

## ***SRA-TOOLKIT on HPC***

### **What is SRA-TOOLKIT?**

SRA (Sequence Read Archive) Tools is a suite of command-line programs developed by the National Center for Biotechnology Information (NCBI) for downloading, converting, and analyzing data from the SRA database. The SRA database is a large repository of high-throughput sequencing data generated by various next-generation sequencing (NGS) platforms, including Illumina, PacBio, Oxford Nanopore, and others.

The SRA Tools allow users to access and download sequence data in its original format or convert it into commonly used file formats such as FASTQ or FASTA. The tools also provide functionality for quality control, filtering, and trimming of raw sequencing data.

In addition, the SRA Tools include programs for alignment and assembly of NGS data, as well as tools for visualization and analysis of sequencing data. The SRA Tools are widely used in the bioinformatics community for accessing and analyzing large-scale genomic data from the SRA database.

Links:

[Official Website](#)

[Wiki](#)

### **Versions Available:**

The following versions are available on the cluster:

- SRA-TOOLKIT 2.9.2

## How to load SRA-TOOLKIT?

To load SRA-TOOLKIT, use the following commands:

```
#Load the SRA-TOOLKIT module
module load bio/sratoolkit/2.9.2
```

To verify if the module is loaded correctly, use the following command,

```
# List all the module loaded in the environment
module list
```

If the module has internet connectivity issue, users have the option to use a singularity container to run sra-toolkit,

```
#Download prebuilt docker container through singularity
module load singularity/3.7.2
# pull sra toolkit image to cwd
singularity pull docker://ncbi/sra-tools

#run the sra-tool container
singularity exec ../sra-tools_latest.sif prefetch SRR000001
```

## How to use SRA-TOOLKIT?

The following website demonstrates few examples of using sra-toolkit,

[SRA examples](#)

Here are few example of using sra-toolkit,

### Example 1:

Downloading a subset of reads from an SRA dataset:

```
fastq-dump --gzip --skip-technical --readids --dumpbase --split-files  
--clip SRR000001
```

This command will download a subset of reads from the SRA dataset with accession number "SRR000001". The "--skip-technical" option skips any technical reads, such as those from adapters or primers. The "--readids" option includes read IDs in the output files. The "--dumpbase" option outputs base calls instead of quality scores. The "--split-files" option creates separate files for paired-end reads. The "--clip" option clips any low-quality bases from the ends of reads.

### Example 2:

Downloading data from a list of SRA accessions:

```
prefetch --option-file accession_list.txt
```

This command will download data from a list of SRA accessions listed in the file "accession\_list.txt". The prefetch tool downloads the data and saves it in a local cache for faster access.

### Example 3:

Generating a report on an SRA dataset:

```
sra-stat SRR000001
```

This command will generate a report on the SRA dataset with accession number "SRR000001", including information on the number of reads, base pairs, and read lengths. The report will be output to the console.

Users can now use the downloaded FASTQ files for downstream analysis, such as read alignment or variant calling.

Use the following slurm script as a template to run SRA toolkit,

```
#!/bin/bash
#SBATCH --job-name=sra_download
#SBATCH --p main
#SBATCH --qos main
#SBATCH --output=sra_download_%j.out
#SBATCH --error=sra_download_%j.err
#SBATCH --ntasks=1
#SBATCH --cpus-per-task= 12
#SBATCH --mem=16G
#SBATCH --time=2:00:00

module load bio/sratoolkit/2.9.2

#fetch data
prefetch SRR000001

#convert to fastq
fastq-dump --gzip --skip-technical --readids --dumpbase --split-files
--clip ~/ncbi/public/sra/SRR000001.sra
```

For more information, please visit the sra website linked above.

### ***Where to find help?***

If you are confused or need help at any point, please contact OIT at the following address.

<https://ua-app01.ua.edu/researchComputingPortal/public/oitHelp>

