

3.2 Genome Browser homework

徐英泽 2024011267

在 mapping_homework 中, 我们通过两种方法 mem 指令与 alt+samse 指令获得了两个 sam 文件, 分别是 THA2-bwa.sam 与 THA2-bwa_new.sam。

```
root@bioinfo_docker:/home/test/mapping/yeast_sacCer3# grep -v "^@" THA2-bwa.sam | awk
'$3 != "*" | cut -f 3 | sort | uniq -c
  7 chrII
  6 chrIII
 11 chrIV
  3 chrIX
  1 chrM
  1 chrV
  2 chrVII
  7 chrVIII
  7 chrX
  5 chrXI
 12 chrXII
  9 chrXIV
  9 chrXV
  1 chrXVI
# 输出最终sam文件部分结果与统计情况
```

图 1. bwa mem 指令获得的 THA2-bwa.sam 统计结果。

```
root@bioinfo_docker:/home/test/mapping/yeast_sacCer3# grep -v "^@" THA2-bwa_new.sam |
awk '$3 != "*" | cut -f 3 | sort | uniq -c
 17 chrI
 54 chrII
 17 chrIII
202 chrIV
 26 chrIX
 18 chrM
 38 chrV
 18 chrVI
129 chrVII
 70 chrVIII
 77 chrX
 60 chrXI
178 chrXII
 72 chrXIII
 59 chrXIV
108 chrXV
 83 chrXVI
# 输出最终sam文件部分结果与统计情况
```

图 2. bwa alt + bwa samse 指令获得的 THA2-bwa_new.sorted.sam 统计结果。

将这两个文件全部导入 IGV 中, 并且在 IGV 中下载 *S. cerevisiae* (sacCer3) 数据库。注意, 在上一个作业中获得的 *S. cerevisiae* (sacCer3) fasta 文件可以导入 IGV, 但是由于 fasta 格式无

基因的注释信息，故不使用 fasta 格式文件导入。

需要注意的是，THA2-bwa_new.sam 文件内部由于未将染色体信息正确排序，使用

```
`samtools sort -O sam THA2-bwa_new.sam -o THA2-bwa_new.sorted.sam`
```

