

1.写质控+过滤运行脚本

vim run_fastp_ZB1.sh

----->编辑

```
#!/bin/bash
```

```
#SBATCH --partition=hpxg
```

```
#SBATCH --nodes=1
```

```
fastp -h 路径/文件名.html -j 路径/文件名.json -i 输入文件的路径/文件名.R1.raw.fastq.gz -l 输入文件的路径/文件名.R2.raw.fastq.gz -o 输出文件路径/文件名.R1.clean.fq.gz -O 输出文件路径/文件名.R2.clean.fq.gz
```

(保存、退出)

2.提交任务

```
sbatch run_fastp_ZB1.sh
```

查看任务运行状态: squeue

取消任务: scancel 任务号

3.组装叶绿体基因组运行脚本

vim assemble_chloroplast_ZB1.sh

----->编辑

```
#!/bin/bash
```

```
#SBATCH --partition=hpxg
```

```
#SBATCH --nodes=1
```

```
get_organelle_from_reads.py -1 /project/weipei/chloroplast_clean_data_gxm/ZB1_R1.clean.fq.gz -2 /project/weipei/chloroplast_clean_data_gxm/ZB1_R2.clean.fq.gz -o /project/weipei/assemble_chloroplast/ZB1(输出文件存放位置, 前缀prefix) -R 15 -k 21,45,65,85,105 -F embplant_pt (固定参数, 不调整)
```

(保存、退出)

备注: embplant_pt是一个数据库, 必须有的。如果没有, 则需要自己手动安装。安装如下:

```
get_organelle_config.py -a embplant_pt --use-local pt数据库文件夹位置/数据库
```

```
get_organelle_config.py -a embplant_mt --use-local mt数据库文件夹位置/数据库
```

=====以下为国重服务器运行脚本=====

```
#!/bin/bash
```

```
#PBS -N YBJ (取个任务名)
```

```
#PBS -l nodes=1:ppn=60 (线程数: 1~64个可选)
```

```
#PBS -l walltime=9999:00:00
```

```
get_organelle_from_reads.py -1 /home/lx_sky6/weipei/yelvti/new_add_7_chloroplast_cleandata/ACX_1_1_R1.clean.fq.gz -2 /home/lx_sky6/weipei/yelvti/new_add_7_chloroplast_cleandata/ACX_1_1_R2.clean.fq.gz -o /home/lx_sky6/weipei/yelvti/add_7_1.assemble_chloroplastgenome/ACX_1_1 -R 15 -k 21,45,65,85,105 -F embplant_pt
```

提交任务: qusb XXXX.sh

查看任务: qstat

取消任务: qdel **38503**.node1(红色改为自己编号)

=====

以下为get_organelle组装叶绿体运行结果文件内的内容

```
-rw-rw-r-- 1 lx_sky6 lx_sky6 158355 Sep 28 20:54 embplant_pt.K105.complete.graph1.1.path_sequence.fasta (这两个可能是正义链、反义链)
-rw-rw-r-- 1 lx_sky6 lx_sky6 158355 Sep 28 20:54 embplant_pt.K105.complete.graph1.2.path_sequence.fasta (这两个可能是正义链、反义链)
-rw-rw-r-- 1 lx_sky6 lx_sky6 131088 Sep 28 20:54 embplant_pt.K105.complete.graph1.selected_graph.gfa
-rw-rw-r-- 1 lx_sky6 lx_sky6 52832526 Sep 28 20:50 extended_1_paired.fq
-rw-rw-r-- 1 lx_sky6 lx_sky6 2932219 Sep 28 20:50 extended_1_unpaired.fq
-rw-rw-r-- 1 lx_sky6 lx_sky6 52832778 Sep 28 20:50 extended_2_paired.fq
-rw-rw-r-- 1 lx_sky6 lx_sky6 2193138 Sep 28 20:50 extended_2_unpaired.fq
-rw-rw-r-- 1 lx_sky6 lx_sky6 348354 Sep 28 20:54 extended_K105.assembly_graph.fastg
-rw-rw-r-- 1 lx_sky6 lx_sky6 5463 Sep 28 20:54 extended_K105.assembly_graph.fastg.extend-embplant_pt-embplant_mt.csv
-rw-rw-r-- 1 lx_sky6 lx_sky6 323924 Sep 28 20:54 extended_K105.assembly_graph.fastg.extend-embplant_pt-embplant_mt.fastg
drwxrwxr-x 10 lx_sky6 lx_sky6 4096 Sep 28 20:54 extended_spades
-rw-rw-r-- 1 lx_sky6 lx_sky6 9446 Sep 28 20:54 get_org.log.txt
drwxrwxr-x 2 lx_sky6 lx_sky6 4096 Sep 28 20:13 seed
```

