

GWH Data Items

Version beta July 2017

Version	Authors	Date	Comments
beta	Meili Chen Yingke Ma Fan Wang	July 2017	Data items summarized

1. Submitter	3
2. General Info.....	3
3. Files.....	5
4. Assignment.....	6
5. Reference	8

1. Submitter

- **Submitter**

- * **First name**

First name of submitter.

- Middle name**

Middle name of submitter.

- * **Last name**

Last name of submitter.

- * **Email**

Primary Email address of submitter.

- Email (secondary)**

Secondary Email address of submitter.

- * **Organization**

Full name of submitter's organization.

- Organization URL**

The URL of submitter's organization.

- * **Department**

The department of submitter.

- Phone**

The phone number of submitter.

- Fax**

The Fax number of submitter.

- * **Street**

The Street.

- * **City**

The City.

- State/Province**

The State/Province.

- * **Postal code**

The Postal code.

- * **Country/Region**

The Country/Region of submitter.

2. General Info

- **BioProject & BioSample**

- * **BioProject accession**

Select the BioProject ID that describes a single research initiative associated your sequencing genome. If you do not have any BioProject ID or do not have a proper BioProject to be assigned to your genome, you can go to create a new BioProject.

- * **BioSample accession**

Select the BioSample ID that describes the biological source material that is sequenced. If you do not have any BioSample ID or do not have a proper BioSample to be assigned to your genome, you can go to create a new BioSample.

- **Release date**

- * **Release date**

Select “Release immediately following curation” or “Release on specified date”.

If you select “Release on specified date”, please provide release date in format (YYYY-MM-DD).

- * **Release Policies and Disclaimers**

Select “I accept it.” or “I don’t accept it.”.

If select “I don’t accept it.”, subsequent operations will not continue.

- **Genome assembly info**

- * ***De novo* genome assembly**

Indicate your genome assembly method is a *de novo* method or not.

- * **Reference assembly**

If this is NOT a *de novo* assembly, you will need to provide the accession.version and/or the assembly name of the genome assembly that was used as the reference guide for this assembly.

- * **Update of existing submission**

Indicate your genome assembly is an update of an existing genome assembly or not.

- * **Existing genome accession**

If this is an update version of a existing assembly, please provide the accession of the genome being updated, when appropriate.

Assembly date

The date for the assembly was made. Date format: YYYY-MM-DD; YYYY-MM; YYYY (eg: 2015-12).

- * **Assembly method**

Select the algorithm/software that is used for the genome assembly. Input the name that is not included in the list.

- * **Program version or release date**

The program version or the date that program was released. Date format: YYYY-MM-DD; YYYY-MM; YYYY (eg: 1.0 or 2015-12).

Assembly name

A short name that describes your assembly method and version or the combination strategy (eg: HGAP2.0PBjelly1.0).

- * **Sequencing technology**

The sequencing platform and instrument model. Input the name that is not included in the list.

- * **Genome coverage**

The sequencing depth that corresponds to the sequencing technology.

Sequencing reads accession

Raw genome sequencing reads accession that reads were used to assemble the genome.

- * **Your sample represents the genome as**

Select the genome represent types for your sample.

- ‘Full genome’ is nearly always for whole genome sequencing;
- ‘Partial genome’ only if a subset of the sample/genome was deliberately selected,

- eg: mitochondria or chloroplast;
- 'Other': a subset of the sample/genome was deliberately selected and the content is not included in the partial genome list.

The composition of your genome

Chromosome

Genome contains the numbers of autosomes and sex chromosomes. Format: autosome+sex chromosome, eg: 22+XY.

Ploidy (n)

The number of sets of chromosomes in a cell.

Plasmid

Genome contains the number of plasmids.

Mitochondrion

Genome contains mitochondrion or not.

Chloroplast

Genome contains chloroplast or not.

Other

Genome contains any other content except chromosome, plasmid, mitochondrion, and chloroplast.

- Other info

*** WGS submission Title**

The title is used to distinguish your submissions in your submission list.

Private message to GWH staff

Input your private requirements or queries to our GWH curation staff. The content will not be public used or released.

3. Files

- **Assembly level**

*** Assembly level**

Select the genome assembly level from the list.

- Completed genome: does not contain any gap sequence, and each sequence has an assignment with complete=true.
- Draft genome in chromosome level: allow gap sequence, and at least one sequence has an assignment with complete=true.

- **Files**

*** Submission method**

Select submission method: online or ftp.

*** Genome sequence file (.fasta)**

- *** Filename**

File name of genome sequence file in fasta format.

- *** MD5**

MD5 checksums are a 32-character alphanumeric string.

For Linux users, use: \$ **md5sum** filename;

For Mac users, use: \$ **md5** filename;

For Windows users, use: \$ **certutil -hashfile** filename **MD5**; and combine the

code by removing the spaces. Or use third party tool.

