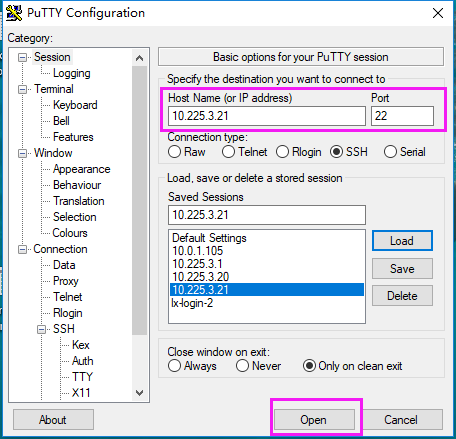
小RNA流程使用说明文档

1. 登录集群账号，登录节点10.225.3.21。

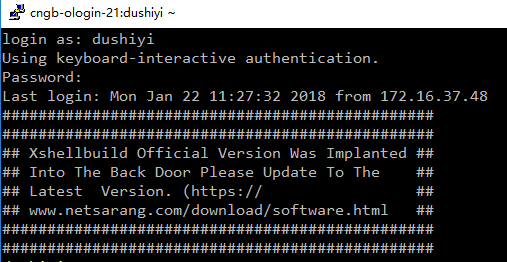
打开putty：在Host Name中输入10.225.3.21，点击Open。



输入用户名：XXX按回车键

输入集群密码：\*\*\*\*\*\*\* 按回车键

如果出现如下界面则表示登录成功。就可以进行下一步。x3Zr8zaMJJ



1. 准备工作：

进入 /ldfssz1/MGI\_BIT/RUO目录下，cd /ldfssz1/MGI\_BIT/RUO

建立自己的文件夹：mkdir yanjuan1

进入自己的文件夹：cd yangjuan1

新建项目文件夹：mkdir smRNA，文件夹一旦建立，下一次可直接使用，不用重复建立。

进入项目文件夹：cd smRNA,该目录作为存放该项目的总目录，在该目录下，再创建每一批次的分析目录：mkdir 23.CL100052781\_L01\_Y34

,为了方便查找和记录，分析目录可以按照这样的格式命名。

进入分析目录：cd 23.CL100052781\_L01\_Y34

将配置文件示例文件拷贝到该路径下：cp /hwfssz2/MGI\_BIT/RUO/PROJECT/P17Z18000N0042/01.smRNA\_new\_kit/23.CL100052781\_L01\_Y34/small\_rna.cf /hwfssz2/MGI\_BIT/RUO/PROJECT/P17Z18000N0042/01.smRNA\_new\_kit/23.CL100052781\_L01\_Y34/main.sh ./

3，修改配置文件

打开需要修改的配置文件：vi small\_rna.cf，按下键盘自己i，进入插入模式。此模式下才能对内容进行编辑，然后将光标移动到需要修改的地方：

每次需要修改最后的数据路径的行，按照q相应的格式将路径替换掉即可。

还需要修改S.REPORT 行，用新的命名替换掉旧的即可。



最后需要注意：S.DIFF行中的bar\_1和bar\_2必须是配置文件中有的barcode。



注意：在vi里面粘贴时，先将光标移动到需要粘贴的位置，在插入模式下（按i进入插入模式），按住shift点击鼠标右键完成粘贴，不能用ctrl+v的方法粘贴。

保存退出：

粘贴完成后，需要保存退出时，先按esc，退出插入模式，再按shift+：（按住shift时再按冒号） 再按wq 回车 就保存退出了，这是所有文件的保存退出方法。如果不想保存只退出就输入“q!”。

查找数据路径：

<http://10.225.3.21/cgi-bin/findRawDataPath/pathFinder.py>

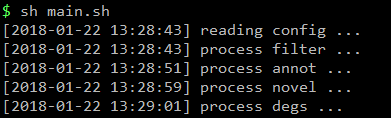
输入 lide号和lane号。复制数据路径

查找相关barcode fastq文件，使用如下命令：大括号里为本次需求的barcode号，粘贴数据路径替换下面的数据路径，（替换掉$i.fq.gz之前的的数据路径，并注意还要添加上slide和lane的格式，格式需要与下面的完全一致）

for i in {1,2,5,7,8,10};do echo "bar\_$i 5 2 /zfssz3/MGI\_NGS/Prod/BGISEQ-500/Zebra1001160211/CL100059898/L01/CL100059898\_L01\_$i.fq.gz";done

