

Cassava Production Pipeline

Terry Casstevens (tmc46@cornell.edu)

May 15, 2014

Requirements

Must have git installed. Check using this command...

```
git --version
```

Must have Java 7 installed. Check using this command...

```
java -version
```

To Initially Download Tassel 5 Standalone Code...

```
git clone https://bitbucket.org/tasseladmin/tassel-5-standalone.git
```

To Update Packages to Latest Source Code...

```
git pull
```

Example Configuration File...

```
<?xml version="1.0" encoding="UTF-8" standalone="no"?>
<TasselPipeline>
  <fork1>
    <ProductionPipeline>
      <inputDirectory>/home/tmc46/CassavaProdPipeline/rawSeqs/</inputDirectory>
      <keyFile>/home/tmc46/CassavaProdPipeline/inputFiles/keyfile.txt</keyFile>
      <enzyme>ApeKI</enzyme>
      <productionTOPM>/export/species/Manihot_esculenta/gbs/topm/productionTOPM_T5_20140327.topm
      </productionTOPM>
      <outputGenotypeFile>/home/tmc46/CassavaProdPipeline/genotypes/CassavaGenotypes.h5</outputG
      enotypeFile>
      <archiveDirectory>/home/tmc46/CassavaProdPipeline/archive/</archiveDirectory>
    </ProductionPipeline>
  </fork1>
  <runfork1/>
</TasselPipeline>
```

Description of Configuration File Parameters...

Usage:

ProductionPipeline <options>

-inputDirectory <Input Directory> : Input directory containing fastq AND/OR qseq files (required)

-keyFile <Key File> : Barcode Key File (required)

-enzyme <Enzyme> : Enzyme used to create the GBS library (required)

-productionTOPM <Production TOPM> : Physical map file containing tags and corresponding variants (production TOPM) (required)

-outputGenotypeFile <Output Genotype File> : Output (target) HDF5 genotypes file to add new genotypes to (new file created if it doesn't exist) (required)

-archiveDirectory <Archive Directory> : Archive directory where to move processed files (required)

To Execute Pipeline...

First symbolic link raw sequence files into rawSeqs directory...

```
ln -s
/export/species/Manihot_esculenta/gbs/fastq/fastq1/c14cgacxx_7_fastq.txt.gz
/home/tmc46/CassavaProdPipeline/rawSeqs/
```

Then execute the pipeline with this command...

```
/home/tmc46/tassel-5-standalone/run_pipeline.pl -Xmx10g -configFile
/home/tmc46/CassavaProdPipeline/inputFiles/CassavaProdPipeline.xml
```

Cron Job (i.e. Execute at 2:30 a.m. every day)

```
30 2 * * * root /home/tmc46/tassel-5-standalone/run_pipeline.pl -Xmx10g
-configFile /home/tmc46/CassavaProdPipeline/inputFiles/CassavaProdPipeline.xml
```