In silico analysis of cysteine content, in six genes, across three marine *Altererythrobacter* strains sampled at varying barometric pressures

Conor Barrie: 13 May 2017.

conorbarrie@gmail.com

Marine microbes experience varying degrees of barometric pressure depending on the depth at which they reside. High barometric pressure can create an extreme environment which favors the survival of barophilic and barotolerant bacteria. Environmental stressors will select for beneficial mutations in microbes that allow their proteins to withstand and function under high barometric pressure. Proteins that function efficiently under such extreme pressures are of commercial interest for the development of high pressure bioreactors (Horikoshi, 1998).

One mechanism by which bacteria stabilize their proteins in extreme, thermophilic, environments is the incorporation of cysteine amino acids, which can form protein stabilizing disulfide bridges (Barton, 2005). As a result, this paper aims to quantify the cysteine content in six genes across three marine *Altererythrobacter* strains that were sampled from varying depth, thus living under corresponding degrees of barometric pressure. The purpose of this study is to test whether marine *Altererythrobacter* strains, living at greater depths, have proportionally more cysteine content, which could improve protein stability under high pressures. It was hypothesized that organisms sampled at greater depths would exhibit higher amounts of cysteine in their proteins as a mechanism to cope with the increased barometric pressure.

This hypothesis was tested by selecting three open source genomes, from three marine *Altererythrobacter* species, which were sampled at different depths. The strains selected were: *A. atlanticus* strain 26DY36 (taxonomy ID: 1267766), *A. epoxidivorans* strain JCS350 (taxonomy ID: 361183) and *A. marensis* strain KCTC22370 (taxonomy ID: 543877). *A. atlanticus* was sampled (at 2500 m depth) and genomically annotated by Wu *et al.* (2015). *A. epoxidivorans* was sampled (at a depth of 120 m) and genomically annotated by Kwon *et al.* (2007). Lastly, *A. marensis* was sampled (at 0 m depth) and genomically annotated by Seo and Lee (2010). The barometric pressures experienced by each of the samples was 256 atm, 13 atm and 1 atm, respectively.

All three strains have similar genomic compositions. In fact, the percentage of identical sequences between each of the strains is as follows: *A. atlanticus* and *A. epoxidivorans*, 92%; *A. atlanticus* and *A. marensis,* 85%; and *A. epoxidivorans* and *A. marensis*, 89% (percentages obtained using NCBI BLAST).

The genes chosen were shared among the three strains. Additionally, an attempt was made to choose genes that were similar in length among each species. The genes chosen code for: NADH dehydrogenase, RNA polymerase sigma rpoD, DNA primase, triose phosphate isomerase, heat shock protein (HSP) 33 and ferrous iron transport B. Translational outputs were determined from the NCBI databases and they were analyzed for their cysteine content. Thus a total of 18 translational outputs were analyzed, where six originated from each of the three *Altererythrobacter* strains of interest. The NCBI protein ID numbers, for each analyzed translational output, can be found in table A1.

Strains sampled from different depths were found to have similar proportions of cysteine (ANOVA; $F= 0.7$; $df= 2, 15$; $P= 0.5$; Table 1). The data were ensured to meet the assumptions of ANOVA with a Levene's test and a Shapiro-Wilk test. All stats were performed using R 3.1.1 (RStudio Team, 2015). This result was similar to a previous study that analyzed the same genes for differences in proline content (Barrie, 2016). The hypothesis that deeper-living strains have greater cysteine content is not supported by this data; however the theory cannot be rejected. This study

was only intended as a preliminary investigation and in order for the hypothesis to be thoroughly supported or rejected a more comprehensive study would be needed. This comprehensive study would involve a larger sample size involving more organisms and more genes across many types of bacterial strains. Future studies into the composition of barophilic and barotolerant proteins is a worthwhile endeavour and should be investigated thoroughly in the future.

Table 1: Table showing the strain (sample depth in m) and the proportion of cysteine in each analyzed gene. The mean proportion of cysteine content is also shown and these values were found to be similar between all three strains (ANOVA; F= 0.7; df= 2, 15; P= 0.5).

strain (depth)	**Proportion of cysteine**						mean proportion of cysteine
	NADH dehydrogenase	RNA pol sigma rpoD	DNA primase	triose P isomerase	HSP 33	ferrous iron transport B	
A. atlanticus (≈2500m)	0.000	0.001	0.008	0.008	0.020	0.005	0.007
A. epoxidivorans (≈120m)	0.017	0.001	0.008	0.017	0.024	0.005	0.012
A. marensis (≈0m)	0.023	0.001	0.008	0.012	0.024	0.005	0.012

Literature Cited

Barrie, C. (2016). Proline content and varying barometric pressure: an analysis of six genes, across three marine *Altererythrobacter* strains. *PeerJ Preprints* 4:e2071v1 https://doi.org/10.7287/peerj.preprints.2071v1

Barton, L. (2005). *Structural and Functional Relationships in Prokaryotes*. Springer: New York, NY.

Horikoshi, K. (1998). Barophiles: deep-sea microorganisms adapted to an extreme environment. *Current Opinion in Microbiology*, 1(3), 291-295.

Kwon, K. K., Woo, J. H., Yang, S. H., Kang, J. H., Kang, S. G., Kim, S. J., & Kato, C. (2007). *Altererythrobacter epoxidivorans* gen. nov., sp. nov., an epoxide hydrolase-active, mesophilic marine bacterium isolated from cold-seep sediment, and reclassification of *Erythrobacter luteolus* Yoon et al. 2005 as *Altererythrobacter luteolus* comb. nov. *International journal of systematic and evolutionary microbiology*, 57(10), 2207-2211.

RStudio Team. (2015). Integrated Development for R. RStudio, Inc., Boston, MA URL http://www.rstudio.com/

Seo, S. H., & Lee, S. D. (2010). *Altererythrobacter marensis* sp. nov., isolated from seawater. *International journal of systematic and evolutionary microbiology*, 60(2), 307-311.

Wu, Y. H., Cheng, H., Zhou, P., Huo, Y. Y., Wang, C. S., & Xu, X. W. (2015). Complete genome sequence of the heavy metal resistant bacterium *Altererythrobacter atlanticus* 26DY36 T, isolated from deep-sea sediment of the North Atlantic Mid-ocean ridge. *Marine genomics*, 24, 289-292.

Appendix

Table A1: Contains the NCBI protein ID number for all 18 of the analyzed proteins. The links will lead directly to the open source data that was used for this experiment.

	Protein name and NCBI ID number					
species (depth)	NADH dehydrogenase	RNA pol sigma rpoD	DNA primase	triose P isomerase	HSP 33	ferrous iron transport B
***A. atlanticus* (≈2500m)**	AKH43871.1	AKH41478.1	AKH41479.1	AKH41653.1	AKH41653.1	AKH43976.1
***A. epoxidivorans* (≈115m)**	ALE16090.1	ALE16372.1	ALE16373.1	ALE16516.1	ALE17990.1	ALE17414.1
***A. marensis* (≈0m)**	WP_047806393.1	WP_047807685.1	WP_047806004.1	WP_047807739.1	WP_047807012.1	WP_047806822.1