



BIG Data Center

国家基因组科学数据中心 科学计划数据汇交 培训会

2021-05-12



中国科学院北京基因组研究所
BEIJING INSTITUTE OF GENOMICS CHINESE ACADEMY OF SCIENCES



National Genomics Data Center

GVM Data Submission

GVM Group

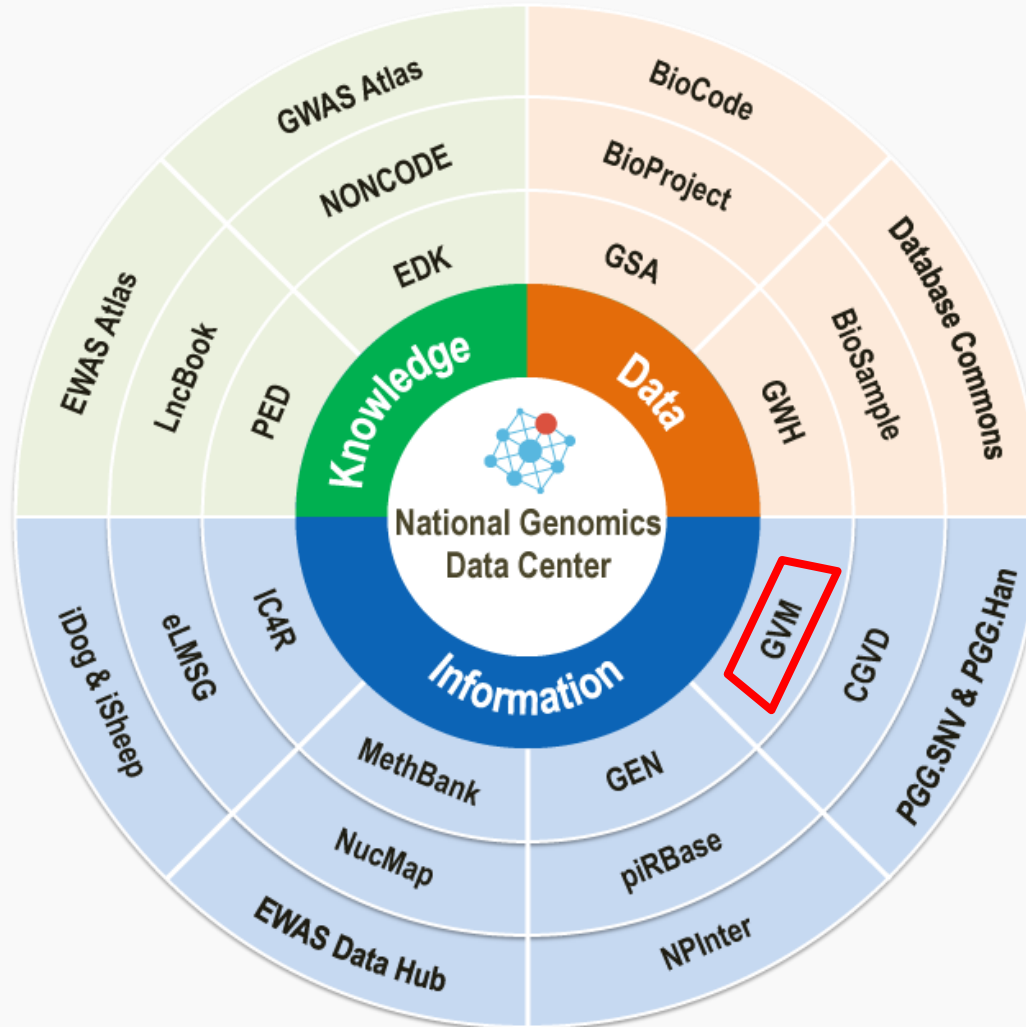
National Genomics Data Center
Beijing, China



