Plant Soil
https://doi.org/10.1007/s11104-017-3492-6

REGULAR ARTICLE

Invasion ecology applied to inoculation of plant growth promoting bacteria through a novel SIMPER-PCA approach

Pedro Beschoren da Costa · Samanta Bolzan de Campos · Andreas Albersmeier · Paul Dirksen · André Luis Pereira Dresseno · Odair José Andrade Pais dos Santos · Karina Maria Lima Milani · Rafael Mazer Etto · André Gustavo Battistus · Andréia Cristina Peres Rodrigues da Costa · André Luiz Martinez de Oliveira · Carolina Weigert Galvão · Vandeir Francisco Guimarães · Alexander Sczyrba · Volker F. Wendisch · Luciane Maria Pereira Passaglia

Received: 6 June 2017 / Accepted: 9 November 2017

Abstract
Aims Plant growth promoting bacteria (PGPB) have been used on crops for years, but inoculants that are efficient in some locations may not be efficient in others. Here, we applied classical invasion ecology theory to PGPB inoculation in order to identify patterns that can be used to predict plant growth promoting (PGP) efficiency. The hypotheses that the inoculant that causes most impact will be the most efficient PGPB, and that the most invasible locations would have higher PGP efficiency, were tested. We also aim to present our statistical approach to analyze SIMPER results.

Responsible Editor: Timothy Cavagnaro.

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s11104-017-3492-6) contains supplementary material, which is available to authorized users.

P. B. da Costa · L. M. P. Passaglia (✉)
Department of Genetics, Federal University of Rio Grande do Sul, Porto Alegre, RS 91501-970, Brazil
e-mail: luciane.passaglia@ufrgs.br

P. B. da Costa · S. B. de Campos · V. F. Wendisch
Faculty of Biology, Bielefeld University, D-33615 Bielefeld, Germany

P. B. da Costa · S. B. de Campos · A. Albersmeier · P. Dirksen · A. Sczyrba · V. F. Wendisch
Center for Biotechnology, Bielefeld University, D-33615 Bielefeld, Germany

A. L. P. Dresseno
Department of Ecology, Federal University of Rio Grande do Sul, Porto Alegre, RS 91540-000, Brazil

O. J. A. P. dos Santos
Department of Agronomy, State University of Londrina, Londrina, PR 10011-000, Brazil

K. M. L. Milani · A. L. M. de Oliveira
Department of Biochemistry and Biotechnology, State University of Londrina, Londrina, PR 10011-000, Brazil

R. M. Etto
Department of Chemistry, State University of Ponta Grossa, Ponta Grossa, PR 84010-919, Brazil

A. G. Battistus · V. F. Guimarães
Agricultural Science Center, Estate University of Western Paraná, Marechal Cândido Rondon, PR 85960-000, Brazil

A. C. P. R. da Costa
Department of Agricultural Sciences, State University of Maringá, Umuarama, PR 87502-970, Brazil

C. W. Galvão
Department of Structural and Molecular Biology and Genetics, State University of Ponta Grossa, Ponta Grossa, PR 84010-919, Brazil

Published online: 14 November 2017

Methods Using next generation sequencing targeting the 16S rDNA gene in metagenomics samples, we analyzed samples of pre-planting bulk soil and rhizosphere of inoculated maize plants. Bacterial communities of inoculated plants were compared to the non-inoculated controls, in order to estimate the inoculant invasion impact. Crop yield was compared to different indexes, and a novel data exploration approach was employed.
Results The most efficient inoculant was not the most invasive, and a nutrient per diversity ratio was unable to predict inoculant efficiency or invasion impact. However, the efficient inoculation treatment presented an enrichment of specific pre-planting taxa.
Conclusions Invasion ecology frameworks could not anticipate field results of inoculated plants. Nonetheless, our data exploration approach, which is explained in detail, can be useful to raise new hypothesis and improve the visualization of dissimilarity data.

Keywords Maize · Metagenomics · Next generation sequencing · PGPB · Rhizosphere

Introduction

Plant growth promoting bacteria (PGPB) have already been used on crops at industrial scale for years (Alves et al. 2003; Castro-Sowinski et al. 2007). However, as plant-bacteria-environment interactions are far from being completely understood, inoculation might not be fully efficient at all times (Berg et al. 2013; Owen et al. 2015). Even commercialized preparations of PGPB might fail to improve crop yield (Otieno et al. 2013; Owen et al. 2015), in part due to interactions with the local microbial community that probably will be competing with the inoculant for nutrients and niches (Bashan et al. 2014). One way to achieve better predictability is the biotechnological intensive construction of very effective and competitive strains (Ryan et al. 2008). Alternatively, ecological interactions in the rhizosphere have to be elucidated and manipulated to maximize plant growth promoting (PGP) efficiency (Castro-Sowinski et al. 2007).

Classical invasion ecology has already been applied to bacteria (van Elsas et al. 2012), but picturing the plant inoculant itself as an invader to the local community had not been considered until recently (Ambrosini et al. 2016). Accurate quantification of the invasion ability can be very difficult (Forrest and Taylor 2002; Margherita and Osborne 2009), as it includes detection of multiple species over time and space, or subjective concepts, like invasibility and invasiveness (Parker et al. 1999; Davis and Pelsor 2001). These will be even more difficult to evaluate with bacterial species and microbial invasions (Mallon et al. 2015). Still, in general, an efficient invader must display a series of traits, such as high dispersal, high reproduction rate, and will cause heavy impact on the environment (Parker et al. 1999; Mallon et al. 2015). Although dispersal and growth rates of inoculants might be very difficult to measure in the field; their impact on the native community is not, thanks to the advancements of Next Generation Sequencing (NGS). Although NGS presents great challenges on appropriate analysis due to the sheer volume of data, several novel evaluation methods are constantly presented. There is no lack of papers describing the best diversity indexes to be used (Haegeman et al. 2013), appropriate multivariate approaches (Rees et al. 2004; Ramette 2007; Thomas et al. 2012), or interactive visualization methods (Ondov et al. 2011). Among these, the Similarity Percentages (SIMPER) test is very powerful (Clarke 1993), but currently underused as its results are usually presented in shortened tables (Forrest and Taylor 2002; Mills et al. 2006; Margherita and Osborne 2009; Gitipour et al. 2013; Purcell et al. 2014; Freedman et al. 2015) or text (Rees et al. 2004; Stevens and Olson 2013).

