

DNA BAR CODES

Life under the scanner

BY CHRISTEN BROWNLEE

Just 2 months ago, Bob Ward took a grueling, 27-hour flight from his hometown in Tasmania, to visit his good friend Paul Hebert in Guelph, Ontario. But rather than bringing his host a souvenir jar of Vegemite sandwich spread or a bottle of fine Australian merlot, Ward arrived totting more unusual cargo. Stored securely within his checked luggage were 940 slices of fish muscle. Each sample, no larger than the head of a pin, had been taken from a different fish species native to Australian waters and carefully packed with a few drops of preservative in a small vial. Altogether, the collection was only the size of a laptop computer, but it represented about 1 percent of the world's fish species.

To Ward's relief, the fragile package survived its long journey intact. "There were no breaks or leaks, thank goodness," says the geneticist from the Commonwealth Scientific and Industrial Research Organization in Hobart.

Some people might have viewed Ward's treasure as a difficult-to-appreciate host gift, but Hebert, a geneticist at the University of Guelph, saw it as precious—another opportunity to record a portion of the world's biodiversity. For the past 2 years, his team has welcomed international visitors and their collections of other types of specimens as part of a vast research project known as DNA bar coding.

Just like the 11-digit Universal Product Code (UPC) that identifies items sold in most countries today, a sliver of DNA could identify every living species on Earth, Hebert and his Guelph colleagues assert. With Ward, they now plan to analyze each fish sample by decoding a certain tiny piece of DNA that can serve as a unique bar code for the species.

The bar codes will enter a master database that scientists consult to identify mystery species, just as a scanner-wielding store clerk gets an accurate ID on mincemeat without having to recognize what it is.

"We're looking at distinguishing the major product lines of life, if you will," says Hebert.

The technique has yet to prove itself fully viable. But if Hebert and his collaborators are correct, DNA bar coding could have a tremendous impact on both scientists and nonscientists with a need to recognize different species. For example, a handheld DNA-bar coding device currently in the works could provide a quick and cheap way for researchers to catalog the world's estimated

10 million animal species, a monumental task that currently overwhelms the relatively few taxonomists trained to undertake it.

DNA bar coding could also benefit farmers who need to identify nonnative species invading their fields, doctors trying to find the correct antivenin for snakebite victims, or schoolchildren curious about a flower they picked on a field trip.

Dan Janzen, a tropical ecologist at the University of Pennsylvania in Philadelphia and one of Hebert's recent collaborators, says that identifying organisms at the flick of a handheld scanner reminds him of the science fiction novels he read as a boy. "When I listen to people talking about bar coding, I think it might not be science fiction anymore," he says.

RAISING THE BAR The ideas behind DNA bar coding have been percolating as long as Hebert has been a population geneticist. Several techniques that promised to separate species using simple molecular biology have come and gone over the past 30 years.

One such approach focused on allozymes, enzymes that vary only slightly among species. However, researchers eventually found that the allozyme technique couldn't distinguish between closely related species. Because allozyme data are subject to interpretation, individual scientists could read the same results in different ways, effectively blurring species boundaries.

"At the back of my mind has always been developing a more effective approach for dealing with life's diversity," Hebert says.

Eventually, he recalls, he and his colleagues hit on the possibility of a single gene being the universal bar code for living things. The order of DNA's nucleotides—A, T, C, and G—

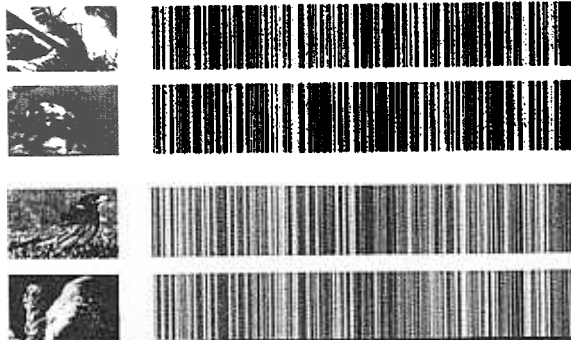
within that gene in each species would fill the role taken by numbers on a UPC in warehouses and stores.

