GWH Submission Quick Start Guide

Overview

[Genome Sequence Submission](#) ([Gsub](#)) offers 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.

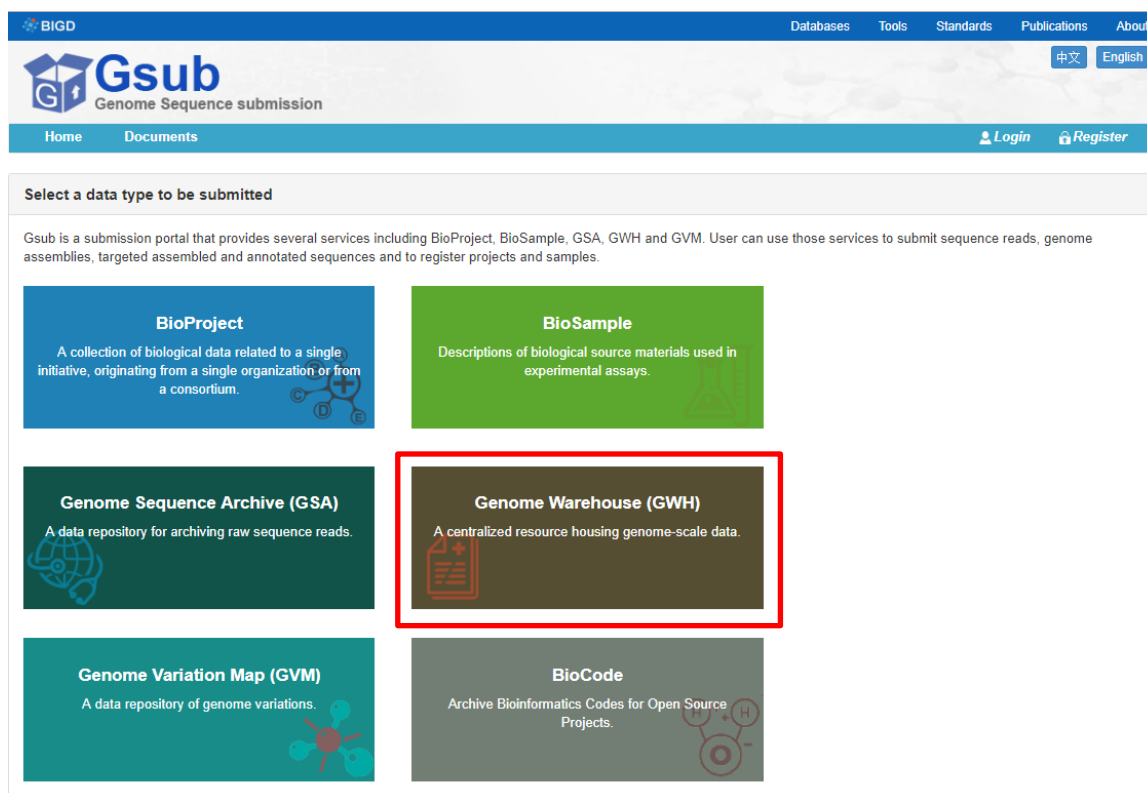[The Genome Warehouse](#) (**[GWH](#)**) stores genome assembly associated data (Figure 1).

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

Login to the Gsub

- Click the login button in [Gsub](#), and then enter your user name and password to login. If you do not have an account already, click the Register button to create one (Figure 2). If you have used an account in the past but no longer see your previous submissions, please contact us at gwh@big.ac.cn for assistance with your account view.

Figure 2. Login Gsub for Submission

- **DO NOT** suggest using the Win10 operating system; Recommend the use of Firefox/Google Chrome browser, other browsers may have bugs.
- After the activation of the login system, use our [Genome Sequence Submission](#) ([Gsub](#)) system and follow steps to finish the submission.

