

ECBOL NEWSLETTER ISSUE 2 DECEMBER 2010

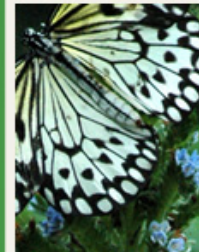
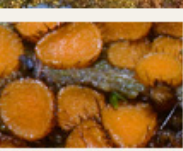
CONTENTS:

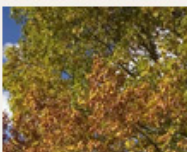
FEATURES

| | |
|--|---|
| DNA Barcoding and the Barcode of Life Data Systems (BOLD) | 1 |
| Barcoding at GenBank | 2 |
| What can the European Nucleotide Archive offer BARCODE submitters and users? | 3 |
| Fungal Barcoding | 4 |
| The mid-term review meeting of the Quarantine Barcode of Life (QBOL)..... | 5 |
| Million Euros for the DNA barcoding factory in Poland..... | 5 |

REGULAR FEATURES

| | |
|--|---|
| Upcoming meetings | 6 |
| Job opportunities | 6 |
| Latest from ecbol.org | 7 |
| Recent publications | 7 |





The newsletter for the European Consortium for the Barcode of Life (ECBOL) is aimed at all members of the ECBOL community and others that are interested in DNA barcoding and biodiversity in Europe and developing countries in Africa. The aim of this newsletter is to keep these communities informed on all issues related to DNA barcoding and biodiversity research and application in Europe. Therefore, this newsletter will feature several short articles on various topics relevant to the ECBOL community. This will also include regular features to keep the community informed about the latest developments.

We would appreciate any contributions and comments regarding features in this newsletter as well as suggestions on improving the content and distribution of information made available in this newsletter.

Feel free to contact us at barcode@cbs.knaw.nl or visit us at www.ecbol.org.

DNA BARCODING AND THE BARCODE OF LIFE DATA SYSTEMS (BOLD).

by Robert Hanner

Early studies using DNA sequences to discriminate species examined a variety of genes and different parts of these genes. This fluidity in analytical targets compromised the use of DNA sequencing for species identification, because some knowledge of species identity was required to select the appropriate primers for the gene region that characterised a particular taxon. DNA barcoding differs from these earlier approaches, being based on the simple yet powerful premise that sequence diversity in short, *standardized* gene regions can provide a sophisticated tool for both the identification of known species and the discovery of new ones.



Dr. Robert Hanner

Standard gene targets are not the only difference between DNA barcoding and prior artisanal approaches that relied on archival databases like DDBJ, EMBL & GenBank (collectively referred to as the International

Nucleotide Sequence Database Collaboration or INSDC), wherein an unknown percentage of such archival sequences derive from vouchered specimens, as nearly all past records lack any explicit connection to vouchers. As a result, cases of sequence disorder among supposedly conspecific individuals, which may reflect specimen misidentifications, cannot be easily resolved. Similarly, the raw sequence data from which INSDC submissions derive are rarely archived. Typically only the sequence assembly itself is deposited, preventing any critical evaluation of the reported nucleotide base calls. Moreover, being archival databases of record, the INSDC does not accept 3rd party annotation or community curation of the data. Thus existing INSDC data is of limited utility for molecular diagnostic applications.

The adoption of 5' COI as a target for DNA barcoding in animals was followed by calls for more stringent data quality standards by the barcode community. In response, the Consortium for the Barcode of Life convened a Database Working Group to address this challenge. A series of consultative meetings with diverse stakeholders (including the INSDC) fostered implementation of a community data standard (see: http://www.barcoding.si.edu/PDF/DWG_data_standards-Final.pdf) which aimed to address the shortcomings of prior approaches. Specifically, the working group determined that animal barcode sequences should be associated with a valid species name and include a bi-directionally sequenced record of at least 500 bp from the 5' end of COI (containing fewer than 1% ambiguous base-calls). Also, reference to the PCR primers used to generate the reference sequences and archival of the underlying electropherogram 'trace files' are required. Details concerning the voucher specimen are also required, including co-ordinates for

the collection locality, collection date, collector, and identifier. Wherever possible, the voucher specimen should be stored in a publicly accessible collection with photographic images of the specimen retained as eVouchers along with other digitized provenance data in order to further document the collection event from which specimens were obtained for barcoding. Together, these combined elements establish "the fitness for use" of barcode data in molecular diagnostic applications – a subtle but important distinction that differentiates barcoding from prior approaches.



