



CONSORTIUM FOR THE BARCODE OF LIFE



INAUGURAL MEEETING,
SMITHSONIAN INSTITUTION, WASHINGTON, DC, MAY 24-25, 2004

MEETING REPORTS AND DOCUMENTS

Draft July 18, 2004

Working Group Reports

Organization	2
Collections 1	3
Collections 2	4
Plants	5
DNA	9
Database	11

Appendices

1. List of Conference Participants	14
2. Conference Agenda	17
3. Draft Memorandum of Cooperation	19
4. Draft Intellectual Property Rights Statement	21
5. Ten Reasons for Barcoding Life	22

BARCODE OF LIFE: A short DNA sequence, from a uniform locality on the genome, used for identifying species.

CONSORTIUM FOR THE BARCODE OF LIFE (CBOL) is an international collaboration of natural history museums, herbaria, biological repositories, and biodiversity inventory sites, together with academic and commercial experts in genomics, electronics, taxonomy, and computer science. The initial organizational support for **CBOL** is provided by a 2.5 year grant from the Sloan Foundation.

The mission of the **Consortium for the Barcode of Life** is to rapidly accelerate compiling of DNA barcodes of known and newly discovered plant and animal species, establish a public database of sequences linked to vouchered specimens and associated biological information, and promote technology development for inexpensive hand-held DNA analysis for species identification.

Rapporteur: Vanessa Pike

CBOL Executive Committee: MODUS OPERANDI

- For the first calendar year the membership will remain as is, namely James Baker, James Edwards, James Hanken, Paul Hebert, Richard Lane & Scott Miller
- Thereafter a regular rotation will be introduced based on a three-year term with two member stepping down per year
- Executive Secretary will act as Chair and fixed member. Administrator will keep minutes and follow up actions
- Ex-officio participation may be called for
- Meet a minimum of once per year will regular phone/email conferences
- When replacements are proposed due consideration will be given to the balance of taxonomic expertise, nature of the institution and the regional representation
- Current members will propose new members and a ballot of names will be drawn, members have the right to comment on the ballot

CBOL Executive Committee: ROLES (in first year)

- Disseminate the final version of the *Memorandum* for signature to May meeting participants and others who have already expressed interest in signing with project datasheet by July 2004
- Agree on location of the scientific symposium and the date & duration by early June 2004 and disseminate the result
- Arrange the recruitment of the two staff positions by 1 August 2004
- Draw up full agenda and required actions for science symposium by end August 2004
- Present Science Plan (based on CoML model) after the scientific symposium in February 2005
- Members to write at least three grant proposals for CoBL projects by January 2005
- Active participation in the scientific symposium (chairing sessions, summing up & reporting) February 2005
- Approve report from the Secretariat on the scientific symposium March 2005
- Approve 1st Annual Report from the Secretariat to the Sloan Foundation April 2005

CBOL Secretariat: ROLES (in first year)

- Recruit a minimum of 25 signatories to the MoU before the scientific symposium
- Generate a tracking database of all the projects contributing to the “10 in 10” objective October 2005
- Identify and build relationship with potential funders of CBOL projects & introduce to Members with projects seeking support August –April 2005
- Assist the hosts in the planning and successful delivery of the scientific symposium Sept-Jan 2005
- Draft report from the Secretariat on the scientific symposium February 2005
- Draft 1st Annual Report from the Secretariat to the Sloan Foundation March 2005

Collections/Specimens Working Group 1 Report CBOL May 25, 2004

Rapporteur: Ole Seberg

The Working Group recommends:

For taxonomic purposes (viz. establishing the link between the taxonomy and the barcodes) it is necessary to voucher all sequences and establish an unambiguous link between the voucher and the sequence.

However, it is not deemed essential to voucher all specimens that are being identified (e.g. in connection with ecological studies etc.) by the barcodes. This is in accordance with established practice for specimens identified by traditional means.

Given the ultimate goals of the CBOL funding of added storage facilities both of vouchers and DNA extractions may become a major economic burden. This is an issue that should be addressed when applying for funding for barcoding projects.

It is essential that a Best Practise is established both for field collection and for storage of DNA extractions/tissues etc. It would be advantageous if this practice is generally available e.g. on the web. It is advisable that a record is kept of the “storage history” of such samples. There is a lot of information in the literature, but no synthesis.

Subject to local legal transfer agreements DNA extraction etc. should be made available to the scientific community.

Pilot project should ideally be linked to institutions where the need infrastructure is in place for storage, databasing etc. It will obviously be beneficial to link pilot projects to ATOL projects.

Solving a number of issues would be of great interest to the use of DNA barcoding (not in prioritised order):

1. Studies of methods to recover damaged DNA (viz. DNA from specimens stored under conditions that are suboptimal for recovering DNA) from museum collections.
2. Studies of levels of intra-/interspecific variation and of the levels of confidence/precision needed in identification
3. Test of implantation of DNA barcoding in taxa without a contemporary taxonomy, e.g. shortcuts to help establish a “working” taxonomy
4. Studies of long-term effects of storage of DNA extractions etc,

A project that turned up during the discussion was establishing of DNA barcodes for commercially important fishes, molluscs, and crustaceans etc. to monitor their global movement of these species in processed form.

