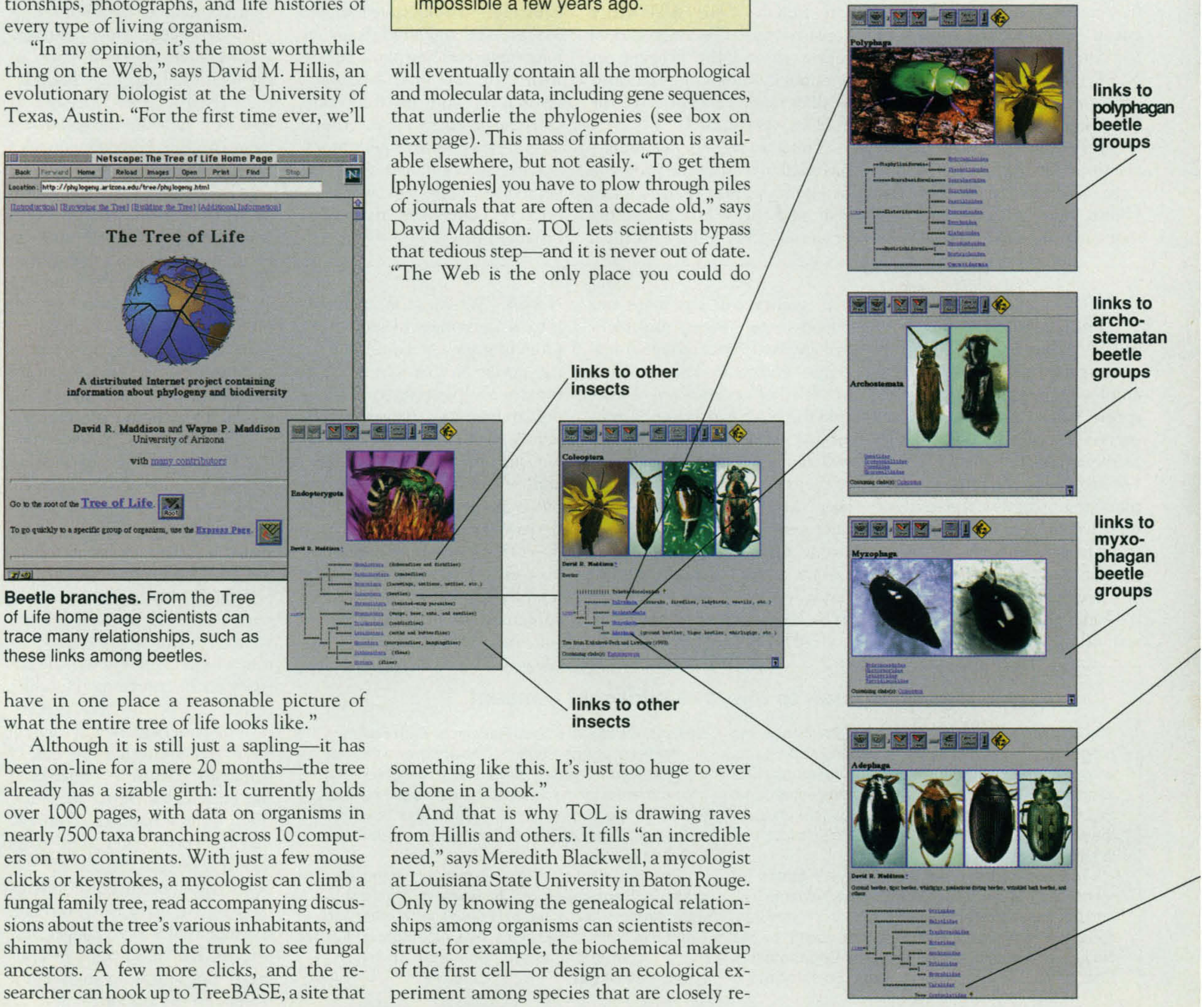
Web-Crawling Up the Tree of Life

Genetic relationships among every type of living organism, an electronic journal, and collaborations—and debates—among far-flung scientists are all coming together at the Tree of Life

The Maddison brothers, David and Wayne, have planted a tree on the World Wide Web, and it is growing faster and further than Jack's proverbial beanstalk. Branches and boughs are spreading far beyond the University of Arizona, Tucson, where both brothers are systematic biologists, to reach institutions around the globe. This prolific plant is nothing less than the Tree of Life (TOL), a vast on-line endeavor that will ultimately include the genetic and evolutionary relationships, photographs, and life histories of every type of living organism.

"In my opinion, it's the most worthwhile thing on the Web," says David M. Hillis, an evolutionary biologist at the University of Texas, Austin. "For the first time ever, we'll have in one place a reasonable picture of what the entire tree of life looks like."


