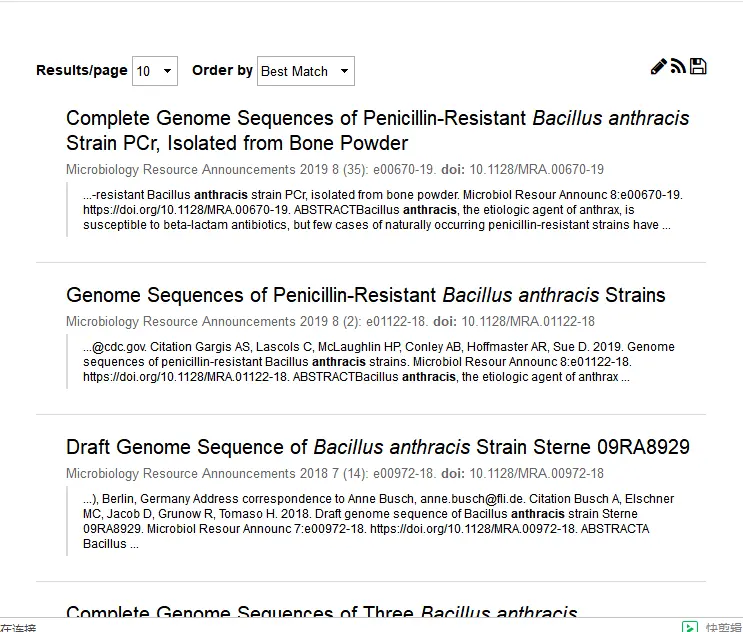
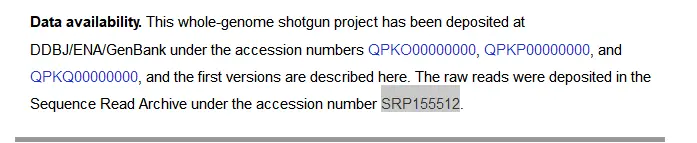
# **一.细菌基因组下载**

##### **1.登陆Genome Announcements网站（[https://mra.asm.org/](https://links.jianshu.com/go?to=https://mra.asm.org/" \t "https://www.jianshu.com/p/_blank)）找一篇细菌基因组文章；**

