Semiconductor Sequencing Reveals the Diversity of Bacterial Communities in an Amazonian Reservoir

Diego Assis das Graças (Corresponding author)

Laboratory of DNA Polymorphism, Federal University of Pará

Rua Augusto Correa 01, Belém, Pará, Brazil

Tel: 55-91-3201-8416   E-mail: diego.a87@gmail.com

Rommel Thiago Jucá Ramos

Laboratory of DNA Polymorphism, Federal University of Pará

Rua Augusto Correa 01, Belém, Pará, Brazil

Tel: 55-91-3201-8416   E-mail: rommelthiago@gmail.com

Pablo Gomes Sá

Laboratory of DNA Polymorphism, Federal University of Pará

Rua Augusto Correa 01, Belém, Pará, Brazil

Tel: 55-91-3201-8416   E-mail: pablogomesdesa@gmail.com

Rafael Azevedo Baraúna

Laboratory of DNA Polymorphism, Federal University of Pará

Rua Augusto Correa 01, Belém, Pará, Brazil

Tel: 55-91-3201-8416   E-mail: r.a.barauna@gmail.com

Rubens Ghilardi Jr.

Superintendência de Meio Ambiente. Centrais Elétricas do Norte do Brasil S/A

Brasília, Distrito Federal, Brazil.

Tel: 55-91-3201-8416   E-mail: rubens.ghilardi@eletronorte.gov.br

Maria Paula Cruz Schneider

Laboratory of DNA Polymorphism, Federal University of Pará

Rua Augusto Correa 01, Belém, Pará, Brazil

Tel: 55-91-3201-8416   E-mail: Paula@ufpa.br

Artur Silva

Laboratory of DNA Polymorphism, Federal University of Pará

Rua Augusto Correa 01, Belém, Pará, Brazil

Tel: 55-91-3201-8416   E-mail: arturluizdasilva@gmail.com

Abstract

The reservoir of the Tucuruí hydropower plant has been reported to be a source of methane by several studies. These results question the Brazilian energy policy that plans to establish many other plants similar to Tucuruí in the Amazon region. In this study, we used a next-generation sequencing platform, the Ion Torrent Personal Genome Machine, and employed a barcode library to investigate the bacterial diversity along a 70-m deep water column in the proximity of the dam. One-litre water samples were taken at each 10 m from the surface to extract total DNA. A total of 1,505,461 reads were generated. The alpha diversity analyses showed 2,733, 2,280, 1,354, 1,338, 2,726, 2,134, 2,017, 1,885 operational taxonomic units (OTUs; 3% cutoff) at depths of 0, 10, 20, 30, 40, 50, 60 and 70 m, respectively. Taxonomic inference analyses identified the most abundant phyla as being Proteobacteria, Cyanobacteria, Bacteroidetes, Acidobacteria and Actinobacteria, but others phyla were also found. This water column seems to produce oxygen in the upper layers due to abundance of Cyanobacteria. In the deeper layers, most of Proteobacterial sequences were assigned to Methanotrophic bacteria, showing that a significant amount of methane is consumed by these communities.

1. Introduction

The understanding of the role of the microbial community in the environment has been significantly increasing since molecular methods became available. These methods have allowed the characterisation of communities in diverse environments without the need for cultivation (Handelsman, 2004). In recent years, technological advancements in sequencing platforms have enabled a more thorough analysis of microbial diversity, as the traditional methods are highly expensive.

The majority of studies have used pyrosequencing to measure microbial diversity (Petrosino, Highlander, Luna, Gibbs, & Versalovic, 2009), but recently, other platforms such as Ion Torrent and Illumina MiSeq and HiSeq have been shown to be efficient alternatives to pyrosequencing for metagenomic analyses (Caporaso et al., 2012; Jünemann et al., 2012; Whiteley et al., 2012). In a recent study comparing the performance of benchtop sequencers, the use of semiconductors was highlighted by their higher yield, as up to 1 Gb can be produced in a 2 h sequencing experiment with one of the lower costs per run ($625 using chip 318) (Loman et al., 2012). The high platform yield along with the increase in the read size can contribute to the study of microbial communities at the genomic scale (R. T. J. Ramos et al., 2012; Scholz, Lo, & Chain, 2012).

These methodologies have been used to characterise the microbial community of various environments. However, few studies have been performed in the Amazon region (da C Jesus, Marsh, Tiedje, & de S Moreira, 2009; Rodrigues et al., 2013), as most of these studies still use the Sanger method (Baraúna et al., 2012; Borneman & Triplett, 1997; Graças et al., 2011, 2013; Navarrete, Cannavan, Taketani, & Tsai, 2010; Pazinato, Paulo, Mendes, Vazoller, & Tsai, 2010; Pureza et al., 2013). If the continental dimensions, large heterogeneity and high complexity of the Amazon region are taken into consideration, we assume that little is known about the diversity of its environments.

One of the Amazonian environments that has been most studied is hydropower plant reservoirs due to the emission of greenhouse gases (GHG) such as methane (CH_4) and carbon dioxide (CO_2) (Barros et al., 2011; Mendonça et al., 2012). According to Barros and colleagues (Barros et al., 2011), GHG emission in hydroelectric reservoirs is affected by the location and age of the reservoir. The Amazonian reservoirs emit high levels of GHGs, with approximately 9 Tg of carbon per year. When investigating the carbon balance in these water bodies, it is important to consider the microbial communities because bacteria and archaea have a key role in the carbon cycle in this type of environment.

