

UNIVERSIDADE FEDERAL DA GRANDE DOURADOS

DANIEL VASCONCELOS RISSI

**ANÁLISE METAGENÔMICA DE SOLOS SOB FLORESTA SEMIDECIDUAL E
SISTEMA PLANTIO DIRETO**

Dourados

2015

DANIEL VASCONCELOS RISSI

**ANÁLISE METAGENÔMICA DE SOLOS SOB FLORESTA SEMIDECIDUAL E
SISTEMA PLANTIO DIRETO**

Trabalho de Conclusão de Curso de Graduação
apresentado para a obtenção do título de Bacharel em
Biotecnologia.

Faculdade de Ciências Biológicas e Ambientais
Universidade Federal da Grande Dourados

Orientador: Prof. Dr. Rodrigo Matheus Pereira

Dourados

2015

Dados Internacionais de Catalogação na Publicação (CIP).

R596a	<p><u>Rissi</u>, Daniel Vasconcelos. Análise <u>metagenômica</u> de solos sob floresta semidecidual e sistema de plantio direto. / Daniel Vasconcelos <u>Rissi</u>. – Dourados, <u>MS</u>: UFGD, 2015. 51f.</p> <p>Orientador: Prof. Dr. Rodrigo Matheus Pereira. Monografia (Graduação em Biotecnologia) – Universidade Federal da Grande Dourados.</p> <p>1. Bioinformática. 2. Mata nativa. 3. Diversidade microbiana. I. Título.</p> <p style="text-align: right;">CDD – 631.51</p>
-------	---

Ficha catalográfica elaborada pela Biblioteca Central – UFGD.

©Todos os direitos reservados. Permitido a publicação parcial desde que citada a fonte.

Daniel Vasconcelos Rissi

**ANÁLISE METAGENÔMICA DE SOLOS SOB FLORESTA SEMIDECIDUAL E
SISTEMA PLANTIO DIRETO**

Trabalho de Conclusão de Curso aprovado como requisito parcial para obtenção do título de Bacharel em Biotecnologia na Universidade Federal da Grande Dourados, pela comissão formada por:

Orientador: Prof. Dr. Rodrigo Matheus Pereira
FCBA – UFGD

Orientador: Prof. Maricy R. Lindenbah Bonfá
FCBA – UFGD

Orientador: Dr. Fabio Martins Mercante
EMBRAPA AGROPECUÁRIA OESTE

Dourados, 13 de Novembro de 2015.

RESUMO

A metagenômica pode ser usada para determinar a diversidade de uma comunidade microbiana, a presença de microrganismos específicos ou dominantes, rotas metabólicas, bem como determinar a simples presença de um gene. No presente trabalho, foram analisadas amostras de solo de dois locais do município de Dourados, Mato Grosso do Sul. Foram analisados solos em cultivo agrícola sob sistema de Plantio Direto e outro sob Mata Nativa de Floresta Semidecidual. Foram inseridas 54.230 sequências a partir do solo de Mata Nativa, no MG-RAST. Dentre estas, 26.944 sequências (49.7%) contêm genes que predizem proteínas de função conhecida e 24.638 sequências (45.4%) contêm genes que codificam para proteínas com função desconhecida. Foram inseridas 69.164 sequências obtidas a partir do solo cultivado sob Sistema Plantio Direto, no MG-RAST. Dentre estas, 43.673 sequências (63.1%) contêm genes codificadores de proteínas com funcionalidade conhecida e 24.776 sequências (35.8%) contêm genes codificadores de proteínas com função desconhecida. Os resultados obtidos através da ferramenta MG-RAST e STAMP apontaram uma presença maior de bactérias em comparação com Archeas, eucariotos e vírus, em ambos os solos. No solo sob Mata Nativa, a porcentagem de bactérias presentes foi de 84,60% - (45.39% Proteobactérias; 15.32% Actinobactérias; 5.20% Firmicutes; 4.78% Acidobactérias e 2.74% de Cyanobactérias), sequências não atribuídas 10.39%; 3.93% de Archeas; 0.77% de Eucariotos e sequências sem classificação 0.23%. Mendes et al. (2015) também encontraram uma abundância maior de proteobactérias (61%) na floresta amazônica. No solo sob Sistema Plantio Direto, 92.74% foram de bactérias (48.72% Actinobactérias; 33.77% Proteobactérias; 3.05% Firmicutes; Chloroflexi 1.78% e 0.97% de Cyanobactérias), sequências não atribuídas de 5.01%; Archeas 1.62%; Eucariotos 0.49% e sequências sem classificação, 0.061%. Apesar da diferença no tratamento do solo, observou-se que os filos de maior abundância estavam presentes em ambos os solos apresentados, tanto em Mata Nativa quanto em Sistema Plantio Direto.

Palavras-chave: Bioinformática; Mata Nativa; Diversidade Microbiana.

AGRADECIMENTOS

Aos meus professores, que ofereceram todo seu conhecimento e se colocaram a disposição dos alunos durante toda a graduação, no meu processo de formação profissional.

Aos meus pais, pelo amor, incentivo e apoio durante toda minha vida. Principalmente, por sua capacidade de acreditar e investir em mim.

Aos meus amigos, pelos seus conselhos e apoio.

A FUNDECT, CAPES, Unesp (FCAV) campos Jaboticabal, Universidade Federal da Grande Dourados (UFGD), EMBRAPA Agropecuária Oeste (CPAO)

SUMÁRIO

1. INTRODUÇÃO.....	9
Objetivos	10
2. REVISÃO BIBLIOGRÁFICA	11
2.1. Diversidade microbiológica	11
2.2. Metagenômica	12
2.3. Etapas da análise metagenômica	14
2.3.1. Coleta do Solo e Extração de DNA.....	14
2.3.2. Sequenciamento Illumina	15
2.3.3. Classificação dos microrganismos (<i>Binning</i>)	15
2.3.3.4. TETRA (2004)	18
2.3.3.5. SOrt-ITEMS (2009).....	18
2.3.3.6. DiScRIBinATE (2010)	18
2.3.3.7. MetaPhyler (2010).....	19
2.3.3.8. MARTA (2010).....	19
2.3.3.9. MEGAN (2007).....	19
2.3.3.10. TACOA.....	19
2.3.3.11. MG-RAST.....	20
2.3.4. STAMP	20
3. METODOLOGIA.....	22
3.1. Coleta das Amostras	22
3.2. Química do Solo	22
3.3. Extração do DNA	23
3.4. Sequenciamento.....	24
3.5. Montagem das Sequencias (Assembly)	24
3.6. Identificação de ORF's	25
3.7. Análise MG-RAST	25
3.8. Análise Estatística.....	26
4. RESULTADOS E DISCUSSÕES.....	27
4.1. Análise Química do Solo	27
4.2. Sequenciamento.....	27
4.3. Diversidade Microbiana.....	28
4.3.1. Curva de Rarefação	29

4.4. Bactérias.....	30
4.4.1. Actinobactérias	31
4.4.2. Proteobactérias	33
4.4.3. Firmicutes	34
4.5. Archeias	34
4.5.1. Thaumarchaeota	35
4.5.2. Euryarchaeota	36
4.5.3. Crenarchaeota	37
4.6. Eucariotos	38
4.6.1. Ascomycota	39
4.6.2. Streptophyta	40
4.6.3. Chordata.....	40
4.7. Vírus	41
5. CONCLUSÕES.....	42
6. REFERÊNCIAS BIBLIOGRÁFICAS	43

LISTA DE FIGURAS

Figura 1. Gráfico Estatístico da Diversidade Microbiana	28
Figura 2. Gráfico de Distribuição da Diversidade Microbiana	28
Figura 3. Curva de Rarefação dos Sistema Analisados	29
Figura 4. Grafico Estatístico do Domínio Bactéria	31
Figura 5. Gráfico de Distribuição do Domínio Bactéria	32
Figura 6. Gráfico Estatístico do Domínio Arqueia.....	35
Figura 7. Grafico de Distribuição do Domínio Arqueia	36
Figura 8. Grafico Estatístico do Domínio Eucarioto	39
Figura 9. Grafico de Distribuição do Domínio Eucarioto.....	39

LISTA DE TABELAS

Tabela 1. Uma visão geral dos vários métodos para classificação taxonômica	17
Tabela 2. Rotação de culturas apresentada no solo de Plantio Direto	22
Tabela 3. Resultado da Análise Química dos Solos.....	23

1. INTRODUÇÃO

Os solos são um dos ambientes onde os microrganismos são mais ativos porém desconhecidos. A complexidade e a heterogeneidade de suas comunidades fazem ser um estudo dos grandes desafios da microbiologia moderna (JANSSON, 2015).

A metagenômica é uma análise genômica independente de cultivo que serve como ferramenta para entender e explorar toda diversidade microbiana de um ambiente. O termo metagenômica é derivado do conceito estatístico de meta-análise (o processo de combinar estatisticamente análises distintas) e genômica (a análise compreensiva do material genético de um organismo) (SCHLOSS et al. 2003). Essa tecnologia possibilita acessar informações em nível de filogenia e o isolamento e clonagem de novos genes da comunidade microbiana responsáveis pelas mais diversas funções (STREIT et al. 2004)

A metagenômica fornece uma maneira muito eficaz para analisar comunidades microbianas complexas. Sequências de genes funcionais amplificadas com base no *Next Generation Sequencing* (NGS) fornece uma perspectiva aprofundada em sua diversidade, distribuição e papéis ecológicos, bem como para as relações entre conjuntos microbianos e particulares funções ecológicas. A tarefa de análise e compreensão das interações planta-microrganismo e os papéis de diversos conjuntos de microrganismos presentes é de vasta dimensão, mas as recompensas ao nível das tecnologias agrícolas mais eficientes e gestão ecológica das terras são proporcionais ao esforço. (HOWARD et al. 2011; LEMA et al. 2014; WANG et al. 2013; WOODHOUSE et al. 2013; ZHENG et al. 2013).

De modo geral, a microbiologia do solo pode ser afetada pelos diferentes tipos de manejo do solo (MERCANTE et al. 2008). Desta forma, cabe entender melhor os impactos causados na microbiota do solo pelo desmatamento da floresta semidecidual nativa para plantio.

A floresta estacional semidecidual, é marcada por intensas chuvas de verão e acentuada seca hiberna, onde 20 a 50% das árvores caducifólias perdem suas folhas (IBGE, 2012), em regiões tropicais com precipitação anual abaixo de 2.000 mm e período de cinco a seis meses de estação seca, com precipitação total menor que 1.000 mm (PENNINGTON et al. 2000). A Floresta Estacional Semidecidual está severamente degradada em toda a área de ocorrência natural, principalmente em função de práticas agropecuárias (DURIGAN et al. 2000).

Como forma de otimização da exploração agropecuária na região, estão sendo implementados sistemas de produção conduzidos em sistema plantio direto com rotação de culturas.

Objetivos

Objetivo Geral: Avaliar a diversidade de comunidades microbianas de solos sob diferentes tipos de manejo através da análise metagenômica total através de sequenciamento Illumina.

Objetivo Específico: Avaliar a diversidade de bactérias, arqueias, eucariotos e vírus em solos de mata nativa e de sistema plantio direto.

2. REVISÃO BIBLIOGRÁFICA

2.1. Diversidade microbiológica

Os microrganismos que habitam o solo possuem funções de grande importância. Pode-se enumerar dentre as mais diferentes funções atribuídas a estes, aquelas mais amplamente conhecidas, como a degradação de compostos orgânicos, e conseguinte ciclagem de nutrientes (GRAHAM et al. 2014), e aquelas mais específicas, como a fixação biológica de nitrogênio e o auxílio às plantas na absorção de nutrientes (MIRANSARI, 2013; CHAGNON et al. 2013).

Aproximadamente 99% dos microrganismos não são passíveis de cultivo por meio de técnicas convencionais de laboratório como crescimento em meio de cultura. Sendo assim, sequenciamento metagenômico é uma abordagem poderosa no estudo da diversidade genética ambiental diretamente, ignorando a limitação do método baseado em cultivo (SHARON et al. 2013).

Devido a sua heterogeneidade física, química e biológica, o solo é um dos maiores reservatórios de biodiversidade microbiana, constituindo um importante recurso para a exploração biotecnológica. A diversidade bacteriana neste ambiente é muito superior à observada para organismos eucariontes, podendo existir milhares de diferentes espécies de bactérias em apenas 1 cm³ de solo (MARON et al. 2011)

A biodiversidade do solo representa um enorme mundo subterrâneo contendo uma vasta gama de organismos, desde arqueias, bactérias e fungos a nematódeos, insetos e vermes. Estes organismos interagem uns com os outros e afetam o funcionamento dos ecossistemas do solo (WAGG et al. 2014). Segundo ARAÚJO et al. (2014), a estrutura da comunidade microbiana do solo é susceptível a variações estacionais e tem respostas diferentes para condições de clima seco e úmido.

Os microrganismos que habitam o solo constituem a base de processos ecológicos, como os ciclos biogeoquímicos e a cadeia trófica, mantendo relações vitais entre si e com os organismos superiores. Diversas funções de importância para a dinâmica funcional do solo são realizadas pelos microrganismos, como a decomposição da matéria orgânica que disponibiliza nutrientes às plantas e a degradação de substâncias xenobióticas. Além disso, a

microbiota do solo atua no controle biológico de patógenos, influencia a solubilização de minerais e contribui para a estruturação e agregação do solo (KUJUR et al. 2012).

2.2. Metagenômica

Em termos de biodiversidade terrestre, o solo é considerado o sistema mais complexo e que sustenta a vida em uma vegetação. A economia do Brasil está fortemente baseada na agricultura e, em 2013/2014, 55,2 milhões de hectares foram cultivados para a produção de grãos em diferentes sistemas de solo e de manejo (CONAB, 2014). A metagenômica permite o estudo da biodiversidade em amostras ambientais de solo, fornecendo grandes ferramentas em potencial para uma melhor compreensão e gestão da paisagem, uma vez que fornecem informações relacionadas não só sobre os organismos que vivem lá, mas também quais funções desempenham e qual é o seu potencial de uso em um contexto sustentável (GOMÉS-MERINO, 2014).