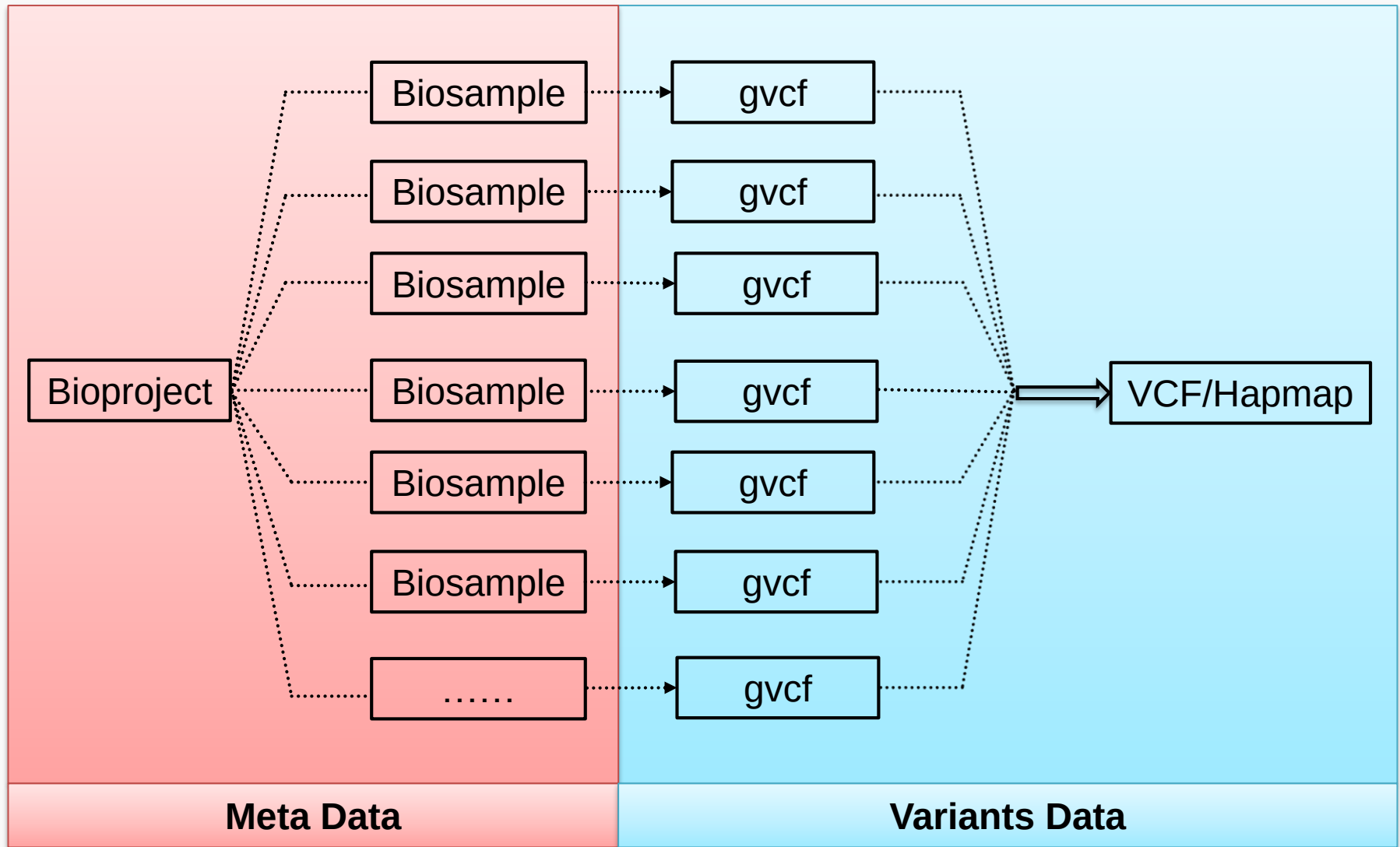
中国科学院北京基因组研究所
BEIJING INSTITUTE OF GENOMICS CHINESE ACADEMY OF SCIENCES

GVM: Genome Variation Map Database

<https://bigd.big.ac.cn/gvm/>



数据结构



准备工作

- Meta 信息，有固定模板，在线填写，不需要提前准备；
- 变异文件，要求是gvcf、vcf或hapmap格式，需提前准备，并计算好文件的MD5值

File Format--VCF

VCF is a text file format (most likely stored in a compressed manner). It contains meta-information lines, a header line, and then data lines each containing information about a position in the genome. The format also has the ability to contain genotype information on samples for each position.

1.1 An example

```
##fileformat=VCFv4.2
##fileDate=20090805
##source=myImputationProgramV3.1
##reference=file:///seq/references/1000GenomesPilot-NCBI36.fasta
##contig=<ID=20,length=62435964,assembly=B36,md5=f126cdf8a6e0c7f379d618ff66beb2da,species="Homo sapiens",taxonomy=x>
##phasing=partial
##INFO=<ID=NS,Number=1,Type=Integer,Description="Number of Samples With Data">
##INFO=<ID=DP,Number=1,Type=Integer,Description="Total Depth">
##INFO=<ID=AF,Number=A,Type=Float,Description="Allele Frequency">
##INFO=<ID=AA,Number=1,Type=String,Description="Ancestral Allele">
##INFO=<ID=DB,Number=0,Type=Flag,Description="dbSNP membership, build 129">
##INFO=<ID=H2,Number=0,Type=Flag,Description="HapMap2 membership">
##FILTER=<ID=q10,Description="Quality below 10">
##FILTER=<ID=s50,Description="Less than 50% of samples have data">
##FORMAT=<ID=GT,Number=1,Type=String,Description="Genotype">
##FORMAT=<ID=GQ,Number=1,Type=Integer,Description="Genotype Quality">
##FORMAT=<ID=DP,Number=1,Type=Integer,Description="Read Depth">
##FORMAT=<ID=HQ,Number=2,Type=Integer,Description="Haplotype Quality">
#CHROM POS ID REF ALT QUAL FILTER INFO FORMAT NA000001 NA000002 NA000003
20 14370 rs6054257 G A 29 PASS NS=3;DP=14;AF=0.5;DB;H2 GT:GQ:DP:HQ 0/0:48:1:51,51 1/0:48:8:51,51 1/1:43:5:.,.
20 17330 . T A 3 q10 NS=3;DP=11;AF=0.017 GT:GQ:DP:HQ 0/0:49:3:58,50 0/1:3:5:65,3 0/0:41:3
20 1110696 rs6040355 A G,T 67 PASS NS=2;DP=10;AF=0.333,0.667;AA=T;DB GT:GQ:DP:HQ 1/2:21:6:23,27 2/1:2:0:18,2 2/2:35:4
20 1230237 . T . 47 PASS NS=3;DP=13;AA=T GT:GQ:DP:HQ 0/0:54:7:56,60 0/0:48:4:51,51 0/0:61:2
20 1234567 microsat1 GTC G,GTCT 50 PASS NS=3;DP=9;AA=G GT:GQ:DP 0/1:35:4 0/2:17:2 1/1:40:3
```

