Larger plants promote a greater diversity of symbiotic nitrogen-fixing soil bacteria associated with an Australian endemic legume

Russell Dinnage1 | Anna K. Simonsen2,3 | Luke G. Barrett4 | Marcel Cardillo1 | Nat Raisbeck-Brown5 | Peter H. Thrall4 | Suzanne M. Prober5

1Macroevolution & Macroecology Group, Research School of Biology, Australian National University, Canberra, ACT, Australia
2Commonwealth Scientific and Industrial Research Organization (CSIRO), Land & Water, Acton, ACT, Australia
3Division of Plant Sciences, Research School of Biology, Australian National University, Canberra, ACT, Australia
4Commonwealth Scientific and Industrial Research Organization (CSIRO), Agriculture and Food, Acton, ACT, Australia
5Commonwealth Scientific and Industrial Research Organization (CSIRO), Land & Water, Wembley, WA, Australia

Abstract

1. A major goal in microbial ecology is to understand the factors that structure bacterial communities across space and time. For microbes that are plant symbionts, community assembly processes can lead to either a positive or negative relationship between plant size or age and soil microbe diversity. Here, we evaluated the extent to which such relationships exist within a single legume species (Acacia acuminata) and their naturally occurring symbiotic nitrogen-fixing bacteria (rhizobia).

2. We quantified the diversity of rhizobia that associate with A. acuminata trees of variable size spanning a large environmental gradient in southwest Australia (72 trees in 24 sites spread across ~300,000 km²), using metabarcoding. We modelled rhizobia diversity using 16S exact genetic variants, in a binomial multivariate statistical framework that controlled for climate and local soil characteristics.

3. We identified two major phylogenetic clades of rhizobia that associate with A. acuminata. Soil sampled at the base of larger Acacia trees contained a higher richness of rhizobia genetic variants. Each major clade responds differently to environmental factors (climate and soil characteristics), but the positive association between tree size and rhizobia genetic diversity was mainly driven by responses from one of the two clades. Overall tree size explained more variation than any other factor, resulting in a ~3-fold increase in total richness and clade diversity from the smallest to the largest trees.

4. Synthesis. Previous studies have shown that plant host species is important in structuring microbial soil communities in the rhizosphere. Our results show that host size or age within a single plant species can also structure diversity of at least one group of soil microbes. A positive relationship between plant host size and rhizobia diversity suggests that hosts may modify the niche space of their surrounding soil (niche construction hypothesis) enabling a higher richness of microbial taxa. That different rhizobial groups responded differently to host size and other ecological factors suggests that rhizobia is not an ecologically uniform group, and that entirely neutral explanations for our results are unlikely. Host
1 | INTRODUCTION

Plants are highly dependent on mutualistic soil bacteria in the rhizosphere for growth and reproduction (Berendsen, Pieterse, & Bakker, 2012; Hayat, Ali, Amara, Khalid, & Ahmed, 2010). Among the key mutualistic microbes in the rhizosphere, symbiotic nitrogen-fixing bacteria (i.e., rhizobia) are critical to ecosystem function due to their intimate associations with legumes and their ability to directly convert atmospheric nitrogen into a plant-available nutrient (Kahindi et al., 1997; van der Heijden et al., 2006). Despite their functional importance, we are only beginning to understand the major ecological forces that structure mutualistic bacterial communities in the rhizosphere. Host plant size and age are important and highly variable traits in most natural plant communities (Hara, 1988; Menges, 2000). Through the simple act of growing, plants modify their surrounding environment (e.g., by increasing shade and leaf litter), which in turn can have important top-down effects on microbial rhizosphere diversity (Philipport, Raaijmakers, Lemanecau, & van der Putten, 2013) by altering numerous ecological and evolutionary processes (i.e., selection, drift, diversification, and dispersal) directly relevant to community assembly patterns (Nemergut et al., 2013; Vellend, 2010). Here, we evaluate the influence of plant size, within a keystone legume species, on the diversity and structure of symbiotic rhizobia communities.

Theory predicts that host plant size and age have important potential impacts on soil microbe diversity. Through selection (i.e., niche-based processes), differences in diversity may be generated because hosts may modify the rhizosphere as they grow and thus change the shape or size of the available microbial niche space in the soil (i.e., "niche construction") (Chaparro, Badri, & Vivanco, 2014; Marques et al., 2014; Meaden, Metcalf, & Koskella, 2016; Micallef, Channer, Shiaris, & Colón-Carmona, 2009; Odling-Smee, Erwin, Palkovacs, Feldman, & Laland, 2013; Wagner et al., 2016). In turn, the potential formation of feedbacks between microbial community assembly processes and pedogenetic soil formation (Verboom & Pate, 2013) could reinforce an increase in niche formation over time. Alternatively, larger hosts, by creating more habitable space compared to smaller hosts, may harbour greater diversity associated with greater immigration (i.e., bigger or smaller island in the context of Island Biogeography Theory or Neutral Theory) (Hubbell, 2001; MacArthur & Wilson, 2015). For example, larger hosts may harbour greater microbial diversity because their surrounding rhizosphere has higher absolute amounts of hospitable space for colonizing soil microbes. Host age, which is expected to be highly correlated with plant size, especially in obligate seeder plant species (e.g., Gosper, Prober, Yates, & Wiehl, 2013), may critically affect rhizosphere diversity due to the accumulation of lineages with slower colonization rates, leading to an increase in diversity (Nemergut et al., 2013).

