CELL PRESERVATION TECHNOLOGY
Volume 5, Number 2, 2007

DOI: 10.1089/cpt.2007.9993

Genomic Diversity Research and the Role of Biorepositories

ROBERT H. HANNER and T. RYAN GREGORY

ABSTRACT

Biodiversity repositories underpin the future of research in the life sciences and biotechnology. However, they represent an extremely heterogeneous assemblage of collections that are lacking a comprehensive index of available resources. A set of "best practices" for biospecimen characterization is proposed for repository biomaterials involving standardized species-level molecular genotyping (DNA barcoding) and the quantification of nuclear DNA content (genome size). This approach has implications for upstream sample collection and preservation methods, as well as downstream implications for highlighting biorepository specimens available for genetic and genomic research. The broad application of the approach here proposed will raise the profile of participating biodiversity repositories, facilitate the compilation of validated reference sequences for molecular species recognition, and drive a deeper understanding of the evolution of the genome.

INTRODUCTION

THE EXTENSIVE COLLECTIONS of specimens contained in natural history museums and other repositories of preserved biological material provide a critical resource for understanding the diversity of life on Earth. The value of these collections is enormous, although they have traditionally targeted the preservation of organismal phenotypes because this is the primary basis for species recognition in most groups of larger organisms. Over the last several decades, however, numerous technological advances have facilitated the direct examination of genes, chromosomes, and increasingly entire genomes. Thus, biodiversity science, like many other disciplines in the life sciences, has undergone drastic changes following the molecular revolution and the dawn of the postgenomic era. Genetic information can aid resolution of species boundaries and also help to elucidate the patterns and rates of evolutionary diversification among species.

The highly productive merger of biodiversity and genomic sciences highlights the need to collect, preserve, and archive biomolecules in a recoverable state from a broad range of species. Here we outline two major areas of genomic diversity research—DNA barcoding and genome size diversity—in which biological repositories can play a significant role in enhancing both genomics and biodiversity science. This includes the use of materials already present in collections as well as the development of both preservation and analysis protocols that could enhance the utility of repository materials for this purpose in the future. In addition, we highlight how these approaches to characterizing genomic diversity could, in turn, be very beneficial for quality control applications in biorepositories.

Department of Integrative Biology and Biodiversity Institute of Ontario, University of Guelph, Guelph, Ontario, Canada.

DNA barcoding

To date, nearly two million species have been formally described out of an estimated 10–100 million species. For most people, information about the biological diversity by which they are surrounded remains inaccessible. What is needed is a simple, cost-effective means of identification that will serve as a portal to the wealth of information that has accrued through the efforts of professional taxonomists. This evident need, combined with the explosive growth of molecular databases, has led to the recognition that short DNA sequences taken from a standardized region of the genome can be used to discriminate related species.[1] In particular, DNA-based species identification attempts to match an unknown specimen to a known reference sample by comparing sequences of genes that are known to vary between species. Mitochondrial genes are obvious targets for this approach given their accelerated rate of evolution relative to nuclear genes and because they typically lack recombination. They are also present in high copy number compared to nuclear DNA, allowing a greater yield of mitochondrial DNA (mtDNA) from a given sample, which enhances the ability to extract sequences from small or degraded specimens. These properties, combined with the existence of highly conserved flanking sites for which "universal" primers have been developed,[2] resulted in the suggestion that the nucleotide sequence of 5′ end of the mtDNA gene cytochrome *c* oxidase I (COI) could be used as molecular "barcode" for most animal life.[3]

The intent of DNA barcoding is to use one or a few reference genes in order to assign newly sampled individuals to species, or to identify cases in which no name yet exists for a given species.[4,5] Despite some recognized limitations,[6] COI has so far succeeded in providing species level resolution across diverse groups of insects,[7] birds,[8] fishes,[9] and primates.[10] COI sequences also demonstrate the ability to differentiate species in other compartments of life, including protists[11–13] and fungi.[14]

At the time of this writing, more than 240,000 COI barcode records are deposited in the Barcode of Life Database (BOLD)[15] with representation from more than 27,000 species. DNA barcoding is a rapidly expanding field of research that is facilitated by an international consortium of biological resource centers and other organizations[16] (see: http://barcoding.si.edu/). Importantly, the COI gene was also validated recently for use in forensic analysis.[17] The key challenge remains the acquisition of vouchered specimens from tens of thousands of species, to which existing biological repositories are able to make a major contribution.

Genome size

In 1948, Boivin et al.[18] noted a "remarkable constancy" in the amount of nuclear DNA among tissues within individual organisms and across individuals within species. At a time before its function or structure were established, this was taken as evidence in favor of DNA as the hereditary molecule. A few years later, surveys of the quantity of DNA per haploid genome indicated that, contrary to expectation, genome size (or "C-value") is not related to the complexity of the organism nor to the presumed number of protein-coding genes.[19] By 1971, this counterintuitive observation had become known as the "C-value paradox."[20] The discovery of noncoding DNA soon resolved the "paradox," but many questions remain to this day regarding genome size diversity. Indeed, genome size and noncoding DNA have become even more relevant in the postgenomic era for both theoretical and practical reasons. In particular, there is increasing interest in the causes and consequences of noncoding DNA spread and loss, and genome size itself is recognized as a key consideration in the choice of future targets for both complete genome sequencing[21–23] and evolutionary developmental biology ("evo-devo") research.[24]

