**UNIVERSIDADE DE SÃO PAULO
DEIBS BARBOSA**

**ANÁLISE COMPUTACIONAL DA VARIAÇÃO DO
POTENCIAL METABÓLICO MICROBIANO EM
METAGENOMAS**

**São Paulo, SP
2015**

**UNIVERSIDADE DE SÃO PAULO
DEIBS BARBOSA**

**ANÁLISE COMPUTACIONAL DA VARIAÇÃO DO
POTENCIAL METABÓLICO MICROBIANO EM
METAGENOMAS**

Tese apresentada ao Programa Interunidades de Pós-Graduação em Bioinformática da Universidade de São Paulo como parte dos requisitos para a obtenção do título de Doutor.
Área de concentração: Bioinformática.

Orientador: Prof. Dr. João Carlos Setubal
Co-orientadora: Profa. Dra. Aline Maria da Silva

**São Paulo, SP
2015**

“(...) claro que quando chegar ao fim do meu passeio saberei mais, mas também é certo que saberei menos, precisamente por mais saber, por outras palavras, a ver se me explico, a consciência de saber mais conduz-me à consciência de saber pouco, aliás, apetece perguntar, que é saber, (...)”
José Saramago, História do Cerco de Lisboa

.

AGRADECIMENTOS

Agradeço aos meus orientadores, professor Setubal e professora Aline, pela acolhida generosa, pela integração imediata à rotina e aos projetos dos grupos, pela liberdade na atividade de pesquisa, e acima de tudo pelo incentivo profissional durante o encerramento dessa etapa.

Agradeço a minha família pelo imensurável e indispensável apoio.

Agradeço aos integrantes do Laboratório de Bioquímica e Biologia Molecular de Microrganismos, a quem estava saindo quando eu estava chegando e a quem estava chegando quando eu estava saindo, pela companhia e pelo companheirismo e acima de tudo por me lembrar que ‘a Termodinâmica é a alma do negócio’...

Agradeço aos integrantes do Laboratório de Bioinformática, onde tive a oportunidade de esclarecer uma série de dúvida sobre bits, bytes e porque não posso ocupar 2 Tb no *storage*...

Agradeço aos professores, em particular Cristina V. Niero e Julio Cezar F. de Oliveira, e alunos do Laboratório de Interações Microbianas da Unifesp – Campus Diadema pelo auxílio na preparação de amostras e pelas sugestões.

Agradeço aos Drs. Paulo Magalhães Bressan e João Batista da Cruz, diretores da Fundação Parque Zoológico de São Paulo. Agradeço também à equipe do parque pela solicitude e valioso auxílio nas diversas coletas.

Agradeço aos demais integrantes do Projeto Metazoo, companheiros de trabalho que muito acrescentaram a minha tese... Luciana Principal Antunes, Layla Farage Martins, Prof. Sergio Verjovski-Almeida, Prof. Luciano Digiampietri, Prof. Ronaldo Quaggio, Gianluca Major, Leandro Lemos, Deyvid Amgarten, Andrew Thomas, Felipe Prata, Raquel Riyuzo e Roberta Verciano.

Agradeço à equipe da Pós-Graduação em Bioinformática, principalmente à Patrícia Martorelli, e aos Profs. Alan Durham e Ricardo Vencio.

Este trabalho foi realizado com apoio financeiro da FAPESP. O autor recebeu bolsa de Doutorado da CAPES.

RESUMO

Este trabalho teve como objetivo analisar, através de uma abordagem metagenômica/computacional, a variação do conteúdo gênico e do potencial metabólico das comunidades microbianas associadas a dois ambientes da Fundação Parque Zoológico de São Paulo: um reservatório artificial, o Lago São Francisco e o sistema de compostagem de resíduos do parque. Para o estudo do primeiro ambiente, amostras foram coletadas mensalmente no Lago São Francisco de outubro de 2012 a setembro de 2013 e submetidas ao sequenciamento metagenômico. Esse estudo mostrou que agrupamentos de amostras de uma mesma estação são estatisticamente suportados. As coocorrências de espécies, com suporte estatístico alto, foram estabelecidas e representadas em uma rede separada em 60 grupos ou módulos. A maioria dos grandes módulos foram formados quase exclusivamente por espécies do mesmo filo, indicando possíveis mecanismos de resposta a fatores ambientais e à presença de nutrientes ao invés da pura interação entre espécies. Os fatores que levaram à variação dos táxons também influenciaram o potencial metabólico da comunidade: embora os módulos metabólicos de forma geral estivessem distribuídos ao longo dos meses, alguns se destacaram pela variação intensa, principalmente na amostra de novembro de 2012. Para o estudo do segundo ambiente, foram analisados dois conjuntos de dados de sequenciamento metagenômico gerados a partir de amostras seriadas coletadas ao longo do processo de compostagem. Nos dias iniciais do processo de compostagem, os mais discrepantes, houve uma super-representação de módulos metabólicos relacionados principalmente ao metabolismo de carboidratos e à síntese de aminoácidos. Em conjunto, os dados obtidos nesse estudo indicaram uma comunidade microbiana com potencial metabólico variando direcionalmente em função dos compostos inicialmente presentes ou oriundos do substrato, em função da presença de oxigênio e em função da etapa do processo.

Palavras-chave: Metagenoma – Potencial metabólico – Lago São Francisco – Compostagem

ABSTRACT

This project aimed to analyze, through a computational approach, the change of gene content and metabolic potential in metagenomes from two microbiomes from São Paulo Zoo Park Foundation: an artificial reservoir, Lake São Francisco, and the compost systems in the park. For the study on the first microbiome, samples were monthly collected in Lake São Francisco from October 2012 to September 2013 and submitted to metagenomic sequencing.

This study showed that clustering of the samples from a same season is statistically supported. The species co-occurrences, with high statistical support, were established and represented in a network composed by 60 groups or modules. The most biggest modules were formed almost exclusively by species of the same phylum, pointing possible response mechanisms to environmental factors and to presence of nutrients instead of the simple species interaction. The factors which lead to taxa variation also influenced the metabolic potential of the community: although the metabolic modules generally are spreaded through the months, some highlighted because the intense variation, mainly in the sample from November 2012. For the latter environment, two metagenomic datasets were analyzed from serial samples collected throughout the composting process. In the initial days of composting process, the most discrepant, there was an uprepresentation of metabolic modules related to carbohydrates metabolism and amino acids synthesis. All data in this study indicated a microbial community with metabolic potential varying according to nutrients, oxygen presence and process stage.

Keywords: Metagenome – Metabolic potential – Lake São Francisco – Composting

SUMÁRIO

CAPÍTULO 1:	10
INTRODUÇÃO GERAL	10
Diversidade Microbiana e o Desenvolvimento da Metagenômica	10
Rotas Metabólicas e Reconstrução a partir de Dados Metagenômicos	13
O Projeto Metazoo	16
OBJETIVO.....	17
REFERÊNCIAS.....	18
CAPÍTULO 2:	23
ANÁLISE COMPUTACIONAL DA COMUNIDADE MICROBIANA NO LAGO SÃO FRANCISCO DA FUNDAÇÃO PARQUE ZOOLOGICO DE SÃO PAULO	23
INTRODUÇÃO	23
METODOLOGIA.....	26
Visão Geral da Sequência de Análises	26
Extração de DNA.....	27
Preparação das Bibliotecas de DNA	27
Análise da Distribuição Taxonômica e das Diversidades α e β das Amostras do Lago São Francisco	29
Análise de Similaridade Local e Construção da Rede de Correlação	30
Identificação e Abundância de e Módulos Metabólicos usando o Programa HUMAnN	31
RESULTADOS E DISCUSSÃO.....	34
Sequenciamento de <i>Amplicons</i> do Gene rRNA 16S e Sequenciamento <i>Shotgun</i> das Amostras do Lago São Francisco	34
Distribuição Taxonômica nas Amostras do Lago São Francisco através do Sequenciamento de <i>Amplicons</i> do Gene rRNA 16S	35
Diversidades α e β das Amostras do Lago São Francisco através de <i>Amplicons</i> do Gene rRNA 16S	44
Coocorrência e Correlação de Espécies nas Amostras do Lago São Francisco através de Sequenciamento <i>Shotgun</i>	56

Análise do Potencial Metabólico do Lago Francisco através de Sequenciamento <i>Shotgun</i>	62
REFERÊNCIAS	68
APÊNDICE 1	74
CAPÍTULO 3:	82
ANÁLISE COMPUTACIONAL DO PERFIL METABÓLICO DO SISTEMA DE COMPOSTAGEM DA FUNDAÇÃO PARQUE ZOOLOGICO DE SÃO PAULO	82
INTRODUÇÃO	82
METODOLOGIA	84
Descrição dos Dois Conjuntos de Dados Metagenômicos	84
RESULTADOS E DISCUSSÃO	86
Variação do Potencial Metabólico no Processo de Compostagem	86
REFERÊNCIAS	98
APÊNDICE 2	101
APÊNDICE 3	108
CAPÍTULO 4:	117
CONSIDERAÇÕES FINAIS E PERSPECTIVAS	117
REFERÊNCIAS	119

CAPÍTULO 1:

INTRODUÇÃO GERAL

Diversidade Microbiana e o Desenvolvimento da Metagenômica

Até recentemente os mais antigos fósseis conhecidos, objetos morfológicamente semelhantes a algumas cianobactérias atuais, eram datados com tendo 3,5 bilhões de anos (SCHOPF, 1993; SCHOPF *et al.*, 2002). Um estudo de 2003, utilizando técnicas como microscopia óptica e eletrônica, análise de imagem digital e espectroscopia micro Raman, reinterpreto esses microfósseis como sendo artefatos secundários formados de grafite amorfo em múltiplas gerações de sílex e vidro vulcânico de circulação hidrotermal metalífera (BRASIER *et al.*, 2002). Uma nova evidência vinda de rochas do éon Arqueano indica a existência de fósseis do que teria sido um microrganismo com metabolismo baseado em enxofre (WACEY *et al.*, 2011). Nesse estudo microestruturas esféricas e elipsoidais exibiram características possivelmente biológicas como lúmens, parede celular carbonácea rica em nitrogênio, degradação tafonômica e organização em cadeias e aglomerados. Os cristais de pirita ao redor das microestruturas seriam um subproduto metabólico de rotas de redução de sulfato ou dismutação de enxofre sugerindo que os primeiros ecossistemas microbianos podem ter sido baseados no metabolismo do enxofre.

Foram as atividades microbianas ocorridas entre 2,7 e 2,4 bilhões de anos que resultaram nas proporções de isótopos estáveis de elementos como carbono, ferro e enxofre encontradas em rochas desde esse período (HAYES & WALDBAUER, 2006; ARCHER & VANCE, 2006). Por volta de 2,7 bilhões de anos aconteceu também a diversificação de cianobactérias associada ao Grande Evento de Oxidação, o aumento considerável da concentração de oxigênio molecular na atmosfera mudando o dramaticamente o curso da vida no planeta (CAVALIER-SMITH *et al.*, 2006; SCHIRMEISTER *et al.*, 2013).

Procariotos são caracterizados pela ausência de compartimentos celulares formados por sistemas de endomembranas com exceção de alguns organismos do

superfilo Planctomycetes-Verrucomicrobia-Chlamydiae (LEE *et al.*, 2009; FUERST *et al.*, 2012). Procariotos incluem os domínios Archaea e Bacteria (WOESE *et al.*, 1990) originalmente diferenciados com base em diferenças no gene rRNA 16S. A relação exata entre os três domínios da vida ainda é objeto de investigação. Questões como, por exemplo, a posição da raiz da árvore da vida ainda é matéria de debate. Fusão de genomas e transferência horizontal de genes tornam um tanto difusa a história evolutiva registrada em sequências gênicas. Em 2004, Rivera e Lake propuseram a origem dos primeiros genomas eucarióticos como resultado de uma fusão de dois genomas procarióticos diversos, um microrganismo fotossintetizante do domínio Bacteria e um microrganismo do domínio Archaea, relacionando os três domínios inicialmente não numa árvore, mas em um anel da vida (RIVERA & LAKE, 2004).

As estimativas atuais indicam que aproximadamente 30 a 50% da biomassa do planeta é constituída por microrganismos distribuídos por uma vasta gama de grupos taxonômicos ocupando todos os ecossistemas já explorados (WHITMAN *et al.*, 1998; TORSVIK *et al.*, 2002).

Nos primórdios da Microbiologia, a identificação dos organismos se dava pelo cultivo laboratorial seguido de múltiplos testes bioquímicos. A comparação do número total de células, estimado por microscopia, e do número de células cultiváveis mostrou uma grande discrepância conhecida como Grande Anomalia da Contagem em Placa (*Great Plate Count Anomaly*) (AMANN *et al.*, 1995). Essa constatação trazia à tona a necessidade de novos métodos de investigação da diversidade microbiana.

As abordagens que seguiram ao cultivo laboratorial se baseavam em informação de marcadores genéticos, usualmente o gene rRNA 16S, para fazer inferências acerca da diversidade (HEAD *et al.*, 1998; MLOT, 2004). Com o advento de técnicas moleculares, obstáculos conceituais e metodológicos que por décadas limitaram o conhecimento acerca das comunidades microbianas foram parcialmente superados, possibilitando a melhor compreensão e uma visão sistêmica da diversidade microbiana e de sua distribuição na biosfera (HEAD *et al.*, 1998; GREEN & BOHANNAN, 2006). Considerando esse panorama, foi proposto em 2010 o projeto norte-americano “The Earth Microbiome” que pretende sequenciar genomas e metagenomas de um grande número de biomas (GILBERT *et al.*, 2010). Além

dessa iniciativa, outras propostas que podem ser citadas são o amplo estudo de 1000 amostras de solo do território da Grã-Bretanha, que resultou no primeiro mapa de distribuição de populações bacterianas de toda a ilha (GRIFFITHS *et al.*, 2011) e o estudo de 98 amostras de solo coletadas ao longo das Américas (FIERER & JACKSON, 2006), os quais sugerem que as diferenças de diversidade e riqueza das comunidades bacterianas do solo podem ser explicadas pelas variações de pH. Mais recentemente foi elaborado o Projeto Microbioma Brasileiro que visa coletar e comparar, de uma forma sistemática, informações sobre a diversidade microbiana em diversos ambientes (PYLRO *et al.*, 2014).

Além de possibilitar uma melhor compreensão da biologia dos microrganismos sob perspectivas evolutiva e ecológica (PACE, 1997; RAPPE & GIOVANNONI, 2003; GREEN & BOHANNAN, 2006), o conhecimento da biodiversidade microbiana tem contribuído para estudos relacionados à preservação e biorremediação, e também para a descoberta de novas biomoléculas (LORENZ & ECK, 2005; SCHMEISSER *et al.*, 2007; SINGH, 2010; SINGH & MACDONALD, 2010).

Atualmente são inquestionáveis a importância e a contribuição da metagenômica na identificação e caracterização de comunidades naturais de microrganismos e na descoberta de novos produtos com aplicação médica e industrial derivados desses microrganismos (BANIK & BRADY, 2010; CRAIG *et al.*, 2010; MOCALI & BENEDETTI, 2010; STEELE *et al.*, 2009). De forma geral, a metagenômica fornece dados em grande escala e permite o levantamento não só de microrganismos já descritos, como também de microrganismos desconhecidos e genes da comunidade microbiana.

O termo metagenoma surgiu em 1998 para designar os “genomas coletivos da microflora de solo” (HANDELSMAN *et al.*, 1998). Originalmente o metagenoma era extraído e clonado e os clones eram analisados em busca de fenótipos de interesse. Com o passar do tempo o termo passou a ser empregado em análises de sequências de DNA obtido de amostras ambientais. A metagenômica tradicional compreende as seguintes etapas: isolamento e purificação do DNA de uma amostra complexa; clonagem do DNA em um vetor apropriado; transformação dos clones recombinantes em uma célula hospedeira com a obtenção de uma biblioteca de DNA metagenômico; sequenciamento dos clones e análise por bioinformática e/ou

seleção dos clones em condições de cultivo que evidenciam uma determinada atividade enzimática (triagem funcional) (DONADIO *et al.*, 2002; RIESENFELD *et al.*, 2004; STREIT & SCHMITZ, 2004; YUN *et al.*, 2004). O marco divisor viria em 2004 com a publicação dos resultados de dois grandes projetos de sequenciamento de amostras ambientais (VENTER *et al.*, 2004; TYSON *et al.*, 2004).

Com o desenvolvimento e acesso facilitado a novas tecnologias de sequenciamento (*Next-Generation Sequencing* – NGS) que dispensam a laboriosa etapa de clonagem, ensaios funcionais passaram a ser cada vez mais raros e limitados a circunstâncias específicas. Finalmente, a metagenômica vem sendo empregada na produção de dados de grande importância para o estabelecimento do ramo da ciência chamada de “ecologia molecular”, campo que exige a integração de um grande número de dados variados, incluindo dados biogeoquímicos e geográficos (RAES *et al.*, 2011).

Com o avanço da tecnologia e projetos de sequenciamento cada vez maiores, surgem desafios computacionais relacionados ao processamento dos dados. O sequenciamento de amostras provenientes de comunidades microbianas complexas pode gerar montagens de *contigs* parciais imprecisas ou pouco informativas devido à cobertura incompleta, levando a uma limitação na atribuição dos fragmentos aos táxons já conhecidos ou na atribuição de função aos fragmentos (PRAKASH & TAYLOR, 2012). Além do gargalo de análise, outras questões figuram nas discussões técnicas, como a detecção eficiente de sequências quiméricas e réplicas artefatuais, estabelecimento da frequência de taxa ou famílias gênicas baseada na contagem de leituras (*reads*) e a reconstrução de rotas metabólicas presentes na amostra (WOOLEY & YE, 2009). Neste contexto, a pesquisa em Bioinformática tem sido cada vez mais relevante, incluindo novas estratégias para mineração da enxurrada de dados na expectativa de novas descobertas (EDWARDS *et al.*, 2002). Assim, a reconstrução de rotas metabólicas a partir de dados metagenômicos emerge como um caminho promissor na exploração dessa vastíssima informação.

Rotas Metabólicas e Reconstrução a partir de Dados Metagenômicos

Os ciclos dos elementos químicos geralmente são estudados separadamente, contudo eles coevoluíram e se influenciam mutuamente (FALKOWSKI *et al.*, 2008).

Rotas metabólicas surgem na história da evolução para utilizar substratos, presentes no ambiente ou produzidos por outros organismos, de duas formas distintas: por modificação de rotas preexistentes ou pela utilização de rotas preexistentes em sentido inverso. O acoplamento de reações e rotas não só possibilita a transposição de barreiras termodinâmicas como também possibilita que, em escala geológica de tempo, a biosfera estabeleça os ciclos de reciclagem de elementos tendo como forças motrizes as dinâmicas internas e externas dos processos geológicos e o fluxo constante de energia solar.

É possível que os primeiros microrganismos tenham se desenvolvido ao redor de aberturas hidrotermais, um ambiente escuro, salino, sob alta pressão e, naquele momento, anaeróbico (SCHLESINGER *et al.*, 2011). O declínio no fluxo de níquel nas aberturas hidrotermais por volta de 2,5 bilhões de anos coincidiu com o aumento de oxigênio na atmosfera e a subsequente evolução dos eucariotos (KONHAUSER *et al.* 2009). Níquel é um elemento essencial para microrganismos metanogênicos e com sua disponibilidade, a quantidade de metano produzida poderia ter propiciado, através do fenômeno da oxidação, o decréscimo da concentração de oxigênio dissolvido.

O surgimento da fotossíntese estabeleceu um ciclo bem definido onde a energia luminosa é empregada na oxidação de H_2O (fotossíntese oxigênica), HS^- , H_2 , ou Fe^{2+} (fotossíntese anoxigênica) e os elétrons e prótons gerados são utilizados na fixação de carbono inorgânico em matéria orgânica. As espécies químicas oxidadas resultantes são utilizadas como aceptores de elétrons no processo de respiração aeróbica ou anaeróbica pelos próprios fotossintetizantes ou por outros organismos não-fotossintetizantes.

Apesar do pouco conhecimento acerca da expansão do processo de fotossíntese, um dos últimos processos metabólicos a surgir foi a fotossíntese oxigênica (FALKOWSKI *et al.*, 2008). Apesar da complexidade do processo, seu surgimento data de 3 bilhões de anos. Com a proliferação de microrganismos fotossintetizantes, a concentração de oxigênio dissolvido e de oxigênio atmosférico aumentou gradativamente nos próximos 500 milhões de anos com a diminuição de metanogênicos (BEKKER *et al.*, 2004).

A produção de nitrato e sua utilização como aceptor de elétrons no processo de respiração surgiram após o desenvolvimento da fotossíntese oxigênica uma vez que sem oxigênio não há nitrato (FALKOWSKI, 1997).

A observação da distribuição de rotas metabólicas aliada a evidências moleculares indicam que a evolução celular nos primórdios da vida contou com um extenso processo de transferência horizontal de genes consistindo no principal modo de evolução de rotas (WOESE, 2002). A distribuição dos genes responsáveis pelos principais processos anabólicos e catabólicos pode ter sido tão extensa ao ponto de formar um reservatório global de genes anterior ao processo de diferenciação celular e à transferência vertical de genes (FALKOWSKI *et al.*, 2008). É válido apontar que a transferência horizontal se aplica tanto para genes individuais como para rotas inteiras (GOGARTEN *et al.*, 2002). A fixação e manutenção de uma reação ou rota metabólica dependem de uma série de fatores; entre eles, pressões seletivas de ordem nutricional ou bioenergética.

Uma forma de reconstruir as possíveis rotas metabólicas a partir de dados metagenômicos envolve a comparação com sequências gênicas pertencentes a rotas já caracterizadas depositadas em bancos públicos. Geralmente a etapa inicial deste método emprega o programa BLAST (ALTSCHUL *et al.*, 1990) ou similar para a comparação das sequências por alinhamento e a obtenção de um suporte estatístico para o resultado da comparação. Atualmente alguns programas automatizam a construção das rotas reunindo as sequências correspondentes a partir do conjunto de dados metagenômicos. Como exemplo é possível citar o PathoLogic (KARP *et al.*, 2002), o KAAS (MORIYA *et al.*, 2007) e o HUMAnN (ABUBUCKER *et al.*, 2012). Esses programas utilizam banco de dados como o BioCyc (CASPI *et al.*, 2008) ou KEGG (KANEHISA *et al.*, 2000) para gerar uma lista de genes ortólogos.

O Projeto Metazoo

O projeto intitulado “Estudos da diversidade microbiana do Parque Zoológico do Estado de São Paulo”, financiado pela Fundação de Amparo à Pesquisa do Estado de São Paulo (Processo FAPESP: 2011/50870-6) e ainda em vigência, tem por objetivo geral coletar, analisar e prospectar dados moleculares de três microbiomas existentes no parque: compostagem de resíduos orgânicos gerados no parque e que incluem material vegetal oriundo da Mata Atlântica, água do Lago São Francisco e fezes de macacos bugios de vida livre e de cativeiro. O projeto, que atendeu à chamada do programa BIOTA-microrganismos, conta com a participação de pesquisadores da Universidade de São Paulo (campi Butantã e Zona Leste), da Universidade Federal de São Paulo (campi Vila Clementino e Diadema) e da Fundação Parque Zoológico de São Paulo, e está sediado no Instituto de Química da USP sob a coordenação do Prof. João Carlos Setubal. Os microbiomas foram escolhidos para a investigação com a intenção de cobrir, ao menos em parte, o espectro de ambientes encontrados no parque: a mata tropical e o ambiente aquático.

Os primeiros dois anos do projeto Metazoo foram dedicados a coleta de amostras da compostagem e do lago visando extração de seu DNA total. Tais amostras foram submetidas ao sequenciamento metagenômico *shotgun* nas plataformas Roche-454 e Illumina-MiSeq. Paralelamente, as mesmas amostras foram utilizadas para obtenção e sequenciamento de *amplicons* do gene *16S* para análise da composição e estruturas destas comunidades microbianas. Os metatranscritomas de algumas amostras do processo de compostagem também foram sequenciados.

A partir de algumas das amostras da compostagem e do lago, também foi realizado o isolamento de bactérias e/ou de consórcios microbianos com fenótipos de interesse biotecnológico. Isolados e/ou consórcios selecionados estão sendo identificados e caracterizados através de diferentes abordagens.

A obtenção e a análise do vastíssimo conjunto de dados metagenômicos gerados no âmbito do projeto Metazoo são tema de dois projetos de Mestrado e de dois de Doutorado, incluindo o que resultou na presente tese, os quais exploram diferentes conjuntos de dados com abordagens distintas. Esta tese é resultado da

aplicação de abordagens computacionais na investigação do potencial metabólico das comunidades microbianas do Lago São Francisco e do sistema de compostagem para a ampliação dos conhecimentos acerca dos processos mantenedores desses ambientes.

OBJETIVO

Este trabalho teve como objetivo analisar, através de uma abordagem metagenômica/computacional, a variação do conteúdo gênico e do potencial metabólico das comunidades microbianas associadas a dois ambientes da Fundação Parque Zoológico de São Paulo: um reservatório artificial, o Lago São Francisco e o sistema de compostagem de resíduos do parque.

A partir dessas análises nos interessa esclarecer: (i) quão variáveis são as rotas metabólicas em amostras coletadas em diferentes períodos e se há uma relação com fatores externos; (ii) se os microbiomas são compostos por uma comunidade específica de microrganismos, generalistas ou especialistas, ou por diferentes agentes variando ao longo do tempo e, (iii) no caso do sistema de compostagem, quais processos no microbioma são semelhantes a outros processos de degradação de biomassa.

Os resultados obtidos serão apresentados a seguir em dois capítulos, os quais serão incorporados em dois manuscritos que estão em fase de redação.

REFERÊNCIAS

- ABUBUCKER, S.; SEGATA, N.; GOLL, J.; SCHUBERT, A. M.; IZARD, J.; CANTAREL, B. L.; RODRIGUEZ-MUELLER, B.; ZUCKER, J.; THIAGARAJAN, M.; HENRISSAT, B.; WHITE, O.; KELLEY, S. T.; METHE, B.; SCHLOSS, P. D.; GEVERS, P.; MITREVA, M.; HUTTENHOWER, C. Metabolic Reconstruction for Metagenomic Data and Its Application to the Human Microbiome. **PLoS Computational Biology**, v. 8, n. 6, p. 1 – 17, 2012.
- ALTSCHUL, S. F.; GISH, W.; MILLER, W.; MYERS, E. W.; LIPMAN, D. J. Basic local alignment search tool. **Journal of Molecular Biology**, v. 215, p. 403 – 410, 1990.
- AMANN, R. I.; LUDWIG, W.; SCHLEIFER, K. H. Phylogenetic identification and in situ detection of individual microbial cells without cultivation. **Microbiological Reviews**, v. 59, n. 1, p. 143 – 169, 1995.
- ARCHER, C.; VANCE, D. Coupled Fe and S isotope evidence for Archean microbial Fe(III) and sulfate reduction. **Geology**, v. 34, n. 3, p. 153 – 156.
- BANIK, J. J.; BRADY, S. F. Recent application of metagenomic approaches toward the discovery of antimicrobials and other bioactive small molecules. **Current Opinion in Microbiology**, v. 13, n. 5, p. 603 – 609, 2010.
- BEKKER, A.; HOLLAND, H. D.; WANG, P.-L.; RUMBLE, D.; STEIN, H. J.; HANNAH, J. L.; COETZEE, L. L.; BEUKES, N. J. **Nature**, v. 427, p. 117 – 120, 2004.
- BRASIER, M. D.; GREEN, O. R.; JEPHCOAT, A. P.; KLEPPE, A. K.; KRANENDONK, M. J. V.; LINDSAY, J. F.; STEELE, A.; GRASSINEAU, N. V. Questioning the evidence for Earth's oldest fossil. **Nature**, v. 416, p. 76 – 81, 2002.
- CASPI, R.; FOERSTER, H.; FULCHER, C. A.; KAIPA, P.; KRUMMENACKER, M.; LATENDRESSE, M.; PALEY, S.; RHEE, S. Y.; SHEARER, A. G.; TISSIER, C.; WALK, T. C.; ZHANG, P.; KARP, P. D. The MetaCyc Database of metabolic pathways and enzymes and the BioCyc collection of Pathway/Genome Databases. **Nucleic Acids Research**, v. 36, p. 623 – 631, 2008.
- CAVALIER-SMITH, T.; BRASIER, M.; EMBLEY, T. M. Introduction: how and when did microbes change the world? **Philosophical Transactions of the Royal Society B**, v. 361, n. 1470, p. 1 – 6, 2006.
- CRAIG, J. W.; CHANG, F.-Y.; KIM, J. H.; OBIAJULU, S. C.; BRADY, S. F. Expanding Small-Molecule Functional Metagenomics through Parallel Screening of Broad-Host-Range Cosmid Environmental DNA Libraries in Diverse Proteobacteria. **Applied and Environmental Microbiology**, v. 76, n. 5, p. 1633 – 1641, 2010.
- DONADIO, S.; MONCIARDINI, P.; ALDUINA, R.; MAZZA, P.; CHIOCCHINI, C.; CAVALETTI, L.; SOSIO, M.; PUGLIA, A. M. Microbial technologies for the discovery of novel bioactive metabolites. **Journal of Biotechnology**, v. 99, n. 3, p. 187 – 198, 2002.

