



## BARCODE OF LIFE DATA SYSTEMS HANDBOOK

November 2011

[www.boldsystems.org](http://www.boldsystems.org) (version 2.5)  
[v3.boldsystems.org](http://v3.boldsystems.org) (version 3.0 Beta)

## Table of Contents

### Getting Started

Introduction	3
Navigation	4
Searches	5
Downloads	5
Registration	5

### Databases

Identification Engine	6
Taxonomy Browser	7
Publication Database	8
Primer Database	9
Public Data Portal	10
Barcode Index Numbers (BINs)	11

### Data Management

Specimen and Sequence Data	12
Data Submission	13
Image Submission	16
Photography Guide	18
Trace Submission	19
Sequence Submission	20
Primer Submission	21
Publication Submission	21

### Managing Data

User Console	22
Create a Project	23
Project Console	24
Searching	25
GenBank/BOLD Publication	26

### Analysis Tools

Image Library	27
Distribution Maps	28
Taxon ID Tree	29
Distance Summary	30
Sequence Composition	30
Nearest Neighbour Summary	31
Accumulation Curve	31
Alignment Viewer	32

This handbook provides details on functionality, data structures and best practices for BOLD version 3.0 (beta). It explains how to use this system to collect, manage and publish Barcode and ancillary data. It also provides details on the integrated analytical tools and web services. At any time while using BOLD, you can access the online documentation by clicking on the “Get Help” link in the footer of every page, or by selecting “Documentation” from the page header.

For an online version of our documentation for BOLD version 2.5, please visit: [www.boldsystems.org/docs/](http://www.boldsystems.org/docs/)  
 For assistance with any feature of BOLD, please email the BOLD Support Team: support@boldsystems.org

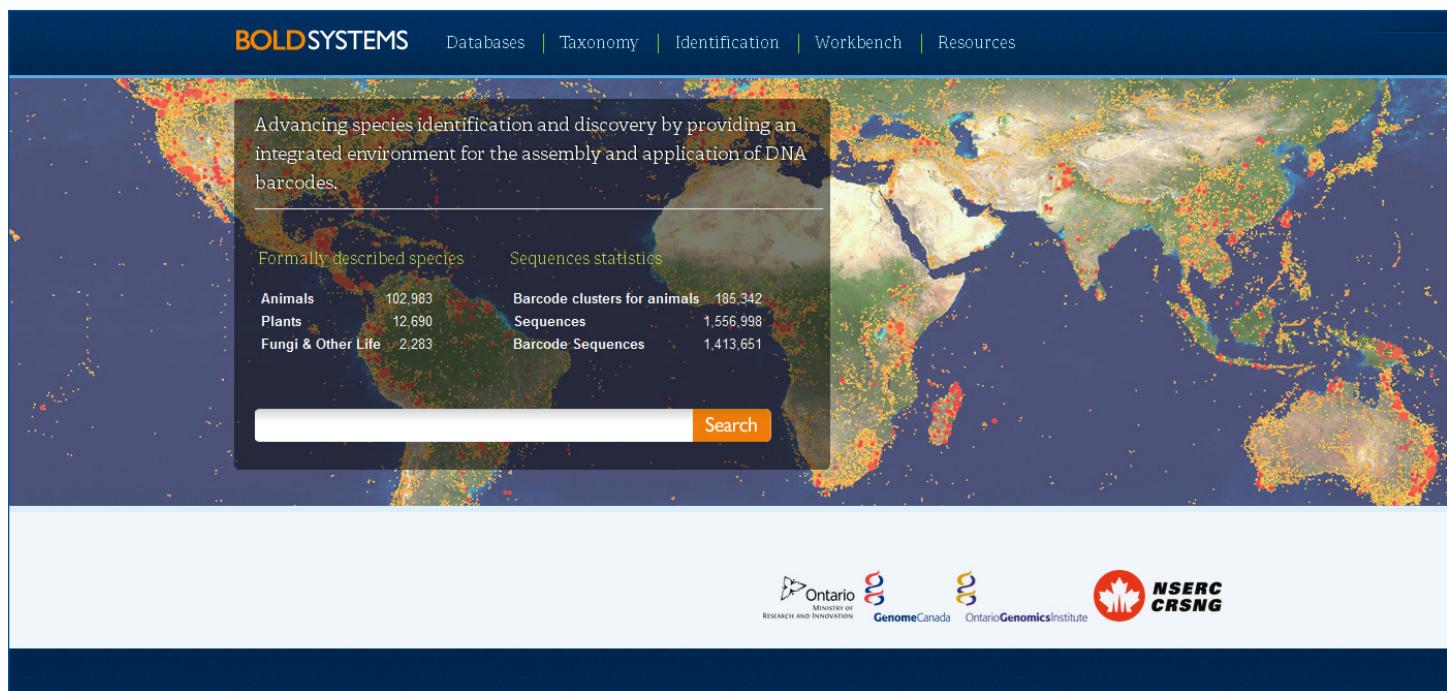
## Introduction

The Barcode of Life Data Systems (BOLD), established in 2005, is a web platform that provides an integrated environment for the assembly and use of DNA barcode data. It delivers an online database for the collection and management of specimen, distributional, and molecular data as well as analytical tools to support their validation. Over the past few years, BOLD has grown to become a powerful online workbench and the central informatics hub of the DNA barcoding community.

BOLD is freely available to any researcher with interests in DNA Barcoding. By providing specialized services, it aids in the publication of records that meet the standards needed to gain BARCODE designation in the international nucleotide sequence databases. Because of its web-based delivery and flexible data security model, it is also well positioned to support projects that involve broad research alliances.

The release of version 3.0 of BOLD in November 2011 represents an evolutionary update to the system. Significant revisions have been made to support an increasing diversity of workflows and an increasing volume of data. A major advance is the activation of Barcode Index Numbers (BINs), an interim taxonomic system for animals, and an annotation framework that supports rapid community based validation of barcode data. As of November 2011, this version is currently in BETA, and the older version of BOLD (2.5) will still be available in parallel.

For more information on the system, please email [info@boldsystems.org](mailto:info@boldsystems.org).



[www.boldsystems.org](http://www.boldsystems.org) (BOLD 2.5)  
[v3.boldsystems.org](http://v3.boldsystems.org) (BOLD 3.0 beta)

The 2.5 and 3.0 versions of the BOLD software access the same database, so changes made in one are accessible in the other.

Subscribe to our social media for updates on our site



@BOLDteam



BOLDsystems



Barcode Connect: BOLD Group

## Getting Started on BOLD

BOLD 3.0's interface improves access to commonly used features and new databases.

From the top menu, you can now access the public databases, the taxonomy browser, the identification engine, the user workbench and resources. In addition, the footer provides access to details on the BOLD organization, community sites and partner sites.

Another change in the BOLD 3.0 interface is that navigating around BOLD outside of the workbench will not log out your user account, so you can move more easily through the system.

Please see the diagram and table below for description of the navigation features.

A publication search page illustrating BOLD 3.0 navigation features

1	Home Page	Clicking on the BOLD logo in the top left corner of any page will bring you back to the BOLD 3.0 home page.
2	Log in/Log out	In the top right corner of any page, you can log in or log out. When logged in, your user name appears in this section.
3	Databases	<p>The Databases link provides access to the following resources that are accessible without a user account.</p> <p><b>Public Data Portal:</b> A database of all of the public sequences on BOLD, including those in the early data release phase of the iBOL project. This database can be used to access and download the associated specimen data and sequences. Search by taxonomic, geographic, institution or identifier keywords.</p> <p><b>BIN Database:</b> Barcode Index Numbers (BINs) are an interim taxonomic system for animals. Barcodes are clustered algorithmically, generating a web page for each cluster which is deposited in this database. Clusters show high concordance with species, which provides a fast-track for documenting diversity where taxonomic resources are limited. Search by taxonomic, geographic, institution or identifier keywords.</p> <p><b>Primer Database:</b> A searchable, database of barcode primers, which includes primer statistics. Search by primer code, submitter name or reference keywords.</p> <p><b>Publication Database:</b> A searchable, community maintained database of barcode papers linked to published datasets. Search by title, abstract or author keywords.</p>
4	Taxonomy	The taxonomy link provides access to the taxonomy browser, a public resource which contains a page that displays the images, distribution map and other details for each taxon on BOLD. Each image uploaded to BOLD has a license applied to it. Images may be used from the taxonomy browser if the image licensed as Creative Commons or No Rights Reserved, following the rules of the license.
5	Identification	The Identification link provides access to the animal, plant and fungal identification engines based on the COI, matK, rbcL and ITS genes. This resource is available without need for a user account.
6	Workbench	The workbench link provides access to the BOLD data analysis and management workbench. The initial page is the user console, which is described on the next page.
7	Resources	The resources link provides access to Site Documentation, Training Materials and Workflow Documents.
8	Sub-header	The sub-header on every page will list the name of the page that you are on.
9	Search Bars	On any page that provides access to a database, a search bar will appear in a blue box at the top of the page. See the next page for more details on these search bars.
10	Page Footer	The page footer provides access to such resources as organizational details about BOLD, community pages, and partner sites.

## Searching

The blue search bars that appear on every database (including Primer, Publication, Public Data Portal, etc), users can enter any combination of keywords. For example, searching “Lepidoptera Canada” in the Public Data Portal will return all of the Lepidoptera records collected in Canada.

For further details and examples for using the search functionality, see the search help section that is available by clicking on the help button to the right of the search bar in each database.

## Downloads

In each database, there is also an option to download the public data available based on your search terms. This includes primer sequences, bibliographies, sequences, traces, taxon ID trees, and specimen data.

## Registering for a User Account

Getting an account on BOLD expands the list of options available to a user of the system beyond access to public data and use of the identification engine. Upon signing in, users can annotate published data, helping to curate and clean the identification library. Moreover, it will be possible to submit data to BOLD and to gain access to other in-progress, private, projects with the permission of the data owners. Once data is on BOLD, a large set of analytical tools are available for validation and generation of reports for publications.

User Console for a new user to BOLD

After you have submitted your registration, a welcome e-mail will be sent to you with the information you need to log in and begin using your BOLD account. With a new user account, upon sign in, you will see the new user console illustrated below. After you start to contribute data and join your collaborators projects, the user console will give you progress statistics and activity feeds (see page 22 for a depiction of the user page after you have access to data).

The new user console provides access to search the public data on BOLD, by project code, title or tag, or using the record search (see page 25 for more details on the record search). A new user can submit primers and bibliographies. After you have access to projects (either by creating them, or gaining access to colleague's projects) you can upload specimen data, images, traces and sequences.

## Identification

The library of sequences collected in BOLD is available for facilitating identification of unknown sequences. The ID engine uses all sequences uploaded to BOLD from private, as well as public projects to locate the closest match. To protect BOLD users, no sequence information from private records is exposed.

### Animal Identification (COI)

The BOLD Identification System (IDS) accepts sequences from the 5' region of the mitochondrial gene COI and returns a species-level identification when possible. Further validation with independent genetic markers is desirable in some forensic applications. BOLD uses the BLAST algorithm to identify single base indels before aligning the protein translation through profile to a Hidden Markov Model of the COI protein. There are four databases within BOLD for use in identification of COI sequences:

#### 1. All Barcode Records Database includes:

Every COI barcode record on BOLD with a minimum sequence length of 500bp (Warning: This is an un-validated database and includes records without species level identification). This includes many species represented by only one or two specimens, as well as all species with interim taxonomy. This search only returns a list of the nearest matches and does not provide a probability of placement to a taxon.

#### 2. Species Level Barcode Database includes:

Every COI barcode record with a species level identification and a minimum sequence length of 500bp (Warning: This is an un-validated dataset). This includes many species represented by only one or two specimens, and all species with interim taxonomy.

#### 3. Public Record Barcode Database includes:

All published COI records from BOLD and GenBank with a minimum sequence length of 500bp. This library is a collection of records from the published projects section of BOLD.

#### 4. Full Length Record Barcode Database includes:

A subset of the Species library with a minimum sequence length of 640bp and containing both public and private records. This library is intended for short sequence identification as it provides maximum overlap with short reads from the barcode region of COI.

