Genome Analysis of a Novel Photoarsenotroph, *Rhodobacter sp.* str. ORIO

Pranav Kirti[1], Isaac Chang[1], Madeline Day[1], Sanjin Mehić[2], and Chad Saltikov[2]

[1i]Lynbrook High School, 1280 Johnson Ave, San Jose, CA 95129
[1ii]Saratoga High School, 20300 Herriman Ave, Saratoga, CA 95070
[1iii]Amador Valley High School, 1155 Santa Rita Rd, Pleasanton, CA 94566
2i, 2ii Department of Microbiology and Environmental Toxicology, University of California, Santa Cruz, 1156 High St, Santa Cruz, CA 95064

***Abstract*— Photosynthetic arsenite oxidation (photoarsenotrophy) is an anoxygenic process where arsenite is used as an electron donor for growth. Through this process, bacteria oxidize arsenite (As(III)) to arsenate (As(V)), a less toxic substance. Extremophiles from hypersaline environments, which are difficult to grow and genetically manipulate, were previously used to determine that photoarsenotrophy is encoded by the *arxB2AB1CD* gene cluster. Here, we analyzed a novel freshwater bacterium, *Rhodobacter sp.* str. ORIO, for its potential to serve as a model organism in the study of photoarsenotrophy. We utilized NCBI BLAST to gather *arx* sequences, EBI Clustal Omega to construct phylogenetic trees, and InterPro, Pfam, Phobius, and SAPS to analyze ORIO's *arx* gene products individually. Lastly, we searched for the *arx* pathway in over 35,000 metagenomes from JGI IMG/MER and mapped the presence of photoarsenotrophs worldwide. Our findings indicate that ORIO contains an *arx* gene cluster similar to those previously studied, and that *arxA-like* genes are ubiquitous in nature, with concentrations in the U.S. and Southeast Asia. In conclusion, these results suggest that ORIO can serve as a model organism for photoarsenotrophy – providing the first step toward bioremediation by detoxifying arsenic-contaminated water and improving water quality.**

I. Introduction

Arsenic contamination affects up to 200 million people worldwide; exposure can have carcinogenic effects and lead to arsenicosis, arsenic poisoning that affects multiple organs [1]. Arsenic-contaminated groundwater is a detrimental problem in lower- and middle-income countries with large geologic arsenic deposits such as Bangladesh, where an estimated 27% of wells exceed the 10 µg/L contaminant limit [2]. The need for clean water will only increase as the human population grows.

Bacteria play a crucial role in cycling arsenic through the environment, as they utilize arsenite as an electron donor and energy source during cellular respiration (oxidizing it into arsenate). Therefore, understanding their function will help researchers remove arsenic from water and mitigate contamination. Investigations on bacteria have led to a hypothesized model of the *arx* gene cluster, which encodes for the anaerobic arsenite oxidase complex and is responsible for photoarsenotrophy [3]. However, photoarsenotrophy has only been researched in extremophiles, which are difficult to grow and genetically manipulate in the laboratory. To circumvent these difficulties and simultaneously fill this research gap, we analyzed the genome of *Rhodobacter sp.* strain ORIO, a freshwater photoarsenotroph, and we surveyed numerous metagenomes for *arx*-like gene clusters. Our findings suggest that ORIO is a suitable model organism for photoarsenotrophy and the *arx* gene cluster is globally ubiquitous in a diverse group of environments (Fig 2B).

II. Methods

Organism: Rhodobacter sp. str. ORIO (Owens River Isolate Oxidizer) is the first non-extremophile photoarsenotroph discovered and was isolated from Owens River (CA) sediments in 2014 by the Saltikov Lab (University of California, Santa Cruz). Sanjin Mehić (PhDc) cultured *Rhodobacter sp.* str. ORIO with PNS medium [4].

Genome Assembly and Annotation: We used raw 150 bp paired-end Illumina MiSeq sequencing reads to assemble the ORIO genome. Forward and reverse reads were uploaded to PATRIC [5], a database for bacterial genome analysis, and its Unicycler pipeline was used to assemble the genome. Next, we used RAST-tk via PATRIC's annotation service [6] to annotate the genome using the default setting.

Phylogenetic Analysis: After generating our genome draft, we created phylogenetic trees for each gene in the *arxXSR* and *arxB1AB2CD* gene clusters. We used the ORIO *arx* sequences obtained from our draft genome with NCBI BLAST [7] to search for the individual *arx* genes, collect the top 25 matches, and create an alignment file using EBI Clustal Omega [8]. Then, we inputted the alignment into EBI Simple Phylogeny [8] to create a phylogenetic tree and used iTOL [9] to customize each tree.

Protein Analyses: To investigate the biochemical properties of the *arx* genes, we analyzed each gene from the *arxXSR* and *arxB1AB2CD* gene clusters with EBI's bioinformatics toolkits: InterPro [10], Pfam [12], Phobius [12], and SAPS [8]. These programs predicted each gene product's molecular weight, protein family/domains, hydrophobicity, and transmembrane topology.

