plot the fits of 100 independent draws from the posterior distribution. The thick lines indicate the mean values of the response distribution. As there was no interaction between the guild and the number of possible interactions, we only show the conditional effect of pollinators.

For We found that for some species, there is a strong negative relationship between stress and specialisation, while for others, there is a strong positive relationship (Figure ??). Interestingly, the slope of this relationship correlates with the species' intercept in the model (Figure ??a and b). Recall that the

~~model estimates the intercept at the mean value for stress across communities (0.68).~~ The mean correlation coefficient was 0.52 [0.33, 0.67]. ~~Therefore, (95% credible intervals shown within square brackets).~~ This correlation indicates that the slope of the stress-specialisation relationship was more likely to be positive for species with ~~a large number of many~~ partners under average stress conditions (and ~~more likely to be~~ negative for species ~~with a smaller number of fewer~~ partners). ~~Extrapolating to no stress conditions:~~ species Recall that the model estimates the intercept ~~at the mean value for stress across communities (0.68).~~ Species that would interact with ~~a small number of few~~ partners under no stress are more likely to interact with more partners as stress increases, whereas those that would interact with ~~a large number of many~~ partners are more likely to interact with ~~less~~~~fewer~~.

~~The variability of the specialisation-stress relationship across species was also evident when looking at overall patterns. We found that environmental stress does not have a consistent effect across species. Indeed, when looking at the fixed effects, stress has virtually no relationship with the normalised degree—our metric of specialisation (Figure ??a).~~



As expected, we found a strong and positive relationship between the number of possible interactions and the number of realised interactions in the community. There was also a large difference of WAIC between the model that included this predictor and that that excluded it. This result indicates that the availability of potential partners—this is, community composition—accounts for a large proportion of the variability in species degree. Importantly, our findings relating to the variability of the stress-specialisation relationship were qualitatively unchanged, whether we included this variable or not.

The standard deviation (in the parameters scale) of the community intercepts was 1.02 [0.85, 1.23] which indicates the importance of the local context when determining specialisation. The standard deviation of the species intercept was 0.54 [0.48, 0.61], and that of the species' stress slope was 0.38 [0.32, 0.44] (95% credible intervals shown within square brackets).

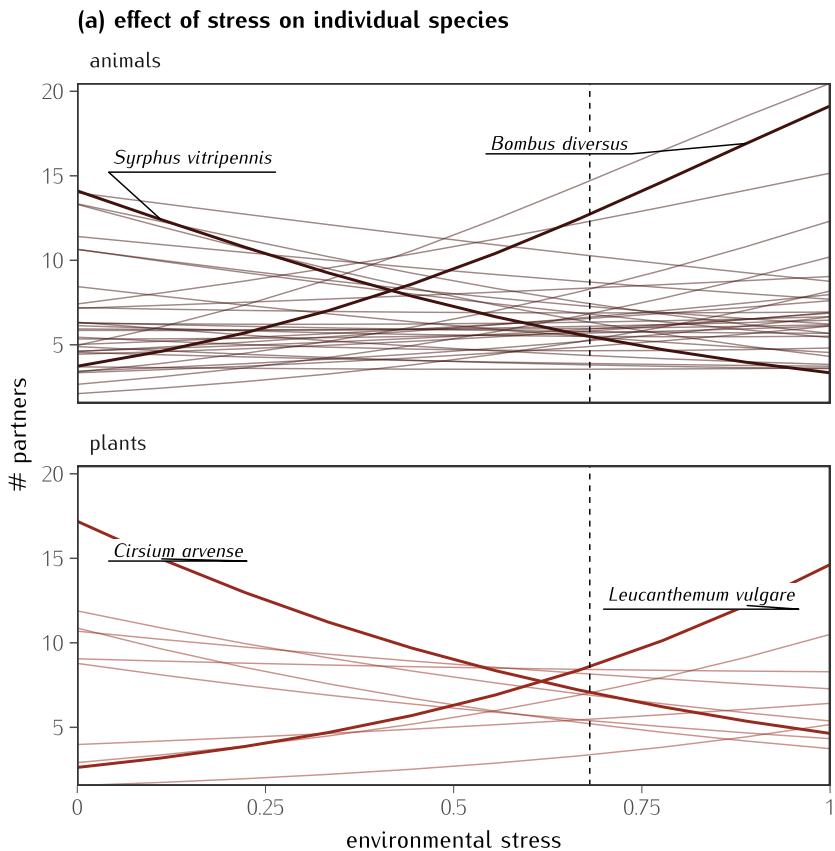


Figure 1.3: Species-level effects of environmental stress (a) Conditional effect of stress for individual species. Each line corresponds to the median relationship for each species. Although we included in the analysis of all species that are present in two or core communities, to facilitate visualisation here, we show only species for which there is suitability information in at least six communities (19 plants and 33 pollinators). As in the previous figure, fitted values assume a hypothetical community of median size. In each panel, we highlight two species for which the relationship between environmental suitability and the normalised degree was particularly strong

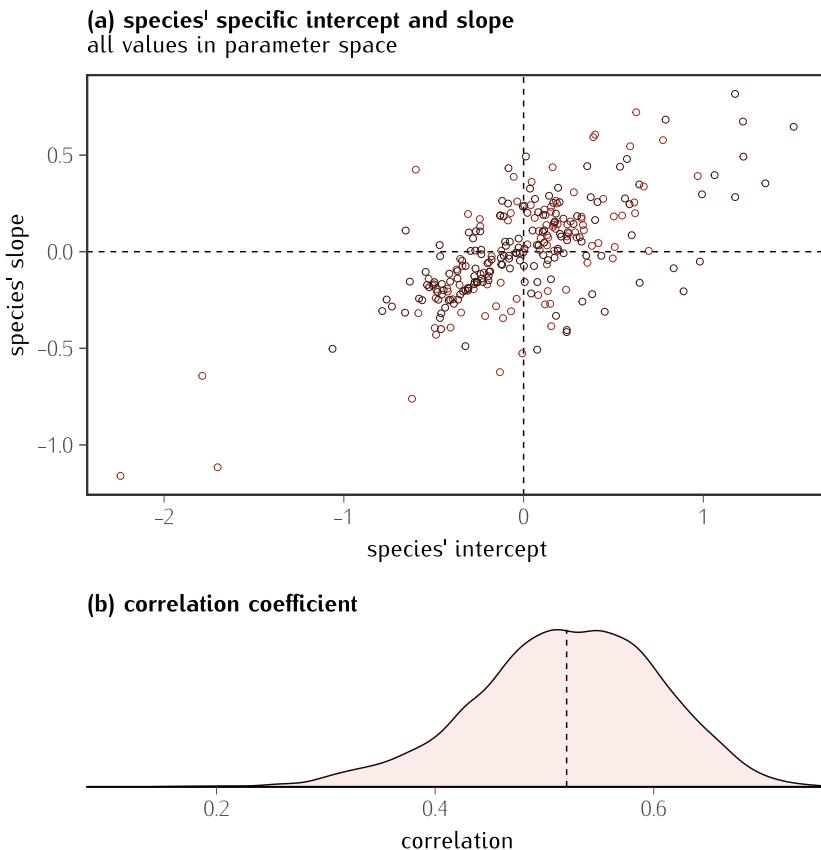


Figure 1.4: The correlation between the species' intercept and the species' slope of suitability was negative. The species' intercept can be interpreted as the relative difference between the number of partners a species has under mean levels of environmental stress and the mean number of partners across all species. Positive values of species' slope indicate a positive relationship between stress and the number of partners and vice-versa.

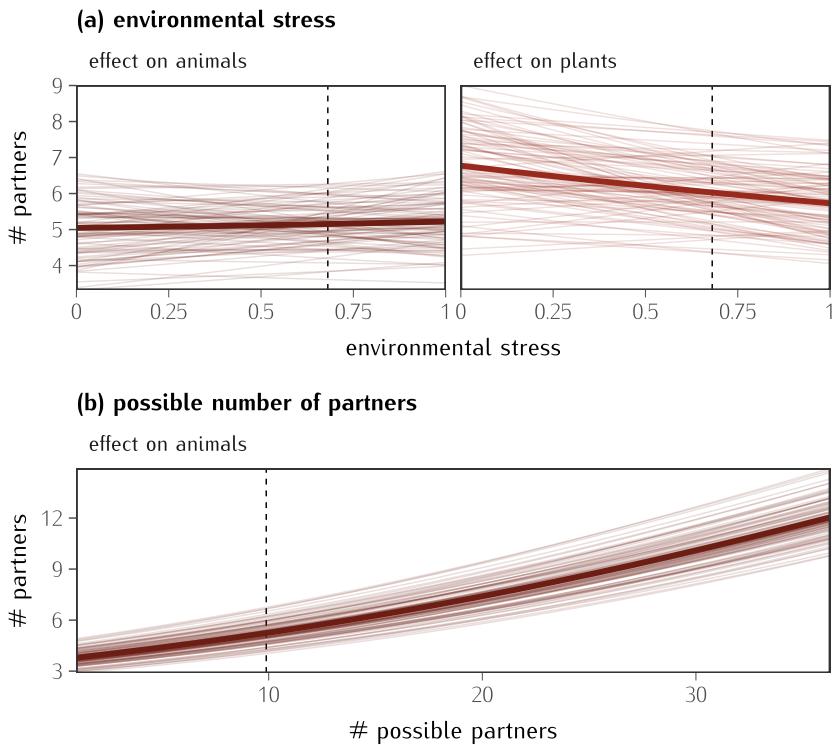


Figure 1.5: Conditional effects of predictors in our baseline model. The shown values are based on predictions for a hypothetical community with 76 and 33 pollinators. These values correspond to the median number of species in each guild across communities. In each panel, we condition on the mean value of the other predictor in the model. We indicate mean values for each predictor with a vertical dashed line. For model fitting, we scaled all predictors to have a mean of zero and unit variance; however, here we show the unscaled predictors to facilitate interpretation. To illustrate the uncertainty around the fitted estimates, we plot the fits of 100 independent draws from the posterior distribution. The thick lines indicate the mean values of the response distribution. As there was no interaction between the guild and the number of possible interactions, we only show the conditional effect of pollinators.

1.4 DISCUSSION

We set out to explore whether and how environmental stress can systematically affect specialisation. After accounting for the pool of potential partners, we found that environmental conditions contribute to determining whether a species is a generalist or a specialist *in their community*. We also found that the particular effect of the environment ~~is strongly dependent~~ depends strongly on the species' inherent tendency to be a specialist or generalist. Based on existing literature, we proposed two alternative hypotheses of how environmental stress may affect species' specialisation, and we found evidence for both of them. Species with a large number of many partners in low-stress communities were more likely to have a negative relationship and hence reduce the number of partners as stress increases. Contrastingly, species in our datasets with a small number of few partners in low-stress communities were more likely to have a larger number of interact with more partners in more stressful communities. In summary, environmental stress pushes species that are flexible enough to change their interaction partners towards intermediate levels of specialisation, a so-called “regression towards the mean”.

Our results suggest that changes in community composition are indeed the primary channel through which the environment determines changes ~~interaction probability~~ in ~~interaction probability, and this is a common mechanism through which ecological networks respond to environmental change~~ ([tylianakis_ecological_2017](#)). However, they also show that, for a large number of species, the environment may also play a substantial role in determining their level of specialisation. Previous research has recognised that environmental factors may help explain the changes in network structure along environmental gradients that cannot be explained by community composition ([tylianakis_habitat_2007](#)). However, how these two factors were linked had been elusive so far ([gravel_bringing_2018](#)). We believe that part of this difficulty could have arisen because species, and ultimately network structure, can respond in multiple, and contrasting, ways depending on the particular bioclimatic variable examined (e.g. temperature or precipitation). Using stress to summarise the effect on species

of multiple environmental gradients allowed us to detect a clear signal of the environment in species' interaction patterns.

~~Although both niche and neutral processes are relevant in determining species interactions, our model suggests that niche processes may be the predominant mechanism through which the environment systematically affects specialisation. First, it is unlikely that environmental stress correlates to local species abundances. Second, even if there is a relationship between stress and abundances, a particular environmental gradient could have a positive effect on the abundance of some species and a negative effect on others. Indeed, we find that within a community there is a wide range of stress values, even for the relatively limited number of species we were able to include in our analysis.~~

Recent research suggests that species are continuously changing their interaction partners wherever environmental conditions change in space or time ([raimundo_adaptive_2018](#)). So far it appears that this rewiring is primarily driven by generalist species ([ponisio_opportunistic_2017](#); [burkle_plant-pollinator_2013](#)), presumably because generalist species are less sensitive to trait matching of their interaction partners ([caradonna_interaction_2017](#)). Our results add two important nuances to these findings. First, because “generalists” seem to focus on [a smaller number of fewer](#) partners as environmental conditions deteriorate, we show that trait matching might still play a role in determining the [less-flexible core](#) interactions of generalist species. Second, and most importantly, our results suggest that only a small proportion of species are “true generalists” or “true specialists”[this; that](#) is, species that interact with a large or small number of partners regardless of the environmental stress, respectively. This pattern implies that rewiring is not exclusive [of species with a large number of to species with many](#) partners. Instead, at least a fraction of the species that appear to be specialist *in their communities* might be as flexible, if not more, than those with a large number of partners, effectively behaving as facultative generalists in the face of environmental change. These [flexible “flexible specialists”](#) might therefore have a more significant role in network persistence than previously expected.

In our model, we can roughly divide species between true specialists, true generalists, and flexible species. However, there is a fourth group that remained invisible to our model but has important implications for network persistence and stability. Species that can vary their interaction partners flexibly and their role in the network are more likely to persist in their community as environmental conditions vary ([gaiarsa_interaction_2019](#)). We propose this fourth group is composed of true specialists that are constrained to interact with partners of high trait-matching and therefore were not likely to be found in more than one community. [Unfortunately it is not straightforward to explicitly account for this fourth group. Accounting for these species would require information that allow us to determine the communities in which the species is not present due to the lack of interaction partners even if climatic conditions are otherwise favourable. How to distinguish these two processes \(often termed Eltonian and Grinellian, niches respectively\) is still an open area of research](#) ([avel Bringing_2018](#))

If species that are not flexible are unlikely to persist over temporal or spatial environmental gradients, we can expect specialised communities that are highly constrained by trait-matching ([vizentin-bugoni_processes_2014](#); [maruyama_morphological_2014](#)) to be far more vulnerable to increased climate change-induced environmental stress and habitat degradation than communities where role and interaction flexibility are more prevalent.

Similarly, if the patterns we see in our models have also played a role during the evolutionary history of pollination communities, our results also help [to](#) explain why only a small fraction of plant-pollinator interactions [show](#) [shows](#) a strong signature of deep co-evolutionary history ([hutchinson_cophylogenetic_2017](#)). The increases in the stress that species are predicted to experience due to rapid environmental change might further erode the co-evolutionary history of specialist species. Communities as a whole might be in a trajectory of even more diffuse co-evolution. For specialists, at least, the longer-term benefits of being able to interact with multiple partners might be more important than the shorter-term benefits of interacting with partners of high trait matching.

The structural implications of the “regression towards the mean” that environmental stress promotes are less clear. However, it is plausible to expect that nestedness, and therefore network stability, might be reduced in the face of rapid environmental change. Determining exactly how the changes in degree caused by environmental stress reflect on systematic changes in network structure would be an interesting avenue of research. Answering this question would require expanding our suitability analysis to all species in the community and compare the degree distribution of networks along a gradient of stress for the community as a whole.

In conclusion, we show that the environment can affect the specialisation level of plants and pollinators in systematic ways beyond community composition. Species that are inflexible with their interaction partners are unlikely to persist under more stressful environmental conditions. However, we show that many species are flexible in-with regards to their specialisation levels and therefore are not inherently generalists or specialists. Instead, the species’ level of specialisation/generalisation should be considered on a relative scale depending on where they are found and the environmental conditions at that environmental stress at a given location.

ACKNOWLEDGEMENTS

We thank Warwick Allen, Marilia Gaiarsa, and Guadalupe Peralta for feedback and valuable discussions. EFC acknowledges the support from the University of Canterbury Doctoral Scholarship and a New Zealand International Doctoral Research Scholarship administered by New Zealand Education. DBS and JMT acknowledge the support of Rutherford Discovery Fellowships (RDF-13-UOC-003 and RDF-UOC-1002) and the Marsden Fund Council (UOC-1705), administered by the Royal Society of New Zealand Te Apārangī.

SUPPLEMENTARY INFORMATION

Additional supporting information may be found online in the Supplementary Material section of the published article at <https://doi.org/10.1101/866772>.

2

TRADE-OFFS OF SHARING POLLINATORS

Los animales todos, en suma, no pueden pasarse sin alimento. El animal cazador no puede tampoco escapar de ser, a su vez, cazado.

Todo animal débil devora al que es más débil que él. Todo animal fuerte, no puede librarse de ser devorado por otro animal más fuerte que él. Y en esto son exactamente igual los hombres, unos respecto de otros, aunque no lleguen a los mismos extremos. Dios, en suma, ha puesto a unos seres como causa de la vida de otros, y reciprocamente a estos como causa de la muerte de aquellos.

— al-Jāhīz,
translated by [palacios_libro_1930](#)

This Chapter has been published as a pre-print by [cagua_trade-offs_2019](#)

SUMMARY

A fundamental feature of pollination systems is the indirect facilitation and competition that arises when plant species share pollinators. When plants share pollinators, [the pollination-service pollination](#) can be dramatically different [to from](#) what it would be without sharing. This depends not only on how many partners plant species share, but also by multiple intertwined factors like the plant species' abundance, visitation, or traits. These factors inherently operate at the community level. However, most of our understanding of how these factors may affect [the pollination-service pollination](#) is based on systems of up to a handful of species. By examining comprehensive empirical data in eleven natural communities, we show here that [the](#)

~~pollination service is—surprisingly—only partially community context plays a central role in determining pollen deposition dynamics. Furthermore, we show that pollination quantity and purity are only weakly influenced by the number of shared pollinators. Instead, the factors that most influence the pollination service (abundance—Abundance and visit effectiveness) also introduce a trade-off between the absolute amount of conspecific pollen received and the amount relative to heterospecific pollen appear to play a larger role, but we also show that both these “strategies” have significant trade-offs between the quantity and the purity of pollination.~~ Importantly, the ways plants appear to balance these trade-offs depend strongly on the community context, as most species showed flexibility in the strategy they used to cope with competition for pollination.

2.1 INTRODUCTION

Animal pollination plays a disproportionately important role in food production and maintenance of global biodiversity (~~bascompte_plant-animal_klein_importance_2007; ollerton_how_2011~~). At a pairwise level, the mutually beneficial relationship between plants and pollinators underpins ~~the pollination service. At a community level pollination. But the former is only an incomplete perspective of plant pollinator relationships. Ultimately, every interaction takes place at the community level. In an ecological community~~, sometimes involving hundreds of species, both plant and pollinator species are connected in a myriad of indirect connections when pollination partners are shared. These indirect connections can ~~dramatically~~ alter the quality of the pollination ~~service~~ ‘~~service~~’ that plants receive because they determine how conspecific and heterospecific pollen is transferred across the community (~~morales_interspecific_2008~~). Generally speaking, there is a trade-off between the benefits gained from a species maximising its number of partners and the costs of sharing them with other plant species (~~waser_interspecific_1978~~). However, due to the large number of factors that operate at the community level, we generally do not know how sharing pollinators affects ~~the pollination service~~ pollination beyond systems

with more than a handful of species. Here we investigate how pollinator sharing affects pollen transfer *in natural communities* and how it compares to other factors known to play a role in community dynamics like abundance, traits, and visitation patterns.

There are two main mechanisms through which sharing pollinators can affect plant fertilisation ([morales_interspecific_2008](#)). The first is by changes in intraspecific pollen transfer. Changes in intraspecific pollen transfer happen, for example, when plants [of different species](#) with more attractive flowers might reduce the number of visits to those less attractive neighbouring plants, and hence reduce the amount of *conspecific pollen* deposited by animals ([yang_pollinator_2011](#)). The second is via interspecific pollen transfer. In that case, even receiving a visit might not necessarily translate into fertilisation ([campbell_mechanism_1985](#)) because a focal plant might receive *heterospecific pollen* or because pollen from the focal plant might be lost to different species. Naturally, the precise effects on female or male plant fitness of conspecific and heterospecific pollen deposition depend on the species involved ([arceo-gomez_invasion_2016](#)) and are unknown for many plant species.