### Fungal (ITS) and Plant (rbcL & matK) Identification

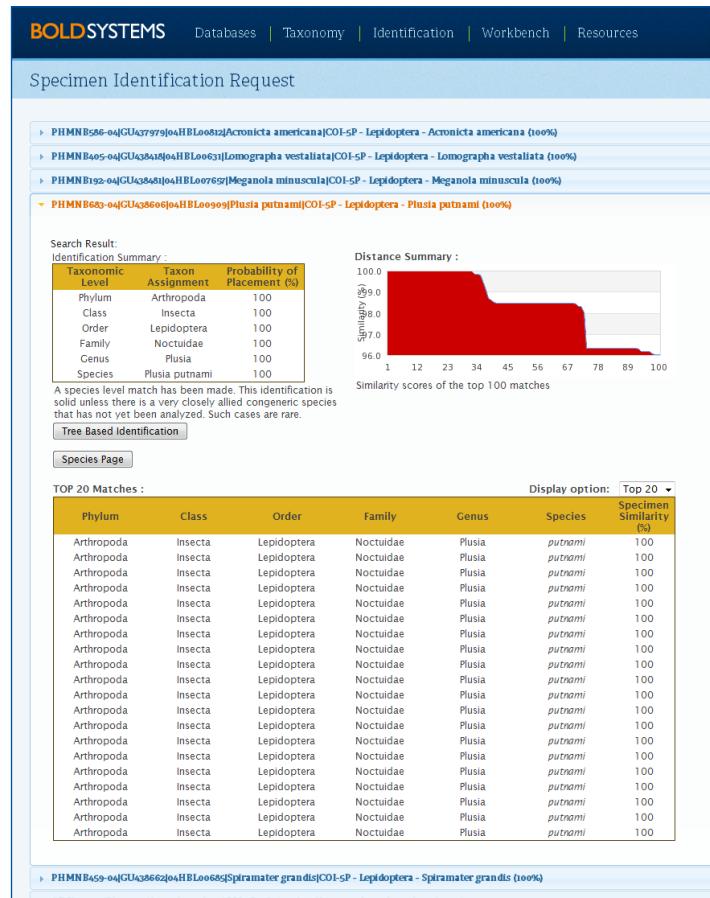
In the BOLD Identification System (IDS), ITS is the default identification tool for fungal barcodes and rbcL and matK are the defaults for plant barcodes. Both return a species-level identification when possible. Further validation with independent genetic markers will be desirable in some forensic applications. The BLAST algorithm is employed in place of BOLD's internal identification engine for these sequences. There are relatively few fungal and plant records on BOLD so most queries will likely not return a successful match. This will improve as sampling efforts continue in these kingdoms. These databases include many species represented by only one or two specimens, as well as all species with interim taxonomy. Both searches only return a list of the nearest matches and do not provide a probability of placement to a taxon.

#### Fungal Database includes:

Every ITS barcode record on BOLD with a minimum sequence length of 100bp (Warning: This is an un-validated database that includes records without species level identification).

#### Plant Database includes:

Every rbcL and matK barcode record on BOLD with a minimum sequence length of 500bp (Warning: This is an un-validated database that includes records without species level identification).



Identification Engine Results Page for batch identification

### Batch Identifications

BOLD 3.0 provides the ability to submit a batch of query sequences for identification. This service is available to users by request. To access this functionality, please email support@boldsystems.org

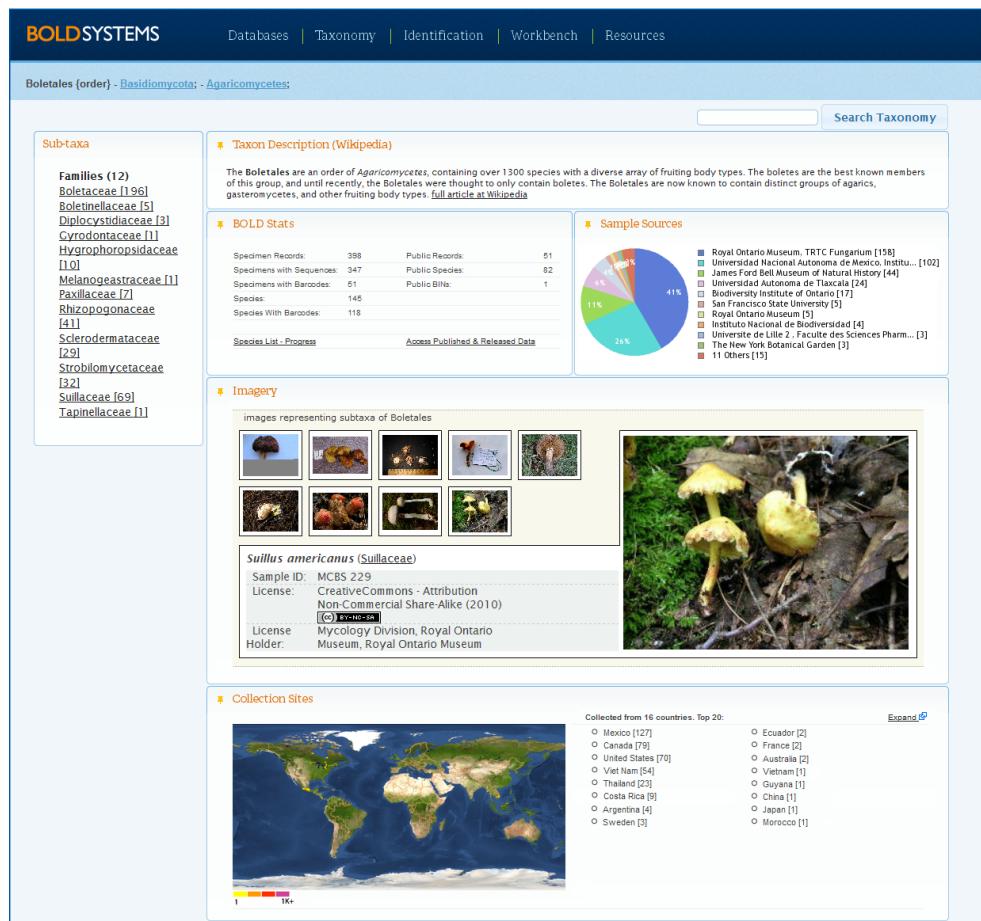
## Taxonomy Browser

The taxonomy browser is a public resource that allows users to examine the progress of DNA barcoding by browsing through the different levels of the taxonomic hierarchy available on BOLD. To access the taxonomy browser, select the Taxonomy link in the header of any main page on BOLD.

Within the taxonomy browser you will be able to select between the Animal, Plant, Fungus, and Protist kingdoms to navigate all the way from phylum to species level. BOLD statistics on the progress of DNA barcoding at each taxon are generated from both public and private data with limited access to private data.

To search for your taxon of interest directly, use the search function by entering any taxonomic name into the search bar on the BOLD main page, or at the top of the taxonomy browser.

The figure to the right depicts the taxonomy browser for a selected taxon. See the table below for descriptions of each section featured on the taxonomy browser taxon page.



BOLD taxonomy browser

Lineage	Displays the taxon name and the higher taxonomic levels.
Sub-Taxonomy	Links to all sub-taxa with number of specimen records for each.
Taxon Description	Displays the description of this taxon from the Wikipedia website.
Statistics	These statistics are compiled by BOLD for this taxon. A species progress list can be download for each rank that has sub-taxa. The published and released sequences for this taxon can be downloaded from this section.

Information available at each taxonomic level within the BOLD taxonomy browser.

Sample Sources	A graph of the top institutions that provided specimens with their specimen tallies.
Imagery	A random selection of the images available for the subtaxa of this taxon. Mousing over an image selects it for higher-resolution display to the right.
Image Details	The taxonomic identifier, the sample identifier, license and attribution are displayed beneath the image that is selected.
Collection Sites	A map of the collection sites including a list of the top countries
Taxon Occurrence	A map of the occurrence data for this taxon worldwide, streaming from the GBIF website.

## Databases: Publication

The publication database is accessible from anywhere in the application

This database indexes title, abstract, year, and authors, allowing for broad searches.

Selecting a publication from the database will provide details on the publication, including a link to the article on the journal's site.

A citation or set of citations can be downloaded from BOLD using the button the the right of the search bar.

Bibliographies can be submitted to this database by following the directions on page 21.

Publication Search

Search:  Go

Showing Records 1 to 50 Page 1 2 3 4 5 6 7 8 9 10 next> Records Per Page 50 Download Publications Selected ▾

1. Tropical montane nymphalids in Mexico: DNA barcodes reveal greater diversity  
Journal: Mitochondrial DNA  
Year: 2011  
Volume: 21  
Issue: S1  
Pages: 30-37  
Url: <http://www.mitogenome.org/article/info%3Adoi%2F10.1371%2Fjournal.pone.0017134>

2. DNA Barcoding the Geometrid Fauna of Bavaria (Lepidoptera): Successes, Surprises, and Questions  
Journal: PLoS One  
Year: 2011  
Volume: 6  
Issue: 2  
Pages: e17134  
Url: <http://www.plosone.org/article/info%3Adoi%2F10.1371%2Fjournal.pone.0017134>

3. DNA extraction techniques for DNA barcoding of minute gall-inhabiting wasps  
Journal: Molecular Ecology Resources  
Year: 2011  
Volume: 11  
Issue: 6  
Pages: 1755-0998  
Url: <http://www.molecular-ecology-resources.com/article/info%3Adoi%2F10.1111/j.1365-294X.2011.00000.x>

4. Building freshwater macroinvertebrate DNA barcode libraries from reference collection material

Download Current Bibliography: RIS EndNote BibTeX

Title: DNA Barcoding the Geometrid Fauna of Bavaria (Lepidoptera): Successes, Surprises, and Questions  
PMID: N/A  
Journal: PLoS One  
Access: public  
Year: 2011  
Volume: 6  
Issue: 2  
Pages: e17134  
PDF URL: <http://www.plosone.org/article/info%3Adoi%2F10.1371%2Fjournal.pone.0017134>  
Open PDF  
DOI: N/A  
Pii: N/A  
Citekey: N/A  
Language: English  
Citations: 0

## Databases: Primer

The primer database is accessible anywhere in the application.

Using the search bar, one can search terms that appear in the primer code, submitter or reference.

Selecting a primer from the database will provide details on the primer, including primer performance statistics derived from data submitted to BOLD.

A primer or set of selected primers can be downloaded in FASTA format using the button the the right of the search bar.

New primers must be registered with BOLD before trace files generated using them are submitted. For details on registering a new primer, see page 21.

Primer Search

Search: Lepidoptera Go

Showing Records 1 to 11 Page 1 Records Per Page 50 Download Selected Primers

1. T-LepF1-short  
Marker: COI-5P  
Submitter: Erik J. van Nieukerken  
Reference: Hebert et al. (2004). Ten species in one: DNA barcoding reveals cryptic species in the neotropical skipper butterfly *Astrapes fulgerator*. *Proceedings of the National Academy of Sciences of the United States of America* 101 (41): 14812-14817 Wahlberg, N., West Wheat, C. (2008). Genomic outposts serve the phylogenomic pioneers: Designing novel nuclear markers for genomic DNA extractions of Lepidoptera. *Systematic Biology* 57 (2): 231-242

2. T-LepR1-short  
Marker: COI-5P  
Submitter: Erik J. van Nieukerken  
Reference: Hebert et al. (2004). Ten species in one: DNA barcoding reveals cryptic species in the neotropical skipper butterfly *Astrapes fulgerator*. *Proceedings of the National Academy of Sciences of the United States of America* 101 (41): 14812-14817 Wahlberg, N., West Wheat, C. (2008). Genomic outposts serve the phylogenomic pioneers: Designing novel nuclear markers for genomic DNA extractions of Lepidoptera. *Systematic Biology* 57 (2): 231-242

3. EF-Nef  
Marker: EF1-alpha  
Submitter: Erik J. van Nieukerken  
Reference: Nieukerken, E.J. van, C. Doorenweerd, FR. Stokvis & D.S.J.

Primer Data

Primer Code(3 to 12 letters):	T-LepF1-short
Primer Name:	null
Alias:	null
Codes(Comma Separated):	
Target Marker:	COI-5P
Cocktail Primer:	No
Primer Sequence (5' to 3'):	TAATACGACTCACTATAGGGATTCAACCAATCTAAAGATAT
Direction:	F
Reference/Citation:	Hebert et al. (2004). Ten species in one: DNA barcoding reveals cryptic species in the neotropical skipper butterfly <i>Astrapes fulgerator</i> . <i>Proceedings of the National Academy of Sciences of the United States of America</i> 101 (41): 14812-14817 Wahlberg, N., West Wheat, C. (2008).

## Databases: Public Data Portal

The BOLD Public Data Portal is a publicly accessible database of all of the public sequences on BOLD, including those in the early data release phase of the iBOL project. This database can be used to access and download the associated specimen data and sequences.

### Searching the Data Portal

By accessing the Public Data Portal search from the Databases link in the header of the BOLD home page, users can search the public database using taxonomy, geography (country and state/province), and institution keywords, or by using Sample ID or BOLD Process ID to find an individual record.

Users can enter any combination of keywords into the search bar. For example, searching “Lepidoptera Canada” will return all of the Lepidoptera records collected in Canada. Searching “Lepidoptera Canada -Ontario” will return the same results with the specimens collected in Ontario removed.

For further details and examples for using the search functionality, see the search help section that is available by clicking on the help button to the right of the search bar.

The search results will display a list of BINs or records based on the options selected. For more information on BIN pages, please see the next page. Clicking on the “Record List” will convert the result list to public records matched only and clicking on “BIN List” will convert the list to all BINs available.

### Specimen Record

The record page gives information on the specimen identifier, taxonomy, specimen details, collection data (including collection site), sequence information, specimen image details, and attribution details. The image to the left shows the details page for a particular record.

