

April 2006

www.universityaffairs.ca

The Barcode of Life takes flight

Many researchers initially scoffed at the idea that you can use a tiny string of an organism's DNA to accurately identify the species to which it belongs. No more. Research on DNA barcoding has exploded, thanks to the backing of heavyweights like the Smithsonian and tens of millions of dollars in funding. The goal is to develop a complete catalogue of Earth's life forms using the novel technique. The man at the centre of this Canadian-initiated effort takes his new-found fame and research wealth with a dose of good humour.

by Stephen Strauss

Stroll through Paul Hebert's bustling laboratory at the University of Guelph and it's easy to see how the past and future of species identification are visually colliding before your eyes. The past is what Dr. Hebert metaphorically describes as a "kaleidoscope of butterflies." That is to say, display cabinets inside which whole butterflies and moths in all their myriad mottles, colours and shapes have been delicately pinned.

Shapes, colours and behaviour of butterflies, not to mention all other living creatures, have been the bedrock of species identification ever since Carolus Linnaeus – the extraordinary Swedish life-classifier – used these characteristics to create the modern science of taxonomy in 1735.



The future, however, looks rather fractured and uncolourful, as elsewhere in the laboratory sit 96 clear plastic vials in which skinny, broken butterfly legs have been immersed in methanol. Over the course of a day, lab technicians will put the legs through a process in which their DNA is removed and the precise order of the 648 chemicals comprising a single mitochondrial gene is determined. This mitochondrial gene, formally known as CO1, is ubiquitous in all living things higher than viruses and bacteria.

What this DNA fingerprinting will uncover is not some random distribution of life's chemicals but a profound orderliness, which to date has been specific to nearly each and every animal species studied. The order constitutes, in Dr. Hebert's felicitous and now world-famous metaphor, a "species barcode."

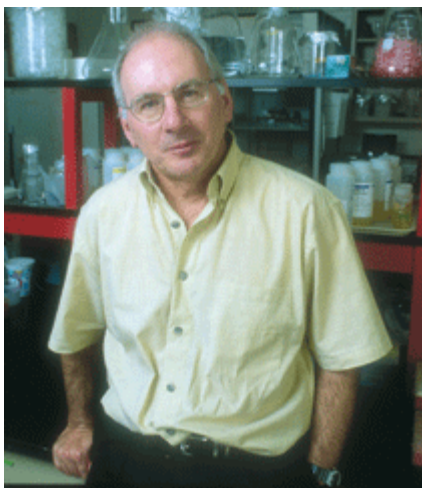
Since publishing the first demonstration of the barcode in 2003, the evolutionary biologist has shown that DNA sequences of CO1 are a breathtakingly 99.75 percent identical within members of a species. "This means that across the entire barcode reading, there are only a couple of nucleotide differences between you, me, an Inuit, or a bushman from Australia," says Dr. Hebert. If the coincidence falls to 97.5 percent, you likely have crossed into some other species.

The original article may be found at:
http://www.universityaffairs.ca/issues/2006/april/barcode_life_01.html

Not only has the DNA shown with great accuracy that the barcode replicates most past Linnaean species divisions, but in some cases it has forced scientists to rethink the degree of refinement of some of their previous classifications. Perhaps the most famous reconfiguration was Dr. Hebert's own 2004 genetic barcoding of 260 North American bird species which, based on DNA evidence alone, soberly pronounced that four so-called varieties of known bird species were actually new species.

What this simple description of barcoding technology and its applications won't convey to you is how important this change really is, because it now looks as if Paul Hebert and his once-small laboratory in Guelph, Ontario, have given birth to that rarest of things in science – a paradigm shift, both for science and for the life of an individual scientist.

Here's what DNA barcoding has meant for Dr. Hebert personally: When his first paper on DNA barcoding appeared in 2003, the professor was in his mid-50s, heading up a laboratory consisting of a handful of graduate students deeply involved in studying the evolution of asexuality, notably in water fleas. He was supported by what seemed then a munificent \$110,000 Discovery Grant from the Natural Sciences and Engineering Research Council of Canada.



Paul Hebert: rejuvenated by his monumental discovery.

