



UNIVERSIDADE FEDERAL DA GRANDE DOURADOS
FACULDADE DE CIÊNCIAS BIOLÓGICAS E AMBIENTAIS
BIOTECNOLOGIA

Rafael Correia da Silva

ANÁLISE METAGENÔMICA DA RIZOSFERA DE *Campomanesia adamantium*
(Cambess) O. Berg

DOURADOS-MS

2016



UNIVERSIDADE FEDERAL DA GRANDE DOURADOS
FACULDADE DE CIÊNCIAS BIOLÓGICAS E AMBIENTAIS
BIOTECNOLOGIA

Rafael Correia da Silva

ANÁLISE METAGENÔMICA DA RIZOSFERA DE *Campomanesia adamantium*
(Cambess) O. Berg

Trabalho de conclusão de curso apresentado como parte dos requisitos para obtenção do título de Bacharel em Biotecnologia, no Curso de Graduação em Biotecnologia da Universidade Federal da Grande Dourados.

Orientador: Prof. Dr. Rodrigo Matheus Pereira

DOURADOS-MS

2016

Dados Internacionais de Catalogação na Publicação (CIP).

S586a Silva, Rafael Correia Da

ANÁLISE METAGENÔMICA DA RIZOSFERA DE *Campomanesia adamantium* (Cambess) O. Berg / Rafael Correia Da Silva, Rodrigo Matheus Pereira -- Dourados: UFGD, 2017.

64f. : il. ; 30 cm.

Orientador: Rodrigo Matheus Pereira

TCC (Graduação em Biotecnologia) - Faculdade de Ciências Biológicas e Ambientais, Universidade Federal da Grande Dourados.

Inclui bibliografia

1. Guavira. 2. Ecologia microbiana. 3. 16S rRNA targeted sequencing. I Rodrigo Matheus Pereira II. Título.

Ficha catalográfica elaborada automaticamente de acordo com os dados fornecidos pelo(a) autor(a).

©Direitos reservados. Permitido a reprodução parcial desde que citada a fonte.

RAFAEL CORREIA DA SILVA

**ANÁLISE METAGENÔMICA DA RIZOSFERA DE *Campomanesia adamantium*
(Cambess) O. Berg**

Trabalho de conclusão de curso apresentado como parte dos requisitos para obtenção do título de Bacharel em Biotecnologia, no Curso de Graduação em Biotecnologia da Universidade Federal da Grande Dourados.

Aprovado em dezessete de outubro de 2016.

BANCA EXAMINADORA

Prof. Dr. Rodrigo Matheus Pereira – Universidade Federal da Grande Dourados



Dr. Luciano Takeshi Kishi – Universidade Estadual Paulista “Júlio de Mesquita Filho”

Prof^ª. Dr^ª. Liliam Silvia Candido – Universidade Federal da Grande Dourados

Prof^ª. Dr^ª. Danielle Marques Vilela – Universidade Federal da Grande Dourados

Ao meu pai, Osmar Correia da Silva, que durante todos esses anos foi e continua sendo a fonte absoluta de inspiração, criatividade e determinação.

AGRADECIMENTOS

Ao orientador, Prof. Dr. Rodrigo Matheus Pereira, do qual emanou a determinação para desenvolvimento deste e de outros trabalhos, através de sua paciência infinita e conhecimento profundo, características do verdadeiro mestre.

À minha namorada, Maria Fernanda Zaneli Campanari, pelo apoio incondicional nos momentos altos e baixos do desenvolvimento do mesmo, pela paciência nos momentos de ausência – é a responsável por inspirar determinação e criatividade através do poder do amor.

Aos meus familiares por serem a pedra angular de todo meu desenvolvimento humano, através do maior de todos os valores – a família, do qual emana o valor do aprendizado e do trabalho duro.

Aos meus amigos da vida acadêmica, Thalles Kraus, Lucas Guarnier, Anderson Souza, Ramir Bavaresco e Mikael Pelluzi, que enriqueceram a experiência da graduação através da amizade.

Ao meu amigo Gabriel de Luccas Ruiz, pela manutenção conjunta do ideal da ciência como objetivo final da existência humana, o qual foi por mim inspirado e me inspirou a enveredar nos caminhos relativos à tecnologia da informação.

À professora Maricy Bonfá, por sua colaboração ao nos ensinar os protocolos relativos à bancada, sem a qual esse experimento não seria possível.

Agradeço também às forças determinadoras do universo, sejam elas relativas a Deus ou não, pois este trabalho é uma dádiva da minha existência, por mais improvável que ela seja perante um universo tão incomensuravelmente grande.

Agradeço também aos colegas do grupo de pesquisa, em especial Suzana Sato e Karina Shirakawa, que muito inspirou através de seu trabalho árduo.

Agradeço à FUNDECT, que através do fomento, permitiu a realização deste trabalho.

Agradeço, por fim, ao Laboratório de Bioquímica de Micro-organismos e Plantas da UNESP/Jaboticabal.

Sou eternamente grato a todos.

RESUMO

O objetivo do presente trabalho foi a análise de bactérias e arqueias da rizosfera através da análise metagenômica da rizosfera de *Campomanesia adamantium* (Cambess) O. Berg... O conhecimento da taxonomia dos micro-organismos mais abundantes presentes nesse ambiente pode nortear a formulação de bioinoculantes, que são compostos de células de micro-organismos capacitados a promover o crescimento de uma planta-alvo, através da fixação de nitrogênio, aquisição de nutrientes e competição. Além disto, o conhecimento da função dos micro-organismos nesse ambiente pode gerar dados para a obtenção de novas enzimas microbianas de interesse biotecnológico. Assim, o presente trabalho buscou a identificação taxonômica e funcional dos micro-organismos da rizosfera de *C. adamantium* através da tecnologia de sequenciamento de nova geração e análises de bioinformática. Solos foram coletados no Horto Florestal da UFGD em Dourados – Mato Grosso do Sul, para extração do material genômico, que foi amplificado para a região hipervariável v4-v5 do RNA ribossomal 16S. O DNA foi sequenciado em parceria com o LMSeq da UNESP em Jaboticabal/SP, através do *Ion Semiconductor – Personal Genome Machine* (PGM). Os dados do sequenciamento foram analisados por dois parâmetros de qualidade através de duas metodologias devido a possíveis discrepâncias em análises de bioinformática – a primeira, sugerida pelo *Brazilian Microbiome Project* e a segunda, MICCA – *Microbial Community Analysis* e, por fim, foi feita a predição funcional através do pacote PICRUSt. Foram encontradas diferenças na comunidades de micro-organismos do solo rizosférico e não-rizosférico, elencando-se 15 gêneros únicos de micro-organismos superabundantes na região de rizosfera, dos quais 05 foram associados em literatura como fixadores de nitrogênio e 03 associados à promoção do crescimento em plantas. Foram preditas 04 proteínas associadas à fixação de nitrogênio e 03 proteínas de interesse biotecnológico abundantes na rizosfera de *C. adamantium*. As análises diferenciadas de bioinformática geraram diferenças na obtenção destes resultados. Espera-se que com estes novos dados sobre essa planta do Cerrado ainda inexplorada, possa se incentivar o estabelecimento de guavirais e a exploração mais intensa do potencial biotecnológico da guavira.

Palavras-chave: Guavira, ecologia microbiana, *16S rRNA targeted sequencing*

ABSTRACT

This work's goal was to explore the biotechnological potential of the plant *Campomanesia adamantium* (Cambess) O. Berg by analyzing the bacteria and archaea community inhabiting its rhizosphere (zone under the influence of the plant roots). Knowledge of these bacterium taxonomy may lead to posterior formulating of a bioinoculante, a composition of bacterial cells that are able to promote plant growth through nitrogen-fixation, nutrient acquisition and competition. Besides, knowing these bacteria function could generate data for obtaining new microbial enzymes with biotechnological application potential. Therefore, the present work sought to identify the taxonomy and predict the functional of the microorganisms living in the rhizosphere of *C. adamantium* through next-generation sequencing and bioinformatics analysis. Soils were sampled from Horto Florestal (Orchard) at UFGD/Dourados – Mato Grosso do Sul, from which the genomic DNA was extracted. The DNA were sequenced in partnership with LMSeq – UNESP/Jaboticabal, by Ion Semiconductor's Personal Genome Machine. The data was analyzed in two parameters of quality control and in two algorithmic methodologies due to possible discrepancies between bioinformatics analysis – the first one was BMP, Brazilian Microbiome Project and MICCA, Microbial Community Analysis. Lastly, the functional prediction was made through the PICRUST package. Differences between the rhizosphere soil and no soil were found (q -value $< 0,05$), listing 15 bacteria and archaea genera who were superabundant in the rhizosphere region. Of these, 05 were associated in literature as nitrogen fixing and 03 associated with growth promotion by other means. 04 predicted proteins related to nitrogen fixation were found and 03 other proteins of biotechnological interest were abundant in *C. adamantium*'s rhizosphere. The multiple analysis of the same dataset generated statistically significant difference between each other. Hopefully, this new data could motivate the establishment of new guavirais and deepen it's biotechnological exploration.

Keywords: Guavira, microbial ecology, 16S rRNA target sequencing

LISTA DE ILUSTRAÇÕES

Figura 1 – O efeito rizosfera e o fenômeno da rizodeposição. Adaptado de (BROCHIER-ARMANET; MOREIRA, 2015).....	22
Figura 2 – Decréscimo dos custos de sequenciamento em relação aos anos 2002-2014. Adaptado de (PEVSNER, 2009).	24
Figura 3 – Rarefação das OTUs em relação ao número de <i>reads</i> , para cada um dos programas testados por Albanese e colaboradores (2015).	28
Figura 4. Curva de rarefação para as diferentes <i>pipelines</i> . De cima para baixo: BMP P1, BMP P2, MICCA P1 e MICCA P2.	41
Figura 5. Análise de componentes principais, gerada através do pacote QIIME. RZ – amostra de rizosfera, NR – amostra não-rizosférica.	43
Figura 6. Micro-organismos superabundantes no solo de rizosfera (laranja), em comparação ao solo não-rizosférico (azul). Gerado através do programa STAMP, <i>pipeline</i> BMP, parâmetro P1.	44
Figura 7. Micro-organismos superabundantes no solo de rizosfera (laranja), em comparação ao solo não-rizosférico (azul). Gerado através do programa STAMP, <i>pipeline</i> BMP, parâmetro P2.	46
Figura 8. Micro-organismos superabundantes no solo de rizosfera (laranja), em comparação ao solo não-rizosférico (azul) a nível de filo. Gerado através do programa STAMP, <i>pipeline</i> BMP, parâmetro P1.....	47
Figura 9. Micro-organismos superabundantes no solo de rizosfera (laranja), em comparação ao solo não-rizosférico (azul). Gerado através do programa STAMP, <i>pipeline</i> MICCA, parâmetro P1.....	49
Figura 10. Micro-organismos superabundantes no solo de rizosfera (laranja), em comparação ao solo não-rizosférico (azul). Gerado através do programa STAMP, <i>pipeline</i> MICCA, parâmetro P2.....	50
Figura 11. Micro-organismos superabundantes no solo de rizosfera (laranja), em comparação ao solo não-rizosférico (azul) a nível de filo. Gerado através do programa STAMP, <i>pipeline</i> MICCA, parâmetro P1.....	51
Figura 12. Predição funcional relativa aos micro-organismos do solo de rizosfera (laranja) e de solo não-rizosférico (azul). Gerado através do pacote PICRUST, com parametrização P1. ...	54

LISTA DE TABELAS

Tabela 1. Alguns inoculantes microbianos comercialmente disponíveis*	19
Tabela 2. Análise química dos solos associado a rizosfera (RZ) e não-rizosférico (NR).	38
Tabela 3. Número e porcentagem de sequências aceitas a diferentes níveis de cortes de qualidade.....	39

LISTA DE ABREVIATURAS E SIGLAS

ACC	1-Aminociclopropano-1-carboxilato
BIOM	<i>Biological Observation Matrix</i>
BMP	<i>Brazilian Microbiome Project</i>
DNA	<i>Desoxyribonucleic Acid</i>
DRBs	<i>Deleterious rhizobacteria</i>
EMP	<i>Earth Microbiome Project</i>
HTS	<i>High-throughput screening</i>
MICCA	<i>Microbial community analysis</i>
NGS	<i>Next-generation sequencing</i>
NTs	Nucleotídeos
OTU	<i>Operational taxonomic unit</i>
PICRUSt	<i>Phylogenetic investigation of communities by reconstruction of unobserved states</i>
PGPR	<i>Plant-growth-promoting rhizobacteria</i>
SAR	<i>Systemic acquired resistance</i>
RNA	<i>Ribonucleic acid</i>
rRNA	<i>Ribosomal ribonucleic acid</i>

Sumário

INTRODUÇÃO.....	13
1 REVISÃO DE LITERATURA.....	16
1.1. C. adamantium e o potencial biotecnológico da biodiversidade de plantas e micro-organismos.....	16
1.2. O potencial da Guavira (Campomanesia adamantium) e a exploração do microbioma da rizosfera	19
1.3. Bioinformática – origens até as tecnologias de sequenciamento de nova geração (NGS)	23
1.4. A abordagem metagenômica para estudo de comunidades de micro-organismos	25
1.5. Controle de qualidade e determinação de parâmetros de 16S	30
2 HIPÓTESE.....	31
3 OBJETIVO GERAL	31
4 MATERIAL E MÉTODOS	32
4.1. Coleta dos Solos – Rizosfera de Guavira (R) e Não-rizosférico (NR).....	32
4.2. Extração do DNA metagenômico e controle de qualidade do material genético extraído	32
4.3. Sequenciamento das regiões hipervariáveis do 16S ribossomal	33
4.3.1. Pipeline de análise BMP - Brazilian Microbiome Project.....	33
4.3.2. Pipeline de análise MICCA – Microbial Community Analysis	34
4.3.3. Análises de Bioinformática –Predição de proteína PICRUSt (v.1.0.0).....	35
4.3.4. Análises de Bioinformática – Validação estatística pelo software STAMP	36
5 RESULTADOS E DISCUSSÕES	37
5.1. Análise química dos solos	37
5.2. Controle de qualidade das reads	39
5.3. Análise taxonômica do 16S: BMP e MICCA.....	42

5.4. Predição funcional por reconstrução de estados não-observados: PICRUSt	53
6 CONCLUSÃO	56
REFERÊNCIAS BIBLIOGRÁFICAS	57
Apêndice A – Micro-organismos abundantes incluindo os não-classificados até gênero pela pipeline MICCA, parâmetro P2	63

INTRODUÇÃO

Os micro-organismos que habitam os solos tem papel fundamental na manutenção de sua estrutura, especialmente quando estes micro-organismos estão associados às plantas em diversos tipos de relações, benéficas, neutras ou detrimenais, executadas por uma diversidade de espécies presentes nestes ambientes.

As plantas obtêm elementos essenciais (como nitrogênio e enxofre) diretamente através de sua interação com o solo, através da região denominada de rizosfera, que acomoda uma diversidade de interações entre a raiz do vegetal e os micro-organismos presentes em sua região de influência.

A principal interação com os micro-organismos presentes é de natureza mutualística – estes oferecem seus serviços de fixação de nitrogênio em troca de condições fisiológicas ótimas na região da rizosfera e fornecimento de carbono, que é abundante o hospedeiro.

Mas esta interação é apenas um dos benefícios que os micro-organismos podem oferecer às plantas – eles também possuem papeis na transferência de outros nutrientes, competem com outros organismos patogênicos promovendo a saúde da planta ou, ainda, promovem o crescimento direto da raiz do vegetal pela produção de fitohormônios.

A comunidade microbiana promotora de benefícios – no caso do presente trabalho, de bactérias e arqueias, pode ser explorada para fins biotecnológicos, baseando-se na definição clássica deste termo – um conjunto de técnicas e práticas que tem por escopo o desenvolvimento de produtos e serviços a fins comerciais e sociais.

Isto pode ocorrer através de um produto chamado de bioinoculante, que se consiste em uma solução líquida composta de células de micro-organismos conhecidamente benéficos à saúde da planta, promotores de seu crescimento. Para isto, é necessário conhecer quais micro-organismos se associam às raízes da planta, determinando-se potenciais candidatos à produção de um bioinoculante.

Assim, no presente trabalho, explorou-se os micro-organismos presentes na rizosfera de *Campomanesia adamantium* (Cambess) O. Berg, também conhecida como guavira ou gabirola, um fruto muito conhecido na região do Cerrado brasileiro e em Mato Grosso do Sul, tanto por

seu sabor único, sua utilização em produtos industrializados e a exploração por suas propriedades medicinais, com o objetivo de incentivar o cultivo deste fruto.

As técnicas para o conhecimento e documentação das bactérias e arqueias são geralmente associadas às técnicas clássicas de microbiologia, baseadas em isolamento e cultivo. Entretanto, conhece-se que muitos micro-organismos não são cultiváveis, o que seria uma desvantagem desta metodologia. Assim, no presente trabalho, foi utilizado um conjunto de metodologias baseado no sequenciamento do material genético do ambiente da rizosfera – técnica comumente descrita como metagenômica.

Este conjunto de técnicas, associadas ao domínio da bioinformática, permite a obtenção da classificação taxonômica dos micro-organismos presentes no solo de rizosfera e com base nesta informação, sua predição funcional, podendo determinar-se a identidade dos micro-organismos presentes no ambiente e também sua função em potencial.

Como forma de identificar àqueles micro-organismos exclusivos da região da rizosfera de *C. adamantium*, também analisou-se o solo do mesmo ambiente, mas não associado a rizosfera, sendo possível determinar os micro-organismos e as funções preditas mais e menos abundantes no ambiente de rizosfera.

