DNA Barcoding, Capacity Building, and CPM’s Mission

David E. Schindel, Executive Secretary
National Museum of Natural History
Smithsonian Institution
SchindelD@si.edu; http://www.barcoding.si.edu
202/633-0812; fax 202/633-2938
DNA barcoding and forest biosecurity

L.M. Humble, J.R. deWaard, R. Hansen and P.D.N. Hebert

Introduction

The ability to distinguish non-indigenous species (NIS) from native species is critical to the success of any surveillance program. Unfortunately, there are numerous problems inherent with detection of NIS including:

- Large samples needed to detect NIS when they are present at low levels
- Invasion life stages often ignored or underestimated
- Often inadequate knowledge of native flora

We provide examples from forest biosecurity in Canada showing how these problems can be circumvented by the application of DNA barcoding (DNAB) for species identification.

What is DNA barcoding?

- DNAB uses the sequence variation in a short, standardized DNA fragment to identify organisms
- DNA fragment used is a part of a 658 base pair region of the mitochondrial cytochrome c oxidase subunit I (COI) gene (Hobert et al. 2003)
- It is composed of a set of consensus regions against a reference library of DNA sequences

Identification of immature life stages

- Egg, larval or pupal stages are the most frequently encountered life stages of many invasive organisms. They usually cannot be identified with traditional methods
- DNA barcoding can identify stages during life history during the larval stage and the adult stage can be identified using DNA sequences

DNA barcoding for NIS

Methods

- DNA barcoding is used in conjunction with UV lights
- Detecting insects using UV lights (Fitt et al. 1997)

Biosurveillance for NIS

- DNA barcoding can be used to sample for NIS
- DNA barcoding is a powerful tool for identifying NIS

Building DNA reference libraries

- The use of DNA barcoding requires a reference library of DNA sequences
- The library includes sequences from known species
- The library can be used to identify new species

DNA barcoding and forest biosecurity

Lymnaea spp. results (cont’d)

- Lymnaea spp. barcodes identified from across 58 species
- Barcodes shared between species
- COI barcodes within L. dumerilii
- Barcodes of Asian subspecies L. d. peregra and L. d. nipponica
- Barcodes of European & North American subspecies L. d. dumerilii

Literature cited

INSECT PEST DIAGNOSTICS & SPECIES DISCOVERY UNDER IBOL: THE CASE OF OROSIS LEAFHOPPERS

Göpurenk, D (1), Mitchell, A (2), Fletcher, MJ (3) & LÖcker, H (4)

(1) Industry & Investment NSW, Wagga Wagga & Orange NSW, Australia
(2) The Australian Museum, Sydney, NSW, Australia

Introduction

Orosis Diastema (Homoptera: Circadellidae) is a leafhopper genus of economic importance in Africa, the Middle East, Asia, Australia and the Pacific. At least half of the eight described species have been reported as vectors of plant pathogens, including phytoplasmas and viruses. Despite their economic importance, the identities and distributions of Orosis species are uncertain and the genus has a checkered taxonomic history. In addition, the many leafhoppers, only adult males of Orosis are identifiable to species. The resulting uncertainty over the identity of vector species has hampered research into phytoplasma transmission and management.

We performed an integrative taxonomic study of the genus in order to provide tools for identification of the vector-carrying species. DNA barcodes are available for members of the eight recognized species, and several putative novel species were found. Phylogenetic analyses were performed on the barcodes. The results are presented in a cladistic analysis using a general mixed Yule-cospeciation (GMYC) model. The results are compared to morphological characters.

Results & Discussion

Multiple-tree MrBayes modeling applied to 97 unique Orosis leafhopper samples identified 16 genetic clusters (Fig. 1). These clusters included the eight previously recognized species, three novel previously recognized species, and four novel species, each represented by 20 or more sequences. Average sequence divergence within morphospecies was 0.79% (95% CI = 0.47–1.11%), whereas divergence between morphospecies was 11.4% (95% CI = 9.9–13.2%). Monophyly of each morphospecies was supported by 100% bootstrap value, and a single small cluster of species had very strong monophyly support. Further, the topology was consistent with morphological data. DNA barcodes were found to be useful for species identification and classification. Of the eight Orosis species that we examined in this study, four were confirmed to be monophyletic, while the remaining four had more complex phylogenetic relationships. DNA barcodes were found to provide robust identification of the species, allowing accurate species identification and species delimitation.

