

Zhang-Lab 生信小课堂 第十二期
Applied Bioinformatics Club (ABC)

和趣求真  秉实生信

(张建伟生物信息学课题组 <https://zhang.hzau.edu.cn>)

JBrowse 搭建基因组浏览器

拥有物种基因组和注释文件，如何直观展示？

2023.05.12 二综一楼C102 15:00 欢迎大家交流学习！

主讲人：张雨露

2023/05/12

为什么要使用 JBrowse ?

rapdb.dna.affrc.go.jp/jbrowse/?data=data%2Ffirgsp1&loc=chr01%3A1..17480&tracks=IRGSP-1.0%2CPredicted%20genes%2CRepresentative%20transcripts%2CMSU%20Osa1%20Rice%20Gene%20Models%2CRepeat%20regions%2C IsoSeq_Leaf_GC&hi...

rap-db The Rice Annotation Project Database

Home · JBrowse

Available Tracks

- filter tracks
- Reference sequence 1
- RAP annotation 4
 - select all from category
 - Gene locus
 - Predicted locus
 - Predicted genes
 - Representative transcripts
- Other rice annotations 4
- Protein alignments 11
- IsoSeq 1
 - select all from category
 - IsoSeq, Leaf, GC
- RNA-Seq 590
- ChIP-Seq 14
- ATAC-Seq 3
- Transcript mapping 6
- RNAs 6
- Masked regions 3
- Markers 1
- OMAP 12
- FSTs 8
- Gene predictions 1
- Organelle 4
- Misc 2

Genome Track View Help

IRGSP-1.0

chr01 chr01:1..17480 (17.48 Kb)

IRGSP-1.0

Predicted genes

Representative transcripts

MSU Osa1 Rice Gene Models

Repeat regions

IsoSeq, Leaf, GC

Os01t0100100-01 RabGAP/TBC domain containing protein.

Os01t0100200-01 Conserved hypothetical protein.

Os01t0100300-00 Cytochrome P450 domain containing protein.

Os01t0100400-01 Similar to Pectinesterase-like protein.

Os01t0100466-00 Hypothetical protein.

LOC_Os01g01010.1

LOC_Os01g01010.2

LOC_Os01g01019.1

LOC_Os01g01030.1

LOC_Os01g01040.1

LOC_Os01g01040.4

LOC_Os01g01040.2

LOC_Os01g01040.3

PB.1.1

PB.1.2

PB.1.3

PB.2.1

PB.2.2

PB.2.3

PB.2.8

PB.2.7

为什么要使用 JBrowse ?

在日常数据处理和分析工作中，根据分析项目的不同我们会面对各种各样的文件，比如mapping得到的SAM或者BAM文件，亦或记录变异信息的VCF文件等。那么，跑了pipeline以后拿到各种文件，如何形象方便地利用这些文件进行后续分析或者向合作者展示信息呢？IGV？

- 1) 程序响应速度慢：若要加载BAM文件、搜具体基因位置等，速度较慢。
- 2) 结果不方便分享：有时候合作者想看十几个基因的表达情况，不得不发一堆截图过去，说不定对方什么时候又想看什么，还得让自己一直处于待机状态。

— 目录 —

01

JBrowse 基因组浏览器特点

02

JBrowse 安装及配置

03

JBrowse 具体使用场景及操作示例

1.JBrowse 基因组 浏览器特点

1.所谓基因浏览器，首先可以把它理解为一个网站（网页），只不过使用它不再需要从HTML和CSS写起，也不用JS来编写各种function。其次它的功能是用来展示各种基因信息，比如基因的外显子内含子，比如转座子和重复序列或者SNP等。这样一来，好处之一是数据呈现的压力从本机转移到了后端服务器；好处之二是其本身的网站属性让你我都可以随时访问查看。

2.完全基于HTML5和Javascript构建的JBrowse通过AJAX技术实现了数据的异步加载，所以响应速度非常快。由于Javascript将大量的计算工作在前端完成，服务器端只需要向浏览器客户端发送静态文件，因此也极大程度减轻了服务器端的负担。