Por fim, as metodologias de análise de metagenômica são por vezes discrepantes, pois existem diferentes abordagens para a execução deste tipo de análise. No presente trabalho, foram comparadas duas rotinas de trabalho de bioinformática, ou seja, duas formas distintas do processamento dos dados genômicos, mas de mesma finalidade.

Através destas metodologias, buscou-se determinar quais a taxonomia dos principais micro-organismos abundantes na região de rizosfera de três plantas de *C. adamantium*, e qual o seu suposto papel nesse ambiente, com o objetivo de determinar micro-organismos alvo para isolamento e formulação de novos bioinoculantes e potenciais proteínas ou enzimas de interesse biotecnológico presentes na rizosfera.

O conhecimento da comunidade de bactérias e arqueias presentes na rizosfera de *C. adamantium* poderá incentivar o estabelecimento e desenvolvimento de guavirais, pois as informações fornecem uma base genômica para o desenvolvimento de bioinoculantes para a planta *C. adamantium* e outras representantes deste gênero, além de elencar micro-organismos abundantes, com potencial para isolamento de novas biomoléculas de interesse humano.

Assim, a determinação do perfil dos micro-organismos e a predição funcional dos mesmos estimulariam a produção desse fruto regional, através de novas informações sobre

possíveis produtos para o incremento da produção (bioinoculantes) como produtos oriundos da planta (bioprospecção de bactérias e arqueias de interesse para a obtenção de biomoléculas).

1 REVISÃO DE LITERATURA

A estrutura desta revisão está disposta em quatro partes. A primeira, discorrerá sobre a função do presente trabalho de forma geral, definindo-o no escopo da Biotecnologia através da exploração da rizosfera, descrevendo-se inclusive o que está disponível em literatura para a planta alvo do presente trabalho.

A seguir, será descrita a natureza das relações da rizosfera, seus principais fundamentadores teóricos e sua importância estratégica como alvo para a obtenção de novos produtos e serviços através de diversas estratégias - promoção de crescimento das plantas, resistência a doenças, adição de tolerância a estresse abiótico, entre outros benefícios.

A parte terceira se refere à fundamentação de nossa metodologia, a bioinformática, uma estratégia altamente tecnológica e promissora para a geração de um grande volume de dados voltados aos nossos objetivos.

Na quarta e última parte, se descreve a Metagenômica (também chamada de "target-sequencing", em nosso caso, pois a análise foi realizada para um gene-alvo), detalhando-se as escolhas executadas, inclusive buscando-se autores que se utilizaram de metodologias similares e exploraram problemas similares ao explorado pelo presente trabalho.

1.1. *C. adamantium* e o potencial biotecnológico da biodiversidade de plantas e micro-organismos

O Brasil é hospedeiro de uma grande diversidade de plantas cujas frutas com características nutricionais e sensoriais são muito peculiares. O consumo destes vegetais tem crescido muito em mercados nacionais e internacionais devido à disseminação do conhecimento de seu valor nutricional e sua relação à saúde humana. Essas frutas, chamadas de tropicais não-convencionais, foram definidas através como uma fonte em potencial de moléculas bioativas, que possuiriam interesse econômico e social (CÂNDIDO; SILVA; AGOSTINI-COSTA, 2015).

Neste contexto, as plantas e frutos do cerrado brasileiro possuem perspectivas de exploração biotecnológica por diversos motivos, sendo a principal perspectiva o teor considerável de compostos com aplicação terapêutica, como vitaminas, antocianinas, carotenoides e compostos fenólicos (CAMPOS et al., 2012).

Um desses frutos, a guavira, oriundo do planta *C. adamantium* (Cambess.) O. Berg, é utilizado na produção de licores, sucos, doces, sorvetes e é utilizado como medicamento popular como um antirreumático, antidiarreico, hipocolesterolêmico e anti-inflamatório (COUTINHO et al., 2009). O extrato hexânico das folhas se mostrou adequado para a obtenção de compostos antioxidantes (RAMOS; CARDOSO; YAMAMOTO, 2008). *C. adamantium* apresentou, ainda, atividade antiiperalgésica e antidepressiva em roedores, sem qualquer toxicidade associada (DE SOUZA et al., 2016).

Apesar destas características, a planta de *C. adamantium*, típica do bioma brasileiro de Cerrado e abundante no Mato Grosso do Sul (LORENZI, 2006), ainda não há estudos de seu microbioma..

Neste ponto, é interessante definir o que é uma aplicação biotecnológica, visto que o presente trabalho abrange uma monografia em Biotecnologia, para então contextualizar o estudo do microbioma. Segundo a referência (FALEIRO; ANDRADE; REIS, 2011), Biotecnologia seria a união conceitual de biologia com tecnologia, consistindo-se em uma série de técnicas, processos e práticas para o desenvolvimento de produtos e serviços com função econômica e social.

A mesma referência ainda cita as seguintes técnicas e práticas como pertencentes ao escopo da Biotecnologia: produção de fármacos, transformação genética e uso de micro-organismos na agricultura. Essas técnicas, no contexto de exploração da biodiversidade de uma planta, teria por finalidade a obtenção de produtos como antibióticos, vitaminas e potenciais bactérias fixadoras de nitrogênio.

Produtos naturais sempre tiveram um papel fundamental no desenvolvimento da medicina contemporânea e ainda executam um enorme papel na descoberta de novos fármacos.

A vasta diversidade genética disponível em plantas e micro-organismos se constituem em um tesouro de possibilidades para o avanço das condições humanas na produção de alimentos, materiais e fármacos. Apesar disto, apenas uma pequena fração do reino Plantae foi analisado para a obtenção de seus potenciais usos terapêuticos (MUKHERJEE, 2015).

Muito da atividade das plantas é dependente ou está associada à sua população de micro-organismos: o microbioma. As plantas possuem um microbioma diferenciado em cada uma de suas regiões - rizosfera (raiz), caulosfera (caule), filosfera (folha), antosfera (flor), carposfera (fruto) e endosfera (semente). Alguns dos micro-organismos são detrimenais às

plantas - os chamados DRBs, *deleterious rhizobacteria*. Outros, são neutros ou possuem uma relação equilibrada não-detrimental.

Mas há aqueles que beneficiam ativamente as plantas, são os PGPRs, *Plant-growth-promoting rhizobacteria*, que podem ser aplicadas diretamente no desenvolvimento das plantas, como um produto: o inoculante microbiano, bioinoculante ou biofertilizante.

Este produto consiste em um formulado composto de micro-organismos benéficos a planta, que pode ser aplicado ao solo, raízes ou sementes, incrementando-se a disponibilidade de nutrientes para a planta através do metabolismo inerente das células bacterianas ali presentes.

As características principais dos biofertilizantes são a capacidade de fornecer às raízes da planta alguns nutrientes, como nitrogênio, fósforo e ferro, através dos mecanismos de fixação biológica de nitrogênio, solubilização de fosfato e sideroforia – solubilização de ferro (PYLRO et al., 2014) .

Entretanto, os inoculantes microbianos podem exercer ainda outros papéis, como por exemplo a fitoestimulação, que se consiste na modulação hormonal da planta através da produção de fitohormônios microbianos, como auxinas, ACC (1-Aminociclopropano-1-carboxilato) deaminase, citocininas e giberelinas, que dependendo-se das condições ambientais podem estimular intensamente o desenvolvimento da planta. Os agentes microbianos podem, ainda, ser utilizados no manejo de fitopatógenos, pois algumas estirpes possuem atividade antibiótica, inibindo a ação de outros micro-organismos patogênicos presentes naquele ambiente (DHANANJAYA PRATAP; HARIKESH BAHADUR, 2016).

Este produto deve ser desenvolvido a partir de micro-organismos específicos, cuja identidade pode ser obtida de várias formas, mas também através da metodologia descrita neste trabalho, configurando-se assim uma etapa de uma aplicação biotecnológica a um problema pontual, da forma como foi definida pelo presente trabalho – através do oferecimento de um serviço.

Existem muitos inoculantes microbianos disponíveis no mundo, compostos de estirpes puras em meio de cultura ou de um consórcio de diferentes micro-organismos. Entretanto, ainda existem poucos inoculantes desenvolvidos para a família das Myrtaceae (como é o caso de *C. adamantium*), e especialmente, para a espécie em análise no presente trabalho, que adota uma estratégia de bioinformática para a prospecção de novas estirpes de compostos microbianos.

A **tabela 1** a seguir exhibe alguns inoculantes microbianos baseados em sistemas bacterianos disponíveis comercialmente.

Tabela 1. Alguns inoculantes microbianos comercialmente disponíveis*

Nome do produto	Empresa responsável	Ingrediente ativo
Bactofil A10	Agro. Bio Hungria	<i>Azospirillum brasilense</i> , <i>azotobacter vinelandii</i> , <i>Bacillus megaterium</i> , <i>Bacillus polymyxa</i> , <i>Pseudomonas fluorescens</i> , <i>Streptomyces albus</i> , entre outros
Bactvipe	International Panacea Ltd., India	<i>Pseudomonas fluorescens</i> <i>Azotobacter vinelandii</i> -B 1795, <i>Bacillus megaterium</i> B1091, <i>Clostridium</i> <i>pasteurianum</i> ,
Microbion UNC	Syn-Bio-Tech Ltd, Hungria	<i>Azospirillum</i> sp., <i>Bacillus</i> <i>subtilis</i> , <i>Rhodobacter</i> sp., <i>Lactobacillus</i> sp., <i>Trichoderma reesei</i> , <i>Saccharomyces cerevisiae</i> , <i>Streptomyces</i> sp.
Nodulator XL	BASF	<i>Rhizobium leguminosarum</i> biovar <i>viceae</i>
Vault SP	BASF	<i>Bradyrhizobium</i> sp.
Poncho/Votivo	Bayer	<i>Bacillus firmus</i>
Jumpstart	Novozymes BioAg	<i>Penicillium bilaiae</i>

*Adaptado de (DHANANJAYA PRATAP; HARIKESH BAHADUR, 2016).

1.2. O potencial da Guavira (*Campomanesia adamantium*) e a exploração do microbioma da rizosfera

Os micro-organismos que habitam as plantas e a região em que habitam são chamados de microbioma, que é considerado uma extensão do genoma do vegetal. Eles possuem um papel

determinante na saúde e produtividade do hospedeiro, podendo estabelecer associações benéficas, neutras ou detrimenais (BERENDSEN, PIETERSE, BAKKER, 2012).

Algumas interações entre micro-organismos e plantas modelo são bem descritas em literatura, como as associações entre rizóbios e leguminosas, mas a maioria dos microbiomas e suas contribuições ao hospedeiro ainda não são completamente elucidados (TURNER; JAMES; POOLE, 2013).

Dos distintos microbiomas, interessa a este trabalho o que se encontra abaixo do solo - a rizosfera. Este termo se refere ao compartimento de solo influenciado pelas raízes da planta, considerado um *hotspot* das interações e atividades microbianas, que são fundamentais em diferentes funções, mas em duas principais: captação de nutrientes e proteção do ataque por patógenos (BAKKER et al., 2013).

Os micro-organismos de solo associados à raiz são muito mais numerosos que aqueles presentes fora da raiz, e isto ocorre devido à enorme presença de compostos carbônicos na região da raiz, que serve como fonte de nutriente para uma grande diversidade de espécies ali presentes.

As regiões do microbioma podem ser manipuladas para atender a uma finalidade econômica ou social, através da redução da incidência de doenças no vegetal, incrementada produtividade, alteração da estrutura química e redução da emissão de gases causadores de efeito estufa.

O microbioma de uma planta é um determinante crucial para sua sobrevivência e produtividade, devido às diversas associações de mutualismo que as plantas realizam com os micro-organismos próximos, ativamente selecionando aqueles benéficos a si, e suprimindo na medida do possível os patogênicos (TURNER; JAMES; POOLE, 2013).

Micro-organismos se associam às plantas em diferentes composições, e já se verificou que a estrutura das comunidades de micro-organismos varia em relação ao microambiente em que se localizam – o estudo de SOUZA e colaboradores verificou que, na cana-de-açúcar, a estrutura da comunidade de bactérias, arqueias e fungos varia para cada microambiente da planta – raiz, talo, colmos e folhas (SOUZA et al., 2016).

O microambiente associado à raiz é chamado de rizosfera, e se consiste na região do solo proximalmente associada às raízes da planta e que está sujeito à influência radicular (BAKKER et al., 2013).

Os micro-organismos do solo não-associado à raiz competem pelos escassos nutrientes que estão disponíveis no solo, e assim, sua proliferação normalmente é baixa, mas quando os micro-organismos estão associados a raiz eles demonstram uma intensa atividade, pois as plantas realizam a exsudação, também conhecida com rizodeposição, e se consiste no acúmulo de compostos biologicamente ativos – secreção de íons, oxigênio livre, água, enzimas, mucilagem, além de uma diversa variedade de compostos carbônicos categorizados em metabólitos primários e secundários (BAIS et al., 2006; DOORNBOS; VAN LOON; BAKKER, 2012).

Assim, os micro-organismos se aproveitam dessa estrutura de nutrientes para seu crescimento, e retornam benefícios às plantas, como por exemplo a nodulação, que auxilia enormemente na fixação de nitrogênio, a solubilização de minerais que contem fósforo, aumentando sua disponibilidade, e antagonismo em relação aos micro-organismos patogênicos (BAIS et al., 2006; BAKKER et al., 2013; TURNER; JAMES; POOLE, 2013). Tais micro-organismos benéficos são chamados de RCPC ou em inglês PGPR – rizobactérias promotoras de crescimento em plantas. (EGAMBERDIEVA; SHRIVASTAVA; VARMA, 2015)

Essa adaptação das populações de PGPR à região da rizosfera, assim como os efeitos de seleção desses indivíduos pela planta e amplificação dessas seletas populações é denominado de efeito rizosfera (BROCHIER-ARMANET; MOREIRA, 2015).

Observa-se esse efeito na **Figura 1**. Os rizodepósitos correspondem aos exsudatos, secreções e substâncias oriundas da lise celular, além do mucigel – uma camada que origina da planta, composta especialmente de mucopolissacarídeos, que são intensamente atraentes aos micro-organismos.

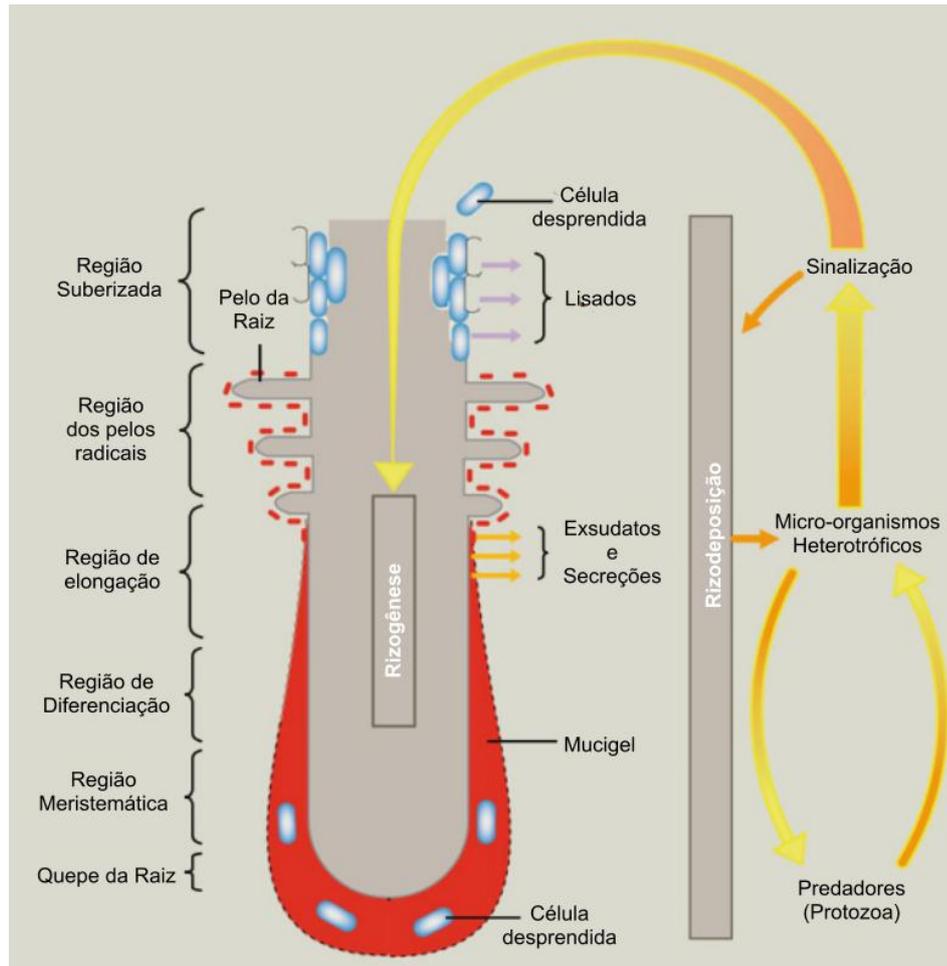


Figura 1 – O efeito rizosfera e o fenômeno da rizodeposição. Adaptado de (BROCHIER-ARMANET; MOREIRA, 2015).

É evidente que nem sempre tais interações são estritamente benéficas as plantas. As chamadas DRBs (*deleterious rhizobacteria*) são detrimenais ao bem-estar das plantas, podendo inibir o crescimento de raiz e caule sem causar outros sintomas visuais, e podem ser responsáveis pela redução do rendimento em cultivares em regime de monocultura contínua (PINTON; VARANINI; NANNIPIERI, 2007).

Estima-se que exista 4×10^7 bactérias por grama de solo (DANIEL, 2005). Essa imensa quantidade de bactérias permite que a biotecnologia utilize algumas bactérias para biofertilização, fitoestimulação, biocontrole e fitoremediação.

