Relationships between community composition, productivity and invasion resistance in semi-natural bacterial microcosms

Matt Lloyd Jones[1†], Damian W. Rivett[1‡], Alberto Pascual-García[1§], Thomas Bell[1¶]

[1]Silwood Park Campus, Imperial College London, UK.

†University of Exeter Medical School, Penryn Campus, Penryn, Cornwall, UK
‡School of Science and the Environment, Manchester Metropolitan University, Manchester, UK
§Institute of Integrative Biology, ETH, Zürich, Switzerland
¶Department of Life Sciences, Imperial College London, Silwood Park Campus, Ascot, UK

Abstract

Experiments with artificial communities have suggested that invasion resistance in microbial communities is often a side-effect of community members' contribution towards overall community productivity (broadly defined as cumulative cell density and/or growth rate). However, few experiments have investigated this in natural microbial communities. We conducted experimental invasions of two bacterial species (*Pseudomonas fluorescens* and *Pseudomonas putida*) into laboratory microcosms inoculated with 680 different mixtures of bacteria derived from naturally-occurring microbial communities collected in the field. Using 16S amplicon sequencing to characterise microcosm starting composition, and high-throughput assays of community phenotypes including productivity and invader survival, we determined that productivity is a key predictor of invasion resistance in natural microbial communities, substantially mediating the effect of composition on invasion resistance. The results suggest that similar general principles govern invasion in artificial and natural communities, and that factors affecting resident community productivity should be a focal point for future microbial invasion experiments.

Introduction

Microbial communities of all types are challenged by the arrival of dispersing microbes, which may displace resident taxa and alter ecosystem functioning (Amalfitano et al., 2015; Fernandez-Gonzalez et al., 2020; Kinnunen et al., 2016; Litchman, 2010; Mallon et al., 2015a; Thakur et al., 2019). A primary way in which established communities defend against such invasion attempts is by achieving a high level of productivity (strictly defined as community members' growth rate but often approximated by standing biomass) in their home environment before invaders arrive (Huston, 2004). High levels of productivity are concomitant with the depletion of resources, the establishment large populations and the occupation of physical space – and consequently, invaders' (in)ability to survive in that environment (Ghoul and Mitri, 2016; Huston, 2004; Stubbendieck et al., 2016).

Studies with simplified, synthetic microbial communities have suggested that community productivity is of such importance to a microbial community's invasion resistance that it is often the main explanation for the effect of that the composition of the community inoculated into the microcosm has on invasion resistance (Eisenhauer et al., 2013; Hodgson et al., 2002; van Elsas et al., 2012; Yang et al., 2018). In these studies, microcosms containing a sterile culture medium (e.g. lab broth, autoclaved soil) are inoculated artificial microbial communities constructed from different combinations of culturable taxa and/or dilutions of natural communities (typically 90% at each dilutions step). Each of these communities - differing in their composition - is then invaded with the same population of microbes to assess the relationship between community composition, productivity and invasion resistance. Such experiments have shown that the presence of individual and/or combinations of resident species (composition) with the highest cell densities and/or growth rates (productivity) explain most of the variation in invasibility between communities (Eisenhauer et al., 2013; Hodgson et al., 2002; van Elsas et al., 2012; Yang et al., 2018). Furthermore, since higher-diversity communities are generally more productive (Bell et al., 2005b), there is often a negative relationship between a resident community's diversity and invasibility (e.g. Hodgson et al., 2002).

However, the relationship between a microbial community's composition, productivity and its invasion resistance is far from clear-cut. For example, Hodgson et al., (2002) noted that whilst the effects of composition on invasion resistance mainly manifested as productivity differences, there was also evidence of productivity-independent effects of composition (e.g. niche complementarity, antagonism, facilitation). Similarly, De Roy et al., (2013) found no clear relationship between composition and productivity in artificial microbial communities grown in complex media, but nonetheless a clear, strong relationship between resident community composition and invasion resistance. These results from experiments with artificial communities suggest that despite the predominance of productivity, there is scope for productivity-independent effects of composition on invasion resistance. Such direct effects of composition might become more apparent in natural communities, where niche space is likely to be more highly packed and productivity differences less pronounced (Eisenhauer et al., 2012). To date, however, the relationships between microbial community composition, productivity and invasion resistance have not been extensively explored in natural microbial communities.

Here, we use natural microbial communities to test the hypotheses that A) resident community productivity is a primary factor underpinning invasion resistance and B) the effect of resident community composition on invasion resistance is mostly mediated by productivity. We conducted experimental invasions into 680 bacterial communities collected from rainwater pools (water-filled beech tree holes/phytotelmata). These types of model communities have been extensively used in previous microcosm experiments (Bell et al., 2005b, 2010; Fiegna et al., 2015a, 2015b; Foster and Bell, 2012; Glücksman et al., 2010; Jones et al., 2017; Lawrence et al., 2012; Rivett et al., 2016; Rivett and Bell, 2018; Scheuerl et al., 2020), and are particularly useful for conducting biodiversity-ecosystem

functioning type experiments because diversity is likely to be determined largely by habitat size in nature (Bell et al., 2005a; Woodcock et al., 2007). We grew and invaded these communities in a common garden, complex medium reflecting their natural growth medium (a beech leaf-based 'tea'), assessing the survival of two invaders (*Pseudomonas fluorescens* & *Pseudomonas putida*) in these communities at 24, 96 and 168 hours (1, 4 & 7 days) post-invasion (Figure 1A).

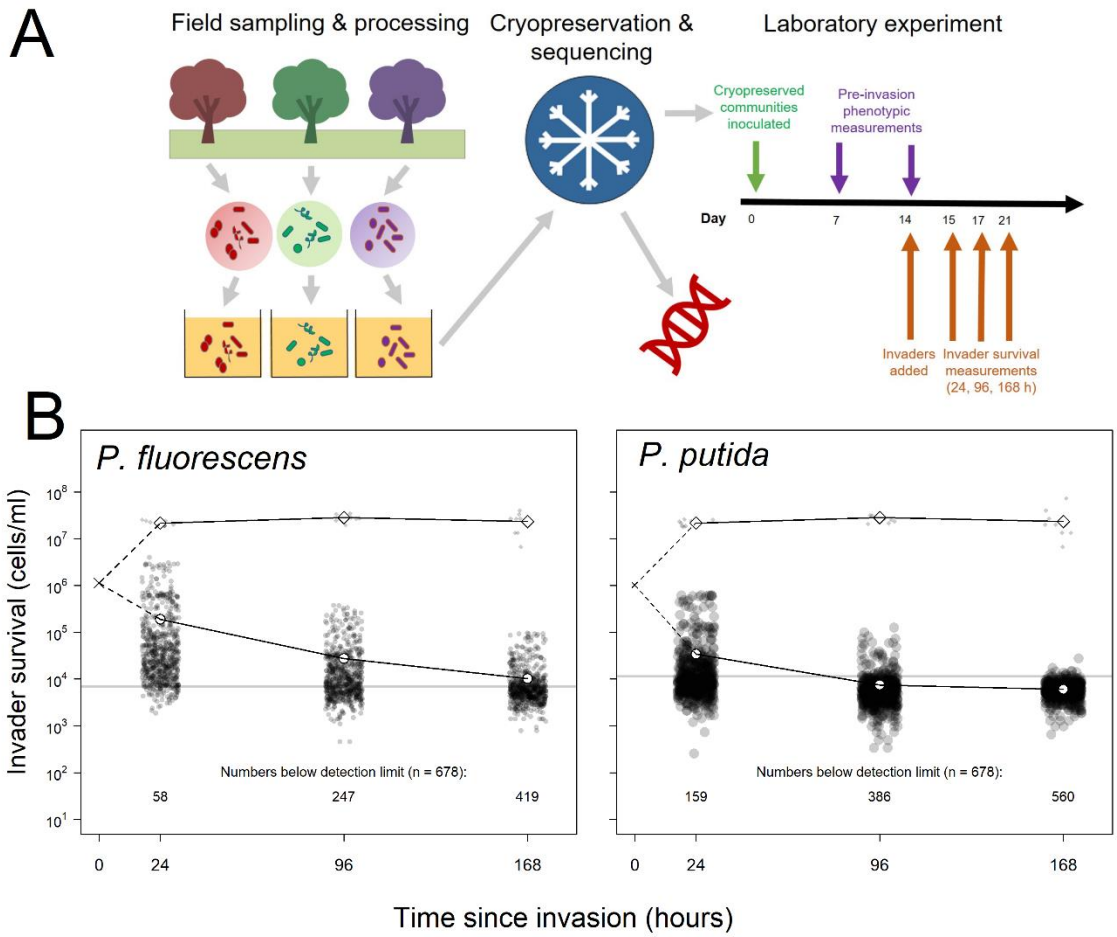
Following previous experiments, we analysed the relationship between invasion resistance and the composition of the community added to each microcosm at the start of the experiment. Unlike previous experiments using synthetic microbial assemblages, the inoculated communities were naturally occurring bacterial assemblages. We quantified the starting community composition of the microcosms using amplicon sequencing (16S rRNA locus) to estimate the sequence abundances of OTUs inoculated into the microcosms (Day 0). As well as estimating the starting community composition in this way (i.e. 16S genotypes present), we also took various phenotypic measurements of the communities at 7 and 14 days of growth in the laboratory microcosms (the latter being the day of invasion). Community productivity was the primary community-level phenotype of interest to us, quantified by measuring cell density and respiration before invasion. Additionally, we took phenotypic measurements related to metabolic activity as alternative hypothesis to productivity, and to further contextualise any results related to composition and productivity. These measurements were the potential metabolic activity (ATP levels) and capacity of the communities to degrade a set of 4 specific substrates that we expected to be important components of these environments (cellulose, chitin, xylose and phosphate). The experiment allowed us to identify the main components of invasion resistance in natural communities, and assess whether the effects of resident community composition on invasion resistance primarily manifest as effects on productivity.

Results

Experimental set-up and general observations about invader survival

Invaders were inoculated into established (resident communities grew for 14 days before invasion), closed (i.e. no new resources added during the experiment) microcosms at a high but naturally feasible propagule size, and thus declined from their initial density upon invasion (Figure 1B). As in the similar experimental set-ups of van Elsas et al., (2012), we thus quantified 'invader survival' as our response variable - or conversely, the ability of the resident communities to eliminate invaders under intense invasion pressure (see Methods: Laboratory techniques for more details). For both invaders, invader survival declined from an inoculum of approximately 10^6 cells/ml to 10^5 cells/ml

Figure 1: Summary of experimental set-up and broad patterns of invader survival across the three sampling points. A) Schematic depicting the sampling and processing of communities (field sampling and growth of lab acclimation of communities, the cryopreservation & sequencing of the lab-acclimated communities, and the set-up and sampling scheme of the laboratory experiment described here. B) Invader survival values for both invaders at each of the three sampling points in monoculture (diamonds) and in communities (circles). Grey line represents the detection limit, dashed line represent inferred trajectories between the inoculation density and the invasion densities, as the inoculation density was measured in the invader culture prior to its inoculation into communities.



1 day after invasion, to ≥10^4 cells/ml at 3 and 7 days after invasion. Survival was higher for *P. fluorescens* than *P. putida*, and increasing numbers of invasions fell below the strict detection limit of the invasion assay, which we account for in our analysis (Figure 1B; Methods: Laboratory techniques).

Most important explanatory variables for invader survival

To select the components of resident community starting community composition and realised community phenotype that best predicted the two invaders' survival in the microcosms across the three sampling points, we compared explanatory variable importance using random forest regressions. These 6 random forests (2 invaders x 3 invader survival sampling points) each contained 37 explanatory variables for invader survival - 23 relating to

the starting community composition and 14 relating to community phenotype in the microcosms prior to invasion. Of the 23 variables relating to starting composition, 3 were diversity metrics calculated from the OTU abundances (see Methods: Computational techniques: Diversity metrics calculated from compositional data), whilst the remaining 20 related to the abundances of 'functional groups' of OTUs – broadly defined as groups of OTUs sharing an equivalent set of significant co-occurrences and segregations with other OTUs across all the communities – and selected as the best dimensionality reduction of the OTU data in preliminary analyses (see Methods: Computational techniques; Supplementary Figure 1). The phenotypic measurements were taken at days 7 and 14 after inoculation into the microcosms (7 days before invasion and immediately prior to invasion, respectively; Figure 1A).

Explanatory power of the models (Figure 2) including all variables relating to starting composition and realised phenotype ranged from 8.14% to 59.37% (pseudo R^2). Models also performed substantially better at 24 compared to 96 and 168 hours - concomitant with invader densities declining from their introduction density over time in this closed system, and accordingly increasing number of invasions falling below the detection limit (Figure 1B; Methods: Laboratory techniques). The models also performed better for the *P. fluorescens* invasion experiments (minimum explained variance: 40.40%) than for the *P. putida* invasion experiments (minimum explained variance: 8.14%) – which also appeared to be associated with more *P. putida* invasions falling below the detection limit. Note that the random forest method somewhat accounts for the difficulty of differentiating noise from true signal below the detection limit by iteratively sub-sampling which samples included in each of its constituent regression trees. However, where there are larger numbers of invasions falling below the detection limit, this becomes more difficult and explanatory power hence drops. Nonetheless the method was sufficient to identify the most important explanatory variables for downstream analysis, where we account for the detection limit problem with a sensitivity analysis.

Across random forests, phenotypic measures related to community productivity before invasion were by far the strongest and most consistent individual predictors of invader survival (Figure 2). Measures of community cell yield and respiration consistently had the highest variable importance values - quantified as the relative increase in the Mean Square Error (% IncMSE) obtained when the data associated to the variable under analysis is absent from a regression (Figure 1). The variable importance of cell yield and respiration was generally higher than that of other variables relating to either phenotype or the starting composition of the microcosms. However, the abundance of Functional Group 18 (containing 12 OTUs assigned to *Cedecea spp., Citrobacter werkmanii, Erwinia persicina, Escherichia shigella spp., Klebsiella pneumoniae, Pantoea agglomerans, Pantoea vagens, Serratia fonticola, Serratia liquefaciens, Serratia quinivorans* & *Trabulsiella spp.)* was sometimes of comparable variable importance

Figure 2: Comparison of total variance explained (top bars) and the variable importance values (bottom heatmap) of the six random forests, computed for each of the two invaders and each of three the invader survival sampling points at 24, 96 and 168 h post-invasion. Total variance explained is calculated as pseudo R-squared: 1-Mean Squared Error/variance(invader survival) of the random forest. Variable importance values are the percentage increase in Mean Squared Error (IncMSE %) when the variable is not permuted i.e. a high (low) value represents a variable of high (low) importance to explaining invasion success. Each column in the variable importance heatmap represents the variable importance values of the random forest using functional Groups represented by the orange bar in the top figure. The heatmap is split into compositional (above split) and functional (below split) variables. Compositional variables labelled "C+number" refer to the functional group ids.