1. JBrowse 基因组浏览器特点

- 1.快速浏览和缩放。可轻松扩展到数十亿碱基的基因组和深度覆盖测序。
- 2.支持GFF3、BED、FASTA、BigWig、BAM、VCF（带tabix）、等。BAM、BigWig和VCF数据直接从压缩的二进制文件中显示，无需转换。
- 3.非常轻的服务器资源需求。事实上，JBrowse没有后端服务器代码，只有用于格式化数据文件以直接通过HTTP读取的工具。
- 4.另外，JBrowse非常容易上手。按照官方提供的Quick-Start Tutorial，只需要十几分钟时间就可以完成基本配置呈现出想要展示的数据。
- 5.随着各种基因组序列以及高通量测序数据爆炸性增长，可以满足基因组可视化、大规模基因组数据分析和应用需要。

2.JBrowse 安装及配置

Download a JBrowse release from GitHub

```
curl -L -O https://github.com/GMOD/jbrowse/releases/download/1.16.11-release/JBrowse-1.16.11.zip
unzip JBrowse-1.16.11.zip
sudo mv JBrowse-1.16.11 /var/www/html/jbrowse
cd /var/www/html
sudo chown `whoami` jbrowse
cd jbrowse
./setup.sh
```

2.JBrowse 安装及配置

Nginx

```
location ^~ /MH63RS3/JBrowse {  
    alias    /home/labuser/ylzhang/jbrowse_MH63RS3/;  
    index   index.html index.htm index.jsp;  
}
```

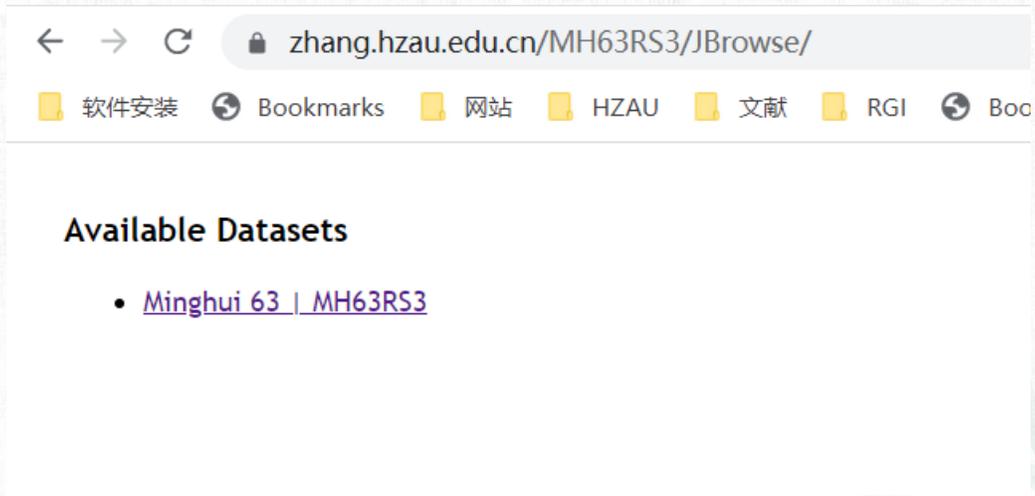
<https://zhang.hzau.edu.cn/MH63RS3/JBrowse>

3.JBrowse 具体使用 场景及操作示例

1) .fasta 及 .gff文件

```
[ylzhang@localhost ~]$ cd jbrowse_MH63RS3/
[ylzhang@localhost jbrowse_MH63RS3]$ ls
bin                extlib            jbrowse_conf.json  package-lock.json  src
browser           genomeList.html  LICENSE            plugins             tests
browserconfig.xml GenomeMenu       Makefile           raw_data           utils
build             index.html       Makefile.PL       README.md          webpack.config.js
CODE_OF_CONDUCT.md index.html.bak   MYMETA.json       release-notes.md   website
css              index.html.test  MYMETA.yml        sample_data        yarn.lock
data             index.html.test  node_modules      setup.log
dist            INSTALL          npm-debug.log     setup.sh
docs            jbrowse.conf    package.json      site.webmanifest
```

```
[datasets.MH63RS3]
url = ?data=data/MH63RS3
name = Minghui 63 | MH63RS3
```