Essa intensa população de micro-organismos do solo possuem um papel fundamental na manutenção dos ciclos biogeoquímicos, além de influenciar na formação da estrutura do solo, decomposição da matéria orgânica, além de agir como um indicador da condição do solo, devido às relações entre diversidade microbiana, qualidade do solo e da planta e sustentabilidade do ecossistema (EGAMBERDIEVA; SHRIVASTAVA; VARMA, 2015).

A exploração da diversidade de micro-organismos de solo foi limitada pelas técnicas clássicas de microbiologia, devido ao fato de que a maioria deles não é cultivável. (NESME et al., 2016) Dessa dificuldade surgiram as tecnologias de sequenciamento dos ácidos nucleicos – em especial, a metagenômica, que estuda a diversidade total de micro-organismos existentes em um ambiente, podendo ser descrita como uma nova disciplina localizada na bioinformática, a fronteira entre a tecnologia da informação e as ciências da vida, que revolucionaram o acesso ao dilúvio de informações advinda da diversidade dos micro-organismos de solo. (NESME et al., 2016).

1.3. Bioinformática – origens até as tecnologias de sequenciamento de nova geração (NGS)

Uma das primeiras definições de Bioinformática, e também uma das mais completas, é baseada no artigo de revisão clássico de LUSCOMBE e colaboradores (2001), que definem a Bioinformática como a conceptualização de dados biológicos, em termos de moléculas, e a aplicação de métodos baseados em tecnologia de informação para compreender e organizar tais dados biológicos em larga escala (LUSCOMBE; GREENBAUM; GERSTEIN, 2001).

Tal definição ainda hoje é muito precisa, no sentido de se englobar uma região interdisciplinar entre as ciências da vida, as ciências da computação e as tecnologias da informação. O nascimento dessa disciplina é associado à década de 1960 – época em que, há poucos anos, Frederick Sanger havia publicado os primeiros sequenciamentos de proteína – ocorrendo-se em seguida o desenvolvimento de modelos computacionais para análise de sequências de proteínas, como os programas baseados na linguagem de programação Fortran para determinar a estrutura de uma proteína a partir de dados de bancada, por Margaret Dayhoff (WÜNSCHIERS, 2013).

Assim, o crescimento exponencial dos dados de sequências de moléculas biológicas cresceu exponencialmente a partir da década de 80, quando as tecnologias de sequenciamento passaram a se tornar amplamente disponíveis (KANEHISA; BORK, 2003), e daí em diante, o termo bioinformática passaria a ser usado de uma maneira mais específica – a utilização de computadores para obtenção, processamento, análise e simulação de informações biológicas. (RAMSDEN, 2015).

Com o advento das tecnologias de sequenciamento de DNA de nova geração e a consequente geração de um grande volume de dados, é dito que a biologia quantitativa moderna

é inundada por um dilúvio de dados – isto é, um número incomensurável de informações é disponibilizado diariamente, que muito supera a capacidade humana de processamento (GREENE et al., 2016; PEVSNER, 2009).

Os avanços dessas tecnologias se justificam pelo desenvolvimento de novas e mais eficientes plataformas de NGS que têm, simultaneamente, custo cada vez menor e capacidade de geração de dados cada vez maior, o que permite cada vez mais uma intensificação em pesquisas que investigam a natureza da vida (MASOUDI-NEJAD; NARIMANI; HOSSEINKHAN, 2013).

Na **Figura 2**, observa-se o decréscimo recente dos custos do sequenciamento. O custo por megabase é relativo a uma sequência de DNA com o mínimo de qualidade Q20, ou 01 erro a cada 100 nucleotídeos enquanto o tamanho do genoma é relativo ao genoma humano. Observa-se que o sequenciamento de um milhão de bases, para o custo mais recente, sairia por menos de 10 centavos de dólar, enquanto que o genoma de um indivíduo custaria menos de dez mil dólares.

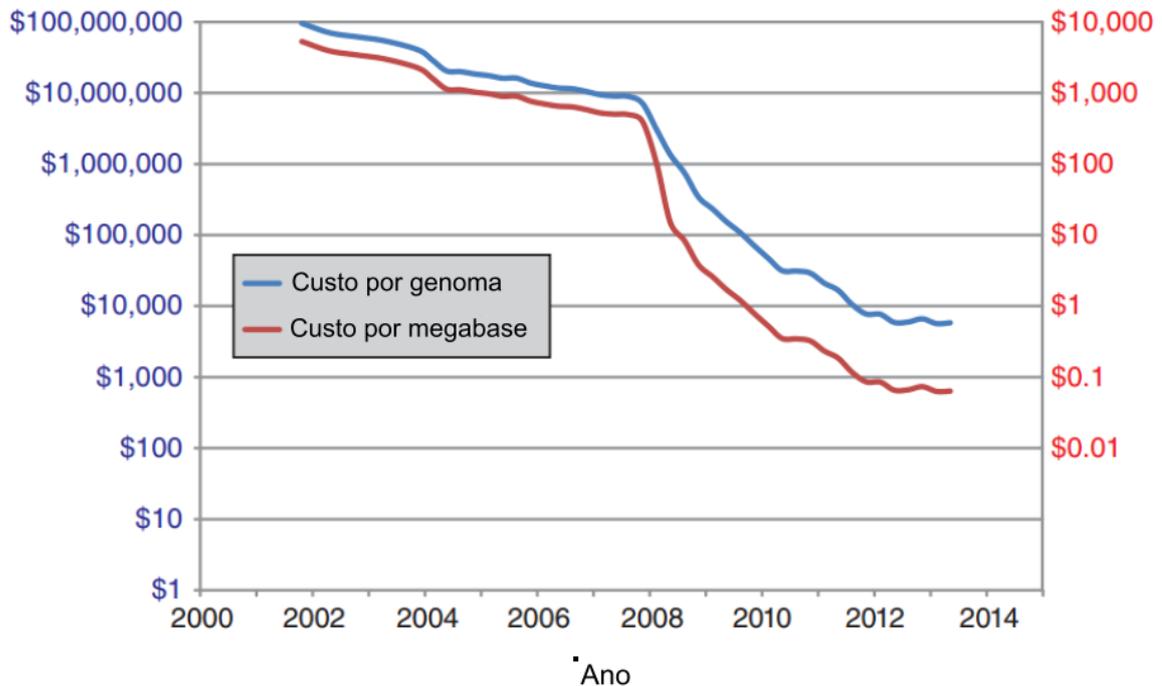


Figura 2 – Decréscimo dos custos de sequenciamento em relação aos anos 2002-2014. Adaptado de (PEVSNER, 2009).

Uma das metodologias de sequenciamento de próxima geração é a Ion Semicondutor ou Sequenciamento Ion Torrent (ThermoFisher), que ao contrário das outras metodologias não utiliza uma reação enzimática, nem detecção fluorescente ou ótica.

O método de detecção de nucleotídeo é baseado em síntese, isto é, o DNA-alvo é linearizado em fita simples, e conforme a fita dupla de DNA é confeccionada pela DNA-polimerase, ocorrem incorporações de nucleotídeos àquela fita, seguido da liberação de um íon de hidrogênio (H^+) que é detectado, pois altera o pH do meio. Essa reação ocorre de forma intensamente paralela ocorre em um microchip, onde deve ser depositado o DNA. (MASOUDI-NEJAD; NARIMANI; HOSSEINKHAN, 2013).

1.4. A abordagem metagenômica para estudo de comunidades de micro-organismos

A literatura descreve que apenas 1% dos micro-organismos que habitam o planeta Terra são cultiváveis pelas metodologias clássicas de microbiologia. Isso significa que, se embasando apenas nas técnicas da microbiologia clássica, a diversidade estrutural e funcional dos micro-organismos é intensamente limitada – e assim, também, todo o potencial para a produção de fármacos e novos produtos industriais permanece inexplorado (STREIT; ROLF, 2010)

Nesse contexto, o nascimento e crescimento vertiginoso da metagenômica nas últimas duas décadas resultou do intenso poder dessa técnica, que tem a capacidade de análise dos micro-organismos de um ambiente, incluindo aqueles não cultiváveis, pelo conhecimento das sequências de DNA oriundos dos genomas dos micro-organismos que estão naquele ambiente (ALLAN, 2014).

A metagenômica como disciplina se define como o estudo de micro-organismos amostrados diretamente de um ambiente, sem necessidade de cultivo. (SHARON; BANFIELD, 2013). Os autores discutem se o sequenciamento que tem por alvo somente o gene 16S pode ser definida como técnica de metagenômica, sugerindo, a essas aplicações o termo *Microbiome-seq* (sequenciamento do microbioma), pois nessa definição metagenômica seria apenas aplicações aos genes de uma amostra (KUMAR, 2014; YE, 2016).

Outros autores, entretanto, se referem às análises de ambiente como análises metagenômicas, independentemente da utilização de sequenciamento direcionado. (LANGILLE et al., 2013). De qualquer forma, considera-se que o objetivo de uma

caracterização metagenômica é responder a duas perguntas fundamentais – quem são os organismos, e qual sua função naquele ambiente. (SCHOLZ; LO; CHAIN, 2012)

O rápido decréscimo e contínuo desenvolvimento de melhorias nas plataformas de sequenciamento (volume de dados e qualidade da saída de bases) da última década permitiu também o nascimento da disciplina de metagenômica e, em especial, a aplicação de sequenciamento de alto desempenho ao gene RNA ribossomal (rRNA) 16S. (JU; ZHANG, 2015).

A compreensão das comunidades de micro-organismos – bactérias e arqueias – em termos de taxonomia é costumeiramente alcançada através do sequenciamento dos genes relativos à sub-unidade ribossomal 16S, relativa ao rRNA 16S (MORI; MARUYAMA; KUROKAWA, 2010).

O RNA ribossomal 16s é um componente da pequena subunidade 30S dos ribossomos procarióticos, sendo uma região considerada altamente conservada ou seja, de baixíssimas taxas de mutação, sendo assim utilizado para identificação em detrimento de outras metodologias tradicionais consideradas menos confiáveis, como a coloração de Gram, análises bioquímicas e testes fenotípicos, devido ao fato de que essas análises são dependentes de cultivo e de crescimento do micro-organismo alvo (WOESE; FOX, 1977; WOO et al., 2008).

O uso de um gene de uma cópia para a análise de uma comunidade de micro-organismos foi uma conquista importante para a ecologia de micro-organismos, devido à sua função essencial, ubiquidade e propriedades relativas à evolução. E assim, as análises relativas às sub-unidades ribossomais pequenas são fundamentais ao estudo da ecologia de micro-organismos, sendo considerado o padrão ouro para análises de natureza filogenética e de estudos da comunidade de micro-organismos (CASE et al., 2007; WALTERS et al., 2016).

O gene 16S é comparado a outros genes já sequenciados, e a partir de um certo nível de similaridade (geralmente de 97 a 98%) é possível dizer que ambos pertencem a uma mesma OTU (unidade taxonômica operacional). A propósito práticos, uma unidade de OTU representaria a identificação de um grupo discreto de micro-organismos extremamente comuns entre si, como um documento de identidade dos mesmos. Esse método é considerado imensamente útil para a classificação de bactérias, mas é um método que é intensamente criticado por seu baixo poder filogenético para identificar bactérias a nível de espécie e baixo poder discriminante a nível de gênero (DELONG et al., 2014).

A análise das comunidades microbianas através das metodologias metagenômica baseadas no 16S costumeiramente ocorrem em três etapas fundamentais – (1) o pré-tratamento dos dados brutos oriundos do sequenciamento, (2) análises da diversidade de micro-organismos e (3) análises avançadas dos dados e metodologias de visualização. (XIA et al., 2013).

Tais metodologias para as análises foram estabelecidas pelo desenvolvimento de alguns projetos colaborativos de larga escala, como o EMP, *Earth Microbiome Project* (GILBERT; JANSSON; KNIGHT, 2014), cuja proposta é a de catalogar a diversidade de micro-organismos não cultiváveis do planeta e o *Human Microbiome Project* (AAGAARD et al., 2013; METHÉ et al., 2012) que de maneira similar, tem por proposta a criação de um conjunto de dados de referência dos micro-organismos de diversas regiões do corpo humano com o objetivo de analisar os efeitos da microbiota sobre a saúde humana.

Esses projetos, além de avançarem os esforços para a caracterização das comunidades de micro-organismos de ambientes diversificados também colaboraram para a padronização dos três passos citados acima – que garantem que a diversidade e a variabilidade entre amostras é atribuída ao ambiente, tentando reduzir as etapas de contaminação da amostragem oriunda da metodologia, especialmente a de análise dos dados no ambiente da bioinformática (YE, 2016).

Devido a linearidade em relação aos passos necessários para uma análise completa do ambiente por metagenômica, e da complexidade dos conjuntos de dados baseados em NGS, uma estratégia indicada para análise de sequências são os *pipelines*, ou canalização, na qual a saída de um programa é a entrada do próximo programa (ALBANESE et al., 2015).

A necessidade da realização das análises em diversas *pipelines* é relativa ao fato de que por vezes, a metodologia de análise oferece resultados discrepantes e opções mais completas, podendo se atribuir méritos a cada uma dessas análises – com frequência, quando uma nova *pipeline* é publicada, seus autores aplicam metodologias para comparar sua performance em relação a outras *pipelines*.

Nesse contexto, é necessário referenciar o trabalho de Albanese e colaboradores (2015), desenvolvedores da *pipeline* MICCA. Para verificação da performance de sua *pipeline* em relação às outras de funcionamento similar, os autores desenvolveram um metagenoma artificial – isto é, um conjunto de dados relativos a um ambiente com uma parametrização totalmente conhecida – de forma que fosse possível comparar a performance do programa com uma simulação da realidade (ALBANESE et al., 2015).

A comparação da rarefação do metagenoma artificial em relação aos diferentes programas de bioinformática pode ser visualizada na **Figura 3**. A técnica do metagenoma artificial para determinação da acurácia dos parâmetros é chamada também de *mock*, sendo utilizada para estudos de *benchmark*, ou seja, determinação de níveis de referência para posterior utilização dos mesmos parâmetros.

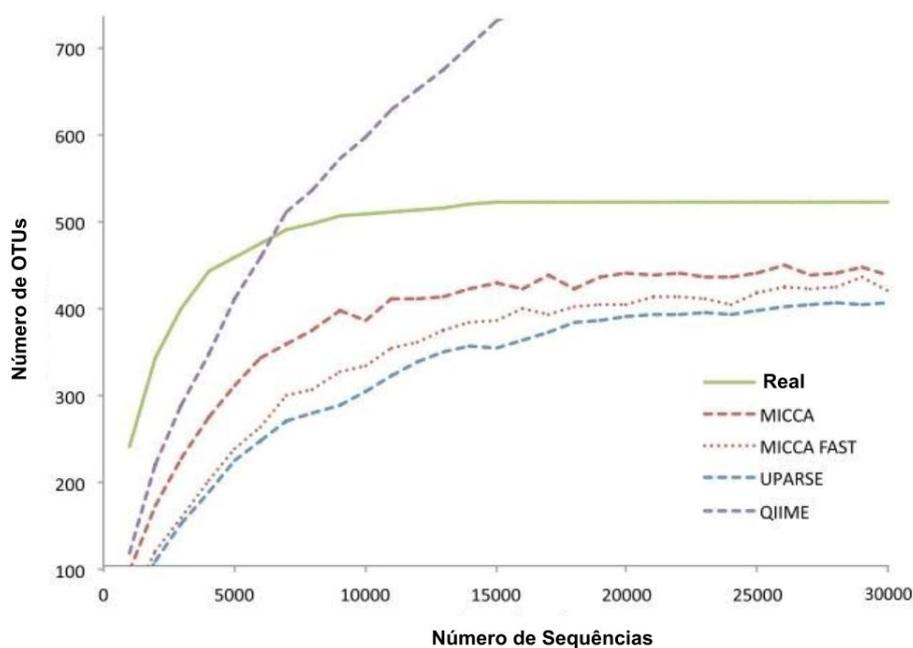


Figura 3 – Rarefação das OTUs em relação ao número de *reads*, para cada um dos programas testados por Albanese e colaboradores (2015).

Em análises de natureza genômica (como a metagenômica), a etapa de sequenciamento e a tecnologia utilizada para isso são fundamentais para a garantia de dados de alta qualidade, que inclusive tenham conseguido assimilar dados que realmente consigam representar suficientemente o ambiente em toda a sua diversidade.

A diversidade em diferentes níveis de resolução é classificada como α , β , e γ -diversidade, na qual a α -diversidade representa a diversidade de micro-organismos em um habitat, enquanto a β -diversidade representa a mudança na composição de espécies em relação a um gradiente, e a γ -diversidade representa a diversidade de micro-organismos em uma região que compreende muitos diferentes habitats. (EGAMBERDIEVA; SHRIVASTAVA; VARMA, 2015), sendo que em análises de 16S rRNA são utilizados principalmente os parâmetros para definição da α e β -diversidade.

No caso da α -diversidade, o teste realizado tenta verificar se toda a diversidade do hábitat foi alcançada pelo sequenciamento, gerando-se uma curva chama curva de rarefação que tenta exaurir uma propriedade de diversidade (como diversidade filogenética, número de OTUs ou outro parâmetro misto, como o *chao1*) em relação ao número de *reads* (sequências).

Assim, um resultado ideal para qualquer sequenciamento seria o alcance do platô ao final do eixo das abcissas, o que significa que por mais que se adicionem sequências à análise, a propriedade de diversidade permanece a mesma – por interpolação, assume-se, assim, que a diversidade continuaria tendendo ao platô, independentemente do número de sequencias.