In fact, some of the metagenomic analyses undertaken for the Tucuruí hydropower plant have shown a large diversity characterised by the abundance of methanogenic archaea and methanotrophic bacteria (Baraúna et al., 2012; Graças et al., 2011, 2013). According to these studies, methanogenic archaea represent 33% of the clone library of the aphotic layer, and the methanotrophic bacteria represent more than 40% of the upper layers of the water column. Other microbial groups also influence the production and consumption of methane in these bodies of water. For example, sulphate-reducing bacteria compete with the methanogenic archaea for some substrates, and therefore, knowledge of the entire microbial community is

the best way to understand the ecological dynamics that exist in the environment.

In this study, we sequenced the bacterial 16S rRNA gene by semiconduction using the Ion Torrent platform with chip 318 to analyse the microbial communities along a 70-m deep water column at a site of the Tucuruí hydropower plant reservoir.

2. Materials and Methods

2.1 Sampling Site

All samples were collected at the end of the rainy season in June 2011 during the day at site M1 (03°51'20.52"S/49°37'23.1"W), which is located very close to the dam (Fig. 1). Site M1 had a 70-m deep water column at this time of the year. This depth is close to the total storage capacity of the reservoir at this location. One litre of water was collected at every 10 m from the surface with the aid of a van Dorn bottle (Alfakit, Florianópolis, Brazil). All samples were filtered twice, first through a pre-filter (8 μm pore size) to eliminate solid particles and then using 0.22-μm-pore-size nitrocellulose membranes (Whatman/GE Healthcare, Kent, UK), and then stored in 5 mL DNA STE (sodium chloride-Tris-EDTA) preservation buffer (50 mM Tris–HCl, 500 mM NaCl, 125 mM EDTA pH 8.0) at -20°C until DNA extraction.
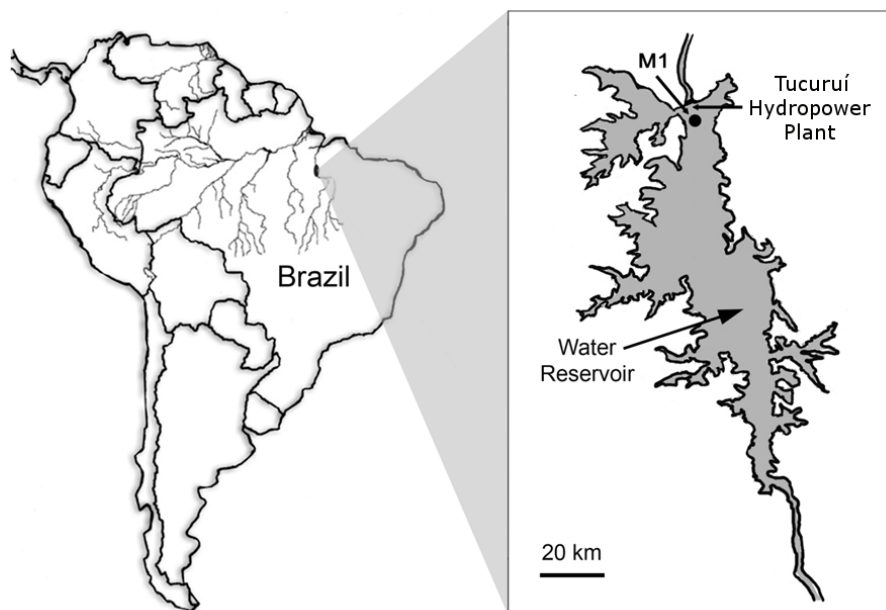

Figure 1. Geographic location of the Tucuruí hydropower plant reservoir

Note: The sampling site (M1) is shown in the insert at the right.

2.2 DNA Extraction and PCR

The initial step of the extraction protocol consisted of the incubation (37°C, 100 rpm) of samples with 10 mg/mL lysozyme overnight. In the second step, 0.5% sodium dodecyl sulphate (SDS) and 60 ng/mL proteinase K were added, and the samples were incubated in a water bath at 50°C for 1 h. Finally, phenol-chloroform extraction was performed (Sambrook

& Russel, 2001). The amplification of the bacterial 16S rRNA gene was performed with an initial PCR using the primers 16S-W8F (5'-AGA GTT TGA T(CT)(AC) TGG CTC AG-3') and 16S-1407R (5'-GAC GGG GGT G(AT)G T(AG)C AA-5'). This first amplification was done because the set of primers has a wider range of match. The PCR conditions were: initial denaturation at 95°C for 5 minutes; 30 cycles of: 95°C for 1 minute, 62°C for 40 seconds and 72°C for 30 seconds; and a final step of 72°C for 10 minutes. The 50 µl reaction mixtures contained 1X buffer, 3 mM MgCl2, 0.2 mM dNTPs, 5 pM of each primer and 2 U of Taq DNA polymerase (Invitrogen). The obtained amplicons were used as templates for a second round of amplification using the primers DHPLC-341F (5'-CCT AGG GGA GGC AGC AG-3') and DHPLC-534R (5'-ATT ACC GCG GCT GCT GG-3'). This second PCR was done to generate smaller amplicons of approximately 200-300 bp, corresponding to the V3 region of the 16S rRNA gene. The PCR conditions were: initial denaturation at 95°C for 5 minutes; 35 cycles of: 95°C for 1 minute, 62°C for 40 seconds and 72°C for 30 seconds; and a final extension step of 72°C for 10 minutes. The reaction was performed in a final volume of 50 µl containing Platinum® PCR Supermix High Fidelity, 50 ng of DNA template and 5 µM of each primer. The obtained amplicons were used for the construction of libraries.

