Check-In Procedure for GBS Data from Novogene

Prerequisites

The following shell script stored in /homes/mlucas/scripts is required:

```
checkin_novogene_gbs_data.sh
```

The shell script references the following programs stored in /homes/mlucas/python3_programs/GBS:

```
rename_gbs_file

compute_gbs_file_table_metadata

generate_barcode_distribution

generate_blank_dna_quantification_report
```

Download GBS Data from Novogene

1. Notification of availability of new data will be received via email from an email address at novogene.com. The email will contain a host address, username and password to access the data.

2. Open a lftp session to the host using the login credentials contained in the email. The command will be of the form:

   ```
   lftp -u 'P202SC18122641-01_20181211_DziSrH,uH9SkL' hwftp.novogene.com
   ```

3. Once you are logged in and see the lftp prompt, execute an ls command to find the name of the directory containing the GBS sequence data. The directory will have a name similar to e.g. `C202SC18122641.` Make a note of the directory name and exit.

4. Create a folder with the name of the directory containing the GBS data from the previous step

   Example: `mkdir /bulk/mlucas/incoming/C202SC18122641`

5. Issue the lftp command to download the GBS data from the Novogene server.

   Example:

   ```
   nohup lftp -e 'mirror C202SC18122641  /bulk/mlucas/incoming/C202SC18122641'
    -u 'P202SC18122641-01_20181211_DziSrH,uH9SkL' hwftp.novogene.com &
   ```

Execute the Script to Download and Check-in the GBS Data

6. Logon to Beocat
7. cd to /homes/mlucas/scripts and locate the script **checkin_novogene_data.sh**
8. Execute the script on the command line with the Novogene directory name noted earlier.

   Example assuming the following parameter values:

   Novogene Directory: C202SC18122641

   ```
   ./checkin_novogene_gbs_data.sh C202SC18122641
   ```

   The script will perform the following steps:

   a. Verify the MD5 checksums of all GBS files that were downloaded.
   b. Update the gbs database table for the associated gbs_id with the flowcell and lane values.
   c. Rename each of the GBS files to a name conforming to the standard GBS file naming standard.

      Example:

      ```
      GBS1442R1xPSIDEC2018StrawberryP01P02_HV2FLBBXX_s_7_fastq.txt
      ```

   d. Compute the MD5 checksum and line count for each GBS file in the set and update the gbs_file table md5sum and num_lines columns for the gbs_file_id associated with each file.

   e. Generate read-barcode distribution report

      This report will allow the user to check % valid reads and % reads found in any blank well in the GBS file.

      The report will have the following naming format:

      GBSnnnn_sample_summary.txt

      Example:

      ```
      GBS1442_sample_summary.txt
      ```

f. Generate DNA quantification report

This report will allow the user to check the DNA quantification values for blank wells in the GBS library

The report will have the following naming format:

GBSnnnn_blank_dna_quant_report.csv

Example:

```
GBS1442_blank_dna_quant_report.csv
```

Review QC Reports and Cleanup

9. Review the GBSnnnn_sample_summary.txt report and verify that the following thresholds have not been exceeded:

   % Valid reads > 90%
   % Reads in any BLANK well < 0.01%

   If either threshold is violated, investigate potential causes:
   i. Incorrect blank well in DNA plate record
   ii. Poor sequencing run quality

10. Review the GBSnnnn_blank_dna_quant_report.csv to make sure that DNA quantification values in the blank wells are within tolerance.

    If the values reported are all NULL, this means that the dnaQuant table has not been updated yet for this GBS plate.

11. Change the group on the GBS file to ksu-plantpath-jpoland and remove write permissions from the file.

```
chgrp ksu-plantpath-jpoland GBS1442R1xPSIDEC2018StrawberryP01P02_HV2FLBBXX_s_7_fastq.txt

chmod a-w GBS1442R1xPSIDEC2018StrawberryP01P02_HV2FLBBXX_s_7_fastq.txt
```

12. Move the GBS file to /bulk/jpoland/sequence directory on Beocat.

```
mv GBS1442R1xPSIDEC2018StrawberryP01P02_HV2FLBBXX_s_7_fastq.txt /bulk/jpoland/sequence/.
```