A record page will reference a BIN when one is available and provides links to GenBank records.

**BOLDSYSTEMS** Databases | Taxonomy | Identification | Workbench | Resources

### Data Portal- BIN List

Search Chordata Go  
Matched Terms: +Chordata[phylum],  
Record List Bin List Specimen Data Sequences Fasta|Trace XML|TSV XML|TSV Combined Show Help

Showing Records 501 to 550 Page <<prev 7 8 9 10 11 12 13 14 15 16 next>> Records Per Page 50 ▾

**BOLD:AAA7123 [Members: 88]**  
Order: Gadiformes (88)  
Species: Macrourus whittoni (33); Macrourus carinatus (19); Macrourus sp. Smith et al. 2011 (18); Macrourus holotrachys (16); (1); Macrourus berglax (1); Distribution: Antarctica (50); Australia (20); New Zealand (10); Argentina (6); (1); Atlantic Ocean (1)

**BOLD:AAA7176 [Members: 51]**  
Order: Cypriniformes (51)  
Species: Cyprinus carpio (39); (12)  
Distribution: (1); Czech Republic (10); Canada (9); Greece (9); Germany (6); South Africa (5); Australia (1)

**BOLD:AAA7195 [Members: 34]**  
Order: Perciformes (34)  
Species: Bathygobius soporator (3)  
Distribution: Belize (2); United States (1)

**BOLD:AAA7196 [Members: 34]**  
Order: Perciformes (34)  
Species: Bathygobius antennalis (24); (10)  
Distribution: Belize (15); Mexico (10); Netherlands Antilles (6); Bahamas (3)

**BOLD:AAA7198 [Members: 11]**

Found 93196 published records, representing 9796 species, forming 12449 BINs (clusters), with specimens from 172 countries, deposited in 244 institutions.

specimen distribution:

Public Data Portal - Retrieved BIN List

**BOLDSYSTEMS** Databases | Taxonomy | Identification | Workbench | Resources

### Data Portal- Barcode Record

Search Go  
Record Details For ASCMT071-011

IDENTIFIERS	Sample ID: BIOUG00818-B02	Museum ID: MAS615-10	Collection Code: BIOUG00818-B02
	Deposited In: Biodiversity Institute of Ontario		
TAXONOMY	Phylum: Arthropoda	Subfamily: Scolopendrinae	
	Class: Insecta	Genus: Cryptopeltis	
	Order: Coleoptera	Species: C. cincta	
	Family: Dytiscidae	BIN(Cluster ID): BOLD:AAC3208	

SPECIMEN DETAILS

Voucher Type: Tissue Type: Brief Note: Detailed Notes:	Reproduction: sexual
	Sex: Male
	Life Stage: adult

COLLECTION DATA

Country: Canada	Date Collected: 2010-07-25
State/Province: Ontario	Collector: Alex Smith
Region/Country: Waterloo Township, near Douglas Ontario	
Sector: Exact Site: Latitude: 45.599	Elevation: Elevation Accuracy:
Latitude: 45.599	Longitude: -76.676
Longitude: -76.676	Depth: Depth Accuracy:
Depth: 658 bp	Depth Accuracy: Depth Accuracy:

SEQUENCE COI-SP [Funding Source: iBOL:WG1.6]  
Sequence ID: ASCMT071-11-COI-SP GenBank Accession: JN288194  
Last Updated: 2011-11-14 Genome: Mitochondrial  
Locus: Cytochrome Oxidase Subunit 1 5' Region  
Number of Bases: 658 bp  
Sequence: (long sequence of DNA bases)

Amino Acid: (long sequence of amino acids)

Illustrative Barcode: (colorful barcode)

ELECTROPHOREGRAM TRACE FILES: Length: 261 Pcr Primers: 344 Sea Primer: Read: Status: Run Date: 2011-07-25 11:45:15

Attribution: Specimen Depositor: Biodiversity Institute of Ontario  
Specimen Depositor: Spencer Walker, Biodiversity Institute of Ontario  
Photographer: Alex Smith  
Collector: Specimen Identification: Funding Source: iBOL:WG1.6

Data Portal - Specimen Record

## Databases: Barcode Index Numbers (BINs)

The Barcode Index Number System is an online framework that clusters barcode sequences algorithmically, generating a web page for each cluster. Since clusters show high concordance with species, this system can be used to verify species identifications as well as document diversity when taxonomic information is lacking.

This system consists of two parts:

- A clustering algorithm employing graph theoretic methods to generate operational taxonomic units (OTUs) and putative species from sequence data without prior taxonomic information,
- A curated registry of barcode clusters integrated with an online database of specimen and taxonomic data with support for community annotations.

Please see the below for information on the elements of each BIN page, and the next page for further details on BINs.

BIN pages display aggregated data in several sections:

Distance Statistics	View BIN details including the member count, average distance between members, and BIN criteria. Also, details of the nearest neighbour BIN, including BIN GUID, taxonomy, and member count are provided
Taxonomy	The taxonomy of the public data is visible for the BIN, with highlighting to indicate taxonomy concordance and discordance.
Collection Locations	List of the collection countries and number of specimens collected per country.
Associated Publications	Details of publications that used sequences contained in the BIN.
Users Responsible for Data	Lists the owners of the public and private sequences contained within a BIN.
Dendrogram of Sequences	For BINS with a smaller number of sequences, a circle tree is displayed, and including the nearest neighbour. A PDF version of the tree is available for all BINs.
Specimen Images	View images for associated records.
Sampling Sites	Displays a map of the collection sites based on GPS values.
Attribution	Lists institutions where specimens are deposited, sequencing centers, photographers, collectors, taxonomists and funding sources.

Descriptions of elements in BIN pages.

**BOLD SYSTEMS** Databases | Taxonomy | Identification | Workbench | Resources

Data Portal - BIN Page

Barcode Index Number Registry For BOLD:AAC0510

**BIN DETAILS**

BIN GUID	BOLD:AAC0510
Member Count	21 [19 Public]
Average Distance	1.65
Criteria	NN + 2.2 edge - COX1
Maximum Distance	3.21
Distance Variance	157.52
Distance to Nearest Neighbor	2.25

**NEAREST NEIGHBOR BIN DETAIL**

Nearest Bin GUID	BOLD:AAC0510
Member Count	21
Nearest Member	ABRMF048-06
Nearest Member	Choristia Adinopterygi Synbranchiformes Synbranchidae Hippocampus Hippocampus kuda

**TAXONOMY**

Phylum	Chondriota [2]
Class	Actinopterygii [2]
Order	Synbranchiformes [20]
Family	Synbranchidae [20]
Subfamily	Hippocampinae [20]
Genus	Hippocampus [10]
Species	Hippocampus kuda [14] Hippocampus reidi [3] Hippocampus capensis [2] Hippocampus histrix [1]

**Comments:** 0  
**Tags:** 0  
**Show Annotations**

**COLLECTION LOCATION**

**Countries:**

- Mexico - [1]
- India - [1]
- South Africa - [5]
- Virgin Islands of the United States - [1]
- China - [1]
- Belize - [1]
- Indonesia - [2]
- Papua New Guinea - [1]
- Vietnam - [4]

**PUBLICATIONS:**

Kawakita R, Ryvola Mira-Masaki, Matsubuchi K, Kohji, Lavery S, Saitoh, Inoue JC, Jun G, Saitoh TP, Takatsu P, Kawaguchi A, Akira, Nishida M, Mutsumi Relationships of the 11 pectoral fin families (sticklebacks, pipefishes, and their relatives): a new perspective based on whole mitogenome sequences from 75 higher teleosts. Mol. Phylogenet. Evol. 2009;01:91-101(122-35) DOI:10.1016/j.ympev.2008.09.020

**DATA MANAGERS:**

**Basic Data:**

Alex V. Borisenko - [7]  
Dirk Steinkne - [6]  
Tara Ziemer - [4]  
Lourdes Vasquez-Yeomans - [1]

**Private Data:**

Tara Ziemer - [1]  
Amy C. Driskell - [1]  
Benjamin Victor - [1]

**TREE RECONSTRUCTION OF BIN & NEAREST NEIGHBOR:**

[PDF tree \(All members and a member of the nearest BIN\)](#)

**Attribution:**

Biodiversity Institute of Ontario - [5]  
El Colegio de la Frontera Sur, Chetumal - [1]  
McGill University, Redpath Museum - [7]  
Museum of Ontario, NCCB - [2]

**Specimen Images:**

HLC-12284

**BIN summary page**

The BIN framework can greatly expedite the evaluation and annotation of described species and putative new ones while reducing the need to generate interim names, a non-trivial issue in barcoding datasets. The BIN algorithm has been effectively tested on a broad set of taxonomic groups and shows potential for applications in species abundance studies and environmental barcoding. The registry employs modern GUID and web service functionality enabling integration with other databases.

BIN pages support community vetting through annotation of individual data elements. Please see the below for details on annotation.

## Annotation

As the volume of barcode data being generated increases rapidly, the need for routine curation has become apparent. BOLD's annotation and notification system supports rapid community based validation of barcode data.

Annotation can occur at the project level, record level and also on specific data elements including taxonomy, images and sequences on BIN pages.

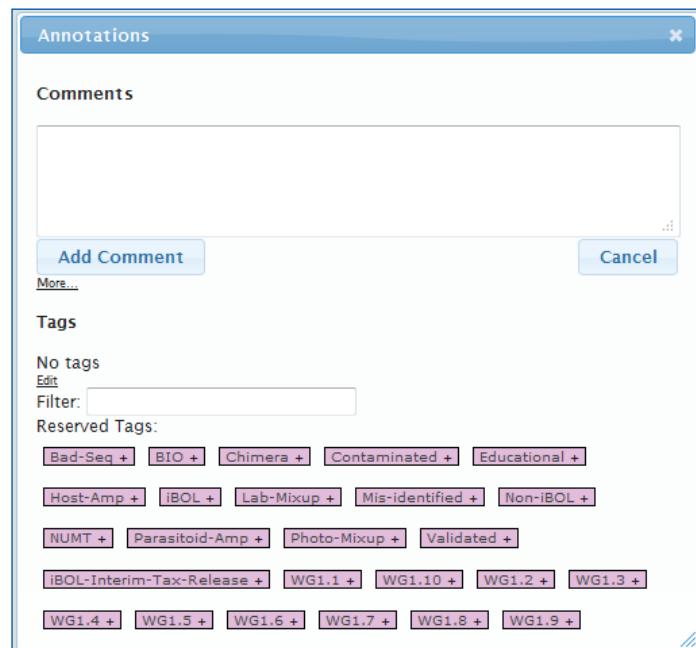
Comments leverage the large user-base and expert knowledge for curation of both private data within collaborative projects and public data through the taxonomy browser and BOLD Public Data Portal.

Crowd sourced comments and tagging of public data utilize the expertise of the taxonomic community.

Tagging allows for categorization using custom and controlled tags.

Both custom and controlled tags can be used for filters, searches, and workflow management.

Comments and tags applied to your data by other users will appear in your activity feed on your User Console and the activity feed on the appropriate Project Console.



Annotation pop-up window

## Specimen Data and Sequence Data Pages

**BOLD** connects specimen data with sequence data in a biphasic record. Please see below for what each part consists of, as well as how to navigate through the pages.

## Specimen Data

The Specimen page stores voucher details, taxonomy, specimen details and collection data for a specimen. Any user with specimen editing permissions can edit the records by selecting "Edit Specimen" from the upper right corner. Selecting the species name will open the appropriate taxonomy browser page for that taxon. There is a world map marked with the location where the specimen was collected. By clicking on the map, you can access the distribution mapping options for that species. The images for the specimen are located at the bottom of the window. By clicking on the images, you will access a larger copy of each image.

# BOLD SYSTEMS

## Arthropods of the Kenai National Wildlife Refuge [KNWRA]

**SPECIMEN IDENTIFIERS**

Sample ID:	KNWR 6776	Museum ID:	
Isolate/Field Num:	KNWR 6776	Collection Code:	
Deposited In:	Kenai National Wildlife Refuge		

**TAXONOMY**

Identifier:		Voucher Type:	
Phylum:	Arthropoda	Tissue Type:	
Class:	Insecta	Extra Info:	
Order:	Diptera	Sex:	
Family:	Syphidae	Reproduction:	
Subfamily:	Syrphinae	Life Stage:	A
Genus:	Parasyphus	Note:	Refuge Voucher
Species:	<a href="#">Parasyphus relitus</a>	Associated Taxa:	Collections   Identifier: Bowser
Taxonomy Note:			

**COLLECTION DATA**

Collectors:	Bowser
Collection Event ID:	
Date Collected:	13-Sep-2010
Date Accuracy:	
Time Collected:	
Country:	United States
State/Province:	Alaska
Region/Country:	Kenai Peninsula Borough
Sector:	Refuge Headquarters
Exact Site:	Ski Hill Road.
Site Code:	
Habitat:	
Sampling Protocol:	
Latitude:	60.465
Longitude:	-151.073
Coord. Source:	
Coord. Accuracy:	
Elevation:	
Elevation Accuracy:	
Depth:	
Depth Accuracy:	
Notes:	

**SPECIMEN DATA ANNOTATIONS**

Comments: 0  
Tags: 0  
[Show Annotations](#)

**PHOTOGRAPHS**



**Caption:** [Dorsal]  
License: CreativeCommons - Attribution Non-Commercial Share-Alike (2011)  
Comments: 0  
Tags: 0  
[Show Annotations](#)

**Caption:** [Lateral]  
License: CreativeCommons - Attribution Non-Commercial Share-Alike (2011)  
Comments: 0  
Tags: 0  
[Show Annotations](#)

## Sequence Data

The Sequence page stores details about the sequence data for a specimen. Different markers can be selected by clicking on their links in the marker section. Trace files can be viewed or downloaded from this window. Primers used can be viewed by clicking on the primer codes. Sequences can also be deleted by users with full sequence access. If desired, the ID engine can be used to identify the sequence from this page.

