



# BARCODE OF LIFE DATA SYSTEMS HANDBOOK

October 2013

[www.boldsystems.org](http://www.boldsystems.org)  
Version 3.6

## Table of Contents

### Getting Started

Introduction	3
Navigation	4
Registration & User Preferences	5

### Databases

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

### Data Submissions

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
Accessing Records	23
Create a Project	24
Project Console, Record List	25
Datasets and DOIs	26
Publication on GenBank/BOLD	27

### Analysis Tools

Image Library	28
Distribution Maps	29
Taxon ID Tree	30
Distance Summary	31
Sequence Composition	31
Barcode Gap Analysis	31
Accumulation Curve	32
Alignment Viewer	32
BIN Discordance Report	33
Diagnostic Character Analysis	33

This handbook provides details on functionality, data structures and best practices for BOLD version 3.6. 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. 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 assistance with any feature of BOLD, please email the BOLD Support Team: [support@boldsystems.org](mailto:support@boldsystems.org)

Copyright © 2013 - All Rights Reserved

BOLD Systems, Biodiversity Institute of Ontario

## 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. Due to its web-based delivery and flexible data security model, it is also well positioned to support projects that involve broad research alliances.

BOLD version 3.6 was released in October 2013. This update addresses a lot of usability gaps from the prior version and has been built from usage models of novice and expert users. It additionally introduces new functionality to further enable investigation and retrieval of barcode data. Find new functionality highlighted throughout this Handbook.

The screenshot displays the BOLD Systems website. At the top, the BOLD SYSTEMS logo is on the left, and navigation links for Databases, Taxonomy, Identification, Workbench, and Resources are in the center. A 'Log In' link with a settings icon is on the right. Below the navigation bar is a world map populated with numerous yellow and red dots representing data points. A small overlay on the map shows 'Barcodes: 2,615,270' and a 'Per Site' filter with buttons for 1000, 100, 10, and 1. Below the map is a search bar with a 'Taxonomy' dropdown and a 'Search' button. The bottom section features four icons and their descriptions:

- Public Data Portal:** A data retrieval interface that allows for searching over 1.7M public records in BOLD using multiple search criteria including, but not limited to, geography, taxonomy, and depository.
- Barcode Index Numbers:** A searchable database of Barcode Index Numbers (BINs), sequence clusters that closely approximate species.
- DNA Barcode Education Portal:** A custom platform for educators and students to explore barcode data and contribute novel barcodes to the BOLD database.
- Workbench:** An integrated data collection and analysis environment that securely supports the assembly and validation of DNA barcodes and ancillary sequences.

[www.boldsystems.org](http://www.boldsystems.org)

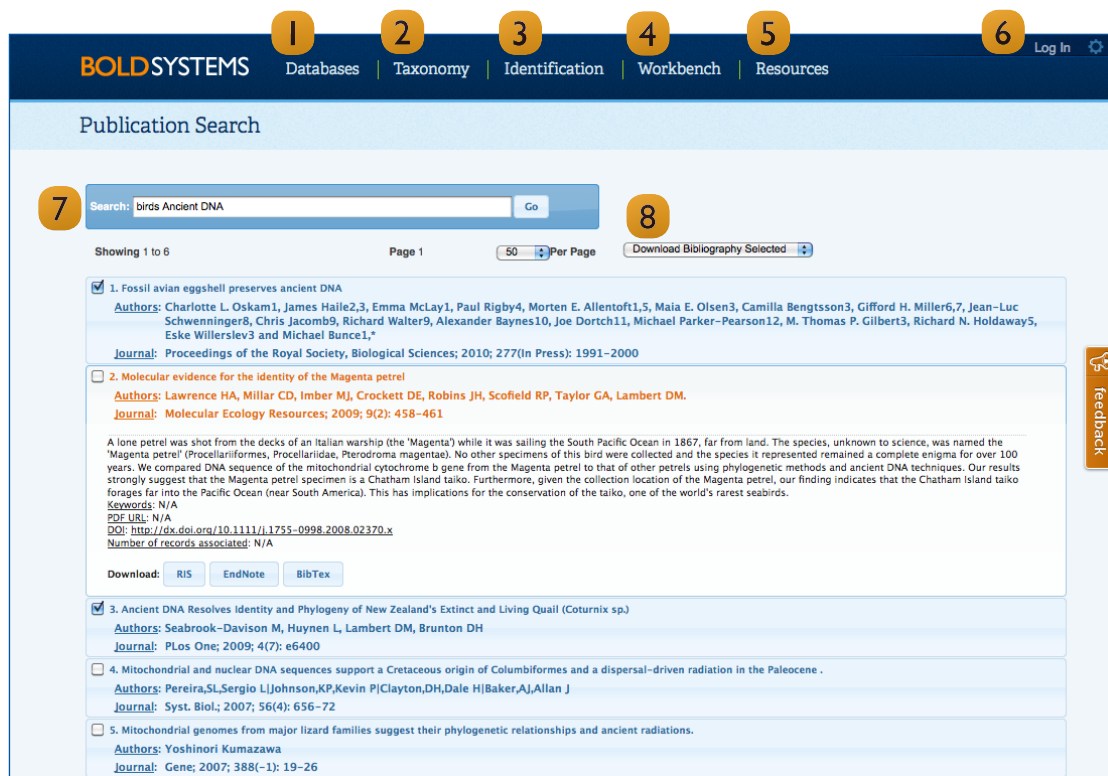
Available from the BOLD home page, the DNA Barcode Education Portal (BOLD-SDP) is an educational module released in 2013. It supports the assembly and analysis of DNA barcode data by students in a classroom environment. Please find more information at [www.boldsystems.org/edu](http://www.boldsystems.org/edu)

## Getting Started on BOLD

The BOLD interface is designed to provide easy access to commonly used features and databases.

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

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



A publication search page illustrating navigation features

Description of features numbered in the image above

1	Databases	<p>The Databases link provides access to the following resources that are accessible without signing in.</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 BINs 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>
2	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 is licensed as Creative Commons or No Rights Reserved, following the rules of the license.
3	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, though advanced features are available for those registered with the system.
4	Workbench	The workbench link provides access to the BOLD data analysis and management workbench. After logging in, the initial page is the User Console.
5	Resources	The resources link provides access to Site Documentation (including an online version of this handbook), Barcoding resources and access to data releases from Barcoding initiatives like iBOL.
6	Log in/out	In the top right corner of any page, users can log in or log out. When logged in, the user's name appears in this section. The gear symbol also provides access to edit account details for users who are logged in.

Table continued on next page...



Table continued from previous page...

7	Search Bars	In the public databases listed on the previous page, a search bar is present at the top of the page. Users can enter any combination of keywords to search within these databases. 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.
8	Downloads	In each database, there is also an option to download the public data returned from a search. This includes barcode sequences, traces, specimen data, bibliographies, and primer sequences. Distribution Maps are available for BINS and Public Data searches.

## Registering on BOLD

The registration form for a BOLD account is available by clicking on the Workbench or Log In link and then selecting “Create Account”. After the registration is submitted, a welcome e-mail will be sent with the information needed to log in and begin using the BOLD workbench. Upon signing in, users will see the User Console. After users start to contribute data and join collaborators projects, the User Console will provide progress statistics and activity feeds (see page 22 for a depiction of the User Console).

The User Console provides access for searching the data on BOLD by project code, title or tag, or by using the record search (see page 25 for more details on the record search within the Workbench). After gaining access to projects (either by creating them or being added to colleague’s projects), users can upload specimen data, images, and sequence data.

### Why register for a User Account?

Getting an account on BOLD expands the list of options available to a user beyond access to public data and use of the identification engine. Users can annotate published data, to help curate and cleanse the barcode library. Moreover, users will be able to submit data to BOLD and gain access to other in-progress, private, projects with the permission of the data owners. Once records are on BOLD, a large set of analytical tools are available for validation and generation of reports for publication. The system will automatically check sequences submitted for Barcode Compliance and provide reports on records lacking pertinent details. BOLD also provides a simple mechanism to release sequences publicly, as well as into partner nucleotide databases such as GenBank.

## User Preferences

User Preference options are available for logged in users to edit their account details. By clicking on the gear symbol in the upper right corner, users can edit their email address, institutional affiliation and name, as well as change the account password. The User Preferences page also provides some profile statistics for each user, such as the count of projects managed and participating in.

## Databases: Identification

The library of sequences collected in BOLD is available for facilitating identification of unknown sequences. The ID engines use 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 for animals 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 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

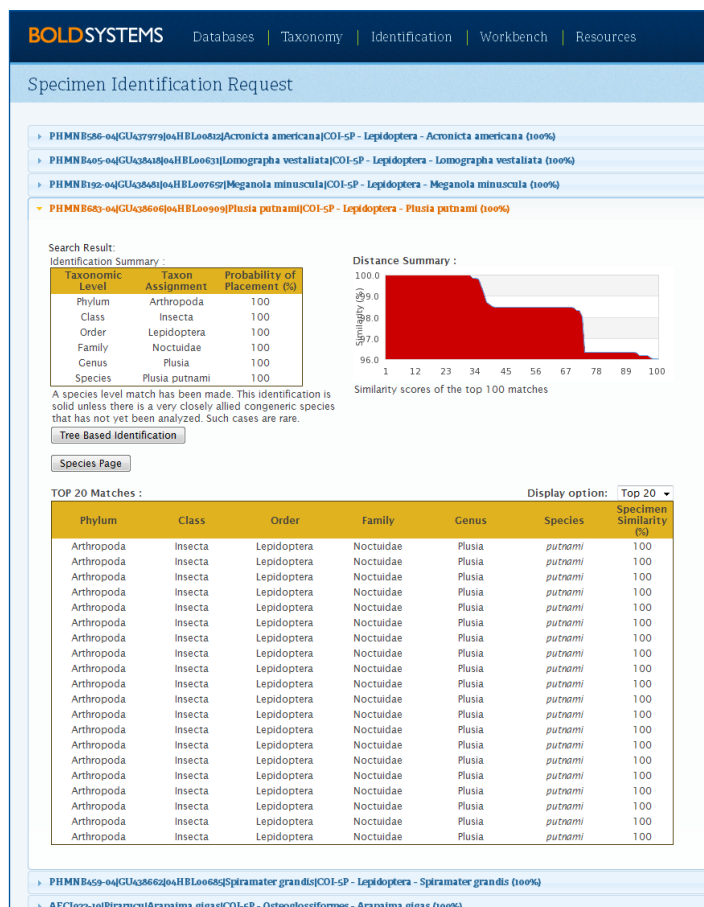
In the BOLD Identification System, ITS is the default identification tool for fungal barcodes and rbcL and matK are the defaults for plant barcodes. Both return a species-level identification when possible. Further validation with independent genetic markers will be desirable in some forensic applications. 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

### NEW! Emailing Identification Results

The newest version of BOLD provides the ability to have identification results emailed when submitting individual sequences or a batch of query sequences. Users must be signed in to utilize the emailing or batch identification features.