Netscape: The Tree of Life Home Page

Location: http://phylogeny.arizona.edu/tree/phylogeny.html

[Introduction] [Browsing the Tree] [Building the Tree] [Additional Information]

The Tree of Life

A distributed Internet project containing information about phylogeny and biodiversity

David R. Maddison and Wayne P. Maddison
University of Arizona

with many contributors

Go to the root of the Tree of Life.

To go quickly to a specific group of organism, use the Express Page.

Beetle branches. From the Tree of Life home page scientists can trace many relationships, such as these links among beetles.

Although it is still just a sapling—it has been on-line for a mere 20 months—the tree already has a sizable girth: It currently holds over 1000 pages, with data on organisms in nearly 7500 taxa branching across 10 computers on two continents. With just a few mouse clicks or keystrokes, a mycologist can climb a fungal family tree, read accompanying discussions about the tree's various inhabitants, and shimmy back down the trunk to see fungal ancestors. A few more clicks, and the researcher can hook up to TreeBASE, a site that will eventually contain all the morphological and molecular data, including gene sequences, that underlie the phylogenies (see box on next page). This mass of information is available elsewhere, but not easily. "To get them [phylogenies] you have to plow through piles of journals that are often a decade old," says David Maddison. TOL lets scientists bypass that tedious step—and it is never out of date. "The Web is the only place you could do something like this. It's just too huge to ever be done in a book."

Computers '96

A special section on computers in science, including News reports and Articles on Internet congestion, the analysis of DNA and protein sequences, and tools for organizing and analyzing large amounts of biological data, begins on page 585. The News report on this page looks at how the World Wide Web is stimulating a collaboration that would have been impossible a few years ago.

And that is why TOL is drawing raves from Hillis and others. It fills "an incredible need," says Meredith Blackwell, a mycologist at Louisiana State University in Baton Rouge. Only by knowing the genealogical relationships among organisms can scientists reconstruct, for example, the biochemical makeup of the first cell—or design an ecological experiment among species that are closely related. There are practical payoffs, too: Developing broad-spectrum fungicides depends on knowing the relationships of the organisms you're trying to kill. For the scientist with new data, it's a place for rapid publication, and in time, the tree will even be a peer-reviewed electronic journal.

The Maddisons' foray into trees and computers began a decade ago with a computer program they developed, called

TreeBASE: The Roots of Phylogeny

Good trees only grow in good soil—and for phylogenetic trees, whose branches show the evolutionary relationships among organisms, that soil is raw data. Gene sequences, physical characteristics, and other basic information allow researchers to perch organism A on one branch and organism B on another. But that data can be scattered across hundreds of journals—leaving biologists with a daunting search if they want to decide which of several published trees is the best, or if they want to replicate and test a colleague's tree-building techniques. That's why two botanists and their colleagues are cultivating a data tree on the World Wide Web.

Data tree. TreeBASE is on the Web at http://phylogeny.harvard.edu/treebase.

TreeBASE, planted on the Web in 1996 by Michael Donoghue of Harvard University and Michael Sanderson of the University of California (UC), Davis, is a prototype of a vast database that will hold the morphological or genetic data behind every version of every phylogenetic tree ever published, as well as the phylogenies themselves; it will also be linked to the Tree of Life (TOL) (see main text). "You can download the same data—all the genetic sequences, for example—other researchers used to create their trees and play with them to your heart's content," says Donoghue, who hopes that scientists will in time submit their phylogenetic data to TreeBASE as a matter of course when they publish them in a journal. Unlike the TOL, which presents a finished (or close-to-finished) phylogeny, TreeBASE has "all the gory details," says David Maddison, a systematic biologist at the University of Arizona, Tuscon, and a co-creator of the TOL. "It's a great tool for people who are interested in how current hypotheses about phylogenies were arrived at."

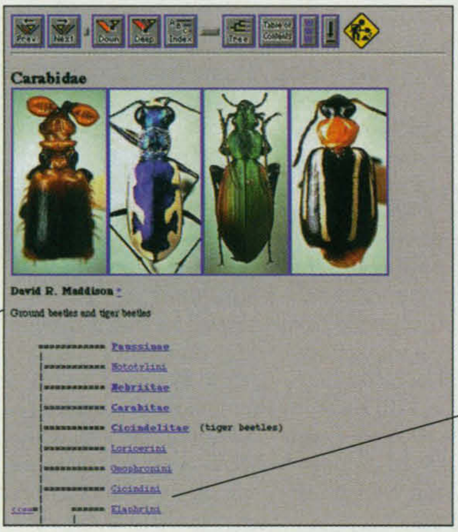
TreeBASE sprouted in 1992, when Sanderson and Donoghue, both then at the University of Arizona—apparently a leading evolutionary-tree farm—developed a stand-alone prototype. "So many phylogenies were—and are—being developed," explains Sanderson. "I think everyone realized something needed to be done to make it easier to access and synthesize these data." The prototype was "very promising," Sanderson says. But the trouble with a stand-alone was that it would only be available to a limited number of people. "So we decided to put it on the Web," says Sanderson.