Then species barcoding broke into the world of science. By the beginning of 2006 Dr. Hebert had received promises of \$25 million to \$30 million in funding from American and Canadian sources, and was awaiting the opening of a new "barcode factory" the size of 10 family homes on the Guelph campus. He has 26 people working for him – mostly permanent staff or postdoctoral fellows – and his staff is likely to double or triple in the near future. In addition, he oversees more than 50 Canadian scientists who are looking to see whether DNA barcoding can differentiate fungi, seaweed, and maybe plants as well.

Dr. Hebert's idea has sparked a revolution in how the world at large is approaching the future of species identification. Within two-and-a-half years, the Consortium for the Barcode of Life has been established at the Smithsonian Institution's National Museum of Natural History in Washington, D.C. More than 100 organizations in 39 countries have pledged to barcode as many of the world's

suspected 100 million-plus species in the next few years.

A significant impetus for change is that the classic Linnaean formula has categorized just 1.7 million species to date. By comparison, the Guelph laboratory alone will be able to barcode 100,000 species a year.

The question today is how the new Linnaeus feels about his personal transformation. And where does he think the field is headed? What remains to be done and what is in doubt?

On a snowy December day in Guelph, Dr. Hebert answered questions while trying to pack for a trip to Australia to collect Lepidoptera (butterflies and moths). His office bulges with a lifetime of papers and books, the walls covered with 19th-century lithographs of various animal lineages (a collage of plankton images is Dr. Hebert's favorite). What is most noticeable about the greying, balding 58-year-old "paradigm shifter" is his energy. Whatever's the opposite of burnout, that's what Dr. Hebert exudes, crackling as he talks, laughing at his own jokes, understanding the unexpressed underpinnings of questions without explanation.

Before we get into what you have done, could you give me a notion of your background?

The original article may be found at:
http://www.universityaffairs.ca/issues/2006/april/barcode_life_01.html

Dr. Hebert: You mean what is my ontogeny? I grew up in Kingston and did my undergraduate work at Queen's. Unlike practitioners in some other sciences, many people in my area of business become passionate about life's diversity very early on. I, for example, began to collect Lepidoptera – butterflies and moths – when I was about six years old.

Moth collecting doesn't sound very DNA-oriented.

Dr. Hebert: Truly. I thought that I would become an ecologist, but then I decided to focus on evolution. However, there was no place in Canada to study that, so I repositioned to Cambridge where I quite unwittingly joined a highly molecular lab. All of a sudden the boy who loved collecting butterflies found himself working on alcohol dehydrogenase in fruit flies, and I thought almost immediately, "Oh my god, this was not what I was meant to do." But in retrospect, maybe the one thing that I was supposed to do was become exposed to molecular technologies.

What then?

Dr. Hebert: I was able to flip out of fruit flies and into water fleas, and through them built a tiny reputation in the evolution of asexuality. What happens to organisms when they give up sex? That was the focus of my doctoral work, and then I left England and went to the University of Sydney and spent four years in Australia advancing understanding of tropical water fleas. I then came back to Canada to study polar water fleas because they have wholly abandoned sex. Until barcoding became my preoccupation, I sustained a small cottage industry focused on using molecules to probe breeding systems and other biological attributes of aquatic life.

What inspired you to attempt a DNA species barcode?

Dr. Hebert: In a word, technological change. When I arrived in Guelph from the University of Windsor in 1990 to do genetics research, I spent much of my start-up funds buying ultra-centrifuges – not to mention that I needed five grams of material to do anything. That prohibited DNA analysis on most life, and even when possible, work was painfully slow. But in the last few years there have been incredible developments, particularly when you can amplify any target gene using what's known to scientists as the polymerase chain reaction. So I guess it was 1998 or 1999, and I was looking at what technology had wrought, and decided that it really was feasible to do species DNA analysis on a large scale.

That is when I decided to reanimate my interest in Lepidoptera on the home front. I deployed some ultraviolet lights in my backyard and began to barcode-test the moths that I could find there. Not many know this but my first barcoding paper reported results from 200 species of Lepidoptera captured in my backyard in Guelph.

You weren't the first to think there could be a species signature in DNA?

Dr. Hebert: Absolutely not. Many, many researchers had recognized this, but it was all very laissez faire on the genomics front. Everyone picked a gene region that was convenient to them – the front end of rDNA here, the back end of something else there. Well, the genome is a big place and this was leading to a Tower of Babel in terms of genetic information. Our major contribution was saying that if you want to accomplish this task in a cost- and time-effective fashion, you have to focus on a single gene region. However, at the time I don't think there was anyone who supposed you could take a single gene region and use it to effectively tell apart nearly all animal species. Nobody anticipated that.

