

Chapter 13

Laboratory Information Management Systems for DNA Barcoding

Meaghan Parker, Steven Stones-Havas, Craig Starger, and Christopher Meyer

Abstract

In the field of molecular biology, laboratory information management systems (LIMSs) have been created to track workflows through a process pipeline. For the purposes of DNA barcoding, this workflow involves tracking tissues through extraction, PCR, cycle sequencing, and consensus assembly. Importantly, a LIMS that serves the DNA barcoding community must link required elements for public submissions (e.g., primers, trace files) that are generated in the molecular lab with specimen metadata. Here, we demonstrate an example workflow of a specimen's entry into the LIMS database to the publishing of the specimen's genetic data to a public database using Geneious bioinformatics software. Throughout the process, the connections between steps in the workflow are maintained to facilitate post-processing annotation, structured reporting, and fully transparent edits to reduce subjectivity and increase repeatability.

Key words: Laboratory information management systems, Workflow, DNA barcoding, Submission, Bioinformatics

1. Introduction

Only recently have powerful, customizable laboratory information management systems (LIMSs) software programs been built and published for public purchase and use (www.limsources.com). Previously, when researchers had a need for a comprehensive method of databasing and tracking specimens, they either had to build a program by hand or contract a programmer to customize a system that would meet their specific needs. Though many of these LIMS programs were unique in one way or another, they were inherently all built to do the same thing: track specimens and workflows. What ensued from the lack of low-cost, customizable LIMS

programs available on the bioinformatics market was an exorbitant amount of energy used by researchers and programmers to build many, slightly different, LIMS software programs.

While the concept behind all LIMS software programs is inherently the same, LIMS programs do need to be customizable because of the individual needs of the scientist. There are now many marketed LIMS software programs that allow the users to customize the program to fit their needs without having to script an entire new program. Further, many LIMS software developers have incorporated into their LIMS program the ability not only to track individual specimens, but also to capture all laboratory workflows associated with each specimen, creating a detailed, transparent electronic notebook for the users and collaborators.

The Moorea Biocode Project (MBP—<http://www.mooreabiocode.org/>) was awarded funds for 3 years from the Gordon and Betty Moore Foundation to genetically index all plant, animal, and fungal macrobiotic species living on the island of Moorea, French Polynesia. It was understood by the project's scientists that the number of specimens to be collected during the 3-year period would be monumental, and that the need for a LIMS software program would be essential. The MBP contracted Biomatters Ltd., a New Zealand bioinformatics company, to cooperatively build a LIMS software program that would track all molecular processes conducted with each individual specimen and link to field specimen metadata. This electronic notebook has allowed project scientists and collaborators from all over the world to view, edit, and share data.

Scientists working on the MBP have, to date, genetically databased over 25,000 specimens using the Geneious LIMS software program. Since 2008, Biomatters Ltd. has worked with the MBP to create a LIMS software program that not only meets all of Biocode's needs, but also the needs of any researcher who, with either large or small datasets, desires the ability to database and track all genetic samples with confidence. There are now many high-quality LIMS software programs available on the bioinformatics market for the single and group users. The Biocode LIMS is intended to fill a niche for people wishing to perform DNA barcoding, but without the capacity to develop or purchase their own LIMS. Moreover, it is adaptable to any similar Sanger-based molecular workflows.

In this chapter, we demonstrate an example workflow of a specimen's entry into the LIMS database to the publishing of the specimen's genetic data to Genbank using Geneious's bioinformatics software and the Biocode PlugIn.

2. Materials

2.1. Downloading and Installing Geneious Software

You need to download Geneious for Windows, Mac, or Linux, available from the Biomatters Web site <http://www.geneious.com/>. To run the Biocode software, you need Geneious version 5.1 or higher. Three files, in addition to Geneious, need to be downloaded in order to use the Biocode plugin: the Biocode plugin itself, the Biocode Genbank Submission plugin (which allows you to submit completed contigs to Genbank), and the MySQL Connector (which allows you to connect to FIMS and LIMS databases). All of these downloads can be retrieved from <http://software.mooreabiocode.org/>.

To install the plugins, open Geneious and drag the plugin onto the Geneious interface. You will see three new icons—two in the toolbar and one in the service tree (Fig. 1).

2.2. Setting Up Databases

Once the Biocode plugin has been installed, the next step is to set up the two required databases, field information management system (FIMS) and LIMS. The FIMS database stores information relating to “field” work (e.g., specimen records, collecting events, and taxonomy). The LIMS stores data relating to lab work (e.g., workflows). The Biocode plugin includes a LIMS that links to data from a FIMS database of your choosing. Since all molecular processes start with a tissue sample, we assume that this is the entry point to LIMS and the tissue’s unique ID can be used to track the appropriate specimen metadata from the FIMS.

Connect to the Biocode plugin by right clicking on the Biocode icon in the service tree. The Biocode login screen will open (Fig. 2). The login screen has two sections: one for FIMS and one for LIMS.

At the bottom of the screen, you need to specify a MySQL Driver (the MySQL connector file that you downloaded with the Biocode plugin file). Click the Browse button and locate the file on your disk.

Your FIMS database can be set up in one of the following four ways: an EXCEL file (see Note 1), a server such as TAPIR (see Note 2), a Google Fusion Table (see Note 3), or a MySQL (see Note 4).

- The EXCEL file is most useful for single users or people who cannot set up a server.

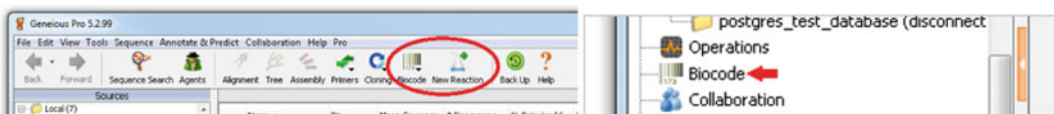


Fig. 1. Biocode plugin icons along the main toolbar (*left*) and Biocode plugin/login icon in service tree (*right*).

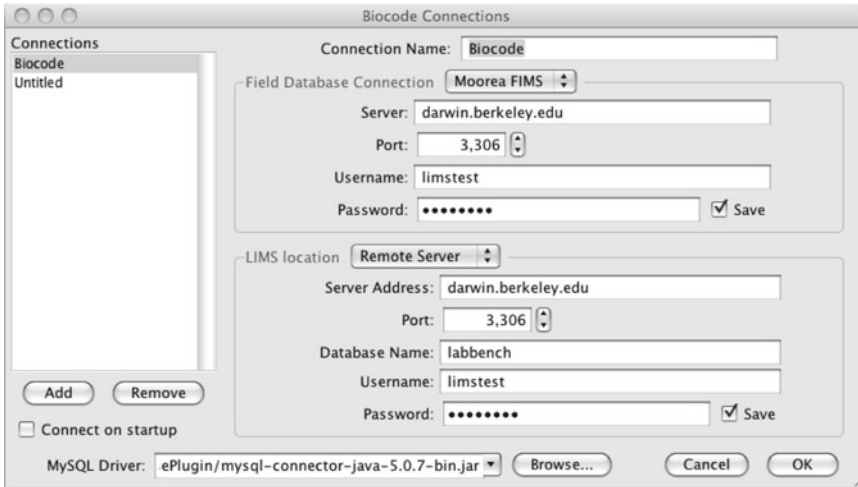


Fig. 2. The Biocode plug-in login screen.

- TAPIR, MySQL, and Google Fusion Tables are most useful for groups who want to store their lab information in a single FIMS or for collaborative projects, where the data are stored on a central server.

You also have the choice to set up your LIMS database in one of the following two ways: using a Remote LIMS (see Note 5) or a Local LIMS (see Note 6).

- Using the Remote LIMS, the plugin stores your lab workflows on a remote MySQL server. This is the best option for labs or collaborative projects.
- Using the Local LIMS, the plugin stores your lab workflows on your local machine. This is intended for single users.

3. Laboratory Information Management System Workflows

Workflows (see Note 7) are an integral part of the LIMS system (Fig. 3). Workflows represent the laboratory path that a tissue extraction takes for a particular locus. We define locus as an expected amplicon region that may contain multiple genes. A workflow is limited to a single extraction, but can have any number of other reactions. Workflows can be queried for tracking progress, troubleshooting, and reports. Information from workflows is also taken by the assembler module to aid in the assembly and analysis of your sequences, so it is important that they are set up correctly. Figure 3 illustrates an example workflow.

We begin with an example set of 96 reactions and demonstrate how these samples move through the LIMS.

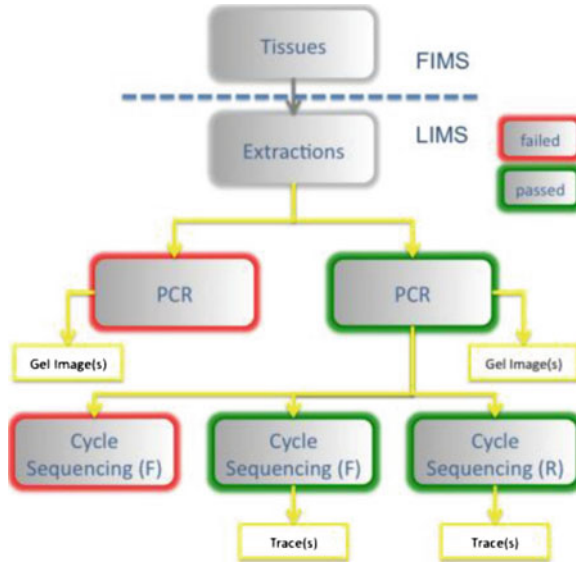


Fig. 3. Laboratory information management systems workflow (see also Note 7). Yellow connectors represent potential paths in a workflow.

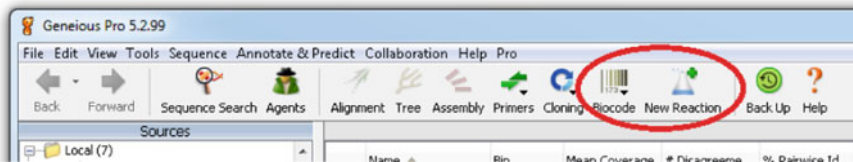


Fig. 4. Geneious main toolbar with Biocode plugin.

3.1. The LabBench Module: Generating Plates and Associated Workflows

3.1.1. Generating the Extraction Plate

Following the LIMS workflow, the first plate generated is the extraction plate. Click the “New Reaction” button on the Geneious main toolbar (Fig. 4). In the drop-down menu of the “New Reaction” window (Fig. 5), you can choose the type of reaction (extraction, PCR, or cycle sequencing) and plate size (one of the three predefined sizes, or a number of individual reactions).

A new window will open displaying a blank extraction plate (or the specific number of samples you have chosen) (Fig. 6). Your FIMS data must be uploaded to the blank extraction plate so that each specimen’s field data are correctly associated with that same specimen’s laboratory workflow. This can be done using the “Bulk Edit” button in the toolbar of the new extraction plate (Fig. 6).

When using the bulk-editor to generate an extraction plate, a number of tools are available in the “Tools” menu within the “Bulk Edit” window (Fig. 7).

- “Get Tissue ID’s from archive plate” allows you to fill your plate with extraction IDs from the FIMS database (either from extraction or tissue archive plates). If you are using an EXCEL,

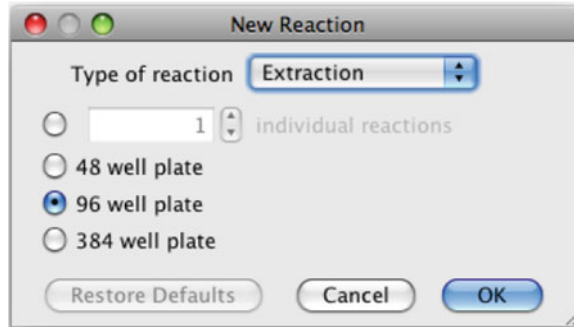


Fig. 5. New reaction window.

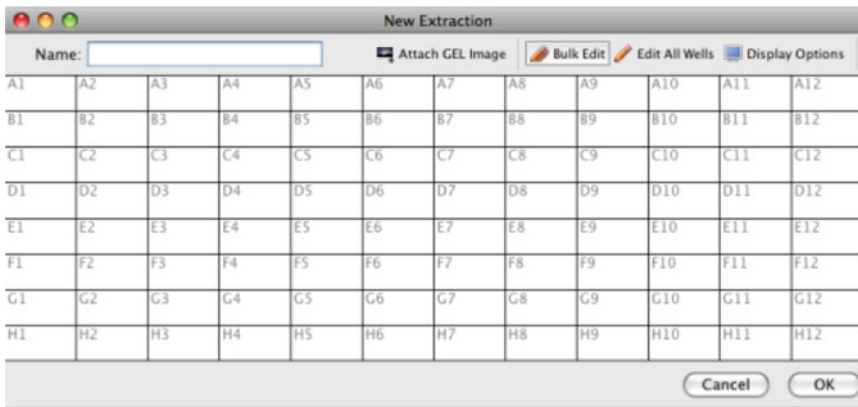


Fig. 6. Blank Plate document.

