



FISH-BOL

Workshop Report

August 26, 2005

**For the workshop held at the
University of Guelph, Ontario, Canada
June 5-8, 2005**

Fish Barcode of Life (FISH-BOL)

June 5-8th, 2005

Conference Sponsorships:



Hosted By:



With Representation From:



Conference Co-Chairs:

Paul Hebert
Biodiversity Institute of Ontario
University of Guelph
Guelph
Canada

Bob Ward
CSIRO Marine and Atmospheric Research
Hobart
Australia



FISH-BOL

Fish Barcode of Life (FISH-BOL)¹

Workshop Report

August 26, 2005

This report addresses the formative stages of an international campaign, fostered by the Consortium for the Barcode of Life and the Census of Marine Life, to establish a comprehensive reference library of DNA barcodes for all fish species. The proposed program of research will enable a fast, accurate, and cost-effective system of molecular identification for all life stages and processed products of the world's ichthyofauna.

Robert Hanner²

University of Guelph

David Schindel

Consortium for the Barcode of Life

Bob Ward³

CSIRO Marine and Atmospheric Research

Paul Hebert³

University of Guelph

1. For additional information visit www.fishbol.org
2. Send comments to RHanner@UoGuelph.CA
3. Workshop Co-Chair

SUMMARY

The societal benefits, scientific rationale and organizational strategy for determining DNA barcodes of all fishes, particularly marine species, were the subject of a “Fish Barcode of Life” (FISH-BOL) workshop held at the University of Guelph, June 5-8, 2005. Major support for this workshop came from the Sloan Foundation, with over 50 participants from 25 nations (section XII).

The goal of this effort is to coordinate the assembly of a standardized reference sequence library for all fish species, one that is derived from voucher specimens with authoritative taxonomic identifications. The benefits of barcoding fishes include facilitating species identification for all potential users, including taxonomists; highlighting specimens that represent a range expansion of known species; flagging previously unrecognized species; and perhaps most importantly, enabling identifications where traditional methods are not applicable.

The workshop included presentations from both the organizers and invited participants. The presentations and ensuing discussions centered on several themes: 1) a background perspective and scientific dialogue surrounding the Barcode of Life initiative and plans to barcode all fish, 2) enabling tools, 3) a review of prior genetic work on fishes relevant to this initiative, 4) enabling organizations, 5) an overview of key taxonomic issues, 6) regional perspectives, 7) organizational issues confronting the FISH-BOL network, 8) administrative structures and 9) funding.

The Fish Barcode of Life effort will create a valuable public resource in the form of an electronic database that contains DNA barcodes, images, and geospatial coordinates of examined specimens. The database will contain linkages to voucher specimens, information on species distributions, nomenclature, authoritative taxonomic information, collateral natural history information and literature citations. FISH-BOL will thus complement and enhance existing information resources, including FishBase and various genomics databases.

Given the estimated \$200 billion USD annual value of fisheries worldwide, FISH-BOL will address socially relevant questions concerning market substitution and quota management of commercial fisheries. For the discipline of ichthyology, FISH-BOL will provide a powerful tool for enhanced understanding of the natural history and ecological interactions of various fish species. The specimens collected and data generated from FISH-BOL will also contribute to an ongoing synthesis concerning the evolutionary history of the most diverse group of vertebrates on Earth. Finally, because the entire edifice of DNA barcoding collapses without accurate taxonomic identifications of reference specimens, the successful execution of FISH-BOL will serve as a powerful demonstration of the immense value of collections, museums and taxonomists to both science and society.

I. BACKGROUND

Historical methods of identifying, naming and classifying fishes are largely based on visible morphology. Modern taxonomic work includes analysis of a host of other traits, including internal anatomy, physiology, behavior, genes and geography; yet morphological traits remain the cornerstone of existing taxonomic treatments. However, there are limitations to relying primarily on morphology when attempting to identify fishes during various stages of their development not considered in original treatments, or when examining fragmentary or processed remains. Even when an intact adult specimen is the subject of identification, the morphological characters and other traits used to discern species are often so subtle and complex that each taxonomist can critically identify only a segment of the global fish fauna.

Multiple taxonomic experts are ordinarily required to identify specimens from even a single biotic survey. Assembling teams of appropriate experts, and/or distributing specimens to them for identification, are both time consuming and expensive tasks. Moreover, accessing existing literature and assessing the validity and priority of various taxon names can be a challenge even for the expert taxonomist. For the non-specialist faced with an assemblage of suboptimal specimens that require species identifications in real time, no method currently exists to bring the sum total of taxonomic knowledge to bear on the problem. This fact is a major impediment to the assessment, conservation and management of global fish biodiversity.

Technological innovation is being harnessed to address this challenge. Large-scale literature digitization projects are enhancing access to existing taxon treatments needed by the global community of taxonomic information consumers. Web-based databases that compile expert-vetted lists of valid taxonomic names and their synonymies, combined with online keys and high-resolution digital images, are further helping to summarize existing knowledge. However, these developments do not address identifications involving larval, juvenile, cryptic or fragmentary specimens.

One of the major benefits of DNA-based identifications is their fast, reliable and accurate characterization across all life stages and species. Early on, the use of DNA sequencing to survey diversity led to the recognition that libraries of reference sequences could be used for species identification in cases of morphological ambiguity, such as with larval stages (Olson et al 1991). DNA, the basic code for all life forms, can be the substance that unifies biological collections of all sorts. In this respect, access to DNA sequences derived from expert-identified voucher specimens can be used to better characterize and broadly identify species. The ensuing catalog of unique genetic sequences or DNA “barcodes” can conceptually unite diverse assemblages of specimens, collections and associated species information under a common registry of sequence accessions. This will enhance online access to information about species and enable a broadly applicable reference database that is essential for performing DNA-based identifications on samples of unknown identity.

