RESEARCH ARTICLE

Microbial Community Profile and Water Quality in a Protected Area of the Caatinga Biome

Fabyano Alvares Cardoso Lopes¹, Elisa Caldeira Pires Catão¹, Renata Henrique Santana², Anderson de Souza Cabral³, Rodolfo Paranhos³, Thiago Pessanha Rangel⁴, Carlos Eduardo de Rezende⁴, Robert A. Edwards⁵,⁶, Cristiane C. Thompson³, Fabiano L. Thompson³☯*, Ricardo Henrique Kruger¹☯*

¹ Cellular Biology Department, Universidade de Brasília (UnB), Brasília, DF, Brazil, ² Genomic Sciences and Biotechnology, Universidade Católica de Brasília (UCB), Brasília DF, Brazil, ³ Institute of Biology, Universidade Federal do Rio de Janeiro (UFRJ), Rio de Janeiro, Brazil, ⁴ Laboratory of Environmental Sciences, Universidade Estadual do Norte Fluminense (UENF), Campos dos Goytacazes, Brazil, ⁵ Computational Science Research Center, San Diego State University (SDSU), San Diego, California, United States of America, ⁶ Department of Computer Science, San Diego State University (SDSU), San Diego, California, United States of America

☯ These authors contributed equally to this work.
* fabianothompson1@gmail.com (FLT); kruger@unb.br (RHK)

Abstract

The Caatinga is a semi-arid biome in northeast Brazil. The Paraguaçu River is located in the Caatinga biome, and part of its course is protected by the National Park of Chapada Diamantina (PNCD). In this study we evaluated the effect of PNCD protection on the water quality and microbial community diversity of this river by analyzing water samples obtained from points located inside and outside the PNCD in both wet and dry seasons. Results of water quality analysis showed higher levels of silicate, ammonia, particulate organic carbon, and nitrite in samples from the unprotected area compared with those from protected areas. Pyrosequencing of the 16S rRNA genes revealed that Burkholderiales was abundant in samples from all three sites during both seasons and was represented primarily by the genus Polynucleobacter and members of the Comamonadaceae family (e.g., genus Limnohabitans). During the dry season, the unprotected area showed a higher abundance of Flavobacterium sp. and Arthrobacter sp., which are frequently associated with the presence and/or degradation of arsenic and pesticide compounds. In addition, genes that appear to be related to agricultural impacts on the environment, as well as those involved in arsenic and cadmium resistance, copper homeostasis, and propanediol utilization, were detected in the unprotected areas by metagenomic sequencing. Although PNCD protection improves water quality, agricultural activities around the park may affect water quality within the park and may account for the presence of bacteria capable of pesticide degradation and assimilation, evidencing possible anthropogenic impacts on the Caatinga.
Introduction
The Caatinga is a semi-arid biome located in the northeast of Brazil (3–17°S to 35–45°W). It occupies almost 900,000 km² of the Brazilian territory and is characterized by its vegetation during the dry season, when the leaves fall and white tree trunks and shrubs remain in the landscape [1, 2]. This unique and important biome is strongly affected by anthropogenic processes. It is estimated that about 50% of the Caatinga biome has been modified by activities related to agriculture, livestock, or coal extraction [3–6]. The dry season in the Caatinga is characterized by water shortage [5], with an annual rainfall of 300–500 mm in the semi-arid zone and 1,500 mm in the mountainous area of Chapada Diamantina [3]. Accordingly, the Paraguaçu River, which is a typical Caatinga river, shows high seasonal volume fluctuations. This river supplies water for agricultural and mining activities, as well as several cities. Despite the great importance of this river, it is subject to pesticide dumping, siltation, and heavy metal pollution [7], but the precise effects of agriculture and mining on the water quality and microbial diversity of this region remain unclear. To prevent environmental degradation, the National Park of Chapada Diamantina (PNCD) was created in 1985. However, this park, which protects part of the Paraguaçu River course, is surrounded by agricultural land, some of which is in direct contact with the Paraguaçu River.

Previous studies have described the isolation and characterization of microbes from the Caatinga [8–12]. Recently, Pacchioni et al. conducted the first metagenomic study of one soil sample from the Caatinga and showed that Actinobacteria and Alphaproteobacteria were the most abundant groups [13]. Pacchioni et al. also described the presence of genes related to stress resistance and the metabolism of DNA, nitrogen, and amino acids. This study suggested that the microbial profile of the Caatinga differed from those of other Brazilian biomes such as the Amazon, savannah (also called Cerrado), and Atlantic forest, but the limited number of samples analyzed hampered a clear distinction. In addition, it is not clear from previous studies how the microbial communities are structured inside and outside protected areas (such as PNCD).

In this study, we evaluated the effect of PNCD protection on the water quality and microbial diversity of the Paraguaçu River. Specifically, we performed taxonomic and metagenomic analyses of bacterial communities from sites inside and outside the PNCD in both wet and dry seasons. We also analyzed water quality by measuring the levels of heavy metals, dissolved organic carbon (DOC), particulate organic carbon (POC), and inorganic nutrients.

Materials and Methods

Ethics statement
This study was approved by Instituto Chico Mendes de Conservação da Biodiversidade (ICMBio) in accordance with the Brazilian law (Permit Number: SISBIO31652-1).

Study area and sample collection
Water samples were collected from three points along the Paraguaçu River (Fig 1). The first sampling point (P1) is located outside the PNCD (13°26'9.11"S 41°20'17.56"W) near Ibicoara (Bahia, Brazil) and was therefore unprotected. The two protected sampling points, P2 and P3, are located just inside the PNCD (13°0'2.60"S 41°23'22.57"W and 12°50'25.91"S 41°19'26.52"W, respectively). Two replicate samples were obtained at each sampling point in the wet season (November 2012) and dry season (February 2013) for a total of 12 samples. For each sampling point in both seasons, we collected approximately 20 L unfiltered water by pump from the water column at a depth of approximately 1 m.
Fig 1. Study area. A) The area under National Park of Chapada Diamantina (PNCD) management is shown in green. The three sampling points on the Paraguaçu River are indicated on the map: P1, unprotected site outside the PNCD; P2 and P3, protected sites within the PNCD. The arrow indicates the direction of water flow. B) Unprotected sampling point P1. C) Protected sampling point P2. D) Protected sampling point P3.

doi:10.1371/journal.pone.0148296.g001
After obtaining water samples, aliquots of each sample (200 ml) were frozen at -20°C for the subsequent measurement of physicochemical parameters. After collecting the water samples, sediments were sampled from the river bottom directly below the water sampling point. The sediment samples were collected using a dredge at a depth of 10 cm inside a perimeter of 2 m and stored at -20°C.

