Mitochondrial DNA Massively Parallel Sequencing (mitoMPS) Analysis using the Qiagen CLC Workbench and AQME Toolkit

1 Purpose

1.1 To compile the sequence data generated by the MiSeq/MiSeq FGx and Verogen Universal Analysis Software (UAS) into a project for analysis, by editing the sequence data and compiling a consensus sequence that can be compared with the revised Cambridge Reference Sequence (rCRS) to determine the mitochondrial DNA type.
Procedure

1.2 The analysis will use two elements of software: Qiagen CLC Genomics Workbench, with the AFDIL-Qiagen mtDNA Expert (AQME) toolkit plugin.

1.3 The data following the MiSeq run will be saved on the PowerWulf server, in individual folders of FASTQ files for analysis. The files will also be stored in an appropriate folder in M:\MPS_Data\.

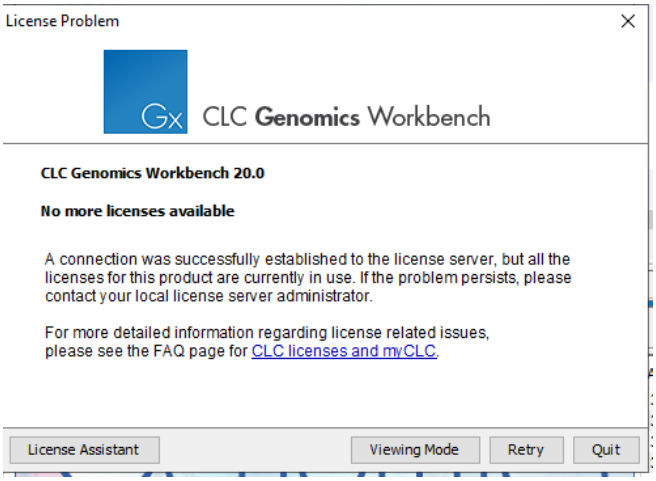
1.4 The files with “R1” filenames will be imported into the CLC software for analysis, using the program’s “Promega NGS mito workflow”. The R1 files will be trimmed and aligned, with variant tables output in excel tab-delimited format.

2 Import data for processing

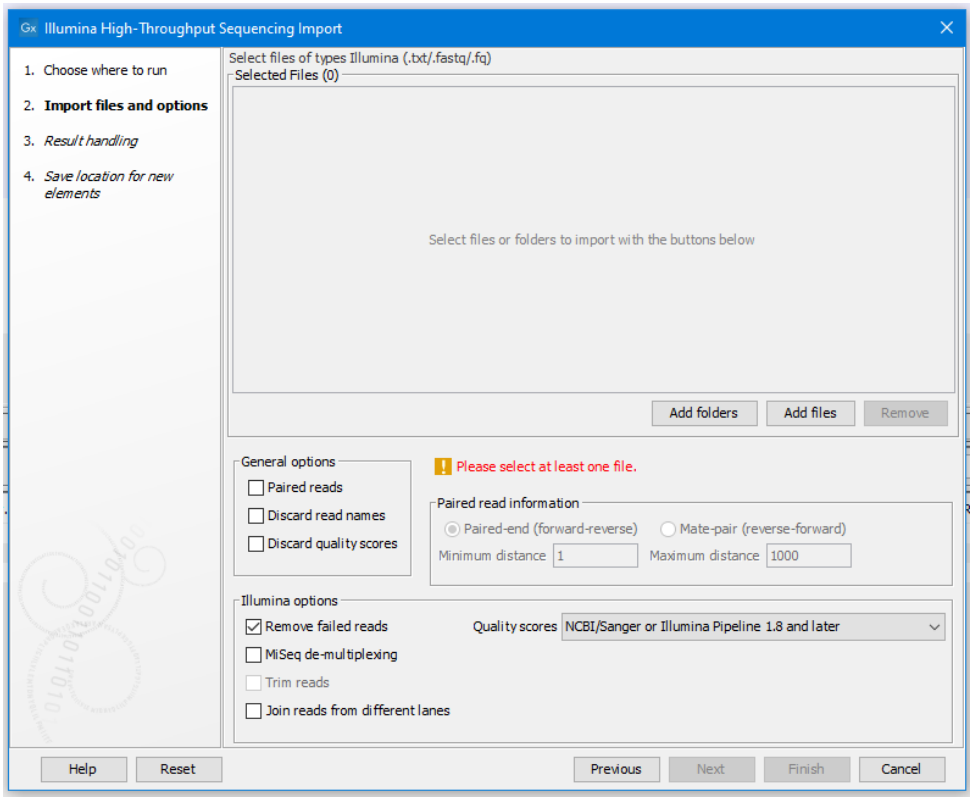
2.1 Open the Qiagen CLC Genomics Workbench from your computer’s start menu.

