

Report of Steering Committee Meeting Tephritid Barcoding Initiative

**Royal Museum of Central Africa, Tervuren, Belgium
27-28 April 2006**

Summary. A Steering Committee met for 1½ days to develop a proposal for a two-year DNA barcoding initiative devoted to tephritid fruit flies. The Committee reviewed the results of a preliminary survey of museum collections and a pilot study on DNA recovery from pinned archival flies that were conducted prior to the meeting. The Committee agreed to submit a proposal to the Executive Committee of the Consortium for the Barcode of Life (CBOL) by 3 July 2006. The initiative would rely mainly on existing collections in 10-15 key collections, and the Committee plans to contact the curators of these collections to obtain their preliminary agreement to participate in the initiative. The Committee will also conduct a follow-on study of DNA recovery to confirm preliminary results that indicate the likely success of using pinned archival specimens.

Background. At the direction of CBOL's Executive Committee in October 2005, CBOL's Executive Secretary conducted an informal email survey of tephritid specialists and formed a Planning Committee of 29 members from 13 countries. This group interacted through an electronic discussion group for several months and based on those discussions a Steering Committee of nine leading specialists from six countries was selected (see Appendix 1 for member list). CBOL's charge to the Steering Committee was to develop a proposal for a two-year project leading to an operational global system for identifying tephritids using DNA barcodes.

Acknowledgements. The Royal Museum of Central Africa (RMCA) in Tervuren, Belgium, hosted the meeting and Marc De Meyer served as local organizer. Support for the meeting was provided by CBOL, RMCA, and a grant to George Roderick (Univ. California Berkeley) from the National Science Foundation's Research Coordination Network Program). Travel support for some participants was provided by the Agricultural Research Service (ARS) and the Animal and Plant Health Inspection Service (APHIS) of the US Department of Agriculture. Administrative support leading up to the meeting was provided by Katia Dewulf (MRCA) and Meg Fritzsche (CBOL). Prior to the meeting, a pilot study of DNA recovery from archival specimens was conducted by Lee Weigt and Amy Driskell (Laboratory of Analytical Biology, National Museum of Natural History, Smithsonian Institution), and data on tephritid species and museum collections were compiled by Jodie Martin (CBOL).

The Meeting. The 1½ day meeting began with a brief overview of CBOL, the DNA Barcode Initiative, and the goals that CBOL's Executive Committee had in mind (see Appendix 2, meeting agenda). CBOL plans to develop a "Demonstrator System" by July 2008 that will provide an operational test of species identification using DNA barcodes. Mosquitoes and tephritids are the two candidate groups for the Demonstrator System. The Executive Committee has asked this Steering Committee, and its counterpart committee for mosquitoes, to submit proposals by 3 July 2006 in which plans to build their Demonstrator Systems are presented.

Following this brief introduction, meeting participants engaged in a discussion of the

scientific and logistical challenges that would be encountered in constructing a tephritid Demonstrator System. A brief conference call with Amy Driskell was held during the meeting to discuss the results of the pilot study on DNA recovery from archival specimens.

The Committee reached the following decisions.

1. Leadership and Organizational Framework. The Committee selected Bruce McPherson (Penn State) as Chair of the Steering Committee. The initiative would operate through four regional centers, each of which would have a Coordinator. Marc De Meyer agreed to be Coordinator for a center at RMCA that would cover Europe and Africa. Allen Norrbom will be Coordinator for a New World center at the Smithsonian. In the coming weeks, the Committee will seek Coordinators and centers for east Asia and for Australasia. If the third and fourth Coordinators and Centers cannot be identified by 1 June, responsibilities for east Asia and Australasia will be divided between the Centers at RMCA and the Smithsonian.

2. Species List. The Steering Committee agreed that the current list of 4473 species provided by Allen Norrbom and (with collaboration of Amnon Freidberg, Ian White, and others) is the best and most current taxonomic list, and should be used as the framework for this initiative.