II. ENABLING TOOLS

For nearly two decades it has been recognized that rapidly evolving mitochondrial genes, punctuated with highly conserved regions, can be recovered via PCR and that the sequences of these regions allow broad phylogenetic application across the animal kingdom (Kocher et al 1989). However, the generation of a DNA-based identification system places new demands. To be cost effective using existing technology, it is imperative for analysis to focus on an easily recovered and standardized region of the genome that provides good taxonomic resolution within a single sequence read. The availability of broad range primers for the amplification of the 5' region of cytochrome *c* oxidase subunit I (COI) from diverse phyla established this gene sequence as a particularly easily recovered segment of the mitochondrial genome (Folmer et al 1994). Hebert et al (2003a, 2003b) have recently demonstrated that this gene region is highly appropriate for discriminating between closely related species across diverse phyla in the animal kingdom, establishing the 5' end of COI as the "DNA barcode" locus for identifying animals, including fish (Ward et al 2005).

Fish comprise nearly half of all vertebrates, yet they are still a manageable group for demonstrating the utility of DNA barcoding, with approximately 20,000 marine and 15,000 freshwater species (FishBase¹). The real challenge is to establish an organizational infrastructure for the task and to develop clear sampling protocols. From an organizational standpoint, the existing species lists associated with nineteen marine and seven inland FAO statistical areas provide an appropriate starting point for directing regional teams with a goal of sampling five specimens from each species across each area. For certain species exhibiting broad geographic distributions perhaps as many as 25 specimens would be sequenced under this scenario. Given the rudimentary knowledge of existing species distributions combined with the nineteen marine FAO areas, an estimated 500,000 specimens will be needed for comprehensive barcoding of all fish species.

There are many species in existing collections, although specimens that have been fixed in formalin are currently difficult to barcode. Thus many new specimens will need to be collected and archived. The barcode reference database must be populated using voucher specimens identified by experts and backed with archival DNA extractions. Delivery of tissues for DNA extraction is currently the rate-limiting factor for the FISH-BOL program, as high throughput sequencing systems are now in place for utilization by the network. DNA banks must be created to enable the deposition of genetic material for each sequence obtained by the initiative.

Assembling the sequence information into a comprehensive DNA barcode library requires the development of a Laboratory Information Management System (LIMS) capable of providing an audit trail for each barcode generated. This piece of software, which is under development at the University of Guelph, will extend the capabilities of the current Management and Analysis System (MAS), which relates a given barcode record to both a voucher specimen and to a broader set of sequences. The existing Barcode of Life Database (BoLD)² serves this function, which among other options generates Neighbor-Joining

¹ FishBase: www.fishbase.org

² BoLD: www.barcodinglife.org

dendrograms of species' barcodes in PDF format. The system can also diagram specimen collection localities on a distribution map with resolution of 1 km/pixel and further facilitates morphological comparison of voucher specimens when appropriate digital images (e.g. eVouchers, sensu Monk and Baker, 2001) are input.

An ongoing Japanese collaboration of the Fish Mitochondrial Research Group has generated whole mitochondrial genome sequences for an impressive number of fishes. The aim of this group is to develop a complete phylogeny of the fishes. The MitoFish database³ compiles both full and partial mitochondrial sequences of fishes and includes full sequences for 250 species that will soon expand to about 750 species. This data set will be a useful reference for primer development when recalcitrant species are encountered in the FISH-BOL analyses. MitoFish also has numerous links to other relevant biological and genetic databases.

FishBase, a 'global public good' developed as a decision support system for the conservation and management of aquatic biodiversity and ecosystems, will help anchor FISH-BOL in a peer-reviewed taxonomic framework of accepted species names. The need is apparent from the fact that there are over 200,000 scientific names for fishes. FishBase currently recognizes approximately 28,000 valid species names and includes over 80,000 synonyms. FishBase maintains relevant literature citations and includes an identification tool based on morphology, complete with digital images of representative specimens. Clearly, a tight integration of FishBase and FISH-BOL will be critical.

The Integrated Taxonomic Information System (ITIS)⁴ represents another initiative to create an easily accessible database with reliable information on species. This program involves a memorandum of understanding among several US federal agencies and in 2001, ITIS joined forces with the UK-based Species 2000 to develop the Catalog of Life, which now boasts a nearly complete inventory of fishes. Standard reports include classification, geography and links to other databases including publications in BioOne and genomic data in GenBank. This taxonomic source for biodiversity information is currently available in four languages on the web. ITIS will be used as the vetted source for valid species names among fishes as part of the FISH-BOL initiative.

The National Institute for Biotechnology Information (NCBI) maintains GenBank and the NCBI Taxonomy Browser databases, which have offered their support for the aims of FISH-BOL. In fact, GenBank has fostered broad support for barcoding among other members of the genomics collaborative that include the DNA Data Bank of Japan (DDBJ) and the European Molecular Biology Laboratory (EMBL) database. The collaborative has agreed to publicly archive DNA sequences from the FISH-BOL project. They have also expanded the fields for core specimen annotation in their database architecture to more effectively serve barcoding. This is primarily related to information pertaining to the voucher specimen from which sequences are derived. GenBank and the collaborative have agreed to annotate sequences with the keyword "BARCODE" when they meet the appropriate

³ MitoFish: mitofish.ori.u-tokyo.ac.jp

⁴ ITIS: www.itis.usda.gov

guidelines which include: a valid species name, at least 500 bp of double stranded sequence (with fewer than 1% ambiguous base-calls) derived from the 5' end of the COI gene, reference to a structured record for the voucher specimen (see below) from which the sequence was derived, and can also include information on coordinates for the collection locality, collection date, collector, and person who performed the identification. In addition, Barcode entries include reference to the PCR primers used to generate the sequence and can link to the raw data or 'trace files' of the sequences themselves, when, as strongly recommended, the traces are deposited in the NCBI Trace Archive. Genbank also make links from specific sequences to specialty databases containing specimen data, literature, and taxonomic databases, including FishBase and ITIS.

