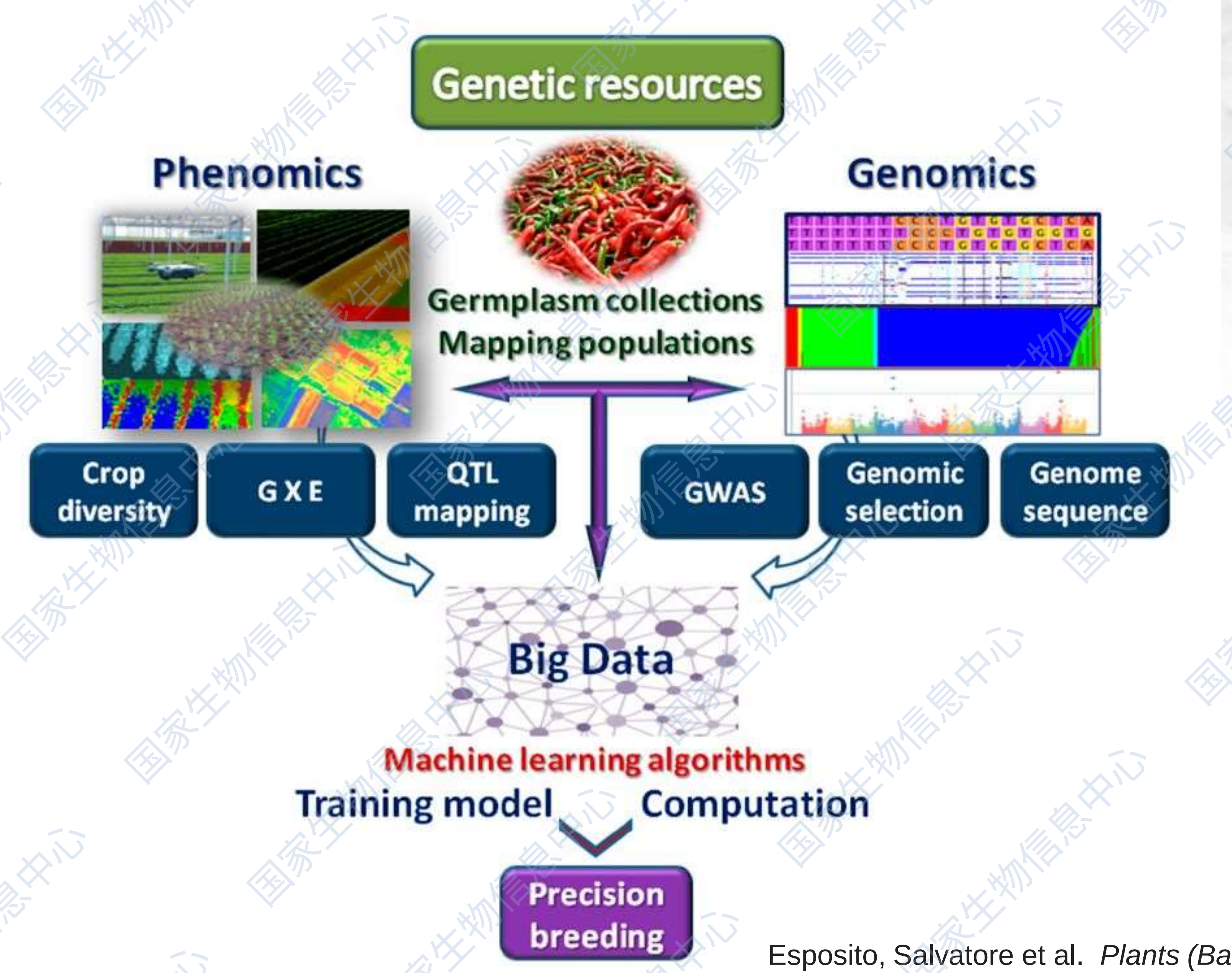
基因组变异与表型相关数据汇交共享

田东梅
CNCN-NGDC
2024年10月24日

数据驱动的生命科学



Esposito, Salvatore et al. *Plants (Basel, Switzerland)*, 2019

遗传变异与表型数据知识体系

建立多物种从基因型到表型再到关联知识的数据知识体系，为科研发现与智慧育种提供支撑

- □ 建立了变异组学数据分析、知识审编整合的标准规范
- □ 形成了完善的变异组学数据汇交、管理与共享服务系统
- □ 实现了遗传变异-表型-知识的关联互通，多维信息可视化应用

基因组变异库GVM

Genome Variation Map has been recommended by Cell Press to provide genetic variation repositories.

The Genome Variation Map (GVM) is a public data repository of genome variations, including single nucleotide polymorphisms and small insertions and deletions, with particular focuses on human as well as cultivated plants and domesticated animals.

https://ngdc.cncb.ac.cn/gvm/

图像表型库OPIA

Open Plant Image Archive (OPIA) is an archive database of plant images and related phenotypic data from high-throughput phenotyping platforms for a diversity of species. View more...

https://ngdc.cncb.ac.cn/opia

知识库GWAS Atlas

https://ngdc.cncb.ac.cn/gwas/

- □ 57个物种，391个研究项目
- □ 83,366个生物学样本
- □ 1.6亿个序列变异位点

- □ 11个物种，56个数据集
- □ 566,225张作物图像
- □ 56种作物表型性状

- □ 31个物种 163,979个变异
- □ 1,724个性状 302,295条G2P
- □ 486条因果变异知识

Nucleic Acids Research (2018, 2020, 2021, 2023, 2024)

提纲

01 | 基因组序列变异数据库GVM

02 | 植物图像及其表型归档库OPIA

03 | G2P关联知识库GWAS Atlas

04 | 用户变异与表型数据汇交与共享

提纲

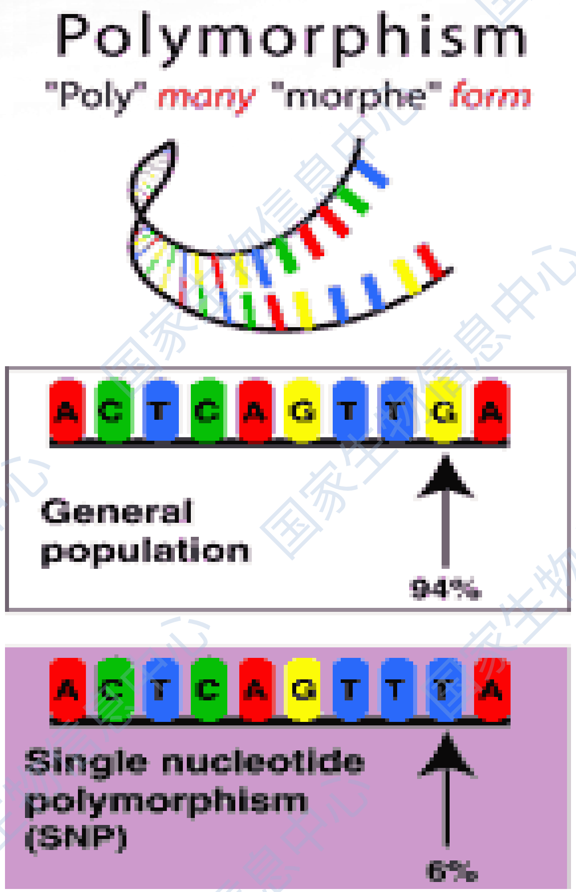
01 | 基因组序列变异数据库GVM

02 | 植物图像及其表型归档库OPIA

03 | G2P关联知识库GWAS Atlas

04 | 用户变异与表型数据汇交与共享

基因组变异



图片来源：genome news network

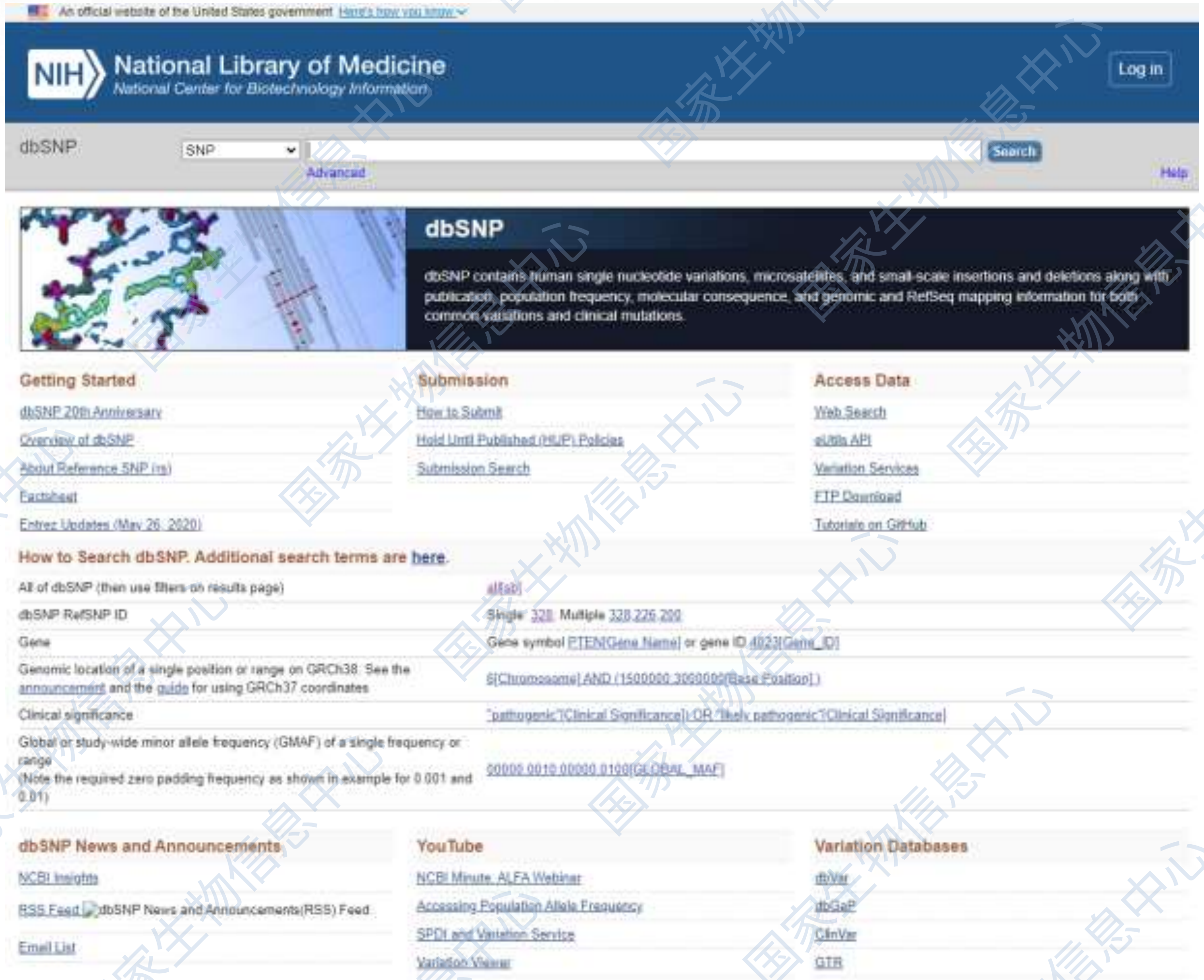
单核苷酸多态性，英文single nucleotide polymorphism，缩写为SNP。SNP主要是指在基因组水平上引起的单个碱基的变异，其在群体中的发生频率不小于1%，包括单碱基的转换、颠换以及单碱基的插入和缺失等。

基因组结构性变异，英文Structure Variations，简称SVs，通常指基因组上大长度的序列变化和位置关系变化。

□ 物种进化研究
□ 表型差异研究
□ 人类疾病研究
□ 药物研发及精准医疗
□ 农业遗传改良育种
□ 生物多样性保护

国际变异资源库

NCBI-dbSNP/dbVar



An official website of the United States government Here's how you know

NIH National Library of Medicine
National Center for Biotechnology Information

Log in

dbSNP SNP Search Advanced Help

dbSNP

dbSNP contains human single nucleotide variations, microsatellites, and small-scale insertions and deletions along with publication, population frequency, molecular consequence, and genomic and RefSeq mapping information for both common variations and clinical mutations.

Getting Started
- dbSNP 20th Anniversary
- Overview of dbSNP
- About Reference SNP (rs)
- Factsheet
- Entrez Updates (May 26, 2020)

Submission
- How to Submit
- Hold Until Published (HUP) Policies
- Submission Search

Access Data
- Web Search
- eUtils API
- Variation Services
- FTP Download
- Tutorials on GitHub

How to Search dbSNP. Additional search terms are here.

All of dbSNP (then use filters on results page)	all[sb]
dbSNP RefSNP ID	Single: 328, Multiple 328,226,200
Gene	Gene symbol PTEN[Gene Name] or gene ID 4023[Gene_ID]
Genomic location of a single position or range on GRCh38. See the announcement and the guide for using GRCh37 coordinates	6[Chromosome] AND (1500000:3000000[Base Position])
Clinical significance	"pathogenic"[Clinical Significance] OR "likely pathogenic"[Clinical Significance]
Global or study-wide minor allele frequency (GMAF) of a single frequency or range (Note the required zero padding frequency as shown in example for 0.001 and 0.01)	00000.0010:00000.0100[GLOBAL_MAF]

dbSNP News and Announcements
- NCBI Insights
- RSS Feed dbSNP News and Announcements(RSS) Feed
- Email List

YouTube
- NCBI Minute: ALFA Webinar
- Accessing Population Allele Frequency
- SPDI and Variation Service
- Variation Viewer

Variation Databases
- dbVar
- dbGaP
- ClinVar
- GTR

EBI-EVA

EMBL-EBI Services Research Training About us EMBL-EBI Hinxton

European Variation Archive

Home Submit Data Study Browser Variant Browser GA4GH API RS Release Help Feedback

EVA > HOME

Overview

The European Variation Archive is an open-access database of all types of genetic variation data from all species.

All users can download data from any study, or submit their own data to the archive. You can also query all variants in the EVA by study, gene, chromosomal location or dbSNP identifier using our Variant Browser.

We will be adding new features to the EVA on a regular basis, and welcome your comments and feedback.

News

Tweets by @evarchive

Search for SNPs

ex: rs123 or ss567 or comma/space separated RS or SS IDs Search

The RS ID release 5 is available in our FTP or through our API . See release page for details.

Statistics

Short genetic variants studies (<50bp)

Structural variants studies (>50bp)

基因组变异数据库GVM

Genome Variation Map has been recommended by Cell Press to provide genetic variation repositories.

The Genome Variation Map (GVM) is a public data repository of genome variations, including single nucleotide polymorphisms and small insertions and deletions, with particular focuses on human as well as cultivated plants and domesticated animals.

If you have any question or would like to give us any suggestion/comment or report a bug, please feel free to contact us.

Email: gvm@big.ac.cn

QQ group: 468638108

https://ngdc.cncb.ac.cn/gvm/

- 我国最大变异资源库，采用标准分析方法鉴定的物种参考变异数据集及功能注释信息
- 我国首个基因组变异数据汇交、管理与共享平台
- 2017年上线，访客数2.7万，累积下载3500+万次

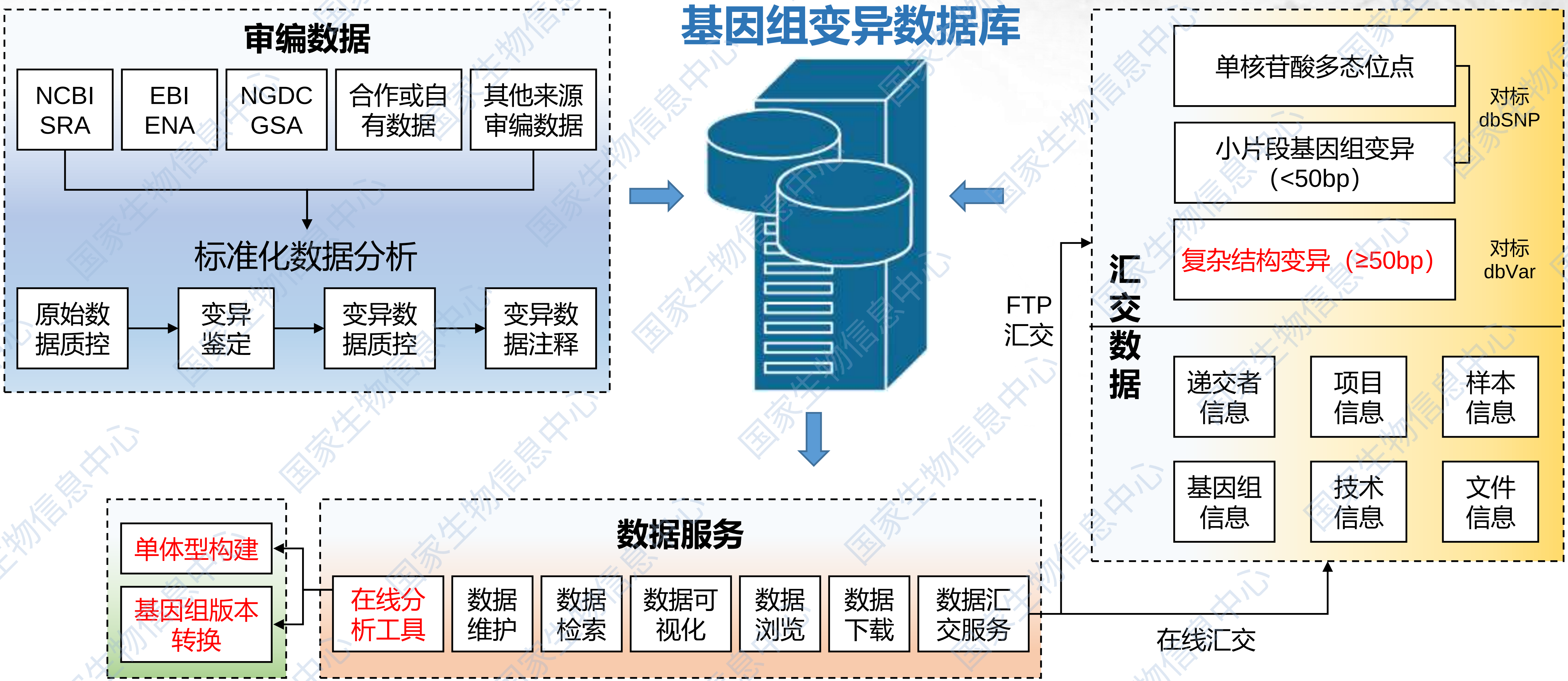
Nucleic Acids Research (2018，2021)

同行评价：

Daniel J. Rigden, *Editor in NAR*: *“A **major new resource** is the Genome Variation Map (58) from the BIG Data Center covering 19 species. **Its arrival is particularly timely** with the announcement that comparable NCBI resources dbSNP and dbVar are to stop accepting non-human.”*

Thomas M. Keane, Team leader of EVA: “The EVA, the NCBI-based dbSNP and dbVar databases, and the Chinese Genome Variation Map resource form **a worldwide network for** brokering submissions, assigning permanent study and locus identifiers and exchanging these identifiers to ensure that consistent data is available at all sites.”

基因组变异数据库组织框架



审编数据

NCBI SRA

EBI ENA

NGDC GSA

合作或自有数据

其他来源审编数据

标准化数据分析

原始数据质控 → 变异鉴定 → 变异数据质控 → 变异数据注释

基因组变异数据库

汇交数据

单核苷酸多态位点

小片段基因组变异（<50bp）

对标 dbSNP

复杂结构变异（≥50bp）

对标 dbVar

递交者信息

项目信息

样本信息

基因组信息

技术信息

文件信息

FTP 汇交

在线汇交

数据服务

在线分析工具

数据维护

数据检索

数据可视化

数据浏览

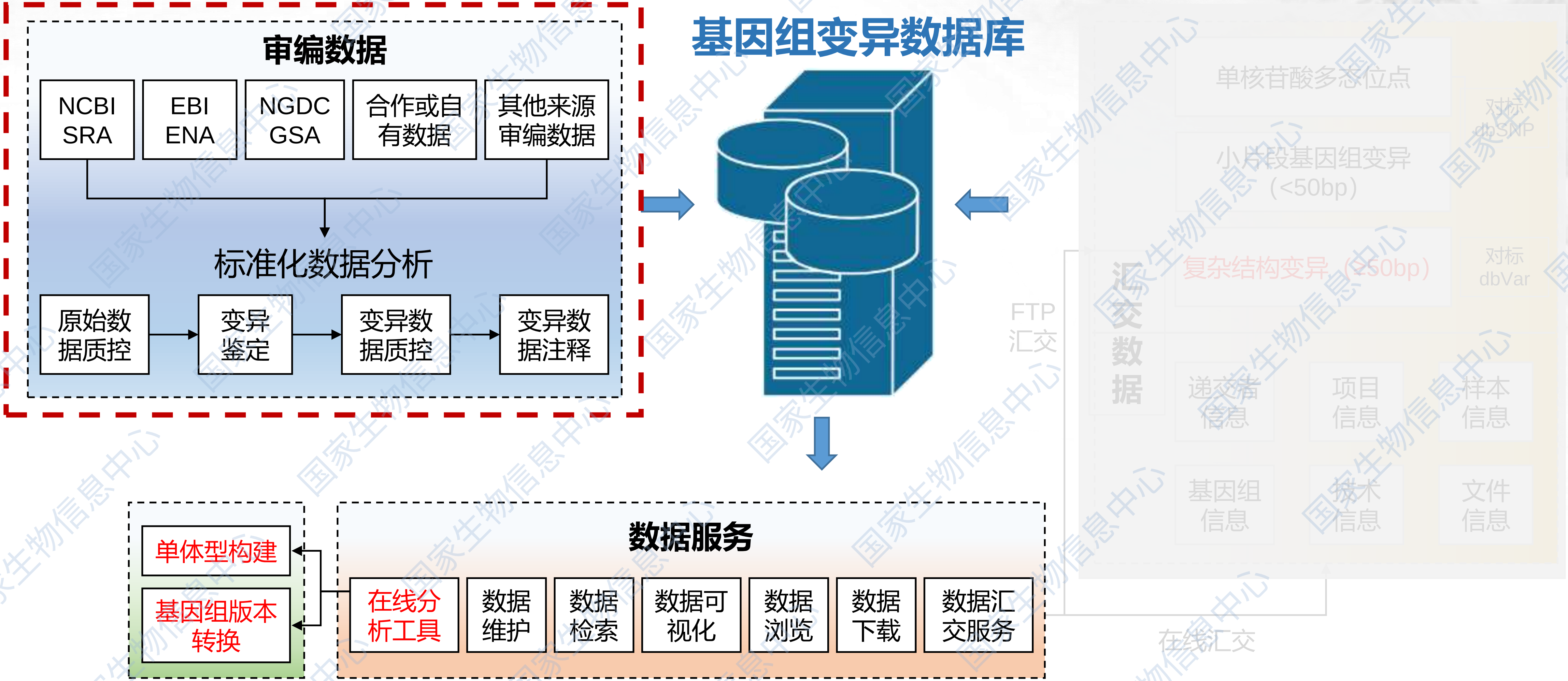
数据下载

数据汇交服务

单体型构建

基因组版本转换

基因组变异数据库组织框架



审编数据

NCBI SRA

EBI ENA

NGDC GSA

合作或自有数据

其他来源审编数据

标准化数据分析

原始数据质控 → 变异鉴定 → 变异数据质控 → 变异数据注释

基因组变异数据库

单核苷酸多态位点

小片段基因组变异（<50bp）

复杂结构变异（≥50bp）

对标 dbSNP

对标 dbVar

汇交数据

递交者信息

项目信息

样本信息

基因组信息

技术信息

文件信息

FTP 汇交

在线汇交

数据服务

在线分析工具

数据维护

数据检索

数据可视化

数据浏览

数据下载

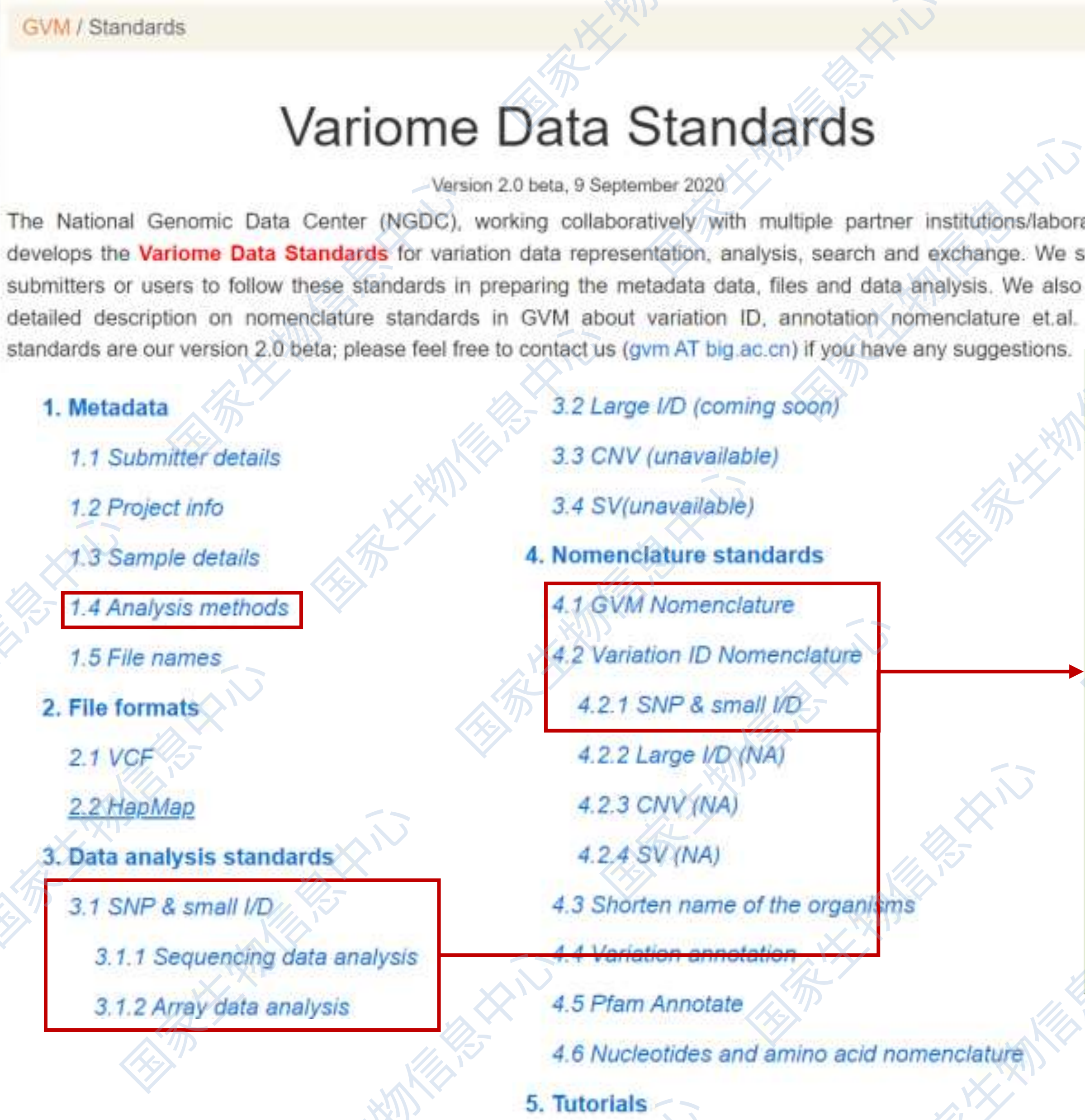
数据汇交服务

单体型构建

基因组版本转换

审编数据分析标准

国家基因组科学数据中心
National Genomics Data Center



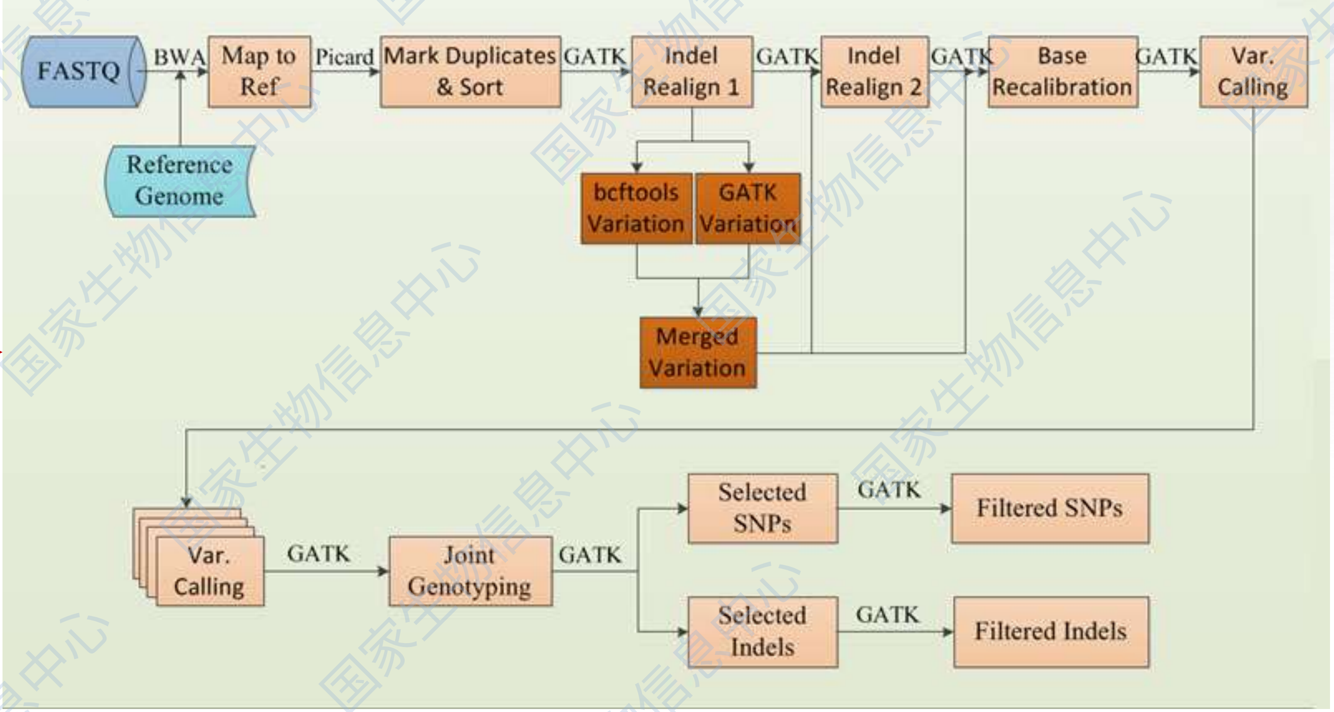
GVM / Standards

Variome Data Standards

Version 2.0 beta, 9 September 2020

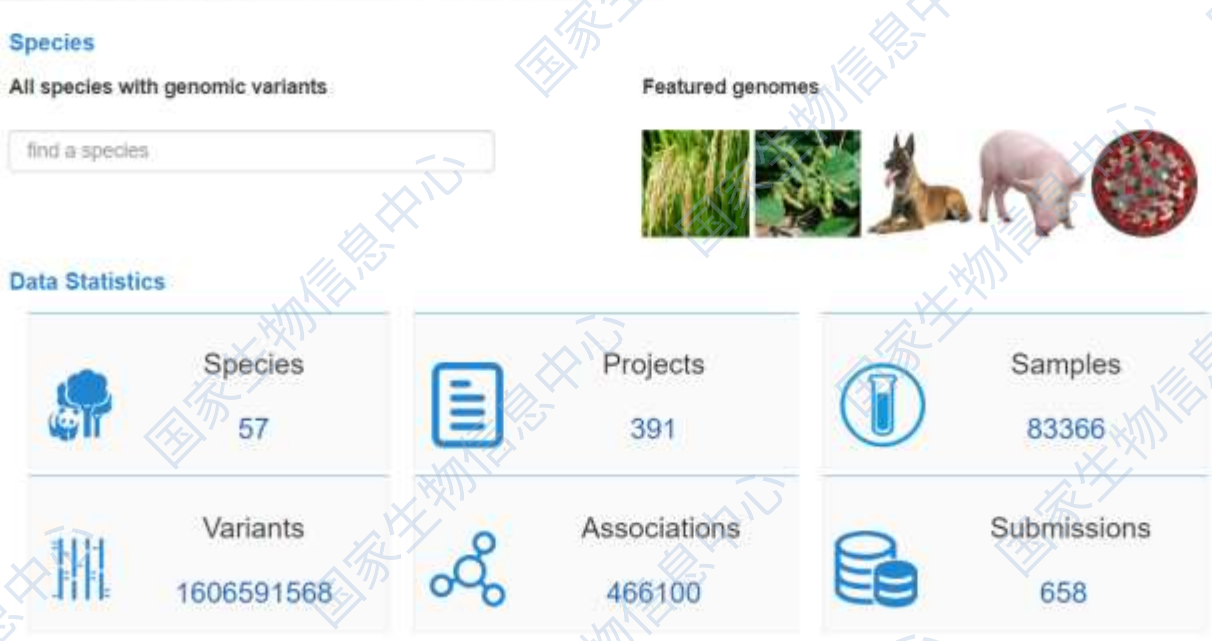
The National Genomic Data Center (NGDC), working collaboratively with multiple partner institutions/laboratories, develops the **Variome Data Standards** for variation data representation, analysis, search and exchange. We suggest submitters or users to follow these standards in preparing the metadata data, files and data analysis. We also give a detailed description on nomenclature standards in GVM about variation ID, annotation nomenclature et.al. These standards are our version 2.0 beta; please feel free to contact us (gvm AT big.ac.cn) if you have any suggestions.