Na Fig. 4; é possível perceber que, apesar do metagenoma artificial ter sido desenhado para o alcance do platô, nem todas as *pipelines* conseguiram alcançar tal resultado – em especial, a *pipeline* do QIIME conseguiu descobrir um número maior de OTUs, sugerindo-se que o sequenciamento foi incompleto, o que não seria possível pois o metagenoma artificial possui um número determinado de OTUs – ou seja, houve uma super-representação de algumas das OTUs. Por outro lado, a *pipeline* dos autores (MICCA) conseguiu alcançar o platô, mas não conseguiu caracterizar o número total de OTUs, havendo assim uma caracterização aquém da esperada.

Um estudo de comparação de *pipelines* realizado por Plummer e colaboradores (2015) demonstra que a atribuição das OTUs a nível de filo é estatisticamente insignificante, mas é notável a nível de gênero. Segundo os autores do estudo, isso não é necessariamente devido à contaminação oriunda da metodologia, mas do fato de que o marcador 16S rRNA é quase idêntico para os micro-organismos muito proximamente relacionados, não sendo possível distingui-los imediatamente apenas com esse marcador (PLUMMER; TWIN, 2015).

Nas análises do trabalho foram utilizadas duas *pipelines* – a BMP, *Brazilian Microbiome Project* (PYLRO et al., 2014) e a MICCA (ALBANESE et al., 2015), além do pacote para predição funcional pICRUST (LANGILLE et al., 2013).

A metodologia principal utilizada às análises de bioinformática foram realizadas de acordo com a parametrização sugerida pelo *Brazilian Microbiome Project*, BMP. O projeto original do BMP tem a ambição de estabelecer um consórcio e bancos de dados relativos aos projetos de metagenômica de ambientes brasileiros, pois considera-se que a biodiversidade de micro-organismos no Brasil ainda é imensamente desconhecida. A associação do BMP permitirá a comparação entre os microbiomas, compreendendo interações entre micro-organismos de mesmo nicho (PYLRO et al., 2014).

1.5. Controle de qualidade e determinação de parâmetros de 16S

O resultado dos sequenciadores automáticos – as sequências em formato digital – não tem qualidade constante ou previsível. Em relação ao sequenciamento da subunidade ribossomal 16S, Edgar e colaboradores (2010) exploraram muito o problema do controle de qualidade das sequências oriundas de NGS. A partir deste ponto, as sequências de DNA serão chamadas de *reads*, ou leituras, aquilo que o sequenciador foi capaz de obter.

O parâmetro de qualidade padrão para análise de sequências é o score ou pontuação *Phred*. Esta pontuação se refere à dois valores: P_i , que seria a probabilidade de se encontrar um erro em um dado nucleotídeo i em uma sequência (informação armazenada no formato *FASTq*) e Q , que se refere à pontuação em relação à P . A relação se dá pela seguinte expressão:

$$Q = -10\log_{10}P_i$$

A taxa de erros típica das máquinas atuais varia de 0,1% até 10% (EDGAR, 2013). No caso da tecnologia de sequenciamento do presente trabalho, são aceitos valores de qualidade superiores a $Q = 20$; o que representaria 1% de erros de acordo com a expressão acima. Logo, este foi o parâmetro utilizado.

Alguns exemplos de referência ao controle de qualidade e determinação de um parâmetro mínimo podem ser apreciados a seguir: em uma análise microbiômica, relativa ao estudo da comunidade fúngica, foi utilizado o controle *Phred* $Q > 20$ (KEMLER et al., 2013). Em um sequenciamento do genoma de *Salmonella entérica*, foi utilizado o controle de qualidade *Phred* $Q > 15$ (DEN BAKKER et al., 2014). Para outro microbioma, obtido de resíduos líquidos, também foi utilizado o parâmetro de corte *Phred* $Q > 15$ (BIBBY; PECCIA, 2013). Em um painel para identificação de SNPs executado através de sequenciamento Ion Torrent foi utilizado também o *Phred* $Q > 20$ (ZHANG et al., 2017).

2 HIPÓTESE

- Hipótese nula (H_0): Não existem diferenças entre os micro-organismos associados à rizosfera de *C. adamantium* (RZ) e residentes no solo local sem associação (NR);
- Hipótese alternativa (H_1): Existem diferenças entre os micro-organismos associados à rizosfera de *C. adamantium* (RZ) e aqueles residentes no solo local (NR).

3 OBJETIVO GERAL

Investigar as bactérias e arqueias associadas à rizosfera de três plantas de *C. adamantium*, especialmente aquelas mais abundantes em relação ao solo local não-rizosférico, estabelecendo a existência de diferença entre os solos e o potencial biotecnológico associado à essa diferença;

3.1 OBJETIVOS ESPECÍFICOS

- Determinar as bactérias e arqueias mais abundantes do solo em rizosfera de *C. adamantium*, obtendo-se sua taxonomia e revisando-se em literatura informações para elencar potenciais micro-organismos para desenvolvimento de inoculante;
- Inferir a função em potencial das bactérias e arqueias em análise, especialmente as mais abundantes em relação ao solo não-rizosférico, determinando-se potenciais proteínas e enzimas de interesse;
- Comparar duas metodologias de bioinformática, acentuando-se as discrepâncias geradas e associando as diferenças analíticas que levaram a isso.

4 MATERIAL E MÉTODOS

4.1. Coleta dos Solos – Rizosfera de Guavira (R) e Não-rizosférico (NR)

Os solos foram coletados no Horto Florestal da Universidade Federal da Grande Dourados (GPS: -22,195805; -54.935606).

Três plantas de *C. adamantium* foram escavadas até sua raiz pivotante, a uma profundidade de 20cm, de onde foi coletado um pequeno volume de solo para cada planta com o auxílio de uma espátula – este solo foi denominado rizosférico (RZ).

Também foi coletada uma amostra adicional de solo do mesmo local, mas com uma distância de 1m de cada planta que teve seu solo de raiz coletado. O objetivo desta coleta é o de determinar a presença de micro-organismos inexistentes na rizosfera, ou seja, aqueles que não sofrem a influência da raiz da planta – este solo foi denominado não-rizosférico (NR).

O material amostrado foi armazenado após a coleta em tubos cônicos esterilizados tipo “falcon” e armazenado em freezer a uma temperatura de -20°C até a extração do material genético.

O solo local também foi coletado para a realização de uma análise química, cujo objetivo foi o de determinar os principais nutrientes (P, K, Al, Ca, Mg), a acidez, atividade de argila e fertilidade. Para isto, foi coletado um volume aproximado de 0,5L de solo do sítio de onde se encontram as plantas, e realizadas as análises químicas.

4.2. Extração do DNA metagenômico e controle de qualidade do material genético extraído

Três amostras de cada solo tiveram seu DNA genômico extraído utilizando-se o kit comercial *PowerSoil® DNA Isolation Kit* (MO BIO Laboratories, EUA), seguindo-se o protocolo de extração conforme as recomendações do fabricante (MOBIO, 2014). O mesmo foi feito para o material genético NR, totalizando-se duas amostras extraídas.

De maneira geral, o kit de extração funciona após a adição de 5g de solo a um tubo, seguida de uma série de reações e centrifugações, em que ocorre a remoção dos inibidores da reação em cadeia da polimerase (PCR), homogeneização da amostra do ambiente (no caso,

solo), lise celular mecânica e química e adesão o DNA à uma membrana de sílica, que é então lavado e eluído, estando pronto para aplicações (*downstream*).

Realizou-se um teste preliminar com o DNA extraído para se determinar a concentração de DNA e o sucesso do protocolo de extração, através de eletroforese em gel de agarose e por espectrofotômetro UV-Vis *BioDrop* (BIODROP Ltd).

Após a confirmação da concentração de DNA, ele foi conservado em gelo e enviado à Jaboticabal/SP, onde ocorreu o sequenciamento em parceria com a *Facility* do Laboratório Multiusuário Centralizado para Sequenciamento de DNA em Larga Escala e Análise de Expressão Gênica da UNESP (LMSeq).

4.3. Sequenciamento das regiões hipervariáveis do 16S ribossomal

As três amostras de cada tratamento (totalizando-se 06 conjuntos de dados) receberam identificação por *barcodes*, uma sequência de DNA conhecida, de pequeno tamanho, para posterior discernimento. As regiões sequenciadas foram as hipervariáveis v4-v5 da região ribossomal 16S. O material foi sequenciado através da metodologia *Ion Torrent - Personal Genome Machine, PGM* (ThermoFisher).

Para tanto, essas regiões foram amplificadas através dos *primers* 515F (5'-TGTGNCAGCMGCCGCGGTAA-3') e 926R (5'-barcode-CCCCGYCAATTYMTT-3'). Estes *primers* foram descritos por Walters e colaboradores (2016) que os consideraram uma alternativa superior para estudos de rRNA 16S em relação aos constructos direcionados somente à região hipervariável v4.

O preparo para o sequenciamento de acordo com os protocolos fornecidos pelo fabricante (ThermoFisher, 2014), e o controle da qualidade do DNA sequenciado foi realizado através de fluorômetro *Qubit* (Invitrogen) e 2100 *Bioanalyzer* (Agilent). Após a corrida de sequenciamento, os artefatos e ruídos foram removidos computacionalmente, com o *software* AmpliconNoise (QUINCE et al., 2011), gerando-se, assim, sequências no formato “*FASTq*”.

4.3.1. Pipeline de análise BMP - Brazilian Microbiome Project

Após análises de controle de qualidade com o programa FastQC e um utilitário da pipeline MICCA, as sequências em formato FASTq foram convertidas para o formato FASTA.

Em seguida, as sequências foram derreplicadas – isto é, removeram-se cópias em excesso das sequências, documentando-se a quantidade das mesmas. Esses passos foram realizados através dos programas do pacote de *software* USEARCH (EDGAR, 2010).

Posteriormente, as sequências foram organizadas por tamanho, e os *singletons* (sequências com um único representante) foram descartados. O próximo passo foi o de se buscar as unidades taxonômicas operacionais (OTUs), através do *software* UPARSE (EDGAR, 2013). Ainda com o *software* USEARCH, foram removidas as quimeras (sequências de DNA associadas a mais de um organismo), utilizando-se como referências o banco de dados RDP-Gold (WANG et al., 2007).

Em seguida, o pacote QIIME (CAPORASO et al., 2010) foi utilizado para alinhamento das sequências, atribuição das taxonomias às sequências e organização dos táxons de acordo com sua distribuição, além da geração das análises de OTU no formato BIOM (MCDONALD et al., 2012). Foi gerada, também, uma árvore de referência pelo pacote QIIME, além de um arquivo de mapeamento das sequências tratadas em relação às sequências originais não-tratadas.

Por fim, tendo disponibilizado os arquivos de entrada – tabela BIOM, arquivo de mapeamento e árvore de referência, foram realizadas as análises principais (através do comando *core_diversity_analyses.py*), que integra o cálculo da rarefação (α -rarefação), diversidade (β -diversidade) e sumarização dos táxons.

4.3.2. Pipeline de análise MICCA – *Microbial Community Analysis*

Tal como a *pipeline* do BMP, o pacote MICCA também é uma *pipeline* de software de dados metagenômicos de 16S que combina as metodologias de controle de qualidade, *clusterização* das OTUs (através do pacote OTUCLUST), geração da taxonomia e inferência da filogenia de um conjunto de dados. A utilização dessa *pipeline* baseou-se em revisão bibliográfica, no qual o MICCA teve uma melhor performance em relação a algumas características dos programas UPARSE e QIIME, nos quais a BMP é baseada.

Foram executadas também duas análises na *pipeline* MICCA. Na primeira, as sequências *Fastq* foram convertidas ao formato FASTA, e truncadas em 250 pares de bases com erro esperado máximo de 2,5 e também em 400 pares de bases com erros esperados 4,0. As OTUs foram definidas utilizando o parâmetro do MICCA denominado “referência aberta”.

A seguir, ocorreu a derreplicação, *clusterização* baseada em distância, filtragem de quimeras e mapeamento das sequências aos representativos, gerando-se, por fim, o arquivo BIOM, a árvore filogenética e as sequências representativas. Por fim, foram geradas as curvas de rarefação e tabelas de taxonomia.

4.3.3. Análises de Bioinformática –Predição de proteína PICRUSt (v.1.0.0)

Além das metodologias para esclarecimentos quanto à taxonomia dos micro-organismos presentes no ambiente-alvo (rizosfera de *C. adamantium*), foi utilizado também o pacote PICRUSt - *phylogenetic investigation of communities by reconstruction of unobserved states* (LANGILLE et al., 2013).

Como já foi dito, o gene 16S rRNA é adequado para caracterização da composição taxonômica e diversidade filogenética de uma amostra ambiental, mas justamente por ser um marcador (isto é, não codificar proteínas) é impossível uma identificação direta do metabolismo ou outras capacidades funcionais dos organismos em análise.

Assim, o algoritmo do PICRUSt se baseia em uma associação entre filogenia e funcionalidade de biomoléculas, utilizando-se de evidências de que a associação entre os dois é fortemente correlacionada, mesmo que imperfeitamente. Nos melhores casos no qual foi aplicado, o algoritmo foi capaz de inferir de 80 a 90% do conteúdo gênico de um metagenoma previamente estudado.

A capacidade de predição de genes a partir de dados de 16S é considerada de grande avanço para obtenção de proteínas, pois para grandes coleções de metagenoma é relativamente simples a obtenção do perfil baseado na diversidade, o que não era possível a partir de genes marcadores (LANGILLE et al., 2013).

Devido a estratégia do PICRUSt ser associada a uma topologia de árvore em que sempre existe um vizinho próximo, o programa irá classificar automaticamente todas as OTUs, mesmo aquelas pouco descritas, o que pode vir a gerar dados relativos à proteínas muito ubíquas

(ASSHAUE et al., 2015). O PICRUSt se baseia no banco de dados KEGG para a obtenção dos genes e proteínas preditos (KANEHISA et al., 2017).

Para a análise com o PICRUSt, foi necessária a seleção de OTUs através de uma referência-fechada – o banco de 16S rRNA *Greengenes*, gerando-se um arquivo no formato BIOM (*Biological Observation Matrix*). A partir da escolha automatizada de OTUs, foi realizada pelo pacote PICRUSt a normalização por número de cópias para regularização do efeito das OTUs mais super-representadas, e, por fim, gerou-se a predição dos metagenomas também com o pacote PICRUSt.

4.3.4. Análises de Bioinformática – Validação estatística pelo software STAMP

As análises de taxonomia e funcionalidade foram validadas através do pacote gráfico STAMP, que permite a aplicação de testes de hipótese estatísticos para análise das diferenças entre conjuntos de amostras organizados em dois ou mais grupos (PARKS et al., 2014).

Para a verificação do teste de hipóteses tanto para a taxonomia gerada pelo BMP, como pelo MICCA, foi utilizado o teste exato de fisher (*Fisher's exact test*) bicaudal para o cálculo da diferença entre as proporções. A correção bicaudal se deve à hipótese de que a população das sequências não é igual, e assim, a diferença entre elas é diferente de zero. A seguir, aplicou-se o modelo de correção *Benjamini-Hochberg FDR (False Discovery rate)*. Foram consideradas apenas os resultados cujo *q-value* foi menor que 0,05; sendo este um valor mínimo aceitável para este teste.

Por fim, para validação do teste de hipótese da predição funcional (PICRUSt) foi utilizado o teste t de Welch bicaudal para dois grupos (Controle e Guavira) e o teste exato de Fisher bicaudal para duas amostras, que permitem inferir qual função é mais abundante em cada solo, com significância estatística. Foi aplicado também o modelo de correção *Benjamini-Hochberg FDR*, desconsiderando-se os resultados cujo valor de *q-value* foi menor que 0,05.

5 RESULTADOS E DISCUSSÕES

Os resultados estão divididos em cinco partes. Inicialmente, discutir-se-á a natureza química dos solos amostrados, caracterizando-os conforme literatura. Tais características, como será exposto, podem exercer efeitos sobre a comunidade microbiana da rizosfera.

A seguir, serão discutidos os parâmetros de controle de qualidade utilizados, pois são os determinantes da riqueza do conjunto de dados, e em si só formam uma metodologia. Assim, serão discutidos os impactos de algumas técnicas específicas à qualidade da análise.

As partes terceira e quarta se referem às *pipelines* de análise da comunidade microbiana – MICCA e BMP, especialmente aos micro-organismos superabundantes atestados pelo programa STAMP, com significância estatística. Estes abundantes são revistos em literatura, associando-se seu impacto à rizosfera.

A quinta e última parte se refere à predição funcional através do programa PICRUSt, que se baseia na reconstrução de estados não-observados daquele solo para predizer os papéis de cada micro-organismo.

5.1. Análise química dos solos

A **Tabela 2** exhibe os resultados quantitativos determinado pela análise química do solo associado à rizosfera (R) e do não-rizosférico (NR). Na maioria dos solos, a presença de fósforo (P) inorgânico é geralmente em concentrações consideravelmente baixas, sendo que a presença deste mineral é dependente do pH e da presença de outros íons negativamente carregados (ânions) que competem com ligantes com o fósforo. Esse íon, quando comparado aos outros nutrientes essenciais às plantas (nitrogênio, enxofre e potássio) é o que possui menor mobilidade nos solos, na maioria das condições (HINSINGER, 2001).

Além disso, esse fósforo não pode ser utilizado diretamente pela planta, que só absorve o fósforo em suas formas solúveis – os íons monobásicos (H_2PO_4^-) e dibásicos (HPO_4^{2-}). Sendo assim, essa solubilização ocorre pela presença de micro-organismos solubilizadores de fosfato, além de exsudatos orgânicos fosforados que a planta libera através das raízes, o que é consistente com a observação de que a concentração de fósforo foi maior no solo associado à rizosfera.

A mobilização de potássio (K) na rizosfera é considerada de mínima importância, mas sugere-se que as plantas possuam um mecanismo de aquisição do potássio a partir da rizosfera, o que poderia explicar a menor concentração encontrada no solo de rizosfera (PINTON; VARANINI; NANNIPIERI, 2007).

