Molecular Ecology Resources (2009) **9** (Suppl. 1), 27–34 doi: 10.1111/j.1755-0998.2009.02629.x

BARCODING METHODOLOGY AND APPLICATIONS

The front-end logistics of DNA barcoding: challenges and prospects

ALEX V. BORISENKO, JAYME E. SONES and PAUL D. N. HEBERT
Canadian Centre for DNA Barcoding, Biodiversity Institute of Ontario, University of Guelph, Guelph, Ontario, Canada N1G 2W1

Abstract

Building a global library of DNA barcodes will require efficient logistics of pre-laboratory specimen processing and seamless interfacing with molecular protocols. If not addressed properly, the task of aggregating specimens may become the biggest bottleneck in the analytical chain. Three years of experience in developing a collection management system to facilitate high-throughput DNA barcoding have allowed the Canadian Centre for DNA Barcoding to recognize and resolve the most common logistical obstacles. Dealing with these challenges on a larger scale will be an important step towards building a solid collection-based foundation for the international DNA barcoding effort.

Keywords: biodiversity, biomaterials, collection management, high-throughput sequencing, natural history, taxonomy

Received 30 September 2008; revision received 27 November 2008; accepted 8 December 2008

Introduction

The concept of DNA barcoding — use of a short standardized genetic marker to provide rapid DNA-based identifications of organisms — gains increasing support as its baseline provisions and benefits to the biodiversity science community (Hebert *et al*. 2003a, b; Hebert & Gregory 2005; Savolainen *et al*. 2005; Schindel & Miller 2005) gain practical verification (e.g. Janzen *et al*. 2005; Hajibabaei *et al*. 2007; Waugh 2007; Borisenko *et al*. 2008; Frézal & Leblois 2008). The applicability of this approach depends on the efficient assembly of a comprehensive reference library of DNA barcodes, which in turn relies on consolidating large-scale efforts of experts in biodiversity science and genomics (Waugh 2007; Frézal & Leblois 2008). The key element of DNA-based approaches to taxonomy lies in establishing solid links to voucher specimens used in analyses, enabling subsequent taxonomic re-examinations (Ruedas *et al*. 2000). This critical prerequisite particularly applies to the assembly of reference DNA barcode data (Hebert & Gregory 2005; Ratnasingham & Hebert 2007; Hubert *et al*. 2008), thereby underpinning the key role of properly curated natural history collections.

Correspondence: A.V. Borisenko, Fax: (519) 824-5703,
E-mail: aborisen@uoguelph.ca

Early efforts to build barcoding capacity focused on designing the high-throughput molecular facilities and informatics platforms needed to create central hubs for the generation and storage of barcode data (Hajibabaei *et al*. 2005; deWaard *et al*. 2008). Current plans call for most samples to be funnelled through large core analytical facilities to aid quality control, adherence to barcode standards, and to minimize analytical costs. By contrast, the required specimens will originate from many collections and collaborating researchers. As a consequence, the core analytical facilities must consolidate samples and data from a multitude of sources and funnel them through a uniform analytical pipeline.

While it is apparent that the formation of a broad international research alliance capable of delivering high volumes of DNA barcode data relies on securing funds and on the coalescence of research agendas, its productivity will require a robust logistical framework. Importantly, the operational task of turning an aggregation of specimens into a laboratory-ready array of samples has received little attention.

The challenges currently facing natural history repositories have received extensive reviews (e.g. Krishtalka & Humphrey 2000; Winker 2004) and are beyond the scope of this study, as are the ethical and philosophical aspects of the adoption of DNA barcoding by museums and the

taxonomic community (e.g. Gregory 2005; Waugh 2007; Ellis 2008). Past studies have also focused on topics such as the general utility of natural history collections in molecular research (e.g. DeSalle & Amato 2004; Suarez & Tsutsui 2004), building genetic resource collections (e.g. Engstrom *et al*. 1999; Prendini *et al*. 2002), and salvaging DNA from archival specimens (e.g. Hajibabaei *et al*. 2006; Wandeler *et al*. 2007). Despite the importance of these issues, we do not revisit them, on the presumption that specimens for barcoding will be provided by contributors supporting the concept, and that samples will have been preserved in a DNA-friendly fashion, allowing the use of high-throughput techniques for DNA recovery (e.g. Ivanova *et al*. 2006; Ivanova & Grainger 2007). In this study, we employ a broad definition of natural history collections, including specialized permanent holdings (e.g. museums), as well as other professionally assembled aggregations of preserved organisms, such as temporary research collections maintained by field stations, small laboratories, and individual investigators. It is implied that all specimens used in barcoding are ultimately destined for archival storage in a specialized repository.