Conclusions

This study provides the keys to the knowledge of the genus and the implications of our findings for the management of Orosis leafhoppers. Our new integrative approach to species identification using DNA barcodes has proved to be highly effective, allowing accurate species identification and species delimitation.
Species Identification
Matters to All Countries

- Food security and safety
- Border inspection and trade agreements:
  - Agricultural pests/beneficial species
  - Disease vectors/pathogens
  - Endangered/protected species
  - Invasive species
- Ensuring ecosystem services
- Environmental quality assessment
- Documenting/developing genetic resources
- University research in biology
A DNA barcode is a short gene sequence taken from standardized portions of the genome, used to identify species.
Associating Life Stages, Processed Parts, Dimorphic Genders
An Internal ID System for All Animals

The Mitochondrial Genome
- D-Loop
- Small ribosomal RNA
- Cytochrome b
- COI
- COII
- COIII
- ATPase subunit 8
- ATPase subunit 6

DNA
- mtDNA
- H-strand
- L-strand
- ND1
- ND2
- ND3
- ND4
- ND4L
- ND5
- ND6

Typical Animal Cell

Mitochondrion

Small ribosomal RNA
- ATPase subunit 8
- ATPase subunit 6
Non-COI regions for other taxa

- **Land plants:**
  - Chloroplast *matK* and *rbcL* approved Nov 09
  - Non-coding plastid and nuclear regions being explored

- **Fungi and protists:**
  - CBOL Working Groups convened
  - Recommendations expected in 2010
“DNA barcoding already meets or exceeds the minimum standards required for diagnostic protocols under ISPM No. 27.”
Barcoding in Diagnostic Protocols

- Applicable to all life stages
- Usable by non-experts
- Well-documented standard lab procedures
- High degree of transparency and repeatability
- Protocols, results, documentation all public and archived
- IDs and specimen comparisons through digital data, objective criteria
- Linkage to reference voucher specimens
Taxonomy

Name, rank and serial number

Biologists want to barcode half a million species in the next five years

The tale of the unknown goby began in 1982 when Benjamin Victor, of the Ocean Science Foundation in Irvine, California, discovered an unusual fish in a reef in Panama. With only a single specimen he was hard pressed to prove it was a new species, so the fish remained, unnamed, on his desk for 25 years. Then, last year, he was sent an unusual fish larva. Using a new kind of DNA identification called barcoding he showed that it was a younger version of his mystery goby and that both specimens were, indeed, a new species.

DNA barcoding was invented by Paul Hebert of the University of Guelph, in Ontario, Canada, in 2003. His idea was to generate a unique identification tag for each species based on a short stretch of DNA. Separating species would then be a simple which there are at least 3,500 species, many of them hard to tell apart.

So far Dr Linton’s team has used the COI gene to distinguish 390 species of mosquito, of which 7% have turned out to be new species. Anopheles oswaldoi, for example, was known to be a carrier of malaria in northern, but not southern, Brazil. That was puzzling. DNA barcoding, however, has shown that A. oswaldoi is actually four species, of which only one carries malaria. That explains the geographical discrepancy and should also assist efforts to curb the disease in Brazil by allowing the real culprit to be studied in detail.

Fly titles
The mosquito initiative has also had a piece of luck. Using some chemical wiz-
as medicines. In doing so, they have had to identify a new kind of barcode, as the COI gene is not found in plants.

Another group that could benefit from barcoding are customs officers, says Mark Blaxter, an evolutionary biologist at the University of Edinburgh. For those struggling to prevent the importation of pests or endangered wildlife, rapid and accurate identification tools are essential—particularly when perishable goods are being held up. America’s Department of Agriculture is creating barcodes for the world’s fruit flies. These are important agricultural pests and often arrive in the country as hard-to-identify larvae, or eggs, on fruit.

Another group at the National Chung Hsing University in Taiwan (where hundreds of newly minted experts in the field have just met for the Second International Barcode of Life Conference) have created a prototype barcoding biochip. This is a collection of miniature DNA test sites on a sliver of glass that will rapidly discriminate between four species of fruit flies.