- **1. Metadata**
  - *1.1 Submitter details*
  - *1.2 Project info*
  - *1.3 Sample details*
  - *1.4 Analysis methods*
  - *1.5 File names*
- **2. File formats**
  - *2.1 VCF*
  - *2.2 HapMap*
- **3. Data analysis standards**
  - *3.1 SNP & small I/D*
    - *3.1.1 Sequencing data analysis*
    - *3.1.2 Array data analysis*
  - *3.2 Large I/D (coming soon)*
  - *3.3 CNV (unavailable)*
  - *3.4 SV(unavailable)*
- **4. Nomenclature standards**
  - *4.1 GVM Nomenclature*
  - *4.2 Variation ID Nomenclature*
    - *4.2.1 SNP & small I/D*
    - *4.2.2 Large I/D (NA)*
    - *4.2.3 CNV (NA)*
    - *4.2.4 SV (NA)*
  - *4.3 Shorten name of the organisms*
  - *4.4 Variation annotation*
  - *4.5 Pfam Annotate*
  - *4.6 Nucleotides and amino acid nomenclature*
- **5. Tutorials**



https://ngdc.cncb.ac.cn/gvm/analysis_standards

资源审编与整合



Species

All species with genomic variants

find a species

Featured genomes

Data Statistics

Species 57

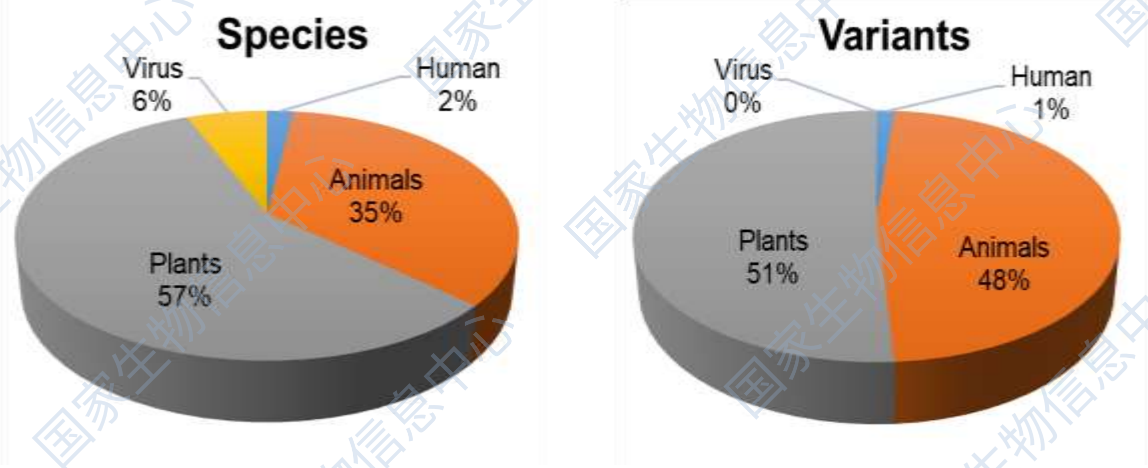
Projects 391

Samples 83366

Variants 1606591568

Associations 466100

Submissions 658

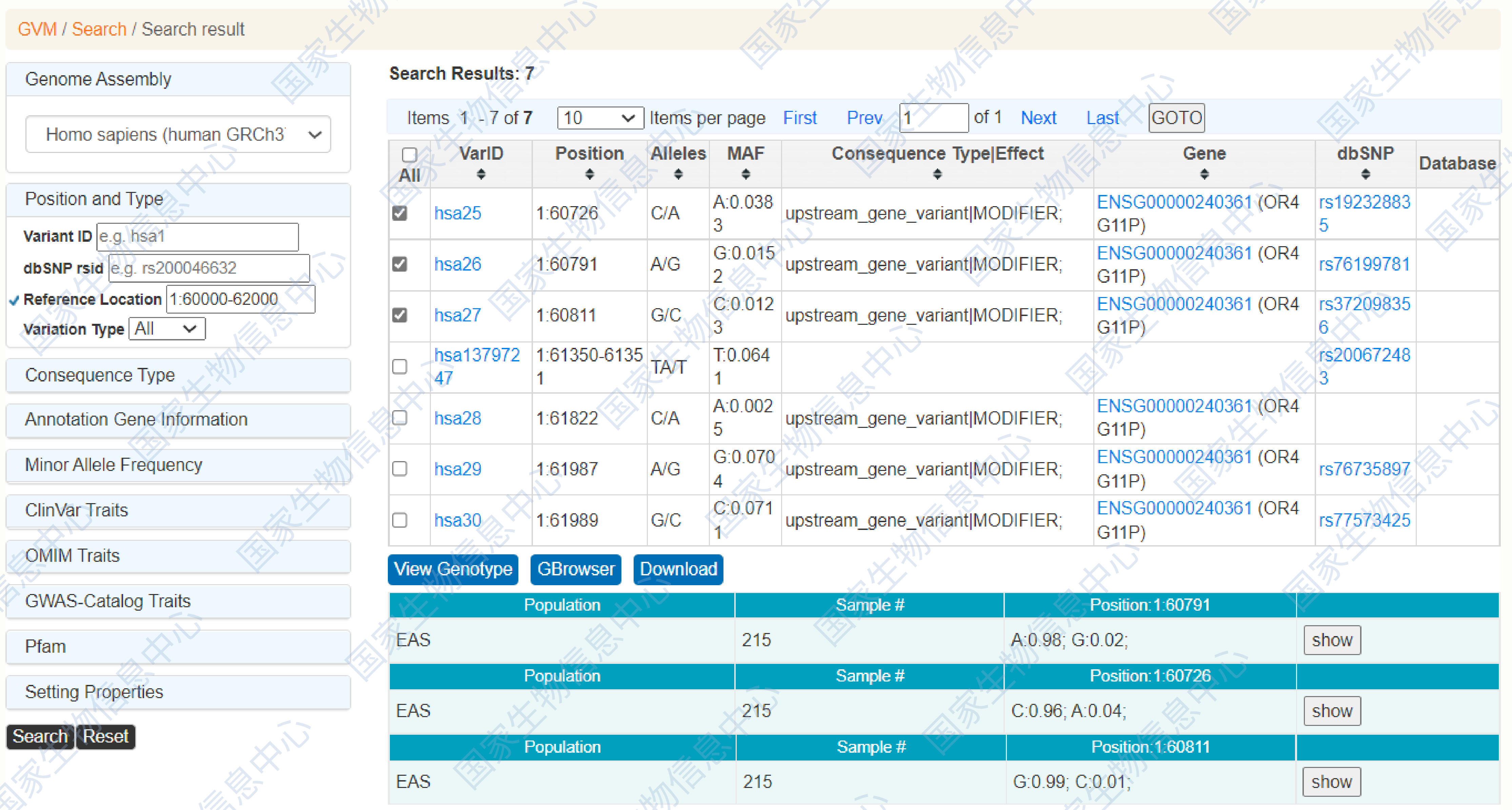


植物数据分布

https://ngdc.cncb.ac.cn/gvm/showSpecies

数据检索与可视化

国家基因组科学数据中心
National Genomics Data Center



GVM / Search / Search result

Genome Assembly

Homo sapiens (human GRCh3

Position and Type

Variant ID: e.g. hsa1
dbSNP rsid: e.g. rs200046632
✓ Reference Location: 1:60000-62000
Variation Type: All

Search Results: 7

Items 1 - 7 of 7 | 10 Items per page | First | Prev | 1 of 1 | Next | Last | GOTO

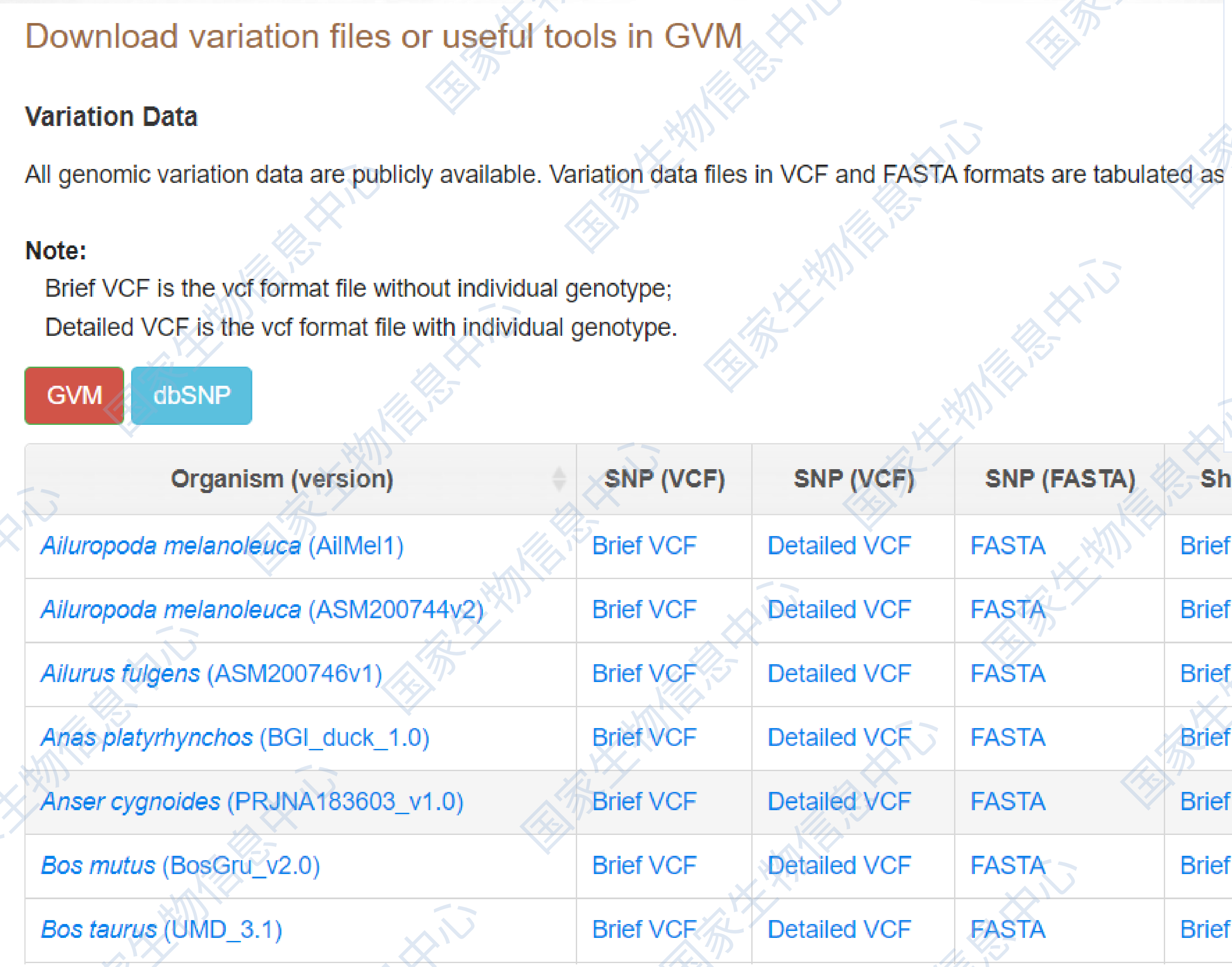
☐ All	VarID	Position	Alleles	MAF	Consequence Type|Effect	Gene	dbSNP	Database
☑	hsa25	1:60726	C/A	A:0.0383	upstream_gene_variant|MODIFIER;	ENSG00000240361 (OR4G11P)	rs192328835	
☑	hsa26	1:60791	A/G	G:0.0152	upstream_gene_variant|MODIFIER;	ENSG00000240361 (OR4G11P)	rs76199781	
☑	hsa27	1:60811	G/C	C:0.0123	upstream_gene_variant|MODIFIER;	ENSG00000240361 (OR4G11P)	rs372098356	
☐	hsa13797247	1:61350-61351	TA/T	T:0.0641			rs200672483	
☐	hsa28	1:61822	C/A	A:0.0025	upstream_gene_variant|MODIFIER;	ENSG00000240361 (OR4G11P)		
☐	hsa29	1:61987	A/G	G:0.0704	upstream_gene_variant|MODIFIER;	ENSG00000240361 (OR4G11P)	rs76735897	
☐	hsa30	1:61989	G/C	C:0.0711	upstream_gene_variant|MODIFIER;	ENSG00000240361 (OR4G11P)	rs77573425	

Population	Sample #	Position:1:60791	
EAS	215	A:0.98; G:0.02;	show

Population	Sample #	Position:1:60726	
EAS	215	C:0.96; A:0.04;	show

Population	Sample #	Position:1:60811	
EAS	215	G:0.99; C:0.01;	show

数据开放与共享



Download variation files or useful tools in GVM

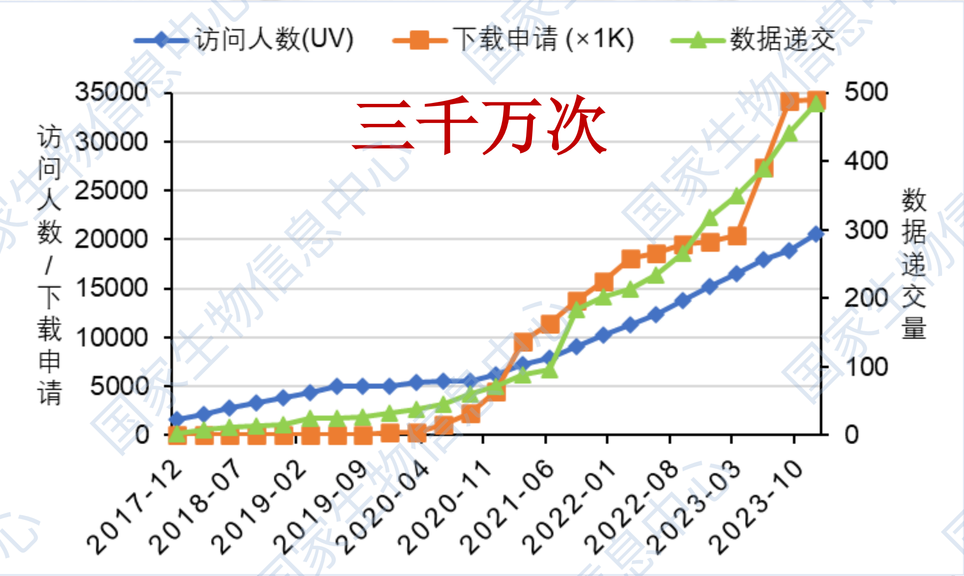
Variation Data

All genomic variation data are publicly available. Variation data files in VCF and FASTA formats are tabulated as

Note:

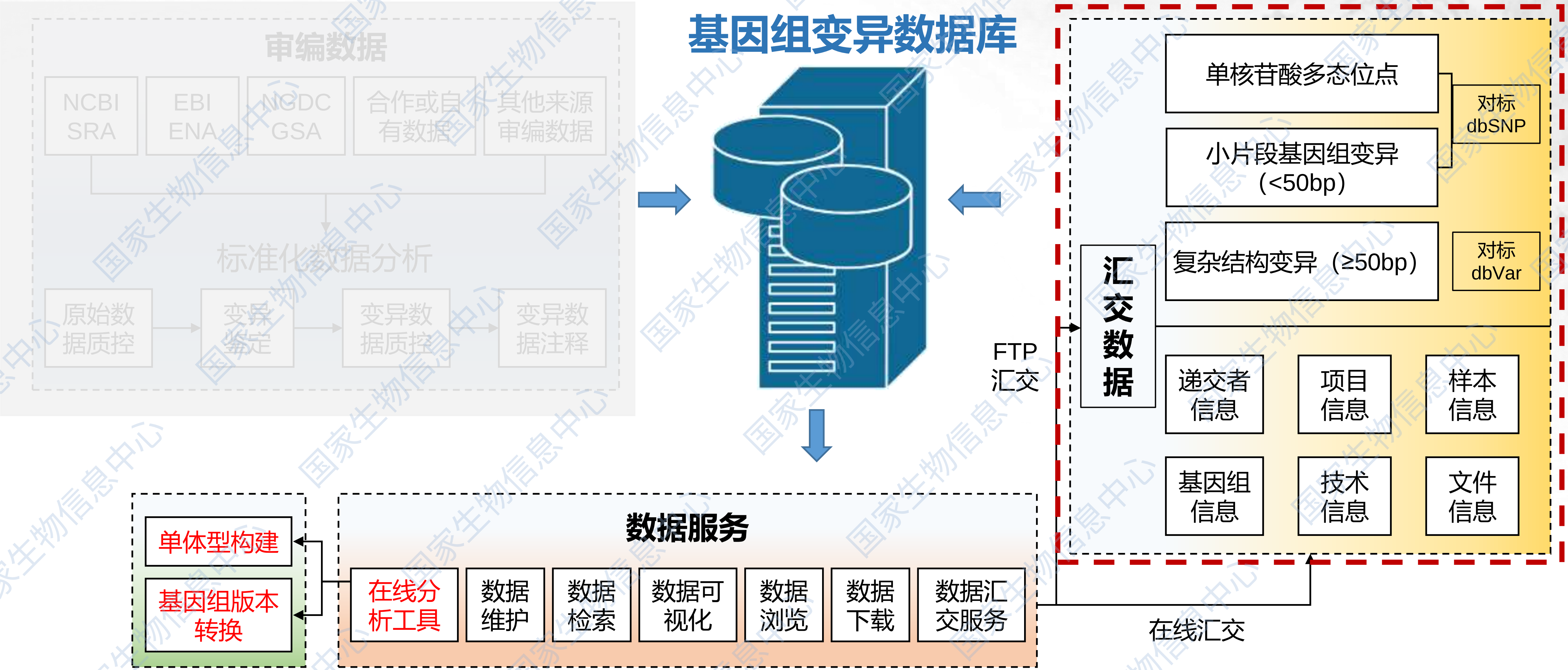
Brief VCF is the vcf format file without individual genotype;
Detailed VCF is the vcf format file with individual genotype.

Organism (version)	SNP (VCF)	SNP (VCF)	SNP (FASTA)	Short INDEL (VCF)	Short INDEL (VCF)	Short INDEL (FASTA)
Ailuropoda melanoleuca (AilMel1)	Brief VCF	Detailed VCF	FASTA	Brief VCF	Detailed VCF	FASTA
Ailuropoda melanoleuca (ASM200744v2)	Brief VCF	Detailed VCF	FASTA	Brief VCF	Detailed VCF	FASTA
Ailurus fulgens (ASM200746v1)	Brief VCF	Detailed VCF	FASTA	Brief VCF	Detailed VCF	FASTA
Anas platyrhynchos (BGI_duck_1.0)	Brief VCF	Detailed VCF	FASTA	Brief VCF	Detailed VCF	FASTA
Anser cygnoides (PRJNA183603_v1.0)	Brief VCF	Detailed VCF	FASTA	Brief VCF	Detailed VCF	FASTA
Bos mutus (BosGru_v2.0)	Brief VCF	Detailed VCF	FASTA	Brief VCF	Detailed VCF	FASTA
Bos taurus (UMD_3.1)	Brief VCF	Detailed VCF	FASTA	Brief VCF	Detailed VCF	FASTA



三千万次

基因组变异数据库组织框架



数据递交与管理

GVM is a repository specialized for archiving genome variantion 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 **English** **Chinese**

中英文帮助文档，仔细阅读，减少错误率！

https://ngdc.cncb.ac.cn/gvm/submit/usersubmit

数据递交与管理

n, including c.175+1G>A, c.73-2A> G, c.2 91+1G>A and c.481delA.

GVM / All Submission / GVM000757

Submission	**GVM000757** 2024-05-20 Sun Yat-sen University Cancer Center
Organism	Homo sapiens
Version	GRCh37
BioProject	PRJCA026167
Sample numbers	228
Abstract	A retrospective study on the identification of gene signature for immunotherapy in urological tumors
Data Accessibility:	Controlled access Request
Release date	2024-05-16
Available data	

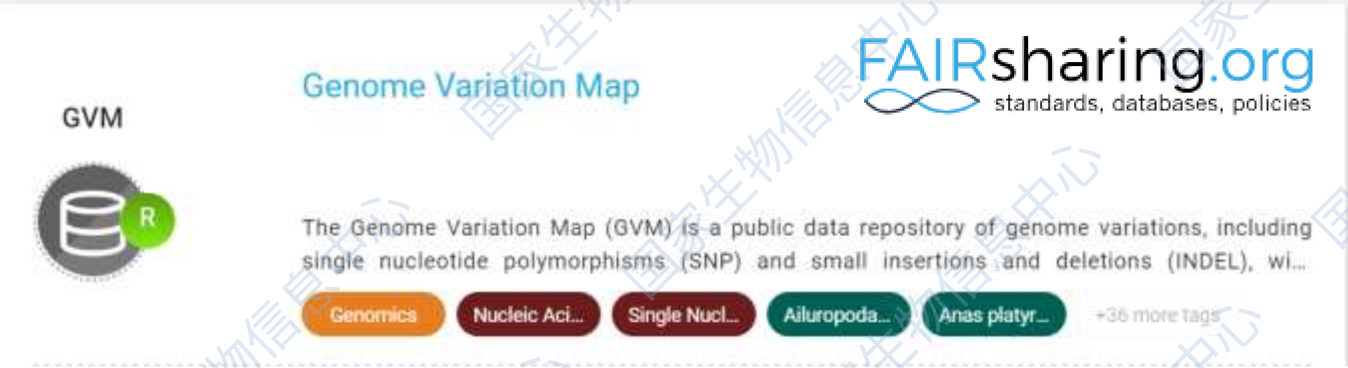
data submitter: shiyx@sysucc.org.cn

2110115928-1_2110115928-1.normal.vep.vcf
2109034337-1_2109034337-1.normal.vep.vcf
2108092984-1_2108092984-1.normal.vep.vcf

人遗相关数据，受控共享
备案后，可选择公开共享

https://ngdc.cncb.ac.cn/gvm/submit/allsubmit

数据递交与管理



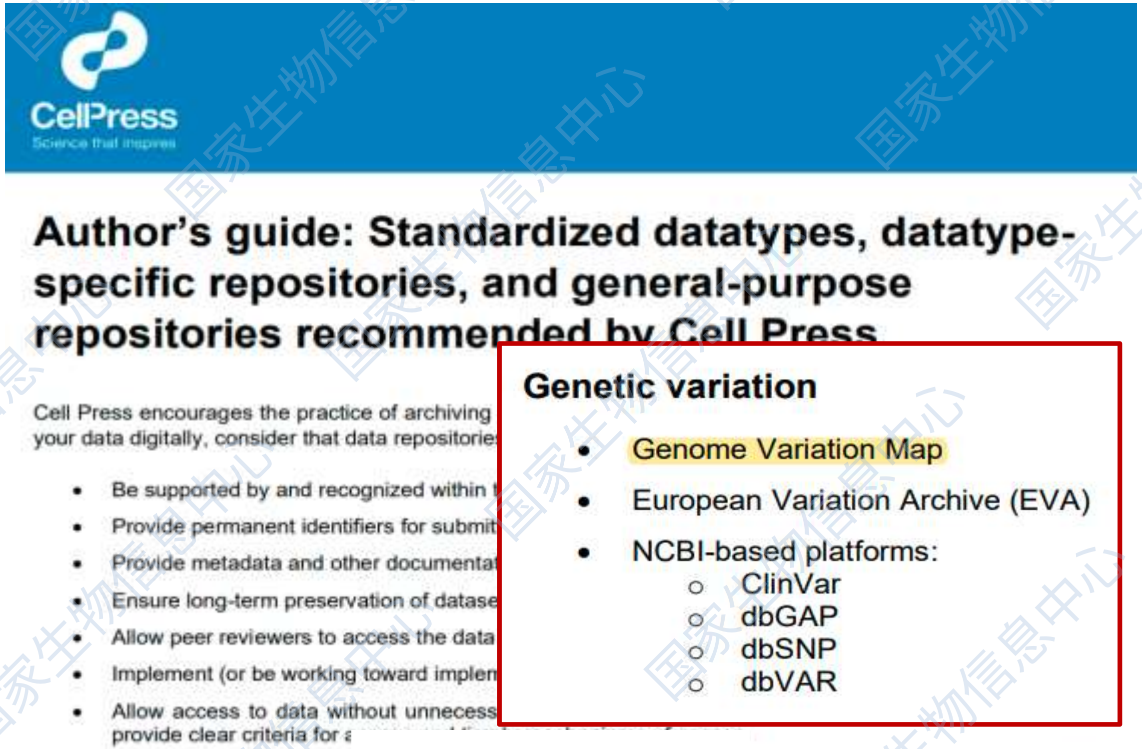
FAIRsharing.org
standards, databases, policies

GVM

Genome Variation Map

The Genome Variation Map (GVM) is a public data repository of genome variations, including single nucleotide polymorphisms (SNP) and small insertions and deletions (INDEL), wi...

https://fairsharing.org/search?q=GVM



Author's guide: Standardized datatypes, datatype-specific repositories, and general-purpose repositories recommended by Cell Press

Cell Press encourages the practice of archiving
your data digitally, consider that data repositorie

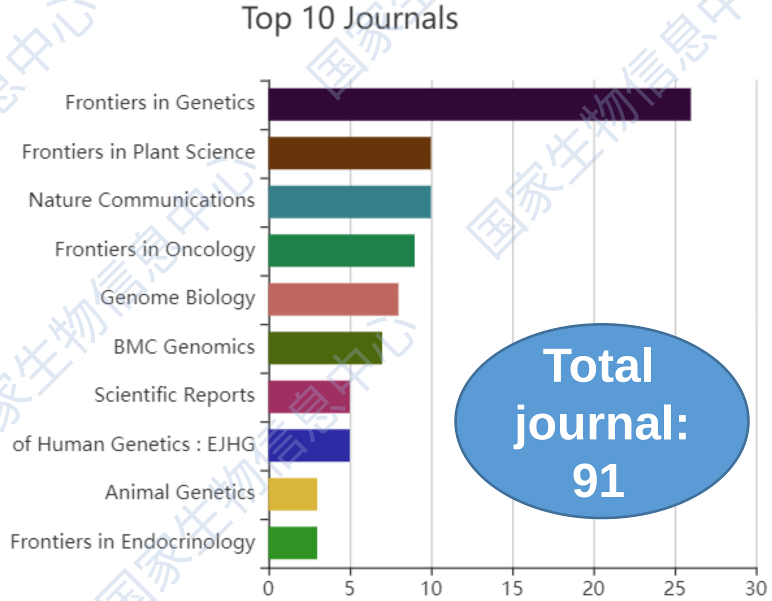
- Be supported by and recognized within t
- Provide permanent identifiers for submit
- Provide metadata and other documentat
- Ensure long-term preservation of datase
- Allow peer reviewers to access the data
- Implement (or be working toward implem
- Allow access to data without unnecess
provide clear criteria for a