Here, we tested several hypotheses that could predict PGP efficiency at an early stage of plant growth: (i) bacterial inoculants causing the most impact on the rhizosphere will be the best PGPB. We consider that the highest impact on the invaded microbial community will be caused by the best invader, and use control-treatment dissimilarities from NGS as a proxy for the impact caused by the invader on this community; (ii) environments with a higher nutrient per diversity ratio will suffer more impact than an environment with a lower nutrient per diversity ratio. For this, we consider that an environment with a higher amount of nutrients but lower bacterial diversity might have more available niches, which is one of the factors that facilitates invasion (Hierro et al. 2005); (iii) environments with a higher nutrient per diversity ratio will allow a better plant growth promoting (PGP) effect. Finally, we

presented a novel approach to visualize metagenomic NGS results with the SIMPER test, focusing on dissimilarities to a standard. This approach can be useful on microbiology research, and it is one of our objectives to explain it in detail so it can be reproduced.

Materials and methods

Field trials design, sampling and soil chemical analysis

Maize (*Zea mays*) plants (hybrid variety 30F53HY, Pioneer) were subjected to a PGPB trial in three different locations in Paraná state, Brazil. The field crops were sown in Londrina (L) (23° 17′ 34″ S, 51° 10′ 24″ W), Marechal Cândido Rondon (M) (24° 33′ 24″ S, 54° 3′ 24″ W), and Ponta Grossa (P) (25° 00′ 50″ S, 50° 09′ 18″ W). All of these locations have climate classified as humid subtropical (Köppen climate classification Cfa). Inoculation treatments were as follow: a non-inoculated control (treatment 1), *Azospirillum brasilense* Ab-V5 (treatment 2) (Hungria et al. 2010), *Achromobacter* sp. VC36 (treatment 3) (Arruda et al. 2013), *Pseudomonas* sp. 4311 (treatment 4), and *Pseudomonas* sp. 4312 (treatment 5) (André Oliveira, personal communication). In order to simplify the nomenclature, we will refer each treatment with the first letter of location followed by the number of inoculated strain. For example: when condition "L3" is cited, we mean *Achromobacter* sp. VC36 (treatment 3) inoculation on Londrina (L) location. Bacteria were grown in liquid LB medium at 28 °C under agitation (200 rpm) for 16 h. Twenty mL of inoculants cultures containing 10^9 cells mL^{-1} were used per kg of seeds. The seeds were exposed to the inoculant twelve to twenty-four hours before sowing. All treatments and controls received 30 kg $hectare^{-1}$ of N fertilizer, and randomized blocks were composed of 4 lines of 10 m (length) × 3.2 m (width) with 0.8 m spacing between each line, following The Brazilian Ministry of Agriculture and Livestock regulations for bacterial inoculant testing. Fields were sown between November 2012 and January 2013. All of the inoculated strains were isolated from maize in previous works. The *A. brasilense* Ab-V5 strain is a well-known PGP bacterium that displays a high N_2-fixing ability in vitro, is able to produce indolic compounds, and had increased the contents of some nutrients in the leaves and grains of maize and wheat, as P, K and Cu, and increased the N contents in the leaves of these crops (Hungria et al. 2010). The *Achromobacter* sp. VC36 strain is able to produce indolic compounds and increased contents of some nutrients in leaves and grains of maize and wheat (Arruda et al. 2013). The two *Pseudomonas* strains have 97% and 95% similarity with the 16 s rDNA gene of *P. koreensis* (for 4311 and 4312, respectively). Both produce indolic compounds but only strain 4312 is able to solubilize P (André Oliveira, personal communication).

Ten days after plant emergence the rhizospheric soil directly attached to the plant roots was scraped from three independent plants and pooled to compose a composite sample. Very young plants were sampled because the effect of inoculation on rhizosphere community composition might be transient or even undetectable after some time (Castro-Sowinski et al. 2007; Chowdhury et al. 2013), and also because PGP efficiency prediction at this early stage could have an important agricultural application. Two independent composite samples (a and b) were used per treatment, allowing us to include all three maize fields in our sequencing effort, which was essential for the experimental design of this collaborative project. Two additional independent composite samples consisting of bulk soil were taken immediately before planting to characterize the pre-planting conditions. The bulk soil samples (treatment 0) were also used for soil chemical analysis, using standard methods (Sparks et al. 1996).

DNA extraction, amplification and sequencing

Rhizospheric and bulk soil DNAs were extracted from 0.3 g of each soil sample using the Nucleo Spin Soil™ kit (Macherey-Nagel). DNA amplification of the V4 region of the bacterial 16S rRNA was performed using the primers F515 and R806 (Caporaso et al. 2011) and purified using MinElute (Qiagen). The protocol for barcoded Illumina pyrosequencing was described by Caporaso et al. (2011). TruSeq DNA protocol (Illumina) was followed for library preparation, starting with the end repair of the fragments. Libraries were sequenced on the MiSeq sequencer (Illumina) with a read length of 2 × 300 nt. MiSeq Control Software v 2.3.0 was used for sequencing and MiSeq Reporter Software v 2.3.32 for demultiplexing and generation of FASTQ files.

Bioinformatics analyses

All sequence analyses were done using QIIME 1.8.0. Read quality control was performed with the FastQC tool, and merging of sequences was performed with the paired end overlap tool, available at the Bioinformatics Resource Facility of the CeBiTec computer cluster (Bielefeld, Germany). QIIME defaults were used for filtering of raw and merged Illumina data, OTUs picking and clustering at 97% identity. Representative sequence alignments were performed with the PyNAST tool, and the greengenes database was used for comparisons through the RDP classifier. The final OTU table was generated in the BIOM format. All sequences are available at NCBI Accession number PRJNA340246.

Statistical analysis

For analysis of metagenomes, OTUs under 200 occurrences were removed from the analysis. Read counts were transformed to relative abundances to normalize the number of valid reads, and then square-root transformed. The SIMPER test produces direct comparisons of 2 or more groups of samples, returning the dissimilarities each taxon is responsible for as percentages of total dissimilarity. Here, SIMPER comparisons were made with the 2 independent samples from the inoculated treatments and the appropriate non-inoculated control from each location. The total SIMPER difference on the treatment-control pairs was used as a proxy for impact of invasion. Before using the SIMPER-PCA approach, all taxa dissimilarities from the SIMPER test were signal-transformed to show if each treatment had more or less of a particular OTU than the appropriate control (explained in supplementary material). The signal-transformed, Bray-Curtis based SIMPER dissimilarities were then processed in a within-group PCA (Principal Component Analysis), grouping per location. Within-group PCA minimizes the differences between groups, and was used to reduce the clustering by location effect, highlighting the control-treatment differences. All tests were performed on Paleontological Statistics (PAST) software (Hammer et al. 2001).