3. Taxonomic Scope. The Committee concluded that complete sampling of all tephritid species in a two-year period is not feasible. However, they agreed that a phased hierarchical approach was realistic and could produce an operational Demonstrator System by July 2008. The first phase of the initiative, lasting two years, would produce barcodes for:

- all economically important (EI) species, including both pest species and beneficial species used for biocontrol (approximately 350 species);
- as many other species as possible from the genera containing pest species. We will attempt to sample at least 75% (1200 spp.) of the approximately 1500 species that are congeneric with pests);
- representative additional species from genera containing beneficial species (we will attempt to sample 25% (50 spp.) of the approximately 200 species)
- representative species from other genera in the subtribes (or tribes without subtribal divisions) that contain pest species. We will attempt to sample at least one species from all these genera; approximately 200 species total; and
- at least one species from all subtribes or other higher taxa in which neither pest nor beneficial species occur (perhaps an additional 50 species); and
- at least one species from the other 7 families of Tephritoidea and additional representative species of Ulidiidae, Platystomatidae and Pyrgotidae, the fly families most closely related to tephritids (20-30 spp. total?).

The scope of this first-phase sampling would allow assignment of unknown tephritids to pest or beneficial species with a very high degree of confidence, and would distinguish them from their closest relatives. Non-EI species would be assigned to genera, subtribes and tribes. Future phases of sampling would fill in species from non-EI genera, subtribes and tribes and would improve the taxonomic resolution of the identifications in these groups.

4. Sampling Density. The Committee agreed that a minimum of five (5) specimens would be analyzed per species, covering as much of the known native geographic range as possible. Additional specimens will be obtained if high levels of COI variability are found. Species known from only a few occurrences or represented by fewer than ten (10) specimens would be low priorities for analysis.

5. DNA Recovery from Pinned Archival Specimens. At CBOL's request, the Smithsonian's Laboratory for Analytical Biology (LAB) conducted a pilot study on DNA recovery from pinned specimens provided by Allen Norrbom. This study explored the amount of 16s and COI DNA provided: (a) by four different extraction methods (two Qiagen protocols, Bio101 Ancient DNA kit, Autogen method); (b) with varying length of extraction (24 versus 48 hours); (c) from different body parts (legs, heads, abdomens, whole flies, puparia); and (d) from specimens varying in age (a few days to 106 years). The study also explored the effectiveness of PCR amplification of: two thermal cycling protocols, several reagents, and the typical primers for the Folmer COI region.

Based on this pilot study, the Committee concluded that destructive sampling of legs from specimens up to ten years old will produce good results. Invasive non-destructive sampling of legs should be tested because curators may be more willing to participate if the sampling is non-destructive.

A cursory survey of two important genera indicated that a majority of species are represented by specimens more than ten years old. Fortunately, the pilot study indicated that there are very good prospects for recovering adequate DNA from specimens up to 50 years old or more using abdomens only. Success rates may drop to 50% for specimens of this age but this would not represent a major barrier.

Amplification of 16S was much more successful than for COI, which led participants to conclude that mitochondrial DNA was also extracted but that the wrong primers were being used. Norman Barr and Ho-Yeon Han suggested trying several of the primers they have been using as well as different polymerase suppliers. LAB staff will be asked to send the DNA extracts to Norman Barr for additional testing.

Results of the pilot study suggested that for older specimens, abdomens could provide adequate amounts of DNA. Participants noted that extracting DNA from abdomens prior to normal clearing and dissection for genitalia preparation would not destroy taxonomically useful morphological characters. This sort of invasive but non-destructive sampling could be much less objectionable to curators.

Participants agreed to ask LAB to conduct a second pilot study of DNA recovery from specimens of varying ages relative to: (a) destructive sampling of one leg versus non-destructive sampling of legs and abdomens; and (b) extraction incubation periods of 24, 36, and 48 hours and 4 days. Pre- and post-extraction digital images will be taken to document the degradation in morphological features cause by extraction. Allen Norrbom and Norman Barr will develop the study design and will send it to LAB.

6. Sources of Specimens. The Committee compiled a list of the 20 most important tephritid collections (see Appendix 3). The Committee concluded that a protocol for invasive non-destructive DNA sampling, with illustrations, should be developed and circulated to curators and managers of the most important tephritid collections. The

Committee would ask the curators and managers for their reactions to the proposed invasive but non-destructive protocol (see item 5, above) and for their preliminary agreement to participate in the initiative by providing access to specimens.

The Committee also acknowledged that in a few cases there would be a need to collect new specimens. A small sum should be reserved for this purpose in the project's budget.