For each freshwater sample, aliquots used to measure microbial abundance were prepared immediately after water sampling. Three 1.5-ml aliquots were dispensed into 2.0-ml cryogenic tubes and fixed with 1% paraformaldehyde and 0.5% glutaraldehyde for the bacterial count [14], 0.5% glutaraldehyde for the viral count [15], and 1% paraformaldehyde for the nanoeukaryotic phytoplankton (NEUK), picoeukaryote (PEUK), and *Synechococcus* spp. analyses [16]. Fixation was performed within 30 minutes after the water samples were collected.

Two liters of water was prefiltered using a 20-μm mesh and then filtered through 0.22-μm Sterivex filters (Millipore), using a peristaltic pump and three Sterivex filters for each sample. The Sterivex filters were stored in SET buffer (20% sucrose, 50 mM EDTA, 0.5 mM Tris-HCl, and pH 8.0) at -80°C until DNA extraction could be carried out.

**Physicochemical and microbial abundance analyses**

Two replicates water samples were analyzed for each physicochemical parameter. Measurements of inorganic nutrients [17] were carried out as follows: ammonia was determined using the indophenol method, nitrite by diazotization, nitrate by Cd-Cu reduction followed by diazotization, total nitrogen by potassium persulfate digestion following nitrate determination, orthophosphate by reaction with ascorbic acid, total phosphorous by acid digestion to phosphate, and silicate by reaction with molybdate. DOC and POC were analyzed as described previously [18].

Each sediment sample was passed through a 2-mm sieve (Millipore) with a stream of water. Sand (coarse, medium and fine), silt, clay, and trace elements were measured in the fraction that passed through the sieve. Quantification of trace elements was carried out by inductively coupled plasma optical emission spectrometry (ICP-OES; Varian Liberty-Series II) using a procedure based on US Environmental Protection Agency Method 3052 [19] and modified by Marques et al. [20]. Approximately 0.5 g of the sieved sediment was used to determine concentrations of the following elements: Al, Ba, Ca, Cd, Cr, Cu, Fe, Mn, Ni, P, Pb, S, Sr, Ti, V, and Zn. Measurements were carried out in triplicate for each sample, and a coefficient of variation between replicates <10% was considered satisfactory.

Microbial and viral abundance in water samples was determined by flow cytometry (FACS-Calibur, BD Biosciences) using nucleic acid affinity fluorophore probes for bacteria [21, 22] and SYBR Green (Life Technologies) for viruses [15]. Picoplankton were detected by the fluorescence of natural photosynthetic pigments [16].

Principal component analysis of physicochemical and microbial parameters was performed using a correlation matrix in Past v.3.01 [23]. Values for physicochemical and microbial parameters from the 12 water samples (two samples taken from three different sites in both seasons) were compared by ANOVA (\(\alpha<0.05\)), followed by the Tukey post hoc test using R statistical software [24].

**Metagenomic DNA extraction**

DNA extraction was performed using lysozyme (1 mg/ml) for 1 h at 37°C, as previously described [25]. Then proteinase K (0.2 mg/ml) and 1% sodium dodecyl sulfate (SDS) were added, and the samples were incubated at 55°C for 60 min with gentle agitation. The lysate was rinsed into a new tube with 1 ml SET buffer. Metagenomic DNA was extracted with one
volume of phenol:chloroform:isoamyl alcohol (25:24:1) and then precipitated with ethanol and sodium acetate (0.3 M final concentration) at -20°C overnight. The DNA was purified using a Power Clean DNA Clean-Up Kit (MO BIO Laboratories) and stored at -20°C.

Polymerase chain reaction, 16S rRNA gene amplicon sequencing, and sequence analysis

Polymerase chain reaction (PCR) amplification of bacterial 16S rRNA genes was performed as described in our previous reports [26]. The hypervariable regions V5 to V9 of the bacterial 16S rRNA gene were amplified using primers 787F (5’-ATTAGATACCCNGGTAG-3’) and 1492R (5’-GNTACCTTGTTACGACTT-3’) [27]. Primers used in this work were designed with the appropriate 454 pyrosequencing adaptor sequences and multiplex identifiers (not shown). Ten 20-μL reactions were carried out for bacteria using 1× buffer, 0.25 mM dNTP mix, 3 mM MgCl₂, 0.175 pmol each primer (forward and reverse), 1.5 U Taq polymerase (Phoneutria, Brazil), 5–10 ng DNA, and deionized ultrapure water. The PCR conditions consisted of an initial denaturing step at 95°C for 3 min; 25 amplification cycles at 95°C for 30 s, 58°C for 30 s, and 72°C for 1.7 min; and a final extension step for 7 min at 72°C. The reactions were pooled and then purified using a QIAquick PCR Purification kit (Qiagen). Pyrosequencing of the amplicons was carried out in two 1/8 picotiter plates by Macrogen Inc. (Seoul, Korea) using a 454 GS-FLX Titanium system (454 Life Sciences; Roche, Basel, Switzerland).

Sequences were analyzed using QIIME 1.7.0 software [28]. Initially the 16S rRNA sequences were demultiplexed, and the reads were renamed according to the sample ID with the `split_libraries.py` script. The sequence quality thresholds used in this step were minimum average quality score of 30, sequence length range of 200–1000 nucleotides, window score of 50 nucleotides, maximum number of ambiguous bases of 6, length of homopolymer run of 6, and maximum number of error in barcodes of 1.5. Primer mismatches were not allowed. The denoising procedure for the output file was performed using the `denoise_wrapper.py` script [29] set to titanium defaults. The script `inflate_denoiser_output.py` reintegrated the denoised data, and the `truncate_reverse_primer.py` script removed the reverse primers and any subsequent sequences. Chimeric sequence identification was performed by `identify_chimeric_seqs.py` script using Chimera Slayer [30], which performs reference-based or de novo (abundance-based) chimera checking. From a total of 165,337 raw sequences obtained by pyrosequencing, 54,423 sequences were removed after the processing steps described above, yielding 110,914 high-quality sequences.

The `pick_de_novo_otus.py` script was used to build an operational taxonomic unit (OTU) table. The parameters used in each step were as follows. (i) OTU picking was performed using uclust with reference sequences from the Greengenes database (May 2013) [31] in the bacterial analyses. Reverse strand matching was enabled, and the sequence similarity threshold was 97%. (ii) The representative set of sequences was chosen based on the most abundant sequences of each OTU. (iii) PyNAST [32] was used to align sequences to the Greengenes core reference alignment, and uclust was used for pairwise alignment; minimum sequence length to include in alignment was 150 nucleotides. (iv) RDP Classifier v.2.2 [33] was used to assign taxonomy based on the Greengenes sequences as references and templates. (v) The phylogenetic tree was built using the FastTree method [34]. (vi) The OTU table showing the relative abundance at each taxonomic level (kingdom, phylum, class, order, family, and genus) was then generated from OTU counts for each sample and their taxonomic assignments.

