Metabolic Network Analysis and Metatranscriptomics of a Cosmopolitan and Streamlined Freshwater Lineage

Joshua J. Hamilton1,\*, Sarahi L. Garcia2, Brittany S. Brown1, Jeffrey R. Dwulit-Smith1, Francisco Moya4, Ben O. Oyserman4, Sarah L.R. Stevens1, Stefan Bertilsson2, Katrina T. Forest1, Susannah G, Tringe3, Tanja Woyke3, and Katherine D. McMahon1,4,\*

1 Department of Bacteriology, University of Wisconsin-Madison, Madison, WI, USA; 2 Department of Ecology and Genetics, Uppsala University, Uppsala, Sweden; 3 United States Department of Energy Joint Genome Institute, Walnut Creek, CA, USA; 4 Department of Civil and Environmental Engineering, University of Wisconsin-Madison, Madison, WI, USA

\* Correspondence: Joshua J. Hamilton, jjhamilton2@wisc.edu and Katherine D. McMahon, trina.mcmahon@wisc.edu

# Abstract

# Introduction

# Materials and Methods

## Single-Cell Genome Generation, Selection, and Sequencing

Single-cell genomes were collected from the top of the water column (depth <1m) from each of two lakes, Mendota (Madison, WI, USA) and Damariscotta (Lincoln County, ME USA), in 2009. Samples were cryopreserved and sent to the Single Cell Genomics Center at the Bigelow Laboratory for Ocean Sciences for sorting, as previously described (Martinez-Garcia et al. 2012, Garcia et al. (2013)). Partial 16S rRNA genes amplified previously (Martinez-Garcia et al. 2012) were phylogenetically classified using a controlled nomenclature for freshwater bacteria (Newton et al. 2011) by insertion into references trees created in the ARB software package (W. Ludwig et al. 2004).

Actinobacterial SAGs used in this study were then sent to the JGI for sequencing and assembly, also as previously described (Ghylin et al. 2014). Briefly, shotgun libraries were constructed for each of the SAGs from re-amplified MDA products and sequenced on an Illumina HiSeq2000. All general aspects of and detailed protocols for library construction and sequencing can be found on the JGI website (http://www.jgi.doe.gov/).

For assembly, raw sequence data was first passed through a filtering program developed at JGI to eliminate known sequencing and library preparation artifacts. Assembly was then performed using Velvet [Zerbino and Birney (2008) and ALLPATHS-LG (Butler et al. 2008). Additional details of the assembly process have been previously described (Ghylin et al. 2014) and are available through the JGI Genome Portal (http://genome.jgi.doe.gov) Genome sequences are available through IMG (https://img.jgi.doe.gov/cgi-bin/mer/main.cgi). Genome-specific information can be accessed in both databases by searching for the IMG Taxon OIDs given in Table 1.

## Metagenome Sampling, Sequencing, Assembly, and Binning

Sample collection, DNA sequencing, metabgenomic assembly, and genomic binning for the Trout Bog samples have been described previously (Bendall et al. 2016), and similar procedures were followed for Lake Mendota samples. A summary is provided here.

For Lake Mendota, depth-integrated water samples were collected from the top 12 meters at 96 time points during ice-free periods from 2008 to 2011. For Trout Bog, depth-integrated water samples were collected from the epilimnion (44 samples) and hypolimnion (45 samples) layers during ice-free periods from 2007 to 2009. All samples were filtered on 0.2 μm polyethersulfone filters (Supor, Pall Corp) prior to storage at -80°C, as described previously (Bendall et al. 2016). DNA was extracted from these filters using the FastDNA kit (MP Biomedicals) and sent to the JGI for sequencing, as described previously (Bendall et al. 2016).

Shotgun libraries were constructed for each of the samples and sequenced on an Illumina GA IIx (four Trout Bog samples) or an Illumina HiSeq2000 (all other samples), following a 2x150 indexed run recipe as previously described (Bendall et al. 2016). All general aspects of and detailed protocols for library construction and sequencing can be found on the JGI website (http://www.jgi.doe.gov/). Metagenomic sequence reads are publicly available on the JGI Genome Portal (http://genome.jgi.doe.gov/) under Proposal ID 394.

Raw sequence data was passed through a filtering program developed at JGI to eliminate known sequencing and library preparation artifacts. Prior to assembly, reads were merged with FLASH (Magoc and Salzberg 2011), as previously described (Bendall et al. 2016). Merged reads were pooled by lake and layer into three co-assemblies using SOAPdenovo (Luo et al. 2012), and contigs from the resulting assemblies were assembled into a final assembly using Minimus (Sommer et al. 2007), as previously described (Bendall et al. 2016). Additional details of the assembly process and metagenomic sequence reads are available through the JGI Genome Portal (http://genome.jgi.doe.gov) under Proposal ID 394.

