

December 14, 2004

A Species in a Second: Promise of DNA 'Bar Codes'

By NICHOLAS WADE

When an astronaut sets foot on an alien planet and sees moving shadows in a nearby wood, he whips out a scanning device that immediately identifies the menacing life-form, flashing up a photo of its species and an assessment of its aggressive intent.

If such devices are standard equipment for visiting distant planets, why can't we have them here at home where we really need them? Less than a fifth of the earth's 10 million species of plants and animals have been cataloged, and taxonomists are backlogged with requests to apply their specialist knowledge to identification problems.

The wait for a simple way of identifying species may not be too much longer, if an idea known as DNA bar coding should prove as good as its advocates say. DNA bar coding depends on analyzing part of just one gene, the same gene in all cases, for every species. If and when a DNA bar code database of all terrestrial plant and animal species is established, a field biologist could take a tiny piece of tissue, like a scale or hair or leaf, from the unknown specimen, and feed it into a hand-held device for analysis. With a cellphone call to the database, the device would identify the species and present its photo and description.

DNA bar coding is the idea of Dr. Paul D. N. Hebert, a population geneticist at the University of Guelph in Ontario. It has attracted support from the Alfred P. Sloan Foundation of New York, where it has been championed by a program officer, Jesse H. Ausubel of Rockefeller University. In April a consortium of major natural history museums and herbariums started the Barcode of Life Initiative, a plan to create a DNA bar coding database that would be linked to identified specimens in their collections.

Such a system, if it works as promised, will help field biologists identify known species and assist immensely in the urgent task of cataloging unknown species before their ranks are decimated by extinction.

Public health authorities could identify whether mosquitoes were of a type likely to carry deadly diseases. Since bar coding works on tissues, inspectors could test animal feed for forbidden items likely to spread mad cow disease.

And because DNA bar codes can be obtained from museum specimens up to 20 years old, curators throughout the world could bar code their collections, making their priceless storehouses of taxonomic knowledge available to everyone and easily accessible.

Despite its promise and quick start, DNA bar coding has not yet won the unanimous support of a core constituency, the taxonomists and biologists who would be its principal creators and users. The resistance stems from doubt that DNA bar codes can distinguish between closely related species.

There may be other sources for the muted enthusiasm, like insiders' habitual reserve toward outsiders' help, and the traditional difference of perspective between those who study whole animals and those interested only in their DNA.

Two reports published in October demonstrated the striking power of DNA bar coding. In an article in *The Proceedings of the National Academy of Sciences*, a team led by Dr. Hebert and Dr. Daniel H. Janzen of the University of Pennsylvania showed that members of a Costa Rican butterfly species, *Astraptes fulgerator*, possessed a total of 10 bar codes. The finding meant that the butterflies were not a single species, as long assumed, but a complex of 10 species occupying overlapping territories.

Researchers had suspected something of the kind because the caterpillars looked quite different and preferred different food plants, as if each had long diverged away from a distant ancestral species. But all looked the same as adults, as if that particular appearance - livery that includes a splash of iridescent blue scales across the body and wings - carried such a survival advantage that none of the descendant species could stray from it.

In another article, in *Public Library of Science Biology*, Dr. Mark Stoeckle of Rockefeller University and colleagues developed DNA bar codes for 260 species of North American birds. They found that all had different DNA bar codes except four species in which two bar codes were present, suggesting in each case that a single species was really two.

Biologists have been using DNA to differentiate between similar species for years. What is new about Dr. Hebert's approach is his proposal to standardize this process by using the same gene for all animal species. He calls the idea DNA bar coding by analogy with the bar codes on supermarket products.

"In a grocery market it occurred to me that we have minimal diversity in universal product codes, yet we can distinguish all these products," Dr. Hebert says. Commercial bar codes generally have just 10 digits. A short length of DNA, a hundred units in length, would be enough to distinguish one species from another, Dr. Hebert says, but since DNA sequencing machines read off the order of DNA units in batches of about 650 units at a time, Dr. Hebert has settled on one "read" as the bar code's length.

The sequence read he has chosen is the first 648 DNA units of a gene called CO1 (for cytochrome c oxidase 1). The gene is one of the few that escape the shuffling of genetic material between generations because it belongs to the mitochondria, energy-producing subunits of the cell that are inherited solely from the mother. CO1 is particularly suitable because in most animal species it has no inserted or deleted DNA units, allowing all of the CO1 sequence reads to be lined up and directly compared.

