Points of View

Syst. Biol. 55(5):844–847, 2006

ISSN: 1063-5157 print / 1076-836X online
DOI: 10.1080/10635150600960079

Who Will Actually Use DNA Barcoding and What Will It Cost?

STEPHEN CAMERON,[1] DANIEL RUBINOFF,[2] AND KIPLING WILL[3]

[1]*Department of Integrative Biology, Brigham Young University, Provo, Utah, 84602, USA; E-mail: stephen_cameron@byu.edu*
[2]*Department of Plant and Environmental Protection Sciences, 310 Gilmore Hall, 3050 Maile Way, University of Hawaii, Honolulu, Hawaii 96822, USA*
[3]*Department of Environmental Science, Policy, and Management, University of California, Berkeley, California, 94720, USA*

It is likely that the mere use of the word barcode is responsible for much of the appeal surrounding DNA barcoding, after all DNA-based identification methods (e.g., DeSalle and Birstein, 1996) used prior to Hebert et al.'s (2002) proposal of the term failed to ignite significant attention from the scientific community and none whatsoever from the general public. The term itself is loaded. Product barcodes are scanned using checkout lasers and indeed the image of the "Star Trek tricorder," a handheld scanner, has been used repeatedly by barcoding proponents in both presentations and papers (Janzen, 2004; Smith, 2005). Savolainen et al. (2005) use just such an allusion to commence their introductory paper of a special issue of *Philosophical Transactions of the Royal Society of London* devoted to DNA barcoding. Although the need for scanners capable of detecting biological weapons will undoubtedly lead to the development of portable DNA scanners at some point (a stated objective of the U.S. Department of Homeland Security, Directorate for Science and Technology), there is, however, absolutely no indication that they will be affordable or practical for the kinds of mass identification uses that barcoding proponents are selling to the general public as the outcome of this research. Further, a critical difference between a "tricorder" and a portable DNA barcoder would be the DNA component. Whereas Mr. Spock need only wave his tricorder in the general direction of an alien to be told what it is, real-world barcoders will need to actually handle that animal, remove tissue from it, and load it into the barcoder to get an identification. DNA barcoding is intrinsically linked to specimens as samples that must be collected for DNA extraction, be it in a molecular biology laboratory as at present or by a handheld barcoder at some point in the future. This small detail neatly circumscribes what barcoding can, and what it can't, achieve.

Who then would use DNA barcoding? Dan Janzen has written eloquently and spoken passionately about the need to improve biological literacy amongst the general public and that without the ability to "read" nature by identifying its contents, biodiversity is doomed to be underappreciated and so destroyed (see Janzen et al., 2005, for full development of this argument). One wonders what the place for even a portable DNA barcoder is in this vision. The majority of the public observes nature; they don't sample it by removing the legs from butterflies and throwing them into a barcoder to get identifications. Wildlife protection authorities expend a lot of effort trying to keep people away from wildlife and certainly not handling or vivisecting them. Will the advent of a portable barcoder result in a complete turn around by wildlife authorities encouraging the wholesale handling of wildlife by the public? For smaller or more delicate plants and animals (which would include those most difficult for an interested amateur to identify without a barcode), being "identified" with a barcoder is likely to be fatal. Not simply due to the handling or picking necessarily, but due to the dissection necessary to get a tissue sample. Further, such barcode identifications will not share the benefits of traditionally collected specimens because an inexperienced member of the public is likely to be simply, and passingly curious, and not inclined to retain each "specimen" as a voucher with rigorous locality data for future study or reference. Although proponents suggest to compensate new species vouchers with free identifications (Janzen, 2004), one needs only think of the vast majority of people who toss aside cans they could recycle for a refund. The inconvenience of handling and sending a specimen in decent condition is far greater than recycling. Clearly such barcodes are unlikely to inspire or benefit the vast majority that needs to be reached to protect biodiversity. Amongst the wider public, the largest group identifying species on a regular basis is birdwatchers. DNA barcoding will be of limited use to them unless it is proposed that a 12-gauge shotgun now become standard equipment for such "twitchers." Even for professional bird identifications, such as those prepared for biological impact studies, trapping is secondary to visual identifications due to the stress that handling inflicts on birds. Mammal identification for the general public is based on direct visual identifications as well as the interpretation of trail signs, scats, or hair. Perhaps some identifications could come through

the analysis of feathers, hair, or scat left behind, but this is unlikely to have the same appeal to the general public that the identification of an animal in hand will present. Are we likely to hear of something similar to the following in the years to come?

