

International barcode of life brings biodiversity science into the molecular age

To understand how any property of the world is changing over time, one must first be able to quantify it.

Imagine, suggested Dr. Paul Hebert, Director of the Biodiversity Institute of Ontario at the University of Guelph, if we had not been monitoring world temperatures over the past century—we never would have noticed the gradual warming trend now at the forefront of climate science and policy. Or if we had not been monitoring the ozone layer—how much damage would have been done before the consequences forced us to take notice?

Unlike climate scientists, researchers concerned with biodiversity and the ways human society impacts and interacts with the rest of life on Earth have lacked access to an easy way to quickly identify species and monitor changes in their distribution and numbers.

Traditional taxonomy, the science of identifying and classifying organisms based on their appearance, requires expert knowledge that can take a lifetime to master, for even a small segment of the world's species. Even today, after 250 years of traditional taxonomic research, scientists have only characterized a small fraction of all species. And even for familiar, well-characterized and cataloged species, humans often encounter them in forms that tax traditional identification schemes, such as eggs, larvae, bits of bone or other body parts, and processed pieces.

An ambitious step

An ambitious international project is now underway to bring species identification into the molecular age. Called the International Barcode of Life (iBOL), the project aims to both simplify and democratize biodiversity monitoring by building a massive database, available online to any interested party, of DNA 'barcodes'—short, standardized regions of the genome derived from expert-identified reference specimens that can be used to accurately identify the species of unknown biological samples.

From idea to reality

The idea of using DNA to survey biological diversity first emerged during the late 1980s and early 1990s, when large-scale gene sequencing became feasible, but early work focused on life such as bacteria that lack clear morphological differences. The development of a DNA-based identification system for animal life was provoked by a paper written in 2003 by Dr. Hebert and his colleagues, which identified a gene region that they believed fit the bill for animals: a gene called the cytochrome c oxidase subunit I (COI) sequence.

"That really set off an explosion" of work in barcoding, said Jesse H. Ausubel, iBOL Board Chair and Director of the Program for the Human Environment at The Rockefeller University in New York City. Although a healthy skepticism existed that the COI sequence would be effective as more researchers tested it in the field, recounted Ausubel, more than a dozen loosely organized campaigns sprung up to test the sequence and start barcoding diverse groups of animal life, including ants, fish, birds, and marine organisms.

In 2008, an international team led by Dr. Hebert applied to establish iBOL as an International Consortium Initiative through Genome Canada, with the goal of assembling 5 million barcodes representing 500,000 species by September 2015. This number "will allow us to use barcoding to identify the vast majority of species that humanity comes into regular contact with," explained Dr. Robert Hanner, Associate Director for the Canadian Barcode of Life Network.

"This will be the first time humanity will be able to 'read life' on our planet," agreed Dr. Hebert. "Our goal is a bar-coded world."

The iBOL project was officially activated in September of 2010, with the participation of researchers and biodiversity organizations in 26 coun-



Dr. Paul Hebert
Director, Biodiversity Institute of Ontario,
University of Guelph

"This will be the first time humanity will be able to 'read' life on our planet."

tries spanning all habited continents. In addition to data streaming in from the official iBOL nodes, the project will continue to collaborate with smaller international projects already underway. The barcodes collected by iBOL and its partners are being deposited in the Barcode of Life Data Systems (BOLD) database (www.boldsystems.org) as well as in GenBank (www.ncbi.nlm.nih.gov/genbank/), both of which are accessible by any interested party.

Just over a year into the project, iBOL team members have added almost 150,000 species to the BOLD library, mainly derived from expert-identified reference specimens. Some early kinks in the process have also been worked out. While the COI gene is highly effective for animals, two alternate barcode markers were selected for plant life in 2009 that can distinguish 70 percent of plants. Additional genes can be used to discriminate the 30 percent that remain ambiguous, explained Ausubel, though using more genes when required makes the process more expensive and time-consuming.

