Taxonomy, DNA, and the Barcode of Life[1]

Meeting held at Banbury Center, Cold Spring Harbor Laboratory, New York, NY
September 10-12, 2003
Sponsored by the Sloan Foundation

Draft Conference Report

December 1, 2003

Imagine a world in which any person, anywhere, at any time can identify any species at little or no cost. That world is technologically upon us. This report addresses the formative stages of an initiative to bring this to society sooner rather than later.

Mark Stoeckle[2]
Rockefeller University

Daniel Janzen
University of Pennsylvania

Winnie Hallwachs
University of Pennsylvania

James Hanken[3]
Harvard University

James Baker[3]
Academy of Natural Sciences, Philadelphia

1. A list of participants, the meeting agenda, and other information related to the Barcode of Life effort is available at http://phe.rockefeller.edu/BarcodeConference/index.html.
2. Send comments to MarkStoeckle@nyc.rr.com
3. Conference Co-Chair

Summary. The scientific rationale, societal benefits, and possible organizational strategy for determining DNA barcodes across all eucaryote domains of life were the subject of a "Taxonomy, DNA, and the Barcode of Life" conference held at Cold Spring Harbor Laboratory, September 10-12, 2003. The goal of this proposed effort is to promote and initiate a process for a rapid and inexpensive method for on-site identification by any user, irrespective of their background training, for the estimated 10 million species of eucaryote life on Earth. The benefits of using a small sequence of a single gene for identification of many large portions of the diversity of life include 1) facilitating species identification for all users, including flagging specimens that may represent species new to an area or new to science, 2) enabling rapid, inexpensive, on-site identifications where traditional methods are unfruitful or too complex, 3) driving technology development for DNA barcode analysis that can be applied in the field, and 4) providing insight into the evolutionary history of life via quick and reliable species-level identifications. In specific cases, this sequence data will also contribute to multi-gene efforts to determine evolutionary relationships.

A short to very short (100-650 bp) sequence of mitochondrial cytochrome c oxidase subunit I (COI) is proving to be highly efficient for barcoding many animal species and possibly a variety of other domains. Identifying effective gene sequences for barcoding plants and other domains is an important priority, as is perfecting forensic methods for quickly extracting adequate sequences from museum specimens stored for decades or centuries under a variety of circumstances. Possible pilot projects are presented—these are focused on select taxonomic groups to further help validate the barcoding approach as an appropriate means of species recognition and discovery. The initial steps toward an International Barcode of Life Initiative, which will draw on the taxonomic expertise and extensive collections of the world's natural history museums, are discussed and it is anticipated that a proposal for a coordinated Secretariat will be forthcoming.

A Barcode of Life effort will create a valuable public resource in the form of an electronic database that contains DNA barcodes linked to voucher specimens, images, and collateral information including taxonomy, geographic distributions, natural history, and relevant medical and economic concerns. In addition to its use in taxonomy and scientific discovery, DNA barcoding has broad potential practical applications in biomedicine (identification of pathogens and vectors), agriculture (recognition of pest species), and national security (remote monitoring for biowarfare agents). By applying technologies of molecular biology to the living world outside as well as in the laboratory, a Barcode of Life Project offers the prospect of deeper understanding, appreciation, and non-destructive use of the diversity of life on Earth, thereby both improving human well-being and raising the chance that a serious fraction of wild biodiversity may be allowed to survive into perpetuity.

I. Background

Current methods of identifying, naming, and classifying organisms are built on the taxonomic system that was developed by Carl Linnaeus 250 years ago and are largely, but not entirely, based on visible morphology. Linnaean typology has been modified by the subsequent recognition of genetic variation among individuals of a species and the insight that, when viewed from an evolutionary time scale, species are not static. Morphological traits are the cornerstone of this taxonomic process, which has described an estimated 1.7 million species, perhaps 20% of eucaryote life on Earth. There are, however, limitations to relying solely or largely on morphology in identifying and classifying life's diversity, and modern taxonomic work includes analysis of a host of other traits, including genes, isoenzymes, physiology, behavior, population biology, and geography. However, the morphological nuances and their less visible partner traits that distinguish species are so complex and subtle that most taxonomists specialize in a single group of closely related organisms. As a result, a multitude of taxonomic experts may be needed to identify specimens from a single biodiversity survey. Finding appropriate experts and distributing specimens can be a time-consuming and expensive process. The non-specialist who needs to identify specimens now, but is far from the few taxonomic centers in the world, is confronted with a nearly impossible task.