Even for species well adapted to pollinator sharing, receiving foreign pollen on stigmas or losing pollen to foreign stigmas is neutral (at best). Indeed, there is [substantial](#) evidence supporting the idea that heterospecific pollen deposition can be detrimental to seed production and plant fitness ([ashman_toward_2013](#); [arceo-gomez_invasion_2016](#)). All else being equal, provided pollen is viable and compatible ([ramsey_pollen_2000](#); [de_jong_plant_1992](#); [dafni_pollen_2000](#)), the higher the quantity of conspecific pollen and its purity (relative to heterospecific pollen), the better the pollination [service](#) received by the focal plant. As such, measuring conspecific and heterospecific pollen deposition provides a good indication of the potential levels of facilitation and competition a plant population might experience.

By definition, intra- and interspecific pollen transfer occur at the community scale. However, with few exceptions ([tur_evaluating_2016](#); [aizen_reproductive_2010](#)), most of what we know about pollen transfer and its relationship with key ecological factors are based

on studies with two plant species, [primarily from experimental settings where the rest of the community has been removed](#). That is partly so because the factors that determine the patterns of pollen deposition at the community scale are tightly intertwined, operate simultaneously, and may lead to emergent phenomena not observed at smaller scales ([flanagan_effects_2011](#)). For instance, recent empirical evidence suggests that [plants pollinators](#) with flowering traits that are “original” relative to others in the community generally have fewer interaction partners ([coux_linking_2016](#)).

This evidence aligns with the notion that a species that interacts with few species does so strongly with each of them whereas a species that interacts with a large number of species does so comparatively weakly, [a concept called asymmetry in pp networks](#) ([thebault_does_2008](#); [bascompte_asymmetric_2006](#);  [luez_species_2007](#)). If evolutionary specialisation occurs by changing traits to focus on fewer but better partners ([caruso_competition_2000](#)), we should expect a reduction of competition for pollinators in species with “original” traits and an increase of competition in species with a large number of interaction partners ([gibson_can_2012](#); [carvalheiro_potential_2014](#)). [Alternatively, it](#) [It](#) might also be the case that abundance (for example, in terms of flower or pollen counts) is the dominant force driving pollen transfer ([seifan_outcome_2014](#)). Abundant plant species might experience a dilution of available pollinators ([feinsinger_effects_1987](#); [feldman_when_2004](#)) but might also receive more effective visits by capitalising on a larger share of both visits and the pollen carried by pollinators ([stavert_plant_2019](#)). In this case, a potential reduction in the absolute amount of conspecific pollen received could be compensated by an increase in the amount of conspecific pollen relative to heterospecific pollen. Altogether, it is clear that these ecological factors can indeed shape pollen deposition at the community level. However, we still do not understand their relative importance and *the trade-offs* that might exist between them.

Here, we investigate pollen-deposition dynamics at the community scale using empirical data from eleven plant-pollinator communities in the Argentinian Pampas. First, we investigate

the relative contribution that four ecological factors make to ~~the pollination service~~pollination. Specifically, we hypothesise that there are trade-offs on how these factors affect the quantity and purity of conspecific pollen deposition. While quantity and purity should decrease for plants that share many pollination partners, other factors like the plant's functional originality, its relative floral abundance, and its visitation patterns should have the potential to compensate for this decrease Second, we examine how these four factors that might affect pollen deposition can change across communities where species are present. Because these factors may affect ~~the pollination service~~pollination in contrasting ways, and a species role is relative to other species in the community, we predict that species present in multiple communities should be flexible enough to compete for pollinators under different community contexts.

2.2 METHODS

We collected data from eleven co-flowering plant communities and their pollinators in three locations in the Argentinian Pampas. In each location, we sampled two restored and two agricultural fragments, except in one located in the Flooding Pampas, where we were only able to sample one restored fragment due to the lack of available sites.

2.3 ~~FACTORS AFFECTING QUANTITY AND PURITY OF POLLINATION SERVICE~~

2.2.1 Factors affecting quantity and purity of pollination

Our first objective was to investigate the relative contribution that different ecological factors have on pollen deposition. Generally speaking, we expect that any factor that increases the amount of conspecific pollen deposited in stigmas, both in quantity and

purity relative to heterospecific pollen, also has a positive effect on the pollination servicepollination. Specifically, we investigated the effect of (*i*) a plant's number of shared pollinator species, (*ii*) a plant's abundance relative to the rest of the community, (*iii*) the mean visit potential—a metric that combines the amount and type of pollen carried by floral visitors and the number of visits it receives from them, and (*iv*) the plant's functional originality (**laliberte_distance-based_2010**). See *Data Analysis* section below for more details on these four factors.

2.2.2 Data collection

2.2.1.1 Data collection

In each of the studied communities, we quantified pollen deposition in a subset of plant species between December 2010 and February 2011.—2011, the core months of the flowering season. This subset comprised between three and nine common insect-pollinated (entomophilous) plant species that were flowering during the sampling period. Based on data from previous years (**marrero_effect_2014**), we chose plant species such that they cover a wide range on a specialization-generalization gradient as well as a wide range of abundances. In each of the selected plants, we removed all flowers leaving only buds that were expected to go into florescence on the next day. Two days after flowering, we collected all remaining flowers and counted the number of conspecific and heterospecific pollen grains in their pistils. More details can be found in **marrero_agricultural_2016**

To obtain the number of shared pollinators for each species, we collected data to construct qualitative and quantitative pollination networks. Qualitative networks were constructed based on ten-hour observations of floral visits in each fragment. Quantitative networks were constructed using two 50 m randomly located transects in each fragment. We counted and collected all floral visitors found in a 2 m wide strip while walking at a pace of 10 m per minute (**marrero_effect_2014**; **memmott_structure_1999**). We visited the transects each month between November 2010

and March 2011. To obtain floral abundance, we counted all units of floral attraction found during an independent sampling of the same transects used to construct the quantitative visitation networks. To estimate visit potential, we need to construct pollen transfer networks in addition to the visitation networks. To do this, we examined the pollen loads present on the floral visitors collected (**marrero_exotic_2017**). When the pollen count on an individual animal was estimated to be less than 2,000 grains, we identified every grain to the species level when possible and to pollen complexes when it was not. When the pollen count was above 2,000 grains, we classified approximately 50% of pollen and total pollen counts were extrapolated (**bosch_plant-pollinator_2009**). Finally, we also recorded morphological traits that relate to plant type (herb, shrub, climber), life cycle (annual, perennial), flower colouration, phenology, and whether the species is native in the study region. More details can be found in Marrero *et al.* (2014 and 2017).

2.2.2 Data analysis

2.2.1.1 Data analysis

To investigate the impact of ecological factors on pollination services, we used two sets of linear mixed models (LMM) with bootstrap resampling. The response variables for these model sets were the number of conspecific and heterospecific pollen grains deposited per stigma in flowers open to animal-mediated pollination. We used LMMs in which pollen loads were log-transformed because these models offered a better fit than equivalent GLMMs with Poisson (or quasi-Poisson) error structure. Models were fitted using the R package `nlme` 3.1-131 (**pinheiro_nlme_2018**).

Because the amount of deposited pollen can vary widely across species, and potentially also across communities, we evaluated two possible structures for the random effects: one that includes a random intercept for plant species, and one that treats measures from species across different communities independently. We

selected the best random structure by comparing the median Akaike Information Criterion for small samples (AICc).

As fixed predictors in the models, we included the four ecological factors described above. Specifically, we calculated the number of shared pollinator species for each plant species by pooling data from the qualitative and quantitative pollination networks. To calculate the plants' relative floral abundance in their community, we aggregated floral counts for each species. We then calculated the mean visit potential of pollinator species i to plant species j as

$$o_{ij} = \frac{v_{ij}}{v_i} \frac{p_{ji}}{p_j},$$

where v_{ij} is the observed number of visits by i to j , p_{ji} is the number of pollen grains from j attached to i , v_i is the total number of visits performed by i , and p_j is the total number of grains ~~carried by~~ of j being carried in the network. We log-transformed the number of shared pollinators, floral abundance, and visit potential before including them in the model.

Finally, functional originality is defined as the distance of a species from the community trait average—the centroid of functional space of the community (**laliberte_distance-based_2010**; **coux_linking_2016**). To include phenological variation, we treated floral abundance in each of the survey months (November to March) as a “trait” in this analysis. To account for the non-independence of floral counts and weight all traits equally, we assigned a weight of 1/5 to these abundances (one for each month). We scaled all traits before calculating the centroid of the functional space and calculated the species-specific functional coordinates using the R package FD 1.0-12 (**laliberte_fd_2014**). Finally, we scaled all four factors to have a zero mean and unit variance.

To estimate the coefficients, perform model selection, and quantify the associated uncertainty, we used a combination of multi-model inference and bootstrap resampling with 99 replicates. Using bootstrap replicates allow us to better understand the uncertainties associated with our estimations. First, we performed model selection using AICc and determined the likelihood of

each candidate model (a particular combination of predictors) by calculating the median ΔAIC_c (relative to the most likely model) for each bootstrap sample. As we wanted model coefficients from more likely candidate models to carry more weight in our results, we sampled the coefficients for our factors proportionally to the likelihood of their candidate model. Finally, we used these distributions of the model coefficients to estimate their mean impact on **the pollination service (pollination)** in terms of quantity and purity of conspecific pollen deposition).

2.3 FLEXIBILITY OF PLANT STRATEGIES

2.2.1 Flexibility of plant strategies

Our second objective was to tease apart whether and how these factors that might affect pollen deposition might change across communities species are present. If community context plays a relatively small role, or species are inflexible in regards to these factors, we would expect plants of the same species to use similar “strategies” across different communities. Alternatively, if the community plays a significant role and plant species are flexible, we should be able to observe differences in the strategy a plant species uses across communities. To test this, we first used a principal component analysis (PCA) of the four ecological factors (number of shared pollinators, floral abundance, visit potential, and trait originality). We scaled factors across the whole study to ensure that the PCA space does not change according to the species present in each community. We define a species’ strategy in a community as its coordinates in PCA space. For each species that was present in two or more communities, we then calculated (*i*) the median distance between the points that correspond to the strategy a species uses in different communities and (*ii*) the area of the convex hull defined by these points in the first two principal components (only for species present in three or more communities). We then compared the magnitude of these two metrics to those obtained with 99 Monte Carlo randomizations

in which we replaced the strategy of the focal plant species by that of another randomly selected species in the dataset.

2.3 RESULTS

2.3.1 Factors affecting quantity and purity of pollination service

We first examined the potential roles played in pollen deposition by four ecological factors (number of shared pollinators, abundance, mean visit potential, and functional originality). We found that our models of pollen deposition had high explanatory power (the coefficient of determination R^2 ranged between 0.76 and 0.93) although a large portion of the explanatory power came from the random effects (Table S3). As determined by AICc, the random structure best supported by the data was the one that fit a separate intercept for each species in each community (as opposed to a common intercept for each species irrespective of the community to which they belong). This structure was best for both the models of conspecific and heterospecific pollen (Table S4).

Of the four factors we considered, we found that a plant's mean visit potential and ~~relative floral abundance~~functional originality were the most important at predicting pollen deposition in plant stigmas (Figure ??). Surprisingly, the number of shared pollinators was comparatively unimportant, particularly for models of heterospecific pollen deposition, as it was only ever included in models with relatively large AICc values (Table S5).

We found that the relationship between each of the ecological factors and pollen deposition was similar for both conspecific and heterospecific pollen. That is, strategies that were associated with an increase in conspecific pollen deposition were also associated with an increase in heterospecific pollen deposition. Specifically, the plants' mean visit potential had a positive effect on pollen deposition (Figure ??a). However, the effect size

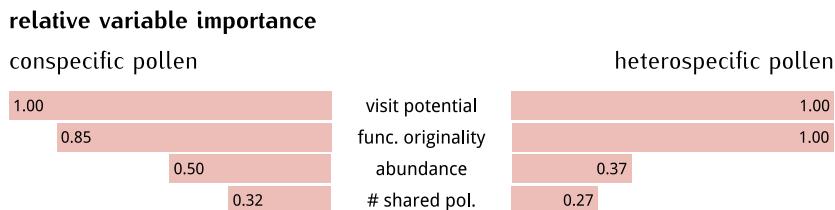


Figure 2.1: Relative importance of ecological factors on pollen deposition. The plant's visit potential and relative floral abundance are the most important factors determining the deposition of conspecific and heterospecific pollen. Meanwhile, the number of shared pollinators was generally less important. The graph shows the relative importance calculated as the sum of the Akaike weights of the candidate models that included the selected factor.

was slightly larger for heterospecific than for conspecific pollen. This larger effect indicates that, although there is a positive association between visit potential and the quantity of pollen deposition, there is a negative relationship with its purity (Figure ??b). In contrast, a plants' relative floral abundance negatively affected its deposition quantity, but the mean difference between the coefficients in the models indicates a positive association with purity (Figure ??b). The third most important factor, functional originality, had a positive, although comparatively smaller, association with both the quantity and purity. Finally, the number of shared pollinators had negative and neutral associations with conspecific and heterospecific pollen, respectively, but these impacts were small when compared to the other factors. Although the ecological factors were positively correlated (Figure S2), the collinearity between predictors did not qualitatively affect our findings (Figure S3).

2.3.2 Flexibility of plant strategies

We used a PCA of the ecological factors–species matrix to investigate whether plants' strategies towards pollen deposition is similar across communities or whether they are flexible and therefore a reflection of the community context. The first two

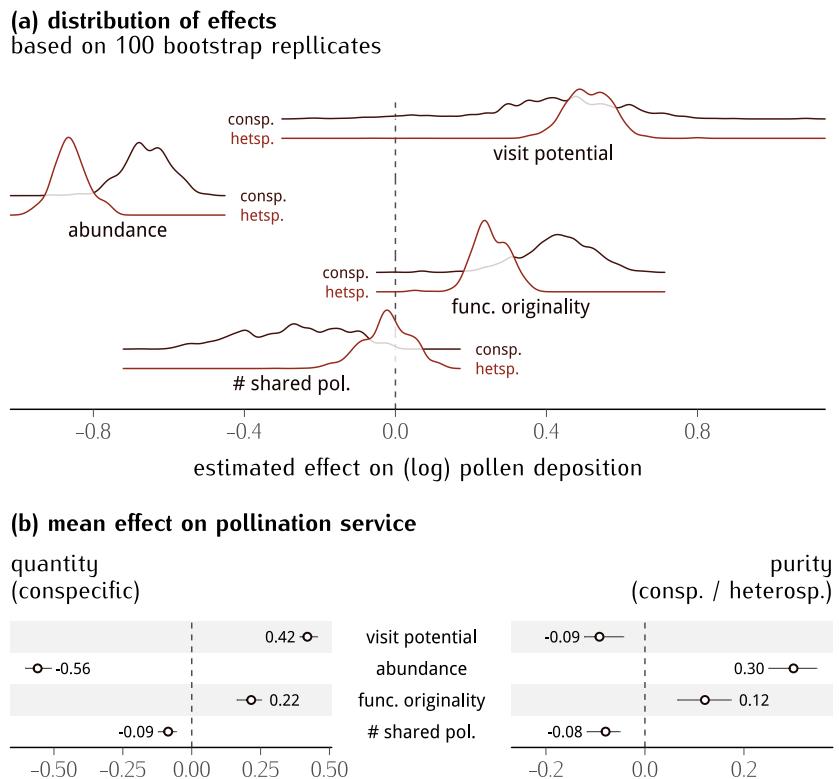


Figure 2.2: Effect of ecological factors on the pollination service. (a) The association between ecological factors and heterospecific pollen (lighter line) tended to align with their association with conspecific pollen (darker line). Visit potential and functional originality had a positive association with pollen deposition, while floral abundance and the number of shared pollinators had a negative association. The plot shows the distribution of the effects (across 99 bootstrap replicates) of the four ecological factors for conspecific and heterospecific pollen. (b) The end result of these associations is that only the plants' functional originality has a positive impact on both the quantity and purity of conspecific pollen deposition (relative to heterospecific pollen). The plot shows the model averaged mean effect (\pm SE of 99 bootstrap replicates).

PCA components explained 75% of the total variance (Figure ??a). The first component was dominated by visit potential and relative abundance while the second component was dominated by the number of shared pollinators and the plant's functional originality. When we locate the species that were sampled in more than one community in the first two PCA components (Figure ??b), we observe that the positions of any given species do not tend to be close to each other. Indeed, when we measured the median distance between the plants' coordinates, we found that it was only significantly smaller than that of randomisations for only two of the twelve species analysed (Figure ??).

2.4 DISCUSSION

Our results suggest that community context plays a central role in determining the pollen deposition dynamics and ultimately the net cost or benefit of sharing pollinators. First, we found that multiple ecological factors can modulate the quality of [the pollination service](#)[pollination](#); however, conspecific and heterospecific pollen deposition are tightly coupled and this creates a clear trade-off between the quantity and purity of pollination ([thomson_effects_2019](#)). Second, we found that the way these factors shape pollen deposition for a species could be dramatically different across communities. For instance, while a plant species in a particular community could show high levels of pollinator sharing and relatively low trait differentiation, the same species in another community can have relatively high trait differentiation and low levels of pollinator sharing. Our findings highlight that trade-offs can at least partially explain the coexistence of facilitative and competitive effects of animal-mediated pollination [in the pollination service](#).

