UNIVERISDADE FEDERAL DA GRANDE DOURADOS

MARIA PRISCILA FRANCO LACERDA

ANÁLISE METAGENÔMICA DA COMUNIDADE BACTERIANA EM SOLOS ANTROPIZADOS

Dourados

2012

UNIVERSIDADE FEDERAL DA GRANDE DOURADOS

Maria Priscila Franco Lacerda

Análise metagenômica da comunidade bacteriana em solos antropizados

Trabalho apresentado ao Curso de Biotecnologia da Universidade Federal da Grande Dourados como parte dos requisitos para obtenção do título de Bacharel em Biotecnologia.

Orientador: Prof. Dr. Rodrigo Matheus Pereira

Dourados

2012

MARIA PRISCILA FRANCO LACERDA

ANÁLISE DA COMUNIDADE BACTERIANA EM SOLOS ANTROPIZADOS

Trabalho apresentado ao Curso de Biotecnologia da Universidade Federal da Grande Dourados como parte dos requisitos para obtenção do título de Bacharel em Biotecnologia.

Aprovado em ____ de _____ de 2012.

BANCA EXAMINADORA

_____

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

_____

Prof. Dr. Marcelo Fossa da Paz
Universidade Federal da Grande Dourados

_____

Prof. Dr. Alessandro Minillo
Universidade Estadual de Mato Grosso do Sul

Dedico esse trabalho a Deus, meu Pai e Criador, Onipotente e presente em todos os dias de minha vida. Aos meus pais, José Hilton Lacerda, e Maria Aparecida Franco Lacerda pela oportunidade de ingressar na Universidade. Ao Professor e Orientador Rodrigo Matheus Pereira pela confiança e pela sua amizade.

AGRADECIMENTOS

Primeiramente a Deus pelas bênçãos dadas e pelas graças alcançadas nesta fase que se encerra.

Em especial ao meu pai José Hilton Lacerda, à minha mãe Maria Aparecida Lacerda e à minha irmã Tatiane Lacerda pelo amor incondicional, incentivo, atenção e paciência dedicados em todos os momentos. Apesar da distância, vocês estão em meus pensamentos todos os dias.

Às minhas amigas Marianna, Jéssica, Bruna, Elaine e Flávia, por serem pessoas tão especiais em minha vida, e que sempre me ajudaram e me apoiaram nos momentos em que mais precisei.

Ao meu namorado pelo amor, carinho e paciência durante essa etapa das nossas vidas.

Ao Professor e Orientador Rodrigo Matheus Pereira pela oportunidade de realizar esse trabalho, pela orientação e amizade.

A todas as pessoas que de alguma forma contribuíram para minha formação acadêmica, o meu muito obrigada.

RESUMO

A metagenômica abriu novos caminhos para a microbiologia. Hoje é crescente o número de metodologias propostas para estudos sobre os microrganismos. Esses organismos possuem grande importância em relação aos ciclos biogeoquímicos do solo. No presente trabalho foram analisadas 2.561 sequências de bactérias usando o gene 16S rRNA, obtidas em trabalhos já publicados através do método de Sanger, em 7 diferentes solos brasileiros que sofreram intensa ação antrópica. Os seguintes códigos foram usados: cultivo de cana – SCC; consórcio de óleo diesel – SCOL; solo cultivado da região da Amazônia – SCA; solo de pasto da região da Amazônia – SPA; solo com cultivo de hortaliças – SCH; solo contaminado com lodo de esgoto – SCLE e solo que foi cultivado por grãos (soja, milho, feijão) - SCG. O objetivo deste trabalho foi fornecer uma visão geral das comunidades bacterianas presentes nestes ambientes. Todas as sequências foram submetidas à ferramenta Web MG-RAST. Os dados foram comparados com o banco de dados do RDP (Ribosomal Database Project) usando um valor máximo de $1e^{-5}$ para e-value, uma identidade mínima de 60%, e um comprimento mínimo de alinhamento de 20 nucleotídeos. Os resultados demonstraram a distribuição de filos do domínio *Bacteria*, o mais representativo foi o filo *Proteobacteria* - cuja presença foi constatada nas sete amostras analisadas. Foi relatada a presença de 13 filos entre as amostras, sendo que em quatro destes houve a presença em comum em seis solos. Os filos são *Bacteroidetes, Firmicutes, Acidobacteria e Actinobacteria*. O presente trabalho permitiu a identificação dos filos *Spirochaetes, Tenericutes* e *Cyanobacteria*, que em trabalhos anteriores não haviam sido identificados.

Palavras-chave: Biorremedição, Bioinformática, 16S rRNA.

LISTA DE ILUSTRAÇÕES

Figura 1. Comparação da distribuição de domínio entre os metagenomas.....27

Figura 2. Comparação de Distribuição de Filo (Bactéria) entre os metagenomas.....28

Figura 3. Distribuição do Filo Proteobacteria.....29

Figura 4. Distribuição da Ordem (Alphaproteobacteria).....30

Figura 5. Distribuição de Ordem (*Rhizobiales*).....31

Figura 6. Distribuição de gênero (*Bradyrhizobiaceae*).....32

Figura 7. Distribuição de espécies (Bradyrhizobium).....33

Figura 8. Distribuição de classes de Firmicutes.....34

Figura 9. Distribuição de classes de Bacteroidetes.....35

Figura 10. Distribuição de classes de *Acidobacterias*.....36

Figura 11. Distribuição de Ordem de *Actinobacteria*.....37

Figura 12. Distribuição de Classes de *Verrucomicrobia*.....38

Figura 13. Distribuição de Classes de *Chloroflexi*.....39

Figura 14. Distribuição de Classes de *Planctomycetes*.....40

Figura 15. Distribuição de Filo (não distribuído).....42

Figura 16. Curvas de rarefação para a riqueza estimada de espécies dos sete solos comparados 45

LISTA DE TABELAS

Tabela 1. Valores de diversidade de espécies de cada metagenoma.....45

SUMÁRIO

1. INTRODUÇÃO 8
1.1 Objetivos Específicos.....10
2. REVISÃO BIBLIOGRÁFICA **11**
2.1. Diversidade biológica dos microrganismos 11
2.2. Metagenoma 13
2.2.1. Panorama Histórico 14
2.2.2.Aplicação da abordagem metagenômica 15
2.3.Contribuição da Bioinformática.....16
2.4. Ferramentas WEB de abordagem metagenômica 18
3. MATERIAL E MÉTODOS **21**
3.1. Sequências de solos impactados 21
3.1.1. Origem das sequências SCG 21
3.1.2. Origem das sequências SCLE **22**
3.1.3. Origem das sequências SCH 22
3.1.4. Origem das sequências SCA e SPA **23**
3.1.5. Origem das sequências SCOL 24
3.1.6. Origem das sequências SCC 24
3.2. Ferramenta Web MG-RAST.....25
3.2.1. Rarefação 25
4. RESULTADOS E DISCUSSÃO **26**
4.1. Distribuição de Domínio 26

4.2.Distribuição de Domínio (Bactéria) 27

4.2.1. Filo *Proteobacteria* 28

4.2.2. Filo *Firmicutes* 33

4.2.3. Filo *Bacteroidetes* 34

4.2.4. Filo *Acidobacteria* 35

4.2.5. Filo *Actinobacteria* 36

4.2.6. Filo *Verrucomicrobia* 37

4.2.7. Filo *Chloroflexi* 38

4.2.8. Filo *Planctomycetes* 39

4.2.9. Filo *Nitrospirae* 40

4.2.10. Filo *Dictyoglomi* 40

4.2.11. Filo *Spirochaetes* 40

4.2.12. Filo *Cyanobacteria* 41

4.2.13. Filo *Tenericutes* 41

4.3. Distribuição de Filo (não distribuído) 42

4.4. Rarefação 43

5. CONCLUSÃO 44

6. REFERÊNCIAS BIBLIOGRÁFICAS 45

INTRODUÇÃO

Hoje é crescente o número de metodologias propostas para estudos sobre os microrganismos do solo e sua ecologia, justificada pela geração de produtos derivados de seus metabolitos secundários que tem sido utilizado para atender às necessidades médicas, industriais e agrícolas (RONDON, 1999), sendo cada vez mais valorizada sua biodiversidade na regulação dos ecossistemas (COPLEY, 2000). A maioria dos metabólitos secundários microbianos em uso hoje se originam dos actinomicetos, entre outros produtores de produtos naturais como *Bacillus* spp., mixobactérias e *Pseudomonas* (LANCINI, 1995; REICHENBACH, 1999).

Pesquisas atuais indicam que mais de 99 % dos microrganismos presentes no ambiente não podem ser facilmente cultivados. Assim, a maioria dos micróbios não tem sido descritos ou estão acessíveis para a biotecnologia. Várias tecnologias baseadas em DNA têm sido desenvolvidas para ultrapassar estas dificuldades. No entanto, apenas uma parte de um gene pode ser amplificada por PCR, e os dados da sequência proporcionam alguns detalhes sobre a informação genética e qualquer papel funcional na comunidade microbiana. Além disso, a amplificação por PCR requer que informações sobre a sequência estejam disponíveis antes do desenho para os iniciadores (KIMURA, 2006).

Os microrganismos possuem grande importância em relação aos atributos físicos do solo, pois os organismos heterotróficos pertencentes ao solo liberam polissacarídeos de origem microbiana, durante a decomposição da matéria orgânica, que são mais persistentes no solo, quando comparados aos polissacarídeos de origem vegetal, assim desempenham fundamental papel na estabilidade de agregados. Tais seres vivos além da geração de novos produtos têm papel-chave no controle biológico de doenças da agricultura e na degradação de compostos sintéticos do solo, entre outros fatores determinantes para a sobrevivência da planta. Assim sendo considerado um vasto campo ainda não explorado para descoberta de novos compostos farmacêuticos e produtos enzimáticos de valor biotecnológico (LOURENTE et al., 2011).

Segundo Gould (1996), a bactéria é o organismo mais abundante e deveria ser o representante da vida em nosso planeta, pois ela reflete a importância da comunidade bacteriana nos diferentes ecossistemas da Terra. Com o progresso de estudos em

metagenômica, é notória a influência dos microrganismos no ambiente e no desenvolvimento de outros seres ligados a eles.

Jo Handelsman foi um dos primeiros pesquisadores a usar o termo “Metagenômica”, e que hoje tem como definição a aplicação de técnicas de genômica moderna para o estudo de comunidades de microrganismos diretamente em seus ambientes naturais, do qual tenham sido extraídas amostras de DNA ou RNA, sem a necessidade de isolamento e cultivo em laboratório das espécies individuais. Um dos pioneiros a trabalhar com pesquisas em metagenoma foi Torsvik, em que um de seus trabalhos foi o estudo da diversidade fenotípica, de cerca de 200 cepas bacterianas isoladas a partir de um solo e foi comparado com o diversidade genotípica de uma mesma população. A metagenômica já abriu novos caminhos de pesquisa, permitindo sem precedentes a análise da heterogeneidade do genoma e evolução em contextos ambientais, e o acesso à diversidade microbiana (HANDELSMAN, 2004).

Segundo Cole et al. (2009) quando amplificadas regiões de um gene comum a todas as bactérias presentes numa dada amostra, o gene em questão é o 16S rDNA. Este gene tem como função ser um marcador de grandes grupos bacterianos, permitindo assim sua classificação – como, por exemplo, a comparação de seqüências de bancos de dados de DNA ribossomal – dos microrganismos presentes no dado amostral. Assim, é possível verificar exatamente quais grupos bacterianos estão presentes em uma amostra de solo, podendo determinar a relevância da presença microbiana com a característica de seu bioma.

No presente trabalho foram analisadas 1.422 sequências de bactérias usando o gene 16S rRNA de 7 diferentes ambientes brasileiros que sofreram intensa ação antrópica. O objetivo deste trabalho foi fornecer uma visão geral das comunidades bacterianas presentes nestes ambientes, além de fornecer elementos a futuros estudos relacionados à biorremediação.

1.1.Objetivos específicos

- Observar se algum tipo grupo bacteriano apresenta-se em comum entre todos os diferentes solos.

- Avaliar se o processo de reanotação de sequências realizado permitiu a identificação de novos filos.

- Avaliar o número de filos que possam estar presentes nas amostras.

2. REVISÃO BIBLIOGRÁFICA

2.1. Diversidade biológica dos microrganismos

O solo forma um ecossistema é constituído de rochas e minerais, sendo habitado por inúmeras espécies. Os microrganismos são parte integrante do solo, desempenhando papel fundamental nas reações bioquímicas, na formação do solo, e na regulação de nutrientes para o crescimento de vegetais. Também podem ser chamados coletivamente de microbiota, sendo representados por quatro grandes grupos: bactérias, fungos, algas e protozoários. As bactérias e fungos correspondem em torno de 90% da atividade microbiana do solo (SILVEIRA *et al*., 2007).

A complexidade e a variabilidade descrevem a diversidade biológica em sua organização, englobando a variabilidade genética dentro de táxons e grupos funcionais nas comunidades. A gama de processos, a complexidade de interações e a quantidade de níveis tróficos são aspectos importantes da diversidade no ecossistema. Diversidade pode ser considerada a distribuição de informação, que é aplicável a diversidade genética total ou complexa de uma comunidade (TORSVIK & OVREAS, 2002).

Segundo Torsvik et al.(1990) em um grama de solo existem 4.000 unidades de genomas bacterianos com base na reassociação DNA-DNA, porém Rosseló-Mora & Amann (2001) em sua revisão estimam que um grama de solo contem cerca de 10 bilhões de procariotos pertencentes a milhares de diferentes espécies.

Em relação ao tamanho e a diversidade de espécies da comunidade microbiana, o solo é provavelmente o maior desafio de ambientes naturais para os microbiologistas. Um grama de solo de floresta pode vir a conter aproximadamente $4x10^7$ células procarióticas, enquanto um grama de solos cultivados e pastos contêm cerca de $2x10^9$ células procarióticas. A partir de DNA isolado de amostras de diferentes solos, o número de genomas procariontes encontra-se entre 2.000 a 18.000 por grama de solo (DANIEL, 2005).

Esses dados demonstram a rica diversidade genética e biológica que ainda podemos explorar na busca de novos genes e produtos como de vias metabólicas (COWAN, 2000). Os microrganismos incultiváveis, provavelmente realizam alguma função vital

para o solo, podem assim Serem considerados uma enorme fonte de aplicação biotecnológica (HANDELSMAN *et al.,* 2004).

A distribuição física de bactérias em níveis micro habitacionais demonstra que, em solos tratados para fertilização, mais de 80% das bactérias estavam localizadas em microporos do solo, já que o tamanho das partículas tem maior impacto na comunidade microbiana que fatores como pH e composto orgânico (2-20 mm) (TORSVIK & OVREAS, 2002). Esses micro habitats oferecem condições mais favoráveis quanto à água e disponibilidade de alimentos.

Há muitas pesquisas que indicam que o tipo e a quantidade de alimento orgânico disponível influenciam a disponibilidade de grupos microbianos e sua diversidade funcional no solo (DE FEDE, 2001). A diversidade de espécies consiste na integração de componentes representados por sua riqueza, equitabilidade e distribuição. Dois grupos de classificação são usados para medir a diversidade microbiana – técnicas bioquímicas e técnicas de base molecular. Atualmente é difícil o estudo da diversidade, pois não temos nenhuma maneira de determinar a precisão dos nossos métodos de extração e detecção. Muitos pesquisadores tentam reduzir a informação recolhida por estudos de descrição da diversidade (KIRKA *et al.,* 2004).

