Analysis Report (Part1): Getting the Data

(Last updated: 11/05/21)

Contents

1. Downloading the Reads

The reads were obtained from SRA in the form of a set of FASTQ files. The FASTQ files were downloaded using **sratoolkit.2.9.6**. The following Bash script was employed:

```
#!/bin/bash

# FETAL SAMPLES

fastq-dump -v --gzip --split-files -O /data/sra  SRR2071348
printf "\n****** SRR2071348 is downloaded  ...\n\n"

fastq-dump -v --gzip --split-files -O /data/sra  SRR2071349
printf "\n****** SRR2071349 is downloaded  ...\n\n"

fastq-dump -v --gzip --split-files -O /data/sra  SRR2071352
printf "\n****** SRR2071352 is downloaded  ...\n\n"

# ADULT SAMPLES

fastq-dump -v --gzip --split-files -O /data/sra  SRR2071346
printf "\n****** SRR2071346 is downloaded  ...\n\n"

fastq-dump -v --gzip --split-files -O /data/sra  SRR2071347
printf "\n****** SRR2071347 is downloaded  ...\n\n"

fastq-dump -v --gzip --split-files -O /data/sra SRR2071350
printf "\n****** SRR2071350 is downloaded  ...\n\n"
```

After saving the above script as *download-sra-reads.sh* in *bash-scripts* directory, the following Bash command was used for running it:

```
nohup sh bash-scripts/download-sra-reads.sh > download-sra-reads.out &
```

2. Reads Quality Control

The quality of the reads was checked using **FastQC-v0.11.8**. The following Bash script was employed:

```
#!/bin/bash
# ALL SAMPLES
fastqc -o fastqc /data/sra/*fastq.gz
```

3. Retrieving the Phenotype Data

The phenotype data of the samples was retrieved from the SRA website and stored in the *phenotype_data.tsv* file.