# Productive bacterial communities exclude invaders

Matt L. Jones[1,2,*], Damian W. Rivett[1,3] Alberto Pascual-García[1,4], and Thomas Bell[1]

[1] Department of Life Sciences. Silwood Park Campus. Imperial College London, Ascot, United Kingdom

[2] University of Exeter Medical School, Penryn Campus, Penryn, Cornwall, UK

[3] School of Science and the Environment, Manchester Metropolitan University, Manchester, UK

[4] Institute of Integrative Biology, ETH Zürich, Zürich, Switzerland

[*] Correspondence: M.L.Jones@exeter.ac.uk

December 18, 2019

## Abstract

Experiments with artificial microbial communities indicate that certain community compositions are more resistant to microbial invasion than others. However, few invasion experiments have used natural microbial communities to investigate how the effect of community composition on invasion resistance relates to different aspects of community growth. We conducted experimental invasions of two bacterial species (*Pseudomonas fluorescens* and *Pseudomonas putida*) into 678 bacterial communities, comparing composition to several aspects of community growth prior to invasion to see how well they predicted invasion. We show that even in these complex microbial assemblages, the effects of resident community composition and growth are largely overlapping - with parameters associated with the productivity of the community (cell yield, community respiration) emerging as the main limiting factors for invasion success. Despite their complexity, these communities can be classified into a few compositional groups that are associated with the main differences in community growth, and thereby invasion resistance.

## Introduction

Microbial communities are challenged by a continuous stream of dispersing cells that may establish populations and displace resident taxa, potentially altering ecosystem functioning [26, 34, 35]. Resident microbial communities can constrain invasions by competing for resources, by producing antimicrobial compounds, or by otherwise altering environmental conditions that are unfavourable to the invader (e.g. through production of inhibitory substances) [21, 36, 37, 49]. Conversely, resident microbial communities might facilitate invasions by producing resources that would otherwise be unavailable, or by favourably altering the environment [22, 36]. The field of microbial invasion ecology has recently emerged with a foundational goal of understanding how the composition of resident microbial communities shapes their resistance to the introduction of new microbes. This would be useful beyond the fundamental science of microbial ecology. For example, with appropriate adjustments to specific contexts [46], such general theory could inform efforts to limit the spread of pathogens and contaminants in agriculture and industry - or conversely, to help increase the establishment success of purportedly probiotic strains in animal and plant hosts [5, 47, 48, 50].

Most studies in microbial invasion ecology have used artificial communities to demonstrate a relationship between resident microbial communities' diversity and their invasion resistance [35]. In such experiments, microcosms containing a sterile culture medium (e.g. lab broth, autoclaved soil) are inoculated with artificial microbial communities with differing levels of diversity, constructed from culturable taxa and/or dilutions of natural communities (typically 90% at each dilutions step). After some period of

growth, these communities are invaded with a new population of microbes, and the relationship between starting diversity and invasion success is determined. There is broad consensus from these experiments that there is a negative relationship between the species richness of the resident community and the growth of invader population [6, 16, 47]. Moreover, diversity appears to enhance invasion resistance because more diverse communities tend to more extensively grow on (and deplete) all the available resources (a 'complementarity effect'), and/or are more likely to contain a species that grows on (and depletes) the same resource as the invader (a 'sampling effect') [12, 22].

In line with these findings, a simpler explanation for invasion resistance is that differences in the growth performance of resident communities drive differences in invaders' success. Communities that grow well are likely to deplete more of the nutrients that would otherwise be available for invaders to exploit - and species with a high growth rate are more likely to be found in more diverse communities, at least across artificially-constructed gradients of diversity. In natural communities with less extreme gradients and much higher levels of diversity, growth performance may even eclipse the effect of composition on invasion resistance.

Here we examined how community composition affects invasion resistance in natural microbial assemblages. We conducted experimental invasions into 678 microbial communities collected from rainwater pools (water-filled beech tree holes). The tree hole micro-ecosystem has been used extensively as a simple natural microcosm [4, 18, 29, 43]. As described previously [43], whole communities were isolated from tree holes, sequenced in order to provide information about diversity and composition, and inoculated into identical standardised environments. Communities were grown for 14 days, during which time we also measured several aspects of community growth: community cell yield (cytometry counts), community activity (adeonisine triphosphate (ATP) and respiration), and the ability to degrade different substrates through the release of exo-enyzmes. The effects of composition and growth on communities' resistance to the invasion of a new bacterial population were then compared. Invasion resistance was experimentally measured for two closely-related, well-studied model invaders: *Pseudomonas fluorescens* SBW25 and *P. putida* KT2440. We used random forests - a machine learning technique - to understand the relative contributions of composition and growth to invasion resistance (see Methods and Materials: Statistical analysis) in our tree hole communities.

# Results

## Relationships between community invasion resistance, composition and diversity

We measured invasion success as the ability of luminescence-marked invaders (*P. fluorescens* and *P. putida*) to survive and grow from stationary phase when inoculated into an established resident microbial community (see Methods and Materials). We first grew both invader strains to carrying capacity (96 h), and then inoculated microcosms containing the communities (grown for 14 days on BLT) with one of the invaders (approximately $10^5$ invader cells per microcosm). We measured short, medium and long-term invasion success by measuring the luminescence of the microcosms at 24, 96 and 168 hours (1, 4 & 7 days) post-invasion. Following previous diversity-invasion resistance experiments, we analysed the relationship between invasion and the diversity and composition of the community that was added to each microcosm at the start of the experiment. Unlike previous experiments using synthetic microbial assemblages, our communities were 678 naturally-occurring bacterial assemblages, and so community diversity and composition were estimated using amplicon sequencing of the 16S rRNA gene - allowing us to estimate a community matrix from Operational Taxonomic Unit (OTU) abundance data (see Methods and Materials).

Unsupervised clustering of the communities according to their $\beta$-diversity with Jensen-Shannon divergence [17] revealed six community types [40]. We represent these classes by projecting the communities on the first two coordinates of a Principle Coordinates Analysis (PCoA) ordination (Figure 1). Two of
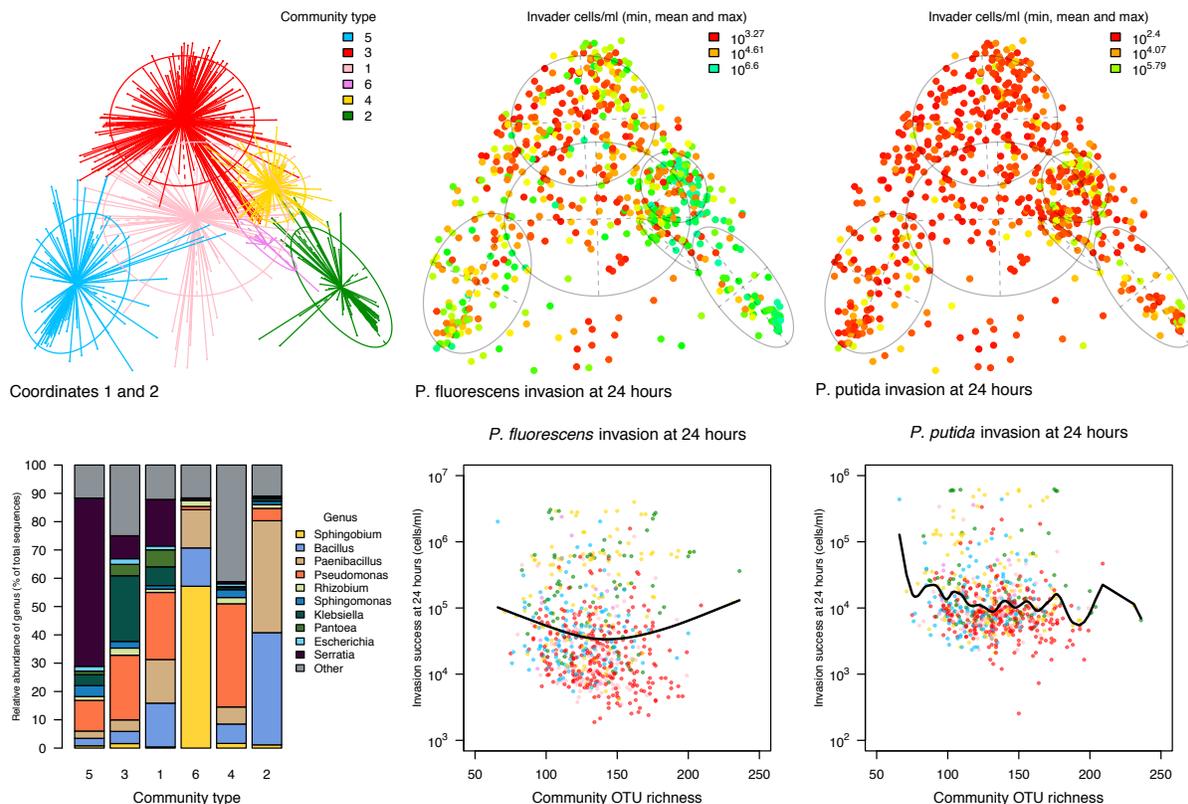
Figure 1: Community composition and invasion success. Top: PCoA of community composition across 678 communities, coloured according to the six communities types identified in [40] (left), *P.fluorescens* invasion success at 24 h (center) and *P.putida* invasion success at 24 h (right). Bottom: Genus-level composition of each of the 6 community types, based on OTUs (left) and relationship between OTU richness and invasion success of *P.fluorescens* (middle) and *P.putida* (right). Smoothed curves (black lines) are fitted with smooth.spline in R.

