Additional sample exam questions:

Dot plots are fast and efficient when examining very large sequences

True False

RNA-Seq measures *relative* RNA abundance

True False

The first free living organism to have their genome sequenced was an isolate of *Escherichia coli*

True False

Multiple sequence alignments

a) are the basis for phylogenetic analyses
b) can aid identification of domains of a protein
c) cannot aid the identification of secondary or tertiary protein structure
d) two of the above are correct (A and B)
e) three of the above are correct

Define: BLASTP

Heuristic sequence comparison using a protein query against a protein database

---

Define: Paired-end sequencing

Sequencing two ends of a contiguous DNA fragment of known size

---

What are the benefits of RNA-seq over Microarray-based technologies for transcriptome analysis?

No reference sequence needed

With microarrays, limited to the probes on the chip

Low background noise

Large dynamic range

High technical reproducibility

What are two pitfalls of the RMSD measure for protein structure comparison?

- all atoms are treated equally, either in the core of the protein or at the surface (when in reality the surface of the protein should have more freedom to vary since such components, like surface exposed loops, can move more)

- does not take into account the attributes of the amino acids (i.e. if they are charged or not)

9. a) Describe the how Amplicon sequencing differs from Metagenomics sequencing for microbiome profiling.

Amplicon sequencing involves profiling a microbiome using amplified marker sequences, such as sequence regions from 16S rRNA genes. It produces a taxonomic profile of the microorganisms in the community being measured that has that particular marker sequence. It is cheaper and faster to perform versus metagenomics sequencing and can utilize a large database of known markers genes to identify the taxa in the sample.

Metagenomics sequencing involves sequencing all DNA in a sample to profile the compilation of genomes from all organisms in a microbiome. It can produce both a taxanomic profile and a gene profile, and is needed when studying communities that do not have a suitable marker sequence for Amplicon sequencing. However it is not as cheap and quick as Amplicon sequencing.

b) Which method would you use if studying a whole viral community? Why?

I would use Metagenomics, since there is no amplicon marker that is conserved in all viruses. So an amplicon-based approach would not characterize a whole viral community, but rather just a subset of virus with that particular marker gene.