The other new aspect of Dr. Hebert's proposal, besides the standardization, is his surprising discovery that it works at all. Biologists would expect the difference in CO1 units between individual members of a species to be quite wide, so wide as to overlap with individuals of closely related species. If such overlaps were common, CO1 would be of little help in identifying species. But in most of the animal species tested so far, CO1 differences within a species have turned out to be surprisingly small - just 2 percent.

Dr. Hebert believes that variation within a species is reduced by the evolutionary mechanism called a selective sweep. Sweeps occur because most changes in the CO1 gene are so important that they rapidly spread through all members of a species. This erasure of natural variability within a species is the critical feature that creates a distinctive genetic signature for each species and makes bar coding possible. But critics say they are not yet convinced that the signatures will be distinctive in every group of animals.

In the 245 years since Linnaeus, taxonomists have identified about 1.7 million species of plants and animals. Assuming there are some 8.3 million species yet to be described, "We will need 1,196 years to complete the job," Dr. Quentin Wheeler of Cornell University laments in a recent issue of *The Systematist*, a taxonomists' newsletter.

So why aren't taxonomists rushing to embrace DNA bar coding? Many doubt that a single gene can serve to

define all species, or substitute for the painstaking study of bodily characteristics and the years of training required for good taxonomy. There is no time to waste on missteps "like arbitrary DNA bar codes," Dr. Wheeler warns.

Nor are ornithologists entirely convinced by Dr. Stoeckle's recent article on the bar coding of birds. "Their sampling is so limited," says Dr. Carla Cicero of the University of California, Berkeley. She thinks too few birds of each species were tested to prove bar coding will provide a distinctive signature, especially for closely related species.

The backers of DNA bar coding are trying hard to win over the taxonomists. The idea is not to supplant taxonomy, they say, but to build on it and expand its power and utility. It is true that one gene is unlikely to distinguish all species perfectly but, come to think of it, traditional taxonomic methods are not so close to perfect either. The sensible criterion, say bar code backers, is practicality, not perfection: does DNA bar coding track traditional taxonomic classifications closely enough to be widely useful?

One taxonomist who believes DNA bar coding will be effective is Dr. James Hanken, director of the Museum of Comparative Zoology at Harvard. Dr. Hanken started out as a critic of bar coding, and indeed believes that it will not work for the particular animal group he studies, that of salamanders. But he is impressed that the bar code advocates have been open to criticism and have adopted standard taxonomic practices like linking bar codes to vouchered museum specimens.

"I think we should give bar coding a try," he said. "I am also taken with the idea of bringing biological diversity out of the realm of professional specialists and making it available to lay people who need information about species identification. Making this knowledge available to others is tremendously important and very attractive."

Dr. Angelique Corthals, the researcher responsible for DNA bar coding at the American Museum of Natural History in New York, said it would be "tremendously helpful in re-sorting and organizing museum collections round the world." But because the approach is still controversial, her museum is not yet an official member of the bar code consortium, she said.

The central idea of bar coding is standardization, but the CO1 gene alone may not have sufficient powers of discrimination for all animal groups. If so, a second mitochondrial gene could be chosen to help with these groups. Plant biologists have already decided that CO1 is not the right gene to distinguish among plant species and may prefer two genes on the chloroplast, the light-trapping subunit of plant cells.

Dr. Hebert said that he had had "immense difficulty" getting his first article about DNA bar coding published and that he had spent four years trying to raise money for the idea. His luck turned in February 2002, when he gave a talk to a group of marine conservationists. Dr. Ausubel of Rockefeller, who was present, was so impressed with the idea that he convened two seminars of senior scientists and museum directors to discuss it. Out of these meetings grew the Consortium for the Barcode of Life, financed by the Sloan Foundation. This month Dr. Hebert received a \$3 million grant from the Gordon and Betty Moore Foundation to churn out bar codes in high volume.

Dr. Ausubel expects that bar coding will find its first uses in commercial applications. He hopes this will drive down costs and stimulate development of the necessary hardware, like the hand-held species identification device for biologists in the field.

Dr. Janzen, for one, can hardly wait. Caterpillars of the Costa Rican forest are his specialty. Most taxonomy is based on identifying adults, not larval forms, but bar coding works equally well with either. "I want something out of my back pocket that for one cent gives me one name in one second," he says firmly.