2.3 Sequencing

The ends of the amplicons obtained in the preceding procedure were repaired and then purified with the Agencourt® AMPure® XP kit (Beckman Coulter, Brea, CA, USA). For each depth, a barcode library was constructed to identify the reads of the sampling site after sequencing. Initially, the adaptors were ligated to the amplicons with the Ion Barcode Adapter Mix (P1-IA), after incubation for 30 minutes at room temperature. After purification, the Ion Primer Mix – BC barcodes were added to each library with Platinum® PCR SuperMix High Fidelity (Life Technologies, Carlsbad, CA, USA). The ligation process was performed in a thermocycler using the following protocol: 72°C for 20 minutes; 95°C for 5 minutes; and 10 cycles of: 95°C for 15 seconds, 58°C for 15 seconds and 68°C for 1 minute. The libraries containing the adapter and barcode were purified, quantified and then mixed in equal concentrations for the emulsion PCR. After library enrichment, the samples were deposited onto chip 318, and the sequencing of the reads was performed in the Ion PGM (Life Technologies) sequencer.

2.4 Data Analysis

The data obtained from the multiplex sequencing with eight barcodes were extracted from the *.sff* file using the script sff_extract (available at http://bioinf.comav.upv.es/sff_extract). The data relative to the samples at a depth of 0 m, 10 m, 20 m, 30 m, 40 m, 50 m, 60 m and 70 m were analysed with the Quality Assessment Long Reads software (R. T. J. Ramos et al., 2012), which was used to apply a Phred 20 quality filter, discarding reads with average quality lower than this value and trimming its ends whenever possible.

2.5 Diversity Analysis

Only sequences larger than 100 bp were used for the diversity analyses. Before diversity analysis, we used Mothur (Schloss et al., 2009) to screen sequences according to alignment

criteria and to filter the alignments. All diversity analyses were performed in the Pyrosequencing Pipeline of the Ribosomal Database Project (RDP) (Cole et al., 2009), consisting of the use of Aligner, Complete Linkage Clustering (using a 3% cut-off), Rarefaction and Shannon Index and Chao1 estimator. In parallel, we also used Classifier for taxonomic inference.

3. Results

The data analyses showed a significant reduction in reads generated by the sequencer for all samples (Table 1). The mean reduction in the number of sequences after trimming and filter was about 36% in the first sequencing round and 70% in the second.

Table 1. Evaluation of data generated by the two sequencing rounds at Ion Torrent platform, before and after application of the Phred 20 quality filter and trimming

Sample	#reads	#bases	#reads (QV20, trimmed, length)	# bases (QV20, trimmed length)
SN1_11 Sequencing /home/rommel/lpdna/metagenoma/SN1-11/QV20/fasta				
70M	177,584	24,011,408	120,149 (67,7%)	20,920,524 (87,1%)
60M	181,790	24,265,995	122,447 (67,4%)	21,260,496 (87,6%)
50M	166,342	20,861,523	102,752 (61,8%)	17,587,688 (84,3%)
40M	188,422	24,428,751	122,322 (64,9%)	21,169,824 (86,7%)
30M	195,821	24,998,146	126,099 (64,4%)	21,689,443 (86,8%)
20M	171,572	21,605,819	105,917 (61,7%)	18,354,835 (85,0%)
10M	182,497	22,723,672	114,738 (62,9%)	19,639,335 (86,4%)
0M	209,204	26,407,981	135,030 (64,5%)	22,916,700 (86,8%)
SN1_20 Sequencing /home/rommel/lpdna/metagenoma/SN1-20/QV20/fasta				
70M	216,942	17,980,473	68,341 (31,5%)	9,333,882 (51,9%)
60M	221,534	18,019,783	70,569 (31,9%)	9,654,499 (53,6%)
50M	147,410	10,488,777	32,547 (22,1%)	4,309,426 (41,1%)
40M	239,614	19,194,807	73,774 (30,8%)	10,060,341 (52,4%)
30M	239,849	19,741,229	78,754 (32,8%)	10,690,328 (54,2%)
20M	227,209	17,704,701	65,479 (28,8%)	8,881,276 (50,2%)
10M	236,429	19,224,767	76,834 (32,5%)	10,647,986 (55,4%)
0M	263,464	22,111,384	89,709 (34,0%)	12,529,224 (56,7%)