## Databases: Taxonomy Browser

The Taxonomy Browser is a synthetic database that allows users to examine the progress of DNA barcoding by browsing through the different levels of the taxonomic hierarchy available on BOLD.

Within the Taxonomy Browser users are able to select between the animal, plant, fungus, and protist kingdoms and navigate from phylum to species level. Statistics on the progress of DNA barcoding at each taxon are generated from both public and private data while protecting private user-owned data.

To look up a specific taxon directly, use the search function by entering a taxonomic name into the search bar at the top of the Taxonomy Browser or on the BOLD main page.

<b>1. Lineage</b>	Displays the taxon name and the higher taxonomic levels.
<b>2. Sub-Taxonomy</b>	Links to all sub-taxa with number of specimen records for each.
<b>3. Taxon Description</b>	Displays the description of this taxon from the Wikipedia website.
<b>4. Statistics</b>	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 in the Public Data Portal can be accessed from this section.
<b>5. Contributors</b>	Graphs depicting the institutions that provided the samples and sequencing for the samples.
<b>6. Imagery</b>	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.
<b>7. Image Details</b>	The taxonomic identifier, the sample identifier and image licensing details are displayed beneath the image that is selected.
<b>8. Collection Sites</b>	A map of the collection sites for records in BOLD, including a list of the top countries
<b>9. Taxon Occurrence</b>	A map of the occurrence data for this taxon worldwide, streaming from the GBIF website.

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

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

**1** Nymphalidae (family) - Arthropoda; Insecta; Lepidoptera;

**2** Sub-taxa

**3** Taxon Description (Wikipedia)

**4** Statistics

**5** Contributors (Specimens & Sequencing)

**6** Imagery

**7** Image Details

**8** Collection Sites

**9** Taxon Occurrence (GBIF)

BOLD taxonomy browser page for the Nymphalid butterflies

## Databases: Publication

The Publication Database is accessible from anywhere in the application by clicking on Databases in the header or Publication Database in the footer. This database indexes title, abstract, year, and authors, allowing for broad searches.

Selecting a publication from the database will provide further details, including a link to the article on the journal's site and access to the records if they are in BOLD.

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

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

Publication Database

## Databases: Primer

The Primer Database is accessible anywhere in the application by clicking on Databases in the header or Primer Database in the footer. Using the search bar, users can search for terms that appear in the primer code, submitter or reference fields.

Selecting a primer 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 to 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 Database







# 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 three 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.
- An annotation framework that allows researchers to review and critique the taxonomic identifications associated with each BIN and notify data owners of errors.

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 URI and web service functionality enabling integration with other databases.

Ratnasingham S, Hebert PDN (2013) A DNA-Based Registry for All Animal Species: The Barcode Index Number (BIN) System. PLoS ONE 8(8): e66213. DOI:10.1371/journal.pone.0066213

BIN pages display aggregated data in several sections:

<b>1. BIN Details</b>	BIN details include BIN identifiers (URI and DOI), the member count, and distributional information. Also, nearest neighbour BIN details are provided, along with the nearest member and the taxonomy of that record.
<b>2. Taxonomy</b>	The taxonomy of the public data is visible for the BIN, with highlighting to indicate taxonomy concordance and discordance. <b>NEW!</b> For each taxon, logged in users can search the records that they have access to by clicking on the magnifying glass icon.
<b>3. Annotation</b>	The Add Tags & Comments button allows for annotation via tagging or comments on several aspects of the BIN page. See more details on Annotation on the next page.
<b>4. Distance Distribution</b>	A histogram plotting the within-BIN species distribution against the Nearest Neighbour BIN distribution.
<b>5. Associated Publications</b>	List of the publications that contain sequences from the BIN.
<b>6. Dendrogram of Sequences</b>	For BINs with 3 - 150 members, a circle tree is displayed which also includes the nearest neighbour. Hovering over taxon names on the circular tree highlights the terminal branch. A PDF version of the tree is available for download for all BINs with more than 2 members.

Descriptions of elements in BIN pages, continued on the next page.

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

Public Data Portal - BIN Page

Barcode Index Number Registry For BOLD:AA85726

**BIN DETAILS:**

BIN URI:	BOLD:AA85726	Average Distance:	0.58% (p-dist)
Member Count:	Pending	Maximum Distance:	2.19% (p-dist)
Nearest Neighbour:	BOLD:ABZ4800	Distance to Nearest Neighbour:	1.76% (p-dist)

**TAXONOMY:**

Phylum:	Arthropoda [83]
Class:	Arachnida [83]
Order:	Araneae [80]
Family:	Opiliones [3]
Subfamily:	Agelenidae [77]
Genus:	Agelenopsis [76]
Species:	Agelenopsis potteri [14]

**DISTANCE DISTRIBUTION:**

**PUBLICATIONS:**

**DENDROGRAM OF BIN & NEAREST NEIGHBOR:**

**HAPLOTYPE NETWORK:**

**SPECIMEN IMAGES:**

**COLLECTION SITES:**

**SPECIMEN DEPOSITORIES:**

**SEQUENCING CENTERS:**

**PHOTOGRAPHERS:**

**COLLECTORS:**

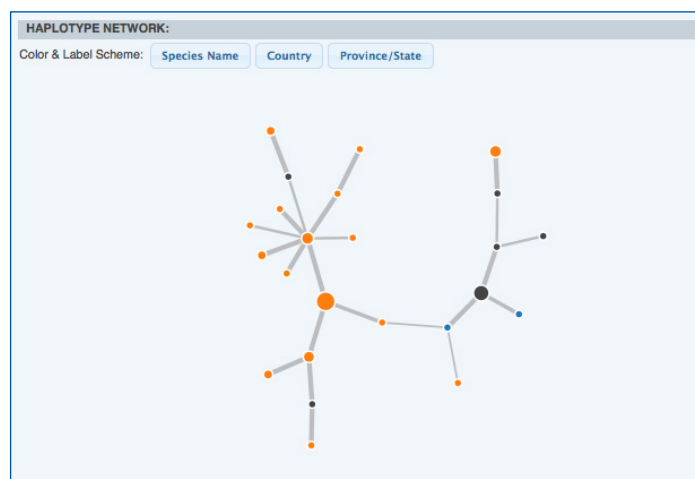
**SPECIMEN IDENTIFICATION:**

**SEQUENCING SUPPORT:**

Example BIN page

Table continued from previous page...

<b>7. Haplotype NEW!</b>	The interactive diagram allows for investigation of the haplotypes in the BIN cluster along species and geographical splits. Hovering over a haplotype node in the diagram reveals details on which species or geographical information are grouped. The larger the node, the more sequences in the haplotype. The thicker the line between nodes, the more closely related those two haplotypes are.
<b>8. Specimen Images</b>	View images for associated records, along with licensing information.
<b>9. Sampling Sites</b>	Displays a map of the collection sites based on GPS values.
<b>10. Attribution</b>	Lists institutions where specimens are deposited, sequencing centres, photographers, collectors, taxonomists and funding sources. <b>NEW!</b> For each institution, logged in users can search the records that they have access to by clicking on the magnifying glass icon.
<b>Further Information</b>	Statistics are available for the collection countries and the data owners (not shown).



A close-up of a BIN Haplotype Network.

## 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. The annotation system leverages the large user-base and expert knowledge for curation of both private data within collaborative projects and public data through the Public Data Portal. 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 data by BOLD users will appear in the activity feed on the User Console and the activity feed on the appropriate Project Console. Comments will persist on the data element with the user's full name and a date stamp. Tags can be removed at any time by any user.

Annotation is available wherever the "Add Tags and Comments" button appears within BOLD. Users must be signed in to BOLD to be able to add tags and comments.

Annotation pop-up window



## Workbench: Specimen Data Submission

The first step to creating records on BOLD is specimen data submission. Each record is assigned a BOLD Process ID when uploaded. After specimen data is uploaded, images, traces, and sequences can then be uploaded. There are two ways to enter records onto BOLD: manually through the online interface or with bulk spreadsheet submissions 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 taxonomy into the BOLD library. Described below are the specimen data fields available on BOLD. See the next page for further details.

### \* Required Fields

Sample ID *	ID associated with the sample being sequenced (often identical to or an extension of Field or Museum ID).
Field ID *	Field number from the collection event or the specimen identifier from a private collection.
Museum ID *	Identifier for specimen assigned by formal collection upon accessioning; also referred to as Voucher ID
Collection Code (only to be filled in if Museum ID is)	Code associated with given collection. Used in conjunction with Museum ID to disambiguate a specimen ID number that might be used in different collections within the same institution.
Institution Storing *	Full name of the institution where specimen or tissue is stored.

