

# NewScientist

26 June 2004 No 2453 Australia \$5.95 (inc.GST) New Zealand NZ\$6.95 (inc.GST) Print Post Approved 290009/0005

## 21<sup>ST</sup> CENTURY ARK TAKING STOCK OF NATURE'S RICHES



### Uncertain universe

Why chance is blocking our view of reality



# Barcode me

Wouldn't it be useful if we could identify any animal on earth simply by reading off a short stretch of its genetic code? Bob Holmes talks to the people who are making this dream a reality

THAT mosquito you just swatted in the act of biting your arm—is it a member of a species that carries West Nile virus or some other nasty disease? With a good entomologist at your elbow, you might get an answer. Ditto if you are one of those rare people who carries a microscope and a manual of mosquito identification. But otherwise, forget it—you will probably never put a name to your tormentor, just as you will fail to identify the ants in your picnic, the moss on the tree trunk and that unusual woodland flower over there.

Fast-forward a few years: instead of just shrugging your shoulders helplessly, you pull a cellphone-sized device out of your pocket and pop the squished insect into its sample port. Seconds later, the device tells you what species you have and links you to a description of its biology. The mosquito is harmless. The flower turns out to be the most toxic plant in the country. The ant is an exotic invader from South America, common further south but never before recorded in this part of the country. And the moss is new to science.

This device—reminiscent of the tricorder from Star Trek—doesn't exist yet, of course, but it may not be as far-fetched as it sounds. Several companies are already working on technology that would enable you to identify any living animal using the same principle that lets supermarket cashiers scan the items in your shopping trolley—only using genes instead of black and white stripes.

But the idea of genetic barcodes is not without its critics. A heated debate is raging in the normally quiet field of taxonomy over whether it can really work. And there's plenty at stake. If barcoding proves its worth, it could give doctors away to quickly identify pathogens, allow agricultural inspectors to spot noxious alien species, and give civil-defence authorities a head start in detecting bioterror agents.

Prominent biodiversity experts are even talking seriously about achieving one of biology's grand ambitions—cataloguing every

species of life on Earth. They hope that the pay-offs from this catalogue will match those of the human genome project, which has inspired whole new ways to think about what makes us and other creatures tick. Knowing every species of life on Earth will help biologists answer the fundamental questions of ecology and evolution. Which species live where? Are most species widespread or narrowly restricted in range? Where are the hottest biodiversity hotspots? And knowing the answers to the basic "which" and "where" questions helps to frame the right "why" questions—the ones we really want answers to.

"The applications of the knowledge are multiplicative," says E. O. Wilson, an evolutionary biologist at Harvard University who is widely considered the godfather of biodiversity. "The more information you have, the more connections you can make, not only in the study of biodiversity itself, but also other fields. Ecology becomes a great deal more precise, the more species you are able to identify. Molecular biologists searching for novel proteins can make surveys much more swiftly." Agricultural scientists, too, will know the full spectrum of wild relatives available for breeding crop plants, and will have a complete

catalogue of species as they search for novel genes. But to achieve this, scientists need an easy way to record all those species, and that's where genetic barcodes come in.

As things stand today, professional ecologists fare little better than anyone else when it comes to identifying species in the field. "I walk out in the forest and take a photograph and there are 500 species of plants in that photograph. Every single one has a Latin name, and I cannot identify a single one," says Dan Janzen, a tropical ecologist at the University of Pennsylvania in Philadelphia. The fact is that once you venture beyond the best-known organisms such as birds, mammals, butterflies and beetles, a given group of organisms may have only one or two specialists in the world—if that—who can identify species with any confidence, and those experts often have a backlog a month deep from scientists asking for help. Even then, in many groups—minuscule parasitic wasps, nematode roundworms and almost all the microscopic world, for example—the vast majority of species are still unnamed and unknown to science.

It is this problem that inspired evolutionary biologist Paul Hebert from the University of Guelph in Ontario, Canada, to look for a solution. "Being an impatient person, I think I had a clear sense of the need for a more rapid and universal ability to identify life," he says. Four years ago he hit on the idea of genetic barcodes, reasoning that if a standard supermarket barcode can use a simple string of numbers to uniquely identify millions of consumer products, then a simple DNA sequence should be able to do the same for species.

So Hebert set out to find a piece of DNA that had all the right features to make a good molecular barcode. He decided to start with the animal kingdom, so clearly he would need a gene that is found in all animals. His attention quickly focused on mitochondria, the cell organelles responsible for energy







production. These carry their own tiny genome, which evolves faster than the nuclear genome—fast enough that even recently diverged species would be likely to carry different DNA sequences. And because mitochondria reproduce without sexual recombination, their genes are less prone to insertions, deletions or other large-scale rearrangements that could scramble the barcode and make it harder to read.

