



Molecular biorepositories and biomaterials management: enhancing the value of high-throughput molecular methodologies for the natural sciences

D. L. Distel*

Ocean Genome Legacy, 240 County Road, Ipswich, Massachusetts 01938, USA

ABSTRACT: Genomics, proteomics, metabolomics and a rapidly growing list of 'omic' methodologies have the capacity to transform our view of biology and ecology by tapping the enormous information content of biomolecules. Although procedurally diverse, these approaches share common elements. All use high-throughput technologies to explore a given class of biomolecules in its totality in a given organism, species or community. All generate large quantities of data from individual and often minute samples. All hold promise to provide new insights into the ontogeny, phylogeny, physiology and ecology of organisms and their communities. However, the full potential of these new technologies is unlikely to be realized unless renewed attention is paid to description, preservation, authentication and distribution of biological source materials: issues that have long been of fundamental concern to traditional biologists but that have received less attention in recent years. This renewed emphasis on biomaterial management will require new methods and new types of biological collections (i.e. molecular biorepositories) specifically designed to address and anticipate the needs of these rapidly evolving technologies. Such biorepositories can help advance the nascent 'omics revolution' by allowing researchers to better access and preserve source materials, disseminate research products and data, share ideas, control financial and environmental costs, integrate with traditional methods and knowledge bases, and extract more meaningful data from biological specimens.

KEY WORDS: Genomic conservation · Biobanking · Biological banking · Genome banking · Genome resource repository · Biorepository · Biodiversity · Bioinformatics · Archiving

Resale or republication not permitted without written consent of the publisher

INTRODUCTION

As molecular methods become more important and more widely applied in the natural sciences, the focus of many biologists and ecologists has shifted from tangible physical specimens (tissues and organisms) to molecules and molecular data. Molecular data provide insights unimaginable just a few years ago, and molecular technologies allow biologists to generate and disseminate data with ever-increasing swiftness and ease. Ongoing development of high-throughput technologies for genomics, proteomics, etc. promises to further accelerate the rate at which we are able to obtain information from biological systems. However, with so much excitement surrounding the generation and analysis of molecular data, less attention is being paid

to describing, verifying and archiving the biological materials that serve as the sources of these data.

If unabated, this trend could have serious consequences for the natural sciences. Traditional practices of specimen management were developed for critical practical purposes—to provide high quality, minimally altered biomaterials for research, ensure open access to these materials, preserve these materials for reanalysis and critical reevaluation, and to maintain a strong link between these materials and their environmental context. Historically, these practices have called on researchers to deposit appropriately preserved source materials in public collections (e.g. museums, botanical gardens, zoological parks, etc.) along with detailed information on the method of their collection and the sites from which they were obtained. This is particularly the

*Email: distel@oglf.org

case for those materials used for description of new taxa or previously undescribed properties of known taxa. Such biorepositories, in turn, provide a variety of functions of value to researchers.

Identification, authentication and validation

The validity of ecological and biological studies depends critically on correct identification of taxa and assignment of specimens to them. Taxonomy is a dynamic science and taxonomic designations are periodically modified to accommodate new information. For this reason, it has long been considered prudent to deposit preserved specimens (vouchers) in public collections so that taxonomic assignments can be revisited and reevaluated. Materials deposited in public collections also serve as references and subject matter for future taxonomic investigation.

Access

Natural environments of research interest are often remote and may be both difficult and expensive to access. This is especially true of marine environments. Biological materials sampled from such environments often remain in the 'private' collections of individual researchers, even when collected at public expense. The practice of depositing valuable research materials in public collections can make them available to a much broader scientific audience. This can maximize sample accessibility, minimize collecting costs and sample loss and reduce the potential environmental impacts of repeated research collection on endangered species and sensitive environments.

Policy and regulation

Biorepositories function to develop and enforce policies that ensure fair, appropriate, equitable and safe use of biomaterials. This not only provides legal, fiscal and health protections for depositors and end users, but also provides legitimacy and safeguards that ease the processes of obtaining funding and permission for collecting biomaterials in foreign states, on private or public lands and from endangered or regulated species and habitats.

Communication, cooperation and transparency

Public biorepositories increase the efficacy of research by fostering communication among researchers and by connecting authenticated specimens with field

observations, experimental results and ecological data. Deposited materials also allow reinterpretation, challenge, dispute or corroboration of published observations and conclusions, thereby improving the rigor and transparency of the scientific process.

Preservation and standards

Biological specimens are perishable and the information that they contain is ephemeral. Centralized public biorepositories provide expertise, standardized methods and quality controls, specialized storage facilities and a level of storage security and continuity that cannot easily be duplicated by independent researchers.

Although modern high-throughput technologies promise fundamental changes in the ways in which research is done in marine ecology, and many other disciplines of natural science, they in no way lessen the need for appropriate specimen management. In fact the opposite is true. Such methods add considerable value to biological source materials and considerable cost to their analysis compared with more traditional methods. This creates a responsibility to ensure the proper management and use of biomaterials.

