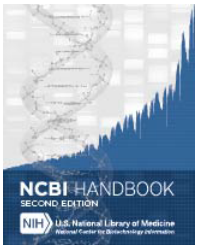
U.S. National Library of Medicine
National Center for Biotechnology Information

NLM Citation: Mizrachi I. GenBank. 2013 Nov 12. In: The NCBI Handbook [Internet]. 2nd edition. Bethesda (MD): National Center for Biotechnology Information (US); 2013-.
Bookshelf URL: https://www.ncbi.nlm.nih.gov/books/



GenBank

Ilene Mizrachi[✉1]

Created: November 12, 2013.

Scope

The GenBank sequence database is an annotated collection of all publicly available nucleotide sequences and their protein translations. This database is produced at the National Center for Biotechnology Information (NCBI) as part of an international collaboration with the European Molecular Biology Laboratory (EMBL) Data Library from the European Bioinformatics Institute (EBI) and the DNA Data Bank of Japan (DDBJ). GenBank and its collaborators receive sequences produced in laboratories throughout the world from hundreds of thousands of distinct organisms. GenBank continues to grow at an exponential rate, doubling every 18 months. Release 197, produced in August 2013, contains over 154 billion nucleotide bases in more than 167 million sequences. GenBank is built by direct submissions from individual laboratories and from large-scale sequencing centers.

Submissions to GenBank include a number of data types: single gene or mRNA sequences, phylogenetic studies, ecological surveys, complete genomes, genome assemblies (WGS), transcriptome assemblies (TSA), and third-party annotation (TPA). In the past, transcript surveys (EST), genome surveys (GSS), and high-throughput genome sequences (HTGS) constituted a significant fraction of submissions, but with the emergence of next-generation sequencing technologies, we have seen a steady decrease in these data types. Submissions to GenBank are prepared using one of a number of submission tools and transmitted to NCBI. Upon receipt of a submission, the GenBank staff reviews the records, assigns an accession number and performs quality-assurance checks prior to release of the data to the public archive. Sequence records once released are retrievable by Entrez, searchable by BLAST, and downloadable by FTP.

Author Affiliation: 1 NCBI; Email: mizrachi@ncbi.nlm.nih.gov.

✉ Corresponding author.