Using the Classifier tool of the RDP, we showed that the majority of the diversity could not be classified in any phylum (Fig. 2). The lowest representation of unclassified sequences was

found in the 70-m layer, with 24% of sequences. However, at 0 and 10 m of depth, almost 40% of sequences could not be classified. Of these sequences, Proteobacteria represented between 30 and 70% of sequences at all depths. Another phylum that was considerably well represented was the Bacteroidetes (0.8 to 7% abundance). The phylum Cyanobacteria was also represented (ranging from 0.8 to 17%), and despite being more abundant in more superficial layers, Cyanobacteria were found up to 30 m deep. The phyla Verrucomicrobia, Acidobacteria, Actinobacteria, Firmicutes and Nitrospira were also identified, but represented less than 10% together. Therefore, others phyla underrepresented were identified as 'Others' in Figure 2.
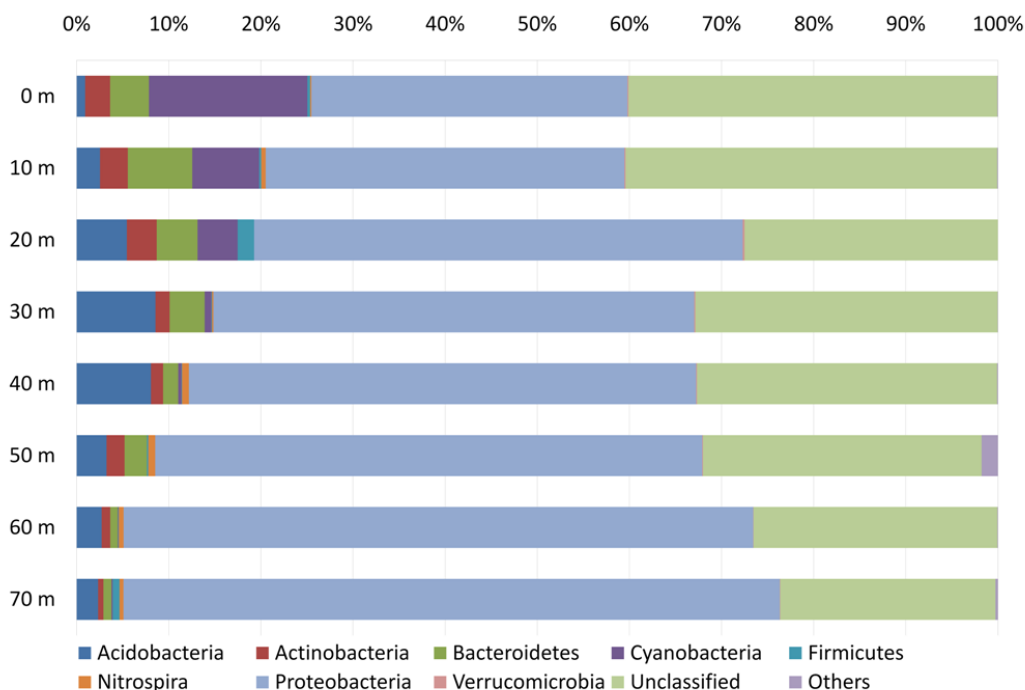

Figure 2. Histogram of bacterial taxonomic diversity found in the various layers

Notes: *Others: phyla representing less than 0.4% each of the total: Chloroflexi, Fibrobacteres, Fusobacteria, Gemmatimonadetes, OP11, Spirochaetes, SR1, TM7, WS3.

The analyses of alpha diversity (Fig. 3) showed 2,735, 2,281, 1,355, 1,339, 2,727, 2,135, 2,018, 1,888 operational taxonomic units (OTUs) at depths of 0, 10, 20, 30, 40, 50, 60 and 70 m, respectively, using a 3% cutoff. Values of 6.11, 5.97, 5.25, 5.14, 6.21, 5.95, 5.60 and 5.63 OTUs were estimated with the Chao1 index for the layers at 0, 10, 20, 30, 40, 50, 60 and 70 m, respectively (Table 2). The Shannon diversity index (H') returned values of 6.10, 5.96, 5.24, 5.13, 6.20, 5.94, 5.59 and 5.62 for the layers at 0, 10, 20, 30, 40, 50, 60 and 70 m, respectively (Table 2). Other diversity indexes were calculated and are shown in table 2.

Table 2. Diversity indexes of all layers along the water column

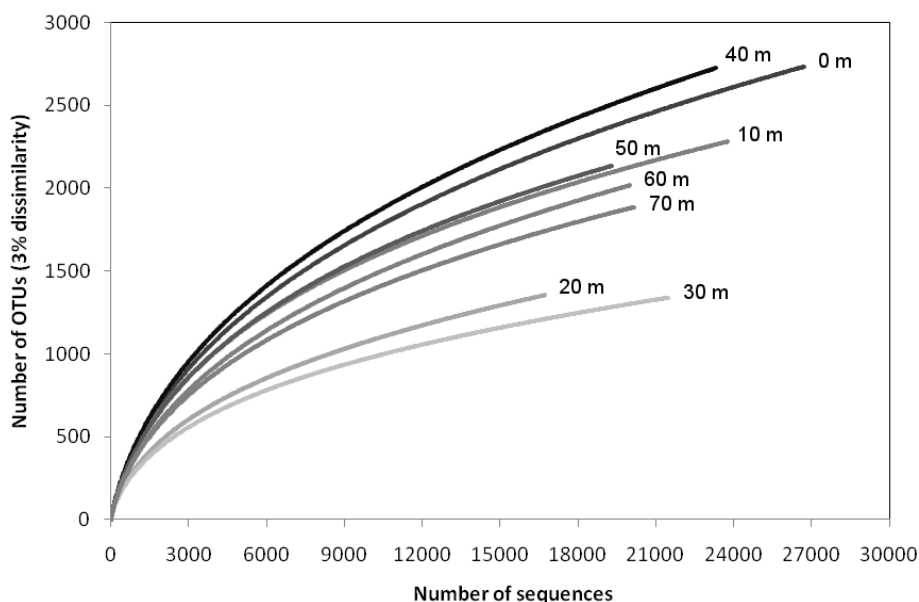| SampleID	Distance	Clusters	Chao	H'	E
0m	0.03	2735	4,437.4	6,11	0.77
10m	0.03	2281	3,570.4	5,97	0.77
20m	0.03	1355	2,211.6	5,25	0.72
30m	0.03	1339	2,188.7	5,14	0.71
40m	0.03	2727	4,528.5	6,21	0.78
50m	0.03	2135	3,233.2	5,95	0.77
60m	0.03	2018	3,401.2	5,60	0.73
70m	0.03	1886	2,998.4	5,63	0.74

