

## NEWS

# Modernizing the Tree of Life

A new generation of systematists seeks to transform its field with the tactics of big science

Bar codes have revolutionized the retail industry, allowing scanners to instantly identify and price everything from beans to beach balls—and allowing retailers to skip the labor-intensive step of having a person actually examine the product. Biologist Paul Hebert of the University of Guelph, Ontario, hopes to adapt that concept to an even more ambitious task: species identification.

Just as cashiers no longer need to know which brand of beans is on sale or whether they are looking at white, brown, or confectioner's sugar, Hebert envisions handheld scanners that automatically read a DNA "bar code"—which conveys differences in the sequence of a single gene—to identify species. By relying on a mitochondrial gene called cytochrome oxidase I to create such bar codes, he claims, "we can make a very efficient engine that would be able to take us through animal life," from barnacles to butterflies.

To experts who have struggled for decades to sort out the identities of closely related organisms, this is a wild claim indeed. "I sputtered about [bar codes] when I saw the first suggestion," recalls invertebrate systematist Frederick Schram of the Institute for Biodiversity and Ecosystem Dynamics at the University of Amsterdam. Adds Hebert, "We were being lampooned in taxonomic corners around the world."

Many researchers doubt that a single gene can resolve all animals into individual species. And many systematists have an almost visceral distaste for the notion of bar-coding animals. But a few, including Schram, are beginning to think Hebert may be on the right track. "As a way of cataloging biodiversity, why not?" says Schram. Even the skeptics are beginning to recognize the need for this sort of broad-brush approach to identifying and classifying great swaths of biological diversity, given the sheer magnitude of the task ahead. Although biologists have cataloged 1.7 million species, they know they have just begun: Estimates of the total number of species on the planet range from 4 million to 100 million.

Naming them all is merely the first step. Researchers need to know where each creature fits into the grand scheme of evolution, from the first single-celled microbe to complicated plants and animals. This scheme is often described as the tree of life, a metaphor proposed 150 years ago by Ger-



**Identification made easy.** These skipper butterflies look similar, but a DNA "bar code" may be able to separate them into two species.

man biologist Ernst Haeckel. Every life form, from krill to whales, seaweed to sequoias, protists to pachyderms, fits somewhere along this tree's twigs and branches. But where?

To find out, more and more biologists are proposing bold new methods that are transforming the practice of taxonomy and systematics. Many of those who find Hebert's bar codes farfetched are impressed with an even more visionary scheme to analyze several genes at once and so allow amateur biologists to both identify an organism and see its evolutionary relationships with the touch of one button. Scores of other researchers are banding together in large collaborative projects and building "super-trees" with data from many studies. Some hope to see their field take on the trappings of a high-tech, big-science endeavor, one in which they tackle the whole tree of life—or at least large chunks of it—at once, in much the same way geneticists now tackle whole genomes. Plant systematists have already embraced such tactics, with notable results (see sidebar on p. 1696).

Not long ago, such ambitions were rare in a field accustomed to much more narrow objectives. But big-science projects are becoming the norm: In 2002, the National Science Foundation (NSF), the chief funder of systematics in the United States, gave out more than \$15 million for a half-dozen multi-investigator projects, each tackling a different group, from bacteria to plants to birds. "Over the last 10 to 15 years, researchers in the systematics community have been working on the small branches on the tree of life," explains NSF plant systematist James Rodman, who runs this program. "We thought it was time to take a more coordinated approach."

Overall, says Michael Donoghue, a botanist at Yale University in New Haven, Connecticut, "we have made good progress in outlining the major groups." It's well accepted, for example, that placental mammals share a common ancestor and that birds are more closely related to lizards than to fish. "But there are 1.7 million species [identified], and we are not anywhere close to putting them all on the tree of life," Donoghue says. Researchers estimate that only about 80,000 species now have a place on their trees.