An illustrative barcode sequence of the species is provided by BOLD, along with a link to the Laboratory Information Management System (LIMS) for the Canadian Centre for DNA Barcoding when available.

## Sequence Data Page

## Workbench: Specimen Data Submission

The data submission is how records are created on BOLD. Each record is assigned a BOLD Process ID when uploaded. Images, traces, and sequences can then be uploaded with the Process IDs. There are two ways to enter records onto BOLD: manually or with bulk uploads through the BOLD Data Managers.

This protocol assists in the submission of bulk data to BOLD through the BOLD Data Managers. This is the easiest way to populate your project with records, as well as the only way to enter new species taxonomy into the BOLD library. Described below is the necessary format of the data that is required for a correct submission.

To submit a batch data submission to the BOLD Data Managers, log into BOLD and select “Specimen Data” under the uploads menu. Click on batch submission and select which project you would like to submit the data to. Upload the excel spreadsheet described below and submit. This will forward the submission to the data management team along with your contact details. You will receive a confirmation email with a reference number for the submission. The data management team will let you know when the submission has been uploaded to BOLD.

The data spreadsheet consists of 4 worksheets; a main specimen identifier worksheet (voucher info) that is linked to three other worksheets: taxonomy, specimen details, and collection data. (Refer to Tables I through 4 for field definitions.) See the next page for more instructions.

\* Minimum required fields for new records.

Sample ID *	Identifier for sample being sequenced, often identical to the Field ID or Museum ID. Sample identifiers are extended when tissue is sub-sampled for secondary analysis. It is important to use a unique and original format for the Sample IDs. If the Sample IDs provided are not original to BOLD, they will need to be changed before the data can go online. Only the following characters may be used in the Sample ID, Field ID, and Museum ID: Numbers, letters, and ^ .:- _ () # All other characters will be removed.
Field ID *	Identifier for specimen assigned in the field. Specimens in personal collections will continue to use this as the primary identifier for the specimen. (This field or the Museum ID field must be filled in)
Museum ID *	Identifier for specimen assigned by formal collection upon accessioning, also referred to as the catalog number. This identifier should be made unique by adding scope of the collection or institution. This is done by following a triplet format <Institution acronym>:<collection code>:catalog number in the case of a museum collection and in the case of a personal collection, Personal:Name of collector:fieldID. (This field or the Field ID field must be filled in)
Collection Code	Code associated with a given collection within an institution and is only applicable when Museum ID is provided. Often used as scope resolution within museums.
Institution Storing *	Full name of the institution handling storage of the specimen or tissue sample. Can be a personal collection.

Table I: Field definitions for Voucher (Specimen) info page.

Sex	Male/female/hermaphrodite <b>only</b> .
Reproduction	Sexual/asexual/cyclic parthenogen <b>only</b> .
Life Stage	Adult/immature <b>only</b> .
Extra Info	User Specified Characteristics (free text) - Can be displayed on a tree or used to sort records. Limited to a maximum of 50 characters. Designate FAO region here.
Notes	Free text or XML tagged text. All XML text should be surrounded by the XML start (<xml>) and stop (</xml>) tags.

Table 2: Field definitions for Specimen Details page.

Collectors	Comma delimited list of collectors.
Collection Date	Date of collection, must be in DD-MMM-YYYY format (e.g. 01-Jan-2005).
Continent/Ocean	ISO Continents and Oceans.
Country *	ISO Countries.
State/Province	States and provinces (according to Getty Geographical Thesaurus).
Region	Park, county, district, lake or river.
Sector	Sector of park or county/city.
Exact Site	Description of collection location.
GPS Coordinates	Latitude & Longitude in “degrees.decimal degrees” format (e.g. 45.837).
Elevation/Depth	Elevation or depth in metres <b>only</b> .

Table 4: Field definitions for Collection Data page.

Full Taxonomy	Full taxonomy consisting of phylum*, class, order, family, subfamily (optional), genus, species in binomial format.
Identifier	Full name of primary individual responsible for providing taxonomic identification of the specimen.
Identifier E-mail	E-mail address of the primary identifier.
Identifier Institution	Institution of the identifier.

Table 3: Field definitions for Taxonomy page on accompanying spreadsheet.

### New Fields

There are a number of new specimen data fields in BOLD 3.0. These include: Collection Event ID, Site Code, Sampling Protocol, Event Time, Date Precision, Habitat, External URLs, Associated Taxa, Associated Specimens, Species Reference and Identification Method. These fields will be available on the spreadsheet template by the end of 2011. The fields are currently available for individual record uploads.

All of the data in BOLD are organized by projects. Related projects can be grouped into containers or temporarily merged with related projects for analysis, etc. An individual entry in the database represents a barcode of a given specimen. The Process ID (assigned by BOLD upon specimen data record upload) uniquely represents a sample in BOLD. This is the identifier that is used to track a sample through the barcoding process: collection, taxonomic identification, sequencing, analysis and final publication of data.

Specimen data can be entered in one of two ways. As outlined here, for larger sets of samples, the data can be entered on the Data Submission Template spreadsheet and sent to BOLD. Data managers will review the data, to ensure that it meets the minimum requirements, and input it into BOLD. For smaller numbers of entries, (ie: 1-10 records) users can enter sample data directly through the website by clicking on “Specimen Data” under the Uploads menu and using the manual interface there.

Here is an example of a properly filled in data submission.

You can get this blank template online from <http://www.boldsystems.org/docs/SpecimenData.xls>

Use the tabs at the bottom of the Excel workbook to navigate through the four pages.

Specimen Info				
Sample ID	Field ID	Museum voucher ID	Collection Code	Institution Storing
Sample-demo01	Sample-demo01			Biodiversity Institute of Ontario
Sample-demo02	Sample-demo02	15466-JUC-ISC	ISC	Burke Museum
Sample-demo03	Sample-demo03			Smithsonian Institution

Figure 1: Example data for Specimen Info

Taxonomy										
Sample ID	Phylum	Class	Order	Family	Subfamily	Genus	Species	Identifier	Identifier Email	Identifier Institution
Sample-demo01	Arthropoda	Insecta	Diptera	Asilidae	H y d r o - psychinae	Efferia	Efferia aestuans	Joe Smith	jsmith@BIO.org	Oxford
Sample-demo02	Arthropoda	Insecta	Diptera	Asilidae		Asilus		Joe Smith	jsmith@BIO.org	Oxford
Sample-demo03	Arthropoda	Insecta	Diptera					Joe Smith	jsmith@BIO.org	Oxford

Figure 2: Example data for Taxonomy

Specimen Details										
Sample ID	Sex	Reproduction	Life Stage	Extra Info	Notes					
Sample-demo01	Female	Sexual	Adult		Commonly called 'Robber Fly'					
Sample-demo02	Male	Sexual	Adult	feeding on fruit						
Sample-demo03	Male	Sexual	Adult							

Figure 3: Example data for Specimen Details

Collection Info											
Sample ID	Collectors	Collection Date	Continent / Ocean	Country	State / Province	Region	Sector	Exact Site	Latitude	Longitude	Elevation
Sample-demo01	Joe Smith	2-Feb-2009	North America	Canada	Ontario	Wellington	Guelph	Riverside Park	43.563	-80.270	325
Sample-demo02	Joe Smith	27-Jul-2007	Asia	Japan	Hokkaido	Soya	Omura		44.671	142.788	95
Sample-demo03	Joe Smith	5-Apr-2007	Central America	Costa Rica	Guanacaste	ACG	Mundo Neuvo	Ricon de la Vieja	10.772	-85.434	305

Figure 4: Example data for Collection Info

## Specimen Data Submissions Continued

There are two types of submissions: “New Submission” and “Update”.

A new submission occurs every time new records are added to a project. An update means to modify records that already exist in a project. If you wish to only update one or two records, please manually select the specimen from the species record listing in your project and click on the “edit” button in the upper right corner. Any details can be edited in this way, except for adding new taxonomy to BOLD. If there is new taxonomy to add to the BOLD library this should be sent by email as a taxonomy update to the BOLD Data Management Team through [submissions@boldsystems.org](mailto:submissions@boldsystems.org).

### New Submissions

New submissions are project specific so that their data can be associated with a project on BOLD. If records are submitted that need to be entered into different projects on BOLD, a separate file for each project needs to be sent. Provide as much detail and additional information as possible with a new submission so that it will take less time later to update the blanks.

The minimal requirements for a new submission on BOLD are:

- Voucher Info Page - Sample ID
- Voucher Info Page - Field ID and/or Museum voucher ID
- Voucher Info Page - Institution Storing
- Taxonomy Page - Phylum
- Collection Page - Country

### Other useful information:

#### Sample IDs:

- It is important to use a unique and original format for the Sample IDs. If the Sample IDs provided are not original to BOLD, they will need to be changed before the data can go online.
- Only the following characters may be used in the Sample ID, Field ID, and Museum ID: Numbers, letters, and ^ .:- \_ () # All other characters will be removed.

#### Specimen Details:

- If the sex, reproduction or life stage values for your specimen do not fit the accepted values for Specimen Details, please move the information to the Extra Info or Notes fields.
- Remember that the “Extra Info” field can be displayed on a Taxon ID Tree on BOLD and thus it is best to enter data there that may help when analyzing the data on a tree.

#### Identifiers/Donors:

- In the case where the donor or identifier is deceased or retired, please make note of that in the email field. It is important to provide this information so we can keep the database up-to-date.

### Updating Specimen Data

The quickest way to update records in bulk is to download the Data Spreadsheet from BOLD containing the records that need to be modified. To do so, click on “Data Spreadsheets” from the Downloads menu on the upper left side of your project. Only download the worksheets and records that will be affected by the update (e.g. if the taxonomy needs to be updated, only download the Taxonomy worksheet; if specimen details and collection data need to be updated, only download the Specimen Details and Collection Data worksheets, etc. Please do not download and submit updates on the progress report.)

Once the worksheets are downloaded, modify the data and copy it into the standard submission spreadsheet. The submitted update must reflect what the data should be on BOLD. Please send this to the Data Management Team through [submissions@boldsystems.org](mailto:submissions@boldsystems.org).

**NOTE:** Any fields left empty will be considered blank and will be removed from BOLD during an update. Do not remove any data from the update sheet if you’d like it to stay on BOLD. The system cannot distinguish between “blank: do not update this field” or “blank: delete the content of this field”.

Updates to Voucher Info are slightly different from updates to Taxonomy, Specimen Details, and Collection Data.

#### a.) Updates to Voucher Info

Identical to new submissions, updates to the voucher info are project specific. The records need to be split into their corresponding project.

#### b.) Updates to Taxonomy, Specimen Details, and Collection Data

Updates to these sections are project independent. Records from any number of projects can be updated in one submission spreadsheet, and the number of records are (in theory) infinite for this type of update.

To select records from more than one project for a Taxonomy, Specimen Details or Collection Data update, you can use the search function, or merge projects in your project list.

## Workbench: Image Submission

Images should be uploaded to BOLD to complete a specimen record. An image provides support for identifications and makes comparisons easier between species.

This protocol outlines the image submission process for BOLD. It describes the necessary format of the images and the ancillary data and the steps required to build the uploadable package required for a successful submission.

### I. Collect Images:

**Group Images:** Group high-quality images of specimens in .jpg format for your project. BOLD accepts high resolution images (up to 20 megapixels), but only displays a greatly reduced thumbnail. The high resolution image is archived but will not be distributed without the submitter's explicit consent. Refer to page 18 for a guide on picture orientation and quality.

### 2. Assemble Package:

The image submission package should consist of all .jpg format images and a spreadsheet with the file names and ancillary data. Make sure that all images in the package are accounted for in the spreadsheet. When submitting more than one image per specimen simply copy the 'Sample ID' and 'Process ID' to the next line with the file name of the consecutive image. You can upload up to 10 images per specimen, depending on organism characteristics. Please photograph several different orientations if needed.

The submission spreadsheet should be named ImageData.xls and contain the columns described in Table 9-1.