Durante muitos anos, os estudos da microbiota basearam-se no isolamento e cultivo de laboratório de bactérias. Este cultivo ainda permanece de grande interesse quando alvos são isolados pela sua capacidade de metabolizar substratos específicos ou degradação de xenobióticos utilizados em biorremediação. A análise molecular da diversidade bacteriana teve como base a reassociação DNA/DNA, e a amplificação de 16S rDNA. No entanto, estes métodos precisam de DNA livres de contaminantes de suas próprias amostras ambientais (ROBE *et al.,* 2003).

2.2. Metagenoma

A detecção e a identificação de microrganismos eram feitas principalmente de acordo com suas exigências nutricionais, com observação direta via microscópio (KENNEDY *et al.,* 1999). Porém, com o uso dessas metodologias tornavam-se limitadas as informações obtidas (ZAK *et. al.,* 1994).

Quando se quer pesquisar a diversidade da microbiota num ambiente, ainda encontram-se limitações nas técnicas tradicionais - como os meios de cultivo

bacterianos, que são seletivos a grupos específicos da espécie. Segundo Amann *et. al*. (1995), 1-10% das espécies bacterianas no mundo já foram identificadas, abrindo caminho para novos estudos em uma área ainda pouco explorada.

A biodiversidade microbiana é a maior guardiã da diversidade de genes. Com o desejo de superação dos limites impostos pelas metodologias convencionais de detecção (VAN ELSAS *et al,* 1998), foi incentivado o desenvolvimento tecnológico na biologia molecular, aliando métodos baseados em ácidos nucléicos e estudo da diversidade microbiana.

Deste modo nasce a metagenômica, uma ciência também chamada de genômica ambiental, genômica de comunidade, ou ainda análise baseada no genoma de comunidades inteiras com grande interação dos organismos em um mesmo contexto. O termo "metagenoma" teve sua primeira definição como o genoma coletivo da microbiota total de um determinado ambiente por Jo Handelsman *et. al.* (1998) (WOOLEY *et al.,* 2010).

Esta abordagem foi originalmente desenvolvida como uma ferramenta para análises de genomas microbianos contida em amostras ambientais (RIESENFELD, 2004; SCHLOSS et al., 2003). Seus primeiros estudos renderam a identificação de características metabólicas, a caracterização dos organismos e a descoberta de novos antibióticos e enzimas (RIESENFELD *et al.,* 2004).

Estudos metagenômicos agora abrangem campos de investigação, incluindo a pesquisa ambiental marinha, vegetal e biotecnologia agrícola, genético humano e diagnóstico de doenças humanas. O domínio das aplicações na metagenômica deve aumentar, passando de estudos de microbioma ambientais, para diagnósticos clínicos (MOKILI *et al.,* 2012).

Muitas pesquisas estao sendo realizadas atribuindo grande importância à metagenômica, Bhatt *et al.* (2012) também desenvolveram um estudo no qual amostras de leite coletadas de duas raças bovinas - uma delas portadoras de mastite subclinica, foram submetidas a análises metagenômicas para elucidar a comunidade microbiana e sua estrutura no leite do gado. Tendo seus perfis metabólicos e filogenéticos gerados pelo MG-RAST, uma ferramenta baseada na web.

2.2.1. Panorama Histórico

A microbiologia tradicionalmente é uma disciplina marginal para a biologia de microrganismos, por seus estudos de interesse ter como foco a diversidade biológica, utilizando a análise genética e molecular nos microrganismos. A utilização de microrganismos, como ferramentas para compreensão da hereditariedade genética, teve como fundamento a convicção e demonstração de que os microrganismos possuem o mesmo material genético e processos biológicos muito semelhantes. Com o estudo de genoma completo feito na década de 1990, os microrganismos tiveram atenção central em estudos aliados a biologia molecular, como o seqüenciamento de organismos como vírus e procariotos (DUPRÉ *et al.,* 2007).

Ao fornecer a capacidade de examinar a relação entre a estrutura do genoma e a funcionalidade de muitas espécies diferentes, os dados de seqüenciamentos já feitos também abriram os campos de genômica comparativa.

Assim, muitas limitações surgiram sobre a funcionalidade de culturas microbianas feitas através de cultivo, devido ao crescimento de informações geradas com os estudos genômicos (WARD & FRASER, 2005). A microbiologia molecular surgiu na tentativa de sanar esses problemas.

A abordagem microbiana ambiental, denominada genomica ambiental ou metagenômica, se estendeu além do seqüenciamento de culturas de microrganismos isolados, para o seqüenciamento de DNA extraído de seus ambientes naturais (AMANN *et al*, 1995).

As seqüências de DNA (proteínas codificantes ou não) foram utilizadas como marcadores e indicadores de espécies. Com o foco em determinados genes foi possível a obtenção de informações sobre características fisiológicas e ecológicas dos organismos (WARD *et al.,* 2006).

A metagenômica foi concebida como uma abordagem das quais muitas limitações podem ser superadas. Pois, em vez de genomas individuais, a metagenômica emprega quantidades generosas de DNA recolhidos de comunidades microbianas em seus ambientes naturais com a finalidade de explorar a diversidade da amostra, interações funcionais e relações evolutivas (DUPRÉ *et al.,* 2007).

2.2.2. Aplicação da abordagem metagenômica

Os microrganismos em seus ambientes de cultivo podem conter genes que codificam ou expressam algum fator biossintetico ou biodegradativo de interesse ainda não estudados independente de métodos de cultura. A vantagem de uma abordagem metagenômica, é permitir pesquisas em filogenia e diversidade funcional dos organismos em uma comunidade (SCHLOSS & HANDELSMAN, 2003; COWAN *et al.*, 2005).

Por muitas décadas, enzimas expressas por microrganismos de solo utilizando meio de cultura, foram colhidas e comercializadas. As bibliotecas de metagenômica ambiental levaram à descoberta de novas enzimas com grandes aplicações industriais. Um resultado positivo disto foi um dos primeiros estudos metagenomicos que envolveu a identificação de celulases de um biorreator "zoolibrary" (HEALY *et al.*, 1995). Há outras enzimas já descobertas pela abordagem metagenômica, como a hidrolase glicosilo multifuncional que foi identificada a partir de uma biblioteca de rúmen.

Outra área de aplicação metagenômica que vem crescendo significativamente é a biorremediação. Técnicas como análises de PCR (Polymerase Chain Reaction) e hibridização do DNA total da comunidade do solo, geram resultados que permitem determinar com maior exatidão qual o melhor processo de biorremediação a ser usado, com o objetivo de utilizar o máximo oferecido de certo grupo de microorganismos de uma comunidade microbiana (CUNHA *et al.*, 2008).

No Brasil, a PETROBRAS/ UNICAMP/ FUNCAMP e o CPQBA, em 2006, deram início à um projeto na linha de metagenômica que tinha por objetivo a prospecção da atividade de degradação de hidrocarbonetos aromáticos em metagenoma microbiano originado de efluente de refinaria de petróleo (BRITO *et al.*, , 2010). Assim como Carvalho *et al.* (2007) que realizaram um estudo em que visavam a detecção de microrganismos redutores de sulfato em amostras ambientais contaminadas por petróleo.

Novas vias catabólicas para a degradação de compostos xenobióticos podem ser isolados com a ajuda da aplicação metagenômica, que pode vir a indicar uma capacidade funcional genética para a degradação de contaminantes. Substratos rotulados

já têm sido utilizados para recuperação de genes de populações envolvidos em processo de degradação (KAVITA *et al.*, 2010).

Outros estudos metagenômicos já identificaram vias catabólicas capazes de codificar nitrilases, que desempenham um papel importante em reações biossintéticas e catabólicas (ROBERTSON *et al.*, 2004), bem como enzimas com propriedades catalíticas que degradam contaminantes orgânicos(KIM *et al.*, 2007).

2.3.Contribuição da Bioinformática

Se tornando quase uma palavra coloquial na comunidade científica, a bioinformática, pode ter sua interpretação muito variada. A palavra tem dois componentes, “bio” e “informática”. A interpretação aceita da parte “bio” é a biologia molecular estudando macromoléculas essenciais para a vida. No entanto, as áreas da biologia consideradas de competência da bioinformática, são o estudo da evolução e da genética. Informática é uma palavra que seu significado coincide com “ciência da computação”. Ao levar em consideração as partes da palavra, pondera- se a resolução de problemas decorrentes da biologia utilizando a metodologia da ciência da computação, aplicando matemática e estatística. O grande objetivo da pesquisa em bioinformática é entender o funcionamento dos organismos vivos a fim de "melhorar a qualidade de vida". Esta melhoria, seria através da identificação de drogas por desenho, de fatores de risco genéticos, terapia genética, modificação genética, entre outras. (BACKOFEN & GILBERT, 2006).

O século 20 foi marcado por um dos grandes desenvolvimentos da área cientifica, o Projeto Genoma Humano que sequenciou o DNA humano completo (WATSON *et al.*, 1990) . Atualmente a principal atividade é a pós-genômica, sendo caracterizada pela disponibilidade de quantidades grandiosas de dados que podem ser usados na Bioinformática, mas que necessita de uma matemática nova e com métodos algorítmicos capazes de extrair qualquer informação contida nesses dados. Muitos modelos computacionais naturais têm sido aplicados na solução de problemas relacionados à visualização de dados, classificação e “clustering”, assim tendo como objetivo mesclar as técnicas para obter diferentes tipos de informação (MASULLI *et al.*, 2009).

A publicação de uma sequência de DNA não fornece informação imediata. Tem-se uma lacuna entre a disponibilidade de dados de sequências novas e uma compreensão científica destas informações. Esta lacuna pode ser compreendida através da utilização da Bioinformática, que atualmente aborda problemas de modelagem estatística, técnicas de aprendizagem de máquina tais como redes neurais artificiais e vetor de suporte de máquina. A complexidade e a quantidade de dados a serem analisados, exige uma estreita colaboração entre biólogos e cientistas da computação. Maior é o valor adicionado a esta relação, se os colaboradores possuem pelo menos um ponto elementar do conhecimento do outro campo (CORNE & FOGEL, 2004).

A metagenômica tem o potencial para o esclarecimento de controles genéticos de uma comunidade microbiana por meio de investigação dos seus principais agentes, seus papéis, e composições na comunidade que podem alterar com a função do clima, o tempo, nutrientes, dióxido de carbono e fatores antrópicos. Estes estudos incluem componentes subtanciais da informática, assumindo um papel computacional complexo e desafios matemáticos. Os microbiologistas tem aproveitado esta técnica moderna, gerando milhares de dados de sequencias, à fim de sua interpretação, ao invés de somente recolhimento de dados (SESHADRI *et al.*, 2007).

Há um número de diferentes sistemas e recursos para análise metagenômica, que são oferecidos na forma de bases de dados, portais web, serviços web, sendo programas básicos, porém muito eficientes (KRAUSE *et al*, 2008; MEYER *et al*, 2008; SESHADRI *et al*, 2007). Esses recursos são, principalmente, centrados na análise de metagenomas individuais. Segundo Mitra *et al*. (2009), atualmente só o MG-RAST, um servidor web (MEYER *et al*, 2008), fornece um serviço prontamente útil para a análise de um conjunto de dados metagenômico novo.

2.4. Ferramentas WEB de abordagem metagenômica

A Web tornou-se uma das fontes mais importantes de informação, a abertura de acesso para uma vasta gama de recursos de pesquisa em todo o mundo. Infelizmente, essa visão ainda não reflete na Bioinformática, pois na situação atual cada instituição implanta um software usando seu próprio acesso – assim não sendo de fácil acesso a usuários externos. A solução mais promissora para este e outros problemas é baseada

na noção de serviços web (WS). WS são aplicações implantadas através da internet que fornecem uma funcionalidade dada a outras aplicações. No contexto da bioinformática, serviços da Web prometem facilitar a integração e interabilidade entre outras aplicações da bioinformática.(FOGEL, 2003)

A metagenômica é um ramo relativamente novo da ciência, e grande parte da pesquisa atual é exploratória. A visualização é um aspecto muito importante do campo. A destilação de dados metagenômicos em representações gráficas, no entanto não é uma tarefa trivial. A base de estudos metagenômicos é a atribuição de valores observados, para taxonomia ou hierarquias funcionais. Os resultados podem ser representados em diagramas para transmitir hierarquia, os gráficos de barras ou torta podem relacionar a abundância de níveis específicos, mas nenhum destes métodos por si só cria uma ilustração completa da análise classificatória (ONDOV *et al.*, 2011)

Um grande número de programas vem sendo desenvolvido atualmente para análise de dados metagenômicos. Muitos já são usados, como Integrated Microbial Genomes (IMG) (MARKOWITZ *et al.*, 2008), que é um sistema de gerenciamento de dados e análise para metagenomas pertencente ao Departamento de Energia (DOE) do Joint Genome Institute (JGI); o Pyrotagger (KUNIN *et al*, 2010), é composto por uma técnica de filtragem de qualidade de leitura, capaz de processar de centenas à milhares de sequencias de RNA em poucos minutos; o Classifier for Metagenomic Sequences (ClaMS) (PATI *et al*, 2011), é uma aplicação Java para contigs reunidos em metagenomas usando conjuntos de treinamento especificados pelo usuário e por parâmetros iniciais; o FAMeS (MAVROMATIS *et al*, 2008), hospeda dados provenientes de um estudo abrangente das metodologias utilizadas para criar OTUs de 16S rRNA, alvo de estudos de comunidades microbianas; Gold (LIOLIOS *et al*, 2006), é uma ferramenta web para o acesso à informação abrangente sobre o seqüenciamento de genoma completo e outros projetos em todo o mundo, proporciona informação relacionada com propriedades do organismo tais como ecotipo, fenótipo, e da doença em questão; o Camera (SESHADRI *et al*, 2007), é um repositório de dados e ferramentas de bioinformática para metagenômica, onde os resultados de pesquisas são pré-computados para apoiar a análise cruzada de amostras ambientais; o webMGA (WU *et al.*, 2011) é um servidor web personalizável para análise metagenômica rápida, que inclui serviços como clusterização de seqüência, controle de qualidade de matérias-

primas de leituras, a remoção contaminações, análise taxonômica, funcional anotação, entre outros.

O MG-RAST (MEYER *et al,* 2008), tem sido um dos programas web metagenômicos mais usados, fornecendo um novo paradigma para anotação e análise de metagenomas. Com suporte integrado para várias fontes de dados, é estável, extensível e livremente disponível para todos os pesquisadores. O MG-RAST foi usado para criar classificações taxonômicas e funcionais de sequencias a partir da leitura da amostra (ONDOV *et al.,* 2011). Uma de suas vantagens é ser uma ferramenta web, na qual não é preciso ser instalado na máquina, sendo de livre acesso em qualquer local e máquina. Outras vantagens são que ele realiza análises de classificação possibilitando comparações através de 15 bancos diferentes, além de realizar análises estatísticas como PCoA, rarefação e até mesmo análises de heatmap.

Realizando uma comparação entre três dos citados programas metagenômicos, o MEGAN, é um programa que roda somente local sendo distribuído gratuitamente. Foi desenvolvido para análises metagenômicas, e resolve muitos problemas por aumentar os diagramas em escalas maiores geradas em gráficos. Este mesmo tipo de visor é também utilizado pela plataforma metagenômica da base Web MG-RAST (MEYER *et al,* 2008). Outra plataforma baseada na web é o METAREP (ONDOV *et al.,* 2011), que caracteriza mapas de interesse naturalmente escalados em abundância.

O ponto comum de MEGAN, MG-RAST, e METAREP é que eles facilitam a comparação direta de vários conjuntos de dados em um só lugar, como metagenomas amostrados de diferentes regiões ou em condições diferentes. Algoritmos de classificação metagenômica são constantemente melhorados, mas seus resultados ainda vêm com um significativo grau de incerteza. Para que as ferramentas de análise metagenômica continuem a ser introduzidas e refinadas, os programas web estão sendo projetados para serem independentes e suficientemente flexíveis para poderem ser adaptados à novas medidas.(FOGEL, 2003).