EDWARDS, J. S.; COVERT, M.; PALSSON, B. Metabolic modelling of microbes: the flux-balance approach. **Environmental Microbiology**, v. 4, n. 3, p. 133 – 140, 2002.

FALKOWSKI, P. G. Evolution of the nitrogen cycle and its influence on the biological sequestration of CO₂ in the ocean. **Nature**, v. 387, p. 272 – 275, 1997.

FALKOWSKI, P. G.; FENCHEL, T.; DELONG, E. F. The Microbial Engines That Drive Earth's Biogeochemical Cycles. **Science**, v. 320, p. 1034 – 1039, 2008.

FIERER, N.; JACKSON, R. B. The diversity and biogeography of soil bacterial communities. **Proceedings of the National Academy of Sciences**, v. 103, n. 3, p. 626 – 631, 2006.

FUERST, J. A.; SAGULENKO, E. Keys to eukaryality: Planctomycetes and ancestral evolution of cellular complexity. **Frontiers in Microbiology**, v. 3, n. 167, p. 1 – 12, 2012.

GILBERT, J. A.; MEYER, F.; JANSSON, J.; GORDON, J.; PACE, N.; TIEDJE, J.; LEY, R.; FIERER, N.; FIELD, D.; KYRPIDES, N.; GLÖCKNER, F. O.; KLENK, H. P.; WOMMACK, K. E.; GLASS, E.; DOCHERTY, K.; GALLERY, R.; STEVENS, R.; KNIGHT, R. The Earth Microbiome Project: Meeting report of the "1 EMP meeting on sample selection and acquisition" at Argonne National Laboratory October 6 2010. **Stand Genomic Science**, v. 3, n. 3, p. 249 – 253, 2010.

GREEN, J.; BOHANNAN, B. J. M. Spatial scaling of microbial biodiversity. **Trends in Ecology & Evolution**, v. 21, n. 9, p. 501 – 507, 2006.

GRIFFITHS, R. I.; THOMSON, B. C.; JAMES, P.; BELL, T.; BAILEY, M.; WHITELEY, A. S. The bacterial biogeography of British soils. **Environmental Microbiology**, v. 13, n. 6, p. 1642 – 1654, 2011.

GOGARTEN, J. P.; DOOLITTLE, W. F.; LAWRENCE, J. G. Prokaryotic evolution in light of gene transfer. **Molecular Biology and Evolution**, v. 19, n. 12, p. 2226 – 2238, 2002.

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. **Chemistry & Biology**, v. 5, n. 10, p. 245 – 249, 1998.

HAYES, J. M.; WALDBAUER, J. R. The carbon cycle and associated redox processes through time. **Philosophical Transactions of the Royal Society B**, v. 361, n. 1470, p. 931 – 950, 2006.

HEAD, I. M.; SAUNDERS, J. R.; PICKUP, R. W. Microbial evolution, diversity, and ecology: a decade of ribosomal RNA analysis of uncultivated microorganisms. **Microbial Ecology**, v. 35, p. 1 – 21, 1998.

KANEHISA, M.; GOTO, S.; KEGG: Kyoto Encyclopedia of Genes and Genomes. **Nucleic Acids Research**, v. 28, n. 1, p. 27 – 30, 2000.

KARP, P. D.; PALEY, S.; ROMER, P. The pathway tools software. **Bioinformatics**, v. 18, p. 1 – 8, 2002.

KONHAUSER, K. O.; PECOITS, E.; LALONDE, S. V.; PAPINEAU, D.; NISBET, E. G.; BARLEY, M. E.; ARNDT, N. T.; ZAHNLE, K.; BALZ S.; KAMBER, B. S. Oceanic nickel depletion and a methanogen famine before the Great Oxidation Event. **Nature**, v. 458, p. 750 – 753.

LEE, K. C.; WEBB, R. I.; JANSSEN, P. H.; SANGWAN, P.; ROMEO, T.; STALEY, J. T.; FUERST, J. A. Phylum Verrucomicrobia representatives share a compartmentalized cell plan with members of bacterial phylum Planctomycetes. **BMC Microbiology**, v. 8, p. 1 – 10, 2009.

LORENZ, P.; ECK, J. Metagenomics and industrial applications. **Nature Reviews Microbiology**, v. 3, p. 510 – 516, 2005.

MLOT, C. Microbial diversity unbound: what DNA-based techniques are revealing about the planet's hidden biodiversity. **Bioscience**, n. 54, p. 1064 – 1068, 2004.

MOCALI, S.; BENEDETTI, A. Exploring research frontiers in microbiology: the challenge of metagenomics in soil microbiology. **Research in Microbiology**, v. 161, n. 6, p. 497 – 505, 2010.

MORIYA, Y.; ITOH, M.; OKUDA, S.; YOSHIKAWA, A. C.; KANEHISA, M. KAAS: an automatic genome annotation and pathway reconstruction server. **Nucleic Acids Research**, v. 35, p. 182 – 185, 2007.

PACE, N. R. A molecular view of microbial diversity and the biosphere. **Science**, v. 276, n. 5313, p. 734 – 740, 1997.

PRAKASH, T.; TAYLOR, T. D. Functional assignment of metagenomic data: challenges and applications. **Briefings in Bioinformatics**, v. 13, n. 6, p. 711 – 727, 2012.

PYLRO, V. S.; ROESCH, L. F.; ORTEGA, J. M.; DO AMARAL, A. M.; TÓTOLA, M. R.; HIRSCH, P. R.; ROSADO, A. S.; GÓES-NETO, A.; DA COSTA DA SILVA, A. L.; ROSA, C. A.; MORAIS, D. K.; ANDREOTE, F. D.; DUARTE, G. F.; DE MELO, I. S.; SELDIN, L.; LAMBAIS, M. R.; HUNGRIA, M.; PEIXOTO, R. S.; KRUGER, R. H.; TSAI, S. M.; AZEVEDO, V. Brazilian Microbiome Project: revealing the unexplored microbial diversity--challenges and prospects. **Microbial Ecology**, v. 67, n. 2, p. 237 – 241, 2014.

SCHIRRMESTER, B. E.; VOS, J. M.; ANTONELLI, A.; HOMAYOUN, C.; BAGHERIA, H. C. Evolution of multicellularity coincided with increased diversification of cyanobacteria and the Great Oxidation Event. **Proceedings of the National Academy of Sciences**, v. 110, n. 5, 1791 – 1796, 2013.

SCHLESINGER, W. H.; COLE, J. J.; FINZI, A. C.; HOLLAND, E. A. Introduction to coupled biogeochemical cycles. **Frontiers in Ecology and the Environment**, v. 9, p. 5 – 8, 2011.

SCHMEISSER, C.; STEELE, H.; STREIT, W. R. Metagenomics, biotechnology with non-culturable microbes. **Applied Microbiology and Biotechnology**, v. 75, n. 5, p. 955 – 962, 2007.

SINGH, B. K. Exploring microbial diversity for biotechnology: the way forward. **Trends in Biotechnology**, v. 28, n. 3, p. 111 – 116, 2010.

SINGH, B. K.; MACDONALD, C. A. Drug discovery from uncultivable microorganisms. **Drug Discovery Today**, v. 15, n. 17, p. 792 – 799, 2010.

RAES, J.; LETUNIC, I.; YAMADA, T.; JENSEN, L. J.; BORK, P. Toward molecular trait-based ecology through integration of biogeochemical, geographical and metagenomic data. **Molecular System Biology**, v. 7, p. 473, 2011.

RAPPE, M. S.; GIOVANNONI, S. J. The uncultured microbial majority. **Annual Review of Microbiology**, v. 57, p. 369 – 394, 2003.

RIESENFELD, C. S.; GOODMAN, R. M.; HANDELSMAN, J. Uncultured soil bacteria are a reservoir of new antibiotic resistance genes. **Environmental Microbiology**, v. 6, n. 9, p. 981 – 989, 2004.

RIVERA, M. C.; LAKE, J. A. The ring of life provides evidence for a genome fusion origin of eukaryotes. **Nature**, v. 431, p. 152 – 155, 2004.

SCHOPF J. W. Microfossils of the Early Archean Apex chert: new evidence of the antiquity of life. **Science**, n. 260, p. 640 – 646, 1993.

SCHOPF, J. W.; KUDRYAVTSEV, A. B.; AGRETI, D. G.; WDOWIAK, T. J.; CZAJA, A. D. Laser-Raman imagery of Earth's earliest fossils. **Nature**, v. 416, p. 73 – 76, 2002.

STEELE, H. L.; JAEGER, K. E.; DANIEL, R.; STREIT, W. R. Advances in recovery of novel biocatalysts from metagenomes. **Journal of Molecular Microbiology and Biotechnology**, v. 16, n. 1 – 2, p. 25 – 37, 2009.

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

TORSVIK, V.; OVREAS, L. Microbial diversity and function in soil: from genes to ecosystems. **Current Opinion in Microbiology**, v. 5, n. 3, p. 240 – 245, 2002.

TYSON, G.; CHAPMAN, J.; HUGENHOLTZ, P.; ALLEN, E.; RAM, R.; RICHARDSON, P.; SOLOVYEV, V.; RUBIN, E.; ROKHSAR, D.; BANFIELD, J. Community structure and metabolism through reconstruction of microbial genomes from the environment. **Nature**, v. 428, p. 37 – 43, 2004.

VENTER, J. C.; REMINGTON, K.; HEIDELBERG, J. F.; HALPERN, A. L.; RUSCH, D.; EISEN, J. A.; WU, D.; PAULSEN, I.; NELSON, K. E.; NELSON, W.; FOUTS, D. E.; LEVY, S.; KNAP, A. H.; LOMAS, M. W.; NEALSON, K.; WHITE, O.; PETERSON, J.; HOFFMAN, J.; PARSONS, R.; BADEN-TILLSON, H.; PFANNKOCH, C.; ROGERS, Y.; SMITH, H. O. Environmental genome shotgun sequencing of the Sargasso Sea. **Science**, v. 304, p. 66 – 74, 2004.

WACEY, D.; KILBURN, M. R.; SAUNDERS, M.; CLIFF, J.; BRASIER, M. D. Microfossils of sulphur-metabolizing cells in 3.4-billion-year-old rocks of Western Australia. **Nature Geoscience**, v. 4, p. 698 – 702, 2011.

WHITMAN, W. B.; COLEMAN, D. C.; WIEBE, W. J. Prokaryotes: the unseen majority. **Proceedings of the National Academy of Sciences**, v. 95, n. 12, p. 6578 – 6583, 1998.

WOESE, C. R.; KANDLER, O.; WHEELIS, M. L. Towards a natural system of organisms: proposal for the domains Archaea, Bacteria and Eucarya. **Proceedings of the National Academy of Sciences**, v. 87, p. 4576 – 4579, 1990.

WOESE, C. R. On the evolution of cells. **Proceedings of the National Academy of Sciences**, v. 99, n. 13, p. 8742 – 8747, 2002.

WOOLEY, J. C.; YE, Y. Metagenomics: Facts and Artifacts, and Computational Challenges. **Journal of Computer Science and Technology**, v. 25, n. 1, p. 71 – 81, 2009.

YUN, J.; KANG, S.; PARK, S.; YOON, H.; KIM, M. J.; HEU, S.; RYU, S. Characterization of a novel amylolytic enzyme encoded by a gene from a soil-derived metagenomic library. **Applied and Environmental Microbiology**, v. 70, n. 12, p. 7229 – 7235, 2004.

CAPÍTULO 2:

ANÁLISE COMPUTACIONAL DA COMUNIDADE MICROBIANA NO LAGO SÃO FRANCISCO DA FUNDAÇÃO PARQUE ZOOLOGICO DE SÃO PAULO

INTRODUÇÃO

O termo metagenoma surgiu em 1998 para designar os “genomas coletivos da microflora de solo” (HANDELSMAN *et al.*, 1998). Originalmente o metagenoma era extraído e clonado e os clones eram analisados em busca de fenótipos de interesse. O grande marco divisor viria em 2004 com a publicação dos resultados de dois grandes projetos de sequenciamento de amostras ambientais (VENTER *et al.*, 2004; TYSON *et al.*, 2004). Com o desenvolvimento e aumento de acessibilidade a novas tecnologias de sequenciamento (*Next-Generation Sequencing* – NGS) que dispensam a laboriosa etapa de clonagem, ensaios funcionais passaram a ser cada vez mais raros e limitados a circunstâncias específicas. Estas tecnologias abriram as portas para a investigação da microbiota de virtualmente todos ambientes terrestres e, dentre eles, ambientes de água doce.

Ecossistemas de água doce vêm despertando cada vez mais atenção, seja por seu papel na manutenção dos ecossistemas adjacentes, seja pela influência que sofrem e exercem nas atividades humanas como recreação e abastecimento (OH *et al.*, 2011). Nos últimos anos alguns estudos apontam para a existência de grupos de microrganismos característicos de água doce e com ampla distribuição geográfica (NEWTON *et al.*, 2011).

Estudos de comunidades microbianas de água doce com abordagens metagenômicas têm aumentado surpreendentemente nos últimos anos, focalizando os diferentes aspectos dos microrganismos e do ambiente no qual estão inseridos. Cada qual com seu processo de amostragem e com sua abordagem computacional de análise, tais estudos têm contribuído significativamente para o avanço do

conhecimento da Ecologia Microbiana. Seguem abaixo três breves descrições de estudos recentes e exemplificam o estado-da-arte desse campo.

O Lago Lanier se localiza em Atlanta, Estados Unidos, e consiste num importante reservatório de água potável. Um estudo iniciado em 2009 por Oh e colaboradores analisou através de sequenciamento *shotgun* na plataforma 454-Roche, amostras de água coletadas após uma tempestade de verão e durante o outono (OH *et al.*, 2011). No levantamento taxonômico feito pelos autores todos os grandes grupos típicos de ambientes lacustres estavam representados: Proteobacteria (37% das sequências do gene rRNA 16S presentes no metagenoma, 447 sequências); Actinobacteria (32%, 134 sequências); Verrucomicrobia (14%, 62 sequências) e Cyanobacteria (9%, 41 sequências). Esse estudo revelou um conteúdo gênico relativamente estável nos períodos de coleta e mais semelhante a conteúdos gênicos de outros ambientes de água doce e de água superficial do oceano do que semelhante a conteúdos gênicos de solo. Uma diferença marcante apontada pelo estudo foi o uso preferencial de mecanismos de osmorregulação baseado em potássio na comunidade microbiana do Lago Lanier em contraste ao uso de mecanismos baseado em sódio em comunidades de água salgada. Os autores também constataram que a comunidade do Lago Lanier apresenta menor riqueza e complexidade do que comunidades oceânicas ou presentes no solo.

Aplicando uma combinação de sequenciamento *shotgun* e sequenciamento do genoma de célula única (*single-cell sequencing*), foi realizado um levantamento de fotoheterotróficos (bactérias contendo rodopsina ou fototróficos aeróbicos e anoxigênicos) e quimioautotróficos em quatro comunidades lacustres dos Estados Unidos (MARTINEZ-GARCIA *et al.*, 2012). Esse estudo mostrou que fotoheterotróficos compõem de 10–23% do bacterioplâncton e que bactérias contendo rodopsinas são a fração dominante. O filo Actinobacteria foi o que mais contribuiu com membros contendo rodopsinas. Outros grupos típicos de água doce também apresentaram representantes: Alphaproteobacteria, Betaproteobacteria, Gammaproteobacteria, Deltaproteobacteria, Verrucomicrobia e Sphingobacteria. *Polynucleobacter* spp. foram os fototróficos aeróbicos anoxigênicos dominantes nos ambientes estudados. Baseados na distribuição do gene codificante da enzima RuBisCO (ribulose 1,5-bisfosfato carboxilase/oxigenase) na comunidade, os autores

propuseram que quimioautotróficos são mais prevalentes entre o bacterioplâncton do que se considerava.

Mais recentemente, foi realizado um estudo comparativo entre dados de sequenciamento metagenômico de lagos suecos (Vättern, Ekoln e Erken), de lagos norte-americanos (Lanier – OH *et al.*, 2011; lago de Yellowstone; Damariscotta, Mendota, Sparkling e Trout – MARTINEZ-GARCIA *et al.*, 2012) e de água salgada (EILER *et al.*, 2014). A análise taxonômica através da recuperação de sequências do gene rRNA 16S constatou a predominância de Proteobacteria nas amostras marinhas e de Actinobacteria nas amostras de água doce. Quando se considerou as classes dentro de Proteobacteria, Alphaproteobacteria mostrou-se dominante nas amostras marinhas, enquanto Betaproteobacteria foi dominante nas amostras de água doce. Como esperado outros filos se mostraram presentes nas amostras dos lagos: Verrucomicrobia, Planctomycetes, Cyanobacteria e Bacteroidetes. Considerando as rotas metabólicas diferencialmente representadas nos dois conjuntos, as discrepâncias são devidas principalmente a funções relacionadas à osmorregulação, transporte ativo, metabolismo de carboidratos e aminoácidos. Esse estudo possibilitou aos autores sugerir a hipótese de que microrganismos oceânicos, comparados a microrganismos de água doce, investem mais em metabolismo de aminoácidos e que mecanismos metabólicos de carboidratos variam entre o primeiro e o último grupo.

Com o propósito de ampliar esse panorama, este trabalho teve como objetivo analisar, através de uma abordagem metagenômica/computacional, a variação do potencial metabólico e a diversidade da comunidade microbiana associada ao Lago São Francisco da Fundação Parque Zoológico de São Paulo, o qual possui uma área de 34.665 m². A escolha deste ambiente, que é um reservatório artificial, foi devido ao fato de que é um ambiente aquático localizado numa interface, fragmento da Mata Atlântica e perímetro urbano, sendo habitado por animais de vida livre incluindo aves migratórias, cuja frequência varia sazonalmente. Uma característica particular deste ambiente é a presença de algas devido ao excesso de matéria orgânica acumulada do ambiente e por isso aspecto esverdeado, sem odor característico. O lago represa água de diversas nascentes do parque e está localizado em um terreno que possui uma ligeira inclinação natural permitindo a confluência das águas recebidas. Diante dessa característica, este ponto foi o local

escolhido para realizar as coletas de água superficial, durante o período de 12 meses.

METODOLOGIA

Visão Geral da Sequência de Análises

Extração de DNA

Coletas de amostras de 2 L de água do Lago São Francisco foram feitas na primeira semana de cada mês, no período de outubro de 2012 à setembro de 2013. As amostras de água provenientes do Lago São Francisco foram previamente filtradas em rede de fitoplâncton de 20 μm , para a retirada de algas. Em seguida, realizou-se filtração em membrana de microfibras de vidro de 8 μm (Milipore), seguida de filtração em membrana de 5 μm (Durapore, Milipore). Após esta etapa, seguiu-se a filtração para extração de DNA de microrganismos. A água foi filtrada em membrana de 0,22 μm de porosidade até saturação da mesma de acordo com as recomendações do kit “MoBio PowerWater DNA Isolation”. Brevemente, após a etapa de concentração em membrana de 0,22 μm de porosidade, esta foi transferida para um tubo contendo *beads* fornecido e processado de acordo com as recomendações do fabricante. Este kit recomenda a filtração de volumes entre 50 e 100 mL de água de acordo com a turbidez da amostra analisada. A pureza/quantidade e integridade do DNA isolado da água do lago foram avaliadas respectivamente por medida da absorbância a 230, 260 e 280nm e eletroforese capilar no equipamento BioAnalyzer. As amostras de DNA foram armazenadas a -20°C até a realização do sequenciamento.

Preparação das Bibliotecas de DNA

Após a análise da pureza e da integridade das amostras de DNA, as bibliotecas foram preparadas para o sequenciamento *shotgun* utilizando o kit comercial “Nextera DNA Sample Prep” (Illumina) segundo o protocolo “Nextera DNA Sample Preparation Guide”. Na realização desse protocolo foram utilizados 35 ng do DNA total diluídos em 25 μL de água. Os fragmentos de DNA foram então purificados utilizando o kit “QIAquick PCR purification” de acordo com protocolo do fabricante resultando 25 μL de suspensão de DNA em “EB buffer”. A etapa posterior consiste em uma amplificação por PCR com a adição de índices exclusivos para cada biblioteca e de adaptadores P5 e P7. Nesta etapa foi utilizada uma alíquota de 20 μL de DNA purificado de cada amostra e os kits empregados foram o “Nextera

PCR Master Mix” para a amplificação e o “AMPure XP beads” para a seleção de fragmentos de tamanho adequado (~700bp). A quantificação das bibliotecas de DNA foi realizada por meio de um ensaio de qPCR utilizando o kit “Kapa Library Quantification” para plataformas de sequenciamento Illumina. Aliquotas contendo 300 pM de cada biblioteca foram reunidas em uma mistura que foi desnaturada por incubação a 98°C por 10 minutos e diluídas em tampão “HT1” (fornecido pelo kit de sequenciamento) até a concentração final de 10 pM de cada uma das amostras. Seiscentos microlitros da mistura foram aplicados no cartucho de sequenciamento, o qual foi inserido no equipamento MiSeq (Illumina). O *workflow* escolhido para a análise após a corrida foi o de “Generate FastQ”. Em todos os casos, os sequenciamentos foram realizados com a estratégia de sequenciamento “Paired-End” usando kits para 500 ciclos.

A preparação de bibliotecas de genes rRNA 16S reações de PCR foram realizadas utilizando um par de primers baseados nas sequências S-D-Bact-0341-b-S-17 e S-D-Bact-0785-a-A-21 (KLINDWORTH *et al.*, 2013), para a amplificação da região variável V3 e V4 do gene rRNA 16S, com a adição de adaptadores sugeridos no *pipeline* da empresa Illumina (Illumina workflow for 16S Metagenomic Sequencing Library Preparation). As sequências dos *primers* que utilizamos está abaixo:

341F 5'-TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGCCTACGGGNGGCWGCAG
785R 5'-GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGGACTACHVGGGTATCTAATCC

Reações de PCR com volume de 25 µL foram montadas com 12,5 ng de DNA metagenômico total utilizando o kit “KAPA HiFi HotStart ReadyMix” (Kapa Biosystems) e 200 nM de cada *primer* a 95°C por 3 minutos seguidos de 25 ciclos compostos por 30 segundos a 95°C, 30 segundos a 55°C e 30 segundos a 72°C. Após a etapa de ciclagem as reações foram incubadas por cinco minutos a 72°C. Um microlitro do produto de PCR foi analisado com o kit Bioanalyzer DNA 1000 chip (Agilent Technologies) para verificar o tamanho dos amplicons (~550 pb). O produto de PCR foi purificado com o kit “AMPure XP beads” (Beckman Coulter, Inc.). Através do kit “Nextera XT Index” (Illumina, Inc., CA), índices foram adicionados a cada biblioteca que foi purificada novamente e a verificação do comprimento esperado das sequências, aproximadamente 630 pb, foi feita novamente pelo kit Bioanalyzer utilizando uma diluição de 1:50. Novamente quantificação das bibliotecas de DNA foi

realizada, por meio de um ensaio de qPCR utilizando o kit “Kapa Library Quantification” para plataformas de sequenciamento Illumina. Aliquotas contendo 300 pM de cada biblioteca foram reunidas em uma mistura que foi desnaturada por incubação a 98°C por 10 minutos e diluídas em tampão “HT1” (fornecido pelo kit de sequenciamento) até a concentração final de 10 pM de cada uma das amostras. Seiscentos microlitros da mistura foram aplicados no cartucho de sequenciamento, o qual foi inserido no equipamento MiSeq (Illumina). O workflow escolhido para a análise após a corrida foi o de “Generate FastQ”. Em todos os casos, os sequenciamentos foram realizados com a estratégia de sequenciamento “Paired-End” usando kits para 500 ciclos.

Análise da Distribuição Taxonômica e das Diversidades α e β das Amostras do Lago São Francisco

Para a análise da distribuição taxonômica e das diversidades α e β , as sequências obtidas no sequenciamento dos *amplicons* dos genes rRNA 16S foram analisadas de acordo com o protocolo padrão do pacote QIIME (CAPORASO *et al.*, 2010) que conta com as etapas de sobreposição de pares (fitas senso e antissenso), retirada de sequências identificadoras e iniciadoras (barcodes e primers respectivamente), eliminação de sequências com baixa qualidade (*Phred Score* menor que 20), derreplicação, agrupamento de OTUs seguido de construção de um conjunto de sequências representativas e atribuição taxonômica das OTUs foi realizada a partir do programa *RDP Classifier*. Foi aplicado um filtro que exclui OTUs cuja contagem tenha sido inferior a 120 considerando as 12 amostras. Para a análise da distribuição taxonômica, foram utilizadas todas as sequências obtidas nas corridas para cada amostra.

Considerando que as análises acima atingem o nível de gênero, o programa MyTaxa foi utilizado para a análise da distribuição taxonômica até o nível de espécie através das sequências obtidas no sequenciamento *shotgun* (LUO *et al.*, 2014). O programa MyTaxa conta com um banco de dados contendo todos os genes de genomas disponíveis, completos e incompletos, com dois valores para cada gene: *D*, o quão bem o gene contribuem para a classificação da *read* no ranque

taxonômico e M quão consistente é a árvore do gene comparada com a árvore das espécies que o contém. O programa recebe o arquivo de saída do programa Usearch contendo os *hits* de cada *read* e utilizando a pontuação *bit score*, D e M calcula a probabilidade da *read*, assinalada a um determinado gene, pertencer a um organismo que possui tal gene. Se a probabilidade for menor que um determinado limite, a *read* não é classificada.

Análise de Similaridade Local e Construção da Rede de Correlação

Através da abundância relativa obtida pela identificação taxonômica com o programa MyTaxa, uma análise de similaridade local foi executada utilizando o programa eLSA (RUAN *et al.*, 2006; XIA *et al.*, 2011). Para duas sequências de mesmo tamanho, séries de OTUs, a medida de similaridade local é definida como a soma máxima do produto das entradas correspondentes de todas as suas subsequências. Para O_1 e O_2 ($O_1 = O_{11}, O_{12}, \dots, O_{1n}$ e $O_2 = O_{21}, O_{22}, \dots, O_{2n}$), a matriz positiva $P_{n \times n}$ e a matriz negativa $N_{n \times n}$ são calculadas da seguinte forma:

- (1) For $i, j = 1, \dots, n$,
 $P_{0,j} = P_{j,0} = 0$ and $N_{0,j} = N_{j,0} = 0$.
- (2) For $i, j = 1, \dots, n$ with $|i - j| \leq D$,
 $P_{i+1,j+1} = \max\{0, P_{i,j} + O_{1,i+1} \cdot O_{2,j+1}\}$ and
 $N_{i+1,j+1} = \max\{0, N_{i,j} - O_{1,i+1} \cdot O_{2,j+1}\}$.
- (3) $P(O_1, O_2) = \max_{1 \leq i, j \leq n} P_{i,j}$ and
 $N(O_1, O_2) = \max_{1 \leq i, j \leq n} N_{i,j}$.
- (4) $\text{MaxScore}(O_1, O_2) = \max(P(O_1, O_2), N(O_1, O_2))$ and
 $\text{Flag}(O_1, O_2) = \text{sgn}(P(O_1, O_2) - N(O_1, O_2))$.

