

# Genome Sequence Archive (GSA)

## Quick Start Guide

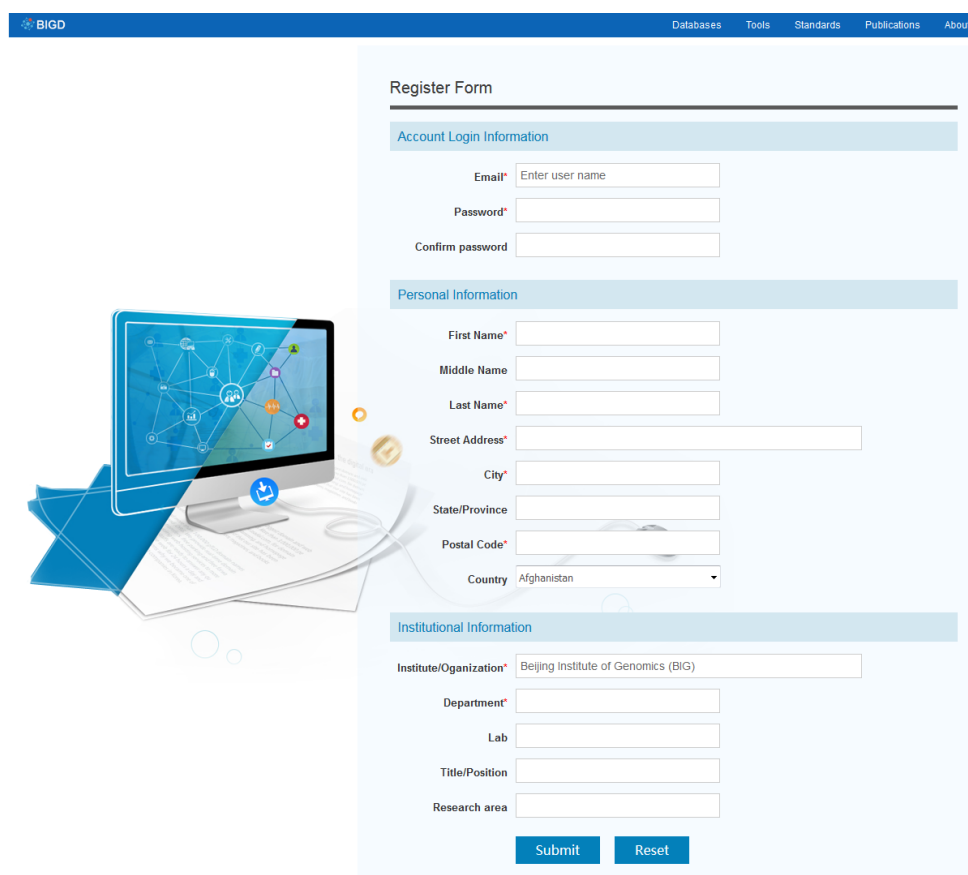
|   |           |
|---|-----------|
| <b>Introduction.....</b>                                | <b>2</b>  |
| <b>Log in to the BIG Submission Portal.....</b>         | <b>2</b>  |
| <b>Create new GSA Submission.....</b>                   | <b>3</b>  |
| <b>How to Edit, Delete or Add New Data .....</b>        | <b>13</b> |
| <b>How to Share Your Data .....</b>                     | <b>15</b> |
| <b>How to Release Your Data.....</b>                    | <b>16</b> |
| <b>Data File Uploading .....</b>                        | <b>17</b> |
| Aspera Command Line .....                               | 17        |
| FTP Upload.....   | 18        |
| Web browser upload via Aspera Connect plugin .....      | 19        |
| <b>Release of linked BioProject/BioSample/GSA .....</b> | <b>21</b> |
| <b>Status and Operation.....</b>                        | <b>22</b> |

## Introduction

The Genome Sequence Archive (GSA) is a data repository for archiving raw sequence reads. It accepts data submissions from all over the world and provides free access to all publicly available data for global scientific communities.

## Log in to the BIG Submission Portal

Login to the BIG Submission Portal (BIG Sub, <https://bigd.big.ac.cn/gsub/>): Click the 'login' tab, then login. If you do not have an account yet, click the 'Register' tab to create one. If you have any problems about your account, please contact [bigd-admin@big.ac.cn](mailto:bigd-admin@big.ac.cn) for assistance.

A screenshot of the 'Register Form' on the BIG Submission Portal. The form is divided into three sections: 'Account Login Information', 'Personal Information', and 'Institutional Information'. The 'Account Login Information' section includes fields for 'Email\*' (with a placeholder 'Enter user name'), 'Password\*', and 'Confirm password'. The 'Personal Information' section includes fields for 'First Name\*', 'Middle Name', 'Last Name\*', 'Street Address\*', 'City\*', 'State/Province', 'Postal Code\*', and a dropdown menu for 'Country' (currently set to 'Afghanistan'). The 'Institutional Information' section includes fields for 'Institute/Organization\*' (pre-filled with 'Beijing Institute of Genomics (BIG)'), 'Department\*', 'Lab', 'Title/Position', and 'Research area'. At the bottom of the form are 'Submit' and 'Reset' buttons. To the left of the form is a decorative graphic of a computer monitor displaying a network diagram, a keyboard, and some papers.

**Notice:** After logged in to the BIG Submission Portal, you can follow the steps below to finish the submission.

## Create new GSA Submission

The BIG Sub provides a browser-based user interface for submitting GSA metadata as well as various options for uploading data files.

The screenshot shows the BIG Submission Portal interface. At the top, there is a navigation bar with 'BIGD' and 'BIG Sub' logos, and a 'Bilingual Mode' selector with '中文' and 'English' options. Below the navigation bar, there is a section titled 'Select a data type to be submitted'. This section contains six cards representing different data types: BioProject, BioSample, Genome Sequence Archive (GSA), Genome Warehouse (GWH), Genome Variation Map (GVM), and BioCode. The 'Genome Sequence Archive (GSA)' card is highlighted with a red border. Below this section, there is a large downward-pointing arrow. At the bottom of the page, there is a 'Create GSA' button, also highlighted with a red border.