Genetic variation

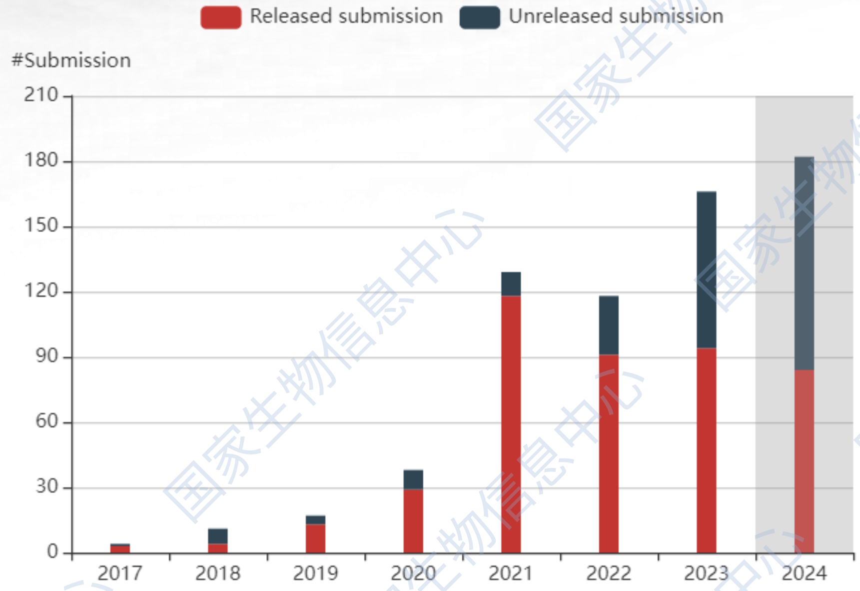
- Genome Variation Map
- European Variation Archive (EVA)
- NCBI-based platforms:
  - ClinVar
  - dbGAP
  - dbSNP
  - dbVAR

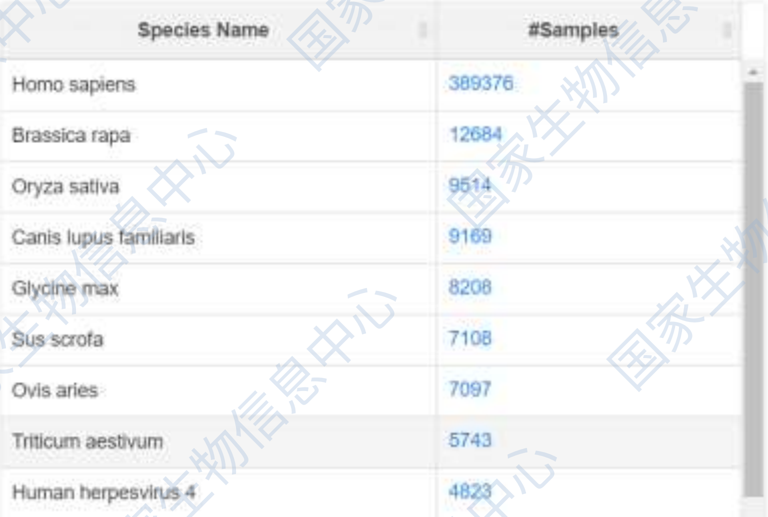
https://www.cell.com/

支持发表文章统计

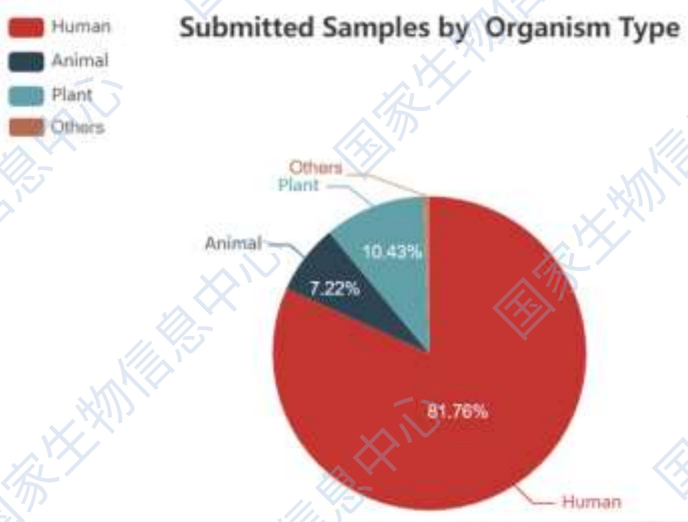


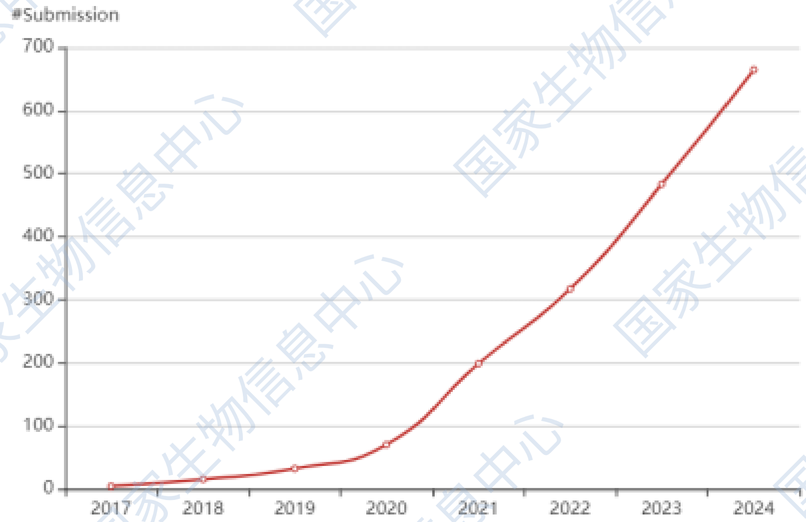
数据递交与管理

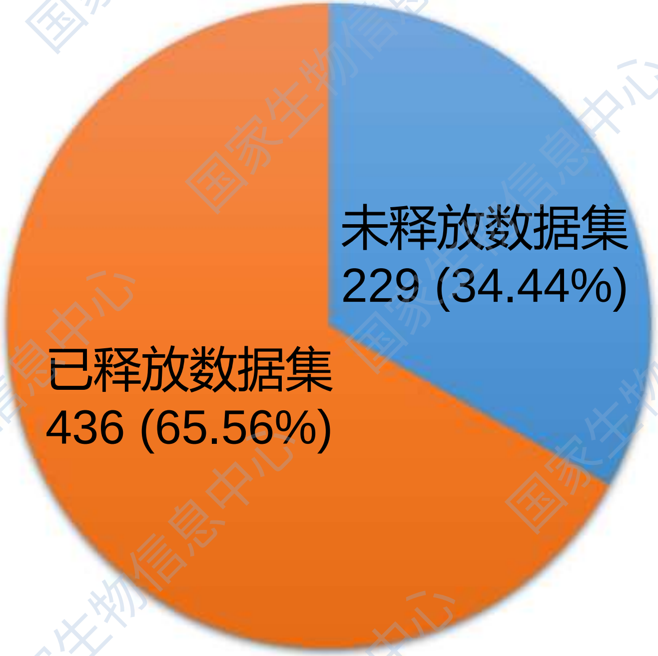




Species Name	#Samples
Homo sapiens	389376
Brassica rapa	12684
Oryza sativa	9514
Canis lupus familiaris	9169
Glycine max	8208
Sus scrofa	7108
Ovis aries	7097
Triticum aestivum	5743
Human herpesvirus 4	4823







数据的**所有权**在数据递交者，未发表的数据**是否公开共享**由数据递交者决定！

文章发表后，需公开数据！

https://ngdc.cncb.ac.cn/gvm/statistics

提纲

01 | 基因组序列变异数据库GVM

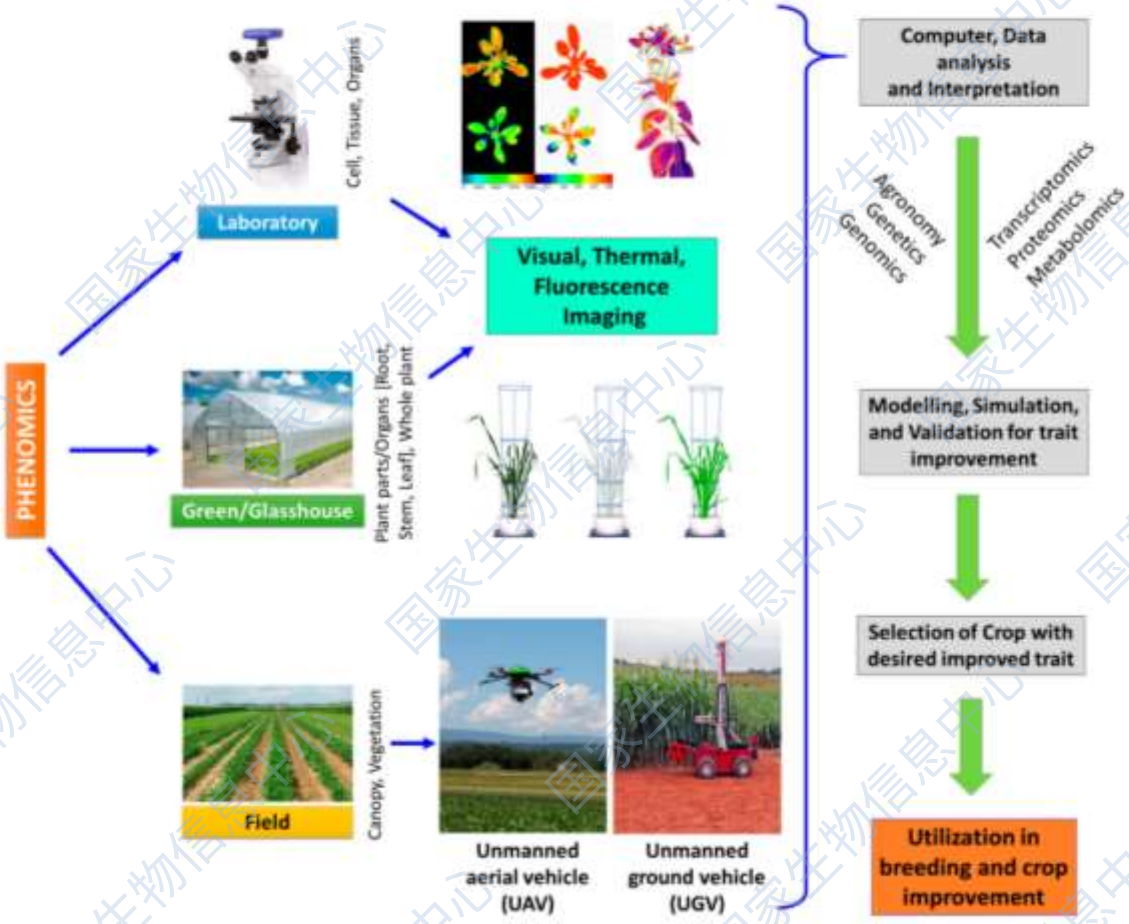
02 | 植物图像及其表型归档库OPIA

03 | G2P关联知识库GWAS Atlas

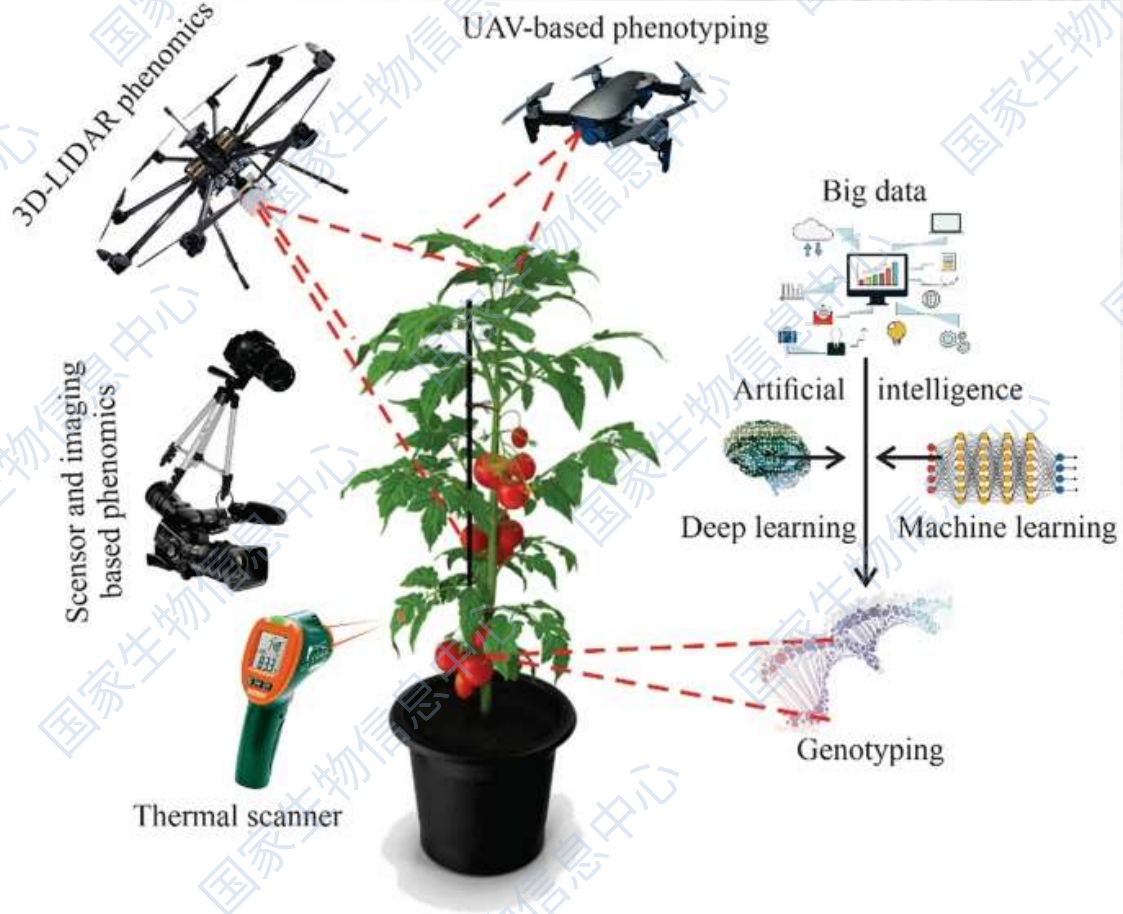
04 | 用户变异与表型数据汇交与共享

植物图像是表型组学中重要组成

植物表型数据是育种的关键，植物图像可以定量描述植物的形态特征、组织结构和发育进程等，基于图像的植物表型数据是植物表型组学中重要的数据来源。



Roychowdhury, Das et al. Genes (2023)

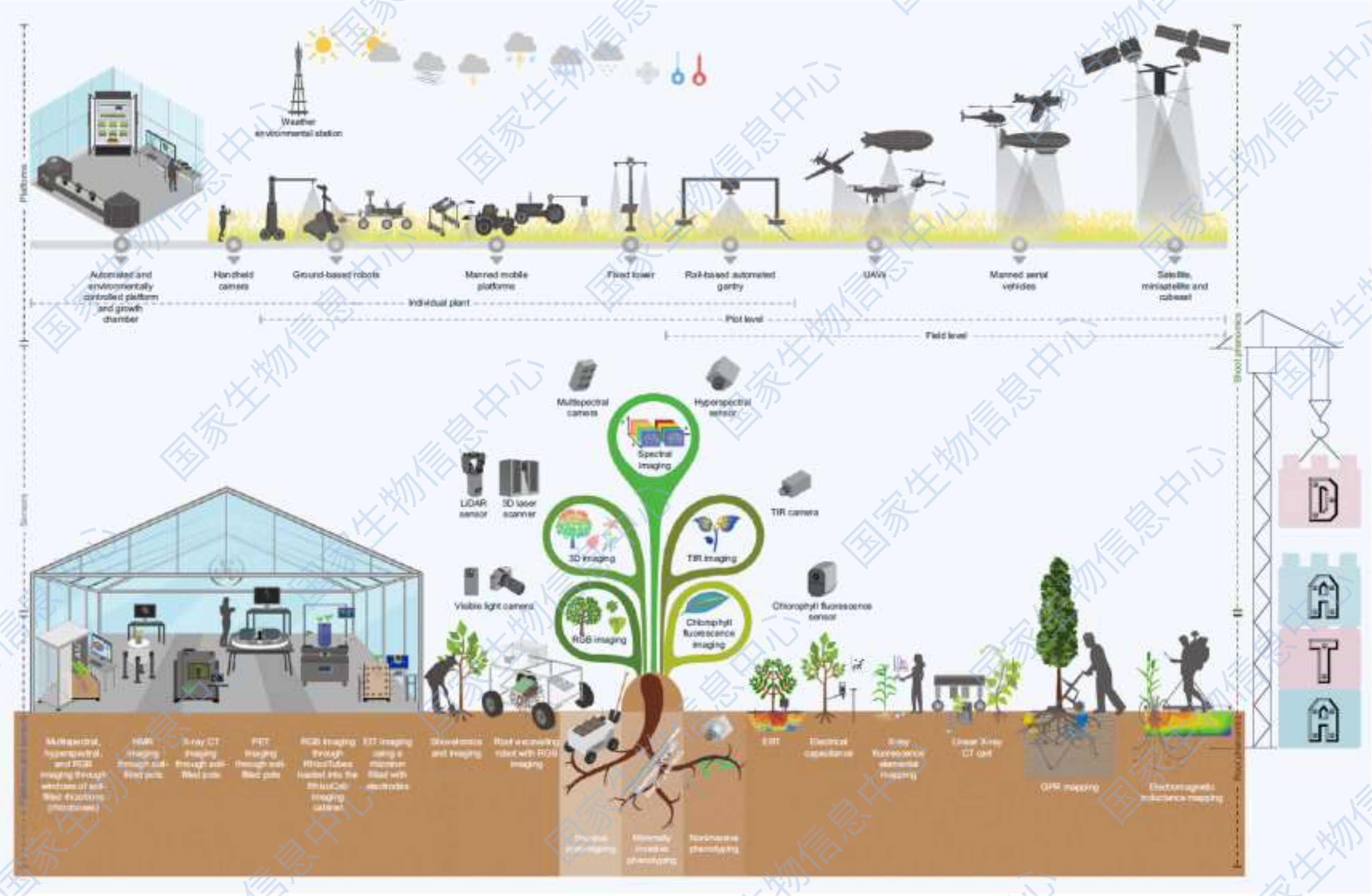


Integration of high-throughtput technologies in phenomics of sustainable agriculture

Waseem, M. et al. eBook (2022)

高通量表型平台获取多尺度图像数据

随着植物表型获取技术的进步，高通量植物表型平台的应用也越来越广泛



Harfouche, Nakhle et al. Trends Plant Sci. 2023

不同表型尺度：

- 单个植株 (individual plant)
- 地块 (plot level)
- 田间 (filed level)

不同相机/传感器：

- 可见光 (RGB imaging)
- 高光谱 (hyperspectral)
- 激光雷达 (LiDAR)
- 热红外 (thermal infrared)
- 叶绿素荧光 (CF)
- ……

开放植物图像表型归档库OPIA

构建植物图像及其相关表型数据归档库，致力于提供植物图像及基于图像获得的表型组学数据共享，为种质筛选、植物病虫害鉴定、农艺性状挖掘等应用提供的重要资源。

Open Plant Image Archive (OPIA) is an archive database of plant images and related phenotypic data from high-throughput phenotyping platforms for a diversity of species. view more...

2023 09-20 Two machine learning models were added.

2023 08-05 I-traits of 2 wheat datasets were onlined.

2023 07-25 20 new image datasets of Wheat were added.

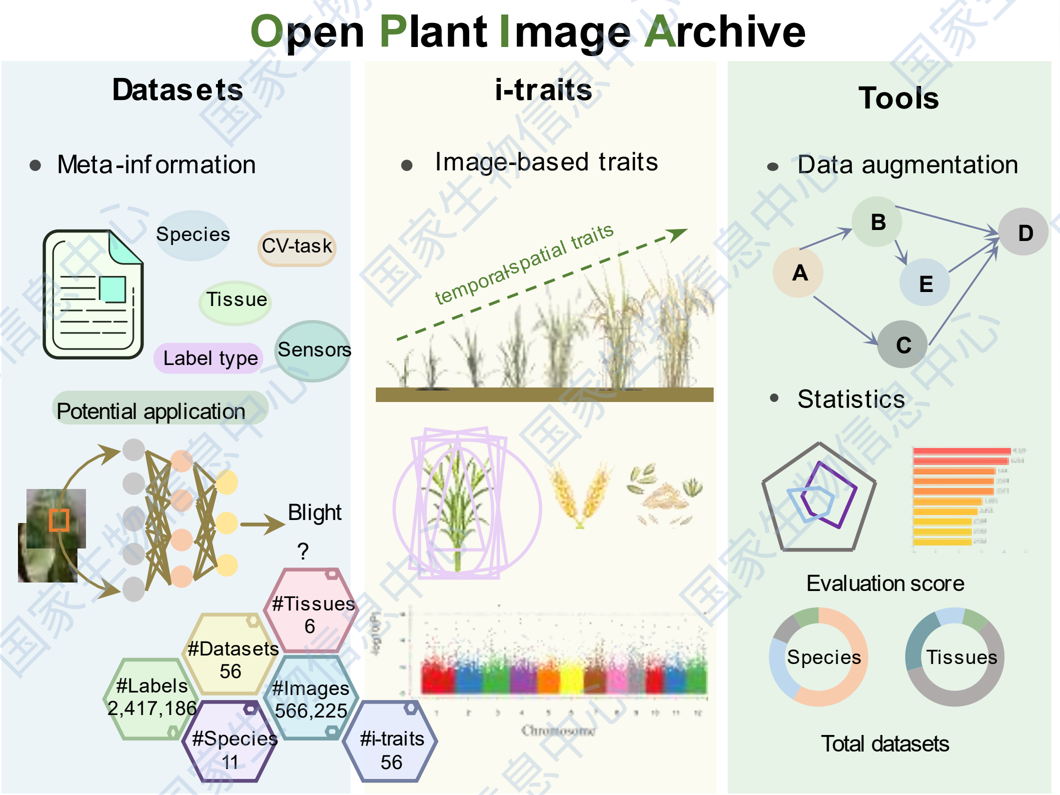
2023 06-21 The OPIA was online, including 36 image datasets.

Title: UAV...

Species: Rice (Oryza sativa)

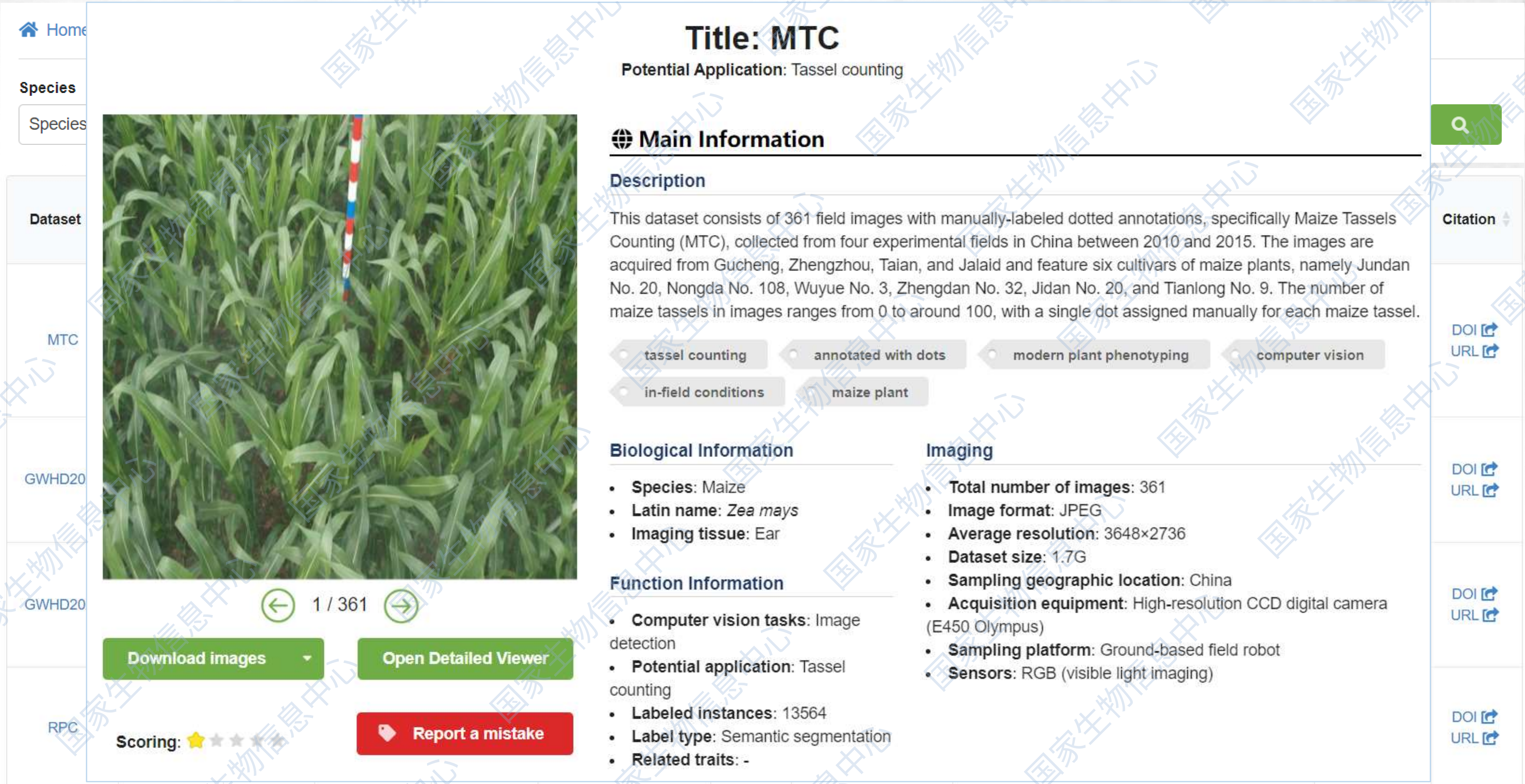
image: 109 slides

Potential application: Plant counting, locating and sizing



https://ngdc.cncb.ac.cn/opia *Nucleic Acids Research* (2024)

数据集检索与可视化



Title: MTC

Potential Application: Tassel counting

Main Information

Description

This dataset consists of 361 field images with manually-labeled dotted annotations, specifically Maize Tassels Counting (MTC), collected from four experimental fields in China between 2010 and 2015. The images are acquired from Gucheng, Zhengzhou, Taian, and Jalaid and feature six cultivars of maize plants, namely Jundan No. 20, Nongda No. 108, Wuyue No. 3, Zhengdan No. 32, Jidan No. 20, and Tianlong No. 9. The number of maize tassels in images ranges from 0 to around 100, with a single dot assigned manually for each maize tassel.

