

Taxonomic and functional profiles of soil samples from Atlantic forest and Caatinga biomes in northeastern Brazil

Ralfo G. Pacchioni¹, Fabíola M. Carvalho², Claudia E. Thompson^{2,3}, André L. F. Faustino¹, Fernanda Nicolini¹, Tatiana S. Pereira¹, Rita C. B. Silva¹, Mauricio E. Cantão⁴, Alexandra Gerber², Ana T. R. Vasconcelos² & Lucymara F. Agnez-Lima¹

¹Department of Cellular Biology and Genetics, UFRN, Natal, Rio Grande do Norte, Brazil

²National Laboratory of Scientific Computing, Quitandinha, Petrópolis, RJ, Brazil

³Centro de Biotecnologia, Universidade Federal do Rio Grande do Sul, Porto Alegre, RS, Brazil

⁴Empresa Brasileira de Pesquisa Agropecuária, Embrapa Suínos e Aves, Concórdia, Santa Catarina, Brazil

Keywords

Atlantic forest, bioinformatic, caatinga, comparative metagenomics, pyrosequencing.

Correspondence

Lucymara Fassarella Agnez-Lima, Department of Cellular Biology and Genetics, CB – UFRN, Campus Universitário, Lagoa Nova, Natal, 59078-970, Rio Grande do Norte, Brazil.
Tel: 55-84-3211-9209;
Fax: 55-84-3215-3346;
E-mail: lfagnez@ufrnet.br

Funding Information

This work was supported by Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) and Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES).

Received: 26 September 2013; Revised: 23 February 2014; Accepted: 25 February 2014

MicrobiologyOpen 2014; 3(3): 299–315

doi: 10.1002/mbo3.169

Introduction

Despite soils being the largest reserve of microbial biodiversity, our knowledge about their genetic pool is still limited (Whitman et al. 1998; Torsvik et al. 2002; Curtis and Sloan 2004; Mocali and Benedetti 2010). In general, soil microbial diversity has been mainly accessed through methods based on 16S gene analysis that allow for estimating the diversity richness, but fail in the identification

Abstract

Although microorganisms play crucial roles in ecosystems, metagenomic analyses of soil samples are quite scarce, especially in the Southern Hemisphere. In this work, the microbial diversity of soil samples from an Atlantic Forest and Caatinga was analyzed using a metagenomic approach. Proteobacteria and Actinobacteria were the dominant phyla in both samples. Among which, a significant proportion of stress-resistant bacteria associated to organic matter degradation was found. Sequences related to metabolism of amino acids, nitrogen, and DNA and stress resistance were more frequent in Caatinga soil, while the forest sample showed the highest occurrence of hits annotated in phosphorous metabolism, defense mechanisms, and aromatic compound degradation subsystems. The principal component analysis (PCA) showed that our samples are close to the desert metagenomes in relation to taxonomy, but are more similar to rhizosphere microbiota in relation to the functional profiles. The data indicate that soil characteristics affect the taxonomic and functional distribution; these characteristics include low nutrient content, high drainage (both are sandy soils), vegetation, and exposure to stress. In both samples, a rapid turnover of organic matter with low greenhouse gas emission was suggested by the functional profiles obtained, reinforcing the importance of preserving natural areas.

of functional attributes. Thus, metagenomic studies have contributed toward the understanding of microbial diversity allowing for the identification of taxa and functional profiles, as the abundance of determinate genes have been used as an indicator of biogeochemical processes (Morales et al. 2010; Brankatschk et al. 2013).

Soil characteristics are important factors that affect the microbial diversity and the dynamics of biogeochemical cycles. As an example, soil pH and nutrient availability have

been considered the main factors driving the composition of soil bacterial communities (Lauber *et al.* 2009; Goldfarb *et al.* 2011; Griffiths *et al.* 2011; Kuramae *et al.* 2012). In addition, high temperature, moisture, low pH, and carbon and nitrogen availability contribute to the increase in N₂O emission from soil, but aeration and drought cause reduction in this process (Sangeetha *et al.* 2009; Cleveland *et al.* 2010; Saggiar *et al.* 2013). Even so, our knowledge about biogeochemical processes is still scarce, mainly in relation to tropical soils, and questions about how the soil characteristics together determine the microbial diversity and biogeochemical processes remains unclear.

With climate changes and the increase in greenhouse gas emission, the understanding of microbial diversity and its involvement in these processes will be important for adequate soil management. The highest emissions of CH₄ and N₂O are from tropical soils, but these soils also present high consumption of CO₂ (Montzka *et al.* 2011). However, the majority of these data are from humid tropical regions and little is known about semiarid tropical soils.

In this work, the microbial diversity of soil samples from Parque das Dunas (PD), as a representative of Atlantic Forest biome, and João Camara city, as representative of Caatinga biome, both located in the State of Rio Grande do Norte (RN, Brazil), were investigated using a metagenomic approach. The Atlantic forest is the third largest Brazilian biome, comprising particular ecosystems such as mangroves and forests which span almost the entire Brazilian coast, and is the second largest humid tropical forest in South America. However, less than 10% of the native forest is preserved (Myers *et al.* 2000) and few works have investigated their microbial diversity. The analysis of 16S gene in soil samples from Atlantic Forest of Paraná and Rio de Janeiro showed a dominance of Acidobacteria and Proteobacteria (Bruce *et al.* 2010; Faoro *et al.* 2010). Concerning to gas emission, the analysis of soil samples of Atlantic Forest from São Paulo showed that N₂O emission and the CH₄ uptake are within the range of other tropical forests of the world. The authors also observed that N₂O and CO₂ emissions were lower at higher altitudes, which may be associated to the temperature decrease (Sousa Neto *et al.* 2011). The PD in Natal city (northeastern Brazil) differs from other representatives of Atlantic Forest due to its soil characteristics. The forest growth on the dunes suggests the occurrence of a specific microbial diversity, differing of the microbiota observed in soil from Atlantic forest from Paraná and Rio de Janeiro, which are more clayey soils.

João Câmara (JC) sampling site is located in the semiarid Caatinga, the only exclusive Brazilian biome. The Caatinga occupies 18% of the Brazilian territory, being the most populous semiarid region of the world. It presents a rich diversity of plants, but only 47% of the native vege-

tation is preserved. Due to natural and anthropogenic factors, the Caatinga is considered a fragile ecosystem subject to desertification. Despite the biological importance, the microbial diversity of Caatinga soils is still unknown, but due to severe climate conditions (high temperature, high UV exposure, and long periods of drought), a low and specialized microbial diversity was estimated (Giongo *et al.* 2011; Menezes *et al.* 2012).

Although the Atlantic Forest and Caatinga in northeastern Brazil have specific soil characteristics (both are sandy soils poor in nutrients) and an endemic flora, we investigated the hypothesis that soil microbes show a distinct taxonomic and functional composition in relation to other biomes.

Reports of microbial diversity are scarce in the Southern Hemisphere, providing an invaluable, interesting, and unexplored field of study for metagenomics. Therefore, to our knowledge, this is the first metagenomic study ever conducted in soils of Caatinga and Atlantic Forest biomes that describe taxonomic and metabolic profiles of the microbial community, as well as a comparative analysis between these two different environments and other public metagenomes from different biomes.

Materials and Methods

Sample collection

Sampling was performed in early October 2009 in two different regions of Rio Grande do Norte state, Brazil: the Parque das Dunas (Park of Dunes, PD) in Natal city, an environmental conservation area of Atlantic Forest biome (1172 ha) covered by native coastal dune vegetation, and João Câmara city (JC), a semiarid area of Caatinga biome 73 km from Natal (Table 1 and Figure S1). The collection site at PD (S5° 50.530' O35° 11.598') is located on uneven ground, is grayish in color, under indirect sunlight, and has many roots. The collection site in JC (S5° 30' 51.81" O35° 54' 17.13") is located on flat ground, is dry, dark brown, under direct sunlight, and has few roots (Table 1). The sampling followed Schneegurt *et al.* (2003) recommendations. In brief, after the removal of roots, ~500 g of soil samples were collected (depth 5–10 cm) using a sterile spatula and immediately transferred to sterile 50 mL tubes kept on ice. Organic and physical–chemical analyses of the soil samples were performed by the Laboratory of Soil, Water and Plants Analyses of EMPARN (Natal, RN, Brazil).

DNA preparation, extraction, purification, and sequencing

Initially, samples were sieved through 2 mm sterile sieves in order to eliminate undesired constituents such as roots

Table 1. General features of the studied areas.

	Parque das Dunas (PD)	João Câmara (JC)
Biome	Atlantic forest	Caatinga
Climate zone	Humid tropical	Semiarid
Temperature (annual average)	22.6–29.2°C	21–32°C
UV	High, but indirect due to canopy	High and direct
Rainfall (annual average)	1600 mm	648 mm
Vegetation	Predominantly arboreal, also presenting shrubs and herbaceous. The main families found are Leguminosae, Myrtaceae, Gramineae (Poaceae), Compositae, Euphorbiaceae, Convolvulaceae, and Rubiaceae (Freire 1990)	Xerophilous species, shrubs, thorny and deciduous small trees (Santos <i>et al.</i> 2010)
Soil	Low compaction, sandy, grayish	Compact, dry, dark brown, with few roots

and tiny stones. Afterwards, 10 g of soil samples were subjected to direct extraction and purification of DNA using Power Max™ Soil DNA Isolation Kit (MoBio Laboratories, Inc., Carlsbad, CA) following the manufacturer's instructions.

For sequencing, the libraries were prepared following the instructions of the GS FLX titanium general library preparation method manual (454-Roche), using 5 µg DNA of each sample. The titration, emulsion PCR, and sequencing steps were performed according to the manufacturer's instructions. A four-region 454 sequencing run was performed on a 70 × 75 PicoTiterPlate (PTP) using the Genome Sequencer FLX System (Roche Applied Science, Sao Paulo, Brasil) (Margulies *et al.* 2005). Each library was loaded onto one quarter of the plate. The sequences were deposited in GenBank with accession numbers SRA026684 and SRA026685 and on the MG-RAST server (4459906.3 and 4459907.3).

The replicated sequences generated as artifact of the 454-based pyrosequencing, were eliminated using the Replicates software (Gomez-Alvarez *et al.* 2009), and the sequences <120 bp were removed by the LUCY program (Chou and Holmes 2001). As a result, 27,618 and 22,611 reads were removed in PD and JC, respectively.

The assembly was conducted using the Newbler Assembler 2.5.3. Reads identified as Partial, Repeat, Outlier, Too-Short, and with high-quality discrepancies were filtered from the dataset. Assembling and filtering cycles were performed until the discrepancies were limited to 1% of the total of the reads filtered out at the first assembly step. However, at the cutoff threshold no contigs were assembled.

Taxonomic distribution and statistical analyses

The taxonomic profiles of the metagenomic reads were assigned using the MG-RAST server. In MG-RAST, the species richness was computed as the antilog of the Shannon diversity (Meyer *et al.* 2008). The abundance data

was identified through the lowest common ancestor (LCA), with the parameters $1e^{-05}$ as the maximum *e*-value, a minimum identity of 60%, and a minimum alignment length of 15 as cutoff. The statistical analysis for distinct taxonomic levels from MG-RAST was conducted using the Statistical Analyses of Metagenomic Profiles (STAMP) (Parks and Beiko 2010) software. The significance of the relative proportion difference in taxonomic distribution of PD and JC samples was performed using the two-sided Fisher's exact test, with Newcombe–Wilson confidence interval method. Because *P*-values were not uniformly distributed using Storey's false discovery rate (FDR), Benjamin–Hochberg FDR was applied for correction. Results with $q < 0.05$ were considered significant and the unclassified reads were removed from analyses. The biological relevance of the statistic taxa was determined applying a difference between the proportions of at least 1% and a twofold ratio between the proportions.