Web-based databases with high-resolution images may help to some extent, but even with modern technological access to the best knowledge, enormous numbers of species cannot be reliably identified except by 1-2 persons in the world. Juvenile forms, which are often more abundant than adults, may have no known species-level distinguishing characteristics and need to be reared to maturity (if that is possible) to be identified. In some species, only one sex can be identified or the sexes cannot be associated. For plants, a specimen may be readily identified from flowers or fruits, while roots, stems, and leaves are indistinguishable without a huge and time-consuming investment of resources. It is difficult to know how many cryptic (i.e., morphologically highly similar) species there are. Assuming that there are another 8 million or so undescribed species of life on Earth, simply determining if a specimen matches an already described species will be increasingly impractical as the encyclopedia of morphologic descriptions expands.

DNA-based species identification is a potentially powerful approach to both fill the gap and build on the taxonomic base already in place (easily 50% of the species ordinarily encountered by most people on earth are already described and can be identified—though often at high cost—with light to intensive application of classical methods). The utility of DNA analysis in taxonomic studies and the identification process was the subject of the initiating conference entitled "Taxonomy and DNA", which was held in March 2003 (Stoeckle 2003). As discussed at that meeting, a remarkably short DNA sequence in a single gene can contain more than enough information to identify 10 million or more species. For example, a 600-nucleotide segment of a protein-coding gene contains 200 nucleotides that are in the third position within a codon. At these sites, substitutions are (usually) selectively neutral and mutations accumulate randomly. Even if a group of organisms was completely biased to either adenosine or thymine (or alternatively, to either guanosine or cytosine) at the third nucleotide positions there would still be 2^{200}, or 10^{60}, possible sequences based on third-position nucleotides alone.

DNA sequence analysis of a single gene sequence to enable species identification is termed DNA barcoding. This use of the word barcoding is meant as an explanatory analogy with the Uniform

Product Code barcodes on manufactured goods, which may be applied to represent a "species" of product (as well as an individual item). The availability of broad-range primers for amplification of a 645 bp fragment of cytochrome c oxidase subunit I (COI) from diverse invertebrate and vertebrate phyla establishes this gene sequence as a particularly promising target for species identification in animals (Folmer et al. 1994), as does the moderate change rate in this sequence and the large numbers of copies per cell owing to COI being a mitochondrial gene. Proof of principle for DNA barcoding is now being provided and probed by comparison of COI sequences among closely related species and across diverse phyla in the animal kingdom (Hebert et al. 2003a, 2003b; Hecht 2003).

A critical test of DNA barcoding is whether it enables discrimination between closely related species. Comparison of COI sequences from 13,000 pairs of congeneric species showed a mean divergence of 11.3%, corresponding to approximately 50 diagnostic substitutions per 500 bp of the COI gene (Hebert et al. 2003b)**.** Furthermore, COI sequence variation within species is ordinarily quite low, less than 2%, and has not been an impediment to species discrimination, including among many assemblages of closely related organisms (Hebert et al. 2003a). Importantly, because COI divergences between even closely related species are ordinarily much higher than intra-specific variation (average 5 to 20-fold higher), COI may be used for provisional detection of undescribed species in taxonomically difficult groups.

COI barcoding will not enable identification of, or discrimination among, all animal species. For example, some (but not all) Cnidaria (sea anemones, corals, and some jellyfish) have very low mitochondrial sequence diversity (Hebert et al 2003a). However, the utility of locating some other single gene sequence that functions for recalcitrant animals, as well as one that works for plants, is undeniable. Species that hybridize regularly and newly emerged species may not exhibit enough sequence diversity to be separated by COI (or by any other single gene), but nonetheless can be distinguished from their less problematic congeners and otherwise confusable species and life forms.

Taken together, the results demonstrate strong congruence between morphology-based taxonomy and COI barcode analysis, and provide confidence that a barcoding initiative will be able to deliver on-site, real-time, species-level identification to the greater public, once the analytic process has been miniaturized to a hand-held device and the sequence libraries are in place. A toll booth system that funds the taxonomic information delivery process is envisioned (Janzen 2004).

