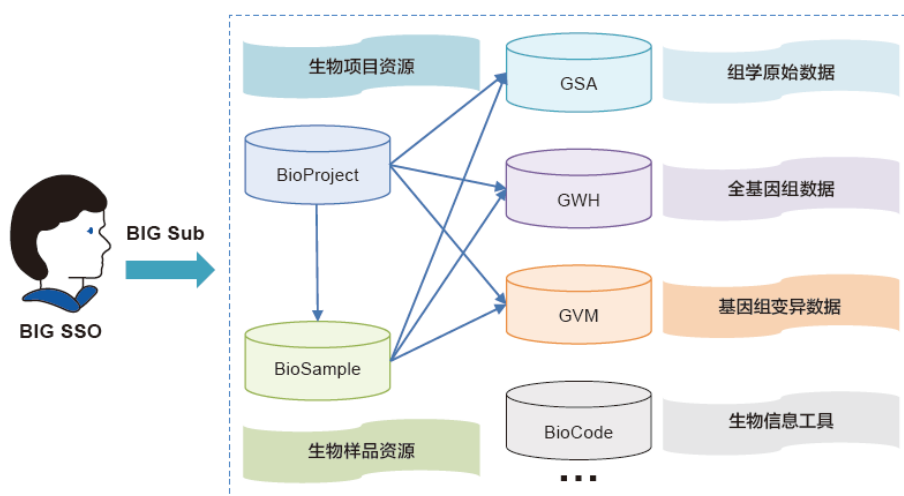


生物数据递交系统（BIG Sub）快速使用说明

系统简介.....	2
用户注册.....	2
使用指南.....	4
数据文件上传.....	7
FTP 上传.....	7
协助上传.....	7

系统简介

生物数据递交系统（BIG Submission, BIG Sub）是大数据中心生物数据统一汇交入口，借助于单点登录（BIG Single Sign-On, BIG SSO）技术，BIG Sub 为用户提供一站式数据递交服务。目前已集成生物项目数据库（BioProject）、生物样本数据库（BioSample）、组学原始数据归档库（Genome Sequence Archive, GSA）、基因组数据库（Genome Warehouse, GWH）、基因组变异数据库（Genome Variation Map, GVM）和生物信息工具库（BioCode）等数据库的提交接口。



用户注册

请您进入生物数据递交系统（BIG Submission, BIG Sub, <http://bigd.big.ac.cn/gsub/>）完成账号注册，建议**使用实验室公共邮箱进行注册**。如果您在账号注册和使用过程中遇到任何问题，请联系 bigd-admin@big.ac.cn。



Register Form

Account Login Information

Email*

Password*

Confirm password

Personal Information

First Name*

Middle Name

Last Name*

Street Address*

City*

State/Province

Postal Code*

Country

Institutional Information

Institute/Organization*

Department*

Lab

Title/Position

Research area



使用指南

The BIG Submission (BIG Sub) is a unified submission portal. It facilitates users to submit data in a single place and accordingly delivers a one-stop service for data submission to BioProject, BioSample, Genome Sequence Archive, Genome Warehouse, Genome Variation Map and BioCode.

- BioProject**
A collection of biological data related to a single initiative, originating from a single organization or from a consortium.
- BioSample**
Descriptions of biological source materials used in experimental assays.
- Genome Sequence Archive (GSA)**
A data repository for archiving raw sequence reads.
- Genome Warehouse (GWH)**
A centralized resource housing genome-scale data.
- Genome Variation Map (GVM)**
A data repository of genome variations.
- BioCode**
Archive bioinformatics codes for open source projects.

1) 生物项目数据库（BioProject）提交入口

BioProject 是收集与共享生物学研究项目信息的资源库，涵盖的项目类型包括常规组学研究的基因组、转录组、表观组和宏基因组等，并针对大型项目提供高效、安全、专业化的项目分级管理。用户可通过此入口进入 BioProject 提交系统，根据信息提示完成研究项目信息的汇交。如果您在使用过程中遇到任何问题，可发邮件至 gsa@big.ac.cn 进行业务咨询。

BioProject is an overall description of a single research initiative; a project will typically relate to multiple samples.