Field trial crop yield was tested with randomized block analysis of variance (ANOVA) for each location. Fold increases in yield where calculated by dividing each treatment yield by the average control yield for each location. Crop yield was calculated considering number of grains per stalk, 100 grain weight and grain humidity. Data from soil nutrient analysis (clay content, P, K, Organic matter, Ca, Mg, S, Cu, B, and Mn) were divided by the pre-planting Shannon diversity index of respective soils. The Shannon index has been chosen as it is considered appropriate to evaluate microbial taxonomic diversity (Haegeman et al. 2013). These values were then transformed to relative percentages per location and averaged, generating a normalized nutrient per diversity ratio. All correlations (nutrient per diversity ratio, SIMPER values and crop yield) were calculated with Pearson test.

Results

The maize field experiments consisted of one non-inoculated control and four inoculated bacterial strains, tested in three different locations in Paraná state, Brazil. Rhizospheric DNA extractions from two independent biological replicates, taken 10 days after plant emergence, were used for 16S rDNA sequencing.

Crop yield

Fold changes in grain yield due to the inoculation of the four different strains in three different locations depended on trial location. Randomized block ANOVA shows that inoculation affected grain yield at location L ($p < 0.001$), but not at M ($p = 0.11$) or P locations ($p = 0.51$) (Fig. 1). At Londrina location, inoculation of maize plants with treatment 5 increased grain yield compared to control by $37.4 \pm 9.8\%$, while the inoculation with treatments 2, 3 and 4 did not affect grain yield significantly. Controls yielded 7.292 ± 521, 10.433 ± 2.723, and 10.481 ± 1.201 kg ha^{-1} for locations L, M and P, respectively.The full dataset regarding crop productivity is available in Online Resource 1.

16S rRNA gene sequencing

Sequencing of amplicons from the V4 region of the bacterial 16S rDNA with the MiSeq platform returned 12.1 million reads, with 5.6 million of unique sequences clustered in 1.3 million OTUs. The 36 sequenced samples provided about 844 thousand singleton sequences, composing 15.1% of sequences and 64% of OTUs. Reads assembling was of $95 \pm 4\%$, except for samples P2b and M3b that resulted in only 37% and 50% assembled reads, respectively. The average number of

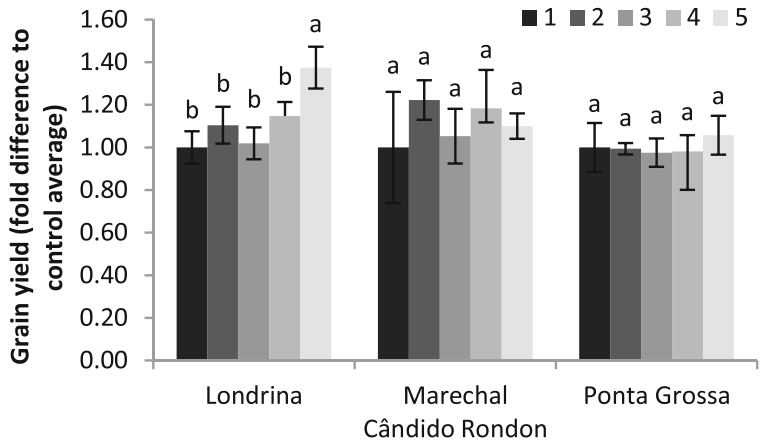

Fig. 1 Fold changes in grain yield due to the inoculation with four different strains in the three different locations. 1 = non-inoculated control; 2 = *Azospirillum brasilense* Ab-V5; 3 = *Achromobacter* sp. VC36; 4 = *Pseudomonas* sp. 4311; and 5 = *Pseudomonas* sp. 4312

valid reads was 313,984, ranging from 49,623 (P2b) to 1,711,823 (M5a). Two samples (P4a and M0b) were entirely removed from analysis because of a large dominance (60 and 40%, respectively) of *Shewanellaceae* sp., a marine bacterium associated to fish spoilage (Satomi 2014).The presence of *Shewanellaceae* was inconsistent with the replicates, and the samples were standing as clear outliers on all tests even when all the *Shewanellaceae* OTUs were removed (data not shown). After cut-offs, a total of 6,457,333 reads distributed in 3179 OTUS was available for analysis. Total reads for each taxon level can be found in Online Resource 2.

The most common phyla found was Proteobacteria (39.32% of reads), followed by Actinobacteria (30.69%), Firmicutes (12.83%), and Acidobacteria (5.55%). Other phyla were less prevalent, together amounting to 11.59% of the reads (Fig. S1). The three different sampling locations could easily be discriminated at the phylum level, with Londrina showing a higher proportion of Actinobacteria, Ponta Grossa showing more Proteobacteria, and Marechal Cândido Rondon presenting more Firmicutes representatives.

Multivariate ordination and SIMPER tests

The Non-Metric Multidimensional Scaling (NMDS) plot at the OTU level (Fig. 2) shows a clear clustering according to location, but no patterns due to the inoculation treatments were observed. Using the Similarity Percentage (SIMPER) test, we quantified the differences on community composition in each treatment-control comparison, easily sorting taxas that would be the most dissimilar (Clarke 1993). The SIMPER value dissimilarity, a percentage of the differences in taxa composition from the treatment-control pair, was used as an indicative of the impact of inoculation, and thus as a snapshot of the invasion process by the strain. Treatment L3 presented the highest SIMPER value at both Phyla (34.68%) and OTU level (59.89%), while condition P4 presented the lowest SIMPER value at Phyla level (11.26%) and condition M4 the lowest SIMPER value at OTU level (26.96%) (Fig. S2). The highest SIMPER values were not always presented by the same inoculant, suggesting that the extent of their impact on the environment depends on the local community. There were no correlations between SIMPER dissimilarities and increases in grain yield due to inoculation (Fig. 3), rejecting one of our hypothesis. The OTUs that were most impacted by inoculation belong to the *Bacillus*, *Burkholderia*, *Pseudomonas,* and *Pseudonocardia* genera, and also to the *Enterobacteriaceae* family (Online Resource 3).

