



Submitting Different Sequence Types using Specific NCBI Submission Resources

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

Submission to the Transcriptome Shotgun Assembly Sequence database (TSA)

Can I submit a Transcriptome Shotgun Assembly (TSA) that I have assembled out of sequences I found in dbEST, the Sequence Read Archive, or the Trace archive?

No, you can submit a [TSA](#) sequence only if you have experimentally determined the primary sequences used to assemble the TSA sequence yourself.

How do I submit Transcriptome Shotgun Assembly (TSA) Sequences?

Complete details of the TSA submission process are available on the [TSA home page](#).

Submission to the High-Throughput Genomic (HTG) sequence division of GenBank

How do I submit to GenBank via the High-Throughput Genomic (HTG) sequence division?

Complete discussion of the HTG division of GenBank is available on the [HTG home page](#). HTG submission instructions are available on the [HTG submission page](#).

The HTG submission system is designed for high-throughput bulk submissions, typically of BAC clone sequences from genome centers. An automated system processes the submissions daily and releases them to GenBank the same day if there are no errors.

- **If you are only sequencing a few BAC clones do not use the HTG system** – submit using the standard [GenBank submission pathway](#).
- The HTG submission system releases all submissions immediately after processing – you cannot set a release date in advance. **If you need to set a release date for your submission, you must submit using the standard [GenBank submission pathway](#).**

If you have questions or are unsure of how to proceed, we recommend you contact htgs-admin@ncbi.nlm.nih.gov for advice before you begin your submission.

Below is a brief overview of what to expect in the HTG submission process:

Before you begin the HTG submission process, review the [HTGS Getting Started page](#) to see assumptions that NCBI has for HTG submitters and their data.

1. **Sequencing center or group must contact htgs-admin@ncbi.nlm.nih.gov and request an FTP account**, where you will transfer your sequence records once they are prepared for submission.

2. Prepare your sequence for HTG submission:

- a. An HTG submission must be in ASN.1 format.
- b. There are two NCBI tools available for creating an ASN.1 formatted HTG submission:
 - i. [Sequin](#)
Sequin contains a setting that allows genome centers to prepare HTG submissions. The user will import a [FASTA](#) formatted sequence file and a [Sequin submission template file](#) (containing contact and citation information) into Sequin, and will then enter annotation for the sequence into the Sequin form. Sequin will then generate the ASN.1 file for HTG submission.
Please see the “Formatting Sequences in FASTA Format” section of the HTG [tbl2asn](#) instructions for more instructions on creating a modified FASTA file that will define segments and gaps.
Please see the “[Using Sequin to prepare a HTG Submission](#)” page for detailed “how to” information.
 - ii. [tbl2asn](#)
Tbl2asn is a command line program that is downloaded to the user's computer. Once tbl2asn is installed, the user runs the tbl2asn software by typing commands that will generate an ASN.1 file for submission from a [FASTA](#) sequence file and a [Sequin submission template file](#) (containing contact and citation information).

Please see the “Formatting Sequences in FASTA Format” section of the HTG [tbl2asn](#) instructions for more instructions on creating a modified FASTA file that will define segments and gaps.

The advantage of using tbl2asn over Sequin for creating HTGS submissions is that it can be set up by the user to create submissions in bulk from multiple files.

Please see the [tbl2asn](#) section of this Quick Start for step-by-step instructions for using tbl2asn. You can find specific instructions for the HTG submission arguments required for HTG submissions on the [HTG page](#) .

- c. Make sure that you include the genome center name and the sequence name in your submission; NCBI will issue the accession number. These three identifiers are required for every update, as described on the [HTGS Getting Started page](#).
3. **Submit your HTG sequence submission ASN.1 files to NCBI into the SEQSUBMIT directory of the FTP account you requested in step 1.**
4. **Note:** HTG sequences submitted to the FTP site are processed on a daily basis.

Submission of Full Length Insert cDNA (FLIC) Submissions

How do I submit Full Length Insert cDNA (FLIC) sequence data to GenBank?

FLICs are processed via an automated FLIC processing system. **Follow the steps below to submit FLIC(s) to GenBank:**

1. To submit sequences in bulk to the FLIC processing system, a center or group must **contact [gb-admin@ncbi.nlm.nih.gov](mailto:contactgb-admin@ncbi.nlm.nih.gov) and request an FTP account for FLIC submissions.**
2. **Use the [tbl2asn](#) program to generate your submission(s)**
Submissions to the FLIC processing system must contain the following identifiers:
 - The genome center tag
(assigned by NCBI and is generally the FTP account login name)

- The sequence name (SeqID)
(a unique identifier that is assigned by the submitter to a particular clone or entry and must be unique within the group's FLIC submissions)
- 3. **Deposit your submission(s) in the FLICSEQSUBMIT directory of your FTP account** and contact gb-admin@ncbi.nlm.nih.gov to let us know your submission(s) is available for processing.
- 4. **Your submission(s) will be processed and assigned an accession number.** The files will be automatically loaded into GenBank if there are no errors in the submission. FLIC submissions cannot be held confidential.
- 5. **Should your submission fail FLIC processing,** you will receive an email from us that describes the problem. Once you have reviewed our email, you will need to submit a corrected entry.
- 6. At the completion of processing, a submission report is automatically generated and deposited in your FTP account.
- 7. All updates to your FLIC submission(s) must include the center tag, sequence name, and accession number, or processing will fail.

