Universidade de Brasília

Instituto de Ciências Biológicas

Departamento de Biologia Celular

Programa de Pós-Graduação em Biologia Microbiana

PERFIL TAXONÔMICO E FUNCIONAL MICROBIANO EM AMBIENTES AQUÁTICOS

Fabyano Alvares Cardoso Lopes

Brasília – DF

Abril de 2016

Universidade de Brasília

Instituto de Ciências Biológicas

Departamento de Biologia Celular

Programa de Pós-Graduação em Biologia Microbiana

PERFIL TAXONÔMICO E FUNCIONAL MICROBIANO EM AMBIENTES AQUÁTICOS

Fabyano Alvares Cardoso Lopes

Orientador:

Dr. Ricardo Henrique Krüger

Co-orientador estrangeiro: **Dr. Rob Edwards**

Tese apresentada ao Programa de Pós-Graduação em Biologia Microbiana da Universidade de Brasília, como requisito parcial para obtenção do título de Doutor em Biologia Microbiana

Brasília – DF

Abril de 2016

Trabalho realizado nos seguintes laboratórios: Laboratório de Enzimologia, Departamento de Biologia Celular do Instituto de Ciências Biológicas da Universidade de Brasília (UnB), Brasília, DF, Brasil; Laboratório de Oceanografia, Departamento de Biologia Marinha, Universidade Federal do Rio de Janeiro (UFRJ), Rio de Janeiro, RJ, Brasil; Bioinformatics Lab, Department of Computer Science, San Diego State University (SDSU), San Diego, CA, Estados Unidos da América.

Orientador:

Dr. Ricardo Henrique Krüger

Co-orientador estrangeiro: Dr. Rob Edwards

Banca Examinadora:

Prof. Dr. Alexandre Siqueira Guedes Coelho (UFG) – Examinador Externo

Profa. Dra. Cristine Chaves Barreto (UCB) – Examinadora Externa

Prof. Dr. Georgios Joannis Pappas Júnior (UnB) – Examinador Interno

Prof. Dr. Helson Mario Martins do Vale (UnB) – Examinador Interno

Prof. Dr. Ricardo Henrique Krüger (UnB) – Orientador

Membro Suplente:

Prof. Dr. Robert Neil Gerard Miller (UnB)

“Knowledge, like air, is vital to life.
Like air, no one should be denied it.”

Alan Moore, V for Vendetta

Agradecimentos

Gostaria de agradecer primeiramente à minha esposa, amiga, companheira Amanda Rafaela. Você foi fundamental em tudo que conquistei até hoje em toda a minha vida. Você esteve comigo por toda a minha caminhada, desde o incentivo para a minha escolha por cursar Ciência Biológicas até a minha travessia por essa etapa tão árdua que é o doutorado. Amanda, você sempre foi, é e será a pessoa mais importante na minha vida. Obrigado, meu bem! Eu te amo.

Agradeço a toda minha família, sempre tão presente, tão amiga, tão preocupada comigo, pessoas que deram o suporte para eu atingir meus objetivos, me apoiando sempre, independente do horário, do clima ou até mesmo da distância entre Goiânia-Brasília-San Diego. Irmão, cunhadas (os), primos, tios, avós, sogros, aqui está o meu muitíssimo obrigado. Pais, muito obrigado por tudo, amo vocês!

Agradeço a todos os integrantes do Laboratório de Enzimologia-UnB, vocês deixaram minha vida muito mais feliz! Em especial quero agradecer ao grupo Metagenômica, todos vocês foram muito especiais em minha vida, me acolheram, me fizeram sentir em casa, mais do que isso, me adotaram como um integrante dessa família tão carinhosa e preocupada um com o outro. Obrigado!

Agradeço a todos os meus amigos, em especial, Andrei e Marcelo, companheiros que compartilhei vários momentos nessa caminhada, dando a força necessária para seguir sempre em frente, além de oferecerem opiniões amigas para o meu bem. Agradeço ao JP (meu eterno pupilo) que apesar da maioria das vezes parecer não falar nada sério, é uma pessoa extremamente leal, íntegra e sábia e ao Alonsito pela sua amizade. Obrigado danados!

Agradeço à Elisa, pessoa que se tornou mais do que uma amiga, uma irmã. Você sabe que nenhum trabalho que faço está completo sem você! Obrigado por ser minha amiga.

Agradeço a todas as pessoas que fizeram parte da minha etapa em San Diego e por tudo que vivi nessa cidade. Daniel, Jeff, Hans, Taylor, Adrian"s", Cynthia, Mike, Doug, Mariu, thanks guys! Em especial, quero agradecer a um amigo que fiz nessa cidade,

grande sereno Geni. Meu velho, sou muito feliz por ter te encontrado, irmão. Obrigado por tudo que fez e que continua fazendo por mim.

Um agradecimento em especial aos meus orientadores Dr. Ricardo Henrique Krüger e Dr. Rob Edwards por terem me acolhido e me oferecido condições de trabalho para mostrar o meu melhor, além de terem contribuído ao meu crescimento científico. Ao Dr. Fabiano L. Thompson, parceiro de trabalho que ofereceu toda a estrutura de seu laboratório, sempre acreditando no meu potencial. Ao prof. Cirano José Ulhoa que sempre me apoiou no meu processo de formação, e novamente reitero, espero me tornar um décimo do que você é profissionalmente, pois você sempre será meu modelo de pesquisador.

Finalmente, agradeço à Universidade de Brasília e ao Programa de Pós-Graduação em Biologia Microbiana pela oportunidade. Agradeço à CAPES pelas bolsas de doutorado e sanduíche, fundamentais para o desenvolvimento desse trabalho.

ÍNDICE

LISTA DE FIGURAS

CAPÍTULO 1 - Microbial community profile and water quality in a protected area of the Caatinga biome

LISTA DE TABELAS

RESUMO GERAL

A mensuração da diversidade microbiana é um dos maiores desafios do campo microbiológico, principalmente por problemas metodológicos. Com o avanço de novas metodologias foi possível observar que a diversidade de microrganismos era maior do que se pensava, assim, possibilitando o estudo desse conjunto de microrganismos. O estudo dos genomas de diversos microrganismos contidos em um dado ambiente é denominado de metagenômica. A metagenômica pode ser utilizada para o estudo de diversos tipos de ambientes, como solo, ar, corpo humano, intestino de cupim, entre outros. Dentro dos ambientes estudados pela metagenômica, o ambiente aquático vem sendo alvo de diversos estudos. Apesar de apresentar diversos estudos descrevendo diferentes profundidades e até mesmo diferentes hábitats (esponja, corais por exemplo), ainda existem inúmeros hábitats no ambiente marinho que ainda não possuem estudos focados sobre a microbiota. Diferente do ambiente marinho, trabalhos focados na descrição da microbiota de água doce são escassos. Além disso, apesar de toda sua importância ecológica, diversos desses corpos d'água estão sendo perturbados por atividades antropogênicas, levando à alteração da microbiota. Alguns trabalhos focaram na detecção de biossensores capazes de detectar impactos antropogênicos no meio ambiente. O objetivo de presente trabalho foi contribuir na descrição da comunidade microbiana em ambientes aquáticos. Para atingir essa meta foram realizados dois estudos, o primeiro estudo foi realizar a análise taxonômica e funcional da comunidade bacteriana do rio Paraguaçú na estação de chuva e seca, além de avaliar o efeito de proteção do Parque Nacional da Chapada Diamantina (PNCD) sobre a qualidade da água e da diversidade da microbiota do rio Paraguaçú. Outro estudo realizado foi o primeiro trabalho metagenômico focado em poças de maré. Esse trabalho deu enfoque na descrição do perfil taxonômico e funcional da microbiota das poças de maré situados em Ocean Beach, San Diego, CA, EUA. Ambos ambientes trabalhados possuem um grande potencial biotecnológico e se apresentaram como fundamentais na determinação do perfil da microbiota encontrada. No rio Paraguaçú foi possível observar uma predominância de genes relacionados à degradação de pesticidas (como o Benzoato), enquanto nas poças de maré foi possível observar uma maior abundância de genes relacionados à tolerância a ambientes com alta concentração salina. Novos estudos devem ser realizados para ambos ambientes buscando elucidar melhor os processos que ocorrem nesses ambientes.

Palavras-chave: Metagenômica; Ambiente aquático; microbiota; Diversidade.

GENERAL ABSTRACT

Determining the true diversity in a microbial community is one of the biggest challenges in microbiology. The development of new techniques has revealed a higher microbial diversity than what was previously thought and provided a new way to study these organisms. Metagenomics, the study of microbial DNA recovered from specific environments, allows the study of microbial communities in soil, air, human body, and termite gut. Among the environments studied using this approach, water is one of the most important. Each of the two types of water environments (seawater and freshwater) has a specific microbial community. Although previous studies have described microbial communities in different water layers and habitats (*e.g.* sponge, coral), the microbial communities of several seawater habitats have not been studied, and studies describing microbial communities in freshwater environments are rare. Moreover, ecologically important freshwater environments are being disturbed by anthropogenic activity, which can change the microbial profile. Previous studies have focused on identifying biosensors that can detect anthropogenic impacts on the environment. The aim of this work was to characterize the microbial communities in two water environments, and this objective was developed in two separate studies. The first described taxonomic and functional analyses of the bacterial community in the Paraguaçú River during both wet and dry seasons. Additionally, we evaluated the protective effect of Parque Nacional da Chapada Diamantina (PNCD) on water quality and microbial diversity in this river. The second study was the first metagenomic study of tide pools, in which we described the taxonomic and functional profiles of microbial communities in tide pools at Ocean Beach, San Diego, CA, USA. Both studies describe environments with great biotechnological potential and showed themselves as fundamental factors shaping the microbial communities. For example, in the Paraguaçú River we observed a high abundance of genes encoding enzymes capable of degrading pesticides such as emamectin benzoate, whereas tide pools showed a high abundance of halotolerance genes. Further studies are needed to elucidate processes in both environments.

Keywords: Metagenomic; Water environment; Microbial community; Diversity.

INTRODUÇÃO

Ecologia microbiana e técnicas aplicadas ao estudo da diversidade microbiana

Os microrganismos estão inseridos no ecossistema que possui diversos hábitats. Hábitat é um dado ambiente compostos por fatores abióticos e bióticos [1], sendo que cada hábitat possui características intrínsecas que são fundamentais na determinação de qual microrganismo o habitará. Assim, certos fatores são cruciais para seleção de microrganismos que irão colonizar um determinado hábitat: nutrientes disponíveis e fatores ambientais (salinidade, pH, temperatura, etc) associados a esse hábitat. Caso a disponibilidade de nutrientes e os fatores ambientais forem favoráveis para diversos microrganismos, isso dará origem a diversas populações originando uma comunidade. A diversidade de microrganismos, além de como se relacionam inter e intraespecificamente e com o ambiente que estão inseridos são estudados na ecologia microbiana [2].

Dentro da ecologia microbiana, existem dois conceitos que mensuram a microbiota de um hábitat, são eles: riqueza e abundância [2]. O conceito de riqueza está relacionado com o número de espécies em um dado local. A riqueza pode ser mensurada numericamente através do índice de diversidade, como o índice de Shannon [3] ou índice de Simpson [4], por exemplo, que possibilita a comparação entre diferentes amostras. Porém, diferente do conceito de riqueza de espécies, o conceito de abundância está vinculado a contribuição de uma espécie em uma comunidade, ou seja, é a proporção em número de indivíduos de uma dada espécie.

Com o avanço de novas metodologias, como contagem direta por microscopia de fluorescência, foi possível observar que a diversidade de microrganismos era maior do que se encontrava em estudos anteriores em placas, fenômeno conhecido como “A grande anomalia da contagem de placas”. Através da contagem direta em microscopia de fluorescência em uma amostra de uma grama de solo, observou-se cerca de 10 bilhões de organismos procariontes enquanto se visualizava em placas um número 100 a 1000 vezes menor [5]. Assim, novas formas de avaliar essa diversidade microbiana têm sido desenvolvidas gerando inúmeras informações, antes desconhecidas.

Inúmeras técnicas podem avaliar a diversidade microbiana, seja diversidade taxonômica ou funcional que podem ser de baixa, intermediária e alta resolução. Dentro das técnicas de baixa resolução, a observação direta da amostra por microscopia consegue oferecer uma noção dos microrganismos presentes em termos morfológicos, além de fornecer o número de microrganismos presentes com uma certa precisão. Outra técnica de baixa resolução é o método de reassociação de DNA da comunidade microbiana, onde a estimativa da diversidade procariótica é dada pela porcentagem de DNA que é renaturada após a desnaturação do DNA da comunidade microbiana, ou seja, quanto menor o teor de DNA renaturado, mais diversa é a amostra, porém a composição da comunidade não é acessada [6,7].

Existem outros métodos de baixa resolução que se baseiam em marcadores bioquímicos como a análise de ácidos graxos. Na análise de ácido graxos é realizada a extração desses compostos provenientes dos microrganismos da amostra após a identificação do composto, se for um marcador bioquímico característico, será possível identificar as espécies, permitindo o monitoramento da diversidade de um ambiente [8]. Todavia esse método é utilizado com algumas ressalvas. Fakruddin e Mannan, em sua revisão, alertam que vários microrganismos possuem os mesmos ácidos graxos o que pode impedir a detecção na mudança da diversidade [9]. Outra ressalva realizada por Fakruddin e Mannan é que a composição de ácidos graxos dos microrganismos pode variar de acordo com as condições de crescimento.

Dentro dos métodos de resolução intermediária estão presentes as técnicas por fingerprint, como DGGE, TGGE, RISA e ARDRA. Essas técnicas consistem na amplificação do gene rRNA 16S (DGGE, TGGE) ou da região ITS (RISA, ARDRA) e separação dos produtos gerados em um gel de eletrofoforese, sendo que o perfil apresentado no gel é utilizado para estimar a diversidade. Godheja *et al.* relatam que apesar das limitações, principalmente pelo desenho dos primers para grupos específicos, essas técnicas têm sido bastante utilizadas em diversos estudos [10]. Porém, essas técnicas não conseguem mensurar de uma forma precisa a abundância e diversidade, principalmente pela dificuldade da reprodutibilidade.

Outro método de resolução intermediária amplamente utilizado é a hibridização por fluorescência *in situ* (FISH) que consiste na hibridização de um marcador filogenético (rRNA 16S por exemplo) com um fluoróforo específico no microrganismo-alvo,

permitindo a detecção e quantificação de células de grupos filogenéticos conhecidos, permitindo a descrição da diversidade da amostra. Porém, conforme Douterelo *et al.* essa técnica possui dificuldade de diferenciar células vivas de células mortas [11]. Sanz e Köchling relatam que esse método requer uma sequência de nucleotídeo conhecida do organismo alvo para o desenho de uma sonda, portanto, permitindo a detecção apenas de microrganismos conhecidos [12].

Para a análise da diversidade funcional, a técnica de microarranjo de fenótipo é utilizada. Esse método consiste no crescimento da comunidade microbiana em diferentes fontes nutricionais (12, 48 ou 96 fontes), exibindo um potencial catabólico característico [13]. Apesar de ser uma alternativa mais barata que o sequenciamento, essa técnica pode superestimar a presença de algumas bactérias, alterando a verdadeira contribuição dessas bactérias na comunidade.

Metagenômica

O estudo dos genomas de diversos microrganismos contidos em um dado ambiente é denominado de metagenômica [14]. Esse estudo pode contemplar tanto a diversidade presente em um ambiente (bactérias, arqueas, fungos, protozoários) quanto a investigação de seu potencial funcional, ou seja, seus recursos genéticos com a facilidade do não-cultivo desses microrganismos.

O termo "metagenômica" foi empregado pela primeira vez por Handelsman *et al.* [14], onde foi levantada uma possibilidade de poder acessar compostos com apelo industrial produzidos por microrganismos desconhecidos de difícil cultivo por meio do sequenciamento do DNA total da comunidade microbiana. Assim, diversos trabalhos foram realizados a partir de diversos tipos de amostra, como solo, água, ar, resíduos industriais, etc.

Inicialmente, o campo da metagenômica era associado à utilização de vetores BAC e fosmídeos [14–17] que eram capazes de manter estavelmente grandes segmentos de DNA (>100kb) exógeno (proveniente da amostra estudada). Esses segmentos de DNA (provenientes de produtos de PCR – gene rRNA 16S ou gene rRNA 18S geralmente – ou

o próprio DNA) poderiam ser facilmente sequenciados fornecendo informações de vias enzimáticas inteiras que podem estar presentes em um único clone BAC e fosmídeos.

Atualmente, os estudos metagenômicos são realizados a partir do sequenciamento *shotgun* direto do DNA "ambiental" [18,19]. Assim, no campo da metagenômica é possível descrever as espécies e/ou genes presentes no ambiente através de identificação realizada por meio do contraste com banco de dados específicos (Greengenes [20], SILVA [21] para taxonomia e SEED [22] para funcional, por exemplo), possibilitando a mensuração da diversidade microbiana de uma forma mais precisa. Por meio do processamento dos dados gerados pelo sequenciamento realizado em softwares e/ou plataformas (QIIME [23], MG-RAST [24]) é possível investigar o potencial gênico de uma dada comunidade. Porém, essa técnica prioriza os membros dominantes de um determinado ambiente, sendo que são necessárias certas modificações e/ou outras técnicas para analisar organismos mais raros.

Metagenômica em ambientes aquáticos

A metagenômica pode ser utilizada para o estudo de diversos tipos de ambientes, como solo [25], ar [26,27], corpo humano [28,29], intestino de cupim [30], entre outros. Dentro dos ambientes estudados pela metagenômica, o ambiente aquático vem sendo alvo de diversos estudos focados na descrição e nas interações que ocorrem na microbiota. Dentro do ambiente aquático, há diversos tipos de ambientes como água do mar e água doce, por exemplo.

Água do mar

Cobrindo cerca de 70% do planeta Terra, o ambiente marinho possui uma grande importância em diversos papeis vitais aos seres que habitam esse planeta, sendo que a maioria dessas atividades são desempenhadas pelos microrganismos. Nos ambientes marinhos, esses microrganismos são encontrados desde a superfície até em profundidades maiores (11000m) com altas pressões (100Mpa) [31]. De acordo com a profundidade, diversos trabalhos têm sido realizados focados na descrição da microbiota de cada camada

na coluna d'água do oceano (superfície, mesopelágico, mar profundo e sedimento marinho [31].

A maioria dos estudos marinhos são realizados nos oceanos Pacífico e no Atlântico Norte [32]. Descrevendo a microbiota presente na superfície marinha do mar de Sargasso (Atlântico Norte), Venter *et al.* relataram que os mais abundantes filos nesse ambiente eram Proteobacteria, Actinobacteria, Cyanobacteria, Firmicutes, entre outros [19]. Nesse mesmo trabalho, Venter *et al.* detectaram mais de 1800 espécies, maioria pertencente a Alphaproteobacteria e Gammaproteobacteria. Em outro estudo, Mueller *et al.* sequenciaram amostras da superfície do mar em Monterey (CA, EUA) no oceano Pacífico [33]. Nesse estudo, eles relataram a predominância dos filos Proteobacteria e Bacteriodetes.

Existem alguns esforços para o preenchimento dessa ausência de conhecimento em outras regiões marinhas. Alves-Junior *et al.* realizaram uma caracterização metagenômica na região sul do oceano Atlântico em diferentes profundidades - superfície, camada profunda de máxima concentração de clorofila (DMC; 48–82m), zona afótica (1200m) [34]. Nesse trabalho eles relataram a predominância do filo Proteobacteria nas amostras e que a abundância de Cyanobacteria foi maior quando comparada a outros oceanos [34]. Eles também relataram que as abundâncias de Alphaproteobacteria e Cyanobacteria diminuíram com o aumento da profundidade, enquanto ocorreu um aumento na abundância de Rhodobacteriales [34].

Além da coluna d'água, o ambiente marinho oferece diversas opções de hábitats para microrganismos. Trindade-Silva *et al.* descreveram a microbiota associada à esponja endêmica *Arenosclera brasiliensis*, onde foi encontrada a prevalência dos gêneros *Burkholderia*, *Pseudomonas* e *Alteromonas* [35]. A microbiota associada à esponja *A. brasiliensis* apresentou uma maior abundância de genes relacionados a transporte de membrana e com o metabolismo de carbono em relação a microbiota presente na coluna d'água. Também foi relatado uma maior abundância de genes relacionados à síntese e degradação de metabólitos secundários, potencial que foi investigado posteriormente por Rua *et al.* [36]. Um outro exemplo de ambiente marinho estudado são os corais. Bruce *et al.* descreveram a microbiota presente em corais no Parque Nacional Marinho de Abrolhos e de corais fora do perímetro do Parque [37]. Ainda existem inúmeros hábitats que ainda não possuem estudos relacionados à microbiota.

Além de estudos ecológicos, os ambientes marinhos também têm fornecido microrganismos alvos de produtos biotecnológicos. Por exemplo, Zhu *et al.* descobriram e modificaram geneticamente uma linhagem de *Marinactinospora thermotolerans* (Actinomycete) em sedimento marinho capaz de produzir um potente agente antibacteriano [38]. A partir de uma microbiota associada à esponja *A. brasiliensis*, Rua *et al.* isolaram microrganismos cultiváveis que apresentam atividade antimicrobiana, possuindo um grande potencial biotecnológico [36].

Poças de maré

Um tipo de ambiente vinculado ao ambiente marinho é a poça de maré. As poças de maré são áreas conhecidas como ambientes sob forte estresse onde os seus habitantes estão sujeitos a oscilações de temperatura e salinidade, além de dissecação e hipoxia [39]. Essas condições de estresse podem ser causadas por diversos fatores como incidência de luz solar (elevando a temperatura), oxigênio, entre outros [40]. Devido à limitação espacial e à facilidade para mensurar nutrientes e outros compostos, esse ambiente oferece a possibilidade de desenvolver estudos ecológicos [41,42]. Em geral, os estudos realizados nas poças de maré são focados em macrorganismos, como alguns crustáceos, algas, camarões e peixes [43–47].

Essas zonas entremarés são também influenciadas pelos seus habitantes. Alguns organismos podem usar os nutrientes em uma proporção diferente dos demais organismos, afetando diretamente a disponibilidade do nutriente em questão nas poças de maré [48,49]. Bjork *et al.* relataram a capacidade da alga *Ulva intestinalis* em aumentar o pH e reduzir a concentração do carbono inorgânico em uma poça de maré [50]. Todas essas mudanças podem ser fundamentais para moldar a microbiota presente nas poças de maré. Estudos relacionados a microrganismos em poças de maré são raros e esses poucos trabalhos estão vinculados a estudos biotecnológicos [51,52]. Nenhum estudo metagenômico foi realizado nesse ambiente.

Água doce

Em relação à água doce, diversos estudos foram realizados focando a descrição da microbiota em rios, lagos e reservatórios, porém são de escalas menores e ainda bastante

escassos quando comparados aos estudos realizados em ambientes marinhos [19,53]. Esses poucos trabalhos já foram capazes de mostrar que a comunidade microbiana presente em água doce é diferente da comunidade presente em água salgada [54]. Esses ambientes de água doce possuem uma grande importância para o homem, além de estarem relacionados a diversos ciclos biogeoquímicos [55].

As caracterizações microbianas são realizadas ao longo da coluna d'água e dos sedimentos desses ambientes de água doce. Ghai *et al.* realizaram uma análise da microbiota do rio Amazonas (Brasil), além de uma caracterização dos parâmetros físico-químicos de sua água [56]. O rio Amazonas apresentou uma microbiota dominada por Actinobacteria, Alphaproteobacteria e Betaproteobacteria, apresentando o gênero *Polynucleobacter* como o mais abundante entre as bactérias. Além dessa descrição, Ghai *et al.* também mostraram a importância do metabolismo heterotrófico nesse rio.

Em outro trabalho, que ajuda a esclarecer a microbiota de água doce, Gibbons *et al.* realizaram um estudo sobre a diversidade taxonômica e funcional da microbiota do sedimento do rio Tongue (EUA) [57]. Os autores relataram a predominância dos filos Proteobacteria, Acidobacteria, Bacteroidetes e Planctomycetes. Nesse mesmo estudo também foi relatado o potencial da microbiota em degradar compostos aromáticos e hidrocarbonetos, como naftaleno, nitrotolueno, benzoato entre outros. Segundo os autores, provavelmente, o potencial de degradação desses compostos na microbiota foi estimulado por estar vinculado à proximidade de uma carvoaria próxima ao corpo d'água.

Apesar de toda sua importância ecológica, diversos desses corpos d'água estão sendo perturbados por atividades antropogênicas, levando à alteração da microbiota [58] e, por sua vez, impactando em diversos ciclos biogeoquímicos, visto que a microbiota tem papel fundamental nessa etapa. Liu *et al.* relataram que no estuário Pearl (China) havia uma alta abundância de Sphingomonadales e Rhizobiales (Alphaproteobacteria), além de uma maior abundância de fitoplâncton. Os autores explicam que esse cenário se deve à alta concentração de nutrientes proveniente de impactos antropogênicos no estuário [59].

Assim, a partir de estudos nesses ambientes é possível realizar a identificação de biomarcadores de atividades antropogênicas para essas áreas que estão em constante contato com o homem. Smith *et al*. (2015) mostraram que comunidades bacterianas de áreas contaminadas por urânio e nitrato ou óleo podem ser diferenciadas de comunidades

bacterianas de ambientes saudáveis, assim, podendo ser utilizadas como biossensores capazes de detectar impactos antropogênicos no meio ambiente [60].

Rio Paraguaçú

O rio Paraguaçú faz parte da bacia do Paraguaçú, sendo que essa bacia está contida em grande parte na Caatinga, assim, possuindo uma grande importância para o estado da Bahia. Essa bacia propicia o desenvolvimento de atividades agropecuárias e mineração, contém o manancial de abastecimento de Salvador e região metropolitana, além de abrigar inúmeras nascentes [61,62].

O curso do rio Paraguaçú está contido inteiramente na Bahia, possuindo 500 km de extensão desde a sua nascente localizada no município de Barra da Estiva (13° 37' 33" S 41° 19' 37" W, 1200m de altitude), passando pela Chapada Diamantina e pela Caatinga do Recôncavo Baiano, chegando à desembocadura na Baía de Todos os Santos (Pereira, 2008). Em seu percurso existem três barragens: do Apertado, Bandeira de Melo e Pedra do Cavalo. A barragem do Apertado está localizada na cidade de Mucugê, antes do rio Paraguaçú entrar no Parque Nacional da Chapada Diamantina, essa barragem possui uma capacidade de 108,89 hm^3 e uma vazão de 8,90 m^3/s, com o propósito de abastecer o município, nas zonas urbana e rural [63].

Apesar de sua extrema importância, o rio Paraguaçú sofre inúmeros impactos ambientais que interferem na sua manutenção, diversidade e qualidade. Esses impactos estão ligados principalmente a atividades agrícolas, como a utilização de pesticidas, desmatamentos e/ou queimadas para formação de pastagem que potencializam as formações de erosões em áreas sem cobertura vegetal com consequente assoreamento do rio [61], além da poluição com metais pesados pela atividade de garimpo [64].

Na Chapada Diamantina, parte do rio Paraguaçú está protegida pelo Parque Nacional da Chapada Diamantina (PNCD). O PNCD possui 152.141 ha e apresenta em sua composição Caatinga (bioma oficial), Cerrado e Mata Atlântica [65]. Sua geologia é constituída pelas formações Paraguaçú, Tombador, Cabloco, Morro do Chapéu e Bebedouro. A formação do grupo Paraguaçú possui as rochas mais antigas da unidade de conservação, compostas por silitos, argilitos, com arenitos e conglomerados [66]. O grupo

Paraguaçú compreende uma sequência metassedimentar pelito-psamítica, cortada por rochas magmáticas [67].

O clima na Chapada Diamantina se alterna de subúmido a seco, sendo que nas regiões mais próximas da nascente pode variar de úmido a subúmido coberto com florestas remanescentes e campos rupestres (S.E.M.A.R.H. e S.R.H., 2005). Segundo o Instituto Nacional de Meteorologia (INMET), entre o período de agosto/2012 a maio/2013, houve variação na temperatura de 17,8 ºC (25/08/2012) a 33 ºC (23/10/2012) e variação da umidade de 43% (30/09/2012) a 100% (23/11/2012), enquanto a precipitação pluviométrica do período de agosto/2012 a maio/2013 chegou a 124 mm (23/01/2013), sendo que o mês de novembro de 2012 apresentou precipitação em quase todos os dias. O período de 25 de janeiro de 2013 a 21 de março de 2013 foi considerado como um dos piores períodos de seca da região possuindo apenas 8 mm de precipitação em dois meses.

O PNCD abriga diversas espécies com risco de extinção, como gavião-pomba, tamanduá-bandeira, tatu-canastra, onça-parda, entre outros [68]. Além disso, o rio Paraguaçú abriga uma espécie endêmica de peixe, *Hypostomus jaguar* [69], o que reforça a preservação e a realização de estudos no rio e na região que o engloba. Apesar da necessidade em caracterizar a região, poucos estudos foram realizados e em sua maioria com intuito de descrever sua flora, fauna e geologia. A Chapada Diamantina possui em sua composição uma fitofisionomia bastante distinta e uma formação geológica muito característica, o que potencializa a presença de micro-organismos característicos da região. Além disso, a avaliação da diversidade microbiana, tanto em áreas naturais como afetas pelo homem pode ajudar na compreensão do panorama ecológico de uma determinada região.

