1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20

Title: **Microbial population genomes from the Amazon River reveal possible modulation of the organic matter degradation process in tropical freshwaters**

Running title: **Microbial population genomes from Amazon River**

Célio Dias Santos-Júnior^{a,b}, Ramiro Logares^{c,#}, Flávio Henrique-Silva^{a,#}

^a Molecular Biology Laboratory, Department of Genetics and Evolution – DGE, Universidade Federal de São Carlos – UFSCar, Rod. Washington Luis KM 235 - Monjolinho, São Carlos, SP, 13565-905, Brazil

^b Institute of Science and Technology for Brain-Inspired Intelligence – ISTBI, Fudan University, Handan Rd 220, Wu Jiao Chang, Yangpu, Shanghai, 200433, China

^c Institut de Ciències del Mar (ICM), CSIC, Departament de Biologia Marina i Oceanografia, Passeig Marítim de la Barceloneta 37-49, ES08003, Barcelona, Catalonia, Spain

[#] Corresponding authors: Flávio Henrique-Silva (dfhs@ufscar.br) and Ramiro Logares (ramiro.logares@icm.csic.es).

Abstract

Rivers connect the carbon cycle in land with that in aquatic ecosystems by transporting and transforming terrestrial organic matter (TeOM). The Amazon River receives huge loads of TeOM from the surrounding rainforest, promoting a substantial microbial heterotrophic activity and consequently, CO₂ outgassing. In the Amazon River, microbes degrade up to 55% of the lignin present in the TeOM. Yet, the main microbial genomes involved in TeOM degradation were unknown. Here, we characterize 51 Population Genomes (PGs) representing some of the most abundant microbes in the Amazon River deriving from 106 metagenomes. The 51 reconstructed PGs are among the most abundant microbes in the Amazon River, and 53% of them are not able to degrade TeOM. Among the PGs capable of degrading TeOM, 20% were exclusively cellulolytic, while the others could also oxidize lignin. The transport and consumption of lignin oxidation by-products seemed to be decoupled from the oxidation process, being apparently performed by different groups of microorganisms. By connecting the genomic features of abundant microbes in the Amazon River with the degradation machinery of TeOM, we suggest that a complex microbial consortium could explain the quick turnover of TeOM previously observed in this ecosystem.

Keywords: Amazon River, population genomes, metagenomics, terrestrial organic matter degradation

1. Introduction

Rivers connect land and ocean ecosystems, carrying about 1.9 Pg of organic carbon per year and performing carbon transformations in their course (Cole et al., 2007). Only 50% of the carbon present in rivers as terrestrial organic matter (TeOM) is delivered to the oceans (Cole et al., 2007; Mayorga et al., 2005), indicating that TeOM is actively consumed in rivers (Hernes & Benner, 2003). The rates of photosynthesis by the phytoplankton in rivers tend to increase from headwaters to estuaries, consequently indicators of heterotrophy tend to decline, e.g. as the relative organic carbon respiration rates (Battin et al., 2008).

The Amazon river basin is the largest freshwater basin in the world, comprising ~38% of continental South America (Mikhailov, 2010). In comparison to other large rivers, like the Mississippi river, where >50% of the particulate organic carbon comes from algae (Kendall, Steven, & Kelly, 2001), algal production in the Amazon river is very low (Wissmar, Richey, Stallard, & Edmond, 1981). This is explained by the turbidity of the Amazon River, which favours heterotrophic microbial activity (Doherty et al., 2017; Ghai, Rodriguez-Valera, et al., 2011; Satinsky, Crump, et al., 2014; Satinsky, Smith, Sharma, Ward, et al., 2017) rather than algal production, turning its waters supersaturated with CO₂ (Mayorga et al., 2005). The TeOM entering this riverine ecosystem comes from the Amazonian rainforest, which is responsible for ~10% of the global primary production (Field, Behrenfeld, Randerson, & Falkowski, 1998; Malhi et al., 2008). The large amounts of TeOM present in the Amazon River generate ecological niches for microorganisms specialized in complex organic matter degradation (Satinsky et al., 2015).

The main components of dissolved TeOM in the Amazon river are lignin and cellulose, accounting for ~60% of all dissolved organic matter (Ertel, Hedges, Devol, Richey, &

Ribeiro, 1986). About 60% of the lignin produced in the Amazon rainforest is channelled to the river (Ward et al., 2013) and continuously decomposed by microbes into monomers, which are subsequently reduced to low molecular weight intermediates and finally remineralized to CO₂. Lignin breakage supports 30-50% of bulk microbial respiration rates in the Amazon River (Ward et al., 2013). This leads to CO₂ outgassing from Amazon River waters (Ward et al., 2016, 2013), which releases 1.4 Tg C per year to the atmosphere (Sawakuchi et al., 2017). In total, <5% of the lignin that the forest produces is stored within the Amazon River basin or delivered to the ocean (Ward et al., 2013), suggesting a fast degradation of this compound by the river microbiota. The rapid degradation of recalcitrant organic compounds, such as lignin, boosted by the presence of labile compounds is called *priming effect* (Bianchi, 2011). Incubation experiments with sets of organic matter (Ward et al., 2016) and measurements of microbial respiration rates (Farjalla, 2014) suggest that confluence river sections are hotspots of bacterial production with CO₂ levels higher than in other regions. It indicates that the turbulence, diversity of substrates and microbes contribute to a *priming effect* in the Amazon River waters contributing to accelerate the degradation of TeOM.

Recently, the Amazon river basin non-redundant microbial gene catalogue (AMnrGC) (Santos-Júnior, Sarmento, de Miranda, Henrique-Silva, & Logares, 2020) indicated a zonation in organic matter processing associated to different river sections. The AMnrGC also revealed the main biochemical machinery used by microbes to degrade plant-derived organic matter, which consisted mainly in glycosyl-hydrolases and laccases. Based on this catalogue, we proposed that a microbial consortium could work for the Amazon River (Santos-Júnior et al., 2020), where two interacting populations (a ligninolytic and a cellulolytic one) work synergistically for the TeOM degradation. Despite the valuable insights that the AMnrGC provided, it is still necessary to link environmental genes to

the genomes from where they originate in order to acquire a more holistic understanding of TeOM degradation in the Amazon River.

The reconstruction of genomes from metagenomes has become a widely used approach during the last years (Baker, Lazar, Teske, & Dick, 2015; Cabello-Yeves, Ghai, et al., 2017; Cabello-Yeves, Haro-Moreno, et al., 2017; Cabello-Yeves et al., 2018; Dick et al., 2009; Ghai, Pašić, et al., 2011; Hugerth et al., 2015; Parks et al., 2017; Tully, Sachdeva, Graham, & Heidelberg, 2017), even though the genomes of the most abundant taxa are typically recovered. These so-called Population Genomes (PGs) or Metagenome-Assembled Genomes (MAGs) are crucial to provide genomic context to environmental genes. In addition, PGs may reveal metabolic adaptations or specific gene arrangements (Georges, El-Swais, Craig, Li, & Walsh, 2014; Muller et al., 2014), being also important to predict the ecological roles played by uncultured microorganisms. PGs have been retrieved from diverse environments (Baker et al., 2015; Cabello-Yeves et al., 2018; Dick et al., 2009; Ghai, Pašić, et al., 2011; Hugerth et al., 2015; Nayfach et al., 2020; Parks et al., 2017; Tully et al., 2017), leading to important findings, such as novel clades (Parks et al., 2017) as well as new insights on light-harvest mechanisms, nutrients uptake and nitrogen fixation in freshwater microbes (Cabello-Yeves, Ghai, et al., 2017; Cabello-Yeves, Haro-Moreno, et al., 2017; Cabello-Yeves et al., 2018). However, to our knowledge, no previous study has extracted PGs from the Amazon River in order to connect them with major biogeochemical cycles, such as the carbon cycle.

Here, we explore 51 abundant PGs extracted from 106 metagenomes retrieved from 30 Amazon River stations in order to address the following questions: What are the PGs functional repertoires to degrade TeOM? Are the systems of lignin oxidation and hemi-/cellulose degradation decoupled? Is the biochemical machinery of lignin-oxidation coupled to the one used for processing lignin-derived aromatic monomers and dimers?

2. Material and Methods

2.1 Metagenome curation and assembly

We analysed 106 metagenomes from 30 stations in the Amazon River originating from previous studies (Santos-Júnior et al., 2017; Santos-Junior, Toyama, Oliveira, Pellon de Miranda, & Henrique-Silva, 2019; Satinsky et al., 2015; Satinsky, Zielinski, et al., 2014; Toyama et al., 2016). Metagenomes included Free-living (FL; 0.2 to 2-5 μm) and Particle-Attached (PA; 2 – 5 to 300 μm) size fractions, covering 5 river sections (Fig. 1a): upstream (U), downstream (D), estuary (E), plume (P) and coastal ocean (O), as we previously defined (Santos-Júnior et al., 2020). Size fractions were not compared in depth, as it was shown that there is no significant difference in their biochemical pathways of TeOM degradation (Santos-Júnior et al., 2020). Metagenomes were downloaded from the Sequence Read Archive (SRA) [projects SRP044326, SRP039390] and from the European Nucleotide Archive (ENA) [project PRJEB25171]. Reads were checked for adapters and poor quality bases ($Q < 24$) using Cutadapt (Martin, 2011). Subsequently, metagenomes were co-assembled with MEGAHIT (v1.0) (D. Li et al., 2016) using meta-large parameters. Co-assemblies were carried out in 30 groups of metagenomes that were defined according to their geographic origin (Supplementary Table A1). See more details about the co-assembly in (Santos-Júnior et al., 2020). Only contigs with > 1 Kbp in length were considered in downstream analyses.