Large-scale endeavors are increasingly necessary, and not just because we're curious to know who else inhabits our planet. The urgent call for systematic information is now coming from several fields at once. "There are a lot of practical questions" that rely on taxonomy, notes Fred Grassle, a marine ecologist at Rutgers University in New Brunswick, New Jersey. Conservationists need to know, for example, whether the Florida scrub jay is really a species—and therefore subject to protection—or just a subspecies. Likewise genomicists trying to interpret the burgeoning array of DNA sequence data are demanding to know where those sequences fit in the evolutionary scheme. The evolutionary distance between, say, zebrafish and nematodes affects what biologists look for as they compare genomes, and it helps track the evolution of particular genes. Even biomedical researchers are embracing taxonomy to help them understand how pathogens become more virulent over time or how new diseases emerge.

Thus taxonomists are finding their science in greater demand than ever before and

in use in far corners of biology. “You pick up any biological journal—it doesn’t matter what field it is—and it will have phylogenetic data,” says David Hillis, an evolutionary biologist at the University of Texas (UT), Austin.

### From morphology to molecules

When Carolus Linnaeus set out to catalog all organisms, he had little idea what he and his future colleagues were in for. Even so, he came up with a system for naming species and set up a classification system that has survived for 250 years and in some quarters is still going strong.

Until recently, it was enough for budding taxonomists to pick a particular group—dung beetles, canines, oaks—and spend their lives gathering and examining specimens. Researchers pored over organisms’ structures, from obvious traits such as the number of legs to subtleties such as the relative heights of spines in a fish’s anal fin. Their quest: to find an effective way to compare creatures within a group, first to see how they differed and later to determine how they were related to each other.

Such taxonomic work has advanced over the centuries, but it moves too slowly for some. Many organisms are scantily described and poorly illustrated in the literature, so systematists must hunt down original specimens in museums in order to compare them to new specimens, complains Scott Miller, an entomologist at the Smithsonian Institution’s National Museum of Natural History (NMNH) in Washington, D.C. If a species is new, researchers spend months or even years painstakingly describing and illustrating it. This workload means that each of the world’s 6000 to 15,000 traditional taxonomists is lucky to describe 250 organisms in a career. At that pace, a complete tree of life is centuries away, says Terry L. Erwin, another NMNH entomologist.

When molecular techniques spread through biology, they promised to circumvent these problems. Taxonomists found they could cover more ground by looking at differences in DNA and ordering species along the tree accordingly. But their results often clashed with traditional morphological classifications. To make matters worse, many early DNA studies were flawed because they were based on an unreliable stretch of DNA or inappropriate organisms, says Andrew Smith, a systematist at the

Natural History Museum in London.

Fortunately, better statistical methods, a better understanding of DNA changes, and cheaper, more accurate sequencing have boosted the credibility of molecular studies. Researchers now use more DNA and go beyond simple differences in bases, assessing small deletions and insertions and other features of the genome’s landscape.

All the same, “molecular workers are realizing that their data sets are not ‘silver bullets’ for revealing the history of life,” says David Lindberg, a systematist at the University of California (UC), Berkeley. Meanwhile, “morphological workers are realizing that their long-held ideas need to be reexamined. The field is ripe for revision and conversion,” he says. Morphologists are invigorating their work with techniques such as electron microscopy and incorporating a broad range of traits, including behavioral differences. This, plus the molecular challenge, has helped many morphologists accept new ways of looking at organisms, notes Michael Lee, a reptile expert at the University of Queensland, Australia.

As a result, the two sides are finding more common ground. For example, in 1999, Tim

morphologists protested, but now, several molecular studies later, “the story is effectively resolved,” Littlewood says. These days, he says, “when incongruities [between molecular and morphological data] exist, we choose to see this as an area requiring additional data, not a barrier to resolving the tree of life.”

Many controversies remain, from the relationship among bacteria, archaea, and eukaryotes to whether the springtail is a primitive insect. These battles get the headlines and the journal pages. But in most cases, the different data point to the same tree, says Lee: “It should be stressed that [these data] agree more often than they disagree.”