我输入的是anthracis,炭疽杆菌的拉丁名，得到

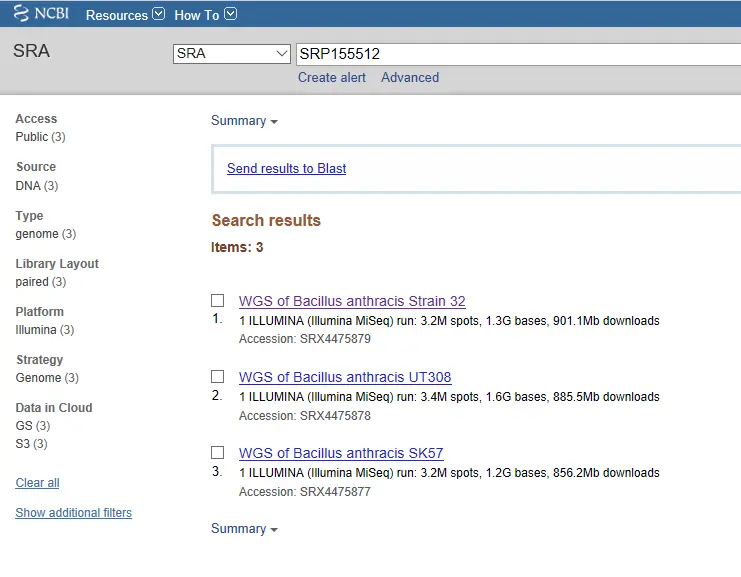


第二个是耐青霉素炭疽桿菌菌株的基因组序列，进去看看

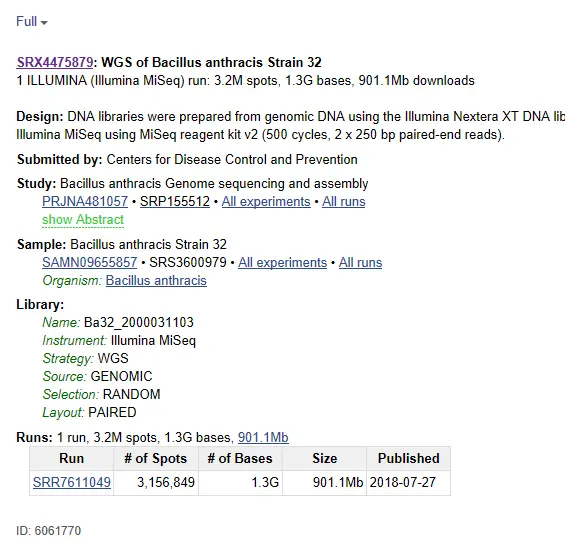


* 在这里，最后一句说这个序列的详细数据记录在SRP155512中

打开NCBI，点击SRA，在搜索框中输入SRP155512



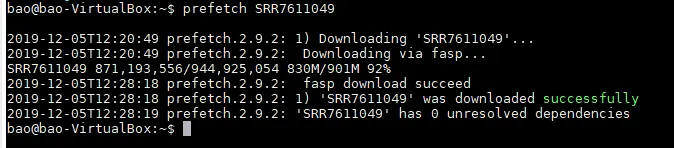
进入第一个，可看到SRA号，SRR7611049



* 然后打开虚拟机，用prefetch下载

prefetch SRR7611049

下载完成！



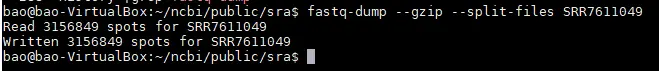
# **二.用Fastq-dump解压SRA文件为fastq**

* 下载得到的文件在~/ncbi/public/sra/SRR7611049下

cd ~/ncbi/public/sra

fastq-dump --gzip --split-files SRR7611049

解压完成！

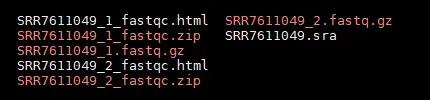


* 由于是双端测序，得到两个文件：  
  SRR7611049\_1.fastq.gz，  
  SRR7611049\_2.fastq.gz

# **三.用Fastqc进行质量控制**

fastqc SRR7611049\_1.fastq.gz SRR7611049\_2.fastq.gz

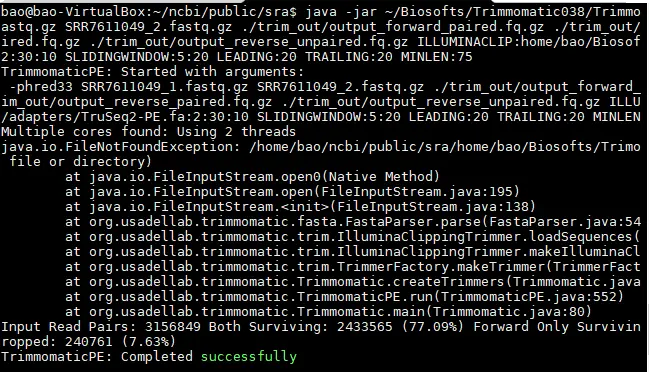
得到



# **四.Trimmomatic去接头**

java -jar ~/Biosofts/Trimmomatic038/Trimmomatic-0.38/trimmomatic-0.38.jar PE -phred33 SRR7611049\_1.fastq.gz SRR7611049\_2.fastq.gz ./trim\_out/output\_forward\_paired.fq.gz ./trim\_out/output\_forward\_unpaired.fq.gz ./trim\_out/output\_reverse\_paired.fq.gz ./trim\_out/output\_reverse\_unpaired.fq.gz ILLUMINACLIP:home/bao/Biosofts/Trimoomatic038/Trimmomatic-0.38/adapters/TruSeq2-PE.fa:2:30:10 SLIDINGWINDOW:5:20 LEADING:20 TRAILING:20 MINLEN:75

结果如图



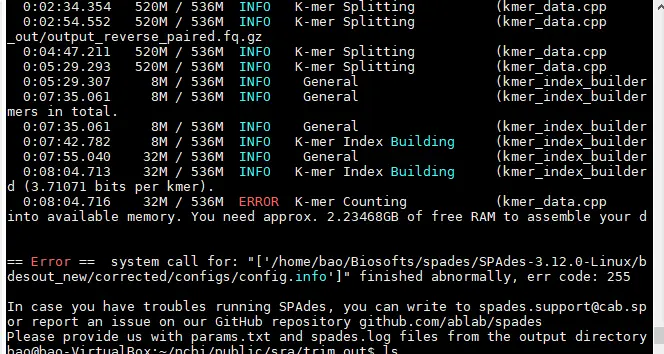
得到如下4个文件

IMG_264

# **五. Spades组装基因组草图**

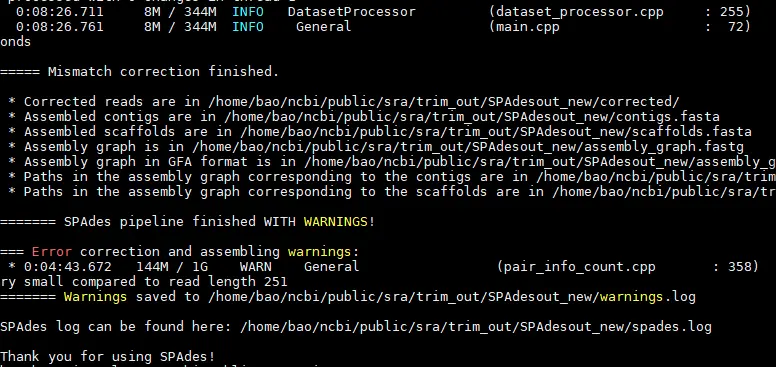
spades.py --careful --pe1-1 output\_forward\_paired.fq.gz --pe1-2 output\_reverse\_paired.fq.gz -o ./SPAdesout\_new