O valor de similaridade local das séries O_1 e O_2 , $LS(O_1, O_2)$, é definido pela divisão do valor $\text{MaxScore}(O_1, O_2)$ descrito acima pelo tamanho da série, n .

Para determinação dos p-valores foram determinadas 2000 réplicas (*bootstrap*). Ao final foram mantidos os pares com valores de correlação menores que -0,9 ou maiores que 0,9 com p-valor abaixo de 2×10^{-3} e taxa de detecção falsa

(*False Discovery Rate* – FDR) 0,08. Os pares de espécies correlacionadas foram convertidos em um grafo não-dirigido através do programa Cytoscape v3 (Shannon *et al.*, 2003) regiões densamente conectadas foram identificadas pelo algoritmo GLay (SU *et al.*, 2010) implementado no plugin clusterMaker2 (MORRIS *et al.*, 2011).

Identificação e Abundância de e Módulos Metabólicos usando o Programa HUMAnN

Rota metabólica pode ser definida como o conjunto de reações químicas subsequentes e interdependentes catalisadas por enzimas e que ocorrem em um sistema biológico. Os produtos intermediários de uma determinada rota metabólica podem servir de substrato para outras rotas compondo uma imensa rede metabólica. Dessa forma é útil dividir essa rede em unidades funcionais chamadas **módulos**. As rotas e módulos discutidos nesse projeto são aqueles que constam na plataforma KEGG - Kyoto Encyclopedia of Genes and Genomes (KANEHISA & GOTO, 2000). Um módulo KEGG é uma coleção de unidades funcionais definida manualmente, identificada por números *M*, utilizado para anotações biológicas de genomas sequenciados. Os módulos KEGG são distribuídos em quatro grupos: módulos de rotas representando unidades funcionais nos mapas metabólicos KEGG; módulos estruturais formando maquinarias celulares; conjuntos funcionais contendo outros elementos essenciais (como Aminoacyl-tRNA synthases, prokaryotes – M00360) e módulos de assinatura contendo marcadores de determinados fenótipos (como EHEC pathogenicity signature, Shiga toxin – M00363) (informações disponíveis em <http://www.genome.jp/kegg/module.html>).

O programa HUMAnN que foi aplicado, acrônimo de *HMP Unified Metabolic Analysis Network* (ABUBUCKER *et al.*, 2012), o qual é composto por uma série de *scripts* escritos em Python que possibilitam identificar com rigor matemático a presença de rotas metabólicas em metagenomas e estimar sua abundância relativa.

As *reads* 1, ou seja, o primeiro conjunto de *reads* sintetizadas no sequenciamento *shotgun* foram submetidas à plataforma *online* MG-RAST, onde foram filtradas, tiveram as regiões de baixa qualidade aparadas, foram derreplicadas segundo o protocolo padrão dessa plataforma (MEYER *et al.*, 2008; GLASS *et al.*,

2010). Após a filtragem, as *reads* foram recuperadas e submetidas ao programa ao programa USEARCH (EDGAR, 2010) para identificação por similaridade a sequências de aminoácidos depositadas na plataforma KEGG organizadas por organismo. O resultado em formato tabular foi utilizado como entrada para a sequência de *scripts* do programa HUMAnN.

Primeiramente, para cada amostra o HUMAnN estabeleceu a abundância de famílias de genes ortólogos através de uma contagem de *hits* segundo a fórmula:

$$w_i = |G_i| \sum_{g \in G_i} \frac{1}{|g|} \sum_j \frac{1 - p_{g,j}}{\sum_{g'} 1 - p_{g',j}}$$

Onde:

w_i – abundância da família de ortólogos de i ;

$|G_i|$ – número de ortólogos da família i ;

$|g|$ – tamanho da sequência g pertencente à família;

$p_{g,j}$ – pontuação do fragmento j comparado à sequência g da família i , calculada a partir da expressão $p = 1 - e^{-E}$.

Os genes pertencentes às famílias de ortólogos identificadas na etapa anterior são submetidos pelo programa HUMAnN ao algoritmo MinPath (YE & DOAK, 2009) que de forma parcimoniosa os atribuiu a rotas metabólicas formando o menor conjunto de reações. Em seguida, com o intuito de eliminar falsos-positivos, o programa HUMAnN estabeleceu uma restrição taxonômica através do estabelecimento da abundância de organismos e eliminou rotas abaixo do valor esperado segundo as fórmulas:

$$w_o = |G_o| \sum_{g \in G_o} \frac{1}{|g|} \sum_j \frac{1 - p_{g,j}}{\sum_{g'} 1 - p_{g',j}} \quad \text{e} \quad e_p = \sum_o \hat{o}_{o,p} w_o$$

Onde:

w_o – abundância do organismo o ;

$|G_o|$ – número de sequências do organismo o ;
 $|g|$ – tamanho da sequência g pertencente ao organismo o ;
 $p_{g,j}$ – pontuação do fragmento j comparado à sequência g do organismo o , calculada a partir da expressão $p = 1 - e^{-E}$;
 ep – abundância esperada para uma determinada rota;
 $\partial_{o,p}$ – 1 se a rota foi anotada no organismo o ou 0 caso contrário.

Para remover os falsos-negativos, o programa HUMAnN estabeleceu uma etapa de “preenchimento de lacunas”. Em cada rota mantida, ortólogos com abundância relativa igual 1,5 x distância interquartil abaixo da mediana teve a abundância relativa elevada a um valor efetivo igual à mediana. Por fim o programa HUMAnN forneceu duas medidas para cada rota segundo as fórmulas:

$$cov_p = \frac{1}{|p|} \sum_{i \in p} \partial(w_{i,p} > \tilde{w}_{i,p}) \quad \text{e} \quad abd_p = \frac{2}{|p|} \sum_{i \in [p/2]} w_{i,p}$$

Onde:

cov_p – cobertura;
 $|p|$ – número de genes da rota p ;
 $w_{i,p}$ – abundância do ortólogo i presente na rota p ;
 $\tilde{w}_{i,p}$ – mediana do ortólogo i presente na rota p ;
 abd_p – abundância.

Uma vez geradas as listas, o valor da abundância de cada rota ou módulo passou por uma correção simples em função do tamanho do metagenoma: o valor foi dividido pelo tamanho do metagenoma e em seguida multiplicado por 10^6 . Os módulos metabólicos das amostras foram representados em gráficos do tipo *heatmap* montados na plataforma R com o pacote Pheatmap com agrupamento hierárquico completo.

RESULTADOS E DISCUSSÃO

Sequenciamento de *Amplicons* do Gene rRNA 16S e Sequenciamento *Shotgun* das Amostras do Lago São Francisco

O Lago São Francisco é um reservatório de águas de outras nascentes. É um ambiente com acúmulo de matéria orgânica e consequentemente eutrófico. Na figura 1 é possível observar o aspecto geral do lago.

FIGURA 1: Vista aérea do Lago São Francisco da Fundação Parque Zoológico de São Paulo. Imagem obtida através do programa Google Earth, altitude do ponto de visão 1.78 km, 23°39'04.72"S, 46°37'05.78"O.

Ao longo de doze meses, amostras foram coletadas e submetidas ao sequenciamento do *amplicon* do gene rRNA 16S e sequenciamento *shotgun*. O resultado dos sequenciamentos está registrado na tabela 1.

TABELA 1: Resultado do sequenciamento metagenômico das amostras do Lago São Francisco.

Amostra	Número total de <i>reads shotgun</i>	Tamanho médio dos <i>reads shotgun</i>	Número de <i>reads amplicon</i> * do rRNA 16S
Out12	5.237.431	180	1.565.448
Nov12	10.479.554	216	2.066.012
Dez12	4.312.046	171	1.120.128
Jan13	6.634.553	185	1.591.527
Fev13	10.463.037	149	1.780.975
Mar13	5.461.386	165	857.425
Apr13	5.520.783	157	1.554.595
Mai13	9.001.569	175	1.251.901
Jun13	4.509.360	169	1.410.121
Jul13	5.281.400	154	1.890.592
Ago13	5.307.017	152	1.147.621
Set13	5.995.897	157	1.497.888

* amplicon de 464 bp, após montagem dos *reads* 1 e 2.

Distribuição Taxonômica nas Amostras do Lago São Francisco através do Sequenciamento de *Amplicons* do Gene rRNA 16S

Os cinco filos mais abundantes foram Proteobacteria, Actinobacteria, Bacteroidetes, Verrucomicrobia e OD1 (FIGURA 2). Os quatro primeiros são comumente os mais abundantes em epilânio de lagos (ZWART *et al.*, 2002; NEWTON *et al.*, 2011).

FIGURA 2: Distribuição taxonômica dos cinco filos mais abundantes do Lago São Francisco.

As *reads* identificadas como pertencentes a Proteobacteria correspondem em média a 37%, variando entre 29% das *reads* provenientes da amostra de março (183.816 *reads*) e 45% das *reads* provenientes da amostra de maio (644.493 *reads*). Como esperado, a classe Betaproteobacteria é o grupo mais abundante: em média 26%, variando entre 17% em março e 33% em julho. A variação mensal dos táxons mais numerosos está registrada na figura 3.

FIGURA 3: Variação da abundância dos sete táxons de Betaproteobacteria mais abundantes no Lago São Francisco.

O gênero *Polynucleobacter* pertencente à família Burkholderiaceae conta com cinco espécies descritas, *P. acidiphobus*, *P. cosmopolitanus*, *P. difficilis*, *P. rarus* e *P. necessarius*, sendo esta última dividida em *P. necessarius necessarius* e *P. necessarius asymbioticus*. Todas as espécies são comuns em corpos d'água doce. Um experimento laboratorial demonstrou que os gêneros *Polynucleobacter* e *Methylophilus* foram onipresentes e dominantes em microcosmos artificiais em diferentes temperaturas (15°C, 25°C e 35°C) e condições nutricionais (controle, adição de nitrogênio, adição de fósforo). Nesse experimento ultramicrobactérias, Actinobacteria, *Polynucleobacter* sp., os agrupamentos LD12 e LD28 e bactérias relacionadas a *Methylophilus*, foram os táxons dominantes da comunidade a altas temperaturas (25°C e 35°C) em qualquer das condições nutricionais (REN *et al.*, 2013). Contudo, um outro estudo sugere que diferentes linhagens de *P. necessarius asymbioticus* podem apresentar respostas adaptativas diferentes a pH, condutividade, carbono orgânico dissolvido, concentração de oxigênio ao invés de uma flexibilidade ampla a esses fatores (JEZBERA *et al.*, 2011). Neste estudo, o gênero *Polynucleobacter* apresentou uma variação considerável na amostra de janeiro de 2013 (FIGURA 3), que pode ter como causa a concentração de substâncias húmicas carregadas dos arredores uma vez que no dia anterior à coleta da amostra foi registrada chuva (<http://www.estacao.iag.usp.br/Boletins/2013.pdf>). A linhagem QLW-P1DMWA-1 de *P. necessarius asymbioticus* é um exemplo de redução genômica: ausência de motilidade, quimiotaxia, além de características compartilhadas com microrganismos marinhos oligotróficos como baixa representação de genes assinalados às categorias COG N (motilidade celular e secreção), T (transdução de sinal), V (defesa), e K (transcrição) e como a alta representação de genes assinalados às categorias COG Q (biossíntese de metabólitos secundários, transporte e catabolismo) e I (transporte de lipídeos e metabolismo) e a utilização de compostos orgânicos de baixo peso molecular derivados da foto-oxidação de substâncias húmicas (HAHN *et al.*, 2012). Além disso, já foi demonstrado que indivíduos do gênero *Polynucleobacter* e outros pertencentes a uma comunidade de água doce desenvolvem-se com a adição de substâncias húmicas naturais (HUTALLE-SCHMELZER *et al.*, 2010). Por algum motivo o gênero *Methylophilus*, também sensível a substâncias húmicas, não mostrou aumento na população.

A classe Alphaproteobacteria variou entre 4% em novembro de 2012 e 8% em junho de 2013. Os gêneros mais representados, identificados até pelo menos o nível taxonômico de família, foram: um gênero não identificado da família Rickettsiaceae (em média 0,5%), *Roseomonas* (família Acetobacteraceae, em média 0,5%), *Rhodobacter* (família Rhodobacteraceae, em média 0,4%), *Phenylobacterium* (família Caulobacteraceae, em média 0,3%) e *Candidatus Pelagibacter* (família SAR11, em média 0,3%). A variação mensal dos táxons está registrada na figura 4.

A classe Alphaproteobacteria, menos estudada e com representação menor em comparação a Betaproteobacteria, possui membros que muitas vezes apresentam mecanismos de evasão de predadores (*grazing resistance*), mecanismos de ajuste e competitividade em condições de restrição de nutrientes, capacidade de degradar compostos orgânicos complexos e ampla distribuição em lagos ao redor do planeta (NEWTON *et al.*, 2011). Um estudo sobre florescimento da diatomácea *Aulacoseira granulata* e das cianobactérias *Microcystis aeruginosa* e *Cylindrospermopsis raciborskii*, mostrou que embora a classe Betaproteobacteria fosse dominante nas coculturas inoculadas com bactérias de lago, decrescia durante o período de incubação enquanto Alphaproteobacteria proliferava (BAGATINI *et al.*, 2014). É possível que o aumento dos táxons relacionados aos gêneros *Roseomonas* e *Phenylobacterium* dessa classe tenham se beneficiado da presença de substâncias húmicas e da disponibilidade do nicho deixado por outras populações (HUTALLE-SCHMELZER, *et al.*, 2010).

FIGURA 4: Variação da abundância dos cinco táxons de Alphaproteobacteria mais abundantes no Lago São Francisco.

As demais *reads* identificadas como pertencentes a Proteobacteria correspondem a 2,4% de Gammaproteobacteria em média, variando entre 14.012 *reads* provenientes da amostra de dezembro de 2012 e 39.417 provenientes da amostra de novembro de 2012, 1% de Deltaproteobacteria em média, variando entre 4.401 *reads* de outubro de 2012 e 24.102 de fevereiro de 2013 e 0,1% de Epsilonproteobacteria em média, variando entre 31 *reads* de março de 2013 e 5.834 de agosto de 2013 (FIGURA 5).

FIGURA 5: Distribuição de *reads* atribuídas às classes Gammaproteobacteria, Deltaproteobacteria e Epsilonproteobacteria no Lago São Francisco.

O filo Actinobacteria foi o segundo mais numeroso e as *reads* identificadas como pertencentes a esse táxon correspondem em média a 18%, variando entre 14% das *reads* provenientes da amostra de dezembro de 2012 (97.784 *reads*) e 25% das *reads* provenientes da amostra de abril de 2013 (280.954 *reads*). Dentro deste filo existe um táxon (ordem Actinomycetales), cujo gênero e família não foram identificados, que contribui em média com 12,4% do total de *reads*, variando entre 10% das *reads* de dezembro de 2012 (70.863 *reads*) e 17% das *reads* de abril de 2013 (197.994 *reads*). Os gêneros identificados como mais abundantes foram *Ilumatobacter* (família Acidimicrobiaceae, em média 2,2%), *Mycobacterium* (família Mycobacteriaceae, em média 1,6%), *Conexibacter* (família Conexibacteraceae, em média 0,7%) e *Iamia* (família Iamiaceae, em média 0,3%) e sua variação está representada na figura 6.

O filo Actinobacteria consiste em um componente autóctone nos ambientes de água doce. Evidências experimentais indicaram o declínio da abundância relativa de táxons desse filo quando a cocultura era enriquecida em carbono orgânico dissolvido. Esse declínio aparentemente não se deve à incapacidade de metabolizar esses compostos e nem a um possível efeito negativo dos mesmos, mas sim à proliferação de microrganismos oportunistas, como membros do filo Proteobacteria, mais adaptados à competição por outros recursos (BURKERT *et al.*, 2003). Além disso, um estudo demonstrou que membros do filo Actinobacteria possuem capacidade limitada de crescer na presença de fenol ou compostos húmicos, reforçando a ideia de baixa competitividade com outros microrganismos (HUTALLE-SCHMELZER *et al.*, 2010). No presente estudo, as principais OTUs pertencentes a esse filo apresentaram uma diminuição drástica nas populações no mês de novembro de 2012. Essa diminuição foi antecederida por um recorde de temperatura máxima para o mês de outubro desde 1933, no dia 31 de outubro, 35,9° (<http://www.estacao.iag.usp.br/Boletins/2012.pdf>). Essa variação populacional pode portanto estar relacionada ao aumento da temperatura da água e à consequente diminuição de oxigênio dissolvido, ao carreamento de compostos húmicos e à proliferação de outros táxons oportunistas mais adaptados a o nicho ecológico estabelecido naquela época do ano.

FIGURA 6: Variação da abundância dos quatro táxons de Actinobacteria mais abundantes no Lago São Francisco.

O filo Bacteroidetes identificado como terceiro mais abundante, contribuiu com 125.846 *reads* (9%) provenientes da amostra de julho de 2013 e 107.334 *reads* (19%) provenientes da amostra de dezembro de 2012, uma contribuição média de 14%. Os gêneros identificados como mais abundantes foram um gênero da família Chitinophagaceae (em média 1,5%), um gênero da família Cryomorphaceae (em média 1,5%) e os gêneros *Flavobacterium* (família Flavobacteriaceae, em média 1,3%), *Fluviicola* (família Cryomorphaceae, em média 0,8%) e *Sediminibacterium* (família Chitinophagaceae, em média 0,7%). A variação desses gêneros está registrada na figura 7. Vários táxons do filo Bacteroidetes apresentam um aumento populacional relacionado ao aumento da concentração de compostos carbônicos dissolvidos, exógenos ou produzidos por organismo fotossintetizantes (EILER & BERTILSSON, 2004; KOLMONEN *et al.*, 2004; EILER & BERTILSSON, 2007; ZEDER *et al.*, 2009). Experimento de enriquecimento em compostos húmicos também mostraram influência em táxons desse filo (HUTALLE-SCHMELZER *et al.*, 2010). Além disso, um levantamento de microrganismos presentes em corpos d'água durante o florescimento da espécie invasiva *Hydrilla verticillata* mostrou que

há uma separação de nicho onde Bacteroidetes são mais propensos ao consumo de carboidratos e, em especial, que o gênero *Sediminibacterium* foi um dos táxons recuperados com maior abundância nesse meio (GORDON-BRADLEY *et al.*, 2015). Assim, os dados da literatura apontam um possível nicho estabelecido nos primeiros meses do ano de 2013, evidenciado principalmente pelo aumento significativo do gênero *Sediminibacterium*.

FIGURA 7: Variação da abundância dos cinco táxons de Bacteroidetes mais abundantes no Lago São Francisco.

O filo Verrucomicrobia foi o quarto mais abundante correspondendo em média a 13,4%, contando com 9,1% do total de *reads* provenientes da amostra de outubro de 2012 (77.099 *reads*) e 23,3% do total de *reads* provenientes da amostra de janeiro de 2013 (102.190 *reads*). O táxon mais numeroso desse filo só pode ser identificado até o nível de classe, Spartobacteria, e contribui em média com 6,5% das *reads*, variando entre 2,8% em julho de 2013 (41.526 *reads*) e 9,9% em março de 2013 (62.546 *reads*). Os demais táxons desse filo, identificados até o nível de gênero, apresentam valores médios abaixo de 1%: o gênero *Opitutus* varia entre

0,11% das *reads* da amostra de dezembro de 2012 e 0,51% das *reads* da amostra de abril de 2013 (em média 0,23% das *reads*), o gênero *Prostheco bacter* varia entre 0,02% das *reads* de setembro de 2012 e 0,27% de fevereiro de 2013 (em média 0,11%) e o gênero *Luteolibacter* varia entre 0,002% das *reads* de novembro de 2012 e 0,049% das *reads* de fevereiro de 2013 (em média 0,01%).

Finalmente, o quinto gênero mais abundante, OD1, é representado por um único táxon não identificado que contribui em média com 4,5% das *reads*, com valores que vão de 1,1% das *reads* da amostra de março de 2013 (6.725 *reads*) e 8,4% das *reads* da amostra de julho de 2013 (120.838 *reads*).

Diversidades α e β das Amostras do Lago São Francisco através de *Amplicons* do Gene rRNA 16S

Para evitar qualquer distorção nas análises estatísticas a seguir em função do número de *reads* obtidas de cada amostra, foi realizado o seguinte processo de normalização: em todas as amostras foram utilizados subconjuntos com o mesmo número de *reads*, ou seja, 687.661 *reads* filtradas. Este número de *reads* equivale ao número de *reads* do menor conjunto, as *reads* obtidas da amostra de março de 2013. Os principais índices utilizados em Ecologia estão ilustrados na tabela 2.

As amostras são comparavelmente ricas e diversas, variando entre 1.250 espécies em agosto de 2013 (índice de Shannon = 7,177; índice de Simpson = 0,982) e 1346 em abril de 2013 (índice de Shannon = 7,521; índice de Simpson = 0,988). Os valores altos de diversidade contidos na tabela 2 indicam que as populações de cada amostra são oriundas de uma comunidade microbiana complexa.

TABELA 2: Índices de diversidade das amostras do Lago São Francisco.

	Chao1	ACE	Shannon	Simpson	Dominância	Equitabilidade	Espécies observadas
Oct12	1.307,52	1.306,94	7,003	0,979	0,021	0,679	1.277
Nov12	1.359,78	1.355,44	7,239	0,982	0,018	0,698	1.327
Dec12	1.384,89	1.371,67	7,451	0,986	0,014	0,717	1.342
Jan13	1.316,67	1.302,37	7,153	0,982	0,018	0,694	1.272
Feb13	1.374,89	1.357,46	7,300	0,984	0,016	0,703	1.337
Mar13	1.352,26	1.340,89	7,412	0,986	0,014	0,716	1.308
Apr13	1.392,63	1.376,01	7,521	0,988	0,012	0,724	1.346
May13	1.383,98	1.368,35	7,152	0,979	0,021	0,689	1.340
Jun13	1.368,26	1.352,07	7,390	0,986	0,014	0,714	1.311
Jul13	1.326,88	1.323,74	7,262	0,984	0,016	0,703	1.292
Aug13	1.306,95	1.282,82	7,177	0,982	0,018	0,698	1.250
Sep13	1.308,63	1.290,35	6,975	0,978	0,022	0,678	1.255

A primeira análise acerca da similaridade foi a elaboração de uma matriz de distância através do algoritmo “Floresta Aleatória” (Random Forest – BREIMAN, 2001), utilizada como entrada em uma análise de coordenadas principais. A análise de coordenadas principais (*Principal Coordinates Analysis* – PCoA) é a forma clássica do escalonamento multidimensional (*Multidimensional scaling* – MDS). Trata-se de uma técnica de visualização com o objetivo de dispor num espaço de dimensionalidade reduzida (geralmente duas ou três dimensões) pontos, representando as amostras, preservando as distâncias tanto quanto possível (DINSDALE *et al.*, 2013). O resultado da análise de coordenadas principais pode ser conferido na figura 8.

A disposição das amostras no espaço bidimensional originou quatro grupos, três deles compostos por amostras de meses sequenciais. O primeiro grupo é formado pelas amostras de novembro e dezembro de 2012. As amostras dos meses de janeiro, fevereiro, março e abril de 2013 formam o segundo grupo. O terceiro grupo é formado pelas amostras de maio, junho e setembro de 2013. O último grupo

é formado por amostras julho e agosto de 2013. A amostra de outubro de 2012 não foi agrupada com nenhum dos grupos anteriores.

As 30 variáveis (OTUs) mais relevantes na obtenção dessa disposição estão representadas na figura 9 e descritas na tabela 3. Oito das variáveis puderam ser identificadas apenas como pertencentes ao domínio Bacteria. O filo mais representado entre essas variáveis foi Proteobacteria, oito ao todo, sendo uma Alphaproteobacteria (Rhizobiales), duas Betaproteobacteria (Comamonadaceae e Methylophilaceae) e uma Deltaproteobacteria (*Peredibacter*). Bacteroidetes foi o segundo filo mais representado, cinco OTUs, sendo duas da família Chitinophagaceae, uma *Ferruginibacter* e uma *Flavobacterium*. Já o terceiro filo mais representado entre as variáveis, Actinobacteria, contou com quatro OTUs, sendo duas *Ilumatobacter*. Outros filios representados foram OD1 (duas OTUs), Chloroflexi (uma OTU da família Anaerolineaceae), Planctomycetes (uma OTU da família Planctomycetaceae) e TM7 (uma OTU). O comportamento das três OTUs mais relevantes ao longo dos meses está registrado na figura 10.

FIGURA 8: Classificação das amostras do Lago São Francisco utilizando a técnica Random Forest e visualização através de multidimensional scaling.

FIGURA 9: Importância das variáveis determinada pela técnica Random Forest usando índice de Gini.

FIGURA 10: Variação da abundância das três OTUs mais relevantes para a classificação das amostras do Lago São Francisco utilizando a técnica Random Forest.

OTU ID	Diminuição Média no Índice de Gini	Táxon
OTU484555707	0,009051349	Bacteria; Proteobacteria
OTU978263524	0,008895492	Bacteria; Bacteroidetes; Sphingobacteria; Sphingobacteriales; Chitinophagaceae
OTU644793471	0,008886841	Bacteria; Bacteroidetes; Sphingobacteria; Sphingobacteriales; Chitinophagaceae
OTU610367246	0,008826746	Bacteria
OTU798462483	0,008799683	Bacteria; Bacteroidetes
OTU569076876	0,008790762	Bacteria; Bacteroidetes; Flavobacteria; Flavobacteriales; Flavobacteriaceae; <i>Flavobacterium</i>
OTU806832448	0,008789413	Bacteria; OD1; OD1_genera_incertae_sedis
OTU836170577	0,00877273	Bacteria; TM7; TM7_genera_incertae_sedis
OTU634756283	0,008751048	Bacteria
OTU804406631	0,00868781	Bacteria
OTU958801264	0,008664349	Bacteria; Proteobacteria
OTU743062318	0,008663254	Bacteria; Planctomycetes; Planctomycetacia; Planctomycetales; Planctomycetaceae
OTU725203347	0,008654635	Bacteria; Proteobacteria; Deltaproteobacteria; Bdellovibrionales; Bacteriovoracaceae; <i>Peredibacter</i>
OTU691143664	0,008653317	Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales; Comamonadaceae
OTU28933715	0,008651365	Bacteria
OTU946727579	0,008615016	Bacteria; OD1; OD1_genera_incertae_sedis
OTU598613314	0,008600698	Bacteria
OTU160455004	0,008599889	Bacteria; Bacteroidetes; Sphingobacteria; Sphingobacteriales; Chitinophagaceae; <i>Ferruginibacter</i>
OTU943159284	0,008528175	Bacteria
OTU19958044	0,008513159	Bacteria; Proteobacteria

TABELA 3: Descrição taxonômica das OTUs mais relevantes para a classificação das amostras do Lago São Francisco utilizando a técnica RandomForest.

OTU ID	Diminuição Média no índice de Gini	Táxon
OTU83993773	0,008488857	Bacteria; Actinobacteria; Actinobacteria; Actinomycetales
OTU107222178	0,008488635	Bacteria; Chloroflexi; Anaerolineae; Anaerolineales; Anaerolineaceae
OTU969027043	0,008467587	Bacteria; Actinobacteria; Actinobacteria
OTU171940722	0,008466698	Bacteria; Proteobacteria
OTU531589351	0,008465032	Bacteria; Actinobacteria; Actinobacteria; Acidimicrobidae_incertae_sedis; <i>Ilumatobacter</i>
OTU599026782	0,008452683	Bacteria
OTU262325809	0,008446143	Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales
OTU265594466	0,008440302	Bacteria; Actinobacteria; Actinobacteria; Acidimicrobidae_incertae_sedis; <i>Ilumatobacter</i>
OTU567628705	0,008435476	Bacteria; Proteobacteria; Betaproteobacteria; Methylophilales; Methylophilaceae
OTU372794185	0,008417317	Bacteria

TABELA 3 cont.: Descrição taxonômica das OTUs mais relevantes para a classificação das amostras do Lago São Francisco utilizando a técnica RandomForest.

