

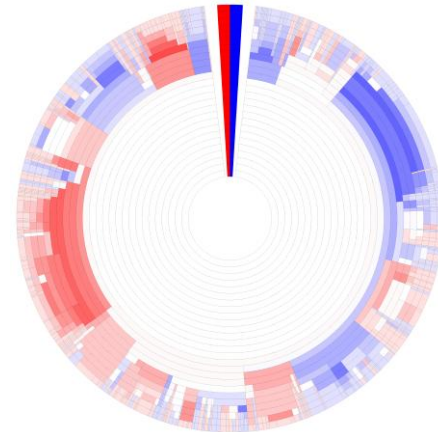
**Elucidating Connections Between Genotypes and Survival Strategies with the Use of Dendritic Heat Maps.**

Matthew Kellom<sup>1</sup> and Jason Raymond<sup>1</sup>, <sup>1</sup>School of Earth and Space Exploration, Arizona State University, Tempe, AZ, USA. matthew.kellom@asu.edu and jason.raymond@asu.edu

**Introduction:** A key advance in microbial ecology in both modern and ancient ecosystems will be connecting genotypic lineages and survival strategies to environmental dynamics. This has become a daunting task in light of the burgeoning repositories of -omics sequence data, calling on entirely new methods for analyzing and visualizing massive biogeochemical datasets [1,2]. The availability of tools for visualization of large-scale sequencing data is lacking. For instance, heat maps, which represent relative abundance of data bins, have been an instrumental part of sequence data visualization [3]. However, heat maps have an inherent problem when it comes to large environmental datasets that have varying levels of biocomplexity. The relative abundance of bins from complex sequence populations can change depending on the level of clustering specificity, defined by the cutoff level, that data is binned and visualized at [4]. Choosing an appropriate specificity level thus is crucial to not obscuring the important features of the data with bins that are too specific or too broad. To better understand data dynamics, we wanted to improve on prevalent figure types in three areas: (1) freedom to scrutinize specific homology levels across a range of biocomplexity, (2) ability to track changes in survival strategies across multiple levels of biocomplexity, and (3) easily visualize how deeply rooted changes in survival strategies are in response to a stressor.

Here, we use dendritic heat maps (DHMs) that simultaneously display multiple heat maps over a range of binning specificities to show the effects of environment on sequence homology and gene expression (Fig. 1). DHMs represent a novel and powerful tool for visualizing correlations in genotypic and phenotypic changes across space and time, and will ultimately help decipher dynamic processes in complex, natural communities, where similarities occur across multitude of scales. We argue that dendritic heat maps will have a significant impact on the study of metatranscriptomes, such as those we are presently obtaining from Yellowstone microbial communities. In this presentation, we ground truth the applicability of DHMs by simulating DNA sequence replication and evolution across large numbers of sequences.

**Figure 1: Representative DHM using simulated data:**



**Results/Discussion:** Simultaneously displaying a range of heat maps that change with specificity level gives a very accurate view of how data is skewed, due to the fact that relative abundance of data bins can change depending on the level of clustering specificity. In DHMs like that of Figure 1, central rings with lower cutoffs would be associated with broad groupings such as phyla or gene superfamilies, while the increasingly strict cutoffs at the peripheral rings would represent more specific identifiers (e.g. same genus/species, gene families/subfamilies).

The aligned configuration of DHM rings preserves the relationships of a dendrogram, where nearby clusters are closer relatives than distant ones. The more central in the diagram a cluster divergence occurs, the more distantly related those clusters are to one another. Combined with the ability to view multiple cutoff levels simultaneously, the dendrogram configuration tracks the specificity levels that relative abundance changes occur. Importantly, tracking changes in relative abundance can be particularly useful for observing the levels at which genotypic divergence (cluster branching) correlates with gene expression (differing heat map skew). Overlaying gene expression on genotype divergence creates a way to visually compare how deeply rooted survival strategies are for an environment, helping to better understand the effects of environment on survival strategies and genotypic lineages.

**References:** [1] Huson *et al.* (2007) *Genome Res*, 17, 377-386. [2] Ondov *et al.* (2011) *BMC Bioinformatics*, 12(385). [3] Wilkinson L. and Friendly M. (2009) *TAS*, 63(2), 179-184. [4] Elser *et al.* (2014) *AEM*, doi: 10.1128/AEM.03160-14.