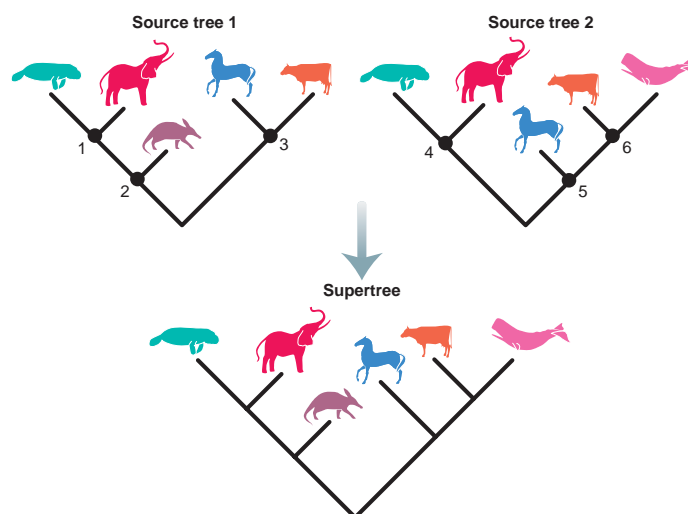
### Banding together

Molecular and morphological types may have achieved détente, but they still have only a rough outline of the evolutionary branching pattern of many groups. For many systematists, that’s not enough. To fill in the tree of life, says NSF’s Rodman, researchers need to figure out how to work together.

Too often, two systematists will take on the same organisms while other taxa go unstudied. And even those working on similar or related groups often lack the benefit of one another’s data, so no one else can take full advantage of their labors, says amphibian expert David Cannatella of UT Austin. That’s why NSF is funding large projects, and why people such as Cannatella are lobbying for joint tree-of-life efforts. “We’d like to get people to actively collaborate and share samples and tissue resources” to build an amphibian tree, he says.

Such collaboration is the first step toward synthesizing systematic information. But it demands a change in the field’s culture. “Systematists have a tradition of working alone,” says Cannatella. “But I think getting together and pooling what we have is the way to go.” Given the biodiversity crisis and the increased use of systematics data by other types

of biologists, he and others would like to see their field take a lesson from the recently completed Human Genome Project. More than a dozen laboratories throughout the world banded together, sharing technological innovations and freely contributing human and other sequence data to a centralized public archive called GenBank. Systematists need such a central archive, says Yale’s Donoghue. “What I think is critical is databasing everything in such a way that it’s accessible,” he says. Then, notes Charles O’Kelly, a systematic biologist at the



**Tree plus tree.** Supertrees combine two or more trees into a single big one by making use of species overlaps.

Littlewood of the Natural History Museum in London analyzed the DNA of a marine worm called the acael, which morphologists had lumped with the parasitic flatworms in part because both have a simple body plan. Littlewood instead proposed that although flatworm morphology evolved from a more complex body plan (as is common in parasites), the acael is primitive: Its body has always been simple. That makes this humble worm a relative of all bilaterally symmetrical organisms, from millipedes to humans (*Science*, 19 March 1999, p. 1823). At first

Bigelow Laboratory for Ocean Sciences in West Boothbay Harbor, Maine, “one can ask about studies at any level. ... It’s a way that we can make the [research] go much faster.”

There are already some attempts at coordinated data sharing. For several years now, plant and animal biologists have been pooling molecular data and submitting both the DNA results and phylogenetic trees to a Web site called TreeBASE ([www.treebase.org](http://www.treebase.org)); more and more journals are requiring authors to deposit their phylogenetic data there. Maureen O’Leary of the State University of New York, Stony Brook, and colleagues are planning MorphoBank, a GenBank-like repository that stores digital images.

These kinds of archives, if they take off, will allow researchers to put together all the data at hand. The challenge would then be to analyze the giant data sets that result, notes Hillis. Mathematically minded biologists are