2.2 Binning and delineation of PGs

Quality-filtered reads were back mapped against contigs using BWA (version 0.7.12-r1039) (H. Li & Durbin, 2009) and resulting sam-bam files were processed with SamTools (version 1.3.1) (H. Li et al., 2009). Contigs from each group were then binned with Metabat (Kang, Froula, Egan, & Wang, 2015) (v2.12.1) using “superspecific” settings. Contig outliers in terms of Kmer and GC composition were eliminated using RefineM (Parks et al., 2017) (version 0.0.23) with default settings. Refined bins were assessed for completeness, contamination, strain heterogeneity and taxonomy (“lineage_wf” and “ssu_finder” modules) using CheckM (version 1.0.11) (Parks, Imelfort, Skennerton, Hugenholtz, & Tyson, 2015). The 16S rRNA-gene sequences extracted from PGs were classified by mapping them against the SILVA SSU Ref NR99 database (version 123) (Quast et al., 2013; Yilmaz et al., 2014) using Usearch (version 9.2) (Edgar, 2010), with an identity cut-off of 97% and a query coverage >70%. Contigs were removed from a bin if they displayed >98% of identity to a hit in the SILVA database that was incongruent with the taxonomic classification obtained via CheckM (Parks et al., 2015) “tree” mode, which uses functional-gene markers to assign taxonomy.

As previously suggested (Bowers et al., 2017; Parks et al., 2017), only bins with genome completeness >50%, contamination <10%, estimated quality ≥ 50 (quality = completeness – 5X contamination (Parks et al., 2017)) and consisting of ≤ 500 contigs with N50 ≥ 10 Kbp were used in downstream analyses (Supplementary Fig. A1). The Average Amino-acid Identity (AAI) between PGs was evaluated with CompareM v.0.0.13 (Parks, 2016) using the “aai_wf” function. PGs with AAI $\geq 99.5\%$ were considered redundant. Pairs of redundant genomes were assessed and the most complete, with lowest contamination and strain heterogeneity, was used in downstream analyses.

2.3 PGs similarity analysis

To check if any of the 51 recovered PGs formed conspecific strains with previously reported genomes a similarity analysis was carried out. PGs were compared, using the Average Nucleotide Identity (ANI) as implemented in FastANI (Jain, Rodriguez-R, Phillippy, Konstantinidis, & Aluru, 2017), against 957 PGs from the TARA Oceans expedition (Eren, 2017), 18 Verrucomicrobial genomes from diverse freshwater reservoirs (Cabello-Yeves, Ghai, et al., 2017), 35 PGs from lake Baikal (Cabello-Yeves et al., 2018), 2 PGs from freshwater *Synechococcus* (Cabello-Yeves, Haro-Moreno, et al., 2017), 3 087 uncultivated bacteria and archaea (UBA) genomes from the Genomes Taxonomy Database – GTDB (Parks et al., 2018) and a collection of 7 520 high-quality, complete, reference genomes from the National Centre for Biotechnology Information – NCBI (<https://www.ncbi.nlm.nih.gov/>). PGs displaying a similarity >96.5% in terms of ANI (considering an aligned fraction (AF) >60%) with another known genome were kept, and the probability of being conspecific strains (p) was calculated (Varghese et al., 2015) and reported when the difference was non-significant ($p \geq 0.9$).

2.4 Estimation of PG abundance

Quality-filtered reads from 106 metagenomes were back mapped to curated PGs using BWA (v0.7.12-r1039) (H. Li & Durbin, 2009) together with sambamba (Tarasov, Vilella, Cuppen, Nijman, & Prins, 2015) (version 0.6.6). After mapping, BAM files were filtered with a custom Perl script (courtesy of Amin Madoui - Genoscope, France), keeping only reads with identity >97% and coverage >80%. The abundance of each PG per metagenome was calculated as the number of reads recruited by it, divided by its size in kilobase pairs and the metagenome size in gigabase pairs (RPKG) (Cabello-Yeves, Haro-Moreno, et al., 2017). The RPKG abundances were then normalized by $\log(\text{RPKG}+1)$

and used to generate a Heatmap with the R packages *ggplot2* (Wickham, 2009), *gplots* (Warnes et al., 2017) and *ColorBrewer* (Neuwirth, 2014).

2.5 Gene prediction and annotation

Genes were predicted using *Prodigal* v2.6.3 (Hyatt et al., 2010) in “normal” mode. Gene annotation was done using *Prokka* v1.11 (Seemann, 2014). Predicted genes annotated with *Prokka* were also annotated with KEGG (Release 2015-10-12) (Kanehisa, Furumichi, Tanabe, Sato, & Morishima, 2017), COG (Release 2014) (Tatusov et al., 2003) and UniProtKB (Release 2016-08) (UniProt Consortium, 2015) using *Diamond* v0.9.17, *Blastp* mode (Buchfink, Xie, & Huson, 2014), with query coverage >50%, amino acid identity >45%, e-value <1e⁻⁵ and score >50. Genes were also annotated using *HMMSearch* v3.1b1 (Eddy, 2011) against *dbCAN* (v5) (Yin et al., 2012), *PFAM* (v30) (Finn et al., 2016) and *eggNOG* (v4.5) (Huerta-Cepas et al., 2016) databases, using an e-value <1e⁻⁵ and mean posterior probability of aligned residues >0.75, with no domain overlaps.

2.6 Taxonomic classification

Phylogenetic trees were inferred using 43 concatenated protein marker families (Supplementary Table A2). Proteins were identified and aligned using *HMMER* v.3.1b1 (Eddy, 2011). Positions present in <50% of taxa or without a common amino acid in ≥25% of taxa were removed. Markers were present as single-copy in ~77% of PGs. The multiple-sequence alignment (MSA) included the concatenated markers from our PGs as well as orthologous dereplicated markers from GTDB (Parks et al., 2018) and RefSeq/GenBank genomes (release 76) (Parks et al., 2017). Database markers were retrieved from the *CheckM* database (Parks et al., 2015) using *CheckM* with option

“tree_qa”. Trees were inferred with FastTree v.2.1.7 (Price, Dehal, & Arkin, 2009) using the JTT+CAT model, and bootstrap support values were calculated using 100 replicates. Newick trees were visualized using Dendroscope v.3 (Huson et al., 2007).

2.7 TeOM degradation

To investigate the TeOM degradation and supplemental metabolism in the PGs, we analysed the pathways for:

- (i) TeOM degradation – hydrolysis of hemi-/cellulose and oxidation of lignin: laccases (PF02578), and glycosyl-hydrolases (GH) annotated using dbCAN and PFAM databases.
- (ii) Degradation of lignin oxidation by-products: using the same methodology and reference genes as in previous work (Kamimura et al., 2017).
- (iii) Alternative carbon source use - Tripartite tricarboxylate transporting (TTT): using the functions *TctA* (PF01970), *TctB* (PF07331) and *TctC* (PF03401).
- (iv) Carbon storage - Polyhydroxy-butyrate and polyhydroxy-alkanoate (PHB/A) synthesis pathway: using the functions *phaA* (annotated as PF00108 and COG0183), *phaB* (annotated as PF00106 and COG1028), *phaC* (annotated as PF07167 and COG3243), *phaE* (PF09712) and *phaR* (annotated as PF05233 and COG5394).

3. Results

3.1 Population genomes (PGs) from the Amazon River

Our dataset contained 106 Amazon River metagenomes from 30 different stations divided into 5 sections: upstream, downstream, estuary, plume and ocean (Fig. 1a). We generated 30 co-assemblies (Supplementary Table A1) and after binning, 54 high-quality PGs were selected featuring ≤ 500 contigs with half of them having >10 Kbp, completeness $>50\%$, contamination $<10\%$, and quality $\geq 50\%$ (quality=completeness – 5 x contamination) [Supplementary Fig. A1]. PGs featuring $>99.5\%$ Average Amino-Acid Identity (AAI) were considered redundant and removed (Supplementary Table A3). A total of 51 non-redundant PGs were kept in our dataset, including 49% (25) high-quality and 51% (26) medium-quality PGs (Table 1 and Supplementary Table A4) according to previously used criteria (Bowers et al., 2017). PGs ranged in size between 0.5 and 7.9 Mbp, the maximum contamination was 6.8%, with $\sim 53\%$ of the PGs displaying a contamination $<1\%$ (Supplementary Table A4). In total, we recovered 25 PGs from the upstream section, 9 from downstream, 6 from the estuary, 9 from the plume and 2 from the ocean (Fig. 1a-b). *Proteobacteria* (39% of PGs, considering Alpha-, Beta- and Gamma- subdivisions as well as PGs classified as *Proteobacteria*), *Bacteroidetes* (15.7%) and non-classified bacteria (15.7%) predominated (Fig. 1b). In contrast, *Cyanobacteria* represented only 4% of the recovered PGs (Fig. 1b). Only two archaeal genomes were retrieved: one belonging to *Thaumarchaeota* (from the Upstream section) and another one from *Euryarchaeota* (from the Plume). PG distribution and abundance along the river was heterogeneous (Fig. 2).

