Genome variations between rhizosphere and bulk soil ecotypes of a Pseudomonas koreensis population: Ecology of P. koreensis in soil habitats

Article *in* Environmental Microbiology · July 2018

DOI: 10.1111/1462-2920.14363

Genome variations between rhizosphere and bulk soil ecotypes of a *Pseudomonas koreensis* population

Lucas Dantas Lopes[a,b], Michele de C. Pereira e Silva[a], Alexandra J. Weisberg[b], Edward W. Davis II[b,c], Qing Yan[b], Camila de S. Varize[a], Chih-Feng Wu[b], Jeff H. Chang[b,c], Joyce E. Loper[b,c], Fernando D. Andreote[a,1]

a. Department of Soil Science, "Luiz de Queiroz" College of Agriculture, University of São Paulo, Piracicaba, SP, Brazil.

b. Department of Botany and Plant Pathology, Oregon State University, Corvallis, OR, 97331, United States of America.

c. Molecular and Cellular Biology Program, Oregon State University, Corvallis, OR, 97331, United States of America.

[1]Corresponding author: Fernando Dini Andreote, Department of Soil Science, "Luiz de Queiroz" College of Agriculture, University of São Paulo. Avenida Padua Dias, 11 – Piracicaba, SP, Brasil. 13420-705. Phone: +55(19)34172123 – E-mail: fdandreo@gmail.com

**Running title:* Ecology of *P. koreensis* in soil habitats

This is the peer reviewed version of the following article: "Lopes, L. D., Pereira e Silva, M. C., Weisberg, A. J., Davis, E. W., Yan, Q., Varize, C. S., Wu, C. F., Chang, J. H., Loper, J. E., Andreote, F. D. (2018). Genome variations between rhizosphere and bulk soil ecotypes of a Pseudomonas koreensis population. Environmental Microbiology 20(12) 4401-4414.", which has been published in final form at "https://onlinelibrary.wiley.com/doi/full/10.1111/1462-2920.14363". This article may be used for non-commercial purposes in accordance with Wiley Terms and Conditions for Use of Self-Archived versions.

Abstract

Bulk soil and rhizosphere are soil compartments selecting different microbial communities. However, it is unknown whether this selection also can change the genome content of specific bacterial taxa, splitting a population in distinct ecotypes. To answer this question we compared the genome sequences of 53 isolates obtained from sugarcane rhizosphere (28) and bulk soil (25). These isolates were previously classified in the *Pseudomonas koreensis* subgroup of the *P. fluorescens* complex. Phylogenomics showed a trend of separation between bulk soil and rhizosphere isolates. Discriminant analysis of principal components (DAPC) identified differences in the accessory genome of rhizosphere and bulk soil sub-populations. We found significant changes in gene frequencies distinguishing rhizosphere from bulk soil ecotypes, *e.g*. enrichment of phosphatases and xylose utilization (*xut*) genes, respectively. Phenotypic assays and deletion of *xutA* gene indicated that accumulation of *xut* genes in the bulk soil sub-population provided a higher growth capacity in a D-xylose medium, supporting the corresponding genomic differences. Despite the clear differences distinguishing the two ecotypes, all 53 isolates were classified in a single 16S rRNA gene OTU. Collectively, our results revealed that the gene pool and ecological behavior of a bacterial population can be different for ecotypes living in neighboring soil habitats.

Key words: population genomics; niche partitioning; fluorescent *Pseudomonas* spp.; evolutionary ecology; soil bacteria

Introduction

Soil is a heterogeneous environment for microbial life, since its physical and chemical characteristics vary in space and time. Microbes thrive in certain soil micro-habitats, for example, attached to aggregate surfaces, in the detritusphere, biopores, and also in niches provided by plants, such as the rhizosphere (Kuzyakov *et al*., 2015). In the rhizosphere, roots drastically change the physical (oxygen levels, aggregation, water content) and chemical (pH, carbon supply) properties of the surrounding soil (Sokolova, 2015). The abundance of low molecular weight carbon compounds in the rhizosphere differentiates this habitat from the bulk soil, which has relatively low readily-available organic content to sustain microbial activity (Dennis *et al*., 2010). Therefore, bulk soil and rhizosphere are neighboring soil compartments without a geographic barrier comprising distinct niches for microbial life.

Studies of soil microbial communities have advanced our knowledge in recent years, revealing a different microbiome assemblage in bulk soil compared to the rhizosphere (Berg and Smalla, 2009; Uroz *et al*., 2010; Mendes *et al*., 2014; Lopes *et al*., 2016), which is enriched in microbes able to promote plant health and growth (Weller *et al*., 2002; Mendes *et al*., 2011: Kwak *et al*., 2013; Phillipot *et al*., 2013). However, little is known about the evolutionary ecology of specific bacterial populations inhabiting these soil compartments. For eukaryotes, population is defined as a set of genetically distinct individuals of the same species coexisting in an environment (Prosser *et al*., 2007). Due to the complexities and controversies for species definition for prokaryotes (Cohan, 2002; Cordero *et al*., 2014), a bacterial population is considered to be a group of coexisting individuals highly clustered at the genotypical and phenotypical level, classified in the same or closely related species (Cordero *et al*., 2014). Ecotypes, in turn, are members of a population optimized by selection for a particular niche, *i.e*. an ecological sub-population (Cohan, 2002; Prosser *et al*., 2007; Cordero *et al*., 2014). While 16S rRNA high-throughput sequencing is the standard method for bacterial community analyses, comparative genomics of cultured strains is a powerful tool to assess the evolutionary ecology of bacterial populations (Falush, 2009).

Pseudomonas spp. is a genus of common soil bacteria containing potential target species for studies of comparative genomics aiming to understand the ecology of a bacterial population in soil habitats, because it is widespread in soils, several species are easily cultured, and some studies suggest that its community is distinct in rhizosphere compared to bulk soil (Costa *et al*., 2007; Garcia-Salamanca *et al*., 2012). In addition, many *Pseudomonas* species – especially those in the *P. fluorescens* complex – contain strains able to support plant growth and health (Mercado-Blanco *et al*., 2007; Loper *et al*., 2012; Hol *et al*., 2013). *Pseudomonas koreensis* is a subgroup of closely-related species inside the *P. fluorescens* complex, containing strains able to suppress plant diseases and help in soil bioremediation (Hultberg *et al*., 2010; Babu *et al*., 2015; Rafikova *et al*., 2016). Thus, knowledge about the ecology of *P*. *koreensis* populations has environmental and agricultural relevance.

A previous study, based on 16S rRNA gene high-throughput sequencing showed that the microbiomes of sugarcane rhizosphere and bulk soil are distinct (Lopes *et al*., 2016). Here, we firstly analyzed the OTUs classified as *Pseudomonas* spp. in that study to check whether the community of this important and large group of bacteria is also distinct in sugarcane rhizosphere compared to bulk soil. Then, we focused in a collection of 53 *P*. *koreensis* strains previously isolated from the same sugarcane rhizosphere and bulk soil samples (Lopes *et al*., 2018), to test the hypothesis that a bacterial population can harbor different ecotypes in neighboring soil habitats, differing in genome content and enriched in niche-specific genes/functions.

Results

Shifts in the communities of *Pseudomonas* spp. from rhizosphere and bulk soil

Our first evaluation was to check whether the whole community of bacteria belonging to the *Pseudomonas* genus was different in bulk soil and rhizosphere of sugarcane. A total of 134 OTUs comprising 601 reads classified as *Pseudomonas* were retrieved from a high-throughput sequencing of the 16S rRNA genes amplified from the DNA directly extracted from bulk soil and rhizosphere samples. These reads are a subset of the sequencing performed for a microbiome analysis showing that total bacterial community is different between sugarcane rhizosphere and bulk soil (Lopes *et al*., 2016). Multivariate analysis clearly indicated that the community structure of *Pseudomonas* spp. is also significantly different between these habitats (Fig. 1A), supported by analysis of similarity (ANOSIM; P=0.002; R=0.74). The ordination of samples in the non-metric multidimensional scaling (NMDS) also showed that rhizosphere communities are more similar between replicates than are bulk soil communities (Fig. 1A).

The abundance of *Pseudomonas* spp. was also significantly different in rhizosphere and bulk soil. Two approaches were used to measure this differential abundance. In the first, we compared the relative proportion of sequences classified as *Pseudomonas* spp. from the microbiome sequencing, which revealed a significantly higher abundance of *Pseudomonas* spp. in the rhizosphere compared to bulk soil (P=0.002, Fig. 1B). In the second approach, we estimated the quantity of 16S rRNA gene copies specific to *Pseudomonas* spp. by qPCR (Bergmark et al., 2013), which resulted in an average of $4.2\text{x}10^3$ and $1.5\text{x}10^5$ copies per gram of soil in the bulk soil and rhizosphere samples, respectively (Fig. 1C, P=0.002). Although technical limitations such as uneven PCR amplification and primer selection could alter the results of individual analyses, the results from both approaches were similar. Together the results confirmed that *Pseudomonas* spp. community in sugarcane rhizosphere and bulk soil has a different structure and abundance, suggesting that specific members of this community are favored in each habitat. If the two habitats select distinct *Pseudomonas* taxa, it is possible that they also select distinct ecotypes with different gene pools inside some of these taxa. To test this hypothesis, we analyzed the genome sequences of isolates comprising a population of *P. koreensis*, which was used as a model in our study.

