**Your submission should be a PDF or a Word document (.pdf, .doc, .docx)**

**For your answers, you are welcome to use any layout you would prefer as long as I can tell which answer matches each question.**

**PART I:**

For this section, you will be using the Basic Local Alignment Search Tool (BLAST), which can be found at the National Center for Biotechnology Information (NCBI) website: <https://blast.ncbi.nlm.nih.gov/Blast.cgi>

BLAST is a program that reports regions of local similarity (at either the nucleotide or protein level) between a query sequence and sequences within a large database. The ability to detect sequence homology allows us to determine if a gene or a protein is related to other known genes or proteins. Detecting sequence homology also facilitates the identification of conserved domains that are shared by multiple genes and the identification of members of a gene family. BLAST is based on a robust statistical framework. This framework allows BLAST to determine if the alignment between two sequences is statistically significant (i.e. whether the probability of obtaining an alignment this good or better by chance alone is low).

For those of you that are unfamiliar with how BLAST works, an interactive tutorial for beginners is available here: <https://digitalworldbiology.com/blast>

On the main BLAST website, there are links to four main types of BLAST searches. Each type has a brief description below the name.

 ***Question 1: What are the four types of BLAST searches and their descriptions?* (2.5 points)**

 ***Question 2: How do these descriptions describe the differences between the BLAST searches?* (2.5 points)**

Run a BLAST-N (Nucleotide BLAST) using the nucleotide sequence below.

Keep all parameters set at default but double check that the Database is set to “Nucleotide collection (nr/nt)” and the Program Selection is Optimized for “Highly similar sequences (megablast)”.

CAGAGAGGGTGCTTGTCCAAGTCCCGGCACTACCCCGATAGTGTAGAAGGGGAGCCAAGGGAAGGTCAGGCAGAGAAGGTCCATCCCCAGGTCCGAGTGCTCTCTGCAGCAGGCATGGCCTCGGTGGTCACACGACCCTTCCCGAGTGCCCCCCTGCATCTCCGCCCACGTCTGTCTCCGTTTCTGCCATGGTCTCCCGCTCACCCTTGCCTCTGCTCATGGTCTGTTCTTGGGTCAGTCAGGTGCCAAGCAGCCAGCACTTCCCCACCACTTTTGGTCCACGGATGCCCTTGGCCATCTGGGAAGCCTGTGGACCCCATCTCAGGAGAATTTTTGCAAACGCATAAAATGAGACCCATAGGATTACAAAGGCAGCAAATTATACTGAAATACAGTTATCAAAGTATTAAACATTCATCAGTAACATAGTCTTTAGTTAAAAGCATTTACTGGCCAGGCTCATACCTGTAATCCCAGCACTTTGGGAGGCTGAGGTGGGAGGACTGCTTGCCTCCAAGAGTTTGAGACCAGCCTGGGCAACATAGTGAGACCTCTTCTCTACAACAAATAAAAACAGCTGGGCGTGGTGGCACACCAGTAGTCCCAGCTACTCAGGAGGCTCAGGCGGGAGGATCGCTTGAGCTCTGGAGGTCAAGGCTGCAGTGAGCTATGATGGCACCACTGCACTCAGCCTGGGCAACAGAGTGAGATTCTGTCTCAAAAAGTAAATAAAAATAAAAGCATGTGTTAAACGTATTAGTGACACCACTCAGTATTAAGGTATTAAATAACAGGATCCCGCCTGACAACCACTGTTATTTCAGAGTAGTGATGAACATAAGTGGTATTCGAACTCTCTGCCACCTCTATGAATTGACAGGAAAACATCTGTGACCTCTCTTGCTGACCGAGTCACGGGTACTGCTAATACTGCCACGTTCATAATGGAAGGAAATGCCCAGTGTCTGTTCGAGGTTGGTGGAAAGAAAGATGTCGTTTTTTCCACCTCAGTCCGTGGAGCCCTGAATTCTGTGTGCAGACGTTTGGGGTCTAAGCAGGACAGTGGGAAGCTTTGCTTCCCACCTTTGCTTTGGCTCAAAGCCCTCATCTGTCTGCTCTCCCCATAGGGATCACAGGTCTCTGGTCTCTGGCCATCATTTCCTGGGAGAGATGGATGGTGGTCTGCAAGCCCTTTGGCAATGTGAGATTTGATGCCAAGCTGGCCATCGTGGGCATTGCCTTCTCCTGGATCTGGGCTGCTGTGTGGACAGCCCCGCCCATCTTTGGTTGGAGCAGGTAAGGGTGCGAGGACGCAAGATGGAGTGGGCAGGGTCAGACTCTGTGACCTTAAGGCAAATCACTTCCTTTCTCTGGGCCCCTCTGAGCGTGCAATGTCTATCAATGTATGAATGTGGCTGCAACATAGGAAAGGCTCTGTGGTCCCCGAACCTCTGGAAACATATTTATCCCAAGCACGATCAGGTCACAGGCGCACACGGAGCTCAGGCCATCAGCACAGCTGTCAGTGAACGCATAGCGTGTTTGCATTCCAGGTCTCTTTCTTGCACACGCTGCCGCACCACGCCCCCCACCTTTCAGAGGCTGCTTGGGTCATAGATCCACCTGGGCCTACAGAGCACATGTCCTGGCCAGGCCAAGCAAGTGGCTCAAATGTTTGATTGGAGTGGACTGGGTGGGACAGCATTTCAC

