Check-In Procedure for GBS Data from Génome Québec

Prerequisites

The following shell scripts and programs that are stored in /homes/mlucas/scripts are required:

checkin_quebec_gbs_data.sh

rename_gbs_file

compute_gbs_file_metadata

generate_barcode_distribution

generate_blank_dna_quantification_report

Download GBS Data from Génome Québec

1. Notification of availability of new data will be received via email from ngs-services@genomequebec.com - Subject will contain Génome Québec Project Name
2. Click on the share link in the email to open a browser with an Illumina login screen to accept the share of the new data.
3. Login to Génome Québec using the link provided in the email which is labelled “now available in Nanuq”. This will take you to the login page.
4. Login. This will take you directly to the Project Data page.
5. Click on the HiSeq Read Sets Tab.
6. Click on the check box in the title line of the project table to select all 8 Read Sets.
7. Click on the Download Read Files Tab.
8. Select the Download Selected Reads button.
9. Select the Text file with URL links button.
10. Click on the Download button. A folder called ReadSetLinks which contains a script to download the GBS sequence files will be downloaded.
11. Copy ReadSetLinks file to Beocat.
12. cd to the ReadSetLinks folder and execute the shell script run_wget.sh. This will download the 8 GBS files in the set as well as files containing the MD5 checksum for each GBS file.