The script `alpha_rarefaction.py` calculated the diversity indices at a single sequencing depth (i.e. number of sequences per sample) with 2910 sequences using the following metrics: non-parametric Shannon [35], Chao1 [36], Observed Species, and Good's coverage [37]. The values...
of the indices from the 12 samples (taken from three sites in both seasons with two replicates each) were compared by ANOVA (α < 0.05), followed by the Tukey post hoc test using R statistical software [24]. The script beta_rarefaction.py was used to calculate beta diversity [38]. Results of the principal coordinates analysis (PCoA) of relative abundances of bacterial taxa (at the order and genus levels) were plotted in EMPor [39] from community similarities values (unweighted UniFrac distance matrices).

We analyzed taxonomic profiles of the water samples using Statistical Analysis of Metagenomic Profiles (STAMP) software [40]. Samples were compared by ANOVA, followed by the Tukey–Kramer post hoc test (p < 0.05) with the Bonferroni correction for multiple comparisons. Taxa with small effect sizes were removed by filtering (effect size = 8.00), and asymptotic confidence intervals (95%) were calculated. For seasonal analysis (wet versus dry), samples were compared by t-test (p < 0.05), followed by the Bonferroni correction for multiple comparisons. Taxa with small effect sizes were removed by filtering (effect size = 8.00), and asymptotic confidence intervals (95%) were calculated.

**Metagenome sequencing and sequence analysis**

Metagenomic DNA libraries were constructed with the Nextera DNA Sample Preparation Kit (Illumina) and 2 × 250 bp paired-end sequencing was carried out by Centro de Genômica de Alto Desempenho UCB (Brasília, Brazil) using an Illumina MiSeq system according to the manufacturer’s instructions. Sequence analysis was performed with 2.2 x 10^7 sequences (S1 Table) using the MG-RAST server [41] using default sequence quality thresholds [42, 43]. The analysis of metagenomic data was based on unassembled reads. Functional annotation was performed against the SEED database [44], and taxonomic profiles were generated using the M5NR database [45].

We compared the abundance of gene and taxa among the water samples using STAMP software [40]. Samples were compared by ANOVA, followed by the Tukey–Kramer post hoc test (p < 0.05) without correction for the gene profile and with the Bonferroni correction for multiple comparisons for the taxonomic profile. Genes and taxa with small effect sizes were removed by filtering (effect size = 8.00), and asymptotic confidence intervals (95%) were calculated. For seasonal analysis (wet versus dry; protected versus unprotected in both seasons), samples were compared by t-test (p < 0.05) without correction (gene profile) and with the Bonferroni correction for multiple comparisons (taxonomic profile). Genes and taxa with small effect sizes were removed by filtering (effect size = 8.00).

The phylogenetic analysis was performed from the taxonomy profile provided at the species level by MEGAN software 5.10.6 [46].

**Accession numbers**

The sequences assessed in this study are available in NCBI Sequence Read Archive (SRA) under the study Accession number PRJNA292014. Metagenomic data sets are available in the MG-RAST server under Biodiversidade Microbiana do Bioma Caatinga project (ID 7927).

**Results**

**Inorganic and organic compounds and microbial cell count**

We tested water samples obtained from an unprotected site outside the PNCD (P1) and two protected areas within the PNCD (P2 and P3) in both wet and dry seasons. Our results showed that levels of silicate, ammonia, POC, and nitrite were higher in samples from the unprotected area, whereas levels of DOC, total nitrogen, orthophosphate, and total phosphorus were higher.
Table 1. General features of water samples obtained from sites on the Paraguaçú River.

<table>
<thead>
<tr>
<th>Geographic location</th>
<th>P1</th>
<th>P2</th>
<th>P3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Altitude (m above sea level)</td>
<td>1100</td>
<td>974</td>
<td>345</td>
</tr>
<tr>
<td>Season</td>
<td>wet</td>
<td>dry</td>
<td>wet</td>
</tr>
<tr>
<td>Depth (m)</td>
<td>± 0.80</td>
<td>± 0.70</td>
<td>± 1.10</td>
</tr>
<tr>
<td>Width (m)</td>
<td>± 4.23</td>
<td>± 2.94</td>
<td>± 36.9</td>
</tr>
<tr>
<td>pH</td>
<td>2.8</td>
<td>4.0</td>
<td>2.9</td>
</tr>
<tr>
<td>Conductivity (µS/ml)</td>
<td>41</td>
<td>44</td>
<td>54</td>
</tr>
<tr>
<td>Coarse sand (%)</td>
<td>2.6</td>
<td>5.5</td>
<td>45.7</td>
</tr>
<tr>
<td>Medium sand (%)</td>
<td>82.3</td>
<td>63.4</td>
<td>48.7</td>
</tr>
<tr>
<td>Fine sand (%)</td>
<td>14.8</td>
<td>28.9</td>
<td>5.6</td>
</tr>
<tr>
<td>Silt (%)</td>
<td>0.3</td>
<td>2.2</td>
<td>0.0</td>
</tr>
<tr>
<td>Clay (%)</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>Clay (%)</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>Bacterial counts (cells/ml)</td>
<td>1.2E+06 ± 1.1E+05</td>
<td>6.9E+06 ± 1.9E+06</td>
<td>1.0E+06 ± 2.0E+04</td>
</tr>
<tr>
<td>Virus counts (CFU/ml)</td>
<td>0.0 ± 0</td>
<td>0.0 ± 0</td>
<td>0.0 ± 0</td>
</tr>
<tr>
<td>Synechococcus (cells/ml)</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>PEUK cells (cells/ml)</td>
<td>0 ± 0</td>
<td>0 ± 0</td>
<td>1.6E+03 ± 2.4E+02</td>
</tr>
<tr>
<td>NEUK cells (cells/ml)</td>
<td>0 ± 0</td>
<td>0 ± 0</td>
<td>0 ± 0</td>
</tr>
<tr>
<td>Total nitrogen (mg/l)</td>
<td>0.76</td>
<td>0.15</td>
<td>3.00</td>
</tr>
<tr>
<td>Total phosphorus (mg/l)</td>
<td>0.10</td>
<td>0.11</td>
<td>1.71</td>
</tr>
<tr>
<td>Orthophosphate (µM)</td>
<td>0.16</td>
<td>0.01</td>
<td>0.27</td>
</tr>
<tr>
<td>Silicate (µM)</td>
<td>0.29</td>
<td>0.45</td>
<td>0.66</td>
</tr>
<tr>
<td>Ammonia (µM)</td>
<td>1.53</td>
<td>1.53</td>
<td>1.53</td>
</tr>
<tr>
<td>Nitrite (µM)</td>
<td>0.97</td>
<td>0.97</td>
<td>0.97</td>
</tr>
<tr>
<td>Nitrate (µM)</td>
<td>1.25</td>
<td>1.25</td>
<td>1.25</td>
</tr>
<tr>
<td>Total nitrogen (µM)</td>
<td>32.15</td>
<td>32.15</td>
<td>32.15</td>
</tr>
<tr>
<td>N/P ratio</td>
<td>1.72</td>
<td>1.72</td>
<td>1.72</td>
</tr>
</tbody>
</table>