At present, genome size estimates are available for more than 10,000 species of animals, plants, and fungi. These data are made freely available through online databases,[25] and have revealed many important insights regarding the large-scale diversity and evolution of eukaryotic genomes. This includes identifying links between bulk DNA content and cell size, cell division rate, body size, metabolic rate, de-

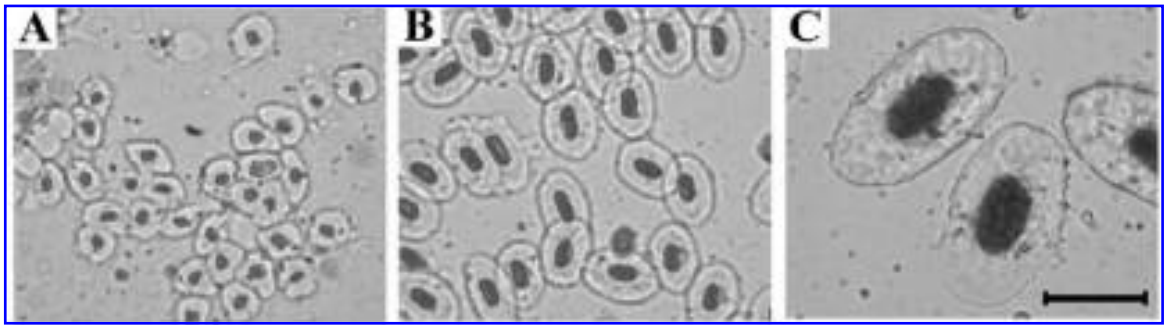

FIG. 1. An example of the strong link between DNA content and red blood cell size. These are photomicrographs taken at the same magnification (40×) from (**A**) Siamese fighting fish (*Betta splendens*) (2C = 1.3 pg), (**B**) Rainbow trout (*Oncorhynchus mykiss*) (2C = 5.2 pg), and (**C**) Australian lungfish (*Neoceratodus forsteri*) (2C = 105 pg). Nuclear DNA is stained dark. Scale bar equals 20 μm.

velopmental rate, and other important features[26,27] (Fig. 1). However, in order to gain a reliable understanding of the causes and effects of genome size diversity, and to maximize the applied utility of this information, it is important to obtain data from as wide a range of organisms as possible. Unfortunately, this has not yet been achieved. Some groups—especially invertebrates and microorganisms—remain very poorly sampled, and a great deal of work remains in cataloging and accounting for the diversity in genome size in such groups. Once again, this is an area in which biorepositories stand to improve the situation dramatically.

METHODS IN GENOMIC DIVERSITY RESEARCH

The methods involved in genomic diversity research are relatively straightforward to implement once equipment and expertise are in place, but both involve particular requirements for tissue preservation to be implemented using existing (and future) collections. Here we outline the basic methods underlying both areas of genomic diversity analysis.

Barcoding biodiversity

DNA barcoding is relatively inexpensive, and can be carried out in most well-equipped molecular biology labs with access to a DNA sequencing facility. The analytical procedure involves six routine steps: DNA extraction, PCR amplification, PCR check (gel electrophoresis), labeling of the PCR product for sequencing, cleanup of the labeled PCR product, and sequencing. Many of these steps can be automated for high-throughput screening,[28] which further reduces the cost per sample.

From an analytical perspective, it is important to distinguish between barcoding an unknown specimen to gain an identification versus barcoding to compile the reference sequence library (the latter is depicted in Fig. 2). Although the analytical steps necessary to produce a DNA sequence are the same for both procedures, building the barcode reference sequence library involves adhering to a set of community consensus-derived BARCODE data standards.[29] The standards involve capturing structured metadata associated with expert-identified voucher specimens held in a repository that are used for barcode analysis. They also require following a set of analytical procedures that includes bidirectional sequencing of the barcode region and retention of the raw data in the form of electropherogram "trace" files for archival with the National Center for Biotechnology Information's (NCBI) Trace Archive. This is critical in order for the companion DNA sequence accession in GenBank to be annotated with the reserved keyword BARCODE, which serves to differentiate records that are compliant with the data standards from other sequence entries in GenBank that lack this level of validation.[29]

Researchers can submit their BARCODE compliant sequence accessions directly to GenBank using the NCBI's Barcode Submission Tool (BarSTool). However, most large-scale