The Toronto CN Tower displaying the DNA barcode of the Canadian Beaver during the launch of iBOL, September 2010.

Launched in September of 2010, the International Barcode of Life Project (see: iBOL.org) is driving the assembly of 5M barcode records from 500K species to further develop a standardized reference DNA sequence library for species identification. In this capacity, iBOL is creating a valuable public resource in the form of an electronic database that contains DNA barcodes, images, and geospatial coordinates for the analysed specimens. This information is initially organized and analysed using the Barcode of Life Data System (BOLD, see www.barcodinglife.org), an informatics workbench aiding the acquisition, storage, analysis and publication of DNA barcode records.



The BOLD System is well positioned to support projects that involve broad research alliances like iBOL because of its web-based delivery and flexible data security model. By

providing specialized services, BOLD aids the assembly of records that meet the standards needed to gain BARCODE designation in the INSDC sequence databases and can be used to submit said data at the click of a button. Global barcode coverage in BOLD has surpassed 1M records (fig). BOLD also supports additional markers for plant barcoding, as well as other secondary markers used to corroborate primary barcode data. By assembling molecular, morphological and distributional data, BOLD bridges an existing bioinformatics chasm. It is freely available to any researcher with interests in DNA Barcoding. The system features a well-developed documentation section, complete with data submission protocols (see: <http://www.bold-systems.org/docs/handbook.php>). It also has a support team to assist with any feature of the system (email: support@boldsystems.org). New projects can be annotated with respect to specific iBOL Working Groups (e.g. WG 1.1 vertebrates) or other customised criteria (e.g. ECBOL) to facilitate aggregate progress reporting.

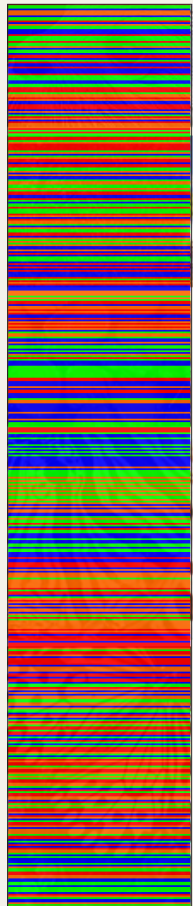
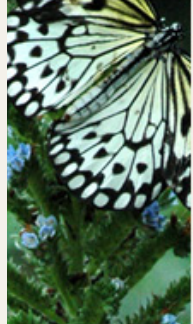
For more information on participation in iBOL and/or ECBOL, please contact Robert Hanner or Lorenzo Lombard at RHanner@uoguelph.ca or l.lombard@cbs.knaw.nl

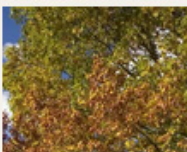
BARCODING AT GenBank

by Conrad Schoch

One of the 4 essential elements required for barcoding is the presence of effective databases (www.barcodeoflife.org/content/about/what-dna-barcoding). The Consortium of the Barcode of Life (CBOL) and GenBank has been involved in laying out the minimum requirements to label a sequence with the BARCODE keyword. The most important of these are the presence of data on voucher specimens and associated meta-data, sequence trace files and an agreed upon gene region. If these conditions are met, a sequence can be submitted in a variety of ways in order to qualify for the BARCODE keyword.

The most straightforward submission method is to use the webinterface, BarSTool at GenBank. The interface can be found here: www.ncbi.nlm.nih.gov/WebSub/index.cgi?tool=barcode. Sequences and their trace files can be uploaded to the nucleotide and trace archive databases and the web interface will take you separately through all the individual steps. There are additional ways for barcodes to be submitted as well. Sequences already submitted by other sub-





mission tools, like Sequin can be requested to be updated as barcodes in the nucleotide database after they meet the barcode criteria. It is important to remember that GenBank comprise the [International Nucleotide Sequence Database Collaboration](#) (INSDC) together with, databases in Europe and Japan, EMBL and DDBJ. All sequences submitted through GenBank will be propagated to all three databases with the BARCODE keyword attached and will be the permanent public repository for barcode data records. The INSDC also houses barcode sequences submitted through third parties, such as the Barcode of Life Database (BOLD), residing at Guelph University Canada.

