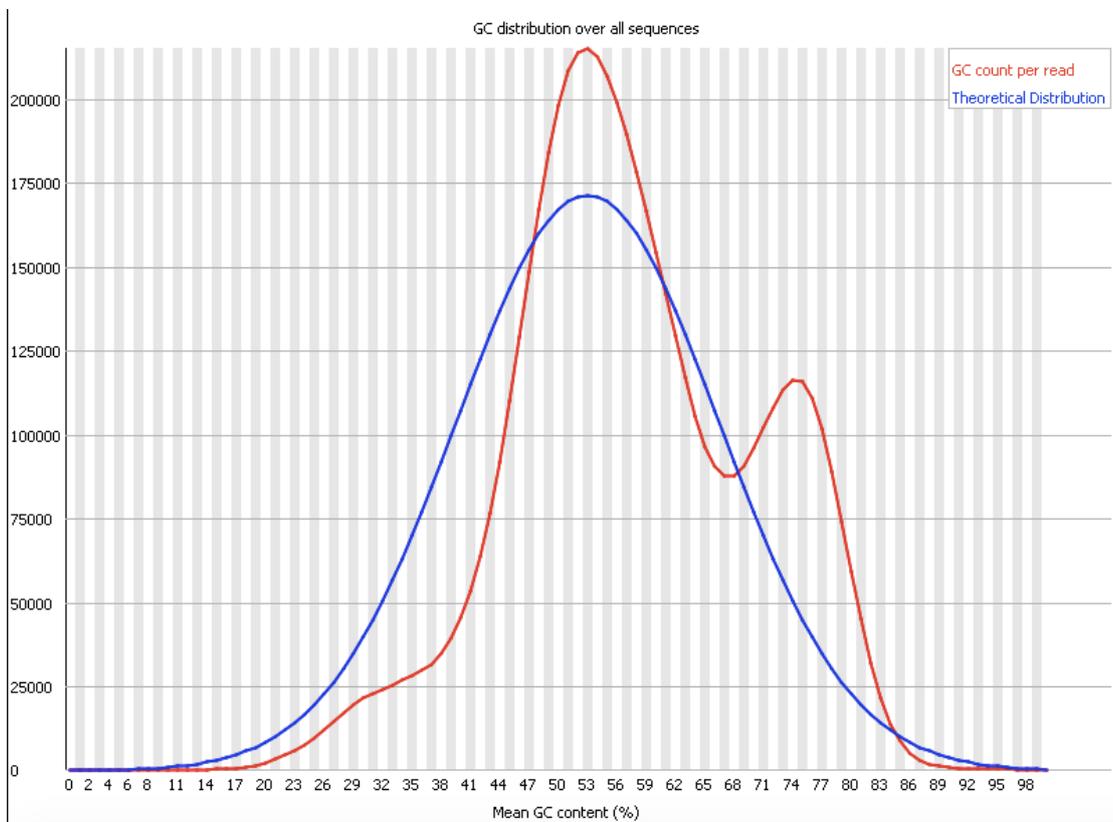


Multiple choice questions day 2 (for recap of day 1, IFAT form #E027)

1: What is the correct order of steps needed for 2nd generation sequencing, e.g., Illumina?

- A: DNA fragmentation, DNA isolation, amplification, primer ligation, sequencing
- B: Amplification, primer ligation, sequencing, DNA fragmentation, DNA isolation
- C: DNA isolation, DNA fragmentation, primer ligation, amplification, sequencing
- D: Sequencing, DNA isolation, DNA fragmentation, primer ligation, amplification
- E: DNA fragmentation, primer ligation, DNA isolation, sequencing, amplification

2: Have a look at the below figure, which shows the FastQC metric “Per sequence GC content”. What does the figure tell you?



- A: Based on the “Per sequence GC content” everything looks fine and there is nothing to be concerned about
- B: The two peaks in the red graph indicate the presence of two organisms with different GC content. I should be concerned about contamination with a different species
- C: The presence of adaptor sequence can cause the two peaks in the red line. The FastQC metric “Adaptor Content” is probably also raising an error.
- D: It looks like the sequencing reaction ran out of A/T nucleotides at some point. The run has to be repeated with a higher concentration of A/T nucleotides
- E: It looks like the sequencing reaction ran out of G/C nucleotides at some point. The run has to be repeated with a higher concentration of G/C nucleotides

3. Which sequencing technique does not require the activity of a DNA polymerase?

- A: Illumina
- B: No sequencing technique require the activity of a DNA polymerase
- C: PacBio
- D: Oxford Nanopore
- E: All sequencing techniques require the activity of a DNA polymerase

4: Which of the following sequencing techniques is considered 3rd generation sequencing (does not require template amplification)?