↓

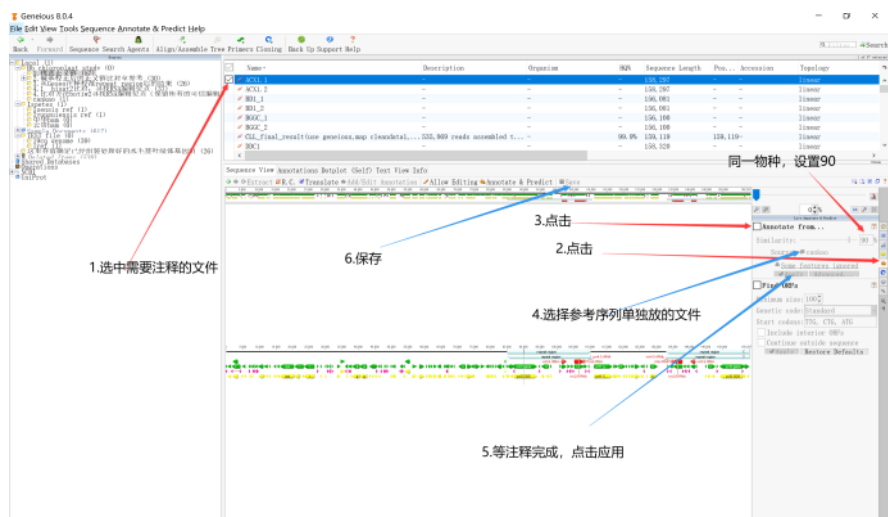
下载这两个文件

↓

=====

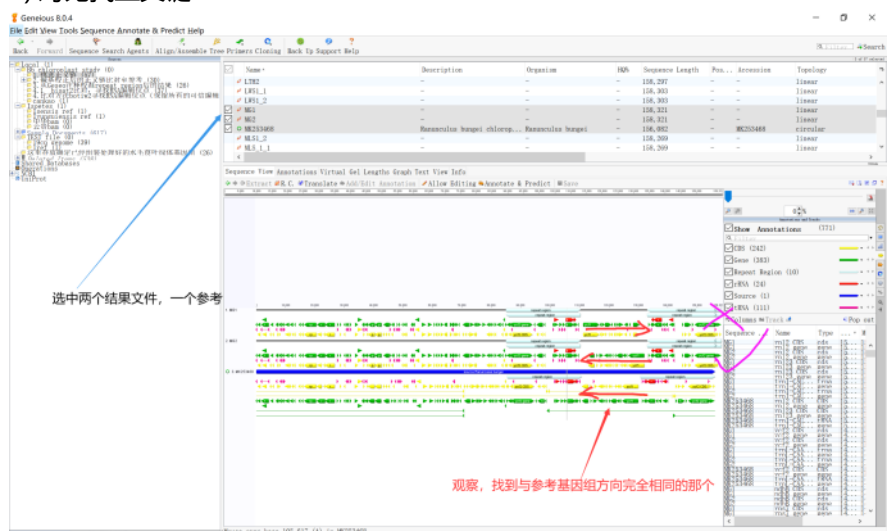
与参考序列对比, **挑选出两个链中的正义链**

1) .结果文件注释: 首先把组装生成的两个文件拖入Geneious



2). 把另一个文件也注释, 按照上述方法

3). 对比找正义链



碱基校正 (需要修改的参数标记为非蓝色)

```
#!/bin/bash
#PBS -N acx2
#PBS -l nodes=1:ppn=60
#PBS -l walltime=1200:00:00
```

`cd /home/lx_sky6/weipei/yelvti/add_7_1.assemble_chloroplastgenome/mytest` (进入存放正义链的文件夹, 路径更换为自己的)