To date the only fungal sequences in GenBank indicated as barcodes consists of CO1, pending the formal declaration of a fungal barcode. One example, submitted via BOLD, can be found here: <http://www.ncbi.nlm.nih.gov/nuccore/EF180377.1>

WHAT CAN THE EUROPEAN NUCLEOTIDE ARCHIVE OFFER BARCODE SUBMITTERS AND USERS?

by RJ Vaughan.

ENA

The European Nucleotide Archive (ENA; <http://www.ebi.ac.uk/ena/>) (ref 1) is the European arm of the International Nucleotide Sequence Database Consortium (INSDC) (ref2), which includes our collaborating partners Genbank and the DNA Database of Japan (DDBJ). As part of INSDC the ENA both collect and display all nucleotide sequence data, from Sanger sequencing to next generation sequencing technologies, and from short, single locus sequences to complete genomes. All data deposited at each of the INSDC databases is distributed to the others daily, therefore data submitted anywhere can be retrieved everywhere. A two way pipeline between the Barcode of Life Database (BOLD) and Genbank ensures that BARCODE data is synchronised, and ENA hopes to build additional integration with barcode data by providing the sequence component of a future European Barcode of Life database. ENA provides a range of submission options, and have developed and continue to extend our data presentation systems.

Submission

BARCODE data can be submitted to ENA by using our template system, by preparing EMBL format flatfiles for upload (using Se-

quin or Artemis for example), or via a custom template if neither of the other options is appropriate. All BARCODE data is checked by expert biological curators for the completeness of mandatory fields, and the sequences are verified for length and quality. Entries which meet the required data standard developed by CBOL are assigned the INSDC controlled BARCODE keyword, allowing barcode databases and users to identify and harvest them. As the BARCODE keyword is assigned by INSDC but controlled by the BARCODE consortium, in cases where the consortium informs us that the entry does not meet the required standards the keyword can be removed. ENA also provides efficient and easy submission routes for raw sequence data, with in depth advice and support for submitters in preparing data and metadata (ref 3).

Data retrieval and search

In addition to our submission systems, we provide a variety of services for retrieving data from the ENA. These include comprehensive search and browse access to all of our data types, including sequence entries, sample information and project records, through interactive web retrieval, webservices, FTP and a number of other services. The ENA browser uses RESTful URLs, and can output data for display or download in a variety of formats. We now also provide a fast, comprehensive search against all nucleotide sequences in ENA, including expressed sequence tag and whole genome shotgun data.

In addition, we have recently launched a prototype service that provides a taxonomic point of entry into ENA data (see <http://www.ebi.ac.uk/ena/data/view/display=html&Taxon:28901> for example). This service, which will be a focus of further development over the coming months as part of our involvement in the European Commission-funded i4Life project, currently allows access to all ENA data holdings for a given taxon and, optionally, all taxa below it. Over time, we expect to support interoperability with Species2000/Catalogue of Life and will extend coverage to include access to non-sequence molecular data held at EBI.

Finally the ENA is keen to develop new collaborations and to provide help and assistance to both submitters and users of our data. As we are constantly developing and extending our tools and services we encour-

age user feedback to ensure that we can develop relevant and useful tools for all biological communities. Please contact us at datsubs@ebi.ac.uk

References

1. The European Nucleotide Archive

Rasko Leinonen, Ruth Akhtar, Ewan Birney, Lawrence Bower, Ana Cerdeno-Tárraga, Ying Cheng, Iain Cleland, Nadeem Faruque, Neil Goodgame, Richard Gibson, Gemma Hoad, Mikyung Jang, Nima Pakseresht, Sheila Plaister, Rajesh Radhakrishnan, Kethi Reddy, Siamak Sobhany, Petra Ten Hoopen, Robert Vaughan, Vadim Zalunin, Guy Cochrane
Nucl. Acids Res. first published online October 23, 2010 doi:10.1093/nar/gkq967