Phylogenomic distribution of the *Pseudomonas koreensis* isolates

Our next goal was to verify whether a specific population of fluorescent *Pseudomonas* sp. has genome variations according to the soil habitat, indicative of niche partitioning between sub-populations. Then, we focused on the analysis of genome sequences of 53 fluorescent *Pseudomonas* strains cultured from the same soil samples addressed in the previous section (Lopes *et al*., 2018). Genomic sequences of these strains isolated from sugarcane rhizosphere (28 isolates) and bulk soil (25 isolates) were compared to 41 reference genomes of the genus *Pseudomonas*, including phylogenetically close (from the *P. fluorescens* group) and distant species (*e.g. P. putida*, *P. syringae*, *P. aeruginosa*). The comparison of these 94 genome sequences yielded 2,064 shared gene clusters. From this total, 1,540 were found as single copies in all 94 genome sequences, and were therefore used for a phylogenomics approach. The phylogenomic tree clustered the 53 isolates to references of the *P. koreensis* subgroup of the *P. fluorescens* complex (Fig. 2), corroborating a previously-published multilocus sequence analysis (MLSA) (Lopes *et al*., 2018). Neither approach (MLSA and phylogenomics) placed the bulk soil and rhizosphere isolates in distinct clades. Three major clades were evident in the phylogenomic tree (Fig. 2). One clade was composed primarily of rhizosphere isolates; a second was composed entirely of rhizosphere isolates; and the third was composed primarily of bulk soil isolates. However, in the phylogenomic tree we observe a higher trend of separation between rhizosphere and bulk soil isolates than MLSA, indicating that the core genome of this population (gene clusters shared by all isolates) have some degree of divergence between rhizosphere and bulk soil habitats (Fig. 2). The slight difference observed between the two approaches (MLSA x phylogenomics) is possibly a result of the genetic markers used for the phylogenetic inferences. While MLSA of the previous study used only few, highly conserved markers (four to seven housekeeping genes), the present phylogenomics approach used more than 1,000 markers, which probably include genes under positive selection that contributed to the higher separation of rhizosphere and bulk soil strains in the phylogenetic tree.

In order to identify the analyzed genomes in the *Pseudomonas* spp. community, the 16S rRNA gene sequences were extracted from the 53 genome sequences and compared to the sequences of the 134 *Pseudomonas* spp. OTUs from the community analysis. The 16S rRNA gene sequence of the reference *P. fluorescens* Pf0-1 was also included in the analysis. Separation of OTUs using UCLUST indicated that all the 53 isolates are comprised in a single OTU, which also contained the sequences of one community OTU and the reference *P. fluorescens* Pf0-1. The close relationship of these 53 isolates in a single OTU confirmed the results of phylogenomics (Fig. 2). Therefore, the following comparison of 53 isolates cultured from bulk soil or rhizosphere corresponds to a single population of closely related individuals, which was detected in the culture-independent approach as a member of the *Pseudomonas* spp. community.

Aiming to support our ecological interpretations, we assessed the pan-genome of this population. Results showed that its pan-genome is open ($\gamma = 0.138$, $\sigma = 5707.900$), but the estimated parameter γ is close to 0, suggesting that the 53 genome sequences analyzed represent a high proportion of the genetic diversity in that population (Fig. 3A). The core genome was estimated as Ω=4,166.581, which could be defined using only 10 genome sequences (Fig. 3B). Therefore, the results suggest that the number of isolates used in our approach was enough to represent the *P. koreensis* population studied and appropriate for the ecological inferences drawn in the following sections.

Genome variations between ecotypes inhabiting rhizosphere or bulk soil

The next step of our study was to determine if the bacterial population studied is split into sub-populations corresponding to rhizosphere and bulk soil ecotypes, and to identify the genes and functions enriched in each ecotype. Using Get Homologues, analysis of genome sequences resulted in 4,200 and 4,289 shared gene clusters in the ecotypes of bulk soil and rhizosphere, respectively (Table 1). The number of accessory gene clusters was slightly higher in the bulk soil (4,246) than in the rhizosphere ecotypes (3,845) (Table 1). When analyzing the 53 isolates together, we identified a total of 9,824 gene clusters, from which 4,124 are shared by all isolates (core genome) and 5,700 comprise the accessory genome (Table 1). The number of core genes estimated using this approach was consistent to the number using non-linear regression, described in the previous section. Regarding the accessory genome, we observed ~800 gene clusters unique to one strain when analyzing the rhizosphere and bulk soil isolates separately (Fig. S1A and B), and ~1,200 gene clusters unique to one strain when the 53 isolates were analyzed together (Fig. S1C). The number of accessory gene clusters present in three or more strains (referred as "shell") was in general less than 500 in both cases (Contreras and Vinuesa, 2013).

We used Discriminant Analysis of Principal Components (DAPC) to check whether the ecotypes from rhizosphere versus bulk soil have genomic distinctions with regard to the accessory genome of the population (gene clusters not shared by all isolates). DAPC resulted in the separation of rhizosphere and bulk soil sub-populations, confirming that some factors (gene clusters) differentiate the ecotypes of the two habitats (Fig. 4A). However, it was necessary to reach the saturation of variance (20 eigenvalues) to separate the ecotypes from each habitat, indicating that only specific gene clusters were differently distributed between the bulk soil and rhizosphere sub-populations (Fig. 4A). DAPC also suggested that the bulk soil sub-population has a higher genetic variation (more peaks in the graphs), supporting the results of Get Homologues (Fig. 4A).

Additional statistical analyses based on Welch's t-test were applied to identify the gene clusters and the molecular functions (GO-terms) enriched in the ecotypes of each habitat. Forty-eight gene clusters showed significant differences, from which 32 and 16 were higher in the bulk soil and rhizosphere ecotypes, respectively (P<0.05) (Fig. 4B). Similarly, 29 and 22 GO-terms were significantly higher (P<0.05) in the ecotypes of rhizosphere and bulk soil (Fig. S2).

Among the functions and gene clusters differentially enriched in each sub-population, some were identified as potentially related to the occupation of the rhizosphere or bulk soil habitats. For example, genomes of the rhizosphere sub-population were enriched for genes coding for alkaline and acid phosphatases. On the other hand, the ecotypes of bulk soil sub-population were highly enriched for genes associated with xylose utilization (Fig. 4B). Consistently, the GO-terms related to the metabolism of organic phosphate or to xylose utilization were enriched in the rhizosphere and bulk soil ecotypes, respectively (Fig. S2). The five xylose utilization genes enriched in the bulk soil ecotypes form the xylose isomerase operon: *xutA, xutR, xutG, xutH* and *xutF* (Liu *et al*., 2015). The acid and alkaline phosphatases genes enriched in the rhizosphere ecotypes were annotated as hypothetical proteins. They were detected based on the occurrence of conserved domains for phosphatase, and thus were considered as putative new phosphatases yet to be properly described. A strong phylogenetic signal was observed for phosphatases and *xut* genes, which were totally separated in distinct phylogenomic clades. The clades containing phosphatases or *xut* genes were dominated by rhizosphere and bulk soil isolates, respectively (Fig. 2).

Besides the highlighted gene clusters/GO-terms, other functions were significantly different between the sub-populations (Fig. S2). GO-terms with regard to polysaccharides catabolism/transport were enriched in the bulk soil ecotypes (Starch binding; Hydrolase activity, acting on glycosyl bonds; Polysaccharide transmembrane transporter activity; Polysaccharide transport; and Hydrolase activity, acting on O-glycosyl compounds). Also, many other functions not immediately recognizable as associated to bulk soil or rhizosphere habitats were revealed providing references for future investigations.

We also generated a heatmap of gene clusters' presence/absence, which revealed a clustering of isolates resembling that of phylogenomics (Fig. S3A). The different presence and absence of genes between sub-populations could be a result of gene losses and/or horizontal gene transfer (HGT) followed by recombination. Aiming to understand the influence of recombination in the differentiation of ecotypes, we used some programs for the prediction of mobile genetic elements and HGT in the 53 genome sequences. Several predicted genomic islands, transposons, insertion sequences, prophages and horizontally transferred genes were detected in the rhizosphere and bulk soil isolates. The number of hits found between the rhizosphere and bulk soil sub-populations was significantly different for some of the predictions (Fig. S3B), suggesting that HGT and recombination could have contributed to ecotypes differentiation.

Higher growth on D-xylose by the bulk soil isolates supports genomic differences

Comparative genomics showed that *xut* genes - enriched in bulk soil isolates - represent ~10% of the significant differences between the ecotypes of each habitat. These genes were already characterized in *P. fluorescens* (Liu *et al.*, 2015). Thus, the xylose utilization function was further analyzed in order to check this difference at the phenotype level. The first evaluation in this subject was to test the growth of 22 isolates (11 from each habitat) - representing all phylogenomic clades (Fig. 2) - by using D-xylose as the single C-source in M9 minimal medium. The isolates containing the *xut* operon, enriched in the bulk soil sub-population, showed higher growth than the ones without it. The few rhizosphere isolates with those genes also showed a similar growth level compared to the bulk soil isolates (Fig. 5A).

Aiming to confirm that the higher growth of bulk soil isolates on D-xylose was a direct effect of *xut* genes, two closely related isolates, named R38 (*xut* $^-$) and B26 (*xut* $^+$), were further investigated. The growth curves indicated that the rhizosphere strain R38 shows a higher growth in a carbon rich environment (LB medium) (Fig. 5B). Strain R38 also showed a higher growth in an M9 minimal medium amended with glucose as the single C-source (Fig. 5C). This pattern shifted when the M9 medium was solely supplemented with D-xylose, where the strain B26 showed a much higher growth than R38, confirming the first assay (Fig. 5A and D).The *xutA* gene which codes for xylose isomerase was deleted in the chromosome of B26 wild type. The generated mutants with the *xutA* deletion B26Δ*xutA* 1 and B26Δ*xutA* 2 lacked the capacity to grow using D-xylose as a C-source, as observed in R38 (Fig. 5E). These results confirmed that the presence of *xut* genes in the bulk soil ecotype directly contributes to its higher growth in D-xylose.

Discussion

Our study initially showed that the community of *Pseudomonas* spp. is distinct between sugarcane rhizosphere and bulk soil, corroborating previous studies on other plant species, *e.g.* strawberry, oilseed rape, flax, tomato (Lemanceau *et al.*, 1995; Costa *et al.*, 2007). Although there is accumulating evidence supporting a different selection of rhizosphere and bulk soil upon the

Pseudomonas spp. community, it is not known whether this effect can influence the genome content of specific bacterial taxa from these habitats. Our study contributed to answer this question. The 53 isolates analyzed here are closely related, composing a population of *P. koreensis* which was classified in a single community OTU. Average nucleotide identity (ANI) was previously calculated for each of these isolates (Lopes et al., 2018). The values ranged from 84 to 99% ANI (Table S1). None of the compared isolates had 100% ANI, suggesting that the 53 isolates are different strains. Nevertheless, it is possible that some of these isolates are identical to one another, since small differences in ANI could be due to sequencing and assembly errors. Despite the close relationship of the 53 *P. koreensis* strains, our results showed that this population does not represent an ecological unit. It is separated into sub-populations according to the soil habitat (distinct ecotypes). In addition, we were able to identify the differences in the gene inventory of ecotypes from rhizosphere and bulk soil. One of the key genomic differences was supported by phenotypic analyses. These results highlight that much care should be taken when making functional inferences based on classification of OTUs in analyses of soil bacterial communities, since an OTU can harbor ecotypes with distinct genome content.

