GenBank Submissions
Handbook

NCBI Help Manual

National Center for
Biotechnology Information

U.S. National Library of Medicine

Submission Wizard for Intergenic Spacer (IGS) Sequences

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

Purpose

The Intergenic Spacer (IGS) Sequence Submission Wizard is for submitting non-rRNA intergenic spacer sequences from pure, cultured fungal strains or vouchered fungal samples, plant, or animal sequences. This wizard will guide you in providing all of the necessary source information for these types of samples and will provide assistance and direction with feature annotation. Examples of source information are provided.

Wizard Import Nucleotide Sequences

Requirements: The Intergenic Spacer (IGS) 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 are submitting more than one sequence, you will be prompted to select the type of submission you are creating. 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 IGS Wizard:

- Pop set (Population study): a set of sequences that were derived by sequencing the same gene from different isolates of the same organism.
- Phy set (Phylogenetic study): a set of sequences that were derived by sequencing the same gene from different organisms.
- Mut set (Mutation study): a set of sequences that were derived by sequencing multiple mutations of a single gene.
- 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. This option will not be available if you imported an alignment.

IGS Wizard Type of Source

Use this page to select the type of organism from which your sequences were obtained.

IGS Wizard Source Information

Requirements: All sequences must have an organism name. Sequences obtained from cultured fungi must have unique strain names. Sequences from vouchered fungal samples must have specimen-vouchers and a unique combination of organism names, specimen-vouchers, and isolate codes. Sequences from all other organisms must have unique organism names or a combination or unique organism names, isolate names, and/or specimen-vouchers.

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.

IGS Wizard Genome

Use this page to select the location from which your sequences were derived. If your sequences were derived from nuclear genomic DNA, select nuclear. If your sequences are derived from an alternative location (i.e. organelle genome), select the appropriate location in this dialog.

IGS Wizard 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 exit the Wizard to complete your submission. However, you cannot return to the Wizard once you have exited.

Intergenic spacer only

Select this option if your sequences only contain intergenic spacer. If your sequences contain part of the flanking gene(s), do not select this option. Once this option is selected, a new dialog will prompt you for the gene symbols of the genes flanking the intergenic spacer at the 5' and 3' ends. Type the gene symbol for the gene flanking the spacer at the 5' end in the text box below 5' gene symbol. Type the gene symbol for the gene flanking the spacer at the 3' end in the text box below 3' gene symbol. You will then need to select the radio button to indicate if your sequences contain the partial or complete intergenic spacer at the 5' and 3' ends.

Intergenic spacer and other features (gene, tRNA) where spans are unknown

Select this option if your sequences contain intergenic spacer and part of the flanking gene(s) and you do not know the exact nucleotide locations of the features in your sequences. Once this option is selected, a new dialog will prompt you for the type of features flanking the spacer. Select the appropriate feature for the 5' end and 3' end. If you select tRNA, you will be prompted to select the name of the tRNA from a list of tRNAs. If you select protein coding gene, you will be provided text boxes to type in the protein name and gene symbol. After you have provided feature information, use the radio buttons to indicate if your sequences contain part of the features at the 5' end and 3' end. If the feature is not present at one or both ends, you will need to indicate if the intergenic spacer is partial or complete.

Intergenic spacers and other features (gene, tRNA) where spans are known

Select this option if the sequences contain intergenic spacer and other features (coding region, tRNA, etc.) and you know the nucleotide locations of the features in your sequences. 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.

Something else

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.