`bwa index (构建索引) /home/lx_sky6/weipei/yelvti/add_7_`

`1.assemble_chloroplastgenome/mytest/embplant_pt.K105.complete.graph1.2.path_sequence.fasta` (正义链, 路径更换为自己的)

```
bwa mem -t 4 /home/lx_sky6/weipei/yelvti/add_7_1.assemble_chloroplastgenome/mytest/embplant_pt.K105.complete.graph1.2.path_sequence.fasta (正义链, 路径更换为自己的) /home/lx_sky6/weipei/yelvti/new_add_7_chloroplast_cleandata/ACX_1_1_R1.clean.fq.gz (过滤+质控后的原始序列文件R1) /home/lx_sky6/weipei/yelvti/new_add_7_chloroplast_cleandata/ACX_1_1_R2.clean.fq.gz (过滤+质控后的原始序列文件R2) | samtools view -@ 4 -bS -F 12 | samtools sort -@ 4 > ./bwa.sort.bam
```

```
samtools index ./bwa.sort.bam
```

```
java -Xmx16G -jar /home/lx_sky6/weipei/yelvti/related_software/pilon-1.23.jar (软件包的路径) --genome /home/lx_sky6/weipei/yelvti/add_7_1.assemble_chloroplastgenome/mytest/embplant_pt.K105.complete.graph1.2.path_sequence.fasta (正义链, 路径更换为自己的) --frags ./bwa.sort.bam --output ./ACX_1_2_final.pilon1 (校正后结果取名, 必须以.pilon1为后缀) --changes --threads 4
```