these classes (types 2 and 4; green and yellow) appeared to drive the invasion patterns observed in our experiments, especially for *P.fluorescens* (Figure 1, top middle and right). Compared to other communities, these two compositional groups were particularly vulnerable to invasion by the two invaders. The most perceptible compositional difference between communities in types 2 and 4 and other communities was that they were dominated by species in the genera Paenibacillus/Bacillus and Pseudomonas, respectively - which represented much smaller proportions of other communities. There was no clear relationship between OTU richness and invasion resistance, and no indication that the community types 2 and 4 had lower levels of OTU richness (Figure 1, bottom row). Qualitatively, community types 5 and 6 appeared to be less even than other types, and when we later considered the relative effects of community OTU evenness (Pielou index) and Community type (PCoA), evenness was found to rank closely behind community type as the second most important compositional driver of invasion success (Figure 4). This supported the conclusion that composition rather than diversity *per se*, drove invasion resistance in our experiments.
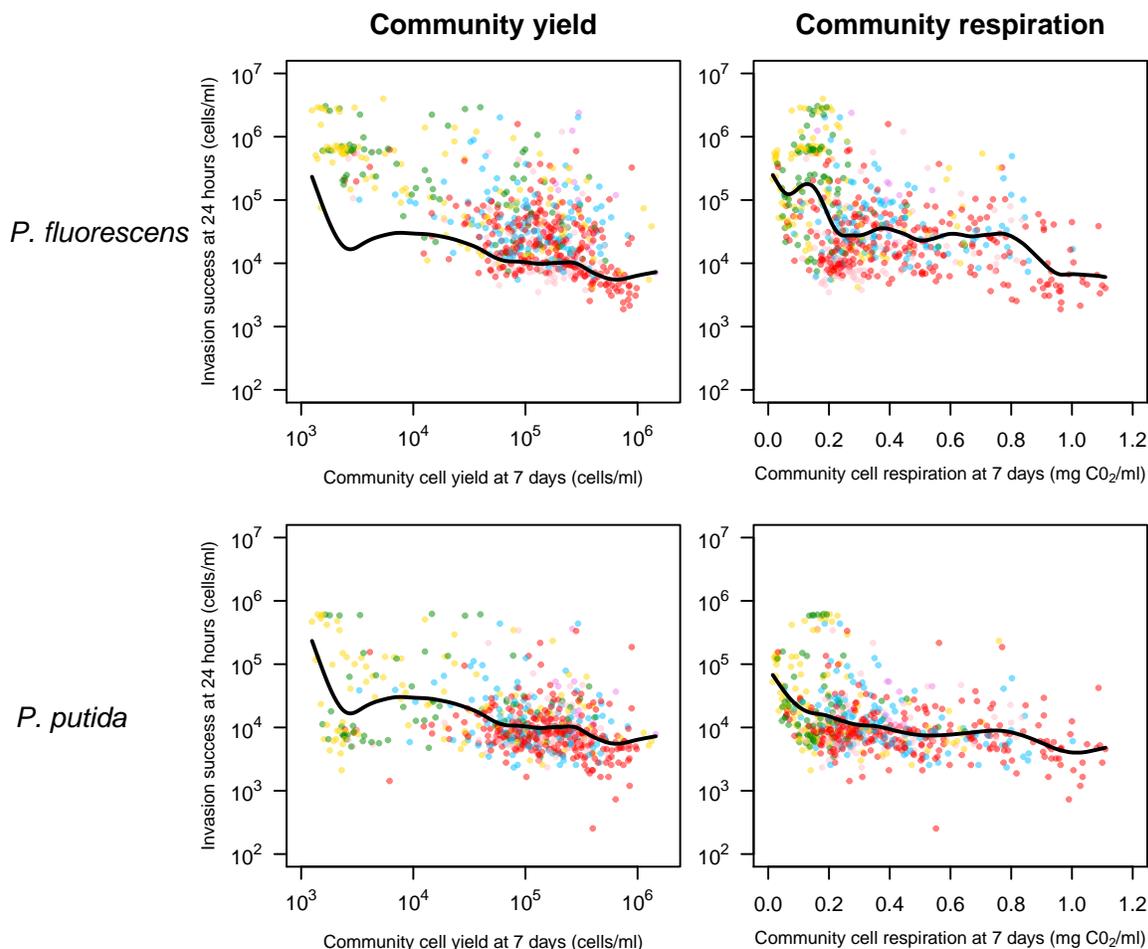
Figure 2: Community growth and invasion success.Relationships between community cell yield at 7 days (top), respiration at 7 days and Simpson index of diversity (bottom) and invasion success of *P. fluorescens (left panels) and P. putida (right)*. Smoothed curves are fitted with *smooth.spline* in R.

## Relationship between community composition, overall growth performance and invasion resistance

As well as starting diversity and composition, we also took measurements related to the growth of the communities before invasions. Growth measures (measured at 7 and 14 days) included those related to the overall growth success of the communities within the microcosms (resident community respiration, cell yield and potential metabolic activity), and the capacity of the communities to metabolise a set of 4 specific substrates that we expected to be important components of the microcosm environment (cellulose, chitin, xylose and phosphate).

Resident community cell yield and respiration both had a clear and strong negative relationship with the invasion success of both invaders, and this effect overlapped with that of composition (Figure 2). High cell yields and respiration levels were associated with lower growth of both invaders in the communities, with a decaying trend. Communities types 1, 3, 5 and 6 (pink, red, blue and violet) achieved cell yields that were tightly skewed toward high cell yield, whilst community types 2 and 4 (green and yellow) spanned
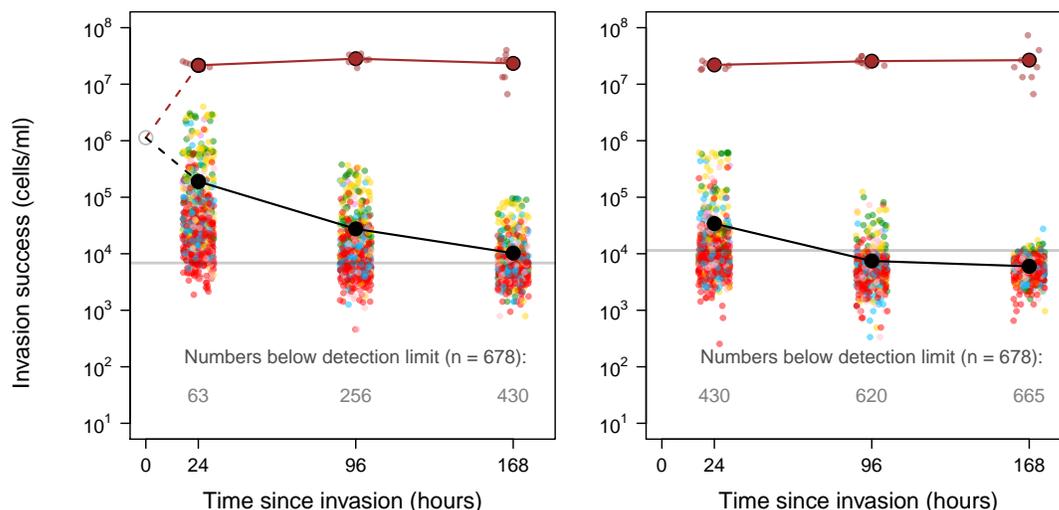
4