The trade-offs involved in attaining high-quality pollination [service](#)(and more broadly between facilitation and competition) are likely to arise when plants simultaneously maximise the deposition of conspecific pollen and minimise that of heterospecific pollen. In the short term, being a specialist and sharing no pollin-

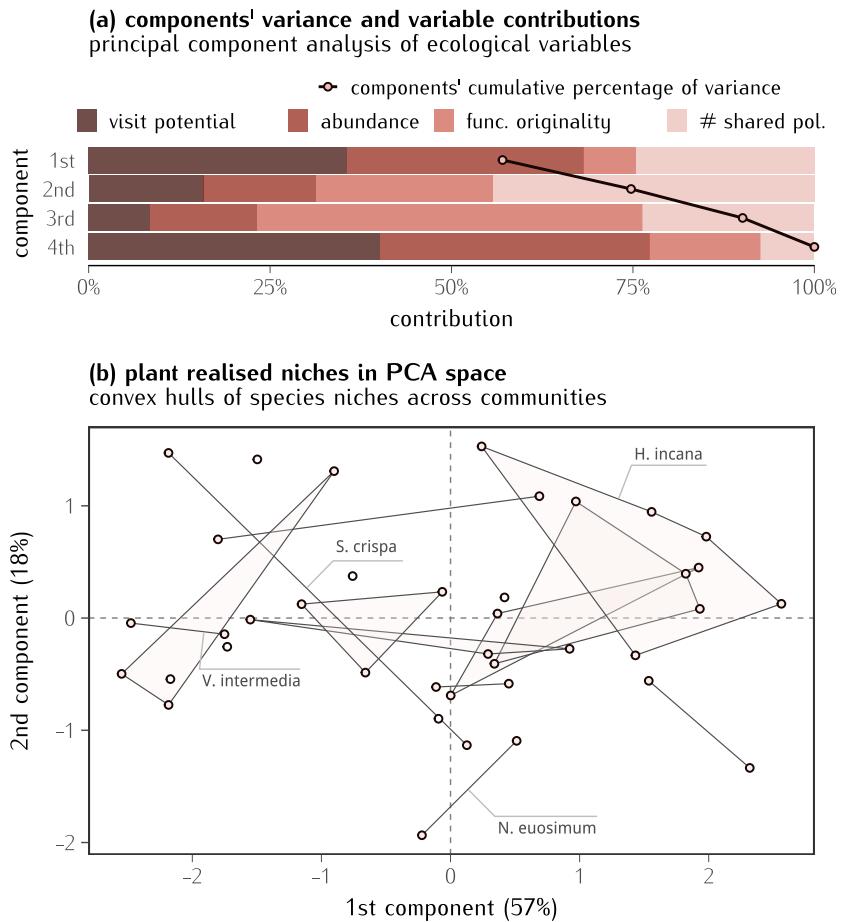


Figure 2.3: PCA of environmental factors. (a) The two first components explain a large proportion of the total variance. (b) When plants that were sampled in more than one community are plotted in terms of these two components, we observe that their points—which represent the strategy (the particular combination of ecological factors) of that species in its community—do not seem to be grouped by plant species.

flexibility of plant's strategies
median distance between plant niches vs. randomisations

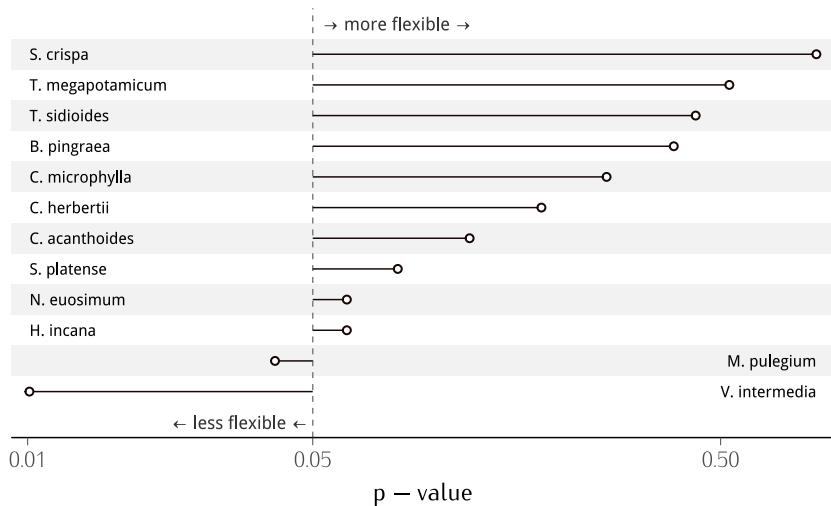


Figure 2.4: The flexibility of plant strategies. The flexibility was confirmed using Monte Carlo randomizations of the median distance between strategies of a plant species. Only two of the examined species had strategies that were less flexible than would be expected at random.

ators might reduce competition ([muchhala_competition_2010](#)) and hence be preferable. This may be due to both costs to male fitness ([muchhala_interspecific_2012](#); [morales_interspecific_2008](#)), and also, as we show here, because sharing pollinators reduces both the quantity and purity of the conspecific pollen deposited. However, over long periods of time, there could be a risk associated with a specialist plant having few pollinators ([ricketts_tropical_2004](#)). To ensure long-term survival, it is thus likely that plants also need to balance this risk with the costs of sharing pollinators ([aizen_specialization_2012](#)). One possible solution is to share pollinators *and* have original traits—as we show that trait originality is generally beneficial to pollen deposition and it is commonly thought that species that are further from others in trait space benefit from reduced competition. Yet, there are two possible caveats to this strategy that highlight the interrelatedness of the ecological factors. First, in a mutualism context, it is also possible that trait originality could come at the cost of being less ‘apparent’ to pollinators ([reverte_pollinators_2016](#)). Second, the negative relationship between originality and generalism ([carvalheiro_potential_2014](#)) has been shown to depend on plant abundance ([coux_linking_2016](#)), with generalist species being able to have original traits only when they are also abundant enough to provide a valuable reward to make visiting worthwhile to pollinators.

Visit potential (high pollen and visits) and floral abundance, which were the most important predictors of pollen deposition here, introduced an even more explicit trade-off between gaining conspecific pollen and avoiding heterospecific pollen. Receiving high visitation increases conspecific pollen deposition but increases heterospecific pollen deposition to a greater extent—even when the visitors are likely to carry a high proportion of conspecific pollen ([fang_paradoxical_2016](#)). Contrastingly, being abundant reduces the amount of conspecific pollen deposited and simultaneously reduces heterospecific pollen at a faster rate. Our results corroborate the importance that two-species studies have ascribed to visitation and abundance ([feldman_when_2004](#); [munoz_presence_2008](#); [morales_interspecific_2008](#)), but they also suggest that ([sargent_role_2006](#)) balancing the [pros-and-cons advantages and disadvantages](#) of sharing pollinators at the com-

munity level is not trivial. The fact that no species can easily out-compete others for pollination might be partially responsible for the diversity of plant-pollinator communities (**benadi_frequency_2018**).

We observed, as expected, that the effects of pollen deposition can vary widely among species. For instance, the fitness of some plant species can be hurt even by low amounts of heterospecific pollen, while the fitness of others can instead be limited by the amount of conspecific pollen (**campbell_mechanism_1985; arceo-gomez_interactive_2019**). Alternatively, plant species can also differ substantially in the extent to which self- vs. outcross-pollen differ in their value for fertilization. The difference can be particularly relevant for species that are not self-fertile or those in which self-fertilization is rarely effective due to a temporary separation in the maturation of the sexes (dichogamy).

Importantly, we show here that the balances between costs and benefits are determined not only by species identity but also by the community to which plants belong. Specifically, most plant species appear to be flexible enough to adopt markedly different “strategies” in different communities. From an evolutionary perspective, our results suggest that selection for a particular strategy might say something about the community in which a species has typically inhabited during its evolutionary history. Furthermore, from a more applied perspective, flowering plants are sometimes introduced to attract pollinators on other nearby plants. On the one hand, our results suggest that plants that increase the relative originality of natives (e.g. through distinct phenology) might have positive effects (**gibson_can_2012**). On the other, because different strategies can lead to different outcomes across communities, our results also highlight the difficulties involved in predicting whether the introduced plant species will facilitate or compete with neighbours (**bartomeus_contrasting_2008**). Other factors that we were unable to measure (e.g. pollinator behaviour and densities or the spatial context) have also been shown to play a role in the outcome of animal-mediated pollination (**ye_competition_2014; flanagan_effects_2011; cariveau_spatially_2009; thomson_effects_2019**). Nevertheless, our results indicate that the strategies a plant might use to successfully minimise competition for pollination (or max-

imise facilitation) must be determined relative to other species in the community, rather than an absolute property of the species itself.

Overall, using empirical data on pollen deposition, we show at the community level that sharing pollinators has a smaller effect on pollen deposition than what we expected based on experimental studies with a handful species. Other factors that underpin community dynamics (abundance, traits, visitation) also influence patterns of pollination quantity and purity. The interrelatedness of these factors, and the flexibility of species to position themselves across them, means that their contributions to the quality of the pollination service pollination cannot be understood in isolation. All of the factors we analysed involve substantial trade-offs in pollen deposition in the short and likely also in the long term. These trade-offs emphasise the inherently competitive nature of pollination. However, many of the widely used theoretical models of plant-pollinator communities do not account for the adverse effects of sharing pollinators ([rohr_structural_2014](#)). We therefore propose that achieving a better understanding of species coexistence and how pollination supports plant biodiversity will require seeing them as both mutualistic and competitive communities ([johnson_coexistence_2019](#)).

ACKNOWLEDGEMENTS

We thank Jamie Stavert, Bernat Bramon Mora, Laís Maia, and Michelle Marraffini for feedback and valuable discussions. We also thank Cátedra de Botánica General, Facultad de Agronomía, Universidad de Buenos Aires, the Agrasar and Bordeu families, and the University of Buenos Aires, for logistical support and permission to conduct this study at estancias Anquilóo, Las Chilcas and San Claudio, respectively. Fieldwork was supported by grants PICT 08–12504 and 0851. EFC acknowledges the support from the University of Canterbury Doctoral Scholarship and a New Zealand International Doctoral Research Scholar-

ship administered by Education New Zealand. DBS and JMT acknowledge the support of Rutherford Discovery Fellowships (RDF-13-UOC-003 and RDF-UOC-1002) and the Marsden Fund Council (UOC-1705), administered by the Royal Society of New Zealand Te Apārangi.

SUPPLEMENTARY INFORMATION

Additional supporting information may be found online in the Supplementary Material section of the published article at <https://doi.org/10.1101/865279>.

3

STRUCTURAL CONTROLLABILITY OF NETWORKS

*However, not everything that can be counted counts,
and not everything that counts can be counted.*

— cameron_informal_1963

This Chapter has been published by **cagua_keystoneness_2019**

SUMMARY

An important dimension of a species' role is its ability to alter the state and maintain the diversity of its community. Centrality metrics have often been used to identify these species, which are sometimes referred as "keystone" species. However, the relationship between centrality and keystone ness is largely phenomenological and based mostly on our intuition regarding what constitutes an important species. Here we introduce structural controllability, an approach that allows us to quantify the extent to which network topology can be harnessed to achieve a desired state. It also allows us to quantify a species' control capacity—its relative importance—and identify the set of species that are critical in this context because they have the largest possible control capacity. We found that species with high control capacity, and in particular critical species, contribute the most to the stable coexistence of their community. This result was true, even when controlling for the species' degree, abundance/interaction strength, and the relative dependence of their partners. Structural controllability is strongly related to the stability of a network and measures the difficulty of managing an

ecological community. It also identifies species that are critical to sustain biodiversity and to change or maintain the state of their community and are therefore likely to be very relevant for management and conservation.

3.1 INTRODUCTION

A major goal in ecology is to understand the roles played by different species in the biotic environment. Within community ecology, a complex-systems approach has led to the development of a variety of analytical and simulation tools with which to compare and contrast the roles of species embedded in a network of interactions ([guimera_cartography_2005](#); [bascompte_assembly_2009](#); [stouffer_evolutionary_2012](#); [coux_linking_2016](#)). A particularly relevant dimension of any species' role is its ability to alter the abundance of other species and the state of the community—since changes of this nature can have knock-on effects on ecosystem function, diversity, processes, and services ([thompson_food_2012](#); [tylianakis_global_2008](#); [tylianakis_conservation_2010](#)). This ability is sometimes referred to as a species' “keystone ness” ([mills_keystone-species](#)).

A significant proportion of the network tools used to estimate species' roles in this context rely on the calculation of a species' centrality—a relative ranking of its positional importance that originally stems from social-network research ([noah_e._friedkin_theoretical_1991](#); [martin_gonzalez_centrality_2010](#)). Generally speaking, central species tend to be better connected and consequently are more likely to participate in the network's “food chains”. Because species that participate in more chains are more likely to affect the abundances of other species, centrality metrics have often been used to identify keystone species in the community ([jordan_quantifying_2007](#)). Centrality metrics have been shown to be useful tools to rank species in regard to their potential to alter the abundances of other species, in particular when estimating the probability of secondary extinctions that may follow the loss of a species ([dunne_network_2002](#); [kaiser-bunbury_robustness_2010](#)).

Despite being conceptually intuitive, the relationship between centrality and a species' presumed impact on the state of the community is largely phenomenological. On the one hand, substantive changes in ecosystem functioning can also occur without complete removal of a species (**mouillot_functional_2013**). On the other, we are often interested in a *specific* state of the community that might be desirable to attain (or preserve) because of its biodiversity, resilience, functioning, or the ecosystem services it provides. In these cases, it might be less useful to understand which species may cause *any* change in the community. Instead, we are better served by understanding how the structure of the network can be harnessed to achieve the desired state and which species may play the largest role in this targeted process. When the state of a community is underpinned by more than a single species (often the case in real communities) and we move beyond single-species removals, we might expect the accuracy of centrality to diminish. As a result, community ecology could arguably benefit from an alternative, perhaps more mechanistically-grounded, approach to understand how species affect each other's abundance.

Species' abundances—and consequently the state of the community as a whole—are influenced both by the structure of their interactions and the dynamics of these interactions, including the mechanisms of self-regulation (**lever_sudden_2014**). However, community and population dynamics can be modelled in innumerable ways, and empirical support for one versus another is often still ambiguous (**holland_population_2002**). The alternative approach should, therefore, ideally acknowledge ecosystem dynamics, but without being overly dependent on the particular choices of how they are characterised. Among the various possibilities *structural controllability*, a branch of network control theory, appears to be a strong candidate (**isbell_human_2013**). Control theory is a widely-studied branch of engineering used to determine and supervise the behaviour of dynamical systems (**motter_networkcontrology_2015**). It is inherently designed to deal with system feedbacks and its application has recently been expanded to complex networks (**lin_structural_1974; liu_control_2016**). Consistent with long-standing ecological questions, advances in structural controllability have established a clear link between the

structure of the network and the way nodes affect each other. Unlike centrality indices, however, this link is not based on a priori assumptions between network metrics and keystone status but is instead based on well-established advances in both dynamical and complex-systems theory (**motter_networkcontrology_2015**).

At its fundamental level, structural controllability first determines whether a system is controllable or not; that is, it asks if a system could ever be driven to a desired state within a finite amount of time. Although the controllability of a network is a whole-system property, it has recently been shown that asking for the controllability of a complex-system is equivalent to finding a particular set of relevant nodes: the set with which is possible to control the state of the whole network (**liu_control_2016**). Importantly, this set of nodes is not always unique for a given network. This implies that an examination of the distinct sets provides a means to connect nodes with their *general* ability to modify the system to which they belong.

Here, we apply methods from structural controllability to a particular ecological problem and show how it can be used to generate insight into the role of species in an ecological network. Specifically, we outline the approach using a set of ten pairs of uninvaded and invaded plant-pollinator communities. We use invaded communities because there is strong empirical evidence showing that invasive species play an important role shaping the abundances of other species, something which is particularly true in these ten networks (**lopezaraiza-mikel_impact_2007**; **bartomeus_contrasting_2008**). This choice thus offers us an opportunity to explicitly contrast our theoretical observations with empirical evidence. Moreover, empirical observations indicate that steering the state of some communities—for example during ecosystem restoration or invasive species removal—can be a very difficult task (**woodford_confronting_2016**). Therefore, we first ask whether there are differences between the controllability of invaded and uninvaded networks. We then expand existing methods from control theory to effectively link the controllability (see Glossary) of a network with the role of its constituent species. We ask—from a control-theoretic perspective—whether there are key differences between species in the role they play

at driving the state of the community and explore the ecological factors related to these differences. This allows us to identify species that might be critical for network control and show that they have a larger than expected impact on the stable coexistence of the community. Finally, we compare the proposed approach to current methods based on species' centrality and show how these methods are indeed valuable but ultimately paint a limited picture in regard to the “keystone” of a species.

3.2 METHODS

We used ten paired pollination communities to apply the control-theoretic approach. Each community pair was composed of a community invaded by a plant and a community free of the invasive species. Four pairs correspond to natural or semi-natural vegetation communities in the city of Bristol, UK ([lopezaraiza-mikel_impact_2007](#)). These communities comprised 19–87 species (mean 55), and non-invaded plots were obtained by experimentally removing all the flowers of the invasive species *Impatiens glandulifera*. The other six pairs were obtained from lower diversity Mediterranean shrublands in Cap de Creus National Park, Spain ([bartomeus_contrasting_2008](#)). These communities comprised 30–57 species (mean 38); in contrast to the above, uninvaded communities were obtained from plots that had not yet been colonised by either of the invasive species *Carpobrotus affine acinaciformis* or *Opuntia stricta*. The structure of all these communities was defined by the pollinator visitation frequency, which has been shown to be an appropriate surrogate for interspecific effects in pollination networks ([vazquez_interaction_2005](#); [bascompte_asymmetric_2006](#)). Full details about the empirical networks can be found in the Supporting Information [Section S1](#).

The first step in applying methods of control theory is to construct a directed network that is able to provide an indication of the extent to which species affect each other's abundance. In some ecological networks, establishing the directionality can be relatively straightforward, for example when links represent bio-

mass transfer or energy flow (**isbell_human_2013**). In pollination networks, however, this directionality is less obvious as both species can, in theory, benefit from the interaction. We overcome that obstacle by noting that the extent to which species i affects species j relative to the extent to which j affects i can be summarised by their interaction asymmetry (**bascompte_asymmetric_2006**). This asymmetry is given by

$$a(i,j) = a(j,i) = \frac{|d_{ij} - d_{ji}|}{\max(d_{ij}, d_{ji})},$$

where the dependence of plant i on pollinator j , d_{ij} , is the proportion of the visits from pollinator j compared to all pollinator visits to plant i . Previous research has shown that mutualistic interactions are often highly asymmetric in natural communities; in other words, if a plant species is largely dependent on a pollinator species, that pollinator tends to depend rather weakly on the plant (and vice versa). We therefore create a directed link from species i to species j when $d_{ij} - d_{ji} \geq 0$ to establish the most likely direction of control between a species pair (Figure ??). Sometimes (2.4% of the observed interactions in our datasets) there is no observed asymmetry between species pairs ($d_{ij} = d_{ji}$), and we cannot infer a dominant direction of control. When this occurs, we deem both species to be equally likely to affect each other and leave a reciprocal interaction between them (a link from i to j and another from j to i). By basing the direction of the links on the asymmetry of their dependence, we are able to generate a network that is consistent with the dynamics of the community while satisfying the requirements of structural controllability. This allows us to calculate the controllability of the networks and investigate whether there are differences between invaded and uninvaded communities.

3.2.1 Controllability

A system is said to be controllable if it is possible to steer it from an initial to an arbitrary final state within finite time (**kalman_mathematical_1963**). A simple version of such a system can be described by $\frac{dx}{dt} = Ax + Bu(t)$, where the change of

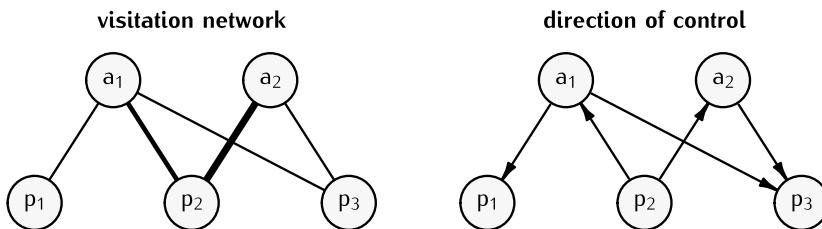


Figure 3.1: The direction of control. To establish the direction of control, we start with a weighted visitation network (on the left). In this network, the width of the links corresponds to the frequency of visitation between animals a_i and plants p_i , with wider links indicating more visits. Plant p_1 is visited exclusively by a_1 but p_1 represents only a small fraction of the floral resources exploited by a_1 . Therefore, the population of p_1 is more likely to be affected by a_1 than vice versa. We represent this with a directed link from a_1 to p_1 in the control network (on the right). The direction of control between all other species pairs can be similarly determined by inspecting the difference between their relative dependences.

its state over time ($\frac{dx}{dt}$) depends on its current state x (for example the species' abundances), an external time-varying input $u(t)$ (the control signal), and two matrices A and B , which encode information about the network structure and how species respond to external inputs, respectively. In classic control theory, determining whether this system is controllable can be achieved by checking that its controllability matrix

$$R = [B \ AB \ A^2B \ \dots \ A^{n-1}B]$$

has full rank. In complex systems, however, employing this rank condition, or numerical approximations of it is infeasible because it is hard to fully parameterise A and B (either because the weight of the links changes over time or because they are difficult to measure). Here, we use an approach based on the structural controllability theorem ([lin_structural_1974](#)), which assumes that we are confident about which elements of A and B have either non-zero or zero values (there is an interaction or not), but that we are less sure about the precise magnitude of the non-zero values. Using this structural approach, we can find out the controllability of a system for every non-zero realisation of the parameters.

