

Telescopes, Microscopes, Macroscopes and DNA Barcodes

Opening Talk

“Using Barcode Data in Studies of Molecular and Evolutionary Dynamics”

The Banbury Center, Cold Spring Harbor Laboratory, 28-31 October 2007

Jesse H. Ausubel

Program Director, Alfred P. Sloan Foundation

ausubel@rockefeller.edu

<http://phe.rockefeller.edu>

Science is now celebrating the 400th anniversary of the start of the diffusion of a pair of mighty innovations, the telescope and its spawn the microscope. The aggressive Galileo is of course most famously associated with the early telescope, a device literally to see far. Astronomers understood they needed instruments with qualities beyond the human eye to observe remote objects. Distance makes objects small, so it was natural that 17th century lens makers would turn their powers not only to remote stars, planets, and moons but also to objects that were close by but also very small, even so minute as to be invisible to our naked eyes.

During centuries of vigorous improvements, both the telescope and the microscope brought societies practical benefits. With versions of telescopes, armies could spy movements of their enemies and signal their allies, and naturalists could enjoy birds through binoculars. With microscopes, medical workers could identify pests and jewelers could assess gems.

Historians of science such as the late Derek de Solla Price have meanwhile impressed us how science itself follows the technology of observation. Galileo's revolutionary views exemplify how new understanding follows technology. Galileo discovered not only the mountains of the moon but a method of artificial revelation that enlarged what was to be explained by science. As de Solla Price observed, techniques are used in every which way in the hope that the finding out will reveal facts of nature that fall outside of the range of what was known before. Historically the arrow of causality is largely from technology to science. It is only mild hyperbole to say that microscopy has driven the advancement of biology.

In 1975, a French writer and systems analyst, Joel de Rosnay, introduced a new word to complement the telescope and microscope, namely the macroscope. De Rosnay identified the macroscope with the computer. Just as some phenomena are too far or too small to see unaided, so some phenomena are too large or complex. For these we need macrosopes.

An example that Alfred P. Sloan Foundation President Ralph Gomory offers is the economy of a nation or the world (<http://web.mit.edu/ipc/sloan05/gomory.pdf>). To observe these exceedingly large and complex objects, economists and others have developed statistical databases such as the systems of national income and product accounts that allow us, for example, to measure national economic growth and perceive

patterns and trends at scales that a single observer could never achieve. If the economic data are the stars in a telescope or the cells under a microscope, then the computer and other instruments of information handling are the macroscope.

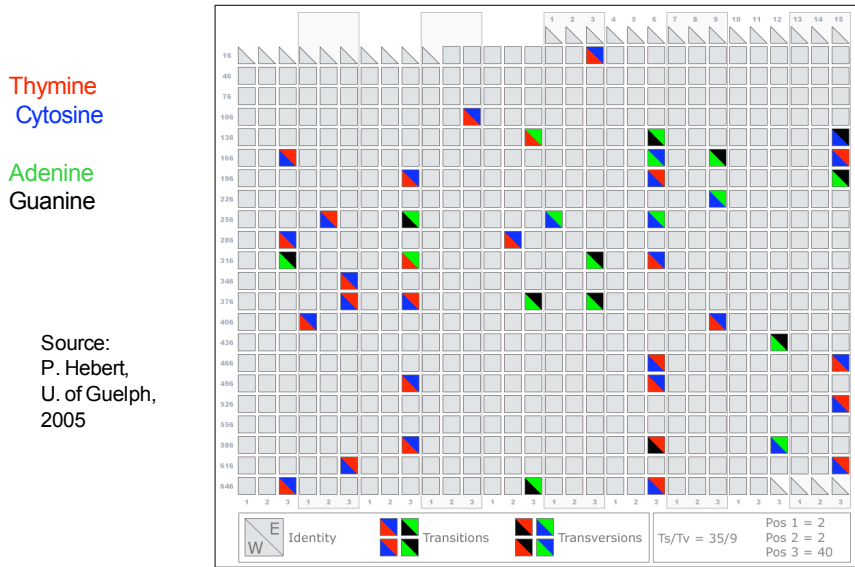
Following economics, meteorology, and other fields, biology has begun during the past 20 years or so to create large databases that permit our ever more powerful macrosopes to go to work. Genomes provide the foremost examples. Computational biologists have tended to follow a strategy of vertical genomics, that is, learning deeply about one genome, and then determining, for example, how many genes a human or a mouse has. We have used macrosopes for comparisons, too, for example, to identify differences between the human and chimpanzee genome.