O solo de rizosfera apresentou um pH maior em três das quatro análises (pH por CaCl_2 , por H_2O e por acidez potencial, H+Al) como pode ser visto na tabela 1.

A saturação por bases (V%) é um indicativo da fertilidade do solo, considerando-se distróficos (inférteis) os solos com saturação menor que 50% (RONQUIM, 2010). Em resposta a esse fator externo, já foram descritas respostas a partir do sistema das raízes: o aumento de volume de solo explorado pela raiz (através de alongação, aumento da área de superfície, de densidade), e também um aumento na atividade de micro-organismos relacionados a planta, através da seleção de micro-organismos benéficos e exclusão de patogênicos (RAO et al., 2016).

Tabela 2. Análise química dos solos associado a rizosfera (RZ) e não-rizosférico (NR).

Propriedade analisada	NR	RZ
P (mg/dm^3)	11,90	15,79
K ($\text{cmol}_c/\text{dm}^3$)	0,78	0,42
Al ($\text{cmol}_c/\text{dm}^3$)	0,12	0,12
Ca ($\text{cmol}_c/\text{dm}^3$)	2,80	3,20
Mg ($\text{cmol}_c/\text{dm}^3$)	1,80	1,90
pH CaCl_2	4,22	4,37
pH H_2O	5,03	5,16
pH SMP	5,94	5,84
H+Al ($\text{cmol}_c/\text{dm}^3$)	5,9	6,6
SB ($\text{cmol}_c/\text{dm}^3$)	5,38	5,52
T ($\text{cmol}_c/\text{dm}^3$)	11,28	12,10
V%	47,6	45,7

NR – rizosférico; RZ – solo sob rizosfera; P – fósforo; K – potássio; Al – Alumínio; Ca – cálcio; Mg – magnésio; CaCl_2 – cloreto de cálcio; SMP – método tampão de Shoemaker-McLean-Pratt; H+Al – acidez potencial; SB – bases trocáveis; T – atividade de argila; V% – saturação por bases.

5.2. Controle de qualidade das *reads*

Como foi descrito, antes do início das análises taxonômicas e funcionais é necessário converter as sequências do formato *FASTq* para o formato *FASTA*, removendo-se as *reads* de baixa qualidade. Os programas utilizados para isto (USEARCH e MICCA) realizam o controle de qualidade através do corte por máximo de erros esperados.

Na **tabela 3** constam os dados relativos ao controle de qualidade das sequências. O sequenciamento *Ion Torrent* produziu 240811 *reads*, de comprimento médio 231,1 nucleotídeos e comprimento máximo 515 nucleotídeos. Uma cobertura adequada para a região sequenciada, a 16S ribossomal, é de 250 a 400 nucleotídeos – sendo assim, estas regiões foram analisadas para a cobertura da análise.

Tabela 3. Número e porcentagem de sequências aceitas a diferentes níveis de cortes de qualidade.

Max_ee (nt)	150nt (nt %)		250nt (nt %)		350nt (nt %)		400nt (nt %)	
1.00	75592	31,4%	20854	8,7%	2190	0,9%	668	0,3%
2.00	124146	51,6%	55768	23,2%	13875	5,8%	7013	2,9%
3.00	150095	62,3%	<u>80456</u>	<u>33,4%</u>	28500	11,8%	17339	7,2%
4.00	161260	67,0%	94734	39,3%	40794	16,9%	<u>27794</u>	<u>11,5%</u>

nt – nucleotídeos; % – porcentagem em relação a todas as *reads*. Sublinhados: parâmetros escolhidos para a análise. Max_ee: Parâmetro de erros das *pipelines* BMP e MICCA.

Foram escolhidos, para dar procedência a análise, os conjuntos de parâmetros: 3 erros para *reads* 250 nucleotídeos (parâmetro 1 ou P1) e 4 erros para *reads* de 400 nucleotídeos (parâmetro 2 ou P2). Convertendo-se a qualidade de P1 para o controle de qualidade *phred*, obtêm-se o valor de $Q = 19,21$. Já para o parâmetro P2, o valor de *phred* seria de exatamente $Q = 20$. Estes parâmetros são tão relaxados quanto os revistos em literatura e ainda assim, se aproveitam apenas 33,4% e 11,5% do total dos dados, respectivamente.

Pode-se observar que dados são perdidos a medida que o tamanho da sequência aumenta. Isto se refere às sequências truncadas, que também é uma medida para se obter sequências de alta qualidade sem que ocorram descartes excessivos.

Descreve-se o processo de truncagem das sequências tomando-se a parametrização P1 como exemplo. O processo funciona excluindo-se os *reads* menores que o tamanho delimitado – em nosso caso, 250 nts. A seguir, busca-se no *read* a região de maior qualidade, e verifica-se

se este *read* não ultrapassou o limite de erros (no caso de P1, máximo de 03 erros). Se isto ocorrer, a sequência é descartada. Caso não ocorra, o *software* passa a adicionar nucleotídeos nas duas caudas da sequência (à jusante e à montante) até o tamanho máximo que não ultrapasse o limite de erros.

Outro procedimento também aplicado ao conjunto de dados é o descarte de *singletons*. O termo *singleton* se refere às *reads* cujo conteúdo é mais de 3% divergente de todas as outras *reads*. Este descarte, de acordo com Edgar (2010), reduz um pouco da sensibilidade da análise, mas são um intenso incremento na especificidade da análise. Para a *pipeline* BMP, os *singletons* são remapeados à uma OTU, e a informação não é perdida.

Desta forma, as *reads* foram analisadas quatro vezes – duas *pipelines* em dois parâmetros de controle de qualidade.

Na **figura 4** estão expostas as curvas de rarefação para estas quatro análises. Esta é uma medida ecológica que avalia a riqueza dos táxons presentes em relação à amostragem. O ideal para tal medida é que se alcance o platô, o que significaria que o número de organismos de diferentes gêneros foi exaurido – todos ou grande parte foi encontrada.

Pode-se perceber que este foi o caso para a *pipeline* BMP (dois gráficos superiores), mas não ocorreu de forma igualmente satisfatória para a *pipeline* MICCA (dois gráficos inferiores). Também observa-se que o número de OTUs encontradas (eixo y) é muito superior para a *pipeline* MICCA, e sempre foi maior para o parâmetro do menor tamanho de sequência (P1, truncado em 250 pares de bases).

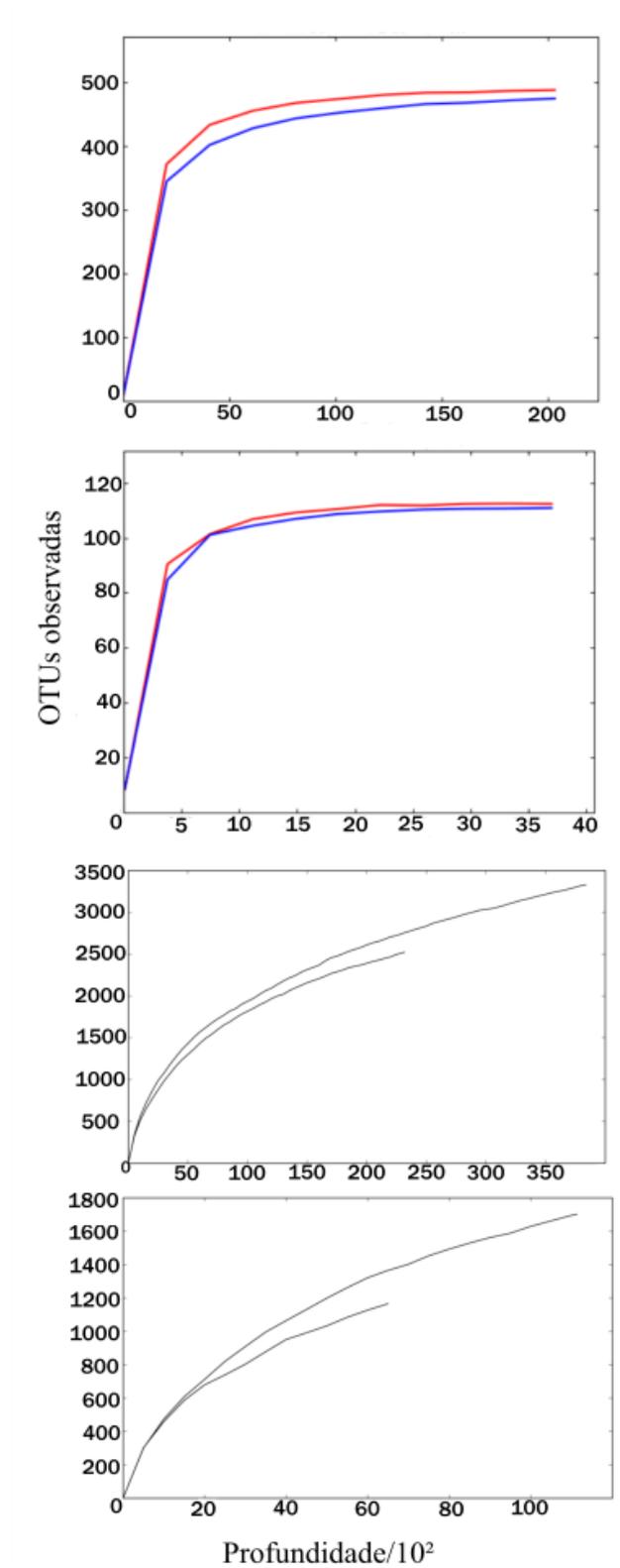


Figura 4. Curva de rarefação para as diferentes *pipelines*. De cima para baixo: BMP P1, BMP P2, MICCA P1 e MICCA P2.

A partir deste ponto, com a parametrização da análise propriamente descrita, se iniciarão os resultados relativos à taxonomia dos micro-organismos, especialmente àqueles mais abundantes.

5.3. Análise taxonômica do 16S: BMP e MICCA

Para visualização do conjunto de dados, as 06 amostras foram dispostas em uma análise de coordenadas principais, que está exposta na **figura 5**. Esta análise foi realizada com os parâmetros P1, através da *pipeline* BMP. Pode-se observar que as amostras associadas ao solo não-rizosférico (NR1 à NR3) estão proximamente agrupadas, enquanto as amostras relativas à rizosfera (RZ1 à RZ3) estão dispersas em relação aos três eixos.

Esta é uma observação consistente com o referencial teórico, que descreve uma diversidade elevada no ambiente de rizosfera, que ocorre devido ao recrutamento dos micro-organismos benéficos pela planta e pela invasão de micro-organismos patogênicos (BERENDSEN; PIETERSE; BAKKER, 2012).

Na seleção destes micro-organismos, o sistema de defesa da planta, conhecido como SAR – *systemic acquired resistance* desempenha um papel importante, pois exerce uma seleção química através dos exsudatos da planta, que tem efeito detrimental aos micro-organismos patogênicos e que tem efeito de promoção dos micro-organismos que estabelecem relações mutualísticas (TURNER; JAMES; POOLE, 2013) .

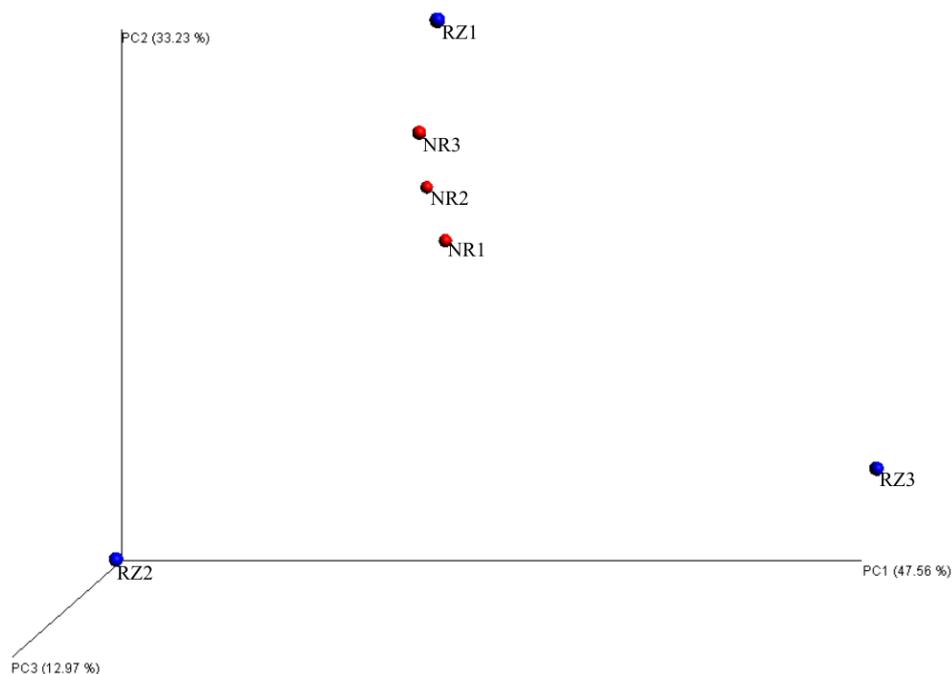


Figura 5. Análise de componentes principais, gerada através do pacote QIIME. RZ – amostra de rizosfera, NR – amostra não-rizosférica.

As três amostras para a rizosfera de guavira (RZ) e solo não-rizosférico (NR) são comparadas, a seguir, através do pacote estatístico STAMP.

Foram omitidos nas próximas ilustrações, relativas aos micro-organismos superabundantes, aqueles não classificados a nível de gênero. Entretanto, estes dados também são parte da análise, e assim, podem ser encontrados no **apêndice único**. Os dados relativos à esses táxons estão contabilizados na normalização dos dados.

A seguir, na **figura 6**, encontram-se os micro-organismos superabundantes para a *pipeline* BMP para o parâmetro de qualidade P1 – tamanho mínimo de 250 nucleotídeos com nível de qualidade ($Q = 19,21$).

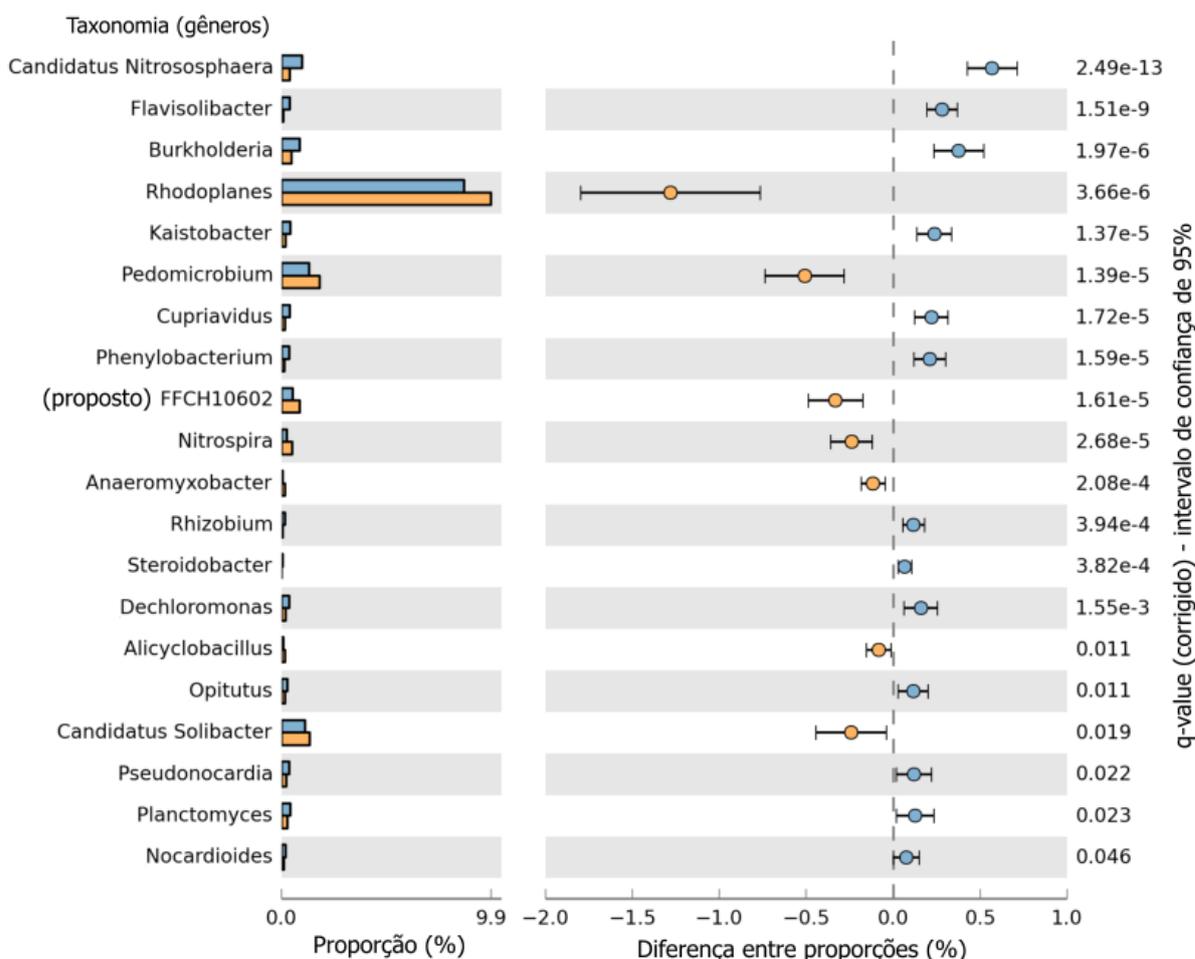


Figura 6. Micro-organismos superabundantes no solo de rizosfera (laranja), em comparação ao solo não-rizosférico (azul). Gerado através do programa STAMP, *pipeline* BMP, parâmetro P1.