The existence of standard symbolic codes for institutional resource collections in Ichthyology (Leviton et al. 1985⁵), paired with the combined taxonomic treatments of FishBase and ITIS, provide an excellent organizational framework for conducting the FISH-BOL campaign. The above ASIH list continues to be updated by Bill Eschmeyer⁶. GenBank will use this information as a source for developing a structured reference to voucher specimens held in existing reference collections, and will also vet barcode submissions against the taxonomic databases to confirm the validity of names associated with submitted barcode sequences.

III. FISH GENES – FROM BARCODES TO TREE OF LIFE

Two ongoing projects have now demonstrated the utility of DNA barcoding in fishes, one at CSIRO and the other at the University of Guelph. CSIRO presented preliminary data gathered on Australian fishes, which began in October of 2004. In a matter of months, via collaboration with the Guelph team, barcodes were generated for nearly 1,200 specimens, representing some 240 species. Results indicate very close congruence between taxa identified on the basis of morphology and barcode partitions (Ward et al 2005). These data are now publicly available on BoLD. Adding to this pool of preliminary data on fish barcodes, the Barcode of Life Initiative at the University of Guelph has now gathered data on 530 species of marine fishes spanning three oceans (representing 3.5% of the global marine species diversity). This work was accomplished using only two sets of primers. The data reveal that a 2% divergence threshold is a useful proxy for discerning existing species recognized on the basis of morphology. While genera were typically punctuated by deep divergences, there were a few congeneric species that exhibited shared barcode sequences, either due to hybridization, recent speciation or possibly incorrect identification. While COI is highly informative for species identification, it might fail to discriminate young species. Data from MitoFish suggest that the more quickly evolving ND4 and ND5 loci might be useful markers to add to COI when seeking to differentiate species that share COI haplotypes.

Current projects of relevance to FISH-BOL include a US National Science Foundation Partnerships for Enhancing Expertise in Taxonomy (PEET) grant targeting a planetary

⁵ www.asih.org/codons.pdf

⁶ www.calacademy.org/research/ichthyology/catalog/abtabr.html

inventory of the catfish biodiversity and a Tree of Life grant for the Cypriniformes. The catfishes comprise some 2,900 described species and from 600 to 1,600 undescribed species. Primarily inhabiting freshwater, the catfishes are globally distributed and are extremely morphologically diverse, with over thirty recognized families. By partnering with this project, it will be possible to assemble barcodes for a large assemblage of fishes that are highly valued as food across their range. The cypriniformes form the largest clade of freshwater fish species, and this project has amassed voucher specimens for over 1,100 species from North America. A charter has been signed defining the terms of biomaterial and data sharing among collaborators, who plan to share tissues and DNA from over 1,000 species. Sequencing them for COI would make a significant contribution to FISH-BOL.

IV. ENABLING ORGANIZATIONS

Two foundations support the aims of FISH-BOL. The Alfred P. Sloan Foundation was the primary sponsor of the workshop, and recognizes the importance of the campaign to existing research initiatives, particularly the Census of Marine Life (CoML)⁷. Early DNA barcoding results piqued the interest of this Foundation, leading them to support establishment of the international Consortium for the Barcode of Life. Support from the Gordon and Betty Moore Foundation to the University of Guelph has played a key role in enabling substantive barcode pilot projects because of the foundation's interests in supporting environmental and conservation research, especially in the marine realm.

The outlook for continued funding of biodiversity research will hinge critically on the collective ability of the network to provide products of taxonomic research for societal benefit. Quick and accurate identifications by non-specialists, as promised by DNA barcodes, meet this criterion. However, it is imperative to resist the temptation to obsess over curious species and to proceed with well-defined species first, in an effort to both calibrate and showcase the power of DNA barcoding for fishes. In this respect, a unified database as a collaborative product of FISH-BOL is highly desirable if not essential.

A number of other organizations are positioned to aid the FISH-BOL campaign. The Census of Marine Life program is an ambitious ten-year project supported by many governmental and private organizations. It aims to facilitate large-scale marine ecological assessment and has adopted a barcoding protocol. The Ocean Biogeographic Information System (OBIS⁸) developed by CoML contains over five million records, relating to roughly 38,000 species. Ongoing CoML surveys offer opportunities to accelerate barcode characterization of marine fishes and to generate linkages to the OBIS database.

The WorldFish Center⁹ has expressed interest in demonstrating the impact of DNA barcoding. Targeting species from Marine Protected Areas could play an important role in this regard. There is a need for accurate species identifications across many different applied projects. The pragmatic value of the barcoding approach lies in the recognition that

⁷ CoML: www.coml.org

⁸ OBIS: www.iobis.org

⁹ WorldFish Center: www.worldfishcenter.org

barcoding can be easily replicated across different labs collaborating on large international projects such as those involving the WorldFish Center laboratories.

The Consortium for the Barcode of Life (CBOL)¹⁰ is the principal enabling organization for all DNA barcode initiatives. CBOL currently includes over 80 member organizations from 35 countries on six continents. The overarching goals of the Consortium include the creation of a reference library for lookup and identification of unknown specimens in support of tangible, high-priority societal needs. The role of the Consortium is primarily organizational, seeking to promote broad standardization and participation in various barcoding campaigns, including FISH-BOL. A major role of the Consortium is to improve protocols by convening working groups to focus on barcode acquisition, databasing and analysis. Other priorities of the Consortium are to push development of more portable, less expensive technologies for sequencing and to improve the overall taxonomic research environment. CBOL views barcoding as an extraordinary resource for testing taxonomic hypotheses and notes that DNA barcoding is not DNA taxonomy, but rather a tool for taxonomists that will eventually serve non-specialists as well.