Table 1: Field definitions for Voucher info page on accompanying spreadsheet.

Sex	Male/female/hermaphrodite <b>only</b> .
Reproduction	Sexual/asexual/cyclic parthenogen <b>only</b> .
Life Stage	The age class or life stage of the specimen(s) at the time of sampling. The field supports free text but is a controlled vocabulary for validation. Example: "Adult", "Immature", "pupa", etc
Extra Info	A brief note or project term associated with the specimen for rapid analysis.
Notes	General notes regarding the specimen
Voucher Status	Status of the specimen in an accessioning process. Controlled vocabulary: <ul style="list-style-type: none"> <li>"Museum Vouchered:Type"</li> <li>"Museum Vouchered:Type Series"</li> <li>"Vouchered:Registered Collection"</li> <li>"To Be Vouchered:Holdup/Private"</li> <li>"E-Vouchered:DNA/Tissue+Photo"</li> <li>"Dna/Tissue Vouchered Only"</li> <li>"No Specimen"</li> </ul>
Tissue Descriptor	A brief description of the type of tissue or material analyzed. Example: "muscle", "leg", "thorax", "liver", "blood", "feces", etc.
Associated Taxa	A list (concatenated and separated) of taxa associated with the taxon at the time of its collection. References to other taxa should be preceded by the relationship. Use of this field implies knowledge of an associated specimen where the identification is an inference. Use of the associated Specimen(below) field is preferred when the associated specimen is databased. Examples: "host: Quercus alba", "prey: caterpillar"
Associated Specimens	A list (concatenated and comma separated) of other specimens associated with the subject specimen at the time of its collection. References to other specimen identifiers should be preceded by the relationship. Examples: "host: PLANT23452, prey: USNM45677" when both prey and host specimens have been captured.
External URLs	Web accessible links that provide additional information about the specimen preceded by a descriptor. Multiple links should be pipe separated (" ").

Table 3: Field definitions for Specimen Details page on accompanying spreadsheet.

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	The full name of the identifier's institutional or organizational affiliation if one exists.
Identification Method	The method(s) used to identify the specimen.
Taxonomy Notes	Additional notes relating to the identification of the organism.

Table 2: Field definitions for Taxonomy page on accompanying spreadsheet.

Collectors	Comma delimited list of collectors.
Collection Date	Date of collection, must be in DD-MMM-YYYY
Continent	ISO Continents
Country/Ocean *	The full, unabbreviated name of the country, major political unit, or ocean.
State/Province	The full, unabbreviated name of the state, province, territory, or prefecture within the given country.
Region	Park, county, district, lake or river.
Sector	Sector of park or county/city.
Exact Site	Exact location of collection site
GPS Coordinates	Latitude & Longitude in "degrees.decimal degrees" format (e.g. 45.837).
Elevation	Elevation of sampling site. Measured in meters relative to sea level. Negative values indicate a position below sea level.
Depth	For organisms collected beneath the surface of a water body. Measured in meters below surface of water.
Elevation Precision	A numerical representation of the precision of the elevation given in meters and is represented as +/- the elevation value.
Depth Precision	A numerical representation of the precision of the depth given in meters and is represented as +/- the depth value.
GPS Source	The source of the latitude and longitude measurements.
Coordinate Accuracy	A decimal representation of the precision of the coordinates given in the decimalLatitude and decimalLongitude.
Event Time	The time or time of day during which the sample was collected.
Collection Date Accuracy	A numerical representation of the precision of the eventDate given in days and is represented as +/- the eventDate value. Default is 0 days.
Habitat	A category or description of the habitat.
Sampling Protocol	The name of, reference to, or description of the method or protocol used during a collection event.
Collection Notes	Comments or notes about the collection event.
Site Code	The name of the sampling location.
Coll. Event ID	A optional event ID.

Table 4: Field definitions for Collection Data page on accompanying spreadsheet.



As outlined below, data can be entered on the Data Submission Template spreadsheet and sent to BOLD. Data managers will review and validate the data, ensure that it meets the minimum requirements, and upload it into BOLD.

## 1. Create Excel file submission

- New submissions are project specific, so that their data can be associated with a project on BOLD. If records need to be entered into different projects on BOLD, a separate excel file for each project needs to be created. BOLD supports the upload of multiple specimen records in a spreadsheet format.
- 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. Tables 1-4 describes the information accepted in the batch specimen data submission. Minimal information can be submitted to start and records can be updated at a later date. Figures 1-4 below illustrates example data filled into the accepted fields for Template 3.0 (This spreadsheet template is available from the online version of this protocol in the Resources tab, or at: [http://www.boldsystems.org/submissionTemplates/SpecimenData\\_v3Transitional.xls](http://www.boldsystems.org/submissionTemplates/SpecimenData_v3Transitional.xls).)
- The minimal requirements for a new specimen record on BOLD are:
  - Voucher Info Page - Sample ID
  - Voucher Info Page - Field ID and/or Museum ID
  - Voucher Info Page - Institution Storing
  - Taxonomy Page - Phylum
  - Collection Page - Country

## 2. Submit file to BOLD for processing

- Open the destination project in BOLD
- Click on "Specimen Data" under the Uploads menu and choose "Initiate Batch Submission", and select "New" for the submission type. This option is available to project managers and project users with edit specimen access.
- In the form, select the Excel file to submit to this project, along with email addresses for collaborators that should be cc'd on further communications regarding the submission, a priority level and note if needed. Then click "submit" to submit the spreadsheet for the first pass of validation.
- If there are any errors detected with the first pass of a validation, please resolve these in the submission and re-submit.
- The Data Management team will contact you if there are any issues during validation, and once the records have been uploaded to your project.

Specimen Info				
Sample ID	Field ID	Museum ID	Collection Code	Institution Storing
demo01	Sample-demo01	15466-JUC-ISC	ISC	Burke Museum

Taxonomy												
Sample ID	Phylum	Class	Order	Family	Subfamily	Genus	Species	Identifier	Identifier Email	Identifier Institution	Identification Method	Taxonomy Notes
demo01	Arthropoda	Insecta	Diptera	Asilidae	Hydro-psychinae	Efferia	Efferia aestuans	Joe Smith	jsmith@BIO.org	Oxford	Morphology	

Specimen Details										
Sample ID	Sex	Reproduction	Life Stage	Extra Info	Notes	Voucher Status	Tissue Descriptor	Associated Taxa	Associated Specimens	External URLs
demo01	Female	Sexual	Adult	Region I	Collected with predator	vouchered: registered collection	leg	Predator: Hornet	Predator: BITK002-12	<a href="http://www.burke.edu/mus/spec15466">www.burke.edu/mus/spec15466</a>

Collection Info											
Sample ID	Collectors	Collection Date	Continent/Ocean	Country	State/Province	Region	Sector	Exact Site	Latitude	Longitude	Elevation
demo01	Joe Smith	2-Jul-2009	North America	Canada	Ontario	Wellington	Guelph	Riverside Park	43.563	-80.270	325m

Depth	Elevation Precision	Depth Precision	GPS Source	Coordinate Accuracy	Event Time	Collection Date Accuracy	Habitat	Sampling Protocol	Collection Notes	Site Code	Collection Event ID
	2m		Garmin	1m	morning	2	dry forest	Malaise	park entrance	#14	#M872a

Figures 1-4: Example Specimen Data Submission.



Here are some important notes on fields for new or update submissions:

#### Sample IDs (Voucher Page):

- It is important to use a unique and original format for the Sample IDs. If the Sample IDs provided are not original on 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.

#### Collection Code (Voucher Page):

- The Collection Code must be used in conjunction with Museum ID in order to disambiguate a ID number that might be used in different collections within the same institution (i.e., a specimen number in a large museum may appear in the bird, mammal and reptile collections). This field is only to be used if Museum ID field is used.

#### Interim Species Names (Taxonomy Page):

- Interim names should contain non-Linnean characters such as numbers, punctuation and/or extra capitalization. Taxonomists are encouraged to append interim names with initials. (Example: Morpho sp. IKHR)

#### Extra Info (Specimen Details Page):

- The “Extra Info” field can be displayed on a Taxon ID Tree on BOLD and thus it is possible to include information that may aid in analysis when illustrated on a tree.

#### Collection Date and Accuracy (Collection Page):

- If there is a date range, the Collection Date should be the mid-point, and the +/- can be entered into the Accuracy field.

### What are Projects and Process IDs?

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.

## Updating Specimen Data

An update means to modify records that already exist in a project. 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.

The quickest way to update a large number of records is to download and revise the Data Spreadsheet from BOLD. To do so:

1. Click on “Data Spreadsheets” from the Downloads menu on the left side of your project, merged projects or record search. (Please note: 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.)
2. Only download the worksheets 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.*)
3. Modify the data on the downloaded worksheets. The submitted update must reflect what the data should be on BOLD.
4. Please send this to the Data Management Team through [submissions@boldsystems.org](mailto:submissions@boldsystems.org) noting the scope of the update, or submit through BOLD if it only affects one project.

If Sample IDs need to be changed after uploaded to BOLD, please contact the support team through [support@boldsystems.org](mailto:support@boldsystems.org).

## 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.

### 1. Collect Images:

Group high-quality images of specimens in .jpg format for your records. 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' 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 must be named ImageData.xls and contain the columns described in the table above. This spreadsheet template is available from the online version of this protocol in the Resources tab on BOLD, or at: <http://www.boldsystems.org/submissionTemplates/ImageData.xls>.

### 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. Please see the next page for an example of the completed ImageData.xls spreadsheet and further steps to upload.