> "Mum is this a grizzly bear or a black bear?"
> "Well Johnnie why don't you go poke your barcoder into it and find out."

The truth is that DNA barcoding will not have any meaningful use for the general public and even when a portable barcoder eventually materializes it will not lead to any increase in the biological literacy of the man on the street. If systematists are truly concerned about increasing biological literacy, they would likely be better served developing well-illustrated virtual field guides (e.g., Janzen and Hallwachs, 2005, online guide to neotropical Lepidoptera of Costa Rica, http://janzen.sas.upenn.edu) or computer based keys like LucID (Cranston, 2005; http://www.lucidcentral.com) that could easily be adapted for devices such as Palm Pilot and Blackberry handheld computers.

So what is barcoding good for? It can help with what are already scientific processes—quarantine interception, forensic identifications, meat or tissue testing, and species discovery. In truth, this last point is probably the most avidly pursued goal of barcoding reflected in their publications, presentations, and even their chosen title—Barcode of Life Initiative. Hajibabaei et al. (2005) clearly lays out the scale of this goal—generating 10 barcode sequences for each of the estimated 10 million species on Earth and that this could be accomplished within a decade if 50 laboratories sequenced 20,000 species a year. Whereas Hajibabaei et al. (2005) do not estimate a cost for this effort, elsewhere Janzen et al. (2005) have listed US$2.50 per sample exclusive of labor as the subsidized cost for barcoding moths. We estimate that the actual cost as more like US$5.00 per barcode to include technician's labor and servicing contracts on automated sequencers (see Appendix 1 for cost calculation). The entire effort would therefore cost half a billion dollars: much cheaper than other "big science" initiatives such as exploring Mars or determining the smallest subatomic particles.

However, it must be noted that this is the cost for barcoding based on a single gene and for only 10 samples per species. Although this is a higher standard than current barcoding papers, which rarely manage more than 5 samples per species (e.g., Ward et al., 2005, Australian fishes, 1 to 15 samples, average of 3.66), 10 samples per species is still an underestimate of the amount of data that would need to be collected. Even proponents of barcoding are advocating larger sample sizes. Matz and Nielsen (2005) note that using 12 samples gives only a 95% confidence level to identifications in the resulting database. Janzen et al. (2005) point out that when barcoding a morphologically unknown biota, one is essentially sequencing in the dark with no idea of possible underlying divisions, so much larger samples are needed. Elsewhere in the same paper, Janzen advocates sequencing at least 10 individuals per location with additional samples in cases of suspected cryptic diversity. For example, after cryptic diversity was identified in the skipper butterfly *Astraptes fulgerator* (Hebert et al., 2004), over 460 individuals were sequenced to identify the 10 putative species occurring in a single region of Costa Rica. It is therefore likely that barcoding life could be never ending due to the constant need to sequence new individuals until the geographic boundaries of a species are established and representative diversity within those boundaries sampled. Although extremely hard to quantify, for argument's sake, assume that the actual number of samples needed would be more like 100 specimens per species rather than the 10 estimated by Hajibabaei et al. (2005).

The second issue is the growing realization that CO1 (mitochondrial cytochrome oxidase *c* subunit 1) is not the magic bullet that Hebert et al. (2002) initially conceived. This is evidenced by some barcoding advocates pushing for the inclusion of additional genes in barcoding efforts, either due to failure of CO1 to resolve taxa of interest (e.g., amphibians Vences et al., 2005; plants Kress et al., 2005) or due to the belief that extra data will improve the accuracy of barcoding (Matz and Nielsen, 2005). Although this matter has yet to be addressed by single gene advocates, for discussion's sake let us assume that 3 barcode genes or approximately 1800 bp is sufficient, as this matches the amount of data suggested by barcoders who advocate multiple gene analyses. Multiplying through—10 million species by 100 specimens by 3 barcodes at US$5.00 per barcode, a more realistic cost for barcoding life is more in the order of US$15 billion. This is all well and good for a war but let's not be under any delusions about this actually getting funded. Although the costs above match those proposed by supporters of DNA barcoding, the amount of data that needs to be collected could be much higher and much more expensive.