This example shows (in order): a good simple SNP, a possible SNP that has been filtered out because its quality is below 10, a site at which two alternate alleles are called, with one of them (T) being ancestral (possibly a reference sequencing error), a site that is called monomorphic reference (i.e. with no alternate alleles), and a microsatellite with two alternative alleles, one a deletion of 2 bases (TC), and the other an insertion of one base (T). Genotype data are given for three samples, two of which are phased and the third unphased, with per sample genotype quality, depth and haplotype qualities (the latter only for the phased samples) given as well as the genotypes. The microsatellite calls are unphased.

File Format--GVCF

格式同VCF相似，利用GATK软件对每个样本进行变异鉴定时生成的中间文件，**一般不需要用户自己手动制作。**

```
##contig=<ID=NW_022611034.1,length=42017>
##contig=<ID=NW_022611035.1,length=140651>
##contig=<ID=NW_022611036.1,length=125530>
##contig=<ID=NC_015139.1,length=153482>
##source=HaplotypeCaller
#CHROM POS ID REF ALT QUAL FILTER INFO FORMAT N-Bulk
NC_024795.2 1 . A <NON_REF> . . END=1030 GT:DP:GQ:MIN_DP:PL 0/0:0:0:0:0,0,0
NC_024795.2 1031 . C <NON_REF> . . END=1032 GT:DP:GQ:MIN_DP:PL 0/0:1:3:1:0,3,24
NC_024795.2 1033 . C <NON_REF> . . END=1033 GT:DP:GQ:MIN_DP:PL 0/0:1:0:1:0,0,0
NC_024795.2 1034 . T <NON_REF> . . END=1074 GT:DP:GQ:MIN_DP:PL 0/0:1:3:1:0,3,24
NC_024795.2 1075 . C <NON_REF> . . END=1161 GT:DP:GQ:MIN_DP:PL 0/0:0:0:0:0,0,0
NC_024795.2 1162 . A <NON_REF> . . END=1162 GT:DP:GQ:MIN_DP:PL 0/0:1:3:1:0,3,27
NC_024795.2 1163 . C <NON_REF> . . END=1164 GT:DP:GQ:MIN_DP:PL 0/0:1:0:1:0,0,0
NC_024795.2 1165 . C <NON_REF> . . END=1165 GT:DP:GQ:MIN_DP:PL 0/0:2:6:2:0,6,51
NC_024795.2 1166 . T <NON_REF> . . END=1168 GT:DP:GQ:MIN_DP:PL 0/0:2:0:2:0,0,0
NC_024795.2 1169 . T <NON_REF> . . END=1169 GT:DP:GQ:MIN_DP:PL 0/0:2:6:2:0,6,50
NC_024795.2 1170 . A <NON_REF> . . END=1170 GT:DP:GQ:MIN_DP:PL 0/0:3:9:3:0,9,69
NC_024795.2 1171 . T <NON_REF> . . END=1171 GT:DP:GQ:MIN_DP:PL 0/0:4:0:4:0,0,16
NC_024795.2 1172 . T <NON_REF> . . END=1191 GT:DP:GQ:MIN_DP:PL 0/0:4:12:4:0,12,82
NC_024795.2 1192 . A <NON_REF> . . END=1195 GT:DP:GQ:MIN_DP:PL 0/0:5:15:5:0,15,115
NC_024795.2 1196 . A <NON_REF> . . END=1196 GT:DP:GQ:MIN_DP:PL 0/0:4:12:4:0,12,101
NC_024795.2 1197 . C <NON_REF> . . END=1199 GT:DP:GQ:MIN_DP:PL 0/0:5:15:5:0,15,123
NC_024795.2 1200 . T <NON_REF> . . END=1200 GT:DP:GQ:MIN_DP:PL 0/0:4:12:4:0,12,105
NC_024795.2 1201 . T <NON_REF> . . END=1212 GT:DP:GQ:MIN_DP:PL 0/0:5:15:5:0,15,111
NC_024795.2 1213 . T <NON_REF> . . END=1213 GT:DP:GQ:MIN_DP:PL 0/0:7:5:7:0,5,132
NC_024795.2 1214 . A <NON_REF> . . END=1216 GT:DP:GQ:MIN_DP:PL 0/0:7:15:7:0,15,225
NC_024795.2 1217 . C <NON_REF> . . END=1217 GT:DP:GQ:MIN_DP:PL 0/0:7:2:7:0,2,130
NC_024795.2 1218 . C <NON_REF> . . END=1225 GT:DP:GQ:MIN_DP:PL 0/0:7:12:6:0,12,180
NC_024795.2 1226 . A <NON_REF> . . END=1226 GT:DP:GQ:MIN_DP:PL 0/0:7:0:7:0,0,125
NC_024795.2 1227 . A <NON_REF> . . END=1230 GT:DP:GQ:MIN_DP:PL 0/0:7:15:7:0,15,225
NC_024795.2 1231 . C <NON_REF> . . END=1232 GT:DP:GQ:MIN_DP:PL 0/0:9:21:9:0,21,315
NC_024795.2 1233 . C <NON_REF> . . END=1233 GT:DP:GQ:MIN_DP:PL 0/0:9:3:9:0,3,168
NC_024795.2 1234 . T <NON_REF> . . END=1235 GT:DP:GQ:MIN_DP:PL 0/0:9:0:9:0,0,160
NC_024795.2 1236 . T <NON_REF> . . END=1236 GT:DP:GQ:MIN_DP:PL 0/0:10:6:10:0,6,209
```



