

**MICROBIOME ANNOUNCEMENT**

**Open Access**

# Metagenomic and metatranscriptomic inventories of the lower Amazon River, May 2011

Brandon M. Satinsky<sup>1</sup>, Caroline S. Fortunato<sup>2</sup>, Mary Doherty<sup>2</sup>, Christa B. Smith<sup>3</sup>, Shalabh Sharma<sup>3</sup>, Nicholas D. Ward<sup>4</sup>, Alex V. Krusche<sup>5</sup>, Patricia L. Yager<sup>3</sup>, Jeffrey E. Richey<sup>4</sup>, Mary Ann Moran<sup>3\*</sup> and Byron C. Crump<sup>6\*</sup>

## Abstract

**Background:** The Amazon River runs nearly 6500 km across the South American continent before emptying into the western tropical North Atlantic Ocean. In terms of both volume and watershed area, it is the world's largest riverine system, affecting elemental cycling on a global scale.

**Results:** A quantitative inventory of genes and transcripts benchmarked with internal standards was obtained at five stations in the lower Amazon River during May 2011. At each station, metagenomes and metatranscriptomes were obtained in duplicate for two microbial size fractions (free-living, 0.2 to 2.0  $\mu\text{m}$ ; particle-associated, 2.0 to 297  $\mu\text{m}$ ) using 150  $\times$  150 paired-end Illumina sequencing. Forty eight sample datasets were obtained, averaging 15  $\times$  10<sup>6</sup> potential protein-encoding reads each (730  $\times$  10<sup>6</sup> total). Prokaryotic metagenomes and metatranscriptomes were dominated by members of the phyla *Actinobacteria*, *Planctomycetes*, *Betaproteobacteria*, *Verrucomicrobia*, *Nitrospirae*, and *Acidobacteria*. The actinobacterium SCGC AAA027-L06 reference genome recruited the greatest number of reads overall, with this single bin contributing an average of 50 billion genes and 500 million transcripts per liter of river water. Several dominant taxa were unevenly distributed between the free-living and particle-associated size fractions, such as a particle-associated bias for reads binning to planctomycete *Schlesneria paludicola* and a free-living bias for actinobacterium SCGC AAA027-L06. Gene expression ratios (transcripts to gene copy ratio) increased downstream from Óbidos to Macapá and Belém, indicating higher per cell activity of Amazon River bacteria and archaea as river water approached the ocean.

**Conclusion:** This inventory of riverine microbial genes and transcripts, benchmarked with internal standards for full quantitation, provides an unparalleled window into microbial taxa and functions in the globally important Amazon River ecosystem.

**Keywords:** Amazon River, Metagenomics, Metatranscriptomics, Internal standards, Microbial communities

## Background

The Amazon River is the world's largest riverine system, formed by a network of tributaries draining Andean and lowland basins [1]. Understanding the fate of materials transported through the Amazon River will help to better quantify its impact on global elemental cycles. As is the case for other large tropical rivers, the Amazon drains extensive floodplains and other continental areas of high primary and secondary production [2]

and contributes significantly to organic matter and nutrient export to the ocean [3, 4]. Processes occurring within the river also drive large fluxes of methane and carbon dioxide to the atmosphere [5–7].

