



The "Submission Category" Page

Michael Fetchko and Adrienne Kitts

Purpose

The BankIt submission tool "Submission Category" page is where you confirm that you actually sequenced the data you are submitting. If you did not sequence the data you are submitting but want to create Third Party Annotation (TPA) submission, you will be asked to provide:

- A description of the biological experiments or other work used as evidence for the new annotation in your TPA submission
- The GenBank Accession number(s) of the **primary** sequence(s) from which your TPA submission was derived.

Original Submissions

The pre-set answer of "Original" indicates that the sequence is an original submission that was sequenced directly by the submitter. If you leave the "Original" button selected and click the "Continue" button at the bottom of the page, you will proceed directly to the next page of the BankIt form.

What is an Original Submission?

All of the following are considered original submissions:

- Data sequenced directly by the submitter

Note: If your sequence is identical to an existing GenBank sequence record, your sequence is still considered original, and must be submitted as new data to GenBank. It will have its own accession number and will be distinct from the sequence record that already exists.

- Synthetic sequences
 - a whole cloning vector that you designed
 - codon-optimized genes/coding sequences for use in specific organisms
- Sequence amplified using PCR primers derived from other sequences

Third Party Annotation (TPA)

If you click the "Third Party Annotation" button, a submission form for the **Third Party Annotation (TPA) sequence database** will appear. A TPA submission adds new feature annotation for primary sequences (i. e. sequences available in GenBank that already have GenBank Accession numbers) that is supplied by experimental or inferential evidence.

A TPA sequence must be built from primary sequence data available in GenBank and identified by a GenBank accession number.

There are specific standards that your data must meet to be included in the TPA database, so read through the information on the [TPA](#) web site carefully to be sure that your data meets these standards before you submit.

Evidence

Provide text describing the evidence that supports your new annotation to the primary sequence.

This evidence can be:

- **Experimental:** Your annotation is supported by wet lab evidence published in a peer-reviewed scientific journal

OR

- **Inferential:** Your annotation is inferred from other work you did, but this work was not by direct experimentation. The supporting information for your inferred annotation must be published in a peer-reviewed scientific journal.

GenBank Accessions

Provide a file of all the GenBank Accession numbers for the primary sequences that you used to build or derive your TPA sequence. The online BankIt Help documentation includes [detailed instructions](#) for creating a TPA file of primary accession numbers.

Primary sequence data includes Whole Genome Shotgun (WGS) data or Trace Archive data, but NOT Reference Sequence (RefSeq) data or data from the [CON](#) (Contig) division of GenBank (you will see the word “CON” in the locus line of these records), since RefSeq and CON data are not primary sequence data.

Common Mistakes Made While Filling Out the “Submission Category” Page

- **Mistake: Entering an incorrect GenBank Accession number**

Fix: If you get an error message saying that one of your GenBank Accessions is invalid, **make sure that the accessions were typed correctly and that:**

- **The accession is NOT for a Reference Sequence (RefSeq) record**

RefSeq accession numbers can be distinguished from GenBank accessions by their format of 2 alphabetic characters followed by an underscore character ('_') and then a series of numbers. For example, a RefSeq mRNA accession is NM_123456.

- **The accession is NOT for a CON (Contig) division record** (you will see the word “CON” in the locus line of these [records](#)).

Con division records are already built from other primary sequence(s) and therefore cannot be cited as a primary sequence)