The varied logistical challenges of managing the specimen flow needed to support increased sequence production have been encountered by the Canadian Centre for DNA Barcoding (CCDB) at the Biodiversity Institute of Ontario (BIO), which began life as a medium-throughput population genetics laboratory. Since 2004, sequence production has scaled up, requiring the development of a collection management system which supports both the laboratory operations and the data requirements of the Barcode of Life Data Systems (BOLD). This activity has forced recognition and resolution of the most common bottlenecks in processing specimens. As a result, current practices provide a good working model of a logistical framework that supports efficient transactions between a core DNA barcoding hub and its external collaborators. This study outlines solutions that have gained operational validation, many of which can be implemented in other barcoding centres and smaller research laboratories. They may also have a broader application in streamlining field collecting efforts destined to supply materials for varied high throughput molecular analyses.

The challenges

Unlike biomedical and environmental biobanks which have well-established protocols for dealing with high-volume tissue exchange (Elliott & Peakman 2008), existing natural history collections are not designed to facilitate such usage. Moreover, each institution or researcher has a 'traditional' way of managing specimens and collection data. Transforming this diversity of approaches into a standard workflow compatible with high-throughput molecular protocols and barcode data requirements is a serious logistical challenge which needs to be addressed either at the point of collection, at the site of initial specimen deposition, or upon arrival of specimens at the DNA barcoding facility. The latter solution demands considerable time and effort from the front-line personnel of the core laboratory.

The first specimens barcoded in 2003 were managed by the Guelph barcoding facility from the point of collection to the generation of the sequence — an approach which continues for some specimens. Initially, the low production volume allowed integration of front-end specimen processing and laboratory analyses into one pipeline. However, as production increased, the coupling of sequence workflows and pre-laboratory stages reduced efficiency. Furthermore, this complicated troubleshooting in the event of human error, due to the necessity of revisiting the entire processing chain to determine where an error occurred.

Front-end (or pre-laboratory) processing as defined here includes the main steps required to transform specimens (or bulk tissue collections) into laboratory-ready arrays of tissue samples. It does not include the tasks involved in specimen collection or associated preparation (e.g. spreading, mounting or preparation of study skins), shipping, taxonomic curation, and long-term storage. Stages included are: preprocessing (accession registry), databasing of collection information, BOLD data upload, specimen arraying (assembly of specimen batches compatible with laboratory processing units), labelling, imaging (photography), tissue sampling, and disassembly of specimen arrays for return or long-term deposition.

Figure 1 illustrates the rise in DNA barcode records generated within the CCDB over a 3-year period and the numbers of voucher specimens processed internally. It is evident that the molecular pipeline outperforms the collection pipeline — a trend particularly apparent in the last 2 years, following the implementation of robotic protocols in 2006 and their fine-tuning in 2007. As opposed to most molecular techniques, pre-laboratory processing cannot be automated, requiring a more or less fixed amount of staff time per specimen. Hence, from today's perspective, it is the front-end stages — and not the laboratory analyses — that are likely to cause bottlenecks in the analytical chain.

The solutions

Internal workflows and organizational structure of a central barcoding hub

An increase in sequencing volume within a barcoding hub implies that its collection processing facilities need to expand proportionally, in order to ensure a sustained flow of samples for analysis; this naturally causes competition

