1. **Contigs** - (A substring of the genome) or a set of overlapping DNA segments that together represent a consensus region of DNA. In bottom-up sequencing projects, a contig refers to overlapping sequence data.
2. **Trie** - (A rooted tree containing all of the patterns in the set) or a data structure that is also known as a digital search tree or a prefix tree.
3. **Hamiltonian path** - a path called that visits each vertex exactly once.
4. **Eulerian path** - a path in a finite graph that visits every edge exactly once (allowing for revisiting vertices).
5. **Codon** – (A triplet of nucleotides) or a series of three nucleotides (a triplet) that encodes a specific amino acid residue in a polypeptide chain or for the termination of translation.
6. What is the Burrows Wheeler Transform of ACTA$ ?
   * AT$AC.
7. Which data structure uses the most memory? Suffix Trie, Suffix Array, BW Transform
   * Suffix Trie.
8. Find the cyclic spectrum of the following peptide: LND

* 0 113 114 115 227 228 229 342

1. Find the linear spectrum of the following peptide: SPV

* 0 87 97 99 184 196 283

1. What is the hamming distance between two strings?

ACTAAGACTCAGGG and ATGCCGACTCGAGG

* + 6.

1. Given the last column of a Burrows Wheeler Transform, what is the original sequence?
   * Last column is t$acc
   * Original sequence is $acct
2. What are non-standard amino acid masses?
   * Masses that are not found in the standard 20 amino acids, but come from other chemical processes.
3. Which experimental spectrum has the highest score compared to the theoretical spectrum?
   * Use Code.
4. How would you use run length encoding to compress the following string of DNA?

AAAAAAAAAACCCTTTTTTTTTTTTTTGGGG

* + 10A3C14T4G

1. What are all possible 4-mers of CGATTACTG?
   * ACTG, ATTA, CGAT, GATT, TACT, TTAC
2. What is the reverse complement of a DNA string?
   * A string encoded in the opposite direction with G’s converted to C’s and A’s converted to T’s.
3. Why do we often test a DNA sequence and its reverse complement when searching for kmers in DNA?
   * In order to improve detection, in some cases the kmer might appear in the original sequence, but in other cases it might be in the reverse complement.
4. What are the modes of action by which antimicrobial peptides kill microbes?
   * Toroidal pore, carpet, and barrel stave.
5. How do you determine if a graph is Eulerian?
   * If it is balanced and strongly connected.
6. What do you call it when ants run away together?
   * Antelope.
7. What is an Okazaki fragment?
   * Short, newly synthesized DNA fragments that are formed on the lagging template strand during DNA replication.
8. What is the central dogma of Molecular Biology?
   * (DNA is converted to RNA which is converted to protein) or the two-step process, transcription and translation, by which the information in genes flows into proteins: DNA → RNA → protein.
9. How many reading frames are there for a stretch of DNA
   * 6.
10. What mechanism is responsible for Transcription?
    * RNA Polymerase.
11. What mechanism is responsible for Translation?
    * Ribosome.
12. How do you calculate the mass of a peptide?
    * Add up the mass of all amino acids in the peptide.
13. How does LeaderboardCyclopeptideSequencing prevent the number of possible solutions from growing exponentially?
    * It only keeps the top possibilities as peptides are extended.
14. What is the Spectral Convolution?
    * The positive difference between every pair of masses in the spectrum.
15. What is ConvelutionCyclopeptideSequencing?
    * Only use the most frequent elements from the convolution between 57 and 200 in the LeaderboardCyclopeptideSequencing algorithm.
16. How much memory is needed to hold a suffix trie given that there are k patterns of maximum length m and a genome of length n?
    * O(n^2).
17. Which molecular machine is responsible for the replication of DNA?
    * DNA Polymerase.
18. What are the differences between a trie and a suffix tree?
    * Trie - A rooted tree containing all of the patterns in the set.
    * Suffix Tree - a compressed trie containing all the suffixes of a given text as their keys and positions in the text as their values.
19. Given the last column of a BWT and the positions of each cyclic rotation, what positions, if any, does a substring occur in the original genome?
    * Use code, inverse function, position is from original BWT starting at zero.