[BioProject Quick Start Guide\(US\)](#) [BioProject Quick Start Guide\(CN\)](#)

Accession	Submission ID	Project title	Release Date	Status	Operation
-----------	---------------	---------------	--------------	--------	-----------

[BioProject 使用说明](#)

2) 生物样本数据库（BioSample）提交入口

BioSample 是收集与共享生物样本信息的资源库，提供生物样本的结构化描述信息递交和发布，涵盖的样本类型包括人、动物、植物、微生物（含环境微生物）、病毒、宏基因组等，提供批量数据上传及离线数据递交服务。用户可通过此入口进入 BioSample 提交系统，根据信息提示完成样本信息的汇交。如果您在使用过程中遇到任何问题，可发邮件至 gsa@big.ac.cn 进行业务咨询

BioSample is a description of the biological source material; each physically unique specimen should be registered as a single BioSample with a unique set of attributes.

[BioSample Quick Start Guide\(US\)](#) [BioSample Quick Start Guide\(CN\)](#)

Accession	Submission ID	Title	Project Accession	Organism	Release date	Status	Operation
-----------	---------------	-------	-------------------	----------	--------------	--------	-----------

BioSample 使用说明

3) 组学原始数据归档库（Genome Sequence Archive, GSA）提交入口

GSA 是组学原始数据汇交、存储、管理与共享系统，是国内首个被国际期刊认可的组学数据发布平台。GSA 围绕国家精准医学和重要战略生物资源，面向全球接收不同测序平台产生的组学原始数据，提供免费的长期存储、管理与共享服务。用户可通过此入口进入 GSA 提交系统，根据信息提示完成实验（Experiment）和测序反应（Run）信息的汇交。如果您在使用过程中遇到任何问题，可发邮件至 gsa@big.ac.cn 进行业务咨询。

GSA is a data repository specialized for archiving raw sequence reads. Before create a new GSA submission, you should create [BioProject](#) (an overall description of a single research initiative) and [BioSample](#) (a description of biological source material) if needed. Please refer to the [GSA Submission Quick Start Guide](#).

[How to share and release your data](#) [如何共享和发布 GSA 数据集](#)

[How to reset your release date](#) [如何延期发布 GSA 数据集](#)

[Create GSA](#)

Accession	Submission ID	Title	Release date	Status	Operation
-----------	---------------	-------	--------------	--------	-----------

GSA 使用说明

4) 基因组数据库（Genome Warehouse, GWH）提交入口

GWH 是面向多物种全基因组数据的公共归档数据库，为基因组数据汇交、存储、释放和共享提供一系列 web 服务。基因组序列和基因组注释信息是组学数据分析的基础，该库汇交了包括动物、植物、真菌、细菌等国家战略资源物种的基因组信息，为组学分析、

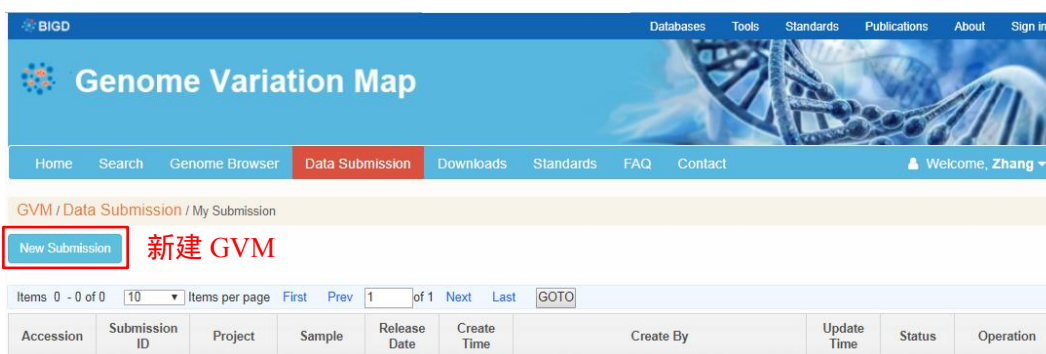
比较基因组分析和分子育种等研究提供重要数据资源。用户可通过此入口进入 GWH 提交系统，根据信息提示完成基因组序列和基因组注释信息的汇交。如果您在使用过程中遇到任何问题，可发邮件至 gwh@big.ac.cn 进行业务咨询。