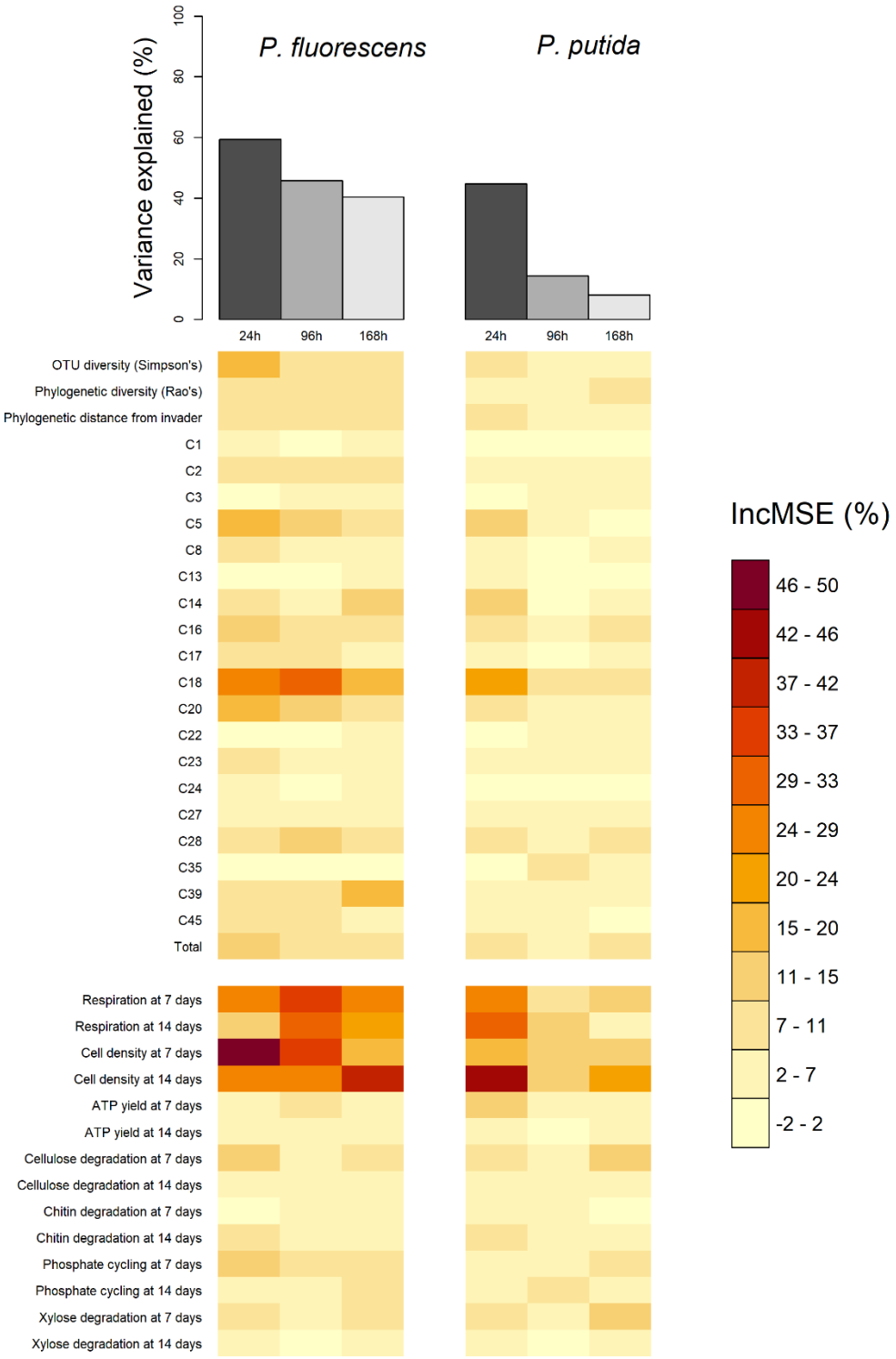
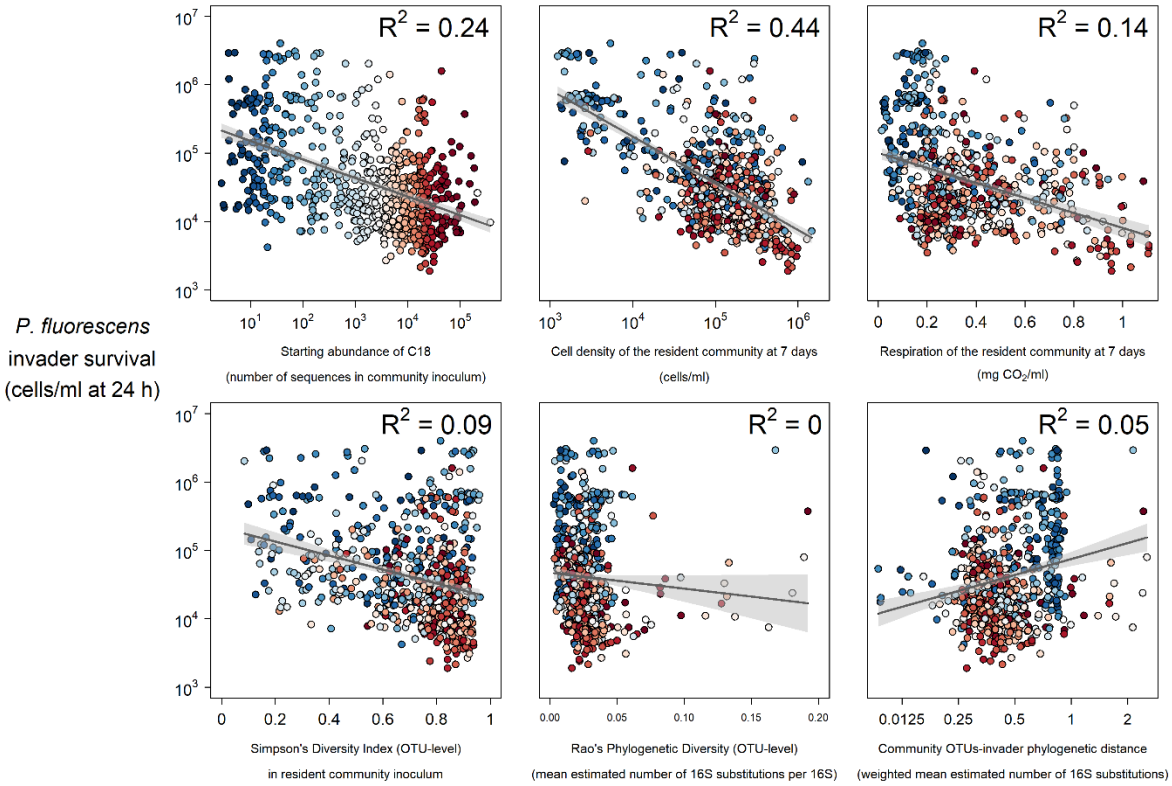


Figure 3: Selected strong (top) and weak (bottom) relationships between explanatory variables and *P. fluorescens* invasion success at 24 hours post-invasion. Colours represent the abundance of Functional Group 4 in each community (blue low, red high).



to the productivity variables. All of these variables were approximately linearly and negatively correlated with invader survival; increasing cell yield, respiration and abundance of Functional Group 18 was correlated with lower invader success (Figure 3).

Other variables relating to starting composition were weaker individual predictors of invader survival. As well as Functional Group 18, other functional groups including Functional Group 5 (*Novispirillum itersonii, Ralstonia pickettii Acinetobacter towneri & Acinetobacter genomospecies 3)* and 20 (*Acidovorax spp., Acinetobacter calcoaceticus, Acinetobacter johnsonii, Aquabacterium spp., Brevundimonas aurantiaca, Caenimonas.spp., Delftia lacustris, Eptothrix spp., Herbaspirillum rubrisubalbicans*, *Herbaspirillum spp., Massilia timonae, Paucimonas spp., Phenylobacterium spp, Pseudomonas balearica, Pseudomonas pseudoalcaligenes & Stenotrophomonas maltophilia)* also appeared to be somewhat important for *P. fluorescens* invasion. There was some evidence of a weak negative relationship between Simpson's diversity and invader survival (Figures 2, 3) - although the abundance of key functional groups was a more reliable indicator of invader survival across time-points and invaders (Figure 2). There was little evidence for a strong effect of phylogenetic diversity (Rao's quadratic entropy; a phylogenetic equivalent to Simpson's diversity index) or phylogenetic distance of the community from the invader (though these results are subject to the reliability of 16S phylogenetic tree; see Discussion).

Overlaying the abundance of the most important functional group (Functional Group 18) on the above described relationships (Figure 3) suggested that the presence of particular 16S genotypes determined overall community phenotype (in terms of productivity and invasion resistance realised in laboratory microcosms). Highly invadable communities with a low cell density and respiration were also those with a low abundance of Functional Group 18 OTUs and similarly, communities with a low abundance of Functional Group 18 had a lower level of respiration (blue colours in Figure 3).

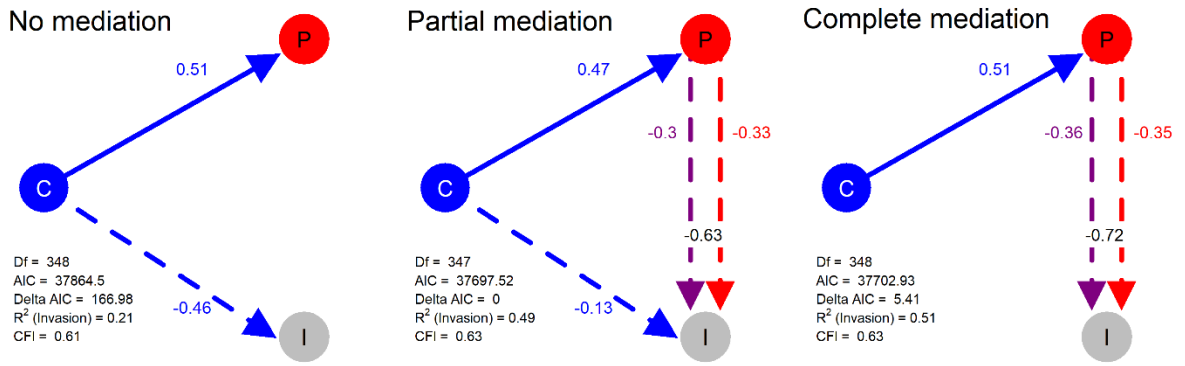
There was little evidence for strong effects of more specific measures of community phenotype, with ATP activity and the potential to degrade specific substrates having a weak effect overall (Figure 2; Supplementary Figure 2).

Mediation of the effect of starting composition by productivity (structural equation models)

Having identified community productivity as the most important predictor of invasion resistance and identified the abundance of functional groups inoculated into the microcosms as the putative cause of this, we aimed to estimate the extent to which the effect of starting composition on invasion resistance was mediated by productivity.

We used structural equation models (SEMs) to identify whether starting community composition was associated with lower invader survival solely because it determined the productivity of communities. For these models, we created a latent variable called “Productivity” including cell yield at 7 and 14 days. We chose these variables because in 2 out of the 6 of experiments (those with the highest average invader survival) cell yield and 7 days was most important, whilst in the other 4 (those with the lowest average invader survival) cell yield at 14 days (the day of invasion) was most important (Figure 2). Although important in the random forests, respiration was not included in the SEMs because we considered that respiration must result from cell density/growth rate, and cell yield measurements were always the most important explanatory variables in random forests. We created another latent variable called “Composition”, including the abundance of the 19 functional groups (+ 1 total mean abundance of remaining OTUs) and the Simpson’s diversity of the OTUs. We included all 19 (+1) functional group abundances and diversity because although Functional Group 18 was clearly the most important determinant of invader survival, the importance values for starting composition variables were more evenly distributed than those for the phenotypic measurements, where productivity measures were clearly the most important (Figure 2). Finally,

Figure 4: Structural models in a mediation test. Nodes in the diagrams represent latent variables (C = Composition (blue), P = Productivity (red), I = Invasion (grey). Arrows between nodes represent regressions between dependent and independent variables (blue = direct effect of composition, red = direct effect of productivity, purple = composition effect mediated by productivity). Direction and value of each path is indicated by arrow type (positive effect = solid line, dotted line = negative effect) and the standardized regression coefficient adjacent to the arrow. In the path P → I the total effect of productivity (sum of direct and composition-mediated effects) is shown between both lines in black.



we created an "Invasion" latent variable encapsulating all measurements of invader survival for both invaders at the three sampling points (24 h, 96 h & 7 days after invasion).

Having specified the latent variables, we sought to understand the mediation of composition by productivity across our experiments by using three different SEMs in which these latent (dependent) variables Composition and Productivity influence the latent (independent) variable Invasion following one of these hypothesis:

1) **No mediation** – only resident community starting composition (directly) affects invaders' survival
2) **Partial Mediation** – resident community starting composition (directly) affects invaders' survival, but also has an indirect effect through its influence on community productivity, which in turn affects invader survival.
3) **Complete Mediation** – resident community starting composition only influences invaders' survival through its indirect effect on community productivity.

Model comparison strongly supported the idea that the effect of starting composition on invader survival was strongly mediated by realised productivity, with the No Mediation model being easily rejected (ΔAIC = 166.98). The Partial Mediation model was the best fitting of the three models and was significantly better than the Complete model (likelihood ratio test, ΔChi^2 =7.81, pval = 0.0052) - though this should be interpreted with caution as the quality of the models is not optimal (CFI ~ 0.63). Based on the estimated coefficients of the best, Partial Mediation model (Figure 4), we estimated that a minimum of 48% of the effect starting composition on invader survival (the product of coefficients C→P x P→I) was mediated by productivity (0.47 x -0.63 = -0.30; which represents a

proportion -0.30/-0.63 = 0.48 of the total effect). Furthermore, comparing the total (direct + indirect) effects of productivity (P→I + C|P→I) and composition (C→I + C|P→I) revealed that the total effect of productivity (-0.33 + -0.3 = -0.63) was much stronger than that of composition (-0.13 + -0.30 = -0.43).

Given the relatively small (but significant) difference between the Partial and Complete mediation model fits, the strong effects of productivity in the Partial model, and the sub-optimal quality of the models, we conducted an additional sensitivity analysis (see Methods: Statistical Methods). This sensitivity analysis tested how sensitive model selection was to the invader survival values fell below the strict detection limit of the invasion assay (see Methods: Laboratory techniques: Invader survival). The sensitivity analysis supported the Complete (the best model in 97.3% of permutations) over the Partial mediation model (the best model in 2.6% of permutations), highlighting the marginality of the model selection result.

Taking everything into account, we believe the most conservative interpretation of these results, therefore, is that most of the effects of composition occurred through productivity.

Discussion

We found that even in complex, species-rich bacterial communities, community productivity is a key determinant of invasion resistance, with much of the effect of community composition manifesting as effects on productivity. This is consistent with the results of earlier experiments with simpler, more artificial communities (Bonanomi et al., 2014; Eisenhauer et al., 2013; Hodgson et al., 2002; van Elsas et al., 2012). Productivity (pre-invasion cell density and to a lesser extent respiration) were certainly the most useful individual explanatory variables for invasion success in our experiments, though with certain compositional variables also having a strong effect (Figures 2, 3). Furthermore, structural equation modelling suggested that much of compositional effect was due to the effect that the bacteria inoculated into that microcosms had on productivity, which in turn effected invasion success (Figure 4). Nonetheless, there is likely a still a (smaller) role for more direct, productivity-independent effects of composition – again, consistent with previous work with artificial communities (De Roy et al., 2013; Hodgson et al., 2002), and for composition-independent effects of productivity.