Scoring: ★☆☆☆☆

Biological Information

- **Species**: Maize
- **Latin name**: *Zea mays*
- **Imaging tissue**: Ear

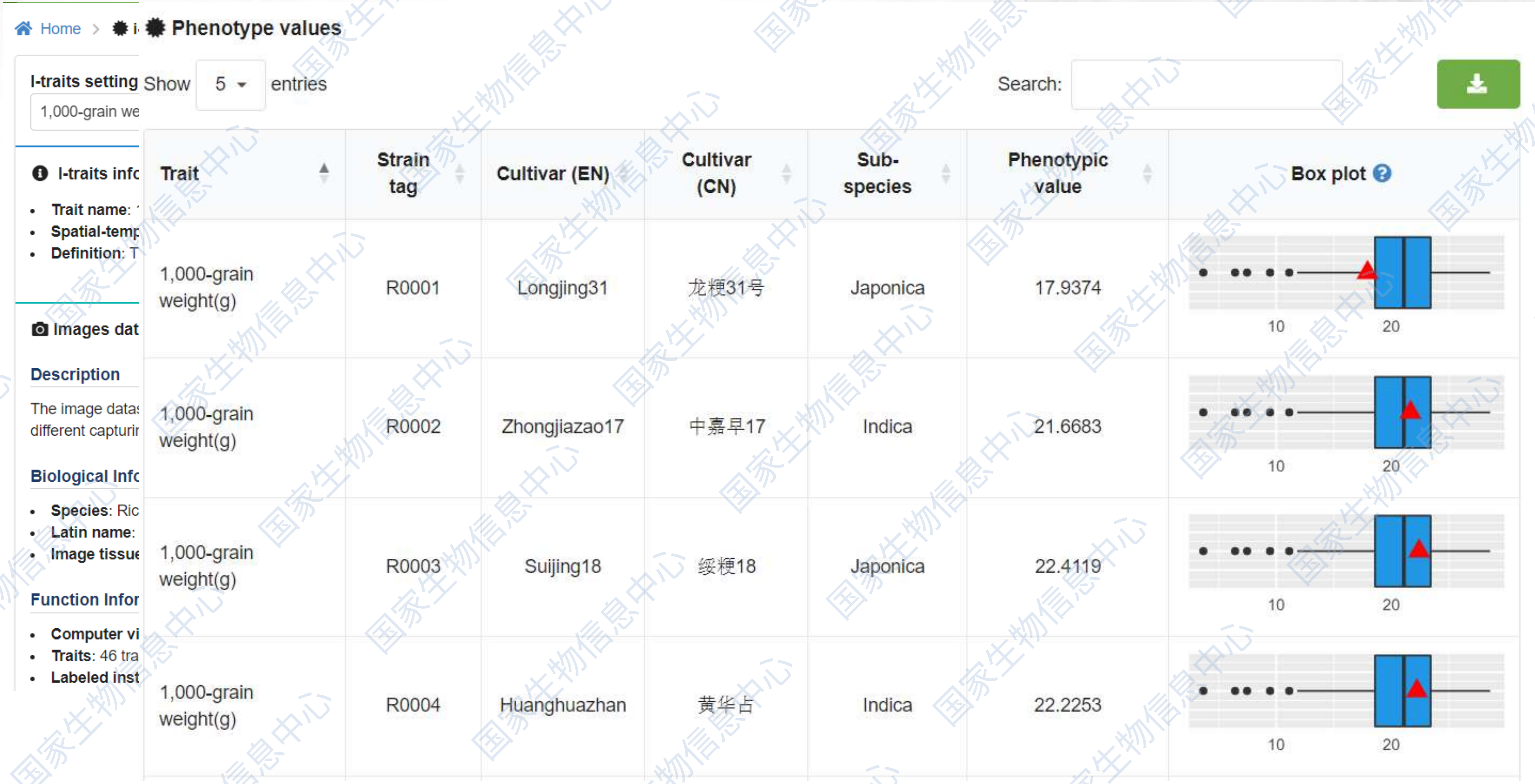
Function Information

- **Computer vision tasks**: Image detection
- **Potential application**: Tassel counting
- **Labeled instances**: 13564
- **Label type**: Semantic segmentation
- **Related traits**: -

Imaging

- **Total number of images**: 361
- **Image format**: JPEG
- **Average resolution**: 3648×2736
- **Dataset size**: 1.7G
- **Sampling geographic location**: China
- **Acquisition equipment**: High-resolution CCD digital camera (E450 Olympus)
- **Sampling platform**: Ground-based field robot
- **Sensors**: RGB (visible light imaging)

i-traits



Home › i › Phenotype values

I-traits setting
1,000-grain we

I-traits info
- **Trait name:**
- **Spatial-temp**
- **Definition:** T

Images dat

Description

The image datas different capturi

Biological Info
- **Species:** Ric
- **Latin name:**
- **Image tissue**

Function Infor
- **Computer vi**
- **Traits:** 46 tra
- **Labeled inst**

Show 5 entries Search:

Trait	Strain tag	Cultivar (EN)	Cultivar (CN)	Sub-species	Phenotypic value	Box plot
1,000-grain weight(g)	R0001	Longjing31	龙粳31号	Japonica	17.9374	
1,000-grain weight(g)	R0002	Zhongjiazao17	中嘉早17	Indica	21.6683	
1,000-grain weight(g)	R0003	Suijing18	绥粳18	Japonica	22.4119	
1,000-grain weight(g)	R0004	Huanghuazhan	黄华占	Indica	22.2253	

图像处理工具

Image Pre-processing Tools

Image Cropping

Image cropping can determine the target object within a small range, thereby reducing computational damage. This tool supports batch cropping of images.

Image Resizing

Image resizing can adjust the size of a single image or compressed image package to the target size. This tool supports batch resizing of images.

Image Flipping

Image flipping is one of the data augmentation methods used to increase the diversity of the dataset. This tool supports batch flipping of images.

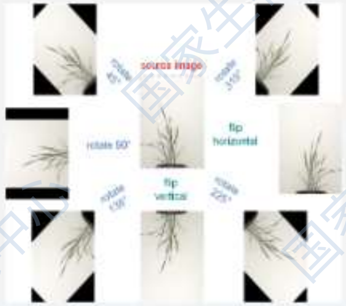


Image Denoising

Image denoising automatically calculates the type of noise in the image and selects the appropriate denoise algorithm for denoising.

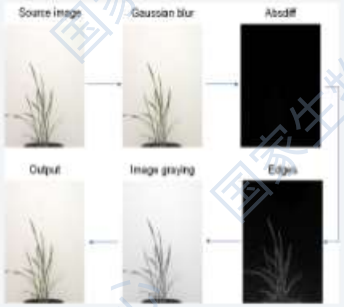
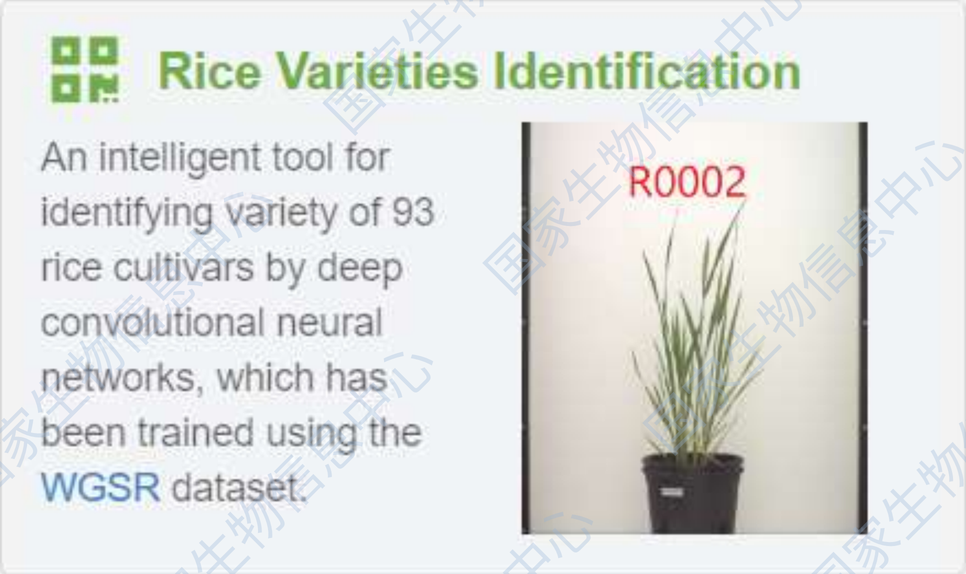


Image Partitioning

Data partitioning divides the original dataset into training set and testing set according to the label category and training set ratio.

图像的智能预测模型



Rice Varieties Identification

An intelligent tool for identifying variety of 93 rice cultivars by deep convolutional neural networks, which has been trained using the WGSR dataset.

Rice Yield Estimation

An intelligent web tool for predicting individual rice yield per plant by deep convolutional neural networks.

Wheat Head Estimation

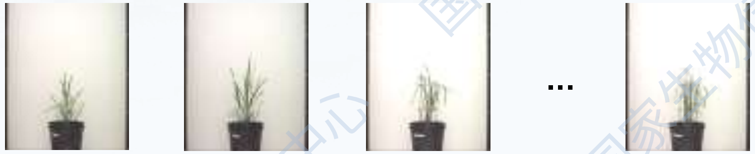
An online web tool for predicting the number of wheat heads, which has been pre-trained using the GWHD2020 dataset.

Model Name	Function	CV-task	Dataset	Download
VGG16	Estimation of rice yield per plant	Image classification	WGSR	Download
FastRCNN	Wheat head counting	Image detection	GWHD2020	Download

https://ngdc.cncb.ac.cn/opia/tools

基于深度学习的水稻品系识别模型

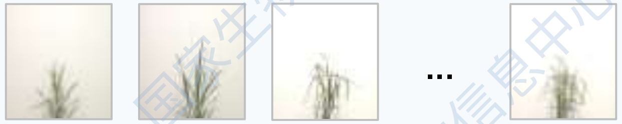
数据预处理



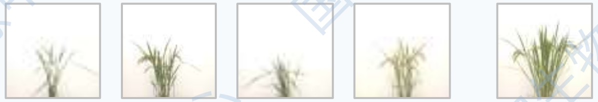
93 种水稻品系（61 japonica 32 indica）

5,585张RGB图像（3,580 japonica 2,005 indica）

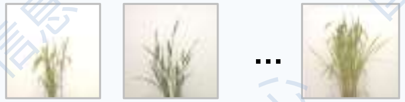
图像裁剪 & 图像分辨率调整



训练集、测试集划分（8 : 2）



训练集 4,468 (2,864 japonica 1604 indica)



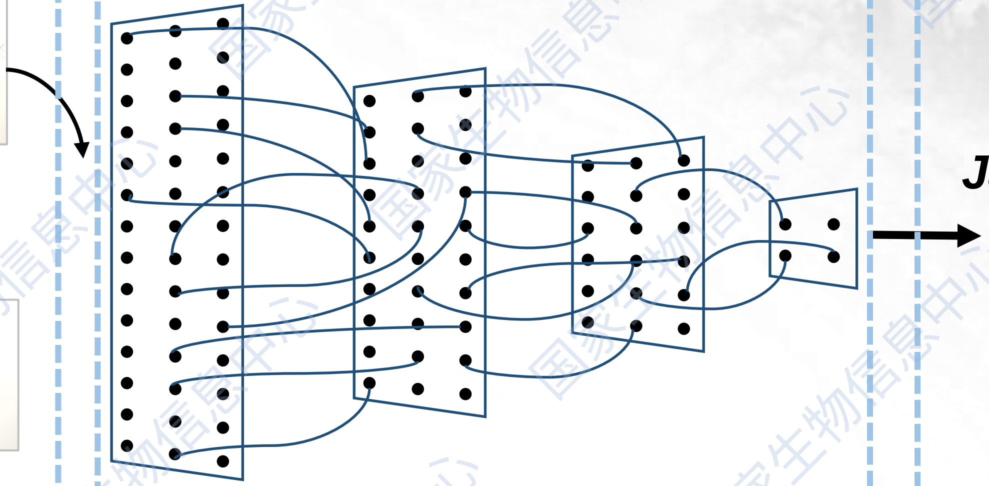
测试集 1,117 (716 japonica 401 indica)

输入

japonica

indica

卷积神经网络（CNN）



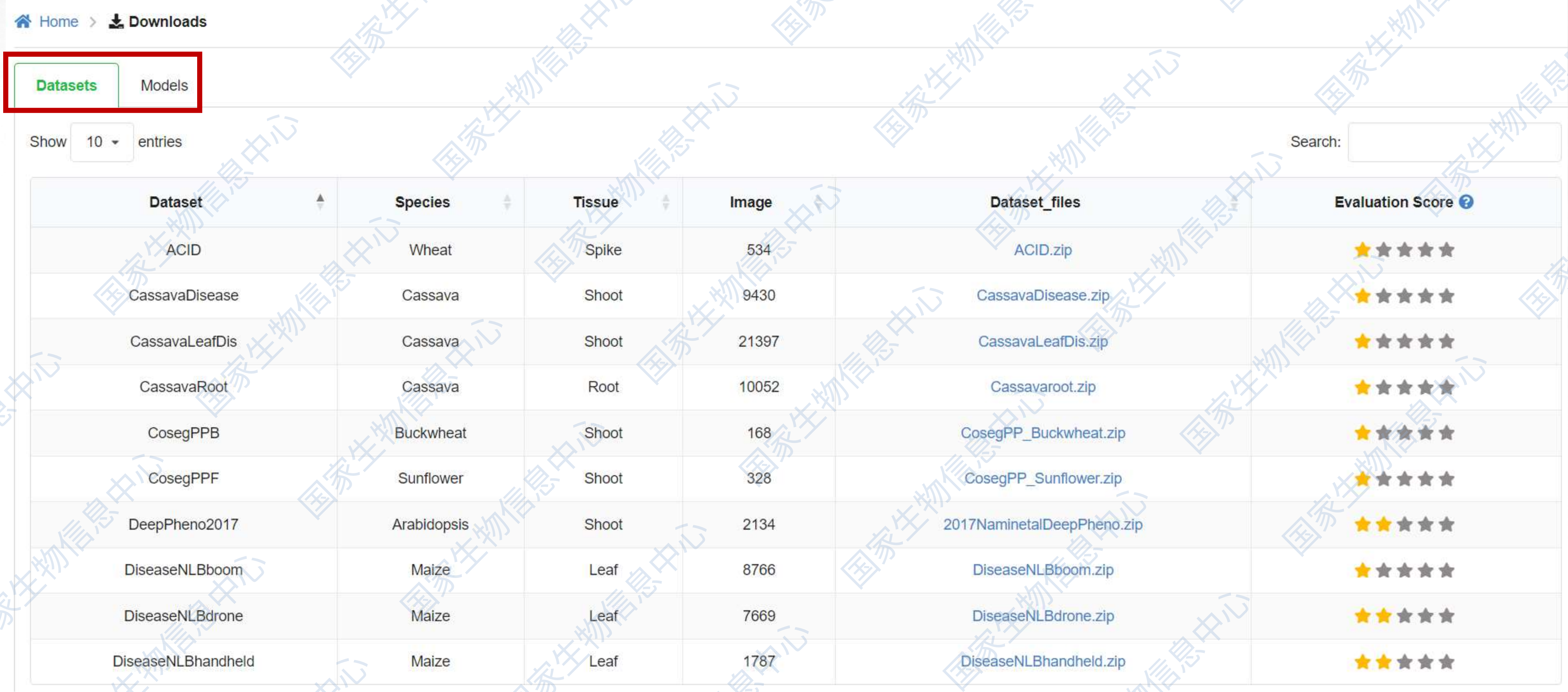
输出

Japonica ？

Indica ？

- ✓ 提供每张图像正确的标签（籼稻、粳稻）
- ✓ 构建深度学习模型
- ✓ Resnet34 Accuracy: 93.7332%

数据下载共享



Dataset	Species	Tissue	Image	Dataset_files	Evaluation Score
ACID	Wheat	Spike	534	ACID.zip	★☆☆☆☆
CassavaDisease	Cassava	Shoot	9430	CassavaDisease.zip	★☆☆☆☆
CassavaLeafDis	Cassava	Shoot	21397	CassavaLeafDis.zip	★☆☆☆☆
CassavaRoot	Cassava	Root	10052	Cassavaroot.zip	★☆☆☆☆
CosegPPB	Buckwheat	Shoot	168	CosegPP_Buckwheat.zip	★☆☆☆☆
CosegPPF	Sunflower	Shoot	328	CosegPP_Sunflower.zip	★☆☆☆☆
DeepPheno2017	Arabidopsis	Shoot	2134	2017NaminetalDeepPheno.zip	★★☆☆☆
DiseaseNLBboom	Maize	Leaf	8766	DiseaseNLBboom.zip	★☆☆☆☆
DiseaseNLBdrone	Maize	Leaf	7669	DiseaseNLBdrone.zip	★★☆☆☆
DiseaseNLBhandheld	Maize	Leaf	1787	DiseaseNLBhandheld.zip	★★☆☆☆

提纲

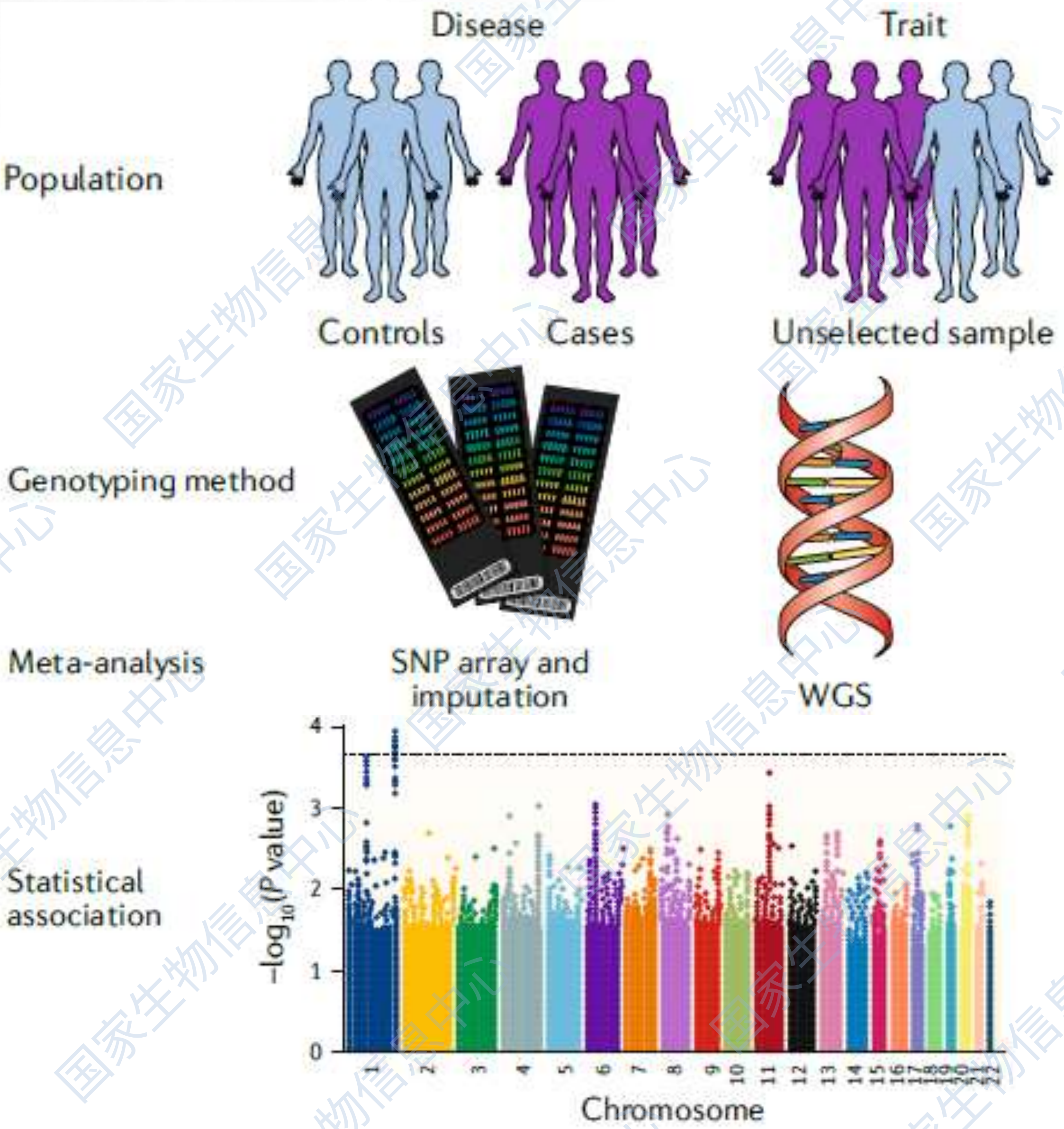
01 | 基因组序列变异数据库GVM

02 | 植物图像及其表型归档库OPIA

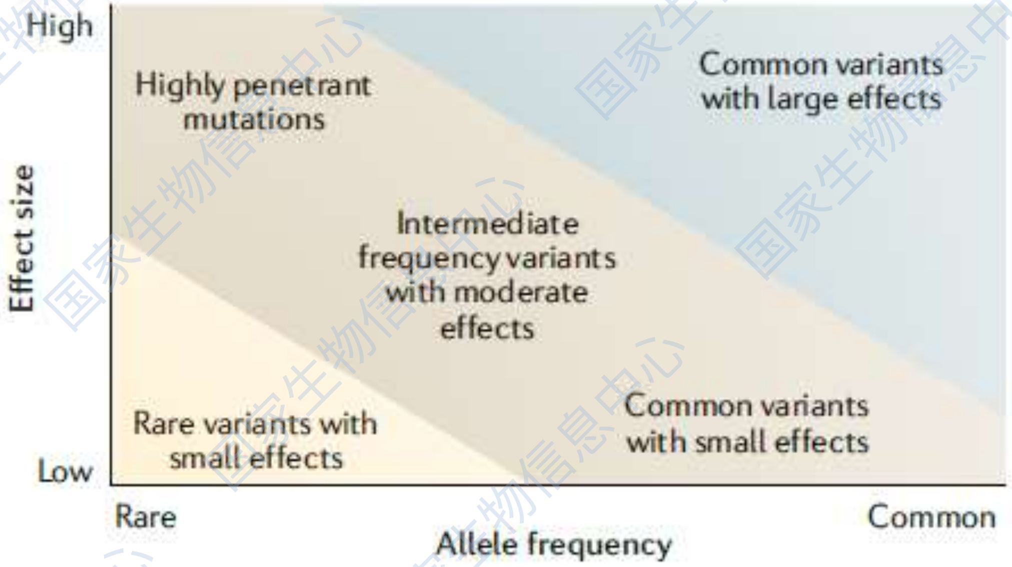
03 | G2P关联知识库GWAS Atlas

04 | 用户变异与表型数据汇交与共享

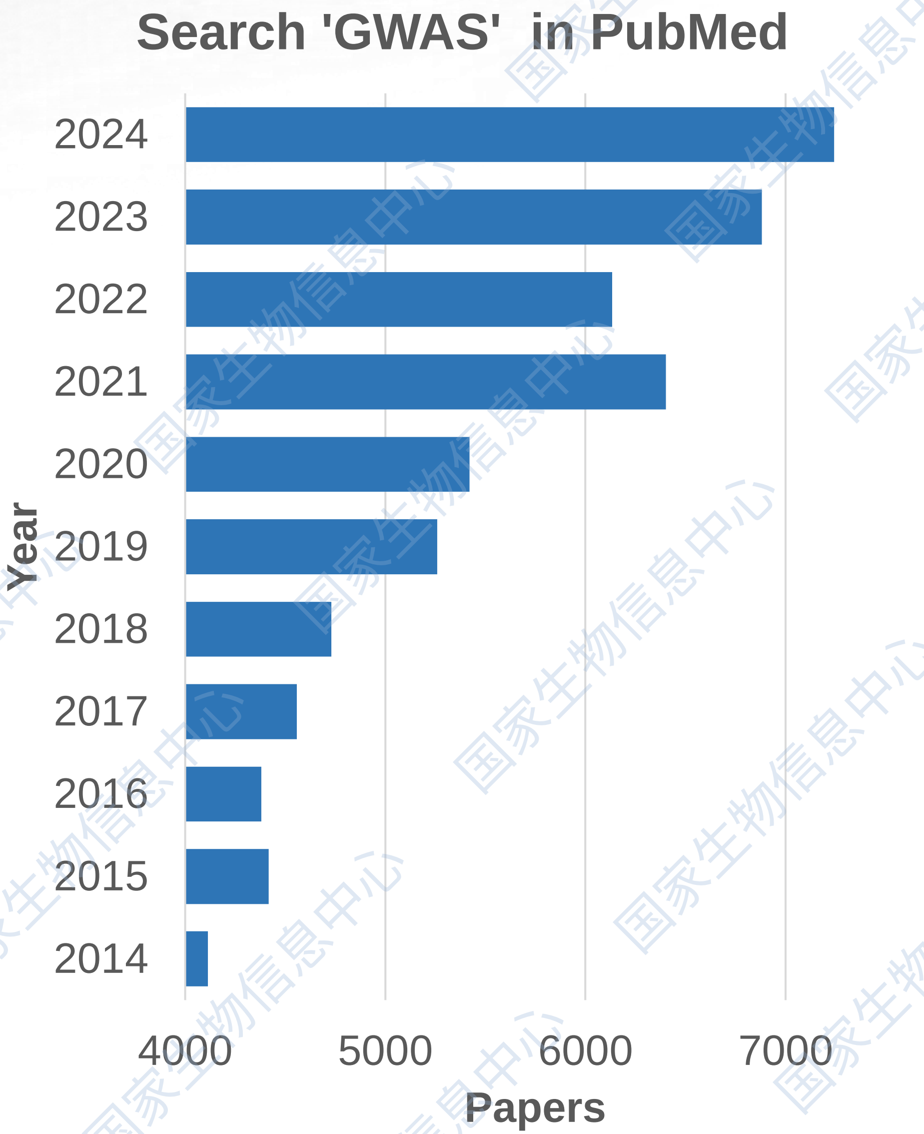
全基因组关联分析(GWAS)



全基因组关联分析（GWAS）是指在全基因组范围内，通过将群体基因型与表型进行关联分析，筛选出与表型相关联的基因型。

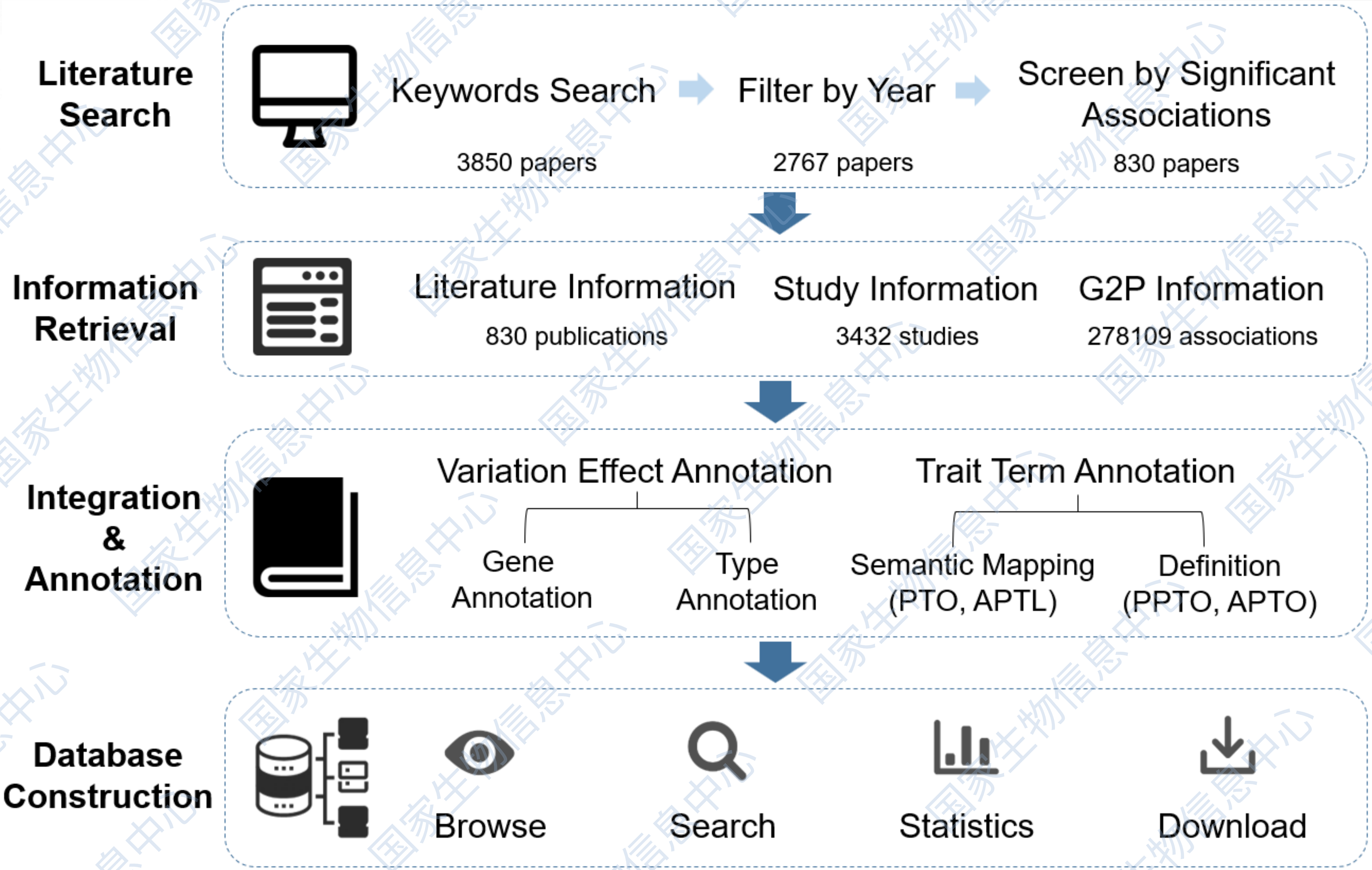


全基因组关联分析(GWAS)



Database Name	Host Institution	Species	Last update
GWAS Catalog	EBI	Human	2024
dbGaP	NCBI	Human	2024
GWAS Central	University of Leicester	Human	2020
GWAS ATLAS	VU University Amsterdam	Human	2019
GWASdb	University of HongKong	Human	2019
LD Hub	University of Bristol	Human	2016
AraGWAS Catalog	ETH Zurich	A. Thaliana	2022

GWAS 数据审编流程



Literature Search

Keywords Search → Filter by Year → Screen by Significant Associations

3850 papers | 2767 papers | 830 papers

Information Retrieval

Literature Information
830 publications

Study Information
3432 studies

G2P Information
278109 associations

Integration & Annotation

Variation Effect Annotation
- Gene Annotation
- Type Annotation

Trait Term Annotation
- Semantic Mapping (PTO, APTL)
- Definition (PPTO, APTO)

Database Construction

Browse

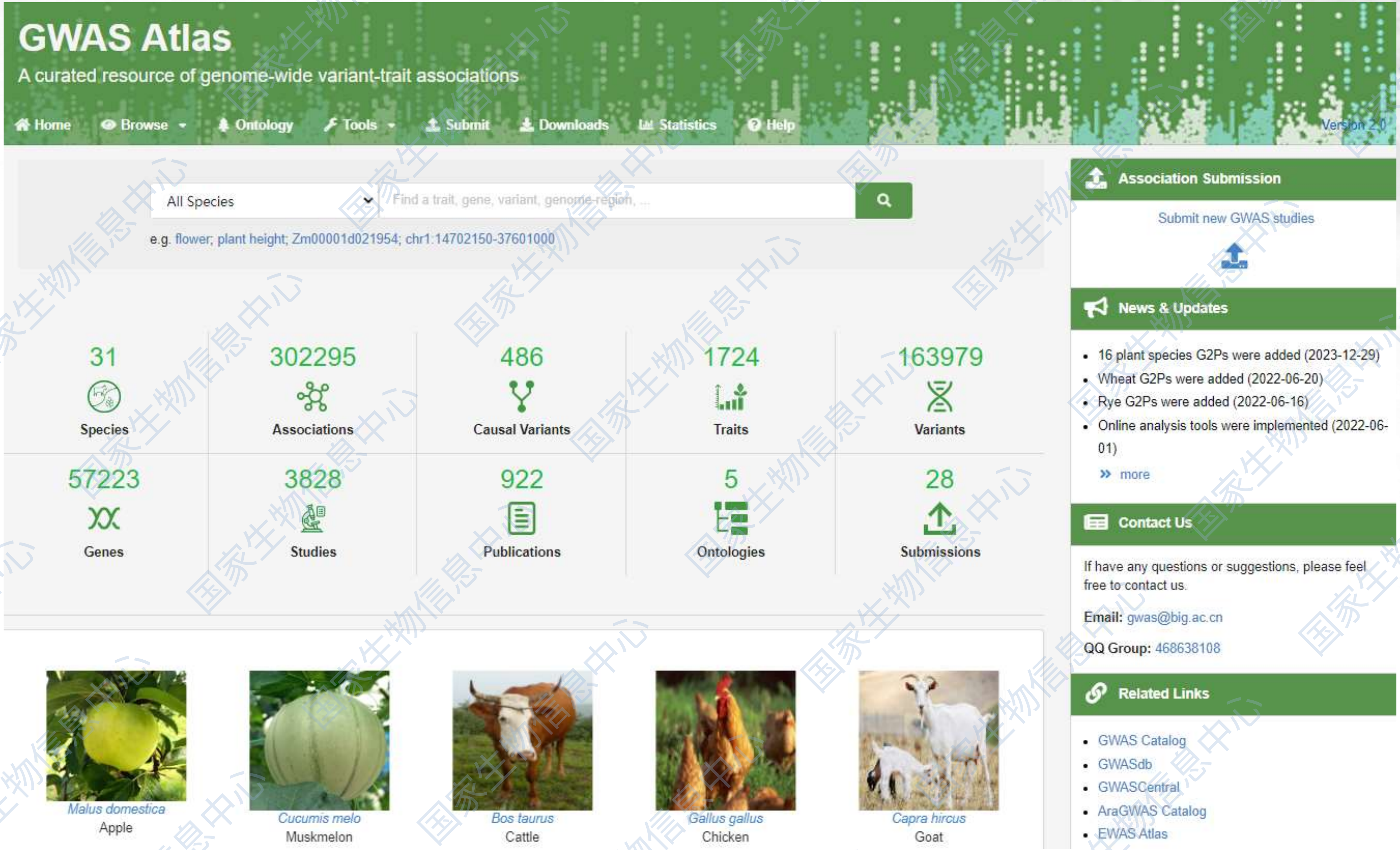
Search

Statistics

Download

基因型与表型关联知识库(GWAS Atlas)

国家基因组科学数据中心
National Genomics Data Center



GWAS Atlas

A curated resource of genome-wide variant-trait associations

e.g. flower; plant height; Zm00001d021954; chr1:14702150-37601000

31 Species	302295 Associations	486 Causal Variants	1724 Traits	163979 Variants
57223 Genes	3828 Studies	922 Publications	5 Ontologies	28 Submissions

Malus domestica Apple

Cucumis melo Muskmelon

Bos taurus Cattle

Gallus gallus Chicken

Capra hircus Goat

Association Submission

Submit new GWAS studies

News & Updates

- 16 plant species G2Ps were added (2023-12-29)
- Wheat G2Ps were added (2022-06-20)
- Rye G2Ps were added (2022-06-16)
- Online analysis tools were implemented (2022-06-01)

» more

Contact Us

If have any questions or suggestions, please feel free to contact us.

