Researchers aim to assemble the tree of life for all 2 million named species

May 21 2012

A new initiative aims to build a grand tree of life that brings together everything scientists know about how all living things are related, from the tiniest bacteria to the tallest tree.

Scientists have been building evolutionary trees for more than 150 years, ever since Charles Darwin drew the first sketches in his notebook. But despite significant progress in fleshing out the major branches of the tree of life, today there is still no central place where researchers can go to browse and download the entire tree.

"Where can you go to see their collective results in one resource? The surprising thing is you can't — at least not yet," said Dr. Karen Cranston of the National Evolutionary Synthesis Center.

But now, thanks to a three-year, $5.76 million grant from the U.S. National Science Foundation, a team of scientists and developers from ten universities aims to make that a reality.

Figuring out how the millions of species on Earth are related to one another isn't just important for pinpointing an aardvark's closest cousins, or determining if hagfish are more closely related to sand dollars or sea squirts. Information about evolutionary relationships has helped scientists identify promising new medicines, develop hardier, higher-yielding crops, and fight infectious diseases such as HIV, anthrax and influenza.

If evolutionary trees are so widely used, why has assembling them across all of life been so hard to achieve? It's not for lack of research, or data. Thanks in large part to advances in DNA sequencing, thousands of new phylogenetic trees are published in scientific journals each year —most of them focused on isolated branches of the tree of life, for everything from birds to botflies.

"There's a firehose of data," said Cranston, principal investigator of the project. "[Over the years] scientists have published tens of thousands of evolutionary trees, but there's been very little work to connect the dots and put them all together into a single resource."

Part of the difficulty lies in the sheer enormity of the task. The largest evolutionary trees built to-date contain roughly 100,000 taxa. Assembling the branches for all two million named species of animals, plants, fungi and microbes — not to mention the countless more still being named or discovered — will require new tools for analyzing large data sets and stitching together vast numbers of published trees.

Another difficulty lies in how scientists typically disseminate their results. A tiny fraction of all evolutionary trees that have been published — researchers estimate a mere 4% —end up in a database in a digital form. Instead, most of that knowledge is locked up in figures in journal articles, as PDFs or other file formats that are impossible for other researchers to download, reanalyze, or merge with new information.

This new initiative — dubbed Open Tree of Life (http://opentreeoflife.org) — aims to change all that.

What makes this project different from previous efforts, the researchers say, is its scope. "This is the first real attempt to put together the entire tree of life," Cranston said.

The team hopes to have a first draft of the complete evolutionary tree — compiled from the evolutionary trees that are already available in existing databases — by August 2013. The first draft that emerges will be far from finished. "There will always be new studies that come out," Cranston said. "There will also be places in the tree where we don't have enough data, or where the data lead to conflicting hypotheses, or where groups of researchers simply disagree."

But with a first draft in hand, scientists will be able to go online and compare their trees to others that have already been published, or download it for further study. They'll also be able to expand the tree, filling in the missing branches and placing newly named or discovered species among their relatives. Eventually, the team's goal is to be able to detect when new trees are published and incorporate them automatically, so that the complete tree can be continuously updated.

If the project is to succeed, one of the biggest challenges will be encouraging more scientists to publish their results in digital form. Growing numbers of scientific journals now require authors to deposit phylogenetic data in a digital database, but many published trees never make it. "We hope to provide infrastructure and tools that will make it easier to do that, such as a more user-friendly interface for submitting data," Cranston said.

"In the long run, we hope this will become the central resource for synthesized phylogenetic data," she added.

Provided by National Evolutionary Synthesis Center (NESCent)

Citation: Researchers aim to assemble the tree of life for all 2 million named species (2012, May 21) retrieved 16 April 2024 from https://phys.org/news/2012-05-aim-tree-life-million-species.html

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.