

# A PHYLOGENOMIC STUDY OF THE RIBOSOME UNCOVERS ITS HISTORY, CHALLENGES THE EXISTENCE OF AN RNA WORLD OF REPLICATORS, AND REVEALS A MYSTERIOUS LINK TO LANGUAGE AND MUSIC

G. Caetano-Anollés<sup>1,2\*</sup>, B. Minhas<sup>1</sup>, D. Caetano-Anollés<sup>1</sup>, K. Shahzad<sup>2</sup>, J.E. Mitten-thal<sup>3</sup>. <sup>1</sup>Evolutionary Bioinformatics Laboratory, Department of Crop Sciences, <sup>2</sup>Illinois Informatics Institute, and <sup>3</sup>Department of Cell and Developmental Biology, University of Illinois, Urbana, IL 61801, USA.

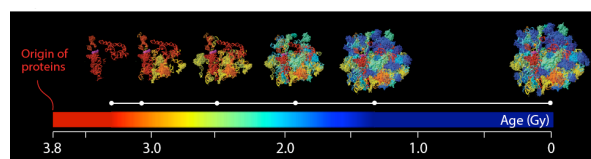
**Background:** Canalization and coevolution are inescapable structuring forces that tailor the makeup of macromolecules and cellular complexes. Here we invoke these principles to find deep phylogenetic imprints in the structure of the ribosome. Analysis of thousands of RNA molecules and millions of proteins reveals the coordinated evolution (coevolution) of ribosomal proteins and RNA components. An “arrow of time” derived from tendencies towards increases in structural domain abundance or decreases in the number of structural conformations defines evolutionary origins without invoking risky assumptions drawn from origins of life models. Phylogenies and timelines of ribosomal history were reconstructed using standard bioinformatics tools of phylogenetic analysis often used in morphometrics [1-4]. The relative ages of structures were derived directly from phylogenetic trees, indexed with structural, functional and molecular contact information, and mapped (by color) onto three-dimensional models of the ribosome (Fig. 1). A molecular clock of folds [5] was used to place relative ages onto the geological record. Coaxial helical stacking patterns and trunk-branch insertion directionalities [6], A-minor interactions [7], tertiary structural motifs, and background knowledge on function and structure supported the detailed phylogenomic-based model of ribosomal evolution. The outcomes of our most recent studies are unexpected and challenge the existence of an ancient RNA world of replicators. Instead, results uncover the gradual accretion of structures and law-like principles operating in language and music.

**Results:** Major findings include (i) Subunit RNA and proteins coevolved tightly, starting with interactions between the oldest proteins (S12, S17, S9, L3) and the oldest rRNA helical segments responsible for decoding and ribosomal dynamics and ending with the rise of a modern multisubunit ribosome; (ii) A major transition in evolution ca. 3.1 billion years ago (Gy) brought evolving ribosomal subunits together by unfolding inter-subunit (bridge) contacts, interactions with full cloverleaf tRNA structures, and tightened intramolecular interactions; (iii) During this transition, a fully fledged peptidyl transferase center (PTC) responsible for protein synthesis appeared by duplication of local helices, supporting an appealing model of PTC origin [8]; (iv) A second evolutionary transition occurred almost concurrently with the “great oxidation

event” (ca. 2.4 Gy) and involved the discovery of the L7/L12 protein complex that stimulates the GTPase activity of elongation factor G and enhances ribosomal efficiency; and (v) The structural accretion process and patterns of molecular growth manifest as trade-offs between economy, flexibility and robustness, with properties that are typical of language-like behavior.

**Conclusions:** The ribosome appeared late in evolution, once genetics was in place but after protein catalysts had already coopted prebiotic chemistries. Our study reveals that the structures and molecular functions of extant ribosomal components hold historical information and mysterious patterns capable of unlocking life’s origins.

**References:** [1] Caetano-Anollés G. (2002) *Nucleic Acids Res.*, 30, 2575-2587. [2] Caetano-Anollés G. et al. (2008) *Front. Biosci.* 13, 5212-5240. [3] Sun F.J. and Caetano-Anollés G. (2009) *J. Mol. Evol.*, 69, 430-443. [4] Harish A. and Caetano-Anollés G. (2012) *PLoS ONE*, 7, e32776. [5] Wang M. et al. (2011) *Mol. Biol. Evol.* 28, 567-582. [6] Petrov A.S. et al. (2014) *Proc. Natl. Acad. Sci. USA*, 111, 10251-10256. [7] Bokov K. and Steinberg S.V. (2009) *Nature*, 457, 977-980. [8] Agmon, I. et al. (2006) *Israel J. Ecol. Evol.* 52, 359-374.



**Fig. 1.** Protein history shows that metabolic enzymes preceded RNA-protein interactions and the ribosome. The rainbow-colored protein timeline was derived from a census of structural domains defined at superfamily level of complexity. The short red segment in the left (~100 million years of evolution) depicts metabolic domain families appearing before the first domains known to interact with RNA. The top scale represents a timeline of ribosomal accretion. The first dot of the timeline shows that the ribosome emerged late in protein evolution. The second dot signals the major transition in ribosomal evolution that is responsible for modern protein biosynthesis. The third dot indicates the second transition that occurred almost concurrently with the great oxidation event of the planet.