Phosphatases were one of the most outstanding genes enriched in the rhizosphere sub-population. A similar result was found in a study assessing another fluorescent *Pseudomonas* sp. population, where phosphate solubilizing genes were enriched in rhizosphere isolates compared to endosphere isolates of the plant *Populus deltoides* (Timm *et al*., 2015). The convergent results suggest that one criterion for fluorescent *Pseudomonas* spp. selection in the rhizosphere is the ability to make phosphate available. Interestingly, one of the main limitations for plant productivity in the tropical soils cropped with sugarcane is phosphate deficiency – promoted by the chemical properties of such acidic soils (Cherubin *et al*., 2016). The low pH where phosphate is unavailable to plants is a characteristic of the soil used in our study (pH = 5.0), which was previously described in Lopes *et al*. (2018).

On the other hand, a remarkable enrichment of *xut* genes was observed in the bulk soil sub-population. These genes are present in most bulk soil strains and absent in most rhizosphere strains. D-xylose is the most abundant sugar of hemicellulose (Lachke, 2002), which is the most labile polymer of lignocellulose, also composed by the recalcitrant fractions cellulose and lignin (Mood *et al*., 2013). The sugarcane soil used in our experiment was managed in a green-harvest system, where the field is permanently covered with high amounts of sugarcane biomass, the straw. Sugarcane straw is predominantly composed of lignocellulose (Carvalho *et al*., 2017). As bulk soil is commonly characterized as oligotrophic, with limited C-sources (Dennis *et al*., 2010), the capacity to assimilate xylose could constitute an advantage to the bulk soil ecotype to live in this habitat. In contrast, as the rhizosphere is characterized by a high and diverse input of labile C-sources from root exudates (Dennis *et al*., 2010), the selective pressures for D-xylose utilization in this compartment might be reduced, decreasing the frequency of *xut* $^{+}$ genotypes. Genetic drift is a possible hypothesis for the decrease of *xut* genes in the rhizosphere sub-population. An alternative hypothesis is that *xut* $^{-}$ strains accumulated in the rhizosphere sub-population by selection. In this case, *xut* genes could represent an extra metabolic burden to the *xut* $^{+}$ genotype, resulting in a fitness advantage to the *xut* $^{-}$ genotype. In the media where the strains were tested with other carbon sources than xylose (*i.e*. sole glucose and LB media), R38 (*xut* $^{-}$) showed a higher growth than B26 (*xut* $^{+}$), which could be due to this fitness advantage (Figure 5 B and C).

In vitro experiments showed that extra metabolic cost can reduce the frequency of a genotype in a population of *P. protegens*, where mutants defective in the regulatory *gacA*/*gacS* genes required for antibiotic production were selected in an environment without interspecific competition (Yan et al., 2018). For our soil population of *P. koreensis*, *xut* $^{+}$ genotypes were selected in the habitat where the capacity to assimilate xylose is advantageous (bulk soil), while the more fit *xut* $^{-}$ genotypes were possibly selected in the habitat where this capacity is not essential (rhizosphere). In a previous microbiome study of the same sugarcane field, CLPP analysis identified D-xylose as a C-source more highly oxidized by the bulk soil than the rhizosphere microbiome (Lopes *et al*., 2016). Those results suggest that the assimilation of xylose could be a characteristic of other microbial taxa inhabiting bulk soil in addition to the ecotype of *P. koreensis* here assessed. It is noteworthy that the reference strains of the *P. koreensis* subgroup (*P. fluorescens* Pf0-1 and *Pseudomonas* sp. MS586) used in our phylogenomic tree also have the *xut* genes and were isolated from the bulk soil of other fields with different crops (Compeau *et al*., 1988; http://www.pseudomonas.com/strain/show/4107).

In addition to xylose utilization, the bulk soil sub-population was significantly enriched in functions associated with polysaccharide binding, transport and catabolism. This is another indication that the bulk soil ecotype needs the genetic framework for the assimilation of complex carbon compounds present in the soil organic matter, which might be dispensable for the rhizosphere ecotype that receives labile C-sources from root exudation.

The change in gene frequencies between the compared sub-populations is a typical case of microevolution, where genotypes with selective advantage in a habitat dominate the population (Falush, 2015). However, since rhizosphere and bulk soil are neighboring soil compartments without a geographic barrier, the genotypes most adapted to a habitat can be dispersed and survive in lower frequencies in the other habitat (*e.g*. *xut* $^{-}$ strains in the bulk soil). That is possibly the reason why we isolated a few rhizosphere-like ecotypes in the bulk soil and *vice-versa*. The evolutionary mechanisms involved in the ecological differentiation of bulk soil and rhizosphere ecotypes need to be elucidated in depth. Horizontal gene transfer (HGT) and recombination are important mechanisms for genetic innovation in *Pseudomonas* spp. genomes, which are highly dynamic and mobile (Silby et al., 2011). In the *P. koreensis* population analyzed, we found some differences in the number of predicted horizontally transferred genes between the rhizosphere and bulk soil sub-populations. The different microbiomes of sugarcane rhizosphere and bulk soil constitute distinct genetic pools for HGT and could have contributed to differentiation of the ecotypes.

We found a relatively low number of genes/functions in different frequencies between the ecotypes of each habitat, *i.e.* less than 0.5% of the total of gene clusters (48/>10,000) in each sub-population (Table 1). These results suggest that the selection of rhizosphere and bulk soil habitats acts on specific genes/functions of this population, not drastically changing the genome structure of each ecotype. This possibility is in agreement with previous findings in a population of *Vibrio* sp. from ocean habitats, where gene-specific rather than genome-wide selection led to the differential occupation of niches (Shapiro *et al.*, 2012). Some studies targeting other bacterial species also detected the separation of a population in ecotypes containing genomic differences. The marine planktonic species *Alteromonas macleodii* was split in ecotypes inhabiting different water column depths (Ivars-Martinez et al., 2008). Genetically distinct ecotypes with psychrotolerant or thermotolerant lifestyles were found in *Bacillus cereus* (Guinebretière et al., 2017). Genomic differences were also observed between pathogenic and non-pathogenic ecotypes of *Gardnerella vaginallis* (Cornejo et al., 2018).

In our previous study we identified a distinct sub-clade dominated by rhizosphere isolates inside this *P. koreensis* population, potentially representing a new species (Lopes *et al.*, 2018). If it is the case, the different selective pressures between rhizosphere and bulk soil might have promoted a sympatric speciation, when new species arise in the absence of a geographic barrier (Hunt *et al.*, 2008). Nevertheless, this sub-clade shows the same pattern of the other *P. koreensis* clades dominated by rhizosphere strains: enrichment of phosphatases and low frequency of *xut* genes. Future studies will shed light on these remaining questions and contribute to the understanding of fluorescent *Pseudomonas* spp. evolutionary ecology in soil.

Conclusions

The present results inaugurate comparative genomics of ecotypes from a bacterial population inhabiting neighboring soil compartments, and revealed that the ecological differences between rhizosphere and bulk soil are not restricted to the structure of microbial communities, but can extend to the population level. For the model population of *P. koreensis* analyzed in our study, xylose utilization and phosphatases were important functions distinguishing the niches of the bulk soil and rhizosphere ecotypes, respectively. Future research can use this approach for other bacterial populations, taking our understanding of soil microbial ecology to a deeper level of resolution.

Experimental Procedures

Sample collection

Sampling was performed in a sugarcane field cultivated for 10 years in a green-harvest cropping system located in Piracicaba-SP, Brazil, in 2014 (22°44´S/47°38´W). Soil samples were collected in two below ground compartments: the soil layer of 1-2 mm adhering to plant roots (rhizosphere); and the soil between the cropping rows (0-10 cm) avoiding roots (bulk soil). Bulk soil and rhizosphere samples were collected in six random points along the sugarcane plantation. In total, 12 samples were used for further analyses, six from each habitat. More sampling details are described in Lopes *et al*. (2018).

Culture-independent analysis of *Pseudomonas* spp. communities in soil samples

A previous study based on 16S rRNA gene high-throughput sequencing (targeting the V3-V4 region – Klindworth et al., 2013) of DNA directly extracted from soil samples supported a publication on major distinctions between the bacterial communities in the bulk soil and rhizosphere of sugarcane (Lopes *et al*., 2016). Libraries were prepared and sequencing was performed on an Illumina MiSeq platform by a third partner service (Laboratory of Functional Genomics Applied to Agriculture and Agroenergy, ESALQ/USP, Brazil). The resulting reads were deposited at the SRA database (NCBI) as BioProject PRJNA319762. BioSample accession numbers are: SAMN04904299 (B1); SAMN04904300 (B2); SAMN04904301 (B3); SAMN04904302 (B4); SAMN04904303 (B5); SAMN04904304 (B6); SAMN04904305 (R1); SAMN04904306 (R2); SAMN04904307 (R3); SAMN04904308 (R4); SAMN04904309 (R5); and SAMN04904310 (R6) (R = Rhizosphere and B = Bulk Soil samples).

Here, this dataset was used to a more specific analysis, where the operational taxonomic units (OTUs) classified as *Pseudomonas* spp. by the Greengenes database were filtered from the OTU table generated by the UCLUST algorithm (97% cutoff) using the command filter_taxa_from_otu_table.py on QIIME software (Caporaso et al., 2010). Thus, a *Pseudomonas* spp. OTU table was obtained, and representative sequences of each OTU classified as *Pseudomonas* spp. were retrieved from the full set of sequences by using the filter_fasta.py command.

Pseudomonas spp. OTU table derived from the 16S rRNA gene high-throughput sequencing was exported to Primer-6 software (Clarke and Gorley, 2006), where data was normalized and a non-metric multidimensional scaling (NMDS) using the Bray-Curtis similarity index was performed in order to assess the ordination of rhizosphere and bulk soil samples. In addition, whole bacterial community OTU table was exported to STAMP software, where a Welch's t-test was performed using the Benjamini-Hochberg *P*-value correction to also verify the differences in relative abundance of reads classified as *Pseudomonas* spp. (Parks *et al*., 2014).