Abbreviations: CFU, colony-forming units; DOC, dissolved organic carbon; NEUK, nanoeukaryotes; N/P, nitrogen/phosphorus; PEUK, picoeukaryotes; POC, particulate organic carbon.

doi:10.1371/journal.pone.0148296.t001

in samples from the protected areas (Table 1; S1–S3 Figs), allowing the segregation of the samples into two groups (Fig 2). Samples obtained in the unprotected area were also clustered according to season, with the vectors for POC and ammonia grouping samples obtained during the wet season, and vectors for silicate, nitrite, and nitrate grouping samples obtained during the dry season (Fig 2). A comparison of samples obtained from within the PNCD shows that DOC and total nitrogen levels were higher in samples from P2, but total phosphorus and orthophosphate levels were higher in samples from P3 (Fig 2). Table 2 summarizes the water analysis results of our study and those of several other metagenomic studies reporting the physicochemical properties of freshwater samples (e.g. rivers, lakes, reservoirs).
We also found that sand and silt percentages and bacterial and eukaryotic cell counts differed among sampling points (Table 1). Although none of the samples contained clay, P1 had a lower percentage of coarse sand than the other sites and was the only site with silt during both seasons. Picoeukaryotes were detected only in the protected areas (in both seasons), whereas *Synechococcus* was detected only in P2 in the dry season (Table 1). Total virus counts were higher in samples obtained during the dry season than in those obtained during the wet season (S1B Fig). In addition, bacterial cell counts were higher in samples obtained from P1 than in...
Table 2. Results of freshwater metagenomics studies.

<table>
<thead>
<tr>
<th>Study classification</th>
<th>Freshwater</th>
<th>Name</th>
<th>Sample</th>
<th>Observation</th>
<th>Bacterial counts (cells/ml)</th>
<th>Depth (m)</th>
<th>pH</th>
<th>DO (mg/l)</th>
<th>Conductivity (mS/cm)</th>
<th>Salinity (PSU)</th>
<th>DOC (mg/l)</th>
<th>POC (mg/l)</th>
<th>Orthophosphate (μM)</th>
<th>Total phosphorus (μM)</th>
<th>Silicate (μM)</th>
<th>Ammonia (μM)</th>
<th>Nitrite (μM)</th>
<th>Nitrate (μM)</th>
<th>Total nitrogen (μM)</th>
<th>N/P ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>Present study</td>
<td>River</td>
<td>Paraguaçu</td>
<td>P1</td>
<td>Wet season</td>
<td>21.49 ± 0.80</td>
<td>2.80</td>
<td>0.04</td>
<td>0.9</td>
<td>7.5</td>
<td>0.2</td>
<td>0.4</td>
<td>16</td>
<td>3.2</td>
<td>0.3</td>
<td>3.3</td>
<td>28.2</td>
<td>79.1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>P2</td>
<td>Wet season</td>
<td>10.10</td>
<td>2.90</td>
<td>0.05</td>
<td>2</td>
<td>2.9</td>
<td>0.1</td>
<td>0.1</td>
<td>0.7</td>
<td>1.1</td>
<td>0.3</td>
<td>0.4</td>
<td>43.8</td>
<td>338.9</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>P3</td>
<td>Wet season</td>
<td>8.73 ± 3.00</td>
<td>3.20</td>
<td>0.03</td>
<td>1.5</td>
<td>1</td>
<td>0.4</td>
<td>0.7</td>
<td>1.4</td>
<td>0.8</td>
<td>0.3</td>
<td>6.5</td>
<td>42.1</td>
<td>60.2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>P1</td>
<td>Dry season</td>
<td>11.88 ± 0.70</td>
<td>4.00</td>
<td>0.04</td>
<td>1.2</td>
<td>2.4</td>
<td>0.3</td>
<td>0.3</td>
<td>11.6</td>
<td>2.6</td>
<td>0.3</td>
<td>5.8</td>
<td>21.2</td>
<td>75.4</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>P2</td>
<td>Dry season</td>
<td>9.86 ± 0.45</td>
<td>3.80</td>
<td>0.06</td>
<td>2.9</td>
<td>0.7</td>
<td>0.2</td>
<td>0.2</td>
<td>0.3</td>
<td>1.5</td>
<td>0.3</td>
<td>0.7</td>
<td>46.6</td>
<td>189.6</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>P3</td>
<td>Dry season</td>
<td>7.13 ± 0.30</td>
<td>4.10</td>
<td>0.04</td>
<td>1.9</td>
<td>0.8</td>
<td>0.2</td>
<td>0.3</td>
<td>0.6</td>
<td>1.6</td>
<td>0.3</td>
<td>8</td>
<td>44.2</td>
<td>130.3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Liu et al. (2015)</td>
<td>River</td>
<td>Pearl Estuary</td>
<td>P01</td>
<td>Surface water</td>
<td>7</td>
<td>0.1</td>
<td>2</td>
<td>21.4</td>
<td>24.6</td>
<td>102.5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>P03</td>
<td>Surface water</td>
<td>1</td>
<td>3.47</td>
<td>0</td>
<td>2</td>
<td>2</td>
<td>21.4</td>
<td>24.6</td>
<td>102.5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>P07</td>
<td>Surface water</td>
<td>1</td>
<td>0.72</td>
<td>0</td>
<td>2</td>
<td>2</td>
<td>21.4</td>
<td>24.6</td>
<td>102.5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>P01</td>
<td>Bottom water</td>
<td>7</td>
<td>0.51</td>
<td>0</td>
<td>4.7</td>
<td>0</td>
<td>2</td>
<td>2</td>
<td>21.4</td>
<td>24.6</td>
<td>102.5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>P03</td>
<td>Bottom water</td>
<td>3</td>
<td>3.14</td>
<td>0</td>
<td>2</td>
<td>2</td>
<td>21.4</td>
<td>24.6</td>
<td>102.5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>P07</td>
<td>Bottom water</td>
<td>19</td>
<td>0.62</td>
<td>0.1</td>
<td>2</td>
<td>2</td>
<td>21.4</td>
<td>24.6</td>
<td>102.5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yan et al. (2015)</td>
<td>River</td>
<td>Three Gorges Reservoir</td>
<td>XXR_E</td>
<td>Lacustine system, Xiangxi River estuary</td>
<td>8.24</td>
<td>7.96</td>
<td>374.9</td>
<td>2.1</td>
<td>0.9</td>
<td>0.7</td>
<td>29.4</td>
<td>7.6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>XXR_M</td>
<td>Lacustine system, Xiangxi River midstream</td>
<td>8.83</td>
<td>12.26</td>
<td>332.3</td>
<td>1.9</td>
<td>1</td>
<td>0.8</td>
<td>15.3</td>
<td>5.3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>WJB</td>
<td>Lacustine system, Wuja Bay</td>
<td>9.28</td>
<td>19.43</td>
<td>266.8</td>
<td>0.1</td>
<td>0.6</td>
<td>0.2</td>
<td>4.4</td>
<td>3.8</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>XXR_U</td>
<td>Riverine system, Xiangxi River upstream</td>
<td>9.52</td>
<td>16.42</td>
<td>282.5</td>
<td>4.1</td>
<td>1.9</td>
<td>0.2</td>
<td>1.5</td>
<td>3.1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>BSR_E</td>
<td>Riverine system, Baisha River estuary</td>
<td>9.25</td>
<td>15.01</td>
<td>266.1</td>
<td>6.8</td>
<td>3.1</td>
<td>0.2</td>
<td>1.3</td>
<td>3.6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>SDR_E</td>
<td>Riverine system, Shendu River estuary</td>
<td>9.24</td>
<td>16.43</td>
<td>271.4</td>
<td>4.6</td>
<td>2.3</td>
<td>0.3</td>
<td>0.4</td>
<td>4.3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tseng et al. (2013)</td>
<td>Reservoir</td>
<td>Feitsui Reservoir (FTR)</td>
<td>M1</td>
<td>Located in North Taiwan</td>
<td>32.46</td>
<td>10</td>
<td>9.08</td>
<td>7.64</td>
<td>0.09</td>
<td>0.04</td>
<td>0.1</td>
<td>0.3</td>
<td>13.8</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>M2</td>
<td>Located in North Taiwan</td>
<td>34.44</td>
<td>10</td>
<td>8.09</td>
<td>9.45</td>
<td>0.08</td>
<td>0.04</td>
<td>0.1</td>
<td>0.3</td>
<td>17.6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>M3</td>
<td>Located in North Taiwan</td>
<td>23.21</td>
<td>10</td>
<td>6.52</td>
<td>7.99</td>
<td>0.06</td>
<td>0.03</td>
<td>0.0</td>
<td>0.3</td>
<td>61</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>M4</td>
<td>Located in North Taiwan</td>
<td>45.76</td>
<td>10</td>
<td>8.78</td>
<td>7.35</td>
<td>0.08</td>
<td>0.04</td>
<td>0.0</td>
<td>0.3</td>
<td>40.2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>M5</td>
<td>Located in North Taiwan</td>
<td>35.33</td>
<td>10</td>
<td>9.13</td>
<td>6.92</td>
<td>0.08</td>
<td>0.04</td>
<td>0.0</td>
<td>0.3</td>
<td>35.8</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>M6</td>
<td>Located in North Taiwan</td>
<td>13.85</td>
<td>10</td>
<td>6.96</td>
<td>6.99</td>
<td>0.06</td>
<td>0.04</td>
<td>0.0</td>
<td>0.3</td>
<td>37.6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