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, 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 (not necessary).	
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: <ul style="list-style-type: none"> <li>• Copyright</li> <li>• No Rights Reserved</li> <li>• CreativeCommons - Attribution</li> <li>• CreativeCommons - Attribution Share-Alike</li> <li>• CreativeCommons - Attribution No Derivatives</li> <li>• CreativeCommons - Attribution Non-Commercial</li> <li>• CreativeCommons - Attribution Non-Commercial Share-Alike</li> <li>• CreativeCommons - Attribution Non-Commercial No Derivatives</li> </ul>	(or) c nrr by by-sa by-nd by-nc by-nc-sa by-nc-nd
License Year*	The year of license declaration (not the year of submission to BOLD).	
License 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

Image File	Original Specimen	View Metadata	Caption	Measurement	Measurement Type	Sample Id	Process Id	License Holder	License	License Year	License Institution	License Contact	Photographer
ROM912D.jpg	yes	Dorsal	skull	15 mm	skull length	ROM 10912	BMI272-03	Jane Beck	Copyright	2010	Biodiversity Institute	photo@bio.org	Joe Smith
ROM912L.jpg	yes	Lateral	lower jaw	7 mm	length	ROM 10912	BMI272-03	Jane Beck	Copyright	2010	Biodiversity Institute	photo@bio.org	Joe Smith
ROM913L.jpg	yes	Lateral	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.

**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 “Images” link in the Uploads menu of the User Console. Select the zipped folder of images to submit to BOLD. The images will appear immediately on the records once the upload is complete.

## 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.
- Select the images and spreadsheet and zip those directly. Zipping the containing folder only may not work.
- 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 or .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).
- 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](mailto:support@boldsystems.org).

## Image Licensing and Use

BOLD assumes no license for images uploaded to the database. The image owner maintains the license and may change the license on the images at any time. Revisions to the given license should move towards making 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 Creative Commons - 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 Browser Taxon Profiles - a random selection of the images on BOLD for each taxon are displayed on the public Taxonomy Browser at a highly reduced size (320 x 240) to create an online profile for each taxon that is stored in BOLD.

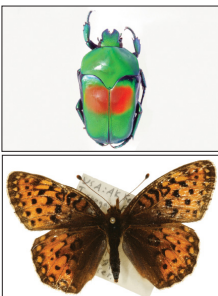

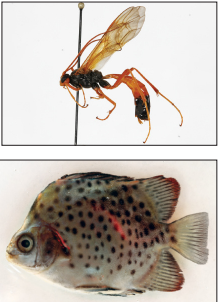



## Photography Guide

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, photos of distinguishing features or habitat are also acceptable.

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.

 <p><b>Dorsal</b></p> <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>	 <p><b>Herbarium Sheets:</b></p> <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"</li> </ul>
 <p><b>Lateral</b></p> <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>	 <p><b>Specific Features:</b></p> <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>
 <p><b>Ventral</b></p> <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>	 <p><b>Whole Specimens:</b></p> <ul style="list-style-type: none"> <li>• These are often live photographs of the organism in it's 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 orientations for photographs.

When entering a new orientation, please capitalize only the first letter and do not add any words (such as "Ventral view"). 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.

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

See page 8 for details on how to search the primer database to confirm your primer is registered with BOLD.

### 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 must be named data.xls and contain the columns described in the table to the right. This spreadsheet template is available from the online version of this protocol in the Resources tab, or at: <http://www.boldsystems.org/submissionTemplates/data.xls>

#### 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, as illustrated below.

**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 menu on the User Console. Select the zipped folder of files, and choose the institution that performed the sequencing to submit to BOLD.

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

### Tips and Troubleshooting For Trace Uploads

- To obtain the Process IDs for a set of records on BOLD, click on "Data Spreadsheets" under the Downloads menu on the left side of a project console. Download the Progress Report (Downloads>Specimen Data) to get the Process IDs that are assigned to each Sample ID submitted to BOLD.
- Primers must be registered before upload. If the primers are not registered, there will be an error.
- 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.
- Select the traces and spreadsheet and zip those directly. Zipping the containing folder only may not work.
- 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 (.ab1, .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 downloaded or viewed BOLD until processed (up to 24 hours after they have been submitted).

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

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

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](mailto:support@boldsystems.org).



## Workbench: Sequence Submission

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

### 1. Assemble Package:

The sequence submission package should consist of aligned sequences in FASTA format referenced by BOLD Process IDs or 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 "Sequences" in the Uploads menu of the User Console. Select the marker and the institution that assembled the sequences. Paste the sequences into the text box. When confirmed, "submit" to upload the sequences. These will appear immediately on the records.

- 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](mailto:support@boldsystems.org) for batch deletions.

**BOLDSYSTEMS**

### Sequence Submission

Required fields for submission are marked with a red asterisk.

Select ID:

Markers:

Run site\*:

To select an institution, begin to type in the name. BOLD will return matches in a drop down box and clicking on a name will select it.

Select institution responsible for generating the sequences being uploaded. If the institution is not listed, please register it by clicking on the button below

[Add New Institution](#)

**PASTE SEQUENCES IN FASTA FORMAT:**

```
>BOLDX001-12
AACATGTACCTAATTTTGGGGCTTGAGCAGCTACTAGGAACAGCAGCTAG
GGTATCATTCGGGCTGAGTAGGTCAACAGGCTCATTAATGGGAATGACCA
AATTATATACCAATTTGAGCGGCGCATGCCCTTATTATATTTTCTTATGGTAT
ACCTATCATATCGGAGGATTGGGAACTGGTTAATCCTTTGATACCTCGGGC
CCCTGATATGGCTTTTCCCGGATAAATAACAAGATTTTGGCTCTCTCCTCT
TCATTAACTCTCTTATATTTGGGATACAGAGGGGGGGGGCGGTACAGGT
TGAACAGATATCCTCCCTCAAGGAATTTCTCACTCCGGAGCAAGGTT
GATCTACCAATTTTCACTTCAATTAGCAGGTGCTAGCTCTATTCTAGAGCTA
TTAAGCTTATCACTACCATATATATATACGAGCAGCAAGAAATATCCCTAGACCGA
ATTCC
>BOLDX002-12
TACTTTGTATTAAATTTTGGAGGTTGATCAGCTAGTTGGCACTGCTTAAGA
GTATATATCGAGGAGCTGCTGCAACAGGCTCACTATTGGAAAGCATCAA
ATCTATAACACTATTGTCACTGCTCATGCTATTGTATATTTCTTATAGTGATA
CCGATCCTGATCGGGGCTTCGGTAAGTGGCTAGTCTCTTATGCTTGGGGC
```

[Submit](#) [Cancel](#)

Sequence Upload Form

## Workbench: Sequence Editor

**NEW!** The online sequence editor is a new feature in BOLD version 3.6. This tool provides the ability to assemble contig sequences from good quality bi-directional trace files. Further, this tool allows users to edit base calls, and upload the sequence directly to the record.

Once trace files have been uploaded to a BOLD record, a user with sequence editing permission can select the traces and choose the Sequence Editor button from the Sequence Page (see page 12 for an illustration of this page).

The Sequence Editor application, illustrated to the right, aligns the forward and reverse traces and generates the contig sequence. Base calls may be inspected by hovering over a peak, which highlights the corresponding peak and nucleotide. The quality of each peak is disseminated by the colour gradient right above the nucleotide bases, with green representing a high quality call and red representing a high degree of ambiguity. Users can choose to edit base calls by clicking on a nucleotide which opens the edit base dialog. The save button in the top right corner allows the user to save the sequence to the BOLD record.



Sequence Editor

## Workbench: Primer Submission

New primers need to be registered on BOLD prior to submitting a trace file package. To register new primers, select “Register Primers” from the User Console. Please note: If a primer sequence has already been registered under a different code, you will be provided with the registered code to be used in your submission. Primers you register on BOLD can be edited at any time after they are created (e.g. to make them public).

Primer Code	Create a code for the 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	Other known code names for the 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	Any 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.

## Workbench: Bibliography Submission

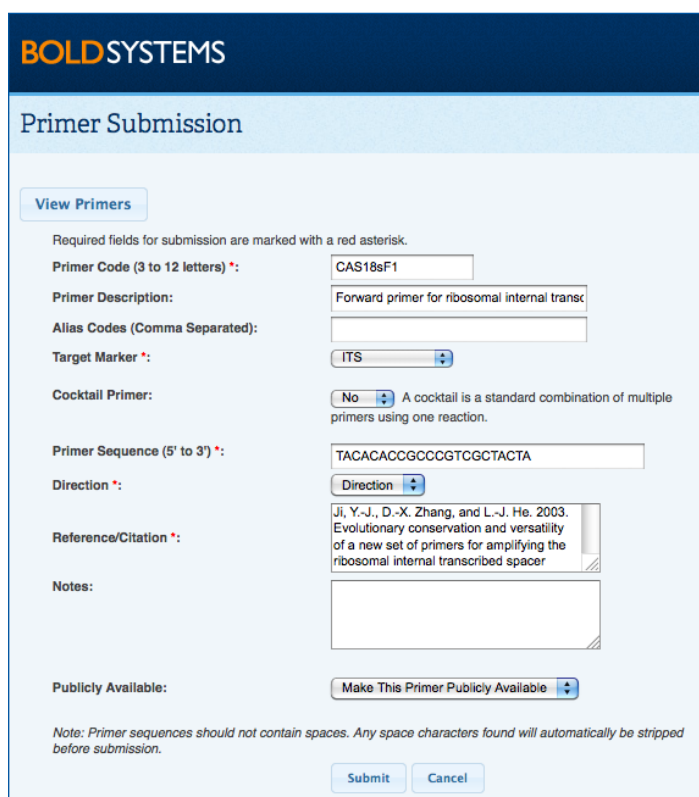
Users can submit bibliographies to BOLD using the Bibliography Submission Form available in three locations:

- From the User Console under Data Uploads
- From a Project Console or Dataset Console
- From a list of searched records (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. These accessions will auto-populate in the form for the records selected if the submission is from the Project Console or Record List. (See page 27 for details on submitting records for GenBank Accessions via BOLD)

The publication details will then appear in the Publication Database, and the citation will appear on each record.



