**Survey\_v1.3**

**Overview**

Survey\_v1.3是在de novo组装中的一个模块，包括全基因组survey和初步组装两个功能。

Survey:通过kmer分析的方法，预测基因组大小，杂合率和重复序列概率。

初步组装：使用SoapDenovo进行初步组装，并对组装之后的Contig文件进行覆盖情况和GC含量的分析。

优化点：

1. 将kmer计数软件由soapKF 替换为jellyfish，缩短了运行时间;
2. 更改了输入方式，以配置文件的方式读入输入文件和参数；增加了一些运行参数；
3. 将流程中的绝对路径改为了相对路径，方便移植；
4. 添加了必要的注释和帮助信息，以及简单的运行日志；

**Installation**

集群路径：

南京集群 */NJPROJ1/RAD/users/zhouyiqi/tar/survey\_v1.3.tar.gz*

天津集群 */ifs/TJPROJ3/RAD/zhouyiqi/tar/survey\_v1.3.tar.gz*

将软件包拷贝到个人目录下，解包：

$ tar -zxvf survey\_v1.3.tar.gz

完成后，进入目录，运行setup.sh:

$ sh setup.sh

更新环境变量：

$ source ~/.bashrc

即可完成环境变量的配置。

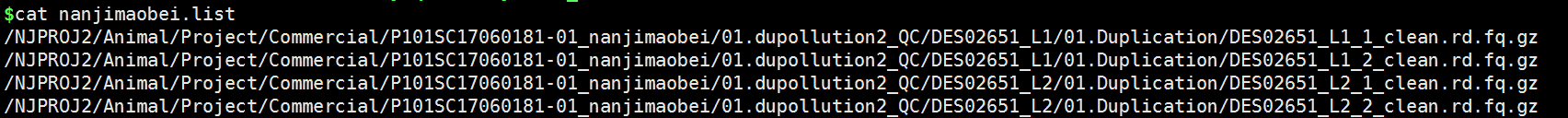
**Requirements**

* Perl 5.10+

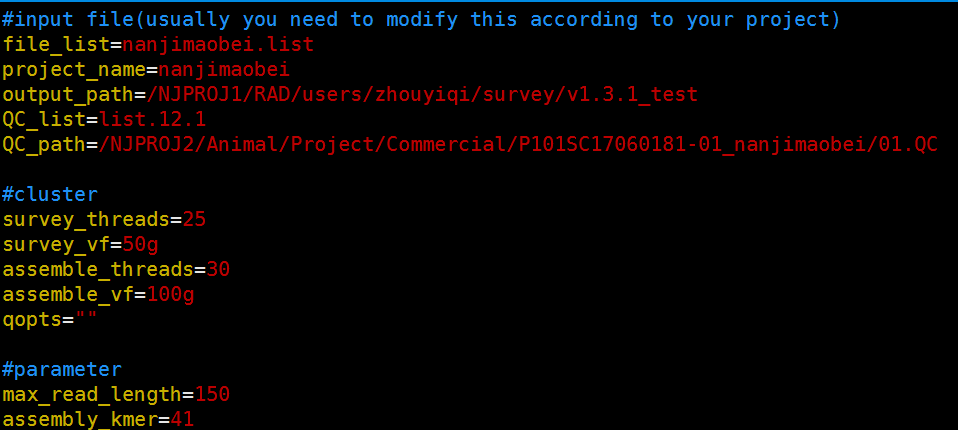
**Usage**

1. 准备好qc文件夹和data.list文件，建立survey输出文件夹。

data.list为 clean\_data的路径，每行一个，可以是.gz文件。示例：



2、将软件包目录下的example.cfg文件拷贝到输出文件夹下，修改其内容：



file\_list:含有clean\_data绝对路径的列表，每行一个数据，注意路径不能是软链接，否则会报错。

project\_name:物种名。

output\_path:输出文件夹。

QC\_list:qc文件夹下list的文件名称，list内容为文库名+文库大小，用tab分隔，每个文库一行。

QC\_path:qc文件夹路径。

survey\_threads：survey线程数。

survey\_vf：survey内存。

assemble\_threads：组装线程数。

assemble\_vf:组装内存。

qopts：其他sge参数，例如-P。

max\_read\_length:read 长度。

assembly\_kmer：组装kmer大小，默认是41。

1. 进入输出文件夹，运行:

$ survey.sh example.cfg

将投递survey和assemble两个任务，运行完毕后会在输出目录下生成report文件夹。

**Contact**

如有任何问题，请联系zhouyiqi@novogene.com