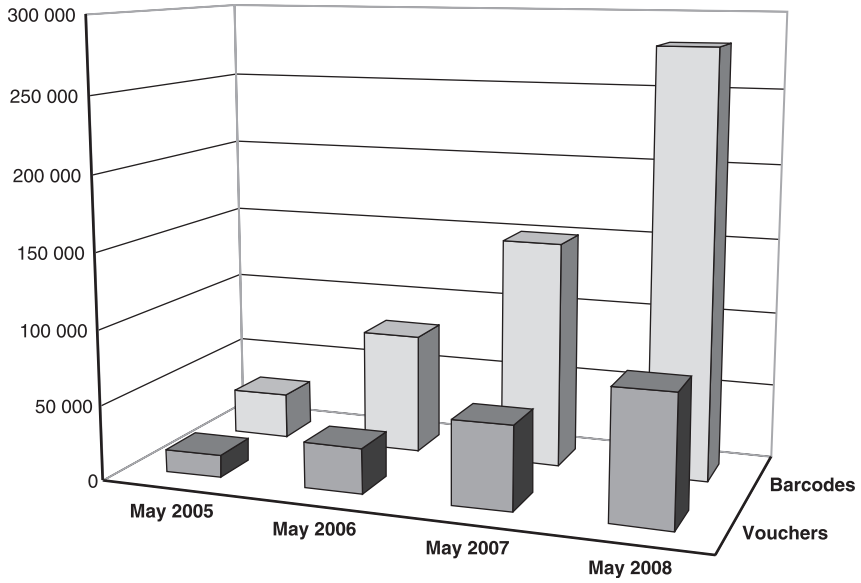

Fig. 1 Comparative accumulation of processed voucher specimens and DNA barcode sequences by the CCDB over a 3-year period.

Table 1 Main types of biomaterials processed at the CCDB and corresponding stages of their processing: Internal, stage carried out within the CCDB; External, stage handled by sample providers; N/A, stage not included in pre-laboratory processing

	Pipeline 1	Pipeline 2	Pipeline 3
Task description	Whole specimens	Arrayed tissue in tube racks	Arrayed tissue in microplates
Specimen collection	N/A	N/A	N/A
Specimen preparation	N/A	N/A	N/A
Preprocessing/accessioning	Internal	External	External
Data entry	Internal	External	External
Array specimens	Internal	External	External
Upload to BOLD	Internal	External	External
Labelling	Internal	External	External
Photography	Internal	External	External
Array tissue samples	N/A	External	External
Tissue sampling/subsampling	Internal	Internal	External
Disassembly/storage/loan	Internal	Internal or N/A	N/A
Taxonomic curation	N/A	N/A	N/A
Storage	N/A	N/A	N/A
Required front-end processing time per 96-well array	5–36 h	1–4 h	0

for resources with the core analytical processes. Alternatively, a barcoding node wishing to maximize its output while minimizing per-unit costs can strive to 'outsource' front-end processing to its sample providers — a solution discussed in more detail below. To maximize flexibility, collection management efforts within the CCDB were split between building an efficient front-end infrastructure and developing routines for external management of pre-laboratory stages. The first critical step was the complete detachment of specimen processing from molecular workflows which was fully implemented in late 2006.

Table 1 summarizes the three main categories of biological materials arriving at the CCDB for analysis and the corresponding analytical pipelines. Time estimates are approximate values based on recorded average technician time spent processing varied animal collections. These estimates vary dramatically, depending on many factors, including the taxonomic group and form of preservation of specimens; however, in any case the amount of investment significantly increases for unarrayed collections.

One of the vital elements in ramping up barcoding production is the adoption of high-throughput molecular protocols involving 96-well microplates (Hajibabaei *et al*. 2005; deWaard *et al*. 2008). While scaling up performance by two orders of magnitude and facilitating robotic compatibility, this format remains 'human-readable'. Correspondingly, the assembly of specimen arrays transferable into these laboratory-friendly batch units (95 samples and

Fig. 2 Example of a specimen array prepared for imaging and tissue sampling: 95 pinned Coleoptera are arranged in a gridded Schmidt box in alphanumerical order according to the sample IDs on their labels. The paper on the box lid contains a printout of the digital plate map which is used as a cross-reference to verify specimen locations during imaging and tissue sampling.

one empty control well) becomes the key stage of front-end processing. Typically, specimens are aggregated in a container in a format matching the future position of their corresponding samples in the microplate (Fig. 2). Their arrangement (plate map) is recorded in a digital spreadsheet (see Appendices I–II for an example) or database, and subsequently relayed to the Laboratory Information Management System (LIMS). The arraying stage is particularly error-sensitive and at the same time human-dependent, demanding robust workflows and well-trained personnel. Ideally, arraying should precede both imaging and sampling to allow tracking of the specimens as a batch through all pre-laboratory stages. Figure 2 provides an example of a specimen sampling/imaging array used in the CCDB.