It is also possible that niche-based processes could lead to a decrease in microbial diversity with increasing plant size. For example, chemical exudation by the host may selectively reduce the subset of microbial taxa that coexist over time as the plant ages (Hu et al., 2018). Alternatively, a decrease in diversity in time may occur due to ecological drift (leading to loss of diversity). In this instance, rare members of a microbial community have a higher likelihood of becoming locally extinct over time, provided that there is a sufficient initial diversity when seeds initially sprout, or other community assembly processes acting in the opposite direction that increase diversity with time.

Given that any of these community assembly processes, in isolation or in combination, may lead to either a positive or negative relationship between plant size and soil microbe diversity, the aim of this study is to evaluate the extent to which such relationships exist in natural mutualistic microbe populations. To gain further insight into the effects of plant size in the face of other important environmental drivers acting at either highly localized scales (i.e., edaphic conditions) or regional scales (i.e., climate), we measured fine-scale genetic diversity of rhizobia communities sampled over a complex heterogeneous landscape. The interaction between legumes and rhizobia provide a good biological system to evaluate relationships between plant size and mutualistic microbial diversity because symbiotic rhizobia also exist as free-living bacteria in the soil and are not entirely dependent on their plant host to reproduce or disperse to other habitats. This potentially allows more dynamic community assembly processes as plants increase in age or size. Furthermore, the functional significance of rhizobia towards legume growth and reproduction is firmly established.

In this study, we examined the effect of host plant size on the diversity of rhizobia that originate from the surrounding soil of natural Acacia acuminata populations, a foundational species of the nationally endangered York gum–jam woodlands of the Western Australian wheatbelt (Prober, Standish, & Wiehl, 2011). It has been well established that host age and size are highly positively correlated in both Acacia (Maslin & McDonald, 2004) and Eucalyptus (Gosper et al., 2013). Using 16S rRNA metabarcoding data of rhizobia isolated from bulk soils in vicinity of individual A. acuminata trees, we investigated the role of host size in influencing the genetic diversity of symbiotic rhizobia. We sampled rhizobia from a standardized volume
of soil collected from the base of *A. acuminata* host plants across the maximum climate gradient of the species range, capturing rhizobia diversity and variable *Acacia* tree size across a broad range of environments.

2 | MATERIALS AND METHODS

2.1 | Site selection and soil sampling

*A. acuminata* is an endemic tall legume shrub or small tree restricted to southwestern Australia. It is typically an obligate seeder (killed by hot fire) with limited resprouting capacity. We sampled rhizobia from 24 sites where *A. acuminata* occur (Figure S2.1 in Supporting Information 2). To select sample sites, we obtained *A. acuminata* occurrence records from the Atlas of Living Australia (ALA, www.alaa.org.au) together with associated mean annual temperatures and mean annual rainfall. Because of the general importance of temperature and rainfall in the distribution of most organisms, we chose sites to optimize three criteria: (a) maximize temperature and rainfall gradient; (b) minimize spatial autocorrelation in rainfall and temperature; and (c) minimize correlation between temperature and rainfall values. Based on the available occurrence records in ALA, we used a brute-force optimization algorithm to generate one million random combinations of potential sites, and chose the combination that best balanced the above criteria (chosen manually based on plots of the criteria). Our final combination of sites ranged in mean annual temperature from 15.4ºC to 20.6ºC, and mean annual rainfall from 287 mm to 612 mm. Temperature and rainfall were not correlated in our final site list (*r* = −0.156, *p* = 0.47).

We sampled soil during the winter season (July) when the soil likely contains higher numbers of viable rhizobia due to higher seasonal rainfall in the region and higher levels of nitrogen fixation between the host and its rhizobial partner (Monk, Pate, & Lonergan, 1981). At each site, we selected three *A. acuminata* trees at least 50 m from the road and at least 50 m apart (where possible). We sampled three (5 cm diameter 10 cm depth) soil cores within 50 cm of the base of each tree to obtain one pooled soil mixture for every sampled tree. We measured the circumference of each sampled tree and converted the measurement to diameter. We measured circumference at the base of each tree to accommodate comparison among different forms of *A. acuminata*, since this species can also occur in more shrub-like forms. While some sampling was from these more shrub-like forms, we refer to each sampled plant as a "tree" throughout this paper. Soil was stored at 4ºC until further processing in the lab.

