LECTURE # 17

BAR CODE OF LIFE

THE BARCODING PIPELINE
BAR CODE OF LIFE

In 2003, Paul Hebert, researcher at the University of Guelph in Ontario, Canada, proposed “DNA barcoding” as a way to identify species. Barcoding uses a very short genetic sequence from a standard part of the genome the way a supermarket scanner distinguishes products using the black stripes of the Universal Product Code (UPC). Two items may look very similar to the untrained eye, but in both cases the barcodes are distinct.

Until now, biological specimens were identified using morphological features like the shape, size and color of body parts. In some cases a trained technician could make routine identifications using morphological “keys” (step-by-step instructions of what to look for), but in most cases an experienced professional taxonomist is needed. If a specimen is damaged or is in an immature stage of development, even specialists may be unable to make identifications. Barcoding solves these problems because even non-specialists can obtain barcodes from tiny amounts of tissue. This is not to say that traditional taxonomy has become less important. Rather, DNA barcoding can serve a dual purpose as a new tool in the taxonomists toolbox supplementing their knowledge as well as being an innovative device for non-experts who need to make a quick identification.

The gene region that is being used as the standard barcode for almost all animal groups is a 648 base-pair region in the mitochondrial cytochrome c oxidase 1 gene (“CO1”). COI is proving highly effective in identifying birds, butterflies, fish, flies and many other animal groups. COI is not an effective barcode region in plants because it evolves too slowly, but two gene regions in the chloroplast, matK and rbcL, have been approved as the barcode regions for plants.

**Barcoding projects have four components:**

The Specimens: Natural history museums, herbaria, zoos, aquaria, frozen tissue collections, seed banks, type culture collections and other repositories of biological materials are treasure troves of identified specimens.

The Laboratory Analysis: Laboratory protocols (pdf; 400Kb) can be followed to obtain DNA barcode sequences from these specimens. The best equipped molecular biology labs can produce a DNA barcode sequence in a few hours. The data are then placed in a database for subsequent analysis.

The Database: One of the most important components of the Barcode Initiative is the construction of a public reference library of species identifiers which could be used to assign unknown specimens to known species. There are currently two main barcode databases that fill this role:

The International Nucleotide Sequence Database Collaborative is a partnership among GenBank in the U.S., the Nucleotide Sequence Database of the European Molecular Biology Lab in Europe, and the DNA Data Bank of Japan. They have agreed to CBOL’s data standards (pdf; 30Kb) for barcode records.

Barcode of Life Database (BOLD) was created and is maintained by University of Guelph in Ontario. It offers researchers a way to collect, manage, and analyze DNA barcode data.

The Data Analysis: Specimens are identified by finding the closest matching reference record in the database. CBOL's Data Analysis Working Group has created the Barcode of Life Data Portal which offers researchers new and more flexible ways to store, manage, analyze and display their barcode data.