TreeBASE opens with an introductory page and a query box. Entering the name of an organism, such as that of the plant genus *Piper* (a group that includes black pepper plants), produces a list of the taxonomic names in the database which include that word. You can then click on the name Piperacae, for example, and call up information about three different phylogenetic studies involving this taxon. Other buttons bring up the actual trees and the data behind them, which are stored in computers at Harvard and UC Davis.

The database is still small at the moment. "We've only entered about 155 data sets," says Donoghue, "and most of these are about plants, because that's what we're familiar with." Now the duo is inviting colleagues from other fields to join their effort. And soon, angiosperm browsers on the TOL will merely need to touch a button at the bottom of an organism's page to jump to the corresponding pages in TreeBASE. The ability to till so much phylogenetic data, Donoghue hopes, will let researchers grow many fertile hypotheses. **–V.M.**

MacClade, that enables scientists to trace the evolution of distinctive traits in organisms. "One of the great changes in evolutionary biology over the past 2 decades has been the recognition that phylogeny—that conduit along which genetic information flows—is a critical part of the history of a group of organisms," explains David Maddison.

links to other adephagan families



Carabidae

David R. Maddison

Ground beetles and tiger beetles

Paussinae
Nototylini
Nebriitae
Carabitae
Cicindelitae (tiger beetles)
Loricerini
Omophronini
Cicindini
Elaphrini

links to carabid beetle groups

MacClade lets researchers study the patterns and processes of the evolution of morphological or molecular characteristics in a phylogenetic context. "For example," says David Maddison, "you can enter a phylogeny of vertebrates and information about those that have limbs for walking and those that have limbs for flying. MacClade will reconstruct which of their ancestral lineages had legs and which had wings." The resulting tree diagram would show that flying evolved three different times (pterosaurs, birds, and bats).

David Maddison wondered if he could then bring a variety of these trees together. "I thought it would be great if you could connect phylogenies with hypertext links on the Web, so that you could jump from a tree on vertebrates, for example, to one on mammals," he says. That small vision became the Tree of Life project, with Web pages authored by experts on each organism, and the tree itself distributed among computers around the world. Says Wayne Maddison, "This way, the tree grows at the same time as the science does, and it doesn't depend on how much the two of us know about terrestrial vertebrates, for example."

To link disparate trees together, in the summer of 1994 the brothers added code and tools to MacClade to automatically produce tree pages that are formatted in hypertext markup language, the Web's own dialect. "That way, people who write these pages don't need to know the language of the Web," explains David Maddison. And it didn't take long for other scientists to want to join in. Some asked the Maddison brothers about contributing, and the brothers invited others to participate; today, there are more than 200 contributors. "The project is hierarchical," says David Maddison. For example, he has invited Harvard University's Michael Donoghue to write the opening page on angiosperms; but Donoghue is responsible for contacting other specialists (who, in turn, may contact other experts) to contribute pages on the orders, families, genera, and species of flowering plants. The Maddisons act as editors and overall project coordinators—and do so on their own nickel; the project has not received outside funding.

A single Tree of Life did, however, raise a thorny issue: What to do with different opinions about which organisms go on which branch? And the problem cropped up at the very root of the tree. There are two major versions of what begot what among the five basic groups of organisms—eubacteria, euryarchaeotes, crenarchaeotes (or eocytes), eukaryotes, and viruses. "Figuring out a way to deal with such controversies gave us a nightmare," says David Maddison, "so we decided that wherever such debates exist, we'd include alternative views." And so the root page holds two competing trees, and will also hold defenses of each version, one authored by University of California (UC), Los Angeles, molecular biologist James Lake and the other by an as-yet-unidentified rival—a public debate that Lake welcomes. "It's absolutely crucial," he says, adding that the TOL "is a wonderful forum for airing this type of important question."

By November 1994, the brothers had planted their prototype on an Arizona computer, and 13 months later announced it as officially germinated.* And while it has shot up since then, the tree is far from being fully leafed out. "There are some gaping holes," says David Maddison. "For instance, mammals have not yet made their appearance," although their pages are being developed.

A user climbing the tree for the first time gets some basic information about the tree and directions for navigating among its branches and leaves. From here, browsers have several choices: They can view sample pages, take express routes to specific organisms, or go to the root page.

