



TAXONOMY

Wanted: A Barcode for Plants

Quick-and-easy DNA identification of animals is under way, but plants are proving harder to pigeonhole

Four years ago, Paul Hebert wowed researchers at the Smithsonian Institution's National Museum of Natural History (NMNH) in Washington, D.C., with the results of a pilot study that he said demonstrated a way to distinguish any animal species from any other, using only a short piece of variable DNA. Hebert, an evolutionary biologist at the University of Guelph in Canada, called it an organism's "barcode." He appealed for a similar effort to find a unique identifier in plants. "He was staring right at me," recalls W. John Kress, a botanist at NMNH. "I took it as a challenge." Kress and his colleagues began what has become a controversial quest for a botanical barcode.

At a meeting in Taipei last month,* hundreds of researchers described their successes in barcoding birds, moths, fish, and other animals, demonstrating rapid progress for this high-tech approach to cataloging biodiversity. Representatives from regulatory agencies outlined plans to use barcodes to track water quality, as well as invasive and endangered species. But despite a strong effort by Kress and dozens of other botanists and systematists, barcoding for plants has yet to gel. "We did not reach consensus" about a few issues, says Ki-Joong Kim, a botanist at Korea University in Seoul, who has come up with his own barcoding scheme.

* The Second International Barcode of Life Conference was held 16 to 21 September 2007 in Taipei, Taiwan.

Debates have been raging about how many and what pieces of DNA it takes to tell one plant from another. Some groups have forged ahead, gathering representative sequences from plants ranging from mosses to daisies, and several teams are developing DNA catalogs of medicinal plants or endangered trees. Yet, for the most part, these data are of little use until everyone can agree on a standard. "Botanists around the world are champing at the bit to get involved in barcoding," says Kenneth Cameron, a plant systematist at the New York Botanical Garden in New York City. "People are very frustrated" by the lack of consensus. And the potential for confusion is rising, as groups pursue selected DNA sequences and different cataloging strategies.

Proposed Plant Barcodes

GROUP	GENE	SPACER
Kress <i>et al.</i>	<i>rbcl</i>	<i>trnH-psbA</i>
Chase <i>et al.</i>	<i>matK, rpoC1, rpoB</i>	
Chase <i>et al.</i>	<i>matK, rpoC1</i>	<i>trnH-psbA</i>
Kim <i>et al.</i>	<i>matK, atpF/H</i>	<i>trnH-psbA</i>
Kim <i>et al.</i>	<i>matK, atpF/H</i>	<i>psbK/I</i>

On the table. Over the past 6 months, researchers have proposed several combinations of DNA regions for barcoding plants.

Identities revealed. Some taxonomists thought these two types of ginger were the same species, but DNA barcoding proved otherwise.

No simple solution

Barcodes on groceries instantly reveal the identity and cost of an item in just a few black and white stripes. In animals, a mitochondrial gene called *COI* seems to work in a similar way, as a kind of species tag. Its sequence varies enough to distinguish most animal lineages but is conserved enough that a single DNA probe works for most organisms. This simplicity has sparked plans to make hand-held sequencers that can provide quick readouts in the field (*Science*, 18 February 2005, p. 1037).

From the start, Kress and others knew that plants would need a different tag. Mitochondrial genes wouldn't work because they evolve more slowly in plants than in animals; too few differences exist between, say, a potato and a tomato to tell them apart. Nuclear genes weren't very appealing either because plant cells often have many copies of a mitochondrial gene but relatively little nuclear DNA. So plant experts turned to a genome not found in animals—that of the chloroplast, the organelle that converts sunlight to chemicals.

As a first pass, Kress and his colleagues scanned the two chloroplast genomes that researchers had already sequenced, picking out nine stretches that varied the most. "The sequences have to be similar enough to be [probed] easily but different enough to distinguish plant species," explains Chang Liu of the University of Hong Kong. Kress's group evaluated these regions. In 2005, at the first international barcode meeting, they nominated about 450 bases, part of a "spacer" sequence between two genes for the plants' barcode, *trnH-psbA*. Spacers tend to be more variable than genes themselves and therefore better identifiers. "So far, it seems to work the best" of all barcodes, Kress insists.

At that meeting, however, "a lot of the botany community said, 'Whoa, there's problems with this,'" recalls Cameron. He and others thought more comprehensive, systematic studies were needed. Representatives from the Alfred P. Sloan and Gordon and Betty Moore foundations—which had financed work in the area—responded with \$900,000 to support further evaluation of barcode candidates. Mark Chase and Robyn Cowan of the Royal Botanic Gardens, Kew, in Richmond, U.K., and researchers from about 10 institutions screened 100 gene and spacer regions in the chloroplast to see which could be pulled out by a single probe. They also checked 96 pairs of species representing

the plant kingdom to see which were variable. And they evaluated the most promising half-dozen in specific plant groups. Kress's Smithsonian group declined to participate; they continued refining the spacer strategy they had proposed.

