

CNSA 数据归档系统操作手册

一、 简介

中国核酸序列归档服务（CNSA）是一个方便快捷在线提交生物研究项目、样本、实验等信息数据的系统，结合国际核酸序列数据库（International Nucleotide Sequence Database Collaboration, INSDC）标准，遵循 Fort Lauderdale Agreement、NHGRI Rapid Data Release Policies、Joint Data Archiving Policy、CC0-No Rights Reserved 等国际数据开放协议，接受来自全球科研的测序研究数据（包括原始数据和其他支撑数据）递交，其数据递交服务可作为文献出版流程的补充，支持早期数据的共享。CNSA 致力于生物测序信息和数据的存储和共享，旨在为全球的研究者提供当前最全面的数据和信息资源，促使研究人员能够便捷深入的访问和使用数据。

二、 数据提交流程

1. 登录数据库

CNGB 所有数据库使用统一的用户注册登录平台，注册账号适用于所有的 CNGB 数据库。用户可通过网址（<http://db.cnbg.org/cnsa>）进入 CNSA 数据库主页（如图 1），请先阅读页面上方的用户须知并点击同意。点击页面右边的“登录/注册”标签进入注册登录页面（如图 2）。（注意：需先注册才能登入，提交数据。）



图 1 主页面



语言 ▼

登录

帐户

密码

<> 验证码 

登录

[注册帐户。](#) [忘记密码？](#)

其他登录方式：

图 2 登录页面

2. 首页

2.1 CNSA 简介

简要概括 CNSA 的目的，意义。

2.2 全局检索

可查询 CNSA 中所有已发布的项目/样本/实验/组装/测序信息（见图 3）。

2.3 数据递交入口

两种递交方式：ENA-BROKER 和 DOI 标准（GigaDB）入口，分别提供原始数据和其他支撑数据的递交和审编（见图 3）。

检索

提供全局检索，可查询 CNSA 中所有已发布的项目 / 样本 / 实验 / 测序 / 组装 信息。不提供GigaDB的检索结果，若要检索GigaDB，请点击<http://gigadb.org>进行查询。

请输入检索关键字。你可以参考下面的示例来获取可用的检索条件。

Q 检索

示例: 编号 `CNP000001/CNS000001/CNX000001/CNR000001/CNA000001` , 项目标题, 样本名称, 实验标题, 物种名称, 数据类型, 样本类型, 描述信息。

提交流程

对以下两种提交流程有任何疑问，可以联系管理员 (datasubs@cngb.org) 寻求帮助。



原始数据提交

原始数据（指测序产生的未经任何过滤的全部下机数据）提交：需要发表文章且符合人类遗传资源条例和伦理的数据默认通过本服务流程同步至EBI-ENA；如果递交的数据需要权限控制，或者需要选择上传数据到 NCBI SRA 或者 DDBJ DRA，请点击[帮助](#)。



其他支撑数据提交

其他支撑数据，作为原始数据的补充，包括但不限于过程和结果数据、分析方法、软件程序、图像文件、音频文件、视频文件、成像文件、电子图表和word文档等。CNSA与Gigascience旗下的GigaDB合作共同审编和归档此类数据，同时通过与DataCite的链接，每个数据集都将被分配一个DOI，可快速被引用。详情请点击[帮助](#)。

图 3 全局检索和数据递交入口

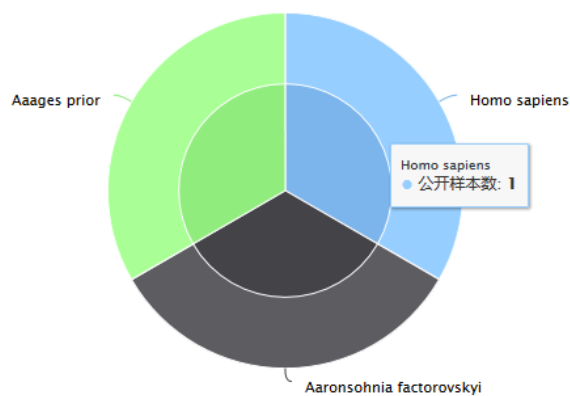
2.4 统计信息

统计用户提交的样本信息中各个物种的分布情况（见图 4）。

数据统计

物种在样本中分布

在下面图表中，每种颜色代表了一个物种。每个颜色的饼块由内圈和外圈构成。内圈代表的是当前物种所有的样本数。外圈可以由深浅两种颜色分成两块，深色块代表的是当前物种中已公开的样本数，浅色块代表的是未公开的样本数，根据样本的公布情况，外圈的两个色块不一定会都存在。