The page tabs presented by the Submission wizard are:

- **Submitter** – the name and email information are auto-filled if you have logged in, and in this case, the system is able to identify the person who is entering the data in the form.

BIG Sub / GSA / New GSA

01 Submitter    02 General    03 Metadata    04 File Upload    05 Overview

Submitter Information

|   |  |                                 |
|---|--|---------------------------------|
| * First name<br>Zhang                               | Middle name<br>middle name                   | * Last name<br>Sisi             |
| * Email<br>zhangss@big.ac.cn                        | Email (secondary)<br>secondary email         |                                 |
| * Organization<br>Beijing Institute of Genomics, Ch | Organization website<br>http://www.big.ac.cn | * Department<br>BIG Data Center |
| Phone   | Fax  |                                 |
| * Street<br>Beichen West Road                       | * City<br>Beijing                            | State/Province                  |
| * Postal code<br>100101                             | * Country/Region<br>China                    |                                 |

Save and forward

- **General information** –this page collects general descriptive information about the GSA’s **Release Date, Title and Description, Project Information** and **Sample Information**. This step links your existing project (PRJCA#) or samples (SAMC#) with your GSA data.

**Notice:** If you have already created GSA related Biological Sample(s) in the [BioSample database](#), please select the ‘**GSA related BioSample information has been created**’. Then follow the wizard to complete the submission.

If you have not created GSA related Biological Sample(s), please select the ‘**No GSA related BioSample information was created**’. Then follow the wizard to complete the submission.

01 Submitter
02 General
03 Metadata
04 File Upload
05 Overview

General Information

Release date

After the approval is passed, release immediately following curation (recommended)  
 Release on specified date  
   
(yyyy-mm-dd)

**\* Release Policies and Disclaimers**

1. A date can be set by authors to withhold the release of new submissions for a specified period of time.
2. The release date can be changed through the GSA submission portal ([http://bigd.big.ac.cn/gsub/submit/gsa/\[substitute your GSA accession number\]/contents](http://bigd.big.ac.cn/gsub/submit/gsa/[substitute your GSA accession number]/contents))
3. If a paper citing the sequence or accession number is published prior to the specified date, the sequence will be released upon publication. Otherwise, GSA will release sequence data on the specified date.
4. As soon as they are available, please send the full publication data--all authors, title, journal, volume, pages and date--to the following address: [gsa@big.ac.cn](mailto:gsa@big.ac.cn)

I accept it.  I don't accept it.

**Release Date** is the date your submission can be publicly accessed. If your submitted information need to be protected for some time, you can set this date to a future date within two years.

**Release Policies and Disclaimers**

Title And Description

\* title

\* description

Project Information

\* Please select Bioproject number

OR Go to create [BioProject](#)

If you have not created BioProject yet, click and go to create [BioProject](#).

If you have created BioProject, select the BioProject Accession number.

Sample Information

No GSA related BioSample information was created

GSA related BioSample information has been created

**Select the 'No GSA related BioSample information was created' as an example.**

**Notice:** If you select '**Release immediately following curation**', the records will be released after the approval passed. If you select '**Release on a specified date**', the GSA will be released on the date you specify.

- **Sample Type** –this page provides a preview of the sample type that submitter is asked to supply during the submission process.

01 Submitter 02 General **03 Sample Type** 04 Attributes 05 Metadata 06 Files 07 Overview

**Sample type**


**Pathogen**  
Used for pathogen samples that are relevant to public health. Required attributes include those considered useful for the rapid analysis and trace back of pathogens.

**Clinical or host-associated**

**Environmental, food or other**

**Microbe**  
Used for bacteria or other microbes when it is not appropriate or advantageous to use for Pathogen or Virus packages.

**Animal**  
Used for multicellular samples or cell lines derived from common laboratory model organisms, e.g., mouse, rat, Drosophila, worm, fish, frog, or large mammals including zoo and farm animals.

**Human**   
Human genetic related raw sequence reads should be submitted to the [GSA for Human](#) database.

**Plant**  
Used for any plant sample or cell line.

**Virus**  
Used for all virus samples not directly associated with disease. Viral pathogens should be submitted using the Pathogen: Clinical or host-associated pathogen package.

**Metagenome/Environmental Sample (GSC MIMS unsupported)**  
Use for metagenome/environmental samples when it is not appropriate or advantageous to use the GSC (Genome Standards Consortium) MIMS (Minimum Information about a MetaGenome Sequence) standards.

**Metagenome/Environmental Sample (GSC MIMS compliant)**  
Describe and standardize sample metadata, defined by the GSC (Genome Standards Consortium) MIMS (Minimum Information about a MetaGenome Sequence) standards for metagenome/environmental samples.

**human-gut**

**soil**

**water**

**Save and forward**

Select the 'Plant' package as an example.

**Notice:** If you determine that your human data must be submitted via **GSA Human database**, please delete your current submission and contact us at [gsa@big.ac.cn](mailto:gsa@big.ac.cn).

- **Attributes** –this page allows you to upload attributes information about multiple samples in a single table. To finish the procedure, you should:

- 1) Download the BioSample submission template table [Plant.us.xlsx](#). For column explanations and examples, please see the [e.g.Plant.us.xlsx](#). For more information, please see the [Help](#).

**Notice:** Downloading new template ensures that you get the most current and correct version.

01 Submitter 02 General 03 Sample Type **04 Attributes** 05 Metadata 06 Files 07 Overview

**Attributes of plant**

\* Upload BioSample batch submission file using Excel format that includes the attributes for each of Samples

Please select file **Upload**

Download BioSample batch submission template [Plant.us.xlsx](#), then fill in and double-check it before uploading.  
For column explanations and examples, please see the [e.g.Plant.us.xlsx](#).  
For more information, please see the [Help](#).

