



Before Starting the Submission Process

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

Choosing the Appropriate Submission Resource

Submission Tools

BankIt vs. Sequin vs. tbl2asn

When should I use BankIt for submissions? When should I use Sequin or tbl2asn?

You should use [BankIt](#) if:

- You prefer to use a web-based submission tool
- You do not require advanced sequence analysis tools

You should use [Sequin](#) if:

- You prefer to work on your submission off-line
- You would like graphical viewing and editing options, including an alignment editor
- You would like the option to have network access to related analytical tools
- You are submitting files containing less than 10,000 sequences. If you have more than 10,000 sequences, you must submit multiple files or use [tbl2asn](#).

You should use [tbl2asn](#) if:

- Your sequence has a lot of annotation
- You are submitting a large [batch](#) of sequences
- You have Whole Genome Shotgun (WGS) submissions or Transcriptome Shotgun Assembly (TSA) submissions
- You have complete genome submissions
- You are submitting FLIC sequences

Once you have decided which of these tools you'd like to use for your submission, see the "[Submitting Sequences using Specific NCBI Submission Tools](#)" section of this Quick Start guide for brief explanations of each of the different submission processes as well as links to useful material.

See the [GenBank Sample Record](#) page, which provides all GenBank Record field definitions.

Expressed Sequence Tags

Which submission tool should I use if I want to submit Expressed Sequence Tags (ESTs)?

ESTs should be submitted through the database of Expressed Sequence Tags ([dbEST](#)) submission system.

You may wish to look at the “[Submitting Sequences using Specific NCBI Submission Tools](#)” section of this Quick Start guide for a brief description of the dbEST submission process, and links to useful material.

I have computationally assembled mRNA sequence reads from primary data such as ESTs, traces, and Next Generation Sequencing Technologies. Where do I submit my assembly?

You can submit your assembly to the Transcriptome Shotgun Assembly (TSA) Sequence Database using the [process](#) described on the [TSA home page](#).

Genome Survey Sequences

Which submission tool should I use if I want to submit Genome Survey Sequences (GSSs)?

GSSs should be submitted through the database of Genome Survey Sequences ([dbGSS](#)) submission system.

dbGSS contains (but is not limited to) genomic sequences from the following types of data:

- Random "single pass read" genome survey sequences
- Single pass reads from cosmid/BAC/YAC ends (these may or may not be chromosome specific)
- Exon trapped genomic sequences
- Alu PCR sequences

You may wish to look at the “[Submitting Sequences using Specific NCBI Submission Tools](#)” section of this Quick Start guide for a brief description of the dbGSS submission process, and for links to useful material.

Barcode Sequences

What are “Barcode” sequences, and where are they submitted?

Barcode sequences, determined as part of the Barcode of Life initiative, are short nucleotide sequences from a standard genetic locus for use in species identification. Currently, the Barcode sequence being accepted for animals is a 5', 658 base pair region of the mitochondrial cytochrome oxidase subunit I (COI) gene.

The [Barcode Submission tool](#) provides for streamlined online submission of Barcode sequences into GenBank. With this tool, you can:

- submit new Barcode sets
- complete your most recent incomplete submission
- download a flat file summary of completed submissions

First-pass sequence data generated from a single cosmid, BAC, YAC, or PAC clone

Where do I submit unfinished DNA (e.g. first-pass sequence data generated from a single cosmid, BAC, YAC, or PAC clone) sequences?

An unfinished collection of DNA sequence data derived from a single cosmid, BAC, YAC or PAC clone that may contain one or more gaps, can be submitted to the [High Throughput Genomic \(HTG\) sequence division of GenBank](#). The HTG division contains unfinished high-throughput clone-based DNA sequences that are available in GenBank and for BLAST similarity searches against the "HTGS" database.

A single accession number is assigned to sequence data generated from a single clone; each HTG record provides a user with the HTG sequence status and a flag that the sequence data are "unfinished" and may contain errors.

If you want to submit data to the HTG division of GenBank, review the [HTG submission documentation](#), as well as [HTG FAQs](#), both of which can be accessed using links located on the [HTG home page](#). There is a [brief overview](#) of the HTG submission process in this Quick Start as well.

Note: 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 would like more information about submitting to the HTG division of GenBank, contact the HTG division at: htgs-admin@ncbi.nlm.nih.gov.

Flatfiles generated using a non-NCBI tool

Can I submit a flat file to GenBank created using a NCBI tool if the file generated looks just like Sequin output or looks just like a GenBank flat file?

GenBank cannot accept flat files created using non-NCBI tools for the following reasons:

- **We cannot accept the flat file format** (even if it is made in Sequin) since the flat file format is a display format only.
- **Sequin is not able to interpret GenBank-style files generated by outside tools** since GenBank requires ASN.1 formatting for proper field specification of features and other elements in a submission, which outside tools are unable to provide.

Please submit properly formatted ASN.1 files.

Learn about GenBank Records before you Submit

How do I find definitions for all the fields in a GenBank Record so I know what they are before I begin the GenBank submissions process?

To explore the definitions the fields in a GenBank Record before you start your submission, **go to our online example of a GenBank record**, and click on any of the light blue links to see the field definitions.

What's the difference between a GenBank record and a RefSeq record?

- **A GenBank record represents** primary sequence data supplied by the original submitter, who has editorial control over that record's data.
- **Reference Sequence (RefSeq) records** are derived from GenBank records, but differ from them in that each RefSeq is a curated synthesis of information about a particular sequence, rather than an archived unit of primary research data like the records in GenBank.
- The RefSeq database aims to provide a non-redundant, well-annotated set of curated sequences to be used as stable references for annotation and for various studies and analyses.
- A RefSeq record will cite the accession numbers of the original GenBank records from which it was derived.
- **RefSeq records may be altered by NCBI staff as needed** to incorporate additional sequence or annotation information. In addition, changes to an original GenBank record by its submitters may be incorporated by NCBI staff into a RefSeq record.

The complete [original publication](#) describing RefSeq is available on PubMed Central.