From the root, tree-climbers can head into any one of the five major divisions by clicking on its name. For instance, touching "Eukaryotes" takes you to the tree showing the major taxa of protists, plants, fungi, and animals. Browsers can jump from here to the branch for vertebrates, and then with another click leap to one of the tree's more complete pages, that of terrestrial vertebrates. Here, Michel Laurin, a paleontologist at UC Berkeley, discusses his view of this group's origins, along with some alternate phylogenies. Touching the names of particular genera takes users into the tree's twigs; and these, in turn, lead to the individual species, or leaves. The tree even has "lichen"—explanatory pages that grow on a branch or tree. Thus, on David Maddison's page about the beetle *Bembidion litorale* and its relatives, browsers can call up photos of all the species and an explanation of their key identifying marks.

Other researchers quickly recognized the tree's merits, particularly the instant access it provides to the most recent phylogenies. "In my field—the evolution of fungi—year-old phylogenetic data aren't any good, and that's usually what you've got in journals," says John Taylor, a mycologist at Berkeley and a TOL author. He notes that his area is particularly hot since mycologists discovered that nucleic acid sequences are far better at resolving issues of relationships than is morphology, and "this is the best and fastest way to publish this data."

Louisiana State's Blackwell, another tree contributor, adds that "I study insect-dispersed fungi where there is a lot of convergent evolution. It's very important to be able to look somewhere and see that the agents of Dutch elm disease and oak wilt are not related, although they were once thought to be." That kind of knowledge, she adds, has practical as well as theoretical implications: "If you're trying to develop a fungicide for these species, you want to know their evolutionary history." Knowing that two species look alike but don't share a lot of genes makes a great difference when developing such a product, she explains.

For still other researchers, the TOL has allowed them to celebrate the little corner of life that they study—and share it in ways previously unimaginable. "I work on a group of poorly known beetles, the Ptiliidae," says W. Eugene Hall, an entomologist at the University of Arizona, Tucson. Only three other researchers in the world work on these beetles—although Hall suspects more would if they knew how "fascinating" they are. And, in fact, his ptiliid page on the TOL, which is jammed with tidbits about odd variants of traits within one species and the males' giant sperm, has sparked a lot of interest. "I've had requests from all kinds of people," he says, including a scientist in South Africa who needed help identifying a ptiliid he had found in a cave.

This near-instantaneous collegial feedback is a big plus for many TOL users and authors. "The greatest advantage is its rapid publication of ongoing and recent research that's up to date," says Berkeley's Laurin, who notes that shortly after posting his pages he received comments from colleagues in Australia and England that helped him refine his ideas.

The Maddisons, in fact, hope that in time their tree will become an electronic peer-reviewed journal. "Some of the data on catfish and jumping spiders appear here for the first time, so it already is primary literature in that sense," says David Maddison. They plan to enlist a board of editors from among their contributors to work out mechanisms for review. Recognizing that contributors would like to have their pages cited, they are also investigating how this can best be done. Currently, pages that do not bear the "under construction" symbol can be cited, but they have not yet dealt with the tree's dynamic nature and how to archive older versions of phylogenies or discussions about them.

But that is all part of TOL's future growth, and it does have a lot of growing to do. "How many millions of organisms are there?" asks David Maddison. "We can't even say we've scratched the surface."

–Virginia Morell

---

* http://phylogeny.arizona.edu/tree/phylogeny.html

GERMAN GENOME PROGRAM

The Right Mix of Form and Function

HEIDELBERG—Last summer, Germany's science minister, Jürgen Rüttgers, made an announcement that many of the country's researchers thought they would never hear: The government would launch a national genome research program, he said, with $130 million by 1999 and more to come later. Convincing the government to take this step took 10 years of lobbying by a few persistent scientists, in a country famous for its fierce public opposition to genetic engineering (*Science*, 16 June 1995, p. 1556). One year and many hard discussions later, with the first projects chosen, the outlines of the enterprise are emerging—although the debate on how to divide up the funds may not be over.

Rüttgers's announcement posed a tough question: How should Germany enter a fast-moving field in which it lags far behind? One camp, which includes some international advisers to the German project, argued that Germany should shoulder its share of high-cost mapping and sequencing efforts already being pursued by other countries. Others, however, felt that Germany had already missed this boat and should instead concentrate on ways to get at the functions of the human genes being uncovered by research and sequencing efforts.

Now, after months of tough deliberations, the program appears to be steering a middle course. The first round of grant proposals was reviewed by the international advisory committee in early March, and the science ministry is now finalizing decisions based on these reviews. Ministry officials say that funds will be split roughly evenly between sequencing, functional studies, and infrastructure development. Although a complete list of projects has not yet been released, committee chair Ernst-Ludwig Winnacker of the University of Munich's Gene Center has provided *Science* with an overview.

The ministry's evenhandedness is unlikely to end the debate over the direction of the program, however. So far they have committed only $50 million, just over one third of the promised total for the first phase, leaving plenty of room for controversy to flare up again. And, although many researchers still have an incomplete picture of the whole program, preliminary reactions are mixed. Some who spoke with *Science* worry that $16 million spread over many functional projects—some potentially very large—