=====

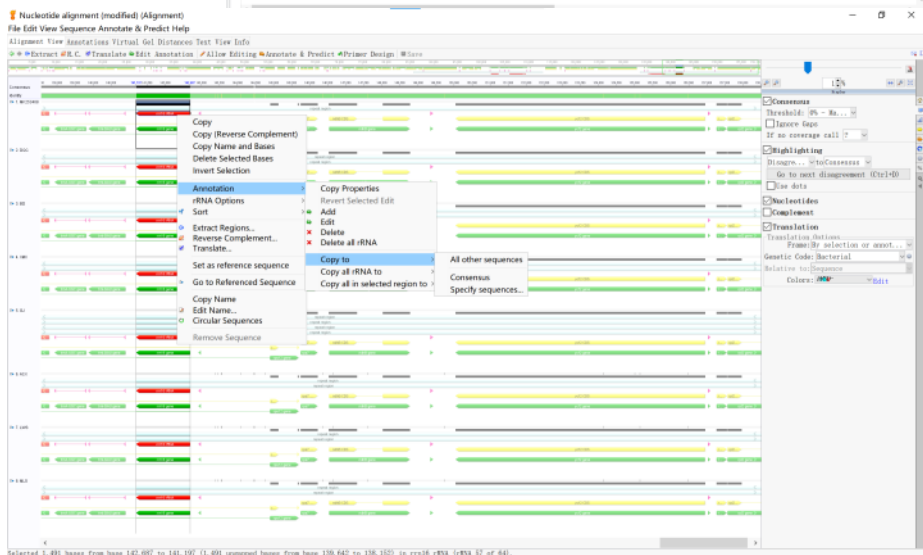
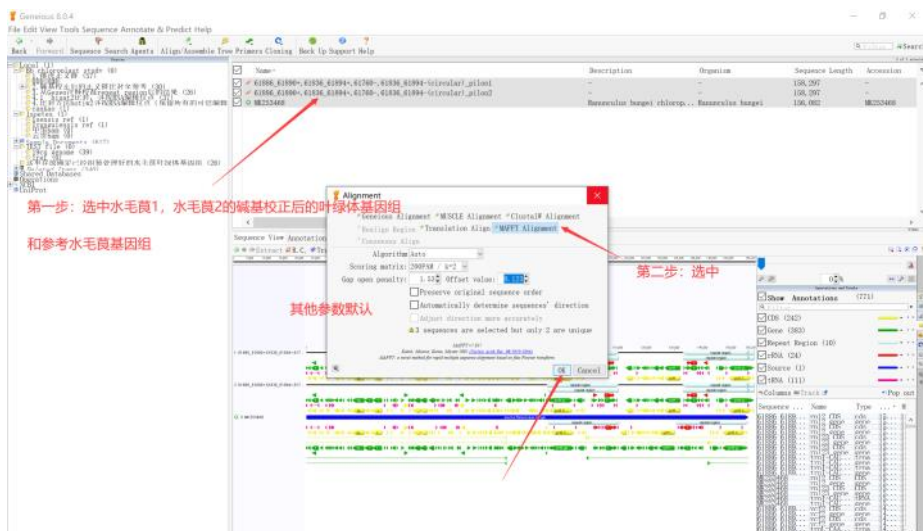
碱基校正后的文件夹如下

```
-rw-rw-r-- 1 lx_sky6 lx_sky6      0 Sep 28 21:44 ACX_1_2_final.pilon1.changes (看他是不是0)
-rw-rw-r-- 1 lx_sky6 lx_sky6 160339 Sep 28 21:44 ACX_1_2_final.pilon1.fasta (如果上面文件为0, 则用ACX_1_2_final.pilon1.fasta继续后续分析)
若不为0, 则把ACX_1_2_final.pilon1.fasta当做正义链, 重新运行碱基校正的脚本, 再跑一次/多次, 直到ACX_1_2_final.pilon1.changes 为0
```

```
-rw----- 1 lx_sky6 lx_sky6   60671 Sep 28 21:44 acx2.e38537
-rw----- 1 lx_sky6 lx_sky6   1162 Sep 28 21:44 acx2.o38537
-rw-rw-r-- 1 lx_sky6 lx_sky6    985 Sep 28 21:37 base_correction.sh
-rw-rw-r-- 1 lx_sky6 lx_sky6 16610321 Sep 28 21:44 bwa.sort.bam
-rw-rw-r-- 1 lx_sky6 lx_sky6    672 Sep 28 21:44 bwa.sort.bam.bai
-rw-rw-r-- 1 lx_sky6 lx_sky6 158355 Sep 28 20:54 embplant_pt.K105.complete.graph1.1.path_sequence.fasta
-rw-rw-r-- 1 lx_sky6 lx_sky6 158355 Sep 28 20:54 embplant_pt.K105.complete.graph1.2.path_sequence.fasta
-rw-rw-r-- 1 lx_sky6 lx_sky6    11 Sep 28 21:37 embplant_pt.K105.complete.graph1.2.path_sequence.fasta.amb
-rw-rw-r-- 1 lx_sky6 lx_sky6    88 Sep 28 21:37 embplant_pt.K105.complete.graph1.2.path_sequence.fasta.ann
-rw-rw-r-- 1 lx_sky6 lx_sky6 158392 Sep 28 21:37 embplant_pt.K105.complete.graph1.2.path_sequence.fasta.bwt
-rw-rw-r-- 1 lx_sky6 lx_sky6 39576 Sep 28 21:37 embplant_pt.K105.complete.graph1.2.path_sequence.fasta.pac
-rw-rw-r-- 1 lx_sky6 lx_sky6 79200 Sep 28 21:37 embplant_pt.K105.complete.graph1.2.path_sequence.fasta.sa
-rw-rw-r-- 1 lx_sky6 lx_sky6 131088 Sep 28 20:54 embplant_pt.K105.complete.graph1.selected_graph.gfa
-rw-rw-r-- 1 lx_sky6 lx_sky6 52832526 Sep 28 20:50 extended_1_paired.fq
-rw-rw-r-- 1 lx_sky6 lx_sky6 2932219 Sep 28 20:50 extended_1_unpaired.fq
-rw-rw-r-- 1 lx_sky6 lx_sky6 52832778 Sep 28 20:50 extended_2_paired.fq
-rw-rw-r-- 1 lx_sky6 lx_sky6 2193138 Sep 28 20:50 extended_2_unpaired.fq
-rw-rw-r-- 1 lx_sky6 lx_sky6 348354 Sep 28 20:54 extended_K105.assembly_graph.fastg
-rw-rw-r-- 1 lx_sky6 lx_sky6   5463 Sep 28 20:54 extended_K105.assembly_graph.fastg.extend-embplant_pt-embplant_mt.csv
-rw-rw-r-- 1 lx_sky6 lx_sky6 323924 Sep 28 20:54 extended_K105.assembly_graph.fastg.extend-embplant_pt-embplant_mt.fastg
drwxrwxr-x 10 lx_sky6 lx_sky6   4096 Sep 28 20:54 extended_spades
-rw-rw-r-- 1 lx_sky6 lx_sky6   9446 Sep 28 20:54 get_org.log.txt
drwxrwxr-x 2 lx_sky6 lx_sky6    4096 Sep 28 20:13 seed
```