File Format--hapmap

用tab键作为分隔符，基因型为实际的碱基，而不是代码。

rs	allele	chrom	pos	strand	assembly	center	protLSID	assayLSID	panel	QCcode	ssc.s1	ssc.s10	ssc.s100
1_250858	C/T	1	250858	NA	NA	NA	NA	NA	CC	CC	CC	CC	CC
1_254337	G/C	1	254337	NA	NA	NA	NA	NA	GG	GG	GG	GG	GG
1_263294	C/T	1	263294	NA	NA	NA	NA	NA	CC	CC	CC	CC	CC
1_263668	C/T	1	263668	NA	NA	NA	NA	NA	CC	CC	CT	CC	CC
1_270055	T/G	1	270055	NA	NA	NA	NA	NA	TT	TT	TT	TT	NN
1_546234	C/T	1	546234	NA	NA	NA	NA	NA	CC	CC	CC	CT	CC
1_565543	G/A	1	565543	NA	NA	NA	NA	NA	GG	GG	GG	GG	NN
1_647688	G/A	1	647688	NA	NA	NA	NA	NA	GA	GG	GG	GG	NN
1_816358	C/T	1	816358	NA	NA	NA	NA	NA	TT	CT	TT	TT	NN
1_863281	G/C	1	863281	NA	NA	NA	NA	NA	GG	GG	GG	GC	NN
1_1076045	G/A	1	1076045	NA	NA	NA	NA	NA	NN	GG	GG	GG	NN
1_1076693	T/C	1	1076693	NA	NA	NA	NA	NA	TT	TT	TT	TT	TT
1_1084677	T/C	1	1084677	NA	NA	NA	NA	NA	TT	TT	TT	TT	TT
1_1090204	A/T	1	1090204	NA	NA	NA	NA	NA	AA	AA	AA	AA	NN
1_1090219	T/A	1	1090219	NA	NA	NA	NA	NA	TT	TT	TT	TT	NN
1_1172064	A/G	1	1172064	NA	NA	NA	NA	NA	AA	AA	AA	AA	AA
1_1356479	G/A	1	1356479	NA	NA	NA	NA	NA	GG	GG	GG	GG	NN
1_1623948	G/A	1	1623948	NA	NA	NA	NA	NA	GG	GG	GG	GA	NN
1_1624127	G/A	1	1624127	NA	NA	NA	NA	NA	GG	GG	GG	GG	GG
1_1630020	T/G	1	1630020	NA	NA	NA	NA	NA	TT	TT	TT	TG	NN
1_1630252	G/A	1	1630252	NA	NA	NA	NA	NA	GG	GG	AA	GA	GG
1_1645291	G/C	1	1645291	NA	NA	NA	NA	NA	GG	GG	GG	GG	NN
1_1844541	T/A	1	1844541	NA	NA	NA	NA	NA	TT	TT	TT	TT	NN
1_1851178	C/T	1	1851178	NA	NA	NA	NA	NA	CC	CC	CT	CT	NN
1_1853179	G/A	1	1853179	NA	NA	NA	NA	NA	GG	GG	GG	GG	GG
1_1859186	C/A	1	1859186	NA	NA	NA	NA	NA	CC	CC	CA	CA	CC
...	...	...	...	...	...	...	...	...	...	...	...	...	...