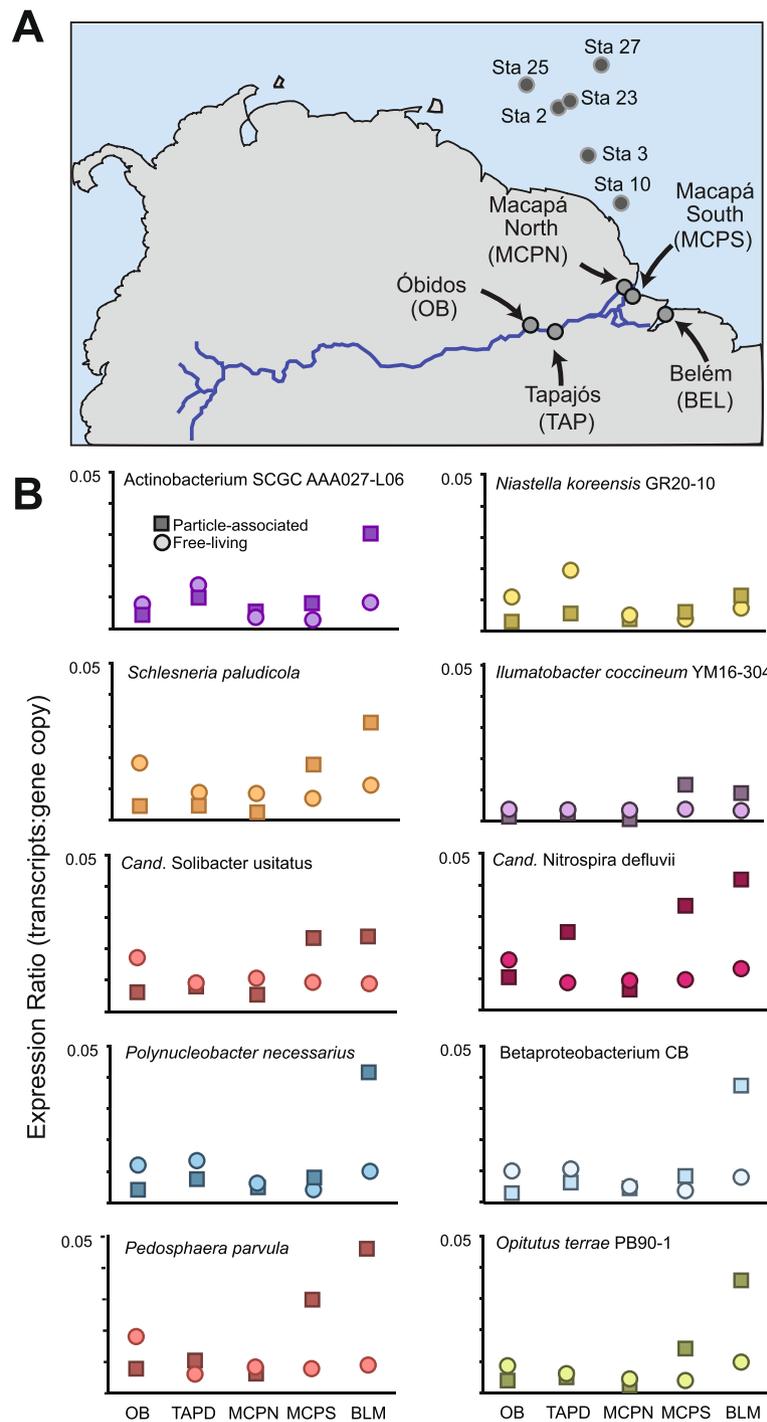
To obtain insights into the microbial processes of the lower Amazon River, we inventoried the genes and transcripts at five stations in May 2011 (Fig. 1). This study complements a previous analysis of the genes and transcripts of the Amazon Plume conducted in 2010, in which microbial expression ratios were determined at six stations across the mid- and outer plume. A previous metagenomic study in the Amazon main stem at a site several hundred kilometers upriver from our starting point at Óbidos revealed an actinobacteria-dominated

\* Correspondence: mmoran@uga.edu; bcrump@coas.oregonstate.edu

<sup>3</sup>Department of Marine Sciences, University of Georgia, Athens, GA 30605-3636, USA

<sup>6</sup>College of Earth, Ocean, and Atmospheric Science, Oregon State University, CEOAS Admin Bldg, Corvallis, OR 97331-5503, USA

Full list of author information is available at the end of the article



**Fig. 1 a** Location of sampling sites in the Amazon River in May 2011. Sampling locations for previous metagenomic and metatranscriptomic datasets in the offshore plume of the Amazon (June, 2010) are also indicated. **b** Expression ratios (transcript to gene copy ratio) for the ten reference genomes recruiting the most reads from the metagenomes. Ratios are shown by station and size fraction (squares, particle-associated; circles, free-living)

ecosystem with a high abundance of genes for heterotrophic carbon processing [8].

Here, metagenomic and metatranscriptomic sequences were obtained by Illumina sequencing, using 150 × 150 bp

overlapping paired-end reads. Whereas community genomic data have typically been analyzed within a relative framework (i.e., percent of metagenome and percent of metatranscriptome), the approach used here incorporated

known copy numbers of internal standards added at the initiation of sample processing [9]. This allowed transcripts and genes to be inventoried within an absolute framework (transcripts  $L^{-1}$ , gene copies  $L^{-1}$ , and transcript to gene copy ratios), facilitating comparisons of gene expression levels and regulatory responses among taxa and between river locations.

Measurements were made at five stations along the lower Amazon River including the historic downstream gauging station, Óbidos, the clear water Tapajós tributary, and the three primary channels near the Amazon River mouth. For each station, metagenomes and metatranscriptomes were obtained in duplicate for two discrete size fractions (free-living, 0.2 to 2.0  $\mu\text{m}$ ; particle-associated, 2.0 to 297  $\mu\text{m}$ ), resulting in 40 datasets (5 stations  $\times$  2 nucleic acid types  $\times$  2 size fractions  $\times$  2 replicates). At the Tapajós station, an additional set of filters were collected from the surface water for comparison with the sample collected at 50 % of the river depth. Following quality control (removal of poor quality reads, removal of ribosomal RNA (rRNAs) from metatranscriptomes, removal of internal standards, and joining of overlapping 150 bp paired ends), 760 million potential protein-encoding reads were obtained and analyzed for taxonomy and function.