2.2 | Rhizobia isolation and DNA extraction

We isolated symbiotic rhizobia from field soil by harvesting legume nodules of *Acacia acuminata* grown in controlled growth chamber conditions inoculated with each field soil sample, where a field soil sample comprised a mixture of three soil cores collected from around the base of one *Acacia* tree. Wild-collected *Acacia acuminata* seeds ("typical form," sourced from native seed collectors Nindethana Seed; www.nindethana.net.au/) were surface sterilized and pregerminated on 1% agar plates for 1 week at room temperature (25ºC) in the dark. Six randomly selected *Acacia* seedling replicates were grown in each field soil sample, in a randomized block design in a growth chamber. For each pot, the field soil was sandwiched between two layers of autoclaved vermiculite and planted with *A. acuminata* seeds. Plants were watered weekly with 10 ml of one-fourth strength McKnight solution and autoclaved water. Plants were grown for 4 weeks, allowing ~2 weeks for nodule formation. To increase the likelihood of targeting symbiotic nitrogen-fixing rhizobia for sequencing, we obtained isolates from pink nodules exhibiting characteristic hallmarks of nitrogen-fixing activity. We sampled five nodules from each *Acacia* seedling, and restricted nodule sampling to below the first layer of autoclaved vermiculite to avoid any potential contaminants. Control pots containing only sterile vermiculite were randomized throughout the growth chamber (n = 10), which had few or negligible nodules at harvest. Surface-sterilized nodules (using commercial bleach) were crushed with sterile forceps and tissue lysate was cultured on Yeast-Mannitol agar. Rhizobia cultures were grown for 7–15 days at 30ºC in the dark. Isolates were replated twice on YMA agar media to obtain a single rhizobia isolate.

The nodule trapping approach done here, while effective at enriching for symbiotic bacteria from legume nodules, is likely to be sampling the most abundant rhizobia strains in the soil. Therefore, any observed differences in richness of genetic diversity among samples is likely to be conservative. In total, our sampled rhizobia collection consisted of 1,834 isolates, where ~25 rhizobia isolates were sampled from each *Acacia* tree (i.e., field soil sample), and ~75 rhizobia isolates were sampled from each site, representing an approximately equal sampling effort of rhizobia isolates for every *Acacia* tree and site.

2.3 | Sequencing

Our unit of metabarcoding sample sequencing in this study was at the *Acacia* tree level (n = 72 trees, 3 trees/site, 24 sites in total). We prepared pooled rhizobia DNA samples at the tree level by suspending freshly grown cells using a 1-l inoculation loop into 500 µl of sterile autoclaved distilled water, pooling ~25 isolates per sequencing sample (i.e., *Acacia* tree). Pooled bacteria isolates were sent to Australian Genome Research Facility (https://www.agrf.org.au/resources/applications/-next-gen-sequencing#Diversity) for DNA extraction and sequencing. Genomic DNA was extracted using MoBio kits (UltraClean Microbial Isolation kit), following the standard kit protocol. Amplicons were generated from the V1–V3 hypervariable region (primer set 27F: AGAGTTTGATCMTGGCTCAG, 519R: GWATTACCGCGGCKGCTG), and were prepared for sequencing on the Miseq platform, following Illumina’s 16S sample preparation guide using paired-end sequencing chemistry and Illumina’s own XT Nextera indices, producing 300-bp paired-ends reads.
2.4 Raw sequence data processing

In order to capture fine-scale diversity, the goal was to generate high-quality exact sequence variants of rhizobia (i.e., 100% identity 16S sequences), which is currently the recommended OTU resolution for microbial community analyses when using hypervariable sequence segments of 16S rRNA (Edgar, 2018). To do this, we ran the reads through standard quality control steps, called exact variants, and then filtered variants to remove any low abundance (less than 0.5% of a sample), as well as any that were not recognized rhizobia genera, which were likely contaminants (see Supporting Information 1, Part A for details).

2.5 Environmental variables

To examine relationships with our community data, we focused on three major classes of ecological and environmental factors: (a) climate; (b) physical and chemical soil characteristics; and (c) host characteristics measured at the site level (CSBP; www.csbp-fertilisers.com.au). Host size was measured as the mean annual precipitation and temperature were obtained from the ALA (www.ala.org.au), and are based on Bioclim (Fick & Hijmans, 2017). For each soil sample, we measured 22 soil chemistry factors (analyses were carried out by CSBP; www.csbp-fertilisers.com.au). Host size was measured as the tree trunk diameter at the base of each tree. In addition to host size, we estimated three additional host characteristics measured at the site level (Acacia acuminata density, Acacia species richness, Acacia phylogenetic diversity). We calculated site-level Acacia acuminata density using all occurrence record data for Acacia acuminata in South West Australia found in the ALA. We estimated Acacia density from point occurrence records using two-dimensional Gaussian kernel estimation, implemented in the kde2d function of the MASS package in R (Venables & Ripley, 2002). Density was estimated within a grid with a rough resolution of 0.37° latitude–longitude within each cell, which is capable of capturing among site variation. Values for each site were then assigned as the density of the cell in which they fell. Because rhizobia diversity could be affected by the presence of related legumes in the area, we obtained site-level Acacia species richness (0.37° resolution), and Acacia phylogenetic diversity (0.5° resolution) estimates from point occurrence data in ALA (Miller, Murphy, Brown, Richardson, & González-Orozco, 2011). All factors, with the exception of climate data and the Acacia community measures, were measured at the soil sample level (i.e., Acacia tree level).