V. TAXONOMIC PERSPECTIVES

Museum taxonomists participating in FISH-BOL commented on the diversity of bony, cartilaginous, and other fishes. They also offered a valuable perspective on the role of voucher specimens and tissue collections in underpinning the FISH-BOL campaign. Major points included recognition that:

- Voucher specimens are a fundamental aspect of all taxonomic studies
- Many type specimens no longer exist
- Many of the vouchers that do exist are often in very poor shape and cannot be transported, which is also a consideration for large specimens
- Existing specimens are scattered across many institutions
- Certain taxa are poorly represented in collections
- Completing the inventory of all fish species with reference to historical collections necessitates access to DNA from formalin preserved specimens
- Simply genotyping museum specimens is insufficient; accurate identifications are critical and involve rechecking identifications by the few remaining experts who know them - as revisions are conducted in the face of barcoding
- The costs associated with re-identification and curation are steep
- All collateral information about a species is accessed via its scientific name, hence misidentifications are positively misleading
- Unambiguous linkage between morphological voucher specimens and tissues is essential
- Tissue type and preservation method should be documented

¹⁰ CBOL: www.barcoding.si.edu

In light of these considerations, a protocol for linking types and existing names to OTUs as revealed by barcoding is necessary. Furthermore, the value of photo documentation and digital imagery for newly collected specimens will be substantial. Finally, designated archival repositories for FISH-BOL should be established for ongoing survey work.

VI. REGIONAL PERSPECTIVES

Fisheries experts working in various regions of the world presented information concerning ongoing programs in their region. Collectively, they presented a global view of fish biodiversity, suggesting that the initial representation of researchers in FISH-BOL is of a sufficient critical mass to commence the project's objective of barcoding all fish.

North American fishes have the benefit of being the best-known, with several recent compilations available to guide the FISH-BOL effort. Scripps is compiling a DNA Bank of California fishes. They have also had recent success in developing a technique to obtain short sequences from formalin preserved specimens, which is proving useful for linking barcode OTUs with named specimens held in reference collections.

Barcoding specimens from Central America, particularly the Isthmus of Panama, will provide an excellent opportunity to compare patterns of morphologic and genetic evolution. Some taxa occurring on either side of the isthmus are considered to be conspecific while other similar species are not. There must be some equivalency in making such taxonomic designations and divergence measures derived from molecular data can help resolve existing discrepancies.

Taxonomic work on South American species is ongoing and the University of Concepcion will establish a regional genetic resource center that could help serve DNA barcoding. Funds are available for defining existing stocks and to support training visits. Capacity building is a major priority for this region and interested researchers should consider becoming involved with the initial phase of FISH-BOL for this region.

European waters have seen a regional decline in species diversity, likely the result of commercial exploitation. A number of relevant EU consortia are well positioned to aid FISH-BOL, including the FishTrace¹¹ project and Fish and Chips¹² which is an array-based approach to species identification. The role of barcoding to management efforts involving local EU fisheries could be exemplified through studies of larval dispersal.

The Oceania region provides an opportunity for generating barcodes from over 6,100 species. Planned work in this region includes the BioCode project, with efforts to barcode the flora and fauna of Moorea and its surrounding waters. Such an initiative will demonstrate many exciting applications of barcoding, including projects aimed at increasing our understanding of community assemblies and food webs.

¹¹ FishTrace: infoweb.jrc.it/fishtrace/web/

¹² Fish and Chips: www.fish-and-chips.uni-bremen.de/PostNuke/html/

New Zealand and Antarctica offer special opportunities for barcoding, as this region contains both cryptic and cosmopolitan species. Barcoding fishes of this region will enable efficient detection of catch substitutions, where low value species are substituted for high value species in the market and will also extend to the detection of quota substitutions. These are universal benefits of fish barcoding. Antarctic waters represent about 10% of the world's ocean, yet there are only about 300 species known from this area suggesting these waters might be a lower priority for FISH-BOL given the expense of collecting in them. However, many of these species will likely be collected via a new CoML project focused in Antarctic waters.

Australian waters range from tropical and temperate to subarctic. While the region contains a large water mass, it is of low productivity. About 4,500 species, or 25% of all marine species, occur in Australian waters, including many endemics. CSIRO has already obtained samples from some 550 species, including 450 that are commercially harvested. An additional 200 species of freshwater fishes must also be considered. In Australia, the recently established National Oceans Office has been charged with developing a plan for the ecologically sustainable use of Australia's marine waters, which should drive a local interest in barcoding to aid management decisions.

The fishes of Asian waters have been the target of intensive genetic work, largely led by the Fish Mitochondrial Research Group in Japan. They have as one of their aims the collecting and archiving of about 80% of the fishes in Japan (about 4000 species), with vouchers located primarily in the Natural Science Museum, Tokyo. While their work will focus on characterization of ND4-5, COI sequences could be collected to aid the FISH-BOL campaign.

Russian waters include approximately 568 marine species and about 400 freshwater species. Russia has taxonomic experts for 22 families that have agreed to help with the problem of specimen collection and identification. Russian fish biologists and geneticists are interested in participating in FISH-BOL and have skilled staff for prompt analysis. They estimate that collection costs would average about \$10 USD per specimen, although projected costs to sequence DNA locally are higher than that.

Indian waters contain approximately 1,500 species. India relies heavily on fisheries, harvesting some 6 million metric tons annually from over 400 species. Several institutions in India are poised to support FISH-BOL and identifiable sources of funding exist for this sampling program. The Zoological Survey of India can help with taxonomic identifications and has plans for a national fish museum that could serve as a regional archive for voucher specimens. The possibility also exists to establish an exchange program for training other Asian colleagues.

The fishes of inland waters in Africa and Madagascar are being surveyed by teams from the American Museum of Natural History and the Royal Belgian Academy, while the South African Institute of Marine Biodiversity has played an early role in archiving marine fishes from this region that have been used in barcode pilot studies. The ongoing surveys include sequencing of COI for selected specimens and also includes the collection of digital images for voucher specimens that will soon be publicly available.

VII. ORGANIZATIONAL ISSUES

1. Assembly and Identification of Specimens:

A master list of species is available from FishBase, with distributions broken down by FAO region, country, and habitat type. This is a logical target list for the initiation of FISH-BOL. Species lists in FishBase and ITIS are different; ITIS determinations are more recent; both have ongoing reviews by contributing taxonomists. The taxonomic information resource for FISH-BOL will be an ITIS/FishBase partnership.