MySQL, or Google Fusion Table FIMS database, you need to have plate and well columns in your spreadsheet, and to set those in the login screen (see Notes 1–4).

- “Import Extraction Barcodes” allows you to import barcode values directly from the output file of the scanner if you are using 2D well barcodes.
- “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.

The well locations are displayed on the left-hand side of each column to make placement easier (Fig. 7). You do not have to enter anything in the workflow ID column. New workflows will be generated when you save the plate. “Swap Direction” allows the user to choose between reading across and down versus down and across along the plate.

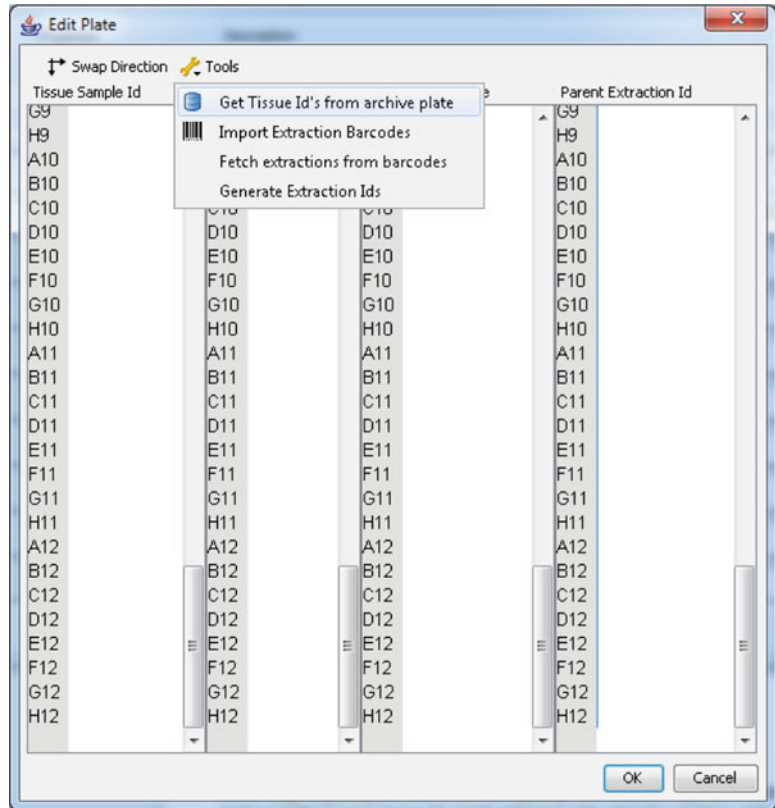


Fig. 7. Bulk-editor window displaying “Tools” drop-down window.

Once the field data have been added to the extraction plate, you can then edit each well. The “Edit All Wells” (also a button in the toolbar of the new plate) is a customizable viewer and editor for your plate documents. It is shown both when creating new plates and when viewing existing plates in the database (Fig. 6). You can select wells in the plate by dragging the mouse across the plate, or select a single well by clicking it. You can hold down the shift and ctrl (command on mac) keys to help you select multiple individual wells. When wells have been selected, click “Edit Selected Wells” to customize those wells. The edit dialog (Fig. 8) has a column of checkboxes on its left-hand side. Values in the checked fields are applied to all selected reactions, and unchecked fields are 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 plate editor can display any number of fields from your FIMS or LIMS database. Click the Display Options button in the toolbar to open the display dialog. The split pane at the center of the dialog allows you to choose the fields to be displayed on your wells (Fig. 9). The available fields are shown in the left-hand pane.

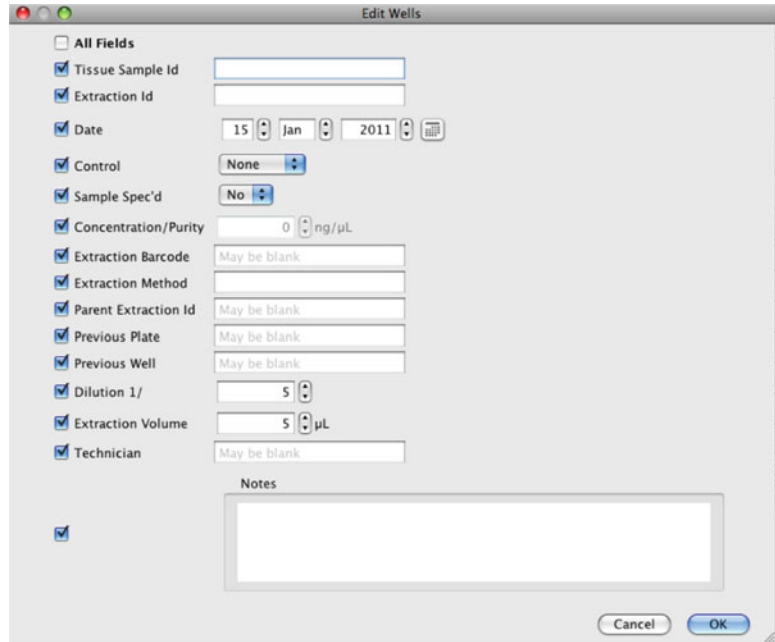


Fig. 8. Edit dialog window within extraction plate.

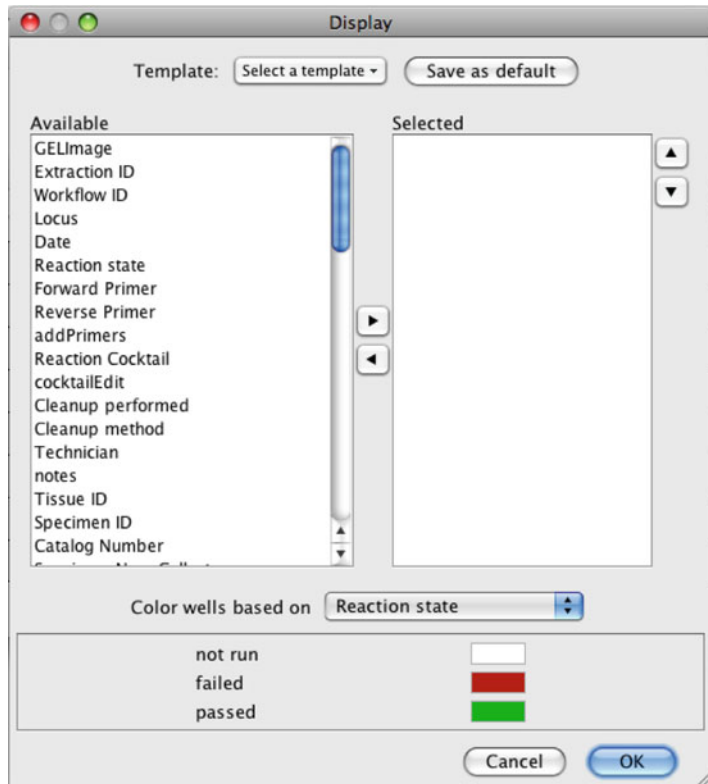


Fig. 9. Display option panel for viewing plates.

Select the fields you want to display, and click the right arrow to move them to the right-hand pane. Select fields in the right-hand pane, click the left arrow to return them to the right-hand panel, and stop them from being displayed in the wells. Once you have chosen your fields to display, you can choose their order using the up/down arrows on the right-hand side of the dialog. The fields will appear in the wells in the order you choose here.

The bottom part of the dialog controls the well coloring. Wells are colored by one field, with each different value for that field being given a different color. Choose the field you are interested in, and all possible values will be displayed below it. You can change the color of any value by clicking it.

You can save any settings you make here as a template by clicking the “Select a template” button at the top of the dialog, and clicking “Create template.” Click the “Save as Default” button to make that template the default (separate defaults are stored for extraction, PCR, and sequencing plates).

Display preferences are not saved with the plate; so in order to preserve the view for your plate, you should save your settings as a template and/or set them to the default for the reaction type.

3.1.2. *Generating a PCR 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, highlight the existing plate (in the plate viewer window, Fig. 10) and click “New Reaction” in the Geneious main toolbar. Select the “Create plate from existing document” checkbox (Fig. 11). If the reaction types are the same (e.g., 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 (e.g., an extraction to a PCR plate), then only the extraction IDs will be copied across. It is possible to create a new 384-well plate from a group of 96-well plates, and to create a group of 96-well plates from a 384-well plate (see Note 8).

When the new PCR plate has been generated (either from an existing document or a new document), it can then be edited using the same commands used to edit the extraction plate. You can select the target locus for each well individually or for all wells at once in the “Edit Wells” dialog box. Select the wells to be edited and click “Edit Selected Wells” in the toolbar. When the “Edit Well” window from the new PCR plate opens (Fig. 12), the edit dialog boxes, again, have a column of checkboxes on the left-hand side. Values in the checked fields are applied to all selected reactions, and unchecked fields are left as they are. You may mark a PCR as not run, run, passed, or failed. Additionally, options to customize your

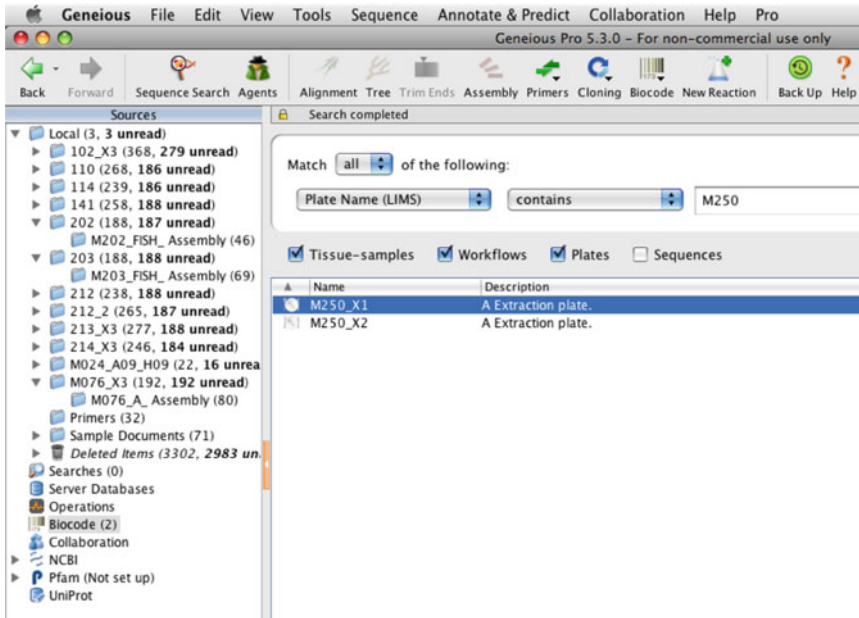


Fig. 10. Plate viewer window.

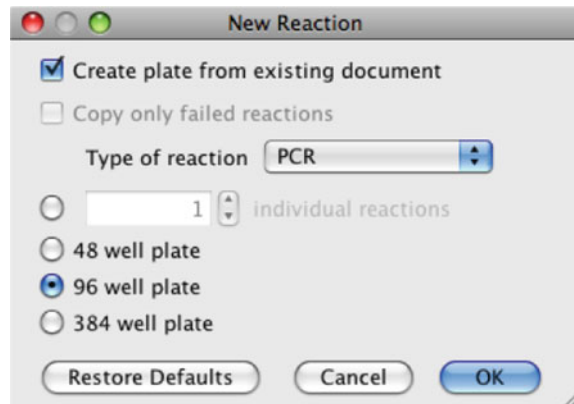


Fig. 11. Generate plate from the existing document checkbox.

PCR cocktails (see Note 9) and thermocycler profiles (see Note 10) are available. You can create your own local primer database or select primers from a global database (see Note 11).

3.1.3. Uploading Gel Images

You can attach GEL images to all types of plates. Click the “Add/Remove GEL Image” button in the plate editor toolbar (Fig. 6), click “Add,” and then browse to the image file on your hard disk. You can also add notes to each GEL image. Geneious accepts images in JPEG, GIF, PNG, and TIFF formats. You can split a GEL image into wells using a draggable grid by clicking the Split GEL button (located above the GEL image in the gel viewer window) (Fig. 13).

Fig. 12. PCR “Edit Wells” window.

