Microbial diversity in freshwater samples and their contaminating human and bovine hosts

Nancy C Stoppe Corresp., [1] , **Tatiana T Torres** [2] , **Fabricio C Leotti** [2] , **Maria Inês Z Sato** [3] , **Laura MM Ottoboni** [1]

[1] CBMEG, Universidade Estadual de Campinas, Campinas, SP, Brazil

[2] Instituto de Biociências, Universidade de São Paulo, São Paulo, São Paulo, Brazil

[3] Departamento de Análises Ambientais, Companhia Ambiental do Estado de São Paulo, São Paulo, São Paulo, Brazil

Corresponding Author: Nancy C Stoppe
Email address: ncstoppe@gmail.com

DNA extracted from feces (human and bovine) and water samples was used for the massive pyrosequencing of the hypervariable V3 region of the 16S rRNA gene, revealing 4296 operational taxonomic units (OTUs). The greatest diversity was observed in samples of cattle feces, and the smallest diversity was found in a pristine water sample. *Firmicutes* was the predominant group in samples of human feces, while in bovine feces the dominant groups were *Firmicutes* and *Bacteroidetes*. The interaction network showed that the stool samples had the greatest diversity and, among the water samples, the one with human pollution source had the highest diversity. The LEfSe method was used to identify host biomarkers. *Actinobacteria*, *Betaproteobacteria*, and *Firmicutes* were identified as human biomarkers, while for cattle, the potential markers were *Bacteroidetes*, *Tenericutes*, and *Spirochaetes*. Host-specific markers were identified, but were not found in the water samples, suggesting either that the tools used did not have the resolution to identify markers in environmental samples, or that the contamination in the water bodies was mixed. Additionally, as the host-specific markers were isolated from non-autochthonous microorganisms, they could be affected by adverse environmental effects including physical-chemical factors and competition with native organisms.

MICROBIAL DIVERSITY IN FRESHWATER SAMPLES AND THEIR CONTAMINATING HUMAN AND BOVINE HOSTS

Stoppe NC[1*], Torres TT[2], Leotti, FC[2], Sato MIZ[3], Ottoboni, LMM[1]

[1]CBMEG – Centro de Biologia Molecular e Engenharia Genética, UNICAMP, Campinas, Brazil

[2]Departamento de Genética e Biologia Evolutiva, IBUSP, São Paulo, Brazil

[3]CETESB – Companhia Ambiental do Estado de São Paulo, São Paulo, Brazil

* Corresponding author: ncstoppe@gmail.com

Abstract

DNA extracted from feces (human and bovine) and water samples was used for the massive pyrosequencing of the hypervariable V3 region of the 16S rRNA gene, revealing 4296 operational taxonomic units (OTUs). The greatest diversity was observed in samples of cattle feces, and the smallest diversity was found in a pristine water sample. *Firmicutes* was the predominant group in samples of human feces, while in bovine feces the dominant groups were *Firmicutes* and *Bacteroidetes*. The interaction network showed that the stool samples had the greatest diversity and that among the water samples the highest diversity was in the one affected by a human pollution source. The LEfSe method was used to identify host biomarkers. *Actinobacteria*, *Betaproteobacteria*, and *Firmicutes* were identified as human biomarkers, while the potential markers for cattle were *Bacteroidetes*, *Tenericutes*, and *Spirochaetes*. Host-specific markers were identified, but were not found in the water samples, suggesting either that the tools used did not have the resolution to identify markers in environmental samples, or that there was mixed contamination in the water bodies. Additionally, as the host-specific markers were isolated from non-autochthonous microorganisms, they could be affected by adverse environmental effects including physical-chemical factors and competition with native organisms.

Keywords: Microbial source tracking, water pollution, pyrosequencing, 16S rRNA.

1. Introduction

Water is essential to sustain life and is liable to fecal contamination from a range of point and nonpoint sources, with potential contributions from wildlife, domesticated animals, and humans (Roslev and Bukh, 2011). Reliable and accurate fecal source identification methods are essential for the development of better management practices for the control of fecal contamination from relevant animal sources, for the protection of recreational water users from waterborne pathogens, and for preserving the integrity of drinking water sources. In microbial source tracking, different library-dependent and independent methods for bacteria have been used for the identification of fecal host markers and for determining the sources of water pollution (USEPA, 2005).

Among these methods, new-generation sequencing techniques can be used as a tool in microbial source tracking. Amplicon libraries with specific barcodes of hypervariable regions of the 16S rRNA gene followed by sequencing are used in studies of comparative microbial ecology (Clingenpeel et al., 2011). This approach allows several samples to be analyzed simultaneously, which reduces the cost. However, few studies have been conducted for the prospection of markers for microbial source tracking (Jeong et al., 2011; Unno et al., 2011).

