Project Summary

Reconstructing the phylogeny of all species has been a grand challenge in biology since Darwin. Recent years have seen great progress in the resolution of many significant clades, and these efforts have produced dramatic evolutionary insights. However, we still lack a comprehensive synthesis of the entire tree of life. Synthesis is currently inhibited by limits of available data, analytical power, and informatics infrastructure. Perhaps more importantly, it is also limited by a lack of compelling means and incentives for community participation. As a result, most phylogenetic knowledge resides as figures in journal articles rather than digital objects in databases. A comprehensive synthesis would yield great benefits across the life sciences, especially if it were self-sustaining, community-driven, and continually updated. We therefore propose to 1) within one year, create the first comprehensive draft tree of life by synthesizing existing phylogenetic and taxonomic knowledge; 2) enable the community to improve, annotate, and expand this tree; 3) initiate a cultural transformation in systematics towards pervasive and ingrained practices of data sharing; and 4) develop novel methods for synthetic tree reconstruction.

Intellectual merit: Systematics is a pillar in the foundation of biological science, tasked with providing a complete account of the origins of all species. This project addresses important empirical and theoretical needs in pursuit of this goal. By assembling a first draft of a comprehensive tree of life that is accessible to both public and scientific audiences, we will stimulate ongoing synthesis of phylogeny through automated and community-driven means. This will be facilitated by our plan to develop novel methods for building large-scale synthetic trees, for analysis and visualization of combined tree and network structures, for estimating incongruence throughout the tree, for incorporating new data into existing analyses, and for identifying gaps in our phylogenetic knowledge. We will build these methods into an open-source software platform that implements a core set of synthesis functions, and a core set of tools designed to incentivize data-sharing by improving the efficiency and productivity of phylogenetic workflows. These include tools for creating semantically enriched, publication-quality tree illustrations, as well as for "one-click" data submission to public archives (e.g., TreeBASE and Dryad).

Broader impacts: The tree of life is a highly compelling metaphor for non-scientists and scientists alike. Research and education across all fields of biology will benefit in fundamental ways from a tree that is easily explored, queried, and downloaded for study. Such a resource will provide a new lens through which to identify and assess global biodiversity and interpret broad-scale patterns and processes of evolution. In fields such as ecology, where phylogeny is being increasingly integrated into community studies, this comprehensive tree will be a central resource for determining evolutionary relationships. It may profoundly accelerate the pace of species discovery by providing a common framework in which to place new taxa. Our series of workshops will ensure engagement of both systematists and the wider scientific community. The three graduate students, ten postdocs, and numerous undergraduates involved in the project will receive diverse research experiences across systematics, bioinformatics, software development, and phylogenetic analysis. Our undergraduate course development will engage an even larger number of students in critical thinking about evolution, biodiversity, and the tree of life. Through programs at our various institutions, we will recruit students from under-served communities. Our public website and social media outreach will engage the general public and K-12 educators. Finally, our software tools and community engagement activities will initiate a transformation of the culture in systematics to one in which data sharing practices are ingrained and broad-scale synthesis is actively pursued.