Another iBOL working group has recently selected the regions of the genome that will best identify fungi, one of the less-studied and less-understood groups of organisms on the planet.

"Early success in barcoding animals is driving the need to establish broad taxonomic coverage, to dramatically expand the range of applications barcoding promises. That's what the whole iBOL project is here to accelerate," said Dr. Hanner. But while this infrastructure is assembled, some early interest from industry and government bodies shows the promise of barcoding-based biosurveillance in an increasingly globalized world.

The utility of a barcoded world

Barcoding grabbed the headlines in May of this year, when the New York Times reported that DNA barcoding had indicated that up to 25 percent of fish sold in North America and Europe was mislabeled—either by accident as the fish moved through the global supply chain, or by purposeful fraud.

Preventing market fraud is a real and immediate application of barcoding technology, said Dr. Hebert—so much so that the Canadian Food Inspection Agency (CFIA) is currently buying DNA sequencers to implement barcode-based surveillance of the food supply throughout Canada.

In another study published in July of this year in *Nature Scientific Reports*, researchers found that up to a third of herbal teas sold in supermarkets in the United States contained ingredients not identified on the labels. Barcoding was able to not only identify the mystery components, but to determine the geographic origin of some of the ingredients.

The Biodiversity Institute of Ontario has already been asked to help with several incidents of tracing contaminants in the food supply. "People are increasingly sourcing raw ingredients from different vendors in different regions of the world—they need to know where contamination and fraud is happening," said Dr. Hanner.

A related application that Dr. Hanner sees potential for is the interception of agricultural pests, which are rife and potentially devastating, with the



introduction of invasive species causing considerable environmental and economic damages in Canada each year. "Published research has shown that some 60 to 90 percent of specimens intercepted for identification by the United States Department of Agriculture and their Japanese counterparts cannot be identified—they're encountering eggs, larvae, or fragmentary remains, or maybe a pest so esoteric that only a few people on the planet even know how to identify it," he elaborated. "So this is a documented demand—building a library these people can use for rapid, accurate species identification is really important."

The Pacific Forestry Centre of Natural Resources Canada has been experimenting with barcoding for identifying introduced and invasive species almost since the idea's inception, said Dr. Leland Humble, a research scientist with the Pacific Forestry Centre and a member of the Canadian barcoding network. Dr. Jeremy deWaard, a graduate student he supervised who had already participated in the initial development of barcoding technology, helped to make barcoding a regular tool in the Centre's work.

Dr. Humble's team has made several interesting discoveries using barcoding to detect introduced species in forest habitats. In one survey, they identified an insect that attacks juniper tree leaves that had been present in British Columbia for more than three decades, but had been misidentified by traditional taxonomic methods when previously collected. Another insect that bores into cottonwood and poplar trees was found to have been misidentified for more than 40 years.

Even in Stanley Park, Vancouver, barcoding helped the team identify four introduced species that had not previously been recorded in British Columbia, "and a couple were new to North America," said Dr. Humble.

"Barcoding detected previously unknown invaders, and this might give us the chance to eradicate them before they become established and save us from their financial impacts, which can be staggering. Early eradication of just one invasive species could more than offset the cost for building the entire global barcode reference sequence library. In this respect, barcoding is perhaps the biggest science bargain on the planet and a shrewd investment by Genome Canada," commented Dr. Hanner.

Barcoding's ability to identify species from eggs and larva has proven particularly helpful to the Forestry Centre, such as in an ongoing project to track the distribution of a recently introduced sawfly. Sawflies can be pests in both forest and agricultural settings. "It's quite difficult to be everywhere at once and get adults that are readily identifiable morphologically, but it's easy to collect larvae," explained Dr. Humble. "You get a much longer time frame in which to work."

Another application that the iBOL team is rolling-out is the use of bar-

coding in environmental monitoring—identifying organisms to assess environmental conditions.