The conversion of specimen data is another critical element of front-end processing, and data quality has a huge impact on the time and effort expended, especially if provenance information has to be digitized from specimen labels or field collection logbooks. Because the transfer of data is detached from the movement of specimens, it is critical to define unique catalogue numbers (sample IDs) linking the vouchers with corresponding tissue samples and data records. This seemingly trivial requirement needs emphasis, because individual numbering of specimens was not common practice in the pre-databasing era (with exception of birds and mammals), and this legacy remains in many natural history collections. Hence, the need to have globally unique specimen numbers is novel to some first-time collection providers.

The flow chart (Fig. 3) illustrates the typical interconnection of specimen processing stages within the CCDB collection management module and their correspondence

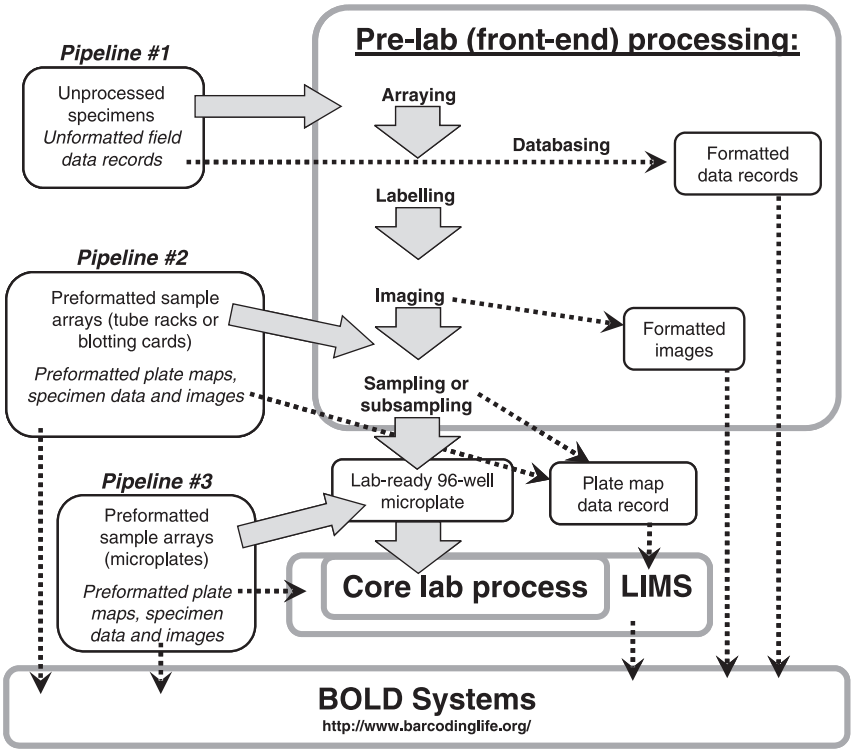

Fig. 3 Flow chart of front-end processing stages at the CCDB and the interfacing between the collection management module, the core analytical facility, BOLD and external collaborators. Pathways of direct interfacing between external collaborator and BOLD are not illustrated.

to the three major pipelines. Versatility is achieved through the establishment of a single point of contact (Pipeline 3) where samples enter the DNA laboratory via a standard operating protocol. If external collaborators supply materials requiring pre-laboratory processing, these stages are orchestrated by the collection management module, but if samples are submitted in laboratory-ready 96-well plates, they can enter Pipeline 3 directly. This enables independent adjustments to workflows within each module without adversely affecting production within the other and allows redeployment of staff to resolve emergent production bottlenecks. Finally, this model of organizational structure provides a core DNA barcoding node with the logistical flexibility to shift responsibility for certain pre-analytical stages to its external sample providers.

Functions of the collection module

Currently, the CCDB collection management module handles all required front-end stages, including preprocessing, databasing, labelling, imaging, and assembly of sampling arrays. It also facilitates specimen preparation and taxonomic curation either externally or in-house.

Since the primary and ultimate goal of a central barcoding hub lies in the generation of molecular data, its need for a collection management module might be questioned. Theoretically, sample providers could assume all front-end stages and simply use the DNA barcoding hub as a sequencing factory and online data repository for their samples. However, practical experience at the CCDB shows that having a collection management module provides a barcoding node with the flexibility required to deal with numerous 'exceptional' cases where additional pre-laboratory processing is required. Furthermore, it allows the hub to address ancillary functions associated with large volumes of specimen transactions.