7. Taxonomic Identifications. Only specimens identified by expert fruit fly taxonomists should be used for extraction. The Committee felt that for most species, especially the economically important ones, such material is already available. In some cases, identifications will either be difficult or will be called into question by the barcode results and will require re-examination or additional study. The budget should include funds to reimburse these specialists for these efforts. Given the fact that all of these experts are middle-aged or older (e.g., the 3 experts on *Dacini* are retired or retirement eligible), the committee also proposed that each regional center should have a post-doc assigned to the project who would be trained by the experts to identify fruit flies using morphological characters.

8. Steering Specimens to Sequencing Facilities. LAB presented a preliminary proposal to process specimens for approximately \$3.20/specimen, on the assumption that the initiative would provide a technician or post-doc to do the work. At this price, LAB would use an extraction method that produces archival DNA extracts that could be sent to a DNA bank for long-term storage and future use. The Committee found this offer compelling and decided to pursue the option of having all analyses done at LAB.

The Committee considered several different models for obtaining tissue samples of specimens and providing them to the Smithsonian. The preferred method would be to train a post-doc at each regional center to remove, photograph (as necessary), and digest abdomens, complete the dissection process and pin the abdomen in a glycerin vial with the original specimen, add a unique museum identifier number, and record all necessary specimen data on site.

The Committee anticipated that some collections would charge bench fees or a fee to supervise the post-doc's work in the collection, and that funds for these fees should be included in the budget. Participants also anticipated that this arrangement would be unacceptable to some curators and collection managers. In these cases, the options of paying for a local technician to remove abdomens should be offered. In some cases, curators may prefer to send specimens on loan rather than having a visiting post-doc do the work on-site. In these cases, the project should have funds to defray the handling costs of large specimen loans.

9. Permitting Issues related to international material transfer. The Committee recognized that many countries have restrictive policies concerning the loan of scientific specimens. Many institutions also have restrictive policies concerning the use and onward shipment of DNA extracts. In polling curators and collection managers about the proposed sampling protocols, the Committee agreed to ask them about international Material Transfer Agreements (MTAs) and the handling of DNA extracted from specimens. Participants felt that MTAs would not turn out to be a major obstacle but that loan policies might.

10. Data Standards. The Committee agreed that for the purpose of this initiative a “voucher specimen” would be needed to confirm the species identity for every specimen from which a reference barcode is obtained. The voucher specimen would be the pinned insect plus the cleared abdomen, preserved in glycerin and stored in a vial with the pinned specimen. If there is a color pattern on the abdomen, dorsal and lateral images of each abdomen would be taken prior to removal and extraction.

Many COI sequences and DNA extracts have already been obtained from specimens that were sampled destructively and for which vouchers are not available. The Committee agreed that these could not be reference records in the Demonstrator because they do not adhere to the BARCODE data standards established by CBOL and GenBank.

The Committee discussed the possibility of creating a repository for DNA extracts resulting from this sampling initiative. Bruce McPherson wrote a White Paper on the desirability of such a DNA Bank of tephritids several years ago. The Committee agreed that this could be an additional and valuable outcome of the Demonstrator Project, but that the loan policies of the participating collections might have restrictions against the onward donation of DNA aliquots to a repository. The Committee agreed to include this question in its survey of curators and collection managers (see item 6, above).

The Committee discussed the option of including the production of web-based “species pages” among the deliverables of the Demonstrator Project. There are already many web-based information resources on tephritids (e.g., the Systematic Entomology Lab’s Fruit Fly Taxonomy Pages). The Committee decided against including species pages in the Demonstrator Project until there is a community-wide standard format for species pages. If such a standard emerges during the project, the Committee will consider having the participating post-docs take on this task, if it will not interfere with their career development. Until that time, the Demonstrator Project can link many barcode records in BOLD and GenBank to existing web pages. In addition, the participating post-docs can capture digital images of representative specimens that can be added to existing data resources.

11. Shared Software Platform. D. Schindel gave a demonstration of the Barcode of Life Data System (BOLD, University of Guelph). The Committee agreed to use BOLD as the workbench for assembling barcode records and submitting them to GenBank.

12. Data Sharing. The Committee agreed on several guiding principles for the Demonstrator:

- Data are owned by lending institution and are controlled by the project participant who assembled the data record;
- Data owners will provide all other participants in the Demonstrator Project with full access to specimen and sequence records in BOLD;
- All participants in the Demonstrator Project agree not to publish or release specimen or sequence BOLD records prior to release on GenBank that are owned by other participants without their permission;
- Records created by participants in the Demonstrator Project will be released to GenBank as soon as possible, when data record has been assembled to BARCODE data standard; and

- Participants in other projects gathering COI data will be invited to join the Demonstrator Project and to submit records to BOLD, with the understanding that they will adhere to above agreements.