研究最热门的10个物种

生物/物种名称	样本	Controlled	已公开
Homo sapiens	1	0	1
Aaronsohnia factorovskiy	1	0	1
Aaages prior	1	0	1

图 4 数据统计

3. 个人中心

3.1 提交流程

先进行项目信息填写，再递交样本信息，然后分为两条路径，一条为提交实验/测序信息；另一条为提交组装信息（见图 5）。

3.2 提交进展

统计当前用户已提交的项目/样本/实验/测序/组装信息（见图 5）。



图 5 提交流程和进展

3.3 提交者信息登记

填写提交者信息，与其后提交的项目/样本/实验/测序/组装信息关联，支持注册用户与提交者非同一人（见图 6）。

提交历史 提交人

基本信息

* 名 中间名 * 姓

* 主要邮箱 备选邮箱

单位信息

* 部门 * 组织/单位名称 组织/单位网址

* 邮编 电话 传真

地址信息

* 街道 * 城市 省/州

* 国家/区

图 6 提交者信息录入

4. 项目提交

4.1 登录项目页面

4.1.1 公开项目信息查询

点击导航栏“项目”进入到项目信息页面，该页展示所有用户公开的项目列表信息。点击编号，可浏览项目的公开详情页（如图 7-8）。



The screenshot shows the CNSA website interface. At the top is the CNSA logo. Below it is a navigation bar with links: 首页, 我的提交, 项目, 样本, 实验/测序, 组装, 帮助. The '项目' link is highlighted. Below the navigation bar, there are two buttons: '已提交项目' (Submitted Projects) and '新的项目' (New Projects). The main content area displays a table of projects under the heading '项目 > 公开项目'. The table has columns for 编号 (ID), 项目标题 (Project Title), 提交人 (Submitter), 生物/物种名称 (Organism), 数据类型 (Data Type), and 公开日期 (Public Date). Below the table is a pagination control with buttons for 首页, 上一页, 1, 下一页, 末页, and a '跳到第' field followed by '页' and '确定'.

编号	项目标题	提交人	生物/物种名称	数据类型	公开日期
CNP0000004	Ant	nancy guo	Aaages prior	Epigenomics	2017-10-24
CNP0000003	Single cells Genome Seque...	lijin you	Homo sapiens	Raw sequence reads	2017-10-23
CNP0000002	search algorithm	nancy guo	Aaages prior	Exome	2017-10-23
CNP0000001	proejct title testing	fan yang	Abarema microcalyx	Genome sequencing and as...	2017-10-22

图 7 项目信息页

项目: CNP0000003

概述

标题: Single cells Genome Sequencing-1
提交人: lijin you; BGI
描述: The pilot study on single cell CNV detection in BGI-Shenzhen. Single cells isolated from blood or embryos.
生物/物种名称: Homo sapiens
数据类型: Raw sequence reads
ENA ID: -

数据

样本: 1
实验: 1
测序: 1
组装: 1
下载: -

图 8 项目公开详情页

4.1.2 用户提交项目

点击“已提交项目”，可查看用户本人提交的项目（图 9），包括已完成提交和未完成提交的项目。支持以项目标题/物种为过滤条件进行查询，点击“编辑”自动跳转到用户上次还未完成的步骤。对于已完成提交的项目，跳转到概览页面。

点击“新的项目”进入到项目提交页面（如图 10），可看到表单。



图 9 我的项目页

4.2 项目基本信息页

项目基本信息页（如图 10），此页要求填写项目的基本信息，如项目名称、项目关联领域、项目公开描述信息、外部链接、关联项目、基金信息、合作信息、数据提供者信息、项目公开日期。注意，星号（*）标识为必填项，其余为选填项。