Figure 3: Invasion success over time. *Pseudomonas fluorescens* SBW25 (left) and *P.putida* KT2440 (bottom) growth/invasion dynamics in monoculture (brown) and communities (other colours) over time (24, 96 and 168 hours post-invasion). Blue, red, pink, violet, yellow and green colours indicate to which type of community class each community belongs (see Figure 1). Inoculum density is represented by time 0, dashed lines representing approximate (unmeasured) growth trajectories.

the continuum of observed yields, reaching into the lower end of the scale. Conversely, communities types 1, 3, 5 and 6 had much wider variation in respiration rate, whilst community types 2 and 4 were tightly skewed toward the lower end of the respiration scale. Collectively, these results suggested that community types 1, 3, 5 and 6 had grown to carrying capacity and were beginning to exhaust the available resources (hence low respiration rates of some communities but consistently high cell numbers), whilst community types 2 and 4 were still growing at the time of invasion.

The strong effect of resident community growth was also evident in the temporal dynamics of invasion success. When grown in monoculture (the absence of a community), *P. fluorescens* densities increased from an inoculum of $(1.13 \pm 0.07) \times 10^6$ cells/mL to an average of $(2.16 \pm 0.96) \times 10^7$ cells/mL by 24 h (Figure 3). *P. putida* densities increased from an inoculum of $(1.02 \pm 0.58) \times 10^6$ cells/mL to $(2.19 \pm 0.11) \times 10^7$ cells/mL per microcosm after 24 h (Figure 3), remaining at these numbers through 96 and 7 days after invasion. In contrast, invaders always experienced a net average decline when inoculated into communities: $(1.05 \pm 0.19) \times 10^5$ *P. fluorescens* cells/mL survived after 24 h (from an inoculum of $(1.13 \pm 0.07) \times 10^6$ cells/mL) (Figure 3, top). For *P. putida*, $(3.41 \pm 0.04) \times 10^4$ cells survived after 24 h (from an inoculum $(1.02 \pm 0.06) \times 10^6$ cells/mL) (Figure 3, bottom). This decline continued through 96 and 168 h in a decaying trend, but was more gradual for the community types 2 (yellow) and 4 (green), in which both invaders persisted longer. Many invasions approached or fell below the detection limit for the luminescence assay ($6.86 \times 10^3$ and $1.14 \times 10^4$ *P. fluorescens* and *P.putida* cells in 250μl microcosm, respectively), with the numbers below the detection limit increasing over time.

## Overall explanatory power of measured compositional and growth variables

In order to quantitatively understand the relationship between resident community composition, growth and invasion resistance, we implemented a random forest regression approach - accounting for non-independent and non-linear effects on invasion resistance (Figure 4). We maintained the ratio of compositional to growth variables in our models by entering the 14 growth potential explanatory variables

5

alongside the abundances of only the top 10 most abundant OTUs across all communities, 2 diversity metrics and 2 compositional metrics (14 compositional variable total). Adding all 581 OTU abundances did not improve model estimations (Appendix 1: Figure 1). Our models showed distinct and consistent patterns in the way community traits explained invasion resistance.

Overall, in our random forest analysis (see Methods and Appendix 2: Table 1) the variables explained up to 59% of the variation in invasion success, but a large proportion of the variation in invasion success remained unexplained across invasion experiments (Figure 4) - despite our large sample size of 678 communities and the inclusion of 28 diverse compositional and growth-related variables in our models. The negative trend relating the number of invasions falling below the detection limit (Figure 3) and the amount of explained variation (Figure 4) - combined with the similarity of variable importances across timepoints (Appendix 2: Table 1) - suggested that decreases in explanatory power for longer-term invasions were due to detection limits, rather than ecological processes (e.g. increasing stochasticity over time).

There were clear and consistent patterns in the relationship between composition and growth and their effect upon invasion. Most notably and as expected, measures of community cell yield and respiration consistently featured high in the variable importance (quantified as the relative increase in the Mean Square Error ($\%IncMSE$) obtained when the data associated to the variable under analysis is permuted, see Methods) rankings across models (Figure 4, Appendix 2: Table 1). However, other variables (including diversity metrics) were generally much less important and had similar levels of variable importance - indicating that their effects may be overlapping with one another. Based on these and our previous results, we therefore hypothesised that invasion resistance was primarily driven by overall community growth, which was partly but not completely driven by composition. We built two additional sets of random forest models including only compositional variables (Figure 4, red bars) and only growth variables (Figure 4, yellow bars) to confirm this hypothesis. The models including only compositional variables explained less variation in invasion success than the full model (composition + growth, blue bars), whilst the models including only growth variables explained a similar amount of variation to their full model counterparts (Figure 4, Supplementary Figures ,). This confirmed that invasion was driven primarily by community growth, which was not completely predictable from the community composition variables relating to community diversity, structure and phylogenetic relationship to the invader.

## Discussion

Our results demonstrate that even in complex, natural microbial assemblages, relatively simple relationships between community composition and growth may drive large portions of the variation in invasion resistance. By the same token, measures of community growth may serve as better predictors of invasion resistance than measures of community composition, because of a minority communities with similar OTU-level compositions that exhibit substantially different growth and invasion resistance patterns. Our results emphasise the importance of measuring overall cell yield and/or respiration prior to invasion success - even where this is not the primary research question - because the effects of community composition on invasion success often manifest as such.

