Multiple choice questions for recap of Day 1

1: Which sequencing method does not make use of optical signals, but rather exploit the fact that addition of a dNTP to a DNA polymer releases an H^+ ion?

A: Oxford nanopore
B: Ion torrent
C: Sanger
D: 454
E: Illumina

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

A: Oxford nanopore
B: Ion torrent
C: Sanger
D: 454
E: Illumina

3: Which file format is seen below?

```
@M10_0139:1:2:18915:1321#ATCACG/1
TATCAAGAAAGATTTTAACAGCATTGACTCTGTTATCGAGTTTCATTTTAAACATAGTTTCCAGTGGT
+M10_0139:1:2:18915:1321#ATCACG/1
_bbeeeccgfgecgiiiihfhchiiiiiiiiihhfhhh^dghhhhf_fffghhhhhhacgeeghgbb]
@M10_0139:1:2:18915:1321#ATCACG/2
AGTTCATAGTGACAAGGTAATATTTGTCAAATTATATCGACCTAAAACGGTAGGATATATAACAAAAT
+M10_0139:1:2:18915:1321#ATCACG/2
a__eceeeeggffhihe^bhfiifh_edeg_agbgd]dd`g`fgdhedffaedadhhchhfhiicfhX
@M10_0139:1:2:12256:1321#ATCACG/1
ACGGGTGAACTGTACGGCATCGAAGCCCTTGCGCGCTGGCACGATCCCCAGCATGGTCATGCCCCCTC
+M10_0139:1:2:12256:1321#ATCACG/1
___`c_c`egge[bfghdeghfhhhhhfiii_ffhhN`ghhfddbcddadcddbccb_bbbcbc^aac
@M10_0139:1:2:12256:1321#ATCACG/2
AATCCGGAAAAGCCCGTACCAAAATCATCTACCGATAAGCCCACGCCCATATCACGCAGGATGAATCG
+M10_0139:1:2:12256:1321#ATCACG/2
a_ZcccWHO_bgadgc_WbaceZefda^f`egd`HO[ega\G\b`F_dggeca_cad`Y]^b__bKYZ
```

A: Sanger format
B: FASTQ format
C: Illumina format
D: ASCII format
E: FASTA format

4: 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

5: An *E. coli* genome with a size of 4,500,000 bp is sequenced. The raw output includes 1,000,000 reads with an average read length of 150 bp. What is the (depth of) coverage?

A: 6.8
B: 333
C: 675
D: 45
E: 33.3

6: Which file format is used for storing the sequence data of assembled genomes (draft and finished)?

A: Sanger format
B: FASTQ format
C: Illumina format
D: ASCII format
E: FASTA format

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: A draft genome consists of seven contigs. Their lengths are listed below. What is the N50 value of the draft genome?

```
Contig1:    2500
Contig2: 150.000
Contig3: 100.000
Contig4:  90.000
Contig5:  25.000
Contig6:     500
Contig7: 130.000
```

A: 498.000
B: 249.000
C: 150.000
D: 100.000
E: 130.000

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

A: In the output table from KmerFinder, the template (reference) organism that has most unique 16mers in common with the input genome is listed in the top.

B: To make the KmerFinder method faster, not all possible 16mers in the input genome and in the template (reference) genomes are considered. Only 16mers that start with a particular sequence (have a particular prefix) are considered in the comparison.

C: The “winner takes it all” scoring method of KmerFinder should not be selected when uploading raw sequence reads, only when uploading draft genomes.

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. 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
arcc	100.00	456	456	0	*arcc-7*
aroe	100.00	456	456	0	*aroe-6*
glpf	100.00	465	465	0	*glpf-28*
gmk	100.00	417	417	0	*gmk_-27*
pta	100.00	474	474	0	*pta_-11*
tpi	100.00	402	402	0	*tpi_-40*
yqil	100.00	516	516	0	*yqil-27*

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: 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".

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: This could never occur.