OBJETIVO GERAL E ESTRUTURA DA TESE

O objetivo geral do presente estudo foi descrever a comunidade microbiana em ambientes aquáticos. Os resultados foram apresentados em dois capítulos, sendo eles:

CAPÍTULO 1 - Microbial community profile and water quality in a protected area of the Caatinga biome

CAPÍTULO 2 - Functional and Taxonomic Description of the Microbial Community from Ocean Beach Tide Pool Area (San Diego, CA)

Referências Bibliográficas

1. Stolp H. Microbial ecology: organisms, habitats, activities. Cambridge University Press; 1988.

2. Xu J. INVITED REVIEW: Microbial ecology in the age of genomics and metagenomics: concepts, tools, and recent advances. Mol Ecol. 2006;15: 1713–1731. doi:10.1111/j.1365-294X.2006.02882.x

3. Shannon CE, Weaver W. The mathematical theory of communication. University of Illinois press; 2015.

4. Simpson EH. Measurement of diversity. Nature. 1949;

5. Staley JT, Konopka A. Measurement of in situ activities of nonphotosynthetic microorganisms in aquatic and terrestrial habitats. Annu Rev Microbiol. 1985;39: 321–346. doi:10.1146/annurev.mi.39.100185.001541

6. Torsvik V, Daae FL, Sandaa RA, Ovreås L. Novel techniques for analysing microbial diversity in natural and perturbed environments. J Biotechnol. 1998;64: 53–62.

7. Griffiths BS, Ritz K, Ebblewhite N, Dobson G. Soil microbial community structure: Effects of substrate loading rates. Soil Biol Biochem. 1998;31: 145–153. doi:10.1016/S0038-0717(98)00117-5

8. Banowetz GM, Whittaker GW, Dierksen KP, Azevedo MD, Kennedy AC, Griffith SM, et al. Fatty acid methyl ester analysis to identify sources of soil in surface water. J Environ Qual. 2006;35: 133–140. doi:10.2134/jeq2005.0048

9. Fakruddin M, Mannan K. Methods for analyzing diversity of microbial communities in natural environments. Ceylon J Sci Biol Sci. 2013;42.

10. Godheja J, Shekhar SK, Modi DR. Advances in Molecular Biology Approaches to Guage Microbial Communities and Bioremediation at Contaminated Sites. Int J Environ Bioremediation Biodegrad Int J Environ Bioremediation Biodegrad. 2014;2: 167–177. doi:10.12691/ijebb-2-4-4

11. Douterelo I, Boxall JB, Deines P, Sekar R, Fish KE, Biggs CA. Methodological approaches for studying the microbial ecology of drinking water distribution systems. Water Res. 2014;65: 134–156. doi:10.1016/j.watres.2014.07.008

12. Sanz JL, Köchling T. Molecular biology techniques used in wastewater treatment: An overview. Process Biochem. 2007;42: 119–133. doi:10.1016/j.procbio.2006.10.003

13. Stefanowicz A. The Biolog plates technique as a tool in ecological studies of microbial communities. Pol J Environ Stud. 2006;15: 669.

14. Handelsman J, Rondon MR, Brady SF, Clardy J, Goodman RM. Molecular biological access to the chemistry of unknown soil microbes: a new frontier for natural products. Chem Biol. 1998;5: R245–9.

15. Shizuya H, Birren B, Kim UJ, Mancino V, Slepak T, Tachiiri Y, et al. Cloning and stable maintenance of 300-kilobase-pair fragments of human DNA in Escherichia coli using an F-factor-based vector. Proc Natl Acad Sci U A. 1992;89: 8794–7.

16. Kim UJ, Birren BW, Slepak T, Mancino V, Boysen C, Kang HL, et al. Construction and characterization of a human bacterial artificial chromosome library. Genomics. 1996;34: 213–8. doi:10.1006/geno.1996.0268

17. Wild J, Hradecna Z, Szybalski W. Conditionally amplifiable BACs: switching from single-copy to high-copy vectors and genomic clones. Genome Res. 2002;12: 1434–44. doi:10.1101/gr.130502

18. Tyson GW, Chapman J, Hugenholtz P, Allen EE, Ram RJ, Richardson PM, et al. Community structure and metabolism through reconstruction of microbial genomes from the environment. Nature. 2004;428: 37–43. doi:10.1038/nature02340

19. Venter JC, Remington K, Heidelberg JF, Halpern AL, Rusch D, Eisen JA, et al. Environmental genome shotgun sequencing of the Sargasso Sea. Science. 2004;304: 66–74. doi:10.1126/science.1093857

20. DeSantis TZ, Hugenholtz P, Larsen N, Rojas M, Brodie EL, Keller K, et al. Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. Appl Environ Microbiol. 2006;72: 5069–5072. doi:10.1128/Aem.03006-05

21. Yilmaz P, Parfrey LW, Yarza P, Gerken J, Pruesse E, Quast C, et al. The SILVA and "All-species Living Tree Project (LTP)" taxonomic frameworks. Nucleic Acids Res. 2014;42: D643–D648. doi:10.1093/nar/gkt1209

22. Overbeek R, Olson R, Pusch GD, Olsen GJ, Davis JJ, Disz T, et al. The SEED and the Rapid Annotation of microbial genomes using Subsystems Technology (RAST). Nucleic Acids Res. 2014;42: D206–D214. doi:10.1093/nar/gkt1226

23. Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, Costello EK, et al. QIIME allows analysis of high-throughput community sequencing data. Nat Methods. 2010;7: 335–336. doi:10.1038/Nmeth.F.303

24. Meyer F, Paarmann D, D'Souza M, Olson R, Glass EM, Kubal M, et al. The metagenomics RAST server - a public resource for the automatic phylogenetic and functional analysis of metagenomes. BMC Bioinformatics. 2008;9. doi:10.1186/1471-2105-9-386

25. Cat&#xe3, O ECP, Lopes FAC, Ara&#xfa, Jo J, F N, et al. Soil Acidobacterial 16S rRNA Gene Sequences Reveal Subgroup Level Differences between Savanna-Like Cerrado and Atlantic Forest Brazilian Biomes, Soil Acidobacterial 16S rRNA Gene Sequences Reveal Subgroup Level Differences between Savanna-Like Cerrado and Atlantic Forest Brazilian Biomes. Int J Microbiol Int J Microbiol. 2014;2014, 2014: e156341. doi:10.1155/2014/156341, 10.1155/2014/156341

26. Behzad H, Gojobori T, Mineta K. Challenges and Opportunities of Airborne Metagenomics. Genome Biol Evol. 2015;7: 1216–1226. doi:10.1093/gbe/evv064

27. Be NA, Thissen JB, Fofanov VY, Allen JE, Rojas M, Golovko G, et al. Metagenomic Analysis of the Airborne Environment in Urban Spaces. Microb Ecol. 2015;69: 346–355. doi:10.1007/s00248-014-0517-z

28. Szalkai B, Grolmusz V. Nucleotide 9-mers characterize the type II diabetic gut metagenome. Genomics. 2016; doi:10.1016/j.ygeno.2016.02.007

29. Wang A-H, Li M, Li C-Q, Kou G-J, Zuo X-L, Li Y-Q. Human colorectal mucosal microbiota correlates with its host niche physiology revealed by endomicroscopy. Sci Rep. 2016;6: 21952. doi:10.1038/srep21952

30. Santana RH, Catão ECP, Lopes FAC, Constantino R, Barreto CC, Krüger RH. The Gut Microbiota of Workers of the Litter-Feeding Termite Syntermes wheeleri (Termitidae: Syntermitinae): Archaeal, Bacterial, and Fungal Communities. Microb Ecol. 2015;70: 545–556. doi:10.1007/s00248-015-0581-z

31. Kennedy J, Flemer B, Jackson SA, Lejon DPH, Morrissey JP, O'Gara F, et al. Marine Metagenomics: New Tools for the Study and Exploitation of Marine Microbial Metabolism. Mar Drugs. 2010;8: 608–628. doi:10.3390/md8030608

32. Arístegui J, Gasol JM, Duarte CM, Herndld GJ. Microbial oceanography of the dark ocean's pelagic realm. Limnol Oceanogr. 2009;54: 1501–1529. doi:10.4319/lo.2009.54.5.1501

33. Mueller RS, Bryson S, Kieft B, Li Z, Pett-Ridge J, Chavez F, et al. Metagenome Sequencing of a Coastal Marine Microbial Community from Monterey Bay, California. Genome Announc. 2015;3. doi:10.1128/genomeA.00341-15

34. Alves Junior N, Meirelles PM, de Oliveira Santos E, Dutilh B, Silva GGZ, Paranhos R, et al. Microbial community diversity and physical-chemical features of the Southwestern Atlantic Ocean. Arch Microbiol. 2015;197: 165–179. doi:10.1007/s00203-014-1035-6

35. Trindade-Silva AE, Rua C, Silva GGZ, Dutilh BE, Moreira APB, Edwards RA, et al. Taxonomic and Functional Microbial Signatures of the Endemic Marine Sponge Arenosclera brasiliensis. PLOS ONE. 2012;7: e39905. doi:10.1371/journal.pone.0039905

36. Rua CPJ, Trindade-Silva AE, Appolinario LR, Venas TM, Garcia GD, Carvalho LS, et al. Diversity and antimicrobial potential of culturable heterotrophic bacteria associated with the endemic marine sponge Arenosclera brasiliensis. PeerJ. 2014;2: e419. doi:10.7717/peerj.419

37. Bruce T, Meirelles PM, Garcia G, Paranhos R, Rezende CE, Moura RL de, et al. Abrolhos Bank Reef Health Evaluated by Means of Water Quality, Microbial Diversity, Benthic Cover, and Fish Biomass Data. PLOS ONE. 2012;7: e36687. doi:10.1371/journal.pone.0036687

38. Zhu Q, Li J, Ma J, Luo M, Wang B, Huang H, et al. Discovery and Engineered Overproduction of Antimicrobial Nucleoside Antibiotic A201A from the Deep-Sea Marine Actinomycete Marinactinospora thermotolerans SCSIO 00652. Antimicrob Agents Chemother. 2012;56: 110–114. doi:10.1128/AAC.05278-11

39. Horn MH, Martin KL, Chotkowski MA. Intertidal fishes: life in two worlds. Academic Press; 1998.

40. Bertness JW, Nybakken JW. Marine Biology, an Ecological Approach. Benjamin Cummings Bolch CJ Reynolds MJ 2002 Species Resolut Glob Distrib Microreticulate Dinoflag Cysts J Plankton Res. 2005;24: 565–578.

41. Bracken MES, Friberg SE, Gonzalez-Dorantes CA, Williams SL. Functional consequences of realistic biodiversity changes in a marine ecosystem. Proc Natl Acad Sci U S A. 2008;105: 924–928. doi:10.1073/pnas.0704103105

42. Altieri AH, Trussell GC, Ewanchuk PJ, Bernatchez G, Bracken MES. Consumers Control Diversity and Functioning of a Natural Marine Ecosystem. PLOS ONE. 2009;4: e5291. doi:10.1371/journal.pone.0005291

43. Longo GO, Morais RA, Martins CDL, Mendes TC, Aued AW, Cândido DV, et al. Between-Habitat Variation of Benthic Cover, Reef Fish Assemblage and Feeding Pressure on the Benthos at the Only Atoll in South Atlantic: Rocas Atoll, NE Brazil. PLoS ONE. 2015;10. doi:10.1371/journal.pone.0127176

44. Barreto FS, Schoville SD, Burton RS. Reverse genetics in the tide pool: knock-down of target gene expression via RNA interference in the copepod Tigriopus californicus. Mol Ecol Resour. 2015;15: 868–879. doi:10.1111/1755-0998.12359

45. Sorte CJB, Bracken MES. Warming and Elevated CO2 Interact to Drive Rapid Shifts in Marine Community Production. PLoS ONE. 2015;10. doi:10.1371/journal.pone.0145191

46. Sampaio HA, Martinelli-Lemos JM, Sampaio HA, Martinelli-Lemos JM. Use of intertidal areas by shrimps (Decapoda) in a brazilian Amazon estuary. An Acad Bras Ciênc. 2014;86: 333–345. doi:10.1590/0001-3765201420120039

47. Godinho WO, Lotufo TMC. Local v. microhabitat influences on the fish fauna of tidal pools in north-east Brazil. J Fish Biol. 2010;76: 487–501. doi:10.1111/j.1095-8649.2009.02501.x

48. Pfister CA. Intertidal invertebrates locally enhance primary production. Ecology. 2007;88: 1647–53. doi:10.1890/06-1913.1

49. Bracken MES, Gonzalez-Dorantes CA, Stachowicz JJ. Whole-Community Mutualism: Associated Invertebrates Facilitate a Dominant Habitat-Forming Seaweed. Ecology. 2007;88: 2211–2219. doi:10.1890/06-0881.1

50. Björk M, Axelsson L, Beer S. Why is Ulva intestinalis the only macroalga inhabiting isolated rockpools along the Swedish Atlantic coast? Mar Ecol Prog Ser. 2004;284: 109–116.

51. Jiao L, Chen H, Meng W, Lei K, Zheng B. PAHs biodegradation in intertidal surface sediment by indigenous microorganisms. Environ Sci Pollut Res. 2014;21: 6463–6471. doi:10.1007/s11356-013-2480-5

52. Yu Z, Zhou N, Qiao H, Qiu J. Identification, Cloning, and Expression of L-Amino Acid Oxidase from Marine Pseudoalteromonas sp. B3. Sci World J. 2014;2014. doi:10.1155/2014/979858

53. Rusch DB, Halpern AL, Sutton G, Heidelberg KB, Williamson S, Yooseph S, et al. The Sorcerer II Global Ocean Sampling Expedition: Northwest Atlantic through Eastern Tropical Pacific. PLOS Biol. 2007;5: e77. doi:10.1371/journal.pbio.0050077

54. Logares R, Bråte J, Heinrich F, Shalchian-Tabrizi K, Bertilsson S. Infrequent transitions between saline and fresh waters in one of the most abundant microbial lineages (SAR11). Mol Biol Evol. 2010;27: 347–357. doi:10.1093/molbev/msp239

55. Tranvik LJ, Downing JA, Cotner JB, Loiselle SA, Striegl RG, Ballatore TJ, et al. Lakes and reservoirs as regulators of carbon cycling and climate. Limnol Oceanogr. 2009;54: 2298–2314. doi:10.4319/lo.2009.54.6_part_2.2298

56. Ghai R, Rodriguez-Valera F, McMahon KD, Toyama D, Rinke R, Cristina Souza de Oliveira T, et al. Metagenomics of the water column in the pristine upper course of the Amazon river. PloS One. 2011;6: e23785. doi:10.1371/journal.pone.0023785

57. Gibbons SM, Jones E, Bearquiver A, Blackwolf F, Roundstone W, Scott N, et al. Human and environmental impacts on river sediment microbial communities. PloS One. 2014;9: e97435. doi:10.1371/journal.pone.0097435

58. Li H, Xing P, Wu QL. Characterization of the bacterial community composition in a hypoxic zone induced by Microcystis blooms in Lake Taihu, China. FEMS Microbiol Ecol. 2012;79: 773–784. doi:10.1111/j.1574-6941.2011.01262.x

59. Liu J, Fu B, Yang H, Zhao M, He B, Zhang X-H. Phylogenetic shifts of bacterioplankton community composition along the Pearl Estuary: the potential impact of hypoxia and nutrients. Front Microbiol. 2015;6: 64. doi:10.3389/fmicb.2015.00064

60. Smith MB, Rocha AM, Smillie CS, Olesen SW, Paradis C, Wu L, et al. Natural Bacterial Communities Serve as Quantitative Geochemical Biosensors. mBio. 2015;6: e00326–15. doi:10.1128/mBio.00326-15

61. S.R.H.S.H., S.R.H. Plano Diretor de Recursos Hídricos da Bacia do Médio e Baixo Paraguaçu. Salvador, BA, Brazil; 1996.

62. Pereira M do CN. Composição do Comitê da Bacia Hidrográfica do rio Paraguaçú-BA: análise da origem geográfica e do setor econômico representado por seus membros como fatores intervenientes na gestão participativa e de recursos hídricos. Universidade de Brasília. 2008.

63. Superintendência e Recursos Hídricos – SRH. 2007;

64. Ribeiro CAO. Participação Social a Gestão de Recursos Hídricos na Bahia: Estudo de caso da Bacia hidrográfica do rio Itapicuru. UFBA. 2006.

65. Velloso AL, Sampaio EVSB, Pareyn FGC. Ecorregiões: Propostas para o Bioma Caatinga. Recife: Associação Plantas do Nordeste/Instituto de Conservação Ambiental; 2002.

66. I.C.M.Bio. Plano de Manejo: Parque Nacional da Chapada Diamantina. 2007;

67. Santos PR, Durães AMM. Hidrogeologia Do Grupo Chapada Diamantina e do Grupo Paraguaçú. Rev Águas Subterrâneas. 1986;

68. Machado ABM, Martins CS, Drummond GM. Lista da Fauna Brasileira Ameaçada de Extinção: Incluindo as Espécies Quase Ameaçadas e Deficientes em Dados. Belo Horizonte: Fundação Biodiversitas; 2005.

69. Zanata AM, Sardeiro B, Zawadzki CH. A new dark-dotted species of Hypostomus Lacepede (Siluriformes: Loricariidae) from rio Paraguacu, Bahia State, Brazil. Neotropical Ichthyol. 2013;11: 247–256.

CAPÍTULO 1 [1] - Microbial community profile and water quality in a protected area of the Caatinga biome

Authors: Fabyano Alvares Cardoso Lopes[1], Elisa Caldeira Pires Catão[1], Renata Henrique Santana[2], Anderson de Souza Cabral[3], Rodolfo Paranhos[3], Thiago Pessanha Rangel[4], Carlos Eduardo de Rezende[4], Robert A. Edwards[5,6], Cristiane C. Thompson[3], Fabiano L. Thompson[3¶*], Ricardo Henrique Kruger[1¶*]

[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

[*] Corresponding authors:

[1] Capítulo publicado como artigo na PLoS One

Fabiano L. Thompson

Laboratório de Microbiologia – Instituto de Biologia,

and Laboratório de Sistemas Avançados de Gestão da Produção (SAGE) – COPPE,

Universidade Federal do Rio de Janeiro (UFRJ), Av. Carlos Chagas

Fo. S/N - CCS - IB - Lab de Microbiologia - BLOCO A (Anexo) A3 - sl 102, Cidade Universitária, Rio de Janeiro, RJ, Brazil 21941-599

E-mail: fabianothompson1@gmail.com

Ricardo H. Kruger

Universidade de Brasília – UnB

Instituto Central de Ciências Sul – Dept. de Biologia Celular

Laboratório de Enzimologia

CEP. 70910-900, Brasília, DF, Brazil.

E-mail: kruger@unb.br

¶ These authors contributed equally to this work.

Short title: National Park Protects Microbial Diversity and Water Quality

Funding: This work was supported by the Conselho Nacional de Desenvolvimento Científico e Tecnológico – CNPq – and Instituto Chico Mendes de Conservação da Biodiversidade – ICMBio – (grant number 551994/2011-8; www.cnpq.br; www.icmbio.gov.br). 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. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript. F.A.C.L. is supported by a Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) fellowship.

Abstract

The Caatinga is a semi-arid biome in northeast Brazil. The Paraguaçú 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.

Keywords: Caatinga; freshwater; bacterial community; metagenomics; pesticide; high-throughput sequencing

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 Paraguaçú 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 Paraguaçú River course, is surrounded by agricultural land, some of which is in direct contact with the Paraguaçú 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 Paraguaçú 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 Paraguaçú River (Fig. 1). The first sampling point (P1) is located outside the PNCD (13°26'9.11"S41°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).

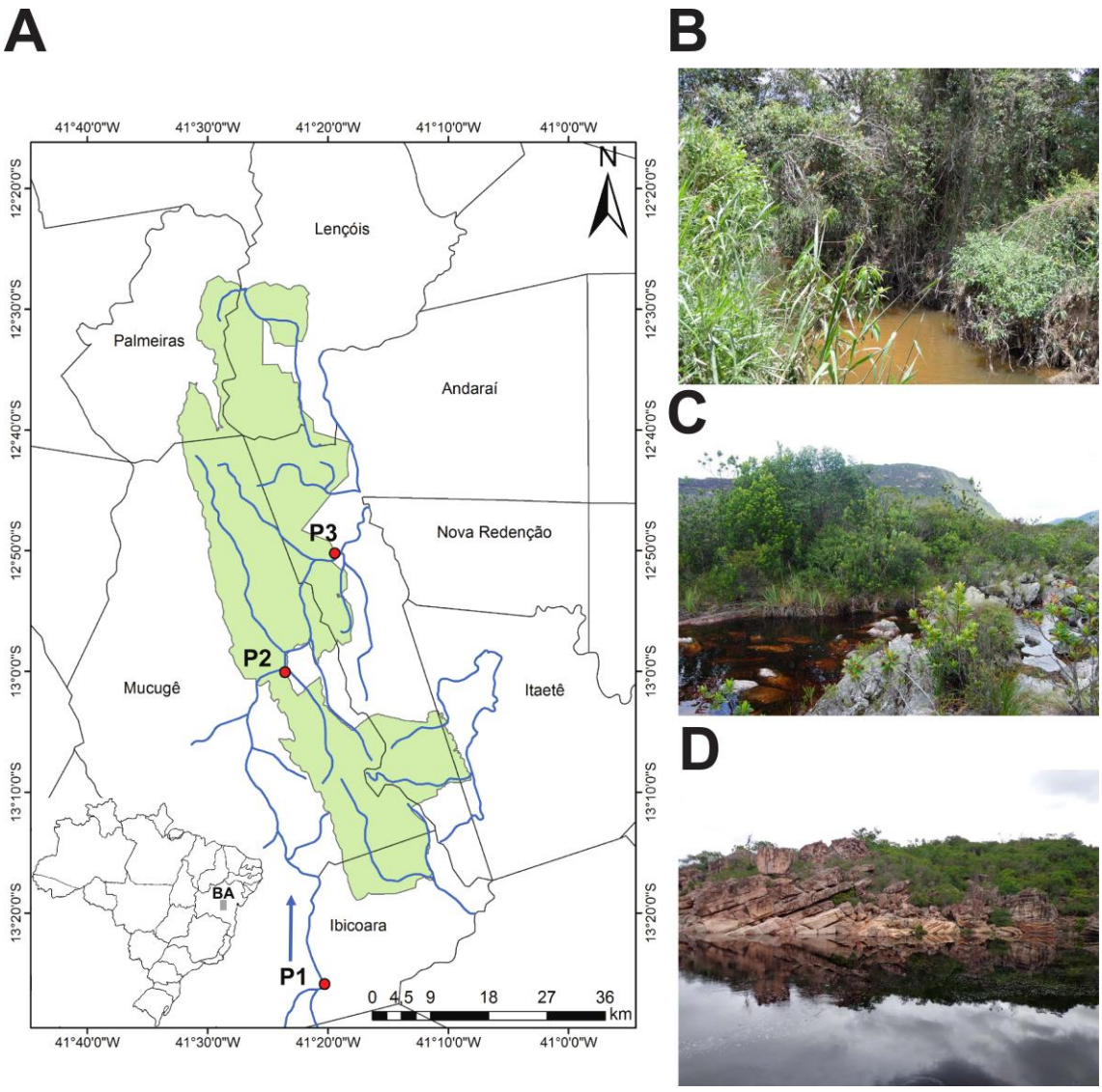
Two replicate samples were obtained at each sampling point in the wet season (November 2012) and dry season (February 2013) for a total of 12 samples. For each sampling point in both seasons, we collected approximately 20 L unfiltered water by pump from the water column at a depth of approximately 1 m.

After obtaining water samples, aliquots of each sample (200 ml) were frozen at -20ºC 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.

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 Paraguaçú 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.



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 (FACSCalibur, 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 $MgCl_2$, 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: nonparametric 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 ($\alpha < 0.05$), 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 EMPeror [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 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).

Fig. 2. Principal component analysis (PCA) of the physicochemical parameters and bacterial community of Paraguaçú River water samples obtained from three sites (P1, unprotected area; P2 and P3, protected areas) during both wet and dry seasons. The ordination analysis was performed with ammonia, nitrite, nitrate, total nitrogen, orthophosphate, total phosphorous, silicate, dissolved organic carbon (DOC), particulate organic carbon (POC), and bacterial counts based on the correlation matrix.

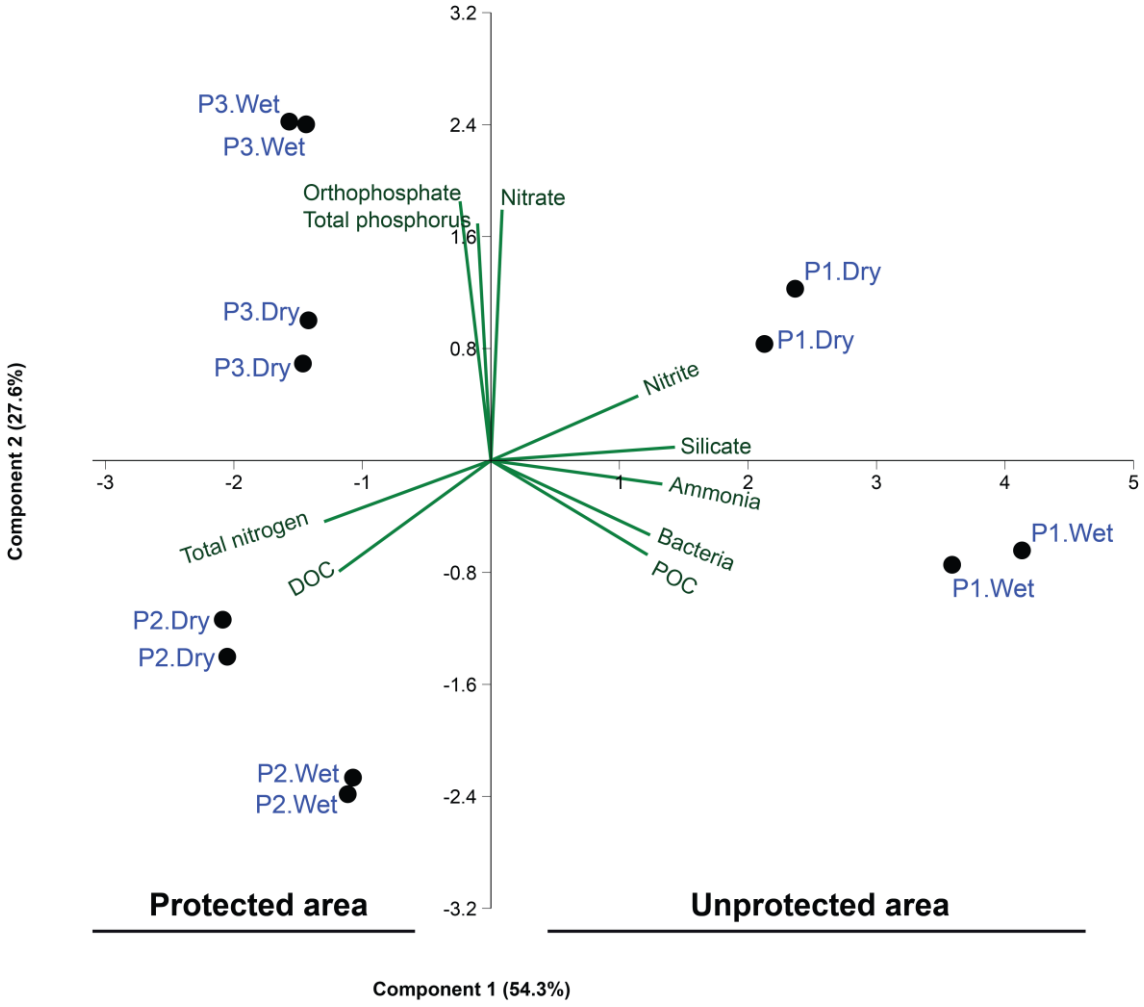


Table 1. General features of water samples obtained from sites on the Paraguaçú River.

	P1				P2				P3			
Geographic location	13°26'9.11"S		41°20'17.56"W		13° 0'2.60"S		41°23'22.57"W		12°50'25.91"S		41°19'26.52"W	
Altitude (m above sea level)	1100				974				345			
Month/Year	Nov 2012		Feb 2013		Nov 2012		Feb 2013		Nov 2012		Feb 2013	
Season	wet		dry		wet		dry		wet		dry	
Depth (m)	± 0.80		± 0.70		± 1.10		± 0.45		> 3.00		> 3.00	
Width (m)	± 4.23				± 2.94				± 36.9			
pH	2.8		4.0		2.9		3.8		3.2		4.1	
Conductivity (µS/ml)	41		44		54		63		27		35	
Coarse sand (%)	2.6		5.5		45.7		41.9		68.2		10.5	
Medium sand (%)	82.3		63.4		48.7		52.4		28.3		78.3	
Fine sand (%)	14.8		28.9		5.6		5.7		3.3		10.7	
Silt (%)	0.3		2.2		0.0		0.0		0.0		0.5	
Clay (%)	0.0		0.0		0.0		0.0		0.0		0.0	
Bacterial counts (cells/ml)	2.1E+06	± 1.1E+05	1.2E+06	± 1.9E+05	1.0E+06	± 2.0E+04	9.9E+05	± 6.8E+04	8.7E+05	± 4.3E+04	7.1E+05	± 2.4E+04
Virus counts (CFU/ml)	6.0E+05	± 4.3E+04	6.9E+06	± 1.9E+06	6.6E+05	± 1.4E+05	1.2E+07	± 5.6E+05	3.7E+05	± 1.2E+05	8.7E+06	± 4.2E+05
***Synechococcus* (cells/ml)**	0	± 0	0	± 0	0	± 0	2.2E+03	± 5.0E+02	0	± 0	0	± 0
PEUK (cells/ml)	0	± 0	0	± 0	1.6E+03	± 2.4E+02	3.1E+03	± 1.7E+02	1.6E+03	± 6.2E+01	2.0E+03	± 9.7E+01
NEUK (cells/ml)	0	± 0	0	± 0	0	± 0	0	± 0	0	± 0	0	± 0
DOC (mg/l)	0.90	± 0.06	1.25	± 0.15	2.01	± 0.14	2.88	± 0.12	1.52	± 0.18	1.95	± 0.09
POC (mg/l)	7.53	± 1.07	2.36	± 0.17	2.89	± 0.22	0.71	± 0.04	0.97	± 0.11	0.76	± 0.03
Orthophosphate (µM)	0.16	± 0.01	0.27	± 0.01	0.05	± 0.00	0.23	± 0.01	0.37	± 0.00	0.24	± 0.03
Total phosphorus (µM)	0.36	± 0.02	0.28	± 0.02	0.13	± 0.01	0.25	± 0.03	0.70	± 0.01	0.34	± 0.01
Silicate (µM)	16.03	± 0.11	11.64	± 0.98	0.66	± 0.01	0.29	± 0.04	1.35	± 0.01	0.56	± 0.01
Ammonia (µM)	3.20	± 0.10	2.55	± 0.11	1.07	± 0.05	1.53	± 0.09	0.83	± 0.06	1.56	± 0.05
Nitrite (µM)	0.31	± 0.01	0.35	± 0.01	0.28	± 0.00	0.27	± 0.01	0.28	± 0.01	0.28	± 0.00
Nitrate (µM)	3.25	± 0.15	5.77	± 0.20	0.45	± 0.01	0.69	± 0.03	6.52	± 0.11	8.03	± 0.12
Total nitrogen (µM)	28.19	± 0.76	21.25	± 0.53	43.82	± 0.16	46.65	± 1.28	42.08	± 1.25	44.15	± 0.89
N/P ratio	79.05	± 4.81	75.39	± 4.07	338.88	± 27.32	189.57	± 20.05	60.17	± 1.86	130.27	± 1.71

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

Table 2. Results of freshwater metagenomics studies.

