Annotating your Sequence for Submission

Created: April 6, 2011; Updated: November 3, 2014.

Why Should I Add Features to my Sequence?

What do you mean by feature annotation and why do I need to annotate my sequences?

Feature annotation is the addition of biological features such as genes and associated coding regions, structural RNA, variation information, exon, introns, etc. to your submitted sequence. The annotation should include the location of the feature (start and stop) and a description of the feature.

The addition of feature annotation to your sequence submission:

- Improves the quality of your submission
- Increases the efficiency with which your submitted sequences are processed by members of the GenBank staff
- Is of far greater use to the scientific community than sequence data alone.

Adding feature annotation will frequently provide an additional tool for reviewing the quality of primary nucleotide sequence data:

For example, annotating protein-coding regions will highlight potential errors in the nucleotide sequence, such as insertion/deletions (in/dels) or improper or uncertain base calls that result from the sequencing reads.

See an alphabetic list **of available Features** in the Sequin Help documentation (These Features can be used in both Sequin and BankIt).

See the "Annotation using BankIt" and "Annotation using Sequin" sections of this Quick Start for information about how to annotate your sequences.

Can I submit a sequence without annotating it?

You must provide some type of annotation with your sequence such as:

- Coding Sequence (CDS), including nucleotide spans and reading frame. Using this information, our software will add the amino acid translations for you.
- structural RNAs such as rRNAs, tRNAs, misc_RNAs (miscellaneous RNAs), with nucleotide spans (if known)
- features which may describe your sequence, such as repeat regions, UTRs, promoters with nucleotide spans

The addition of feature annotation to your sequence submission:

- Improves the quality of your submission

- Increases the efficiency with which your submitted sequences are processed by members of the GenBank staff
- Is of far greater use to the scientific community than sequence data alone.

Adding feature annotation will also provide an additional tool for reviewing the quality of primary nucleotide sequence data.

For example, annotating protein-coding regions will frequently highlight potential errors in the nucleotide sequence, such as insertion/deletions (in/dels) or improper or uncertain base calls that result from the sequencing reads.

See an alphabetic list **of available Features** in the Sequin Help documentation (These Features can be used in both Sequin and BankIt).

See the "Annotation using BankIt" and "Annotation using Sequin" sections of this Quick Start for information about how to annotate your sequences.

Feature Annotation Using BankIt

How do I annotate features in my submission using BankIt?

1 **At step 8 ("Features") of the BankIt submission process, you will choose between:**
   a. Uploading a 5-column, tab-delimited table file containing your sequence features (select: "File" button)

      OR
   b. Picking feature categories and feature types for your sequence from a list provided in an online BankIt form (select: "Form" button).

Benefits of using File data upload:

- Good for different features on multiple sequences
- Helpful for adding many multiple features on a single sequence or on multiple sequences
- Uses the five-column, tab-delimited feature table format
- Multiple tables can be uploaded in a single file

Benefits of using Form input:

- Good for a single feature or a few features applied to a single sequence
- Good for applying a single feature to all sequences in a set or batch, or for applying a few of the same features to all sequences in a set or batch
- Features can be added across an entire sequence or by specific intervals within a sequence
- One or more modifiers can be chosen to apply to each feature
   2. **If you select the File upload button**:
      a. **Click the "Browse" button and select the feature table .txt** file you would like to upload.
      b. **Click the "Upload File" button** and upload your file.
      c. **Find the "Current Features" section of the Features (Overview) page**, where you will see a list of the features created from the table you just uploaded. Next to each feature on the list are buttons that allow you to either edit a feature or remove it entirely.
      d. Scroll to the bottom of the "Current Features" section and **click "Continue" to go to the next step of the submission process** once the features have been entered to your satisfaction.
      e. Features can be edited or deleted before continuing to the Review and Submit steps.

3. **If you select the "Form" button:**
   a. **Select one category from the five general feature categories presented:**
      - CDS/gene/mRNA
      - structural RNAs
      - Gene
      - Repeat Region (for simple repeats, mobile elements, and satellites)
      - Other (e.g.: D-loop, misc_feature, polyA_site, variation, etc.)
   b. **Select the appropriate feature type if presented with a choice** after selecting the feature category.
   c. **Click the "Add" button**. On the new page that appears, provide the specific details for the feature (e.g. nucleotide intervals, protein name, rRNA name, gene name, etc.).
   d. **Click the "Accept" button** at the bottom of the page.
   e. **On the "Features (Overview)" page, find the "Current Features" section**. You will see a list of the features created using the data you provided. Next to each feature on the list are buttons that allow you to either edit a feature or remove it entirely before continuing to the Review and Submit steps.
   f. **Go to the bottom of the "Current Features" section of the page, and click "Continue"** to go to the next step of the submission process once the features have been entered to your satisfaction.

