

Summer Project: Notoriously difficult early animals phylogeny **Dissecting important steps in the analysis**

Project description: Phylogeny is the research of relationships between organisms. It is based on the notion that life on Earth originated from a common ancestor, therefore we can reconstruct the relationships based on various pieces of evidence. The revolution of molecular and computational biology allows both generating and analyzing vast datasets sampled from the diversity of life on Earth. We can even attempt to specify the timing of evolutionary events: based on the idea of a molecular clock and using fossils to calibrate it. This summer project will focus on one of the most difficult phylogenetic questions: relationships between Metazoa (animals).

Why is it important? Resolving this tree is crucial for understanding such fundamental questions as the evolution of the nervous system. Researchers are not sure about that: heated dispute on whether Ctenophores or sponges are the deepest branching lineage is still ongoing! Interestingly, while the branching order is controversial, the timing of the diversification consistently points to deeper origin than is classically accepted in paleontology. Should one be looking for fossils in a different place?

The aim of this project is gathering data that will contribute to resolving these questions. Phylogenetic endeavors are full of methodological pitfalls that make striving for reproducibility and transparency all the more important. We will tackle these problems using lessons learned from resolving the Tree of Life (Zaremba-Niedzwiedzka et al. Nature. 2017; 541(7637) Asgard archaea illuminate the origin of eukaryotic cellular complexity).

Specifically our starting point is available phylogenetic datasets, conveniently gathered by Li et al in a github repository (Mol Biol Evol. 2021 38(10), Rooting the Animal Tree of Life). We will inspect this data closely (looking at alignments, trimming, checking orthology, taxonomic coverage, sequence biases, contribution of phylogenetic signal from single genes) and build a bait in the form of profiles of the best taxonomic markers to be able to screen environmental datasets. Is it possible that organisms affiliated with the troublesome notorious long branches (Ctenophora, Placozoa) are lurking in various environments? I want to be able to explore existing datasets for possible Metazoa environmental DNA (eDNA) that could have been obtained in projects not necessarily focused on Metazoa diversity.

Necessary bioinformatic background: working with Linux tools, no specific phylogenetic experience required;

Available supervision: I will supervise the student personally and provide training in non-trivial phylogenetic analysis and interdisciplinary approaches to reconstructing evolution. Specifically I am available all weeks except 30: on campus (weeks 23-25, 28-29, 31-35) for introductory period and setting up the analysis, as well as discussing and summarizing the results at the end, and online (weeks 26-27, 29) to consult on daily basis the analysis driven by the student. This project is stand-alone but connected to analysis of molecular dating of early animal evolution and inspecting sources of bias in the Metazoa phylogenies. Following student's interest there are possibilities to present results at the Molecular Evolution department and continue this research.

Project duration: 6-8 weeks in the period between 2022-06-13 and 2022-08-28.