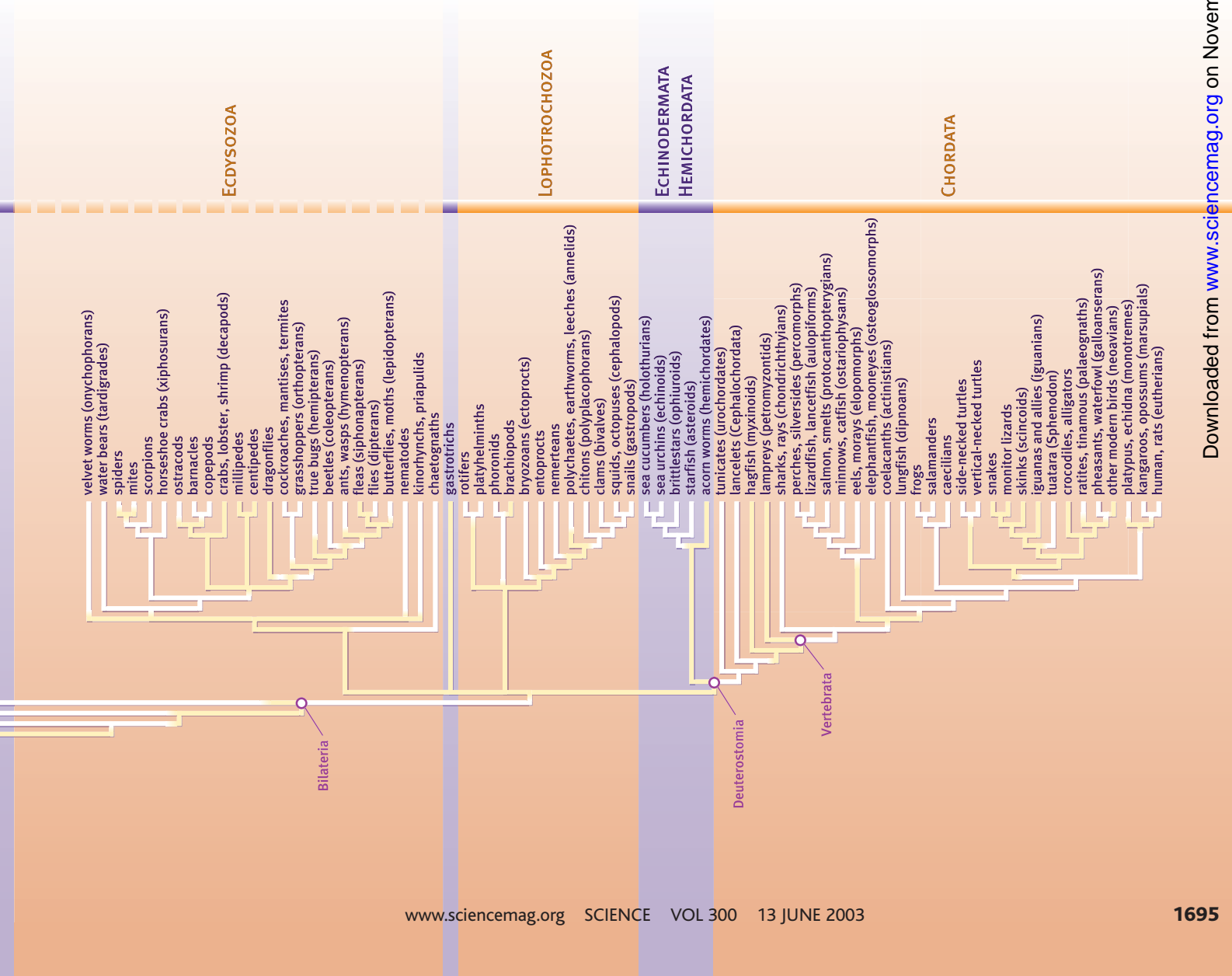
already busy making computer programs to meet the demand.

One approach is to build a supertree, in which a computer program compiles smaller trees of life into a gigantic one, merging branches of the trees without reanalyzing the original data. “There have been any number of areas in biology, especially evolutionary biology, that have really wanted large phylogenies at the species level,” says John Gittleman, an evolutionary biologist at the University of Virginia, Charlottesville. “What we need is the complete tree.”

Supertrees fulfill that need, and they are gaining momentum as more phylogenies become available online. Gittleman estimates that a few years ago, only five researchers were braving supertree analysis; now there are at least 100. “If we really want to get a big picture of the tree of life—of say a million species—then we really need a method of piecing the snippets together,” says

Donoghue. A supertree analysis looks among smaller trees—which might use either morphological or molecular data—for a few species in common, then it uses the overlap to come up with a bigger branching pattern. Supertrees “provide another line of insight into what can be very intimidating mountains of data,” says the University of Amsterdam’s Schram.