MD5 Code

- Windows: 直接百度搜索“MD5计算”，随便找一个小工具，即可计算。

(<http://www.metools.info/other/o21.html>)



The screenshot shows a web-based MD5 calculator interface. At the top, there are four tabs: "MD5加密", "文件MD5计算", "SHA1加密", and "文件SHA1/SHA256校验". The "文件MD5计算" tab is selected. Below the tabs, there is a "选择文件" button, a text input field containing "C:\fakepath\2020-09-08.rep", and a green "MD5计算" button. Below this, the "文件MD5值:" label is followed by a text input field containing the MD5 hash "8626236ab7aaa6085c9825d41b63aa79", which is highlighted with a red rectangular box. At the bottom, there is a section titled "执行过程:" containing a list of execution details: "计算耗时: 84ms", "计算成功, MD5值: 8626236ab7aaa6085c9825d41b63aa79", "加载数据: 第2部分, 总2部分", "加载数据: 第1部分, 总2部分", and "开始计算, 文件名: (2020-09-08.report)".

- linux: `md5sum filename`

务必正确记录每个文件的32位MD5值！

账号注册/登录

◆ Login or Register links:

<https://bigd.big.ac.cn/sso/login?service=https://bigd.big.ac.cn/gvm//userlogin>

The screenshot shows the CNGB-NGDC Genome Variation Map (GVM) homepage. The header includes the CNGB-NGDC logo and navigation links: Databases, Tools, Standards, Publications, and About. The main banner features the text "Genome Variation Map" and a background image of a DNA double helix. Below the banner is a navigation bar with links: Home, Browse, Search, Genome Browser, Submit, Downloads, Statistics, Standards, FAQ, Login (highlighted with a red box), and Register. The main content area includes a description of the GVM as a public data repository of genome variations. It features a search bar with a dropdown menu set to "GVM" and a search button. Below the search bar, there is a section for "Species" with a link to "All species with genomic variants" and a search bar for "find a species". To the right of the species section is a "Featured genomes" section with images of various organisms. On the right side of the page, there is a "Data Submission" section with a link to "Deposit data into GVM" and a "Problems or Questions?" section with contact information: Email: gvm@big.ac.cn and QQ group: 468638108.

NGDC Central Authentication Service

The screenshot shows the NGDC Central Authentication Service login page. The header includes the NGDC logo and the text "Central Authentication Service". The main heading is "Enter your Username and Password". Below this are input fields for "Email" and "Password". The "Password" field has a "Forgot password?" link. Below the password field is a "Check code" field with a CAPTCHA image showing the code "8c7d". There is a checkbox labeled "Keep me signed in". At the bottom are three buttons: "Login", "Reset", and "Register".

New Submission—Submitter Information

GVM / Data Submission / My Submission

GVM is a repository specialized for archiving genome variation data (VCF, GVCF or HapMap format). Creating a new GVM submission, three steps are needed:

1. Create a BioProject (an overall description of a single research initiative) in [BioProject submission portal](#).
2. Create a BioSample (description of the biological source material) in [BioSample submission portal](#).
3. Create a GVM Submission by clicking 'New Submission' below.

Any question, problems or suggestions, please feel free to contact us

Email: gvm@big.ac.cn

QQ group: 468638108

[GVM Submission Wizard Help](#)

[New Submission](#)



GVM / Data Submission / New Submission

Step 1
Submitter

Step 2
General information

Step 3
Meta-data Information

Step 4
Overview

Submitter

• First name

li

Middle name

• Last name

cp

• Email address

licp@big.ac.cn

Telephone number

• Laboratory/Centre

NGDC

• Company/Institute/University

Beijing Institute of Genomics, Chinese Academy of Sciences

[Save and next](#)



New Submission—New BioProject

A project must be resisted first in BioProject (<https://bigd.big.ac.cn/bioproject/>), create it if not.

GVM / Data Submission / New Submission

Step 1
Submitter

Step 2
General information

Step 3
Meta-data Information

Step 4
Overview

BioProject

+ BioProject accession

OR Go to create [BioProject](#)

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

Project title

Abstract information

If you have already created BioProject for the project successfully, please skip this step (Page 13-17) and jump to page **18** straightly.



New Submission—New BioProject

BIG Sub / BioProject / New BioProject

01
Submitter

02
General

03
Project Type

04
Publication

05
Overview & Submit

Submitter Information

* First name

Middle name

* Last name

* Email

Email (secondary)

* Organization

Organization website

* Department

Phone

Fax

* Street

* City

State/Province

* Postal code

* Country/Region

Save and forward



New Submission—New BioProject

[Home](#) [Documentation](#) Welcome, li

[BIG Sub](#) / [BioProject](#) / [New BioProject](#)

01
Submitter

02
General

03
Project Type

04
Publication

05
Overview & Submit

General information

Release Date

☒ After the approval is passed, release immediately following curation (recommended)
☐ Release on specified date

Umbrella project

--

Project

* Project title

GVM Submission

* Relevance

Agricultural

* Description

GVM data submission test.

When all required items have been filled in correctly, click “save and forward” at the bottom of the page.