Independentemente da abordagem de sequenciamento utilizada para gerar os dados colocados para análise, os primeiros passos na análise de qualquer metagenoma envolve a comparação dessas sequencias com sequencias conhecidas em bases de dados. Esta tarefa computacionalmente intensiva fornece os tipos básicos de dados para muitas análises subsequentes, incluindo comparações filogenéticas, anotações funcionais,

classificação de seqüências, perfis filogenômicos e reconstruções metabólicas (MEYER *et al*, 2008).

3. MATERIAL E MÉTODOS

Foram utilizadas 1.422 sequências parciais do gene 16S rRNA. Todas as sequências foram extraídas originalmente de sete amostras de solos sob diferentes tratamentos como cultivo intensivo, pastagem e contaminados com óleo, em diferentes regiões do território brasileiro. Todos os solos analisados apresentaram em comum o fato de terem sofrido algum tipo de intervenção antrópica. As amostras utilizadas foram inicialmente trabalhadas em diferentes pesquisas, realizadas por Val-Moraes *et al*. em 2009(135), Pedrinho *et al*. em 2009(76), Pereira *et al*. em 2006(85), Jesus *et al.* em 2009(79 crops)(80 pasture), Cury *et al.* em 2011(1139), Paixão *et al.* em 2010(431) e Pisa *et al.*, em 2011(536)

3.1. Sequências de solos impactados

3.1.1. Origem das sequências SCG

Foram originalmente analisadas no trabalho "MOLECULAR CHARACTERIZATION OF BACTERIAL POPULATIONS OF DIFFERENT SOILS", Pereira *et al*. em 2006.

As amostras de solo foram coletadas na Fazenda Lagoa do Fogão, localizada na cidade de Guaíra, no norte do Estado de São Paulo. Suas coordenadas geográficas são latitude 20 º 20 'S e longitude 48 º 23 'W, com uma altitude de 500 metros. A média anual de precipitação é 1,330 milímetros e as temperaturas médias podem variar em mínima de 17 º C a 30 º C máxima. A área cultivada está em uso há mais de dez anos. O sistema aplicado é o cultivo direto com as culturas sucessivas de tomates ou feijão no inverno, de maio a junho, seguido pelo milho, a partir de agosto a setembro, e soja em dezembro. A agricultura é bastante intensiva e faz uso de todos os recursos possíveis (fertilizantes, pesticidas e irrigação) para a produção durante o ano inteiro. O solo foi classificado como "latossolo vermelho". As suas características físicas e químicas são apresentadas na Tabela 1. O DNA metagenômico foi utilizado em reacções de PCR para a amplificação do gene 16S rRNA, usando primers específicos. A diversidade de nucleotídeos foi estimada a partir do número de posições variáveis para as sequências alinhadas em uma determinada biblioteca (Pereira *et al.*, 2006). As sequências obtidas

das analises metagenômicas do solo podem ser acessadas pelo Código GenBank |AY734433.1| a |AY734310.1|, sendo em 85 sequências.

3.1.2. Origem das sequências SCLE

Foram originalmente analisadas no trabalho "AVALIAÇÃO DO IMPACTO DO LODO DE ESGOTO NA MICROBIOTA DO SOLO UTILIZANDO O GENE 16S rRNA", Pedrinho *et al*. em 2009.

As amostras de solo foram obtidas de áreas da Empresa Brasileira de Pesquisa Agropecuária de Jaguariúna, SP, Brasil, localizada na latitude 22º 41' sul, longitude 47º W e altitude de 570 m e classificadas como Latossolo Vermelho Distroférrico (textura argilosa). As amostras foram submetidas a análises químicas e granulométricas. O lodo de esgoto utilizado no experimento foi obtido na Estação de Tratamento de Esgoto de Barueri, SP, que trata esgoto domiciliar e industrial. A extração do DNA metagenômico das comunidades microbianas de cada solo, a amplificação, o sequenciamento do gene 16S rRNA e a análise das sequências observadas foram realizados conforme proposto por PEREIRA et al. (2006) e SILVEIRA et al. (2006) (Pedrinho et al., 2009). As sequências obtidas do solo podem ser acessadas pelo Código GenBank |DQ084325.1| a |DQ084251.1|, sendo 76 sequências.

3.1.3. Origem das sequências SHC

Foram originalmente analisadas no trabalho "DIVERSIDADE DE BACTÉRIAS DO SOLO SOB VEGETAÇÃO NATURAL E CULTIVO DE HORTALIÇAS", Val-Moraes *et al*. em 2009.

A coleta do solo supressivo foi realizada em uma mata situada na microbacia do Córrego Taquara Branca, município de Sumaré (22° 49' 13" S; 47° 16' 08" W), no estado de São Paulo e a supressividade dos solos foi avaliada pelo método do crescimento do patógeno (Rhizoctonia solani) nas amostras de solo segundo Ghini; Zaroni (2001). O solo conducente com cultivo comercial de hortaliças foi coletado na mesma microbacia. As análises dos solos foram realizadas no Departamento de Solos e Adubos da Universidade Estadual Paulista, Campus de Jaboticabal, como pode ser

observado na Tabela 1. As análises químicas de macronutrientes; micronutrientes, e granulométricas com fracionamento de areia foram efetuadas de acordo com o Programa de Controle de Qualidade Análise de Solo - Sistema IAC. (RAIJ et al., 2001) (Val-Moraes *et al*., 2009). As sequências obtidas do solo podem ser acessadas pelo Código GenBank |AY694570.1| a |AY694434.1|, sendo no total 135 sequências.

3.1.4. Origem das sequências SCA e SPA

Foram originalmente analisadas no trabalho "CHANGES IN LAND USE ALTER THE STRUTURE OF BACTERIAL COMMUNITIES IN WESTERN AMAZON SOILS", Jesus *et al*., em 2009.

Os locais de coleta das amostras foi o município de Benjamin Constant, do Estado do Amazonas, Brasil, entre as coordenadas geográficas 41210 e 41260 para o sul e 691360 e 70110 oeste, ao longo do rio Solimões. O clima na região tem temperatura média anual de 25,7 1C e precipitação média de 2562 milímetros. Inceptisols é o tipo de solo predominante com Gleissolos e Alisols também encontrado (Coelho et al., 2005). A taxa de desmatamento nesta região é baixo devido principalmente à dificuldade de acesso e baixa densidade populacional. Comunidades locais usam a terra para corte e queima, e agricultura, no entanto, algumas áreas de pastagens também apresentam-se como uma conseqüência de políticas governamentais implementadas durante a década de 1970. Amostras compostas foram recolhidas a partir de culturas e pastagens em março de 2004. Todos locais eram originalmente floresta serrana. Espécies cultivadas nos locais de colheita foram cultivo de mandioca, milho, banana, cana e abacaxi. As bibliotecas foram construídas para as amostras de pastagem (Código GenBank. FJ037233-FJ037389) e culturas (Código GenBank FJ037390-FJ037544), ao todo somam 159 sequências.

3.1.5. Origem das sequências de SCOL

Foram originalmente analisadas no trabalho "MOLECULAR ANALYSIS OF THE BACTERIAL DIVERSITY IN A SPECIALIZED CONSORTIUM FOR DIESEL OIL DEGRADATION", Paixão et al., em 2010.

O solo usado para a obtenção da amostra foi coletado a partir de um local contaminado localizado em Ribeirão Preto, Estado de São Paulo (Brasil), a 21 ° 06 'S, 47 ° 49 'O, no passado, tinha uma indústria que fabricava lubrificantes automotivos. A extração de DNA metagenômico foi feita a partir da amostra de solo e amplificação do gene 16S Rna. As sequências obtidas podem ser acessadas pelo código GenBank FJ978048 a FJ978478, sendo 431 sequências.

3.1.6. Origem das sequências de SCC

Foram originalmente analisadas no trabalho "DIVERSITY OF 16S RRNA GENES FROM BACTERIA OF SUGARCANE RHIZOSPHERE SOIL", Pisa et al., em 2011.

A área de estudo do qual foram obtidas as sequências foi a Estação Experimental da Universidade Federal do Paraná (UFPR), localizado na a cidade de Paranavaí (23 º 05'30 "S/052 º 29'04" W), Estado de Paraná (Brasil). Foram coletadas amostras de solo aderido às raízes de cana (Saccharum spp; ordem Poales; família Poaceae), que foram submetidos à fertilização. O solo aderido às raízes de espécimes vegetais foi utilizado para a amplificação do gene 16S rRNA e para construção da biblioteca. Um total de 1g de solo de cada amostra foi utilizada para a extracção de ácido nucleico. As sequências obtidas podem ser acessadas pelo código GenBank HQD18044 à HQD18581, sendo 536 sequências.

3.2. Ferramenta Web: MG-RAST

Todas as sequências usadas neste trabalho foram submetidas à ferramenta Web MG-RAST. O banco de dados adotado para esse trabalho foi o RDP database disponível dentro do MG-Rast. Os parâmetros selecionados foram valor para e-value, que é a probalidade de ocorrência de falsos positivos, e valor de identidade, que é o número de resíduos similares identificados.

O valor máximo de corte para o e-value foi de $1e^{-5}$. Tal "nota de corte" foi estabelecida de acordo com o padrão aceito para uma primeira análise de sequências,

em que são aceitos apenas e-values menores que e-5 e e-10. O valor mínimo de identidade foi de 60%, pois foi trabalhado com uma margem de erro aceita.

As sequências foram comparadas a outras sequências contidas em bancos de dados explorados pelo programa. Diante da comparação feita pelo MG-RAST podemos obter informações quantitativas das populações microbianas com base nos dados das sequências.

3.2.1. Rarefação

A Rarefação é um método usado na comparação entre o número de espécies e comunidades quando o tamanho da amostra é diferente (MORENO, 2001). Tendo por objetivo verificar se o esforço amostral foi suficiente para representar a abundância e inferir a diversidade das amostras.

A análise de rarefação serve para demonstrar se as amostras utilizadas neste trabalho foram suficientes pra representar todas as comunidades bacterianas presentes nos solos. A ferramenta MG-RAST fornece uma análise de rarefação apenas em nível de espécies, tornando necessário o sequenciamento de um maior número de indivíduos para que a análise se torne mais robusta. Se a análise fosse ao nível de gênero e filo, provavelmente as curvas seriam diferentes, porque é mais difícil obter toda a diversidade de espécies de um ambiente do que obter toda a diversidade de gêneros e filos de um determinado ambiente.

4. RESULTADOS E DISCUSSÃO

Os dados relativos às sequências foram calculados para metagenomas com os seguintes códigos 4490035.3 (cultivo de cana - SCC), 4478966.3 (consórcio de óleo diesel - SCOL), 4487439.3 (solo cultivado da região da Amazônia (SCA), 4487440.3 (solo de pasto da região da Amazônia - SPA), 4476476.3 (solo com cultivo de hortaliças - SCH), 4488219.3 (solo contaminado com lodo de esgoto - SCLE), e 4476241.3 (solo que foi cultivado por grãos (soja, milho, feijão) (SCG)). Os dados foram comparados com o banco de dados do RDP usando um valor máximo de 1e-5 para e-value, uma identidade mínima de 60%, e um comprimento mínimo de alinhamento de 20 nucleotídeos. Os dados foram normalizados para valores entre 0 e 1, para permitir a comparação de amostras de diferentes tamanhos.

4.1. Distribuição de Domínio

A primeira figura gerada diz respeito às bactérias que puderam ser classificadas. A grande maioria foi classificada como bactéria como era esperada, no entanto, uma pequena quantidade de sequências foi identificada como sendo pertencente ao domínio taxonômico *Eucaryota*. Este resultado inesperado significa que essas sequências provavelmente apresentavam baixa qualidade, pois só foram selecionadas sequências provenientes do gene 16S rRNA, exclusivo de bactérias. Como os dados foram coletados do genbank, somente foi possível trabalhar com sequências fastas prontas, sem acesso aos eletroferogramas o que possibilitaria fazer uma rigorosa análise da qualidade das mesmas. Além disso, todas as sequências foram obtidas pelo método de Sanger, que como qualquer tipo de sequenciamento também está sujeito a falhas.

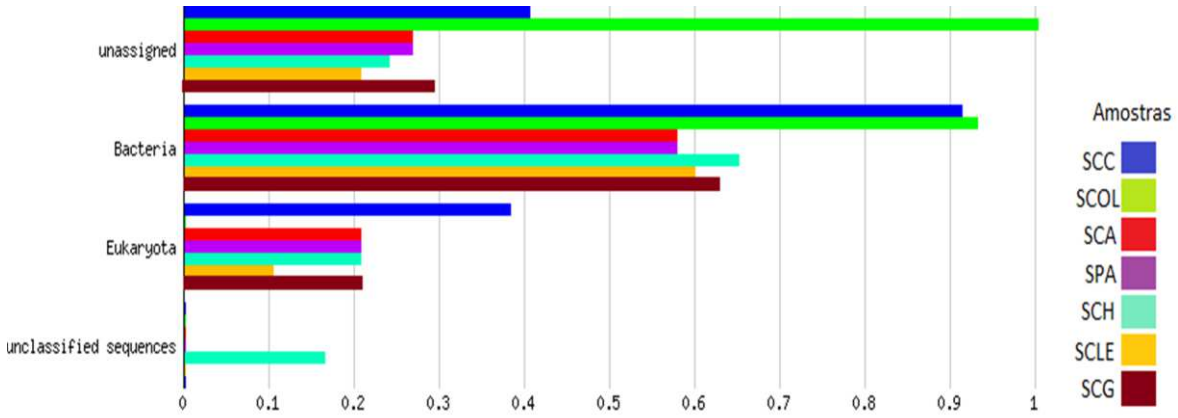

Figura 1. Comparação da distribuição de domínio entre os metagenomas, realizada pelo programa web MG-RAST. Os dados foram normalizados para valores entre 0 e 1. A amostra SCC (cultivo de cana) está representada pela cor azul; SCOL (consórcio de óleo diesel) pela cor verde; SCA (solo cultivado da região da Amazônia pela cor vermelha; SPA (solo de pasto da região da Amazônia) pela cor roxa; SCH (solo com cultivo de hortaliças) pela cor azul celeste; SCLE (solo contaminado com lodo de esgoto) pela cor amarela e SCG (solo que foi cultivado por grãos (soja, milho, feijão)) pela cor marrom.

4.2.Distribuição de Filo (Bactéria)

Observando a figura 2, referente à distribuição de filos do domínio *Bacteria*, o mais representativo é a *Proteobacteria* - cuja presença foi constatada nas sete amostras analisadas, sendo seguido dos filos *Firmicutes. Bacteriodetes, Acidobacteria* e *Actinobacteria*, respectivamente, estes presentes em cinco das sete amostras analisadas no presente trabalho. Os ambientes que mais se destacaram foram SCC e SCOL, respectivamente. Maiores detalhes sobre os filos encontrados podem ser observados no Anexo 1.