Após essa primeira fase de análise acerca da similaridade, mais testes foram executados para verificar a consistência desse primeiro resultado. Em seguida foi elaborada uma matriz de distância de Bray-Curtis entre as amostras (amplamente utilizada em Ecologia) e utilizada como entrada em uma análise de coordenadas principais. O resultado da análise de coordenadas principais utilizando a distância de Bray-Curtis pode ser conferido na figura 11.

FIGURA 11: Classificação das amostras do Lago São Francisco através de análise de coordenadas principais a partir de distância Bray Curtis.

Ao todo 60% da variação foi representada nas três primeiras coordenadas: 32%, 15% e 13% respectivamente. É possível destacar quatro grupos, três deles compostos por amostras de meses sequenciais. O primeiro grupo é formado pelas amostras de novembro e dezembro de 2012. As amostras dos meses de janeiro, fevereiro, março e abril de 2013 formam o segundo grupo, composto por dois subgrupos, janeiro e março e fevereiro e abril. O terceiro grupo é formado pelas

amostras de junho e setembro, agosto e julho de 2013. O último grupo é formado por amostras de meses não sequenciais, outubro de 2012 e maio de 2013.

A distância UniFrac (LOZUPONE & KNIGHT, 2005) é um método de estabelecimento de distância de amostras amplamente utilizado em Ecologia Microbiana e reúne os requerimentos formais como métrica de distância: sempre fornece valores não-negativos, apresenta simetria entre as amostras e satisfaz a desigualdade triangular e amostras idênticas têm valor igual a 0 (LOZUPONE *et al.*, 2011). O método apresenta duas versões, distância não-ponderada e ponderada, sendo que a primeira não considera e a segunda considera a abundância relativa de cada OTU nas amostras e portanto é mais sensível aos táxons mais abundantes.

O resultado da análise de coordenadas principais utilizando a matriz de distância UniFrac não-ponderada pode ser conferido na figura 12A. As três coordenadas principais compreendem 65% da variação: 39%, 15% e 11% respectivamente. As amostras apresentaram disposição mais ampla em relação à análise utilizando distância de Bray-Curtis, contudo é possível notar que amostras de meses subsequentes tendem a ser posicionadas umas próximas das outras. A amostra de outubro de 2012 apresentou a maior distância em relação às demais, em seguida, as amostras de novembro e dezembro de 2012 agruparam-se bastante próximas entre si e ligeiramente mais próximas das amostras de janeiro e fevereiro de 2013 do que da de outubro. Começando pela amostra de março de 2013, mais próxima da de fevereiro, apresentam-se de forma sequencial março, abril e maio de 2013. As amostras de junho, julho, agosto e setembro de 2013 formam um subgrupo separado das demais, sendo junho a mais próxima da de maio, a julho equidistante da de junho e da de agosto. Curiosamente a amostra de setembro de 2013 apresentou mais proximidade com a amostra de junho de 2013.

FIGURA 12: Classificação das amostras do Lago São Francisco através de análise de coordenadas principais. A: Agrupamento obtido a partir de distância UniFrac não-ponderada. B: Agrupamento obtido a partir de distância UniFrac ponderada.

O maior nível de variação captado por análise de coordenadas foi obtido a partir de matriz de distância UniFrac ponderada (FIGURA 12B). As três coordenadas principais compreendem 69% da variação: 45%, 14% e 10% respectivamente. As amostras apresentaram disposição tão ampla quanto à análise anterior, porém as amostras mais próximas nem sempre foram de meses subsequentes, como foi o caso das amostras de abril e de agosto de 2013 e outubro de 2012, maio e junho de 2013. Assim como na análise anterior as amostras de novembro e dezembro de 2012 e janeiro, fevereiro e março de 2013 formaram grupos, com algumas das distâncias similares às aquelas da análise anterior. As amostras de julho e setembro de 2013 não agruparam com as demais. As amostras provenientes de novembro e dezembro de 2012, bem como as de janeiro, fevereiro e março de 2013 agruparam-se nas análises de coordenadas principais independentemente das matrizes de distâncias utilizadas como entrada. E ainda, as amostras provenientes de julho e agosto de 2013 agruparam-se nas análises de coordenadas principais em três das quatro matrizes de distâncias utilizadas (exceção para UniFrac ponderada).

Para verificar se amostras coletadas na mesma estação do ano apresentam mais similaridade entre si do que com as demais (FIGURA 13), foi aplicada a análise de similaridade (*Analysis of similarity* – ANOSIM; CLARKE, 1993) implementada no pacote Qiime. O resultado da análise de similaridade resultou $R = 0,6512$ ($p\text{-valor} < 5 \times 10^{-5}$) e $R = 0,6259$ ($p\text{-valor} < 4,2 \times 10^{-3}$) utilizando as distâncias UniFrac não-ponderada e ponderada respectivamente. Esses valores suportam estatisticamente a conclusão de que a comunidade microbiana presente no Lago São Francisco em um determinado mês se assemelha mais às presentes em meses da mesma estação do ano. Muitos estudos de microbiomas aquáticos apontam uma variação sazonal dos grupos microbianos relacionada à disponibilidade de nutrientes ou a fatores ambientais (RAM *et al.*, 2005; ALONSO-SÁEZ & GASOL, 2007; VILA-COSTA *et al.*, 2007; WILKES *et al.*, 2009). A variação sazonal da abundância de organismos autotróficos no Mar Cantábrico, Espanha, foi identificada entre 2002 e 2003 onde a ausência de *Prochlorococcus* na primavera e verão mostrou-se relacionada com a hidrologia do que as condições nutricionais enquanto a abundância de bactérias heterotróficas foi menos variável (CALVO-DÍAZ & MORÁN, 2006). Um outro estudo realizado no Mar Mediterrâneo demonstrou que nas condições oligotrófica do verão bactérias livres nadantes dominaram enquanto

bactérias aderidas a partículas foram muito menos abundantes embora predominantemente responsáveis por atividades celulares específicas e relativamente constantes, indicando uma colonização de partículas por espécies ubíquas (GHIGLIONE *et al.*, 2007). Um estudo realizado em águas temperadas na costa do Reino Unido entre os anos de 2003 e 2008 identificou com maior frequência as ordens Rickettsiales e Rhodobacterales, ambas pertencentes à classe Alphaproteobacteria, além de padrões recorrentes na comunidade bacteriana como um todo com pico de diversidade no inverno (GILBERT *et al.*, 2012).

Além da duração dos dias e da temperatura média, a pluviosidade também é um fator contribuinte para a caracterização das estações. O suporte estatístico obtido na separação das amostras por estações não foi obtido quando separamos as amostras em estação chuvosa (outubro de 2012 a março de 2013) e seca (abril a setembro de 2013): $R = 0,4074$ ($p\text{-valor} < 2,5 \times 10^{-2}$) e $R = 0,1$ ($p\text{-valor} < 1,8 \times 10^{-1}$) para distâncias UniFrac não-ponderada e ponderada respectivamente. O baixo suporte estatístico pode ser atribuído a variação anual da pluviosidade bem como a definição pouco marcante de um período chuvoso e um seco em um clima temperado relativamente constante.

FIGURA 13: Classificação das amostras do Lago São Francisco através de análise de coordenadas principais. A: Agrupamento obtido a partir de distância UniFrac não-ponderada. B: Agrupamento obtido a partir de distância UniFrac ponderada. Verde: primavera; vermelho: verão; amarelo: outono; ciano: inverno.

Coocorrência e Correlação de Espécies nas Amostras do Lago São Francisco através de Sequenciamento *Shotgun*

Na análise de similaridade local (LSA), 3056 das coocorrências envolvendo 556 espécies identificadas apresentaram um valor de correlação bastante significativo: 3007 apresentaram valor de correlação acima 0,9 (p-valores entre 0 e $2,0 \times 10^{-3}$) envolvendo 545 espécies e 49 apresentaram valor de correlação abaixo de -0,9 (p-valores entre 0 e $5,0 \times 10^{-4}$) envolvendo 60 espécies.

Para extrair algum significado biológico da coocorrência estatisticamente relevante das espécies, o resultado da análise de correlação foi disposto em um grafo não-dirigido onde os nós representam as espécies e as arestas as coocorrências, formando uma rede composta por 51 componentes (FIGURA 14). Ao todo 22 filos foram representados, Proteobacteria (236 espécies), Actinobacteria (124 espécies), Bacteroidetes (67 espécies), Firmicutes (45 espécies), Cyanobacteria (22 espécies) e Spirochaetes (10 espécies), Deinococcus-Thermus (9 espécies), Acidobacteria (7 espécies), Euryarchaeota e Chlorobi (5 espécies), Planctomycetes e Chlamydiae (4 espécies), Deferribacteres (3 espécies), Synergistetes, Ignavibacteriae, Chloroflexi e Aquificae (2 espécies), Verrucomicrobia, Thermotogae, Nitrospirae, Gemmatimonadetes e Chrysiogenetes (1 espécie). As espécies *Candidatus* Cloacamonas acidaminovorans e *Candidatus* Methylomirabilis oxyfera ainda não possuem filo definido. Uma inspeção visual da rede mostra regiões densas, com muitos nós interconectados. Para analisar com mais propriedade tais regiões, a rede foi dividida em módulos. O processo resultou em 60 componentes e a distribuição do número de espécies está registrada na figura 15. Curiosamente, com exceção da maior sub-rede, com 127 espécies, todas as sub-redes com mais de 10 espécies são quase exclusivamente formadas por um filo (FIGURA 16).

FIGURA 14: Rede de espécies correlacionais presentes no Lago São Francisco. Os nós representam as espécies sendo o diâmetro proporcional à abundância da espécie. As arestas preenchidas representam correlação acima de 0,9 e as arestas hachuradas representam correlação abaixo de -0,9. Vermelho: Proteobacteria; azul: Actinobacteria; amarelo: Bacteroidetes; verde: Cyanobacteria; lilás: Firmicutes.

FIGURA 15: Distribuição das sub-redes de espécies correlacionas presentes no Lago São Francisco.

FIGURA 16: Sub-redes de espécies correlacionas presentes no Lago São Francisco. Vermelho: Proteobacteria; azul: Actinobacteria; amarelo: Bacteroidetes.

O fato de as sub-redes serem compostas quase exclusivamente por populações de espécies filogeneticamente relacionadas é um indicativo do papel de fatores ambientais no Lago São Francisco ao invés da pura interação entre espécies. Estudos já apontaram a predisposição de alguns grandes grupos por determinados substratos (SALCHER *et al.*, 2008; PÉREZ *et al.*, 2010; ECKERT *et al.*, 2012).

Um estudo sobre o consumo de dois substratos, leucina e timidina, por comunidades bacterianas de dois lagos alpinos austríacos detectou que 70 a 80% das células bacterianas captavam moléculas de leucina, enquanto apenas 15 a 43% captavam timidina (PÉREZ *et al.*, 2010). No caso das moléculas de timidina, acima de 80% das células consumidoras pertenciam ao filo Actinobacteria. Após o florescimento do fitoplâncton em um lago pré-alpino durante a estação da primavera, detectou uma transição de uma população predominante composta de espécies de Cytophaga–Flavobacteria e de Betaproteobacteria para uma população predominante composta por espécies de Actinobacteria resistentes à predação ou filamentosas, com impacto direto no consumo de N-acetil-glicosamina (ECKERT *et al.*, 2012). Nesse caso, com o florescimento do fitoplâncton e com o consequente aumento de nanoflagelados heterotróficos, espécies dos filos Actinobacteria e LD2 beneficiariam-se, não somente pela maior disponibilidade de N-acetil-glicosamina oriunda de restos celulares de bactérias consumidas, mas também pela eliminação de competição (ECKERT *et al.*, 2013). Nesse contexto é possível assumir que os mecanismos de obtenção de nutrientes, a susceptibilidade ou resistência à predação por protozoários e à infecção viral e os mecanismos de resposta a fatores ambientais são compartilhados desde o nível de filo, sendo portanto características evolutivas conservadas.

Como caso de estudo, a sub-rede composta por espécies do gênero *Mycobacterium* foi escolhida (FIGURA 17). O gênero *Mycobacterium* variou entre aproximadamente 2% e 9% das sequências classificadas como Bacteria (13% e 34% das sequências de Actinobacteria) nas amostras colhidas em dezembro de 2012 e julho de 2013 respectivamente. Em média, a espécie identificada com maior frequência foi *Mycobacterium smegmatis* (1,32%), seguida por *Mycobacterium gilvum* (0,51%) e *Mycobacterium rhodesiae* (0,47%). Embora *Ilumatobacter coccineum* seja de longe a espécie mais abundante de Actinobacteria, a abundância

de *Mycobacterium smegmatis*, *Mycobacterium gilvum* e *Mycobacterium rhodesie*, e do gênero *Mycobacterium* como um todo, é atípica (NEWTON *et al.*, 2011).

FIGURA 17: Sub-rede de espécies do gênero *Mycobacterium* presentes no Lago São Francisco. Os nós representam as espécies sendo o diâmetro proporcional à abundância da espécie. As arestas preenchidas representam correlação acima de 0,9 e as arestas hachuradas representam correlação abaixo de -0,9.

A variação da população de *Mycobacterium smegmatis*, *Mycobacterium gilvum* e *Mycobacterium rhodesie* está registrada na figura 18. Houve uma queda brusca da população entre outubro e novembro de 2012, seguida de uma queda sutil em dezembro. Após um crescimento sutil em janeiro de 2013, a população aumentou de forma acentuada até maio, apresentou uma ligeira queda em junho e atingiu o pico do ciclo em julho. Em agosto houve uma queda acentuada seguida de uma ligeira recuperação da população.

FIGURA 18: Variação da abundância das três espécies mais abundantes do gênero *Mycobacterium* presentes no Lago São Francisco.

Por conter uma série de espécies patogênicas o gênero *Mycobacterium* é bastante estudado sob a perspectiva médica, mas pouco se sabe sobre seu papel nas comunidades microbianas ambientais. As espécies mais estudadas, entre elas *Mycobacterium tuberculosis*, são descritas na literatura em relação a sua capacidade de responder ao estresse causado por hipóxia (RAO *et al.*, 2008; GUMBER *et al.*, 2009), pH altamente ácidos (GEBHARD *et al.*, 2008; VANDAL *et al.*, 2009; LEEAH *et al.*, 2010; ABRAMOVITCH *et al.*, 2011), temperatura (ANUCHIN *et al.*, 2009) e resistência à predação (RØNN *et al.*, 2001; BLAND *et al.*, 2005). A baixa abundância das espécies do gênero pode ser atribuída a um aumento da temperatura da água do lago e do ar diminuindo a concentração de oxigênio dissolvido (BLAND *et al.*, 2005; NEWTON *et al.*, 2011). Os meses de outubro, novembro e dezembro do ano de 2012 apresentaram temperatura média do ar acima da média histórica segundo o boletim climatológico anual da estação meteorológica do Instituto de Astronomia, Geofísica e Ciências Atmosféricas da Universidade de São Paulo (<http://www.estacao.iag.usp.br/Boletins/2012.pdf>) e um relatório técnico de dezembro de 2012 fornecido pela FPZSP indica uma concentração de oxigênio dissolvido de 1,82 mg O₂/L.

Análise do Potencial Metabólico do Lago Francisco através de Sequenciamento *Shotgun*

Utilizando o programa HUMAnN foi possível estabelecer um panorama geral do potencial metabólico e identificar diferenças importantes nas amostras coletadas ao longo do período de um ano. O agrupamento hierárquico das amostras em função do mês de coleta resultou em dois grandes grupos (FIGURA 19). O primeiro é constituído pelas amostras de junho e de janeiro de 2013, seguidas por outubro de 2012 e julho de 2013 e finalmente novembro de 2012. O segundo grande grupo é constituído pelas demais amostras na seguinte ordem: maio e março de 2013, fevereiro e agosto de 2013 agrupadas com abril de 2013, dezembro de 2012 e setembro de 2013.

Visando identificar a variação do potencial metabólico da microbiota do lago, refletida na abundância relativa de módulos em cada amostra, e ainda quão relevante é esta variação, valores discrepantes de abundância (*outliers*) foram identificados como quaisquer valores abaixo ou acima de 1.5 x distância interquartil.

A maioria dos valores *outliers* superiores foram referentes à amostra de novembro de 2012 (TABELA 4). Esses *outliers* podem ser distribuídos em 15 categorias KEGG: metabolismo central de carboidrato (M00001, M00002, M00003 e M00005), metabolismo de cisteína e metionina (M00021), metabolismo de pirimidina (M00052), metabolismo de aminoácidos aromáticos (M00024 e M00025), metabolismo de lipopolissacarídeo (M00063 e M00064), metabolismo de ácido graxo (M00083), síntese de cofatores e vitaminas (M00115, M00119, M00123 e M00125), síntese de poliaminas (M00133), síntese de ATP (M00144, M00151 e M00156), fixação de carbono (M00169), sistemas de transporte de fosfato e aminoácidos (M00222), sistemas de transporte de cátion metálico, ferro-sideróforo e vitamina B12 (M00240 e M00242), sistemas ABC-tipo 2 e outros transportadores (M00250, M00254, M00255, M00259 e M00320), aminoacil tRNA (M00359 e M00360), síntese de terpenóides (M00364). Como mencionado acima, a estrutura da comunidade microbiana sofreu uma modificação drástica entre setembro e novembro de 2012, possivelmente relacionada a fatores climáticos e essa modificação influenciou profundamente o potencial metabólico dessa amostra em particular.

FIGURA 19: *Heatmap* representando a abundância dos módulos KEGG nas amostras do Lago São Francisco. Cada um dos 101 módulos presentes nas amostras está representado horizontalmente enquanto as amostras estão representadas verticalmente. O gráfico foi gerado através da plataforma R, módulo *heatmap*, cada módulo foi normalizado de forma que a média dos valores correspondesse a zero e o desvio padrão a 1.

		Oct12	Nov12	Dec12	Jan13	Feb13	Mar13	Apr13	May13	Jun13	Jul13	Aug13	Sep13
M00001	Glycolysis (Embden-Meyerhof pathway)	146	175	131	148	127	134	133	131	159	132	126	110
	Glycolysis, core module involving 3C compounds	190	217	168	179	158	162	167	156	186	170	162	133
M00002													
M00003	Gluconeogenesis	167	191	152	159	144	139	136	136	172	151	140	119
M00005	PRPP biosynthesis	195	222	190	178	171	172	172	171	189	188	188	148
M00021	Cysteine biosynthesis	117	130	94	90	92	84	92	91	110	105	91	72
M00024	Phenylalanine biosynthesis	83	92	64	59	57	49	48	52	67	65	52	38
M00025	Tyrosine biosynthesis	91	98	68	65	65	52	52	59	74	73	58	45
M00052	Pyrimidine ribonucleotide biosynthesis	185	216	163	175	150	161	140	147	160	168	145	125
M00063	CMP-KDO biosynthesis	95	126	86	109	79	76	76	85	89	88	78	64
M00064	ADP-L-glycero-D-manno-heptose biosynthesis	105	129	89	88	80	72	67	71	84	82	76	60
M00083	Fatty acid biosynthesis, elongation	157	194	150	155	134	140	118	134	148	137	126	105
M00115	NAD biosynthesis	141	169	117	143	119	125	113	113	136	120	121	102
M00119	Pantothenate biosynthesis	92	111	84	94	73	78	75	84	86	74	61	55
M00123	Biotin biosynthesis	83	95	62	59	61	57	48	62	70	64	54	47
M00125	Riboflavin biosynthesis	145	182	129	144	121	123	120	123	156	130	122	106
M00133	Polyamine biosynthesis	32	45	28	36	29	28	22	32	30	31	29	29
M00144	Complex I (NADH dehydrogenase)	177	206	160	172	153	155	164	150	182	169	154	128
M00151	Complex III (Cytochrome bc1 complex)	102	123	89	81	88	59	72	73	86	97	70	52
M00156	Complex IV (Cytochrome c oxidase)	47	75	49	45	48	39	40	54	49	73	29	33
M00169	CAM (Crassulacean acid metabolism), light	79	108	90	77	79	75	68	67	80	71	71	65
M00222	Phosphate transport system	164	190	147	144	144	137	139	137	177	156	141	131
M00240	Iron complex transport system	30	41	29	23	21	19	20	25	27	23	21	18
M00242	Zinc transport system	21	36	20	25	21	20	21	22	34	25	23	25
M00248	Antibiotic transport system	80	97	68	71	62	66	61	61	73	69	61	51
M00250	Lipopolysaccharide transport system	88	101	78	80	73	78	79	69	84	75	79	63
M00254	ABC-2 type transport system	55	79	61	70	53	60	50	55	56	49	50	43
M00255	Lipoprotein-releasing system	90	116	92	101	81	79	68	80	86	84	76	60
M00259	Heme transport system	74	95	64	62	55	40	42	49	63	67	51	40
M00320	Lipopolysaccharide export system	83	102	71	67	65	44	45	57	68	75	60	42
M00359	Aminoacyl-tRNA biosynthesis, eukaryotes	153	198	158	166	146	151	143	141	152	150	146	120
M00360	Aminoacyl-tRNA biosynthesis, prokaryotes	174	214	170	178	160	164	159	153	171	167	160	132
M00364	C10-C20 isoprenoid biosynthesis, bacteria	34	47	31	50	27	44	30	35	30	27	32	26

TABELA 4: Distribuição dos valores *outliers* de módulos KEGG super-representados na amostra de novembro de 2012 no Lago São Francisco.

Para ampliar o conhecimento acerca do potencial metabólico da comunidade microbiana do lago e identificar suas particularidades, dados metagenômicos diversos foram recuperados da plataforma MG-RAST (MEYER *et al.*, 2008) (TABELA 5).

Tabela 5: Dados metagenômicos recuperados da plataforma MG-RAST para comparação metabólica.

Metagenoma	MG-RAST id	Localização	Referência
Lake Gatun	4441590.3	Panama	Rusch <i>et al.</i> , 2007
Lake Ace	4443684.3	Antarctica	Ng <i>et al.</i> , 2010
Albufera0P8	4516289.3	Spain	Ghai <i>et al.</i> , 2012
Albufera3P0	4516290.3	Spain	Ghai <i>et al.</i> , 2012

Após a aplicação do programa HUMAnN foi possível estabelecer um panorama bastante diferente do anterior (FIGURA 20). A abundância relativa dos módulos da comunidade microbiana do Lago Gatun é bem maior que das demais amostras, seguida das abundâncias dos módulos das comunidades de Albufera e do Lago Ace. Contudo, o agrupamento hierárquico das amostras do Lago São Francisco comportou-se de forma bastante semelhante: dois grandes grupos, o primeiro formado pelas amostras de abril, março, maio, agosto, fevereiro e setembro de 2013 e o segundo formado pelas amostras de junho de 2013, julho, outubro e janeiro de 2013, dezembro e novembro de 2012.

Muitos dos módulos super- ou sub-representados se encontram nas amostras adicionadas, evidenciando a grande diferença entre esses ambientes e o Lago São Francisco.

FIGURA 20: *Heatmap* representando a abundância dos módulos KEGG nas amostras do Lago São Francisco e outros metagenomas de amostras de água doce. Cada um dos 119 módulos presentes nas amostras está representado horizontalmente enquanto as amostras estão representadas verticalmente. O gráfico foi gerado através da plataforma R, módulo *heatmap*, cada módulo foi normalizado de forma que a média dos valores correspondesse a zero e o desvio padrão a 1.

A amostra de novembro de 2012 apresentou apenas nove módulos super-representados distribuídos em sete categorias KEGG: metabolismo de aminoácidos aromáticos (M00024), metabolismo de pirimidinas (M00051), metabolismo de lipopolissacarídeo (M00064), síntese de cofator e vitamina (M00117), síntese de

ATP (M00144 e M00156), fixação de carbono (M00169) e aminoacil-tRNA (M00359 e M00360).

Alguns módulos merecem destaque por estarem sub- ou super-representados em todos os metagenomas adicionados à análise e serem assim particularmente úteis para o entendimento da comunidade microbiana do Lago São Francisco. Entre eles é possível destacar os módulos referentes à síntese de β -caroteno (M00097), ao ciclo redutivo das pentoses-fosfato (M00167) e à fixação de nitrogênio (M00175): a super-representação em outros ambientes indicam que a comunidade microbiana do Lago São Francisco é possivelmente menos capaz de realizar fixação de carbono e de nitrogênio em decorrência do grau de eutrofização.

Os dados apontaram o Lago São Francisco como um ambiente extremamente variável, sujeito a uma série de fatores ambientais. As análises indicaram também que se trata de um ambiente diverso onde as populações de microrganismos variam mais em função de tais fatores do que em função da interação entre espécies. Futuras análises deverão incluir novas abordagens para detectar as interações interespecíficas e a contribuição dos táxons no potencial metabólico do Lago São Francisco.

REFERÊNCIAS

- ABUBUCKER, S.; SEGATA, N.; GOLL, J.; SCHUBERT, A. M.; IZARD, J.; CANTAREL, B. L.; RODRIGUEZ-MUELLER, B.; ZUCKER, J.; THIAGARAJAN, M.; HENRISSAT, B.; WHITE, O.; KELLEY, S. T.; METHE, B.; SCHLOSS, P. D.; GEVERS, P.; MITREVA, M.; HUTTENHOWER, C. Metabolic Reconstruction for Metagenomic Data and Its Application to the Human Microbiome. **PLoS Computational Biology**, v. 8, n. 6, p. 1 – 17, 2012.
- ABRAMOVITCH, R. B.; ROHDE, K. H.; HSU, F-F.; RUSSELL, D. G. aprABC: a Mycobacterium tuberculosis complex-specific locus that modulates pH-driven adaptation to the macrophage phagosome. **Molecular Microbiology**, v. 80, n. 3, p. 678 – 694, 2011.
- ALONSO-SÁEZ, L.; GASOL, J. M. Seasonal Variations in the Contributions of Different Bacterial Groups to the Uptake of Low-Molecular-Weight Compounds in Northwestern Mediterranean Coastal Waters. **Applied and Environmental Microbiology**, v. 73, n. 11, p. 3528 – 3535, 2007.
- ANUCHIN, A. M.; MULYUKIN, A. L.; SUZINA, N. E.; DUDA, V. I.; EL-REGISTAN, G. I.; KAPRELYANTS, A. S. Dormant forms of Mycobacterium smegmatis with distinct morphology. **Microbiology**, v. 155, n. 4, p. 1071 – 1079, 2009.
- BAGATINI, I. L.; EILER, A.; BERTILSSON, S.; KLAVENESS, D.; TESSAROLLI, L. P.; VIEIRA, A. A. H. Host-Specificity and Dynamics in Bacterial Communities Associated with Bloom-Forming Freshwater Phytoplankton. **PLoS ONE**, v. 9, n. 1, p. 1 – 13, 2014.
- BLAND, C. S.; IRELAND, J. M.; LOZANO, E.; ALVAREZ, M. E.; PRIMM, T. P. Mycobacterial ecology of the Rio Grande. **Applied and Environmental Microbiology**, v. 71, n. 10, p. 5719 – 5727.
- BREIMAN, L. Random Forests. **Machine Learning**, v. 45, n. 1, p. 5 – 32, 2001.
- BURKERT, U.; WARNECKE, F.; BABENZIEN, D.; ZWIRNMANN, E.; PERNTHALER, J. Members of a Readily Enriched β -Proteobacterial Clade Are Common in Surface Waters of a Humic Lake. **Applied And Environmental Microbiology**, v. 69, n. 11, p. 6550 – 6559, 2003.
- CALVO-DÍAZ, A.; MORÁN, X. A. G. Seasonal dynamics of picoplankton in shelf Waters of the southern Bay of Biscay. **Aquatic Microbial and Ecology**, v. 42, p. 159 – 174, 2006.
- CLARKE, K. R. Non-parametric multivariate analyses of changes in community structure. **Australian Journal of Ecology**, v.18, p. 117 – 143, 1993.
- CAPORASO, J. G.; KUCZYNSKI, J.; STOMBAUGH, J.; BITTINGER, K.; BUSHMAN, F. D.; COSTELLO, E. K.; FIERER, N.; PEÑA, A. G.; GOODRICH, J. K.; GORDON, J. I.; HUTTLEY, G. A.; KELLEY, S. T.; KNIGHTS, D.; KOENIG, J. E.; LEY, R. E.;

LOZUPONE, C. A.; MCDONALD, D.; MUEGGE, B. D.; PIRRUNG, M.; REEDER, J.; SEVINSKY, J. R.; TURNBAUGH, P. T.; WALTERS, W. A.; WIDMANN, J.; YATSUNENKO, T.; ZANEVELD, J.; KNIGHT, R. QIIME allows analysis of high-throughput community sequencing data. **Nature Methods**, v. 7, p. 335 – 336, 2010.

