

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：

```
$cat nanjimaobei.list
/NJPROJ2/Animal/Project/Commercial/P101SC17060181-01_nanjimaobei/01.dupollution2_QC/DES02651_L1/01.Duplication/DES02651_L1_1_clean.rd.fq.gz
/NJPROJ2/Animal/Project/Commercial/P101SC17060181-01_nanjimaobei/01.dupollution2_QC/DES02651_L1/01.Duplication/DES02651_L1_2_clean.rd.fq.gz
/NJPROJ2/Animal/Project/Commercial/P101SC17060181-01_nanjimaobei/01.dupollution2_QC/DES02651_L2/01.Duplication/DES02651_L2_1_clean.rd.fq.gz
/NJPROJ2/Animal/Project/Commercial/P101SC17060181-01_nanjimaobei/01.dupollution2_QC/DES02651_L2/01.Duplication/DES02651_L2_2_clean.rd.fq.gz
```

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=nanjimaobei.list
project_name=nanjimaobei
output_path=/NJPROJ1/RAD/users/zhouyiqi/survey/v1.3.1_test
QC_list=list.12.1
QC_path=/NJPROJ2/Animal/Project/Commercial/P101SC17060181-01_nanjimaobei/01.QC


survey_threads=25
survey_vf=50g
assemble_threads=30
assemble_vf=100g
qopts=""


max_read_length=150
assembly_kmer=41

```

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。

3. 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