Study	Freshwater classification	Name	Sample	Observation	Bacterial counts (cells/ml)	Depth (m)	pH	DO (mg/l)	Conductivity (mS/cm)	Salinity (PSU)	DOC (mg/l)	POC (mg/l)	Orthophosphate (µM)	Total phosphorus (µM)	Silicate (µM)	Ammonia (µM)	Nitrite (µM)	Nitrate (µM)	Total nitrogen (µM)	N/P ratio
Present study	River	Paraguaçú	**P1**	Wet season	$21.49 . 10^5$	± 0.80	2.80		0.04		0.9	7.5	0.2	0.4	16	3.2	0.3	3.3	28.2	79.1
			P2	Wet season	$10 . 10^5$	± 1.10	2.90		0.05		2	2.9	0.1	0.1	0.7	1.1	0.3	0.4	43.8	338.9
			P3	Wet season	$8.73 . 10^5$	> 3.00	3.20		0.03		1.5	1	0.4	0.7	1.4	0.8	0.3	6.5	42.1	60.2
			P1	Dry season	$11.88 . 10^5$	± 0.70	4.00	4.64	0.04		1.2	2.4	0.3	0.3	11.6	2.6	0.3	5.8	21.2	75.4
			P2	Dry season	$9.86 . 10^5$	± 0.45	3.80	4.7	0.06		2.9	0.7	0.2	0.2	0.3	1.5	0.3	0.7	46.6	189.6
			P3	Dry season	$7.13 . 10^5$	> 3.00	4.10	6.21	0.04		1.9	0.8	0.2	0.3	0.6	1.6	0.3	8	44.2	130.3
Liu et al. (2015) [47]	River	Pearl Estuary	**P01**	Surface water		1	7.01	0.22		0			4.8			139.7	14.8	78.6		
			P03	Surface water		1	7.24	3.47		0			2			21.4	24.6	102.5		
			P07	Surface water		1	6.99	0.72		0			2			34.4	43.1	105		
			P01	Bottom water		7	7.01	0.51		0			4.7			140.1	14.7	96.1		
			P03	Bottom water		3	7.21	3.14		0			2.2			30.6	26.4	103		
			P07	Bottom water		19	6.98	0.62		0.1			2			31.1	41.9	106.4		
Yan et al. (2015) [48]	River	Three Gorges Reservoir	**XXR_E**	Lacustrine system, Xiangxi River estuary			8.24	7.96	374.9				2.1	0.9		0.7		29.4	7.6	
			XXR_M	Lacustrine system, Xiangxi River midstream			8.83	12.26	332.3				1.9	1		0.8		15.3	5.3	
			WJB	Lacustrine system, Wujia Bay			9.28	19.43	266.8				0.1	0.6		0.2		4.4	3.8	
			XXR_U	Riverine system, Xiangxi River upstream			9.52	16.42	282.5				4.1	1.9		0.2		1.5	3.1	
			BSR_E	Riverine system, Baisha River estuary			9.25	15.01	266.1				6.8	3.1		0.2		1.3	3.6	
			SDR_E	Riverine system, Shendu River estuary			9.24	16.43	271.4				4.6	2.3		0.3		0.4	4.3	
Tseng et al. (2013) [49]	Reservoir	Feitsui Reservoir (FTR)	**M1**	Located in North Taiwan	$32.46 . 10^5$		9.08	7.64	0.09	0.04			0.1				0.3	13.8		
			M2	Located in North Taiwan	$34.44 . 10^5$		8.09	9.45	0.08	0.04			0.1				0.3	17.6		
			M3	Located in North Taiwan	$23.21 . 10^5$		6.52	7.99	0.06	0.03			0.1				0.1	61		
			M4	Located in North Taiwan	$45.76 . 10^5$		8.78	7.35	0.08	0.04			0				0.3	40.2		

			M5	Located in North Taiwan	$35.33 . 10^5$		9.13	6.92	0.08	0.04	0		0.3	35.8
			M6	Located in North Taiwan	$13.85 . 10^5$		6.96	6.99	0.06	0.04	0		0.1	37.6
Ghai et al. (2011) [50]	River	Solimões-Amazon		Whitewater streams with high sediment concentrations		8	7.09	5.05	109.7					
Hemme et al. (2010) [51]	Groundwater	Oak Ridge	**FW301**	Pristine groundwater			± 7.00							6.5
		Field Research Center	**FW106**	Uranium contamination	10^4–10^5	± 12.00	3.70	0.26						10046.6

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

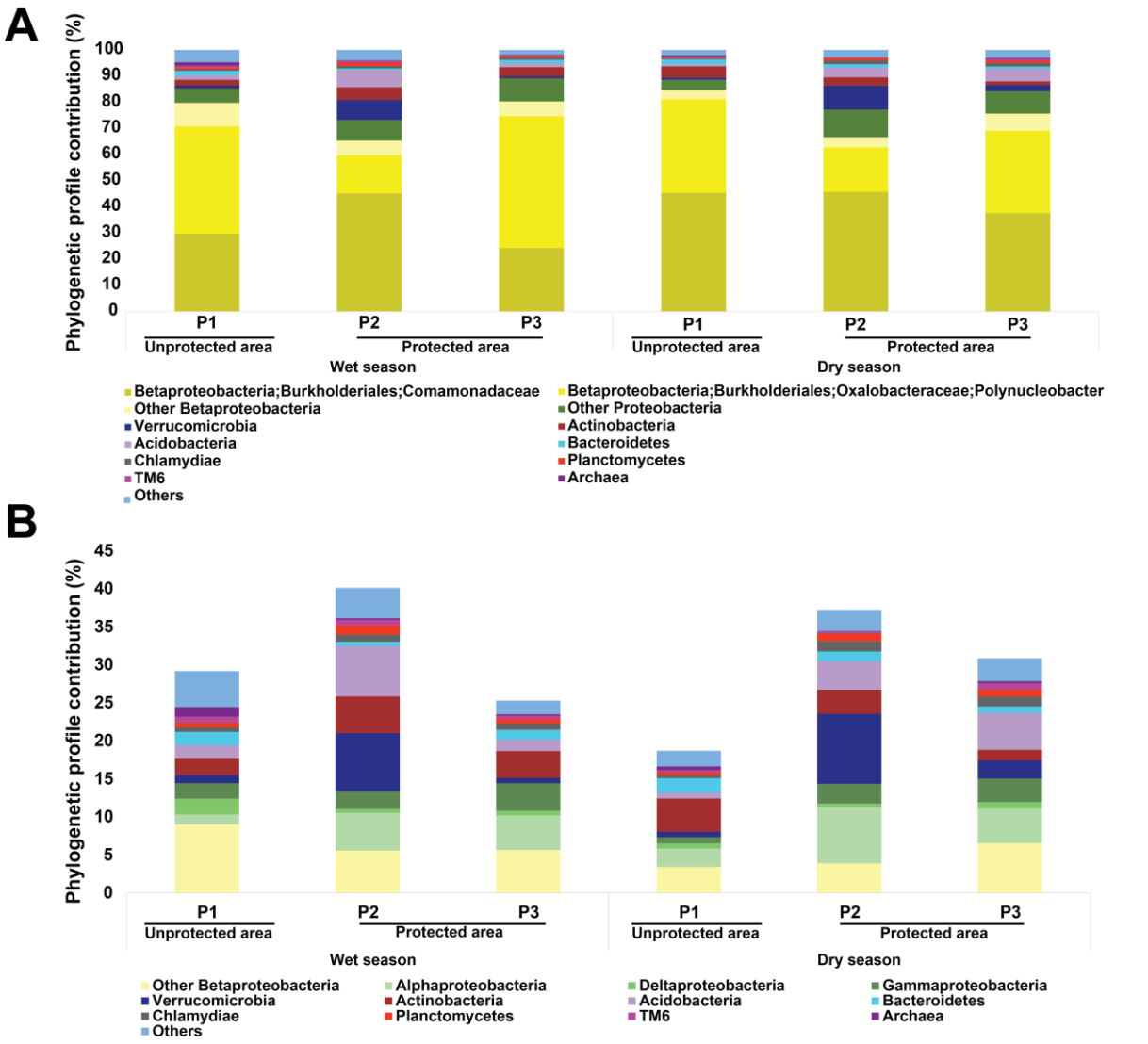
We also found that sand and silt percentages and bacterial and eukaryotic cell counts differed among sampling points (Table 1). Although none of the samples contained clay, P1 had a lower percentage of coarse sand than the other sites and was the only site with silt during both seasons. Picoeukaryotes 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 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.

Fig. 3. Taxonomic classification based on sequence analysis of 16S rRNA genes detected in Paraguaçú 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 Comamonadaceae 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).


A
Phylogenetic profile contribution (%)
100
90
80
70
60
50
40
30
20
10
0
P1
P2
P3
P1
P2
P3
Unprotected area
Protected area
Unprotected area
Protected area
Wet season
Dry season
Betaproteobacteria;Burkholderiales;Comamonadaceae
Betaproteobacteria;Burkholderiales;Oxalobacteraceae;Polynucleobacter
Other Betaproteobacteria
Other Proteobacteria
Verrucomicrobia
Actinobacteria
Acidobacteria
Bacteroidetes
Chlamydiae
Planctomycetes
TM6
Archaea
Others
B
Phylogenetic profile contribution (%)
45
40
35
30
25
20
15
10
5
0
P1
P2
P3
P1
P2
P3
Unprotected area
Protected area
Unprotected area
Protected area
Wet season
Dry season
Other Betaproteobacteria
Alphaproteobacteria
Deltaproteobacteria
Gammaproteobacteria
Verrucomicrobia
Actinobacteria
Acidobacteria
Bacteroidetes
Chlamydiae
Planctomycetes
TM6
Archaea
Others
C
P3.Wet
PC2 (14%)
P3.Dry
P3.Wet
P1.Dry
P1.Wet
P2.Wet
P3.Dry
P2.Wet
P2.Dry
P2.Dry
PC1 (21%)
P1.Wet
P1.Dry
PC3 (11%)
Protected area
Unprotected area
1. Betaproteobacteria; Burkholderiales
2. Betaproteobacteria; Methylophiales
3. Gammaproteobacteira; Xanthomonadales
4. Actinobacteria; Acidimicrobiales
5. Alphaproteobacteria; Sphingomonadales
6. Verrucomicrobia; Pedosphaerales
7. Acidobacteria; Acidobacteriales
8. Verrucomicrobia; Chthoniobacterales
9. Actinobacteria; Actinomycetales
10. Betaproteobacteria; Rhodocyclales

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 4.2% 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), Sphingomonadales (Alphaproteobacteria), Pedosphaerales (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.

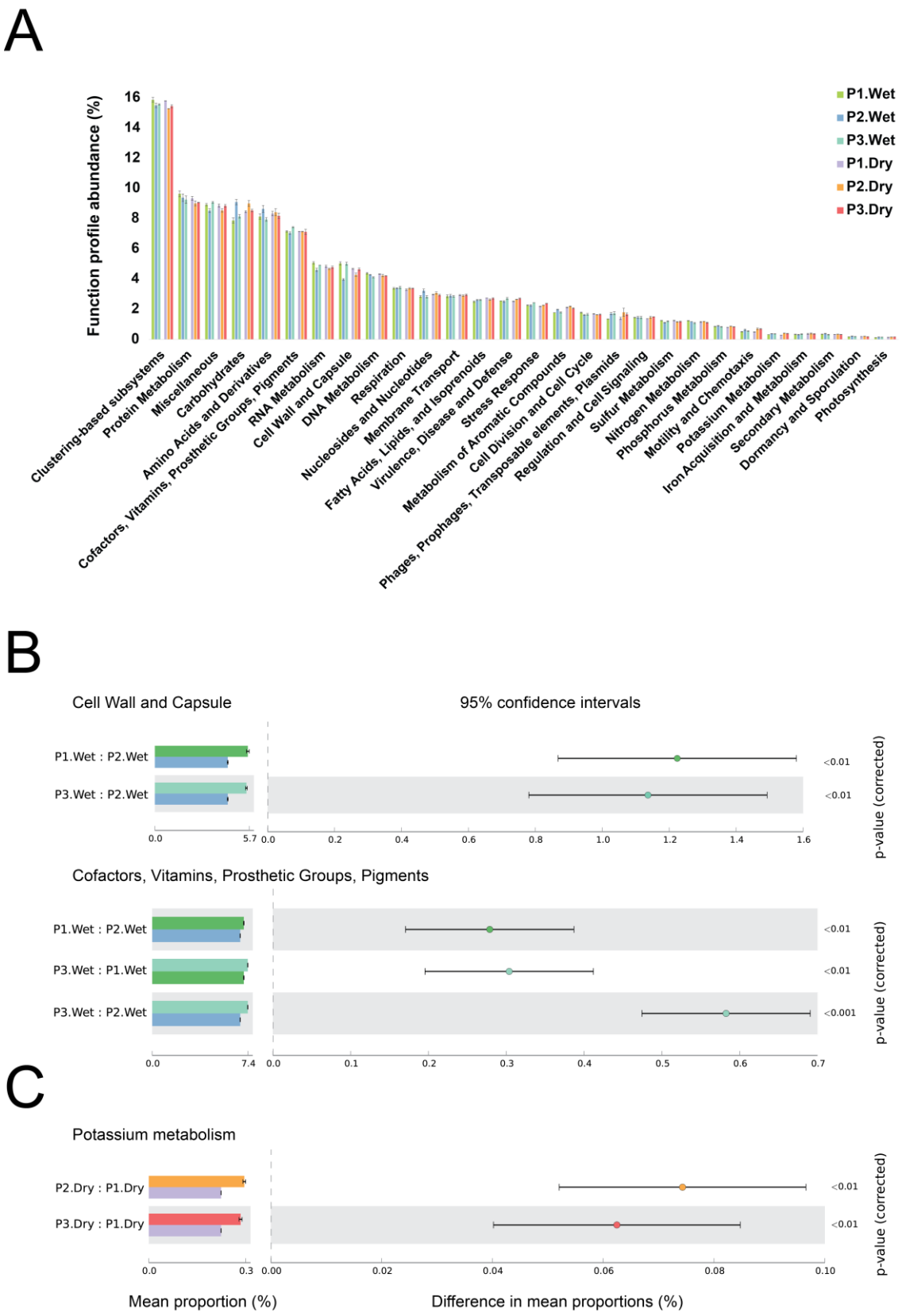
The taxonomic profile from metagenomic data (S6 Fig.) revealed that *Polynucleobacter necessaries* was the species most frequently detected, contributing 17.12% of all OTUs (S5 Table). The most abundant genera according to percentage of total OTUs were

Polynucleobacter (17.12%), *Acidovorax* (5.94%), *Burkholderia* (3.22%), *Polaromonas*, (3.07%), and *Ralstonia* (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 (1.56×10^7) 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.).

Fig. 4. Functional diversity of the Paraguaçú 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 Paraguaçú River water samples obtained in the wet season. C) Comparative analysis of functional profiles of Paraguaçú 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.



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-monooxygenase (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).

Fig.5. Comparative analysis of pesticide-related genes in Paraguaçú 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.

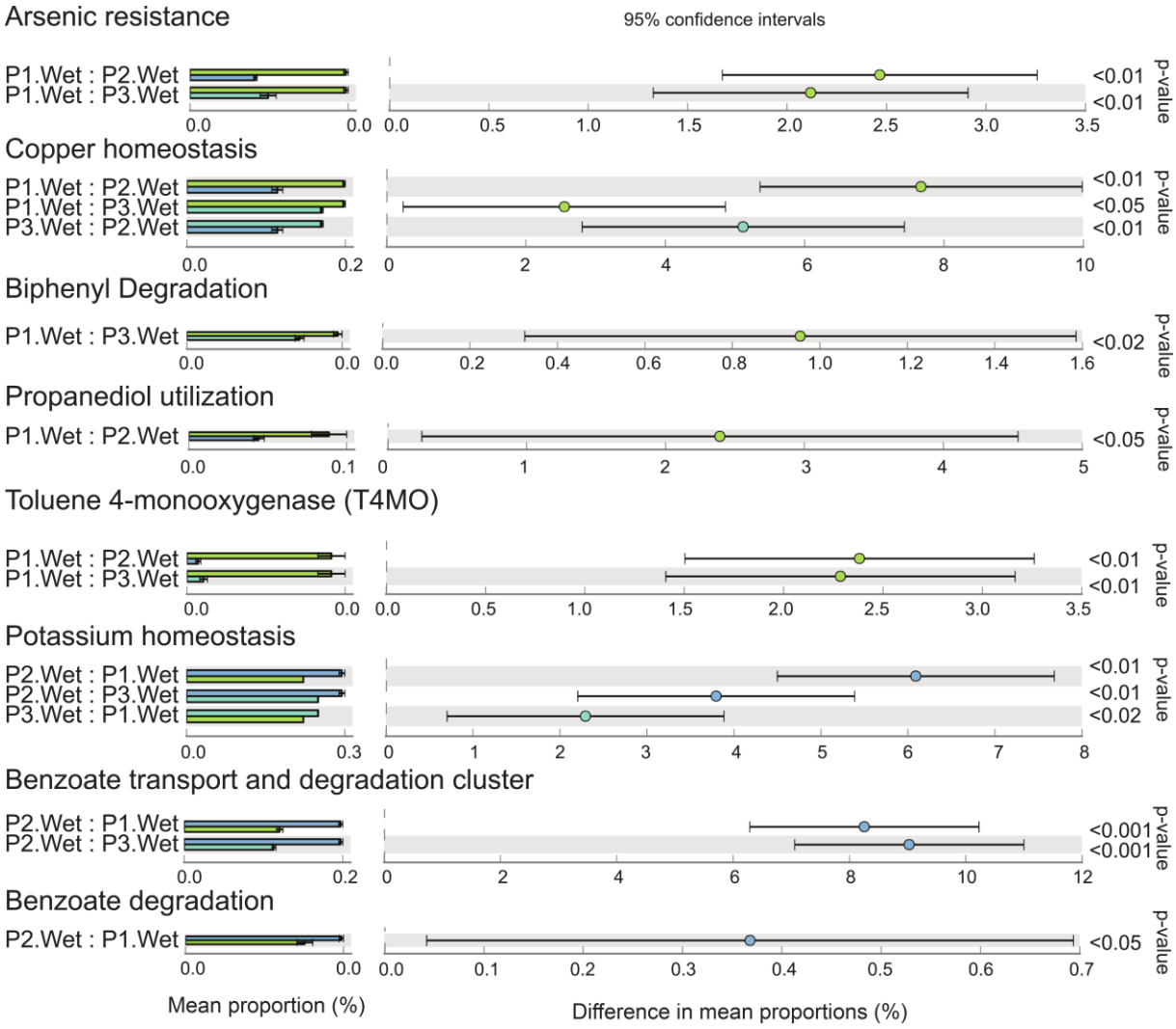
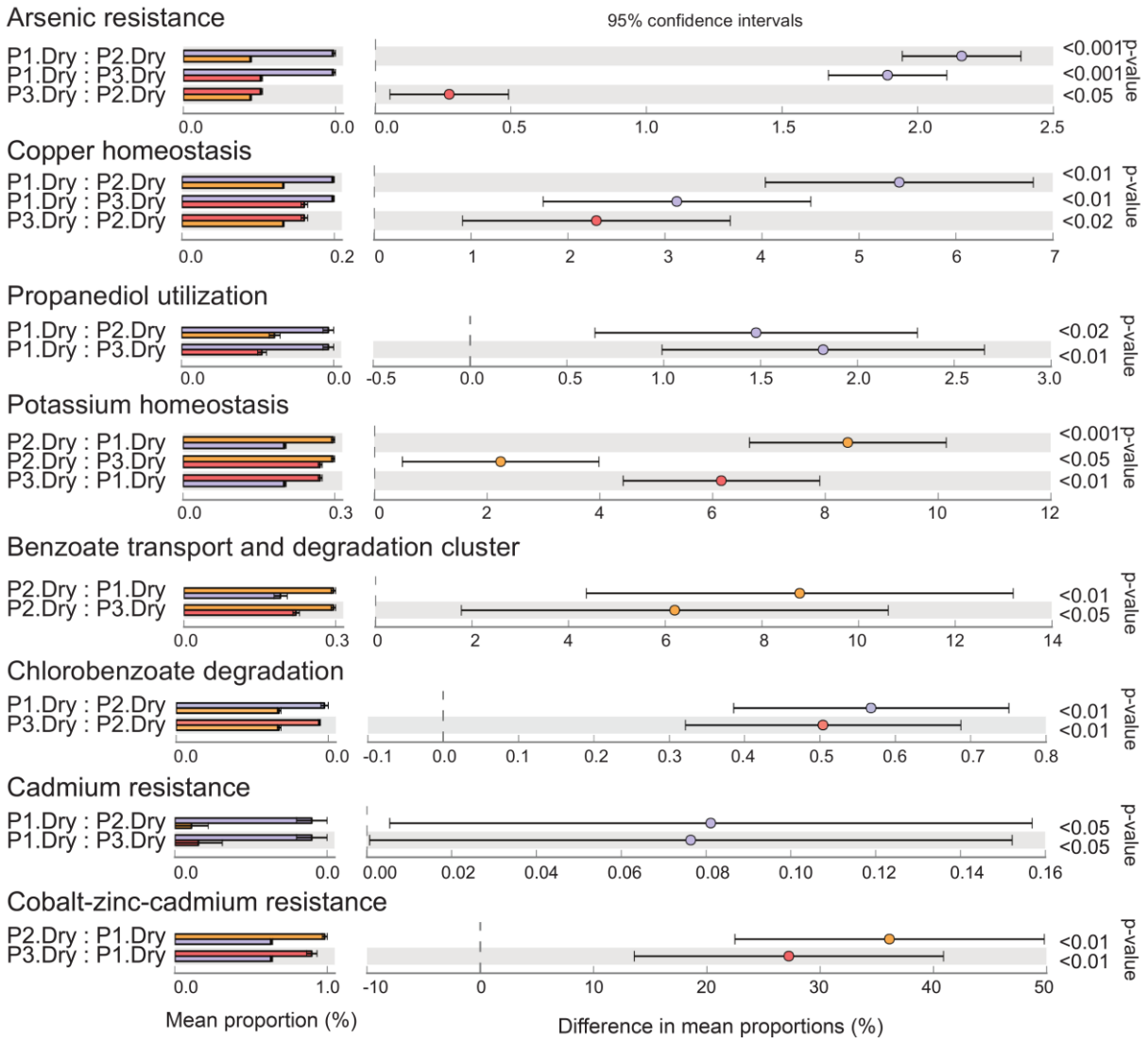


Fig. 6. Comparative analysis of pesticide-related genes in Paraguaçú River water samples obtained in the dry season. Samples were compared by ANOVA, followed by the Tukey–Kramer post hoc test ($p < 0.05$) without correction for multiple comparisons. Taxa with small effect sizes were removed by filtering (effect size = 8.00) using STAMP software.



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 Paraguaçú River. However, we cannot rule out that the abundance of this group is due to i)

runoff 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) [57]. *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 (20ºC –25ºC) [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 Paraguaçú 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 pickettiia* 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 physicochemical 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.

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.

References

1. De Albuquerque SG, Bandeira GRL. Effect of thinning and slashing on forage phytomass from a caatinga of Petrolina, Pernambuco, Brazil. Pesquisa Agropecuária Brasileira. 1995;30:885-.
2. Padro DE. As Caatingas da América do Sul. In: Leal IR, da Silva JMC, editors. Ecologia e conservação da caatinga: Universidade Federal de Pernambuco; 2003.
3. Giulietti AM, Harley R, Queiroz LP, Rapini A. To set the scene. In: Queiroz LP, Rapini A, Giulietti AM, editors. Towards greater knowledge of the Brazilian semi-arid biodiversity. Brasília: Ministério de Ciência e Tecnologia; 2006. p. 11–5.
4. MMA. Plano de ação para prevenção e controle ao desmatamento na Caatinga (versão premilinar). In: desmatamento DdPpoca, editor. Brasil: Ministério do Meio Ambiente; 2010.
5. Giongo V, Galvão SRS, Mendes AMS, Gava CAT, Cunha TJF. Soil organic carbon in theBrazilian semi-arid tropics. Dynamic soil, dynamic plant. 2011;5:12-20.
6. Menezes RS, Sampaio EV, Giongo V, Perez-Marin AM. Biogeochemical cycling in terrestrial ecosystems of the Caatinga Biome. Brazilian journal of biology = Revista brasleira de biologia. 2012;72(3 Suppl):643-53. PubMed PMID: 23011295.
7. Ribeiro CAO. Participação Social a Gestão de Recursos Hídricos na Bahia: Estudo de caso da Bacia hidrográfica do rio Itapicuru. Salvador: UFBA; 2006.
8. Gorlach-Lira K, Coutinho HDM. Population dynamics and extracellular enzymes activity of mesophilic and thermophilic bacteria isolated from semi-arid soil of northeastern Brazil. Braz J Microbiol. 2007;38(1):135-41. doi: Doi 10.1590/S1517-83822007000100028. PubMed PMID: WOS:000246576400028.
9. Monteiro JM, Vollu RE, Coelho MR, Alviano CS, Blank AF, Seldin L. Comparison of the bacterial community and characterization of plant growth-promoting rhizobacteria from different genotypes of Chrysopogon zizanioides (L.) Roberty (vetiver) rhizospheres. Journal of microbiology. 2009;47(4):363-70. doi: 10.1007/s12275-009-0048-3. PubMed PMID: 19763409.
10. Marcon AE, Ferreira DD, de Moura MDV, Campos TFC, do Amaral VS, Agnez-Lima LF, et al. Genotoxic analysis in aquatic environment under influence of cyanobacteria, metal and radioactivity. Chemosphere. 2010;81(6):773-80. doi: DOI 10.1016/j.chemosphere.2010.07.006. PubMed PMID: WOS:000283700500013.
11. Nessner Kavamura V, Taketani RG, Lanconi MD, Andreote FD, Mendes R, Soares de Melo I. Water regime influences bulk soil and Rhizosphere of Cereus jamacaru bacterial communities in the Brazilian Caatinga biome. PloS one. 2013;8(9):e73606. doi: 10.1371/journal.pone.0073606. PubMed PMID: 24069212; PubMed Central PMCID: PMC3775785.
12. Castro Wde O, Torres-Ballesteros AM, Nakayama CR, Melo IS, Pellizari VH, Silva A, et al. Draft Genome Sequence of Haloferax sp. Strain ATB1, Isolated from a Semi-Arid Region in the Brazilian Caatinga. Genome announcements. 2014;2(4). doi: 10.1128/genomeA.00812-14. PubMed PMID: 25125649; PubMed Central PMCID: PMC4132625.
13. Pacchioni RG, Carvalho FM, Thompson CE, Faustino AL, Nicolini F, Pereira TS, et al. Taxonomic and functional profiles of soil samples from Atlantic forest and Caatinga biomes in northeastern Brazil. MicrobiologyOpen. 2014;3(3):299-315. doi: 10.1002/mbo3.169. PubMed PMID: 24706600; PubMed Central PMCID: PMC4082704.
14. Marie D, Vaulot D, Partensky F. Application of the novel nucleic acid dyes YOYO-1, YO-PRO-1, and PicoGreen for flow cytometric analysis of marine prokaryotes. Applied and environmental microbiology. 1996;62(5):1649-55. PubMed PMID: WOS:A1996UJ22800026.