Multivariate ordination on SIMPER tests

As the intensity of the changes in the bacterial rhizospheric community could not explain differences regarding productivity and PGP efficiency (Fig. 3), we considered that the identities of these changes might explain such differences. However, as it was difficult to extract meaningful information from such results (Online Resource 4) due to the multiple experimental factors and taxa, we devised the SIMPER-Principal Component Analysis (PCA) plot (Fig. 4) to facilitate such analysis. The SIMPER-PCA plot does not show the composition of the communities such as the NMDS (Fig. 2), but instead it shows the differences each inoculant has to its appropriate controls. Strictly speaking, it simply shows

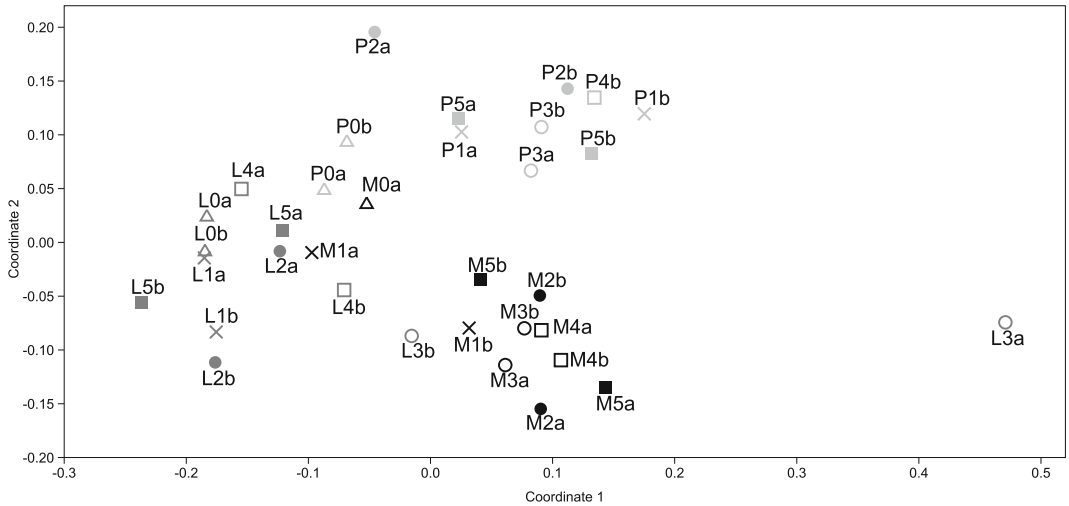

Fig. 2 Non-Metrical Multidimensional Scaling (NMDS) of OTUs based on Bray-Curtis distance. The three different sampling locations clustered together while the five different inoculation treatments did not group. Different colors show samples from different locations: dark grey = Londrina (L); light grey = Ponta Grossa (P); black = Marechal Cândido Rondon (M). Empty triangle = bulk soil from pre-planting condition (0); X = non-inoculated control (1); full circle = *Azospirillum brasilense* Ab-V5 (2); empty circle = *Achromobacter* sp. VC36 (3); empty square = *Pseudomonas* sp. 4311 (4); and full square = *Pseudomonas* sp. 4312 (5). a = first replicate; b = second replicate

the data from Online Resources 3 and 4 in a more graphical way. As a regular PCA, it shows linear correlations drawn from eigenvalues and eigenvectors of a covariance matrix. Figure 4 shows the SIMPER-PCA results at OTU level. Treatments L2, L4, and L5 from Londrina location clustered very closely, showing that most differences to control were associated to the presence of *Comamonadaceae* family OTU representatives. Treatment L3, however, had different dissimilarities to the control at Londrina, mostly associated to the presence of *Burkholderia* and *Enterobactereaceae* OTU representatives. Conditions P2, P3, and P5 from Ponta Grossa location also clustered together, with differences more associated to the presence of Betaproteobacteria, *Pseudonocardia,* and *Micrococcaceae* OTU representatives, and the absence of Enterobacteriaceae representatives. Treatment P4, however, also had a different set of differences, more highly associated to the presence of a *Pseudomonas* sp. OTU. In the sense of impacting the environment in a different manner than other inoculants at the same location, condition P4 is similar to L3, but with a lower SIMPER value (38.03%). Although data points from Marechal Cândido Rondon location were more closely clustered than those from other locations, conditions M2 and M5 were separated from conditions M3 and M4 by the 2° Principal Component (PC). All this information can also be extracted from Online Resource 4, but it is much more difficult to notice such correlations and associations analyzing data from that Table directly.

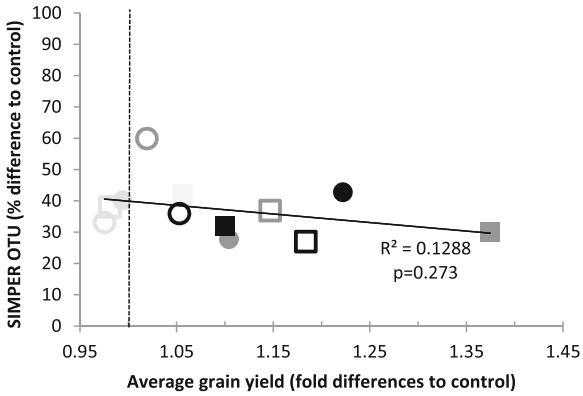

Fig. 3 Fold changes in average crop yield due to inoculation correlated to average treatment-control SIMPER dissimilarities at OTU level. R^2 and *p* values of the correlation are shown next to regression line, and the dashed line represents non-inoculated control yield. Different colors show samples from different locations: dark grey = Londrina; light grey = Ponta Grossa; black = Marechal Cândido Rondon; full circle = *Azospirillum brasilense* Ab-V5; empty circle = *Achromobacter* sp. VC36; empty square = *Pseudomonas* sp. 4311; and full square = *Pseudomonas* sp. 4312

Although the dataset was better explored with the SIMPER-PCA plot, changes in the identities of the bacterial rhizospheric community due to invasion could not explain PGP efficiency. Treatment L5, that had remarkable PGP efficiency, had no remarkable

