Debarcer: De-Barcoding and Error Correction

Paul Krzyzanowski
Ontario Institute for Cancer Research,
Toronto, Ontario, Canada
paul.krzyzanowski@oicr.on.ca

February 9, 2016

1 Location

You can export or checkout the whole package from the following location:

```
  $ svn export file:///u/pkrzyzanowski/svn/repo/projects \
/EAC/molecular_barcoding/trunk/debarcer .
  $ svn co file:///u/pkrzyzanowski/svn/repo/projects \
/EAC/molecular_barcoding/trunk/debarcer .
```

2 Using Debarcer

2.1 Requirements

- Ensure you have 16G memory available
- Note that a bam file for the current alignment will be generated in your work directory
- Currently, you need to have access to 'qsub' through SGE or OGS

2.2 Running the package

To generate a set of positional compositions, you can export Debarcer, set the BHOME environment variable, and run it like so:

```
  $ svn export file:///u/pkrzyzanowski/svn/repo/projects/EAC/ \
molecular_barcoding/trunk/Debarcer .
  $ export BHOME $PWD
  $ runDebarcer.sh [FASTQFILE] [SAMPLENAME]
```

Or if you're set up for using modules, you can load debarcer from anywhere and run it:

```
  $ module load debarcer/trunk
  $ runDebarcer.sh Sample_6.R1.fastq.gz Sample6
```

Using modules is handy for running via qsub:

```
  qsub -N "BCS" -b y -cwd -l h_vmem=16g "module load debarcer/trunk; \
runDebarcer.sh Sample_6.R1.fastq.gz Sample6"
```

To run a test set of data, do:

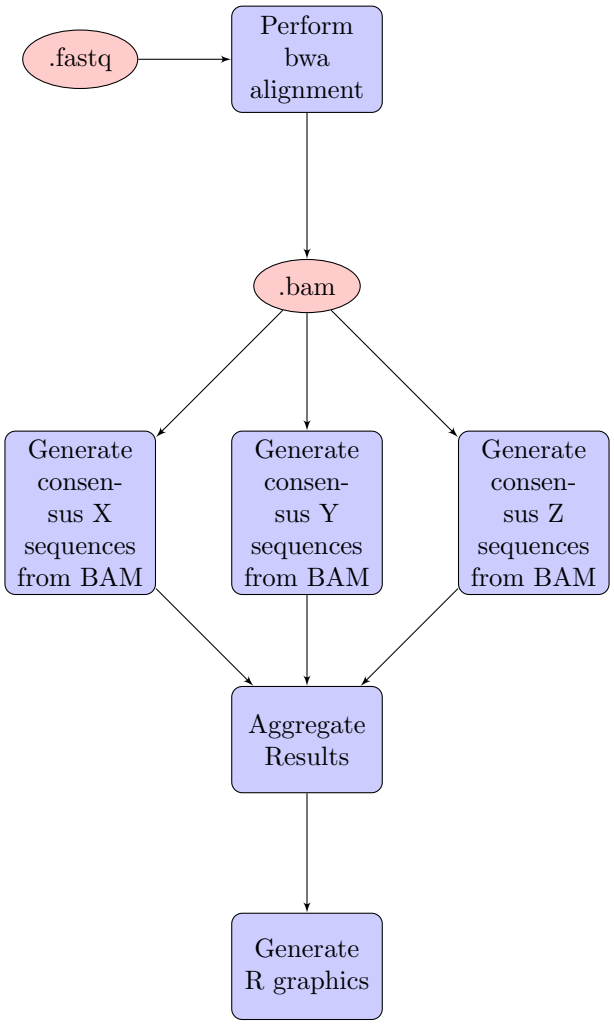
```
# ls /u/pkrzyzanowski/projects/simsenseq/debarcer/demodata # for more test files
cp /u/pkrzyzanowski/projects/simsenseq/debarcer/demodata/Sample_9297.R1.fastq.gz .
module load debarcer/trunk
runDebarcer.sh Sample_9297.R1.fastq.gz Sample_9297
```

Sample_9297 is a 1-plex and Sample_9292 is a 5-plex.

To help automate this whole process, there is a script in the debarcer 'tools' directory that will create a results hierarchy based on a directory of fastq files and write the appropriate run scripts:

```
mkdir -p ResultsRoot\fastqs
cd ResultsRoot\fastqs
find [location of the the fastq files] -name "*fastq.gz" -exec ln -s {} \;
cd ..
svn export file:///u/pkrzyzanowski/svn/repo/projects/EAC/ \
molecular_barcoding/trunk/debarcer/tools/populateWorkDirectory.pl
# Make any necessary changes to populateWorkDirectory.pl
perl populateWorkDirectory.pl ./fastqs
# Setup any debarcer.conf files in the results directories, if needed
./relaunchAllDebarcers.sh
```

3 Workflow



4 Output

The output contains several different types of files.

Package results files with a tar command that wraps up the files found by:

```
find . -name *.pdf -o -name "*bamPositionComposition*" -o \
-name "*UID*" -o -name "*SummaryStatistics.txt" -o -name "*log"
```

i.e.

```
tar cvfz results_file_name.tar.gz ‘[find command goes in backticks]‘
```

5 Miscellaneous

Requirements: Consensus calling script needs at least 16G memory.

5.1 Revisions

0.2.0: July 27, 2015. Svn Revision: 347.