15. Brussaard CPD, Marie D, Bratbak G. Flow cytometric detection of viruses. J Virol Methods. 2000;85(1-2):175-82. doi: Doi 10.1016/S0166-0934(99)00167-6. PubMed PMID: WOS:000085701900018.
16. Marie D, Simon N, Vaulot D. Phytoplankton Cell Counting by Flow Cytometry. Algal Culturing Techniques. 1 ed: Academic Press; 2005.
17. Grasshoff K, Kremling K, Ehrhardt MG, Science AIoM. Methods of Seawater Analysis (3rd Edition): VCH Publishers.; 1999. 632 p.
18. Rezende CE, Pfeiffer WC, Martinelli LA, Tsamakis E, Hedges JI, Keil RG. Lignin phenols used to infer organic matter sources to Sepetiba Bay – RJ, Brasil. Estuarine, Coastal and Shelf Science. 2010;87(3):479-86. doi: http://dx.doi.org/10.1016/j.ecss.2010.02.008.
19. EPA S-. Microwave assisted acid digestion of siliceous and organically based matrices. Test Methods for Evaluating Solid Waste. 3rd ed. Washington DC: US Environmental Protection Agency; 1995.
20. Marques JSJ, Rangel TP, Brito FP, Almeida MG, Salomão MSMB, Gobo AAR, et al. Geoquímica de Metais em Sedimentos da Zona Estuarina do Complexo Industrial Porto de Suape, PE - Brasil. Revista de Gestão Costeira Integrada. 2011;11(4):379-87.
21. Gasol JM, Del Giorgio PA. Using flow cytometry for counting natural planktonic bacteria and understanding the structure of planktonic bacterial communities. Sci Mar. 2000;64(2):197-224. PubMed PMID: WOS:000088019800007.
22. Andrade L, Gonzalez AM, Araujo FV, Paranhos R. Flow cytometry assessment of bacterioplankton in tropical marine environments. Journal of microbiological methods. 2003;55(3):841-50. PubMed PMID: 14607430.
23. Hammer. Ø, Harper. DAT, Ryan. PD. Paleontological statistics software package for education and data analysis. Paleontologia Eletronica 2001;4(1):1-9.
24. R.Core.Team. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing; 2015.
25. Bruce T, Meirelles PM, Garcia G, Paranhos R, Rezende CE, de Moura RL, et al. Abrolhos bank reef health evaluated by means of water quality, microbial diversity, benthic cover, and fish biomass data. PloS one. 2012;7(6):e36687. doi: 10.1371/journal.pone.0036687. PubMed PMID: 22679480; PubMed Central PMCID: PMC3367994.
26. Santana RH, Catao EC, Lopes FA, Constantino R, Barreto CC, Kruger RH. The Gut Microbiota of Workers of the Litter-Feeding Termite Syntermes wheeleri (Termitidae: Syntermitinae): Archaeal, Bacterial, and Fungal Communities. Microbial ecology. 2015. doi: 10.1007/s00248-015-0581-z. PubMed PMID: 25749937.
27. Armougom F, Raoult D. Exploring microbial diversity using 16S rRNA high-throughput methods. Journal of Computer Science & Systems Biology 2009;2:74-92.
28. Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, Costello EK, et al. QIIME allows analysis of high-throughput community sequencing data. Nat Methods. 2010;7(5):335-6. doi: Doi 10.1038/Nmeth.F.303. PubMed PMID: WOS:000277175100003.
29. Reeder J, Knight R. Rapidly denoising pyrosequencing amplicon reads by exploiting rank-abundance distributions. Nat Methods. 2010;7(9):668-9. doi: DOI 10.1038/nmeth0910-668b. PubMed PMID: WOS:000281429200006.
30. Haas BJ, Gevers D, Earl AM, Feldgarden M, Ward DV, Giannoukos G, et al. Chimeric 16S rRNA sequence formation and detection in Sanger and 454-pyrosequenced PCR amplicons. Genome Res. 2011;21(3):494-504. doi: DOI 10.1101/gr.112730.110. PubMed PMID: WOS:000287841100015.
31. DeSantis TZ, Hugenholtz P, Larsen N, Rojas M, Brodie EL, Keller K, et al. Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. Applied and environmental microbiology. 2006;72(7):5069-72. doi: Doi 10.1128/Aem.03006-05. PubMed PMID: WOS:000238961000071.

32. Caporaso JG, Bittinger K, Bushman FD, DeSantis TZ, Andersen GL, Knight R. PyNAST: a flexible tool for aligning sequences to a template alignment. Bioinformatics. 2010;26(2):266-7. doi: DOI 10.1093/bioinformatics/btp636. PubMed PMID: WOS:000273696900018.
33. Wang Q, Garrity GM, Tiedje JM, Cole JR. Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. Applied and environmental microbiology. 2007;73(16):5261-7. doi: Doi 10.1128/Aem.00062-07. PubMed PMID: WOS:000248825900024.
34. Price MN, Dehal PS, Arkin AP. FastTree 2-Approximately Maximum-Likelihood Trees for Large Alignments. PloS one. 2010;5(3). doi: Artn E9490

Doi 10.1371/Journal.Pone.0009490. PubMed PMID: WOS:000275328800002.
35. Chao A, Shen T-J. Nonparametric estimation of Shannon's index of diversity when there are unseen species in sample. Environmental and Ecological Statistics. 2003;10(4):429-43. doi: 10.1023/A:1026096204727.
36. Chao A. Nonparametric Estimation of the Number of Classes in a Population. Scandinavian Journal of Statistics. 1984;11(4):265-70. doi: 10.2307/4615964.
37. Good IJ. The population frequencies of species and the estimation of population parameters. Biometrika. 1953;40(3-4):237-64. doi: 10.1093/biomet/40.3-4.237.
38. Lozupone C, Knight R. UniFrac: a new phylogenetic method for comparing microbial communities. Applied and environmental microbiology. 2005;71(12):8228-35. doi: Doi 10.1128/Aem.71.12.8228-8235.2005. PubMed PMID: WOS:000234417600073.
39. Vazquez-Baeza Y, Pirrung M, Gonzalez A, Knight R. EMPeror: a tool for visualizing high-throughput microbial community data. GigaScience. 2013;2(1):16. doi: 10.1186/2047-217X-2-16. PubMed PMID: 24280061.
40. Parks DH, Beiko RG. Identifying biologically relevant differences between metagenomic communities. Bioinformatics. 2010;26(6):715-21. doi: DOI 10.1093/bioinformatics/btq041. PubMed PMID: WOS:000275243500002.
41. Meyer F, Paarmann D, D'Souza M, Olson R, Glass EM, Kubal M, et al. The metagenomics RAST server - a public resource for the automatic phylogenetic and functional analysis of metagenomes. Bmc Bioinformatics. 2008;9. doi: Artn 386

Doi 10.1186/1471-2105-9-386. PubMed PMID: WOS:000260079800001.
42. Gomez-Alvarez V, Teal TK, Schmidt TM. Systematic artifacts in metagenomes from complex microbial communities. The ISME journal. 2009;3(11):1314-7. doi: 10.1038/ismej.2009.72. PubMed PMID: 19587772.
43. Cox MP, Peterson DA, Biggs PJ. SolexaQA: At-a-glance quality assessment of Illumina second-generation sequencing data. Bmc Bioinformatics. 2010;11:485. doi: 10.1186/1471-2105-11-485. PubMed PMID: 20875133; PubMed Central PMCID: PMC2956736.
44. Overbeek R, Begley T, Butler RM, Choudhuri JV, Chuang HY, Cohoon M, et al. The subsystems approach to genome annotation and its use in the project to annotate 1000 genomes. Nucleic acids research. 2005;33(17):5691-702. doi: 10.1093/nar/gki866. PubMed PMID: 16214803; PubMed Central PMCID: PMC1251668.
45. Wilke A, Harrison T, Wilkening J, Field D, Glass EM, Kyrpides N, et al. The M5nr: a novel non-redundant database containing protein sequences and annotations from multiple sources and associated tools. Bmc Bioinformatics. 2012;13:141. doi: 10.1186/1471-2105-13-141. PubMed PMID: 22720753; PubMed Central PMCID: PMC3410781.
46. Huson DH, Auch AF, Qi J, Schuster SC. MEGAN analysis of metagenomic data. Genome Res. 2007;17(3):377-86. doi: 10.1101/gr.5969107. PubMed PMID: 17255551; PubMed Central PMCID: PMC1800929.

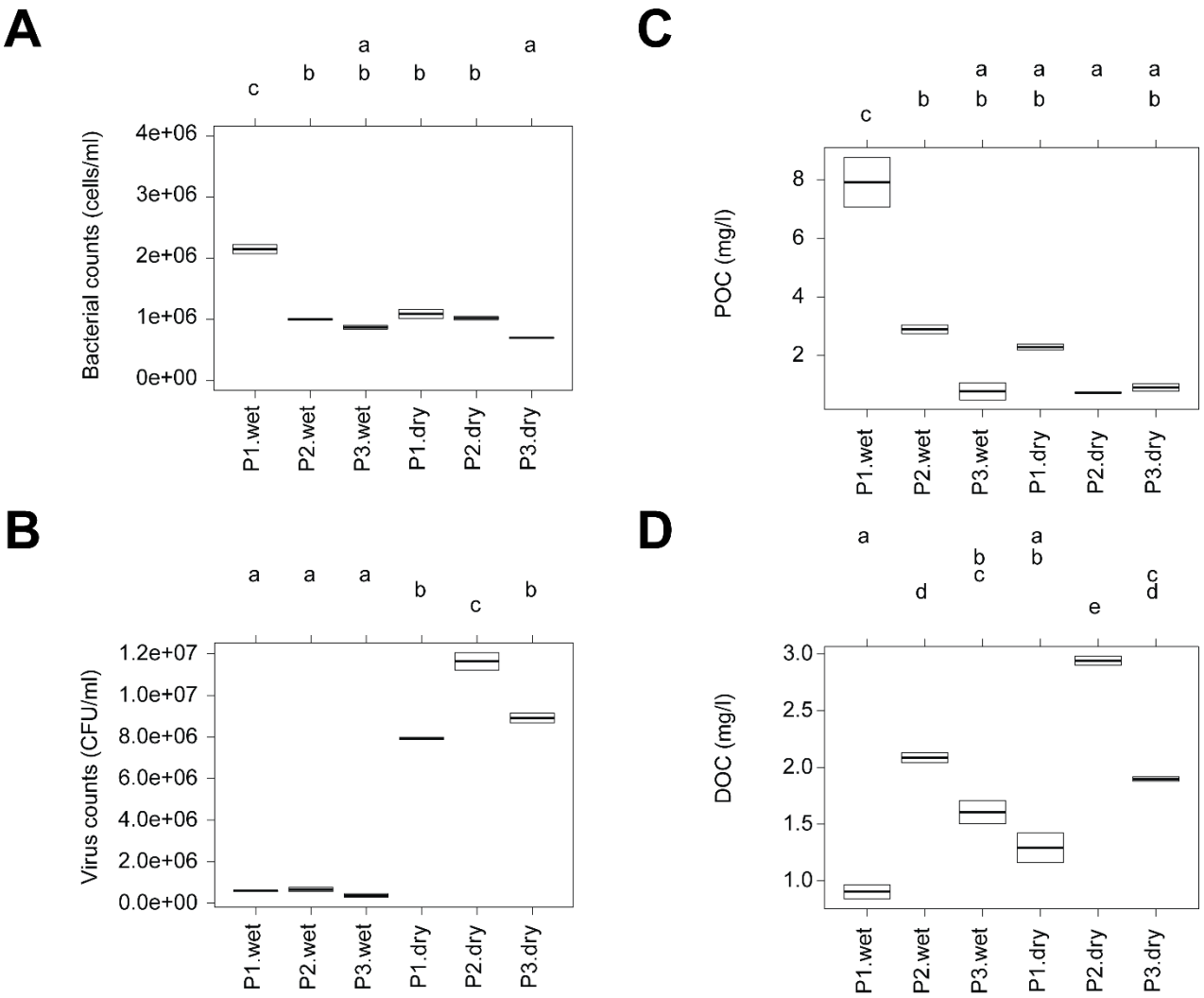
47. Liu J, Fu B, Yang H, Zhao M, He B, Zhang XH. Phylogenetic shifts of bacterioplankton community composition along the Pearl Estuary: the potential impact of hypoxia and nutrients. Frontiers in microbiology. 2015;6:64. doi: 10.3389/fmicb.2015.00064. PubMed PMID: 25713564; PubMed Central PMCID: PMC4322608.
48. Yan Q, Bi Y, Deng Y, He Z, Wu L, Van Nostrand JD, et al. Impacts of the Three Gorges Dam on microbial structure and potential function. Scientific reports. 2015;5:8605. doi: 10.1038/srep08605. PubMed PMID: 25721383; PubMed Central PMCID: PMC4342553.
49. Tseng CH, Chiang PW, Shiah FK, Chen YL, Liou JR, Hsu TC, et al. Microbial and viral metagenomes of a subtropical freshwater reservoir subject to climatic disturbances. The ISME journal. 2013;7(12):2374-86. doi: 10.1038/ismej.2013.118. PubMed PMID: 23842651; PubMed Central PMCID: PMC3834851.
50. Ghai R, Rodriguez-Valera F, McMahon KD, Toyama D, Rinke R, Cristina Souza de Oliveira T, et al. Metagenomics of the water column in the pristine upper course of the Amazon river. PloS one. 2011;6(8):e23785. doi: 10.1371/journal.pone.0023785. PubMed PMID: 21915244; PubMed Central PMCID: PMC3158796.
51. Hemme CL, Deng Y, Gentry TJ, Fields MW, Wu LY, Barua S, et al. Metagenomic insights into evolution of a heavy metal-contaminated groundwater microbial community. Isme Journal. 2010;4(5):660-72. doi: DOI 10.1038/ismej.2009.154. PubMed PMID: WOS:000277164000007.
52. van-Elsas JD, Torsvik V, Hartmann A, Øureås L, Jansson. JK. The bacteria and archaea in soil. In: Elsas JDv, Jansson JK, Trevors JT, editors. Modern soil microbiology II. Boca Raton, Florida: CRC Press; 2006. p. 83-106.
53. Compant S, Nowak J, Coenye T, Clement C, Ait Barka E. Diversity and occurrence of Burkholderia spp. in the natural environment. FEMS microbiology reviews. 2008;32(4):607-26. doi: 10.1111/j.1574-6976.2008.00113.x. PubMed PMID: 18422616.
54. Hahn MW, Lang E, Brandt U, Wu QL, Scheuerl T. Emended description of the genus Polynucleobacter and the species Polynucleobacter necessarius and proposal of two subspecies, P. necessarius subsp. necessarius subsp. nov. and P. necessarius subsp. asymbioticus subsp. nov. International journal of systematic and evolutionary microbiology. 2009;59(Pt 8):2002-9. doi: 10.1099/ijs.0.005801-0. PubMed PMID: 19567561; PubMed Central PMCID: PMC2956138.
55. Hahn MW, Lang E, Brandt U, Lunsdorf H, Wu QL, Stackebrandt E. Polynucleobacter cosmopolitanus sp. nov., free-living planktonic bacteria inhabiting freshwater lakes and rivers. International journal of systematic and evolutionary microbiology. 2010;60(Pt 1):166-73. doi: 10.1099/ijs.0.010595-0. PubMed PMID: 19648339; PubMed Central PMCID: PMC2957076.
56. Fang H, Cai L, Yang Y, Ju F, Li X, Yu Y, et al. Metagenomic analysis reveals potential biodegradation pathways of persistent pesticides in freshwater and marine sediments. The Science of the total environment. 2014;470-471:983-92. doi: 10.1016/j.scitotenv.2013.10.076. PubMed PMID: 24239819.
57. Simek K, Kasalicky V, Jezbera J, Jezberova J, Hejzlar J, Hahn MW. Broad habitat range of the phylogenetically narrow R-BT065 cluster, representing a core group of the Betaproteobacterial genus Limnohabitans. Applied and environmental microbiology. 2010;76(3):631-9. doi: 10.1128/AEM.02203-09. PubMed PMID: 19948856; PubMed Central PMCID: PMC2812994.
58. Jezbera J, Jezberova J, Koll U, Hornak K, Simek K, Hahn MW. Contrasting trends in distribution of four major planktonic betaproteobacterial groups along a pH gradient of epilimnia of 72 freshwater habitats. FEMS microbiology ecology. 2012;81(2):467-79. doi: 10.1111/j.1574-6941.2012.01372.x. PubMed PMID: 22452571; PubMed Central PMCID: PMC4268498.

59. Starliper CE. Bacterial coldwater disease of fishes caused by Flavobacterium psychrophilum. Journal of Advanced Research. 2011;2(2):97-108. doi: http://dx.doi.org/10.1016/j.jare.2010.04.001.
60. 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. PubMed PMID: 25799109; PubMed Central PMCID: PMC4370401.
61. Carpenter SR. Eutrophication of aquatic ecosystems: bistability and soil phosphorus. Proceedings of the National Academy of Sciences of the United States of America. 2005;102(29):10002-5. doi: 10.1073/pnas.0503959102. PubMed PMID: 15972805; PubMed Central PMCID: PMC1177388.
62. Hillebrand H, Gruner DS, Borer ET, Bracken ME, Cleland EE, Elser JJ, et al. Consumer versus resource control of producer diversity depends on ecosystem type and producer community structure. Proceedings of the National Academy of Sciences of the United States of America. 2007;104(26):10904-9. doi: 10.1073/pnas.0701918104. PubMed PMID: 17581875; PubMed Central PMCID: PMC1904146.
63. Sant'Anna CL, de Carvalho LR, Fiore MF, Silva-Stenico ME, Lorenzi AS, Rios FR, et al. Highly toxic Microcystis aeruginosa strain, isolated from Sao Paulo-Brazil, produce hepatotoxins and paralytic shellfish poison neurotoxins. Neurotoxicity research. 2011;19(3):389-402. doi: 10.1007/s12640-010-9177-z. PubMed PMID: 20376712.
64. Paerl H. Nutrient and other environmental controls of harmful cyanobacterial blooms along the freshwater-marine continuum. Advances in experimental medicine and biology. 2008;619:217-37. doi: 10.1007/978-0-387-75865-7_10. PubMed PMID: 18461771.
65. Fiedler S, Holl BS, Freibauer A, Stahr K, Drosler M, Schloter M, et al. Particulate organic carbon (POC) in relation to other pore water carbon fractions in drained and rewetted fens in Southern Germany. Biogeosciences. 2008;5(6):1615-23. PubMed PMID: WOS:000262411100011.
66. Kiaune L, Singhasemanon N. Pesticidal copper (I) oxide: environmental fate and aquatic toxicity. Reviews of environmental contamination and toxicology. 2011;213:1-26. doi: 10.1007/978-1-4419-9860-6_1. PubMed PMID: 21541846.
67. Wang Q-Y, Zhou D-M, Cang L. Microbial and enzyme properties of apple orchard soil as affected by long-term application of copper fungicide. Soil Biology and Biochemistry. 2009;41(7):1504-9. doi: http://dx.doi.org/10.1016/j.soilbio.2009.04.010.
68. Ryan MP, Pembroke JT, Adley CC. Ralstonia pickettii in environmental biotechnology: potential and applications. Journal of applied microbiology. 2007;103(4):754-64. doi: 10.1111/j.1365-2672.2007.03361.x. PubMed PMID: 17897177.
69. Riedel SL, Lu J, Stahl U, Brigham CJ. Lipid and fatty acid metabolism in Ralstonia eutropha: relevance for the biotechnological production of value-added products. Applied microbiology and biotechnology. 2014;98(4):1469-83. doi: 10.1007/s00253-013-5430-8. PubMed PMID: 24343766.
70. Hundt K, Wagner M, Becher D, Hammer E, Schauer F. Effect of selected environmental factors on degradation and mineralization of biaryl compounds by the bacterium Ralstonia pickettii in soil and compost. Chemosphere. 1998;36(10):2321-35. PubMed PMID: 9566302.
71. Zarcinas BA, Ishak CF, McLaughlin MJ, Cozens G. Heavy metals in soils and crops in Southeast Asia. 1. Peninsular Malaysia. Environmental geochemistry and health. 2004;26(4):343-57. doi: 10.1007/s10653-005-4669-0. PubMed PMID: 15719158.
72. Hussen A, Westbom R, Megersa N, Mathiasson L, Bjorklund E. Selective pressurized liquid extraction for multi-residue analysis of organochlorine pesticides in soil. Journal of chromatography A. 2007;1152(1-2):247-53. doi: 10.1016/j.chroma.2007.02.076. PubMed PMID: 17400237.

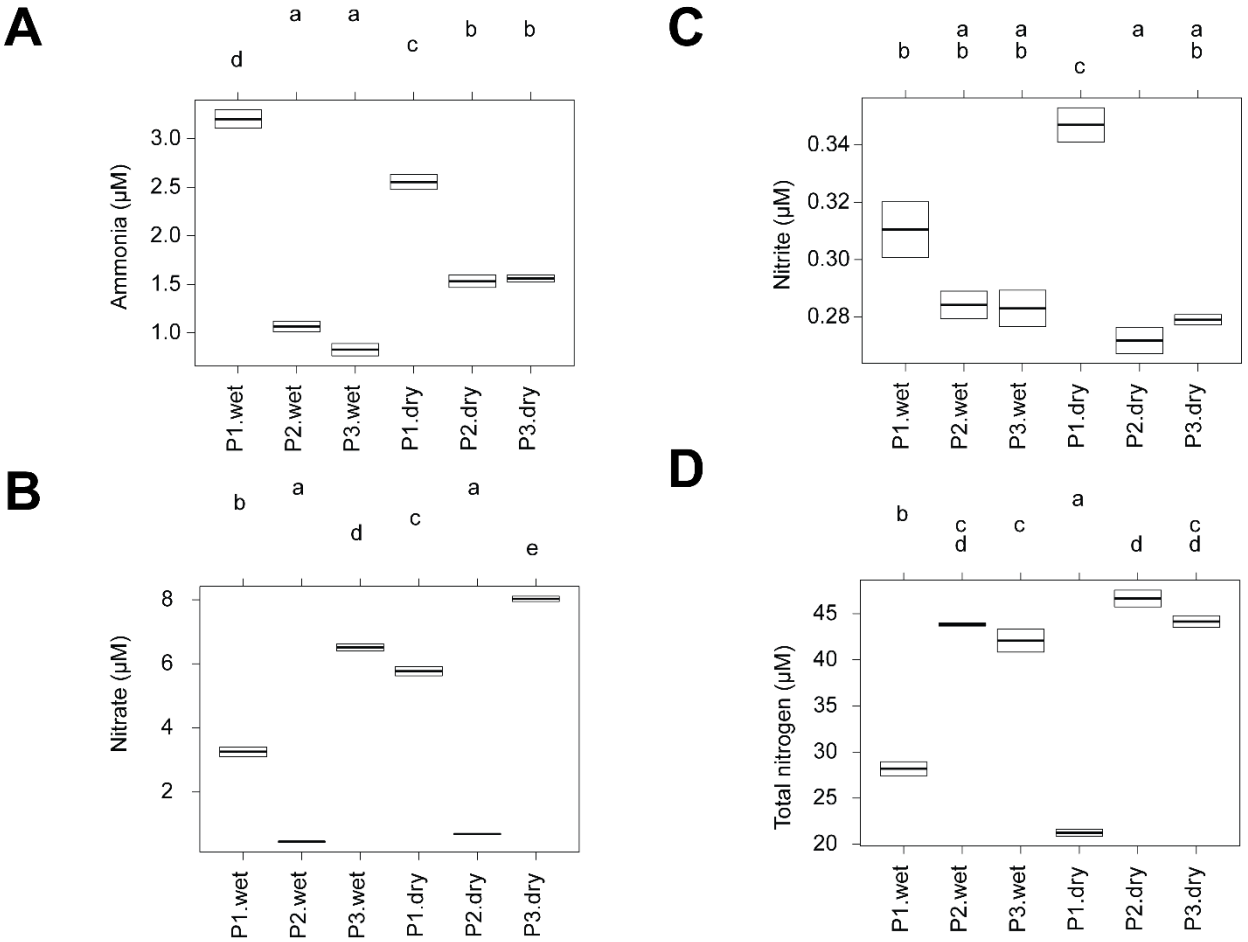
73. Shegunova P, Klanova J, Holoubek I. Residues of organochlorinated pesticides in soils from the Czech Republic. Environmental pollution. 2007;146(1):257-61. doi: 10.1016/j.envpol.2006.03.057. PubMed PMID: 17045372.
74. Zhang JY, Qiu LM, He J, Liao Y, Luo YM. Occurrence and congeners specific of polychlorinated biphenyls in agricultural soils from Southern Jiangsu, China. Journal of environmental sciences. 2007;19(3):338-42. PubMed PMID: 17918597.
75. Zhao FJ, Ma JF, Meharg AA, McGrath SP. Arsenic uptake and metabolism in plants. The New phytologist. 2009;181(4):777-94. doi: 10.1111/j.1469-8137.2008.02716.x. PubMed PMID: 19207683.
76. Cao L, Jiang H, Yang J, Fan L, Li F, Huang Q. Simultaneous Determination of Benzene and Toluene in Pesticide Emulsifiable Concentrate by Headspace GC-MS. Journal of analytical methods in chemistry. 2013;2013:121783. doi: 10.1155/2013/121783. PubMed PMID: 23607048; PubMed Central PMCID: PMC3626381.
77. Jadhav P, Skorczynki S, Skorczynksi S. Agricultural pesticide formulation. Google Patents; 2014.
78. Sagarkar S, Bhardwaj P, Storck V, Devers-Lamrani M, Martin-Laurent F, Kapley A. s-triazine degrading bacterial isolate Arthrobacter sp. AK-YN10, a candidate for bioaugmentation of atrazine contaminated soil. Applied microbiology and biotechnology. 2015. doi: 10.1007/s00253-015-6975-5. PubMed PMID: 26403923.
79. Wang Q, Xie S, Hu R. Bioaugmentation with Arthrobacter sp. strain DAT1 for remediation of heavily atrazine-contaminated soil. International Biodeterioration & Biodegradation. 2013;77:63-7. doi: http://dx.doi.org/10.1016/j.ibiod.2012.11.003.
80. Mattes TE, Alexander AK, Richardson PM, Munk AC, Han CS, Stothard P, et al. The genome of Polaromonas sp. strain JS666: insights into the evolution of a hydrocarbon- and xenobiotic-degrading bacterium, and features of relevance to biotechnology. Applied and environmental microbiology. 2008;74(20):6405-16. doi: 10.1128/AEM.00197-08. PubMed PMID: 18723656; PubMed Central PMCID: PMC2570305.
81. Boyd RA. Herbicides and herbicide degradates in shallow groundwater and the Cedar River near a municipal well field, Cedar Rapids, Iowa. The Science of the total environment. 2000;248(2-3):241-53. PubMed PMID: 10805243.
82. Fanigliulo A, Sacchetti M. Emamectin benzoate: new insecticide against Helicoverpa armigera. Communications in agricultural and applied biological sciences. 2008;73(3):651-3. PubMed PMID: 19226807.
83. Willis KJ, Ling N. The toxicity of emamectin benzoate, an aquaculture pesticide, to planktonic marine copepods. Aquaculture. 2003;221(1–4):289-97. doi: http://dx.doi.org/10.1016/S0044-8486(03)00066-8.
84. Kuo JN, Buday C, van Aggelen G, Ikonomou MG, Pasternak J. Acute toxicity of emamectin benzoate and its desmethyl metabolite to Eohaustorius estuarius. Environmental toxicology and chemistry / SETAC. 2010;29(8):1816-20. doi: 10.1002/etc.209. PubMed PMID: 20821636.
85. Zorb C, Senbayram M, Peiter E. Potassium in agriculture--status and perspectives. Journal of plant physiology. 2014;171(9):656-69. doi: 10.1016/j.jplph.2013.08.008. PubMed PMID: 24140002.

Supporting information

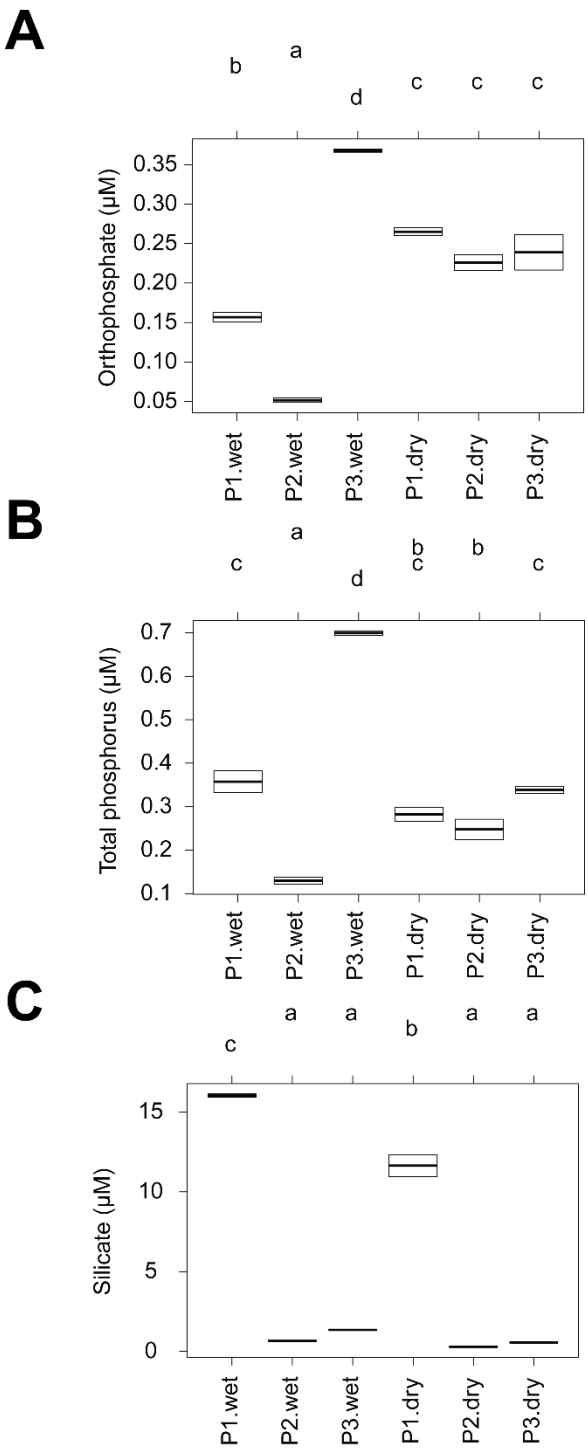
S1 Fig. Microorganism counts and organic carbon content in water samples obtained from the Paraguaçú River. A) Bacterial counts. B) Virus counts. C) Particulate organic carbon (POC). D) Dissolved organic carbon (DOC). Samples were compared by ANOVA ($\alpha < 0.05$), followed by the Tukey post hoc test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



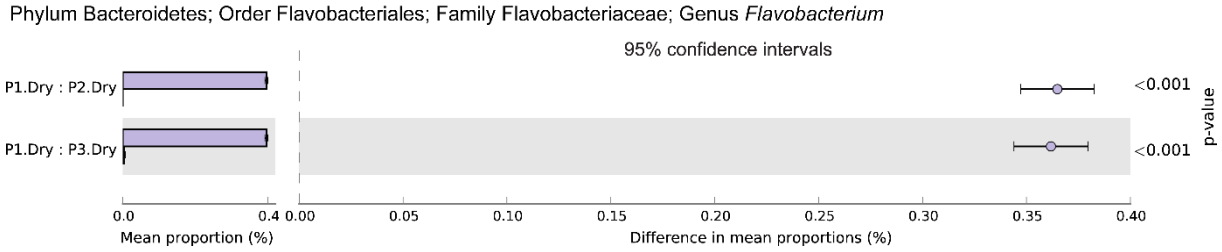
S2 Fig. Nitrogen content in water samples obtained from the Paraguaçú River. A) Ammonia. B) Nitrate. C) Nitrite. D) Total nitrogen. Samples were compared by ANOVA ($\alpha < 0.05$), followed by the Tukey post hoc test using R statistical software.