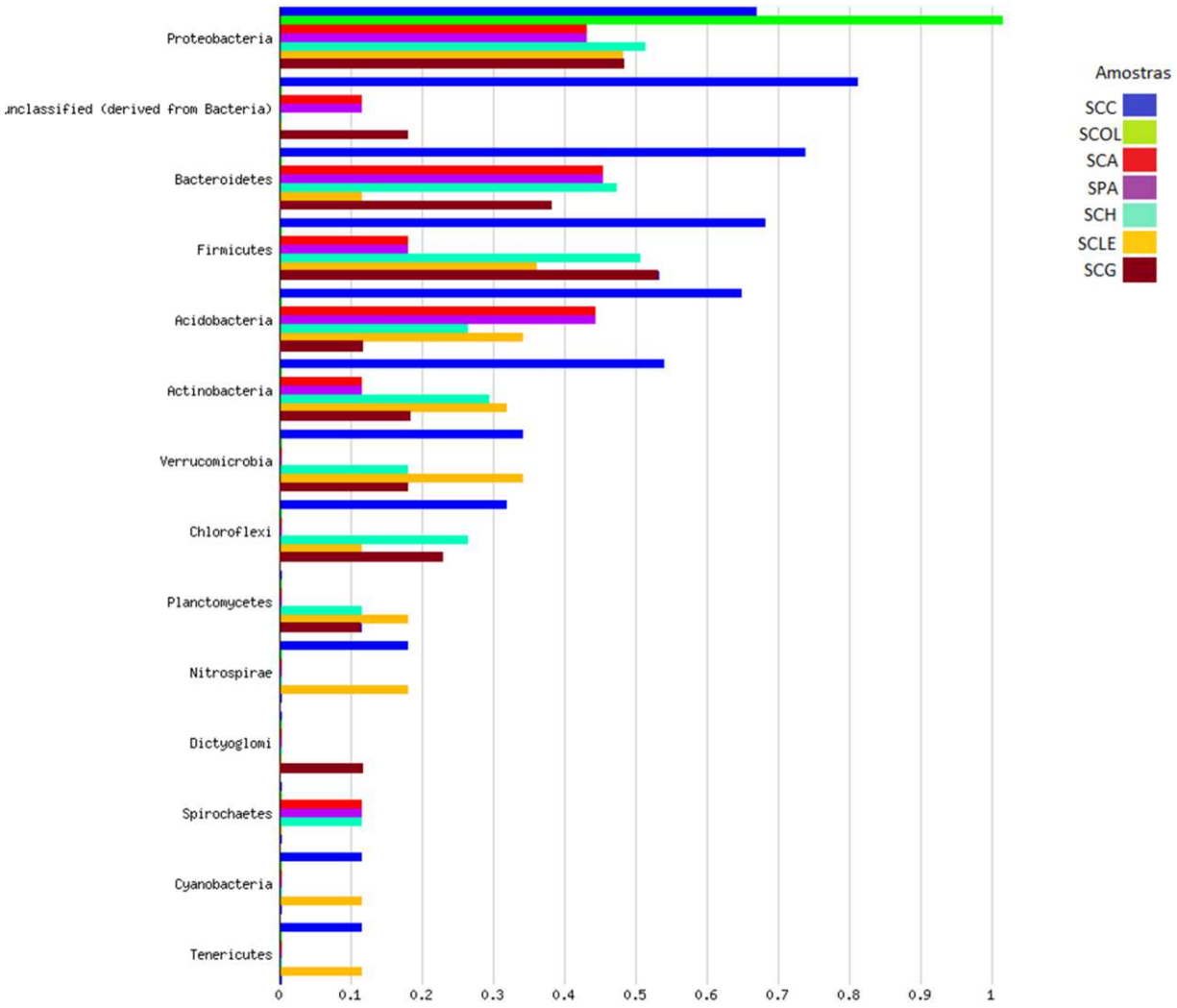

Figura 2. Comparação de Distribuição de Filo (Bactéria) entre os metagenomas. Os dados foram normalizados para valores entre 0 e 1. A amostra SCC (cultivo de cana) está representada pela cor azul; SCOL (consórcio de óleo diesel) pela cor verde; SCA (solo cultivado da região da Amazônia pela cor vermelha; SPA (solo de pasto da região da Amazônia) pela cor roxa; SCH (solo com cultivo de hortaliças) pela cor azul celeste; SCLE (solo contaminado com lodo de esgoto) pela cor amarela e SCG (solo que foi cultivado por grãos (soja, milho, feijão)) pela cor marrom.

4.2.1. Filo *Proteobacteria*

O filo *Proteobacteria* é o maior e mais variado grupo de bactérias cultivadas, apresentando uma grande diversidade metabólica e morfológica, com aproximadamente 528 gêneros já descritos segundo Garrity *et al*, (2004). São muitas as fontes de obtenção

de energia deste filo, como organismos quimiolitotróficos, quimiorganotróficos, e fototróficos, entre outras vias metabólicas (CANHOS *et al.*, 1997).

Por apresentar tais características, muitos estudos observaram sua presença em vários ambientes, porém ocorre mais frequentemente em solos cultivados, como descritos por Pereira *et al.* (2006). Como pode ser visto na figura 3, os ambientes que obtiveram maior predominância deste filo foram o SCOL e SCC, estes alcançaram um valor muito significativo, já que são solos impactados por óleo diesel e cultivo de cana, respectivamente com grande circulação de matéria orgânica, sendo assim um habitat para diversas espécies de microrganismos.

Maiores detalhes desse filo podem ser observados na figura 3, onde é possível verificar que as classes *Alphaproteobacteria*, *Gamaproteobacteria* e *Betaproteobacteria* estão presentes em todas as amostras, demostrando a grande diversidade referente à *Proteobacteria*. Considerando as classes *Alphaproteobacteria e Gamaproteobacteria*, podemos observar na figura que os ambientes que mais se sobressaíram foram novamente SCC e SCOL, sendo os mais próximos do valor de 1. Em *Alphaproteobacteria* são seguidos por SCA, que também está em número significativo entre os demais, com 0.314 do valor de 1. Já a classe *Betaproteobacteria* não apresentou grandes variações entre as amostras.

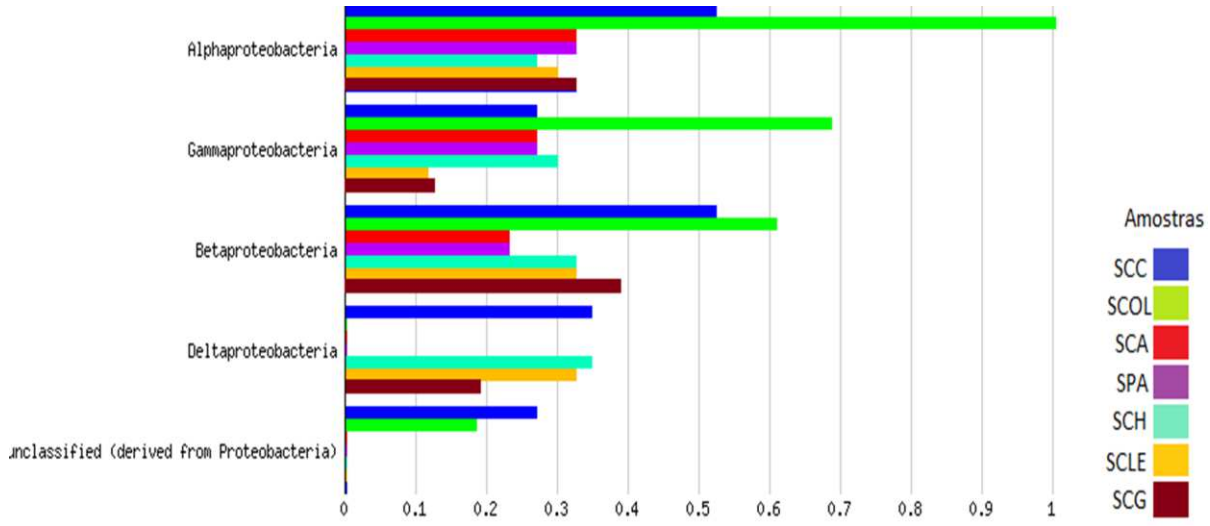

Figura 3. Distribuição do Filo Proteobacteria. Os dados foram normalizados para valores entre 0 e 1. A amostra SCC (cultivo de cana) está representada pela cor azul; SCOL (consórcio de óleo diesel) pela cor verde; SCA (solo cultivado da região da Amazônia pela cor vermelha; SPA (solo de pasto da região da Amazônia) pela cor roxa; SCH (solo com cultivo de hortaliças) pela cor azul celeste; SCLE (solo contaminado com lodo de esgoto) pela cor amarela e SCG (solo que foi cultivado por grãos (soja, milho, feijão)) pela cor marrom.

Na distribuição de ordens em *Alphaproteobacteria*, visto na figura 4, que foi a classe que mais se destacou na distribuição de *Proteobacteria*, verificamos que a ordem *Rhizobiales* atinge a numeração máxima com a amostra SCOL, e ainda tendo sua importância destacada por estar presente em todas as amostras. O ambiente SCC atinge um valor significativo, porém abaixo da média, com 0.389. Em *Sphingomonadales*, SCOL também se destaca atingindo 0.7, entre demais amostras presentes nessa ordem, que são SCC, ARR, SCA e SPA.

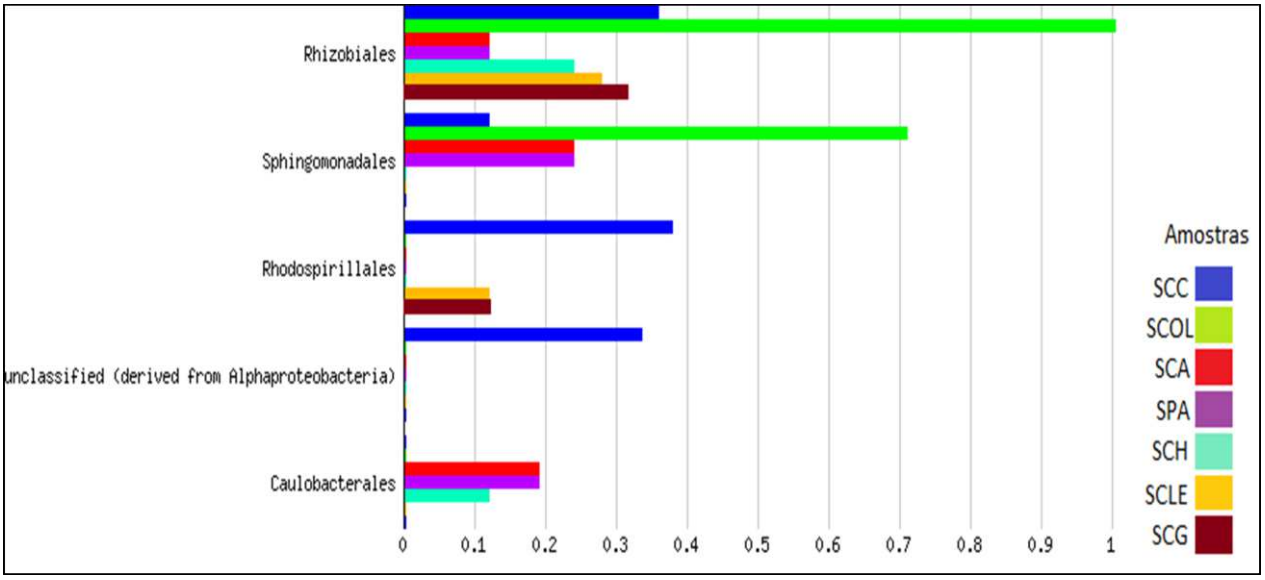

Figura 4. Distribuição da Ordem (Alphaproteobacteria). Os dados foram normalizados para valores entre 0 e 1. A amostra SCC (cultivo de cana) está representada pela cor azul; SCOL (consórcio de óleo diesel) pela cor verde; SCA (solo cultivado da região da Amazônia pela cor vermelha; SPA (solo de pasto da região da Amazônia) pela cor roxa; SCH (solo com cultivo de hortaliças) pela cor azul celeste; SCLE (solo contaminado com lodo de esgoto) pela cor amarela e SCG (solo que foi cultivado por grãos (soja, milho, feijão)) pela cor marrom.

Em destaque à importância da ordem *Rhizobiales*, a figura 5 ressalta a importância de SCOL. *Methylocystaceae* é uma família de bactérias que são capazes de obtenção de carbono e energia a partir de metano, ela está presente somente em SCOL, chegando sua presença ao valor máximo. *Beijerinckiaceae* também está presente somente em SCOL, porém em um alcance bem menor na figura comparado à *Methylocystaceae*, pois chega à 0.123. *Phyllobacteriaceae* tem sua presença em SCOL chegando à 0.738, sendo seguido dos solos SCA e SPA com valores idênticos de 0.123. A terceira família que também se destaca é Bradyrhizobiaceae, estando presente em cinco ambientes, os quais são SCC, SCOL, SCH, SCLE e SCG.

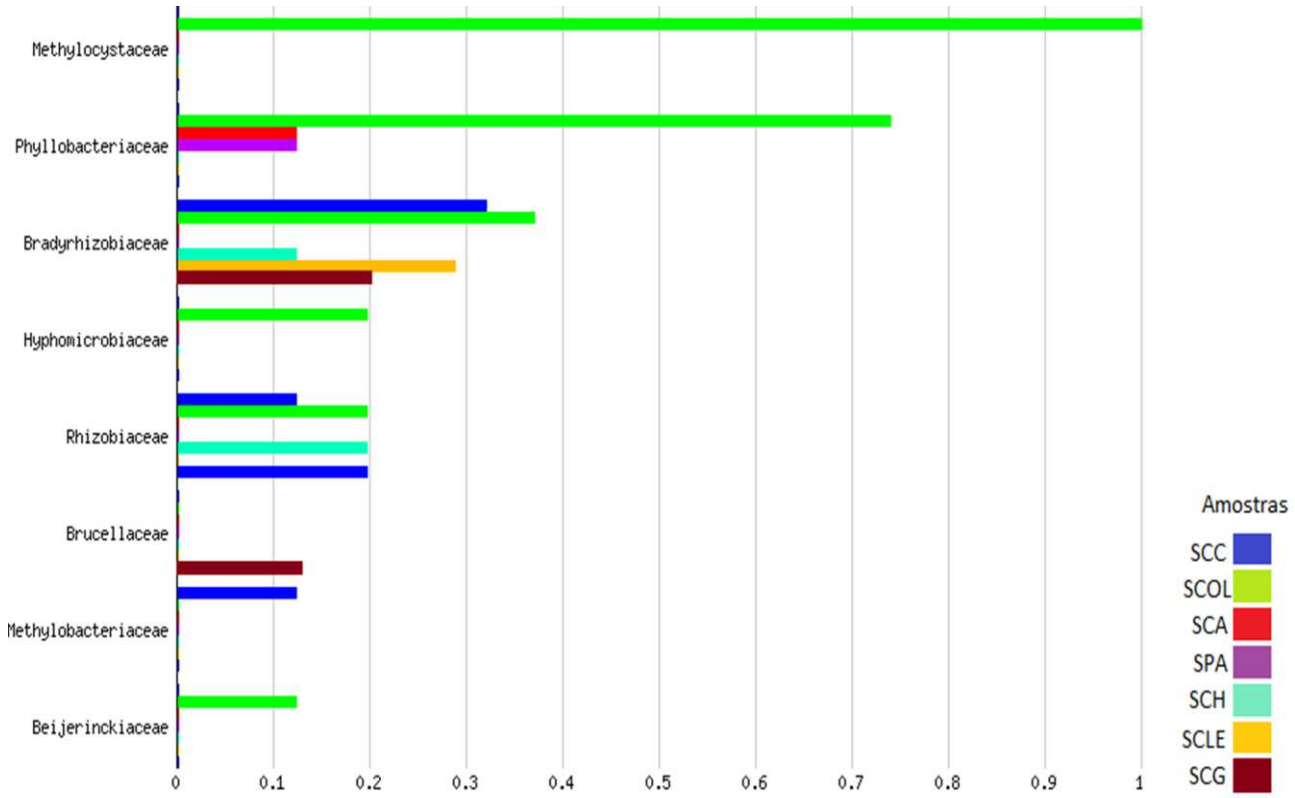

Figura 5. Distribuição de Ordem (*Rhizobiales*). Os dados foram normalizados para valores entre 0 e 1. A amostra SCC (cultivo de cana) está representada pela cor azul; SCOL (consórcio de óleo diesel) pela cor verde; SCA (solo cultivado da região da Amazônia pela cor vermelha; SPA (solo de pasto da região da Amazônia) pela cor roxa; SCH (solo com cultivo de hortaliças) pela cor azul celeste; SCLE (solo contaminado com lodo de esgoto) pela cor amarela e SCG (solo que foi cultivado por grãos (soja, milho, feijão)) pela cor marrom.