Collections Working Group 2 Report

CBOL May 25, 2004

Rapporteur: John La Salle, CSIRO Entomology

Present:

Angelique Corthals, American Museum of Natural History

Po-Feng Lee, National Taiwan University

Shannon Hackett, Field Museum

Phil Hastings, University of California San Diego

Ian Poiner, CSIRO Marine Research

George Roderick, University of California Berkeley

Gary Rosenberg, Academy of Natural Sciences

Pablo Luis Tubaro, Museum Argentino de Ciencias Naturales

Terms of Reference

1. Discuss issues of importance in terms of Natural History Collections and the Consortium for the Barcode of Life.
2. List action items and recommendations in regards to resolving these issues.
3. Discuss potential pilot projects, and give examples of suitable projects.

Issues of importance in terms of Natural History Collections and the CBOL

1) The most important issues were in regards to standards and quality control around data gathering. There was agreement on the following principles;

- all specimens from which Barcodes were derived should be represented by voucher material;
- all specimens from which Barcodes were derived should have complete specimen information in an accessible database;
- digital images were desirable to accompany the specimen information;
- DNA material from which Barcodes were derived should be archived;
- standards to ensure quality control should be available to all participants;
- these standards were to include information on collecting material to facilitate later DNA retrieval, as well as current "best protocols" for retrieving DNA from Collection specimens.

We were concerned that these data standards and best practices were not readily available to participants.

Action Items

The executive committee should request that the Database Committee and the DNA Committee (preferably through discussion with each other and with the Collections Committee) make available on the web

- a list of data standards and fields for the database information to accompany barcode vouchers;
- a list of minimum standards and desired formats for digital images to accompany barcode vouchers;

- information on standards for the extraction of the DNA for the barcode, as well as .pdf files of any relevant articles concerning extraction of DNA from natural history collection specimens.

2) There is a need for a person to track information about the various collections, and promote the spread of information and collaboration between institutions. This person might perform the following tasks:

- maintain a web based list of participating institutions;
- maintain a web based list of institutions willing to receive and care for voucher material and/or archived DNA associated with barcodes;
- maintain a web based list of projects, including the institutions involved and the target organisms.

Action Item

The Executive Council to recruit a person willing to volunteer to perform these activities.

3) There was some concern about the impact on natural history collections of needing to house large numbers of voucher specimens and/or genetic material. The ambitious 10 x 10 goal (10 million specimens in 10 years) would mean that 10 million voucher specimens need to be prepared, curated, databased, digitally imaged, and stored in natural history collections. Archived genetic material associated with the vouchers may or may not be housed in the same institution. This represents a significant amount of time, effort, and cost, and it is unreasonable to expect natural history collections to supply these resources.

Action Item

The Executive Council to publicize and encourage the need for strategic planning in regards to voucher specimens associated with barcoding.

Potential pilot projects

There was considerable discussion around potential pilot projects. Projects could be enabling, taxon based, regional, or functional in nature. The following are examples of potential pilot projects.

DNA extraction from museum specimens (enabling)

This project would look at improving the current methods of DNA extraction from specimens housed in collections. There are currently many problems associated with extractions from museum specimens, including age of specimen, method of curation (e.g. in formalin), and the need for non-destructive sampling.

Utility of barcode methodology (enabling)

This project would look at sampling strategies (e.g. number of specimens needed per taxon), the levels of intra- and inter-specific variation, and the utility of barcodes as a tool for taxonomy and biodiversity studies

Global Fisheries Network (functional)

This is a collaborative project which will build on individual initiatives and develop a global barcode library for discrimination of important marine creatures used in the fisheries industry. Champions: Phil Hastings, Ian Poiner, Gary Rosenberg.

Invasive species (functional)

This project will build a global barcode library to allow rapid and accurately discriminate the most important invasive species.

Champion: George Roderick.

Barcodes as a Rapid Assessment in Conservation (functional)

This project will explore barcodes as a rapid assessment tool to monitor genetic diversity in endangered habitats.

Although no specific project was identified, it was felt that there was the need for a taxon based pilot project as a proof of concept to establish that barcodes are applicable in various aspects of taxonomy.

It was also felt that there was a real need to generate a small pool of money for the Executive Committee to fund some of the most relevant pilot projects.

Plant Working Group Report

CBOL, May 25, 2004

Participants: Jesse Ausebel, Alfred P. Sloan Foundation,
Robyn Cowan, Royal Botanic Gardens, Kew,
Gail Reeves, National Biodiversity Institute, Kirstenbosch Research Centre, South
Africa
B. McGill, Missouri Botanic Garden
John Kress, NMNH, Smithsonian
Ole Seberg, Institute of Biology, Denmark
Jesus Ugalde, INBio, Costa Rica
Michel Veuille, Museum National d'Histoire Naturelle, Paris

In discussions on which DNA marker(s) may be appropriate to use for the CBOL programme it was concluded that it will probably be necessary to use a two gene combination, although ideally a single marker could be used to obtain circa 90% of species identification, particularly in the seed plants. ITS, *rbcl* and *matK* were all mentioned as potential candidates. Kenneth Wurdack and colleagues at the Smithsonian are continuing to search for/evaluate hypervariable regions and should have more information in a few months time. There may also be information available that is not in GenBank e.g. at RBG Kew and participants were asked whether they could consider this within their own institutes. Issues such as hybridisation, species complexes and pseudogenes also need to be dealt with.

We agreed to propose to the secretariat that a Plant Working Group be established, and the importance of encouraging representation and participation from other botanists/institutes/regions was also highlighted. We also agreed that a meeting of the proposed Working Group in five to six months may be desirable in order to review the results of the efforts to find an appropriate marker, and to establish one or more pilot projects. Some suggestions put forward for the pilot project were cycads, orchids, palms and pteridophytes. Pilot projects should be achievable within two years.

We felt that geographical information on the database would greatly assist in species identification, and that coordination and links with projects and databases such as *iPLANTS* would be useful.