We identified 10 PGs that had a high similarity ($>97\%$ ANI) and non-significant ($p > 0.05$) differences to other known genomes (Supplementary Table A5): *Richelia*

intracellularis-A with AM_2804, *Trueperella pyogenes* with AM_0546, *Acinetobacter junii* with AM_0608, *Methylopusillus* sp1 with AM_0507 / AM_0219, *Coccinistipes* sp. with AM_2208, *Sphingobium* sp2 with AM_1603, the taxonomically unannotated *UBA11236* with AM_1606, *Xanthomonas fuscans* with AM_0519 as well as a few characterized species from *Rokubacteria*, *GWA2-73-35* sp1 with AM_2207. Other PGs from the *TARA-Oceans* expedition (Eren, 2017) or freshwater environments (Cabello-Yeves, Ghai, et al., 2017; Cabello-Yeves, Haro-Moreno, et al., 2017; Cabello-Yeves et al., 2018) did not display high similarity to the Amazon PGs. Eighty percent of the Amazon PGs had no close genomic relative in databases or published datasets. The lowest taxonomic rank of these PGs was: Kingdom, 14% of PGs; Phylum, 14%; Class, 8%; Order, 16%; Family, 26%; and Genus 4%. The remaining 18% PGs could not be taxonomically assigned.

3.2 Cellulose and lignin oxidation

Only 24 PGs out of the recovered 51 had the potential to degrade TeOM (Fig. 3). Laccase genes were present in all taxa, except *Bacteroidetes*. All PGs having laccase genes also had glycosyl hydrolase (GH) genes, suggesting that the systems of hemi-/cellulose degradation and lignin oxidation are coupled. Furthermore, there were few cellulolytic PGs (~20%) that did not have lignin oxidation potential, pointing to two assemblages, one that besides being cellulolytic is also ligninolytic, and another one that performs only cellulose degradation. Overall, the PGs with the highest potential for TeOM degradation were AM_0519 (*Xanthomonas fuscans*), AM_0876 / AM_0936 (both unclassified bacteria), and AM_1603 (*Sphingobium* sp2), according to our criterion of having a minimum of two protein families related to TeOM degradation, with at least two different genes.

3.3 Decoupling lignin-oxidation by-products from TeOM degradation

After lignin oxidation, small aromatic compounds are formed and need to be internalized into the cell via transmembrane transporters to complete lignin degradation. Among the PGs having transporter genes for lignin oxidation by-products (Supplementary Table A6) only two of them (AM_0630 and AM_0902) were also potential lignin oxidizers.

PGs were analysed also for genes required to process aromatic compounds produced after lignin oxidation (Supplementary Table A7). Only two PGs (AM_0519 and AM_1603) may be able to both degrade lignin-derived aromatic compounds and oxidize lignin. The PGs potentially able to degrade mono-/di-aryls derived from lignin did not possess genes for cellulose degradation or lignin oxidation. Therefore, there is an apparent decoupling of functions related to the oxidation of cellulose, lignin and the processing of by-products of lignin oxidation.

3.4 Alternative carbon sources and carbon storage

For an effective carbon flux in aquatic environments, transporter systems present in microbes are crucial to ensure that alternative carbon sources can be used, such as tricarboxylates, mono- and di-aryls generated during lignin oxidation. In the Amazon River, there are two main carbon contributors: The TeOM and the less complex compounds, such as humic acids and tricarboxylates. In particular, tricarboxylates are good examples of alternative carbon sources, being constituted by molecules containing three carboxyl functional groups (-COOH), e.g. citrate. Tripartite tricarboxylate transporters (TTT) use substrate binding proteins to sequester their ligands from the extracellular milieu and to import them into the cytoplasm (Fig. 4a).

Only seven PGs may use tricarboxylates via the TTT system (Fig. 4b). The PGs containing the complete TTT system included *Alphaproteobacteria* (AM_0275) as well

as *Betaproteobacteria*, mainly from the *Burkholderiales* family. One important characteristic of the TTT system is the specificity of each substrate-binding protein to a certain substrate (Fig. 4a). This promotes a high diversity of *tctC* genes, which were found to range from tens to hundreds across PGs (Fig. 4b). In contrast, <10 genes appeared to be needed for the membrane attached parts (*tctA* and *tctB*) of this system (Fig. 4b). PGs containing a complete TTT system seem incapable of TeOM degradation, except for AM_0630, a *Burkholderiales* member containing laccase and GH8 genes. Interestingly, all PGs containing the TTT system (except AM_0630 and AM_0233) also had the biochemical machinery to process aromatic compounds derived from lignin oxidation.

One of the most important carbon storage systems is the polyhydroxy-butyrate (PHB) metabolism performed by a few enzymes (Fig. 4c). PHB biosynthesis enzymes were searched in PGs to evaluate their potential to store carbon via this polymer (Fig. 4d). Almost all PGs displaying the complete PHB pathway (*phaA-C*) (Fig. 4d) included also the TTT system, except for AM_0528 and AM_1603, which were found to be possible TeOM degraders and did not have the TTT system. Yet, the largest number of genes related to the PHB pathway was found in the TeOM degrader PG AM_1603, a *Sphingobium*. The largest gene diversity was related to the initial steps of PHB biosynthesis (genes *phaA* and *phaB*), which are not crucial for PHB production as they perform non-specific transformations, but ensure monomer availability. However, a few gene variants encoded the last steps performed by the *phaC* gene (Fig. 4d), which is the last and crucial step for PHB formation. The gene *phaR*, a transcription regulator protein also related to the accumulation of PHB, was present in 7 out of 13 PGs presumed to produce PHB (Fig. 4d). Only AM_1111 may produce other polymers different than PHB, the polyhydroxy-alkanoate/butyrate, as it contains the *phaE* gene that allows this species to produce alternative monomers (Fig. 4d).

4. Discussion

Almost half of the analysed PGs from the Amazon River seemed capable of TeOM degradation, however, as isolates are absent, we do acknowledge the fact that some of those lacking known genes related to TeOM degradation also may perform it (Fig. 3). Among the protein families involved in TeOM degradation, laccases seemed to be present as single copy genes in almost all genomes, except in *Bacteroidetes*. The diversity of these genes was much lower than among soil microorganisms (López-Mondéjar, Zühlke, Becher, Riedel, & Baldrian, 2016), indicating that even though Amazon River PGs could potentially degrade lignin and cellulose, their potential is modest when compared to soil microbes.

The number of PGs containing genes related to TeOM degradation potential decreased in downstream and estuary sections of the river (Fig. 1b, 2 and 3). These results could reflect a technical artefact, given that metagenomes had a heterogeneous representation in the different river sections (Fig. 2). Specifically, sequencing depth per metagenome decreased towards the ocean, while the number of libraries increased (Fig. 1a; Table S1). Overall, even though the analysed PGs do not represent the entire set of genes present in the community, the heterogeneous distribution of population genomes agrees with the negative correlation between TeOM degradation genes and the linear geographic distance of samples to the Amazon River source in Peru previously observed (Santos-Júnior et al., 2020). Furthermore, salinity and the diversity/abundance of copiotrophs (microorganisms with high nutritional demands) are key-players in the lacustrine TeOM transformations (Yang et al., 2020).

The fact that almost half of the genomes binned in this study lack machinery to degrade allochthonous organic matter just reinforces the possible role of rare microorganisms in

this process. Rare microbes are essential to geochemical cycles by providing necessary traits or participating in important interspecific interactions (Jousset et al., 2017). Interestingly, bio-incubation experiments demonstrated an accumulation of aromatic and oxidized compounds when the diversity decreases along permafrost thawing gradients (Zhou et al., 2020). Therefore, as previously observed (Yang et al., 2020), the TeOM transformation seems to be a synergistic process of cooperation among multiple species. Especially, the limiting step is the metabolism of monomers that may enroll in inhibition of hemi-/cellulose degradation. This is corroborated as obligate monomer specialists being relatively rare in aquatic ecosystems (Mayali & Weber, 2018).