A survey of on-line data resources of tissue samples discovered that many institutions do not post their data. The AMCC (Ambrose Monell Cryo Collection at the American Museum of Natural History), Guelph, CSIRO and ROM (Royal Ontario Museum) provided data on their holdings of marine fish species, representing approximately 960 species. When this list was filtered against FishBase up to 30% of names were mismatches. While it is very difficult to determine if tissue samples are associated with morphological vouchers using existing online data resources, these specimens will be useful for generating a first-pass reference sequence.

The FishBase distribution lists can be imported into BoLD and used as templates for campaign management. However, a capability to modify this 'shopping list' will be required. The BoLD workbench is available to support the process of barcode assembly and submissions to GenBank and other databases, as well as assisting in the preparation of data for publication. Some COI data will be gathered as a part of larger projects (e.g. sequencing other regions, Tree of Life) and as such, might be managed elsewhere. However, there would be value in having such projects assemble their barcode data on BoLD to create a unified fish barcode database, as envisaged by the workshop participants and deemed critical by potential funding sources.

Sources of Specimens:

It is desirable to piggy-back the specimen collections required for FISH-BOL on current or planned taxonomic and faunal surveys where possible, to save funds, even if this requires FISH-BOL to deploy designated collectors. New fisheries collections provide an ongoing source of material, with on-board observers as possible sample and data collectors. Possible sources of funding for recruitment and training of such collectors will be needed. Other opportunities for obtaining specimens involve non-taxonomic field studies (esp. ecologists), sport fishing tournaments, commercial fishermen, public donations, and markets (food and ornamentals). Limited quality assurance on associated data may apply to such classes of collections however.

Taxonomic identification:

Error rates in museum collections and catalogs can be significant. Thus, the inclusion of an identification precision index used to standardize relative levels of confidence in identifications is highly desirable. The Australian ranking system provides a useful template:

IDENTIFICATION LEVELS

As of July 1993, specimens in the CSIRO Fish Collection were identified to one of five levels of reliability depending on the taxonomic expertise of the identifier involved and their intentions. A general definition of these levels follows:

Level 1: Highly reliable identification — Specimen identified by (a) an internationally recognised authority of the group, or (b) a specialist that is presently studying or has reviewed the group in the Australian region.

Level 2: Identification made with high degree of confidence at all levels — Specimen identified by a trained identifier who had prior knowledge of the group in the Australian region or used available literature to identify the specimen.

Level 3: Identification made with high confidence to genus but less so to species — Specimen identified by (a) a trained identifier who was confident of its generic placement but did not substantiate their species identification using the literature, or (b) a trained identifier who used the literature but still could not make a positive identification to species, or (c) an untrained identifier who used most of the available literature to make the identification.

Level 4: Identification made with limited confidence — Specimen identified by (a) a trained identifier who was confident of its family placement but unsure of generic or species identifications (no literature used apart from illustrations), or (b) an untrained identifier who had/used limited literature to make the identification.

Level 5: Identification superficial — Specimen identified by (a) a trained identifier who is uncertain of the family placement of the species (cataloguing identification only), (b) an untrained identifier using, at best, figures in a guide, or (c) where the status & expertise of the identifier is unknown.

Implementation of such a ranking system is being adapted for the purpose of establishing confidence in specimen identifications. BARCODE records in GenBank include an 'Identified by:' field and wherever possible, the name of the person associated with a given identification should be captured. To be thorough, the taxon concept used by the identifier should also be recorded if available (i.e. original description, field guide, etc).

The need for traveling specialists to provide/confirm identifications has intrinsic appeal. Travel funds are therefore needed in each region, but this proposal lacks visible productivity and career rewards and thus might be difficult to put into practice. Digital images are generally not adequate for species level identifications and moving large lots of specimens between countries for identification are costly, unwieldy, and dependent on transfer permitting. Thus, appropriations for a traveling specialist system might be the best way to achieve desired results.

Local experts should make initial determinations, enter the source of their taxon concept and cross-check IDs with the barcode, where possible. Taxonomists within the region should check misidentifications/conflicted IDs whenever possible. Requesting help from outside the region, initially using digital images, is the logical next step. Barring that, it is important

to weigh relative costs/benefits of specialist versus specimen travel and judge the timing/expense of specialist travel versus leaving a possible misidentification in BoLD and adjusting the ID confidence level.

2. Vouchering: Protocols and Repositories

Tissue Preservation:

Concerns exist about whether ethanol-preserved collections more than 10-20 years old have suffered DNA degradation. Freezing EtOH preserved tissues might minimize this problem. However, alcohol varies in several ways, (e.g. hydration levels, possible contaminants). Airtight seals on containers are critical for minimizing evaporation and/or hydration of the sample. The volume of EtOH to specimen is also an important consideration with a threefold or higher relative volume of EtOH to tissue desirable. Ultimately, ETOH is flammable and difficult to transport. Other preservatives are DNA-friendly such as RNA Later (Ambion), lysis buffer (Seutin et al 1991), and FTA cards (see below). For a discussion on DMSO uses and contraindications see section XI.

Filter paper that has been treated with detergents to break up cells and nuclease inhibitors and antioxidants to stabilize DNA are suitable for DNA storage (Makowski, 1996; Kline et al., 2002) and purportedly extend its shelf life (see manufacturer's descriptions for FTA paper (Whatman, Florham Park, NJ)). Tissue blots on specialized filter paper should be stored at 30% relative humidity to maximize shelf life of samples (Genvault, Carlsbad, CA). The filter paper platform was designed for storage at ambient temperatures; however, storage at lower temperatures may increase shelf life according to the same principles that govern stability of dry organisms (Walters et al., 2005).

Surplus DNA extracts from barcoding could be archived at participating sequencing facilities using this platform to voucher the sequence run. However, this does not overcome the need for archiving morphological voucher specimens or tissue samples required for further comparative genetic analysis to confirm the authenticity of a suspect DNA extract and/or associated barcode sequence.