Even if all the shortcuts that barcoding advocates propose are taken and the entire barcoding effort only cost half a billion, what are the implications for the real-world funding climate in taxonomic biology? The United States currently has one of the best funding regimes for taxonomic science with the National Science Foundation's twice yearly Systematics panel, biennial Partnerships for Extending Expertise in Taxonomy (PEET), and the Planetary Biotic Inventories (PBI) program, which collectively award approximately US$8 million per year. Due to the absence of exclusive grant programs for taxonomy in most countries, it is difficult to extend these numbers to a global taxonomic expenditure, but anecdotal evidence suggests much lower funding rates (e.g., House of Lords Report on Taxonomy in the United Kingdom transcript available online: http://www.systass.org/systass-lords-transcript.html). Again, for the sake of round numbers let us assume that global expenditure is approximately US$10 million annually. Even if every dollar currently spent on taxonomy or systematics were to be diverted to barcoding, it would still consume 50 years worth of funding at

the extremely low price tag suggested by Hajibabaei et al. (2005) or 1500 years at our more realistic rate. Neither estimate is even close to the 10-year time frame suggested by Hajibabaei et al. (2005).

The current barcoding plan also assumes that all of life's variation is readily available for barcoders to work on. It is probable that less than half of life has ever been collected. Although museum and herbarium collections for a few locations such as western Europe and North America have collections accounting for almost all free-living multicelled species occurring within those areas, they cover a small and, in biodiversity terms, almost insignificant portion of the planet. The majority of species occur in the tropics, in less developed countries, whose biotas have been poorly studied. Even for moderately well-studied biotas such as that of Australia, curators estimate that their collections account for less than half of the true fauna (C. Burwell, Queensland Museum, personal communication). As recognized by Lord May (2004), massive collecting, sorting, and curation efforts will be a necessary component of any global barcoding effort. Even if all species were present in museum collections, the barcoding success rate with museum material is poor. Full-length barcodes were amplified from less than 30% of moth specimens only 2 years old (Hajibabaei et al., 2005). Use of more sensitive extraction protocols, repair enzymes, and amplification of multiple small overlapping sections of the barcode had higher success rates (circa 70%); however, the supply, labor, and time cost and complexity of such approaches are considerably higher than those quoted above. It would probably be easier to collect such samples anew. However, the cost of collecting new material has never been included in estimates of what barcoding all of life would cost and could very easily surpass the sequencing costs by several orders of magnitude.

Are the other uses of DNA barcoding any more practical as alternatives to currently used methods? Although DNA approaches to identifying quarantine or pest species intercepts shows promise (Armstrong and Ball, 2005) and the smaller number of species that need to be identified makes the assembly of a high-quality reference database of barcodes much easier than for all of life, would it be practical to switch to an entirely DNA-based quarantine system? Personnel costs would probably be similar with DNA technicians replacing current curatorial staff. Equipment costs would be high, approximately US$400,000 per laboratory to cover the cost of purchasing an automated sequencer, PCR machines, and additional specialist molecular biological equipment. Savings on equipment could be made by consolidating quarantine barcoding in a central facility, but this would come at a cost to efficiency, slowing down the rate at which identifications can be performed due to the time taken to transport samples from field inspection sites to the laboratory. Centralization of facilities also raises the cost per barcode determined due to transportation costs, which can be very high for overnight or other forms of rapid delivery. The most time-efficient approach would be to simply replace existing morphological identification units with barcoding ones. Identifications would then still take at least a day to perform, PCRs being run during the day and sequencing overnight. This compares poorly to the rapid identifications that can be achieved by specialists based on morphological characters. Given the pressure to release international trade in a timely fashion, speedy identifications are at a premium and even next day time frames are more costly than they need be. There is also the question of the ongoing costs associated with a DNA-based quarantine identification system. At US$5.00 a barcode, a busy international port can spend a lot of money on identifications. The Australian Quarantine and Inspection Service (AQIS) identifies 20,000 intercepted specimens per year (J. Nielsen, personal communication). A change to a barcoding system would thus cost an additional 100,000 dollars each year. Although there are certainly species complexes that are difficult to identify based on morphological criteria alone and for which DNA barcoding is a logical application, the empirical evidence is that barcoding in its current form (single gene, distance-based discrimination) doesn't resolve species complexes very well (Armstrong and Ball, 2005; Meyer and Paulay, 2005; Hajibabaei et al., 2006). There are also circumstances in which species identifications are not enough; identifying resistant strains within a species may be more important than the species identification alone, especially if nonresistant strains occur in a region but resistant strains of a pest species do not and keeping the resistant strains out is the most important activity. In these circumstances, it is therefore probable that the best approach is to develop specialized identification tools to address the particular questions posed by each important pest species or species complex.