#### Steps:

**A.** Fill in the ImageData.xls spreadsheet with all the data related to the images in the submission package. To easily create the list of image files in a folder, open a terminal window (Start > Run > cmd in Windows), navigate to the folder containing the image files, and run one of the following commands:

- Windows                    `dir /b *.jpg>list.txt`
- MacOS                    `ls *.jpg*JPG>list.txt`
- Linux/Unix                `ls *.jpg*JPG>list.txt`

These commands will generate a list of all the files in the current folder and save it in a document called 'list.txt' that will appear in the current folder. You can then open list.txt and move the data into the Image File column.

Obtain the Process IDs by clicking on "Data Spreadsheets" under the Downloads menu on the left side of a project console. Download the Progress Report to get the Process IDs that are assigned to each Sample ID.

Image File *	Complete (incl. extension) and identical file name (case sensitive) of images.
Original Specimen *	Enter "Yes" if the image shows the actual specimen for this record. Otherwise enter "No".
View Metadata *	Controlled vocabulary term to group media depicting a specific set of features of the organism or related environment. Dorsal, Lateral, Ventral, Frontal, etc.
Caption	Free text description of the subject. Short descriptions are recommended, such as: part of organism photographed, life stage, sex, etc. (400 Characters)
Measurement	Any single relevant measurement that was taken in metric units.
Measurement Type	Item or feature that was measured.
Sample ID *	Sample ID for record, which must match Sample ID in BOLD.
Process ID *	Process ID for record, which must match Process ID in BOLD.
License Holder*	The primary individual holder of the license. This is less critical when using creative commons licenses.
License*	Pick one of the following license types or short-forms: • Copyright (c) • No Rights Reserved (nrr) • CreativeCommons - Attribution (by) • CreativeCommons - Attribution Share-Alike (by-sa) • CreativeCommons - Attribution No Derivatives (by-nd) • CreativeCommons - Attribution Non-Commercial (by-nc) • CreativeCommons - Attribution Non-Commercial Share-Alike (by-nc-sa) • CreativeCommons - Attribution Non-Commercial No Derivatives (by-nc-nd)
License Year*	The year of license declaration (not the year of submission to BOLD).
Licensee Institution*	The primary license holder's institutional or organizational affiliation. Decisions regarding use of material falls to the institution when the individual is unreachable or unresponsive.
License Contact*	Contact information for the license holder. Can be an email address, mailing address, phone number, or all of the above.
Photographer	The individual or team responsible for photographing and editing the media prior to submission.

Field definitions for accompanying image submission spreadsheet.

#### \* Required Fields

**B.** These two components (Image files and Spreadsheet) need to be placed in a single folder. Compress them all into a single file before submitting. The following free tools are available to provide this functionality, however, most modern operating systems have built-in functionality for zipping:

- » WinZip - <http://www.winzip.com>
- » WinRar - <http://www.rarsoft.com>
- » MacZipIt - <http://www.maczipit.com>

**C.** BOLD will accept a maximum zipped file size of 190 MB. Upload the images to BOLD by clicking on the “Specimen Images” link in the Uploads menu of the desired project. Select the zipped folder of images and then hit “submit”. Do not navigate away from the project or close the pop-up window until the successful upload message is displayed.

Image File	Original Specimen	View Metadata	Caption	Measurement	Measurement Type	Sample Id	Process Id	License Holder	License	License Year	License Institution	License Contact	Photographer
ROM912-D.jpg	yes	Dorsal	skull	15 mm	skull length	ROM 10912	BMI272-03	Jane Beck	Copyright	2010	Biodiversity Institute	photo@bio.org	Joe Smith
ROM912-L.jpg	yes	Lateral	lower jaw	7 mm	length	ROM 10912	BMI272-03	Jane Beck	Copyright	2010	Biodiversity Institute	photo@bio.org	Joe Smith
ROM913-L2.jpg	yes	Lateral	skull	15 mm	skull length	ROM 10913	BMI273-03	Jane Beck	Copyright	2010	Biodiversity Institute	photo@bio.org	Joe Smith
ROM913-V.jpg	yes	Ventral	skull	15 mm	skull length	ROM 10913	BMI273-03	Jane Beck	Copyright	2010	Biodiversity Institute	photo@bio.org	Joe Smith

Image Submission Spreadsheet (ImageData.xls) completed with sample data.

## Tips and Troubleshooting for Image Uploads

- Zipped files must be under 190MB in size. If the upload fails to initialize, the zipped file may be too large. Break it into more than one upload, each with its own spreadsheet.
- The spreadsheet cannot contain any formulas.
- If the upload program cannot find the image files, it is possibly because it can not read the names. Make sure that the spreadsheet contains text values only.
- Full filenames must be used in the spreadsheet. The extension (.jpg) must be included in the image file name which is case sensitive.
- Spreadsheet must be named ImageData.xls. If the upload program can not find the spreadsheet, confirm that it is named correctly (case sensitive).
- Verify that the data length in the free text fields and make adjustments if necessary
- Data must start on the second line of the spreadsheet. There is only one line for the column headers.
- Adding extra columns to the sheet will cause errors.
- Image names can not contain the characters “&” or “-”. Please rename your images so that they do not have these characters.

You can upload more images in separate batches to any record at any time. If you wish to delete images for a record, please contact the BOLD Support Team through support@boldsystems.org.

## Image Licensing and Use

BOLD assumes no license for images uploaded to the site. The image owner maintains the license and may change the license from the images at any time. Revisions to the given license should be to make the license more liberal over time as it is very difficult to retract an open license and make it more restrictive.

If no license is chosen for an image, by default BOLD will forward all requests for that image to the owner for response. Adding a license reduces that burden and makes access easier. BOLD encourages the use of CreativeCommons - Attribution Non-Commercial Share-Alike, as this license has a good balance of protection and access.

There are three reasons for having images uploaded for each specimen on the BOLD database.

1. Quality Assurance - images can be used to confirm the taxonomic identification of organisms during sequence analysis.
2. Peer Review/Quality Assurance - once records are made public, peers can utilize your images and sequences to assist in their own validation of related specimens.
3. Taxonomy Brower Taxon Profiles - a random selection of the images on BOLD for each taxon are displayed on the public Taxonomy Brower at a highly reduced size (320 x 240) to create an online profile for each taxon that is stored in BOLD.

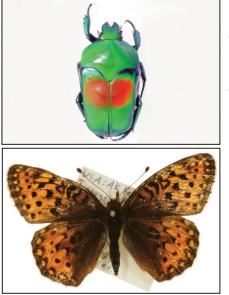
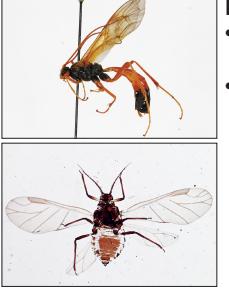
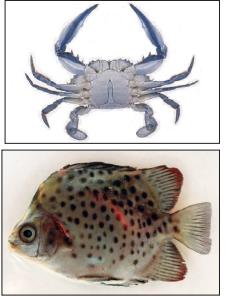
## Photography Guide

This guide has been developed with assistance from the Canadian Centre for DNA Barcoding in an effort to provide best practices for taking voucher photographs. The BOLD database can accept up to ten images per specimen, so besides photos of a mounted or live organism, feel free to photograph and submit photos of distinguishing features or habitat.

To provide the best specimen image for viewing on the web, the following guidelines should be adhered to when appropriate:

- Please take pictures using the high quality mode on your camera (please note that BOLD can accept up to 20mp photos).
- The specimen should be centered in the image frame.
- Photos should be taken as close-up to the specimen as possible, leaving very little gap around the edges.
- It is often beneficial to include a measurement scale in the image to provide a size reference or a colour scale to provide colour reference.
- Background should be a contrasting colour where possible
- Standardizing the aspect ratio during specimen photography for your project(s) will make your images easier to compare
- Standardizing the position/orientation of each specimen will make it much easier to compare specimens across a project or BOLD.

See below for some common standardized orientations for animals, plants and fungi.

	<b>Dorsal</b> <ul style="list-style-type: none"> <li>• The anterior of the specimen should be facing the top of the image frame.</li> <li>• The specimen should be face-down, with the dorsal aspect of the head visible.</li> </ul>		<b>Herbarium Sheets:</b> <ul style="list-style-type: none"> <li>• The full sheet should be included in the frame</li> <li>• The sheet should be oriented so that text is legible in photo</li> <li>• Label as "Herbarium Specimen" in the view metadata</li> </ul>
	<b>Lateral</b> <ul style="list-style-type: none"> <li>• The anterior of the specimen should be facing the left side of the image frame.</li> <li>• The specimen should be oriented with the feet towards the bottom of the image.</li> </ul>		<b>Specific Features:</b> <ul style="list-style-type: none"> <li>• These are often live photographs with focus on distinguishing characteristics</li> <li>• For plants, these may include opened fruit, adaxial veins, male and female components</li> <li>• Macro focus is recommended if available</li> <li>• Label as "Leaf", "Fruit", "Buds", "Bark", etc</li> </ul>
	<b>Ventral</b> <ul style="list-style-type: none"> <li>• The anterior of the specimen should be facing the top of the image frame.</li> <li>• The specimen should be face-up, with the ventral aspect of the head visible.</li> </ul>		<b>Whole Specimens:</b> <ul style="list-style-type: none"> <li>• These are often live photographs of the organism in its natural location</li> <li>• The specimen should be centered in the frame and provide information such as shape of plant, etc</li> <li>• If shooting outside, ideal environment has no wind and is slightly overcast (to avoid overexposure)</li> <li>• Label as "in situ", "Plant", etc</li> </ul>

Suggested formats for photographs.

When entering a new orientation, please capitalize only the first letter and do not add any words (such as "view") to the end. If your specimen does not fit into any of these categories, then please feel free to create a new category of view. (Displayed Specimen Images: All Rights Reserved)

## Workbench: Trace Submission

Trace files provide support for sequences and should be uploaded for every specimen record. They can be uploaded once the data submission step is completed and BOLD has assigned a Process ID to each record.

This protocol assists in the submission of trace files to BOLD. It describes the necessary format of the files and the ancillary data that is required for the correct submission.

### I. Confirm primers are registered on BOLD.

See page 8 for details on how to search the primer database.

### 2. Assemble Package:

The submission package consists of trace files (.ab1 or .scf), corresponding phred (score) files if available (.phd.l) and a spreadsheet with the file names and ancillary data. The submission spreadsheet should be named data.xls and contain the columns described in the table to the right.

#### Trace File Upload Steps:

##### A. Fill in the data.xls sheet with all the data about your files.

To easily create the list of the files in a folder, you need to open a terminal window (Start > Run > cmd in Windows), navigate to the folder where the trace and score files have been placed and run one set of the following commands:

- Windows    dir /b \*.ab1>ab1.txt and    dir /b \*.phd.l >phd.txt
- MacOS      ls \*.ab1>ab1.txt    and    ls \*.phd.l > phd.txt
- Linux/Unix    ls \*.ab1>ab1.txt    and    ls \*.phd.l > phd.txt

These commands will generate lists of all the files in the current folder. They will be saved as ab1.txt and phd.txt text files. You can then open the text files and move the data into the appropriate columns.

Obtain the Process IDs by clicking on “Data Spreadsheets” under the Downloads menu on the left side of a project console. Download the Core Lab Book to get the Process IDs that are assigned to each Sample ID.

**B.** These components (Trace files, Score files and Spreadsheet) need to be placed in a single folder. Compress them all into a single zipped file before submitting.

**C.** BOLD will accept a maximum file size of 190MB. Upload the traces to BOLD by clicking on the link “Trace Files” in the Uploads panel of the desired project. Select the zipped folder of files, the run site and then hit “submit”. Do not navigate away from the project or close the pop-up window until the successful upload message is displayed.

You can upload more traces in separate batches to any record at any time. If you wish to delete any traces for a record, please contact the BOLD Support Team through support@boldsystems.org.

Trace File *	Complete (including extension) and identical file name (case sensitive).
Score File	Complete (including extension) and identical file name (case sensitive).
PCR Primers Fwd/Rev *	Primer codes are case sensitive. Both must be filled in.
Sequence Primer	Primer codes are case sensitive.
Read Direction *	Forward or Reverse.
Process ID *	Process ID of record, which must match Process ID in BOLD.
Marker  (2 blank columns must be left after the Process ID column)	If sequencing multiple genes, the marker needs to be filled in to match the short form marker code in your project, such as one of the following: • COI-5P • ITS • rbcLa • matK

Field definitions for accompanying trace submission spreadsheet.