**BOLD SYSTEMS**

### Primer Submission

[View Primers](#)

Required fields for submission are marked with a red asterisk.

Primer Code (3 to 12 letters) \*: CAS18sF1

Primer Description: Forward primer for ribosomal internal transcribed spacer

Alias Codes (Comma Separated):

Target Marker \*: ITS

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

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

Direction \*: Direction

Reference/Citation \*: Ji, Y.-J., D.-X. Zhang, and L.-J. He. 2003. Evolutionary conservation and versatility of a new set of primers for amplifying the ribosomal internal transcribed spacer

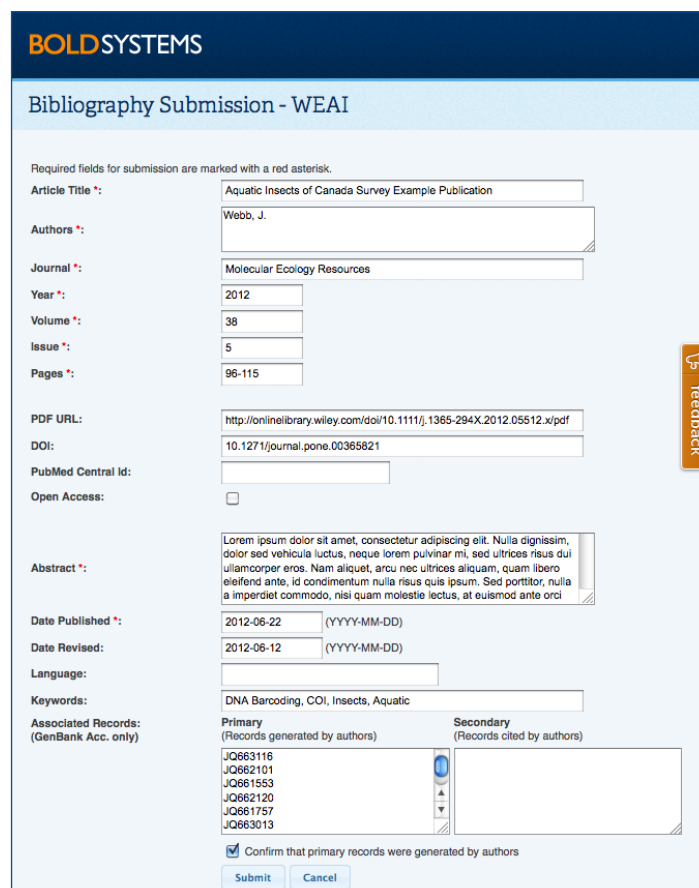
Notes:

Publicly Available: Make This Primer Publicly Available

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

[Submit](#) [Cancel](#)

BOLD Primer Submission Form



**BOLD SYSTEMS**

### Bibliography Submission - WEAI

Required fields for submission are marked with a red asterisk.

Article Title \*: Aquatic Insects of Canada Survey Example Publication

Authors \*: Webb, J.

Journal \*: Molecular Ecology Resources

Year \*: 2012

Volume \*: 38

Issue \*: 5

Pages \*: 96-115

PDF URL: http://onlinelibrary.wiley.com/doi/10.1111/j.1365-294X.2012.05512.x/pdf

DOI: 10.1271/journal.pone.00365821

PubMed Central Id:

Open Access: ☐

Abstract \*: Lorem ipsum dolor sit amet, consectetur adipiscing elit. Nulla dignissim, dolor sed vehicula luctus, neque lorem pulvinar mi, sed ultrices risus dui ullamcorper eros. Nam aliquet, aro nec ultrices aliquam, quam libero eleifend ante, id condimentum nulla risus quis ipsum. Sed portitor, nulla a imperdiet commodo, nisi quam molestie lectus, at euismod ante orci

Date Published \*: 2012-06-22 (YYYY-MM-DD)

Date Revised: 2012-06-12 (YYYY-MM-DD)

Language:

Keywords: DNA Barcoding, COI, Insects, Aquatic

Associated Records: (GenBank Acc. only)

Primary (Records generated by authors)

Secondary (Records cited by authors)

JQ663116  
JQ662101  
JQ661553  
JQ662120  
JQ661757  
JQ663013

☒ Confirm that primary records were generated by authors

[Submit](#) [Cancel](#)

BOLD Publication Submission Form

## Workbench: User Console

The BOLD User Console is the landing page for users who have logged in. This console allows for rapid access to frequently accessed projects, near real-time reports on project activity as well as powerful search tools.

### Data Uploads

Upload sequences, traces, images, primers, and bibliographies directly from user console.

### Real-time Activity Reporting

BOLD maintains a detailed log of all the actions taken by users in the system. This log is transformed into a report where events pertinent to each user are extracted and displayed on this console. As users work together on projects, submitting and refining data, the Activity Report allows them to stay informed on the steps taken by their colleagues. Logs can be downloaded, allowing users to keep personal records and perform additional analysis. Users can also filter recent activity for a particular user or action by using the search field at the top of the Most Recent Activities panel.

### Searching

The newest version of BOLD introduces an integrated search bar accessible from the User Console, Project Console, and Project List. It includes a project search, where a user can jump directly to a project by entering the code in the project search field. If the code is not known, the user can generate a short list by entering a project tag or part of the project title.

The second search functionality, available by clicking on Records: "Search", opens the workbench Record Search interface, which allows users to search for 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. See the next page for further details on the Record Search for the workbench.

### Datasets

The User Console provides access to the Dataset framework, including the options to create new Datasets and view existing public datasets. For more details, please see page 26.

**User Console**

Project Search:  Code Tag Title Records: Search

**Project Management**

Full Project List List all the projects you have access to

New Project Create a new project

**Your Data**

Campaign Name	Records You Manage	Records You Have Access To
Arctic Life (PROBE)	0	95 [1 projects]
Barcoding Applications	26 [1 projects]	39 [3 projects]
Barcoding at the African Center (ACDB)	0	2781 [18 projects]
Educational Barcoding Initiatives	0	0
Finnish Barcoding of Life	0	3407 [10 projects]
General Projects	3738 [9 projects]	11929 [40 projects]
Lepidoptera of NA Phase II	0	1735 [2 projects]
Marine Life (MarBOL)	0	1471 [17 projects]
Peats and Invasive Species	0	1116 [2 projects]
Plants of Canada	0	10935 [37 projects]
Problematic Samples	98 [6 projects]	1229 [6 projects]
WG1.2 Land Plants	14943 [19 projects]	17335 [57 projects]
WG1.8 Marine Bio-Surveillance	0	34075 [137 projects]
<b>Total Count</b>	<b>18805 [35 projects]</b>	<b>86147 [331 projects]</b>

**Data Uploads**

Sequences Traces Images Primer Bibliography

**Projects You Recently Accessed**

Code	Title	Specimens	Sequences
BCIL	Barro Colorado Island Lianas	716	matK:327, trnH-psbA:329, rbcLa:390
SEQLD	SE QLD rainforest DNA barcoding project	0	matK:0, trnH-psbA:0, rbcLa:0
ACCCX	Five critical groups of the Mexican flora - TRASH	273	rbcL:2, matK:1, trnH-psbA:12
MHPAC	Plants of ACG3 (in progress)	7149	matK:3190, rbcLa:5010
MEDPL	Medicinal Plants	1824	matK:0, trnH-psbA:0, rbcLa:0
WABLK	Wabikon Lake Forest Plot	815	matK:0, trnH-psbA:54, rbcLa:592
ANGMD	Angraecums of Madagascar	475	ITS2:111, matK:313, rbcLa:360
MKCRA	Crassulaceae in Mexico	94	ITS2:90, matK:93, rbcLa:94
NNOR	Northern Norwegian Vascular plants	1140	ITS2:626, matK:0, rbcLa:871, rbcL-like:2
IDRCG	Trees of Ghana - project IDRC	1519	ITS2:0, matK:1075, rbcLa:1271
SAFH	Southern Africa Flora Holding Project	2347	matK:1427, rbcLa:1798

**Your Datasets**

New Dataset

**Public Datasets**

**Frequent Collaborators**

Person	Shared Projects
Gary Saunders	154 projects
Tanya E. Moore	154 projects
Daniel C. McDevitt	93 projects
Kyatt R. Dixon	50 projects
Steven G. Newman	31 projects
Ragupathy Subramanyam	30 projects
Anuar Rodrigues	29 projects
Alex V. Borisov	27 projects

**Most Recent Activities:**

Show 25 entries Search:

Timestamp	Who	Action	Project
Oct-15, 2013 12:23	Gary Saunders	Delete-Sequences (1)	ABMMC
Oct-15, 2013 12:22	Gary Saunders	Modify-Sequences (2)	ABMMC
Oct-15, 2013 12:22	Gary Saunders	Modify-Sequences (3)	GIGAR
Oct-15, 2013 10:59	Bruce A. Bennett	Modify-Specimens (1)	BBYUK
Oct-15, 2013 10:57	Bruce A. Bennett	Modify-Specimens (1)	BBYUK
Oct-15, 2013 10:40	Gary Saunders	New-Sequences (1)	GIGAR
Oct-14, 2013 10:57	Masha L. Kuzmina	Modify-Sequences (25)	BCIL
Oct-14, 2013 10:09	Masha L. Kuzmina	Delete-Sequences (14)	BCIL
Oct-14, 2013 00:12	Masha L. Kuzmina	New-Sequences (341)	BCIL
Oct-11, 2013 16:38	Anuar Rodrigues	New-Traces (190)	FBPL
Oct-11, 2013 16:37	Anuar Rodrigues	New-Traces (95)	FBPL
Oct-11, 2013 16:36	Anuar Rodrigues	New-Traces (95)	FBPL

User Console for an example user

**NEW!** In the Project Search bar (available on the User Console, Project Console and Record List), users can search for multiple projects at once by entering more than one project code, separated by commas. For example, searching "NEOSI,LHSMI" will merge those two projects together and provide a consolidated project console.