Additionally, real-time PCR (qPCR) analysis was performed using a primer pair that anneals with high specificity (99%) to a fragment of the 16S rRNA gene of *Pseudomonas* spp. (Bergmark *et al*., 2012). The reaction contained Sybr Green (1X), 1 µL of template DNA and 385 nM of each primer (PseF/PseR) for a total volume of 25 µL (Bergmark et al., 2013). Amplification was conducted using a StepOne Real-Time System (Applied Biosystems) under the following conditions: 1 cycle of 95°C for 10 min; and 40 cycles of 95°C for 30 s and 60°C for 1 min, followed by a melting curve analysis. For the standard curve, soil-derived amplicons of the same 16S rRNA gene fragment were diluted from 10^{-1} to 10^{-8} and quantified. These dilutions were submitted to amplifications under similar conditions as described above. Results of *Pseudomonas* spp. 16S rRNA gene sequences quantification was exported to Past software (Hammer *et al*., 2001), where data was normalized and a Tukey pairwise comparison test was performed aiming to assess the existence of significant differences ($P<0.05$) between bulk soil and rhizosphere samples.

Genome analysis of fluorescent *Pseudomonas* strains from bulk soil and rhizosphere

The genome sequences of 76 isolates obtained from the same rhizosphere and bulk soil samples were previously used for phylogenetic classification and disclosing of a new *Pseudomonas* spp. diversity (Lopes *et al*., 2018). Briefly, genomes were sequenced using the Illumina HiSeq 3000 platform in the Center for Genome Research and Biocomputing (CGRB-OSU, USA), draft genomes were assembled using SPAdes 3.7.0 and annotations were obtained using the Prokka software (Bankevich *et al*., 2012; Seeman, 2014). This Whole Genome Shotgun project (BioProject PRJNA383709) has been deposited at DDBJ/ENA/GenBank under the accessions NEHD00000000-NEKA00000000. The versions described here are NEHD01000000-NEKA01000000. More information with respect to isolation, culturing and genome sequencing are provided in Lopes et al. (2018). The accession numbers for the genome sequences of the isolates used in the present study are supplied in Table S2.

Here we focused on 53 of the 76 isolates, which were classified in the *P. koreensis* subgroup of the *P. fluorescens* complex based on multilocus sequence analysis (Lopes *et al*., 2018). The remaining 23 isolates were classified in other *P. fluorescens* subgroups or in the *P. putida* group (Lopes *et al*., 2018). The reason for choosing these 53 isolates is because they form a single population of *P. koreensis* and thus are suitable for a population genomics study comparing rhizosphere to bulk soil isolates. Putative orthologues present in the genome sequences of the 53 isolates and in other 41 relevant reference strains were inferred by clustering genes using the OrthoMCL algorithm on Get Homologues software (Contreras-Moreira and Vinuesa, 2013). The default settings of Get Homologues were used to group sequences in gene clusters, *i.e*. all-vs-all BLASTp with a minimum coverage of 75% and an e-value cutoff of 1x10^{-5}. Gene clusters with single copies and present in all 94 genome sequences were individually aligned using Mafft 7.271 (L-insi-i algorithm). The resulting alignments of gene sequences were concatenated and RAxML 8.2.8 was used to generate a maximum likelihood phylogenetic tree, further visualized using the ItoL platform (Katoh and Standley, 2013; Stamatakis, 2014; Letunic and Bork, 2016).

The same approach of Get Homologues was used to define the core genome of the 53 isolates composing the *P. koreensis* population. The gene clusters present in 100% of the isolates were defined as part of the core genome, while the gene clusters not present in 100% of the isolates (shared by 52 isolates or less) were defined as part of the accessory genome of the population. The matrix of gene clusters' abundance among the 53 isolates generated by Get Homologues was used for statistical analyses. To avoid bias from a single interpretation based on gene cluster annotation performed by Get Homologues, InterProScan 5 software was used to identify conserved protein domains in the amino acid sequences based on Pfam and TIGRFAM protein databases (Jones *et al*., 2014). After the identification of conserved domains, GO-terms were attributed to each inferred function based on the identification of conserved domains in the amino acid sequences. The abundance of GO-terms in each genome sequence was organized into a matrix, which was also further used for statistical analyses. All functional classes of GO-terms were considered in this analysis.

The gene clusters' matrix generated on Get Homologues was exported to R software, where a discriminant analysis of principal components (DAPC) was performed using the adegenet package (Jombart, 2008; Jombart *et al*., 2010). The same matrix, as well as the one of GO-terms were also exported to STAMP software, where Welch's t-test was performed using the Bonferroni *P*-value correction, aiming to identify the gene clusters and functions significantly different between rhizosphere and bulk soil ecotypes ($P<0.05$).

The 16S rRNA gene sequences were extracted from the genome assemblies and compared to sequences of the OTUs classified as *Pseudomonas* spp. in the community analysis. Mafft 7.271 (L-insi-i algorithm) was used to align sequences, and the 16S rRNA gene sequences were trimmed to match with the aligned region of the OTU sequences (V3-V4 region). Amplicon and genome derived sequences were separated in OTUs using UCLUST algorithm and a 99% cutoff in the USEARCH 6.0 software (Edgar, 2010), a similar approach used in the community analysis, in order to check whether the genome sequences match to any community OTU, and in how many OTUs the 53 isolates are split. The 16S rRNA gene sequence of the reference strain *P. fluorescens* Pf0-1 was also included in this analysis.

The core and pan genome sizes were estimated for the population analyzed in our study. For that, a non-linear regression was performed on the permutated data (100 replicates) to estimate the core and pan sizes according to Tettelin *et al*. (2008). The free parameters of the models were estimated using the R nls function of the stats v3.4.1 package, calculated with the following formulas - Pan-genome: $n=\sigma(\#\ genomes)^{\gamma}$. Where $0< \gamma <1$ indicates an open pan-genome, and $\gamma <0$ indicates a closed pan-genome. Core genome: $F(n)=\kappa * \exp(-(\#\ \text{genomes})/\tau)+\Omega$, where Ω is the estimate of the core genome size.

In order to detect mobile genetic elements and horizontally transferred genes in the 53 genome sequences, some platforms were used, including Alien Hunter (Vernikos and Parkhill, 2006), PHAST (Zhou et al., 2011), IslandViewer v.3 (Dhillon et al., 2015) and IS Finder (Siguier et al., 2006). The number of hits in the genome sequences of the rhizosphere and bulk soil sub-populations was statistically compared by Tukey test in Past software (Hammer et al., 2001).

Physiological analyses of the isolates and mutagenesis assay

Twenty-two strains were used to test their capacity to grow on D-xylose (11 from each habitat, representing each phylogenomic clade of the population). Most of the strains containing the *xut* operon are from bulk soil, but rhizosphere strains containing *xut* genes as well as bulk soil strains lacking *xut* genes were also evaluated for growth on M9 minimal medium containing D-xylose as the sole C-source (Atlas, 1993).

The M9 medium was prepared by mixing four individual sterilized solutions: 10.0 mL of 5X M9 salts (autoclaved); 0.1 mL of 1M $MgSO_4$ (autoclaved); 0.005 mL of 1M $CaCl_2$ (autoclaved); 5.0 mL of D-xylose 10% (filter sterilized); and 0.05 mL of a micronutrients solution (filter sterilized) for a final volume of 50 mL. After mixing the solutions, the medium was buffered by addition of 20 g/L 3-(N-morpholino)-propanesulfonic acid (MOPS), adjusted to pH 7, and filter sterilized (Table S3). The micronutrients solution was prepared based on the preparation of the medium for *E. coli* (Cam *et al*., 2016). Overnight cultures of each strain were centrifuged and washed twice with sterilized deionized water in order to remove all carbon residues. After optical density (O.D.) adjustment to 0.03 (600 nm), 10 µL of each culture was inoculated into 90 µL of M9 or LB media in a well of a microtiter plate (Atlas, 1993). OD values were recorded every hour at 600 nm for a period of ~65 hours using a TECAN microplate reader (Männedorf, Switzerland). Each strain was grown in triplicate cultures, and blanks consisted of non-inoculated media.

Two phylogenetically related isolates B26 (*xut* $^{+}$) and R38 (*xut* $^{-}$), isolated from bulk soil and rhizosphere, respectively, were used as models to assess the capacity to grow on D-xylose. Alternatively, growth curves were tested with glucose replacing D-xylose in the M9 medium or with LB medium. A *xutA* deletion mutant was made in strain B26. The primer pairs xutA-F1/xutA-Ovlp-R1 and xutA-R2/xutA-Ovlp-F1 were used to PCR amplify two DNA fragments flanking the *xutA* gene (Table S4). The two fragments were joined together in a second PCR and the product was digested with *Hind*III and *Eco*RI, generating a 982-bp DNA fragment. This DNA fragment was ligated to pEX18Tc, previously digested and with compatible sticky ends, to create construct p18Tc-Δ*xutA*. This construct was transferred into strain *E. coli* S17-1 and introduced by bi-parental mating into strain B26 wild type. The deletion mutants were generated via allelic exchange mutagenesis, and confirmed by PCR using primers xutA-F3/xutA-R3. Two B26Δ*xutA* mutant strains were independently generated using this approach. Growth curves in the M9 medium with D-xylose were assessed using the mutant strains B26Δxut1/B26ΔxutA2 and the wild type strains R38/B26 with the same conditions described above.

Acknowledgements

We acknowledge CAPES for the scholarship support provided to LDL and MCPS; FAPESP and CNPq for the financial support to FDA, project 2013/50353-7 and grant no. 302168/2014-1, respectively. This work was also supported by the National Institute of Food and Agriculture, US Department of Agriculture award 2014-51181-22384 to JHC and 2011-67019-30192 to JEL; and by the National Science Foundation Graduate Research Fellowship grant no. DGE-1314109 to EWD. We are grateful to Dr. Jeff Anderson (OSU-USA) and Dr. Luiz C. Basso (USP -Brazil) for making their laboratories available for Tecan analyses; Dr. Tal Pupko (Tel Aviv University-Israel) for helpful discussions; Dr. Brendan Bohannan (University of Oregon-USA) and Dr. Francisco D. Andreote (University of Groningen-Netherlands) for reviewing the manuscript.

Authors' contributions

LDL collected soil samples, performed molecular, computational, phylogenetic, statistical and physiological analyses, and wrote the manuscript; MCPS collected soil samples, performed isolation and computational analyses. EWD and AJW performed computational, phylogenetic and statistical analyses; QY conducted the mutagenesis assay; CSV and CFW performed physiological analyses; JHC, JEL and FDA conceptualized and delineated the study, and contributed to the writing and editing of the manuscript.