It is still a challenge to associate host molecular markers with environmental samples containing the natural microbiota as well as possible mixed sources of pollution. Another challenge has been the use of reliable computational tools to identify specific biomarkers in metagenomics data. Segata et al. (2011) proposed the linear discriminant analysis effect size method (LEfSe), which uses linear discriminant analysis for the statistically significant separation of two or more classes. This classification is checked by statistical tests that evaluate the biological consistency. The method has been successfully used for comparison of the mice gut microbiota of obese and gastric bypass individuals (Liou et al., 2013).

Here, next generation sequencing of the V3 region of the 16S rRNA gene was used in host and water samples with the aim of identifying specific molecular markers from contaminating hosts and detecting their presence in water samples affected by different sources of pollution.

2. Material and Methods

2.1. Sample collection

Fresh feces were collected from six healthy humans (H1 to H6) (three males and three females, aged from 21 to 58 years) living in São Paulo city. None of them had received antibiotics within at least six months before sampling. The Research Ethics Committee of the State University of Campinas School of Medical Sciences approved the present study (Permission 046/11), and all participants gave their informed written consent. Feces were also collected from six bovines (B1 to B6) from three different sites in São Paulo State: two cattle sheds, at Serra do Cambara farm (23°30'20'' S; 47°39'45'' W) and Zanella farm (23°3'58'' S; 47°49'7'' W), and the riparian area of the Tietê River in Laranjal Paulista (22°47'25'' S; 47°49'23'' W) (Figure 1). The samples were divided into 1 g portions that were placed in sterilized bags and kept at -80 ^{o}C until processed.

Water samples (5-10 L) were collected from two sites on the Tietê River, one located near to the source of the river, in a pristine environment (W1), and the other located in an agricultural area with large areas of pasture for raising cattle (W3). The coordinates of these sites were 23°33'54'' S; 46°00'57'' W and 22°47'25'' S; 47°49'23'' W, respectively. A water sample affected mainly by human fecal contamination (W2) was collected from the Billings Reservoir, at 23°46'37'' S; 46°32'1'' W. An additional sample from a pristine site (W4) was collected at the Ipiranga River, located in an environmental protection area (23°20'9.4"S; 45°08'1.4"W) (Figure 1). Water samples were collected in fresh bottles that had been specially treated (washing with DNA-removing disinfectant, rinsing with DNA-free water, and autoclave sterilization). The samples were filtered onto a 0.22 µm membrane filter, which was used for DNA extraction (Haugland et al., 2005).

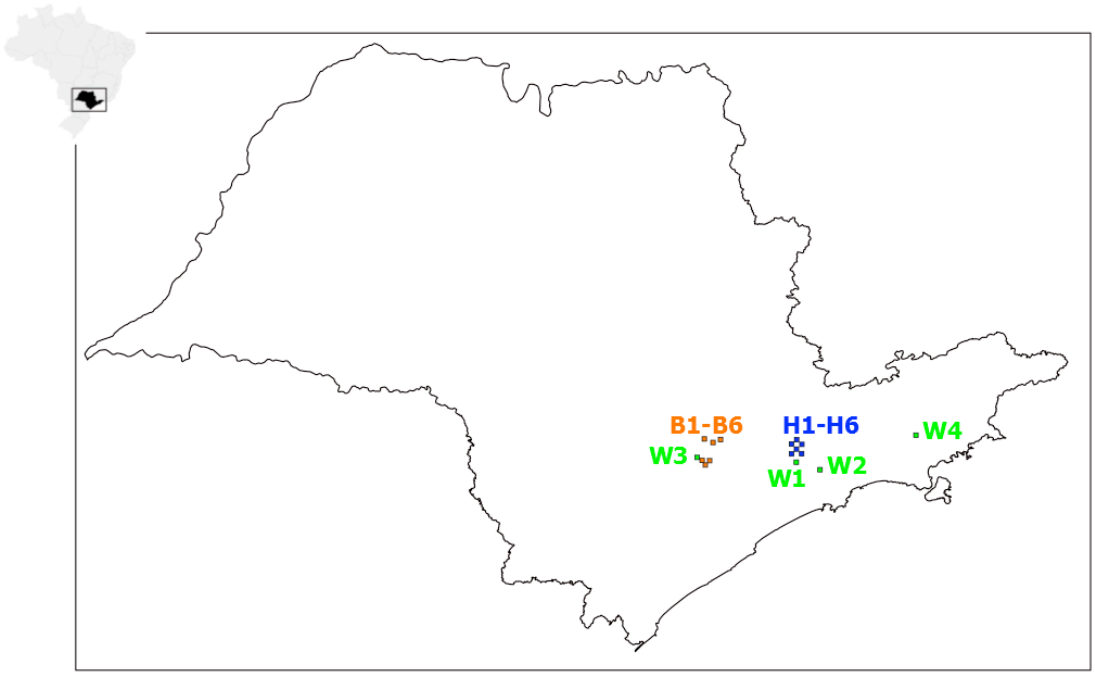

Figure 1 – Sampling sites. H1-H6 are the human feces sample collection sites (blue). B1-B6 are the bovine feces sample collection sites (orange). W1 to W4 are the water sample collection sites (green). W1 is located near the source of the Tietê River (pristine environment). W2 is located at the Billings Reservoir (human fecal contamination). W3 is located in an agricultural area with large areas of pasture for raising cattle (bovine fecal contamination). W4 is located at the Ipiranga River, within an environmental protection area (pristine environment).