We are often able to estimate A in ecological networks, as this matrix represents the interactions between species. Part of the control problem thus resides in estimating a supportable estimation of B , which represents the links between external inputs and species. Naively, any ecological community (and any system for that matter) could be controlled if we control the state of every species independently, but such an approach is typically impractical. Here, we are interested in finding a minimum driver node set (effectively finding B) with which to make the system controllable. The brute-force search for this minimum driver-node set is computationally prohibitive for most networks as it involves the evaluation of 2^N different controllability matrices where N is the number of species in the community. We therefore instead employ a recently-developed approach that shows that the control problem of finding the minimum driver-node set can be mapped into a graph-theoretic problem: maximum matching ([liu_controllability_2011](#); [liu_control_2016](#)).

maximum matching is a widely studied topic in graph theory and is commonly used in multiple applications, ranging from dating apps and wireless communications to organ transplant allocation and peer-to-peer file sharing. A matching in an unweighted directed graph is defined as a set of links that do not share common start or end nodes; the largest possible matching is called a maximum matching. For example, in a network composed of jobs and job applicants, a matching is any pairing between applicants and positions that satisfies one basic constraint: an applicant can be assigned to at most one position and vice versa. Consequently, a maximum matching is an optimal pairing, one that maximises the number of applicants with jobs and the number of positions filled. Admittedly, the link between matchings and structural controllability may appear far from straightforward.

This link becomes apparent after examining the graphical interpretation of structural controllability: from a topological perspective, a network is structurally controllable if there are no inaccessible nodes—that is, nodes without incoming links—or dilations—expansions of the network (Figure ??; Supporting Information [Section S2](#)). The key is to note that these two fundamental conditions of structural controllability imply that there is a one-to-one relationship between *superior* and *subordinate* nodes just like the one-to-one relationship between jobs and applicants (Figure ??, left). We thus use the maximum-matching algorithm to find an optimal pairing of *superior nodes* (those that can control another node) and subordinate nodes (those that can be controlled by another node) in a manner consistent with the controllability conditions (Supporting Information [Section 3.1](#)). Given the result, we can further decompose the matching into a set of paths that reveal how a control signal can flow across the links in a network to reach every node within it. As recently shown ([liu_controllability_2011](#)), the minimum driver-node set—those to which an external control input should be applied to make the system controllable—corresponds exactly to the *unmatched* nodes in the network (Figure ??).

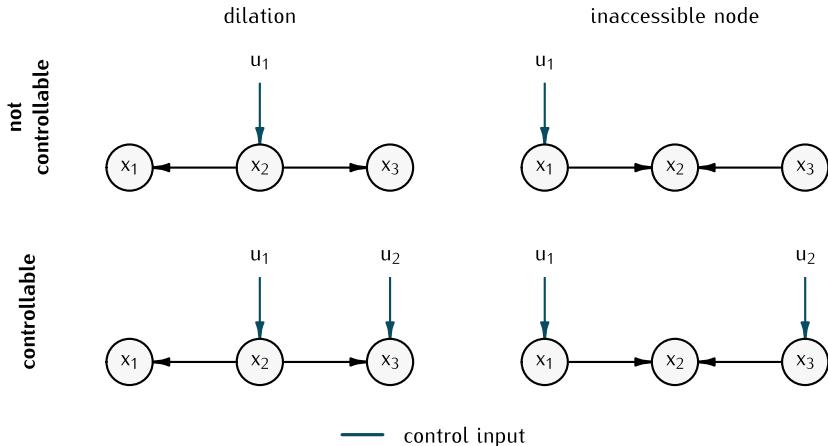


Figure 3.2: Controllability conditions. Once we have established the directions of control, we can determine whether the network is controllable or not. Any system defined by a directed network (with state nodes x_i ; species' populations in an ecological context) and external control inputs (nodes u_i , orange links) is structurally controllable if it satisfies two conditions: it has no dilations (expansions in the network) and no inaccessible nodes. The system on the top left is not controllable because there is a dilation since node x_2 is being used to control two nodes simultaneously; in other words, there are fewer superiors (x_2) than subordinates (x_1 and x_3). The network on the top right is not controllable because node x_3 is inaccessible for the only input node u_1 in the system. Both systems can be made controllable by adding an extra input node (u_2 in both bottom networks).

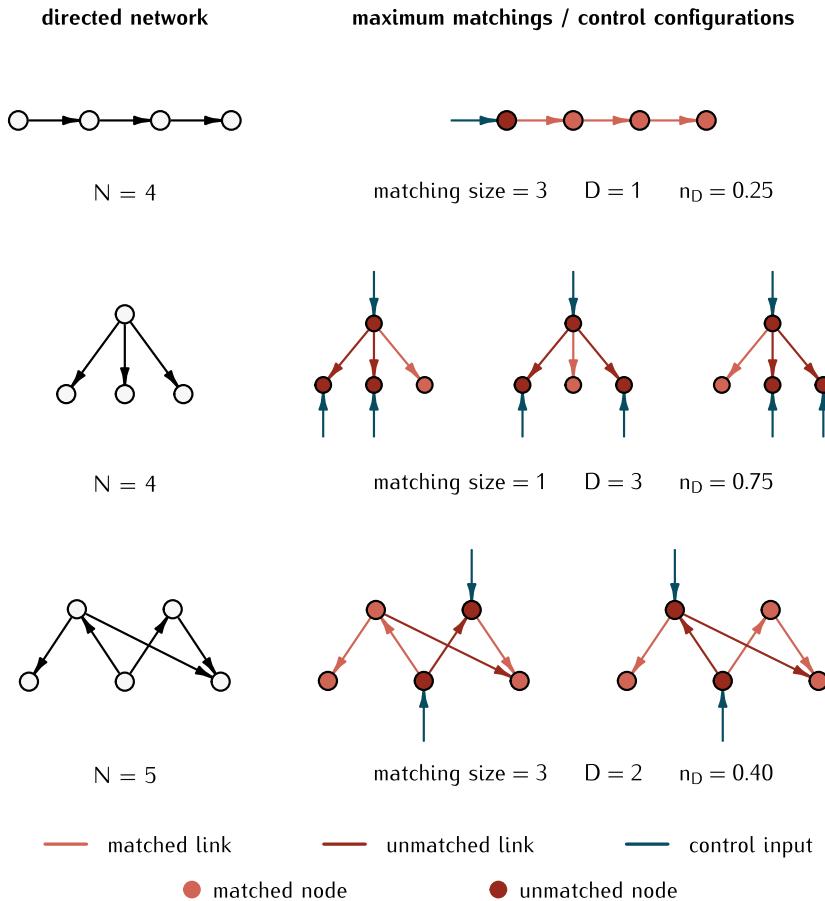


Figure 3.3: Maximum matchings and control configurations. In directed networks, a maximum matching is the largest possible set of links that do not share start or end nodes (dark purple). Maximum matchings are not necessarily unique; instead, each of them is related to a possible minimum driver-node set in the network (the nodes to which an external control input, in orange, should be applied in order to ensure controllability). The size of the minimum driver-node set D corresponds exactly to the number of unmatched nodes (the number of nodes in the network N minus the matching size). To account for network size, we use the size of the minimum driver-node set relative to the total number of nodes $n_D = D/N$ as a measure of the extent to which the network structure can be harnessed to control the system.

3.2.2 Differences between invaded and uninvaded networks

Our first objective is to investigate whether the controllability of a community is associated with invasion status or not. Finding out exactly how difficult it is to control a network depends strongly on the particularities of the desired control trajectory (i.e. the path to the desired final state) as well as the dynamical relationship between nodes. However, we are interested in understanding the controllability of a network in a more general sense, such that it can be applied even when the precise control scenario is known only incompletely. To this end, we chose an indicator that follows directly from our approach: the size of the minimum driver-node set. This simple metric provides a general indication of how difficult controlling a network might be, as systems that require a large number of external inputs to be fully controlled are intuitively more difficult or costly to manage. For instance, achieving full control in a “network” in which species do not interact at all is relatively more difficult as we would require an intervention for every single species. Conversely, the structure of a linear trophic chain can be harnessed to achieve full control using just one intervention targeted at the top species; a suitable control signal could then cascade through the trophic levels and reach other species in the community. Specifically, drawing from the structural-controllability literature, we use the size of the minimum driver-node set relative to the total number of species $n_D = \frac{D}{N}$ as a measure of the *controllability* of a network—the extent to which the network structure can be harnessed to control the community. The lower n_D the more controllable the community. In an ecological context, external inputs can be thought of as management interventions that modify the abundance of a particular species.

After finding the minimum driver-node set in each of our networks, we wanted to test whether invasion status or other predictors are correlated to controllability. We do this using a set of generalised linear models with Gaussian errors and a logit link function. The response variable was the relative size of the minimum driver-node set n_D of the twenty empirical networks (ten invaded and ten uninvaded), and we included invasion status

as a predictor. As predictors, we also include the network connectance, the network nestedness (NODF), the number of species (**menge_indirect_1995**), the network asymmetry (an indication of the balance between plant and pollinator diversity), and the interaction strength asymmetry (**bluthgen_specialization_2007**). We compared models using the Akaike Information Criterion for small sample sizes (AICc).

In addition, we also explored whether real networks differ in their architecture from random ones in a concerted way that could impact these results. Specifically, we used two null models each with 99 randomisations per network. In the first, we followed **vazquez_species_2007** and maintained the connectance of the network but randomised the visits across species such that the relative probabilities of interactions were maintained. We then re-estimated the direction of control and the corresponding size of the minimum driver-node set, n_D . For the second null model, we used the empirical directed network described above and randomly shuffled the direction of control between a species pair prior to re-estimating the size of the minimum driver-node set.

3.2.3 Species' roles

Our second objective is related to how species differ in their ability to drive the population dynamics of the community. We in turn examine whether these differences are also reflected in the role species play at supporting the stable coexistence of other species in the community. Ecologically, these differences are relevant because resources and data are limited, and therefore full control is infeasible. While calculating the size of the minimum drive-node set can measure the controllability of an ecological community, it does not provide information about the roles that particular species play.

To answer this question, we harness the fact there may be multiple maximum matchings for a given network, and each of these maximum matchings indicates a unique combination of species with which it is possible to control the network. Moreover, some

species belong to these combinations more often than do others. We call this property a species' control capacity, ϕ . The higher a species' control capacity, the greater the likelihood that it would need to be directly managed to change (or maintain) the ecological state of their community. Therefore, a species' control capacity provides an estimation of its relative importance at driving the state of the community ([jia_control_2013](#)).

To calculate a species' control capacity ϕ , we must first enumerate all possible maximum matchings (Supporting Information [Section S3.2](#)). Unfortunately, enumerating all maximum matchings is extremely expensive from a computational perspective, a network with a couple dozen species has several hundred million unique maximum matchings. To solve this problem, we employ a recently-developed algorithm that reveals the control correlations between the nodes in the graph while requiring considerably less computational resources ([zhang_input_2016](#)). Using this algorithm, we are able to identify species that are possible control inputs—those that belong to the minimum driver-node set in at least one of the possible [control configurations](#). Here, we extend this algorithm such that it is possible to calculate a highly accurate approximation of the control capacity ϕ of every species in the network (Supporting Information [Section S3.3](#)). In the networks that contained reciprocal links (because there was no asymmetry in the dependences of a species pair), we averaged a species' control capacity ϕ across every possible “non-reciprocal” version of the network (Supporting Information [Section S3.4](#)).

We then examined how species-level properties were related to control capacity using a set of candidate generalised linear models with binomial error structure. These models included five predictor variables that mirror the network-level predictors. First, the species' contribution to nestedness, which has been proposed as a key feature that promotes stability and robustness in mutualistic networks ([saavedra_strong_2011](#)). Second, the species' strength (the sum of a species' visits), which quantifies the strength of a species' associations and is indirectly related to its abundance ([poisot_comparative_2012](#)). Third, the direction of asymmetry which quantifies the net balance in dependencies; that is, it indicates if a species affects other species more than

what they affect it or not (`vazquez_species_2007`). Fourth, the species' degree in order to account for the intrinsic centrality of a species. Finally, we included a categorical variable for the species' trophic level (plant or pollinator) and an interaction term between trophic level and the previous four variables. To facilitate comparison between predictors, degree and visitation strength were log-transformed and all four continuous variables were scaled to have a mean of zero and a standard deviation of one. In these models, species from all networks were analysed together. We initially included random effects to account for possible variation across communities. Specifically, we tested structures that allowed for a random intercept for the network, site, and the study it comes from. However, we found that in all cases the among-group variance was effectively zero, and therefore we did not include any random effect in further analyses. We then generated all possible candidate models across the space of models with all, some, and none of the predictor variables. To identify the models that were best supported by the data, we first determined the most parsimonious random structure using the AICc. The relative importance of variables was then assessed by looking at their effect sizes in the top-ranked models and the cumulative weight of the models in which they are present.

In addition, we wanted to understand how a species' control capacity ϕ described above relates to metrics of keystone ness based on centrality. Specifically, in each network, we calculated the species' degree, betweenness, closeness centrality (`martin_gonzalez_centrality_2010`), page rank (`mcdonald-madden_using_2016`), and eigen centrality (`jordano_ecological_2006`). We then calculated the spearman correlation coefficient between control capacity and each of these centrality metrics.

Our analysis revealed that some species have a control capacity $\phi = 1$. These species are critical to controlling their community because they are part of the minimum driver-node set in *every* control scenario. In other words, it is theoretically impossible to drive the state of the community to a desired state without directly managing the abundance of these species. We thus anticipate that these species have a disproportionately large impact on the community dynamics. To test this hypothesis, we identified these

critical species in each of the networks and investigated whether they have a larger than average impact on the stable coexistence of species in the community. Within mutualistic networks, one useful measure of stable coexistence is called structural stability ([rohr_structural_2014](#)). Mathematically, the structural stability of a network represents the size of the parameter space (i.e., growth rates, carrying capacities, etc.) under which all species can sustain positive abundances ([saavedra_nested_2016](#)). The contribution of any given species i to stable coexistence can be estimated by calculating the structural stability of the community when the focal species i is removed. To allow comparison across communities, the structural stability values were scaled within each network to have a mean of zero and a standard deviation of one. Given these species-specific estimates of structural stability, we then used a t-test to compare the contribution to stable coexistence of critical and non-critical species. More details about the calculation of structural stability can be found in the Supporting Information [Section S4](#).

3.2.4 Testing assumptions

Just like the centrality metrics, the information obtained by applying structural controllability depends on the ability of the network to accurately represent the ecological community. We thus tested the sensitivity of our approach to two fundamental assumptions. First, we tested that visitation is an appropriate proxy to infer interspecific effects by comparing the results obtained using visitation to two alternative metrics in a separate dataset that lacked invasive species ([ballantyne_constructing_2015](#)). Specifically, we also calculated the controllability (the size of the minimum driver node-set) and the control capacity of networks constructed using pollinator efficiency (which measures the pollen deposition of an interaction) and pollinator importance ([neeman_framework_2010](#)). See Supporting Information [Section S5](#) for more details.

Second, because interspecific dependencies themselves depend on the network topology and consequently on the accurate

sampling of interactions, we tested the robustness of structural controllability to the uncertainty involved with the sampling of interactions. Here, we compared the results obtained when using the full network and when randomly removing interactions from the weakest links in the network. This effectively removed the rare interactions from the networks (more details in the Supporting Information [Section S6](#)).

3.3 RESULTS

3.3.1 Controllability

The size of the minimum driver-node set relative to the number of species in each network n_D ranged between $n_D = 0.58$ and $n_D = 0.88$ (median 0.74).

3.3.2 Differences between invaded and uninvaded networks

We found that the relative size of the minimum driver-node set of invaded communities was not significantly different from that of communities that have not been invaded (Figure ??a). In contrast, there was a large negative relationship between n_D and the network asymmetry (Figure ??b). Furthermore, there were also negative, albeit weaker, relationships between n_D and connectance, nestedness and species richness (Table [S3](#)). The relative size of the minimum driver-node set n_D of empirical networks did not differ from that of a null model that roughly preserved the degree distribution and fully preserved the network connectance ($p = 0.66$; Figure ??c). However, empirical networks had a larger n_D than null models that preserved the interactions but shuffled the direction of control of the empirical network ($p = 2.4 \times 10^{-7}$).

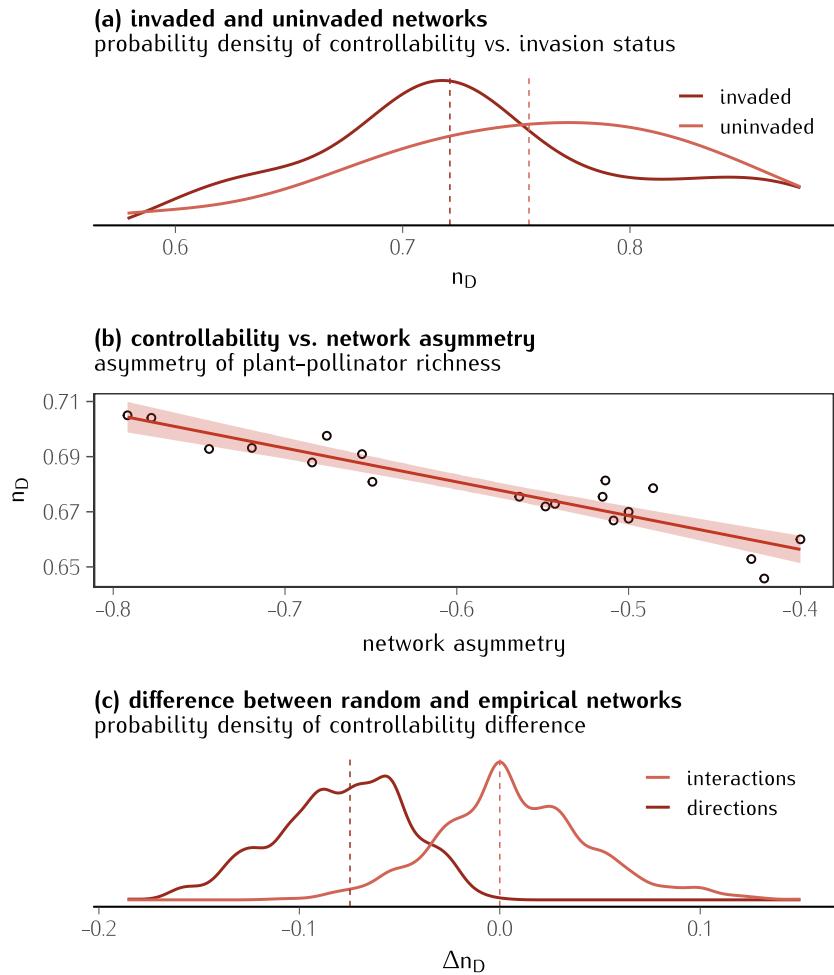


Figure 3.4: Drivers of network controllability. (a) Probability density of the relative size of the minimum driver-node set n_D in the invaded (light) and uninvaded (dark) empirical networks. (b) Relationship between the asymmetry plant/pollinator richness and n_D . (c) Probability density of the difference between the relative size of the minimum driver-node set of random networks and that of empirical networks. We randomised either the species visitation patterns (light line) or randomised the direction of control between a species pair (dark line). The vertical dashed lines in (a) and (c) indicate the median values of the distributions.

3.3.3 Species' roles

Species varied widely in their control capacity (Figure ??). Pollinators had, in average, larger control capacities than plants. That said, almost no pollinator was critical for network control, (where a species is critical for control if it has control capacity $\phi_i = 1$). Plants had a multimodal distribution of control capacity with maxima at both extremes of the distribution (Figure ??a). Intriguingly, every invasive species was critical for network control in each of their communities. The species-level models identified a positive relationship between control capacity ϕ and a species' contribution to nestedness, visitation strength, and the asymmetry of its dependences (Table ??; Figure ??; Table S4). Comparatively, species' degree was only weakly associated with control capacity (Table S5). In fact, many species with a low degree, especially pollinators, exhibited a large control capacity in their communities (Figure S10a).

Species' control capacity ϕ was only weakly correlated with commonly used centrality metrics. The Spearman correlation between these ranged between -0.14 (with betweenness centrality) and 0.42 (with eigen centrality), see Figure S11a. The correlation coefficient with degree was 0.13, however most species with high degree also tended to attain a high control capacity (Figure S10a).

Finally, we found that critical species have a particularly large impact on species coexistence when compared to non-critical species. Indeed, the structural stability of the networks where critical species were removed was considerably lower than those where non-critical species were removed ($p = 2 \times 10^{-5}$; Figure ??; Supporting Information S4).