Conflicts of interest

The authors declare no conflicts of interest.

References

Atlas, R.M. (1993) *Handbook of microbiological media*. CRC Press, Inc.: Boca Raton, FL, USA.

Babu, A.G., Shea, P.J., Sudhakar, D., Jung, I-B., Oh, B-T. (2015) Potential use of *Pseudomonas koreensis* AGB-1 in association with *Miscanthus sinensis* to remediate heavy metal(loid)-contaminated mining site soil. *J Environ Manage* **151**:160-166.

Bankevich, A., Nurk, S., Antipov, D., Gurevich, A., Dvorkin, M., Kulikov, A., et al. (2012) SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J Comput Biol* **19**:455–477.

Berg, G., Smalla, K. (2009) Plant species and soil type cooperatively shape the structure and function of microbial communities in the rhizosphere. *FEMS Microbiol Ecol* **68**:1-13.

Bergmark, L., Poulsen, P.H.B., Al-Soud, W.A., Norman, A., Hansen, L.H., Sorensen, S.J. (2012) Assessment of the specificity of *Burkholderia* and *Pseudomonas* qPCR assays for detection of these genera in soil using 454 pyrosequencing. *FEMS Microbiol Lett* **333**:77-84.

Cam, Y., Alkim, C., Trichez, .D, Trebosc, V., Vax, A., Bartolo, F., Besse, P., Francois, J.M., Walther, T. (2016) Engineering of a synthetic metabolic pathway for the assimilation of (D)-D-xylose into value-added chemicals. *ACS Synth Biol* **5**: 607-618.

Caporaso, J.G., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F.D., Costello, E.K., et al. (2010) QIIME allows analysis of high-throughput community sequencing data. *Nat Methods* **7**:335–336.

Carvalho, J.L.N., Nogueirol, R.C., Menandro, L.M.S., Bordonal, R.C., Borges, C.D., Cantarella, H. et al. (2017) Agronomic and environmental implications of sugarcane straw removal: a major review. *GCB Bioenergy* **9**:1181-1195.

Cherubin, M.R., Franco, A.L.C., Cerri, C.E.P., Karlen, D.L., Pavinatto, P.S., Rodrigues, M., Davies, C.A., Cerri, C.C. (2016) Phosphorus pools responses to land-use change for sugarcane expansion in weathered Brazilian soils. *Geoderma* **265**: 27-38

Clarke, K.R. & Gorley, R.N. (2006) *PRIMER v6: User Manual/Tutorial*. Plymouth: PRIMER-E.

Cohan, F.M. (2002) What are bacterial species? *Annu Rev Microbiol* **56**:457-487.

Compeau, G., Al-Achi, B.J., Platsouka, E., Levy, S.B. (1988) Survival of rifampicin-resistant mutants of *Pseudomonas fluorescens* and *Pseudomonas putida* in soil systems. *Appl Environ Microbiol* **54**:2432-2438.

Contreras-Moreira, B., Vinuesa, P. (2013) GET_HOMOLOGUES, a versatile software package for scalable and robust microbial pangenome analysis. *Appl Environ Microbiol* **79**(24):7696-7701.

Cordero, O.X., Polz, M.F. (2014) Explaining microbial genomic diversity in light of evolutionary ecology. *Nat Rev Microbiol* **12**(4):263-273.

Cornejo, O.E., Hickey, R.J., Suzuki, H., Forney, L.J. (2018) Focusing the diversity of *Gardnerella vaginallis* through the lens of ecotypes. *Evol Appl* **11**(3): 312-324.

Costa, R., Gomes, N.C.M., Krogerrecklenfort, E., Opelt, K., Berg, G., Smalla, K. (2007) *Pseudomonas* community structure and antagonistic potential in the rhizosphere: insights gained by combining phylogenetic and functional gene-based analyses. *Environ Microbiol* **9**(9):2260-2273.

Dennis, P.G., Miller, A.J., Hirsch, P.R. (2010) Are root exudates more important than other sources of rhizodeposits in structuring rhizosphere bacterial communities? *FEMS Microbiol Ecol* **72**:313-327.

Dhillon, H., Laird, M.R., Shay, J.A., Winsor, G.L., Lo, R., Nizam, F., Pereira, S.K., Waglechner, N., McArthur, A.G., Langille, M.G., Brinkman, F.S. (2015) IslandViewer 3: more flexible, interactive genomic island discovery, visualization and analysis. *Nucleic Acids Res* **43**(W1): W104-108.

Edgar, R.C. (2010) Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* **26**(19):2460-2461.

Falush, D. (2009) Toward the use of genomics to study microevolutionary change in bacteria. *PLoS Genet* **5**(10):e1000627.

Garcia-Salamanca, A., Molina-Henares, M.A., Dillewijn, P., Solano, J., Pizarro-Tobias, P., Roca, A., et al. (2012) Bacterial diversity in the rhizosphere of maize and the surrounding carbonate-rich bulk soil. *Microb Biotechnol* **6**(1):36-44.

Guinebretière, M-H., Loux, V., Martin, V., Nicolas, P., Sanchis, V., Broussolle, V. (2017) Draft genome sequences of 18 psychrotolerant and 2 thermotolerant strains representative of particular ecotypes in the *Bacillus cereus* group. *Genome Announc* **5**(5): e01568-16.

Hammer, Ø., Harper, D.A.T., Ryan, P.D. (2001) PAST: paleontological statistics software package for education and data analysis. *Paleontol Electron* **4** 1–9.

Hol, W.H., Bezemer, T.M., Biere, A. (2013) Getting the ecology into interactions between plants and the plant growth-promoting bacterium *Pseudomonas fluorescens*. *Front Plant Sci* **4**:81.

Hultberg, M., Alsberg, T., Khalil, S., Alsanius, B. (2010) Suppression of disease in tomato infected by *Pythium ultimum* with a biosurfactant produced by *Pseudomonas koreensis*. *BioControl* **55**(3):435-444.

Hunt, D.E., David, L.A., Gevers, D., Preheim, S.P., Alm, E.J., Polz, M.F. (2008) Resource partitioning and sympatric differentiation among closely related bacterioplankton. *Science* **320**(5879):1081-1085.

Ivars-Martinez, E., Martin-Cuadrado,A.B., D'auria, G., Mira, A., Ferriera, S., Johnson, J., Friedman, R., Rodriguez-Valera, F. (2008) Comparative genomics of two ecotypes of the marine planktonic copiotroph *Alteromonas macleodii* suggests alternative lifestyles associated with different kinds of particulate organic matter. *ISME J* **2**(12): 1194-1212.

Jombart, T. (2008) adegenet: a R package for the multivariate analysis of genetic markers. *Bioinformatics* **24**(11):1403-1405.

Jombart, T., Devilled, S., Bollox, F. (2010) Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. *BMC Genet* **11**:94.

Jones, P., Binns, D., Chang, H.Y., Fraser, M., Li, W., McAnulla, C., et al. (2014) InterProScan 5: genome-scale protein function classification. *Bioinformatics* **30**(9):1236-1240.

Katoh, K., Standley, D.M. (2013) MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol* **30**:772–780.

Klindworth, A., Pruesse, E., Schweer, T., Peplies, J., Quast, C., Horn, M., et al. (2013) Evaluation of general 16S ribosomal RNA gene PCR primers for classical and next-generation sequencing-based diversity studies. *Nucleic Acids Res* **41**:e1 10.1093/nar/gks808.

Kuzyakov, Y., Blagodatskaya, E. (2015) Microbial hotspots and hot moments in soil: concept and review. *Soil Biol Biochem* **83**:184-199.

Kwak, Y.S., Weller, D.M. (2013) Take-all of wheat and natural disease suppression: a review. *Plant Pathol J* **29**(2):125-135.

Lachke, A. (2002) Biofuel from D-xylose: the second most abundant sugar. *Resonance* **7**(5):50-58.

Lemanceau, P., Corberand, T., Gardan, L., Latour, X., Daguerre, G., Boeufgras, J.M., et al. (1995) Effect of two plant species, flax (*Linum usitatissinum* L.) and tomato (*Lycopersicon esculentum* Mill.), on the diversity of soil borne populations of fluorescent Pseudomonads. *Appl. Environ Microbiol* **61**(3):1004-1012.

Letunic, I., Bork, P. (2016) Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees. *Nucleic Acids Res* **44**(W1):W242-5.

Liu, Y., Rainey, P.B., Zhang, X.X. (2015) Molecular mechanisms of xylose utilization by *Pseudomonas fluorescens*: overlapping genetic responses to xylose, xylulose, ribose and mannitol. *Mol Microbiol* **98**(3):553-570.

Loper, J.E., Hassan, K.A., Mavrodi, D.V., Davis, E.W., Lim, C.K., Shaffer, B.T., et al. (2012) Comparative genomics of plant-associated *Pseudomonas* spp.: Insights into diversity and inheritance of traits involved in multitrophic interactions. *PLoS Genet* **8**(7):e1002784.

Lopes, L.D., Davis, E.W.II., Pereira e Silva, M.C., Weisberg, A.J., Bresciani, L., Chang, J.H., et al. (2018) Tropical soils are a reservoir for fluorescent *Pseudomonas* spp. biodiversity. *Environ Microbiol* **20**(1):62-74.

Lopes, L.D., Pereira e Silva, M.C., Andreote, F.D. (2016) Bacterial abilities and adaptation toward the rhizosphere colonization. *Front Microbiol* **7**:134.

Mendes, L.W., Kuramae, E.E., Navarrete, A.A., van Even, J.A., Tsai, S.M. (2014) Taxonomical and functional microbial community selection in soybean rhizosphere. *ISME J* **8**:1577–1587.

Mendes, R., Kruijt, M., Bruijn, I., Dekkers, E., Voort, M., Schneider, J.H.M., et al. (2011) Deciphering the rhizosphere microbiome for disease-suppressive bacteria. *Science* **332**(6033):1097-100.

Mercado-Blanco, J., Bakker, P.A. (2007) Interactions between plants and beneficial *Pseudomonas* spp.: exploiting bacterial traits for crop protection. *Antonie Van Leeuwenhoek* **92**(4):367-389.