Email: gwas@big.ac.cn

QQ Group: 468638108

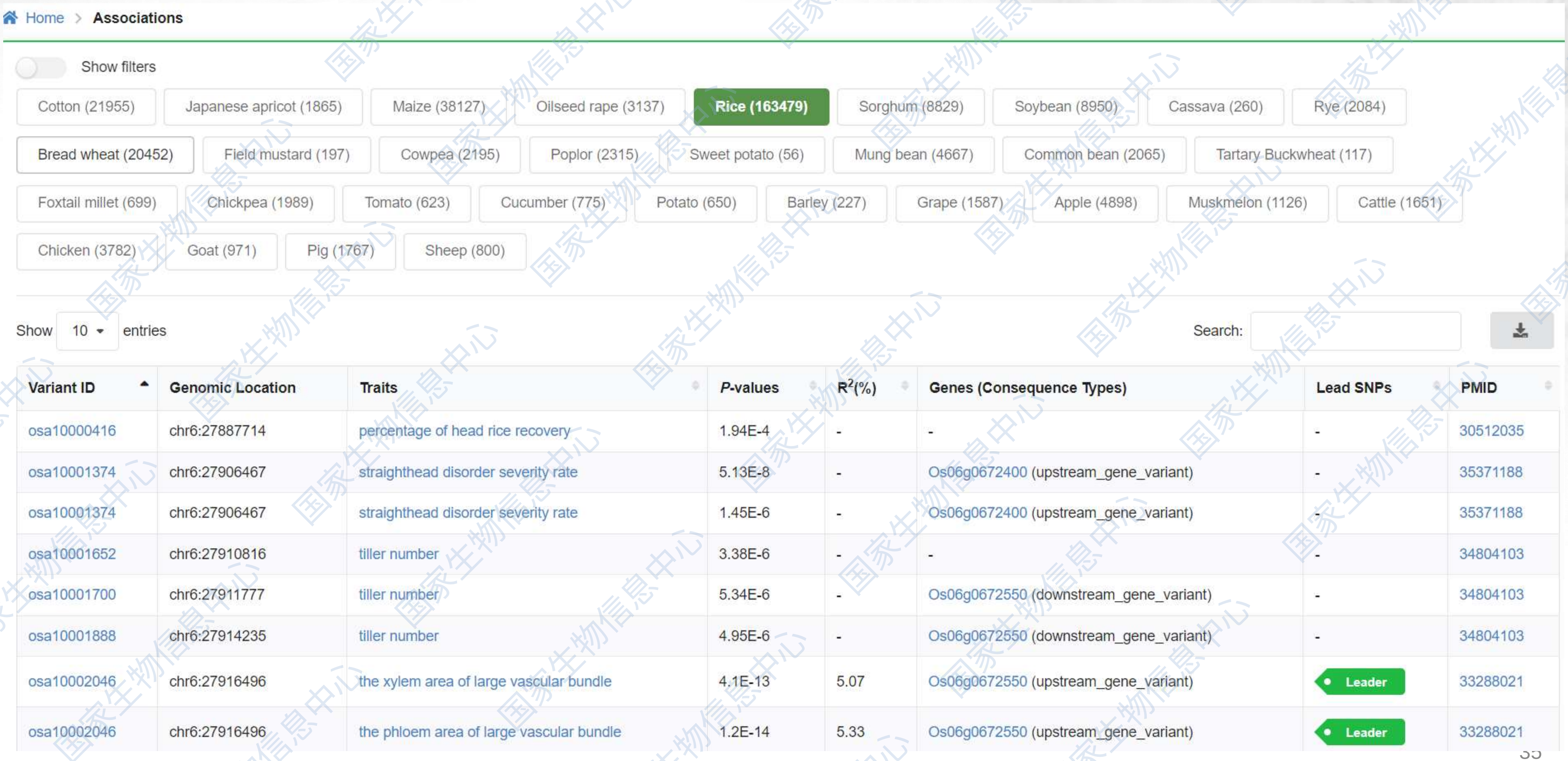
Related Links

- GWAS Catalog
- GWASdb
- GWASCentral
- AraGWAS Catalog
- EWAS Atlas

Liuxn et.al. 2022 NAR;   Tiandm et.al. 2020 NAR   https://ngdc.cncb.ac.cn/gwas

GWAS关联知识

国家基因组科学数据中心
National Genomics Data Center



Home > **Associations**

Show filters

Cotton (21955) | Japanese apricot (1865) | Maize (38127) | Oilseed rape (3137) | **Rice (163479)** | Sorghum (8829) | Soybean (8950) | Cassava (260) | Rye (2084)

Bread wheat (20452) | Field mustard (197) | Cowpea (2195) | Poplor (2315) | Sweet potato (56) | Mung bean (4667) | Common bean (2065) | Tartary Buckwheat (117)

Foxtail millet (699) | Chickpea (1989) | Tomato (623) | Cucumber (775) | Potato (650) | Barley (227) | Grape (1587) | Apple (4898) | Muskmelon (1126) | Cattle (1651)

Chicken (3782) | Goat (971) | Pig (1767) | Sheep (800)

Show 10 entries

Search:

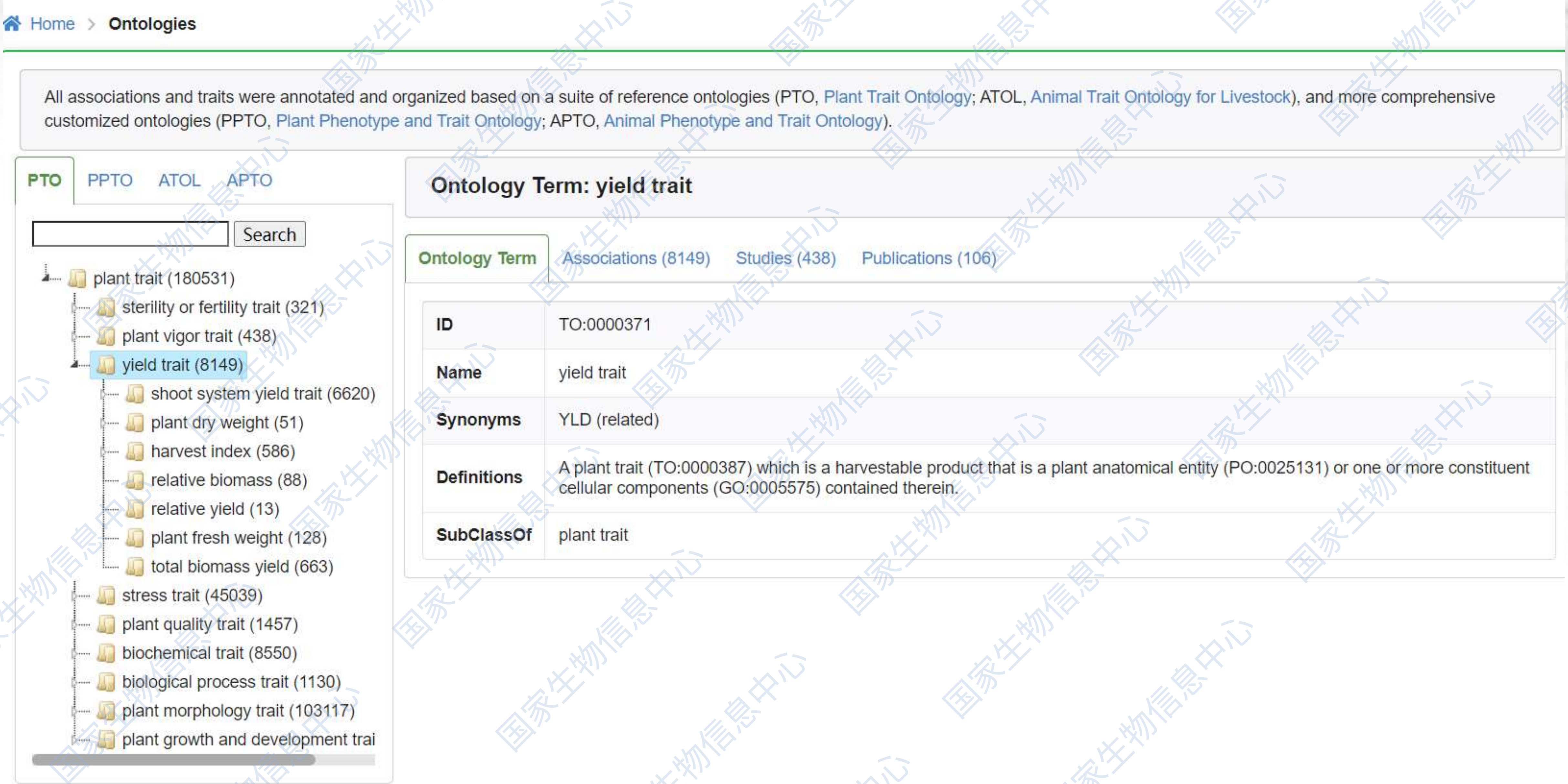
Variant ID	Genomic Location	Traits	*P*-values	R^2(%)	Genes (Consequence Types)	Lead SNPs	PMID
osa10000416	chr6:27887714	percentage of head rice recovery	1.94E-4	-	-	-	30512035
osa10001374	chr6:27906467	straighthead disorder severity rate	5.13E-8	-	Os06g0672400 (upstream_gene_variant)	-	35371188
osa10001374	chr6:27906467	straighthead disorder severity rate	1.45E-6	-	Os06g0672400 (upstream_gene_variant)	-	35371188
osa10001652	chr6:27910816	tiller number	3.38E-6	-	-	-	34804103
osa10001700	chr6:27911777	tiller number	5.34E-6	-	Os06g0672550 (downstream_gene_variant)	-	34804103
osa10001888	chr6:27914235	tiller number	4.95E-6	-	Os06g0672550 (downstream_gene_variant)	-	34804103
osa10002046	chr6:27916496	the xylem area of large vascular bundle	4.1E-13	5.07	Os06g0672550 (upstream_gene_variant)	Leader	33288021
osa10002046	chr6:27916496	the phloem area of large vascular bundle	1.2E-14	5.33	Os06g0672550 (upstream_gene_variant)	Leader	33288021

变异与表型因果关联知识

Show 10 entries Search:

Variant ID	Genomic Location	Effect Allele	Trait	Trait Impact	Allele Effect	Causal Type	Gene (name)	Consequences	PMID
osa9978294	chr6:27484286	T	grain number per plant	increasing grain number	superior	causal	Os06g0665400 (APO1)	upstream_gene_variant	20151298, 21119645
osa9978170	chr6:27481339	C	stem strength	increasing culm strength	superior	potential causal	-	-	21119645
osa9978136	chr6:27480778	C	lodging resistance	increasing lodging resistance	superior	potential causal	Os06g0665400 (APO1)	5_prime_UTR_variant	20151298, 21119645
osa9934720	chr6:26592357	T	average grain weight	increasing grain weight	superior	causal	Os06g0650300	3_prime_UTR_variant	25535376
osa9934720	chr6:26592357	T	plant height	increasing plant height	inferior	causal	Os06g0650300	3_prime_UTR_variant	25535376
osa9472656	chr6:17161791	T	blast disease resistance	increasing blast disease resistance	superior	potential causal	Os06g0494100 (PID2)	missense_variant	16709195
osa9235218	chr6:13057570	G	blast disease resistance	increasing blast disease resistance	superior	causal	Os06g0330100 (PID3)	5_prime_UTR_variant	19506306, 21621742, 23384860
osa9235061	chr6:13055819	G	blast disease resistance	increasing blast disease resistance	superior	causal	Os06g0330100 (PID3)	intron_variant	19506306, 21621742, 23384860
osa9087598	chr6:10389734	A	blast disease resistance	increasing blast disease resistance	superior	causal	Os06g0286700 (PIZ, PIZT)	missense_variant	16387888, 17073304, 26183036, 30967012
osa9037425	chr6:9338243	T	days to heading	early heading date under long day condition	superior	causal	Os06g0275000 (SE1)	stop_gained	11148291, 19246394

性状本体



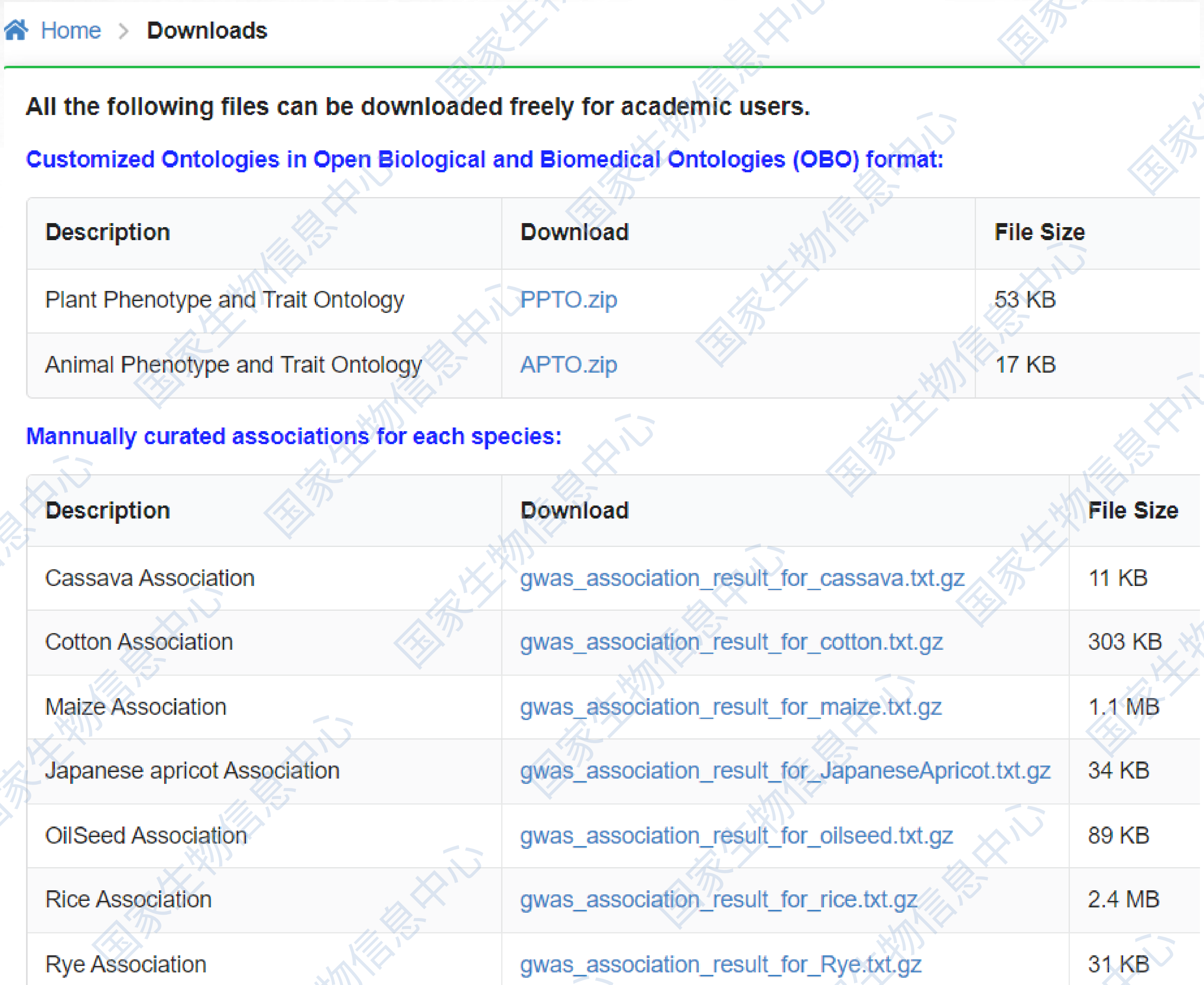
All associations and traits were annotated and organized based on a suite of reference ontologies (PTO, Plant Trait Ontology; ATOL, Animal Trait Ontology for Livestock), and more comprehensive customized ontologies (PPTO, Plant Phenotype and Trait Ontology; APTO, Animal Phenotype and Trait Ontology).

- plant trait (180531)
  - sterility or fertility trait (321)
  - plant vigor trait (438)
  - yield trait (8149)
    - shoot system yield trait (6620)
    - plant dry weight (51)
    - harvest index (586)
    - relative biomass (88)
    - relative yield (13)
    - plant fresh weight (128)
    - total biomass yield (663)
  - stress trait (45039)
  - plant quality trait (1457)
  - biochemical trait (8550)
  - biological process trait (1130)
  - plant morphology trait (103117)
  - plant growth and development trai

Ontology Term: yield trait

ID	TO:0000371
Name	yield trait
Synonyms	YLD (related)
Definitions	A plant trait (TO:0000387) which is a harvestable product that is a plant anatomical entity (PO:0025131) or one or more constituent cellular components (GO:0005575) contained therein.
SubClassOf	plant trait

数据共享



Home > Downloads

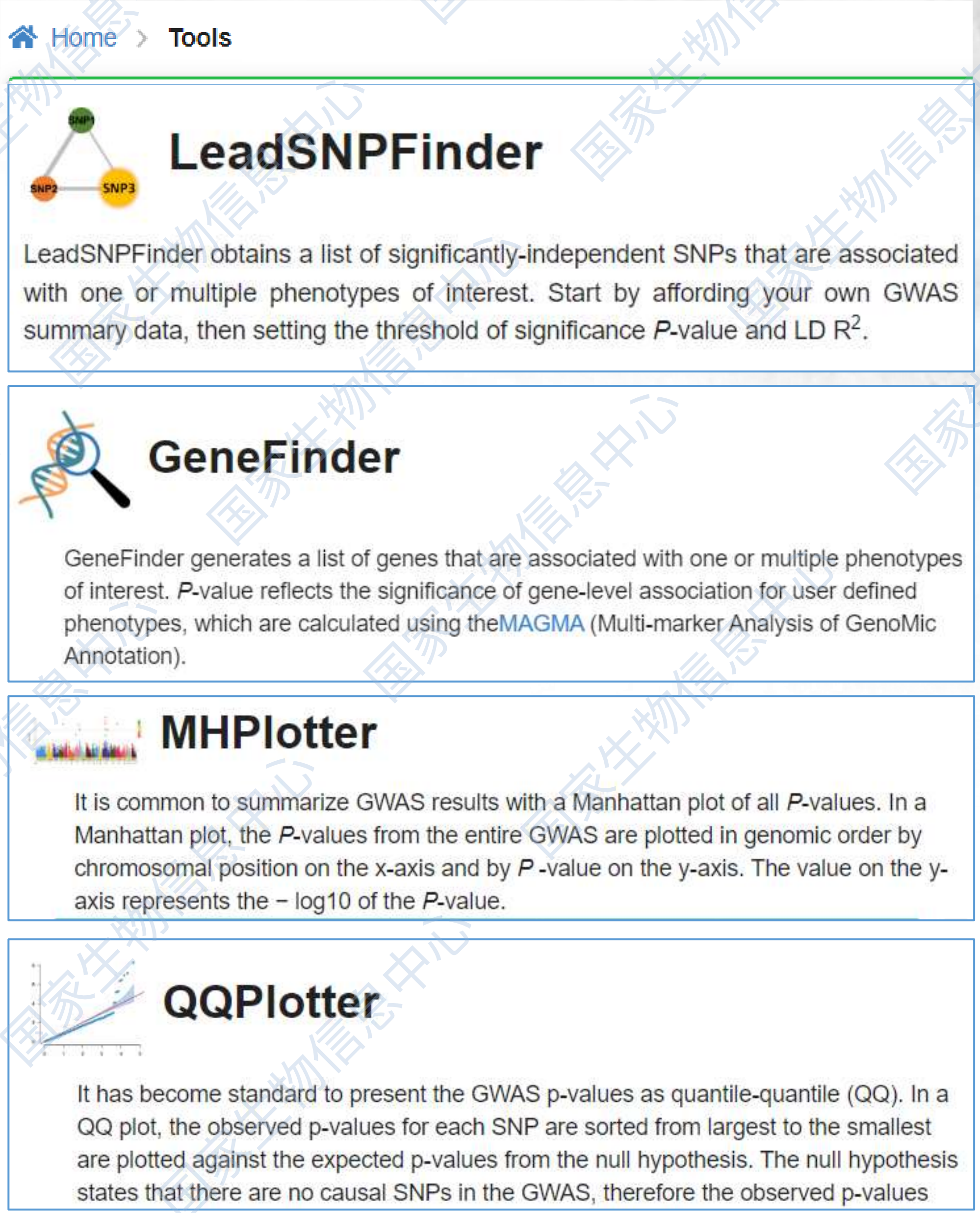
All the following files can be downloaded freely for academic users.

Customized Ontologies in Open Biological and Biomedical Ontologies (OBO) format:

Description	Download	File Size
Plant Phenotype and Trait Ontology	PPTO.zip	53 KB
Animal Phenotype and Trait Ontology	APTO.zip	17 KB

Mannually curated associations for each species:

Description	Download	File Size
Cassava Association	gwas_association_result_for_cassava.txt.gz	11 KB
Cotton Association	gwas_association_result_for_cotton.txt.gz	303 KB
Maize Association	gwas_association_result_for_maize.txt.gz	1.1 MB
Japanese apricot Association	gwas_association_result_for_JapaneseApricot.txt.gz	34 KB
OilSeed Association	gwas_association_result_for_oilseed.txt.gz	89 KB
Rice Association	gwas_association_result_for_rice.txt.gz	2.4 MB
Rye Association	gwas_association_result_for_Rye.txt.gz	31 KB



Home > Tools

LeadSNPFinder

LeadSNPFinder obtains a list of significantly-independent SNPs that are associated with one or multiple phenotypes of interest. Start by affording your own GWAS summary data, then setting the threshold of significance *P*-value and LD R^2.

GeneFinder

GeneFinder generates a list of genes that are associated with one or multiple phenotypes of interest. *P*-value reflects the significance of gene-level association for user defined phenotypes, which are calculated using theMAGMA (Multi-marker Analysis of GenoMic Annotation).

MHPlotter

It is common to summarize GWAS results with a Manhattan plot of all *P*-values. In a Manhattan plot, the *P*-values from the entire GWAS are plotted in genomic order by chromosomal position on the x-axis and by *P* -value on the y-axis. The value on the y-axis represents the − log10 of the *P*-value.

QQPlotter

It has become standard to present the GWAS p-values as quantile-quantile (QQ). In a QQ plot, the observed p-values for each SNP are sorted from largest to the smallest are plotted against the expected p-values from the null hypothesis. The null hypothesis states that there are no causal SNPs in the GWAS, therefore the observed p-values

提纲

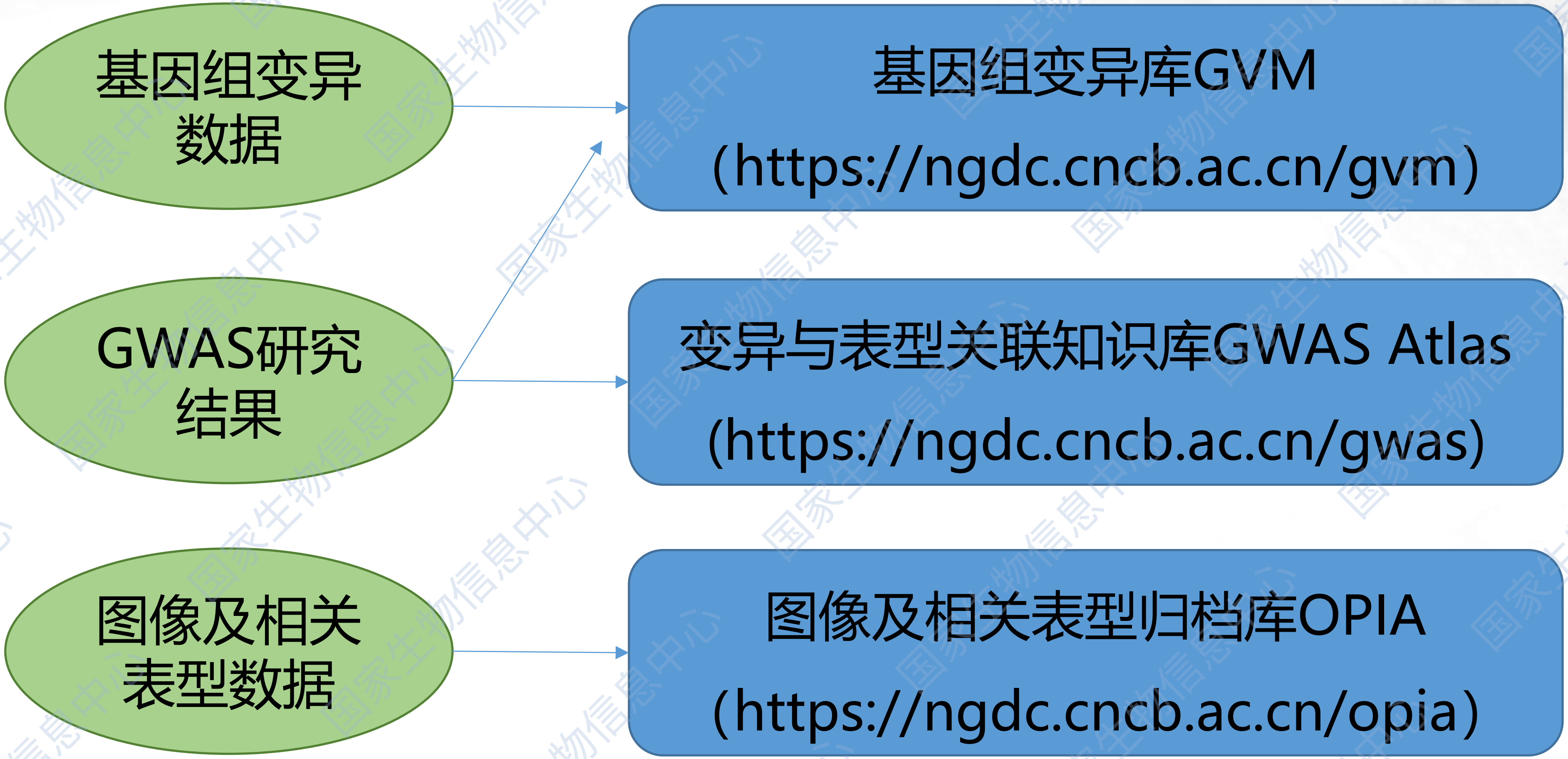
01 | 基因组序列变异数据库GVM

02 | 植物图像及其表型归档库OPIA

03 | G2P关联知识库GWAS Atlas

04 | 用户变异与表型数据汇交与共享

数据汇交类型



基因组变异数据

GWAS研究结果

图像及相关表型数据

基因组变异库GVM
(https://ngdc.cncb.ac.cn/gvm)

变异与表型关联知识库GWAS Atlas
(https://ngdc.cncb.ac.cn/gwas)

图像及相关表型归档库OPIA
(https://ngdc.cncb.ac.cn/opia)

