Key for BioConcept Questions, Chapter 8

1. In Sanger sequencing, why does a newly synthesized strand of DNA terminate when DNA polymerase inserts a dideoxy nucleotide? How are these terminated DNA strands used to "read" the nucleotide sequence of the original DNA molecule?

   A dideoxy nucleotide is exactly like a normal DNA nucleotide (A, T, C or G), except that it lacks the 3′ –OH group needed to attach the next nucleotide in the chain. Thus, when DNA polymerase inserts one of these dideoxy nucleotides, nothing further can happen with that chain and synthesis terminates.

   With a low concentration of dideoxy nucleotides and many DNA polymerases attempting to copy many copies of the DNA template strand, eventually we will have a dideoxy-terminated strand corresponding to every possible nucleotide in the original DNA strand (see Figure 8.11). Because they come out in order on a gel and because each type of dideoxy nucleotide has a unique fluorescence wavelength, the terminating dideoxy bases can be "read" in order as DNA fragments move down a gel.

2. Why is shotgun sequencing so much faster than the directed approach originally taken by the IHGP? Why is it more dependent on computer power and bioinformatics?

   The directed approach requires reading one segment of DNA sequence in order to know the sequence of a primer that should be synthesized to read the next segment as we move in order down a chromosome. The shotgun approach merely sequences any random fragment of the chromosome without regard to order and allows a computer program to assemble the sequences to give the original chromosome sequence. This is much faster but couldn't be done until computing power and algorithm development made it possible to assemble a continuous sequence from a huge number of random fragments.

3. If the entire human genome were cleaved into a single set of small, non-overlapping fragments, we could not determine the genome sequence by sequencing the fragments. Explain why this is the case.

   There would in this case be no way to know whether fragment A is followed by fragment B or perhaps by C or D or E: any fragment could be placed next to any other fragment with no means of deciding which assembly is correct.

4. How do next-generation sequencing techniques extend and improve upon the shotgun sequencing technique? What are their disadvantages?

   Next-generation sequencing is essentially shotgun sequencing in which far more sequences can be generated at once from a single run of a sequencing instrument. But, the length of the reads is sacrificed to get high throughput, meaning that there will be fewer and smaller overlaps and correct assembly becomes even more dependent on a good assembly algorithm. Additionally, some next-generation techniques offer notably lower accuracy in generating their sequences and so become much more dependent on high coverage.

5. Complex genomes often contain many repeated sequences. For example, there are many STR (short tandem repeat) sites in the human genome, where a short sequence such as GATA might be repeated anywhere from a few to dozens of times. Why would an STR region potentially pose a problem for sequencing? Are next-generation techniques more or less susceptible to errors resulting from repeated sequences?

   Suppose the original DNA sequence is CCCCCGATAGATAGATAGATACCCCCCC. If we get fragments like CCCCCGATAGATA and GATAGATACCCCCCC, we wouldn't know how to assemble them: we might overlap them by both of the GATA repeats or by one of the repeats, both of which would lead to wrong assemblies in this case. The problem is mitigated if we get fragments that span the STR region and include unique sequences on both sides, but this is less likely to occur with next-generation techniques that produce very short reads.