PHYLOGENETIC RECONSTRUCTION OF THE EARLIEST DIVERGENCES AMONG AMINOACYL-TRNA SYNTHETASE PROTEIN FAMILIES. M. Cantine[1] and G.P. Fournier[1]. [1]Department of Earth, Atmospheric and Planetary Sciences, Massachusetts Institute of Technology, (g4nier@mit.edu)

Introduction: One of the most important and highly conserved parts of the translation system for synthesis of genetically encoded polypeptides are aminoacyl-tRNA synthetases (aaRS), which specifically recognize cognate tRNA and aminoacylate their acceptor stems with the correct activated amino acid. With a few exceptions, there is a 1:1 relationship between aaRS and cognate amino acids, and each aaRS family is found across the entire diversity of life. aaRS come in two major, nonhomologous types, Class I and Class II, with about 10 specific aaRS families found within each class. As such, the diversification of aaRS protein families represent very ancient events in the early history of life, preceding the Last Universal Common Ancestor (LUCA), and possibly in some cases even involved with the evolution of the genetic code itself[1,2]. Several previous studies have attempted to resolve the deep evolutionary history of aaRS paralogs in each class[3,4], a challenge given the limited phylogenetic information within such divergent sequences and the abundance of non-homologous structural subunit diversity, which forces misalignment. This is especially true within Class II aaRS, which, in come cases, retain very little sequence conservation, and even more dramatic structural modifications across families. We attempt to resolve the deep phylogenetic relationships between aaRS families within each class using (1) decomposition of aligned sequences into homologous blocks, preserving well-aligned regions within subsets of aaRS families, while avoiding misalignment of nonhomologous regions between aaRS families; (2) for Class II aaRS, structure-informed alignment of highly divergent sequences to recover deeper signals of site homology; and (3) parsimony-informed rooting of aaRS Classes using the phylogenetic distribution of previously identified conserved motifs across subsets of aaRS families.

Results: Structure-based sequential alignment of Class II aaRS greatly improved alignment quality and the amount of phylogenetically informative sites for tree reconstruction, preventing multiple systemic misalignments that were observed for traditional methods. In both Class I and Class II aaRS, alignment decomposition further avoided misalignment while improving the size and quality of the site character matrixes available for phylogenetic reconstruction. Applying both maximum-likelihood and Bayesian tree reconstruction methods to these improved alignments gave more consistent and well supported deep topologies, more closely constraining the earliest divergences in aaRS history. For the first time, we also report support for a sister relationship between Class I LysRS and ArgRS protein families, suggesting a more ancient origin for positively charged amino acids, and a closer mapping of aaRS evolution and the physiochemical properties of their cognate amino acids. Reduced phylogenetic uncertainty for the deepest relationships within these families, combined with more tightly constrained possible rootings, permit a more accurate and informative reconstruction of the earliest stages of the evolution of the universal translation system.

Broader implications: An improved understanding of the earliest history of aaRS evolution across both Classes provides a nearly unique window into the pre-LUCA evolution of the translation system, in parallel to its other major component, the ribosome. These results also provide critical data for the sequence reconstruction and synthetic resurrection of the earliest aaRS ancestors, raising the possibility of experimental elucidation of their ancient origins and specificities, and the selective process by which they diversified into their current forms.

Acknowledgements: This work was made possible by funding from the Simons Collaboration on the Origins of Life, and NASA Astrobiology Institute *Foundations of Complex Life* (MIT).

References:

[1] Fournier G.P., Alm E.J. (2015) *J Mol Evol., 80,* 171-185. [2] Woese C. R., et al. (2000) *Microbiol Mol Biol Rev. 64,* 202-236. [3] Nagel G. M., Doolittle R.F. (1995) *J Mol Evol., 40,* 487-498. [4] Wolf Y.I et al. (1999) *Genome Res. 9,* 689-710.