(Continued)
<table>
<thead>
<tr>
<th>Study</th>
<th>Freshwater classification</th>
<th>Name</th>
<th>Sample</th>
<th>Observation</th>
<th>Bacterial counts (cells/ml)</th>
<th>Depth (m)</th>
<th>pH</th>
<th>DO (mg/l)</th>
<th>Conductivity (mS/cm)</th>
<th>Salinity (PSU)</th>
<th>DOC (mg/l)</th>
<th>POC (mg/l)</th>
<th>Orthophosphate (μM)</th>
<th>Total phosphorus (μM)</th>
<th>Silicate (μM)</th>
<th>Ammonia (μM)</th>
<th>Nitrite (μM)</th>
<th>Nitrate (μM)</th>
<th>Total nitrogen (μM)</th>
<th>N/P ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ghai et al. (2011)</td>
<td>River</td>
<td>Solimões-Amazon</td>
<td>Whitewater streams with high sediment concentrations</td>
<td></td>
<td>8</td>
<td>7.09</td>
<td>5.05</td>
<td>109.7</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hemme et al. (2010)</td>
<td>Groundwater</td>
<td>Oak Ridge</td>
<td>FW301 Pristine groundwater</td>
<td></td>
<td>± 7.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>6.5</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Field Research Center</td>
<td>FW106 Uranium contamination</td>
<td></td>
<td>10^-1 – 10^2</td>
<td>± 12.00</td>
<td>3.70</td>
<td>0.26</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>10046.6</td>
</tr>
</tbody>
</table>

Abbreviations: DO, dissolved oxygen; DOC, dissolved organic carbon; N/P, nitrogen/phosphorus; POC, particulate organic carbon; PSU, practical salinity units.

doi:10.1371/journal.pone.0148296.t002
samples from the protected sampling points (Fig 2 and S1A Fig). Analysis of trace elements is shown in S2 Table.

Microbial community structure

The 454 pyrosequencing of the 16S rRNA gene revealed similar profiles at high taxonomic levels (phylum and some classes) among the samples (Fig 3 and S3 Table), and statistical differences were not found between any of the diversity indices tested (S4 Table). The phylum Proteobacteria (primarily Betaproteobacteria) accounted for 79.54% of the OTUs and was observed in all samples analyzed (Fig 3A). Burkholderiales, an order within Betaproteobacteria, was most frequently detected, contributing 70% of all OTUs (S3 Table). The OTUs from this order belonged mainly to the families Comamonadaceae (represented by the genus *Limnohabitans* and other genera) and Burkholderiaceae (represented by the genus *Polynucleobacter*). The abundance of Comamonadaceae was similar across sampling points during the wet season (P1 (30.09%), P2 (44.38%), P3 (24.10%), p > 0.05) and dry season (P1 (45.44%), P2 (46.07%), P3 (36.60%), p > 0.05). The abundance of the genus *Limnohabitans* was similar between seasons (p > 0.05) and across sampling sites (p > 0.05): P1 (0.81% and 2.29% in the wet and dry seasons, respectively), P2 (2.88% and 1.53%), and P3 (1.65% and 1.04%). The abundance of the genus *Polynucleobacter* was also similar between seasons (p > 0.05) and across sampling sites (p > 0.05): P1 (40.96% and 35.82% in the wet and dry seasons, respectively), P2 (14.6% and 16.98%), and P3 (50.43% and 31.42%). In contrast, genus *Flavobacterium* sp. (phylum Bacteroidetes) was more abundant during the dry season in P1 than in P2 and P3 (p < 0.05, S4 Fig). The other orders within Betaproteobacteria, such as Methylophilales and candidate division MND1, comprised 5.71% of the total sequences.