## Methods

Detailed sample collection and processing methodology can be found in Additional file 1. The five sampling sites were located at Óbidos, the Tapajós River confluence, the north and south Macapá channels, and Belém (Fig. 1a, Additional file 2). Water samples were collected at 50 % of water column depth at each station, which ranged from 10–33 m among the stations, and microbial cells were collected by filtration and preserved in RNAlater (Applied Biosystems, Austin, TX). During sample processing, internal standards consisting of two different  $\sim$ 1000 base RNA standards [10, 11] and *T. thermophilus* HB8 genomic DNA standard [9] were added to each sample prior to cell lysis. The samples collected for metatranscriptomics were processed by extracting total RNA from the collected filters following the removal of RNAlater, treating the extracted total RNA with DNase to remove residual DNA, depleting rRNA through subtractive hybridization with community specific biotinylated nucleotide probes, linearly amplifying the remaining transcripts, and making double-stranded complementary DNA (cDNA) for library preparation and sequencing. The metagenomic samples were processed by extracting DNA and removing residual proteins and RNA. Following sample processing, cDNA or DNA was sheared and libraries were constructed for paired-end sequencing (150  $\times$  150) using the Illumina HiSeq 2500 platform.

From a total of 48 samples, we obtained  $1.27 \times 10^9$  raw sequences. Following quality control,  $0.94 \times 10^9$  reads

with a mean length of 212 bp were obtained. Internal standards were quantified and removed, along with any remaining rRNA sequences, leaving  $0.73 \times 10^9$  possible protein-encoding reads. These were annotated against the RefSeq protein database using RAPSearch2 [12], and abundance per liter was calculated based on internal standard recovery following methods in Satinsky et al. [9] (Additional file 2).

Biological and chemical data measured concurrently with sample collection included temperature, depth, conductivity, bacterial abundance, dissolved inorganic carbon concentrations, and chlorophyll concentrations (Additional file 2). Datasets describing the organic chemistry and bacterial respiration were collected at each of the lower stations [13–15].

## Quality assurance

The PANDAseq program [16] was used to join the paired-end Illumina reads using all default parameters except for the threshold score, which was set at a value stricter than the default (0.8). The FASTX-Toolkit [[http://hannonlab.cshl.edu/fastx\\_toolkit/index.html](http://hannonlab.cshl.edu/fastx_toolkit/index.html)] was used for quality control of the paired reads. Ribosomal RNA and internal standard sequences were identified in the metatranscriptomes using a Blastn search against a custom database containing representative rRNA sequences and internal standard sequences; sequences with a bit score  $\geq 50$  were identified as either rRNA or internal standards and removed from the datasets. Internal standards were identified in metagenomes by first performing a Blastn search (bit score cutoff  $\geq 50$ ) against the *T. thermophilus* HB8 genome. Hits were subsequently queried against the RefSeq protein database using Blastx (bit score cutoff  $\geq 40$ ) to identify and quantify *T. thermophilus* HB8 protein-encoding reads, and these reads were removed from the datasets prior to data analysis and deposition.

## Initial findings

Metagenomic and metatranscriptomic reads from the five Amazon River stations were assigned to bacterial, archaeal, eukaryotic, and viral taxa based on the best hits to reference genomes in the RefSeq protein database. Bacterial reads dominated the sequence collections, accounting for >90 % of quality-controlled reads in the metagenomes and >60 % in the metatranscriptomes (Table 1). Eukaryotic reads amounted to only 3 % of the genes but 9 % of the free-living and 30 % of the particle-associated transcript pools. Archaea accounted for 3 % of the genes and about 10 % of transcripts. Viral-like sequences (presumably representing ongoing viral infections; few free viral particles would be captured on the filters used for sample collection) accounted for 3 % of the community genes and 1 % of the transcripts (Table 1).

**Table 1** Library types and reads obtained in the Amazon River May 2011 microbial gene and transcript inventories, as part of the Amazon Continuum Project

Data type	Metagenomes (community DNA)		Metatranscriptomes (community mRNA)	
Stations sampled	6		6	
Size fractions sampled	2		2	
Replicates	2		2	
Samples	24		24	
Raw reads	8.85 × 10 <sup>8</sup>		1.73 × 10 <sup>9</sup>	
Joined reads post QC	3.54 × 10 <sup>8</sup>		6.17 × 10 <sup>8</sup>	
Average joined read length (bp)	218		222	
rRNA reads	–		1.52 × 10 <sup>8</sup>	
Potential protein-encoding reads	3.52 × 10 <sup>8</sup>		3.78 × 10 <sup>8</sup>	
Average abundance (genes L <sup>-1</sup> or transcripts L <sup>-1</sup> )	Free-living	Particle-associated	Free-living	Particle-associated
Bacterial	3.70 × 10 <sup>12</sup>	6.18 × 10 <sup>12</sup>	3.24 × 10 <sup>10</sup>	6.17 × 10 <sup>10</sup>
Archaeal	1.25 × 10 <sup>11</sup>	1.94 × 10 <sup>11</sup>	4.70 × 10 <sup>9</sup>	8.07 × 10 <sup>9</sup>
Eukaryotic	9.56 × 10 <sup>10</sup>	1.88 × 10 <sup>11</sup>	3.44 × 10 <sup>9</sup>	3.04 × 10 <sup>10</sup>
Viral	1.39 × 10 <sup>11</sup>	1.31 × 10 <sup>11</sup>	4.50 × 10 <sup>8</sup>	5.42 × 10 <sup>8</sup>