Tricarboxylates are molecules often found in TeOM and humic environments (Ertel et al., 1986; Satinsky et al., 2015; Westergaard & Hansen, 1998) and can be generated during lignin processing (Kamimura et al., 2017). They represent alternative carbon sources and are metabolized after being transferred from the environment to the intracellular milieu. For this transfer, microbes usually recur to the TTT system, which was recently shown to be widespread in the Amazon River (Ghai, Rodriguez-Valera, et al., 2011; Santos-Júnior et al., 2020) and correlated with lignin and hemicellulose degradation (Santos-Júnior et al., 2020). Only a small fraction of our PGs (13.5%) had the TTT system. Six out of 7 PGs containing the TTT system belonged to *Betaproteobacteria*, mainly *Burkholderiales*, agreeing with another study that suggested a predominance of this transporter system in *Betaproteobacteria* (Antoine et al., 2003). The presence of tens to hundreds of gene variants of *tctC* per PG (Fig. 4) concurs with earlier findings. For example, the genome of *Bordetella pertussis* has 90 *tctC* copies (Antoine et al., 2003). Given that each *tctC* gene has a high affinity for its substrate (Antoine et al., 2005, 2003; Winnen, Hvorup, & Saier, 2003), our results also point to multiple substrates in the Amazon River. In the analysed PGs, the decoupling between

the TTT system and the TeOM degradation apparatus suggests that organisms incapable of degrading TeOM may specialize in degrading tricarboxylates. This could reflect a general differentiation in carbon use among Amazon River microbes.

Bacteria have developed impressive mechanisms to cope with adversity. Fluctuations in the water level, change in the concentration of nutrients and seasonality, are common disturbances in the Amazon river. The production and intracellular accumulation of nutritive polymers, later used to prevent starvation during unfavourable conditions, represent an important trait in multiple microbes. Specific mechanisms, such as carbon storage, are relevant also to understand the flux of carbon in ecosystems. The polyhydroxy-butyrate (PHB) metabolism (Fig. 4b) can be used by microbes to store excess of carbon and avoid starvation when conditions are unfavourable. We found that the complete pathway (*phaA-C*) tended to be present in organisms with the TTT system, suggesting a coupling between these systems. Most of those PGs were identified as *Betaproteobacteria*, suggesting that this group is also important in carbon storage. The gene redundancy found in the PHB biosynthesis was much lower than that in the TTT system, and pathway limiting reactions were performed by enzymes encoded by single genes or few gene variants. This points to a potential pathway disruption in case of gene loss, and also a direction of resources inside the cell. The transcriptional repressor gene *phaR*, which coordinates the accumulation of PHB (Quelas, Mesa, Mongiardini, Jendrossek, & Lodeiro, 2016), was present in more than half of the Amazon River PGs featuring the PHB pathway (Fig. 4d). This points to a microbial assemblage specialized in accumulating PHB that could represent a sink in the carbon cycle that needs to be further explored.

The *priming effect* in the Amazon River was largely documented with strong experimental evidence, correlating the main steps of TeOM processing to different taxa

(Santos-Júnior et al., 2020; Sawakuchi et al., 2017; Ward et al., 2016). This work shows two communities possibly involved in TeOM degradation, one responsible for hemi-/cellulose degradation and another one responsible for lignin degradation. Our data suggest that the biochemical machinery to perform cellulose hydrolysis and lignin oxidation are usually coupled, while both being decoupled in terms of lignin-derived aromatic compounds consumption.

We propose that different microbial assemblages act in synchrony to degrade TeOM (Fig. 5). Given that the biochemical machinery for metabolizing lignin-oxidation by-products is decoupled from the one performing TeOM degradation, TeOM-degrading assemblages may not be able to consume these by-products, leading to their accumulation in the environment. This accumulation can be toxic, and sterically prevent cellulase reactions (Qin et al., 2016), inhibiting the TeOM-degrading consortium and decreasing its growth (Fig. 5). The finding that some PGs are able to transport and metabolize lignin oxidation by-products, being putatively not able to oxidize lignin or hydrolyse cellulose, suggests that there is another microbial assemblage using alternative carbon sources, such as tricarboxylates, which may eliminate the above-mentioned inhibition to the hemi-/cellulose hydrolysis (Fig. 5). The microbial consortium using alternative carbon sources could be characterized by transporters specialized in tricarboxylates (TTT system) as well as genes related to lignin-derived aromatic compounds degradation. Interestingly, this assemblage also may be capable of intracellular carbon storage via PHB biosynthesis, representing a reversible sink in the carbon cycle.

The taxonomic composition of PGs (Fig. 1b and 2) was coherent with that observed by other authors in the Amazon River using 16S rRNA amplicons (Reis et al., 2019; Toyama et al., 2017) and reads binning (Doherty et al., 2017; Satinsky, Smith, Sharma, Ward, et al., 2017). The Amazon River's PGs were dominated by *Actinobacteria* and

Proteobacteria (39% of PGs) agreeing with other studies (Ghai, Rodriguez-Valera, et al., 2011; D Toyama et al., 2017). Salinity changes appeared as the main factor influencing PGs distributions (Fig. 2), in agreement with previous studies indicating the importance of salinity in structuring the Amazon River microbiota (Satinsky, Crump, et al., 2014; Satinsky, Zielinski, et al., 2014). The upstream river section displayed the most taxonomically diverse PGs, when compared with other sections. This may be related to a deeper sequencing depth per metagenome in the upstream section, although other metagenomes from the downstream section included more libraries per section, compensating the total sequencing coverage. Yet, a shallower sequencing depth across multiple libraries could be translated into recovering more information for abundant taxa, and less information for less abundant counterparts. In turn, concentrating all the sequencing effort in fewer samples could be translated into recovering more information for less abundant taxa, that eventually could lead to high quality PGs. Given that this work uses a heterogeneous metagenome dataset, we cannot control for the different amounts of information recovered for less abundant taxa.

Even though *Synechococcus* was previously reported as an Amazon freshwater dominant phototrophic genus (Reis et al., 2019), we did not find it among our high quality PGs. One possibility is that *Synechococcus* did not assemble or did not constitute high quality bins due to its microdiversity (Haverkamp et al., 2009; Mackey et al., 2017). Instead, we recovered high quality PGs from *Richelia* and *Anabaena*. *Richelia* was previously identified by reads binning (Hilton, Satinsky, Doherty, Zielinski, & Zehr, 2015), and although it was reported to occupy preferentially the plume section (Del Vecchio & Subramaniam, 2004), we found it to be abundant in the estuary and ocean sections. Furthermore, the *Anabaena*-related PG AM_0902 was found to be more abundant in downstream and plume sections. This is expected, as photosynthesis

increases as the river approaches the ocean, mainly due to a decrease in particulate matter (Hilton et al., 2015; Satinsky et al., 2015; Satinsky, Smith, Sharma, Landa, et al., 2017; Satinsky, Smith, Sharma, Ward, et al., 2017).

Here, we binned population genomes of prokaryotes from the Amazon River. Our results showed that half of the sampled genomes may be able to perform TeOM degradation, pointing to a widespread trait. The lower gene diversity related to cellulose processing in PGs recovered from plume and ocean zones points to a potential compartmentalization of TeOM degradation steps in the river sections. Furthermore, there is a decoupling between the TTT system and the TeOM degradation apparatus, suggesting a general differentiation in the carbon use among Amazon River microbes. Microbes able to use tricarboxylates are also specialized in accumulating PHB, representing a reversible sink in the carbon cycle. Altogether, our results expand the comprehension of microbial heterotrophy in tropical waters by adding a layer of complexity related to the microbial consortium working in the degradation of TeOM after eliminating the feedback inhibition caused by its by-products.

6. Acknowledgements

Bioinformatics analyses were performed at the MARBITS platform of the Institut de Ciències del Mar (ICM; <http://marbits.icm.csic.es>) as well as in MareNostrum (Barcelona Supercomputing Center) via grants obtained from the Spanish Network of Supercomputing (RES) to RL. We thank Pablo Sánchez (ICM) for his assistance with bioinformatics analyses and support. We also thank the EMM group (<https://emm.icm.csic.es>) at the ICM-CSIC for all the support and cordiality during the development of this work, specially Lidia Montiel.