3. JBrowse 具体使用 场景及操作示例

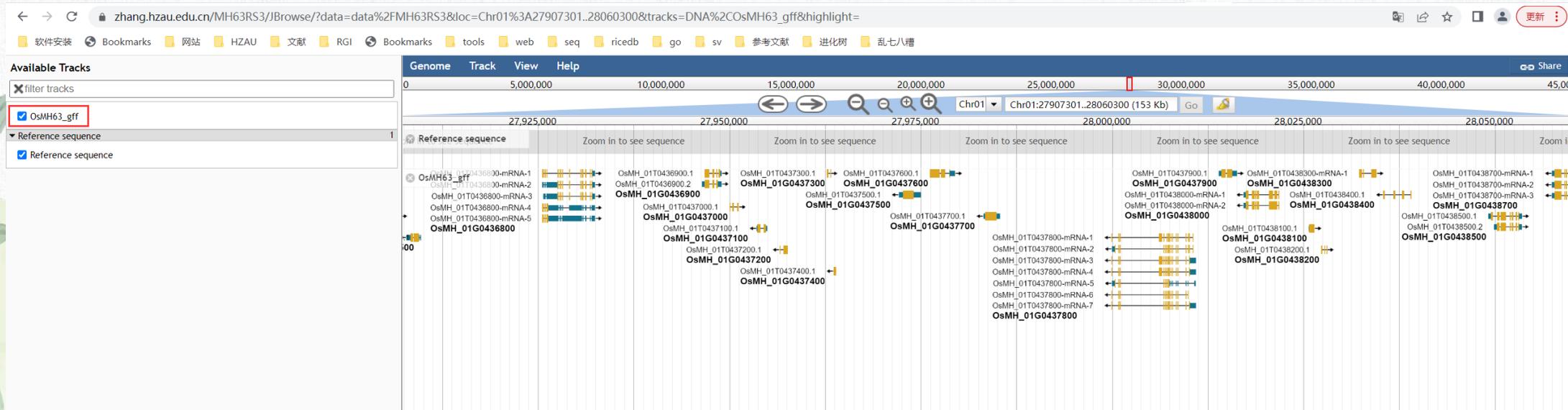
1) .fasta 及 .gff文件

```
[ylzhang@localhost jbrowse_MH63RS3]$ cd raw_data/  
[ylzhang@localhost raw_data]$ ls  
MH63RS3.fasta MH63RS3.gff3
```

```
[ylzhang@localhost jbrowse_MH63RS3]$ cd data/  
[ylzhang@localhost data]$ ls  
MH63RS3
```

```
[ylzhang@localhost jbrowse_MH63RS3]$ bin/prepare-refseqs.pl --fasta ./raw_data/MH63RS3.fasta --  
out ./data/MH63RS3/
```

```
[ylzhang@localhost jbrowse_MH63RS3]$ bin/flatfile-to-json.pl --gff ./raw_data/MH63RS3.gff3 --trackType  
CanvasFeatures --trackLabel OsMH63_gff --out ./data/MH63RS3/
```