NANNIPIERI (2014) argumenta que o solo é um sistema biológico único, onde há uma grande diversidade de organismos que desempenham várias funções essenciais ao ecossistema. Recentemente, a identificação de novos genes de microrganismos não cultivados em meios dependentes de cultivo permitiu uma compreensão mais profunda de novas vias metabólicas que se desenvolvem nestes ecossistemas. Outros avanços que estão permitindo o rápido desenvolvimento da ciência do solo são as ciências da genômica como metagenômica, metatranscriptômica, o metaproteômica e metabolômica entre outros. Novos desafios para a biologia de sistemas é integrar informações que são obtidas por meio de tecnologias ômicas, a fim de explicar mais plenamente a importância e aplicabilidade da biologia do solo em benefício da sustentabilidade da vegetação.

A metagenômica fornece uma valiosa ferramenta para a descoberta de novos genes, vias metabólicas e produtos importantes na biotecnologia, farmacêutica e de elevada relevância médica (CULLIGAN et al. 2014). O pirosequenciamento surgiu como uma técnica valiosa para investigar mais a fundo a ecologia microbiana e tornar acessível uma enorme diversidade microbiana complexa não cultivável presente nos ambiente. Nos últimos tempos, estratégias metagenômicas têm sido responsáveis pela exploração da imensa diversidade microbiana em todo o mundo (PRAMANIK et al. 2015).

As comunidades bacterianas têm sido tradicionalmente identificadas através de técnicas de cultivo e de microscopia que se mostram úteis a estudos taxonômicos e fisiológicos baseando-se em suas reações metabólicas, características da colônia e morfologia celular que, apesar de se mostrarem válidos por diversos anos, são questionados nos dias atuais em relação à quantidade de microrganismos que se adaptam aos meios de cultivo. Desta forma, a necessidade de técnicas cultivo-independentes se mostrou necessária, tendo como a principal alternativa o desenvolvimento da biologia molecular que propicia a extração de material genético de amostras ambientais, sua amplificação e análise (HANDELSMAN, 2004)

O estudo da metagenômica é de extrema importância, pois durante muito tempo cientistas se dedicaram ao estudo de microrganismos cuja cultura podia ser feita em laboratório, devido ao poder e precisão dos estudos de bactérias nessas condições. Entretanto, percebeu-se que a cultura não resultava na grande diversidade microbiana existente, que se encontra principalmente entre os microrganismos em que não é possível fazer a cultura. Portanto, neste contexto, a metagenômica tornou-se uma peça central para se obter conhecimento sobre a fisiologia e genética desses micro-organismos (HANDELSMAN, 2004).

O crescente uso da metagenômica vem propiciando, cada vez mais, um maior acesso a diversidade microbiana proveniente de amostras ambientais, mostrando cada vez mais a importância desta técnica para o descobrimento de novos microrganismos. Segundo Hugenholtz (1998), apenas 10% dos microrganismos provenientes de amostras ambientais eram cultivados. No entanto, trabalhos mais recentes, mostram que cerca 1% apenas dos microrganismos são cultiváveis em laboratório (MARTINY et al., 2006; CHAUHAN et al. 2011). Entretanto, um dos principais desafios voltado para futuras investigações do solo é a falta de representantes cultiváveis, e conseqüentemente, a baixa disponibilidade de genomas de referência (HOWE et al. 2014).

Análises metagenômica são mais frequentemente realizada por sequenciamento da subunidade 16S do RNA ribossomal (rRNA) ou por sequenciamento total metagenômica (*shotgun*), normalmente em uma plataforma de pirosequenciamento. Estas tecnologias exploram comunidades que independem de métodos microbiológicos tradicionais de cultivo e têm melhorado nossa compreensão das ricas comunidades microbianas que estão presentes em determinados habitats (PREIDIS, 2015) e também têm sido aplicadas na busca de genes e produtos gênicos com interesses biotecnológicos, que possam ser utilizados e incorporados em processos industriais, visando melhorias e redução de custos nesses processos. No que diz

respeito à biotecnologia, a metagenômica veio como uma ferramenta alternativa no âmbito de proporcionar recursos para a descoberta de novas enzimas, compostos químicos e isolamento de genes para o uso em biotecnologia (SCHMEISSER et al. 2007).

Em estudos com fins biotecnológicos, como identificação de novas substâncias com aplicações médicas e industriais, além da necessidade de um DNA livre de contaminantes, o método de extração tem que proporcionar uma quantidade e qualidade de DNA suficientes para as reações. E nestas situações, logicamente, quanto mais DNA, melhor, visto que a representatividade da amostra aumenta, elevando a possibilidade de se encontrar um novo gene e conseqüentemente, uma nova substância (LAMMLE et al., 2007).

A rápida redução no custo nos *Next Generation Sequencing* foi dramaticamente acelerada pelo desenvolvimento de sequenciamentos metagenômicos. Na verdade, o número de sequenciamentos de metagenomas utilizando dados de *shotgun* vem aumentando nos últimos anos e virá a ser uma grande ferramenta para muito laboratórios e cientistas que trabalham no campo da microbiologia ambiental (THOMAS, 2012).

2.3. Etapas da análise metagenômica

2.3.1. Coleta do Solo e Extração de DNA

Metagenômica de solo é pesquisado atualmente por estudos descritivos que analisam metagenomas em amostras de solo tiradas em um único ponto de tempo e localização, possivelmente combinados com a análise das características do solo e / ou quando comparados com diferentes locais ou tratamentos (por exemplo, além de fertilizantes, pH e tipo de solo) (PROSSER, 2015).

O material genético pode ser obtido a partir de amostras por vários métodos, que incluem a lise das células, seguido por extração e purificação de ácidos nucleicos (DNA ou RNA). No entanto, a quantidade de matéria não biológica associada com o material biológico pode interferir com os processos de quantificação, extração e amplificação. Assim, diferentes amostras vão requerer diferentes métodos de extração. Uma boa representação da diversidade biológica depende da eficiência da extração dos ácidos nucleicos. O material genético extraído deve ser livre de inibidores de amplificação e atenção especial deve ser dada quando se lida com amostras obtidas a partir de locais poluídos (CARDOSO et al. 2010).

2.3.2. Sequenciamento Illumina

Estratégias baseadas em illumina são capazes de gerar a maior quantidade de dados de sequência por dólar investido. Dependendo da plataforma e dos reagentes, pode-se obter de 300 a 500 ciclos de dados de sequências nas plataformas HiSeq e MiSeq, respectivamente. Estes ciclos são comumente divididos em duas leituras, fornecendo pareamento dos *reads* do mesmo fragmento de DNA. As plataformas também variam em seu rendimento; o MiSeq foi capaz de gerar 8,5 Gbp utilizando pareamento de reads de 250-nucleotídeos (ou seja, 17 milhões de pares de leituras) (KOZICH et al. 2013). A plataforma HiSeq tornou-se o método padrão para o sequenciamento *shotgun* de metagenomas por causa de sua profundidade aumentada de leitura; no entanto, o MiSeq tem um maior potencial para utilização com 16S rRNA em estudos de sequenciamento de genes porque gera sequências de *reads* mais longas, por exemplo, o HiSeq2500 e MiSeq produzem *reads* 2x150bp e 2x250bp respectivamente, que após a fusão pode gerar *reads* de até 290 e 490 bp (LOGARES et al. 2014), e o seu desempenho e custo correspondem com as necessidades dos pesquisadores (CAPORASO et al. 2012).

O desenvolvimento de ferramentas de bioinformática para a acurácia de sequências geradas nas plataformas Illumina tem sido limitado. Alguns estudos sugerem que a extensa curadoria de sequência e formação de *contigs* chegam a ser desnecessárias (CAPORASO et al., 2014; WERNER, et al. 2012). A diversidade microbiana utilizando o Miseq foram empregados em estudos como o de CAPORASO et al. 2012; SALIPANTE et al. 2014; ZHANG et al. 2015)

2.3.3. Classificação dos microrganismos (*Binning*)

Um dos principais desafios em pesquisa metagenômica é a identificação da origem taxonômica para cada *read* ou gene da metagenômica por *shotgun* e, ultimamente, a reconstrução dos genomas dos membros de cada táxon diretamente desses *reads*. Um conjunto diversificado de métodos desenvolvido para analisar dados de *shotgun* e obter perspectivas sobre os táxons subjacentes. Estes métodos podem ser amplamente divididos em

várias categorias distintas, incluindo: alinhamento para referenciar genomas, classificação taxonômica, *assembly* e *binning* (CARR, 2013)

Um passo fundamental na recuperação do genoma a partir de dados de sequência metagenômicos é a classificação de sequências montadas a partir dos *reads* em unidades distintas, referidos como “bins”. Estas “bins” representam genomas compostos de populações individuais que compõem a comunidade microbiana. Um grande número de abordagens foi desenvolvida para sequências “bins”, montadas a partir de sequências de dados metagenômicos (MACKELPRANG et al. 2011; SHARON et al. 2013). Dentre essas técnicas, uma das mais utilizadas é a auto-organização emergente de mapas (ESOMs), que tem sido usada para montagem de sequências por uma frequência de tetranucleotídeo e ler níveis de *reads* (BAKER, 2004)

Primeiramente, “bins” composicionais fazem uso do fato de que os genomas tem conservado a composição de nucleotídeos (por exemplo, um certo GC ou a distribuição de abundância particular de k-mers) e isso também será refletido em fragmentos da sequência dos genomas. Em segundo, o fragmento de DNA desconhecido pode codificar para um gene e a similaridade deste gene com genes conhecidos de uma base de dados de referência, podem ser usados para classificar e portanto, obtermos a classificação da sequência (THOMAS, 2012).

Existem dois tipos fundamentais de ferramentas de análise de dados metagenômicos, denominados *assembly* e *binning*. Ferramentas de *assembly* tentam reconstruir os genomas existentes na amostra. Ferramentas de *binning* tentam agrupar o *reads* de NGS de espécies semelhantes. (WANG, 2014)

Em um caso ideal, se pudéssemos montar cada espécie em uma amostra metagenômica, poderíamos resolver o problema da anotação de forma relativamente fácil. No entanto, as ferramentas de montagem existentes estão longe de ser o caso ideal e agrupar os dados metagenômicos ainda é um problema desafiador e não resolvido, embora ferramentas de *assembly* metagenômicos podem construir contigs mais longos (WANG, 2014)

Entretanto, os maiores avanços têm sido feitos para as análises de *binning*. As estratégias de *binning* existentes podem ser divididas em duas categorias: métodos supervisionados (também chamados métodos baseados em similaridade) e métodos não supervisionados (também chamados de métodos baseados em composição) (WANG, 2014)

Os métodos de *binning* supervisionados (BRADY, 2009; MCHARDY, 2006) são as abordagens mais comuns para análise de amostras metagenômicas. Eles fazem uso de

genomas conhecidos e buscam similaridade através de *reads* ou *contigs* (após o *assembly*). Alguns métodos supervisionados usam recursos genéricos, tais como pequena subunidade 16S rRNA, *recA* e *rpoB*, para classificar os fragmentos. No entanto, uma grande porcentagem (> 99%) de *reads* (ou *contigs*) não tem estes recursos (GARCIA, 2008). Além disso, uma espécie pode ter vários marcadores e várias espécies podem compartilhar o mesmo marcador (CASE, 2007).

Métodos não supervisionados - que não dependem de informações sobre o genoma conhecido - e geralmente são agrupados em *reads* e *contigs* juntos com base em três observações: (A) a frequência k-mer de leituras, onde, $k \cong 16$ geralmente linearmente proporcional à abundância da espécie correspondente (FOFANO, 2004); (B) w-mer são suficientemente longas, onde $w \geq 36$, tem uma probabilidade muito alta para ser única para cada espécie (CHATTERJI, 2008); e (C) reads e contigs suficientemente longos da mesma ou de espécies semelhantes, tendem a ter distribuições semelhantes de q-mer curtas, em que $q = 4$ ou 5, (LEUNG, 2011).

A Tabela 1 mostra alguns programas utilizados para classificação taxonômica com seus respectivos métodos para a classificação de microrganismos.

Tabela 1: Uma visão geral dos vários métodos para classificação taxonômica. Os checkmarks indicam se os métodos são baseados em composição ou baseado em comparação- e se eles usam um classificador Baesiano ou de máxima verossimilhança (ML) (GERLACH, 2012 ADAPTADO).

Método	Composição	Comparação	Baesiano	ML
TETRA	✓			
Phylopythia	✓			
TACOA	✓			
RAlphy	✓			
NBC	✓		✓	
PhymmBL	✓		✓	
GSTaxClassifier	✓		✓	
BLAST		✓		
MG-RAST		✓		
MEGAN		✓		
CARMA 1		✓		
AMPHORA		✓		
SOrt-ITEMS		✓		
Sphinx	✓	✓		
DiScRIBinATE		✓		
Metaphyler		✓		
MARTA		✓		
EPA		✓		✓
Pplacer		✓		✓
MLtreeMap		✓		✓

2.3.3.4. TETRA (2004)

TETRA (Teeling, 2003. TEELING, 2004) usa a distribuição de tetranucleotídeos dos diferentes fragmentos metagenômicos para calcular os coeficientes de correlação de par-a-par, que podem ser utilizados como uma estimativa para a probabilidade de que dois fragmentos metagenômicos originaram no mesmo genoma. Uma comparação das sequências de um metagenoma com a distribuição de padrões de sequências de tetranucleotídeo com conhecidos padrões taxonômicos fornece uma caracterização taxonômica da metagenoma. Os autores relatam que este método requer sequências com cerca de 40 kb para uma melhor análise, enquanto sequências abaixo de 1 kb não são adequadas para a análise.

2.3.3.5. SOrt-ITEMS (2009)

Sort-ITEMS (HANQUE, 2009) é um método que estende o método LCA e utiliza técnicas adicionais para reduzir o número de predições de falsos positivos. Os resultados mostram que a ferramenta consistentemente faz um número significativamente menor de falsas previsões do que o MEGAN em todos os cenários de avaliação. Os números das previsões falsas em ranques taxonômicos inferiores não são fornecidos.

2.3.3.6. DiScRIBinATE (2010)

DiScRIBinATE é uma variante modificada do Sort-ITEMS. O método envolve uma reclassificação de *reads* que tenham sido atribuídas à taxa com poucos *reads* atribuídos. A avaliação mostra uma redução do tempo de processamento e uma sensibilidade um pouco maior em relação ao Sort-ITEMS.

