三代扩增子流程使用说明

|  |  |  |
| --- | --- | --- |
| 编制人：涂成芳 | 审核人：张阳 | 批准人： |
| 编制日期：20230602 | 审核日期： | 批准日期： |
| 管理部门：质量管理组 | | |
| 发放范围： | | |
| □正本 □ 副本 NO.：v2.1.0 | | |

## 1目的

三代扩增子流程主要适用于 三代16s,18s,Its分析。

## 2范围

适用于北京238节点

## 3 定义

分析异常：在项目分析中，出现了意外的流程中断。

## 4 职责

信息负责人：如期完成标准分析及个性化分析，分析结果达到质控要求。参考《数据分析标准操作规程》、《信息搜集表评估标准操作规程》

## 5 工作程序

### 5.1信息搜集表准备及说明

信息搜集表为一个excel文件，分为三个部分，每个部分之间需要一个空行做区分。

必须命名为xlsx格式的文件。

#### 5.1.1 信息收集表部分

项目编号：子项目编号

任务单名称：任务单名称

分析类型：16s, 18s， ITs 三选一



#### 5.1.2 样品信息表部分

样品名称：对应的下机路径的样品名称，即过滤完成后的样本名；

样品编号：样品对应的文库号；

结题报告中样品名称：结题报告中样本对应的名称，结题报告中样品名称列字符长度小于10,以字母开头且只能包含字母、数字、下划线

样品描述：描述中尽量不要使用特殊符号；

分组：样本对应的组名。

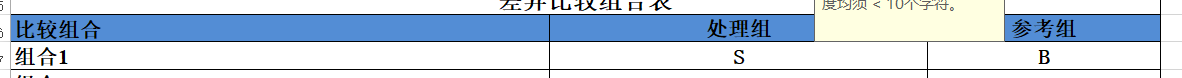


#### 5.1.2 差异比较组合表

比较组合：比较组合的名字，一般无义

处理组：处理组的组名

参考组：参考组的组名。



#### 5.1.3 信搜传到集群

将信息搜集表修改为xlsx结尾的文件并放到info目录下。

该目录下不能有多个信搜。

### 5.2 数据准备

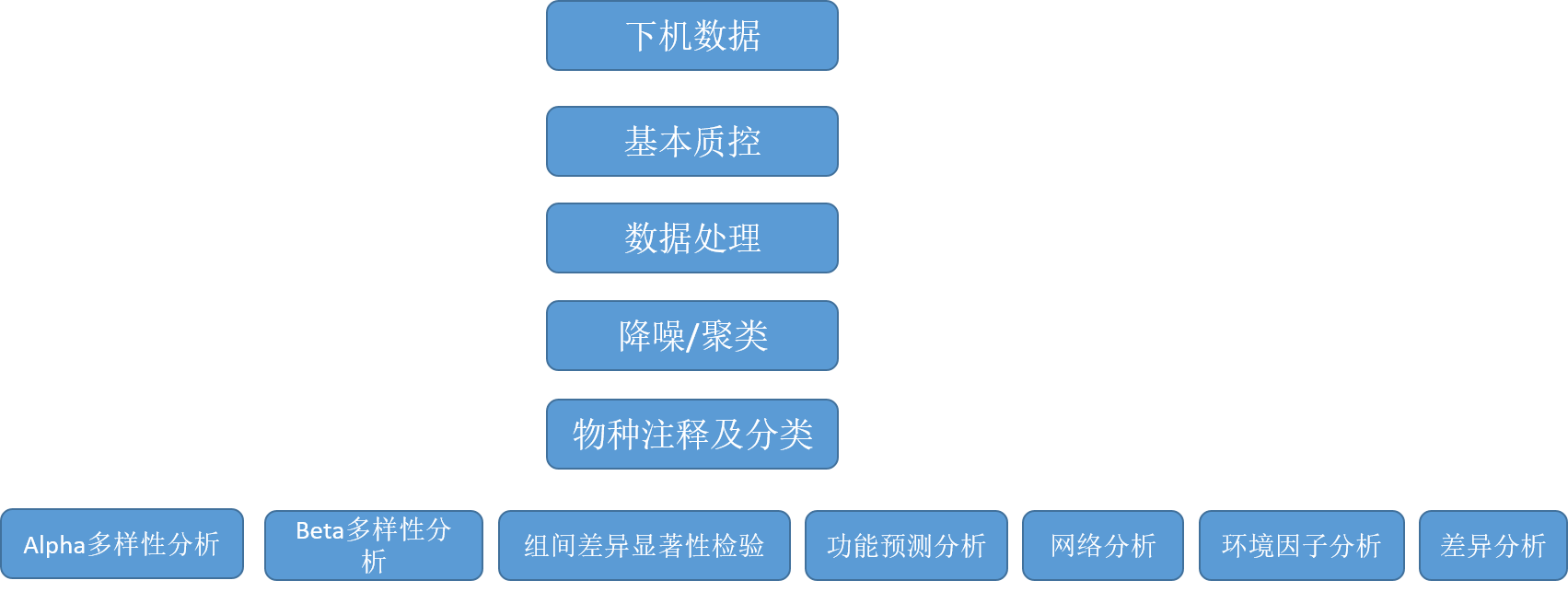
5.2.1 数据准备

数据目录支持两种模式：三代过滤流程整理的交付 Filter\_Result/ANNO\_all\_data