The potential value of CBOL to governments and organisations involved in the regulation of international trade in plant groups such as orchids, cycads, and sustainable timber was pointed out, as was the possibility of approaching these organisations for support in the programme.

Two main objectives came out of this meeting:

1. To have pooled all available information on potential DNA regions to use for barcoding purposes and to establish which of these to proceed with, within the next few months.
2. To establish at least one pilot project on a group of plants to be completed within two years.

Plant DNA banking was also discussed, and this working group would like to suggest that a special session on DNA banking approaches and techniques be held at the CBOL conference in

2004/2005. Substantial collections of plant DNAs and/or preserved material are held in long-term storage by RBG Kew, NBI South Africa and Missouri Botanical Garden. Initiatives such as these should be encouraged by all participants in plant DNA barcoding as a resource for present and future endeavour in this field.

Participants:

Allan Baker
Ann Bucklin
Ann Bucklin
Robyn Cowan
Robert Hanner

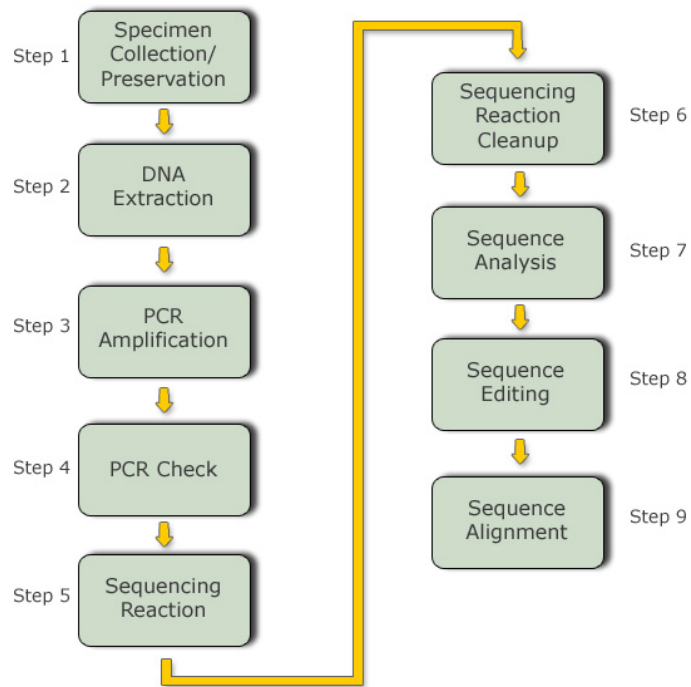
Paul Hebert (rapporteur)
Chris Meyer
Elisabeth Raleigh
Erik Verheyen
Haile Yancy

Gene Selection:

The group began its discussions by considering potential target genes. It was accepted that existing pilot studies have confirmed that a COI-based system will usually generate species-level resolution in most animal groups. However, it was felt that that it was important to demonstrate that COI and *cyt b* generate similar results, especially for the vertebrates where the latter gene has been widely studied. It was also felt that further work was required to identify candidate gene(s) for use in the development of a plant barcoding system.

The Analytical Train:

Nine steps are involved in moving from a specimen to an aligned sequence (Figure 1) and members of this working group felt that CBOL could play an important role in aiding the development of both standardized and optimized operating procedures. It was recognized that optimal protocols will vary, particularly in stages 2 and 3, depending on the preservation history of specimens and their age. For example, the optimal protocols for amplifying DNA from a 100-year insect specimen will vary dramatically from those used to amplify DNA from a recently collected individual of the same species. The task of identifying



optimal protocols is complicated by the wide and growing range of options. Moreover, there are often tradeoffs between the speed/ease of a protocol and its cost. Our discussions revealed that members of the DNA Working Group are actively engaged in evaluating varied protocols and it was felt that CBOL could provide a very useful service to members of the barcode community by assembling this collective experience. It was decided that an initial effort in this regard would be made in time for distribution at the planned Barcode Conference.

Scale of Analysis:

It was felt that the assembly of a barcode library was likely to be advanced in two ways. It is probable that some high volume facilities (>100K sequences per year) will be established that will focus solely on the acquisition of barcode data. However, it was recognized that much useful information would also be generated if the community of researchers engaged in phylogenetic studies research could be persuaded to include barcode sequences as a routine element of their investigations.

Private-Sector Partners:

It was felt that the development of a large-scale program of DNA barcoding will undoubtedly reveal problems that could be resolved with new reagents/kits. It was felt that CBOL could serve a useful role in channeling these problems to private sector partners. As well, it was felt that members of CBOL could aid in the beta testing of newly developed products.

Sequence Quality:

There was considerable discussion directed toward the issue of sequence quality. The suggestion to require bidirectional sequencing for all sequences placed in the Barcode of Life library did not receive support. Similarly, a proposal to exclude manual editing of sequences was not supported. However, there was support for both attaching the sequence traces and Phred scores to all sequence records.

Facility Design:

It was recognized that there is substantial room to accelerate sequence acquisition by automation. Full-blown efforts towards automation will only be feasible in high-volume facilities. However, it was felt that CBOL could play a very useful role in designing optimal equipment configurations (lab-in-as-box) for facilities with varied production goals.

Recognition:

It was felt that it was important to recognize in a very clear way the institutions and individuals involved in aiding assembly of the Barcode of Life library. Hence names and organizational affiliations should be attached to both specimen and sequence pages.