## Workbench: Record Search

The record search interface in the BOLD workbench allows for rapid access to large numbers of records by searching across a number of parameters.

To search across all records that a user has access to, select the “Search Records” button in the User Console or Project List. The form also gives the option to add in all public records on BOLD. Within a Project Console or Record List page, the search allows users to choose from searching within the project, or across the entire workbench.

**NEW!** The newest version of BOLD includes several new search parameters, including specimen depository, collection date ranges, and GPS-based polygons.

The results from the search will appear in a Record List like the one illustrated on page 25.

Taxonomy	Searches specific taxonomic names. Use quotes to surround Species name and separate multiple taxa by spaces. Example: “Bos taurus” “Bos indicus”
Geography	Searches the country and province names. Use quotation marks to surround multi-word names and separate multiple terms by spaces. Example: “Costa Rica” Mexico “United States”
Marker and Sequence Length	Filter results by the presence of a sequence for a specific marker, as well as defined length range.
Tags	Enter terms found in tags on records
Depository	Accepts depository institution names.
Extra Info	Accepts terms that appear in the “Extra Info” field on specimen records.
Collection Dates	Allows for a date range to be specified.
Region of Map	Searches for records with GPS coordinates, which are found within a user-defined polygon area on the map. A polygon can consist of many points but it must contain at least 3 points on the map and be close-looped. The returned results will have all of the records which have GPS coordinates within the polygon and to which the user has access to.
List of Identifiers (2nd Tab)	Search by Sample IDs, Process IDs, GenBank accessions and BIN URIs. Can be used in conjunction with the above fields to return the intersection of records.
Search within project	If accessed from within a project, the record search will remain restricted to that project unless the checkbox is unchecked.
Include public records	Checking the box at the bottom will allow the search to go across all public data on BOLD, as well as the records in projects the user has direct access to.

Search Parameters Available

Record Search window, showing a polygon defining the boundaries of a National Park.



## Workbench: Creating a project

To upload Barcode records to BOLD, a project must first be created in which to house the records. 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.

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 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 records)</li> <li>Folder Project (contains other projects)</li> </ul>
Primary Marker*	Select the primary marker. Primary marker options are: COI-5P, rbcl, matK, ITS, 18S-3P, 18S-V4, trnH-psbA
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	User-defined or controlled tags may be selected.
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 Public 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 access levels are described below. <p><b>Sequence Access:</b></p> <ul style="list-style-type: none"> <li>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).</li> <li>View &amp; Download - user can view or download the sequence data, as well as analyze.</li> <li>Edit Sequences - user can upload trace files, upload, edit and delete sequences, as well as view and analyze.</li> </ul> <p><b>Specimen Access:</b></p> <ul style="list-style-type: none"> <li>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.</li> </ul> <p><b>NEW! Dataset Access</b></p> <ul style="list-style-type: none"> <li>Add to Dataset - users can add records to datasets, which can be used for analysis or publication purposes.</li> </ul>

Field definitions for BOLD project creation form.

\* Required Fields

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

### New Project Submission

Required fields for submission are marked with a red asterisk.

Project Title \*:

Project Code \*:  (A unique 3-5 letter code)

Project Type: ☒ Data Project (contains specimen & sequence records) ☐ Folder Project (contains other projects)

Primary Marker \*:  (COI is the only marker that holds BARCODE status in GenBank at this time)

Supporting Marker(s):

Campaign:

Place in container:

Tags:  or enter tag:

Project Description \*:

Bounding Box: Top Left:  Lat:  Lon:  Bottom Right:  Lat:  Lon:

Project Access: ☐ Make this project publicly visible (All BOLD users will be allowed to view, analyze, and download sequence data)

Project Manager:

Assign Users:	User	Sequence Access	Specimen Access	Dataset Access
<input checked="" type="checkbox"/>	Rodriguez Anuar	<input checked="" type="checkbox"/> Analyze only <input checked="" type="checkbox"/> View, Download <input checked="" type="checkbox"/> Edit	<input checked="" type="checkbox"/> Edit Specimens	<input checked="" type="checkbox"/> Add to Dataset
<input checked="" type="checkbox"/>	Bennett Bruce A.	<input checked="" type="checkbox"/> Analyze only <input checked="" type="checkbox"/> View, Download <input checked="" type="checkbox"/> Edit	<input checked="" type="checkbox"/> Edit Specimens	<input checked="" type="checkbox"/> Add to Dataset
<input checked="" type="checkbox"/>	Hades Peter	<input checked="" type="checkbox"/> Analyze only <input checked="" type="checkbox"/> View, Download <input checked="" type="checkbox"/> Edit	<input checked="" type="checkbox"/> Edit Specimens	<input checked="" type="checkbox"/> Add to Dataset
<input type="checkbox"/>		<input type="checkbox"/> Analyze only <input type="checkbox"/> View, Download <input type="checkbox"/> Edit	<input type="checkbox"/> Edit Specimens	<input type="checkbox"/> Add to Dataset
<input type="checkbox"/>		<input type="checkbox"/> Analyze only <input type="checkbox"/> View, Download <input type="checkbox"/> Edit	<input type="checkbox"/> Edit Specimens	<input type="checkbox"/> Add to Dataset

New Project Form

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 required is not on the list, please contact BOLD support staff to register one through support@boldsystems.org.





## Workbench: Datasets and DOIs

Managing data with projects on BOLD is very efficient for coordinating sequencing activities and digitization of specimen records. Projects allow for the effective management of permissions by ensuring that any single record can exist only in one project. Datasets extend this functionality by allowing the creation of named collections of records by taxonomic, geographic or thematic criteria.

Any user can create datasets and add records they have Dataset Level access to. A record may appear in multiple datasets, but is still linked to the original record. Changes made to the original record are automatically synchronized with those in datasets. A single dataset can hold up to 25,000 records, for assembly of large collections for publication in GenBank or public release. In addition to providing accessions through GenBank for records in a dataset, BOLD, through a partnership with DataCite, can provide a single DOI for each dataset. The assigned DOI can be used in associated manuscripts in place of supplementary tables and can be referenced when used by others.

### Creating A New Dataset

Once logged into BOLD, select the New Dataset button in “Your Datasets” section of the User Console. By choosing to make the dataset public, the user is provided with the opportunity to request a DOI.

The screenshot shows the BOLD SYSTEMS 'New Dataset Submission' form. The form includes fields for Dataset Title, Dataset Code, Dataset Description, Dataset Access, and Bounding Box. A 'Publish Confirmation' modal is open on the right, asking for confirmation to make the dataset public and offering the option to request a DOI.

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

### New Dataset Submission

Required fields for submission are marked with a red asterisk.

Dataset Title \*:

Dataset Code \*: DS-  (A unique 4-8 character alphanumeric code)

Dataset Description \*:  (15 characters minimum)

Dataset Access: ☒ publicly visible  
(All BOLD users will be allowed to view, analyze, and download sequence data)

Bounding Box: **Top Left:** Lat:  Lon:   
(For geographically oriented datasets) **Bottom Right:** Lat:  Lon:

**Publish Confirmation**

You have elected to make this dataset public. Please confirm that you want to proceed.

If public, this dataset will be accessible in the BOLD Public Data Portal within 48 hours. The records in this dataset can be retrieved using the following permanent URL: [http://www.boldsystems.org/index.php/Public\\_SearchTerms?query=<DATASET CODE>](http://www.boldsystems.org/index.php/Public_SearchTerms?query=<DATASET CODE>).

BOLD now supports minting of **Digital Object Identifiers (DOI)\*** for datasets so that they can easily be used in publications and subsequently cited. Please request a DOI if this feature would be useful.

*\* DOI's are persistent identifiers used to identify and cite publications but are also very useful in citing data directly.*

The BOLD Dataset creation form with DOI request shown.

### Adding And Removing Records

Users can search for and add records to a dataset in batches of up to 2,500 records at a time from a record list page.

**NEW!** The ability to remove records from a Dataset is now available in the Record List for a particular dataset. Users can select individual records to remove from the dataset. For further assistance with BOLD Datasets, please email support through support@boldsystems.org.

## Workbench: Publication of Records on BOLD and GenBank

### Publication through BOLD

To submit records to the BOLD Public Data Portal, the Project Manager can make a Project or Dataset public. This option is available in the Project Properties form which the Project Manager can access by clicking on “Modify Project Properties” within a Project or Dataset.

Once GenBank accessions have been obtained, users can submit a bibliography to the BOLD database following the directions on page 21. This will associate the publication with the records using the GenBank Accession numbers. This citation will appear on the records in the workbench and the Public Data Portal.

### Publication to GenBank, NCBI

BOLD shares a tightly integrated data exchange pipeline with NCBI (GenBank) that allows for the automatic submission of data to GenBank. Users are only required to fill in the author and publication information and which is sent to GenBank along with the specimen, sequence, and trace data which has been transformed to the required formats. GenBank responds directly to the user with the accessions for their records to be included in publications. Accessions are also sent to BOLD to ensure bidirectional linkage.

The data exchange pipeline is further utilized to send GenBank updates to records. 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.

To submit a set of sequences to GenBank, the Project Manager can access the form shown to the right via the “Submit to GenBank” button within a Project or Dataset Console.

All records within the project 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 or copied to a BOLD Dataset which can be submitted.

GenBank accession numbers are generally returned by email to the corresponding author within five business days. The accession numbers will be associated with the records on BOLD once they are available.

Submissions to GenBank from BOLD 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.

Contact the BOLD support team through [support@boldsystems.org](mailto:support@boldsystems.org) for assistance with any aspect of publication.

