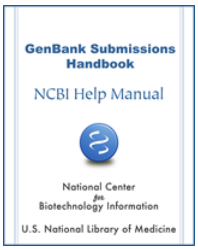
U.S. National Library of Medicine
National Center for Biotechnology Information

NLM Citation: The GenBank Submissions Handbook [Internet]. Bethesda (MD): National Center for Biotechnology Information (US); 2011-. Submission Wizard for Uncultured Samples. 2012 May 15 [Updated 2014 Jan 16].
Bookshelf URL: https://www.ncbi.nlm.nih.gov/books/



Submission Wizard for Uncultured Samples

Created: May 15, 2012; Updated: January 16, 2014.

Purpose

The Uncultured Sample Submission Wizard is for submitting sequences obtained from an uncultured source/ environmental samples only. Do not use this wizard for sequences from purified bacterial or fungal strains. This wizard will guide you in providing all of the necessary source information for uncultured samples and will provide assistance and direction with feature annotation. Examples of source information are provided.

Wizard Import Nucleotide Sequences

Requirements: The Uncultured Sample Sequence Submission requirements are listed in the Sequences tab of the Wizard Import Nucleotide Sequences dialog box.

Sequence Format: You may import your sequences in FASTA format or you may import an alignment. Use the Import Nucleotide FASTA button to import your properly formatted FASTA file. For help with how to format the FASTA file click the FASTA Format Help button. The Sequences tab will display the information about the imported sequence(s). Please check the number of sequences, Sequence IDs (SeqIDs) and length of each sequence to make sure this information is correct. You may also import a nucleotide alignment file in any of the Sequin compatible formats (fasta+gap, Nexus, Phylip). See examples of these formats here.

If the sequences contain a significant number of ambiguous bases near the 5' or 3' end, you may be prompted to trim or remove these sequences from your submission.

Trim Vector Contamination: It is highly recommended that you perform a vector screen on your sequences and trim vector contamination by clicking the Vector Trim Tool button.

Delete Sequences: You can remove sequences from your submission using the Sequence Deletion Tool under the Edit Menu. This tool will assist you in removing any sequences from your file that you need to delete or that do not meet GenBank minimum sequence length requirements.

Sequencing Method

If you are submitting over 500 sequences or your sequences were generated using next-generation sequencing technology, the information in this form is required.

Sequencing Method: Use the check boxes at the top of the form to select the sequencing technology type(s) used to obtain the sequences. Multiple types can be selected, if appropriate. If you used technology that is not listed in the form, please select other and use the free text box to provide the information.

Assembly Program: After selecting the sequencing technology, select the radio button to indicate if your sequences are raw sequence reads or sequence assemblies. If you are submitting assemblies using next-generation sequencing technology, the name of the assembly program and program version or date the assemblies were made are required in the free text boxes. If multiple assembly programs were used, Click on Add More Assembly Programs and complete the provided spreadsheet.

Raw sequence reads from next generation sequencing technologies should not be submitted to GenBank.

Submission Type

If you imported a nucleotide FASTA file and you are submitting more than one sequence, you will be prompted to select the type of submission you are creating. This dialog will not appear if you imported an alignment. If you select a set, all of the sequences in the set must have the same release date. The following submission types are available in the uncultured Wizard:

- Environmental Set: a set of sequences that were derived by sequencing the same gene from a population of unclassified or unknown organisms.
- Batch: related sequences that are not part of a population, mutation, or phylogenetic study. The sequences should be related in some way, such as coming from the same publication or organism.

Uncultured Sample Wizard Source Information

Requirements: All sequences must have an organism name, isolation-source or host, and unique clone name. Optional modifiers can be added to provide additional information, if known. The organism name should not contain the entire lineage information. Please review examples of uncultured sample organism names.

The clone name is used for both traditional clones and PCR product sample IDs. The clone names must be unique within the submission. In addition, information about the environment from which the sequences were isolated should be supplied within the isolation-source field. If the source organism was isolated from within a host organism, this should be supplied in the host field.

How to add source information: There are three ways to add the source information: 1) directly type into this form, 2) import a tab-delimited source table, or 3) automatically populate the form if source information was included in the FASTA definition lines.

You can set the same source qualifier value for all sequences by filling in the top row of boxes and using the appropriate Apply button. Use the Copy from SeqID button to apply the sequence IDs to the qualifier indicated in this table if this information was used as the sequence IDs in the original FASTA file.

Click on the Source Table Help button to open a text dialog with information on making a tab-delimited source table.

If you entered all required source information in the FASTA definition lines, minimal input will be necessary on this form.

