24/6/2020

**From:** Noelani Kamelamela <[hilo1@luria.mit.edu](mailto:hilo1@luria.mit.edu)>  
**Sent:** Friday, June 12, 2020, 5:19 PM

The data from your experiment are ready to be downloaded from the BioMicro servers.   
  
Here is a summary of the samples from this flowcell (4500T) sequenced as:  
300 + 300 bases pair-end run with 8 + 8 nucleotide indexes  
  
200219Seg/  
    D20-160001 : Axenic1 (GAACAATTCCTAGAGTTGGA)  
    D20-160002 : Axenic2 (TGTGGTCCGGAGAGCACTAG)  
    D20-160003 : Axenic3 (CTTCTAAGTCACTCTACAGG)  
    D20-160004 : Axenic4 (AATATTGCCACGGTGACACC)  
    D20-160005 : 2A1 (TCGTGCATTCGCGTTGGTAT)  
    D20-160006 : 2A2 (AAGATACACGTGTGCTAACA)  
    D20-160007 : 2C1 (TGCAATGAATCCAGAAGTAA)  
    D20-160008 : empty (CTATGAAGGACTTATACCTG)  
    D20-160009 : 2C2 (GAAGACTAGAACTAGAACTT)  
    D20-160010 : 5A1 (AGGAGTCGAGTTAGGCTTAC)  
    D20-160011 : 5A2 (TTCACTCACTTATCATGAGA)  
    D20-160012 : 5B1 (GGTCCGCTTCCTCACACAAG)  
    D20-160013 : 5B2 (CAACGAGAGCGAATTGAGTG)  
    D20-160014 : 5C1 (ATTGAGGTCCCGGATTATAT)  
    D20-160015 : 5C2 (GGAGAGACTCTTGAAGCAGA)  
    D20-160016 : 10A1 (CCGCTCCGTTTACGGCGAAG)  
    D20-160017 : 10A2 (ATACATCACATCTCCATTGA)  
    D20-160018 : 10B1 (TAGGTATGTTCGAGACCAAG)  
    D20-160019 : 10B2 (CACCTAGCACTGCTGGACAT)  
    D20-160020 : 10C1 (TTCAAGTATGGATGGTATCG)  
    D20-160021 : 10C2 (TTAAGACAAGGGCTTAATTG)

Sequences are downloaded from:

\* Use a SSH/SFTP client to connect to [bmc-150.mit.edu](http://bmc-150.mit.edu) with the following credentials  
username: segre\_ill  
password: an247bn4  
  
Download instructions are available at <https://openwetware.org/wiki/BioMicroCenter:FAQ#DOWNLOADING_DATA>  
  
The samples have been processed using the BMC/BCC 1.7 pipeline updated on 01/11/2019.  
- Upgraded cellranger from 2.0.2 to 3.0.1 for the 10X pipeline  
Details can be found at <http://openwetware.org/wiki/BioMicroCenter:Software#BMC-BCC_Pipeline>

**From:** Stuart Levine <[slevine@mit.edu](mailto:slevine@mit.edu)>  
**Sent:** Monday, June 22, 2020 9:47:16 PM  
**To:** Osborne, Melisa <[melosbor@bu.edu](mailto:melosbor@bu.edu)>