Notes: H': Shanonn index, E: Evenness.

Figure 3. Rarefaction curves from all samples sequenced

Note: The 3% cutoff was used for clustering.

4. Discussion

Although the Ion Torrent platform generates larger reads (on average 120 bp from the protocol used in this work), the data quality is low when compared with high-throughput sequencers such as SOLiD and Illumina. Thus, the application of more stringent filters is necessary to avoid alignment and read evaluation errors (Li & Homer, 2010). For data from the 454 sequencer, the filter can strictly eliminate low-quality ends because the average size of reads is 700 bp. However, for the Ion Torrent PGM, trimming the sequences can lead to significant losses of information (R. T. J. Ramos et al., 2012).

The analyses of alpha diversity showed a high number of OTUs in the water column and estimate an even higher diversity using the Chao and Shannon indexes. Graças and colleagues (Graças et al., 2013) previously studied the diversity of this water column using

denaturing high-performance liquid chromatography (DHPLC) as an analysis tool in addition to Sanger sequencing. The authors showed that the bacterial and archaeal diversity exhibited a pattern in which diversity tended to increase with water column depth. The diversity of an anoxic region of the reservoir was also studied and the aphotic region displayed a higher species richness and diversity than the more superficial photic region (Graças et al., 2013). However, a different pattern was encountered in the present work, as the most diverse layer were 0 and 40 meters depth.

The few studies that have evaluated the microbial diversity of the Tucuruí hydropower plant reservoir showed that it is an environment of high diversity with a high abundance of organisms that have not yet been classified in a phylum. In the studies by Graças (Graças et al., 2011, 2013) and Baraúna (Baraúna et al., 2012), the percentage of unclassified bacteria in the samples was lower than 20%, but at this study the values were higher. The size of the reads may be an explanation for this result, as a higher number of bases to be analysed leads to a higher probability of assigning a sequence to a taxon. However, it should be kept in mind that the Classifier recommends the use of a 50% bootstrap cut-off level for sequences with sizes in the range of 50-250 bp, as this value is sufficient to classify the sequences even at the genus level (Cole et al., 2009).

Cyanobacteria were identified in high abundance in the layers at 0, 10 and 20 m deep and were mainly of the genus *Synechococcus*. The abundance and classification of *Synechococcus* was reported in the reservoir by Dall'Agnol *et al.* (Dall'agnol et al., 2012). These Cyanobacteria are found in several environments but are more abundant in freshwater and low-temperature habitats (Becker, Richl, & Ernst, 2007; Dall'agnol et al., 2012). The growth of Cyanobacteria is favoured by high luminosity, making the water column an important oxygen source in these layers. These results corroborate other analyses undertaken at the Tucuruí hydropower plant that showed that the upper layers exhibited a high number of Cyanobacteria (Baraúna et al., 2012; Graças et al., 2011, 2013). The Proteobacteria found in these upper layers were primarily assigned to the β-Proteobacteria class as well as to unclassified Proteobacteria. Among the β-Proteobacteria, only two orders were identified, Burkhoderiales and Methylophilales. The order Burkhoderiales was represented exclusively by the family Comamonadaceaei, which was composed of physiologically heterogeneous microorganisms that included autotrophs, heterotrophs and chemolithotrophs that use hydrogen (Woods & Sokol, 2006). The order Methylophilales was only represented by Methylophilaceae, which comprises methane- and methanol-oxidising microorganisms typical of freshwater environments (Lapidus et al., 2011). In the upper layers (up to 20 m), *α*- and *γ*-Proteobacteria were also abundant, but only unclassified representatives were identified. Firmicutes, Bacteroidetes and Actinobacteria were also identified in lower amounts. These results demonstrate that the upper layers of the water column are probably strong oxygen sources. This oxygen can then be used in methane and methanol oxidation by methylotrophic organisms.