DINSDALE, E. A.; EDWARDS, R. A.; BAILEY, B. A.; TUBA, I.; AKHTER, S.; MCNAIR, K.; SCHMIEDER, R.; APKARIAN, N.; CREEK, M.; GUAN, E.; HERNANDEZ, M.; ISAACS, K.; PETERSON, C.; REGH, T.; PONOMARENKO, V. Multivariate analysis of functional metagenomes. **Frontiers in Genetics**, v. 4, p. 1 – 25, 2013.

ECKERT, E. M.; SALCHER, M. M.; POSCH, T.; EUGSTER, B.; PERNTHALER, J. Rapid successions affect microbial N-acetyl-glucosamine uptake patterns during a lacustrine spring phytoplankton bloom. **Environmental Microbiology**, v. 14, n. 3, p. 794 – 806, 2012.

ECKERT E. M.; BAUMGARTNER, M.; HUBER, I. M.; PERNTHALER, J. Grazing resistant freshwater bacteria profit from chitin and cell-wall-derived organic carbon. **Environmental Microbiology**, v. 15, n. 7, p. 2019 – 2030, 2013.

EDGAR, R. C. Search and clustering orders of magnitude faster than BLAST. **Bioinformatics**, v. 26, n.19, p. 2460 – 2461, 2010.

EILER, A.; BERTILSSON, S. Composition of freshwater bacterial communities associated with cyanobacterial blooms in four Swedish lakes. **Environmental Microbiology**, v. 6, p. 1228 – 1243, 2004.

EILER, A.; BERTILSSON, S. 2007. Flavobacteria blooms in four eutrophic lakes: linking population dynamics of freshwater bacterioplankton to resource availability. **Applied and Environmental Microbiology**, v. 73, p. 3511 – 3518, 2007.

EILER, A., ZAREMBA-NIEDZWIEDZKA, K.; MARTÍNEZ-GARCÍA, M.; MCMAHON, K. D.; STEPANAUSKAS, R.; ANDERSSON, S. G. E.; BERTILSSON, S. Productivity and salinity structuring of the microplankton revealed by comparative freshwater metagenomics. **Environmental Microbiology**, v. 16, p. 9, p. 2682 – 2698, 2014.

GEBHARD, S.; HÜMPEL, A.; MCLELLAN, A. D.; COOK, G. M. The alternative sigma factor SigF of *Mycobacterium smegmatis* is required for survival of heat shock, acidic pH and oxidative stress. **Microbiology**, v. 154, n. 9, p. 2786 – 2795, 2008.

GHAJ, R.; HERNANDEZ, C. M.; PICAZO, A.; MIZUNO, C. M.; ININBERGS, K.; DÍEZ, B.; VALAS, R.; DUPONT, C. L.; MCMAHON, K. D.; CAMACHO, A.; RODRIGUEZ-VALERA, F. Metagenomes of Mediterranean Coastal Lagoons. **Scientific Reports**, v. 2, n. 490, p. 1 – 13, 2012.

GHIGLIONE, J. F.; MEVEL, G.; PUJO-PAY, M.; MOUSSEAU, L.; LEBARON, P.; GOUTX, M. Diel and Seasonal Variations in Abundance, Activity, and Community Structure of Particle-Attached and Free-Living Bacteria in NW Mediterranean Sea. **Microbial Ecology**, v. 54, n. 2, p. 217 – 231, 2007.

GILBERT, J. A.; STEELE, J. A.; CAPORASO, J. G.; STEINBRÜCK, L.; REEDER, J.; TEMPERTON, B.; HUSE, S.; MCHARDY, A. C.; KNIGHT, R.; JOINT, I.; SOMERFIELD, P.; FUHRMAN, J. A.; FIELD, D. Defining seasonal marine microbial community dynamics. **The ISME Journal**, v. 6, p. 298 – 308, 2012.

GLASS, E. M.; WILKENING, J.; WILKE, A.; ANTONOPOULOS, D.; MEYER, F. Using the metagenomics RAST server (MG-RAST) for analyzing shotgun metagenomes. **Cold Spring Harbor protocols 2010**, pdb prot5368, 2010.

GORDON-BRADLEY, N.; LI, N.; WILLIAMS, H. N. Bacterial community structure in freshwater springs infested with the invasive plant species *Hydrilla verticillata*. **Hydrobiologia**, v. 742, p. 221 – 232, 2015.

GUMBER, S.; TAYLOR, D. L.; MARSH, I. B.; WHITTINGTON, R. J. Growth pattern and partial proteome of *Mycobacterium avium* subsp. *paratuberculosis* during the stress response to hypoxia and nutrient starvation. **Veterinary Microbiology**, v. 133, n. 4, p. 344 – 357, 2009.

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. **Chemistry & Biology**, v. 5, p. 245 – 249, 1998.

HAHN, M. W.; SCHEUERL, T.; JEZBEROVÁ, J.; KOLL, U.; JEZBERA, J.; ŠIMEK, K.; VANNINI, C.; PETRONI, G.; WU, Q. L. The Passive Yet Successful Way of Planktonic Life: Genomic and Experimental Analysis of the Ecology of a Free-Living *Polynucleobacter* Population. **PLoS ONE**, v. 7, n. 3, p. 1 – 17, 2012.

HUTALLE-SCHMELZER, K. M. L.; ZWIRNMANN, E.; KRÜGER, A.; GROSSART, H-P. Enrichment and cultivation of pelagic bacteria from a humic lake using phenol and humic matter additions. **FEMS Microbiology Ecology**, v. 72, n. 1, p. 58 – 73, 2010.

JEZBERA, J.; JEZBEROVÁ, J.; BRANDT, U.; HAHN, M. W. Ubiquity of *Polynucleobacter necessarius* subspecies *asymbioticus* results from ecological diversification. **Environmental Microbiology**, v. 13, n. 4, p. 922 – 931, 2011.

KANEHISA, M.; GOTO, S.; KEGG: Kyoto Encyclopedia of Genes and Genomes. **Nucleic Acids Research**, v. 28, n. 1, p. 27 – 30, 2000.

KLINDWORTH, A.; PRUESSE, E.; SCHWEER, T.; PEPLIES, J.; QUAST, C.; HORN, M.; GLOCKNER, F. O. Evaluation of general 16S ribosomal RNA gene PCR primers for classical and next-generation sequencing-based diversity studies. **Nucleic Acids Research**, v. 41, n. 1, p. 1 – 11, 2013.

KOLMONEN, E.; SIVONEN, K.; RAPALA, J.; HAUKKA, K. Diversity of cyanobacteria and heterotrophic bacteria in cyanobacterial blooms in Lake Joutikas, Finland. **Aquatic Microbial Ecology**, v. 36, p. 201 – 211, 2004.

LEE, E-S.; YOON, T-H.; LEE, M-Y.; HAN, S-H.; KA, J-O. Inactivation of environmental mycobacteria by free chlorine and UV. **Water Research**, v. 44, n. 5, p. 1329 – 1334, 2010.

LOZUPONE, C.; KNIGHT, R. UniFrac: a New Phylogenetic Method for Comparing Microbial Communities. **Applied and Environmental Microbiology**, 2005 Dec; v. 71, n. 12, p. 8228 – 8235, 2005.

LOZUPONE, C.; LLADSER, M. E.; KNIGHTS, D.; STOMBAUGH, J.; KNIGHT, R. UniFrac: an effective distance metric for microbial community comparison. **The ISME Journal**, v. 5, n. 2, p. 169 – 172, 2011.

LUO, C.; RODRIGUEZ-R, L. M.; KONSTANTINIDIS, K. T. MyTaxa: an advanced taxonomic classifier for genomic and metagenomic sequences. **Nucleic Acids Research**, v. 42, n. 8, p. 1 – 12, 2014.

MARTINEZ-GARCIA, M.; SWAN, B. K.; POULTON, N. J.; GOMEZ, M. L.; MASLAND, D.; SIERACKI, M. E.; STEPANAUSKAS, R. High-throughput single-cell sequencing identifies photoheterotrophs and chemoautotrophs in freshwater bacterioplankton. **The ISME Journal**, v. 6, p. 113 – 123, 2012.

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**, v. 9, n. 386, 2008.

MORRIS, J. H.; APELTSIN, L.; NEWMAN, A. M.; BAUMBACH, J.; WITTKOP, T.; SU, G.; BADER, G. D.; FERRIN, T. E. clusterMaker: a multi-algorithm clustering plugin for Cytoscape. **BMC Bioinformatics**, v. 12, n. 436, p. 1 – 14, 2011.

NEWTON, R. J.; JONES, S. E.; EILER, A.; MCMAHON, K. D.; BERTILSSON, S. A Guide to the Natural History of Freshwater Lake Bacteria. **Microbiology and Molecular Biology Reviews**, v. 75, n. 1, p. 14 – 49, 2011.

NG, C.; DEMAERE, M. Z.; WILLIAMS, T. J.; LAURO, F. M.; RAFTERY, M.; GIBSON, J. A. E.; ANDREWS-PFANNKUCH, C.; LEWIS, M.; HOFFMAN, J. M.; THOMAS, T.; CAVICCHIOLI, R. Metaproteogenomic analysis of a dominant green sulfur bacterium from Lago Ace, Antarctica. **The ISME Journal**, v. 4, p. 1002 – 1019, 2010.

OH, S.; CARO-QUINTERO, A.; TSEMENTZI, D.; DELEON-RODRIGUEZ, N.; LUO, C.; PORETSKY, R.; KONSTANTINIDIS, K. T. Metagenomic Insights into the Evolution, Function, and Complexity of the Planktonic Microbial Community of Lake Lanier, a Temperate Freshwater Ecosystem. **Applied and Environmental Microbiology**, v. 77, n. 17, p. 6000 – 6011, 2011.

PÉREZ, M. T.; HÖRTNAGL, P.; SOMMARUGA, R. Contrasting ability to take up leucine and thymidine among freshwater bacterial groups: implications for bacterial production measurements. **Environmental Microbiology**, 2010 v. 12, n. 1, p. 74 – 82, 2010.

RAM, A. S. P.; BOUCHER, D.; SIME-NGANDO, T.; DEBROAS, D.; ROMAGOUX, J. C. Phage Bacteriolysis, Protistan Bacterivory Potential, and Bacterial Production in a Freshwater Reservoir: Coupling with Temperature. **Microbial Ecology**, v. 50, p. 64 – 72, 2005.

RAO, S. P. S.; ALONSO, S.; RAND, L.; DICK, T.; PETHE, K. The protonmotive force is required for maintaining ATP homeostasis and viability of hypoxic, nonreplicating *Mycobacterium tuberculosis*. **Proceedings of the National Academy of Sciences**, v. 105, n. 33, p. 11945 – 11950, 2008.

REN, L.; HE, D.; ZENG, J.; WU, Q. L. Bacterioplankton communities turn unstable and become small under increased temperature and nutrient-enriched conditions. **FEMS Microbiology Ecology**, v. 84, n. 3, p. 614 – 624, 2013.

RØNN, R.; GRUNERT, J.; EKELOUND, F. Protozoan response to addition of the bacteria *Mycobacterium chlorophenolicum* and *Pseudomonas chlororaphis* to soil microcosms. **Biology and Fertility of Soils**, v. 33, n. 2, p. 126 – 131, 2001.

RUAN, Q.; DUTTA, D.; SCHWALBACH, M. S.; STEELE, J. A.; FUHRMAN, J. A.; SUN, F. Local similarity analysis reveals unique associations among marine bacterioplankton species and environmental factors. **Bioinformatics**, v. 15, n. 22, p. 2532 – 2538, 2006.

RUSCH, D. B.; HALPERN, A. L.; SUTTON, G.; HEIDELBERG, K. B.; WILLIAMSON, S.; YOOSEPH, S.; WU, D.; EISEN, J. A.; HOFFMAN, J. M.; REMINGTON, K.; BEESON, K.; TRAN, B.; SMITH, H.; BADEN-TILLSON, H.; STEWART, C.; THORPE, J.; FREEMAN, J.; ANDREWS-PFANNKUCH, C.; VENTER, J. E.; LI, K.; KRAVITZ, S.; HEIDELBERG, J. F.; UTTERBACK, T.; ROGERS, Y. H.; FALCÓN, L. I.; SOUZA, V.; BONILLA-ROSSO, G.; EGUIARTE, L. E.; KARL, D. M.; SATHYENDRANATH, S.; PLATT, T.; BIRMINGHAM, E.; GALLARDO, V.; TAMAYO-CASTILLO, G.; FERRARI, M. R.; STRAUSBERG, R. L.; NEALSON, K.; FRIEDMAN, R.; FRAZIER, M.; VENTER, J. C. The Sorcerer II Global Ocean Sampling expedition: northwest Atlantic through eastern tropical Pacific. **PLoS Biology**, v. 5, n. 3, p. 398 – 431, 2007.

SALCHER, M. M.; PERNTHALER, J.; ZEDER, M.; PSENNER, R.; POSCH, T. Spatio-temporal niche separation of planktonic Betaproteobacteria in an oligo-mesotrophic lake. **Environmental Microbiology**, v. 10, n. 8, p. 2074 – 2086, 2008.

SHANNON, P.; MARKIEL, A.; OZIER, O.; BALIGA, N. S.; WANG, J. T.; RAMAGE, D.; AMIN, N.; SCHWIKOWSKI, B.; IDEKER, T. Cytoscape: A Software Environment for Integrated Models of Biomolecular Interaction Networks. **Genome Research**, v. 13, p. 2498 – 2504, 2003.

SU, G.; KUCHINSKY, A.; MORRIS, J. H.; STATES, D. J.; MENG, F. GLay: community structure analysis of biological networks. **Bioinformatics**, v. 26, n. 24, p. 3135 – 3137, 2010.

TYSON, G.; CHAPMAN, J.; HUGENHOLTZ, P.; ALLEN, E.; RAM, R.; RICHARDSON, P.; SOLOVYEV, V.; RUBIN, E.; ROKHSAR, D.; BANFIELD, J.

Community structure and metabolism through reconstruction of microbial genomes from the environment. **Nature**, v. 428, p. 37 – 43, 2004.

VANDAL, O. H.; ROBERTS, J. A.; ODAIRA, T.; SCHNAPPINGER, D.; NATHAN, C. F.; EHRT, S. Acid-Susceptible Mutants of *Mycobacterium tuberculosis* Share Hypersusceptibility to Cell Wall and Oxidative Stress and to the Host Environment. **Journal of Bacteriology**, v. 191, n. 2, p. 625 – 631, 2009.

VENTER, J. C.; REMINGTON, K.; HEIDELBERG, J. F.; HALPERN, A. L.; RUSCH, D.; EISEN, J. A.; WU, D.; PAULSEN, I.; NELSON, K. E.; NELSON, W.; FOUTS, D. E.; LEVY, S.; KNAP, A. H.; LOMAS, M. W.; NEALSON, K.; WHITE, O.; PETERSON, J.; HOFFMAN, J.; PARSONS, R.; BADEN-TILLSON, H.; PFANNKUCH, C.; ROGERS, Y.; SMITH, H. O. Environmental genome shotgun sequencing of the Sargasso Sea. **Science**, v. 304, p. 66 – 74, 2004.

VILA-COSTA, M.; PINHASSI, J.; ALONSO, C.; PERNTHALER, J.; SIMÓ, R. An annual cycle of dimethylsulfoniopropionate-sulfur and leucine assimilating bacterioplankton in the coastal NW Mediterranean. **Environmental Microbiology**, v. 9, p. 1 – 13, 2007.

WILKES, G.; EDGE, T.; GANNON, V.; JOKINEN, C.; LYAUTEY, E.; MEDEIROS, D.; NEUMANN, N.; RUECKER, N.; TOPP, E.; LAPEN, D. R. Seasonal relationships among indicator bacteria, pathogenic bacteria, *Cryptosporidium* oocysts, *Giardia* cysts, and hydrological indices for surface waters within an agricultural landscape. **Water Research**, v. 43, p. 2209 – 2223, 2009.

XIA, L. C.; STEELE, J. A.; CRAM, J. A.; CARDON, Z. G.; SIMMONS, S. L.; VALLINO, J. J.; FUHRMAN, J. A.; SUN, F. Extended local similarity analysis (eLSA) of microbial community and other time series data with replicates. **BMC Systems Biology**, v. 5, p. 1 – 12, 2011.

YE, Y.; DOAK, T. G. A Parsimony Approach to Biological Pathway Reconstruction/Inference for Genomes and Metagenomes. **PLoS Computational Biology**, v. 5, n. 9, p. 1 – 8, 2009.

ZEDER, M.; PETER, S.; SHABAROVA, T.; PERNTHALER, J. A small population of planktonic Flavobacteria with disproportionately high growth during the spring phytoplankton bloom in a prealpine lake. **Environmental Microbiology**, v. 11, p. 2676 – 2686, 2009.

ZWART, G.; CRUMP, B. C.; AGTERVELD, M. P. K. V.; HAGEN, F.; HAN, S. K. Typical freshwater bacteria: an analysis of available 16S rRNA gene sequences from plankton of lakes and rivers. **Aquatic Microbiology and Ecology**, v. 28, p. 141 – 155, 2002.

APÊNDICE 1

Tabela de Abundância Relativa de Módulos no Lago São Francisco do Parque Zoológico

ID	NOME	OCORRÊNCIAS POR MILHÃO											
		Out12	Nov12	Dez12	Jan13	Fev13	Mar13	Abr13	Mai13	Jun13	Jul13	Ago13	Set13
M00001	Glycolysis (Embden-Meyerhof pathway), glucose => pyruvate	146,29	174,65	131,34	148,16	127,50	134,18	133,47	131,26	159,44	132,41	125,73	110,49
M00002	Glycolysis, core module involving three-carbon compounds	190,15	216,97	168,26	179,11	158,44	161,79	167,17	156,09	186,30	170,21	161,68	132,75
M00003	Gluconeogenesis, oxaloacetate => fructose-6P	167,10	191,19	152,06	158,80	144,11	138,70	136,22	136,27	171,65	150,85	140,01	119,45
M00004	Pentose phosphate pathway (Pentose phosphate cycle)	99,71	102,17	76,23	98,38	86,40	95,87	85,29	96,05	113,28	90,40	80,72	78,47
M00005	PRPP biosynthesis, ribose 5P => PRPP	194,55	221,61	190,22	178,23	171,11	171,92	172,43	171,47	189,11	188,11	187,70	147,70
M00006	Pentose phosphate pathway, oxidative phase, glucose 6P => ribulose 5P	75,59	73,33	53,88	72,26	67,14	76,00	63,59	74,99	96,81	69,23	60,04	62,78
M00007	Pentose phosphate pathway, non-oxidative phase, fructose 6P => ribose 5P	146,44	168,37	130,27	154,06	121,15	129,79	129,46	133,58	136,49	130,23	123,13	104,60
M00008	Entner-Doudoroff pathway, glucose-6P => glyceraldehyde-3P + pyruvate	47,43	48,12	35,05	36,68	37,76	37,97	35,40	50,55	50,64	37,86	34,77	24,72
M00009	Citrate cycle (TCA cycle, Krebs cycle)	175,06	188,67	161,87	174,35	156,91	161,19	158,85	145,13	186,36	162,72	161,40	136,94
M00010	Citrate cycle, first carbon oxidation	162,90	177,88	158,93	178,22	155,68	174,36	163,51	145,57	183,39	154,93	159,69	142,52
M00011	Citrate cycle, second carbon oxidation	183,27	195,80	163,68	172,11	157,65	154,20	156,19	144,86	188,19	167,78	162,44	133,80
M00012	Glyoxylate cycle	139,13	146,19	121,34	123,77	118,27	113,84	116,78	106,97	134,73	131,02	116,24	94,70

M00015	Proline biosynthesis, glutamate => proline	142,09	163,71	110,90	123,75	111,48	106,58	111,22	113,90	145,64	122,64	106,94	89,80
M00016	Lysine biosynthesis, aspartate => lysine	136,96	157,51	122,68	131,07	116,66	116,20	114,30	121,40	148,10	124,43	116,83	99,70
M00018	Threonine biosynthesis, aspartate => homoserine => threonine	124,42	145,80	114,27	121,96	101,58	117,79	114,13	107,82	134,88	108,96	102,03	94,37
M00020	Serine biosynthesis, glyceralate-3P => serine	154,42	189,40	132,95	156,68	133,95	147,76	132,93	138,08	164,43	139,66	132,61	113,00
M00021	Cysteine biosynthesis, serine => cysteine	116,59	129,71	94,38	90,06	92,25	84,28	91,68	90,83	109,51	104,86	90,94	71,82
M00022	Shikimate pathway, phosphoenolpyruvate + erythrose-4P => chorismate	145,70	167,02	117,72	125,37	117,86	110,24	111,92	116,63	139,52	128,48	116,08	95,04
M00023	Tryptophan biosynthesis, chorismate => tryptophan	152,09	171,97	126,89	127,65	124,68	116,20	122,61	115,89	149,39	136,19	118,11	97,26
M00024	Phenylalanine biosynthesis, chorismate => phenylalanine	82,92	91,92	63,96	58,89	57,22	48,63	47,89	52,49	66,92	65,11	51,66	37,54
M00025	Tyrosine biosynthesis, chorismate => tyrosine	90,71	98,48	67,95	65,48	64,98	52,25	51,73	58,89	73,78	72,88	58,04	44,91
M00026	Histidine biosynthesis, PRPP => histidine	91,66	100,51	77,32	87,67	68,02	69,55	68,60	66,87	99,69	75,31	56,18	57,96
M00027	GABA (gamma-Aminobutyrate) shunt	26,89	17,59	11,09	9,53	18,50	19,96	24,31	27,19	26,69	26,43	21,90	21,19
M00028	Ornithine biosynthesis, glutamate => ornithine	72,19	90,80	62,76	82,52	65,61	73,17	56,84	67,33	82,29	62,35	64,74	59,87
M00034	Methionine salvage pathway	29,61	28,80	17,05	30,79	24,35	25,92	20,41	29,04	30,51	29,33	25,71	27,08
M00035	Methionine degradation	101,51	117,77	107,05	121,85	90,43	110,82	98,33	114,08	110,17	95,76	95,17	100,89
M00036	Leucine degradation, leucine => acetoacetate + acetyl-CoA	50,59	66,69	55,82	66,53	45,08	52,29	47,27	52,51	57,70	44,01	48,12	42,56

M00045	Histidine degradation, histidine => N-formiminoglutamate => glutamate	28,04	41,74	28,30	40,84	25,72	34,73	35,38	39,30	37,82	24,84	31,13	22,12
M00048	Inosine monophosphate biosynthesis, PRPP + glutamine => IMP	184,02	202,22	152,47	171,87	158,18	157,51	148,60	159,72	190,09	169,88	154,18	136,16
M00049	Adenine nucleotide biosynthesis, IMP => ADP/dADP,ATP/dATP	77,41	84,99	57,19	37,12	41,80	33,12	55,45	73,31	33,09	58,81	56,79	39,24
M00050	Guanine nucleotide biosynthesis, IMP => GDP/dGDP,GTP/dGTP	76,24	84,64	56,92	37,05	41,39	32,87	55,00	72,48	33,02	58,16	56,03	39,24
M00052	Pyrimidine ribonucleotide biosynthesis, UMP => UDP/UTP,CDP/CTP	184,75	215,87	162,67	174,81	149,75	160,82	140,39	146,78	159,65	168,09	144,61	124,81
M00053	Pyrimidine deoxyribonucleotide biosynthesis, CDP/CTP => dCDP/dCTP,dTDP/dTTP	92,32	102,57	68,81	46,98	51,82	41,99	66,83	87,03	41,45	71,55	68,88	48,98
M00063	CMP-KDO biosynthesis	94,78	126,12	85,88	108,84	79,38	76,39	76,24	85,43	88,87	88,03	78,11	63,78
M00064	ADP-L-glycero-D-manno-heptose biosynthesis	105,26	129,27	88,50	88,44	80,37	71,75	67,08	70,72	83,63	81,52	76,28	60,33
M00083	Fatty acid biosynthesis, elongation	156,52	194,43	149,78	154,88	133,57	139,56	118,30	133,98	148,39	136,61	126,15	104,70
M00086	beta-Oxidation, acyl-CoA synthesis	134,91	170,48	137,48	147,23	111,73	126,32	127,82	131,28	149,42	118,16	108,80	89,70
M00087	beta-Oxidation	77,30	85,00	72,50	79,58	64,26	74,71	75,83	79,09	90,84	74,95	68,83	63,72
M00093	Phosphatidylethanolamine (PE) biosynthesis, PA => PS => PE	110,17	132,71	93,78	101,53	86,27	83,10	74,71	83,59	98,39	105,06	85,21	61,18
M00096	C5 isoprenoid biosynthesis, non-mevalonate pathway	78,91	101,84	67,85	103,03	63,98	90,06	69,29	75,70	68,48	64,08	72,44	65,48

M00115	NAD biosynthesis, aspartate => NAD	141,19	169,41	116,82	143,22	118,81	124,58	112,78	112,87	135,86	119,83	120,55	102,35
M00118	Glutathione biosynthesis, glutamate => glutathione	72,96	91,82	68,16	69,68	57,83	51,94	59,89	58,63	71,97	72,20	53,94	38,26
M00119	Pantothenate biosynthesis, valine/L-aspartate => pantothenate	91,98	111,27	84,33	94,36	72,83	77,51	74,98	84,09	86,20	73,80	60,73	54,57
M00120	Coenzyme A biosynthesis, pantothenate => CoA	145,72	170,30	114,32	137,37	116,94	121,46	116,61	122,68	146,26	121,55	115,04	99,92
M00121	Heme biosynthesis, glutamate => protoheme/siroheme	53,40	60,88	32,23	41,66	40,81	42,23	37,98	47,36	52,37	45,79	37,00	38,62
M00123	Biotin biosynthesis, pimeloyl-CoA => biotin	82,92	95,43	61,73	59,45	60,83	56,58	48,36	62,22	69,50	64,06	53,66	47,27
M00125	Riboflavin biosynthesis, GTP => riboflavin/FMN/FAD	145,02	181,89	129,49	143,60	120,94	123,39	119,76	123,31	155,74	129,92	122,26	106,47
M00126	Tetrahydrofolate biosynthesis, GTP => THF	106,64	135,97	102,65	106,67	82,61	86,11	87,07	96,27	99,88	85,40	81,69	73,59
M00133	Polyamine biosynthesis, arginine => agmatine => putrescine => spermidine	31,53	44,68	27,92	36,28	29,27	27,84	22,38	31,80	29,70	31,49	29,48	29,30
M00134	Polyamine biosynthesis, arginine => ornithine => putrescine	24,24	35,29	26,99	36,14	22,21	30,28	23,54	22,61	23,62	20,22	23,83	22,35
M00135	GABA biosynthesis, eukaryotes, putrescine => GABA	29,99	27,35	20,29	23,00	22,99	30,10	23,65	28,07	38,31	29,84	23,28	23,26
M00140	C1-unit interconversion, prokaryotes	104,09	128,63	98,03	133,56	108,69	113,30	118,35	104,13	132,78	102,52	109,80	93,94
M00144	Complex I (NADH dehydrogenase), NADH dehydrogenase I	176,92	206,16	159,51	171,58	153,47	155,28	164,24	149,96	181,87	168,80	154,00	127,90