3.3.4 Testing assumptions

We found that using visitation as a proxy for the strength of species' interactions leads to similar results than those obtained using pollinator importance (`neeman_framework_2010`), see Sup-

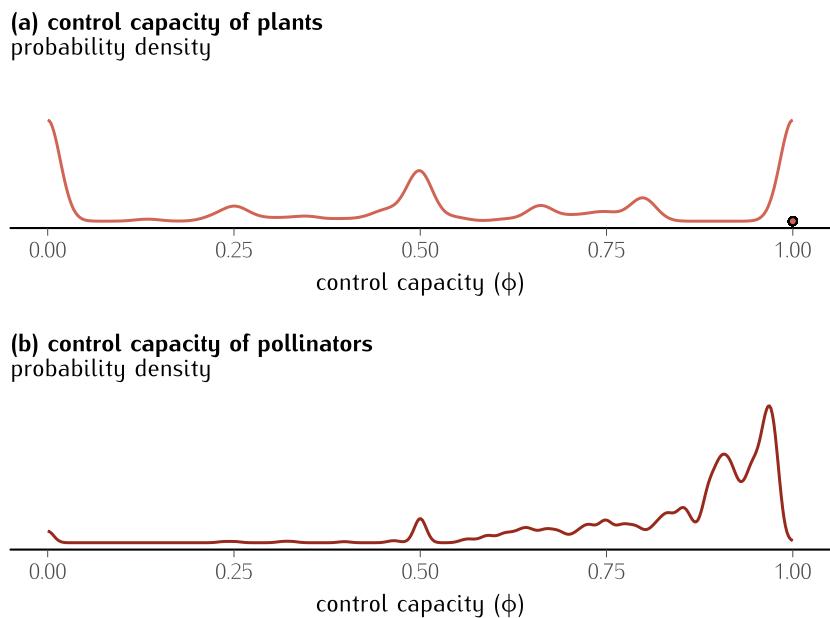


Figure 3.5: Probability density of the control capacity ϕ of (a) plants and (b) pollinators across all networks. The control capacity of all invasive species is $\phi = 1$ and is depicted with solid circles.

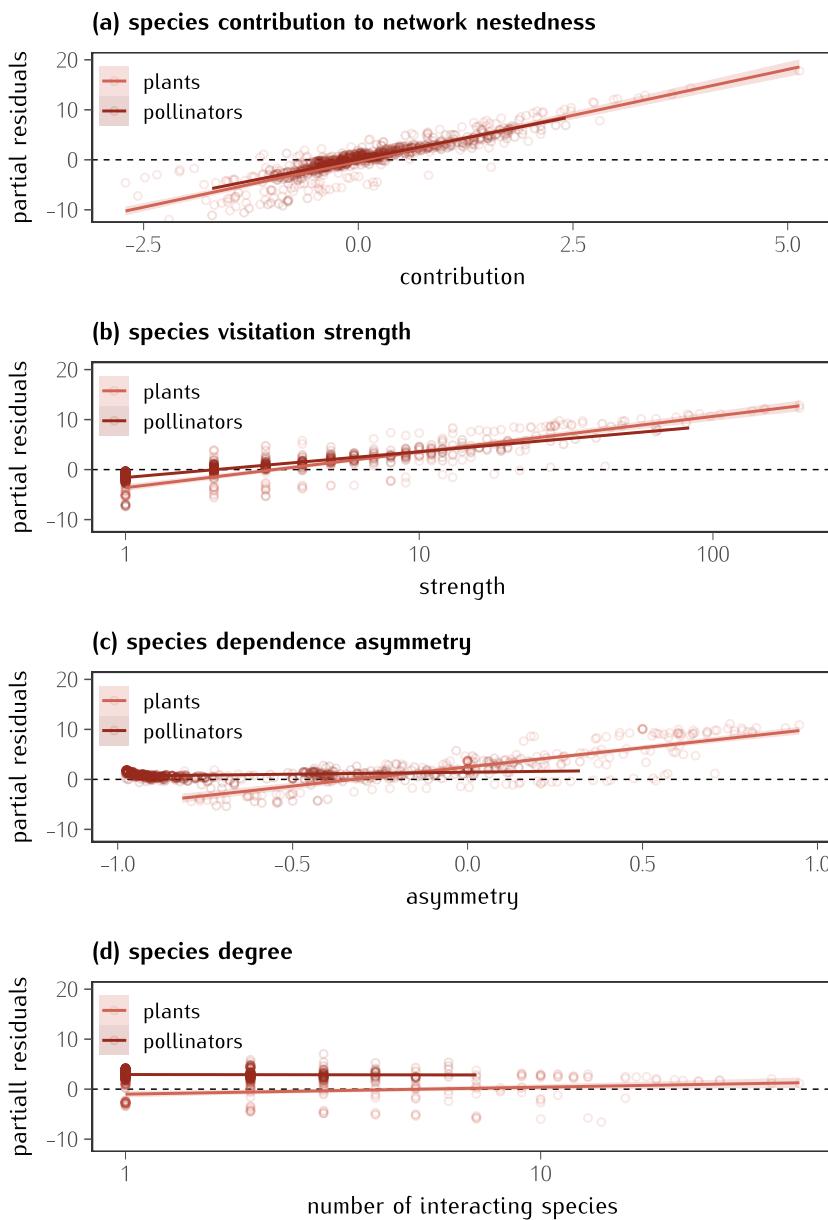


Figure 3.6: Partial-residual plots for the independent variables: (a) contribution to nestedness, (b) visitation strength, (c) asymmetry of dependences, and (d) degree. Partial-residual plots show the relationship between control capacity and each of the independent variables while accounting for all other remaining variables. Plotted values correspond to the predictions of the models weighted average.

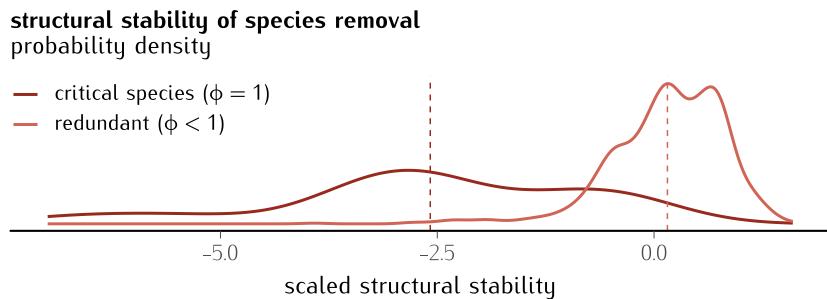


Figure 3.7: Probability density of the structural stability of the communities after a single focal species is removed. Mathematically, the structural stability of a network represents the size of the parameter space (i.e., growth rates, carrying capacities, etc.) under which all species can sustain positive abundances. The structural stability of communities in which critical species have been removed (darker line) is considerably smaller than that of communities in which non-critical species have been removed. This indicates that critical species contribute more to the stable coexistence of their communities. To allow comparison across communities, the structural stability values were scaled within each network to have a mean of zero and a standard deviation of one. Here, we assume values of the mutualistic trade-off and mean interspecific competition of $\delta = 0$ and $\rho = 0.01$ respectively. However, the choice of these parameters does not affect the results (Supporting Information S4).

Table 3.1: Selection table of the binomial generalised linear models of species' control capacity, ϕ . Only models with a weight larger or equal to 0.01 are shown.

Model terms											
int.	k	l	a	n	s	k:l	l:a	l:n	l:s	ΔAICc	W.
-1.20		+	0.80	0.15	0.29		+	+		0.00	0.48
-1.19		+	0.76	0.13	0.35		+	+	+	1.52	0.22
-1.26	-1.24	+	1.44	0.39	1.07	+	+		+	4.09	0.06
-1.37	-0.66	+	1.03		1.06	+	+		+	4.39	0.05
-1.27	-1.15	+	1.37	0.33	1.07	+	+	+	+	4.92	0.04
-1.37	-0.10	+	0.90		0.43	+	+			6.36	0.02
-1.25	-0.28	+	1.24	0.40		+	+			6.47	0.02
-1.24	-0.62	+	1.29	0.38	0.40	+	+			6.50	0.02
-1.39	0.30	+	0.83			+	+			6.72	0.02
-1.28	-0.17	+	1.16	0.32		+	+	+		7.03	0.01
-1.26	-0.53	+	1.23	0.32	0.39	+	+	+		7.42	0.01
-1.02		+	0.69	0.30	0.31		+			7.48	0.01

Note. Terms: intercept (int), degree (k), trophic level (l), asymmetry (a), contribution to nestedness (n), visitation strength (s). Model weight indicated with (W.)

porting Information [Section S5](#). Importantly, we also found that structural stability is robust to incomplete sampling of interactions. Indeed, we found strong agreement between results obtained using the complete empirical networks and those obtained by randomly removing the weakest interactions (Supporting Information [Section S6](#)). Despite removing rare interactions and species, the relative size of the minimum driver-node set, the superior species, and the relative rankings of control capacity were generally maintained. Of particular note, we found that critical species in the full network were also critical in the vast majority of rarefied networks.

3.4 DISCUSSION

Our main goal was to understand the role that species play in both modifying the abundance of the species they interact with and the state of the community as a whole. To achieve that goal we applied *structural controllability*, a field at the intersection between control and complex theory that allowed us to obtain two key pieces of information: the *controllability* of a network and a species' *control capacity*. We found that the controllability of a network does not depend on its invasion status and that the species that are critical to altering the state of the community are also the ones that most sustain the stable coexistence of species in their communities.

Our results indicate that fully controlling ecological networks might currently be out of reach for all but the smallest communities (**motter_networkcontrology_2015**). Indeed, the median size of the relative minimum driver-node set in our dataset was $n_D = 0.74$, a high value when compared to other complex systems in which controllability has been investigated (the lower n_D the more controllable the community). For instance, only gene regulation networks appear to achieve similar levels of controllability while most social, power transmission, Internet, neuronal, and even metabolic networks seem to be “easier” to control ($0.1 < n_D < 0.35$) (**liu_controllability_2011**). Structural controllability provides solid theoretical rationale for the many difficulties encountered in the management and restoration of natural communities (**woodford_confronting_2016**). Nevertheless, structural controllability might be helpful at identifying communities in which changes in the ecological state are more likely to occur.

The differences between the controllability across networks are likely to arise from differences in their structure rather than their invasion status. Specifically, when controlling for network structure, we found no difference between the controllability of invaded and uninvaded networks. Instead controllability is almost completely constrained by the patterns of species richness at each trophic guild and their degree distributions (**melian_complex_2002**; **bluthgen_specialization_2007**). These

two factors are particularly relevant because they govern the asymmetric nature of mutual dependences, which themselves provide the foundation of structure and stability in mutualistic networks (**bascompte_asymmetric_2006**; **memmott_tolerance_2004**; **astegiano_robustness_2015**).

Accordingly, our results suggest that structural controllability is closely related to the persistence of an ecological community based on two lines of evidence. First, we found a comparatively small but thought-provoking negative relationship between the controllability of a network and its nestedness. Previous studies indicate that nestedness promotes species coexistence and confers robustness to extinction (**bastolla_architecture_2009**; **memmott_tolerance_2004**) even at the expense of the dynamic stability of the mutualistic community (**saavedra_nested_2016**). These observations are in agreement with our results, as we would expect the dynamic stability (the ability to return to equilibrium after a perturbation in species abundances) of a community to be correlated to the difficulty to control it. Second, species' control capacity was strongly correlated to their contribution to nestedness and critical species had the largest impact to the stable coexistence of species in their communities. Therefore, species that play a key role at determining the state of the community might also be more key to "maintain the organization and diversity of their ecological communities", one of the hallmarks of keystone species (**mills_keystone-species_1993**).

When controlling for a species' visitation strength (the sum of a species' visits), which is indirectly a proxy of its abundance, and the net balance of its dependencies, we found that control capacity could not be easily predicted by species' degree or other metrics of centrality. For instance, some species with a low degree achieved the maximum control capacity and were critical for control in their communities. At first glance, our findings challenge numerous studies that highlight the role that central species play in the dynamics of their communities and their utility at predicting species extinctions (**jordan_keystone_2009**). However, further inspection shows that our results do not contradict these findings; most species with a large degree also have a large control capacity and all of them were classified as superior

nodes which corroborates the utility of classic centrality metrics. Putting these observations together, our results therefore take previous findings one step further and suggest that centrality might paint an incomplete picture of the relevance of species.

Other conceptual differences between structural controllability and centrality metrics provide three key insights into the conservation of ecological networks. First, structural controllability emphasizes that the effect a species has on other species is not independent of the effects that the other species have in the community. The rankings provided by centrality metrics and other heuristics fail to account for the collective influence of several species at once. Second, it demonstrates that to ensure the persistence of a community it is often necessary to consider the abundances of more than a single species, even when full control is infeasible or undesired (for example 90% of our communities contained more than one critical species). Third, structural controllability explicitly acknowledges the existence of multiple management strategies and some will be better than others depending on the context. Approaches to prioritise species for conservation and reintroduction based on traits or centrality are still useful and are likely to overlap with species' control capacity ([pires_friendship_2017](#); [devoto_understanding_2012](#)). Stepping back, our results also provide support to the idea that management decisions should not be based on a single technique but indicate that focusing on ecosystem processes and interactions may be more effective than traditional ranking-based approaches ([harvey_bridging_2017](#)). As much potential as any metric or metrics to summarise species' importance might appear to have, it's clear that we also need more empirical studies in different types of networks in order to build intuition and ground truth their usefulness.

Our choice of studying invaded/uninvaded networks was based on a desire to contrast the extensive empirical evidence of the role of invasive plants with our theoretical results. We found that invasive plants were always critical for network control and as such our results were in line with our expectations. Invasive plants have been previously found to exacerbate the asymmetries in their communities ([aizen_invasive_2008](#); [bartomeus_contrasting_2008](#);

`henriksson_strong_2016`) and to attain high centrality in their communities (`vila_invasive_2009`; `palacio_generalist_2016`). We found, however, that it is not that invasive plants have some different mechanism for influencing the community compared to their native counterparts (`emer_species_2016`; `stouffer_how_2014`). Both native species and invasive plants tended to attain a high control capacity if they were important to network persistence, were abundant, and depended little on other species. Furthermore, our observation that changes in the abundance of invasive plants (and presumably all critical species) are crucial to modify the state of the community agrees with recent evidence showing that ecosystem restoration focused on the eradication of invasive plants can have transformative desirable effects in plant-pollinator communities (`kaiser-bunbury_ecosystem_2017`). However, our results also suggest that removals must be exercised with caution. Not only it is hard to predict the direction in which the system will change, but we also show that critical species can underpin the coexistence of species and therefore some communities may be acutely vulnerable to their eradication (`traveset_invaders_2013`; `albrecht_consequences_2014`).

Structural controllability assumes that the networks can be approximated using linear functional responses (`liu_control_2016`). The ramifications of this assumption imply that, while structural controllability is useful to identify species that are relevant for network control, it cannot be used to design the *exact* interventions that should be applied to these species in order to achieve a desired state. In an ideal scenario, we would completely incorporate the species' dynamics into the controllability analysis (`cornelius_realistic_2013`); the reality is that such information is rarely available in most ecological scenarios. In contrast, structural controllability only requires a quantitative approximation of the network's interactions to gain valuable insight from the community. Furthermore, while the relationship between centrality and keystone species is based on an intuitive understanding of what a keystone species is, the assumptions of structural controllability are explicit and the estimation of a species' importance arises from a mechanistic understanding of the population dynamics between species. By accounting for network dynamics (even if in a simple way), structural stability incorporates more ecological

realism, especially in the extreme scenario in which the structure of interactions within the community only marginally affects the community's state.

ACKNOWLEDGEMENTS

The authors thank Jane Memmott and co-authors, and everyone that has made their data available to us, Takeuki Uno for the insight provided to find the set of all maximum matching algorithms, and Jason Tylianakis, Bernat Bramon Mora, Guadalupe Peralta, Rogini Runghen, Michelle Marraffini, Mark Herse, Warwick Allen, Matthew Hutchinson, and Marilia Gaiarsa for feedback and valuable discussions. E.F.C. acknowledges the support from the University of Canterbury Doctoral Scholarship, the University of Canterbury Meadow Mushrooms Postgraduate Scholarship, a New Zealand International Doctoral Research Scholarship, and a travel grant from the European Space Agency. D.B.S. acknowledges the support of a Marsden Fast-Start grant and a Rutherford Discovery Fellowship, administered by the Royal Society Te of New Zealand Aparangi.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article (<https://doi.org/10.1111/1365-2745.13147>).

DATA ACCESSIBILITY

All data used in this manuscript have already been published by [lopezaraiza-mikel_impact_2007](#) [bartomeus_contrasting_2008](#) and [ballantyne_constructing_2015](#). The reader should refer to the original sources to access the data.

A

COPHYLOGENY OF POLINATION INTERACTIONS

Después reflexioné que todas las cosas le suceden a uno precisamente, precisamente ahora. Siglos de siglos y sólo en el presente ocurren los hechos; innumerables hombres en el aire, en la tierra y el mar, y todo lo que realmente pasa me pasa a mí...

— borges_jardin_1941

This Chapter has been published by **hutchinson_cophylogenetic_2017**

SUMMARY

That evolutionary history can influence the way that species interact is a basic tenet of evolutionary ecology. However, when the role of evolution in determining ecological interactions is investigated, focus typically centers on just one side of the interaction. A cophylogenetic signal—the congruence of evolutionary history across both sides of an ecological interaction—extends these previous explorations and provides a more complete picture of how evolutionary patterns influence the way species interact. To date, cophylogenetic signal has most typically been studied in interactions that occur between fine taxonomic clades that show high intimacy. In this study, we took an alternative approach and made an exhaustive assessment of cophylogeny in pollination interactions. To do so, we assessed the strength of cophylogenetic signal at four distinct scales of pollination interaction: (i) across plant-pollinator associations globally, (ii) in local pollination communities, (iii) within the modular structure of those

communities, and (iv) in individual modules. We did so using a globally-distributed dataset comprised of 54 pollination networks, over 4000 species, and over 12,000 interactions. Within these data, we detected cophylogenetic signal at all four scales. Cophylogenetic signal was found at the level of plant-pollinator interactions on a global scale and in the majority of pollination communities. At the scale defined by the modular structure within those communities, however, we observed a much weaker cophylogenetic signal. Cophylogenetic signal was detectable in a significant proportion of individual modules and most typically when within-module phylogenetic diversity was low. In sum, the detection of cophylogenetic signal in pollination interactions across scales provides a new dimension to the story of how past evolution shapes extant pollinator-angiosperm interactions.

A.1 INTRODUCTION

Populations do not exist in isolation but are instead constantly interacting with each other. Each of these interactions can impact the fitness of individuals and hence lead to selection for amplification or avoidance of future interactions (**gervasi_real-time_2017**; **thompson_geographic_2005**). Furthermore, when interactions directly influence the reproductive isolation of one or both species, selection can be powerful enough to cause speciation or extinction events that can potentially intertwine the evolutionary trajectories of pairs of taxa and their descendants (**thompson_geographic_2005**). Where selection, regardless of its origin, is strong enough to drive coupled speciation in two interacting clades, the resultant macroscopic pattern is synonymously referred to as cophylogeny, co-speciation, or parallel cladogenesis (**thompson_geographic_2005**; **page_tangled_2003**).

A cophylogenetic signal implies two observations: that the phylogenies of interacting clades are congruent in structure and that extant interactions occur between evolutionarily coupled taxa (**page_tangled_2003**; **desdevises_cophylogeny_2007**; **balbuena_paco_2013**). A cophylogenetic signal suggests that contemporary ecological

associations among species are the product of coupled evolutionary history such that ancestral forms of each species experienced and responded to shared selection pressures ([page_tangled_2003](#); [aizen_phylogenetic_2016](#)). In contrast, the current paradigm of phylogenetic signal of species interactions suggests only that more closely related species interact in more similar ways ([rezende_non-random_2007](#); [gomez_ecological_2010](#); [rafferty_phylogenetic_2013](#)). Therefore, cophylogenetic signal in ecological networks would suggest that coupled evolutionary history as well as relatedness can determine species interactions, thereby providing additional insights into the role past evolution plays in determining contemporary ecological associations.