Supertrees have yet to incorporate as many species as Donoghue envisions, but even when small, these trees can highlight which groups are well studied and which have conflicting phylogenies. For example, one supertree of mammals, which synthesized data from 315 papers and 400 phylogenetic trees, supported most of the existing mammalian classifications but questioned the relationship between megabats and microbats and also the ancestry of insectivores (*Science*, 2 March 2001, p. 1786).





## Plants Find Their Places on the Tree of Life

Researchers trying to piece together the tree of animal life are hacking through dense foliage, barely able to see the top branches, never mind the distant twigs (see main text). But their colleagues studying plants have many of their phylogenetic trees neatly pruned and manicured. Whereas the animal and microbial types are wrestling with new techniques and beginning to talk about collaborations, botanists have already embraced the culture and methods of big science.

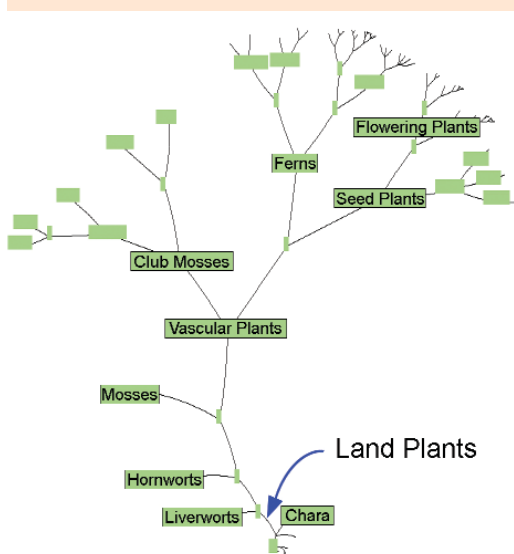
Over the past decade, 200 plant taxonomists from a dozen countries have been analyzing and refiguring the evolutionary history of their favorite flora in an effort called Deep Green. For other systematists, the endeavor has become one to aspire to (*Science*, 13 August 1999, p. 990). "The plant people have made major advances as far as I can tell" and are moving faster than animal-centric researchers, says Frederick Schram, a barnacle expert at the University of Amsterdam.

By coming together, Deep Green researchers were able to identify poorly studied groups and holes in the data. They then parceled out the work to fill those holes. Although it sounds simple, Deep Green depended on the vision of several systematists who rallied their colleagues, says James Rodman, a plant systematist at the U.S. National Science Foundation (NSF). Others say the plant people are succeeding because their field and their trees—filled with a mere 300,000

species—are smaller. As a result, it seems "that the plant people have a good handle on all sorts of data, almost to the point of being truly comprehensive," says John Gittleman, an evolutionary biologist at the University of Virginia, Charlottesville.

For the next step, the Deep Green researchers have been busy figuring out the best way to combine their data into that one tree—"our most accurate representation of the history of green plants," says Charles O'Kelly, a systematic biologist at the Bigelow Laboratory for Ocean Sciences in West Boothbay Harbor, Maine. The work of wrestling their very large data sets into a single tree is part of a 5-year, \$2.7 million NSF grant involving six U.S. teams. At the same time, there's an increased push to put all these data into public databases.

Thus for plant taxonomists the next 5 years promise to be a data gold rush. Several other projects, independently funded but interconnected, are delving into less well-covered aspects of the field. One, called Deep Gene, will help plant experts make use



**Green and growing.** On the Deep Green Web site ([ucjeps.berkeley.edu/TreofLife/hyperbolic.php](http://ucjeps.berkeley.edu/TreofLife/hyperbolic.php)), this tree of plants gets more detailed with a click of the mouse.

of plant genomics information and vice versa. "Deep Time" researchers will blend plant fossil finds with modern botany. And "Deepest Green" delves into the base of the plant tree to sort out the relationships among the green algae.

Such a multipronged, coordinated assault makes some animal systematists green with envy. "I think that the animal [researchers] would profit by a similar approach rather than the winner-take-all rodeo that seems to prevail right now in zoology," says Schram. If his colleagues' competitive spirits could be converted to cooperation, he says, "in a few years, we could achieve wonders."

—E. P.