Importance of community productivity for invasion resistance

There is ample evidence that resident community productivity reduces the potential for the growth of invading species in microbes and non-microbes alike (Crawley et al., 1999; Hodgson et al., 2002; Jousset et al., 2011;

Kinnunen et al., 2016). In our system too, resident community productivity before invasion was a key predictor of invasion resistance in our study system. Random forest regressions showed that when explanatory variables were considered individually, productivity measures had the strongest correlations with invader survival. Plotting these relationships showed that both cell density and respiration before invader survival had clear negative relationships with both invaders' success (Figure 3). For example, for *P. fluorescens* invader survival at 24 hours (the response variable most well-explained by the explanatory variables), for every order-of-magnitude increase in resident community cell density at 7 days, there was an approximately order-of-magnitude decrease in the invader density (Figure 3). Similarly, for every milligram per ml of CO_2 produced by the resident community by 7 days, there was an approximately order-of-magnitude decrease in the invader density (Figure 3).

Cell density and respiration were likely important predictors of invasion resistance because they were good proxies for how much the resident community had used the available resources that could otherwise have been used by the invader i.e. resource limitation. Interestingly, productivity at 7 days before (rather than immediately prior to) invasion had the most explanatory power in the models - suggesting that resource limitation may have operated in complex ways. One possibility is that those communities that could degrade the resource base most rapidly could also do so most efficiently, leaving less resource for the invader at 14 days. Alternatively, communities that have been at carrying capacity for longer periods may exhibit behaviours that select against invasion. For example, limited resources and/or metabolic stress can select for antibiotic production (Craney et al., 2013) or the production of inhibitory secondary metabolites (Watrous et al., 2013).

Productivity-mediated effects of composition

Structural equation modelling further revealed that productivity significantly mediates the effect of starting composition on invasion success - further emphasising the extent to which invasion resistance is mainly a side-effect of productivity. Model comparison demonstrated good support both for scenarios assuming Partial (48%) mediation of the effect of composition by productivity, and for scenarios assuming Complete (100%) mediation of the effect of composition by productivity, with only marginal differences when a sensitivity analysis was considered. This demonstrates that the presence of specific genotypes in microbial communities may affect community-level phenotype in quite general ways; largely affecting the invasion resistance of the community by bolstering the overall population size and/or growth rate.

This result is similar to that of previous studies with much more artificial microbial microcosms. Hodgson et al., (2002) demonstrated a strong negative correlation between *P. fluorescens* SBW25 invasion success and resident *P. fluorescens* SBW25 community diversity (R^2 = 0.7) in 2-5 strain communities grown in a defined laboratory

medium, driven by the presence of the most productive *P. fluorescens* SBW25 strain type. Another more recent study with five-strain communities of *Ralstonia* grown in defined laboratory media, showed that resistance to invasion by another *Ralstonia* strain was determined by the presence of one or two strains with the fastest growth rate (Yang et al., 2018). What is remarkable about our results is not that we found a similar result to these studies *per se*, but that we found a similar result in microcosms containing a complex growth medium (beech leaf tea) inoculated with communities with richnesses of 66-236 OTUs. In our system as in more artificial microcosm experiments, it appeared that the starting abundance of particular group of bacteria – in our case, a putative 'functional group' containing several *Enterobacterales (*Functional Group 18) - determined the level of invasion resistance realised by the community (i.e. a dominance/selection effect).

The most parsimonious explanation for this effect in our system was simply that this was the most productive functional group, and a higher starting abundance gave this group of bacteria a better chance of outcompeting other bacteria and raising the level of growth in the community (i.e. a priority effect). Alternatively, the starting abundance of particular functional groups in a microcosm may determine community productivity in more complex ways – for example, because the most abundant early colonisers cause local environmental changes that affect the growth potential of other colonising bacteria (e.g. through cross-feeding and competition). In dental biofilms it has been demonstrated that a full biofilm can only be achieved with a certain order of colonisation, with metabolically-similar groups of early colonisers causing local environmental changes that allow for subsequent bacteria to colonise, allowing the full development of the biofilm (Mazumdar et al., 2013). Similarly, in a study system more similar to our own, it has been shown that the ability bacterial species to colonise a new phyllosphere environment depends on the local density (and by implication, identity) of other neighbouring bacteria colonising the environment at the same time - likely because more intense competition for resources reduces the average growth success of the population (Remus-Emsermann et al., 2012). Such species interactions/succession-type dynamics are likely to be at least partly driving productivity-mediated effects of composition in our system also - especially as our previous work has shown our 'functional groups' correspond to distinct metabolic profiles inferred from predicted metagenomes (Pascual-García and Bell, 2020b).

Productivity-independent effects of composition

There was also some evidence that community composition impacted invasion resistance independently of the impacts of community productivity, though in a more minor way. Under the Partial Mediation scenario, 54% of the effect of composition was attributable to more direct effects of composition on invasion. Despite the predominant effect of productivity, therefore, the relationship between composition and invasion resistance in natural

communities is likely to go beyond than productivity - as has been seen in a more limited sense in previous studies with artificial communities (De Roy et al., 2013; Hodgson et al., 2002).

One explanation for this is that groups of bacteria which were abundant in some communities but not others were functionally distinct; specialising on resource-poor or resource-rich environments (Pascual-García and Bell, 2020a) and degrading harder-to-access and/or rarer components of the resource pool. Previous work using the same study system has also shown that the starting composition of communities affects the extent to which recalcitrant substances are able to be used during subsequent community growth (Rivett et al., 2016).

Another possible explanation is that particular resident species acted as ecosystem engineers (Pascual-García et al., 2020). We have already mentioned the possibility of bacteria-driven environmental changes setting limits on the community productivity and leaving certain communities more vulnerable to invasion. Additionally, there may be other ways in which particular species engineer the environment to affect invader survival more directly - such as by modifying the pH of the environment or producing reactive oxygen species. Regarding pH modification, bacterial acidification has been shown to be an important factor in recent microbial invasion experiments in soil microcosms, which showed that *Pseudomonas* may be prevented from invading by the presence of species that alter the growth medium towards more acidic pH values (Amor et al., 2020). Regarding the production of reactive oxygen species, this has recently been demonstrated to be important in in-host studies of invasion resistance; *E. faecalis* has been shown to protect *C. elegans* nematodes from *Staphylococcus aureus* invasion by the producing reactive oxygen species which act as an antimicrobial to kill the invaders (Ford and King, 2020). However, whilst these more specialised mechanisms of invasion resistance are likely to be fairly common, we emphasise again that they are likely to play a more minor role than more general factors such as community growth rate.

Sub-hypotheses with weaker or no support

We found some evidence for the expected negative relationship between the diversity (Simpson's diversity of OTUs) of the inoculated community and the invaders' survival (Figure 3) in this semi-natural system, though this was relatively weak ($R^2 = 0.09$ for *P. fluorescens*). Although it is hard to make a direct comparison, superficially at least, this weak relationship contrasted with previous experiments using artificial communities constructed from isolates or created by dilution-to-extinction of natural communities, where the relationship was clearer (De Roy et al., 2013; Eisenhauer et al., 2013; van Elsas et al., 2012). Nonetheless, given that previous experiments often found that the negative diversity-invasion relationship plateaued at higher diversities (Bonanomi et al., 2014; Eisenhauer et al., 2013; van Elsas et al., 2012), the weak negative slope observed in our high-diversity experiment

may simply be because our communities all had relatively high levels of diversity. This suggests that species-like diversity is likely to be a less important predictor of invasion resistance in natural rather than artificial communities.

Phylogenetic diversity metrics had even poorer explanatory power, though this result should be interpreted with caution as gene trees based solely on the 16S gene have substantial limitations and estimating the phylogenetic diversity of natural communities is still a challenge (Rajendhran and Gunasekaran, 2011). Regardless, some preliminary conclusions about the importance of 16S phylogenetic diversity can be made. First, a phylogenetic equivalent of Simpson's diversity (Rao's quadratic entropy) was not related to invader survival, suggesting 16S phylogenetic alpha diversity is not predictive of invader survival. Secondly, the 16S-based phylogenetic distance of the communities from the invader (estimated number of mutations to between per nucleotide on their 16S gene) and the invaders' success only had a very weak positive relationship with invader survival (Figure 2). Darwin's 'naturalisation hypothesis' posits that invaders are more likely to establish when invading communities composed of resident species that are, on average, more phylogenetically distant from them, because of greater supposed niche differences (Darwin, 1859). However, evidence from previous bacterial microcosm experiments with artificial communities to support this hypothesis is mixed (Gu et al., 2019; Jiang et al., 2010; Kinnunen et al., 2018; Li et al., 2019), and we did not find strong evidence to support this hypothesis in natural communities. Somewhat related to the phylogenetic hypothesis, by using two different invaders in separate experiments, we also explored whether resident communities were better at resisting an invader when they had a high abundance of conspecific species (as was the case for *P. putida;* see Methods: Laboratory techniques: Choice of invaders*)*. Qualitatively, these results corroborate the idea that invaders with conspecific residents (*P. putida)* are less successful - though the general mechanisms of invasion resistance appear similar in both *P. fluorescens* and *P.* putida experiment (distribution of the variable importance values in Figure 2, shape of relationships between invader survival and the explanatory variables in Figure 3). A more dedicated study with more invader treatments would be needed to study this sub-hypothesis properly.

Finally, despite the importance of productivity measures for predicting invader survival, more specific measures of community phenotype related to bacterial metabolism were not informative. Again, this suggested that invasion resistance is primarily the result of general rather than specific mechanisms.

Unexplained variation

Finally, although we successfully explained much of the variation in invasion using the measured variables, a large component of the variation remained unexplained across the six sub-experiments. Given the predominant role of

composition and productivity, we expect that some of this unexplained variation is due to the limitations of our techniques in capturing these key components of the resident communities.

Regarding composition, firstly, 16S amplicon sequencing was likely inadequate at capturing the microbial composition of the resident community inocula, and metagenomic/shotgun sequencing might have better characterised the compositional and functional diversity of communities and correlated with invasion. Secondly, although we tried to ensure that only bacteria were present in our communities (through filtering, passage and anti-fungal treatment, visual and flow-cytometric observation; see Methods: Field sampling of communities), it is possible that some non-bacterial microorganisms such as fungi or phage survived and could have affected invader survival. This is difficult to avoid and so future studies might improve explanatory power by keeping the inoculated communities even more intact, characterising compositional effects on invasion at multiple taxonomic and trophic levels simultaneously (again using metagenomic sequencing). Thirdly, additionally characterising community composition nearer to the point of invasion may have improved explanatory power and better contextualised the results. Our unpublished work with a subset of these communities suggests that despite our lab acclimation period (see Methods: Field techniques: Field sampling of communities), some more minor compositional changes still occur in these communities after inoculation (as should be expected), with communities converging towards a similar composition but maintaining a high level of diversity. Characterising these changes and accounting for them in models of future experiments might therefore help confirm, for example, whether differences in communities' invasion resistance are the result of the rate at which certain functional groups of bacteria take over the community during growth.

Regarding productivity, the strong predictive power of productivity at 7 days (i.e. 7 days prior to invasion) implies that the temporal dynamics of the community growth is an important component of invasion resistance. Increasing the temporal resolution of growth measurements may therefore have improved the predictive power, as would methods to better distinguish active vs dormant proportions of the population.

Aside from productivity and composition, measuring additional variables in the growth period including pH (Amor et al., 2020) and better inferring functional performance using metagenomics and/or metabolomics could have also helped identify the role of environmental modification in preventing invasions. We also do not exclude the possibility that the unexplained variation may result from the stochastic nature of invasions, since stochastic processes can partly govern invader survival in microbial communities (Amor et al., 2020; Kinnunen et al., 2018).

Conclusion

Experiments with artificial communities have suggested that the composition of microbial communities mostly affects invasion resistance by affecting their productivity (Bonanomi et al., 2014; Eisenhauer et al., 2013; Hodgson et al., 2002; van Elsas et al., 2012; Yang et al., 2018). Our experiment lends support to the extension of this claim to natural communities; semi-natural microcosms of bacteria mostly achieve invasion resistance primarily (though not wholly) as the result of their productivity, which is the result of the identity and abundance of the bacteria with which they are inoculated.

A new generation of microbial invasion ecology experiments with more natural microbial communities (Bell, 2019) might therefore benefit from placing a greater emphasis on productivity where diversity has historically been the focus of experimental designs. Productivity is arguably likely to be more variable than diversity among natural communities, because natural communities are already saturated with species and the population density of a community is more easily affected by environmental disturbances (e.g. dilution by rainfall, environmental pollution). Several lines of evidence in experimental microbial ecology exploring the effect of disturbances on invasion already suggest that disturbances primarily affect invasion success by affecting the population size of the resident community, for example (e.g. Lear et al., 2020; Mallon et al., 2015b). Equally, work outside microbial invasion ecology with pathogens frequently suggests that these ‘invaders’ frequently gain entry into host microbiomes by reducing the density of resident competitors (e.g. Brown et al., 2007, Wei et al., 2018). Efforts to characterise resident community growth trajectories before invasion - and what drives differences and disturbances to them - should thus be continued and extended in order to better understand microbial invasion resistance. We believe that taking resident community productivity as a null hypothesis for invasion resistance in this way will reveal more clearly the way invasion resistance emerges and is disrupted in natural microbial communities of all types (Amalfitano et al., 2015; Fernandez-Gonzalez et al., 2020; Kinnunen et al., 2016; Litchman, 2010; Mallon et al., 2015a; Thakur et al., 2019).

Methods and Materials

Field techniques

Field sampling of communities

We used naturally-occurring bacterial communities collected from 680 tree hole communities across Southern England between August 2013 and April 2014. The term 'tree hole' refers to the naturally occurring, semi-permanent pools of rainwater that collect in the pans formed by the buttress roots of *Fagus sylvatica* beech trees (Bell et al., 2005). Bacterial communities inhabit these phytotelmata, thriving on the organic matter (the bulk of which is beech leaf litter) that collects in them. Communities were sampled in the field by actively searching for pools of water in the buttress roots of beech trees. Once located, GPS coordinates and date of collection was recorded, the tree holes were homogenised using a sterile plastic Pasteur pipette, and 1ml was transferred into a sterile 1.5 ml centrifuge tube (Starlab; Milton Keynes, UK). Samples were transported to the lab at an ambient temperature within 24 hours of collection, where they were diluted 1:4 in sterile phosphate buffered saline and filtered (pore size 20-22 µm, Whatman 4 filter paper) to remove debris and large organisms. 500 µl of this filtrate was then used to inoculate 5mL of beech leaf tea medium (see 'Laboratory culturing of communities' for a description) containing 200 µg ml^{-1} cyclohexamide (Sigma-Aldrich) to remove fungi, and the communities allowed to grow for 7 days at 22°C to reach stationary phase and acclimatise to laboratory conditions. Communities were then frozen in 60% glycerol with NaCl solution (Sigma-Aldrich; Gillingham, UK) to produce a frozen community archive, and revived from these frozen stocks for further experiments.

