

# DNA Banking for the 21<sup>st</sup> Century

## *A White Paper of Recommendations from the U.S. Workshop on DNA Banking*

The importance of DNA data to modern studies of systematics and evolution cannot be overstated. The use of sequencing or other molecular data has come to play a fundamental and critical role in the vast majority of research in those fields. Further, the nascent field of DNA barcoding offers the hope of greatly reducing the “taxonomic impediment,” in which the limited availability of taxonomic expertise needed to identify specimens hinders research in other fields, inventory projects, and biological resource management. Technological breakthroughs in sequencing are also making it possible to obtain entire organellar and large amounts of nuclear sequence data at low cost, revolutionizing phylogenomics.

The United States has been a world leader in developing these methodologies, and our scientists continue to contribute to the development of data-rich new methods such as megasequencing. Hundreds of high-quality research papers in these subject areas are published by U.S. labs each year. To continue to produce cutting-edge research, researchers and students must have access to a wide range of biological samples. Phylogenetic studies need to include as many species as possible, especially isolated or rare lineages not included in prior studies that may prove to occupy critical phylogenetic positions. The practical utility of barcoding datasets obviously depends crucially on the inclusion of taxa that field workers may encounter. The cost per base pair of sequencing has dropped greatly, while field work has become more costly and sometimes more difficult to conduct. As a result, obtaining necessary samples is increasingly the rate-limiting step in the conduct of systematics and barcoding research.

DNA banking holds the promise of minimizing the sample-collection bottleneck and facilitating a more efficient research workflow. It may be defined as the long-term preservation for future use in DNA-based research of biological samples capable of supplying high-quality DNA, whether aliquots of isolated DNA or tissue samples from which DNA can be isolated. It implies an intention to make samples available as a public service, not only to researchers from the host institution and their collaborators, but to bona fide researchers from other institutions, ideally around the world.

The identified needs of the U.S. systematics and evolutionary biology research community related to DNA and tissue banking fall into three categories. Firstly, many institutions and collectors who could operate or provide samples to DNA banks are not aware of the importance of this practice or do not know how best to organize a bank or collect or store samples for the long term; they need ready access to reliable information and other sources of assistance. Secondly, researchers who might utilize these banks often have no easy way to know what samples are stored at other institutions; improved networking and access to sample databases are needed to facilitate increased use by researchers. Thirdly, if DNA banks are to continue to facilitate the conduct

of highly informative research, their taxonomic coverage must continue to expand, including all major taxonomic groups and, to the greatest extent possible, multiple geographic regions. In the last category, it is particularly recommended that every named and recognized species occurring in the U.S. be made available in DNA banks to facilitate complete barcoding of the indigenous U.S. flora and fauna and all known invasive species, an effort that will require broad support from the biological research community.

## INFRASTRUCTURE, METHODOLOGY, AND INFORMATION

At the present time, many institutions with active field collecting activities do not operate nor contribute to DNA banks, and some institutions that have collections they would be willing to make accessible to the research community do not know how to set up and maintain a well-functioning DNA bank or what the costs of doing so might be. Awareness is the first and most critical step in the process. All institutions with collections that are not being used to their full potential should consider how those collections might be made more useful, and all collectors should think of supplying DNA samples to a repository as being an activity on par with contributing specimens to a museum or herbarium. As more institutions and individuals participate in DNA banking, it is important that their activities are planned to maximize efficiency and quality standards are maintained so that money or samples are not wasted. Participants in the Workshop considered it important for current or potential DNA banks to be aware of the following points:

- It may not be desirable for every institution whose staff actively collect field samples to operate a DNA bank, which requires at least modest infrastructure, staffing, and an ongoing expenditure of resources for maintenance. Instead, collectors should be encouraged to donate samples to one of the large existing repositories that can handle them most efficiently.
- DNA banks that want to receive donations of material from outside collectors should provide guidance on sample collection, set explicit standards for high specimen quality and legal collection practices, and possibly provide free collecting kits and assistance with shipping in some cases. Several large repositories have already made instructions for the collection of tissue samples and voucher specimens available in PDF format online.
- Banked animal tissues should always be stored frozen at  $-80^{\circ}\text{C}$  or below; amplifiable DNA can sometimes be extracted from dry or FAA- or ethanol-preserved specimens of certain taxa, but significant loss of quality can be expected. Liquid nitrogen storage is preferred to freezer storage because of the risk of power failures and the eventual certainty of freezer failure.
- There is less agreement regarding ideal or acceptable storage conditions for plant tissue (e.g.,  $-20^{\circ}\text{C}$ ,  $-80^{\circ}\text{C}$ , or perhaps room temperature, with silica gel removed or included); the preferred conditions may vary among plant families. Fungi have been stored frozen or lyophilized and stored at room temperature.