For a discussion on tissue collection considerations, refer to Prendini et al (2002) and references therein [pdf available online¹³] and guidelines posted on the AMCC website¹⁴. An overview of preferred protocols for collecting, preserving, and curating specimens, including tissues, can be found on-line as well¹⁵. Frozen tissue will have more uses than just DNA analysis (e.g., toxicology, epidemiology) and represents the 'gold standard' when combined with another redundant preservation technique (alcohol, buffer, FTA card, etc.). When subsampling small specimens for tissue, Scripps recommends eye removal as one possibility.

Vouchering:

Gold standard: formalinized specimen with associated frozen tissue, tissue samples in alcohol or other buffer, and DNA extract. Digital images need scale and color calibration.

¹³ research.amnh.org/amcc/papers.html

¹⁴ research.amnh.org/amcc/help_prot.html

¹⁵ clade.acnatsci.org/rosenberg/archiving/method/methods.html

Standard views for digital imaging – Copeia normal standards; some non-standard views for large specimens (see CoML photographic standards online¹⁶)

Wherever possible, five specimens should be collected and barcoded from any single FAO region, and if it is not possible to retain all these as voucher specimens, specimens should be retained until barcoding is complete. This will facilitate the resolution of any contentious issues, especially regarding specimen identification, should they arise.

3. Sequence Analysis and DNA Repositories

Regions need to assess their capacity for sample processing and DNA archiving. Assistance and lab capacity can be obtained through FISH-BOL and CBOL. Inter-country transfer of genetic material will in most cases require permits and may be problematic, driving the need for national/regional sequencing centers. One way around this is the development of special agreements such as that between Costa Rica and the Smithsonian. This MOU is a model that allows the country of origin to retain control of vouchers while sending PCR products outside for sequencing. The exported sample is completely consumed or destroyed, while specimen data, digital image(s) and sequence data are made public.

4. Data Management

The workshop concluded that the use of BoLD as the workbench for assembly of all FISH-BOL sequences is desirable. When data are ultimately submitted to GenBank, both the submitter and BoLD share the ability to correct data. CBOL also has authority to request removal of the “BARCODE” keyword from records in GenBank if the submitter cannot be located to update an erroneous record.

5. Publication Policy and Data Sharing

Data ownership – It was accepted that there would be important benefits in sharing data across the FISH-BOL community as this would enable rapid assessments of species coverage and appraisals of species congruence across regions. However, it was recognized that the first opportunity to publish data should lie with those who gathered the information. The BoLD analytical platform has the capability to allow analyses of sequence information without actually releasing the sequence data, protecting data from premature release, but allowing its use to guide other projects. Data within BoLD projects are password-protected. FISH-BOL participants are assumed to opt in to a ‘read-only’ global virtual project, to gain advantages of checking IDs. There are four phases of privacy: individual access, visible throughout FISH-BOL, submitted to GenBank but hold until published, released on GenBank. Infractions of privacy policy (intentional or unintentional) will be resolved should they arise. Currently, BoLD permits open-access of its ‘Identify Animal’ code from any submitted sequence, but no stored sequence data are revealed. Contributing projects deserve recognition – as co-authors (like genomics projects) or as acknowledged contributors. Early stage publications at identified milestones and synthetic papers (with advantage of first

¹⁶ clade.acnatsci.org/rosenberg/archiving/taxa/fish.html

access to FISH-BOL members) are envisioned and authors need to negotiate co-authorship (no general policy). Release via GenBank is normally tied to manuscript publication, where data passes through a peer review process.

VIII. ADMINISTRATIVE STRUCTURE

A potential administrative structure for the FISH-BOL campaign was discussed at the workshop. It was decided that the primary work would be led by ten Working Groups that would take responsibility for overseeing collections, identifications and barcoding of the fish faunas in their region.

These regional Working Groups (WGs) included:

- Africa
- Australia
- Europe/Russia
- Indian subcontinent/Central Asia
- Northeast Asia
- Southeast Asia
- Oceania/Antarctic
- South America
- Meso America
- North America

Each WG will include both fresh water and marine partitions, although each WG will have a single Chair. A number of individuals at the workshop expressed a willingness to participate in the formation of the WGs and the FISH-BOL co-organizers will soon seek individuals to act as an interim chair until each region can call a meeting of its members. Chair announcements will be posted to the campaign website¹⁷ as they are established.

The WGs will each assemble a team of researchers, nominate a leader, review the list of species generated for their area using FAO data (with FishBase assistance), keep records of barcoding involvement (collections, vouchers, sequencing), minimize duplication, and seek funding.

The global FISH-BOL campaign will be overseen by a Scientific Committee with 14 members:

- Co-Chairs (Paul Hebert and Bob Ward)
- Campaign Coordinator (Robert Hanner)
- Taxonomic Committee (FishBase & ITIS, TBD)
- 10 WG Chairs (TBD)

¹⁷ www.fishbol.org

Among other duties, the Scientific Committee will synthesize WG reports and generate a summary report on overall progress. It will provide advice to FISH-BOL members, organize the next global meeting, seek funding to support FISH-BOL administration, provide informal linkage/communication with OBIS and CoML and other organizations with a stake in the FISH-BOL campaign.