Na figura 6 é ressaltada a presença de apenas um gênero à família *Bradyrhizobiaceae*, o *Bradyrhizobium.* O gênero *Bradyrhizobium,* microrganismo importante na fixação biológica de nitrogênio prevalece atingindo valor máximo de um em SCOL, seguido de SCC e SCLE, que atingiram 0.862 e 0.774, respectivamente. Já na figura 7, podemos observar a distribuição de espécies do gênero *Bradyrhizobium,* onde destacamos a presença de *Bradyrhizobium japonicum* em três solos em comum, o SCOL, SCH e SCLE. *Bradyrhizobium sp. RSB6* e *Bradyrhizobium sp. SEMIA 6099* constam somente em SCC; *Bradyrhizobium sp. CCBAU 33039* somente em SCLE e *Bradyrhizobium.elkani* somente em SCG.

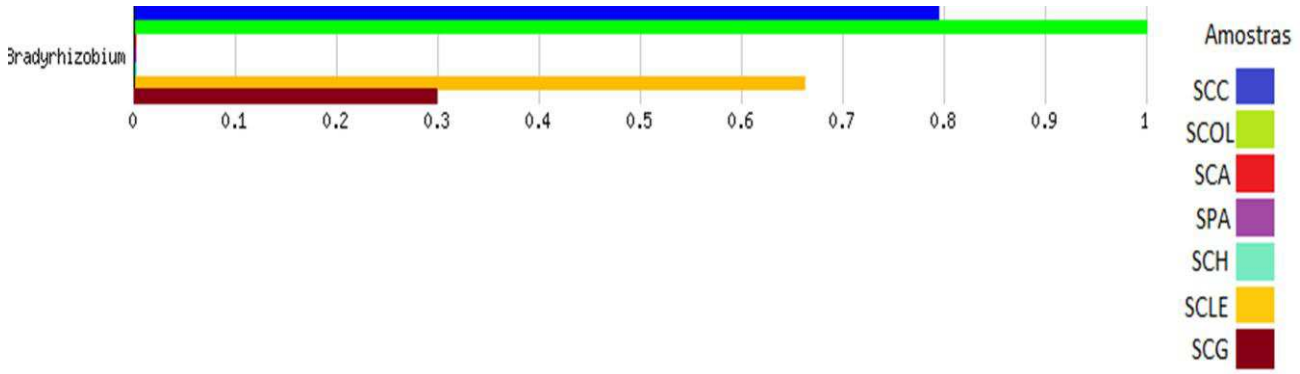

Figura 6. Distribuição de gênero (*Bradyrhizobiaceae*). Os dados foram normalizados para valores entre 0 e 1. A amostra SCC (cultivo de cana) está representada pela cor azul; SCOL (consórcio de óleo diesel) pela cor verde; SCA (solo cultivado da região da Amazônia pela cor vermelha; SPA (solo de pasto da região da Amazônia) pela cor roxa; SCH (solo com cultivo de hortaliças) pela cor azul celeste; SCLE (solo contaminado com lodo de esgoto) pela cor amarela e SCG (solo que foi cultivado por grãos (soja, milho, feijão)) pela cor marrom.

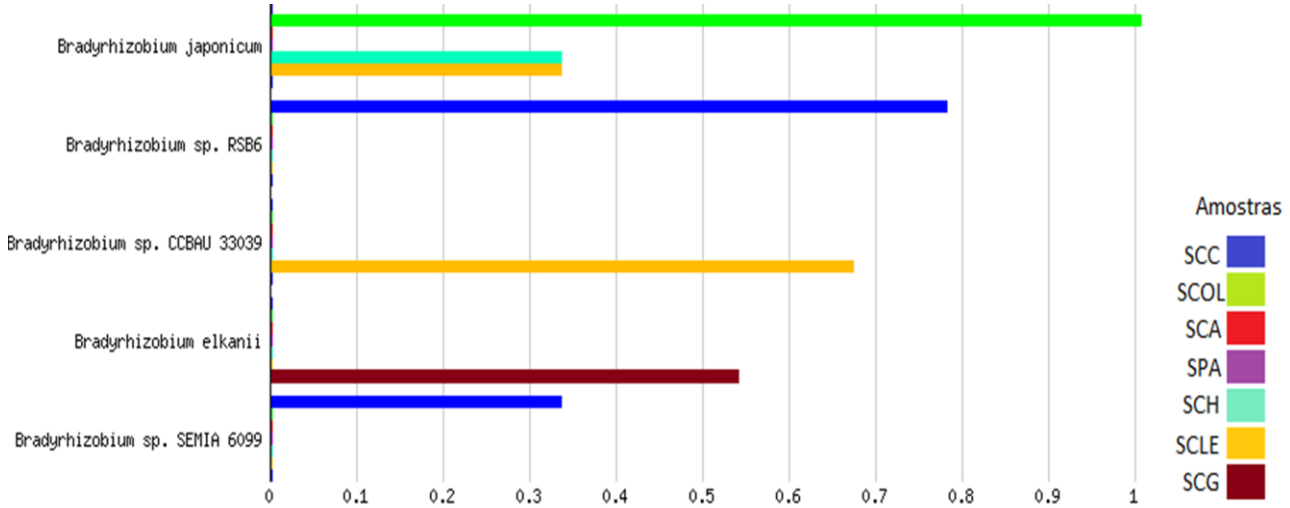

Figura 7. Distribuição de espécies (Bradyrhizobium)**.** Os dados foram normalizados para valores entre 0 e 1. A amostra SCC (cultivo de cana) está representada pela cor azul; SCOL (consórcio de óleo diesel) pela cor verde; SCA (solo cultivado da região da Amazônia pela cor vermelha; SPA (solo de pasto da região da Amazônia) pela cor roxa; SCH (solo com cultivo de hortaliças) pela cor azul celeste; SCLE (solo contaminado com lodo de esgoto) pela cor amarela e SCG (solo que foi cultivado por grãos (soja, milho, feijão)) pela cor marrom.

4.2.2. Filo *Firmicutes*

O filo *Firmicutes* está presente em seis dos sete ambientes estudados, sendo estes SCC, SCA, SPA, SCH, SCLE e SCG. O solo SCG, foi a amostra em que mais foi constatada sua presença. Lembrando que este é um solo que foi cultivado por diferentes grãos (feijão, milho e soja) em determinadas épocas, pode-se estimar que os representantes deste filo se sobressaíram por serem estrategistas "r", cuja estratégia de sobrevivência é sua rápida multiplicação no meio em que vive, quando há nutrientes em quantidades abundantes (ATLAS, 1999). Eles são encontrados em meios instáveis caracterizados por várias transições. Os solos cultivados sofrem ação antrópica sucessivamente, como isso provoca contínuas mudanças na microbiota, o crescimento de *Firmicutes* é favorecido (PEDRINHO *et al.*,2009). Como podem ser observadas na figura 8, as classes predominantes nas cinco amostras foram Bacilli e Clostridia. Os membros destas classes possuem a função de fixação de nitrogênio, o que podem promover o crescimento de determinadas espécies de plantas (KUHNER *et al.*, 2000).

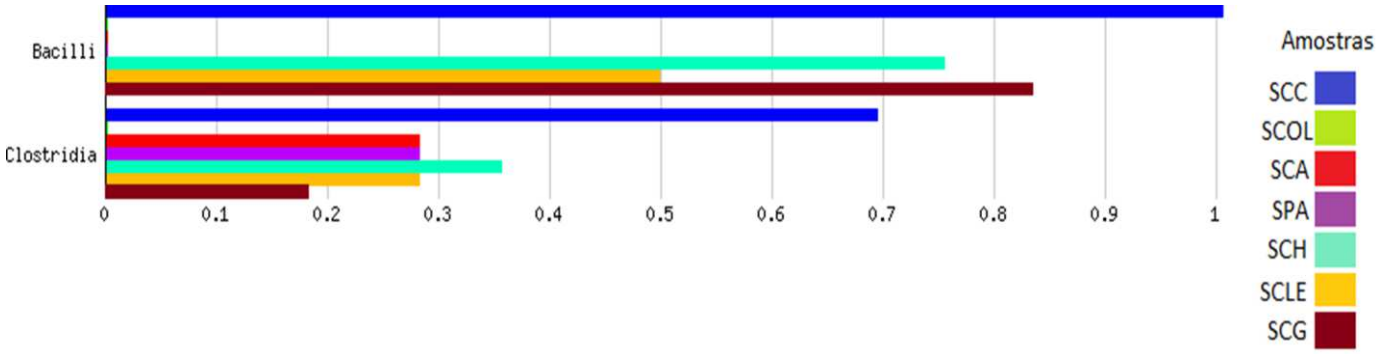

Figura 8. Distribuição de classes de Firmicutes. Os dados foram normalizados para valores entre 0 e 1. A amostra SCC (cultivo de cana) está representada pela cor azul; SCOL (consórcio de óleo diesel) pela cor verde; SCA (solo cultivado da região da Amazônia pela cor vermelha; SPA (solo de pasto da região da Amazônia) pela cor roxa; SCH (solo com cultivo de hortaliças) pela cor azul celeste; SCLE (solo contaminado com lodo de esgoto) pela cor amarela e SCG (solo que foi cultivado por grãos (soja, milho, feijão)) pela cor marrom.

4.2.3. Filo *Bacteroidetes*

O filo *Bacteroidetes* também abrange seis ambientes, os quais são SCC, SCA, SPA, SCH, SCLE e SCG. A amostra que mais se destacou foi SCC, seguido dos demais ambientes. Bacteroidetes é composto por três classes: *Bacteroidetes, Flavobacteria* e *Sphingobacteria*. A classe *Bacteroidetes* foi encontrada em SCA e SPA. Já a classe *Flavobacteria* teve constatada sua presença em SCC, SCA, SPA, SCH. De acordo com Kirchman (2002), esta classe é predominante em ambientes marinhos, como pode ser observado na figura 9, no qual Flavobacteria predomina igualmente com 0.554 em SCA e SPA, amostras que foram trabalhadas em conjunto, porém extraídas de cultivos diferentes. Estas bactérias heterotróficas são consideradas fundamentais na captação e degradação de matéria orgânica (VANTRAPPEN *et al.*, 2004).

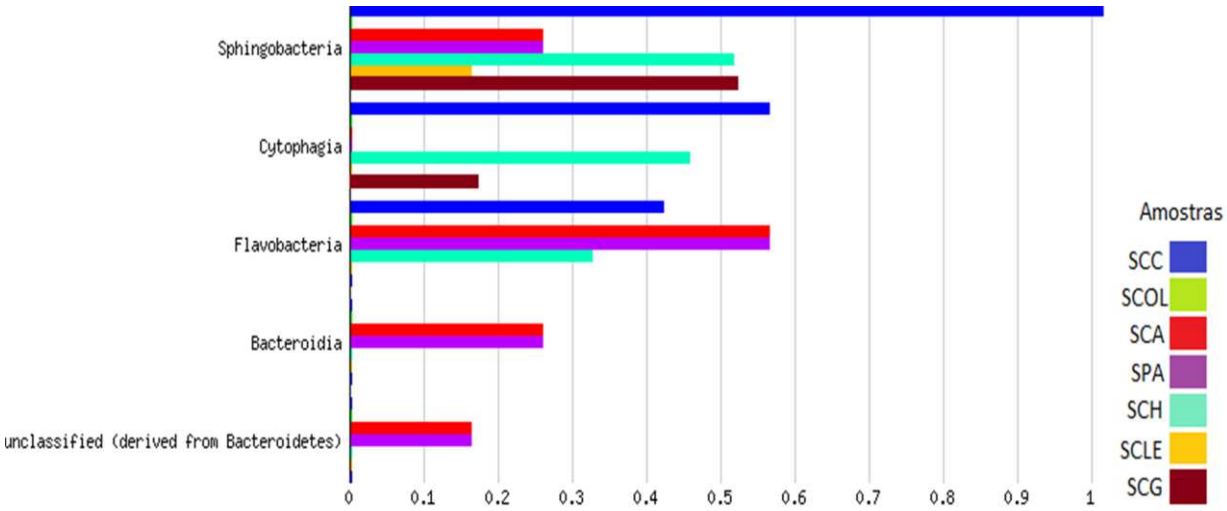

Figura 9. Distribuição de classes de Bacteroidetes Os dados foram normalizados para valores entre 0 e 1. A amostra SCC (cultivo de cana) está representada pela cor azul; SCOL (consórcio de óleo diesel) pela cor verde; SCA (solo cultivado da região da Amazônia pela cor vermelha; SPA (solo de pasto da região da Amazônia) pela cor roxa; SCH (solo com cultivo de hortaliças) pela cor azul celeste; SCLE (solo contaminado com lodo de esgoto) pela cor amarela e SCG (solo que foi cultivado por grãos (soja, milho, feijão)) pela cor marrom.

4.2.4. Filo *Acidobacteria*

Este filo está presente em seis amostras das analisadas no presente trabalho, as quais são SCC; SCA; SPA; SCH; SCLE e SCG. Em destaque na figura 10 em

referência à este filo, temos a amostra SCC, cujo solo foi cultivado por cana de açúcar sob diferentes condições de adubação de nitrogênio. Este solo tinha como pH 5.2 e 5.1, podendo ser uma possível explicação para maior ocorrência de Acidobactéria, já que a presença deste filo está diretamente relacionada com o pH do solo (SAIT *et al.*, 2006). As Acidobactérias são componentes ecológicos de grande importância nos ecossistemas, sendo geralmente encontrados em amostras ambientais como fontes termais, ambientes marinhos, terrestres e animais (BARNS *et al.*, 2007), mas principalmente em comunidades de solo (HUGENHOLTZ *et al.*, 1998), podendo constituir de 30 a 50% do total da população bacteriana obtida em distintas amostras de solo (SILVEIRA *et al.*, 2006). Este filo foi agrupado e denominado em estudos realizados por Hirashi et al. (1995), utilizando o gene 16S rRNA, possuindo atualmente três representantes cultiváveis (PEDRINHO *et al.*, 2009).