A taxonomic analysis was also conducted using the MEGAN4 (Huson *et al.* 2011) software. The given reads were compared against the NR and NT NCBI databases using the BLASTX and BLASTN algorithms (Altschul *et al.* 1997), respectively. Statistical tests on the taxonomic data were also performed with MEGAN. The PD and JC counts were normalized to produce data sets of 100,000 reads. The analysis was performed comparing distinct hierarchical levels and directed homogeneity test was applied to highlight the significant differences in the sample comparisons. The highlighting thickness is logarithmically proportional to its significance; that is the thickness is an integer value of $2\log x$ when $P = 1.0ex$ (Mitra *et al.* 2009). Multiple testing correction analysis was not applied and all unassigned reads were ignored.

Functional analyses

Functional profiles were identified using the SEED subsystems annotation source of the MG-RAST, with $1e^{-05}$ as

maximum e-value, a minimum identity of 60%, and a minimum alignment length of 15. Distinct functional levels from MG-RAST were statistically analyzed in the STAMP, using the same parameters above described for taxonomic distribution. Through the workbench tool from MG-RAST server, we generated subsets of the reads annotated in a functional subsystem for taxonomic identification. The metabolic pathways of biogeochemical cycles were generated using the Kyoto Encyclopedia of Genes and Genomes (KEGG) from MG-RAST server, with $1e^{-05}$ maximum e-value cutoff, 60% of minimal identity, and a minimal alignment length of 15.

A functional analysis using the SEED (Overbeek *et al.* 2005) and KEGG (Kanehisa *et al.* 2012) databases was conducted using the MEGAN4 software. Each sequence was related to its SEED functional role using the best BLAST score to protein sequences without known functional roles. A similar procedure was used to match each sequence to a KEGG orthology (KO) accession number.

A database was constructed using proteins' Refseqs for a number of key enzymes of different and important metabolic pathways. A screening of functional key enzymes was conducted by BLAST using the Refseqs against the PD and JC metagenome. Matches with alignment scores higher than 80 were retained.

Comparative metagenomic analysis

The taxonomic and SEED subsystems profiles of metagenomic samples from soil, water, and host-associated samples were obtained from MG-RAST server. Samples belonging to rhizosphere, temperate and tropical forests, marine habitat, host-associated, and desert biomes were included. The criteria applied for inclusion were a maximum e-value cutoff of $1e^{-05}$, a minimum identity of 60%, and a minimum alignment length of 15. The metagenomes included in the analysis were 4440463.3, 4440939.3, 4444130.3, 4444164.3, 4444165.3 (animal-associated habitat), 4465556.3, 4449956.3 (rhizosphere biome), 4477805.3, 4477872.3, 4477901.3, 4477904.3 (des-

ert biome), 4443713.3, 4441057.4, 4441586.3, 4441578.3 (marine habitat), 4477876.3, 4477877.3, 4477899.3 (temperate forest), 4477807.3 and 4477875.3 (tropical forest). Trends in the abundance of the taxonomy and the SEED subsystems were examined using Principal Component Analysis (PCA) through the multiple groups analysis of STAMP, in which the statistical test applied was analysis of variance with Games–Howell post hoc test and Benjamin–Hochberg FDR for correction. For the comparison between two groups, the Welch's t-test, the Welch's inverted test for confidence interval method and Benjamin–Hochberg FDR for correction were applied. The relative proportion difference in functional distribution of PD and JC samples was considered significant when $q < 0.05$. The unclassified reads were removed from analyses.

Results

Organic and physical–chemical characteristics of soil samples

The organic and physical–chemical parameters of PD and JC soil samples are summarized in Table 2. Both samples are classified as very acidic soils (pH < 5.0) and exhibit a large portion of sand, particularly in PD. However, the PD soil shows lower values for all the components analyzed

Table 3. Taxonomic profile of PD and JC samples to domain level, computed by MEGAN and MG-RAST.

Domain	MEGAN		MG-RAST	
	PD	JC	PD	JC
Archaea	266	662	232	366
Bacteria	88,786	106,396	74,299 ¹	93,371 ¹
Eukaryota	7489	794	5312 ¹	701 ¹
Viruses	11	25	5 ¹	21 ¹

¹Differences statistically significant between PD and JC samples ($P < 1e^{-15}$) by STAMP.

Table 2. Organic and physical–chemical parameters of PD and JC soil samples.

Sample	pH in water	Ca	Mg	Al	H + Al	P	K	Na	N	Organic matter	C:N
		(cmol.dm ⁻³)			(mg.dm ⁻³)	(mg.dm ⁻³)	(g.dm ⁻³)				
PD	4.99	0.30	0.14	0.10	1.32	2	12	4	0.31	11.26	21:1
JC	4.60	9.5	23.5	2.25	7.76	4	305	199	1.10	24.6	12.7:1
Granulometry (%)		Sand		Clay		Silt		Texture classification			
PD		97.9		2		0.1		Sandy			
JC		58.2		12		29.8		Sandy loam			

and higher C:N ratio when compared to the JC sample, an indicator of a deficiency in nutrients and minerals.

General characteristics of the metagenomes

The soil DNA sequencing resulted in 147,278 and 151,274 reads from PD and JC, with a total of 64,214,298 and 68,328,253 bb, an average length of 436 ± 63 and 451 ± 60 bp, and the GC content of 61 ± 8 and $66 \pm 7\%$, respectively (Table S1). Reads identified as artificial duplicate was removed by the MG-RAST. After the quality control, 155,805 proteins were predicted for PD sample and ~53% of the total reads were annotated as proteins functionally assigned. For JC metagenome, 164,449 proteins were predicted and 61.6% of the total reads were identified functionally (Table S1). The algorithm implemented by MEGAN assigned more sequences when compared to MG-RAST and it was able to identify a higher number of sequences related to *Bacteria*, *Archaea*, *Eukarya*, and *Viruses* (Tables S1 and 3), probably due to the higher number of available sequences in the reference database. The species richness estimated through the α -diversity index showed that the metagenome from JC presents α -diversity of 448.587, whereas for PD the species richness is 441.864.

Comparative taxonomic profiles

For phylum level, the microbiota profile generated by MG-RAST was similar in the PD and JC samples. The more abundant phylum in both samples were Actinobacteria (27.8% and 36.4%), Proteobacteria (26.1% and

24.8%), and Acidobacteria (9.1% and 2.4%). However, the statistic difference was observed for Proteobacteria, Acidobacteria, and Chlamydia, more frequent in PD sample, and Actinobacteria, Bacteroidetes, and Cyanobacteria with highest frequency in JC. Concerning Archaea and Eukarya phyla, the highest occurrence of Thaumarchaeota was observed in JC sample, while Ascomycota was predominant in PD sample. The MEGAN analysis showed similar results for phylum representation (data not shown).

The Actinobacteria (27.8% and 36.4%) and Alphaproteobacteria (14.86% and 12.41%) were the predominant classes in PD and JC. The classes Alphaproteobacteria, Solibacteres, and Acidobacteria are more frequent in PD sample (Fig. 1), with the highest representation of the orders Rhizobiales, Solibacterales and Acidobacteriales according to STAMP analysis. In JC sample, Actinobacteria, Deltaproteobacteria, and unclassified Cyanobacteria were predominant (Fig. 1), with overrepresentation of the orders Actinomycetales, Sphingomonadales, and Mixococcales. The statistical analyses implemented by MEGAN showed that PD metagenome has a significantly higher number of reads related to Acidobacteria, Alphaproteobacteria, and Planctomycetia classes. Moreover, Betaproteobacteria, Gammaproteobacteria, and Deltaproteobacteria are statistically more represented in the JC metagenome (Figure S2).

Both metagenomes showed a similar microbial composition considering some of the most frequently found genera, despite the difference in their representation (Fig. 2). The PD sample showed predominance of *Candidatus solibacter*, *Candidatus koribacter*, *Acidobacterium*

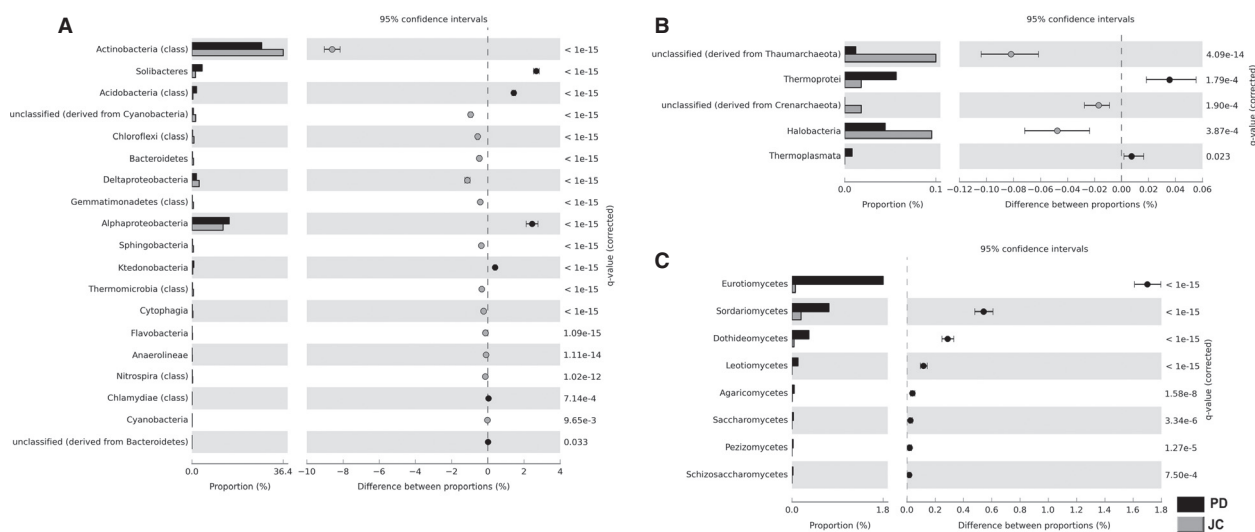


Figure 1. Comparative taxonomic profile of the PD and JC samples at class level, computed by MG-RAST. Classes with significant biological differences ($P < 0.05$, difference between the proportions $> 1\%$ and twofold of ratio between the proportions, STAMP) for the Bacteria domain (A); for the Archaea domain (B); and for the Eukaryota domain (C).