Other classes of Proteobacteria that were detected in our samples included Gammaproteobacteria and Deltaproteobacteria, which represented about 2.4% and 0.86% of the sequences, respectively (Fig 3B and S3 Table). The class Alphaproteobacteria accounted for about 3.63% of total sequences. Verrucomicrobia was the second most abundant phylum, comprising about 3.63% of the total sequences. The phyla Actinobacteria and Acidobacteria were less abundant (3.25% and 3.18% of the total sequences, respectively). Taken together, the relative abundance of the other phyla was about 7.0% (Fig 3B and S3 Table). No taxonomical differences were found between the wet and dry seasons.

The microbial structure of samples obtained from the protected areas differed from that of the unprotected area (Fig 3C). Although the relative abundance of Burkholderiales (Betaproteobacteria) was similar across samples, the three sampling points were clustered according to order. The following orders were more abundant in the protected areas P2 and P3: Methylophilales (Betaproteobacteria), Xanthomonadales (Gammaproteobacteria), Acidimicrobiales (Actinobacteria), Sphingomonadaceae (Alphaproteobacteria), Pedosphaeraceae (Verrucomicrobia), Acidobacteriales (Acidobacteria), and Chthoniobacterales (Verrucomicrobia) (Fig 3C). On the other hand, the orders Actinomycetales (Actinobacteria) and Rhodocyclales (Betaproteobacteria) were more abundant in the unprotected area P1. No segregation was observed according to season (dry versus wet). A similar pattern was seen in the OTU-level analysis (S5 Fig). However, the OTU-level analysis shows the influence of genus Candidatus *Rhodoluna* (order Actinomycetales) and genus C39 (order Rhodocyclales) in the clustering of samples from the unprotected area P1.

The taxonomic profile from metagenomic data (S6 Fig) revealed that *Polynucleobacter necessaries* was the species most frequently detected, contributing 17.12% of all OTUs (S5 Table). The most abundant genera according to percentage of total OTUs were *Polynucleobacter* (17.12%), *Acidovorax* (5.94%), *Burkholderia* (3.22%), *Polaromonas*, (3.07%), and *Ralstonia*
Fig 3. Taxonomic classification based on sequence analysis of 16S rRNA genes detected in Paraguaçu River water samples obtained from three sampling points (P1, unprotected area; P2 and P3, protected area) during both wet and dry seasons. A) Total relative abundance of taxonomic groups (kingdom, phylum, class, family, and genus). B) Relative abundance of taxonomic groups (kingdom, phylum, and class) without the contribution of the Comamonadaeae family and Polynucleobacter genus. C) Principal coordinates analysis (PCoA) of relative abundances of taxonomic groups (at the order level). PCoA plot of community similarities (unweighted UniFrac distance matrices). The relative abundance of each taxonomic level (kingdom, phylum, class, and order) for each sample and their taxonomic assignments were determined by QIIME 1.7.0 software with the Greengenes database (May 2013).

doi:10.1371/journal.pone.0148296.g003
(1.80%). Differences among samples were observed only in the dry season, during which *Arthrobacter* sp. was more abundant in P1 than in P2 and P3 (S7 Fig).

**Functional classification of metagenome data**

After quality control filtering, approximately 82.54% of the sequences (1.56 × 10⁷) were annotated with an assigned function using the SEED database (Fig 4A). Most of the annotated sequences (51.4%) were assigned to one of the following categories: protein metabolism; carbohydrates; amino acids and derivatives; cofactors, vitamins, prosthetic groups, pigments; RNA metabolism; or miscellaneous. Proteins with unknown function accounted for about 15.5% of the annotated sequences. In the wet season, sequences assigned to the cell wall and capsule category and the cofactor, vitamins, prosthetic groups, pigments category were more abundant in sampling points P1 and P3 than in P2 (Fig 4B), whereas in dry season sequences related to potassium metabolism were more abundant in P2 and P3 than in P1 (Fig 4C). A comparison of samples by season showed that annotated sequences related to the metabolism of aromatic compounds were more abundant in the dry season (S8 Fig).

Although they were not abundant, several genes that appear to be related to pesticide degradation were detected (Figs 5 and 6). In both seasons, genes related to arsenic resistance, copper homeostasis, and propanediol utilization were more abundant in P1. Genes related to biphenyl degradation and toluene 4-monoxygenase (T4MO) were also more abundant in P1 but only during the wet season (Fig 5). Genes related to cadmium resistance were more abundant in P1 during the dry season (Fig 6).

In contrast, the protected area P2 showed a higher abundance of genes related to potassium homeostasis and the benzoate transport and degradation cluster in both seasons, compared with the other sampling sites (Figs 5 and 6). Similarly, genes related to cobalt-zinc-cadmium resistance were more abundant in both protected areas during the dry season compared to the unprotected area. However, the abundance pattern of genes involved in the chlorobenzoate degradation pathway was different, with higher abundance during the dry season in P1 and P3 compared with P2.

Genes related to nitrogen, sulfur, and phosphorus metabolism were also evaluated (S9 and S10 Figs). Regarding genes involved in nitrogen uptake and assimilation (S9A and S10A Figs), ammonia assimilation, nitrogen fixation, and nitrosative stress genes were more abundant in P1 during the wet season. The relative abundance of nitrate and nitrite ammonification genes was higher in P2 than in P3 during the wet season, but was higher in P2 than in P1 during the dry season. During the dry season, urea decomposition and urease subunits genes were most abundant in P2.

Regarding sulfur-related genes (S9B and S10B Figs), P1 showed the highest abundance of inorganic sulfur assimilation and alkanesulfonate utilization in the wet season and the highest abundance of sulfur oxidation and alkanesulfonate utilization genes in the dry season. Although sulfur oxidation genes were more abundant in P1 than in P2 in the dry season, genes for alkanesulfonate assimilation were more abundant in P2 in the wet season.

Finally, regarding phosphorus metabolism (S9C and S10C Figs), genes involved in alkyphosphonate utilization were most abundant in P3 in wet season. In the dry season, genes involved in phosphate metabolism were more abundant in P2 than in P1.