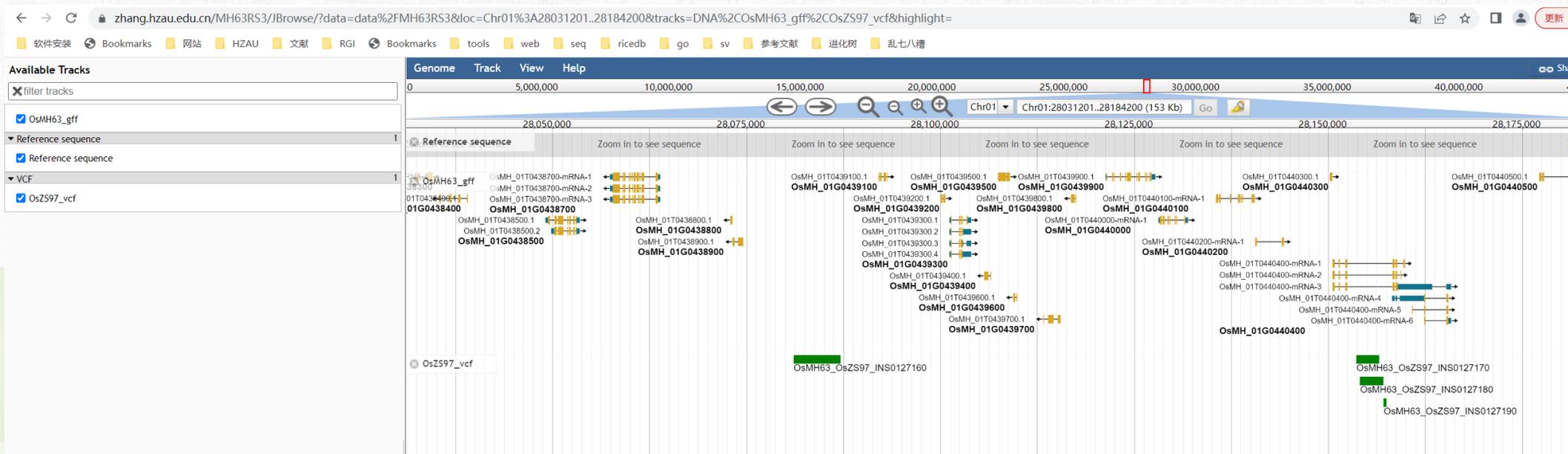


3. JBrowse 具体使用 场景及操作示例

2) vcf文件

```
[ylzhang@localhost raw_data]$ ls
MH63RS3.fasta MH63RS3.gff3 OsZS97.vcf
[ylzhang@localhost raw_data]$ bgzip OsZS97.vcf
[ylzhang@localhost raw_data]$ tabix -p vcf OsZS97.vcf.gz
[ylzhang@localhost raw_data]$ ls
MH63RS3.fasta MH63RS3.gff3 OsZS97.vcf.gz OsZS97.vcf.gz.tbi
```

```
[ylzhang@localhost MH63RS3]$ cat tracks.conf
[tracks.OsZS97_vcf]
storeClass=JBrowse/Store/SeqFeature/VCFTabix
urlTemplate=../../raw_data/OsZS97.vcf.gz
type=JBrowse/View/Track/CanvasVariants
key=OsZS97_vcf
```



3.JBrowse 具体使用 场景及操作示例

4) 分组

tracks.conf:

```
[tracks.OsZS97_vcf]
storeClass=JBrowse/Store/SeqFeature/VCFTabix
urlTemplate=../../raw_data/OsZS97.vcf.gz
type=JBrowse/View/Track/CanvasVariants
key=OsZS97_vcf
category = VCF
```

```
[tracks.MH63_seedling_totalRNA_rep1]
urlTemplate=../../raw_data/13248071.sorted.cram
key = MH63_seedling_totalRNA_rep1
category = RNA_seq
```

```
[ylzhang@localhost MH63RS3]$ cat trackList.json
{
  "compress" : 0,
  "key" : "OsMH63_gff",
  "label" : "OsMH63_gff",
  "storeClass" : "JBrowse/Store/SeqFeature/NCList",
  "style" : {
    "className" : "feature"
  },
  "trackType" : "CanvasFeatures",
  "type" : "CanvasFeatures",
  "urlTemplate" :
  "tracks/OsMH63_gff/{refseq}/trackData.json",
  "metadata":{"category" : "GFF"}
}
]
```


3.JBrowse 具体使用 场景及操作示例

5) 建立索引方便搜索

The screenshot displays the JBrowse web interface in a browser window. The address bar shows the URL: `zhang.hzau.edu.cn/MH63RS3/JBrowse/?data=data%2FMH63RS3&loc=Chr01%3A27031271..27193480&tracks=DNA%2COsMH63_gff%2COsZS97_vcf%2CMH63_seedling_totalRNA_rep1&highlight=`. The interface includes a navigation bar with 'Genome', 'Track', 'View', and 'Help' menus. A search bar at the top right contains 'Chr01' and 'Chr01:27031271..27193480 (162.21 Kb) Go'. The main view shows a genomic track with a reference sequence and various gene models (e.g., OsMH_01G0421800.1, OsMH_01G0422000.1, etc.) represented by colored bars and arrows. The left sidebar lists 'Available Tracks' with checkboxes for 'MH63_seedling_totalRNA_rep1', 'OsMH63_gff', 'OsZS97_vcf', and 'Reference sequence'.