IX. FUNDING

At present, the FISH-BOL initiative has no direct support. However, an application will be assembled in late 2005 for funding needed for the core business activities of the Network (approximately \$100K per year will enable operation of the Scientific Committee, the Working Groups and the Taxonomic Committee). Some resources are currently available to support activation of the goals of FISH-BOL. For example, the Guelph Barcode of Life Laboratory has indicated its capability to analyze 10K specimens for FISH-BOL participants during 2005-2006. The Food and Drug Administration in the USA has indicated its ability to provide some reagent support. As well, data acquisition efforts are already underway as a consequence of internal support (e.g. CSIRO in Australia). The demonstration of early progress was viewed as the best way to both establish the credibility of the enterprise and to aid access to other funding sources. It was felt that the assembly of barcode records for 3000 fish species by the end of 2006 would be an important and feasible goal. A variety of potential funding sources were discussed including national fisheries (e.g. NMFS) and genomics organizations (e.g. Genome Canada) as well as international programs (e.g. EU FP7) and organizations (World Bank, ICES, PISCES)

X. ACTION ITEMS

The meeting co-chairs were left with 8 action items to pursue. These included:

- Generate an application for funding needed to support FISH-BOL administration
- Recruit individuals to chair the 10 working groups and a taxonomy coordinator
- Establish a website (www.fishbol.org) to provide a forum for project participants and information on the project.
- Create a campaign interface for BoLD that summarizes fish barcode records
- Ascertain best protocols for specimen transfers
- Identify repositories in each region willing to accept voucher specimens
- Expand FISH-BOL participation to new nations
- Identify a campaign co-ordinator¹⁸

¹⁸ Robert Hanner

XI. DMSO USES AND CONTRAINDICATIONS

The effect of DMSO on genetic material has been a perennial question. The physiologic and pharmacologic properties and effects of DMSO are incompletely understood (Brayton 1986). DMSO is a hydrogen-bond disrupter, cell-differentiating agent, hydroxyl radical scavenger, intercellular electrical uncoupler, intracellular low-density lipoprotein-derived cholesterol mobilizing agent, cryoprotectant, and solubilizing agent used in many sample preparations (Santos et al 2003).

DMSO is a powerful surfactant and scavenges free radicals. DNA primary and secondary structure should be neutral to DMSO, because DNA is held together by very strong covalent bonds plus lots of ionic interactions (not hydrophobic or weak hydrophilic interactions where DMSO interacts). DMSO might even preserve DNA primary and secondary structure since DNA is very susceptible to oxidative damage. Yet, DMSO may disrupt tertiary structure a little, because those interactions are weakly hydrophilic.

In combination with glycerol, DMSO cryoprotectant cocktails exhibit properties that allow cells to undergo osmotic stress without major volume changes. These cryoprotectants alter the freezing properties of water thereby preventing freeze/thaw cycles known to be detrimental to most biological materials. The use of DMSO as a solvent for fixatives is thought to enhance preservation of cellular ultrastructure, however, Malinin & Malinin (2004) have shown that DMSO alters the ultrastructural integrity of glutaraldehyde fixed cells.

Most cryoprotectant cocktails are applied at 0°C (on ice), because of the toxicity observed at room temperature. The cryoprotectant cocktail is lethal within 10 min at 25°C. At 0°C, tissues are rarely exposed for more than an hour in cryoprotectants before being frozen. Upon recovery, the tissues are washed immediately to get the cryoprotectants out. Prolonged exposure to DMSO is not recommended (though wearing gloves when handling DMSO is). It is believed that glycerol is the toxic agent (not DMSO), but that DMSO increases the toxicity by allowing other solutes to penetrate cells. Researchers are looking for alternative, less toxic combinations as DMSO increases the toxicity of everything because it increases the rate at which chemicals are absorbed (Walters, pers com).

Thus, DMSO is added to cryoprotective cocktails at the time of freezing to help maintain cellular viability and due to the toxicity of such cocktails at room temperature (to both specimens and researchers), they should probably not be used in a generalized collection buffer for specimens collected primarily for DNA analysis and maintained (even for a short while) at ambient temperatures. However, when collecting rare and valuable tissues in combination with a liquid nitrogen vapor shipper, cryoprotective cocktails might enable recovery of viable cells back at the lab, and while desirable this represents a specialized research program in its own right.

