

---

# **BWP Informatics Documentation**

***Release 1.0***

**Niamh Redmond, Mike Trizna**

**Jul 06, 2018**



<b>1</b>	<b>Updates</b>	<b>3</b>
1.1	Workflow	3
1.2	Software components	5
1.3	Barcode Data Standard	5
1.4	BWP Data Standard Extension	9
1.5	Naming conventions	11
1.6	Populating the FIMS spreadsheet	12
1.7	FIMS spreadsheet validation and upload	13
1.8	Geneious introduction	15
1.9	Connecting with the Geneious Biocode Plugin	19
1.10	Generating extraction plates	24
1.11	Generating PCR Plates	28
1.12	Generating cycle sequencing plates	35
1.13	Downloading traces from LIMS	37
1.14	Assembling contigs	39
1.15	Quality control of assembled contigs	45
1.16	Marking assemblies as Pass or Fail in LIMS	49
1.17	GenBank Upload	49
1.18	Annotating for ITS GenBank Upload	60
1.19	LIMS Connection Troubleshooting	62
1.20	Migrating the BWP LIMS Database	68



This documentation site describes the best practices for the informatics pipeline of the [Barcode of Wildlife Project](#).



# CHAPTER 1

---

## Updates

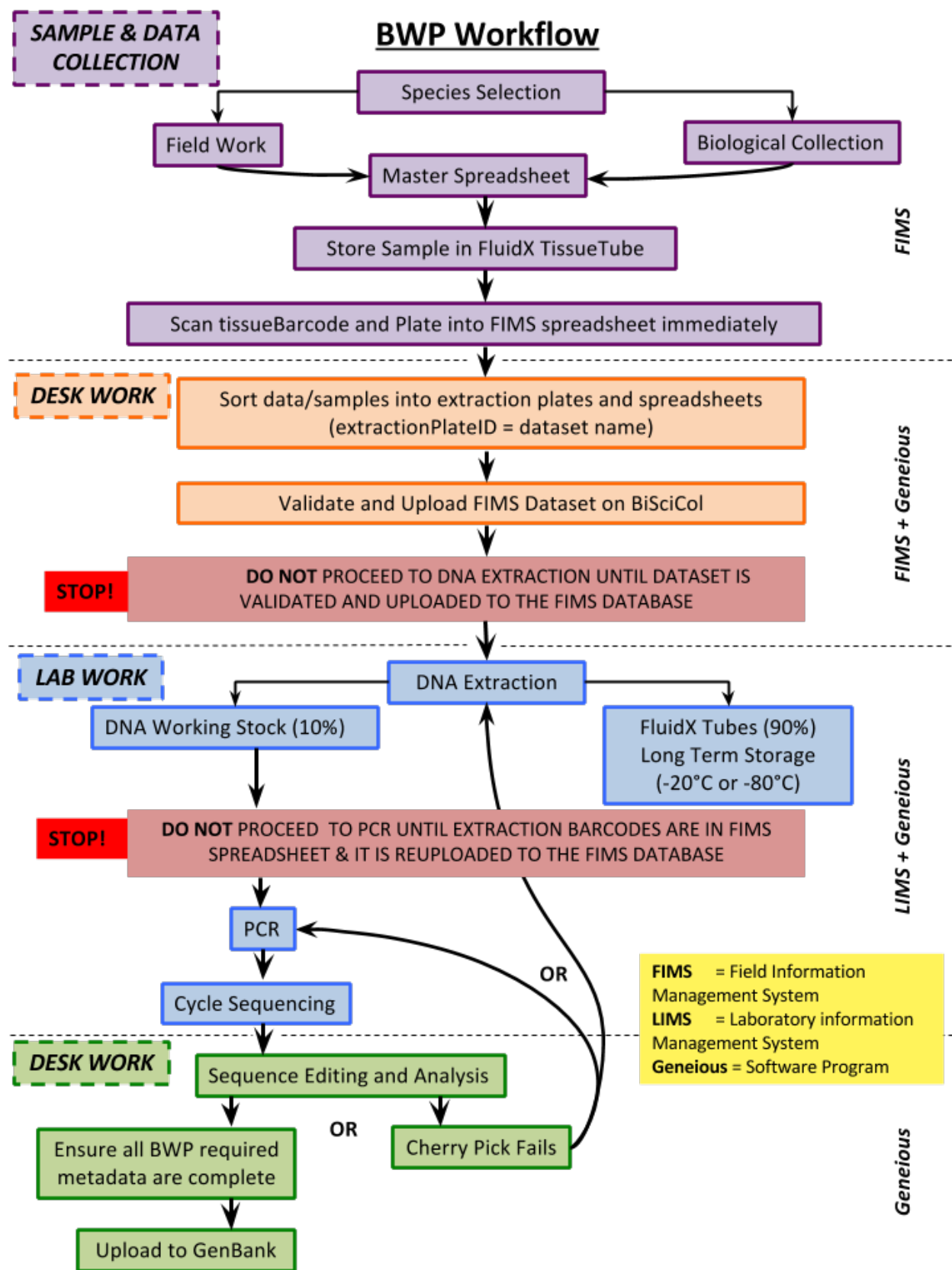
---

- 11 September, 2017** Biocode plugin updated to fix problem where the Biocode FIMS only retrieved field definitions from the Barcode of Wildlife project instead of the project that was specified. This would result in empty tissue fields.
- 20 September, 2016** The GenBank submission plugin was updated to version 1.6.5 with many minor bug fixes. The most important one to the BWP project is that the plugin will now attempt to replace non-ASCII characters (such as accented characters) with equivalent ASCII characters. This has been a problem in the past, since the previous default action was to simply omit these characters.
- 15 June, 2016** The Biocode Plugin was updated to version 3.0.0, and the GenBank Upload Plugin was updated to version 1.6.4. You can read about the updates to the Biocode Plugin [here](#). You can read about the updates to the GenBank Upload Plugin [here](#). Additionally, the back-end LIMS database was split up by country and new login information was provided to each BWP partner via email.

## 1.1 Workflow

### 1.1.1 Detailed workflow for the Barcode of Wildlife Project (BWP)

Here is the beginning (species selection) to end (publication to GenBank) workflow that BWP uses for the generation of high-quality DNA barcode records.



The workflow above is annotated on the right side with the names of the software systems that handle each step of the process. A more detailed explanation of each component is found in the next page, *Software components*.



## 1.2 Software components

### 1.2.1 Field Information Management System (FIMS) - BiSciCol Biocode

The purpose of a field information management system, or FIMS, is to store information relating to fieldwork (e.g. specimen records, collecting events, identification, etc.) The FIMS being used for the BWP is the Biocode FIMS, which is being developed by the Biological Science Collections (BiSciCol) project. The BWP is using a cloud-hosted version that is available at <http://biscicol.org>. All that you need to use the Biocode FIMS is a spreadsheet program (Excel, Google Sheets, etc.), and a browser (Firefox, Chrome, etc.).

### 1.2.2 Geneious software program

The BWP uses the Geneious molecular biology software package to run the Laboratory Information Management System (LIMS) portion of the project, and for assembly of sequences and submission to GenBank. Geneious is a cross-platform (works equally well on Windows, Mac, and Linux) program that provides a simple GUI (Graphical User Interface) to many classic molecular biology tools in a layout similar to an e-mail client. Geneious is a commercial software that is not free, but BWP is providing institutional licenses to partner institutions. You can find more information about Geneious at their website: <http://www.geneious.com/>.

### 1.2.3 Laboratory Information Management System (LIMS)

The LIMS allows you to manage your laboratory and sequence editing workflows.

A wiki containing information on the plugin and an online manual is also found at [http://software.mooreabiocode.org/index.php?title=Main\\_Page](http://software.mooreabiocode.org/index.php?title=Main_Page)

## 1.3 Barcode Data Standard

### 1.3.1 Introduction

The Barcode Data Standard was established by the Consortium of the Barcode of Life soon after the first scientific paper by Dr. Paul Hebert was published that proposed the method of DNA Barcoding. You can find a link to the official data standard document on the CBOL website at: [http://www.barcodeoflife.org/sites/default/files/DWG\\_data\\_standards-Final.pdf](http://www.barcodeoflife.org/sites/default/files/DWG_data_standards-Final.pdf).

# Barcode of Life

Identifying Species with DNA Barcoding

[About](#) [Community](#) [Resources](#) [Events](#) [Partners](#) [News](#)

[Publications](#)

[Special Issues](#)

[Standards and Guidelines](#)

[Outreach Materials](#)

[Meeting Presentations & Reports](#)

[FAQs](#)

## Standards and Guidelines

Adherence to community standards and guidelines is what allows the products of individual barcoding projects around the world to be integrated into the global DNA barcoding landscape.

### **BARCODE Data Standard**

Learn how to link barcode sequence records to voucher specimens, valid species names and taxonomic publications

### **Guidelines for non-CO1 Selection**

Sequence records in GenBank, EMBL and DDBJ can only be labeled as BARCODE records if they are from a gene region that has been approved by CBOL. As of June 2007, only the Folmer region of COI has been approved for BARCODE status. CBOL will use these guidelines to evaluate proposals to give BARCODE status to other gene regions.

### **Barcode Data Submission Tool (BarSTool)**

Submit your barcode data directly to GenBank

The data standard consists of several required and strongly recommended elements that have to do either with specimen metadata or sequence data. We will go through each element, give a brief explanation, and try to highlight any commonly seen mistakes.

<b>Specimen Metadata</b>		
<b>Text from Standard</b>	<b>GenBank Field</b>	<b>Required or Recommended?</b>
“a unique identifier for the voucher specimen using a structured field specified by CBOL and NCBI”	specimen_voucher	Required
“the name of a formally described species or a provisional label for an unpublished species”	organism	Required
“Country-Code using the controlled vocabulary used by GenBank”	country	Required
“Latitude and longitude”	lat_lon	Strongly recommended
“Name of the identifier”	identified_by	Strongly recommended
“Name of the collector”	collected_by	Strongly recommended
“Date of collection”	collection_date	Strongly recommended
<b>Sequence Metadata</b>		
<b>Text from Standard</b>	<b>GenBank Field</b>	<b>Required or Recommended?</b>
“Come from a gene region accepted by CBOL as an effective barcode” . . . “Include the name of the region used”	gene	Required
“the sequences of all forward and reverse primers used”	PCR_primers	Required
“the names of the forward and reverse primers”	PCR_primers	Strongly recommended
<b>1.3. Barcode Data Standard</b>		<b>7</b>
“trace files for the forward and	NCBI Trace Archive	Required

## 1.3.2 Specimen metadata

### Collection metadata

**Country – Required** The GenBank field name “Country” is slightly confusing – not just because the INSDC country controlled vocabulary list (<http://www.insdc.org/country.html>) includes oceans and seas in addition to countries – but because the country name is often concatenated with a colon to provide more specific location information about where a specimen was collected.

**Latitude and Longitude – Strongly Recommended** The geographical coordinates of the location of where a specimen was collected are stored in the “lat\_lon” field in decimal format. GenBank uses the specific format “d[d.ddd] N|S d[dd.ddd] W|E”. An example of this is “38.891262 N 77.026093 W” for the Smithsonian Natural History Museum.

**Collector Name – Strongly Recommended** The name of the person(s) or institute that collected the specimen. GenBank does not provide any guidance on how to structure name (“Give Name Surname” vs. “Surname, Given Name”) or how to group multiple names, but at least be consistent.

**Collection Date – Strongly Recommended** The date(s) on which the specimen was collected. Date ranges are supported by providing two collection dates from among the supported value formats, delimited by a forward-slash character.

Here are the supported value formats, with examples:

- “DD-Mmm-YYYY”: 01-Jan-2016
- “Mmm-YYYY”: Jan-2016
- “YYYY”: 2016
- “YYYY-MM-DD”: 2016-01-01
- “YYYY-MM”: 2016-01

### Voucher metadata

**Specimen Voucher – Required** The specimen voucher field is the most important portion of the Barcode Data Standard, because it serves as the bridge between genetic data and specimen data. This field is even more important for plants, because the plant barcode consists of more than one gene region. The two sequences that make up a plant barcode are published as two separate GenBank records, so a unique specimen voucher field is the only thing that asserts that these sequences came from the same individual.

Not only is a unique identifier required for the specimen voucher, but it also needs to be in a specific format. It is very easy to miss since this format is specified in a footnote, but the data standard document specifies that the voucher specimen identifier should use a triplet structure based on elements of the Darwin Core (DwC) separated by a colon:

`institutionCode:collectionCode:catalogNumber`

To ensure that specimen voucher identifiers are unique and traceable, CBOL maintains the Global Registry of Biorepositories ([GRBio.org](http://GRBio.org)), which is a community-curated directory of all biological collections in natural history museums, herbaria, and other biorepositories.

**Organism – Required** The scientific name of the organism that provided the sequenced genetic material. The text from the data standard reads “the name of a formally described species or a provisional label for an unpublished species”, which allows for the exception of allowing for organism names only identified to the Order or Family level. It is recommended by GenBank to give provisional names the values of the specimen voucher for reproducibility reasons.

**Identifier Name – Highly Recommended** The name of the person(s) or institute that identified the specimen. Just as with Collector Name, GenBank does not provide any guidance on how to structure name (“Give Name Surname” vs. “Surname, Given Name”) or how to group multiple names, but at least be consistent.

### 1.3.3 Sequence metadata

**Nucleotide Sequence – Required** This is the DNA sequence of the barcode record.

**PCR Primer Sequence(s) – Required** This refers to the sequences for the PCR primers used to amplify the DNA Barcode region. All sequences should be presented in 5’>3’ order.

Fill this in with complicated examples of cocktail primers vs. multiple pcr attempts.

**PCR Primer Name(s) – Highly Recommended** This refers to the “common names” of the primer sequences. Unfortunately this field is optional, and the vast majority of BARCODE keyword records do not have primer names listed.

**Trace Files – Required** Trace files for the forward and reverse sequencing runs must be submitted to the NCBI Trace Archive or the Ensembl Trace Server.

## 1.4 BWP Data Standard Extension

The Barcode of Wildlife Data Standard Extension was created by the BWP Specimen Informatics Working Group to hold the barcode records from BWP to the higher level required for using as evidence in court cases.

### 1.4.1 Relation to Barcode Data Standard

The first component of the BWP Data Standard states that all “strongly recommended” elements from the Barcode Data Standard are now required. These elements include:

- latitude/longitude
- name of identifier
- name of collector
- collection date
- PCR primer names

### 1.4.2 New fields in the BWP Structured Comment

In addition to the newly required fields listed above, BWP worked with GenBank to create a standardized BWP “structured comment”, which will contain these extra fields in BWP Barcode GenBank records.

COMMENT	<pre>##BWP:1.0-START## basisOfIdentification      :: materials examined, reference                            specimens, literature identificationConfidence  :: High identificationDate        :: 31-May-2013 typeStatus                :: None ##BWP:1.0-END##</pre>
FEATURES	Location/Qualifiers
source	<pre>1..659 /organism="Brachypelma hamorii" /organelle="mitochondrion" /mol_type="genomic DNA" /specimen_voucher="IBUNAM:CNNA Ar003614" /db_xref="taxon:1750707" /country="Mexico: Colima, Tecoman" /collection_date="01-Dec-2012" /collected_by="E. Goyer, E. Hijmensen, D. Ortiz" /identified_by="Jorge I. Mendoza Marroquin" /PCR_primers="fwd_name: LCO1490, fwd_seq: gggtcaacaaatcataaagatattgg, rev_name: HCO2198, rev_seq: taaacttcagggtgaccaaaaaaatca" &lt;1..&gt;659 /gene="COI"</pre>

**basisOfIdentification – Required** Free-text field describing what rationale was used to identify the specimen voucher as representing the Taxon. Can include materials examined or reference specimens, literature, or field guides.

**identificationConfidence – Required** The subjective appraisal by the identifier of the reliability of the identification at the species level. The reliability of a species identification can be High, Medium or Preliminary, based on the identifier’s expertise in the taxonomic group, the nature and quality of the specimen, and the reference material used to make the identification.

**identificationDate – Required** The date on which the specimen voucher was identified as representing the Taxon.

**typeStatus – Required** A list (concatenated by semicolons) of nomenclatural types (type status, typified scientific name, publication) applied to the subject. DarwinCore recommends a controlled vocabulary of none, holotype, paratype, neotype, syntype, lectotype, paralectotype, hapantotype, other.

**chainOfCustody – Optional** Whether or not a specimen voucher has been collected, transported, and accessioned into a secure repository following chain of custody rules that prevent tampering, contamination, mislabeling or loss. Chain of custody refers to the chronological documentation of the sequence of physical locations of a voucher specimen and the people who had custody of it.

**coordinatePrecision – Optional** A decimal representation of the precision of the coordinates given in the decimalLatitude and decimalLongitude fields. This field is REQUIRED when you are giving “fuzzed” GPS coordinates. Make sure the actual real coordinates are stored in your specimen database.

**Example:** “1.0 (nearest degree)” for when real decimalLatitude and decimalLongitude values of 38.891351 and -77.026072 are “fuzzed” to 39 and -77 to protect location information.

## 1.5 Naming conventions

### 1.5.1 FIMS naming conventions

Once specimens are available to be processed in the lab, you must divide the specimens into a batch, or batches, of specimens that will be processed together through the full laboratory pipeline (extraction through to sequencing). Each batch is also known as a “plate” or “dataset”, terms that will be used repeatedly throughout our instructions for the FIMS and LIMS. Specimen spreadsheets must be completed for each plate of specimens prior to their processing in the lab.

**Create your plate/dataset name from [Institution Code][User Initials][Sequential Number]** Example:

CBOL\_MGT\_001 for the first plate/batch from Michael G Trizna of CBOL

This name will then be used for the:

- Spreadsheet filename (CBOL\_MGT\_001.xlsx)
- Dataset Code in FIMS

- extractionPlateID field in spreadsheet

	C	
	extractionPlateID	e
	CBOL_MGT_001	

- and will also be incorporated in the plate names in the LIMS system (see below)

### 1.5.2 LIMS naming conventions

**Extraction Plate:** *[dataset code]\_X[extraction attempt number]*

Example: CBOL\_MGT\_001\_X01 for the first extraction plate from the plate/batch of samples in the dataset named “CBOL\_MGT\_001”. If you were to re-extract DNA from all of the samples in CBOL\_MGT\_001, then this second extraction plate would be called CBOL\_MGT\_001\_X02.

**PCR Plate:** *[dataset code]\_PCR[PCR attempt number]\_[fwd primer]\_[rev primer]*

Example: CBOL\_MGT\_001\_PCR01\_dgLCO1490\_dgHCO2198 for a PCR reaction of the CBOL\_MGT\_001 plate using primers dgLCO1490 and dgHCO2198. Subsequent PCRs should be named in chronological order regardless of the primer set used. For example if a second PCR was completed for this plate it would be named CBOL\_MGT\_001\_PCR02\_Fprimername\_Rprimername.

**Cycle Sequencing Plate:** *[dataset code]\_Seq[Sequencing attempt number]\_[primer used]\_[F or R]*

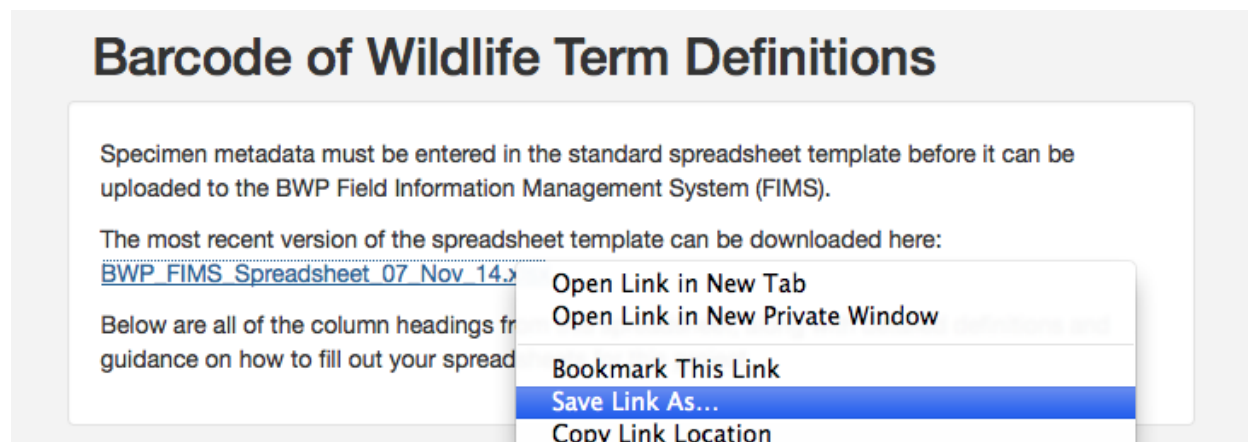
Examples: CBOL\_MGT\_001\_Seq01\_dgLCO1490\_F for the cycle sequencing plate for the forward strand using the dgLCO1490 primer; and CBOL\_MGT\_001\_Seq02\_dgHCO2198\_R for the reverse strand using the dgHCO2198 primer.

**IMPORTANT** There are different scenarios for dealing with re-runs (extractions, PCRs and sequencing) of subsets of samples from an individual dataset. Please contact us to discuss these scenarios to ensure that you continue to use a manageable workflow in the LIMS.

## 1.6 Populating the FIMS spreadsheet

### 1.6.1 Downloading and renaming the FIMS spreadsheet

Go to the Barcode of Wildlife Term Definitions web page at <http://barcodeofwildlife.org/fims/terms.html> and download a FIMS spreadsheet template.



It is a good idea to immediately rename this spreadsheet file with the name of your dataset according to the *FIMS naming conventions*. For this example, I'm going to name my spreadsheet "CBOL\_MGT\_001.xlsx". This is because my institution is CBOL, my initials are MGT, and this is the first dataset I am uploading to the FIMS.

#### What is a dataset?

A dataset in the Biocode FIMS is a collection of specimen records that will go through the entire laboratory pipeline together (DNA extraction through to sequencing). They are often grouped together taxonomically, but this is not the only factor considered for creating the dataset. It is important to choose an appropriate size for your dataset. For example, at the Smithsonian we run extractions, PCR, and sequencing reactions in 96-well plates. Because of the 94 specimen batch size (96 - 2 controls), all FIMS datasets contain 94 specimen records.

#### What do these columns mean? Where did they come from?

The FIMS spreadsheet contains a number of columns that will gather information on each specimen such as taxonomy, collection details, life history.

