mitoCalc Tutorial

Prepared by Krista Opsahl-Ong, Yong Qian, TJ Butler, and Jun Ding
Updated on July 15, 2015

---

1. Compile BaseCoverageCalc.cpp file to generate the executable “BaseCoverageCalc”:
   `g++ BaseCoverageCalc.cpp -o BaseCoverageCalc`

2. Run the following command with your own specifications to analyze one individual’s sequencing data (i.e., one bam file):
   `perl mitoCalc.pl <in.bam> <workDir> <BaseCoverageCalcCommand>`
   - **mitoCalc.pl** is the perl file you will be running to execute the mitoCalc program. It requires the three arguments shown below (<in.bam>, <workDir>, and <BaseCoverageCalcCommand>). If you wish to run the program outside of your mitoCalc folder you can also run “perl /Path/To/Command/mitoCalc.pl” as the first part of the command.
   - **<in.bam>** is the individual bam file to be analyzed, and it can be in the format of /Path/To/bam/File/ID.bam. "ID" will be used as the identifier for output files.
   - **<workDir>** is the working directory, also formatted as “/Path/To/workDir”. This is a folder you should create before running the program, and is where the intermediate files and final output files will be generated.
   - **<BaseCoverageCalcCommand>** is the path to the executable "BaseCoverageCalc" in the format of “/Path/To/Command/BaseCoverageCalc”.

3. This full version of mitoCalc analyzes the mitochondrial genome and all 22 autosomal chromosomes to estimate mitochondrial DNA (mtDNA) copy numbers. It takes ~2 hours to analyze a bam file generated by whole-genome low-pass sequencing at ~4X average coverage for nuclear DNA.

4. After the analysis is done, intermediate files will be deleted automatically and the output file will be saved in <workDir> with file name “<ID>MTData.txt”. Here is a detailed explanation of the output:

mt_copy_number_avg: the estimated mtDNA copy number, calculated using the average of autosomal DNA coverage of 20 chromosomes, excluding the lowest and highest chromosomal coverages as outliers (In practice, this estimator has a correlation >0.999 with the one calculated using the average of all 22 chromosomes).

mt_copy_number_low, mt_copy_number_high: the lower and upper ends of the 90% confidence interval for mtDNA copy number. Twenty-two mtDNA copy number estimators can be calculated by using the DNA coverage for each of the 22 chromosomes. Therefore, a rough 90% (to be exact, 20/22 = 90.9%) confidence interval can be obtained by selecting the 2nd lowest and 2nd highest mtDNA copy number estimators.

mt_coverage: the average mtDNA coverage

autosomal_coverage: the autosomal coverage, calculated by averaging the coverage of 20 autosomal chromosomes, excluding the lowest and highest coverages as outliers.

chromNo avgCoverage: a list of 22 autosomal chromosomes and their corresponding DNA coverages (chromosomes are ranked by their DNA coverages from lowest to highest)