New Submission—New BioProject

BIG Sub / BioProject / New BioProject

1

01
Submitter

02
General

03
Project Type

04
Publication

05
Overview & Submit

Project type

* Project data type

- ☒ Whole genome sequencing
- ☐ Clone ends
- ☐ Epigenomics
- ☐ Exome
- ☐ Map
- ☐ Metagenome
- ☐ Phenotype or Genotype
- ☐ Random survey
- ☐ Targeted Locus (Loci)
- ☐ Transcriptome or Gene expression
- ☐ Variation
- ☐ Genome sequencing and assembly
- ☐ Raw sequence reads
- ☐ Genome sequencing
- ☐ Assembly
- ☐ Metagenomic assembly
- ☐ Targeted loci cultured
- ☐ Targeted loci environmental
- ☐ Other

* Sample scope

Multiisolate

2

Sample scope choices ?

Monoisolate: a single animal, cultured cell-line, inbred population (or possibly a heterogeneous population when a single genome assembly is generated from the pooled sample; not preferred).

Multiisolate: multiple individuals, a population (representative of a species).

Multi-species: sample represents multiple species.

Environment: the species content of the sample is not known.

Synthetic: the sample is synthetically created by a machine.

Single cell: the sample is single cell sequencing.

Other: specify the sample scope that was used.

Save and forward



New Submission—New BioProject

BIG Sub / BioProject / New BioProject

01
Submitter

02
General

03
Project Type

04
Publication

05
Overview & Submit

Publication

PubMed ID

OR

DOI

29069473



+ Add another publication

Save and forward



New Submission—New BioProject

BIG Sub / BioProject / New BioProject

1

01
Submitter

02
General

03
Project Type

04
Publication

05
Overview & Submit

Overview

Submitter Information

Submitter	li cp licp@big.ac.cn
Organization	Beijing Institute of Genomics, Chinese Academy of Sciences
Department	BIGD
Country/Region	China
Address	NO.1 Beichen West Road, Chaoyang District Beijing , China
Postal code	100101

Project type

2

Project data type	Whole genome sequencing
Sample scope	Multiisolate

Publication

PubMed ID	29069473
DOI	
Journal name	
Article title	
Year	

General information

Project title	GVM Submission test
Relevance	Agricultural
Description	GVM data submission test.
Release Date	2020-08-25
Grants	

Agency

Program

Grant ID

Grant Title

No funding support

Submit



New Submission—BioProject

GVM / Data Submission / New Submission

Step 1
Submitter

Step 2
General information

Step 3
Meta-data Information

Step 4
Overview

BioProject

• BioProject accession

PRJCA003320



OR Go to create [BioProject](#)

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

Project title

GVM Submission test

Abstract information

GVM data submission test.

Select the right BioProject
accession for the project.



New Submission—New BioSample

BioSample

- ☒ **No GVM related BioSample information was created**
☐ **GVM related BioSample information has been created**

If you have not created GVM related Biological Sample(s), please select the 'No GVM related BioSample information was created' and create BioSample(s) in this wizard.
If you have already created GVM related Biological Sample(s), please select the 'GVM related BioSample information has been created'.

GVM

Release Date for GVM

- ☒ Release immediately following curation (recommended)
☐ Release on specified date

Abstract for the variation data

If you have already created BioSample for each sample, please select the “GVM related BioSample information has been created”, then click “Save and next” and jump to page **26** straightly.

Save and next



New Submission—New BioSample

BIG Sub / BioSample / New BioSample

01
Submitter

02
General

03
Sample Type

04
Attributes

05
Overview & Submit

Submitter Information

* First name

li

Middle name

middle name

* Last name

cp

* Email

licp@big.ac.cn

Email (secondary)

secondary email

* Organization

Beijing Institute of Genomics, Ch

Organization website

http://www.big.ac.cn

* Department

BIGD

Phone

Fax

* Street

NO.1 Beichen West Road, Chao

* City

Beijing , China

State/Province

* Postal code

100101

* Country/Region

China



Save and forward



New Submission—New BioSample

BIG Sub / BioSample / Submission: subSAM092255

01
Submitter

02
General

03
Sample Type

04
Attributes

05
Overview & Submit

General information

Release date

- ☒ After the approval is passed, release immediately following curation (recommended)
- ☐ Release on specified date

Submission type

- ☒ Batch BioSamples (recommended)
 - ☐ Single BioSample
-  If the submission contains multiple samples, please select the "Batch BioSample".

Select the right BioProject accession for the project.

General information

* Project accession

PRJCA003320

OR Go to create [BioProject](#)

Save and forward



New Submission—New BioSample

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



New Submission—New BioSample

Attributes of human

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

Please select file

Upload

1

Download BioSample batch submission template **Human.us.xlsx**, then fill in and double-check it before uploading.

For column explanations and examples, please see the [e.g.Human.us.xlsx](#).

For more information, please see the [Help](#).