Because the main function of large natural history repositories is to store biological materials and to loan them to multiple external researchers, they act as major donors. Typical users of these collections are individual researchers who request a small number of samples for projects with a tight taxonomic focus. By contrast, a central DNA barcoding facility acts as a major recipient of biomaterials from multiple providers. Although it has a well-delimited research focus, the volume and taxonomic scope of specimens received is orders of magnitude greater than that of an average research laboratory, matching that of a large repository. The volume of incoming loans to the central barcoding hub requires it to have a special unit for facilitating these transactions. Just as collection repositories develop loan policies and Biological Material Transfer Agreements (BMTA) that govern the outflow of specimens, a DNA barcoding facility requires clear-cut guidelines stipulating how incoming samples will be stored and utilized. One important early task facing the CCDB collection module involved the development of a tissue policy statement and BMTA (Appendices IV and V). Because loan conditions differ, depending on the agreement with each donor, a barcoding hub must track the history and ownership rights of the specimens and the DNA extracts that it holds. Aside from developing internal protocols and policies for sample management and establishing agreements with its external research partners, a DNA barcoding node must ensure than any secondary loans comply with applicable regional, federal, and international regulations. This, combined with the taxonomic diversity of specimens processed, can place barcoding centres under considerable administrative pressure. These issues are best dealt with by specialized staff handling biomaterial transactions. While this function is probably unique to large barcoding hubs, the overall organizational model and routine workflows of the CCDB collection management module can be applied with minimal adjustments in medium and large-scale field-based operations collecting samples for DNA barcoding and other high-throughput molecular analyses.

External management of front-end processing

Since the CCDB gained activation, the volume of tissue samples has continually scaled up (Fig. 1). This trend makes a central DNA barcoding hub increasingly dependent on its network of external sample providers. While acknowledging the complex ethical and economical aspects of building collaborative partnerships underlying this process of 'front-end outsourcing' [which are addressed by the emerging International Barcode of Life Project (iBOL)], we only focus here on its logistical framework. Aside from apparent advantages to the core facility, decentralization of pre-laboratory management offers benefits to external collaborators, which include, for example, retaining possession of voucher specimens and bulk tissue, removing the risk of specimen damage during transfer and reducing shipping restrictions.

The key element in transferring the initial stages of specimen processing to external collaborators involves strict adherence to standard protocols for sample and data submission set by the core facility (Hajibabaei *et al*. 2005), which apparently have to be balanced with their own collection management needs. To ensure 'compliance', the central hub must provide comprehensive information packages and guidance to its sample donors to minimize labour and human error. As BOLD projects become organized into campaigns with a certain taxonomic and/or geographical focus, the need builds for campaign-based guidelines for contributors, outlining best practices for specimen and data management, specific to certain taxonomic or ecological groups. Such specialization is particularly important for facilitating pre-analytical screening

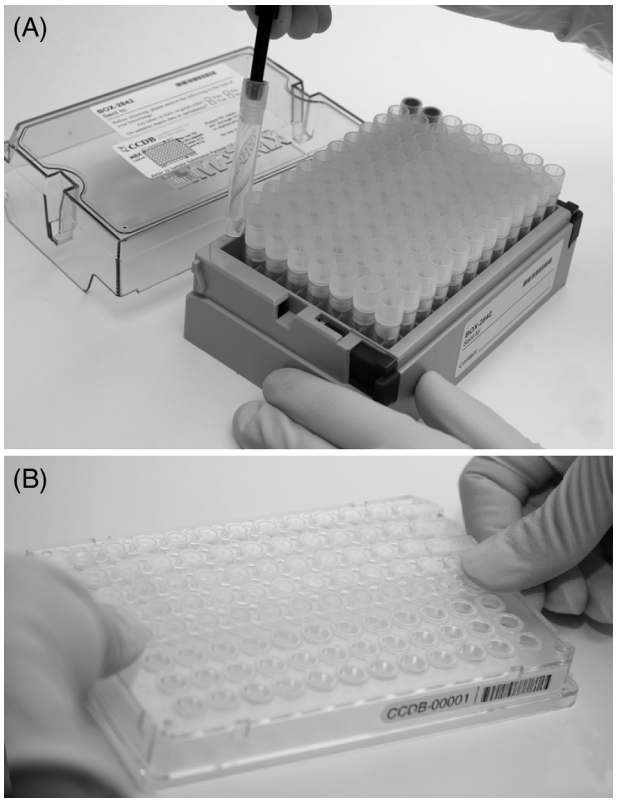

Fig. 4 Standard sampling media compatible with the high-throughput 96-well format, adopted by the Canadian Centre for DNA Barcoding (CCDB): (A) Tube rack; (B) Microplate.

for DNA quality, which must account for taxon-dependent variation in commonly used preservation methods. At the same time, it is important for central nodes to reduce divergence between their basic requirements to allow the use of standardized sample and data submission protocols across the barcoding community.