**Save and forward**

- 2) Fill in the template table and double-check it before uploading. Use the Selection box to select your completed table.

01 Submitter 02 General 03 Sample Type 04 Attributes 05 Metadata 06 Files 07 Overview

Attributes of plant

\* Upload BioSample batch submission file using Excel format that includes the attributes for each of Samples

Please select file Upload

Download BioSample batch submission template [Plant.us.xlsx](#), then fill in and double-check it before uploading.  
For column explanations and examples, please see the [e.g.Plant.us.xlsx](#).  
For more information, please see the [Help](#).

Save and forward

- 3) Then click the 'Check' button to verify the submitted batch information online.

01 Submitter 02 General 03 Sample Type 04 Attributes 05 Metadata 06 Files 07 Overview

Attributes of plant

\* Uploaded BioSample batch submission file using Excel format that includes the attributes for each of Samples

[Plant.xlsx](#) 24KB Delete Check

Please click the "Check" button to verify the submitted batch information online. If the file checked OK, please click the "Save and forward" to complete your submission. If not, please delete the file first, then modify and reload until it passed the examination.

Save and forward

- 4) If the file has passed the examination, please click the 'Save and forward' button to complete your submission. If not, please click the 'Delete' button. You should edit and re-upload the file until it is correct.

01 Submitter 02 General 03 Sample Type 04 Attributes 05 Metadata 06 Files 07 Overview

Attributes of plant

\* Uploaded BioSample batch submission file using Excel format that includes the attributes for each of Samples

[Plant.cn.xlsx](#) 24KB Delete Check ✘ ERROR:

Please download the error file: [error.txt](#)

Error in Sample sheet  
row 11, column 7: "biomaterial\_provider" is empty  
row 12, column 7: "biomaterial\_provider" is empty  
row 13, column 7: "biomaterial\_provider" is empty  
row 14, column 7: "biomaterial\_provider" is empty

Please click the "Check" button to verify the submitted batch information online. If the file checked OK, please click the "Save and forward" to complete your submission. If not, please delete the file first, then modify and reload until it passed the examination.

Save and forward



• **Metadata** – this page allows you to upload metadata information about the raw sequences, including Experiments and Runs. To finish the procedure, you should:

- 1) Download the GSA submission template table [GSA\\_Template.cn.xlsx](#). For column explanations and examples, please see the [e.g.GSA\\_Template.cn.xlsx](#). For more information, please see the [Help](#).

**Notice:** Downloading new template ensures that you get the most current and correct version

- 2) Fill in the template table and double-check it before uploading. Use the Selection box to select your completed table.

- 3) Then click the **'Check'** button to verify the submitted batch information online.



01 Submitter 02 General 03 Sample Type 04 Attributes 05 Metadata 06 Files 07 Overview

GSA Metadata Information

\* Uploaded GSA batch submission file.

plant\_ExR.xlsx 32KB Delete **Check**

Please click the "Check" button to verify the submitted batch information online. If the file checked OK, please click the "Save and forward" to complete your submission. If not, please delete the file first, then modify and reload until it passed the examination.

Save and forward

- 4) If the file has passed the examination, please click the 'Save and forward' button to go to the next step of the submission. If not, please click the 'Delete' button. You should edit and re-upload the file until it is correct.

01 Submitter 02 General 03 Sample Type 04 Attributes 05 Metadata 06 Files 07 Overview

GSA Metadata Information

\* Uploaded GSA batch submission file.

plant\_ExR.xlsx 32KB Delete Check ✔ Checked OK.

Please click the "Check" button to verify the submitted batch information online. If the file checked OK, please click the "Save and forward" to complete your submission. If not, please delete the file first, then modify and reload until it passed the examination.

**Save and forward**

- **File Upload** – this page allows you to select the file upload method, including **FTP**, **Aspera Command Line** (Recommend) and **Web browser upload via Aspera Connect plugin**. You can also upload your files before the metadata submission if you choose the **FTP** or **Aspera Command Line** method.

**Notice:**

- 1) Please remember to check the names and MD5 checksums of the sequence files, which must be the same as those you filled in the batch submission table. Otherwise, your files cannot be archived correctly.
- 2) **If you choose the Aspera Command Line to upload files, please write down the Aspera Command Line information.**

01 Submitter   02 General   03 Sample Type   04 Attributes   05 Metadata   06 File Upload   07 Overview

**File upload mode**

**Notice:**  
 Each file **must be listed in the GSA metadata file you uploaded.**  
**Unique file names should be used for all files.**  
 Files must be compressed using **gzip** or **bzip2**.

**\* How do you want to provide files for submission?**

**FTP**  
 Use FTP client to upload files. FTP Username and Password are same to the BIG Sub account.  
 You may upload your files to our FTP site using an FTP client software (such as the [FileZilla Client](#)) or using the FTP command.  
**Address:** ftp://submit.big.ac.cn  
**Username:** Same as you login the BIG Sub  
**Password:** Same as you login the BIG Sub  
 After you connected to the FTP server, please **navigate to the GSA directory and upload your files here. DO NOT upload files to the root directory.**  
**NOTE:** Please use the **binary mode** to transfer files. If you are using the FTP client software to log in to the FTP server, follow the softwares instruction to set the transfer mode. If you are using ftp command, **type the binary command before the mput command**

**Aspera Command Line**  
 Use Aspera Command Line to upload files.

**Web browser upload via Aspera Connect plugin**  
 Use the Aspera Connect plugin to upload files. If you dont have Aspera Connect plugin installed in your computer, please [Download Aspera Connect](#) and install it first.

Auto finish submission

**Save and forward**

- **Overview** – this page presents a summary of the provided information. On this step, please be careful to check the details of your submission. If you need to make changes, go back by clicking the relevant tab then edit. If everything is correct, click the **'Submit'** button to complete the submission.

