# Answers to Deliverable VI

1. What does multiplex sequencing mean? Why do people use multiplex sequencing? (2 point)

Multiplex sequencing allows sequencing libraries to be pooled and sequenced simultaneously during a single run on most high-throughput instrument.

Running several samples on one sequencing run

– Saves money & time

– Eliminates sequencing run as variable

1. Give at least two reasons as to why we cannot just do exact string matching when aligning our reads to reference sequences (2 points)

We have mismatches and indels. It would be way too slow

1. The Burrows-Wheeler Transform creates an alfabetically sorted suffix array and a Full-text index in Minute space (FM-index). Mention some important effects of having the data sorted and compressed (2 points)

Sorted data is faster to search and compressed means less storage & memory usage

1. BWA is a program that uses the Burrows-Wheeler Transform in two different implementations: aln and mem. How does each method work? drawing is encouraged (3 points)

bwa aln: First ~30nt of read as seed

bwa mem: Multiple short seeds across the read

1. Why does the average sequencing depth we observe when mapping reads to a reference genome not always reflect the actual coverage of the reference genome? (1 point)

Often we have VERY uneven coverage meaning that even with an average of 90 times coverage we can observe that 50% of the genome is not covered at all by reads.

1. Why is the 16s rRNA gene used for microbiome phylogenetic analysis? (1 point)

It is found in all bacteria and archaea. It has conserved regions perfect for amplicon primers and variable regions allowing the separation of different bacteria.

1. Why does different primers for 16s rRNA amplicon analysis matter so much? (1 point)

Even the very conserved regions are not identical and different primers will have different affinity to different bacteria

1. What is the difference between an Operational Taxonomic Unit (OTU) and Amplicon Sequence Variant (ASV) (2 points)

OTU describes a cluster of similar sequences (often 97% identical) while ASVs are the result of denoised data and clustering is not needed.

1. In the following figure. Are we looking at alpha or beta diversity? Have we sequenced deep enough to describe the different samples? ( 2 points)

Species richness is a measure of alpha diversity within-a-sample. Some of the rarefaction curves are still climbing, which indicates that we need to sequence deeper to get a full picture

1. Shannon index incorporates the species ... and ... of a sample (2 points)

Richness and eveness

1. When accounting for different sample sizes. What does it mean to rarefy samples? What effect does sample rarefication not take into account (2 points)

Rarefication or downsizing means to resamples an equal amount of observations or reads from each samples. The more reads we keep the more sensitive. We may have to remove samples with few counts. We might throw away a lot of data

Rarefication does not take into account that we are doing a compositional analysis, meaning that two environmentally independent species can affect the read count of each other.