



GSA Submission Quick Start Guide

| | |
|---|----------|
| Overview | 2 |
| Login to the Genome Sequence Submission..... | 2 |
| Create a GSA Submission | 2 |
| <i>Creating and Describing New GSA Submission</i> | <i>2</i> |
| <i>Creating and Describing Experiment</i> | <i>3</i> |
| <i>Creating Runs and Describing Run.....</i> | <i>4</i> |
| Data Transfer..... | 6 |
| Quality Control and Feedback..... | 6 |

Overview

We offer a number of services through which data can be submitted to the BIGD, You can use this service to submit sequence reads, genome assemblies, targeted assembled and annotated sequences and to register projects and samples.

The Genome Sequence Archive (GSA) stores raw sequence read data from "next-generation" sequencing technologies (Figure 1).

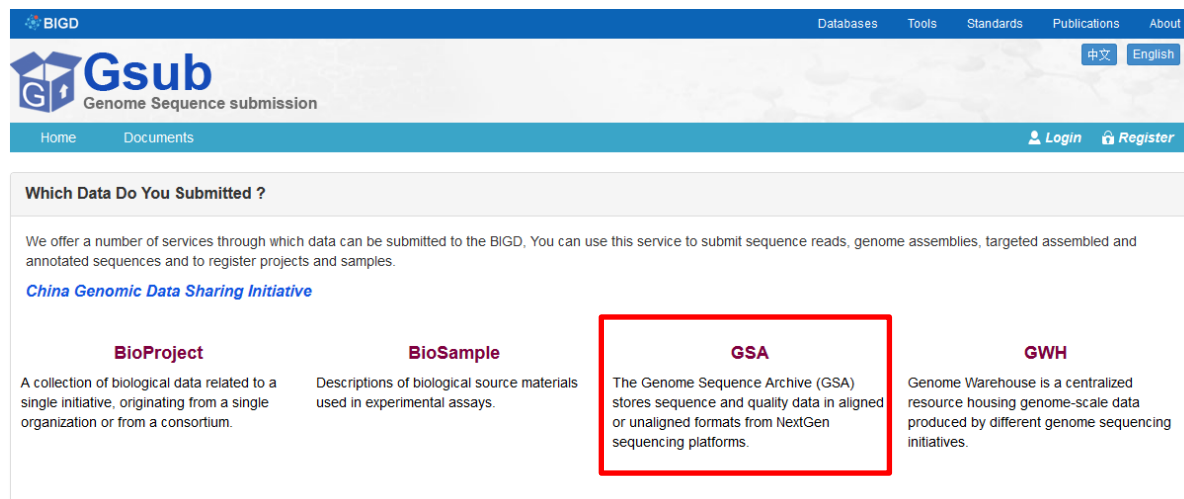


Figure 1. Home page of Genome Sequence Submission (Gsub)

Login to the Genome Sequence Submission (Gsub)

- Click the login tab, and then login. If you do not have an account already, click the Register tab to create one (see Figure 2). If you have used an account in the past but no longer see your previous submissions, please contact us at gsa@big.ac.cn for assistance with your account view.



Figure 2. from the 'login' tab/ 'register' tab, click to login/register for Submission.

- **Do NOT** suggest using the Win10 operating system; Recommend the use of Firefox browser version, other browsers may have bugs.
- After the activation of the login system, use our Submit Reads Data System—Gsub and follow steps to finish the submission.

Create a GSA Submission

• *Creating and Describing New GSA Submission*

Click the 'Create GSA' button to create a new GSA Submission (Figure 3).

The GSA Accession No. is assigned with prefix 'CRA'. The corresponding description is like: "The sequence data reported in this paper have been deposited in the Genome Sequence Archive of Beijing Institute of Genomics, Chinese Academy of Sciences".

| Accession | Submission ID | Title | Date Released | Status | Operation |
|------------|------------------------------|--------|---------------|---------------------------|------------------------|
| Unassigned | subCRA000019 | Tumor2 | 2017-12-12 | Unchecked Confidential | Delete |
| Unassigned | subCRA000018 | Tumor1 | 2017-12-12 | Unchecked Confidential | |

Figure 3. Click the ' Create GSA ' tab to register GSA

Note: Alias is an ID used by submitters to track the submission of a set of Experiments and Runs. The release date is required for all submissions. It is advisable to enter a release date before loading any data into a Submission. This will prevent accidental early release of data. Dates may be set for up to two year in the future in anticipation of a publication release date (Figure 4).

New GSA Submission

* **Alias**

* **Date Released**

 (yyyy-mm-dd)

Figure 4. New GSA Submission

• **Creating and Describing Experiment**

An Experiment describes a sequencing library and instrument. An Experiment references 1 BioProject and 1 BioSample. Click the 'Add Experiment' button to begin creating an Experiment (Figure 5a), and then filling the Meta Information and Additional descriptions of library(Figure 5b).