Mood, S.H., Golfeshan, A.H., Tabatabaei, M., Jouzani, G.S., Najafi, G.H., Gholami, M., et al. (2013) Lignocellulosic biomass to bioethanol, a comprehensive review with a focus on pretreatment. *Renewable and Sustainable Energy Reviews* **27**:77-93.

Parks, D.H., Tyson, G.W., Hugenholtz, P., Beiko, R.G. (2014) STAMP: statistical analysis of taxonomic and functional profiles. *Bioinformatics* **30**:3123–3124.

Philippot, L., Raaijmakers, J.M., Lemanceau, P., van der Putten, W.H. (2013) Going back to the roots: the microbial ecology of the rhizosphere. *Nat Rev Microbiol* **11**:789-799.

Prosser, J.I., Bohannan, B.J.M., Curtis, T.P., Ellis, R.J., Firestone, M.K., Freckleton, R.P., et al. (2007) The role of ecological theory in microbial ecology. *Nat Rev Microbiol* **5**(5):384-392.

Rafikova, G.F., Korshunova, T.Y., Minnebaev, L.F., Chetverikov, S.P., Loginov, O.N. (2016) A new bacterial strain, *Pseudomonas koreensis* IB-4, as a promising agent for plant pathogen biological control. *Microbiology* **85**(3):333-341.

Seeman, T. Prokka: rapid prokaryotic genome annotation. *Bioinformatics* **30**(14):2068-2069 (2014).

Shapiro, B.J., Friedman, J., Cordero, O.X., Preheim, S.P., Timberlake, S.C., Szabó, G., et al. (2012) Population genomics of early events in the ecological differentiation of bacteria. *Science* **336**:48-51.

Silby, M.W., Winstanley, C., Godfray, S.A., Levy, S.B., Jackson, R.W. (2011) *Pseudomonas* genomes: diverse and adaptable. *FEMS Microbiol Rev* 35(4): 652-680.

Siquier, P., Perochon, J., Lestrade, L., Mahillon, J., Chandler, M. (2006) ISfinder: the reference centre for bacterial insertion sequences. *Nucleic Acids Res* **34**: D32-36.

Sokolova, T.A. (2015) Specificity of soil properties in the rhizosphere: analysis of literature data. *Eurasian Soil Sci* **9**:1097-1111.

Stamatakis, A. (2014) RAxMLversion 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**:1312-1313.

Tettelin, H., Riley, D., Cattuto, C., Medini, D. (2008) Comparative genomics: the bacterial pan-genome. *Curr. Opin. Microbiol.* **11**(5):472-477.

Timm, C.M., Campbell, A.G., Utturkar, S.M., Jun, S.R., Parales, R.E., Tan, W.A., et al. (2015) Metabolic functions of *Pseudomonas fluorescens* strains from *Populus deltoides* depend on rhizosphere or endosphere isolation compartment. *Front Microbiol.* **6**:1118.

Uroz, S., Buée, M., Murat, C., Frey-Klett, P., Martin, F. (2010) Pyrosequencing reveals a contrasted bacterial diversity between oak rhizosphere and surrounding soil. *Environ Microbiol Rep* **2**:281–288.

Vernikos, G.S., Parkhill, J. (2006) Interpolated variable order motifs for identification of horizontally acquired DNA: revisiting the *Salmonella* pathogenicity islands. *Bioinformatics* **22**(18): 2196-2203.

Weller, D.M., Raaijmakers, J.M., Gardener, B.B., Thomashow, L.S. (2002) Microbial populations responsible for specific soil suppressiveness to plant pathogens. *Annu Rev Phytopathol* **40**:309-348.

Yan, Q., Lopes, L.D., Shaffer, B.T., Kidarsa, T.A., Vining, O., Philmus, B., Song, C., Stockwell, V.O., Raaijmakers, J.M., McPhail, K.L., Andreote, F.D., Chang, J.H., Loper, J.E. (2018) Secondary metabolism and interspecific competition affect accumulation of spontaneous mutants in the GacS-GacA regulatory system in *Pseudomonas protegens*. *mBio* **9**(1): e01845-17.

Zhou, Y., Liang, Y., Lynch, K.H., Dennis, J.J., Wishart, D.S. (2011) PHAST: a fast phage search tool. *Nuclei Acids Res* **39**: W347-52.

Tables

Table 1. Number of gene clusters in the ecotypes from rhizosphere or bulk soil

	Bulk soil isolates (25)	Rhizosphere isolates (28)	All isolates (53)
Shared gene clusters	4,200	4,289	4,124
Total gene clusters	8,446	8,134	9,824
Accessory gene clusters	4,246	3,845	5,700

Figures

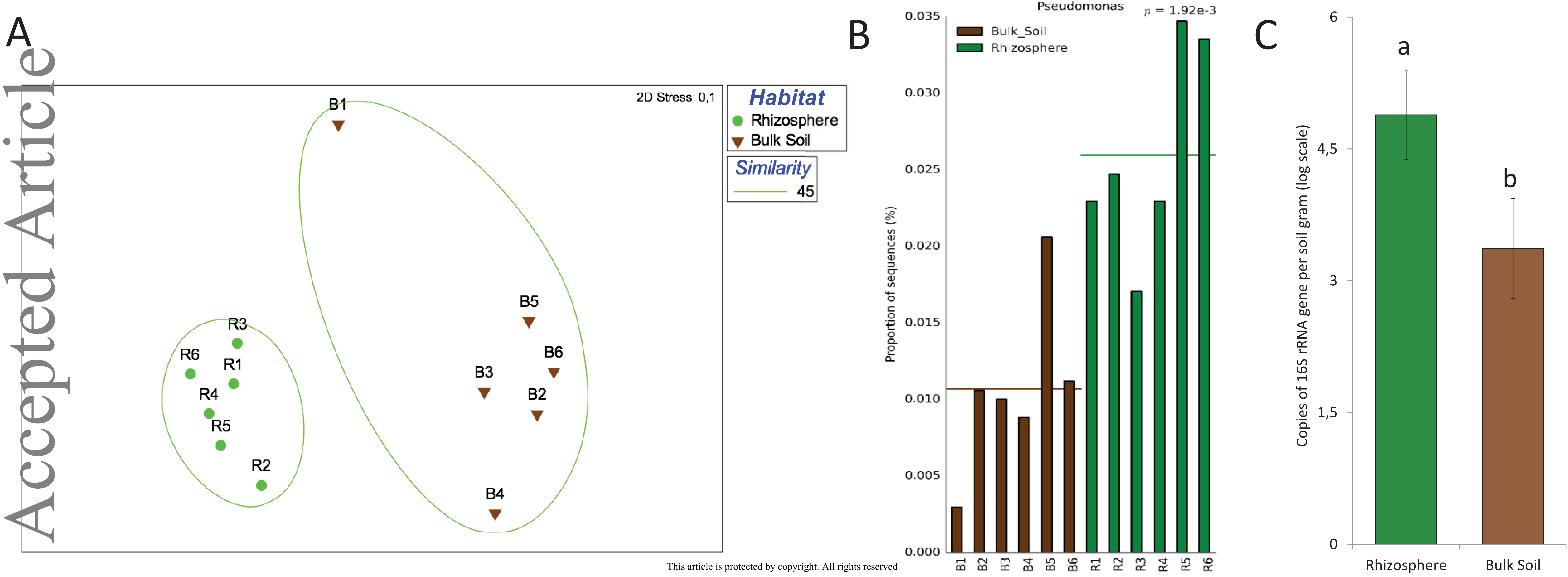
Fig. 1. Differences in *Pseudomonas* community between rhizosphere and bulk soil. A) Changes in community structure shown by NMDS based on the *Pseudomonas* spp. OTU table from 16S rRNA gene high-throughput sequencing; B) Differences in the relative abundance of reads classified as *Pseudomonas* spp. from 16S rRNA gene high-throughput sequencing; C) Differences in 16S rRNA gene copy numbers specific to *Pseudomonas* spp. analyzed by qPCR. R1-R6 and B1-B6 refer to different rhizosphere and bulk soil samples, respectively.

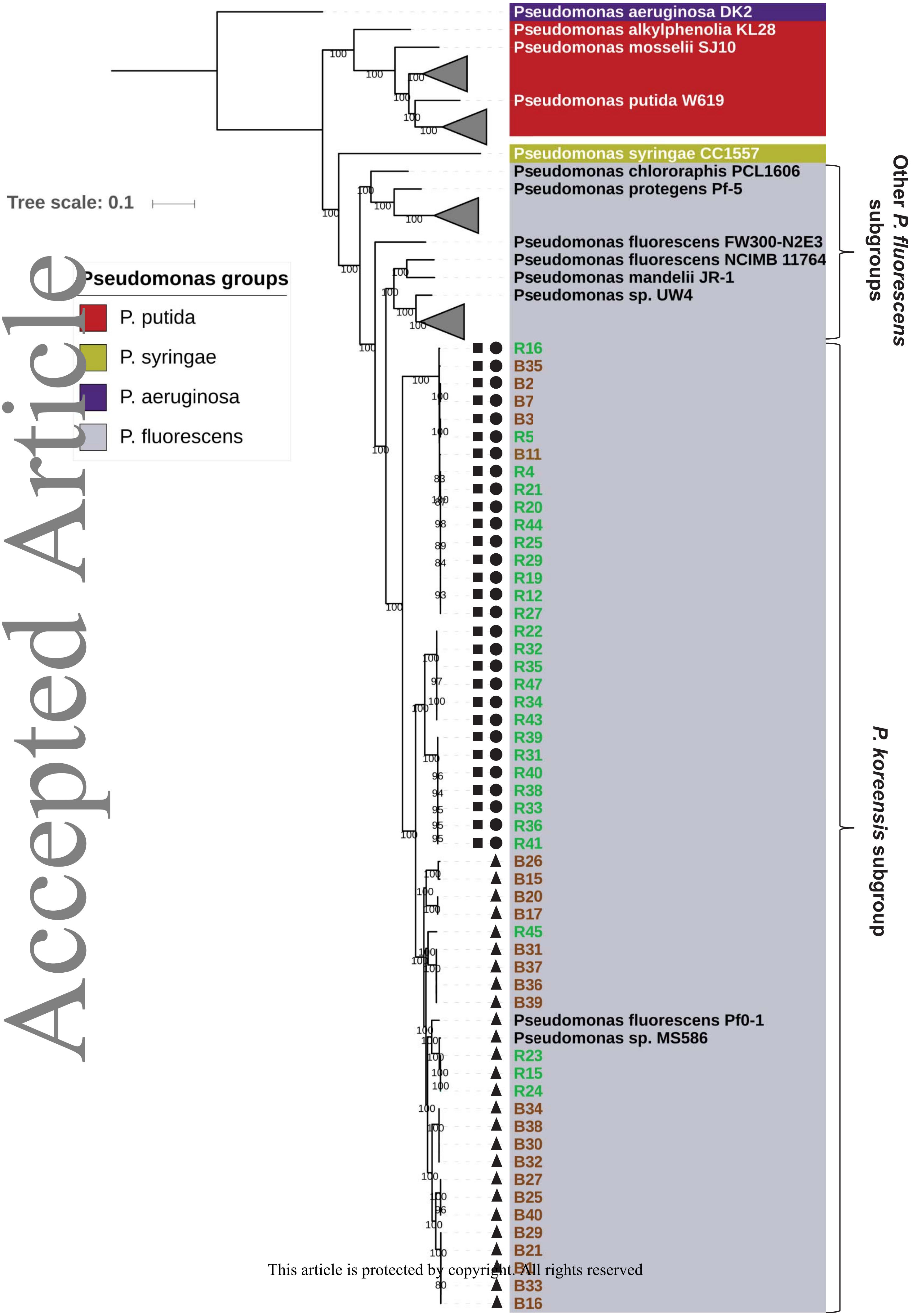
Fig. 2. Maximum likelihood phylogenetic inference based on phylogenomics. 1,540 orthologous gene sequences present in single copies in all 53 isolates of this study and 41 reference strains were identified, concatenated and used to generate the phylogenetic tree. The isolates of our study are labeled in green (rhizosphere) or brown (bulk soil), while the reference strains are labeled in black/white font. The presence of selected gene clusters in a genome is denoted with a symbol: triangle, xylose utilization; circle, acid phosphatase; rectangle, alkaline phosphatase genes. The numbers below each branch are the bootstrap values (100 bootstrap tests). Genomes classified in the *P. koreensis* subgroup and other subgroups inside the *P. fluorescens* complex are highlighted.