DNA Archives:

There was general agreement that it was desirable to archive DNA extracts so that they would be available for broader sequence characterization. It was also felt that it would be best if these desirable to see the extracts migrated to national or regional facilities that could best ensure their proper curation. Notwithstanding this position, members of the DNA group felt that requiring DNA extract deposition could stand as a serious barrier to the assembly of the barcode library (because of concerns linked to uncontrolled access to DNA sequence information).

Database Working Group Report

CBOL May 24-25, 2004

Participants:

Takashi Abe
Allen Allison
Jim Edwards
Scott Federhen

Robert Hanner
Mike Ruggiero
Larry Speers.

Central questions addressed by the working group included:

- *What should the DNA barcode database strive to do?*

The ultimate goal of this working group is to develop an informatics infrastructure that will enable scientists to transform a collection of individual DNA sequences into a “data rich” environment for species identification and discovery.

- *What should the DNA barcode database include?*

At present, the core information required is the COI sequence and the provenance data associated with the voucher specimen from which the sequence was derived. A more robust data standard (such as the Darwin Core or ABCD schema) remains to be agreed upon. Images of the voucher specimen are deemed highly desirable.

- *What should the DNA barcode database link to?*

Obvious linkages include the National Center for Biotechnology Information (NCBI) federation of databases (i.e. GenBank and PubMed), GBIF, ITIS and TDWG. Other initiatives, such as “Tree of Life” have also convened groups with taxonomic expertise, which are conducting projects that could be conceptually linked to the CBOL database. Linking to images of voucher specimens will require more development, as no national or international facility has yet emerged for image deposition, nor have standards for linkage been accepted. Note however, that GBIF has proposed the creation of a “Global Unique Identifier” (GUID) or “Life Science ID” (LSID) to facilitate this endeavor.

Short-term Goals (1-2 yrs):

1. Enable pilot projects. The working group deemed that a separate database is not immediately required for pilot projects to proceed. However, such projects should make an effort to host an online catalog of the specimens being examined and then make explicit links from their catalog to the sequence data, which should be deposited in GenBank along with appropriate annotation documenting the source of the sequence.
2. Convene another meeting of the database working group as soon as possible. It was suggested by Scott Federhen that this meeting might be hosted by the NCBI (where the possibility exists that some funds for travel could be made available through their “visitor’s program”). GenBank needs to develop a mechanism to flag “Barcode” sequences, which would be a focal topic of this meeting. The NCBI “Ref Seq” database will be examined as a model “curated” sequence database and utility of the GUID or LSID concept should also be addressed.
3. Call for proposals on barcode database architecture at the preliminary CBOL scientific meeting planned for the end of this year.

4. Engage other organizations involved in database development projects, from both the public and private sector.
5. Plan a workshop to develop recommendations for standards and architecture of a formal barcode database (\$10,000 pledged by Jim Edwards of GBIF to help fund workshop).
6. Apply for funding to construct a prototype DNA barcode database.
7. Establish a “repository impact factor” based on the contributions of various collections to the barcode database. This index will serve as a positive incentive for participation in the CBOL initiative and should serve as further justification for funding of biodiversity collections.

Longer Term Goals (5yrs):

1. Develop a curated DNA barcode database, preferably in collaboration with the NCBI and/or other stakeholders (such as the NBII and GBIF) who might be willing to help maintain or otherwise support the database.
2. Track hits and use statistics on the web database in order to determine who the primary classes of users are, in an effort to justify requesting financial support from them.
3. Integrate the Barcode database with other databases hosting catalogs of valid species names, eVouchers (and other “rich media” such as audio files, for example), taxonomic publications, digital gazetteers, etc.
4. Form the backbone of the *Linne*’ system; a proposed virtual instrument for taxonomic and systematic research, which has been the focus of several recent NSF-supported workshops. This infrastructure, combined with the Barcode database and ensuing CBOL technology development initiative, will revolutionize taxonomy and greatly enhance participation in the taxonomic enterprise at levels, resulting in global bioliteracy.

Time Line (provisional):

- Summer 04: Second meeting of the CBOL Database Working Group, hosted by NCBI.
- Fall 04: Call for proposals on Barcode database architecture with the announcement of the CBOL Barcode Conference.
- Winter 04: CBOL Barcode Conference.
- Spring 05: Workshop on Barcode database architecture.
- Summer 05: submit database development proposal to the NSF Biological Databases and Informatics (BDI) panel.

Identify next steps:

- Confirm participation of current working group members and identify possible new members for inclusion in the next meeting.
- Confirm willingness of NCBI to host the next meeting of the working group. Establish an agenda, specific date and timeline for the meeting.

Issues for the next meeting of the working group:

1. Evaluate existing systems such as BoLD, as well as the Monell and IPBIR databases which are functionally linked to GenBank.

2. Identification of the necessary field structure for the database:

Scientific name
Source of scientific name
Determiner (source of identification)
Sequence of bar code
Trace from sequence
Link to voucher
Or
Link to observational database
Preservation type
Protocol for preparation
Attribution of source of sequence
Audit trail

Reversionary capacity
Sampling history

3. Identification of linkage fields needed to link to other databases
4. Single centralized database vs distributed system or hybrid
5. How will the information in the database be curated
 - a. Audit trail
 - b. Discussion threads
6. Questions of how to link to publications – electronic publication or curation of database
7. Questions about legal issues around data use – disclaimers, liabilities
8. Need for quality of identifications – taxonomic impediment
9. How do we supply feedback to (update information in) other databases
10. Are any special search engines or tools needed – possibly locality
11. Identify likely funding sources for database development
12. Identify other potential partners (both public and private) for database development
13. Prioritize development efforts in relation to pilot projects
14. Sieving existing data in GenBank – is it worth the effort?
15. Plan for the upcoming:

“Workshop on standards and architecture for the barcode of life databases”



CONSORTIUM FOR THE BARCODE OF LIFE



INAUGURAL MEETING,
SMITHSONIAN INSTITUTION, WASHINGTON, DC, MAY 24-25, 2004

List of Participants

Takashi Abe

National Institute of Genetics, Japan
Center for Information Biology and DNA
Data Bank of Japan
1111 Yata, Mishima, Shizuoka 411-8540
JAPAN
takaabe@genes.nig.ac.jp

Allen Allison

Bishop Museum
1525 Bernice ST
Honolulu, HI 96817 USA
allison@hawaii.edu

Jesse Ausubel

Alfred P. Sloan Foundation
630 Fifth Avenue, Suite 2550
New York, NY 10111-0242 USA
ausubel@mail.rockefeller.edu

Roger Baird

Canadian Museum of Nature
P.O. Box 3443, Station D
Ottawa, Ontario K1P 6P4 CANADA
rbaird@mus-nature.ca

Allan Baker

Royal Ontario Museum
100 Queen's Park
Toronto, Ontario M5S 2C6 CANADA
allanb@rom.on.ca

James Baker

Academy of Natural Sciences
1900 Benjamin Franklin Parkway
Philadelphia, PA 19103-1195 USA
baker@acnatsci.org

Ann Bucklin

Ocean Process Analysis Laboratory
University of New Hampshire
142 Morse Hall
Durham, NH 03824 USA
ann.bucklin@unh.edu

John Burns

National Museum of Natural History
Smithsonian Institution
PO Box 37012 MRC 127
Washington, DC 20013-7012 USA
burns.john@nmnh.si.edu

Angelique Corthals

American Museum of Natural History
Central Park West at 79th ST
New York, NY 10024 USA
corthals@amnh.org

Robyn Cowan

Royal Botanic Gardens, Kew
Jodrell Laboratory
Richmond, Surrey, TW9 3AB
UNITED KINGDOM
r.cowan@rbgkew.org.uk

James Edwards

Global Biodiversity Information Facility
Universitetsparken 15
DK-2100 Copenhagen Ø DENMARK
jedwards@gbif.org

David Evans

National Museum of Natural History
Smithsonian Institution
Washington, DC 20013-7012 USA
evans@si.edu

Scott Federhen

National Center for Biotechnology Information
National Library of Medicine, NIH
45 Center Drive, MSC 6510
Bethesda, MD 20892-6510 USA
federhen@ncbi.nlm.nih.gov

Bo-Feng Lee

Institute of Ecology and Evolution
National Taiwan University
TAIWAN
firefly@gate.sinica.edu.tw

Robert Friedman

The Center for the Advancement of Genomics
1901 Research Blvd. 6th Floor
Rockville, MD 20850 USA
Robert.Friedman@tcag.org

Matthew Greenstone

USDA/ARS, Insect Biocontrol Laboratory
BARC-West, Bldg. 011A, Rm 214
Beltsville, MD 20705-2350 USA
greenstm@ba.ars.usda.gov

Shannon Hackett

Field Museum
1400 South Lake Shore Drive
Chicago, IL 60605 USA
shackett@fieldmuseum.org

James Hanken

Museum of Comparative Zoology
Harvard University
26 Oxford Street
Cambridge, MA 02138 USA
hanken@oeb.harvard.edu

Robert Hanner

Coriell Institute
403 Haddon Avenue
Camden, NJ 08103 USA
rhanner@coriell.umdnj.edu

Philip Hastings

U.C. San Diego
Scripps Institution of Oceanography - 0208
9500 Gilman Dr.
La Jolla, CA USA
phastings@ucsd.edu

Paul Hebert

Dept of Zoology
University of Guelph
Guelph, Ontario N1G 2W1 CANADA
phebert@uoguelph.ca

Leonard Hirsch

National Museum of Natural History
Smithsonian Institution
Washington, DC 20013 USA
lph@si.edu

Daniel Janzen

Department of Biology
University of Pennsylvania
Philadelphia, PA 19104 USA
djanzen@mail.sas.upenn.edu

John LaSalle

CSIRO Entomology
GPO Box 1700
Canberra, ACT 2601 AUSTRALIA
john.lasalle@csiro.au

Christopher Meyer

Florida Museum of Natural History
373 Dickinson Hall
University of Florida
Gainesville, FL 32611 USA
cmeyer@flmnh.ufl.edu

Scott Miller

National Museum of Natural History
Smithsonian Institution
PO Box 37012 MRC 105
Washington, DC 20013-7012 USA
Miller.scott@nsmnh.si.edu

Vanessa Pike

The Natural History Museum, London
Cromwell Road
London SW7 5BD UNITED KINGDOM
v.pike@nhm.ac.uk

Ian Poiner

CSIRO Marine Research
Hobart TAS 7000 AUSTRALIA
ian.poiner@csiro.au

Elisabeth Raleigh

New England Biolabs
32 Tozer Road
Beverly, MA 01915-5510 USA
raleigh@neb.com

Gail Reeves

National Botanical Institute
Kirstenbosch Research Centre
Capetown, SOUTH AFRICA
reeves@nbi.ac.za

George Roderick

Environmental Sciences
201 Wellman Hall MC 3112
University of California
Berkeley, CA 94720-3112 USA
roderick@nature.berkeley.edu

Yu-Hui Rogers

The Center for the Advancement of Genomics
1901 Research Blvd., 6th Floor
Rockville, MD 20850 USA
Yu-Hui.Rogers@ventersciencejtc.org