Metagenomic Analysis and Mapping: We blasted ORIO's *arxA* gene in over 35,000 metagenomes on JGI IMG/MER [13, 14] to determine the *arx* genes' environmental relevancy. For each metagenome hit, we searched for the hypothesized "GRGWG" active site in [15] the protein translation of *arxA*. We created a map that illustrates where *arxA* and other *arx* genes can be detected (Fig. 2).

III. Data Analysis

Genome Assembly and Annotation: PATRIC's genome assembly feature provided us with the summary statistics found in Table 1. ORIO's genome size and GC content falls into the range commonly seen in *Rhodobacter* species. We found the entire *arx* gene cluster on a 27.5 kB contig.

Table 1. ORIO GENOME CHARACTERISTICS

GC Content	*Contigs*	*Base Pairs*	*N50*
68.2%	160	4,601,101	217,804

a. The assembled and annotated genome had a course consistency of 98.0%, a fine consistency of 96.8%, 3089 protein-encoding genes with functional assignment, and 1623 protein-encoding genes without functional assignment.

Phylogenetic Analysis: We created protein phylogenetic trees for each of the *arx* genes. The phylogenetic trees indicate that the genera *Rhodobacterales*, *Rhodobacteraceae*, and *Rhodospirillales* contain closely related *arx* sequences to ORIO. Furthermore, Table 2 includes bacteria closely related to ORIO based on the *arx* genes' phylogeny.

Table 2. CLOSELY RELATED BACTERIA TO ORIO

ORIO Gene	*Closest Relative*
arxA	*Thiocapsa rosea*
arxB1	*Rhodovulum steppense*
arxB2	*Rhodovulum steppense*
arxC	*Ectothiorhodospira sp.* PHS-1
arxD	*Azoarcus sp.* CIB

a. Selected bacteria are sister taxa of ORIO based on each gene's phylogeny.

Protein Domain Analyses: InterPro predicted protein domains and conserved sequences present in each *arx* gene translation; Pfam generated predictions for the protein encoded by each *arx* gene; Phobius created a graph conveying predictions for transmembrane topology and signal peptides; and SAPS produced a list of protein sequence properties, including molecular weight. This data was used to reinforce the current hypothesized photoarsenotrophy model (Fig. 1). For example, it was hypothesized that *ArxC* is a transmembrane protein, which our Phobius results demonstrated. Specifically, we verified the location and function of *ArxA*, *ArxB*, *ArxB2*, and *ArxC*.

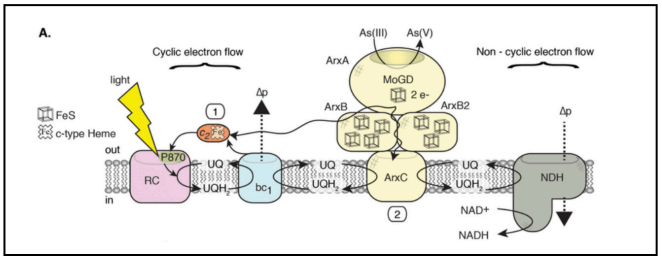

Figure 1. The current hypothesized model of photoarsenotrophy. This model was developed by the Saltikov Lab using extremophiles [16].

Metagenomic Analysis and Mapping: We searched *arxA*'s protein sequence in BLAST, which yielded 371 hits for bacteria with *arx*-related genes within their metagenomes (Fig. 2A). Specifically, we filtered our results by searching for the *ArxA* active site sequence (GRGWG). Our map exemplifies the ubiquity of *arx* genes and indicates that *arx* genes can be found on all seven continents. Also, we found that bacteria containing the *ArxA* active site were distributed among 11 primary environments, which included hot springs, wastewater, and freshwater areas (Fig. 2B).

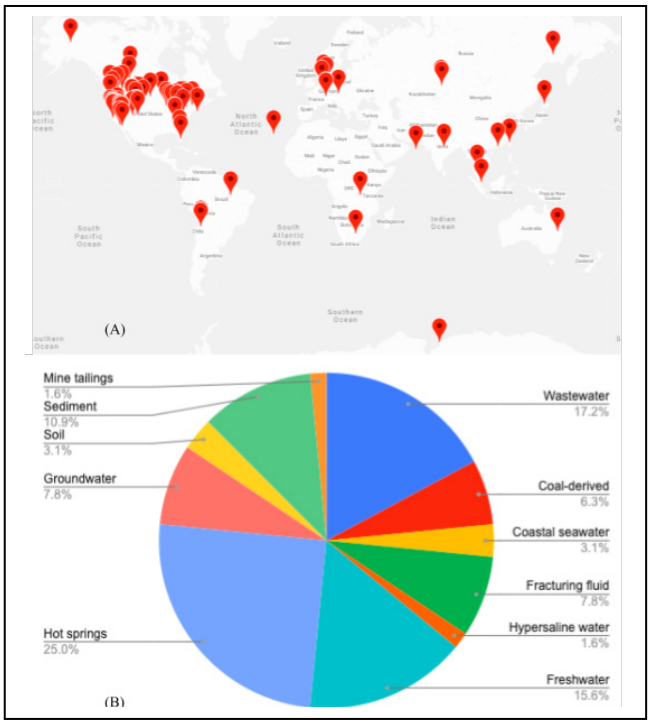

Figure 2. (A) Locations with the *ArxA* active site present. (B) A pie chart depicting the environments where *arx* genes were found.