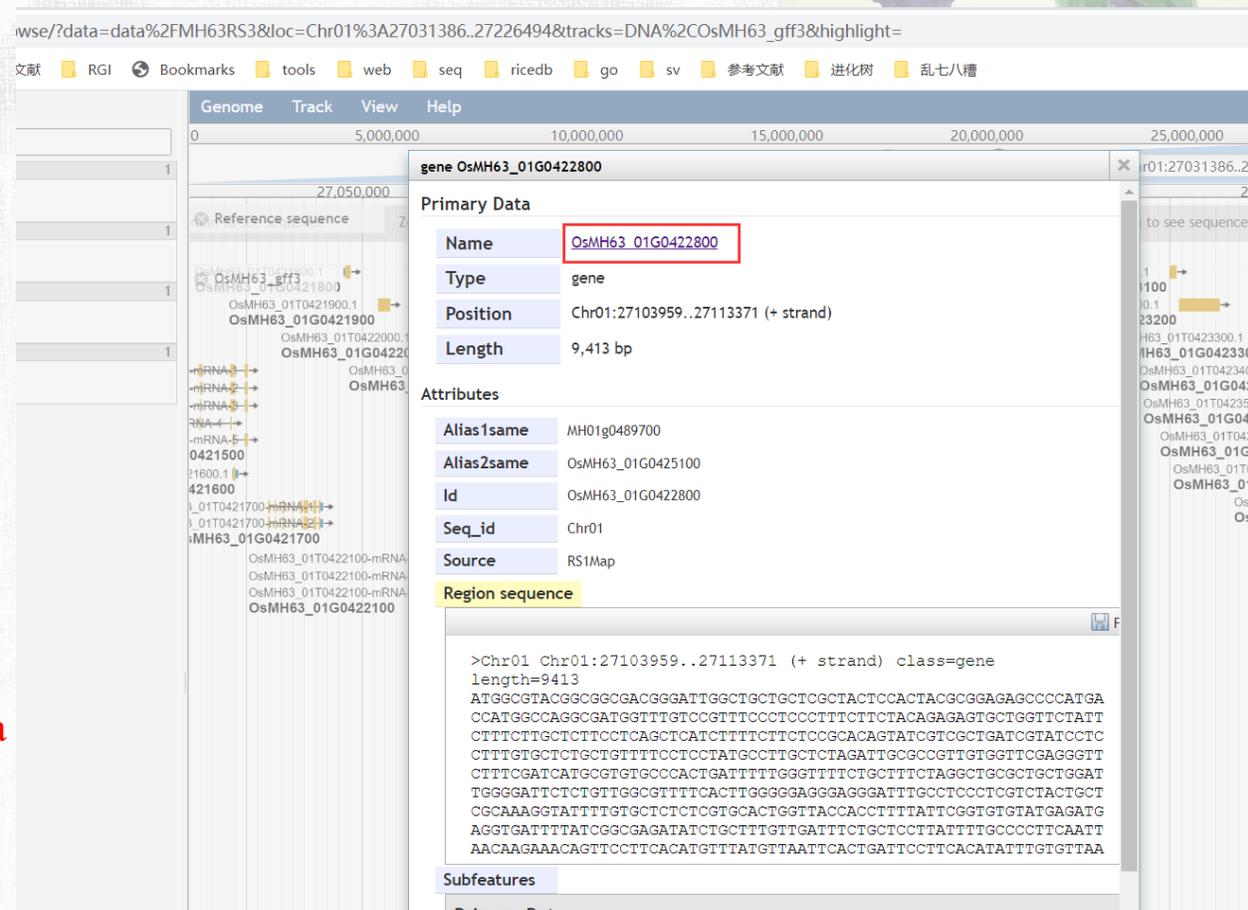
```
[ylzhang@localhost jbrowse_MH63RS3]$ bin/generate-names.pl -v --out ./data/MH63RS3
```

3. JBrowse 具体使用 场景及操作示例

6) 加超链接

trackList.json

```
{  
  "compress" : 0,  
  "key" : "OsMH63_gff3",  
  "label" : "OsMH63_gff3",  
  "storeClass" : "JBrowse/Store/SeqFeature/NCLList",  
  "style" : {  
    "className" : "feature"  
  },  
  "trackType" : "CanvasFeatures",  
  "type" : "CanvasFeatures",  
  "urlTemplate" :  
  "tracks/OsMH63_gff3/{refseq}/trackData.json",  
  "fmtDetailValue_Name": "function(name) { return '<a  
  href=\"https://riceome.hzau.edu.cn/genecard/'+name+'\",  
  target=\"_blank\">'+name+'</a>; }",  
  "metadata": {"category" : "GFF"}  
}
```



谢谢观看

— Small fresh work report —

汇报人：张雨露 时间：2023.05.12