2. The International Nucleotide Sequence Database Collaboration

Guy Cochrane, Ilene Karsch-Mizrachi and Yasukazu Nakamura
Nucl. Acids Res – In Press

3. The sequence read archive

Rasko Leinonen, Hideaki Sugawara, Martin Shumway
Nucl. Acids Res. first published online November 9, 2010 doi:10.1093/nar/gkq1019

FUNGAL BARCODING

by Conrad Schoch

The route towards the declaration of a formal a fungal barcode has already been described in detail by Keith Seifert in a 2009 paper (Seifert 2009). An important part of this process was a meeting on May 2007 consisting of 37 mycologists from 12 countries at Front Royal, Virginia. At this meeting most participants declared the 5.8S nuclear ribosomal gene and its flanking spacer regions (ITS) to be the most likely candidate for a universal fungal barcode despite some promising but mixed results for the chosen universal gene, COI (Seifert et al. 2007). This is often misunderstood as the point at which a fungal barcode was declared. However, the group which coordinates barcoding and determines the standards for GenBank to use in application of a barcode tag, the Consortium for the Barcode of Life (CBOL) requires additional steps for any gene other than COI. For fungi this includes the publication of a proposal where ITS is compared against a set of other candidates. Much progress has now been

made towards such a publication, slated for the submission in the second half of 2011.

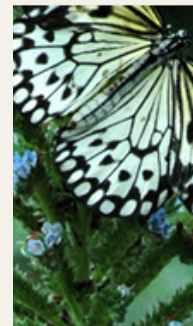
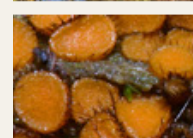
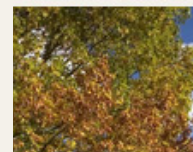
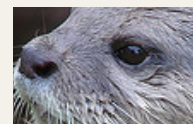
This summer two sessions at mycological conferences yielded a good response from attendees and proved that a sizeable part of the fungal community remains enthusiastic about fungal barcoding. At the Mycological Society of America and the International Symposium of Endophytes on Grasses in Lexington Kentucky, an overview on barcoding was presented and several speakers presented data on several of the main fungal groups as well as the utility of barcoding for environmental sampling with a lively discussion thereafter. This session was followed up in an international context in August at the ninth International Mycology Conference in Edinburgh, Scotland. The overview and handouts presented at the introductory talks to the barcoding sessions are available at www.fungalbarcoding.org. At both these meetings volunteers stepped forward who were willing to provide samples and sequences for a paper that will compare ITS and 3 other genes across the kingdom fungi.



Until now we received commitments for more than 40 clusters of 5 sibling species or more, spread over 12 of the main fungal lineages. The final paper looks set to involve more than 50 coauthors from all continents but participation remains open – we only ask that the deadlines be respected and that complete gene data be provided. More details can be found at the discussion group 'Fungi' on connect.barcodeoflife.net.

Seifert KA, 2009. Progress towards DNA barcoding of fungi. *Molecular Ecology Resources* 9, 83-89.

Seifert KA, Samson RA, Dewaard JR, Hou-





braken J, Levesque CA, Moncalvo JM, Louis-Seize G, Hebert PDN, 2007. Prospects for fungus identification using CO1 DNA barcodes, with *Penicillium* as a test case. *Proceedings of the National Academy of Sciences of the United States of America* 104, 3901-3906.

THE MID-TERM REVIEW MEETING OF THE QUARANTINE BARCODE OF LIFE (QBOL)

by *Lorenzo Lombard*

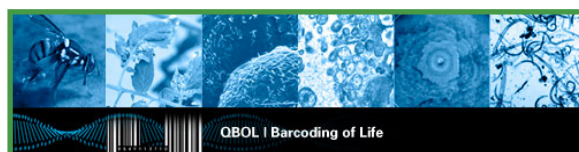


On 21-22 October 2010, the 4th meeting of the European Union (EU) the 7th Framework Programme (2007-13) funded project "Quarantine Barcode of Life" (QBOL), was held in Bologna, Italy, hosted by the University of Bologna. Participants included representatives of various partner institutions and organisations as well as two members of the Advisory Board and two external EU project reviewers. The purpose of the meeting was to highlight the progress of the various work packages in QBOL in view of the mid-term report and review that has to be submitted to the EU Scientific officer by the end of November 2010.