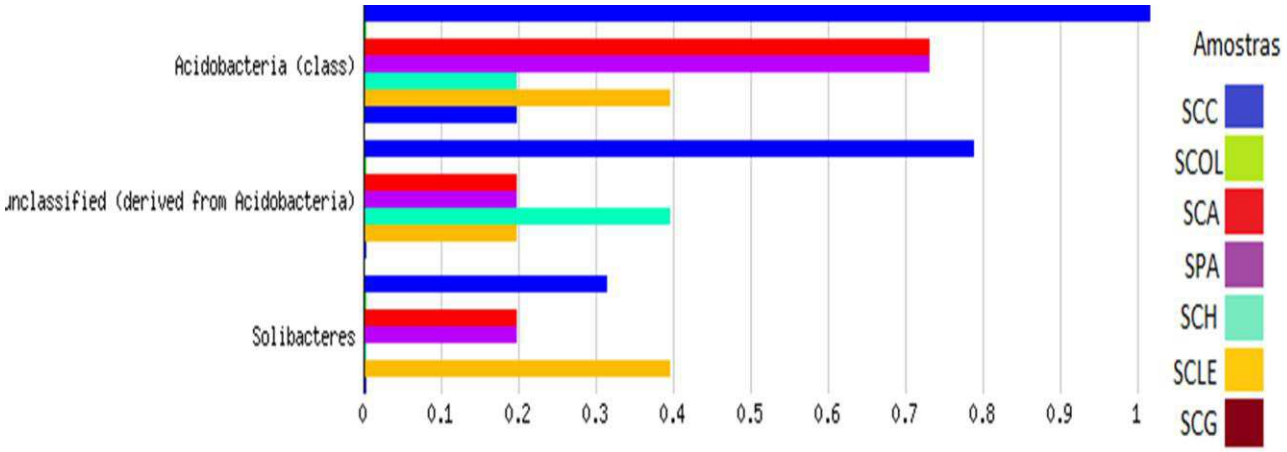

Figura 10. Distribuição de classes de *Acidobacterias*. Os dados foram normalizados para valores entre 0 e 1. A amostra SCC (cultivo de cana) está representada pela cor azul; SCOL (consórcio de óleo diesel) pela cor verde; SCA (solo cultivado da região da Amazônia pela cor vermelha; SPA (solo de pasto da região da Amazônia) pela cor roxa; SCH (solo com cultivo de hortaliças) pela cor azul celeste; SCLE (solo contaminado com lodo de esgoto) pela cor amarela e SCG (solo que foi cultivado por grãos (soja, milho, feijão)) pela cor marrom.

4.2.5. Filo *Actinobacteria*

Na figura 2, referente à distribuição de Filos, podemos verificar que Actinobacteria se fez presente em 6 solos, com maior abundancia em SCLE com 0.529; com presença igual para SCA e SPA com 0.111 e SCH, SCLE e SCG com 0.287, 0.312 e 0.176 respectivamente. Na figura 11, três ordens foram referente à este filo,

Actinomycetales, Bifidobacteriales e Rubrobacterales. *Actinomycetales* está presente em todos os solos exceto em SCOL, e ainda em SCC atinge valor máximo de um. *Bifidobacteriales* está presente somente em SCC com 0.436, esta ordem é considerada um componente importante da microbiota devido ao elevado nível de oligossacáridos complexos disponíveis. A sua utilização tem sido importante para a inibição de agentes patogénicos no intestino, como suplemento alimentar e tratamentos doenças utópicas (MIQUELLETO *et al.,* 2011). *Rubrobacterales* está presente somente em SCG com um índice de 0.218, esta ordem é conhecida por sua alta resistência à radiação ultravioleta e ionizante, e, por conseguinte, exposta à radiação UV forte (COURADEAU *et al.*, 2011). As actinobactérias possuem especial importância à indústria farmacêutica por sua capacidade ilimitada de produzir metabólitos secundários possuindo diversas estruturas químicas e atividades biológicas, sendo um mercado promissor para estudos de novas drogas farmacêuticas. Juck *et al.* (2000) observaram que, em solo contaminado 64 % das bactérias pertenciam à ordem *Actinomycetales*. A capacidade desse grupo de microrganismo degradar hidrocarbonetos tem sido reportada na literatura indicando que dispõe de mecanismos de adaptação e que pode ser valioso para o desenvolvimento de novas estratégias de biorremediação (BRITO *et al.*, 2006).

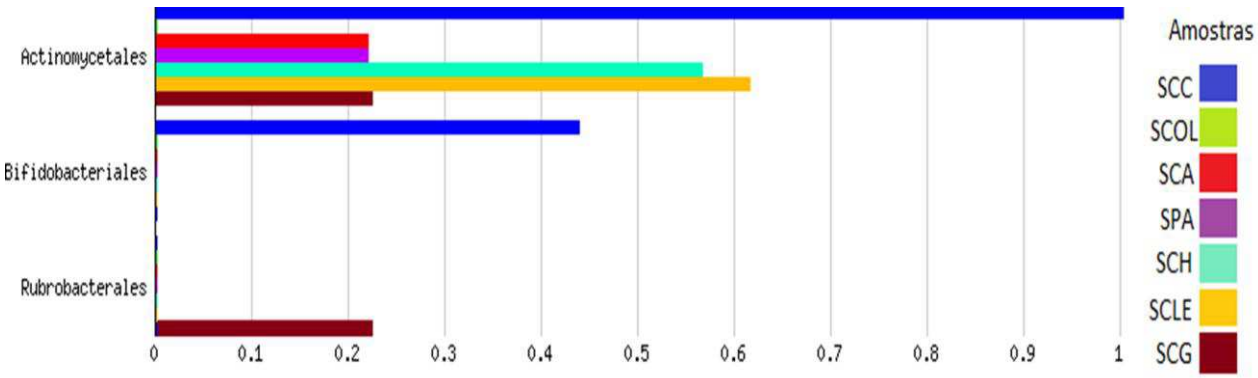

Figura 11. Distribuição de Ordem de *Actinobacteria.* Os dados foram normalizados para valores entre 0 e 1. A amostra SCC (cultivo de cana) está representada pela cor azul; SCOL (consórcio de óleo diesel) pela cor verde; SCA (solo cultivado da região da Amazônia pela cor vermelha; SPA (solo de pasto da região da Amazônia) pela cor roxa; SCH (solo com cultivo de hortaliças) pela cor azul celeste; SCLE (solo contaminado com lodo de esgoto) pela cor amarela e SCG (solo que foi cultivado por grãos (soja, milho, feijão)) pela cor marrom.

4.2.6. Filo *Verrucomicrobia*

Na figura 2 de distribuição de Filos, o filo *Verrucomicrobia* atinge quatro solos, sendo eles SCC, SCLE, SCH e SCG, com valores de 0.334 e 0.176 respectivamente por dupla de amostras. Apesar de atingirem pouca evidência nos solos citados, foram distinguidas três classes predominantes deste filo nas análises, sendo *Spartobacteria* predominante em SCC; *Verrucomicrobiae* em SCC, SCH, SCLE e SCG; e *Opitutae* somente em SCLE. A evidência disponível indica, portanto, que os membros do filo *Verrucomicrobia* são um grupo bem distribuído, abundante e ativo de bactérias do solo. Porém, muito pouco se sabe sobre a biologia dos membros deste filo, a informação sobre suas funções e ecologia é baseado apenas em correlações de sua abundância, com base no 16S rRNA e parâmetros ambientais (SANGWAN *et al.*, 2005).

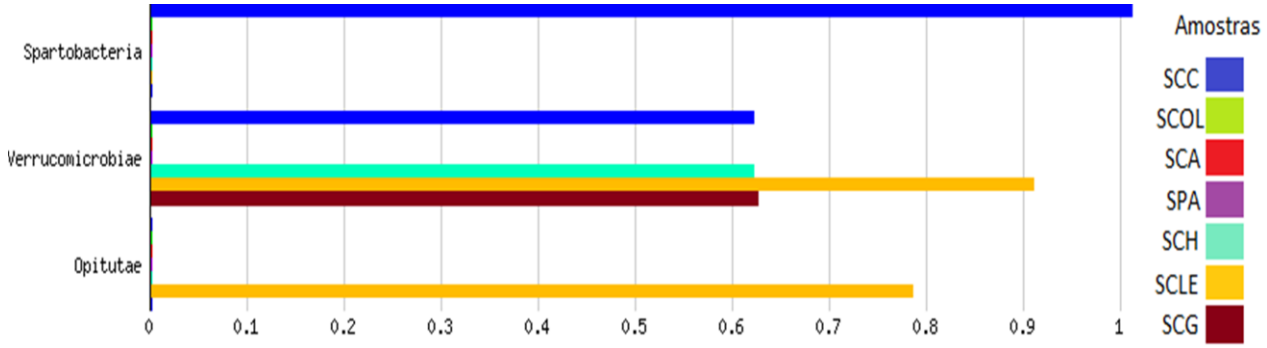

Figura 12. Distribuição de Classes de *Verrucomicrobia.* Os dados foram normalizados para valores entre 0 e 1. A amostra SCC (cultivo de cana) está representada pela cor azul; SCOL (consórcio de óleo diesel) pela cor verde; SCA (solo cultivado da região da Amazônia pela cor vermelha; SPA (solo de pasto da região da Amazônia) pela cor roxa; SCH (solo com cultivo de hortaliças) pela cor azul celeste; SCLE (solo contaminado com lodo de esgoto) pela cor amarela e SCG (solo que foi cultivado por grãos (soja, milho, feijão)) pela cor marrom.

4.2.7. Filo *Chloroflexi*

Chloroflexi está presente em quatro solos, os quais são SCC, SCH, SCLE e SCG – com 0.312, 0.258, 0.111, 0.222, respectivamente - como pode ser observado na figura 2. Na figura 13, são ressaltadas as classes de *Chloroflexi* predominantes nestes solos, como *Ktedonobacteria, Chloroflexi (class), Thermomicrobia (class) e Dehalococcoidetes.*

O Filo *Chloroflexi* é composto por espécies filamentosas que estão comumente presentes em lodo ativado de instalações de tratamento de esgoto (KRAGELUND *et al.*, 2007). É um organismo que sintetiza carotenoides, porém não possui quaisquer informações detectadas em seu genoma sobre genes que possam estar presentes no processo de biossíntese da clorofila ou fotossíntese. No entanto, muitos genes possivelmente envolvidos na produção de metabólitos secundários foram encontrados em seu genoma, deste modo é possível que alguns genes possam conferir a capacidade para fazer novos compostos (PS, U.S).

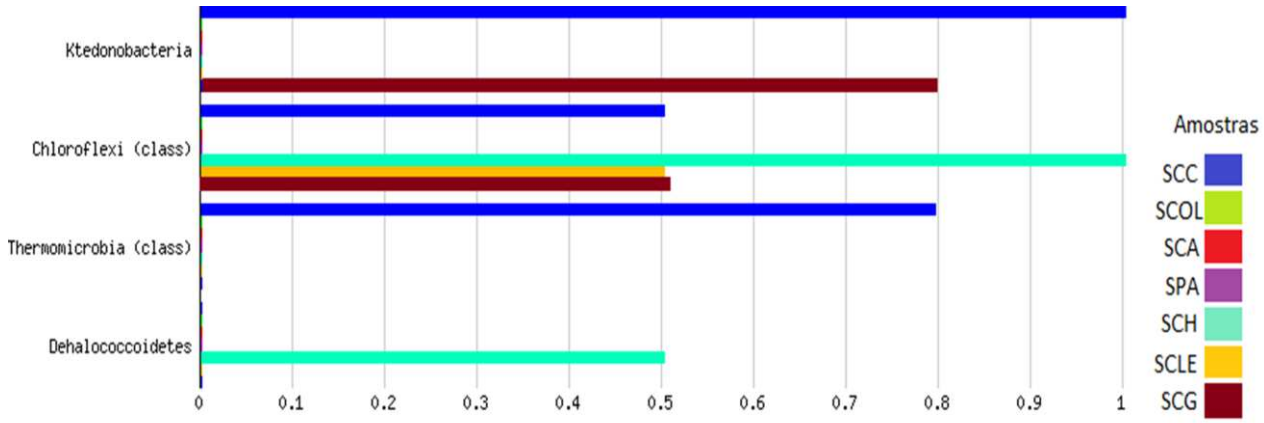

Figura 13. Distribuição de Classes de *Chloroflexi***.** Os dados foram normalizados para valores entre 0 e 1. A amostra SCC (cultivo de cana) está representada pela cor azul; SCOL (consórcio de óleo diesel) pela cor verde; SCA (solo cultivado da região da Amazônia pela cor vermelha; SPA (solo de pasto da região da Amazônia) pela cor roxa; SCH (solo com cultivo de hortaliças) pela cor azul celeste; SCLE (solo contaminado com lodo de esgoto) pela cor amarela e SCG (solo que foi cultivado por grãos (soja, milho, feijão)) pela cor marrom.

4.2.8. Filo *Planctomycetes*

Planctomycetes está presente em três solos, SCH, SCLE, e SCG. Mesmo sendo em um valor baixo possui maior predominância em SCLE, com 0.176, o que confirma sua abundância ser mais destacada em ambientes aquáticos (AXELROOD *et al.*, 2002). Porém, desde sua descoberta em 1992, os membros de *Planctomycetes* foram encontrados em bibliotecas bacterianas geradas a partir de uma variedade de solos, como solo cultivado, pastagem, floresta, tundra, e até mesmo em solos térmicos (BUCKLEY *et al.*, 2006). A escassez de informações disponíveis sobre *Planctomycetes*, em soma com sua abundância no solo, se torna um alvo de muitas

pesquisas sobre a diversidade microbiana dos mesmos. Na figura 14, podemos notar a predominância de *Planctomycetacia* em somente em SCLE, muito possivelmente porque o lodo de esgoto é proveniente de estações de tratamento de água o que facilita a presença de microrganismos desta classe.

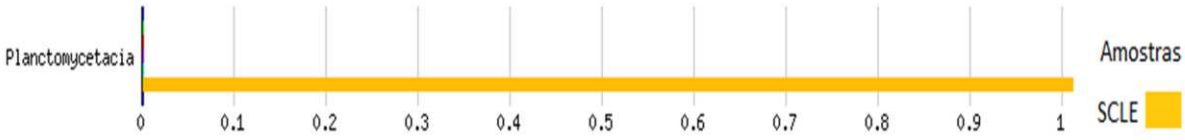

Figura 14. Distribuição de Classes de *Planctomycetes* Os dados foram normalizados para valores entre 0 e 1. A amostra SCLE (solo contaminado com lodo de esgoto) é representada pela cor amarela.

4.2.9. Filo *Nitrospirae*

Nitrospirae, como visto na figura 2, somente presencia-se em um único solo, o SCLE, com alcance de 0.176. Deste modo, temos este solo como uma região de interesse para extração de material para futuros estudos, já que *Nitrospirae* são bactérias nitrito-oxidantes, sendo considerados os principais oxidantes de nitrito em estações de tratamento de águas residuais e de reatores em escala de laboratório (HOLGER *et al.*, 2001).

4.2.10. Filo *Dictyoglomi*

Dictyoglomi é um filo presente somente em SCG, possuindo um alcance de 0.111 na figura 2. Mesmo havendo a classificação deste filo na distribuição entre as amostras de solos, não houve classificações posteriores como classes, famílias ou gêneros. Possivelmente, por ainda ser um filo ainda pouco explorado pela comunidade cientifica, mesmo sendo um filo bacteriano anaeróbio termofílico caracterizado pela capacidade de produzir enzimas úteis, tais como amilase, mananase, e xilanase (NISHIDA *et al.*, 2011).

4.2.11. Filo *Spirochaetes*

Spirochaetes está presente em três solos, SCA, SPA e SCG, estes aparecem na mesma proporção, com um alcance de 0.111 de um. Mesmo havendo a classificação deste filo na distribuição de filos entre os solos citados, não houve classificações posteriores como classes, famílias ou gêneros. É um filo que possivelmente não detém de uma grande significância em análises do solo, já que na literatura não se encontra quaisquer informações neste sentido, pois somente é citado como um filo que contém espécies patogénicas para humanos e animais (EGGERS *et al*., 2000).

4.2.12. Filo *Cyanobacteria*

Cyanobacteria é um filo presente em dois solos, os quais são SCC e SCLE, possuindo ainda o mesmo valor de alcance de 0.111. Mesmo havendo a classificação deste filo na distribuição de filos entre os solos citados, não houve classificações posteriores como classes, famílias ou gêneros. Mas, possui uma interessante importância ecológica, na qual são consideradas como uma das principais provedoras de nitrogênio para cadeias tróficas dos mares, sendo ainda muito utilizada em pesquisas na alimentação humana e produção de biocombustíveis (OREN *et al.*, 2004).