II. What is a DNA barcode? Most simply, it is

A short DNA sequence that identifies a species

More completely, it is a short DNA sequence(s) that enables species identification in a particular domain of life. This definition is meant to include the fact that there is intra-specific variation, thus a species may typically be characterized by a collection of sequences, albeit very similar sequences (intra-specific divergences are ordinarily less than 2%). Also, the appropriate target gene(s) may be different in different domains of life.

Supplementary comments:

1. DNA barcode data will be derived by the professional community, which will establish protocols and standards for quality control.

2. In constructing sequence libraries, each DNA barcode sequence will be linked to a voucher specimen accessioned in an institutional repository (e.g. museum or hebarium collection). DNA barcodes and related taxonomic and collection data will be made widely and publicly available through electronic databases and will be linked to other complementary efforts (e.g. MorphoBank, GenBank, GBIF).

3. As currently envisioned, barcoding is being tested and developed for eucaryote organisms. While in theory barcoding can obviously be applied to procaryotes (including viruses), the sequences and technology currently available demand that this be a parallel or even second phase effort. Equally, the current initiative is focused on identifying samples taken from a single organism, rather than community-level analysis of pooled samples, though the latter is in the foreseeable future.

What is it not?

There are many possible (mis)interpretations of what a DNA barcode project is and is not intended to do.

1. DNA barcoding is intended to complement, facilitate, and enhance—not supplant or invalidate—existing taxonomic practice. It does not duplicate or compete with existing efforts to resolve deep phylogeny, i.e. the history of life on planet Earth.

2. It is expected that DNA barcodes will contribute to the discovery and formal recognition of new species. However, DNA barcodes should not be used as the sole criterion for description of new species, which instead require analysis of diverse data, including morphology, ecology, and behavior, as well as genetics.

III. Societal Benefits

Allan Bromley, in the June 2003 issue of *Physics Today* asks the question: what criteria should be used to establish funding priorities for science programs (Bromley 2003)? His criteria can be usefully applied to the Barcode of Life effort as we look to develop the best case for funding. These are all criteria that have been and are being used by funding agencies to make decisions about large new programs. There are seven questions to answer that relate to the intrinsic value of the research, the extrinsic impacts, and structural needs:

1. To what extent does the research proposed have the potential of providing fundamental new understanding of our universe?

The full diversity of life on Earth is largely unknown. We simply do not know, even within an order of magnitude, how many species exist. More importantly, because we cannot easily

recognize species, our rate of discovery and our ability to organize and draw on what we know is very weak. A Barcode of Life Project offers an efficient (rapid and inexpensive) method to map and read our biological universe. In addition, analysis of a small set of genes across species will facilitate the quest for understanding of the evolutionary history of life.

2. To what extent does the research have the potential of affecting other areas of scientific research?

The results of this research will be of central importance to conservation biology, ecology, biomedicine, agriculture, biodiversity prospecting, and ecosystem management—it will earn prominent pages in every basic biology textbook. It will be enormously useful for biodiversity surveys, in which large numbers of organisms from diverse taxa need to be identified rapidly and cheaply. It will enable identification of eggs and immature forms to determine life cycles, plant roots in soil samples for plant physiology and soil science research, and analysis of stomach contents to determine food webs.

3. To what extent does the research have the potential of leading to new generic technologies?

This project will drive development of robust techniques for inexpensive, fast, and field-friendly DNA analysis that will have widespread scientific and industrial uses.

4. To what extent does the research contribute to national security, economic competitiveness, or improvement in our quality of life?

DNA barcoding will have broad practical applications, including in biomedicine (identification of pathogens and vectors), agriculture (recognition of pest species), environment (identification of species that are endangered or whose trade or exploitation is restricted), and national security (remote monitoring for biowarfare agents). In addition, by giving the entire populace the ability to actually "read" wild and domestic biodiversity at the species level, the results of this project will enhance efforts to conserve and restore nature, a highly popular goal that will improve our quality of life.

5. To what extent does the research hold promise of significant return on earlier scientific facility investment (e.g., major equipment, new laboratories, etc.)?