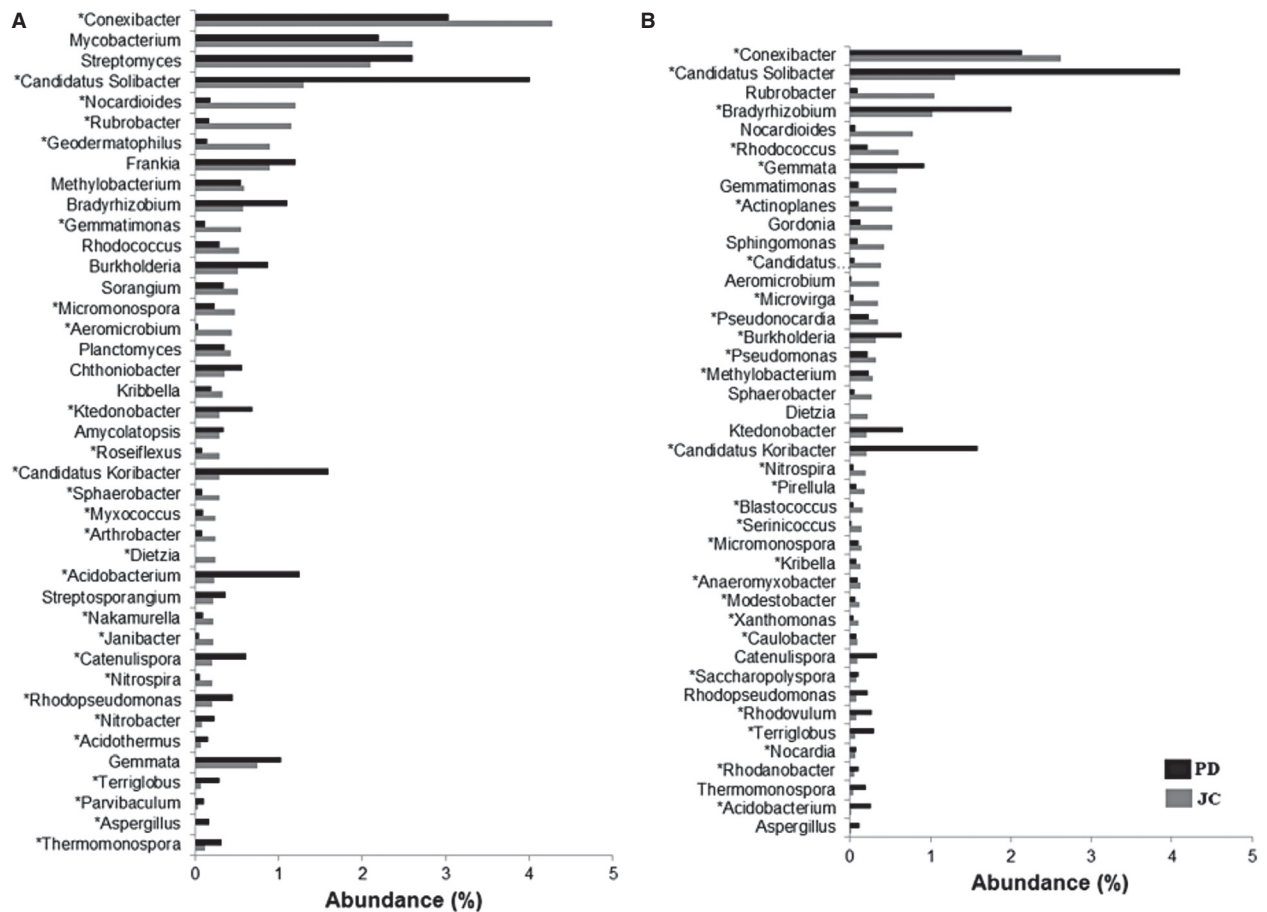


Figure 2. Comparative taxonomic profile of the PD and JC samples at genus level, computed by MG-RAST (A) and MEGAN (B). The twenty most abundant genera found in each samples are shown. *Significant differences between PD and JC samples ($P < 0.05$, difference between the proportions $>1\%$ and twofold of ratio between the proportions).

(Acidobacteria), *Ktedonobacter* (Chloroflexi), and *Catenulispora* (Actinobacteria). In JC sample, the genera with the significant representation were *Conexibacter*, *Nocardioides*, *Rubrobacter*, *Geodermatophilus* (Actinobacteria), and *Gemmatimonas* (Gemmatimonadetes) (Fig. 2). For PD, *Bradyrhizobium*, *Burkholderia*, *Microvirga* (Proteobacteria), and *Gemmata* (Planctomycetes) were significant only considering the MEGAN analysis. Additionally, *Rhodococcus*, *Actinoplanes* (Actinobacteria), and *Candidatus choracidobacterium* (Acidobacteria) were indicated as statistically meaningful for JC metagenome by MEGAN (Fig. 2).

Among Archaea classes, Halobacteria (Euryarchaeota phylum) and Thermoprotei (Crenarchaeota phylum) showed significant results in JC and PD metagenomes, respectively, although with low occurrence (Fig. 1).

The PD and JC samples presented a wide variation in hits for Eukaryota domain and low occurrence of viruses. Despite that there was a similar representation of the most abundant phyla in the two metagenomes, the number of sequences assigned to Eukaryota was around

10-fold higher in PD. The major contribution to Eukaryota microbiota in PD came from Ascomycota phylum, with predominance of Eurotiomycetes, followed by Sordariomycetes and Dothideomycetes classes (Fig. 1). Among Eurotiomycetes, *Aspergillus* was the most frequent genus, with 0.171% of all hits for PD, according to the MG-RAST analyses. The most abundant Ascomycota in the JC sample was *Nectria haematococca* mp VI 77-13-4, which is a plant pathogen and it is the teleomorph (sexual reproductive stage) of *Fusarium solani*.

As the tropical forest and arid soil representatives, the PD and JC samples were compared with public metagenomes. Interestingly, the principal component analysis showed that PD and JC were more similar to each other than to other metagenomes (Fig. 3), and were considered as one group in order to compare with other groups such as rhizosphere and desert.

The taxonomic profile indicates that the PD and JC group clustered near desert, with 80.2% and 72.4% of variance being explained by the first two components for

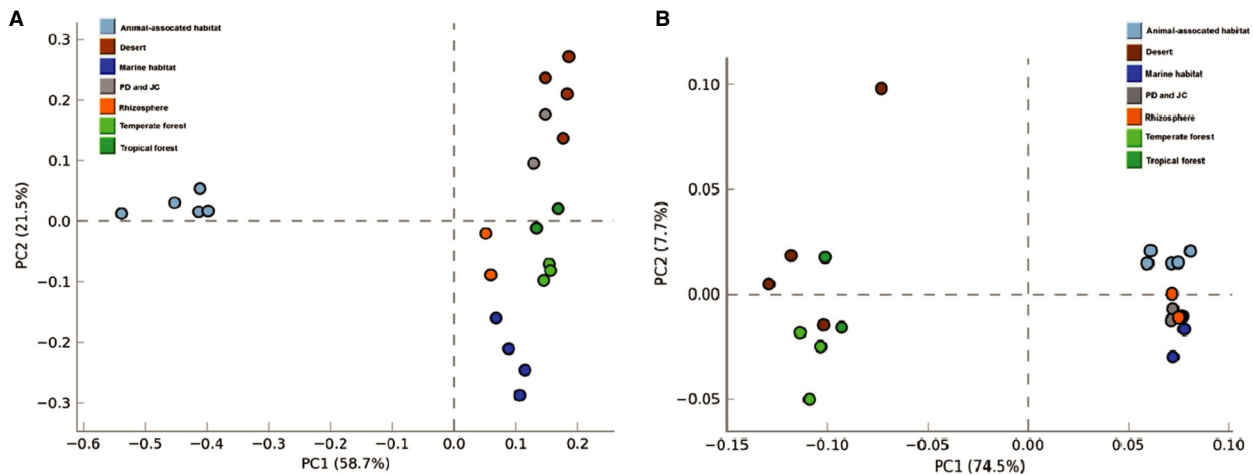


Figure 3. Trends in the PD and JC taxonomy for the class level (A) and for SEED subsystems at level 1 (B) examined using Principal Component Analysis (PCA) through the STAMP software, based on multiple group analysis, applying ANOVA test, Games–Howell post hoc test for confidence interval method and Benjamin–Hochberg FDR for correction.

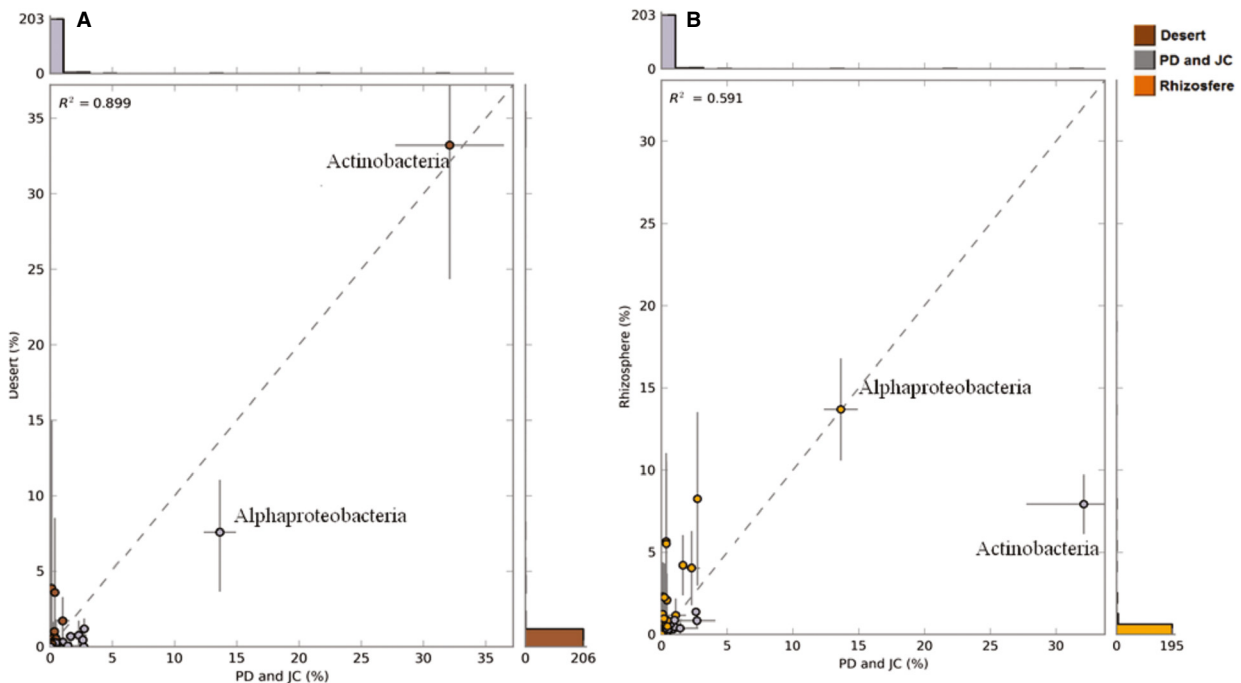


Figure 4. Comparison computed using two group analysis at class taxonomic level for PD and JC versus deserts (A) and PD and JC versus rhizospheres (B). The Welch's *t*-test, the Welch's inverted test for confidence interval method and Benjamin–Hochberg FDR for correction were applied. Data related to relative frequency. Significant differences were not observed.

class and genus, respectively (Fig. 3A). Although not significant, the PD and JC samples differ from desert soils in the Alphaproteobacteria proportion, which is two times higher in the Brazilian biomes (13.6% and 7.5%) (Fig. 4). In contrast, the desert samples showed a slight increase in Actinobacteria (33.2% and 32.1%) (Fig. 4), especially in relation to the *Rubrobacter* genus (3.3% and 0.6%), and a lower *Mycobacterium* representativeness (0.2% and 2.4%)

(Fig. 5). Moreover, the analyses showed that in PD and JC there is an overrepresentation of *Methylobacter*, *Nitrococcus*, and *Psychromonas* genera (Gammaproteobacteria class), which are statistically relevant (Fig. 5). In deserts, these genera are practically absent.