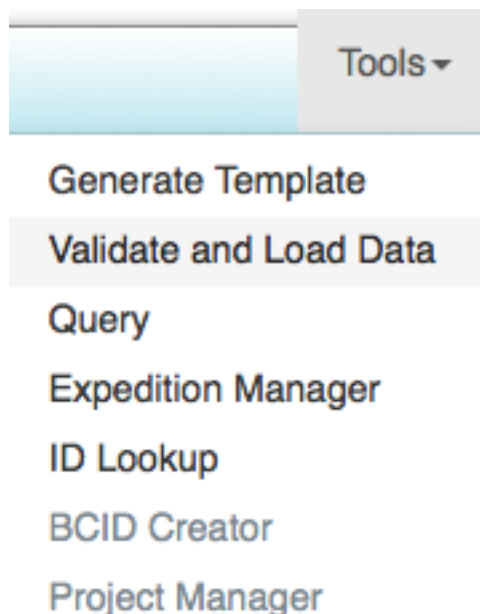
The Barcode of Wildlife Term Definitions web page at <http://barcodeofwildlife.org/fims/terms.html>, contains in depth descriptions and information on all of the columns, including what columns are required for a successful upload to the FIMS and if there are formatting restrictions on the column entry.



## 1.7 FIMS spreadsheet validation and upload

### 1.7.1 Validating and then uploading the spreadsheet

Once we have a spreadsheet fully populated, our next step will be to validate and upload the spreadsheet to the FIMS. The validation step checks every specimen record to make sure that they meet the set of rules established by the project. Go to <http://biscicol.org> and click the Login button at the top. Log in with your pre-assigned BiSciCol username and password. (You do not need to log in to validate your spreadsheet, but you must be logged in to upload it to the database). After logging in, click on the Tools button at the top, and then choose Validation.



On the Validate and Load Data page, for 'FIMS Data' click the Choose File (or Browse, depending on what browser you are using) button to select your spreadsheet. Select your Project from the Project drop-down. If you are ready to upload your spreadsheet, select the Upload tick box. Once you check that box, the Dataset Code dropdown will appear. You can either leave it as Create New Expedition to upload the spreadsheet as a new dataset, or you can select one of your previously created datasets to overwrite. Ensure that the Public checkbox is ticked, otherwise the LIMS software will not be able to pull in the information. Press Submit.

## Validate and Load Data

FIMS Data	<input type="button" value="Browse..."/>	CBOL_MGT_001.xlsx
FASTA Data (optional)	<input type="button" value="Browse..."/>	No file selected.
Choose Project	<input type="text" value="Barcode of Wildlife Training"/>	<input type="checkbox"/> (include public projects)
Upload	<input checked="" type="checkbox"/>	
Dataset Code	<input type="text" value="Create New Expedition"/>	
Public	<input checked="" type="checkbox"/>	
<input type="button" value="Submit"/>		

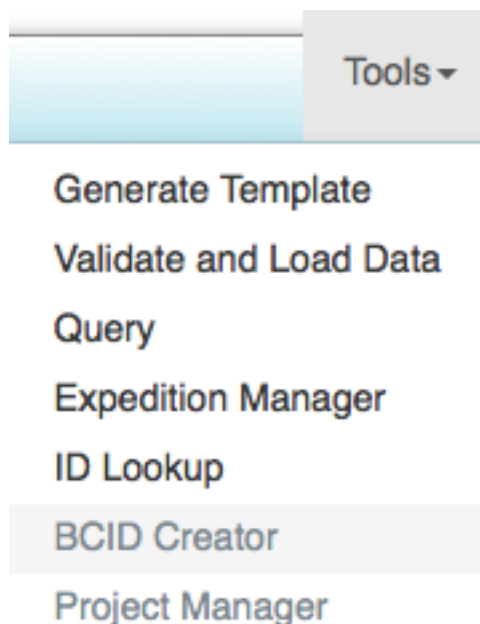
If you chose to Create New Expedition, you will be asked to enter an code. Remember to follow the BWP *FIMS naming conventions*, and make sure not to repeat a previously created expedition code. Click Create.

### Can I add a few records to an existing dataset by just uploading those new records?

NO. When you choose to upload to a previously created dataset, the old dataset will be completely deleted and replaced by the data in the new spreadsheet. Therefore you should add the new records to your existing dataset, save it, and validate and upload this updated spreadsheet.

### 1.7.2 Check that the spreadsheet was successfully uploaded

Even though we got a success message, let's make sure your data made it into the database successfully. This will also show how to retrieve data from the FIMS if you ever want to make changes. Click on the Tools button up top, and choose Query.



On the Query page, you will be able to view the data from any dataset on any of the projects in the Biocode FIMS. This is why it is so important not to include sensitive information (such as highly accurate GPS coordinates) in spreadsheet uploads. To see the dataset we just uploaded, choose Barcode of Wildlife Training from the Choose Projects dropdown. Once the project is chosen, the list of datasets will be populated. Scroll through the list of datasets until you find the one you just created. If you do not see it in this list, then your data was not successfully uploaded.

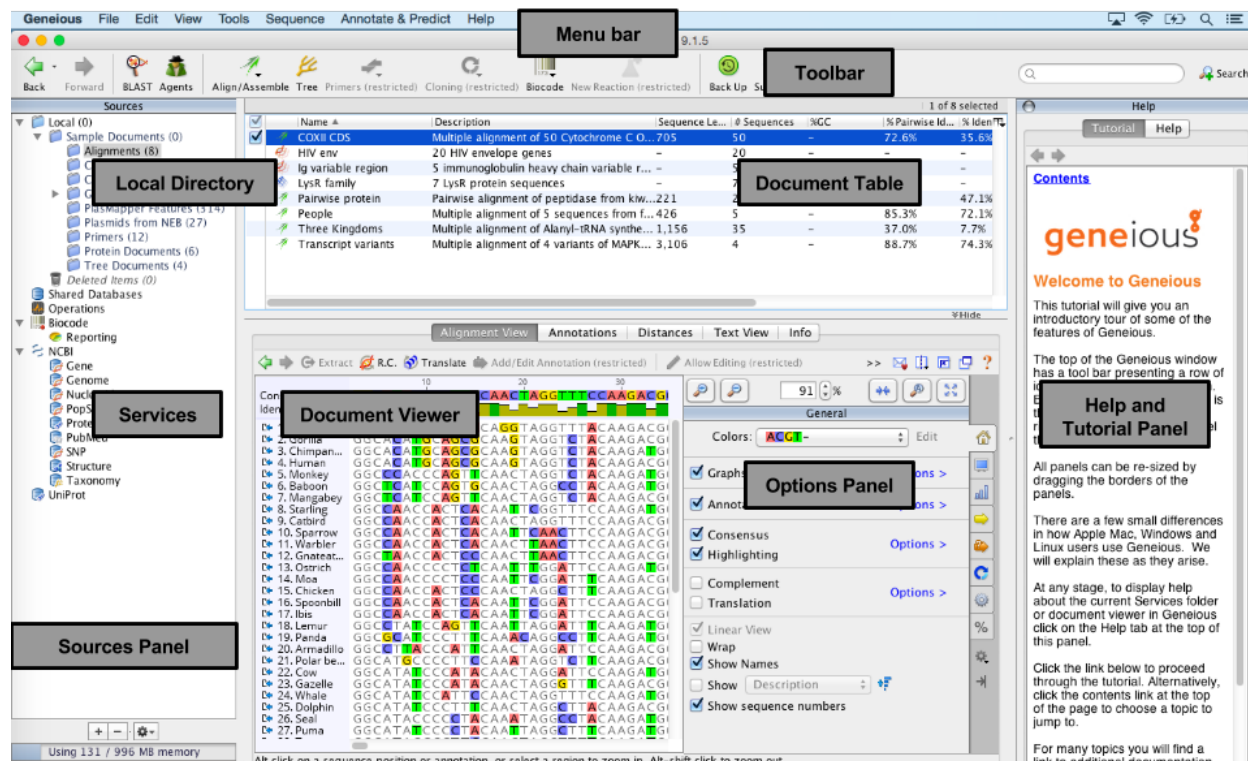
Select the dataset you just uploaded and click the Table button. This will display an HTML table view of the specimen records in this dataset. It will be incredibly hard to read and scroll through, but be happy it made it in there.

You can click the Excel button to download a fresh Excel spreadsheet of this dataset. However, be advised that the wells are out of order and that you will lose any styling (colors, bold/italics, etc.) from when you originally uploaded the spreadsheet. Also, the default spreadsheet name is “biocode-fims-output.xls”, so be sure to rename it with your dataset name immediately.

## 1.8 Geneious introduction

### 1.8.1 Geneious layout

Here is the layout of the Geneious window.



The **Toolbar** is found at the top of the Geneious window, and contains a variety of buttons including Back, Forward, Tree, Support and so on.

The Geneious window is further divided into a number of panels including the **Document Table**, **Document Viewer**, **Options Panel**. The size of each panel can be adjusted by clicking on the border and dragging to the desired size.

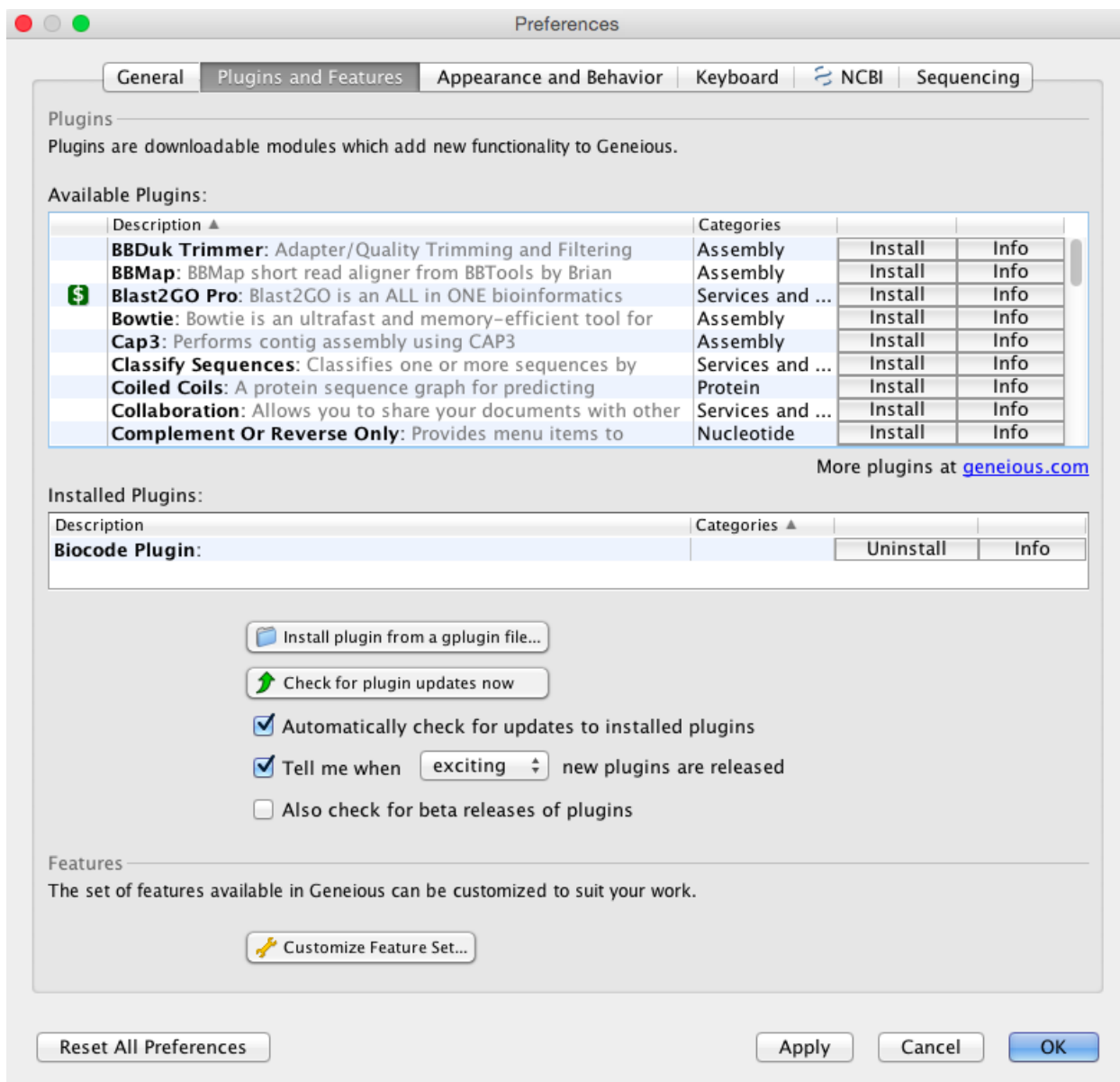
On the right-hand side is the **Help and Tutorial Panel**, where you can read a number of tutorials and also search for help in other topics.

On the left-hand side is the **Sources Panel** which includes the **Local Directory** and list of **Services**. The **Local Directory** is where you permanently store your documents such as primers, raw traces, contigs, alignments, trees etc.

## 1.8.2 Geneious plugins

Plugins are downloadable modules that can add new functionality to the program.

Click on “Tools” in the Menu bar then select “Plugins” from the dropdown menu. A new window will open and the “Plugins and Features” tab contains information on the available and installed plugins.



One way to install a plugin is to click on the “Install” button for the relevant plugin in the available list, or alternatively, you can download the plugin from an outside source and then click on “Install plugin from a gplugin file”. Within this window you can also uninstall installed plugins, get plugin information such as versions, check for plugin updates etc.

### 1.8.3 Creating primers

To add a primer to your Local directory, you should first highlight the destination folder for the primer file. If you need to create a new folder in the directory, highlight the Local folder in the Sources panel, Click on “File” in the Menubar, and enter the new folder name.

To create a new primer, highlight the destination folder, then in the Toolbar select “Sequence” followed by “New Sequence” from the dropdown menu. A “New Sequence” window will open. Here, you must enter the primer sequence, name and in the “Type” dropdown menu select “Primer”.

New Sequence

Type or paste in sequence:

1-27 INSERT PRIMER SEQUENCE HERE

Name: INSERT PRIMER NAME HERE

Description: Optional

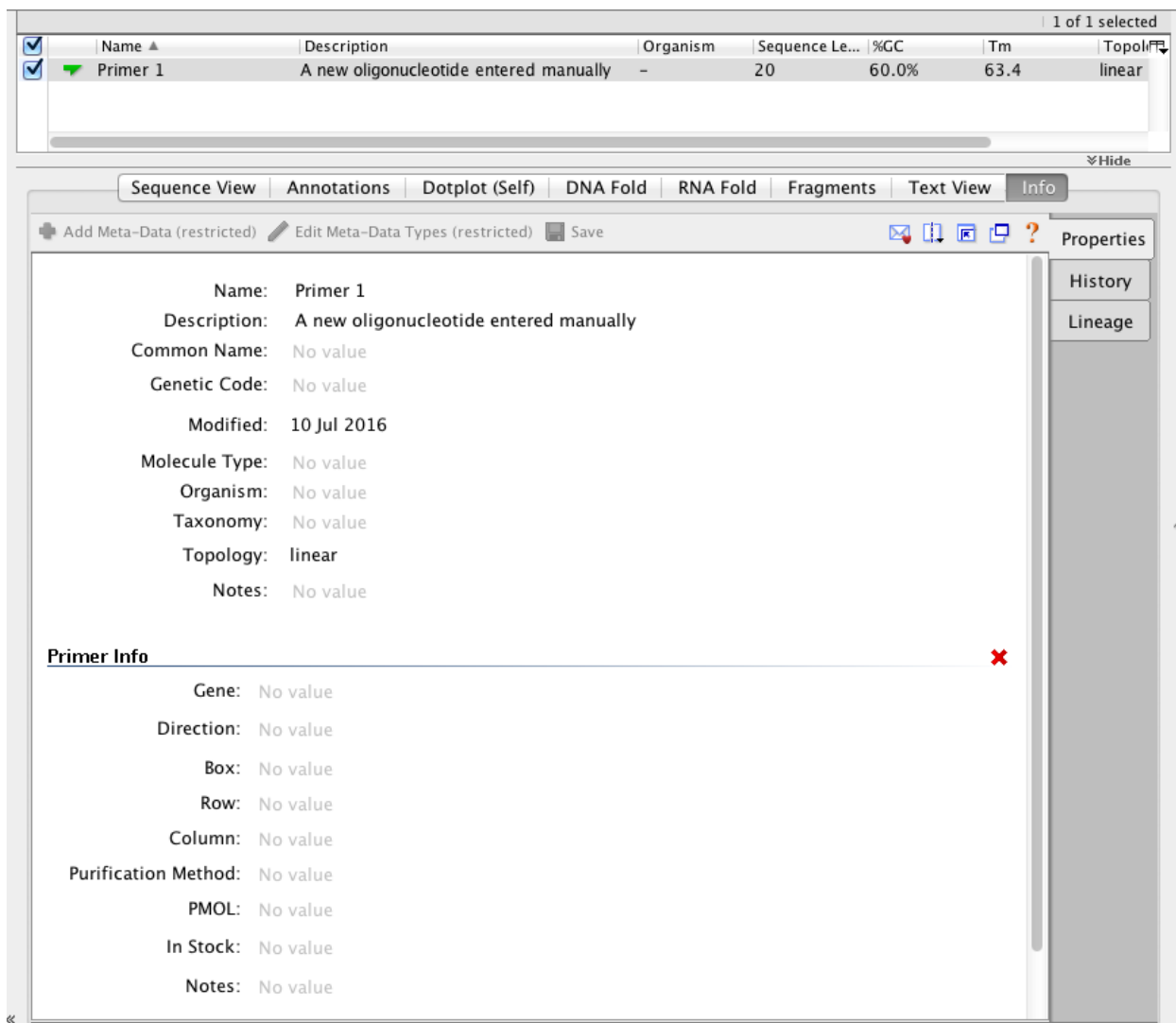
Organism: Optional

Type:

- Nucleotide: New nucleotide sequence
- Protein: New protein sequence
- Primer: A new oligonucleotide entered manually
- Probe: A new oligonucleotide entered manually

Cancel OK

Once a primer has been created, you can edit the associated information. Highlight the primer in the Document Table, and go to the “Info” tab in the Document Viewer. Here you can add various meta-data such as Gene, Organism, Direction etc. We recommend that you fill in as much information as possible for future reference.



A very convenient tool is the ability to export and import folders of primers. To export a folder from the Local, simply select the folder, go to “File” in the Menu bar, followed by “Export” in the dropdown menu. There are multiple options for the export, however in this case, select “Export folder” and enter a name and destination for the export.

There are two options for importing a folder or file into Geneious. One way is to go to “File” in the Menu bar, followed by “Import” in the dropdown menu, and then select what you are importing (folder, file etc.). You will need to select the type format followed by the location of what you want to import.

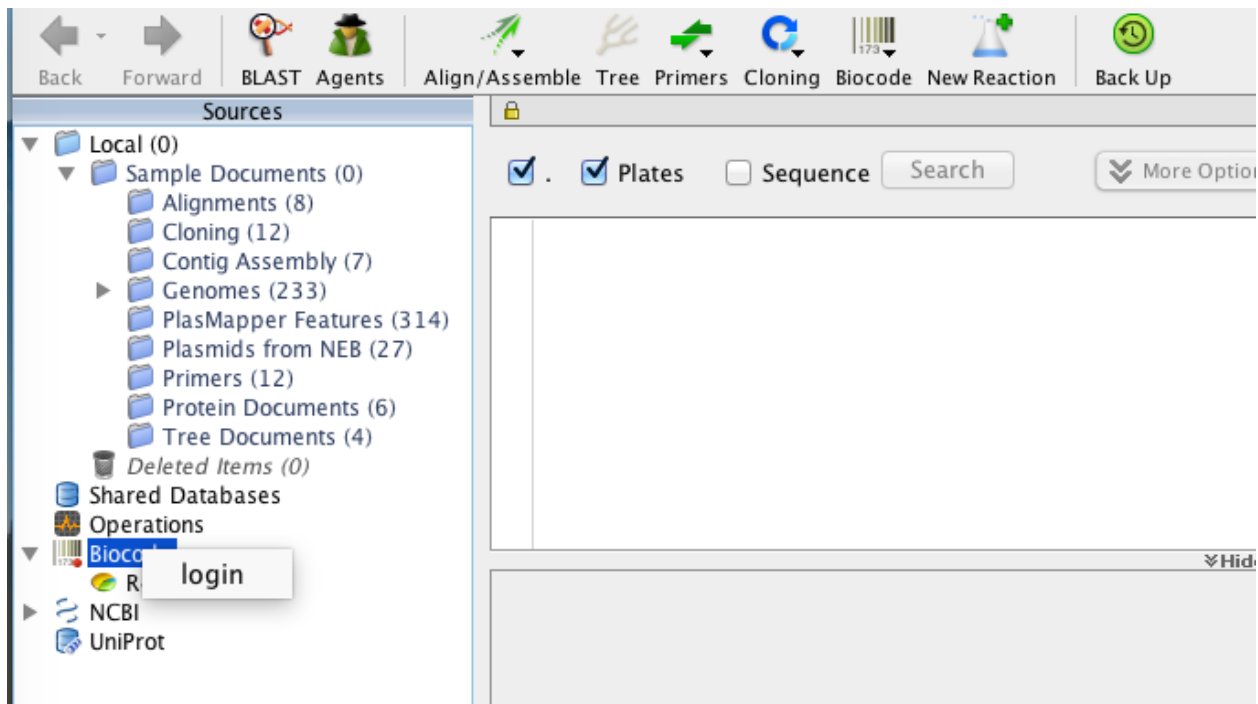
Alternatively, you can move folders, files etc. into the destination folder by dragging and dropping.

## 1.9 Connecting with the Geneious Biocode Plugin

### 1.9.1 BWP Biocode Plugin connection details

Install the latest version of the Biocode Plug-in. Check the [Updates](#) section of the Home Page to see the latest version available. If you do not have the latest version, then follow the instructions in the [Geneious plugins](#) section of the previous page.

Once the Biocode plugin has been successfully installed, you can now connect by right-clicking on the Biocode icon in the Sources panel and click “Login”.



This will open the Biocode Connections window that looks like this:

### Connection Name and Method

Enter a value in the Connection Name so that the plug-in saves this information. Connection Method should be “Direct FIMS/LIMS”.

### Field Database Connection

In the Field Database Connection section, start out by selecting “Biocode FIMS” from the dropdown.

This will auto-fill the Host box with “<http://biscicol.org>”. **You will need to manually change this entry to “<http://www.biscicol.org>”. This new address will be fixed in future versions of the Biocode plug-in, but this manual correction is necessary until then.** Enter the FIMS Username and Password that Mike Trizna assigned to you during training. Make sure to click the “Save” box, and then the “Authenticate” button to ensure you can connect to FIMS.

After authenticating, the Project drop-down will now be populated with only the FIMS projects that you have access to. Choose the project that corresponds with your country.

**Tissue ID field** For “Tissue ID field”, select “tissueBarcode”.

**Specimen ID field** For “Specimen ID field”, select “voucherID”.

Make sure to check the box next to “The FIMS database contains plate information”.

**Plate name field** For “Plate name field”, select “extractionPlateID”.

**Well field** For “Well field”, select “extractionWell”.



**Biocode Connections**

Connections: **Untitled**

Connection Name:

Field Database Connection: **Excel**

Choose the location of your excel file.  
The first row should be column headers, and it should have at least a tissue and specimen column.