Genome Submissions

Which GenBank resource do I use to submit an incomplete genome that was sequenced using clone-based sequencing, and which resource do I use to submit one that was sequenced using the Whole Genome Shotgun (WGS) approach?

Genome submissions are comprised of genomic DNA sequences representing either complete or incomplete genomes from both prokaryotes and eukaryotes. Viral, phage or complete organellar sequences should be submitted as regular GenBank records. In addition, data consisting of only a subset of genes (eg 16S ribosomal RNA or "your gene of interest"), should be submitted as regular GenBank submissions. Unassembled sequences from next-generation sequencing platforms should be submitted to the Sequence Read Archive ([SRA](#)).

WGS genomes should be submitted to the [Whole Genome Shotgun \(WGS\) division](#). The clones (usually BACs) of genomes sequenced using traditional clone-based sequencing should be submitted to:

- The High-Throughput Genome (HTG) division of GenBank if they do not need to be confidential

OR

- GenBank if they do need to be kept confidential.

Whole Genome Shotgun Submissions

How do I submit a genome to GenBank via the Whole Genome Shotgun (WGS) division?

Complete submission details for the WGS division of GenBank are available on the [WGS submission page](#). If you are unsure about the type of data submitted to the WGS division, visit the [WGS List](#) for example projects. You must register your project with the [BioProject](#) and [BioSample](#) databases. For further details on how to create a WGS submission, see the "How to Submit WGS Genomes" page of the WGS site".

Complete Genome Submissions

How do I submit a complete genome to GenBank?

Details details for complete, prokaryotic genomes are available in the [Prokaryotic Annotation Guide](#). You must register your project with the [BioProject](#) and [BioSample](#) databases.

Metagenome Submissions

How do I submit a Metagenome to GenBank?

Metagenomics is the culture-independent genomic analysis of a community of microorganisms. Complete details for submitting Metagenomes to GenBank are available from the [Metagenomes Submission Guide](#). You must register your project with the [BioProject](#) and [BioSample](#) databases.

Submission of Third Party Annotation Records

What is the Third Party Annotation (TPA) database, and what kind of data can be submitted to it?

The Third Party Annotation (TPA) database contains nucleotide sequences that are derived or assembled from primary sequence data that is housed in other International Nucleotide Sequence Collaboration (INSDC) databases (TPA is part of INSDC).

- **The INSDC databases (DDBJ,EMBL, andGenBank) containprimary sequence data** and corresponding annotations submitted by the laboratories that completed the original sequencing.
- **The TPA subset database contains nucleotide sequences built from the existing INSDC primary sequence data** that has new feature annotation described in a peer-reviewed scientific journal.

Each TPA record can be one of two types:

- **Experimental:** annotation is supported by wet-lab evidence published in a peer-reviewed scientific journal
- **Inferential:** annotation is inferred from other work by the submitter, but is not the subject of direct experimentation itself. Supporting information for the inferred annotation has been published in peer-reviewed scientific journal(s).

For additional information about what constitutes an Experimental or an Inferential TPA record, see the [TPA FAQ](#).

A brief list of TPA sequence submission examples can be found on the [TPA home page](#) (scroll down a little to see the list) as can a list of sequences/annotation that should not be submitted to the TPA (scroll to bottom of page).

How do I submit annotation to TPA?

You can submit sequence and new annotation to the TPA database using either BankIt or Sequin, but be sure to adhere to the following guidelines regardless of which method you use to submit to TPA:

- The entire sequence submitted to TPA must be derived or built from primary sequence data.
- There is no limit on the number of overlapping/adjoining primary sequences that can be cited for a TPA submission.
- If sections of a sequence submitted to TPA have been newly determined by the submitter, those sections of sequence (if they are more than 50 nucleotides) must first be submitted to GenBank, processed, and released to the public before they can be cited as primary sequences for TPA.
- Each TPA sequence must cite the same organism as the primary sequence data used to build or derive it.

To submit to TPA using BankIt:

1. Fill out or verify your contact information on the “Contact” page, and then use the “Reference” and “Nucleotide” pages enter your publication information and your Nucleotide data.
2. On the Submission Category page, choose Third Party Annotation.