M00149	Complex II (succinate dehydrogenase / fumarate reductase), succinate dehydrogenase	167,57	168,51	127,00	139,82	130,91	120,38	122,44	131,12	162,67	149,83	133,12	120,82
M00151	Complex III (Cytochrome bc1 complex)	101,86	122,77	89,22	80,66	87,69	59,35	72,00	73,01	85,93	97,40	70,30	52,14
M00153	Cytochrome bd complex	36,30	41,74	29,11	34,87	29,47	31,28	32,54	32,53	36,31	39,89	36,79	28,87
M00156	Complex IV (Cytochrome c oxidase), cytochrome c oxidase, cbb3-type	47,26	74,71	49,32	45,36	47,55	38,81	39,95	53,53	48,58	73,21	29,46	32,77
M00157	F-type ATPase, bacteria	183,62	204,60	148,64	160,24	143,20	132,69	133,21	139,79	168,42	160,29	134,18	106,15
M00164	ATP synthase	183,62	204,60	148,64	160,24	143,20	132,69	133,21	139,79	168,42	160,29	134,18	106,15
M00168	CAM (Crassulacean acid metabolism), dark	167,21	180,75	143,73	154,12	135,98	131,66	137,47	134,13	156,84	141,93	137,17	106,77
M00169	CAM (Crassulacean acid metabolism), light	79,21	107,63	89,79	77,20	78,82	75,39	68,47	67,34	79,77	71,35	70,82	64,74
M00170	C4-dicarboxylic acid cycle, phosphoenolpyruvate carboxykinase type	57,86	80,46	70,83	94,47	58,66	74,45	61,72	66,92	54,25	52,35	70,97	60,11
M00178	Ribosome, bacteria	229,67	267,86	206,44	223,80	191,82	196,85	203,85	182,67	227,52	214,26	194,13	156,98
M00183	RNA polymerase, bacteria	234,33	255,02	202,08	204,99	201,56	180,85	193,15	185,95	241,52	221,64	200,58	171,26
M00185	Sulfate transport system	93,88	89,36	70,74	44,37	68,28	52,54	52,08	66,45	73,59	76,36	57,90	49,10
M00188	Sulfonate/nitrate/taurine transport system	41,28	54,72	46,94	54,07	38,45	43,89	42,85	41,99	50,25	36,93	37,47	34,34
M00189	Molybdate transport system	52,67	74,43	55,14	57,21	46,89	48,12	53,44	60,99	72,41	54,07	44,94	39,30
M00190	Iron(III) transport system	77,16	94,26	71,73	81,90	67,31	66,44	61,98	64,66	80,97	67,51	61,64	54,23
M00193	Putative spermidine/putrescine transport system	33,63	37,48	30,19	39,52	30,97	37,78	37,50	31,92	44,44	32,24	34,31	30,00
M00207	Multiple sugar transport system	26,49	29,99	22,31	21,61	20,41	22,34	23,25	25,11	24,53	22,90	18,26	15,38
M00210	Putative ABC transport	117,58	117,74	85,23	79,08	80,99	81,33	78,99	90,00	96,44	108,96	75,62	68,17

	system												
M00221	Simple sugar transport system	29,17	37,34	29,92	28,32	27,39	28,33	31,60	29,78	35,29	27,70	25,47	23,81
M00222	Phosphate transport system	164,04	189,78	147,35	144,45	143,57	137,40	138,75	137,33	177,07	156,05	141,16	130,67
M00223	Phosphonate transport system	30,77	43,26	42,48	39,89	23,40	31,90	26,60	28,42	33,57	19,60	21,50	21,40
M00230	Glutamate/aspartate transport system	32,01	33,82	28,05	28,66	22,58	19,86	26,68	26,21	24,85	27,55	21,55	13,11
M00236	Polar amino acid transport system	45,71	51,67	34,60	37,40	39,05	36,93	40,19	40,68	53,32	45,84	40,14	35,68
M00237	Branched-chain amino acid transport system	96,44	117,12	87,29	83,95	77,78	76,46	86,79	93,69	102,76	87,53	71,70	57,82
M00239	Peptides/nickel transport system	68,06	88,08	72,81	72,56	57,82	67,80	60,03	60,20	74,16	57,03	57,20	53,40
M00240	Iron complex transport system	29,51	41,43	29,45	22,81	20,92	19,30	19,68	24,64	27,26	22,97	21,19	17,84
M00242	Zinc transport system	20,97	35,92	19,62	25,40	21,48	20,08	21,07	21,78	34,36	25,13	22,69	25,24
M00244	Zinc/manganese transport system	32,37	35,63	20,33	23,74	24,42	22,11	20,93	23,05	34,74	22,56	22,89	20,25
M00248	Antibiotic transport system	79,94	96,83	67,75	70,93	62,12	66,36	60,79	60,93	73,34	68,60	60,80	50,77
M00250	Lipopolysaccharide transport system	88,50	101,08	77,72	80,04	73,23	77,97	78,51	68,85	83,93	74,92	78,69	63,20
M00252	Lipooligosaccharide transport system	45,75	55,50	55,52	41,87	41,77	45,03	59,61	46,60	48,25	40,00	45,59	24,85
M00254	ABC-2 type transport system	55,25	79,35	60,93	69,54	53,13	59,86	50,40	54,61	56,17	49,09	49,94	43,26
M00255	Lipoprotein-releasing system	89,65	116,30	91,83	101,20	81,02	78,61	68,26	80,38	85,90	83,61	76,47	59,74
M00256	Cell division transport system	77,43	95,19	72,87	86,80	74,14	81,16	89,80	78,80	96,15	84,45	86,11	67,14
M00258	Uncharacterized ABC transport system	65,35	105,92	80,22	88,30	60,96	73,24	59,10	64,60	71,54	55,65	53,02	51,08
M00259	Heme transport system	73,61	94,90	63,58	62,07	54,97	39,91	42,07	48,93	63,30	66,58	51,32	40,24
M00299	Spermidine/putrescine	92,56	85,49	62,39	93,10	83,72	92,44	91,29	85,13	118,96	89,83	92,22	85,45

	transport system												
M00300	Putrescine transport system	21,01	28,17	24,42	28,98	19,18	24,29	20,24	19,32	28,66	18,93	20,86	19,27
M00307	Pyruvate oxidation, pyruvate => acetyl-CoA	134,60	159,58	129,97	128,69	119,34	112,21	107,73	111,72	141,68	127,63	114,30	98,94
M00313	indolepyruvate:ferredoxin oxidoreductase	31,99	37,35	31,75	29,50	24,47	25,02	26,38	30,78	31,21	27,04	20,04	16,12
M00320	Lipopolysaccharide export system	82,90	101,59	70,86	66,94	64,88	43,72	45,07	56,78	67,78	75,43	59,58	42,02
M00338	Cysteine biosynthesis, homocysteine + serine => cysteine	25,53	32,09	27,68	34,46	25,89	32,11	28,42	29,46	32,22	25,28	28,83	27,18
M00342	Bacterial proteasome	30,58	29,32	27,19	30,96	29,53	29,72	33,21	30,22	47,85	36,87	37,26	37,97
M00349	Microcin C transport system	29,45	35,69	35,51	40,48	25,12	27,80	29,03	26,63	32,61	28,67	23,22	20,16
M00359	Aminoacyl-tRNA biosynthesis, eukaryotes	152,72	198,48	158,03	165,80	146,13	151,32	143,27	141,25	151,87	149,82	146,09	120,10
M00360	Aminoacyl-tRNA biosynthesis, prokaryotes	173,55	213,61	169,81	178,32	160,13	164,31	158,81	153,16	170,82	167,16	160,05	131,76
M00364	C10-C20 isoprenoid biosynthesis, bacteria	33,85	46,71	31,32	49,65	27,37	43,99	29,98	34,97	29,53	27,06	31,67	26,31
M00365	C10-C20 isoprenoid biosynthesis, archaea	20,70	28,37	19,00	29,73	16,86	27,10	20,89	22,16	18,93	18,21	22,23	19,18

CAPÍTULO 3:

ANÁLISE COMPUTACIONAL DO PERFIL METABÓLICO DO SISTEMA DE COMPOSTAGEM DA FUNDAÇÃO PARQUE ZOOLOGICO DE SÃO PAULO

INTRODUÇÃO

Compostagem é um processo de tratamento de resíduos orgânicos de origens diversas, onde a matéria orgânica é convertida em um composto rico em ácidos húmicos e fúlvicos, amplamente utilizado com fertilizante. Assim, o processo de compostagem constitui uma alternativa ecológica para a mitigação de impactos relacionados à destinação de lixo.

Os sistemas empregados na degradação da biomassa e obtenção do composto também variam grandemente em relação à escala, duração do processamento e modo de processamento, variando de sistemas abertos de empilhamento dos resíduos, semiabertos, onde a disposição dos resíduos é feita em câmaras ou fechados com a utilização de reatores (RYNK *et al.*, 1992; FITZPATRICK *et al.*, 2005; GAJALAKSHMI & ABBASI, 2008).

É comum o processo de compostagem apresentar três fases distintas: fase bio-oxidativa, fase de resfriamento e fase de maturação (PARTANEN *et al.*, 2010; KUMAR, 2011). Geralmente a fase bio-oxidativa é caracterizada por um aumento rápido de temperatura e degradação intensa de composto simples como mono-, di- e oligossacarídeos, bem como carboidratos de reserva. Nesta fase, essencialmente termofílica, a temperatura pode atingir temperaturas entre 60 – 80°C devido ao intenso metabolismo microbiano (KUMAR, 2011). Nas fases de resfriamento e de maturação, a temperatura decresce gradualmente, refletindo o esgotamento dos compostos simples e o início da degradação de compostos mais complexos como hemicelulose, lignina e proteínas (LÓPEZ-GONZÁLEZ *et al.*, 2013). Dessa forma se estabelece uma influência dos microrganismos no meio e as variações em sua composição e/ou natureza promovem a sucessão dos grupos microbianos comumente associados à compostagem.

Técnicas moleculares são comumente empregadas para a identificação dos grupos responsáveis pelo processo. Entre tais técnicas podemos citar RFLP (do inglês, *Restriction Fragment Length Polymorphism* – ISHII *et al.*, 2000; ALFREIDER *et al.*, 2002), microarranjos de DNA (FRANKE-WHITTLE *et al.*, 2009) e perfil de ácidos graxos (STEGER *et al.*, 2007). Estes estudos apontaram grupos de microrganismos predominantemente associados a diferentes fases do processo de compostagem (ALFREIDER *et al.*, 2002). A logística envolvida na gerência de um parque zoológico é uma área capital dada sua grande complexidade. Além das acomodações, da alimentação e dos procedimentos ligados à saúde e bem-estar dos animais, existe o desafio de lidar com os resíduos orgânicos gerados no cotidiano. Diante disso, foi inaugurado em 2003 no Parque Zoológico do Estado de São Paulo um sistema de compostagem contando com uma área total de 1500 m², 44 células em alvenaria medindo 8 m³ e capacidade de produzir 600 toneladas de composto/ano (MARTINS *et al.*, 2013). Um estudo metagenômico recente de duas amostras independentes do sistema de compostagem do Parque Zoológico foi realizado por Martins e colaboradores (MARTINS *et al.*, 2013). Esses resultados ampliaram o conhecimento acerca do processo de degradação de biomassa nesse sistema complexo, incluindo o papel preponderante de bactérias, como por exemplo as do grupo Clostridiales e Actinomycetales, na degradação de compostos recalcitrantes. Os autores concluíram que o papel de fungos nos processos de degradação foi menor que o esperado possivelmente devido às condições anaeróbicas e termofílicas do sistema. Além disso, os autores constataram que as categorias funcionais dos genes identificados coincidem com as categorias encontradas em outros metagenomas de ambientes de degradação, à despeito das diferenças taxonômicas entre eles.

Dada sua singularidade, foi postulado que o sistema de compostagem do Parque Zoológico abrigaria uma grande diversidade microbiana com enorme potencial biotecnológico, o que foi corroborado pela análise metagenômica inicial. Um importante passo para o aprofundamento do conhecimento do processo de compostagem é conhecer os aspectos metabólicos potencialmente relevantes na degradação da biomassa e obtenção do composto maduro.

METODOLOGIA

Descrição dos Dois Conjuntos de Dados Metagenômicos

Os conjuntos de dados metagenômicos (TABELA 1) analisados neste trabalho correspondem a sequências (*reads*) geradas pelo sequenciamento *shotgun* do DNA total extraído de amostras coletadas ao longo do processo de compostagem. O primeiro conjunto de dados são oriundos de amostras coletadas da composteira denominada ZC3 após 1, 30, 64, 78 e 99 dias de sua montagem. O segundo conjunto de dados foi obtido da composteira denominada ZC4 a partir de amostras coletadas após 1, 3, 7, 15, 30, 64, 67, 78 e 99 dias de sua montagem. Em ambos os casos a coleta e o processamento das amostras foram realizados como previamente descrito (MARTINS *et al.*, 2013). O sequenciamento *shotgun* foi realizado com as plataformas 454-Roche e MiSeq-Illumina para as amostras de ZC3, e apenas com a plataforma MiSeq-Illumina no caso de ZC4. Ambos conjuntos de dados foram obtidos como parte do projeto de Doutorado de Luciana Principal Antunes realizado no Instituto de Química da USP, sob a orientação de Aline Maria da Silva e co-orientação de João Carlos Setúbal (Tese de doutorado em fase de redação).

TABELA 1: Número total de sequências (*reads*). A: 454-Roche; B: MiSeq-Illumina (*paired-end reads*/500 ciclos). Dados fornecidos por Luciana Principal Antunes.

Amostra	ZC3A	ZC3B	ZC4B
D1	520.074	2.270.265	4.106.932
D3	-	-	2.481.164
D7	-	-	4.556.562
D15	-	-	3.574.653
D30	737.772	2.145.651	3.164.693
D64	771.427	1.265.241	5.055.450
D67	-	-	2.825.621
D78	1.063.197	2.169.689	11.245.102
D99	711.081	1.509.658	8.323.524

Os dados obtidos no sequenciamento foram analisados processados através do programa HUMAnN de acordo com a metodologia descrita no capítulo 1. A amostras que possuem *reads* 454 e Illumina tiveram os conjuntos unidos depois do passo de recuperação da plataforma MG-RAST, antes da aplicação do programa.

RESULTADOS E DISCUSSÃO

Variação do Potencial Metabólico no Processo de Compostagem

Através da aplicação do programa HUMAnN foi possível estabelecer um panorama geral do potencial metabólico e identificar diferenças importantes nas amostras coletadas em dias diferentes do processo de compostagem tanto para composteira ZC3 como para ZC4.

O agrupamento hierárquico dos resultados obtidos (abundância relativa de rotas metabólicas) para cada amostra de ZC3 relaciona as amostras dos dias 64 e 78 seguidas pelas amostras dos dias 30, 99 e 1 respectivamente (FIGURA 1). O mesmo padrão se repete quando se consideram os módulos (FIGURA 2). Visando identificar a dinâmica do processo de compostagem e quão relevante é a variação ao longo do processo, valores discrepantes da abundância relativa (*outliers*) foram identificados como quaisquer valores abaixo ou acima de 1,5 x distância interquartil. A maioria dos valores *outliers* foram referentes ao dia 1 e/ou ao dia 99 (Apêndices 2 e 3), o que corrobora o padrão dos gráficos *heatmaps*.

No dia 1 do processo de compostagem existe uma super-representação de módulos relacionados ao metabolismo de carboidratos, possivelmente refletindo a composição do substrato rico em matéria vegetal e fezes. A diversidade de tais carboidratos presentes nos primeiros estágios do processo reflete-se nos módulos listados na tabela 1, dentre os quais se pode destacar os módulos relacionados à degradação de pectina, a sistemas transportadores de maltose/maltodextrina, L-Arabinose, metil-galactosídeo e sistemas de fosfotransferase (PTS) específicos para maltose e glicose, N-acetilglicosamina, *arbutin-like* (arbutina e salicina), sacarose, trealose, beta-glucosídeos, N-acetilgalactosamina. O módulo referente à gliconeogênese, super-representado na comunidade, pode garantir a algumas espécies um suprimento adequado de glicose em micro nichos anóxicos.

Ainda no 1º dia, a super-representação do módulo referente à biossíntese de coenzima M indica a abundância de metanogênicos, grupo frequentemente associado a microbiota de alguns animais. Os módulos referentes a 2-oxoglutarato:ferredoxina oxidorreductase e a indolpiruvato:ferredoxina oxidorreductase

muitas vezes associadas ao metabolismo anabólico na ausência de oxigênio, também se mostraram super-representados.

FIGURA 1: *Heatmap* representando a abundância das rotas metabólicas nas amostras da composteira ZC3. Cada uma das 174 rotas presentes no banco KEGG está representada horizontalmente enquanto as amostras dos dias 1 (ZC3D1), 30 (ZC3D30), 64 (ZC3D64), 78 (ZC3D78) e 99 (ZC3D99) de processo estão representadas verticalmente. O gráfico foi gerado através da plataforma R, módulo *heatmap*, cada rota foi normalizada de forma que a média dos valores correspondesse a zero e o desvio padrão a 1.

FIGURA 2: *Heatmap* representando a abundância dos módulos KEGG nas amostras da composteira ZC3. Cada um dos 259 módulos presentes no banco KEGG está representado horizontalmente enquanto as amostras dos dias 1 (ZC3D1), 30 (ZC3D30), 64 (ZC3D64), 78 (ZC3D78) e 99 (ZC3D99) de processo estão representadas verticalmente. O gráfico foi gerado através da plataforma R, módulo *heatmap*, cada módulo foi normalizado de forma que a média dos valores correspondesse a zero e o desvio padrão a 1.

Com base na abundância relativa dos módulos, é possível inferir que a síntese dos aminoácidos prolina, lisina, metionina, treonina, fenilalanina, tirosina, histidina (M00015, M00016, M00017, M00018, M00024, M00025, M00026 respectivamente) manteve-se estável durante todo o processo. Os módulos relativos

à síntese do aminoácido triptofano (M00023) e ao transporte de glutamina (M00219) foram os únicos que apresentaram abundância relativamente baixa no primeiro dia de processo comparado aos demais. A diminuição da abundância do módulo referente à síntese de leucina coincide com o aumento dos módulos referentes ao transporte de aminoácidos de cadeia ramificada e à degradação de leucina, respectivamente (FIGURA 3).

FIGURA 3: Abundância relativa de módulos KEGG relacionados a metabolismo de aminoácidos.

De forma geral é possível afirmar que, a despeito do esgotamento de compostos simples, da diminuição do substrato e do aumento das condições gerais de estresse, a comunidade mantém-se ativa e estável com o deslocamento do potencial metabólico. Esse deslocamento pode ser ilustrado pelo aumento ou a manutenção de módulos relacionados à rota das pentoses-fosfato (M00004 e M00006), ao ciclo do citrato (M00009, M00010 e M00011), ao ciclo do glioxilato (M00012), à síntese de ácidos graxos (M00060) e a proteassomos bacterianos (M00331). Embora haja a liberação de dióxido de carbono, é possível que parte

deste gás seja fixado através do metabolismo CAM e C4 (M00168, M00169 e M00170).

Para comparar o potencial metabólico das amostras do sistema de compostagem, quatro outros metagenomas disponíveis na plataforma *on-line* MG-RAST foram submetidos ao mesmo processo de análise: microbiota intestinal de cupim, microbiota de rúmen bovino, microbioma do solo da floresta de Luquillo em Porto Rico e biofilme em terreno de drenagem de mineração (MG-RAST Ids 4442701.3, 4441682.3, 4446153.3, 4441138.3 respectivamente). Os três primeiros metagenomas foram escolhidos por estarem relacionados ao processo de degradação de biomassa, enquanto o quarto foi escolhido por se tratar de um ambiente extremo.

O agrupamento hierárquico baseado na abundância de rotas metabólicas nesse novo conjunto de metagenomas gerou dois grupos: o primeiro formado pelas amostras dos dias 64 e 78 seguidas pelas amostras dos dias 30, 99 e os metagenomas do solo de Luquillo e do biofilme respectivamente. Já o segundo grupo foi formado pelo metagenoma de rúmen juntamente com a amostra do primeiro dia de compostagem (FIGURA 4). Contudo, quando se considera a abundância de módulos, surge um padrão tripartido: o primeiro grupo formado pelos metagenomas relacionados ao rúmen e ao intestino de cupim, o segundo formado pelas amostras da compostagem e pelo metagenoma de Luquillo e finalmente, distante dos demais, o metagenoma do biofilme (FIGURA 5).

FIGURA 4: *Heatmap* representando a abundância das rotas metabólicas nas amostras da composteira ZC3. Cada uma das 174 rotas presentes no banco KEGG está representada horizontalmente enquanto as amostras dos dias 1 (ZC3D1), 30 (ZC3D30), 64 (ZC3D64), 78 (ZC3D78) e 99 (ZC3D99) de processo estão representadas verticalmente. O gráfico foi gerado através da plataforma R, módulo *heatmap*, cada rota foi normalizada de forma que a média dos valores correspondesse a zero e o desvio padrão a 1.

FIGURA 5: Heatmap representando a abundância dos módulos KEGG nas amostras da composteira ZC3. Cada um dos 259 módulos presentes no banco KEGG está representado horizontalmente enquanto as amostras dos dias 1 (ZC3D1), 30 (ZC3D30), 64 (ZC3D64), 78 (ZC3D78) e 99 (ZC3D99) de processo estão representadas verticalmente. O gráfico foi gerado através da plataforma R, módulo *heatmap*, cada módulo foi normalizado de forma que a média dos valores correspondesse a zero e o desvio padrão a 1.

Com o intuito de aprofundar o entendimento da variação no potencial metabólico da comunidade microbiana relacionada ao processo de compostagem, nove amostras seriadas foram coletadas de uma segunda célula (chamada de ZC4) nos dias 1, 3, 7, 15, 30, 64, 67, 78, 99 após a montagem.

Diferente da composteira ZC3, que apresentou a grande maioria de módulos sub- e super-representados no primeiro e no último dia do processo, a composteira ZC4 apresentou uma distribuição de módulos mais uniforme. Contudo, a adição que mais quatro pontos de amostragem podem ser responsáveis por essa distribuição (TABELA 2).

TABELA 2: Distribuição dos módulos KEGG sub- e super-representados nas amostras do Lago São Francisco.

	Super-representados	Sub-representados
ZC4D1	7	5
ZC4D3	3	2
ZC4D7	7	4
ZC4D15	1	0
ZC4D30	2	2
ZC4D64	0	0
ZC4D67	0	0
ZC4D78	2	1
ZC4D99	2	7

O sistema de compostagem do Parque Zoológico é semiestático: o composto é revirado por volta do 60º dia após o preenchimento de cada célula de compostagem. Para ZC3, a revira aconteceu no 65º dia, enquanto para ZC4 aconteceu no 63º. Através do agrupamento hierárquico das amostras em função da abundância dos módulos metabólicos, foi possível identificar um padrão não reconhecido em ZC3: a variação do potencial metabólico da comunidade microbiana em função da presença de oxigênio. Essa constatação foi possível devido ao maior número de pontos amostrais. Assim, no processo de agrupamento, amostras dos dias 1, 3 e 64, um dia após a revira, formaram um grupo, bem como amostras dos dias 30, amostra anterior à revira, e 99. As amostras dos dias 7 e 67 formaram um grupo e as amostras dos dias 15 e 78 formaram o último grupo (FIGURA 6).

Para assegurar que esse agrupamento hierárquico é estatisticamente robusto, uma matriz de distância euclidiana foi obtida a partir da tabela de abundância dos módulos e submetida a uma análise de similaridade (*Analysis of similarity* – ANOSIM; CLARKE, 1993), resultando em um valor $R = 0,8556$ com $p\text{-valor} < 0.0008$. Se se considera apenas os dois grandes grupos formados, dias 1, 3, 7, 64 e

67 e dias 15, 30, 78 e 99, o resultado da análise de similaridade foi $R = 0,7313$ com $p\text{-valor} < 0.008$.

FIGURA 6: *Heatmap* representando a abundância dos módulos KEGG nas amostras da composteira ZC4. Cada um dos 127 módulos presentes no banco KEGG identificados nas amostras de ZC4 está representado horizontalmente enquanto as amostras dos dias 1 (ZC4D1), 3 (ZC4D3), 7 (ZC4D7), 15 (ZC4D15), 30 (ZC4D30), 64 (ZC4D64), 67 (ZC4D67), 78 (ZC4D78) e 99 (ZC4D99) de processo estão representadas verticalmente. O gráfico foi gerado através da plataforma R, módulo *heatmap*, cada módulo foi normalizado de forma que a média dos valores correspondesse a zero e o desvio padrão a 1.

Entre os módulos super-representados nos primeiros sete dias do processo, é possível destacar o módulo referente à glicólise no dia 7 (M00001 – rota de Embden-Meyerhof-Parnas). Um estudo comparativo recente constatou que 97% dos

microrganismos anaeróbicos têm exclusivamente a rota de Embden-Meyerhof-Parnas como rota de metabolismo de glicose, enquanto 71% dos aeróbicos facultativos e 55% dos aeróbicos estritos utilizam exclusivamente essa rota (FLAMHOLZ *et al.*, 2012). Esse mesmo estudo mostrou que 10% dos aeróbicos facultativos e 21% dos aeróbicos estritos têm como única rota de metabolismo de glicose a rota de Entner-Doudoroff. Essas observações podem explicar a variação na abundância dos módulos referentes a essas rotas nos primeiros dias de ZC4: a super-representação do módulo referente à rota de Embden-Meyerhof-Parnas coincide com a menor abundância da rota de Entner-Doudoroff no dia 7, indicando uma possível transição de táxons.

O módulo referente ao sistema de fosfotransferase (PTS) específico para frutose (M00273) estava super-representados no dia 1 e, além dele, módulos referentes ao transporte de carboidratos como o sistema de transporte de maltose/maltodextrina (M00194) e de açúcar (M00197). Sistemas de fosfotransferase transportam essencialmente monossacarídeos e dissacarídeos e a fosforilação do substrato durante o transporte torna os passos seguintes do metabolismo mais eficientes, além de constituir uma forma de regulação da utilização preferencial de determinado substrato por repressão catabólica inibindo, por exemplo, permeases específicas de outros açúcares (BRÜCKNER & TITGEMEYER, 2002; BIDDLE *et al.*, 2013). A prevenção da ativação de genes transportadores e catabólicos também é reconhecida em culturas: quando há mais de um tipo de carboidrato, bactérias tendem consumir somente um tipo por vez (BRÜCKNER & TITGEMEYER, 2002). Por outro lado, transportadores tipo ABC tendem a ter especificidade por oligossacarídeos em detrimento de hexoses (BIDDLE *et al.*, 2013). Os dados obtidos neste estudo reforçam a ideia de uma comunidade microbiana com metabolismo variado e modulada pelos compostos presentes ou oriundos do substrato, capaz de realizar o metabolismo de carboidratos diversos.

Novamente a abundância de módulos referentes à síntese dos aminoácidos prolina, lisina, treonina (M00015, M00016 e M00018 respectivamente) manteve-se estável durante todo o processo.

O módulo referente à síntese do aminoácido triptofano (M00023), assim como em ZC3, estava sub-representado na amostra do dia 1 de ZC4. Embora a atividade

da corismato sintase, essencial na rota dos aminoácidos aromáticos, seja limitada na presença de oxigênio, devido à grande concentração de coenzima oxidada (KITZING *et al.*, 2004; ELY *et al.*, 2008), os módulos referentes à síntese dos outros dois aminoácidos aromáticos (M00024 e M00025) estavam super-representados no dia 1 e ambos aumentaram ligeiramente após a revira, embora esse aumento não possa ser considerado super-representação (FIGURA 7). Vale destacar que o módulo referente à síntese de histidina (M00026) estava sub-representados nas amostras dos dias 30 e 99, os dias mais anóxicos (FIGURA 7).

FIGURA 7: Abundância relativa de módulos KEGG relacionados a metabolismo de aminoácidos.