Excel file location:  **Browse**

Tissue ID field: **No Columns**

Specimen ID field: **No Columns** ☐ Specimen photos on Flickr **?**

☐ The FIMS database contains plate information

Plate name field: **No Columns**

Well field: **No Columns**

Specify your taxonomy fields, in order of highest to lowest **Autodetect**

**No Columns** **+** **-**

FIMS Timeout (seconds):

LIMS location: **Built-in MySQL Database**

Choose database: **No databases** **Add Database** **Remove Database**

☐ Connect on startup

LIMS Timeout (seconds):

**Cancel** **OK**

Field Database Connection: **Excel: Read field information from an excel worksheet**

**Google Fusion Tables:** Use a remote Fusion Table as your FIMS

**Remote MySQL Database:** Use a MySQL database as your FIMS

**Moorea FIMS:** A connection to the Moorea FIMS database

**TAPIR:** Connect to a TAPIR provider

**Biocode FIMS:** Connection to the new Biocode FIMS (<https://github.com/biocodellc/biocode-fims>) **Browse**

Finally, click the “Autodetect” button to automatically populate the different taxonomy fields.

The completed Field Database Connection should look like this:

Field Database Connection **Biocode FIMS**

Host:

Username:

Password:  ☐ Save

---

Project: **Barcode of Wildlife Training**

Tissue ID field:

Specimen ID field:  ☐ Specimen photos on Flickr

☒ The FIMS database contains plate information

Plate name field:

Well field:

Specify your taxonomy fields, in order of highest to lowest

<input type="text" value="kingdom"/>		
<input type="text" value="phylum"/>		
<input type="text" value="class"/>		
<input type="text" value="order"/>		
<input type="text" value="family"/>		
<input type="text" value="genus"/>		
<input type="text" value="scientificName"/>		

## LIMS Database Connection

In the LIMS Location section, start out by selecting “Remote MySQL Database” from the dropdown.

**Server Address** For “Server Address”, copy and paste in “db.bwplims.info”.

**Port** For “Port”, make sure it is set to “3,306”.

**Database Name** For “Database Name”, enter either “kenya\_lims”, “mexico\_lims”, or “south\_africa\_lims” depending on which partner country you are affiliated with.

**Username and Password** Finally, for the “Username” and “Password” boxes, enter the LIMS database Username and Password that Mike Trizna assigned to you via email.

Make sure that the checkbox next to “Save” is selected.

Your fully-completed Biocode Connection window should look like this:

Biocode Connections

Connection Name:

Field Database Connection:

Host:

Username:

Password:  ☒ Save

---

Project:

Tissue ID field:

Specimen ID field:  ☐ Specimen photos on Flickr  ☒ The FIMS database contains plate information

Plate name field:

Well field:

Specify your taxonomy fields, in order of highest to lowest

<input type="button" value="kingdom"/>	<input type="button" value="kingdom"/>	<input type="button" value="kingdom"/>	<input type="button" value="kingdom"/>
<input type="button" value="phylum"/>	<input type="button" value="phylum"/>	<input type="button" value="phylum"/>	<input type="button" value="phylum"/>
<input type="button" value="class"/>	<input type="button" value="class"/>	<input type="button" value="class"/>	<input type="button" value="class"/>
<input type="button" value="order"/>	<input type="button" value="order"/>	<input type="button" value="order"/>	<input type="button" value="order"/>
<input type="button" value="family"/>	<input type="button" value="family"/>	<input type="button" value="family"/>	<input type="button" value="family"/>
<input type="button" value="genus"/>	<input type="button" value="genus"/>	<input type="button" value="genus"/>	<input type="button" value="genus"/>
<input type="button" value="scientificName"/>	<input type="button" value="scientificName"/>	<input type="button" value="scientificName"/>	<input type="button" value="scientificName"/>

FIMS Timeout (seconds):

LIMS location:

Server Address:

Port:

Database Name:

Username:

Password:  ☒ Save

LIMS Timeout (seconds):

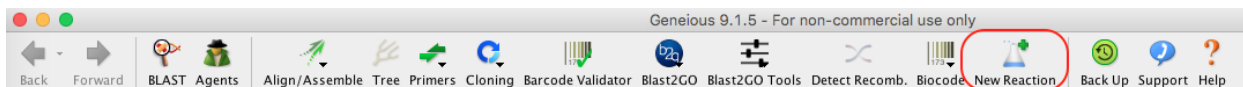
## 1.10 Generating extraction plates

### 1.10.1 Creating the plate (96 wells)

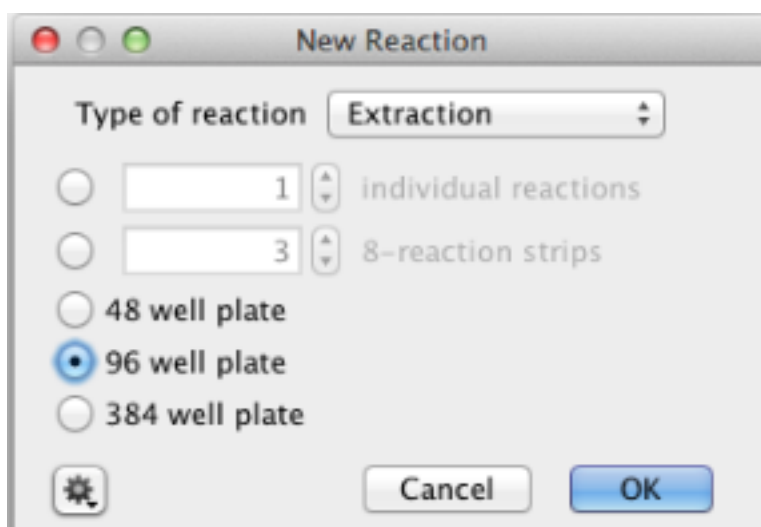
The first step of the LIMS workflow is the generation of a DNA extraction plate. In LIMS terms, an “extraction plate” is a set of DNA extractions following identical protocols and can include anywhere from a single extraction up to a plate of 384 extractions. This set of extraction will be moved through the entire pipeline together.

To create an extraction plate:

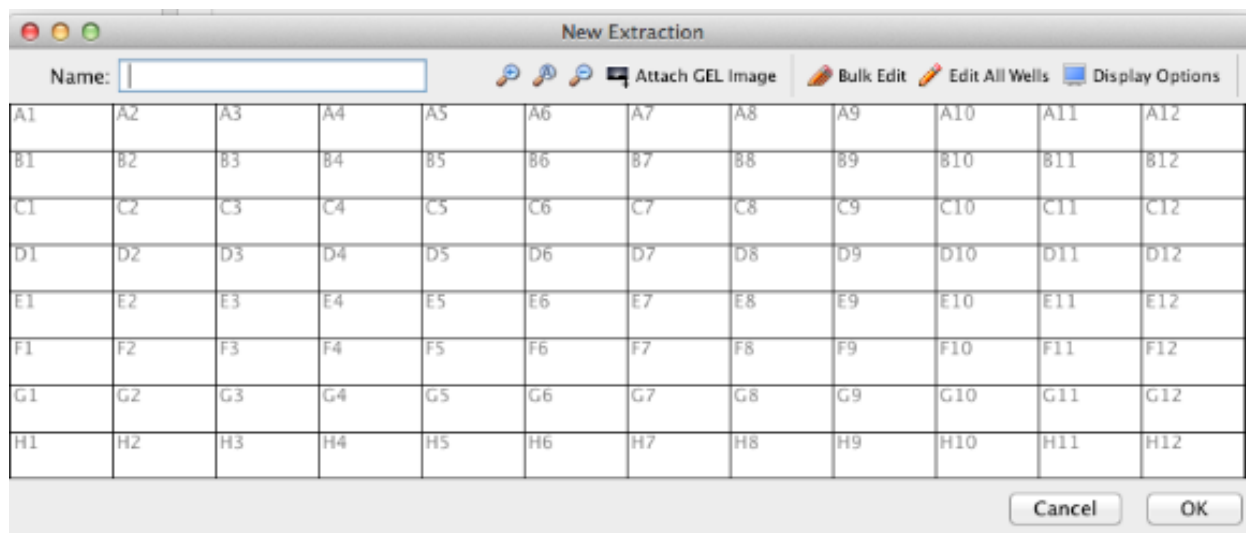
- Click the “New Reaction” button on the Geneious Toolbar.



- In the drop-down menu of the “New Reaction” window, choose “Extraction” from the “Type of reaction” drop-down menu. Select the “96 well plate” size option.



- A new window will open displaying a map of a 96 well extraction plate. The FIMS data must be uploaded to an empty extraction plate so that the each specimen’s field data is correctly associated with that same specimen’s laboratory workflow.
- Insert a name for the extraction plate into the relevant field in the window.
- Click the “Bulk Edit” button (found in the toolbar, as seen in image above). This will allow you to attach the FIMS meta-data from all tissues to each tissue’s associated workflow in LIMS.
- A new “Edit Plate” window will appear.
- Click on “Tools” and select “Get Tissue ID’s From Archive Plate” from the dropdown menu.
- Another new window “Get FIMS plate” will appear.
- Here you must enter the extractionPlateID value that was included in the FIMS spreadsheet for this plate/dataset. Click “OK” once complete.
- A new window will appear indicating the LIMS is fetching the tissue ID’s from the FIMS database. When successful, the “Tissue Sample Id” fields in the “Edit Plate” window will be populated. Alternatively, if you receive a message indicating that the plate cannot be found in the FIMS, check:

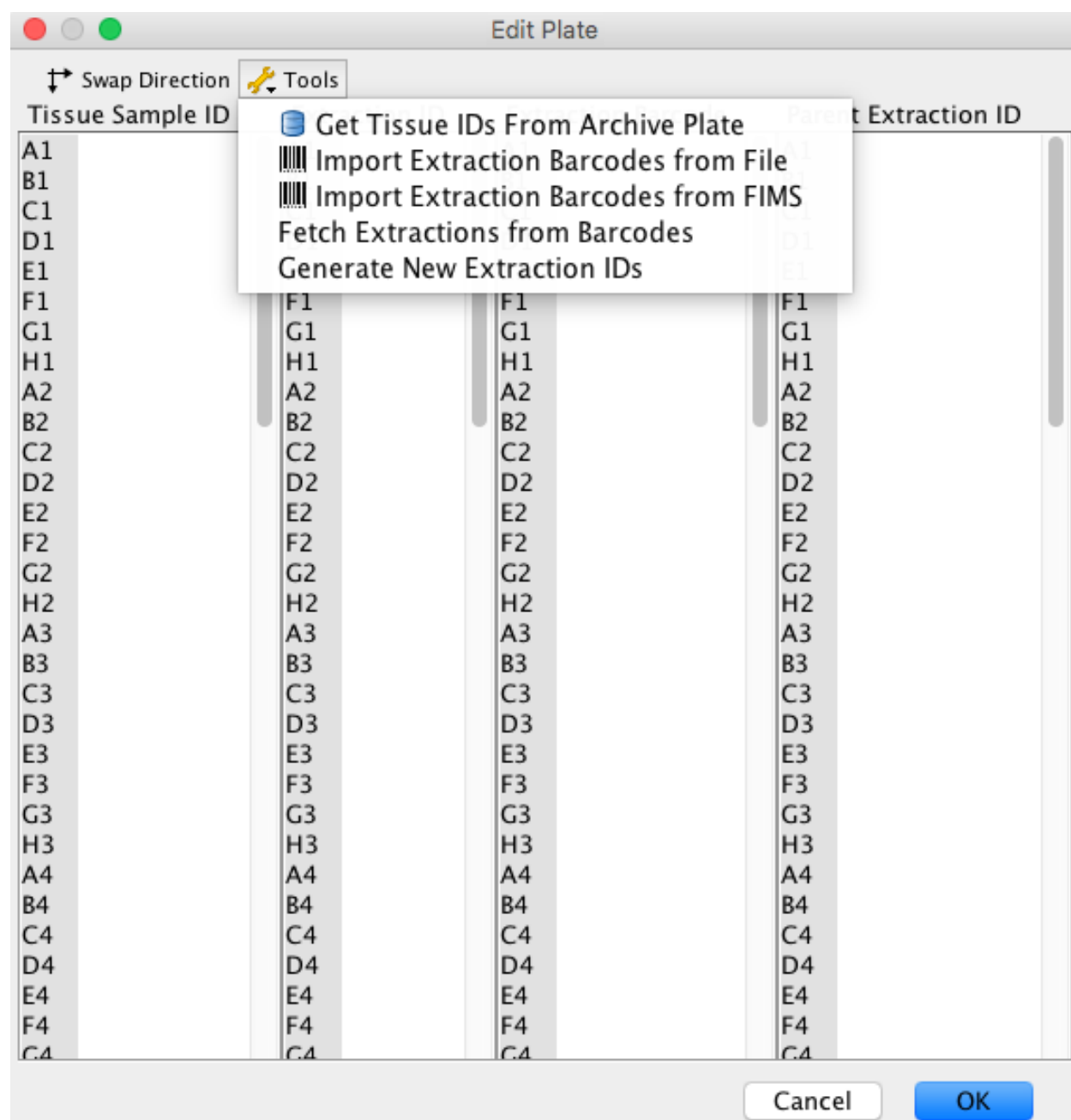


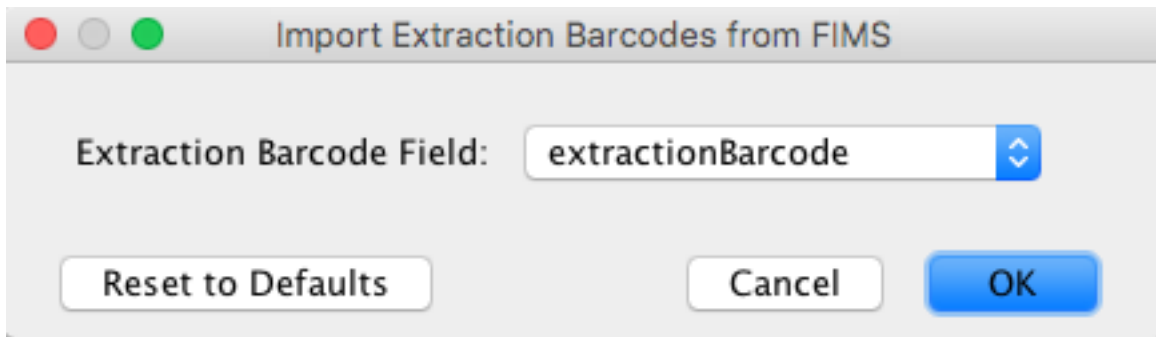
- that the plate ID is spelled exactly as it was included in the FIMS spreadsheet.
  - that the FIMS spreadsheet was successfully uploaded to the FIMS; and
  - that when you logged in to the LIMS that you selected the correct FIMS field for the “Plate name field” of the log-in window.
- After successful upload, select “Generate New Extraction IDs” from the “Tools” dropdown menu. The LIMS will automatically populate the second column of the window (Extraction Id). The .1 indicates this is the first extraction in the LIMS for that particular tissue, .2 would indicate it is the second and so on.
  - Finally you want to import the Extraction Barcode’s for each DNA extract. This can be done in two different ways:
    - *Option 1:* This is our preferred method. Select “Import Extraction Barcodes from FIMS” from the “Tools” dropdown menu. A new window will appear, and select the FIMS column that contains the extraction barcodes from the dropdown menu - in our case it is “extractionBarcode”. For this to work, you must ensure that your FIMS spreadsheet contains the unique extraction barcodes.
    - *Option 2:* Select “Import Extraction Barcodes from File” from the “Tools” dropdown menu. Select the file generated from the plate scanner or alternatively you can copy and paste the data from the file generated from the plate scanner into that column in the window. Since copy and pasting is open to user error, we strongly advise against using it unless absolutely necessary.
  - Once this is complete Click “OK” and you will be brought back to the “New Extraction” window.
  - Here, click on “Edit All Wells” to open a new window.
  - This window allows you to enter information about the extraction - date, method, technician etc.
  - Once complete, click “OK” which will bring you back to the “New Extraction” window.
  - Click “OK” in this window for the LIMS to save the extraction plate.

## 1.10.2 Additional Information:

### More information on the “Edit Plate” Window:

In this window, the well locations are displayed on the left hand side of each column to make placement easier. The “Swap Direction” button allows the user to choose between reading the plate horizontally or vertically. Clicking “OK”

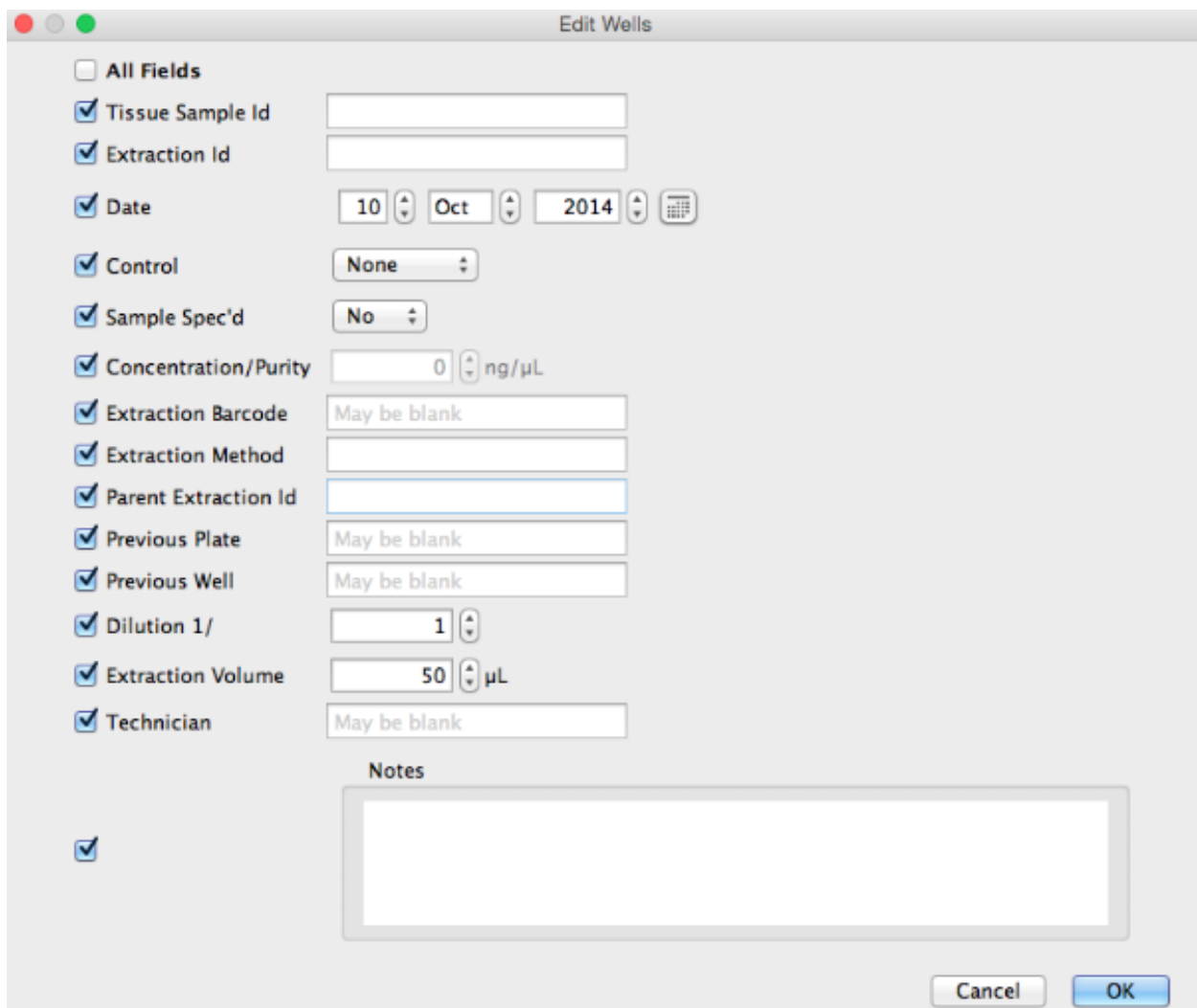




A dialog box titled "Import Extraction Barcodes from FIMS". It features a label "Extraction Barcode Field:" followed by a text input field containing "extractionBarcode" and a dropdown arrow. At the bottom, there are three buttons: "Reset to Defaults", "Cancel", and "OK".

Import Extraction Barcodes from FIMS

Extraction Barcode Field:



An "Edit Wells" dialog box with a list of fields on the left, each with a checkbox. The fields and their values are: "All Fields" (unchecked), "Tissue Sample Id" (empty), "Extraction Id" (empty), "Date" (10 Oct 2014), "Control" (None), "Sample Spec'd" (No), "Concentration/Purity" (0 ng/μL), "Extraction Barcode" (May be blank), "Extraction Method" (empty), "Parent Extraction Id" (empty), "Previous Plate" (May be blank), "Previous Well" (May be blank), "Dilution 1/" (1), "Extraction Volume" (50 μL), and "Technician" (May be blank). At the bottom left is an unchecked checkbox. To the right of the checkboxes is a "Notes" section with a large text area. At the bottom right are "Cancel" and "OK" buttons.

Edit Wells

☐ All Fields

☒ Tissue Sample Id

☒ Extraction Id

☒ Date

☒ Control

☒ Sample Spec'd

☒ Concentration/Purity

☒ Extraction Barcode

☒ Extraction Method

☒ Parent Extraction Id

☒ Previous Plate

☒ Previous Well

☒ Dilution 1/

☒ Extraction Volume

☒ Technician

☒

Notes

will attach the FIMS meta-data to the new extraction plate.

Under the “Tools” menu within the “Bulk Edit” window there are a number of options, described below in order of the dropdown menu, available.

- The “Get Tissue ID’s from archive plate” option allows the extraction plate to be filled with extraction IDs from the FIMS data source.
- “Import Extraction Barcodes from File” allows the extraction barcode values to be directly imported from the output file of the scanner if 2D well barcodes are being used.
- “Import Extraction Barcodes from FIMS” allows the extraction barcode values to be directly imported from the FIMS.
- “Fetch extractions from barcodes” is used during “cherry picking” to populate newly reconstituted plates from prior plate locations.
- “Generate Extraction IDs” automatically generates appropriate extraction IDs based on the tissue sample IDs.

### **More information on editing wells in the plate:**

The “Edit All Wells” or “Edit Selected Wells” button, found in the toolbar of the “New Extraction” plate window, opens a customizable viewer and editor for plate documents. It is shown both when creating new plates, and when viewing existing plates in the database. Wells can be selected in the plate by dragging the mouse across the plate to select a number of wells, or holding down the shift and ctrl (command on mac) keys to help select multiple individual wells. When a well, or wells, have been selected, click “Edit Selected Wells” to customize those wells.

The “Edit Wells” dialog window will open (see image above), and it has a column of checkboxes on its left hand side. Values in the checked fields will be applied to all selected reactions, and unchecked fields will be left as they are. Most values can simply be entered into a dialog box. Make sure to save the plate after making any edits to it.