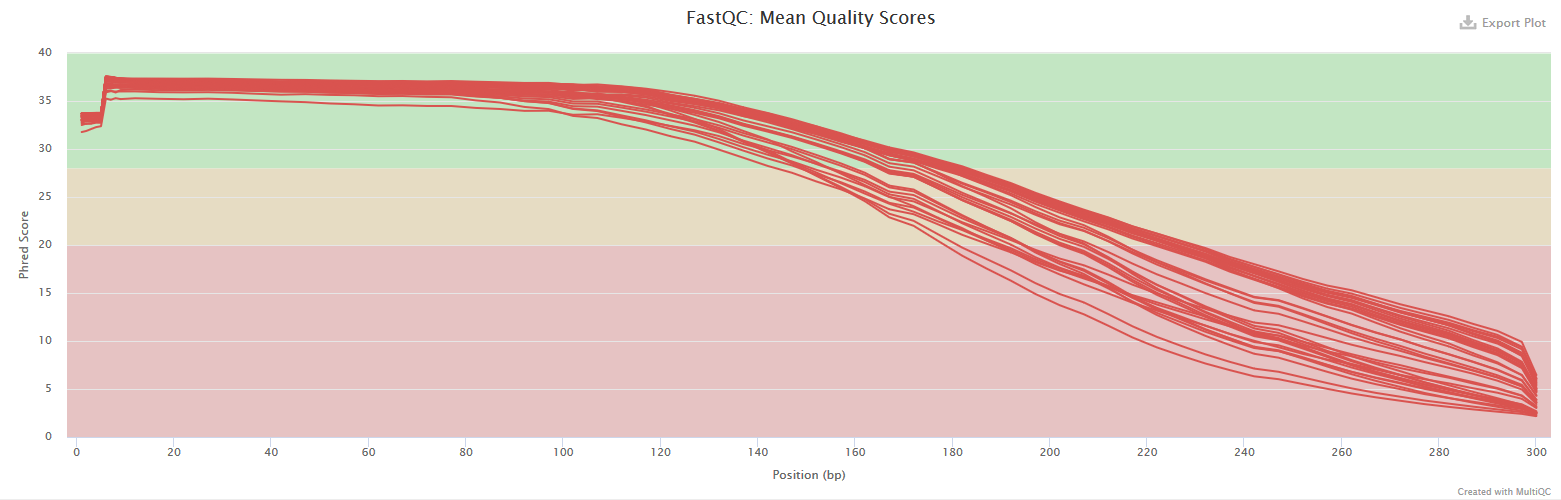
Fastqc should indicate fail as the samples are not shotgun human genome - we run it as well. Samples were prepped with Flex, which is about the best we can do for longer inserts - the solid matrix helps control insert size. I'm attaching our internal QC as a 7zip file. We do not do any trimming / pair merging of reads. This is not atypical for long miseq runs, several samples had very low input and did not perform well as Austin indicated. The run itself was robust in terms of error rate (QC plot at the top).

Samples:

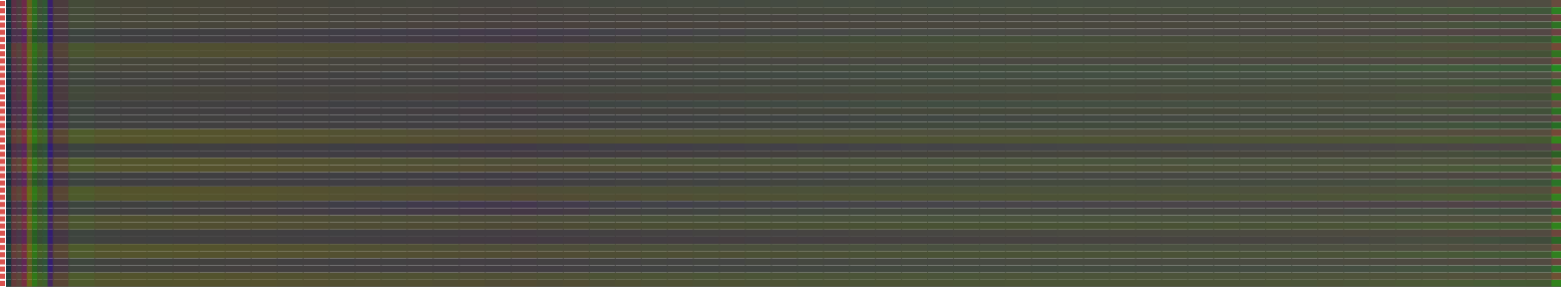
|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **#** | **sample** | **PRO** | **ALT** | **exp** | **days** | **ng/ul** | **final volume** |
| 1 | Axenic | 1A3 |  |  |  | 30 | 100 |
| 2 | Axenic | DE |  |  |  | 10 | 100 |
| 3 | Axenic | 9313 |  |  |  | 0.6 | 100 |
| 4 | Axenic | MIT0604 |  |  |  | 6 | 100 |
| 5 | 2A | 9313 | 1A3 | 2 | 100 | 0.7 | 100 |
| 6 | 2A | 9313 |  | 6 | 440 | 1.2 | 100 |
| 7 | 2C | 9313 | 1A3 | 2 | 100 | 1.5 | 100 |
| 8 | empty | empty | empty | empty | empty | empty | empty |
| 9 | 2C | 9313 | 1A3 | 6 | 440 | 8 | 100 |
| 10 | 5A | MIT0604 | 1A3 | 2 | 100 | 4.28 | 100 |
| 11 | 5A | MIT0604 | 1A3 | 6 | 440 | 16 | 100 |
| 12 | 5B | MIT0604 | 1A3 | 2 | 100 | 3.75 | 100 |
| 13 | 5B | MIT0604 | 1A3 | 6 | 440 | 7.3 | 100 |
| 14 | 5C | MIT0604 | 1A3 | 2 | 100 | 9.3 | 100 |
| 15 | 5C | MIT0604 | 1A3 | 6 | 440 | 19.3 | 100 |
| 16 | 10A | MIT0604 | DE | 2 | 100 | 5.85 | 100 |
| 17 | 10A | MIT0604 | DE | 6 | 440 | 14 | 100 |
| 18 | 10B | MIT0604 | DE | 2 | 100 | 1.6 | 100 |
| 19 | 10B | MIT0604 | DE | 6 | 440 | 6.25 | 100 |
| 20 | 10C | MIT0604 | DE | 2 | 100 | 5.34 | 100 |
| 21 | 10C | MIT0604 | DE | 6 | 440 | 18 | 100 |

