Week 10

Chapter 15:

Questions on Chapter 15 material: the answers for these will be covered in lecture.

1. Why does repetitive DNA pose problems for genome sequencing? There has been a major push to develop technologies that will sequence longer fragments. How could this help?

ANS: We sequence genomes by first breaking copies of the genome into many short fragments, and sequencing those fragments. It will be difficult to assemble our short fragments in repetitive regions of the genome, as our sequenced fragments may map to multiple locations in the genome. Longer fragments are less likely to align to multiple places in the genome, especially if they are longer than the length of repeats, and so are easier to assemble.

2. You sequence a 100kb region of the *Bacillus anthracis* genome (the bacteria that causes anthrax) and a 100kb region from the *Gorilla gorilla* genome. What differences or similarities might you expect to see in the annotation of the sequences- for example, in number of genes, gene structure, regulatory sequences, repetitive DNA?

ANS: The 100-kb *Bacillus* genome sequence will contain many more annotated genes than the 100-kb *Gorilla* sequence because prokaryotic genes are more compact, contain short regulatory sequences, do not have introns, and are packed tightly together; in contrast, eukaryotic genes are larger due to the presence of large introns, contain larger regulatory regions, and are separated by large intergenic DNA sequences. The *Bacillus* contain some operons and little repetitive DNA, whereas the *Gorilla* segment will lack operons and contain interspersed repetitive DNA.

General concepts: Bacterial genes are short and do not contain introns, some bacterial genes are organized in operons where multiple genes have shared regulatory sequences, the gene regulatory sequences are short, and the genes are closely packed together with little to no intergenic DNA. Bacterial genomes do not contain much repetitive DNA. Eukaryotic genes are large and consist often of short exons separated by large introns; there are no operons and the genes contain larger, complex regulatory regions. Eukaryotic genomes contain many interspersed, repetitive DNA elements.

3. Genome size does not correlate with organismal complexity. How does the existence of transposable elements and other sources of repetitive DNA help explain this observation?

ANS: Many eukaryotic organisms have genomes that are bloated with transposable elements (TEs), and other sources of repetitive DNA. The number of these repeats is the best prediction of genome size, large genomes are often due to large expansions of TEs.

4. Imagine that you sequence one organism with a 300 million basepair (Mb) genome and one with a 300 billion base pair (Gb) genome. If you generate one mutation at random in both of these genomes, in which species is it more likely to have functional consequences?

ANS: The small genome, as more of it is functional (less of it is TEs and other repetitive elements, see previous answers), and so mutations are likely to be deleterious.

5. Below is a short region of homologous sequence from 3 species:

```
Species 1: ATTCGAATTCGAACCTAGT
Species 2: GTCCGAATACGAAGCGAGG
Species 3: ATCCGAATTCGAAGCTAGT
```

A) How many changes have occurred between species 1 & 2?
B) Which two species are likely to more closely related?
C) You sequence a fourth species in this genomic region and find that it is shorter as this species has a deletion in this region. The sequence you obtain for species 4 is: `ATCTGAAGAAGGTAGT`. What is the best alignment of this sequence to the others? Which species is it closest to?

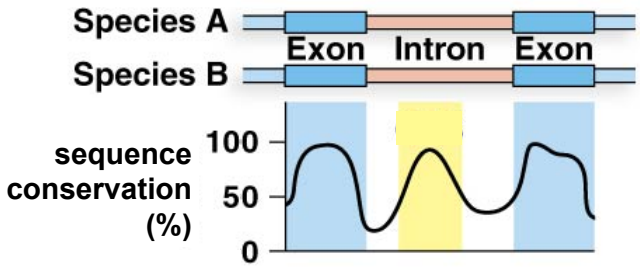
ANS: **A)** 6 ; **B)** Species 1 and Species 3. There are 6 differences between Species 1 and Species 2, but only 2 between Species 1 and Species 3. **C)** Best alignment is below, this new species (Species 4) is closest to species 3. Disregarding the deletion in Species 4, there are 2 differences between Species 4 and Species 3, and 4 differences between Species 4 and Species 2, and all the positions where Species 1 and 3 differ from one another, Species 4 has the same base as Species 3.

```
Species 1: ATTCGAATTCGAACCTAGT
Species 2: GTCCGAATACGAAGCGAGG
Species 3: ATCCGAATTCGAAGCTAGT
Species 4: ATCTGAA---GAAGGTAGT
```

6. How can evolutionary conservation be used to identify functionally important regions of the genome?

ANS: Sequences that are conserved in the genomes of two or more species are more likely to be functional than sequences that are not conserved. If a sequence is functional, mutations will be more likely to have functional consequences, and thus will be more likely to be removed by selection. This will result in higher sequence conservation of this region.

7. From the figure below, what can you infer about the sequence present in the intron between these two exons? Propose a possible function for this sequence in the intron.



ANS: The sequence in the intron is likely to be functional, as it has a high degree of sequence conservation between the two species. The level of conservation is as high as the exons of the gene. This sequence is likely to be important for gene expression, perhaps there is an enhancer sequence present in this region, since enhancers can be present in introns.

8. You sequence two species, species A and species B. In species A you find two genes A1 & A2 which are homologous to a gene B1 in species B.
A) Are A1 & A2 better described as paralogs or orthologs?
B) Are A1 & B1 better described as paralogs or orthologs?
C) A1 & B1, have introns, and are found at the same syntenic position in the genomes of species A and species B. But the gene A2 resides on a different chromosome and lacks introns. What is the most likely mechanism of duplication of the A2 gene?

ANS: A) paralogs, B) orthologs, C) retrotransposition

9. What are three potential evolutionary fates of a newly duplicated gene? Briefly describe each one.

ANS: 1) Most newly duplicated genes will be non-functional, due to failure to also duplicate important regulatory regions, or even the whole coding region of the gene. These duplicates will degrade (acquire mutations) over evolutionary time. 2) If a gene performs multiple functions, one possibility after duplication would be subfunctionalization, where the two copies evolve to divide the ancestral functions between them. 3) Another potential evolutionary outcome after duplication is that one copy can maintain the function of the original gene, whereas the other can evolve a novel function. This is referred to as neofunctionalization.