Once the BLAST is completed, click “Back to Traditional Results Page” to see a page similar to the tutorial.

 ***Question 3: How long is the query sequence that was used to search the database?* (1 points)**

***Question 4: List the following information for the top two sequences that matches your query the best:* (3 points)**

1. ***Sequence Name***
2. ***Query Coverage***
3. ***E-value***
4. ***Percent Identity***
5. ***Accession Number***

 ***Question 5: What is the difference between “Query Cover” and “Percent Identity”?* (2 points)**

 ***Question 6: What is the significance of the “E-value”? Which is better: higher or lower e- values?* (2 points)**

Click on the Accession Number of the top sequence (this number should start “NG\_”).

 ***Question 7: What species is this sequence from?*  (1 point)**

 ***Question 8: How many base pairs are in this DNA sequence?*  (1 point)**

 ***Question 9: What protein does this gene encode?*  (2 points)**

 ***Question 10: What happens if there is a mutation in this gene?*  (2 points)**

Run a second BLAST-N (Nucleotide BLAST) using the same nucleotide sequence. Under ‘Choose Search Set’ > ‘Organism’ enter “Pipidae (taxid:8352)”. Set the Program Selection Optimization to “More dissimilar sequences (discontiguous megablast)”. Be sure that the Database is set to “Nucleotide collection (nr/nt)”.

 ***Question 11: How many sequences did your query match with?*  (1 point)**

***Question 12: List the following information for two of the sequences that matches your query:* (3 points)**

1. ***Sequence Name***
2. ***Query Coverage***
3. ***E-value***
4. ***Percent Identity***
5. ***Accession Number***

Click on the Accession Number of the bottom sequence (this number should start “NM\_”).

 ***Question 13: What species is this sequence from?*  (1 point)**

 ***Question 14: How many base pairs are in this sequence?*  (1 point)**

 ***Question 15: Is this sequence DNA, mRNA or protein?*  (1 point)**

 ***Question 16: Which of the two research articles (references) listed for this sequence is more specifically focused on understanding this gene?*  (2 points)**

 ***Question 17: Using the translated amino acid sequence given on this page, how long (number of amino acids) are present in the polypeptide this gene encodes?* (2 points)**

Run a BLAST-P (Protein BLAST) using the amino acid sequence below.