FastQC/multiQC on all samples:

### Sequence Quality Histograms



**Per Base Sequence Content**



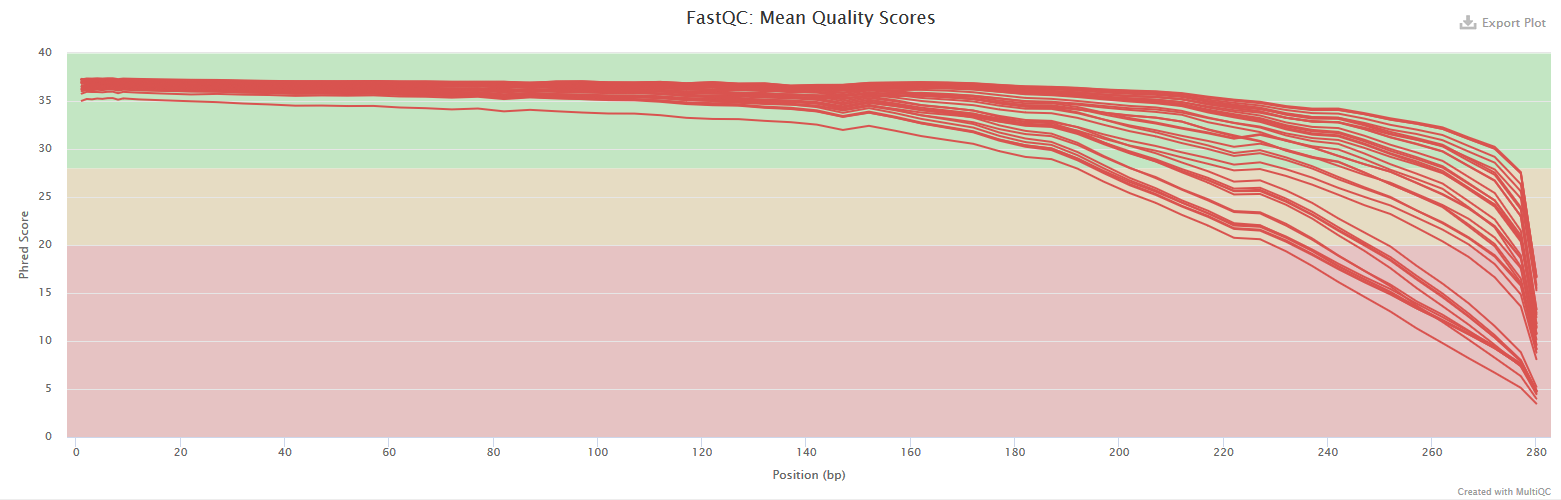
Decided to trim the first 20 bases. And not to trim the low quality at the end as advice from bwa mem developer is not to trim low quality and let bwa do soft clipping during run time.

