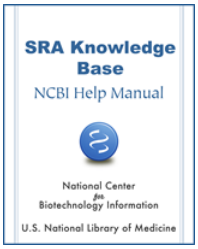
U.S. National Library of Medicine
National Center for Biotechnology Information

NLM Citation: Sequence Read Archive Submissions Staff. Problems Downloading Data. In: SRA Knowledge Base [Internet]. Bethesda (MD): National Center for Biotechnology Information (US); 2011-.
Bookshelf URL: https://www.ncbi.nlm.nih.gov/books/



Problems Downloading Data

Sequence Read Archive Submissions Staff

FTP Download Quits before Complete

I am trying to download files from the SRA, but the FTP process quits before the files are completely downloaded.

Sometimes this can be a temporary problem; wait and try your download again in an hour or two. If you continue to have problems, here are a few suggestions:

- If you are using an internet browser as your transfer client, you may want to consider a dedicated FTP client. There are many free FTP clients like FileZilla, Cyberduck, and NcFTP available.
- Do you regularly use FTP at your location without issue? If you are unable to access and download data from the Mozilla Foundation's FTP site, FTP traffic may be blocked or restricted by your institution.
- Consider using Aspera Connect to conduct your download since:
  - The amount of data for SRA projects can exceed 10 gigabytes and traditional FTP may be too slow to download your data effectively.
  - FTP performance degrades proportionally with the number of hops or switches the data must take to get to you. Aspera performance does not degrade with distance.

See the "Performance comparison of FTP and ascp downloads" section of this Knowledge Base for information on the differences between FTP and Aspera for SRA downloads.

No "Download Data" Link in Entrez Search Results

What if the data download link is missing from either an experiment or a run I found using Entrez?

If the download links are missing, then there is a technical issue with the data. Please contact sra@ncbi.nlm.nih.gov and provide us with the Experiment or Run accession.

No Download Link for Specific File Types

I want to download the data I've found in a particular format, but I only see a download link for .sra files.

The SRA archive format (".sra files") can be converted to several standardized file formats, including fasta, fastq, sam/bam, sff, ABI colorspace fasta/qual, and Illumina native. Please note that not all SRA data can be converted into all of the above data formats (this is determined by the submitted format: data submitted as fastq does not contain the information necessary to build an sff file, for example). Please see this page for more information. If you are only interested in fasta- or fastq- formatted data, please see the following question.

We store data in the SRA format due to storage space constraints — storing the data in different formats (.fastq, .sff, etc.) requires far more storage space than saving the data in a single archive format and allowing users to generate multiple different file types from the same source.

I am only interested in fasta- or fastq- format data. How can I download this directly?

From this page, you may enter an Experiment accession (SRX accession), or a comma-separated list of Experiment accessions, that will take you to a page like this on which you may specify (1) from which Run(s) you would like to download data, (2) the formatting (fasta or fastq), and (3) options relating to sequence filtering, trimming, and clipping. The latter page may also be reached through the Run Browser by selecting the "Reads" tab and then clicking "Filtered download" (leave the filter empty to request all reads).

fastq-dump outputs color space fastq containing numbers in the sequence reads instead of "normal" base space fastq.

Fastq files that have numbers instead of letters are in color space format. This format is a feature unique to the methodology used by the ABI SOLiD sequencer. To learn more about color space and its relationship to base space (ATCG) see the ABI webinar on the on the fundamentals of 2 Base Encoding and Color Space, and the ABI documentation "Color Space Analysis in the SOLiD™ System: the Theory, Advantages and Solutions".

In order to dump a "normal" base space fastq file, please run the fastq-dump command with the "-B" option. Conversely, to dump base space data as color space fastq, please use the "-C" option.