2.3.3.7. MetaPhyler (2010)

MetaPhyler (GHOSH, 2010) restringe a classificação taxonômica para 31 genes com marcadores similares aos da AMPHORA para criar um perfil filogenético. Isto reduz as atribuições taxonômicas para a melhor combinação do BLAST para um maior nível taxonômico, dependendo dos parâmetros como *bit score* e de tamanho dos segmentos pareados de alta pontuação.

2.3.3.8. MARTA (2010)

A classificação taxonômica do MARTA baseia-se na taxonomia do melhor valor de BLAST e usa parâmetros de alinhamento (por exemplo, a porcentagem de identidade) como limites para fazer atribuições maiores nos raques taxonômicos, se necessário. MARTA pode ser utilizado para classificar as sequências metagenômicas shotgun, bem como sequências de 16S-rDNA.

2.3.3.9. MEGAN (2007)

O MEGAN (HUSON, 2007) é baseado no menor ancestral comum (LCA). A utilização do BLAST melhora a aproximação da classificação, pois utiliza somente o melhor hit encontrado.

2.3.3.10. TACOA

O programa aplica a ideia intuitiva do k-nearest neighbor (k-NN) aproximada e o combina com uma função de kernel (HATIE, 2002; TRAM, 2006). Em comparação com

outras abordagens menos intuitivas e mais complexas, os métodos baseados k-NN provaram produzir resultados competitivos com um grande número de problemas de classificação (BRADBURY, 2006; ZHU, 2008).

2.3.3.11. MG-RAST

O *MG-RAST* ou *Anotação Rápida usando Tecnologia de Subsistemas para Metagenômica* consiste em um sistema disponível gratuitamente e acessível on-line para o processamento de dados de sequências metagenômicas e envolve alguns tipos de análise, dentre as quais se destacam as comparações filogenéticas, anotações funcionais, entre outras, e cerca de 500 genomas já foram processados por sua versão beta. É um sistema de código aberto baseado em um quadro SEED para genômica comparativa. A submissão das sequências é feita no formato FASTA e estas são processadas automaticamente. A anotação de um ou mais genomas e metagenomas, comparação de metabolismos e filogenética são alguns dos métodos fornecidos para acessar os diferentes tipos de dados existentes. O acesso aos dados é protegido por uma senha e todos os resultados gerados são disponibilizados para serem baixados no computador do usuário. Cada usuário efetua um registro para limitar o acesso ao seu conjunto de dados e obter informações se acaso ocorra algum problema no processamento dos mesmos. Quando está conectado, o usuário pode acessar os metagenomas que enviou, os dados que proprietários permitiram acessar, os metagenomas públicos disponíveis e também pode disponibilizar seus dados a todos os usuários do sistema (MEYER et al. 2008).

2.3.4. STAMP

STAMP é um software gráfico que fornece testes de hipóteses estatísticas e gráficos exploratórios para analisar perfis taxonômicos e funcionais. Ele oferece suporte para comparação de pares de amostras ou amostras organizadas em dois ou mais grupos de

tratamento. Dimensões de efeito e intervalos de confiança são fornecidos para permitir uma avaliação crítica da relevância biológica dos resultados dos testes. Uma interface gráfica amigável permite fácil exploração de resultados estatísticos e geração de gráficos com qualidade de publicação (PARKS &BEIKO, 2010).

3. METODOLOGIA

3.1. Coleta das Amostras

As amostras foram coletadas na Embrapa Agropecuária Oeste (CPAO) localizada no município de Dourados-MS. O solo classificado como Latossolo Vermelho distroférico típico, com 70% de argila. Os sistemas de manejo foram instaurados em 1995 com diferentes tipos de preparo do solo e sequência de culturas. As amostras de solo foram coletados em: Mata Nativa (MN) (Floresta Semidecidual; S22°28'2231'', W54°81'4571'') e Sistema Plantio Direto (S22.279410, W54.812593). O manejo avaliado no presente estudo contempla sistema plantio direto (SPD) onde no verão é cultivado milho ou soja e no inverno estas culturas são rotacionadas com trigo, aveia e nabo forrageiro. Na Tabela 2, está apresentada a sequência de culturas estabelecidas no SPD.

Os solos foram coletados a uma profundidade de 0-10 cm. 10 amostras simples para se obter uma amostra composta representativa de cada sistema. Essas amostras foram armazenadas em sacos plásticos estéreis, mantidos sob refrigeração e transportados para a UNESP Jaboticabal, para sequenciamento genético, e ao laboratório de análise de solos da Universidade Federal da Grande Dourados para análise química.

Tabela 2: Rotação de culturas apresentada no solo de Plantio Direto A:aveia; S: soja; T: trigo; N: nabo; M: milho nas estações de inverno(INV) e verão (VER).

ANO	2010/11		2011/12		2012/13		2013/14	
ESTAÇÃO	INV	VER	INV	VER	INV	VER	INV	VER
Plantio Direto	A	S	T	S	N	M	A	S

3.2. Química do Solo

Para as análises de nutrientes do solo foram utilizados: Extrator Melich: K - P - Fe - Mn - Cu e Zn, Extrator KCl: Ca - Mg - Al, Extrator HCl 0,05 N: B, Extrator Fosfato de Cálcio: S, Extrator Dicromato de sódio: Carbono

Tabela 3: Resultado da Análise Química dos Solos do Sistema Plantio Direto e Mata Nativa (Cmol/dm³)

Gleba	pH	Matéria Orgânica	P	S	K	Ca	Mg	H+Al	Al
Plantio Direto	4,80	22,81	28,76	0,0	0,50	5,50	2,01	7,20	0,23
Mata Nativa	5,30	24,15	4,70	0,0	0,73	13,83	4,56	5,35	0,00

3.3. Extração do DNA

Foi realizada a extração do DNA através do DNA SPIN KIT (Mobio, USA) de acordo com instruções do fabricante. Posteriormente o DNA teve sua quantidade e qualidade verificado nos equipamentos nanodrop e quibit. Utilizou-se o seguinte procedimento:

- ✓ Adicionou-se até 500 mg de amostra de solo para um tubo de Lise Matrix E.
- ✓ Adicionou-se 978 uL de tampão fosfato de sódio para provar em tubo Lysing Matrix E.
- ✓ Adicionou-se 122 ul MT Buffer.
- ✓ Transferiu-se o sobrenadante para um tubo de microcentrífuga de 2,0 ml limpo. Adicionou-se 250 ul de PPS (Protein Precipitation Solution)
- ✓ Centrifugou-se durante 5 minutos para sedimentar o precipitado. Transferiu-se o sobrenadante para um tubo de 15 ml limpo.
- ✓ Ressuspendeu-se a suspensão Matrix Binding e adicionou-se o sobrenadante em tubo de 15 ml.
- ✓ Inverteu-se as amostras durante 2 minutos para permitir a ligação de DNA.
- ✓ Colocou-se o tubo em uma cremalheira durante 3 minutos para permitir a resolução de matriz de sílica.
- ✓ Descartou-se o sobrenadante
- ✓ Ressuspendeu-se Encadernação Matrix no montante remanescente do sobrenadante.
- ✓ Transferiu-se cerca de 600 ul da mistura para um filtro de centrifugação.

- ✓ Esvaziou-se o tubo de captura e adicionou-se a mistura restante para o Filtro de SPIN™ e centrifugou-se.
- ✓ Adicionou-se 500 ul preparado SEWS-M e ressuspendeu-se o sedimento
- ✓ Centrifugou-se e posteriormente esvaziou-se o tubo
- ✓ Suspendeu-se a Encadernação Matrix (acima do filtro SPIN) em 50-100 ul de DES (DNase /-Pyrogen isenta de água).
- ✓ Armazenou-se a 4 °C até à sua utilização.

3.4. Sequenciamento

A qualidade do sequenciamento foi verificada pelo programa FastQC (ANDREWS, 2010). Sequências ruins que apresentavam baixa qualidade ou que eram muito pequenas foram retiradas através do programa prinseq-lite (SCHMIEDER et al. 2011)

3.5. Montagem das Sequencias (Assembly)

Segundo Peng (2012), o IDBA-UD supera SOAPdenovo, Velvet e Meta-IDBA em todos os aspectos. N50 é o tamanho do maior contig de tal modo que todos os contigs maiores que este contig, cobre pelo menos metade do genoma a ser montado (44.879) contigs construídos por IDBA-UD é cerca de 10 vezes maior do que o segundo melhor N50 (4588) de contigs construídos pela Meta-IDBA. IDBA-UD também tem a melhor cobertura (99,15%), que é cerca de 10% maior do que o segundo mais alto obtido pelo Meta-IDBA. Os contigs construídos por IDBA-UD cobriam quase toda a região de três espécies, a profundidade desigual não afeta a qualidade da montagem de IDBA-UD. Semelhante ao assembly de uma única célula, andaimes em todas as assembleis produziram contigs mais longos, mas com uma cobertura menor. Quanto para a taxa de erro de substituição, IDBA-UD e Meta-IDBA têm muito maior precisão e IDBA-UD construindo um número mínimo de contigs errôneos gerados no assembly. Os tempos de execução de todas as assembleis são semelhantes, mas IDBA-UD e Meta-IDBA usaram cerca da metade da memória como SOAPdenovo e Velvet.

3.6. Identificação de ORF's

A identificação de orfs é necessária, pois o sequenciamento por NGS sequêcia DNA total. Desta forma, muito material que não é éxon também é sequenciado, para poder realizar o processo de anotação adequadamente foi realizada a identificação de orfs com o FragGeneScan (RHO et al. 2010).

3.7. Análise MG-RAST

No MG-RAST, os parâmetros padrão foram utilizados para a classificação taxonômica.

Foi utilizado o banco M5RN (Meta 5 Non-Redundant) associado ao MG-RAST para a recuperação de dados em vários bancos primários. Esta análise foi feita por meio de *Best Hit Classification – All annotation source: M5NR – e-value: 1e-5, Identify: 60%, Alignment Length Cutoff: 15.*

Com o objetivo de verificar se o esforço amostral realizado foi o suficiente para capturar os *taxa* mais abundantes e para inferir o quão diverso é cada uma das amostras, os arquivos SPD e MN, foram analisados através de curvas de rarefação. Construir uma curva de rarefação consiste em fracionar aleatoriamente e repetitivamente uma amostra para que a média do número de espécies encontradas em cada sub amostra possa ser calculada. Desta forma, é possível saber se foi possível ter acesso à uma boa amostragem da biodiversidade local.

Diversos trabalhos utilizaram o MG-RAST para estudos metagenômicos (MENDES et al. 2015; NAVARRETE et al. 2015; REAVY, et al. 2015; SOUZA, et al. 2014)

3.8. Análise Estatística

Foi realizada análise estatística dos solos dos dois sistemas utilizando a ferramenta STAMP (Statistical Analysis of Metagenomics Profiles), versão 2.1.3; por meio do teste estatístico G-test (W/Yates') + Fischer'scom taxa de confiabilidade de 95% para p-value maiores que 0,05.

4. RESULTADOS E DISCUSSÕES

4.1. Análise Química do Solo

Como pode ser visto na tabela 3, o solo de mata possui maior concentração de Cálcio do que o solo de Plantio Direto (13,83% e 5,50% respectivamente). Entretanto, o solo de Plantio Direto possui maior concentração de Fósforo do que o solo de Mata Nativa (28,76% e 4,70% respectivamente).

4.2. Sequenciamento

Foram inseridas 54.230 sequências a partir do solo de Mata Nativa no MG-RAST. 2.033 sequências (3.7%) falharam no teste de QC. Entre as sequências que foram aprovadas no teste QC, 119 sequências (0.2%) contêm genes de RNA ribossomal. Do restante, 26.944 sequências (49.7%) contêm genes que predizem proteínas de função conhecida e 24.638 sequências (45.4%) contêm genes que codificam para proteínas com função desconhecida. 496 (0.9%) das sequências que foram aprovadas no teste QC não possuem genes rRNA ou genes codificadores de proteínas.

Foram inseridas 69.164 sequências obtidas a partir do solo sob Sistema Plantio Direto no MG-RAST, sendo que 19 sequências falharam no teste de QC. 159 sequências (0.2%) contêm genes de RNA ribossomal. Do restante, 43.673 sequências (63.1%) contêm genes codificadores de proteínas com funcionalidade conhecida e 24.776 sequências (35.8%) contêm genes codificadores de proteínas com função desconhecida. 537 (0.8%) das sequências que passaram pelo teste QC não possuem genes de rRNA ou genes que predizem proteínas.

4.3. Diversidade Microbiana

As amostras ambientais representam um enorme reservatório de diversidade genética de arqueia, bactérias, eucariotos e vírus (SHARON et al. 2013).

Os resultados obtidos por meio da ferramenta MG-RAST e STAMP, apontaram uma presença maior de bactérias em comparação com arqueias, eucariotos e vírus em ambos os solos. Observando-se uma alta presença de bactérias não classificadas maior do que grandes grupos como arqueias e vírus (Figuras 1 e 2).

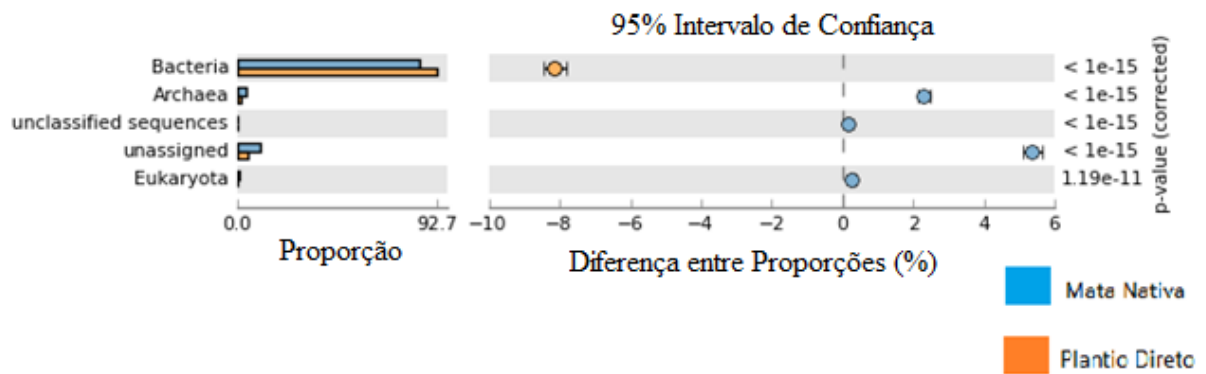


Figura 1: Gráfico estatístico da diferença de proporções entre os domínios. Obtido através do programa STAMP para solo sob Mata Nativa e solo sob Sistema Plantio Direto com intervalos de confiança de 95%.