The lower layers - at 30-60 m deep - also exhibited the same diversity pattern, as all contained essentially Proteobacteria (55% on average) and unclassified bacteria (30% on average). Four classes of bacteria were represented in all of the layers, with the γ- and

β-Proteobacteria being the most abundant. The majority of the β-Proteobacteria (~80%) could not be classified in any order. Nonetheless, the remaining sequences were assigned to the orders Burkholderiales, Methylophilales and Rhodocyclales. Among the γ-Proteobacteria, the order Methylococcales was predominant, with half of the sequences assigned to the genus *Methylomonas*. Microorganisms of the genus *Methylomonas* are type I methanotrophs (Hanson & Hanson, 1996) and typical of freshwater environments such as sediments, rice paddy soils and wastewaters (Bowman, 2005; Ogiso et al., 2012). The high abundance of microorganisms of the families Methylococcaceae and Methylophilaceae in freshwater environments with high use of methane has already been reported, indicating cooperation between these two families; however, the nature of this interaction remains unclear (Beck et al., 2013).

The deepest layer (70 m) showed the second largest phyla diversity, with the identification of the phyla Proteobacteria, Bacteroidetes, Firmicutes, Acidobacteria, Actinobacteria and Nitrospira, beyond another 13 phyla in lower abundance (<5%). A significant amount of the sequences could not be assigned to a phylum (25%). The most abundant phylum was the Proteobacteria, with approximately 70% of sequences. The α-Proteobacteria represented approximately 10% of the Proteobacteria, with the orders Rhizobiales and Rhodospirillales the most representative (~50% of the class). Among the Rhizobiales, the majority of sequences were assigned to the genus *Methylosinus*, which corresponds to type II methanotrophs (Hanson & Hanson, 1996). From the order Rhodospirillales, only individuals of the family Acetobacteraceae, which are characterised by the ability to oxidise ethanol to acetate in strictly aerobic conditions, were detected. Thus, we speculate that even at a depth of 70 m, there is still a significant amount of oxygen, but physicochemical analyses would be important. This contrasts with data from site C1 reported by Graças *et al.* (Graças et al., 2011). The existence of methanotrophic aerobic microorganisms at deep layers suggests that methanogenic archaea are only present in the sediment or even that the reservoir contains a large amount of dissolved methane at the deepest part of the lake, as described in various studies (Barros et al., 2011; Mendonça et al., 2012; F. M. Ramos et al., 2009).

The β-Proteobacteria represented approximately 32% of the Proteobacteria, with the orders Burkholderiales, Methylophilales and Rhodocyclales the most representative, as in the layer at 60 m. The class Burkholderiales was predominantly unable to be classified into a family, but some were assigned to the genus *Incertae sedis* found in aquatic environments and considered to be plant pathogens (Benítez & Gardener, 2009). This may be due to the large amount of vegetation on the lakebed. The bacteria of the order Rhodocyclales were assigned only to the genus *Methyloversatilis*, comprising methylotrophic microorganisms that also interact with bacteria of the order Burkholderiales and have been found in freshwater environments (Kalyuzhnaya et al., 2006).

The γ-Proteobacteria represented approximately 34% of Proteobacteria, and as in the layers at 30-60 m, the most abundant order was the Methylococcales (55% of the class). However, a significant amount of the sequences were assigned to the orders Pseudomonadales and Enterobacteriales (25% of the class). Despite the order Pseudomonadales being known for its pathogens, none of the assigned sequences exhibited high similarities with pathogenic

microorganisms. The sequences could not be assigned to any family, preventing any conclusion on the role of these orders in the environment. The Δ-Proteobacteria represented approximately 4% of Proteobacteria and belonged predominantly to the class Bdellovibrionales. None of the sequences could be assigned to a family, but they did have high similarity with order members thought to be predators (Davidov, Huchon, Koval, & Jurkevitch, 2006).

5. Conclusion

In this study, the use of the Ion Torrent Personal Genome Machine with chip 318 is shown to be a powerful alternative for the analysis of microbial communities. Semiconductor sequencing is fast, inexpensive and efficient, even if it has the disadvantage of generating shorter reads. The water-column bacterial community of this reservoir site is mainly composed of Proteobacteria, but there are a significant amount of unclassified bacteria whose role is not yet clear. Most of the bacteria that could be assigned to a taxon are involved in the degradation of organic matter along with the production and consumption of gases such as oxygen and primarily methane.

References

Baraúna, R. A., Graças, D. A., Miranda, P. R., Ghilardi, R., Barbosa, M. S., Schneider, M. P. C., & Silva, A. (2012). Prokaryotic Diversity of the Tucuruí Hydropower Plant Reservoir in the Brazilian Amazon. *Aquatic Science and Technology*, *1*(1), 181-199. http://dx.doi.org/10.5296/ast.v1i1.2981

Barros, N., Cole, J. J., Tranvik, L. J., Prairie, Y. T., Bastviken, D., Huszar, V. L. M., … Roland, F. (2011). Carbon emission from hydroelectric reservoirs linked to reservoir age and latitude. *Nature Geoscience*, *4*(9), 593-596. http://dx.doi.org/10.1038/ngeo1211

Beck, D. a. C., Kalyuzhnaya, M. G., Malfatti, S., Tringe, S. G., Glavina del Rio, T., Ivanova, N., … Chistoserdova, L. (2013). A metagenomic insight into freshwater methane-utilizing communities and evidence for cooperation between the Methylococcaceae and the Methylophilaceae. *PeerJ*, *1*, e23. http://dx.doi.org/10.7717/peerj.23

Becker, S., Richl, P., & Ernst, A. (2007). Seasonal and habitat-related distribution pattern of