2.6 Data analyses

The goal of our analyses was twofold: firstly, to determine what environmental factors are associated with fine-scale rhizobia genetic diversity, as measured by the probability of observing an exact 16S sequence variant (genetic variant richness) and secondly, to determine whether these responses to environmental factors depend on the major phylogenetic clades we observed. We first identified the major clades in our rhizobia sequence by inspecting the phylogeny. We used mothur to align our 16S sequence variants to the SILVA 16S reference database (Quast et al., 2013; Schloss et al., 2009). We then constructed a Bayesian ultrametric phylogeny using MrBayes (Ronquist & Huelsenbeck, 2003). We identified four rhizobia families and five genera in our samples using an RDP classifier implemented in the r package DADA2 (Callahan et al., 2016). The phylogeny revealed these taxonomic groups fell into three distinct well-supported clades with high posterior probability (Figure 1), which we called clade Br (containing Bradyrhizobium from the family Bradyrhizobiaceae), clade MeShRh (containing the genera Rhizobium and Shinella from the family Rhizobiaceae, and Mesorhizobium from the family Phyllobacteriaceae), and clade Bu (containing the genera Burkholderia and Paraburkholderia in the family Burkholderiaceae). Bradyrhizobium had the greatest number of genetic variants (62), followed by Rhizobium (22), then Shinella (5), Mesorhizobium (4), and Burkholderia-Paraburkholderia (4). With the exception of clade Bu, at least one genetic variant from each major clade occurred in all soil samples (i.e., at every sampled Acacia tree across all 24 sites), while clade Bu was only detected in two soil samples across two sites. Due to the limited diversity and distribution of clade Bu variants, we excluded this clade from our presented analyses (but inclusion or exclusion of this clade in any subsequent analysis did not alter any inferences for clade Br and clade MeShRh).

For soil (chemical and physical) characteristics, we drew out two variables of particular interest (soil salinity and soil pH) because previous work has shown that rhizobial taxa respond differently to these factors (Sprent, Ardley, & James, 2017). We collapsed the remaining set of 20 soil variables into a smaller number using principal components analysis. For our downstream analyses, we used the first three PC axes, which cumulatively explained ~66% of variation. In total, we analysed 12 predictors in our full model (mean annual temperature, mean annual rainfall, soil salinity, soil pH, soil PCs 1–3, tree diameter, latitude, Acacia acuminata density, Acacia species richness, and Acacia phylogenetic diversity).

2.7 Statistical model

To model the 16S sequence variant data, we used a multivariate model-based approach. The object of the analysis was to model the distribution of exact 16S sequence variants with respect to a set of 12 environmental factors and the major clade they belonged to. We used a GLMM with binomial errors to model the presence or absence of exact 16S sequence variants in individual tree samples. Although our data also included a measure of the abundance of each genetic variant in each sample, we did not use abundance data in our main model because it is not comparable between samples due to uncontrolled sources of variation in the total number of reads measured per identified genetic group. Specifically, preliminary DNA extraction trials on a subset of individual isolates showed large differences in total DNA yield, despite approximately equal cell inputs in the extraction protocol. Therefore, we considered only the presence or absence of a genetic variant in a sample to be reasonably reliable data. Given this, our model estimates the effect of environment and clade on the probability of a genetic variant in an Acacia tree-level sample. An increased
probability of variants in a tree-level sample is indicative of higher genetic variant richness, and so this is a model of genetic variant richness within the surrounding soil of an Acacia tree. We included random effects for the tree, the site, and the genetic variant in all of our models to fully account for any nonindependence due to these hierarchical factors. Fixed effects include the effect of the clade to which a variant belonged, and the clade by environment interactions, encoding the effects of each environmental factor for each clade

FIGURE 1 Bayesian phylogeny of all exact 16S sequence variants detected in this study, with major clades and genera labelled. Nodes are labelled with the posterior clade credibility.
independently. We fit the model using a Bayesian approach implemented in the \texttt{R} package \texttt{INLA} (Martins, Simpson, Lindgren, & Rue, 2013; Rue, Martino, & Chopin, 2009; see Supporting Information 1, Part B for full details of the statistical model).

Because we did not know a priori which of our variables would be most important in explaining rhizobia richness, we used a model selection and model averaging approach in our analysis. Specifically, we ran models for all possible combinations of our 12 environmental variables (for a total of 4,096 models). We then used pseudo-Bayesian model averaging to estimate a posterior distribution for each of the model's parameters, which accounts for uncertainty both in the model selection process, and in the individual model's estimation (Yao, Vehlari, Simpson, & Gelman, 2017). We then used hierarchical partitioning to measure variable importance, and chose the top five best variables for further statistical analysis (Supporting Information, Part C). Specifically, we refit our main model using only the top five variables, and assessed model fit and adequacy, then ran three auxiliary models to explore several alternative model assumptions. These three auxiliary models examined: (a) the effect of our choice of rhizobia groupings; (b) the effect of including genetic variant abundance data in the model; and (c) the effect of spatial structure on our results (see Supporting Information 1, Part G for full methods for these models).