Create a GWH Submission

- ### Enter GWH submission system
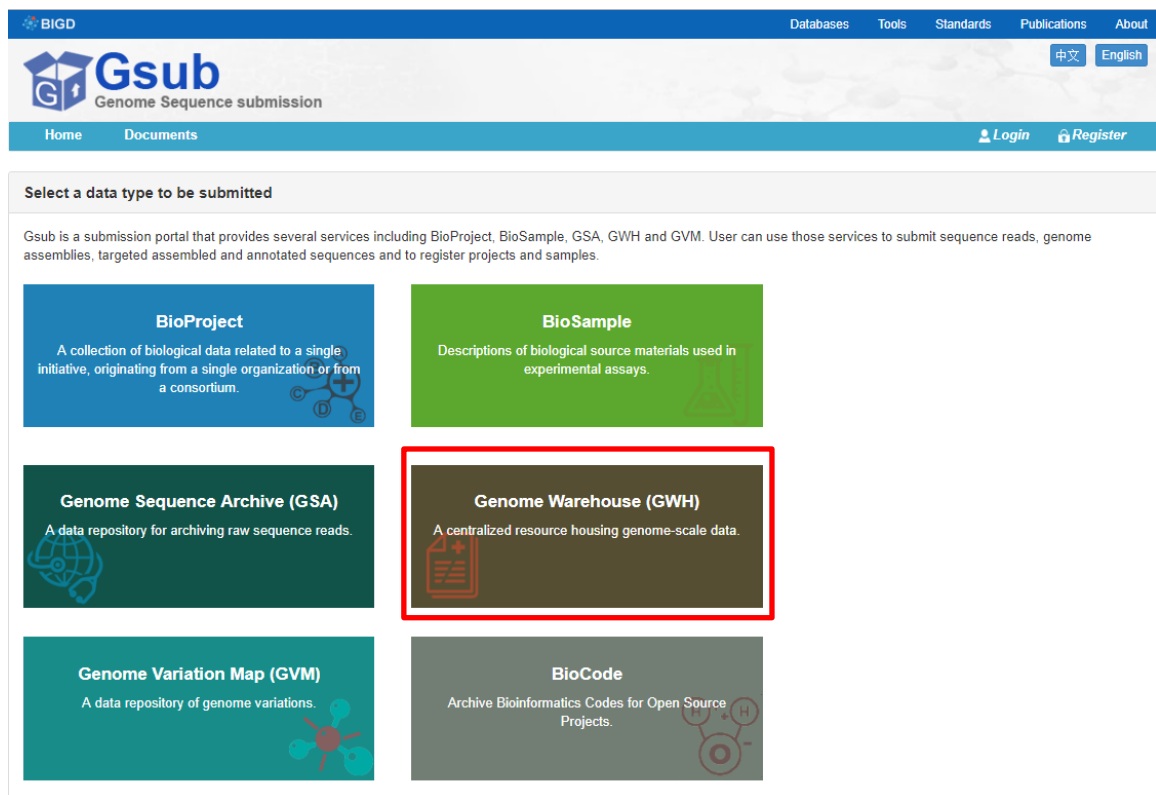 Click **GWH** to enter GWH submission system (Figure 3).

Figure 3. Login GWH submission system

- ### Create and Describe New GWH Submission

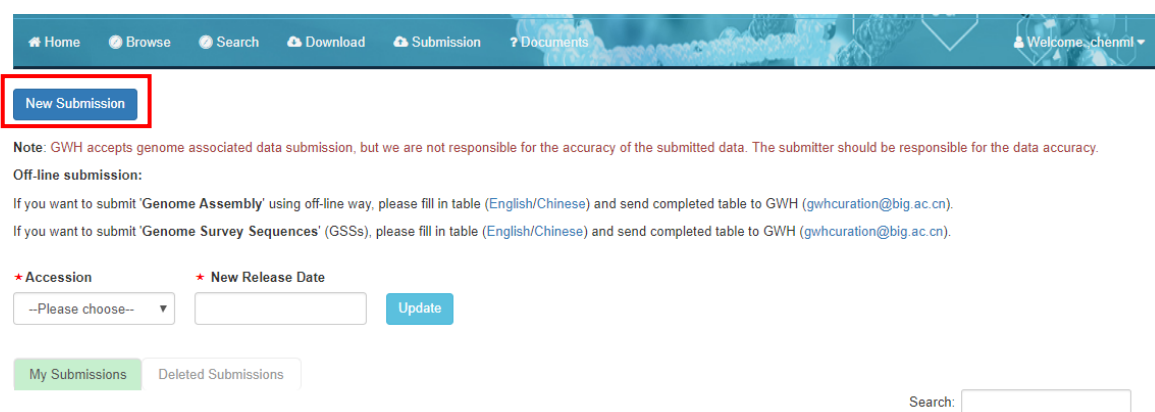 Click the 'New Submission' button to create a new **GWH** Submission (Figure 4).