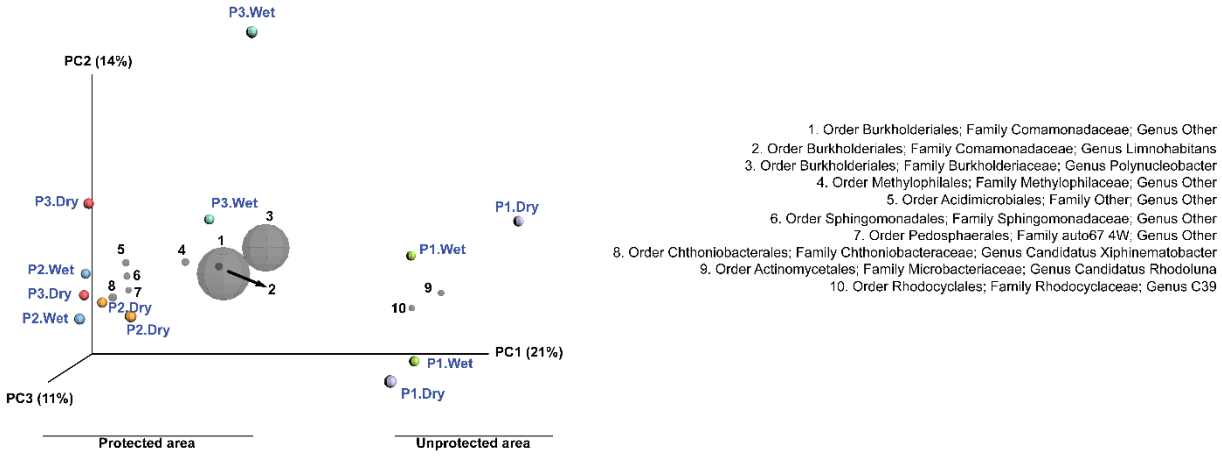
S3 Fig. Phosphorus and silicate content in water samples obtained from the Paraguaçú River. A) Orthophosphate. B) Total phosphorus. C) Silicate. Samples were compared by ANOVA ($\alpha < 0.05$), followed by the Tukey post hoc test using R statistical software.



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.



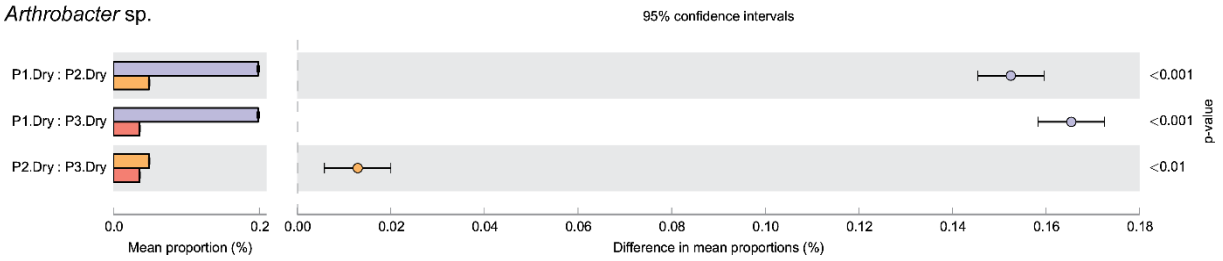
S5 Fig. Principal coordinates analysis (PCoA) of relative abundances of taxonomic groups (at the genus level). The relative abundance of each genus for each sample and their taxonomic assignments was performed using QIIME 1.7.0 software and the Greengenes database (May 2013).



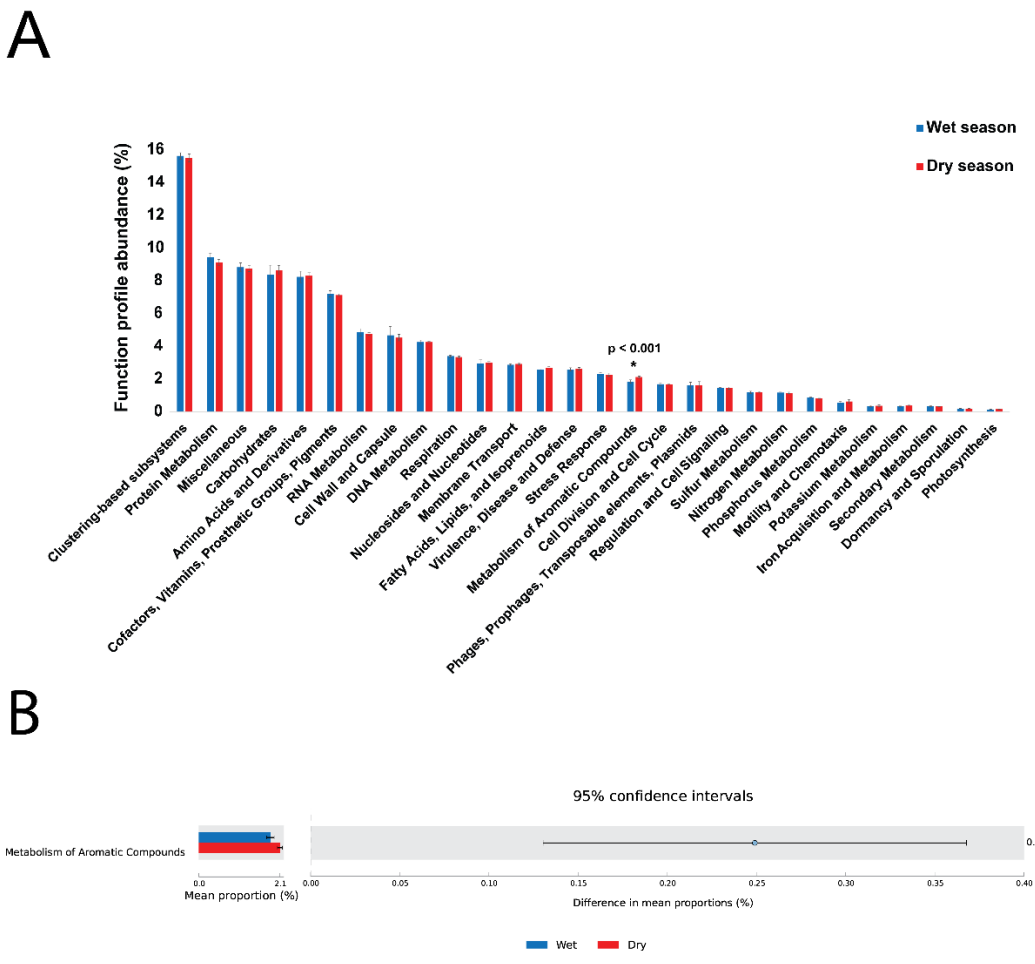
S6 Fig. Microbial community profile from metagenome sequencing. OTUs were classified at the species level in MG-RAST using the M5NR database with default sequence quality thresholds and assigned with MEGAN 5.10.6.

Figure available only online.

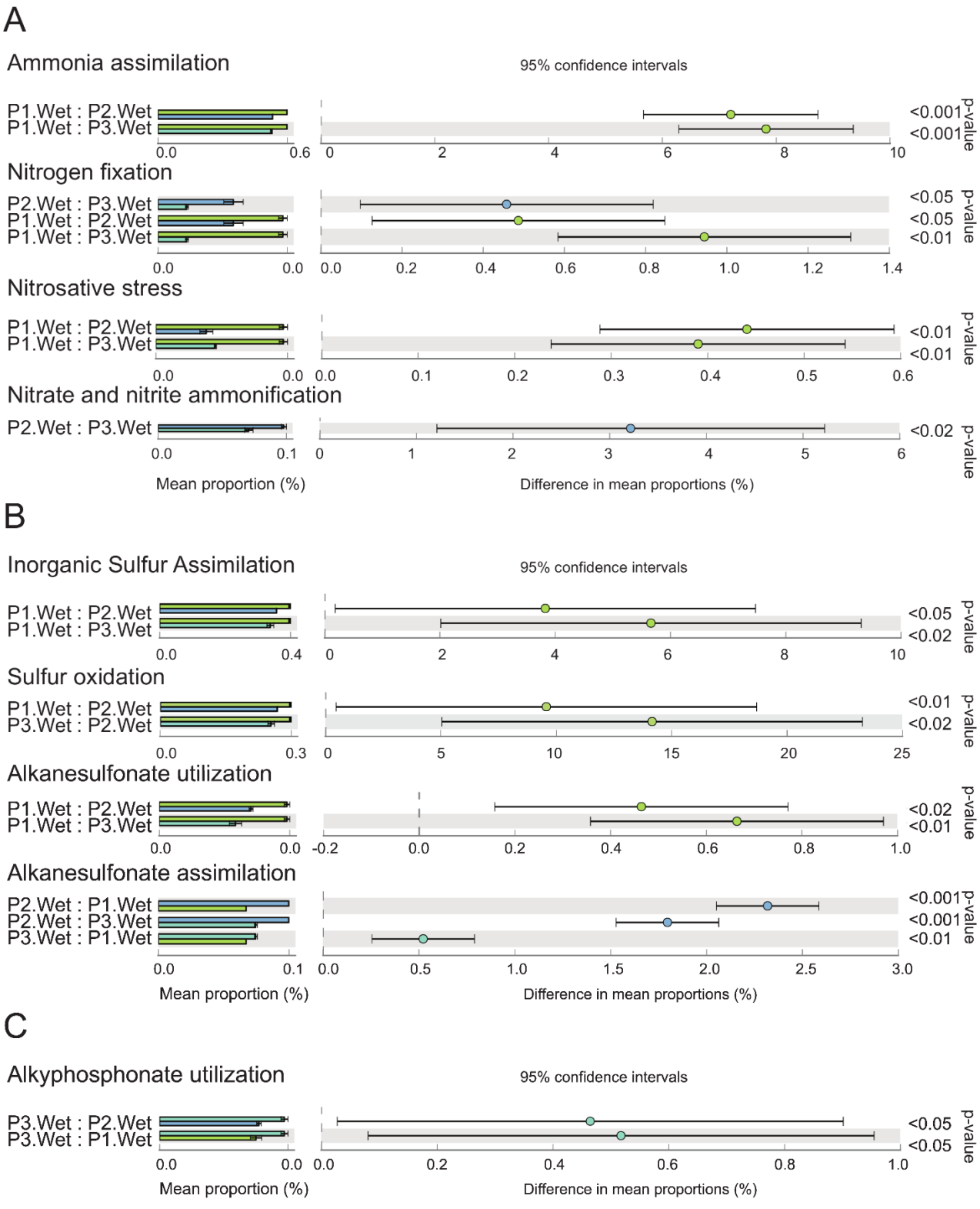
S7 Fig. Comparative analysis of metagenome taxonomic profiles of Paraguaçú River water samples obtained in the dry season. Samples were compared by ANOVA, followed by the Tukey–Kramer post hoc test ($p < 0.05$) and Bonferroni correction for multiple comparisons. Taxa with small effect sizes were removed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



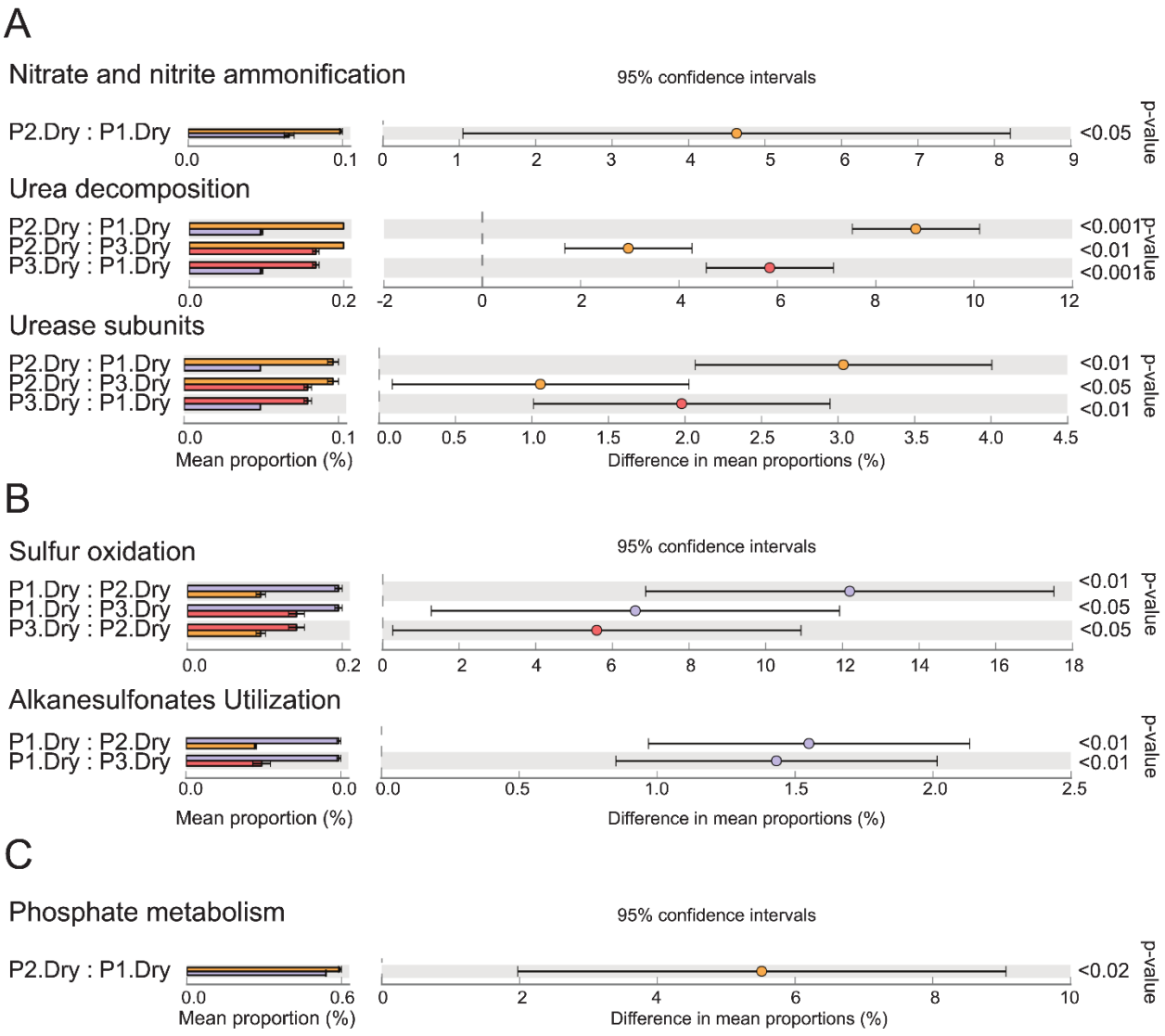
S8 Fig. Functional diversity of the Paraguaçú River's metagenomes in both wet and dry seasons. Classification was based on the SEED database level 1 in the MG-RAST server. A) Relative abundance of genes grouped by functional role. B) Comparison of genes involved in the metabolism of aromatic compounds 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.



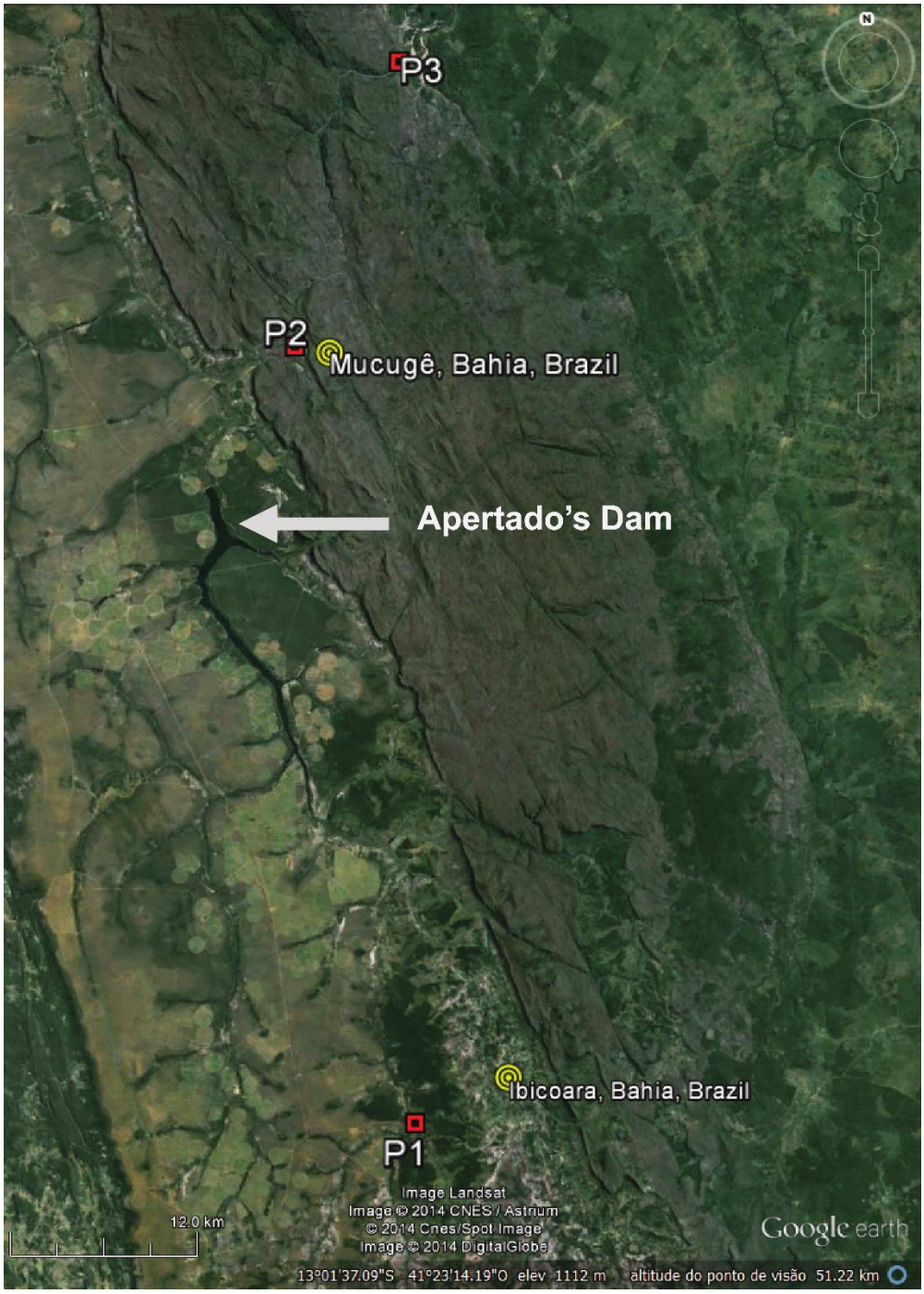
S9 Fig. Comparative analysis of functional profiles of Paraguaçú River water samples obtained in the wet season. Samples were compared by ANOVA, followed by the Tukey–Kramer post hoc test ($p < 0.05$) without correction for multiple comparisons. Taxa with small effect sizes were removed by filtering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



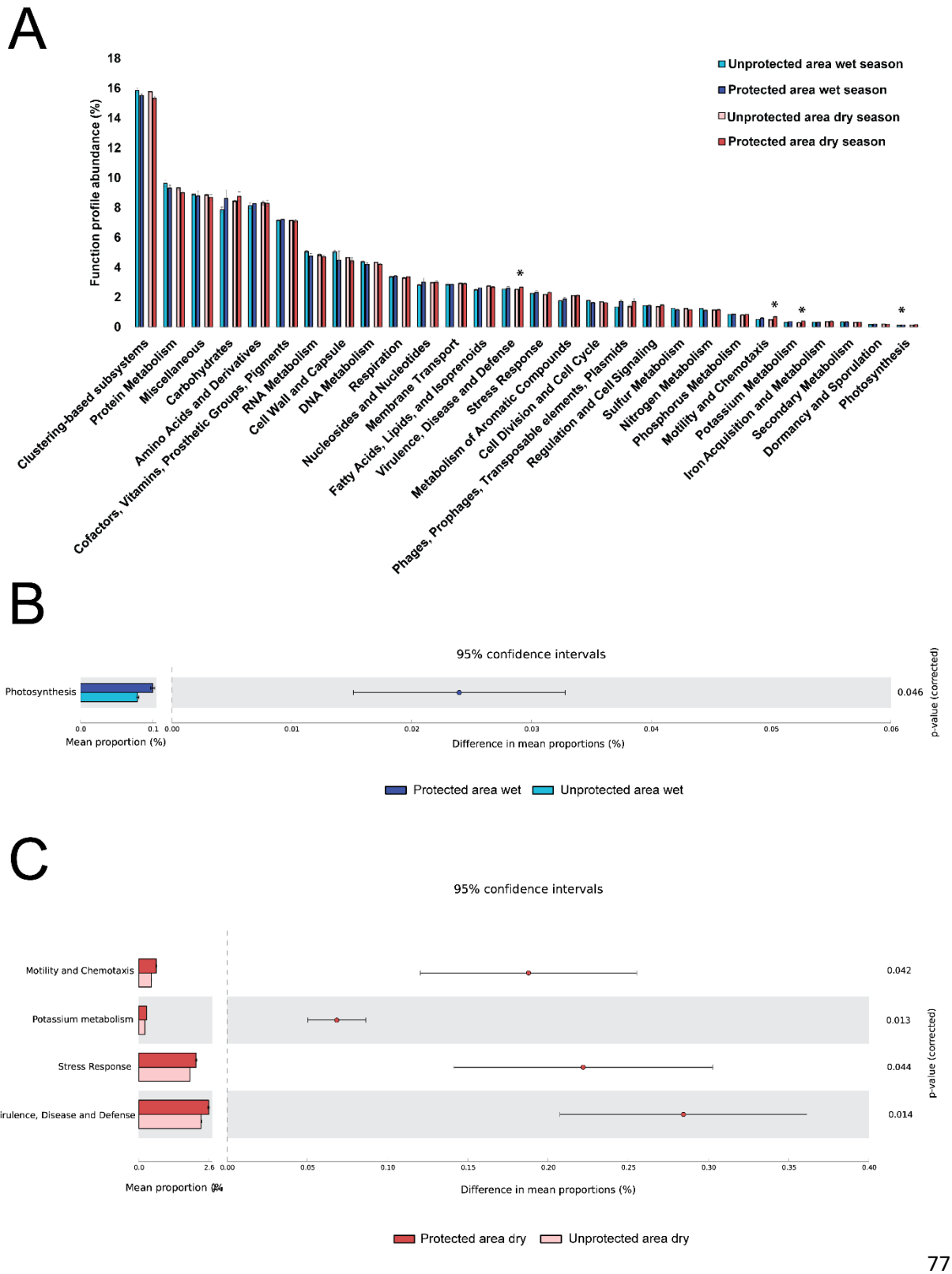
S10 Fig. Comparative analysis of functional profile of Paraguaçú River samples obtained in the dry the season. Samples were compared by ANOVA, followed by the Tukey–Kramer post hoc test ($p < 0.05$) without correction for multiple comparisons. Taxa with small effect sizes were removed by filtering (effect 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 Paraguaçú 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.

Season	P1.1 (Number of sequences)	P1.2 (Number of sequences)	P2.1 (Number of sequences)	P2.2 (Number of sequences)	P3.1 (Number of sequences)	P3.2 (Number of sequences)	Total (Number of sequences)
Wet	2.60E+06	1.00E+06	3.00E+06	1.10E+06	1.90E+06	6.20E+05	1.00E+07
Dry	3.40E+06	1.40E+06	2.70E+06	1.00E+06	2.40E+06	7.50E+05	1.20E+07
Total	-	-	-	-	-	-	2.20E+07

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

Compounds	P1						P2						P3			
	Nov 2012			Feb 2013			Nov 2012			Feb 2013			Nov 2012			Feb 2013*
Al (µg/g)	3457.96	±	33.41	6121.50	±	1851.28	2051.65	±	25.90	1143.23	±	61.71	3199.14	±	191.82	3228.818
Ba (µg/g)		ND		10.98	±	1.09		ND			ND			ND		5.507
Ca (µg/g)	51.46	±	5.17	41.12	±	8.42	29.46	±	0.76	23.94	±	4.90	58.47	±	2.78	28.521
Cd (µg/g)	0.20	±	0.03		ND		0.24	±	0.01		ND		0.33	±	0.05	ND
Cr (µg/g)	1.10	±	0.11	3.82	±	0.03	0.44	±	0.03	0.62	±	0.00	1.18	±	0.19	2.248
Cu (µg/g)	0.53	±	0.04	1.76	±	0.23		ND		0.64	±	0.08		ND		0.864
Fe (µg/g)	1082.18	±	25.86	1666.64	±	443.40	778.06	±	120.02	88.09	±	86.72	668.61	±	63.24	722.313
Mn (µg/g)		ND		5.63	±	0.07		ND		1.61	±	0.30		ND		3.693
Ni (µg/g)	1.50	±	0.18	2.49	±	0.09	1.22	±	0.22	2.13	±	0.14	1.62	±	0.09	1.966
P (µg/g)	21.06	±	2.50	16.97	±	0.91	19.73	±	1.75	10.86	±	0.02	25.96	±	2.93	23.264
Pb (µg/g)		ND		8.40	±	0.28		ND		8.05	±	0.08		ND		7.731
S (µg/g)	8.37	±	1.02	28.28	±	7.07	21.34	±	0.40	38.70	±	4.52	44.18	±	6.71	34.001
Sr (µg/g)		ND		2.02	±	0.21		ND		9.33	±	0.61	0.54	±	0.32	17.169
Ti (µg/g)	163.41	±	14.46	164.32	±	29.57	129.51	±	11.43	31.77	±	9.17	266.63	±	27.33	310.633
V (µg/g)	3.61	±	0.19	4.91	±	0.44	2.52	±	0.21	1.23	±	0.11	3.65	±	0.32	4.490
Zn (µg/g)	1.81	±	0.06	0.41	±	0.99	0.31	±	0.20	1.23	±	0.35	2.77	±	0.19	1.564

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.

Table available only online.

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

Samples	Season	OTUs observed	Chao1	Shannon index	Singletons	Doubletons	Coverage
P1		339.2 ± 22.5	910.084 ± 57.11	3.948 ± 0.21	237.4 ± 16.8	48.8 ± 3.7	0.91 ± 0.0
P2	Wet	307.65 ± 31.2	633.591 ± 75.35	4.262 ± 0.15	185.9 ± 21	52.5 ± 4.9	0.93 ± 0.0
P3		239.75 ± 37	510.443 ± 67.05	3.377 ± 0.44	146.5 ± 17.7	38.9 ± 5	0.95 ± 0.0
P1		205.7 ± 12.3	588.732 ± 13.66	3.267 ± 0.05	143.8 ± 8.3	26.9 ± 3	0.95 ± 0.0
P2	Dry	229.95 ± 27.8	465.003 ± 64.66	3.828 ± 0.23	136.4 ± 17.4	39.2 ± 4.1	0.95 ± 0.0
P3		348.2 ± 33.7	684.439 ± 31.14	4.198 ± 0.36	203.2 ± 13	61.2 ± 8.1	0.92 ± 0.0

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.

Table available only online.

CAPÍTULO 2[2] - Functional and Taxonomic Description of the Microbial Community from Ocean Beach Tide Pool Area (San Diego, CA)

Authors: Fabyano Alvares Cardoso Lopes[1], John Matthew Haggerty[2], Genivaldo Gueiros Z. Silva[3], Elisa Caldeira Pires Catão[1], Renata Henrique Santana[4], Cristiane C. Thompson[5], Fabiano L. Thompson[5], Elizabeth Ann Dinsdale[2], Ricardo Henrique Kruger[1¶*], Robert A. Edwards[3,6¶*]

[1] Cellular Biology Department, Universidade de Brasília (UnB), Brasília, DF, Brazil

[2] Biology Department, San Diego State University (SDSU), San Diego, CA, United States of America

[3] Computational Science Research Center, San Diego State University (SDSU), San Diego, CA, United States of America

[4] Genomic Sciences and Biotechnology, Universidade Católica de Brasília (UCB), Brasília DF, Brazil

[5] Institute of Biology, Universidade Federal do Rio de Janeiro (UFRJ), Rio de Janeiro, Brazil

[6] Department of Computer Science, San Diego State University (SDSU), San Diego, California, United States of America

[*] Corresponding authors:

Ricardo H. Kruger

Universidade de Brasília – UnB

Instituto Central de Ciências Sul – Dept. de Biologia Celular

Laboratório de Enzimologia

CEP. 70910-900, Brasília, DF, Brazil.

[2] Draft a ser submetido para PLoS One posteriormente.