图 10 项目基本信息页

4.3 项目详情页

项目基本信息填完后，点击“保存并继续”按钮，进入到项目详情页（如图 11），本页面要求填写项目类型、样本范围等信息。注意，星号（*）标识为必填项，其余为选填项。

1 2 3

项目类型

*** 项目数据类型**

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

*** 样本范围**

Multisolate

研究对象

* 分类编号	* 生物/物种名称	
<input type="text" value="请先输入然后从下拉列表中选择匹配的信息"/>	<input type="text" value="Cyclocynips tumorvirgae"/>	
分型/品种	品种/育种	栽培品种
<input type="text" value="请输入"/>	<input type="text" value="请输入"/>	<input type="text" value="请输入"/>

图 11 项目详情页

4.4 项目概览页

项目类型信息填完后，点击“保存并继续”按钮，进入到概览页（如图 12），本页面是对前面两个表单信息的汇总。如发现有误，可进入到前面任何一个页面进行相应的修改，如核对无误，可点击“提交”进行提交，项目提交完之后，会自动生成一个访问号如（CNP0000003）。可通过访问号查看详细的项目信息。注意，点击“提交”后也可进行修改。在整个项目信息填写过程中，中途退出系统会保留上一次的填写结果。

样本 ▶ 公开样本

[我的样本](#)
[新的样本](#)

编号	样本名称	提交人	生物/物种名称	项目	公开日期
CNS0000002	Typographic variants	nancy guo	Aaages prior	CNP0000002	2017-10-23
CNS0000001	BGI_Single_Cell	lijin you	Homo sapiens	CNP0000003	2017-10-23

[首页](#) [上一页](#) [1](#) [下一页](#) [末页](#) 跳到第 页 [确定](#)

图 13 样本信息页

样本: CNS0000001

概述

名称: BGI_Single_Cell
提交人: lijin you
生物/物种名称: Homo sapiens
ENA ID: -

属性

样本名称: BGI_Single_Cell
样本标题: -
分类编号: 9606
生物/物种名称: Homo sapiens
通用名称: man
分离个体: Homo sapiens
描述: -
年龄: 20
生物材料提供者: bgi
性别: female
组织类型: blood
细胞系: -
细胞亚群: -
细胞类型: -
菌种保藏: -
发育阶段: -

图 14 样本公开详情页

5.1.2 用户提交样本

点击“我的样本”，可查看用户本人提交的样本（图 15），包括已完成提交和未完成提交的样本。支持以样本名称/物种为过滤条件进行查询，点击“编辑”自动跳转到用户上次还未完成的步骤。对于已完成提交的样本，跳转到概览页面。

点击“新的样本”进入到样本提交页面，可看到表单。

检索字段：提交历史，编号，样本名称，物种分类编号，物种



提交历史	编号	样本名称	分类编号	生物/物种名称	状态	最后更新日期	操作
sub000005	-	-	-	-	processing, a...	2017-10-23 14:54:55	查看 编辑 删除
sub000004	CNS0000001	BGI_Single_Cell	9606	Homo sapiens	finished	2017-10-23 14:03:07	查看 编辑
sub000003	-	-	-	-	processing, a...	2017-10-23 12:01:46	查看 编辑 删除

[首页](#) [上一页](#) [1](#) [下一页](#) [末页](#) 跳转到 页 [确定](#)

图 15 我的样本页

5.2 样本基本信息页

样本基本信息页（如图 16），此页要求填写样本的基本信息，如项目编号，样本公开日期。注意，星号（*）标识为必填项，其余为选填项。

[首页](#) [我的提交](#) [项目](#) [样本](#) [实验/测序](#) [组装](#) [帮助](#)

样本 ▶ 提交信息 ▶ 新的样本

1 基本信息
2 样本类型
3 属性
4 概览

基本信息

* 项目编号 ?

请输入并从弹出的建议列表中选择。

样本发布

* 公开日期(北京时间) ?

2017-10-25
≡

保存并继续

图 16 样本基本信息页

5.3 样本类型页

样本基本信息填完后，点击“保存并继续”按钮，进入到样本类型页（如图 17），本页面要求选择样本类型。

样本 > 提交信息 > 新的样本



样本类型

- Pathogen affecting public health
 - Clinical or host-associated pathogen(v1)
 - Environmental, food or other pathogen(v1)
- Microbial sample(v1)
- Model organism or animal sample(v1)
- Metagenome or environmental sample(v1)
- Invertebrate(v1)
- Human sample(v1)
- Plant sample(v1)
- Virus sample(v1)
- GSC MixS environmental sample