2.1.1 If the software does not appear to be installed, submit a HelpDesk ticket for installation.

2.1.2 If the following message appears, then another user is currently using the department's network license.



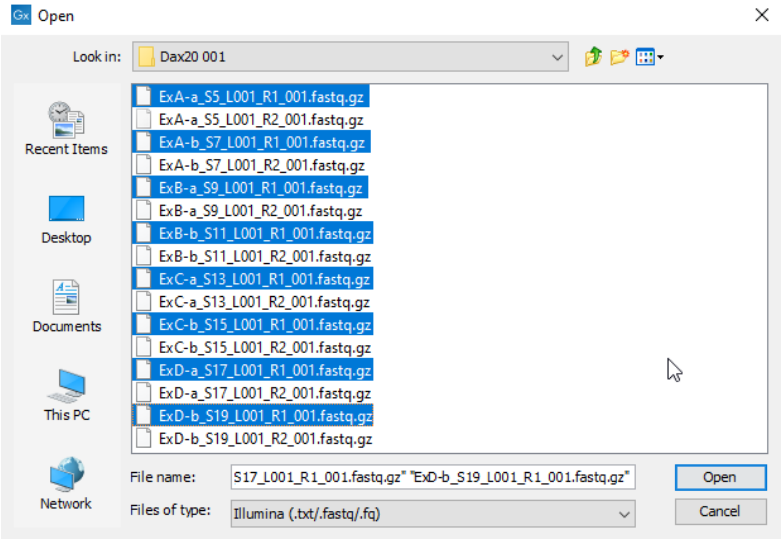
2.2 Go to File -> Import -> Illumina and ensure that the settings are as below:



2.3 Navigate to the PowerWulf subdirectory that contains the FASTQ files for analysis.

2.3.1 If the PowerWulf server is not mapped to a network drive on your computer, please contact a supervisor for the IP address and login credentials.

2.4 Select the R1 sample files from the directory, and click **Open**.



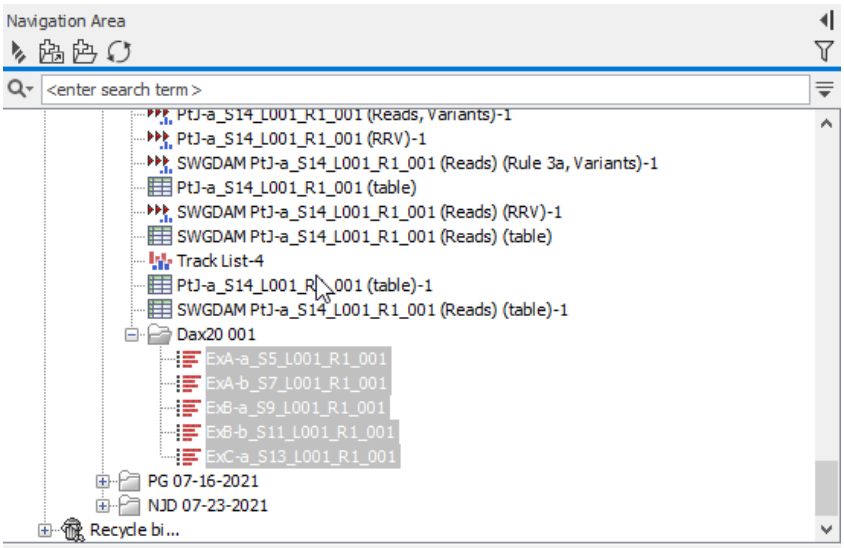
2.5 Confirm the selected R1 files are listed in the subsequent screen. Click **Next.**

2.6 Under **Result handling**, choose the **Save** radio button, and be sure that **Create subfolders per batch unit** is NOT checked, then click **Next.**

2.7 On the directory navigation screen, navigate to the casework folder and click **New Folder** to create a subdirectory for your run. Enter the run name as the name of the new folder, then click **OK**.