Dentre os gêneros encontrados, pode-se observar que 07 foram superabundantes à rizosfera ($q < 0,05$) enquanto 13 foram mais abundantes em solo não-rizosférico ($q < 0,05$). O gênero *Rhodoplanes* representa micro-organismos gram-negativos, geralmente em formato de bastonetes e fototróficos, e que realizam metabolismo de compostos fenólicos como carotenoides – este gênero já foi isolado da região rizosférica de arrozais (SRINIVAS; SASIKALA; RAMANA, 2014).

Os micro-organismos deste gênero possuem afinidade a pH neutro (como foi o caso da rizosfera, que apresentou pH mais neutro em relação ao solo não-rizosférico) e também são documentados como formadores de biofilmes (CAVANAUGH et al., 2014).

Este gênero foi encontrado em superabundância na rizosfera de um arbusto perene, *Caragana jubata*, através também de uma metodologia de análise de 16S através de uma metodologia muito similar à BMP, associando UPARSE e QIIME. Em contraste, este gênero

foi encontrado como menos abundante em rizosfera de carvalho, *Quercus* sp. (UROZ et al., 2010).

O gênero superabundante seguinte para a rizosfera, *Pedomicrobium*, está associado ao metabolismo oxidativo de ferro e manganês – nutrientes importantes ao desenvolvimento da planta, e se utilizam de fontes orgânicas de nitrogênio, geralmente associados a biofilmes microbianos de ambientes aquáticos e marinhos (CAVANAUGH et al., 2014). Este gênero já foi associado como promotor de crescimento na raiz de macieira, em um estudo metagenômico que associou a presença e abundância microbiana em relação a um gradiente de crescimento das árvores (FRANKE-WHITTLE et al., 2015).

O próximo gênero abundante na rizosfera é o proposto FFCH10602, pertencente à classe *Ktedonobacteria* do filo *Chloroflexi*. Este gênero é recém-descrito devido à sua dificuldade de cultivo, mas foi documentado como heterotrófico, degradador de polissacarídeos, e foi obtido em uma cultura pura pela primeira vez em ambiente de solo (YOKOTA, 2012).

O próximo gênero abundante em rizosfera, o *Nitrospira*, é um dos três gêneros pertencentes à família *Nitrospiraceae* do filo Nitrospirae, micro-organismos associados à oxidação de nitritos para nitratos (ALBERS; SIEBERS, 2014). Em uma análise metagenômica de milho associada a um gradiente temporal, este gênero apresentou maior abundância no solo não-rizosférico comparado à rizosfera, e que em ambos os ambientes, a abundância desse gênero diminuiu após seis meses (LI et al., 2014). Este gênero também foi encontrado em uma análise metagenômica de feijão, *Phaseolus vulgaris* (SUYAL et al., 2015).

A seguir, o próximo gênero de micro-organismos abundantes em rizosfera foi o *Anaeromyxobacter*, um micro-organismo anaeróbio facultativo que, quando exerce este metabolismo, se utiliza de compostos como nitrato, fumarato e clorofenólicos como aceptores de elétrons (ROSENBERG et al., 2014a).

Diversas espécies deste gênero foram encontradas em superabundância na rizosfera de arroz, *Oryza sativa* (CABEZAS et al., 2015). Este micro-organismo também foi encontrado em superabundância intensa de junco-solto, *Juncus effusus*, em um estudo que investigou as diferenças nas comunidades microbianas de solo rizosférico e não-rizosférico em uma região sob inundações periódicas (KAPLAN et al., 2016).

O penúltimo gênero superabundante nesta análise, *Alicyclobacillus*, é um gênero de bactérias gram-positivas esporuladas comumente associadas à contaminações por sua capacidade de sobreviver à processos de pasteurização, causando problemas de sabor e odor

em alimentos processados – entretanto, o habitat comum desse gênero é o solo, e o metabolismo de seus representantes não inclui açúcares como fontes de carbono (ROSENBERG et al., 2014b).

Este micro-organismo foi ocorrente na rizosfera de eucalipto geneticamente modificado, *Pinus radiata* D., em um estudo metagenômico que comparou a comunidade de micro-organismo deste em relação ao eucalipto comum. (LOTTMANN et al., 2010).

Por fim, os micro-organismos pertencentes ao gênero *Solibacter* não foram associados a estudos de rizosfera por literatura.

A seguir, a **figura 07** representa a mesma comparação que a figura anterior, porém neste momento, com a parametrização mais rígida (P2), que como já foi descrito, retornou um menor volume de dados relativos às OTUs encontradas nos dois ambientes.

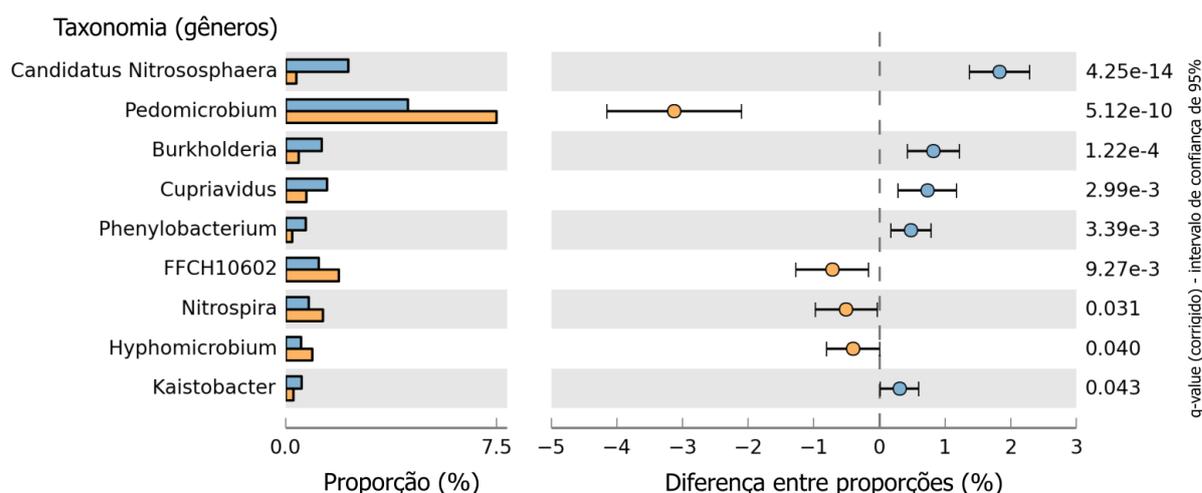


Figura 7. Micro-organismos superabundantes no solo de rizosfera (laranja), em comparação ao solo não-rizoférico (azul). Gerado através do programa STAMP, *pipeline* BMP, parâmetro P2.

Pode-se observar na **figura 07** que alguns dos micro-organismos abundantes na parametrização P1 não foram encontrados, enquanto outros (*Pedomicrobium*, *FFCH10602*, *Nitrospira*) se mantiveram ($q < 0,05$).

Um gênero adicional foi encontrado nesta análise, o *Hyphomicrobium*, também pertencente a classe das alfaproteobactérias, este micro-organismo é amplamente distribuído (considerado ubíquo) e utiliza amônio como fonte de nitrogênio, solubilizando-o até amônia

(CAVANAUGH et al., 2014). Este gênero foi encontrado como abundante em um estudo da rizosfera de uma planta leguminosa chinesa chamada *Caragana jubata* (NA et al., 2016).

Estes resultados discrepantes são decorrentes ao filtro mais estrito dos dados obtidos do pré-processamento, que removeram uma quantidade maior de sequências que não foram sequenciadas até no mínimo 400 pares de bases, que cobriria quase completamente a região sequenciada do rRNA 16S. Entretanto, devido aos diversos passos relativos à remoção de erros do pré-processamento, os dados da **figura 07** podem ser considerados tão relevantes quanto aos da **figura 06**.

Considerando a validade das duas metodologias, na **figura 08** a seguir foi gerado para a parametrização P1 através da *pipeline* BMP a distribuição da comunidade de micro-organismos em relação ao nível taxonômico de filo. Pode-se observar que, exceto para os filios candidatos (aqueles que micro-organismos cultiváveis ainda não foram encontrados), somente os micro-organismos pertencentes ao filo *Acidobacteria* foram encontrados como superabundantes na rizosfera de *C. adamantium*.

A maior proporção de sequências no total foi associada ao filo Proteobactéria, apesar da maior abundância neste nível taxonômico superior estar associada ao solo não-rizosférico.

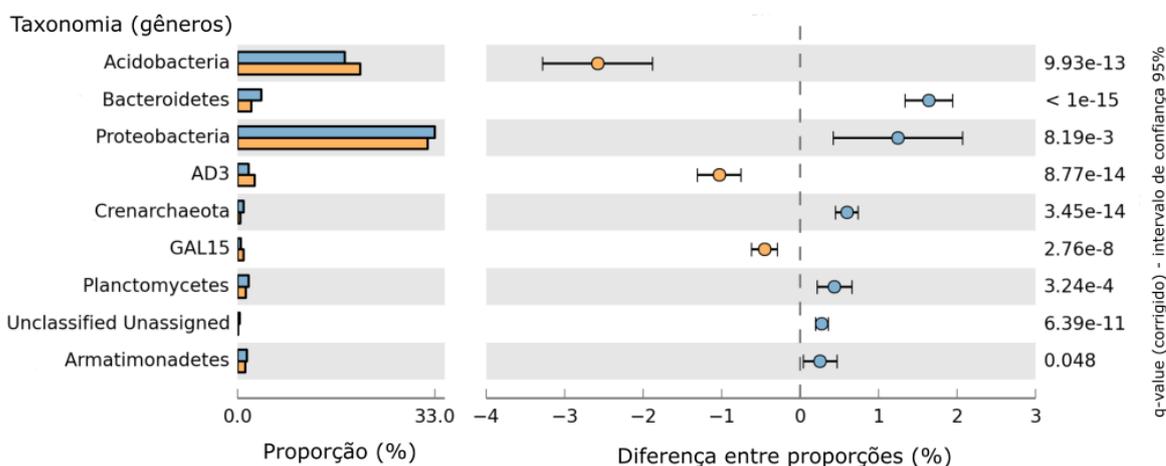


Figura 8. Micro-organismos superabundantes no solo de rizosfera (laranja), em comparação ao solo não-rizosférico (azul) a nível de filo. Gerado através do programa STAMP, *pipeline* BMP, parâmetro P1.

A seguir, na **figura 09**, encontram-se os micro-organismos superabundantes em rizosfera discernidos pela *pipeline* MICCA no parâmetro P1.

Esta análise pode ser considerada a mais rica em dados em relação tanto ao programa utilizado quanto à parametrização, revelou 10 gêneros de bactérias abundantes relativas à rizosfera de *C. adamantium* ($q < 0,05$) e 15 gêneros superabundantes ao solo não-rizosférico ($q < 0,05$), totalizando-se 05 gêneros a mais que a análise através da *pipeline* MICCA.

Um dos três gêneros únicos abundantes na região de rizosfera presentes na **figura 09** foi o *Microlunatus*, um micro-organismo isolado originalmente em efluentes, capazes de armazenar polifosfatos (que são nutrientes importantes ao crescimento das plantas) e biopolímeros como polihidroxialcanoatos e polihidroxitiratos (HEDDERICH; WHITMAN, 2006). Apesar deste micro-organismo também habitar o solo, não foram encontradas associações anteriores à solos de rizosfera.

O penúltimo gênero abundante na rizosfera de *C. adamantium* foi o *Bacillus*. Este gênero é amplamente conhecido como benéfico ao crescimento de plantas, e também por sua capacidade de colonizá-las, havendo-se ainda, inclusive, pedido de patentes para bactérias deste gênero capazes de induzir a produção de compostos orgânicos voláteis em plantas (KLOEPPER, J. W. *et al*, 2016).

O último gênero abundante no solo de rizosfera desta análise foi o *Candidatus Koribacter*, pertencente à família *Acidobacteriaceae*, e foi isolado inicialmente de um solo sob pastagem, na Austrália, mas que ainda não possui descrição taxonomicamente válida, mas evidências encontradas em seu genoma sugerem um papel na redução de nitratos, nitritos e óxido nitroso, e por vezes, sua presença foi dominante em bibliotecas de rRNA 16S baseadas em ambientes ricos em ferro (ALBERS; SIEBERS, 2014)

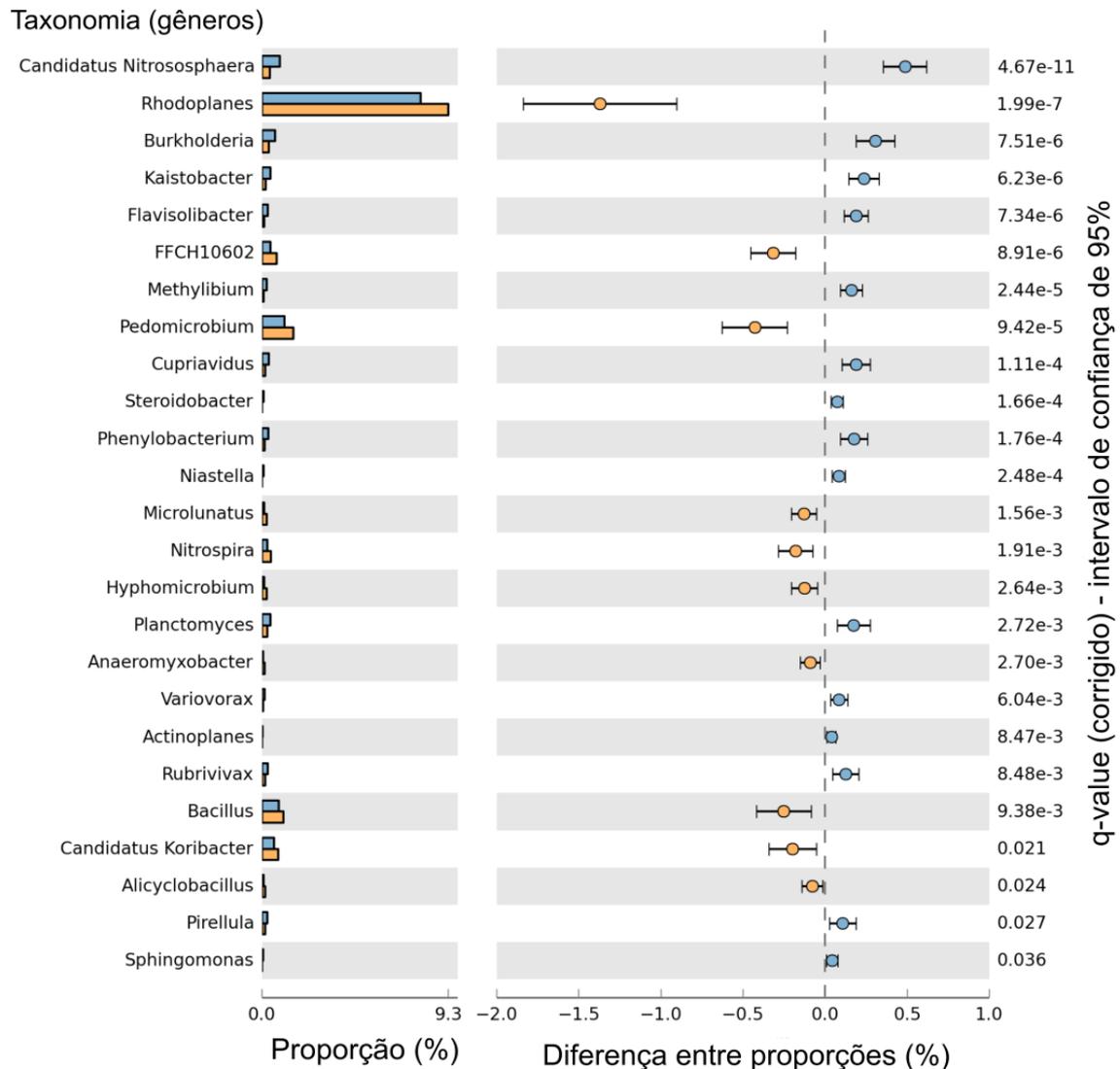


Figura 9. Micro-organismos superabundantes no solo de rizosfera (laranja), em comparação ao solo não-rizosférico (azul). Gerado através do programa STAMP, *pipeline* MICCA, parâmetro P1.

Por fim, na análise a seguir relativa à **figura 10**, utilizou-se o parâmetro P2 para a *pipeline* MICCA, com o objetivo de se encontrar gêneros abundantes de micro-organismos da rizosfera com sequências de tamanho mínimo de 400 pares de bases.

Ao contrário da análise relativa ao programa BMP, todos os micro-organismos abundantes relativos à rizosfera na parametrização P2 foram encontrados na parametrização P1. Isto pode ser associado ao maior poder de análise da *pipeline* MICCA, que utiliza-se de um algoritmo mais robusto para associar os dados do rDNA 16S aos dados de taxonomia. Entretanto, nesta análise, o tamanho do efeito para os dois gêneros abundantes (*Pedomicrobium* e *Rhodoplanes*) foi maior em relação à parametrização P1.