#### \* Required Fields

Trace File	Score File	PCR Fwd	PCR Rev	Seq Primer	Read Direction	Process ID	blank	blank	Marker
KKBNA001-04.ab1	KKBNA001-04.phd.l	BirdF1	BirdR1	BirdF1	Forward	KKBNA001-04			COI-5P
KKBNA001-04r.ab1	KKBNA001-04r.phd.l	BirdF1	BirdR1	BirdR1	Reverse	KKBNA001-04			COI-5P
KKBNA002-04.ab1	KKBNA002-04.phd.l	BirdF1	BirdR1	BirdF1	Forward	KKBNA002-04			COI-5P
KKBNA002-04r.ab1	KKBNA002-04r.phd.l	BirdF1	BirdR1	BirdR1	Reverse	KKBNA002-04			COI-5P

Trace File Submission Spreadsheet (data.xls) completed with sample data.

## Tips and Troubleshooting For Trace Uploads

- Primers must be registered before upload. If the primers are not registered, there will be an error. Please refer to the section below for details on how to register new primers.
- Zipped file must be under 190MB in size. If the upload fails to initialize, it is probably because the zipped file is too large. Try breaking it into more than one upload, each with its own spreadsheet.
- The spreadsheet cannot contain any formulas.
- If the upload program can not find the files, it is possibly because it can not read the names. Make sure that you have text values only in the spreadsheet.
- Full filenames must be used in spreadsheet. The extension (.abl, .scf, .phd.l) must be included in the file name. These extensions are case sensitive.
- The spreadsheet must be named data.xls. If the upload program can not find the spreadsheet, confirm that it is named correctly (case sensitive).
- Data must start on the second line of the spreadsheet. There is only one line for the column headers.
- Do not add extra columns to the spreadsheet.
- Trace files may not be downloadable from BOLD until 24 hours after they have been submitted.

## Workbench: Sequence Submission

This protocol outlines the DNA sequence submission process on BOLD. It describes the sequence format and steps required for a successful submission. Only users with sequence editing access on a project may upload sequences.

### I. Assemble Package:

The sequence submission package should consist of aligned sequences in FASTA format referenced by BOLD Process IDs or your Sample IDs.

To upload with Process IDs, the FASTA header line must conform to the following format: it should begin with a ‘>’ followed by the Process ID, with any additional information separated by either a bar (‘|’), an underscore (‘\_’) or a space (‘ ’). There can be no spaces before the end of the Process ID.

To upload with Sample IDs, the FASTA header line must conform to the following format: it should begin with a ‘>’, followed by the Sample ID, with any additional information separated by a bar (‘|’). Do not use a space or an underscore to separate information from the Sample ID.

### 2. Upload Package:

You can include up to 1000 sequences into one upload. Upload the sequences to BOLD by clicking on the link “Sequences” in the Uploads menu of the desired project. Select the marker and sequencing institution. Paste the sequences into the text box.

When confirmed, click on “submit”.

- » If you wish to replace a sequence on BOLD, simply upload the new one with the same Process ID or Sample ID.
- » To delete an individual sequence, you can do so by using the Delete button within a record’s sequence data page (for more info on Sequence pages, please see page 12). Contact the BOLD support team through support@boldsystems.org for batch deletions.

Sequence Upload Window

## Workbench: Primer Submission

Be sure that your primers are registered with BOLD before assembling the submission package. To register new primers, select “Register Primers” from the Project Options menu in your project on BOLD or from the primer list page. Please note: If the primer sequence you used has already been registered under a different code, you will be provided with the registered code to be used in your submission. Primers you own can be edited at any time after they are created (e.g. to make them public).

**BOLDSYSTEMS**

**Primer Submission**

[View Primers](#)

Primer Code (3 to 12 letters)\*:

Primer Description:

Alias Codes (Comma Separated):

Target Marker\*:  Marker

Cocktail Primer:  No  Yes A cocktail is a standard combination of multiple primers using one reaction.

Primer Sequence (5' to 3')\*:

Direction\*:  Direction

Reference/Citation\*:

Notes:

Publicly Available:

\*Primer sequences should not contain spaces. Any space characters found will automatically be stripped before submission.

BOLD Primer Submission Form

Primer Code	Create a code for your primer. If the primer is already published in a manuscript, please use the code that is in press.
Primer Description	A description of what the primer is used for.
Alias Codes	Fill in any other known code names for your primer, separated by commas.
Target Marker	Select the target marker from the controlled list (e.g. ITS, COI 5', matK, etc.).
Cocktail Primer	Select “Yes” if it is a cocktail primer. This will create extra fields to add multiple sequences.
Primer Sequence	Fill in the sequence(s), 5' to 3'.
Direction	Select the direction of the sequence.
Reference/Citation	List references and/or citations.
Notes	Notes about the primer.
Publicly Available	If the primer has already been published, or if you wish to make it publicly available, this should be left public. The other option is to keep the primer private until publication.

Field definitions for Primer Submission Form.

## Workbench: Bibliography Submission

Users can add bibliographies to BOLD using the Bibliography Submission Form available in three locations:

- From the user console under Data Uploads
- From a project console under Uploads
- Within a record list page

Any user with edit sequence or edit specimen permissions to records will have the ability to submit a bibliography connected to those records as primary or secondary associations.

The primary and secondary GenBank accessions can be filled in here, separated by a line. If the user submitting this bibliography does not have edit access to the primary records on BOLD, then they won't be able to upload the linked publication.

**BOLDSYSTEMS**

**Bibliography Submission**

Article Title:

Authors\*:

Journal\*:

Year\*:

Volume\*:

Issue\*:

Pages\*:

DOI:

PubMed Central Id:

Open Access:

Abstract:

Date Published\*:

Date Revised:

Language:

Keywords:

Associated Records: (GenBank Acc. only)

Primary (Records generated by authors)  Secondary (Records cited by authors)

Confirm that primary records were generated by authors

BOLD Publication Submission Form

## Workbench: User Console

BOLD 3.0 provides a new user centric management console as the landing page when users log in. This new console allows for rapid access to frequently accessed projects, near real-time reports on activities on projects they are involved in as well as powerful search tools.

### Stats and Progress Feed

As users create projects, upload records and gain access to collaborators projects within BOLD, the user data stats and activity feed will activate. The information provided in these tables provide a way to monitor progress in projects and campaigns, as well as view recent activities by your collaborators. These activity logs are downloadable for analysis.

### Data Uploads

For users that have created projects, or have editing access on collaborators projects, there will be options or upload specimens, sequences, traces and/or images. Every user has the ability to upload primers and bibliographies.

The screenshot shows the BOLD Systems User Console. At the top, there's a navigation bar with links to Databases, Taxonomy, Identification, Workbench, and Resources. Below the navigation is a header "User Console".

**Stats and Progress Feed:**

- Project Search:** A text input field with three buttons: Code, Tag, and Title.
- Project Management:** Buttons for Project List, New Project, and Record Search.
- Your Data:** A table showing campaign names, records managed, and records accessible. Examples include ACG Parasitoids (0 managed, 2 accessible), All Birds Barcoding Initiative (0 managed, 0 accessible), and All Caddis DNA Barcoding (338 managed, 0 accessible).

**Data Uploads:**

- Data Uploads:** Buttons for Sequences, Specimens, Traces, Images, Primer, and Bibliography.
- Most Recent Activities:** A table listing timestamp, who, action, and project for recent uploads. For example, on Nov-22, 2011 at 20:54, Chris C. Y. Ho performed a "New-Comment (1)" on the LIMC project.
- Previous Activities:** Buttons for "Previous 100 Activities" and "Previous 200 Activities".
- Downloads:** Buttons for "Download last week's activity" and "Download last month's activity".

### Searching

This console provides two ways to perform searches. The first is a project search, where a user looking to open a project can jump directly to it by entering the code in the project search bar. If the code is not known, they can generate a short list by entering a tag on the project or part of the project title. As an alternate, users can get a complete list of projects they have access to by clicking on the

The second search functionality is one that generates a list of records based on search terms consisting of geography, taxonomy, tags, sequence length, and pasted lists of identifiers. Records retrieved from a search can be downloaded or analyzed right on the system.

## Workbench: Creating a project

From the User Console, select the 'New Project' button from the Project Management menu.

All project details can be edited at any time (with the exception of Project Code, Project Type and Project Manager) by simply clicking on 'Modify Project Properties' in the Project Options menu of the Project Console. Only the Project Manager is able to modify these details.

Table: Field definitions for BOLD project creation form.

Project Title	Please create a descriptive name.
Project Code	A 3-5 letter code that needs to be unique across BOLD. A good approach is to use your initials and 2 or 3 other letters as an acronym for the title.
Project Type	Choose between the following options: <ul style="list-style-type: none"><li>• Data Project (contains specimen &amp; sequence records)</li><li>• Folder Project (contains other projects)</li></ul>
Primary Marker	Select your primary marker. COI is the default. Primary marker options are: <ul style="list-style-type: none"><li>• Cytochrome Oxidase Subunit I 5' (COI-5P)</li><li>• Interspacer Region (ITS)</li><li>• Ribulose-bisphosphate carboxylase (rbcL)</li><li>• Maturase K (matK)</li></ul>
Supporting Markers	Select as many secondary markers as needed from the list of registered markers.
Campaign	Select the name of the campaign the project is part of if desired.
Place in Container	Select the name of the Folder Project if desired.
Tags	Please enter annotations that you would like to appear as Project Tags on the Project List and Project Console pages. These Project Tags can make it easier to organize or define relationships between your projects.
Project Description	Enter a summary of the use and intention of the project. 15 - 500 characters.
Bounding Box	Define the bounding box of the collection area covered by the project using GPS coordinates.
Project Access	Check to make project publicly visible on BOLD and submit to the BOLD Data Portal.
Project Manager	The person who creates a project is automatically the Project Manager.
Assign Users	Other BOLD users can be added to a project. Different levels of access are possible, and are described below.  <b>Sequence Access:</b> Analyze Only - user can perform analysis on the data, but cannot view more than a summary of the data (sequence and related information remain hidden). View & Download - user can view or download the sequence data, as well as analyze. Edit Sequences - user can upload trace files, upload, edit and delete sequences, as well as view and analyze.  <b>Specimen Access:</b> Edit Specimens - user has control over sample identifiers, taxonomy, collection data, and images of the specimens: this edit permission level is intended for project managers, collectors, and taxonomists.

The screenshot shows the 'Project Properties' section of the BOLD Systems User Console. It includes fields for Project Title, Project Code (with a note about uniqueness), Project Type (radio buttons for 'Data Project' or 'Folder Project'), Primary Marker (dropdown for 'Select Primary Marker' with 'COI' as the default), Supporting Marker(s) (dropdown for 'Select Supporting Markers' with a note about contacting support), Campaign (dropdown for 'None (General Project)'), Place in container (dropdown for 'Independent Project'), Tags (text input for annotations with a note about project tags), Project Description (text input), Bounding Box (input fields for Top Left Lat/Lon and Bottom Right Lat/Lon with a note about 15 character minimum), Project Access (checkbox for public visibility with a note about allowing all BOLD users to view, analyze, and download sequence data), and Assign Users (table for assigning users with checkboxes for Analyze only, View & Download, Edit, Sequence Access, and Specimen Access).

### New Project Form.

#### \* Required Fields

Please note that the person who creates a project is automatically assigned as the **Project Manager**. To change the project manager, the current Project Manager must send a request to the BOLD support staff through support@boldsystems.org.

Supporting markers are added upon request. If a marker you require is not on the list, please contact BOLD support staff to register one through support@boldsystems.org.

The Bounding Box is a new field for projects in BOLD 3.0, that is used to define the collection area covered by the project.

## Workbench: Managing Data

Once your project has been populated with the specimen data, images, traces and sequences that you have uploaded to BOLD, it will resemble the figure on the right.

### Project Console

The project console is the main summary of a project which reports on progress and lists all actions and analysis open to a project member.

The console displays a report of the number of specimens, along with tallies of any missing components of the records. Also included are graphs to provide a quick visual overview of the project, as well as a list of all the users with access to the project. The links to the left provide access to uploads, downloads and various analysis tools.

Project Managers will see the “Modify Project Properties” button with which they can change the project title and description, add or remove markers, and add, remove or modify permissions of users at any time. The Project Manager also has access to publish the records in the project to GenBank. (See page 26 for more details on GenBank submissions.)

To access a list of the records within each project, click on “View All Records” in the project options menu.

### Record List

A Project Record List is the full list of all records within a project, along with the actions and tools open to a project member.

The record list gives access to individual specimen and sequence data for each record.

You can select specific records for analysis or download using the checkboxes.

#### Flags

- Icons appearing next to a record indicate the presence of certain characteristics of a record; see legend to the right for more details.
- A red-highlighted, bolded sequence length is a warning that the sequence contains more than 1% ambiguity and won't meet the Barcode Standard.

The red arrows along the column headers can be used to sort the records by header.

The Project Manager or a user with full edit permissions can move records from one project to another by selecting the records needed and then clicking on “Move records to Another Project” in the Options menu. The destination projects that will appear in the list will be ones in which the user has full permissions to. Click on the Sample ID or the Process ID to access the Specimen Data and Sequence Data respectively, for each record. These are illustrated on the next page.

