

PALEOPHENOTYPE RECONSTRUCTION OF CARBON FIXATION PROTEINS AS A WINDOW INTO HISTORIC BIOLOGICAL STATES. B. Kacar¹, L. Guy², E. Smith³, J. Baross³, Harvard University, Cambridge, MA, ²Uppsala University, Sweden, ³ Earth-Life Science Institute, Tokyo, Japan and ⁴ University of Washington, Seattle, WA.

Introduction: Earth's past environments were essentially alien worlds, and past organisms may have found different optima than are possible under modern environmental conditions, but it's difficult to infer what those optima are a priori. One way to reconstruct reventant metabolisms and pathways is to seek guidance from the Earth's environmental and biological past by using ancient DNA as a proxy to understand life's origins and early evolution. What can the phenotypes of inferred ancient proteins tell us about the origins of critical metabolic pathways? How can we reconstruct ancient biological functions representing major evolutionary innovations of our planet's alien past? To answer these questions, here we present a new approach to animate ancient DNA (reventant genes). We present two examples, ancient Rubisco [1] and Carbonic anhydrase [2] proteins and how reconstructing past biological components through paleophenotype reconstruction help address questions related to understanding biology in deep time. Reconstructing bacteria by engineering their genomes with ancestral DNA is a fundamentally new methodology with which to study the origins of life. This approach allows reconstructing macroevolutionary phenotypical trends across geologic time to ultimately infer conditions that approach LUCA and possibly life's origins.

Experimental Approach: Reconstruction of ancestral Rubisco proteins: PhyloBot software (<http://www.phylobot.com>) [3] was used for reconstruction. Briefly, orthologs of the Rubisco family were identified by BLAST search based on the 167 amino acid sequences of group IA and IB Rubisco in *Synechococcus elongatus* and group IV 168 Rubisco in *Baccillus*, using the NCBI BLAST Tool. Using 81 curated 169 sequences, multiple sequence alignments were inferred using MSAProbs and MUSCLE with the default settings. Both of these alignments were best-fit by the 171 PROTCATWAG model with model fitness 172 assessed using the Akaike Information Criterion.

Reconstruction of Carbonic Anhydrase proteins: Representative for different taxonomic groups were selected using the software phyloSkeleton (available at <https://bitbucket.org/lionelguy/phyloskeleton>) as follows: one for each species of *Streptococcus*, one for each family in Proteobacteria and Cyanobacteria, and one for each order otherwise. After manual curation of

poorly classified genomes, 388 genomes were retained. RAxML 8.2.8 with the PROTCATLG model was used to infer the phylogenetic tree depicted in Figure 2. A hundred parametric bootstraps were drawn to estimate branch reliability. Trees were visualized and edited with FigTree 1.4.3 (Andrew Rambaut, available from <http://tree.bio.ed.ac.uk/software/figtree/>).

Results: Reconstructed carbonic anhydrase results demonstrate that events of horizontal gene transfer in an evolutionary tree for a given gene need to be recognized prior to paleophenotype reconstruction. The Rubisco tree is highly supported and recapitulates organismal phylogeny. Our analysis of amino acid substitution enrichment along the selected phylogenetic transect suggests possible relationships between the geochemical history of the Earth and the mutational history of Rubisco protein.

References:

- [1] Kacar B, Hanson-Smith V, Adam ZR, Boekelheide N Constraining the timing of the Great Oxidation Event within the Rubisco Phylogenetic Tree. *Geobiology* (in revision)
- [2] Kacar B, Guy L, Smith E, Baross J Paleophenotype reconstruction of carbon fixation proteins as a window into historic biological states. *Proc Roy Soc A* (in revision)
- [3] Hanson-Smith V, Johnson A PhyloBot: A Web Portal for Automated Phylogenetics, Ancestral Sequence Reconstruction, and Exploration of Mutational Trajectories. *PLoS Comput Biol.* 2016, 29;12(7):e1004976