**Discussion**

The high abundance of Burkholderiales in samples obtained from all three sampling sites in this study suggests that these microbes are members of the natural microbiome of the Paraguaçu River. However, we cannot rule out that the abundance of this group is due to i) runoff...
Fig 4. Functional diversity of the Paraguaçu River’s metagenomes (P1, unprotected area; P2 and P3, protected area) in both wet and dry seasons. Classification was based on the SEED database level 1 in the MG-RAST server. A) Relative abundance of genes grouped by functional role. B) Comparative analysis of functional profiles of Paraguaçu River water samples obtained in the wet season. C) Comparative analysis of functional profiles of Paraguaçu River water samples obtained in the dry season. Samples were compared by t-test (p < 0.05), followed by the Bonferroni correction using STAMP software.

doi:10.1371/journal.pone.0148296.g004
from agriculture lands or ii) enrichment of nutrients and presence of pesticides that could promote their growth in the river. Some Burkholderiales members (e.g. genus *Burkholderia*) are associated with plant roots [52] and could be related to disease resistance or plant disease [53]. *Polynucleobacter* (Burkholderiaceae) and *Comamonas* (Comamonadaceae), which form symbiotic relationships with plants, are found in fresh water [54, 55] and possess genes related to the degradation of pesticides used in agriculture [56]. The genus *Limnohabitans* (family Comamonadaceae) prefers freshwater environments that are nonhumic and nonacidic (pH > 6)
Limnohabitans often coexists with genus Polynucleobacter, which prefers acidic habitats (pH < 6), but they show different abundances according to the habitat features [58]. Although no differences were observed between ecological parameters of samples collected in the dry season, the genus Flavobacterium (phylum Bacteroidetes) was more abundant in samples collected outside the PNCD (P1) in the dry season. The Flavobacterium genus includes both nonpathogenic species and several species that cause fish diseases such as columnaris.
disease, which is caused by *Flavobacterium columnare* and typically occurs in warm waters (20°C –25°C) [59]. This genus has also been detected in arsenic-contaminated groundwater [60].

The protected sampling points located just inside the PNCD (P2 and P3) exhibited lower nutrient concentrations in both wet and dry seasons compared with the unprotected location (P1), which is near agriculture land and may therefore be affected by chemical fertilizers (S11 Fig). The concentrations of several nutrients (e.g. ammonia, nitrite, nitrate, and POC) and the bacterial cell count were highest in samples obtained from P1. Soil fertilization is a common practice in agriculture to increase plant development, but if those nutrients reach the river, as is frequently observed in the wet season, they can stimulate bacteria growth and eutrophication, promoting changes in biodiversity [61, 62]. The process of eutrophication can render unfeasible the use of water for drinking and agricultural activities (e.g. irrigation) because of the presence of bacterial toxins (e.g. microcystin, saxitoxin) [63]. Eutrophication also affects several food webs, thereby affecting nutrient cycling [64]. Several genes related to the utilization of nutrients such as nitrogen (wet season) and sulfur (both seasons) were more abundant in P1 than in the protected sites. Agricultural management could also increase the level of particulate organic matter (POM), quantified by POC, in the river. Microorganisms, which have been shown to colonize POM particles for use as a carbon source, play a role in POM dynamics, leading to increased POM levels in the water [65]. Thus our findings suggest a microbial community response to agriculture. In addition, the high amount of copper in Paraguaçu River sediment at the unprotected (P1) site suggests the presence of agricultural pesticides. Copper can persist in the environment for many years and can be toxic to aquatic organisms in high concentrations [66]. Genes related to copper homeostasis indicate a possible influence on the microbiota (e.g. development of copper resistance) [67]. Furthermore, the genus *Ralstonia* (family Burkholderiaceae) was one of the most abundant genera detected. The species *Ralstonia pickettii* has been reported in areas with high concentrations of copper, iron, nickel, and zinc and is able to degrade aromatic hydrocarbons and chlorinated phenolic compounds [68–70]. In the protected sites, the scarcity of available organic carbon and other compounds (e.g. ammonia and nitrite) may explain, at least in part, the increase in stress response genes during the dry season (S12 Fig).

Pesticide-related genes, such as those involved in arsenic resistance, propanediol utilization, biphenyl degradation, T4MO, and cadmium resistance, were more abundant in P1 than in the protected sites. These genes are related to the degradation of compounds commonly used in agriculture such as fungicides, bactericides, pesticide solvents, and fertilizers [71–77]. Moreover, our results showed a higher abundance of the genus *Arthrobacter* in P1 in the dry season. *Arthrobacter* (phylum Actinobacteria) is able to degrade atrazine, one of the most commonly used pesticides in several countries. Several studies have described the bioremediation of atrazine-contaminated soil with *Arthrobacter* spp. [78, 79]. Other genera detected include *Acidovorax* and *Polaromonas* (both from family Comamonadaceae), which are known for their arsenite-oxidizing abilities [60, 80]. Genes related to sulfur oxidation and alkanesulfonate utilization were also more abundant in P1 in both seasons, suggesting that agricultural activities increase the amount of sulfonate compounds in the environment and lead to their accumulation in freshwater [81]. Several of these sulfur-related genes were identified as belonging to Actinomycetales, and the presence of Candidatus *Rhodoluna* (order Actinomycetales) helps separate the unprotected sampling site (P1) from the protected sites (P2 and P3). The most protected site (P3) showed lower levels of pesticide-related genes than the other sampling points, indicating the efficacy of PNCD protection. However, even with PNCD protection, signs of anthropogenic activity were observed within the park. For example, P2 had a relatively high abundance of benzoate-related and potassium-related genes. Benzoate is an anion present
in emamectin benzoate, an insecticide used against the Lepidoptera *Helicoverpa armigera* [82]. This compound can persist in the environment by binding to particulate material or surfaces, with detrimental effects in aquatic environments [83, 84]. Potassium is used in agriculture because of its importance in several physiological processes in plants and in the water-holding capacity of soils [85]. Runoff may be a source of agriculture-related benzoate and potassium, influencing the bacterial community present in this theoretically protected environment.

We compared our results with those of other metagenomic studies reporting the physico-chemical properties of freshwater samples. Despite evidence of anthropogenic actions in the PNCD, water quality parameters of the Paraguaçú River are more similar to those of the Feitsui reservoir [49] than the Pearl estuary [47]. The Feitsui reservoir supplies water to several cities and is used to generate electricity, reflecting its relatively good water quality, whereas the Pearl estuary is intensely polluted with sewage and industrial waste. However, few parameters were evaluated across all studies, making detailed comparisons difficult.

This is the first metagenomic study conducted within the PNCD, which is part of the Caatinga biome. Our results showed important changes in microbial community structure and gene content in an unprotected area near the park, illustrating the importance of park protection to maintain microbial diversity and water quality. Although the PNCD provides some protection of water quality, agricultural activities around the park are still able to affect water quality within the park and may account for the presence of bacteria capable of pesticide degradation and assimilation. Thus, the present study provides evidence of the anthropogenic impact on the Caatinga and demonstrates the need for additional protection of areas near the border of PNCD.

**Supporting Information**

S1 Fig. Microorganism counts and organic carbon content in water samples obtained from the Paraguaçú River. A) Bacterial counts. B) Virus counts. C) Particulate organic carbon (POC). D) Dissolved organic carbon (DOC). Samples were compared by ANOVA ($\alpha < 0.05$), followed by the Tukey post hoc test using R statistical software.