### **More information on the Display Options in the plate:**

Clicking the “Display Options” button, found in the toolbar of the New Extraction plate window, opens the “Display” dialog window (below). The split-pane allows the user to choose any number of fields from the FIMS or LIMS database for display in the wells.

The available fields are shown in the left hand pane, and when fields are in the right hand pane they are displayed in the wells. To move a field between the two panes, select it and click the right or left arrow depending on the direction of the move. Once the fields to display have been decided upon, their display order can be altered using the up/down arrows on the right hand side of the dialog box. The field in the top position of the list will be displayed more prominently in the well, as it will be in larger, bold text.

Each well can be color-coded according to a particular field value. To select the field for color-coding use the “Color wells based on” drop-down menu found at bottom of the dialog window. All possible values for that field will be displayed and a color can be assigned to each of the values using the color chart.

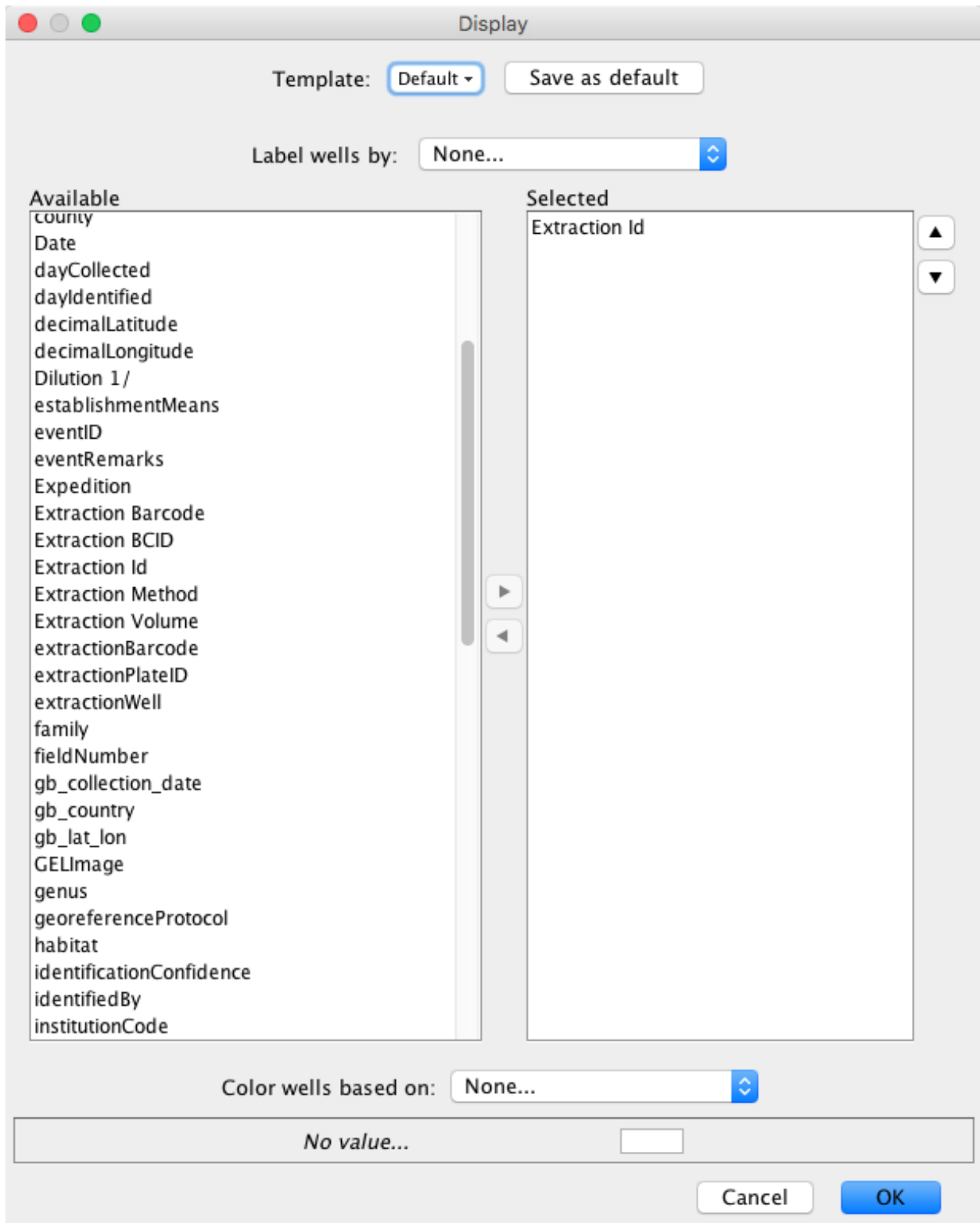
A display setting can be saved as a template by clicking the “Select a template” button at the top of the dialog window and clicking “Create template”. Click the “Save as Default” button to make that template the default. Separate defaults are stored for extraction, PCR, and cycle sequencing plates.

## **1.11 Generating PCR Plates**

### **1.11.1 Searching for existing plates**

Biocode LIMS searches can return four types of documents:





- Tissue sample documents – Each of these represents a tissue sample in the field database. Tissue documents contain collection and specimen information.
- Plate documents – These represent a plate in the LIMS, and can be extraction, PCR, or cycle sequencing plate types.
- Workflow documents – These contain a set of linked reactions performed on an extraction.
- Sequence documents - Sequences entered into the LIMS when traces/assemblies are marked as passed/failed.

Searches are performed by entering text into the search box (found above the Display Table) and will return all matching documents.

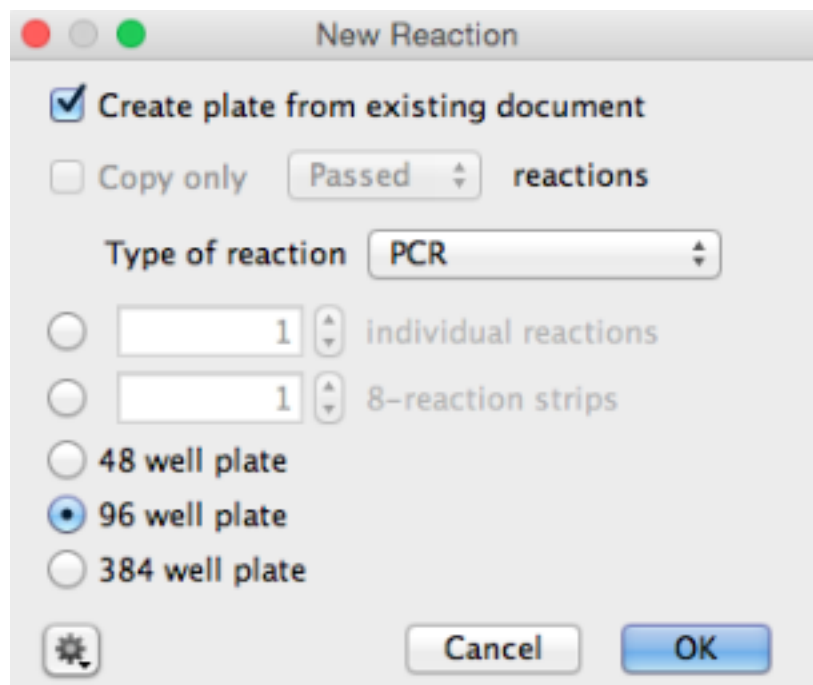
Searches are restricted to particular types of documents by deselecting the checked boxes in the search window. To explicitly search against particular fields, click the “More Options” button. Click the “+” and “–” buttons on the right hand side to add or remove fields from the search. Choose the search fields of interest using the left-most drop-down menu and choose the middle drop-down menu to constrain search conditions.

### 1.11.2 Creating a PCR Plate from an existing extraction plate

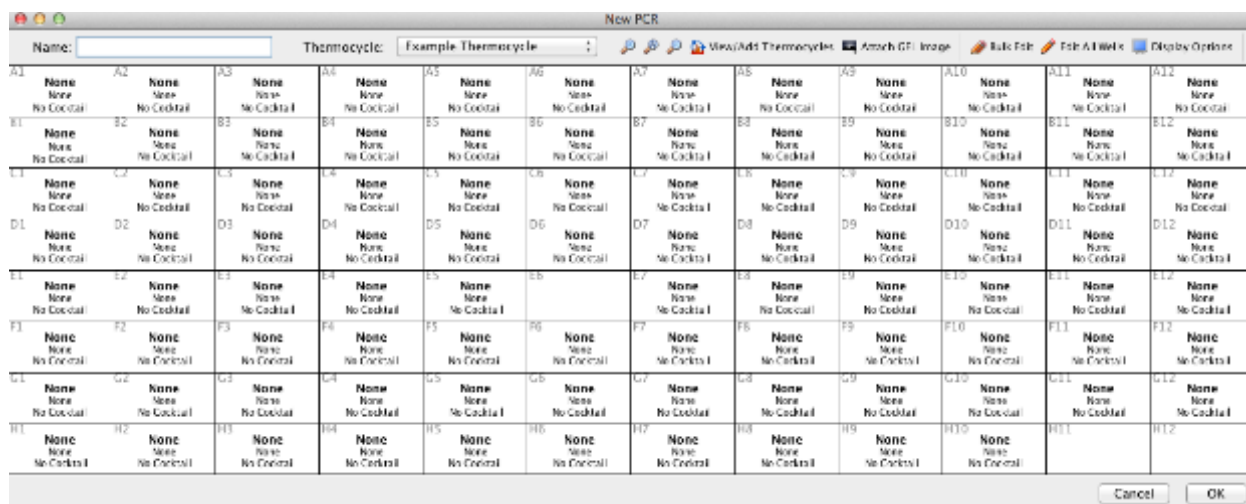
From this point forward, generating new plates, whether PCR or cycle sequencing, is very similar to the way the extraction plate is generated. Geneious can use an existing plate as a guide to create a new plate, so that new reactions are appended onto corresponding workflows. To do this, select an existing plate from the Biocode LIMS plug-in Search Results Window found in the Document Table.

Name	Description
M167_CYC01_RBCLM13_F	A CycleSequencing plate.
M167_CYC01_RBCLM13_R	A CycleSequencing plate.
M167_PCR01_RBCLM13	A PCR plate.
M167_X1	A Extraction plate.
M167_X2	A Extraction plate.

- Click “New Reaction” in the Geneious Toolbar and check the “Create plate from existing document” checkbox.



- If the reaction types are the same (for example, creating an extraction plate as a working aliquot from an archival extraction plate), then all reaction parameters will be copied to the new plate. If the reaction type is different (for example, creating a PCR plate from an existing extraction plate), then only the extraction ID's will be copied across.
- Select the Type of reaction as PCR, click OK and a “New PCR” plate window will appear.



- Enter a plate name in the box provided. This name should be descriptive and unique to allow easy and successful searching in the LIMS. (See [LIMS naming conventions](#))
- The next step is to select the thermocycling protocol that was used for this PCR plate. The Geneious LIMS stores thermocycling profiles and these are found in the “Example Thermocycling” dropdown menu in the New PCR plate Toolbar. Creating new profiles is described in the Additional Information section at the end of this document.
- Next, to edit all wells on the PCR plate, select “Edit All Wells”. Alternatively, to edit individual wells, highlight

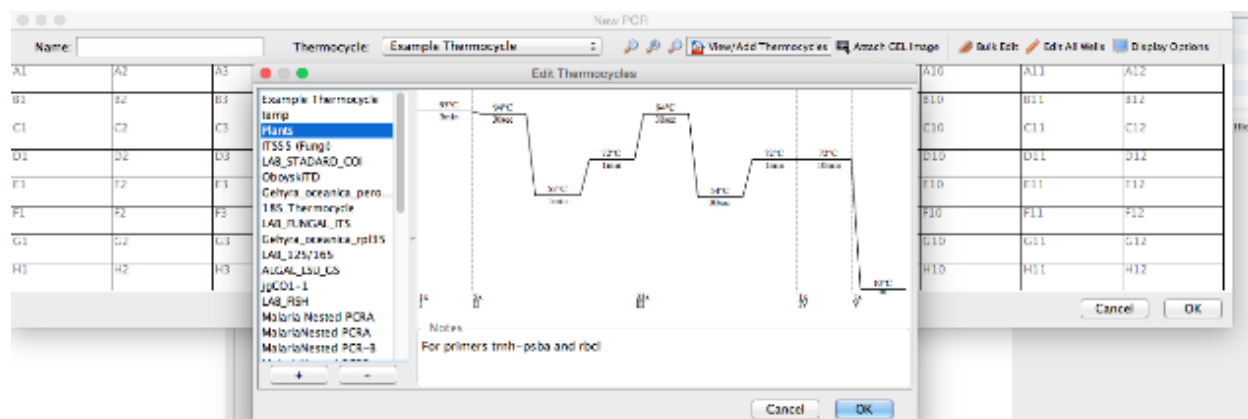
the selected wells in the plate and click “Edit Selected Wells” in the plate toolbar. In both cases the “Edit Wells” window opens.

- This window has a column of checkboxes on its left hand side. Values in the checked fields will be applied to all selected reactions, and unchecked fields will be left as they are.
- Most values can simply be entered into a dialog box, with the exception of PCR cocktails, cycle sequencing cocktails, thermocycler profiles, and primers which are set elsewhere.
- The target locus can be selected in the “Locus” drop-down list. This list is pre-populated, however, it can be manually edited to include additional loci for use in the LIMS.
- A PCR reaction can be marked as “not run”, “run”, “passed” or “failed”.
- The forward and reverse PCR primers are selected using the “Choose” button which points to a selected folder on the Local Directory where primers are stored by the user. (See [Creating primers](#))
- The reaction cocktail can be chosen from a list of existing cocktails, or by creating a new cocktail, described in the Additional Information section at the end of this document.
- Gel images can be uploaded by clicking the “Gel Images” button in the plate window and pointing to the relevant file. More information on attaching a gel image and scoring is available in the Additional Information section below.
- Once all editing of the plate is complete, Click OK.

### 1.11.3 Additional Information

#### Creating thermocycling protocols:

- To create custom thermocycler profiles, click “View/Add Thermocycles” in the “New PCR” window.
- The “Edit Thermocycles” window opens (below) and this displays all of the stored profiles in the left section of the window and a diagram of the selected profile in the right section.



- Click the “+” button on the lower left-hand corner of the window to add a new profile.
- The “New Thermocycle” window will open. The temperatures and cycles associated with the new program can be entered using the dialog boxes and “Edit Cycles” buttons.
- The “+” and “-” buttons to the right of the “Edit Cycle” button allow for addition or removal of steps in the thermocycling profile.

**Note:** It is not possible to modify thermocycler profiles once they are created and profiles can only be deleted if they are not associated with a workflow. As a result, double check any new profiles you create before saving them.

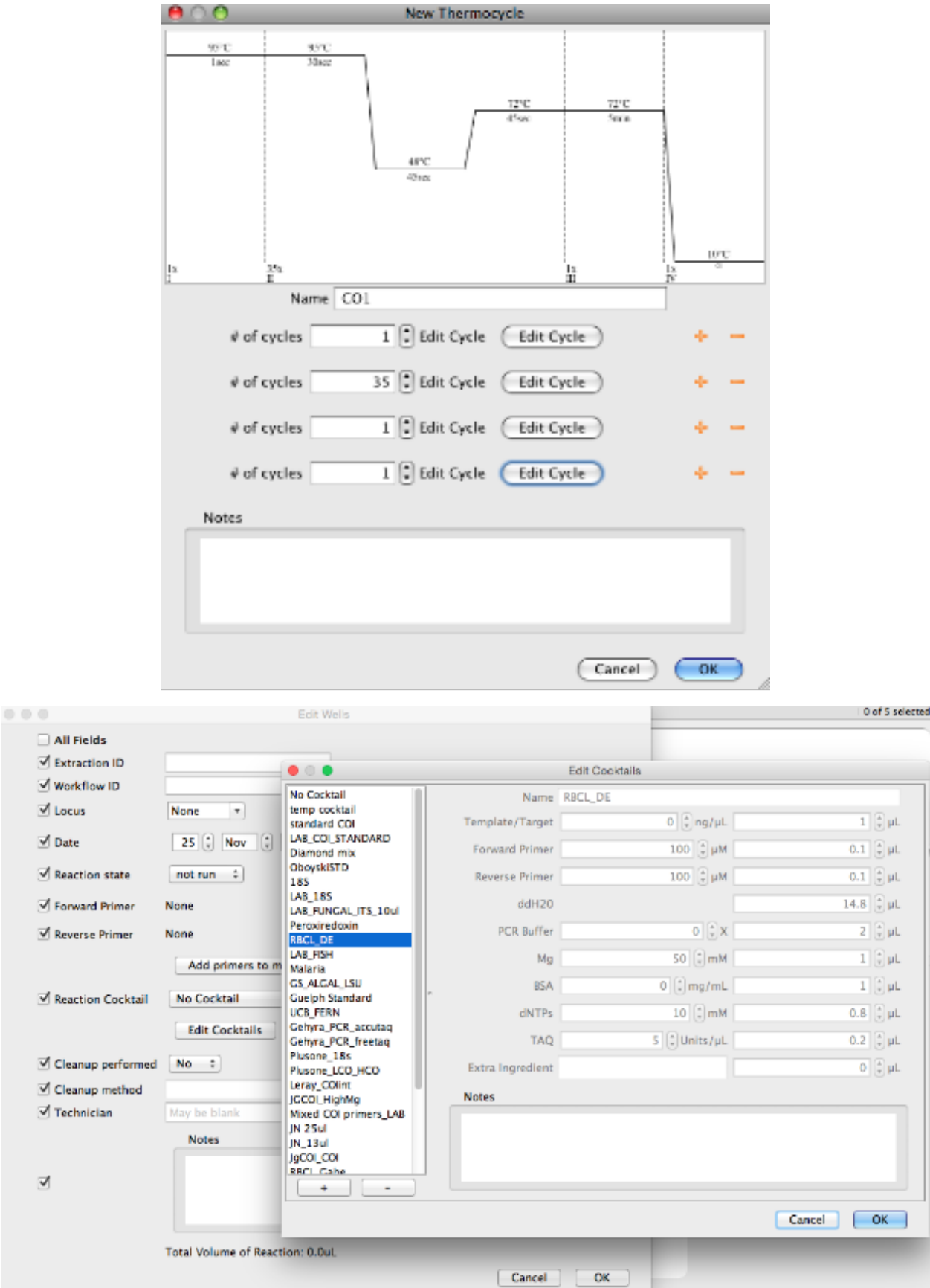
#### Creating custom PCR cocktails:

- Click “Edit Cocktails”, then click the + button in the dialog, and enter the volumes and concentrations in the fields provided.
- There is space to store one extra ingredient (both concentration and volume) that isn’t in the provided ingredient list.
- Any additional information about the new cocktail can be stored in the Notes field.

**Note:** It is not possible to modify cocktails once they are created, and cocktails can only be deleted if they are not associated with a workflow. Cocktails are available also to all users. As a result, double check that the cocktail you require doesn’t already exist prior to creating it.

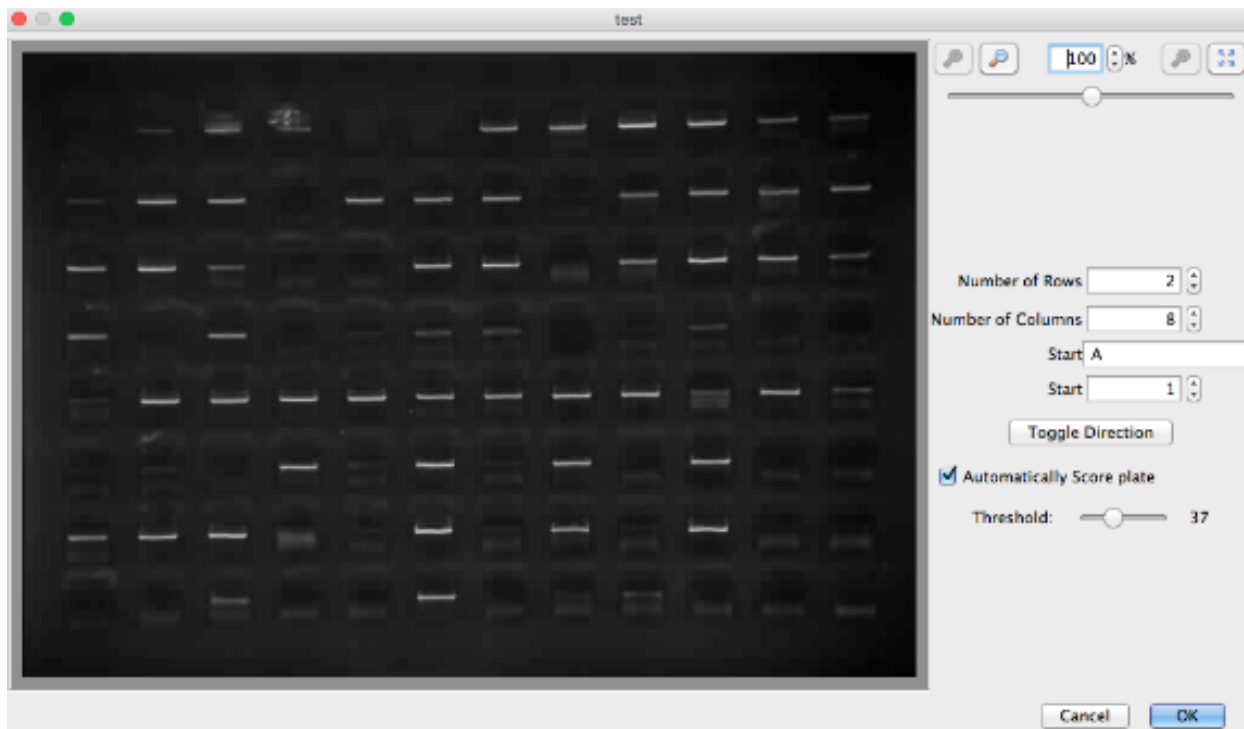
#### Attaching and scoring gel images:

Gel images can be attached to all types of plates at various points in the pipeline. If attaching a gel (or gels) when creating a new plate click the “Attach GEL image” button found in the toolbar of the plate. When attaching a gel or



gels to an existing plate, click “GEL images” in the plate’s toolbar. In both cases the “Edit Gel Images” window will appear. Click “Add”, then browse to find the image or image files to be attached. Geneious will accept images in JPEG, GIF, PNG, and TIFF formats. Notes can also be added each gel image at this point.

Once the image or images are attached, click the “Split GEL” button located above the image in the gel viewer window. To map the samples on the gel to the workflows in the plate, populate the four dialog boxes and choose the read direction of the plate i.e., vertical or horizontal. To generate the grid, drag the mouse/pointer over the image. If the grid is misplaced, start again by dragging the mouse over the image. If the gel image is not straight causing the grid to line up incorrectly, use the sliding bar in the top right of the window to level the image. Automated calling of pass/fail reactions in the gel is possible by checking “Automatically Score plate”. Click “OK” once the grid is in the correct position.