Gary Rosenberg

Academy of Natural Sciences
1900 Benjamin Franklin Parkway
Philadelphia, PA 19103-1195 USA
rosenberg@acnatsci.org

Mike Ruggiero

National Biological Information Infrastructure
and Catalogue of Life
National Museum of Natural History
Washington, DC USA
ruggiero.michael@nsmnh.si.edu

Cristian Samper

National Museum of Natural History
Smithsonian Institution
Washington, DC 20013 USA
SamperC@si.edu

Michael Schauff

Systematic Entomology Laboratory, USDA
National Museum of Natural History
Washington, DC. 20560-0168 USA
mschauff@Sel.barc.USDA.gov

Ole Seberg

Institute of Biology, Denmark
Dept. of Evolutionary Biology
Gothersgade 140
Copenhagen K DK-1123 DENMARK
oles@bot.ku.dk

Larry Speers

Global Biodiversity Information Facility
Universitetsparken 15
DK-2100 Copenhagen Ø DENMARK
LSpeers@gbif.org

Mark Stoeckle

Rockefeller University
1270 York Avenue
New York, NY 10021 USA
MarkStoeckle@nyc.rr.com

Hans-Dieter Sues

National Museum of Natural History
Smithsonian Institution
Washington, DC USA
sues.hans@nsmnh.si.edu

Evonne Tang

The National Academies
Board on Life Sciences
Washington, DC USA
ETang@nas.edu

Barbara Thiers

The New York Botanical Garden
200th Street and Kazimiroff Blvd.
Bronx, NY 10458-5126 USA
bthiers@nybg.org

Pablo Luis Tubaro

Museum Argentino de Ciencias Naturales
Bernardino Rivadavia
Av. Angel Gallardo 470
C1405DJR - Buenos Aires ARGENTINA
ptubaro@interlink.com.ar

Jesus Ugalde

INBio
Apdo.Postal 22-3100
Santo Domingo, Heredia COSTA RICA
jugalde@inbio.ac.cr

Miguel Vences

University of Amsterdam, Zoological Museum
Institute for Biodiversity and Ecosystem
Dynamics (IBED)
PO Box 94766
1090 GT Amsterdam NETHERLANDS
vences@science.uva.nl

Erik Verheyen

Royal Belgian Institute of Natural Sciences
Vertebrate Department
Vautierstaat 29
1000 Brussels BELGIUM
erik.verheyen@naturalsciences.be

Michel Veuille

Museum National d'Histoire Naturelle, Paris
16 rue Buffon
Paris 75005 FRANCE
veuille@mnhn.fr

Haile Yancy

U.S. Food and Drug Administration
Center for Veterinary Medicine
8401 Muirkirk Road
Laurel, MD 20708 USA
hyancy@cvm.fda.gov



CONSORTIUM FOR THE BARCODE OF LIFE



INAUGURAL MEETING,
SMITHSONIAN INSTITUTION, WASHINGTON, DC, MAY 24-25, 2004

BARCODE OF LIFE: A short DNA sequence from a uniform part of the genome used for identifying species.

The goal of this meeting is to formally establish the **CONSORTIUM FOR THE BARCODE OF LIFE (CBOL)**, an international collaboration of natural history museums, herbaria, biological repositories, and biodiversity inventory sites, together with academic and commercial experts in genomics, electronics, taxonomy, and computer science. The initial organizational support for **CBOL** is provided by a 2.5 year grant from the Sloan Foundation.

The mission of the **Consortium for the Barcode of Life** is to rapidly accelerate compiling of DNA barcodes of known and newly discovered plant and animal species, establish a public database of sequences linked to vouchered specimens and associated biological information, and promote technology development for inexpensive hand-held DNA analysis for species identification.

The specific aims of this meeting are establishing the organizational structures that will enable the **Consortium for the Barcode of Life** to achieve the scientific, educational, technological, and financial goals of its mission.

Monday, 24 May 2004 Executive Conference Room, National Museum of Natural History*

Co-Chair James Baker, President and CEO, Academy of Natural Sciences, Philadelphia
Co-Chair James Hanken, Director, Museum of Comparative Zoology, Professor of Biology, Harvard University

8:30 AM Coffee

9:00 AM Welcome
David Evans, Undersecretary for Science, Smithsonian Institution
Cristián Samper, Director, National Museum of Natural History

9:15 AM What DNA barcoding can do: a new and powerful diagnostic tool for rapid species identification and discovery (James Hanken)

10:00-10:15 AM Why we need the Consortium: scientific, educational, technological, and financial goals (James Baker)

10:15-10:30 PM Specific Goals and Working Groups—Overview (Mark Stoeckle)

10:30-12:00 PM Specific Goals and Working Groups (30-45 minutes per topic)

1. SPECIMENS: Institutions, Organism groups, Sampling Approaches

2. DNA : Extraction, Archiving, Sequencing

3. TECHNOLOGY DEVELOPMENT

12:00 PM Working Lunch (Provided)

1:00-3:00PM Specific Goals and Working Groups (Continued)

4. DATABASING: Standards (data fields), Analytic Tools, Database Host

5. OUTREACH: Education, Website, Publication

- 3:00 PM Organizational and financial issues: reprise
- 4:00 PM Special Issues (Breakout Groups)
- Plants
 - International outreach and capacity building
 - Intellectual property rights
- 6:00 PM Dinner at Atrium Café, National Museum of Natural History
 Speakers (*Before dinner*) Dan Janzen, University of Pennsylvania
 Speaker (*After dinner*): Jesse Ausubel, Sloan Foundation

*Tuesday, 25 May 2004 Kerby Room**, East Court 340, National Museum of Natural History*