The introduction of TrakMate sample tube racks (a.k.a. 'Matrix boxes' — Fig. 4A) in 2004 was the first step towards adjusting sampling protocols to a 96-well high-throughput analytical format. However, as production scaled up, it became apparent that this medium was not optimal, because of the time required to subsample tissues arriving in tube racks (Table 1). As well, the boxes are bulky, adding costs in shipment. Arraying samples into 96-well format can easily become a significant barrier to sequence production as noted by others (Richardson *et al*. 2006). In June 2008, CCDB sampling protocols were altered to unify the array order across all sampling media and to make it more intuitive. One key element of this transition involved the option of sampling directly in 96-well microplates (Fig. 4B). While requiring little additional effort from donors, this significantly reduces pre-processing time and removes a step from the analytical chain, excising one potential source of error. An incentive is now in place to have microplate-based protocols become the central form of sample transactions across the iBOL community. Appendices SI–SII provide more details on the new sampling procedures.

Similar approaches are needed to streamline data submission. For large contributors possessing their own collection databases, such as museums, development of data conversion routines can greatly enhance data flow. However, it is arguably even more important to produce custom data curation solutions for medium and small-scale field collecting operations which do not maintain their own databases and data standards. These solutions have to be simple and intuitive to allow their efficient use by individuals with limited databasing background, but at the same time provide sufficient benefits to justify their adoption, such as the ability to output data in a variety of formats (e.g. BOLD, Google Earth, or Darwin Core) or the option to generate specimen labels instantly after data entry. In our experience, templates built on the platform of electronic spreadsheets (see Appendix SIII for an example) offer the required combination of simplicity and versatility, while greatly enhancing data quality and reducing time spent on conversion, validation, and parsing data entries.

The prospects

The first 'barcoding sweep' of global taxonomic diversity is likely to be accomplishable with the collection resources of relatively few large contributors, which hold the majority of natural history collections. However, the filling of taxonomic gaps will become increasingly dependent on small research groups or experts focused on 'rare' taxa or 'exotic' geographical areas. This will imply a dramatic increase in the number of providers submitting small batches of samples. If all these transactions are handled directly by core DNA barcoding laboratories, they risk getting overwhelmed or will need to expand their front-end processing modules — a shift that will adversely affect their analytical capacity. A more feasible and natural strategic solution will involve 'outsourcing' the coordination of regional barcoding projects to a network of satellite hubs specialized in pre-laboratory sample processing. These hubs should be based on existing biodiversity research facilities which will agree to adopt barcoding as part of their operation. The strategic trend towards decentralizing both front-end logistics and research administration will involve a more complex structure of hierarchical relationships between sample providers than envisioned earlier. It also creates a different perspective on the concept of satellite laboratories proposed in the early years of barcoding (Hajibabaei *et al*. 2005), which would be supplying core centres with PCR products. Emphasis was then placed on building the analytical infrastructure for these hubs, which today could be achieved with a relatively modest investment. It is now becoming clear

that pre-laboratory stages require similar attention and dedicated effort, and that the possession of molecular facilities is insufficient to ensure successful integration into barcoding workflows. In effect, willingness to assume the logistical burden of front-end processing and efficiency in coordinating research projects will determine the potential for becoming a barcoding hub.

As the DNA barcoding community expands and becomes more structured, satellite hubs will naturally crystallize from research groups or institutions willing to take on functions of pre-laboratory specimen processing with a particular regional or taxonomic focus. These nodes will be ideally positioned to coordinate the acquisition of samples at regional and/or national levels. Over time, these hubs will, as well, have the highest potential to develop their molecular laboratory capacities bottom to top by gradually incorporating additional analytical stages until eventually they function as core barcoding laboratories. These facilities will then take a lead role in the practical application of DNA barcodes on a large scale and contribute to further refinement of the reference library.