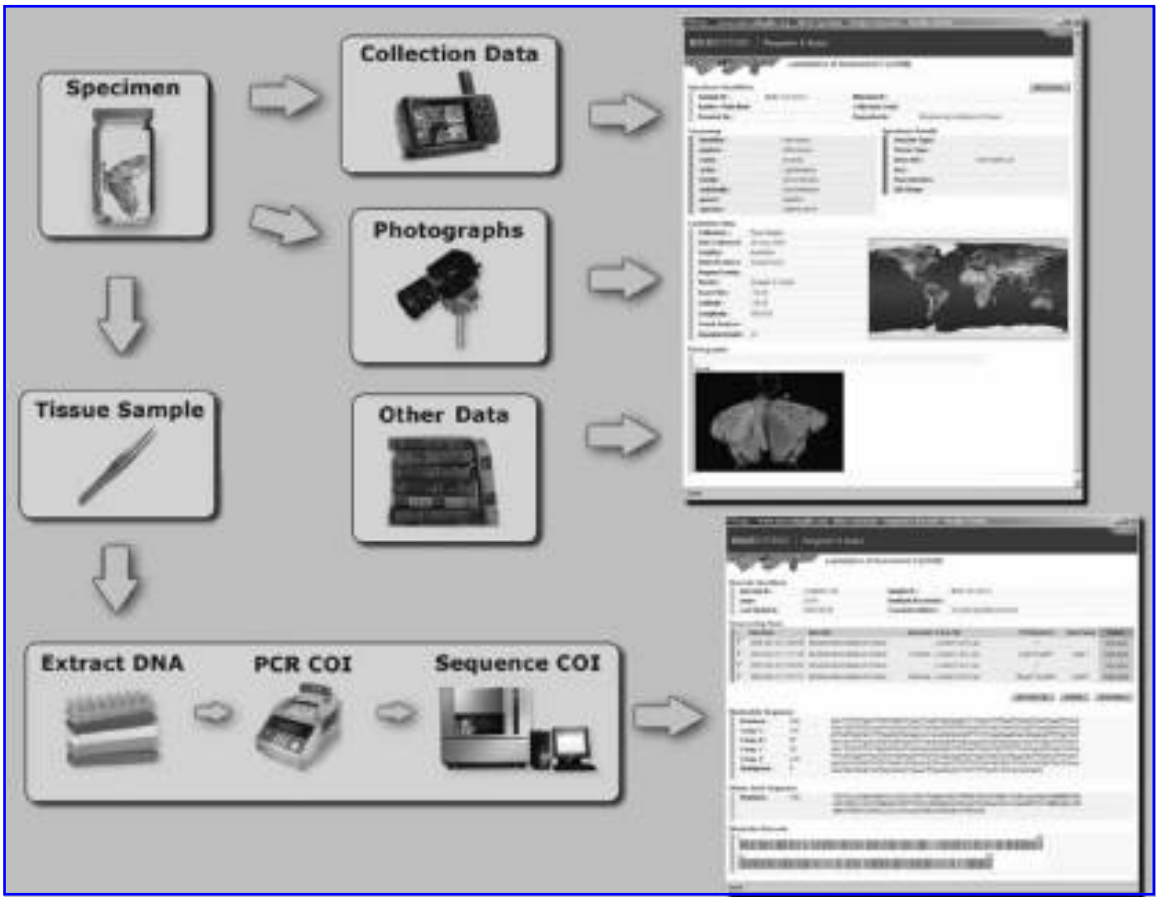

FIG. 2. A graphical depiction of the analytical chain involved in the generation of a DNA barcode reference library. DNA is extracted from the specimen and the COI gene is amplified through PCR and sequenced. The sequence data, collection, voucher specimen details, and other information are included in the reference database, which can be accessed freely online via the Barcode of Life Data Systems web page (http://www.barcodinglife.org/) and used to provide identifications of representatives of the same species collected subsequently.

barcoding initiatives have adopted the BOLD platform[15] as a workbench for assembling their projects, since BOLD supports the data structure necessary for gaining BARCODE annotation and can submit data to the GenBank automatically.

Once a DNA barcode library has been assembled, users need only input the COI sequence of an unidentified specimen to perform a query of the database. This returns either a species identification, or indicates that no barcode sequence currently exists for the species in question. Although simple from a molecular biology standpoint and much more accessible than traditional means of identification (i.e., consultation with a physical reference collection or a taxonomic expert), this process remains limited to those with access to DNA sequencing equipment (either their own or that of a sequencing center). In the future, DNA identifications are expected to be made widely available to nonscientists through the continued reduction in size and cost of DNA sequencing hardware. In principle, it should even be possible to develop a hand-held "DNA barcoder" capable of returning an accurate identification of any small amount of tissue within minutes at a cost of a few cents per identification.

Sizing up genomes

The quantification of nuclear DNA is generally accomplished via one of two well-established methods: Feulgen densitometry or flow cytometry. These methods differ in their physical and chemical underpinnings, but both are