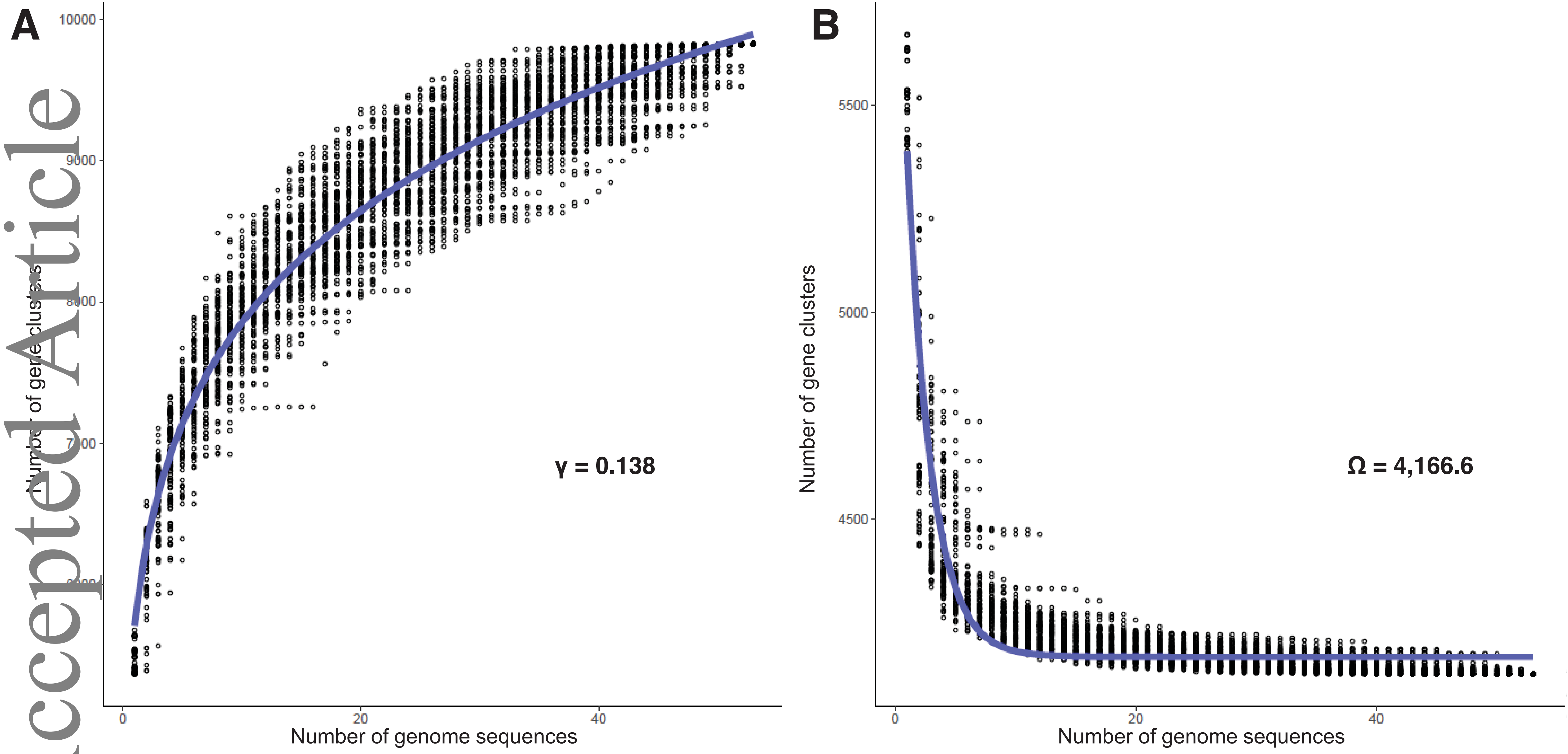
Fig. 3. Estimation of the core and pan genome sizes of the population based on a non-linear regression. A) Pan-genome accumulation curve. B) Core genome rarefaction curve. Analysis was performed considering the 53 genome sequences. $\gamma < 0$ indicates a closed pan-genome, while $0 < \gamma < 1$ indicates an open pan-genome; Ω is the estimation of the core genome size.

Fig. 4. Statistical analyses comparing rhizosphere versus bulk soil ecotypes. A) Discriminant analysis of principal components (DAPC) contrasting the rhizosphere and bulk soil isolates (sub-populations). At the top left are the PCA eigenvalues needed to discriminate without overlap the sub-populations of the two habitats. The bulk soil sub-population is represented in brown, while the rhizosphere sub-populations is represented in green. B) List of gene clusters showing the most significant differences between rhizosphere and bulk soil ecotypes ($P < 0.01$). Statistical analysis was performed using the Welch's t-test with the Bonferroni *P*-value correction. Asterisks in brown and green highlights genes selected in the bulk soil and rhizosphere sub-populations, respectively.

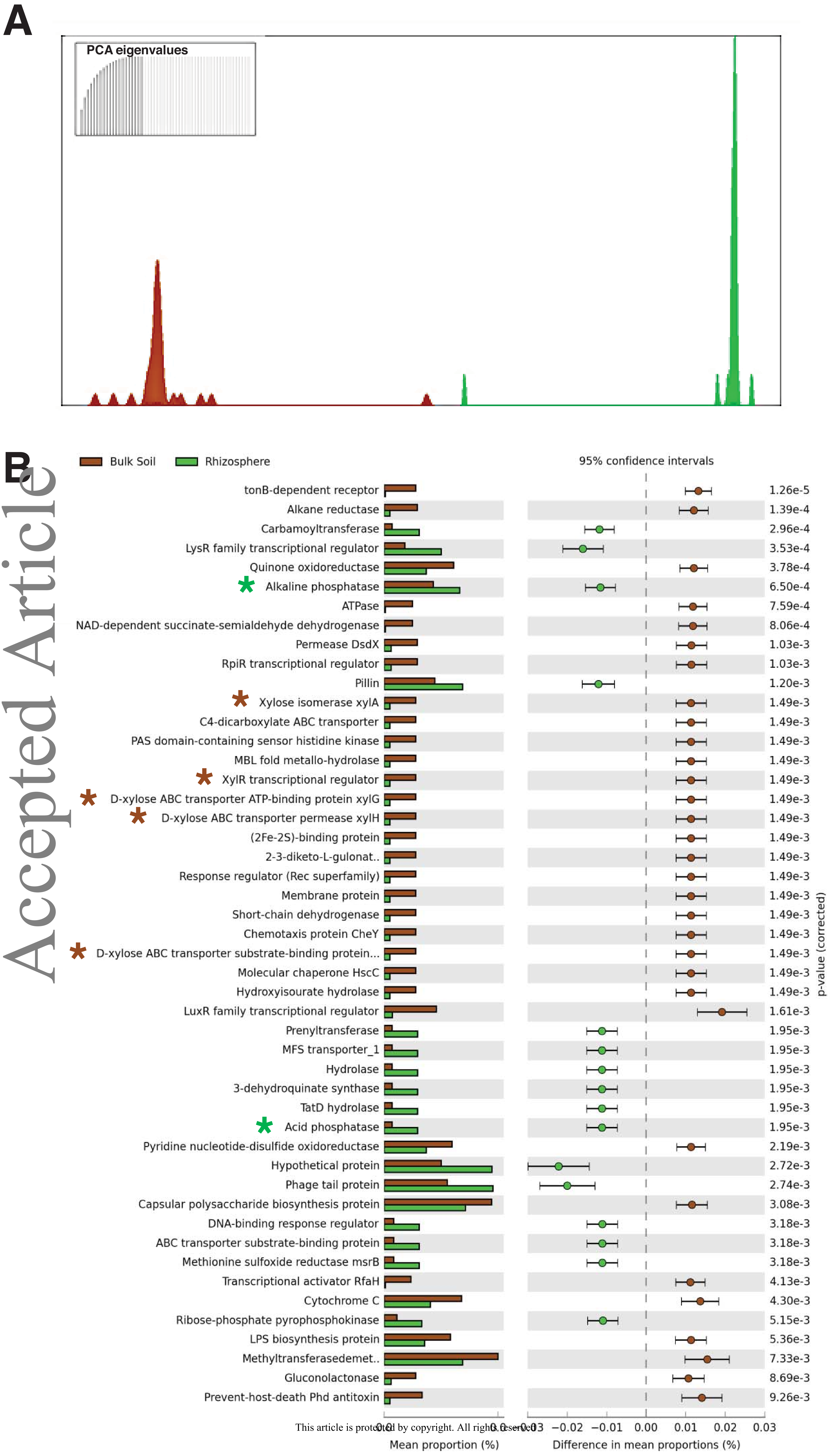
Fig. 5. Differential growth of bulk soil and rhizosphere isolates on D-xylose. A) Representative strains of each phylogenomic clade growing on M9 medium with D-xylose as the single C-source. Rhizosphere or bulk soil isolates are represented by green and brown bars, respectively. Triangles indicate the presence of *xut* genes (xut^{+} strains). The O.D. shown in this graph was measured after 60 hours of growth, when curves were stable (stationary phase). B-D) Growth of isolates containing (B26) or not containing (R38) the *xut* genes on B) LB medium; C) M9 minimal medium with glucose as the single carbon source (C-source); and D) M9 medium with D-xylose as the single C-source. E) Two *xutA* deletion mutants (B26Δ*xutA* 1; B26Δ*xutA* 2) compared to B26 wild type (WT) and R38 on M9 medium with D-xylose as the single C-source. The growth curves of the mutants and R38 are overlapping in the graph.