Synechococcus genotypes in Lake Constance. *FEMS Microbiology Ecology*, *62*(1), 64-77. http://dx.doi.org/10.1111/j.1574-6941.2007.00366.x

Benítez, M.-S., & Gardener, B. B. M. (2009). Linking sequence to function in soil bacteria: sequence-directed isolation of novel bacteria contributing to soilborne plant disease suppression. *Applied and Environmental Microbiology*, *75*(4), 915-24. http://dx.doi.org/10.1128/AEM.01296-08

Borneman, J., & Triplett, E. W. (1997). Molecular microbial diversity in soils from eastern Amazonia: evidence for unusual microorganisms and microbial population shifts associated with deforestation. *Applied and Environmental Microbiology*, *63*(7), 2647-53. Retrieved from http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=168563&tool=pmcentrez&rendertype=abstract

Bowman, J. P. (2005). Order VII. Methylococcales ord. nov. In D. J. Brenner, R. N. Krieg, & T. J. Staley (Eds.), *Bergey's Manual of Systematic Bacteriology* (2nd ed., pp. 248-270). New York: Springer.

Caporaso, J. G., Lauber, C. L., Walters, W. a, Berg-Lyons, D., Huntley, J., Fierer, N., … Knight, R. (2012). Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms. *The ISME Journal*, *6*(8), 1621-4. http://dx.doi.org/10.1038/ismej.2012.8

Cole, J. R., Wang, Q., Cardenas, E., Fish, J., Chai, B., Farris, R. J., … Tiedje, J. M. (2009). The Ribosomal Database Project: improved alignments and new tools for rRNA analysis. *Nucleic Acids Research*, *37*(Database issue), D141-145. http://dx.doi.org/10.1093/nar/gkn879

Da C Jesus, E., Marsh, T. L., Tiedje, J. M., & de S Moreira, F. M. (2009). Changes in land use alter the structure of bacterial communities in Western Amazon soils. *The ISME Journal*, *3*(9), 1004-11. http://dx.doi.org/10.1038/ismej.2009.47

Dall'agnol, L. T., Ghilardi-Junior, R., Mcculloch, J. a., Schneider, H., Schneider, M. P. C., & Silva, a. (2012). Phylogenetic and gene trees of Synechococcus: choice of the right marker to evaluate the population diversity in the Tucurui Hydroelectric Power Station Reservoir in Brazilian Amazonia. *Journal of Plankton Research*, *34*(3), 245-257. http://dx.doi.org/10.1093/plankt/fbr109

Davidov, Y., Huchon, D., Koval, S. F., & Jurkevitch, E. (2006). A new alpha-proteobacterial clade of Bdellovibrio-like predators: implications for the mitochondrial endosymbiotic theory. *Environmental Microbiology*, *8*(12), 2179-88. http://dx.doi.org/10.1111/j.1462-2920.2006.01101.x

Graças, D. a, Jesus, E. C., Filho, L. C. F., Jr, R. G., Barbosa, M. S., Ramos, R. T. J., … Silva, A. (2013). Changes in Microbial Communities along a Water Column in an Amazonian Flooded Area. *Aquatic Science and Technology*, *1*(1), 9-29. http://dx.doi.org/10.5296/ast.v1i1.2508

Graças, D. a, Miranda, P. R., Baraúna, R. a, McCulloch, J. a, Ghilardi, R., Schneider, M. P. C.,

& Silva, A. (2011). Microbial Diversity of an Anoxic Zone of a Hydroelectric Power Station Reservoir in Brazilian Amazonia. *Microbial Ecology*, 853-861. http://dx.doi.org/10.1007/s00248-011-9906-8

Handelsman, J. (2004). Metagenomics: application of genomics to uncultured microorganisms. *Microbiol. Mol. Biol. Rev.*, *68*(4), 669-685.

Hanson, R. S., & Hanson, T. E. (1996). Methanotrophic bacteria. *Microbiological Reviews*, *60*(2), 439-71. Retrieved from http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=239451&tool=pmcentrez&rendertype=abstract

Jünemann, S., Prior, K., Szczepanowski, R., Harks, I., Ehmke, B., Goesmann, A., … Harmsen, D. (2012). Bacterial community shift in treated periodontitis patients revealed by ion torrent 16S rRNA gene amplicon sequencing. *PloS One*, *7*(8), e41606. http://dx.doi.org/10.1371/journal.pone.0041606

Kalyuzhnaya, M. G., De Marco, P., Bowerman, S., Pacheco, C. C., Lara, J. C., Lidstrom, M. E., & Chistoserdova, L. (2006). Methyloversatilis universalis gen. nov., sp. nov., a novel taxon within the Betaproteobacteria represented by three methylotrophic isolates. *International Journal of Systematic and Evolutionary Microbiology*, *56*(Pt 11), 2517-22. http://dx.doi.org/10.1099/ijs.0.64422-0

Lapidus, A., Clum, A., Labutti, K., Kaluzhnaya, M. G., Lim, S., Beck, D. a C., … Chistoserdova, L. (2011). Genomes of three methylotrophs from a single niche reveal the genetic and metabolic divergence of the methylophilaceae. *Journal of Bacteriology*, *193*(15), 3757-64. http://dx.doi.org/10.1128/JB.00404-11

Li, H., & Homer, N. (2010). A survey of sequence alignment algorithms for next-generation sequencing. *Briefings in Bioinformatics*, *11*(5), 473-83. http://dx.doi.org/10.1093/bib/bbq015

Loman, N. J., Misra, R. V, Dallman, T. J., Constantinidou, C., Gharbia, S. E., Wain, J., & Pallen, M. J. (2012). Performance comparison of benchtop high-throughput sequencing platforms. *Nature Biotechnology*, *30*(5), 434-9. http://dx.doi.org/10.1038/nbt.2198

Mendonça, R., Kosten, S., Sobek, S., Barros, N., Cole, J. J., Tranvik, L., & Roland, F. (2012). Hydroelectric carbon sequestration. *Nature Geoscience*, *5*(December), 838-840.