Save and forward

A	B	C	D	E	F	G	H	I	J	K
This is a submission template for batch deposit of "Human sample" to the BioSample database.										
GREEN fields are mandatory. Your submission will fail if any mandatory fields are not completed.										
GREY fields are optional. Leave optional fields empty if no information is available.										
YELLOW columns have drop-down menus that allow you to select from a controlled vocabulary. Once specified for one row, these values can be copied-and-pasted down.										
CAUTION: DO NOT delete or insert column!										
Be aware that Excel may automatically apply formatting to your data. In particular, take care with dates, incrementing autofills and special characters like / or -. Doublecheck that your text file is acc										
1	2	3	4	5	6	7	8	9	10	11
*sample_name	*public_description	*project_accession	*sample_title	*isolate	age	age_unit	*biomaterial_provider	*sex	*tissue	disease



New Submission—New BioSample

Attributes of plant

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

Plant.us.xlsx

Upload

2

1 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,

Attributes of plant

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

[Plant.us.xlsx](#)

24KB

Delete

Check

3

1 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.

Attributes of plant

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

[Plant.us.xlsx](#)

24KB

Delete

Check

✓ Checked OK.

1 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



New Submission—New BioSample

BIG Sub / BioSample / Submission: subSAM092255

01 Submitter 02 General 03 Sample Type 04 Attributes 05 Overview & Submit

Overview

Submitter Information

Submitter	li cp licp@big.ac.cn
Organization	Beijing Institute of Genomics, Chinese Academy of Sciences
Department	BIGD
Country/Region	China
Address	NO.1 Beichen West Road, Chaoyang District Beijing , China
Postal code	100101

General information

Project accession	PRJCA003320
Sample type	Plant sample
Sample count	5
Create time	2020-08-25
Release date	2020-08-25
Modify time	2020-08-25

Attributes

Metadata File: [Plant.us.xlsx](#)

Sample Name	Organism	Title	Description
SSS1	Maize	SSS1	GVM submission test
SSS2	Maize	SSS2	GVM submission test
SSS3	Maize	SSS3	GVM submission test
SSS4	Maize	SSS4	GVM submission test
SSS5	Maize	SSS5	GVM submission test

Total Items: 5 | Items of 1 - 5 | Page size: 20 | Page 1/1 | << First | Last >> | Jump To: 1 | GO

Submit



New Submission—Meta Information

GVM / Data Submission / New Submission

Step 1
Submitter

Step 2
General information

Step 3
Meta-data Information

Step 4
Overview

GVM Meta-data Information

BioProject: PRJCA003421

★ Please upload GVM template file

Select file...

Browse ...

Download the GVM Submission [Template](#)

This is a submission template for batch deposit of samples to the GVM database.

GREEN fields are mandatory. Your submission will fail if any mandatory fields are not completed.

GREY fields are optional. Leave optional fields empty if no information is available.

YELLOW columns have drop-down menus that allow you to select from a controlled vocabulary. Once specified for one row, these values can be copied-and-pasted down.

CAUTION: Be aware that Excel may automatically apply formatting to your data. In particular, take care with dates, incrementing autofills and special characters like / or -. Doublecheck that your text file is accurate before send

1	2	3	4	5	6	7	8	9	10
ID	*sample_name	project_accession	geographic_location	latitude_longitude	collector	collection_date	trait	phenotype	...
S1	zhang1	PRJCA003421	China, Beijing						
S2	li2	PRJCA003421	China, Beijing						
S3	wang3	PRJCA003421	China, Beijing						

This is a submission template for batch deposit of variants' files to GVM database.
GREEN fields are mandatory. Your submission will fail if any mandatory fields are not completed.
GREY fields are optional. Leave optional fields empty if no information is available.
YELLOW columns have drop-down menus that allow you to select from a controlled vocabulary. Once specified for one row, these values can be copied-and-pasted down.

column 5 (md5): The MD5 code can be calculated by md5sum in linux or same other online tools in web.

CAUTION: Be aware that Excel may automatically apply formatting to your data. In particular, take care with dates, incrementing autofills and special characters like / or -. Doublecheck that your text file is accurate before send

1	2	3	4	5
analysis_alias	file_name	file_description	file_type	md5
1	Human.vcf.gz	Variants file of 1000 Chinese in vcf format.	vcf	a11111111111111111111111111111111
2	S1.g.vcf.gz	Variants file of sample S1.	gvcf	a33333333333333333333333333333333
3	Human.variants.hapmap	Variants file of 1000 Chinese in hapmap format.	hapmap	a44444444444444444444444444444444

column 2, 3 and 4: If the sample is genotyped by array, you should select "Genotyping by array" from the drop sequencing strategy from the drop-down list in column 2 and select the sequencing platform from the drop-down

CAUTION: Be aware that Excel may automatically apply formatting to your data. In particular, take care with data

1	2	3	4	5	6	7	8	9	10	11
ID	strategy	platform	array_name	coverage	reference	reference_version	reference_link	software	parameters	...
S1	WGS	Illumina HiSeq 2000			Homo sapiens	GRCh37	http://grch37.ensem	BWA; GATK	mem; HC	
S2	WGS	Illumina HiSeq 2000			Homo sapiens	GRCh37	http://grch37.ensem	BWA; GATK	mem; HC	
S3	WGS	Illumina HiSeq 2000			Homo sapiens	GRCh37	http://grch37.ensem	BWA; GATK	mem; HC	

New Submission—Meta Information

• Please upload GVM template file

Select file...