01 Submitter 02 General 03 Sample Type 04 Attributes 05 Metadata 06 Files 07 Overview

**Basic Information**

Submission ID: subCRA003595  
 title: Full-length transcriptome sequencing from multiple tissues of annual ryegrass  
 description: Full-length transcriptome sequencing of annual ryegrass (*Lolium multiflorum* Lam.)  
 Release date: 2020-07-29

**Submitter Information**

Submitter: Zhang Sisi  
 Organization: [zhangss@big.ac.cn](mailto:zhangss@big.ac.cn)  
 Beijing Institute of Genomics, Chinese Academy of Sciences  
 Department: BIG Data Center  
 Country/Region: China  
 Address: Beichen West Road Beijing  
 Postal code: 100101

**Metadata**

Sample type: Plant  
 Biosample attributes file: [Plant.xlsx](#)  
 Metadata file: [plant\\_ExR.xlsx](#)

| Experiment Title | Platform                                | Sample Name | Taxon Name                | Selection |
|------------------|---|-------------|---------------------------|-----------|
| CK1              | PacBio Sequel                           | CK1         | <i>Lolium multiflorum</i> | PCR       |
| Run Alias        | Run Data File Information               |             |                           | File Type |
| CK1              | File: m54211_180613_064306.subreads.bam |             |                           | bam       |
| CK2              | PacBio Sequel                           | CK2         | <i>Lolium multiflorum</i> | PCR       |
| Run Alias        | Run Data File Information               |             |                           | File Type |
| CK2              | File: m54212_180609_084342.subreads.bam |             |                           | bam       |
| DK1              | PacBio Sequel                           | DK1         | <i>Lolium multiflorum</i> | PCR       |
| Run Alias        | Run Data File Information               |             |                           | File Type |
| DK1              | File: m54213_180607_091704.subreads.bam |             |                           | bam       |
| DK2              | PacBio Sequel                           | DK2         | <i>Lolium multiflorum</i> | PCR       |
| Run Alias        | Run Data File Information               |             |                           | File Type |
| DK2              | File: m54214_180609_073921.subreads.bam |             |                           | bam       |

Total Items: 4 | Items of 1 - 4 | Page size: 20 | Page 1/1 | << First | Last >> | Jump To: 1 | GO

**Submit**



Create GSA GSA Quick Start Guide (US) GSA Quick Start Guide (CN)

| Accession  | Submission ID | Title   | Release date | Status                    | Operation |
|------------|---------------|---|--------------|---------------------------|-----------|
| Unassigned | subCRA003595  | Full-length transcriptome sequencing from multiple tissues of annual ryegrass | 2020-07-29   | Unchecked<br>Confidential | Delete    |

After completing the submission, please wait for data curation. We will check both metadata and the sequence files and send feedbacks to your registered Email if they are not perfectly correct. So, please **pay attention to your mail feedbacks**. After the curation, your data will be archived to a single GSA set and the assigned accession number will be shown in your GSA list.

| Accession | Submission ID | Title   | Release date | Status                     | Operation |
|-----------|---------------|---|--------------|----------------------------|-----------|
| CRA002926 | subCRA003595  | Full-length transcriptome sequencing from multiple tissues of annual ryegrass | 2020-07-29   | Checked OK<br>Confidential | Delete    |

**Notice:**

- 1) Each new submission receives a temporary **Submission ID** in the form of sub#, like subCRA003595. Please provide this ID when contacting the GSA Working Team. DO NOT use the temporary **Submission ID** in the publication or [BIG Search](#).
- 2) After the submission, you will get the **GSA Accession** numbers in the form of CRA#, like CRA002926. **Please use this number in a publication or [BIG Search](#).**

## How to Edit, Delete or Add New Data

Before the GSA data are archived, you can click the **Submission ID** to enter the Overview page. On this page, you can 1) update the **Release date** and **Title**; 2) edit the **Submitter information**; 3) Append data by clicking the **'Add Data'** button, for more information, please see ['Create new GSA Submission'](#); 4) edit or delete Metadata information for each submitted Experiment or Run; and 5) upload or update data files by clicking **'Upload File'** button.

**Notice:** For more detail about submission status and the available operations, please go to ['Status and Operation'](#).

**Basic Information**

Submission of GSA: subCRA000633 / Release Date: 2020-12-31 / Project: **PRJCA000644** Related BioProject

Status: 1 Runs are waiting for check  
2 Runs are waiting for files

Title: Human dataset Update Release Date and Title

Date Released: 2020-12-31 Update

[Contact us](#)

**Submitter Information**

Submitter: Zhang Sisi  
Organization: Beijing Institute of Genomics, Chinese Academy of Sciences  
Department: BIG Data Center  
Country/Region: China  
Address: Beichen West Road Beijing  
Postal code: 100101 Edit Edit Submitter information

**Metadata**

Sample type: Human sample Related BioSample(s)

BioSample Submission ID: **subSAM023445**

Metadata file: 5.Human.cn.test5.xlsx added on 2019-07-30  
exp\_run-Human.cn.test5.xlsx added on 2019-07-30

Update File Add Data

| Experiment Accession | Experiment Title     | Organism                          | Platform            | Layout                    | Sample Accession: Sample Name    | Experiment Status       | Operation                                |
|----------------------|----------------------|-----------------------------------|---------------------|---------------------------|----------------------------------|-------------------------|--|
| CRX022919            | scRNA-seq of WJMSC27 | Homo sapiens                      | Illumina HiSeq 2500 | FRAGMENT                  | SAMC021024: scRNA-seq of WJMSC27 | Unchecked Confidential  | <span>Edit</span><br><span>Delete</span> |
| Run Accession        |                      | Run Title                         |                     | Run Data File Information |                                  |                         | Operation                                |
| CRR022892            | scRNA-seq of WJMSC27 | CS300_TGACCA_L002_R1_001.fastq.gz |                     |                           | Status: Unchecked                |                         | <span>Edit</span><br><span>Delete</span> |
| CRX022920            | scRNA-seq of WJMSC28 | Homo sapiens                      | Illumina HiSeq 2500 | FRAGMENT                  | SAMC021025: scRNA-seq of WJMSC28 | Checked OK Confidential | <span>Edit</span><br><span>Delete</span> |
| Run Accession        |                      | Run Title                         |                     | Run Data File Information |                                  |                         | Operation                                |
| CRR022893            | scRNA-seq of WJMSC28 | CS300_TGACCA_L002_R1_002.fastq.gz |                     |                           | Status: Checked OK               |                         | <span>Edit</span><br><span>Delete</span> |
| CRX022921            | scRNA-seq of WJMSC29 | Homo sapiens                      | Illumina HiSeq 2500 | FRAGMENT                  | SAMC021026: scRNA-seq of WJMSC29 | Checked OK Confidential | <span>Edit</span><br><span>Delete</span> |
| Run Accession        |                      | Run Title                         |                     | Run Data File Information |                                  |                         | Operation                                |
| CRR022894            | scRNA-seq of WJMSC29 | CS300_TGACCA_L002_R1_001.fastq.gz |                     |                           | Status: OK                       |                         | <span>Edit</span><br><span>Delete</span> |