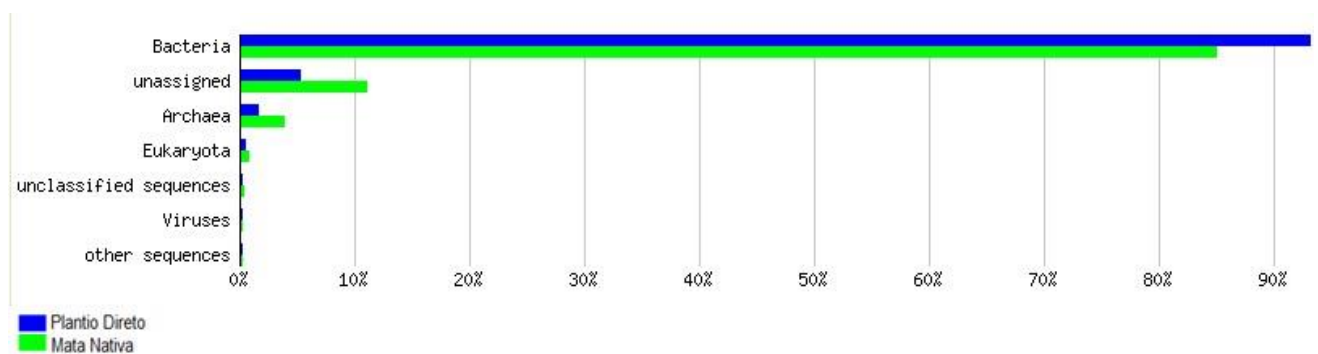


Figura 2: Gráfico de distribuição para os domínios encontrados, obtido através da ferramenta MG-RAST para Mata Nativa e Sistema Plantio Direto.

No solo sob Mata Nativa, a porcentagem de bactérias presentes foi de 84.60% - (45.39% Proteobactérias; 15.32% Actinobactérias; 5.20% Firmicutes; 4.78% Acidobactérias e 2.74% de Cianobactérias), sequências não atribuídas 10.39%; 3.93% de arqueias; 0.77%

eucariotos e sequências sem classificação 0.23%. Mendes e colaboradores (2015), também encontraram uma abundância maior de proteobactérias (61%) na floresta amazônica.

No solo sob Sistema Plantio Direto, 92.74% foram de bactérias (48.72% Actinobactérias; 33.77% Proteobactérias; 3.05% Firmicutes; Chloroflexi 1.78% e 0.97% de Cianobactérias), sequências não atribuídas de 5.01%; arqueias 1.62%; eucariotos 0.49% e sequências sem classificação, 0.061%.

Através da Figura 2, nota-se a maior proporção de bactérias no solo sob Sistema Plantio Direto em relação ao solo sob Mata Nativa. Entretanto, os solos sob Mata Nativa possuíram maior proporção dos demais domínios (eucarioto, arqueias, vírus e de sequências se classificação).

4.3.1. Curva de Rarefação

Todas as amostras apresentaram curvas de rarefação pouco íngremes à esquerda e quase planas à direita, mostrando que estava próxima de atingir o platô no caso de mata nativa, como apresentado na Figura 3, o que indica que a amostragem e o sequenciamento foram responsáveis por uma boa cobertura das espécies que compõem esse ambiente.

O platô não foi totalmente atingido, pois o gráfico está avaliando a distribuição de espécies algo extremamente diverso na natureza.

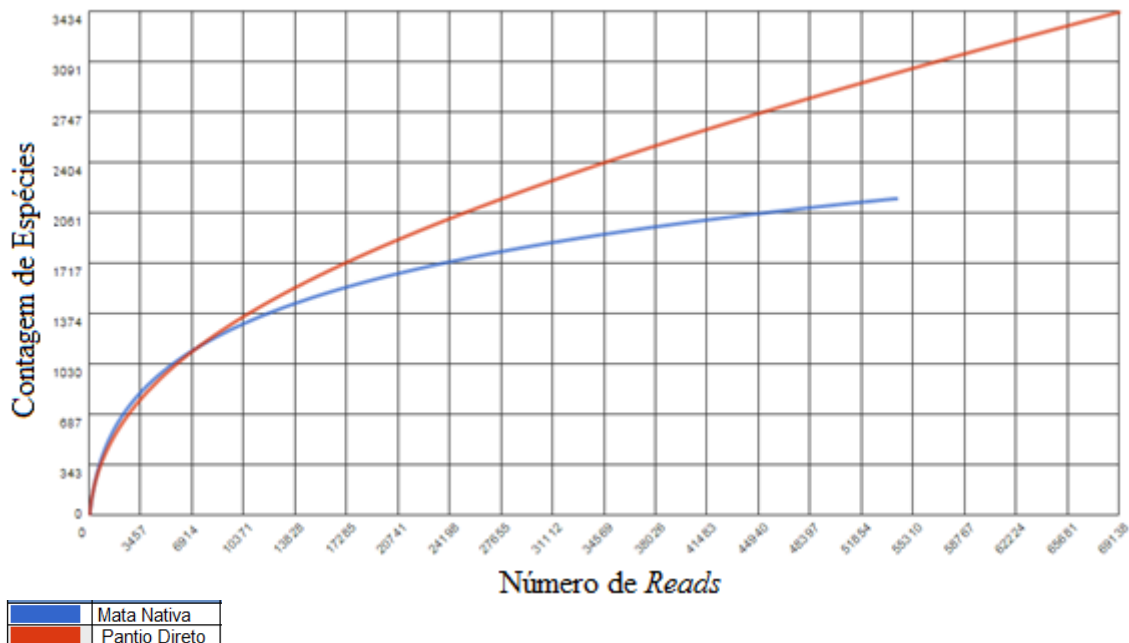


Figura 3: Curva de Rarefação de número de espécies x número de reads, obtida através da ferramenta MG-RAST

4.4. Bactérias

Dentre as Bactérias, encontramos maior presença do filo Actinobactéria em solos sob sistema de plantio direto. As Proteobactérias foram o segundo filo com maior presença no solo sob sistema plantio direto seguidos pelas Acidobactérias e Firmicutes (Figura 4 e 5).

No solo sob mata nativa, se encontra em maior número as Proteobactérias, que são grande colaboradoras no ciclo do nitrogênio, carbono e enxofre no solo.

ZHALNINA et al. (2014) realizaram um trabalho em diferentes concentrações de pH de solo e encontrou 97.5% de seqüências que foram classificadas como bactérias, 0.3% como arqueias, e 2.2% permanecem desconhecidas.

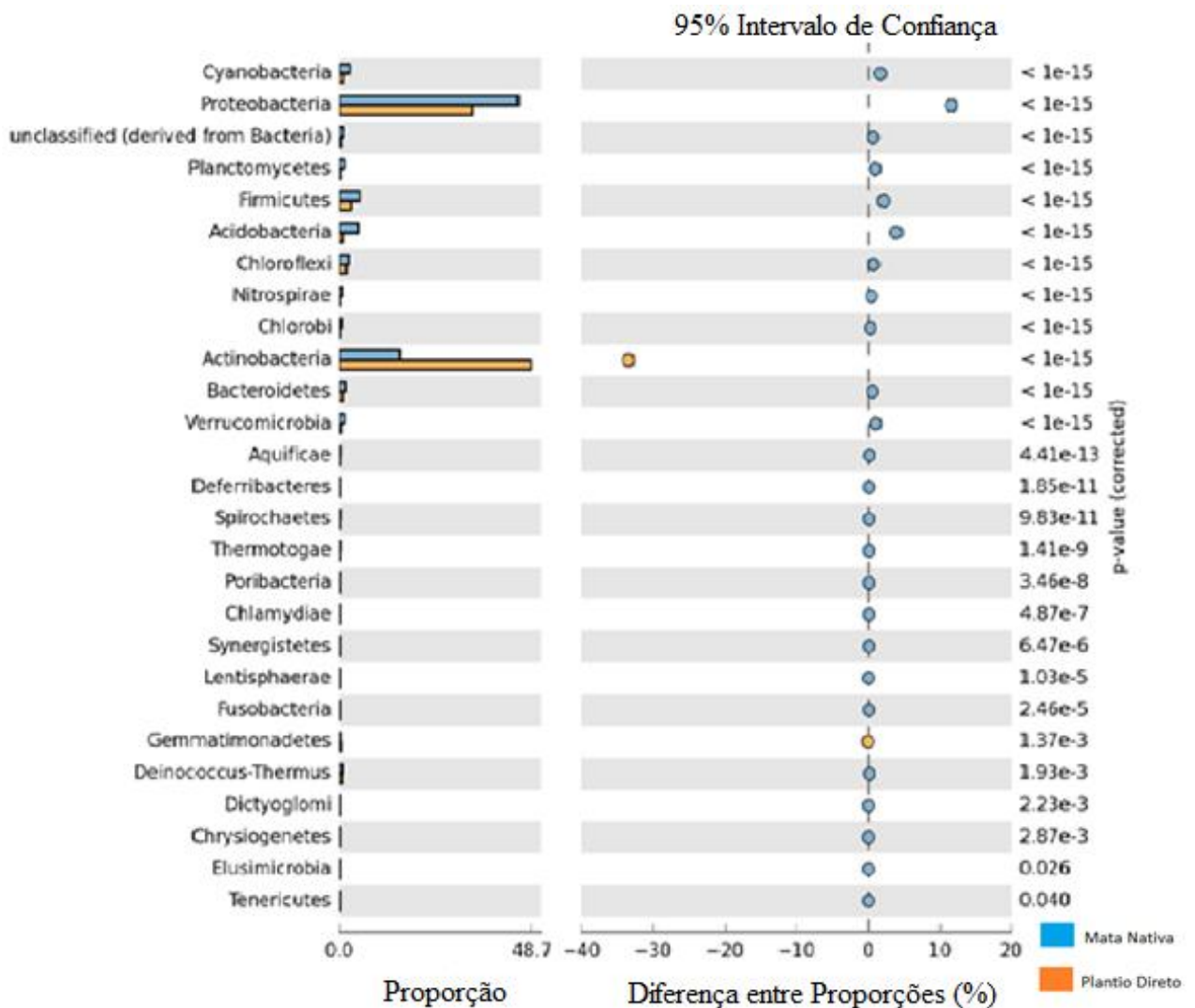


Figura 4: Gráfico estatístico com a diferença de proporções entre os filamentos do domínio bactéria obtidos através do programa STAMP para solo sob Mata Nativa e solo sob Sistema Plantio Direto com intervalo de confiança de 95%.

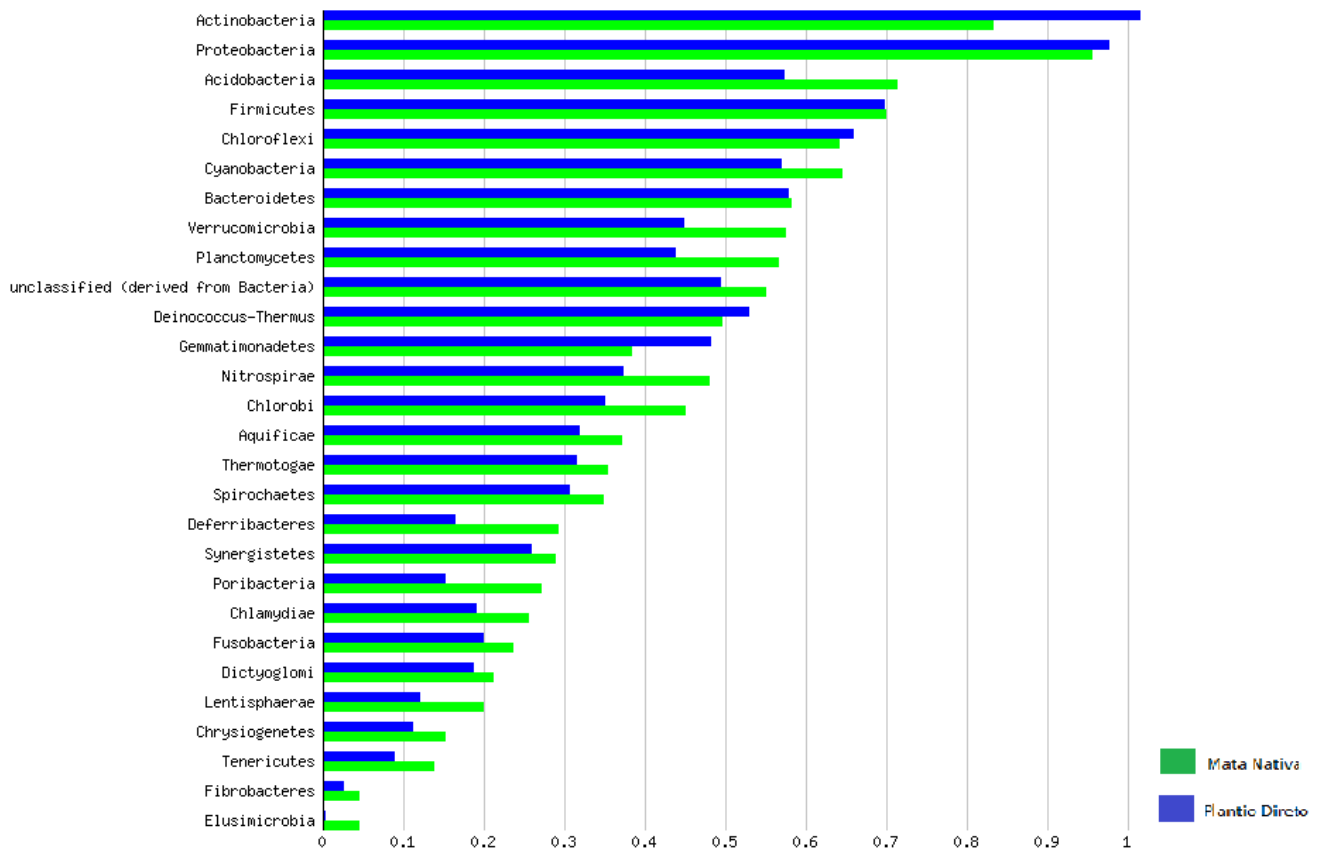


Figura 5: Gráfico de distribuição para os filos encontrados no domínio bactéria, obtido através da ferramenta MG-RAST para solos sob Mata Nativa e sob Sistema Plantio Direto.

