Question 2.1 - What scores and statistics BLAST use? What does each one means and how is it calculated?

- The bit score is a measure of how good the alignment is. The higher the score the better the alignment. The bit score is calculated using the substitution matrix. Bit scores are normalized, so they can be compared no matter the alignment or scoring matrix. The E-value or expected value reflects the statistical significance of the alignment. It takes into account both the size of the database and the scoring system used. The specific E-values reflect the probability that a given alignment can occur alone.

Question 2.2 - What is the difference between BLAST, BLAST 2, PSI-BLAST, and FASTA? When should each of these been used (trade-offs)?

- BLAST is the traditional report designed for readability by a user rather than a program. It contains a header, a one-line description of each database, and the alignments of each database sequence matched. BLAST 2 provides the user with a more formal report that integrates the hit table. The hit table is more simple and clean since it only contains information that is absolutely necessary. BLAST 3 integrates a structured output that allows for automatic and rigorous checks for syntax errors and changes. PSI-BLAST is used for iterative protein-sequence similarity searches. It implements a position-specific scoring matrix. It is an adaptation to BLAST. FASTA is the first widely used algorithm used for similarity searches of protein or DNA sequences. It is also a format of a nucleic acid or protein structure.

Question 2.3 - Explain what is the p-value for the top match of the following query

- The E-value of the query is 2e-117, which is very close to 0. So, this means that the probability of getting a match alone or on accident is miniscule. So, the significance of the match is very high. If we want to find the p-value, we use the equation Dat HD:private:var:folders:kh:0s3vhc653sqb1c9sbrby3w6h0000gn:T:TemporaryItems:(5).gif

which gives us a value very close to zero, so we can assume it is zero.