- Co-Chairs Vanessa Pike, The Natural History Museum, London
 Mark Stoeckle, Rockefeller University, New York
- 9:00 AM Consortium Organization and Memorandum of Understanding
 Secretariat office: Staffing, timeline
 Late 2004 scientific symposium: dates and venue
- 10:00 AM CBOL committees and working groups: personnel, timetables
 CBOL Steering committee
 CBOL protocol, product development, database, outreach groups
- 12:00 PM Working lunch (provided)
 CBOL Partnerships/Links: US and International NGOs, US Government, Other
- 1:00 PM Summary session
- 2:00-5:00 PM **CBOL Open Scientific Symposium**
- 2:15 PM **Introduction to the Consortium and the Symposium**
 Scott Miller, Smithsonian Institution
- 2:45 PM **DNA Barcodes and Biodiversity**
 Paul Hebert, University of Guelph
- 3:30 PM **Does DNA Barcoding Work? Empirical Tests from Three Gastropod Groups**
 Christopher Meyer, Florida Museum of Natural History
- 4:00 PM **Species Identification and Discovery in the Pelagic Realm**
 Ann Bucklin, University of New Hampshire
- 4:30 PM **Going for the Green: Prospects for Plant DNA Barcoding**
 Kenneth J. Wurdack, Lee Weigt, Elizabeth A. Zimmer, W. John Kress
 National Museum of Natural History, Smithsonian Institution
- 5:00 PM Adjournment and Reception in the Executive Conference Room, NMNH

*You reach the Executive Conference Room (ECR) via the Constitution Avenue lobby elevators to the First Floor. The ECR is immediately to the left.

**The Kerby Room is in a non-public area in the East Court of NMNH. You must have a badge from the Security Office in the Constitution Avenue lobby. If you need an escort please contact Karie Darrow at 357-2126.

MEMORANDUM OF COOPERATION

to develop a

CONSORTIUM AND SECRETARIAT FOR THE BARCODE OF LIFE

Definition of the barcode of life: A short DNA sequence from standardized portions of the genome, used as an aid in identifying species.

The Consortium for the Barcode of Life (CBOL) is an international collaboration of natural history museums, herbaria, biological repositories, research agencies and biodiversity inventory sites, together with academic and commercial experts in genomics, electronics, taxonomy, and computer science¹.

Mission:

To rapidly accelerate compiling of DNA barcodes of known and newly discovered plant and animal species; establish a public database of sequences linked to vouchered specimens and associated biological information; promote technology development for efficient methods for DNA barcode determinations, including low cost, portable devices for field use; and promote the technical and intellectual development of DNA barcoding, including fostering communication within the academic community and among diverse users throughout society. CBOL intends to work closely and inclusively with the entire global taxonomic community.

The purpose of this Memorandum of Cooperation (MOC) is to establish an international framework of cooperation amongst the members to develop, continuously improve, and maintain a CBOL. The MOC provides a foundation for the members to work together on issues of common interest and upon which the members can jointly plan and carry out mutually beneficial programs, projects, and activities, including the operations of the Secretariat. The members agree that it is in their common interest and to their mutual benefit to work cooperatively through CBOL in a manner consistent with each member's mission and objectives. Coordinated activity allows the members to share expertise, facilities, equipment, and data, and make efficient use of funds and provide consistent information to the public. The members see CBOL as contributing to the goals of the Global Taxonomy Initiative of the Convention on Biological Diversity.

¹ Sloan Foundation has provided a grant (starting 1st May 2004 for 30 months) to establish the CBOL structure and function.

Membership:

The membership is expected to be international. Members of the CBOL will be organizations or programs that have an interest in actively contributing to the advancement of the barcoding of life. Members will support and promote the goals of CBOL and will be able to offer investment of resources. Contributions may be via existing or new projects funded from core institutional sources, existing or new externally funded projects and/or access to specimens or other resources.

Executive Committee:

To promote timely and efficient management, an Executive Committee of the members will oversee the operations of the Secretariat.

Secretariat:

A Secretariat will be the day-to-day operating mechanism for CBOL. Primary functions of the Secretariat will include representation of stakeholders' interests to funding agencies, beneficiaries and other partners, promotion of communication among members, recruitment of members and encouragement of development of "best practices." Subject to funding becoming available, the Secretariat may disburse small grants for targeted projects directly linked to the stated objectives of CBOL.

Any reimbursement or contribution of funds by one signatory of the MOC to another will be handled in accordance with applicable laws, regulations, and procedures.

This MOC may be modified or amended upon the request of any signatory party with the concurrence of all others. Any signatory party may terminate its participation in this MOC with written notice to the Secretariat 60 days prior to such withdrawal.

This MOC will be effective from 1st September 2004 for a period of five years at which point it will be subject to review by the Executive Committee and re-commitment by the signatories.

Date of signature:

Signature:

Printed name:

Position:

On behalf of (name the institution/organisation):

25 May 2004

Consortium for the Barcode of Life

Statement on Intellectual Property Rights

DNA barcodes are short DNA sequences from standardized and well-known portions of the genome, used as an aid in identifying species. The mission of the Consortium for the Barcode of Life is to rapidly accelerate compiling of DNA barcodes of known plant and animal species; establish a public database of sequences linked to vouchered specimens and associated biological information; promote technology development for efficient methods for DNA barcode identifications; and promote the technical and intellectual development of DNA barcoding by fostering communication within the academic community and among diverse users throughout society.

The core information that the Consortium expects participants to make publicly available is the DNA barcode data (a short sequence of a well-known gene) and data about the voucher specimen (if one exists). The Consortium will not centralize other genomic data or voucher specimens.