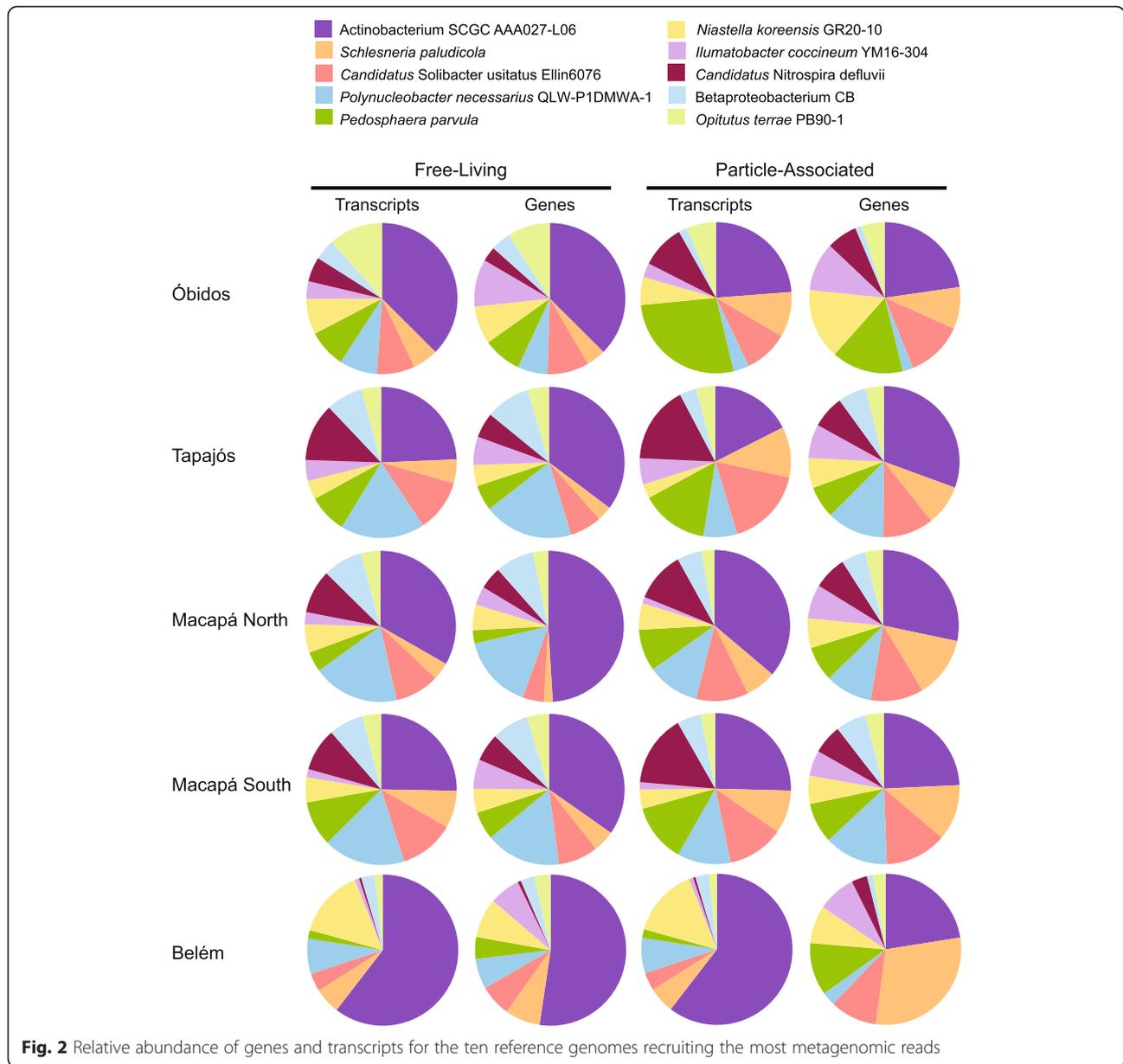
The ten reference genomes recruiting the highest number of reads were in the phyla *Actinobacteria* (actinobacterium SCGC AAA027-L06, *Ilumatobacter coccineum* YM16-304), *Planctomycetes* (*Schlesneria paludicola*), *Betaproteobacteria* (*Polynucleobacter necessarius* subsp. *asymbiolicus* QLW-P1DMWA-1, betaproteobacterium CB), *Verrucomicrobia* (*Pedosphaera parvula*, *Opitutus terrae* PB90-1), *Nitrospirae* (*Candidatus Nitrospira defluvii*), and *Acidobacteria* (*Candidatus Solibacter usitatus* Ellin6076) (Fig. 2). Reads binning to the actinobacterium SCGC AAA027-L06 genome were common in every sample and were particularly abundant at the upriver stations. At Tapajós, for example, SCGC AAA027-L06 contributed  $\sim 7 \times 10^{11}$  genes L<sup>-1</sup> and  $\sim 1 \times 10^{10}$  transcripts L<sup>-1</sup>, or 7 and 11 % of the prokaryotic metagenome and metatranscriptome at that site. The *Actinobacteria* phylum is frequently the most numerically dominant in freshwater environments, and the acI-B1 tribe that includes SCGC AAA027-L06 [17] is one of the more abundant clades [18]. Multiple thaumarchaeote and methylo-troph genomes were also among the higher-recruiting bins.