Under almost all conditions in our experiments, established microbial communities reduced the success of the invader relative to its growth in monoculture (Figure 3). Two compositional types of communities representing approximately a quarter (26.85%) of all communities were associated with a reduction in the invaders' growth success. Across experiments, up to approximately 60% of the variation in this invasion success could be explained by all of the explanatory variables considered, and this explained variation behaved in a simple way with regard to compositional and growth-related effects, as further analysis revealed.
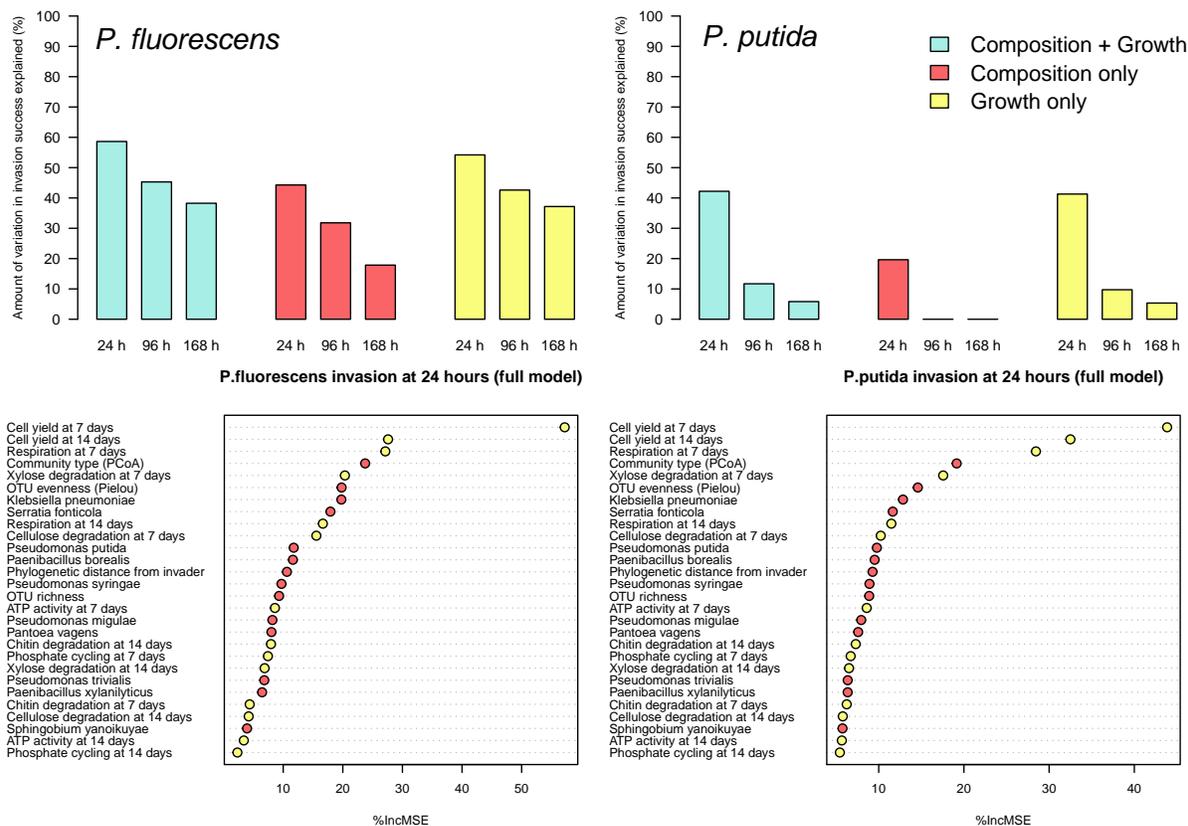
Figure 4: Results of random forest models for the two invaders' (*P. fluorescens* left, *P.putida* right) invasion success in 678 communities (n = 678). The differences in explanatory power between the full model including all 28 explanatory variables (blue), the model including the 14 compositional variables only (red) and the model the 14 including growth variables only (yellow) are shown on the top row. The variable importance values (relative increase in Mean Square Error %IncMSE) for all variables in the full model for both invaders at 24 hours (model with highest explanatory power) are shown on the bottom row, with colours indicating whether the variables are compositional (red) or growth variables (yellow).

## Compositional effects

We established that, whilst starting community composition clearly drives invasion resistance differences between communities (as expected under a common garden design), invasion resistance could more easily be predicted by measuring aspects of community growth directly. Even when all OTUs were accounted for, community composition was only able to explain up to 48.13% of the variation in invasion success (*P.fluorescens* invasion at 24 hours - Appendix 1: Figure 1). This was because the complexity of community composition is hardly encoded in a few measures and additionally, compositional differences mirrored the variation in community cell yield and respiration. This results in a lack of resolution (i.e. some communities with similar compositions had different yield/respiration levels), making cell yield and respiration more predictive measures. These effects appeared to be associated with the two invasion-vulnerable community types being dominated by genera (Pseudomonas and Paenibacillus/Bacillus) that were of lower abundance in other communities. However, there was no straightforward association between the abundance of particular resident species and the invaders - rather, the community type as a whole was a better predictor of invasion success (Figure 4), suggesting the importance of species interac-

tions. The unevenness of the two invasion-vulnerable community types is one way in which such species interactions might manifest, and indeed OTU evenness ranked fairly highly in the variable importance rankings. However, community types 5 and 6 were more uneven (>50% dominated at the genus level by single genera Serratia and Sphingobium, respectively) than invasion-vulnerable types 2 and 4 - suggesting that species interactions did not simply manifest as unevenness and the identity of interacting species was also important. We also tested Darwin's naturalisation hypothesis, in which communities with less similar species to the invader are predicted to be more vulnerable to invasion because they are less likely to occupy the same niche space as the invader [15]. Evidence from previous bacterial microcosm experiments regarding this effect is mixed [20, 23, 27, 30]. In our common garden experiment with 678 natural communities and 2 invaders, however, we found no evidence of a relationship between the mean phylogenetic distance between each OTU in a community and the invader (see Methods: Compositional analyses), with this metric scoring low in all variable importance rankings (Figure 4, Appendex 2: Table ). Furthermore, one of the most invasion vulnerable community types (4) had the highest relative abundance of the *Pseudomonas* genus amongst the community types.

Furthermore, we found no strong effect of community diversity on invasion resistance. Most existing experiments using constructed or manipulated microbial communities have suggested that biodiversity (defined in various ways) is a good indicator of invasion resistance in microbial communities [6, 16, 47]. Although diversity (particularly evenness) emerged as more important than some of the measures of the community metabolic activity, we consistently identified community cell yield as a much more important predictor of invasion resistance. Previous reviews have predicted that diversity-functioning relationships are likely to be less prominent in natural communities because variation in diversity between natural communities is small relative to those in synthetic communities [3]. If the relationship between diversity and invasion success is also decelerating as has been observed in several previous experiments [13, 16, 47], then one would expect to find the kind of weak relationship between diversity and invasion resistance observed in our experiments.

## Growth-related effects

The most consistent and strong effect observed in our experiment was the negative effect of community cell yield on invasion success. Community cell yield and respiration at 7 and 14 days emerged as the most important predictors of invasion success across our models. Communities demonstrated negative relationships between invasion success and cell yield and respiration. This negative relationship was approximately linear until invasion success reached $10^4$ cells/microcosm, with invasion success declining approximately one order of magnitude for every order of magnitude increase in resident community cell yield. Consistent with these findings, in previous work we observed that a complex relation between substrate-degrading enzymes and ATP production influenced both cell yield and respiration, which appeared as the main community-level responses [40]. It is then expected that cell yield and respiration emerge as the most predictive variables, resource limitation being the most parsimonious explanation for this yield-invasion relationship. Communities that reached a high cell yield before invasion likely had a low vulnerability to invasion because they monopolised most of the available resources in that environment. Although more composition-specific factors such as invader-targeting toxin and/or antibiotic production may play a role in invasion resistance - as observed in previous studies [9, 21] - we did not find strong evidence that composition had any substantial effect on invasion success independent of the yield effect. The observation that low yield ($<10^{4.5}$ cells/microcosm) communities tended to belong to Community types 2 and 4 reflected this.

Community cell yield has been neglected in favour of hypotheses about diversity and composition in microbial invasion ecology - even though it has long been understood that the effects of composition often manifest as productivity, driving differences in invasion resistance [11, 22, 25, 26]. The most adept microbial invaders have often evolved to exploit this  as exemplified by the triggering of inflammation to reduce the density of resident competitors by invading pathogens of the gut [7] and reductions of antagonistic populations during *Ralstonia* invasion into the rhizosphere [49]. Furthermore, in natural systems,

8

community cell yield may play an even more central role, with short-term disturbances opening the door for invaders unless communities are able to quickly recover to carrying capacity [24]. This windows of opportunity framework for invasions emphasises that the invasibility of a community is temporally variable, not an intrinsic feature of an ecological community, which has to some extent been an implicit assumption of existing microbial invasion experiments [14].

Other measures of growth all cumulatively contributed to invasion resistance but did not have a substantial effect in and of themselves. The capacity to degrade xylose and cellulose at 7 days appeared to be the most important of these measures, in fitting with the idea that the long-term growth success of the communities was dependant upon metabolising nutrients in the leaf-litter medium [44]. However in order to reject the hypothesis that the metabolism of specific substrates was not an important component in invasion resistance, community functioning would have to be measured more widely.

One surprising component of our results is the extent of overlap between how community composition and community growth affect invasions. We found the explained component of invasion success to be almost entirely predictable from growth-related measures alone, with composition only contributing a small amount of explained variation independent of our 14 functional measures. There are three likely reasons for this:

1. The effect of composition on invasion resistance genuinely manifests almost solely as growth success, captured by our 14 measures. To expand explanatory power, measures of other functional components of communities, such as antibiotic production or pH modification, would need to be measured.