- A: Only Oxford Nanopore
- B: Only PacBio
- C: Illumina AND Oxford Nanopore
- D: Only Illumina
- E: Oxford Nanopore AND PacBio

5: What is the main reason the quality of the basecalls generated by Illumina sequencing decreases with increased sequence length?

- A: This is due to the DNA polymerase becoming more and more sloppy as time goes by.
- B: The DNA strands on each spot gets increasingly out of phase due to incomplete/missed base incorporation during previous sequencing rounds
- C: You should not see decreased quality of the basecall with increased sequence length. If you do, the sequencing run must be repeated.
- D: This is likely to be caused by the reagents having been used up (reagents exhaustion)
- E: This is due to the DNA strands gradually being degraded by the exposure to the laser

6: A file with raw sequence reads in FASTQ format contains 1.000.000 lines. How many reads does it contain?

- A: 4.000.000
- B: 1.000.000
- C: 500.000
- D: 250.000
- E: 100.000

7: Which of the following statistical measures describes the quality of a draft assembly?

- A: The phred score
- B: The FASTQ file
- C: The ASCII value
- D: The N50 value
- E: The homopolymer value

8. Below, the output when running whole genome sequence data from an *S. aureus* isolate through the MLST web-service, is shown. Although perfectly matching alleles are identified for all loci, the sequence type is reported as unknown. What is the likely cause of this?

MLST-1.6 Server - Typing Results

Sequence Type: *Unknown ST*

Locus	% Identity	HSP Length	Allele Length	Gaps	Allele
<i>arcc</i>	100.00	456	456	0	<i>arcc-7</i>
<i>aroe</i>	100.00	456	456	0	<i>aroe-6</i>
<i>glpf</i>	100.00	465	465	0	<i>glpf-28</i>
<i>gmk</i>	100.00	417	417	0	<i>gmk_-27</i>
<i>pta</i>	100.00	474	474	0	<i>pta_-11</i>
<i>tpi</i>	100.00	402	402	0	<i>tpi_-40</i>
<i>yqil</i>	100.00	516	516	0	<i>yqil-27</i>

A: The chosen MLST configuration does not match the species. This is likely not a *S. aureus* isolate.

B: This can only be explained by a bug in the method.

C: This could never occur.

D: This is likely due to too low initial quality of the sequence data. This can be confirmed by examining the N50 value of the draft genome.

E: Although all the alleles have been seen before and are included in the MLST database, the *combination* of alleles is previously unseen and hence reported as "unknown".

9: Which of the following statements about KmerFinder is NOT true?

A: The "Depth" metric is only relevant when draft assemblies have been uploaded and should be disregarded, when raw reads are used as input.

B: In the output table from KmerFinder, the template (reference) organism that has most unique kmers in common with the input genome is listed in the top.

C: To make the KmerFinder method faster, not all possible kmers in the input genome and in the template (reference) genomes are considered. Only kmers that start with a particular sequence (have a particular prefix) are considered in the comparison.

D: KmerFinder uses a larger proportion of the input genome for predicting the species than 16S rRNA based methods.

E: The query coverage is the proportion of unique, co-occurring kmers between the input genome and the template (reference) genome divided by the total number of unique kmers in the input genome.

10. The first figure below shows the output when uploading an *S. aureus* strain to ResFinder. The second figure shows the results, when the same file has been uploaded to PlasmidFinder. Which plasmid is the *blaZ* gene located on?

ResFinder-2.1 Server - Results

Aminoglycoside						
No resistance genes found.						
Beta-lactam						
Resistance gene	%Identity	Query/HSP length	Contig	Position in contig	Predicted phenotype	Accession number
<i>blaZ</i>	99.65	846 / 846	NODE_3_length_25945_cov_153.070877	8513..9358	Beta-lactam resistance	AP003139
Fluoroquinolone						
No resistance genes found.						
Fosfomycin						
No resistance genes found.						
Fusidic Acid						
No resistance genes found.						
Glycopeptide						
No resistance genes found.						

PlasmidFinder - Gram-positiv						
Plasmid	%Identity	Query/HSP length	Contig	Position in contig	Note	Accession number
<i>rep19</i>	99.69	975 / 975	NODE_3_length_25945_cov_153.070877	18648..19622	repA(pWBG759)	GQ900401.1
<i>rep7</i>	99.76	821 / 821	NODE_16_length_556392_cov_36.716381	211008..211828	rep(MSSA476)	BX571857.1
<i>rep16</i>	99.73	744 / 744	NODE_29_length_261626_cov_52.975273	10640..11383	CDS6(pSJH101)	CP000737
<i>rep5</i>	100.00	861 / 861	NODE_11_length_25953_cov_148.769653	12630..13490	rep(pMW2)	NC005011
<i>rep20</i>	98.89	990 / 990	NODE_18_length_19533_cov_249.624481	4639..5628	rep(pTW20)	FN433597.1

- A: rep19
- B: rep7
- C: rep16
- D: rep5
- E: rep20