Edit or delete Meta information for each submitted Experiment or Run

Total Items: 6 | Items of 1 - 6 | Page size: 20 | Page 1/1 | << First Last >> | Jump To: 1 GO

After the GSA data are archived (Status is Checked OK; confidential), you can click the **Submission ID** to enter the Overview page. On this page, you can 1) update the **Release Date** and **Title**; 2) edit the **Submitter information**; 3) Append data by clicking the ‘**Add Data**’ button, for more information, please see ‘[Create new GSA Submission](#)’; and 4) upload or update data file by clicking the ‘**Upload File**’ button. If you still want to change the **Metadata information** for each submitted Experiment or Run, please contact us at [gsa@big.ac.cn](mailto:gsa@big.ac.cn).

**Notice:** For more detail about submission status and the available operations, please go to ‘[Status and Operation](#)’.

[Create GSA](#)

| Accession | Submission ID | Title              | Release date | Status                     | Operation           |
|-----------|---------------|--------------------|--------------|----------------------------|---------------------|
| CRA000532 | subCRA000595  | Human dataset test | 2020-12-31   | Checked OK<br>Confidential | ReleaseNow<br>Share |

↓

**Basic Information**

Submission of GSA: subCRA000595 / Release Date: 2020-12-31 / Project: PRJCA000644

Status: Data Archived

Title:

Date Released:

[Update](#)

[Contact us](#)

**Submitter Information**

Submitter: Zhang Sisi  
[zhangss@big.ac.cn](mailto:zhangss@big.ac.cn)

Organization: Beijing Institute of Genomics, Chinese Academy of Sciences

Department: BIG Data Center

Country/Region: China

Address: Beichen West Road Beijing

Postal code: 100101

[Edit](#)

**Metadata**

Sample type: Human sample

Metadata file: [exp\\_run-Human.cn.test4.xlsx](#) added on 2019-06-17

[Update File](#) [Add Data](#)

| Experiment Accession | Experiment Title      | Organism                          | Platform            | Layout   | Sample Accession: Sample Name     | Experiment Status   | Operation |
|----------------------|-----------------------|-----------------------------------|---------------------|----------|-----------------------------------|---|-----------|
| CRX020324            | scRNA-seq of WJMISC21 | Homo sapiens                      | Illumina HiSeq 2500 | FRAGMENT | SAMC018351: scRNA-seq of WJMISC21 | Processed Succeed<br>Confidential                           |           |
| <b>Run Accession</b> | <b>Run Title</b>      | <b>Run Data File Information</b>  |                     |          |                                   | <b>Operation</b>  |           |
| CRR022201            | scRNA-seq of WJMISC21 | CS300_TGACCA_L002_R1_001.fastq.gz |                     |          |                                   | Status: <span style="color: blue;">Processed Succeed</span> |           |
| CRX020325            | scRNA-seq of WJMISC22 | Homo sapiens                      | Illumina HiSeq 2500 | FRAGMENT | SAMC018352: scRNA-seq of WJMISC22 | Processed Succeed<br>Confidential                           |           |
| <b>Run Accession</b> | <b>Run Title</b>      | <b>Run Data File Information</b>  |                     |          |                                   | <b>Operation</b>  |           |
| CRR022202            | scRNA-seq of WJMISC22 | CS300_TGACCA_L002_R1_002.fastq.gz |                     |          |                                   | Status: <span style="color: blue;">Processed Succeed</span> |           |
| CRX020326            | scRNA-seq of WJMISC23 | Homo sapiens                      | Illumina HiSeq 2500 | FRAGMENT | SAMC018353: scRNA-seq of WJMISC23 | Processed Succeed<br>Confidential                           |           |
| <b>Run Accession</b> | <b>Run Title</b>      | <b>Run Data File Information</b>  |                     |          |                                   | <b>Operation</b>  |           |
| CRR022203            | scRNA-seq of WJMISC23 | CS300_TGACCA_L002_R1_003.fastq.gz |                     |          |                                   | Status: <span style="color: blue;">Processed Succeed</span> |           |

Total Items: 6 | Items of 1 - 6 | Page size: 20 | Page 1/1 | [<< First](#) [Last >>](#) | Jump To: 1 [GO](#)

## How to Share Your Data

After logging into the BIG Sub system and entering the GSA submission portal, please find the **'Share'** button in the **'Operation'** column of this list as shown below.

| Create GSA |               | GSA Quick Start Guide (US) GSA Quick Start Guide (CN)                         |              |                            |                     |
|------------|---------------|---|--------------|----------------------------|---------------------|
| Accession  | Submission ID | Title   | Release date | Status                     | Operation           |
| CRA002926  | subCRA003595  | Full-length transcriptome sequencing from multiple tissues of annual ryegrass | 2020-07-29   | Checked OK<br>Confidential | ReleaseNow<br>Share |

Click the **'Share'** button, you can get the **'Shared URL'** as shown in the figure below. Send this URL to the editors for peer review.