Thus, these communities were 'natural communities' to the extent that they had only been subject to manipulations for practical purposes – dilution for filtering, the removal of fungi, acclimation to the laboratory growth medium, and freezing/thawing. We consider that the 1:4 dilution for filtering was minor compared to that typically used in dilution-to-extinction experiments (typically 90% at each dilution step). The removal of fungi as a practical simplification of the communities is perhaps most contentious manipulation, as fungi are likely to be important decomposers of leaf litter in this system. However, previous experiments also typically worked with bacteria-only resident communities and so our results are comparable. Finally, freezing the communities and re-growing them from frozen stocks for each experiment was a practical step to allow repeatable experiments using the same, fully sequenced starting community. To ameliorate this, the whole procedure of re-growing, functionally characterising and invading communities was repeated in 4 separate assays and the means of these 4 pseudo-replicated measurements used for the final analysis.

Laboratory techniques

Amplicon sequencing of inocula communities

Communities were sequenced at the point of freezing/archiving and thus, when we talk about ‘composition’ in this paper, we are referring to the composition of the inocula added to the microcosms at the start of the experiment - as is conventional for microbial diversity/composition-invasion resistance experiments. Briefly, immediately prior to archiving, DNA was extracted from all communities using ZR-96 DNA Soil extraction kits (Zymo Research Ltd, Irvine, CA, USA). Extractions were amplified on the V4 region of the 16S rRNA gene, using the barcoded 515F and 806R PCR primers using the HotStarTaq Plus Master Mix Kit (Qiagen; Valencia, CA, USA). The PCR cycle used was: 94°C for 3 minutes, followed by 28 cycles of 94°C for 30 seconds, 53°C for 40 seconds and 72°C for 1 minute, followed by a final elongation step at 72°C for 5 minutes. Sequencing was performed by MR DNA using the MiSeq platform (www.mrdnalab.com; Shallowater, TX, USA), with primers and barcodes removed according to their standard procedure.

Laboratory culturing of communities

The common garden culture medium selected for this experiment was a ‘beech leaf tea’ (BLT), designed to approximate the predominant nutrient conditions found in tree holes. To produce this medium, 50 g of autumn-fall beech leaves were autoclaved with 500 ml of distilled water to create a concentrated solution that was diluted 32-fold to produce the final BLT culture medium. To begin the common garden experiment, communities were revived from the acclimatised stocks by adding 50 µl of each stock to 1.8 ml of BLT in 1.8 ml deep-well 96 well plates and growing for 14 days at 22°C before invasion treatments were applied.

Productivity measures

During the 14 day growth period prior to invasion, cell density and respiration were characterised at 7 (7 days before invasion) and 14 days (the day of invasion) as estimates of productivity.

Cumulative density of the communities was measured by sub-sampling the communities and staining with Thiazole Orange (Sigma-Aldrich; Gillingham, UK) at concentration of 100nM for 15 minutes and analysing cultures using flow cytometry (BD Accuri C6; BD Biosciences, San Jose, CA, USA). Thiazole Orange is a membrane permeant nucleic acid stain, staining both live and dead bacteria and thus this assay represents the total cumulative yield in each community. Fluorescence gating against negative (beech tea) controls was used to count the number of fluorescent cells in each community (cumulative cell density).

Respiration of each community was measured by sub-sampling and using the MicroResp system (The James Hutton Institute; Aberdeen, UK). Briefly, agar-set indicator gels were suspended above growing sub-cultures at 7 and 14 days in a deep-well 96 well plate in an airtight system. As growing cells respired, CO_2 released was absorbed by the indicator gels above the cultures, and the gel colour changed from pink to purple. Colour change is read via a spectrophotometer prior to suspension above cultures (OD_{600} at 0 h), and after 24 hours of respiration/after suspension above cultures. The change in colour is used to calculate $mgCO_2$ released in the 24 hour period using a standard curve.

Additional phenotypic measurements

We also took phenotypic measurements related to the metabolic activity of communities at 7 and 14 days by performing assays of ATP and enzyme activity. ATP activity assays provide a general estimate of how metabolically active communities are, and we expected that communities that were more metabolically active (i.e. producing more ATP) would be more able to actively defend against invaders (e.g. through resource competition or direct competition).

ATP activity assays were performed at 7 and 14 days of community growth prior to invasion using the BacTiter-Glo™ Cell Viability Assay (Promega; Madison, WI, USA). This assay is a two-step process that releases ATP stored inside cells in order that it can bind to bind to ATP-activated luciferase present in the formula. The maximum luminescence generated in the immediate 5-minute reading period is thus proportional to the amount of ATP in the sample, which is calculated exactly by converting into concentration in nM using a standard curve.

Enzyme activity assays measured the degree to which communities were degrading particular substrates present in the BLT. We measured the metabolism of the following substrates by the following enzymes; hemi-cellulose by xylosidase, chitin by N-acetyl glucosamine, cellulose by |-1-4-glucosidase and phosphate groups by phosphatase. These substrates are common in beech leaf litter (Aneja et al., 2006), the complex substrate constituting the BLT medium. Substrates labelled with the fluorescent moiety 4-methylumbelliferone (MUB) (Sigma-Aldrich; Gillingham, UK) were incubated with communities at a working concentration of 400 M for 1 hour, as per Frossard et al., (2013). Fluorescence is generated when labelled substrates are cleaved and deprotonated by bacterially-produced enzymes. Fluorescence detected spectrophotometrically is thus used to calculate mg/ml of each enzyme associated with each substrate present in each community – a measure of the capacity of communities to metabolise certain substrates in the BLT.

Invasion assays

After the 14-day growth period during which productivity and metabolic activity were characterised, two lux-tagged, IPTG-inducable bacteria - *Pseudomonas fluorescens* SBW25 and *Pseudomonas putida* KT2440 – were separately introduced into each of the 680 communities. To perform these two invasion assays, communities were homogenised and aliquoted out twice into 237.5 µl volumes in sterile white microtitre plates. Each invader was grown in 10 ml of BLT medium for 96 hours prior to invasion, and 10 µl added to the community aliquots with 2.5 µl of 100 mM IPTG solution (final concentration 1 mM IPTG). Invaders reached mean densities of 2.5 X 10^7 ± cells/mL in 48 hours, and so this represented a high invasion pressure of approximately 10^5 invader cells entering each community containing an average of 10^5 cells according to cytometry readings.

Such invasion conditions are similar to that used by van Elsas et al., (2012), in which an established community was inoculated with an approximately equal density of invading cells. Such conditions are probably relatively common in microbial communities (Mallon et al., 2018; Rillig et al., 2015) where an established community may frequently face the invasion of an entire microbial community (e.g. during faecal deposition on soil, leaf fall into soil or water, flooding of terrestrial ecosystems by aquatic ecosystems). For example, as noted by Mallon et al., (2018), deposition of faeces into soil may introduce as much as 10^9 *E. coli* cells per gram (Tenaillon et al., 2010). For our tree hole system in which autumn leaf fall is a main input (beech tree holes were often found to be full or overflowing with leaves during fieldwork), there is evidence that in nature there are 10^6-10^9 bacteria per gram of beech leaves (Holm and Jensen, 1972). During litterfall, these bacteria would be carried on leaves into the phytotelma, which contain approximately 10^6 bacteria per ml in nature (Walker et al., 1991). Furthermore, *Pseudomonas* is a significant component of these bacterial flora (e.g. Aneja et al., 2006; our own data in Supplementary Table 1), especially just prior to the beginning of litterfall where they represent approximately 10^6 cells per gram according to one available estimate (Holm and Jensen, 1972). Thus, we believe our conditions are a somewhat realistic representation of the high-magnitude, low frequency autumn invasion of bacteria into tree hole microbial communities that have established over the previous year.

Choice of invaders

Pseudomonas were chosen as invaders both because of the aforementioned evidence of their abundance on beech leaves, and because they are fast-growing, metabolically versatile bacteria that are common colonisers, contaminants and pathogens in microbial communities (Nikel et al., 2014). Furthermore, *Pseudomonas* is one of the most cosmopolitan bacterial genera, with its members being found in many very diverse environments

(Pascual-García et al., 2014). Combined with the high invasion pressure, we therefore considered them good candidates to be successful invaders, allowing us to study invasion resistance under conditions where it was likely to be most possible and detectable, and invader-side factors were less limiting.

The choice of two closely-related species was initially motivated by the intention to explore, in a minor sense, whether invasion resistance and its mechanisms were different between communities that did not have a high abundance of conspecifics to the invader and those that did. None of the OTUs present in the resident communities was aligned to *Pseudomonas fluorescens* whilst one was aligned to *Pseudomonas putida* and had a mean abundance of 2.35% ± 0.02 (and so *P. fluorescens* represented a more 'true' invader). As detailed in our Results, anecdotally our results support the idea that invasion resistance is greater where there is a high abundance of conspecifics, but that the mechanisms of invasion resistance are similar regardless. However, there could have been many more reasons for this and species-level resolution for OTUs is unreliable (Janda and Abbott, 2007; Johnson et al., 2019; Knight et al., 2018). In order to support this hypothesis robustly a more dedicated study would need to be done, probably supported by metagenomic sequencing approaches.

Invader survival

Luminescence of lux-tagged bacterial strains is directly related to the density of metabolically active cells (Close et al., 2012). Therefore, luminescence values of each microcosm were used to calculate the number of metabolically active invader cells (invader survival) produced from the original invader inoculum. As described previously (Jones et al., 2017), we calibrated the luminescence values by performing growth assays of the invader in BLT medium with IPTG (1 mM) prior to the experiment, measuring luminescence and plating cultures onto LB agar to obtain cell numbers. There was a strong relationship between $\log_{10}$ luminescence and $\log_{10}$ plate counts (R_2 = 0.87 and 0.82 for *P. fluorescens* and *P. putida*, respectively) during log phase, so we used the calibration curve to convert luminescence values into invader cell densities. Invader survival was defined as the density of metabolically active invader cells/ml^{-1} in the resident communities after some defined period since invasion (24, 96 and 168 hours). The upper detection limit of the luminescence assay was 12 lumens (the maximum background luminescence of 1000 sterile BLT microcosms), which translated into upper invader survival detection limits of 6.86 x 10_3 cells/mL for *P. fluorescens* and 1.14 x 10_4 cells/mL for *P. putida*. These detection limits were accounted for in our sensitivity analysis (see Statistical Methods: Structural equation models) which suggested that values below the detection limit contained some real invader signal as well as noise.

In total, therefore, we assayed 680 communities (n = 680) for invasion resistance to 2 invaders with invader survival measured at 3 time points (680 x 2 x 3 = 4063 measurements total). 680 was the number of communities remaining

from the original set of 753 communities after the removal of communities with less than 10,000 sequences and without all growth assay and/or invasion assay data. This entire 3-week long assay of 4063 measurements (2 weeks of community growth, followed by invasion and measurement of invader survival up until 1 week after invasion) was repeated 4 times (technical replicates = 4) from frozen stocks to the last measurement (invader survival at 168 hours after invasion). Values used in the experimental analysis are the means calculated from these technical replicates. No explicit power analysis was used as our study was not straightforwardly hypothesis-driven (or at least there were many possible hypotheses) and we had a very large number of replicates (n = 680).

Computational techniques

Diversity metrics calculated from compositional data

Sequenced communities were characterised with bioinformatics implemented at the sequencing center (Molecular Research DNA; www.mrdnalab.com) as described previously (Rivett and Bell, 2018). Briefly, sequences under 150 bp and/or with ambiguous calls were removed, and then the remaining sequences were subjected to de-noising and chimera removal. Sequences were then classified into operational taxonomic units (OTUs) at 97% similarity. Subsequently, extremely rare OTUs occurring in less than 10 samples or with under 100 reads across all samples were removed. We also removed communities with less than 10,000 reads/sequences per sample. This process produced an OTU abundance table of 581 OTUs, with the abundance of OTUs being the number of sequences per OTU. We used this OTU abundance table for all downstream analysis relating to community composition. OTU diversity (Simpon's Index) was computed from the species abundance table using the diversity function from the R package vegan (Oksanen et al., 2019). Simpson's Index was chosen over Shannon's because it gives more weight to relative abundance vs. species richness, and many OTUs had single digit abundances in some communities. Inspection of exploratory plots of invasion against the two diversity indices also revealed that there was a clearer correlation between invader survival and Simpson's diversity.

The two phylogenetic metrics were based were calculated based on the OTU abundance table and a phylogenetic tree we constructed using the 16S sequence associated with each of the 581 OTUs as well as the two sequences associated with each of the two invaders (obtained from NCBI). These sequences were then trimmed to 250bp and aligned using MUSCLE multiple sequence alignment in Geneious 2.0. A phylogenetic tree was created with a Maximum Likelihood method of phylogenetic inference using RaxML with a threshold of 100 bootstraps using the GTR+G sequence evolution model. This tree was rooted in the Archael outgroup Halobacterium salinarum. It is important to briefly note the limitations of using such a 16S gene tree when interpreting our results about phylogenetic effects. Whilst 16S is arguably the optimum single-locus choice for microbial phylogeny analysis, 16S

tree trees' efficacy is limited by problems such as horizontal gene transfer, different versions of 16S genes within a single organism and the large differences in the similarity of 16S genes among species depending on the genus considered (Rajendhran and Gunasekaran, 2011). Our phylogenetic tree had broad consensus with the assigned genus-level taxonomy of the OTUs, but nonetheless there were inconsistencies and the presence of six extremely long branch lengths (>99th percentile) suggested that a better phylogeny is needed to make firmer conclusions about the importance of phylogenetic composition for invasion resistance.

Using this 16S phylogenetic gene tree, a phylogenetic diversity metric (Rao's quadratic entropy) was computed from the species abundance table using the diversity function from the function rao.diversity from the R package SYNCSA (Debastiani and Pillar, 2012). Rao's quadratic entropy is a phylogenetic equivalent to Simpson's diversity, giving the mean phylogenetic distance between any two randomly chosen individuals (in this case OTUs) in the community, taking into account their relative abundance (Rao, 1982). Phylogenetic distance between the invader and community was calculated as follows. First, the cophenetic function of the picante package (Kembel et al., 2010) was used to calculate the pairwise distances between all OTUs in the master phylogenetic tree and each of the invaders, resulting in two vectors of the phylogenetic distance (number of branch lengths; number of inferred substitutions per 16S gene) between each invader and all OTUs. Second, these vectors were used to calculate an abundance weighted invader-community phylogenetic distance metric using the weighted.mean function (stats package; R Core Team, 2020) to calculate the phylogenetic distance between the invader and each OTU, weighted by the relative abundance of that OTU in each community.

Dimensionality reduction approaches tested

Given the difficulty of analysing the effect of 581 OTUs with only 680 samples, we opted to reduce the dimensionality of the starting composition data before proceeding with our main analyses, testing seven approaches.