2.2. DNA extraction

Genomic DNA was extracted from the feces and water samples using mechanical and chemical lyses, according to Ahlroos and Tynkkynen (2009), with minor modifications. The samples were thawed and resuspended in 1:10 EDTA (50 mmol/L), and mechanical lysis was performed in a Stomacher for 5-10 min. The solution was centrifuged at 16,000*g* for 5 min and the precipitate was resuspended in a solution containing final concentrations of 25 µmol/mL EDTA, 120 µg/mL lysozyme, and 0.6 U/mL mutanolysin, followed by incubation for 1 h at 37 °C. The mixture was centrifuged at 16,000*g* for 5 min and the pellet was submitted to DNA extraction using the Wizard Genomic DNA purification kit (Promega).

2.3. V3 region of 16S rDNA amplification

PCR amplification of the hypervariable V3 region of the 16S rRNA gene was performed as described by Clingenpeel et al. (2011), with minor modifications. The template was amplified using 338F (5'-ACTCCTACGGGAGGCAGCAG-3') and 533R (5'-TTACCGCGGCTGCTGGCAC -3') primers for 15 cycles, with addition of 10 ng of DNA to the amplification mix containing 1U AccuPrime *Pfx* DNA polymerase (Invitrogen), 5 µL Accuprime reaction mix, and 0.6 µM of each primer, in a final volume of 50 µL. The PCR conditions were an initial denaturation at 94 ^{o}C for 1 min, 15 cycles of 94 ^{o}C for 15 s, 62 ^{o}C for 30 s, and 72 ^{o}C for 30 s, and a final extension at 72 ^{o}C for 2 min. After that, five further cycles were performed using a 1:2.5 or 1:5 dilution of the first PCR product as a template. The same PCR amplification conditions were employed, but using specific barcoded 338F primers and the adaptor A for each sample, and the same adaptor B for all samples. The PCR products were submitted to 2.0% agarose gel electrophoresis and the 265 bp DNA band was excised, purified using Illustra GFX (GE Healthcare), and checked for purity using 1.5% agarose gel electrophoresis. Quantification was performed using a Qubit fluorometer (Invitrogen). Amplicons from all the samples were pooled in equimolar concentrations to a final concentration of 23.4 ng/µL, for subsequent 454 pyrosequencing carried out using an FLX 454 system at the DNA Facility of the University of Iowa.

2.4. Data analysis

The QIIME v. 1.5.0 pipeline (Caporaso et al., 2010a) was used for data analysis, with sequences selected according to quality (minimum of 25) and size (150-240 nt). The sequences were clustered into operational taxonomic units (OTUs) using UCLUST, with a 97% similarity threshold. Representative sequences from each OTU cluster were aligned using PyNAST (Caporaso et al., 2010b), and taxonomy was assigned using the Ribosomal Database Project (RDP) Classifier Program (Wang et al., 2007). A minimum sequence confidence of 80% was used for OTU classification. The OTU table, containing the OTU abundance in each sample, was exported to the Cytoscape program to build a network (Shannon et al., 2003). The samples were attributed as source interactions, and the OTUs were the target interactions. A rarefied OTU table was built for estimation of microbial diversity. The linear discriminant analysis effect size method (LEfSe) was used to find specific OTUs (biomarkers) for differentiation of the human and bovine samples, as previously described by Segata et al. (2011), where the non-parametric factorial sum-rank test was used to detect features with significant differential abundance with respect to the class of interest. Further, biological significance was

investigated with a set of pairwise tests among subclasses, using the (unpaired) Wilcoxon rank-sum test. Finally, linear discriminant analysis was used to estimate the effect size of each differentially abundant feature.

3. Results and Discussion

Although a 10 L water sample was collected at site W4, the amount of DNA obtained was insufficient for PCR analysis. This could have been due to the oligotrophic environment, characterized by low levels of organic matter. However, in the case of the other pristine site (W1), 10 L of water provided sufficient material for analysis, while only 5 L amounts were needed from sites W2 and W3. In total, 91,056 filtered sequences (150 to 240 nt) were obtained, ranging from 1,111 (W1) to 17,978 (H4).