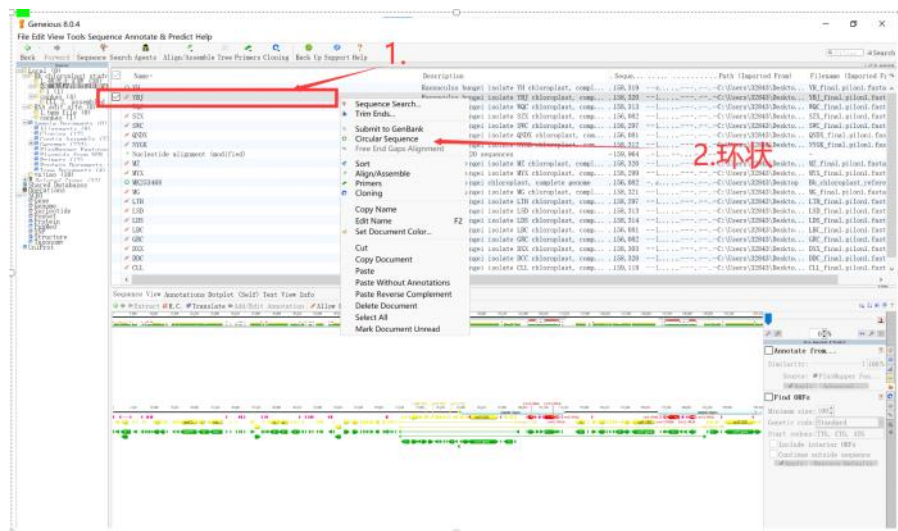
=====

手动校准自己的研究对象



调整完成后保存。

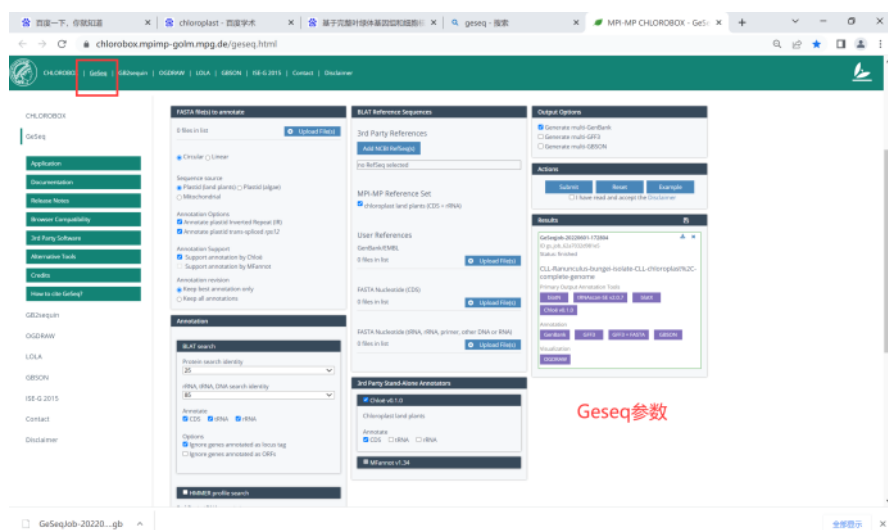
↓然后环状保存每一个文件



然后，将序列导出为.fasta格式

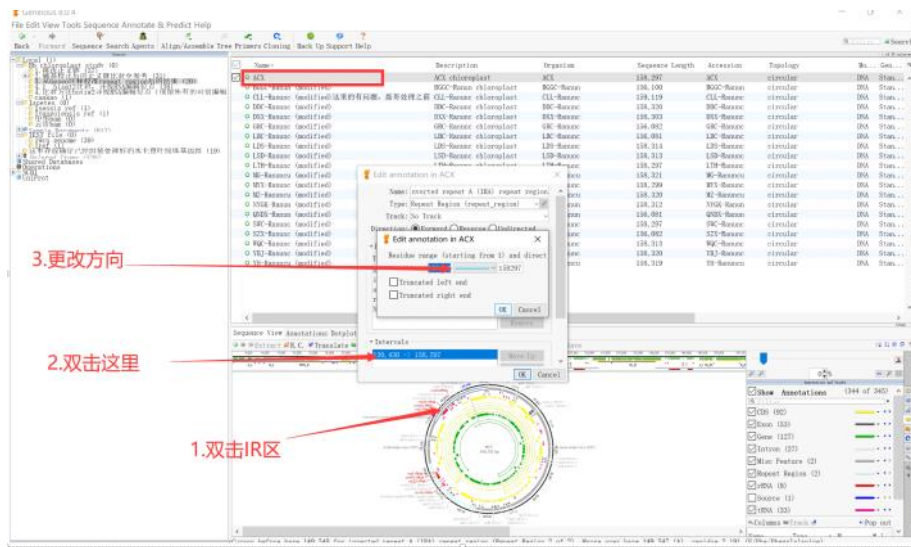
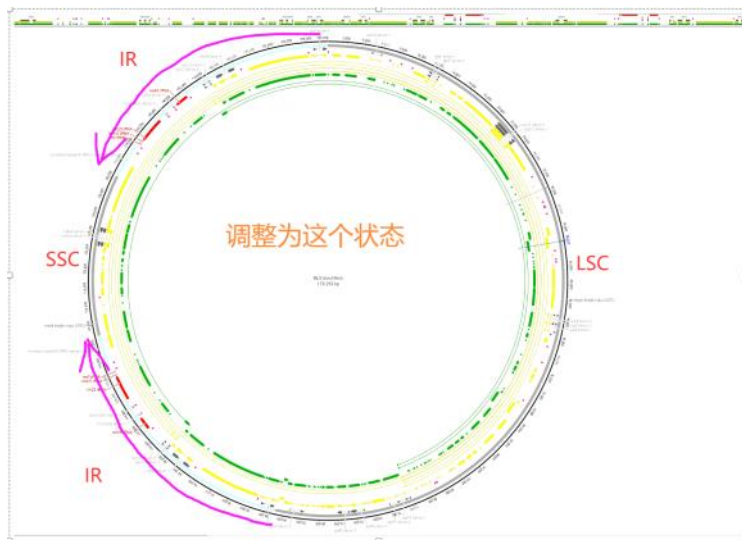
下一步：用线网站进行IR区的校正和进一步校准（参数勾选如下图）