E-mail: kruger@unb.br

Rob Edwards

San Diego State University

5500 Campanile Drive

PS 106

92182, San Diego, CA, United States of America

E-mail: redwards@cs.sdsu.edu

¶ These authors contributed equally to this work.

Short title: Functional and Taxonomic Description of Microbial Community from Tide pools

Funding: This work was supported by the Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq - www.cnpq.br). We are grateful for the support offered by Fundação de Amparo à Pesquisa do Estado do Rio de Janeiro (FAPERJ). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript. F.A.C.L. is supported by a Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) fellowship. RHK is partially supported by CAPES Senior fellowship.

Abstract

Tide pools represent unique environments that are susceptible to temperature and salinity variations, desiccation, and hypoxia. No published studies have described the microbial communities in this environment. In this work we performed the first metagenomic study of the tide pool environment, describing the taxonomic and functional characteristics of the microbial community in the Ocean Beach Tide Pool Area (San Diego, CA). Sequencing revealed that the most abundant species were *Alteromonas macleodii*, *Pseudomonas stutzeri*, and *Glaciecola psychrophila*, which belong to Gammaproteobacteria. The most abundant SEED level 1 functional categories were carbohydrates, amino acids and derivatives, and protein metabolism. No species profile pattern was found in the tide pool metagenomes. At SEED level 3 functional classification, the tide pool group (containing 15 of 19 metagenomes) is clearly distinguished, supported by a bootstrap value of 94%. The variables responsible for clustering of the tide pool group are related to stress conditions. Moreover, when the tide pool group was compared with other metagenomes (seawater and freshwater), several genes from the membrane transport, motility and chemotaxis, and stress response categories were more abundant than other groups. However, the functional diversity index was similar among these communities. Sunlight, nutrient starvation, and salinity may be crucial in determining the microbial community structure in this environment, primarily regarding the functional profile, because tide pools are disconnected from the sea for extended periods during low tide, which results in evaporation and altered nutrient recycling. Thus, these environmental stressors could be fundamental in shaping the tide pool community.

Keywords: Tide pool; stress environment; metagenomics; microbial community; functional profile.

Introduction

Tide pools represent a stressful environment that exposes its inhabitants to temperature and salinity variations, desiccation, and hypoxia [1,2]. Because tide pools are small and accessible, this environment is amenable to ecological studies [3,4]. Previous studies of tide pools generally focused on macroorganisms such as copepods, algae, shrimp, and fish [5–9].

Intertidal zones are influenced by their inhabitants. Organisms utilize nutrient sources in different ratios, directly affecting nutrient availability in tide pools [10,11]. For example, Bjork *et al.* reported that the alga *Ulva intestinalis* raises the pH in tide pools and decreases the concentration of inorganic carbon [12], which could be fundamental in shaping microbial communities. The few studies that have characterized microorganisms in tide pools focused on biotechnological applications [13,14] rather than describing the microbial communities.

The aim of this work was to perform the first metagenomic study of the tide pool environment. Specifically we described the taxonomic and functional profiles of the microbial community in the Ocean Beach Tide Pool Area (San Diego, CA). In addition, we compared the tide pool metagenomes with seawater and freshwater metagenomes to identify specific features of tide pool microorganisms.

Materials and Methods

Study area and samples collection

Microbial communities were collected from the intertidal region (tide pool area) of Ocean Beach, San Diego, CA (32.747, -117.254). Samples were collected on September 8, 2014 (water level, 11 cm) and September 11, 2014 (water level, 5 cm).

A total of 19 samples were collected from tide pools and intertidal organisms between the high tide mark and the waterline (Table 1). Samples from tide pools were captured in 0.22-µm Sterivex filters (Millipore) with 60-µL syringes. Microbes associated with tide pool organisms were collected on site using two methods. Microbes from organisms larger than 3 cm in diameter were isolated with a two-way syringe: sterile seawater was loaded into the syringe and washed over the target organism, dissociating microbes, which were aspirated back

into the syringe. Organisms smaller than 3 cm in diameter were placed into a 50-mL conical tube of sterile seawater and inverted approximately 10 times to dissociate the surface microbes. In both methods, the dissociated microbes were captured in 0.22-µm Sterivex filters, which were placed on ice and stored at -80°C until the DNA was extracted.

Table 1. Source of metagenomic DNA from the tide pool area and number of sequences obtained by metagenome sequencing. Metagenomic DNA libraries were prepared with the Nextera XT DNA Library Preparation Kit (Illumina) and 2 × 300 bp paired-end sequencing using an Illumina MiSeq.

Metagenomes	Number of sequences before applying the quality threshold	Number of sequences after applying the quality threshold
Anemone 1	5.90E+05	4.08E+05
Anemone 2	3.68E+06	3.09E+06
Brittle Star	2.03E+06	1.77E+06
Brown Algae	4.25E+06	3.01E+06
Crab 1	2.63E+05	2.31E+05
Crab 2	5.63E+06	4.26E+06
Limpet 1	1.77E+06	1.54E+06
Limpet 2	1.37E+06	1.30E+06
Mussel	1.97E+05	1.76E+05
Nudibranch	1.90E+06	1.69E+06
Sculpin	2.45E+06	2.13E+06
Seagrass 1	1.58E+06	1.40E+06
Seagrass 2	1.07E+06	7.62E+05
Sea Snail	4.07E+06	2.57E+06
Small Hermit Crab 1	4.61E+05	4.17E+05
Small Limpet	3.73E+05	3.19E+05
Small Hermit Crab 2	6.27E+04	5.51E+04
Velella	2.20E+06	1.95E+06
Water	2.21E+06	1.85E+06
Total	3.62E+07	2.89E+07

DNA extraction, metagenome sequencing, and sequence analysis

The metagenomic DNA was extracted using the NucleoSpin Tissue Kit (MACHEREY-NAGEL, Bethlehem, PA), according to the manufacturer's instructions. Metagenomic DNA

libraries were prepared with the Nextera XT DNA Library Preparation Kit (Illumina) and 2 × 300-bp paired-end sequencing using an Illumina MiSeq according to the manufacturer's instructions. The output generated by Illumina MiSeq sequencing was 3.62×10^7 sequences (Table 1).

The paired-end merge was performed using PEAR v.0.9.6 [15] using default parameters (minimum overlap size, 10; minimum possible length of the assembled sequences, 50; p-value, 0.01). Output files of each sample (PEAR assembled sequences + unassembled P1 + unassembled P2) were then merged into a single file. Quality analysis was performed using Prinseq-lite v.0.20.4 [16] with the following parameters: minimum sequence length, 75; minimum mean quality score, 30; maximum percentage of Ns, 1; trim 10 nucleotides from left; trim 10 nucleotides from right. Lastly, the metagenome data (2.89×10^7 sequences) were analyzed at the species level with functional annotation (Table 1).

Species profiles were generated by FOCUS v.0.29 [17] against all completed genomes in the RefSeq database [18] using default parameters (*k-mer* frequency, 7; minimum relative abundance, 1%). Functional annotation was performed using SUPER-FOCUS v.0.25 [19] against the SEED database [20] with the RAPSearch2 aligner [21] and default parameters (minimum identity, 60%; minimum alignment, 15 amino acids; e-value, 0.00001; database, 98%).

We performed principal component analysis (PCA) of taxonomic and functional profiles (species and SEED level 3, respectively) using a correlation matrix with the factoextra package [22] in R statistical software [23]. In addition, we performed cluster analysis for each profile (species and SEED level 3) by Euclidian distance, Unweighted Pair-Group Method with Arithmetic mean (UPGMA) algorithm, and bootstrap of 1,000,000. The cluster analysis was performed in R statistical software [23] using the pvclust package [24]. Only the clusters with a bootstrap value > 90% were selected.

The relative abundance of functional profiles (SEED level 1 and 3) between two groups was compared by Welch's t-test ($p < 0.05$), asymptotic confidence intervals (95%), and the Bonferroni correction in STAMP software [25]. Classification labels with small effect sizes were removed by filtering (difference between proportions < 1.0, or ratio of proportions < 2.00 for level 1 and < 4.00 for level 3). All outliers were removed in these analyses.

Functional and taxonomic diversity of tide pools

The diversity index was calculated in R software [23] using the vegan package [26]. We calculated the Shannon–Wiener index (H') [27] for taxonomic and functional profiles (species and SEED level 3, respectively) for each group described in the SEED level 3 cluster analysis. The indices of all groups analyzed (both profiles) were compared by factorial ANOVA ($p < 0.05$), followed by the Tukey post hoc test in R software [23].

Other metagenome data

To detect specific features of the tide pool microbial community, we compared Pacific seawater metagenomes from Sunagawa *et al.* [28] (S1 Table) with the tide pool metagenomes. Metagenomes from three different water layers were used for this comparison: surface water layer (SRF), deep chlorophyll maximum layer (DCM), and mesopelagic zone (MES). Four freshwater metagenomes from the Paraguaçú River in the wet season (two metagenomes) and dry season (two metagenomes) from Lopes *et al.* [29] were used as outgroups (S1 Table).

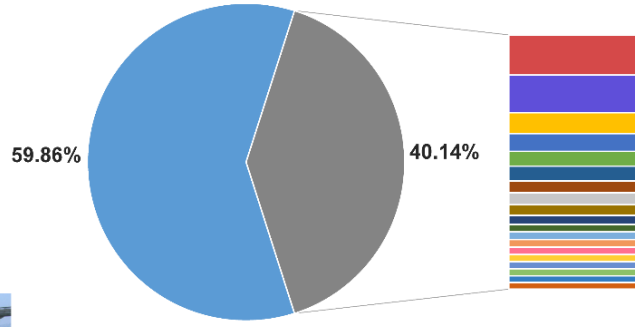
Results

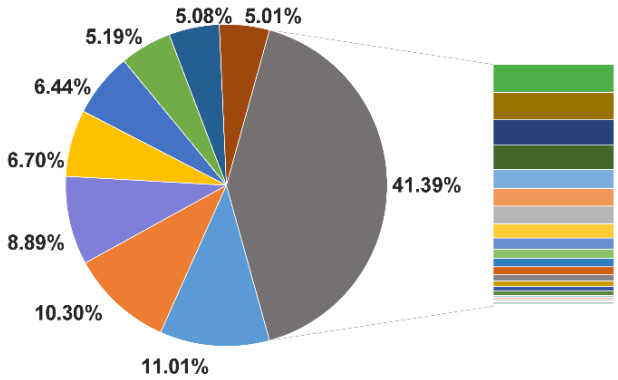
Tide pool microbial profiles

Results of metagenomic DNA sequencing revealed the most common species and functional categories in the tide pool samples (Fig. 1; S2 and S3 Tables). The relative abundance of *Alteromonas macleodii* was the highest among all species annotated (6.31%). The following five species also showed a relative abundance > 2.00%: *Pseudomonas stutzeri* (5.96%), *Glaciecola psychrophila* (3.30%), *Staphylococcus pseudintermedius* (2.82%), *Flavobacterium branchiophilum* (2.33%), and *Leptospira biflexa* (2.30%) (Fig. 1 and S2 Table). The sum of the relative abundances of the remaining species was 76.97%. Regarding functional annotation (Fig. 1 and Table S3), most sequences were classified into one of the following categories: carbohydrates (11.01%); amino acids and derivatives (10.30%); protein metabolism (8.89%); cofactors, vitamins, prosthetic groups, pigments (6.70%); DNA metabolism (5.19%); membrane transport (5.08%); or respiration (5.01%). Proteins with unknown function accounted for approximately 6.44% of the annotated sequences. The sum of the relative abundances of other SEED level 1 categories was 41.39%. SEED level 1 classification for each tide pool metagenome showed that the most abundant categories were carbohydrates, amino

acids and derivatives, and protein metabolism in all metagenomes (S1 Fig.); however, Limpet 1 had a standard profile that differed from the other metagenomes, with protein metabolism as the most abundant category.

Fig. 1. Species and functional classification based on all tide pool metagenomes. Species classification was performed in FOCUS v.0.29 against all completed genomes in the RefSeq database using default parameters. Functional classification was based on the SEED database level 1 in SUPER-FOCUS v.0.25 using RAPSearch2 aligner and default parameters.





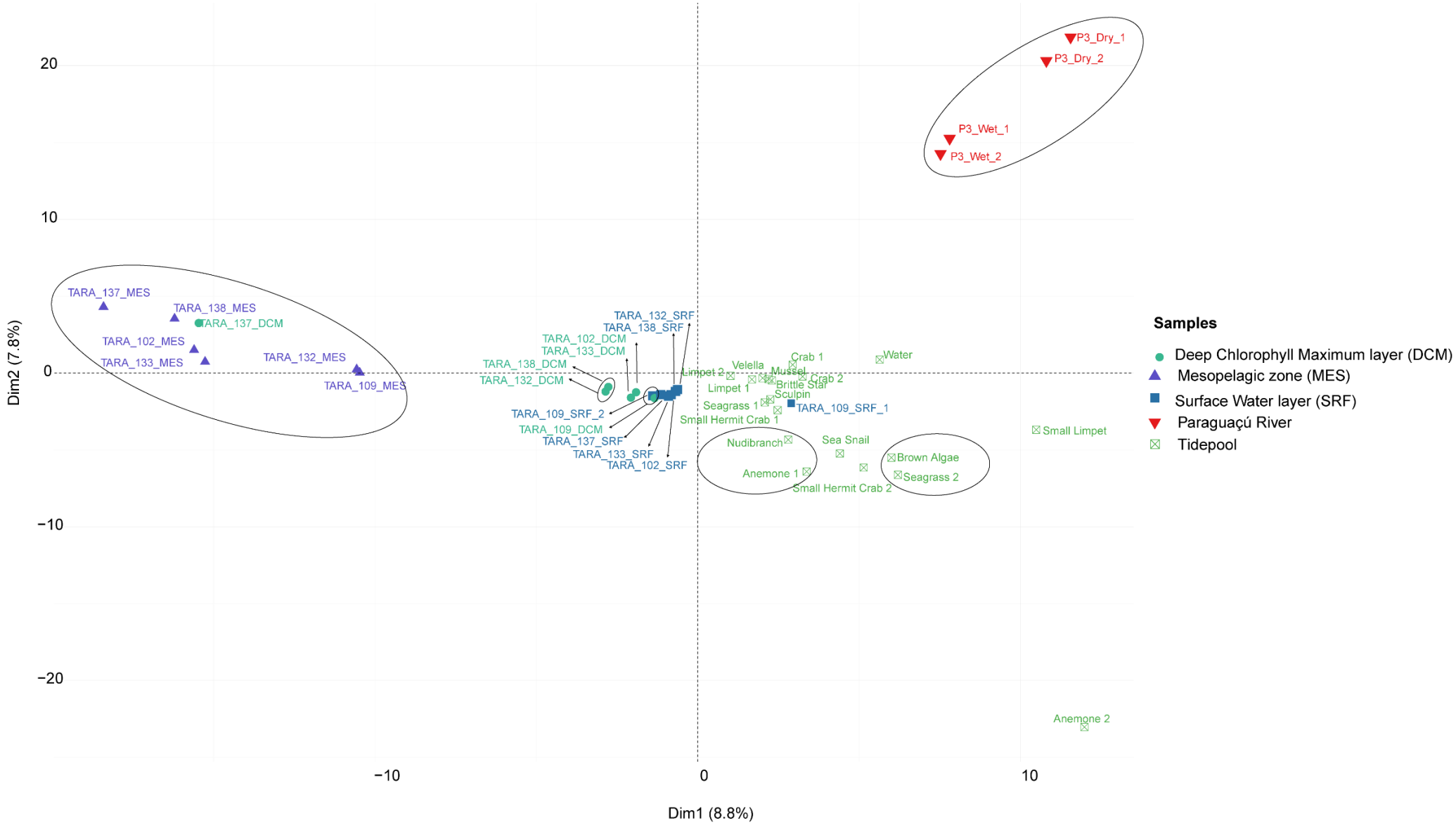
Species	Relative Abundance (%)
Others	59.86
Alteromonas macleodii	6.31
Pseudomonas stutzeri	5.96
Glaciecola psychrophila	3.30
Staphylococcus pseudintermedius	2.82
Flavobacterium branchiophilum	2.33
Leptospira biflexa	2.30
Candidatus *Nitrosoarchaeum koreensis*	1.89
Pseudoalteromonas atlantica	1.89
Anaplasma marginale	1.70
Octadecabacter antarcticus	1.40
Leuconostoc gelidum	1.20
Alcanivorax borkumensis	1.20
Hirschia baltica	1.20
Candidatus *Portiera aleyrodidarum*	1.16
Rhodopirellula baltica	1.14
Haloquadratum walsbyi	1.12
Pyrobaculum islandicum	1.10
Coprothermobacter proteolyticus	1.07
Methanobacterium sp. AL-21	1.04

Level 1 classification	Relative Abundance (%)
Carbohydrates	11.01
Amino Acids and Derivatives	10.30
Protein Metabolism	8.89
Cofactors, Vitamins, Prosthetic Groups, Pigments	6.70
Clustering-based subsystems	6.44
DNA Metabolism	5.19
Membrane Transport	5.08
Respiration	5.01
Stress Response	4.80
Cell Wall and Capsule	4.59
Virulence	4.45
RNA Metabolism	4.16
Fatty Acids, Lipids, and Isoprenoids	3.22
Nucleosides and Nucleotides	3.05
Miscellaneous	3.01
Motility and Chemotaxis	2.43
Regulation and Cell signaling	1.90
Nitrogen Metabolism	1.57
Phosphorus Metabolism	1.49
Sulfur Metabolism	1.38
Potassium metabolism	1.06
Metabolism of Aromatic Compounds	0.96
Cell Division and Cell Cycle	0.77
Iron acquisition and metabolism	0.71
Phages, Prophages, Transposable elements, Plasmids	0.36
Predictions based on plant-prokaryote comparative analysis	0.35
Virulence, Disease and Defense	0.23
Central metabolism	0.23
Photosynthesis	0.21
Transcriptional regulation	0.20
Dormancy and Sporulation	0.17
Secondary Metabolism	0.10
Plant cell walls and outer surfaces	0.00
Arabinose Sensor and transport module	0.00

Clustering of tide pool, sea, and freshwater metagenomes

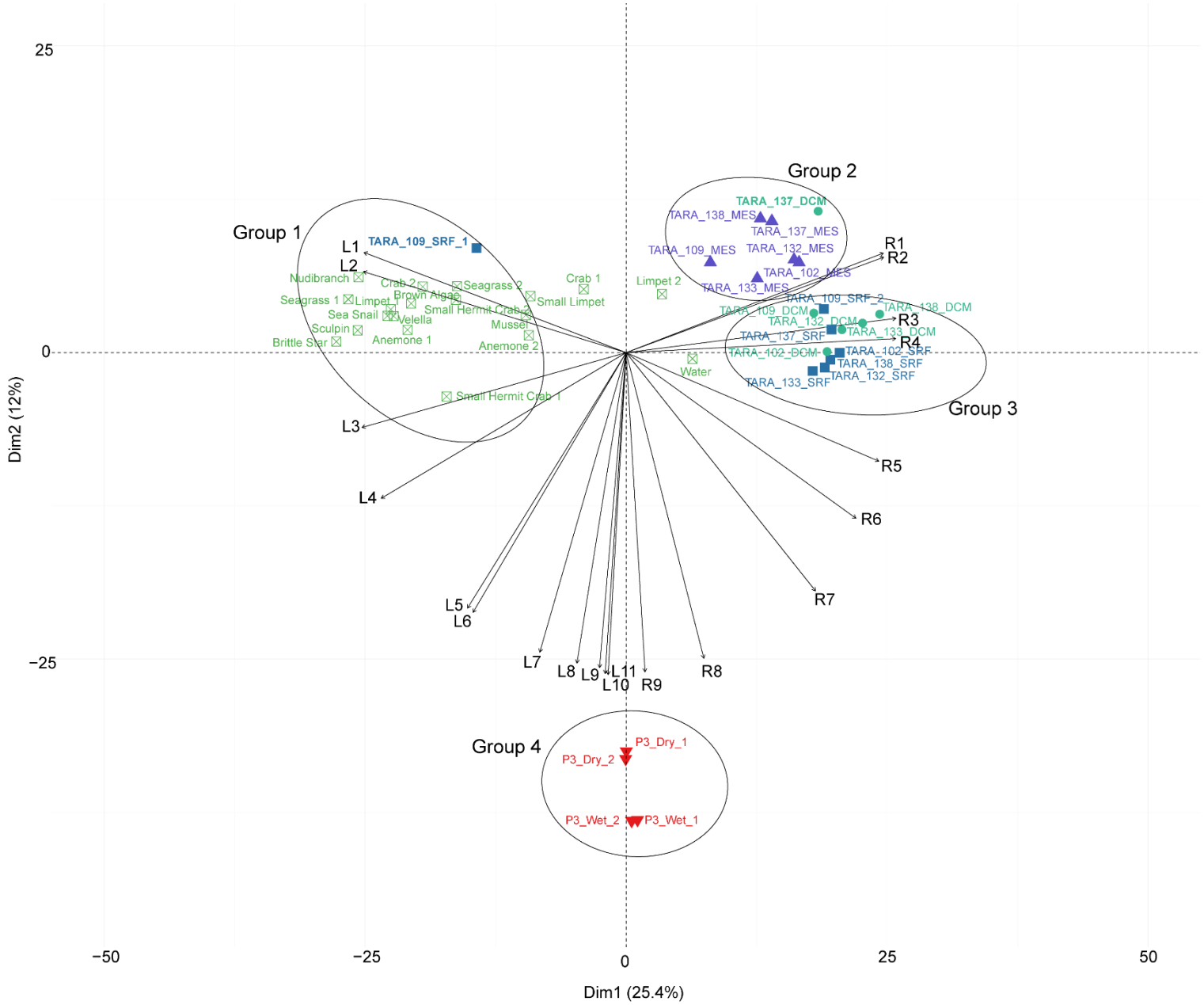
The tide pool metagenomes showed no pattern regarding species profiles (Figs. 2 and S2); however, some clusters were apparent. Brown Algae and Seagrass 2 clustered together with bootstrap support of 100% (S2 Fig.). Another cluster according to PCA and cluster analysis was Anemone 1 and Nudibranch, with bootstrap support of 100% (Figs. 2 and S2). Among the seawater metagenomes, no metagenomes from the DCM or SRF were grouped. Within the DCM, TARA_138_DCM clustered with TARA_132_DCM, and TARA_109_DCM clustered with TARA_109_SRF_2, with bootstrap support > 90%. MES metagenomes clustered together with bootstrap support of 100%, and the TARA_137_DCM metagenome clustered with the MES group (Figs. 2 and S2). Freshwater metagenomes were clustered with bootstrap support of 100% (Figs. 2 and S2), with two groups representing the wet and dry seasons within the freshwater cluster.

Fig. 2. Principal component analysis (PCA) of relative species abundances from the tide pool, seawater, and freshwater metagenomes. PCA was performed using a correlation matrix with factoextra package in R statistical software. Ellipses indicate bootstrap support > 90%.


20
10
0
−10
−20
Dim2 (7.8%)
−10
0
10
Dim1 (8.8%)
P3_Dry_1
P3_Dry_2
P3_Wet_1
P3_Wet_2
TARA_137_MES
TARA_138_MES
TARA_137_DCM
TARA_102_MES
TARA_133_MES
TARA_132_MES
TARA_109_MES
TARA_132_SRF
TARA_138_SRF
TARA_102_DCM
TARA_133_DCM
TARA_138_DCM
TARA_132_DCM
TARA_109_SRF_2
TARA_109_DCM
TARA_137_SRF
TARA_133_SRF
TARA_102_SRF
TARA_109_SRF_1
Limpet 2
Velella
Crab 1
Water
Mussel
Crab 2
Brittle Star
Limpet 1
Sculpin
Seagrass 1
Small Hermit Crab 1
Nudibranch
Sea Snail
Brown Algae
Seagrass 2
Anemone 1
Small Hermit Crab 2
Small Limpet
Anemone 2
Samples
Deep Chlorophyll Maximum layer (DCM)
Mesopelagic zone (MES)
Surface Water layer (SRF)
Paraguaçú River
Tidepool

Regarding SEED level 3 functional classification, our results showed several groups with bootstrap support > 90% (Figs. 3 and S3). Most tide pool metagenomes (15 metagenomes) were clustered with bootstrap support of 94% (S3 Fig.), and the TARA_109_SRF_1 metagenome was grouped with tide pool metagenomes. Within this tide pool group (Group 1) were three small groups. One group consisted of Sculpin, Brittle Star, and Seagrass 1 (bootstrap support of 98%), and the other groups consisted of Seagrass 2 and Brown Algae (bootstrap support of 100%), and Nudibranch and Anemone 1 (bootstrap support of 98%). Variables responsible for the clustering of Group 1 included arsenic resistance, bacterial chemotaxis, cobalt-zinc-cadmium resistance, and biogenesis of cbb3−type cytochrome c oxidases (Fig. 3). Four tide pool metagenomes did not cluster with other groups (Fig. 3). Two metagenomes were closer to Group 1 (Crab 1 and Small Limpet), and the other two metagenomes were closer to seawater metagenomes (Limpet 2 and Water).

Fig. 3. Principal component analysis (PCA) of relative abundances of functional content (SEED database level 3 classification) from tide pool, seawater, and freshwater metagenomes. PCA was performed using a correlation matrix with factoextra package in R statistical software. The 20 most influential variables are shown. Ellipses indicate groups with bootstrap support > 90%.



Samples

- Deep Chlorophyll Maximum layer (DCM)
- Mesopelagic zone (MES)
- Surface Water layer (SRF)
- Paraguaçú River
- Tidepool

Variables from Left

L1. Arsenic resistance;
L2. Bacterial Chemotaxis;
L3. Cobalt−zinc−cadmium resistance;
L4. Biogenesis of cbb3−type cytochrome c oxidases;
L5. CBSS−176299.4.peg.1292;
L6. CBSS−318161.14.peg.2599;
L7. YrdC−YciO−Sua5 and associated protein families;
L8. CBSS−208964.1.peg.1012;
L9. CBSS−232721.5.peg.360;
L10. Tricarballylate Utilization;
L11. Single−Rhodanese−domain proteins;

Variables from Right

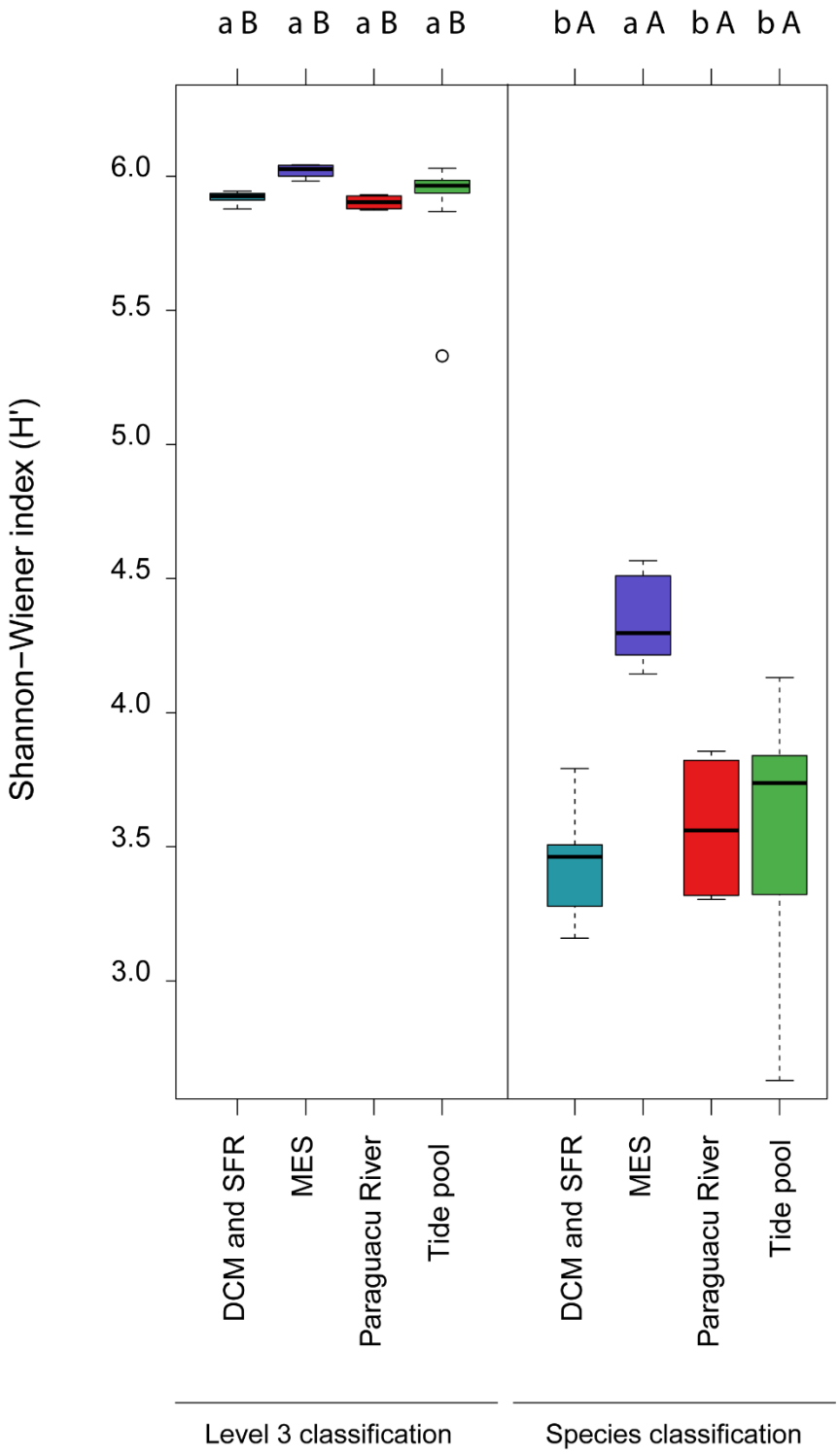
R1. Fatty Acid Biosynthesis FASII;
R2. Modification of eukaryotic initiation factor 5A;
R3. Purine conversions;
R4. Isoprenoid Biosynthesis;
R5. Nonmevalonate Branch of Isoprenoid Biosynthesis;
R6. TenI−like tautomerase;
R7. COG1836;
R8. Pterin metabolism 3;
R9. tRNA aminoacylation, Thr;

Other groups formed by cluster analysis (S3 Fig.) included seawater metagenomes that clustered into two groups (Fig. 3). One seawater group consisted of all MES metagenomes and one DCM metagenome (TARA_137_DCM), with bootstrap support of 100% (Group 2). Another group consisted of the remaining seawater metagenomes (DCM and SRF), with bootstrap support of 100% (Group 3). Variables responsible for the clustering of Group 2 were type II fatty acid biosynthesis (FASII) and modification of eukaryotic initiation factor 5A. Variables responsible for the clustering of Group 3 variablewere purine conversions and isoprenoid biosynthesis. Freshwater metagenomes clustered together with bootstrap support of 100% (Figs. 3 and S2) (Group 4), with two groups representing the wet and dry seasons. Almost all variables responsible for the clustering of Group 4 were related to putative genes (Fig. 3). Other variables influencing the clustering of Group 4 were YrdC−YciO−Sua5 and associated protein families, tricarballylate utilization, single-rhodanese–domain proteins, nonmevalonate branch of isoprenoid biosynthesis, TenI-like tautomerase, pterin metabolism, and tRNA aminoacylation, Thr.