This model for the development of regional barcoding nodes conflicts with current priorities to develop high-tech capacity. It is symptomatic that many institutions are primarily interested in assembling cutting-edge molecular technology, with little regard for the prospective suppliers of tissues needed to maintain production in these facilities. This is not to mention the decreasing popularity of taxonomy and collection-based research and the diminishing public appreciation of the value of natural history collections (Dalton 2003; Gropp 2003; Suarez & Tsutsui 2004). Resolution of these obstacles must be taken very seriously by the international DNA barcoding community. Recognizing and addressing the challenges of front-end logistics represents an important first step in this direction.

Acknowledgements

We thank BIO-CCDB staff for their contribution at all stages of evolution of the collection management system. Robert Dooh developed original standards for shipping tissue samples and later elaborated specimen imaging workflows. Brianne St. Jacques has managed sample and specimen transactions and tested the efficiency of specimen transaction protocols. Natalia Ivanova, Chris Grainger and Janet Topan helped to develop solutions for interfacing between the laboratory and front-end processing. Sujeevan Ratnasingham, Gregory Downs, Megan Milton and Pia Marquardt provided valuable insights on BOLD data requirements and facilitated BOLD data submission and validation. Rick Turner developed and/or refined standard imaging setups for varied kinds of specimens and assisted with illustrative photos. Robert Hanner advised on general questions pertaining to the logistics of high throughput tissue analyses. Xin Zhou, Rodolphe Rougerie, Evgeny Zakharov, David Porco, Tomislav Terzin, and Alex Smith evaluated the utility of collection management workflows. Suzanne Bateson aided in the design of standard sampling instructions and the flowchart for this paper. Sarah Adamowicz offered helpful feedback on form and structure of the manuscript; Kevin Winker and an anonymous referee made further comments. Special thanks to our collaborators, particularly Daniel Janzen, Jean-François Landry, Tanya Dapkey, and Andrea Brauner for field testing the sampling protocols and providing valuable user feedback.

Pre-laboratory specimen processing at the Biodiversity Institute of Ontario was supported by grants to PDNH from the Gordon and Betty Moore Foundation, Genome Canada through the Ontario Genomics Institute, the Canada Foundation for Innovation, the Ontario Innovation Trust and NSERC.

Conflict of interest statement

The authors have no conflict of interest to declare and note that the funders of this research had no role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript.

References

Borisenko AV, Lim BK, Ivanova NV, Hanner RH, Hebert PDN (2008) DNA barcoding in surveys of small mammal communities: a field study in Suriname. *Molecular Ecology Resources*, **8**, 471–479.

Dalton R (2003) Natural history collections in crisis as funding is slashed. *Nature*, **423**, 575.

DeSalle R, Amato G (2004) The expansion of conservation genetics. *Nature Reviews Genetics*, **5**, 702–712.

Elliott P, Peakman TC (2008) The UK Biobank sample handling and storage protocol for the collection, processing and archiving of human blood and urine. *International Journal of Epidemiology*, **37**, 234–244.

Ellis R (2008) Rethinking the value of biological specimens: laboratories, museums and the Barcoding of Life Initiative. *Museum and Society*, **6**, 172–191.

Engstrom MD, Murphy RW, Haddrath O (1999) Sampling vertebrate collections for molecular research: practice and policies. In: *Managing the Modern Herbarium: An Interdisciplinary Approach* (eds Metsger D, Byers S), pp. 315–330. Elton-Wolf, Vancouver, Canada.

Frézal L, Leblois R (2008) Four years of DNA barcoding: Current advances and prospects. *Infection, Genetics and Evolution*, **8**, 727–736.

Gregory TR (2005) DNA barcoding does not compete with taxonomy. *Nature*, **434**, 1067–1067.

Gropp RE (2003) Are University natural science collections going extinct? *Bioscience*, **53**, 550.

Hajibabaei M, deWaard JR, Ivanova NV *et al*. (2005) Critical factors for assembling a high volume of DNA barcodes. *Philosophical Transactions of the Royal Society B: Biological Sciences*, **360**, 1959–1967.

Hajibabaei M, Singer GAC, Hebert PDN, Hickey DA (2007) DNA barcoding: how it complements taxonomy, molecular phylogenetics and population genetics. *Trends in Genetics*, **23**, 167–172.

Hajibabaei M, Smith MA, Janzen DH *et al*. (2006) A minimalist barcode can identify a specimen whose DNA is degraded. *Molecular Ecology Notes*, **6**, 959–964.