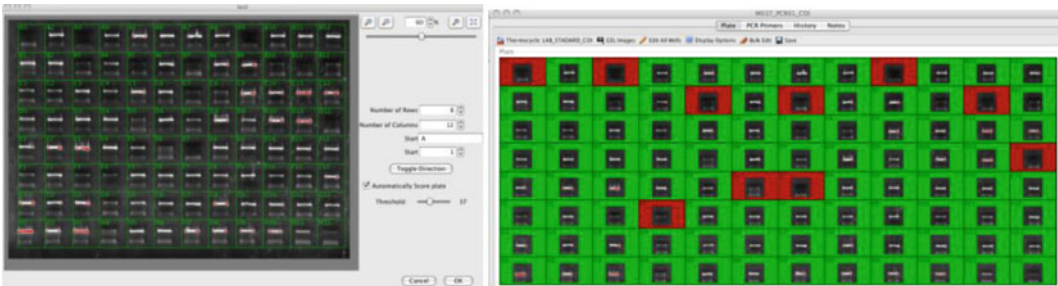


Fig. 13. Splice tool for scoring PCRs via gel image.

Choose the number of rows and columns, the start row and column, and the direction of the wells in the grid, and then drag your mouse on the image to generate the grid. If you misplace the grid, you can start again by dragging the mouse. Click “ok” when done. Automated calling of pass/fail is possible, but should be verified through various display options of the PCR plate (Fig. 13).

3.1.4. Generating

Cycle Sequencing Plates

To append new cycle sequencing plates onto existing workflows, highlight the existing PCR plate (in the plate viewer window, Fig. 10) and click “New Reaction,” followed by checking the box labeled “Create plate from existing document” (Fig. 11). A plate very similar to the PCR plate will be generated and you will go

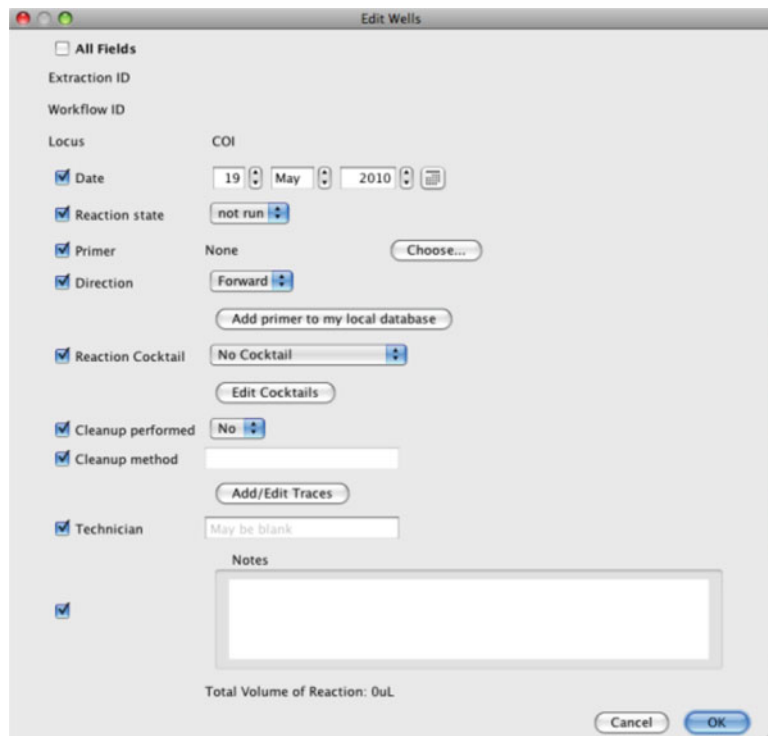


Fig. 14. Cycle sequencing “Edit Wells” window.

about editing the plate in the same manner you edited both extraction and PCR plates; all associated metadata will carry through.

The “Edit Wells” (Fig. 14) window should now begin to look familiar, and here, again, you can customize the dialog boxes as well as add new cycle sequencing cocktails, cycle sequencing thermocycler profiles, and primers (see Notes 9–11).

3.1.5. Adding Traces

Once the plate or samples have been sequenced, you can add the traces to your cycle sequencing plates or reactions. To add traces to individual reactions in the plate view, open the reaction well in which you would like to add a trace file to (double click on the specific well), and click “Add/Edit Traces” in the “Edit Wells” dialog (Fig. 14). You can then click the “Add Sequences” button in the upper left corner to add one or more sequences to the well. To remove one or more sequences from the well, select the sequences you want to remove and click “Remove Sequence(s).” You can also import the sequences into Geneious as documents by clicking Import Sequence(s) into Geneious.

You also have the option to bulk upload traces onto the cycle sequencing reactions or plate. To bulk upload traces to the cycle sequencing plate or reactions, open the appropriate cycle sequencing plate. Click “Bulk Add Traces” on the plate’s toolbar and click

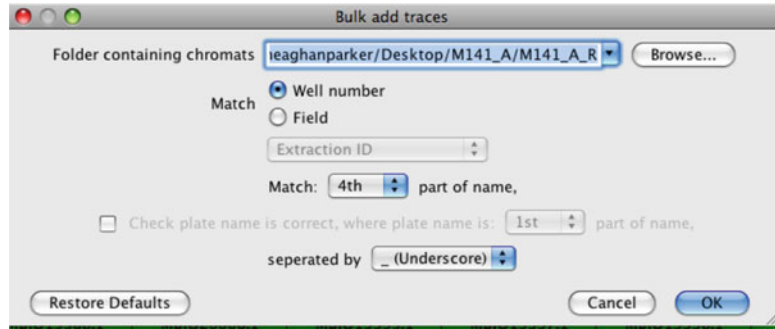


Fig. 15. Bulk add traces from cycle sequencing plate toolbar.

“Browse” to locate the trace files on your hard drive. Traces from each cycle sequencing plate should be in a separate folder. The traces are matched to the appropriate well by field or well number, which are found in the filename of the trace file. You need to tell Geneious where in the filename to look for either “Well number” or “Field.” For example, if you want Geneious to link the sample’s trace file to that sample’s well position on the cycle sequencing plate and you had named that trace file “3726294_A01_capture.ab1”, you would select “Match 2nd” (from drop-down menu) part of name and “separated by Underscore” (from drop-down menu) and click “OK” (Fig. 15). Geneious would then attach the sample’s trace file to its appropriate reaction well. After all traces have been attached and saved, they are ready to be downloaded into the Geneious Assembler. When the download is complete, all FIMS and LIMS metadata will be attached to the correct sample, and each sample will be ready for editing and assembly.

3.2. “The Assembler” Module: Downloading, Assembling, and Exporting Edited Data

3.2.1. Importing Traces

The Assembler workflow is graphically represented in Fig. 16. Once trace files have been downloaded into Geneious’s assembler module, they can then be edited, saved, exported, and ultimately published.

To import the trace files into Geneious for assembly, first generate a new folder for the traces in your “Local” database (located in the service tree). To generate a new folder, highlight “Local” in the service tree and then select File > New Folder under the main toolbar.

The easiest way to import your traces into Geneious is to download them directly from the LIMS database (if you have previously attached your trace files to a sequencing plate). If downloading from the LabBench, traces will have all associated metadata. If you import traces from disk, you will have to manually set parameters such as read direction (see Note 12) and choose “Annotate from FIMS/LIMS Data” to link in other metadata (see Note 13).

To download your traces from the LIMS, select the sequencing plates you want to download in the Biocode plugin search, and

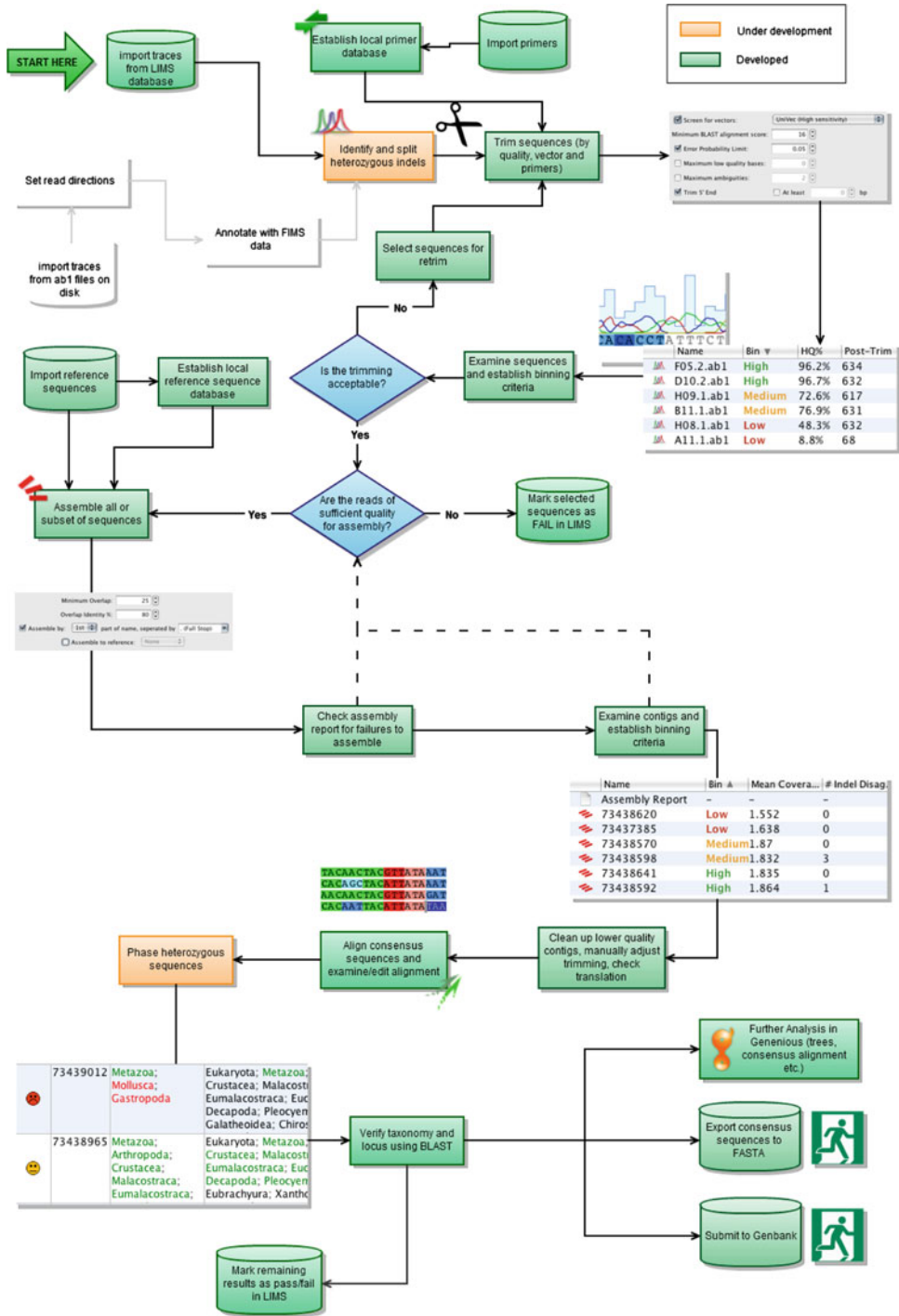


Fig. 16. Workflow in the assembler module.

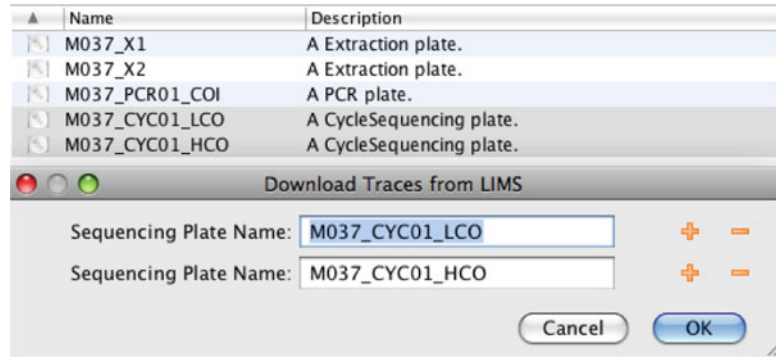


Fig. 17. An example of downloading two sequencing plates (one for each direction).

then select Biocode>Download Traces from LIMS (Fig. 17). Typically, you would choose one forward- and one reverse-sequencing plate for a set of sequences. You may select a single plate if both your forward and reverse reads are contained on it. Click “OK “and Geneious will ask you to choose a folder, and begin downloading your sequences. Once complete, you will have all of the traces from the plates you entered and they will already have their read directions set and be annotated with the necessary data from the FIMS. If you know the names of your sequencing plates, you can download them directly without having to perform a search. Select the folder you want to download to in Geneious, then select Biocode>Download Traces from LIMS, and enter your plate names manually.

3.2.2. Batch Rename

If you want to change the names of your reads to reflect aspects of the FIMS data, from the main toolbar select Edit>Batch Rename to copy your choice of fields into the name column. This feature is also available in renaming assemblies.