This effort builds on five decades of molecular biology research including DNA sequencing technologies derived from the Human Genome Project and extends the value of GenBank, the likely repository of the gene sequences, from health to environment. It makes use of the huge effort undertaken over the past 200 years to build collections in natural history museums and comparable institutions. .

6. To what extent is the research at or near the international frontiers of work in the field?

The key evidence of the efficacy of short sequences for species identification was published only in January 2003 (Hebert 2003a). There are now small-scale efforts for DNA barcoding of restricted sets of organisms occurring all over the world, with many other countries and researchers rapidly perceiving the opportunity being presented. TThe initiative originated in Canada, but the fact that the USA holds the largest specimen collections and its mass of

technologic and taxonomic ability argue for a lead role for the USA in carrying forward the initiative and extending the frontier.

7. Are enough scientists with the relevant expertise available to carry out the program?

A large group of willing and enthusiastic researchers who have the technical expertise is available. Funds and organization are needed.

IV. Scientific Benefits

The scientific benefits of sequencing a particular gene across large blocks of the diversity of life lie in at least four easily visualized areas: 1) greatly facilitating on-site, real-time species identification by anyone, including flagging specimens that may represent new species, 2) enabling identifications where traditional methods are unrevealing, 3) driving technology development for DNA barcode analysis that can be applied cheaply and quickly in the field, and 4) facilitating insight into the evolutionary history of life on Earth through allowing easy identifications of any set of specimens.

1. Facilitating species identification. As a uniform, practical method for species identification, DNA barcoding will be of great utility in biodiversity surveys, where large numbers of specimens from diverse taxa need to be identified quickly and cheaply by people with little or no taxonomic training or inclination. Additionally, once a comprehensive set of DNA barcodes has been established, any set of specimens could be rapidly "scanned", and those with novel barcodes, which might represent new species, selected for further analysis. By streamlining the process of specimen identification, barcoding will allow taxonomic resources to be focused on scientific discovery, rather than on the drudgery of routine identifications.

Regarding the role of DNA barcoding in diagnosing new species, it will be a useful addition to the existing tools, but it is not visualized as replacing them. In many groups, alpha taxonomy requires data from morphology, behavior, ecology, natural history, and geographic variation. These data will certainly be enhanced by complementary information on associated DNA sequences.

The aim of DNA barcoding is to analyze the smallest sequence that will provide the required information, thus reducing cost and facilitating automation. It appears that a small portion of a COI may enable identification of most animal species (and a sequence of some other gene is very likely to work as well for plants, fungi, etc). For some specimens, analysis of the primary target may provide a less precise identification, perhaps just to a portion of a genus (which will nonetheless be useful). In these situations, additional gene target(s) or other traits of the organism will be needed for species determination. However, routinely sequencing additional genes will add to the complexity and cost of the procedure, and is not anticipated in setting up barcoding sequence libraries. Analysis of multiple sequences may be particularly difficult in samples in which DNA is not well preserved, especially with museum specimens. On the other hand, any and all DNA captured for DNA sequence libraries should be saved for later analysis of other genes where the objective merits it.

2. Identifications where morphology is inconclusive. Together with flagging specimens that may represent new species, the most important uses of DNA barcoding will be in enabling identification where traditional methods are unrevealing. Potential applications include identifications of immature forms (to explore life cycles as well as to know what juvenile is in hand), analysis of stomach contents to determine food webs, diagnosis of cryptic species (these may be more common than is generally realized), and identification of roots in soil samples. DNA barcoding could be used to recognize products prepared from protected species, to identify animal reservoirs for human disease (e.g. determination of the avian sources of West Nile virus infection by analyzing mosquito blood meals), and to identify pest species in imported goods and at agricultural sites (perhaps using automated detection systems).