数据提交入口

- 入口URL: https://ngdc.cncb.ac.cn/gsub

生物数据递交系统(BIG Submission, BIG Sub)是国家基因组科学数据中心生物数据统一汇交入口，为用户提供一站式数据递交服务。

BioProject
收集与共享生物学研究项目信息的资源库

BioSample
收集与共享实验相关的生物样本信息的资源库

Genome Sequence Archive
组学原始数据汇交、存储、管理与共享系统

GSA for Human
人类遗传资源组学原始数据库

Genome Warehouse
多物种基因组组装序列和注释数据归档库

GenBase
核酸和蛋白序列及注释数据归档库

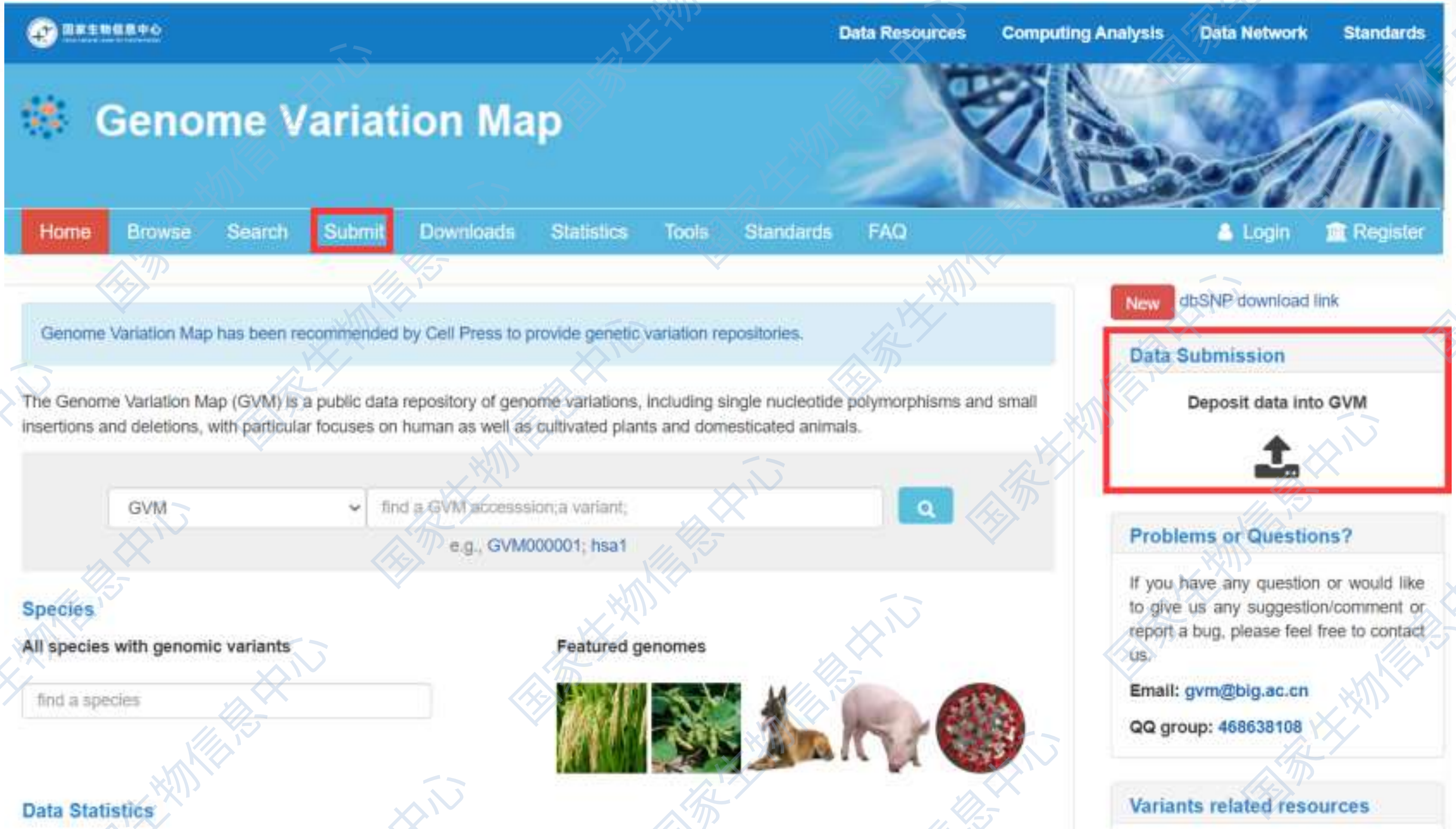
OMIX
收集和共享表达谱、甲基化谱、蛋白组学、代谢组学等数据的资源库

Genome Variation Map
基因组序列变异信息汇交、管理与检索的资源库

BioCode
整合开源生物信息软件工具的数据库

数据提交入口

- **入口URL:** https://ngdc.cncb.ac.cn/gvm



Genome Variation Map has been recommended by Cell Press to provide genetic variation repositories.

The Genome Variation Map (GVM) is a public data repository of genome variations, including single nucleotide polymorphisms and small insertions and deletions, with particular focuses on human as well as cultivated plants and domesticated animals.

e.g., GVM000001; hsa1

Species

All species with genomic variants

Featured genomes

Data Statistics

Data Submission

Deposit data into GVM

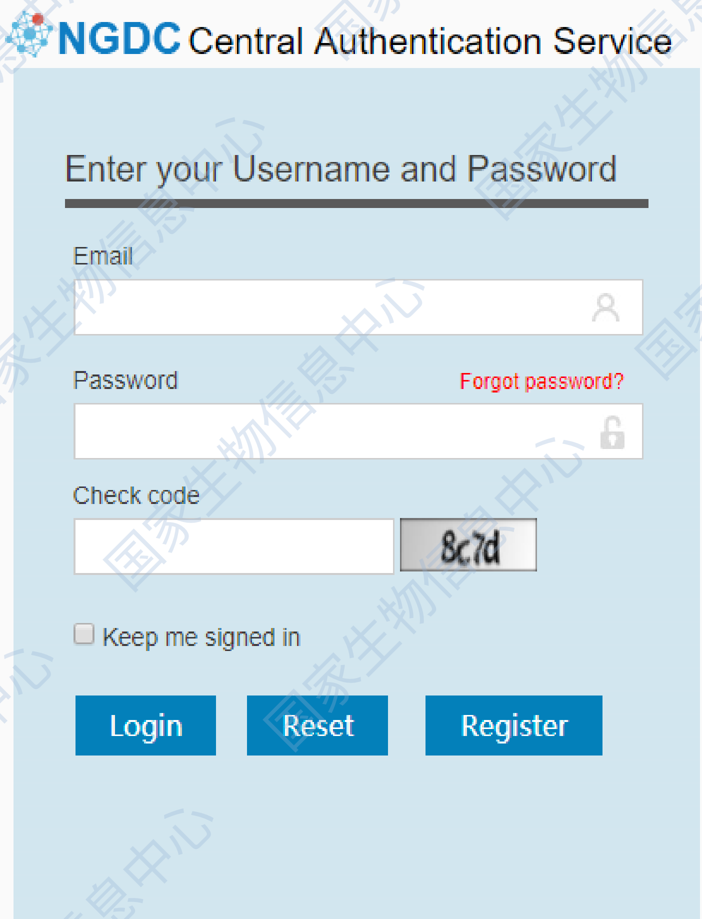
Problems or Questions?

If you have any question or would like to give us any suggestion/comment or report a bug, please feel free to contact us.

Email: gvm@big.ac.cn

QQ group: 468638108

Variants related resources

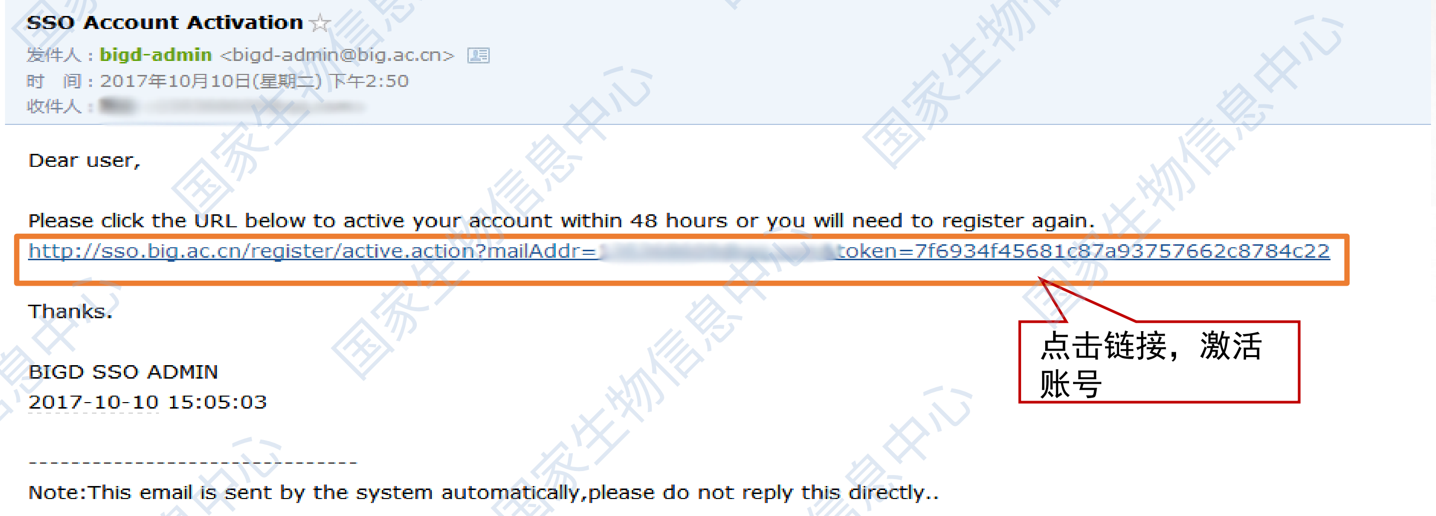


账号注册

注意：由于需要通过邮件接收激活链接，请确保该邮箱属于自己并能够登录进去点击激活链接！否则账号无法激活和登录！同时请合理设置垃圾邮件判断策略，检查激活邮件是否被邮件系统转移到垃圾邮箱中，或者被邮件系统拒收！

注意：密码必须同时包含大写字母、小写字母、数字、特殊字符，且长度在8~30位之间！

激活邮件



SSO Account Activation ☆

发件人：**bigd-admin** <bigd-admin@big.ac.cn>

时　间：2017年10月10日(星期二) 下午2:50

收件人：[redacted]

Dear user,

Please click the URL below to active your account within 48 hours or you will need to register again.

http://sso.big.ac.cn/register/active.action?mailAddr=[redacted]token=7f6934f45681c87a93757662c8784c22

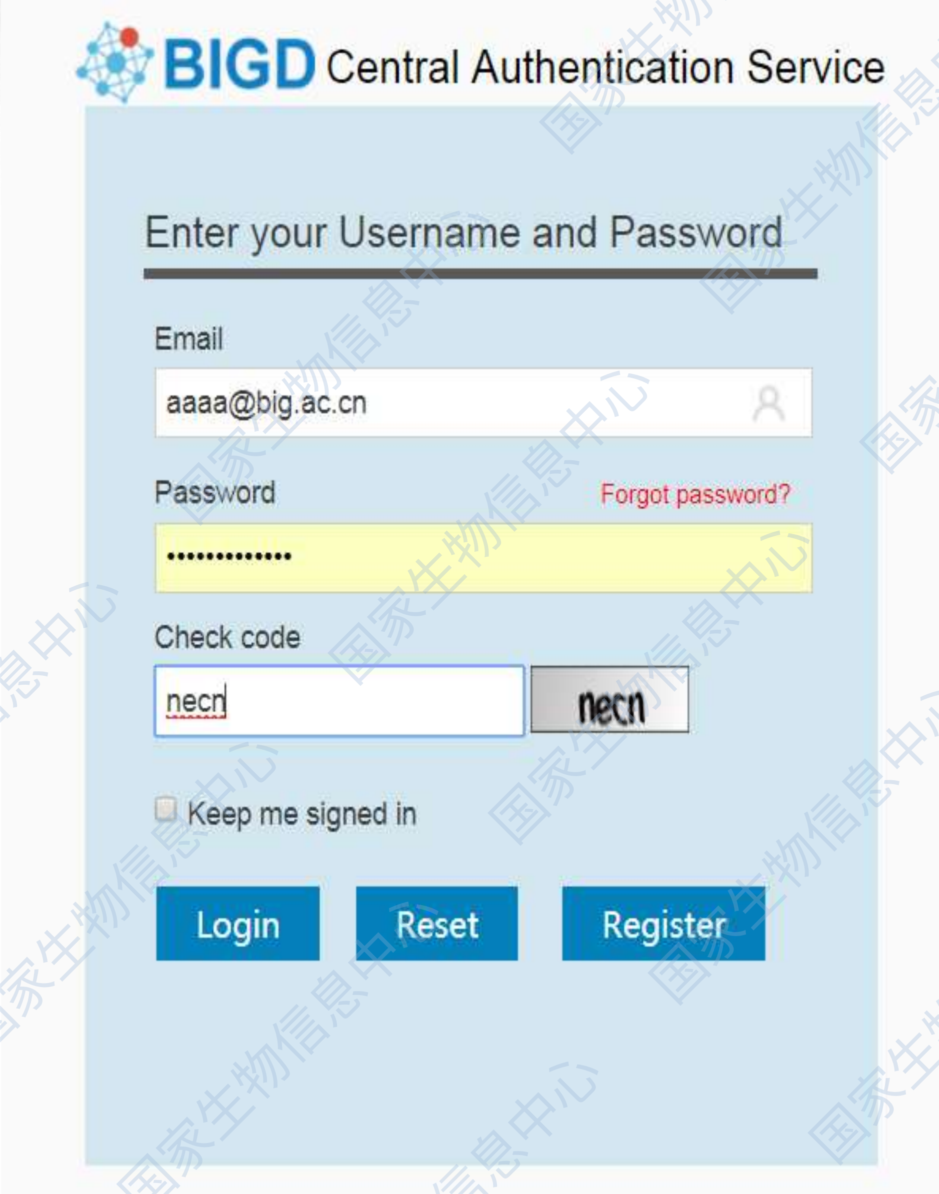
Thanks.

BIGD SSO ADMIN
2017-10-10 15:05:03

Note:This email is sent by the system automatically,please do not reply this directly..

点击链接，激活账号

账号登录


BIGD Central Authentication Service

Enter your Username and Password

Email: aaaa@big.ac.cn

Password: ••••••••••••••

Check code: necn

GVM / Submit / My Submission

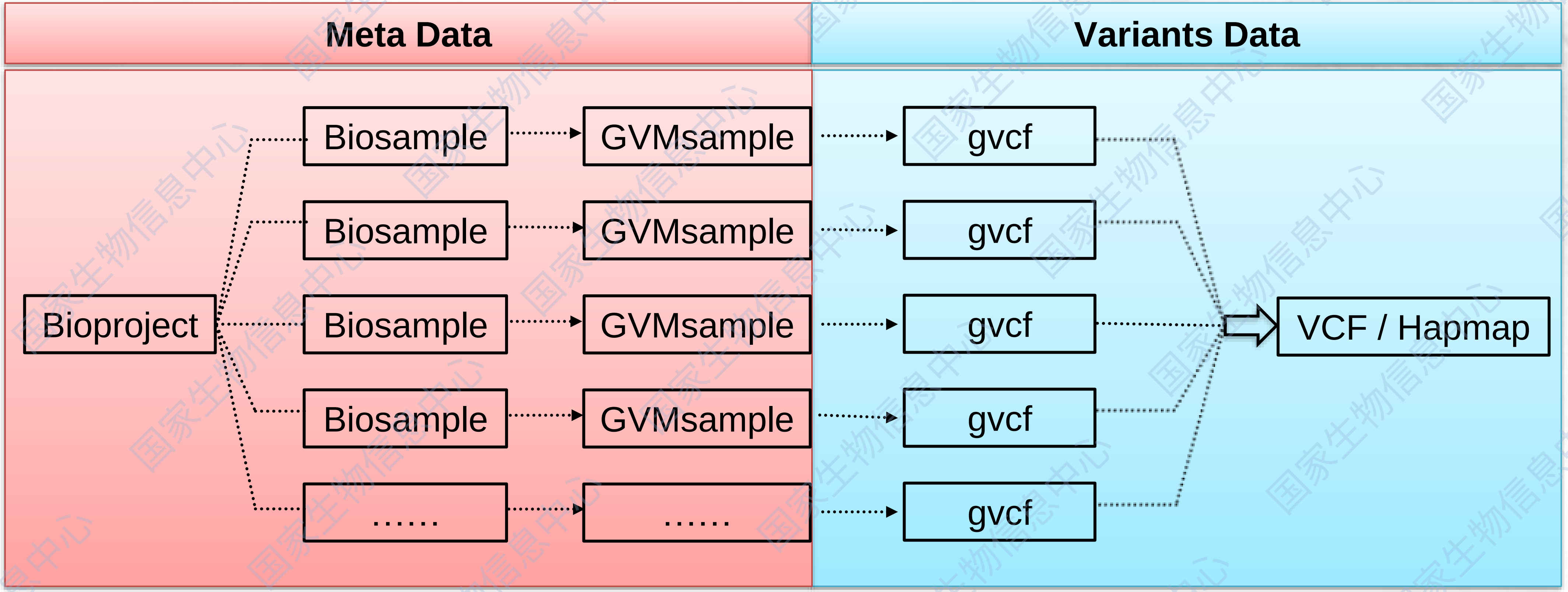
GVM is a repository specialized for archiving genome variantion data (VCF, GVCF or HapMap format). Creating a new G

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 English Chinese

数据结构



- □ 项目信息
- □ 样本采集信息
- □ 样本变异分析参数等信息

在线填写

- □ 实体数据文件
- □ 支持VCF、GVCF、Hapmap等数据格式

FTP上传

四个环节

1. 账号注册/登录
2. 元信息在线填写
3. 实体数据上传
4. 数据审核与引用

在线填写汇交申请

GVM / Submit / My Submission

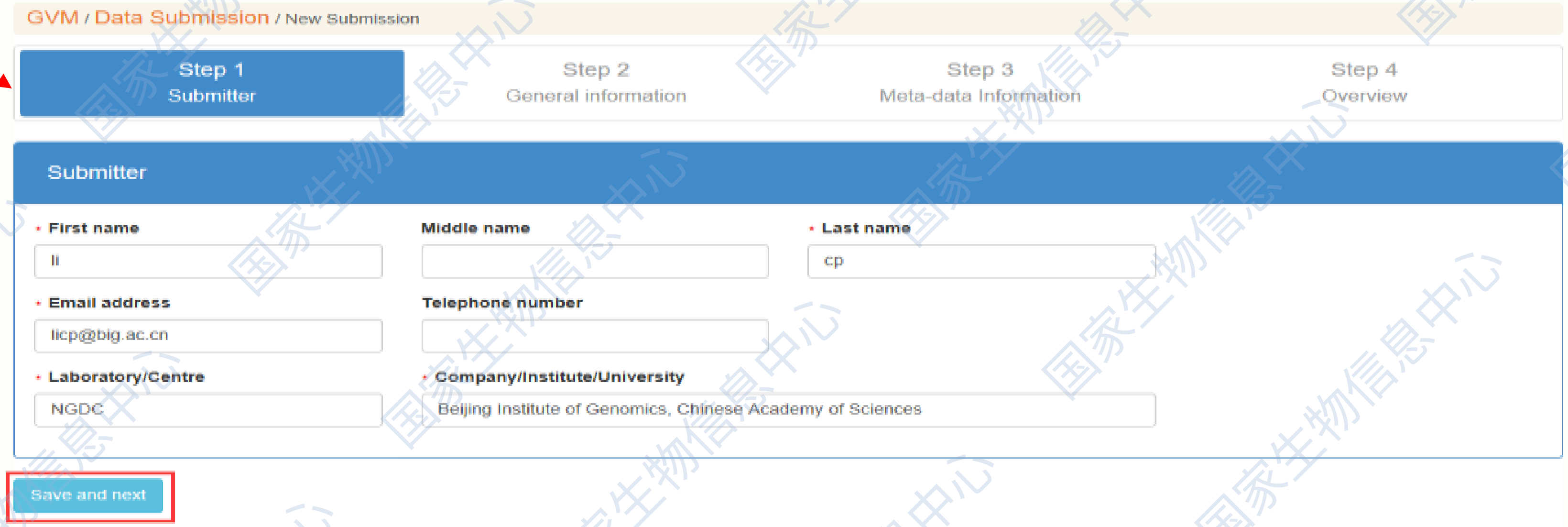
GVM is a repository specialized for archiving genome variantion data (VCF, GVCF or HapMap format). Creating a new GVM submission, three steps are needed:

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 English Chinese

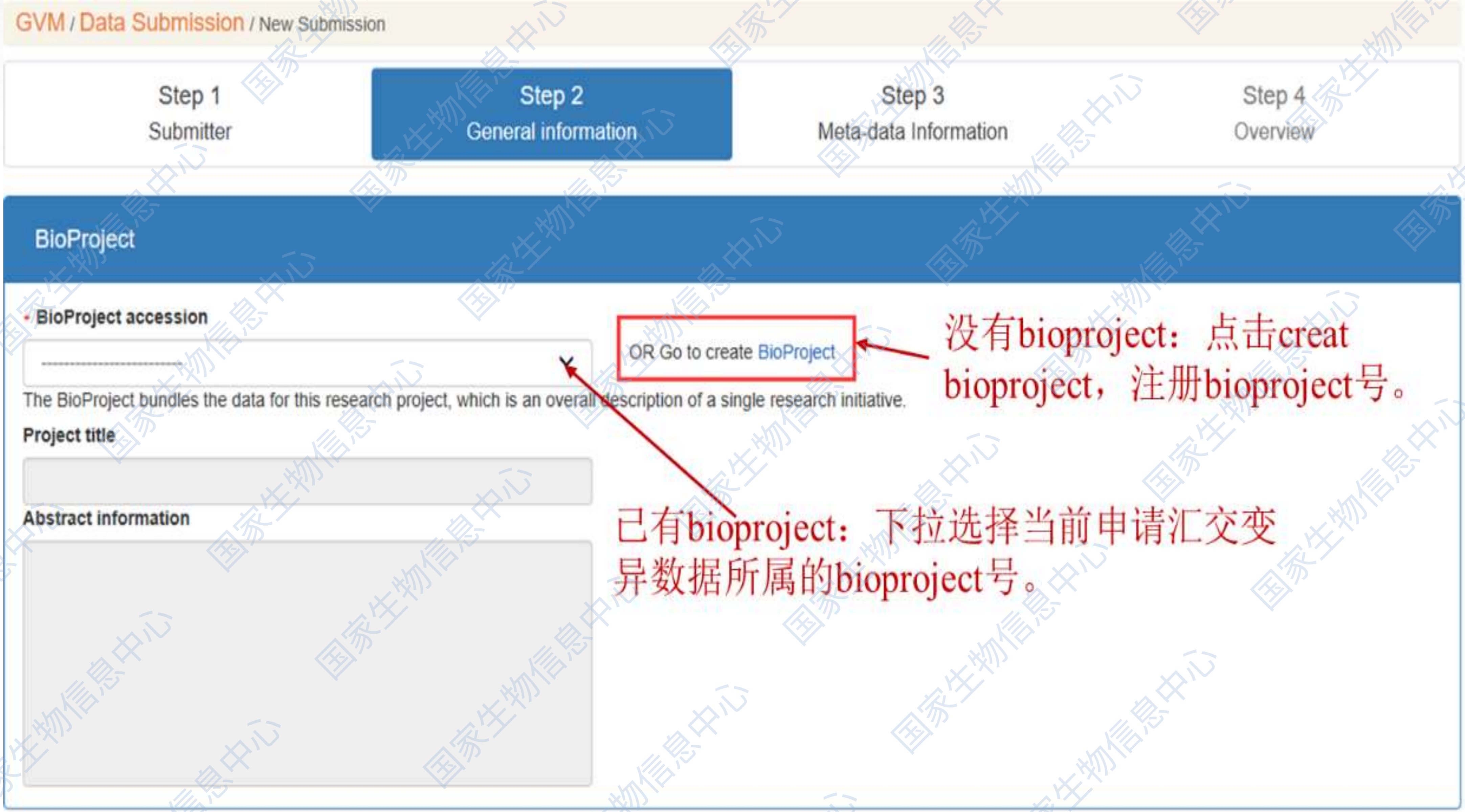
中英文帮助文档，仔细阅读，减少错误率！


GVM / Data Submission / New Submission

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	

项目信息填写



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

没有bioproject：点击creat bioproject，注册bioproject号。

已有bioproject：下拉选择当前申请汇交变异数据所属的bioproject号。

项目信息填写

提交者信息

* 名	中间名	* 姓氏
Dongmei	middle name	Tian
* 邮箱	备用邮箱	
tiandm@big.ac.cn	secondary email	
* 单位	单位网址	* 部门
Beijing Institute of Genomics, Chi	http://www.big.ac.cn	Big Data Center
手机	传真	
* 街道	* 城市	州/省
NO.1 Beichen West Road, Chaoy	Beijing	beijing

样本信息填写

★BioSample

● **No GVM related BioSample information was created** ← 没有BioSample

○ **GVM related BioSample information has been created** ← 已有BioSample

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

选择数据释放日期，立即释放 / 择期释放

★Abstract for the variation data

变异信息描述

★Short title for the variation data (less than 150 characters)

简短标题

样本信息填写

提交者信息

* 名: Dongmei	中间名: middle name	* 姓氏: Tian
* 邮箱: tiandm@big.ac.cn	备用邮箱: secondary email	
* 单位: Beijing Institute of Genomics, Chi	单位网址: http://www.big.ac.cn	* 部门: Big Data Center
手机:	传真:	
* 街道: NO.1 Beichen West Road, Chaoy	* 城市: Beijing	州/省: beijing

元信息填写

GVM / Data Submission / New Submission

GVM Meta-data Information

BioProject: PRJCA003421

* **Please upload GVM template file**

Download the GVM Submission Template

下载离线模板

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 menues 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
ID	*sample_name	project_accession	geogra
S1	zhang1	PRJCA003421	Ch
S2	li2	PRJCA003421	Ch
S3	wang3	PRJCA003421	Ch

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

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

1	2	3	4
ID	strategy	platform	array_nan
S1	WGS	Illumina HiSeq 2000	
S2	WGS	Illumina HiSeq 2000	
S3	WGS	Illumina HiSeq 2000	

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 menues that allow you to select from a controlled vocabulary. Once specified for one row, these values can

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

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	a1111111111111111111111111111111
2	S1.g.vcf.gz	Variants file of sample S1.	gvcf	a3333333333333333333333333333333
3	Human.variants.hapmap	Variants file of 1000 Chinese in hapmap format.	hapmap	a4444444444444444444444444444444

元信息上传

国家基因组科学数据中心
National Genomics Data Center

- Please upload GVM template file

Select file...

选择已填好的离线模板

Browse ...

Download the GVM Submission Template

- Please upload GVM template file

GVM Submission Template V3.0.xlsx

上传已填好的离线模板

Remove

Upload

Browse ...

Download the GVM Submission Template

- Please upload GVM template file

GVM_Submission_Template_V3.0.xlsx

Check

Remove

校验元信息格式及内容

- Please upload GVM template file

GVM_Submission_Template_V3.0.xlsx

Check

Remove

Validate Success !