Browse ...

2

Download the GVM Submission [Template](#)

• Please upload GVM template file

GVM Submission Template V3.0.xlsx

Remove

Upload

Browse ...

3

Download the GVM Submission [Template](#)

• Please upload GVM template file

[GVM_Submission_Template_V3.0.xlsx](#)

Check

Remove

4

• Please upload GVM template file

[GVM_Submission_Template_V3.0.xlsx](#)

Check

Remove

Validate Success !

Save and next

5



New Submission—Submit

You now at: [Submit](#) > [Submission](#)

Step 1
Submitter

Step 2
General information

Step 3
Meta-data Information

Step 4
Overview

Overview

Please transmitting your data files to the GVM FTP site ([download FileZilla client](#))

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

User: Same as you login the GVM

Password: Same as you login the GVM

Finish

点击链接，下载文件上传工具 FileZilla。

<https://filezilla-project.org/download.php?type=client>

New Submission

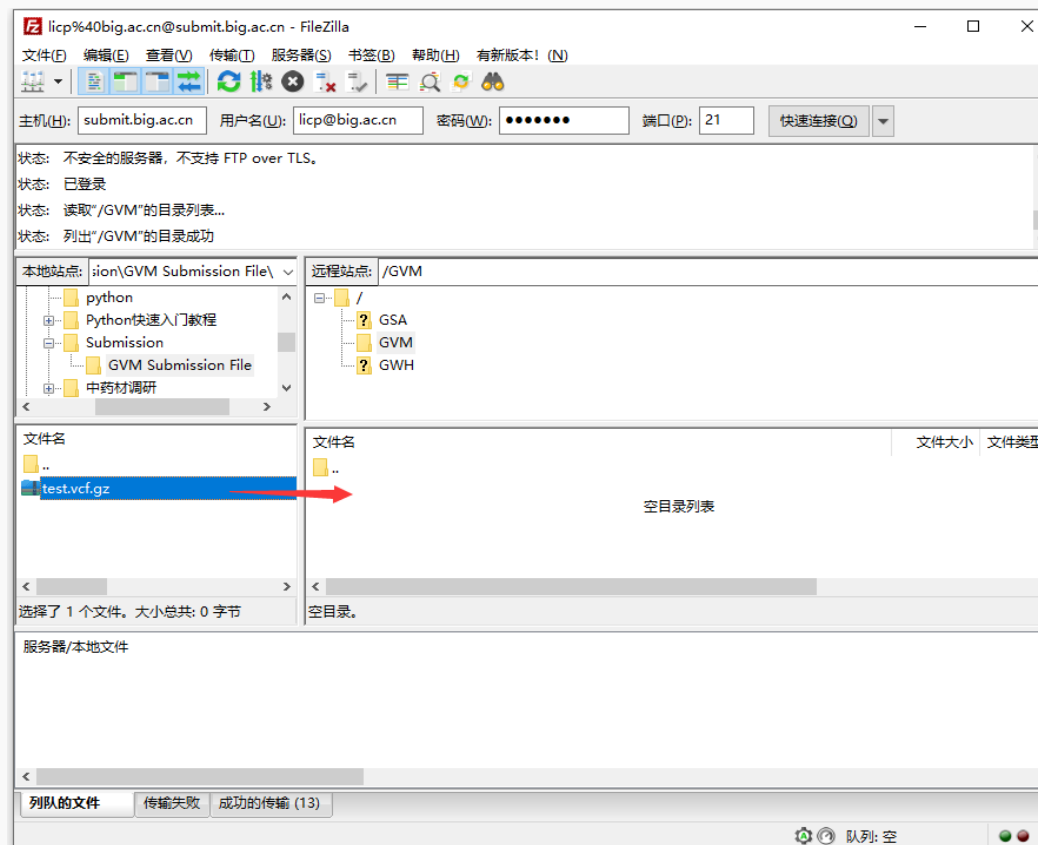
Items 1 - 10 of 177 Items per page [First](#) [Prev](#) of 18 [Next](#) [Last](#) [GOTO](#)

Accession	Submission ID	Project	Sample	Release Date	Create Time	Create By	Update Time	Status		User Operation	Admin Operation
Unassigned	sub000254	PRJCA003421	SAMC237135, SAMC237136 more	2020-09-30	2020-09-09 20:06:24	licp@big.ac.cn	2020-09-09 20:47:09	Meta-data Finished	Confidential		audit



New Submission—Data File Submit

- Please transmitting your data files to the GVM FTP site ([download FileZilla client](#))
- Address: **ftp://submit.big.ac.cn**
- User: Same as you login the GVM
- Password: Same as you login the GVM
- Port: **21**



Contact us

If you have any questions, problems and suggestions, please contact us.

- E-Mail: gvm@big.ac.cn
- QQ group: 468638108



欢迎大家递交数据至GVM

谢谢！

GVM Group
2021/05/12