The meeting was opened by Peter Bonants, coordinator of QBOL, with a welcoming of the partners as well as members of the Advisory Board and the EU reviewers. Members of the Advisory Board were André Lévésque, Agriculture and Agri-Food, Canada, and Françoise Petter, European and Mediterranean Plant Protection Organization (EPPO). The EU reviewers were Pietro Piffanelli, "Parco tecnologico Padano", Lodi, Italy, and Juli Mylona, EU Commission representative. As from 1 December 2010, Juli Mylona will act as the EU scientific officer for QBOL; however, she attended the meeting as an external reviewer only.

The meeting consisted of presentations from the group leaders of the eleven work packages in QBOL and highlighted the tasks and deliverables achieved at the mid-point of

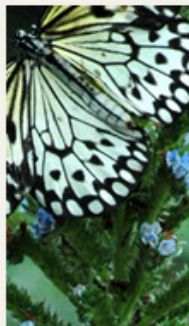
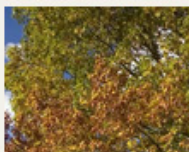
the project for each work package. Based on these presentations, it was clear that majority of these tasks and deliverables have been achieved with the remainder being achieved in the near future. A problem highlighted by several of the work packages was the availability of some of the quarantine organisms that still needs to be barcoded and for these an "Urgently Wanted" list is being compiled and will be available on the QBOL website. Both the Advisory Board and the EU external reviewers were very satisfied by the targets achieved for each work package and by the overall progress of the project. However they recommended a more aggressive approach of actively contacting people to obtain the final few quarantine species or isolates of species still needed for barcoding in the project. For more information on the work packages in QBOL and other QBOL activities, please visit www.qbol.org.



MILLION EUROS FOR THE DNA BARCODING FACTORY IN POLAND.

by *Wieslaw Bogdanowicz*

The Museum and Institute of Zoology of the Polish Academy of Sciences in Warsaw has been granted one million Euros by the Ministry of Science and Higher Education to finish building the DNA barcoding factory and metagenomic facility in Poland. The primary use of this funding will be to buy missing hardware (incl. a new generation sequencer) and to complete the Laboratory Information Management System. The Institute is already involved in DNA barcoding of nematodes occurring in Poland and of mammals from the Caucasus Mountains (a global biodiversity hotspot).



UPCOMING MEETINGS:

- The International Conference on Biodiversity – Senckenberg, Germany – 1-3 December 2010.
- EDIT Year 5 Event – Museum National D’Histoire Naturelle, Paris – 17-19 January 2011.
- Scripting Life: the science behind ViBRANT – Grande Galerie de l’Evolution, Paris – 20-21 January 2011.
- Phylogenetic Systematics and Molecular dating (Course) – University of Copenhagen – 17-28 January 2011.
- Integrative taxonomy and taxonomic expertise in the framework of the DNA-barcoding initiative (Course) – Muséum National d’Histoire Naturelle, Paris – 7-11 February 2011.
- German Society for Tropical Ecology Conference – Frankfurt, Germany – 21-25 February 2011.
- BioSystematics Berlin 2010 – Berlin, Germany – 21-27 February 2011.
- Fungal Barcoding Workshop – CBS-KNAW Fungal Biodiversity Centre, Utrecht, The Netherlands – 17-18 April 2011

If you are planning a DNA barcoding or biodiversity related meeting, workshop or course, please contact us at barcode@cbs.knaw.nl to spread the word to other members of the ECBOL community!!