Two mitochondrial genes looked promising: *cytochrome c oxidase (COI)* and *cytochrome b*, both of which play central roles in converting chemical energy into the chemical ATP that cells use to power their daily activities. Because COI proved easier to isolate from a wide range of organisms, Hebert opted

for that gene as his barcode—in particular, a stretch 645 bases long near the beginning of the gene.

To test this choice, Hebert and his colleagues compared COI sequences for over 13,000 pairs of closely related animal species found in the GenBank public sequence database to see how much they differed. Last year they reported that over 98 per cent of sequences differed by more than 2 per cent of sequences (*Proceedings of the Royal Society B*, vol 270, p S96). In contrast, sequences from different individuals of the same species tended to differ by much less than 1 per cent. In other words, you can almost always draw a clear line between the amount of barcode variation within a species and the amount

of variation between two species.

Each of 200 moth species from the Guelph area, for example, turned out to have distinctly different barcodes. Likewise, barcodes correctly identified almost all of 421 specimens of about 100 Costa Rican moth species provided by Janzen, and revealed that one "species" of skipper butterfly may really be 10 separate species with different feeding habits and distinct barcodes. "All the evidence to date points to the ability to resolve the immense diversity of animal life," says Hebert, who has now barcoded more than a thousand species. "This year we'll see a small flotilla of papers looking at different taxonomic groups."

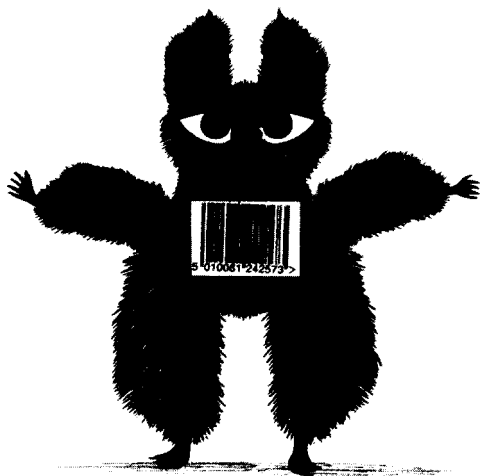
COI barcodes should also work in fungi and unicellular organisms, Hebert says.

But barcoders will need to find a different gene or set of genes for plants, because their mitochondrial DNA evolves too slowly to distinguish between species. And so far, Hebert has steered clear of bacteria and archaeans, whose great diversity and willingness to trade DNA will make finding a barcode for them much more difficult.

Not everyone shares Hebert's enthusiasm, however. "A supermarket barcode is very useful, isn't it?" says James Mallet, an evolutionary biologist at University College London. "All 500-millilitre bottles of orange juice made by the same company have the same barcode. But it isn't that way for species." Although Hebert's studies show more barcode variation between different species than between individuals of the same species, the distinction may not always be so clear-cut. Felix Sperling, a taxonomist at the University of Alberta in Edmonton, Canada, points out that the GenBank records Hebert used are biased towards organisms whose DNA sequences fall into clearly demarcated species groups. Along with other critics, Sperling worries that these preliminary studies might paint too rosy a picture, and that in future barcoders may find themselves mired in uncertainty.

"You go out in the field, get a barcode sequence, and it could be six bases different, say, from anything else you've got. Does it belong to another species? You don't know," says Chris Humphries, a botanist at the Natural History Museum in London. As many as a quarter of all species might fall into grey areas of this sort, says Sperling, who has used COI as one of many characters in sorting problematic groups of insect species. The moral, he says, is that barcodes are most likely to fail in precisely the cases where they would be most useful—closely related species that are hard to tell apart visually.

Such uncertainties don't bother Hebert. After all, he says, traditional taxonomists often have trouble deciding whether some evolving populations represent the same or different



**"Once you venture beyond the best-known organisms, any given group may have only one or two specialists who can confidently identify species"**

species. The classic definition of a species—in which two populations represent the same species if they can interbreed, and different species if they cannot—is difficult to apply in practice. So taxonomists must fall back on morphology, behaviour and ecology, which can often be ambiguous. Why should we expect barcoding to do any better in these problem cases, Hebert asks.

Besides, even the "failures" may be close enough for many purposes. "Failure isn't an absolute failure of no information. It's that you only have partial information. If you could even identify [any organism] to genus, that would be a tremendously useful system," says Scott Miller, an insect systematist at the Smithsonian Institution's National Museum of Natural History in Washington DC, who is a leading advocate of barcoding.

Only further studies will reveal whether the failure rate of barcoding is closer to an acceptable 2 per cent or a much more worrisome 25 per cent. "All we can say is, go out and collect data on your favourite group of organisms, and look at the results," says Hebert. He and his collaborators are doing just that. Within five years, Hebert expects to have barcoded every economically important animal species in Canada—some 10,000 in all. By the end of this year, he and his colleagues should have all North American birds in the bag, and within a few more years, every bird in the world.