4，运行：

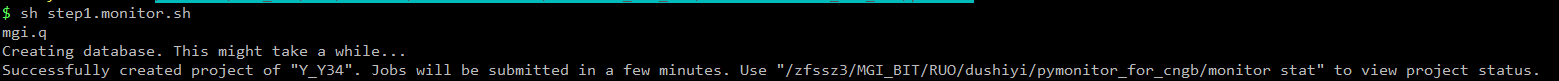
在工作目录下运行：sh main.sh



然后进入刚刚生成的process目录中：cd process 再运行：

sh step1.monitor.sh

出现Successfully则表示成功。



饱和度分析：

1. 进入到process目录下，创建饱和度分析的目录：
2. mkdir -p 01.saturation/15M 01.saturation/20M 01.saturation/25M 01.saturation/30M
3. 分别进入到这四个目录中cd 01.saturation
4. cd 15M
5. 拷贝 cp ../../../small\_rna.cf ../../../main.sh ./
6. 修改small\_rna.cf文件（使用vi打开）：vi small\_rna.cf 按i进入插入模式，将光标移动到第10行中的S.FILTER = -c 0处，并修改为S.FILTER = -c 15，同理，修改的数字表示截取的数据量。再修改S.REPORT = Y\_Y34为Y34\_15M,
7. 保存退出，再运行 sh main.sh
8. 再运行：cd process && sh step1.monitor.sh
9. cd ../.. 返回到饱和度分析目录
10. 重复第四步，使用cd命令进入20M，25M，30M目录下，重复以上步骤。即可完成。

饱和度分析方法二：

1. 进入到process 目录下，
2. /ldfssz1/MGI\_BIT/Public/Software/miniconda2/bin/python /zfssz3/MGI\_BIT/RUO/dushiyi/script/Program/03.forSmallRNA/smallRNA\_pipeline/bin/saturation/saturation.py -d ./ -c "15M 20M 25M 30M" 复制粘贴这段话将自动创建饱和度分析的目录和脚本，无需修改任何东西。
3. 然后进入cd 01.saturation，再进入每个目录，运行sh main.sh
4. 再运行：cd process && sh step1.monitor.sh

5，查看运行状态：

使用命令qstat查看任务运行结果，如果无任何内容显示，则表示运行结束。

使用/zfssz3/MGI\_BIT/RUO/dushiyi/pymonitor\_for\_cngb/monitor stat 即可查看任务的执行情况。

6，查看结果

结果放在分析目录下的process目录中，使用cd，进入该目录：

就能看到xls后缀的文件，使用filezilla下载即可查看。或者在集群中是用命令：

less output.miRNA.xls 即可查看内容。

结果有三个：output.filter.xls，output.miRNA.xls，output.mapping.xls

7，其他物种：

如果需要做其他物种的分析，需要将DATABASE = /zfssz3/MGI\_BIT/RUO/dushiyi/script/Program/03.forSmallRNA/database\_for\_2017a/hg19\_UCSC这一行替换成其他物种的database即可。

拟南芥：/ldfssz1/MGI\_BIT/RUO/dushiyi/scripts/Program/03.forSmallRNA/database\_for\_2017a/Arabidopsis

水稻：/ldfssz1/MGI\_BIT/RUO/dushiyi/scripts/Program/03.forSmallRNA/database\_for\_2017a/Oryza\_sativa

小鼠：/ldfssz1/MGI\_BIT/RUO/dushiyi/scripts/Program/03.forSmallRNA/database\_for\_2017a/Mus\_musculus

7，补充：

统计文库信息时可以用脚本：

python /ldfssz1/MGI\_BIT/RUO/dushiyi/scripts/Program/03.forSmallRNA/for\_qc.py

将路径跟在后面即可。Example：

python /ldfssz1/MGI\_BIT/RUO/dushiyi/scripts/Program/03.forSmallRNA/for\_qc.py /zfssz3/MGI\_NGS/Prod/BGISEQ-500/1001150023/CL100056649/L01