Home Documents sisi Logout

Gsub / GSA / subCRA000018

Basic information

Submission of GSA: subCRA000018 / Tumor1 / release time : 2017-12-12

* Alias: Tumor1 * Date Released: 2017-12-12

Experiments & Runs

| Experiment Accession | Title / Alias | Taxon Name | Platform | Project | Sample | Release Date | Experiment Status | Operation |
|----------------------|---|--------------|-----------|--------------------------|-------------------------|--------------|---------------------------|--|
| CRX000176 | Evolutionary genomics in Hepatocellular carcinoma | Homo sapiens | 454 GS 20 | PRJCA000258 CRP000406 | SAMC000319 CRS010007 | 2017-12-12 | Unchecked Confidential | <input type="button" value="Add Run"/> |

Figure 5a. Click the 'Add Experiment' button to begin creating an Experiment.

Gsub / GSA / subCRA000018 / Experiment

Meta Information

* Platform: 454 GS 20 * Alias: Some description of the experiment alias * Title: Some description of the experiment title

* Project Accession: * Sample accession:

* Library Construction / Experiment design: Some description of the library design

Library

Library name: Some description of the library name * Strategy: WGA * Source: GENOMIC * Selection: unspecified

* Layout: FRAGMENT, FRAGMENT, PAIRED

Processing

* Layout: PAIRED * Insert size (bp): Insert size Nominal size (bp): Nominal size Nominal standard deviation (bp): Nominal standard deviation

Link to BioProject ID and BioSample ID you created in BioProject and BioSample Submission step, please see detail in BioProject Submission Quick Start Guide and BioSample Submission Quick Start Guide.

There are two options for the Layout, 'FRAGMENT' and 'PAIRED', and is closely related to the number of Run files.

Figure 5b. Creating and Describing Experiment

• **Creating Runs and Describing Run**

Runs describe the files that belong to the previously created Experiments. They specify the data files for a specific sample to be processed by GSA. Experiments may contain many Runs depending on how many sequencer runs were involved in data acquisition (Figure 6).

Home Documents sisi Logout

Gsub / GSA / subCRA000018

Basic information

Submission of GSA: subCRA000018 / Tumor1 / release time : 2017-12-12

* Alias: * Date Released:

Experiments & Runs Add Experiment

| Experiment Accession | Title / Alias | Taxon Name | Platform | Project | Sample | Release Date | Experiment Status | Operation |
|----------------------|---|--------------|-----------|--------------------------|-------------------------|--------------|---------------------------|--|
| CRX000176 | Evolutionary genomics in Hepatocellular carcinoma | Homo sapiens | 454 GS 20 | PRJCA000258 CRP000406 | SAMC000319 CRS010007 | 2017-12-12 | Unchecked Confidential | <input type="button" value="Add Run"/> |

Figure 6a. Click the 'Add Run' button to begin creating an Run.

Home Documents sisi Logout

Gsub / GSA / subCRA000018 / Run

Run Submission of Experiment: CRX000176 / Tumor1

General Information

* Alias:

* Run data file type: (dropdown menu with options: fastq, bam, vcf, sff, Data Blocks)

* File Name: * MD5 for file:

Transmitting your data files to the GSA FTP site
 Address: ftp://submit.big.ac.cn
 User: Same as you login the GSA
 Password: Same as you login the GSA

The storage format (fastq, BAM, sff, vcf) of the sequence data being submitted and Support zip and gz format of compressed files.

A checksum or hash sum generated for the file listed in 'File Name' that is used to detect errors introduced through storage or transfer.

Figure 6b. Creating and Describing Run

Note: Runs can only be updated until data has been loaded for the Run. Once there is data in a Run, it will be locked from further updates. Contact GSA for changes to be made to locked Runs (Figure 7).

Gsub Genome Sequence submission 中文 English

Home Documents sisi Logout

Gsub / GSA

The GSA Accession No. is assigned with prefix 'CRA'. The corresponding description is like: "The sequence data reported in this paper have been deposited in the Genome Sequence Archive of Beijing Institute of Genomics, Chinese Academy of Sciences".

| Accession | Submission ID | Title | Date Released | Status | Operation |
|------------|------------------------------|--------|---------------|---------------------------|---------------------------------------|
| Unassigned | subCRA000019 | Tumor2 | 2017-12-12 | Unchecked Confidential | <input type="button" value="Delete"/> |
| Unassigned | subCRA000018 | Tumor1 | 2017-12-12 | Unchecked Confidential | |

Figure 7. Summary Run display

Data Transfer

After the metadata is entered, data may be uploaded to the Gsub. Use FTP client software (such as FileZilla Client) to log in FTP serve.

Transmitting your data files to the Gsub FTP site

Address: ftp://submit.big.ac.cn

User: Same as you login the Gsub

Password: Same as you login the Gsub

Quality Control and Feedback

After finished all above tasks, Gsub system will check your information and files, and give your feedback.