**NEW!** To access the list of GenBank accessions for a set of records on BOLD, download the Progress Report spreadsheet from the Downloads Menu, under Data Spreadsheet. Alternatively, select the “Summary - Specimens, Localities, and GenBank” within a Project Console or Record List, to view accessions.

**BOLDSYSTEMS**

**Project Publication - KALLY**

Required fields for submission are marked with a red asterisk. Please use only English characters.

Publication Title \*: A comparison of two DNA barcode markers for species discrimination in the red algal family Kallymenia

Author List (\* At least one author is required)

First Name	M.I.	Last Name	Suffix
Brigette	E.	Clarkston	
Gary	W.	Saunders	

[Add Author](#)

Corresponding Author (Must also be in author list)

First Name *: Brigitte	Phone *: 495-555-1548
M. I.: E.	Fax *: 495-555-1549
Last Name *: Clarkston	Email *: bclarks@unb.ca
Suffix:	Street Address *: 50 Main Street
	City: Fredericton
Institution *: University of New Brunswick	State / Province *: New Brunswick
Department *: Botany Department	Zip/Postal Code *: T2H4Q1
	Country *: Canada

Please double check your data before requesting submission to GenBank. Your submission request will put in queue for final quality checks and will be forwarded to GenBank once validated.

☐ Submit an update of these records to GenBank if they already have accessions.

[Submit Request](#) [Close](#)

GenBank Submission Form

\* Required Fields

## 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
- Barcode Gap Analysis: Evaluate the Barcode gap
- Accumulation Curve: Review sampling efficiency
- Alignment Browser: Diagnose unaligned sequences
- Diagnostic Characters: Examine polymorphism
- BIN Discordance Report: Utilize BINs to highlight possible issues

### Tip: Run Multiple Analytic Tools in Parallel

Use the email analysis option to run multiple analyses and have the results emailed 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 a project, it is possible to 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 via the Image Library in order to view a group of specimens, as shown below.

The Image Library displays one image per specimen and allows users to choose a specimen orientation view in order to compare morphological differences between specimens. This tool is useful for diagnosing contamination or misidentifications as taxonomy is displayed below each image.

**NEW!** The specimens can be displayed sorted by identification or BIN. In the general options, users can also select how many images to display per row and per page.

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

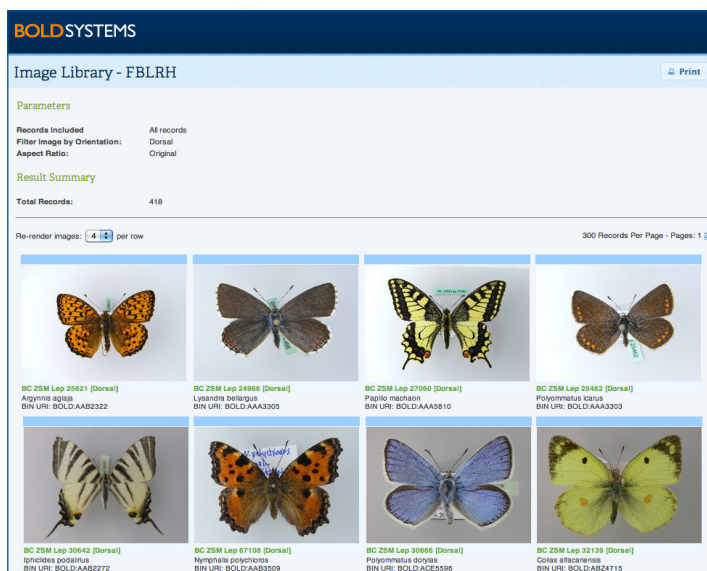


Image Library for Lepidoptera (Displayed Specimen Images: Copyright 2010, Bavarian State Collection of Zoology)



## Workbench: Distribution Map Analysis

The Distribution Map plot the collection points for a selected set of specimens when geographic reference data is available. There are two versions available for use, including the Interactive Map which is new in BOLD version 3.6.

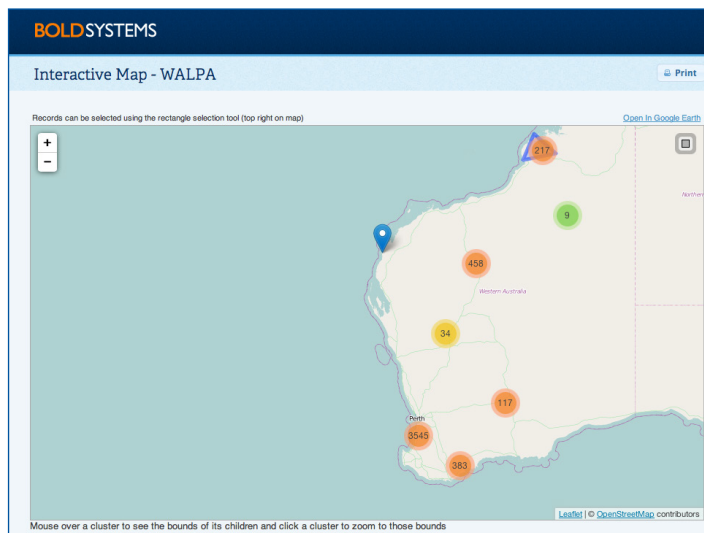
### NEW! Interactive Map

GPS points are clustered on the map in coloured circles, and single records are signified by a blue marker. The circles indicate that there are multiple GPS points clustered there - hovering over these circles will display the bounds of the cluster. Green circles indicate that there is more than 1 point, yellow circles have more than 10 points, and orange circles have more than 100 points.

The square symbol in the upper right corner of the map is a search tool that allows users to draw a rectangle on the map to select all the records that fall within that region. The records can then be searched on BOLD in order to allow users to download or run further analysis on that subset of records. It also allows users to edit records to quickly correct GPS points, if they are found to be incorrect.

Users can zoom in out or on the map using their mouse scroll wheel, or with the plus/minus buttons provided in the upper left hand side. Clicking on one of the GPS cluster markers will zoom in to that cluster.

The Interactive Map is built on the open-access platform OpenStreetMap and therefore can be published, re-used, and modified at the user's discretion.

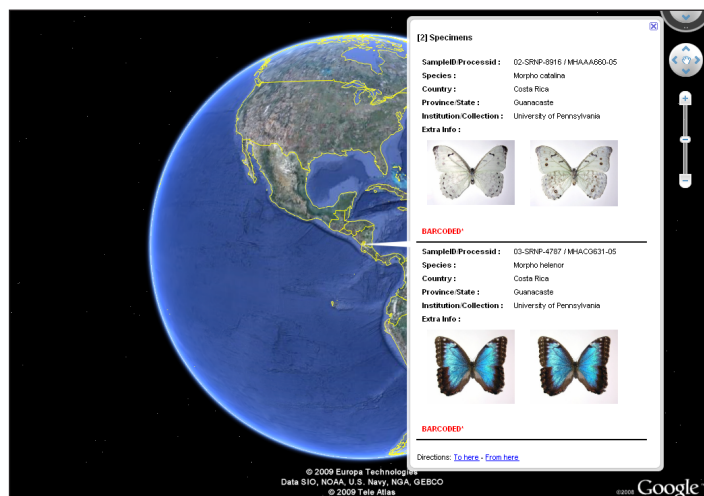


Interactive Map

### 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 Google Earth map is a portable KML file download which can be shared among colleagues. It is available by clicking on the "Open in Google Earth" link at the top of the Interactive Map. This file has embedded specimen images, along with specimen identifiers, country, province/state, institution/collection information, and the Extra Info field.



Google Earth Map featuring BOLD records

## Workbench: Taxon ID Tree

The Taxon ID Tree functionality allows for the generation of dendrograms from sequencing using the Neighbour Joining algorithm. Sequence alignment and multiple labelling options are available. This tool can be accessed by clicking on "Taxon ID Tree" under the sequence analysis panel on the Project Console and Record List pages.

Result of the analysis include a tree in PDF format for easy dissemination and publication, Newick format for rendering in other formats and layouts (i.e. circular), and a taxonomy report that provides a breakdown of the data used. When the option to include matching images and spreadsheet is selected, the system returns a spreadsheet with specimen identifiers and full taxonomy in the same order the records appear in the tree as well as the images of the specimens in the same order as the tree. Utilizing these options help catch data entry and lab errors where samples may be mixed up. They are also useful for detecting misidentified specimens.

**NEW!** BOLD now provides the ability to built larger Taxon ID trees, up to 12,000 COI sequences. For other markers and unaligned sequences, trees are limited to 3,000 sequences.

Sequence Data	Nucleotide or Amino Acid for coding markers
Distance Model	Kimura 2 Parameter, Jukes Cantor, Pairwise Distance distance models are available.
Tree Building Method	Neighbour Joining is the only method at this time.
Marker	Select the marker from a multi-gene dataset.
Alignment Options	Users can utilize the BOLD aligner (amino-acid based HMM), Kalign, MUSCLE or use unaligned sequences
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.
Filters	Can be applied to disregard sequences below a given length (since very short sequences can skew the results), omit sequences with more than 1% ambiguous bases, or to exclude problematic sequences.
Colour	By default, problematic sequences are highlighted (blue for stop codons and red for contaminations or user-applied flags). Other colourization options include: taxonomy, geography, extra info, sequence age and BINs.
Ambiguous Base/ Gap Handling	Complete or Pairwise Deletion option.
Result Options	Choose to view the results immediately or to have the results emailed when completed.
<b>NEW!</b> Page Orientation	Trees can now be displayed in landscape orientation.