In comparison with metagenomes from the rhizosphere, PD and JC showed high divergence in the Actinobacteria abundance (32.1% and 7.9%, respectively)

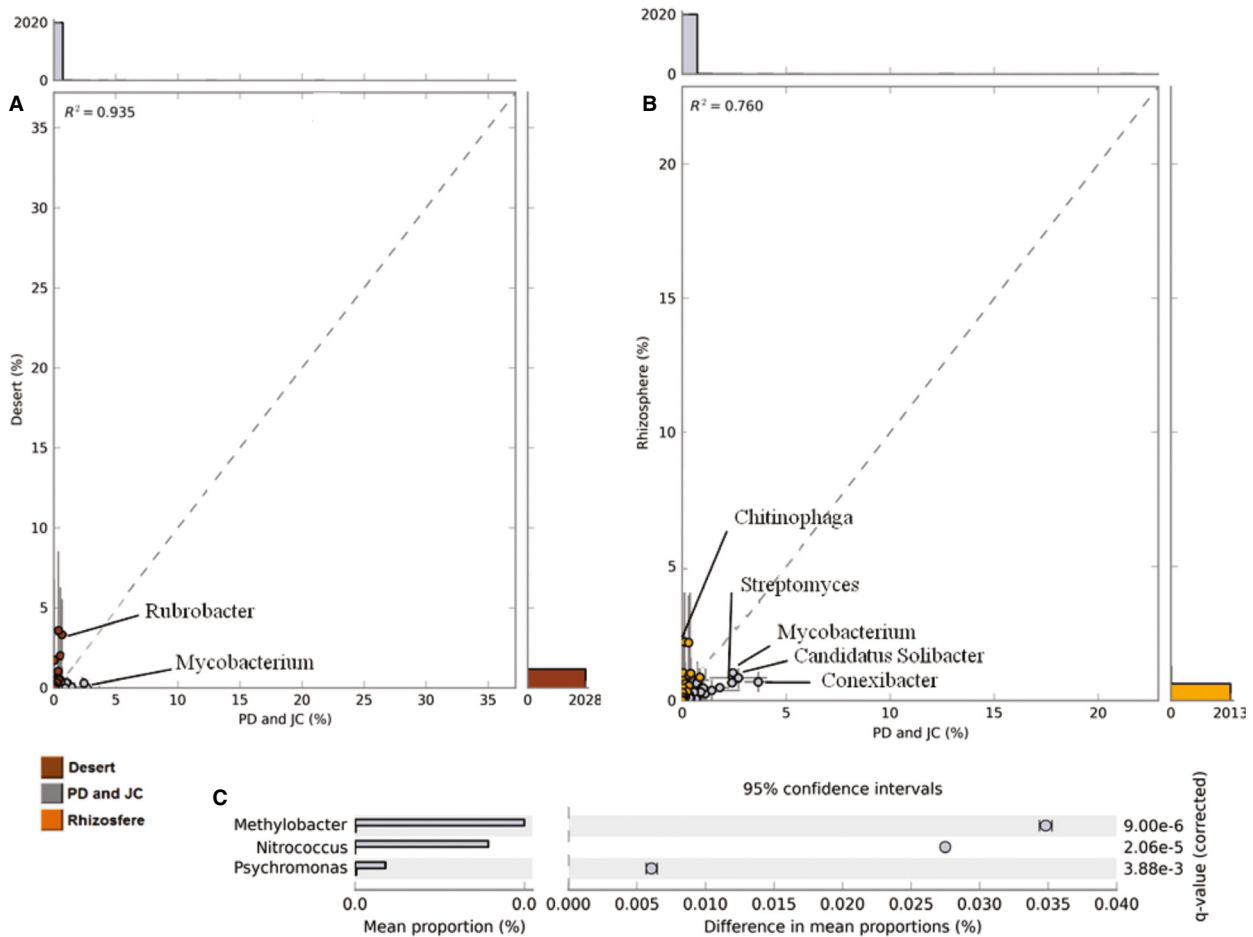


Figure 5. Comparison computed using two groups analysis at genus level for PD and JC versus deserts (A). The significant genera are shown in (B). Genera with difference of relative frequency between PD and JC versus desert (C). The Welch's *t*-test, the Welch's inverted test for confidence interval method and Benjamin–Hochberg FDR for correction were applied.

(Fig. 4), particularly with *Conexibacter* (3.6% and 0.6%), *Mycobacterium* (2.4% and 1.0%), and *Streptomyces* (2.3% and 0.6%) genera (Fig. 5). In contrast, Rhizosphere presented an elevated proportion of Betaproteobacteria (8.2% and 2.7%), Gammaproteobacteria (4.2% and 1.6%), and Deltaproteobacteria (4% and 2.2%) (Fig. 4). At genus level, a higher abundance of *Chitinophaga* was evident in rhizosphere (2.1% and 0.1%, respectively) (Fig. 5). Despite the divergence observed in the class abundance, it was not statistically significant.

Comparative functional profiles

The functional profile obtained using MG-RAST and STAMP did not show discrepant difference in proportion for subsystems at level 1 (considering differences of at least 1% and a twofold ratio between the proportions) (Fig. 6). However, some subsystems of level 1 present a higher representation in PD or JC with a $P < 0.05$

(Fig. 6). In addition, significant differences in proportion were observed at levels 2 and 3. At level 1, the carbohydrates subsystems and clustering-based subsystem, (which groups hypothetical protein families based on conserved colocalization across multiple genomes), were the most abundant in both JC and PD sample. At level 3, the subsystems serine-glyoxylate cycle and YgfZ showed the highest representation. Subsystems such as amino acids and derivatives, DNA metabolism, stress response and nitrogen metabolism are highlighted in JC sample, mainly in relation to genes involved with degradation of amino acids, DNA repair and replication, ammonia assimilation and nitrate and nitrite ammonification. PD sample shows the highest representation of the subsystems virulence, disease and defense, metabolism of aromatic compounds, and phosphorus metabolism, with significant representation of functions related to resistance to antibiotics and toxic compounds, benzoate degradation, and phosphorus uptake (Fig. 6). The analysis of taxonomical distribution

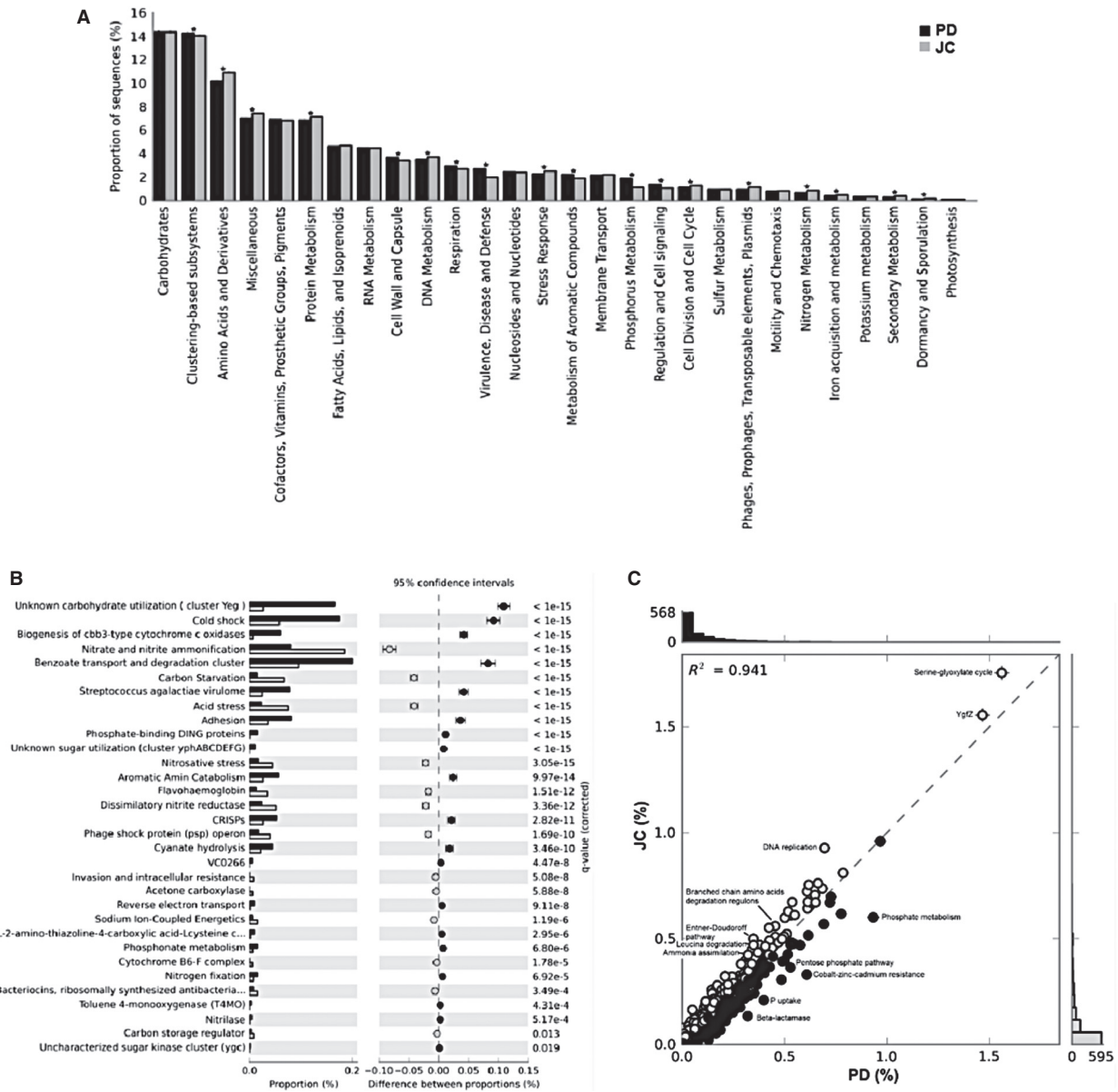


Figure 6. Comparative functional profile of PD (in black) and JC (in gray) samples identified by MG-RAST and statistically analyzed by STAMP. (A) Subsystems at level 1 (* $P < 0.05$). (B) Subsystems at level 2 ($P < 0.05$, difference between the proportions $> 1\%$ and a twofold ratio between the proportions). (C) Abundance of subsystems at level 3.

among the reads annotated in each subsystem showed that in PD the genera *Rhodospseudomonas*, *Candidatus solibacter*, *Candidatus koribacter*, *Bradyrhizobium*, *Burkholderia*, *Nitrobacter*, and *Frankia* present the highest contribution for the functional differences observed, while in JC the main genera found were *Mycobacterium*, *Nocardioidea*, *Rubrobacter*, and *Burkholderia* (Figure S3). The functional profile obtained using MEGAN was similar to the MG-RAST data (data not shown).

