Quick Overview:
Public Databases for Genomic Data

GREEN GATE GENOMICS

NCBI National Center for Biotechnology Information

ENA European Nucleotide Archive

DDBJ DNA Data Bank of Japan

Depicted logos are trademarks of the National Center for Biotechnology Information, the European Molecular Biology Laboratory and the Japanese National institute of Genetics

International Sequence Data Collaboration (INDSC)

	Standard Sequences	Next-Gen sequencing reads	High Quality curated DNA and protein record
What type of genomic data do I find in which archive?	- GenBank – US Sequence Database at NCBI - European Nucleotide Archive (ENA) at EMBL-EBI - Data Databank for Japan (DDNJ) at NIG	Sequence Read Archive (SRA)	- NCBI Reference Sequences (RefSeq) - UniProt/Swiss-Prot - Pfam - Others

Important NCBI Databases

GenBank: is a comprehensive database of nucleotide sequences, including DNA and RNA sequences. It provides access to both annotated and unannotated sequences, along with associated metadata such as sequence features, organism information, and literature references. It includes **complete genome assemblies, gene sequences, mRNA transcripts**, and other types of genetic data.

Sequence Read Archive (SRA): is a database for storing and sharing raw sequencing data generated by next-generation sequencing (NGS) technologies. It serves as a repository for sequencing reads produced by platforms such as Illumina, Ion Torrent, and PacBio.
SRA contains **raw sequencing reads in their original format, typically stored as FASTQ files**.

RefSeq: is a curated database of reference sequences for **genomes**, **transcripts, proteins, and other biomolecules**. It provides a comprehensive, high-quality collection of annotated sequences derived from a diverse range of organisms.

Important EMBL-EBI Databases

European Nucleotide Archive (ENA): The main ENA database serves as a comprehensive repository for **nucleotide sequences, including DNA, RNA, and viral sequences**. It provides access to annotated sequences, experimental metadata, and related information.

Metagenomics and Environmental Sequences (MGnES): MGnES is a specialized database within ENA dedicated **to metagenomic and environmental sequencing data**. It contains sequences from environmental samples, microbiome studies, and other metagenomic projects, along with associated metadata and analysis tools.

Assembly Archive: Assembly Archive stores genome assemblies and related data submitted by researchers. It includes **assembled genomes, scaffolds, contigs, and associated metadata.**

Need more information?
Contact us at info@greengategenomics.com

GREEN GATE GENOMICS

greengategenomics.com