2. The effect of composition is happening at a deeper taxonomic resolution than the OTU level, and is therefore not detectable beyond its coarse effect on resource metabolism. We tested the effects Amplicon Sequence Variant (ASV)-level composition [8] in a separate analysis and found it to have poorer explanatory power, but non-16S amplicon approaches might change our results. To expand explanatory power, deeper sequencing of the communities (e.g. metagenomics) and/or approaches using other amplicons more specifically related to functioning is likely needed.

3. Contrary to the results of experiments with artificial communities, starting composition of communities is insufficient to predict their eventual resistance to invasion after growth. To expand explanatory power, composition would need to be measured throughout the growth period prior to invasion.

It is likely that a mix of these three mechanisms underpins our results and therefore future work should consider additionally testing components of community functioning not obviously related to the resource niche, using deeper sequencing methods on a smaller scale, and sequencing more frequently throughout the experiment. The observations that OTU level composition overlaps with growth performance, and diversity has little impact on invasion resistance in natural communities, therefore shifts our perspective on microbial invasions, suggesting new questions need to be asked in this field of research and with a different methodological approach in order to understand microbial invasions better. Our results suggest that furthering our understanding of microbial invasions will require looking beyond diversity and composition when hypothesising and designing experiments. For example, our experiment - like those before it - was conducted in closed microcosms where the culture medium was not refreshed and invader densities thus declined after initial introduction. In nature, environmental stochasticity is likely to be higher, with factors like nutrient levels fluctuating and resulting in more complex invasion dynamics. As we have argued before [24], we hypothesise that such environmental stochasticity will primarily affect invasion success by causing resident community cell yield to fluctuate over time, opening and closing windows for invasion success. We therefore encourage more studies with natural communities like ours but with more dynamic, open systems that enable community yield to be disturbed in varying temporal and spatial patterns, and the resulting effects on invasion success quantified more accurately.

9

## Conclusion

The field of microbial invasion ecology has recently emerged with the goal of understanding which characteristics of microbial communities best predict their risk of invasion by other microbes. Building from the foundations of biodiversity-ecosystem functioning theory, microbial invasion ecologists have repeatedly and convincingly argued that the diversity of a community or at least its composition drives its invasion resistance. However, our results suggest that whilst the general risk of invasion from microbes in natural communities does result from its composition, it is currently difficult to reject the hypothesis that this effect may often primarily operate almost exclusively through limiting communities ability to grow well an effect which best captured by measuring the yield and/or respiration of the community being directly. The high dimensionality of community composition hinders a fair comparison between any metric aiming to quantify it, that necessarily must reduce its dimensionality, and more simple metrics such as cell yield are thus often more useful. We suggest that more research is needed to investigate which are the most appropriate compositional metrics, and to test its performance against the (more parsimonious) hypothesis that cell yield and respiration are better indicators of invasion success. This would require measuring cell yield and respiration before and after invasion and testing the effects of temporally-variable environmental conditions. It is encouraging to see forthcoming publications in this direction [1,45] and we hope this trend will continue in order to uncover the general principles of microbial invasions and adapt them to applied problems.

# Methods and Materials

## Field sampling of communities

We used naturally-occurring bacterial communities collected from 678 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 [4]. Bacterial communities inhabit these pools, 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 water-filled holes in the buttress roots of beech trees. Once located, tree holes were homogenised using a sterile plastic Pasteur pipette and 1ml transferred into a sterile 1.5 ml centrifuge tube (Starlab; Milton Keynes, UK). Samples were transported to the lab on the day of collection, where they were frozen in 60% glycerol with NaCl solution (Sigma-Aldrich; Gillingham, UK). GPS coordinates and the date of collection was recorded for each tree hole.

## 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. Frozen stocks of field communities were grown in BLT for 14 days used to create fresh, laboratory-acclimatised glycerol stocks prior to any experimentation. This break-in period was designed to allow communities to adjust to laboratory conditions and produce sufficient stocks for subsequent experiments. Next-generation sequencing of communities was performed immediately after this break-in period, and thus sequencing results represent the inoculum composition of communities, as is conventional for biodiversity-invasion resistance experiments (see Community composition). For 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 a further 14 days at 22C before invasion treatments were applied.

## Invasions

Two lux-tagged, IPTG-inducable common soil bacteria, *Pseudomonas fluorescens* SBW25 and *Pseudomonas putida* KT2440 were selected as model invaders. Pseudomonads are fast-growing, metabolically versatile bacteria that are common colonisers, contaminants and pathogens in microbial communities of human interest [38]. They have also been considered as candidates for biocontrol (e.g. of plant health) - a field that would also benefit from general information about the circumstances under which these strains can colonise new communities.

These particular strains were selected because they were closely related and had similar growth characteristics in monoculture. Conspecifics had different relative abundances in communities being invaded - 1 OTU representing *Pseudomonas fluorescens* (mean abundance in communities = 0% 0) and 1 representing *Pseudomonas putida* (mean abundance in communities = 6.5% 0.35) was found in the communities. The choice of these two species therefore allowed us to qualitatively compare the invasion success of a true invader and an introduced species that has conspecific residents already in the community, with the result that the latter (*P.putida*) struggled to take hold.

The two invaders were invaded separately into each of the 678 communities. Having been grown in a BLT for 14 days and functionally profiled (see below), 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 communities 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 $10^4$ cells according to cytometry readings. Such a high invasion pressure was chosen in order to ensure that propagule pressure (the number of individuals invading) was not limiting invasion success [24], allowing us to focus on the resident communities resistance to invasion.

## Invasion success

As previously described [24], luminescence of lux-tagged bacterial strains is directly related to the density of metabolically active cells [10]. Therefore, luminescence values of each microcosm were used to calculate the number of metabolically active invader cells produced from the original invader inoculum. As described previously [24], 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. Invasion success 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). All microcosms with luminescence values below the maximum background luminescence of 1000 sterile BLT microcosms (12 lumens) were removed from the dataset before further analysis. Accordingly, the detection limits for each invader were 6.86 x $10^3$ cells/mL for *P. fluorescens* and 1.14 x $10^4$ cells/mL for *P. putida*.

In total, therefore, we assayed 678 communities (n = 678) for invasion resistance to 2 invaders with invasion success measured at 3 timepoints (678 x 2 x 3 = 4063 measurements total). This entire 3-week long assay (2 weeks of community growth, followed by invasion and measurement of invasion success up until 1 week after invasion) was repeated 4 times (technical replicates = 4) in order to obtain a more accurate mean invasion success per community. No explicit power analysis was used as our study was not hypothesis-driven (or at least there were many possible hypotheses) and we had a very large number of replicates at our disposal (678).

## Amplicon sequencing

Communities were sequenced and characterised with bioinformatics as described in [43]. Briefly, after collection in the field, communities were grown for 1 week in BLT medium and DNA was then 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: 94C for 3 minutes, followed by 28 cycles of 94C for 30 seconds, 53C for 40 seconds and 72C for 1 minute, followed by a final elongation step at 72C 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. For this experiment, the final OTU table included 580 OTUs.

## Compositional analyses

The OTU abundance table was used for all downstream analysis relating to community composition. OTU richness and evenness (Pielou index) was computed from the species abundance table using the R package vegan and custom functions [39]. Phylogenetic distance between the invader and community was calculated as follows. First, a phylogenetic tree was constructed. This was done by taking each 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 500bp 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. Secondly, the cophenetic function of the picante package 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 between each invader and all OTUs. Finally, these vectors were used to calculate an abundance weighted invader-community phylogenetic distance metric using the weighted.mean function (stats package) to calculate the phylogenetic distance between the invader and each OTU, weighted by the relative abundance of that OTU in each community.

We calculated the compositional dissimilarity between all 678 communities with Jensen-Shannon divergence [17]. Communities were then clustered using the Partition Around Medoids (PAM) method implemented via the *pam* function of the R package cluster [42]. Clustering was repeated for a broad range of number of output clusters $K$, using the Calinski-Harabasz index ($CH$) to select the optimal classification $k_{opt} = \arg\max_k(CH)$. Community similarity was visually represented by reducing the dimensionality of the data with Principal Coordinate Analysis (R function *dudi.pco*, package ade4), and projecting the communities into the first two coordinates. Communities were then coloured according to the community type identified using the PAM method.