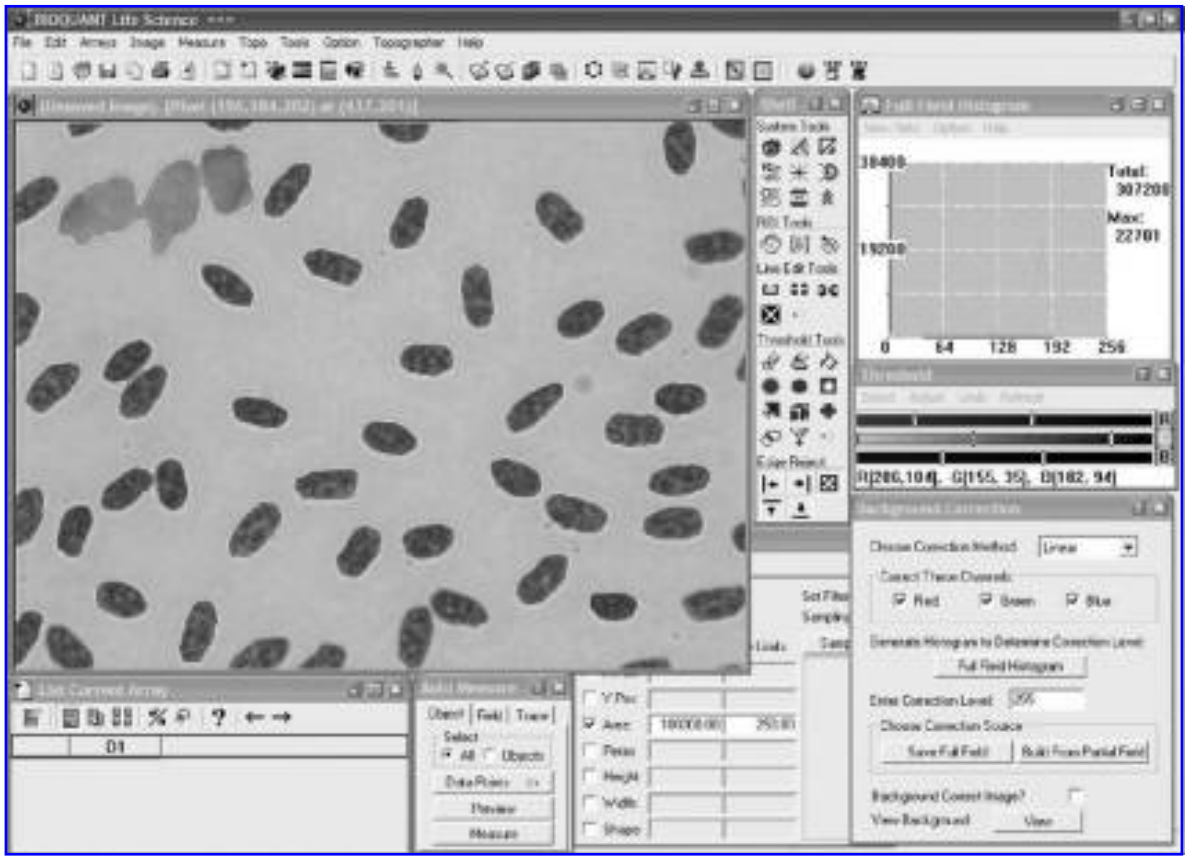

FIG. 3. Screenshot of the BIOQUANT Life Science software, one of several image analysis packages that can be used for genome quantification by Feulgen image analysis densitometry.[31] This figure shows an analysis being performed on Feulgen-stained chicken erythrocytes. The method is applicable to a wide range of cell types, so long as intact nuclei of known ploidy can be prepared in a monolayer and stained for analysis. Photo is reproduced courtesy of BIOQUANT Image Analysis Corporation.

based on the same general approach of quantifying the amount of bound stain or dye relative to a standard of known DNA content.

The Feulgen reaction has been used in the analysis of DNA content for nearly 6 decades, and involves depurinating DNA by hydrolysis with strong acid (usually 5N HCl) followed by staining with Schiff reagent, which binds to the aldehyde groups freed during hydrolysis and turns from colorless to pink. The density (absorbance) of the stain in the nucleus is then calculated from comparisons of transmitted light through the nucleus versus in an unobstructed beam. Because nuclei differ in size and vary in their compaction and distribution of DNA, it is necessary to examine a series of individual point densities, the sum of which is given by the integrated optical density (IOD) of the nucleus. Mean IODs for unknowns and standards are then compared to provide an estimate of the total DNA content per nucleus of the unknown. Traditionally, IOD measurements have been carried out with specialized densitometry equipment, but more recently this has been largely replaced by computerized image analysis systems that use pixels in captured images of microscopic fields as individual point densities and can return results much more quickly by analyzing multiple nuclei simultaneously (Fig. 3).

Flow cytometry does not involve a measure of light absorption, but rather the light emitted by a fluorescent dye excited by an intense light source. In this case, cells or (more commonly) free nuclei are prepared in suspension and treated with a fluorochrome prior to being passed through a flow cytometer. Depending on the fluorochrome, light of a specific wavelength (e.g., a laser or UV mercury arc lamp) is used to stimulate fluorescence of individual

nuclei in flow, with a series of detectors present to quantify the intensity of emitted light. As with Feulgen image analysis densitometry (FIA), this is compared against a known standard; however, in this case the standard and unknown are analyzed together in the same preparation (Fig. 4).

There are advantages and disadvantages to both FIA and flow cytometry (FCM), such that these methods should be considered complementary rather than in competition. FIA employs prepared samples on microscope slides which require only air drying, which makes the sampling of some tissues (e.g., vertebrate blood) very straightforward in the lab as well as in the field and facilitates shipping among institutions. A relatively small number of nuclei (dozens to a few hundred) are required to obtain accurate measurements, and tissue-specific analyses are easily accomplished because the nuclei are visualized on screen while measurements are being carried out. Feulgen-stained slides can be kept for long periods of time and measured repeatedly without ill effects on accuracy. However, preparing slides for some tissues can be difficult[30] and differences in DNA compaction level among nuclei can have a significant impact on the resulting density estimate.[31] FCM is more automated than FIA, and in some cases can be more rapid as a result. FCM uses large samples of nuclei (thousands) and is amenable to a wider variety of tissues as the only requirement is that individual cells/nuclei be freed; this is usually accomplished by simple mechanical and chemical processing of whole tissues.[32] On the other hand, FCM requires fresh or frozen samples, and these can only be analyzed once because fluorescence fades relatively quickly. FCM is also less conducive to tissue-specific analyses, as intact tissues are not examined with this method.

