

?? ?? ??

SRA Lite ☐☐☐ ☐☐ ☐☐

WGS of Streptococcus agalactiae: strain SGEHI2015-31 (SRR6327875)

🔍 Metadata 🧑‍🔬 Analysis 📄 Reads 📁 Data access 📄 FASTA/FASTQ download

SRA archive data

SRA archive data is normalized by the SRA load process and used by the [SRA Toolkit](#) to read and produce formats like FASTQ, SAM, etc. The default toolkit configuration enables it to find and retrieve SRA runs by accession.

Public SRA files are now available from GCP and AWS cloud platforms as well as from NCBI. Access to most data in the cloud requires a user account with the cloud service provider. The user's account will incur costs for cloud compute or to copy data outside of the specified cloud service region.

Type	Version	Created	Size	Location	Name	Free Egress	Access Type
SRA Normalized	1	2017-11-29	665.1M	AWS	https://sra-pub-run-odp.s3.amazonaws.com/sra/SRR6327875/SRR6327875	worldwide	anonymous
SRA Lite	1	2020-06-23	365.9M	NCBI	https://sra-downloadb.be-md.ncbi.nlm.nih.gov/sos5/sra-pub-zq-14/SRR006/327/SRR6327875.sralite.1	worldwide	anonymous
				AWS	s3://sra-pub-zq-5/SRR6327875/SRR6327875.sralite.1	s3.us-east-1	aws identity
				GCP	gs://sra-pub-zq-6/SRR6327875/SRR6327875.zq.1	gs.US	gcp identity

Original format

The original files submitted to SRA. These files may require specific software to open, read and interpret data.

Type	Version	Created	Size	Location	Name	Free Egress	Access Type
fastq	1	2020-03-13	454.7M	AWS	s3://sra-pub-src-8/SRR6327875/WEB312-combined_R1.fastq.gz	-	Use Cloud Data Delivery
fastq	1	2020-03-13	473.5M	AWS	s3://sra-pub-src-8/SRR6327875/WEB312-combined_R2.fastq.gz	-	Use Cloud Data Delivery

👉 📁 📁 📁 📁 📁 📁 , Cloud9 📁 (EC2 📁)📁 📁 EC2 📁
SSH 📁 📁 📁 . (SSH 📁 📁 EC2 📁 📁)

?? ??? ??

📁

```
cd /mnt/volume1
wget https://sra-downloadb.be-md.ncbi.nlm.nih.gov/sos5/sra-pub-zq-14/SRR006/327/SRR6327875.sralite.1
```

R1, R2 📁 📁

📁 : c5.4xlarge 📁 📁 📁 📁 6📁 30📁 📁 📁 .

```
fastq-dump --split-files --gzip SRR6327875.sralite.1
```

Registry of Open Data on AWS 📁

SRR6327875 📁 Registry of Open Data on AWS📁 📁 📁 📁 📁 .
fastqerq-dump📁 📁 📁 📁 📁 📁_📁 📁 .

```
mkdir -p /tmp/fastq/SRR6327875
cd /tmp/fastq/SRR6327875

# note here the --no-sign-request makes an anonymous request to this public S3 bucket
aws s3 sync --no-sign-request s3://sra-pub-run-odp/sra/SRR6327875/ /tmp/fastq/SRR6327875/
```

```
# convert the sra formatted file to fastq, then gzip them and clean up
fasterq-dump ./SRR6327875
gzip SRR6327875_1.fastq
gzip SRR6327875_2.fastq
rm -f SRR6327875
```

download failed: s3://sra-pub-run-odp/sra/SRR6327875/SRR6327875 to ./SRR6327875 [Errno 28] No space left on device

```
df -h
df -h /tmp
```

!

```
SRR6327950
SRR6327950
```

(Optional) AWS CLI Amazon S3

```
--bucket {bucket} \
S3
```

```
brandon-20230927
```

```
aws s3api create-bucket \
  --bucket {bucket} \
  --region us-east-1
```

AWS

S3 (AWS S3)

Buckets

Access Points

Object Lambda Access Points

Multi-Region Access Points

Batch Operations

IAM Access Analyzer for S3

Block Public Access settings for this account

Storage Lens

Dashboards

AWS Organizations settings

Feature spotlight 7

AWS Marketplace for S3

▼ Account snapshot

Storage lens provides visibility into storage usage and activity trends. [Learn more](#)[View Storage Lens dashboard](#)

Total storage

⌚ Pending

Object count

⌚ Pending

Average object size

⌚ Pending

You can enable advanced metrics in the "default-account-dashboard" configuration.