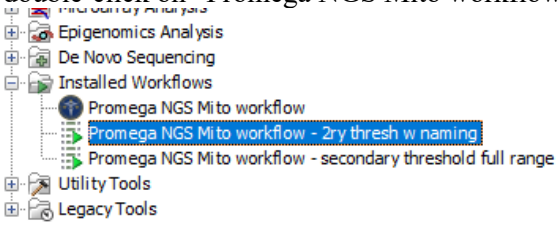
2.8 Click **Finish**.

2.9 Confirm that the selected files are now present in the Navigation Area in the top left of your screen.



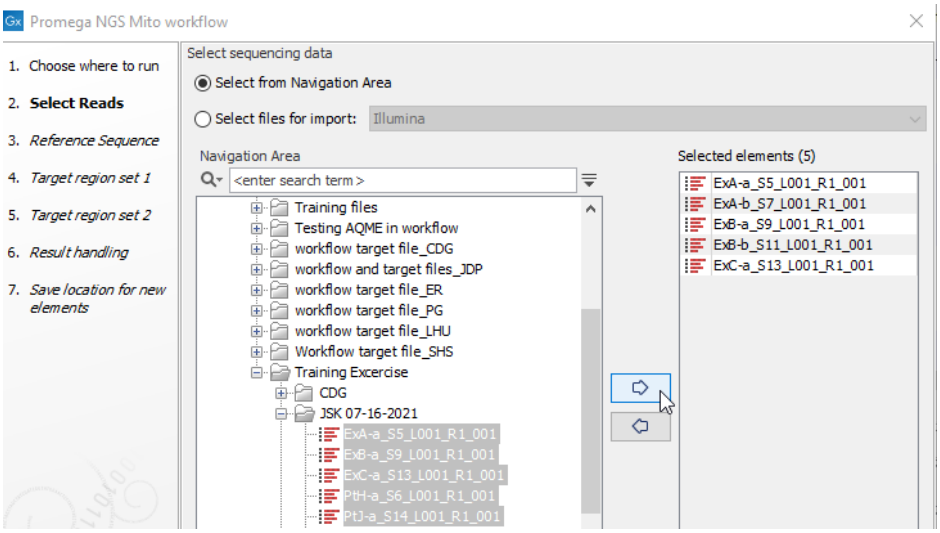
3 Run the analysis workflow

3.1 In the Toolbox -> Toolbox menu, click the + to expanded the Installed Workflows items, and double-click on “Promega NGS Mito workflow – 2ry thresh w naming”.

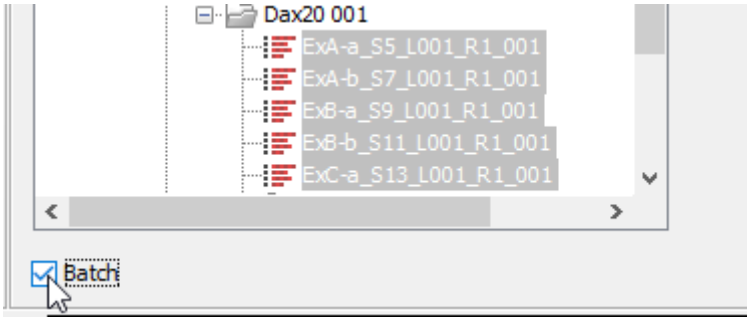


3.1.1 If this workflow is not installed, you may find the installer in the M:\MPS_Data\ directory. The current file is “Promega NGS Mito workflow – 2ry thresh w naming-0.1.cpw”

3.2 If your R1 files are not already present in the Selected Elements area on the right of the screen, click to highlight your R1 files and click the **Right Arrow** to add them to the Selected Elements.



3.3 Check the **Batch** button under the Navigation area, then click **Next**.

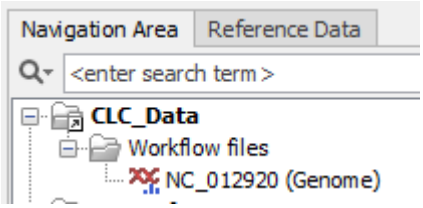


3.4 For Configure batching, the **Use organization of input data** radio box should be selected. Click **Next**.

3.5 For Batch overview, the selected R1 files should be in the Units list on the left. Click **Next**.

3.6 For Workflow Input, confirm that **NC_012920 (Genome)** is selected by default.

3.6.1 If it is not selected, click on the icon and select **NC_012920 (Genome)** from the CLC_Data/Workflow files in the Navigation Area on the left side, click the **Right arrow** to add to the Selected Elements, and click **OK**.