Figure 4. Click the 'New Submission' button to create new **GWH** submission

Step 1	Step 2	Step 3	Step 4	Step 5	Step 6
Submitter	General Info	Files	Assignment	Reference	Overview

Submitter

* First name
Meili

Middle name

* Last name
Chen

* Email
chenml@big.ac.cn

Email (secondary)

* Organization
Beijing Institute of Genomics, Chir

Submit organization Url
http://

* Department
The CAS Key Laboratory of Geno

Phone

Fax

* Street
No.1 Beichen West Road, Chaoya

* City
Beijing

State/Province

* Postal code
100010

* Country/Region
China

Save and forward

| Step 1 Submitter | Step 2 General Info | Step 3 Files | Step 4 Assignment | Step 5 Reference | Step 6 Overview |

BioProject & BioSample

Pre-Register BioProject ID | Register a BioProject if you havn't it.

* **BioProject accession**

--Please choose-- ▼   OR Go to create **BioProject**

The BioProject bundles the data for this research project, which is an overall description of a single research initiative.

* **BioSample accession**

[   ] ▼   OR Go to create **BioSample**

The BioSample is a description of the biological source material that was sequenced.

Pre-Register BioSample ID | Register a BioSample if you havn't it.

Release Date

○ Release immediately following curation (**recommended**)
○ Release on specified date

Note : Release of BioProject and BioSample is also triggered by the released of WGS-associated data.

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 genome portal (**http://bigd.big.ac.cn/gwh/submit/submission**).
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, GWH will release sequence data on the specified date.
4. As soon as it is available, please *send the full publication data*--all authors, title, journal, volume, pages and date--to the following address: **gwh@big.ac.cn**.

⦿ I accept it.   ○ I don't accept it.

Genome assembly info

* **Is it an update of existing submission?**
  ○ Yes   ○ No

* **Is it a *de novo* assembly?**
  ○ Yes   ○ No

Assembly date ⍰
[   ]

* **Assembly method** ⍰   * **Program version or release date** ⍰
[   ]   [   ]   ⊕Add ⊖Delete

* **Assembly name** ⍰
[   ]

* **Sequencing technology** ⍰   * **Genome coverage** ⍰
[   ]   [   ]   ⊕Add ⊖Delete

Sequencing reads accession
[   ]

Fill in with corresponding assembly program and version/released date.

Fill in with corresponding platform of sequencing and genome coverage.

* **Your sample represents the genome as:**
  ○ Full genome (For a prokaryotic genome, the assembly may not include plasmid)
  ○ Partial genome
  ○ Other

The composition of your genome

The composition of your genome:

Chromosome	Ploidy (n) ⍰	Plasmid	Mitochondrio
eg: 22+XY	eg: 2	eg: 2	---

The detail composition of your genome

Other info

* **WGS Submission Title** ⍰
[   ]

Set a title to make it Easy to be recognized

Private message to GWH staff ⍰
[   ]

[Save and next]

Step 1	Step 2	Step 3	Step 4	Step 5	Step 6
Submitter	General Info	Files	Assignment	Reference	Overview

★ Assembly level

- ○ Completed genome
- ● Draft genome in chromsome level
- ○ Draft genome in scaffold level
- ○ Draft genome in contig level

Choose assembly level for your submitted genome

★ Files

★ Submission method ⑦

- ○ Online   ● ftp

★ Genome sequence file (.fasta, .fa, .fsa, .gz or .bz2):

★ Filename
Please Input File Name

★ MD5 ⑦

You can submit your genome-associated files by online directly or ftp way.

Generation method of MD5 code: we suggest to use md5sum + file command with Linux users, and use md5sums.exe or WinMD5Free tools with windows users.