Os custos metabólicos da síntese de aminoácidos aromáticos são os mais altos, seguidos pelo custo de síntese da histidina (AKASHI & GOJOBORI, 2002). É possível que o oxigênio possibilite a proliferação de microrganismos com metabolismo energeticamente mais custoso capazes de sintetizar os aminoácidos ao invés de microrganismos mais adaptados à captação desses aminoácidos do ambiente.

Os resultados desse trabalho ampliam os conhecimentos acerca do metabolismo da comunidade microbiana envolvida no processo de compostagem. Mesmo se tratando de composteiras diferentes, é possível afirmar que os perfis assemelham-se muito e as diferenças encontradas podem ser atribuídas à natureza dos materiais que compõem o substrato e suas proporções, ao tempo de preenchimento das células de compostagem e de variações climáticas nos anos de coleta.

REFERÊNCIAS

- ABUBUCKER, S.; SEGATA, N.; GOLL, J.; SCHUBERT, A. M.; IZARD, J.; CANTAREL, B. L.; RODRIGUEZ-MUELLER, B.; ZUCKER, J.; THIAGARAJAN, M.; HENRISSAT, B.; WHITE, O.; KELLEY, S. T.; METHE, B.; SCHLOSS, P. D.; GEVERS, P.; MITREVA, M.; HUTTENHOWER, C. Metabolic Reconstruction for Metagenomic Data and Its Application to the Human Microbiome. **PLoS Computational Biology**, v. 8, n. 6, p. 1 – 17, 2012.
- AKASHI, H.; GOJOBORI, T. Metabolic efficiency and amino acid composition in the proteomes of *Escherichia coli* and *Bacillus subtilis*. **Proceedings of the National Academy of Sciences**, v. 99, n. 6, p. 3695 – 3700, 2002.
- ALFREIDER, A.; PETERS, S.; TEBBE, C. C.; RANGGER, A.; INSAM, H. Microbial community dynamics during composting of organic matter as determined by 16S ribosomal DNA analysis. **Compost Science & Utilization**, v. 10, p. 303 – 312, 2002.
- ALTSCHUL, S. F.; GISH, W.; MILLER, W.; MYERS, E. W.; LIPMAN, D. J. Basic local alignment search tool. **Journal of Molecular Biology**, v. 215, p. 403 – 410, 1990.
- BECKER, S. A.; PALSSON, B. Ø. Genome-scale reconstruction of the metabolic network in *Staphylococcus aureus* N315: an initial draft to the two-dimensional annotation. **BMC Microbiology**, v. 5, n. 8, p. 1 – 12, 2005.
- BIDDLE, A.; STEWART, L.; BLANCHARD, J.; LESCHINE, S. Untangling the Genetic Basis of Fibrolytic Specialization by Lachnospiraceae and Ruminococcaceae in Diverse Gut Communities. **Diversity**, v. 5, n. 3, p. 627 – 640, 2013.
- BORENSTEIN, E.; KUPIEC, M.; FELDMAN, M. W.; RUPPIN, E. Large-scale reconstruction and phylogenetic analysis of metabolic environments. **Proceedings of the National Academy of Sciences**, v. 105, n. 38, p. 14482 – 14487, 2008.
- BRÜCKNER, R.; TITGEMEYER, F. Carbon catabolite repression in bacteria: choice of the carbon source and autoregulatory limitation of sugar utilization. **FEMS Microbiology Letters**, v. 209, n. 2, p. 141 – 148, 2002.
- CLARKE, K. R. Non-parametric multivariate analyses of changes in community structure. **Australian Journal of Ecology**, v. 18, p. 117 – 143, 1993.
- COTTRET, L.; MILREU, P. V.; ACUÑA, V.; MARCHETTI-SPACCAMELA, A.; STOUGIE, L.; CHARLES, H.; SAGOT, M. F. Graph-Based Analysis of the Metabolic Exchanges between Two Co-Resident Intracellular Symbionts, *Baumannia cicadellinicola* and *Sulcia muelleri*, with Their Insect Host, *Homalodisca coagulata*. **PLoS Computational Biology**, v. 6, n. 9, p. 1 – 13, 2010.
- EDGAR, R. C. Search and clustering orders of magnitude faster than BLAST. **Bioinformatics**, v. 26, n. 19, p. 2460 – 2461, 2010.

ELY, F.; NUNES, J. E. S.; SCHROEDER, E. K.; FRAZZON, J.; PALMA, M. S.; SANTOS, D. S.; BASSO, L. A. The *Mycobacterium tuberculosis* Rv2540c DNA sequence encodes a bifunctional chorismate synthase. **BMC Biochemistry**, v. 9, n. 13, p. 1 – 16, 2008.

FEIST, A. M.; HENRY, C. S.; REED, J. L.; KRUMMENACKER, M.; JOYCE, A. R.; KARP, P. D.; BROADBELT, L. J.; HATZIMANIKATIS, V.; PALSSON, B. Ø. A genome-scale metabolic reconstruction for *Escherichia coli* K-12 MG1655 that accounts for 1260 ORFs and thermodynamic information. **Molecular Systems Biology**, v. 3, n. 121, p. 1 – 18, 2007.

FITZPATRICK, G. E.; WORDEN, E. C.; VENDRAME, W. A. Historical Development of Composting Technology during the 20th Century. **HortTechnology**, v. 15, n. 1, p. 48 – 51, 2005.

FLAMHOLZ, A.; NOOR, E.; BAR-EVEN, A.; LIEBERMEISTER, W.; MILO, R. Glycolytic strategy as a tradeoff between energy yield and protein cost. **Proceedings of the National Academy of Sciences**, v. 110, n. 24, p. 10039 – 10044, 2012.

FRANKE-WHITTLE, I. H.; KNAPP, B. A.; FUCHS, J.; KAUFMANN, R.; INSAM, H. Application of COMPOCHIP microarray to investigate the bacterial communities of different composts. **Microbial Ecology**, v. 57, n. 3, p. 510 – 521, 2009.

GAJALAKSHMI, S. & ABBASI, S. A. Solid Waste Management by Composting: State of the Art. **Critical Reviews in Environmental Science and Technology**, v. 38, p. 311 – 400, 2008.

KANEHISA, M.; GOTO, S.; KEGG: Kyoto Encyclopedia of Genes and Genomes. **Nucleic Acids Research**, v. 28, n. 1, p. 27 – 30, 2000.

KITZING, K.; AUWETER, S.; AMRHEIN, N.; MACHEROUX, P. Mechanism of Chorismate Synthase: Role Of The Two Invariant Histidine Residues In The Active Site. **The Journal Of Biological Chemistry**, v. 279, n. 10, p. 9451 – 9461, 2004.

KREIMER, A.; DORON-FAIGENBOIM, A.; BORENSTEIN, E.; FREILICH, S. NetCmpt: a network-based tool for calculating the metabolic competition between bacterial species. **BIOINFORMATICS APPLICATIONS NOTE**, v. 28, n. 16, p. 2195 – 2197, 2012.

KUMAR, S. 2011. Composting of municipal solid waste. **Critical reviews in biotechnology**, v.31, p. 112 – 136, 2011.

LEVY, R.; BORENSTEIN, B. Metabolic modeling of species interaction in the human microbiome elucidates community-level assembly rules. **Proceedings of the National Academy of Sciences**, v. 110, n. 31, p. 12804 – 12809, 2013.

LÓPEZ-GONZÁLEZ, J. A.; LÓPEZ, M. J.; VARGAS-GARCÍA, M. C.; SUÁREZ-ESTRELLA, F.; JURADO, M.; MORENO, J. Tracking organic matter and microbiota dynamics during the stages of lignocellulosic waste composting. **Bioresource Technology**, v. 146, p. 574 – 584, 2013.

MARTINS, L. F.; ANTUNES, L. P.; PASCON, R. C.; OLIVEIRA, J. C. F.; DIGIAMPIETRI, L. A.; BARBOSA, D.; PEIXOTO, B. M.; VALLIM, M. A.; VIANANIERO, C.; OSTROSKI, E. H.; TELLES, G. P.; DIAS, Z.; CRUZ, J. B.; JULIANO, L.; VERJOVSKI-ALMEIDA, S.; SILVA, A. M.; SETUBAL, J. C. Metagenomic Analysis of a Tropical Composting Operation at the São Paulo Zoo Park Reveals Diversity of Biomass Degradation Functions and Organisms. **PLoS ONE**, v. 8, n. 4, p 1 – 13, 2013.

OH, Y. K.; PALSSON, B. O.; PARK, S. M.; SCHILLING, C. H.; MAHADEVAN, R. Genome-scale Reconstruction of Metabolic Network in *Bacillus subtilis* Based on High-throughput Phenotyping and Gene Essentiality Data. **The Journal of Biological Chemistry**, v. 282, p. 28791 – 28799, 2007.

PARTANEN P.; HULTMAN, J.; PAULIN, L.; AUVINEN, P.; ROMANTSCHUK, M. Bacterial diversity at different stages of the composting process. **BMC Microbiology**, v. 10, n. 94, p. 1 – 11, 2010.

RYNK, R.; KAMP, M; WILLSON, G. B.; SINGLEY, M. E.; RICHARD, T. L.; KOLEGA, J. J.; GOUIN, F.R.; LALIBERTY JR. L.; KAY, D; MURPHY, D. W.; HOITINK, H. A. J.; BRINTON, W. F. On-Farm composting handbook. **Northeast Regional Agricultural Engineering Service**, Ithaca, N.Y., 1992.

STEGER, K.; SJÖGREN, A. M.; JARVIS, A; JANSSON, J. K.; SUNDH, I. Development of compost maturity and Actinobacteria populations during full-scale composting of organic household waste. **Journal of Applied Microbiology**, n. 103, v. 2, p. 487 – 498, 2007.

YE, Y.; DOAK, T. G. A Parsimony Approach to Biological Pathway Reconstruction/Inference for Genomes and Metagenomes. **PLoS Computational Biology**, v. 5, n. 9, p. 1 – 8, 2009.

APÊNDICE 2

Tabela de Abundância Relativa de Rotas Matabólicas no Sistema de Compostagem do Parque Zoológico

SIGLA	NOME	Nº DE OCORRÊNCIAS POR MILHÃO				
		ZC3D1	ZC3D30	ZC3D64	ZC3D78	ZC3D99
ko00010	Glycolysis / Gluconeogenesis	12187,61	11183,53	11268,71	11041,95	11117,27
ko00020	Citrate cycle (TCA cycle)	10487,65	13612,40	13572,10	13364,54	13399,84
ko00030	Pentose phosphate pathway	16234,60	14296,45	14048,91	14024,98	13302,64
ko00040	Pentose and glucuronate interconversions	8012,89	7540,90	7155,55	6804,17	6811,47
ko00051	Fructose and mannose metabolismo	10346,51	7884,22	7962,12	8032,81	8062,04
ko00052	Galactose metabolismo	8139,35	7118,70	6762,13	6556,28	6472,17
ko00053	Ascorbate and aldarate metabolismo	3423,04	4311,64	4045,35	4502,72	4058,11
ko00061	Fatty acid biosynthesis	13727,70	18150,10	18461,55	18101,94	17000,28
ko00071	Fatty acid metabolismo	4389,00	5931,39	5921,47	5945,16	6774,54
ko00072	Synthesis and degradation of ketone bodies	828,86	1671,30	1970,27	2063,89	2103,49
ko00100	Steroid biosynthesis	271,42	412,00	402,04	462,57	477,55
ko00121	Secondary bile acid biosynthesis	5368,85	1903,87	1881,12	2620,33	1601,99
ko00130	Ubiquinone and other terpenoid-quinone biosynthesis	6106,07	9670,75	9275,59	9438,64	10259,15
ko00140	Steroid hormone biosynthesis	500,04	650,72	539,21	516,53	759,95
ko00190	Oxidative phosphorylation	5777,56	7886,37	8000,25	7905,16	7323,66
ko00195	Photosynthesis	258,72	747,69	859,55	685,91	735,66
ko00196	Photosynthesis - antenna proteins	160,79	190,96	198,63	281,49	326,09
ko00230	Purine metabolismo	10792,71	9431,79	9326,42	9334,75	9143,65
ko00240	Pyrimidine metabolismo	14394,61	12036,41	12096,57	12145,99	11778,79
ko00250	Alanine, aspartate and glutamate metabolismo	13470,07	12269,08	12203,09	12077,80	13165,47
ko00253	Tetracycline biosynthesis	4,00	57,22	28,21	43,55	59,01
ko00260	Glycine, serine and threonine metabolism	11345,65	10815,62	10856,32	10491,90	11233,64
ko00270	Cysteine and methionine metabolismo	11135,60	10721,82	10905,42	10578,09	11197,93
ko00280	Valine, leucine and isoleucine degradation	5202,94	7769,62	7944,32	7916,30	8472,06
ko00281	Geraniol degradation	676,63	5790,96	5811,36	6054,51	6848,71
ko00290	Valine, leucine and isoleucine biosynthesis	17817,78	19509,47	20019,43	19127,24	19553,93
ko00300	Lysine biosynthesis	14020,71	13540,05	13344,39	12661,28	14112,06
ko00310	Lysine degradation	2915,90	4457,70	4533,24	4556,42	4954,95
ko00311	Penicillin and cephalosporin biosynthesis	1328,05	2287,09	2212,01	1898,43	2266,65
ko00312	beta-Lactam resistance	1000,19	644,96	696,73	659,63	385,32
ko00330	Arginine and proline metabolismo	7569,38	8123,51	8063,87	7736,66	8345,78

ko00331	Clavulanic acid biosynthesis	44,51	6,00	10,74	29,28	12,57
ko00340	Histidine metabolismo	13403,49	12729,81	13013,27	13044,58	13121,79
ko00350	Tyrosine metabolismo	4979,17	5279,17	5189,60	5157,49	6152,21
ko00360	Phenylalanine metabolismo	4038,54	4676,45	4721,41	4389,64	5245,10
ko00361	Chlorocyclohexane and chlorobenzene degradation	87,30	1339,86	1306,68	1423,09	1268,83
ko00362	Benzoate degradation	2946,13	3742,26	3721,86	3788,00	4234,17
ko00363	Bisphenol degradation	983,31	731,92	622,27	585,11	986,51
ko00364	Fluorobenzoate degradation	17,52	64,03	28,66	23,57	30,46
ko00380	Tryptophan metabolismo	2869,72	4283,95	4305,23	4335,56	4626,33
ko00400	Phenylalanine, tyrosine and tryptophan biosynthesis	11126,10	11637,02	11550,17	11448,71	11729,35
ko00410	beta-Alanine metabolismo	5510,19	7350,87	7040,61	7005,10	7009,14
ko00430	Taurine and hypotaurine metabolismo	6844,54	6898,73	6775,08	6598,13	6356,48
ko00440	Phosphonate and phosphinate metabolismo	2819,56	2975,41	2679,75	2676,79	2117,29
ko00450	Selenocompound metabolismo	12760,33	12803,06	12731,10	12691,62	14003,80
ko00460	Cyanoamino acid metabolismo	6939,53	6029,66	6021,92	5765,70	6083,69
ko00471	D-Glutamine and D-glutamate metabolismo	26465,82	20004,73	20209,11	19305,73	17402,43
ko00472	D-Arginine and D-ornithine metabolism	704,15	927,78	619,01	583,96	1607,87
ko00473	D-Alanine metabolismo	17610,49	16028,09	16894,74	16758,19	18053,98
ko00480	Glutathione metabolismo	6364,42	8159,32	7983,12	8404,75	8288,60
ko00500	Starch and sucrose metabolismo	9088,48	7073,90	6684,92	6794,34	7510,64
ko00510	N-Glycan biosynthesis	1123,05	641,93	695,17	745,96	973,02
ko00511	Other glycan degradation	4734,56	2548,69	2336,40	2163,63	2351,17
ko00520	Amino sugar and nucleotide sugar metabolismo	11223,70	9546,69	9392,23	9237,45	9758,51
ko00521	Streptomycin biosynthesis	17063,12	15228,34	14783,99	15134,47	16408,00
ko00522	Biosynthesis of 12-, 14- and 16-membered macrolides	35,71	113,64	93,73	95,46	151,91
ko00523	Polyketide sugar unit biosynthesis	5543,81	3881,72	3757,83	4179,57	3799,98
ko00531	Glycosaminoglycan degradation	2892,21	1497,46	1293,91	1119,95	1443,06
ko00532	Glycosaminoglycan biosynthesis - chondroitin sulfate	25,15	12,06	17,07	20,09	37,44
ko00540	Lipopolysaccharide biosynthesis	12228,79	9546,44	9110,34	10349,26	9816,16
ko00550	Peptidoglycan biosynthesis	18979,62	16437,44	16607,24	16321,41	16844,39
ko00561	Glycerolipid metabolismo	5480,76	5311,77	5317,04	5000,05	4878,86
ko00562	Inositol phosphate metabolismo	3512,19	3982,40	3966,14	3698,01	4603,02
ko00563	Glycosylphosphatidylinositol(GPI)-anchor biosynthesis	75,19	33,11	31,09	34,92	224,81

ko00564	Glycerophospholipid metabolismo	7655,97	6575,74	6283,28	6450,46	6286,14
ko00565	Ether lipid metabolismo	324,67	355,04	300,86	342,26	436,53
ko00590	Arachidonic acid metabolismo	1828,04	2184,35	2011,54	2365,76	42,41
ko00591	Linoleic acid metabolismo	201,22	365,11	320,47	314,11	195,86
ko00600	Sphingolipid metabolismo	494,64	638,56	576,68	564,92	599,92
ko00601	Glycosphingolipid biosynthesis - lacto and neolacto series	2,41	0,64	0,93	0,76	23,90
ko00620	Pyruvate metabolismo	10906,80	11959,86	12205,77	11979,18	12504,68
ko00621	Dioxin degradation	25,23	31,63	27,77	31,03	29,46
ko00622	Xylene degradation	26,93	28,96	35,72	50,38	45,45
ko00623	Toluene degradation	512,60	54,34	52,64	64,63	68,38
ko00624	Polycyclic aromatic hydrocarbon degradation	2070,28	2783,72	2477,28	2572,41	3255,81
ko00625	Chloroalkane and chloroalkene degradation	4800,32	5453,10	5408,09	5601,38	6422,33
ko00626	Naphthalene degradation	4152,85	4271,03	4172,23	4273,59	5309,62
ko00627	Aminobenzoate degradation	3189,48	3982,82	3891,30	3852,18	4309,58
ko00630	Glyoxylate and dicarboxylate metabolismo	6290,04	7483,07	7298,04	7375,53	8220,51
ko00633	Nitrotoluene degradation	1725,26	1520,30	1690,47	1600,99	1460,26
ko00640	Propanoate metabolismo	6821,95	9267,60	9452,88	8817,62	8901,78
ko00642	Ethylbenzene degradation	41,14	42,90	56,22	57,95	61,60
ko00643	Styrene degradation	33,28	81,79	95,28	69,22	110,20
ko00650	Butanoate metabolismo	6889,00	8850,13	8972,97	8697,83	9315,12
ko00660	C5-Branched dibasic acid metabolism	11281,89	19055,57	19455,09	17900,14	17572,53
ko00670	One carbon pool by folate	18926,18	16170,82	16919,06	17197,49	17806,62
ko00680	Methane metabolismo	5461,83	4442,88	4574,65	4480,72	4535,58
ko00710	Carbon fixation in photosynthetic organisms	2081,89	1400,67	1460,56	1414,59	1885,37
ko00720	Carbon fixation pathways in prokaryotes	10555,01	12072,29	12364,53	11865,91	12223,04
ko00730	Thiamine metabolismo	13914,92	15697,16	16270,66	15000,63	15487,21
ko00740	Riboflavin metabolismo	10062,82	11344,20	11046,06	11105,92	10781,02
ko00750	Vitamin B6 metabolism	12042,78	11431,42	11480,06	11172,97	12353,15
ko00760	Nicotinate and nicotinamide metabolismo	12993,61	11658,46	11259,96	11817,28	11271,54
ko00770	Pantothenate and CoA biosynthesis	17648,28	18340,27	18456,95	18281,19	17511,78
ko00780	Biotin metabolismo	13894,86	13379,19	13048,89	13224,79	13755,47
ko00785	Lipoic acid metabolismo	15994,82	18228,72	18556,58	19073,17	17672,85
ko00790	Folate biosynthesis	13015,10	14277,91	14494,07	15564,10	15122,60

ko00791	Atrazine degradation	976,71	969,22	894,63	801,64	1198,83
ko00830	Retinol metabolismo	1160,56	1496,54	1583,29	1598,89	1770,10
ko00860	Porphyrin and chlorophyll metabolismo	5981,52	7771,02	7920,15	7864,92	7579,97
ko00900	Terpenoid backbone biosynthesis	13952,50	11975,60	12484,98	11915,73	12072,78
ko00901	Indole alkaloid biosynthesis	59,55	287,10	292,34	325,98	369,76
ko00903	Limonene and pinene degradation	3645,74	4610,84	4144,21	4347,47	5631,58
ko00906	Carotenoid biosynthesis	1028,10	1578,97	1300,32	1343,84	1317,87
ko00908	Zeatin biosynthesis	8157,95	5741,76	5739,86	5893,13	6447,82
ko00909	Sesquiterpenoid biosynthesis	18,93	68,86	21,41	34,58	25,63
ko00910	Nitrogen metabolismo	6154,84	5871,38	5935,51	6115,88	6282,44
ko00920	Sulfur metabolismo	7758,19	9012,92	9108,10	9444,54	9982,31
ko00930	Caprolactam degradation	504,11	1152,51	1094,24	1050,03	1393,14
ko00940	Phenylpropanoid biosynthesis	2280,54	2435,07	2155,43	2082,39	2386,25
ko00941	Flavonoid biosynthesis	525,65	1033,34	923,30	1150,66	1320,78
ko00943	Isoflavonoid biosynthesis	23,77	68,00	38,63	76,87	113,21
ko00960	Tropane, piperidine and pyridine alkaloid biosynthesis	4807,14	3930,60	3835,43	3634,92	4672,98
ko00965	Betalain biosynthesis	444,25	959,72	1098,81	1246,53	1281,70
ko00970	Aminoacyl-tRNA biosynthesis	19337,52	16013,49	16500,45	16171,42	16791,37
ko00983	Drug metabolism - other enzymes	18005,93	12745,53	12756,91	12371,73	10465,97
ko01040	Biosynthesis of unsaturated fatty acids	3397,85	5119,59	4655,02	5061,41	5367,24
ko01053	Biosynthesis of siderophore group nonribosomal peptides	1343,70	1249,94	1239,20	1233,85	1550,19
ko01055	Biosynthesis of vancomycin group antibiotics	26521,01	17740,64	16939,75	19996,80	18874,04
ko01056	Biosynthesis of type II polyketide backbone	107,38	762,60	970,43	833,28	1221,84
ko01057	Biosynthesis of type II polyketide products	24,39	76,61	128,47	76,49	109,52
ko02010	ABC transportes	6027,03	7613,63	7498,72	7036,33	6710,73
ko02020	Two-component system	4184,81	5330,95	5191,49	5262,03	5660,06
ko02030	Bacterial chemotaxis	6545,72	11762,67	11969,13	11145,46	10813,99
ko02040	Flagellar assembly	6818,28	13505,77	13356,34	12724,01	11298,62
ko02060	Phosphotransferase system (PTS)	4720,03	5196,18	4224,48	4661,82	3269,50
ko03008	Ribosome biogenesis in eukaryotes	1043,33	803,09	804,73	927,54	936,90
ko03010	Ribosome	34784,71	28502,83	28202,24	28534,67	19465,87
ko03013	RNA transport	926,83	598,87	591,41	680,00	733,12
ko03015	mRNA surveillance pathway	116,09	68,17	82,93	74,28	87,40

ko03018	RNA degradation	6768,01	7070,80	6922,76	7238,68	6343,12
ko03020	RNA polymerase	7779,29	6822,51	6658,55	6794,25	6753,81
ko03022	Basal transcription factors	282,39	69,42	133,99	130,38	153,54
ko03030	DNA replication	15091,42	13555,77	13707,62	14024,07	14078,85
ko03050	Proteasome	253,29	386,78	286,69	346,75	437,88
ko03060	Protein export	18349,12	17884,76	18557,18	19027,48	14946,12
ko03070	Bacterial secretion system	10569,14	10285,69	10823,04	11250,98	9118,59
ko03410	Base excision repair	11461,87	11842,83	12166,63	12334,61	13032,09
ko03420	Nucleotide excision repair	7755,72	6237,07	6387,31	6556,77	7746,18
ko03430	Mismatch repair	18884,21	16724,76	16685,78	16890,17	16532,54
ko03440	Homologous recombination	18137,86	15038,32	15437,60	15488,81	15535,24
ko03450	Non-homologous end-joining	241,60	1091,86	1265,75	1275,14	1248,25
ko04112	Cell cycle - Caulobacter+C155	12135,86	12063,40	12175,94	12123,63	12091,84
ko04113	Meiosis – yeast	129,86	140,92	135,34	131,11	192,54
ko04122	Sulfur relay system	14444,50	17494,99	18809,34	18532,06	17283,04
ko04141	Protein processing in endoplasmic reticulum	522,86	396,48	395,42	453,37	528,57
ko04142	Lysosome	631,51	361,34	320,62	291,24	336,58
ko04146	Peroxisome	2547,22	3063,54	3045,79	3251,92	3389,15
ko04210	Apoptosis	242,13	163,74	138,00	154,00	144,13
ko04621	NOD-like receptor signaling pathway	710,95	281,78	241,69	320,30	362,53
ko04626	Plant-pathogen interaction	1302,36	906,78	933,01	944,70	986,37
ko04910	Insulin signaling pathway	959,96	521,00	521,92	570,69	665,78
ko04962	Vasopressin-regulated water reabsorption	4,03	28,74	19,21	17,34	26,36
ko04974	Protein digestion and absorption	550,56	356,83	414,73	418,37	430,95
ko04976	Bile secretion	85,28	91,14	78,63	69,99	98,65
ko05010	Alzheimer's disease	623,06	626,65	624,76	619,76	591,33
ko05012	Parkinson's disease	290,77	457,76	446,69	452,05	418,16
ko05100	Bacterial invasion of epithelial cells	928,91	237,75	230,12	300,13	184,40
ko05110	Vibrio cholerae infection	49,44	47,80	34,44	54,21	71,31
ko05111	Vibrio cholerae pathogenic cycle	2323,85	2334,81	2489,37	2702,03	2665,73
ko05120	Epithelial cell signaling in Helicobacter pylori infection	705,48	246,50	222,47	216,35	174,42
ko05130	Pathogenic Escherichia coli infection	16,32	34,96	27,77	38,98	23,09
ko05131	Shigellosis	16,50	23,74	12,80	22,66	8,93

ko05142	Chagas disease (American trypanosomiasis)	178,24	179,61	178,24	217,07	210,72
ko05143	African trypanosomiasis	629,36	439,12	377,17	447,12	502,65
ko05145	Toxoplasmosis	15,68	96,82	81,63	87,07	71,37
ko05146	Amoebiasis	331,35	335,61	352,11	315,38	350,07
ko05150	Staphylococcus aureus infection	366,22	395,44	344,59	382,81	155,07
ko05322	Systemic lupus erythematosus	15,46	23,39	31,77	60,37	77,54

APÊNDICE 3

Tabela de Abundância Relativa de Módulos no Sistema de Compostagem do Parque Zoológico