Parameters available for Taxon ID Tree

Sequence Data:

Nucleotide

Distance Model:

Kimura 2 Parameter

Tree Building Method:

Neighbor Joining

Marker:

COI-5P - Cytochrome Oxidase Subunit 1 5' Region

Align Sequences \*:

BOLD Aligner (Amino Acid based HMM)

Select Terminal Branch Labels:

☐ Extra info  
☐ Phylum ☒ Family  
☐ Class ☐ Subfamily  
☐ Order ☐ Tribe  
☐ No Taxonomy

Geography:

☐ Country & Province/State ☐ Sector  
☒ Region ☐ Exact Site  
☐ No geography

☐ 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  
☐ BIN URI ☒ No BIN details

Apply Filters:

☐ Matching specimen photographs and spreadsheet  

Sequence Length > 200 bp

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

Colorize Tree Based on:

Highlight Problematic Records

Ambiguous Base/Gap Handling:

☒ Pairwise Deletion ☐ Complete Deletion

Result Options:

Email me when the results are available

Store the results for 3 Days

Page Orientation:

☒ Portrait ☐ Landscape

Build Tree

Taxon ID Tree Parameter Page



Standard Taxon ID Tree (Lepidoptera)

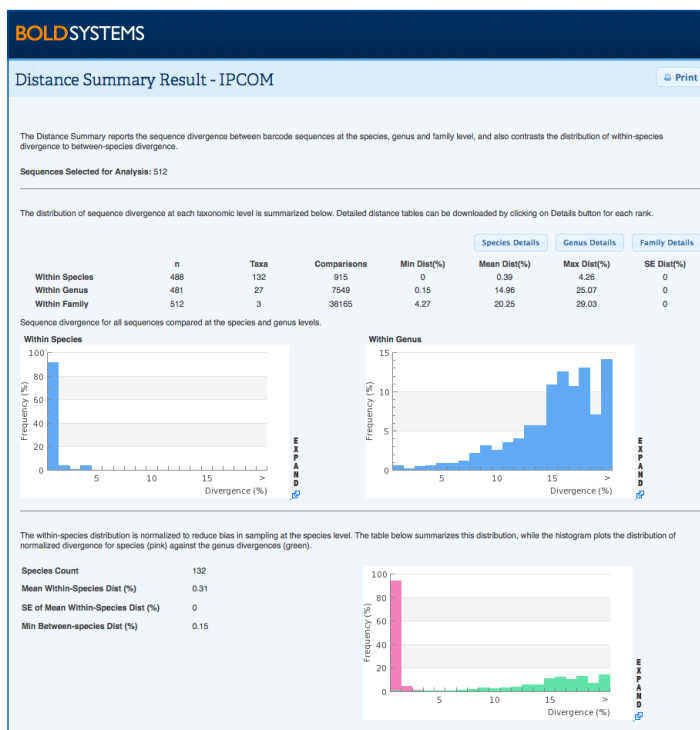
## Workbench: Distance Summary

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 conspecific and congeneric levels.

Distance values are calculated using a user selected distance metric. 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 levels of species, genus and family are available by clicking on the links in the top right corner of the Distance Summary results.

### NEW! Complete Deletion

BOLD 3.6 includes the option to analyze sequences using Complete Deletion for handling ambiguous bases and gaps. This is available in the Distance Summary, Barcode Gap and Alignment Browser.

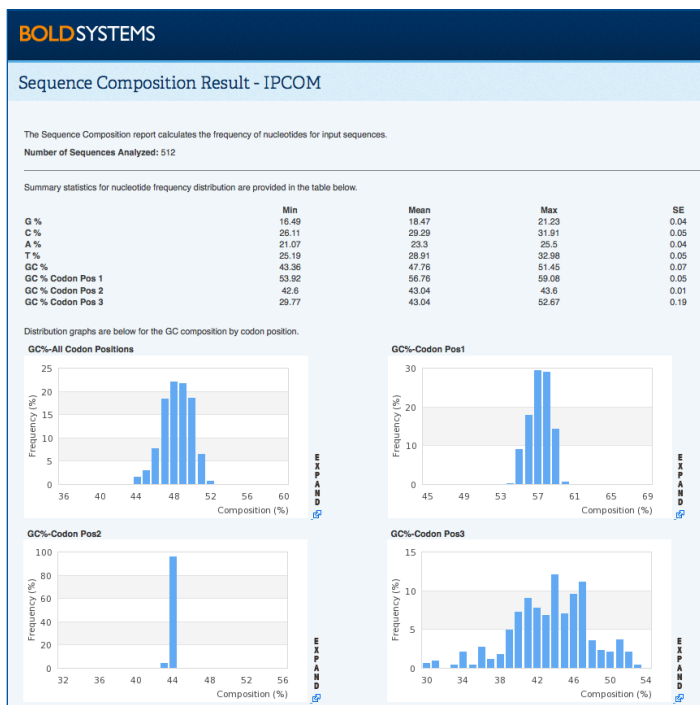


Distance Summary Analysis

## Workbench: Sequence Composition

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.

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.



Sequence Composition Analysis





## Workbench: Diagnostic Characters

The Diagnostic Character analysis provides a means to examine nucleotide or amino acid polymorphism between sets of sequences that are grouped by taxonomic or geographic labels. More specifically, this tool identifies consensus bases from each group, compares them to those from the remaining sequences in other groups, then characterizes how unique each consensus base is. The tool categorizes consensus bases by their diagnostic potential.

Changes in the input sequences, filtering options or the analysis parameters can affect the consensus sequences in each group and the diagnostic potential will be different between analyses. In general, having more sequences per group will provide a more accurate diagnosis of each group, as it reduces the problem caused by small sample size.

**Diagnostic Character Result - WWGSL**

Analyses Requested: Group sequences by species

Analyses Parameters:

- Sequence data type = nucleotide
- Marker = Cytochrome Oxidase Subunit 1 5' Region
- Alignment = original alignment
- Minimum number of sequences per group = 3
- Codon of interest = 1, 2, 3
- Excludes a sequence if:
  - has < 400 bases

Analyses Performed: Note: sequences are retrieved if ALL data required for their analyses are present.

Number of sequences initially retrieved for analysis = 460  
Number of sequences included (after filtering) for analysis = 460

All analyses requested were completed with results. Please choose the options below to obtain the results.

Detailed Results: [species](#) [Display](#) [Download Report](#)

**Summary Results**

Group Name (# sequences)	# Diagnostic Characters	# Diagnostic or Partial Characters	# Partial Characters	# Partial or Uninformative Characters	# Invalid Characters
Americocheilia longicornis (10)	0	0	1	0	0
Americocheilia megastipitata (4)	1	0	1	0	0
Ampelicia eschrichtii (12)	3	0	0	0	0
Amphioxe longimana (3)	0	0	0	0	0
Amphioxe rubricata (5)	0	0	0	0	0
Anonyx rugax (10)	0	0	0	0	0
Anonyx sarai (13)	1	0	0	0	0
Argis dentata (5)	0	0	0	0	0
Boreomyia arctica (5)	0	0	0	0	0
Callipus laeviusculus (29)	0	0	0	0	0
Cancer inornatus (8)	0	0	0	0	0
Caprella insularis (8)	1	0	0	0	0
Caprella mutica (4)	0	0	0	0	0
Caprella septentrionalis (12)	1	0	1	0	0
Carcinus maenas (8)	0	1	0	0	0
Chironomites opilio (6)	0	0	1	0	0

Diagnostic Characters Analysis

## Workbench: BIN Discordance Report

The Barcode Index Number module analyzes new COI sequences and assigns them to an existing or a new BIN. Besides generating BIN pages, this system acts as a rapid check of the validity of taxonomic designation on specimen records. The BIN Discordance report facilitates this check by comparing the taxonomy on selected records against all others in the BINs they are associated with.

The results are sorted by the degree of conflict, displaying those records in BINs where there is a phylum level conflict first (likely the result of cross-contamination) down to species level conflicts. Users can pull up records from this page to examine ancillary data or to edit the taxonomy where there is an error.

**BIN Discordance Report - WWGSL**

Overview

The Barcode Index Number system clusters COI sequence data into OTUs (called BINs) independent of prior taxonomic assignment. As such, it provides a means of confirming the concordance between barcode sequence clusters and species designations. This report performs this validation by comparing the taxonomy on input records against all others in the same BINs, including those submitted and managed by other users.

Data Selected: **460** Records **458** Records with BINs Representing **86** BINs

Breakdown of Data:

	BINs	Records
Taxonomically Discordant:	15	77
Taxonomically Concordant:	70	380
Singletons:	1	1

Detailed Results

Discordant BINs are sorted by rank of discordance (phylum level conflicts at the top and species level at the bottom). Concordant and singleton BINs are separated into secondary tabs. Member records for a BIN can be retrieved by selecting the checkboxes next to Process IDs of interest and clicking "Record Search".

Discordant BINs Concordant BINs Singletons

**BINs with Taxonomic Discordance**

Record Search

Process ID	Identification	Conflicting Taxon in BIN	Rank of Conflict	BIN	BIN Total Members	BIN Tax Variation
<input type="checkbox"/> WW561-08	Carcinus maenas	Arthropoda	Phylum	BOLD:AAA7687	48	Arthropoda[125], Annelida[1]
<input type="checkbox"/> WW562-08	Carcinus maenas	Arthropoda				
<input type="checkbox"/> WW564-08	Carcinus maenas	Arthropoda				
<input type="checkbox"/> WW565-08	Carcinus maenas	Arthropoda				
<input type="checkbox"/> WWCSL137-08	Carcinus maenas	Arthropoda				

BIN Discordance Report

Last modified: October 2013

**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 ©2013 - All Rights Reserved  
BOLD Systems, Biodiversity Institute of Ontario

Authored by: Megan Milton ([mmilton@boldsystems.org](mailto:mmilton@boldsystems.org))  
Paola Pierossi ([ppierossi@boldsystems.org](mailto:ppierossi@boldsystems.org))  
Sujeewan Ratnasingham ([sratnasi@boldsystems.org](mailto:sratnasi@boldsystems.org))

Designed by: Suzanne Bateson, Biodiversity Institute of Ontario