## Respiration and yield

We measured traits related to the productivity of communities by performing assays of community respiration and yield production. Respiration of each community was measured at 14 days, providing a measure of the productivity of the communities. Assays were performed using the MicroResp system (The James Hutton Institute; Aberdeen, UK). Briefly, agar-set indicator gels are suspended above the growing cultures in a deep-well 96 well plate in an airtight system. As growing cells respire, $CO_2$ released is absorbed by the indicator gels above the cultures, and the gel colour changes from pink to purple. Colour change is read via a spectrophotometer ($OD_{400}$ at 0 (prior to suspension above cultures) and after 24 hours of respiration. The change in colour is used to calculate $mgCO_2$ released in the 24-hour period using a standard curve.

Cumulative yield of the communities was measured at 14 days by staining with Thiazole Orange (Sigma-Aldrich; Gillingham, UK) at concentration of 100nM for 15 minutes before being read 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 yield).

## Metabolic activity

We measured traits related to the metabolic activity of communities by performing assays of ATP and enzyme activity. ATP activity assays provide a general estimate of how metabolically active communities are, with the hypothesis 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 the end of 14 day period of community growth prior to invasion using the BacTiter-Glo$^{\mathrm{TM}}$ 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, the complex substrate of BLT [2]. Substrates labeled 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 [19]. 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.

## Statistical analysis

All statistical analyses were conducted in R [41]. To determine which of the many measured functional variables were the best predictors of *P. fluorescens* and *P. putida* invasion success, we implemented a random forest regression approach using the randomForest package [31]. Random forest regression is an ensemble machine learning technique that works by constructing a set of independent regression trees on subsamples of the entire dataset. In each regression tree, a subset of all samples (approximately two thirds) and a subset of the explanatory variables are used to predict values of the response variable in the remaining third of the dataset (here *P. fluorescens* or*P. putida* invasion success). The results are then combined in order to summarise the ability of each of the explanatory variables to predict the remaining data. The effect of each explanatory variable may be direct or due to simple or complex interactions with other variables. There are two main advantages of the random forest method. Firstly, it utilises regression trees ability to handle non-linear and complex relationships between dependent and independent variables. Secondly, by using a subset of the explanatory variables in each tree, it ameliorates problems associated with model over-fitting that are commonly encountered when entering a large number of explanatory variables into a single regression tree. To estimate the predictive power of each variable $i$, we first computed the Mean Square Error $MSE$ obtained following an "Out of the Bag" procedure [32]. The $MSE$ was then recalculated when the data associated with each variable was permuted ($pMSE(i)$). The importance of a variable was quantified as the relative increase of the Mean Square Error: $\%IncMSE(i) = (MSE - pMSE(i))/MSE$.

This approach allowed us to identify the variables most important to invasion success, accounting for overlapping and non-linear effects. We first explored how 28 composition and growth-related measurements of resident communities affected their invasion resistance by creating separate random forests for each of the *P. fluorescens* and *P. putida* invasion experiments, for each the invasion successes at 24, 96 and 7 days since invasion (i.e. 2 x 3 = 6 random forests total). Random forests included 28 variables

related to starting composition (14 variables) and all growth-related variables measured at 2 timepoints prior to invasion (14 variables) (i.e. 7 days and 14 days). Secondly and thirdly, we made a new set of random forests using only the 14 compositional and 14 growth-related variables, respectively - allowing us to determine the extent of overlap in explained variance between composition and growth (revealing that growth was driving almost all of the variation in invasion success).

In Appendix 2: Table 1, we report the total variance explained by each random forest (% variance explained) and the increase in prediction error when each variable is not permuted ($\%IncMSE$ - i.e. the variable importance). These statistics reflect the overall predictive power of the entire set of metrics and of each individual metric, respectively. A high value of $\%IncMSE$ indicates predictor makes a large contribution to the overall predictive power of the model. As noted above, error estimates are incorporated into these statistics given that they are calculated from average predictive error, as described by [33].

All forests were validated by checking that each forest was grown to a large enough size (ntree = 10000) that error stabilised. Adjusting the number of explanatory variables (mtry) entered in each iteration of the random forest to the optimum as calculated by the trainControl function of the caret R package [28] did not affect results, and so we kept the default value of the number of variables divided by 3.

# Acknowledgments

# Data Availability

Data and R code is available via Dryad (https://doi.org/10.5061/dryad.qz612jm9k). Sequence data is available from the NCBI Short Read Archive under accession number SRP14503. Sequence data is available from the NCBI Short Read Archive under accession number SRP14503.

# References

[1] Michaeline B.N. Albright, Sanna Sevanto, La Verne Gallegos-Graves, and John Dunbar. Principles for engineering microbial composition and ecosystem functioning Michaeline. *bioRxiv*, pages 1–21, 2019.

[2] Manish Kumar Aneja, Shilpi Sharma, Frank Fleischmann, Susanne Stich, Werner Heller, Günther Bahnweg, Jean Charles Munch, and Michael Schloter. Microbial colonization of beech and spruce litter - Influence of decomposition site and plant litter species on the diversity of microbial community. *Microbial Ecology*, 52(1):127–135, 2006.

[3] Thomas Bell, Andrew K Lilley, Andy Hector, Bernhard Schmid, Lindsay King, and Jonathan a Newman. A Linear Model Method for Biodiversity Ecosystem Functioning Experiments. *The American naturalist*, 174(6):836–849, dec 2009.

[4] Thomas Bell, Jonathan A. Newman, Bernard W. Silverman, Sarah L. Turner, and Andrew K. Lilley. The contribution of species richness and composition to bacterial services. *Nature*, 436(7054):1157–1160, 2005.

[5] Michael D. Besmer, Jürg A. Sigrist, Ruben Props, Benjamin Buysschaert, Guannan Mao, and Frederik Hammes. Laboratory-scale simulation and real-time tracking of a microbial contamination event and subsequent shock-chlorination in drinking water. *Frontiers in Microbiology*, in press(October):1–11, 2017.

[6] Giuliano Bonanomi, Manuela Capodilupo, Guido Incerti, Salvatore A. Gaglione, and Felice Scala. Fungal diversity increases soil fungistasis and resistance to microbial invasion by a non resident species. *Biological Control*, 72:38–45, 2014.

[7] Sam P. Brown, Ludovic Le Chat, and François Taddei. Evolution of virulence: Triggering host inflammation allows invading pathogens to exclude competitors. *Ecology Letters*, 11(1):44–51, 2008.

[8] Benjamin J. Callahan, Paul J. McMurdie, Michael J. Rosen, Andrew W. Han, Amy Jo A. Johnson, and Susan P. Holmes. DADA2: High-resolution sample inference from Illumina amplicon data. *Nature Methods*, 13(7):581–583, 2016.

[9] Emilie Chapelle, Rodrigo Mendes, Peter A.Hm Hm Bakker, and Jos M. Raaijmakers. Fungal invasion of the rhizosphere microbiome. *The ISME journal*, 10(May):1–4, 2015.

[10] Dan Close, Tingting Xu, Abby Smartt, Alexandra Rogers, Robert Crossley, Sarah Price, Steven Ripp, and Gary Sayler. The evolution of the bacterial luciferase gene cassette (lux) as a real-time bioreporter. *Sensors*, 12(1):732–752, 2012.

[11] Michael J. Crawley. Chance and Timing in Biological Invasions. In J.A. Drake, H.A. Mooney, F. di Castri, R.H. Groves, F.J. Kruger, M. Rejmánek, and M. Williamson, editors, *Biological invasions: a global perspective*, chapter 18, pages 407–423. Wiley-Blackwell, New York, New York, USA, 1989.

[12] Michael J. Crawley, Susan L. Brown, Matt S. Heard, and Grant R. Edwards. Invasion resistance in experimental grassland communities: species richness or species identity? *Ecology Letters*, 2(3):140–148, 1999.