代码，排序，并获得可以导入 IGV 的 THA2-bwa_new.sorted.sam 文件

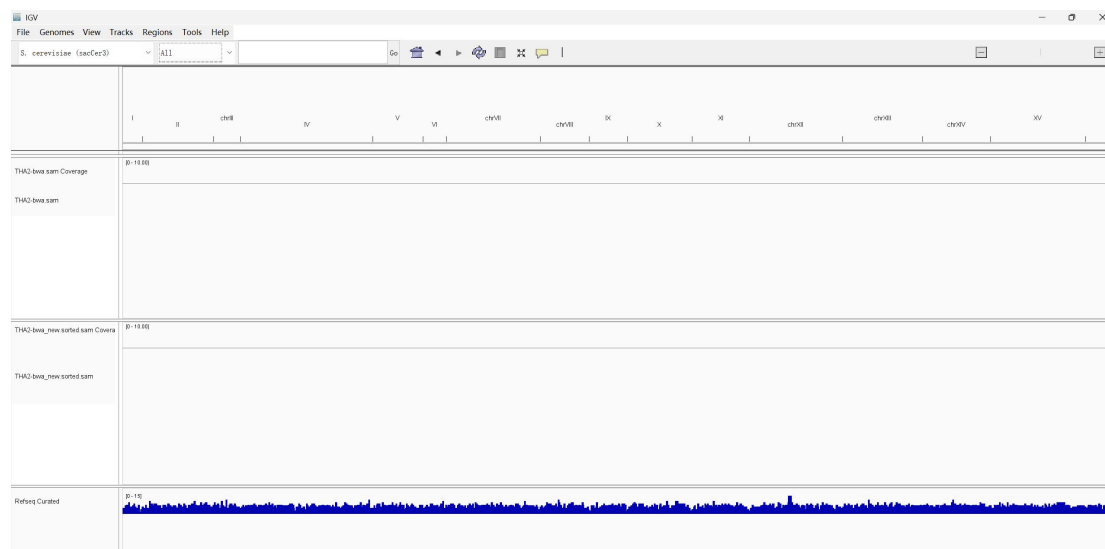


图 3. IGV 中导入文件，从上至下共 3 个 track，分别为 THA2-bwa.sam、THA2-bwa_new.sorted.sam 与基因组文件。

寻找有意思的区域，选取 chrM:46,109-47,107 区域进行查看。

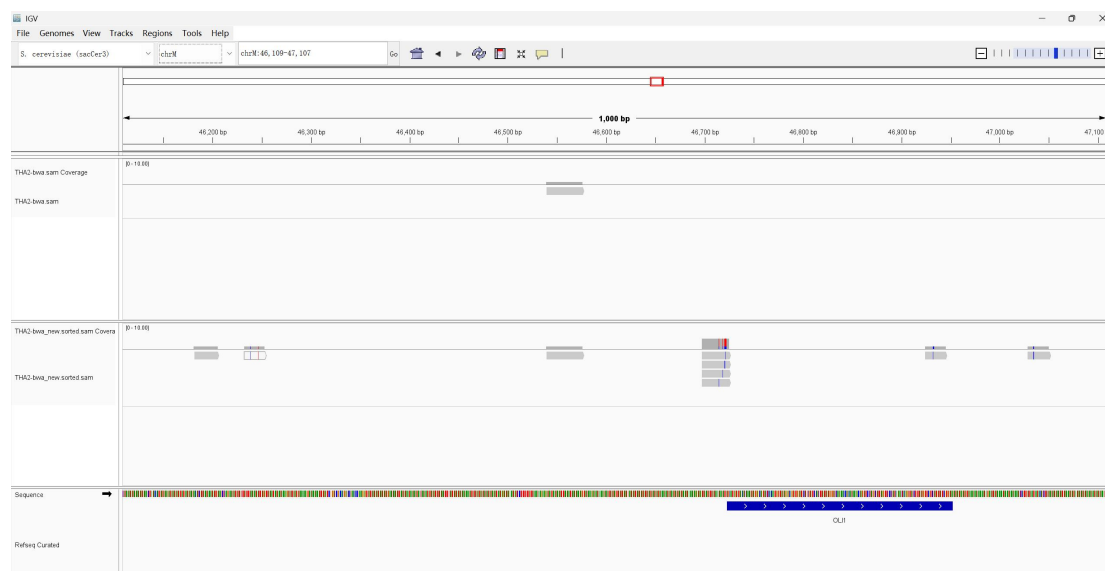


图 4. chrM:46,109-47,107 区域比对结果与基因信息。

直观上来说，THA2-bwa.sam 中比对的结果远少于 THA2-bwa_new.sorted.sam，符合统计结果。

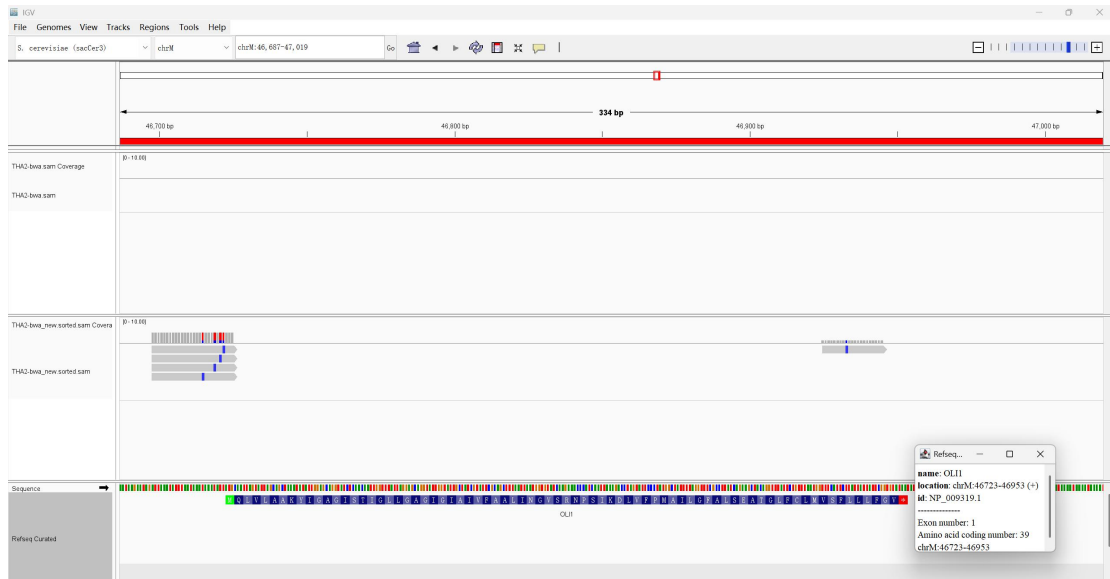


图 5. 进一步查看 OLI1 的信息。

THA2-bwa_new.sorted.sam 在 OLI1 基因附近成功比对到了多条序列，图中可见序列信息以及深度。