To date, the vast majority of studies of cophylogeny have centered on host-parasite relationships ([hafner_phylogenetic_1988](#); [de_vienne_cospeciation_2013](#)) where the focus is often at the relatively fine scale of families and genera ([weckstein_biogeography_2004](#); [desdevises_cophylogeny_2007](#); [hughes_multiple_2007](#); [page_tangled_2003](#); [chomicki_m韆croevolutionary_2015](#)). If cophylogenetic signal can also be considered a relevant predictor of ecological interactions, it should be detectable in other types of ecological associations. In particular, there is an increasing focus on the mutualistic assemblages of flowering plants and their pollinators as another system in which one might expect to detect a cophylogenetic pattern. Several recent studies have demonstrated the presence of cophylogeny at the scales of fig-wasp ([marussich_host-specificity_2007](#); [jousselin_one_2008](#); [cruaud_extreme_2012](#)) and yucca-moth ([althoff_geographic_2012](#)) pollination interactions ([aizen_phylogenetic_2016](#)). Findings such as these—coupled with the facts that the association of angiosperms and pollinators dates back to the Cretaceous period ([crane_origin_1995](#); [grimaldi_co-radiations_1999](#)) and that there is reasonable evidence to expect that at least some angiosperms and pollinators co-diverged ([grimaldi_co-radiations_1999](#))—suggest a cophylogenetic signal may be widely identifiable between these groups (Figure ??). On the other hand, perfect congruence of speciation patterns cannot be expected across such rich and diverse groups after more than 100mya of association. Nevertheless, even a weak cophylogenetic signal between angiosperms and their pollinators (i.e., at taxonomic scales above the species level) would

provide important evidence that coupled evolution between taxa is an important correlate of their tendency to interact.

The most relevant scale at which cophylogenetic signal could characterize ecological interactions is that of the community. It is already well established that elements of community composition, such as evenness, functional trait diversity and interaction structure have an element of phylogenetic determinism ([webb_phylogenies_2002](#); [eklof_phylogenetic_2016](#); [harmon-threatt_filtering_emerson_phylogenetic_2008](#); [vamosi_emerging_2009](#)). Hence, even when recognizing the various caveats of a phylogenetic approach to community ecology ([mayfield_opposing_2010](#); [losos_seeing_2011](#)), there are clear implications should cophylogenetic signal be detectable in ecological communities. For instance, many studies have focused on how the evolutionary history of a single group (e.g. forest trees, pollinators) influences community structure ([cavender-bares_phylogenetic_2006](#); [kembel_phylogenetic_2006](#); [danieli-silva_pollination_2012](#)); in contrast, detectable cophylogenetic signal would suggest that observed interactions tend to occur between taxa that show coupled evolutionary history and therefore that the structure of pollination networks is, at least partially, the by-product of this evolutionary coupling of taxa (Figure ??). Although recent findings for a set of related networks indicate that cophylogenetic signal may be detectable in pollination networks ([aizen_phylogenetic_2016](#)), it is currently unclear whether or not this is generally the case.

At the same time, ecological communities can also be stochastic assemblages of species and interactions ([hubbell_unified_2001](#); [cottenie_integrating_2005](#)) and the value of examining smaller groups of closely interacting species has been shown time and again ([paine_food_1966](#); [estes_sea_1974](#); [olesen_modularity_2007](#); [rezende_compartments_2009](#)). Therefore, it is entirely possible that a cophylogenetic signal also permeates to finer scales within a community. Indeed, ecological networks are known to have identifiable structural features including being characterized by modules of closely interacting subsets of the community ([barber_modularity_2007](#); [thebault_stability_2010](#)). Modules—a pervasive feature in pollination networks ([olesen_modularity_2007](#))—are thought to play crucial roles in ecological community resili-

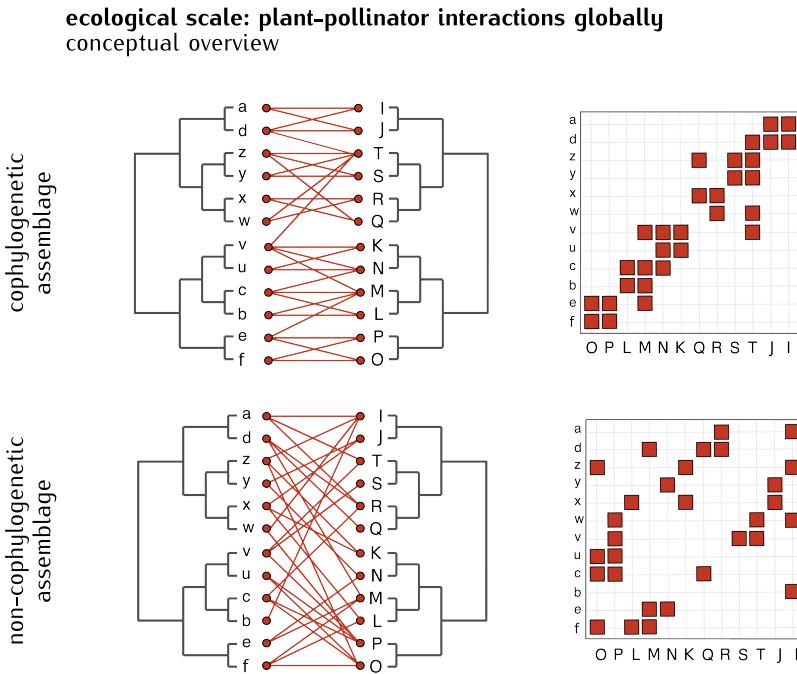


Figure A.1: The first scale of investigation is plant and pollinator interactions globally. We aggregate 54 local pollination communities into a single network to provide a global view of pollination interactions. Here and in the following conceptual figures, we provide a representative example of pollination-interaction structures that are cophylogenetic and non-cophylogenetic. Across all four scales (shown here and in Figure ??, ??, and ??) the key to cophylogenetic signal is that the observed interactions tend to occur between species that show coupled evolutionary history (i.e., their speciation patterns match). Though the general idea is similar across them, our hypothesis tests at each scale address a slightly different question regarding the presence of cophylogenetic signal in pollination interactions.

ecological scale: plant-pollinator communities
conceptual overview

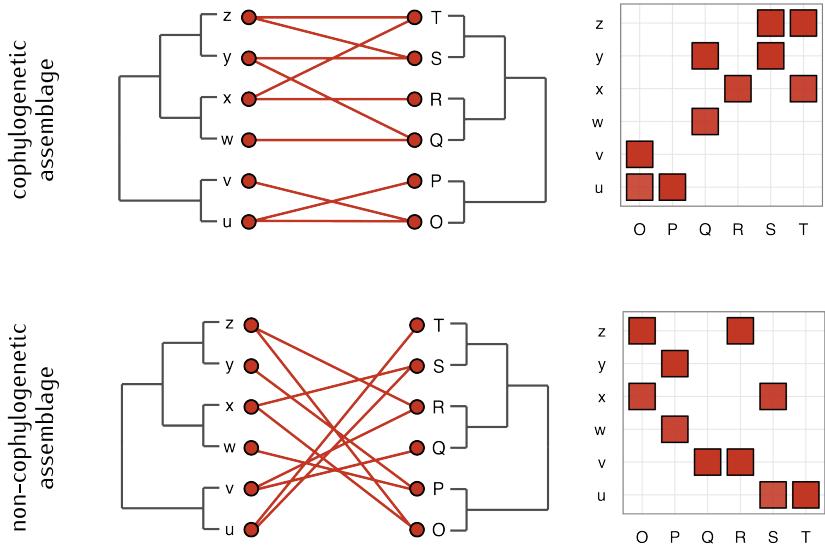


Figure A.2: The second scale of investigation are local communities of interacting plants and pollinators that co-occur in a given location. Each community is a subset of the network in ???. At this level, we are asking if those interactions that are realized in local communities tend to occur between plants and pollinators with the most congruent evolutionary histories.

ence (**thebault_stability_2010**) and may represent a fundamental unit of coevolution (**olesen_modularity_2007; bascompte_mutualistic_2014**). As such, these tight sets of interacting species may provide a more ecologically relevant scale at which cophylogenetic signal could act. On the one hand, a network's modular structure may show the clearest cophylogenetic signal (Figure ??). On the other hand, many different processes have been causally attributed to ecological module formation (**olesen_modularity_2007; rezende_compartments_2009; krasnov_phylogenetic_2012; rohr_structural_2014; schleuning_ecological_2014**), implying that some individual modules within networks may be better characterized by cophylogenetic signal than others (Figure ??).

For a network's modular structure to show a cophylogenetic pattern, two constraints must be satisfied (Figure ??). First, the modular structure should embody the phylogenetic congruence of the network such that modules represent groupings of closely related species on each side of the network. However, this mapping of each side of the modules to phylogenies does not take into account the degree to which those interactions within modules occur between evolutionarily-coupled plants and pollinators. Thus second, a network's modules should also be comprised of the interactions that contribute most to the cophylogenetic signal of the network while interactions that contribute less should tend to fall between modules. At the even finer scale of individual modules, a module could be considered to show cophylogenetic signal when just its interactions show greater cophylogenetic signal than expected by chance (Figure ??).

Here, we explore cophylogenetic signal between plants and their pollinators in 54 pollination networks from around the world that together provide a taxonomically and geographically diverse dataset. In particular, we leverage these data to quantify the evidence of cophylogenetic signal at the four distinct scales mentioned previously (Figure ??, ??, ??, and ??): (*i*) cophylogenetic signal between angiosperms and pollinators globally, (*ii*) community cophylogenetic signal—where evolutionary congruence between species should be embodied by the interactions of the entire network, (*iii*) cophylogenetic signal of a network's modular structure—where the modular structure of a network should

ecological scale: modular structure of plant-pollinator networks
conceptual overview

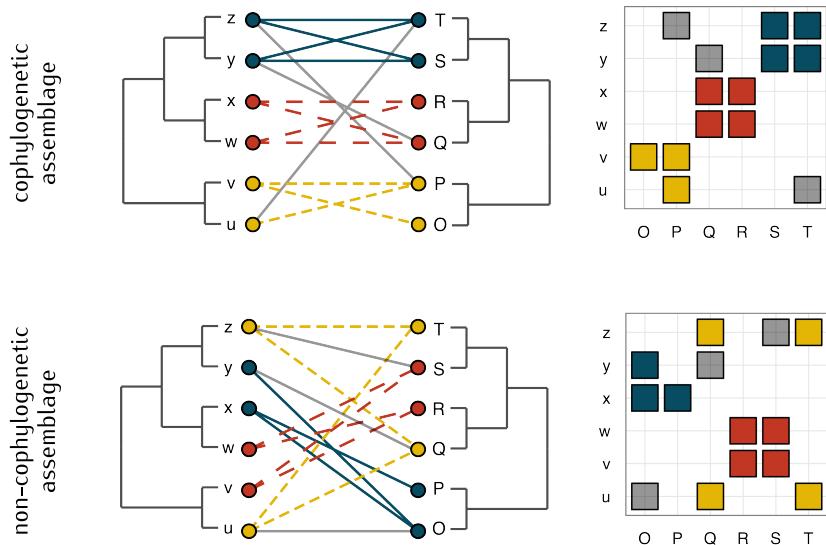


Figure A.3: The third scale of investigation is with the same local communities as in Figure ?? but examined at the perspective of their modules, where modules group together species based on how likely they are observed to interact. At this scale, we examine how the modular structure of those communities captures cophylogenetic signal. Our hypothesis here derives from the suggestion that the module is the fundamental unit of coevolution and asks whether interactions within modules tend to occur between those species with congruent evolutionary history and vice versa for interactions that fall between modules.

ecological scale: individual modules within plant-pollinator networks
conceptual overview

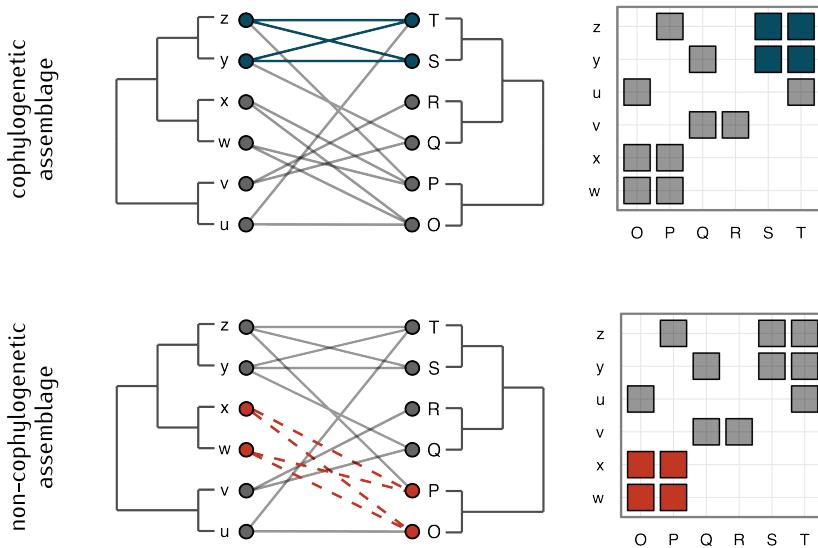


Figure A.4: The fourth and final scale of investigation is at individual modules (groups of frequently interacting plants and pollinators) extracted from Figure ???. They are also subsets of the local communities in Figure ???. Many factors are thought to contribute to module formation and cophylogenetic signal may therefore not be observable across all modules in a community. As a result, we assess cophylogenetic signal in individual modules where the interactions of a module should occur between species with congruent evolutionary history more than would be expected by chance.

tend to contain more closely related plant species, more closely related pollinator species, and the most evolutionarily congruent interactions between them, and (iv) individual module cophylogenetic signal—where interactions within a module should be more cophylogenetic than expected by chance. Detectable cophylogenetic signal across these four scales suggests that, at least in plant-pollinator associations, the evolutionary determinants of extant interactions are a product of both interacting species rather than arising from only the phylogenetic relatedness on one side of the interaction.

A.2 METHODS

A.2.1 Empirical data and phylogeny construction

We analyzed a dataset comprised of 54 binary, plant-pollinator mutualistic networks from a wide range of locations around the globe and with diverse species assemblages (Supporting Information [Sections S4 and S5](#)). In each of the networks, the presence or absence of interactions is based on observed visitation of flowering plants by their animal pollinators. In total, these networks include 1,318 species of flowering plants, 2,930 species of pollinators, and over 12,000 unique interactions.

Studying cophylogenetic signal between two sets of interacting species, such as the flowering plants and pollinators that we examine here, requires an understanding of the evolutionary history of both groups. We followed several steps to generate sufficiently well-resolved phylogenies of flowering plants and their pollinators. First, to ensure all species identifications were up-to-date, we verified all species' names in the original interaction matrices. Plant names were checked and corrected with the NCBI database (<http://www.ncbi.nlm.nih.gov/>) whereas we corrected animal names with the resolve function in the R package taxize, which accessed a range of taxonomic databases ([chamberlain_taxize_2014](#)). We combined these species lists with published mega-phylogenies of plants ([zanne_three_2014](#)) and

insects ([misof_phylogenomics_2014](#)) to generate dated phylogenetic trees of our data. The two published trees provide a backbone for the construction of our phylogenies in that divergence patterns and dates of major lineages can be used as the basis of the phylogenies specific to our data. For the plants, major nodes are fully resolved down to the family level (with some dating below the family level). For the insect pollinators, the backbone tree provides resolution to the order level. It is most important for cophylogenetic analysis that these major nodes are dated accurately as it is thought that matching at these deeper phylogenetic scales drives observable cophylogenetic signal ([aizen_phylogenetic_2016](#)). Below the family- and order-level, respectively, we largely rely on taxonomic information to infer evolutionary relationships between taxa. Full details of phylogeny construction are available in the Supporting Information [Section S1](#)

A.2.2 Measuring cophylogenetic signal

To conduct a direct assessment of cophylogenetic signal between angiosperms and pollinators across these four scales, we implemented a recently developed Procrustean method: Procrustean Approach to Cophylogeny ([balbuena_paco_2013](#)). This approach addresses the cophylogeny problem by optimizing the fit of the phylogeny-interaction graphs of each network ([balbuena_paco_2013](#)). The cophylogenetic signal of each individual interaction is given by the squared residual distance (r^2) between the two corresponding points in the phylogenetic graphs. PACo thus returns a quantification of the global fit of the phylogenetic objects based on observed interactions as the sum of squared residual distances ($\mathcal{R} = \sum r^2$) between phylogenetic-interaction graphs ([balbuena_paco_2013](#)). As in regression, the smaller the residual distance, the better the fit of the two phylogenies to each other and the more support for a hypothesis of cophylogenetic signal as reflected by the extant interactions.

PACo, as implemented in R, offers several configuration options ([hutchinson_paco_2017](#)). In this study, we have focused on the

results of the symmetric method where the normalized plant graph is superimposed on the normalized pollinator graph. This means that we assessed cophylogenetic signal in terms of the plant phylogeny tracking the pollinator phylogeny since insect lineages preceded angiosperms ([misof_phylogenomics_2014](#)) and pollinators have been shown to drive the evolution of plants ([gervasi_real-time_2017](#)). It also implies that \mathcal{R} is standardized with respect to the two phylogenies rather than in units relative to the pollinator phylogeny (as it would with an asymmetric superimposition). Importantly, our results do not differ qualitatively when selecting alternative configuration of PACo arguments (Supporting Information [Section S6](#)).

Currently, there are a range of tools available for undertaking cophylogenetic analysis, and these methods can mainly be sorted into event-based and global-fit methods. Event-based methods reconcile one phylogenetic tree with the other by directly assessing evolutionary events (i.e., cospeciation, duplication, host switches) that are explicit in the tree topology ([conow_jane_2010](#); [drinkwater_towards_2016](#)). Conversely, global-fit methods, such as ParaFit and PACo, aim to assess cophylogeny based on the congruence of observed interactions relative to the phylogenies rather than with the specifics of the phylogenetic topologies ([balbuena_paco_2013](#); [legendre_statistical_2002](#)). An important consequence of this difference in approaches is the manner in which significance of the observed cophylogenetic statistic tends to be inferred. The randomization approach implemented in PACo (and other global-fit methods) maintains the topology of the phylogeny of each group while shuffling the associations (i.e., interactions) between species to generate random instances of the observed data ([balbuena_paco_2013](#); [hutchinson_paco_2017](#)). Conversely, event-based methods such as Jane ([conow_jane_2010](#)) and CoRe-PA ([merkle_parameter-adaptive_2010](#)) instead permute the topology of the phylogenies due to their explicit focus on specific events. Both approaches have limitations ([balbuena_paco_2013](#); [drinkwater_towards_2016](#)) and we take a global-fit approach here because it is the most amenable to the data we have and network-centric questions we explore.

Consequently, the approach that we adopt to study cophylogeny necessitates the shuffling of the association matrix to estimate the null distribution (**balbuena_paco_2013**; **hutchinson_paco_2017**). Rather than allow the results to be driven by variation in species' specificity or generalism, we also constrain this randomization so that each species' number of interactions is maintained when the associations are shuffled (**fortuna_nestedness_2010**). For all instances of shuffling the association matrix, we use 1000 permutations of the data to generate the null (and hence to infer significance of the observed pattern); we determined that this number of permutations was sufficient, here and in the related tests that follow below, by assessing convergence of the resulting p-values used for inference (Supporting Information [Section S11](#)).

A.2.3 Global-scale cophylogenetic signal

We first used PACo to make a global assessment of cophylogenetic signal between flowering plants and pollinators using the global-scale phylogenies and the aggregate interaction network of our dataset, where all observed interactions between angiosperms and pollinators are represented. We assessed the significance of observed cophylogenetic signal at a global scale with a Monte Carlo approach whereby the observed (\mathcal{R}) was compared to the same value (\mathcal{R}^*) from an ensemble of 1000 randomizations of the aggregated interaction network. In each of these randomizations, we conserved the number of interactions for each species as well as the total number of interactions in the network (**fortuna_nestedness_2010**). A conservative null model such as this preserves any influence on cophylogenetic signal of the total number of species interactions or the degree distribution meaning that deviation from the null model can be interpreted as due to properties of the particular species that interact rather than network topology. We considered the cophylogenetic signal of the observed network to be significant if \mathcal{R} was smaller than the null distribution of \mathcal{R}^* , at $\alpha = 0.05$. Since our hypothesis of significant cophylogenetic signal at the global scale implies that

\mathcal{R} should be significantly smaller than the null expectation, we performed one-tailed tests for this global analysis.