The biogeochemical cycle analyses evaluated the nitrogen and methane metabolism (Fig. 7), showing that the

metagenomes have similar profiles in relation to the enzymes involved in these metabolic pathways, with some exceptions. Concerning to the nitrogen metabolism (Fig. 7A), singular patterns in the proportion of hits were observed in JC. A higher representation of enzymes associated to ammonia production and conversion into amino acids was found. In the methane metabolism cycle (Fig. 7B), PD showed a higher occurrence of the carbon monoxide dehydrogenase (ferredoxin) (EC 1.2.99.2), an enzyme involved in the oxidation of CO to CO₂, while JC had a higher abundance of catalase (EC 1.11.1.6) and

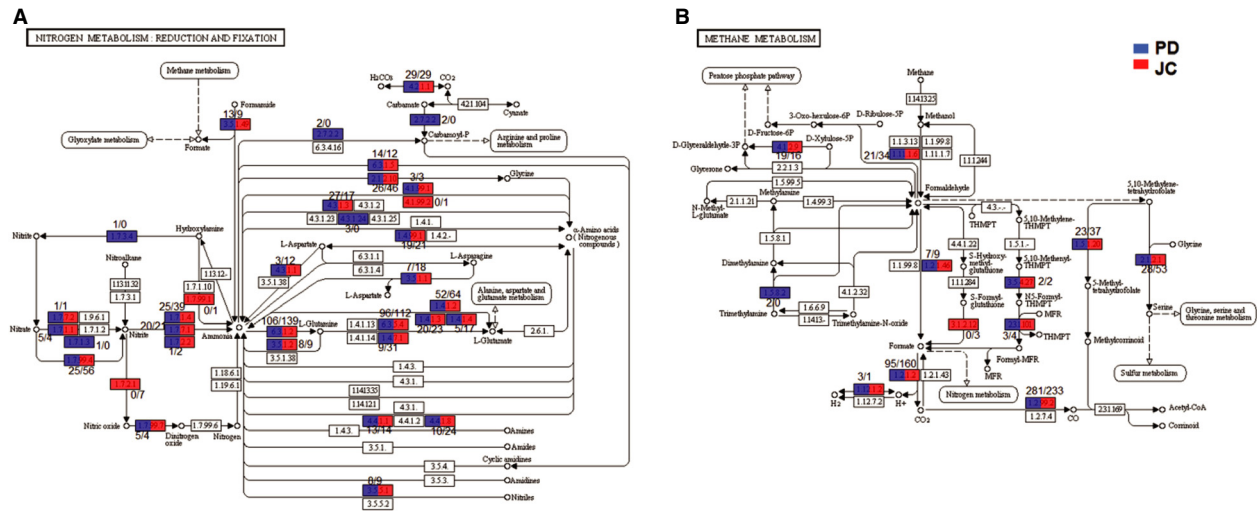


Figure 7. Nitrogen (A) and Methane (B) metabolic pathways performed by KEGG mapper from MG-RAST, with the hit number obtained for each EC number in relation to PD (blue) and JC (red) (adapted by Kanehisa – www.genome.jp/kegg).

glycine hydroxymethyltransferase (EC 2.1.2.1), which acts in the methylenetetrahydrofolate to serine conversion.

The screening of the functional key enzyme (Fig. 8) showed a dominance of Proteobacteria and Actinobacteria hits. Although Acidobacteria was one of the dominant phyla in the PD sample, few hits assigned to this phylum have been identified among the key genes analyzed. The highest number of hits was observed for virulence and pathogenicity, CO oxidation, and acidity. Differences between the samples were identified in the carbon fixation, CO oxidation, and acidity resistance categories, which were predominant in the PD sample, while the virulence, pathogenicity, and nitrate respiration were predominant in the JC sample.

When comparing PD and JC reads with public metagenome considering the functional profiles, the principal component analysis showed that PD and JC were more similar to the rhizosphere metagenome (Fig. 3B), differing from the taxonomic profile which showed PD and JC metagenomes close to desert microbiota (Fig. 3A).

As a group, PD and JC did not present significant categories in comparison with the rhizosphere samples. However, when comparing a rhizosphere sample to PD or JC individually, the rhizosphere metagenomes have more sequences related to the DNA metabolism and iron acquisition and metabolism at level 1, whereas PD presented a higher number of hits related to the carbohydrate and phosphorus metabolism. In the JC sample, the amino

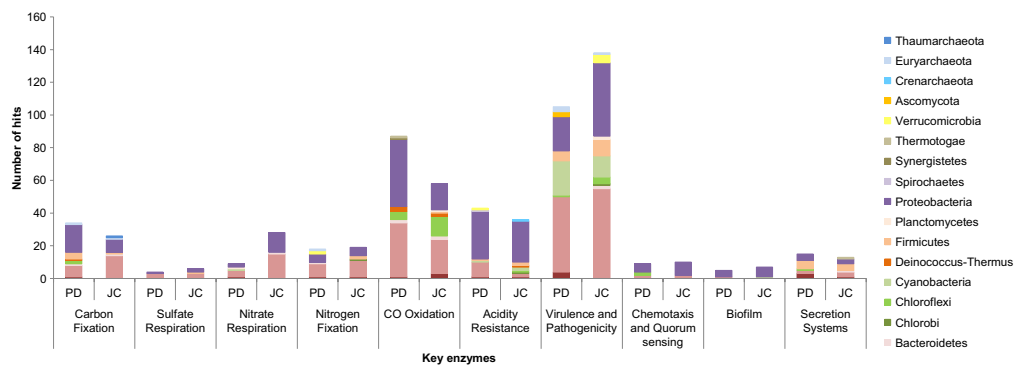


Figure 8. Comparison between PD and JC samples for key enzymes coding genes of distinct metabolic pathways: Carbon Fixation (*abfD*, *cbfQ*, *cbfO*, *cbfX*, *cbfR*), Sulfate Respiration (*dsrA*), Nitrate Respiration (*nirK*, *nirS*, *narG*, *narH*, *napA*, *napB*, *napC*, *nosZ*), Nitrogen Fixation (*fixA*, *fixB*, *fixC*, *nifA*, *nifH*), CO Oxidation (*coxL*, *coxM*, *coxS*), Acidity Resistance (*idhD*, *atpA*, *aldB*, *cfa*, *groEL*, *pstS*, *phoBDRSU*), Virulence and Pathogenicity (*katG*, *aroG*, *ohr*, *cpkA*, *prfA*, *palFA*, *sdh1*, *mpg1*, *pth11*, *acr1*, *ace1*, *pela*, *cut1*, *pth3*, *pth8*, *pkS1-15*, *tri5*, *fsr1*, *magB*, *mgv1*, *mps1*, *pmk1*), Chemotaxis and Quorum sensing (*pilG*, *pilH*, *pilJ*), Biofilm (*hnsP*), and Secretion Systems (TIII Type – *escV*, *pscJ*, *fliI*, *flhA/ATIV* Type – *virB4*, *virB11*, *cpaF*).

acids and derivatives, fatty acids, lipids, and isoprenoids categories are increased (Figure S4).

A significant functional variation was observed at level 1 related to cell wall and capsule; motility and chemotaxis, amino acids and derivatives, nucleosides and nucleotides, and membrane transport, which were all predominant in the PD and JC group (Figure S5) when compared to desert metagenomes.

Discussion

The soil samples analyzed in this work are classified as sandy soil with very low nutrient content. The data obtained from JC sample are similar to that observed in the majority of the soils from Caatinga biome (recently reviewed by Giongo *et al.* (2011) and Menezes *et al.* (2012)). In contrast, PD sample differs from other Atlantic Forest soils that show a higher organic matter content and clay proportion with higher water holding capacity (Faoro *et al.* 2010; Sousa Neto *et al.* 2011; Vieira *et al.* 2011). These differences may be associated to the distinct taxonomic and functional profiles discussed below, as soil characteristics such as nutrient availability, moisture, and texture are determinants for microbial diversity and consequent biogeochemical processes (Arias *et al.* 2005; Chau *et al.* 2011; Saggari *et al.* 2013).

The microbial diversity found in PD soil show a distinct profile compared to the Atlantic Forest of Rio de Janeiro and Paraná, also differing from soils of Amazon and Cerrado, where Acidobacteria phylum was dominant, accounting for between 29 and 63% of 16S rRNA sequences obtained (Jesus *et al.* 2009; Bruce *et al.* 2010; Faoro *et al.* 2010; Araujo *et al.* 2012). Moreover, a high occurrence of Acidobacteria was observed in soils from other subtropical or tropical moist forests (Kim *et al.* 2012; Nie *et al.* 2012). It has also been one of the main groups found in arid or semiarid, ranging from 4% to 18% of the sequences of 16S (Chanal *et al.* 2006; Bachar *et al.* 2010; Aguirre-Garrido *et al.* 2012), contrasting with the JC sample, where only 2.5% of the sequences corresponded to the Acidobacteria phylum.

The genome analysis of three Acidobacteria species indicated the presence of cellulose synthesis genes and excreted proteins suggesting potential traits for desiccation resistance (Ward *et al.* 2009). However, the Acidobacteria physiology is still relatively unknown and data about temperature and ultraviolet resistance are scarce. Furthermore, the discrepancy between our data and those obtained in other Brazilian biomes soils may be related to sand content in PD and JC samples, since the abundance of Acidobacteria was found to be higher in the clay than in the sand or silt fractions (Liles *et al.* 2010; Russo *et al.* 2012). Other aspect that is important to consider is the

different methodology used for taxonomic analysis. In previous works about Atlantic forest, Amazon and Cerrado, the analysis was based on 16S sequences (Jesus *et al.* 2009; Bruce *et al.* 2010; Faoro *et al.* 2010; Araujo *et al.* 2012). In this work, our analysis was based on LCA approach of the MG-RAST and MEGAN using the total DNA sequences. The 16S rRNA analysis has been efficient for taxonomic identification, however, expanding the number of markers to include other highly conserved genes has improved the phylogenetic resolution. Methods based on LCA or other parsimonious evolutionary principles are useful to reduce false-positives generated by tools based on homology, increasing the analysis robustness and permitting more precise taxa abundance estimation (Clemente *et al.* 2010; Guo *et al.* 2013; Segata *et al.* 2013).

Our data also showed that Actinobacteria was the dominant phylum in both samples. A high occurrence of Actinobacteria in semiarid soils (20–50%) has been previously reported (Chanal *et al.* 2006; Bachar *et al.* 2010; Köberl *et al.* 2011; Aguirre-Garrido *et al.* 2012), as observed in JC sample (36.4%). However, its occurrence in forest is generally low (<15%) (He *et al.* 2006; Lin *et al.* 2010; Nie *et al.* 2012), especially in Atlantic forest and Amazon soils (<5%) (Jesus *et al.* 2009; Bruce *et al.* 2010; Faoro *et al.* 2010), differing from PD sample, in which Actinobacteria recorded 27.8% of the hits. Some authors have proposed that in vegetated soil, rhizosphere zone or under plant canopy, the Proteobacteria occurrence is high, while barren soils are characterized by Actinobacteria or Acidobacteria abundance (Bernard *et al.* 2007; Thomson *et al.* 2010; Bachar *et al.* 2012). In sandy soil, Actinobacteria is one of dominant classes found (Russo *et al.* 2012). Soil water content is a determinant factor for Actinobacteria abundance, which increases in arid soil due to the resistance of several species to drought stress (Connon *et al.* 2007; Brockett *et al.* 2012).

Comparative metagenomics analyses have indicated that the substrate (*i.e.*, soil or water) plays a fundamental role in determining the taxonomic and functional profiles of microbial communities. Therefore, soil samples tend to be more similar to each other in relation to taxonomy and the presence of environment-specific genes than samples from other environments (Tringe *et al.* 2005; Jeffries *et al.* 2011). In PD, sand and nutrients contents and vegetation seems to be the most important factors for microbial diversity, while in JC, in addition to these same factors, stress conditions (caused by temperature, UV and drought) also affect the microbial diversity. Our samples differ from other soils due to high content of sand and differ from desert biomes due to high occurrence of superior plants. These characteristics may explain the data obtained in our PCA analysis, which showed taxonomic

profile of PD and JC similar to desert while the functional profiles were similar to rhizosphere biomes.

Among the genera more represented in PD and JC samples, there are mainly members of Proteobacteria, Actinobacteria, and Acidobacteria phyla. Species of the genera as *Rhodopseudomonas*, *Bradyrhizobium*, *Candidatus solibacter*, *Candidatus koribacter*, *Nocardioides*, *Rubrobacter*, and *Geodermatophilus* are described as important in N and C fixation and organic matter degradation, and some of them, as *Rubrobacter xylanophilus* and *Geodermatophilus obscurus* (mainly found in JC sample), are among the most resistant species to gamma and UV radiation (Jothmani *et al.* 2003; Iwai *et al.* 2005; Starkenburg *et al.* 2008; Ward *et al.* 2009; Ivanova *et al.* 2010; Chikere *et al.* 2011; Torres *et al.* 2011; Yuan *et al.* 2012). Compared to desert biomes, PD and JC showed a higher occurrence of the genera *Nitrococcus*, *Psychromonas*, and *Methylobacter*. These genera are described as involved in N and C cycles and stress resistant, including resistance to desiccation and salt stress (Bowman *et al.* 1993; Koops and Pommerening-Röser 2001; Riley *et al.* 2008; Ward 2008).