Annotation file (.tbl, .gff, .gz or .bz2)

Filename
Please Input File Name

MD5 ⑦

Genetic code ⑦ ❶

Set a genetic code to make validation for genome annotation more accuracy

★ Assembly level for the AGP file

★ Do you submit AGP file ?

- ○ Yes   ○ No
- ○ Contig→Scaffold→Chromosome
- ● Contig→Chromosome

Choose corresponding AGP file type if your genome assembly level is draft genome in chromosome/scaffold level.

Assembly level   **AGP file**

Chromosome from contigs:

★ Filename
Please Input File Name

★ MD5 ⑦

Note :
Sequences concatenated in unknown order are not allowed.
Chromosome from contigs level means minimum unit of sequence in the AGP file is contig.

★ Do you need our validation check processes?

- ○ Yes. Then our staff will do and feedback contamination and annotation validation che... resubmitted.
- ○ No. Then validation check processes will be skipped. The release data will be marked...

Decide to do strict quality control processing or not.

Save and next

Step 1	Step 2	Step 3	**Step 4**	Step 5	Step 6
Submitter	General Info	Files	**Assignment**	Reference	Overview

Gaps assignment

* **Whether N's in your genome sequences represent gaps?**

◉ Yes   ○ No

* **The Ns length of a gap sequence is:**

○ Real length   ◉ Estimated length

* **Minimum number of N's sequence represents an estimated length of gap:**

[10 ▼]

* **The type of linkage evidence that assemble two sequences with a gap.**

◉ Pair-ends   ○ Gene   ○ Others

> Illustrate gap information about what the N's represent if your genome sequence contains it.

Chromosome assignment

* **Do any sequences can be assigned to an exact chromosome?**

◉ Yes   ○ No

* **Do you upload csv files of the assignments? (chr.csv format and example)**

◉ Yes   ○ No

* **Filename**   * **MD5**

[chr.csv]   [7a93c3efab1e199b7cfca96688fba6a ✔]

> Indicate any sequence that is/belongs to a chromosome by the way of online or submitting a file with ftp.

Plasmid assignment

* **Do any sequences can be assigned to an exact plasmid?**

◉ Yes   ○ No

* **Do you upload csv files of the assignments? (plasmid.csv format and example)**

○ Yes   ◉ No

* **Sequence ID**	* **Plasmid name**	* **Complete**	* **Circular**	**Add**	**Delete**
[seq2]	[pB2]	[true ▼]	[false ▼]	●	●

> Indicate any sequence that is/belongs to a plasmid by the way of online or submitting a file with ftp.

Organella assignment

* **Do any sequences can be assigned to a mitochondrion or chloroplast?**

◉ Yes   ○ No

* **Do you upload csv files of the assignments? (organella.csv format and example)**

○ Yes   ◉ No

* **Sequence ID**	* **Type**	* **Complete**	* **Circular**	**Add**	**Delete**
[seq1]	[Mitochondrion ▼]	[true ▼]	[true ▼]	●	●

> Indicate any sequence that is/belongs to a organella by the way of online or submitting a file with ftp.

[Save and forward]

⬇

Step 1	Step 2	Step 3	Step 4	Step 5	Step 6
Submitter	General Info	Files	Assignment	Reference	Overview

★ Sequence authors

Indicate the authors of your submitting genome sequence.

★ First name	Middle name	★ Last name	★ Email	
Meili ✔		Chen ✔	chenml@big.ac.cn ✔	⊕Add ⊖Delete

★ Reference

★ Publication status

○Unplished   ⦿In-press   ○Published   ✔

Indicate the paper that you (plan to) publish your genome sequence.

