

NOW IS THE TIME

Daniel H. Janzen

Philosophical Transactions of the Royal Society : in press

The space ship lands. He steps out. He points it around. It says "friendly - unfriendly - edible - poisonous - dangerous - living - inanimate". On the next sweep it says "Quercus oleoides - Homo sapiens - Spondias mombin - Solanum nigrum - Crotalus durissus - Morpheus - serpentine. This has been in my head since reading science fiction in 9th grade a half a century ago. I am sure it was in the heads of Linnaeus, Alexander the Great, and Timid the Mastodont Stomper. And it has been on the wish list of every other human confronted with the bewildering blizzard of wild biodiversity on the edge, middle and center of society.

Imagine a world where every child's backpack, every farmer's pocket, every doctor's office and every biologist's belt has a gadget the size of a cell phone. For free. Pop off a leg, pluck a tuft of hair, pinch a piece of leaf, swat a mosquito, and stick it in on a tuft of toilet tissue. One minute later the screen says *Periplaneta americana*, *Canis familiaris*, *Quercus virginiana*, or West Nile Virus in *Culex pipiens*. A chip the size of your thumbnail could carry 30 million species-specific gene sequences and brief collaterals. Push the collateral information button once, the screen offers basic natural history and images for that species - or species complex - for your point on the globe. Push it twice, and you are in dialogue with central for more complex queries. Or, the gadget, through your cell-phone uplink, says "this DNA sequence not previously recorded for your zone, do you wish to provide collateral information in return for 100 identification credits?". Imagine what maps of biodiversity would look like if they could be generated from the sequence identification requests of from millions of users.

Such a gadget would allow access to true bioliteracy for all humanity. Such a gadget would be to biodiversity as the printing press was to literacy (and reading glasses, chairs, newspapers, the Library of Congress, and the computer). The blessing of information access through such a gadget is what the taxosphere - the collective intellectual might of

taxonomists, museums, collections and their centuries of literature - has within its power to offer society, global society, everyone. Will it? If it does not, wild biodiversity will continue its inexorable decline into the pit under the human heel, and the taxosphere will continue its accelerating slide into the realm of quaint esoteria shared among a very few enthusiasts who love their bugs, ferns and birds.

The gadget requires two things and a third. Thing one is the economic and social selective pressure to engineeringly collapse what today occupies two tabletops of machinery and a technician down to the size of a cell phone, reusable and cheap. This collapse is technically feasible in any one of many industrial centers of the world. \$1 million and five bright people. It has not happened in the past decade because no one saw any particular reason to do it. But there is a reason. Real bioliteracy requires on-site real-time hand-held cheap identification of hundreds of thousands of species, even though any one person at any one time may care about only one organism in one place.

Thing two is the global library of partial DNA sequences of a few cleverly selected target genes that among them carry species-specific combinations of nucleotides. Such a library is constructible in two phases running concurrently. The world's great biodiversity collections - museums, herbaria and microbe depositories - have on the shelf, somewhat in order, easily half (if not more) of the species of wild biodiversity encountered daily and consciously by 99% of the world's people. One phase is to quite straightforwardly organize and fund SWAT teams to simultaneously polish the taxonomic organization and extract the DNA samples for target gene species-level sequencing for these shelves of things. The other phase is to simultaneously reinforce the on-going biodiversity inventory of the world, and its taxonomic processing, so as to be able to sequence and characterize the as-yet uncollected wild biodiversity.

Both the in-house taxonomic processing and

outdoors inventory must be congruent with the agendas of the taxon- and site-focused primary users, so that as the sequence libraries emerge from the great collections, these same collections are also receiving and taxonomically processing the stream of new material (much of which may be sequenced as collected). The cross-phase potential for mutual and iterative reinforcement between the taxasphere and building the sequence library - and populating its collaterals - is enormous. I cannot overemphasize the necessity for collaterals. A phone number does you no good if there is nothing at the other end.

Thing three is the commercial-entrepreneurial process such that each of the millions of times that the gadget processes a sequence, a penny drops into a bucket that fuels the taxasphere to do what it does best and with such joy, and fuels the conservation community to actually conserve that which is being sequenced in its wild homes. Such a feedback system is imperative to saving the present and future biodiversity Library of Congress, so to speak. The goal is not to support yet another guild of biodiversity administration and consultancies, but rather to deliver bioliteracy to the world. And once people can read, ensure that there still be books to be read.

