GENEWATCH

SMUGGLERS, POACHERS, AND DNA BARCODING
By CRG staff - interview with Mark Stoeckle

Mark Stoeckle, MD, is an Adjunct Faculty Member in the Program for the Human Environment at The Rockefeller University. He has been involved in the DNA Barcoding Initiative since its beginnings in 2003. DNA barcoding is a technique for identifying species from DNA samples using a short genetic marker at a standard and agreed-upon position in the genome.

**GeneWatch**: What do you need in order to scan an organism’s DNA barcode? How easy is it to do?

**Mark Stoeckle**: DNA barcoding is just a simple, standardized way of identifying species by DNA. With animals, for instance, you analyze one specific gene region and you try to match that sequence to your reference library. That makes it important to have a really good reference library—and that's where most of the effort is, in building that library, so that when you get a DNA sequence from a sample there is something in the library to match it to. It's like a database of fingerprints: you need fingerprints in the database in order to identify your sample.

The sequencing technology is pretty simple, and it's getting simpler. It still requires a laboratory, but I can imagine it will only get easier. That part is a straightforward technology. You can get a result in about a day, and it could be faster in the future. Then you match that result to the library just like you would use a Google search to look something up. Since these are public databases, you're just entering the sequence and seeing what in the database it's matched to.

**GeneWatch**: With people and labs all over the world doing this, how do you make sure everyone is looking at the right location on the genome?

**Mark Stoeckle**: It's a standardized approach, so the idea is that everybody is going to use the same gene region. There's no rule exactly; it's kind of a social agreement among scientists that it makes sense, if you're analyzing a new species or going through your museum specimens, to analyze this specific gene or portion of the gene and to put that into the public databases. There's a social agreement that for DNA barcoding, we all use the same gene region.

**GeneWatch**: What is the margin of error? Is it very easy to confuse two very similar species, or would an unexpected mutation make it harder to identify species?

**Mark Stoeckle**: For animals, I would say something like 95% of the time it's very straightforward and there is no ambiguity about the boundaries between species. The groups that you would get from just doing the DNA sequence...
alone fall under sets of similar sequences, and those turn out to match very closely one-to-one with the same groups that biologists have identified as being the species. So it's amazingly close.

There are maybe 5% of cases where two species are genetically very similar, and there it might be harder but not impossible to tell them apart; and there are some cases of organisms that biologists call different species, but in this particular gene region they are identical. Biologists call them two different species, maybe they don't interbreed, but by analyzing this gene region there's just not enough information to tell them apart. You know it's one of the two species, but you don't know which one it is.

Mutations haven't turned out to be much of an issue. For instance, if you were to analyze this same region of the genome among people, any two individuals might differ in one or two positions in the barcode; but people differ from chimpanzees by about 50 positions. So you wouldn't be confusing one with the other.

So, in practical terms, it's very good. The limitation right now is the library. There are lots of species—there are two million named species of plants and animals, and the most recent estimate is there are another eight million that we haven't named yet.

So DNA barcoding can be used to identify species, but it doesn't go beyond that.

Right, it's not good for identifying individuals. Obviously crime labs use DNA to identify individuals, but you have to analyze more gene regions in order to do that. The goal here is to make it as simple and small as possible, accepting the fact that in the few percent of cases you're not going to be able to distinguish closely related species from this gene region alone.

DNA barcoding has been used to identify bushmeat species and seafood and even tea. How far can it be taken? How about, say, alligator boots?

You know, DNA is an amazingly hardy molecule. Scientists have recovered DNA from very ancient specimens that are tens of thousands of years old. More recently, people have tried things that are very processed—leather is certainly one of those. I think we're just beginning to look at that. I know that George Amato's group at the American Museum of Natural History has retrieved DNA, and specifically DNA barcodes, out of leather products—information that the U.S. Fish and Wildlife Service has then used in prosecution of people importing products made from endangered species.

So is DNA barcoding generally being accepted as evidence in court?

The FDA, as recently as last fall, published DNA barcoding as their official method for seafood identification, and the FDA does investigate seafood fraud. That's the first government agency that I know of that has said, "This is our legal standard," but I think that's going to increase. FDA is a model for agencies in other countries, and I know that other government agencies like USDA are looking at this. INTERPOL is also looking at it for detecting commercial fraud. In individual cases, the U.S. Fish and Wildlife Service has certainly used it in court. So yes, it has been used in court cases, and I think it will get adopted by more agencies.