1. Analysis/样本/filter/clean/样本\_R1.fq.gz
2. ANNO\*/Cleandata/样本/样本\_R1.fq.gz（只能有一个ANNO\*开头的目录，如果有多个会报错）
3. 如果有以上两种，优先识别1）
4. 数据目录下需要有STAT\_Result.xls文件

将以上两种目录放到Filter\_Result目录下，并链接到Filter目录下。

### 5.3常规流程使用说明



#### 5.3.1 参数说明

说明：此流程目前在北京238节点投递(qsub投递)。

流程目录为：

$Bin=/annoroad/data1/bioinfo/PROJECT/RD/Cooperation/RD\_Group/tuchengfang/Pipeline/Stable\_Pipeline/TGS\_16s

主脚本：

$BIN/bin/Generate\_pipeline/generate\_pipeline.py

流程参数说明：

-i,--indir:【必需】输入路径（分析路径），为Filter,info所在路径

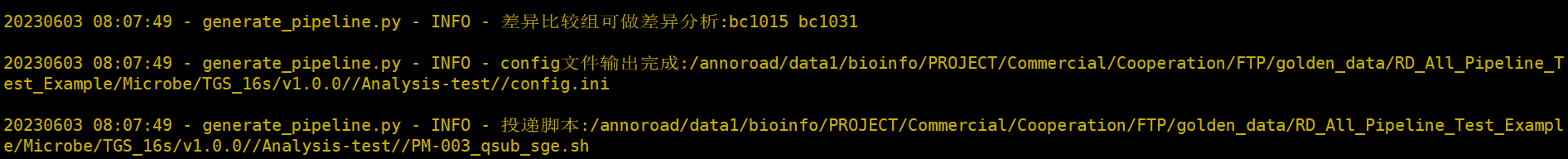
-o,--outdir:【非必需】输出路径，如果不给定，则默认为indir

-c,--config:【非必需】流程配置文件，默认为 $BIN/config/config.txt

-r,--run:【非必需】是否自动投递参数，如果给定，则自动流程；如果不给定，需要手动投递

使用示例：

/annoroad/data1/software/bin/miniconda/envs/python3\_base/bin/python3 $BIN/bin/Generate\_pipeline/generate\_pipeline.py –i /annoroad/data1/bioinfo/PROJECT/Commercial/Cooperation/FTP/golden\_data/RD\_All\_Pipeline\_Test\_Example/Microbe/TGS\_16s/v1.0.0/