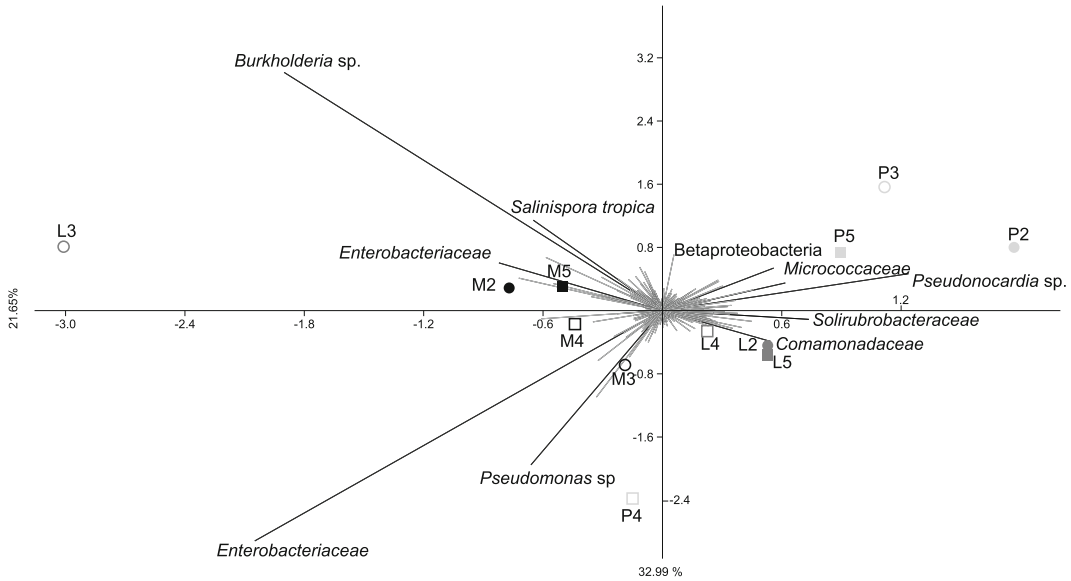

Fig. 4 SIMPER-PCA approach based on OTU taxa level showing differences in treatment-control pairs. In this plot, a cluster of samples present a similar set of differences compared to the control, while a sample far from the cluster have a different set of differences to the control. Each line represents an OTU, with dark grey explaining more variance than the light grey lines. Objects plotted in the direction of the lines presented more representatives of those particular taxa than the control, and those in the opposite direction of the line presented less representatives of that taxa than the non-inoculated control. Different colors show samples from different locations: dark grey = Londrina (L); light grey = Ponta Grossa (P); black = Marechal Cândido Rondon (M); full circle = *Azospirillum brasilense* Ab-V5 (2); empty circle = *Achromobacter* sp. VC36 (3); empty square = *Pseudomonas* sp. 4311 (4); and full square = *Pseudomonas* sp. 4312 (5)

positioning in the plot or association to taxa dissimilarities in this plot.

Analysis of pre-planting conditions

Soil chemical and nutrient analysis for each soil sampling at pre-planting condition and the nutrient per diversity ratio, are shown in Table 1. The nutrient per diversity ratio, compared to SIMPER values (both at OUT level), is shown on Fig. 5. It is clear that there was no correlation, suggesting that higher control-treatment SIMPER differences, our measure of inoculant invasion impact, are independent of the nutrient per diversity ratio. Similarly, the pre-planting nutrient per diversity ratio was not correlated to fold increases in grain yield due to inoculation (Fig. 6).

The SIMPER analysis was also used on pre-planting conditions at OTU level (Online Resource 5), using Londrina as a standard compared to Marechal Cândido Rondon and Ponta Grossa locations (39.05% and 38.43% dissimilarity, respectively). The 20 most dissimilar OTUs from both comparisons showed that Londrina had more *Actinobacteria* representatives, especially from the *Thermoleophilia* Class, and less representatives of the Bacillaceae family.

Discussion

The SIMPER-PCA approach: Visualizing control-treatment dissimilarities

The clustering of the OTUs by location observed in NMDS (Fig. 2) was largely expected, as geographic distance is one of the key limiting factors for bacterial composition in different environments (Hanson et al. 2012). Although we cannot take more conclusions from these plots with our experimental design due to the straightforwardness of clustering per community identity, the SIMPER results (Online Resource 3) provide an extensive description of the treatment-control differences, which could be correlated to PGP effectiveness. Several publications, including studies of invasions, reported combined analysis of the tabled SIMPER results and ordination methods or permutation tests (Kassen et al. 2000; Forrest and Taylor 2002; Rees

Table 1 Pre-planting (0) soil chemical characterization, nutrients, and Shannon diversity for raw values (top), raw nutrient per diversity ratios (middle) and normalized nutrient per diversity ratios (bottom)

Location	Clay	OM	P	K	Ca	Mg	S	Zn	Cu	B	Mn	pH	Shannon diversity (OTU)
	%		mg.dm-^{3}		(nmolc.dm-^{3})								
L0	5.9	3.3	6.4	339	6.7	2	11	3.2	13	0.5	30	5.7	5.891
M0	6	3.2	11	173	5	2.6	25	0.9	4.1	0.7	25	5.4	6.25
P0	2.2	7.2	7.9	292	5.3	3.6	15	4.1	0.4	0.4	7	6.1	5.9775
Location	Soil Nutrients / Shannon												
L0	1.00	0.56	1.08	57.54	1.13	0.34	1.86	0.54	2.20	0.08	5.09	–	
M0	0.96	0.51	1.76	27.68	0.80	0.41	4.00	0.14	0.65	0.11	4.00	–	
P0	0.36	1.20	1.32	48.85	0.88	0.60	2.50	0.68	0.06	0.06	1.17	–	
													Average per location
Location	Normalized soil nutrient / Shannon (relative percentages)												(nutrient per diversity ratio)
L0	0.43	0.24	0.26	0.42	0.40	0.25	0.22	0.39	0.75	0.32	0.49	–	0.383
M0	0.41	0.22	0.42	0.20	0.28	0.30	0.47	0.10	0.22	0.42	0.39	–	0.316
P0	0.15	0.52	0.31	0.36	0.31	0.44	0.30	0.50	0.02	0.25	0.11	–	0.301

et al. 2004; Mills et al. 2006; Margherita and Osborne 2009; Thomas et al. 2012; Gitipour et al. 2013; Wilkins et al. 2013; Stevens and Olson 2013; Wang et al. 2014; Purcell et al. 2014; Freedman et al. 2015). However, to the best of our knowledge, this is the first attempt to visualize SIMPER results on ordination methods.