7. References

- Antoine, R., Huvent, I., Chemlal, K., Deray, I., Raze, D., Loch, C., & Jacob-Dubuisson, F. (2005). The Periplasmic Binding Protein of a Tripartite Tricarboxylate Transporter is Involved in Signal Transduction. *Journal of Molecular Biology*, 351(4), 799–809. doi: 10.1016/j.jmb.2005.05.071
- Antoine, R., Jacob-Dubuisson, F., Drobecq, H., Willery, E., Lesjean, S., & Loch, C. (2003). Overrepresentation of a gene family encoding extracytoplasmic solute receptors in *Bordetella*. *Journal of Bacteriology*, 185(4), 1470–1474. doi: 10.1128/JB.185.4.1470-1474.2003
- Baker, B. J., Lazar, C. S., Teske, A. P., & Dick, G. J. (2015). Genomic resolution of linkages in carbon, nitrogen, and sulfur cycling among widespread estuary sediment bacteria. *Microbiome*, 3(1), 14. doi: 10.1186/s40168-015-0077-6
- Battin, T. J., Kaplan, L. A., Findlay, S., Hopkinson, C. S., Marti, E., Packman, A. I., ... Sabater, F. (2008). Biophysical controls on organic carbon fluxes in fluvial networks. *Nature Geoscience*, 1(2), 95–100. doi: 10.1038/ngeo101
- Bianchi, T. S. (2011). The role of terrestrially derived organic carbon in the coastal ocean: A changing paradigm and the priming effect. *Proceedings of the National Academy of Sciences of the United States of America*, 108(49), 19473–19481. doi: 10.1073/pnas.1017982108
- Bowers, R. M., Kyrpides, N. C., Stepanauskas, R., Harmon-Smith, M., Doud, D., Reddy, T. B. K., ... Woyke, T. (2017). Minimum information about a single amplified genome (MISAG) and a metagenome-assembled genome (MIMAG) of bacteria and archaea. *Nature Biotechnology*, 35(8), 725–731. doi: 10.1038/nbt.3893

- 514 Buchfink, B., Xie, C., & Huson, D. H. (2014). Fast and sensitive protein alignment
515 using DIAMOND. *Nature Methods*, 12(1), 59–60. doi: 10.1038/nmeth.3176
- 516 Cabello-Yeves, P. J., Ghai, R., Mehrshad, M., Picazo, A., Camacho, A., & Rodriguez-
517 Valera, F. (2017). Reconstruction of Diverse Verrucomicrobial Genomes from
518 Metagenome Datasets of Freshwater Reservoirs. *Frontiers in Microbiology*, 8,
519 2131. doi: 10.3389/fmicb.2017.02131
- 520 Cabello-Yeves, P. J., Haro-Moreno, J. M., Martin-Cuadrado, A.-B., Ghai, R., Picazo,
521 A., Camacho, A., & Rodriguez-Valera, F. (2017). Novel Synechococcus
522 Genomes Reconstructed from Freshwater Reservoirs. *Frontiers in Microbiology*,
523 8, 1151. doi: 10.3389/fmicb.2017.01151
- 524 Cabello-Yeves, P. J., Zemskaya, T. I., Rosselli, R., Coutinho, F. H., Zakharenko, A. S.,
525 Blinov, V. V., & Rodriguez-Valera, F. (2018). Genomes of Novel Microbial
526 Lineages Assembled from the Sub-Ice Waters of Lake Baikal. *Applied and*
527 *Environmental Microbiology*, 84(1), e02132-17. doi: 10.1128/AEM.02132-17
- 528 Cole, J. J., Prairie, Y. T., Caraco, N. F., McDowell, W. H., Tranvik, L. J., Striegl, R.
529 G., ... Melack, J. (2007). Plumbing the Global Carbon Cycle: Integrating Inland
530 Waters into the Terrestrial Carbon Budget. *Ecosystems*, 10(1), 172–185. doi:
531 10.1007/s10021-006-9013-8
- 532 Del Vecchio, R., & Subramaniam, A. (2004). Influence of the Amazon River on the
533 surface optical properties of the western tropical North Atlantic Ocean. *Journal*
534 *of Geophysical Research*, 109(C11), C11001. doi: 10.1029/2004JC002503
- 535 Dick, G. J., Andersson, A. F., Baker, B. J., Simmons, S. L., Thomas, B. C., Yelton, A.
536 P., & Banfield, J. F. (2009). Community-wide analysis of microbial genome
537 sequence signatures. *Genome Biology*, 10(8), R85. doi: 10.1186/gb-2009-10-8-
538 r85

- 539 Doherty, M., Yager, P. L., Moran, M. A., Coles, V. J., Fortunato, C. S., Krusche, A.
540 V., ... Crump, B. C. (2017). Bacterial Biogeography across the Amazon River-
541 Ocean Continuum. *Frontiers in Microbiology*, 8, 882. doi:
542 10.3389/fmicb.2017.00882
- 543 Eddy, S. R. (2011). Accelerated Profile HMM Searches. *PLoS Computational Biology*,
544 7(10), e1002195. doi: 10.1371/journal.pcbi.1002195
- 545 Edgar, R. C. (2010). Search and clustering orders of magnitude faster than BLAST.
546 *Bioinformatics*, 26(19), 2460–2461. doi: 10.1093/bioinformatics/btq461
- 547 Eren, A. M. (2017). TARA-NON-REDUNDANT-MAGs. *Figshare*. doi:
548 <https://doi.org/10.6084/m9.figshare.4902923.v1>
- 549 Ertel, J. R., Hedges, J. I., Devol, A. H., Richey, J. E., & Ribeiro, M. de N. G. (1986).
550 Dissolved humic substances of the Amazon River system1. *Limnology and*
551 *Oceanography*, 31(4), 739–754. doi: 10.4319/lo.1986.31.4.0739
- 552 Farjalla, V. F. (2014). Are the mixing zones between aquatic ecosystems hot spots of
553 bacterial production in the Amazon River system? *Hydrobiologia*, 728(1), 153–
554 165. doi: 10.1007/s10750-014-1814-8
- 555 Field, Behrenfeld, Randerson, & Falkowski. (1998). Primary production of the
556 biosphere: Integrating terrestrial and oceanic components. *Science (New York,*
557 *N.Y.)*, 281(5374), 237–240.
- 558 Finn, R. D., Coghill, P., Eberhardt, R. Y., Eddy, S. R., Mistry, J., Mitchell, A. L., ...
559 Bateman, A. (2016). The Pfam protein families database: Towards a more
560 sustainable future. *Nucleic Acids Research*, 44(D1), D279-85. doi:
561 10.1093/nar/gkv1344
- 562 Georges, A. A., El-Swaiss, H., Craig, S. E., Li, W. K., & Walsh, D. A. (2014).
563 Metaproteomic analysis of a winter to spring succession in coastal northwest

- 564 Atlantic Ocean microbial plankton. *The ISME Journal*, 8(6), 1301–1313. doi:
565 10.1038/ismej.2013.234
- 566 Ghai, R., Pašić, L., Fernández, A. B., Martin-Cuadrado, A.-B., Mizuno, C. M.,
567 McMahon, K. D., ... Rodríguez-Valera, F. (2011). New Abundant Microbial
568 Groups in Aquatic Hypersaline Environments. *Scientific Reports*, 1(1), 135. doi:
569 10.1038/srep00135
- 570 Ghai, R., Rodríguez-Valera, F., McMahon, K. D., Toyama, D., Rinke, R., de Oliveira,
571 T. C. S., ... Henrique-Silva, F. (2011). Metagenomics of the water column in the
572 pristine upper course of the Amazon river. *PLoS ONE*, 6(8), e23785. doi:
573 10.1371/journal.pone.0023785
- 574 Haverkamp, T. H. A., Schouten, D., Doeleman, M., Wollenzien, U., Huisman, J., &
575 Stal, L. J. (2009). Colorful microdiversity of *Synechococcus* strains
576 (picocyanobacteria) isolated from the Baltic Sea. *The ISME Journal*, 3(4), 397–
577 408. doi: 10.1038/ismej.2008.118
- 578 Hernes, P. J., & Benner, R. (2003). Photochemical and microbial degradation of
579 dissolved lignin phenols: Implications for the fate of terrigenous dissolved
580 organic matter in marine environments. *Journal of Geophysical Research:*
581 *Oceans*, 108, 3291.
- 582 Hilton, J. A., Satinsky, B. M., Doherty, M., Zielinski, B., & Zehr, J. P. (2015).
583 Metatranscriptomics of N₂-fixing cyanobacteria in the Amazon River plume.
584 *The ISME Journal*, 9(7), 1557–1569. doi: 10.1038/ismej.2014.240
- 585 Huerta-Cepas, J., Szklarczyk, D., Forslund, K., Cook, H., Heller, D., Walter, M. C., ...
586 Bork, P. (2016). eggNOG 4.5: A hierarchical orthology framework with
587 improved functional annotations for eukaryotic, prokaryotic and viral sequences.
588 *Nucleic Acids Research*, 44(D1), D286–D293. doi: 10.1093/nar/gkv1248