Buckets (1) Info

Buckets are containers for data stored in S3. [Learn more](#)

Copy ARN

Empty

Delete

Create bucket

Find buckets by name

< 1 > ⌕

Name	AWS Region	Access	Creation date
○ brandon-20230927	US East (N. Virginia) us-east-1	Bucket and objects not public	September 27, 2023, 13:29:16 (UTC+09:00)

S3

```
aws s3 cp SRR6327875.sralite.1_1.fastq.gz s3://{bucket}/raw/SRR6327875_1.fastq.gz
aws s3 cp SRR6327875.sralite.1_2.fastq.gz s3://{bucket}/raw/SRR6327875_2.fastq.gz
```

AWS S3 CLI

1) S3 recursive listing (-r flag)

```
. aws s3 ls
. aws s3 ls s3://aws-lab-james
. aws s3 ls --recursive s3://aws-lab-james
```

2) local directory upload

```
. aws s3 cp --recursive ./video/ s3://aws-lab-james/video/
```

3) S3 recursive download

```
. aws s3 cp --recursive s3://aws-lab-james/video/ ./video/
```

4) local directory sync upload (cp -r flag)

```
. aws s3 sync ./video/ s3://aws-lab-james/video/
```

5) S3 recursive download (cp -r flag)

```
. aws s3 sync s3://aws-lab-james/video/ ./video/
```

S3 bucket name and region

For this tutorial, we will use the S3 bucket `brandon-20230927` in the `us-east-1` region.

1. List the S3 bucket contents:

```
aws s3 ls s3://{bucket-name}/raw/SRR6327875_1.fastq.gz --region us-east-1
```

```
) aws s3 ls s3://brandon-20230927/raw/SRR6327875_1.fastq.gz --region us-east-1
```

The output shows the file `SRR6327875_1.fastq.gz` in the `raw` directory of the `brandon-20230927` bucket.

2. Create a local directory for the files:

```
mkdir -p /tmp/fastq/SRR6327875
```

```
cd /tmp/fastq/SRR6327875
```

3. AWS CLI command to download the S3 bucket contents:

```
aws s3 cp s3://{bucket-name}/raw/SRR6327875_1.fastq.gz .
```

```
aws s3 cp s3://{bucket-name}/raw/SRR6327875_2.fastq.gz . --region us-east-1
```

4. MD5 checksum command to verify the files:

```
md5sum *.fastq.gz
```

```
)
```

```
20|ubuntu@ip-172-31-47-43:/tmp/fastq/SRR6327875$ md5sum *.fastq.gz
cbfecacca82edd95f2ffb20364f3dc76  SRR6327875_1.fastq.gz
ce3c2255d4280565a73af018f222d317  SRR6327875_2.fastq.gz
21|ubuntu@ip-172-31-47-43:/tmp/fastq/SRR6327875$ md5sum ~/.fastq.gz
cbfecacca82edd95f2ffb20364f3dc76  /home/ubuntu/SRR6327875.sralite.1_1.fastq.gz
ce3c2255d4280565a73af018f222d317  /home/ubuntu/SRR6327875.sralite.1_2.fastq.gz
```

5. S3 bucket for NCBI SRA repository ([Registry of Open Data on AWS](#)) bucket name and region:

```
mkdir -p /tmp/fastq/SRR6327950
```

```
cd /tmp/fastq/SRR6327950
```

```
# note here the --no-sign-request makes an anonymous request to this public S3 bucket
```

```
aws s3 sync --no-sign-request s3://sra-pub-run-odp/sra/SRR6327950/ /tmp/fastq/SRR6327950/
```

```
# convert the sra formatted file to fastq, then gzip them and clean up
```

```
fasterq-dump ./SRR6327950
```

```
gzip SRR6327950_1.fastq
```

```
gzip SRR6327950_2.fastq
```

```
rm -f SRR6327950
```

□□ □ □□ □□□ □□□ □□□ □□□□ .

Revision #27

Created 27 September 2023 02:10:38 by Hyunmin Kim

Updated 9 May 2024 03:24:10 by Hyunmin Kim