We estimated average genome equivalents per liter of river water for the ten most dominant taxa based on genes L<sup>-1</sup> in the reference genome bin (calculated from recoveries of the internal genomic standard) and the number of protein-encoding genes in each genome (Table 2). Genomes of actinobacterium SCGC AAA027-L06 and betaproteobacterium *P. necessarius* subsp. *asymbiolicus* QLW-P1DMWA-1 were the most abundant, with  $2.0 \times 10^9$  and  $3.5 \times 10^8$  genomes, respectively, in an average liter of Amazon River water. Most of the dominant taxa contributed more genomes to the particle-associated community than the free-living community. The most extreme

examples were planctomycete *S. paludicola*, which was fourfold more abundant in the larger size class, and nitrite-oxidizing *Candidatus Nitrospira defluvii*, which was threefold more abundant. This pattern is consistent with knowledge of the biology of these groups, as *Planctomycetes* have been observed on marine aggregates and other surfaces [19, 20] and are considered specialists in degrading complex organic matter, whereas members of the genus *Nitrospira* form cell clusters during active growth [21]. Looking across all dominant taxa, there was a positive correlation between genome size and extent of association with particulate material, a pattern consistent with previous studies of marine planktonic bacteria [22, 23].

We were able to assess the accuracy of gene count calculations based on the internal genomic standard by comparing genome equivalent estimates to direct cell count data obtained by epifluorescence microscopy. The average number of genome equivalents in Amazon River water was estimated to be  $3.80 \times 10^9$  genomes L<sup>-1</sup>, calculated by extrapolating from the sum of the genome equivalents of the top ten taxa ( $6.38 \times 10^8$ ; Table 2) and assuming they account for 16.8 % of the total genome equivalents (the same as their percent contribution to total genes; Table 2). In excellent agreement with this internal standard-based calculation, direct cell counts indicated an average of  $3.46 \times 10^9$  cells L<sup>-1</sup> (Additional file 2).

Calculations of the expression ratio (defined as the ratio of transcripts to gene copy) allowed comparisons of transcriptional activity among taxa, and by river location and size fraction. Of the ten most dominant prokaryotic reference genomes, all but nitrite-oxidizing *Candidatus Nitrospira defluvii* showed higher average transcripts to gene



copy ratios in the larger size class, indicative of more active cells when associated with aggregates or particulate material. Within this overall pattern, however, expression ratios were consistently higher for free-living cells in the upriver stations at Óbidos and Tapajós and at the Macapá-South station. At Macapá-North and Belém, as well as in offshore Amazon Plume stations sampled in a previous study [11], particle-associated prokaryotic cells were considerably more transcriptionally active (Fig. 1b). An additional collection was made of surface water at the Tapajós station to compare with the 50 % water depth sample, in order to assess depth-related differences in river microbial communities that could indicate water column sub-structure. The composition of the surface and 50 %

depth metagenomes and metatranscriptomes were highly similar (Additional file 2: Figure S1).