IV. RESULTS/DISCUSSION

Overall, our findings are fundamental to understanding photoarsenotrophy and have established *Rhodobacter sp.* str. ORIO as an excellent model organism for photoarsenotrophy. This is because its *arx* genes are closely related to previously studied photoarsenotrophs. Protein domain analyses validated that each ORIO *arx* gene product contains the same features that were previously described in photoarsenotroph studies. For example, we detected hypothesized molybdenum-containing domains in ArxA, iron-sulfur clusters in the ArxB2 and ArxB1 putative proteins, and several transmembrane domains in ArxC (Fig. 1). Our discovery will allow researchers to more easily and efficiently genetically manipulate *arx* genes. This will expedite the overall study of photoarsenotrophy and drive novel experimentation.

Lastly, we are the first group to detect *arxA*-like sequences in metagenomes on all seven continents (Fig. 2A). Freshwater, groundwater, and wastewater environments accounted for 40.2% of *arxA* hits, whereas hot spring and saline environments accounted for 29.7% (Fig. 2B). This is interesting since nearly all previous photoarsenotrophy studies have been conducted in hot spring and saline environments. When taken into consideration with ORIO's genetic malleability [17], ORIO emerges as the leading model organism for studying photoarsenotrophy.

Future studies should aim to complete the genome sequence of *Rhodobacter sp.* str. ORIO and perform gene knockouts to further enhance our knowledge of the mechanisms behind photoarsenotrophy. In the long term,

the goal is to harness the unique oxidative ability of these bacteria to develop compact technology that will improve water quality through bioremediation.

V. Acknowledgements

This project was made with the support of our primary mentor, Sanjin Mehić (PhDc), and faculty advisor, Dr. Chad Saltikov (PhD). We would also like to thank the Science Internship Program and the University of California, Santa Cruz for providing us with an opportunity to conduct open-ended research.

VI. References

[1] Wang Q, et al., Environ Microbiol, 2018, 20: 1782-1793. doi:10.1111/1462-2920.14108
[2] Islam MS, Islam F, Accessed 2020, IWA Publishing. https://www.iwapublishing.com/news/arsenic-contamination-groundwater-bangladesh-environmental-and-social-disaster
[3] Hernandez-Maldonado J, et al., Environ Microbiol, 2017 Jan;19(1):130-141. doi: 10.1111/1462-2920.13509.
[4] Lindquist, J. 2014 July;21. https://www.splammo.net bact102/102pnsb.html
[5] Davis JJ, et al., Nucleic Acids Res. 2020 Jan 8;48(D1):D606-D612. doi: 10.1093/nar/gkz943.
[6] Brettin, T., et al., Sci Rep. 2015. 5: 8365.
[7] BlastP [internet]. Bethesda (MD): National Library of Medicine (US), National Center for Biotechnology Information; 2004 – [cited 2020 October 6]. Available from: https://www.ncbi.nlm.nih.gov/blast/
[8] Madeira F, Park YM, Lee J, et al., Nucleic Acids Research, 2019 Jul;47(W1):W636-W641. DOI: 10.1093/nar/gkz268.
[9] Ivica Letunic, et al., Nucleic Acids Research, Volume 47, Issue W1, 02 July 2019, Pages W256–W259, https://doi.org/10.1093/nar/gkz239
[10] Alex L Mitchell, et al., Nucleic Acids Research, Jan 2019, (doi: 10.1093/nar/gky1100)
[11] Sara El-Gebali, et al., Nucleic Acids Research, Volume 47, Issue D1, 2019, Pages D427–D432, https://doi.org/10.1093/nar/gky995.
[12] Käll L, Krogh A, Sonnhammer EL. J Mol Biol. 2004 May 14;338(5):1027-36. doi: 10.1016/j.jmb.2004.03.016.
[13] Chen IA, et al., Nucleic Acids Res. 2019 Jan 8;47(D1):D666-D677. doi: 10.1093/nar/gky901.
[14] Supratim Mukherjee, et al., Nucleic Acids Research, Volume 47, Issue D1, 08 January 2019, Pages D649–D659, DOI:10.1093/nar/gky977
[15] Glasser, Nathaniel R., et al. Proceedings of the National Academy of Sciences, vol. 115, no. 37, National Academy of Sciences, Sept. 2018, pp. E8614–23. doi:10.1073/pnas.1807984115.
[16] Hernandez-Maldonado J., 2017. Dissertation. University of California Santa Cruz
[17] Mehic S and Saltikov CW. ASM Microbe. June 20-24, 2019. Poster. San Francisco, CA. DOI: 10.13140/RG.2.2.13409.66403/1