The total number of OTUs was 4,296, ranging from 192 (W1) to 1172 (B5), with the bovines presenting the greatest diversity and the water samples the least. Nevertheless, W2 presented a higher number of OTUs, compared to the human samples.

Most of the OTUs were singletons, as can be seen in the OTU heatmap (available at http://biologia.ib.usp.br/torres/Nancy/otu_table.html).

The phylum distribution showed a predominance of *Firmicutes* among the human samples (average of 81.4%), while *Firmicutes* and *Bacteriodetes* were the most common phyla in the bovine samples (43 and 16.2%, respectively). The microbial communities present in the digestive tract are influenced by its anatomical structure. Ruminants and humans have different types of digestive tract, but share many bacterial groups (Krause and Kafipour, 2011). Our results also showed this similarity at lower hierarchical levels, but at the genus level *Faecalibacterium* and *Blautia* were more frequent in the human samples. Some groups were abundant in bovine feces, but neither of them was classified at this level. At the order level, *Bacillales* accounted for 14.4% in one of the bovine samples (B2), but contributed less than 1% in the other samples. Individual differences in the microbial community are frequently observed in humans, mainly among those with different feeding habits (Arumugam et al., 2011; Lozupone et al., 2012). These results indicated that even animals with less variable feeding habits could present intra-specific differences in their microbiota.

In the case of the water samples, *Proteobacteria* was a common phylum (W1 - 26%, W2 - 27.5%, and W3 - 67.7%). The subsequent most common phyla were dependent on the degree of pollution. The

pristine site (W1) presented a high level of *Cyanobacteria* (12.6%), while the human polluted site (W2) and site W3 showed the *Actinobacteria* group as a common phylum (18.7 and 14.8%, respectively). In general terms, the microbial communities from the pristine site (W1) and the human polluted site (W2) showed similar compositions that were even observed at the genus level for the most abundant OTUs as well as the less prevalent ones (Figure 2).

