

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.

Project Data : CIMMYT_BW17GSBL-B01_GBS1349-GBS1356

Project

- General Information

Project Details | Sample Submission (1) | Samples (0) | Libraries (8) | HiSeq Read Sets (8) | Documents (0) | Analyses (0)

CSV | Download Read Files | Help with icons | BAM Files Help

Read Sets (8 elements) Add/Remove Column

Name	Alias	Barcode	Library Barcode	Quote(s)	Library	Library Type	Type of Sequencing	Run Concentration (�M)	Concentration Method	Adaptor	Run	Region	QC	Run Start Date	Read Set Id	Status	Number of Reads	Number of Bases	Average Quality	% Duplicate	Reads Fastq R1	Reads Fastq R2	Reads BAM	Quality Offset
GBS1349	GBS1349	MP512343224-D09	MP512343224-D09	SCI021972	Library	Shotgun	Illumina HiSeq 4000 SR100	200	qPCR	Custom	4709	1	QC	2018-05-28	HI.4709.001.GBS1349	2018-05-30	355,347,377	35,534,737,700	39	96.307	(18230MB)		33	
GBS1350	GBS1350	MP512343224-E09	MP512343224-E09	SCI021972	Library	Shotgun	Illumina HiSeq 4000 SR100	200	qPCR	Custom	4709	2	QC	2018-05-28	HI.4709.002.GBS1350	2018-05-30	361,333,193	36,133,319,300	39	95.838	(18842MB)		33	
GBS1351	GBS1351	MP512343224-F09	MP512343224-F09	SCI021972	Library	Shotgun	Illumina HiSeq 4000 SR100	200	qPCR	Custom	4709	3	QC	2018-05-28	HI.4709.003.GBS1351	2018-05-30	318,561,661	31,856,166,100	39	95.020	(15661MB)		33	
GBS1352	GBS1352	MP512343224-G09	MP512343224-G09	SCI021972	Library	Shotgun	Illumina HiSeq 4000 SR100	200	qPCR	Custom	4709	4	QC	2018-05-28	HI.4709.004.GBS1352	2018-05-30	339,507,743	33,950,774,300	39	96.176	(17432MB)		33	
GBS1353	GBS1353	MP512343224-H09	MP512343224-H09	SCI021972	Library	Shotgun	Illumina HiSeq 4000 SR100	200	qPCR	Custom	4709	5	QC	2018-05-28	HI.4709.005.GBS1353	2018-05-30	352,503,902	35,250,390,200	39	95.796	(18176MB)		33	
GBS1354	GBS1354	MP512343224-A10	MP512343224-A10	SCI021972	Library	Shotgun	Illumina HiSeq 4000 SR100	200	qPCR	Custom	4709	6	QC	2018-05-28	HI.4709.006.GBS1354	2018-05-30	353,478,520	35,347,852,000	39	96.448	(18293MB)		33	
GBS1355	GBS1355	MP512343224-B10	MP512343224-B10	SCI021972	Library	Shotgun	Illumina HiSeq 4000 SR100	200	qPCR	Custom	4709	7	QC	2018-05-28	HI.4709.007.GBS1355	2018-05-30	367,497,030	36,749,703,000	39	96.699	(18840MB)		33	
GBS1356	GBS1356	MP512343224-C10	MP512343224-C10	SCI021972	Library	Shotgun	Illumina HiSeq 4000 SR100	200	qPCR	Custom	4709	8	QC	2018-05-28	HI.4709.008.GBS1356	2018-05-30	344,217,715	34,421,771,500	39	95.821	(18442MB)		33	

Figure 1 Project Data Page

Download Type

Download all read files for the technology

Download files from selected reads

Download files from selected reads

NOTE: If one or more files have to be retrieved from the tape archive, the download may take a few minutes longer to complete.

You may choose between 3 types of download:

- Compressed File:** contains all the files of the chosen type for the selected reads.
- Text File:** contains URL links to files of the chosen type for the selected reads including md5 files.
- Md5 File:** contains all the md5 signature of the chosen type for the selected reads.

Type of Download:

Compressed file

Text file with URL links (UNIX/Linux/Mac OS X users only)

Md5 File

Type of Files to Download:

Fastq R1 Fastq R2 Bam Bam Index Index 1 Fastq Index 2 Fastq

Close
Download

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 GBS1370FxFStrawberryP01P02_HVY7JBGX7_s_0_fastq.txt.gz
```

```
chmod a-w GBS1370FxFStrawberryP01P02_HVY7JBGX7_s_0_fastq.txt.gz
```

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

```
mv GBS1370FxFStrawberryP01P02_HVY7JBGX7_s_0_fastq.txt.gz /bulk/jpoland/sequence/.
```