结果遇到了一点问题，报错了

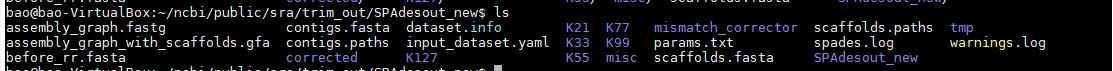


* 尝试退出虚拟机然后在设置中加虚拟机的内存
* 重新输入命令，这个过程等待了很长时间，应该是找的基因组太大的原因

等了俩小时，终于成功啦！



产生的文件如下

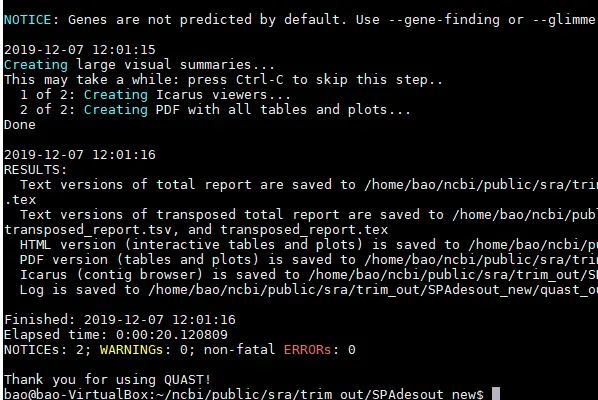


# **六. Quast评价组装的基因组效果**

* 输入

quast.py contigs.fasta -o quast\_out

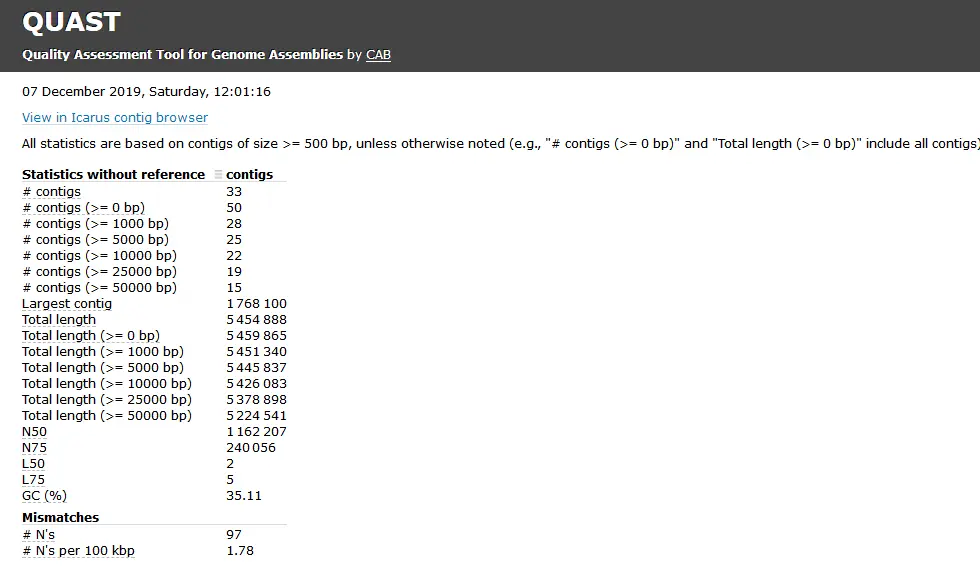
得到结果

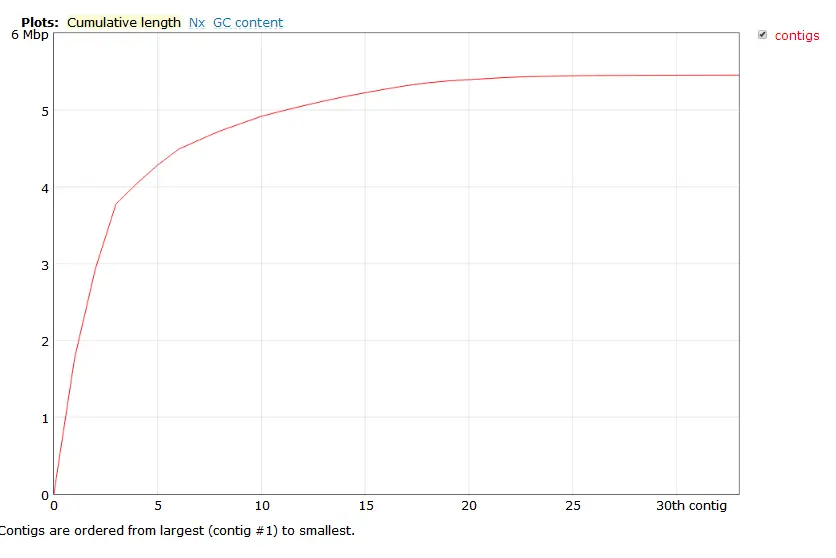


打开quast\_out 文件夹，得到

IMG_269

将report.html下载到windows上查看结果





最后编辑于 ：2023.01.12 19:26:31