XII. WORKSHOP PARTICIPANTS

	Name	Affiliation	Email address
1	Ablan Menchie	WorldFish Center, Malaysia	m.ablan@cgiar.org
2	Ausubel Jesse	Alfred P. Sloan Foundation, USA	ausubel@mail.rockefeller.edu
3	Bailly Nicolas	WorldFish Center, Philippines	n.bailly@cgiar.org
4	Bermingham Biff	STRI, Panama	eb@naos.si.edu
5	Blohm Dietmar	University of Bremen, Germany	dhb@biotec.uni-bremen.de
6	Boutillier Jim	Pacific Biological Station, DFO, Canada	BoutillierJ@pac.dfo-mpo.gc.ca
7	Burks Christian	Ontario Genomics Institute, Canada	cburks@ontariogenomics.ca
8	Carpenter Kent	Old Dominion University, USA	kcarpent@odu.edu
9	Carvalho Gary	University of Bangor, UK	g.r.carvalho@bangor.ac.uk
10	Chow Seinen	National Research Institute Far Seas Fisheries, Japan	chow@affrc.go.jp
11	Collette Bruce	National Marine Fisheries Services, NMNH, USA	collett@si.edu
12	Connell Allan	CSIR, South Africa	allan@artisteoils.co.za
13	Cooper Andrew	DFO, Canada	CooperA@DFO-MPO.GC.CA
14	Corthals Angélique	American Museum of Natural History, USA	corthals@amnh.org
15	Costa Filipe	Institute of Marine Research, Portugal	fic@fct.unl.pt
16	Crête-Lafrenière Alexis	Université Laval, Canada	alexis.crete-lafreniere@giroq.ulaval.ca
17	Diaz Nelson	University of Chile, Chile	ndiaz@uchile.cl
18	Diaz de Astarloa Juan	Universidad Nacional de Mar del Plata, Argentina	astarloa@mdp.edu.ar
19	Federhen Scott	NCBI, USA	federhen@ncbi.nlm.nih.gov
20	Galleguillos Ricardo	University of Concepcion, Chile	rgalleg@udec.cl
21	Hanner Robert	University of Guelph, Canada	rhanner@uoguelph.ca
22	Hastings Phil	Scripps Institution, USA	phastings@ucsd.edu
23	Hebert Paul	University of Guelph, Canada	phebert@uoguelph.ca
24	Holmes Bronwyn	CSIRO Marine Research, Australia	Bronwyn.Holmes@csiro.au
25	Kartavtsev Yuri	Institute of Marine Biology, Russia	kartavtsev_yu48@hotmail.com
26	Kim Choong-Gon	Ocean Research & Development Institute, South Korea	kimcg@kordi.re.kr
27	Kullander Fang Fang	FishBase, Swedish Museum of Natural History, Sweden	fang.kullander@nrm.se
28	Lakra Wazir	Fish Genetics and Biotechnology, CIFE, India	lakr@wazir@hotmail.com
29	Last Peter	CSIRO Marine Research, Australia	Peter.Last@csiro.au
30	Lowenstein Jake	Researcher - Area de Conservación Guanacaste, Costa Rica	jhlowenstein@hotmail.com
31	Lundberg John	Academy of Natural Sciences, USA	lundberg@acnatsci.org
32	Mayden Rick	St Louis University, USA	maydenr@slu.edu
33	Miya Masaki	Chiba University, Japan	miya@chiba-muse.or.jp
34	Nishida Mutsumi	Ocean Research Institute, Japan	mnishida@ori.u-tokyo.ac.jp
35	Noren Michael	FishTrace, Swedish Museum of Natural History, Sweden	michael.noren@nrm.se
36	Omura Jim	Gordon and Betty Moore Foundation, USA	Jim.Omura@moore.org
37	Orrell Tom	ITIS, USA	orrellt@si.edu
38	Oyarzun Ciro	Universidad de Concepcion, Chile	coyarzun@udec.cl
39	Planes Serge	University of Perpignan, France	planes@univ-perp.fr
40	Ratnasingham Sujeevan	University of Guelph, Canada	sratnasi@uoguelph.ca
41	Sarder Rafiqul	Bangladesh Agricultural University, Bangladesh	sarderri@royalten.net
42	Schelly Robert	American Museum of Natural History, USA	schelly@amnh.org
43	Schindel David	Consortium for the Barcode of Life	schindeld@si.edu
44	Shao Kwang-Tsao	Academia Sinica, Taiwan	zoskt@gate.sinica.edu.tw
45	Smith Peter	NIWA, New Zealand	p.smith@niwa.co.nz
46	Steinke Dirk	Department of Biology, Univ of Konstanz, Konstanz, DE	Dirk.Steinke@uni-konstanz.de
47	Tringali Mike	Florida Fish and Wildlife Conservation Commission, USA	mike.tringali@myfwc.com
48	Van Guelpen Lou	Atlantic Reference Centre, New Brunswick	arc@mar.dfo-mpo.gc.ca
49	Vasilieva Ekaterina	Zoological Museum of Moscow University, Russia	Vas_katerina@mail.ru
50	Ward Bob	CSIRO Marine Research, Australia	Bob.Ward@csiro.au
51	Weigt Lee	National Museum of Natural History, USA	lweigt@lms.si.edu
52	Winterbottom Richard	Royal Ontario Museum, Canada	rickw@rom.on.ca
53	Yancy Haile	FDA, USA	HYancy@CVM.FDA.GOV
54	Zemlak Tyler	University of Guelph, Canada	tzemlak@uoguelph.ca

XIII. REFERENCES

- Brayton CF. 1986. Dimethyl sulfoxide (DMSO): a review. *Cornell Vet.* 76: 61-90.
- Folmer et al. 1994. DNA primers for amplification of mitochondrial cytochrome c oxidase subunit I from diverse metazoan invertebrates. *Mol. Mar. Biol. Biotech.* 3: 294-299.
- Hebert et al. 2003a. Biological identifications through DNA barcodes. *Proc. R. Soc. Lond. B Biol. Sci.* 270: 313–321.
- Hebert et al. 2003b. Barcoding animal life: Cytochrome c oxidase subunit 1 divergences among closely related species. *Proc. R. Soc. Lond. B Biol. Sci.* 270: S596–S599.
- Kline et al. 2002. Polymerase chain reaction amplification of DNA from aged blood stains: Quantitative evaluation of the “suitability for purpose” of four filter papers as archival media. *Analyt. Chem.* 74: 1863-1869.
- Kocher et al. 1989. Dynamics of mitochondrial DNA evolution in animals: amplification and sequencing with conserved primers. *Proc. Natl. Acad. Sci. USA* 86: 6196-6200.
- Leviton et al. 1985. Standards in ichthyology and herpetology: Part I. Standard symbolic codes for institutional resource collections in herpetology and ichthyology. *Copeia* 1985: 802-832.
- Makowski et al. 1996. The effect of storage on Guthrie cards: implications for deoxyribonucleic acid amplification. *Ann. Clin. Lab. Science* 26: 458-469.
- Malinin GI, Malinin TI. 2004. Effects of dimethylsulfoxide on the ultrastructure of fixed cells. *Biotech. Histochem.* 79: 65-9. (Official publication of the Biological Stain Commission)
- Monk RR, Baker RJ. 2001. eVouchers and the use of digital imagery in natural history collections. *Museology, Museum of Texas Tech University* 10: 1-8.
- Olson et al. 1991. Whose larvae? *Nature* 351: 357-358.
- Prendini et al. 2002. Obtaining, storing and archiving specimens for molecular genetic research. pp176-248. In DeSalle, R. Giribet, G. and Wheeler, W. (eds.) *Techniques in Molecular Systematics and Evolution. Methods and Tools in Biosciences and Medicine*, Birkhauser, Basel.
- Santos et al. 2003. Multidisciplinary utilization of dimethyl sulfoxide: pharmacological, cellular, and molecular aspects. *Biochem Pharmacol.* 65: 1035-41.
- Seutin et al. 1991. Preservation of avian blood and tissue samples for DNA analysis. *Can. J. Zool.* 69: 82–90.
- Walters et al. 2005. Dying while dry: Kinetics and mechanisms of deterioration in desiccated organisms. *J. Comp. Biol.* 45: (in press).
- Ward et al. 2005. Barcoding Australia’s fish species. *Phil. Trans. Royal Soc. Lond.* (in press).