3. New technologies. A large-scale DNA barcoding effort will drive technology development for DNA analysis, including robust methods for DNA isolation from various specimens and rapid, inexpensive sequencing techniques. "Faster, better, cheaper" methods are critically important to the widespread application of barcoding throughout society as well as scientists more "narrow" applications, as it will be repeatedly employed whenever specimens are collected (and in this way it is very different from a genomic sequencing project). Technologies that can be applied in the field are highly desirable. It is envisioned that advances in instrumentation and electronics can be harnessed to produce an IPod-sized, low-cost unit that will process field specimens (a leaf, an animal hair, an insect leg), do sequence analysis and comparison to stored sequences, and be wirelessly linked to a public database for uploading sequence and collection data and dialoguing about collaterals (Janzen 2004). Regardless of the pace of miniaturization, the Barcode of Life effort will create a lasting public resource in the form of a web-accessible database that contains DNA barcodes linked to specimen and taxonomic information.

4. Evolutionary insights. Sequencing a uniform gene from each "leaf" on the tree of life is likely to provide important insights into evolution by greatly facilitating the recognition of the existence of those "leaves" and their traits. A barcode gene, i.e., one that is divergent enough to enable species identification, may not be effective at reconstructing the deep branches, but it will help in understanding the process of speciation—that after all, did occur at each of those deep branches--by comparisons within groups of closely related organisms.

V. Implementation: Blueprint for International Barcode of Life Project

Mission Statement. The mission of the International Barcode of Life Project is to establish a practical and efficient DNA-based method for identification of the estimated 10 million eucaryote species of life (with equal capacity for perhaps an equal number of prokaryote species being conceivable by an integrated parallel mission). By applying technologies of molecular biology to the living world outside the laboratory, a Barcode of Life Project offers the prospect of world-wide and much deeper understanding and appreciation of the diversity of life for everyone.

Long-range goals. The long-range goals of the Barcode of Life project are:

1) To determine and compile the DNA barcodes of all eucaryotic forms of life on Earth

2) To establish an electronic database that links DNA barcodes of vouchered specimens with corresponding taxonomic and collection data

3) To facilitate, through user toll booths or other resource gathering processes, support of the taxonomic community to provide their part in linking species-level identification to what is known about that species

4) To develop efficient methods for DNA barcode determination, including a low-cost, portable device for field use

Background. The Barcode of Life Project will draw on the taxonomic expertise and extensive collections of the world's natural history museums and herbaria. From the outset, the project will facilitate the growth and development of the taxonomic community of these institutions to provide and link to the accumulated and accumulating collateral information on species world-wide.

Timetable. What follows is a brief outline of possible initial steps. Developing a more complete implementation plan will involve additional input from conference participants and other interested experts.

January 2004. Barcode of Life working group submits proposal for a coordinating Secretariat for a Biodiversity Barcode Consortium of stakeholder institutions to be based at a major taxonomic center. Proposal requests support for 2 full-time positions (1 academic, 1 administrative) for 2 years, and for semi-annual Consortium meetings and for website construction and development of associated software. The proposal outlines a framework for a Consortium and Secretariat for the Barcode of Life including tentative structure, participants, funding, and timetable.

February 2004. Meeting at U. S. National Museum of Natural History, Smithsonian Institution, Washington, DC, to explore interests of US government agencies in participating in and funding the initiative.

April 2004. Foundational meeting of the International Barcode of Life Consortium, probably at the U. S. National Museum of Natural History, Smithsonian Institution, Washington, DC. The conference will mark the formal initiation of the Consortium, and refine its charter, goals, and operations. The meeting will focus on achieving balanced international participation, and explore relations with the Global Biodiversity Information Facility and other potential partners. The Conference will discuss actual and potential Demonstration Projects that could be largely completed in 2004 to prove the value of barcoding. It will explore how museums might integrate sequencing projects and coordinate them with non-museum efforts. It will set 5 and 10-year goals and timetables (including cost estimates) and develop and prioritize projects. It will assign responsibilities for writing of proposals to potential funding sources. It will set the agenda for the First International Barcode of Life Conference.

Late 2004. First International Barcode of Life Conference. It will include experts in plant and animal taxonomy, forensic sequencing, miniaturization of the sequencing process, collateral information management, and uses of biodiversity information, and draw on the taxonomic expertise and vast collections of the world's natural history museums. The conference will review and advance the state of the art, expand the worldwide community of interest in barcoding, and harmonize research efforts.

Late 2004. First meeting of the Barcode Secretariat and Executive Committee of the Consortium. After the Conference, with the staffing of the Secretariat completed, the leadership will meet to revise priorities for 2005 and to make the needed plans and proposals for 2006 and beyond.