Thus it seems that some of the proposed applications of DNA barcoding suffer from serious impracticalities. DNA barcoding will have almost no direct use to the general public—even if a portable barcoder becomes available—due to the absolute necessity of handling specimens which need to be identified. Other institutional uses of barcoding will carry a hefty price tag that has yet to be considered. Both the cost of setting up a barcode database to cover all life and the on going expense of using barcodes as the frontline identification technique are vastly greater than the proponents of barcoding have estimated. Even if one were to ignore all of the theoretical questions that hang over using DNA barcoding as a primary approach to discovering biodiversity (e.g., Lipscomb et al., 2003; Will and Rubinoff, 2004; Moritz and Cicero, 2004; Will et al., 2005), it is still too expensive to form a practical alternative to current approaches to species discovery or identification. The financial implications of full implementation of the barcoding manifesto should be given greater consideration in future discussions of DNA barcoding's impact on taxonomic practice.

ACKNOWLEDGEMENTS

The authors would like to thank C. Burwell and J. Nielsen for their advice. SLC was supported by NSF grants DEB0120718 and DEB0444972.

References

Armstrong, K. F., and S. L. Ball. 2005. DNA barcodes for biosecurity: Invasive species identification. *Phil. Trans. R. Soc. Lond. B* 360:1813–1823.

Cranston, P. S. 2005. Ancient and modern—-Toolboxes for e-bugs. *Syst. Entomol.* 30:183–185.

DeSalle, R., and V. J. Birstein. 1996. PCR identification of black caviar. *Nature* 381:197–198.

Hajibabaei, M., J. R. deWaard, N. V. Ivanova, S. Ratsasingham, R. T. Dooh, S. L. Kirk, P. M. Mackie, and P.D.N. Hebert. 2005. Critical factors for assembling a high volume of DNA barcodes. *Phil. Trans. R. Soc. Lond. B* 360:1959–1967.

Hajibabaei, M., D. H. Janzen, J. M. Burns, W. Hallwachs, and P. D. N. Hebert. 2006. DNA barcodes distinguish species of tropical Lepidoptera. *Proc. Nat. Acad. Sci. USA* 103:968–971.

Hebert, P. D. N., A. Cywinska, S. L. Ball, and J. R. deWaard. 2002. Biological identifications through DNA barcodes. *Proc. R. Soc. Lond. B* 270:313–321.

Hebert, P. D. N., E. H. Penton, J. M. Burns, D. H. Janzen, and W. Hallwachs. 2004. Ten species in one: DNA barcoding reveals cryptic species in the neotropical skipper butterfly *Astraptes fulgerator*. *Proc Nat. Acad. Sci. USA* 101:14812–14817.

Janzen, D. H. 2004. Now is the time. *Phil. Trans. R. Soc. Lond. B* 359:731–732.

Janzen, D. H., M. Hajibabaei, J. M. Burns, W. Hallwachs, E. Remigio, and P. D. N. Hebert. 2005. Wedding biodiversity inventory of a large and complex Lepidoptera fauna with DNA barcoding. *Phil. Trans. R. Soc. Lond. B* 360:1835–1845.

Janzen, D. H., and W. Hallwachs. 2005. Dynamic database for an inventory of the macrocaterpillar fauna, and its food plants and parasitoids, of Area de Conservacion Guanacaste (ACG), northwestern Costa Rica. http://janzen.sas.upenn.edu.

Kress, J. W., K. J. Wurdack, E. A. C. Zimmer, L. A. Weigt, and D. H. Janzen. 2005. Use of DNA barcodes to identify flowering plants. *Proc. Nat. Acad. Sci. USA* 102:8369–8374.

Lipscomb, D., N. Platnick, and Q. Wheeler. 2003. The intellectual content of taxonomy: A comment on DNA taxonomy. *Trends Ecol. and Evol.* 18:64–66.

Matz, M. V., and R. Nielsen. 2005. A likelihood ratio test for species membership based on DNA sequence data. *Phil. Trans. R. Soc. Lond. B* 360:1969–1974.