- 589 Hugerth, L. W., Larsson, J., Alneberg, J., Lindh, M. V., Legrand, C., Pinhassi, J., &
590 Andersson, A. F. (2015). Metagenome-assembled genomes uncover a global
591 brackish microbiome. *Genome Biology*, 16(1), 279. doi: 10.1186/s13059-015-
592 0834-7
- 593 Huson, D. H., Richter, D. C., Rausch, C., DeZulian, T., Franz, M., & Rupp, R. (2007).
594 Dendroscope: An interactive viewer for large phylogenetic trees. *BMC*
595 *Bioinformatics*, 8, 460. doi: 10.1186/1471-2105-8-460
- 596 Hyatt, D., Chen, G.-L., LoCascio, P. F., Land, M. L., Larimer, F. W., & Hauser, L. J.
597 (2010). Prodigal: Prokaryotic gene recognition and translation initiation site
598 identification. *BMC Bioinformatics*, 11(1), 119. doi: 10.1186/1471-2105-11-119
- 599 Jain, C., Rodriguez-R, L. M., Phillippy, A. M., Konstantinidis, K. T., & Aluru, S.
600 (2017). High-throughput ANI Analysis of 90K Prokaryotic Genomes Reveals
601 Clear Species Boundaries. *BioRxiv*, 225342. doi: 10.1101/225342
- 602 Jousset, A., Bienhold, C., Chatzinotas, A., Gallien, L., Gobet, A., Kurm, V., ... Hol, W.
603 H. G. (2017). Where less may be more: How the rare biosphere pulls ecosystems
604 strings. *The ISME Journal*, 11(4), 853–862. doi: 10.1038/ismej.2016.174
- 605 Kamimura, N., Takahashi, K., Mori, K., Araki, T., Fujita, M., Higuchi, Y., & Masai, E.
606 (2017). Bacterial catabolism of lignin-derived aromatics: New findings in a
607 recent decade: Update on bacterial lignin catabolism. *Environmental*
608 *Microbiology Reports*, 9(6), 679–705. doi: 10.1111/1758-2229.12597
- 609 Kanehisa, M., Furumichi, M., Tanabe, M., Sato, Y., & Morishima, K. (2017). KEGG:
610 new perspectives on genomes, pathways, diseases and drugs. *Nucleic Acids*
611 *Research*, 45(D1), D353–D361. doi: 10.1093/nar/gkw1092

- 612 Kang, D. D., Froula, J., Egan, R., & Wang, Z. (2015). MetaBAT, an efficient tool for
613 accurately reconstructing single genomes from complex microbial communities.
614 *PeerJ*, 3. doi: 10.7717/peerj.1165
- 615 Kendall, C., Steven, R. S., & Kelly, V. J. (2001). Carbon and nitrogen isotopic
616 compositions of particulate organic matter in four large river systems across the
617 United States. *Hydrological Processes*, 15, 1301–1346.
- 618 Li, D., Luo, R., Liu, C. M., Leung, C. M., Ting, H. F., Sadakane, K., ... Lam, T. W.
619 (2016, June 1). MEGAHIT v1.0: A fast and scalable metagenome assembler
620 driven by advanced methodologies and community practices. *Methods*, Vol.
621 102, pp. 3–11. doi: 10.1016/j.ymeth.2016.02.020
- 622 Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with Burrows-
623 Wheeler transform. *Bioinformatics*, 25(14), 1754–1760. doi:
624 10.1093/bioinformatics/btp324
- 625 Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., ... 1000 Genome
626 Project Data Processing Subgroup. (2009). The Sequence Alignment/Map
627 format and SAMtools. *Bioinformatics*, 25(16), 2078–2079. doi:
628 10.1093/bioinformatics/btp352
- 629 López-Mondéjar, R., Zühlke, D., Becher, D., Riedel, K., & Baldrian, P. (2016).
630 Cellulose and hemicellulose decomposition by forest soil bacteria proceeds by
631 the action of structurally variable enzymatic systems. *Scientific Reports*, 6,
632 25279. doi: 10.1038/srep25279
- 633 Mackey, K. R. M., Hunter-Cevera, K., Britten, G. L., Murphy, L. G., Sogin, M. L., &
634 Huber, J. A. (2017). Seasonal Succession and Spatial Patterns of *Synechococcus*
635 Microdiversity in a Salt Marsh Estuary Revealed through 16S rRNA Gene

- 636 Oligotyping. *Frontiers in Microbiology*, 8, 1496. doi:
637 10.3389/fmicb.2017.01496
- 638 Malhi, Y., Roberts, J. T., Betts, R. A., Killeen, T. J., Li, W., & Nobre, C. A. (2008).
639 Climate change, deforestation, and the fate of the Amazon. *Science*, 319(5860),
640 169–172. doi: 10.1126/science.1146961
- 641 Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput
642 sequencing reads. *EMBnet.Journal*, 17(1), 10. doi: 10.14806/ej.17.1.200
- 643 Mayali, X., & Weber, P. K. (2018). Quantitative isotope incorporation reveals substrate
644 partitioning in a coastal microbial community. *FEMS Microbiology Ecology*,
645 94(5). doi: 10.1093/femsec/fiy047
- 646 Mayorga, E., Aufdenkampe, A. K., Masiello, C. A., Krusche, A. V., Hedges, J. I., Quay,
647 P. D., ... Brown, T. A. (2005). Young organic matter as a source of carbon
648 dioxide outgassing from Amazonian rivers. *Nature*, 436, 538.
- 649 Mikhailov, V. N. (2010). Water and sediment runoff at the Amazon River mouth. *Water*
650 *Resources*, 37(2), 145–159. doi: 10.1134/S009780781002003X
- 651 Muller, E. E. L., Pinel, N., Laczny, C. C., Hoopmann, M. R., Narayanasamy, S.,
652 Lebrun, L. A., ... Wilmes, P. (2014). Community-integrated omics links
653 dominance of a microbial generalist to fine-tuned resource usage. *Nature*
654 *Communications*, 5, 5603. doi: 10.1038/ncomms6603
- 655 Nayfach, S., Roux, S., Seshadri, R., Udworthy, D., Varghese, N., Schulz, F., ... Elloe-
656 Fadrosch, E. A. (2020). A genomic catalog of Earth's microbiomes. *Nature*
657 *Biotechnology*, 1–11. doi: 10.1038/s41587-020-0718-6
- 658 Neuwirth, E. (2014). CRAN - Package ColorBrewer Palettes. Retrieved November 7,
659 2017, from <https://cran.r-project.org/web/packages/RColorBrewer/index.html>

- 660 Parks, D. H. (2016). CompareM. Retrieved from GitHub website:
661 https://github.com/dparks1134/CompareM/blob/master/users_guide.pdf
- 662 Parks, D. H., Chuvochina, M., Waite, D. W., Rinke, C., Skarshewski, A., Chaumeil, P.-
663 A., & Hugenholtz, P. (2018). A proposal for a standardized bacterial taxonomy
664 based on genome phylogeny. *BioRxiv*, 256800. doi: 10.1101/256800
- 665 Parks, D. H., Imelfort, M., Skennerton, C. T., Hugenholtz, P., & Tyson, G. W. (2015).
666 CheckM: assessing the quality of microbial genomes recovered from isolates,
667 single cells, and metagenomes. *Genome Research*, 25(7), 1043–1055. doi:
668 10.1101/gr.186072.114
- 669 Parks, D. H., Rinke, C., Chuvochina, M., Chaumeil, P.-A., Woodcroft, B. J., Evans, P.
670 N., ... Tyson, G. W. (2017). Recovery of nearly 8,000 metagenome-assembled
671 genomes substantially expands the tree of life. *Nature Microbiology*, 2(11),
672 1533–1542. doi: 10.1038/s41564-017-0012-7
- 673 Price, M. N., Dehal, P. S., & Arkin, A. P. (2009). FastTree: Computing Large Minimum
674 Evolution Trees with Profiles instead of a Distance Matrix. *Molecular Biology
675 and Evolution*, 26(7), 1641–1650. doi: 10.1093/molbev/msp077
- 676 Qin, L., Li, W.-C., Liu, L., Zhu, J.-Q., Li, X., Li, B.-Z., & Yuan, Y.-J. (2016). Inhibition
677 of lignin-derived phenolic compounds to cellulase. *Biotechnology for Biofuels*,
678 9(1), 70. doi: 10.1186/s13068-016-0485-2
- 679 Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., ... Glöckner, F.
680 O. (2013). The SILVA ribosomal RNA gene database project: Improved data
681 processing and web-based tools. *Nucleic Acids Research*, 41(Database issue),
682 D590-6. doi: 10.1093/nar/gks1219
- 683 Quelas, J. I., Mesa, S., Mongiardini, E. J., Jendrossek, D., & Lodeiro, A. R. (2016).
684 Regulation of Polyhydroxybutyrate Synthesis in the Soil Bacterium