★ Reference title

Whole genome sequencing of wolf

★ Journal title

Scientific reports   ✔

★ Year	Volumn	Issue	Pages from	Pages to
2018	14 ✔		e10 ✔	

★ Reference authors
- ⦿ Same as sequence authors
- ○ New authors:

[Save and forward]

- **Data Transfer**

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

> **Transmitting your data files to the Gsub FTP site**
> **Host address**: ftp://submit.big.ac.cn
> **User name**: Same as you login the Gsub
> **Password**: Same as you login the Gsub
> **Path**: /GWH/WGSXXXXXX (submission ID)
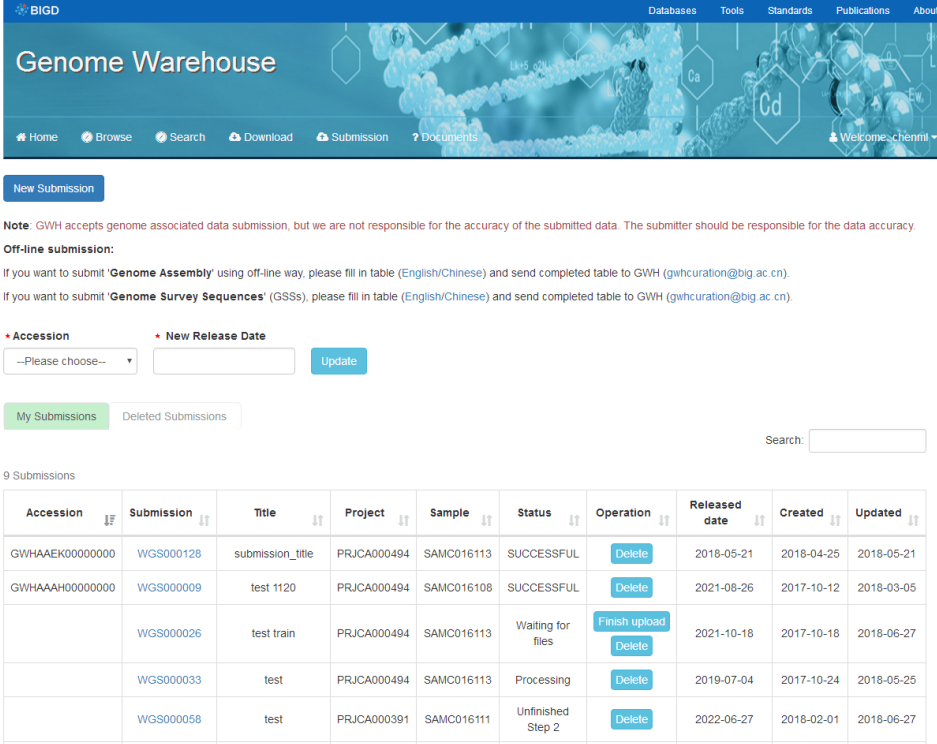
Note: Please do not update your file during uploading. If you submit files by ftp, please check the upload status and make sure that it is "sucessful" (or "finished"). After file submission, please log back into genome submission portal, click the "**Finish upload**" button in "My Submissions" list to inform us to check your submission.

Submission List

After the metadata is completed, the submission list is displayed on the page (Figure 5, http://bigd.big.ac.cn/gwh/submit/submission).

Figure 5. GWH Submission list

Notes:

[1] Release Date is the date your submission can be public access. If your genome data needs to be **protected** for some times, please set this date two or more years later. If a paper citing the sequence or accession number is published *prior to the specified date*, the sequence will *be released upon publication*. Otherwise, GWH will release sequence data on the specified date.If your genome data is released, the release of **BioProject** and **Biosample** are also **Trigged**.

[2] The release date can be changed through the genome portal (http://bigd.big.ac.cn/gwh/submit/submission).

[3] You can revise your information in any step before you submit in Overview page. However, you are **NOT** able to revise your information after submitting. In this way, please check carefully before clicking the 'Finish' button. Please send email to gwh@big.ac.cn if you have any question or revision about your submission.

[4] Please do not update your file during uploading. If you submit files by ftp, please check the upload status and make sure that it is "sucessful" (or "finished"). After file submission, please log back into genome submission portal, click the "**Finish upload**" button in "My Submissions" list to inform us to check your submission.

Quality Control and Feedback

After finished all above tasks, Gsub system will check your submitted files and send an email for feedback validation result in a few days.

(1) If the submit files pass quality control system, the submission status will turn into successful. And GWH will assign an unique genome assembly accession No. for the submission.

(2) On the contrary, the submission status will turn into "unfinished". And GWH will send error report to submitter through email. Submitter needs to check and be sure that all errors in the attached report files are fixed, and resubmits the corrected files at your convenience.