Fungi are also an essential component of terrestrial ecosystems by acting as organic matter decomposers, pathogens and plant-mutualists (Anderson *et al.* 2003; Hunt *et al.* 2004; Buée *et al.* 2009). The higher number of reads found in PD when compared to the JC sample may be explained by the plant diversity found in this biome, since the most represented genera identified are saprophytic or plant pathogens. It has been proposed that fungi are more important for the degradation of complex C source such as cellulose and lignin, while bacteria are more competitive in degrading simple C source. Usually, Basidiomycota is the dominant phylum found in soil samples (Hunt *et al.* 2004; Buée *et al.* 2009). This is in contrast to our data that indicated Ascomycota as more frequent, which is mainly attributed to the genus *Aspergillus*, known as saprophytic species and also an opportunistic pathogen (Klich 2002; Horn 2003; Amaike and Keller 2011).

The rapid turnover of organic matter is observed in soils with high temperature and intermediary moisture that are the best conditions for aerobic decomposition, leading to a lower organic matter accumulation and high mineralization rate (Sayer 2006; Fierer *et al.* 2007), favoring the growth of copiotrophic bacteria as Proteobacteria (Bernard *et al.* 2007; Thomson *et al.* 2010). The characteristics of vegetation are an important factor that affects the organic matter decomposition since recalcitrant compounds are more resistant to microbial degradation (Allison and Vitousek 2004; DeAngelis *et al.* 2011). As representative of the Atlantic forest, PD has a rich diversity of plants, predominantly arboreal, although herbaceous vegetation such as grasses is also found (Freire 1990). Caatinga biome has a lower diversity of plants

characterized by deciduous shrubs and xerophilous species (Santos and Santos 2008; Santos *et al.* 2010, 2012). These characteristics may contribute to the taxonomic and functional profiles found in JC and PD.

The N and P contents in soils are limiting factors, as microorganisms and plants compete for the nutrients. In soils presenting a C:N ratio less than 20:1, the organic matter decomposition occurs quickly, while in soils presenting C:N ratio greater than 20:1, the decomposition is slow (Peng *et al.* 2002; Rennenberg *et al.* 2009; Richardson and Simpson 2011). In this condition, the growth of N₂-fixing bacteria may be favored as observed in the PD sample. Other limiting factor observed in the PD soil is the low phosphorous content that may be related to the highest occurrence of hits annotated in phosphorus metabolism in PD compared to the JC metagenome. These differences may be attributed to genera *Rhodopseudomonas*, *Candidatus solibacter*, *Candidatus koribacter*, *Bradyrhizobium*, *Burkholderia*, *Nitrobacter*, and *Fankia*, which presented the highest representation among the hits related to N, C, and P metabolisms compared to JC (Figure S3).

The low nutrient retention in sandy soils suggests that plants play an important role in maintaining the biological diversity due organic matter degradation and/or roots' exudates in rhizosphere that includes compounds that may be used as carbon sources by microorganisms (Bais *et al.* 2006; Jones *et al.* 2009). This may explain the highest occurrence of hits associated to the aromatic compounds metabolism and the bacterial defense against toxic compounds, found mainly in PD samples.

In JC soil, the stress caused by high temperatures, UV exposure, and long drought periods seems to be an important trigger factor for microbial diversity. This explains the occurrence of bacteria resistant to stress such as many Actinobacteria genera, as *Mycobacterium*, *Rubrobacter*, and *Nocardioides*. In fact, the highest frequency of functional categories related to the DNA metabolism, mainly DNA repair and DNA replication, and oxidative and osmotic stress were found JC metagenome. Concerning the N cycle, the nitrate and nitrite ammonification and ammonia assimilation are more represented in the JC metagenome. These are processes possibly related to organic matter decomposition and mineralization (Rennenberg *et al.* 2009).

Genes related to ammonia oxidation (nitrification-amoA, B and C subunits) were not identified (data not shown) in PD and JC samples, despite the involvement of Bacteria and Archaea species in this N cycle step (Leininger *et al.* 2006; Di *et al.* 2009). The low occurrence of genes related to N₂O production as *NirS* (only 1 hit was found in PD and 1 in JC) suggest a low potential for greenhouse gas production in these soils, as the

abundance of this gene was proposed as an indicator of greenhouse gas emission (Morales et al. 2010). Corroborating this hypothesis, a high occurrence of genes involved in nitrite and ammonia assimilation suggests the retention of N and consequently avoiding the loss by denitrification, which is important in an N poor environment.

Additionally, it is interesting to observe that both samples have a good representation of bacteria families described as N₂-fixer. However, in the PD sample the highest abundance of these bacteria was found, especially when considering the Alphaproteobacteria class, Rhizobiales order. The occurrence of *Bradyrhizobium*, *Rodopseudomonas*, and *Nitrobacter* genus infers the important role of these microorganisms in N and C fixation and in biodegradation of aromatic compounds (Jothimani et al. 2003; Starkenburg et al. 2008; Torres et al. 2011; Yuan et al. 2012).

Moreover, the CO oxidation genes, especially the CO dehydrogenase, are more represented in the PD sample. The CO oxidation has been used by microorganism as a source of energy and carbon (King and Weber 2007). Soil carbon stocks are affected by addition/decomposition of organic matter. Acidic soils, as observed for PD and JC samples, are generally the most active to remove CO from air (Inman et al. 1971; Bartholomew and Alexander 1979). This explains the occurrence of CO oxidation and acidity-related genes, particularly for the PD sample which has the lowest organic matter content. Furthermore, the higher CO concentration prevents the growth of nitrate-respiring organisms, and the lower oxygen concentration in the soil samples enables some aerobic CO-oxidizers to obtain energy in an anaerobic and nitrate-independent manner (King 2006). An important source for volatile compounds such as CO and CO₂ is the photodegradation of organic matter (Schade et al. 1999; Brandt et al. 2009), which is a dominant process in semi-arid ecosystems during exposure to solar radiation (Austin and Vivanco 2006; Brandt et al. 2007; Day et al. 2007; Gallo et al. 2009). In agreement, JC sample showed the highest occurrence of CO₂-fixation hits, mainly related to the *Mycobacterium* genus (Figure S3).

Another curious finding is the high occurrence of the gene *abfD* (4-hydroxybutyryl-CoA dehydratase) involved in 3-Hydroxypropionate/4-hydroxybutyrate cycle, a CO₂ fixation process firstly identified in Archaea species (Berg et al. 2007). It seems to be frequent in Bacteria phyla, while the classical bacteria RuBisCO genes are poorly represented (only 3 hits were found). However, this data should be viewed with caution since the role of these bacterial counterparts still remains unclear (Ettema and Andersson 2008; Ivan et al. 2008).

The subsystem serine-glyoxylate cycle, which is another pathway for C fixation, is well represented in the JC and PD samples and may be associated to organic matter deg-

radation. This is an alternative pathway for mono carbon (C1) assimilation in methylotrophic bacteria as some *Mycobacterium* species. Methanol is very abundant in soil due to degradation of pectin and lignin (Kolb 2009). In addition, YgfZ, a folate-dependent regulatory protein involved in one-carbon metabolism (Teplyakov et al. 2004), is also well represented in both JC and PD. Other indications of this process is the occurrence of genes related to Entner–Doudoroff pathway, as this alternative path for catabolism of glucose to pyruvate was also associated to pectin degradation in some bacteria (Paster and Canale-Parola 1985; Slov  kov   et al. 2002).

In conclusion, even with metagenomics being a powerful tool in the study of microbial biodiversity, much remains to be understood about the biogeochemical processes in soils, requiring a multidisciplinary approach. Although there is a high similarity between the PD and JC samples considering the higher taxonomic and functional levels, significant differences were found in lower hierarchical categories. This was mainly related to the habitat-specific characteristics such as nutrient level, vegetation, and stressful conditions. In both samples, a rapid turnover of organic matter with low greenhouse gas emission was suggested by functional profiles obtained, reinforcing the importance of preserving natural areas. Our data contribute to the understanding of soil microbial diversity in seldom assessed environments up to date.

Acknowledgments

We would like to thank the staff of LNCC for insightful discussions and comments. This work was supported by Conselho Nacional de Desenvolvimento Cient  fico e Tecnol  gico (CNPq) and Coordena  o de Aperfei  oamento de Pessoal de N  vel Superior (CAPES).

Conflict of Interest

None declared.

References

- Aguirre-Garrido, J. F., D. Montiel-Lugo, C. H  rnandez-Rodr  guez, G. Torres-Cortez, V. Mill  n, N. Toro, et al. 2012. Bacterial community structure in the rhizosphere of three cactus species from semi-arid highlands in central Mexico. *Antonie Van Leeuwenhoek* 101:891–904.
- Allison, S. D., and P. M. Vitousek. 2004. Rapid nutrient cycling in leaf litter from invasive plants in Hawai'i. *Oecologia* 141:612–619.
- Altschul, S. F., T. L. Madden, A. A. Sch  ffer, J. Zhang, Z. Zhang, W. Miller, et al. 1997. Gapped BLAST and