Annotation of Coding Regions using BankIt

How do I add annotation for coding regions in my submission using BankIt?

The easiest way to add annotation for coding regions in your submission is to:

- Provide the coding region spans when prompted by your submission program to do so; **the submission toolwill automatically translate the feature span for you.** (we prefer this means of generating the translation)

but, you can also:

- Import the amino acid sequence and ask the submission program to predict the coding region spans for you.
  - Once you import the protein sequence, BankIt will process the translation for you, and will inform you if there are errors in the translated intervals. If you are notified of errors, you will then be prompted to make corrections to your sequence before proceeding with your submission.
  - Although we will accept this means of generating a translation, **we would prefer that you submit the span information**.

Do I have to submit the translated sequence when I annotate my submission using BankIt?

Generally, **there is no need to provide the translations yourself since**:

- **If you provide the span information** for the feature when prompted to do so during your submission, the submission tool will automatically translate the feature span for you.
  - This is easiest way to add annotation for coding regions in your submission.
  - **We prefer this means of generating the translation.**
- **If you do not have span information**, you can import the protein and have the submission program translate the protein to get the information for the feature you wish to annotate.

- Once you import the protein sequence, BankIt will process the translation for you, and will inform you if there are errors in the translated intervals. If you are notified of errors, you will then be prompted to make corrections to your sequence before proceeding with your submission.
- Although we will accept this means of generating a translation, **we would prefer that you submit the span information**.

Feature Annotation Using Sequin

How do I annotate features in my submission using Sequin?

As Sequin has a number of annotation options, the answer to this question depends on the nature of the annotation you wish to add to your sequence:

- **If you are submitting a single sequence and wish to annotate a single** feature **to it:**
  You can either enter them when prompted by Sequin during the submission process, or you can use the "Record Viewer Annotate Menu", which will provide you with a list of annotation options.

For more information on how to use the Record Viewer Annotate Menu, see the Features section of the Sequin Help Documentation.

- **If you are submitting a set of similar sequences, and want to annotate the same feature across the entire span of each:**
  Use the Batch Feature Apply option *if the feature you wish to annotate spans the entire nucleotide sequence of each member of the set. You cannot annotate specific nucleotide locations using this option.*

For more information about the Batch Feature Apply option, see the Annotate Menu section of the Sequin Help Documentation.

- **If you are submitting an aligned set of sequences, and want to annotate the same feature that you added to one sequence to all the other sequences within the set:**
  Use the Feature Propagate option. For more information about the Feature Propagate option, see the Feature Propagate section of the Sequin Help Documentation.
- **If you are submitting an aligned set of sequences, and want to annotate** features to each sequence in the set:
  Use the Alignment Assistant. For more information about the Alignment Assistant, see the Alignment Assistant section of the Sequin Help Documentation.
- **If you wish to annotate features to sequence records in Sequin, but prefer not to use the options mentioned above:**
  You can create a five-column, tab-delimited feature table and import it into Sequin.

See the step-by-step instructions for making a tab-delimited table in this Quick Start.

See the "Submission of Annotation Using a Table" page of the Sequin help documentation for additional information about the use of annotation tables in Sequin.

Annotation of Coding Regions using Sequin

How do I add annotation for coding regions in my submission using Sequin?

The easiest way to add annotation for coding regions in your submission is to:

- **Provide the coding region spans** when prompted by your submission program to do so; the submission tool will automatically translate the feature span for you. (we prefer this means of generating the translation)

but, you can also:

- Import the amino acid sequence and ask the submission program to predict the coding region spans for you.
  - If you choose to proceed using this means of generating a translation, **you must validate your submission to make sure the predicted spans are correct before you submit**. Please check any validation warnings generated by Sequin, and correct any errors that you find.
  - Although we will accept this means of generating a translation, we would prefer that you submit the span information.

Do I have to submit the translated sequence when I annotate my submission using Sequin?

Generally, **there is no need to provide the translations yourself since**:

- **If you provide the span information** for the feature when prompted to do so during your submission, the submission tool will automatically translate the feature span for you
  - This is easiest way to add annotation for coding regions in your submission.
  - **We prefer this means of generating the translation.**
- **If you do not have span information**, you can import the protein and have the submission program translate the protein to get the information for the feature you wish to annotate.
  - If you choose to proceed using this means of generating a translation, **you must validate your submission to make sure the predicted spans are correct before you submit**. Please check any validation warnings generated by Sequin, and correct any errors that you find.
  - Although we will accept this means of generating a translation, we would prefer that you submit the span information.