4.4.1. Actinobactérias

Resultados de análise estatística do STAMP demonstraram uma concentração de Actinobactérias de 15.32% e no MG-RAST de 16.7%. Dentro do domínio Bactéria, as Actinobactérias representaram 18.11% e 17.89% segundo o STAMP (Figura 4) e o MG-RAST (Figura 5) respectivamente no sistema MN.

No SPD, o STAMP (Figura 4) encontrou uma concentração de Actinobactérias de 48.72% e no MG-RAST de 52.1%. Dentro do domínio Bactéria, as Actinobactérias representaram 52.53%, 53.68% no STAMP e no MG-RAST, respectivamente.

Portanto, esse filo se apresentou em maior proporção no solo sob SPD do que no solo sob MN (floresta semidecidual). Este filo costuma ser um dos principais observados em diversos estudos metagenômicos (NACKE et al, 2014; RAMIREZ et al. 2014; BARH et al.

2015). CARBONETTO (2014) o filo Actinobactéria como aquele mais comumente encontrado em solos não cultiváveis. O manejo do sistema plantio direto promove o aumento do filo Actinobactéria, que pode ser explicada pela acumulação de resíduos de plantas (MELNICHUK et al. 2015)

As Actinobactérias são bactérias Gram positivas filamentosas, formadoras de esporo, com um alto conteúdo G+C em seu DNA e que estão amplamente distribuídas na natureza, variando aproximadamente de 55% a 70% em bactérias do gênero *Streptomyces* (BERVANAKIS, 2008). Ocorrem frequentemente no solo, em especial na rizosfera (PROCÓPIO et al., 2012; VASCONCELLOS et al., 2010). Estas bactérias filamentosas apresentam ciclo de vida saprofítico e a maioria é aeróbica, entretanto, alguns gêneros são anaeróbios facultativos ou obrigatórios, como o gênero *Actinomyces* (BERNARDO, 2012).

Dentre os microrganismos benéficos, as Actinobactérias pertencentes ao gênero *Streptomyces* desempenham um papel importante na rizosfera de plantas, por influenciarem o crescimento das plantas e atuarem no controle de fitopatógenos, devido à produção de antibióticos e outros metabólitos secundários (DAMASCENO, 2011)

As Actinobactérias, por apresentar características de promover o crescimento de plantas, têm sido amplamente estudadas nas últimas décadas como uma alternativa ao uso de defensivos agrícolas (BETTIOL; GHINI, 1995; OLIVEIRA et al., 2003).

Apresentam diferentes propriedades fisiológicas e metabólicas, como a produção de enzimas extracelulares e a formação de uma grande variedade de metabólitos secundários, dos quais muitos são potentes antibióticos (VENTURA et al. 2007; GENILLOUD et al. 2011; GEORGE et al. 2012; GOODFELLOW & FIEDLER, 2013).

Cerca de 75% dos antibióticos conhecidos são naturalmente produzidos a partir de Actinobacteria e particularmente os que pertencem ao gênero *Streptomyces*. Eles também detêm uma posição de destaque devido a sua capacidade de produzir diversos outros metabólitos secundários de interesse comercial (antivirais, agentes antitumorais, imunossuppressores, biopesticidas, agentes anti-helmínticas, enzimas etc.) (BOUIZGARNE et al. 2014)

4.4.2. Proteobactérias

Resultados de análise estatística do STAMP demonstraram uma concentração de Proteobactérias de 45.39% e no MG-RAST de 50.3%. Dentro do domínio Bactéria, as Proteobactérias representaram 53.65% e 55.00% segundo o STAMP e o MG-RAST, em mata nativa (Figuras 4 e 5).

No sistema sob plantio direto, o STAMP encontrou uma concentração de Proteobactérias de 33.77% e no MG-RAST de 34.1%. Dentro do domínio Bactéria, as Proteobactérias representaram 36.41% e 36.40% no STAMP e no MG-RAST, respectivamente. Resultados semelhantes também foram encontrados nos estudos de ORTIZ (2014).

Apesar da concentração de Proteobactérias também ser alta no solo sob plantio direto, as Proteobactérias foram o segundo grupo mais abundantes, no solo de mata nativa, as Proteobactérias foram o grupo de organismos maior ocorrência.

FIERER et al. (2007) relataram a abundância de Proteobactérias correlacionadas com as taxas de carbono mineralizado em regiões quentes. Da mesma forma, outros estudos sugerem que as Proteobactérias englobam uma enorme diversidade ao nível morfológico, fisiológico e metabólico, os quais desempenham um papel significativo nos ciclos do carbono, nitrogênio e enxofre do solo (KERSTERS et al., 2006; ACOSTA-MARTÍNEZ et al., 2010)

ACOSTA-MARTÍNEZ et al. (2010) também encontraram Actinobactéria e Proteobactéria como filos mais abundantes em solos de desertos, onde esses grupos são associados ao valor baixo de pH do solo e a altas temperaturas. Além disso, são encontrados na natureza, grupos de Proteobactérias desempenhando papel ativo no ciclo do nitrogênio, principalmente as β -proteobactéria e γ -proteobactéria amônio-oxidantes, sendo representadas as *Nitrosomonas*, *Nitrospira* e *Nitrosococcus* (PURKHOLD et al., 2003, COBO-DIAZ et al. 2015). As Proteobactérias se destacam por estarem envolvidas no processo de fixação biológica de nitrogênio (VAL-MORAES, 2008).

TRIVEDI et al. 2012 concluíram que entre, os filos colonizadores da rizosfera, as proteobactérias foram os mais abundantes na região de rizosferas saudáveis, enquanto filos como Acidobactéria, Actinobactéria, e Firmicutes foram mais dominantes no solo em massa.

4.4.3. Firmicutes

Resultados de análise estatística do STAMP demonstram uma concentração de Firmicutes de 5.20% e no MG-RAST de 5.7%. Dentro do domínio Bactéria, os Firmicutes representaram 6.15% e 5.19% segundo o STAMP e o MG-RAST respectivamente no sistema sob mata nativa (Figuras 4 e 5).

No sistema plantio direto, o STAMP demonstrou uma concentração de Firmicutes de 3.05% e no MG-RAST de 3.3%. Dentro do domínio Bactéria, os Firmicutes representaram 3.29% e 2.75% no STAMP e no MG-RAST, respectivamente (Figuras 4 e 5).

Apesar do número aproximado de bactérias desse grupo ser encontrado em ambos os solos, o solo sob sistema de plantio direto teve, nesse grupo, como o terceiro maior filo de bactérias apresentado, enquanto no solo sob Mata Nativa verificou-se uma maior concentração de Acidobactérias do que Firmicutes.

RODRIGUES et al. (2012), que estudaram a conversão da floresta para pasto na Amazônia, sugeriram que o filo Firmicutes é resistente a alta amplitude das variações na temperatura do solo, tais como os que ocorrem naturalmente ao longo do dia e as estações, indicando que esta poderia ser uma característica geral observada em áreas de pastagem.

COBO-DÍAZ (2015), em sua pesquisa de metagenômica em solo após queima, afirma que os Firmicutes possuem um importante papel na fixação de nitrogênio. Ao comparar a biodiversidade de deserto com terra cultivável, KÖBERL et al. (2012) observaram que houve uma correlação entre a abundância de Firmicutes e a disponibilidade de água.

4.5. Arqueias

Em solo sob mata nativa, o domínio arqueia teve uma maior proporção quando comparado ao solo sob sistema plantio direto, como pode se observar na figura 6.

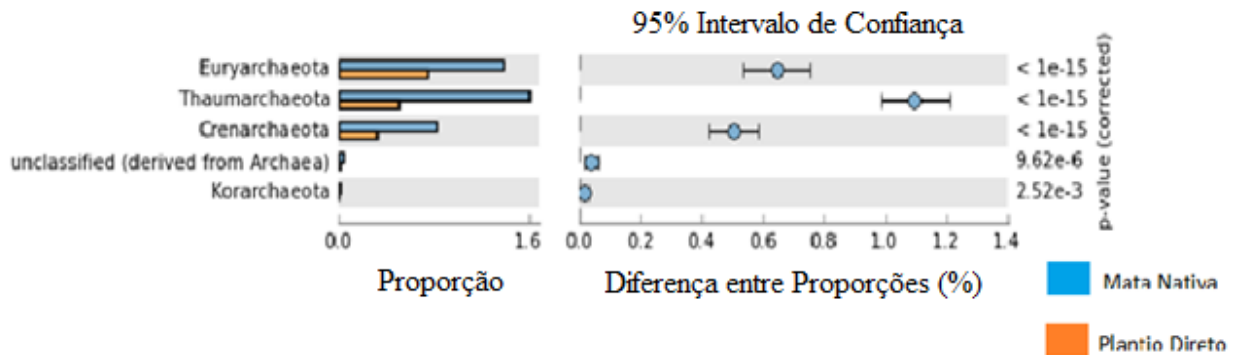


Figura 6: Gráfico estatístico com a diferença de proporções entre os filões do domínio arqueia obtidos através do programa STAMP para solo sob Mata Nativa e solo sob Sistema Plantio Direto com intervalo de confiança de 95%.

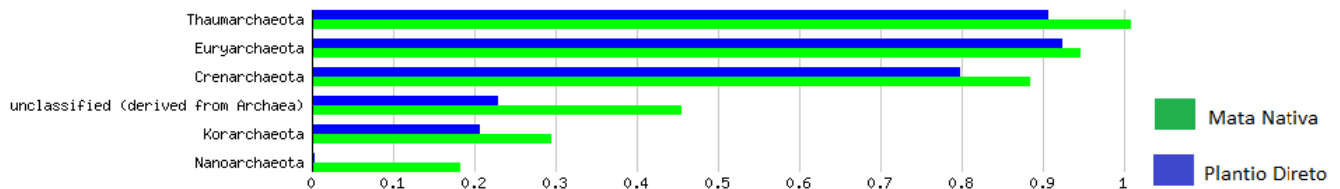


Figura 7: Gráfico de distribuição do domínio arqueia nos solos sob Mata Nativa e sob Sistema Plantio Direto apresentados pelo programa MG-RAST

4.5.1. Thaumarchaeota

Resultados de análise estatística do STAMP demonstrou uma concentração de Thaumarchaeotas de 1.61% e no MG-RAST de 1.9%. Dentro do domínio arqueia, as Thaumarchaeotas representaram 41.02% e 45.93% segundo o STAMP e o MG-RAST respectivamente em solo sob mata nativa (Figuras 6 e 7).

No solo sob sistema plantio direto, o STAMP encontrou uma concentração de Thaumarchaeotas de 0.51% e no MG-RAST de 0.5%. Dentro do domínio arqueias, as Thaumarchaeotas representaram 31.87% e de 37.66% no STAMP e no MG-RAST, respectivamente (Figura 6 e 7).

O filo Thaumarchaeota foi um dos filões mais encontrados dentro das arqueias. Sua presença foi maior no solo sob mata nativa do que no solo sob sistema plantio direto.

Em pesquisas de metagenoma realizados na floresta amazônica (TAKETANI, 2010) e na mata atlântica (ETTO et al. 2012), a Thaumarchaeota foi um dos filões de arqueias mais

comumente encontrados. A comunidade de arqueias no latossolo do bioma Cerrado é predominantemente composta pelo filo Thaumarchaeota. Esses solos têm um maior teor de matéria orgânica, mas são descritos como ácido (CATÃO et al. 2013).

Em trabalho realizado por CALDWELL (2015), foram encontrados resultados semelhantes onde filo Thaumarchaeota representou a maior parte das sequências de arqueias analisadas (89.7%), seguido pelas de Euryarchaeotas (8.3%). BATHIA (2015) apresentou amostras de Archaeas apenas do filo, Thaumarchaeota. Este filo também foi o maior representante do grupo encontrado no estudo realizado por PACCHIONI (2014) na cidade de João Câmara no estado de Rio Grande do Norte.

Os Thaumarchaeota, são reconhecidos como sendo um dos mais importantes filios responsáveis pela nitrificação no mundo (MONTEIRO et al. 2014). Parece também haver uma associação entre os membros do filo Thaumarchaeota com ambientes ácidos. No entanto, tanto o pH quanto as condições de nutrientes parecem ser importantes condutores do nicho de diferenciação dentro do Thaumarchaeota (AUGUET & CASAMAYOR, 2013).

Com o avanço de tecnologias de sequenciamento de última geração, abordagens metagenômicas e o potencial para a cultura do Thaumarchaeota espera-se compreender melhor tanto as atividades fisiológicas e os mecanismos da seleção microbiana dentro destes ecossistemas (BARTON, et al. 2014).

4.5.2. Euryarchaeota

Resultados de análise estatística do STAMP demonstraram uma concentração de Euryarchaeotas de 1.40% e no MG-RAST de 1.5%. Dentro do domínio arqueia, as Euryarchaeotas representaram 35.65% e 30.84% segundo o STAMP e o MG-RAST respectivamente em solo sob mata nativa (Figuras 6 e 7).

No solo sob sistema plantio direto, o STAMP resultou numa concentração de Euryarchaeotas de 0.75% e no MG-RAST de 0.7%. Dentro do domínio arqueia, as Euryarchaeotas representaram 46.60% e de 41.93% no STAMP e no MG-RAST respectivamente (Figuras 7 e 8).

O filo Euryarchaeota foi o de maior ocorrência no solo sob sistema plantio direto. Entretanto, sua concentração foi próxima ao filo Thaumarchaeota, segundo mais encontrado no solo sob sistema plantio direto e o principal filo encontrado no solo sob mata nativa.

HU et al. (2013) encontrou em seu estudo realizado em terrenos fofos (paddy lands) 64% de Euryarchaeota, sendo um dos mais abundantes no estudo de arqueias realizado por ele.

O Euryarchaeota engloba tanto mesófilos, termófilos e psicrófilos, incluindo acidófilos, alcalifilos e halófilos. As arqueias metanogênicas também estão incluídas dentro das Euryarchaeotas, vivendo em associação com bactérias e usando produtos metabólicos bacterianos como um substrato para a metanogênese (OFFRE et al. 2013; BOENIGK et al. 2015).