2.8 | Separating within- and between-site effects of environment

Our analysis was explicitly hierarchical, in that the data were measured within individual trees, which were found at different sites. Although we used hierarchical random effects in a mixed model to account for the nonindependence induced by this sampling strategy, it is still interesting to ask how much of the effects we see can be attributed to within-site differences among trees, compared to between-site differences. This information is lost in a simple linear mixed model, as it can only test for an overall effect. To examine within-site and between-site effects of the environment, we used contextual analysis (Snijders & Bosker, 2011). In contextual analysis, the effects of a variable on the response is modelled as a regression of both the individual measured variable, and the mean of the variable for the group that the individual data point belongs to. Here, we modelled the presence or absence of a clade in a tree sample within a site as a function of the environment measured at the individual tree, as well as the mean environment for the site, for the top five variables according to their importance in our main model (see Supporting Information, Part E for more details).

3 | RESULTS

Out of the total input of 1,834 cultured isolates across all samples, we obtained 97 exact 16S sequence variants classified as rhizobia after filtering sequence data for known symbiotic rhizobia genera (Shamseldin, Abdelkhaled, & Sadowsky, 2017), which have all previously been found in association with Australian Acacias (Hoque, Broadhurst, & Thrall, 2011; Lafay & Burdon, 1998). These could be placed in two main phylogenetic clades (Figure 1), which we called clade Br (containing \textit{Bradyrhizobium} from the family \textit{Bradyrhizobiaceae}), and clade MeShRh (containing the genera \textit{Rhizobium} and \textit{Shinella} from the family \textit{Rhizobiaceae}, and \textit{Mesorhizobium} from the family \textit{Phyllobacteriae}), whose distinct distributions across environmental factors we modelled using generalized linear mixed models.

The top five variables for predicting each clade's genetic variant richness, according to our hierarchical partitioning importance score, included the following environmental variables: tree host diameter, soil salinity, mean annual rainfall, soil pH, and soil PC3. Importance score dropped off after these top five, which is reflected in the fact that standardized coefficients were low and their model averaged credible intervals mostly overlapped zero for the remaining factors (Figure 2).

We found that the model refit to the top five variables explained the genetic variant data fairly well when we examined its residuals and its predicted values. We used the Dunn–Smyth residuals, which are designed to deal with non-Gaussian integer responses (Dunn & Smyth, 1996; Warton, Stoklosa, Guillera-Arroita, MacKenzie, & Welsh, 2017). There was no evidence of violation of the assumptions of the GLMM (Figure S2.2 in Supporting Information 2), with a good fit to normality of errors, and no evidence of nonhomogeneity of variance. Predicted genetic variant richness from the models, when summarized at individual tree samples, showed a strong relationship with the observed genetic variant richness (Pseudo-$R^2 = 0.87$; Figure S2.3 in Supporting Information 2).

3.1 | Host tree diameter predicts rhizobia diversity within and among sites

We found that total genetic variant richness was higher from soil sampled at the base of larger \textit{A. acuminata} trees (Figure 3). Furthermore, the genetic variant richness within the MeShRh clade increased as tree size increased (Figures 2 and 3), but clade Br genetic variant richness did not increase. There was also an increase in clade diversity (Gini–Simpson index) with host tree size (Figure 3). These analyses generally indicate that diversity among and within clades was higher in larger trees, where changes in clade diversity patterns were being driven entirely by the increase in genetic variants within clade MeShRh (leading to higher evenness among the clades).

We further analysed the possibility that clade MeShRh was more likely to occur at higher richness in larger tree hosts within or between sites. In other words, was there evidence that higher genetic variant richness of clade MeShRh observed in larger trees was because some sites generally had larger \textit{Acacia} trees, or was there evidence that even within sites of variable tree size, a higher genetic variant richness of rhizobia could still be observed in larger trees? As with the previous analyses, our results consistently show strong genetic variant richness responses in clade MeShRh at the site level (Figure 4). However, we also found a greater genetic variant richness of clade MeShRh (Figure 4) for larger trees that occurred within a given site (Figure 4). In total, our analyses on the effects of tree
FIGURE 2  Main model results. Model is a GLMM with binomial errors, modelling the probability of genetic variants within samples taken from individual Acacia trees, fit using an Integrated Nested Laplace Approximation (INLA) to the Bayesian posterior distribution. Points represent model averaged posterior means of model parameters, and error bars represent the model averaged 95% credible interval. The posterior is derived by combining 4096 different models representing all possible combinations of 12 environmental predictors, using Bayesian model averaging (see methods for details). All predictors were standardised prior to modelling, so plotted values are standardised coefficients. Parameters whose credible intervals do not overlap zero are plotted with larger points and higher opacity. The top panel plots hyper-parameters of the random effects in the model, estimating the standard deviation of the Tree, Site, and Genetic Variant random effects. The middle panel plot the fixed intercept terms, and the bottom panel plot the fixed clade by environment terms (with each clade’s terms plotted in different colours). Note the different scale of the x axis in each panel. At the right of the environmental terms is plotted an estimate of the importance of each environmental predictor in the model, based on hierarchical partitioning.
size on rhizobia diversity show with confidence that clade MeShRh is much more likely to be present in larger trees, and this effect held even within sites when other highly localized factors (soil characters) are also included in the analyses.