3. Follow the instructions and enter:
 - a. A brief explanation of the work done as evidence to support the new feature annotation.
 - b. The GenBank/INSDC accession numbers of all primary data used to assemble or derive the sequence.
4. Be sure to add all new feature annotation on the Features pages.
5. Confirm your submission on the “Review and Correct” page and click the “Finish” button to submit your sequence and annotation.
6. The submission will be labeled as a TPA record and processed accordingly after it is successfully submitted.

To submit to TPA using Sequin:

1. Begin the Sequin submission process as directed, and complete the “Submitting Authors” form. When you complete the final (Affiliation) section of the “Submitting Authors” form, click the “Next Form” button at the bottom of the page. You will go to the “Sequence Format” form.
2. Select the radio button that marks “Third Party Annotation” as your selection from the submission category section in the Sequence Format window.
3. Provide a brief explanation of the work done as evidence to support the new feature annotation in the TPA Evidence box.
4. Enter the GenBank/INSDC accession numbers of all primary sequences used to assemble or derive the TPA sequences into the “Assembly Tracking” box that appears with the flatfile display following the Annotation page.
5. Click on Accept; a new COMMENT field will appear in the flatfile, which will list the primary sequence Accession Numbers.
6. Complete your submission as directed by Sequin.
7. When you have completed the submission process, you must email the .sqn submission files generated by the Sequin program to gb-sub@ncbi.nlm.nih.gov since Sequin does not automatically transmit the completed file for you at the end of the Sequin process (a dialog box will appear at the end of the Sequin submission process instructing you to email your submission files).
8. Include a note in the email that contains the .sqn file that the submission is intended for TPA.

Submission to the database of Expressed Sequence Tags (dbEST)

What are ESTs and where do I submit them?

GenBank defines Expressed Sequence Tags (ESTs) as short (300-500 bp) single reads from mRNA (cDNA) that are usually produced in large numbers. They represent a snapshot of what is expressed in a given tissue, and/or at a given developmental stage. They also represent tags (some coding, others not) of expression for a given cDNA library.

You can submit ESTs to [dbEST](#).

In addition to short sequence reads, **dbEST also includes** sequences that are longer than the traditional ESTs or are produced as single sequences or in small batches — including **products of differential display experiments and RACE experiments**.

dbEST is reserved for single-pass reads. **Assembled sequences should not be submitted through dbEST**, but through the [TSA](#) (Transcriptome Shotgun Assembly) division of GenBank. See the “Starting a Submission to TSA sequence database” section of this Quick Start for more information on submitting to TSA.

How do I submit to dbEST?

1. All EST sequences must be submitted using the custom streamlined submission procedures as described on the [dbEST submission page](#). The file format used to submit to dbEST is outlined on the [file format page](#), where you will find detailed examples of this format in use following each file format description.
2. Information (if applicable) you will be required to provide for your EST submissions will include:
 - Clone name
 - Clone library [catalog number, reference, lab source, and/or specific (in-house) name or number]
 - Tissue type
 - Developmental stage
3. Generally speaking, no annotation is expected in an EST record regardless of length, quality, or quantity of sequence submitted.
4. Once you have completed your submission files as described in the dbEST submission documentation, you will send them to: batch-sub@ncbi.nlm.nih.gov either attached to a single email message, or you can include the files in the body of the email message. Be sure that the files are in plain text (ASCII) format.

dbEST offers a specified release date option for your data should you require it. See the “Assignment of GenBank Accession Numbers and Release of Data” section of the [dbEST submission page](#) for more details.

Submission to the database of Genome Survey Sequences (dbGSS)

What are GSSs and where do I submit them?

dbGSS contains (but is not limited to) genomic sequences produced from the following:

- Random "single pass read" genome survey sequence experiments.
- Single pass reads of cosmid/BAC/YAC ends (these may or may not be chromosome specific)
- Exon trapping experiments
- Alu PCR experiments

How do I submit Genome Survey Sequences?

1. **All GSS sequences must be submitted using the custom streamlined submission procedures as described on the [dbGSS submission page](#).** The file format used to submit to dbGSS is outlined on the [file format page](#), where you will find detailed examples of this format in use following each file format description.
2. **Generally speaking, no annotation is expected in a GSS record** regardless of length, quality, or quantity of sequence submitted.
3. Once you have completed your submission files as described in the dbGSS submission documentation, send them to: batch-sub@ncbi.nlm.nih.gov either attached to a single email message, or you can include the files in the body of the email message. Be sure that the files are in plain text (ASCII) format.

dbGSS offers a specified release date option for your data should you require it. See the “Assignment of GenBank Accession Numbers and Release of Data” section of the [dbGSS submission page](#) for more details.