[13] Karen de Roy, Massimo Marzorati, Andrea Negroni, Olivier Thas, Annalisa Balloi, Fabio Fava, Willy Verstraete, Daniele Daffonchio, and Nico Boon. Environmental conditions and community evenness determine the outcome of biological invasion. *Nature communications*, 4(1383):1–5, 2013.

[14] Brian G. Degasperis and Glenn Motzkin. Windows of opportunity: Historical and ecological controls on Berberis thunbergii invasions. *Ecology*, 88(12):3115–3125, 2007.

[15] Richard P. Duncan and Peter A. Williams. Ecology: Darwin's naturalization hypothesis challenged. *Nature*, 417(6889):608–609, 2002.

[16] Nico Eisenhauer, Wiebke Schulz, Stefan Scheu, and Alexandre Jousset. Niche dimensionality links biodiversity and invasibility of microbial communities. *Functional Ecology*, 27(1):282–288, feb 2013.

[17] Dominic M. Endres and Johannes E. Schindelin. A New Metric for Probability Distributions. *IEEE Transactions on Information Theory*, 49(7):1858–1860, 2003.

[18] Francesca Fiegna, Alejandra Moreno-Letelier, Thomas Bell, and Timothy G. Barraclough. Evolution of species interactions determines microbial community productivity in new environments. *The ISME Journal*, 9(5):1235–1245, 2015.

[19] Aline Frossard, Linda Gerull, Michael Mutz, and Mark O. Gessner. Shifts in microbial community structure and function in stream sediments during experimentally simulated riparian succession. *FEMS Microbiology Ecology*, 84(2):398–410, 2013.

[20] Yabing Gu, Delong Meng, Sheng Yang, Nengwen Xiao, Zhenyu Li, and Zhenghua Liu. Invader-resident community similarity contribute to the invasion process and regulate biofertilizer effectiveness. *Journal of Cleaner Production*, 241:118278, 2019.

[21] Xuesong He, Jeffrey S. McLean, Lihong Guo, Renate Lux, and Wenyuan Shi. The social structure of microbial community involved in colonization resistance. *The ISME journal*, 8(3):564–74, 2014.

[22] David J. Hodgson, Paul B. Rainey, and Angus Buckling. Mechanisms linking diversity, productivity and invasibility in experimental bacterial communities. *Proceedings of the Royal Society of London Series B-Biological Sciences*, 269(1506):2277–2283, nov 2002.

[23] Lin Jiang, Jiaqi Tan, and Zhichao Pu. An experimental test of Darwin's naturalization hypothesis. *The American Naturalist*, 175(4):415–23, apr 2010.

[24] Matt L. Jones, Josep Ramoneda, Damian W. Rivett, and Thomas Bell. Biotic resistance shapes the influence of propagule pressure on invasion success in bacterial communities. *Ecology*, 98(7):1743–1749, 2017.

[25] Alexandre Jousset, Wiebke Schulz, Stefan Scheu, and Nico Eisenhauer. Intraspecific genotypic richness and relatedness predict the invasibility of microbial communities. *The ISME Journal*, 5(10):1108–1114, 2011.

[26] Marta Kinnunen, Arnaud Dechesne, Caitlin R Proctor, Frederik Hammes, David Johnson, Marcos Quintela-Baluja, David W Graham, Daniele Daffonchio, Stelios Stilianos Fodelianakis, Nicole Hahn, Nico Boon, and Barth F. Smets. A conceptual framework for invasion in microbial communities. *The ISME Journal*, 10(12):1–7, 2016.

[27] Marta Kinnunen and Barth F Smets. Stochastic processes govern invasion success in microbial communities when the invader is phylogenetically close to resident bacteria. *The ISME Journal*, pages 2748–2756, 2018.

[28] Max Kuhn, Jed Wing, Steve Weston, Andre Williams, Chris Keefer, and Allan Engelhardt. Caret: Classification and Regression Training, 2012.

[29] Diane Lawrence, Francesca Fiegna, Volker Behrends, Jacob G. Bundy, Albert B. Phillimore, Thomas Bell, and Timothy G. Barraclough. Species interactions alter evolutionary responses to a novel environment. *PLoS Biology*, 10(5):e1001330, 2012.

[30] Shao-peng Li, Jiaqi Tan, Xian Yang, Chao Ma, and Lin Jiang. Niche and fitness differences determine invasion success and impact in laboratory bacterial communities. *ISME Journal*, 13(2):402–412, 2019.

[31] Andy Liaw. Package 'randomForest', 2018.

[32] Andy Liaw and Matthew Wiener. Classification and Regression by randomForest. *R News*, 2(3):18–22, 2002.

[33] Nicole Liew, Maria J.Mazon Moya, Claudia J. Wierzbicki, Michael Hollinshead, Michael J. Dillon, Christopher R. Thornton, Amy Ellison, Jo Cable, Matthew C. Fisher, and Serge Mostowy. Chytrid fungus infection in zebrafish demonstrates that the pathogen can parasitize non-amphibian vertebrate hosts. *Nature Communications*, 8:1–10, 2017.

[34] Elena Litchman. Invisible invaders: Non-pathogenic invasive microbes in aquatic and terrestrial ecosystems. *Ecology Letters*, 13(12):1560–1572, 2010.

[35] Cyrus Alexander Mallon, Jan Dirk Van Elsas, Jan Dirk van Elsas, and Joana Falcão Salles. Microbial invasions: The process, patterns, and mechanisms. *Trends in Microbiology*, 23(11):719–729, 2015.

[36] Cyrus Alexander Mallon, X. Le Roux, G. S. van Doorn, F. Dini-Andreote, F. Poly, and J. F. Salles. The impact of failure: unsuccessful bacterial invasions steer the soil microbial community away from the invader's niche. *The ISME Journal*, 12(3):728–741, 2018.

[37] Cyrus Alexander Mallon, Franck Poly, Xavier Le Roux, Irene Marring, Jan Dirk van Elsas, Joana Falcão Salles, Xavier Le Roux, Irene Marring, Jan Dirk van Elsas, and Joana Falcão Salles. Resource pulses can alleviate the biodiversity invasion relationship in soil microbial communities. *Ecology*, 96(4):915–926, 2015.

[38] Pablo I Nikel, Esteban Martínez-García, and Víctor de Lorenzo. Biotechnological domestication of pseudomonads using synthetic biology. *Nature reviews. Microbiology*, 12(5):368–79, 2014.

[39] Jari Oksanen, F. Guillaume Blanchet, Michael Friendly, Roeland Kindt, Pierre Legendre, Dan McGlinn, Peter R. Minchin, R. B. O'Hara, Gavin L. Simpson, M. Peter Solymos, Henry H. Stevens, Eduard Szoecs, and Helene Wagner. Package 'vegan', 2019.

[40] Alberto Pascual-Garcia and Thomas Bell. Community-level signatures of ecological succession in natural bacterial communities. *bioRxiv*, 2019.

[41] R Core Team. A Language and Environment for Statistical, 2017.

[42] A P Reynolds, G Richards, and B D E L A Iglesia. Clustering Rules : A Comparison of Partitioning and Hierarchical Clustering Algorithms. *Journal of Mathematical Modelling and Algorithms*, 5:475–476, 2006.

[43] Damian W Rivett and Thomas Bell. Abundance determines the functional role of bacterial phylotypes in complex communities. *Nature Microbiology*, 2018.

[44] Damian W. Rivett, Thomas Scheuerl, Christopher T. Culbert, Shorok B. Mombrikotb, Emma Johnstone, Timothy G. Barraclough, and Thomas Bell. Resource-dependent attenuation of species interactions during bacterial succession. *The ISME Journal*, 10(9):2259–2268, 2016.

[45] Kati Saarinen and Leena Lindström. Invasion triple trouble : environmental fluctuations, fluctuation-adapted invaders and fluctuation-mal-adapted communities all govern invasion success. *BMC Evolutionary Ecology*, 19(1):1–8, 2019.