Annotation file (.tbl or .gff)

- **Filename**
File name of genome annotation file in gff/tbl format.
- **MD5**
MD5 checksums are a 32-character alphanumeric string.
For Linux users, use: \$ **md5sum** filename;
For Mac users, use: \$ **md5** filename;
For Windows users, use: \$ **certutil -hashfile** filename **MD5**; and combine the code by removing the spaces. Or use third party tool.
- **Genetic code**
Indicate the set of your genetic material (DNA or mRNA sequences) translation rules. Please see the detail explanation of [genetic code table](#) on NCBI.

Assembly level for the AGP file

- *** Do you submit AGP file?**
Indicate your genome assembly contains any AGP file that is genome sequence ordering and orientation information file or not.
 - *** Filename**
File name of genome sequence ordering and orientation information file in AGP format.
 - *** MD5**
MD5 checksums are a 32-character alphanumeric string.
For Linux users, use: \$ **md5sum** filename;
For Mac users, use: \$ **md5** filename;
For Windows users, use: \$ **certutil -hashfile** filename **MD5**; and combine the code by removing the spaces. Or use third party tool.
- * Do you need our validation check processes?**
Select your choice to do initial validation and sequence contamination processing or not.

4. Assignment

- **Gap assignment**
 - * Whether N's in your genome sequences represent gaps?**
Decide N-sequence in your genome sequences represents gap or variation sequence.
 - * The Ns length of a gap sequence is**
Decide the N's length of a gap represents real length or estimated length.
 - * Minimum number of N's sequence represents an estimated length of gap**
Select the minimum number of N's sequence represents an estimated length of a gap.
 - * The type of linkage evidence that assemble two sequences with a gap**
Linkage evidence type.
 - * Others**
Input a linkage evidence type that is not included the selection list.
- **Chromosome assignment**

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

Select 'yes' if some of your genome sequence can be assigned to an exact chromosome.

*** Do you upload csv files of the assignments?**

Select 'yes' if the chromosome assignments were filled in a table file; Select 'No' if the chromosome assignments would be filled online.

*** Sequence ID**

Sequence that can be assigned to an exact chromosome. The sequence ID should be from genome sequence ID and be unique in the assignments.

*** Chromosome name**

Assign a chromosome name that is a positive integer/capital letter.

*** Complete**

The chromosome sequence is complete or not.

*** Circular**

The chromosome sequence is circular or not.

• **Plasmid assignment**

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

Select 'yes' if some of your genome sequence can be assigned to an exact plasmid.

*** Do you upload csv files of the assignments?**

Select 'yes' if the plasmid assignments were filled in a table file; Select 'No' if the plasmid assignments would be filled online.

*** Sequence ID**

Sequence that can be assigned to an exact plasmid. The sequence ID should be from genome sequence ID and be unique in the assignments.

*** Plasmid name**

Assign a plasmid name:

(1) allowed characters: digits, dots, underscores, and letters;

(2) start with lower case 'p' or 'unnamed';

(3) not include the word 'plasmid';

(4) length: ≤ 20 characters.

*** Complete**

The plasmid sequence is complete or not.

*** Circular**

The plasmid sequence is circular or not.

• **Organelle assignment**

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

Select 'yes' if some of your genome sequence can be assigned to an exact organelle.

*** Do you upload csv files of the assignments?**

Select 'yes' if the organelle assignments were filled in a table file; Select 'No' if the organelle assignments would be filled online.

*** Sequence ID**

Sequence that can be assigned to an exact organelle. The sequence ID should be from genome sequence ID and be unique in the assignments.

*** Type**

Select organella type.

* **Complete**

The organella sequence is complete or not.

* **Circular**

The organella sequence is circular or not.

5. Reference

- **Sequence authors**

- * **First name**

- First name of sequence author.

- Middle name**

- Middle name of sequence author.

- * **Last name**

- Last name of sequence author.

- * **Email**

- Email address of sequence author.

- **Reference**

- * **Publication status**

- Select status of the reference that reports the submitted genome from the list.

- * **Publication source**

- Select source of the reference from the list.

- * **PubMed ID**

- Input the PubMed ID of the reference.

- * **URL**

- Input the URL address of the reference.

- * **Reference title**

- The title of the reference.

- * **Journal title**

- Journal title of the reference.

- * **Year**

- Reference public release date (format: yyyy).

- * **Volume**

- Volume No. of the reference.

- Issue**

- Issue No. of the reference.

- * **Pages from**

- The start page No. of the reference.

- Pages to**

- The end page No. of the reference.

- * **Reference authors**

- Select the source of reference author information.

- New authors:**

- * **First name**

- First name of reference author.

- **Middle name**
Middle name of reference author.
- * **Last name**
Last name of reference author.
- * **Email**
Email address of reference author.