The SIMPER-PCA approach (Fig. 4) facilitated the interpretation of the SIMPER results (Online Resources 3 and 4), and allowed formulating new hypotheses that were not evident on NMDS (Fig. 2). Since this approach can visualize the difference related to a defined standard (such as the control), potentially it may be applied to a large number of studies that rely on cluster analysis from ordination methods. Three issues must be addressed to apply this approach: (i) the SIMPER test is based on the Bray-Curtis distance, so for consistency it is appropriate to use this distance on other tests (such as ANOSIM, PERMANOVA or NMDS) (Wang et al. 2014); (ii) Taxa frequency should be square root transformed to adjust for weights of dominant taxa (Clarke and Warwick 2001); and (iii) the SIMPER dissimilarities per taxa

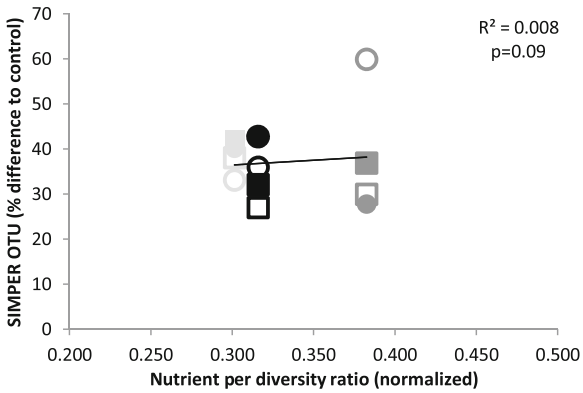

Fig. 5 Nutrient per diversity ratios of pre-planting conditions correlated to SIMPER values of control-treatments pairs after inoculation, at OTU level. Different colors show samples from different locations: dark grey = Londrina; light grey = Ponta Grossa; black = Marechal Cândido Rondon; full circle = *Azospirillum brasilense* Ab-V5; empty circle = *Achromobacter* sp. VC36; empty square = *Pseudomonas* sp. 4311; and full square = *Pseudomonas* sp. 4312

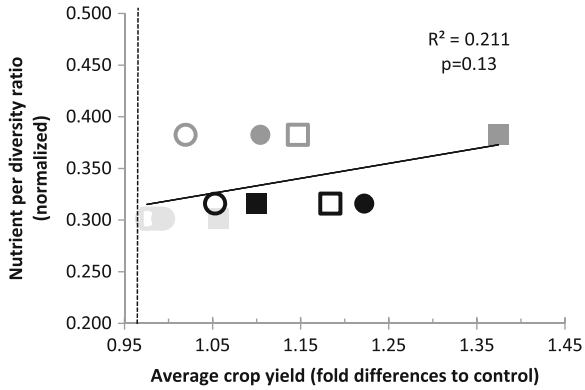

Fig. 6 Fold changes in average crop yield due to inoculation correlated to nutrient per diversity ratio at OTU level. R^2 and *p* values of the correlation are shown next to regression line, and the dashed line represents non-inoculated control yield. Different colors show samples from different locations: dark grey = Londrina; light grey = Ponta Grossa; black = Marechal Cândido Rondon; full circle = *Azospirillum brasilense* Ab-V5; empty circle = *Achromobacter* sp. VC36; empty square = *Pseudomonas* sp. 4311; and full square = *Pseudomonas* sp. 4312

must be multiplied by −1 when the average taxa abundance of the treatment is smaller than that of standard. This is further explained in Supplementary Information.

The SIMPER data could potentially be used on other ordination methods. Categorical Principal Component Analysis (CatPCA) can evaluate non-linear correlations with ordinal and nominal data (Linting et al. 2007); and Canonical Correspondence Analysis (CCA) can include explanatory environmental variables (Ramette 2007) – or even use the SIMPER data as the explanatory variables. Care must be taken to avoid constructing dissimilarity plots based on the dissimilarity data, like a Bray-Curtis NMDS with the SIMPER data, because such approach will generate artifacts (Fig. S3). Applying our approach, the control samples will not appear on the SIMPER-PCA plot, as they are used as the standard.

Impact on community does not predict PGP effectiveness

The observed phyla distribution (Fig. S1) was typical for rhizospheric soil (Berg et al. 2013), so we can assume that the bacterial community composition was not unusual. The impact of inoculation on the rhizospheric native bacterial community has already been studied to some detail (Gilbert et al. 1996; Ambrosini et al. 2016). This impact, and also PGP effectiveness, is known to depend on many factors (Castro-Sowinski et al. 2007; Owen et al. 2015). Evaluation of invasion ability as observable community impact has also already been considered (Parker et al. 1999), but, to the best of our knowledge, SIMPER differences have not been used to measure invasion impact.

Based on the SIMPER-PCA approach (Fig. 4), it can be observed that no single inoculant induced the same set of changes in the different locations, which would characterize a consistent invasion effect. Also, based on the SIMPER-PCA approach, we raised the hypothesis that treatment L3 would have a higher agronomical productivity, since it was the treatment with the highest SIMPER value and also had a different set of differences from the other inoculants at the same location. Since treatment P4 also had a different set of alterations to control when compared to other inoculation treatments at the same location, we raised the hypothesis that treatment P4 could also have a higher productivity, compared to treatments in the same location. Finally, since the samples from Marechal Cândido Rondon location were divided by the 2° PC, we raised the third hypothesis that this division would be also noticed on agronomical productivity.

None of the three hypotheses mentioned above was supported by field data, as can be seen on Fig. 1. Only treatment L5 had a significant plant growth promotion effect, and it showed no remarkable features on the SIMPER-PCA plot. This suggests that unique changes in the bacterial rhizospheric communities due to inoculation were not associated to grain yield. Likewise, higher SIMPER values also did not correlate to higher grain yield by inoculation, as seen on Fig. 3. While it can be concluded that the inoculant that caused the highest impact was not the best PGPB, we must also consider that the invasion impact measurement in this work, the SIMPER differences in control-treatment pairs, is oversimplified. There is much active research on the traits of invasive species and definitions of invisibility (Davis et al. 2000; Davis and Pelsor 2001), that could be applied to microbiological invasions, but were not explored here. It is also possible that all strains used as inoculants had very poor survival in the fields, preventing them from exerting more detectable impacts in the community; this would explain why we could not detect their genera in larger proportions in the NGS data (Online Resource 2). In addition, crop yield improvements of inoculant 5 at Londrina, but not at other locations, are not enough evidence to support inoculant 5 as an efficient PGP in this work.

More accessible niches do not predict PGP efficiency, but specific antagonists may play a role

Our expectation was that a higher nutrient per diversity ratio, which represent more available niches, could facilitate inoculant survival, plant colonization or display PGP traits due to a lower competition in the rhizosphere (Compant et al. 2010), inducing a higher crop yield. However, there were no correlations between the pre-planting nutrient per diversity ratio and higher SIMPER values or improved crop yield (Fig. 6). This suggests that this ratio, as obtained in this work, is not enough to predict inoculant invasion impact or PGP efficiency. However, our data is not sufficient to suggest that inoculant addition in field trials will not be subjected to niche and nutrient competitions constrains. The results that we observed could be due to the fact that the inoculants were not very effective in promoting plant growth overall, that native communities were well

adapted and prevented invasion (Mallon et al. 2015), or that existing biological effects were too subtle to be detected with our sampling design.