3.2 | Rainfall predicts rhizobia clade diversity

We generally found strong patterns of rhizobia diversity associated with rainfall, with clade Br and clade MeShRh showing differential responses (Figures 2 and 5). When we examined the richness patterns of genetic variants within each clade, we found that clade MeShRh had a greater richness of genetic variants at sites with low rainfall. In contrast, our fitted model shows a greater richness of unique clade Br genetic variants at high rainfall. Both clade’s effects were marginal with credible interval slightly overlapping zero, with the negative effect of clade MeShRh being stronger—95.5% credible interval did not overlap zero—however, since the effects act in opposite directions for each clade, rainfall is an overall good predictor of rhizobia clade composition (hence the high importance score). When examining the net effect of all genetic variants, we found higher total genetic variant richness at wetter sites (Figure 5). In total, our rainfall results are similar to tree size in that the probability of a clade MeShRh genetic variant occurring in a sample appears to be limited to more restricted conditions (i.e., drier sites); while clade Br occurs along the entire rainfall gradient, showing higher levels of richness at wetter sites, the net effect translates into higher total genetic variant richness at wetter sites. This pattern also led to clade diversity being higher at drier sites (Figure 5), since the genetic variant richness of both clades diverge in opposite directions with increasing rainfall, leading to lower Gini-Simpson diversity at wetter sites.

3.3 | Soil characteristics predict rhizobia diversity within and among sites

We found strong patterns of diversity associated with soil characteristics. Similar to host tree diameter responses, genetic variant richness tended to respond only within clade MeShRh, whereas clade Br did not respond, leading to a predicted change in clade-level diversity along several soil variables. Three soil variables were in the top five most important explanatory variables, soil salinity, soil pH, and soil PC3 (Figure 2). Both soil salinity and soil pH were positively related to the genetic variant richness of clade MeShRh, but clade Br did not respond to either (Figures 2 and 5).

Again, clade MeShRh showed a positive genetic variant richness response to soil PC3, but clade Br showed no positive nor negative response (Figures 2 and 5). Soil PC3 was associated with soil with high concentrations of aluminium, iron, and nitrate (Figure S2.4 in Supporting Information 2).
FIGURE 4  Contextual analysis model results. Model is a GLMM with binomial errors, modelling the probability of genetic variants within samples taken from individual Acacia trees, fit using an Integrated Nested Laplace Approximation (INLA) to the Bayesian posterior distribution. Contextual analysis models environmental factors as between and within site differences. Between site predictors are site-level means, within site predictors are tree-level deviations from the site-level means. Points represent posterior means of model parameters, and error bars represent the 95% credible interval. Only the top 5 most important environmental factors from the main model (see Figure 2) were included. Parameters whose credible intervals do not overlap zero are plotted with larger points and higher opacity. The top panel plots hyper-parameters of the random effects in the model, estimating the standard deviation of the Tree, Site, and Genetic Variant random effects. The middle panel plot the fixed intercept terms, and the bottom panel plot the fixed clade by environment terms (with each clade's terms plotted in different colours, and between and within site terms plotted with different shapes). Note the different scale of the x axis in each panel. Note that mean annual rainfall, as a climatic factor, does not vary within sites, and so only has a between site term.
The three auxiliary models we fit showed that our results are robust to a number of assumptions we made in our model. Our first auxiliary model used four genus-level clade groupings instead of the two used in the main model, and showed this model did not fit better than our main model (ΔWAIC = 10.02), and that genera within our clade MeShRh all trended in the same direction as the clade as a whole in their response to the environment (Figure S1.1 in Supporting Information 1, Part G).

Our second auxiliary model fit a zero-inflated Poisson model to our data, which included estimates of the abundance of each
genetic variant found in each sample (Figure S1.2 in Supporting Information 1). The model results were completely consistent with our main model, with the abundance part of the model being consistent with and reinforcing the binomial (presence/absence) part of the model.

Our third auxiliary model accounted for the possibility of spatial nonindependence affecting our results by including a spatial random effects. The host size effect captured in our main model was unchanged by including space, and all other environmental variables at least trended in the same direction as our main model (see Supporting Information 1, Part G for full results of all auxiliary models).

4 | DISCUSSION

In this study, we evaluated the effect of plant size on the diversity of rhizobial symbiont communities across variable abiotic climate and soil conditions within a single dominant legume species. In summary, we found that soil at the base of larger *Acacia* trees hosted a higher diversity of rhizobia. Specifically, we found that larger trees had a higher Gini-Simpson diversity index at the phylogenetic clade level as well as higher genetic variant richness in total and within at least one of the two clades. Of the two major phylogenetic clades we identified in this study (Figure 1), clade Br was equally likely to be present across all host plant sizes. Therefore, the increase in the clade diversity was primarily driven by the genetic variant richness of clade MeShRh, which was much more likely to occur at the base of larger *Acacia* trees. Furthermore, when we examined the relationship between tree size and rhizobial genetic variant richness within each clade, we found a higher richness of genetic variants within clade MeShRh from soil found at larger trees, indicating that tree size also affected cryptic genetic variation. While accounting for other abiotic factors that also potentially covaried with host plant size, we found that climate (mean annual rainfall) and edaphic conditions were strongly associated with genetic variant richness shifts and that clade MeShRh in particular was highly responsive to changes in environmental conditions. This suggests that clade MeShRh generally has a much narrower environmental range compared to clade Br.