(TIF)

S2 Fig. Nitrogen content in water samples obtained from the Paraguaçú River. A) Ammonia. B) Nitrate. C) Nitrite. D) Total nitrogen. Samples were compared by ANOVA ($\alpha < 0.05$), followed by the Tukey post hoc test using R statistical software.

(TIF)

S3 Fig. Phosphorus and silicate content in water samples obtained from the Paraguaçú River. A) Orthophosphate. B) Total phosphorus. C) Silicate. Samples were compared by ANOVA ($\alpha < 0.05$), followed by the Tukey post hoc test using R statistical software.

(TIF)

S4 Fig. Comparative analysis of 16S rRNA taxonomic profiles of Paraguaçú River water samples obtained in the dry season. Samples were compared by ANOVA, followed by the Tukey–Kramer post hoc test ($p < 0.05$) and Bonferroni correction for multiple comparisons. Taxa with small effect sizes were removed by filtering (effect size = 8.00) using STAMP software. OTUs were classified at the order level (3% dissimilarity) using the Greengenes database (May 2013) and QIIME 1.7.0 software.

(TIF)

S5 Fig. Principal coordinates analysis (PCoA) of relative abundances of taxonomic groups (at the genus level). The relative abundance of each genus for each sample and their taxonomic
assignments was performed using QIIME 1.7.0 software and the Greengenes database (May 2013).

**S6 Fig. Microbial community profile from metagenome sequencing.** OTUs were classified at the species level in MG-RAST using the M5NR database with default sequence quality thresholds and assigned with MEGAN 5.10.6.

**S7 Fig. Comparative analysis of metagenome taxonomic profiles of Paraguaçu River water samples obtained in the dry season.** Samples were compared by ANOVA, followed by the Tukey-Kramer post hoc test ($p < 0.05$) and Bonferroni correction for multiple comparisons. Taxa with small effect sizes were removed by filtering (effect size = 8.00) using STAMP software. OTUs were classified at the order level in MG-RAST using the M5NR database and default sequence quality thresholds.

**S8 Fig. Functional diversity of the Paraguaçu River’s metagenomes in both wet and dry seasons.** Classification was based on the SEED database level 1 in the MG-RAST server. A) Relative abundance of genes grouped by functional role. B) Comparison of genes involved in the metabolism of aromatic compounds in water samples obtained in the wet and dry seasons. Samples were compared by $t$-test ($p < 0.05$), followed by the Bonferroni correction using STAMP software.

**S9 Fig. Comparative analysis of functional profiles of Paraguaçu River water samples obtained in the wet season.** Samples were compared by ANOVA, followed by the Tukey-Kramer post hoc test ($p < 0.05$) without correction for multiple comparisons. Taxa with small effect sizes were removed by filtering (effect size = 8.00) using STAMP software. A) Nitrogen metabolism-related genes. B) Sulfur metabolism-related genes. C) Phosphorus metabolism-related genes.

**S10 Fig. Comparative analysis of functional profile of Paraguaçu River samples obtained in the dry the season.** Samples were compared by ANOVA, followed by the Tukey-Kramer post hoc test ($p < 0.05$) without correction for multiple comparisons. Taxa with small effect sizes were removed by filtering (effect size = 8.00) using STAMP software. A) Nitrogen metabolism-related genes. B) Sulfur metabolism-related genes. C) Phosphorus metabolism-related genes.

**S11 Fig. Satellite image of the study area.**

**S12 Fig. Functional diversity of the Paraguaçu River’s metagenomes in samples obtained from an unprotected area (P1) and protected areas (P2 and P3) in both wet and dry seasons.** Classification was based on the SEED database level 1 in the MG-RAST server. A) Relative abundance of genes grouped by functional role. B) Comparison of genes involved in photosynthesis detected in water samples from the unprotected and protected areas obtained in the wet season. C) Comparison of genes involved in motility and chemotaxis; potassium metabolism; stress response, and virulence, disease, and defense detected in water samples from the unprotected and protected areas obtained in the dry season. Samples were compared...
by t-test (p < 0.05), followed by the Bonferroni correction using STAMP software.

**S1 Table. Number of sequences obtained by metagenome sequencing.** Metagenomic DNA libraries were constructed with the Nextera DNA Sample Preparation Kit (Illumina) and 2 × 250 bp paired-end sequencing by Illumina MiSeq system according to the manufacturer’s instructions.

**S2 Table. Metal, sulfur, and phosphorus concentrations of sediments in water samples from the Paraguaçú River.** ND = Not detected. " Data from one replicate.

**S3 Table. Relative abundance of OTUs and sequence information of Paraguaçú River microbiota.** OTUs were classified at the level of genus using the Greengenes database (May 2013) at 3% dissimilarity performed by QIIME 1.7.0 software.

**S4 Table. Diversity indices of water samples obtained from the Paraguaçú River at 3% dissimilarity for the 16S rRNA gene.**

**S5 Table. Relative abundance of OTUs and sequence information of Paraguaçú River metagenome.** OTUs were classified at the species level using the M5NR database with default sequence quality thresholds performed using the MG-RAST server.

**Acknowledgments**

This work was supported by grants from the Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) and Instituto Chico Mendes de Conservação da Biodiversidade (ICMBio), Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES). We are grateful for the support offered by Fundação de Amparo à Pesquisa do Estado do Rio de Janeiro (FAPERJ) and National Park of Chapada Diamantina. F.A.C.L. is supported by a CAPES fellowship.

**Author Contributions**

Conceived and designed the experiments: FACL RAE FLT RHK. Performed the experiments: FACL ASC TPR. Analyzed the data: FACL ECPC RHS TPR RAE. Contributed reagents/materials/analysis tools: RP CER CTT RAE FLT RHK. Wrote the paper: FACL FLT RHK.

**References**


58. Jezbera J, Jezberova J, Koll U, Hornak K, Simek K, Hahn MW. Contrasting trends in distribution of four...
59. Hillebrand H, Gruner DS, Borer ET, Bracken ME, Cleland EE, Elser JJ, et al. Consumer versus resource...
60. Starliper CE. Bacterial coldwater disease of fishes caused by Flavobacterium psychrophilum. Journal...
61. Haust D, Adukauskiene N, Caron M, Hahn MW. Phylogenetic evidence for Polynucleobacter cosmopolitanus...
70. Paul D, Kazy SK, Gupta AK, Pal T, Sar P. Diversity, metabolic properties and arsenic mobilization potential of indigenous bacteria in arsenic contaminated groundwater of West Bengal, India. PloS one. 2015; 10(3):e0118735. doi:10.1371/journal.pone.0118735 PMID: 25799108; PubMed Central PMCID: PMC4370401.