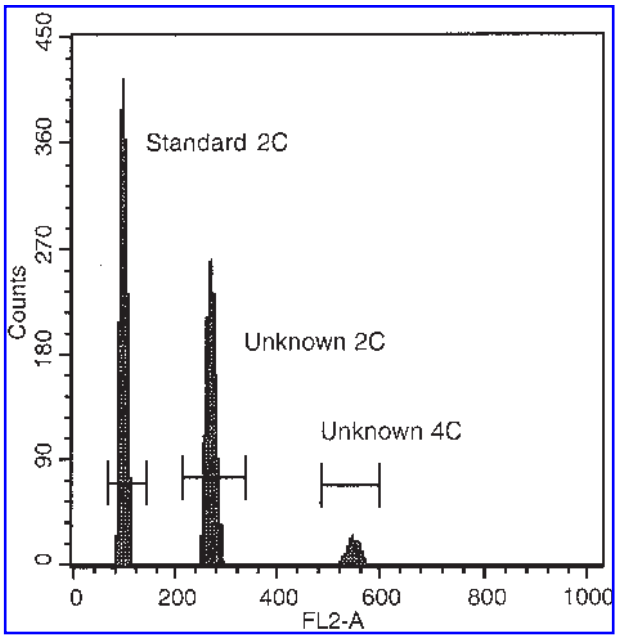

FIG. 4. Sample output of a flow cytometry estimate of genome size in an insect. Hundreds or thousands of nuclei can be assessed in an automated fashion with a flow cytometer, and the mean fluorescence of these is compared against a standard of known genome size prepared, stained, and analyzed along with the unknown sample. 2C indicates nuclei with two copies of the genome, and 4C represents nuclei with four copies.

Advances in technology have contributed to a reduction in both the size and cost of densitometry and flow cytometry equipment,[33] and the analysis of nuclear DNA contents is relatively inexpensive once the requisite infrastructure is in place. If implemented according to best practice guidelines,[31,32] the resolution of both Feulgen image analysis densitometry and flow cytometry is excellent. Indeed, both methods have been rigorously tested for use in medical applications, in which small differences in DNA content can be useful in cancer diagnosis.[31]

BIOREPOSITORIES AS GENOMIC DIVERSITY RESOURCES

Biological resources derived from museum tissue collections, seed banks, culture collections, and other living collections (including aquaria, botanical gardens, and zoos) are crucial for genomic diversity research. The combination of taxonomic expertise, worldwide sampling strategies, and attention to vouchering make them ideal sources of samples for genomic diversity research. Here we outline some approaches to maximize the utility of these source collections for both DNA barcoding and genome quantification.

DNA barcoding

While DNA barcoding is not technically difficult, major challenges do exist. Gaining access to expert-identified reference material is a primary concern. For some groups of organisms there are few if any taxonomists available to identify specimens for constructing the reference sequence library. Although some success has come from the analysis of recently collected (e.g., <20 years old) museum specimens that were either dried or preserved in ethanol, older material and formalin-fixed tissues are poor templates for PCR amplification of the entire 648-bp fragment of the COI barcode region. Frozen tissue collections provide an excellent resource; even specimens that have been compromised by extended thawing or freezer meltdown can still prove useful for genetic characterization.[34]

As currently practiced, barcoding relies heavily on new collections, which is, in turn, strongly dependent upon the availability of funding and the ability to acquire collecting permits. These obstacles can be extremely difficult to overcome for certain taxa or when attempting to work in some nations that impose restrictions on sampling native biodiversity. However, barcoding provides a tool for nations to identify their sovereign genetic resources wherever they occur, and for this reason even countries with restricted access policies are expected to participate in this initiative. In this regard, even small, local repositories can be imoprtant participants in the DNA barcoding initiative.

There are certain difficulties that barcoding must address, both socially and scientifically. Although COI has demonstrated an ability to differentiate closely related species of fungi,[14] there is a cultural inertia in the fungal community around the use of ribosomal genes and intergenic spacer sequences for species identification.[35] It may be difficult to overcome discipline-specific norms concerning marker choice until a sufficient quantity of data exists to demonstrate the validity of switching to a new crossdisciplinary standard as proposed by proponents of COI barcoding. Also, because COI evolves too slowly to discriminate some closely related species of plants, the optimal genic targets for plant barcoding are still being explored and may ultimately involve more than one marker.[36]

Ways in which biological repositories can participate most effectively in the DNA barcoding program include the following: (1) maintaining tissue collections in a manner compatible with DNA extraction such as (at least) ethanol preservation or (preferably) deep freezing; (2) make small subsamples of existing collections available for sequencing; (3) assist in the development of standardized protocols for DNA barcoding, including gene choice and specimen processing; (4) facilitate collaborations with field collectors, local repositories, governmental and nongovernmental agencies, and help to develop access to species that are difficult to acquire or transport. Contributions along any of these lines can greatly assist in the effort to provide rapid, accurate, inexpensive, and universally accessible information about the world's biodiversity.

Genome size

Genome quantification methods, as noted above, are relatively straightforward to implement and can be conducted at high rates under the right circumstances. However, two major impediments to high throughput genome sizing are (1) insufficient access to accurately identified specimens and (2) a lack of sample preservation and preparation protocols for previously unstudied groups. Biorepositories have the potential to assist greatly on both of these fronts.