4.2.13. Filo *Tenericutes*

Tenericutes é um filo presente em dois solos, os quais são SCC e SCLE, possuindo ainda o mesmo valor de alcance de 0.111. Mesmo havendo a classificação deste filo na distribuição de filos entre os solos citados, não houve classificações posteriores como classes, famílias ou gêneros. Na literatura científica, ainda são escassas as fontes de informações que possam trazer alguma importância a este filo.

4.3. Distribuição de Filo (não distribuído)

A figura 15 traz a distribuição de filo não distribuído, que ocorre pela não utilização de parâmetros rígidos, pois os dados foram comparados usando um valor máximo de 1e-5 para e-value, identidade de 20%, e comprimento de alinhamento de 20%, tais parâmetros são considerados mínimos. A escolha desses parâmetros foi embasada no fato de sete das amostras serem sequenciadas pelo método de Sanger e apresentar pequeno número de sequencias levando-se em consideração as novas tecnologias de sequenciamento utilizadas atualmente.

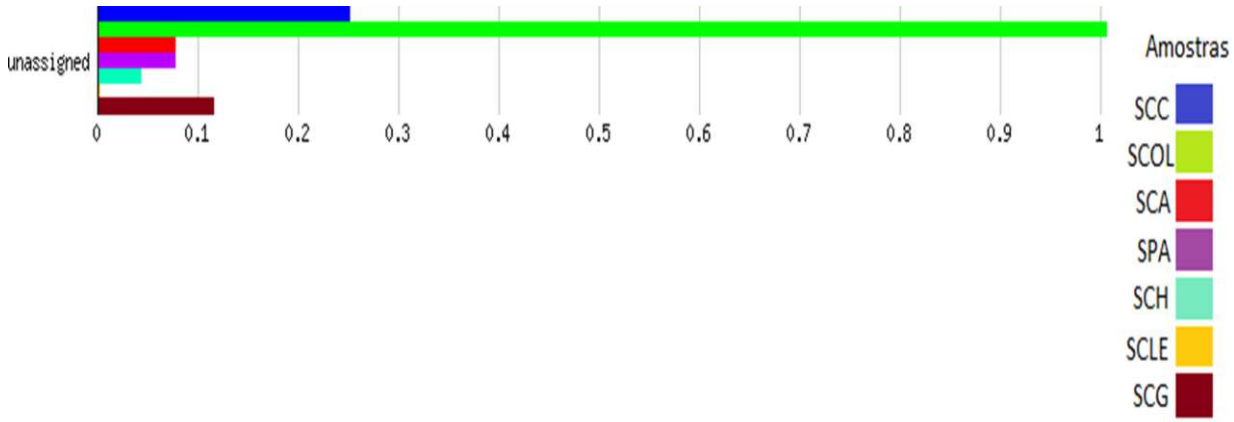

Figura 15. Distribuição de Filo (não distribuído)**.** Os dados foram normalizados para valores entre 0 e 1. A amostra SCC (cultivo de cana) está representada pela cor azul; SCOL (consórcio de óleo diesel) pela cor verde; SCA (solo cultivado da região da Amazônia pela cor vermelha; SPA (solo de pasto da região da Amazônia) pela cor roxa; SCH (solo com cultivo de hortaliças) pela cor azul celeste; SCLE (solo contaminado com lodo de esgoto) pela cor amarela e SCG (solo que foi cultivado por grãos (soja, milho, feijão)) pela cor marrom.

4.4. Rarefação

Os solos sequenciados há mais tempo apresentam um número menor de sequências, como SCOL, SCA, SPA e SCC, sendo o motivo os custos de sequenciamento da época, quando comparados com solos provenientes de trabalhos mais recentes, como SCH, SCG e SCLE, respectivamente. Portanto, os solos com menor número de sequências possivelmente foram prejudicados na amostragem de populações. Deste modo, torna-se ainda necessário o sequenciamento de um maior número de indivíduos para que a análise se torne mais robusta.

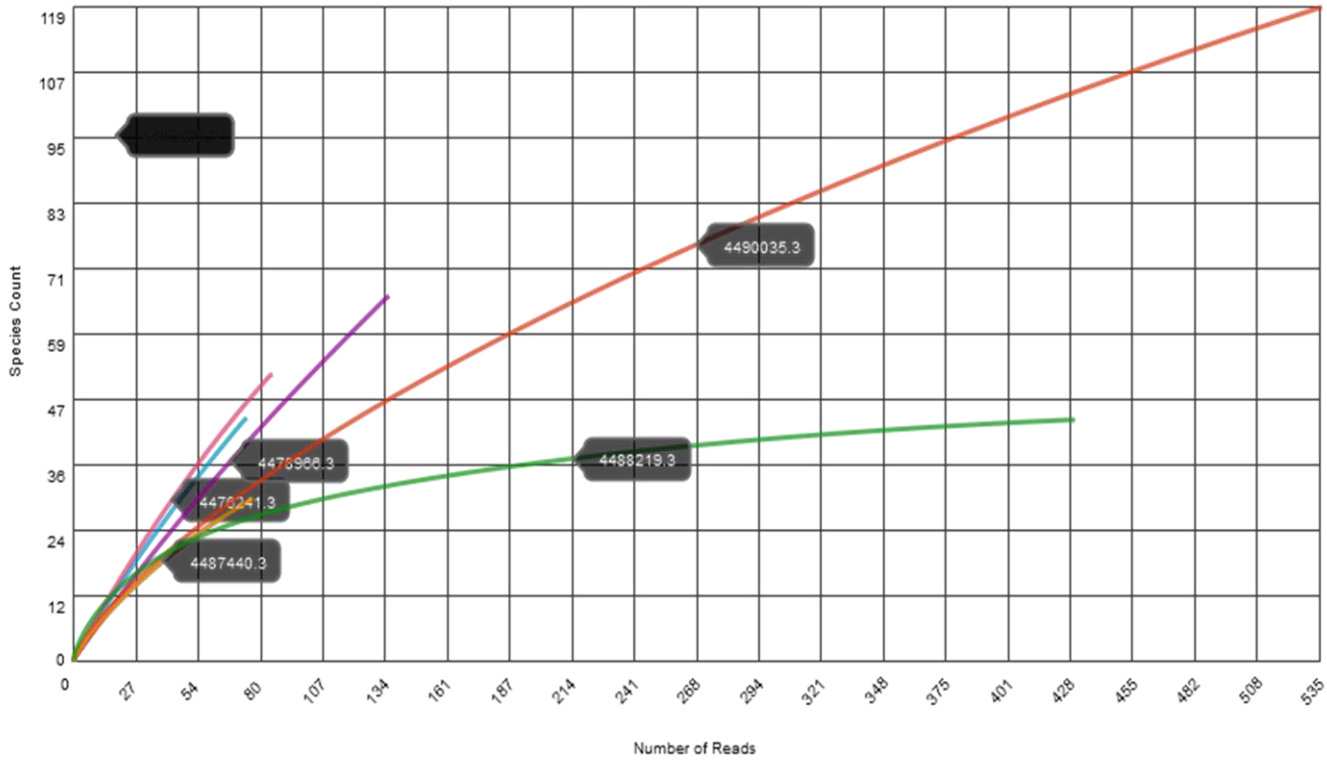

Figura 16. Curvas de rarefação para a riqueza estimada de espécies dos sete solos comparados, obtidas pela ferramenta Web MG-Rast.

Curvas	Metagenoma	Diversidade
	SCC	32,68
	SCOL	16,39
	SCA	20,89
	SPA	20,89
	SCH	60,93
	SCLE	41,54
	SCG	49,37

Tabela 1. Valores de diversidade de espécies de cada metagenoma gerados pelo MG-RAST.

5. CONCLUSÃO

Através da análise dos resultados obtidos foi possível concluir que:

1. Foi encontrado um total de 13 filos, sendo eles: *Proteobacteria, Acidobacteria, Firmicutes, Verrucomicrobia, Actinobacteria, Bacteroidetes, Chloroflexi, Planctomycetes, Spirochaetes, Tenericutes, Nitrospirae, Cyanobacteria* e *Dictyoglomi.*
2. O Filo *Proteobacteria* estava presente em todas as amostras ambientais.
3. O presente trabalho permitiu a identificação dos filos *Spirochaetes, Tenericutes* e *Cyanobacteria*, que em trabalhos anteriores não haviam sido identificados.

6. CONSIDERAÇÕES FINAIS

- O tipo de atividade antrópica sobre o solo promoveu um padrão de arranjo entre os filos encontrados no presente trabalho.
- O filo *Proteobacteria* apresentou grande diversidade com muitas espécies, provavelmente por serem estrategistas “k”, em que adaptam-se com grande facilidade competindo melhor por nutrientes.

Recomendações à futuros trabalhos

- É de sugestão à trabalhos que venham a ser realizados com objetivos semelhantes o uso de um número maior de sequências para cada amostra, visando um resultado de maior qualidade e abrangência.

6. REFERÊNCIA

AMANN, R.I.; LUDWIG, W.; SHELEIFER, K-H. Phylogenetic identification and in situ detection of individual microbial cells without cultivation. **Microbiological Reviews**, v. 59, p. 143-169, 1995.

AMRITA, P.; LENWOOD, S.; NIKOS, C. K.; IVANOVA, N. CLAMS: A Classifier for Metagenomic Sequences. **Standards in Genomic Sciences,** v.5, p. 248-253, 2011.

ATLAS, R.M.; BARTHA, R. **Microbial Evolution and Biodiversity** : The origins of life. 4 ed. , Portland: Book News, 1999.

AXELROOD, P. E.; CHOW, M. L.; RADOMSKI, C. C.; MCDERMOTT, J. M.; DAVIES, J. Molecular characterization of bacterial diversity from British Columbia forest soils subjected to disturbance. **J. Microbiol**, v. 48, p. 655-674, 2002.

BACKOFEN, R.; GILBERT, D. Bioinformatics and Constraints. Handbook of Constraint Programming, Chapter 26. 2006.

BARNS, S.M.; CAIN, E.C.; SAMMERVILLE, L.; KUSKE, C.R. Acidobacteria phylum sequences in uraniumcontaminated sediments greatly expand the known diversity within the phylum. **Applied Environmental Microbiology**, v.73, n.9, p.3113-3116, 2007.

BHATT, D.; VIRAL, B.; PRAKASH, G.; SUBHASH, J.; DHRAMSHI, N.; NAURIYAL, S.; ANJU, P.; CHAITANYA, G. Milk microbiome signatures of subclinical mastitis-affected cattle analyzed by shotgun sequencing. **Journal of Applied Microbiology**, p 1-32, 2012.

BRITO, E.M.S.; GUYONEAUD, R.; GÕNI-URRIZA, M.; RANCHOU-PEYRUSE, A.; VERBAERE, A.; CRAPEZ, M.A.C.; WASSERMAN, J.C.A.; DURAN, R. Characterization of hydrocarbonoclastic bacterial communities from mangrove sediments in Guanabara Bay, Brazil. **Res. Microbiol**., v. 157, p. 752-762, 2006.

BRITO, G. C. B.; SOUZA, D. B.; VASCONCELOS, F. C. W.; BRAGA, L. C. A importância da bioprospecção de microrganismos em áreas contaminadas com produtos derivados do petróleo. **Revista em Agronegócios e Meio Ambiente**, v.3, n.3, p. 291-310, 2010.

BUCKLEY, D. H. ; VARISA, H.; TYRRELL, A.; NELSON, A.; RUMBERGER, A.; THIES, J. E. Diversity of *Planctomycetes* in Soil in Relation to Soil History and Environmental Heterogeneity. **Appl Environ Microbiol**, v. 72, n. 7, p. 4522-4531**, 2006.**

CANHOS, V.P.; VAZOLLER, R.F. Diversidade no domínio bactéria. In: JOLY, C.A.; BICUDO, C.E.M. **Biodiversidade do Estado de São Paulo, Brasil**: síntese do conhecimento ao final do século XX. São Paulo: Fapesp, 1997. p.1-13.

CARVALHO, C. M.; MEDEIROS, S. R. B.; LIMA, L. F. A.; BLAHA, C. A. G. **Detecção de procariotos redutores de sulfato (prs) em amostras ambientais do manguezal de diogo lopes, bacia petrolífera potiguar**. 4°PDPETRO, 2007.

COELHO, M.R.; FIDALGO, E.C.C.; , ARAÚJO, F. DE O.; MENDONÇA, S. M.L. Solos das áreas-piloto do Projeto BiosBrasil, Município de Benjamim Constant, Estado do Amazonas Rio de Janeiro: **Embrapa Solos**. — (Boletim de pesquisa e desenvolvimento/Embrapa Solos, ISSN 1678-0892; 67).

COLE, J. R.; WANG, Q.; CARDENAS, E.; FISH, J.; CHAI, B.; FARRIS, R.J.; KULAM-SYED-MOHIDEEN, A.S.; MCGARRELL, D.M.; MARSH, T.; GARRITY, G.M.; TIEDJE, J.M.. The Ribosomal Database Project: improved alignments and new tools for rRNA analysis. **Nucleic Acids Research**, v. 37, p.141–145, 2009.

COPLEY, J. Ecology goes underground. **Nature**, London, v. 406, p. 452-454, 2000.

CORNE, D.W.; FOGEL, G.B. An Introduction to Bioinformatics for Computer Scientists. In: FOGEL, G.B., CORNE, D.W. (eds.) **Evolutionary Computation in Bioinformatics**, 2004, p. 3–18.

COURADEAU, E.; BENZERARA, K.; MOREIRA, D.; GE´RARD, E.; KAZ´MIERCZAK, J.; TAVERA, R. Prokaryotic and Eukaryotic Community Structure in Field and Cultured Microbialites from the Alkaline Lake Alchichica (Mexico). **PLoS ONE**, v. 6, n. 12, p. 1-16, 2011.

COWAN, D.; MEYER, Q.; STAFFORD, W.; MUYANGA, S.; CAMERON, R.; WITTWER, P. Metagenomic gene discovery: Past, present and future. **Trends in Biotechnology**, v.23, p.321–329, 2005.

COWAN, D.A. Microbial genomes-the untapped resource. **Trends Biotechnology**, v.18, p. 14-16, 2000.

CUNHA, C. D.; LEITE, S. G. F.; ROSADO, A. S.; ROSÁRIO, M. **Biorremediação de água subterrânea contaminada com gasolina e análise molecular da comunidade bacteriana presente**. Rio de Janeiro, RJ: CETEM/MCT, 2008.

CURY, J. C.; ARAUJO, F. V.; COELHO-SOUZA, S. A.; PEIXOTO, R. S.; OLIVEIRA, J. A. L. Microbial Diversity of a Brazilian Coastal Region Influenced by an Upwelling System and Anthropogenic Activity. **PLoS ONE**, v. 6, n. 1, p. 1-13, 2011.

DANIEL, R. The Metagenomics of soil. **Nature Reviews /Microbiology**, v. 3, p. 470-478, 2005.

De Fede K. L.; Panaccione D. G.; Sexstone A. J. Characterization of dilution enrichment cultures obtained from size-fractionated soil bacteria by BIOLOGR community-level physiological profiles and restriction analysis of 16S rRNA genes. **Soil Biol Biochem**, *v. 33, p. 1555-1562*, 2001.

DUPRE J.; O'MALLEY M. A. Metagenomics and biological ontology. **Studies in History and Philosophy of Biological and Biomedical Sciences**, p. 834–846, 2007.

FOGEL, G. B. Internet resources for bioinformatics data and tools. IN: FOGEL, G. B. & CORNE, D. W. (Eds.) **Evolutionary computation in bioinformatics**. Amsterdam, 2003, p. 367-372.

GARRITY, G.M.; BELL, J.A.; LILBURN, T.G. **Bergey's manual of systematic**: bacteriology Taxonomic outline of the Prokaryotes. 2.ed. New York: Springer, 2004.