To further examine any large-scale geographic patterns, we also constructed continent-specific phylogenies and networks. With the same statistical approach as for the global data, we examined cophylogenetic signal between plants and pollinators at the continental scale. The extent to which the interactions between plants and pollinators at a regional scale showed cophylogenetic signal was assessed for the six continents from which the empirical networks originate (Africa, Asia, Europe, North America, Oceania, South America). We followed the same approach as for the global scale. We assessed whether empirical \mathcal{R} at the regional scale was smaller than its null expectation (i.e., the same null model as above), at $\alpha = 0.05$.

A.2.4 Community cophylogenetic signal

We next explored cophylogenetic signal at the scale of pollination communities from two angles. First, we assessed the cophylogenetic signal of the species assemblages themselves while maintaining the observed interaction structure—i.e., do the species present in each network represent a more cophylogenetic assemblage than if we were to draw the same number of species randomly from our full phylogenies? Second, we assessed the cophylogenetic signal of the interaction network itself—i.e., do the interactions observed in each network occur between more evolutionarily coupled species than we would expect by chance? To address these questions, we first quantified the cophylogenetic signal of each of the 54 empirical networks.

A.2.4.1 *Cophylogenetic signal based on interaction network*

The first assessment of cophylogenetic signal at the community scale was the degree to which realized interactions between plants and pollinators showed a greater cophylogenetic signal than expected by chance. If coupled evolution is an important determinant of species interactions then we would expect that

the interactions observed in a network show a stronger cophylogenetic signal than expected by chance. To make this assessment in our dataset, we constructed a null model whereby observed interactions between pollinators and plants within a network were shuffled to create a randomization of the empirical network. The shuffling procedure—identical to that used for the global-scale analysis of cophylogenetic signal—fixes the total number of interactions as well as the degree of plants and pollinators. We used a Monte Carlo approach to compare the cophylogenetic signal of observed communities to the same (\mathcal{R}^*) from an ensemble of 1000 randomizations of the network. We considered the cophylogenetic signal of the observed interaction network to be significant if its \mathcal{R} was smaller than the null distribution, at $\alpha = 0.05$.

A.2.4.2 *Cophylogenetic signal based on species assemblage*

As mentioned earlier, a cophylogenetic signal may be detectable between two species purely because of coupled evolution of their ancestors (rather than the two species themselves). Therefore, it was also important to assess whether cophylogenetic signal seen at the community scale was different to that produced by the chance co-occurrence of species and evolutionary coupling between their ancestral states. To do so, we constructed a null model whereby pollinators and plants were randomly sampled from the full phylogenies (in equivalent numbers to those observed in the assemblage) and assigned to the interaction network. We used a Monte Carlo approach to assess whether or not the cophylogenetic signal seen in observed communities was different to those created by chance with the comparison of an empirical network's \mathcal{R} to the same (\mathcal{R}^*) from an ensemble of 1000 randomizations of that network following the outlined approach. We considered the cophylogenetic signal of an empirical species assemblage to be significant if its \mathcal{R} was smaller than the null distribution, at $\alpha = 0.05$.

A.2.5 Modules in pollination networks

To test whether the modular structure of each network and individual modules themselves show a cophylogenetic signal, we first needed to identify modules in each of the interaction networks. To do so, we followed the approach proposed by **barber_modularity_2007** and implemented in MODULAR (**marquitti_modular_2014**) where nodes in binary bipartite networks are partitioned across modules via a stochastic-optimization procedure—simulated annealing—to maximize the modularity measure, Q_B (**marquitti_modular_2014**). While there are several such methods to assess modularity, the method employed here has been shown to perform as well or better than other contemporary module detection algorithms in binary bipartite networks (**thebault_identifying_2013**). Although by and large we use binary interaction networks in our analysis, quantitative networks can provide additional information with which to identify modules (**dormann_method_2014**). To assess the influence of a quantitative interaction network approach, we also studied the quantitative versions available for 15 of the networks in our dataset. For each of these, we followed the same approach to cophylogenetic signal at the modular scale as for binary networks. We undertook all of the analyses that we outline in the following methodological sections for both our full dataset of binary networks and this subset of quantitative networks.

A.2.6 Cophylogenetic signal of a network's modular structure

While the degree to which communities as a whole show cophylogenetic signal could be established purely with PACo (**balbuena_paco_2013**), an assessment of the extent to which cophylogenetic signal is manifest in the modular structure of those networks required subsequent analysis. To do so, we tested two aspects of a network's modular structure: i) the degree to which closely related species co-occur in modules (for both plants and pollinators) and ii) the degree to which interactions within

modules tend to show a stronger cophylogenetic signal than interactions between modules. Finally, we used Chi-squared tests to assess the degree to which a significant test statistic for modular structure cophylogenetic signal, plant module phylogenetic signal, or pollinator module phylogenetic signal was related to a significant test statistic in the other two measures.

A.2.6.1 Phylogenetic congruence of module assignments

The first step we took was to quantify how plant and pollinator modules reflect the evolutionary history of each group. To do so, we inferred the phylogenetic signal present in species' module assignments using a likelihood-ratio test (**cadotte_phylogenies_2018**). In the case of phylogenetic signal of a discrete trait, such as module assignment, significant phylogenetic signal is based on the comparison of two candidate models of trait evolution using Pagel's lambda (**pagel_inferring_1999**). In the first model, λ is optimized based on the observed tree and observed traits. In the second model, the tree is first transformed based on $\lambda = 0$ (i.e. the tree is transformed into a star phylogeny, or, a single, large polytomy) and λ is again optimized. The degree of phylogenetic signal in the trait—module assignment—can then be inferred with a likelihood-ratio test (LRT) that compares how well each model, or version of the tree, explains the trait data. The LRT, therefore, assesses the degree to which the observed tree topology provides a better explanation of module assignment than an uninformative phylogeny. We considered the module assignment of the plant or pollinator species in a network to show significant phylogenetic signal if the fit of the model with the observed tree was significantly better than the fit of the model with the star phylogeny, at $\alpha = 0.05$. Both models and tree transformations were implemented with the `ape::fitDiscrete` function in R with the lambda transformation and equal-rates model (**paradis_ape_2004**) and followed the approach suggested by **pagel_inferring_1999** and reiterated by **cadotte_phylogenies_2018**.

A.2.6.2 Cophylogenetic signal within and between modules

The previous analysis describes the degree to which closely-related plants and/or pollinators co-occur in modules. However, it does not consider the degree to which the interactions within those modules occur between evolutionarily-coupled taxa. Therefore, alongside our assessment of module phylogenetic signal, we also assessed the degree to which the modular structure of a pollination network is characterized by cophylogenetic interactions. If the modular structure of a network is characterized by a cophylogenetic signal, we expected interactions within modules to have a higher degree of congruence (*i.e.*, have smaller residuals r on average) than interactions between species in different modules. To assess whether there was in fact a relationship between r and modular structure in each network, we calculated the average residual distance of interactions within modules \bar{r}_w , the average residual distance of interactions that occur between modules \bar{r}_b and defined a test statistic $d_m = \bar{r}_w - \bar{r}_b$. We then compared the value of this test statistic to the same (d_m^*) for 1000 instances of the empirical network with randomized module assignments. Since our hypothesis of significant cophylogenetic signal in the modular structure of a network implies that the empirical d_m should be significantly smaller than the null expectation, we performed a one-tailed test here.

To generate null expectations for both of these analyses, we randomized the species' module assignments using two approaches. In the first, more conservative, approach we created random modules for each network by maintaining the observed number of modules and the number of species within each of them (*same-sizes* null model). This approach preserves the modular structure of the network and just shuffles species between modules. In the second approach, we created random modules for each network by allowing for a random number of species in each module and a random number of modules (*all-sizes* null model). An approach such as this reconfigures the network's modular structure by potentially changing the number and size of its modules. Differences between approaches are not substantial and do not qualitatively affect our results or conclusions, therefore we present the results of the first, more conservative

approach here. Results for the second approach can be found in the Supporting Information [Section S6](#).

A.2.7 Cophylogenetic signal of individual modules

Finally, we wanted to understand whether and how individual modules within the same network varied in their cophylogenetic signal. To do so, we assessed the degree to which each empirical module tends to show a greater cophylogenetic signal than expected by chance. For each empirical module, we first generated a distribution of the relationships between the cophylogenetic signal of its interactions and the cophylogenetic signal of interactions in each of 1000 random modules with an equivalent number of randomly sampled interactions. Random modules were drawn from randomizations of the networks that preserved connectance and degree distribution (the same null model described earlier for our assessment of cophylogenetic signal the global and community scales). In each case, we assessed whether the empirical module shows a stronger cophylogenetic signal than an equivalent random module by comparing the r of all the interactions within the empirical module (r_i) to the same for a random module (r_i^*) with a Wilcoxon-signed rank test. If our hypothesis of stronger cophylogenetic signal in the empirical module can be supported, then r_i should tend to be smaller than r_i^* . From this distribution of 1000 comparisons of the empirical module with a random module, we were able to characterize a module as significantly cophylogenetic if its interactions tended to be smaller than their random counterparts, at $\alpha = 0.05$. Since our hypothesis of significant cophylogenetic signal within a module implies that empirical modules should always show a greater cophylogenetic signal than chance, we performed a one-tailed test here.

A.2.7.1 *Determinants of cophylogenetic signal in individual modules*

To better understand variation in cophylogenetic signal between modules, we quantified several module characteristics. For each module, we calculated module size (total number of species, plant richness, pollinator richness), module degree (the number of interactions within the module and the total number of interactions of participant species), and the phylogenetic diversity of the module for both plants and pollinators (**faith_conservation_1992**). After checking for correlation or near co-linearity between explanatory variables (Pearson's $r < 0.7$), we used a logistic regression to assess the relationship, at $\alpha = 0.05$, of (i) overall species richness of the module, (ii) the proportion of participant species' interactions that occur within the module, and (iii) phylogenetic diversity (**faith_conservation_1992**) of both flowering plants and pollinators with module cophylogenetic signal.

A.2.7.2 *Distribution of participation in cophylogenetic modules across the phylogeny*

Lastly, we examined how the species that made up these modules were distributed across the plant and pollinator phylogenies to assess whether participation in a cophylogenetic module is phylogenetically clustered for either the plants or pollinators. To do so, we treated the participation of species in a significantly cophylogenetic module as a binary trait—0 if never found in such a module and 1 if found in at least one such module. Based on this trait, we constructed a distance matrix m_d of cophylogenetic module assignment across all plant species, and the same across all pollinator species. Note that, in these symmetric distance matrices, species pairs are assigned a 0 if they both participated in any cophylogenetic module or if they both did not, and a 1 otherwise. We then used a Mantel test to gauge whether or not cophylogenetic module participation as given by the distance matrix m_d was predicted by the phylogenetic relatedness of plants and pollinators, respectively, as captured by the phylogenetic variance-covariance matrix m_v . We then compared the Z-statistic of the observed matrices to the

same (Z^*) from an ensemble of 1000 matrix randomizations which preserved the non-independence of the underlying distances with `ape::mantel.test` (`paradis_ape_2004`). A significant Z -statistic (at $\alpha = 0.05$) here indicates that participation in co-phylogenetic modules is clustered on the phylogeny more than expected by chance. As participation in cophylogenetic modules may either clustered or over-dispersed on the phylogeny, we undertook a two-tailed test here.

A.2.8 The effect of exotic species on cophylogenetic signal

The final analyses we undertook were an exploration of the effect that exotic species had on the cophylogenetic signal that we detect. To do so, we identified those networks in our analysis that contained exotic species. Specifically, we found references to particular exotic species in the original publications of these networks and used that subset of our data to answer this question. In all, nine networks out of the full set of 54 explicitly identified exotic species. For these networks, we removed the exotic species identified by the original authors and re-analyzed the data at all scales. We assessed cophylogenetic signal at the community scale, at the scale of the modular structure of the community, and within individual modules. These results do not show meaningful differences from the case where exotic species are included, we present the corresponding results in the Supporting Information [Section S10](#).

A.3 RESULTS

We first present cophylogenetic analysis of pollination interactions at the global scale where local interaction networks were aggregated into a single conglomerate network. While our dataset is extensive, it does not fully encompass all flowering plant and pollinator species. In terms of representativeness, our global-scale plant phylogeny contains 38 out of 68 recognized orders,

133 out of 489 recognized families, and 761 genera. The global-scale pollinator tree contains 24 orders, 263 families, and 1595 genera. More detail regarding taxonomic diversity can be found in the Supporting Information [Section S5](#). Despite leaving some taxa unrepresented, there is nonetheless, a significant cophylogenetic signal between plants and pollinators at the global scale (Monte Carlo test, $p < 0.001$). Moreover, we split this global dataset into continent specific trees and networks to examine regional patterns in cophylogenetic signal. In each region, we see that interactions between pollinators and plants show the same pattern of significant cophylogenetic signal that is seen at the global scale (Monte Carlo tests; all $p < 0.001$).

At the community level, our results show that most pollination communities exhibit a significant cophylogenetic signal. The observed interaction patterns of a majority of empirical pollination networks show significant cophylogenetic signal compared to an ensemble of random network structures (38 out of 54 networks; Monte Carlo test, $p < 0.05$; Figure ??). Similarly, a large proportion of empirical networks show significantly stronger cophylogenetic signal than assemblages of plants and pollinators randomly sampled from the full phylogenies (24 out of 54 pollination networks, Monte Carlo test, $p < 0.05$; Figure ??).

The frequent observation of significant cophylogenetic signal at a network scale, such as this, suggests that cophylogenetic signal should also be manifest in the modules of these networks and perhaps more so, given that modules are groups of closely interacting species. If module participation is a direct result of cophylogenetic association, we first expected that modules based on who interacts with whom are consistent with the evolutionary histories of both flowering plants and pollinators. Here, we instead see that the modular structure of a network is consistent with the pollinator and plant phylogenies in only 43% and 17% of networks, respectively (Monte Carlo test, $p < 0.05$; Figure ??). Our second consideration to assess the cophylogenetic signal of a network's modular structure was the degree to which within-module interactions are more congruent with a cophylogenetic hypothesis than those interactions between modules. Here, we find that the modular structure of a network shows a signific-

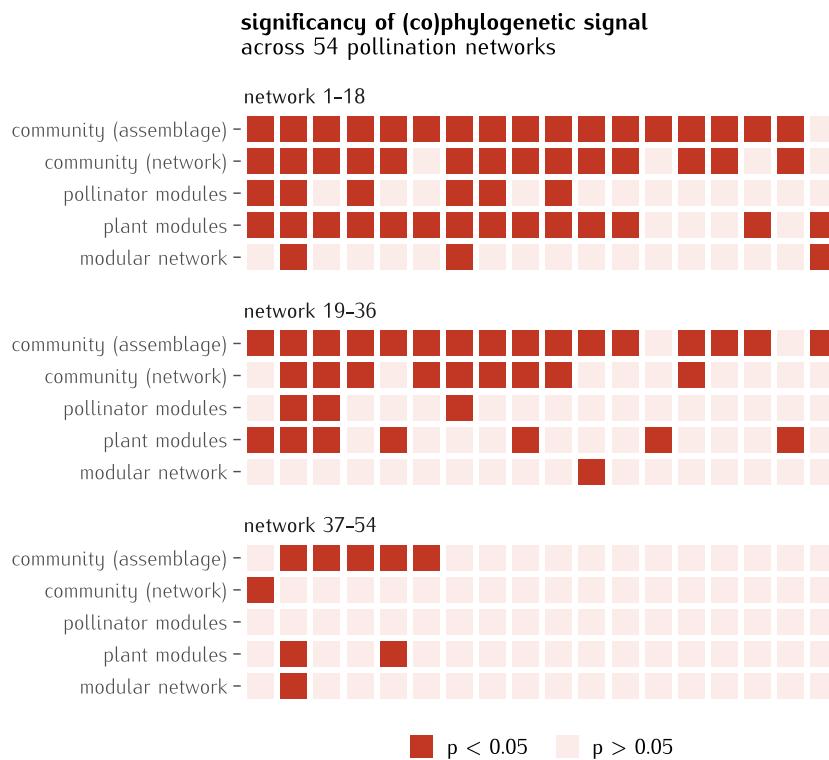


Figure A.5: The degree to which each of the 54 networks in our data set show cophylogenetic signal (CS) at the scale of local communities and their modular structure. On the y-axis, we show these measures of cophylogenetic signal starting with community level cophylogenetic signal of the observed species interaction network and community level cophylogenetic signal of the observed species assemblage. The measures that follow, the phylogenetic signal of both plants and pollinators in the observed modular structure of a network, and the degree to which a network's modular structure is cophylogenetic, pertain to the assessment of cophylogenetic signal in a community's modular structure. In each case, a light red square indicates that a network is no different than a null expectation with respect to the measure of interest while a darker red square indicates that the empirical network shows a significantly higher level of the measure than expected by chance. Networks were ordered by overall

ant cophylogenetic signal in only 9% of the empirical networks (Monte Carlo test, $p < 0.05$; Figure ??).

Across our dataset, networks tend to vary in the degree to which their modular structure shows cophylogenetic signal (Figure ??). Phylogenetic signal of modules for both pollinators and flowering plants is observed at a greater frequency than would be expected at random (χ^2 test, $p < 0.001$ in both cases); however, just two networks appear to satisfy all three constraints for cophylogenetic signal at the scale of a network's modular structure (Figure ??). As such, a significant result for one aspect of cophylogenetic signal at this scale does not make it more likely for other aspects to also support a hypothesis of cophylogenetic signal (χ^2 tests of both phylogenies conserved, $p < 0.001$; plant phylogeny conserved and cophylogenetic grouping of interactions, $p = 0.662$; pollinator phylogeny conserved and cophylogenetic grouping of interactions, $p = 0.417$; all three constraints, $p = 0.662$; Figure ??). When put together, our results provide rather limited evidence that the entire modular structure of a pollination network is the product of a cophylogenetic association.

A lack of signal at the scale of a network's modular structure does not imply that individual modules within those networks are also poorly characterized by a cophylogenetic signal. Indeed, most networks show significant cophylogenetic congruence in at least one module (Figure ??). Similarly, significantly more modules (88 out of 349; χ^2 test, $p < 0.05$) show a detectable cophylogenetic signal, when aggregating across networks, than would be expected at random (Figure ??). Our exploration of the characteristics of cophylogenetic modules with a logistic regression shows that module size and the proportion of participant species' interactions that are within the module are not related to module cophylogenetic signal ($z = 0.136$, $p = 0.892$ and $z = -0.920$, $p = 0.358$, respectively). However, the phylogenetic diversities of both flowering plants and pollinators do significantly influence module cophylogenetic signal with increased diversity making cophylogenetic signal less likely ($z = 2.225$, $p = 0.026$ and $z = 4.426$, $p < 0.001$, respectively). Furthermore, we see that the probability of species appearing in cophylogenetic modules is not equivalent for plants and pollinators. For the pol-

linators there is a significant correlation between cophylogenetic module participation and phylogeny (Mantel test; $p < 0.001$) indicating that cophylogenetic module participation is clustered on the phylogeny. Indeed, although the 1151 pollinator species that participate in cophylogenetic modules come from 8 orders, 946 of those species are either hymenopterans or dipterans. For the flowering plants, the opposite is true: participation in cophylogenetic modules is not influenced by phylogeny and hence is well distributed across the phylogeny (Mantel test; $p = 0.650$).