JOB OPPORTUNITIES:

- Research and PhD positions related to fungal barcoding in the Netherlands (www.cbs.knaw.nl).
- Lectureships, Imperial College London (www3.imperial.ac.uk/employment/lectureship)
- Postdoctoral Researcher at the University of Sheffield (www.sheffield.ac.uk/jobs)
- Assistant Professor/Research Associate in Plant Ecology at Uppsala University (www.personalavd.uu.se/ledigaplatser/engindex.html)
- Associate Professor/Senior Lecturer in Animal Conservation Biology at Uppsala University (www.personalavd.uu.se/ledigaplatser/engindex.html)

If you know of or have any positions available regarding DNA barcoding or biodiversity and wish to advertise these positions in this newsletter or on the ECBOL website, please contact us at barcode@cbs.knaw.nl.

LATEST FROM ECBOL.ORG:

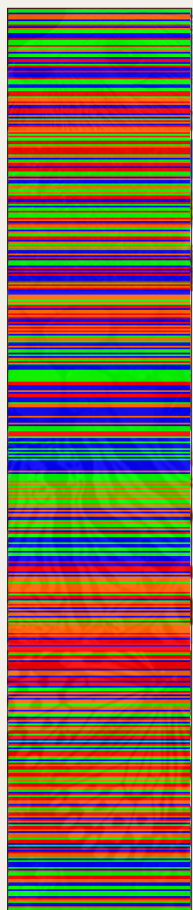
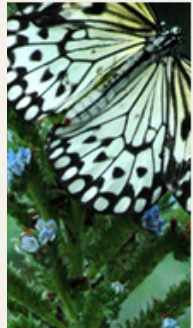
The official ECBOL website, www.ecbol.org, features several new pages that include the latest information on:

- Funding opportunities
- Publications
- Upcoming meetings and courses
- Job opportunities
- Activities on Connect (<http://www.connect.barcodeoflife.net>)

If you wish to contribute to or comment on any of these new features and other features or wish to have more information made available on www.ecbol.org, please contact Lorenzo Lombard at barcode@cbs.knaw.nl.

RECENT PUBLICATIONS:

- Pedersen ML. 2010. DNA barcode profiling: a new platform for the investigation of genome integrity. *Genome Biology* **11**(Suppl. 1): P30.
- Pniewski FF, Friedl T, Latala A. 2010. Identification of diatome isolates from the Gulf of Gdansk: testing of species identifications using morphology, 18S rDNA sequencing and DNA barcodes of strains from the Culture Collection of Baltic Algae (CCBA). *Oceanological*





and Hydrobiological Studies **39**(3): DOI 10.2478/v10009-010-0031-7.

- Raupach MJ, Astrin JJ, Hannig K, Peters MK, Stoeckle MY, Wagele J-W. 2010. Molecular species identification of Central European ground beetles (Coleoptera: Carabidae) using nuclear rDNA expansion segments and DNA barcodes. *Frontiers in Zoology* **7**:26
- Rougerie R, Smith MA, Fernandez-Triana J, Lopez-Vaamonde C, Ratnasingham S, Hebert PDN. 2010. Molecular analysis of parasitoid linkages (MAPL): gut content of adult parasitoid wasps reveal larval host. *Molecular Ecology* doi: 10.1111/j.1365-294X.2010.04918.x
- Zhao P, Luo J, Zhuang W-Y. 2010. Practice towards DNA barcoding of the nectriaceous fungi. *Fungal Diversity* doi: 10.1007/s13225-010-0064-y
- Van Zuydam NR, Paciura D, Jacobs K, Wingfield MJ, Coetzee MPA, Wingfield BD. 2010. Barcoding and microcoding using "identiprimers" with *Leptographium* species. *Mycologia* **102**(6): 1274-1287.

All these scientific publications and more are available on www.ecbol.org.

SPONSORS:



The ECBOL Newsletter will be published quarterly by the European Consortium for the Barcode of Life as part of the EDIT network.

Design and Layout: Lorenzo Lombard & Manon van den Hoeven-Verweij

Please send your contributions, comments or suggestions to barcode@cbs.knaw.nl or to:

Dr. L. Lombard
 Coordinator: European Consortium for the Barcode of Life
 CBS-KNAW Fungal Biodiversity Centre
 Uppsalalaan 8
 3584 CT Utrecht
 The Netherlands
 Tel: +31 (0)30 212 2687
 Fax: +31 (0)30 251 2097