- All DNA and tissue samples must be properly vouchered. Vouchers should be stored in a publicly accessible institution, and the bank should be able to inform sample recipients of the voucher location. It is ideal for the banking institution to receive duplicate vouchers when samples are donated by outside researchers. Some institutions are willing to store vouchers but not frozen collections, or possibly vice versa, so a division of labor among reliable collaborating institutions may facilitate material supply. In a few circumstances, such as when a species is small in size and very rare, photo-vouchering may be unavoidable; if so, voucher photos should be made accessible to material recipients.
- Relatively large quantities of high-quality DNA are needed for some whole-genome, next-generation sequencing methods; expected sample size and required handling methods should take that into account, and should be routinely revisited to respond to rapid changes in sequencing technology.
- The mass of tissue needed to supply the required amount of DNA can vary among taxa by an order of magnitude (e.g., algae are said to provide poor DNA yield). If experience with a particular group is limited, the necessary sample size should be generously estimated. Some taxa may be particularly likely to degrade when tissue is stored even under ideal conditions; these should be extracted promptly and stored as isolated DNA.
- Rolling circle amplification is sometimes used to replenish samples of isolated DNA that cannot be replaced, but must be used with caution as it may introduce DNA errors. Banks should formulate criteria to determine when samples will be replenished, replaced, or discarded.
- Banks must make efforts to ensure that samples are collected and supplied to them and thence to researchers in accordance with all national and international laws, including CITES and, for samples of international origin, the Convention on Biological Diversity. Persons requesting samples should explain the nature of the planned research; research on international samples with commercial applications cannot be supported unless the samples can be proven to have been collected under permits that allowed such activities.
- Banks should create a standard Material Transfer Agreement, to be signed by parties requesting material, that limits the scope of research and prevents transfer of samples or extracts thereof to third parties without permission. The MTA can also be used to clarify institutional expectations such as the submission of sequences to GenBank and perhaps of alignments and/or trees to Dryad, TreeBASE, or other databases; acknowledgement of the repository and collectors in publications; and payment of fees to help defray costs of maintaining the facility.
- Enforcement of MTAs is acknowledged to be a problem. Many researchers do not follow through on their agreement to credit repositories in publications or supply copies of publications (if any were ultimately produced), making it difficult for the repositories to demonstrate their importance to the conduct of research. Expectations for follow-up contacts from recipients should be increased; recipients

who have not reported any publications within a few years might be contacted and queried about their activities.

## NETWORKING AND AWARENESS

The collection of samples is only half of the necessary equation; researchers must also request and use the samples if they are to be of value. Over 50 DNA and tissue collections have been identified by and listed in iDigBio. Though a few large, relatively well-known repositories receive many requests for samples each year, many researchers, especially students, are unfamiliar with the full range of institutions that may hold relevant samples. Greater awareness and publicity would lead to increased usage. Workshop participants envision an ideal future in which a single Internet portal, familiar to all researchers, could be easily utilized to locate and obtain specimens.

- U.S. DNA banks should be better networked to make searching for samples of desired taxa as efficient as possible. The iDigBio site has begun by compiling a searchable bibliography of collections with links to their websites (<https://www.idigbio.org/genetic-resources>). Institutions with collections that would be made available to outside researchers but that are not yet listed on this page should make themselves known to iDigBio.
- Many repositories have not yet made databases of their holdings available online. They are encouraged to do so to make their collections more accessible, as some will assume that a repository that lists no holdings is unlikely to supply needed material.
- Ultimately, the ability to search all U.S. collections from a single search page would be desirable. Creating this infrastructure would require resolution of issues including data compatibility and security. Both GBIF and iDigBio use “doi”s for samples.

## VISIONS FOR THE FUTURE

More efficient curation and increased utilization of existing collections will provide substantial support for systematics, evolution, and barcoding research. However, because the ability of DNA banks to facilitate research is directly proportional to the diversity of organisms that can be obtained from those banks, expansion of those collections will be required to maximize their value. It is therefore recommended that deliberate efforts to improve taxonomic representation in DNA banks be planned and implemented.

Broad-scale sample collection in some parts of the world has become so challenging, for reasons including legal restrictions on DNA sample export, that it probably will not be possible for U.S. DNA banks to make available all lineages that should be included in future research. Workshop participants emphasized that it will be important to encourage and facilitate DNA banking in biodiverse foreign nations, so that molecular studies can be accomplished in the future through the inclusion of collaborators from those countries. On the

other hand, few regulatory impediments exist to collection of the vast majority of species occurring domestically, and ensuring that every species in the U.S. is available to the research community is seen as an ambitious, but entirely feasible and highly desirable goal.

The Workshop's recommendations for improving U.S. DNA banks to meet the needs of the future include the following:

- A goal should be set of making all species native to or naturalized in the U.S. available in DNA banks. The reliable availability of these taxa would facilitate cutting-edge research on U.S. flora and fauna, which should not be hindered in the future by lack of biological samples, as well as the development of barcoding methods that could be used to track invasives, rare species, and so forth. Attainment of this goal would require long-term financial support for an approach that involved multiple student projects and publishable units of research.
- Taxonomic groups that are most underrepresented in current collections should be targeted for increased representation. While large terrestrial animals are relatively well represented, many small invertebrates (e.g., plankton) are poorly known and poorly represented. Fungi are noted to be a major taxonomic group needing urgent improvement. It is not known what native and invasive fungi exist in the U.S. now, making it impossible to detect changes in the mycoflora; barcoding would enormously enhance our ability to deal with fungi. Publicly accessible fungal collections are very limited and poorly publicized.
- Projects such as the Open Tree of Life should be used to identify important underrepresented lineages and large taxonomic gaps in the worldwide flora and fauna that should be filled if possible. These gaps should be publicly noted in some fashion so that researchers who may work on the taxa involved or have the opportunity to collect them can be made aware of their potential importance.
- The most efficient way to obtain rare lineages, in terms of numbers of samples obtained and per-sample cost, is through broad-scale general collecting of one or more major taxonomic groups from a given area. Rare and threatened habitats should be the focus of community effort, whereby experts in multiple taxonomic groups collaborate in sampling. Barriers to this sort of work include lack of funding for general collection and lack of professional credit for collectors. Mechanisms to support general fieldwork in unique habitats should be developed.
- There should be community pressure for researchers to donate surplus material to DNA banks when projects are completed. Appropriate placement of DNA or tissue samples is included in NSF's guidelines for data management plans but is usually overlooked. Reviewers should be encouraged to criticize the absence of such plans, and NSF should encourage researchers to include money in their budgets to support curation of samples in DNA banks.
- Proper disposal of leftover samples and voucher specimens should be included in ethics training.