Larger *Acacia* trees (on average) tend to be found at wet sites (Figure S2.5 in Supporting Information 2), but even after incorporating rainfall into our analyses, tree size was still a strong determinant of rhizobia communities. In fact, our analyses showed that rainfall was negatively associated with clade MeShRh genetic variant richness, such that the direct effect of rainfall and its indirect effect through its positive association with host size seem to be driving rhizobia diversity in opposing directions. We also showed that the tree size effect was as strong at the local within-site level as it was at the regional between-site level using contextual analysis (Figure 4), making it less likely that the effect is driven by a large-scale unmeasured confounding factor, and consistent with a mechanism driven by local-scale processes. We are not aware of many previous uses of this statistical method in ecology (but see Bradford et al., 2017), but a related method has been used in evolutionary biology—contextual selection analysis—designed to distinguish group selection from individual-level selection coefficients (Goodnight, Schwartz, & Stevens, 1992). However, the method is well established in the social sciences (Bell, Jones, & Fairbrother, 2017; for example, Davis, Spaeth, & Huson, 1961; Snijders & Bosker, 2011), and would appear to have great potential in ecology, given that it allows modelling and testing of different effects of predictors at different scales. Rhizobial diversity patterns also responded strongly to local soil chemistry, in addition to larger scale factors acting at the regional site level (i.e., mean annual rainfall) (Figure 4).

The finding that larger legumes have higher symbiotic microbe diversity in the surrounding soil has several implications for the community assembly processes of mutualistic microbes, which we discuss below.

Our results are consistent with the hypothesis that as trees grow in size or age, they modify the habitat of their surrounding soil, generating niches that may support higher rhizobial diversity. This may therefore constitute an empirical example of niche construction, where an organism modifies its existing habitat and subsequently imposes selective forces on other interacting organisms (e.g., bacteria in the rhizosphere) at the community level (Odling-Smee et al., 2013). Modification of a plant’s own rhizosphere is by definition a highly local process and previous studies have shown that rhizobial population dynamics are also highly localized in that rhizobial population density is much higher in the immediate vicinity of legume roots, while soil meters away contains significantly fewer rhizobia (Parker, Malek, & Parker, 2006). Consistently, we find that the influence of tree size on rhizobial community assembly is also highly localized, since our results indicate that rhizobia (and clade MeShRh in particular), are strongly responsive to withinsite variation in tree size. Previous pilot studies in the woodland systems evaluated here suggest that soil sampled under *Acacia* species likely have much higher rhizobia populations or symbiotic potential with rhizobia (as indicated by higher nodule numbers in soil inoculation experiments; see Supporting information 3). On the other hand, we cannot completely rule out the possibility that the causation is reversed—that *Acacia* grow more quickly in soils with high rhizobia diversity, and this is driving the association we find. We find this less likely as it would require that rhizobia diversity is a persistent and long-term characteristic of the soil, even in the absence of any hosts. Additionally, the only experimental quantification of the effects of rhizobia diversity on host growth that exists showed that rhizobia diversity appeared to reduce host growth rate rather than increase it (Barrett, Bever, Bissett, & Thrall, 2015; Simonsen, Chow, & Stinchcombe, 2014).

How *Acacia acuminata* or other plants modify their rhizosphere requires further investigation. However, our data indicate a shift in soil chemical composition as a potential modification pathway,
since we found that soil PC2 was correlated with Acacia tree size (Figure S2.5 in Supporting Information 2), and that one of the unique characteristics of PC2 is a high loading of ammonia (NH₃) and iron (Figure S2.4 in Supporting Information 2). In a previous pilot study, we found that soil sampled at the base of Acacia acuminata trees exhibited unique chemical profiles (i.e., higher ammonia, higher iron, and lower pH) and higher nodule numbers (indicative of higher rhizobia populations or higher symbiosis potential with rhizobia) compared to nearby unvegetated bare ground and Eucalyptus phloxeleoa soil sampled within the same sites (see Supporting Information 3). Similarly, Prober et al. (2011) also found high levels of ammonia and iron at the base of Acacia acuminata sampled from soil in the same woodland habitats (i.e., York gum-jam woodlands). Given that iron has been identified as a critical and limiting element for nitrogen fixation (Brear, Day, & Smith, 2013), and that ammonia is a known product of biological nitrogen fixation reaction, our results suggest that these chemical modifications are specific to legume nitrogen-fixing activity. However, our study also demonstrates that other unmeasured changes associated with increasing tree size are likely to be taking place, given that tree size is still strongly explanatory even after controlling for soil conditions in our analyses. For example, large trees may have higher root mass or density at the base, which may be playing a critical role in altering the physical structure of soil in the rhizosphere.