Keep all parameters set at default but double check that the Database is set to “Non-redundant protein sequences (nr)” and the Program Selection Algorithm for “blastp (protein-protein BLAST)”.

FDEWWCCWQEGFWHFKSCLWEWHFKIGWKKVMFFHGGMSFHGRFPKQRKWGCKQLFHRF

 ***Question 18: How many sequences did your query match with?*  (1 point)**

***Question 19: List the following information for the sequence that matches your query the best:* (1.5 points)**

1. ***Sequence Name***
2. ***Query Coverage***
3. ***E-value***
4. ***Percent Identity***
5. ***Accession Number***

 ***Question 20: What is a conserved domain?*  (3 points)**

 ***Question 21: What putative conserved domain(s) were found in this query?*  (2 points)**

 ***Question 22: What basic genetic process would a protein containing this domain be involved in?*  (2 points)**

Run a BLAST-P (Protein BLAST) using the amino acid sequence below.

Keep all parameters set at default but double check that the Database is set to “Non-redundant protein sequences (nr)” and the Program Selection Algorithm for “blastp (protein-protein BLAST)”.

MIENLRQRKAAYVRYFNSSAFFVVFLVLLLRKIFIKGIITTISFCIWAVQTWYDSLGAINKIQQKYKTLEYNLTTTEVVMSNGDDSLFFSNFSLLGTPVDINFKIERGLAVAGSTGAGMVIMGELSEGKIKHSGRISFCWIMPGNVSYDEYRYRSTIKEVIKACQLEEDISGITLSGGQRISLRAVYKDADLYLLDSPFGYLDVLTEKEIFESCVCKLMARILVTSKKLILHEGSYFYGT

***Question 23: List the following information for the top sequence that matches your query the best:* (1.5 points)**

1. ***Sequence Name***
2. ***Query Coverage***
3. ***E-value***
4. ***Percent Identity***
5. ***Accession Number***

***Question 24: List at least four different species (not including the best match) that have significant E-values. Then list the lowest E-value and the percent identity for each of your four organisms.*  (4 points)**

Click on the Accession Number of the best sequence (this number should start “XP\_”).

 ***Question 25: What is the common name of the species this sequence is from?*  (1 point)**

 ***Question 26: Is this sequence DNA, mRNA or protein and how long is it?*  (2 points)**

 ***Question 27: How was this sequence predicted?* (2 points)**

**PART II:**

For this section, you will be using the Online Mendelian Inheritance in Man (OMIM) database, which can be found at: <http://omim.org/>. OMIM is a searchable database of human genes, genetic traits, and disorders that is created and edited by researchers at Johns Hopkins University.

Search the term “Lewy Body” at the OMIM website.

 ***Question 28: How many search results did you find?*  (1 point)**

Click on the first search result (“# 127750. DEMENTIA, LEWY BODY; DLB”).

***Question 29: What is the name of the two genes that cause dementia with Lewy bodies when mutated?*  (2 points)**

***Question 30: What chromosomes are these genes located on?*  (1 point)**

***Question 31: What type of inheritance do mutant alleles of these genes have?*  (2 points)**

***Question 32: List at least three phenotypes found in individuals with dementia with Lewy bodies.*  (3 points)**

***Question 33: What mutation was the cause of dementia with Lewy bodies in one of the Japanese families studied by Ishikawa et al. (1997)?*  (2 points)**

Click on the link for the gene page of SNCA (“163890”).

 ***Question 34: What type of cells is this protein abundant in?*  (1 point)**

***Question 35: How many exons does this gene contain?*  (1 point)**

***Question 36: What is the approximate length of this gene?*  (1 point)**

***Question 37: What conclusion did Tabrizi et al. (2000) make about the link between mutations in SNCA and dopamine toxicity?*  (3 points)**

**PART III:**

For this section, you will be using the Genetics Home Reference (GHR) database, which can be found at: <https://ghr.nlm.nih.gov/>.