For the first five approaches, we simply aggregated the abundances of the OTUs at each of five taxonomic ranks (360 genera, 177 families, 90 orders, 42 classes, 18phyla), according to the species-level taxonomic assignment of the OTU. This approach was intended to account for the fact that higher-level taxonomic groupings may drive compositional effects (e.g. abundance of a clade with a particular functional role) and relatedly, take account for the fact that species-level resolution for OTUs is unreliable (Janda and Abbott, 2007; Johnson et al., 2019; Knight et al., 2018).

For the sixth approach, we performed a principal coordinates analysis of the distance matrix of the Jensen-Shannon divergence (Endres & Schindelin, 2003) of OTU abundances using the 'dudi.pcoa' function from the ade4 package (Chessel et al., 2004), selecting the first 5 coordinates by inspecting the scree plot.

For the seventh and final approach, we calculated a network representing significant correlations between OTU abundances (Pascual-García and Bell, 2020b) estimated with SparCC (Friedman and Alm, 2012). We inferred significant correlations between species abundances with SparCC (Friedman and Alm, 2012), considering only significant links (pseudo p-value<0.01, absolute value of the correlation coefficient > 0.2) leading to a network whose links represent either segregations (negative correlations) or aggregations (positive correlations) between species. Note that these links do not necessarily reflect ecological interactions, since correlations may be caused by bacteria's environmental preferences (Pascual-García et al., 2014). We next partitioned the network into functional groups using functionInk (Pascual-García and Bell, 2020b). Briefly, functionInk is a community-detection method that classifies the nodes in the network clustering together those nodes sharing approximately the same number and type of links (aggregations and segregations in our case) with respect to the same neighbours. Since the nodes have the same links with the same neighbours, we say they have the same functional role in the network, hence the name of functional groups. In previous work, we observed that a classification of the samples according to their beta-diversity distance, led to the identification of six community-classes, and that these classes had distinct functional and metagenomic signatures (Pascual-García and Bell, 2020a). To identify these functional groups (Supplementary Table 1), we run functionInk on the network finding an optimal partition at the maximum of the total partition density (Pascual-García and Bell, 2020b) with 63 functional groups. For the final analysis we selected groups that had three or more members, plus a group of functional groups of *Paenibacillus* species that had less than three members. This was done on the basis of the high abundance of *Paenibacillus borealis* across the communities, and the high connectivity of *Paenibacillus* OTUs.

Selection of a dimensionality reduction approach

To select the most appropriate dimensionality reduction approach for our analysis, we compared their explanatory power for invader survival using random forest regressions. To do this, before our main analyses, we computed 7 sets of random forest regressions using only these dimensionality reductions of starting composition as the explanatory variables, and *P. fluorescens* and *P. putida* invader survival at 24, 96 and 168 hours (1, 3 and 7 days since invasion) as the response variable (Supplementary Figure 1). In total there were therefore 21 random forests compared (7 dimensionality reduction approaches x 2 invaders x 3 sampling points). We compared the dimensionality-reduction techniques by plotting the total variance explained against the number of variables

permuted in each of the 21 random forests (Supplementary Figure 2). This revealed functional groups approach (Pascual-García and Bell, 2020b) to be the most appropriate dimensionality reduction of composition to explain invasion, offering the best compromise in terms of variance explained vs. number of variables permuted. We therefore proceeded with using the abundances of the 19+1 functional groups as explanatory variables in the main analysis random forests, adding the phenotypic explanatory variables alongside them to explain invader survival.

Statistical techniques

All statistical analyses were performed in the R programming language (R Core Team, 2017). Before carrying out analyses, we visually inspected plots of the two-way relationships between invader survival and each of the explanatory variables under various transformations to identify appropriate transformations. The response variable, invader survival, was $\log_{10}$ transformed after adding a pseudocount (+1) to each of the density/ml estimates, in order to deal with zero values. All the explanatory variables except for the three diversity indices (Simpson's diversity and phylogenetic distance from the invader) were $\log_{10}$ transformed.

Random forest regressions

We used random forest regressions implemented using the 'randomForest' package (Liaw and Wiener, 2002) to find the most important individual explanatory variables for invader survival (see main text) among many. The main advantage of the random forest approach for this was that it is able to deal well with overfitting associated with having many explanatory variables, by randomly shuffling the explanatory variables (and samples) permuted in each constituent regression tree and calculating their mean importance across the 'forest'. It also makes few assumptions about the relationships between dependent and independent variables than parametric approaches, accounting somewhat for non-independence and non-linearity, for example.

Three types of random forest models were built for each of the 6 sub-experiments (2 invaders x 3 time points). For each community, starting composition was computed using the selected dimensionality reduction approach i.e. the mean sequence abundance (across species in the group) in each of the 19 functional groups (+ 1 variable quantifying the total number of reads in the sample). Additionally, the 14 phenotypic measurements of the communities were included as explanatory variables, totalling 37 explanatory variables. We used the variable importance values from the random forests to narrow down the 37 explanatory variables to the best predictors of *P. fluorescens* and *P. putida* invader survival.

Structural equation models

The random forest approach was useful for identifying the key explanatory variables for invader survival in our system, but as a black box approach, it is difficult to use to understand the shape of the relationships between explanatory and response variables, and assess fit. In our second analysis therefore, we sought to estimate the putative chain of causation between composition, productivity and invasion resistance using structural equation modelling (SEMs). This approach allowed us to test three scenarios to estimate the extent by which the productivity mediated the effect of composition on invasion (see main text). SEMs were implemented using the lavaan package (Rosseel, 2012).

The chosen explanatory variables (the abundance of the 19 functional groups + 1 total abundance, Simpson's diversity, and 2 measures of resident community cell density at 7 and 14 days) were entered into these models to explain the 6 invader survival end-points (response variables), after the same $\log_{10}$-transformations applied previously (ensuring the variances were approximately equal/within two orders of magnitude). The three structural equation models described in Results were then built and analysed with the R package 'lavaan' (Rosseel, 2012). Although the fit of the three models could be improved specifying residual correlations guided by the computation of modification indexes (Lin et al., 2017), we omitted this step. This optimisation should be done for each model independently and it is unclear how the process could be performed in parallel in a way in which a fair comparison between the final (optimized) models is guaranteed. Hence, we compared the three basic models in terms of their AIC values, and an ANOVA test was performed to evaluate if the reduction in degrees of freedom of the partial mediation model with respect to the complete mediation model implied a significant reduction in the AIC value (Barrett, 2007). We also verified that other metrics such as the Comparative Fit Index were also improved (Bentler, 1990). SEMs were visualised using base functions (R Core Team, 2020), the package semPlot (Epskamp, 2015) and the Arrows function from 'shape' (Soetaert, 2020).

We conducted sensitivity analyses as part of the SEM analysis, in order to validate that the results of the main analysis on the raw data were not an artefact of random noise below the detection threshold of the luminescence assay (which was converted into CFU/ml units). For this analysis, we randomly shuffled the under detection limit invader survival values 999 by sampling the CFU-converted under detection limit data without replacement using the 'sample' function (base package). To assess sensitivity, we examined the average model fit parameters for the Complete, Partial and No Mediation models fit to the 999 generated datasets.

When values below the detection limit were randomly shuffled, the model comparison very marginally supported the Complete (mean ΔAIC = 0.97) over the Partial mediation model (mean ΔAIC = 0.02). However, the AIC values of all three models also increased by 1383.71

~1400 (i.e. fit decreased) - indicating that at least some the values below the detection limit were informative in the original models. Despite the sensitivity of the result therefore, the nature of the change was consistent with the idea that productivity-independent effects of composition played a role in invasion resistance alongside effects of productivity.

Other R packages used

Other R packages used include 'rstudioapi' for setting working directory (Ushey et al., 2020), 'tibble', 'stringr' and 'plyr' for data wrangling (Müller and Wickham, 2020; Wickham, 2019, 2011), 'plotrix' for standard error calculation (Lemon, 2006), 'caret for extracting variable importance values from random forests (Kuhn, 2020), RColorBrewer for figure production (Neuwirth, 2014), ape for phylogenetic tree wrangling (Paradis and Schliep, 2019) and 'grateful' for generating a bibliography of the packages used (Rodríguez-Sánchez and Hutchins, 2020).

Acknowledgments

The funders of this research were the European Research Council (ERC StG 311399-Redundancy), and APG was also funded by the Simons Collaboration: Principles of Microbial Ecosystems (PriME, award number 542381). Thank you to the funders and the EU more generally for making this PhD and research possible. Thank you to members of the Bell, Barraclough and Raymond Labs for support throughout this project. Thank you to Professors Tim Barraclough and Jonathan Jeschke, who provided useful comments to improve this work as PhD viva examiners. Finally, thank you to the manuscript's reviewers, who provided very helpful feedback to significantly improve this manuscript.

Data availability

Data and R code is available via Dryad (XXXX) and Github (https://github.com/befriendabacterium/communityinvasion).

References

Amalfitano, S., Coci, M., Corno, G., Luna, G.M., 2015. A microbial perspective on biological invasions in aquatic ecosystems. Hydrobiologia 746, 13–22. https://doi.org/10.1007/s10750-014-2002-6

Amor, D.R., Ratzke, C., Gore, J., 2020. Transient invaders can induce shifts between alternative stable states of microbial communities. Sci. Adv. 6, eaay8676. https://doi.org/10.1126/sciadv.aay8676

Aneja, M.K., Sharma, S., Fleischmann, F., Stich, S., Heller, W., Bahnweg, G., Munch, J.C., Schloter, M., 2006. Microbial Colonization of Beech and Spruce Litter—Influence of Decomposition Site and Plant Litter Species on the Diversity of Microbial Community. Microb. Ecol. 52, 127–135. https://doi.org/10.1007/s00248-006-9006-3

Barrett, P., 2007. Structural equation modelling: Adjudging model fit. Personal. Individ. Differ. 42, 815–824. https://doi.org/10.1016/j.paid.2006.09.018

Bell, T., 2019. Next-generation experiments linking community structure and ecosystem functioning. Environ. Microbiol. Rep. 11, 20–22. https://doi.org/10.1111/1758-2229.12711

Bell, T., Ager, D., Song, J.-I., Newman, J.A., Thompson, I.P., Lilley, A.K., van der Gast, C.J., 2005a. Larger Islands House More Bacterial Taxa. Science 308, 1884–1884. https://doi.org/10.1126/science.1111318

Bell, T., Bonsall, M.B., Buckling, A., Whiteley, A.S., Goodall, T., Griffiths, R.I., 2010. Protists have divergent effects on bacterial diversity along a productivity gradient. Biol. Lett. 6, 639–642. https://doi.org/10.1098/rsbl.2010.0027

Bell, T., Newman, J.A., Silverman, B.W., Turner, S.L., Lilley, A.K., 2005b. The contribution of species richness and composition to bacterial services. Nature 436, 1157–1160. https://doi.org/10.1038/nature03891

Bentler, P., 1990. Comparative fit indexes in structural models. Psychol. Bull. 107 2, 238–46. https://doi.org/10.1037/0033-2909.107.2.238

Bonanomi, G., D'Ascoli, R., Scotti, R., Gaglione, S.A., Caceres, M.G., Sultana, S., Scelza, R., Rao, M.A., Zoina, A., 2014. Soil quality recovery and crop yield enhancement by combined application of compost and wood to vegetables grown under plastic tunnels. Agric. Ecosyst. Environ. 192, 1–7. https://doi.org/10.1016/j.agee.2014.03.029

Brown, S.P., Le Chat, L., Taddei, F., 2007. Evolution of virulence: triggering host inflammation allows invading pathogens to exclude competitors. Ecol. Lett. 0, 071129084308001-??? https://doi.org/10.1111/j.1461-0248.2007.01125.x

Chessel, D., Dufour, A.-B., Thioulouse, J., 2004. The ade4 Package – I: One-Table Methods. R News 4, 5–10.

Close, D., Xu, T., Smartt, A., Rogers, A., Crossley, R., Price, S., Ripp, S., Sayler, G., 2012. The Evolution of the Bacterial Luciferase Gene Cassette (lux) as a Real-Time Bioreporter. Sensors 12, 732–752. https://doi.org/10.3390/s120100732

Coates, J., Park, B.R., Le, D., Şimşek, E., Chaudhry, W., Kim, M., 2018. Antibiotic-induced population fluctuations and stochastic clearance of bacteria. eLife 7, e32976. https://doi.org/10.7554/eLife.32976

Craney, A., Ahmed, S., Nodwell, J., 2013. Towards a new science of secondary metabolism. J. Antibiot. (Tokyo) 66, 387–400. https://doi.org/10.1038/ja.2013.25

Crawley, Brown, Heard, Edwards, 1999. Invasion-resistance in experimental grassland communities: species richness or species identity? Ecol. Lett. 2, 140–148. https://doi.org/10.1046/j.1461-0248.1999.00056.x

Darwin, C., 1809-1882, 1859. On the origin of species by means of natural selection, or preservation of favoured races in the struggle for life. London : John Murray, 1859.

De Roy, K., Marzorati, M., Negroni, A., Thas, O., Balloi, A., Fava, F., Verstraete, W., Daffonchio, D., Boon, N., 2013. Environmental conditions and community evenness determine the outcome of biological invasion. Nat. Commun. 4, 1383. https://doi.org/10.1038/ncomms2392

Debastiani, V., Pillar, V., 2012. SYNCSA - R tool for analysis of metacommunities based on functional traits and phylogeny of the community components. Bioinformatics 28, 2067–2068.

Eckert, E.M., Quero, G.M., Cesare, A.D., Manfredini, G., Mapelli, F., Borin, S., Fontaneto, D., Luna, G.M., Corno, G., 2019. Antibiotic disturbance affects aquatic microbial community composition and food web interactions but not community resilience. Mol. Ecol. 28, 1170–1182. https://doi.org/10.1111/mec.15033

Eisenhauer, N., Scheu, S., Jousset, A., 2012. Bacterial Diversity Stabilizes Community Productivity. PLOS ONE 7, e34517. https://doi.org/10.1371/journal.pone.0034517

Eisenhauer, N., Schulz, W., Scheu, S., Jousset, A., 2013. Niche dimensionality links biodiversity and invasibility of microbial communities. Funct. Ecol. 27, 282–288. https://doi.org/10.1111/j.1365-2435.2012.02060.x

Endres, D.M., Schindelin, J.E., 2003. A new metric for probability distributions. IEEE Trans. Inf. Theory 49, 1858–1860. https://doi.org/10.1109/TIT.2003.813506

Epskamp, S., 2015. semPlot: Unified Visualizations of Structural Equation Models. Struct. Equ. Model. Multidiscip. J. 22, 474–483. https://doi.org/10.1080/10705511.2014.937847

Fernandez-Gonzalez, N., Braz, G.H.R., Regueiro, L., Lema, J.M., Carballa, M., 2020. Microbial invasions in sludge anaerobic digesters. Appl. Microbiol. Biotechnol. https://doi.org/10.1007/s00253-020-11009-9

Fiegna, F., Moreno-Letelier, A., Bell, T., Barraclough, T.G., 2015a. Evolution of species interactions determines microbial community productivity in new environments. ISME J. 9, 1235–1245. https://doi.org/10.1038/ismej.2014.215

Fiegna, F., Scheuerl, T., Moreno-Letelier, A., Bell, T., Barraclough, T.G., 2015b. Saturating effects of species diversity on life-history evolution in bacteria. Proc. R. Soc. B Biol. Sci. 282, 20151794. https://doi.org/10.1098/rspb.2015.1794

Ford, S.A., King, K.C., 2020. In Vivo Microbial Coevolution Favors Host Protection and Plastic Downregulation of Immunity. Mol. Biol. Evol. msaa292. https://doi.org/10.1093/molbev/msaa292

Foster, K.R., Bell, T., 2012. Competition, Not Cooperation, Dominates Interactions among Culturable Microbial Species. Curr. Biol. 22, 1845–1850. https://doi.org/10.1016/j.cub.2012.08.005

Friedman, J., Alm, E.J., 2012. Inferring Correlation Networks from Genomic Survey Data. PLOS Comput. Biol. 8, 11.