| Create GSA |               | GSA Quick Start Guide (US) GSA Quick Start Guide (CN)                         |              |                            |  |
|------------|---------------|---|--------------|----------------------------|--|
| Accession  | Submission ID | Title   | Release date | Status                     | Operation  |
| CRA002926  | subCRA003595  | Full-length transcriptome sequencing from multiple tissues of annual ryegrass | 2020-07-29   | Checked OK<br>Confidential | ReleaseNow<br>Shared URL:<br><a href="http://bigd.big.ac.cn/gsa/s/T31EvCtu">http://bigd.big.ac.cn/gsa/s/T31EvCtu</a><br>Cancel Share |

**Notice: The Shared URL is a temporary link only for editors' data review. For your data security, please do not disclose this link to the public. Please click on the 'Cancel Share' button to cancel the data sharing when the data review is finished.**

---

## How to Release Your Data

After the article is published, you can click on the **'Release Now'** button in the **'Operation'** column of the list as shown below.

| Create GSA |               | GSA Quick Start Guide (US) GSA Quick Start Guide (CN)                         |              |                            |                     |  |
|------------|---------------|---|--------------|----------------------------|---------------------|--|
| Accession  | Submission ID | Title   | Release date | Status                     | Operation           |  |
| CRA002926  | subCRA003595  | Full-length transcriptome sequencing from multiple tissues of annual ryegrass | 2020-07-29   | Checked OK<br>Confidential | ReleaseNow<br>Share |  |

Click **'Yes'** in the **'confirmation box'** to trigger the release. The release of GSA will trigger the release of the related BioProject and BioSample(s), so you DO NOT need to release BioProject and BioSample in their respective system again.

Confirmation Box

Are you sure to **RELEASE** cra subCRA003595.

**Tip:**  
The system will spend several hours to archive files; temporarily you cannot find your data entry from our search system.  
后台将对文件归档处理，需要数小时，暂时不能搜索

Cancel Yes

It will take several hours to release a GSA dataset, depending on its data size. After they are released, all the data of the GSA dataset can be retrieved from the BIG Search portal within 14 hours.



---

## Data File Uploading

Three methods are offered for data uploading: Aspera Command Line, FTP and Aspera Connect plugin. Please choose one to upload your data. If you need any help during data file uploading, please contact the GSA Working Team at [gsa@big.ac.cn](mailto:gsa@big.ac.cn) or QQ group: [548170081](https://www.qq.com/group/548170081).

### NOTICE:

1. Unique file names should be used for all files, and each file must be listed in the GSA metadata file you uploaded.
2. Files must be compressed using gzip or bzip2.
3. Uploaded files will be removed after they are archived.

## Aspera Command Line

Use Aspera Command Line to upload files. You may use the following command to upload files via **Aspera Comand Line**:

```
[path/to/ascp/] -P33001 -i [path/to/key/file] -QT -l100m -k1 -d [path/to/folder/containing/files]  
aspsub@submit.big.ac.cn:uploads/[user dir]
```

### Where:

**[path/to/ascp/]:**

**Microsoft Windows:** C:\Program Files\Aspera\Aspera Connect\bin\ascp.exe

or

C:\users\[username]\AppData\Local\Programs\Aspera\Aspera Connect\bin\ascp.exe

**Mac OS X:** /Applications/Aspera/Connect.app/Contents/Resources/ascp (for admin's installation)

or

/Users/[username]/Applications/Aspera/Connect.app/Contents/Resources/ascp (for non-admin's installation)

---

**Linux:** /opt/aspera/bin/ascp or /home/[username]/aspera/connect/bin/ascp

**[path/to/key/file]** must be an absolute path, e.g.: /home/keys/aspera.openssh

**[path/to/folder/containing/files]** needs to specify the local folder that contains all the files to upload.

**[user dir]** user directory. You can click the **Submission ID** to enter the Overview page. On this page, click the 'Add Data' button and enter the "04 Files" page to find the **user directory** information.

### [Get the key file](#)

#### **Notice:**

- 1) Please make a new subdirectory for each new submission. Your submission subfolder is a temporary holding area and will be removed once the whole submission is complete.
- 2) Do not upload complex directory structures or files that do not contain sequence data.

## **FTP Upload**

We recommend to use an FTP client software to upload files. FTP Username and Password are same to the BIG Sub account.

**NOTICE:** Please use the binary mode to transfer files. If you are using the FTP client software (such as [FileZilla Client](#)) to log in to the FTP server, follow the tool's instruction to set the transfer mode; If you are using FTP command, type the **binary** command before the **mput** command.

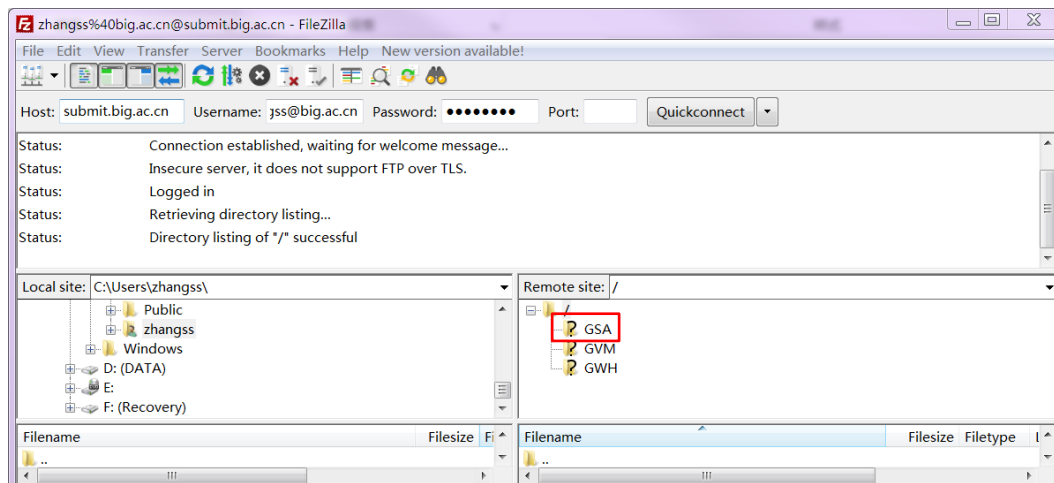
### **Transmitting your data files to the GSA FTP site**

**Address:** ftp://submit.big.ac.cn  
**User:** Same as you login the BIG Sub  
**Password:** Same as you login the BIG Sub

---

## After you connected to the FTP server:

1. Navigate to the GSA folder.
2. Upload your files here.



## Web browser upload via Aspera Connect plugin

Use the Aspera Connect plugin to upload files. If you do not have Aspera Connect plugin installed in your computer, please [Download Aspera Connect](#) and install it first.