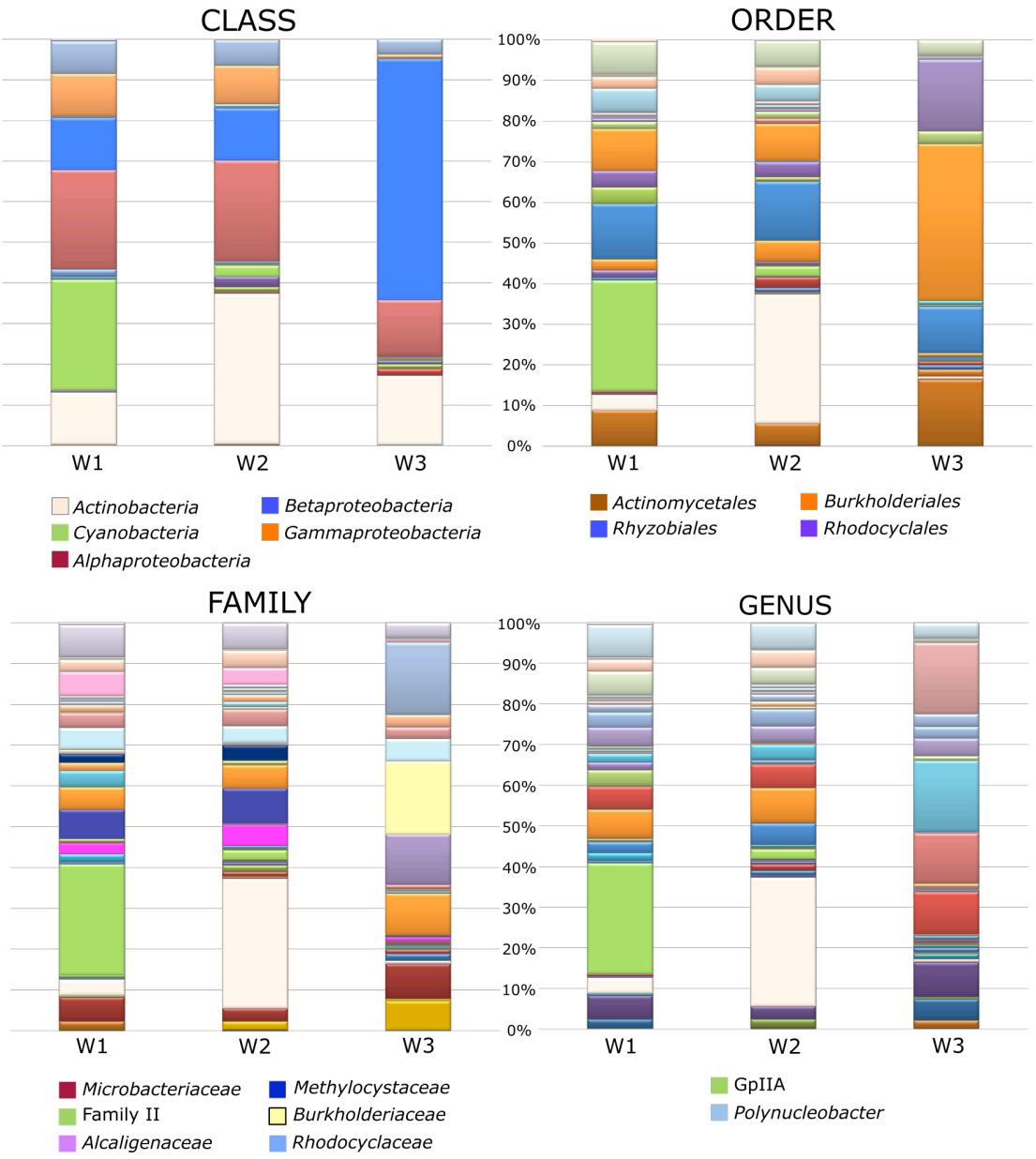

Figure 2 – Microbial communities in water samples, at different taxonomic levels (class, order, family, and genus). The legends show the most frequent groups (>5%).

Cyanobacteria are frequently found in aquatic environments and are associated with biogeochemical cycling and photosynthesis, as well as with bloom events that produce toxins. Among them, GpIIA was the group most frequently observed in the pristine sample (W1). This group is an important contributor to the primary production in oceans (Friedline et al., 2012), and it has been found in freshwaters in China (Cheng et al., 2011). Hence, the highest abundance of this group at site W1, together with its decrease at the polluted sites (W2 and W3), suggests that it might be suitable for use as a marker for pristine environments. Its decreased abundance could be due to higher levels of organic matter, which would encourage the growth of heterotrophic microorganisms, while impairing the growth of cyanobacteria.

The *Actinobacteria* class was found in both water and feces samples, although its frequency in feces samples was less than 1%. Actinobacteria comprise a broad group that includes commensal species of medical and clinical interest, as well as autochthonous environmental species, supporting the association of this group with the water samples.

In contrast, sample W3 presented a different bacterial distribution, with 51.5% of the OTUs belonging to the *Betaproteobacteria* class, and the *Burkholderiaceae* (15.5%), *Rhodocyclaceae* (15.3%), *Alcaligenaceae* (10.8%), and *Microbacteriaceae* (6.2%) being the most common families. These groups are frequently found in water samples, and high levels of *Betaproteobacteria* have been associated with anthropogenic activities (Patel et al., 2014). W3 was greatly affected by bovine fecal pollution, which could have been responsible for the increased abundance of *Betaproteobacteria*. However, neither of these families was found in the bovine feces samples. The second most common phylum was *Actinobacteria* (14.8%), and at the order level, *Actinomycetales* was most common. At the family and genus levels, the majority of the OTUs were unclassified (Figure 2). 'Candidatus Planktophila limnetica' is the group most frequently found in freshwaters worldwide (Warnecke et al., 2005; Jezbera et al., 2009). However, the alignment of unclassified OTUs was clustered with 89% sequence similarity with Candidatus. The sequence similarity of the ribosomal gene currently used for definition of the same species was below 97%, suggesting that the unclassified OTUs belonged to another species.

The W1 and W2 sampling sites were located in the same watershed, under the influence of the Metropolitan Area of São Paulo. However, they were 100 km distant from each other and the watercourses, and they were not interrelated. Furthermore, the former (near the source of the Tietê

River) was in a lotic system in a pristine environment, while the latter (at the artificial Billings Reservoir) was in a lentic system highly impacted by different pollution sources. These sites have been monitored by the São Paulo State Environment Agency for the last 40 years, and in the last ten years have presented water quality index values classified as good (W1) and bad (W2), respectively. The water quality index (WQI) is derived from a combined set of variables including pH, dissolved oxygen, biological oxygen demand, *E. coli*, water temperature, total nitrogen, total phosphorus, total suspended matter, and turbidity (CETESB, 2015).

The interaction network showing the distribution and relationships between the OTUs and the samples is provided in Figure 3. The human and bovine feces samples showed the greatest OTU diversity. Among the water samples, the sample with human pollution (W2) presented the greatest diversity. Exclusive connections can be seen between the human samples and W2, comprising *Bacteroides* and *Bifidobacteriaceae*. These groups have been extensively described as indicators of human contamination and have been used as human markers (Kildare et al., 2007; Layton et al., 2006; Lee et al., 2011; The Human Microbiome Project Consortium, 2012). The bovine and W3 samples showed *Acetobacteriaceae* as a potential bovine marker, despite the fact that it was not one of the most common groups found in bovine fecal microbiota (Dowd et al., 2008). It is notable that the exclusive OTUs were exclusively shared by humans and W2, and by bovines and W3, suggesting that these OTUs could be used as biomarkers. Nevertheless, the low frequency of these OTUs raises doubts about their suitability as good biomarkers.