- PSI-BLAST: a new generation of protein database search program. *Nucleic Acids Res.* 25:3389–3402.
- Amaike, S., and N. P. Keller. 2011. *Aspergillus flavus*. *Annu. Rev. Phytopathol.* 49:107–133.
- Anderson, I. C., C. D. Campbell, and J. I. Prosser. 2003. Diversity of fungi in organic soils under a moorland – Scots pine (*Pinus sylvestris* L.) gradient. *Environ. Microbiol.* 5:1121–1132.
- Araujo, J. F., A. P. de Castro, M. M. Costa, R. C. Togawa, G. J. Júnior, B. F. Quirino, et al. 2012. Characterization of soil bacterial assemblies in Brazilian savanna-like vegetation reveals acidobacteria dominance. *Microb. Ecol.* 64:760–770.
- Arias, M. E., J. A. Gonzalez-Perez, F. J. Gonzalez-Vila, and A. S. Ball. 2005. Soil health – a new challenge for microbiologists and chemists. *Int. Microbiol.* 8:13–21.
- Austin, A. T., and L. Vivanco. 2006. Plant litter decomposition in a semiarid ecosystem controlled by photodegradation. *Nature* 442:555–558.
- Bachar, A., A. Al-Ashhab, M. I. Soares, M. Y. Sklarz, R. Angel, E. D. Ungar, et al. 2010. Soil microbial abundance and diversity along a low precipitation gradient. *Microb. Ecol.* 60:453–461.
- Bachar, A., M. I. Soares, and O. Gillor. 2012. The effect of resource islands on abundance and diversity of bacteria in arid soils. *Microb. Ecol.* 63:694–700.
- Bais, H. P., T. L. Weir, L. G. Perry, S. Gilroy, and J. M. Vivanco. 2006. The role of root exudates in rhizosphere interactions with plants and other organisms. *Annu. Rev. Plant Biol.* 57:233–266.
- Bartholomew, G. W., and M. Alexander. 1979. Microbial metabolism of carbon monoxide in culture and in soil. *Appl. Environ. Microbiol.* 37:932–937.
- Berg, I. A., D. Kockelkorn, W. Buckel, and G. Fuchs. 2007. A 3-hydroxypropionate/4-hydroxybutyrate autotrophic carbon dioxide assimilation pathway in Archaea. *Science* 318:1782–1786.
- Bernard, L., C. Mougél, P. A. Maron, V. Nowak, J. Lévêque, C. Henault, et al. 2007. Dynamics and identification of soil microbial populations actively assimilating carbon from ¹³C-labelled wheat residue as estimated by DNA- and RNA-SIP techniques. *Environ. Microbiol.* 9:752–764.
- Bowman, J. P., L. I. Sly, P. D. Nichols, and A. C. Hayward. 1993. Revised taxonomy of the methanotrophs: description of *Methylobacter* gen. nov., emendation of *Methylococcus*, validation of *Methylosinus* and *Methylocystis* species, and a proposal that the family Methylococcaceae includes only the group I methanotrophs. *Int. J. Syst. Bacteriol.* 43: 735–753.
- Brandt, L. A., J. Y. King, and D. G. Milchunas. 2007. Effects of ultraviolet radiation on litter decomposition depend on precipitation and litter chemistry in a short grass steppe ecosystem. *Glob. Change Biol.* 13:2193–2205.
- Brandt, L. A., C. Bohnet, and J. Y. King. 2009. Photochemically induced carbon dioxide production as a mechanism for carbon loss from plant litter in arid ecosystems. *J. Geophys. Res.* 114:G02004.
- Brankatschk, R., T. Fischer, M. Veste, and J. Zeyer. 2013. Succession of N cycling processes in biological soil crusts on a Central European inland dune. *FEMS Microbiol. Ecol.* 83:149–160.
- Brockett, B. F. T., C. E. Prescott, and S. J. Grayston. 2012. Soil moisture is the major factor influencing microbial community structure and enzyme activities across seven biogeoclimatic zones in western Canada. *Soil Biol. Biochem.* 44:9–20.
- Bruce, T., I. B. Martinez, O. Maia Neto, A. C. Vicente, R. H. Kruger, and F. L. Thompson. 2010. Bacterial community diversity in the Brazilian Atlantic forest soils. *Microb. Ecol.* 60:840–849.
- Buée, M., M. Reich, C. Murat, E. Morin, R. H. Nilsson, B. Uroz, et al. 2009. 454 Pyrosequencing analyses of forest soils reveal an unexpectedly high fungal diversity. *New Phytol.* 184:449–456.
- Chanal, A., V. Chapon, K. Benzerara, M. Barakat, R. Christen, W. Achouak, et al. 2006. The desert of Tataouine: an extreme environment that hosts a wide diversity of microorganisms and radiotolerant bacteria. *Environ. Microbiol.* 8:514–525.
- Chau, J. F., A. C. Bagtzoglou, and M. R. Willig. 2011. The effect of soil texture on richness and diversity of bacterial communities. *Environ. Forensics* 12:333–341.
- Chikere, C. B., G. C. Okpokwasili, and B. O. Chikere. 2011. Monitoring of microbial hydrocarbon remediation in the soil. *3 Biotech.* 1:117–138.
- Chou, H. H., and M. H. Holmes. 2001. DNA sequence quality trimming and vector removal. *Bioinformatics* 17:1093–1104.
- Clemente, J. C., J. Jansson, and G. Valiente. 2010. Accurate taxonomic assignment of short pyrosequencing reads. *Pac. Symp. Biocomput.* 3–9.
- Cleveland, C. C., W. R. Wieder, S. C. Reed, and A. R. Townsend. 2010. Experimental drought in a tropical rain forest increases soil carbon dioxide losses to the atmosphere. *Ecology* 91:2313–2323.
- Connon, S.A., E. D. Lester, H. S. Shafaat, D. C. Obenhuber, and A. Ponce. 2007. Bacterial diversity in hyperarid Atacama Desert soils. *J. Geophys. Res.* 112:G04S17.
- Curtis, T. P., and W. T. Sloan. 2004. Prokaryotic diversity and its limits: microbial community structure in nature and implications for microbial ecology. *Curr. Opin. Microbiol.* 7:221–226.
- Day, T. A., E. T. Zhang, and C. T. Ruhland. 2007. Exposure to solar UV-B radiation accelerates mass and lignin loss of *Larrea tridentata* litter in the Sonoran Desert. *Plant Ecol.* 193:185–194.
- DeAngelis, K. M., M. Allgaier, Y. Chavarria, J. L. Fortney, P. Hugenholtz, B. Simmons, et al. 2011. Characterization of trapped lignin-degrading microbes in tropical forest soil. *PLoS ONE* 6:e19306.

- Di, H. J., K. C. Cameron, J. P. Shen, C. S. Winefield, M. O'Callaghan, S. Bowatte, et al. 2009. Nitrification driven by bacteria and not archaea in nitrogen-rich grasslands soil. *Nat. Geosci.* 2:621–624.
- Ettema, T. J. G., and S. G. E. Andersson. 2008. Comment on “A 3-hydroxypropionate/4-hydroxybutyrate autotrophic carbon dioxide assimilation pathway in Archaea”. *Science* 321:342.
- Faoro, H., A. C. Alves, E. M. Souza, L. U. Rigo, L. M. Cruz, S. M. Al-Janabi, et al. 2010. Influence of soil characteristics on the diversity of bacteria in the Southern Brazilian Atlantic Forest. *Appl. Environ. Microbiol.* 76:4744–4749.
- Fierer, N., M. A. Bradford, and R. B. Jackson. 2007. Toward an ecological classification of soil bacteria. *Ecology* 88:1354–1364.
- Freire, M. S. B. 1990. Levantamento florístico do Parque Estadual das Dunas do Natal. *Acta. Bot. Bras.* 4:41–59.
- Gallo, M. E., A. Porras-Alfaro, K. J. Odenbach, and R. L. Sinsabaugh. 2009. Photoacceleration of plant litter decomposition in an arid environment. *Soil Biol. Biochem.* 41:1433–1441.
- Giongo, V., S. R. S. Galvão, A. M. S. Mendes, C. A. T. Gava, and T. J. F. Cunha. 2011. Soil organic carbon in the Brazilian semi-arid tropics. *Dyn. Soil Dyn. Plant.* 5:12–20.
- Goldfarb, K. C., U. Karaoz, C. A. Hanson, C. A. Santee, M. A. Bradford, K. K. Treseder, et al. 2011. Differential growth responses of soil bacterial taxa to carbon substrates of varying chemical recalcitrance. *Front Microbiol.* 2:94.
- Gomez-Alvarez, V., T. K. Teal, and T. M. Schmidt. 2009. Systematic artifacts in metagenomes from complex microbial communities. *ISME J.* 3:1314–1317.
- Griffiths, R. I., B. C. Thomson, P. James, T. Bell, M. Bailey, and A. S. Whiteley. 2011. The bacterial biogeography of British soils. *Environ. Microbiol.* 13:1642–1654.
- Guo, F., F. Ju, L. Cai, and T. Zhang. 2013. Taxonomic precision of different hypervariable regions of 16S rRNA gene and annotation methods for functional bacterial groups in biological wastewater treatment. *PLoS ONE* 8: e76185.
- He, J., Z. Xu, and J. Hughes. 2006. Molecular bacterial diversity of a forest soil under residue management regimes in subtropical Australia. *FEMS Microbiol. Ecol.* 55:38–47.
- Horn, B. W. 2003. Ecology and population biology of aflatoxigenic fungi in soil. *J. Toxicol.* 22:351–379.
- Hunt, J., L. Boddy, P. F. Randerson, and H. J. Rogers. 2004. An Evaluation of 18S rDNA Approaches for the Study of Fungal Diversity in Grassland Soils. *Microb. Ecol.* 47:385–395.
- Huson, D. H., S. Mitra, H. J. Ruscheweyh, N. Weber, and S. C. Schuster. 2011. Integrative analysis of environmental sequences using MEGAN4. *Genome Res.* 21:1552–1560.
- Inman, R. E., R. B. Ingersoll, and E. A. Levy. 1971. Soil: a natural sink for carbon monoxide. *Science* 172:1229–1231.
- Ivan, A. B., D. Kockelkorn, W. Buckel, and G. Fuchs. 2008. Response to comment on “a 3-hydroxypropionate/4-hydroxybutyrate autotrophic carbon dioxide assimilation pathway in archaea”. *Science* 321:342.
- Ivanova, N., J. Sikorski, M. Jando, C. Munk, A. Lapidus, T. Glavina Del Rio, et al. 2010. Complete genome sequence of *Geodermatophilus obscurus* type strain (G-20). *Stand. Genomic Sci.* 30:158–167.
- Iwai, S., A. Yamazoe, R. Takahashi, F. Kurisu, and O. Yagi. 2005. Degradation of mono-chlorinated dibenzo-p-dioxins by *Janibacter* sp. strain YA isolated from river sediment. *Curr. Microbiol.* 51:353–358.
- Jeffries, T. C., J. R. Seymour, J. A. Gilbert, E. A. Dinsdale, K. Newton, S. S. Leterme, et al. 2011. Substrate type determines metagenomic profiles from diverse chemical habitats. *PLoS ONE* 6:e25173.
- Jesus, E. C., T. L. Marsh, J. M. Tiedje, and F. M. S. Moreira. 2009. Changes in land use alter the structure of bacterial communities in Western Amazon soils. *ISME J.* 3:1004–1011.
- Jones, R. T., M. S. Robeson, C. L. Lauber, M. Hamady, R. Knight, and N. Fierer. 2009. A comprehensive survey of soil acidobacterial diversity using pyrosequencing and clone library analyses. *ISME J.* 3:442–453.
- Jothimani, P., G. Kalaichelvan, A. Bhaskaran, D. A. Selvaseelan, and K. Ramasamy. 2003. Anaerobic biodegradation of aromatic compounds. *Indian J. Exp. Biol.* 41:1046–1067.
- Kanehisa, M., S. Goto, Y. Sato, M. Furumichi, and M. Tanabe. 2012. KEGG for integration and interpretation of large-scale molecular data sets. *Nucleic Acids Res.* 40:D109–D114.
- Kim, M., D. Singh, A. Lai-Hoe, R. Go, R. Abdul Rahim, A. N. Ainnuddin, et al. 2012. Distinctive phyllosphere bacterial communities in tropical trees. *Microb. Ecol.* 63:674–681.
- King, G. M. 2006. Nitrate-dependent anaerobic carbon monoxide oxidation by aerobic CO-oxidizing bacteria. *FEMS Microbiol. Ecol.* 56:1–7.
- King, G. M., and C. F. Weber. 2007. Distribution, diversity and ecology of aerobic CO-oxidizing bacteria. *Nat. Rev. Microbiol.* 5:107–118.
- Klich, M. A. 2002. Biogeography of *Aspergillus* species in soil and litter. *Mycologia* 94:21–27.
- Köberl, M., H. Müller, E. M. Ramadan, and G. Berg. 2011. Desert farming benefits from microbial potential in arid soils and promotes diversity and plant health. *PLoS ONE* 6: e24452.
- Kolb, S. 2009. Aerobic methanol-oxidizing bacteria in soil. *FEMS Microbiol. Lett.* 300:1–10.
- Koops, H. P., and A. Pommerening-Röser. 2001. Distribution and ecophysiology of the nitrifying bacteria emphasizing cultured species. *FEMS Microbiol. Ecol.* 37:1–9.
- Kuramae, E. E., E. Yergeau, L. C. Wong, A. S. Pijl, J. A. van Veen, and G. A. Kowalchuk. 2012. Soil characteristics more strongly influence soil bacterial communities than land-use type. *FEMS Microbiol. Ecol.* 79:12–24.

