BOTANY

Plant Bar Code Soon to Become Reality

DNA bar-coding, the ambitious idea of using a short piece of DNA to tell every species in the world apart, is already a powerful tool for scientists studying animals. But for plant biologists, the idea has for the most part remained a pipe dream, stalling systematic studies of plants and efforts to conserve flora.

So far, every plant DNA sequence proposed as a marker has had problems, and the search for the "right" plant bar code has proven controversial (*Science*, 12 October 2007, p. 190). But at last, a solution appears within reach. A paper published this week in the *Proceedings of the National Academy of Sciences* (*PNAS*) proposes two genetic sequences, or loci, taken from chloroplast genes called *matK* and *rbcL*, as the official plant bar code. Although the new bar code works for some kinds of plants better than others, it identified 72% of all species on average and grouped 100% of plants into the correct genus.

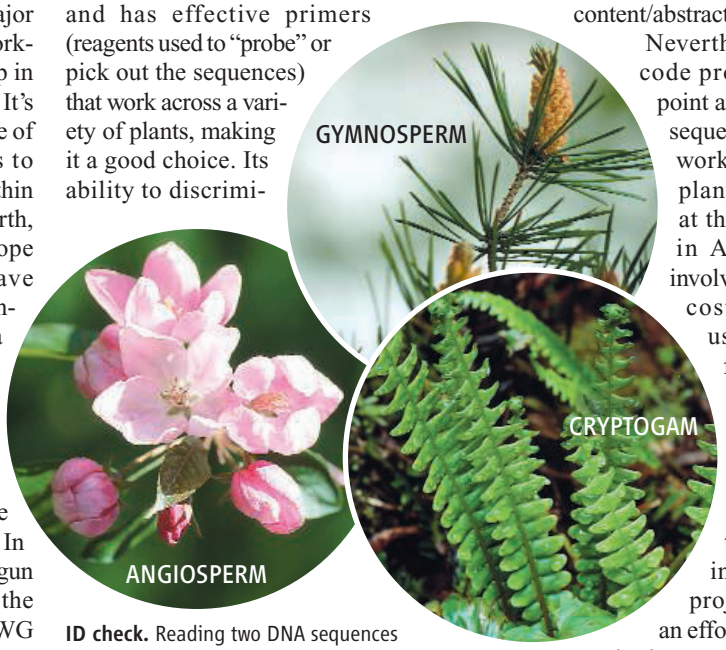
The results comprise data from 25 major bar-coding labs, all part of the Plant Working Group (PWG), an organization set up in 2005 to help resolve the plant problem. It's now up to the Consortium for the Barcode of Life (CBOL) and a board of experts to endorse or reject the plant ID strategy within the next 2 months, says Peter Hollingsworth, chair of PWG. "I have a very strong hope that they will approve it, and we have worked very closely with CBOL throughout this process," says Hollingsworth, a conservation geneticist at the Royal Botanic Garden Edinburgh.

Plant bar-coding centers are eager for that approval. They could then finally "tag" wide varieties of plant species and share that information online through the Barcode of Life Data Systems Web site. In fact, many optimistic researchers have begun bar-coding using *matK* and *rbcL*, after the two loci were informally proposed at a PWG meeting in September 2008.

But PWG admits that the two-locus bar code is far from perfect, and a minority within the group believes the search is not over. W. John Kress, a co-author of the *PNAS* paper, endorses the proposed bar code, but he says his lab will continue to add another DNA sequence called *trnH-psbA* to *matK* and *rbcL*. From these three, the best two-locus bar code would then be selected at a later date. "We're finding that we still get some pretty good results with the three-locus bar code, and personally I'm not ready to give that up yet," says Kress, a botanist at the Smithsonian Institution in Washington, D.C.

The *PNAS* paper reports on seven candidate loci, including sections of genes and spacers (regions between genes), taken from work led by Kress; Ki-Joong Kim, a botanist at Korea University in Seoul; and Mark Chase at the Royal Botanic Gardens, Kew, in the United Kingdom. Each sequence was assessed on its ability to tell species apart and on factors such as its ability to be read rapidly by automated sequencers, says Robyn Cowan, a co-author and conservation geneticist at the Royal Botanic Gardens, Kew. Ultimately, four were ruled out, for reasons such as low discrimination rates or because the sequences were hard to read efficiently.

Of the remaining three, no individual sequence came up to scratch, so the loci were combined to bump up bar-code quality. The *rbcL* gene provides good-quality sequences and has effective primers (reagents used to "probe" or pick out the sequences) that work across a variety of plants, making it a good choice. Its ability to discriminate species hovers above the average at a rate of 61%. Both *matK* and *trnH-psbA* were better at telling species apart, with success rates of 66% and 69%, respectively. But *trnH-psbA* produced poorer quality sequences due to the presence of long mononucleotide repeats, in which one DNA base recurs excessively within the spacer. Such repeats can cause the sequence to be misread. On the other hand, *matK* doesn't have primers that allow it to be isolated in all plant groups. Although there is no immediate way to resolve the issues with *trnH-psbA*, the group expects the quality of *matk* primers to improve over time. Already, the current primers work well in angiosperms (flowering plants), being able to latch onto the loci in 90% of species, but are less successful with gymnosperms (cone-producing plants) at 83% and extremely poor with cryptogams (lower plants such as mosses and ferns) at 10%.



ID check. Reading two DNA sequences can tell most plants apart.

The proposed plant bar code remains less powerful than its animal counterparts, a gene called *CO1*, which averages upward of 95% discrimination. But PWG hopes to push up species discriminating power in the near future. One way of doing this is by using further supplementary bar codes in certain plant groups. For example, in protea, a group of flowering plants, botanists will use the ITS ribosomal spacer, says Chase. (For a Perspective on challenges in plant bar-coding by Chase and Fay, see www.sciencemag.org/cgi/content/abstract/1176906.)

Nevertheless, the proposed bar code provides a good "starting point and a reference for all subsequent plant DNA bar-coding work," says Andrew Lowe, a plant conservation biologist at the University of Adelaide in Australia who was not involved with the paper. As the cost of sequencing drops, using several genetic loci for bar-coding will become a more manageable exercise, he adds.

If the bar code is approved, it will open the way for a variety of plant tagging projects, including large-scale systematic projects such as Tree-BOL, an effort to bar code all trees, and the international grass bar-coding project. While Tree-BOL has been collecting *rbcL* and *matK* bar-code loci for some time, the formal approval of this bar code would make funding applications and publishing results easier, says Damon Little, chair of Tree-BOL and a plant biologist at the New York Botanical Garden in New York City.

For now, the plant community eagerly awaits CBOL's decision. "We [could] at long last move forward with sequencing" on a larger scale, says Little.

–CLAIRE THOMAS

CREDITS: JUPITERIMAGES

Published by AAAS