To verify the scoring of the gel image, re-open the plate view. It is possible to manually change the call by selecting the well in the plate viewer and changing the reaction state in the drop-down menu. The program may score primer dimers in the gel, so if they are present, extra care should be taken and the calls double checked.

## 1.12 Generating cycle sequencing plates

### 1.12.1 Creating the plate

- To append new cycle sequencing reactions onto existing PCR workflows, use the Geneious LIMS search window to find the corresponding PCR plate.
- Highlight the PCR plate in the Search Results Window and click “New Reaction” followed by “Create plate from existing document”.
- Select “cycle sequencing” from the drop-down menu for plate type. A cycle sequencing plate, similar to a PCR plate, will be generated.



- Cycle sequencing plates can be edited in the same manner as PCR plates are edited using the “Edit Wells” window.
- As described for PCR plates, set the thermocycling profile, reaction cocktail, primer, date etc.
- Also set the direction as Forward or Reverse.
- If your cycle sequencing plate contains reactions for both directions, then highlight the forward reactions, click on “Edit Selected Wells” and set the direction and primer. Do the same for the remaining reactions in the opposite direction.
- Save plate once complete.

### 1.12.2 Attaching raw traces to a cycle sequencing plate

Once cycle sequencing is complete, the raw trace files can be uploaded to the corresponding cycle sequencing plate or set of reactions.

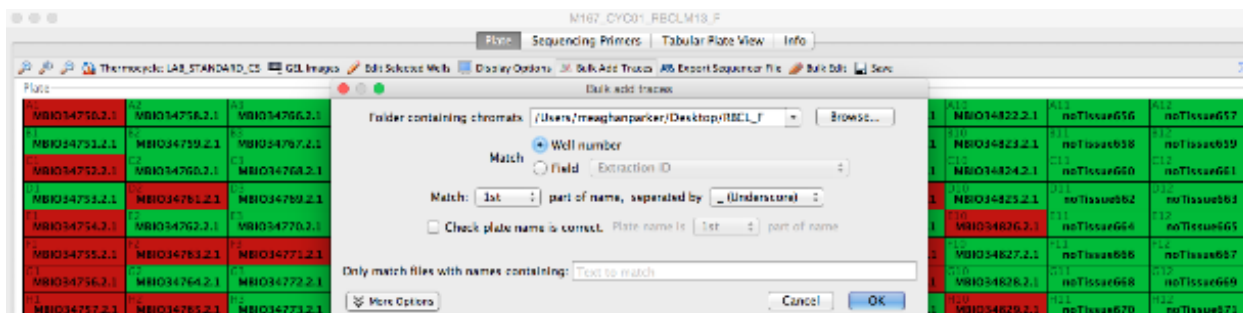
To add a trace file to a single cycle sequencing reaction:

- Click open the reaction well of interest.
- In the “Edit Wells” dialog window that now appears, select the “Add/Edit Traces” button.
- A new window will appear with an “Add Sequence(s)” button located in the upper left corner of the window.
- Click the button and direct Geneious to the location of the raw trace file. Be sure to add only the correct trace (e.g. forward or reverse) to each reaction.
- To remove one or more traces from a well, select the trace or traces in question and click “Remove Sequence(s)”.

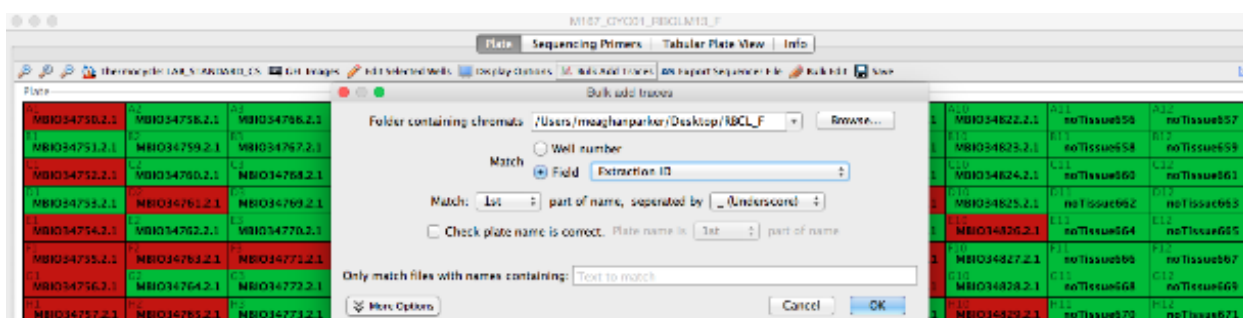
To bulk upload traces to a cycle sequencing plate or set of reactions:

- Open the appropriate cycle sequencing plate and click “Bulk Add Traces” on the plate’s toolbar.
- Click “Browse” to point Geneious to the location of the trace files. Traces are matched to their corresponding cycle sequencing reactions based on components found in the trace file name (i.e., well number or field) along with name delimiters. For example, to attach a sample’s trace file to it’s corresponding well position based on the well position in the trace file name (e.g., 3726294\_\*\*A01\*\*\_capture.ab1), the user would select the “Well number” button followed by “Match 2nd part of name”, “separated by\_(Underscore)” in the “Bulk add traces” dialog window.





- To attach a sample's trace file based on a field (in this case, extraction barcode **3726294\_A01\_capture.ab1**), the user would select the "Field" button followed by "Extraction Barcode" from the drop-down menu along with "Match 1st part of name, separated by \_(Underscore)".



- After all traces have been attached, click "OK" to save.
- To double check that the traces have successfully attached the user can change the display options of the cycle sequencing plate so that # Traces is displayed. If the traces have attached correctly, wells should display "# Traces: 1".

Plate											
A1	A2	A3	A4	A5	A6	A7	A8	A9	A10	A11	A12
# Traces: 1	# Traces: 1	# Traces: 1	# Traces: 1	# Traces: 1	# Traces: 1	# Traces: 1	# Traces: 1	# Traces: 1	# Traces: 1	# Traces: 1	# Traces: 1
B1	B2	B3	B4	B5	B6	B7	B8	B9	B10	B11	B12
# Traces: 1	# Traces: 1	# Traces: 1	# Traces: 1	# Traces: 1	# Traces: 1	# Traces: 1	# Traces: 1	# Traces: 1	# Traces: 1	# Traces: 1	# Traces: 1
C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	C11	C12
# Traces: 1	# Traces: 1	# Traces: 1	# Traces: 1	# Traces: 1	# Traces: 1	# Traces: 1	# Traces: 1	# Traces: 1	# Traces: 1	# Traces: 1	# Traces: 1
D1	D2	D3	D4	D5	D6	D7	D8	D9	D10	D11	D12
# Traces: 1	# Traces: 1	# Traces: 1	# Traces: 1	# Traces: 1	# Traces: 1	# Traces: 1	# Traces: 1	# Traces: 1	# Traces: 1	# Traces: 1	# Traces: 1
E1	E2	E3	E4	E5	E6	E7	E8	E9	E10	E11	E12
# Traces: 1	# Traces: 1	# Traces: 1	# Traces: 1	# Traces: 1	# Traces: 1	# Traces: 1	# Traces: 1	# Traces: 1	# Traces: 1	# Traces: 1	# Traces: 1
F1	F2	F3	F4	F5	F6	F7	F8	F9	F10	F11	F12
# Traces: 1	# Traces: 1	# Traces: 1	# Traces: 1	# Traces: 1	# Traces: 1	# Traces: 1	# Traces: 1	# Traces: 1	# Traces: 1	# Traces: 1	# Traces: 1
G1	G2	G3	G4	G5	G6	G7	G8	G9	G10	G11	G12
# Traces: 1	# Traces: 1	# Traces: 1	# Traces: 1	# Traces: 1	# Traces: 1	# Traces: 1	# Traces: 1	# Traces: 1	# Traces: 1	# Traces: 1	# Traces: 1
H1	H2	H3	H4	H5	H6	H7	H8	H9	H10	H11	H12
# Traces: 1	# Traces: 1	# Traces: 1	# Traces: 1	# Traces: 1	# Traces: 1	# Traces: 1	# Traces: 1	# Traces: 1	# Traces: 1	# Traces: 1	# Traces: 1

- With the traces now attached to their corresponding cycle sequencing reactions, they are ready to be downloaded into the Geneious Assembler module.

## 1.13 Downloading traces from LIMS

The Geneious Assembler Module is used to edit, save, export and ultimately publish the sequence data.

You have two options for importing the raw traces to the Assembler Module:

- Downloading the raw traces from the LIMS database
- Directly importing the traces from disk

The main advantage of the first option is that all of the specimen's associated FIMS and LIMS meta-data are attached to its trace file. Alternatively, if importing traces from disk, none of the specimen's associated FIMS or LIMS meta-data will be attached, and it can prove more difficult to attach this meta-data after import. As a result, you should utilize the first option whenever possible.

### **1.13.1 Downloading the raw traces from the LIMS database**

- Log in to the Biodcode LIMS plug-in.
- Search for the relevant cycle sequencing plate(s) in the Biocode search window.
- Select the relevant plate(s) from the search results.
- Click the “Biocode” button in the Toolbar and choose “Download Traces from LIMS” to begin the operation.
- A new window will open, prompting you to choose a destination folder in the local directory for the downloaded traces. Either select an existing folder or create a new folder. To generate a new folder, highlight the folder in the Local Directory where you want the new folder to be located, select “New Folder”, and provide a name.
- Once the destination folder has been selected, click “OK” and Geneious will begin downloading the trace files into this folder.

Alternatively, if the exact names of cycle sequencing plates are known, it is possible to download the plates directly without having to perform a search. To do this:

- Log in to the Biodcode LIMS plug-in.
- Highlight a destination folder under the Local Directory.
- Select the “Biocode” button in the Toolbar followed by the “Download Traces from LIMS” option and enter the cycle sequencing plate name manually.
- If you want to download traces from more than one plate, use the “+” button to add more fields for the additional plate names.
- Once this is complete, click “OK” and Geneious will begin downloading the trace files into this folder.

### **1.13.2 Directly importing the traces from disk**

There are two ways to do this:

#### **Using the Menu Bar**

- In the local directory, create a destination folder for the raw traces, and highlight it.
- In the main toolbar, click on “File” > “Import” > “From Multiple Files”.
- A new window appears where you can point to the location of the raw traces for import.
- Once complete, click “OK” and the import will begin.

## Drag-and-Drop

- Outside of Geneious, navigate in your computer's file explorer to find the location of your trace files.
- Select the trace files to import, and drag them into Geneious and into a selected folder in the local directory.

Once the raw trace files have been imported, it is necessary to define the read direction. To do this:

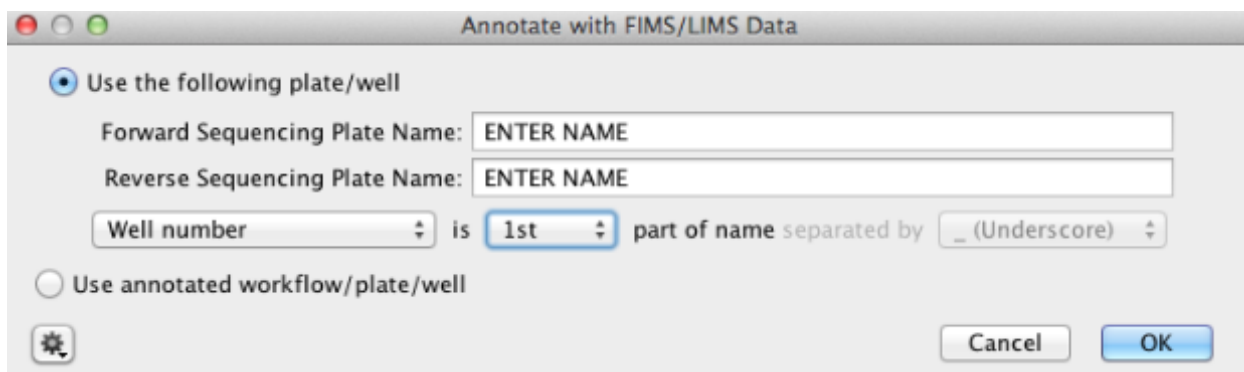
- Choose either the forward or reverse trace files and select "Sequence" in the Menu bar followed by "Set Read Direction".
- Choose the read direction and click "OK". It is only necessary to choose the direction for one set of reads because the other set of reads will be assigned the opposite direction by default.
- After performing this task, an extra column will be added to the Document Table titled, "Is Forward Read" with a value of true or false.

If the forward and reverse traces are in different folders, it's easiest to import all of the traces from one folder, set the read direction for that folder and then import the second folder. The Search/Filter dialog box in the upper right-hand corner of the Geneious window can be used to locate a particular direction of reads based on trace names if both forward and reverse traces are imported in one folder.

Traces imported into Geneious directly from disk have none of the specimen's associated FIMS or LIMS meta-data attached. It is possible to annotate traces with the associated meta-data from the FIMS, but this must be done pre-assembly (with the traces) because forward and reverse traces can come from different sequencing plates.

To attach the associated meta-data:

- Click the "Biocode" button in the Toolbar, then select "Annotate from FIMS/LIMS Data".
- A new window appears where you enter the forward and reverse sequencing plate names (from the LIMS) that correspond to the traces. You must also identify which part of the raw trace name identifies the well location.
- Click "OK" and the operation will run



## 1.14 Assembling contigs

Once sequencing is complete, the raw data needs to be assembled into bidirectional contigs. The Geneious assembler module is used to edit, save, export and ultimately publish the data.

Two slightly different pipelines can be employed for trimming the raw traces and assembling them into contigs:

1. Incorporating the trimming of traces into the assembly function
2. Trimming all traces first, then assembling into contigs

---

**Note:** Any settings or parameter values outlined below are guidelines. It is recommended that the user evaluates different trim and assembly settings to establish what is optimal for the group and/or marker being analyzed.

---

### 1.14.1 OPTION 1: Trimming and assembling at the same time

- Select all the raw traces you want to assemble into a contig.
- Click on “Align/Assemble” in the toolbar.
- Select “De Novo Assemble” in the dropdown menu.

#### De Novo Assemble window

- In the “Data” section, specify the unique part of the trace filename you want to use to match the forward and reverse direction traces. This will depend on how you labelled your raw traces.
- In the “Method” section, choose the Assembler and Sensitivity. We recommend the options displayed below, however, it is also possible to choose “Custom Sensitivity” from the dropdown menu, and choose custom parameters (for example, minimum overlap). If you select “Custom Sensitivity” many of the options in the “Advanced” section of the window will now be available for selection.
- In the “Results” section, insert an Assembly Name. After the assembly is complete all contigs will be deposited in a folder in the local directory with this title. “Save assembly report” and “Save results in a new subfolder” should both be selected.
- In the “Advanced” section, ensure that the “Circularize contigs with matching ends” option is unchecked.
- In the “Trim Sequences” section select “Trim Sequences” and then click on the Option button. A “Trim Options” window appears and this is discussed further below.

#### Trim Options window

When creating your sequencing assemblies, be sure to select the correct options in the “Trim Options” window.

The first option to choose between is “Annotate new trimmed regions” or “Remove trimmed regions from sequences”. If “Annotate new trimmed regions” is selected, then the information in the trimmed region of the trace is annotated and not deleted. The underlying raw data is maintained throughout downstream analyses for possible adjustment later in the pipeline. Assembly and other analyses automatically take the trims into account, and exclude these regions in all calculations.

If either “Remove new trimmed regions from sequences” or “Remove existing trimmed regions from sequences” is selected (the latter option will only show in this window when the selected traces have been previously trimmed), then the trimmed regions are deleted and the associated information will not be available for downstream analyses.

“Trim vectors” uses a clone of NCBI’s VecScreen tool (Altschul et al, 1997) to screen for vector contaminants from the NCBI UniVec database.

Check “Trim Primers” to trim the appropriate PCR primers from each raw trace. Clicking the “Choose” button points you to the list of primers you have saved in your local directory. It is possible to choose as many primers for trimming as required.

- If you are using M13 tagged PCR primers, you have two options for trimming. You can trim the full primer(s) including the M13 tag or alternatively you can just trim the untagged portion of the primers. We would lean towards the latter option, since the chances of finding the longer tagged version of the primer is slimmer than finding the untagged version.

De Novo Assemble

**Data**

☒ Assemble by: 1st part of name, separated by \_ (Underscore)   
☐ Assemble each sequence list separately

**Method**

Assembler: Geneious   
Sensitivity: Highest Sensitivity / Slow ?   
Memory Required: 24 MB of 7.5 GB


**Trim Sequences**


☐ Use existing trim regions   
☐ Remove existing trim regions from sequences   
☒ Trim sequences Options   
☐ Do not trim

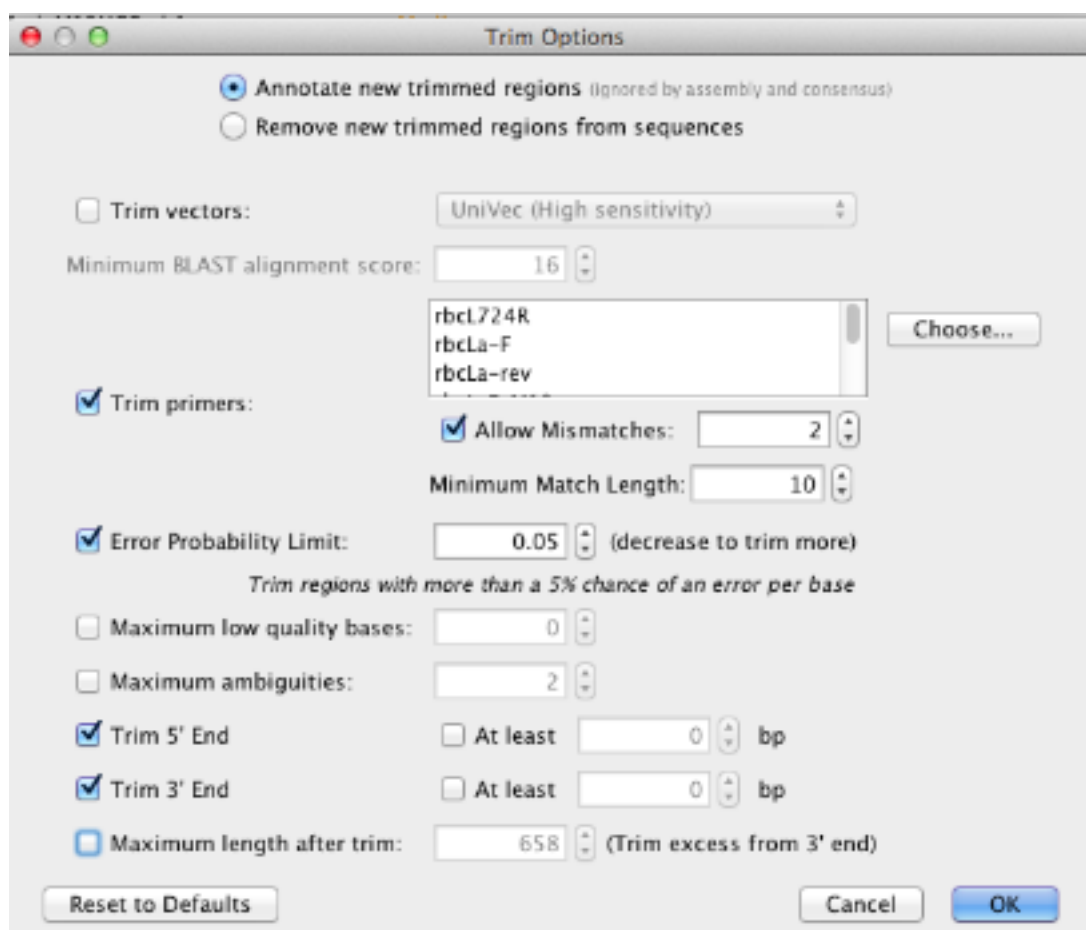
**Results**

Assembly Name INSERT NAME   
☒ Save assembly report   
☐ Save list of unused reads   
☒ Save in sub-folder   
☒ Save contigs   
☐ Save consensus sequences Options

**Advanced**

☒ Don't merge variants with coverage over approximately 6 ☒ Merge homopolymer variants   
☐ Produce scaffolds ☐ Circularize contigs with matching ends   
*To specify any of the settings below, choose the Custom Sensitivity method*   
☒ Allow Gaps Maximum Per Read: 20 % Maximum Gap Size: 5   
☐ Minimum Overlap: 25 ☐ Minimum Overlap Identity: 80 %   
Word Length: 10 Index Word Length: 10   
☐ Ignore words repeated more than 1,000 times Reanalyze threshold: 0   
Maximum Mismatches Per Read: 50 % Maximum Ambiguity: 16   
Low Memory Use  Speed   
☒ Use paired read distances to improve assembly   
☐ Only use paired hits during assembly

 Fewer Options Cancel OK



The image shows a 'Trim Options' dialog box with the following settings:

- ☒ Annotate new trimmed regions (Ignored by assembly and consensus)
- ☐ Remove new trimmed regions from sequences
- ☐ Trim vectors: UniVec (High sensitivity)
- Minimum BLAST alignment score: 16
- ☒ Trim primers:
  - ☒ Allow Mismatches: 2
  - Minimum Match Length: 10
- ☒ Error Probability Limit: 0.05 (decrease to trim more)  
*Trim regions with more than a 5% chance of an error per base*
- ☐ Maximum low quality bases: 0
- ☐ Maximum ambiguities: 2
- ☒ Trim 5' End
  - ☐ At least 0 bp
- ☒ Trim 3' End
  - ☐ At least 0 bp
- ☐ Maximum length after trim: 658 (Trim excess from 3' end)

Buttons: Reset to Defaults, Cancel, OK

- For Allow Mismatches we suggest you select “2” - if you insert too high a number here it will “find” the primers at the incorrect locations of the read.
- For a primer of 15-20 bp in length, we suggest you select “10” for Minimum Match Length. If you select too low a number here it will “find” the primers at incorrect locations of the read.

“Error Probability Limit” value of 0.05 is an appropriate starting value for trimming. This option works by trimming the trace sequence to find the longest possible untrimmed region, which has an overall error probability less than 5%. To trim more aggressively, decrease the limit or manually edit the trim by clicking and dragging either end of the annotation in the Sequence View.

- Select both Trim 5’ End and Trim 3’ End options, but leave “At least” UNCHECKED.
- Leave “Maximum length after trim” UNCHECKED.
- Once all settings are complete hit “OK”. This will bring you back to the “De Novo Assemble” window where clicking “OK” will begin the trim/assembly process. All the assemblies will be deposited into the folder you specified/named in your local directory. When the operation is finished, an Assembly Report and list of Consensus Sequences will also be added to the folder. Geneious will generate a new subfolder each time an assembly is run.

### 1.14.2 OPTION 2:Trimming and assembling in two separate steps

To trim traces:

- Highlight all relevant traces in the Document Table.
- Select “Annotate & Predict” from the Menu bar.
- Choose “Trim Ends” from the drop-down menu.

A “Trim Ends” window will open and this is essentially the same as the “Trim Options” window described earlier. Proceed with directions as laid out for that window.