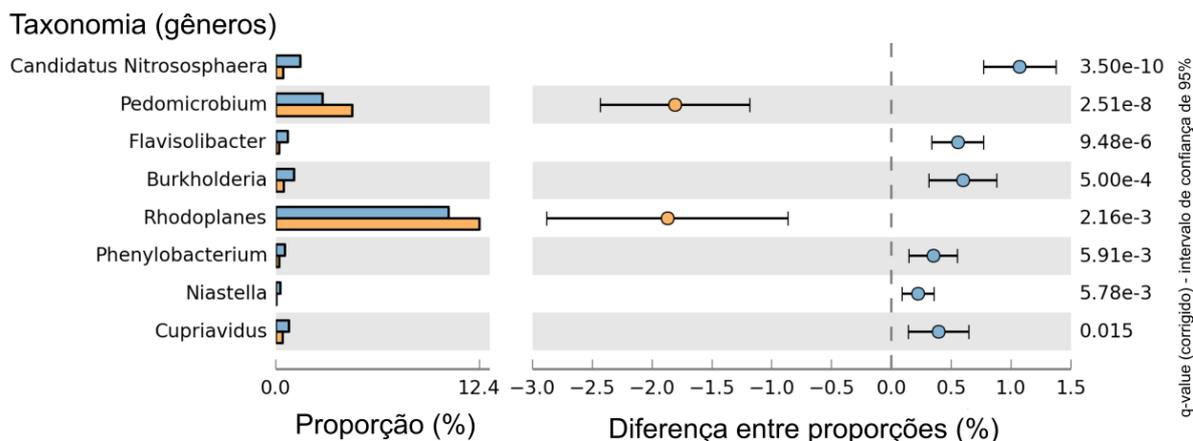


Figura 10. Micro-organismos superabundantes no solo de rizosfera (laranja), em comparação ao solo não-rizosférico (azul). Gerado através do programa STAMP, *pipeline* MICCA, parâmetro P2.

Dos micro-organismos superabundantes, *Rhodoplanes* pode ser considerado um candidato para o isolamento e formulação de um bioinoculante devido ao seu metabolismo de composto fenólico e carotenoides, que são interessantes à promoção do crescimento das plantas.

O mesmo pode ser dito para o próximo gênero, *Pedomicrobium*, devido à sua capacidade de assimilação de ferro e outro micronutriente, o manganês. O próximo gênero que pode ser considerado para a formulação de um bioinoculante é o *Nitrospira*, devido à seu metabolismo documentado associado à oxidação de nitritos, tornando-os acessíveis à assimilação pela raiz da planta.

Também podem ser utilizados os micro-organismos do gênero *Anaeromyxobacter* para a inoculação de *C. adamantium*, devido às suas associações com compostos nitrogenados e fenólicos. O gênero *Hyphomicrobium*, encontrado na análise BMP P2 também poderá ser usado como compositor do produto devido ao metabolismo de amônio.

Ainda no metabolismo de nitrogênio, podem ser utilizados os micro-organismos do gênero *Candidatus Koribacter* após seu isolamento devido ao seu suposto papel na assimilação de nitrogênio. Os micro-organismos do gênero *Microlunatus* podem ser considerados para a formulação do produto devido à sua capacidade de acumulação de fosfatos. Por fim, seria interessante o isolamento de uma estirpe de *Bacillus* única para *C. adamantium*, pois já existem estirpes patenteadas para utilização na raiz de outras plantas.

Por fim, apesar da superabundância de *Alicyclobacillus* em relação ao solo não-rizosférico, sua resistência a processos industriais de sanitização torna-o um candidato não tão

atraente quanto os outros encontrados para a formulação de um composto inoculante. A seguir, na figura 11, encontra-se a distribuição das comunidades superabundantes a nível de filo, pela parametrização P1 da *pipeline* MICCA. Verifica-se que, para este nível taxonômico, as duas *pipelines* tiveram a mesma conclusão (o predomínio de *Acidobacteria*, *AD3* e *GAL15*) no solo de rizosfera, apesar da presença de um menor volume dados relativos ao solo não-rizosférico – 4 filios foram abundantes no solo não-rizosférico para MICCA P1, enquanto 5 filios foram abundantes ao solo não-rizosférico no BMP P1, além dos micro-organismos não atribuídos).

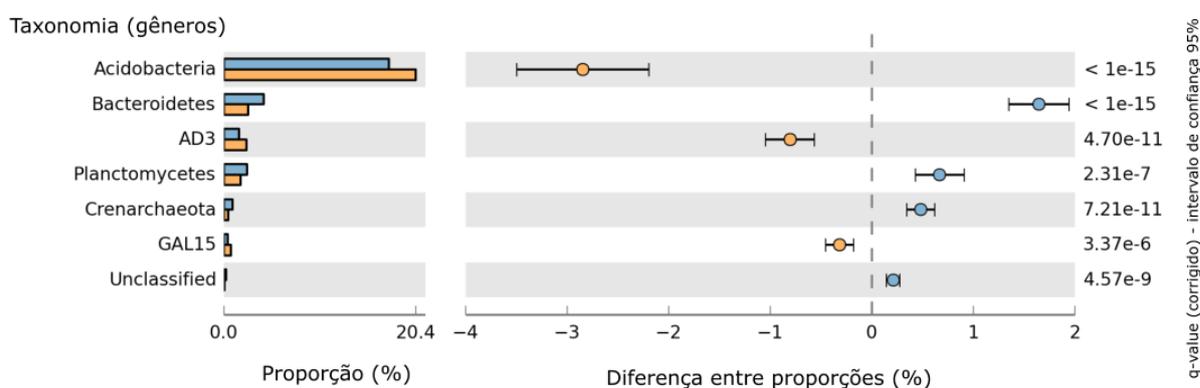


Figura 11. Micro-organismos superabundantes no solo de rizosfera (laranja), em comparação ao solo não-rizosférico (azul) a nível de filo. Gerado através do programa STAMP, *pipeline* MICCA, parâmetro P1.

Assim, com a obtenção dos micro-organismos mais abundantes na rizosfera (até o nível de gênero), é possível delimitar estratégias para isolar e cultivar estes micro-organismos especificamente, de forma a testá-los em relação à sua capacidade de promoção de crescimento, definindo-se quais são os mais impactantes na saúde e desenvolvimento da planta *C. adamantium*.

Em relação às metodologias de análise, a *pipeline* MICCA para o parâmetro P1 conseguiu inferir um maior número de OTUs a nível de gênero, enquanto que o parâmetro P2, a *pipeline* BMP foi capaz de produzir um maior número de verificações quanto aos micro-organismos superabundantes, o que sugere que a utilização de mais de uma metodologia de análise é interessante para desvendar mais relações entre as comunidades de bactérias e arqueias.

5.4. Predição funcional por reconstrução de estados não-observados: PICRUSt

A **figura 12** a seguir representa a predição funcional gerada através do programa PICRUSt, que tenta inferir o papel dos micro-organismos através de dados do RNA ribossomal 16S.

Foram consideradas as 30 primeiras predições com maior tamanho de efeito (ou seja, os 30 mais significativos) para ($q < 1e-15$), isto é, com probabilidade de ocorrência ao acaso tão próxima de zero quanto possível. O padrão para este tipo de análise é ($q < 0,05$), que gerou mais de 2000 predições, mas devido às ressalvas relativas à utilização da predição funcional, foi realizado este delineamento adicional (*q-value* mais estrito) como tentativa de remover os dados associados a um maior grau de incerteza.

Os dados foram gerados utilizando-se como pré-processamento os parâmetros P1, devido à sua validação anterior para a análise taxonômica. Os dados foram concatenados em relação ao tamanho do efeito, sendo que em laranja estão os dados relativos ao solo de rizosfera e, em azul, o solo não-rizosférica.

Algumas das funções preditas foram: antiportador/co-transporte de poliaminas (K03294), difosfato sintase de geranilgeranil (K13789), metabolismo de xenobiótico (K03927) e regulador de resposta (NarL) a nitritos e nitratos (K02479) que podem ter papel na promoção de crescimento da planta.

No caso da primeira, K03294 e da última, K02479, demonstra que a comunidade de micro-organismos exerce esses metabolismos associados ao ciclo do nitrogênio com maior intensidade, enquanto a K13789 está associada ao metabolismo de fosfatos mais intenso no solo de rizosfera (este nutriente estava mais presente no solo de rizosfera, como exposto na tabela 1). Outro composto, K03927, é descrito no banco de dados como uma esterase degradadora de xenobiótico.

As duas principais funções preditas em tamanho de efeito foram: permease de transporte tipo ABC (K02004) e sub-unidade RBP 11 da RNA-pol II (K03088), duas proteínas relativas ao metabolismo a um nível extremamente conservado, relativo às funções básicas dos micro-organismos.

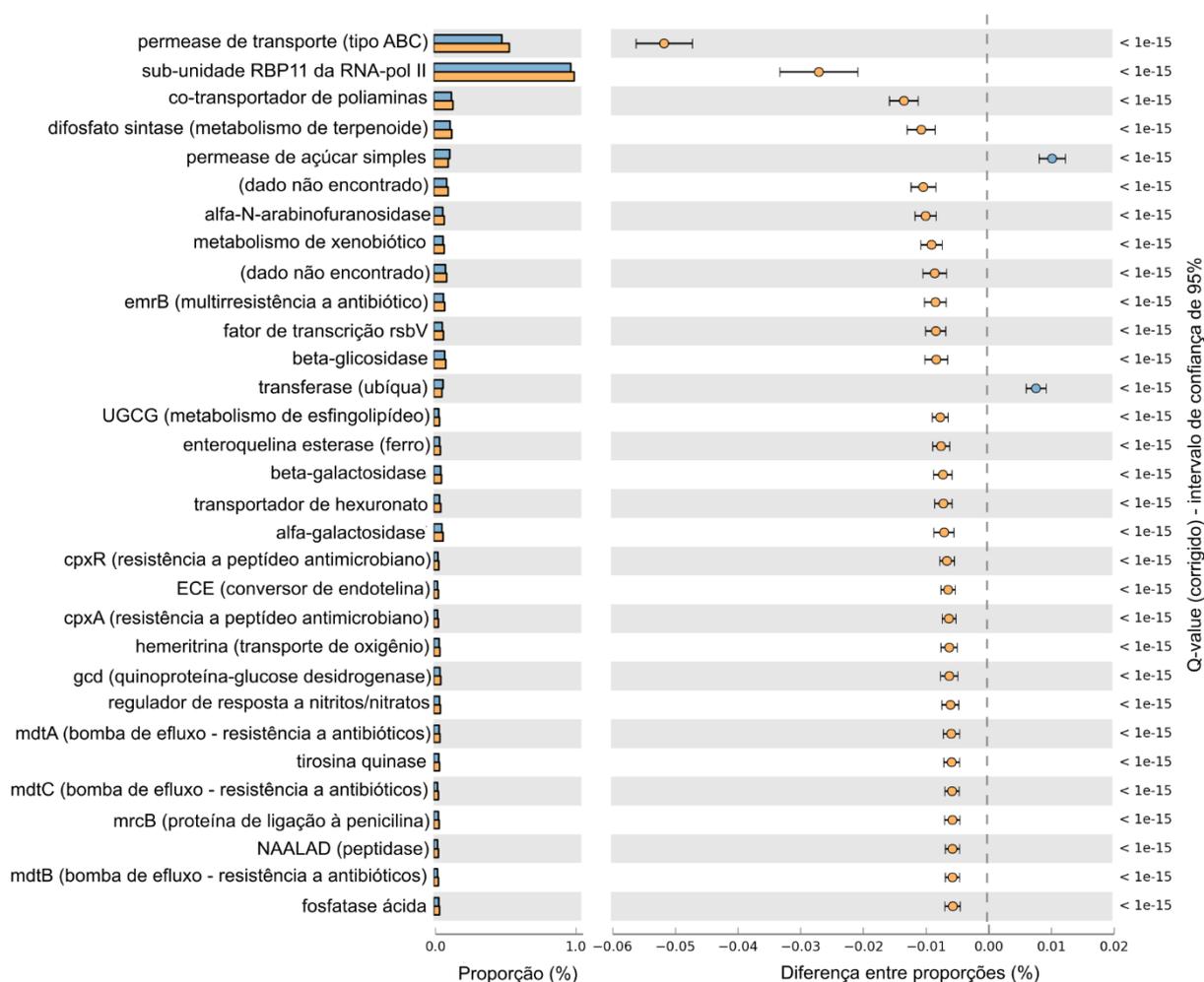


Figura 12. Predição funcional relativa aos micro-organismos do solo de rizosfera (laranja) e de solo não-rizosférico (azul). Gerado através do pacote PICRUST, com parametrização P1.

Chama-se atenção também à presença de múltiplas proteínas associadas à resistência a antibióticos e peptídeos antimicrobianos superabundantes na rizosfera, além da predição de compostos microbianos de interesse biotecnológico, como beta-glicosidase, beta-galactosidase e alfa-galactosidase, sugerindo-se que a rizosfera pode ser uma fonte superior de bioprospecção desses micro-organismos (sem a finalidade de obtenção de estirpes para inoculantes) para aplicações biotecnológicas.

A última observação relativa à predição funcional é a de que no solo não-rizosférico, apenas funções muito ubíquas ou gerais foram superabundantes (permease de açúcar simples – K02057 e transferase ubíqua – K00924), demonstrando-se a pressão seletiva que a rizosfera pode exercer na comunidade microbiana para que exerça papéis específicos relativos a associação com a planta, que fornece condições fisiológicas mais interessantes aos micro-organismos próximos, como já foi descrito anteriormente neste trabalho.

Entretanto, como a obtenção destas proteínas (e assim, funções) não foi realizada experimentalmente (em bancada), são necessários testes adicionais que podem dar ênfase a estas funções preditas, norteados futuros trabalhos para obtenção de novas enzimas microbianas.

6 CONCLUSÃO

Através da metodologia metagenômica, foram encontradas diferenças na comunidade dos micro-organismos do solo associado a rizosfera em comparação ao solo não-rizosférico, rejeitando-se a hipótese nula e aceitando-se a hipótese alternativa.

Desta comunidade, foi possível encontrar 15 gêneros únicos de micro-organismos superabundantes na região de rizosfera, potenciais candidatos para a formulação de um bioinoculante. Dentre estes gêneros, 05 foram associados à promoção do crescimento de plantas – *Nitrospira*, *Hyphomicrobium*, *Bacillus*, *Rhodoplanes*, e *Pedomicrobium*, onde os 03 primeiros foram associados em literatura ao metabolismo de nitrogênio e os 02 últimos foram associados a outras características de interesse.

Em relação às metodologias de análise do conjunto de dados, foram definidos dois parâmetros (P1 e P2) para as duas metodologias em análise, e verificou-se que um maior volume de informações é gerado através do parâmetro menos rígido P1. As *pipelines* BMP e MICCA tiveram performance similar, mas MICCA conseguiu verificar mais conclusões que BMP a nível de gênero, o que é mais interessante para análise da comunidade de bactérias e arqueias para a obtenção dos micro-organismos mais abundantes.

Em relação à predição funcional da comunidade microbiana, foi possível verificar 04 proteínas preditas (K03294, K13789, K03297, K02479) associadas à fixação de nitrogênio e promoção de crescimento em plantas, além de 03 outras funções (alfa-galactosidase, beta-galactosidase e beta-glicosidase) interessantes para a exploração do potencial biotecnológico dos micro-organismos presentes na região de rizosfera da planta.

Com estes dados, espera-se contribuir com futuros esforços de isolamento e caracterização fisiológica e morfológica da comunidade de micro-organismos da rizosfera de *C. adamantium* para o incentivo da produção de um maior número de plantas de guavira e uma exploração mais intensa de seu potencial biotecnológico.

REFERÊNCIAS BIBLIOGRÁFICAS

- AAGAARD, K. et al. The Human Microbiome Project strategy for comprehensive sampling of the human microbiome and why it matters. **The FASEB Journal**, v. 27, n. 3, p. 1012–1022, mar. 2013.
- ALBANESE, D. et al. MICCA: a complete and accurate software for taxonomic profiling of metagenomic data. **Scientific reports**, v. 5, p. 9743, 2015.
- ALBERS, S.-V.; SIEBERS, B. The Prokaryotes: Other Major Lineages. In: **The Prokaryotes**. Berlin, Heidelberg: Springer Berlin Heidelberg, 2014. v. 9783642389p. 323–346.
- ALLAN, E. Metagenomics. **Virulence**, v. 5, n. 3, p. 397–398, 2014.
- ASSHAUE, K. P. et al. Tax4Fun: Predicting functional profiles from metagenomic 16S rRNA data. **Bioinformatics**, v. 31, n. 17, p. 2882–2884, 2015.
- BAIS, H. P. et al. the Role of Root Exudates in Rhizosphere Interactions With Plants and Other Organisms. **Annual Review of Plant Biology**, v. 57, n. 1, p. 233–266, 2006.
- BAKKER, P. A. H. M. et al. The rhizosphere revisited: root microbiomics. **Frontiers in plant science**, v. 4, n. May, p. 165, 2013.
- BERENDSEN, R. L.; PIETERSE, C. M. J.; BAKKER, P. A. H. M. The rhizosphere microbiome and plant health. **Trends in Plant Science**, v. 17, n. 8, p. 478–486, 2012.
- BIBBY, K.; PECCIA, J. Prevalence of respiratory adenovirus species B and C in sewage sludge. **Environmental Science: Processes & Impacts**, v. 15, n. 2, p. 336, 2013.
- BROCHIER-ARMANET, C.; MOREIRA, D. **Environmental Microbiology: Fundamentals and Applications**. Dordrecht: Springer Netherlands, 2015.
- CABEZAS, A. et al. Geobacter, Anaeromyxobacter and Anaerolineae populations are enriched on anodes of root exudate-driven microbial fuel cells in rice field soil. **Environmental Microbiology Reports**, v. 7, n. 3, p. 489–497, 2015.
- CAMPOS, R. P. et al. Conservação pós-colheita de Guavira (*Campomanesia* sp.). **Revista Brasileira de Fruticultura**, v. 34, n. 1, p. 041–049, 2012.
- CÂNDIDO, T. L. N.; SILVA, M. R.; AGOSTINI-COSTA, T. S. Bioactive compounds and antioxidant capacity of buriti (*Mauritia flexuosa* L.f.) from the Cerrado and Amazon biomes. **Food Chemistry**, v. 177, p. 313–319, 2015.
- CAPORASO, J. G. et al. QIIME allows analysis of high-throughput community sequencing data Intensity normalization improves color calling in SOLiD sequencing. **Nature Publishing Group**, v. 7, n. 5, p. 335–336, 2010.