Late 2004. After the Conference, visits with U.S. Government support (NSF, EPA, DOE, Dept Agriculture, possibly other agencies) and funders outside the US to gain support for the emerging Plan of the International Barcode of Life project.

VI. Pilot Projects

The goal of the Barcode of Life Project is to determine DNA barcodes all species of life, beginning now with eucaryotes. Pilot projects at key points will help guide the later, larger effort. Important criteria include scientific and technical value (e.g. validity of barcoding vis-à-vis existing taxonomy, new species discovery, evaluation of methods for DNA recovery from dried or otherwise preserved museum specimens), public appeal, economic or medical importance, and availability of funding. These initial forays should generate results within 1-2 years, and will depend heavily on initiatives already in motion for other agendas and on having PIs with a burning desire to move in this direction. The primary goal of pilot studies is to demonstrate to potential funding agencies that a large-scale DNA barcoding effort is worth investing in.

VII. References

1. Bromley DA. 2003. What criteria should be used to establish funding priorities? Physics Today
2. Folmer O, Black M, Hoeh W, Lutz R, Vrijenhoek R. 1994. DNA primers for amplification of mitochondrial cytochrome c oxidase subunit I from diverse metazoan invertebrates. Molecular Marine Biology and Biotechnology. 1994 3:294-299.
3. Hebert PDN, Cywinska A, Ball SL, deWaard JR. 2003a. Biological identifications through DNA barcodes. Proceedings of the Royal Society of London, Series B 270:313-322.
4. Hebert PDN, Ratnasingham S, deWaard JR. 2003b. Barcoding animal life: cytochrome c oxidase subunit 1 divergences among closely related species. Proceedings of the Royal Society of London, Series B 270: S596-599.
5. Hecht J. 2003. Every species has a 'barcode'. New Scientist 22 March 2003.
6. Janzen DH, 2004. Now is the time. Transactions Royal Society of London: in press.
7. Stoeckle M, 2003. Taxonomy, DNA, and the Barcode of Life. BioScience 53: 796-797.

Appendix 1. Summary of Proposed Pilot Projects

Important caveat: The following summaries are based on discussion at the meeting, and may include details that have not yet been verified. More information is available on the conference website (http://phe.rockefeller.edu/BarcodeConference/index.html) under "Criteria for Selecting Pilot Projects"

The following projects are arranged by species number.

Primates (world)

Species	300
New Species	Few
Scientific Value	Validation (comparison to mature taxonomy)
Technical Value	Chordate primers
Practical Application	Customs/Wildlife
Public interest	High
Conservation Priority	High
Taxonomic Expertise	High
Specimen availability	High (including frozen tissues)
Database	IPBIR database (contains species data? specimen data?)

Turtles (world)

Species	300
New Species	Possibly many (preliminary DNA studies show each species seems to be composed of 2 or more species)
Scientific Value	Validation; test of intra-specific variation by detailed population sampling; test of detection of hybrids
Technical Value	Chordate primers
Public interest	High
Conservation Priority	High in some areas
Taxonomic Expertise	High (relevant scientific community organized and eager to participate)
Specimen availability	High (including frozen tissues of virtually all recognized species)
Database	Digital database (of species ?) available

Skippers (ACG, Costa Rica)

Species	400+
New Species	Likely many sibling species
Scientific Value	Test of barcoding related to subspecies, intraspecific variation; proof of concept for linking adults and larva; proof of concept for DNA barcoding application to biodiversity survey
Technical Value	Demonstration project for international collaborative effort
Practical Application	Some are agricultural pests
Public interest	Some
Conservation Priority	ACG is a conservation site
Taxonomic Expertise	willing enthusiasts; pilot projects underway; interested institutions (USNM/SI; INBio, NHM); could integrate with TOL and PBI
Specimen availability	High
Database	All species, all specimens databased

Salamanders (world)

Species	500+
New species	Many
Scientific Value	Species discovery; test of barcoding re: high intraspecific variation)
Technical Value	chordate primers
Practical Application	
Public Interest	Some
Conservation Priority	High (eg CI Global Amphibian Assessment)
Taxonomic Expertise	High; existing research infrastructure (AmphibiaTree, HerpNET, AmphibiaWeb)
Specimen availability	High (70% of named species in frozen tissue collections; many already sequenced)
Database	