Collectively, the world's frozen tissue collections, culture collections, and other biological repositories contain samples from thousands of species, the majority of which have not been assayed with respect to genome size, and a substantial fraction of which will be very difficult to collect again for this purpose. As a result, simply making existing materials accessible for genome quantification could have a major impact on the global genome size dataset. More broadly, it should be possible, with only minor effort, to implement sampling protocols that will allow routine genome sizing for newly accessioned material.

For vertebrates, adding genome quantification to the analytical chain could be as simple as preparing a few blood smears at the time of collection for subsequent analysis by FIA. With protists, air-dried samples on microscope slides could serve a similar function. This may be more difficult with plants or invertebrates in which preparations are more involved[37] and in which fresh, fixed, or frozen tissues are required. But here, too, the expertise housed within biorepositories may be extremely useful, as knowledge of tissue preservation techniques could be employed in the development of efficient genome size-friendly sampling methods. Unfortunately, little work has been done to date on the development of specimen preservation protocols that are conducive to genome quantification, a fact that greatly limits the collaborative capacity between those with access to materials and those with genome sizing capabilities. Biorepositories, which are focused on both biodiversity and preserving the utility of samples for future genomic research, could play an important role in bridging this current gap.

GENOMIC DIVERSITY METHODS AND QUALITY CONTROL IN BIOREPOSITORIES

Biological resource collections of various kinds underpin the future of life science and biotechnology research, but the heterogeneous assemblage of collections remains to be coordinated on a global scale.[38] This challenge can be addressed by cataloging specimens in existing collections under a common registry of genetic sequence accessions (i.e., DNA barcodes).[39] Genetic data can also be of use on the scale of individual biological repositories. The taxonomic identity of specimens held in most biological resource collections is assumed to be correct, and although some repositories are actively engaged in taxonomic research, others simply record the species identity as provided by the submitter of the specimen. Even in repositories where taxonomic research is a focus there are typically gaps in the expertise of the staff, which results in varying degrees of certainty in the identifications made across broad taxonomic divisions. Routine protocols like barcoding and genome size quantification enhance the authentication of available resources, while the existence of the repository enhances the value of emerging high-throughput methodologies that are taking hold in the natural sciences.[40]

DNA identification

Biological resource centers have both a need and an obligation to validate the taxonomic identity of their holdings. In particular, they have a need to validate the identity of submitted materials or processed products that are prepared for distribution, such as DNA harvested from cells in culture.[10] DNA barcoding is very appropriate for these quality assurance and quality control applications. Moreover, by making the sequence data from their specimens public via genetic databases, repositories can promote an increased awareness of their holdings and reap the associated benefits that come with increased utilization of their collections.

The application of DNA barcoding as a quality control tool would be relatively simple, requiring only standard molecular equipment or, if that is not feasible, a relatively small budget to cover costs of sequencing by collaborating labs or large-scale barcoding centers. At present, the cost of DNA barcoding a single specimen is on the order of a few dollars, but this is expected to decrease significantly as sequencing expenses continue to plummet. In the future, bench-top DNA barcoding systems should be available, but this is most likely to occur sooner if there is a demonstrated interest in such technology among potential users.

DNA quantification

To the extent that biological repositories are intended to facilitate and enhance large-scale genomic research in the future, the inclusion of genome size data represents an easily acquired "value added" parameter. These data will be influential in shaping the future course of both exploratory and applied genomics, and could allow more efficient decision-making processes regarding which biorepository samples should be targeted for analyses along these lines.

More generally, the application of genome quantification methods to biorepository mate-

rials could provide important insights into the effects of various field and laboratory preservation methods on the structure of entire nuclei. As genomics continues to move from comparisons of individual gene sequences to wholescale comparisons of genomes at large, knowledge regarding procedures that preserve intact nuclei will become increasingly valuable.

Finally, genome quantification approaches could be used as quality control tools in cell culture facilities as well. It is well known that long-term culture of individual cell lines can involve breakages and other chromosome-level changes, potentially including duplications or losses of genomic material. So long as individual genes are the focus of study, quantifying shifts in DNA content may be of only minor concern. However, determining the effects of long-term culture on whole genomes may prove to be important in the future. As noted, flow cytometric analyses are capable of detecting minor changes in DNA quantity, and could be pressed into service in this capacity with minimal difficulty, especially if genome sizing becomes standard practice within biorepositories for the other reasons outlined above.

FUTURE DIRECTIONS

Modern research techniques are placing new and increasing demands on biological repositories.[41] Emerging case studies[42] and "Best Practices"[43] are beginning to address these challenges for human biomaterial banking operations. However, in spite of the fact that the value of museum-based frozen tissue collections has long been recognized,[44] many biodiversity researchers have yet to embrace newer recommendations for specimen collection, preservation, and archival[45] that can extend the value of collected specimens. We advocate a closer collaboration between genomic diversity researchers and biorepositories, in an effort to optimize both sample and data acquisition procedures for ongoing and future research. New approaches toward biodiversity banking are emerging[46] although advanced preservation techniques[47] that support multiple analytical paradigms are critical to their success and more research in this area is needed. We seek to maximize the information available from existing collections and to promote research initiatives that build and maintain biological resource centers with a unified sense of purpose. We believe that DNA barcoding and genome size estimation are two important initiatives that can help to achieve this goal.

