



Submission Wizard for Viruses

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

Purpose

The Virus Sequence Submission Wizard is for submitting virus and viroid sequence submissions only. This wizard will guide you in providing all of the necessary source information for different types of viruses and will provide assistance and direction with feature annotation. Examples of source information are provided.

Wizard Import Nucleotide Sequences

Requirements: The Virus 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 Virus 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.

Virus Wizard Type of Virus

Use this page to select the type of virus sequence(s) you are submitting. There are specific source requirements for certain virus types, including:

- Norovirus, Sapovirus (Caliciviridae)
- Foot-and-mouth disease virus
- Influenza virus
- Rotavirus

If the virus type is not listed or you are submitting sequences from a mixed set of different viruses, select the "Not listed above or mixed set of different viruses" radio button.

Virus Wizard Source Information

Requirements: Each type of virus submission has specific source requirements. Please see the sub-section below for specific requirements for each virus type. In addition to the specific requirements listed below, you will need to provide unique source information (such as unique strain or isolate names/IDs) for each sequence if all of your sequences are from the same gene/region.

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. Click the Export This Table button to export a tab-delimited template file. You must maintain the tab-structure of the table in order to correctly import the data back into the submission wizard. Do not use spaces between the columns.

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.

Norovirus, Sapovirus (Caliciviridae) Requirements

Norovirus or Sapovirus (Caliciviridae) submissions must include the following information:

1. organism name
2. isolate
3. collection-date
4. country
5. host or isolation-source

Use the Add isolation-source button if the source of the virus is better described as an isolation source rather than host.

Foot-and-mouth Disease Virus Requirements

Foot-and-mouth disease virus submissions must include the following information:

1. organism name
2. isolate

Use the other fields in the table to provide additional information about the source (country, collection-date, host, etc).

Influenza Virus Requirements

Influenza virus submissions must include the following information:

1. organism name
2. properly formatted strain
3. collection-date
4. host or isolation-source

5. segment
6. country
7. Influenza A viruses also must list the serotype

Use the Add isolation-source button if the source of the virus is better described as an isolation source rather than host. Passage history is optional.

Rotavirus Requirements

Rotavirus submissions must include the following information:

1. organism name
2. isolate
3. collection-date
4. country
5. host or isolation-source

Use the Add isolation-source button if the source of the virus is better described as an isolation source rather than host. Please use the other qualifiers in the table to provide additional information about the source.

Not listed above or mixed set of different viruses Requirements

All other virus sequences must include the following information:

1. organism name
2. isolate
3. country (optional)
4. collection-date (optional)
5. host (optional)

The country, collection-date, and host are optional fields, however we urge you to provide this information for any virus submission. Use the Add isolation-source button if the source of the virus is better described as an isolation source rather than host. You may be prompted at a later date for country, collection-date, and host/isolation-source information if you do not provide it in this table.

Virus Wizard Molecule Information

Use this page to select the molecule type that was isolated and sequenced in your experiment. The topology of the molecule can also be changed in this window. Only set the topology to circular if you are submitting a complete, circular viral genome or segment. Single genes or fragments of viral genomes should not be set to a circular topology.

If you selected mRNA as the molecule type you will be prompted for more information about your samples.

Virus 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 do so to complete your submission. However, you cannot return to the Wizard once you have exited.

Note about Influenza Annotation: If you selected Influenza virus as the type of source in a previous dialog, please select “Multiple features per sequence” and follow the dialog instructions for uploading a feature table made using the [NCBI Influenza Genome Annotation Tool](#).

Single coding region across the entire sequence

Select this option if the sequences contain the same, single coding region across the entire length of all of the sequences. 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.

Single non-coding feature across the entire sequence

Select this option if the sequences contain a single non-coding feature, such as a UTR or LTR, across the entire length of all of the sequences. Once you have selected this button, a new dialog will appear listing common types of non-coding features. Select the appropriate radio button. If none of the choices listed are appropriate, select "Something else" and a free text box will appear for you to type a description of what the sequences contain.

Multiple features per sequence (coding regions, LTRs, etc.)

Select this option if the sequences contain more than one feature and you know the nucleotide spans of each feature. Once this option is selected, a dialog will open with instructions for importing a five-column, tab-delimited feature table containing all of the feature locations and you will be prompted to exit the wizard and open the record viewer. 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.

If you are submitting Influenza sequences and you selected Influenza virus as the type of virus in a previous dialog, you will be prompted to make and upload a feature table in the dialog that follows. Please follow the instructions for uploading a feature table made using the [NCBI Influenza Genome Annotation Tool](#).