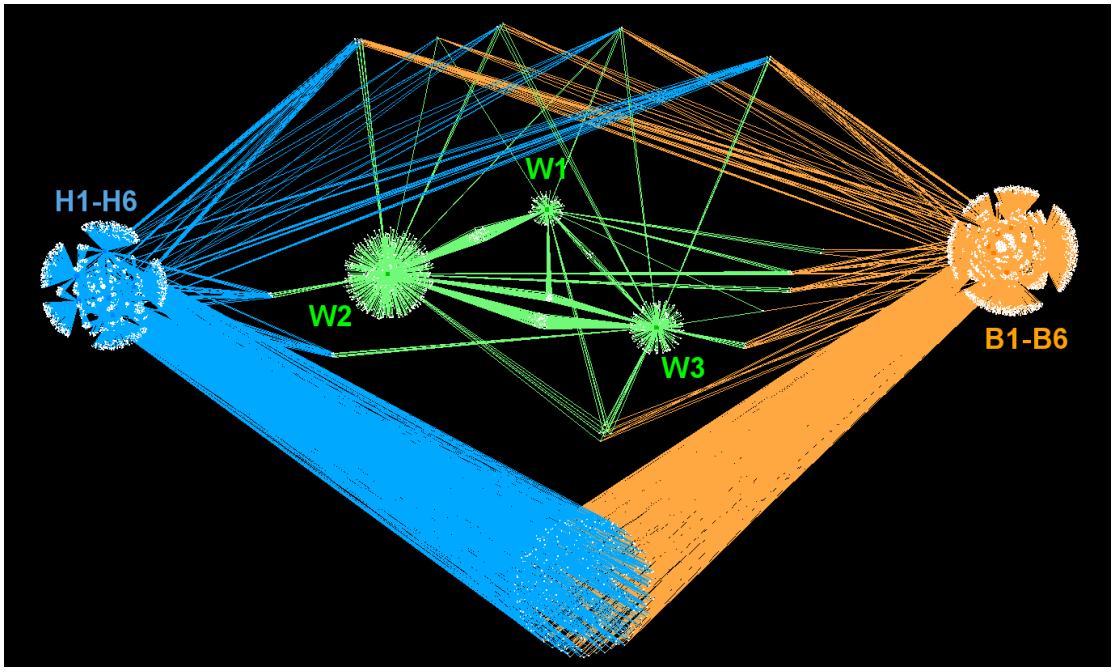

Figure 3 – Network interaction among the samples studied. Blue squares represent the human samples, orange squares represent the bovine samples, and green squares represent the water samples. White circles represent the different OTUs, and the lines represent the interactions between the samples and the OTUs.

Richness was evaluated using the rarefied OTUs table, and the metrics showed similar results, with a clear asymptote for the majority of the samples indicating that OTU coverage was achieved, with the exception of W1 and B6.

Beta diversity showed separate clusters for humans and bovines, highlighting that these samples were clearly different from each other. For the three water samples, no clusters were observed, showing that the microbiota were unique in each sample. The water samples presented different characteristics, providing an explanation for the differences among the microbial communities.

Application of the LEfSe method (Segata et al., 2011) showed the presence of specific biomarkers for feces samples. For humans, the following groups were identified: *Actinobacteria*, *Betaproteobacteria*, and *Firmicutes* (*Clostridia*). The groups found for bovines were *Bacteroidetes* (*Flavobacterium* and *Bacteroidia*), *Tenericutes*, and *Spirochaetes* (Figure 4).