3.7 Click **Next**.

3.8 For Target region set 1, confirm that NC_012920 (Genome) selection set 1 -600 threshold (Misc. feature) is selected by default.

3.8.1 If it is not selected, click on the icon and select **NC_012920 (Genome) selection set 1 -600 threshold (Misc. feature)** from the CLC_Data/Workflow files in the Navigation Area on the left side, click the **Right arrow** to add to the Selected Elements, and click **OK**.

3.9 Click **Next**.

3.10 For Target region set 2, confirm that **NC_012920 (Genome) selection set 1 -600 threshold (Misc. feature)** is selected again by default.

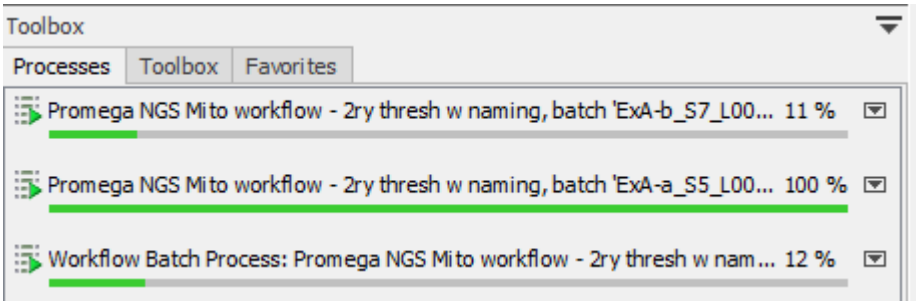
3.10.1 If it is not selected, click on the icon and select **NC_012920 (Genome) selection set 1 -600 threshold (Misc. feature)** from the CLC_Data/Workflow files in the Navigation Area on the left side, click the **Right arrow** to add to the Selected Elements, and click **OK**.

3.11 Click **Next**.

3.12 For Result handling, check the **Save in specified location** radio button, and check **Create workflow result metadata**. Click **Next**. (If the radio button only says "Save", use the "Previous" button to go back to the "Select Reads" step and make sure that the "Batch" box is checked off.)

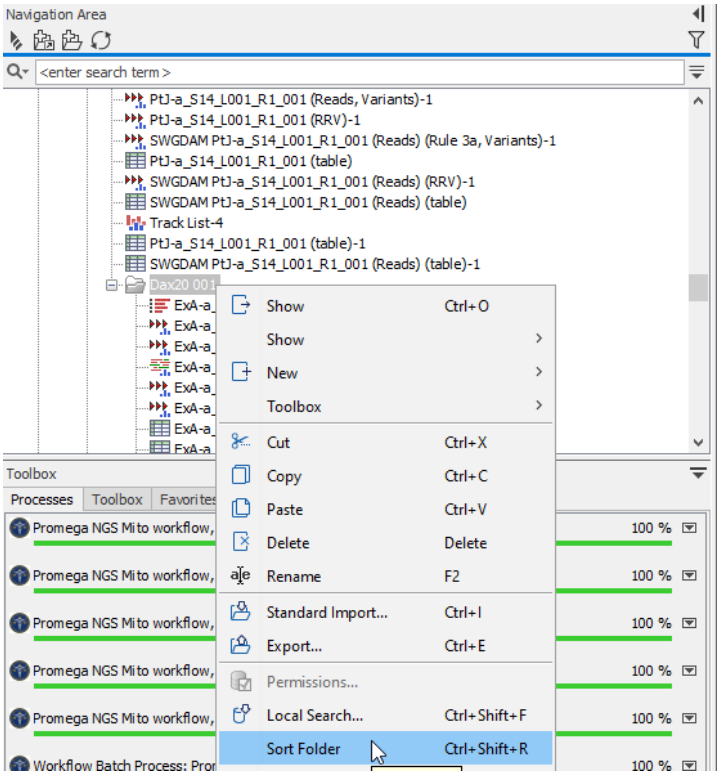
3.13 For Save location for new elements, select the run folder in the navigation area and click **Finish**.