- 685 Bradyrhizobium diazoefficiens. *Applied and Environmental Microbiology*,
686 82(14), 4299–4308. doi: 10.1128/AEM.00757-16
- 687 Reis, M. C. dos, Bagatini, I. L., Vidal, L. de O., Bonnet, M.-P., Marques, D. da M., &
688 Sarmento, H. (2019). Spatial heterogeneity and hydrological fluctuations drive
689 bacterioplankton community composition in an Amazon floodplain system.
690 *PLOS ONE*, 14(8), e0220695. doi: 10.1371/journal.pone.0220695
- 691 Santos-Júnior, C. D., Kishi, L. T., Toyama, D., Soares-Costa, A., Oliveira, T. C. S., de
692 Miranda, F. P., & Henrique-Silva, F. (2017). Metagenome Sequencing of
693 Prokaryotic Microbiota Collected from Rivers in the Upper Amazon Basin.
694 *Genome Announcements*, 5(2), e01450-16. doi: 10.1128/genomeA.01450-16
- 695 Santos-Júnior, C. D., Sarmento, H., de Miranda, F. P., Henrique-Silva, F., & Logares,
696 R. (2020). Uncovering the Genomic Potential of the Amazon River Microbiome
697 to Degrade Rainforest Organic Matter. *Microbiome*, 8, 151. doi:
698 10.1186/s40168-020-00930-w
- 699 Santos-Junior, C. D., Toyama, D., Oliveira, T. C. S., Pellon De Miranda, F., &
700 Henrique-Silva, F. (2019). Flood Season Microbiota from the Amazon Basin
701 Lakes: Analysis with Metagenome Sequencing. *Microbiology Resource*
702 *Announcements*, 8(17), e00229-19.
- 703 Satinsky, B. M., Crump, B. C., Smith, C. B., Sharma, S., Zielinski, B. L., Doherty,
704 M., ... Moran, M. A. (2014). Microspatial gene expression patterns in the
705 Amazon River Plume. *Proceedings of the National Academy of Sciences of the*
706 *United States of America*, 111(30), 11085–11090. doi:
707 10.1073/pnas.1402782111
- 708 Satinsky, B. M., Fortunato, C. S., Doherty, M., Smith, C. B., Sharma, S., Ward, N.
709 D., ... Crump, B. C. (2015). Metagenomic and metatranscriptomic inventories

- 710 of the lower Amazon River, May 2011. *Microbiome*, 3(1), 39. doi:
711 10.1186/s40168-015-0099-0
- 712 Satinsky, B. M., Smith, C. B., Sharma, S., Landa, M., Medeiros, P. M., Coles, V. J., ...
713 Moran, M. A. (2017). Expression patterns of elemental cycling genes in the
714 Amazon River Plume. *The ISME Journal*, 11(8), 1852–1864. doi:
715 10.1038/ismej.2017.46
- 716 Satinsky, B. M., Smith, C. B., Sharma, S., Ward, N. D., Krusche, A. V., Richey, J.
717 E., ... Moran, M. A. (2017). Patterns of Bacterial and Archaeal Gene Expression
718 through the Lower Amazon River. *Frontiers in Marine Science*, 4, 253. doi:
719 10.3389/fmars.2017.00253
- 720 Satinsky, B. M., Zielinski, B. L., Doherty, M., Smith, C. B., Sharma, S., Paul, J. H., ...
721 Moran, M. A. (2014). The Amazon continuum dataset: Quantitative
722 metagenomic and metatranscriptomic inventories of the Amazon River plume,
723 June 2010. *Microbiome*, 2, 17. doi: 10.1186/2049-2618-2-17
- 724 Sawakuchi, H. O., Neu, V., Ward, N. D., Barros, M. de L. C., Valerio, A. M., Gagne-
725 Maynard, W., ... Richey, J. E. (2017). Carbon Dioxide Emissions along the
726 Lower Amazon River. *Frontiers in Marine Science*, 4, 76. doi:
727 10.3389/fmars.2017.00076
- 728 Seemann, T. (2014). Prokka: Rapid prokaryotic genome annotation. *Bioinformatics*,
729 30(14), 2068–2069. doi: 10.1093/bioinformatics/btu153
- 730 Tarasov, A., Vilella, A. J., Cuppen, E., Nijman, I. J., & Prins, P. (2015). Sambamba:
731 Fast processing of NGS alignment formats. *Bioinformatics*, 31(12), 2032–2034.
732 doi: 10.1093/bioinformatics/btv098

- 733 Tatusov, R. L., Fedorova, N. D., Jackson, J. D., Jacobs, A. R., Kiryutin, B., Koonin, E.
 734 V, ... Natale, D. A. (2003). The COG database: An updated version includes
 735 eukaryotes. *BMC Bioinformatics*, 4(1), 41. doi: 10.1186/1471-2105-4-41
- 736 Toyama, D, Santos-Júnior, C., Kishi, L., Oliveira, T., Garcia, J. W., Sarmiento, H., ...
 737 Henrique-Silva, F. (2017). A snapshot on prokaryotic diversity of the Solimões
 738 River basin (Amazon, Brazil). *Genetics and Molecular Research*, 16(2),
 739 gmr16029567. doi: 10.4238/gmr16029567
- 740 Toyama, Danyelle, Kishi, L. T., Santos-Júnior, C. D., Soares-Costa, A., Souza De
 741 Oliveira, T. C., Pellon De Miranda, F., & Henrique-Silva, F. (2016).
 742 Metagenomics Analysis of Microorganisms in Freshwater Lakes of the Amazon
 743 Basin. *Genome Announc*, 4(6), 1440–16. doi: 10.1128/genomeA.01440-16
- 744 Tully, B. J., Sachdeva, R., Graham, E. D., & Heidelberg, J. F. (2017). 290 metagenome-
 745 assembled genomes from the Mediterranean Sea: A resource for marine
 746 microbiology. *PeerJ*, 5, e3558. doi: 10.7717/peerj.3558
- 747 UniProt Consortium. (2015). UniProt: A hub for protein information. *Nucleic Acids*
 748 *Research*, 43(D1), D204–D212. doi: 10.1093/nar/gku989
- 749 Varghese, N. J., Mukherjee, S., Ivanova, N., Konstantinidis, K. T., Mavrommatis, K.,
 750 Kyrpides, N. C., & Pati, A. (2015). Microbial species delineation using whole
 751 genome sequences. *Nucleic Acids Research*, 43(14), 6761–6771. doi:
 752 10.1093/nar/gkv657
- 753 Ward, N. D., Bianchi, T. S., Sawakuchi, H. O., Gagne-Maynard, W., Cunha, A. C.,
 754 Brito, D. C., ... Keil, R. G. (2016). The reactivity of plant-derived organic
 755 matter and the potential importance of priming effects along the lower Amazon
 756 River. *Journal of Geophysical Research: Biogeosciences*, 121(6), 1522–1539.
 757 doi: 10.1002/2016JG003342

- 758 Ward, N. D., Keil, R. G., Medeiros, P. M., Brito, D. C., Cunha, A. C., Dittmar, T., ...
759 Richey, J. E. (2013). Degradation of terrestrially derived macromolecules in the
760 Amazon River. *Nature Geoscience*, 6(7), 530–533. doi: 10.1038/ngeo1817
- 761 Warnes, G., Bolker, B., Bonebakker, L., Gentleman, R., Huber, W., Liaw, A., ...
762 Venables, B. (2017). gplots: Various R Programming Tools for Plotting Data
763 (Version 3.1.0) [R package]. Retrieved from [https://cran.r-](https://cran.r-project.org/web/packages/gplots/index.html)
764 [project.org/web/packages/gplots/index.html](https://cran.r-project.org/web/packages/gplots/index.html)
- 765 Westergaard, B., & Hansen, H. C. B. (1998). Determination of anions in soil solutions
766 by capillary zone electrophoresis. *Analyst*, 123(4), 721–724. doi:
767 10.1039/A707497B
- 768 Wickham, Hadley. (2009). *Ggplot2: Elegant graphics for data analysis*. Springer.
- 769 Winnen, B., Hvorup, R. N., & Saier, M. H. (2003). The tripartite tricarboxylate
770 transporter (TTT) family. *Research in Microbiology*, 154(7), 457–465. doi:
771 10.1016/S0923-2508(03)00126-8
- 772 Wissmar, R. C., Richey, J. E., Stallard, R. F., & Edmond, J. M. (1981). Plankton
773 Metabolism and Carbon Processes in the Amazon River, Its Tributaries, and
774 Floodplain Waters, Peru-Brazil, May-June 1977. *Ecology*, 62(6), 1622–1633.
775 doi: 10.2307/1941517
- 776 Yang, J., Jiang, H., Liu, W., Huang, L., Huang, J., Wang, B., ... Tolic, N. (2020).
777 Potential utilization of terrestrially derived dissolved organic matter by aquatic
778 microbial communities in saline lakes. *The ISME Journal*, 14(9), 2313–2324.
779 doi: 10.1038/s41396-020-0689-0
- 780 Yilmaz, P., Parfrey, L. W., Yarza, P., Gerken, J., Pruesse, E., Quast, C., ... Glöckner, F.
781 O. (2014). The SILVA and “all-species Living Tree Project (LTP)” taxonomic
782 frameworks. *Nucleic Acids Research*, 42(D1), D643–8. doi: 10.1093/nar/gkt1209

Yin, Y., Mao, X., Yang, J., Chen, X., Mao, F., & Xu, Y. (2012). DbCAN: a web resource for automated carbohydrate-active enzyme annotation. *Nucleic Acids Research*, 40(Web Server issue), W445-51. doi: 10.1093/nar/gks479

Zhou, L., Zhou, Y., Yao, X., Cai, J., Liu, X., Tang, X., ... Jeppesen, E. (2020). Decreasing diversity of rare bacterial subcommunities relates to dissolved organic matter along permafrost thawing gradients. *Environment International*, 134, 105330. doi: 10.1016/j.envint.2019.105330

8. Data Accessibility

This paper was made available as a preprint at Research Square (DOI 10.21203/rs.3.rs-32535). Metagenomes used to construct the Amazon River PGs are available at NCBI-SRA database as projects SRP044326, PRJEB25171 and SRP039390. The PGs' nucleotide sequences were deposited at the European Nucleotide Archive - ENA, in the project: PRJEB25176. The PGs' annotations are available at Zenodo (DOI 10.5281/zenodo.1484509).