But the right gene would have to meet strict specifications. To keep costs low and accuracy high, the target-gene region must be no more than 650 nucleotides long. With today's technology, regions this short can be sequenced quickly with few mistakes.

Next, the gene would have to be easy to extract from an organism's vast collection of genetic material.

Most important, the gene must have mutated fast enough to be distinguishable from species to species, but not so fast that it differs greatly among individuals within a species.

After years of trial and error, Hebert's team settled on a mitochondrial gene known as *cytochrome c oxidase subunit I*, or *COI*. Not only does *COI* seem to fill all requirements, but it's also present in every animal species on Earth, Hebert says. That's because *COI* is necessary for animals' cellular respiration, the process by which cells use oxygen to create energy.



PRODUCT LINES — Each color represents one of DNA's four nucleotides on these bar codes for the honeybee, bumblebee, American robin, and hermit thrush (top to bottom).

On the other hand, the gene isn't a suitable bar code for plants or bacteria. For those organisms, other genes will have to be selected. Regardless of which gene regions researchers eventually choose, that DNA must be standardized for a large number of species to fall under the bar coding umbrella.

TEN IN ONE Although the focus on *COI* is new, DNA bar coding relies on methods that are already commonly used in molecular biology labs. When Hebert receives a new set of samples, such as the Australian fish-muscle collection, the researchers start by extracting all DNA from each bit of flesh. They then pull out just the *COI* gene from that DNA. To obtain enough of the gene to work with, the researchers make multiple copies of it using a technique called polymerase chain reaction. Then, the researchers put the product into a machine that automatically reads the string of nucleotides in DNA. Finally, they sit back and wait for the results.

"If we start in the morning, by evening we could have those sequences finished, all without resorting to any exotic technology," says Hebert. "There's not much rocket science involved in this."

With such a speedy turnaround, Hebert and his supporters estimate that if the tissue samples were available, the world's 10 million animal species could all be accurately bar coded by a single high-volume sequencing lab in about a year. At current costs, the effort to collect tissues from animals and bar code these samples would be only about \$1 billion. Although that investment would be substantial, it would be significantly less than the cost of describing each organism using traditional taxonomy or more-intensive genetic methods.

The first paper that Hebert's group published on the use of *COI* as a bar code came out last year, prompting a smattering of papers published soon thereafter by other groups. For the past 2 years, Hebert and his lab-mates have been working with other geneticists, taxonomists, and ecologists from around the world to start the overall task of bar coding animals.

In the Sept. 28 *PLOS Biology*, they reported on their ongoing effort to bar code all bird species in North America. Their findings include the discovery of at least four new bird species.

For another study, published in the Oct. 12 *Proceedings of the National Academy of Sciences*, Hebert worked with the staff of Janzen's lab and John Burns of the Smithsonian Institution in Washington, D.C., to bar code a species of skipper butterfly known as *Astrartes fuligator*. Scientists had previously suspected that it might actually include more than one species because the butterflies cover a particularly large territory, ranging from the southern United States to northern Argentina, and feed on an unusually wide variety of plants.

Sure enough, bar-coding data indicates that the group identified as *A. fuligator* is made up of at least 10 species of identical-looking butterflies. Janzen and his colleagues contend that the butterflies might be part of a large "mimicry complex" whose various species have evolved to look the same as a particularly fast-flying butterfly that birds don't bother to chase.

OUT OF LINE Despite these seeming successes, many researchers suspect that Hebert, Janzen, and their far-flung colleagues are wasting their time. According to Diana Lipscomb, a protist taxonomist at George Washington University in Washington, D.C., DNA bar coding simply hasn't been around long enough to have established its value.