3.2.3. Trimming

Geneious treats trimming as an annotation class so that information is not lost once a sequence is trimmed. The underlying raw data are 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. To trim sequences, highlight all of the sequences you are going to assemble, and from the main toolbar, select Sequence>Trim Ends. You can also add trim ends to the main toolbar by right clicking on the toolbar and turning on Trim Ends (Fig. 18). For most applications, the default of Error Probability Limit 0.05 is a good start. This option works by trimming the sequence to find the longest possible untrimmed region, which has an overall error probability <0.05. Decrease the limit to trim more aggressively. Other options include screening for vectors, which

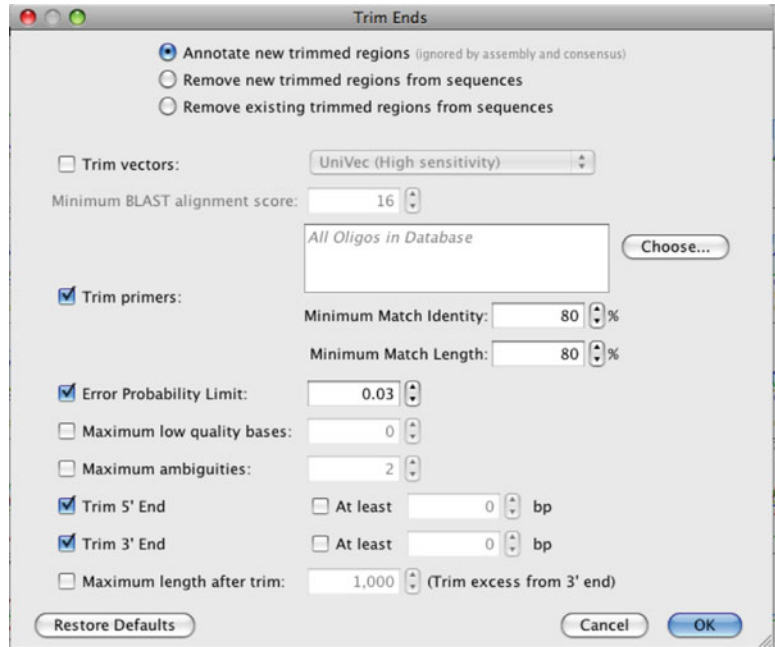


Fig. 18. The Geneious trimming dialog.

uses a clone of NCBI’s VecScreen tool, screening for primers, and basic limiting options, such as minimum amount to trim.

When you click “OK” to run trimming, it will add annotations to each of the sequences, which correspond to the trimmed regions. You can flick through all of the trimming results in the Sequence View before saving the changes. If there are reads which are obviously not of good enough quality for assembly, then you can mark these as failed in the LIMS, but it is easier to let the assembly report pick these out for you (described in Subheading 3.2.5).

Before using the primer screening feature, you need to establish a database of primers in Geneious (see Note 11). This is done by creating an oligo-type sequence that can be stored anywhere in your local folders (or you can generate a specific “Primer” folder if you want to store them all in one place). Oligo sequences are generated in one of the following ways.

- Extract a primer/probe annotation from a sequence.
- Select “Sequence” → “New Sequence” from the menu and choose Primer or Probe as the type of the new sequence.
- Select one or more existing primer sequences (these may be imported from a file, e.g., a fasta file) and then click “Primers” → “Primer Characteristics” to transform them into oligo-type sequences.

Once you have generated oligo sequences for all of your primers, the “Screen for Primers” option will search for matches with

any of them. Geneious uses a custom Smith–Waterman search to locate primer matches in sequences. You can retrim sequences using different parameters at any stage. To do this, just select the sequences for retrimming and follow the steps above for trimming. The only difference is you should select the “Annotate new trimmed regions” option to have the new trims replace the old one. When a sequence is retrimmed, it currently stores the history of trims that were used in the Notes tab for each sequence.

Sequences can be manually trimmed as well by selecting a region at the end of a sequence in the Sequence View and clicking “Annotate” and choosing “Trimmed” for the annotation type. If a sequence has more than one trimmed annotation at its end, then the largest trimmed annotation will be used. You can also manually edit any trimmed region by clicking and dragging either end of the trim annotation in the sequence view.

3.2.4. Binning

“Binning” is used to group traces and assemblies into three categories (high, medium, and low) based on various measures of quality (see Binning Parameters below). The purpose of binning is to speed up processing by summarizing the properties of sequencing results. The Bin column is hidden by default. To view, click on the small icon in the top-right corner of the table and then check the “Bin” item (this can also be done by right clicking on the table header). Documents can be sorted according to Bin by clicking on the Bin table header.

The binning parameters define thresholds for each bin (Fig. 19). They cover metrics, such as the percentages of high- and low-quality bases in your sequences, sequence length, number of ambiguities, and, in the case of assemblies, coverage (see Note 14). For information on any of the binning parameters, hold the mouse over the option to get a description.

There are three levels at which binning parameters can be set: globally (for all local and server documents), per-folder (all documents inside a particular folder or any subfolder), and per-document. To set the global binning parameters, from the main toolbar, select Tools > Preferences and go to the Sequencing tab. To set per-folder or per-document parameters, select the folder or documents you want to change and, again, from the main toolbar select Sequence > Set Binning Parameters. The most specific parameters are used in favor of less specific ones. Per-document parameters are used over any per-folder or global parameters that are set.

To help in the detection of frameshifts, you can set the number of stop codons as an optional binning parameter. The number of stop codons is calculated for the specified genetic code, and is defined as the minimum count of stop codons in the consensus sequence for all frames (i.e., we check frames 1, 2, 3 in the forward and reverse direction, count the number of stop codons in each, and then take the frame with the minimum number of stops).

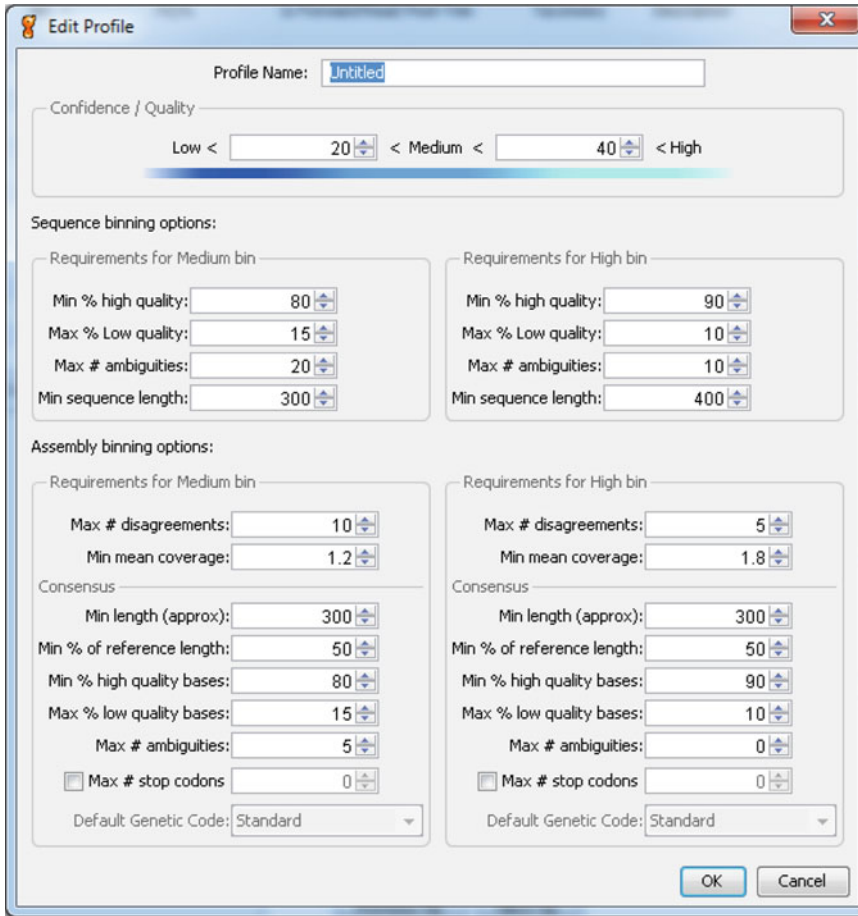


Fig. 19. The binning parameters dialog box.

As an example, when looking at assembly results, the bins could be used in the following way (if the parameters as set up as such).

- High = There is probably no need to look at these assemblies.
- Medium = These assemblies may need to be edited.
- Low = Fail: These assemblies are likely beyond rescue and should be marked as failed.

To set up the parameters in this way, you need to have strict parameters for the high bin, for example 0 ambiguities, 658 consensus length, and 1.8 coverage (for the COI barcode). The medium bin can be quite relaxed depending on how many assemblies you want to examine. Users can create binning profiles and share these with collaborators or use them for standard operating procedures to ensure repeatability.

3.2.5. Assembly

Select all of the reads you are going to assemble (and a reference sequence or list if you have one) and then click the “Assembly”

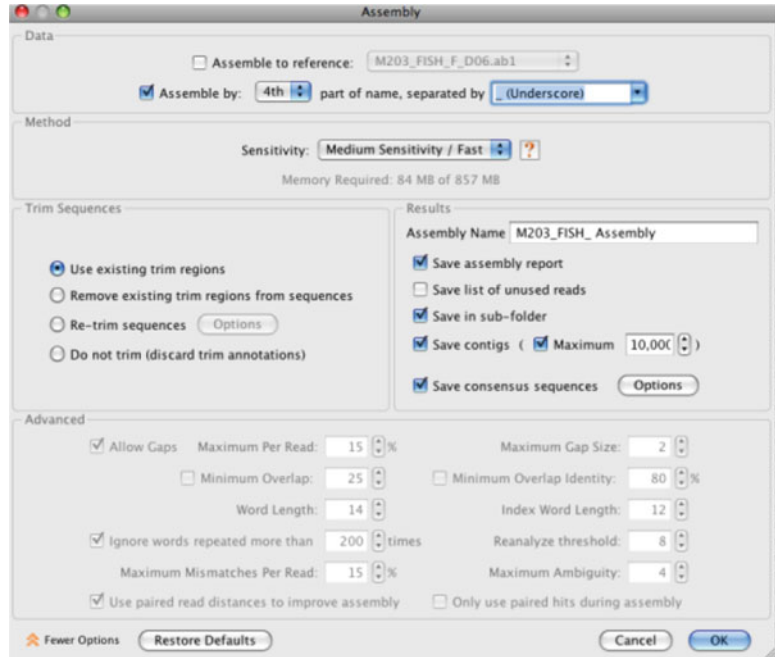


Fig. 20. The assembly dialog box.

button in the main toolbar. To assemble pairs of reads by name, check “Assemble by name” and choose the appropriate delimiters (Fig. 20). The recommended Sensitivity is “Highest Sensitivity/Slow.” It is also possible to choose “Custom Sensitivity,” and choose your own parameters (e.g., minimum overlap). If you have already trimmed your sequences, select “Use existing trim regions”; otherwise, specify trim options. “Save assembly report” and “Save results in a new subfolder” should both be selected. After clicking “OK,” a new subfolder called “Assemblies” will be generated and assemblies will be added to it as the operation runs. When the operation is finished, an Assembly Report and list of Consensus Sequences will also be added to the folder. Geneious generates a new subfolder each time an assembly is run.

The assembly report (Fig. 21) provides a record of which reads were assembled successfully and which reads failed. For example, click the blue hyperlink next to the red “X” to select all reads which failed to assemble and use the “Mark as Failed in LIMS” tool to mark these reads for resequencing.

3.2.6. Viewing And Editing Assemblies

As with the traces, assemblies are each assigned a bin based on various quality criteria. By default, Geneious uses a Highest Quality consensus which rarely generates ambiguities because the highest quality base call is used automatically. However, ambiguities are generated in situations, where the qualities of conflicting bases are similar.

Assembly Report

✓ 88 of [188 reads](#) were assembled to produce [44 contigs](#)
 ✗ [100 reads](#) were not assembled

Assembled using Geneious assembler. [Show Options](#)
 Assembly Duration: 50.59 seconds (1 minute and 9 seconds CPU time)

✗ Not Assembled

A01.ab1 - [2 Reads](#)

- [M202_FISH_F_A01.ab1](#)
- [M202_FISH_R_A01.ab1](#)

A02.ab1 - [2 Reads](#)

- [M202_FISH_F_A02.ab1](#)
- [M202_FISH_R_A02.ab1](#)

A03.ab1 - [2 Reads](#)

- [M202_FISH_R_A03.ab1](#)
- [M202_FISH_F_A03.ab1](#)

A04.ab1 - [2 Reads](#)

- [M202_FISH_F_A04.ab1](#)
- [M202_FISH_R_A04.ab1](#)

A05.ab1 - [2 Reads](#)

- [M202_FISH_R_A05.ab1](#)
- [M202_FISH_F_A05.ab1](#)

Fig. 21. The assembly report (showing assembled and unassembled reads).