Project Console

Record List

	GPS coordinates present for record
	Country data present for record
	Images present for specimen
	The number of traces present
	Stop codons present in sequence
	Contamination present in sequence
	Flagged record, not in ID engine
	Sequence is Barcode Complaint

BOLD Record List icon legend

## Workbench: Record Search

There is one common interface for the three ways to search the BOLD workbench.

To search from all projects that a user has access to, select the “Search Records” button in the User Console or Project List. The form gives the option to add in all public records on BOLD. To search from within a merged set of projects or a single project, select the “Search Records” button from the Project Console.

The form contains the following fields:

Taxonomy	Searches specific taxonomic names
Geography	Searches the country and province names
Marker	A selection of the markers on BOLD
Min Seq Length	Define minimum number of base pairs desired.
Tags	Searches tags on specimen or sequence records
Identifiers	Searches Sample IDs, Process IDs, GenBank accessions and BIN GUIDs.

The screenshot shows the "Record Search" dialog box. It includes fields for Taxonomy, Geography (Country or Province), Marker (with a dropdown menu showing "None"), Minimum Sequence Length, Tags, and Identifiers (with radio buttons for Process ids, Genbank Accession, Sample ids, and BIN GUID\*, where BIN GUID\* is selected). There is also a checkbox for "Include public records" and a note explaining what a BIN GUID is. At the bottom are "Search Records" and "Cancel" buttons.

Record Search Engine pop-up window

## Workbench: GenBank Submission

BOLD has an automatic submission tool for Project Managers to publish sequences to GenBank.

All records within the project opened will be submitted to GenBank. If only a subset of records need to go to GenBank, then the records should be moved to a new project which can be submitted.

GenBank accession numbers are generally returned by email to the project manager within five business days. The accession numbers will be associated on BOLD with your records for quick reference.

### After Publication - BOLD

Once your paper is published, you need to make your BOLD project public to release them into the BOLD Data Portal. The Project Manager can do this by clicking on "Modify Project Properties" within the project and checking off the box that says "Make this project publicly visible". Please submit a bibliography to the BOLD database following the directions on page 21. This will associate your publications with the records on BOLD using the GenBank Accession numbers.

### After Publication - GenBank

Submissions to GenBank are automatically locked for 1 year to allow time for publication. If the publication is released sooner than one year, the corresponding author should contact GenBank directly to request public release at the time of publication.

The screenshot shows a web-based form titled 'BOLD Systems - Project Publication'. At the top, it says 'BOLDSYSTEMS' and 'BOLD Systems - Project Publication'. A note indicates that fields marked with an asterisk (\*) are required and must be in English characters. The form includes sections for 'Publication Title' (a single input field), 'Author List' (multiple input fields for first name, middle initial, last name, and suffix), and 'Corresponding Author' (multiple input fields for first name, middle initial, last name, suffix, phone, fax, email, street address, city, state/province, zip/postal code, and country). There is also a note about double-checking data before submission. At the bottom, there are 'Submit Request' and 'Close' buttons.

GenBank Submission Form  
\* Required Fields

### GenBank Update Channel

The BOLD environment allows users to maintain and update specimen and sequence records, which has created opportunities to disseminate current and up-to-date information to sister databases that replicate BOLD data. The first instance of an update system has been developed in partnership with GenBank, resulting in changes made to records on BOLD being reflected in GenBank flat files.

Identifications of records submitted through BOLD to GenBank can still be refined and updated as new information is obtained. Changes to the taxonomy of BOLD records are automatically sent to GenBank on a weekly basis so that GenBank has the most current and up to date information. Updates made to all GenBank compatible fields are made available to GenBank staff so that they can integrate changes into the published records. This means that records only need to be modified in BOLD and the update is automatically distributed to partner databases. GenBank staff may contact you to confirm any changes to your records which seem unusual before they modify your records.

Contact the BOLD support team through support@boldsystems.org if necessary with any aspect of publication.

## Workbench: Analytical Tools

BOLD includes core and extended tools to analyze specimen and sequence data:

### Core Analysis Tools

- Image Library: Compare morphological characteristics
- Distribution Maps: Interact with geographical data
- Taxon ID Tree: Visualize a neighbour joining tree with matching images
- Barcode Index Numbers (BINs): Barcode clusters (see page 10)
- Identification Engine: Locate closest matches to an unknown sequence (see page 6)

### Extended Analysis Tools

- Distance Summary: Browse sequence divergence at multiple taxonomic levels
- Sequence Composition: Explore compositional variation at all codon positions
- Nearest Neighbour Summary: Evaluate the Barcode gap
- Accumulation Curve: Review sampling efficiency
- Alignment Browser: Diagnose unaligned sequences



When the “Expand” icon (shown to the left) appears next to a graph, the graph is expandable for a better quality version that can be used in publications.

### Run Multiple Analytic Tools in Parallel

Use the newly available option to run multiple analyses and have the results emailed to you when the analysis is finished

Results can be stored for up to 4 weeks, saved for future comparison, and links to the results can be shared between collaborators.

Find this option on the parameters page for most analysis tools.

## Workbench: Image Library

Once images have been uploaded to your project, you can view them in two ways. The first is by opening an individual record where any corresponding images will be displayed below the specimen data. The second is the Image Library for viewing a group of specimens, shown in Figure 15-1.

The Image Library allows you to sort by orientation so you can compare morphological differences between specimens. This tool is useful for diagnosing contamination or misidentifications as taxonomy is displayed below each image.

Image Licensing is viewable upon mouse rollover. To view the attribution and further details on specific images, open the specimen data for that record from the record listing page.

**BOLD SYSTEMS**

Number of records : 9  
Filter by orientation : By default, a single image is picked for each specimen, filter options override the default behavior.  
 Dorsal(9)  
 Filter  Clear  Fix Aspect Ratio (3:2)

Re-render images :  3 per page      300 Images Per Page - Pages : 1

BC ZSM Lep 11319 [Dorsal] Danielaparra fragmentata	BC ZSM Lep 11663 [Dorsal] Ennada pellicata	BC ZSM Lep 11665 [Dorsal] Eupithecia horismoides
BC ZSM Lep 11127 [Dorsal] Nebula bellissima	BC ZSM Lep 11659 [Dorsal] Orthonama pliemyrata	BC ZSM Lep 11281 [Dorsal] Physoloba multivirgulata
BC ZSM Lep 11382 [Dorsal] Rheumaptera exacta	BC ZSM Lep 11276 [Dorsal] Synpelurga corralensis	BC ZSM Lep 11325 [Dorsal] Triptiloides laeta

Image Library (Lepidoptera)

(Displayed Specimen Images: Copyright 2010, Axel Hausmann, Bavarian State Collection of Zoology)

## Workbench: Distribution Map Analysis

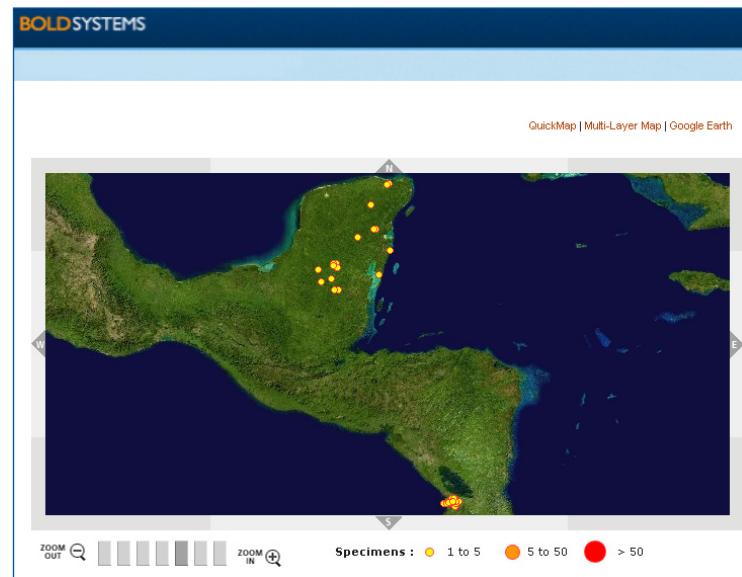
BOLD's Distribution Maps plot the collection points for a selected set of specimens when geographic reference data is available. The multiple mapping tools available on BOLD are described below.

### Quick Map

The Quick Map is built using the NASA Blue Marble Project. They are open access and therefore can be re-used and modified at the user's discretion.

Users can zoom in to a max of 1 km per pixel by using the scale at the bottom or clicking on a region in the map and can pan the map in 4 directions by clicking on the N/E/S/W directions in the frame.

The collection points are shown on the map using markers specifying the density of sampling at each point. Larger markers are placed beneath smaller ones so all points can be visible.

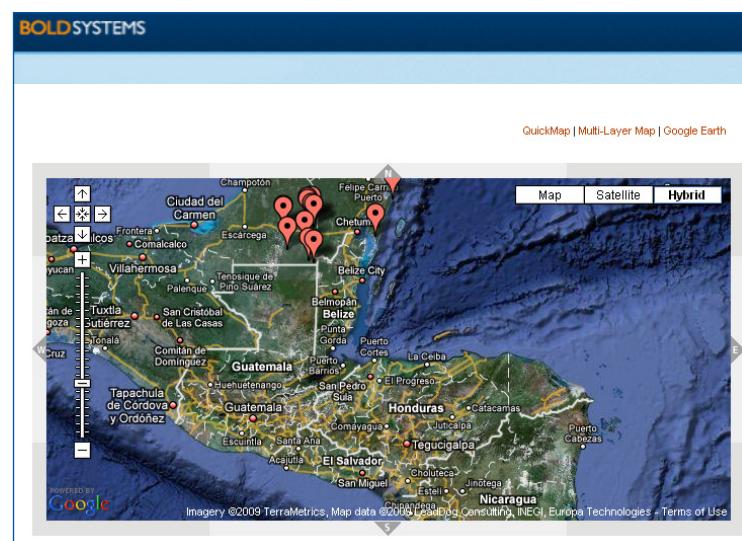


### Multi-Layer Map

The Multi-Layer Map is based on Google maps. (Google gives permission for use in publications as long as the Google logo remains on the image.) The layers include political boundaries with regional names, as well as a satellite map of the world. These can be viewed individually or combined, which is shown to the right.

The markers on this map are active and can be clicked to retrieve a list of the specimens from BOLD.

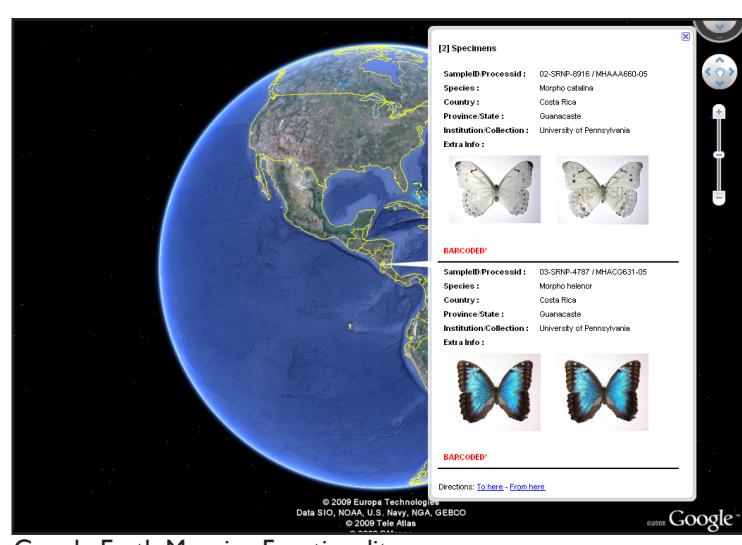
This record list is also active, meaning users can open specimen records, which can be edited directly. This is a great way to validate and correct GPS data.



### Google Earth

The Google Earth map is a display of the specimen collection points in the program Google Earth. This is free software that can be downloaded from the web. (Google gives permission for use in publications as long as the Google logo remains on the image.) The benefit of this type of map is that it provides a portable KML file download which can be shared among colleagues.

This map is embedded with specimen images, along with specimen identifiers, country, province/state, institution/collection, and extra info.



## Workbench: Taxon ID Tree

This section describes how to use the Taxon ID Tree. The user can access this feature by clicking on “Taxon ID Tree” under the Sequence Analysis panel on the project console or record list. If the tree is only desired for a selection of records, it may be accessed from the Project Record List.

Below is an example of what the tree will look like. The user can view and save a modifiable PDF of the tree, export the tree to Newick format and view a taxonomy report, as well as view the matching image library and the accompanying data spreadsheet.

Once a tree is built, it can be used to compare specimens. Users can identify specimens, as well as catch misidentifications and contaminations.