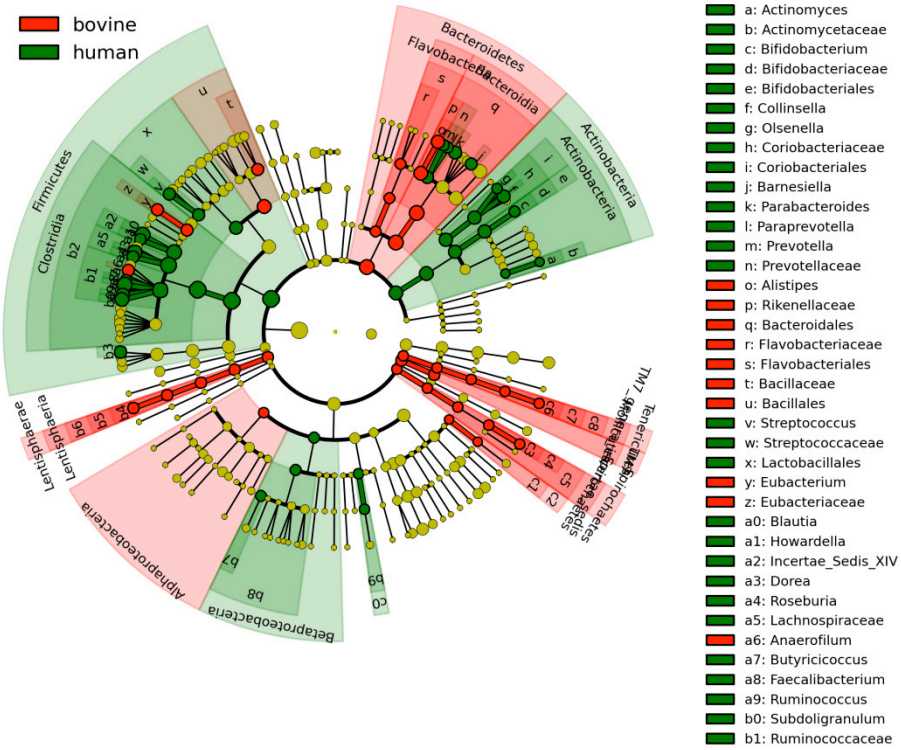

Figure 4 – Cladogram showing specific markers for human and bovine feces, using the LEfSe method. Red colors indicate the bovine markers, and green colors indicate the human markers.

For the water samples, no biomarkers could be identified using LEfSE. As water bodies are dynamic environments, replicate samples should be used, which might enable the identification of specific markers.

The identification of specific markers found in hosts suggests that they could be used as a tool for source tracking, it is recommended that further studies at other sites should therefore be undertaken to test their applicability.

4. Conclusions

The microbiota was similar among human samples and among bovine samples. The interaction network among samples showed shared OTUs between human feces and water with human contamination (W2), as well as between bovine feces and water with bovine contamination. However, the OTUs were found in low frequencies and could not be used as host markers or for tracking pollution sources. Even at the W2 site, which received huge discharges of untreated wastewater, the

main groups found in human samples were not observed. Despite the high level of human pollution, the autochthonous water microbiota seems to have had a competitive advantage. Unexpectedly, sites W1 (pristine environment) and W2 (human contamination) presented similar microbial communities, suggesting that the degree of pollution played a minor role in determining microbiota composition in the water. Host-specific markers were found that might be able to be used for microbial source tracking. Nevertheless, they were not able to identify water samples according to the source of pollution, suggesting either that the low levels of the markers in water samples hampered their detection, or that adverse environmental effects (such as physical-chemical factors and the native microbiota) restricted the presence of these markers. The results indicated that source tracking should not use a single tool, reflecting the complexity of environmental samples.

Acknowledgements

We thank the São Paulo State Research Foundation (FAPESP) for supporting this work. NCS received a fellowship from FAPESP (project no. 2010/15222-0). TTT received a research fellowship from the National Council for Technological and Scientific Development (CNPq). We thank Elayse Maria Hachich from CETESB for providing the use of microbiological laboratory facilities.

References

Ahlroos T, Tynkynen S (2009) Quantitative strain-specific detection of *Lactobacillus rhamnosus* GG in human faecal samples by real-time PCR. J Appl Microbiol 106: 506-514

Arumugam, M, Raes, J, Pelletier E, Lepaslier D, Yamada T, Mende DR, Fernandes GR, Tap J, Bruls T, Batto JM, Gertalan M, Borruel N, Caseelas F, Fernandez L, Gautier L, Hansen T, HattoriI M, Haushi, T, Kleerebenzem M, Kurosawa K, Leclerc M, Levenez, F, Manichanh C, Nielsen HB, Nielsen T, Pons K, Poulain, J, Qin J, Sicheritz-Ponten T, Tims S, Torrents D, Ugarte E, Zoetendal EG, Wang J, Guarner F, Pedersen O, DeVos WM, Brunak S, Dore J, Metahit Consortium, Weissenbach J, Ehrlich SD, Bork P (2011) Enterotypes of the human gut microbiome. Nature 473: 174-180

Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, Costello E, Fierer N, Peña AG, Goodrich JK, Gordon JI, Hutley GA, Kelley ST, Knights D, Koenig JE, Ley RE, Lozupone CA, McDonald D, Muegge BD, Pirrung M, Reeder J, Sevinsky JR, Turnbaugh PJ, Waters WA, Widmann J,

Yatsunenko T, Zaneveld J, Knight R (2010a) QIIME allows analysis of high-throughput community sequencing data. Nat Meth 7: 335-336

Caporaso JG, Bittinger K, Bushman FD, DeSantis TZ, Andersen GK, Knight R (2010b) PyNAST: a flexible tool for aligning sequences to a template alignment. Bioinformatics 26: 266-267

CETESB (2015) http://aguasinteriores.cetesb.sp.gov.br/wp-content/uploads/sites/32/2013/11/agua-doce-parte1-corrigido.pdf. Access 18.nov.15.