To assemble previously trimmed raw traces into a contig: \* Select all of the traces to assemble (and a reference sequence or list if applicable) then click the “Align/Assemble” drop-down button in the Toolbar. \* Select “De Novo Assemble”. \* The “De Novo Assemble” window opens. See previous section for description of the options available in this window. The only difference is noted below. \* Since the trace sequences were previously trimmed, select “Use existing trim regions” in the Trim Options part of the window.

### 1.14.3 Additional Information

It is possible to re-trim trace sequences using different parameters at any stage. To do so select the traces for re-trimming and follow the steps outlined above. The only difference is the “Annotate new trimmed regions” option should be selected to have the new trims replace the old trims. When a trace sequence is re-trimmed, it stores the history of trims in the “Info” tab for each trace.

Manually trimming traces is also an option. To manually trim a trace, select a region at the end of the trace in the Sequence View, click “Annotate” and choose “Trimmed” for the annotation type. If a trace has multiple trimmed annotations for the same region, the largest trimmed annotation will be used.

The assembly report is found in the assemblies folder, regardless of implementing Option 1 or Option 2. It provides a record of which traces were assembled successfully and which traces failed. The blue hyperlink next to the green checkmark at the top of the report links to all traces that assembled successfully. The blue hyperlink next to the red “X” points the user to all traces that failed to assemble. Click the hyperlink and select the “Mark as Failed in LIMS” tool to mark these failed traces for re-sequencing.

The screenshot displays the Geneious 7.1.4 software interface. The left sidebar shows a tree view of sources, including 'Local (0)', 'RBCL\_M167 (192, 189 unread)', 'M167\_M13\_RBCL\_Assembly (60)', 'TEST (3745, 922 unread)', 'Deleted Items (60)', 'Shared Databases', 'Operations', 'Biocode (5)', and 'NCBI'. The main window is titled 'Geneious 7.1.4 - For non-comm' and contains a toolbar with icons for Back, Forward, Sequence Search, Agents, Align/Assemble, Tree, Primers, Cloning, Biocode, New Reaction, Back Up, Support, and Help.

The 'Assembly Report' is displayed in the main window. It shows a summary of the assembly process and a list of reads that were not assembled.

### Assembly Report

✓ 116 of 192 reads were assembled to produce 58 contigs  
 ✗ 76 reads were not assembled

Assembled using Geneious assembler. [Show Options](#)  
 Assembly Duration: 6.01 seconds (8.15 seconds CPU time)

#### ✗ Not Assembled

**F05.ab1 - 2 Reads**

- [M167\\_M13\\_RBCL\\_F\\_F05.ab1](#)
- [M167\\_M13\\_RBCL\\_R\\_F05.ab1](#)

**H12.ab1 - 2 Reads**

- [M167\\_M13\\_RBCL\\_F\\_H12.ab1](#)
- [M167\\_M13\\_RBCL\\_R\\_H12.ab1](#)

**H11.ab1 - 2 Reads**

- [M167\\_M13\\_RBCL\\_R\\_H11.ab1](#)
- [M167\\_M13\\_RBCL\\_F\\_H11.ab1](#)

**H08.ab1 - 2 Reads**

- [M167\\_M13\\_RBCL\\_F\\_H08.ab1](#)
- [M167\\_M13\\_RBCL\\_R\\_H08.ab1](#)

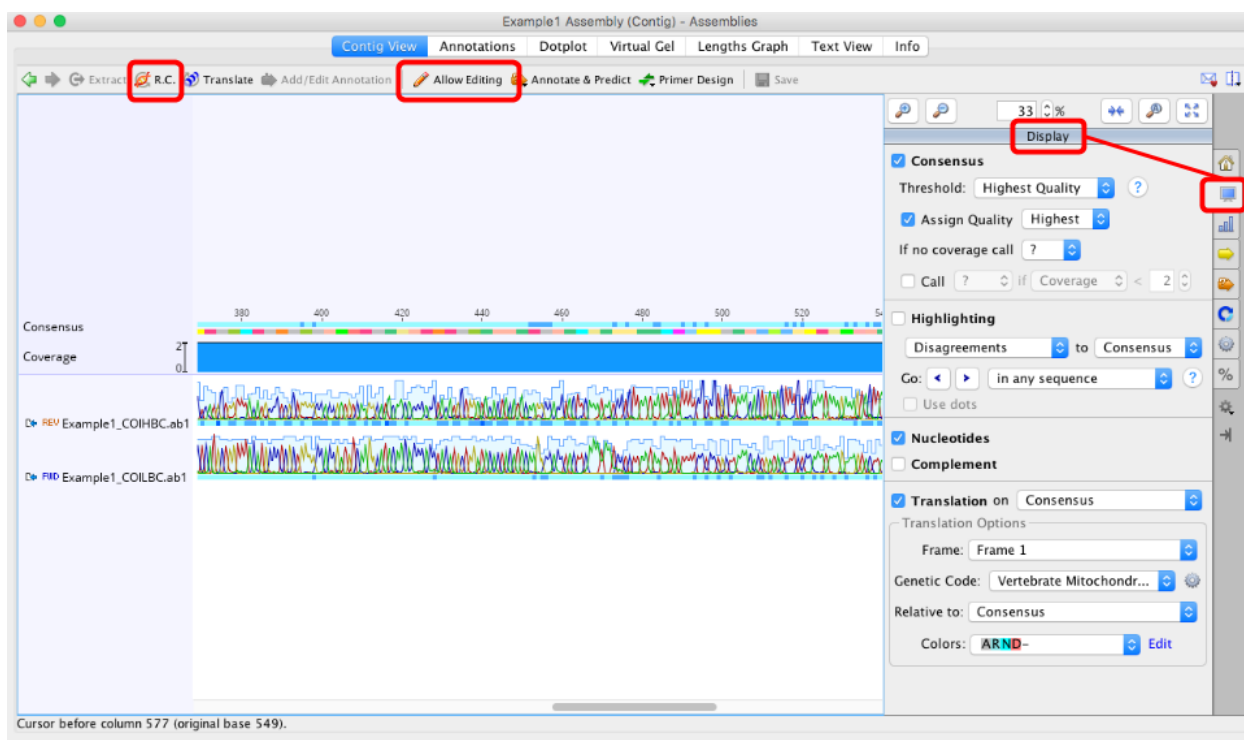


## 1.15 Quality control of assembled contigs

Go to the relevant folder in the local directory where the assembled bidirectional contigs are stored. Select and open an assembly.

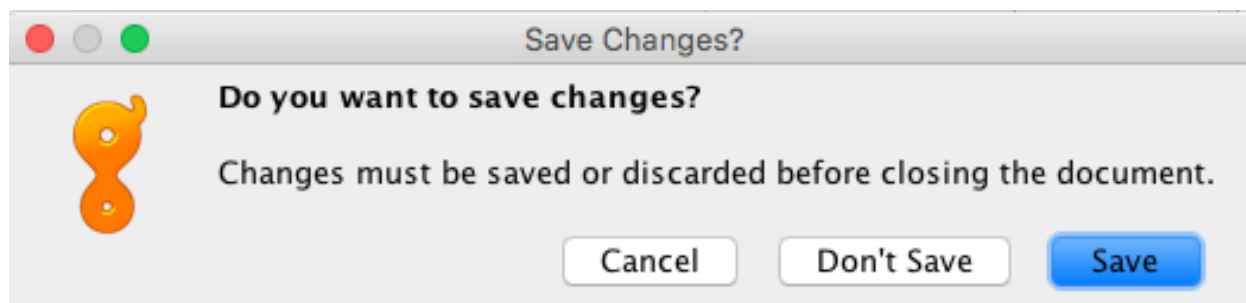
### 1.15.1 Manually editing assemblies

- **If you are editing protein coding genes**, check the “Translation” option in the right hand menu of the Display window
  - Set the correct genetic code (“Vertebrate Mitochondrial” or “Invertebrate Mitochondrial” for COI, or “Bacterial” for rbcL and matK) and
  - Select the correct reading frame. Black dots = stop codons, so we do not want any of these. If stop codons are present double-check the following:
    - \* the correct genetic code is selected,
    - \* the assembly is in the correct orientation (Use “R.C.” button in top left of contig window if you need to reverse complement it),
    - \* whether insertions or deletions are present, and/or
    - \* check BLAST to verify it is not a contaminant

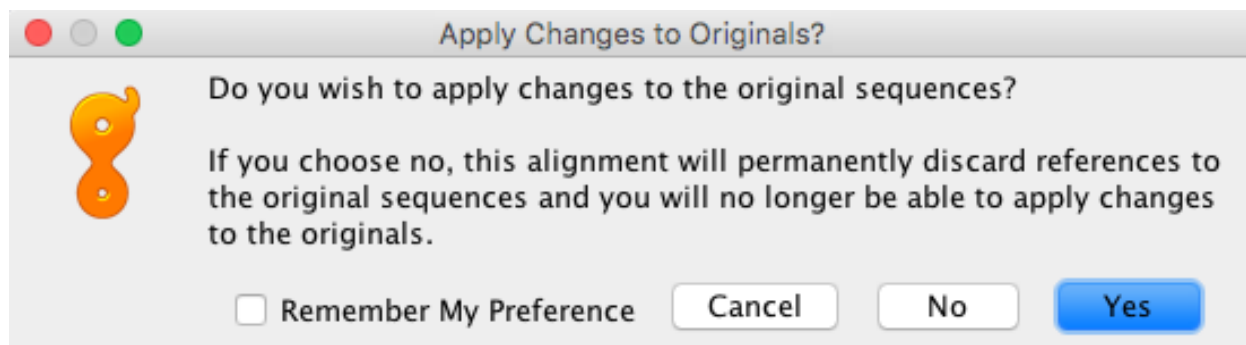


- Quickly scan through the individual assemblies and assess whether or not each disagreement (if present) needs a manual edit.
- A manual edit **ONLY** needs to be made if you feel the consensus sequence has been called incorrectly (or there is a gap that needs to be deleted). If Geneious calls the consensus sequence correctly, **NO** changes should be made to individual traces.

- To manually edit an assembly, the “Allow Editing” button in the toolbar of the contig window should be clicked on (see image above).
- If you are unhappy with the trimmed portions, you can edit these manually by clicking on, and dragging, the red bar indicating the trimmed region.
- **Do not forget to save your edits.** You will be prompted to do this when you try to close the assembly.



- In addition, another prompt window will ask if you want to apply changes to the original sequences. **ALWAYS Click “Yes”, because you risk losing connection to reference sequences.**



### 1.15.2 Checking sequence quality with alignments

A second quality check is made by aligning your sequences based on the gene - align COI sequences together, rbcL together, etc.

For alignments of protein coding regions Geneious’s Translation Alignment program doesn’t work like it should, so use an alternate program such an online program called TranslatorX (<http://translatorx.co.uk>) to create an alignment.

---

**Note:** It’s important to note that TranslatorX only checks the forward reading frames, so you need to Reverse-Complement the matK sequences before putting them into this alignment program otherwise you will receive errors.

---

- Export the consensus sequences (of good assemblies only) as a FASTA file then import this file into the program. We suggest you leave the Protein Alignment Option method selected as “Muscle”. In the Genetic Code box select the relevant reading frame and be sure to check the “Guess most likely reading frame” option. Then hit Submit Query.
- If the program runs OK and doesn’t encounter any errors, it will return an alignment of the nucleotides and also an alignment of the amino acids. You may download the fasta file of both, however, the alignment of amino acids is what will be used for the second quality check. Import the fasta file(s) of the alignments into Geneious for further analyses.

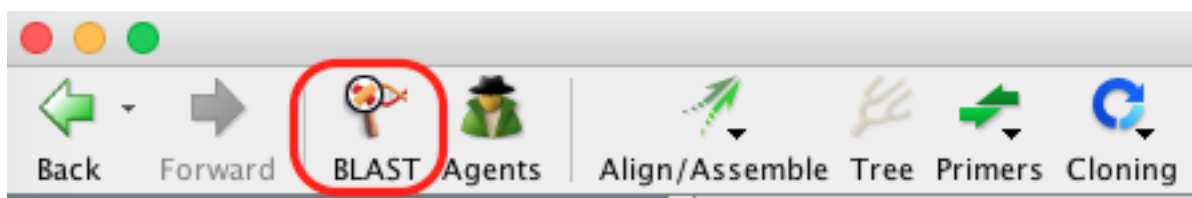
- Use the alignment to address any issue that you can see i.e. a clear difference between one sequence to the others (Remember this can be possible if the sequences are distantly related but still cross reference the alignment to the individual assemblies). Also, gaps must be assessed and resolved. Major differences in the alignment may also indicate that one or more of the sequences are contaminants (use BLAST to determine this).
- You may need to repeat the alignment step a number of times as you cross reference the assemblies and make edits. Save the edits, re-export all the consensus sequences and create a new alignment with these new consensus fasta files.
- If more than a handful of edits need to be made to the consensus sequence, the assembly should be discarded and the sample re-sequenced. You need to make a judgement call on this.

### 1.15.3 BLAST

Geneious provides the ability to BLAST your sequences from within the program. You can read more about BLAST on the NCBI BLAST website here: <http://blast.ncbi.nlm.nih.gov/Blast.cgi>.

To use BLAST, follow these directions:

- Select a single or batch of contigs you want to compare to the NCBI public DNA sequence database – we recommend small batches of less than 15 sequences, since the process can be quite slow otherwise – and click on the “BLAST” button in the Toolbar.



- The “BLAST” window appears and has multiple options for consensus, GenBank database, program etc. The selections in the image below are our recommendations for querying COI sequences, however we encourage all users to look into the various options and decide what works best for your dataset.
- The consensus options allows you to choose how the program will call the consensus sequence of each assembly.
- Once you have made your selections, click “Search” button in the “BLAST” window.
- The search progress appears in the Document Window. If this is too slow, or you want to exit the search for whatever reason, click on the “Stop” button in the top left of the Document Window.
- Once complete, the results are saved in a subfolder (folder name ends with “- nr Megablast”) within the folder containing your query sequence(s). If you did a batch search, there will be further subfolders containing BLAST results for each of the sequences you entered into the BLAST search.
- In the results folder the BLAST results are displayed in the “Hit Table” tab. Various information is included e.g. Hit Accession number, Query coverage, % Pairwise Identity, etc. You can chose what is displayed by clicking on the manage columns icon found in the upper right of the table. Further information is found in the other tabs of the folder (Query Centric View, Annotations, Distances, Info).
- To get more information about the individual BLAST hits, select one of the hits and the information about that sequence appears in the Document Viewer. Any of the columns can be sorted, rearranged, or resized as usual.

BLAST

☒ Batch search of 2 consensus sequences Consensus Options

Query: ☐ Selected sequences (select several to batch search)  
☐ Enter unformatted or FASTA sequence

Database:  ▼ Add/Remove Databases

Program:

Results:  ?

Retrieve:

Maximum Hits:

☐ Low Complexity Filter Word Size:

☐ Human Repeats Filter Max E-value:

☒ Mask for lookup table Gap cost (Open Extend):

Scoring (Match Mismatch):

Other Arguments:

Entrez Query:

Fewer Options Cancel Search

Consensus Options

Threshold:  ?

☐ Ignore Gaps

☒ Assign Quality

If no coverage call

☐ Call  if Coverage  <

☐ Split into separate sequences around '?' calls

☐ Append text to name of alignment

Fewer Options Reset to Defaults Cancel OK

Hit Table   Query Centric View   Annotations   Distances   Info						
<input type="checkbox"/>	Query ▲	Name	Accession	Query coverage	% Pairwise Identity	Sequence Length
	Test1 Assembly	DQ433631	DQ433631	98.23%	99.8%	665
	Test1 Assembly	DQ433632	DQ433632	98.23%	99.8%	665
	Test1 Assembly	DQ433633	DQ433633	98.23%	100.0%	665
	Test1 Assembly	EU525404	EU525404	100.00%	99.6%	677
	Test1 Assembly	JX516072	JX516072	100.00%	99.7%	677

## 1.16 Marking assemblies as Pass or Fail in LIMS

### 1.16.1 Using the “Mark as Pass in LIMS” tool

- Log in to the Biodcode LIMS plug-in.
- Select the assemblies that you want to pass in LIMS.
- Select the ‘Biocode’ Icon on the toolbar and then ‘Mark as Pass in LIMS’. A new window will appear
- Leave the “Remove previous final sequencing results” checked (see below)
- Uncheck “Also attach raw traces to sequencing reactions in LIMS”
- Add any relevant notes in the notes field under your name
- Leave the rest of the fields as the default selection, and press “OK” for the operation to run.

### 1.16.2 Using the “Mark as Fail in LIMS” tool

Once you have marked your final sequences as ‘Passed’ in LIMS, you need to mark your failed sequences in LIMS in the same manner:

- Select your failed sequences.
- Select the “Biocode” icon on the toolbar and then “Mark as FAIL in LIMS”.
- Uncheck “Also attach raw traces to sequencing reaction in LIMS”.
- Fill in the “Reason Details” and “Notes” fields with any relevant information.
- Keep the Consensus sequence settings at their default, and press “OK” for operation to run.

## 1.17 GenBank Upload

### 1.17.1 Installing the Geneious Plug-in

---


**Note:** The most up-to-date GenBank Submission plug-in version is 1.6.4, which can be found [here](#).

---

We will be using Geneious’s GenBank Submission plug-in to submit completed sequences to GenBank. The Geneious GenBank Submission Plug-in does all of the hard work of bundling together the various parts of a GenBank submission – sequence data, specimen metadata, trace files, etc.

Mark as Pass in LIMS

The generated consensus sequences will be uploaded to the LIMS as the final sequencing result.

 The sequences being uploaded should be ready for submission to public databases.  
(Of sufficient quality, have had all edits completed etc)

☒ Remove previous final sequencing results

Traces

☐ Also attach raw traces to sequencing reaction in LIMS

Note: If you downloaded your chromatograms from the LIMS, you do not need to add them again.

Details

Your name

Notes

Consensus

Threshold:


☐ Ignore Gaps

☒ Assign Quality

If no coverage call 

☐ Call 

☐ Split into separate sequences around '?' calls



Mark as Fail in LIMS

☒ Remove previous final sequencing results

Traces

☐ Also attach raw traces to sequencing reaction in LIMS

Note: If you downloaded your chromatograms from the LIMS, you do not need to add them again.

Details

Your name

Reason

Reason Details

Notes

Consensus

Threshold:


☐ Ignore Gaps

☒ Assign Quality

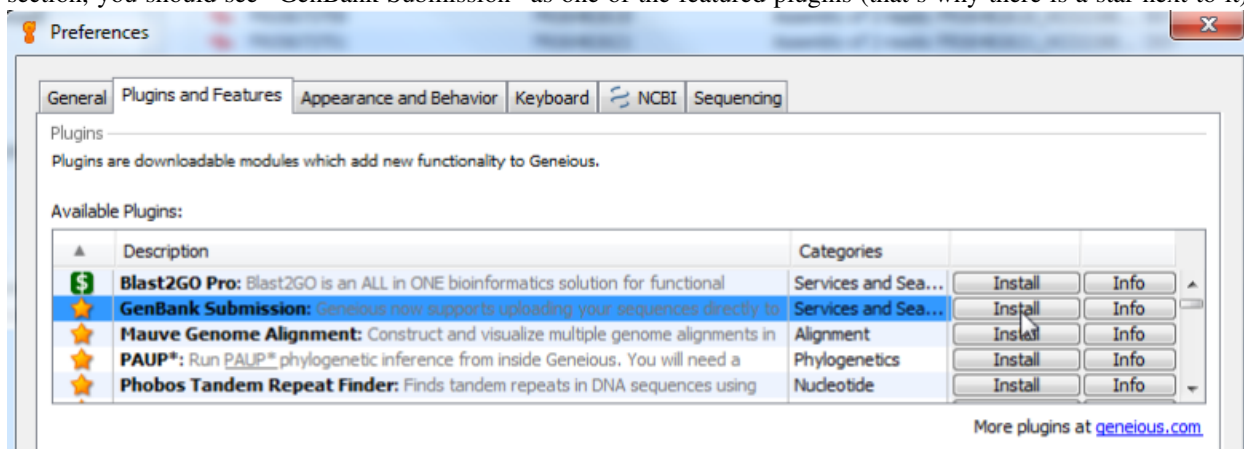
If no coverage call 

☐ Call

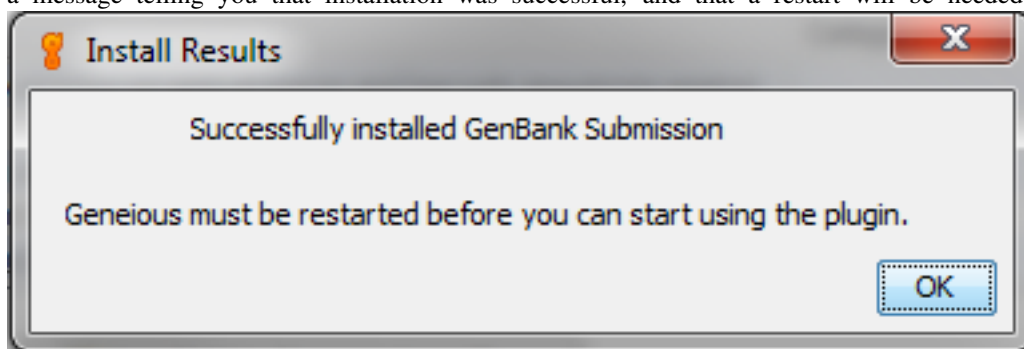
☐ Split into separate sequences around '?' calls



The easiest way to install the plug-in is within Geneious. Go to Tools > Plugins. In the top “Available Plugins” section, you should see “GenBank Submission” as one of the featured plugins (that’s why there is a star next to it).



Click “Install”, and Geneious will start downloading it from the Internet. If all goes well, you should see a message telling you that installation was successful, and that a restart will be needed. Restart Geneious.



### 1.17.2 Using the plug-in

Because traces are a required part of a BARCODE keyword record on GenBank, **use the Assembly as the basis of GenBank submission.**

Organize the assemblies you want to submit, and then go to Tools > Submit to GenBank. Make sure to only select assemblies from one gene at a time. For example, you will have to make a separate submission package for rbcL and matK sequences.

You will see a window appear that has the following sections. Each section is detailed below.



**Submit to GenBank**

**You are about to create a submission of a single sequence.**

**Submission Details**

Submission Name:  [Edit Publisher Details...](#)

Submission Type:

☒ Save a local file (.tar)

☐ Upload New Submission

☐ Update Existing Submission BankIt ID:

Account:

☒ Submit using the Geneious BankIt FTP account

☐ Submit with your own BankIt FTP Account [Configure](#)

**Fields Mapping**

Map fields on your document to GenBank fields ?

Project Name:  Country:

Specimen Voucher:  Sequence ID:

Identified by:  Collection Date:

Collected by:  Organism:

Molecule Type:  Genetic Code:

Genetic Location:

☐ Include Features/Annotations ☐ Also include annotations in tracks ?