Taxonomic and functional diversity of tide pool, sea, and freshwater metagenomes

Based on the results of cluster analysis, we calculated the Shannon-Wiener diversity index (H') for each profile (Fig. 4). Regarding species profiles, Group 2 (MES) had the highest index value among groups formed by cluster analysis (4.34 ± 0.17, $p < 0.05$); however, no differences between groups were found regarding functional profiles. All functional diversity indices were greater than their respective species diversity indices ($p < 0.05$) (Fig. 4).

Fig. 4. Taxonomic and functional diversity indices of tide pool metagenomes compared with those of seawater and freshwater metagenomes. Diversity indices were calculated using R software and the vegan package. The plot was based on the Shannon–Wiener index (H') for taxonomic and functional profiles (species and SEED level 3 classification, respectively) for each group based on SEED level 3 cluster analysis. The index values from all groups analyzed (both profiles) were compared by factorial ANOVA ($p < 0.05$), followed by the Tukey post hoc test. Groups with the same lower letter did not differ significantly inside each profile. Groups with the same capital letter did not differ significantly regarding taxonomic diversity index and species diversity index.



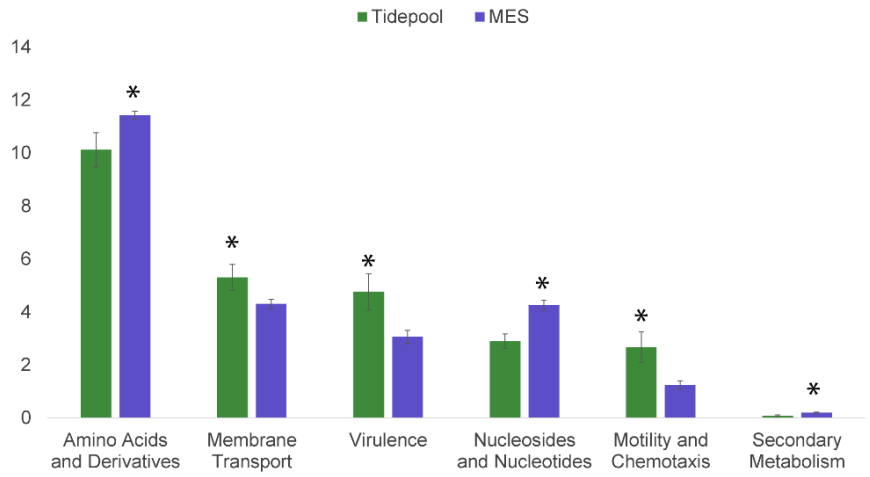
Comparison of tide pool, sea, and freshwater metagenomes

Based on the results of cluster analysis, we compared Group 1 (tide pool group) with other groups and with tide pool metagenomes that were not clustered at SEED level 1 or level

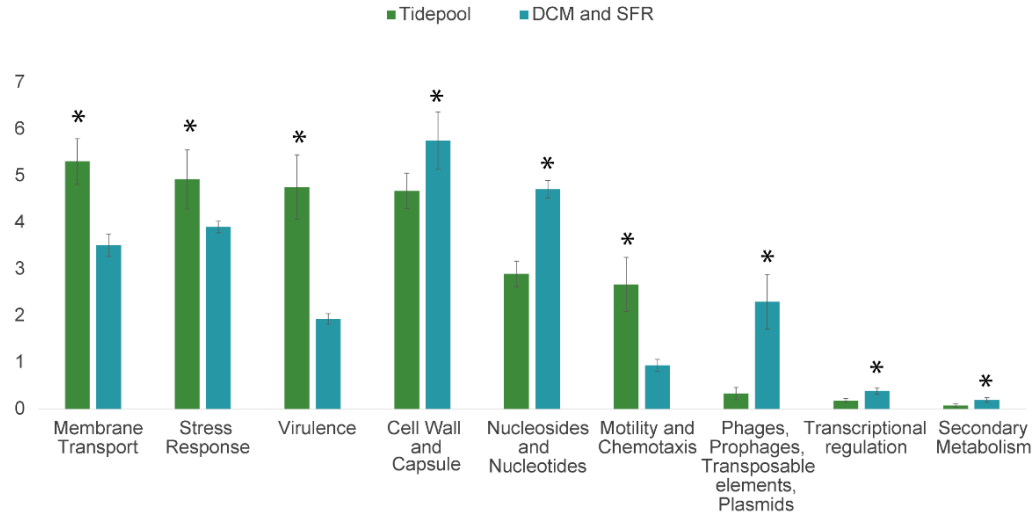
3 classification (Figs. 5 and S4–S7). Compared with Group 2 (MES) (Fig. 5A), Group 1 had more sequences assigned to the membrane transport ($p < 0.05$), virulence ($p < 0.05$), and motility and chemotaxis ($p < 0.05$) categories, whereas Group 2 had more sequences assigned to the amino acids and derivatives ($p < 0.05$), nucleosides and nucleotides ($p < 0.05$), and secondary metabolism ($p < 0.05$) categories. A comparison of Group 1 and Group 2 (S4 Fig.) showed that categories more abundant in Group 1 were sodium hydrogen antiporter ($p < 0.05$), biogenesis of cbb3-type cytochrome c oxidases ($p < 0.05$), DNA-binding regulatory proteins, strays ($p < 0.05$), alkylphosphonate utilization ($p < 0.05$), ectoine biosynthesis and regulation ($p < 0.05$), universal stress protein family ($p < 0.05$), curli production ($p < 0.05$), seqA and co-occurring genes ($p < 0.05$), tetrathionate respiration ($p < 0.05$), glutathionylspermidine and trypanothione ($p < 0.05$), and Pseudomonas quinolone signal (PQS) ($p < 0.05$). Compared with Group 3 (DCM and SRF) (Fig. 5B), Group 1 had more sequences assigned to the membrane transport ($p < 0.05$), stress response ($p < 0.05$), virulence ($p < 0.05$), and motility and chemotaxis ($p < 0.05$) categories, whereas Group 3 had more sequences assigned to cell wall and capsule ($p < 0.05$), nucleosides and nucleotides ($p < 0.05$), phages, prophages, transposable elements, plasmids ($p < 0.05$), transcriptional regulation ($p < 0.05$), and secondary metabolism ($p < 0.05$) categories. Regarding SEED level 3 categories (S5 Fig.), Group 1 had more sequences assigned to cobalt-zinc-cadmium resistance ($p < 0.05$), bacterial chemotaxis ($p < 0.05$), copper homeostasis ($p < 0.05$), bacterial hemoglobins ($p < 0.05$), nitrate and nitrite ammonification ($p < 0.05$), lipopolysaccharide assembly ($p < 0.05$), type IV pilus ($p < 0.05$), type VI secretion systems ($p < 0.05$), general secretion pathway ($p < 0.05$), orphan regulatory proteins ($p < 0.05$), and DNA repair, bacterial RecBCD pathway ($p < 0.05$).

Fig. 5. Comparative analysis (SEED level 1 classification) of tide pool metagenomes with seawater and freshwater metagenomes. A) Tide pool (Group 1) versus MES (Group 2). B) Tide pool versus DCM and SFR (Group 3). C) Tide pool versus Paraguaçú River (Group 4). Samples were compared by Welch's t-test ($p < 0.05$), asymptotic confidence intervals (95%), and Bonferroni correction in STAMP software [11]. Level classification labels with small effect sizes were removed by filtering (difference between proportions < 1.0 or ratio of proportions < 2.00). All outliers were removed from these analyses.

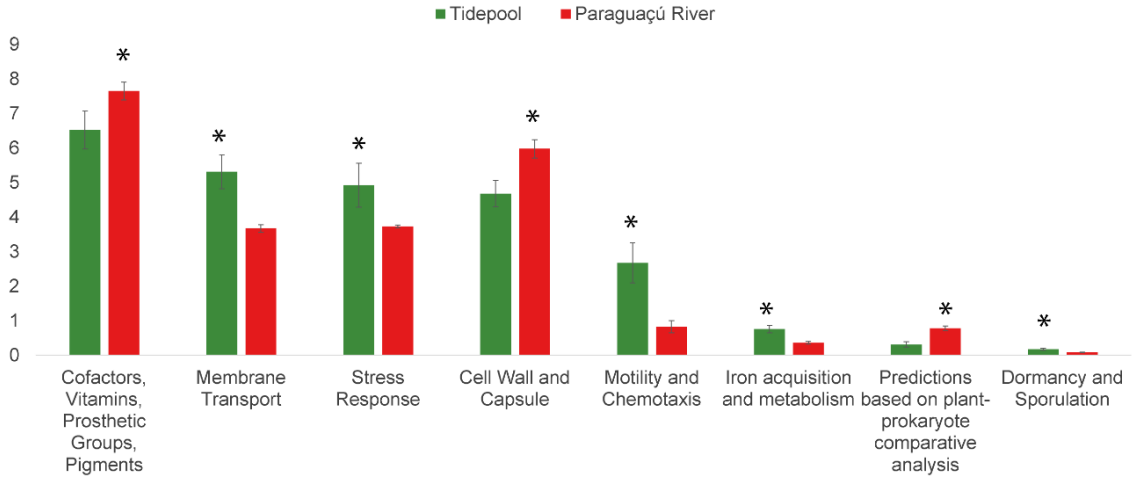
A


Tidepool
MES
14
12
10
8
6
4
2
0
Amino Acids and Derivatives
Membrane Transport
Virulence
Nucleosides and Nucleotides
Motility and Chemotaxis
Secondary Metabolism

B


Tidepool
DCM and SFR
7
6
5
4
3
2
1
0
Membrane Transport
Stress Response
Virulence
Cell Wall and Capsule
Nucleosides and Nucleotides
Motility and Chemotaxis
Phages, Prophages, Transposable elements, Plasmids
Transcriptional regulation
Secondary Metabolism

C


Tidepool
Paraguaçú River
9
8
7
6
5
4
3
2
1
0
Cofactors, Vitamins, Prosthetic Groups, Pigments
Membrane Transport
Stress Response
Cell Wall and Capsule
Motility and Chemotaxis
Iron acquisition and metabolism
Predictions based on plant-prokaryote comparative analysis
Dormancy and Sporulation

A comparison of Group 1 with freshwater metagenomes (Group 4) at SEED level 1 (Fig. 5C) showed that Group 1 had more sequences assigned to the membrane transport ($p < 0.05$), stress response ($p < 0.05$), motility and chemotaxis ($p < 0.05$), iron acquisition and metabolism ($p < 0.05$), and dormancy and sporulation ($p < 0.05$) categories, whereas Group 4 had more sequences assigned to the cofactors, vitamins, prosthetic groups, pigments ($p < 0.05$), cell wall and capsule ($p < 0.05$), and predictions based on plant-prokaryotes comparative analysis ($p < 0.05$) categories. Regarding SEED level 3 categories (S6 Fig.), Group 1 had more sequences assigned to the Ton and Tol transport systems ($p < 0.05$), bacterial chemotaxis ($p < 0.05$), Na (+)- translocating NADH-quinone oxidoreductase and rnf-like group of electron transport complexes ($p < 0.05$), bacterial hemoglobins ($p < 0.05$), type VI secretion systems ($p < 0.05$), lipopolysaccharide assembly ($p < 0.05$), DNA repair, bacterial RecBCD pathway ($p < 0.05$), and denitrification ($p < 0.05$) categories.

Discussion

The most abundant species in the tide pool metagenomes belonged to the class Gammaproteobacteria. The three most abundant species (*Alteromonas macleodii*, *Pseudomonas stutzeri*, and *Glaciecola psychrophila*) are commonly associated with stressful environments. *A. macleodii* is frequently detected in marine environments, mainly the DCM and MES layers [30]; however, *A. macleodii* has also been detected in surface layers [31], which has conditions similar to those found in tide pools. Ivars-Martinez *et al.* suggested that the presence of the photolyase gene in *A. macleodii* could be related to high sunlight exposure [31]. *P. stutzeri* is a species known for its denitrifying potential [32,33]. Moreover, this species is able to grow in environments with low oxygen concentrations, which allows *P. stutzeri* to fix nitrogen [34]. *G. psychrophila*, which was first isolated from sea-ice samples, was proposed as a new species because its physiology differed from that of other *Glaciecola* species [35]. The *G. psychrophila* genome was published recently [36], but no extensive studies about this new species have yet been published. However, *G. psychrophila* does not show the ability to reduce nitrate [35].

Sequences belonging to the categories carbohydrates, amino acids and derivatives, and protein metabolism are important for nutrient utilization, which may explain why they are the most abundant categories; however, the membrane transport and motility and chemotaxis categories seem more important to tide pool microorganisms. Tide pools are disconnected from

the ocean for extended periods during low tide, which increases sunlight exposure and decreases nutrient levels. Sunlight increases the evaporation of tide pool water, thereby increasing Na^+ concentration. The sodium hydrogen antiporter plays a role in cell homeostasis, pH regulation, and osmoregulation by removing excess Na^+ from the cell to the extracellular environment [37]. Fernández *et al.* reported that metagenomes from hypersaline environments demonstrate C and N cycling capacity, the use of light as an energy source by bacteriorhodopsins, and the potential to synthesize osmoprotective compounds such as glycine betaine, ectoine, and trehalose [38].

Tide pool areas are good environments for biofilms, which may account for the abundance of sequences related to motility and chemotaxis. The curli protein, which enables adhesion, colonization, and biofilm production [39], was more abundant in the tide pool group than in the other groups. In stressful environments such as tide pools, chemotaxis is extremely important for microorganisms. The Pseudomonas quinolone signal (PQS), which was also more abundant in the tide pool group, mediates the delivery of antimicrobials and toxins and DNA transport, and may improve survival upon challenge with stressing agents [40,41]. Although PQS enhances DNA fragmentation, this compound also increases fitness under UV radiation [42]. Häussler *et al*. suggested that the DNA fragmentation could act as a sensor for stressful conditions to stimulate DNA repair machinery [42].

In oxygen-poor environments, microorganisms have several specialized features. Some sulfur microorganisms (e.g. *Thioploca* spp.) require nitrate to oxidize sulfide, and at low nitrate concentrations these sulfur microorganisms exhibit chemotaxis toward nitrate [43]. Tide pool microorganisms show a higher abundance of sequences related to bacterial hemoglobins and biogenesis of cbb3-type cytochrome c oxidases. Bacterial hemoglobins are associated with oxygen limitation and oxidative/nitrosative stress [44]. Cbb3-type cytochrome c oxidases are copper-dependent enzymes that are crucial for the onset of many anaerobic biological processes, such as anoxygenic photosynthesis or nitrogen fixation [45,46].

Another indicator of stressful conditions is the presence of universal stress protein (*USP*) genes. *USP* genes are induced under conditions of nutrient starvation, drought, high salinity, extreme temperatures, and presence of toxic metals [47,48]; however, the mechanisms by which USP allows organisms to cope with these environmental stressors remain unclear. Gustavsson *et al.* showed that USP has an important role in protection against ultraviolet (UV) irradiation [49], which may be an important feature for survival in bright environments.

Nitrogen, which is limiting nutrient in several marine ecosystems [50], was a factor differentiating the tide pool group from the DCM and SFR group. The gene encoding dissimilatory nitrite reductase and genes involved in nitrate and nitrite ammonification, denitrification, and nitrosative stress were more abundant in the tide pool group. Although tide pools can be disconnected from the ocean for long periods of time without its supply of nitrogen, some tide pool invertebrates excrete ammonium, which can be used as a nitrogen source [51,52]. Prister showed that even in tide pools emergent for hours, the presence of mussels, crabs, fish, limpets, and invertebrates can prevent nitrogen limitation [10]. Several steps of the marine nitrogen cycle are mediated by bacterial chemotaxis, such as nitrogen fixation, nitrification, and denitrification [43,53,54].

Arsenic is readily detected in many marine environments [55]. Its sources include leaching from soil and rocks [56] and active ridges that release arsenic from the inner Earth [56]. Algae in superficial water layers can accumulate arsenic and are involved in arsenic cycling [57,58]. Although researchers have studied arsenic-resistant microorganisms [56,59], their role in arsenic recycling remains unclear. It is likely that arsenic and other heavy metals (e.g. cobalt, zinc, and cadmium) are more concentrated in tide pools because of evaporation.

Although functional diversity was similar between environments in our study, the relative gene abundances differed. Nutrient scarcity is an important factor influencing species composition and abundance [50]. Previous studies reported that the same clusters were identified by PCA whether taxonomic or functional profiles were used [60,61]; however, other studies showed that environments shape communities regarding functional content, particularly extreme environments (e.g. halophilic, acid, cold, and deep subterranean environments) [62,63].

Conclusion

Tide pools represent unique environments characterized by nutrient starvation, hypersalinity, and bright sunlight throughout the day when they are isolated from the sea. This type of environment is more influenced by weather variations than large environments (e.g. forest, sea). These weather variations may play a crucial role in determining community structure, particularly regarding the functional profile. Genes related to stress conditions, communication between microorganisms, and homeostasis of metals and sodium are more

abundant in this stressed environment compared with the other metagenomes analyzed in this study (especially several seawater metagenomes), suggesting a strong selection for these features in the tide pool environment. Moreover, almost all tide pool metagenomes were clustered based on functional profiles (SEED level 3), but no consistent tide pool cluster was observed based on species profiles. This finding suggests that the influence of this environment on its microbial community involves functional content more than species composition.

Acknowledgments

This work was supported by grants from the Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) and 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). F.A.C.L. is supported by a CAPES fellowship. RHK is partially supported by a CAPES Senior fellowship.

References

1. Horn MH, Martin KL, Chotkowski MA. Intertidal fishes: life in two worlds. Academic Press; 1998.

2. Bertness JW, Nybakken JW. Marine Biology, an Ecological Approach. Benjamin Cummings Bolch CJ Reynolds MJ 2002 Species Resolut Glob Distrib Microreticulate Dinoflag Cysts J Plankton Res. 2005;24: 565–578.

3. Bracken MES, Friberg SE, Gonzalez-Dorantes CA, Williams SL. Functional consequences of realistic biodiversity changes in a marine ecosystem. Proc Natl Acad Sci U S A. 2008;105: 924–928. doi:10.1073/pnas.0704103105

4. Altieri AH, Trussell GC, Ewanchuk PJ, Bernatchez G, Bracken MES. Consumers Control Diversity and Functioning of a Natural Marine Ecosystem. PLOS ONE. 2009;4: e5291. doi:10.1371/journal.pone.0005291

5. Longo GO, Morais RA, Martins CDL, Mendes TC, Aued AW, Cândido DV, et al. Between-Habitat Variation of Benthic Cover, Reef Fish Assemblage and Feeding Pressure on the Benthos at the Only Atoll in South Atlantic: Rocas Atoll, NE Brazil. PLoS ONE. 2015;10. doi:10.1371/journal.pone.0127176

6. Barreto FS, Schoville SD, Burton RS. Reverse genetics in the tide pool: knock-down of target gene expression via RNA interference in the copepod Tigriopus californicus. Mol Ecol Resour. 2015;15: 868–879. doi:10.1111/1755-0998.12359

7. Sorte CJB, Bracken MES. Warming and Elevated CO2 Interact to Drive Rapid Shifts in Marine Community Production. PLoS ONE. 2015;10. doi:10.1371/journal.pone.0145191

8. Sampaio HA, Martinelli-Lemos JM, Sampaio HA, Martinelli-Lemos JM. Use of intertidal areas by shrimps (Decapoda) in a brazilian Amazon estuary. An Acad Bras Ciênc. 2014;86: 333–345. doi:10.1590/0001-3765201420120039

9. Godinho WO, Lotufo TMC. Local v. microhabitat influences on the fish fauna of tidal pools in north-east Brazil. J Fish Biol. 2010;76: 487–501. doi:10.1111/j.1095-8649.2009.02501.x

10. Pfister CA. Intertidal invertebrates locally enhance primary production. Ecology. 2007;88: 1647–53. doi:10.1890/06-1913.1

11. Bracken MES, Gonzalez-Dorantes CA, Stachowicz JJ. Whole-Community Mutualism: Associated Invertebrates Facilitate a Dominant Habitat-Forming Seaweed. Ecology. 2007;88: 2211–2219. doi:10.1890/06-0881.1

12. Björk M, Axelsson L, Beer S. Why is Ulva intestinalis the only macroalga inhabiting isolated rockpools along the Swedish Atlantic coast? Mar Ecol Prog Ser. 2004;284: 109–116.

13. Jiao L, Chen H, Meng W, Lei K, Zheng B. PAHs biodegradation in intertidal surface sediment by indigenous microorganisms. Environ Sci Pollut Res. 2014;21: 6463–6471. doi:10.1007/s11356-013-2480-5

14. Yu Z, Zhou N, Qiao H, Qiu J. Identification, Cloning, and Expression of L-Amino Acid Oxidase from Marine Pseudoalteromonas sp. B3. Sci World J. 2014;2014. doi:10.1155/2014/979858

15. Zhang J, Kobert K, Flouri T, Stamatakis A. PEAR: a fast and accurate Illumina Paired-End reAd mergeR. Bioinformatics. 2014;30: 614–620. doi:10.1093/bioinformatics/btt593

16. Schmieder R, Edwards R. Quality control and preprocessing of metagenomic datasets. Bioinformatics. 2011;27: 863–864. doi:10.1093/bioinformatics/btr026

17. Silva GGZ, Cuevas DA, Dutilh BE, Edwards RA. FOCUS: an alignment-free model to identify organisms in metagenomes using non-negative least squares. PeerJ. 2014;2: e425. doi:10.7717/peerj.425

18. Tatusova T, Ciufo S, Fedorov B, O'Neill K, Tolstoy I. RefSeq microbial genomes database: new representation and annotation strategy. Nucleic Acids Res. 2014;42: D553–D559. doi:10.1093/nar/gkt1274

19. Silva GGZ, Green KT, Dutilh BE, Edwards RA. SUPER-FOCUS: A tool for agile functional analysis of shotgun metagenomic data. Bioinformatics. 2015; btv584. doi:10.1093/bioinformatics/btv584

20. Overbeek R, Begley T, Butler RM, Choudhuri JV, Chuang H-Y, Cohoon M, et al. The Subsystems Approach to Genome Annotation and its Use in the Project to Annotate 1000 Genomes. Nucleic Acids Res. 2005;33: 5691–5702. doi:10.1093/nar/gki866

21. Zhao Y, Tang H, Ye Y. RAPSearch2: a fast and memory-efficient protein similarity search tool for next-generation sequencing data. Bioinformatics. 2012;28: 125–126. doi:10.1093/bioinformatics/btr595

22. Kassambara A. factoextra: Extract and visualize the results of multivariate data analyses [Internet]. 2015. Available: http://www.sthda.com

23. R Core Team. R: A Language and Environment for Statistical Computing [Internet]. Vienna, Austria: R Foundation for Statistical Computing; 2015. Available: https://www.R-project.org/

24. Suzuki R, Shimodaira H. pvclust: Hierarchical Clustering with P-Values via Multiscale Bootstrap Resampling [Internet]. 2014. Available: http://CRAN.R-project.org/package=pvclust

25. Parks DH, Tyson GW, Hugenholtz P, Beiko RG. STAMP: statistical analysis of taxonomic and functional profiles. Bioinformatics. 2014;30: 3123–3124. doi:10.1093/bioinformatics/btu494

26. Oksanen J, Blanchet FG, Kindt R, Legendre P, Minchin PR, O'Hara RB, et al. vegan: Community Ecology Package [Internet]. 2016. Available: http://CRAN.R-project.org/package=vegan

27. Hill MO. Diversity and Evenness: A Unifying Notation and Its Consequences. Ecology. 1973;54: 427–432. doi:10.2307/1934352

28. Sunagawa S, Coelho LP, Chaffron S, Kultima JR, Labadie K, Salazar G, et al. Structure and function of the global ocean microbiome. Science. 2015;348: 1261359. doi:10.1126/science.1261359

29. Lopes FAC, Catão ECP, Santana RH, Cabral A de S, Paranhos R, Rangel TP, et al. Microbial Community Profile and Water Quality in a Protected Area of the Caatinga Biome. PLoS ONE. 2016;11: e0148296. doi:10.1371/journal.pone.0148296

30. García-Martínez J, Acinas SG, Massana R, Rodríguez-Valera F. Prevalence and microdiversity of Alteromonas macleodii-like microorganisms in different oceanic regions. Environ Microbiol. 2002;4: 42–50. doi:10.1046/j.1462-2920.2002.00255.x

31. Ivars-Martinez E, Martin-Cuadrado A-B, D'Auria G, Mira A, Ferriera S, Johnson J, et al. Comparative genomics of two ecotypes of the marine planktonic copiotroph Alteromonas macleodii suggests alternative lifestyles associated with different kinds of particulate organic matter. ISME J. 2008;2: 1194–1212. doi:10.1038/ismej.2008.74

32. Zumft WG. Nitric oxide reductases of prokaryotes with emphasis on the respiratory, heme–copper oxidase type. J Inorg Biochem. 2005;99: 194–215. doi:10.1016/j.jinorgbio.2004.09.024

33. Zumft WG. Biogenesis of the Bacterial Respiratory CuA, Cu-S Enzyme Nitrous Oxide Reductase. J Mol Microbiol Biotechnol. 2005;10: 154–166.