The procedure for checking disagreements in an assembly is as follows.

- Turn on Allow Editing in the Viewer toolbar.
- Select an assembly to display an overview. Disagreements are shown as small black marks on the sequences and the trimmed regions can be seen.
- Highlight disagreements or ambiguities in the Display tab of the control panel to the right of the viewer. Ctrl+D on Windows and Linux or Command+D on Mac OS jumps between highlighted bases. If this is the first assembly you have looked at, you should zoom in to a level you are comfortable with. Geneious remembers this zoom level for the next assembly (Fig. 22).
- If you agree with a call or an ambiguity in the consensus, then you can go to the next disagreement because the call has already been made.
- If you disagree, you can resolve the conflict by editing either of the traces or by editing the consensus (editing the consensus is a shortcut for changing all base calls at that position).
- Continue editing through the disagreements until you have looked at all of them. Save the assembly and repeat for the next assembly.

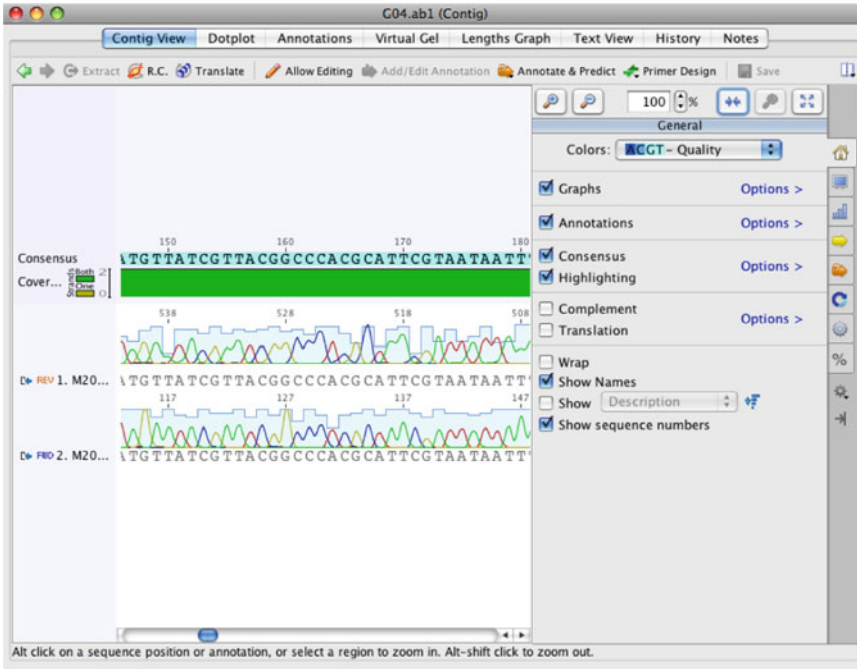


Fig. 22. A sample assembly view in Geneious.

If you decide that some assemblies are not good enough despite having assembled correctly, then you can mark these as failed at this point. Select the assemblies that have failed and go to “Mark as Failed in LIMS.” It is a good idea to move the failed assembly to a new subfolder (e.g., named “fail”).

3.2.7. Alignment of Consensus Sequences

An alignment of consensus sequences is a useful tool for checking and correcting assembly accuracy, especially near the ends of traces, where there might be poor coverage. To generate an alignment of consensus sequences:

- Select all of the assemblies you want to align and click the “Alignment” button in the toolbar.
- In the Alignment dialog box (Fig. 23), click “Consensus Align,” select “Generate alignment of consensus sequences only,” and choose an alignment algorithm (e.g., MUSCLE, MAFFT, and ClustalW).
- Click “OK” and an alignment is generated as a new document (Fig. 24).

The alignment retains all information from the original assemblies. Clicking the small blue arrow button to the left of each name brings you to the associated assembly (Fig. 24). Geneious currently does not propagate changes in the alignment back to the original assembly, but you can use the alignment for downstream steps so that alignment edits are not lost.

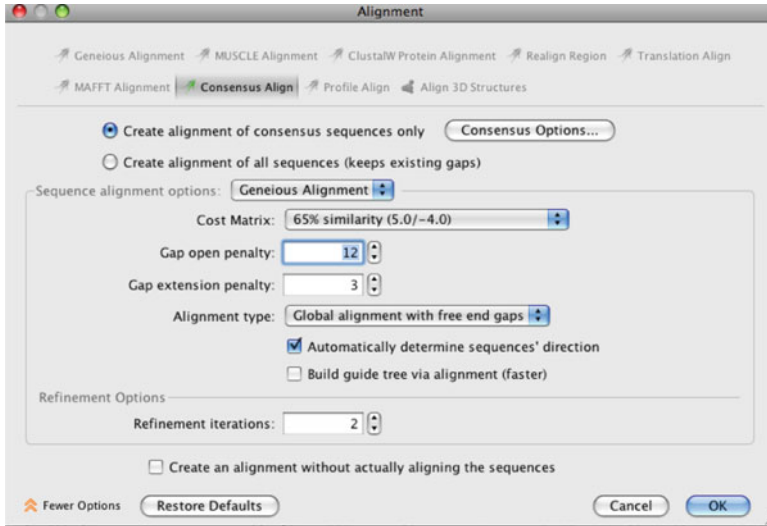


Fig. 23. Alignment dialog box.

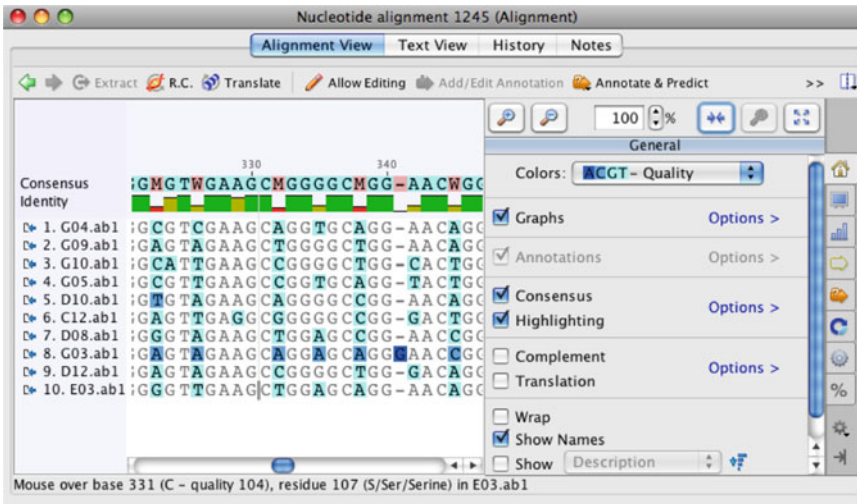


Fig. 24. Nucleotide alignment consensus dialog box.

To view the alignment translation, follow these steps.

- In the options to the right of the alignment view, change the Colors option to “By Translation.”
- Turn off the Highlighting option.
- Open the Complement and Translation section and set up the appropriate translation options, such as *Genetic Code* and *Frame*. We recommend that “Translation” is set “Relative to Consensus.” You can also set the amino acid *Color* scheme here (e.g., MacClade).

- You should also turn off Annotations so that editing history annotations do not interfere with the layout.
- Check the alignment for frameshifts and stop codons (binning should have identified these previously).

Clicking on Help in the toolbar while viewing an alignment displays documentation on editing and shortcut keys.

3.3. Verify Taxonomy and Loci

To help verify taxonomy annotated in the FIMS and identify contaminants (high-quality sequences, but wrong targets), Geneious can run a specialized batch *BLAST* against the NCBI public DNA sequence database. This can be run on any selection of contigs and alignments of contigs. If you have performed an alignment as above, then you should use the alignment to make sure that you are using the edited consensus sequence.

- Select an assembly, a list of assemblies, or an alignment.
- From the main toolbar, select Biocode>Verify Taxonomy (Fig. 25). This brings up the standard BLAST options. It is required that “Fully annotate hit summaries” is turned on but the rest of the options can be modified as necessary. Click “OK” to begin the search. This can take quite a long time to run due to BLAST.

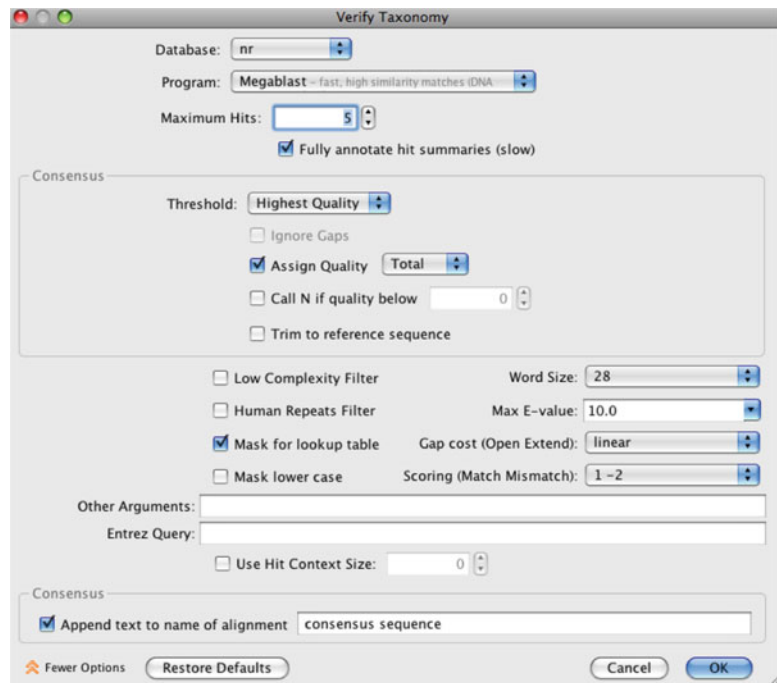


Fig. 25. Verify taxonomy dialog box.

Bin	Qu...	Query Tax...	Hit Taxon	Keywor...	Hit Definition	Hit Length	Hit Identity	Assembly Bin
E02.a		Metazoa; Chordata; Craniata; Gnathostom; Actinopteryg; Acanthopteryg; Perciforme; Percoidei; Sillaginidae	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei; Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes; Ostraciidae; Lactoria	COI, cytochr	Lactoria cornuta voucher HLC-TSZ-006 cytochrome oxidase subunit 1 (COI) gene, partial cds; mitochondrial	87%	85.5%	Low
D06.		Metazoa; Chordata; Craniata; Gnathostom; Actinopteryg; Acanthopteryg; Perciforme; Percoidei; Apoqonidae	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei; Acanthomorpha; Acanthopterygii; Percomorpha; Perciformes; unclassified Perciformes	COI, cytochr	Perciformes sp. BOLD:AAG7087 voucher MFLV13423 cytochrome oxidase subunit 1 (COI) gene, partial cds; mitochondrial	93%	96.3%	Low
F11.a		Metazoa; Chordata;	Eukaryota; Metazoa; Chordata; Craniata;	COI, cytochr	Abudedefduf vaigiensis			Low

Fig. 26. A sample verify taxonomy table; see “Note columns” for explanation of headers.

- When the process is complete, a “Verify Taxonomy Results” document will be produced (Fig. 26). This displays a table, which has a row for each of the queries comparing them with each of their top hits returned from BLAST. As with traces and assemblies, customizable binning options are available for efficient reporting on the results (see Note 15).
- Rows can be selected in the table by clicking/dragging and holding shift/ctrl/command while clicking. Click on “Go To Queries” to jump to the assemblies associated with the selected rows. Click “Show Other Hits” to see additional hits that were downloaded for the selected row. “Show Other Hits” is only enabled when one row is selected. Double clicking on a row also shows other hits.

The Verify Taxonomy Results may reveal that some sequences do not match the expected taxonomy. If you decide that the sequencing was a failure (possibly due to contamination), you can go back to the assemblies and “Mark as Fail in LIMS” and list the reason in the notes section. Also, as mentioned above, it is always a good idea to move any failed assemblies to a new subfolder (e.g., named “fail” or “contaminants”).

3.4. Mark Sequences as Pass in LIMS

Once you have verified taxonomy, assured that all sequence quality parameters are acceptable, and trimmed the primers, select either the assemblies themselves from the Assembly folder or the aligned consensus sequences from the alignments and select “Mark As Pass

in LIMS” under the Biocode button. This action writes the following data to the LIMS database:

- The extraction ID.
- The consensus sequence (with sequence quality values).
- The parameters used to trim and assemble the reads.
- The average coverage of the assembly.
- The number of disagreements in the assembly.
- A record of any edits made to the sequences in the assembly.
- The assembly bin.

By marking a sequence as Pass, this operation saves the consensus sequence of your assembly to the LIMS. This is the sequence that you submit to public sequence databases. You should make sure that the sequence is of sufficient quality and that you have completed all edits before you Mark as Passed.