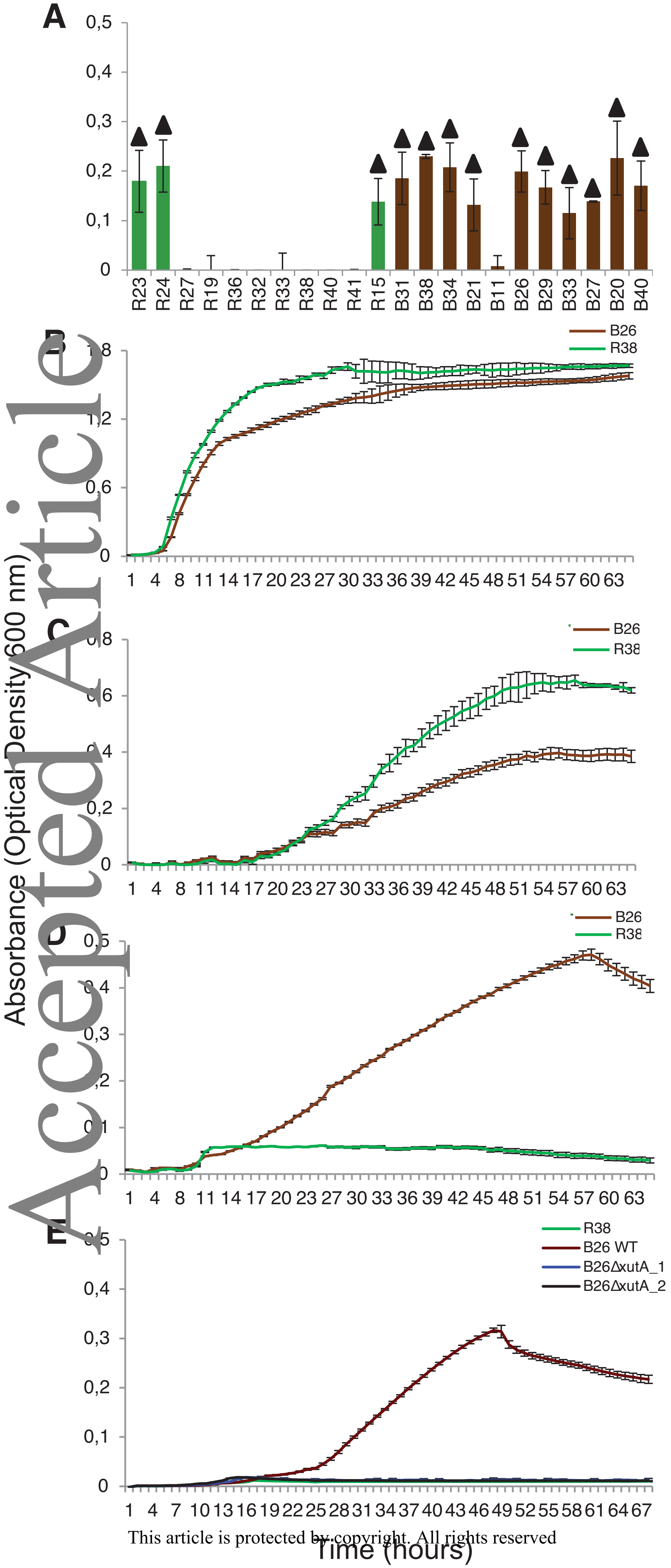

A
2D Stress: 0,1
Habitat
● Rhizosphere
▼ Bulk Soil
Similarity
45
B1
B5
B6
B3
B2
B4
R3
R6
R1
R4
R5
R2
B
Pseudomonas
p = 1.92e-3
Bulk_Soil
Rhizosphere
Proportion of sequences (%)
0.035
0.030
0.025
0.020
0.015
0.010
0.005
0.000
B1
B2
B3
B4
B5
B6
R1
R2
R3
R4
R5
R6
C
Copies of 16S rRNA gene per soil gram (log scale)
6
4,5
3
1,5
0
a
b
Rhizosphere
Bulk Soil


Tree scale: 0.1
Pseudomonas groups
P. putida
P. syringae
P. aeruginosa
P. fluorescens
Pseudomonas aeruginosa DK2
Pseudomonas alkylphenolia KL28
Pseudomonas mosselii SJ10
Pseudomonas putida W619
Pseudomonas syringae CC1557
Pseudomonas chlororaphis PCL1606
Pseudomonas protegens Pf-5
Pseudomonas fluorescens FW300-N2E3
Pseudomonas fluorescens NCIMB 11764
Pseudomonas mandelii JR-1
Pseudomonas sp. UW4
R16
B35
B2
B7
B3
R5
B11
R4
R21
R20
R44
R25
R29
R19
R12
R27
R22
R32
R35
R47
R34
R43
R39
R31
R40
R38
R33
R36
R41
B26
B15
B20
B17
R45
B31
B37
B36
B39
Pseudomonas fluorescens Pf0-1
Pseudomonas sp. MS586
R23
R15
R24
B34
B38
B30
B32
B27
B25
B40
B29
B21
B1
B33
B16
Other *P. fluorescens* subgroups
P. koreensis subgroup


A
10000
9000
8000
7000
6000
Number of gene clusters
γ = 0.138
0
20
40
Number of genome sequences
B
5500
5000
4500
Number of gene clusters
Ω = 4,166.6
0
20
40
Number of genome sequences

Lorem ipsum


A
PCA eigenvalues
B
Bulk Soil
Rhizosphere
95% confidence intervals
tonB-dependent receptor 1.26e-5
Alkane reductase 1.39e-4
Carbamoyltransferase 2.96e-4
LysR family transcriptional regulator 3.53e-4
Quinone oxidoreductase 3.78e-4
* Alkaline phosphatase 6.50e-4
ATPase 7.59e-4
NAD-dependent succinate-semialdehyde dehydrogenase 8.06e-4
Permease DsdX 1.03e-3
RpiR transcriptional regulator 1.03e-3
Pillin 1.20e-3
* Xylose isomerase xylA 1.49e-3
C4-dicarboxylate ABC transporter 1.49e-3
PAS domain-containing sensor histidine kinase 1.49e-3
MBL fold metallo-hydrolase 1.49e-3
* XylR transcriptional regulator 1.49e-3
* D-xylose ABC transporter ATP-binding protein xylG 1.49e-3
* D-xylose ABC transporter permease xylH 1.49e-3
(2Fe-2S)-binding protein 1.49e-3
2-3-diketo-L-gulonat.. 1.49e-3
Response regulator (Rec superfamily) 1.49e-3
Membrane protein 1.49e-3
Short-chain dehydrogenase 1.49e-3
Chemotaxis protein CheY 1.49e-3
* D-xylose ABC transporter substrate-binding protein... 1.49e-3
Molecular chaperone HscC 1.49e-3
Hydroxyisourate hydrolase 1.49e-3
LuxR family transcriptional regulator 1.61e-3
Prenyltransferase 1.95e-3
MFS transporter_1 1.95e-3
Hydrolase 1.95e-3
3-dehydroquinate synthase 1.95e-3
TatD hydrolase 1.95e-3
* Acid phosphatase 1.95e-3
Pyridine nucleotide-disulfide oxidoreductase 2.19e-3
Hypothetical protein 2.72e-3
Phage tail protein 2.74e-3
Capsular polysaccharide biosynthesis protein 3.08e-3
DNA-binding response regulator 3.18e-3
ABC transporter substrate-binding protein 3.18e-3
Methionine sulfoxide reductase msrB 3.18e-3
Transcriptional activator RfaH 4.13e-3
Cytochrome C 4.30e-3
Ribose-phosphate pyrophosphokinase 5.15e-3
LPS biosynthesis protein 5.36e-3
Methyltransferasedemet.. 7.33e-3
Gluconolactonase 8.69e-3
Prevent-host-death Phd antitoxin 9.26e-3
p-value (corrected)
0.0
0.0
Mean proportion (%)
−0.03
−0.02
−0.01
0.00
0.01
0.02
0.03
Difference in mean proportions (%)


A
0
0,1
0,2
0,3
0,4
0,5
R23
R24
R27
R19
R36
R32
R33
R38
R40
R41
R15
B31
B38
B34
B21
B11
B26
B29
B33
B27
B20
B40
B
0
0,6
1,2
1,8
B26
R38
1 4 8 11 14 17 20 23 27 30 33 36 39 42 45 48 51 54 57 60 63
C
0
0,2
0,4
0,6
0,8
B26
R38
1 4 8 11 14 17 20 23 27 30 33 36 39 42 45 48 51 54 57 60 63
D
0
0,1
0,2
0,3
0,4
0,5
B26
R38
1 4 8 11 14 17 20 23 27 30 33 36 39 42 45 48 51 54 57 60 63
E
0
0,1
0,2
0,3
0,4
0,5
R38
B26 WT
B26ΔxutA_1
B26ΔxutA_2
1 4 7 10 13 16 19 22 25 28 31 34 37 40 43 46 49 52 55 58 61 64 67
Absorbance (Optical Density 600 nm)
Time (hours)