Nonetheless, the Londrina location had the highest nutrient per diversity ratio and was the single location with PGP efficiency. Since we could see that these variables are not correlated, another factor should be responsible for this effect. The known PGP abilities of the *Pseudomonas* sp. 4312 strain (production of indolic compounds and P solubilization) are unlikely to be sufficient for this effect. Strains 2 and 3 also present indolic compounds production and P solubilization abilities, and if such abilities were enough to provide improved crop yield, it would be observable in other locations. Networks of microbial species can easily affect bacteria survival and activity (Zelezniak et al. 2015), so interactions of the inoculant with the local community should have played a key role on PGP effectiveness. With the SIMPER approach on pre-planting soils using Londrina as a standard (Online Resource 5), we could argue that the higher abundance of *Actinobacteria* from the Thermoleophilia Class facilitated the display of plant growth promotion by L5. Alternatively, (or in addition to) it was the lower abundance of *Bacillales* that facilitated the plant growth effect, as it might have been acting as a specific antagonist to the inoculant. Both *Actinobacteria* and *Bacillus* can act as antagonist in soils and produce antibiotics (van Elsas et al. 2012), so they have the potential to largely influence rhizosphere interactions. These synergic or antagonistic effects to inoculant 5 (*Pseudomonas* sp. 4312) could be tested in greenhouse trials, by enriching or suppressing these taxa. While supported by the literature, limitation in our sample size limits the confidence we have to propose such interactions, making additional tests necessary to provide sufficient evidence for our suggestions.

Conclusions

With the results obtained in this work, the hypothesis that the most impacting invader would be the best PGPB was rejected. Inoculant invasion impact, measured as dissimilarity of the inoculated treatments compared to the control, was not correlated to an effective PGP effect or to more invasible conditions. The pre-planting nutrient per diversity ratio was unable to predict PGP field efficiency or invasion impact. Thus, we were unable to predict PGP efficiency by sampling very young plants and pre-planting soils. We conclude that our invasion ecology approach could not anticipate field results of inoculated plants. Despite rejecting our hypothesis, we provide full raw data and extensive details of our statistical approach, which might be useful for other authors interested in focusing on treatment-control differences.

We urge that more attention is given to microbial communities before planting, especially with NGS tools. If the microbial community is to be manipulated by inoculation, its structure cannot be ignored. Considering that NGS costs are reducing drastically over time, and food prices and the need for food production increase over time, NGS might be a standard tool for farmers within less than 30 years. Just like today farmers take pH measurements of their soils to calculate the right inputs for optimal productivity, they could use NGS at pre-planting stages or before addition of inputs to calculate the proper microbial management strategy. This would be a clear, long term goal that could help direct applied research in the area.

Acknowledgements This work was financed by grant and fellowships from the Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq/Brazil) and Instituto Nacional de Ciência e Tecnologia (INCT) da Fixação Biológica do Nitrogênio (Brazil) for Brazilian researchers; Deutscher Akademischer Austauschdienst (DAAD/Germany) for PBdC; Bielefeld Young Researchers Fund for SBC. The bioinformatics support of the BMBF-funded project "Bielefeld-Giessen Center for Microbial Bioinformatics – BiGi (Grant number031A533)" within the German Network for Bioinformatics Infra-structure (de.NBI) is gratefully acknowledged.

Compliance with ethical standards

Conflict of interest The authors declare no conflict of interest.

References

Alves BJR, Boddey RM, Urquiaga S (2003) The success of BNF in soybean in Brazil. Plant Soil 252:1–9

Ambrosini A, de Souza R, Passaglia LMP (2016) Ecological role of bacterial inoculants and their potential impact on soil microbial diversity. Plant Soil 400:193–207. https://doi.org/10.1007/s11104-015-2727-7

Arruda L, Beneduzi A, Martins A et al (2013) Screening of rhizobacteria isolated from maize (Zea Mays L.) in Rio Grande do Sul state (South Brazil) and analysis of their potential to improve plant growth. Appl Soil Ecol 63:15–22. https://doi.org/10.1016/j.apsoil.2012.09.001

Bashan Y, De-Bashan LE, Prabhu SR, Hernandez J-P (2014) Advances in plant growth-promoting bacterial inoculant technology: formulations and practical perspectives (1998–2013). Plant Soil 378:1–33. https://doi.org/10.1007/s11104-013-1956-x

Berg G, Zachow C, Müller H et al (2013) Next-generation bio-products sowing the seeds of success for sustainable agriculture. Agronomy 3:648–656. https://doi.org/10.3390/agronomy3040648

Caporaso JG, Lauber CL, Walters WA et al (2011) Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample. PNAS 108:4516–4522. https://doi.org/10.1073/pnas.1000080107/-/DCSupplemental.www.pnas.org/cgi/doi/10.1073/pnas.1000080107

Castro-Sowinski S, Herschkovitz Y, Okon Y, Jurkevitch E (2007) Effects of inoculation with plant growth-promoting rhizobacteria on resident rhizosphere microorganisms. FEMS Microbiol Lett 276:1–11. https://doi.org/10.1111/j.1574-6968.2007.00878.x

Chowdhury SP, Dietel K, Rändler M et al (2013) Effects of bacillus amyloliquefaciens FZB42 on lettuce growth and health under pathogen pressure and its impact on the rhizosphere bacterial community. PLoS One 8:e68818. https://doi.org/10.1371/journal.pone.0068818

Clarke KR (1993) Non-parametric multivariate analyses of changes in community structure. Aust J Ecol 18:117–143

Clarke KR, Warwick RM (2001) Change in marine communities: an approach to statistical analysis and interpretation, 2nd editio. PRIMER-E, Plymouth

Compant S, Clément C, Sessitsch A (2010) Plant growth-promoting bacteria in the rhizo- and endosphere of plants: their role, colonization, mechanisms involved and prospects for utilization. Soil Biol Biochem 42:669–678. https://doi.org/10.1016/j.soilbio.2009.11.024

Davis MA, Pelsor M (2001) Experimental support for a resource-based mechanistic model of invasibility. Ecol Lett 4:421–428. https://doi.org/10.1046/j.1461-0248.2001.00246.x

Davis MA, Grime JP, Thompson K (2000) Fluctuating resources in plant communities: a general theory of invasibility. J Ecol 88:528–534