Why did you choose a gene in the mitochondrion, which is outside the cell nucleus, to base a species barcode on?

Dr. Hebert: The obvious place you might think to begin is the nuclear genome, but it is an incredibly complex place, a very big place that is easy to get lost in. Think of it as a genetic New

The original article may be found at:
http://www.universityaffairs.ca/issues/2006/april/barcode_life_01.html

York City but, as opposed to its urban counterpart, the nuclear genome is very slow-moving and changes are too sluggish to give a good reading about species differences. On the other hand, you have this tiny little gene town – the mitochondrial genome is really just a little loop of DNA – whose sequences change very quickly. So my thinking was if you seek combinatorial diversity in DNA, go to a small place where things happen quickly.

Isn't it also true that if you were looking for a place where a species signature might exist, not to mention speciation might occur, a gene in that part of a cell which controls cell energy would be a logical place to look?

Dr. Hebert: Not a bad place to look at all. I can think of many other genic regions that would be worse candidates than the prime energy generation region of life, but it would be a bit presumptuous to say that little gene bit has played a key role in speciation. You could say it if you wanted to but not me. I have enough nails through my hands from critics to not want to go in that direction just now.

Was crucifying you the general response from the community when you started presenting your barcoding results?

Dr. Hebert: Well, I would say some people reacted in a “wow if that works, it is really, really important” way. These were people who had some exposure to genetics and knew its impact on many areas of biology, but who were not sufficiently immersed in molecular evolution to be concerned by potential roadblocks that led many senior evolutionary biologists to adopt positions that were beyond skeptical. I will not name names, but early on I gave a talk at the University of Toronto, and afterwards one of Canada's most eminent evolutionists walked by another of his colleagues and said in a not-so-quiet voice: “What a load of shit that was. There is no way that it can work.” That was a pretty common response, a resistance that is also shown by the fact that I have never submitted a paper on DNA barcoding that hasn't seen initial rejection. Even today, if you mention the words DNA barcode in front of some biologists, it is akin to waving a red flag in front of a bull.

Why so much resistance?

Dr. Hebert: Well, there is a typological ring to barcoding. If it were really and truly the case that each species was indelibly barcoded with an invariant DNA marker, that would have some very drastic consequences for theories of the origin of life. If there was only significant variation between species – and none within species – that could imply creationism.

I will come back to that. How did you come up with the term DNA barcoding initially?

Dr. Hebert: Serendipity. I was striding through the supermarket one day, as everyone does, and I was suddenly taken aback by a simple observation: You have all these products on the shelf and they have these codes which identify them. But these codes are incredibly short – just 11 different numbers. If short numeric strings were able to identify every supermarket product, shouldn't there be a similar combinatorial diversity in DNA which tells species apart?

If barcoding was so controversial, how the heck did you go from zero to where you are today?

Dr. Hebert: Again, serendipity. In 2002, I went to this modest-sized conference in eastern Canada, organized by the Census of Marine Life, and presented my thinking and some very early findings. A representative of the New York City-based Sloan Foundation, Jesse Ausubel, happened to be at this meeting. I think that the rest of the audience thought it was an interesting presentation, but Jesse came up afterwards and said, “we have to talk. If what you said is true, this could be an immense advance in biodiversity science. I want you to visit my colleagues at the foundation.”

The original article may be found at:
http://www.universityaffairs.ca/issues/2006/april/barcode_life_01.html

Well I couldn't immediately because I had to be in Australia the next week, but a month later I was in New York City. And three days after that the foundation had invested in bringing people together at a couple of conferences in 2003, all in advance of the first paper we had being published.

What were those conferences like?

Dr. Hebert: I didn't really have much data, so I pitched the combinatorial [barcoding] idea and the fact there should be plenty of signal there and that I have looked at all of 300 specimens and things seem to be working. But Jesse assembled a lot of heavy hitters, and after the 2003 conferences he said there was enough traction and enough scientific support, so he encouraged a small group of us to assemble a proposal. Within weeks, the Sloan Foundation committed funding, nearly a million dollars, to establish the Consortium for the Barcode of Life that is now based at the Smithsonian Institution in Washington, and we did get moving.

