**Sequence Alignment, Homology, and Paralogy**

**1) BLAST**

* BLAST= Basic Local Alignment Search Tool
* Simply analyzes local parts of unknown genes by aligning them against already known sequences.
* Query- unknown sequence
* Subject- known sequence
* Involves the use of Bioinformatics to compare the query to other already published similar sequences that have already been discovered (the GenBank).
* By comparing genomes we have recently discovered that genome size doesn’t correlate with an organism’s complexity
* ex: the Whisker fern has more genes than humans, but it isn’t more complex
* The results from BLAST help us create phylogenetic trees to classify organisms.
* BLAST was mentioned before when comparing mRNA with the DNA it was transcribed from. This showed us that mRNA contains missing segments, called introns, that were found in the DNA
	1. **BLAST Steps**
1. The sequence is broken up into short “words” that are around 11 bases long. The longer the sequence, the more chances there are of finding a match due to its increased specificity.
2. The sequence is then compared to the the massive amounts of subject sequences already in the GenBank.
3. After finding matches, the search is extended above (upstream) and below (downstream) the sequences. Look at the worksheet from class to see an example of this.
4. An E value is then given to each alignment which indicates how good of a match the sequences were.
	1. **BLAST terms**
* Accession numbers- unique identifiers given to a sequence when it is entered into the GenBank.
* Identity- percent of perfect matches between subject and query
* E Value- the “expectation value” - number of alignments with this query sequence that would be expected to occur by chance with this database (similar to P value – the smaller the better)
* The closer to 0, the less errors were found, the more related the organisms are.
* Indicates whether or not the organisms are from the same species or share a common ancestor.
	1. **Reasons Why the Sequences Don’t Match**
* SNP’s- Single Nucleotide Polymorphisms
* INDEL’s- Insertions or deletions of nucleotides
	1. **Homology**
* Homologies- having shared ancestry due to similar structures or genes found in different species. “Descent with modification”
* Morphological Homologies- similar bone structures suggest a common ancestor.
* Homologous Structures
* There is an evolutionary relationship from an ancestral species.
* Examples: Forelimbs, Noses
* Homologies between Tetrapods: Humerus, radius, ulna
* Homologies between animal and Plant cells: Cell membrane, mitochondria, ribosomes, vacuole, nucleus, etc. (EXCEPT Cell wall, Chloroplasts, Centrioles)
* Homologous Bases: individual nucleotides that match in the alignment
* Ex: CO1 gene is homologous for every organism with a mitochondria
* Ex: Similiar nucleotides sequences between primates (gorillas, Homo, Pongos)
* Vestigial Structures: homologous features that lose their usefulness
* Dew claws, birds with teeth, teeth of fetal baleen whales.
	1. **Paralogy**
* Paralog: different gene families within the same species
* When a gene duplication occurs and genes evolve.
* Replication causes mutation.
* Paralogs are a type of homology.
* Example in Sickle Cell Anemia
* Alpha and Beta Hemoglobin are paralogs
* Hemoglobin and Myoglobin are paralogs because myoglobin was created by a duplication in the production of hemoglobin.
	1. **Analogy**
* Analogy: having a structure of similar function between unrelated species.
* There is no evolutionary relationship between the two organisms because they evolved independently of one another.
* Convergent Evolution: when two organisms evolve parallel of one another
* when organisms evolve similar solutions to similar problems
* There is no common ancestor
* Same analogous structure
* Example: Bat and Bird wings, the streamlined bodies of dolphins and sharks

**Essay Questions**

1. Explain the steps of how BLAST works and how it contributes to bioinformatics.
2. Explain the difference between homologies and analogies while also giving examples for both.

**Multiple Choice Questions**

1. What does BLAST stand for?
2. Basic Legitimate Alignment Sequences Tool
3. Basic Ligament Alignment Search Tool
4. Basic Local Alignment Search Tool
5. Bioinformatics Local Alignment Search Tool
6. The higher the E value, …
7. The more the organisms are related
8. The less amount of SNP’s and INDEL’s were found
9. The less likely the organisms are related
10. The more sequence base nucleotides were used
11. A specific example of a paralog are
12. bat and bird wings
13. hemoglobin and myoglobin
14. cats and human forelimbs
15. the streamlined structure of dolphins and sharks
16. What is a homology?
17. When organisms share similar anatomical structures, but are not from the same common ancestor.
18. When organisms come from the same common ancestor, and share similar anatomical features.
19. When organisms evolved convergent of one another.
20. When organisms come from the same species, but have different gene families.
21. When genes are BLASTed, the unknown sequences known as the \_\_\_\_\_\_\_ is aligned against the known sequences referred to as the \_\_\_\_\_\_\_\_\_.
22. subject, query
23. query, subject
24. RNA sequence, DNA sequence
25. DNA sequence, RNA sequence

**KEY**

Multiple Choice

1. c
2. c
3. b
4. b
5. b

Essay Question 1

Part 1: BLAST Steps

* Use PCR to determine the DNA gene sequence
* Type in the gene sequence, known as the query
* BLAST database aligns the query against subject sequences to determine similarity between organisms
	+ Compares small part of sequence and expands it to find organisms most specific to query, upstream extends the search above and downstream extends the search below
	+ Expanded sequences help specify the similarity between sequences in organisms
* E value is the “expectation value” that predicts the number of alignments with a certain query sequence that would occur by chance within the database, the smaller the E value the better
* Identity value is the percentage of perfect matches between subject and query

Part 2:

* BLAST sequences help in bioinformatics by providing DNA evidence of various organism’s relationship to each other
* Help create phylogenetic trees based on the SNPs and Indels present between gene sequences
* E values and Identity values also help determine how related an organism is

Essay Question 2

Intro

* Due to advances in technology, like DNA barcoding, scientists have been able to classify a multitude of organisms.
* They can be separated into groups as big as domains, or as small as species.
* However, before DNA sequencing the way scientists classified organisms was based on their anatomical structures.
* This technique was helpful in some cases, but not for others.

Body Paragraph I : Homologies

* Observing anatomical features was useful when the structures were homologous.
* This is because a homology is when organisms from different species have similar structures (or genes) and share a common ancestor.
* Therefore the anatomical evidence can help prove two organisms have an evolutionary relationship which can help accurately place them on a phylogenetic tree.
* An example of this is the forearms of cats and humans. Both have similar bone structures which indicates they have a common ancestor.
* In this case, that is true, however this does not always happen.

Body Paragraph II : Analogies

* Sometimes organisms appear to have similar anatomical features, but they don’t have an evolutionary relationship. This is called an analogy.
* What happens in this case is referred to as convergent evolution which is when organisms evolve parallel of on another.
* The reason why this occurs is because two separate species evolve similar solutions to the same problems.
* An example of this are the wings of bats and birds. The bone structures are the same because they both adapted the ability to fly. However, it has been proven that they don’t come from a common organism.

Conclusion

* Due to this, it shows that anatomical features can be helpful to a certain point if they are homologous structures. However, in situations where analogies occur, it is best to refer to DNA sequencing to accurately classify organisms.