9. Author Contributions

CDSJ, FHS and RL designed the study. CDSJ compiled and curated the data and performed bioinformatics analysis. CDSJ, FHS and RL interpreted the results. FHS and RL supervised and administered the project and provided funding. The original draft was written by CDSJ. All co-authors contributed substantially to manuscript revisions. All author(s) read and approved the final manuscript.

Table 1. Population genomes (PGs) identified in this study. PGs are described in terms of the river section from where they originate, their GC% content, size in 10⁶ bp (Mbp), completeness (C), contamination (Cx), quality classification and taxonomy (lowest possible rank).

PG	Section	GC (%)	Size (Mbp)	C (%)	Cx	Classification	Taxonomy
AM_0118	Upstream	37.77	1.35	81.87	0	Medium-Quality Draft	Streptomycetaceae
AM_0219	Upstream	36.37	1.05	76.13	0.71	Medium-Quality Draft	Methylophilus
AM_0226	Upstream	63.28	2.99	98.79	0.34	High-Quality Draft	Rhizobiales
AM_0228	Upstream	56.89	2.92	97.56	2.42	High-Quality Draft	Burkholderiales
AM_0233	Upstream	53	2.98	96.26	0	High-Quality Draft	Proteobacteria
AM_0240	Upstream	60.07	3.02	91.32	0.89	High-Quality Draft	Bacteria
AM_0244	Upstream	41.41	2.10	93.23	0.76	High-Quality Draft	Gammaproteobacteria
AM_0256	Upstream	53.45	2.66	88.24	2.15	Medium-Quality Draft	Bacteria
AM_0268	Upstream	52.62	2.91	86.69	0.42	Medium-Quality Draft	Bacteria
AM_0275	Upstream	52.7	2.41	72.7	1.41	Medium-Quality Draft	Rhizobiales
AM_0466	Upstream	41.13	1.03	62.5	0	Medium-Quality Draft	Bacteria
AM_0507	Upstream	36.3	1.00	79.25	0.64	Medium-Quality Draft	Betaproteobacteria
AM_0510	Upstream	62.61	2.00	98.08	1.1	High-Quality Draft	Bacteria
AM_0519	Upstream	65.04	4.82	96.74	0.07	High-Quality Draft	Gammaproteobacteria
AM_0528	Upstream	48.07	2.08	93.43	0.34	High-Quality Draft	Rhizobiales
AM_0546	Upstream	58.47	1.47	70.69	0.07	Medium-Quality Draft	Actinobacteria
AM_0608	Upstream	38.52	3.23	99.28	1.07	High-Quality Draft	Moraxellaceae
AM_0615	Upstream	36.39	1.56	99.51	0.97	High-Quality Draft	Thaumarchaeota
AM_0616	Upstream	59.76	5.51	98.66	0.59	High-Quality Draft	Betaproteobacteria
AM_0619	Upstream	42.57	2.83	91.18	2.52	High-Quality Draft	Bacteria
AM_0621	Upstream	56.53	7.87	90.33	0	High-Quality Draft	Bacteria
AM_0630	Upstream	52.99	2.89	95.43	2.17	High-Quality Draft	Proteobacteria
AM_0643	Upstream	56.96	2.89	91.49	2.2	High-Quality Draft	Burkholderiales
AM_0729	Upstream	53.85	2.45	84.4	1.26	Medium-Quality Draft	Bacteria
AM_0764	Upstream	50.2	3.89	86.77	1.1	Medium-Quality Draft	Bacteria
AM_0832	Downstream	44.61	3.47	96.39	1.64	High-Quality Draft	Bacteroidetes
AM_0849	Downstream	65.31	2.11	86.21	6.78	Medium-Quality Draft	Actinobacteria
AM_0854	Downstream	52.59	0.53	73.22	0	Medium-Quality Draft	Bacteria
AM_0876	Downstream	52.57	4.47	90.28	2.88	High-Quality Draft	Bacteria
AM_0902	Downstream	37.07	4.06	96.49	0	High-Quality Draft	Cyanobacteria
AM_0936	Downstream	62.13	2.69	86.63	4.45	Medium-Quality Draft	Bacteria
AM_1003	Downstream	59.74	6.26	95.93	0.91	High-Quality Draft	Betaproteobacteria
AM_1104	Downstream	59.11	3.00	78.96	3.08	Medium-Quality Draft	Bacteria
AM_1111	Downstream	38.28	2.82	90.95	6.68	Medium-Quality Draft	Bacteroidetes
AM_1205	Estuary	47.76	0.71	65.39	0	Medium-Quality Draft	Bacteria
AM_1312	Estuary	61.53	2.17	85.71	1.82	Medium-Quality Draft	Bacteria
AM_1409	Estuary	39.04	0.61	70.98	1.72	Medium-Quality Draft	Bacteria
AM_1503	Estuary	67.21	2.09	63.57	0.73	Medium-Quality Draft	Bacteria
AM_1603	Estuary	63.07	4.54	98.42	1.19	High-Quality Draft	Sphingomonadales
AM_1606	Estuary	55.47	2.24	96.4	0.61	High-Quality Draft	Bacteria
AM_1801	Plume	45.14	1.20	67.56	0	Medium-Quality Draft	Euryarchaeota

AM_1811	Plume	43.37	2.27	85.47	3.72	Medium-Quality Draft	Bacteria
AM_2104	Plume	60.62	2.11	84.85	4.55	Medium-Quality Draft	Bacteria
AM_2116	Plume	46.6	0.69	52.31	0	Medium-Quality Draft	Bacteria
AM_2124	Plume	40.54	1.53	87.19	1.49	Medium-Quality Draft	Flavobacteriaceae
AM_2202	Plume	47.45	1.80	97.31	0	High-Quality Draft	Bacteria
AM_2207	Plume	43.09	1.79	85.25	1.19	Medium-Quality Draft	Flavobacteriaceae
AM_2208	Plume	53.01	1.62	92.99	0.36	High-Quality Draft	Bacteria
AM_2324	Plume	37.43	3.41	90.78	0.99	High-Quality Draft	Bacteroidetes
AM_2502	Ocean	39.43	1.43	89.99	1.18	Medium-Quality Draft	Flavobacteriaceae
AM_2804	Ocean	33.5	3.29	93.56	0.27	High-Quality Draft	Cyanobacteria

812
813
814
815
816
817
818
819
820
821
822
823
824
825
826

For Review Only

Figure 1. Metagenomes and Population Genomes (PGs) from the Amazon River. Distribution of the 106 metagenomes along the five sections of the Amazon River: Upstream, Downstream, Estuary, Plume and Ocean, from where the PGs derive (a). The size of the red circles indicates the sequencing coverage per metagenome (Cov./Gbp). Number of curated PGs that were reconstructed per river section (b)

Figure 2. Abundance of PGs along Amazon River sections. PG abundances are indicated as Log(RPKG+1).

836

Figure 3. Presence of cellulose degradation functions in the Amazon River PGs. The y-axis shows main glucosyl hydrolase (GHs) families known to participate in cellulose degradation. The x-axis shows the PGs. The size of the bubble indicates number of genes per protein family.

Figure 4. Pathways involved in tricarboxylates usage and carbon storage. Amazon River PGs were analysed for their potential to use tricarboxylates a) and only genomes containing the complete TTT system b) are reported (IM – inner membrane; OM – outer membrane). The polyhydroxy-butyrates/alkanoates production was further investigated c) to assess the potential for carbon storage in Amazon River PGs. Those PGs with the potential to store carbon are shown in d), indicating the number of different Pha genes. NB: the enzyme phaE allows in the presence of the enzyme phaC, the biosynthesis of polyhydroxy-alkanoate/butyrate, a hybrid biopolymer (Poly3HB-co-4HB). The protein phaR regulates the accumulation of polyhydroxy-butyrates in granules inside the cell.

Figure 5. Genome-based model for TeOM degradation in the Amazon River. Green arrows indicate beneficial effects for assemblages in terms of cell growth, blue arrows represent the secretion of ectoenzymes and the degradation of TeOM, while the black arrow indicates lignin oxidation generating low molecular weight aromatic by-products. The red arrow indicates growth inhibition and the beige arrow its suppression. In this model, three microbial assemblages interact. Two of them are responsible for exposing and processing TeOM. Besides providing structural carbon and energy for cell growth, by-products of this metabolism inhibit TeOM degradation. The third assemblage prevents TeOM degradation to stop by consuming these by-products and storing the resulting carbon intracellularly.