☒ Add Gene & CDS features using fields: [Choose](#) ? **Gene and CDS Features**

☐ Include quality scores (if possible)

☒ Include extra fields: [Choose...](#)

☐ Include structured comments: [Choose...](#) **BWP Structured Comments**

**Consensus and Primers**

Consensus ☒ Traces ☒ Sequencing Primers ☒ Primers

Threshold:

☐ Ignore Gaps

☒ Assign Quality

If no coverage call

☐ Call ? if Coverage <

☐ Split into separate sequences around '?' calls

**Traces and Sequencing Primers**

[OK](#) [Cancel](#)

### 1.17.3 Submission Details

The first part of the GenBank submission deals with filling out the contact details and attributions for your sequence submission, as well as choosing how to submit to GenBank. Give the

“Submission Name” field a descriptive name for your submission. This entry will not show up in the GenBank record. Also, be sure to select the option of “Save a local file (.tar)”.

Submission Name:

Submission Type:

☒ Save a local file (.tar)

☐ Upload New Submission

☐ Update Existing Submission BankIt ID

Account

☒ Submit using the Geneious BankIt FTP account

☐ Submit with your own BankIt FTP Account

Click on the “Edit Publisher Details...” button to bring up the Publisher Details dialog box.

**Publisher Details**

Contact Information

First Name:  Last Name:

Phone:  Fax:

Email:  (not in GenBank record)

Affiliation

Institution:  Department:

Street Address:  City:

State/Province:  ZIP/Postal code:

Country:

Sequence Authors:

First Name:  Middle Initial:  Last Name:  Suffix:

First Name:  Middle Initial:  Last Name:  Suffix:

Consortium:

Publication Status: ☒ Unpublished ☐ In-Press ☐ Published

Reference Title:

Reference Authors: ☒ Same as Sequence Authors ☐ Specify new authors

**Contact Information** Fill out the top section with the contact information for your submission. This will be the information that GenBank staff will use to contact the submitter with questions or updates for the submission.

**Affiliation** Fill out the relevant information for the institution that produced these sequences. The entries in this section will show up in the GenBank record, so be sure to provide accurate and consistent information.

**Sequence Authors** List as many people who were involved in the production of these sequences as you can think of. Keep in mind that only these people will be authorized to make changes to the GenBank record.

**Consortium** Always fill this in as “Barcode of Wildlife Project”. This will allow for easier searching and filtering of all BWP-generated sequences.

**Publication Status and Title** We will be generating several different scientific publications about the sequences generated from the Barcode of Wildlife Project, but for now select “Unpublished” for the Publication Status, and

“Barcode of Wildlife Project [country]” (where [country] is the name of the BWP partner country that produced the sequence.) Again, this allows for easier searching and filtering, and we will be able to add publications to sequences as they are published.

LOCUS	ABC123_COI	655 bp	DNA	linear	INV 16-SEP-2015
DEFINITION	Sample GenBank Record				
ACCESSION					
VERSION					
KEYWORDS	BARCODE.				
SOURCE	mitochondrion				
ORGANISM	Genus species				
	Kingdom; Phylum; Class; Order; Family; Genus; Species.				
REFERENCE	1 (bases 1 to 655)				
AUTHORS	Trizna,M.G. and Bauer,J.				
TITLE	Barcode of Wildlife Project [Country]				
JOURNAL	Unpublished				
REFERENCE	2 (bases 1 to 655)				
AUTHORS	Trizna,M.G. and Bauer,J.				
TITLE	Direct Submission				
JOURNAL	Submitted (16-SEP-2015) CBOL, Smithsonian Institution, 10th and Constitution Ave NW, Washington, DC, USA				

Fig. 1: This portion of a sample GenBank record shows how the different parts of the Publisher Details section will appear when the record is published.

#### 1.17.4 Fields Mapping

The next part of the GenBank submission, will be to map all of the different specimen metadata fields to your GenBank record.

**Project Name** Just like the “Submission Name” field at the beginning, this entry won’t end up in the GenBank record, but should be a meaningful name used to organize your sequences.

**Country** This will become the “country” field in GenBank. It corresponds with the FIMS field “countryOrOcean”, which has already been validated to be part of the NCBI country list.

**Specimen Voucher** This will become the “specimen\_voucher” field in GenBank. It corresponds with the FIMS field “voucherID”, which should be a colon-separated triplet comprised of [institution-Code]:[collectionCode]:[catalogNumber].

**Sequence ID** This field will not be published as part of the GenBank record, but it is very important because this field will connect the specimen data, sequence data, and trace data. Select the LIMS field “Workflow Name” for this.

**Identified by** This will become the “identified\_by” field in GenBank. It corresponds with the FIMS field “identifiedBy”. It is required for the BWP Data Standard, but if it is unknown you can select *None*.

**Collection Date** This will become the “collection\_date” field in GenBank. We separated this into “yearCollected”, “monthCollected”, and “dayCollected” fields in FIMS so that each could be validated. However, Geneious *should* automatically combine these fields into one “Collection Date” field if your assemblies are annotated correctly.

**Collected by** This will become the “collected\_by” field in GenBank. It corresponds with the FIMS field “collectedBy”. It is required for the BWP Data Standard, but if it is unknown you can select *None*.

Map fields on your document to GenBank fields

Project Name: Test Project

Specimen Voucher: voucherID

Identified by: identifiedBy

Collected by: collectedBy

Molecule Type: Genomic DNA

Genetic Location: Genomic

Country: countryOrOcean

Sequence ID: Workflow Name

Collection Date: Collection Date

Organism: scientificName

Genetic Code: Standard

☐ Include Features/Annotations ☐ Also include annotations

☒ Add Gene & CDS features using fields: Choose...

☐ Include quality scores (if possible)

☒ Include extra fields: Choose...

☒ Include structured comments: Choose...

Standard: transl\_table: 1

Vertebrate Mitochondrial: transl\_table: 2

Yeast Mitochondrial: transl\_table: 3

Mold Protozoan Mitochondrial: transl\_table: 4

Invertebrate Mitochondrial: transl\_table: 5

Ciliate: transl\_table: 6

Echinoderm Mitochondrial: transl\_table: 9

Euplotid Nuclear: transl\_table: 10

Bacterial: transl\_table: 11

Alternative Yeast: transl\_table: 12

Ascidian Mitochondrial: transl\_table: 13

Flatworm Mitochondrial: transl\_table: 14

Blepharisma Nuclear: transl\_table: 15

Fig. 2: This screenshot shows the appropriate FIMS fields to select for each of the GenBank fields.

**Organism** This field corresponds with the “scientificName” field from FIMS. It will be checked against the NCBI taxonomy database, so make sure that it is already in the database, or be prepared to create a new entry in the database. The name should only be the binomial name (or trinomial if subspecies), and should not include the taxonomic name authority.

**Molecule Type** This will always be “Genomic DNA” for DNA Barcode records.

**Genetic Code** For COI barcode sequences, this will be either “Vertebrate Mitochondrial” or “Invertebrate Mitochondrial”. (*Make sure to separate vertebrates and invertebrate submissions, as you can only choose 1.*) Plant barcode sequences (matK and rbcL) will always be “Bacterial” (the full name that Geneious abbreviated is “The Bacterial, Archaeal, and **Plant** Plastid Code”).

**Genetic Location** For COI barcode sequences, this will be “Mitochondrion”. For plant barcode sequences (matK and rbcL), this will be “Chloroplast”.

### 1.17.5 Gene and CDS Features

The next step will be to let GenBank know which gene was sequenced. As you can see in the snippet from a sample GenBank record below, this will also provide enough information for Geneious to automatically generate the protein amino acid sequence as well.

```

gene      <1..>655
          /gene="COI"
CDS       <1..>655
          /gene="COI"
          /codon_start=2
          /transl_table=5
          /product="cytochrome oxidase subunit 1"
          /protein_id="Protein ID"
          /db_xref="GI:Nucleotide GI"
          /translation="TLYFIFGWAGMVGTSLSILIRMEISTPGSLIGDDQIYNTIVTA
          HAFIMIFFMVMPIMIGGFGNWLVPMLGAPDMAFPRLNNMSFWLLPSSLILLISSIV
          ENGAGTGWTVYPPLSSNIAHSGSSVDLAIFSLHLAGISSILGAINFITTIINMRVNNL
          SFDQMSLFIWAVGITALLLLSLPVLGAITMLLTDRNLNTSFFDPAGGGDP"

```

Since DNA barcodes are not full gene sequences, select “Partial” for both Gene Feature and CDS Feature.

Enter your own text or choose a field to use the associated value on your documents.

Gene Feature

Sequence: ☒ Partial ☐ Complete

Gene: COI

☒ Add CDS Feature

Sequence: ☒ Partial ☐ Complete

Product: cytochrome oxidase subunit 1

Translation frame will be automatically detected.

Reset to Defaults OK Cancel

The following table will show the corresponding Gene and CDS Product name for each DNA barcode region. You can copy and paste directly from here.

Gene	CDS Product
COI	cytochrome oxidase subunit 1
matK	maturase K
rbcL	ribulose-1,5-bisphosphate carboxylase/oxygenase large subunit

## Non-BARCODE region sequences

If you are creating submission files for sequences for protein-coding regions that are not part of the DNA Barcode Data Standard, you can still use the Gene and CDS Features. However, it is very important that you ensure that “Experimental Strategy” in the Traces tab is set to “TARGETED LOCUS”. This is reiterated in the “Traces and Sequencing Primers” section below.

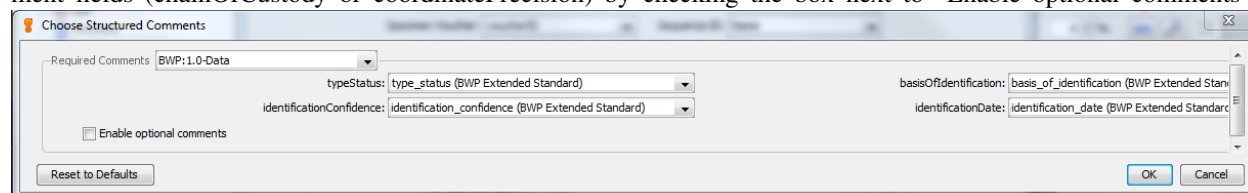
If you are creating submission files for sequences that are NOT protein-coding, follow the instructions laid out in the *Annotating for ITS GenBank Upload* special SOP.

Here are the Gene and corresponding CDS Product for common non-barcode regions. If you are unsure, look at existing sequence on GenBank.

Gene	CDS Product
CytB	cytochrome b

### 1.17.6 BWP Structured Comments

This window will appear when you check the button next to “Include structured comments”, and then click the “Choose...” button. These are the Barcode of Wildlife Project special fields that set apart BWP GenBank records from regular BARCODE keyword records. The screenshot below shows the appropriate FIMS field to choose for each structured comment field. You can also include additional non-required structured comment fields (chainOfCustody or coordinatePrecision) by checking the box next to “Enable optional comments”.



```
COMMENT      ##BWP:1.0-START##
              typeStatus          :: none
              basisOfIdentification :: Complete Adult
              identificationConfidence :: High
              identificationDate    :: 07-Aug-2014
              chainOfCustody        :: Yes
              ##BWP:1.0-END##
```

Fig. 3: This screenshot of a sample BWP GenBank record shows how the BWP structured comment fields will appear.

### 1.17.7 Consensus and Primers

Since we are submitting from an assembly of traces, we need to specify to Geneious how to calculate the sequence to submit to GenBank. Keep the default settings.

Consensus ☒ Traces ☒ Sequencing Primers ☒ Primers

Threshold:

☐ Ignore Gaps

☒ Assign Quality

If no coverage call 

☐ Call

☐ Split into separate sequences around '?' calls

PCR Primers are a required component of the Barcode Data Standard. You will need to tell Geneious which of your fields holds the PCR primer names, and PCR primer sequences. The appropriate fields should be populated automatically.

☒ Include structured comments: 

Consensus ☒ Traces ☒ Sequencing Primers ☒ Primers

☒ Include Primers

Forward PCR Primer Name  Reverse PCR Primer Name

Forward PCR Primer Sequence  Reverse PCR Primer Sequence

Forward PCR Primer Name (Sequencing Primer)  
Forward PCR Primer Sequence (Sequencing Primer)  
Forward Sequencing Primer Name (Sequencing Primer)  
Forward Sequencing Primer Sequence (Sequencing Primer)  
Reverse PCR Primer Name (Sequencing Primer)  
Reverse PCR Primer Sequence (Sequencing Primer)  
Reverse Sequencing Primer Name (Sequencing Primer)  
Reverse Sequencing Primer Sequence (Sequencing Primer)

### 1.17.8 Traces and Sequencing Primers

**Experimental Strategy** Choose “BARCODE” for this field if you are creating a submission for one of the official BARCODE gene regions (COI for animals, or rbcL and matK for plants). If you are submitting another region, then choose “TARGETED LOCUS”.

**Sequencing Strategy** Always choose “PCR”, even for non BARCODE gene regions.

**Center Project Name** Enter the name of the location where the traces were generated.

**Base Calling Program** If you generated your traces with an Applied Biosystems sequencer (your trace files will all end with “.ab1”), then enter “KB Basecaller”. This is the name of the software that is on all ABI sequencers that decides what each base in your trace files are.

**DNA Source Type** This will always be “Genomic DNA”.

**Trace End** Leave this as the default value of “*Let Geneious determine*”

Just like for the previous “PCR Primers” section, these fields should all be populated by Geneious automatically.



Consensus ☒ Traces ☒ Sequencing Primers ☒ Primers

☒ Include Traces

Experimental Strategy:  Sequencing Strategy:

Center Project Name:  Base Calling Program:

DNA Source Type:  Trace End:

☐ Include extra fields:

Fig. 4: If your version of the Geneious plug-in does not show this exact set of field options for “Traces”, then you need to upgrade to version 1.6.4.

Consensus ☒ Traces ☒ Sequencing Primers ☒ Primers

☒ Include Sequencing Primers

Forward Sequencing Primer Name:

Forward Sequencing Primer Sequence:

Reverse Sequencing Primer Name:

Reverse Sequencing Primer Sequence:

## 1.18 Annotating for ITS GenBank Upload

Since ITS1 and ITS2 are not genes and do not encode proteins, you cannot use the automated “Add Gene & CDS features” button. You are going to have to individually annotate each sequence.

Here are a couple of example ITS records that show what we will be trying to achieve: <http://www.ncbi.nlm.nih.gov/nuccore/JN253283> and <http://www.ncbi.nlm.nih.gov/nuccore/KT960182>

### 1.18.1 Adding the annotation

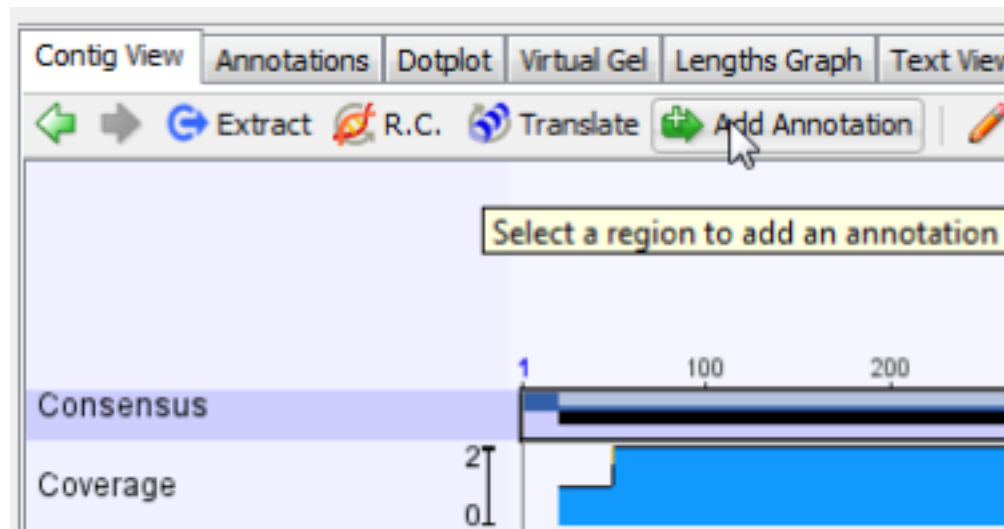
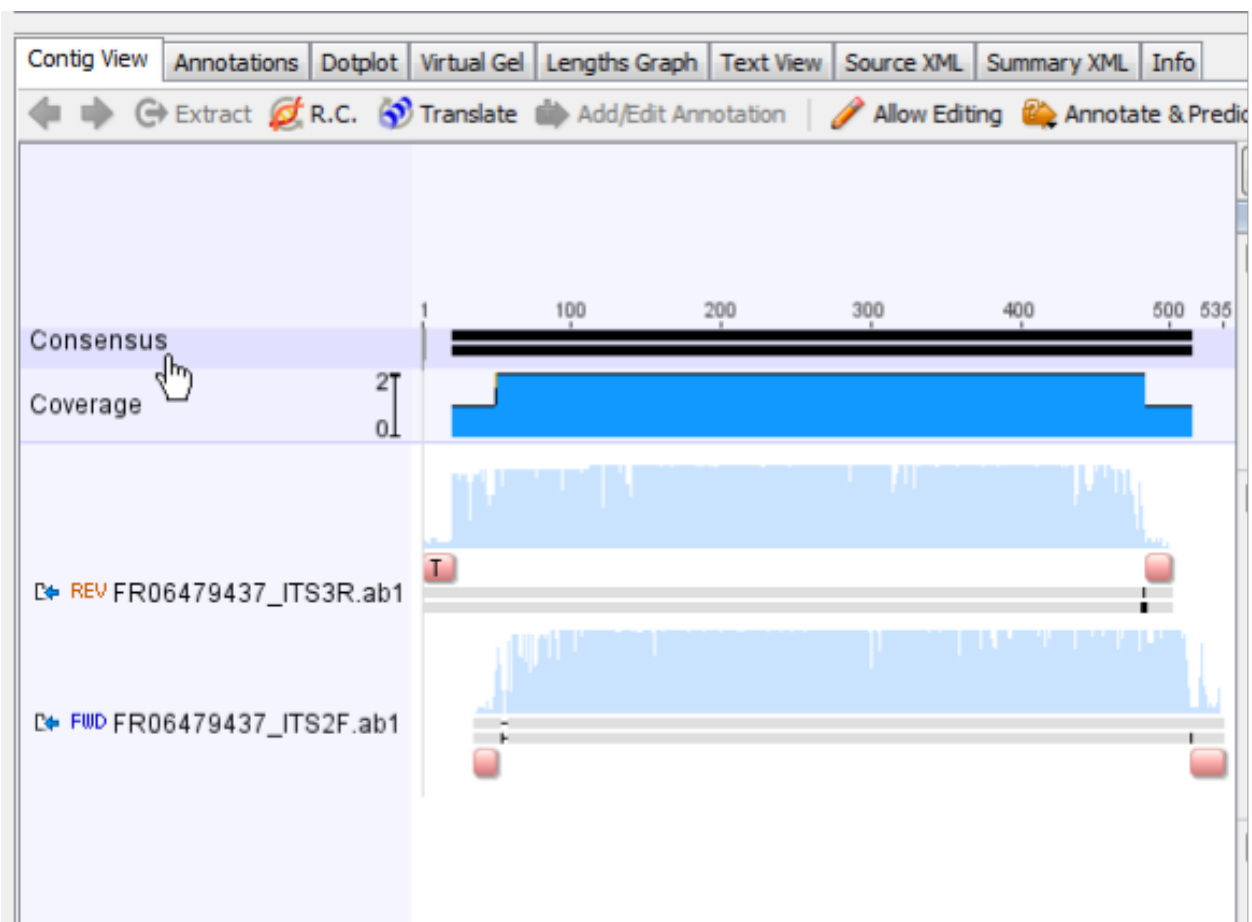
To annotate an assembly, make sure you are on the Contig View tab, and click select the consensus sequence by clicking the Consensus label.

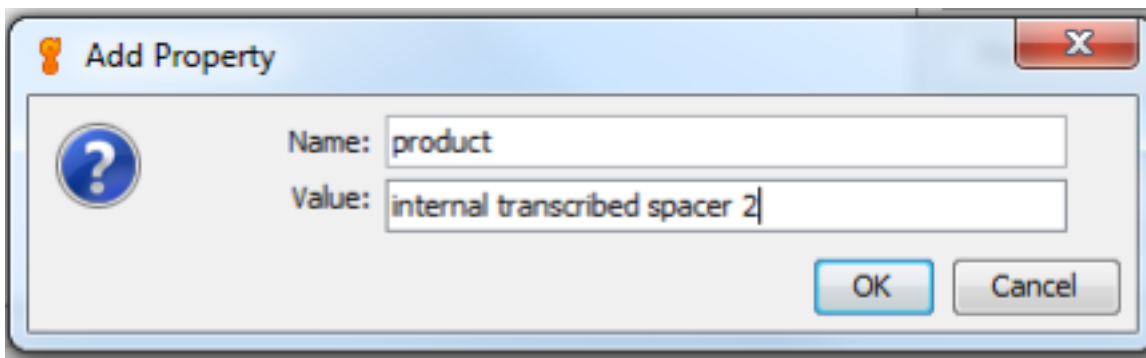
Once the consensus sequence is selected, click the “Add/Edit Annotation” button.

This will bring up the “Add annotation” window. In the Name box, enter any name that describes this annotation. For this example, I named it “ITS2 Annotation”. For Type, select “Misc\_RNA”. Then make sure that Track is “No Track”, and select “Undirected” for Direction.

In the Properties section, click the Add button to bring up the “Add Property” window. For “Name”, enter “product”, and for “Value” enter “internal transcribed spacer 2”. Click OK.







Finally, in the Intervals section, select the number range that is already there, and click “Edit”. Make sure that both the “Truncated left end” and “Truncated right end” options are selected. Click OK.

Your completed window should now look like this:

Click OK, and the ITS2 annotation will be added to the consensus portion of the assembly.

The annotation has been added to the consensus sequence, however it extends beyond the region of the trimmed sequence. We will need to shorten it on both ends. Zoom in on the sequence using the Magnifying glass “+” button, and scroll to the left end of the sequence.

Drag the end of the annotation so that it lines up with the end of the trimmed consensus sequence.

The left end of the annotation should now look like this:

Repeat the annotation trimming procedure on the right side.

Finally, click the Save button to save this Annotation.

### 1.18.2 Submitting annotated sequences to GenBank

Make sure to follow this manual annotation procedure for all of the ITS sequences you will be submitting to GenBank.

To submit to GenBank, follow the same procedure detailed here: [http://bwp-informatics.readthedocs.org/en/latest/gb\\_upload.html](http://bwp-informatics.readthedocs.org/en/latest/gb_upload.html), with the following changes:

- In the “Field Mapping” section, Genetic Location should be “Genomic”
- Rather than selecting “Add Gene & CDS features using fields”, select the option above it: “Include Features/Annotations”. This will ensure the ITS Annotation you just added will be submitted with the sequence. Make sure that “Also include annotations in tracks” is NOT selected.
- In the Traces tab, select “TARGETED LOCUS” for “Experimental Strategy”. All other selections will stay the same.