VALUE IN BIOMOLECULES

High-throughput methodologies add value to biological materials in a variety of ways. Chief among these is the capacity to extract more information from less material. Microgram or even nanogram quantities of biomaterials can often yield information from millions of base pairs of DNA or RNA, or tens of thousands of genes, proteins, and metabolites. Moreover, such tiny specimens are often called on to serve as proxies for anything from individual cells and organisms to entire communities. The resultant molecular data are easily and widely disseminated via electronic media, and can be used without modification across a broad range of research disciplines.

Molecular methods also make it possible to generate many new types of functional and informative derivatives of biomolecules, e.g. DNA or cDNA clone libraries, gene or whole genome amplification products, synthetic DNAs, microarrays and a variety of functional genes and gene products. These derivatives can be broadly distributed and used by others as primary research materials.

Additionally, molecular methods lend new value to biological samples by making it possible to preserve not only morphological information, but also to preserve, replicate and propagate biochemical information and biological functions. Thus, appropriately pre-

served biomaterials can now serve as sources of potentially valuable biochemical agents, pathways and biomarkers and as historical records of genetic and metabolic diversity, evolution, function and potential.

Finally, molecular methods add conservation value to preserved biomolecules by allowing these materials to be used to establish historical baselines for assessment of species, population and ecosystem change (Ryder et al. 2000, Ryder 2005). This may be of considerable importance to ecologists given the high rates of physical and biotic change and species and population extinction suspected in many parts of the ocean. In this respect, preservation of biomaterials may be regarded as a form of *ex situ* conservation that can contribute information valuable for the protection of threatened species and ecosystems (Ryder 2005). Beyond this, preserved biomolecules may also contribute to the resurrection of extinct genes and genomes. Indeed, complete functional genomes of extinct viruses have already been reconstructed from archival materials (Lamb & Jackson 2005), and partial genome sequences of extinct cave bears, woolly mammoths and Neanderthal man have been reconstructed from fossil tissue and bone fragments (Noonan et al. 2005, Dalton 2006, Poinar et al. 2006). Clearly, these new technologies can add extraordinary value to preserved biomaterials derived from thriving, threatened and extinct species and ecosystems.

COSTS AND BENEFITS

While genomic, proteomic and other 'omic' methods add value, they also contribute to costs of research. On the one hand, compared with traditional methods, high-throughput molecular analyses (e.g. massively parallel sequencing technologies, robotics, array-based hybridization and mass spectrometry methods) require large capital investments and typically generate high incidental and amortized costs per experiment. On the other hand, as these technologies have evolved, the unit costs of data have fallen precipitously. For example, genome sequencing costs have declined >100-fold since the start of the human genome project in 1990, and there is reasonable expectation that this trend will continue (Lander & Austin 2002). These cost decreases, however, are economies of scale that may not extend to small-scale projects and small institutions.

Given this trend, the per-datum sample analysis cost for high-throughput methods may become comparatively small relative to other research costs. However, no such cost decreases can be expected in sample collection and management. If anything, the costs of collecting biological materials, particularly in the marine realm, can be expected to increase with

increasing energy costs, decreasing abundance of key taxa, increasing regulation of collection in national and international waters, and increasing need to explore more remote and inaccessible environments. Costs associated with identification and description of specimens can also be expected to rise as traditional taxonomic expertise becomes more rare.

This combination of high initial capital costs, high per-experiment costs, high costs of sample collection and limited access to marine environments will likely change the way research is conducted and funded in marine ecology, favoring large research institutions over small. By broadening access to biomaterials, data and ideas, and by providing low cost services such as DNA sequencing or library construction, molecular biorepositories can extend economy of scale to small-scale users.

In summary, high-throughput methods now make it possible to derive, disseminate and utilize more information and value from individual biological specimens than ever before, albeit often at greater total cost. Thus, these new technologies greatly increase value of individual biological specimens and so proportionately increase the potentially harmful consequences of their misidentification, incorrect documentation, mishandling or loss. The obvious conclusion is that new technologies call for increasing emphasis on appropriate specimen management rather than the decreasing trend that has been evident in recent years.

PROBLEMS AND SOLUTIONS

Surprisingly, there are few incentives for modern biologists to practice good specimen management, even as support has grown for analogous improvements in data management practices. For example, many funding agencies enforce strict policies requiring submission of sequence data, trace files, assemblies and quality evaluations to existing public data repositories as a prerequisite for funding of large-scale genomics projects. Similarly, many journals will not publish results of genomics investigations without sequence accession numbers assigned by public databases. However, few organizations enforce similar policies with regard to specimens, source materials and derivatives from genomics projects. As a result, numerous genome-sequencing projects have been completed, at considerable public expense, with no genomic DNAs or voucher specimens on deposit in public collections and little information made available describing collection, identification, locale and ecological or physiological context of source materials. A relatively small investment in proper biomaterial management could not only produce more meaningful data from such costly projects,

but could also provide the opportunity for retrospective analyses if suspicion of taxonomic misidentification or mishandling of source materials arises.