The Smithsonian has agreed to host an international barcoding secretariat to coordinate barcoding efforts. And in April the New York-based Sloan Foundation promised **\$670,000** to kick-start the Barcode of Life Initiative, a central repository of barcode data. "This is less than a year from first publication. I'm staggered at how fast this is moving," says Hebert.

But he is after bigger game yet. Barcoding, he says, offers biologists their best shot at building the longed-for comprehensive catalogue of every living species on Earth.



Certainly, barcoding would seem tailor-made for such a massive project. It is quick and simple, and a well-trained technician can collect and sequence barcodes to flag new species without needing expert knowledge of specific groups of organisms. However, those very strengths are also barcoding's biggest weakness.

"If you just did barcoding, you could count the species, but you would only have a partial genome and a name," notes Wilson. "Everything in between—what are organisms like, how do they behave, how do they interact with other organisms, what do they even look like—you'd have to do that with straight, direct, boots-on-the-ground fieldwork. You're going to have to do it the old-fashioned way."

Indeed, Wilson and others have been planning just such an effort, though they have yet to secure funding.

Then, too, there's the problem of placing all the new species on the tree of life. By most estimates, the 1.7 million species recognised or named to date represent barely one in five of Earth's total, and the millions of unnamed ones will need to be slotted into their proper place. The barcode sequence suggests where that should be—near other species with similar sequences. However, Hebert's barcode system leaves a lot to be desired when it comes to placing organisms on branches. His initial test runs turned up several glaringly misplaced organisms—ladybird beetles classified as wasps, arthropods stuck in with molluscs, and the like. In one test using North American

birds, only 88 per cent of species ended up in their proper place in the evolutionary tree. Any unknown specimen sorted by barcode alone would run the risk of similar misclassification.

"The biggest problem is, we don't know when we're wrong," says Kipling Will, an insect systematist at the University of California, Berkeley. That means barcoding's supposed advantages in speed and efficiency are mostly illusory, he says, because you still need a taxonomist to review the results and catch errors. Like many taxonomists, Will worries that if too many people jump on the barcoding bandwagon, there won't be enough taxonomists—or money—left to do that checking, or to accumulate the rich biological detail about the organisms being tallied. "We have to be willing to do the hard work, to train natural historians who really can identify what they see," he says. Such an effort would require perhaps three times as many taxonomists worldwide as are working today, but Wilson estimates it could complete a global biodiversity survey in about 25 years.

Most barcoding advocates agree that their tool works best alongside rather than in place of more traditional research. They say barcoding can free traditional taxonomists from the drudgery of routine identifications and let them use their limited time more effectively. And for many groups of organisms it may not be necessary to describe every species in detail immediately, says Miller. Barcodes could be used to estimate the number of species in a genus, leaving taxonomists to describe in detail only the few of economic importance on their first pass. Indeed, some Australian entomologists at CSIRO in Canberra are already beginning to adopt this strategy, says Miller.

All of this makes barcoding a tool well worth serious thought for anyone interested in drawing up an inventory of life. But if that seems too esoteric, the approach also has more commercial potential. At Canon US Life Sciences in Alexandria, Virginia, for example, Rita Colwell, the former director of the



National Science Foundation, has just taken charge of an effort to develop a rapid, portable DNA sequencing device to allow doctors to identify disease-causing microbes, or to let civil defence sentinels nip bioterror attacks in the bud. "It's a difficult problem, but it's very doable," Colwell says. She expects a device within five years.

Another biotech company, US Genomics of Woburn, Massachusetts, is hoping to target a similar market with a somewhat different DNA identification method. Their technique, which they call DNA mapping, involves using a series of tags that bind to specific short sequences of a species' DNA. They then use a laser to read the location of each tag along the DNA molecule. Each species will produce a different binding profile for the tags, allowing quick identification of the species.

Technologies like these promise to be small and fast enough for scientific use, but what about Joe Public? Ianzen, for one, thinks the day is not far off when barcoders could be small enough, quick enough and cheap enough that anyone who wants one could have one. "There are more cellphones in the world now than there are regular ones," he says. "How long did that take? How much does a cellphone cost you now?" Providers may someday choose to simply give away barcoders and earn their profit by charging a few pennies for each identification, he says. "To me the social drivers are there for this to happen very, very fast."

If it does take off, barcoding technology has the potential to transform the way we relate to the natural world. As more and more people learn about the life that surrounds them, Janzen thinks, they will also learn to care. "To a person who's illiterate, a library is a large stack of firewood," he says. "To a person who cannot read biodiversity, it's just green. And what do we do with green? We just push it out of the way. If we want to keep that stuff out there, people have to see it as more than just a green blob."

**"Barcoding offers biologists their best shot at building the longed-for comprehensive catalogue of every living species on Earth"**