- CASE, R. J. et al. Use of 16S rRNA and rpoB genes as molecular markers for microbial ecology studies. **Applied and Environmental Microbiology**, v. 73, n. 1, p. 278–288, 2007.
- CAVANAUGH, C. M. et al. **The Prokaryotes: Alphaproteobacteria and Betaproteobacteria**. Berlin, Heidelberg: Springer Berlin Heidelberg, 2014. v. 3
- COUTINHO, I. D. et al. Gas Chromatography-Mass Spectrometry (GC-MS) and evaluation of antioxidant and antimicrobial activities of essential oil of *Campomanesia adamantium* (Cambess.) O. Berg (Guavira). **Brazilian Journal of Pharmaceutical Sciences**, v. 45, n. 4, p. 767–776, 2009.
- DANIEL, Rolf. The metagenomics of soil. **Nature Reviews Microbiology**, v. 3, n. 6, p. 470–478, 2005.
- DE SOUZA, J. C. et al. Toxicological analysis and antihyperalgesic, antidepressant, and anti-inflammatory effects of *Campomanesia adamantium* fruit barks. **Nutritional neuroscience**, p. 1–9, 2016.
- DELONG, E. F. et al. **The Prokaryotes**. Berlin, Heidelberg: Springer Berlin Heidelberg, 2014.
- DEN BAKKER, H. C. et al. Rapid whole-genome sequencing for surveillance of salmonella enterica serovar Enteritidis. **Emerging Infectious Diseases**, v. 20, n. 8, p. 1306–1314, 2014.
- DHANANJAYA PRATAP, S.; HARIKESH BAHADUR, S. **Microbial inoculants in sustainable agricultural productivity part II**. [s.l.: s.n.]. v. 1
- DOORNBOS, R. F.; VAN LOON, L. C.; BAKKER, P. A. H. M. Impact of root exudates and plant defense signaling on bacterial communities in the rhizosphere. A review. **Agronomy for Sustainable Development**, v. 32, n. 1, p. 227–243, 2012.
- EDGAR, R. C. Search and clustering orders of magnitude faster than BLAST. **Bioinformatics**, v. 26, n. 19, p. 2460–2461, 2010.
- EDGAR, R. C. UPARSE: highly accurate OTU sequences from microbial amplicon reads. **Nature Methods**, v. 10, n. 10, p. 996–8, 2013.
- EGAMBERDIEVA, D.; SHRIVASTAVA, S.; VARMA, A. **Plant-Growth-Promoting Rhizobacteria (PGPR) and Medicinal Plants**. Cham: Springer International Publishing, 2015. v. 42
- FALEIRO, F. G.; ANDRADE, S. R. M. DE; REIS, F. B. Biotecnologia: estado da arte e aplicações na agropecuária. **Planaltina, DF: Embrapa Cerrados**, 2011.
- FRANKE-WHITTLE, I. H. et al. Rhizosphere bacteria and fungi associated with plant growth in soils of three replanted apple orchards. **Plant and Soil**, v. 395, n. 1–2, p. 317–333, 2015.
- GILBERT, J. A.; JANSSON, J. K.; KNIGHT, R. The Earth Microbiome project: successes and aspirations. **BMC Biology**, v. 12, n. 1, p. 69, 2014.

- GREENE, A. C. et al. Adapting bioinformatics curricula for big data. **Briefings in Bioinformatics**, v. 17, n. 1, p. 43–50, 2016.
- HEDDERICH, R.; WHITMAN, W. B. **The Prokaryotes: Actinobacteria**. New York, NY: Springer New York, 2006.
- HINSINGER, P. Bioavailability of soil inorganic P in the rhizosphere as affected by root-induced chemical changes: A review. **Plant and Soil**, v. 237, n. 2, p. 173–195, 2001.
- JU, F.; ZHANG, T. 16S rRNA gene high-throughput sequencing data mining of microbial diversity and interactions. **Applied Microbiology and Biotechnology**, p. 4119–4129, 2015.
- KANEHISA, M. et al. KEGG: new perspectives on genomes, pathways, diseases and drugs. **Nucleic Acids Research**, v. 45, n. D1, p. D353–D361, 4 jan. 2017.
- KANEHISA, M.; BORK, P. Bioinformatics in the post-sequence era. **Nature genetics**, v. 33 Suppl, n. march, p. 305–310, 2003.
- KAPLAN, D. I. et al. Unique Organic Matter and Microbial Properties in the Rhizosphere of a Wetland Soil. **Environmental Science and Technology**, v. 50, n. 8, p. 4169–4177, 2016.
- KEMLER, M. et al. Ion Torrent PGM as tool for fungal community analysis: A case study of endophytes in eucalyptus grandis reveals high taxonomic diversity. **PLoS ONE**, v. 8, n. 12, 2013.
- KLOPPER, Joseph W. et al. Inoculants Including Bacillus Bacteria for Inducing Production of Volatile Organic Compounds in Plants. **U.S. Patent Application** n. 15/014,397, 3 fev. 2016.
- KUMAR, S. (ED.). **The Role of Bioinformatics in Agriculture**. [s.l.] CRC Press, 2014.
- LANGILLE, M. et al. Predictive functional profiling of microbial communities using 16S rRNA marker gene sequences. **Nature biotechnology**, v. 31, n. 9, p. 814–21, 2013.
- LI, X. et al. Dynamics of the bacterial community structure in the rhizosphere of a maize cultivar. **Soil Biology and Biochemistry**, v. 68, p. 392–401, 2014.
- LORENZI, H. **Frutas brasileiras e exóticas cultivadas: de consumo in natura**. [s.l.] Instituto Plantarum de Estudos da Flora, 2006.
- LOTTMANN, J. et al. Bacterial and fungal communities in the rhizosphere of field-grown genetically modified pine trees (*Pinus radiata* D.). **Environmental Biosafety Research**, v. 9, n. 1, p. 25–40, 2010.
- LUSCOMBE, N. M.; GREENBAUM, D.; GERSTEIN, M. What is bioinformatics? An introduction and overview. **Yearbook of Medical Informatics**, p. 83–100, 2001.
- MASOUDI-NEJAD, A.; NARIMANI, Z.; HOSSEINKHAN, N. **Next Generation Sequencing and Sequence Assembly**. 1. ed. New York, NY: Springer New York, 2013. v. 4
- MCDONALD, D. et al. The Biological Observation Matrix (BIOM) format or: how I learned

- to stop worrying and love the ome-ome. **GigaScience**, v. 1, n. 1, p. 7, 2012.
- METHÉ, B. A. et al. A framework for human microbiome research. **Nature**, v. 486, n. 7402, p. 215–221, 2012.
- MOBIO. **PowerSoil DNA Isolation Kit Protocol**, 2014. Disponível em: <https://www.biotechniques.com/protocols>.
- MORI, H.; MARUYAMA, F.; KUROKAWA, K. VITCOMIC: visualization tool for taxonomic compositions of microbial communities based on 16S rRNA gene sequences. **BMC bioinformatics**, v. 11, p. 332, 2010.
- MUKHERJEE, J. (ED.). **Biotechnological Applications of Biodiversity**. Berlin, Heidelberg: Springer Berlin Heidelberg, 2015. v. 147
- NA, X. et al. Bacterial diversity in the rhizosphere of two phylogenetically closely related plant species across environmental gradients. **Journal of Soils and Sediments**, p. 1–11, 2016.
- NESME, J. et al. Back to the future of soil metagenomics. **Frontiers in Microbiology**, v. 7, n. February, p. 1–5, 2016.
- PARKS, D. H. et al. STAMP: Statistical analysis of taxonomic and functional profiles. **Bioinformatics**, v. 30, n. 21, p. 3123–3124, 2014.
- PEVSNER, J. **Bioinformatics and Functional Genomics, 2nd Ed.** [s.l.: s.n.].
- PINTON, R.; VARANINI, Z.; NANNIPIERI, P. **The Rhizosphere**. Boca Ratón: CRC Press, 2007.
- PLUMMER, E.; TWIN, J. A Comparison of Three Bioinformatics Pipelines for the Analysis of Preterm Gut Microbiota using 16S rRNA Gene Sequencing Data. **Journal of Proteomics & Bioinformatics**, v. 8, n. 12, p. 283–291, 2015.
- PYLRO, V. S. et al. Brazilian Microbiome Project: Revealing the Unexplored Microbial Diversity-Challenges and Prospects. **Microbial Ecology**, v. 67, n. 2, p. 237–241, 2014.
- QUINCE, C. et al. Removing Noise From Pyrosequenced Amplicons. **BMC Bioinformatics**, v. 12, n. 1, p. 38, 2011.
- RAMOS, D. D.; CARDOSO, C. A. L.; YAMAMOTO, N. T. Avaliação do potencial citotóxico e atividade antioxidante em *Campomanesia adamantium* (Cambess.) O. Berg (Myrtaceae). **Revista Brasileira de Biociências**, v. 5, n. 2, p. 774–776, 2008.
- RAMSDEN, J. **Bioinformatics**. London: Springer London, 2015. v. 21
- RAO, I. M. et al. Root adaptations to soils with low fertility and aluminium toxicity. **Annals of Botany**, v. In press, 2016.
- RONQUIM, C. Conceitos de fertilidade do solo e manejo adequado para regiões tropicais. **Boletim de Pesquisa e Desenvolvimento**, v. 8, p. 26, 2010.

- ROSENBERG, E. et al. (EDS.). **The Prokaryotes: Deltaproteobacteria and Epsilonproteobacteria**. Berlin, Heidelberg: Springer Berlin Heidelberg, 2014a.
- ROSENBERG, E. et al. (EDS.). **The Prokaryotes: Firmicutes and Tenericutes**. Berlin, Heidelberg: Springer Berlin Heidelberg, 2014b.
- SCHOLZ, M. B.; LO, C. C.; CHAIN, P. S. G. Next generation sequencing and bioinformatic bottlenecks: The current state of metagenomic data analysis. **Current Opinion in Biotechnology**, v. 23, n. 1, p. 9–15, 2012.
- STREIT, Wolfgang R.; DANIEL, Rolf. Metagenomics. **Humana Pr Inc**, 2010.
- SHARON, I.; BANFIELD, J. F. Microbiology. Genomes from metagenomics. **Science (New York, N.Y.)**, v. 342, n. 6162, p. 1057–8, 2013.
- SOUZA, R. S. C. DE et al. Unlocking the bacterial and fungal communities assemblages of sugarcane microbiome. **Scientific Reports**, n. February, p. 1–15, 2016.
- SRINIVAS, A.; SASIKALA, C.; RAMANA, C. V. Rhodoplanes oryzae sp. nov., a phototrophic alphaproteobacterium isolated from the rhizosphere soil of paddy. **International Journal of Systematic and Evolutionary Microbiology**, v. 64, n. PART 7, p. 2198–2203, 2014.
- SUYAL, D. C. et al. Bacterial diversity and community structure of Western Indian Himalayan red kidney bean (*Phaseolus vulgaris*) rhizosphere as revealed by 16S rRNA gene sequences. **Biologia**, v. 70, n. 3, p. 305–313, 2015.
- TURNER, T. R.; JAMES, E. K.; POOLE, P. S. The plant microbiome. **Genome Biology**, v. 14, n. 6, p. 209, 2013.
- UROZ, S. et al. Pyrosequencing reveals a contrasted bacterial diversity between oak rhizosphere and surrounding soil. **Environmental Microbiology Reports**, v. 2, n. 2, p. 281–288, 2010.
- WALTERS, W. et al. Improved Bacterial 16S rRNA Gene (V4 and V4-5) and Fungal Internal Transcribed Spacer Marker Gene Primers for Microbial Community Surveys. **mSystems**, v. 1, n. 1, p. e00009-15, fev. 2016.
- WANG, Q. et al. Naïve Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. **Applied and Environmental Microbiology**, v. 73, n. 16, p. 5261–5267, 2007.
- WOESE, C. R.; FOX, G. E. Phylogenetic structure of the prokaryotic domain: the primary kingdoms. **Proceedings of the National Academy of Sciences of the United States of America**, v. 74, n. 11, p. 5088–5090, 1977.
- WOO, P. C. Y. et al. Then and now: Use of 16S rDNA gene sequencing for bacterial

identification and discovery of novel bacteria in clinical microbiology laboratories. **Clinical Microbiology and Infection**, v. 14, n. 10, p. 908–934, 2008.

WÜNSCHIERS, R. **Computational Biology**. Berlin, Heidelberg: Springer Berlin Heidelberg, 2013.

XIA, Y. et al. Mining of Novel Thermo-Stable Cellulolytic Genes from a Thermophilic Cellulose-Degrading Consortium by Metagenomics. **PLoS ONE**, v. 8, n. 1, 2013.

YE, S. Q. (ED.). **Big Data Analysis for Bioinformatics and Biomedical Discoveries**. Boca Raton: CRC Press, 2016.

YOKOTA, A. Cultivation of Uncultured Bacteria of the Class Ktedonobacteria in the Phylum Chloroflexi. **Makara Journal of Science**, v. 1, p. 1–8, 2012.

ZHANG, S. et al. Developmental validation of a custom panel including 273 SNPs for forensic application using Ion Torrent PGM. **Forensic Science International: Genetics**, v. 27, p. 50–57, 2017.

Apêndice A – Micro-organismos abundantes incluindo os não-classificados até gênero pela *pipeline* MICCA, parâmetro P2

Na **figura A-1** a seguir encontram-se os micro-organismos abundantes segundo a *pipeline* MICCA nos parâmetros P2, incluindo-se aqueles não classificados até gênero. Enquanto na **figura 10** foram identificados apenas 08 diferentes gêneros com relações de superabundância estatisticamente significativas, na figura abaixo foram encontradas 23 relações de superabundância.

Entretanto, como o nível de anotação e caracterização é menor para níveis taxonômicos superiores é mais generalizado, foi considerado menos interessante para o presente trabalho (que busca aplicações biotecnológicas pontuais e específicas) e, desta forma, estes dados foram omitidos.

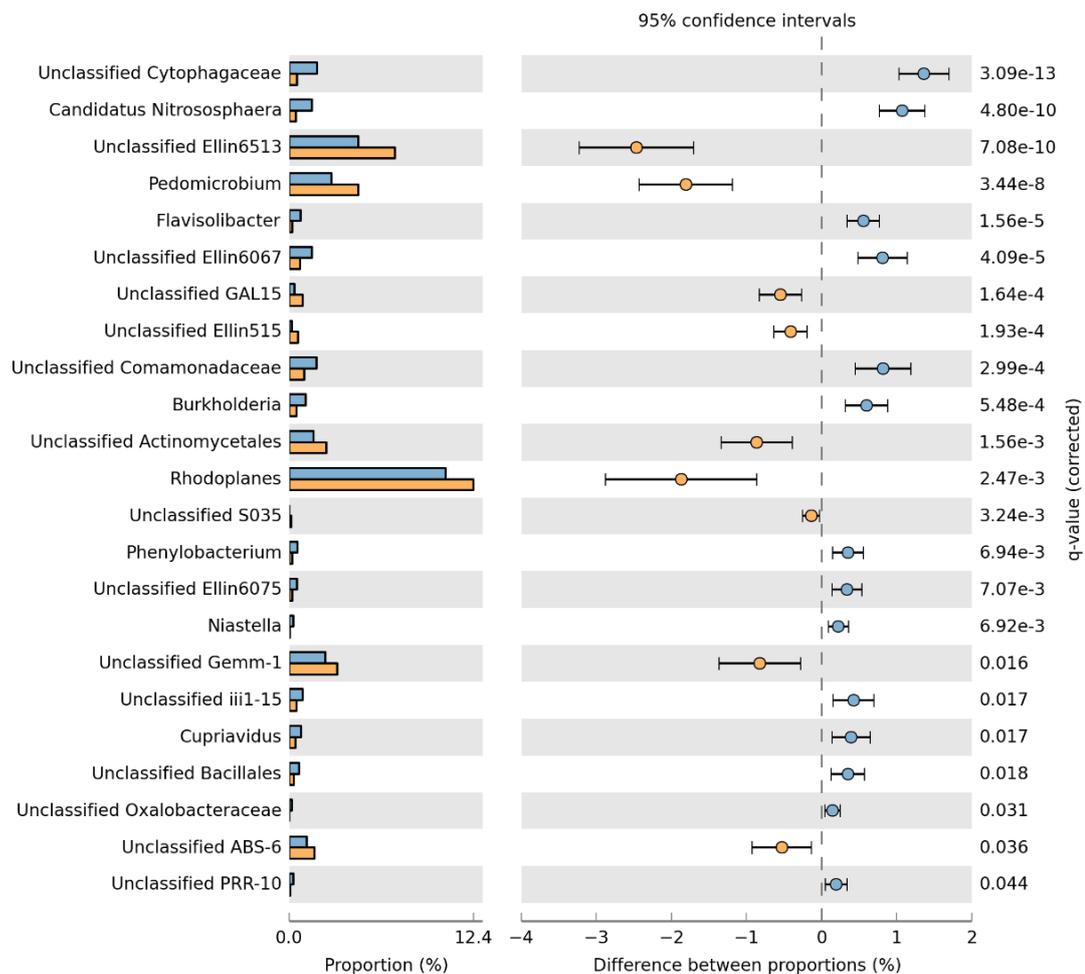


Figura A-1. Micro-organismos superabundantes a nível de gênero na rizosfera (laranja) em relação ao solo não rizosférico (azul), incluindo-se os micro-organismos não classificados até nível de gênero. Gerado pelo programa STAMP, através da *pipeline* MICCA com parâmetros P2.