

### Interpreting the Interpreters of Life: On the Evolution of Aminoacyl-tRNA Synthetases.

A. Poppinga<sup>1,2</sup>, R. Bouckaert<sup>1,3</sup>, A. J. Drummond<sup>1,2</sup>, and P. R. Wills<sup>4</sup>, <sup>1</sup>Department of Computer Science, University of Auckland, Auckland, New Zealand (alexpopinga@gmail.com, a.poppinga@auckland.ac.nz), <sup>2</sup>Allan Wilson Centre for Molecular Ecology and Evolution, New Zealand (alexei@cs.auckland.ac.nz), <sup>3</sup>Machine Learning Group, University of Waikato, Hamilton, New Zealand (remco@cs.waikato.ac.nz), <sup>4</sup>Department of Physics, University of Auckland, Auckland, New Zealand (p.wills@auckland.ac.nz).

**Introduction:** For genetic information to be preserved and transmitted, a system must yield machinery that carries out the reproductive process above a threshold of error [1]. Aminoacyl-tRNA synthetases (aaRSs) are examples of such machinery [2]. Significantly, major dynamic transitions in the cores of these molecules are observable within their evolutionary histories, which are representable through phylogenetic inference. All aaRSs, of all specificities from all currently known organisms, fall into two main classes. Each class has a core structure that has been highly conserved since the origin of life on Earth 3.8 billion years ago. The aaRSs have evolved from the simplest structures, which provided the basic functional requirements for interpreting a loose binary code, into the complex present day structures that are capable of interpreting the highly specific 61-codon to 20-amino acid universal genetic code. Throughout this process, thermodynamic instabilities drove bifurcation events in the aaRS phylogeny, expanding the amino acid alphabet by one binary digit at each bifurcation. Therefore, in order to analyze the cophylogenies of Class I and Class II aaRSs before the LUCA event, it is necessary to consider epochs in which there were fewer than 20 classes of distinguishable amino acids, invalidating the use of 20 x 20 substitution matrices throughout the phylogenetic inference [3]. To tackle this problem, we have developed an extension of the Bayesian software package BEAST, using matrices of reduced  $2 \leq n \leq 20$  dimensions for each pre-organismal epoch. Results from the phylogenetic analysis are used in complement with empirical structural data to examine the transitions of specificity in these self-constructing genetic interpreters and provide insight into the evolution of functional meaning.

informatic basis of biosemiosis. *Biosemiotics* DOI 10.1007/s12304-013-9196-2.

**References:** [1] Eigen M. (1971) Self-organization of matter and the evolution of biological macromolecules. *Naturwissenschaften* 58: 465—523.

[2] Wills P. R. (2009) Informed Generation: physical origin and biological evolution of genetic codescript interpreters. *J Theor Biol* 257: 345

[3] Wills P. R. (2013) Genetic information, physical interpreters and thermodynamics; the material-