Frossard, A., Gerull, L., Mutz, M., Gessner, M.O., 2013. Shifts in microbial community structure and function in stream sediments during experimentally simulated riparian succession. FEMS Microbiol. Ecol. 84, 398–410. https://doi.org/10.1111/1574-6941.12072

Ghoul, M., Mitri, S., 2016. The Ecology and Evolution of Microbial Competition. Trends Microbiol. 24, 833–845. https://doi.org/10.1016/j.tim.2016.06.011

Glücksman, E., Bell, T., Griffiths, R.I., Bass, D., 2010. Closely related protist strains have different grazing impacts on natural bacterial communities. Environ. Microbiol. 12, 3105–3113. https://doi.org/10.1111/j.1462-2920.2010.02283.x

Gu, Y., Meng, D., Yang, S., Xiao, N., Li, Z., Liu, Z., Li, L., Zeng, X., Zeng, S., Yin, H., 2019. Invader-resident community similarity contribute to the invasion process and regulate biofertilizer effectiveness. J. Clean. Prod. 241, 118278. https://doi.org/10.1016/j.jclepro.2019.118278

Hodgson, D.J., Rainey, P.B., Buckling, A., 2002. Mechanisms linking diversity, productivity and invasibility in experimental bacterial communities. Proc. R. Soc. Lond. B Biol. Sci. https://doi.org/10.1098/rspb.2002.2146

Holm, E., Jensen, V., 1972. Aerobic Chemoorganotrophic Bacteria of a Danish Beech forest. Microbiology of a Danish Beech Forest I. Oikos 23, 248. https://doi.org/10.2307/3543413

Huston, M.A., 2004. Management strategies for plant invasions: manipulating productivity, disturbance, and competition. Divers. Distrib. 10, 167–178. https://doi.org/10.1111/j.1366-9516.2004.00083.x

Janda, J.M., Abbott, S.L., 2007. 16S rRNA Gene Sequencing for Bacterial Identification in the Diagnostic Laboratory: Pluses, Perils, and Pitfalls. J. Clin. Microbiol. 45, 2761–2764. https://doi.org/10.1128/JCM.01228-07

Jiang, L., Tan, J., Pu, Z., 2010. An Experimental Test of Darwin's Naturalization Hypothesis. Am. Nat. 175, 415–423. https://doi.org/10.1086/650720

Johanesen, P.A., Mackin, K.E., Hutton, M.L., Awad, M.M., Larcombe, S., Amy, J.M., Lyras, D., 2015. Disruption of the Gut Microbiome: Clostridium difficile Infection and the Threat of Antibiotic Resistance. Genes 6, 1347–1360. https://doi.org/10.3390/genes6041347

Johnson, J.S., Spakowicz, D.J., Hong, B.-Y., Petersen, L.M., Demkowicz, P., Chen, L., Leopold, S.R., Hanson, B.M., Agresta, H.O., Gerstein, M., Sodergren, E., Weinstock, G.M., 2019. Evaluation of 16S rRNA gene sequencing for species and strain-level microbiome analysis. Nat. Commun. 10, 5029. https://doi.org/10.1038/s41467-019-13036-1

Jones, M.L., Ramoneda, J., Rivett, D.W., Bell, T., 2017. Biotic resistance shapes the influence of propagule pressure on invasion success in bacterial communities. Ecology 98, 1743–1749. https://doi.org/10.1002/ecy.1852

Jousset, A., Schulz, W., Scheu, S., Eisenhauer, N., 2011. Intraspecific genotypic richness and relatedness predict the invasibility of microbial communities. ISME J. 5, 1108–1114. https://doi.org/10.1038/ismej.2011.9

Kembel, S.W., Cowan, P.D., Helmus, M.R., Cornwell, W.K., Morlon, H., Ackerly, D.D., Blomberg, S.P., Webb, C.O., 2010. Picante: R tools for integrating phylogenies and ecology. Bioinformatics 26, 1463–1464.

Kinnunen, M., Dechesne, A., Albrechtsen, H.-J., Smets, B.F., 2018. Stochastic processes govern invasion success in microbial communities when the invader is phylogenetically close to resident bacteria. ISME J. 12, 2748–2756. https://doi.org/10.1038/s41396-018-0202-1

Kinnunen, M., Dechesne, A., Proctor, C., Hammes, F., Johnson, D., Quintela-Baluja, M., Graham, D., Daffonchio, D., Fodelianakis, S., Hahn, N., Boon, N., Smets, B.F., 2016. A conceptual framework for invasion in microbial communities. ISME J. 10, 2773–2779. https://doi.org/10.1038/ismej.2016.75

Knight, R., Vrbanac, A., Taylor, B.C., Aksenov, A., Callewaert, C., Debelius, J., Gonzalez, A., Kosciolek, T., McCall, L.-I., McDonald, D., Melnik, A.V., Morton, J.T., Navas, J., Quinn, R.A., Sanders, J.G., Swafford, A.D., Thompson, L.R., Tripathi, A., Xu, Z.Z., Zaneveld, J.R., Zhu, Q., Caporaso, J.G., Dorrestein, P.C., 2018. Best practices for analysing microbiomes. Nat. Rev. Microbiol. 16, 410–422. https://doi.org/10.1038/s41579-018-0029-9

Kuhn, M., 2020. caret: Classification and Regression Training.

Lawrence, D., Fiegna, F., Behrends, V., Bundy, J.G., Phillimore, A.B., Bell, T., Barraclough, T.G., 2012. Species Interactions Alter Evolutionary Responses to a Novel Environment. PLOS Biol. 10, e1001330. https://doi.org/10.1371/journal.pbio.1001330

Lear, L., Hesse, E., Shea, K., Buckling, A., 2020. Disentangling the mechanisms underpinning disturbance-mediated invasion. Proc. R. Soc. B Biol. Sci. 287, 20192415. https://doi.org/10.1098/rspb.2019.2415

Lemon, J., 2006. Plotrix: a package in the red light district of R. R-News 6, 8–12.

Li, S., Tan, J., Yang, X., Ma, C., Jiang, L., 2019. Niche and fitness differences determine invasion success and impact in laboratory bacterial communities. ISME J. 13, 402–412. https://doi.org/10.1038/s41396-018-0283-x

Liaw, A., Wiener, M., 2002. Classification and Regression by randomForest 2, 5.

Lin, L.-C., Huang, P.-H., Weng, L.-J., 2017. Selecting Path Models in SEM: A Comparison of Model Selection Criteria. Struct. Equ. Model. Multidiscip. J. 24, 855–869. https://doi.org/10.1080/10705511.2017.1363652

Litchman, E., 2010. Invisible invaders: non-pathogenic invasive microbes in aquatic and terrestrial ecosystems. Ecol. Lett. 13, 1560–1572. https://doi.org/10.1111/j.1461-0248.2010.01544.x

Mallon, C.A., Elsas, J.D. van, Salles, J.F., 2015a. Microbial Invasions: The Process, Patterns, and Mechanisms. Trends Microbiol. 23, 719–729. https://doi.org/10.1016/j.tim.2015.07.013

Mallon, C.A., Le Roux, X., van Doorn, G.S., Dini-Andreote, F., Poly, F., Salles, J.F., 2018. The impact of failure: unsuccessful bacterial invasions steer the soil microbial community away from the invader's niche. ISME J. 12, 728–741. https://doi.org/10.1038/s41396-017-0003-y

Mallon, C.A., Poly, F., Le Roux, X., Marring, I., van Elsas, J.D., Salles, J.F., 2015b. Resource pulses can alleviate the biodiversity-invasion relationship in soil microbial communities. Ecology 96, 915–926. https://doi.org/10.1890/14-1001.1

Mazumdar, V., Amar, S., Segre, D., 2013. Metabolic proximity in the order of colonization of a microbial community. PLoS ONE Electron. Resour. 8, e77617. https://dx.doi.org/10.1371/journal.pone.0077617

Müller, K., Wickham, H., 2020. tibble: Simple Data Frames.

Neuwirth, E., 2014. RColorBrewer: ColorBrewer Palettes.

Nikel, P.I., Martínez-García, E., Lorenzo, V. de, 2014. Biotechnological domestication of pseudomonads using synthetic biology. https://doi.org/10.1038/nrmicro3253

Oksanen, J., Blanchet, F.G., Friendly, M., Kindt, R., Legendre, P., McGlinn, D., Minchin, P.R., O'Hara, R.B., Simpson, G.L., Solymos, P., Stevens, M.H.H., Szoecs, E., Wagner, H., 2019. vegan: Community Ecology Package.

Paradis, E., Schliep, K., 2019. ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. Bioinformatics 35, 526–528.

Pascual-García, A., Bell, T., 2020a. Community-level signatures of ecological succession in natural bacterial communities. Nat. Commun. 11, 2386. https://doi.org/10.1038/s41467-020-16011-3

Pascual-García, A., Bell, T., 2020b. functionInk: An efficient method to detect functional groups in multidimensional networks reveals the hidden structure of ecological communities. Methods Ecol. Evol. 11, 804–817. https://doi.org/10.1111/2041-210X.13377

Pascual-García, A., Bonhoeffer, S., Bell, T., 2020. Metabolically cohesive microbial consortia and ecosystem functioning. Philos. Trans. R. Soc. B Biol. Sci. 375, 20190245. https://doi.org/10.1098/rstb.2019.0245

Pascual-García, A., Tamames, J., Bastolla, U., 2014. Bacteria dialog with Santa Rosalia: Are aggregations of cosmopolitan bacteria mainly explained by habitat filtering or by ecological interactions? BMC Microbiol. 14. https://doi.org/10.1186/s12866-014-0284-5

R Core Team, 2020. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.

R Core Team, 2017. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.

Rajendhran, J., Gunasekaran, P., 2011. Microbial phylogeny and diversity: Small subunit ribosomal RNA sequence analysis and beyond. Microbiol. Res. 166, 99–110. https://doi.org/10.1016/j.micres.2010.02.003

Rao, C.R., 1982. Diversity and dissimilarity coefficients: A unified approach. Theor. Popul. Biol. 21, 24–43. https://doi.org/10.1016/0040-5809(82)90004-1

Remus-Emsermann, M.N.P., Tecon, R., Kowalchuk, G.A., Leveau, J.H.J., 2012. Variation in local carrying capacity and the individual fate of bacterial colonizers in the phyllosphere. ISME J. 6, 756–765. https://doi.org/10.1038/ismej.2011.209

Rillig, M.C., Antonovics, J., Caruso, T., Lehmann, A., Powell, J.R., Veresoglou, S.D., Verbruggen, E., 2015. Interchange of entire communities: microbial community coalescence. Trends Ecol. Evol. 30, 470–476. https://doi.org/10.1016/j.tree.2015.06.004

Rivett, D.W., Bell, T., 2018. Abundance determines the functional role of bacterial phylotypes in complex communities. Nat. Microbiol. 3, 767–772. https://doi.org/10.1038/s41564-018-0180-0

Rivett, D.W., Scheuerl, T., Culbert, C.T., Mombrikotb, S.B., Johnstone, E., Barraclough, T.G., Bell, T., 2016. Resource-dependent attenuation of species interactions during bacterial succession. ISME J. 10, 2259–2268. https://doi.org/10.1038/ismej.2016.11

Rodríguez-Sánchez, F., Hutchins, S.D., 2020. grateful: Facilitate citation of R packages.

Rosseel, Y., 2012. lavaan: An R Package for Structural Equation Modeling. J. Stat. Softw. 48, 1–36.

Scheuerl, T., Hopkins, M., Nowell, R.W., Rivett, D.W., Barraclough, T.G., Bell, T., 2020. Bacterial adaptation is constrained in complex communities. Nat. Commun. 11. https://doi.org/10.1038/s41467-020-14570-z

Soetaert, K., 2020. shape: Functions for Plotting Graphical Shapes, Colors.

Stubbendieck, R.M., Vargas-Bautista, C., Straight, P.D., 2016. Bacterial Communities: Interactions to Scale. Front. Microbiol. 7. https://doi.org/10.3389/fmicb.2016.01234

Tenaillon, O., Skurnik, D., Picard, B., Denamur, E., 2010. The population genetics of commensal Escherichia coli. Nat. Rev. Microbiol. 8, 207–217. https://doi.org/10.1038/nrmicro2298

Thakur, M.P., van der Putten, W.H., Cobben, M.M.P., van Kleunen, M., Geisen, S., 2019. Microbial invasions in terrestrial ecosystems. Nat. Rev. Microbiol. 17, 621–631. https://doi.org/10.1038/s41579-019-0236-z

Ushey, K., Allaire, J.J., Wickham, H., Ritchie, G., 2020. rstudioapi: Safely Access the RStudio API.

van Elsas, J.D., Chiurazzi, M., Mallon, C.A., Elhottova, D., Kristufek, V., Salles, J.F., 2012. Microbial diversity determines the invasion of soil by a bacterial pathogen. Proc. Natl. Acad. Sci. 109, 1159–1164. https://doi.org/10.1073/pnas.1109326109

Walker, E.D., Lawson, D.L., Merritt, R.W., Morgan, W.T., Klug, M.J., 1991. Nutrient Dynamics, Bacterial Populations, and Mosquito Productivity in Tree Hole Ecosystems and Microcosms. Ecology 72, 1529–1546. https://doi.org/10.2307/1940953

Watrous, J.D., Phelan, V.V., Hsu, C.-C., Moree, W.J., Duggan, B.M., Alexandrov, T., Dorrestein, P.C., 2013. Microbial metabolic exchange in 3D. ISME J. 7, 770–780. https://doi.org/10.1038/ismej.2012.155

Wei, Z., Hu, J., Gu, Y., Yin, S., Xu, Y., Jousset, A., Shen, Q., Friman, V.-P., 2018. Ralstonia solanacearum pathogen disrupts bacterial rhizosphere microbiome during an invasion. Soil Biol. Biochem. 118, 8–17. https://doi.org/10.1016/j.soilbio.2017.11.012

Wickham, H., 2019. stringr: Simple, Consistent Wrappers for Common String Operations.