3.5. Searching the Database

Biocode searches return four types of documents as follows.

- Tissue sample documents—each of these represents a tissue sample in the field database. Tissue documents contain collection information, and optionally taxonomy and photographs.
- Plate documents—these represent a plate in the lab database, and contain a diagram of the wells, as well as the plate’s thermocycle and attached GEL images if available.
- Workflow documents—these contain a linked set of reactions performed on an extraction.
- Sequence documents—sequences entered into the LIMS when traces/assemblies are marked as pass/fail.

You can perform either a basic search or an advanced search. Basic searches are performed by entering text into the search box, and return all documents that have a field with a similar value to the text entered (Fig. 27). You can restrict searches to particular types of documents by unchecking some of the checkboxes to the right of the search box.

Advanced searches explicitly search against particular fields. They are performed by clicking the *More Options* button. Click the + and – buttons to add and remove fields from the search.

Match of the following:

<input type="text" value="Collector"/>	<input type="text" value="contains"/>	<input type="text" value="Indiana Jones"/>	<input type="button" value="+"/>	<input type="button" value="-"/>
<input type="text" value="Common Name"/>	<input type="text" value="contains"/>	<input type="text" value="Nazi Gold"/>	<input type="button" value="+"/>	<input type="button" value="-"/>

Fig. 27. Search box.

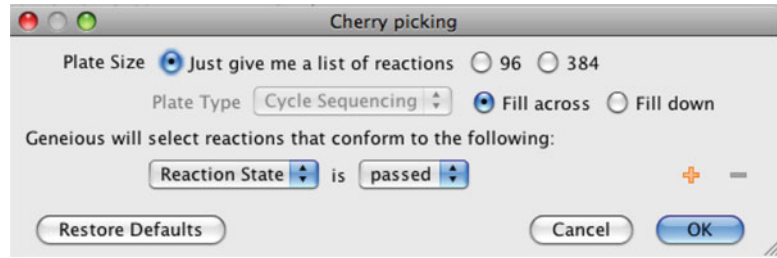


Fig. 28. The cherry picking window.

Choose the fields you want to search using the leftmost drop-down, and choose the search condition using the drop-down box to its right (see Note 16).

3.6. Cherry Picking

The Cherry Picking function (Fig. 28) is available in the Biocode drop-down menu and allows you to select reactions from one or more plates, based on the criteria that you specify (e.g., failed reactions for second attempts or extractions based on taxonomy for additional genes). You can use these selected reactions to create a new plate (or plates) or have them returned to you as a list. To perform Cherry Picking, select the plates containing the reactions you want to pick and click on Cherry Picking in the Biocode toolbar menu. Choose your destination, and then choose the criteria to select your reactions. You can add additional criteria using the orange “+” button on the right.

3.7. Preparing Your Documents for Genbank Submission

Genbank has stringent requirements for submitted sequences. It is important that you correctly prepare your sequences before you begin the submission process. This section outlines the fields you need for your submission, and how to attach them to your Geneious documents. All fields that are a part of your Genbank submission need to be either entered in the submission dialog or annotated on your sequences. In order to receive the BARCODE keyword in Genbank, the following fields need to be annotated on your sequences:

- Specimen Voucher/ID.
- Sequence ID.
- Target locus.
- Collector.
- Collection date.
- Identified By.
- Organism.

For non-barcode submissions, requirements vary depending on what type of sequences you are submitting (e.g., AFLPs, SNPs).

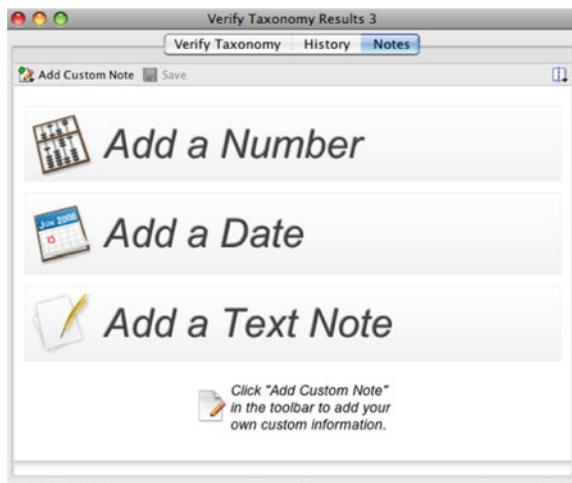


Fig. 29. The document notes tab, from a group of selected assembly documents (showing primer information attached).

nDNA, etc.). We recommend that you check with the NCBI Trace Archive: <http://eutils.ncbi.nlm.nih.gov/Traces/trace.fcgi?cmd=show&if=rfc>.

3.7.1. Attaching Data to Your Sequences

The easiest way to attach data to your sequences is to have included it in your FIMS database. When you download the traces from LIMS (or annotate the sequences with FIMS/LIMS data; see Note 13), the information will be automatically attached to your sequences. If you are using the submission tool without a FIMS or LIMS or you have extra information you want to attach, then you can use Document Notes. Document Notes appear as the rightmost viewer for any selected document(s), and enable you to store arbitrary information on your sequences (Fig. 29). When you click on the notes tab, you will see a list of the notes currently added to your documents, displayed in name/value pairs. To add a note, click the “Add Custom Note” button in the toolbar to see a list of predefined note types. “The Genbank Submission” note type contains the fields most commonly used by submitters. Any notes (and note fields) added to your documents will be able to be attached to your Genbank submissions.

If you do not see a note type that meets your needs, you can generate your own by clicking “Edit note types.” Click the “Generate Note Type” button in the note types dialog, and click the orange + buttons to add some note fields. We recommend choosing “Text” as the field type for Genbank fields. Once you have generated your note type, add it to your selected documents by selecting it from the “Add Custom Note” drop-down menu.

Any values you enter in the viewer are applied to all selected documents when you click save.

3.8. Genbank Submission

The Genbank Submission plugin allows you to submit your contigs to Genbank once you have completed all edits.

3.8.1. Preparing Your Submission

If you did not install the Genbank submission plugin when you set up the Biocode plugin, do so now (<http://software.mooreabiocode.org/index.php?title=Download>). Once the plugin is installed, select the contigs and/or sequences that you want to submit to Genbank, and click Tools>Submit to Genbank. Fill in the options and fields (Fig. 30) according to the following guidelines.

- The submission name is a free-form field and does not affect the results of your submission, so can be filled in as desired.
- Click *Edit Publisher Details* to edit your author/publication information. This information is preserved between submissions; so for many cases, it does not need to be changed between submissions.
- The next set of options matches fields annotated to your documents. You may choose a field from the drop-down to map a field on your documents to a Genbank submission field. All fields displayed in the main dialog are required. If you want to add optional fields, click the “Additional Source Fields” button. You can choose the fields you want from the drop-down menu, and click the + button to add more fields.
- If you have selected documents with traces and want to include them in your submission (required for BARCODE keyword), click the “Include Traces” checkbox. The required fields are variable, so the options you see will change depending on what values are selected. You can use the additional fields button to add optional fields to your submission.
- Check the “Include Primers” checkbox to include primers. If you are submitting sequences annotated from the LIMS, then primers will have been annotated on your sequences as document notes. If not, you can annotate the primers yourself by clicking the notes tab when viewing the selected documents and choosing “Sequencing Primer” in the “Add Custom Note” drop-down.
- If you have selected assemblies, you can choose the options used for building the consensus sequence (the passed consensus is what is submitted to Genbank).
- If you have chosen BARCODE as your experimental strategy, then you are able to enter the target locus (gene) of your sequences. This will be included in your submission as gene annotations spanning the entire length of your sequences.

Fig. 30. The Genbank submission options. Some of these options may not be displayed, depending on your selected documents.

3.8.2. Validation and Submission

You may either generate a submission file or upload the submission directly to NCBI (you need a BankIT FTP account to do this: see Note 17). You can make your choice at the top of the submission options dialog. If you are updating an existing submission, you can choose the update option and enter the BankIt ID in the field provided. Otherwise, choose “Upload New Submission” and a new submission ID is generated for you.

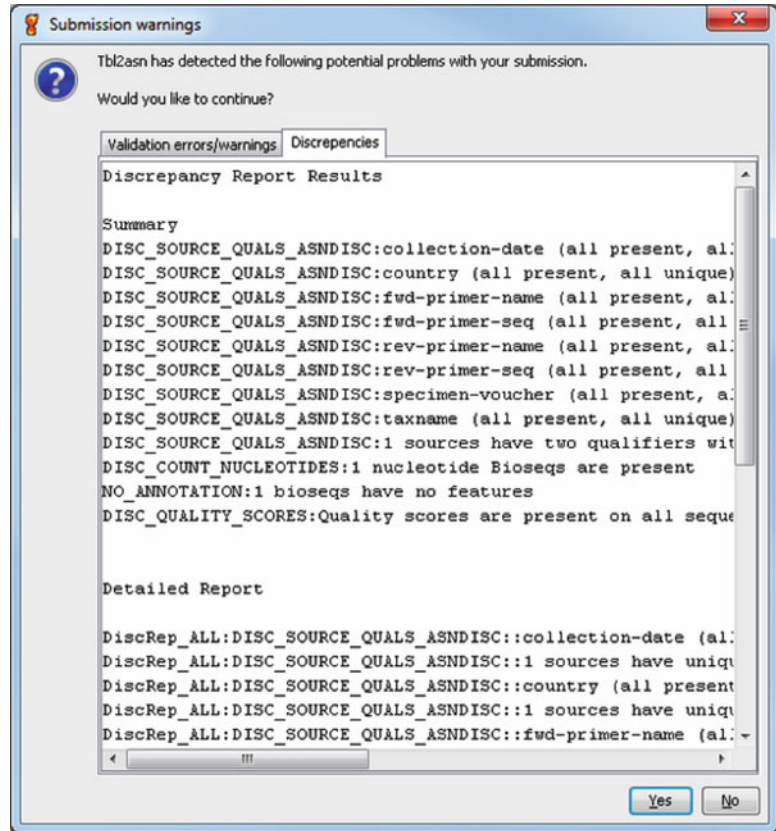


Fig. 31. The Genbank validation dialog.

Your submission is validated using tbl2asn, and you will be shown any problems before the submission is commenced. The validation result window has two tabs. (1) The Validation errors/warnings tab shows you a list of errors that may prevent your submission from being accepted. (2) The Discrepancies tab shows you potential errors that you may have made, based on common errors made in Genbank submissions. It is recommended that you thoroughly review the information in both tabs before proceeding.

If you have chosen to automatically upload your submission, further validation will take place on the server once the upload is complete. Geneious informs you whether your submission has been Accepted, Accepted with Warnings, or Rejected (Fig. 31). You should also receive an e-mail from Genbank detailing your submission. Once your submitted sequence(s) have been assigned accession numbers, you should receive a further e-mail from Genbank with the details.

3.9. Getting Help

The Biocode plugins are complex, and while learning how to use them can seem like a daunting task, help is available. Your first port of

call should be the Geneious introduction video (<http://www.biometers.com/assets/demonstrations/biocode.html>), which walks you through the plugins, and an instruction manual is available at <http://software.mooreabiocode.org>.

A user community and technical support are available from <http://connect.barcodeoflife.org/group/lims>. Here, you can engage with the wider community, get help from experienced users, and make suggestions about how to improve the plugins.

If you have any questions or suggestions that you do not want to post to the community, you can e-mail at support@mooreabiocode.org.

4. Notes

1. EXCEL FIMS

For single users, or users who cannot set up a field information management systems (FIMS) server, the EXCEL FIMS is the easiest way to connect to a specimen/tissue database. Geneious will read data from an excel workbook, and convert the rows into specimen/tissue records. It is assumed that all molecular processes start with tissue and this is the entry point into the LIMS.

Your excel workbook must conform to the following:

Your workbook should have only one sheet. If you have more than one sheet, only the first sheet will be read.

Each column corresponds to a data field in your database. You can have as many columns as you like, but you must have at least a specimen ID column and a tissue sample ID column.

The first row of the table should be the names of the columns. The other rows should contain the data.

Right click on the biocode icon in the Geneious service panel (on the left-hand side of the main window), and click login. Choose “EXCEL FIMS” in the uppermost drop-down select box, and enter the location of the EXCEL file. Choose the columns that contain the specimen and tissue ID’s from the drop-downs (if you use only one ID for specimens and tissues, enter the same column in both drop-downs). Also enter the taxonomy fields in your excel file, in order of highest to lowest, using the + and – buttons to add and remove fields. You can also specify plate name and well columns in your sheet if you keep your tissues in plates. Just check the “The FIMS database contains plate information” checkbox, and enter your plate and well fields. You will then be able to make a direct copy of the tissue plate when making new extraction plates.

