**MIP 280A4: Microbial Sequence Analysis**

**Mapping, part 1, In-class exercise questions**

1. What are the two main inputs to read mapping software? (1 pt)
2. The diagram below shows Illumina reads mapped to a virus reference sequence. (1 pt each)



* 1. What is the coverage depth of position 20 in the reference sequence?
  2. What is the average (mean) coverage depth of positions 1-10 in the refrence sequence?

1. Answer the following questions about the paper-based mapping exercise (1 pt each)
   1. What was the approximate average coverage depth for your mapped reads?
   2. What was the maximum coverage depth?
   3. What was the minimum coverage depth?
   4. Was coverage across the ‘genome’ even?
   5. What percent (approximately) of the genome was covered by at least one read?
   6. Were all of your reads mappable?
   7. What was the source of unmappable reads?
   8. In a real sequencing dataset, what is one possible source on unmappable reads?
   9. What fraction (approximately) of reads mapped unambiguously (uniquely)?
   10. Did you identify any sequencing errors?
   11. Did you identify any variants (SNPs)?
   12. How would you distinguish a real variant from a sequencing error?
   13. What was your approximate mapping speed (how many reads per minute did you map)?
2. From the example shown in the lecture of 186,000 reads mapped to the *D. melanogaster* genome answer the following questions (1 pt each):
   1. 10.9% of the reads didn’t map. What are two possible sources of these reads?
   2. 42% of the reads mapped non-uniquely. What is the likely source of these reads?
3. If a read maps to a reference genome sequence with a mapping quality of 40, what is the probability that it is not mapped to correct location? (1 pt)
4. If a read maps to a reference genome sequence with a mapping quality of 10, what is the probability that it is not mapped to correct location? (1 pt)