<https://chlorobox.mpimp-golm.mpg.de/geseq.html>



(finished后下载保存结果文件)

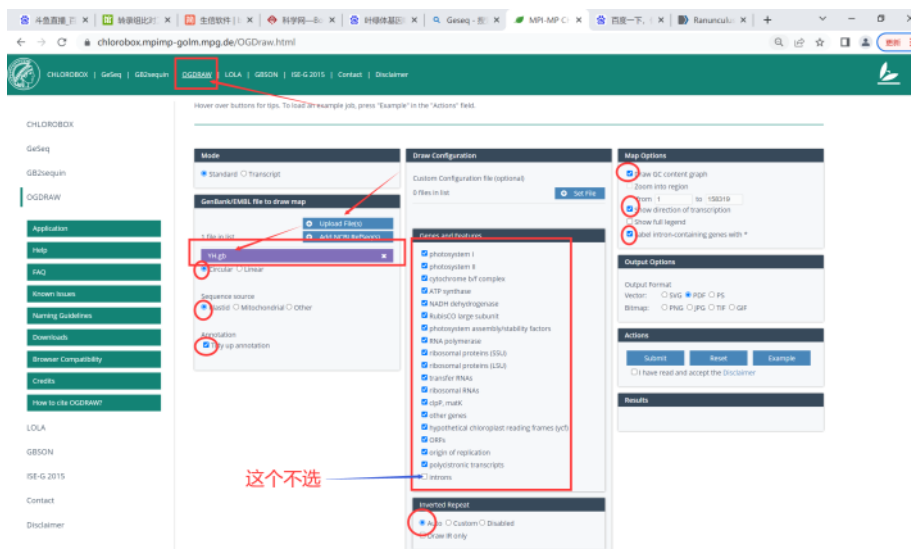
将结果文件重新导入 Geneious 中手动调整IR区方向



保存 ----->导出为 .gb格式

下一步：用线上网站进行叶绿体基因组草图绘制（参数如下）

<https://chlorobox.mpimp-golm.mpg.de/geseq.html>



下载结果。