An exciting macroscopic initiative, a cousin to DNA barcoding, is the Encyclopedia of Life project (www.eol.org), which will create a web page for all 1.8 million or so named animals, plants, and fungi over the next 5-10 years. Like DNA barcoding, the Encyclopedia of Life will serve as a macroscope for comparative and evolutionary biology. Imagine, for example, exploring lifespan across tens or hundreds of thousands of species, an opportunity the Encyclopedia of Life should provide.

In 2003 here at Banbury the biology community first appreciated the power of short DNA sequences from a uniform locality on the genome for identifying the species of a specimen, even a tiny fragment, reliably and cheaply. We now also appreciate DNA barcoding opens the field of horizontal genomics on a massive scale. For me the image that demonstrated the macroscopic potential of DNA barcoding was the University of Guelph lab's display of transitions and transversions between

the nucleotides of the COI barcode regions of the Eastern and Western Marsh Wren (Figure 1).

Diagnostic Nucleotide Characters between Eastern & Western Marsh Wren
 Imagine exploring alignments in 10 million sequences!



A macroscope applied to the barcode library will allow not only comparison of an Eastern and a Western Wren or a human and a chimp but comparison of thousands, tens of thousands, hundreds of thousands of specimens and species.

An aligned, searchable database of barcode differences of not only two wrens but all birds or all vertebrates or all animals thrills my imagination, in good part because I believe the practical difficulties of working with more than a small number of specimens or species has formed a major limit to knowledge in systematic, comparative, and evolutionary biology. Just as Galileo’s telescope expanded the explicandum of astronomy by orders of magnitude and in many and often fortuitous directions, the barcode database and its macrosopes could take comparative and

evolutionary biology into vast new territories, and those territories are the subject of this third Banbury Conference on DNA barcoding.

In fact, for me the questions for our Conference are essentially two:

- 1) What are exciting macroscopic questions that the growth of the DNA barcode database allow us to explore?
- 2) What are the additional tools for visualization and analysis that we need to perceive and communicate the insights, patterns, and trends we might discover?

Let us accept as a point of departure rapid, even exponential growth of DNA barcode databases. The DNA barcoding movement began only four years ago with a pair of meetings in March and October 2003 here at Banbury, organized by 2007 attendees Scott Federhen and Mark Stoeckle and others who invented the Consortium for the Barcode of Life (CBOL), which has organized this meeting. At that time Genbank had COI sequences of any length or interval for about 6,000 animal species. Few of the sequences would meet the 2007 barcode data standard, but the evidence was good enough to allow the animal community to standardize on COI as its barcode region. Today the Barcode of Life Database, BOLD, reports more than 318,000 specimens with COI barcodes from almost 34,000 species. BOLD's taxonomy browser

(http://www.barcodinglife.org/views/taxbrowser_root.php) shows the initiation of broad taxonomic coverage and impressive spatial coverage on land.

Numerous efforts brighten the prospects for growth of the barcode library. The communities working on birds, marine & freshwater fish, and mosquitoes have aggressive campaigns underway to complete these groups in the next few years. A new marine barcoding effort, not only fish, should

add barcodes for 50,000 or more of the 230,000 or so named marine species by the end of 2010 with extensive spatial coverage of the oceans. The plant barcoding community verges on announcing its strategy, likely involving the use of three DNA segments, a coding region (part of the *matK* gene), and two intergenic spacer regions, *atpF/atpH* and *trnH-psbA*. An initiative to develop the barcodes for the 100,000 or so trees of the world is already funded and about to launch. The fungal community is likely to launch the Fungal Barcode Initiative, FBI, in the summer of 2008, using the non-coding ITS region of the nuclear rDNA and, I hope, a coding region as well, perhaps even COI.

The partial use of spacer regions by the plant and fungal barcoding communities, by the way, may open macroscopic avenues for research about the relative evolutionary importance of indels and single nucleotide polymorphisms. I hope this Conference will explore this and other questions associated with the growth of libraries for non-coding regions.