1. On opening the File Upload page, your browser may ask you for permission to execute an external application called 'IBM Aspera Launcher', click 'Launch Application' to allow it.
2. After selecting to upload files via Aspera, click the 'Click to Upload Files' button to select files on your computer. You can select multiple files at one time.
3. After your files have been selected, a prompt will appear asking you to confirm the Aspera connection to BIGD, click 'allow' to start the file transfer.
4. You can select the 'Auto finish submission' checkbox to allow backend upload. This will allow you to turn to the next page before the uploading is completed. DO NOT close the Aspera Connect Transfer window when you are still uploading your files.

🔗 Web browser upload via Aspera Connect plugin

Use the Aspera Connect plugin to upload files. If you don't have Aspera Connect plugin installed in your computer, please [Download Aspera Connect and install it first](#).

On opening this page, your browser may ask you for permission to execute an external application called "IBM Aspera Launcher", click "Launch Application" to allow it.

After selecting to upload files via Aspera, click the "使用Aspera Connect插件上传文件" button to select files on your computer. You can select multiple files at one time.

After your files have been selected, a prompt will appear asking you to confirm the Aspera connection to BIGD, click "allow" to start the file transfer.

You can select the "Auto finish submission" checkbox to allow backend upload. This will allow you to turn to the next page before files are uploaded. DO NOT close the Aspera Connect Transfer window when you are still uploading your files.

**NOTE:** Only one Aspera connection should be created at any given moment.

Click the 'Click to Upload Files' button to select files on your computer.

All uploaded files:

| File Name | File Size (Bytes) |
|-----------|-------------------|
|-----------|-------------------|

Auto finish submission 'Auto finish submission' to allow backend upload.

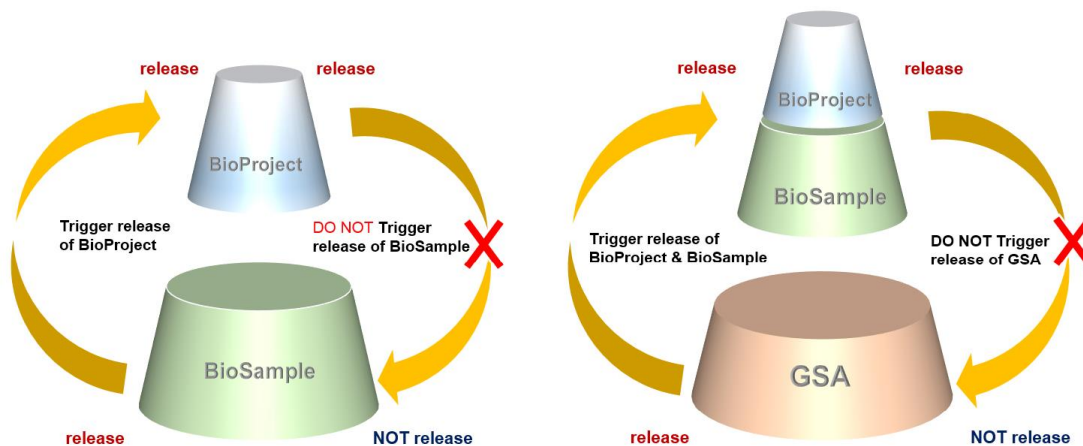
**Notice:** Only one Aspera connection should be created at any given moment.

---

## Release of linked BioProject/BioSample/GSA

Release rules of linked BioProject, BioSample, and GSA are as follows:

1. The release of the BioProject records DO NOT trigger the release of the other linked data.
2. The release of the BioSample records JUST triggers the release of its BioProject.
3. The release of the GSA nucleotide sequence data DO trigger the release of the linked BioProject and BioSample records.



**Notice:** Therefore, please carefully fill in the '**release time**' of a BioProject, BioSample and GSA. Once published, the representative data or information can be retrieved or downloaded by other users.

## Status and Operation

### GSA Status and Operation

[Create GSA](#)

| Accession                 | Submission ID                | Title              | Release date | Status  | Operation   |
|---------------------------|------------------------------|--------------------|--------------|---|---|
| Unassigned                | <a href="#">subCRA000654</a> | human              | 2019-08-06   | <i>Unfinished at the OverView step<br/>Confidential</i>   | <a href="#">Delete</a>                              |
| Unassigned                | <a href="#">subCRA000653</a> | Human dataset 1    | 2020-12-31   | <i>Unchecked<br/>Confidential</i>                         | <a href="#">Delete</a>                              |
| Unassigned                | <a href="#">subCRA000652</a> | microbe dataset    | 2019-08-05   | <i>Unfinished at the Attributes step<br/>Confidential</i> | <a href="#">Delete</a>                              |
| Unassigned                | <a href="#">subCRA000651</a> | Human dataset 2    | 2020-12-31   | <i>Deleted<br/>Confidential</i>                           |   |
| <a href="#">CRA000532</a> | <a href="#">subCRA000695</a> | Human dataset test | 2020-12-31   | <i>Checked OK<br/>Confidential</i>                        | <a href="#">ReleaseNow</a><br><a href="#">Share</a> |