GHINI, R.; ZARONI, M. M. H. Relação entre coberturas vegetais e supressividade de solos a Rhizoctonia solani. **Fitopatologia Brasileira**, v. 26, n. 01, p. 10-15, 2001.

GOULD S.J. Full House: The Spread of Excellence from Plato to Darwin. **Harmony Books**, p. 244, 1996.

HANDELSMAN J. Metagenomics: application of genomics to uncultured microorganisms. **Microbiology and molecular biology reviews**, v. 68, n. 4, p. 669-85, 2004.

HANDELSMAN, J.; RONDON, M. R.; BRADY, S. F.; CLARDY, J.; GOODMAN, R. M. Molecular biological access to the chemistry of unknown soil microbes: a new frontier for natural products. **Chem Biol**, v. 5, p. 245–249, 1998.

HEALY, F.G.; RAY, R.M.; ALDRICH, H.C.; WILKIE, A.C.; INGRAM, L.O.; SHANMUNGAM, K.T. Direct isolation of functional genes enconding celulases from the microbial consortia in a thermophilic, anaerobic digester maintained on lignocellulose. **Appl. Microbiology Biotechnol,** v. 43, p. 667-674, 1995.

HIRASHI, A.; KISHIMOTO, N.; KOSAKO, Y.; WAKAO, N.; TANO, T. Philogenetic position of the menaquinone-containing acidophilic chemoorganotroph *Acidobacterium capsulatum.* **FEMS Microbiology Letters**, v.132, p.91-94, 1995.

HOLGER, D.; JEPPE, L.; NIELSEN, P. H.; SCHLEIFER, N.; WAGNER, M. In Situ Characterization of *Nitrospira*-Like Nitrite-Oxidizing Bacteria Active in Wastewater Treatment Plants. **App Environ Microbiol** , v.67, n. 11, p. 5273-5284, 2001.

HUGENHOLTZ, P.; GOEBEL, B.M.; PACE, N.R. Impact of culture-independent studies on the emerging phylogenetic view of bacterial diversity. **Journal of Bacteriology,** v. 180, p. 4765-4774, 1998.

JESUS, E. C.; MARSH, T. L.; TIEDJE, J. M.; MOREIRA, F. M. S. Changes in land use alter the structure of bacterial communities in Western Amazon soils. **The ISME Journal,** p. 1-8, 2009.

JUCK, D.; CHARLES, T.; WHYTE, L.; GREER, C. Polyphasic microbial community analysis of petroleum hydrocarbon-contaminated soils from two northern canadian communities. **FEMS Microbiol. Ecol,** v. 33, p. 241–249, 2000.

KAVITA, S.; KAKIRDE, LARISSA C.; PARSLEY, MARK R. LILES. Size does matter: Application-driven approaches for soil metagenomics. **Soil Biology & Biochemistry**, v. 42 p. 1911-1923, 2010.

KENNEDY, A.C. Bacterial diversity in agroecosystems. **Agriculture, Ecosystems and Environment**, v. 74, p. 65-76, 1999.

KIM, B.S.; KIM, S.Y.; PARK, J.; PARK, W.; HWANG, K.Y.; YOON, Y.J.; OH, W.K.; KIM, B.Y.; AHN, J.S. Sequence-based screening for self-sufficient P450 monooxygenase from a metagenome library. **Journal of Applied Microbiology**, v.102, p. 1392-1400, 2007.

KIMURA, N. Metagenomics: Acess to unculturable Microbes in the Environment. **Microbes Environments**, vol. 21, n. 4, 201-215, 2006.

KIRCHMAN, D. L. The ecology of Cytophaga-Flavobacteria in aquatic environments. **FEMS Microbiology Ecology**, v. 39, n. 2, p. 91-100, 2002.

KIRKA J. L.; Beaudettea; Miranda H.; Peter M.; John N.; Klironomosb, H. L.; Jack T. T. Methods of studying soil microbial diversity. **Journal of Microbiological Methods**, v. 58, p. 169– 188, 2004.

KRAGELUND, C.; LEVANTESI, C.; BORGER, A.; THELEN, K.; EIKELBOOM, D.; TANDOI, V.; KONG, Y. VAN DER.; WAARDE, J.; KROONEMAN, J.; ROSSETTI, S.; THOMSEN, T. R.; NIELSEN, P. H. Identity, abundance and ecophysiology of filamentous Chloroflexi species present in activated sludge treatment plants. **FEMS Microbiol Ecol,** v. 59, n. 3, p. 671-82, 2007.

KRAUSE, L.; DIAZ, N. N.; GOESMANN, A.; KELLEY, S.; NATTKEMPER, T. W.; ROHWER, F.; EDWARDS, R. A.; STOYE, J. Phylogenetic classification of short environmental DNA fragments. **Nucleic Acids Res**, v. 36, p. 2230–2239, 2008.

KUHNER, C.H.; MATTHIES, C.Clostridium akagii sp. nov. and Clostridium acidisoli sp. nov.: acid-tolerant, N(2)-fixing clostridia isolated from acidic forest soil and litter. **Int J. Syst. Evol. Microbiol,** v. 50, p. 873-881, 2000.

KUNIN, V.; HUGENHOLTZ, P. PYROTAGGER. A fast, accurate pipeline for analysis of rRNA amplicon pyrosequence data. **The Open Journal**, n.1, p. 1-8, 2010.

LANCINI, G.; PARENTI, G.; GALLO, G. G. **Antibiotics: A Multidisciplinary Approach**. 1. Ed. Springer, 1995. 288 p.

LIOLIOS, K.; TAVERNARAKIS, N.; HUGENHOLTZ, P.; KYRPIDES, N. C. The Genomes On Line Database (GOLD) v.2: a monitor of genome projects worldwide. **Nucleic Acids Research**, v. 34, p. 1-3, 2006.

LOURENTE, E.; MERCANTE F. M.; ALOVISI A. M. T.; GOMES C. F.; GASPARINI A. S.; NUNES M. C. Atributos Microbiológicos, químicos e físicos de solo sob diferentes sistemas de manejo e condições de cerrado. **Pesq. Agropec. Trop**., Goiânia, v. 41, n. 1, p. 20-28, jan./mar. 2011.

MARKOWITZ, V. M.; IVANOVA, N. N.; SZETO, E.; PALANIAPPAN, K.; CHU, K.; DALEVI, D.; CHEN, A.; GRECHKIN, Y.; DUBCHAK ; INNA, A. I.; LYKIDIS, A.; MAVROMATIS, K.; HUGENHOLTZ, P.; KYRPIDES, N. C. IMG/M: a data management and analysis system for metagenomes. **Nucleic Acids Research**, v. 36, p. 1-5, 2008.

MASULLI, F.; MITRA, S. Natural computing methods in bioinformatics: A survey. **Information Fusion**, p. 211–216, 2009.

MEYER, F.; PAARMANN, D.; D'SOUZA, M.; OLSON, R.; GLASS, E. M.; KUBAL, M.; PACZIAN, T.; RODRIGUEZ, A.; STEVENS, R.; WILKE, A.; WILKENING, J.; EDWARDS, R. A. The metagenomics RAST server – a public resource for the automatic phylogenetic and functional analysis of metagenomes. **BMC Bioinformatics**, p. 1-8, 2008.

MIQUELETTO, P. B.; ANDREOTE, F. D.; DIAS, A. C. F.; FERREIRA, J. C.; NETO, E. V.; OLIVEIRA, V. M. Cultivation-independent methods applied to the microbial prospection of oil and gas in soil from a sedimentar basin in Brazil. **AMB Express**, v.1, n. 35, p. 1-16, 2011.

MITRA, S.; KLAR, B.; HUSON, D. H. Visual and statistical comparison of metagenomes. **Bioinformatics**, v.25, n. 15, p. 1849-1855, 2009.

MOKILI J. L.; ROHWER F.; DUTILH B. E. Metagenomics and future perspectives in virus Discovery**. Current Opinion in Virology**, v. 2, p. 1–15, 2012.

MORENO, C. Métodos para medir la biodiversidad. **Rev. biol. Trop**, vol.49, n.3-4, pp. 1300-1302 . 2001.

NISHIDA, H.; BEPPU, T.; UEDA, K. Whole-genome comparison clarifies close phylogenetic relationships between the phyla Dictyoglomi and Thermotogae. **Genomics**, v. 98, n. 5, p. 370-375, 2011.

ONDOV, B. D.; BERGMAN, N. H.; PHILLIPPY, A. M. Interactive metagenomic visualization in a Web Browser. **BMC Bioinformatics,** v. 12, n. 385, p. 1-9, 2011.

OREN, A. A proposal for further integration of the cyanobacteria under the Bacteriological Code. **Int J Syst Evol Microbiol**, v. 54, n. 5, p. 1895-1902, 2004.

PAIXÃO, D. A. A.; DIMITROV, M. R.; PEREIRA, R. M.; ACCORSINI, F. R.; VIDOTTI, M. B.; LEMOS, E.G.M. Molecular analysis of the bacterial diversity in A specialized consortium for diesel oil degradation. **R. Bras. Ci. Solo,** v. 34, p. 773-781, 2010.

PEDRINHO, E.A.N..; LEMOS, E.G.M..; PEREIRA, R.M..; SCAQUITTO, D.C..; SILVEIRA, É.L. DA.; VAL-MORAES, S.P..; ALVES, L.M.C..; WICKERT, E.; VALARINI, M.J.. Avaliação do impacto do lodo de esgoto na microbiota do solo utilizando o gene 16S rRNA. **Arq. Inst. Biol**., v.76, n.3, p.443-448, 2009.

PEREIRA, R. M.; SILVEIRA, E. L.; SCAQUITTO, D. C.; PEDRINHO, E. A. N.; VAL-MORAES, S. P.; WICKERT, E.; CARARETO-ALVES, L. M.; LEMOS, E. G. M. Molecular characterization of bacterial populations of different soils. **Brazilian Journal of Microbiology,** v. 37, p. 439-447, 2006.

PISA, G.; MAGNANI, G.S.; WEBER, H.; SOUZA, E.M.; FAORO, H.; MONTEIRO, R.A.; DAROS, E.; BAURA, V.; BESPALHOK,J.P.; PEDROSA, F.O.; CRUZ. L.M. Diversity of 16S rRNA genes from bacteria of sugarcane rhizosphere Soil. **Brazilian Journal of Medical and Biological Research**, v. 44, n. 12, p. 1215-1221, 2011.

PS, U.S. Penn State University. **Chlorofexi.** Disponível em: http://bmb-it-services.bmb.psu.edu/bryant/lab/Project/Chloroflexi/index.html . Acesso em: 10 de agosto de 2012.

RAIJ, B. V. et al. Análise química para avaliação da fertilidade de solos tropicais. São Paulo: **Instituto Agronômico**, 284p, 2001.

REICHENBACH H.; HÖFLE G. Myxobacteria as producers of secondary metabolites. **Drug Discovery from Nature**. Berlin, Springer, p. 79-149, 1999.

RIESENFELD C. S.; SCHLOSS P. D.; HANDELSMAN J. Metagenomics: genomic analysis of microbial communities. **Annu Rev Genet**, v. 38, p. 525-552, 2004.

ROBE P.; NALIN R.; CAPELLANO C.; VOGEL T. M.; SIMONET P. Extraction of DNA from soil. **European Journal of Soil Biology**, v. 19, p. 183–190, 2003.

ROBERTSON, D. E.; CHAPLIN, J. A.; DESANTIS, G.; PODAR, M.; MADDEN, M.; CHI, E.; RICHARDSON, T.; MILAN, A.; MILLER, M.; WEINER, D. P. Exploring nitrilase sequence space for enantioselective catalysis. **Appl Environ Microbiol,** v.70, p. 2429-2436, 2004.

RONDON, M. R.; GOODMAN, R. M.; HANDELSMAN, J. The Earth's bounty: assessing and accessing soil microbial diversity. **Trends in Biotechnology**, v.17, p. 403-410, 1999.

ROSSELÓ-MORA R.; AMANN R. The species concept for prokaryotes. **FEMS Microbiology Review**, Amsterdam, v. 25, n. 1, p. 39-67, 2001.

SAIT, M.; DAVIS, K. E. R.; JANSSEN, P. H. Effect of pH on isolation and distribution of members of subdivision 1 of the phylum Acidobacteria occurring in soil. **Applied and Environmental Microbiology**, v. 72, n. 03, p. 1852–1857, 2006.

SANGWAN, P.; KOVAC, S.; DAVIS, K.; SAIT, M.; JANSSEN, P. H. Detection and Cultivation of Soil Verrucomicrobia. **Appl. Environ. Microbiol**, v. 71, n. 12, p. 1-10, 2005.

SCHLOSS P. D.; HANDELSMAN J. Biotechnological prospects from metagenomics. **Current Opinion Biotechnology**, v. 14, p. 303-310, 2003.

SESHADRI, R.; KRAVITZ, S. A.; SMARR, L.; GILNA, P.; FRAZIER, M. CAMERA: A Community Resource for Metagenomics. **PLoS Biology**, v.5, p.394-397, 2007.

SILVEIRA A.; FREITAS S. **Microbiota do solo e qualidade ambiental**. Campinas. Instituto Agronômico, p. 312, 2007.

SILVEIRA, E.L.; PEREIRA, R.M.; SCAQUITTO, D.C.; PEDRINHO, E.A.N.; VAL-MORAES, S.P.; WICKERT, E.; CARARETO-ALVES, L.M. & LEMOS, E.G.M. Bacterial diversity of soil under eucalyptus assessed by 16S rDNA sequencing analysis. **Pesq. Agropec. Bras**., v. 41, p. 1507-1516, 2006.

TORSVIK V.; GOKSOYR, J.; DAAE F. L. High diversity in DNA of soil bacteria. **Appl. Environ. Microbiol**, v. 56, p. 782–787, 1990.

TORSVIK, V.; OVREAS, L. Microbial diversity and function in soil: from genes to ecosystems. **Ecology and industrial microbiology**, v. 5, p. 240–245, 2002.

VAL-MORAES, S.P.; VALARINI, M.J.; GHINI, R.; LEMOS, E.G.M.; CARARETO-ALVES, L. M. Diversidade de bactérias de solo sob vegetação natural e cultivo de Hortaliças. **Revista Ciência Agronômica**, v. 40, n. 1, p. 7-16, 2009.

VAN TRAPPEN, S.; VANDECANDELAERE, I.; MERGAERT, J.; SWINGS, J. Flavobacterium degerlachei sp. nov., Flavobacterium frigoris sp. nov. and Flavobacterium micromati sp. nov., novel psychrophilic bacteria isolated from microbial mats in Antarctic lakes. **International Journal of Systematic and Evolutionary Microbiology,** v. 54, p. 85–92, 2004.

WARD, N. New directions and interactions in metagenomics. **FEMS Microbial Ecology**, v. 55, p. 331–338, 2006.

WARD, N.; FRASER, C. M. How genomics has affected the concept of microbiology. **Current Opinion in Microbiology**, v. 8, p. 564–571, 2005.

WATSON, J.D. The human genome project: past, present, and future. **Science**, v. 248, p. 44–49, 1990.

WOOLEY, J. C.; GODZIK, A.; FRIEDBERG, I. A Primer on Metagenomics. **PLoS Comput Biol,** v. 6, n. 2, p. 1-13, 2010.

WU, S.; ZHU, Z.; FU, L.; NIU, B.; LI, W. WebMGA: a customizable web server for fast metagenomic sequence analysis. **BMC Genomics,** v. 12, n. 444, p. 1-9, 2011.

ZAK, J.C.; WILLIG, M.R.; MOORHEAD, D.L.; WILDMAN, H.G. Functional diversity of bacterial communities: a quantitative approach. **Soil Biology and Biochemistry**, v.26, p. 1101-1108, 1994.