Cheng C, Zaichaob Z, Aizhonga D, Jiayan W, Jingfab X, Yujiao S (2011) Bar-coded pyrosequencing reveals the bacterial community during *Microcystis* water bloom in Guanting Reservoir, Beijing. Procedia Engineering 18: 341-346

Clingenpeel S, Macur R, Kan J, Inskeep WP, Lovalvo D, Varley J, Mathur E, Nealson K, Gorby Y, Jiang H, LaFracois T, McDermott TR (2011) Yellowstone Lake: high-energy geochemistry and rich bacterial diversity. Environ Microbiol 13: 2172-2185

Dowd SE, Callaway TR, Wolcott RD, Sun Y, McKeehan T, Hagevoort RG, Edrington TS (2008) Evaluation of the bacterial diversity in the feces of cattle using 16S rDNA bacterial tag-encoded FLX amplicon pyrosequencing (bTEFAP). BMC Microbiol 8: 125

Friedline CJ, Franklin RB, McCallister SL, Rivera MC (2012) Bacterial assemblages of the eastern Atlantic Ocean reveal both vertical and latitudinal biogeographic signatures. Biogeosciences 9: 2177–2193

Haugland RA, Siefring SC, Wymer LJ, Brenner KP, Dufour AP (2005) Comparison of *Enterococcus* measuments in freshwater at two recreational beaches by quantitative polymerase chain reaction and membrane filter culture analysis. Water Res 39: 559-568

Jeong JY, Park HD, Lee KH, Weon HY, Ka JO (2011) Microbial community analysis and identification of alternative host-specific fecal indicators in fecal and river water samples using pyrosequencing. The J Microbiol 49: 585-594

Jezbera J, Sharma AK, Brandt U, Doolittle WF, Hah MW (2009) 'Candidatus Planktophila limnetica', an actinobacterium representing one of the most numerically important taxa in freshwater bacterioplankton. Int J Syst Evol Microbiol 59: 2864–2869

Kildare BJ, Leutenegger, CM, McSwain BS, Bambic DG, Rajal VB, Wuertz S (2007) 16SrRNA-based assays for quantitative detection of universal, human-, cow-, and dog-specific fecal *Bacteriodales*; a Bayesian approach. Water Res 41: 3701-3715

Krause DO, Khafipour E (2011) The fecal environment, the gut. In: Sadowsky MJ, Whitman RL, (Eds.), The Fecal Bacteria. ASM Press, Washington, DC, pp.1-21.

Layton A, McKay L, Williams D, Garrett V, Gentry R, Sayler G (2006) Development of *Bacteroides* 16S rRNA gene TaqMan-based real-time PCR assays for estimation of total, human, and bovine fecal pollution in water. Appl Environ Microbiol 72: 4214-4224

Lee JE, Le S, Sung J, Ko GP (2011) Analysis of human and animal fecal microbiota for microbial source tracking. The ISME J 5: 362-365

Liou A, Paziuk M, Luevano Jr JM, Machineni S, Turnbaugh PJ, Kaplan LM (2013) Conserved shifts in the gut microbiota due to gastric bypass host weight and adiposity. Sci Transl Med 5: 178ra41

Lozupone CA, Stombaugh JI, Gordon JI, Jansson JK, Knight R (2012) Diversity, stability and resilience of the human gut microbiota. Nature 489: 221-230

Patel V, Munot H, Shouche YS, Madamwar D (2014) Response of bacterial community structure to seasonal fluctuation and anthropogenic pollution on coastal water of Alang-Sosiya ship breaking yard, Bhavnagar, India. Biores Technol 161: 362-370

Roslev P, Bukh AS (2011) State of the art molecular markers for fecal pollution source tracking in water. Appl Microbiol Biotechnol 89: 1341-1355

Segata N, Izard J, Waldron L, Gevers D, Miropolsky L, Huttenhower C (2011) Metagenomic biomarker discovery and explanation. Genome Biol 12: R60

Shannon P, Markiel A, Ozier O, et al (2003) Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res 13: 2498-2504

The Human Microbiome Project Consortium (2012) Structure, function and diversity of the healthy human microbiome. Nature 486: 207-214

Unno T, Jang J, Han D, Kim JH, Sadowsky MJ, Kim OS, Chun J, Hur HG (2011) Use of barcoded pyrosequencing and shared OTUs to determine sources of fecal bacteria in watersheds. Environ Sci Technol 44: 7777-7782

USEPA (2005) Microbial source tracking guide document. Office of Research and Development, Washington, DC EPA-600/R-05/064

Wang Q, Garrity GM, Tiedje JM, Cole JR (2007) Naïve Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. Appl Environ Microbiol 73: 5261-5267

344 Warnecke F, Sommaruga R, Sekar R, Hofer JS, Pernthaler J (2005) Abundances, identity, and growth
345 state of Actinobacteria in mountain lakes of different UV transparency. Appl Environ Microbiol 71:
346 5551-5559