2. TAPIR FIMS

TAPIR, or TDWG Access Protocol for Information Retrieval, is a standard protocol for sharing specimen data. The TAPIR FIMS connection reads in tissue data from a TAPIR server. Reliably integrating museum collection management systems (CMS) or FIMS to a LIMS can be difficult. Often, data needs to be exported and then reimported into each system. The TAPIR protocol is an attempt to standardize the way in which collection databases communicate, and to remove the difficulties associated with collaborative collection management.

Setting up a TAPIR provider with LIMS extension

These instructions assume the use of the TapirLink software (written in PHP), a collections management database with tissue records, stored in a TapriLink-compatible relational database, and the free version of Geneious (biomatters software), with the Biocode plug-in installed.

Step 1: Set up TAPIR.

There are several TAPIR software installation tools (<http://wiki.tdwg.org/twiki/bin/view/TAPIR/TapirSoftware>). This tutorial assumes that you are using the TapirLink software available (<http://wiki.tdwg.org/twiki/bin/view/TAPIR/TapirLink>). TapirLink requires a Web server running PHP, with an appropriate module to connect to your FIMS database. We recommend Apache, although this is not required. Download the TapirLink installation archive, and follow the installation instructions included.

Step 2: Incorporate the LIMS extension into the TAPIR provider.

While setting up your TAPIR provider (or afterwards), be sure to include the LIMS extension as an additional schema (available at http://biocode.berkeley.edu/schema/lims_extension.xml). You will be asked to map your local database fields to the LIMS extension fields.

Step 3: Point the Geneious LIMS system to your TAPIR provider link.

Right click on the biocode icon in the service tree on the left-hand side of the main Geneious window, and click “login.” Choose “TAPIR” from the field database drop-down at the top of the connection dialog, enter the address of your TAPIR server. Choose an LIMS database (see Remote LIMS; Note 3 and Local LIMS; Note 4), and click OK.

3. Google Fusion Tables FIMS

Google Fusion tables is ideal for groups that want to have a collaborative shared FIMS database, but do not want to set up their own server. You may enter data directly into Fusion

Tables, or upload excel spreadsheets (see <http://www.google.com/fusiontables/public/tour/index.html>). As with the Excel FIMS, it is assumed that all molecular processes start with tissue and this is the entry point into the LIMS.

Right click on the biocode icon in the Geneious service panel (on the left-hand side of the main window), and click login. Choose Google Fusion Tables in the uppermost drop-down. When viewing a fusion table on the Web, the url will contain the phrase `dsrcid=XXXX` (with XXXX being a number). This number is your table's ID. Enter the ID in the space provided, and click Update. Choose the columns that contain the specimen and tissue ID's from the drop-downs (if you use only one ID for specimens and tissues, enter the same column in both drop-downs). Also enter the taxonomy fields, in order of highest to lowest. You may press the autodetect button to do this for you, or use the + and - buttons to add and remove fields. You can also specify plate name and well columns in your table if you keep your tissues in plates. Just check the "The FIMS database contains plate information" checkbox, and enter your plate and well fields. You will then be able to make a direct copy of the tissue plate when making new extraction plates.

4. MySQL FIMS

If you are already using a MySQL database for your FIMS but do not want to set up a TAPIR server (see Note 2), then you can connect directly to the MySQL database.

Right click on the biocode icon in the Geneious service panel (on the left-hand side of the main window), and click login. Choose MySQL Database in the uppermost drop-down. Enter your server URL, port, username, password, and database in the fields provided, and click Update. Choose the columns that contain the specimen and tissue ID's from the drop-downs (if you use only one ID for specimens and tissues, enter the same column in both drop-downs). Also enter the taxonomy fields, in order of highest to lowest. You may press the autodetect button to do this for you, or use the + and - buttons to add and remove fields. You can also specify plate name and well columns in your table if you keep your tissues in plates. Just check the "The FIMS database contains plate information" checkbox, and enter your plate and well fields. You will then be able to make a direct copy of the tissue plate when making new extraction plates.

5. Remote mySQL LIMS

The remote LIMS is intended for labs and research groups, where the lab data needs to be shared between and edited by a large number of users. To set up a remote LIMS server, you will need MySQL, available from <http://www.mysql.com>. Download and install the MySQL server software on a server

which is accessible by everyone who needs to use the LIMS. The next step is to create a blank schema. You can store multiple LIMS database on one server, but each database must have its own schema. Create a schema, and then run the following script to create a blank LIMS database: http://www.biomatters.com/assets/plugins/biocode/labbench_latest_mysql.sql.

You will need to create at least one user account with read/write access to this schema. To connect to the LIMS database, you need to choose “Remote Server” in the biocode login screen within Geneious.

6. Local LIMS

This is intended for single or small-scale users, and is a database that exists within your local copy of Geneious. To create a new database, click on the “Add Database” button, enter a name for your new database, and click “Ok.” A new, empty database will be created for you. If you have already created a database, select it from the drop-down (other users of Geneious will not be able to connect to any LIMS databases that you create as a local database). To create a LIMS database that you can use to share data with other people, please see Remote LIMS (Note 5).

The Local LIMS databases are stored within your local Geneious user directory, so they will be backed up if you choose to do a Geneious backup (choose the “Back Up” button from the main Geneious toolbar). It is suggested that you back up your Geneious data regularly to avoid losing data.

7. Workflow considerations

One extraction can belong to many workflows (as extractions are often used as stock for many reactions).

You can have any number of failed reactions in a workflow, and you can have any number of passed reactions in a workflow. The passed or failed status of a workflow for a given reaction type is taken from the most recent reaction of that type.

While each workflow can only contain reactions of a single locus, each locus can have any number of workflows (useful if multiple people are working independently on the same locus for the same extraction).

Workflows are created with reactions. Any reaction (apart from extractions) that has an empty workflow field when saved will have a new workflow created for it. That means that it is particularly important that you fill in the workflow field correctly for all reactions that you save. Fortunately, this is easily accomplished in the “Bulk editor” (see Subheading 3.1.1 and Fig. 7). Clicking autodetect workflows in the tools drop-down will automatically fill in the workflow field for any reactions that have an available workflow (i.e., one with a matching extraction and locus).

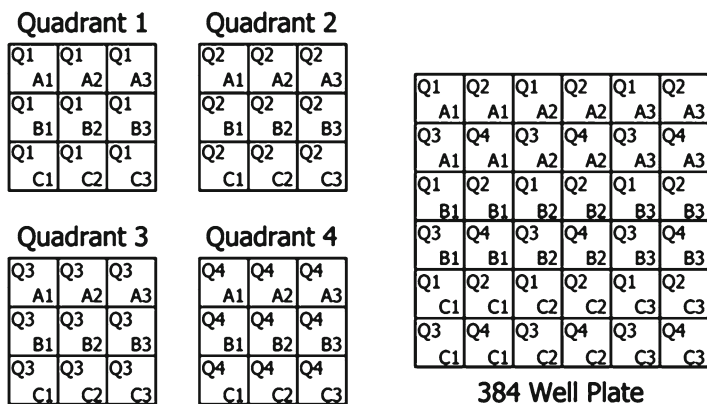


Fig. 32. A graphical illustration representing how four 96-well plates are converted into a single 384-well plate.

If more than one matching workflow exists, the most recent one will be chosen.

If no matching workflow exists, the workflow field will remain blank, and a new workflow will be created when you save the reaction.

Reactions are entered into the Lab Bench Database by plates. Plates come in a number of sizes, 48, 96, or 384 wells, and also in a grouping of any number of reactions. Creating a plate opens that plate in the plate viewer.

8. Converting between 96- and 384-well plates

It is possible to create a new 384-well plate from a group of 96-well plates, and to create a group of 96-well plates from a 384-well plate (Fig. 32). Each 96-well plate corresponds to one quadrant in the 384-well plate. To create a 384-well plate, select up to four 96-well plates in Geneious, and click “New Reaction.” Select the “Create plate from existing document” checkbox, and choose 384-well plate. A panel will appear at the bottom of the dialog which will allow you to choose to which quadrant each 96-well plate corresponds.

9. Creating custom cocktails

Cocktails are a recipe for the ingredients that will go into a reaction (excluding the primer). You can choose from a list of existing cocktails, or create your own. To create your own cocktail, click “Edit Cocktails,” then click new in the dialog, and enter the volumes and concentrations in the fields provided (Fig. 33). There is space for you to store one extra ingredient (both concentration and volume). Any additional information about your cocktail can be stored in the notes field. For safety reasons, you cannot modify or delete cocktails once they are created. You can create a copy of an existing cocktail by selecting it in the view, and then clicking “Add.”

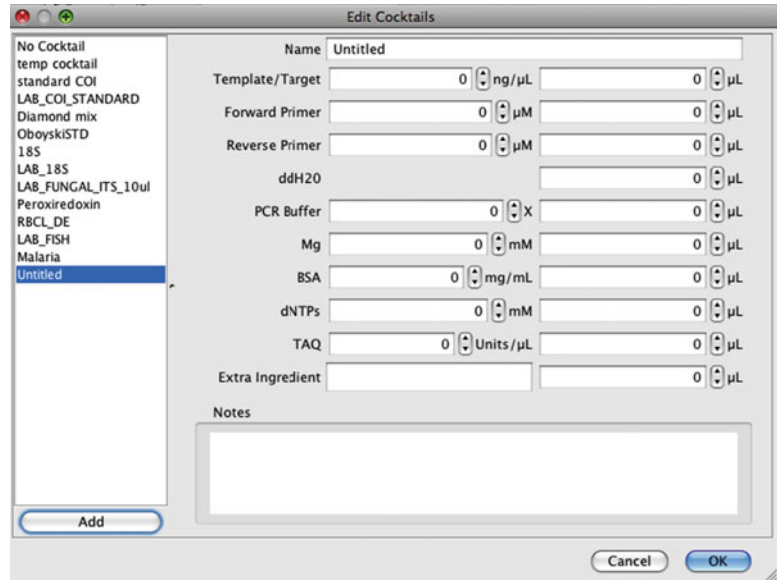


Fig. 33. “Edit Cocktails” window.

The new cocktail will have all the same volumes and concentrations as the one you selected.

10. Creating thermocycler profiles

To create custom thermocycler profiles, click “View/Add Thermocycles” in the New PCR plate toolbar. When the “Edit Thermocycles” window opens, click the “Add” button on the lower left-hand corner of the window (Fig. 34).

A New Thermocycle window will open and here you will be able to customize temperatures and cycles using the dialog boxes and “Edit Cycles” buttons (Fig. 35).

11. Creating Primers

To create a new primer in Geneious, click “Sequence” along the top of the main toolbar. In the Sequence drop-down menu, click “New Sequence.” The New Sequence window will open and at the bottom of the window choose “Primer” from the Type drop-down menu. Then, enter your sequence and identifying information in the dialog boxes and click ok (Fig. 36). Primers set on reactions will be saved to the lab bench database so that they can be viewed by others without you needing to send them your primer library.

If you have a large number of primers, you may want to organize your primers by type. You can do this by storing your primers in a folder structure in the Geneious service tree (for example, you could store all your primers for a particular locus in the same folder, or store your primers by taxonomy). This

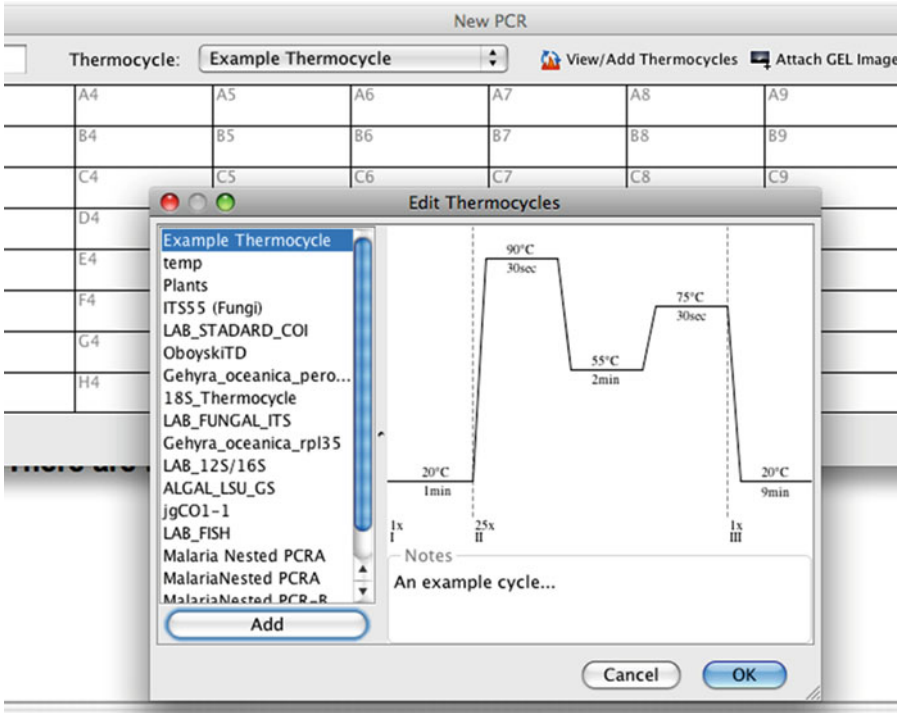


Fig. 34. "Edit Thermocycles" window from a new PCR plate.