The current shift in priorities to favor data management over sample management likely has multiple causes. It may be due in part to changes in the ways that biologists are trained. New technologies have relaxed barriers that formerly separated scientific disciplines, allowing many researchers to enter new research fields without benefit of exposure to the distinct traditional practices of those fields. This problem is likely exacerbated by shortages of mentors with classical training in such areas as systematics, taxonomy, anatomy and histology. However, a more important contributing factor may be the lack of appropriate biorepositories and methods suitable to meet the developing needs of modern environmental and natural sciences. Few existing biorepositories, museums and natural history collections are equipped to provide inexpensive long-term storage of non-medical biomaterials in ways that preserve informative biomolecules. Methods for preservation, storage, propagation and distribution of biomaterials and informative biomolecules have not been adequately developed, tested and standardized. Broadly accepted policies that regulate ownership and liability for natural biomaterials, their derivatives and the intellectual property that stems from them are lacking. Standards have not been established to ensure responsible collection and trade of research biomaterials from threatened or endangered species and habitats. In short, infrastructure, practices, policies and methods in biomaterials management have failed to anticipate or to keep pace with the rapid changes occurring in biological and ecological research.

Nonetheless, the prospects for developing new standards and infrastructure for molecular biomaterial management are encouraging. Already, a large number of biological specimen repositories have emerged to fill the specialized needs of medical, veterinary, agricultural and toxicological research. Established repositories store and manage blood, fluids, tumors, cell cultures, reproductive products, embryos, cord blood, nucleic acids, proteins, metabolites and bioactive compounds. These biorepositories increase the speed and efficacy of research by providing broad access to limited biological materials, by fostering communication among researchers and by connecting authenticated biospecimens with clinical, experimental and epidemiological observations. Efforts are also afoot at museums, zoological parks, botanical gardens and private non-profit institutions to establish biomolecular resource repositories dedicated to the natural sciences. Examples include the Ambrose-Monell Cryo-collection (American Museum of Natural History, <http://research.amnh.org/amcc/>), the Center for Reproduction of Endangered Species (San Diego Zoo, <http://cres.sandiegozoo.org/index.html>),

the RBG Kew Plant DNA Bank (Royal Botanic Gardens, Kew, <http://rbgkew.org.uk/data/dnaBank/homepage.html>) and the Ocean Genome Resource (Ocean Genome Legacy, <http://oglf.org>).

CONCLUSIONS

Clearly, good biomaterial management practices and appropriate biorepositories can contribute significantly to the successful exploitation of high-throughput technologies in the natural sciences. However, participation by individual researchers and research communities will be needed to develop the infrastructure, methods and scientific consensus that will make this a reality. Marine ecologists, like researchers in many fields of natural science, are only just beginning to add genomic, proteomic, metabolomic and other high-throughput methodologies to their arsenal of investigative techniques. Many are well acquainted with and appreciate the importance of sample management practices traditional to this field, and so are well positioned to recognize, advocate and contribute to the development of modernized biorepositories and biomaterial management methods. This can be done by following, promoting and teaching good specimen management practices, by utilizing existing collections and by urging funding agencies to support biorepositories and biorepository-related research through funding and policy decisions.

Acknowledgements. I thank Debra Pittman, Nate Eckborg and Yvette Luyten for helpful comments on the text. This work was supported by the Ocean Genome Legacy, a non-profit private research institution and genome resource biorepository dedicated to exploring and preserving the biological diversity of marine environments.

LITERATURE CITED

- Dalton R (2006) Neanderthal DNA yields to genome foray. *Nature* 441:260–261
- Lamb RA, Jackson D (2005) Extinct 1918 virus comes alive. *Nature Med* 11:1154–1156
- Lander E, Austin R (2002) Sequencing and resequencing workshop summary: NHGRI large-scale sequencing workshop. National Human Genome Research Institute, National Institutes of Health, Bethesda, MD
- Noonan JP, Hofreiter M, Smith D, Priest JR and 6 others (2005) Genomic sequencing of pleistocene cave bears. *Science* 309:597–599
- Poinar HN, Schwarz C, Qi J, Shapiro B and 9 others (2006) Metagenomics to paleogenomics: large-scale sequencing of mammoth DNA. *Science* 311:392–394
- Ryder OA (2005) Conservation genomics: applying whole genome studies to species conservation efforts. *Cytogenet Genom Res* 108:6–15
- Ryder OA, McLaren A, Brenner S, Zhang YP, Benirschke K (2000) DNA banks for endangered animal species. *Science* 288:275–277