| No. | Status                                     | Description  | Operation  |
|-----|--|--|--|
| 1   | <b>Unfinished at the General Info step</b> | Finished the Submitter step and enter the general info step.   | Edit <sup>[1]</sup> ; Delete   |
| 2   | <b>Unfinished at the Sample Type step</b>  | Finished the General info step. If not created GSA related Biological Sample(s), enter the Sample type step. | Edit <sup>[1]</sup> ; Delete   |
| 3   | <b>Unfinished at the Attributes step</b>   | Finished Sample type step, enter the Attributes step   | Edit <sup>[1]</sup> ; Delete   |
| 4   | <b>Unfinished at the Metadata step</b>     | Finished the Attributes step, enter the GSA metadata step.   | Edit <sup>[1]</sup> ; Delete   |
| 5   | <b>Unfinished at the File Upload step</b>  | Finished the GSA metadata step, enter the File Upload step.  | Edit <sup>[1]</sup> ; Delete   |
| 6   | <b>Unfinished at the Overview step</b>     | Enter the overview step.   | Edit <sup>[1]</sup> ; Delete   |
| 7   | <b>Unchecked</b>                           | All the information are submitted, waiting for check.  | Edit <sup>[1]</sup> ; Delete   |
| 8   | <b>Checked failed</b>                      | Data file(s) processed error.  | Edit <sup>[1]</sup> ; Delete; Reload data file via FTP or Aspera Command |

|    |                   |  |                     |
|----|-------------------|--|---------------------|
|    |                   |  | Line <sup>[2]</sup> |
| 9  | <b>Checked OK</b> | Data file(s) Processed succeed and GSA Accession number is assigned. | Release Now; Share  |
| 10 | <b>Deleted</b>    | Deleted  |                     |

[1]: You can click the **GSA Submission ID** to enter the Overview page to edit GSA related metadata. For more detail, please see '[How to Edit, Delete or Add New Data](#)'.

[2]: For more details for data file upload, please see '[Data File Upload](#)'.

## Experiment Status and Operation

| Experiment Accession | Experiment Title     | Organism                          | Platform            | Layout   | Sample Accession: Sample Name    | Experiment Status             | Operation                                      |
|----------------------|----------------------|-----------------------------------|---------------------|----------|----------------------------------|-------------------------------|--|
| CRX022949            | scRNA-seq of WJMISC1 | <i>Homo sapiens</i>               | Illumina HiSeq 2500 | FRAGMENT | SAMC021054: scRNA-seq of WJMISC1 | <b>Unchecked Confidential</b> | <a href="#">Edit</a><br><a href="#">Delete</a> |
| Run Accession        | Run Title            | Run Data File Information         |                     |          |                                  | Operation                     |  |
| CRR022922            | scRNA-seq of WJMISC1 | CS300_TGACCA_L002_R1_001.fastq.gz |                     |          |                                  | Status: <b>Unchecked</b>      | <a href="#">Edit</a><br><a href="#">Delete</a> |

| No. | Status                | Description                               | Operation                    |
|-----|-----------------------|---|------------------------------|
| 1   | <b>Unchecked</b>      | Metadata submitted and waiting for check. | Edit <sup>[1]</sup> ; Delete |
| 2   | <b>Checked OK</b>     | Metadata Checked OK                       | Edit <sup>[1]</sup> ; Delete |
| 3   | <b>Checked failed</b> | Metadata Checked failed                   | Edit <sup>[1]</sup> ; Delete |
| 5   | <b>Deleted</b>        | Deleted                                   |                              |

[1]: You can click the **GSA Submission ID** to enter the Overview page to edit the Experiment metadata. For more details, please see '[How to Edit, Delete or Add New Data](#)'.

## Run Status and Operation

| Experiment Accession | Experiment Title     | Organism                          | Platform            | Layout   | Sample Accession: Sample Name    | Experiment Status             | Operation                                      |
|----------------------|----------------------|-----------------------------------|---------------------|----------|----------------------------------|-------------------------------|--|
| CRX022949            | scRNA-seq of WJMISC1 | <i>Homo sapiens</i>               | Illumina HiSeq 2500 | FRAGMENT | SAMC021054: scRNA-seq of WJMISC1 | <b>Unchecked Confidential</b> | <a href="#">Edit</a><br><a href="#">Delete</a> |
| Run Accession        | Run Title            | Run Data File Information         |                     |          |                                  | Operation                     |  |
| CRR022922            | scRNA-seq of WJMISC1 | CS300_TGACCA_L002_R1_001.fastq.gz |                     |          |                                  | Status: <b>Unchecked</b>      | <a href="#">Edit</a><br><a href="#">Delete</a> |

| No. | Status           | Description                               | Operation                    |
|-----|------------------|---|------------------------------|
| 1   | <b>Unchecked</b> | Metadata submitted and waiting for check. | Edit <sup>[1]</sup> ; Delete |

|   |                          |  |  |
|---|--------------------------|--|--|
| 2 | <b>Checked OK</b>        | Metadata Checked OK                                    | Edit <sup>[1]</sup> ; Delete   |
| 3 | <b>Checked failed</b>    | Metadata Checked failed                                | Edit <sup>[1]</sup> ; Delete   |
| 4 | <b>Uploaded Succeed</b>  | Data file(s) uploaded succeed, waiting for processing. |  |
| 5 | <b>Processing</b>        | Data file(s) under processing.                         |  |
| 6 | <b>Processed succeed</b> | Data file(s) processed succeed                         |  |
| 7 | <b>Processed error</b>   | Data file(s) processed error                           | Edit <sup>[1]</sup> ; Delete; Reload data file via FTP or Aspera Command Line <sup>[2]</sup> |
| 8 | <b>Deleted</b>           | Deleted  |  |

<sup>[1]</sup>: You can click the **GSA Submission ID** to enter the Overview page to edit Run metadata. For more detail, please see '[How to Edit, Delete or Add New Data](#)'.

<sup>[2]</sup>: For more detail for data file upload, please see '[Data File Upload](#)'.