Genomes were binned from each metagenomic co-assembly using MetaBat (D. D. Kang et al. 2015), as described previously (Bendall et al. 2016). Briefly, contigs were classified into bins using tetranucleotide frequency and coverage patterns across the time-series and then manually curated, as previously described (Bendall et al. 2016). Genome sequences are available through IMG (https://img.jgi.doe.gov/cgi-bin/mer/main.cgi) by searching for the IMG Taxon OIDs given in Table 1. Genomes were classified using taxonomic assignments from a set of 37 highly-conserved single-copy marker genes using Phylosift (Darling et al. 2014), as previously described (Bendall et al. 2016). Final bin size and number of contigs are reported in Table 1.

## Metatranscriptome Sampling and Sequencing

Four samples were collected from the top of the water column (depth <1m) from Lake Mendota (Madison, WI, USA) over a twenty-four hour period on August 20 and 21, 2015. For each sample, between 200 and 400 mL lake water was filtered onto a 0.2 μm polyethersulfone filter (Supor, Pall Corp) and stored at -80°C until extraction.

Prior to extraction, three samples were spiked with an internal standard to enable quantification of total transcript abundance, following an established protocol (Satinsky et al. 2014). Briefly, a 970-nucleotide-long mRNA standard was synthesized using a a T7 RNA polymerase and the Riboprobe In Vitro Transcription System (Promega, Madison, WI), according to the manufacturer’s protocol. A fixed quantity of each standard (1.172 x 10^10 copies) was added independently to each lysis tube immediately prior to the addition of the sample filter.

Samples were subject to TRIzol-based RNA extraction (Thermo Fisher Scientific, Waltham, MA) followed by on-column DNAse digestion and RNA purification using an RNeasy Mini Kit (Qiagen, Venlo, Netherlands). RNA was then sent to the University of Wisconsin-Madison Biotechnology Center (https://www.biotech.wisc.edu) for sequencing. There, samples were prepared for sequencing using the TruSeq RNA Library Prep Kit v2 (Illumina, San Diego, CA), with the addition of a step for selective ribosomal RNA depletion using the Ribo-Zero rRNA Removal Kit (Bacteria) (Illumina). The resulting cDNA libraries were pooled in an equimolar ratio, and sequenced on an Illumina HiSeq2500.

Raw paired-end reads were then merged using FLASH (Magoc and Salzberg 2011) using default parameters. Finally, additional rRNA and ncRNA sequences were removed using SortMeRNA (Kopylova, Noe, and Touzet 2012) using default parameters. SortMeRNA was run using eight built-in databases for bacterial, archaeal, and eukaryotic small and large ribosomal subunits and ncRNAs, derived from the SILVA 119 (Quast et al. 2013) and RFAM (Nawrocki et al. 2015) databases.

Additional information, including all protocols and scripts for RNA analysis, can be found on Github (https://github.com/McMahonLab/OMD-TOILv2). Raw RNA sequences can be found on the National Center for Biotechnology Information (NCBI) website under BioProject PRJNA######.

# Results

# Discussion

# Acknowledgements

# Conflict of Interest

The authors declare no conflict of interest.

# References

Bendall, Matthew L, Sarah LR Stevens, Leong-Keat Chan, Stephanie Malfatti, Patrick Schwientek, Julien Tremblay, Wendy Schackwitz, et al. 2016. “Genome-wide selective sweeps and gene-specific sweeps in natural bacterial populations.” *The ISME Journal* 10 (7). Nature Publishing Group: 1589–1601. [doi:10.1038/ismej.2015.241](http://doi.org/10.1038/ismej.2015.241).

Butler, Jonathan, Iain MacCallum, Michael Kleber, Ilya A Shlyakhter, Matthew K Belmonte, Eric S Lander, Chad Nusbaum, and David B Jaffe. 2008. “ALLPATHS: De novo assembly of whole-genome shotgun microreads.” *Genome Research* 18 (5): 810–20. [doi:10.1101/gr.7337908](http://doi.org/10.1101/gr.7337908).

Darling, Aaron E, Guillaume Jospin, Eric Lowe, Frederick A Matsen, Holly M Bik, and Jonathan A Eisen. 2014. “PhyloSift: phylogenetic analysis of genomes and metagenomes.” *PeerJ* 2 (January): e243. [doi:10.7717/peerj.243](http://doi.org/10.7717/peerj.243).

Garcia, Sarahi L, Katherine D McMahon, Manuel Martinez-Garcia, Abhishek Srivastava, Alexander Sczyrba, Ramunas Stepanauskas, Hans-Peter Grossart, Tanja Woyke, and Falk Warnecke. 2013. “Metabolic potential of a single cell belonging to one of the most abundant lineages in freshwater bacterioplankton.” *The ISME Journal* 7 (1). Nature Publishing Group: 137–47. [doi:10.1038/ismej.2012.86](http://doi.org/10.1038/ismej.2012.86).