**Future directions**

The Amazon basin plays a central role in global nutrient cycling, and the rainforest surrounding the river is responsible for nearly 10 % of global primary production [24]. At its mouth, the Amazon discharges water at a rate greater than the next six largest global rivers combined. To better understand the diversity and metabolic activity of microbial communities within this extremely large river system and its oceanic plume, four high-throughput metagenomic and metatranscriptomic sequence datasets are being produced as part of the ANACONDAS and

**Table 2** Gene and transcript inventories for the ten reference genome bins recruiting the highest number of metagenome reads

Reference genome	Phylum	Mean transcripts L <sup>-1</sup>	Mean genes L <sup>-1</sup>	Mean % of transcripts	Mean % of genes	Genome size <sup>a</sup>	PA genome equivalents L <sup>-1</sup>	FL genome equivalents L <sup>-1</sup>	Total genome equivalents L <sup>-1</sup>
Actinobacterium SCGC AAA027-L06	<i>Actinobacteria</i>	$4.84 \times 10^9$	$4.97 \times 10^{11}$	11.2	7.1	1244	$1.92 \times 10^8$	$2.08 \times 10^8$	$4.00 \times 10^8$
<i>Schlesneria paludicola</i>	<i>Planctomycetes</i>	$1.30 \times 10^9$	$1.51 \times 10^{11}$	1.7	2.5	7033	$1.71 \times 10^7$	$4.41 \times 10^6$	$2.15 \times 10^7$
<i>Polynucleobacter necessarius</i> asym.	<i>Betaproteobacteria</i>	$1.55 \times 10^9$	$1.49 \times 10^{11}$	1.1	1.3	2077	$4.19 \times 10^7$	$2.73 \times 10^7$	$6.91 \times 10^7$
<i>Candidatus Solibacter usitatus</i> Ellin6076	<i>Acidobacteria</i>	$1.16 \times 10^9$	$1.44 \times 10^{11}$	1.4	0.9	7998	$1.31 \times 10^7$	$5.53 \times 10^6$	$1.86 \times 10^7$
<i>Pedosphaera parvula</i>	<i>Verrucomicrobia</i>	$2.00 \times 10^9$	$1.16 \times 10^{11}$	1.0	1.2	6510	$1.25 \times 10^7$	$5.36 \times 10^6$	$1.79 \times 10^7$
<i>Niastella korensis</i> GR20-10	<i>Bacteriodetes</i>	$1.04 \times 10^9$	$1.12 \times 10^{11}$	2.7	1.4	7136	$9.54 \times 10^6$	$6.09 \times 10^6$	$1.56 \times 10^7$
<i>Ilumatobacter coccineum</i> YM16-304	<i>Actinobacteria</i>	$4.44 \times 10^8$	$1.10 \times 10^{11}$	0.2	1.2	4233	$1.56 \times 10^7$	$1.04 \times 10^7$	$2.60 \times 10^7$
<i>Candidatus Nitrospira defluvii</i>	<i>Nitrospirae</i>	$1.44 \times 10^9$	$7.42 \times 10^{10}$	0.7	0.3	3938	$1.45 \times 10^7$	$4.34 \times 10^6$	$1.88 \times 10^7$
Betaproteobacterium CB	<i>Betaproteobacteria</i>	$5.38 \times 10^8$	$7.27 \times 10^{10}$	0.5	0.4	2034	$2.11 \times 10^7$	$1.46 \times 10^7$	$3.57 \times 10^7$
<i>Opitutus terrae</i> PB90-1	<i>Verrucomicrobia</i>	$6.94 \times 10^8$	$6.75 \times 10^{10}$	0.4	0.5	4612	$7.61 \times 10^6$	$7.04 \times 10^6$	$1.46 \times 10^7$
Sum		$1.50 \times 10^{10}$	$1.49 \times 10^{12}$	20.8	16.8				$6.38 \times 10^{8b}$

Data were averaged across five locations (Fig. 1a) sampled in the lower Amazon River in May 2011

<sup>a</sup>Number of protein-encoding genes in the reference genome

<sup>b</sup>The top ten taxa averaged 16.8 % of the genes in Amazon River water. Applying this same proportion to the number of genome equivalents contributed by the top ten taxa ( $6.38 \times 10^8$ ), the total genome equivalent is estimated at  $3.80 \times 10^9$  genomes L<sup>-1</sup>. By comparison, direct epifluorescence microscopy counts indicate  $3.46 \times 10^9$  cells L<sup>-1</sup>

ROCA projects (<http://amazoncontinuum.org>). The June 2010 plume dataset has been published [11]. Two additional datasets consisting of concurrently sampled plume and river collections in July 2012 are in progress. These high-coverage, size-discrete, and replicated datasets are all benchmarked with internal genomic and messenger RNA (mRNA) standards. Future analysis will focus on expression of biogeochemically relevant genes mediating key transformations in the carbon, nitrogen, and phosphorus cycles and the physiological and environmental factors regulating expression levels [10, 25].

### Availability of supporting data

Sequences from this May 2011 Amazon Continuum study are available from NCBI under accession numbers SRP039390 (metagenomes) and SRP037995 (metatranscriptomes). The NCBI sequences are fastq files from which internal standard sequences (metagenomes and metatranscriptomes) and rRNA sequences (metatranscriptomes only) have been removed prior to deposition. Meta-data accompanying the omics datasets are provided in Additional file 2. ANACONDAS and ROCA project data are also available at the BCO-DMO data repository (<http://www.bco-dmo.org/project/2097>).

