BioOne COMPLETE

Assembling an Online Tree of Life of Two Million Species

Author: Watanabe, Myrna E.

Source: BioScience, 63(1) : 64

Published By: American Institute of Biological Sciences

URL: https://doi.org/10.1525/bio.2013.63.1.16

BioOne Complete (complete.BioOne.org) is a full-text database of 200 subscribed and open-access titles in the biological, ecological, and environmental sciences published by nonprofit societies, associations, museums, institutions, and presses.

Your use of this PDF, the BioOne Complete website, and all posted and associated content indicates your acceptance of BioOne's Terms of Use, available at www.bioone.org/terms-of-use.

Usage of BioOne Complete content is strictly limited to personal, educational, and non - commercial use. Commercial inquiries or rights and permissions requests should be directed to the individual publisher as copyright holder.

BioOne sees sustainable scholarly publishing as an inherently collaborative enterprise connecting authors, nonprofit publishers, academic institutions, research libraries, and research funders in the common goal of maximizing access to critical research.

Downloaded From: https://complete.bioone.org/journals/BioScience on 29 Mar 2024
Terms of Use: https://complete.bioone.org/terms-of-use

Assembling an Online Tree of Life of Two Million Species

MYRNA E. WATANABE

The Open Tree of Life project is moving forward with a first draft of the tree expected this August. The Open Tree of Life (*http://opentreeoflife.org*) will synthesize all extant phylogenetic trees and will eventually include every known organism, living or extinct, and will leave room for those that are, as yet, unknown.

The project is part of a natural evolution of the National Science Foundation's Assembling the Tree of Life (AToL) program, which funded studies on branches of the tree, such as green algae, Lepidoptera, and other groups of organisms. "I view this as the biodiversity equivalent of the human genome project," said Doug Soltis, professor in the Department of Biology at the University of Florida, Gainesville, who is one of the lead researchers on the project. The goal, explained Keith Crandall, founding director of the Computational Biology Institute of George Washington University, Virginia Science and Technology Campus, in Ashburn, is "to graft together trees from experts on individual groups." The approach, he continued, may be genetic, morphological, or both.

A synthetic tree of life is not new. But, as principal investigator Karen Cranston, informatics projects manager at the National Evolutionary Synthesis Center, at Duke University, said, "doing it through large-scale data integration that allowed for both curation and automated updating had not been done before." In an e-mail, Cranston wrote, "We aim to synthesize the thousands of published phylogenetic trees into a comprehensive tree of all species."

It will be created in open-source software, and the data will have Creative Commons licenses, said Cranston. Crandall said that the project will "allow researchers to not only add their own phylogenetic estimates to the tree but also to calculate statistics and conduct analyses with the tree." It will also be accessible to the public, wrote Stephen Smith, a computational biologist from University of Michigan, who is a leader of the project.

The tree starts with many missing pieces. "We recognize [that] we don't know most of it," explained Laura Katz, professor of biological sciences at Smith College. "Of the approximately two million named species, there are 250,000 on GenBank," Katz noted.

Eleven researchers from 10 institutions are collaborating on the project, including those who study green plants (Soltis), arthropods (Crandall), fungi (David Hibbett, professor of biology at Clark University), and Amoebozoa (Katz). The team includes computational experts and even a graphic artist and journalist, whose job is to help visualize the tree.

To select projects to fund under its Assembling, Visualizing, and Analyzing the Tree of Life (AVAToL) solicitation in August 2011, the NSF hosted an unusual five-day, intensive Ideas Lab workshop "to stimulate transformative approaches to building, visualizing, and analyzing an interactive tree of life," according to the program announcement. Instead of AToL's approach of creating information on "relationships within major groups of organisms," the concept was to take all the information about the tree's branches and place them in context in a comprehensive tree of life for all named species.

Prospective participants submitted two-page preliminary proposals in response, according to Tim Collins, who was one of the NSF program directors for AVAToL and is now professor and chair of the Department of Biological Sciences at Florida International University. An organizational psychologist helped decide who would best work together both cooperatively and competitively in an Ideas Lab.

The selected group included molecular biologists who work on molecular systematics, morphologists, artists and designers from the Rhode Island School of Design, paleontologists, and others with interest in the tree of life. Facilitators from Knowinnovation (KI), which started in the United Kingdom and now has a US presence in Buffalo, New York, had the job to "stir… creative juices" among the selected participants, Collins said. KI has significant experience in doing just that, having organized similar events in the United Kingdom in a process called a *sandpit*. They also managed the Ideas Lab and made sure the process stayed on track.

Over five days, the participants broke into various groupings, all the while jotting down ideas on notecards that were stuck onto a board for comments. Soltis said that he and Hibbett, "both organismal biologists," and Cranston and Rick Ree ("computer folks"), the latter of the Field Museum of Natural History in Chicago, "independently had a similar idea for how we might build a tree of life and engage the community in a Wikipedia-type fashion."

By the final morning, after some ideas were dropped and some groups reorganized, said Collins, "The groups that are left give their final presentation and a short prospectus of what they want to do, who the personnel will be, and a rough idea of the budget." The mentors and program directors soon decided which teams would be invited to submit a full proposal. The Open Tree of Life was one of three teams funded. (The others were Comparative Analysis Workflows for the Tree of Life and Next Generation Phenomics for the Tree of Life.)

The $5.76 million Open Tree of Life project extends for three years.

Myrna E. Watanabe (mew@99main.com) is a science and grant writer in Patterson, New York.

doi:10.1525/bio.2013.63.1.16

Downloaded From: https://complete.bioone.org/journals/BioScience on 29 Mar 2024
Terms of Use: https://complete.bioone.org/terms-of-use