- Lauber, C. L., M. Hamady, R. Knight, and N. Fierer. 2009. Pyrosequencing-based assessment of soil pH as a predictor of soil bacterial community structure at the continental scale. *Appl. Environ. Microbiol.* 75:5111–5120.
- Leininger, S., T. Urich, M. Schloter, L. Schwark, J. Qi, G. W. Nicol, et al. 2006. Archaea predominate among ammonia-oxidizing prokaryotes in soils. *Nature* 442:806–809.
- Liles, M. R., O. Turkmen, B. F. Manske, M. Zhang, J. M. Rouillard, I. F. George, et al. 2010. A phylogenetic microarray targeting 16S rRNA genes from the bacterial division *Acidobacteria* reveals a lineage-specific distribution in a soil clay fraction. *Soil Biol. Biochem.* 42:739–747.
- Lin, Y. T., Y. J. Huang, S. L. Tang, W. B. Whitman, D. C. Coleman, and C. Y. Chiu. 2010. Bacterial community diversity in undisturbed perhumid montane forest soils in Taiwan. *Microb. Ecol.* 59:369–378.
- Margulies, M., M. Egholm, W. E. Altman, S. Attiya, J. S. Bader, L. A. Bemben, et al. 2005. Genome sequencing in open microfabricated high density picoliter reactors. *Nature* 437:376–380.
- Menezes, R. S., E. V. Sampaio, V. Giongo, and A. M. Pérez-Marin. 2012. Biogeochemical cycling in terrestrial ecosystems of the Caatinga Biome. *Braz. J. Bio.* 72:643–653.
- Meyer, F., D. Paarmann, M. D'Souza, R. Olson, E. M. Glass, M. Kubal, et al. 2008. The metagenomics RAST server – a public resource for the automatic phylogenetic and functional analysis of metagenomes. *BMC Bioinformatics* 9:386.
- Mitra, S., B. Klar, and D. H. Huson. 2009. Visual and statistical comparison of metagenomes. *Bioinform* 25:1849–1855.
- Mocali, S., and A. Benedetti. 2010. Exploring research frontiers in microbiology: the challenge of metagenomics in soil microbiology. *Res. Microbiol.* 161:497–505.
- Montzka, S. A., E. J. Dlugokencky, and J. H. Butler. 2011. Non-CO₂ greenhouse gases and climate change. *Nature* 476:43–50.
- Morales, S. E., T. Cosart, and W. E. Holben. 2010. Bacterial gene abundances as indicators of greenhouse gas emission in soils. *ISME J.* 4:799–808.
- Myers, N., R. A. Mittermeier, C. G. Mittermeier, G. A. da Fonseca, and J. Kent. 2000. Biodiversity hotspots for conservation priorities. *Nature* 403:853–858.
- Nie, M., H. Meng, K. Li, J. R. Wan, Z. X. Quan, C. M. Fang, et al. 2012. Comparison of bacterial and fungal communities between natural and planted pine forests in subtropical China. *Curr. Microbiol.* 64:34–42.
- Overbeek, R., T. Begley, R. M. Butler, J. V. Choudhuri, H. Y. Chuang, M. Cohoon, et al. 2005. The subsystems approach to genome annotation and its use in the project to annotate 1000 genomes. *Nucleic Acids Res.* 33:5691–5702.
- Parks, D. H., and R. G. Beiko. 2010. Identifying biologically relevant differences between metagenomic communities. *Bioinformatics* 26:715–721.
- Paster, B. J., and E. Canale-Parola. 1985. *Treponema saccharophilum* sp. nov., a large pectinolytic spirochete from the bovine rumen. *Appl. Environ. Microbiol.* 50:212–219.
- Peng, C., J. Liu, Q. Dang, M. J. Apps, and H. Jiang. 2002. TRIPLEX: a generic hybrid model for predicting forest growth and carbon and nitrogen dynamics. *Ecol. Model.* 153:109–130.
- Rennenberg, H., M. Dannenmann, A. Gessler, J. Kreuzwieser, J. Simon, and H. Papen. 2009. Nitrogen balance in forest soils: nutritional limitation of plants under climate change stresses. *Plant Biol.* 1:4–23.
- Richardson, A. E., and R. J. Simpson. 2011. Soil microorganisms mediating phosphorus availability update on microbial phosphorus. *Plant Physiol.* 156:989–996.
- Riley, M., J. T. Staley, A. Danchin, T. Z. Wang, T. S. Brettin, L. J. Hauser, et al. 2008. Genomics of an extreme psychrophile. *Psychromonas ingrahamii*. *BMC Genomics* 9:210.
- Russo, S. E., R. Legge, K. Weber, E. Brodie, K. C. Goldfarb, A. Benson, et al. 2012. Bacterial community structure of contrasting soils underlying Bornean rain forests: inferences from microarray and next-generation sequencing methods. *Soil Biol. Biochem.* 55:48–59.
- Saggari, S., N. Jha, J. Deslippe, N. S. Bolan, J. Luo, D. L. Giltrap, et al. 2013. Denitrification and N(2)O:N(2) production in temperate grasslands: processes, measurements, modelling and mitigating negative impacts. *Sci. Total Environ.* 465:173–195.
- Sangeetha, M., R. Jayakumar, and C. Bharathi. 2009. Nitrous oxide emission from soils – a review. *Agric. Rev.* 30:94–107.
- Santos, A. M. M., and B. A. Santos. 2008. Are the vegetation structure and composition of the shrubby Caatinga free from edge influence? *Acta. Bot. Bras.* 22:1077–1084.
- Santos, M. V. F., M. A. Lira, J. C. B. Dubeux Junior, A. Guim, A. C. L. Mello, and M. V. Cunha. 2010. Potential of Caatinga forage plants in ruminant feeding. *R. Bras. Zootec.* 39:204–215.
- Santos, R. M., A. T. Oliveira-Filho, P. V. Eisenlohr, L. P. Queiroz, D. B. Cardoso, and M. J. Rodal. 2012. Identity and relationships of the Arboreal Caatinga among other floristic units of seasonally dry tropical forests (SDTFs) of north-eastern and Central Brazil. *Ecol. Evol.* 2:409–428.
- Sayer, E. J. 2006. Using experimental manipulation to assess the roles of leaf litter in the functioning of forest ecosystems. *Biol. Rev. Camb. Philos. Soc.* 81:1–31.
- Schade, G. W., R. M. Hormann, and P. J. Crutzen. 1999. CO emissions from degrading plant matter. *Tellus B Chem. Phys. Meteorol.* 51:899–908.
- Schneegurt, M. A., S. Y. Dore, and C. F. Jr Kulpa. 2003. Direct extraction of DNA from soils for studies in microbial ecology. *Curr. Issues Mol. Biol.* 5:1–8.
- Segata, N., D. Boernigen, T. L. Tickle, X. C. Morgan, W. S. Garrett, and C. Huttenhower. 2013. Computational

- meta'omics for microbial community studies. *Mol. Syst. Biol.* 9:666.
- Slováková, L., D. Dusková, and M. Marounek. 2002. Fermentation of pectin and glucose, and activity of pectin-degrading enzymes in the rabbit caecal bacterium *Bifidobacterium pseudolongum*. *Lett. Appl. Microbiol.* 35:126–130.
- Sousa Neto, E., J. B. Carmo, M. Keller, S. C. Martins, L. F. Alves, and S. A. Vieira. 2011. Soil-atmosphere exchange of nitrous oxide, methane and carbon dioxide in a gradient of elevation in the coastal Brazilian Atlantic forest. *Biogeosciences* 8:733–742.
- Starkenburg, S. R., F. W. Larimer, L. Y. Stein, M. G. Klotz, P. S. Chain, L. A. Sayavedra-Soto, et al. 2008. Complete genome sequence of *Nitrobacter hamburgensis* X14 and comparative genomic analysis of species within the genus *Nitrobacter*. *Appl. Environ. Microbiol.* 74:2852–2863.
- Tepljakov, A., G. Obmolova, E. Sarikaya, S. Pullalarevu, W. Krajewski, A. Galkin, et al. 2004. Crystal structure of the YgfZ protein from *Escherichia coli* suggests a folate-dependent regulatory role in one-carbon metabolism. *J. Bacteriol.* 186:7134–7140.
- Thomson, B. C., N. Ostle, N. McNamara, M. J. Bailey, A. S. Whiteley, and R. I. Griffiths. 2010. Vegetation affects the relative abundances of dominant soil bacterial taxa and soil respiration rates in an upland grassland soil. *Microb. Ecol.* 59:335–343.
- Torres, M. J., E. Bueno, S. Mesa, E. J. Bedmar, and M. J. Delgado. 2011. Emerging complexity in the denitrification regulatory network of *Bradyrhizobium japonicum*. *Biochem. Soc. Trans.* 39:284–288.
- Torsvik, V., L. Ovreas, and T. F. Thingstad. 2002. Prokaryotic diversity—magnitude, dynamics, and controlling factors. *Science* 296:1064–1066.
- Tringe, S. G., C. von Mering, A. Kobayashi, A. A. Salamov, K. Chen, H. W. Chang, et al. 2005. Comparative metagenomics of microbial communities. *Science* 308:554–557.
- Vieira, S. A., L. F. Alves, P. J. Duarte-Neto, S. C. Martins, L. G. Veiga, M. A. Scaranello, et al. 2011. Stocks of carbon and nitrogen and partitioning between above- and belowground pools in the Brazilian coastal Atlantic Forest elevation range. *Ecol. Evol.* 1:421–434.
- Ward, B. B. 2008. Nitrification in marine systems. Pp. 199–261 *Nitrogen in the marine environment*, 2nd ed. Academic Press, San Diego.
- Ward, N. L., J. F. Challacombe, P. H. Janssen, B. Henrissat, P. M. Coutinho, M. Wu, et al. 2009. Three genomes from the phylum acidobacteria provide insight into the lifestyles of these microorganisms in soils. *Appl. Environ. Microbiol.* 75:2046–2056.
- Whitman, W. B., D. C. Coleman, and W. J. Wiebe. 1998. Prokaryotes: the unseen majority. *Proc. Natl Acad. Sci. USA* 95:6578–6583.
- Yuan, H., T. Ge, C. Chen, A. G. O'Donnell, and J. Wu. 2012. Significant role for microbial autotrophy in the sequestration of soil carbon. *Appl. Environ. Microbiol.* 78:2328–2336.

Supporting Information

Additional Supporting Information may be found in the online version of this article:

Figure S1. Localization of the sampling points: João Câmara city and Parque das Dunas (Natal city), which are situated in Rio Grande do Norte state (Brazil).

Figure S2. Analysis of the composition of the PD (black) and JC (gray) communities, showing the Bacteria classes diversity of metagenomic sequences. The data were computed by MEGAN based on a BLASTX using an e-value cutoff of $1e-5$. Light gray highlighting on the left side of a node indicates that the up-test of directed homogeneity test showed a significant difference. The thickness of the highlighting is logarithmically proportional to the significance. The size of the bars is scaled logarithmically to represent the number of reads assigned to each taxon.

Figure S3. Taxonomical distribution relative to the SEED subsystems that showed differences between PD (black bars) and JC (gray bars). Only genus with the highest representation and significant differences are shown.

Figure S4. Comparison of the functional data between: PD versus Holm-Oak rhizosphere (A); PD versus rice rhizosphere (B); JC versus Holm-Oak rhizosphere (C); and JC versus rice rhizosphere; (D) considering subsystem level 1. The relative proportion difference in functional distribution of PD and JC samples was considered significant when $q < 0.05$.

Figure S5. Comparison computed using two groups analysis at subsystem level 1 for PD and JC versus deserts. Only significant differences are shown (by Welch's *t*-test, the Welch's inverted test for confidence interval method and Benjamin-Hochberg FDR for correction were applied).

Table S1. General characteristics of the PD and JC metagenomes.