The value of making species-level identifications broadly accessible via a fast, accurate, inexpensive, and potentially automated procedure is palpable. However, issues remain with defining what constitutes a "match" to the reference database when a query sequence is not identical to any of the reference sequences. Fortunately, new posterior probability-based methods are emerging to address this issue.[48] The recognition that a comprehensive DNA sequence library is essential for identification with DNA barcodes[49] is ultimately driving this international mega-science initiative toward completion, offering exceptional opportunity for biodiversity repositories in the process.

Likewise, the rapidly expanding field of comparative genomics will undoubtedly require the assistance of biological repositories as genomics becomes increasingly focused on nonmodel organisms. Because genome size is directly relevant to the choice of sequencing targets, and as it represents a key subject of comparative genomics in its own right, enabling genome quantification will represent yet another important role for biorepositories in the future.

Biological repositories have played a critical role in efforts to understand biodiversity for centuries. As we move further into the postgenomic era, their importance remains undiminished. In fact, they stand poised to make a major contribution to the future of both genomic and biodiversity science. Recognizing the mutual benefits to be gleaned by increased interaction between biological repositories and researchers in genomic diversity represents a crucial first step in what is certain to be an exciting journey of discovery.

ACKNOWLEDGMENTS

We would like to thank John G. Baust for inviting the submission of this manuscript and

for providing helpful comments. Rob Dooh and Alex Ardila Garcia assisted with figures. This work was partially supported through funding to the Canadian Barcode of Life Network from Genome Canada (through the Ontario Genomics Institute), NSERC, and other sponsors listed at www.BOLNET.ca.

REFERENCES

1. Hebert PDN, Cywinska A, Ball SL, et al. Biological identifications through DNA barcodes. Proc R Soc Lond B 2003;270:313–321.
2. Folmer O, Black M, Hoeh W, et al. DNA primers for amplification of mitochondrial cytochrome *c* oxidase subunit I from diverse metazoan invertebrates. Mol Marine Biol Biotechnol 1994;5:294–299.
3. Hebert PDN, Ratnasingham R, deWaard JR. Barcoding animal life: Cytochrome *c* oxidase subunit I divergences among closely related species. Proc R Soc Lond B 2003;270(Suppl. 1):96–99.
4. Hebert PDN, Penton EH, Burns JM, et al. Ten species in one: DNA barcoding reveals cryptic species in the neotropical skipper butterfly *Astraptes fulgerator*. Proc Natl Acad Sci USA 2004;101:14812–14817.
5. Smith MA, Wood DM, Janzen DH, et al. DNA barcodes affirm that 16 species of apparently generalist tropical parasitoid flies (Diptera, Tachinidae) are not all generalists. Proc Natl Acad Sci USA 2007;104: 4967–4972.
6. Moritz C, Cicero C. DNA barcoding: Promise and pitfalls. PLoS Biol 2004;2:1529–1531.
7. Hajibabaei M, Janzen DH, Burns JM, et al. DNA barcodes distinguish species of tropical Lepidoptera. Proc Natl Acad Sci USA 2006;103:968–971.
8. Hebert PDN, Stoeckle MY, Zemlak TS, et al. Identification of birds through DNA barcodes. PLoS Biol 2004;2:1657–1663.
9. Ward RD, Zemlak TS, Innes BH, et al. DNA barcoding Australia's fish species. Philos Trans R Soc B 2005;360:1847–1857.
10. Lorenz JG, Jackson WE, Beck JC, et al. The problems and promise of DNA barcodes for species diagnosis of primate biomaterials. Philos Trans R Soc B 2005; 360:1869–1877.
11. Barth D, Krenek S, Fokin SI, et al. Intraspecific genetic variation in *Paramecium* revealed by mitochondrial cytochrome *c* oxidase I sequences. J Eukaryot Microbiol 2006;53:20–25.
12. Lynn DH, Struder-Kypkej MC. Species of *Tetrahymena* identical by small subunit rRNA gene sequences are discriminated by mitochondrial cytochrome *c* oxidase I gene sequences. J Eukaryotic Microbiol 2006;53:385–387.
13. Saunders G. Applying DNA Barcoding to red macroalgae: A preliminary appraisal holds promise for future applications. Philos Trans R Soc B 2005;360: 1879–1888.
14. Seifert KA, Samson RA, deWaard JR, et al. Prospects for fungus identification using CO1 DNA barcodes, with *Penicillium* as a test case. Proc Natl Acad Sci USA 2007;104:3901–3906.
15. Ratnasingham S, Hebert PDN. BOLD: The Barcode of Life Data Systems (*www.barcodinglife.org*). Mol Ecol Notes 2007;7:355–364.
16. Savolainen V, Cowan RS, Vogler AP, et al. Towards writing the encyclopedia of life: An introduction to DNA barcoding. Philos Trans R Soc B 2005;360:1805–1811.
17. Dawnay N, Ogden R, McEwing R, et al. Validation of the barcoding gene COI for use in forensic genetic species identification. Forensic Sci Int (in press)
18. Boivin A, Vendrely R, Vendrely C. L'acide désoxyribonucléique du noyau cellulaire dépositaire des caractères héréditaires; arguments d'ordre analytique. Comp R Acad Sci 1948;226:1061–1063.
19. Mirsky AE, Ris H. The desoxyribonucleic acid content of animal cells and its evolutionary significance. J Gen Physiol 1951;34:451–462.
20. Thomas CA. The genetic organization of chromosomes. Annu Rev Genet 1971;5:237–256.
21. Evans JD, Gundersen-Rindal D. Beenomes to Bombyx: Future directions in applied insect genomics. Genome Biol 2003;4:107.101–107.104.
22. Gregory TR. Genome size evolution in animals. In: Gregory TR, ed. *The Evolution of the Genome*. San Diego: Elsevier; 2005:3–87.
23. Pryer KM, Schneider H, Zimmer EA, et al. Deciding among green plants for whole genome studies. Trends Plant Sci 2002;7:550–554.
24. Jenner RA, Wills MA. The choice of model organisms in evo—devo. Nat Rev Genet 2007;8:311–314.
25. Gregory TR, Nicol JA, Tamm H, et al. Eukaryotic genome size databases. Nucleic Acids Res 2007; 35(Suppl. 1):D332–D338.
26. Bennett MD, Leitch IJ. Genome size evolution in plants. In: Gregory TR, ed. *The Evolution of the Genome.* San Diego: Elsevier: 2005:89–162.
27. Gregory TR. Synergy between sequence and size in large-scale genomics. Nat Rev Genet 2005;6:699–708.
28. Hajibabaei M, deWaard JR, Ivanova NV, et al. Critical factors for assembling a high volume of DNA barcodes. Philos Trans R Soc B 2005;360:1959–1967.
29. Hanner R, Schindel D, Federhen HS, et al. DNA barcodes: Indexing the encyclopedia of life. (In review).
30. Gregory TR, Hebert PDN. Genome size variation in lepidopteran insects. Can J Zool 2003;81:1399–1405.
31. Hardie DC, Gregory TR, Hebert PDN. From pixels to picograms: A beginners' guide to genome quantification by Feulgen image analysis densitometry. J Histochem Cytochem 2002;50:735–749.
32. DeSalle R, Gregory TR, Johnston JS. Preparation of samples for comparative studies of arthropod chromosomes: Visualization, *in situ* hybridization, and