Barcoding’s ease of use is also attracting interest from other government agencies. America’s Federal Aviation Administration and its air force are working on bird
July 2010 Technical Panel on Diagnostic Protocols
Washington, DC
How Barcoding Works

First, build a barcode reference library:
- Well-identified specimen
- Tissue subsample
- DNA extraction, PCR amplification
- DNA sequencing
- Data submission to GenBank

Second, use it to identify unknowns:
- Any unidentified juvenile, adult, fragment, product
- Tissue sample, DNA, sequencing
- Comparison with sequences in reference library
The Barcoding Pipeline
From specimen to sequence to species

Voucher Specimen
Collecting
DNA extraction
CO1 gene
DNA sequencing
Trace file
Database of Barcode Records
Current Norm: High throughput
Large labs, hundreds of samples per day

 ABI 3100 capillary automated sequencer

Large capacity PCR and sequencing reactions
- US$100-165K purchase
- 2-3 hours processing time
- 150-500 samples per day
- US$3-5 per sample
Technology Development Partnership Goal

The DNA Sequencing Lab of 2013?
Producing Barcode Data: 201?
Barcode data anywhere, instantly

- Data in seconds to minutes
- Pennies per sample
- Link to reference database
- A taxonomic GPS
- Usable by non-specialists
BARCODE Records in INSDC

Specimen Metadata
- Georeference
- Habitat
- Character sets
- Images
- Behavior
- Other genes

Voucher Specimen

Barcode Sequence
- Trace files
- Primers

Literature
- (link to content or citation)

Species Name
- Indices
  - Catalogue of Life
  - GBIF/ECAT
- Nomenclators
  - Zoo Record
  - IPNI
  - NameBank
- Publication links
  - New species
  - Databases
    - Provisional sp.

Other Databases
- Phylogenetic
- Pop’n Genetics
- Ecological

CONSORTIUM FOR THE BARCODE OF LIFE
1 Million+ records, 100K+ species
Sequence Webpages
Ostichthys kaianus (Günther, 1880)

Deepwater soldier

Species recognized by FishBase, R. Froese & D. Pauly (eds) in Catalogue of Life
IUCN RED LIST STATUS: NOT EVALUATED

IMAGES

COPYRIGHT: Some rights reserved
SUPPLIER: FishBase
SOURCE: John E. Randall

fish market: BPBM 10045 Locality: Naha
GenBank, EMBL, and DDBJ
Global, Open Access to Barcode Data

International Nucleotide Sequence Database Collaboration

- The International Nucleotide Sequence Databases (INSD) have been developed and maintained collaboratively between DDBJ, EMBL, and GenBank for over 18 years.
- The INSDC advisory board, the International Advisory Committee, is made up of members of each of the databases’ advisory bodies. At their most recent meeting, members of this committee unanimously endorsed and reaffirmed the existing data-sharing policy of the three databases that make up the INSDC, which is stated below.
- Individuals submitting data to the international sequence databases should be aware of INSDC policy.

How to submit data

- For full details of how to submit data to the databases, please select a collaborating partner.
- DDBJ, EMBL, GenBank
- The INSDC Feature Table Definition Document is available here.
Link from GenBank to Museums

Mammal Collection
University of Alaska Museum of the North

Specimen Search | Publication/Project Search | Advanced Features

UAM Mamm 86887
Orcinus orca
Details
BerkeleyMapper

Location
North America, United States, Alaska, Cordova Quad
Oktue Spit, near Kayak Island
Lat/Long: 60° 3' 32.9" N 144° 10' 48" W ± 1 km
Collecting Date: 28 Jul 2006
Collectors: Tim Lebling
Preparators: Pam Tuomi

Used By: Canadian Barcode of Life Network

Parts: liver, heart, muscle; kidney

Individual Attributes

Sex: unknown

Standard Measurements:
- Total length: 584 cm
- Tail length: 82 cm
- Hind foot: 85 cm
- Ears: 9 cm
- Weight: 281 kg

Remarks: Necropsy by Pam Tuomi; ASK.

Identifying Numbers
- Original identifier: 00-0002
- AF: 50332
- GenBank: EU139289

Mammal Collection
University of Alaska Museum at the University of Alaska Fairbanks, Fairbanks, AK 99775-6980.