In the 6 June issue of *PloS One*, Kress and his colleagues reported their results: They adopted a more complex strategy to conduct a survey of 50 plant species. "We all wanted the ideal—a single region," Kress recalls. But as his team looked beyond flowering plants to mosses, liverworts, and other distant kin, they ran into too much variation. Although researchers could line up and compare sequences in closely related plants, those in unrelated plants such as ginger and tomato were too different. The remedy they suggest is a "two-locus barcode," says Kress: both the *trnH-psbA* spacer and part of a gene called *rbcL*. Adding the gene, which has changed much more slowly over evolutionary time, is useful for distinguishing distantly related plants.

"It's a concept that I actually like," says Cameron. But Chase and Cowan haven't been eager to buy into the strategy. Earlier, their team turned away from *rbcL*, which codes for a key enzyme involved in capturing carbon dioxide for photosynthesis, because they couldn't come up with a universal probe for pulling out short, easy-to-sequence pieces. As for the *trnH-psbA* spacer, Chase says their results suggest that its variability limits its utility as a universal barcode.

In the May issue of *Taxon*, Chase's team instead proposed a barcode using three DNA regions. A triple probe is needed, Chase explains, because "no one of them works universally." His group had not quite settled on which probes work the best; different ones help distinguish certain groups of plants (see table).

Meanwhile, Kim had struck out on his own in search of the best plant barcode. His group sequenced the chloroplasts from nine plants, including seven ginseng species, discovering several regions that provided unique species signatures. Kim's group also decided on a three-region barcode—a gene and two spacers—and could discriminate flowering plants belonging to 10 other genera, including dandelions, lilacs, and *Cardamine*. The gene they chose is *matK*, one of Chase and Cowan's choices. Using this method, Kim has already barcoded 500 Korean species.

In all, about a half-dozen proposals came up during the Taipei meeting; discussions were intense. The Korean strategy bubbled up as quite promising, says plant systematist Sean Graham of the University of British Columbia in Vancouver, Canada. But "a final

set of markers was not quite decided on," he notes. Most of the researchers agree that Kim and Chase's *matK* and Kress's spacer should be used. And most are calling for a third region, likely one of the two other spacers proposed by Kim. Graham and his colleagues are going to evaluate these four candidates and report back later this fall on how well they work.

Kress, however, left frustrated. He points out that several papers presented at the meeting supported his choices for a barcode, whereas there's little published data supporting other scenarios. He's hesitant about any three-gene scenario because it would create "an order of magnitude more work." Anyway, he says, "we're moving ahead" for now using his two-barcode regions.

Conflicting needs

Part of the problem is that plant researchers have different needs. For example, a unique barcode may not be all that critical for cataloging the plants in a given habitat, where typically the species aren't closely related. "Less than three, and often one, gene will work quite well" in a local survey, Graham

points out. Indeed, last year, Pierre Taberlet of Joseph Fourier University in Grenoble, France, and colleagues found they could use just one DNA snippet—a noncoding part of a gene—to distinguish half of 132 Arctic species studied and the onion, potato, and leek ingredients in a dried soup mix. The snippet also worked on plant matter extracted from a 20,000-year-old frozen human fecal sample, the group reported online 14 December 2006 in *Nucleic Acids Research*. They suggest that this limited barcode would suffice for tracking plants used in the food and cosmetic industries.

In contrast, taxonomists need more depth within a genus—enough DNA to reveal the degree of relatedness. Introns and spacer regions don't always do that; multiple genes are needed. And some systematists argue that nuclear genes will eventually have to be part of the barcode mix. "It's a Catch-22 situation," says Graham. "The criteria to pick these markers are somewhat contradictory."

But there's a growing need to come up with a solution. Right now, the Barcode of Life Data Systems (BOLD) provides one-stop

shopping for anyone seeking animal barcodes. But neither BOLD nor public databases that archive DNA sequences will accept plant barcodes until there is a single agreed-upon standard. Furthermore, BOLD will need to develop new bioinformatics to accommodate barcodes that include multiple DNA regions. "I am worried that if we don't start thinking about this database [problem], suddenly we will have thousands of sequences and no place to put them," Kress says.

The potential chaos is reflected in barcoding for medicinal plants. The Smithsonian group has developed a barcode library for 750 medicinal plants. But until recently, Kress wasn't aware that Liu has been using yet



Intense debate. In Taipei, plant researchers wrangled over potential barcode regions, making headway but not reaching full agreement.

another barcode combination to catalog Chinese medicinal plants.

And some researchers aren't waiting for the standard to be decided upon. Kress and his collaborators are barcoding the 300 tree species in a long-term study site in Panama, and they plan to do the same at 16 other study sites around the world. Chase's group is developing a barcode database of endangered tropical trees for the United Kingdom to use in detecting illegal timber imports. The Sloan Foundation has asked Cameron to draft a plan to coordinate the barcoding of the world's tree species. And Genome Canada is 500 plants into a scheme to develop barcodes for the country's 5000 plant species.

Yet despite all this activity, David Schindel, executive secretary for the Consortium for the Barcode of Life based in Washington, D.C., argues for patience. "This process has taken longer than anticipated and certainly longer than what we hoped," he points out. "But, at the end of the day, the data will reveal which are the most effective high-performing regions."

—ELIZABETH PENNISI