BARCODING LIFE: TEN REASONS

IDENTIFYING SPECIES BY DNA

BARCODE OF LIFE: A short DNA sequence, from a uniform locality on the genome, used for identifying species.

DNA sequences from a uniform locality on genomes can be a *barcode of life* for identifying species, always the front line in discovery, monitoring and research. Since Linnaeus, biologists have used distinguishing features in taxonomic keys to apply binomial species names, such as *Homo sapiens*. Then, as a master key opens all the rooms in a building, the binomial species name accesses all knowledge about a species.

From insects to birds, evidence now shows that short DNA sequences from a uniform locality on genomes can also be a distinguishing feature. As a Linnaean binomial is an abbreviated label for the morphology of a species, the short sequence is an abbreviated label for the genome of the species.

The *barcode of life* thus provides an additional master key to knowledge about a species. Compiling a public library of sequences linked to named specimens, plus faster and cheaper sequencing, will make this new barcode key increasingly practical and useful.

What additional power does barcoding offer?

1. Works with fragments. Barcoding can identify a species from bits and pieces. When established, barcoding will quickly identify undesirable animal or plant material in processed foodstuffs and detect commercial products derived from regulated species. Barcoding will help reconstruct food cycles by identifying fragments in stomachs and assist plant science by identifying roots sampled from soil layers.



Unwanted animal material in livestock feed can transmit BSE (mad cow disease)

2. Works for all stages of life. Barcoding can identify a species in its many forms, from eggs and seed, through larvae and seedlings, to adults and flowers.



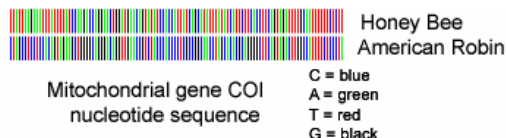
Life stages of Spicebush Swallowtail Butterfly

3. Unmasks look-alikes. Barcoding can distinguish among species that look alike, uncovering dangerous organisms masquerading as harmless ones and enabling a more accurate view of biodiversity.



Anopheles mosquitoes are vectors for human malaria. Only a few of the 430 known species transmit infection.

4. Reduces ambiguity. Written as a sequence of four discrete nucleotides - CATG – along a uniform locality on genomes, a barcode of life provides a digital identifying feature, supplementing the more analog gradations of words, shapes and colors. A library of digital barcodes will provide an unambiguous reference that will facilitate identifying species invading and retreating across the globe and through centuries.



5. Makes expertise go further. The bewildering diversity of about 2 million species already known confines even an expert to morphological identification of only a small part of the plant and animal kingdoms. Foreseeing millions more species to go, scientists can equip themselves with barcoding to speed identification of known organisms and facilitate rapid recognition of new species.



Known species:			
>150,000	>250,000	>300,000	>30,000
flies, mosquitoes	flowering plants	beetles	crabs, lobsters

6. Democratizes access. A standardized library of barcodes will empower many more people to call by name the species around them. It will make possible identification of species whether abundant or rare, native or invasive, engendering appreciation of biodiversity locally and globally.

7. Opens the way for an electronic handheld field guide, the Life Barcoder. Barcoding links biological identification to advancing frontiers in DNA sequencing, miniaturization in electronics, and computerized information storage. Integrating those links will lead to portable desktop devices and ultimately to hand-held barcoders. Imagine the promise of a schoolchild with a barcoder in hand learning to read wild biodiversity, the power granted to a field ecologist surveying with a barcoder and global positioning system, or the security imparted by a port inspector with a barcoder linked to a central computer!



A handheld barcoder, such as the one envisioned here, would have many uses. Promoting technology development of portable devices for field use is a major goal of this initiative.

8. Sprouts new leaves on the tree of life. Since Darwin, biologists seeking a natural system of classification have drawn genealogical trees to represent evolutionary history. Barcoding the similarities and differences among the nearly 2 million species already named will provide a wealth of genetic detail, helping to draw the tree of life on Earth. Barcoding newly discovered species will help show where they belong among known species, sprouting new leaves on the tree of life.

9. Demonstrates value of collections. Compiling the library of barcodes begins with the multimillions of specimens in museums, herbaria, zoos and gardens, and other biological repositories. The spotlight that barcoding shines on these institutions and their collections will strengthen their ongoing efforts to preserve Earth's biodiversity.



The National Museum of Natural History has over 600,000 bird specimens representing more than 80% of the world's known species.

10. Speeds writing the encyclopedia of life. Compiling a library of barcodes linked to vouchered specimens and their binomial names will enhance public access to biological knowledge, helping to create an on-line encyclopedia of life on Earth, with a web page for every species of plant and animal.

CONSORTIUM FOR THE BARCODE OF LIFE (CBOL) is an international collaboration of natural history museums, herbaria, biological repositories, and biodiversity inventory sites, together with academic and commercial experts in genomics, electronics, taxonomy, and computer science. The initial organizational support for CBOL is provided by a 2.5 year grant from the Sloan Foundation.

The mission of CBOL is to rapidly accelerate compiling of DNA barcodes of known and newly discovered plant and animal species, establish a public database of sequences linked to vouchered specimens and associated biological information, and promote technology development for inexpensive hand-held DNA analysis for species identification.

More information is available at: <http://barcoding.si.edu/> and <http://phe.rockefeller.edu/BarcodeConference/index.html>

Mark Stoeckle The Rockefeller University
Paul Waggoner Connecticut Agricultural Experiment Station
Jesse Ausubel Alfred P. Sloan Foundation

v2.0 June 23, 2004