Errors: Any problems or missing information will be listed on the right side of the form. If you have made any changes on this form, please use the Recheck Errors button to validate the new information. Use the Show only sequences with errors radio button to list only those sequences that did not pass the validation.

Are you unable to pass the Source Information window? If you have not provided some required source information, the issue will be listed in the ***Problems*** column. After fixing any problems, click the Recheck Errors Button to determine if all issues have been fixed. You may display only the entries with problems by selecting the radio button next to Show only sequences with errors.

Do you not see a source qualifier in the table that you want to use in your submission? You may add columns for some commonly added source qualifiers using the buttons below the table. Other optional modifiers can be added to provide additional information using the "Apply/See More Source Information" button or "Import Source Table" button. A window with instructions for creating a source table can be viewed by clicking Source Table Help.

Did you have source information in your FASTA file that is not displayed in this table? This table only displays the required source qualifiers for each type of submission. It does not display all source information. If your FASTA definition lines were correctly formatted, the extra source information you provided in the FASTA definition lines will be imported. You will be able to review this information in the record viewer.

Uncultured Sample Wizard Primer Type

Use this page to select the type of primers used in PCR amplification of the samples.

Select "universal primers" if the primers amplify DNA from a broad range of organisms.

Select "species-specific primers" if primers amplify DNA from a single species.

Uncultured Sample Annotation

Use the radio buttons to select the option that best describes the sequences. After completing all dialogs for each section, you will be directed to leave the Wizard and transferred to the record viewer. You must do so to complete your submission. However, you cannot return to the Wizard once you have exited.

Single rRNA, ITS, or IGS

Select this option if the sequences contain a single ribosomal RNA, one internal transcribed spacer, or one intergenic spacer across the entire sequence. Once you have selected this button, a new dialog will appear with radio buttons to select the type of organism from which the sequences were derived. Once the organism type is selected, a dialog listing common types of rRNA, ITS, or IGS will appear. Select the appropriate radio button. If none of the choices are appropriate, select Something else and type in a description of what the sequences contain.

Multiple rRNA, ITS, or IGS regions where spans are unknown

Select this option if the sequences contain more than one ribosomal RNA, internal transcribed spacer, and/or intergenic spacer and you are uncertain of the nucleotide locations of each feature. Once you have selected this button, a new dialog will appear with radio buttons to select the type of organism from which the sequences were derived. Once the organism type is selected, a dialog listing common types of rRNA, ITS, and IGS will appear. Select the appropriate checkboxes. If none of the choices are appropriate, select Something else and type in a description of what the sequences contain.

Multiple rRNA, ITS, or IGS where spans are known

Select this option if the sequences contain more than one ribosomal RNA, internal transcribed spacer, and/or intergenic spacer and you know the nucleotide spans of each feature. Once this option is selected, you will be prompted to exit the wizard and the record viewer will open. A text dialog will open with instructions for importing a five-column, tab-delimited table containing the feature locations. You may also apply annotation using the Annotate menu options in the record viewer. Alternately, if you imported an alignment you may use Feature Propagate or the Alignment Assistant to add feature annotation to your submission.

Intergenic spacer (not rRNA-IGS)

Select this option if your sequences contain intergenic spacer that is not ribosomal intergenic spacer. The dialogs that follow will prompt you for more information about the flanking genes and if these genes are present in the sequences. Please see more documentation here.

Coding Region (CDS)

Select this option if the sequences encode the same, single protein across the entire length of the sequence. Once you have selected this button, a new dialog will appear with text boxes to input the protein name, protein description, gene symbol, and comments. Only the protein name is required, other fields are optional. If the coding region is partial, check the appropriate 5' or 3' boxes near the top of the dialog as appropriate.

Something else, multiple features

Select this option if the sequences do not contain one of the feature types listed in the annotation dialog or you wish to apply annotation in the record viewer. Upon selecting this option, you will be prompted to exit the wizard and the record viewer will open. A text dialog will open with instructions for importing a five-column, tab-delimited table containing the feature locations. You may also apply annotation using the Annotate menu options in the record viewer. Alternately, if you imported an alignment you may use Feature Propagate or the Alignment Assistant to add feature annotation to your submission.

Wizard rRNA Chimera Checking

If you selected the annotation of 16S ribosomal RNA sequences from bacteria or archaea, an additional dialog will appear asking if the sequences have been screened with a chimera check program. If you have used a chimera check program, please provide the name and version if applicable. Note that BLAST is not a chimera check program. If you have screened the sequences for chimeras, please be sure to remove all suspected chimeric sequences before submission.