### Additional files

**Additional file 1: Detailed methods.** Description of metagenome and metatranscriptome sample processing, sequencing, and data analysis, including internal standard additions and analyses. **Figure S1.** Hierarchical clustering of the Bray-Curtis dissimilarities in the taxonomic binning of transcripts from two microbial size fractions from 50 % water depth at each of five locations in the Amazon River and from the surface water at the Tapajós station.

**Additional file 2: Metadata.** Metadata accompanying the metagenomic and metatranscriptomic datasets, including station locations, environmental conditions, and library sizes and statistics.

### Abbreviations

bp: base pairs; cDNA: complementary DNA; mRNA: messenger RNA; rRNA: ribosomal RNA.

### Competing interests

The authors declare that they have no competing interests.

### Authors' contributions

BMS participated in the conception and design of protocols, sample collection, sample processing, data analysis, writing, and final approval of the manuscript. CSF participated in the sample collection, sample processing, critical revision, and final approval of the manuscript. MD participated in the sample processing, protocol design, critical revision, and final approval of the manuscript. CBS participated in the sample processing, protocol design, critical revision, and final approval of the manuscript. SS participated in the data analysis, critical revision, and final approval of the manuscript. NDW participated in the field campaign design, sample collection, and final approval of the manuscript. AVK participated in the field campaign design, sample collection, and final approval of the manuscript. PLY participated in the field campaign design and administration, critical revision, and final approval of the manuscript. JER participated in the field campaign design and administration, sample collection, critical revision, and final approval of the manuscript. MAM participated in the conception and

design of protocols, data analysis, writing, and final approval of the manuscript. BCC participated in the conception and design of protocols, data analysis, writing, and final approval of the manuscript. All authors read and approved the final manuscript.

### Acknowledgements

We appreciate the assistance of Roger Nilsen with library preparation and thank the scientists of the ROCA and ANACONDAS projects, Henrique O. Sawakuchi (CENA, Piracicaba), Alan Cavalcanti da Cunha (UNIFAP, Macapá), Dáimio Brito (UEAP, Macapá), Troy P. Beldini (UFOPA, Santarem), José Mauro (UFOPA, Santarem), and Rodrigo da Silva (UFOPA, Santarem) for the supporting data and insights. This research is funded by the Gordon and Betty Moore Foundation through Grants GBMF 2293 and 2928 to PL Yager and Grant GBMF 538.01 to MA Moran. Resources and technical expertise were provided by the University of Georgia's Georgia Advanced Computing Resource Center.

### Author details

<sup>1</sup>Department of Microbiology, University of Georgia, Athens, GA 30602, USA. <sup>2</sup>Horn Point, Laboratory University of Maryland Center for Environmental Science, Cambridge, MD 21612, USA. <sup>3</sup>Department of Marine Sciences, University of Georgia, Athens, GA 30605-3636, USA. <sup>4</sup>School of Oceanography, University of Washington, Seattle, WA 98112, USA. <sup>5</sup>CENA-USP, Avenida Centenário 303, 13416-000 Piracicaba, São Paulo, Brazil. <sup>6</sup>College of Earth, Ocean, and Atmospheric Science, Oregon State University, CEAS Admin Bldg, Corvallis, OR 97331-5503, USA.