Save and next

下一步

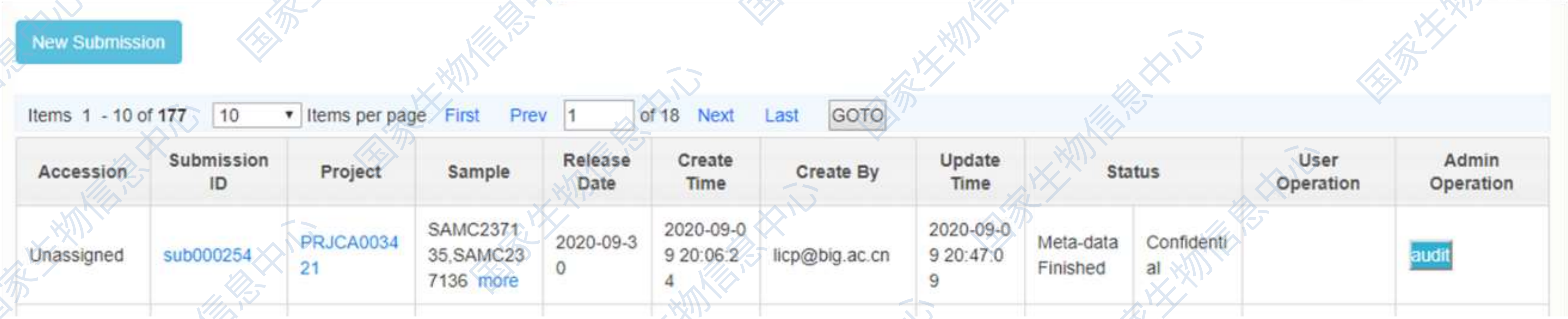
元信息提交

You now at: Submit > Submission

Overview

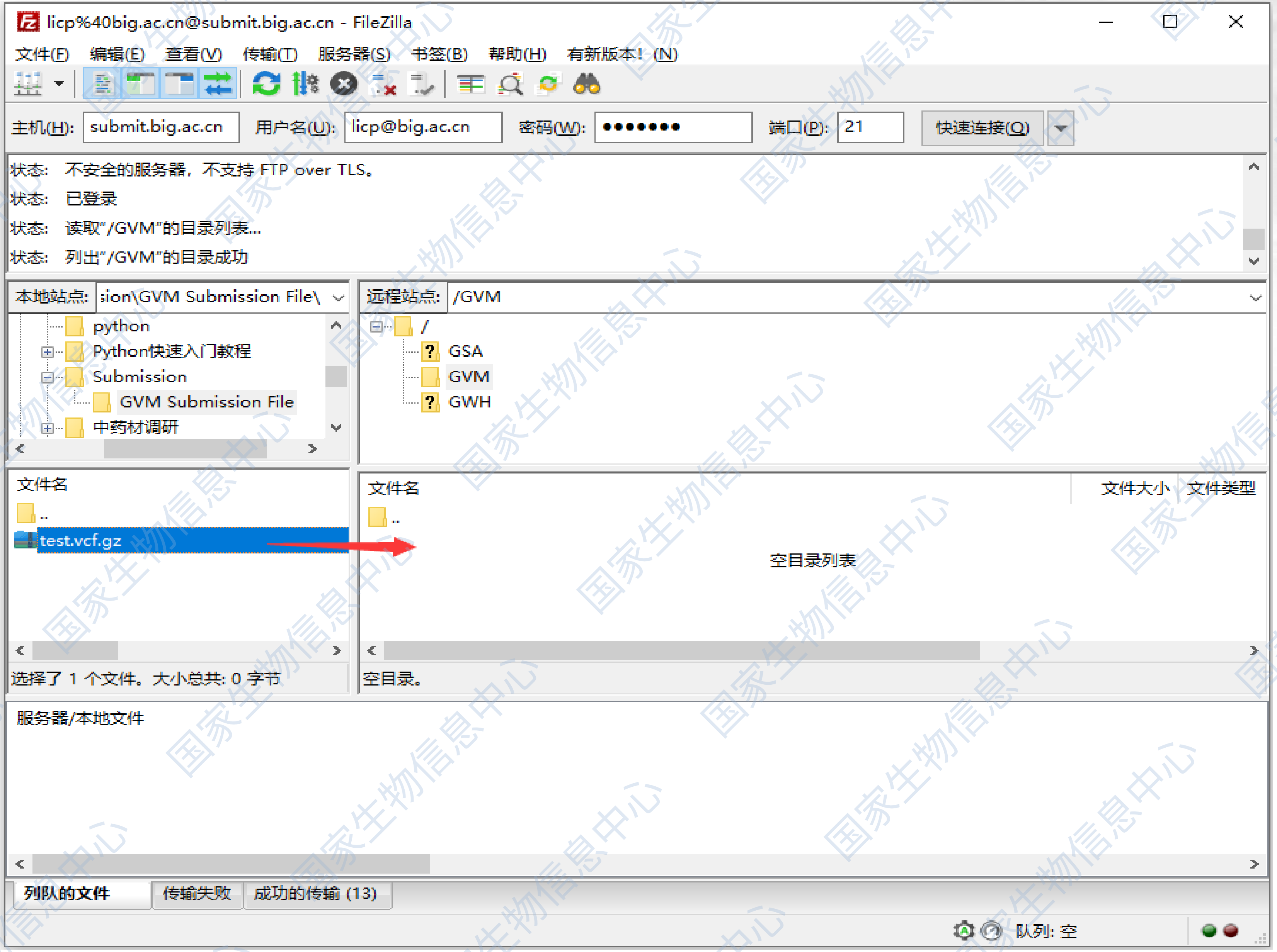
Please transmitting your data files to the GVM FTP site (download FileZilla client)
Address: ftp://submit.big.ac.cn
User: Same as you login the GVM
Password: Same as you login the GVM

点击链接，下载文件上传工具 FileZilla
https://filezilla-project.org/download.php?type=client



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

变异数据上传



主机(H): submit.big.ac.cn 用户名(U): licp@big.ac.cn 密码(W): ●●●●●●● 端口(P): 21

```
状态:	不安全的服务器，不支持 FTP over TLS。
状态:	已登录
状态:	读取"/GVM"的目录列表...
状态:	列出"/GVM"的目录成功
```

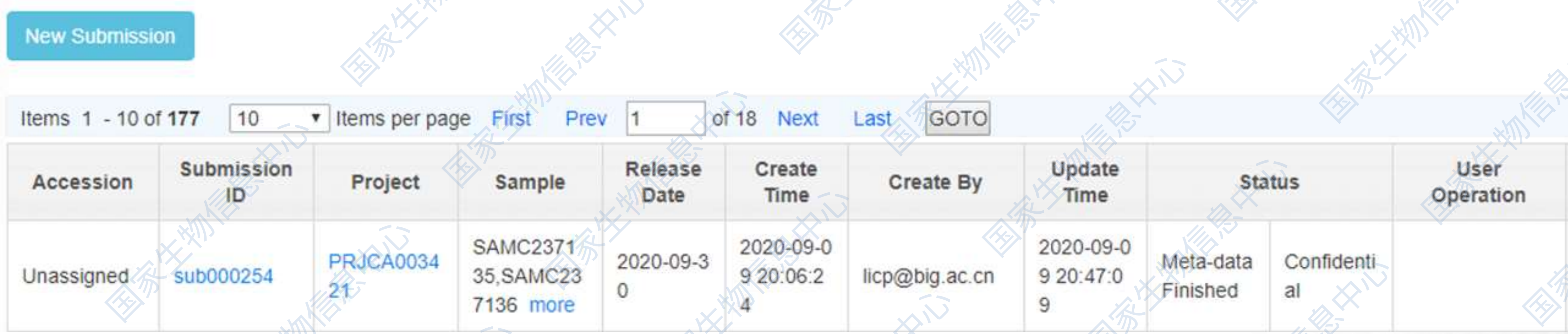
本地站点: sion\GVM Submission File\

远程站点: /GVM

选择了 1 个文件。大小总共: 0 字节

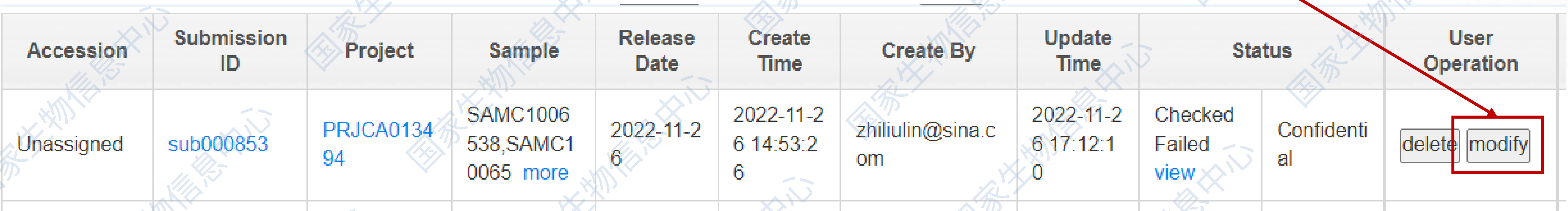
空目录。

等待审核



Accession	Submission ID	Project	Sample	Release Date	Create Time	Create By	Update Time	Status		User 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	

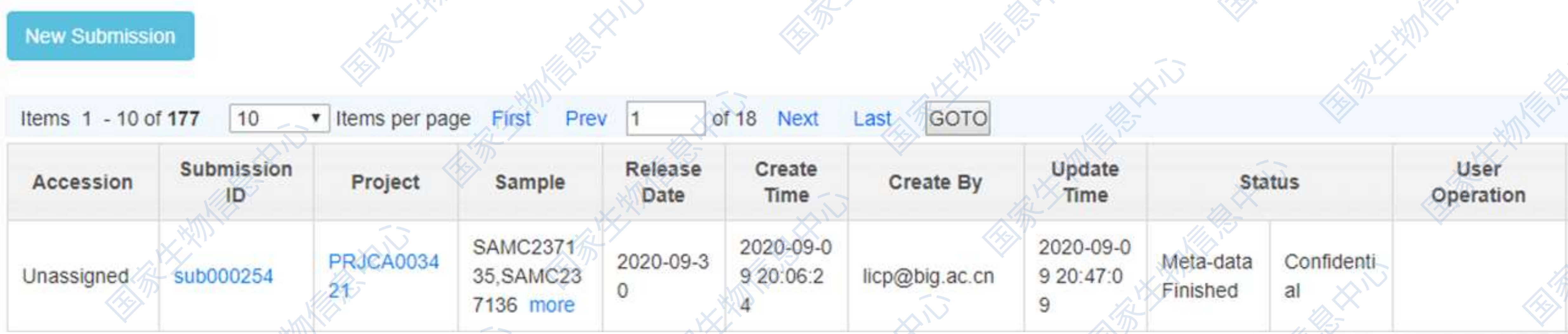
审核未通过，修改完善



Accession	Submission ID	Project	Sample	Release Date	Create Time	Create By	Update Time	Status		User Operation
Unassigned	sub000853	PRJCA013494	SAMC1006538,SAMC10065 more	2022-11-26	2022-11-26 14:53:26	zhiliulin@sina.com	2022-11-26 17:12:10	Checked Failed view	Confidential	delete modify

等待审核

国家基因组科学数据中心
National Genomics Data Center



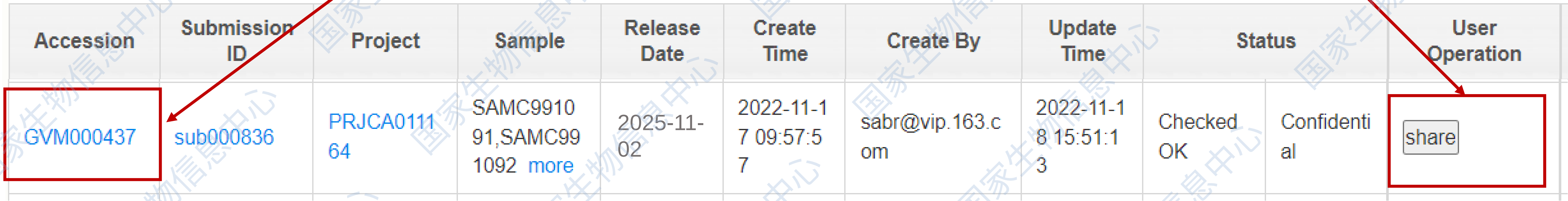
New Submission

Items 1 - 10 of 177 10 ▾ Items per page First Prev 1 of 18 Next Last GOTO

Accession	Submission ID	Project	Sample	Release Date	Create Time	Create By	Update Time	Status		User 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	

审核通过，获得编号

未公开数据，共享链接



Accession	Submission ID	Project	Sample	Release Date	Create Time	Create By	Update Time	Status		User Operation
GVM000437	sub000836	PRJCA011164	SAMC991091,SAMC991092 more	2025-11-02	2022-11-17 09:57:57	sabr@vip.163.com	2022-11-18 15:51:13	Checked OK	Confidential	share

数据库引用

- **引用示例：**

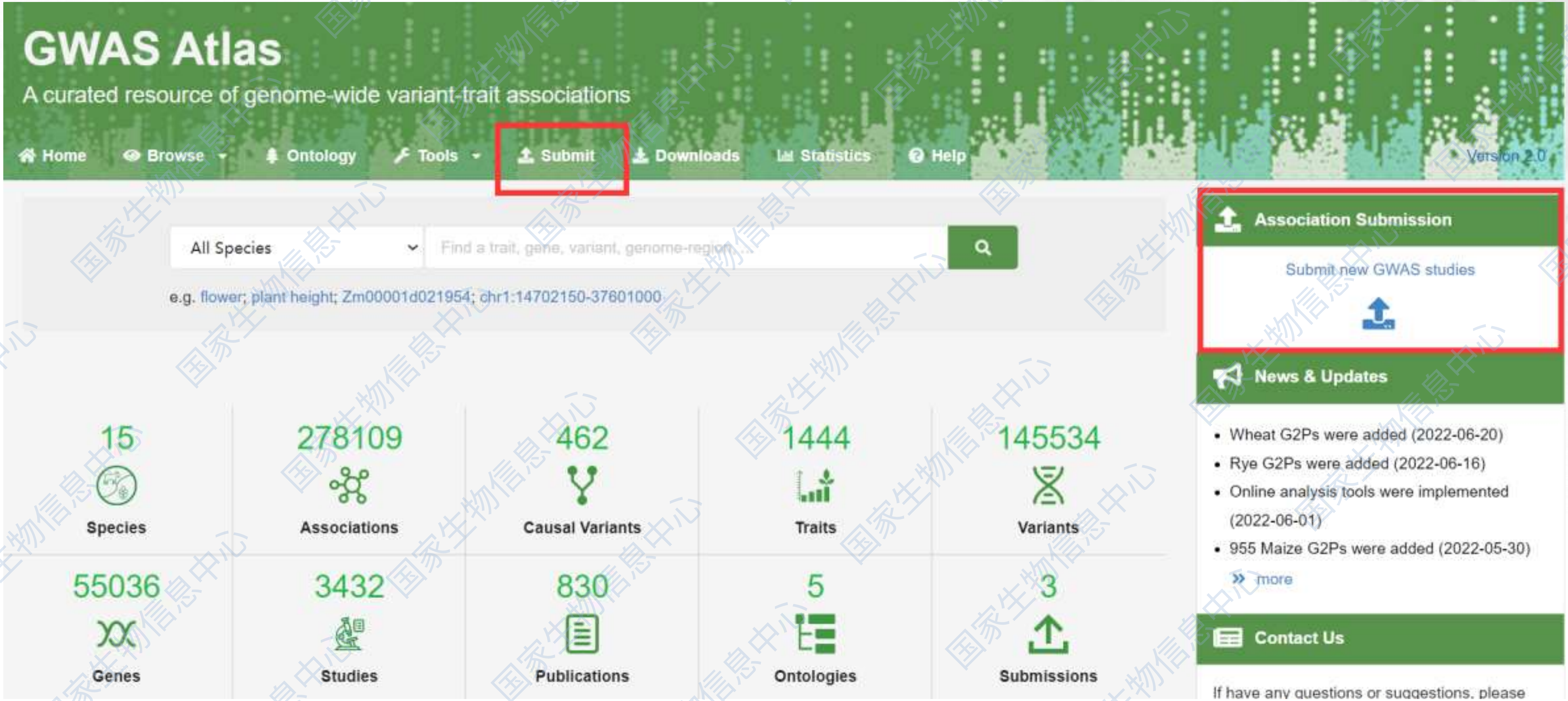
The variation data reported in this paper have been deposited in the Genome Variation Map (GVM) [1] in National Genomics Data Center, Beijing Institute of Genomics, Chinese Academy of Sciences and China National Center for Bioinformation, under accession number GVMXXXXXX.

- **References:**

[1] Genome Variation Map: a worldwide collection of genome variations across multiple species. *Nucleic Acids Res* 2021, 49(D1):D1186-D1191. [PMID=33170268].

GWAS 数据递交

- **入口URL:** https://ngdc.cncb.ac.cn/gwas



GWAS Atlas

A curated resource of genome-wide variant-trait associations

e.g. flower; plant height; Zm00001d021954; chr1:14702150-37601000

15	278109	462	1444	145534
Species	Associations	Causal Variants	Traits	Variants
55036	3432	830	5	3
Genes	Studies	Publications	Ontologies	Submissions

Association Submission

Submit new GWAS studies

News & Updates

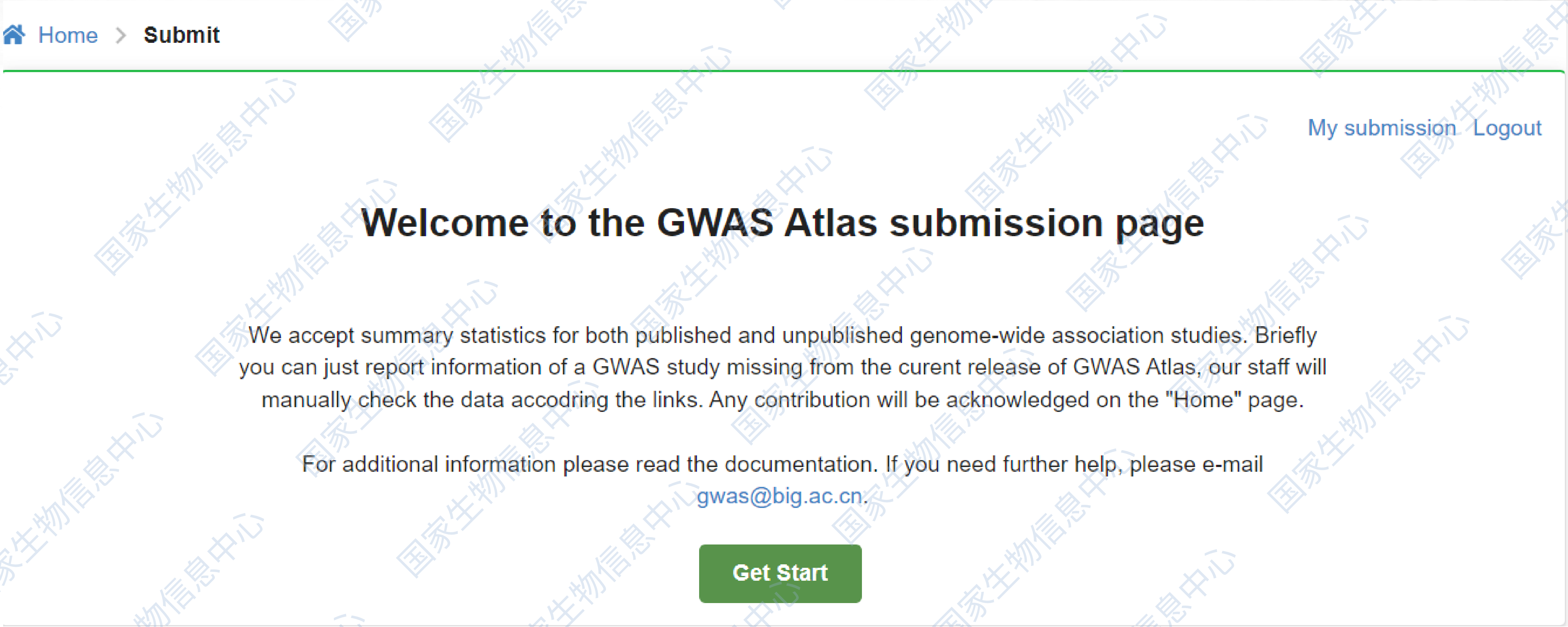
- Wheat G2Ps were added (2022-06-20)
- Rye G2Ps were added (2022-06-16)
- Online analysis tools were implemented (2022-06-01)
- 955 Maize G2Ps were added (2022-05-30)

» more

Contact Us

If have any questions or suggestions, please

GWAS 数据递交



Welcome to the GWAS Atlas submission page

We accept summary statistics for both published and unpublished genome-wide association studies. Briefly you can just report information of a GWAS study missing from the curent release of GWAS Atlas, our staff will manually check the data accodring the links. Any contribution will be acknowledged on the "Home" page.

For additional information please read the documentation. If you need further help, please e-mail gwas@big.ac.cn.

GWAS 数据递交

New Submission

Publications

Is it a published study? *
Select

Publication title *
Title

GWAS Study

Upload GWAS meta data
Download meta data template study_template.xlsx
选择文件 未选择任何文件

Upload GWAS summary files (multiple file please compressed in .gz)
选择文件 未选择任何文件

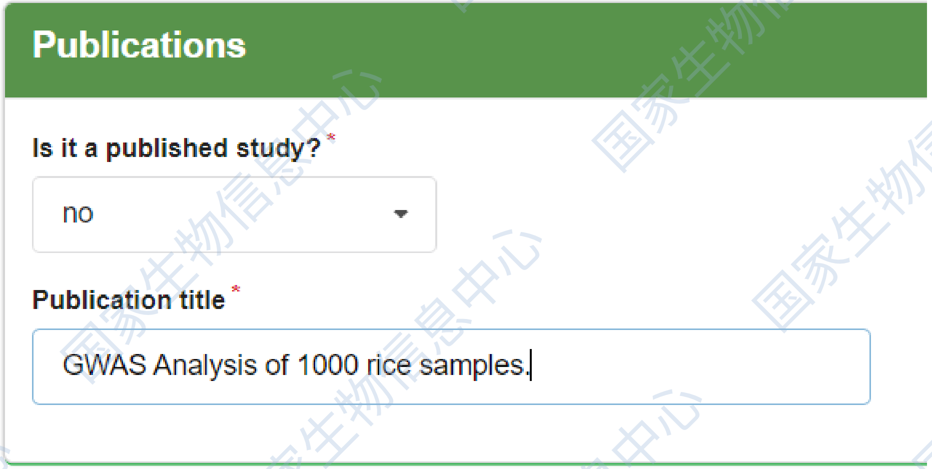
Release date
● Release immediately (recommended)
○ Release on specified date
2022-11-29

Publications
信息填写

Study信息及
文件上传

预览&提交

GWAS 数据提交



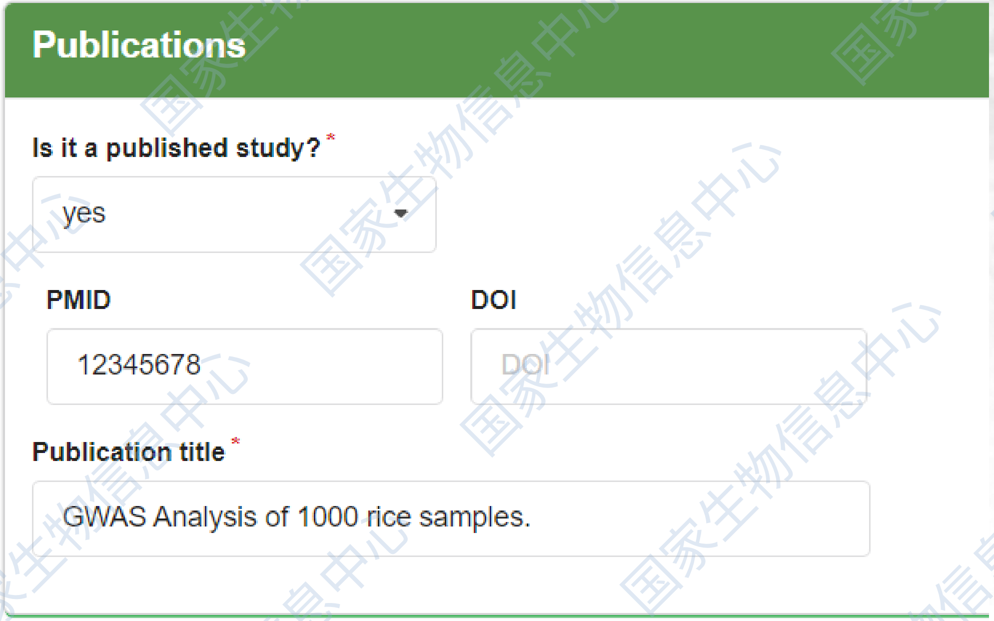
Publications

Is it a published study? *

no

Publication title *

GWAS Analysis of 1000 rice samples.



Publications

Is it a published study? *

yes

PMID: 12345678

DOI: DOI

Publication title *

GWAS Analysis of 1000 rice samples.

研究内容没有发表：

Is or not: No；

Publication title: 简短题目。

研究内容已发表：

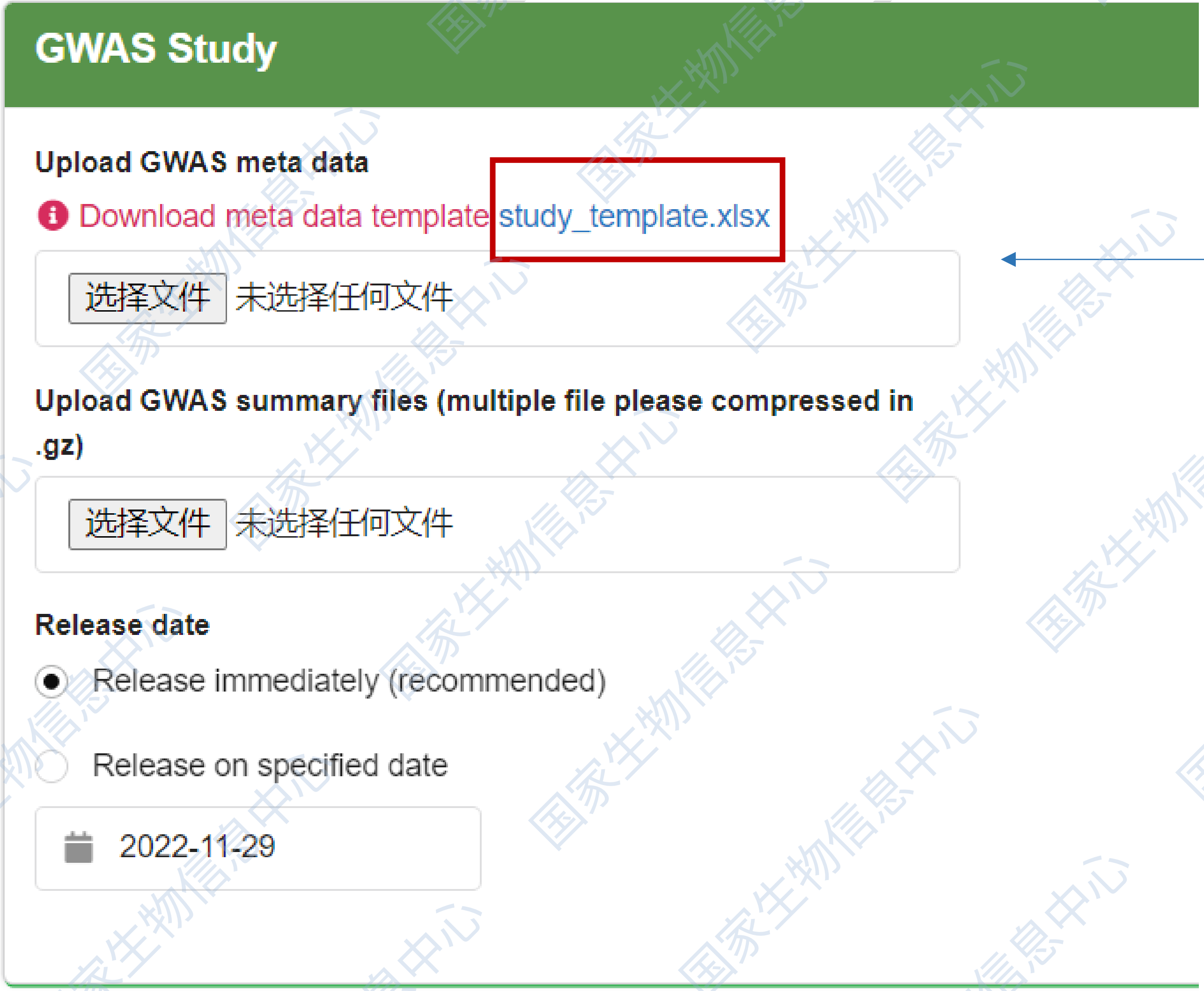
Is or not: Yes；

PMID 或者DOI必选其一填写；

Publication title: 简短题目。

GWAS 数据提交

Publications → **GWAS Study** → Submit



GWAS Study

Upload GWAS meta data

Download meta data template study_template.xlsx

Upload GWAS summary files (multiple file please compressed in .gz)

Release date

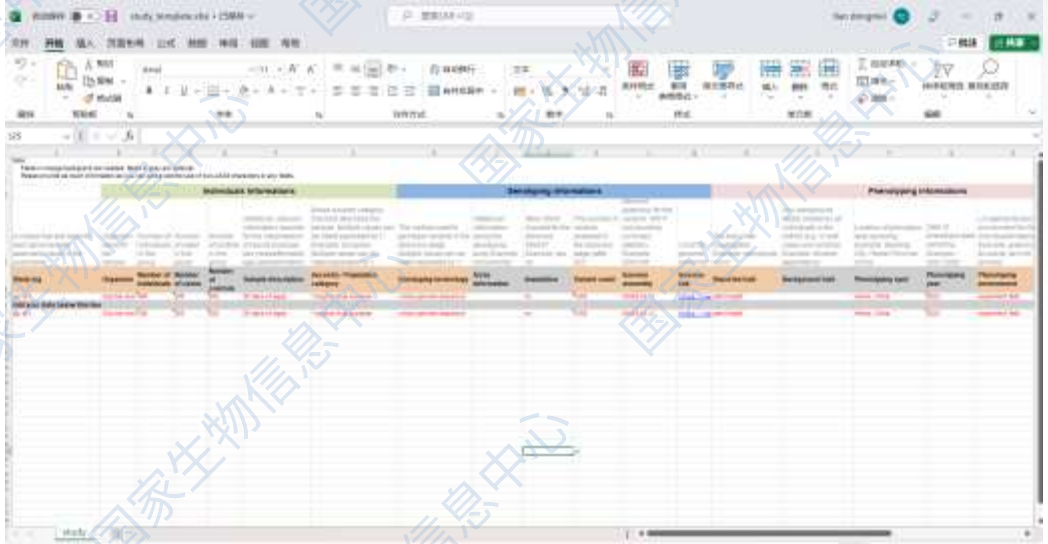
● Release immediately (recommended)