genome size estimation. Methods Enzymol 2005;395: 460–488.
33. Bonetta L. Flow cytometry smaller and better. Nat Methods 2005;2:785–795.
34. Hanner R, Corthals A, Dessauer HC. Salvage of genetically valuable tissues following a freezer failure. Mol Phylogenet Evol 2005;34:452–455.
35. Koljalg U et al. UNITE: A database providing web-based methods for the molecular identification of ectomycorrhizal fungi. New Phytologist 2005;166:1063–1068.
36. Newmaster SG, Fazekas AJ, Ragupathy S. DNA barcoding in land plants: Evaluation of rbcL in a multigene tiered approach. Can J Botany 2006;84:335–341.
37. Gregory TR. The C-value enigma in plants and animals: A review of parallels and an appeal for partnership. Ann Botany 2005;95:133–146.
38. Organisation for Economic Co-operation and Development. *Biological Resource Centers. Underpinning the Future of Life Sciences and Biotechnology.* Paris: OECD Publications; 2001.
39. Walters C, Hanner R. Platforms for DNA banking. In: DeVicente MC, Andersson MS, eds. *DNA Banks—Providing Novel Options for Gene Banks? Topical Reviews in Agricultural Biodiversity.* Rome: International Plant Genetic Resources Institute; 2006:25–35.
40. Distel DL. Molecular biorepositories and biomaterials management: Enhancing the value of high-throughput molecular methodologies for the natural sciences. Marine Ecol Prog Series 2007;332:307–310.
41. Edwards SV, Birks S, Brumfield RT, et al. Future of avian genetic resource collections: Archives of evolutionary and environmental history. The Auk 2005; 122:979–984.
42. Eiseman E, Bloom G, Brower J, et al. *Case Studies of Existing Human Tissue Repositories. "Best Practices" for a Biospecimen Resource for the Genomic and Proteomic Era.* Arlington, VA: Rand Corporation; 2003.
43. International Society for Biological and Environmental Repositories. Best practices for repositories I: Collection, storage, and retrieval of human biological materials for research. Cell Preserv Technol 2005; 3:5–48.
44. Dessauer HC, Hafner MS, eds. *Collections of Frozen Tissues. Value, Management, Field and Laboratory Practices, and Directory of Existing Collections.* Lawrence, KS: Association of Systematics Collections: 1984.
45. Prendini L, Hanner R, DeSalle R. Obtaining, storing and archiving specimens and tissue samples for use in molecular studies. In: DeSalle R, Giribet G, Wheeler W, eds. *Techniques in Molecular Systematics and Evolution.* Basel: Birkhauser Verlag; 2002:176–248.
46. Savolainen V, Powell MP, Davis K, et al., editors. *DNA and Tissue Banking for Biodiversity and Conservation: Theory Practice and Uses.* Richmond, Surrey: Royal Botanic Gardens, Kew; 2006.
47. Molin DGM, Van Dam AJ. A preservation method supporting multipurpose analysis of long-stored samples. Cell Preserv Technol 2006;4:46–50.
48. Abdo Z, Golding GB. A step toward barcoding life: A model-based, decision theoretic method to assign genes to preexisting species groups. System Biol 2007;56:44–56.
49. Ekrem T, Willassen E, Stur E. A comprehensive DNA sequence library is essential for identification with DNA barcodes. Mol Phylogenet Evol 2007;43:530–532.

Address reprint requests to:
Robert H. Hanner
Department of Integrative Biology and
Biodiversity Institute of Ontario
University of Guelph
Guelph, Ontario, N1G 2W1 Canada

E-mail: rhanner@uoguelph.ca