4.5.3. Crenarchaeota

Resultados de análise estatística do STAMP demonstrou uma concentração de Crenarchaeotas de 0.83% e no MG-RAST de 0.6%. Dentro do domínio arqueias, as Crenarchaeotas representaram 21.16% e 21.00% segundo o STAMP e o MG-RAST, respectivamente, na mata nativa.

No sistema plantio direto, o STAMP demonstrou uma concentração de Crenarchaeotas de 0.32% e no MG-RAST de 0.3%. Dentro do domínio arqueia, as Crenarchaeotas representaram 20.27% e de 19.38% no STAMP e no MG-RAST, respectivamente. TIAN (2015) encontrou em seu estudo, as Crenarchaeotas como o filo mais abundante das arqueias em análise de lodo.

O filo Crenarchaeotas foi o terceiro filo mais encontrado em ambos os solos, possuindo maior concentração na mata nativa (Figuras 6 e 7). O filo Crenarchaeota, é conhecido por ser altamente abundante em solo ácido da floresta temperada (DING et al. 2015).

Crenarchaeota é um dos quatro filos de arqueias. As espécies hipertermófilos de Crenarchaeota tendem a se agrupar em conjunto e ocupam ramos curtos no gene na árvore filogenética de 16S rRNA. Estes organismos são considerados bons modelos para os iniciais de arqueias. Filogeneticamente, é uma espécie de evolução rápida. A maioria dos Crenarchaeota hipertermófilos foi isolado de solos e águas aquecidas geotérmicas, contendo enxofre e sulfuretos elementares, ou fontes hidrotermais. Entre eles, podemos encontrar membros das ordens Sulfolobales, Thermoproteales, e Desulfurococcales. Uma espécie de *Sulfolobus*, *S. acidocaldarius*, foi a primeira arqueia hipertermófila descoberta (AMILS, 2015). Recentemente, os Crenarchaeotas mesofílicos, foram classificados em Thaumarchaeota. Desta

forma, o filo Crenarchaeota ficou restrito a uma única classe: Thermoprotei. Esta classe é normalmente associada a ambientes extremos, por exemplo, e ambientes quentes e ácidos. Entretanto, em algumas bases de dados, os Thaumarchaeotas ainda são classificados como Crenarchaeotas (OFFRE et al. 2013).

Segundo SHEN (2013), uma investigação mais completa da função microbiana do solo, particularmente do domínio arqueia, a respeito de sua abundância, diversidade e fatores de condução, é muito necessária para uma melhor compreensão do uso da terra e seus impactos nos ecossistemas do solo. As Archaea ainda podem ser fortes bioindicadores de qualidade do solo (SOUZA et al. 2013).

4.6. Eucariotos

Os eucariotos tiveram uma maior proporção em solo sob mata nativa do que no solo sob sistema plantio direto, como pode se observar na figura 8.

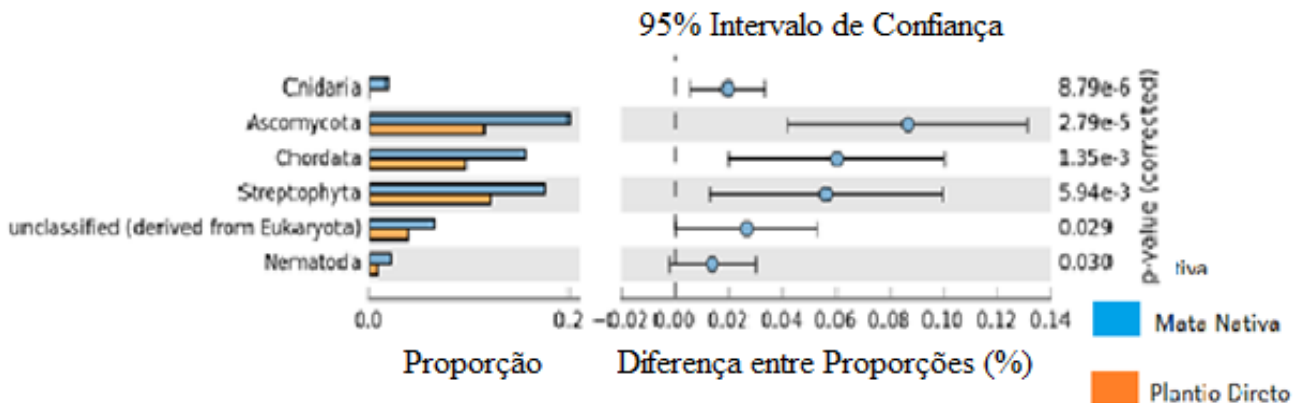


Figura 8: Gráfico estatístico com a diferença de proporções entre os filamentos do domínio eucariótico obtidos através do programa STAMP para solo sob Mata Nativa e solo sob Sistema Plantio Direto com intervalo de confiança de 95%.

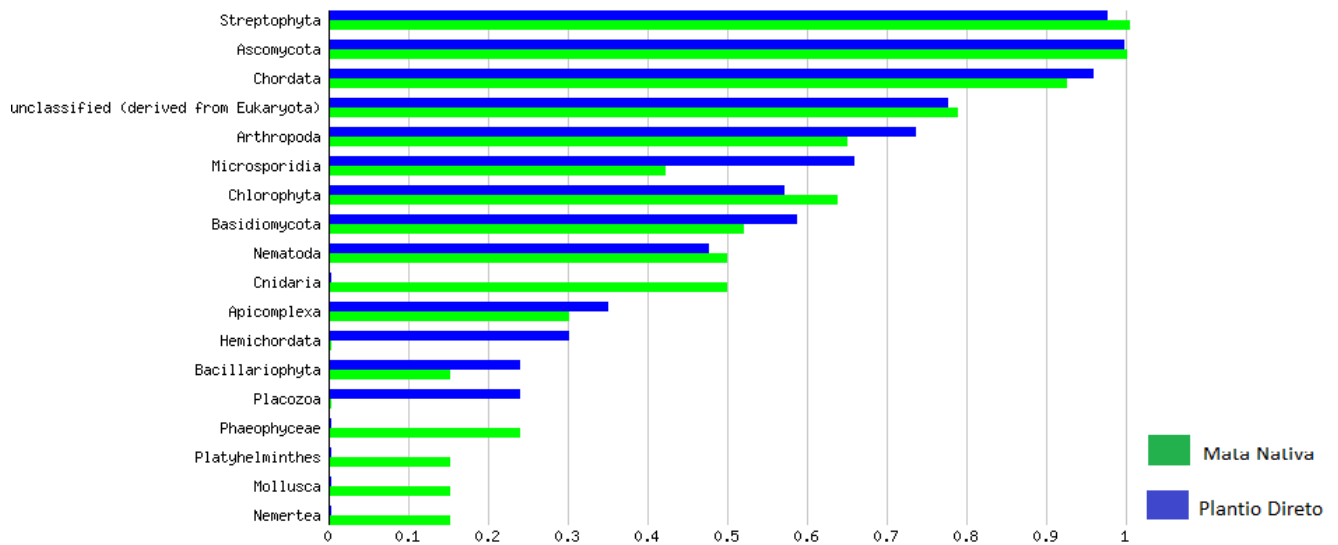


Figura 9: Gráfico de distribuição do domínio eucarioto nos solos sob Mata Nativa e sob Sistema Plantio Direto apresentados pelo programa MG-RAST

4.6.1. Ascomycota

Resultados de análise estatística do STAMP no solo sob mata nativa, indicou uma concentração de Ascomycotas de 0.20% e no MG-RAST de 0.2%. Dentro do domínio Eucarioto, as Ascomycotas representaram 26.33% e 25.70%, segundo o STAMP e o MG-RAST, respectivamente (Figuras 8 e 9).

No sistema plantio direto, o STAMP demonstrou uma concentração de Ascomycotas de 0.11% e no MG-RAST de 0.2%. Dentro do domínio Eucarioto, as Ascomycotas representaram 23.57% e de 24.63% no STAMP e no MG-RAST, respectivamente (Figuras 9 e 10).

As Ascomycotas foram o filo de Eucariotos mais encontrado no solo sob sistema plantio direto e o segundo mais encontrado no solo de Mata Nativa. Entretanto, as concentração desse filo foram semelhantes em ambos os solos.

Os Ascomycotas são um grupo irmão das Basidiomycotas dentro dos fungos quitinosos. São predominantemente terrestre, embora algumas espécies também pode ser encontradas em ambientes aquáticos (BOENIGK et al. 2015).

4.6.2. Streptophyta

Resultados de análise estatística do STAMP na mata nativa, indicaram uma concentração de Streptophytas de 0.12% e no MG-RAST de 0.2%. Dentro do domínio Eucarioto, as Streptophytas representaram 23.34% e 26.46% segundo o STAMP e o MG-RAST, respectivamente (Figuras 8 e 9).

No sistema plantio direto, o STAMP indicou uma concentração de Streptophytas de 0.005% e no MG-RAST de ~0.0%. Dentro do domínio Eucarioto, as Streptophytas representaram 25.00% e de 22.39% no STAMP e no MG-RAST, respectivamente (Figuras 9 e 10).

Entre as Streptophytas, o filo de Eucariotos foi mais encontrado no solo sob mata nativa. Entretanto, percebe-se sua abundância em ambos os solos sendo o segundo mais abundante no solo sob sistema plantio direto.

Estudo realizado por KHODAKOVA (2014), com sequenciamento *shotgun* e WGA (*whole-genome amplification*), encontrou Streptophyta como um dos principais eucariotos no solo analisado em estudo forense e TIAN (2015) observa como Streptophyta o maior filo de maior ocorrência dos eucariotos estudados em seu trabalho.

As Streptophytas incluem todas as plantas terrestres (Embriófitas), bem como uma coleção diversificada de algas verdes de água doce, como o Mesostigmatophytina, Chlorokybophytina, Klebsormidiophytina, Zygnematophytina, Coleochaetophytina, e Charophytina (BOENIGK et al. 2015)

4.6.3. Chordata

Resultados de análise estatística do STAMP na mata nativa, indicaram uma concentração de Chordata de 0.09% e no MG-RAST de 0.2%. Dentro do domínio Eucarioto, as Chordata representaram 20.55% e 18.07%, segundo o STAMP e o MG-RAST respectivamente (Figura 8 e 9).

No sistema plantio direto, o STAMP verificou uma concentração de Chordata de 0.001% e no MG-RAST de ~0.0%. Dentro do domínio Eucarioto, as Chordata representaram 19.91% e de 20.65% no STAMP e no MG-RAST, respectivamente (Figuras 9 e 10). Sendo

assim, foi observado que o filo Chordata, foi o terceiro mais filo de Eucariotos mais encontrados no solo de Plantio Direto e um dos principais encontrados ainda, em solo da mata nativa. KHODAKOVA (2014); GIOINGO (2012) e SCHNEIDER (2015) também encontram o filo Chordata como um dos principais eucariotos em seus estudos.

4.7. Vírus

Segundo MG-RAST, foram encontrados ainda, 0.03% e 0.04% de sequências derivadas de vírus da mata nativa e sob sistema plantio direto, respectivamente. Já o STAMP indicou 0.02% (17 sequências) e 0.03% (36 sequências) respectivamente para mata nativa e sistema plantio direto, de sequências derivadas de vírus.

Os vírus nos solos podem ter, pelo menos, duas funções importantes. Em primeiro lugar, como agentes patogênicos que têm a capacidade de modular as populações de microrganismos do solo e, por conseguinte, alterar a biologia e funções dos solos. Em segundo lugar, por agentes da transferência horizontal de genes. Os vírus podem facilitar a rápida adaptação de bactérias do solo às mudanças nas condições pela transferência de genes benéficos (REAVY, 2015). ORTIZ (2014) também encontrou número de sequências correspondentes a vírus extremamente baixas, 0,13%.

5. CONCLUSÕES

Os dados sugerem que mudanças no uso da terra, assim como o tipo de cobertura vegetal podem mudar propriedades químicas do solo e, conseqüentemente, alterar grupos microbianos que ali vivem. Essas mudanças na composição microbiana podem ter efeitos sobre a funcionalidade do ambiente. Sendo assim, efeito à longo prazo da mudança no uso da terra nas comunidades microbianas é de suma importância para estudos de diversidade e para maior compreensão dos ciclos biogeoquímicos e é, em última instância, necessário para avaliar a sustentabilidade de diferentes sistemas de manejo.

Apesar da diferença no tratamento do solo, observou-se que os filos de maior abundância estavam presentes em ambos os sistemas apresentados, tanto na mata nativa quanto no sistema plantio direto, variando basicamente apenas em suas concentrações.

Além disso, o grande número de sequências que não puderam ser atribuídas, chegando a 10.39% demonstra que os dois sistemas de manejo avaliados possuem uma grande riqueza microbiana inexplorada e o grande caminho que a metagenômica ainda tem a percorrer para desvendar a vasta diversidade microbiana nos ecossistemas e as suas interações.