○ Release on specified date

2022-11-29

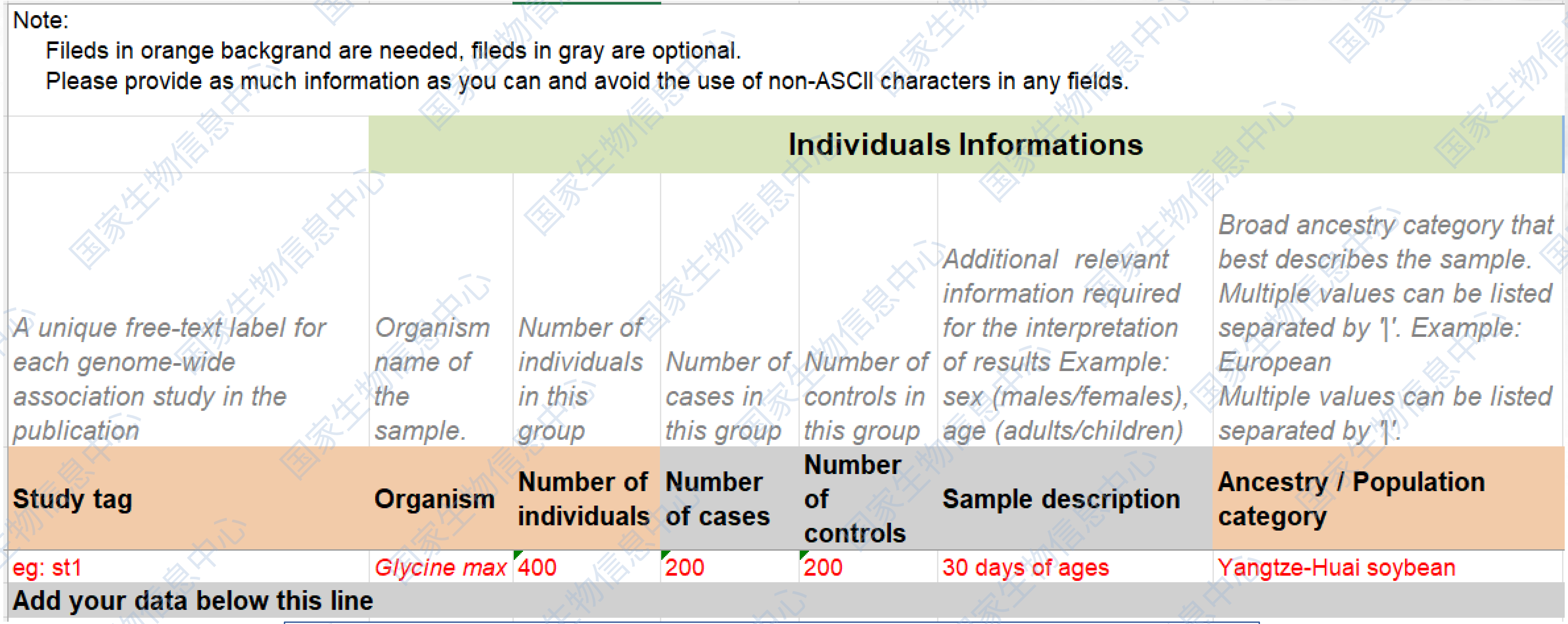
GWAS元信息填写并上传

下载模板，填写GWAS研究相关信息，注意*号内容必填项，尽可能信息完整。



GWAS 数据提交

Publications → **GWAS Study** → Submit



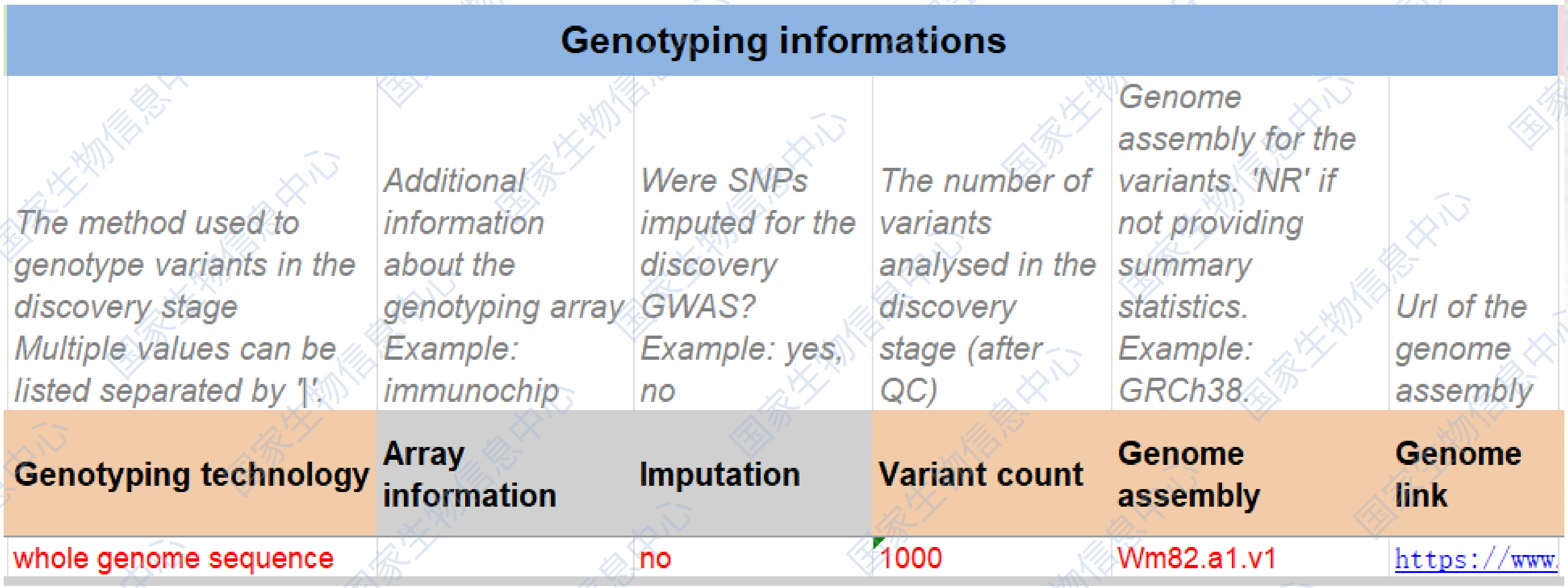
Note:
Fileds in orange backgrand are needed, fileds in gray are optional.
Please provide as much information as you can and avoid the use of non-ASCII characters in any fields.

	Individuals Informations					
A unique free-text label for each genome-wide association study in the publication	*Organism name of the sample.*	*Number of individuals in this group*	*Number of cases in this group*	*Number of controls in this group*	*Additional relevant information required for the interpretation of results Example: sex (males/females), age (adults/children)*	*Broad ancestry category that best describes the sample. Multiple values can be listed separated by '|'. Example: European Multiple values can be listed separated by '|'.*
Study tag	**Organism**	**Number of individuals**	**Number of cases**	**Number of controls**	**Sample description**	**Ancestry / Population category**
eg: st1	*Glycine max*	400	200	200	30 days of ages	Yangtze-Huai soybean
Add your data below this line						

以单个GWAS研究为单元

GWAS 数据提交

Publications → **GWAS Study** → Submit

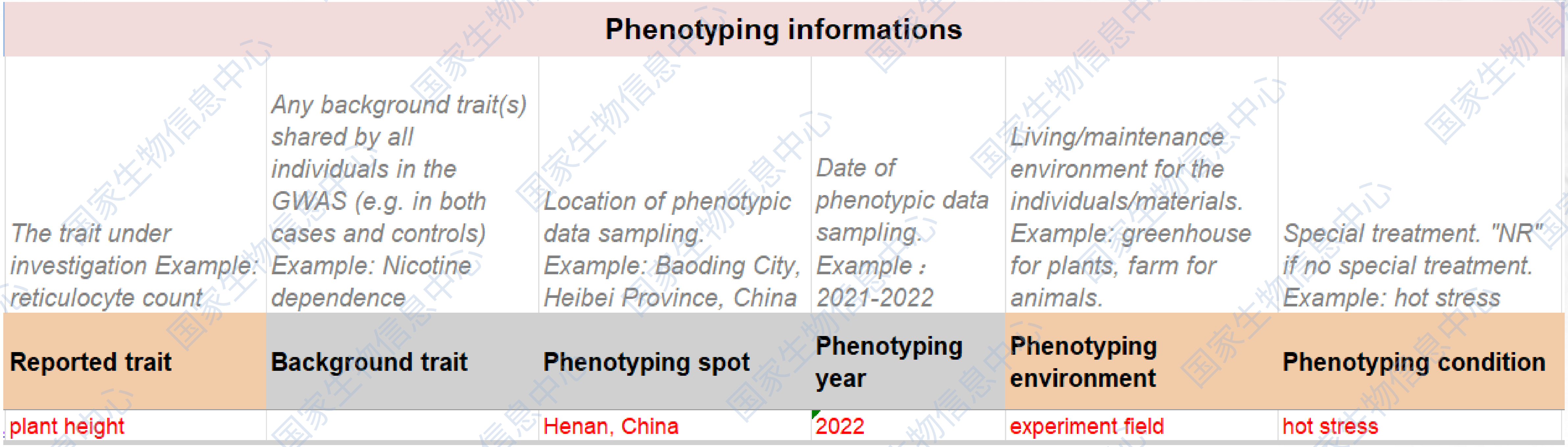


Genotyping informations					
The method used to genotype variants in the discovery stage Multiple values can be listed separated by '|'.	*Additional information about the genotyping array Example: immunochip*	*Were SNPs imputed for the discovery GWAS? Example: yes, no*	*The number of variants analysed in the discovery stage (after QC)*	*Genome assembly for the variants. 'NR' if not providing summary statistics. Example: GRCh38.*	*Url of the genome assembly*
Genotyping technology	**Array information**	**Imputation**	**Variant count**	**Genome assembly**	**Genome link**
whole genome sequence		no	1000	Wm82.a1.v1	https://www.

以单个GWAS研究为单元

GWAS 数据提交

Publications → **GWAS Study** → Submit

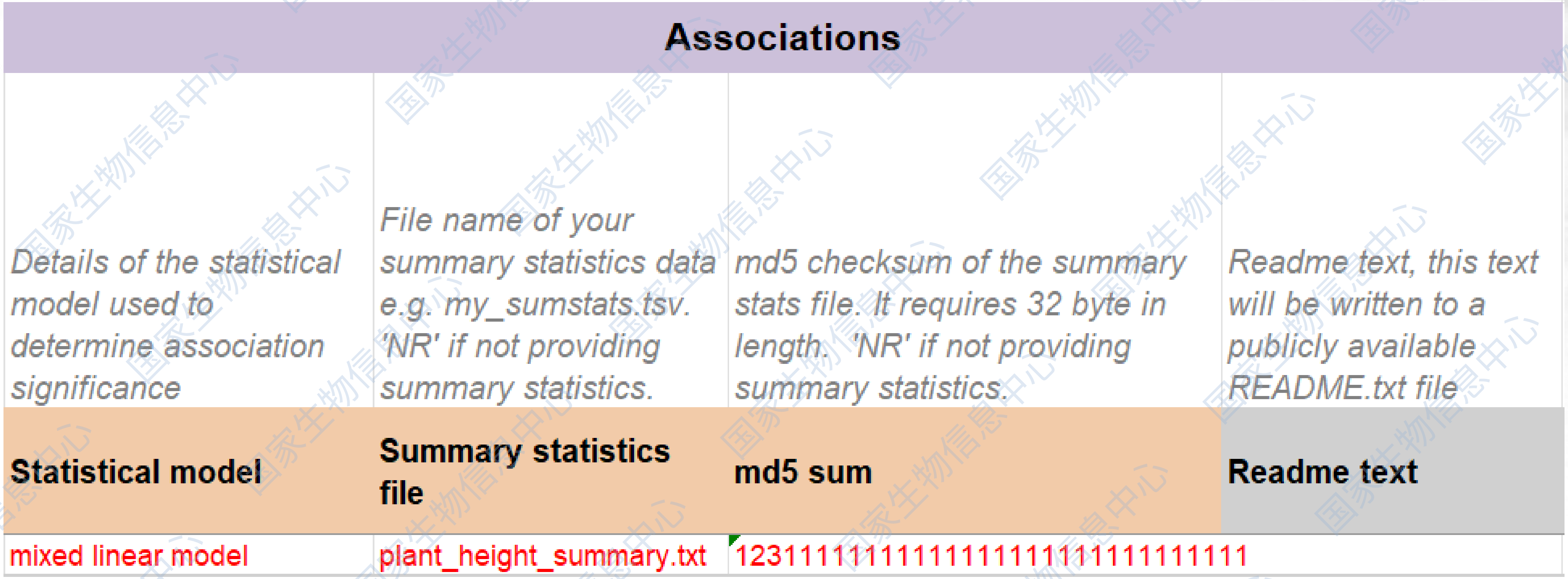


Phenotyping informations					
The trait under investigation Example: reticulocyte count	*Any background trait(s) shared by all individuals in the GWAS (e.g. in both cases and controls) Example: Nicotine dependence*	*Location of phenotypic data sampling. Example: Baoding City, Heibei Province, China*	*Date of phenotypic data sampling. Example：2021-2022*	*Living/maintenance environment for the individuals/materials. Example: greenhouse for plants, farm for animals.*	*Special treatment. "NR" if no special treatment. Example: hot stress*
Reported trait	**Background trait**	**Phenotyping spot**	**Phenotyping year**	**Phenotyping environment**	**Phenotyping condition**
plant height		Henan, China	2022	experiment field	hot stress

以单个GWAS研究为单元

GWAS 数据提交

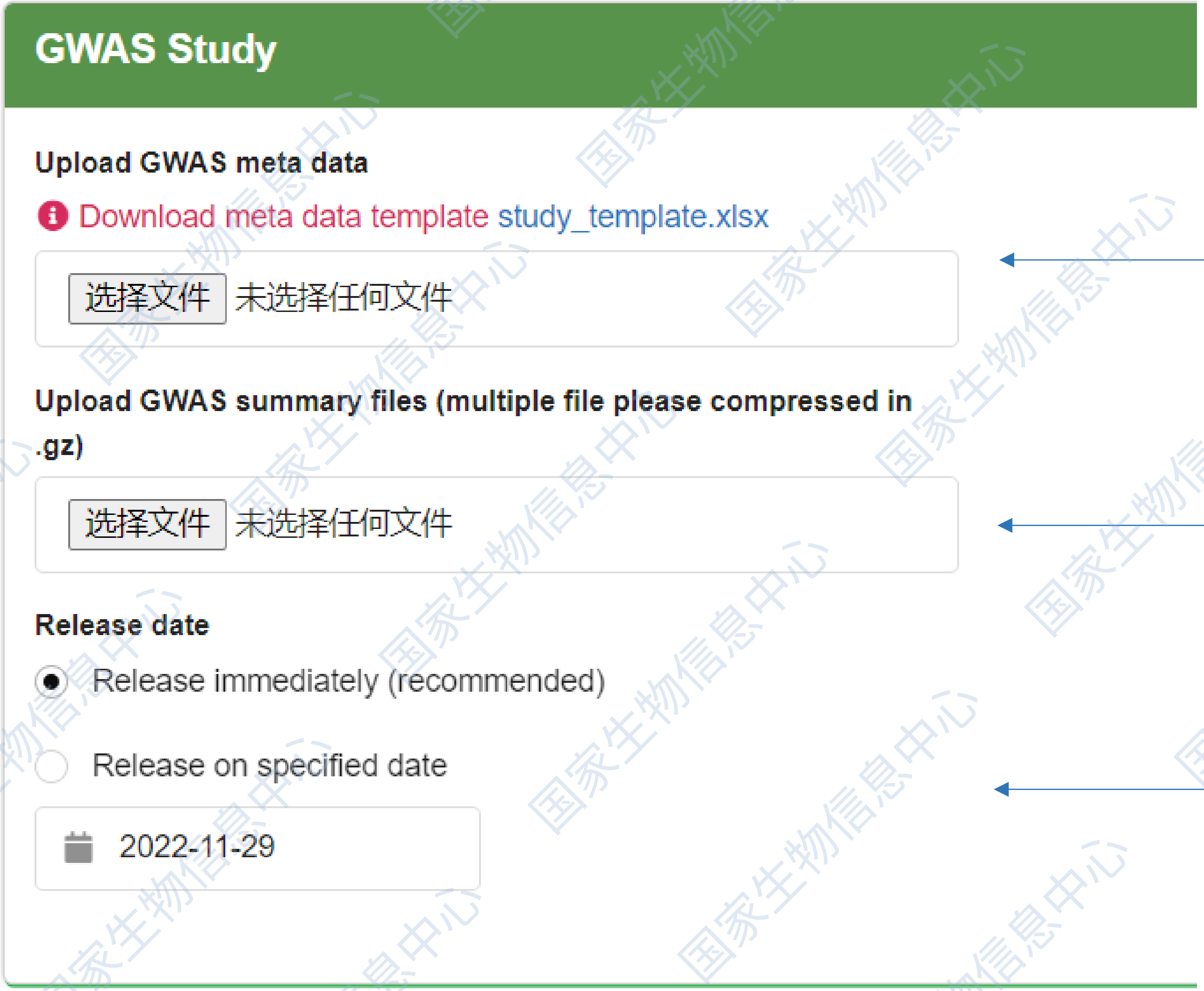
Publications → **GWAS Study** → Submit



Associations			
Details of the statistical model used to determine association significance	*File name of your summary statistics data e.g. my_sumstats.tsv. 'NR' if not providing summary statistics.*	*md5 checksum of the summary stats file. It requires 32 byte in length. 'NR' if not providing summary statistics.*	*Readme text, this text will be written to a publicly available README.txt file*
Statistical model	**Summary statistics file**	**md5 sum**	**Readme text**
mixed linear model	plant_height_summary.txt	12311111111111111111111111111111	

以单个GWAS研究为单元

GWAS 数据提交



GWAS Study

Upload GWAS meta data

Download meta data template study_template.xlsx

选择文件 未选择任何文件

Upload GWAS summary files (multiple file please compressed in .gz)

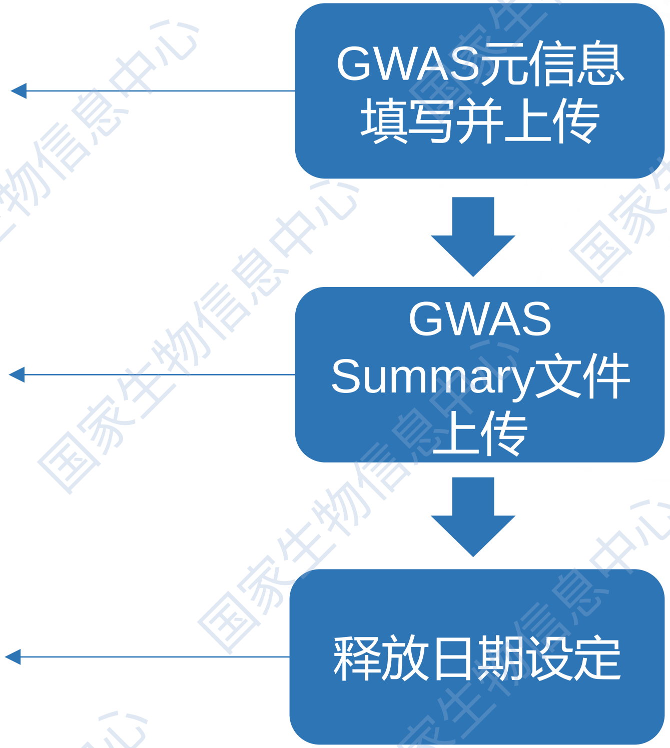
选择文件 未选择任何文件

Release date

● Release immediately (recommended)

○ Release on specified date

2022-11-29



GWAS元信息填写并上传

下载模板，填写GWAS研究相关信息，注意*号内容必填项，尽可能信息完整。

GWAS Summary文件上传

GWAS Summary 文件，说明文档等，多文件要.gz压缩。

释放日期设定

选择立即释放，数据审核通过后其他用户可下载。择期释放，则研究信息可见，数据文件共享权限由数据递交者控制。

GWAS 数据提交

Publications → GWAS Study → **Submit**

Overview

Publications information

Is the study published? *

no

Publication title *

GWAS Analysis of 1000 rice samples.

GWAS Study

Setting GWAS meta information

000039_study.xlsx

Upload GWAS summary (multiple file please compressed .gz)

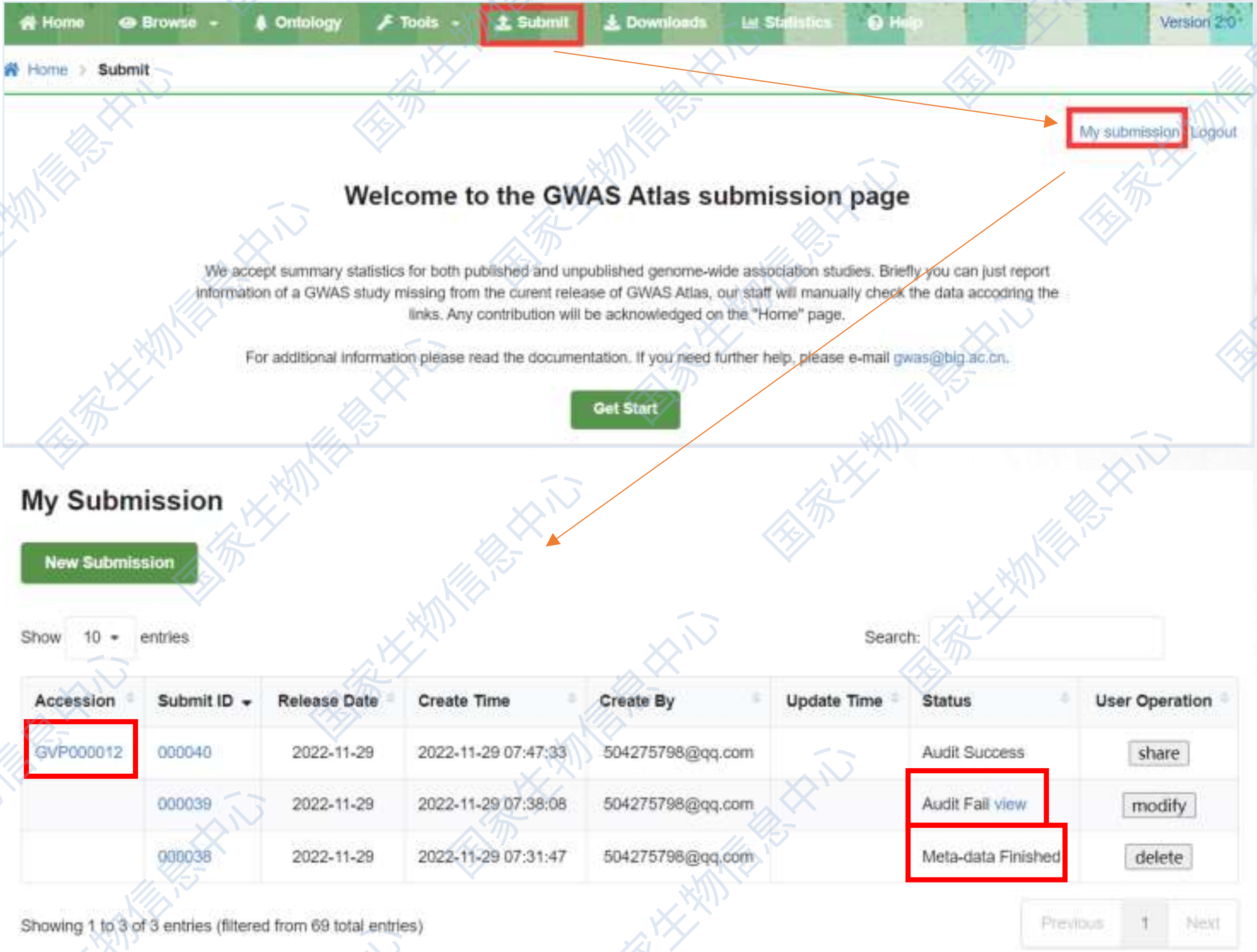
000039_summary..txt

Release Date for GWAS

● Release immediately following curation(recommended)

○ Release on specified date

2022-11-29



Welcome to the GWAS Atlas submission page

We accept summary statistics for both published and unpublished genome-wide association studies. Briefly you can just report information of a GWAS study missing from the curent release of GWAS Atlas, our staff will manually check the data accoding the links. Any contribution will be acknowledged on the "Home" page.

For additional information please read the documentation. If you need further help, please e-mail gwas@big.ac.cn.

My Submission

Show 10 entries Search:

Accession	Submit ID	Release Date	Create Time	Create By	Update Time	Status	User Operation
GVP000012	000040	2022-11-29	2022-11-29 07:47:33	504275798@qq.com		Audit Success	share
	000039	2022-11-29	2022-11-29 07:38:08	504275798@qq.com		Audit Fail view	modify
	000038	2022-11-29	2022-11-29 07:31:47	504275798@qq.com		Meta-data Finished	delete

Showing 1 to 3 of 3 entries (filtered from 69 total entries)

数据库引用

- **引用示例：**

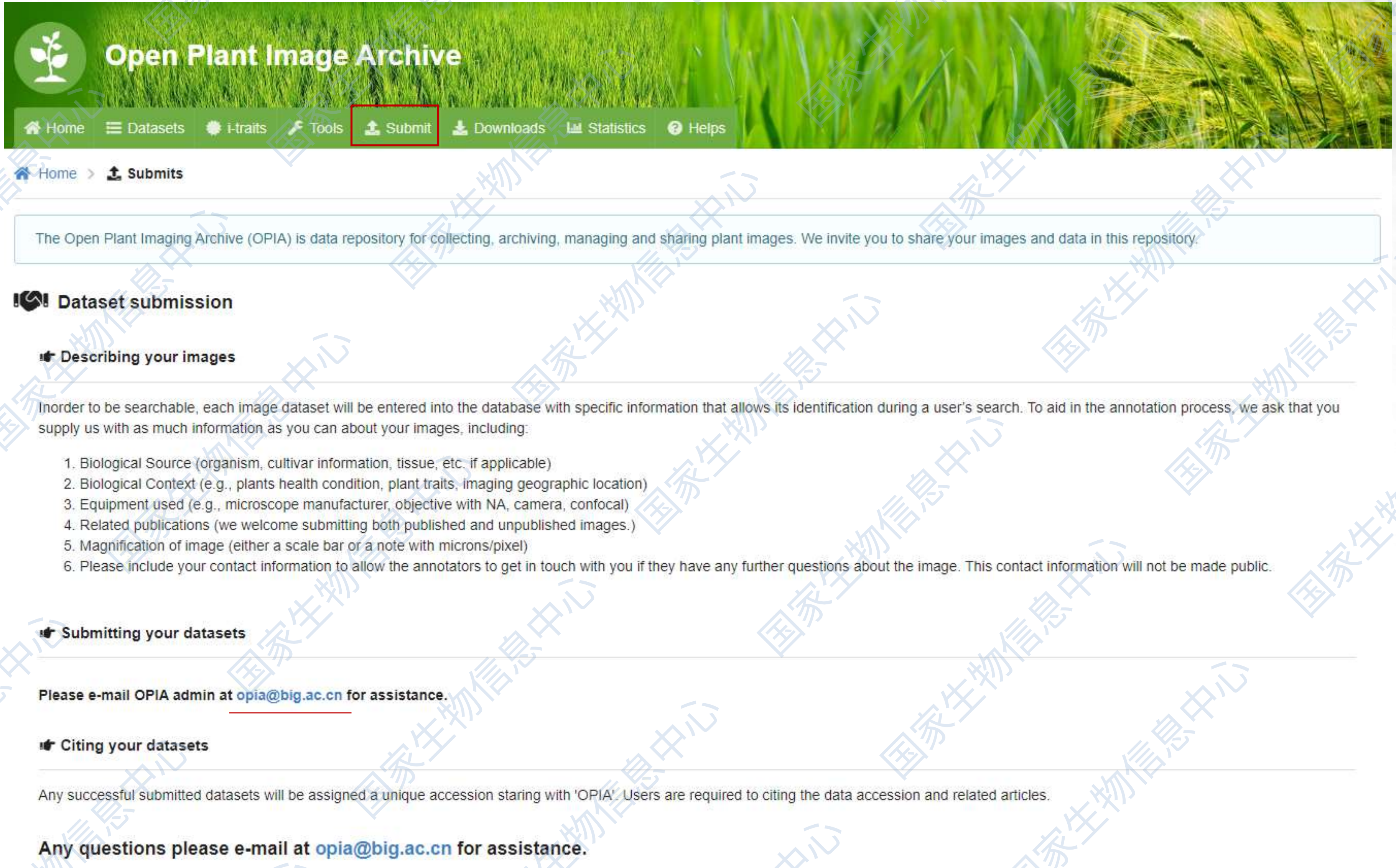
The gwas association data reported in this paper have been deposited in the GWAS Atlas database [1] in National Genomics Data Center, Beijing Institute of Genomics, Chinese Academy of Sciences and China National Center for Bioinformation, under accession number GVPXXXXXX.

- **References:**

[1] GWAS Atlas: an updated knowledgebase integrating more curated associations in plants and animals. *Nucleic Acids Res* 2022. [PMID=36263826].

高通量图像数据汇交

◆ 入口URL: https://ngdc.cncb.ac.cn/opia/submit



The Open Plant Imaging Archive (OPIA) is data repository for collecting, archiving, managing and sharing plant images. We invite you to share your images and data in this repository.

Dataset submission

Describing your images

Inorder to be searchable, each image dataset will be entered into the database with specific information that allows its identification during a user's search. To aid in the annotation process, we ask that you supply us with as much information as you can about your images, including:

1. Biological Source (organism, cultivar information, tissue, etc. if applicable)
2. Biological Context (e.g., plants health condition, plant traits, imaging geographic location)
3. Equipment used (e.g., microscope manufacturer, objective with NA, camera, confocal)
4. Related publications (we welcome submitting both published and unpublished images.)
5. Magnification of image (either a scale bar or a note with microns/pixel)
6. Please include your contact information to allow the annotators to get in touch with you if they have any further questions about the image. This contact information will not be made public.

Submitting your datasets

Please e-mail OPIA admin at opia@big.ac.cn for assistance.

Citing your datasets

Any successful submitted datasets will be assigned a unique accession staring with 'OPIA'. Users are required to citing the data accession and related articles.

Any questions please e-mail at opia@big.ac.cn for assistance.

常见问题

- ◆ Biosample、GVMsample与变异实体数据文件中的样本名称不一致；
- ◆ 多个样本的变异数据，只填写一行样本元信息；
- ◆上传至服务器的变异文件名与元信息表格中填写的文件名不一致；MD5与元信息表格中填写的MD5不一致；
- ◆参考基因组的链接没有直接链接到序列文件或链接无法正常打开；
- ◆ 参考基因组版本与变异数据不符；
- ◆ 文章未发表前，不想释放数据：选择择期释放。到期后可在数据递交页面自行修改延后释放日期。如果文章发表，也请联系我们关联您的文献。

工作团队及联系方式

宋述慧	**负责人**	**songshh@big.ac.cn**
田东梅	系统开发与数据分析	tiandm@big.ac.cn
白雪	生物审编与数据分析	baix@big.ac.cn
唐碧霞	系统开发	tangbx@big.ac.cn
工作邮箱		gvm@big.ac.cn
QQ交流群		468638108

GVM信息交流

扫一扫二维码，加入群聊。