System Administrator is Dusty McDonald.
How Barcoding Works

First, build a barcode reference library:

- Well-identified specimen
- Tissue subsample
- DNA extraction, PCR amplification
- DNA sequencing
- Data submission to GenBank

Second, use it to identify unknowns:

- Any unidentified juvenile, adult, fragment, product
- Tissue sample, DNA, sequencing
- Comparison with sequences in reference library
How Complete is the Barcode Library?

- More than 1 million records in BOLD
- More than 100,000 species represented
- Projects underway in all major groups
- Focus on groups with commercial and societal importance:
  - Agricultural pests
  - Disease vectors
  - Endangered species
Barcode of Life Community

- Promote barcoding as a global standard
- Build participation
- Working Groups
- BARCODE standard
- International Conferences
- Increase production of public BARCODE records

Networks, Projects, Organizations

- ECBOL
- BOLDSYSTEMS
- ABBI
- FISH-BOL
- MBI
- TBI
- CCDB
- BIO

Canadian Barcode of Life Network
Investments in Barcoding

- ~US $5 million per year
  - Smithsonian Laboratories for Analytical Biology
  - Smithsonian barcoding projects
  - Sloan Foundation support for CBOL
  - Project support by USDA, EPA, FDA, FAA...
  - Barcoding in NSF-funded biodiversity grants
Adoption by Regulators

- USDA, Belgian research projects on fruit flies
  - Plans for submission of Diagnostic Protocols
- Food and Drug Administration
  - Reference barcodes for commercial fish
- Environmental Protection Agency
  - $250K pilot test, water quality bioassessment
- NOAA/NMFS
  - $100K for Gulf of Maine pilot project
  - FISH-BOL workshop with agencies, Taipei, Sept 2007
- Federal Aviation Administration – $500K for birds
Investments in Barcoding

- ~US $5 million per year
- CAN $80 million over 2005-2015
- Commitments of ~CAN $75 million from iBOL partners over 2010-2015
  - 5 million specimens
  - 500K species
  - 25 partner countries
  - Canada, US, EU, China are “central nodes”
Consortium for the Barcode of Life (CBOL)

- Established May 2004 with Sloan Foundation grant
- Secretariat hosted by Smithsonian Institution
- Now in its fourth two-year funding period
- Workshops, Working Groups, networking, representation/marketing
- Now an international affiliation of 200+ members in 50+ countries:
  - Natural history museums, biodiversity organizations
  - Users: e.g., government agencies
  - Private sector biotech companies, database providers
CBOL Member Organizations: 2010

- 200+ Member organizations, 50 countries
- 35+ Member organizations from 20+ developing countries
Building the Community

- Internal communication through Community Network (http://connect.barcodeoflife.net)
- Outreach communication through
  - www.barcodeoflife.org
  - CBOL Webinars
- Coordination with other barcoding projects through CBOL’s Implementation Board
- Steering Committee planning meetings
- Assistance in preparing and submitting proposals
Welcome to ConnectBarcodeOfLife.net

New to our online community? Here are few things you can do to get started:
- Complete your profile
- Add yourself to the member map
- Browse our Forum and ask a question or leave a comment
- Join a Group or start your own
- Blog about barcoding here or
- Let us know if you'd like to share your own blog through our rss pages
- Have a question? Ask one of our hosts or Check out our FAQ.
Barcoding Projects

There are many international barcoding activities dedicated to the development of targeted public reference BARCODE sequence libraries.

FEATURED PROJECTS

Sort By: Title

All Birds Barcoding Initiative (ABBI)

ABBI, the All Birds Barcoding Initiative, is a campaign to collect DNA barcodes from 5 or more individuals of all of the approximately 10,000 bird species in the world. The ABBI DNA barcode library will help speed discovery of new species, open new avenues for scientific investigation, and provide a forensic tool for identifying specimens, including for example tissue fragments from bird-airplane collisions and avian blood samples from biting insects that harbor West Nile virus or other human disease agents.

Project Site

All Fungi Barcoding

All Fungi Barcoding provides up-to-date information on fungal barcoding and facilitates communication and collaboration among researchers interested in fungi.