Dr. Peter Miller was recruited to the Biodiversity Institute of Ontario and iBOL from the Southern California Coastal Water Research Project (SCCWRP) to serve as Director of Barcode Applications, to help interested parties from industry with using barcoding to understand how their activity is altering the environments in which they work.

"DNA barcoding for species identification for use in environmental bioassessment is brand-new, cutting edge," said Dr. Miller. "The trend in biomonitoring has more and more been to start looking at populations of organisms living in an environment and assess whether or not there might be some stress [from human activity], but identifying organisms is often very difficult."

In his position at SCCWRP, Dr. Miller realized that barcoding might be a useful tool to meet their environmental monitoring objectives. His interactions with the team at the University of Guelph eventually led to his being recruited as a liaison with interested parties in the private sector. "One of my jobs involves contacting companies whose activities impact the environment and showing them how they could be using DNA barcoding as an effective biomonitoring tool," explained Dr. Miller.

Several environmental consulting companies have already sent environmental samples directly to the Biodiversity Institute for barcoding. One such group, explained Dr. Miller, is working with a company that is extracting gravel from river beds, and they need to assess the effect of the industry on fish and fish reproductive success. "Identifying fish eggs and larvae to species is difficult or impossible, so they wanted to use DNA barcoding, which is well-suited to identifying all life stages of organisms," he elaborated.

Other consulting companies—which provide assessments for a wide variety of industries including mining, petroleum, forestry, power generation, and sewage treatment—have enquired more generally about barcoding for species identification for use in environmental assessment, he continued. If one can count the number the species in an environmental sample, that information can be plugged into mathematical models that can assess how impacted an environment is by industrial activity.

"Enumerating species in samples is the state-of-the-art for environmental assessment, but the difficulty is that it takes considerable taxonomic expertise to identify organisms, as well as months of effort," said Dr. Miller. "DNA barcoding can provide more accurate identifications with finer taxonomic resolution, and is much faster—a week or two instead of months." This barcoding application will be predicated on building the reference library in part-

nership with taxonomic experts.

Encouraging to Dr. Miller and the rest of the iBOL staff has been how positively people working in traditional taxonomy have responded to the new technology. "As we initiated some of our early DNA barcoding monitoring projects, we feared that the traditional taxonomists we were working with would view barcoding skeptically, and perhaps feel threatened by the new technology. Their response has actually been very different from what we expected; they've been excited, saying that 'if only we'd had this tool 10 or 20 years ago, it would have saved us so much hard work.'"

Part of this excitement stems from the new knowledge of the complexity of diversity of species and their tangled evolutionary relationships that barcoding has already provided. "It's highlighting areas where we need to invest more heavily in clarifying species boundaries—we're finding cryptic species using barcoding in everything from fishes to insects, birds, and mammals," said Dr. Hanner. (Cryptic species are animals—or plants or insects or other forms of life—that look virtually identical but are genetically quite different, to the point where they cannot interbreed.)

But on the flip side, said Ausubel, barcoding has confirmed that the idea of species remains valid; they've seen a strong 'barcode gap' between nearly every species cataloged so far. Within a species, you might have 1 percent variation in the genetic code of the COI gene, so maybe 1 to 6 of the 600 DNA nucleotides are different. "Then there's a gap, and the next species will have 30 to 60 different nucleotides. Barcoding has been a way of quantifying the difference between species," he explained.

Tools for reading life

Looking to the future, Dr. Hebert hopes to see technology catch up with the volumes of information being funneled into the BOLD database. "We anticipate a world where there are two kinds of access to the iBOL data," he said. One would be a hand-held device that could provide an instant barcode identification for any tiny sample fed into it; not just a Star Trek concept, as the time for sequencing an entire human genome—over 4 million times larger than a single barcode—has fallen to just over a week, and continues to fall monthly. The second would be laboratory machines that could reveal the diversity of organisms in a large environmental sample—for example, 5 cubic meters of water at once.

"We don't really know what's happening to life on this planet right now," concluded Dr. Hebert, "but that's where we're headed—automated systems for reading life, on scales we can't even imagine today."