6. REFERÊNCIAS BIBLIOGRÁFICAS

- ACOSTA-MARTÍNEZ, V.; DOWD, S.E.; SUN, Y.; WESTER, D.; ALLEN, V.G. Pyrosequencing analysis for characterization of soil bacterial populations as affected by an integrated livestock-cotton production system. **Applied Soil Ecology**, Amsterdam, v. 45, p. 13-25, 2010.
- AMILS, R. Crenarchaeota. **Encyclopedia of Astrobiology**. DOI: 10.1007/978-3-642-27833-4_368-2. 2015.
- ANDREWS, S. “FastQC: A quality control tool for high throughput sequence data”. <http://www.bioinformatics.bbsrc.ac.uk/projects/fastqc/.2010>.
- ARAÚJO, A.S.F.; TSAI, S.M.; BORGES, C.D.; CESARZ, S.; EISENHAUER, N. Soil bacterial diversity in degraded and restored lands of Northeast Brazil. **Antonie Van Leeuwenhoek** 106:891–899. 2014.
- AUGUET, J.C.; CASAMAYOR, E.O.; Partitioning of Thaumarchaeota populations along environmental gradients in high mountain lakes. **FEMS Microbiol. Ecol.** 84, 154–164. doi: 10.1111/1574-6941.12047. 2013.
- BAKER, G.C., COWAN, D.A.; 16S rDNA primers and the unbiased assessment of thermophile diversity. **Biochem Soc Trans**, 32(Pt 2):218–221, Apr 2004.
- BARH, D.; et al. PlantOmics: The Omics of Plant Science. **Springer India**. DOI 10.1007/978-81-322-2172-2_22. 2015.
- BARTON, H.A.; GIARRIZZO, J.G.; SUAREZ, P.; ROBERTSON, C.E.; BROERING, M.J.; BANKS, E.D.; VAISHAMPAYAN, P.A.; VENKATESWARAN, K.; Microbial diversity in a Venezuelan orthoquartzite cave is dominated by the Chloroflexi (Class Ktedonobacterales) and Thaumarchaeota Group I.1c. **Frontiers in Microbiology**. 2014.
- BERNARDO, G.R.B.; “Atividade Antifúngica de Actinobactérias Da Rizosfera de Terminalia fagifolia (Bioma Caatinga) Ativas Contra Candida Spp.” **Trabalho de Conclusão de Curso. Universidade Federal de Pernambuco**. 2012.
- BERVANAKIS G.; Detection and Expression of Biosynthetic Genes in Actinobacteria. **Tese de mestrado em Ciências**. Departamento de Biotecnologia Médica, Escola de Medicina, Faculdade de Ciências da Saúde. Flinders University. 2008.
- BETTIOL, W.; GHINI, R. Controle Biológico. In: BERGAMIN, A.F.; KIMATI, H.; AMORIN, L. “Manual De Fitopatologia. Princípios E Conceitos.” 3ª ed. São Paulo: **Agronômica Ceres**, p. 717-728. 1995.
- BOUIZGARNE, B., AOUAMAR, A.A.T., Bacterial Diversity in Sustainable Agriculture, Sustainable Development and Biodiversity. **Springer**. DOI 10.1007/978-3-319-05936-5_3. 2014.

BRADBURY, D.B.; DUBITZKY, I.; Instance-based concept learning from multiclass DNA microarray data. **BMC Bioinformatics**. **7**:73. 2006.

BRADY, A. SALZBERG, S.L.; Phymm and PhymmBL: metagenomic phylogenetic classification with interpolated Markov models. **Nat Methods**. 2009.

CALDWELL, A.C.; SILVA, L.C.F.; DA SILVA, C.C.; OUVENERY, C.C.; Prokaryotic Diversity in the Rhizosphere of Organic, Intensive, and Transitional Coffee Farms in Brazil. **PLoS ONE** **10**. doi:10.1371/journal.pone.0106355. 2015.

CAPORASO, J.G.; LAUBER, C.L.; WALTERS, W.A.; BERG-LYONS, D.; FIERER, H.J.; OWENS S.M.; BETLEY J.; FRASER, L.; BAUER, M.; GORMLEY, N.; GILBERT, J.A.; SMITH, G.; KNIGHT, R.; Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms. **ISME J**. **6**:1621–1624. 2012.

CARBONETTO, B.; RASCOVAN, N.; ÁLVAREZ, R.; MENTABERRY, A.; VÁZQUEZ, M.P.; Structure, Composition and Metagenomic Profile of Soil Microbiomes Associated to Agricultural Land Use and Tillage Systems in Argentine Pampas. **Plos One**. 2014.

CARDOSO A.M.; CLEMENTINO M.M.; VIEIRA R.P.; CAVALCANTI J.J.V.; ALBANO R.M.; MARTINS O.B.; Archaeal Metagenomics: Bioprospecting Novel Genes and Exploring New Concepts. **Metagenomics: Theory, Methods, and Applications**. Wymondham: Caister Academic Press. v. 1, p. 159-169. 2010.

CARR, R.; SHAI S.; SHEN-ORR; BORENSTEIN, E. Reconstructing the Genomic Content of Microbiome Taxa through Shotgun Metagenomic Deconvolution. **PLOS Computational Biology**. 2013.

CASE, R.J.; BOUCHER, Y.; DAHLLOF, I.; HOLMSTROM, C.; DOOLITTLE, W.F.; KJELLEBERG, S.; Use of 16S rRNA and rpoB genes as molecular markers for microbial ecology studies. **Appl. Environ. Microbiol.** vol. 73 no.1 p278-288. 2007.

CATÃO, E.; CASTRO, A. P.; BARRETO, C. C.; KRUGER, R.H.; KYAW, C. M.; Diversity of Archaea in Brazilian savanna soils. **Arch Microbiol**. 195:507–512. 2013.

CHAGNON, P.L.; BRADLEY, R.L.; MAHERALI, H.; KLIRONOMOS, J.N. A trait-based framework to understand life history of mycorrhizal fungi. **Trends in Plant Science**, **18**: 484-491, 2013.

CHATTERJI, S.; YAMAZAKI, I.; BAI, Z.; EISEN, J.A.; CompostBin: A DNA composition-based algorithm for binning environmental shotgun reads. **Springer-Verlag**; p17-28. 2008.

CHAUHAN, P.S.; CHAUDHRY, V.; MISHRA, S.; NAUTIYAL, C.S.; Uncultured bacterial diversity in tropical maize (*Zea mays* L.) rhizosphere. **J. Basic Microbiol.**, v. 51, p. 15-32, 2011.

COBO-DÍAZ, J.F.; FERNÁNDEZ-GONZÁLEZ, A.J.; VILLADAS, P.J.; TORO, N.; FERNÁNDEZ-LÓPEZ M.; Metagenomic Assessment of the Potential Microbial Nitrogen Pathways in the Rhizosphere of a Mediterranean Forest After a Wildfire. **Springer**.2015.

CONAB (Companhia Nacional de Abastecimento), 2014. Available at: www.conab.gov.br

CULLIGAN, E.P.; SLEATOR, R.D.; MARCHESI, J.R.; HILL, C.; Metagenomics and novel gene discovery: Promise and potential for novel therapeutics. **Virulence**, 5, pp. 399–412. 2014.

DALEVI, D.; IVANOVA, N.N.; MAVROMATIS, K.; HOOPER, S.D.; SZETO, E.; HUGENHOLTZ, P.; KYRPIDES, N.C.; MARKOWITZ, V.M. Annotation of metagenome short reads using proxygenes. **Bioinformatics**, 24(16):i7–13, Aug. 2008.

DAMASCENO, J.C.A.; Actinobactérias na promoção de crescimento e controle de *Meloidogyne javanica* em mudas de tomateiro. **Dissertação (Mestrado em Fitotecnia)**. Mestrado em Ciências Agrárias, Universidade Federal do Recôncavo da Bahia, Cruz das Almas. 2011.

DING, J.; ZHANG, Y.; DENG, Y.; CONG, J.; LU, H.; SUN, X.; YANG, C.; YUAN, T.; NOSTRAND, J.D.V.; LI, D.; ZHOU, J.; YANG, Y.; Integrated metagenomics and network analysis of soil microbial community of the forest timberline. **SCIENTIFIC REPORTS** | 5 : 7994 | DOI: 10.1038/srep07994. 2015.

DURIGAN, G.; FRANCO, G. A. D. C.; SAITO, M.; BAITELLO, J. B. Estrutura e diversidade do componente arbóreo da floresta na Estação Ecológica dos Caetetus, Gália, SP. **Revista Brasileira de Botânica**, São Paulo, v. 23, n. 4, p. 371-383, 2000.

ETTO, R.M.; CRUZ, L.M.; JESUS, E.C.; GALVÃO, C.W.; GALVÃO, F.; SOUZA, E.M.; PEDROSA, F.O.; STEFFENS, M.B.R.; Prokaryotic communities of acidic peatlands from the southern Brazilian Atlantic Forest. **Braz J Microbiol** 43(2):1–8 S. 2012.

FIERER, N.; BRADFORD, M.A.; JACKSON, R.B. Toward an ecological classification of soil bacteria. **Ecology**, Washington, DC v. 88, p. 1354-1364. 2007.

GENILLOU, O.; GONZÁLEZ, I.; SALAZAR, O.; MARTÍN, J.; TORMO, J.R.; VICENTE, F.; Current approaches to exploit actinomycetes as a source of novel natural products. **Ind Microbiol Biotechnol** 38: 375–89. 2011.

GEORGE, M.; ANJUMOL, A.; GEORGE, G.; HATHA, A.A.M.; Distribution and bioactive potential of soil actinomycetes from different ecological habitats. **Afr J Microbiol Res** 6(10): 2265–2271. 2012.

GERLACH, W.; Taxonomic Classification of Metagenomic Sequences. **Tese de Doutorado**. Submetido na Faculdade de Tecnologia, Bielefeld University, Germany. 2012.

GHOSH, T.S.; HAQUE, M.M.; MANDE, S.S.; DiScRIBinATE: a rapid method for accurate taxonomic classification of metagenomic sequences. **BMC Bioinformatics**, 11 Suppl 7:S14, 2010.

GÓMEZ-MERINO, F.C.; GARCÍA-ALBARADO, J.C.; TREJO-TÉLLEZ, L.I.; VÁZQUEZ, A.P.; SILVA-ROJAS, H.V.; VELASCO-VELASCO J.; Genomic sciences, soil biodiversity and landscape: interactions for sustainability. **Revista Mexicana de Ciencias Agrícolas**. Pub. p. 1771-1780. 2014.

- GOODFELLOW, M.; FIEDLER, H.P.; A guide to successful bioprospecting: informed by actinobacterial systematics. **Antonie van Leeuwenhoek** 98: 119–142. 2010.
- GRAHAM, E.B.; WIEDER, W.R.; LEFF, J.W.; WEINTRAUB, S.R.; TOWNSEND, A.R. *et al.* Do we need to understand microbial communities to predict ecosystem function? A comparison of statistical models of nitrogen cycle processes. **Soil Biology Biochemistry**, 68: 279-282. 2014.
- HANDELSMAN, J.; Metagenomics: application of genomics to uncultured microorganisms. **Microbiology and Molecular Biology Reviews** 68: 669–685. 2004.
- HOWARD, E.C.; SUN, S.; REISCH, C.R.; DEL VALLE, D.A.; BÜRGMANN, H.; KIENE, R.P.; MORAN, M.R.; Changes in dimethylsulfoniopropionate demethylase gene assemblages in response to an induce phytoplankton bloom, **Appl. Environ. Microbiol.**, 77, 524e531. 2011.
- HOWE, A.C.; JANSSON, J.K.; MALFATTI, S.A.; TRINGE, S.G.; TIEDJE, J.M.; BROWN, C.T.; Tackling soil diversity with the assembly of large, complex metagenomes. **PNAS**. April 22, 2014.
- HU, H.W.; ZHANG, L.M.; YUAN, C.L.; HE, J.Z.; Contrasting Euryarchaeota communities between up land and paddy soils exhibited similar pH-imposed biogeographic patterns. **Soil Biology & Biochemistry**. 64. 18e27. 2013.
- HUGENHOLTZ, P.; PITULLE, C.; HERSHBERGER, K. L.; PACE, N. R.; Novel division level bacterial diversity in a Yellowstone hot spring. **Journal of Bacteriology**, v. 180, p. 366-76, 1998.
- HUSON, D.H.; AUCH, A.F.; QI, J.; SCHUSTER, S.C.; MEGAN analysis of metagenomic data. **Genome Res**, 17(3):377–386, Mar 2007.
- IBGE. (INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA). Manual técnico da vegetação brasileira. 2. ed. Brasília: IBGE, 2012.
- JANSSON, J.; Soil Metagenomics. **Encyclopedia of Metagenomics**. pp 600-609. 2015.
- KERSTERS, K.; DE VOS, P.; GILLIS, M.; SWINGS, J.; VANDAMME, P.; STACKEBRANDT, E. Introduction to the proteobacteria. In: DWARKIN, M.; FALKOW, S.; ROSENBERG, E.; SCHLEIFER, K.H.; STACKEBRANDT, E. (Ed.). The prokaryotes. 3rd ed. **New York: Springer**, v. 5, p. 3-37, 2006.
- KHODAKOVA, A.S.; SMITH, R.J.; BURGOYNE, L.; ABARNO, D.; LINACRE, A.; Random Whole Metagenomic Sequencing for Forensic Discrimination of Soils. **PLoS ONE** 9(8): e104996. doi:10.1371/journal.pone.0104996. 2014.
- KÖBERL, M.; MÜLLER, H.; RAMADAN, E.M.; BERG, G.; Desert Farming Benefits from Microbial potential in arid soils and promotes diversity and plant health. **PLoS One**. 6. 2012

KOZICH, J.J.; WESTCOTT, S.L.; BAXTER, N.T.; HIGHLANDER, S.K.; SCHLOSS, P.D.; Development of a Dual-Index Sequencing Strategy and Curation Pipeline for Analyzing Amplicon Sequence Data on the MiSeq Illumina Sequencing Platform. **Appl. Environ. Microbiol.** vol. 79 no. 175112-5120. 2013.

KUJUR, M.; GARTIA, S.K.; PATEL, A.K. Quantifying the contribution of different soil properties on enzyme activities in dry tropical ecosystems. **ARPN Journal of Agricultural and Biological Science**, v.7, p. 763-772. 2012.

LAMMLE, K.; ZIPPER, H.; BREUER, M.; HAUER, B.; BUTA, C.; BRUNNER, H. e RUPP, S. Identification of novel enzymes with different hydrolytic activities by metagenome expression cloning. **Journal of Biotechnology**. Vol. 127. p.575–592. 2007.

LEMA, K. A.; WILLIS, B. L.; AND BOURNE, D. G.; Amplicon pyrosequencing reveals spatial and temporal consistency in diazotroph assemblages of the Acroporamillepora, **Environ. Microbiol.**, 16, 3345e3359. 2014.

LEUNG, H.C.M.; YIU, S.; YANG, B.; PENG, Y.; WANG, Y.; LIU, Z.; CHEN, J.; QIN, J.; LI, R.; CHIN, F.Y.L.; A robust and accurate binning algorithm for metagenomic sequences with arbitrary species abundance ratio. **Bioinformatics**. **27**(11):1489-1495. 2011.