[46] Madhav P Thakur, Wim H van der Putten, Marleen M P Cobben, Mark van Kleunen, Stefan Geisen, Wim H Putten, Marleen M P Cobben, Mark Kleunen, and Stefan Geisen. Microbial invasions in terrestrial ecosystems. *Nature Reviews Microbiology*, 2019.

[47] Jan Dirk van Elsas, Mario Chiurazzi, Cyrus A Mallon, Dana Elhottova, Václav Kristufek, and Joana Falcão Salles. Microbial diversity determines the invasion of soil by a bacterial pathogen. *Proceedings of the National Academy of Sciences*, 109(4):1159–64, 2012.

[48] Jens Walter, María X. Maldonado-Gómez, and Inés Martínez. To engraft or not to engraft: an ecological framework for gut microbiome modulation with live microbes. *Current Opinion in Biotechnology*, 49:129–139, 2018.

[49] Zhong Wei, Jie Hu, Yi'an Gu, Shixue Yin, Yangchun Xu, Alexandre Jousset, Qirong Shen, and Ville-Petri Petri Friman. Ralstonia solanacearum pathogen disrupts bacterial rhizosphere microbiome during an invasion. *Soil Biology and Biochemistry*, 118(December):8–17, 2018.

[50] Douglas C. Woodhams, Hannelore Brandt, Simone Baumgartner, Jos Kielgast, Eliane Küpfer, Ursina Tobler, Leyla R. Davis, Benedikt R. Schmidt, Christian Bel, Sandro Hodel, Rob Knight, and Valerie McKenzie. Interacting symbionts and immunity in the amphibian skin mucosome predict disease risk and probiotic effectiveness. *PLoS ONE*, 9(4):Supporting Information, 2014.
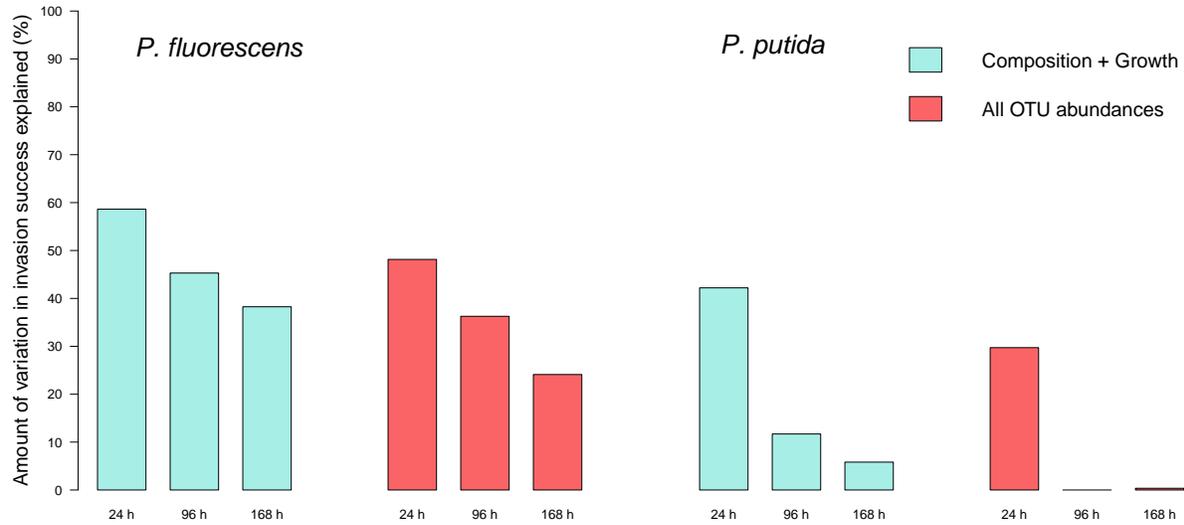
# Appendix 1



Figure 1: Comparison of full model (14 composition + 14 growth explanatory variables) and the model with the abundances of the 581 OTUs found across the 678 communities. Although including all OTUs improves variance explained relative to including only 14 composition variables, explained variance still drops significantly between the full model and the 581 OTUs model, unlike for the full model versus the growth variables only model (Figure 4).

# Appendix 2

Table 1: Random forest results ($\%IncMSE$) for *P. fluorescens* SBW25 and *P.putida* KT2440 invasions for full model including all 14 growth-related measurements and 14 compositional metrics.

| | P. fluorescens SBW25 | | | P. putida KT2440 | | |
|---|---|---|---|---|---|---|
| | 24 h | 96 h | 168 h | 24 h | 96 h | 168 h |
| **Respiration at 7 days** | 27.13 | 33.88 | 27.28 | 28.45 | 7.02 | 9.08 |
| **Respiration at 14 days** | 16.65 | 28.28 | 22.18 | 32.5 | 11.06 | 5.86 |
| **Cell yield at 7 days** | 57.22 | 34.1 | 19.28 | 19.15 | 9.35 | 10.85 |
| **Cell yield at 14 days** | 27.59 | 29.02 | 43.07 | 43.85 | 14.29 | 23.4 |
| **ATP activity at 7 days** | 8.61 | 8.2 | 7.75 | 11.48 | 4.81 | 8.54 |
| **ATP activity at 14 days** | 3.39 | 2.52 | 4.72 | 8.59 | 0.71 | 4.58 |
| **Cellulose degradation at 7 days** | 15.56 | 8.62 | 7.53 | 6.24 | 4.35 | 13.99 |
| **Cellulose degradation at 14 days** | 4.21 | 8.29 | 4.34 | 6.36 | 4.49 | 4.1 |
| **Chitin degradation at 7 days** | 4.38 | 0.17 | 4.47 | 6.71 | 4.75 | 2.28 |
| **Chitin degradation at 14 days** | 7.95 | 4.04 | 4.11 | 9.52 | 5.04 | 3.38 |
| **Xylose degradation at 7 days** | 20.35 | 13.62 | 14.65 | 9.78 | 6.05 | 7.09 |
| **Xylose degradation at 14 days** | 6.88 | 6.78 | 6.16 | 7.95 | 9.2 | 7.97 |
| **Phosphate cycling at 7 days** | 7.43 | 6.45 | 11.66 | 8.88 | 5.12 | 14.3 |
| **Phosphate cycling at 14 days** | 2.32 | 0.79 | 5.1 | 6.36 | 3.59 | 5.25 |
| **OTU richness** | 9.31 | 7.73 | 2.86 | 5.75 | 0.28 | 1.22 |
| **OTU evenness (Pielou)** | 19.8 | 11.26 | 13.74 | 11.65 | 4.18 | 5.59 |
| **Community type (PCoA)** | 23.75 | 19.48 | 14.05 | 12.85 | 6.88 | 5.66 |
| **Phylogenetic distance from invader** | 10.62 | 11.58 | 8.5 | 5.67 | 4.1 | 2.1 |
| **Sphingobium yanoikuyae** | 3.93 | 5.65 | 10.02 | 6.51 | 2.58 | 9.44 |
| **Klebsiella pneumoniae** | 19.76 | 18.08 | 15.76 | 14.59 | 9.51 | 7.2 |
| **Pantoea vagens** | 8.03 | 8.11 | 10.77 | 9.28 | 1.05 | 5.09 |
| **Paenibacillus borealis** | 11.63 | 11.91 | 7.06 | 8.92 | 4.87 | 1.35 |
| **Pseudomonas putida** | 11.76 | 9.44 | 9.99 | 10.23 | 5.14 | 10.89 |
| **Pseudomonas syringae** | 9.72 | 8.07 | 12.21 | 7.58 | 4.28 | 6.44 |
| **Pseudomonas migulae** | 8.19 | 8.94 | 9.72 | 5.44 | 5.11 | 4.75 |
| **Serratia fonticola** | 17.92 | 11.46 | 9.62 | 17.57 | 8.81 | 6.61 |
| **Paenibacillus xylanilyticus** | 6.46 | 6.76 | 5.06 | 7.3 | 6.32 | 1.22 |
| **Pseudomonas trivialis** | 6.83 | 5.41 | 8.8 | 5.78 | 1.43 | 5.02 |