GHR is a searchable database of consumer information about genetic conditions and the genes or chromosomes associated with those conditions put out by the National Library of Medicine.

Search the term “Lewy Body” at the GHR website.

 ***Question 38: How many search results did you find?*  (1 point)**

Click on the first search result (“Dementia with *Lewy bodies*”).

 ***Question 39: What organ system is affected in by dementia with Lewy bodies?* (1 points)**

 ***Question 40: What may be the first symptom of patients with dementia with Lewy bodies?*  (2 points)**

 ***Question 41: What can you conclude about how often dementia with Lewy bodies occurs in humans?*  (3 points)**

 ***Question 42: List the four genes that are directly or indirectly involved in dementia with Lewy bodies. Then list how mutations in the genes can lead to the formation of Lewy bodies. (Note: the mechanism is unknown for one of these genes)*  (5 points)**

 ***Question 43: Based on what you wrote about mutations in the SNCA gene in the previous question: why is dementia with Lewy bodies a dominant (gain-of-function) disorder?* (3 points)**

 ***Question 44: Which website (OMIM or GHR) do you prefer to find information about dementia with Lewy bodies, and why?*  (3 points)**

**PART IV:**

For this section, you will be using the Protein Data Bank (PDB), which can be found at: <http://www.rcsb.org/pdb/home/home.do>.

PDB is an international archive of 3-D structural information for biological macromolecules. PDB’s structure records provide access to several interactive molecular graphics program. You will be using the Protein Workshop tool to view and generate images of protein structures.

For those of you that are unfamiliar with PDB, an introduction to PDB can be found at: <http://pdb101.rcsb.org/learn/guide-to-understanding-pdb-data/introduction>

Using the Search feature, enter “2QUG” into the search box. (There should only be one result from this query.)

 ***Question 45: What is the name of this protein, how is this protein classified, and what organism is it found in?*  (3 points)**

 ***Question 46: What method was use to map this 3D structure, and what resolution was used?* (3 points)**

On the left side of the “2QUG” page, click on “Protein Workshop” under “Standalone Viewers”. You will be prompted to download a JNLP file called “RCSB-ProteinWorkshop”. By opening this file, you will be looking at a protein workshop window containing the 3D structure for “2QUG”. JAVA needs to be installed on your computer to open this file.

If you have trouble with this application, go to the Protein Workshop Help file available from PDB: <https://www.rcsb.org/pdb/staticHelp.do?p=help/viewers/proteinWorkshop_viewer.html>

At the top of the control panel, you should see four tabs: Tools, Shortcuts, Options, and Help and Credits. If you need to reset the structure to its original configuration at any time during this activity, select the Options tab and click “Reset”.

 ***Question 47: For this question, paste a picture of the default protein structure of “2QUG”.*  (2 points)**

 ***Question 48: What two different secondary structures are shown in this image?* (2 points)**

 ***Question 49: Using the Shortcuts tab, “Recolor the Backbone” by selecting “By Compound” then clicking “Enact”. This will color each part of the protein based on what amino acids are present. Paste a picture of this change to the protein structure.*  (3 points)**

 ***Question 50: Using the Shortcuts tab, “Recolor the Backbone” by selecting “Hydrophobicity” then clicking “Enact”. This will color each part of the protein based on how hydrophobic or hydrophilic the amino acids are. Paste a picture of this change to the protein structure.*  (3 points)**

 ***Question 51: Using the Tools tab, select the “Colors” tool and select “Ribbons” as what you want to affect. Color the entire protein a single color (of your choice) by clicking on “2QUG” under “Choose items…”.***

 ***Then select a different color, and color the following amino acids:***

 ***29 ASN; 30 LEU; 31 ALA; 32 GLU***

 ***Then choose a third color, and color the following amino acids:***

 ***356 ILE; 357 PRO; 358 MET; 359 SER***

 ***Paste a picture of this change to the protein structure.*  (3 points)**