A supertree's overview may also help make sense of a particular family tree. For example, for 50 years, biologists have assumed that foxes and weasels—diminutive carnivores—split into an unusually large number of species because their small size lets them exploit more parts of their environments, says Gittleman. But a supertree analysis "found very little support for this hypothesis," he says. The most diverse groups—those with the densest "foliage"—did not have the smallest mammals. Instead,

he theorizes that faster rates of reproduction account for the species-richness of these animals, an idea that has yet to be tested.

But this method, too, has its detractors. "Supertrees don't have much value," says Ward Wheeler, an invertebrate systematist at the American Museum of Natural History in New York City. He sees them as summaries of summaries, two steps removed from real data.

That's why some researchers are using another meta-analysis approach that digs

even deeper than the supertree. Called the supermatrix, these analyses merge the original data from a host of smaller trees. The supermatrix may consider 100 or more species on one axis and upward of 10,000 or more traits on the other; the information is usually gathered from previous studies and includes both morphological and molecular traits. A computer then looks for patterns in all these data that reveal the relationships among species. "It's a way of incorporating all the information into one picture of evolution and diversification," says Wheeler.

### Brave new taxonomy

All this work on very large data sets is producing larger and larger trees. But even these big trees are not enough to satisfy a few researchers, who have their eyes fixed on the grand goal of an all-encompassing tree of life. Traditional systematists start with a small group of organisms and work to fill in the neighboring branches and twigs of the tree. But an increasing number of researchers envision methods that tackle organisms from all or most of the tree of life's branches at once—and that don't even require a systematist's expertise to apply.

Hebert's species identification plan is one example. He envisions a handheld device that requires just a small sample, such as an insect leg, to read the sequence of at least 645 bases of the mitochondrial gene cytochrome oxidase I, which is involved in energy production. According to Hebert, fewer than a dozen base substitutions in this gene can distinguish one species from another, yet there's enough variation in the sequence to allow for the discrimination of hundreds of millions of species. He also contends

that the gene's sequence doesn't appear to vary among individuals of the same species, most likely because the gene plays some essential but species-specific role.

If those claims pan out, the gene could be a powerful tool for classification. The bar-code device could compare sample DNA to a database of sequences from thousands of other species. If there was no match, then the sample would become a new addition to the database, thereby reducing the number of unknown organ-

isms by one.

No such device exists yet, but Hebert claims that results to date are promising. In a report published online early this year by the *Proceedings of the Royal Society of London B*, Hebert and his team determined the distinguishing cytochrome oxidase I sequence for seven phyla, eight insect orders, and 200 species of moths. They then used these sequences to classify DNA from 300 organisms from all walks of nature. It worked, says Hebert: The bar code correctly identified each of 150 moths of various species, and almost all the other test organisms wound up in the proper group. Working with Daniel Janzen at the University of Pennsylvania in Philadelphia and John Burns of NMNH, Hebert is now tackling a group of butterflies called skippers, some of which look enough alike to be one species. Burns suspects that they represent about a dozen species, distinguished by slight color differences and feeding preferences. In a pilot project, the bar codes differentiated the various skippers while placing individuals of the same species together.

Hebert has managed to get a few other systematists excited about the potential of his scheme. "There's enormous value in gathering a library of these sequences," says Janzen, who is trying out bar codes with Hebert's help. Rutgers's Grassle, too, has embarked on bar coding, as part of the Census of Marine Life project. "If we can do this on a very large scale, it may accelerate

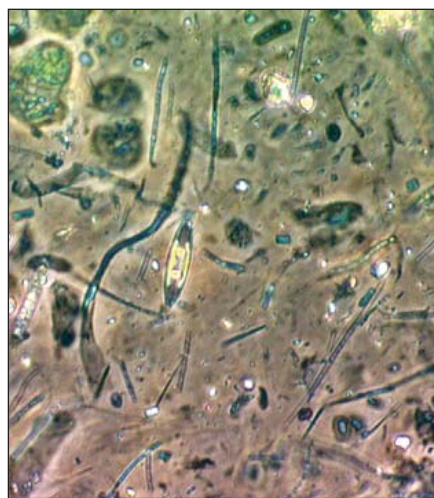
the study of [neglected] species," he says.