ID	NOME	OCORRÊNCIAS POR MILHÃO				
		ZC3D1	ZC3D30	ZC3D64	ZC3D78	ZC3D99
M00001	Glycolysis (Embden-Meyerhof pathway), glucose => pyruvate	16040,82	13448,68	13466,32	12865,38	13643,72
M00002	Glycolysis, core module involving three-carbon compounds	14926,37	13027,90	13381,90	12768,31	13384,36
M00003	Gluconeogenesis, oxaloacetate => fructose-6P	16489,42	10391,60	10863,60	10438,66	10672,07
M00004	Pentose phosphate pathway (Pentose phosphate cycle)	8397,73	10066,09	10008,51	9703,27	9876,54
M00005	PRPP biosynthesis, ribose 5P => PRPP	16577,97	13604,60	12253,90	14575,14	14253,69
M00006	Pentose phosphate pathway, oxidative phase, glucose 6P => ribulose 5P	6049,80	8624,21	8213,06	8423,09	8667,33
M00007	Pentose phosphate pathway, non-oxidative phase, fructose 6P => ribose 5P	13724,05	12086,90	12808,57	11442,33	11477,84
M00008	Entner-Doudoroff pathway, glucose-6P => glyceraldehyde-3P + pyruvate	3703,11	5729,68	3799,84	5382,76	4908,05
M00009	Citrate cycle (TCA cycle, Krebs cycle)	6694,97	10465,16	9942,06	10466,25	11142,77
M00010	Citrate cycle, first carbon oxidation	7878,93	10614,34	9677,95	10782,53	11760,57
M00011	Citrate cycle, second carbon oxidation	6599,42	10377,65	10107,55	10285,23	10802,30
M00012	Glyoxylate cycle	4600,14	6719,93	6690,65	6780,66	8202,28
M00015	Proline biosynthesis, glutamate => proline	10684,07	11795,77	11185,26	10347,03	11622,96
M00016	Lysine biosynthesis, aspartate => lysine	10649,71	11188,72	10837,88	10294,59	11821,45
M00017	Methionine biosynthesis, aspartate => homoserine => methionine	7078,16	7644,58	7091,88	6826,88	7411,19
M00018	Threonine biosynthesis, aspartate => homoserine => threonine	12008,47	12486,16	12146,13	11086,05	12194,84
M00019	Leucine biosynthesis, pyruvate => 2-oxoisovalerate => leucine	1682,32	0,00	0,00	0,00	7581,95
M00020	Serine biosynthesis, glyceraldehyde-3P => serine	11169,46	8226,61	8337,94	8938,48	9607,89
M00021	Cysteine biosynthesis, serine => cysteine	14360,86	11954,84	12467,78	13631,16	12001,34
M00022	Shikimate pathway, phosphoenolpyruvate + erythrose-4P => chorismate	13203,81	12022,99	11948,47	11201,50	11624,81
M00023	Tryptophan biosynthesis, chorismate => tryptophan	7668,57	11330,15	10932,55	10686,71	12593,78

M00024	Phenylalanine biosynthesis, chorismate => phenylalanine	3055,66	2365,14	2716,83	3538,74	4408,29
M00025	Tyrosine biosynthesis, chorismate => tyrosine	4075,46	3091,73	3390,81	4353,24	5367,45
M00026	Histidine biosynthesis, PRPP => histidine	9342,31	9828,32	10445,14	10125,95	9859,88
M00027	GABA (gamma-Aminobutyrate) shunt	529,06	486,22	136,61	581,00	825,74
M00028	Ornithine biosynthesis, glutamate => ornithine	7061,10	7438,84	7500,89	7349,62	8659,67
M00033	Ectoine biosynthesis	732,28	2889,46	2205,32	2658,20	3979,69
M00034	Methionine salvage pathway	1158,06	2282,85	2927,13	3002,54	874,38
M00035	Methionine degradation	5414,82	6107,31	5713,66	5949,37	7538,84
M00036	Leucine degradation, leucine => acetoacetate + acetyl-CoA	1940,96	4419,26	4395,32	4690,17	4963,51
M00045	Histidine degradation, histidine => N-formiminoglutamate => glutamate	6256,18	5112,36	5037,86	4997,11	5469,24
M00048	Inosine monophosphate biosynthesis, PRPP + glutamine => IMP	15000,77	13799,72	14039,00	13480,16	13950,20
M00049	Adenine nucleotide biosynthesis, IMP => ADP/dADP,ATP/Datp	11858,35	7992,96	8035,36	9415,87	9974,14
M00050	Guanine nucleotide biosynthesis, IMP => GDP/dGDP,GTP/dGTP	12328,68	8435,73	8403,74	9682,81	10191,07
M00051	Uridine monophosphate biosynthesis, glutamine (+ PRPP) => UMP	14114,05	0,00	10117,75	8116,11	0,00
M00052	Pyrimidine ribonucleotide biosynthesis, UMP => UDP/UTP,CDP/CTP	12956,75	14914,82	15405,03	15060,59	15739,17
M00053	Pyrimidine deoxyribonucleotide biosynthesis, CDP/CTP => dCDP/dCTP,dTDP/dTTP	10913,68	9291,47	9518,91	10989,67	11067,68
M00060	Lipopolysaccharide biosynthesis, KDO2-lipid A	2923,67	1634,02	304,80	3487,72	105,05
M00061	Uronic acid metabolismo	3235,36	1285,29	562,05	1,60	2094,03
M00063	CMP-KDO biosynthesis	8168,97	6767,83	6576,65	7521,84	6805,72
M00064	ADP-L-glycero-D-manno-heptose biosynthesis	5634,17	3297,62	3166,23	4264,14	3469,35
M00081	Pectin degradation	206,34	0,03	0,00	0,00	0,00
M00083	Fatty acid biosynthesis, elongation	11247,56	15593,69	15739,67	15153,10	16006,82
M00086	beta-Oxidation, acyl-CoA synthesis	14639,66	11824,81	11610,91	11624,81	14272,89
M00087	beta-Oxidation	2244,62	5152,27	5178,13	5040,38	5849,92

M00088	Ketone body biosynthesis, acetyl-CoA => acetoacetate/3-hydroxybutyrate/acetone	989,75	39,77	0,11	73,98	103,75
M00091	Phosphatidylcholine (PC) biosynthesis, PE => PC	0,00	0,07	1,52	0,00	1102,84
M00093	Phosphatidylethanolamine (PE) biosynthesis, PA => PS => PE	13400,37	9042,68	9224,55	9107,70	8594,75
M00095	C5 isoprenoid biosynthesis, mevalonate pathway	2892,22	848,00	670,71	76,47	0,63
M00096	C5 isoprenoid biosynthesis, non-mevalonate pathway	12664,66	10227,06	10366,48	10150,82	10781,98
M00097	beta-Carotene biosynthesis, GGAP => beta-carotene	496,89	1818,89	1117,56	953,57	1517,65
M00098	Acylglycerol degradation	559,65	37,27	12,47	5,82	733,07
M00115	NAD biosynthesis, aspartate => NAD	9732,88	10287,41	10127,99	9818,81	9912,12
M00116	Menaquinone biosynthesis, chorismate => menaquinone	0,00	343,07	1,64	2430,36	2565,32
M00117	Ubiquinone biosynthesis, prokaryotes, chorismate => ubiquinone	3133,37	5108,43	1150,62	5936,84	5674,59
M00118	Glutathione biosynthesis, glutamate => glutathione	4544,07	5332,62	6147,80	5866,16	6446,81
M00119	Pantothenate biosynthesis, valine/L-aspartate => pantothenate	10598,03	11286,57	11266,11	11566,10	9960,10
M00120	Coenzyme A biosynthesis, pantothenate => CoA	15455,97	12612,45	13156,42	11910,59	14715,26
M00121	Heme biosynthesis, glutamate => protoheme/siroheme	5835,09	9717,53	9684,26	9716,38	9873,96
M00123	Biotin biosynthesis, pimeloyl-CoA => biotin	6744,57	6654,66	6564,94	7415,46	6882,64
M00124	Pyridoxal biosynthesis, erythrose-4P => pyridoxal-5P	2908,11	23,29	25,78	294,99	405,36
M00125	Riboflavin biosynthesis, GTP => riboflavin/FMN/FAD	11306,41	13041,75	12532,56	12688,67	12849,98
M00126	Tetrahydrofolate biosynthesis, GTP => THF	10797,65	10628,60	10463,49	10331,69	11285,99
M00127	Thiamine biosynthesis, AIR => thiamine-P/thiamine-2P	7031,81	8217,63	8514,12	7884,57	7377,38
M00133	Polyamine biosynthesis, arginine => agmatine => putrescine => spermidine	4486,83	7289,90	7457,79	7288,44	7544,14
M00134	Polyamine biosynthesis, arginine => ornithine => putrescine	1630,46	2564,16	2783,62	2915,51	2534,71

M00135	GABA biosynthesis, eukaryotes, putrescine => GABA	1377,55	4372,87	4073,47	4317,77	4746,65
M00140	C1-unit interconversion, prokaryotes	13312,93	8485,20	9901,47	8903,21	8458,21
M00144	Complex I (NADH dehydrogenase), NADH dehydrogenase I	5650,40	10799,54	10989,97	10066,64	10364,75
M00149	Complex II (succinate dehydrogenase / fumarate reductase), succinate dehydrogenase	7216,65	9891,43	10126,69	9948,55	8747,74
M00150	Complex II (succinate dehydrogenase / fumarate reductase), fumarate reductase	3,56	2058,54	152,62	11,94	0,01
M00151	Complex III (Cytochrome bc1 complex)	3542,99	4864,04	4533,17	4858,72	4798,66
M00153	Cytochrome bd complex	4172,54	5762,51	5607,86	5414,63	5539,78
M00155	Complex IV (Cytochrome c oxidase), cytochrome o ubiquinol oxidase/cytochrome c oxidase/quinol oxidase polypeptide	0,00	0,60	1844,64	627,39	417,76
M00156	Complex IV (Cytochrome c oxidase), cytochrome c oxidase, cbb3-type	0,00	18,39	4413,82	5330,73	5493,93
M00157	F-type ATPase, bactéria	15530,83	16746,13	16290,53	16604,27	14354,25
M00159	V-type ATPase, prokaryotes	5686,97	534,95	1606,59	1464,10	0,00
M00164	ATP synthase	15530,83	16746,13	16290,53	16604,27	14354,25
M00166	Reductive pentose phosphate cycle, RuBP + CO2 => glyceraldehyde-3P	0,00	944,94	0,22	729,32	205,66
M00168	CAM (Crassulacean acid metabolism), dark	4650,17	7406,18	6894,12	7400,20	8738,45
M00169	CAM (Crassulacean acid metabolism), light	5924,32	5402,40	4840,43	5285,72	6760,38
M00170	C4-dicarboxylic acid cycle, phosphoenolpyruvate carboxykinase type	6664,17	7250,41	7372,37	6931,29	8590,43
M00174	Methane oxidation, methylotroph, methane => CO2	0,07	11,96	17,71	0,02	29,69
M00175	Nitrogen fixation, N2 => ammonia	1207,42	0,00	0,05	1,49	155,16
M00178	Ribosome, bactéria	29868,09	24033,72	23946,11	24000,45	17444,37
M00183	RNA polymerase, bactéria	19632,80	14199,70	13844,32	13827,93	14530,02
M00185	Sulfate transport system	3207,85	4869,92	4750,89	5110,91	4503,46
M00186	Putative tungstate transport system	46,93	1072,73	74,57	74,51	195,25
M00188	Sulfonate/nitrate/taurine transport system	5954,92	8579,99	8444,76	8025,65	8166,43
M00189	Molybdate transport system	5758,14	6969,39	6554,42	7047,42	6788,65
M00190	Iron(III) transport system	6704,95	7104,84	7232,24	6774,75	6752,35

M00191	Thiamine transport system	199,13	917,95	555,62	267,56	755,22
M00193	Putative spermidine/putrescine transport system	1223,96	1684,16	1727,50	1530,30	1734,24
M00194	Maltose/maltodextrin transport system	7136,17	5518,99	5191,56	4727,24	4970,13
M00197	Putative sugar transport system	8647,80	6178,58	5783,60	5154,12	5653,94
M00198	sn-Glycerol 3-phosphate transport system	2050,27	4491,63	4159,39	3618,95	4379,04
M00200	Sorbitol/mannitol transport system	8,73	240,53	4,55	17,09	18,82
M00206	Cellobiose transport system	3,87	249,31	64,65	7,19	149,60
M00207	Multiple sugar transport system	5025,64	6222,79	5777,54	5241,47	6246,12
M00208	Glycine betaine/proline transport system	2551,55	6276,81	6686,27	5496,38	7182,48
M00209	Osmoprotectant transport system	3591,55	7952,34	8438,62	7093,93	6302,42
M00210	Putative ABC transport system	5888,97	7756,48	7095,37	8252,85	6435,84
M00211	Putative ABC transport system	2272,57	2729,19	2865,49	2829,00	2112,94
M00212	Ribose transport system	5704,38	7008,87	6371,40	5609,80	6199,10
M00213	L-Arabinose transport system	166,22	0,01	0,00	0,00	0,00
M00214	Methyl-galactoside transport system	2032,03	130,36	0,00	0,18	0,17
M00215	D-Xylose transport system	1850,81	2477,30	2298,25	2259,64	2506,12
M00216	Putative multiple sugar transport system	1727,02	179,62	35,82	7,61	433,63
M00219	AI-2 transport system	89,00	0,05	0,14	0,04	0,00
M00221	Simple sugar transport system	8427,99	5660,72	5353,30	5022,59	5093,57
M00222	Phosphate transport system	9502,53	13303,78	12915,30	11952,67	11854,31
M00223	Phosphonate transport system	2813,83	3731,65	3410,16	3818,87	3499,05
M00224	Putative phosphonate transport system	74,03	1,90	0,10	0,59	25,69
M00225	Lysine/arginine/ornithine transport system	10,48	21,59	0,00	0,38	0,00
M00227	Glutamine transport system	630,31	1422,38	1472,72	1406,47	1127,78
M00228	Putative glutamine transport system	1645,50	2002,58	1436,08	1709,53	1864,11
M00230	Glutamate/aspartate transport system	327,46	798,33	291,02	122,81	1,50
M00232	General L-amino acid transport system	163,59	3293,64	3415,67	3074,44	2642,42
M00233	Glutamate transport system	44,74	525,88	145,89	2,34	357,61
M00234	Cystine transport system	1557,16	1493,30	1526,91	930,84	831,12
M00236	Polar amino acid transport system	10525,38	6969,83	6670,34	6485,62	5535,33
M00237	Branched-chain amino acid transport system	5984,73	10704,61	11114,55	9159,70	9934,53
M00238	D-Methionine transport system	6808,46	6180,73	6586,34	5657,93	5781,84
M00239	Peptides/nickel transport system	9841,93	12737,07	12596,57	11410,43	10968,32

M00240	Iron complex transport system	10132,25	12422,83	12420,10	11233,37	10625,64
M00241	Vitamin B12 transport system	1,35	51,62	5,10	0,00	0,05
M00242	Zinc transport system	11096,31	8340,92	9032,16	8851,84	7251,28
M00243	Manganese/iron transport system	1407,75	3741,99	3469,24	3278,53	3477,27
M00244	Zinc/manganese transport system	2419,83	2821,09	2471,83	2845,04	3008,97
M00245	Cobalt transport system	345,41	4208,00	3811,75	5,53	352,17
M00246	Nickel transport system	0,00	0,01	0,00	0,00	71,81
M00247	Putative ABC transport system	5185,97	5741,48	5796,00	4002,85	3728,69
M00248	Antibiotic transport system	11043,07	13794,05	13227,33	14038,51	14550,57
M00249	Capsular polysaccharide transport system	68,86	35,46	1,56	1,24	3,15
M00250	Lipopolysaccharide transport system	6017,77	4166,23	4053,92	4628,41	4517,62
M00251	Teichoic acid transport system	543,44	1995,60	921,26	967,42	2,29
M00252	Lipooligosaccharide transport system	77,36	2087,97	2423,26	2022,59	2792,27
M00253	Sodium transport system	528,78	1704,88	1643,29	1872,09	1783,54
M00254	ABC-2 type transport system	12690,17	14936,40	14998,06	14099,42	13162,93
M00255	Lipoprotein-releasing system	9797,52	7649,25	7345,30	8183,04	9053,65
M00256	Cell division transport system	13797,86	9188,16	9498,89	10659,71	10349,99
M00258	Uncharacterized ABC transport system	16042,53	14404,60	14359,61	14514,14	13076,46
M00259	Heme transport system	1472,90	6143,36	6039,42	5692,73	4658,10
M00264	DNA polymerase II complex, Archaea	482,81	0,01	0,06	0,00	34,47
M00265	PTS system, glucose-specific II component	2099,85	1438,36	1445,51	1743,85	326,99
M00266	PTS system, maltose and glucose-specific II component	337,95	3,57	5,50	5,76	46,76
M00267	PTS system, N-acetylglucosamine-specific II component	1692,81	49,49	56,10	25,05	104,86
M00268	PTS system, arbutin-like II component	72,90	0,03	1,39	0,00	0,00
M00269	PTS system, sucrose-specific II component	580,87	0,35	20,72	1,16	4,45
M00270	PTS system, trehalose-specific II component	2766,94	845,24	1345,86	1709,78	1048,27
M00271	PTS system, beta-glucosides-specific II component	2873,55	99,44	51,57	16,67	52,48
M00273	PTS system, fructose-specific II component	4012,24	3595,27	3752,36	4046,82	3119,26
M00274	PTS system, mannitol-specific II component	1120,40	1324,55	1679,79	2183,21	1793,88
M00275	PTS system, cellobiose-specific II component	3275,84	3157,69	3758,15	3753,75	2078,34
M00276	PTS system, mannose-specific II component	2322,97	2238,78	2538,12	2507,39	1183,81

M00277	PTS system, N-acetylgalactosamine-specific II component	141,72	0,26	0,00	4,12	0,02
M00279	PTS system, galactitol-specific II component	1,54	2520,01	1916,36	2455,02	1345,79
M00280	PTS system, glucitol/sorbitol-specific II component	101,82	146,26	323,48	411,52	1177,96
M00281	PTS system, lactose-specific II component	100,88	60,01	153,02	203,81	0,05
M00282	PTS system, D-glucosamine-specific II component	235,17	38,02	92,09	24,84	161,89
M00283	PTS system, ascorbate-specific II component	2106,47	3788,61	3867,73	4725,18	3000,28
M00299	Spermidine/putrescine transport system	10504,33	6517,57	6468,65	6161,67	4655,86
M00300	Putrescine transport system	3403,02	3470,85	3037,64	3167,22	2778,51
M00307	Pyruvate oxidation, pyruvate => acetyl-CoA	9445,17	13230,97	13420,95	13300,43	13423,80
M00308	Semi-phosphorylative Entner-Doudoroff pathway, gluconate => glyceraldehyde-3P + pyruvate	225,54	85,38	8,16	13,16	167,58
M00309	Non-phosphorylative Entner-Doudoroff pathway, gluconate => glyceraldehyde + pyruvate	19,94	0,00	0,00	0,00	0,00
M00310	Pyruvate:ferredoxin oxidoreductase	1596,99	3288,14	3286,74	2394,37	1857,98
M00311	2-oxoglutarate:ferredoxin oxidoreductase	8891,38	3019,35	3780,93	3239,37	2038,00
M00313	indolepyruvate:ferredoxin oxidoreductase	7879,24	1160,97	1233,18	704,20	946,35
M00314	Bacitracin transport system	29,36	1,96	1,28	0,02	0,00
M00317	Manganese/iron transport system	199,89	1734,13	1123,34	638,56	1073,58
M00318	Iron/zinc/copper transport system	90,63	1,37	0,00	0,00	0,00
M00319	Manganese/zinc/iron transport system	16,09	2035,38	1791,94	1167,71	1352,66
M00320	Lipopolysaccharide export system	5547,07	6231,51	5291,76	5888,12	6641,54
M00324	Dipeptide transport system	1338,70	2905,37	2260,60	2103,23	2937,61
M00336	Twin-arginine translocation (Tat) system	649,90	4911,66	6038,49	8845,82	3914,04
M00338	Cysteine biosynthesis, homocysteine + serine => cysteine	315,08	1491,47	1518,90	1773,09	2911,75
M00342	Bacterial proteasome	19,04	644,61	541,67	739,02	1941,27
M00343	Archaeal proteasome	262,99	54,11	0,95	1,72	70,64
M00344	Formaldehyde assimilation, xylulose monophosphate pathway	343,89	1000,24	1930,23	2164,37	3538,61
M00345	Formaldehyde assimilation, ribulose monophosphate pathway	4808,78	264,12	3823,45	3688,35	774,09
M00348	Glutathione transport system	1149,18	3540,27	3231,10	2888,25	2936,82
M00349	Microcin C transport system	1682,76	2133,24	1773,89	2094,94	1670,54

M00358	Coenzyme M biosynthesis	16,22	0,00	0,00	0,00	0,00
M00359	Aminoacyl-tRNA biosynthesis, eukaryotes	16705,38	10829,76	10944,98	10985,25	12025,23
M00360	Aminoacyl-tRNA biosynthesis, prokaryotes	17023,04	12144,03	12360,02	12065,56	13408,29
M00364	C10-C20 isoprenoid biosynthesis, bactéria	7003,71	5470,47	6285,88	5642,67	5317,91
M00365	C10-C20 isoprenoid biosynthesis, archaea	3205,96	3157,86	2952,22	2944,09	3341,34

CAPÍTULO 4:

CONSIDERAÇÕES FINAIS E PERSPECTIVAS

As comunidades microbianas portam a maioria das rotas bioquímicas e são responsáveis diretas pelos ciclos biogeoquímicos no planeta. Essas rotas têm uma distribuição ampla nos diversos táxons conhecidos.

Dados metagenômicos produzidos por sequenciamento *shotgun* em larga escala, resultando em milhões de fragmentos de algumas centenas de pares de bases, têm fornecido subsídio para a identificação dos componentes das comunidades microbianas e para montagens de genomas parciais ou completos, mas trouxe o desafio de integrar todos esses dados de forma a fornecerem informações acerca das relações entre espécies e relações com o ambiente.

Os microbiomas objetos desta tese são bastante diferentes. De um lado o Lago São Francisco, um reservatório artificial que consiste no maior corpo d'água do parque, sujeito a fatores climáticos, receptor de água de nascentes, residência de aves migratórias e não-migratórias e mamíferos. De outro lado o sistema de compostagem de resíduos orgânicos produzidos no parque, um ambiente confinado, com substrato complexo de várias origens: podas, restos de alimentos dos animais e fezes.

Este trabalho foi desenvolvido para estabelecer uma metodologia de estudos do potencial metabólico de microbiomas que possa ser aplicada em dados metagenômicos seriais. O artigo “Metabolic Reconstruction for Metagenomic Data and Its Application to the Human Microbiome” (ABUBUCKER *et al.*, 2012) foi um dos pilares da abordagem deste projeto. Neste artigo os autores descrevem o funcionamento do programa HUMAnN bem como sua aplicação no estudo de microbiomas associados a humanos. A consistência dos resultados descritos nos estimulou a aplicar uma abordagem semelhante aos metagenomas oriundos de ambientes da Fundação Parque Zoológico.

De forma geral a aplicação do programa HUMAnN gerou resultados significativos, não só em relação ao potencial metabólico das comunidades

microbianas mas também em relação a futuros estudos metagenômicos e metatranscritômicos integrados. Todos os conjuntos de dados puderam ser analisados de forma simples e sem distinção da tecnologia de sequenciamento 454 ou Illumina, mas para evitar distorções foi aplicada uma normalização simples em função do tamanho do metagenoma em questão.

A metodologia aplicada permitiu reconhecer a variação do potencial metabólico na forma de variação da abundância relativa de módulos KEGG. Vários módulos tiveram variação mais ou menos significativa e a variação de alguns deles coincidiu, como descrito, com a variação de táxons ou de fatores ambientais, abrindo caminho para investigações futuras sobre quais táxons estão contribuindo com quais rotas. Já a análise computacional do potencial metabólico compostagem, por se tratar de um processo complexo desempenhado por um consórcio de microrganismos num sistema fechado, teve como etapa crítica o estabelecimento do que é super-representado ou sub-representado em um determinado momento. É factível que todas as rotas metabólicas conhecidas estejam representadas em maior ou menor grau, em parte devido à complexidade do “inóculo” inicial, ou seja, o conjunto de microrganismos presentes no material utilizado na montagem da composteira.

Como perspectiva futura, com mencionado acima, um estudo metatranscritômico será realizado com aplicação do programa HUMAnN e outras análises estatísticas com o intuito de esclarecer quais genes previstos no potencial metabólico estão efetivamente expressos e se há alguma correlação entre a abundância de uma rota e a expressão de algum de seus componentes. Para tanto amostras de RNA mensageiro foram sequenciadas e serão num futuro próximo analisadas.

Para averiguar se há possíveis relações de interdependência entre as espécies identificadas nos microbiomas do Parque Zoológico, considerando produtos metabólicos produzidos por uma espécie (sementes) e utilizados e não produzido por outras, será empregado o algoritmo NetSeed proposto por Borenstein e colaboradores (BORENSTEIN *et al.*, 2008; COTTRET *et al.*, 2010), que se baseia em teoria dos grafos e constrói redes metabólicas para os organismos com informações depositadas em bancos de dados como KEGG. Dois outros artigos que merecem destaque são “Large-scale reconstruction and phylogenetic analysis of

metabolic environments” (BORENSTEIN *et al.*, 2008) e “Metabolic modeling of species interaction in the human microbiome elucidates community-level assembly rules” (LEVY & BORENSTEIN, 2013). No primeiro consta a descrição do algoritmo NetSeed e a análise do conjunto de compostos exógenos necessários para a sobrevivência de 478 espécies de microrganismos bem como as implicações para a determinação do habitat. O segundo descreve a aplicação da mesma abordagem a dados de microbiomas humanos com um delineamento da coocorrência de microrganismos, suas relações e o papel do hospedeiro em seu estabelecimento. Esta abordagem pode ser adaptada para outros dados metagenômicos e no caso do processo de compostagem pode fornecer dados importantes para a compreensão do estabelecimento e do papel dos microrganismos atuantes.

Deste trabalho resultou até o momento uma coautoria num artigo descrevendo resultados preliminares da análise de duas composteiras (MARTINS *et al.*, 2013). Em breve deveremos submeter um manuscrito de primeira autoria correspondente ao capítulo 2 desta tese e outro manuscrito de coautoria em que nossa contribuição é o que consta aqui como capítulo 3.

REFERÊNCIAS

ABUBUCKER, S.; SEGATA, N.; GOLL, J.; SCHUBERT, A. M.; IZARD, J.; CANTAREL, B. L.; RODRIGUEZ-MUELLER, B.; ZUCKER, J.; THIAGARAJAN, M.; HENRISSAT, B.; WHITE, O.; KELLEY, S. T.; METHE, B.; SCHLOSS, P. D.; GEVERS, P.; MITREVA, M.; HUTTENHOWER, C. Metabolic Reconstruction for Metagenomic Data and Its Application to the Human Microbiome. **PLoS Computational Biology**, v. 8, n. 6, p. 1 – 17, 2012.

BORENSTEIN, E.; KUPIEC, M.; FELDMAN, M. W.; RUPPIN, E. Large-scale reconstruction and phylogenetic analysis of metabolic environments. **Proceedings of the National Academy of Sciences**, v. 105, n. 38, p. 14482 – 14487, 2008.

COTTRET, L.; MILREU, P. V.; ACUÑA, V.; MARCHETTI-SPACCAMELA, A.; STOUGIE, L.; CHARLES, H.; SAGOT, M. F. Graph-Based Analysis of the Metabolic Exchanges between Two Co-Resident Intracellular Symbionts, *Baumannia cicadellinicola* and *Sulcia muelleri*, with Their Insect Host, *Homalodisca coagulata*. **PLoS Computational Biology**, v. 6, n. 9, p. 1 – 13, 2010.

LEVY, R.; BORENSTEIN, B. Metabolic modeling of species interaction in the human microbiome elucidates community-level assembly rules. **Proceedings of the National Academy of Sciences**, v. 110, n. 31, p. 12804 – 12809, 2013.

MARTINS, L. F.; ANTUNES, L. P.; PASCON, R. C.; OLIVEIRA, J. C. F.; DIGIAMPIETRI, L. A.; BARBOSA, D.; PEIXOTO, B. M.; VALLIM, M. A.; VIANANIERO, C.; OSTROSKI, E. H.; TELLES, G. P.; DIAS, Z.; CRUZ, J. B.; JULIANO, L.; VERJOVSKI-ALMEIDA, S.; SILVA, A. M.; SETUBAL, J. C. Metagenomic Analysis of a Tropical Composting Operation at the São Paulo Zoo Park Reveals Diversity of Biomass Degradation Functions and Organisms. **PLoS ONE**, v. 8, n. 4, p 1 – 13, 2013.