图 17 样本类型页

5.4 样本属性页

选择样本类型后，点击“保存并继续”按钮，进入到样本属性页（如图 18），不同的样本类型要求填写不同的属性信息，如人类样本属性页要求填写样本名称、分离描述、年龄、生物材料提供者、性别、组织。支持增加更多的属性信息。注意，星号（*）标识为必填项，其余为选填项。

[首页](#)
[我的提交](#)
[项目](#)
[样本](#)
[实验/测序](#)
[组装](#)
[帮助](#)

[样本](#) > [提交信息](#) > [新的样本](#)

[基本信息](#)
[样本类型](#)
[属性](#)
[概览](#)

基本信息

*** 样本名称**
样本标题

*** 分类编号**
*** 生物/物种名称**
通用名称

请至少填写这些字段中的一个：分型/品种, 分离个体。

分型/品种
分离个体

描述

*** 经纬度**
菌株保藏
基因型

宿主年龄
宿主其他描述
宿主疾病结果

宿主疾病阶段
宿主健康状况
宿主性别

宿主识别编号
宿主取样组织
传代史

致病型
血清型
血清变型

样本凭证
亚群
亚型

自定义属性 + 增加

名称	值	操作
没有数据。		

上一步
保存并继续

图 18 样本属性页

5.5 概览页

样本属性信息填完后，点击“保存并继续”按钮，进入到概览页（如图 19），本页面是对前面三个表单信息的汇总。如发现有误，可进入到前面任何一个页面进行相应的修改，如核对无误，可点击“提交”进行提交，样本提交完之后，会自动生成一个访问号如（CNS0000006）。可通过访问号查看详细的样本信息。注意，点击“提交”后也可进行修改。在整个样本信息填写过程中，中途退出系统会保留上一次的填写结果。

首页 我的提交 项目 样本 实验/测序 组装 帮助

样本 > 提交信息 > 新的样本

基本信息 样本类型 属性 概览

1 2 3 4

基本信息

- * 样本编号: CNP0000003
- 公开日期: 2017-10-23

样本类型

- * 样本类型: Human sample

属性

- * 样本名称: BGI_Single_Cell
- 样本标题: -
- * 分类编号: 9606
- * 生物/物种名称: Homo sapiens
- 通用名称:

图 19 样本概览页

6. 实验/测序提交

6.1 登录实验/测序页面

输入账号和密码后，将进入到 CNSA 页面，点击导航栏“实验/测序”进入到实验/测序信息页（如图 20）。点击“Run”，可进入测序的列表页面（见图 21）查看提交的测序的公开列表；点击带有 CNR 前缀的编号，可看到测序的详细公开信息，与该测序相关联的项目，样本，实验信息以及可供下载的数据（见图 22）。

CNSA

首页 我的提交 项目 样本 实验/测序 组装 帮助

实验/测序 > 公开实验

我的实验 新的实验

编号	实验标题	提交人	平台	项目	公开日期	操作
CNX0000005	start_guo	nancy guo	454 GS 20	CNP0000002	2017-10-24	Run
CNX0000004	data_sync_test	nancy guo	454 GS 20	CNP0000002	2017-10-24	Run
CNX0000003	nancy.guo	nancy guo	454 GS 20	CNP0000002	2017-10-24	Run
CNX0000002	test guo	nancy guo	454 GS 20	CNP0000002	2017-10-23	Run
CNX0000001	Illumina HiSeq 2000 seq...	lijin you	Illumina HiSeq 2000	CNP0000003	2017-10-23	Run

首页 上一页 1 下一页 末页 跳到第 页 确定


图 20 实验/测序信息页

CNX0000005的测序 x

编号	文件类型
CNR0000005	bam

[首页](#) [上一页](#) [1](#) [下一页](#) [末页](#) [跳到第](#) [页](#) [确定](#)

图 21 测序列表页面



[首页](#) [我的提交](#) [项目](#) [样本](#) [实验/测序](#) [组装](#) [帮助](#)

测序: CNR0000005

概述

别名 -
 生物/物种名称 Aages prior
 ENA ID -

数据

项目 CNP0000002
 样本 CNS0000002
 实验 CNX0000005
 公开日期 -

文件名	文件大小	状态	编号	下载
start_guo.bam	317	uploaded	CND0000005	下载

图 22 测序概览页

点击“我的实验/测序”标签（如图 23）可查看用户本人提交的实验/测序。点击“新的实验/测序”进入到实验/测序提交页面。点击“Run”，可进入测序的个人递交列表页面（见图 24），查看提交的测序的详细列表及数据上传进度；点击“查看”按钮，可查看提交的详细信息（见图 25）。