Project Site
International Barcode Conferences

- Natural History Museum, London: 2005
- Academia Sinica, Taipei: 2007
- UNAM, Mexico City: 2009
- University of Adelaide, Australia: 2011
- All-Africa Conference: 2012

- 30-60 Travel Bursaries awarded for participants from developing countries
Challenges

- Raising awareness about barcoding
Outreach Activities

- Cape Town, South Africa, April 2006, SANBI
  - Scale insects in African agriculture
- Nairobi, Kenya, October 2006
  - Commercial fisheries in Rift Valley lakes
- Brazil, March 2007
  - Hardwood tree species
  - Endangered mammals, reptiles, amphibians
- Taiwan, September 2007
- Nigeria, October 2008
- Beijing, May 2009
- India, November 2010
Challenges

✓ Raising awareness about barcoding

■ Buy-in by national/international authorities
  – Access to study specimens for international research under the Convention on Biological Diversity
DNA Barcoding: A New Tool for Identifying Biological Specimens and Managing Species Diversity
ABS Workshop, Museum Koenig
17-19 November 2008
**51 Participants from 24 Countries**

<table>
<thead>
<tr>
<th>Sector</th>
<th>Research</th>
<th>Agency</th>
<th>Other</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>29</td>
<td>10</td>
<td>12</td>
</tr>
<tr>
<td></td>
<td>56.9%</td>
<td>19.6%</td>
<td>23.5%</td>
</tr>
</tbody>
</table>

**Geographic Representation**

<table>
<thead>
<tr>
<th>Region</th>
<th>OECD</th>
<th>Africa</th>
<th>Latin America</th>
<th>Asia</th>
<th>Pacific</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>28</td>
<td>8</td>
<td>4</td>
<td>9</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>54.9%</td>
<td>15.7%</td>
<td>7.8%</td>
<td>17.6%</td>
<td>3.9%</td>
</tr>
</tbody>
</table>
The supreme decision-making body of the United Nations Convention on Biological Diversity (CBD) meets in Nagoya, Japan, on 18–29 October 2010 for its tenth biennial conference. One of the most important items on the agenda is a new protocol which, if enacted, would specify how countries that are parties to the convention control access to their “genetic resources” (including whole organisms, tissue samples and DNA extracts) and what benefits they can expect from sharing them. The negotiators focus on genetic resources used to develop commercial products has left non-commercial academic research in a precarious position. \[...\]

A researcher prepares to analyze plant samples from the biodiversityrich region around Kauai.

**Biology without borders**

Fundamental research must not be hampered by an international agreement on sharing the benefits from national biodiversity, says David Schindel.
CBD International Regime for Access and Benefit Sharing

In the development and implementation of their national legislation on access and benefit-sharing, [and on the basis of the sovereign right of Parties who regulate access to genetic resources and its derivatives,] Parties shall:

(a) Create conditions to **promote and encourage research** which contributes to the conservation and sustainable use of biological diversity, particularly in developing countries, including through **simplified measures on access for non-commercial research purposes**, taking into account the need to address a change of intent for such research.
Challenges

✓ Raising awareness about barcoding

▪ Buy-in by national/international authorities

▪ Start-up funding

  ✓ Mexican national barcoding network, equipment and project grants
  ✓ Brazilian national funding program
  ✓ India national initiative
  ✓ South African national network
Challenges

✓ Raising awareness about barcoding

■ Buy-in by national/international authorities

■ Start-up funding

■ Training

✓ CBOL training opportunities for researchers, students

✓ Annual short courses: Buenos Aires, Johannesburg, Paris

– Needs to be scaled up

– Needs to be extended to regulatory officials, other users
Challenges

- Raising awareness about barcoding
- Buy-in by national/international authorities
- Start-up funding
- Training
- Capacity building
  - Specimen repositories
  - Small labs for DNA extraction
  - National/Regional sequencing centers
  - Informatics capabilities
Welcome

The Consortium for the Barcode of Life and the University of Adelaide invite you to join us in Adelaide, Australia from 28 November - 3 December 2011 for the Fourth International Barcode of Life Conference. Barcoding has seen extraordinary growth since the Mexico City Conference in November 2009 so join participants from around the world for the biggest barcoding event ever!

The organizers have developed this website to provide potential participants, co-sponsors, and other stakeholders with information about the conference. The conference organizers are also eager to have your feedback as we plan the conference so please share your ideas through Connect, the DNA Barcoding network. You can do this by using the links found throughout this website.

Important Dates

University of Adelaide
South Australia
28 November – 3 December 2011