Forrest BM, Taylor MD (2002) Assessing invasion impact: survey design considerations and implications for management of an invasive marine plant. Biol Invasions 4:375–386. https://doi.org/10.1023/A:1023613428351

Freedman ZB, Romanowicz KJ, Upchurch RA, Zak DR (2015) Differential responses of total and active soil microbial communities to long-term experimental N deposition. Soil Biol Biochem 90:275–282. https://doi.org/10.1016/j.soilbio.2015.08.014

Gilbert GS, Clayton MK, Handlesman J, Parke J (1996) Use of cluster and discriminant analysis to compare rhizosphere bacterial communities following biological perturbation. Microb Ecol 32:123–147

Gitipour A, El Badawy A, Arambewela M et al (2013) The impact of silver nanoparticles on the composting of municipal solid waste. Environ Sci Technol 47:14385–14393

Haegeman B, Moriarty J, Neal P et al (2013) Robust estimation of microbial diversity in theory and in practice. ISME J 7:1092–1101. https://doi.org/10.1038/ismej.2013.10

Hammer Ø, Harper DAT, Ryan PD (2001) PAST: Paleontological statistics software package for education and data analysis. Palaeontol Electron 4(1):9

Hanson CA, Fuhrman JA, Horner-Devine MC, Martiny JBH (2012) Beyond biogeographic patterns: processes shaping the microbial landscape. Nat Rev Microbiol 10:497–506. https://doi.org/10.1038/nrmicro2795

Hierro JL, Maron JL, Callaway RM (2005) A biogeographical approach to plant invasions: the importance of studying exotics in their introduced and native range. J Ecol 93:5–15. https://doi.org/10.1111/j.1365-2745.2004.00953.x

Hungria M, Campo RJ, Souza EM, Pedrosa FO (2010) Inoculation with selected strains of Azospirillum brasilense and A. Lipoferum improves yields of maize and wheat in Brazil. Plant Soil 331:413–425. https://doi.org/10.1007/s11104-009-0262-0

Kassen R, Buckling A, Bell G, Rainey P (2000) Diversity peaks at intermediate productivity in a laboratory microcosm. Nature 406:508–512

Linting M, Meulman JJ, Groenen PJF, Van der Kooij AJ (2007) Nonlinear principal components analysis: introduction and application. Psychol Methods 12:336–358. https://doi.org/10.1037/1082-989X.12.3.336

Mallon CA, Elsas JD Van, Sales JF (2015) Microbial invasions : the process, patterns, and mechanisms. Trends Microbiol 23:719–729

Margherita G, Osborne B (2009) Assessing the impact of plant invasions on soil seed bank communities : use of univariate and multivariate statistical approaches. J Veg Sci 20:547–556. https://doi.org/10.1111/j.1654-1103.2009.01054.x

Mills DK, Entry JA, Voss JD, et al (2006) An assessment of the hypervariable domains of the16S rRNA genes for their value in determining microbial community diversity : the paradox of traditional ecological indices. 57:496–503. doi:https://doi.org/10.1111/j.1574-6941.2006.00135.x

Ondov BD, Bergman NH, Phillippy AM (2011) Interactive metagenomic visualization in a web browser. BMC Bioinf 12:385. https://doi.org/10.1186/1471-2105-12-385

Otieno BN, Culhane J, Germaine KJ et al (2013) Screening of large collections of plant associated bacteria for effective plant growth promotion and colonisation. Asp Appl Biol 120:23–28

Owen D, Williams AP, Griffith GW, Withers PJA (2015) Use of commercial bio-inoculants to increase agricultural production through improved phosphrous acquisition. Appl Soil Ecol 86:41–54. https://doi.org/10.1016/j.apsoil.2014.09.012

Parker IM, Simberloff D, Lonsdale WM et al (1999) Impact : toward a framework for understanding the ecological effects of invaders. Biol Invasions 1:3–19

Purcell SW, Clarke KR, Rushworth K, Dalton SJ (2014) Defining critical habitats of threatened and endemic reef fishes with a multivariate approach. Conserv Biol 28:1688–1698. https://doi.org/10.1111/cobi.12343

Ramette A (2007) Multivariate analyses in microbial ecology. FEMS Microbiol Ecol 62:142–160. https://doi.org/10.1111/j.1574-6941.2007.00375.x

Rees GN, Baldwin DS, Watson GO, Perryman S (2004) Ordination and significance testing of microbial community composition derived from terminal restriction fragment length polymorphisms: application of multivariate statistics. Antonie Van Leeuwenhoek 86:339–347

Ryan RP, Germaine K, Franks A et al (2008) Bacterial endophytes: recent developments and applications. FEMS Microbiol Lett 278:1–9. https://doi.org/10.1111/j.1574-6968.2007.00918.x

Satomi M (2014) The family Shewanellaceae. In: Rosenberg E, Delong EF, Thompson F (eds) The prokaryotes, 4th edn. Springer, Berlin, pp 597–619

Sparks DL, Page AL, Helmke PA, Loeppert RH (1996) Methods of soil analysis: part 3. Chemical methods and processes. Soil Science Society of America, Madison, USA

Stevens JL, Olson JB (2013) Invasive lionfish harbor a different external bacterial community than native Bahamian fishes. Coral Reefs 32:1113–1121. https://doi.org/10.1007/s00338-013-1072-7

Thomas T, Gilbert J, Meyer F (2012) Metagenomics - a guide from sampling to data analysis. Microb Inf Exp 2:3. https://doi.org/10.1186/2042-5783-2-3

van Elsas JD, Chiurazzi M, Mallon CA et al (2012) Microbial diversity determines the invasion of soil by a bacterial pathogen. Proc Natl Acad Sci U S A 109:1159–1164. https://doi.org/10.1073/pnas.1109326109

Wang J, Linnenbrink M, Künzel S, et al (2014) Dietary history contributes to enterotype-like clustering and functional metagenomic content in the intestinal microbiome of wild mice. PNAS 111:1–8. doi:https://doi.org/10.1073/pnas.1402342111

Wilkins D, Lauro FM, Williams TJ et al (2013) Biogeographic partitioning of Southern Ocean microorganisms revealed by metagenomics. Environ Microbiol 15:1318–1333. https://doi.org/10.1111/1462-2920.12035

Zelezniak A, Andrejev S, Ponomarova O et al (2015) Metabolic dependencies drive species co-occurrence in diverse microbial communities. Proc Natl Acad Sci U S A 112:E7156. https://doi.org/10.1073/pnas.1522642113