What's the alternative? What's it replacing? What is, say, USDA using instead of DNA barcoding?

They were using something called isoelectric focusing and protein electrophoresis for identifying seafood. It's really not a very robust method. And I think for a lot of seafood, there just isn't any other method besides DNA barcoding. Once you cut up a fish, you don't know what it is. Once you cut a fin off of a shark, no one can identify it. In that area, barcoding is a completely new technology.

How is DNA barcoding proving most useful for conservation purposes?

I think we're at the beginning of the practical uses of it. Major uses are trade in products of regulated or endangered species, such as fish and bushmeat, where you need to be able to identify which species the product comes from. Most of those samples are from a product in the form that people use—they're hard to identify because they're cut up into pieces or processed in some way.

I think another area where barcoding could be useful in conservation is for conducting biosurveys—trying to figure out what lives in a certain area. Say you get a thousand samples of invertebrates and you send the crickets to a cricket specialist and the moths to a moth specialist to try to figure out what they are. Instead, you could run the DNA on all of them and you wouldn't need an expert. It would potentially be an easier, cheaper and faster way to do biosurveys. Barcoding is already being used that way for freshwater quality surveys. The best indicator of the health of a watershed, what's most sensitive, is the life forms in the pond or the stream. And those are hard even for experts to identify, so that's where people are using DNA barcoding.

It has also been mentioned as a potential tool for invasive species. How would that work?

One of the ways we get invasive species is in ballast water in ships. You're supposed to be checking the ballast water to make sure it doesn't have certain species in it. That's hard to do, but it might be easier to test it by just taking a water sample—don't even look at what's in there—just sort of spin it down and get some DNA out of it.

Another way that's just starting to be used with water, along the same lines, is to not try to collect organisms, just collect water. For instance, in Europe, the American bullfrog is an invasive species. If you want to know if there are bullfrogs in the pond, you can just collect a water sample and see if there is American bullfrog DNA in the water sample. Again, that's capitalizing on this ability to use very small amounts of DNA.

You have worked with high school students to use this tool for some conservation sleuthing—first "Sushigate," when you helped your daughter and a friend uncover mislabeled sushi, and now with the Urban Barcode Project. That raised a question for me: How easy is it for someone to do their own DNA detective work?
The Urban Barcode Project is a really fun project, run by Cold Spring Harbor Laboratories. It's a competition among high schools in New York City, mainly public schools, to use DNA barcoding to do an investigation that they think is interesting.

It's pretty simple. In this project, the students are thinking of what they want to know and they're collecting the samples. They then bring them to a laboratory in a classroom that's set up with the right equipment, where they go through the steps to isolate the DNA and amplify the barcode gene. The equipment costs a few thousand dollars—it's not ten thousand or a hundred thousand—and students can sort of walk in and, if they have someone supervising them, they can do it on the spot. It doesn't require extensive training. Then the samples are sent to a lab that does sequencing, sort of like how we used to send film for processing. So the students send the DNA to lab to do the actual sequencing. It's really pretty simple. It's not a toaster yet, but it's getting there.

So you send out the sample with the DNA isolated and the barcode section amplified, the lab sends you back the sequence, and you go and match it against the reference library.

Right, and that's done on the Internet. They send you sequence, and the mapping is something you do just like you'd do a Google search, using GenBank or the Barcode of Life database.

There's nothing magical about DNA barcoding. It aims to be a toaster: a technology that you don't have to read the instructions to use. We're not quite there yet, but the basic principles of it are there. The work for the scientific community is to build up the reference library, because the method is only as good as the library. High school students like my daughter were able to see that sushi labeled as white tuna was actually tilapia, and they could only do that because researchers had deposited sequences of tuna and tilapia in the public database.

The education potential is very big. High school students—anybody, but say high school students—can discover things that no one else knows. Most science projects, the teacher knows the answer; but with this technology, students can think of an investigation, collect the samples, and until you do the DNA investigation, you really don't know what you have. And that's just such a fun thing ... it's discovery.