LOGARES, R.; SUNAGAWA, S.; SALAZAR, G.; CORNEJO-CASTILLO, F.M.; FERRERA, I.; SARMENTO, H.; HINGAMP, P.; OGATA, H.; DE VARGAS, C.; LIMA-MENDEZ, G.; RAES, J.; POULAIN, J.; JAILLON, O.; WINCKER, P.; KANDELS-LEWIS, S.; KARSENTI, E.; BORK, P.; ACINAS, S.G.; Metagenomic 16S rDNA Illumina tags are a powerful alternative to amplicon sequencing to explore diversity and structure of microbial communities. **Environmental Microbiology**, 16: 2659–2671. doi: 10.1111/1462-2920.12250. 2014.

MACKELPRANG, R.; WALDROP M.P.; DEANGELIS, K.M.; DAVID, M.M.; CHAVARRIA, K.L.; BLAZEWCZ, S.J.; RUBIN, E.M.; JANSSON, J.K.; Metagenomic analysis of a permafrost microbial community reveals a rapid response to thaw. **Nature**, 480:368–371. 2011.

MARON, P.A.; MOUGEL, C.; RANJARD, L. Soil microbial diversity: Methodological strategy, spatial overview and functional interest. **Comptes Rendus Biologies**, v. 334, p. 403-411, 2011.

MCHARDY, A.C.; MARTÍN, H.G.; TSIRIGOS, A.; HUGENHOLTZ, P.; RIGOUTSOS, I.; Accurate phylogenetic classification of variable-length DNA fragments. **Nature Methods** - **4**, 63 – 72. 2007.

MELNICHUK T.N.; ANDRONOV E.E.; KAMENEVA I.A.; ABDURASHYTOV S.F.; CHEKALINA U.V.; ALEKSEENKO N.V.; RADCHENKO L.A.; PARCHOMENKO T.YU.; YAKUBOVSKAYA A.I.; PASHTETSKYI V. S.; CHAIKOVSKYI V.A.; Changes Of Microbial Association Of Chernozem Southern In Different Farming Systems. **SWorld Journal**, Vol.J11509 (Scientific world, Ivanovo, 2015) – URL: <http://www.sworldjournal.com/e-journal/j11509.pdf>. Página 51. 2015.

MENDES, L.W.; TSAI, S.M.; NAVARRETE, A.A.; HOLLANDER, M.; VEEN, J.A.V.; KURAMAE, E.E.; Soil-Borne Microbiome: Linking Diversity to Function. **Microb Ecol.** DOI 10.1007/s00248-014-0559-2. 2015.

MERCANTE, F.M.; SILVA, R.F.; FRANCELINO, C.S.F.; CAVALHEIRO, J.C.T.; OTSUBO, A.A. Biomassa microbiana, em um Argissolo Vermelho, em diferentes coberturas vegetais, em área cultivada com mandioca. **Acta Scientiarum Agronomy**. Maringá, v. 34, n. 4, p. 479-485, 2008.

MEYER, F.; et al. The metagenomics rast server - a public resource for the automatic phylogenetic and functional analysis of metagenomes. **BMC Bioinformatics**, 9. 386. 2008.

MG-RAST. Disponível em: <http://blog.metagenomics.anl.gov/>

MIRANSARI, M.; Soil microbes and the availability of soil nutrients. **Acta Physiologiae Plantarum**, 35: 3075-3084, 2013.

MONTEIRO, M.; SENECA, J.; MAGALHAES, C.; The history of aerobic ammonia oxidizers: from the first discoveries to today. **J. Microbiol.** 52, 537–547. doi: 10.1007/s12275-014-4114-0. 2014.

NACKE, H.; FISCHER, C.; THÜRMER, A.; DANIEL, R.; Land Use Type Significantly Affects Microbial Gene Transcription in Soil. **Microb Ecol.** DOI 10.1007/s00248-014-0377-6

NANNIPIERI, P. et al.; Soil as a biological system. Caister Academic Press, Firenze, Italy. p 200. **Nature methods**. 2014.

NAVARRETE, A.A.; TSAI, S.M.; MENDES, L.W.; FAUST, K.; HOLLANDER, M.; CASSMAN, N.A.; VEEN, J.A.; KURAMAE, E.E.; Soil microbiome responses to the short-term effects of Amazonian deforestation. **Molecular Ecology**. 24, 2433-2448. 2015.

ORTIZ, M.; LEGATZKI, A.; NEILSON, J.W.; FRYSLIE, B.; NELSON, W.M.; WING, R.A.; SODERLUND, C.A.; PRYOR, B.M.; MAIER, R.M.; Making a living while starving in the dark: metagenomic insights into the energy dynamics of a carbonate cave. **The ISME Journal** 8, 478–491; doi:10.1038/ismej.2013.159. 2014.

PACCHIONI, R.G.; CARVALHO, F.M.; THOMPSON, C.E.; FAUSTINO, A.L.F.; NICOLINI, F.; PEREIRA, T.S.; SILVA, R.C.B.; CANTÃO, M.E.; GERBER, A.; VASCONCELOS, A.T.R.; AGNEZ-LIMA L.F.; Taxonomic and functional profiles of soil samples from Atlantic forest and Caatinga biomes in northeastern Brazil. **Microbiology Open**; 3(3): 299–315. 2014.

PARKS, D.H.; BEIKO, R.G.; Identifying biologically relevant differences between metagenomic communities. **Bioinformatics** 26: 715–721. 2010.

PENG, Y.; HENRY, C. M.; LEUNG, S.; YIU, M.; FRANCIS, Y.L.; IDBA-UD: a de novo assembler for single-cell and metagenomic sequencing data with highly uneven depth. **Oxford Journals**. 2012.

PENNINGTON, R.T.; PRADO, D.E.; PENDRY, C.A.; Neotropical seasonally dry forests and Quaternary vegetation changes. **Journal of Biogeography** 27:261-273. 2000.

PRAMANIK, A.; BASAK, P.; BANERJEE, S.; SENGUPTA, S.; CHATTOPADHYAY, D.; BHATTACHARYYA, M.; Pyrosequencing based profiling of the bacterial community in the Chilika Lake, the largest lagoon of India. **Genomics Data** 4 112–114. 2015.

PROCÓPIO, R. E. L.; SILVA, I. R.; MARTINS, M. K.; AZAVEDO, J. L.; ARAÚJO, J. M. “Antibiotics Produced by Streptomyces” **The Brazilian Journal of Infectious Diseases**, 16 (5): 466–471. 2012.

PROSSER, J.I. Dispersing misconceptions and identifying opportunities for the use of ‘omics’ in soil microbial ecology. **Nature Reviews Microbiology**. 2015.

PURKHOLD, U. et al. Biogeographic patterns in below-ground diversity in New York City's Central Park are similar to those observed globally. **Proc. R. Soc. B** 20141988; DOI: 10.1098/rspb.2014.1988. Published 1 October. 2014.

REAVY, B.; SWANSONA, M.M.; COCKA , P.; DAWSONB, L.; FREITAGB, T.E.; SINGHC, B.K.; TORRANCE, L.; MUSHEGIAN, A.R.; TALIANSKY, M.; Distinct circular ssDNA Viruses Exist in Different Soil Types. **Appl. Environ. Microbiol.** Doi:10.1128/AEM.03878-14. 2015.

RHO, M.; TANG, H.; YE, Y.; FragGeneScan: predicting genes in short and error-prone reads. **Nucleic Acids Research**, 2010.

RODRIGUES, J.L.M.; PELLIZARI, V.H.; MUELLER, R.; BAEK, K.; JESUS, E.D.; PAULA, F.S.; MIRZA, B.; HAMAOU, G.S.; TSAI, S.M.; FEIGL, B.; TIEDJE, J.M.; BOHANNAN, B.J.M.; NÜSSLEIN, K.; Conversion of the Amazon rainforest to agriculture results in biotic homogenization of soil bacterial communities. **Proc. Natl. Acad. Sci. U.S.A.** 110, 988–993. 2012.

SALIPANTE, S.J.; KAWASHIMA, T.; ROSENTHAL, C.; HOOGESTRAAT, D.R.; CUMMINGS, L.A.; SENGUPTA, D.J.; HARKINS, T.T.; COOKSON, B.T.; HOFFMAN N.G.; Performance Comparison of Illumina and Ion Torrent Next-Generation Sequencing Platforms for 16S rRNA-Based Bacterial Community Profiling. **Appl. Environ. Microbiol.** vol. 80 no. 247583-7591. 2014.

SCHLOSS, P. D.; HANDELSMAN, J. Biotechnological prospects from metagenomics. **Curr. Opin. Biotechnol.**, v. 14, p.303–310, 2003.

SCHMEISSER, C.; STEELE, H.; STREIT, W. R.; Metagenomics, biotechnology with nonculturable microbes. **Appl Microbiol Biotechnol** 75:955–962, 2007.

SCHNEIDER, D.; REIMER, A.; HAHLBROCK, A.; ARP, G.; DANIEL, R.; Metagenomic and Metatranscriptomic Analyses of Bacterial Communities Derived From a Calcifying Karst Water Creek Biofilm and Tufa. **Geomicrobiology Journal**, 32:3-4, 316-331, DOI: 10.1080/01490451.2014.907376. 2015.

SHARON I.; BANFIELD J.F.; Genomes from metagenomics. **Science** 342, 1057–1058. 10.1126/science.1247023. 2013.

SHARON, I.; MOROWITZ. M.J.; THOMAS, B.C.; COSTELLO, E.K.; RELMAN, D.A.; BANFIELD, J.F.; Time series community genomics analysis reveals rapid shifts in bacterial species, strains, and phage during infant gut colonization. **Genome Res**, 23:111–120. 2013.

SOUZA, R.C. et al. Archaea in biogeochemical cycles. **Applied Soil Ecology** 72. 49–61. 2013.

SOUZA, R.C.; HUNGRIA, M.; CANTÃO, M.E.; VASCONCELOS, A.T.R.; NOGUEIRA, M.A.; VICENTE, V.A.; Metagenomic analysis reveals microbial functional redundancies and specificities in a soil under different tillage and crop-management regimes. **Applied Soil Ecology** 86 106–112. 2014.

STREIT, W.R.; SCHMITZ, R.A.; Metagenomics-the key to the uncultured microbes. **Current Opinion Microbiology**., v. 7, n. 5, p. 492-498, 2004.

TAKETANI, R.G.; TSAI, S.M.; The influence of different land uses on the structure of archaeal communities in Amazonian anthrosols based on 16S rRNA and amoA genes. **Microb Ecol** 59(4):734–743. 2010.

TEELING, H.; MEYERDIERKS, A.; BAUER, M.; AMANN, R.; GLÖCKNER. F.O.; Application of tetranucleotide frequencies for the assignment of genomic fragments. **Environ Microbiol**, 6(9):938–947, Sep 2003.

TEELING, H.; WALDMANN, J.; LOMBARDOT, T.; BAUER, M.; GLÖCKNER F.O.; TETRA: a webservice and a stand-alone program for the analysis and comparison of tetranucleotide usage patterns in DNA sequences. **BMC Bioinformatics**, 5:163, Oct 2004.

TIAN, M., et al., The first metagenome of activated sludge from full-scale anaerobic/anoxic/oxic (A2O) nitrogen and phosphorus removal reactor using..., **J. Environ. Sci.**, <http://dx.doi.org/10.1016/j.jes.2014.12.027>. 2015.

TRIVEDI, P.; HE, Z.; VAN NOSTRAND, J.D.; ALBRIGO, G.; ZHOU. J.; WANG, N.; Huanglongbing alters the structure and functional diversity of microbial communities associated with citrus rhizosphere. **ISME J** 6(2):363–383. doi:10.1038/ismej.2011.100. 2012.

VASCONCELLOS, R. L. F.; SILVA, M. C. P.; RIBEIRO, C. M.; CARDOSO, E. J. B. N. “Isolation and Screening for Plant Growth-promoting (PG) Actinobacteria from Araucaria angustifolia Rhizosphere Soil.” **Scientia Agricola**, 67 (6): 743–746. 2010.

WAGG, C.; BENDER, S.F.; WIDMER, F.; VAN DER HEIJDEN, M.G.A.; Soil biodiversity and soil community composition determine ecosystem multifunctionality. **Proceedings of the National Academy of Sciences of the United States of America** 111, 5266-5270. 2014.

WANG, Y.; LEUNG, H.X.M.; YIU, S.M.; YUK, F.; CHIN, L.; MetaCluster-TA: taxonomic annotation for metagenomic data based on assembly-assisted binning. **BMC Genomics**, 15 (Suppl 1):S12. 2014.

- WERNER, J.J.; ZHOU, D., CAPORASO, J.G.; KNIGHT, R.; ANGENENT, L.T.; Comparison of Illumina paired-end and single-direction sequencing for microbial 16S rRNA gene amplicon surveys. **ISME J.** 6:1273–1276. 2012.
- WOODHOUSE, J.N.; FAN, L.; BROWN, M.V.; THOMAS, T.; NEILAN, B.A.; Deep sequencing of non-ribosomal peptide synthetases and polyketide synthases from the microbiomes of Australian marine sponges, **ISME J.**, 7, 1842e1851 (2013).
- ZHALNINA, K.; DIAS, R.; DE QUADROS, P.D.; DAVIS-RICHARDSON, A.; CAMARGO, F.A.O.; CLARK I.M.; MCGRATH, S.P.; HIRSCH, P.R.; TRIPLETT, E.W.; Soil pH Determines Microbial Diversity and Composition in the Park Grass Experiment. **Springer Science+Business Media**. New York. 2014.
- ZHANG, X.; QU, Y.; MA, Q.; ZHANG, Z.; LI, D.; WANG, J.; et al. Illumina MiSeq Sequencing Reveals Diverse Microbial Communities of Activated Sludge Systems Stimulated by Different Aromatics for Indigo Biosynthesis from Indole. **PLoS ONE** 10(4): e0125732. doi:10.1371/journal.pone.0125732. 2015.
- ZHENG, H.; BODINGTON, D.; ZHANG, C.; MIYANAGA, K.; TANJI, Y.; HONGO, Y.; XING, X.H.; Comprehensive phylogenetic diversity of [Fe-Fe]-hydrogenase genes in termite gut microbiota, **Microbes Environ.**, 28, 491e494. 2013.
- ZHU, M.Z.Z.; HIRDES J.P.; STOLEE P: Using machine learning algorithms to guide rehabilitation planning for home care clients. **BMC Medical Informatics and Decision Making**. 2007.