Thing three is obviously the most difficult, given that Homo sapiens is notorious for not reinvesting its gains, ill-gotten or otherwise, in the raw material source of those gains. Yes, startup capital will be required, but rather than get this from classical venture capitalism, this is a time for the world philanthropic capitalists to focus their energy. And will the gadget user pay a penny an identification once the system is in place? Yes, if "one minute one sequence one name" - and serious amounts of collaterals are available. Will the users feed new collaterals back into the information source to accompany their old or new sequence submitted? Yes, if they get identification credits, and as they see the value of retrieving their own submissions years down the road, to say nothing of the value of examining each others' submissions in real time and across geography.

The blending of these three things within their software glue and matrix is easily

attainable in less than a decade with the technological and sociological understanding in hand, for a total budget in the \$1-5 billion range. The process can be put in motion as a proof-of-concept for a tiny fraction of this.

The viewpoint in this commentary was inspired by reading Paul Hebert's enthusiasm and foresight in targeting just a part of the DNA sequence of a single gene as a species' "barcode" (Hebert et al 2003a,b), by recent planning efforts (Stoeckle 2003), and by witnessing the clarity with which a portion of the CO1 gene sequence can discriminate among a large number of species of butterflies and moths, bees, birds and mammals. Simultaneously it has its roots in decades of attempting to process millions of neotropical insects and plants through biodiversity inventory for a multiplicity of agendas.

And I am frustrated by working for a half century in the field, nurtured and guided at long distance by the worlds' best taxonomists, among hundreds of thousands of species of organisms, most of which are actually known to science yet can be identified in the field at best by only a very few persons. Neither I nor the other millions of wild biodiversity users can carry in their pocket the tens of thousands of pages of taxonomic descriptions, keys and images, and their authors. And even if all were to be collapsed down into a single chip, I still could not connect the beast in hand to its information as I stumble through the mud, rain and green of a 200,000-species-rich patch of Costa Rican rainforest. No one can learn the scientific language to read and hear the taxasphere's collective wisdom and facts for identification at the moment the bug is in hand or the leaf in mouth - even if we have the best access uplink to Google. And if each of us makes the long trek, which each won't, to the doors of any one of the great collections, in a matter of seconds the total taxasphere will be overwhelmed. And the gadget has huge potential for relieving the taxasphere of the drudgery of routine identifications at those places where even today's bioilliterate populace already knows that it needs to know what it is - the farmer's field, ports of entry, doctor's office, environmental monitoring, the kitchen, school science class, etc. Imagine what happens if the environmental monitor can know in a few minutes on-site the

hundreds of species of insects, mites, fungi and protista in an environmental sample.

The answer does not lie in better keys, more keys, more images on the web, more web sites, species pages, more descriptions, more phylogenies, more specimens, more anythings. Those are necessary collaterals, but not sufficient. The answer lies in a process that will for the first time connect the collective species-level biodiversity knowledge of the world to any and all users, on the spot, in real time, now. Fast, cheap and on-site single (or very few) gene sequencing has the potential to deliver the species-specific linkage between the species and its human-known collaterals. There is a huge opportunity for the taxasphere to thrust itself into a position of friendly social prominence - just as have education, agriculture, medicine and communication.

We must move wild biodiversity from the category of something to be removed to make room for the extended human genome to a book to be read, and read, and read. To an illiterate people, a library is just neatly stacked firewood. We must move the taxasphere from a "woe is us" mode to "here is what we can offer at society's negotiating table". It is within the technological power of the taxasphere to choose to move into a mutualism with directed molecular biology, mineaturizing engineering and entrepreneurialism. Praise and support taxonomists to be taxonomists and let us all be able to read wild biodiversity. The time is ripe for a barcoder. Godfray (2002) noted "in 10 to 20 years' time it will be simpler to take an individual organism and get enough sequence data to assign it to a 'sequence cluster' (equivalent to species) than to key it down using traditional methods". We do not have to wait 1-2 decades. Please do it now.

LITERATURE CITED.

GODFRAY, H.C.J. 2002. Challenges for taxonomy. *Nature* 417:17-19.

HEBERT, P. D. N., CYWINSKA, A., BALL, S. L. & DEWAARD, J. R. 2003a. Biological identifications through DNA barcodes. *Proc. R. Soc. Lond. B* 270: 313-322.

HEBERT, P. D. N., RATSINGHAM, S., & DEWAARD, J. R. 2003b. Barcoding animal life: cytochrome C oxidase subunit 1 divergences among closely related species. *Proc. R. Soc. Lond. B* 270: supplement O3BL0066.1-4.

STOECKLE, M. 2003. Taxonomy, DNA, and the bar code of life. *BioScience* 53:796-797.