Training 6:
Process of submitting sequences to GenBank & making them Public on BOLD

By D. Jean Lodge

We aim to make sequences publicly available as soon as the submitter and ID validator agree
Fast Facts on availability of sequences, when to submit them to GenBank and release them in BOLD

Sequences that are not public in GenBank are not detected by BLAST searches

Sequences that are not public in BOLD show up in BOLD ID searches as ‘private’ in BOLD meaning you see the % match, but you can’t see the details of the record (who, where)

Sequences can be made public in BOLD are detected using BOLD ID, but are not detected by GenBank BLAST. When you create a new Dataset containing records, √ the box to make Dataset Publicly Available.

Sequences produced by BOLD under our agreement for obtaining a reduced sequencing rate in return for augmenting BOLD’s fungal library must be submitted to GenBank within 1 yr of production (but researchers/students can request delay)

Sequences submitted to GenBank via BOLD are automatically kept private for 1 yr

You can request a DOI (publication digital object identifier web link) from BOLD to be sent to GenBank
Define the Project & who is in it

- Some projects consist of many contributors and others are just one or two people.
- Some projects have had their records moved from the NAMPA to a **New Project** (from the Main Console: Projects > Add a New Project).
- Most users have their data embedded in NAMPA (in Data Console > View all Records > Options > Move records to a Dataset > **Create a New Dataset**).
- Our BOLD records are linked to iNaturalist or Mushroom Observer, and the names of the submitter of those records is usually not their real name. You will need to request the real names and email addresses of all the contributors in a project from the project leader.
- All contributors to a project will be included as co-authors of GenBank sequences. We can’t change the author list for individual records in a project.
- Sequence co-authors have to include the **corresponding author** as FunDis Coordinator (coordinator@fundis.org) so communications with GenBank can be directed to the person in charge.
- If you edited the sequences in a project, you should be the **Corresponding Author** after the Project Leader unless you and the FunDiS Coordinator are one and the same/redundant.
- Co-authors Dr. Bitty Roy (roy.bitty@gmail.com) and D. Jean Lodge (dlodgester@gmail.com) need to be included at the end of the co-author list for trouble-shooting.
To Move records to a New Dataset: Mark the records belonging to a project, then un-mark sequences flagged ‘failed’ or ‘contaminant’, and verify that name updates have been made.

- For sequences flagged as ‘failed’ or ‘contaminant’, ensure the sequence is erased from the file, then un-mark the record so it is not exported.
- √ for flags saying Add Genus species to BOLD DB. If it matches the Identification, remove the flag. If not, follow instructions on requesting name additions or changes in BOLD.
- Sequences at least 150 bases long are still useful, so leave them in. BOLD generally doesn’t put short sequences in the sequence box, but someone may have recovered enough sequence to BLAST search and uploaded it into the file. BOLD may not process it.

<table>
<thead>
<tr>
<th>Select</th>
<th>Identification</th>
<th>Specimen Page</th>
<th>Sequence Page</th>
<th>Extra Info</th>
<th>BIN</th>
<th>Record Flags</th>
<th>Legend</th>
<th>Bases [Ambig]</th>
<th>Tags</th>
</tr>
</thead>
<tbody>
<tr>
<td>✅</td>
<td>Thelephora</td>
<td>iNat32779494</td>
<td>NAMPA666-21</td>
<td>1053</td>
<td>1 2</td>
<td></td>
<td></td>
<td>665[1n]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Russula foetens</td>
<td>iNat32789196</td>
<td>NAMPA667-21</td>
<td>1053</td>
<td>1 2</td>
<td></td>
<td></td>
<td>0</td>
<td>failed</td>
</tr>
<tr>
<td>✅</td>
<td>Lactarius Yazooensis</td>
<td>iNat33755572</td>
<td>NAMPA668-21</td>
<td>1053</td>
<td>1 2</td>
<td></td>
<td></td>
<td>713[0n]</td>
<td>Identified</td>
</tr>
</tbody>
</table>
Move the marked records that have good ITS sequences to a Dataset

This can be done in two steps (recommended for beginners) or use 1-step shortcut with experience because the Create a Dataset program is glitchy.

2-step process: Create a New Dataset first from the main Console, then move marked records to the New Dataset

1-step shortcut: Mark the records, choose Options > + Add Records to Dataset, and hit the green New Dataset button and fill out the form, then hit the Add Records button
Create a Data Set (1-step) is accessed on the Data Console, with View All Records opened (1st Sort Records by hitting Sequence Page header)

Within the project with your marked records, choose Options, then select the 2nd option + Add Records to Dataset
This is the Create a Data Set Pop-up Form

You can only use 4-8 letters Unique code

Check this box to make public in BOLD

The Manager is the creator

Add others to give them access

Include Maria Kuzmina, aka Masha, as an editor

Hit Save
If you make the Dataset Public you will get this message with instructions on how to find the DOI’s for each record (you can paste in a list on NAMPA codes from the spreadsheet)
Send an email to GenBank with their assigned numbers and the DOI corresponding to each as ‘Publications’

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 6 business days. The records in this dataset can be retrieved using the following permanent URL:

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.
Create a New Dataset 1st (2-step)

In the **Main Bold Console**
Datasets are listed on the left
Hit the **green New Dataset**
button to Create a New Dataset
You will see the Create a Data Set Pop-up Form

- You can only use 4-8 letters Unique code
- Check this box to make public in BOLD
- The Manager is the creator
- Add others to give them access
- Hit Save to include Maria Kuzmina, aka Masha, as an editor
Move Data to New Dataset (2-step)

Within NAMPA
with your marked records,
Choose the 2nd Option
+ Add Records to Dataset
The Popup shows the Data source and New Data Set if using the 2-step method and you already created the Dataset

If the correct dataset shows up (or if you can select it in the pull-down list), proceed with Add Records

Hit Add Records
Submitting a Dataset to GenBank from BOLD

Go to the BOLD Main Console to select your Dataset. BOLD added DS-in front of the code.

Hit the blue hot-link for the Dataset you want to submit to GenBank.
Submitting a Dataset to GenBank from BOLD (2)

Then select Publication 1st choice: Submit to GenBank
GenBank submission form in BOLD

Publication Title
Use Project Title or ask Leader
Use same title in Publication requesting a DOI

Fill in author’s names, mark Corresponding Author & fill in their contact information below

See slide 3 for who should be co-authors & order

Include Maria Kuzmina, aka Masha, as an editor

Hit green Submit Request at bottom right