Received: 26 March 2015 Accepted: 12 August 2015

Published online: 10 September 2015

### References

- Richey JE, Devol AH, Wofsy SC, Victoria R, Riberio MNG. Biogenic gases and the oxidation and reduction of carbon in amazon river and floodplain waters. *Limnol Oceanogr.* 1988;33(4):551–61.
- Nebel G, Dragsted J, Vega AS. Litter fall, biomass and net primary production in flood plain forests in the Peruvian Amazon. *Forest Ecol Manag.* 2001;150(1–2):93–102.
- Hedges JJ, Clark WA, Quay PD, Richey JE, Devol AH, Santos UD. Compositions and fluxes of particulate organic material in the Amazon river. *Limnol Oceanogr.* 1986;31(4):717–38.
- Spencer RGM, Hernes PJ, Aufdenkampe AK, Baker A, Gulliver P, Stubbins A, et al. An initial investigation into the organic matter biogeochemistry of the Congo River. *Geochim Cosmochim Acta.* 2012;84:614–27.
- Devol AH, Richey JE, Clark WA, King SL, Martinelli LA. Methane emissions to the troposphere from the Amazon floodplain. *J Geophys Res-Atmos.* 1988;93(D2):1583–92.
- Richey JE, Melack JM, Aufdenkampe AK, Ballester VM, Hess LL. Outgassing from Amazonian rivers and wetlands as a large tropical source of atmospheric CO<sub>2</sub>. *Nature.* 2002;416(6881):617–20.
- Sawakuchi HO, Bastviken D, Sawakuchi AO, Krusche AV, Ballester MVR, Richey JE. Methane emissions from Amazonian Rivers and their contribution to the global methane budget. *Global Change Biol.* 2014;20(9):2829–40.
- Ghai R, Rodriguez-Valera F, McMahon KD, Toyama D, Rinke R, Cristina Souza De Oliveira T, et al. Metagenomics of the water column in the pristine upper course of the Amazon river. *PLoS One.* 2011;6(8): e23785.
- Satinsky BM, Gifford SM, Crump BC, Moran MA. Use of internal standards for quantitative metatranscriptome and metagenome analysis. *Method Enzymol.* 2013;531:237–50.
- Satinsky BM, Crump BC, Smith CB, Sharma S, Zielinski BL, Doherty M, et al. Microspatial gene expression patterns in the Amazon River Plume. *Proc Natl Acad Sci U S A.* 2014;111(30):11085–90.
- Satinsky BM, Zielinski BL, Doherty M, Smith CB, Sharma S, Paul JH, et al. The Amazon continuum dataset: quantitative metagenomic and metatranscriptomic inventories of the Amazon River plume, June 2010. *Microbiome.* 2014;2:17.
- Zhao Y, Tang H, Ye Y. RAPSearch2: a fast and memory-efficient protein similarity search tool for next-generation sequencing data. *Bioinformatics.* 2012;28(1):125–6.
- Seidel M, Yager PL, Ward ND, Carpenter EJ, Gomes HR, Krusche AV, et al. Molecular-level changes of dissolved organic matter along the Amazon River-to-ocean continuum. *Mar Chem.* 2015; doi:10.1016/j.marchem.2015.06.019.

14. Ward ND, Keil RG, Medeiros PM, Brito DC, Cunha AC, Dittmar T, et al. Degradation of terrestrially derived macromolecules in the Amazon River. *Nat Geosci.* 2013;6(7):530–3.
15. Ward ND, Krusche AV, Sawakuchi HO, Brito DC, Cunha AC, Moura JMS, et al. The compositional evolution of dissolved and particulate organic matter along the lower Amazon River—Óbidos to the ocean. *Mar Chem.* 2015; doi:10.1016/j.marchem.2015.06.013.
16. Masella AP, Bartram AK, Truszkowski JM, Brown DG, Neufeld JD. PANDAseq: paired-end assembler for illumina sequences. *BMC Bioinformatics.* 2012;13:31.
17. Garcia SL, McMahon KD, Martinez-Garcia M, Srivastava A, Szczyrba A, Stepanauskas R, et al. Metabolic potential of a single cell belonging to one of the most abundant lineages in freshwater bacterioplankton. *ISME J.* 2013;7(1):137–47.
18. Newton RJ, Jones SE, Eiler A, McMahon KD, Bertilsson S. A guide to the natural history of freshwater lake bacteria. *Microbiol Mol Biol Rev.* 2011;75(1):14–49.
19. Delong EF, Franks DG, Alldredge AL. Phylogenetic diversity of aggregate-attached Vs free-living marine bacterial assemblages. *Limnol Oceanogr.* 1993;38(5):924–34.
20. Fuchsmann CA, Staley JT, Oakley BB, Kirkpatrick JB, Murray JW. Free-living and aggregate-associated Planctomycetes in the Black Sea. *Fems Microbiol Ecol.* 2012;80(2):402–16.
21. Spieck E, Hartwig C, McCormack I, Maixner F, Wagner M, Lipski A, et al. Selective enrichment and molecular characterization of a previously uncultured Nitrospira-like bacterium from activated sludge. *Environ Microbiol.* 2006;8(3):405–15.
22. Giovannoni SJ, Tripp HJ, Givan S, Podar M, Vergin KL, Baptista D, et al. Genome streamlining in a cosmopolitan oceanic bacterium. *Science.* 2005;309(5738):1242–5.
23. Luo H, Moran MA. Evolutionary ecology of the marine roseobacter clade. *Microbiol Mol Biol Rev.* 2014;78(4):573–87.
24. Field CB, Behrenfeld MJ, Randerson JT, Falkowski P. Primary production of the biosphere: integrating terrestrial and oceanic components. *Science.* 1998;281(5374):237–40.
25. Hilton JA, Satinsky BM, Doherty M, Zielinski B, Zehr JP. Metatranscriptomics of N<sub>2</sub>-fixing cyanobacteria in the Amazon River plume. *ISME J.* 2015;9(7):1557–69.

**Submit your next manuscript to BioMed Central and take full advantage of:**

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at  
[www.biomedcentral.com/submit](http://www.biomedcentral.com/submit)