May, R. M. 2004. Tomorrow's taxonomy: Collecting new species in the field will remain the rate-limiting step. *Phil. Trans. R. Soc. Lond. B* 359:733–734.

Meyer, C. P., and G. Paulay. 2005. DNA Barcoding: Error rates based on comprehensive sampling. *PLoS Biol.* 3:e422.

Moritz, C., and C. Cicero. 2004. DNA barcoding: Promise and pitfalls. *PLoS Biol.* 2:1529–1531.

Savolainen, V., R. S. Cowan, A. P. Vogler, G. K. Roderick, and R. Lane. 2005. Towards writing the encyclopaedia of life: An introduction to DNA barcoding. *Phil. Trans. R. Soc. Lond. B* 360:1805–1811.

Smith, V. S. 2005. DNA barcoding: Perspectives from a "Partnerships for Enhancing Expertise in Taxonomy" (PEET) debate. *Syst. Biol.* 54:841–844.

Vences, M., M. Thomas, R. M. Bonett, and D. R. Vieites. 2005. Deciphering amphibian diversity through DNA barcoding: Chances and challenges. *Phil. Trans. R. Soc. Lond. B* 360:1859–1868.

Ward, R. D., T. S. Zemlak, B. H. Innes, P. R. Last, and P. D. N. Hebert. 2005. DNA barcoding Australia's fish species. *Phil. Trans. R. Soc. Lond. B* 360:1847–1857.

Will, K. W., B. D. Mishler, and Q. D. Wheeler. 2005. The perils of barcoding and the need for integrative taxonomy. *Syst. Biol.* 54:844–851.

Will, K. W., and D. Rubinoff. 2004. Myth of the molecule: DNA barcodes for species cannot replace morphology for identification and classification. *Cladistics* 20:47–55.

First submitted 01 February 2006; reviews returned 14 March 2006; final acceptance 16 April 2006

Associate Editor: Marshal Hedin

Appendix 1

Barcode Cost Calculations

Our calculations for the cost per barcode are based on current cost estimates from suppliers for the consumables minimally necessary for generating a barcode. We do not factor in the costs of infrastructure as it is impossible to estimate how many barcodes would be produced before a particular piece of equipment needs to be replaced. Infrastructure costs are better considered as one-off, up-front costs. Similarly, we assume that any serious DNA barcoding facility would run all of its own equipment, so we do not include any wages either. Producing a DNA barcode requires a minimum of four steps: DNA extraction, PCR, PCR purification, and sequencing. Hajibabaei et al. (2005) tried multiple DNA extraction methods and found that the NucleoSpin96 kit had the highest success rate for both fresh and archival specimens. This kit costs US$1.90 per extraction. PCR costs could vary considerably depending on the quality of *Taq* polymerases used in the reaction. Because of the range of template qualities likely to be encountered in a barcoding program, it would be advisable to use recombinant or second generation *Taq* such as the Applied Biosystems AmpliTaq, which costs US$0.37 per unit, 1 unit per PCR; however, hot-start PCR enzymes such as AmpliTaq, Gold have higher success rates with less handling at a higher per unit cost of US$0.43. Most molecular biologists would then verify the results of the PCR by agarose electrophoresis; however, this is not essential, success can be determined directly by sequencing. PCR purification can take many forms; however, vacuum-driven column purification methods such as the Millipore Montage PCR system have excellent capacity for handling high numbers of reactions rapidly. Purification using the MontagePCR96 system costs US$0.28 per reaction. Sequencing costs vary widely between facilities at different universities, from as high as US$10 per sample for facilities that perform the entire sequencing reaction to much lower costs for groups that do more of the work themselves. Currently the senior author's home institution charges US$0.50 per lane, which covers service contracts on the automated sequencers but not technician wages to run them. The senior author usually uses 1 μl of ABI Big Dye ver3 dye terminator sequencing chemistry—US$0.68—for each sequencing reaction. For double-stranded sequencing, a minimum of two sequencing reactions are needed for a total sequencing cost of US$2.36. Use of robotic sequencing approaches to reduce the amount of sequencing chemistry used per reaction could reduce this cost but at a very large additional cost to infrastructure.

The costs are therefore: extraction, US$1.90; PCR, US$0.37; PCR purification, US$0.28; and sequencing, US$2.36, for a total of US$4.91. Additional minor laboratory supplies such as buffers, gels, etc., would take the sum to very close to the US$5.00 quoted in the text.