[首页](#) [我的提交](#) [项目](#) [样本](#) [实验/测序](#) [组装](#) [帮助](#)

实验/测序 [提交信息](#)
[我的实验](#) [新的实验](#)

检索字段: 编号, 实验标题, 项目

编号	实验标题	项目	最后更新日期	公开日期	文件提交	操作
CNX0000001	Illumina HiSeq 2000 sequencin...	CNP0000003	2017-10-23 14:20:25	2017-10-23	Uncompleted	Run

[首页](#) [上一页](#) [1](#) [下一页](#) [末页](#) [跳到第](#) [页](#) [确定](#)

图 23 我的实验/测序页

CNX0000001的测序 ×

编号	文件类型	最后更新日期	已发布	状态	操作
CNR0000001	fastq	2017-10-23 14:10:25	Yes	<div style="width: 100%; height: 10px; background-color: #ccc;"></div>	查看 编辑

● 上传完毕 ● 上传中 ● 上传失败 ● 未上传

[上一页](#) [下一页](#) 跳到第 页 [确定](#)

图 24 个人递交测序列表页面

首页 我的提交 项目 样本 **实验/测序** 组装 帮助

实验/测序: CNR0000001

文件列表

- * 文件类型: fastq
- * 文件名称: BGI_Single_Cell_HUMafrRHYDFAAPEI-1_110812.gz
- * MD5: 38b8c2c1093dd0fec383a9d9ac940515
- * 状态: not upload
- 是否为曾提交的实验?: 是 --

基本信息

- * 项目编号: CNP0000003
- * 样本编号: CNS0000001
- * 标题: Illumina HiSeq 2000 sequencing; Single cell CNV
- * 平台: Illumina HiSeq 2000

文库信息

- * 文库名称: HUMafrRHYDFAAPEI-1
- * 文库构建策略: WGS
- * 文库来源: GENOMIC
- * 文库选项: PCR

图 25 实验/测序查看页面

6.2 实验/测序提交页面

进入实验/测序提交页面后（如图 26），该页要求首先填写数据文件类型，文件名及 MD5 值，MD5 值用来确定数据传输的完整性。其次，填写相关的实验信息，如元信息，包括测序平台、实验名称、项目编号、样本编号；建库信息，包括建库名称、建库策略、

样本来源、选择和富集方式、测序设计（单末端/双末端测序）；流程信息，包括使用程序及版本；发布日期。需要注意的是，在提交实验信息之前，需要先提交项目信息和样本信息，有相应的项目编号和样本编号，才能提交实验信息。填写完成后，数据库会反馈 FTP 账户、密码，用户可通过此账户密码将数据文件传到 FTP 服务器，完成数据提交。上传到 FTP 的数据，会通过 MD5 来检查和确认数据传输的完整性。填写并检查无误后，可点击“提交”进行提交，信息提交完之后，会自动生成两个访问号如（CNX0000006/CNR000001）。可通过访问号查看详细的实验/测序信息。注意，星号（*）标识为必填项，其余为选填项。点击“提交”后均可进行修改。在整个实验/测序信息填写过程中，中途退出系统不会保留上一次的填写结果。

首页 我的提交 项目 样本 实验/测序 组装 帮助

实验/测序 > 提交信息 > 新的实验

* 文件类型 ?
请选择

* 文件列表 ? + 增加

文件名称	MD5	状态	操作
没有数据。			

* 是否为曾提交的实验？
 是 否

基本信息

* 项目编号 ?

* 样本编号 ?

* 平台 ?

* 标题 ?