34. Lalucat J, Bennasar A, Bosch R, García-Valdés E, Palleroni NJ. Biology of Pseudomonas stutzeri. Microbiol Mol Biol Rev. 2006;70: 510–547. doi:10.1128/MMBR.00047-05

35. Zhang D-C, Yu Y, Chen B, Wang H-X, Liu H-C, Dong X-Z, et al. Glaciecola psychrophila sp. nov., a novel psychrophilic bacterium isolated from the Arctic. Int J Syst Evol Microbiol. 2006;56: 2867–2869. doi:10.1099/ijs.0.64575-0

36. Yin J, Chen J, Liu G, Yu Y, Song L, Wang X, et al. Complete Genome Sequence of Glaciecola psychrophila Strain 170T. Genome Announc. 2013;1: e00199–13. doi:10.1128/genomeA.00199-13

37. Melo AMP, Felix NAM, Carita JN, Saraiva LM, Teixeira M. The Na+/H+ antiporter of the thermohalophilic bacterium Rhodothermus marinus. Biochem Biophys Res Commun. 2006;348: 1011–1017. doi:10.1016/j.bbrc.2006.07.134

38. Fernández AB, Ghai R, Martin-Cuadrado A-B, Sánchez-Porro C, Rodriguez-Valera F, Ventosa A. Prokaryotic taxonomic and metabolic diversity of an intermediate salinity hypersaline habitat assessed by metagenomics. FEMS Microbiol Ecol. 2014;88: 623–635. doi:10.1111/1574-6941.12329

39. Dyer JG, Sriranganathan N, Nickerson SC, Elvinger F. Curli Production and Genetic Relationships Among Escherichia coli from Cases of Bovine Mastitis. J Dairy Sci. 2007;90: 193–201. doi:10.3168/jds.S0022-0302(07)72620-6

40. McBroom AJ, Kuehn MJ. Release of outer membrane vesicles by Gram-negative bacteria is a novel envelope stress response. Mol Microbiol. 2007;63: 545–558. doi:10.1111/j.1365-2958.2006.05522.x

41. Mashburn LM, Whiteley M. Membrane vesicles traffic signals and facilitate group activities in a prokaryote. Nature. 2005;437: 422–425. doi:10.1038/nature03925

42. Häussler S, Becker T. The Pseudomonas Quinolone Signal (PQS) Balances Life and Death in Pseudomonas aeruginosa Populations. PLOS Pathog. 2008;4: e1000166. doi:10.1371/journal.ppat.1000166

43. Stocker R, Seymour JR. Ecology and Physics of Bacterial Chemotaxis in the Ocean. Microbiol Mol Biol Rev MMBR. 2012;76: 792–812. doi:10.1128/MMBR.00029-12

44. Frey AD, Kallio PT. Bacterial hemoglobins and flavohemoglobins: versatile proteins and their impact on microbiology and biotechnology. FEMS Microbiol Rev. 2003;27: 525–545. doi:10.1016/S0168-6445(03)00056-1

45. Ekici S, Pawlik G, Lohmeyer E, Koch H-G, Daldal F. Biogenesis of cbb3-type cytochrome c oxidase in Rhodobacter capsulatus. Biochim Biophys Acta. 2012;1817: 898–910. doi:10.1016/j.bbabio.2011.10.011

46. Ekici S, Yang H, Koch H-G, Daldal F. Novel Transporter Required for Biogenesis of cbb3-Type Cytochrome c Oxidase in Rhodobacter capsulatus. mBio. 2012;3. doi:10.1128/mBio.00293-11

47. Siegele DA. Universal Stress Proteins in Escherichia coli. J Bacteriol. 2005;187: 6253–6254. doi:10.1128/JB.187.18.6253-6254.2005

48. Liu W-T, Karavolos MH, Bulmer DM, Allaoui A, Hormaeche RDCE, Lee JJ, et al. Role of the universal stress protein UspA of Salmonella in growth arrest, stress and virulence. Microb Pathog. 2007;42: 2–10. doi:10.1016/j.micpath.2006.09.002

49. Gustavsson N, Diez. A., Nyström T. The universal stress protein paralogues of Escherichia coli are co-ordinately regulated and co-operate in the defence against DNA damage. Mol Microbiol. 2002;43: 107–117. doi:10.1046/j.1365-2958.2002.02720.x

50. Moore CM, Mills MM, Arrigo KR, Berman-Frank I, Bopp L, Boyd PW, et al. Processes and patterns of oceanic nutrient limitation. Nat Geosci. 2013;6: 701–710. doi:10.1038/ngeo1765

51. Bracken MES, Nielsen KJ. Diversity of intertidal macroalgae increases with nitrogen loading by invertebrates. Ecology. 2004;85. Available: http://escholarship.org/uc/item/47t5n0vn

52. Bracken MES. Invertebrate-Mediated Nutrient Loading Increases Growth of an Intertidal Macroalga1. J Phycol. 2004;40: 1032–1041. doi:10.1111/j.1529-8817.2004.03106.x

53. Van de Broek A, Lambrecht M, Vanderleyden J. Bacterial chemotactic motility is important for the initiation of wheat root colonization by Azospirillum brasilense. Microbiology. 1998;144: 2599–2606. doi:10.1099/00221287-144-9-2599

54. Kennedy MJ, Lawless JG. Role of Chemotaxis in the Ecology of Denitrifiers. Appl Environ Microbiol. 1985;49: 109–114.

55. Neff JM. Ecotoxicology of arsenic in the marine environment. Environ Toxicol Chem. 1997;16: 917–927. doi:10.1002/etc.5620160511

56. Chen S, Shao Z. Isolation and diversity analysis of arsenite-resistant bacteria in communities enriched from deep-sea sediments of the Southwest Indian Ocean Ridge. Extremophiles. 2008;13: 39–48. doi:10.1007/s00792-008-0195-1

57. Francesconi KA, Edmonds JS. Arsenic species in marine samples. Croat Chem Acta. 1998;71: 343–359.

58. Dembitsky VM, Levitsky DO. Arsenolipids. Prog Lipid Res. 2004;43: 403–448. doi:10.1016/j.plipres.2004.07.001

59. Takeuchi M, Kawahata H, Gupta LP, Kita N, Morishita Y, Ono Y, et al. Arsenic resistance and removal by marine and non-marine bacteria. J Biotechnol. 2007;127: 434–442. doi:10.1016/j.jbiotec.2006.07.018

60. White RA, Chan AM, Gavelis GS, Leander BS, Brady AL, Slater GF, et al. Metagenomic Analysis Suggests Modern Freshwater Microbialites Harbor a Distinct Core Microbial Community. Front Microbiol. 2016;6. doi:10.3389/fmicb.2015.01531

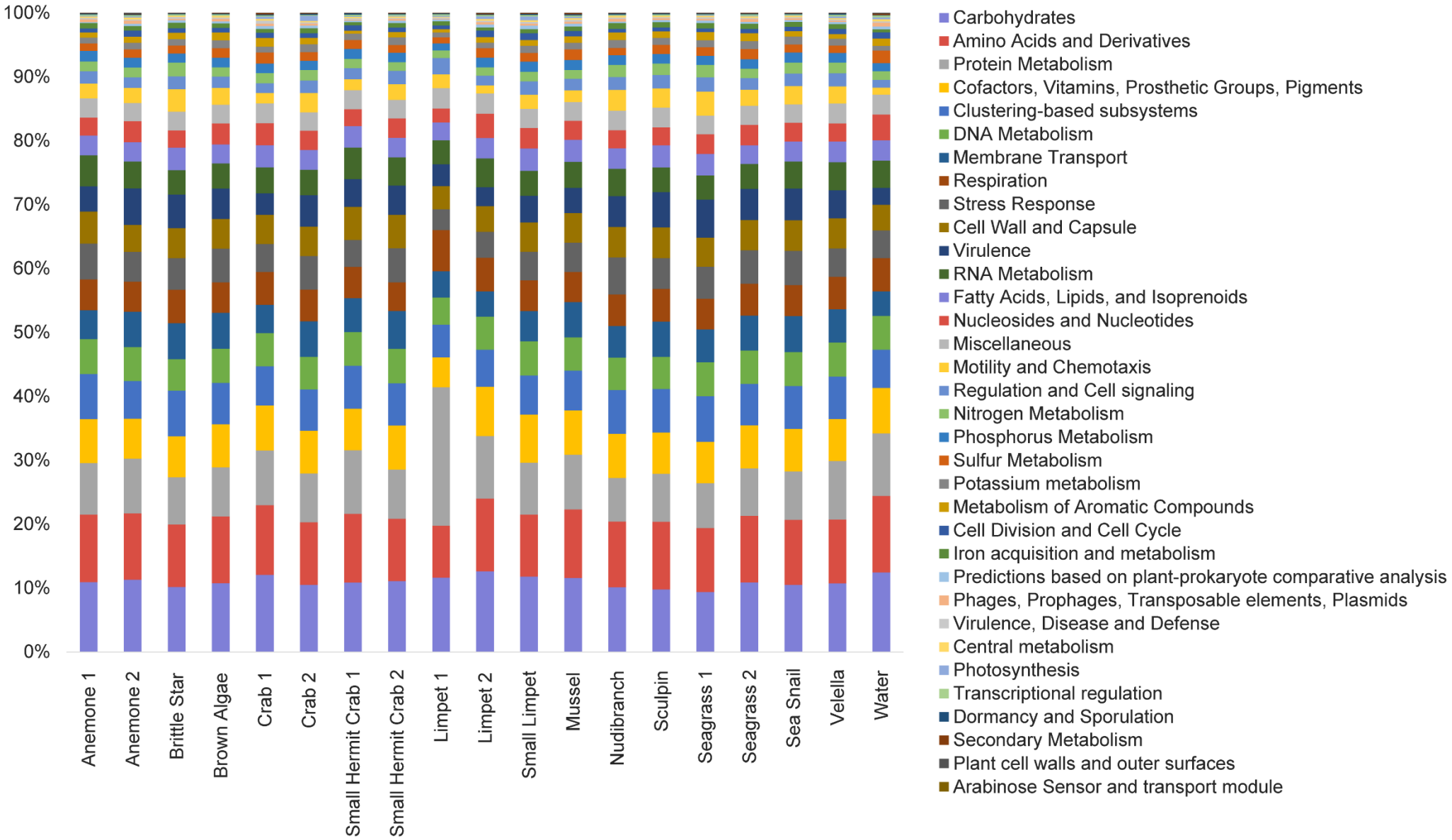
61. Luo G, Fotidis IA, Angelidaki I. Comparative analysis of taxonomic, functional, and metabolic patterns of microbiomes from 14 full-scale biogas reactors by metagenomic sequencing and radioisotopic analysis. Biotechnol Biofuels. 2016;9. doi:10.1186/s13068-016-0465-6

62. Cowan D, Ramond J-B, Makhalanyane T, De Maayer P. Metagenomics of extreme environments. Curr Opin Microbiol. 2015;25: 97–102. doi:10.1016/j.mib.2015.05.005

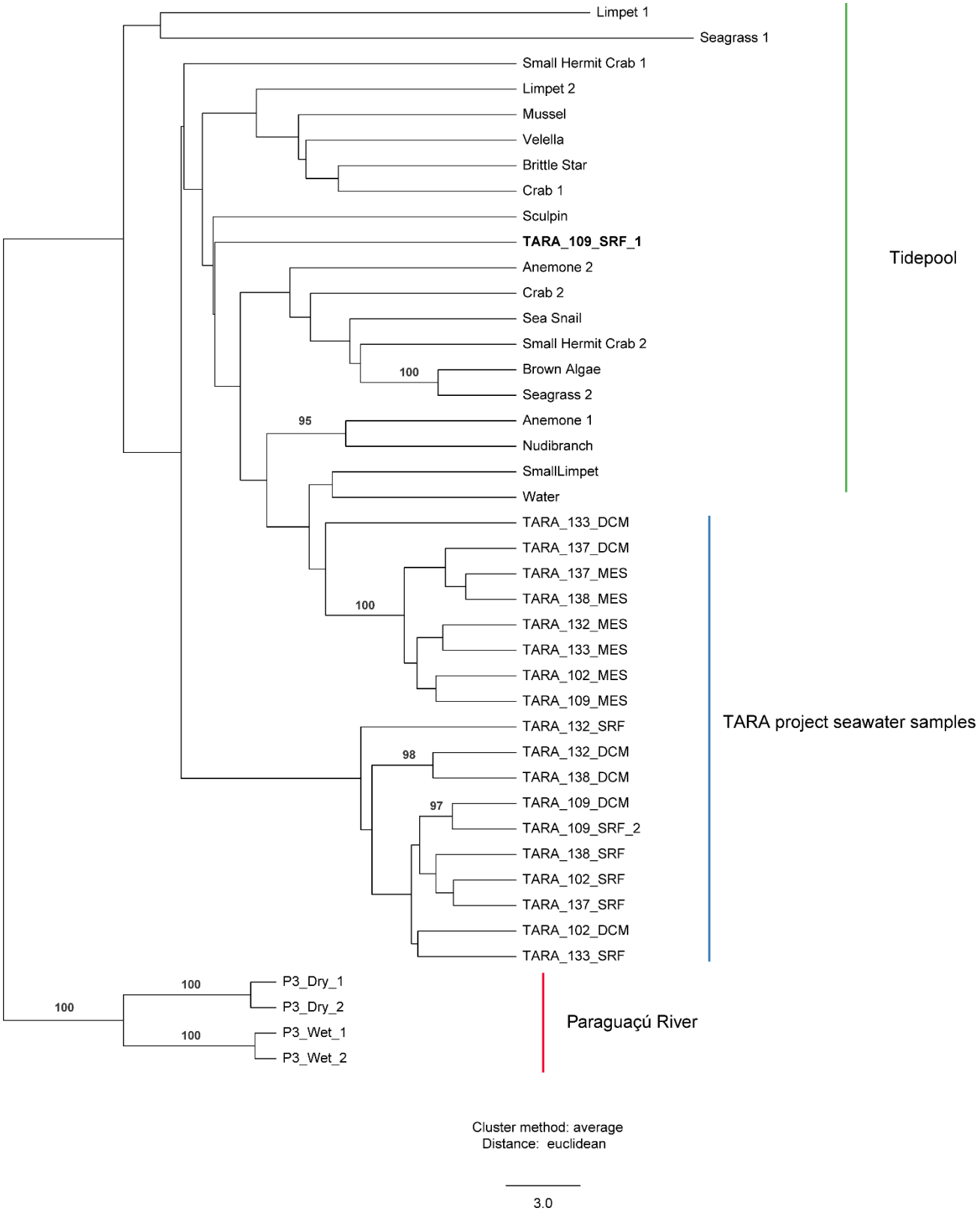
63. Graves CJ, Makrides E, Schmidt V, Giblin A, Cardon Z, Rand D. Functional responses of salt marsh microbial communities to long-term nutrient enrichment. Appl Environ Microbiol. 2016; AEM.03990–15. doi:10.1128/AEM.03990-15

Supporting information

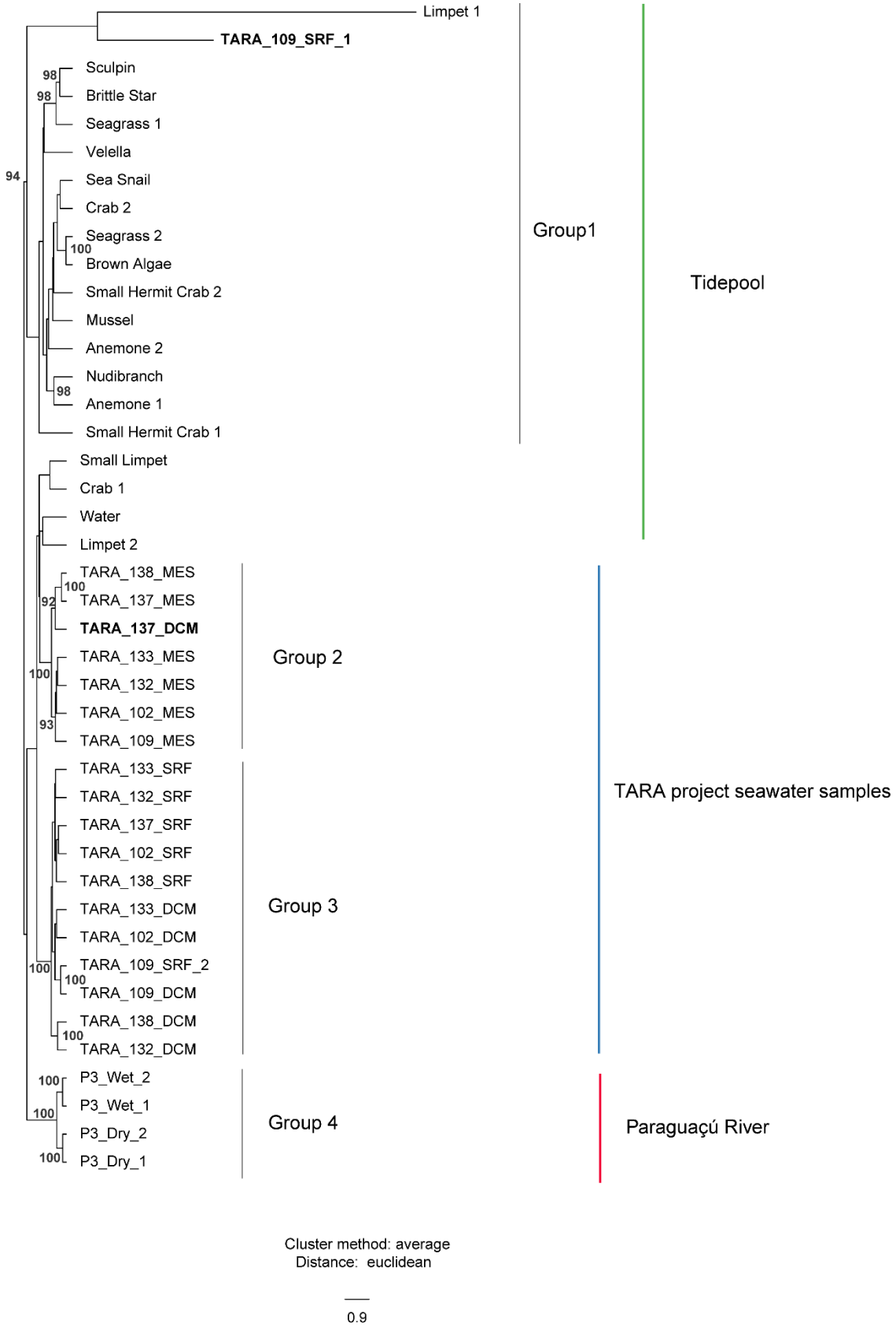
S1 Fig. Functional classification of tide pool metagenomes. The functional classification was based on the SEED database level 1 in SUPER-FOCUS v.0.25 using RAPSearch2 aligner and default parameters.



S2 Fig. Species cluster analysis of tide pool, seawater, and freshwater metagenomes. Cluster analysis of the relative abundances of species in each profile by Euclidian distance, UPGMA algorithm, and bootstrap of 1,000,000 in R statistical software using the pvclust package. Only clusters with bootstrap support > 90% were selected.



S3 Fig. SEED level 3 cluster analysis of tide pool, seawater, and freshwater metagenomes. Cluster analysis based on the relative abundance of SEED level 3 classifications of each profile by Euclidian distance, UPGMA algorithm, and bootstrap of 1,000,000 in R statistical software using the "pvclust" package. Only clusters with bootstrap values > 90% were selected.



S4 Fig. Comparative analysis of functional content (SEED level 3) of tide pool metagenomes (Group 1) against MES seawater metagenomes (Group 2). Samples were compared by Welch's t-test ($p < 0.05$), asymptotic confidence intervals (95%), and Bonferroni correction in STAMP software [11]. Level classification labels with small effect sizes were removed by filtering (difference between proportions < 1.0 or ratio of proportions < 4.00). All outliers were removed from these analyses.

Figure available only online.

S5 Fig. Comparative analysis of functional content (SEED level 3) of tide pool metagenomes (Group 1) against DCM and SRF seawater metagenomes (Group 3). Samples were compared by Welch's t-test ($p < 0.05$), asymptotic confidence intervals (95%), and Bonferroni correction in STAMP software [11]. Level classification labels with small effect sizes were removed by filtering (difference between proportions < 1.0 or ratio of proportions < 4.00). All outliers were removed from these analyses.

Figure available only online.

S6 Fig. Comparative analysis of functional content (SEED level 3) of tide pool metagenomes (Group 1) against Paraguaçú River freshwater metagenomes (Group 4). Samples were compared by Welch's t-test ($p < 0.05$), asymptotic confidence intervals (95%), and Bonferroni correction in STAMP software [11]. Level classification labels with small effect sizes were removed by filtering (difference between proportions < 1.0 or ratio of proportions < 4.00). All outliers were removed from these analyses.

Figure available only online.

S1 Table. *Outgroup* metagenomes used in this study.

Study	Water classification	Name	Location	Sample	Observation
Lopes *et al*. (2016)	Freshwater	Paraguaçú River	Bahia, Brazil	P3_Wet_1	Wet season
				P3_Wet_2	Wet season
				P3_Dry_1	Dry season
				P3_Dry_2	Dry season
Sunagawa *et al*. (2015)	Seawater	Pacific Ocean	Pacific Equatorial Divergence Province	TARA_102_DCM	DCM
				TARA_102_MES	MES
				TARA_102_SRF	SRF
			Chile−Peru Current Coastal Province	TARA_109_DCM	DCM
				TARA_109_MES	MES
				TARA_109_SRF_1	SRF
				TARA_109_SRF_2	SRF
			North Pacific Subtropical and Polar Front Provinces	TARA_132_DCM	DCM
				TARA_132_MES	MES
				TARA_132_SRF	SRF
			North Pacific Subtropical and Polar Front Provinces 2	TARA_133_DCM	DCM
				TARA_133_MES	MES
				TARA_133_SRF	SRF
			North Pacific Equatorial Countercurrent Province	TARA_137_DCM	DCM
				TARA_137_MES	MES
				TARA_137_SRF	SRF
			North Pacific Equatorial Countercurrent Province 2	TARA_138_DCM	DCM
				TARA_138_MES	MES
				TARA_138_SRF	SRF

DCM, deep chlorophyll maximum layer; MES mesopelagic zone; SRF, surface water layer

S2 Table. Relative species abundance in tide pool, seawater and freshwater metagenomes. Taxonomic profiles were generated by FOCUS v.0.29 against all completed genomes of the RefSeq database using default parameters.

Table available only online.

S3 Table. Relative abundance of functional categories in tide pool, seawater, and freshwater metagenomes. Functional annotation was performed using SUPER-FOCUS v.0.25 against the SEED database using RAPSearch2 aligner and default parameters.

Table available only online.

CONSIDERAÇÕES FINAIS

Os resultados apresentados nesse presente trabalho foram importantes para descrever a microbiota em ambientes aquáticos, até então, inéditos. Ambos resultados foram os primeiros estudos metagenômicos para cada ambiente. Apesar de ambos os ambientes possuírem um grande potencial biotecnológico, não haviam trabalhos para o rio Paraguaçú, rio que possui grande parte de seu curso na Caatinga, e para as poças de maré, ambiente que sofre com a escassez de nutrientes e outros fatores.

Em ambos os trabalhos desenvolvidos foi possível observar a influência do ambiente sobre a microbiota, atuando sobre o potencial funcional da comunidade microbiana. Genes vinculados a condições adversas foram observados em situações de stress nutricional, luminosidade, aumento na concentração de íons em geral. Por exemplo, no rio Paraguaçú foi possível observar uma maior abundância de genes relacionados à degradação de pesticidas (como o Benzoato), enquanto nas poças de maré foi possível observar uma maior abundância de genes relacionados à tolerância a ambientes com alta concentração salina. Assim, esses ambientes selecionam seus habitantes pelo seu recurso funcional, e não pela espécie em si.

Novos estudos devem ser realizados para ambos ambientes. Para o estudo realizado no rio Paraguaçú ainda é necessário investigar quais são as funções das proteínas desconhecidas, categoria mais abundante entre as demais do nível 1 do banco de dados do SEED. Além disso, genomas serão montados a partir de metagenomas. Em relação ao estudo nas poças de maré, será realizado um estudo de metatranscriptoma para diferentes períodos, assim, podendo realizar um monitoramento do momento em que os genes, principalmente, relacionados a stress são transcritos.

ARTIGOS E CAPÍTULOS PUBLICADOS DURANTE O DOUTORADO

1. Lopes FAC, Catão ECP, Santana RH, Cabral A de S, Paranhos R, Rangel TP, et al. Microbial Community Profile and Water Quality in a Protected Area of the Caatinga Biome. PLoS ONE. 2016;11: e0148296. doi:10.1371/journal.pone.0148296

2. Catão ECP, Lopes FAC, Rubini MR, Nardoto GB, Prosser JI, Krüger RH. Short-term impact of soybean management on ammonia oxidizers in a Brazilian savanna under restoration as revealed by coupling different techniques. Biol Fertil Soils. 2016; 1–12. doi:10.1007/s00374-015-1086-0

3. Santana RH, Catão ECP, Lopes FAC, Constantino R, Barreto CC, Krüger RH. The Gut Microbiota of Workers of the Litter-Feeding Termite Syntermes wheeleri (Termitidae: Syntermitinae): Archaeal, Bacterial, and Fungal Communities. Microb Ecol. 2015;70: 545–556. doi:10.1007/s00248-015-0581-z

4. Catão ECP, Lopes FAC, Silva MRSS da, Santana RH, Bustamante MM da C, Krüger RH. Diversity and Function of Bacterial Assemblages in Savanna Vegetation Soils. In: Maheshwari DK, editor. Bacterial Diversity in Sustainable Agriculture. Springer International Publishing; 2014. pp. 147–165. Available: http://link.springer.com/chapter/10.1007/978-3-319-05936-5_6

5. Lopes FAC, Catão, ECP, Araújo, JF, et al. Soil Acidobacterial 16S rRNA Gene Sequences Reveal Subgroup Level Differences between Savanna-Like Cerrado and Atlantic Forest Brazilian Biomes, Soil Acidobacterial 16S rRNA Gene Sequences Reveal Subgroup Level Differences between Savanna-Like Cerrado and Atlantic Forest Brazilian Biomes. Int J Microbiol Int J Microbiol. 2014;2014, 2014: e156341. doi:10.1155/2014/156341, 10.1155/2014/156341

6. Monteiro VN, Steindorff AS, Almeida FB dos R, Lopes FAC, Ulhoa CJ, Félix CR, et al. Trichoderma reesei Mycoparasitism against Pythium ultimum is coordinated by G-alpha Protein GNA1 Signaling. J Microb Biochem Technol. 2014;2015. doi:10.4172/1948-5948.1000173

7. Qualhato TF, Lopes FAC, Steindorff AS, Brandão RS, Jesuino RSA, Ulhoa CJ. Mycoparasitism studies of Trichoderma species against three phytopathogenic fungi:

evaluation of antagonism and hydrolytic enzyme production. Biotechnol Lett. 2013;35: 1461–1468. doi:10.1007/s10529-013-1225-3

8. Geraldine AM, Lopes FAC, Carvalho DDC, Barbosa ET, Rodrigues AR, Brandão RS, et al. Cell wall-degrading enzymes and parasitism of sclerotia are key factors on field biocontrol of white mold by Trichoderma spp. Biol Control. 2013;67: 308–316. doi:10.1016/j.biocontrol.2013.09.013

ARTIGOS E CAPÍTULOS NO PRELO

[3]Draft genome sequence of *Cylindrospermopsis raciborskii* (Cyanobacteria) strain ITEP-A1 isolated from a Brazilian semi-arid freshwater body: evidence of saxitoxin and cylindrospermopsin synthetase genes

Running title: Draft genome sequence of a filamentous heterocyst-forming cyanobacterium

Adriana Sturion Lorenzi[1,2], Genivaldo Gueiros Z. Silva[2], Fabyano Alvares Cardoso Lopes[3], Mathias Ahii Chia[1,4], Robert A. Edwards[2,5,6], and Maria do Carmo Bittencourt-Oliveira[1,*]

[1]Laboratory of Cyanobacteria, Department of Biological Sciences, Luiz de Queiroz College of Agriculture, University of São Paulo (USP), Piracicaba, SP, Brazil, [2]Computational Science Research Center, San Diego State University, San Diego, California, United States of America, [3]Laboratory of Enzymology, Department of Cell Biology, University of Brasília (UNB), Brasília, DF, Brazil, [4]School of Marine and Atmospheric Sciences, Stony Brook University, Southampton Campus, New York, United States of America, [5]Department of Computer Science, San Diego State University, San Diego, California, United States of America, [6]Division of Mathematics and Computer Science, Argonne National Laboratory, Argonne, Illinois, United States of America

* Corresponding author: mbitt@usp.br

[3] Artigo aceito para publicação pela Genome Announcements

[4]Title: An agile functional analysis of metagenomic data using SUPER-FOCUS

Authors: Genivaldo Gueiros Z. Silva[1], Fabyano A. C. Lopes [4], and Robert A. Edwards[1,2,3*]

[1]Computational Science Research Center, [2]Department of Biology, and [3]Department of Computer Science, San Diego State University, 5500 Campanile Drive, San Diego, CA 92182, USA, [4] Cellular Biology Department, Universidade de Brasília (UnB), 700910-900 Brasília, DF, Brazil.

*For correspondence please contact Dr. Robert Edwards at redwards@mail.sdsu.edu.

Summary

One of the main goals in metagenomics is to identify the functional profile of a microbial community from unannotated shotgun sequencing reads. Functional annotation is important in biological research because it enables researchers to identify the abundance of functional genes of the organisms present in the sample, answering the question, "What can the organisms in the sample do?" Most currently available approaches do not scale with increasing data volumes, which is important because both the number and lengths of the reads provided by sequencing platforms keep increasing. Here, we present SUPER-FOCUS, SUbsystems Profile by databasE Reduction using FOCUS, an agile homology based approach using a reduced reference database to report the subsystems present in metagenomic datasets and profile their abundances. SUPER-FOCUS was tested with real metagenomes, and the results show that it accurately predicts the subsystems present in the profiled microbial communities, is computationally efficient, and up to 1,000 times faster than other tools. SUPER-FOCUS is freely available at http://edwards.sdsu.edu/SUPERFOCUS.

Keywords

bioinformatics, metagenomics, functional profiling, agile tool, sensitive, SEED

[4] Capítulo aceito para publicação pela Methods in Molecular Biology

[5]Título: Saúde do rio Paraguaçú dentro e fora do Parque Nacional da Chapada Diamantina (PNCD) avaliada por meio de qualidade de água e metagenômica

Fabyano Alvares Cardoso Lopes

Departamento de Biologia Celular, Programa de Pós-graduação em Biologia Microbiana, Universidade de Brasília, CEP 70919-970, Brasília, DF, Brasil. e-mail: fabyanoalvares@gmail.com

Elisa Catão Caldeira Pires

Departamento de Biologia Celular, Programa de Pós-graduação em Biologia Microbiana, Universidade de Brasília, CEP 70919-970, Brasília, DF, Brasil. e-mail: elisaccp@gmail.com

Renata Henrique Santana

Programa em Ciências Genômicas e Biotecnologia, Universidade Católica de Brasília, CEP 70790-160, Brasília, DF, Brasil. e-mail: renatarhs@gmail.com

Anderson de Souza Cabral

Departamento de Biologia Marinha, Universidade Federal do Rio de Janeiro, CEP 21941-599, Rio de Janeiro, RJ, Brasil. e-mail: andersonscabral@yahoo.com.br

Rodolfo Paranhos

Departamento de Biologia Marinha, Universidade Federal do Rio de Janeiro, CEP 21941-599, Rio de Janeiro, RJ, Brasil. e-mail: rodolfo@biologia.ufrj.br

Thiago Pessanha Rangel

Laboratório de Ciências Ambientais, Universidade Estadual do Norte Fluminense Darcy Ribeiro, CEP 28013-602, Campos dos Goytacazes, RJ, Brasil, e-mail: thiprangel@yahoo.com.br

Carlos Eduardo de Rezende

Laboratório de Ciências Ambientais, Universidade Estadual do Norte Fluminense Darcy Ribeiro, CEP 28013-602, Campos dos Goytacazes, RJ, Brasil, e-mail: crezende@uenf.br

Cristiane C. Thompson

Departamento de Biologia Marinha, Universidade Federal do Rio de Janeiro, CEP 21941-599, Rio de Janeiro, RJ, Brasil. e-mail: thompsoncristiane@gmail.com

Ricardo Henrique Krüger

Departamento de Biologia Celular, Programa de Pós-graduação em Biologia Microbiana, Universidade de Brasília, CEP 70919-970, Brasília, DF, Brasil. e-mail: kruger@unb.br

Fabiano L. Thompson # autor correspondente

Departamento de Biologia Marinha, Universidade Federal do Rio de Janeiro, CEP 21941-599, Rio de Janeiro, RJ, Brasil. e-mail: fabianothompson1@gmail.com

[5] Capítulo aceito no livro organizado pelo CNPq/ICMBio