The main impetus for barcoding may have come from practical needs for species identifications, just as lens makers honed their skills by addressing lucrative markets for eyeglasses and the military. Still, the subject for Banbury III ranges far beyond spectacles that allow highly reliable species identification. The challenge is to prepare the microscope for scientific discovery to explore and probe the massive database now accumulating. The meeting will be a success if attendees, as individuals or teams, leave Banbury with exciting ideas for research, ideas that will become proposals and papers and important discoveries in comparative and evolutionary biology and related fields.

Let me also say what this conference is not. A couple of years ago, agronomist Paul Waggoner and I made a list of 9 worries about barcoding. As barcoding grew, worries disturbed minds, some worries evoked by broad concerns and some prompted by perceptible limits. Some worries concern any DNA taxonomy and some concern only COI barcoding. While time has lessened the worries, I believe it may help to acknowledge the worries and to identify those that should form part of our agenda.

1. Barcoding is reductionist and Orwellian. Biologists struggle for a “feeling for the organism” and to maintain academic departments concerned with holistic and integrative studies, not only molecular and genetic ones. Barcoding reduces species, the fundamental units of biology, to a brief digital code. Will barcoding make humans see other animals and plants as ciphers and devalue them?

I neither believe this worry is being realized nor that our group would contribute much by discussing it.

2. Barcoding denigrates hard-won knowledge. During more than two centuries of painstaking observation biologists have amassed wonderful knowledge about the morphology and behavior of species that can be accessed with the master key of a binomial name. If DNA barcodes become the arbiter of species, will they invalidate the master key of binomial names and waste the amassed knowledge?

To build the barcode library, the barcode movement relies on and proceeds arm-in-arm with more traditional approaches. The fall-out, both positive and troublesome, from the success of barcoding might be a popular meal-

time subject but not important for how to improve and employ the microscope.

3. Barcoding threatens jobs. Publications and taxonomic keys backed by archived and vouchered specimens form the visible capital of knowledge about species. Less obvious capital is in the human form of experience and skill embodied in biologists themselves. An example demonstrates the reality of the human capital: Only examination and confirmation by an expert authority establish the identity of a rare specimen, and especially, only the recognized authority can confirm the discovery of a new species. Will barcoding dissipate this human capital, eliminating occupations for systematists as new technology unemployed blacksmiths and celestial navigators?

On the contrary, we believe CBOL is a huge boost for systematic biology. But the subject of this conference is not the health of the profession except to the extent we may enhance it by adding excitement to the research agenda.

4. Barcoding is a mere slogan and an inadequate analogy. Since the 1970s biologists have been accumulating examples of DNA sequences that differ among groups of species and provide clues to evolutionary connections. Is the phrase “barcode of life” only a stratagem for winning funds by implying, for example, that the accumulating DNA sequences will grant users and managers new powers as in retail enterprises? Does “barcode of life” wrongly imply that species are static? Does it imply that species are invariant and neither the outcome of evolution nor susceptible to future variation?

I believe the success of the analogy is evident from the quantity and quality of public interest, and displays like the Wren comparison are among the most vivid ways to convey molecular evolution.

5. Evidence already shows animal species that COI cannot identify. A threshold of 2.7% between COI barcodes has been proposed for distinguishing bird species. Sequences of another mitochondrial gene were compared among several recently evolved sister species of birds, i.e., species evolved from a common ancestor. Several of the distances among sequences of the other genes in the sister species were shorter than the proposed 2.7% threshold for COI. Will there be enough exceptions to disqualify barcoding at the COI location?

With every month, the craft of barcoding improves, and problematic groups seem to be rarer rather than more frequent. But the microscope is a great tool for isolating and explaining exceptions. Let's consider ways to use the growing barcode of life database to consider the population of exceptions.

6. COI cannot be a universal barcode. Agreement on a uniform system of bars and lines as much as the optics and computers in a supermarket imparts power to commercial barcodes. While a sequence of about 650 base pairs of COI barcode has distinguished species of insects and birds, botanists know that a different gene locality and possibly two or three will be required to identify species of plants. Could so many sequences be required to cover the book of life that the lack of universality of a single, discriminating, short sequence restricts the power of barcoding species?

Evidence is overwhelmingly encouraging that one or few regions suffice, and in fact that once a reference specimen with a full barcode is in the

library, a mini-barcode of only 100 or so basepairs may suffice to achieve identification. As part of this conference, let's inquire why it works so well.

7. It is too soon to standardize on a very few localities. Historically, domains of technology have sometimes locked into an adequate standard such as the QWERTY keyboard that time proved sub-optimal. Should we postpone standardization while much more extensive and systematic testing is carried out on the genome to choose a locality likely to withstand all challengers?