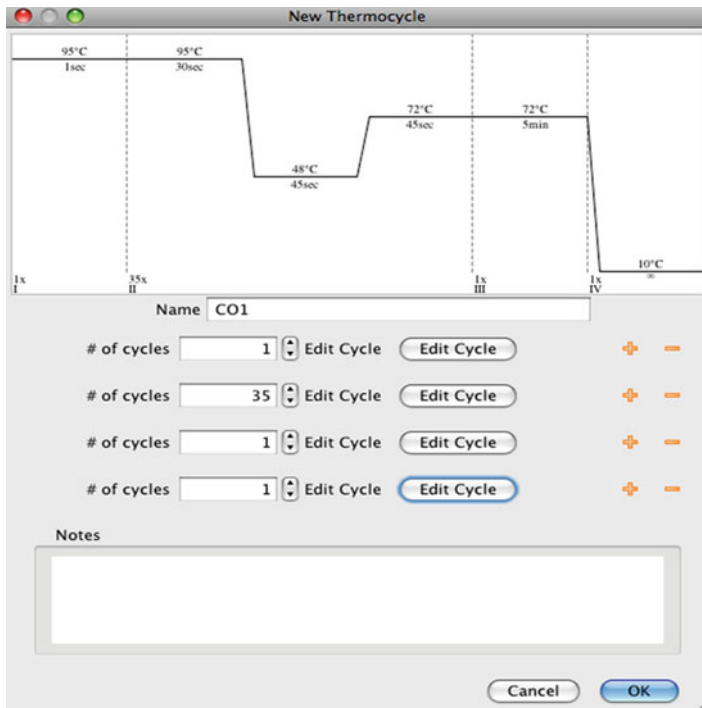


Fig. 35. A "New thermocycle" profile entry.



Fig. 36. "New Sequence" window for adding custom primers.

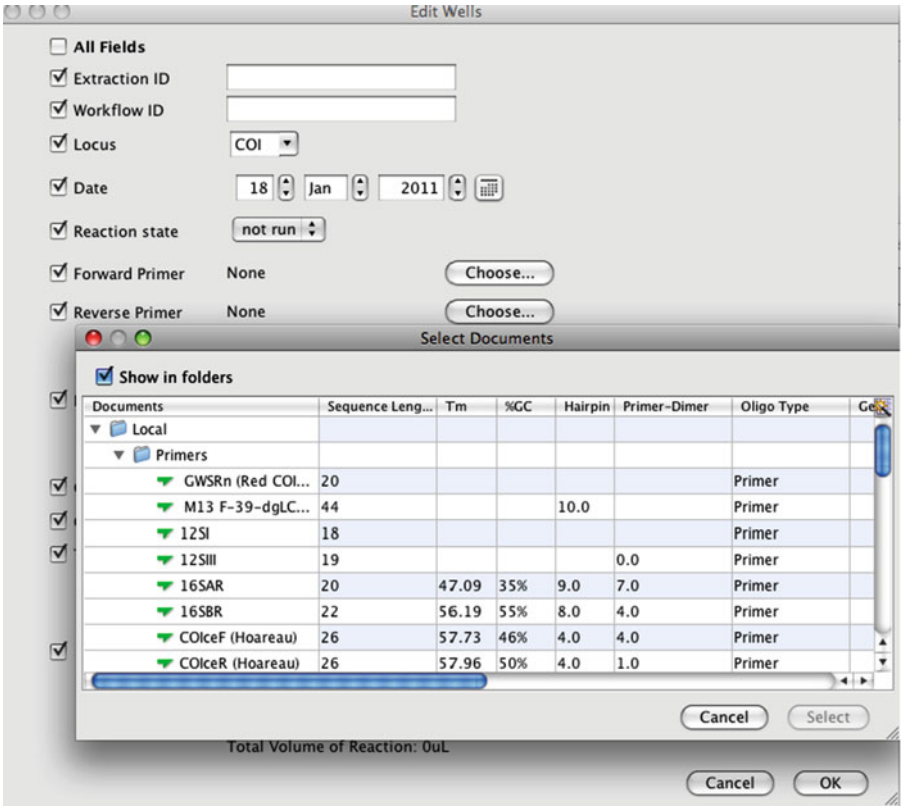


Fig. 37. Primer database accessed from the "Choose" button in the "Edit Wells" window.

folder structure will be displayed when you choose your primers when editing wells in your plate.

To choose a primer (or primers) for your wells in the plate editor, select the wells you want to edit, and click “edit selected wells.” Select the primer you want to add to the reaction (Primer fields display a list of primers in your local database), and click the “Choose” button (Fig. 37). You can choose any primer from your database, and it will be applied to all the selected wells. Only primers you have set on wells are stored in the LIMS database. Primers that have not been set on wells exist only in your local copy of Geneious and cannot be seen by others accessing the LIMS.

12. Importing .abi files from disk, setting read directions, and batch renaming

To import traces from a disk, locate the .ab1 or .scf files on disk and then click and drag them from the file manager on to the new folder in Geneious. Alternatively, you can import from inside Geneious using File > Import > From File... in the menu.

Once you have imported the raw trace files, it is currently necessary to tell Geneious which reads are in the forward or reverse direction. To set read directions, select all of either the forward or reverse reads from the ones you have imported and select Biocode > Set Read Direction in the toolbar. Choose either Forward or Reverse for the read direction and click OK. It is only necessary to mark either the forward or reverse reads; Geneious will work out the rest by process of elimination (this is so that the correct read is reversed during assembly and downstream steps are able to identify the direction of the reads).

After performing this task, an extra column will be added to the reads named “Is Forward Read” with a value of true or false.

If your forward and reverse reads are in different folders, it is easiest to import all of the reads from one folder, then set the read direction for those, and then import the second folder.

If you want to change the names of your reads to reflect some aspect of the FIMS data, from the main toolbar select Edit > Batch Rename to copy your choice of fields into the name column. This feature is also available in renaming assemblies. You can also use Edit > Batch Rename... to add _F or _R to the names of your reads if the names do not have any indication of direction (not required).

If you have imported both forward and reverse reads into Geneious before setting read direction, you can use Search or Filter in the top right corner of the Geneious window to locate a particular direction of read based on names.

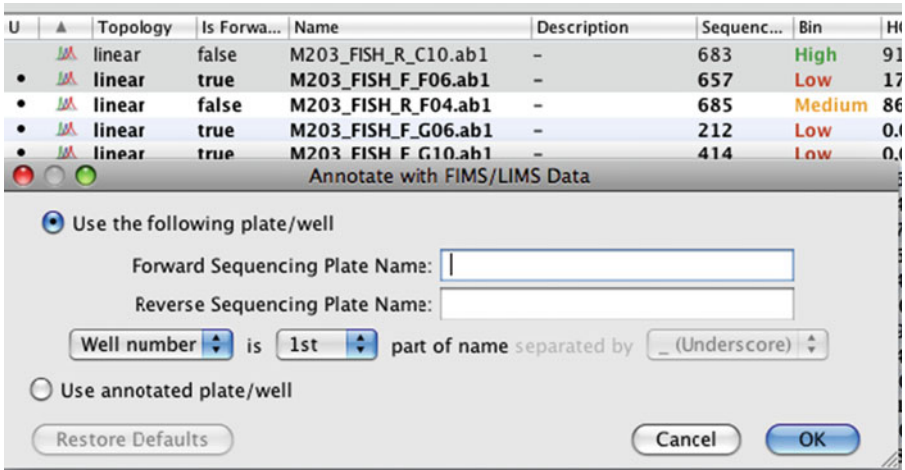


Fig. 38. “Annotate with FIMS/LIMS data” screen.

13. Annotating with FIMS/LIMS data

You can either enter a forward and reverse plate or use the annotated plate and well if you are updating sequences you have previously annotated.

To aid downstream analysis and submission, it is extremely useful to annotate sequences with the associated data from the FIMS. This must be done pre-assembly (with the reads) because forward and reverse reads can come from different sequencing plates. Annotating is the first step in the assembly pipeline that utilizes the FIMS/LIMS database, so you will need to connect to the Biocode service before proceeding. To do this, right click on Biocode in the source panel on the left-hand side and select login.

Select all of the reads which you imported and go to Biocode > Annotate with FIMS/LIMS Data in the toolbar. If you have plate data in your FIMS database and you do not wish to enter reaction information for your data in the LIMS, choose “Biocode > Annotate with FIMS data only...” (see below), and enter the name of your FIMS plate.

You need to enter the forward and reverse sequencing plate names (from the LIMS) which correspond to your reads and identify which part of the sequence names identify the well location. If both forward and reverse reads are on a single plate, then you can leave the reverse plate field blank or enter the same name twice. Click OK and the operation will add many new columns to the table for each of the reads (Fig. 38). These include things like Specimen ID, Taxonomy, and Collector. The values should be identical for each forward and reverse pair of reads.

Often, there will be reads which do not have entries in the FIMS due to sequencing results coming through from wells



Fig. 39. Example of Mean Coverage.

which were essentially empty. This operation will tell you about any of these and the extra columns will be left blank.

Annotating with FIMS data only

Please note that you will not get primer information for your sequences using this method, so you may have to annotate those yourself if you want to use the sequences to generate a genbank submission.

Tip: To get the empty well reads out of your way, you can easily select them all by sorting the table by one of the FIMS attributes (e.g., Tissue ID) and then selecting the ones with no value. You can then either delete them or create a new subfolder called “empties” and move them into there. If you want to change the names of your reads to reflect some aspect of the FIMS data, you can use Edit > Batch Rename... to copy your choice of fields into the name column.

14. Mean Coverage

Mean coverage is one of the binning criteria for assemblies and is also available as a column in the table. It is also the least intuitive value, so here is a description.

Coverage is the number of sequences that cover a given position in an alignment/assembly. Mean coverage is, therefore, the mean of this value across all positions in the alignment/assembly (Fig. 39).

For this alignment above, the first two positions have a coverage of 1. The next five positions have a coverage of 2 and the last three have coverage 1 again. Mean coverage is, therefore, $(2 \times 1 + 5 \times 2 + 3 \times 1) / 10 = 1.5$. The mean coverage will be between 1 and the number of sequences in the alignment/assembly. For a pairwise assembly, that means 2 is full coverage and 1 is no coverage.

15. Taxonomic Verification Binning

Similar to the bin column that has been used for reads and assemblies, Bin columns in the Verify Taxonomy Results window summarize properties of the verification process by assigning each result a High, Medium, or Low value (in the form of a smiley).

Query: The name of the query assembly.

Query Taxon: The taxonomy of the query from the FIMS. The verify operation fills in higher taxonomy by searching

NCBI taxonomy. If the taxon could not be found in NCBI, this will be noted and result will be marked as Low bin.

Hit Taxon: The taxonomy of the top hit from BLAST. Levels in the taxonomies are marked as green or red depending on whether they match with the query.

Keywords: A user-defined list of keywords which are expected in the hit definition from BLAST. These are highlighted red or green depending on whether they are found in the definition.

Hit Definition: The definition of the top hit returned from BLAST with matching keywords highlighted.

Hit Length: Length of the hit alignment from BLAST, highlighted according to binning parameters (red, orange, or green).

Hit Identity: Identity of the hit alignment from BLAST, highlighted according to binning parameters (red, orange, or green).

Assembly Bin: The bin that was assigned to the assembly according to the previously mentioned binning parameters.

You can sort by any of the columns as usual and rearrange/resize them.

16. Useful example searches

Last Modified (LIMS) | Greater Than | 01 May 2010—all work done after the beginning of May.

Plate Name (LIMS) | Contains | “Plate1”—all plates which have the phrase “Plate1” somewhere in their name.

Locus | Contains | “COI”—all COI workflows and plates.

17. Most users should use the Geneious BankIt FTP account when submitting sequences. Larger research groups or sequencing centers may wish to create their own submission account, which can be done by contacting gb-sub@ncbi.nlm.nih.gov.

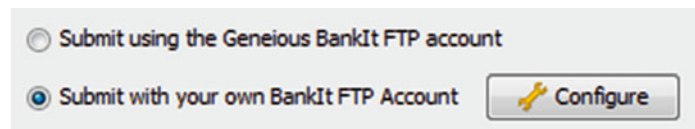


Fig. 40. Genbank Submission Account.