Sequence Data	<ul style="list-style-type: none"> <li>Nucleotide</li> <li>Amino Acid</li> </ul>
Distance Model	<ul style="list-style-type: none"> <li>Kimura 2 Parameter</li> <li>Jukes Cantor</li> <li>Pairwise Distance</li> </ul>
Tree Building Method	Neighbour Joining is the only method at this time.
Marker	Select the marker from a multi-gene dataset.
Alignment Options	<ul style="list-style-type: none"> <li>No Alignment</li> <li>Allow BOLD align sequences</li> <li>Kalign</li> <li>Muscle</li> </ul>
Select Terminal Branch Labels	Many options for labels to add to the end of each branch including taxonomy, geography, identifiers, sequence details, BIN identifiers
Photographs	Option to include matching specimen photographs and spreadsheet for comparison.
Codon Positions Included	1st, 2nd and 3rd Codon Positions are included as default but may be excluded
Apply Filters	Can be applied to disregard sequences below a given length (since very short sequences can skew the results) or to analyze only sequences with less than 1% ambiguous bases. Exclude problematic sequences.
Colourize Tree Based on	<ul style="list-style-type: none"> <li>Problematic Sequences</li> <li>Taxonomy: Class</li> <li>Taxonomy: Order</li> <li>Taxonomy: Family</li> <li>Taxonomy: Subfamily</li> <li>Location: Country</li> <li>Extra Info</li> <li>Sequence Age</li> <li>BINs</li> </ul>
Result Options	Choose to view the results immediately or to have the results emailed to your account when available.

Parameters available for Taxon ID Tree

**BOLD SYSTEMS**

Sphingidae - Haxaire collection 5 [SOWE]

Sequence Data: Nucleotide ▾

Distance Model: Kimura 2 Parameter ▾

Tree Building Method: Neighbor Joining

Marker: COI-5P - Cytochrome Oxidase Subunit 1 5' Region ▾

Alignment Options: No alignment ▾

Select Terminal Branch Labels:  Extra info

Phylum  Family  
 Class  Subfamily  
 Order

Geographic Region (Country & State/Province)  
 Collection Date  
 Sex/Gender

Specimen Sample ID  
 Field Number/Isolate  
 Voucher Number/Museum Number  
 Sequence/Process ID

Include Sequence Length in label  
 Include GC Composition in label

Matching specimen photographs and spreadsheet

BIN Name  GUID  
 1st  2nd  3rd

Sequence Length > 200 bp ▾

Exclude Contaminants  
 Exclude Records With Stop Codons  
 Exclude Records Flagged as Misidentifications or errors

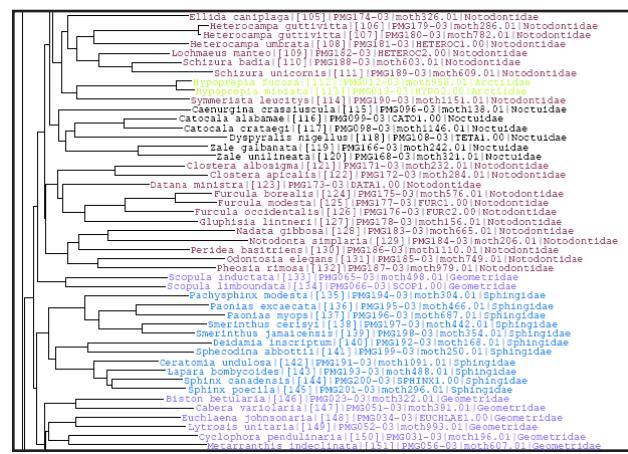
Only sequences with < 1% ambiguous bases

Highlight Problematic Records ▾

Colorize Tree Based on:  
Result Options:  
View the results immediately ▾

**Build Tree**

Taxon ID Tree Parameter Page



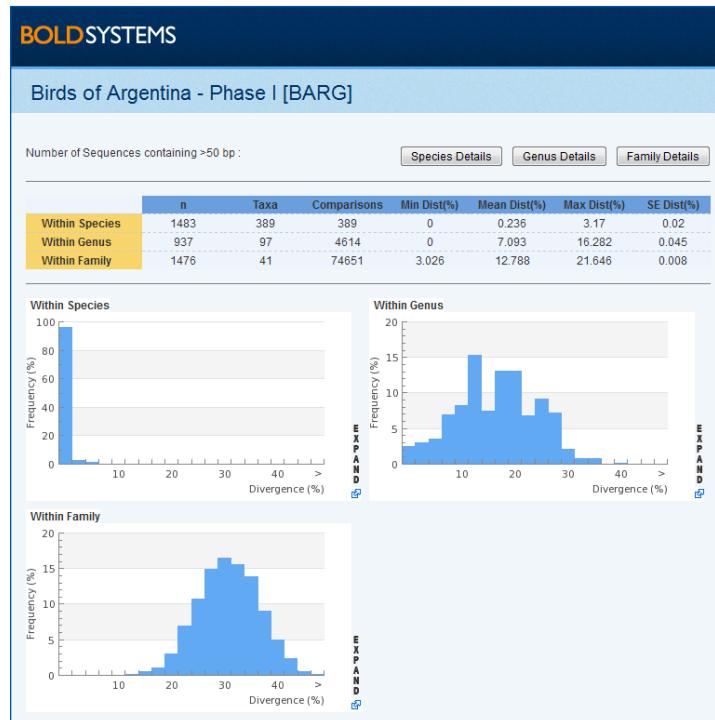
Standard Taxon ID Tree (Lepidoptera)

## Workbench: Distance Summary

It is desirable for barcodes to show very low sequence divergence within a species, with significantly higher sequence divergence at higher taxonomic levels.

The Distance Summary tool gives a report of sequence divergence between barcode sequences at the level of species, genus, family, order, and class.

Distance values are calculated using the Kimura 2 Parameter method. Comparisons are performed between the given taxonomic levels with the frequency plotted as shown in the figure to the right. Details for the comparisons done at the level of species, genus, and family are available by clicking on the links in the top right corner of the Distance Summary browser.

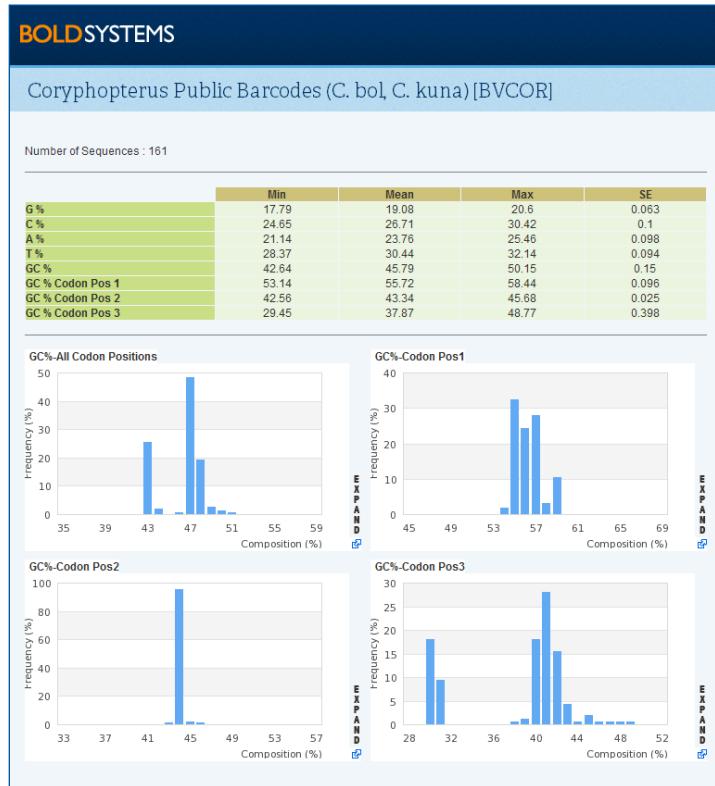


Distance Summary Page

## Workbench: Sequence Compostion

The frequency of DNA bases, observed with emphasis on GC-content, can be a useful metric for evolutionary biologists. GC-content within the barcoding region of COI has been correlated with GC-content of the entire mitochondrial genome for many species.

Using the Sequence Composition tool allows the user to view the frequency of each base, G, C, A and T, as well as graphics for GC content on all codon positions. This information includes overall sequence composition, as well as for codon positions 1, 2, and 3. "Detailed View" tabulates the results for each specimen.



Sequence Composition Results Page

## Workbench: Nearest Neighbour Summary

The Nearest Neighbour Summary presents users with an examination of the distance to the nearest neighbour for each of the species in the list of specimens. Distances are highlighted if the nearest neighbour is less than 2% divergent, or when the distance is less than the intra-specific distance. Warnings presented by this tool may be summarized by clicking on the link in the top right corner of the Nearest Neighbour browser.

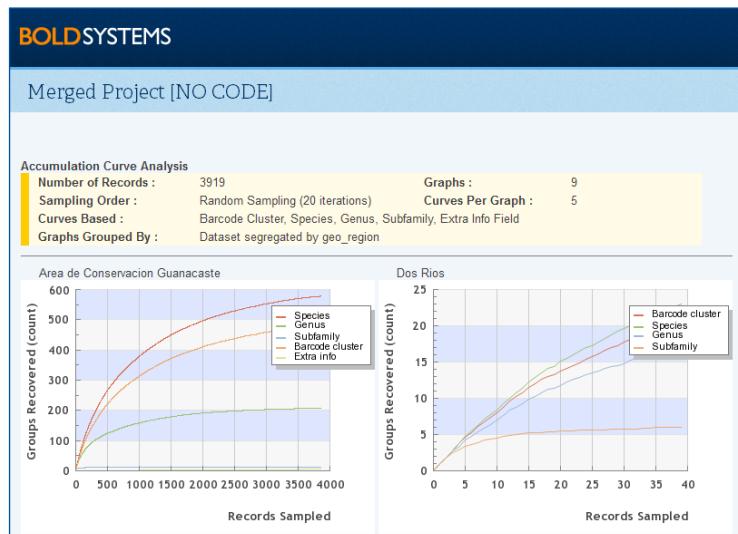


Nearest Neighbour Summary Page

## Workbench: Accumulation Curve

An accumulation curve of standardized DNA barcodes and related features provides a clear, transparent and reproducible estimate of the diversity and sampling efficiency of areas or collections. Should one wish to characterize a region's invertebrate fauna, it is clear that we need to accelerate the sampling process if we are to understand how well we have sampled the community.

This tool also allows users to quickly compare sampling efficiency at multiple regions by multiple taxonomic levels.

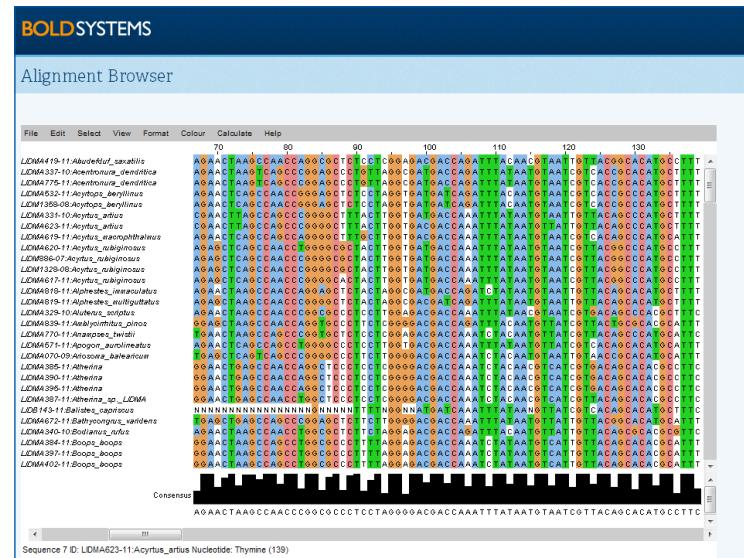


Accumulation Curves Page

## Workbench: Alignment Browser

Managing sequence alignments and base calls are a critical step in any barcode analysis. To prevent the inconvenience of importing sequences into 3rd party software to analyze and often edit, BOLD provides an integrated alignment browser that will include many features popular in other packages. In BOLD 3.0, the updated alignment browser can handle thousands of sequences and will soon support direct editing to the database.

Multiple alignment options such as Muscle and Kalign algorithms as well as colourization options are now available.



Alignment Browser Page

## Notes

## Notes



Last modified: November 2011

**BOLD SYSTEMS**

For online version, please visit: [www.boldsystems.org/docs/](http://www.boldsystems.org/docs/)  
For support with any feature of BOLD, please email: [support@boldsystems.org](mailto:support@boldsystems.org)

Barcode of Life Data Systems  
Biodiversity Institute of Ontario  
University of Guelph  
50 Stone Road East  
Guelph, Ontario, Canada  
N1G 2W1

Copyright ©2011 Biodiversity Institute of Ontario

Authored by:Megan Milton ([mmilton@boldsystems.org](mailto:mmilton@boldsystems.org))  
Mallory Van Wyngaarden ([mvanwyng@boldsystems.org](mailto:mvanwyng@boldsystems.org))  
Taika von Königslöw ([tvonkoni@boldsystems.org](mailto:tvonkoni@boldsystems.org))  
Sujeevan Ratnasingham ([sratnasi@boldsystems.org](mailto:sratnasi@boldsystems.org))

Designed by:Suzanne Bateson, Biodiversity Institute of Ontario