Apart from niche-driven differences, our results could also be explained by differences in dispersal capacity or competitive ability among rhizobial clades or taxa (Bissel, Richardson, Baker, Wakelin, & Thrall, 2010). Given the wide distribution of clade Br, it is possible that clade Br has a much higher dispersal, establishment, or competition capability and thus may be more important for new Acacia seedling recruitment, which is also consistent with the observation that Bradyrhizobium is more frequently found across Australia compared to Rhizobium (Lafay & Burdon, 1998). If clade MeShRh has a slower dispersal or recruitment rate, it may only begin to become common in older plants, which have had more time to accumulate rhizobia lineages. If clade MeShRh does have a slower dispersal rate, it is possible we could find evidence for this if there is greater geographic structure (also known as distance decay) in genetic variant composition within clade MeShRh as compared with clade Br. However, a multiple regression on dissimilarity matrices using a permutation test (Lichstein, 2007) found no overall relationship between geographic distance and genetic variant turnover (as measured by Simpson’s dissimilarity; coefficient = 8.98x10⁻⁶, p = 0.128), nor did it indicate a significant difference between the two clades in their rate of genetic variant distance decay (coefficient = -1.33x10⁻⁴, p = 0.620). However, there was a very high turnover in general in genetic variants, which could have affected our ability to detect differences (most of the turnover values were at the maximum value of one: no overlap in genetic variant composition).

The explanations above suggest that rhizobial taxa exhibit different characteristics or particular habitat preferences. In contrast, species equivalency is one of the key assumptions of metacommunity dynamics in the context of neutral theory (Hubbell, 2001), and is of increasing interest in microbial ecology (Dumbrell, Nelson, Helgason, Dytham, & Fitter, 2010). Generally, we find the assumption of equivalency among our identified rhizobial clades to be unrealistic. More specifically, under the assumption of neutrality, and its prediction of high diversity with larger area, we would expect a relationship with total genetic variants and host size, but no relationship between clade diversity and host size, since differential clade level responses imply nonneutral, higher level phylogenetic organization driving community assembly patterns. However, we found both genetic variant and clade-level diversity associated with host size and strong clade-specific responses to different abiotic conditions (climate and soil characteristics), suggesting that clades are not equivalent (although equivalency may hold within clades). More specifically, clade MeShRh appears to occur over a much narrower range of edaphic and climate conditions. Previous studies also find that differences in abiotic growth conditions is variable among rhizobial taxa and strains (Han et al., 2009; Li et al., 2011; Thrall, Bever, & Slattery, 2008; Vuong, Thrall, & Barrett, 2017), which together imply that species equivalency is unlikely in natural rhizobia populations. Together, these results build a stronger case for niche-driven differences that will benefit from further experimental work.

Few studies, thus far, have sought to explicitly examine the impact of plant size on microbial diversity among natural plant populations (although see Krah et al., 2018; Meaden et al., 2016; Patiño, Gómez-Rodríguez, Pupo-Correia, Sequeira, & Vanderpoorten, 2018), despite the recognition that variation in plant size (reflected by either the productivity or age of the habitat) is a prominent and highly dynamic feature of plant communities (i.e., natural forest stands are always a mixture of different age stages). Previous manipulative transplant work has shown that soil microbe communities change during the course of plant development (Philippot et al., 2013; Wagner et al., 2016), implying that plant age is likely to be generally important for microbes. At the same time, other manipulative studies have shown that inoculating legumes with a higher diversity of rhizobial strains leads to decreased host plant performance (Barrett et al., 2015; Simonsen et al., 2014). If diversity of rhizobia does reduce plant growth (perhaps due to interference effects) this opens the interesting possibility that the increase in diversity with larger plants could lead to a negative feedback effect that will eventually limit the ability of an Acacia tree to grow larger.

To our knowledge this is the first ecological study to demonstrate a positive relationship between legume plant size and naturally occurring symbiotic rhizobial diversity. In addition to demonstrating a positive relationship between plant size and symbiont diversity, our study also demonstrates that the effect of tree size on microbial diversity is sufficiently strong that it can even be detected across a large complex landscape of high climate and soil variability. Together, these studies contribute to a growing body of evidence that plant size is important in structuring the diversity of soil microbe communities (Krah et al., 2018; Meaden et al., 2016) and highlight the potentially important role of host size and age in modifying ecological processes in the rhizosphere. Finally, these studies provide additional considerations that can be
in informative for future sampling design wishing to measure patterns of microbial diversity in natural populations.

ACKNOWLEDGEMENTS

We thank Cathryn O’Sullivan and Alan Richardson for helpful comments on the manuscript. This research was funded in part by an Ignition Grant from the Centre for Biodiversity Analysis, Canberra, Australia.

AUTHORS’ CONTRIBUTIONS

R.D. and A.K.S. contributed equally to this study and manuscript. A.K.S. created the sampling design and experimental idea, in consultation with S.M.P. and N.R.-B. A.K.S. and R.D. collected the data, analysed the data, and led the writing of the manuscript. R.D. generated the figures. All authors contributed critically to the drafts and gave final approval for publication.

DATA ACCESSIBILITY

All raw sequences and processed data necessary to reproduce the analyses in this paper are archived at Dryad Digital Repository: https://doi.org/10.5061/dryad.37nh37f (Dinnage et al., 2018).

ORCID

Russell Dinnage http://orcid.org/0000-0003-0846-2819
Anna K. Simonsen http://orcid.org/0000-0002-5091-261X

REFERENCES

Krah, F.-S., Seibold, S., Brandl, R., Baldrian, P., Müller, J., & Bässler, C. (2018). Independent effects of host and environment on the diversity...


SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.