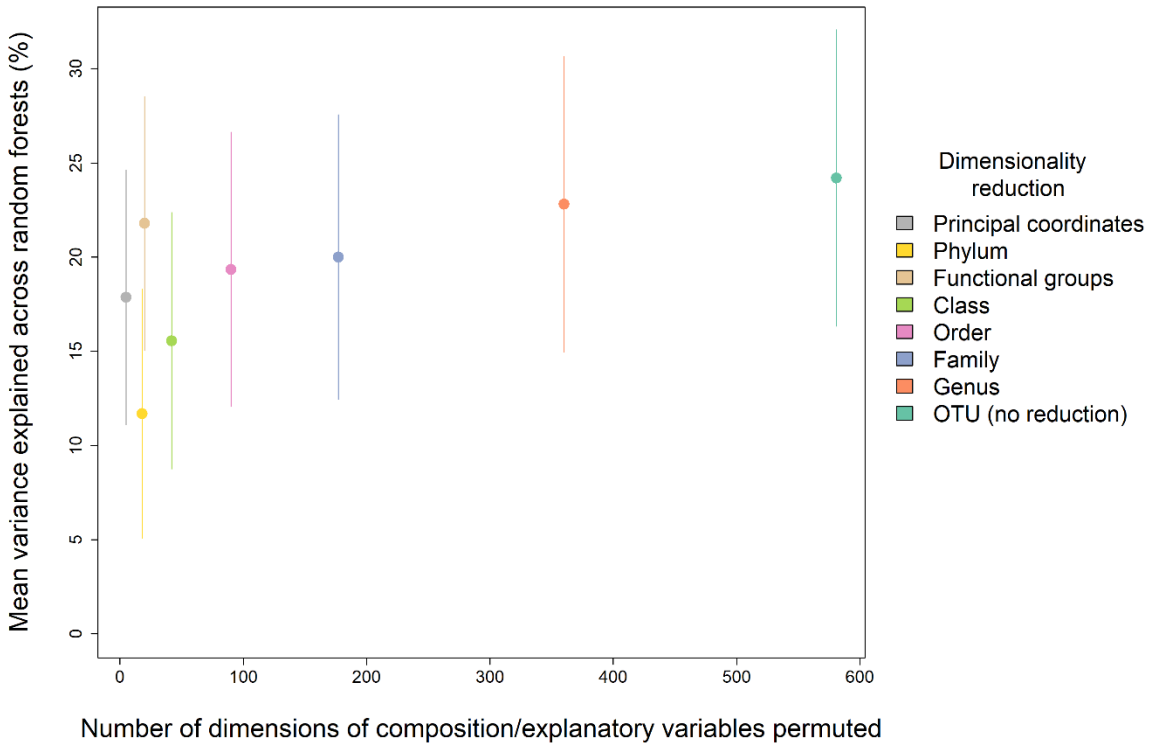
Wickham, H., 2011. The Split-Apply-Combine Strategy for Data Analysis. J. Stat. Softw. 40, 1–29.

Woodcock, S., Head, I.M., Sloan, W.T., 2007. Neutral assembly of bacterial communities. FEMS Microbiol Ecol 10.

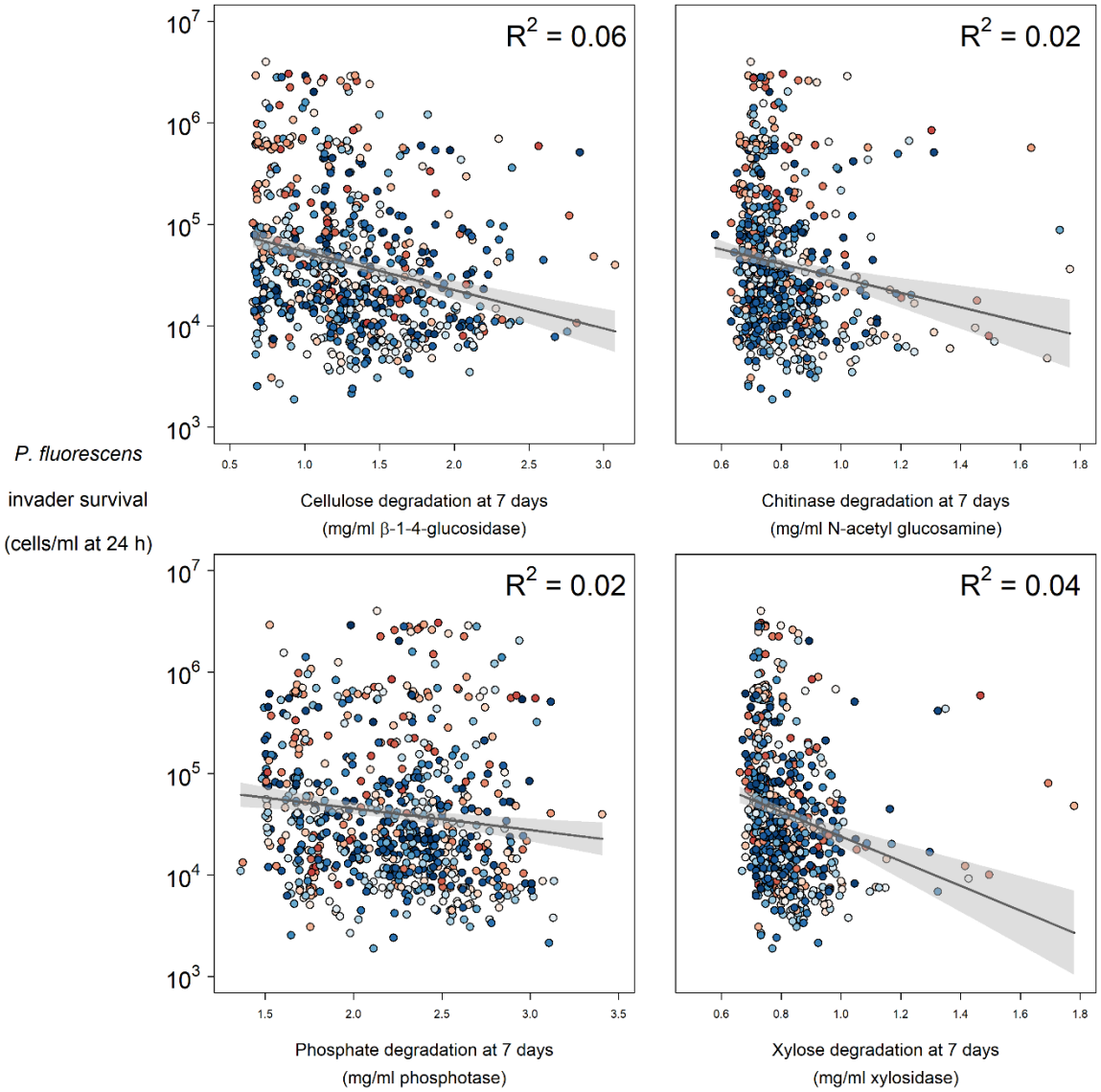
Yang, T., Han, G., Yang, Q., Friman, V.-P., Gu, S., Wei, Z., Kowalchuk, G.A., Xu, Y., Shen, Q., Jousset, A., 2018. Resource stoichiometry shapes community invasion resistance via productivity-mediated species identity effects. Proc. R. Soc. B Biol. Sci. 285, 20182035–20182035. https://doi.org/10.1098/rspb.2018.2035

Supplementary Materials

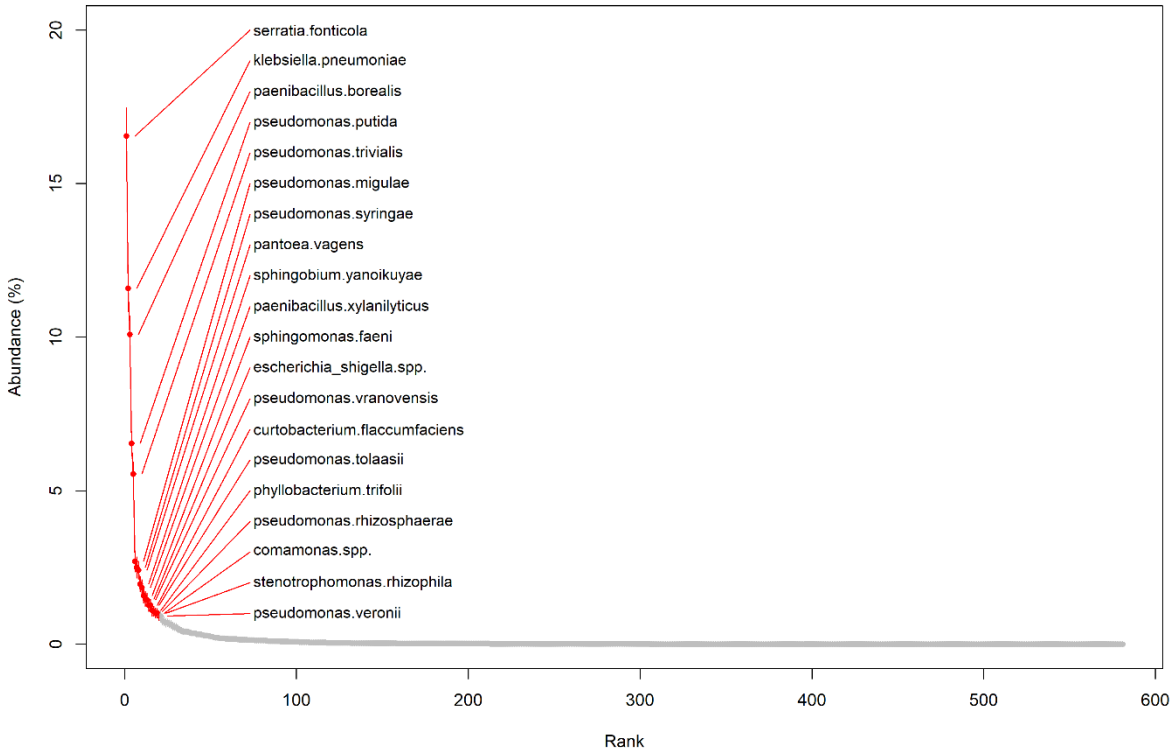
Supplementary Figure 1: Comparison of the performance of different dimensionality reductions of the starting composition data, shown as mean variance explained vs the number of dimensions in each of the tested reductions. The functional groups approach had a disproportionate explanatory power for its number of dimensions, comparable to no dimensionality reduction, and so we opted for this method for our main analyses.



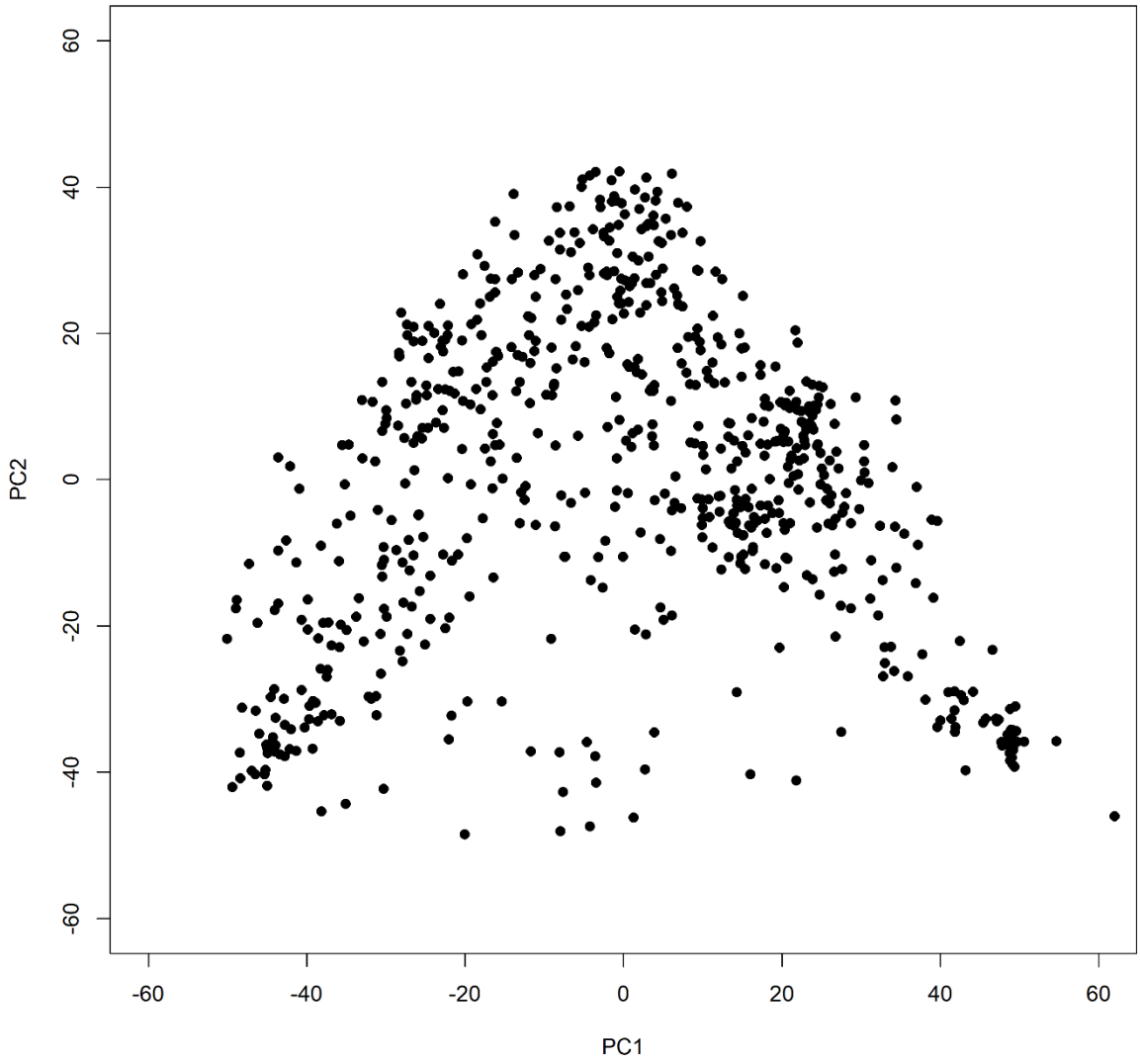
Supplementary Figure 2: Relationships between enzyme activity and *P. fluorescens* invader survival