Navarrete, A. a., Cannavan, F. S., Taketani, R. G., & Tsai, S. M. (2010). A Molecular Survey of the Diversity of Microbial Communities in Different Amazonian Agricultural Model Systems. *Diversity*, *2*(5), 787-809. http://dx.doi.org/10.3390/d2050787

Ogiso, T., Ueno, C., Dianou, D., Huy, T. Van, Katayama, A., Kimura, M., & Asakawa, S. (2012). Methylomonas koyamae sp. nov., a type I methane-oxidizing bacterium from floodwater of a rice paddy field. *International Journal of Systematic and Evolutionary Microbiology*, *62*(Pt 8), 1832-7. http://dx.doi.org/10.1099/ijs.0.035261-0

Pazinato, J. M., Paulo, E. N., Mendes, L. W., Vazoller, R. F., & Tsai, S. M. (2010). Molecular Characterization of the Archaeal Community in an Amazonian Wetland Soil and

Culture-Dependent Isolation of Methanogenic Archaea. *Diversity*, *2*(7), 1026-1047. http://dx.doi.org/10.3390/d2071026

Petrosino, J. F., Highlander, S., Luna, R. A., Gibbs, R. a, & Versalovic, J. (2009). Metagenomic pyrosequencing and microbial identification. *Clinical Chemistry*, *55*(5), 856-66. http://dx.doi.org/10.1373/clinchem.2008.107565

Pureza, L. M., Graças, D. a, Gonçalves, E. C., Correa, J. A., Silva, A., & Schneider, M. P. (2013). Bacterial Diversity in an Amazonian Mangrove Ecosystem. *Aquatic Science and Technology*, *1*(1), 66-85. http://dx.doi.org/10.5296/ast.v1i1.2725

Ramos, F. M., Bambace, L. a. W., Lima, I. B. T., Rosa, R. R., Mazzi, E. a., & Fearnside, P. M. (2009). Methane stocks in tropical hydropower reservoirs as a potential energy source. *Climatic Change*, *93*(1-2), 1-13. http://dx.doi.org/10.1007/s10584-008-9542-6

Ramos, R. T. J., Carneiro, A. R., Soares, S. D. C., Santos, A. R. dos, Almeida, S., Guimarães, L., … Silva, A. (2012). Tips and tricks for the assembly of a Corynebacterium pseudotuberculosis genome using a semiconductor sequencer. *Microbial Biotechnology*, *6*(2), 150-156. http://dx.doi.org/10.1111/1751-7915.12006

Rodrigues, J. L. M., Pellizari, V. H., Mueller, R., Baek, K., Jesus, E. D. C., Paula, F. S., … Nüsslein, K. (2013). Conversion of the Amazon rainforest to agriculture results in biotic homogenization of soil bacterial communities. *Proceedings of the National Academy of Sciences of the United States of America*, *110*(3), 988-93. http://dx.doi.org/10.1073/pnas.1220608110

Sambrook, J., & Russel, D. W. (2001). *Molecular Cloning: A laboratory manual. Molecular Cloning: A laboratory manual* (Third., pp. 6.4-6.11). Cold Spring Harbor Laboratory Press.

Schloss, P. D., Westcott, S. L., Ryabin, T., Hall, J. R., Hartmann, M., Hollister, E. B., … Weber, C. F. (2009). Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Applied and Environmental Microbiology*, *75*(23), 7537-41. http://dx.doi.org/10.1128/AEM.01541-09

Scholz, M. B., Lo, C., & Chain, P. S. G. (2012). Next generation sequencing and bioinformatic bottlenecks: the current state of metagenomic data analysis. *Current Opinion in Biotechnology*, *23*(1), 9-15. http://dx.doi.org/10.1016/j.copbio.2011.11.013

Whiteley, A. S., Jenkins, S., Waite, I., Kresoje, N., Payne, H., Mullan, B., … O'Donnell, A. (2012). Microbial 16S rRNA Ion Tag and community metagenome sequencing using the Ion Torrent (PGM) Platform. *Journal of Microbiological Methods*, *91*(1), 80-8. http://dx.doi.org/10.1016/j.mimet.2012.07.008

Woods, D. E., & Sokol, P. E. (2006). The genus Burkholderia. In M. Doworkin, S. Falkow, E. Rosenberg, K.-H. Schleifer, & E. Stackebrandt (Eds.), *The Prokaryotes* (pp. 848-860). Springer. http://dx.doi.org/10.1007/0-387-30745-1_1

Copyright Disclaimer