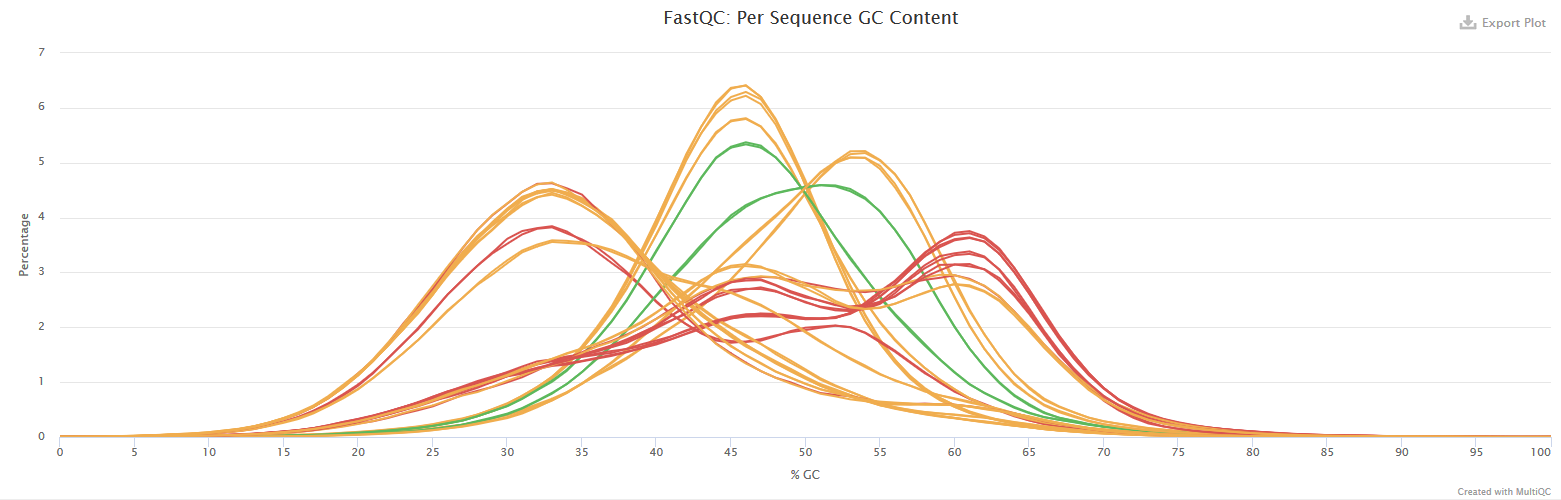
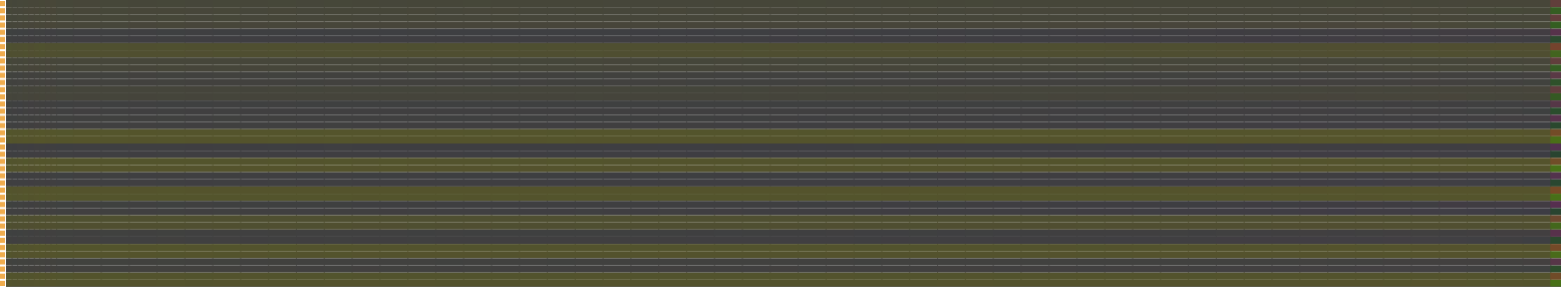
Ran the following command to trim adapters and barcodes from all fastq files:

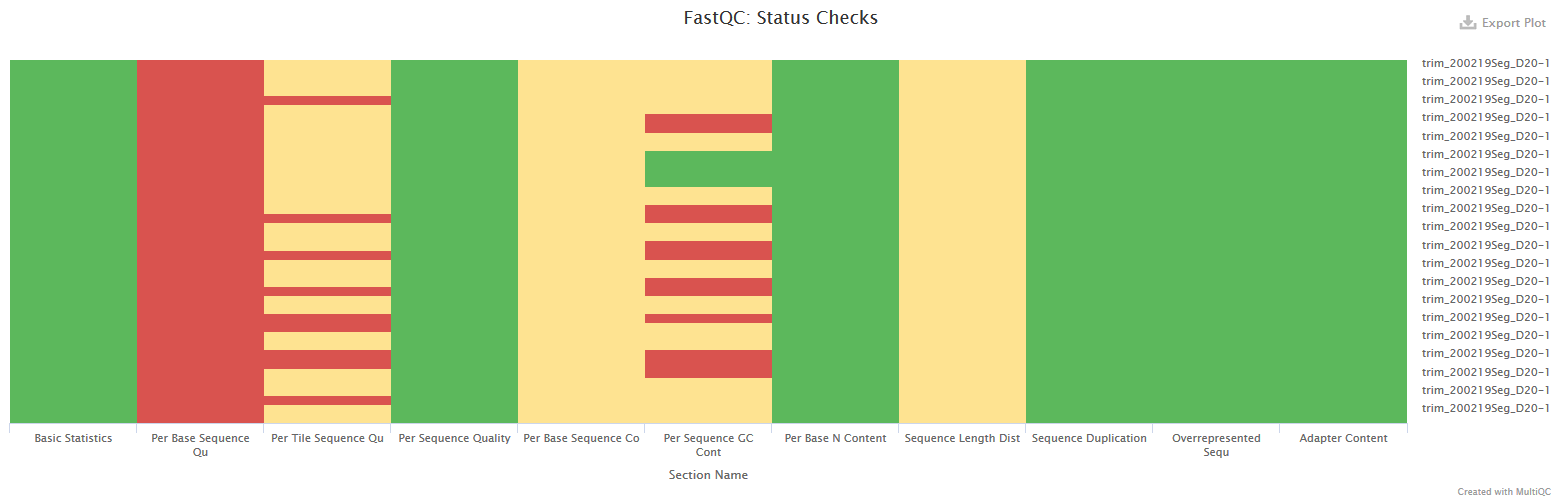
adapter=~/anaconda3/pkgs/trimmomatic-0.39-1/share/trimmomatic/adapters/**NexteraPE-PE.fa**

trimmomatic PE -threads 10 -trimlog ${outdpath}/trimmomatic.log -summary ${outdpath}/trimmomatic.log -validatePairs ${fastq1path} ${fastq2path} ${pout1} ${uout1} ${pout2} ${uout2} **ILLUMINACLIP:${adapter}:2:30:10:1:TRUE HEADCROP:20**

After trimming:







Download genomes:

Genomes were downloaded from NCBI:

|  |  |  |  |
| --- | --- | --- | --- |
| Species | Page | Accession | URL |
| DE | Alteromonas mediterranea DE, complete sequence | NC\_011138.3 | <https://www.ncbi.nlm.nih.gov/nuccore/NC_011138.3> |
| 1A3 | Alteromonas macleodii strain HOT1A3 plasmid pAM1A3, complete sequence | NZ\_CP012203.1 | <https://www.ncbi.nlm.nih.gov/nuccore/NZ_CP012203.1> |
| 1A3 | Alteromonas macleodii strain HOT1A3 chromosome, complete genome | NZ\_CP012202.1 | <https://www.ncbi.nlm.nih.gov/nuccore/NZ_CP012202.1> |
| MIT0604 | Prochlorococcus sp. MIT 0604, complete genome | NZ\_CP007753.1 | <https://www.ncbi.nlm.nih.gov/nuccore/NZ_CP007753.1> |
| MIT9313 | Prochlorococcus marinus str. MIT 9313, complete genome | NC\_005071.1 | <https://www.ncbi.nlm.nih.gov/nuccore/NC_005071.1> |

All genomes were indexed by

bwa index <genome>.fna