Hebert claims that the technique can distinguish among species, even closely related ones, with about 97 percent accuracy. But without significantly more testing, Lipscomb counters, it's impossible to say whether *COI* can distinguish all animal species with the low margin of error that Hebert promises.

"I'm concerned that they're throwing away the old rigor [of taxonomy] and adopting [DNA bar coding] as if it was perfect from the beginning," she says.

Craig Moritz, an evolutionary biologist at the University of California, Berkeley, also has concerns. "There are many cases where you might not get clean partitioning," the definite separation of species, he says. Moritz is coauthor of an editorial evaluating DNA bar coding in the Sept. 28 *PLOS Biology*.

For example, contends Moritz, in a species that has evolved from other animals within only the past 10,000 years, DNA bar coding based on only one gene might not show the species as being an independent group. Similarly, a single-gene approach may not distinguish hybrids, the products of crosses between members of different species, from their parent species.

Other problems arise, says Moritz, because *COI* is a mitochondrial gene, passed down only from mothers to their offspring. Consequently, if females from a population don't migrate and mate as widely as males do, changes in their *COI* may not reflect the entire population.

Some scientists have more general worries that *COI* simply isn't specific enough to work as a marker for every animal species. Jim Mallet, a butterfly taxonomist at University College London, contends that DNA bar coding assumes "that species are like items you can buy in the supermarket, and that's not true of living organisms." Because organisms are constantly experiencing evolutionary changes, scientists can't transfer the rules governing toothpaste to living things, Mallet adds.

With DNA bar coding growing in popularity, Mallet says that he and other traditional taxonomists are concerned that the technique will divert attention and grant money from tried-and-true taxonomic studies. "It's less easy to hype [taxonomy] and make it really exciting. Everybody loves a magic bullet, but in science, magic bullets just don't work," he says.

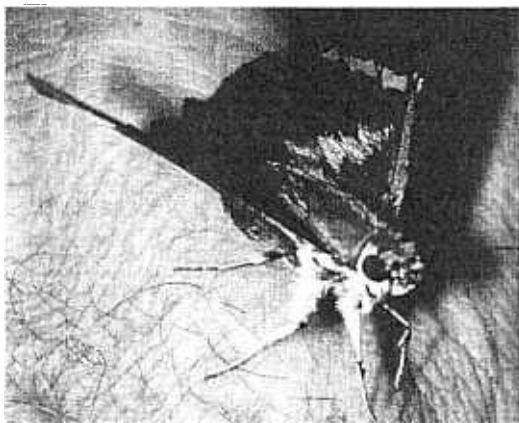
CHECK OUT These concerns have alienated some taxonomists from DNA bar coding, but Hebert says that his goal isn't to replace traditional techniques. Instead, he expresses hope that creating a bar code catalog of the world's species will give taxonomists a timesaving, first-pass guess at the identity of a species.

Meanwhile, more and more taxonomists seem to be moving beyond the traditionalists' concerns to join Hebert in gathering samples of animal species or developing new techniques for bar coding. John Kress, a plant taxonomist at the Smithsonian Institution in Washington, D.C., for example, has developed a potential method for bar coding plant species that he and his colleagues plan to describe in a publication soon.

Kress declines to name which gene or genes are used in this technique, but he notes that his team will soon be testing it "in a big way" by bar coding all known plant species in Costa Rica.

Plans by other researchers to bar code mussels, plankton, and even some pathogenic bacteria are also in the works.

Hebert predicts that society will eventually see substantial returns if it invests in the bar-code technology. "This is one of the grand science bargains on the planet," he says. "Compare it with \$30 billion for the International Space Station, \$10 billion for the next cyclotron. We're asking for just \$1 billion. Will you give it to us today?" ■



FLUTTERING FACSIMILE — Scientists using DNA bar coding found that this skipper butterfly, currently called *Astrartes fuligator*, belongs to one of at least 10 species that hadn't been previously distinguished.