We also explored the degree to which modules show cophylogenetic signal when those modules are identified from quantitative rather than bipartite interaction networks. In all cases, we see qualitatively the same results between binary and quantitative networks. In terms of a network's modular structure, the partitioning of species into modules is at least slightly different between the two approaches (Supporting Information [Section S7](#)). However, for cophylogenetic signal at the scale of a network's modular structure, the results are nearly identical. Cophylogenetic signal at the scale of a network's modular structure was assessed with three separate analyses for each network (Methods). In the 45 analyses across these 15 networks, there are only five qualitative changes in the result of an analysis between binary and quantitative versions of a network (i.e., from significant to non-significant or vice versa; Supporting Information [Section S7](#)). At the scale of individual modules, we see similar results. The proportion of a network's modules that show significant cophylogenetic signal is often different between binary and quantitative networks but there does not appear to be much consistency to this difference (e.g. four networks show more cophylogenetic modules in the binary version than the quantitative, seven show the opposite pattern, and four are identical; Supporting Information [Section S7](#)).

Last, we saw that exotic species do not appear to have a marked effect on the cophylogenetic signal that we see in pollination communities or in their modular structure (Supporting Information [Section S10](#)). The only exception is that the presence of exotic species does appear to dampen cophylogenetic signal at the scale of individual modules (Supporting Information [Section](#)

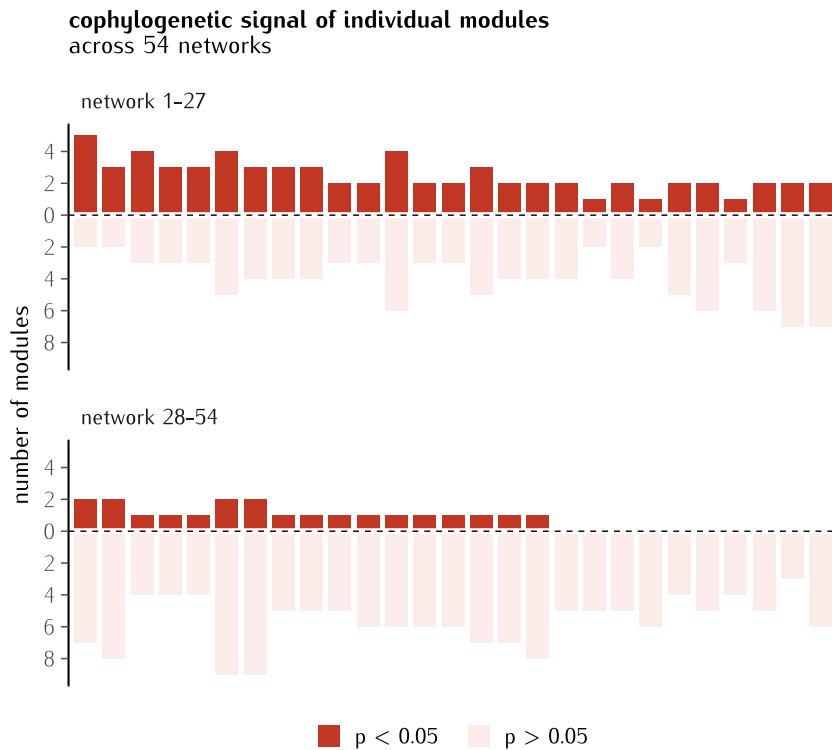


Figure A.6: There is substantial variation in the extent to which individual modules within networks show cophylogenetic signal. For each of the 54 networks (x-axis), we show the number of modules with significant cophylogenetic signal (darker red) and the number that are non-significant (lighter red). The majority of modules in almost all networks do not show a cophylogenetic signal (261 of 349). However, in 44 of 54 networks at least one module shows a stronger cophylogenetic signal than expected by chance.

S10). However, these analyses should likely be treated as exploratory since exotic species were only found in a small subset of the networks that we studied here.

A.4 DISCUSSION

The primary goal of this study was to examine the potential role of cophylogeny as an evolutionary determinant of ecological interactions. Specifically, we have set out to quantify the degree to which a cophylogenetic signal is manifest between flowering plants and their pollinators across a broad dataset and at a range of ecological scales. We found that cophylogenetic signal in pollination interactions appears quite commonplace, from the global level of interactions between flowering plants and pollinators to the scale of ecological communities and their internal structure. In particular, we find that local communities tend to exhibit a greater degree of cophylogenetic signal than both randomly assembled communities of plants and pollinators with the same network structure and observed pollination communities with a shuffled network structure. As such, it appears that the role that evolutionary history plays in determining pollination interactions is not just on one side of the interaction or the other, but can instead be the product of both taxa and the coupled evolutionary history they share.

In its most basic sense, a cophylogenetic pattern is a macro-evolutionary signature of coupled evolutionary divergence between interacting taxa. In some cases—for instance between toucans and chewing lice (**weckstein_biogeography_2004**) or amphibians and *Polystoma* (**bentz_historical_2006**)—the congruence in phylogenies of interacting clades has been attributed to shared biogeographical pressure such as vicariance (**weckstein_biogeography_2004**) or habitat acquisition (**bentz_historical_2006**). Given the purported role of biogeographical forces in producing phylogenetic signal (**cavender-bares_merging_2009**), the significant cophylogenetic signal that we see at the global scale may be explained by the co-occurrence of taxa that have diversified in the same regions

and interact due to proximity. On the other hand, if clumping of lineages by biogeographical filtering is a strong determinant of the signal that we see, we might expect to see a greater proportion of local communities showing significant signal when compared to random assemblages of species. Our results instead suggest that the proportion of local communities showing significant cophylogenetic signal is greater when the null expectation is a randomization of the observed community's interactions rather than when it is a random assemblage of species.

It has also been hypothesized that cophylogenetic signal is most parsimoniously explained by a coevolutionary process ([thompson_geographic_2008](#); [smith_distinguishing_2008](#); [godsoe_divergence_2009](#); [aizen_phylogenetic_2003](#)), even when reciprocal selection need not always result in cospeciation ([thompson_geographic_2005](#)). Indeed, [page_tangled_2003](#) suggests that “it is difficult to imagine that cospeciation can occur without at least some degree of coevolution”. Clearly, the process or processes that underpin cophylogenetic signal remain an open question. As such, we focus here on the various implications of cophylogenetic signal across a gradient of ecological scales rather than speculate about the underlying mechanism.

Pollination is a comparatively less intimate and a more variable interaction type ([bascompte_mutualistic_2014](#)) than other systems in which cophylogeny has been studied ([hafner_phylogenetic_1988](#); [weckstein_biogeography_2004](#); [desdevises_cophylogeny_2007](#); [hughes_multiple_2007](#)). Accordingly, the strength of cophylogenetic signal observed here is not as extreme as levels seen in other systems ([hafner_phylogenetic_1988](#)). Nonetheless, empirical associations of plants and pollinators still tend to show a significant cophylogenetic signal. Indeed, even when exotic species—who presumably have little coupled evolutionary history with native species—are present, a cophylogenetic signal can still be observed either due to potential evolutionary matching of deep phylogenetic branches between exotic and native taxa ([aizen_phylogenetic_2016](#)), or due to the qualitative nature of our assessment of cophylogenetic signal (i.e., significant vs. non-significant). Interestingly however, the dampening of cophylogenetic signal by non-native species was more obvious at the scale of individual modules of plants and pollinators.

Across ecological scales, we observe cophylogenetic signal between plants and pollinators. At all scales of pollination association, this implies that while interactions may be predicted by co-occurrence ([gotelli_species_2002](#)) or functional traits ([dehling_functional_2014](#)), it is also important to consider the evolutionary coupling of two taxa as a determinant of their likelihood to interact. Perhaps most importantly, when a community does show cophylogenetic signal it also implies that there is phylogenetic congruence on both sides of its pollination interactions rather than only one. Our results therefore extend previous conclusions that closely related species in ecological networks tend to interact in similar ways ([rezende_non-random_2007](#); [gomez_ecological_2010](#); [fontaine_comparing_2015](#)) and that phylogenetic relatedness influences community assembly ([emerson_phylogenetic_2008](#)) by suggesting that the role of past evolution in determining species interactions is not limited to one side of the interaction but can instead traverse the interaction.

The scale at which we see a less clear-cut cophylogenetic signal is at the level of modules. Undoubtedly, modules have fundamental roles in ecological networks as they describe groups of tightly-bound interaction partners ([olesen_modularity_2007](#)). However, a network's modular structure appears to poorly reflect a cophylogenetic signal. Perhaps this is unsurprising. We know that modules in ecological networks can be the product of a suite of processes ranging from ecological to evolutionary and back again ([olesen_modularity_2007](#); [rezende_compartments_2009](#); [krasnov_phylogenetic_2012](#); [rohr_structural_2014](#); [schleuning_ecological_2014](#)). Therefore, the weak cophylogenetic signal that we observe in the modular structure of pollination networks may be due to the fact that the modules of a community can be the result of a melting pot of ecological and evolutionary processes ([olesen_modularity_2007](#); [krasnov_phylogenetic_2012](#); [traveset_invaders_2013](#); [schleuning_ecological_2014](#)). Having said that, recent work that highlights the differences in the determinants and characterization of modules within a network ([olesen_modularity_2007](#); [rezende_compartments_2009](#)) suggests the need to focus less on modular structure and more on individual modules themselves. In particular, if individual modules can be thought of as distinct entities then examining co-phylogenetic signal across the modules of an entire community

may unnecessarily blur the patterns of cophylogenetic signal present in the modules themselves and the community as a whole.

Accordingly, we find substantial variation in cophylogenetic signal across individual modules. While a significant proportion of observed modules are cophylogenetic, they may be closer to the exception than the rule. These cophylogenetic modules—that typically show low phylogenetic diversity on both sides of the interaction—appear to be tightly interacting and closely-related groups of flowering plants and pollinators that exhibit both historical *and* contemporary associations. The constituent species in these modules appear to be more phylogenetically constrained for the case of pollinators than for plants as most pollinators that participate in cophylogenetic modules belong to the orders diptera and hymenoptera.

Given that the pollinators in pollination syndromes are thought to be more phylogenetically delimited than plants ([fenster_pollination_2004](#)), that hymenoptera and diptera account for 4 out of 11 pollination syndromes recognized by [ollerton_phenotype_2000](#) and that these syndromes are thought to be represented by modules ([olesen_modularity_2007](#)), we conclude that future work should aim to understand whether or not the cophylogenetic modules we observe are in fact the manifestation of distinct pollination syndromes. Such work could also incorporate the traits of these species alongside their evolutionary history and ecological associations. A clear pattern of cophylogenetic signal and trait-matching within modules may provide the clearest evidence to date in support of [olesen_modularity_2007](#) hypothesis that modules represent a fundamental unit of coevolution in pollination networks.

Our results contribute a new consideration to the prediction of ecological interactions. In the face of accelerating global change, the ability to understand why species interact in the way they do has become particularly imperative to ecologists ([tylianakis_global_2008](#)). The breakdown of ecological networks that occurs when species become locally extinct or upon the addition of species through introduction poses a similar challenge to

ecologists: how will the community respond? In an attempt to address these questions, much work has focused on predicting species interactions and network rewiring ([lopezaraiza-mikel_impact_2007](#); [memmott_global_2007](#); [tylianakis_global_2008](#); [kaiser-bunbury_robustness_2010](#); [aizen_phylogenetic_2016](#)). The cophylogenetic signal that we have seen in pollination interactions suggests that the coupled evolutionary history of taxa may play an important role in determining whether or not they can and will interact as the community around them changes. Given that pollination interactions represent a key ecosystem service, it is particularly important to improve our understanding of why and how particular species interact in the way that they do. We expect that our observation of cophylogenetic signal between plants and pollinators provides another valuable step in this process.

ACKNOWLEDGEMENTS

All authors contributed substantially to the study. M. C. Hutchinson and D. B. Stouffer conceived the study; M. C. Hutchinson and E. F. Cagua undertook analyses; M. C. Hutchinson, E. F. Cagua, and D. B. Stouffer drafted and edited the manuscript. The authors thank all those without whose data this work would be impossible. We also thank T. Poisot, J. M. Olesen, S. L. Nuismer, B. Bramon Mora, M. P. Gaiarsa, S. Saavedra, and four anonymous reviewers for invaluable comments on an early version of the manuscript. The authors acknowledge T. Poisot for his role in developing the software integral to this study. M. C. Hutchinson acknowledges the support of the University of Canterbury Summer Scholarship program. E. F. Cagua acknowledges the support of both a University of Canterbury Doctoral Scholarship and a University of Canterbury Meadow Mushrooms Postgraduate Scholarship. D. B. Stouffer acknowledges the support of a Rutherford Discovery Fellowship, administered by the Royal Society of New Zealand.

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of this article at <http://onlinelibrary.wiley.com/doi/10.1002/ecy.1955/suppinfo>

DATA ACCESSIBILITY

Data available from the Dryad Digital Repository: <https://doi.org/10.5061/dryad.8n3q3>

CONCLUSION

*It's not the numbers that are interesting.
It's what they tell us about the lives behind the numbers.*

— rosling_factfulness_2018

In this thesis, I explored how biotic and abiotic factors may affect the network of interactions between flowering plants and their pollinators in an ecological community, and how this network may influence the dynamics their constituent species. First, I explored how the structure can be influenced by the stresses that the environment exerts on plant or pollinator species. Second, I explored how network structure can be influenced by the trade-offs that manifest when plants share pollinators and compete for their pollination service. Third, I explored how network structure can be harnessed to design effective management interventions.

SUMMARY OF RESULTS

In ??, I found that the environment can play a role in determining the number of partners a species may have in an ecological community. On the one hand, it can limit the number of partners a species can interact with by determining the potential partners present in the community. This contribution, via community composition, was, in fact, the most substantial effect of the environment. However, a species might not necessarily interact with a potential partner, even if the partner is present in the community. On the other hand, we found that the stress the environmental conditions may impose on species can explain this difference between the number of potential partners and the actual number of partners. This is, the environment can directly influence the

level of specialisation of a species in its community. Specifically, I found that species with a relatively small number of interacting partners under optimal environmental conditions are more likely to interact with more species as environmental conditions degrade. This is, they act as facultative generalists that interact with a larger proportion of available species. Contrastingly, species with a relatively large number of interacting partners were more likely to interact with fewer species as environmental stress increases. The fact that this second group of species specialises by favouring interactions with a subset of possible species suggests that they might be focusing on fewer but better partners.

In ?? I found that the number of shared pollinators, which is generally correlated to the number of partners a species had a surprisingly low influence on the quality of pollination. Other factors like visit potential, functional originality, and abundance had a larger impact on both the quantity and purity of the pollination service. Specifically, high visitation and pollen production appear to have a positive effect on the quantity of pollination (the amount of conspecific pollen received), but a negative effect on the purity of pollination (the amount of conspecific pollen relative to heterospecific pollen). The relative abundance of a species had the opposite trend; a negative effect on pollination quantity and a positive effect on purity. Functional originality had a positive, albeit smaller effect on both pollination quantity and purity. Finally, the number of shared pollinators had a negative effect on both quantity and purity.

These results indicate that there is no free lunch when species compete for better pollination service. Some factors that may be beneficial for the quantity of pollen received may be detrimental to its purity. Those that are beneficial for both in the short term, like being specialised or having original traits, might be a risky strategy when biotic or abiotic conditions change. Another remarkable result was that species do not seem to play the same “strategies” across different communities. In other words, plants seem to balance the factors affecting the trade-offs inherent to the pollination service in different ways. For example, while a species may enjoy trait originality and therefore relatively good pollination service in a community, it might be less original in

another community but might be able to compensate by other strategies.

Finally, in ??, I found that species with a large number of interaction partners are also likely to have a high control capacity and, therefore, likely to be critical for controlling the ecological community. Interestingly, however, we found that being a generalist is not the determining factor for having a high control capacity. Indeed some species that interact with only a small number of species were also determined to be critical for network control. Our analysis indicated that the strongest predictor of control capacity was a species contribution to nestedness. Although controlling an ecological community is still not feasible with the current theoretical and practical limitations, identifying these species with high control capacity provides some lessons on the stability of the communities. Indeed, we found that critical species, which have the highest possible control capacity, were more likely to disproportionately contribute to the stable coexistence of species in their community.

GENERAL IMPLICATIONS

Having a large number of partners might be beneficial because it provides insurance against their loss in the long term (**walker_biodiversity_1992; yachi_biodiversity_1999**). In ??, however, we learnt that the number of pollinators a plant species shares has a negative, albeit small, effect on the pollination service. This small effect might indicate that the benefits of having a large number of partners in the long term might outweigh the disadvantages in the short term. The detriment to pollination service appears to be proportional to the number of partners, and hence there might be a point at which this trade-off becomes too large. This might explain, at least in part the intuition that pollination networks are relatively generalised compared to other ecological interactions, like parasitism, or even other mutualisms, like fruit-dispersal. But it also explains why pollination networks are not perfectly connected,

which would be the expected outcome if there were no trade-offs in the pollination mutualism (**bastolla_architecture_2009**).

From an evolutionary perspective, species can limit the number of interaction partners by developing traits that allow them to interact more effectively with certain species while rendering the interaction with certain others less likely (**caruso_competition_2000**). These traits, which together define the species' trophic niche, can be shaped by the environment, as we learnt from ???. Taken together, the results from ?? and ?? suggest that the point at which the trade-offs of sharing pollinators become too large might not only be species-dependent but might also be influenced by the environment; specifically, the stress it induces on a species. In other words, it is possible that the differences among the "strategies" with which a species approaches the competition for pollination across different communities (which we studied in ??) could be partially explained by environmental differences between communities.

These two chapters also provide some insights about the consequences of climate change for pollination communities. Specifically, climate change might result in pollination communities that are less robust, less diverse, and, paradoxically, less competitive. The robustness of pollination communities would be compromised because increased environmental stress would reduce the number of partners of generalists and increase the number of partners of "flexible" specialists. This "regression to the mean" would likely compromise the nested structure that has been shown to confer robustness to pollination networks (**burgos_why_2007; okuyama_network_2008**). The diversity of pollination communities would be compromised because less nestedness might also mean it is harder for species to coexist (**bastolla_architecture_2009**). Additionally, increased environmental stress might also increase the risk of extinction of specialists that are not flexible enough to adapt. Pollination networks might be less competitive because less nestedness and more connectance might imply higher levels of pollinator sharing. As found by **bastolla_architecture_2009** higher levels of pollinator sharing would favour species coexistence *if facilitation predominates*. This is consistent with the findings of

lopezaraiza-mikel_impact_2007 who found that facilitation is more predominant than competition in the harsh environment prevalent in high altitude pollination communities.

Nestedness offers an additional point of connection with ???. Recall that species with a high control capacity are critical to managing an ecological community because they are able to influence the abundance of other species (**cagua_keystoneness_2019**). The control capacity of a species and the manageability of the community as a whole are strongly dependent on the structure of the network. If climate change is able to modify the structure of pollination networks (for example by increasing the stress that species experience on average within a community) we could expect dramatic changes on species control capacity and network manageability. However, we need more research to determine the exact nature of these changes accurately.

FINAL WORDS

In this doctoral thesis, I harnessed the network of interactions between plant and pollinators as a tool to understand the ecological communities. The first two chapters provided insight into how disturbances might affect these communities, the services they provide, and the processes that govern them. While ecological networks have provided so far a useful framework to embrace the complexities of ecological communities, I believe we are still at the infancy of what could be achieved using networks to gain ecological insight. More data about the species interactions and their traits, and also more work that bridges complex thinking with applied problems (as I attempted in ??) will hopefully contribute to solving one of the main challenges of network tools, and ecology as a whole: applying the general results of theoretical and computational ecology to the needs of conservation management on the ground.

BIBLIOGRAPHY

Hypothesen sind Netze, nur der wird fangen, der auswirft

— novalis_novalis_1837

LIST OF FIGURES

Storytellers of all stripes must regularly compress all of the possible information their stories could contain into a manageable number of relatable details

— Micahel Austin

The sole aim of a methaphor is to call up a visual image

— George Orwell 1946

LIST OF TABLES

COLOPHON

This document was compiled using the R package `bookdown` ([xie_bookdown_2018](#)). It was typeset using the typographical look-and-feel `classicthesis` developed by André Miede and Ivo Pletikosić for L^AT_EX. All text and code can be found in:

<https://github.com/efcaguab/phd-thesis>

Feedback and comments can be directed to:

fernando@cagua.co

Final Version as of 21st June 2020 (Version 1.2.0).