

**PHYLOGENOMICS OF COLD ADAPTATION IN BACTERIA AND ARCHAEA.** R. E. Collins<sup>1</sup>, J. Goordial<sup>2</sup>, I. Raymond-Bouchard<sup>2</sup>, Yevgen Zolotarov<sup>2</sup>, Martina Stromvik<sup>2</sup>, L. Whyte<sup>2</sup>, <sup>1</sup>Institute of Marine Science, University of Alaska Fairbanks, 905 N Koyukuk Dr, Fairbanks, Alaska, USA, <sup>2</sup>Department of Natural Resource Sciences, Macdonald Campus of McGill University, 21111 Lakeshore, Ste. Anne de Bellevue, Quebec, Canada.

**Introduction:** Extremophiles that grow at sub-zero temperatures are found widely throughout the tree of life, and the adaptations used to survive low temperature vary among these different groups. A paucity of genomic data on psychrophiles, cryophiles and their mesophilic relatives has to date made it difficult or impossible to assess the generality of some commonly identified molecular adaptations to low temperature. Yet it is vitally important to compare closely related genomes in order to make robust assertions about the importance of various potential mechanisms of cold adaptation.

Here we compare two genomic pipelines for assessing cold adaptation using various proposed indices at the amino acid level, including the ratio of arginine to lysine residues, the frequency of acidic residues, aliphatic residues, and proline residues, the aliphatic index, and the Grand Average of Hydropathicity (GRAVY). Additional gene family information is included to investigate the abundance and distribution of clusters of genes that may play a role in cold adaptation.

The first pipeline uses as its base the HIMA database (<http://cryomics.org/hima>), a manually-curated set of metadata that includes information on genomes of microbes from low temperature environments. The closest relatives of database members with complete genome sequences were included as non-cold-adapted reference sequences. A phylogenetic tree of concatenated core proteins from each genome was inferred using PhyML. All-versus-all BLAST scores were calculated for all of the genes in all of the genomes, which were then clustered by single-linkage clustering into gene families. The indices of adaptation were calculated individually for each translated gene in each family.

To adjust for variability among widely divergent taxonomic groups, the adaptation index scores were normalized by the phylogenetic distance-weighted of all of the other genes in the gene family. This results in an index of 'surprise' that suggests whether the score is 'unusual' relative to its nearest neighbors. The figure at the bottom shows a 10kb region of several *Shewanella* genomes, with a number of 'cold' adapted genes in blue, and 'non-cold' adapted genes in red.

The second pipeline is focused on closely related genomes. The data set consists of the genomes of a number of bacteria isolated at sub-zero temperatures from permafrost. For each cryophile genome, the genomes of the 3-14 closest mesophilic relatives were identified in public databases for comparison using the same set of adaptation indices. In this pipeline, proteins from each cryophilic strain were compared to their mesophilic relatives using stand alone blastp. The top hit from each mesophile was taken for downstream analyses. All proteins were then assessed for cold adaptation using the indices described above. The results were averaged for the mesophilic proteins and compared using a one sample t-test to the values obtained from the cryophile genome. Cryophilic proteins which were found to be significantly different from the mesophiles were then classified as either cold or hot adapted depending on the direction of the change. Enrichment of cold adapted genes was calculated by determining the ratio of the total number of cold adapted genes to the total number hot adapted genes for a given index in the cryophilic genomes and a chi square ratio test was employed to determine if the differences were significant.

The two complimentary approaches utilized here allows insight into the convergent evolution of cold adaptive traits, as well as identify any taxa specific cold adaptation.