Ghylin, Trevor W, Sarahi L Garcia, Francisco Moya, Ben O Oyserman, Patrick Schwientek, Katrina T Forest, James Mutschler, et al. 2014. “Comparative single-cell genomics reveals potential ecological niches for the freshwater acI Actinobacteria lineage.” *The ISME Journal* 8 (12). Nature Publishing Group: 2503–16. [doi:10.1038/ismej.2014.135](http://doi.org/10.1038/ismej.2014.135).

Kang, Dongwan D, Jeff Froula, Rob Egan, and Zhong Wang. 2015. “MetaBAT, an efficient tool for accurately reconstructing single genomes from complex microbial communities.” *PeerJ* 3: e1165. [doi:10.7717/peerj.1165](http://doi.org/10.7717/peerj.1165).

Kopylova, Evguenia, Laurent Noe, and Helene Touzet. 2012. “SortMeRNA: fast and accurate filtering of ribosomal RNAs in metatranscriptomic data.” *Bioinformatics* 28 (24): 3211–17. [doi:10.1093/bioinformatics/bts611](http://doi.org/10.1093/bioinformatics/bts611).

Ludwig, Wolfgang, Oliver Strunk, Ralf Westram, Lothar Richter, Harald Meier, Yadhukumar, Arno Buchner, et al. 2004. “ARB: a software environment for sequence data.” *Nucleic Acids Research* 32 (4): 1363–71. [doi:10.1093/nar/gkh293](http://doi.org/10.1093/nar/gkh293).

Luo, Ruibang, Binghang Liu, Yinlong Xie, Zhenyu Li, Weihua Huang, Jianying Yuan, Guangzhu He, et al. 2012. “SOAPdenovo2: an empirically improved memory-efficient short-read de novo assembler.” *GigaScience* 1: 18. [doi:10.1186/2047-217X-1-18](http://doi.org/10.1186/2047-217X-1-18).

Magoc, Tanja, and Steven L Salzberg. 2011. “FLASH: fast length adjustment of short reads to improve genome assemblies.” *Bioinformatics* 27 (21): 2957–63. [doi:10.1093/bioinformatics/btr507](http://doi.org/10.1093/bioinformatics/btr507).

Martinez-Garcia, Manuel, Brandon K Swan, Nicole J Poulton, Monica Lluesma Gomez, Dashiell Masland, Michael E Sieracki, and Ramunas Stepanauskas. 2012. “High-throughput single-cell sequencing identifies photoheterotrophs and chemoautotrophs in freshwater bacterioplankton.” *The ISME Journal* 6 (1). Nature Publishing Group: 113–23. [doi:10.1038/ismej.2011.84](http://doi.org/10.1038/ismej.2011.84).

Nawrocki, Eric P., Sarah W. Burge, Alex Bateman, Jennifer Daub, Ruth Y. Eberhardt, Sean R. Eddy, Evan W. Floden, et al. 2015. “Rfam 12.0: Updates to the RNA families database.” *Nucleic Acids Research* 43 (D1): D130–37. [doi:10.1093/nar/gku1063](http://doi.org/10.1093/nar/gku1063).

Newton, Ryan J, Stuart E Jones, Alexander Eiler, Katherine D McMahon, and Stefan Bertilsson. 2011. “A guide to the natural history of freshwater lake bacteria.” *Microbiology and Molecular Biology Reviews* 75 (1): 14–49. [doi:10.1128/MMBR.00028-10](http://doi.org/10.1128/MMBR.00028-10).

Quast, Christian, Elmar Pruesse, Pelin Yilmaz, Jan Gerken, Timmy Schweer, Pablo Yarza, Jörg Peplies, and Frank Oliver Glöckner. 2013. “The SILVA ribosomal RNA gene database project: improved data processing and web-based tools.” *Nucleic Acids Research* 41 (D1): D590–96. [doi:10.1093/nar/gks1219](http://doi.org/10.1093/nar/gks1219).

Satinsky, Brandon M, Brian L Zielinski, Mary Doherty, Christa B Smith, Shalabh Sharma, John H Paul, Byron C Crump, et al. 2014. “The Amazon continuum dataset: quantitative metagenomic and metatranscriptomic inventories of the Amazon River plume, June 2010.” *Microbiome* 2 (1): 17. [doi:10.1186/2049-2618-2-17](http://doi.org/10.1186/2049-2618-2-17).

Sommer, Daniel D, Arthur L Delcher, Steven L Salzberg, and Mihai Pop. 2007. “Minimus: a fast, lightweight genome assembler.” *BMC Bioinformatics* 8: 64. [doi:10.1186/1471-2105-8-64](http://doi.org/10.1186/1471-2105-8-64).

Zerbino, Daniel R, and Ewan Birney. 2008. “Velvet: Algorithms for de novo short read assembly using de Bruijn graphs.” *Genome Research* 18 (5): 821–29. [doi:10.1101/gr.074492.107](http://doi.org/10.1101/gr.074492.107).