## 1.19 LIMS Connection Troubleshooting

The most common error report we have been receiving with the Barcode of Wildlife Project informatics is the inability to log in with the Biocode plug-in in Geneious.

Specifically, many users encounter issues with connecting to the BWP Remote LIMS database. Here is an example:

The most likely cause for this error is that a firewall is blocking you from connecting with the LIMS database.

---

**Note: Technical Explanation**

**Add annotation**

Name:

Type:

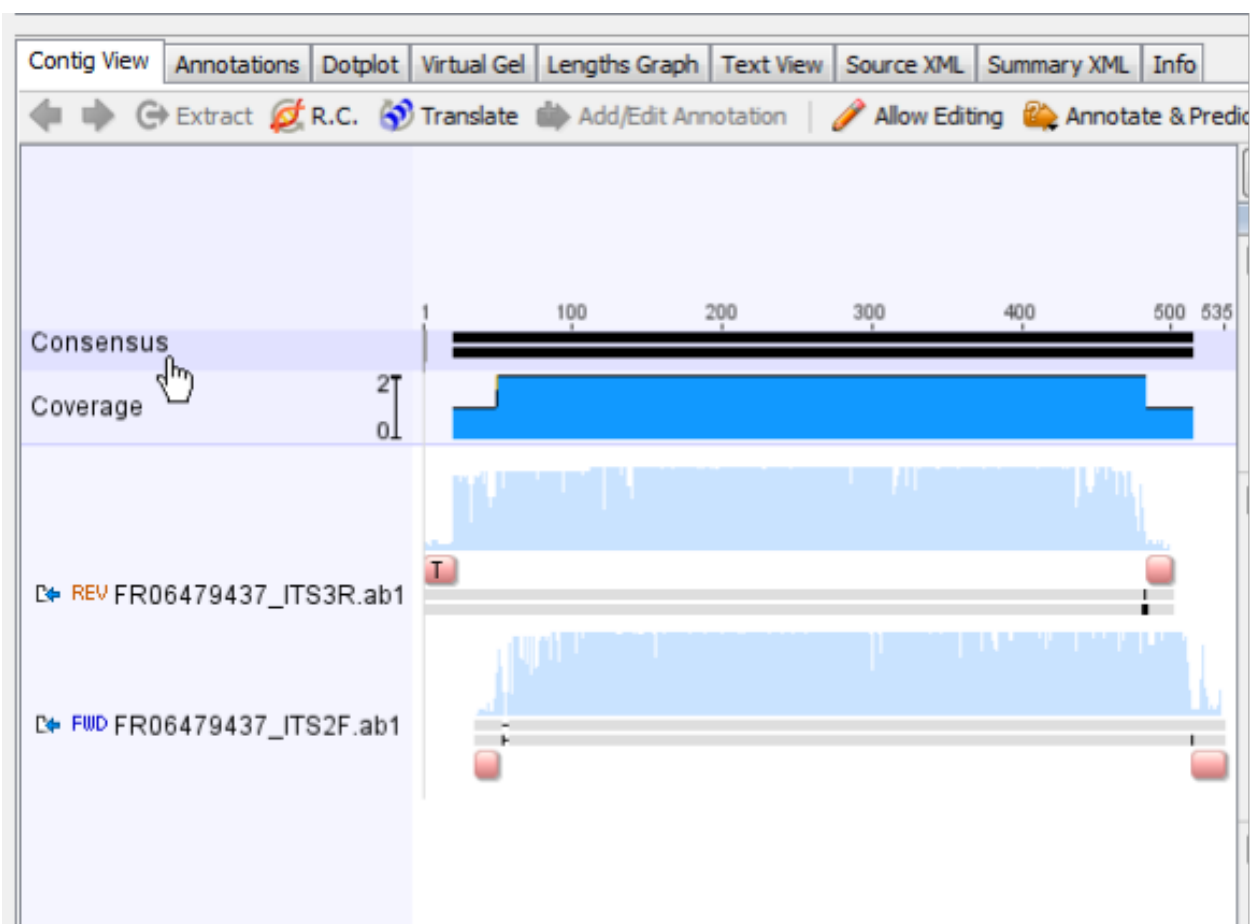
Track:

Direction: ☐ Forward ☐ Reverse ☒ Undirected

▼ Properties

Type Color: Override Color:

▼ Intervals





Without getting too technical, your computer uses different “TCP” port numbers to connect to other servers. There are 2 specific port numbers used to read web pages on the World Wide Web (Port 80 for sites starting with “http,” and Port 443 for sites starting with “https”), and several for sending and receiving e-mail. Geneious uses TCP port number 3306 to connect with the database, and a lot of the time this port will be blocked by an institutional firewall.

### 1.19.1 How to check if Port 3306 is open

We will be using a program called “Telnet client” to check if your computer can use Port 3306 to connect with the BWP LIMS.

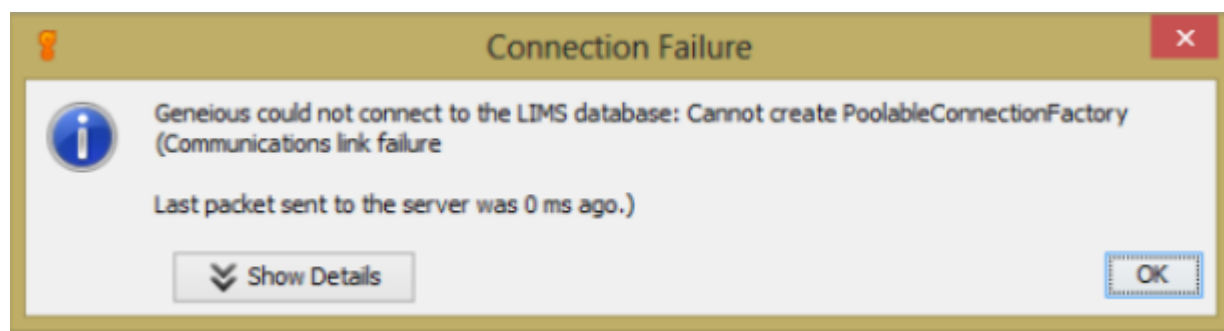
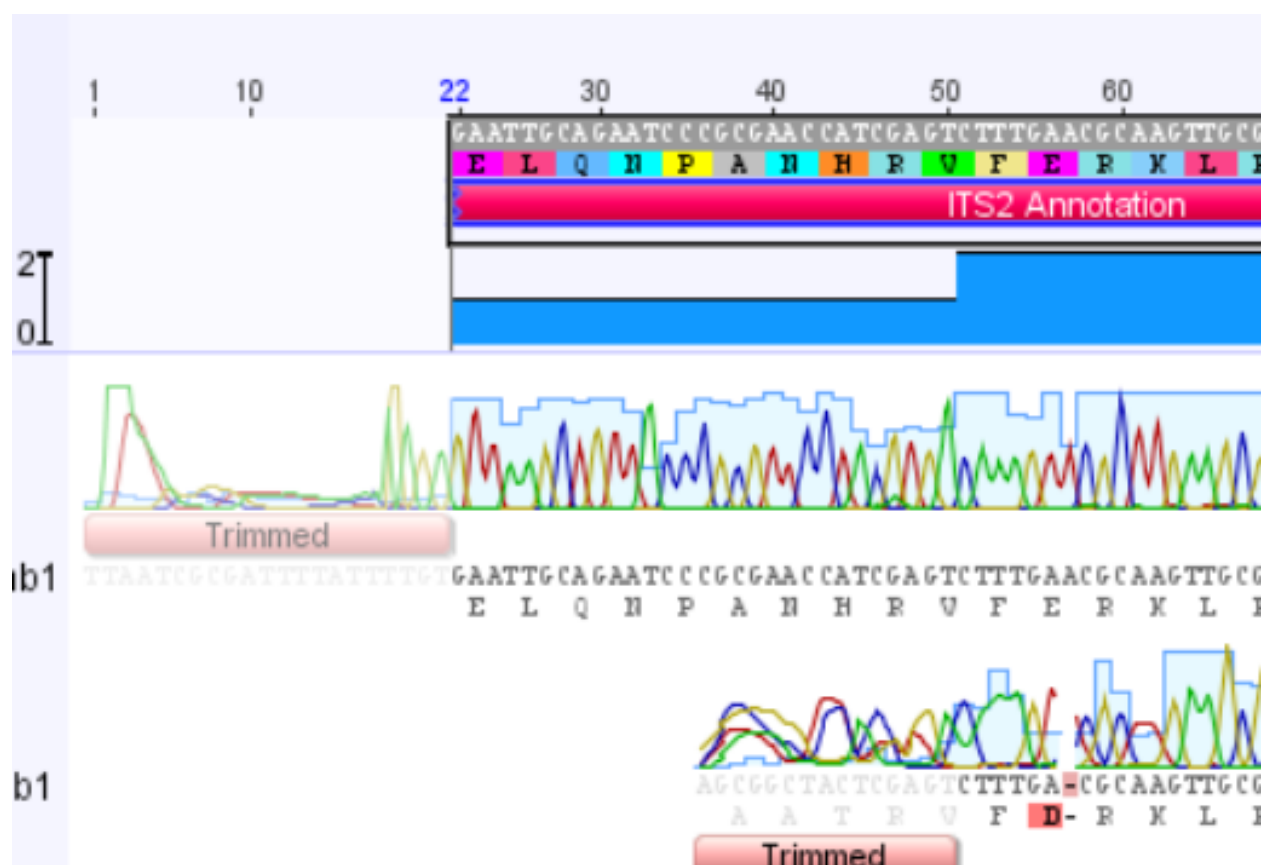
#### Install Telnet

If you are using a Mac or Linux computer, then the Telnet client will already be installed on your computer. If you are on a Windows computer, then Telnet client will likely be “turned off” by default. Follow the instructions here: <http://social.technet.microsoft.com/wiki/contents/articles/910.windows-7-enabling-telnet-client.aspx> to enable Telnet client.

#### Open a terminal window

*Windows*

- Go to “All Programs” > “Accessories” > “Command Prompt”



*Mac*

- Open “terminal” from your Apps folder (or press Apple-Space) and type “terminal” to launch the terminal program.

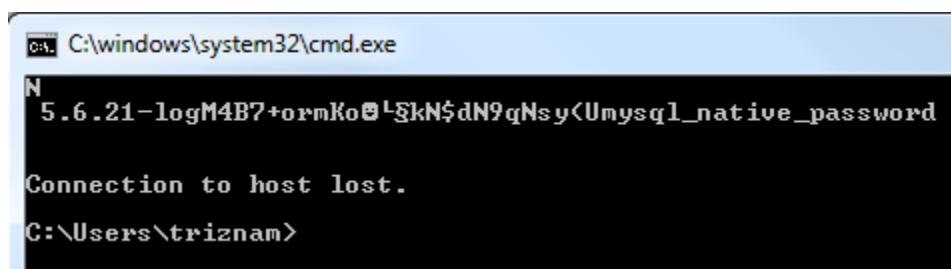
**Use telnet to check if BWP LIMS port 3306 is accessible**

- In the terminal, copy the following line, paste it into your terminal.
- Press Enter.

Two possible things will now happen.

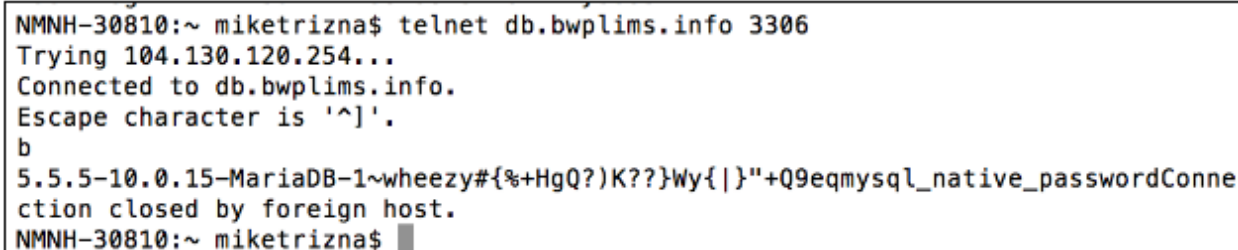
**Option 1**

The first option is that the terminal will clear, and you will get a cryptic message mentioning “mysql\_native\_password,” and then after several seconds it will tell you “Connection to host lost”.



```
C:\windows\system32\cmd.exe
N
5.6.21-logM4B7+ormKo@L$KNSdN9qNsyUmysql_native_password
Connection to host lost.
C:\Users\triznam>
```

Fig. 5: Viewed on Windows



```
NMNH-30810:~ miketrizna$ telnet db.bwplims.info 3306
Trying 104.130.120.254...
Connected to db.bwplims.info.
Escape character is '^]'.
b
5.5.5-10.0.15-MariaDB-1~wheezy#{%+HgQ?)K??}Wy{|}"+Q9eqmysql_native_passwordConne
ction closed by foreign host.
NMNH-30810:~ miketrizna$
```

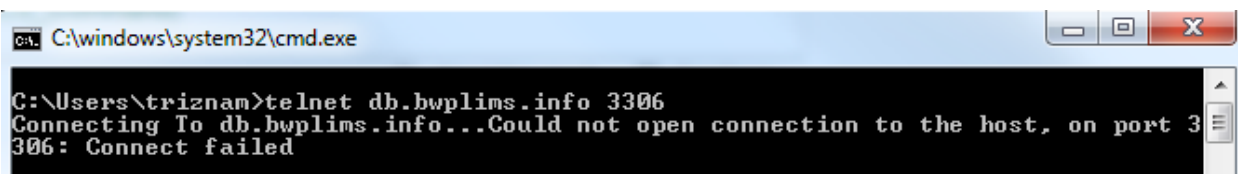
Fig. 6: Viewed on Mac

If this happens, then the most likely reason you are not able to log in to Geneious Biocode is because your password is incorrect. Please go back to your country-specific log-in information document, and check that you are using the correction information. E-mail Mike Trizna at [triznam@si.edu](mailto:triznam@si.edu) immediately if you are still having issues.

**Option 2**

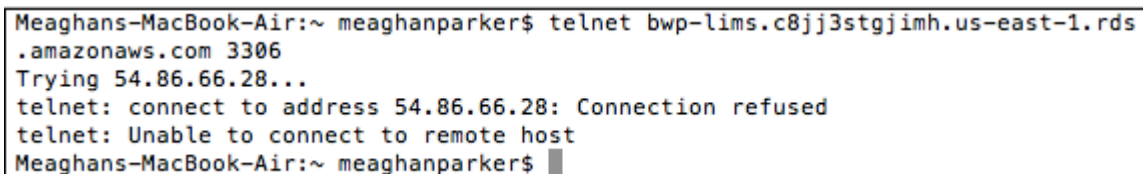
The second option is that telnet will report that the connection failed.

If this happens, then port 3306 is not accessible to you, and you will need to arrange with your IT department to have it opened. I put together the following e-mail below. Send it to the IT department at your institution and make sure to “cc” Mike Trizna ([triznam@si.edu](mailto:triznam@si.edu)).



```
C:\windows\system32\cmd.exe
C:\Users\triznam>telnet db.bwplims.info 3306
Connecting To db.bwplims.info...Could not open connection to the host, on port 3306: Connect failed
```

Fig. 7: Viewed on Windows



```
Meaghans-MacBook-Air:~ meaghanparker$ telnet bwplims.c8jj3stgjmh.us-east-1.rds.amazonaws.com 3306
Trying 54.86.66.28...
telnet: connect to address 54.86.66.28: Connection refused
telnet: Unable to connect to remote host
Meaghans-MacBook-Air:~ meaghanparker$
```

Fig. 8: Viewed on Mac

Hello,

We are experiencing difficulties connecting with a remote MySQL database used for collaboration on the Barcode of Wildlife Project.

The project's informatics lead, Mike Trizna, says that this is probably being caused by the database server's MySQL port (3306) not being accessible.

Here are all of the connection details:

Server Address: db.bwplims.info

Port: 3306

Username: connection\_test

Password: DNABarcodes

Please contact Mike Trizna at [triznam@si.edu](mailto:triznam@si.edu) if you have any questions about the database, and he will be happy to help you.

## 1.20 Migrating the BWP LIMS Database

CBOL has generously supported hosting the Laboratory Information Management System (LIMS) database since the beginning of the Barcode of Wildlife Project, but can no longer maintain this arrangement. Here is a guide to exporting your country's data and setting it up locally.

### 1.20.1 About the LIMS database

In the overall workflow of the Barcode of Wildlife Project, the LIMS database stores all of the lab-specific data associated with barcode sequencing of samples. A typical lab workflow starts with DNA extraction, then amplification of targeted DNA regions via PCR, followed by sequencing. The LIMS database also stores final version of sequences submitted to GenBank through this platform. Users document the progress of this workflow in Geneious using the specially-designed Biocode plugin, but all supporting data is stored in the LIMS database.

The LIMS database itself is a relational MySQL database designed by the developers of Geneious for this purpose. It is made up of 20 tables, connected together by several foreign keys. Here is a diagram of the tables in the database:



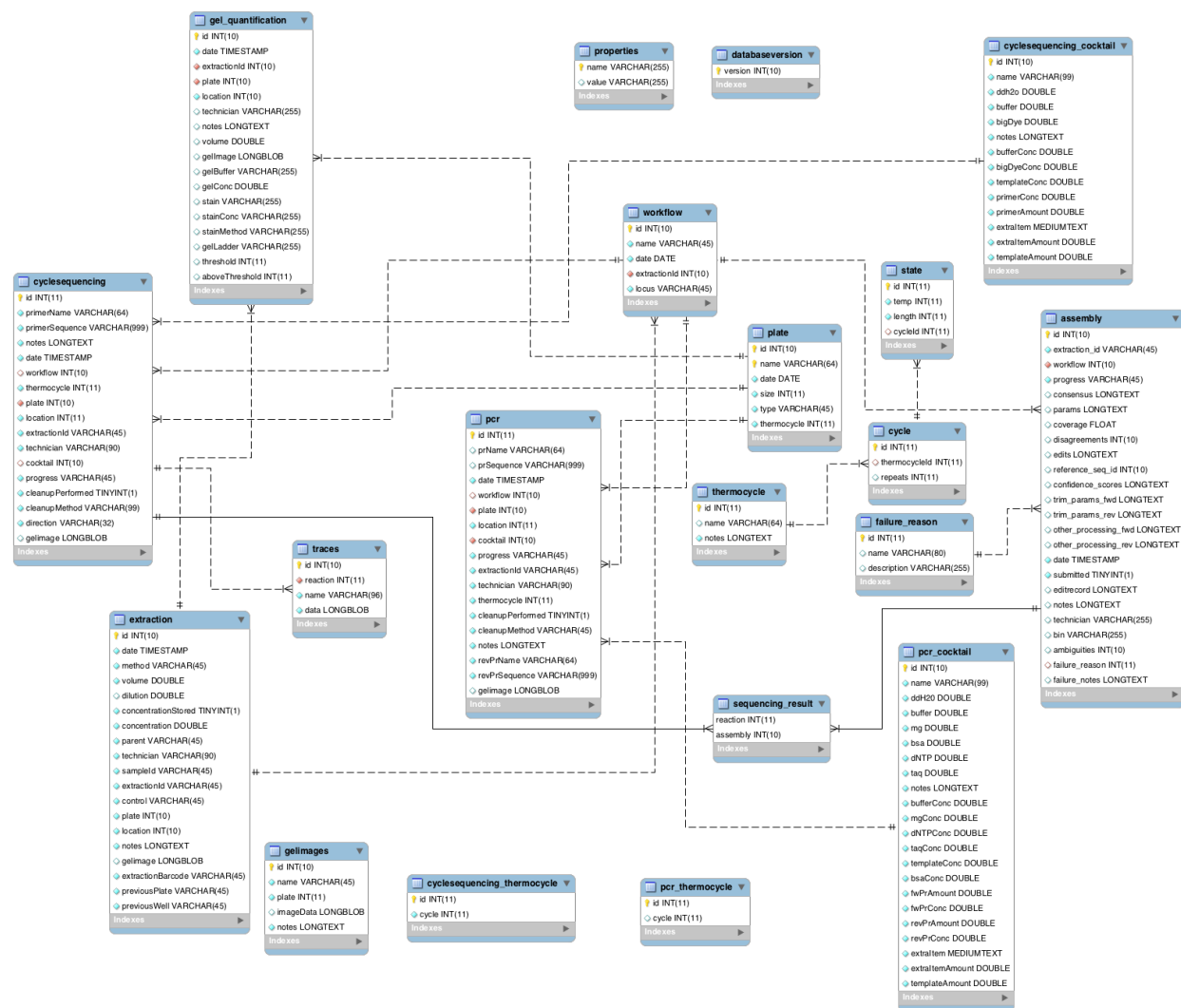


Fig. 9: This is an Entity-Relationship diagram of the BWP LIMS database. You can click on it to see the full-size image.

## 1.20.2 Dumping and importing the database

The same user account that was set up for troubleshooting connections to the database can be used to “dump” the contents for migration.

Use the tool “mysqldump”, which should be installed if you have MySQL installed on your computer. The command is:

```
mysqldump -h db.bwplims.info -u connection_test -p [DATABASE NAME] > lims_dump.sql
```

The country-specific database names are “kenya\_lims”, “mexico\_lims”, and “south\_africa\_lims”. You will then be prompted for a password, which is “DNABarcodes”.

I have prepared database dumps of each database on January 19, 2018 to profile the amount of space required, but it is a good idea to run the mysqldump command immediately prior to importing so that you get the “freshest” data.

Database dump filename	Uncompressed file size	Compressed file size	# of plates	# of work-flows
<a href="#">kenya_lims_2018_01_19.sql.gz</a>	1.28 GB	433.7 MB	442	3284
<a href="#">mexico_lims_2018_01_19.sql.gz</a>	2.69 GB	1.03 GB	429	11323
<a href="#">south_africa_2018_01_19.sql.gz</a>	1.1 GB	384.8 MB	755	2402

To import this SQL dump file into your own database, first set up an empty database in MySQL using the “CREATE DATABASE [DATABASE NAME]” command. It would probably be easier to keep the name the same, but you can change it to whatever you like.

After the database is created, you can import the SQL file with the following command:

```
mysql -u [USERNAME] -p [DATABASE NAME] < lims_dump.sql
```

Again, you will be prompted for a password, but this time it will be for the USERNAME that you have set up on your local MySQL instance.

## 1.20.3 User management

Unfortunately the Biocode plug-in and LIMS database system are configured in a way such that all users are able to log in and modify ANY data contained in the database. You can create user accounts for each user, but we found that it’s easier to just create 1 account for each country.

To create a user in the LIMS database with the correct permissions needed to connect and use the BioCode plugin, use the following command after logging into the database as root:

```
GRANT SELECT, UPDATE, INSERT, DELETE
ON [DATABASE NAME].*
TO '[USERNAME]'@'%'
IDENTIFIED BY '[PASSWORD]';
```