13. Publication of Results. The Committee discussed the possible forms in which results of the Demonstrator Project could be published. This will be particularly important for the career development of the participating post-docs and younger researchers. The following possibilities emerged from the discussion:

- Evaluations of the effectiveness of barcodes as an identification tool;
- Articles in applied journals on cost-effectiveness and reliability of barcodes;
- DNA recovery from archival specimens and other curatorial/laboratory methodologies used;
- Taxonomic revisions of selected species groups;
- Co-authorship based on the use of COI data from the Demonstrator Project in phylogenetic and other studies (to be negotiated with authors on an individual basis);
- Reports of progress and preliminary results in the new barcode section of Molecular Ecology Notes; and
- Online publication of acknowledgements of data and specimen contributors.

14. Budget Estimate. The Committee compiled a very conservative preliminary budget estimate of US\$1.27 million. This estimate included four full-time post-doctoral fellows and a 25% overhead rate, and did not include any participant cost-sharing. Participants agreed that a target budget of \$1 million for two years was realistic.

15. Communication Plan. The Committee agreed to continue to use the NBII Portal for its electronic discussion and document sharing. In addition, an informational website will be developed as soon as possible. Initially this website will be on the Smithsonian server and if it is approved as the Demonstrator Project, it will migrate to an independent site in the “.org” web domain, similar to the FISH-BOL and All Birds campaign websites.

The Committee agreed that two documents must be developed immediately: (1) a brief prospectus describing the project to potential supporters, funders, curators and collection managers, and (2) the protocols for sampling, extraction, DNA analysis, vouchering, and data handling for curators, collection managers, and other potential participants. In order to assess the feasibility of the proposed sampling protocols prior to submitting the proposal, the Committee agreed that the Coordinators should send the prospectus and protocols to curators and managers of the most important collections, with a request for their tentative agreement to grant access to their collections. These documents should also be sent to organizations like the Dutch Plant Protection Agency and the Australia Biosecurity Fund with an invitation to participate in the Demonstrator Project.

16. Funding Strategies. The Committee compiled a list of potential funding sources, consisting primarily of government regulatory agencies, research and private foundations, and associations in the agricultural industry (see Appendix 4). Participants agreed that it would be beneficial to develop a short prospectus that could be used to initiate discussions with these potential funders.

17. Supporters. CBOL’s Executive Committee recently developed the criteria it would use to evaluate the tephritid and mosquito proposals to become the Demonstrator Project.

Among the criteria are statements of support from potential users. The Committee compiled a list of regulatory agencies, biotech companies, and international organizations concerned with invasive and pest species that might provide such statements of support (see Appendix 5). Several of these might be interested in conducting an evaluation of the Demonstrator System at the end when it is operational in 2008, and statements to this effect might strengthen the proposal to the Executive Committee.

18. Timetable. The Committee constructed a detailed timetable leading to submission of the proposal to CBOL's Executive Committee on or before 3 July 2006, and a more general timeline for a two-year project ending by July 2008. The major milestones during the project are:

- September 2006: Outreach presentations at fruit fly workshops in Brazil, Japan
- October 2006: Submit funding proposals; hire a Smithsonian intern and/or technician and first post-doc to begin sequencing on-hand samples from voucher specimens and available DNA
- February 2007: Second International Barcode Conference: Present progress and preliminary results; invasive non-destructive protocols; hold meeting of Steering Committee
- March 2007: hire other post-docs
- July 2007: complete sequencing of on-hand samples (1000 species, 3000 specimens, 75% of pest species)
- October 2007: meeting of Steering Committee to review progress, contribute to Sloan renewal proposal

APPENDIX 1: Tephritid Steering Committee Meeting

Armstrong, Karen, Lincoln University, New Zealand, Armstron@lincoln.ac.nz

Barr, Norman, AHIS/PPQ laboratory, USA, norman.b.barr@aphis.usda.gov

De Meyer, Marc, Museum of Central Africa, Belgium, marc.de.meyer@africamuseum.be

Freidberg, Amnon, Tel Aviv, Israel, afdipter@post.tau.ac.il

Han, Ho-Yeon, Yonsei University, Korea, hyhan@yonsei.ac.kr

McPheron, Bruce, Penn State University, USA, bam10@psu.edu

Norrbom, Allen, USDA-SEL, USA, anorrbom@sel.barc.usda.gov

Roderick, George, UC Berkeley, USA, roderick@berkeley.edu

White, Ian, Natural History Museum, UK, white.dacus@btinternet.com

APPENDIX 2: Meeting Agenda

Tephritid Steering Committee Meeting, Tervuren, Belgium

Thursday, 27 April 2006:

- 12:00:** Lunch, welcome and introductions, Dr. Marc De Meyer
- 1:00:** Welcome, Overview of Museum, Mr. Guido Gryseels, Director, Royal Museum for Central Africa
- 1:20:** Formal selection of Steering Committee Chair
- 1:30:** Overview of the Barcode Initiative, Consortium for the Barcode of Life (CBOL) A Tephritid Demonstrator System, Goals of the meeting: David Schindel
- 2:00:** Views of the CBOL Scientific Advisory Board, Karen Armstrong
- 2:15:** Discussion of meeting goals
- 3:15:** Overview of Scientific and Logistical Issues, Experiences of CBOL Campaigns David Schindel
- 3:30:** Moderated discussion of scientific issues
- 19.** Species List
 - 20.** Taxonomic Scope
 - 21.** Sampling Density
 - 22.** DNA Recovery from Pinned Archival Specimens: Results of pilot project, Allen Norrbom
- 6:00:** Adjourn for Day

Friday, 28 April 2006:

- 9:00:** Moderated discussion of logistical issues, continued
- 23.** Sources of Specimens, obtaining loan permissions
 - 24.** Taxonomic Identifications
 - 25.** Steering Specimens to Sequencing Facilities
 - 26.** Permitting Issues related to international material transfer
 - 27.** Organizational Framework
 - 28.** Data Standards
- 12:30:** Lunch
- 1:30:** Moderated discussion resumes
- 29.** Shared Software Platform, Demonstration of BoLD
 - 30.** Strategies for Data Sharing and Publication
 - 31.** Budget Estimate
 - 32.** Funding Strategies
- 4:00:** Formulation of proposal to CBOL Executive Committee
Next steps, division of task assignments
- 6:00:** Adjourn

APPENDIX 3: List of Important Tephritid Collections

The most important systematic collections are listed below, grouped according to the regional centers:

Smithsonian for New World

- Museum Sao Paolo
- INBio
- CDFA
- Florida State Collection Arthropods
- Instituto de Ecología, Veracruz, Mexico
- Utah State University

Tervuren for Europe/Africa

- BMNH (Richard Lane)
- Geneva Museum (MHNG)
- RMCA
- National Museums of Kenya
- Plant Protection Research Institute, National Collection of Insects
- TAU
- Schmalhausen Institute of Zoology, Ukrainian Academy of Sciences

Korea (Yonsei University?) for east Asia

- Institute of Zoology, Chinese Academy of Sciences
- Monash University Malaysia
- Academia Sinica, Taiwan?
- Japan: Osaka? Tsukuba?
- Kyushu Research Center Forestry and Forest Product Research Institute
- Bogor, Indonesia

(Griffith University?) for Australasia/Pacific Basin:

- Queensland Department of Primary Industries
- Queensland Museum (check with Paul DeBarro)
- Bishop Museum

APPENDIX 4: Potential Funding Sources

- Regulatory agencies
 - APHIS, ARS, CSREES of USDA
 - INRA, CIRAD
 - Dutch Plant Protection Agency
 - Australian Biosecurity Fund
 - Australian Center for International Agricultural Research
 - European Plant Protection Organization
 - IAEA, DoE
 - CDFA Sacramento
 - National Assoc. State Departments of Agriculture
 - Horticulture Australia
 - Agriculture Canada
 - FAO
- Associations of growers
 - California Citrus Research Board
 - Florida Citrus Industry
- Large companies
 - HSBC Bank
- Foundations
 - NSF PEET
 - National research agencies
 - USDA – Binational Agriculture Research and Development
 - USAID
 - APEC

APPENDIX 5: Potential Users and Supporters

- APHIS
- Affymetrix
- New Zealand Ministry of Agriculture and Forestry
- UK Ministry – Central Science Lab, York
- FAO Working Group
- APEC
- Israel Plant Protection & Quarantine
- ICIPE, other CGIARs
- LUCID
- BioNET INTERNATIONAL
- GBIF
- ISSG, GISP, IABIN