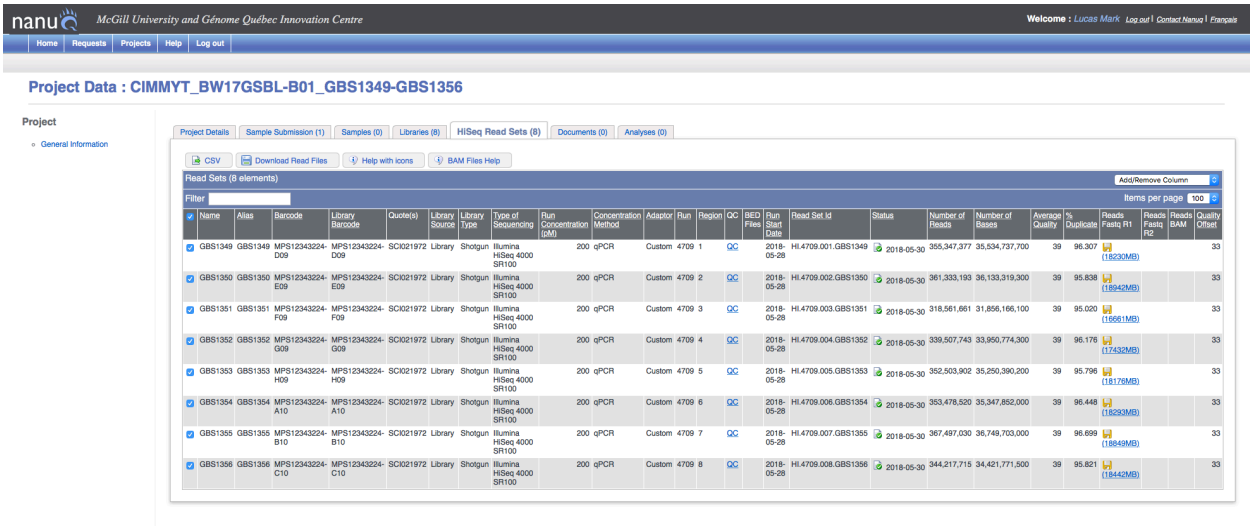

Figure 1 Project Data Page

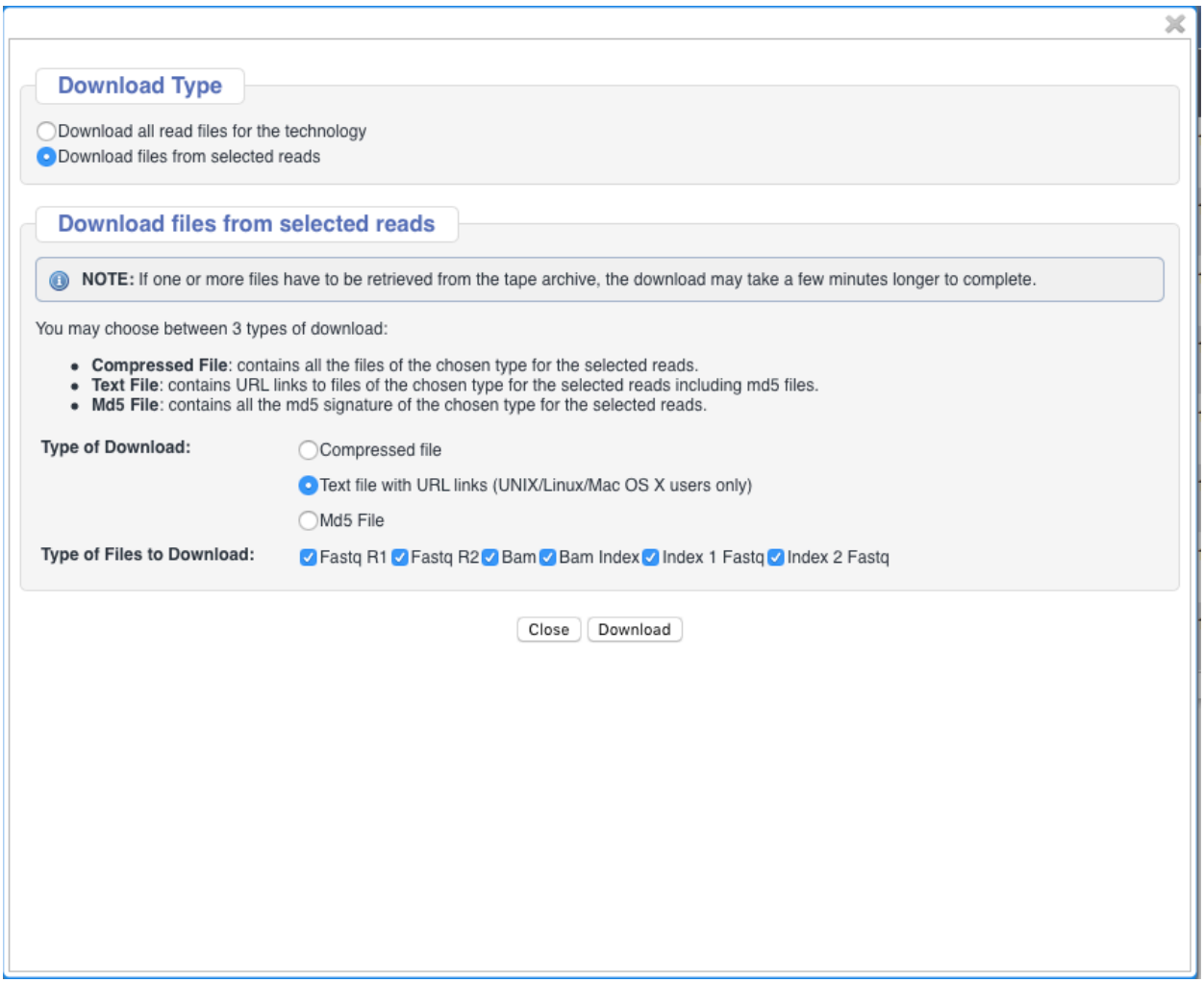

Figure 2 Download Read Files Page

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

13. Logon to Beocat
14. cd to /homes/mlucas/scripts and locate the script **checkin_ksu_quebec_data.sh**
15. Execute the script on the command line with the Illumina project name
    Example:

    ./checkin_quebec_gbs_data.sh readSetLinks_GBS1360-GBS1367

    The script will perform the following steps:

    a. Verify the MD5 checksums of the 8 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 8 GBS files to a name conforming to the standard GBS file naming standard.

       Example:

       GBS1370xStrawberryP01P02_HVY7JBGX7_s_0_fastq.txt.gz

    d. Compute the MD5 checksum and line count for each GBS file in the set and update the gbs table ms5sum and num_lines columns for the gbs_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:

       GBS1370_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:

```
GBS1370_blank_dna_quant_report.csv
```

Review QC Reports and Cleanup

16. 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

17. 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.

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

```
chgrp ksu-plantpath-jpoland GBS1370FxStrawberryP01P02_HVY7JBGX7_s_0_fastq.txt.gz
chmod a-w GBS1370FxStrawberryP01P02_HVY7JBGX7_s_0_fastq.txt.gz
```

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

```
mv GBS1370FxStrawberryP01P02_HVY7JBGX7_s_0_fastq.txt.gz /bulk/jpoland/sequence/.
```