#### 5.3.2 使用说明

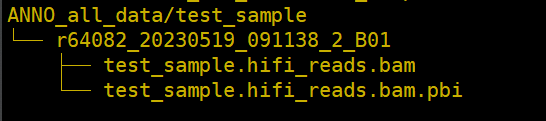
1. indir要求：目录下必须有info,Filter路径

2. Filter路径要求：

1）将三代过滤的Filter\_Result/链接到该目录下

2）Filter\_Result目录下必须有ANNO\_all\_data目录（如果没有需要自己造）

3）目录结构为：



3. info目录要求

信息搜集表必须以xlsx结尾

### 5.4结果说明

1. 在outdir/Analysis-test/目录下生成 Analysis和prepare目录，并将项目的配置文件都放到prepare目录下

2. 任务投递：nohup sh 编号\_qsub.sh &

#### 5.4.2 Analysis目录

├── anosim：组间差异显著性检验分析

├── association\_analysis：相关性分析，包括环境因子分析

├── Diff：差异分析，按照差异分析方法建立目录

├── Diversity：多样性分析，按照alpha和beta分别建立目录

├── Merge\_bam：按照样本合并过滤后的bam文件

├── Network：网络分析

├── PB\_16s\_result：pb\_tools运行的结果，对应2\_0分析

├── Picrust2：功能预测分析

├── QC：合并后的bam的质控

├── Report：结题报告

├── shell：shell脚本

├── Stat：数据去掉接头，Primer之后的数据统计

└── Taxnomy：物种注释统计和绘图结果

## 6. 异常问题说明

问题1：5\_0\_diff.sh步骤中断

可能原因：没有差异结果

排查方法:查看.e文件的输出信息，是否没有差异

问题2；资源不够

解决方案：如果报错资源不够，可以拷贝job\_config.txt文件并修改，然后重新投递 prepare/子项目编号\_qsub.sh

## 6 相关文件

### 6.2 注意事项

### 7 常见问题及解决方法

Q1：支线任务断了咋办？

A：支线任务断了，主线任务会继续运行，而整个流程不受影响。但是查看show\_process的时候，可以看到break的状态。而break的等级是高于run和hold的，因此，需要自行判定程序是否完成。

Q2：主线任务断了咋办？

A：主线断了，那么就断了，需要重新投递。如果主线任务断了，而支线任务没有完成，那么主程序会等候支线任务完成才会实现最终的退出，所以会导致程序一直在运行的假象。这个时候有两个处理办法：

1. 把所有进程都杀掉，然后重新投递

2. 只杀掉主进程（pipeline.py)，并且在log.txt中人为添加支线任务的finish标识（防止再次运行pipeline.py时重新投递），之后重新运行该任务。

Q3：如何监控项目运行状态？

A：运行/annoroad/bioinfo/PMO/liutao/pipeline\_generate/bin/v5/show\_process.py会显示项目的状态。项目运行状态分为running, break, plan, end 四种。其中running表示正在运行，break表示中断，plan表示准备运行，end表示运行完成,hold表示磁盘不够，任务挂起。

Q4：发现任务状态是break，该咋办？

A：当发现任务状态是break的时候，首先需要确定break掉的任务是否是主线任务。

如果是主线任务，如果主程序自然退出(在top或者ps的时候没有发现pipeline.py），则可以重新投递任务；

如果没有自然退出（主程序pipeline.py还在运行），那么可能是有之前的支线任务未完成，可以参照Q2来进行操作；

如果是支线任务，如果查看log.txt发现主线任务全部完成或者正常运行，如果时间允许，可以等待所有任务完成后再投递任务；

如果加急，可以把该步对应的sh文件修改后，手动投递该任务；

如果主线任务也断掉了，那么修改脚本后，重新投递所有任务。

Q5：监控项目的记录文件在哪？

A：程序会在您的home目录下，生成一个记录文件，路径为~/.mission/.pipeline.log，记录了每个项目的分析目录。如果不想显示某个项目，可以对相应的行进行删除或者编辑。

Q6：如果程序断了，咋办？

A：如果程序由于各种因素中断了，仔细检查脚本，如果脚本没错，确定只是中断，那么重新运行一次之前的脚本，默认会把断掉的模块全部重头运行；如果不想将该模块内已经完成的样品重新运行，可以加上-c参数，那么会只运行没有运行成功的样品。

Q7：如何发现配额不够？

A：如果配额不够了，会把所有的任务挂起，使用show\_process查看时，会发现Hold状态；或者使用qstat的时候，会发现hqw，hr，ht等，或者没有任务在运行（因为配额不够，程序会自动不投递任务）

Q8：配额不够了，咋办？

A：第一，找系统组修改配额 ； 第二，修改相应的sh.\*.log文件，加入一行DISK\_QUOTA\*\*G ;之后，程序会自动的release。但需要注意的是，因为之前在程序里设置了较小的配额，所以之后每一步都会被hold。所以需要之后每个log文件都加上 DISK\_QUOTA \*\*G，来每次进行更改；或者杀掉重新来。 或者删除文件来释放空间，这样的话，后面可以不用修改就可以运行。

Q9： 如何杀掉程序？

A:

1. 杀掉所有的子进程 守护进程qsub\_sge.pl，否则的杀掉的任务会重新投递；

2. 杀掉所有的任务 qdel掉

Q10：如何精准的杀掉守护进程？

A：

1. 在重新投递之前，使用ps -f -u name |cat 然后仔细的判别，获得进程ID

2. 查看shell后面的数字，在sh 和log直接的数字，是进程ID，使用kill -9可以杀掉

## 8更新记录

|  |  |  |
| --- | --- | --- |
| **版本号** | **修订内容** | **修订日期** |
| v0.0.1 | 搭建完成 | **20190326** |
| V2.1.0 | 修改比对软件为pbmm2 | **20211021** |
|  |  |  |