The plant community has now employed systematic testing. In the absence of strong evidence that markedly superior alternatives exist for animals or plants, delay seems unjustified.

8. Barcoding obscures the nuance of the species concept. Biodiversity is not arranged for biologists' convenience. Nature separates some clusters of life by deep valleys and others by shallow ones. Much judgment has gone into giving the rank of species to groups of living things and calling them by name. Mechanical rules based on absolute numbers of nucleotide differences or ratios do not comprehend the bewildering diversity of nature. Will barcoding endanger our nuanced appreciation of species?

The light that millions of barcodes might shed on the species concept seems a wonderful subject for our conference.

9. Barcodes cannot grow a tree of life. While barcoding sprouts leaves on the twigs, resolving the evolution of entire genomes along the branches of the tree of life also employs DNA but typically requires more and longer sequences. Will concentration on the barcodes comprised of short sequences

at a uniform location confuse and retard resolution of the tree of life? Will success by barcoding take resources from other studies needed for older parts of the tree of life?

CBOL has convened joint activities with the tree-of-life community concerned with the trunk and branches, and we look forward to more activities that join trunk, branches, twigs, and leaves. Together with the Encyclopedia of Life, we would like zoomable trees with all named forms of life. A perfect macroscopic challenge for our Conference is how to create and display very large trees.

Before concluding, let me thank all of you for coming and the organizers of the conference, Mark Blaxter, Paul Gilna, Richard Harrison, David Rand, Michel Veuille, and David Schindel, for assembling an exciting program. David provides extraordinary leadership for CBOL, and we are all deeply indebted to him. Jan Witkowski, the able leader of the Banbury Center, is away this evening but will join the meeting tomorrow.

We will miss two expected attendees. Visualization master Paul Gilna from the University of California at San Diego had to cancel on Friday afternoon to manage the recovery of his Institute from the fire threat in that region. Barcoding innovator Paul Hebert was in a serious car accident in Ottawa a few days back. Although Paul sustained only some cuts and bruises, he was advised not to travel for a few days. Happily, his Guelph team is strongly represented. Tomorrow morning I need to attend a meeting in Manhattan at the Sloan Foundation but will rejoin you at lunchtime, and stay until Tuesday afternoon.

In conclusion, let me recur to the telescope. Leonardo da Vinci's notebooks show Leonardo envisioned the telescope one hundred years earlier than Galileo and other innovators usually named. The *Codex Atlanticus* of Leonardo contains drawings and phrases such as "Make glasses to see the moon magnified." Evidence is also quite strong that Leonardo did actually construct the telescope and use it in the garden of the Belvedere in Rome between 1513-1515, while employed by Pope Leo X.

In Rome Leonardo employed two German mechanics, Mastro Giorgio and Mastro Giovanni degli Specchi (masters of the mirrors), whom he appeared not to like or trust. Leonardo wrote "He [Giorgio] never did any work without discussing it every day with Giovanni, who then spread the news of it and proclaimed it everywhere, stating that he was a master of the art, and as regards the part which he did not understand he said that I did not know what I wanted to do" (*Cod Atl* fol 182 v-c). Perhaps anticipating Galileo's fate, Leonardo chose to be secretive, which was anyway his habit.

If Leonardo conceived and built a telescope, then where did he record, or hide, his observations? As a hobby for many years I have assisted Italian physicist, Cesare Marchetti, in studies of Leonardo, and we propose an answer to where Leonardo secreted his celestial observations. A portion of a painting that forms the dark blue fabric worn by the Madonna when magnified a bit and viewed without bias appears instead as a starry night highlighted by Jupiter and its moons, among the most salient phenomena a telescope reveals. More prudent than Galileo, Leonardo safely preserved his results for eternity, we believe.

Like the telescope and microscope, the macroscope can be dangerous by upsetting our world view. Our task for the next few days is to enjoy danger, to identify ways to magnify the power of our macroscope for DNA

barcodes, and to conjecture about attainable knowledge so upsetting that our first instinct, like Leonardo, may be to hide it inside a lady's garment.

Thank you.

Bibliography

Rosnay, Joël de. *Le microscope: vers une vision globale*, Paris, Seuil, 1975.

Solla Price, Derek de. *Little Science, Big Science...and Beyond*, New York, Columbia University Press, 1986.