Hebert is calling for taxonomists to use the cytochrome oxidase I gene to develop bar codes for museum specimens throughout the world. NMNH is considering asking Congress for \$15 million in the 2005 budget for a 5-year bar-coding project. "We're very encouraged [of the technique]," says museum director Cristián Samper. "We feel it's an important tool."

But there are many, many skeptics, including UC Berkeley's Brent Mishler, who think it's highly unlikely that the cytochrome oxidase I gene can discriminate among all species. Mishler calls the approach "basically terrible and arbitrary," and he worries that spurious counts of biodiversity will result.

Hebert predicts that a global bar-code project would require \$1 billion over the next 20 years. So far, NSF officials are lukewarm at best. "It's not research," says NSF's Rodman, who notes that bar codes are still unproven.

As Hebert tries to drum up support for his vision, other researchers are also dream-



**Hidden diversity.** Researchers have only just begun to catalog the many kinds of microbes on the tree of life.

quickly, which help identify species, and some that evolve slowly, which reveal ancient mutations and evolutionary history. "[Together] they allow you to place an organism within the larger context of the tree of life," says Hillis. "Even if you don't know about something, the [method] instantaneously classifies it."

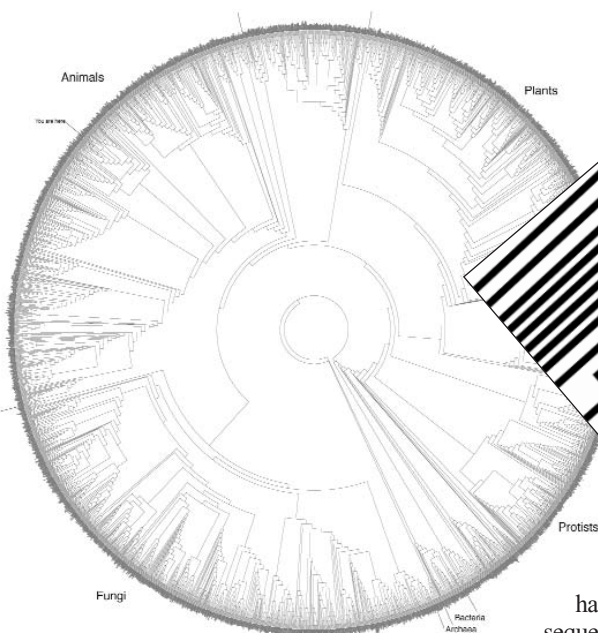
Someday, both Hebert and Hillis hope to have devices that fit in the

palm of the hand and require little expertise to operate. "People could go out anywhere and identify any organism," says Hillis. Such scanners may sound like science fiction. In fact, both Hebert and Hillis liken their imagined devices to tricorders, the whirring handheld scanners that instantly classified alien life forms on the *Star Trek* series. Better yet, says Hillis, call them biocorders.

Very few of their colleagues expect these 21st century tricorders to show up on the lab bench anytime soon. "I am a little doubtful whether it will come down to something that simple," says NMNH's Burns.

Yet whether or not Hebert and Hillis's dreams ever become reality, the existence of such radical schemes signals a field on the edge of transformation. There's a new sense that knowledge of the complete tree of life, once considered an all-but-unreachable goal, may indeed be within biologists' grasp. Three years ago, NSF hosted information-gathering workshops to decide how to accelerate systematic work. Rodman was surprised to find that the several dozen participants were willing to commit to a deadline. They predicted that with enough resources, there could be a rough draft of a tree of life in the next 10 to 20 years, although getting to the species level would take longer. "I think we can have a tree everyone can live with, but that there will be fine-tuning for decades and occasionally some startling shifts," Schram predicts. That fine-tuning and those shifts are part of the process, adds Littlewood. Otherwise, "we wouldn't be engaged in the science."

—ELIZABETH PENNISI



**Roundabout.** David Hillis has incorporated 3000 species into this circular tree of life, best viewed when enlarged to a 1.5-meter diameter.

ing of new devices to automate taxonomic tasks. UT Austin's Hillis envisions a handheld species analyzer that would sequence parts of several genes, not just one, and offer an instant phylogenetic analysis as well as a species designation. He plans to use some genes that mutate relatively