Hebert PDN, Cywinska A, Ball SL, deWaard JR (2003a) Biological identifications through DNA barcodes. *Proceedings of the Royal Society B: Biological Sciences*, **270**, 313–321.

Hebert PDN, Gregory TR (2005) The promise of DNA barcoding for taxonomy. *Systematic Biology*, **54**, 852–859.

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 B: Biological Sciences*, **270**, S96–S99.

Hubert N, Hanner R, Holm E *et al*. (2008) Identifying Canadian freshwater fishes through DNA barcodes. *PLoS ONE* **3**, e2490.

Ivanova NV, deWaard JR, Hebert PDN (2006) An inexpensive, automation-friendly protocol for recovering high-quality DNA. *Molecular Ecology Notes*, **6**, 998–1002.

Ivanova NV, Grainger CM (2007) CCDB protocols, sequencing. Retrieved from http://www.dnabarcoding.ca/CCDB_DOCS/CCDB_Sequencing.pdf on 24 September 2008.

Janzen DH, Hajibabaei M, Burns JM *et al*. (2005) Wedding biodiversity inventory of a large and complex Lepidoptera fauna with DNA barcoding. *Philosophical Transactions of the Royal Society B: Biological Sciences*, **360**, 1835–1845.

Krishtalka L, Humphrey PS (2000) Can natural history museums capture the future? *Bioscience*, **50**, 611–617.

Prendini L, Hanner R, DeSalle R (2002) Obtaining, storing and archiving specimens and tissue samples for use in molecular studies. In: *Techniques in Molecular Systematics and Evolution* (eds DeSalle R, Giribet G, Wheeler W), pp. 176–248. Birkhauser Verlag, Basel, Switzerland.

Ratnasingham S, Hebert PDN (2007) BOLD: the Barcode of Life Data System (www.barcodinglife.org). *Molecular Ecology Notes*, **7**, 355–364.

Richardson DE, Vanwye JD, Exum AM, Cowen RK, Crawford DL (2006) High-throughput species identification: from DNA isolation to bioinformatics. *Molecular Ecology Notes*, **7**, 199–207.

Ruedas LA, Salazar-Bravo J, Dragoo JW, Yates TL (2000) The importance of being earnest: what, if anything, constitutes a 'specimen examined?' *Molecular Phylogenetics and Evolution*, **17**, 129–132.

Savolainen V, Cowan RS, Vogler AP, Roderick GK, Lane R (2005) Towards writing the encyclopaedia of life: an introduction to DNA barcoding. *Philosophical Transactions of the Royal Society B: Biological Sciences*, **360**, 1805–1811.

Schindel DE, Miller SE (2005) DNA barcoding a useful tool for taxonomists. *Nature*, **435**, 17–17.

Suarez AV, Tsutsui ND (2004) The value of museum collections for research and society. *BioScience*, **54**, 66–74.

deWaard JR, Ivanova NV, Hajibabaei M, Hebert PDN (2008) Assembling DNA barcodes: analytical protocols. In: *Environmental Genomics, Methods in Molecular Biology*, Vol. 410 (ed. Martin CC), pp. 275–283. Humana Press, Totowa, New Jersey.

Wandeler P, Hoeck PEA, Keller LF (2007) Back to the future: museum specimens in population genetics. *Trends in Ecology & Evolution*, **22**, 634–642.

Waugh J (2007) DNA barcoding in animal species: progress, potential and pitfalls. *Bioessays*, **29**, 188–197.

Winker K (2004) Natural history museums in a postbiodiversity era. *BioScience*, **54**, 455–459.

Supporting information

Additional supporting information may be found in the online version of this article:

Appendix SI Standard sampling instructions for 96-well microplates distributed by BIO to its external collaborators.

Appendix SII Standard CCDB record file for entering plate map information.

Appendix SIII Example of a field data entry sheet designed to facilitate basic in-field collection management, taxonomic curation and data validation. Data output module allows easy conversion of specimen collection information into BOLD data submission spreadsheets. Such spreadsheets are provided upon request to specimen sample providers who lack designated collection management facilities and databases.

Appendix SIV BIO-CCDB tissue policy statement.

Appendix SV BIO-CCDB Biological Material Transfer Agreement template.

Please note: Wiley-Blackwell are not responsible for the content or functionality of any supporting materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.