Now, the Gordon and Betty Moore Foundation, Genome Canada, NSERC, the Canada Foundation for Innovation, the Ontario Innovation Trust and others have committed millions. How does this happen?

Dr. Hebert: Find a good friend in the U.S. and write proposals. Jesse has been an ardent supporter. He has traveled to Ottawa on behalf of barcoding, emphasizing its importance and the fact that it was born here. Not only has he been endlessly enthusiastic, but also foundations such as the Sloan have the ability to make decisions with remarkable speed. The early funding from the Sloan Foundation added serious credibility and organizational capacity to the barcode movement. This in turn spurred a major [\$3 million] investment from the Moore Foundation that allowed us to develop and run a barcode factory in Guelph. In turn, this enterprise spurred an alliance of many Canadian researchers with interests in biodiversity science, and allowed us to develop compelling proposals that have now gained support from varied Canadian funding agencies.

Does it surprise you how quickly things have taken off?

Dr. Hebert: It's quite bizarre and I shake my head occasionally and say, "I can't really believe this has happened," especially when – one must say – taxonomy is not the place where you go when you are looking for a fast response to anything.

What are the personal impacts?

Dr. Hebert: I guess I always knew science is prepared to consume you, but it really wants to consume me at this moment in time. In the past I set my own schedule and was relatively regimented and rigorous about dealing with things in the order I wanted. Now contingencies rule my life. It is very, very hard to predict my immediate future because this meeting comes up, and that one, and on and on.

What motivated you to do this in sort of the after-bloom of a career?

Dr. Hebert: You mean did I have a mid-life crisis or something like that? Well the analogy I would make would not be to a mid-life crisis but to prospectors who spend the better part of their lives pounding rocks and hoping they will find diamonds or gold. In my 50s, it looks like I found the species vein of gold I have always been looking for.

Scientifically what has most surprised you about what has been happening?

Dr. Hebert: What's really surprising is that a few years ago I never thought that I would be whining about a lack of animal diversity on this planet. The fact is we are going to finish

The original article may be found at:
http://www.universityaffairs.ca/issues/2006/april/barcode_life_01.html

barcoding things. Within five years we are going to have codified the majority of lepidopteran species on this continent, as well as all of the bird and fish species on this planet. If you can do it for these groups, you can do it for beetles and likely for any animal group. So, the sort of depressing thought today is that there isn't enough biodiversity to keep us very busy for very long.

Let's return to creationism again. You said that one of the surprising things is that there is so little difference between individuals in a species.

Dr. Hebert: The explanation, when this lack of mitochondrial diversity showed up in humans, was a population bottleneck – the collapse of our species to perhaps 1,000 individuals about 150,000 years ago in Africa and out of that bottleneck modern humans emerged, stripped of mitochondrial diversity. But am I supposed to believe that every aphid species on the planet, every water flea, every fish went through a similar bottleneck at the same time? That is impossible. Honestly, I sometimes wake up at night and think, "If I was a creationist, this would be my proudest moment."

Since you are not a creationist, what do you think is happening?

Dr. Hebert: I think that it can be explained scientifically by there being a scouring mechanism that cleanses mitochondrial genomes of their diversity on a regular basis.

But how does natural selection wash away the diversity?

Dr. Hebert: At this point in time, that is a science mystery.

Is there a practical application for species barcoding?

Dr. Hebert: I'll tell you a cool case that we just resolved – the mystery of the pasta mouse. A food processing company sent us a tray of pasta with chicken that had a mouse head in its midst. It turns out much of the frozen chicken used in these products comes from the Far East, and the company wanted to know where the mouse originated so they could tell if mouse heads in pasta were an in-house problem or one that originated elsewhere. Our barcode studies revealed first that it was indeed the severed head from a house mouse, but since mouse DNA shows regional genetic divergence, we could further establish that the errant mouse head derived from Southeast Asia.

Some claim barcoding is going to replace Linnaean classification as the way to determine a species.

Dr. Hebert: There is a community of individuals that are very concerned DNA is going to take over the business of delineating species, but then I am told that there was earlier concern that printing presses would take over from papyrus scrolls. The bottom line is that we think that Linnaeus would be celebrating.

Why are you so sure?

Dr. Hebert: This was a guy whose mission was to codify life on this planet, and to do so he developed the system we are parameterizing today. No matter what some of his present acolytes say about DNA barcoding, I can't see him being upset with anything we do that makes species identification better and quicker.