Vertebrate parasites (ACG, Costa Rica)

Species	940 (this is the number of vertebrate hosts)
New Species	Many (new genera and new species)
Scientific Value	Proof of concept for associating parasite life stages
Technical Value	Demonstration of international collaborative effort at biodiversity site involving diverse taxa
Practical Application	
Public Interest	Low
Conservation Priority	Low
Taxonomic Expertise	High (inventoried since 1995; existing network of >30 taxonomists; well-trained parataxonomists on site
Database	All specimens databased and on web (http://brooksweb.zoo.utoronto.ca/index.html)

Gulf of Maine megafauna

Species	1000
New Species	Few (except in benthic fauna)
Scientific Value	
Technical Value	Demonstration of geographic all-taxa survey
Practical Application	Many species of commercial importance; analysis of stomach contents for food webs
Conservation Priority	Interest in fish stock management
Taxonomic Expertise	
Specimen Availability	
Database	

Sphinx Moths (world)

Species	2000+ (given need to assess local and geographic variation, need 20 specimens each)
New Species	Likely
Scientific Value	identification of cryptic species
Technical Value	invertebrate primers, forensic techniques for museum DNA recovery
Practical Application	larva identification for life cycle, food webs (in bird meals)
Public interest	Some (enthusiastic amateurs)
Conservation Priority	?
Taxonomic Expertise	High (willing experts include and dedicated amateurs)
Specimen availability	High (essentially all species in 2 collections, NHM, other)
Database	World picture guide, checklist; NHM specimen collection is databased

Mosquitos (world)

Species	3500
New Species	Likely many
Scientific Value	Taxa have broad range of evolutionary histories, ages
Technical Value	Invertebrate primers, forensic techniques for DNA recovery from museum specimens
Practical Application	High; vectors for human disease; identification of larval forms for control programs
Public interest	Low
Conservation Priority	Low
Taxonomic Expertise	willing research community
Specimen availability	representative specimens in 2 museums (NHM, NMNH)
Database	World taxonomic digital catalog linked to online pdf files of most taxonomic literature; keys
Other	Funding opportunities related to disease including diagnostics for larva identification

Tephritid Fruitflies (world)

Species	4400
New Species	Many
Scientific Value	
Practical Application	Identification of agricultural pests and biological control agents
Public Interest	Med-fly is well-known.
Conservation Priority	Low
Taxonomic Expertise	High (also willing)
Specimen Availability	Possibly limited
Database	Digital names catalog soon (other info already at www.sel.barc.usda.gov/diptera/tephriti/tephriti.htm/

Birds (world)

Species	10,000
New Species	Few
Scientific Value	Validation; identification of cryptic species
Technical Value	Chordate primers
Practical Application	Identification of West Nile Virus (and other avian zoonooses) reservoir by analysis of mosquito blood meals Non-destructive identification of juvenile/non-breeding birds in banding studies
Public interest	High
Conservation Priority	Varied (identification of cryptic species may change conservation priorities)
Taxonomic Expertise	High
Specimen availability	High (much is available in a small number of collections)
Database	World checklists (Wells, Sibley); no specimen databases

Costa Rica/INBio national inventory

Species	200,000+
New Species	Many
Scientific Value	Species discovery
Technical Value	Demonstration of international collaborative biodiversity survey including parataxonomists
Practical Application	Other biodiversity surveys
Public Interest	Costa Rica has high name recognition as tropical conservation site
Conservation Priority	High
Taxonomic Expertise	Local and collaborative expertise available (400+ in dozens of countries)
Specimen Availability	High
Database	Entire collection is already databased (done at time of collection)

Nematodes

Species	Unknown; possibly >1,000,000
New Species	Vast; only 80,000 species are described
Scientific Value	High; including species discovery, ecology, human disease, agriculture
Technical Value	
Practical Application	Detection of human/plant disease agents
Public Interest	Low
Conservation Priority	Low
Taxonomic Expertise	DNA-based taxonomy effort underway (Blaxter, others); well organized research network (eg see http://nematode.unl.edu/)
Specimen availability	Likely to involve collecting in diverse environments
Database	