[New Submission](#) 新建 GWH

5) 基因组变异数据库（Genome Variation Map, GVM）提交入口

GVM 是国际领先、国内规模最大的基因组序列变异库。该库基于人工审编整合了包括人、畜牧动物、主要农作物和其他资源物种在内的大量基因组序列单核苷酸多态位点和小的插入与删除变异信息，是基因组序列变异信息汇交、管理与检索的专业资源库。用户可通过此入口进入 GVM 提交系统，根据信息提示完成基因组序列变异信息的汇交。如果您在使用过程中遇到任何问题，可发邮件至 gvm@big.ac.cn 进行业务咨询。



6) 生物信息工具库（BioCode）提交入口

BioCode 是整合开源生物信息软件工具的数据库，其主要功能是分类搜集整理生物信息软件工具源代码、软件包以及重要元数据信息，包括软件名称、功能描述、分类、发表文章、文章引用情况、开发人员信息及单位信息等。用户可通过此入口进入 BioCode 提交系统，根据信息提示完成开源生物信息软件工具的汇交。

数据文件上传

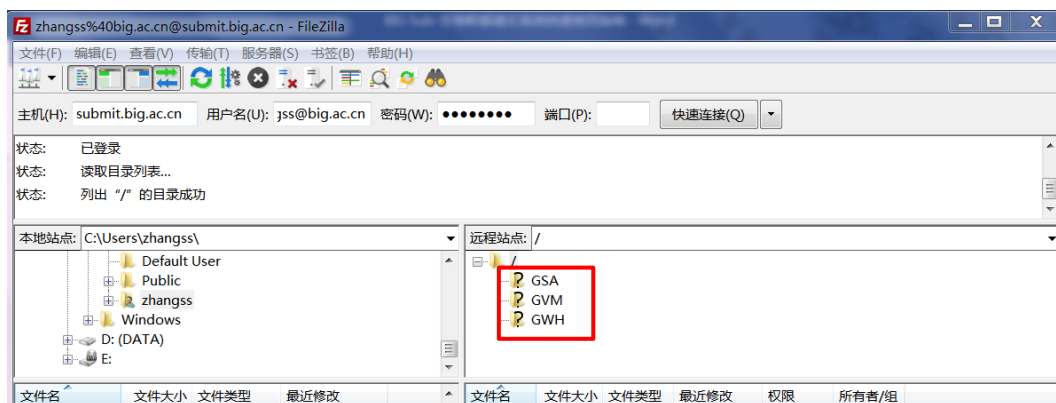
FTP 上传

请使用 FTP 客户端软件（比如 [FileZilla Client](#)）登录 FTP 服务器，用户账号与 BIG sub 账号一致。

FTP 服务器地址:

<ftp://submit.big.ac.cn>

注意：用户登录自己的FTP路径后，先cd到相应数据库目录下（如：/GSA, /GWH, /GVM）再上传文件。请采用二进制模式上传：如果是用FTP软件上传，请参考软件说明进行设置；如果是用FTP指令上传，请在mput指令前，先运行binary指令设置二进制传输方式。



此外，数据上传完毕，并收到您的元数据信息后，后台系统将对您的数据进行相应的审核，请您耐心等待并密切关注系统和邮箱的情况反馈。

协助上传

考虑到大体量数据递交用户的需求，开启了硬盘寄送和协助上传的绿色通道，详情请垂询各数据库工作组。

数据库工作组	工作邮箱
Genome Sequence Archive	gsa@big.ac.cn
Genome Warehouse	gwh@big.ac.cn
Genome Variation Map	gvm@big.ac.cn