3.13.1 The analysis progress can be seen in real time by clicked the Processes tab in the Toolbox, on lower left side of the screen.

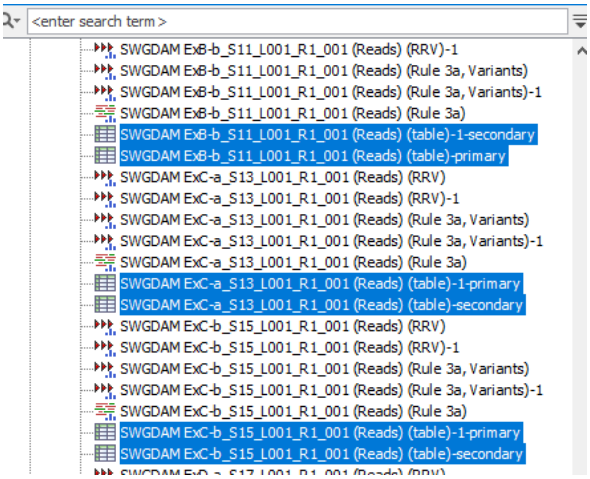


4 Exporting results

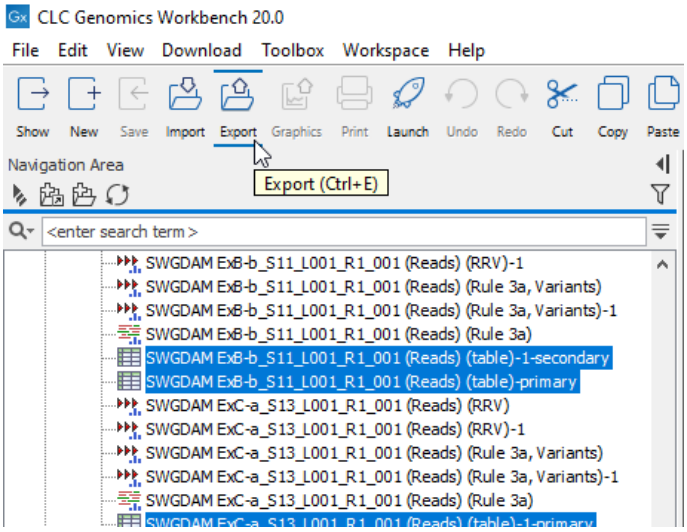
4.1 In the Navigation area, right-click on the folder name of your project, and select **Sort folder**.



4.2 Scroll down the output file list, until you reach the files beginning with **SWGDAM**. Click and select all SWGDAM files with the variant table icon . Each sample should have two tables whose names end in "-primary" and "-secondary", respectively. These correspond with the 600-read and 100-read analysis thresholds respectively.



4.3 With the files selected, click the **Export** icon along the top menu.



4.4 Select the output format for the tables. Select **Excel 97-2007** and click the **Select** button.

4.5 Confirm there are two SWGDAM files for each template library in the Selected Elements table on the right side of the screen, one for each of the two analysis thresholds used to analyze each library.

4.5.1 If any are absent from the Selected elements list, locate the missing file(s) in the Navigation area, click to select, and click the **right arrow** to populate them into the Selected elements list.

4.6 Check the **Batch** button under the Navigation area, then click **Next**.

4.7 Under Batch overview, the selected pairs of variant table, with names ending in **(table)** and **(table)-1** should be present in the Units box.

4.8 Under Specify export parameters, Output as a single file should NOT be checked. Export all columns. (Deselect to specify columns in next step) SHOULD be checked. Click Next.

4.9 Under Select output folder, navigate to the desired directory and click **Finish**.

5 Importing Control data into LIMS

5.1 Negative Controls

5.1.1 For each amplification negative control and extraction negative control, open the primary threshold analysis file for each sample, which has a filename ending in "-primary".

5.1.2 Click on the "Report" tab. Print/save the "Report" tab to a PDF file

5.1.3 Upload all negative control PDFs to the "Attachments" tab of the MiSeq test batch.

5.2 Positive controls

5.2.1 Open the secondary threshold analysis file for the positive control, which has a filename ending in "-secondary".

5.2.2 Click on the "Report" tab. Print/save the "Report" tab to a PDF file

5.2.3 Upload all positive control PDFs to the "Attachments" tab of the MiSeq test batch.