



## What Kind of Data Can be Submitted to GenBank?

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

### What kind of data will GenBank accept?

GenBank is a nucleotide sequence database and will accept primary sequence data that was directly determined by the submitter.

**Below are examples of submission types included in GenBank:**

- mRNA Sequences
- Prokaryotic Genes
- Eukaryotic Genes
- rRNA and/or ITS
- Viral Sequences
- Transposon or Insertion Sequences
- Microsatellite Sequences
- Pseudogenes
- Cloning Vectors
- Phylogenetic or Population Sets
- Non-coding RNAs

**The following submission types are accepted by GenBank, but should be submitted using their own submission tools** (see below):

- Expressed Sequence Tags (**EST**) should be submitted directly to [dbEST](#) (the EST division of GenBank)
- Genome Survey Sequences (**GSS**) should be submitted directly to [dbGSS](#) (the GSS division of GenBank)
- Transcriptome Shotgun Assembly (**TSA**) should be submitted directly through the [submission portal](#) according to these [directions](#).

If your submission does not fall into one of the above categories, contact [info@ncbi.nlm.nih.gov](mailto:info@ncbi.nlm.nih.gov) to determine which NCBI resource would be most appropriate for your submission.

**For help beginning the submissions process** to GenBank, see the “[Submitting Sequences using Specific NCBI Submission Tools](#)” section of this Quick Start.

### What kind of data will GenBank NOT accept?

**The following submission types are not accepted** by GenBank:

- Sequences <200 bp long. Unassembled sequences from next-generation sequencing platforms should be submitted to the Sequence Read Archive ([SRA](#))

- A genomic sequence of multiple exons joined together without the sequence of the intervening introns or without a 'gap' of internal nns representing the missing sequence
- Primer only sequences (These sequences can be submitted directly to NCBI's [Probe database](#))
- Protein only sequences
- Sequences containing a mix of genomic and mRNA sequence represented as a single sequence
- Sequences without a physical counterpart (consensus sequences)

**For help beginning the submissions process** to GenBank, see the “[Submitting Sequences using Specific NCBI Submission Tools](#)” section of this Quick Start.

### **Can I submit a sequence contig to GenBank?**

**The answer to this question depends upon the sequence contig you intend to submit.**

- Sequence contigs assembled from sequence already present in International Nucleotide Sequence Database Collaboration ([INSDC](#)) sites should be submitted to NCBI's Third Party Annotation sequence database ([TPA](#)).
- Sequence contigs assembled from nucleotides that you have sequenced yourself and assembled using sequence overlap can be submitted to GenBank. If there are gaps in your contig assembly, they must be filled with internal nns that represent any missing sequence.
- Computer-derived mRNA assemblies should be submitted to [TSA](#).

For help beginning the submissions process to GenBank, see the “[Submitting Sequences using Specific NCBI Submission Tools](#)” section of this Quick Start.

### **How do I submit a large number of cosmid, BAC, or YAC derived genomic clones to GenBank?**

The best way to submit a large number of cosmid, BAC, or YAC derived genomic clones to GenBank is to submit through our [High-Throughput Genomic \(HTG\) sequence division](#).

The HTG division contains unfinished high-throughput DNA sequences that are available for BLAST similarity searches against the "HTGS" database.

**Note:** Sequences submitted via the HTG automated system are made public immediately and cannot be held for release at a later date.

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](mailto:htgs-admin@ncbi.nlm.nih.gov).

### **Can I submit protein-only sequences to GenBank?**

**GenBank** is a nucleotide sequence database and therefore **does not accept protein-only sequence submissions**.

**If you do not have nucleotide sequence(s) for your protein(s)**, but would still like to submit your directly sequenced protein to a public database, see [The Universal Protein Resource \(UniProt\)](#).

### **Can I submit primer sequences to GenBank?**

GenBank does not accept primer sequences, but you can submit primers to NCBI's [Probe database](#), which is a public registry of nucleic acid reagents designed for use in a wide variety of biomedical research applications.

The Probe database also includes information on reagent distributors and probe effectiveness, as well as computed sequence similarities. The Probe database [submission documentation](#) provides a simple overview of the required information you will need, as well as some basic rules to follow during your submission.