图 26 实验/测序提交页面

7. 组装提交

7.1 登录组装页面

点击导航栏“组装”进入到组装公开信息页（如图 27），该页展示所有用户公开的组装列表信息。点击编号，可浏览组装的公开详情页（如图 28）。该页面包含公开的组装信息，以及关联的项目，样本和可供下载的数据等信息。

组装 > 公开组装

我的组装

新的组装

编号	标题	提交人	组装方法	公开日期
CNA0000001	single cell scaffold	lijin you	SOAPdenovo	2017-10-24

[首页](#)
[上一页](#)
[1](#)
[下一页](#)
[末页](#)
 跳到第 页 [确定](#)

图 27 组装信息页

组装: CNA0000001

概述

标题: 组装
组装名称: single cell scaffold assembly
分子类型: genomic DNA
测序技术: Illumina HiSeq
组装方法: SOAPdenovo(1.0)
ENA ID: -

数据

项目: CNP0000003
样本: CN50000001
公开日期: 2017-10-24

文件名称	文件类型	状态	文件大小	编号	下载
无可用数据。					

图 28 组装概览页

点击“我的组装”标签，出现用户本人提交的组装信息，可看到用户提交的数据文件状态（见图 29）。点击“查看”，可查看提交的详细信息（如图 30）；点击“编辑”，可编辑更新已提交信息（如图 31）。点击“新的组装”进入到组装提交页面（如图 32）。

组装 > 提交信息

我的组装

新的组装

检索字段：编号，标题 [Q](#)

编号	标题	最后更新日期	状态	操作
CNA0000001	single cell scaffold	2017-10-24 16:12:32	<div style="width: 100%; height: 10px; background-color: #ccc;"></div>	查看 编辑

[首页](#)
[上一页](#)
[1](#)
[下一页](#)
[末页](#)
 跳到第 页 [确定](#)

● 上传完毕
 ● 上传中
 ● 上传失败
 ● 未上传

图 29 我的组装页

首页 我的提交 项目 样本 实验/测序 组装 帮助

组装详情: CNA0000001

基本信息

- * 项目编号: CNP0000003
- * 样本编号: CNS0000001

文件列表

- * 文件类型: scaffold_fasta
- * 文件名称: HUMAN_SCAFFOLD.fasta.gz
- * MD5: e3b5dd475c449300dd11f258538ff493
- * 状态: not upload

组装元数据

- * 标题: single cell scaffold
- * 组装名称: -
- * 分子类型: genomic DNA
- * 覆盖度: 120
- * 测序技术: Illumina HiSeq

组装方法

- * 组装方法: SOAPdenovo(1.0)
- * 组装方法描述: A method of assembling a plurality of individual parts to manufacture a workpiece.
- * 部分基因组: 否

图 30 组装查看页面



图 31 组装编辑页面

7.2 组装提交页面

进入组装提交页面后（如图 32），该页要求填写相关的组装信息，如项目编号、样本编号、文件类型，名称，状态；组装元信息，包括组装标题、组装名称、分子类型、覆盖度、测序技术、组装方法，需要注意的是，在提交组装信息之前，需要先提交项目信息和样本信息，有相应的项目编号和样本编号，才能提交组装信息。填写完成信息后，数据库下方会反馈 FTP 账户、密码，用户可通过此账户密码将数据文件传到 FTP 服务器，完成数据提交。上传到 FTP 的数据，会通过 MD5 来检查和确认数据传输的完整性。填写并检查无误后，可点击“提交”进行提交，组装信息提交完之后，会自动生成一个访问号如（CNA0000006）。可通过访问号查看详细的组装信息。注意，星号（*）标识为必填项，其余为选填项。点击“提交”后，对于提交信息可以进行修改。在整个组装信息填写过程中，中途退出系统不会保留上一次的填写结果。

组装 > 提交信息 > 新的组装


基本信息

* 项目编号 

请输入然后从下拉列表中选择

* 样本编号 

请输入然后从下拉列表中选择

* 文件列表 

+ 增加

注意:
类型为 `contig_fasta` 和 `contig_flatfile` 的数据不能同时提交。
类型为 `scaffold_fasta` 和 `scaffold_agp` 的数据不能同时提交。
类型为 `chromosome_fasta` 和 `chromosome_agp` 的数据不能同时提交。
必须先提交类型为 `contig_fasta` 或 `contig_flatfile` 的数据, 才能提交类型为 `unlocalised_contig_list` 的数据。
必须先提交类型为 `scaffold_fasta` 或 `scaffold_flatfile` 或 `scaffold_agp` 的数据, 才能提交类型为 `unlocalised_scaffold_list` 的数据。
所有类型的数据只能提交一次。

文件类型	文件名称	MD5	状态	操作
没有数据。				

图 32 组装提交页面

三、 帮助

其他帮助请参考 CNSA 的帮助页面里的 FAQ。