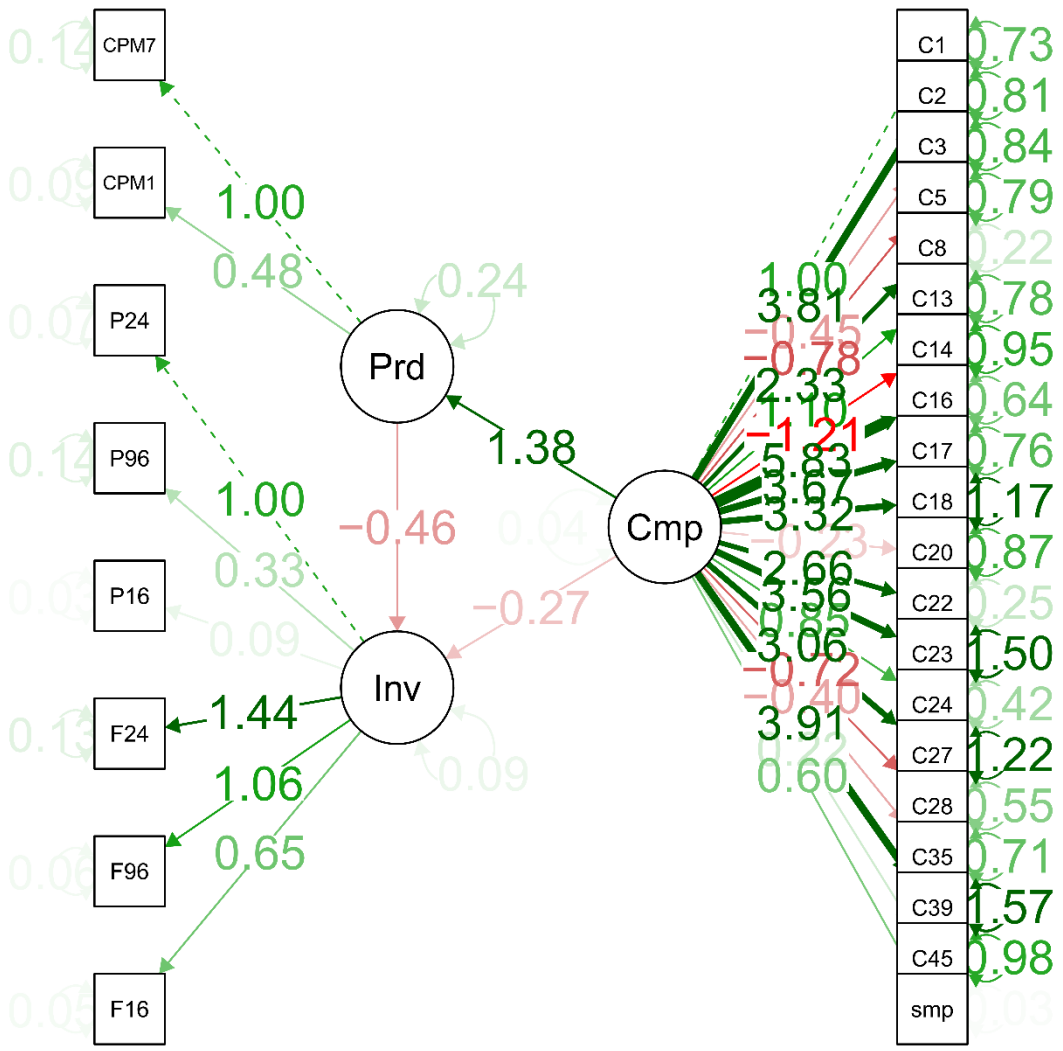
Supplementary Figure 3: Rank abundance plot of OTUs (mean and standard error of each OTU's abundance in all communities)



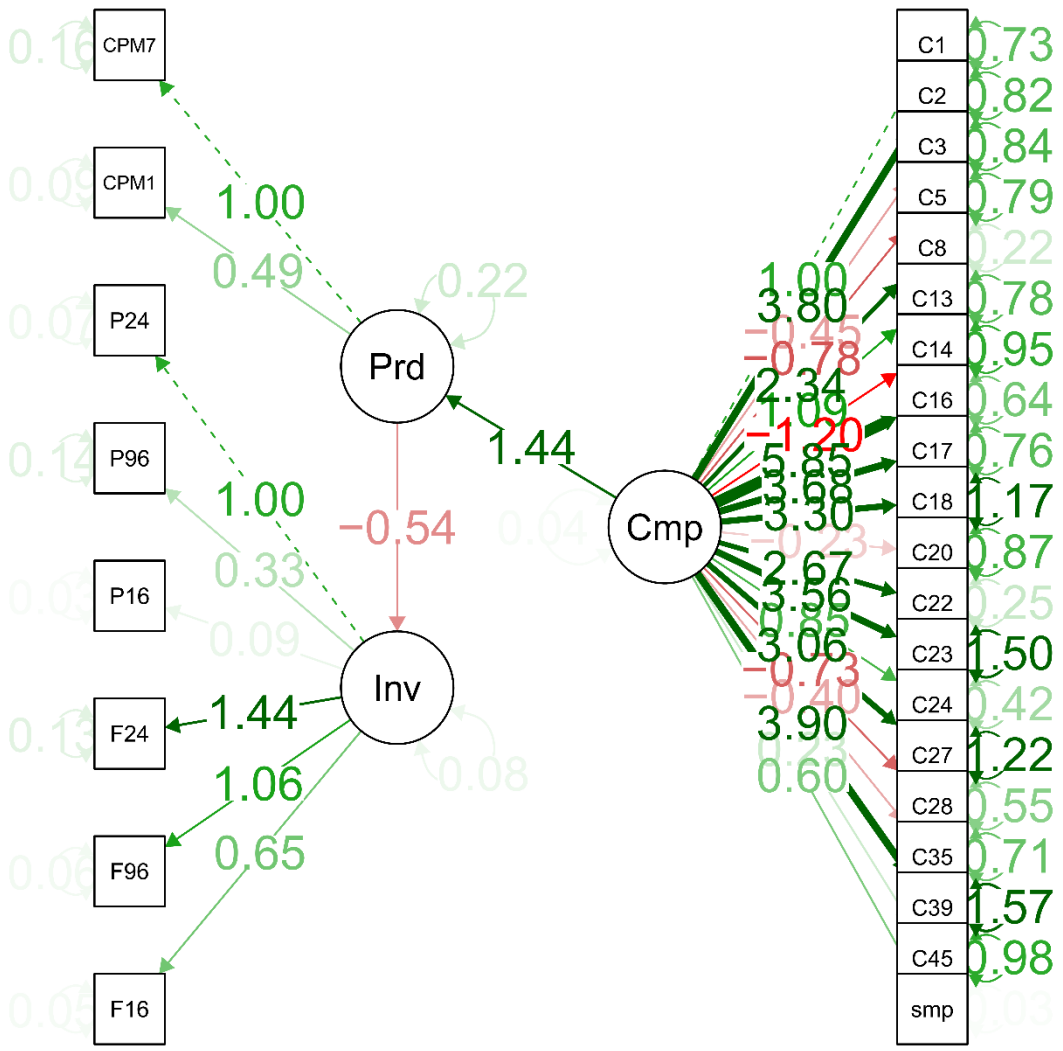
Supplementary Figure 4: PCoA (first two coordinates of five used)



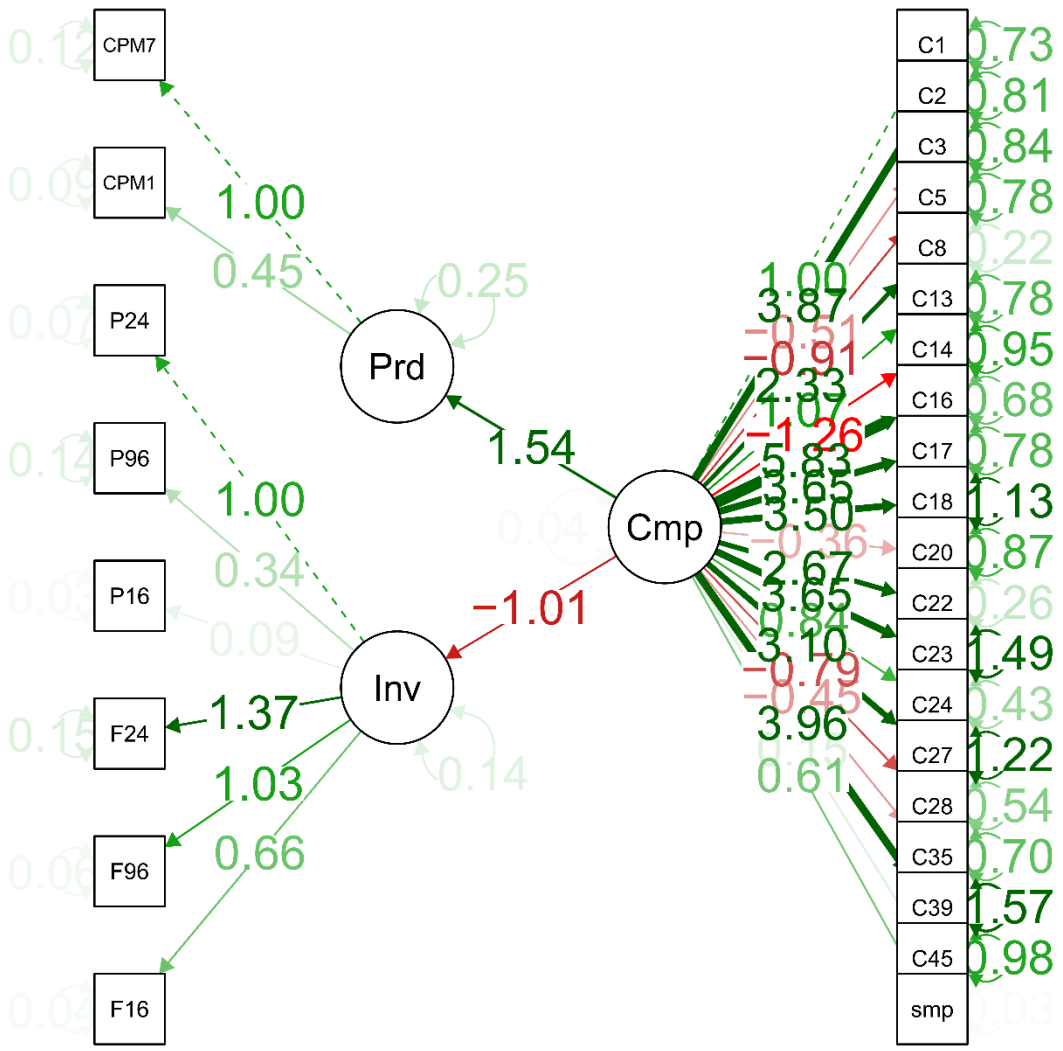
Supplementary Figure 5: Full structure of Partial Mediation structural equation model



Supplementary Figure 6: Full structure of Complete Mediation structural equation model



Supplementary Figure 7: Full structure of No Mediation structural equation model



Supplementary Table 1: Functional group membership

Cluster	OTUs (Nearest species-level designation)
1	***1.*** *Arthrobacter sulfureus* ***2.*** *Arthrobacter protophormiae* ***3.*** *Arthrobacter.spp arthrobacter.sp.*
2	***1.*** *Pseudomonas migulae* ***2.*** *Pseudomonas tolaasii* ***3.*** *Pseudomonas trivialis* ***4.*** *Pseudomonas veronii* ***5.*** *Pseudomonas putida* ***6.*** *Pseudomonas thivervalensis* ***7.*** *Pseudomonas vranovensis* ***8.*** *Pseudomonas rhizosphaerae* ***9.*** *Pseudomonas lutea* ***10.*** *Pseudomonas marincola* ***11.*** *Pseudomonas plecoglossicida* ***12.*** *Comamonas spp.* ***13.*** *Stenotrophomonas rhizophila* ***14.*** *Pseudomonas syringae*
3	***1.*** *Cloacibacterium spp.* ***2.*** *Bacillus thermoamylovorans* ***3.*** *Corynebacterium diphtheriae*
5	***1.*** *Novispirillum itersonii* ***2.*** *Ralstonia pickettii* ***3.*** *Acinetobacter towneri* ***4.*** *Acinetobacter genomospecies 3*
8	***1.*** *Pedobacter aquatilis* ***2.*** *Chryseobacterium gregarium* ***3.*** *Dysgonomonas spp.* ***4.*** *Pedobacter spp.*
13	***1.*** *Hylemonella spp.* ***2.*** *Acidovorax konjaci* ***3.*** *Variovorax spp.*

14	***1.*** *Streptomyces sanglieri* ***2.*** *Streptomyces xanthochromogenes* ***3.*** *Streptomyces viridochromogenes*
16	***1.*** *Novosphingobium subarcticum* ***2.*** *Pedobacter daejeonensis* ***3.*** *Janthinobacterium lividum*
17	***1.*** *Massilia spp.* ***2.*** *Acinetobacter piperi* ***3.*** *Pedobacter cryoconitis*
18	***1.*** *Erwinia rhapontici* ***2.*** *Cedecea spp.* ***3.*** *Trabulsiella spp.* ***4.*** *Pantoea agglomerans* ***5.*** *Citrobacter werkmanii* ***6.*** *Klebsiella pneumoniae* ***7.*** *Escherichia_shigella* ***8.*** *Erwinia persicina* ***9.*** *Pantoea vagens* ***10.*** *Serratia fonticola* ***11.*** *Serratia liquefaciens* ***12.*** *Serratia quinivorans*
20	***1.*** *Leptothrix spp.* ***2.*** *Delftia.lacustris* ***3.*** *Acidovorax spp.* ***4.*** *Aaenimonas spp.* ***5.*** *Stenotrophomonas maltophilia* ***6.*** *Pseudomonas balearica* ***7.*** *Aquabacterium spp.* ***8.*** *Herbaspirillum spp.* ***9.*** *Pseudomonas pseudoalcaligenes* ***10.*** *Brevundimonas aurantiaca* ***11.*** *Phenylobacterium spp.* ***12.*** *Acinetobacter.calcoaceticus*

	13. *Acinetobacter johnsonii* ***14.*** *Massilia timonae* ***15.*** *Paucimonas spp.* ***16.*** *Herbaspirillum rubrisubalbicans*
22	***1.*** *Pedobacter wanjuense* ***2.*** *Epilithonimonas lactis* ***3.*** *Chryseobacterium soldanellicola*
23	***1.*** *Rhizobium leguminosarum* ***2.*** *Rhizobium giardinii* ***3.*** *Rhizobium cellulosilyticus*
24	***1.*** *Haemophilus parainfluenzae* ***2.*** *Neisseria subflava* ***3.*** *Streptococcus salivarius*
27	***1.*** *Brevundimonas variabilis* ***2.*** *Brevundimonas bullata* ***3.*** *Bosea thiooxidans*
28	***1.*** *Corynebacteriales spp.* ***2.*** *Vogesella spp.* ***3.*** *Nocardioides spp.* ***4.*** *Pleomorphomonas spp.*
35	***1.*** *Tepidimonas spp.* ***2.*** *Weeksella spp.* ***3.*** *Pseudoxanthomonas taiwanensis*

39	***1.*** *Rhizobium sullae* ***2.*** *Rhizobium spp.* ***3.*** *Agrobacterium tumefaciens*
45 (combined with clusters 48, 49, 56)	***1.*** *Paenibacillus xylanilyticus* ***2.*** *Paenibacillus favisporus* ***3.*** *Paenibacillus wynnii* ***4.*** *Paenibacillus borealis* ***5.*** *Paenibacillus lautus* ***6.*** *Paenibacillus.spp.* ***7.*** *Paenibacillus chondroitinus*