Biol3950 – Biotechnology Spring 2024 Dr Carr

Environmental DNA Bioinformatics Exercises

1. Identify these unknown sequences using **nucleotide BLAST** (default parameters)
    a. How are the results sorted? What do the different values mean?
    b. What does the alignment view tell you?
    c. What does the taxonomy view tell you?
    d. For each sequence, what is the **species identity**? What are some ways to deal with ambiguous matches (multiple equally likely matches)?
2. Search these sequences on **BOLD**. Do you get the same information? https://www.boldsystems.org/
    a. What is a **BIN**?
3. These species were found in Newfoundland. Are any these species of special concern? E.g., invasive, threatened or endangered, or of commercial interest?
4. For one of the ambiguous results, use **MEGA** to identify a potential locus for a new marker that would distinguish the species from each other (if there are a lot of species in the top hits, pick the top 4).
    a. Either within MEGA's browser or through your regular browser, search GenBank for reference sequences for your species. You will need to find a gene that has at least one (but preferably more) reference sequences for each species. HINT: most likely going to find this for mitochondrial genes such as **COI**, **Cytb**, **16S**, **12S**, **Dloop**
    b. Download these sequences and add them to the alignment viewer. Rename the sequences for clarity.
    c. Align the sequences. If they won't all align, you may need to reverse complement or trim some sequences. Try aligning just two sequences at a time if this happens to isolate the issue. It's ok to delete sequences from the alignment that won't align, but you need to ensure that at least one per species is present for the analysis.
    d. Once all aligned, trim the ends of your alignment so that there are no gaps and you are only looking at the section where you have sequence for all species.
    e. Use the tools in **MEGA** to look for a section of the gene with a unique sequence for each species. Is there a suitable primer binding site upstream and downstream of this location that is the same (conserved) for all species?

1	TTTATCAAGTATACAAGCTCACTCAGGGGGATCGGTGGATATGGCAATCTTTAGTCTTCATTTAGCTGGGATATCTTCAATATTGGGAGCTATGAATTTTATTACAACGATTATAAATATGAGAGCACCAGGAATCACAATGGACCGAATGCCTTTATTTGTGTGATCTGTTTTAGTAACTGCCGTTTTGTTATTGTTATCTTTACCGGTATTGGCAGGAGCTATAACAATGCTTTTGACAGATCGAAATTTTAATACCGCGTTCTTTGATCCAGCGGGGGGAGGAGATCCTATTTTATATCAACACCTTTTT
2	ACTATATTTTATCTTTGGGGCTTGAGCTGGTATAGTAGGGACTTCTTTGAGTCTTATTATTCGAGCTGAATTAGGGCAGCCAGGAACTTTAATCGGTAACGACCAAATTTATAACGTTGTTGTAACTGCTCATGCTTTTGTAATAATTTTTTTCATAGTAATACCAATTATAATTGGAGGATTTGGTAATTGACTTGTACCTCTAATATTAGGAGGGCCAGATATA
3	TTTATCTGGACCACAAACACATTCTGGTGGTTCTGTAGATATGGCTATATTTAGTTTACATTGTGCTGGTGCCTCTTCAATTATGGGGGCTATAAATTTTATTACAACAATAATTAATATGAGAGCACCTGGATTAACAATGGATAAATTACCATTGTTCGTGTGATCTGTGTTAATAACAGCTGTATTACTACTACTATCTTTACCTGTTTTAGCAGGCGCAATAACAATGTTATTAACAGATCGTAATTTCAACACAACATTTTTTGACCCGGCCGGAGGTGGAGATCCAGTTCTTTATCAACATTTATTT
4	CACCGCGGTTATACGAGAGGCCCTAGTTGATAAATACCGGCGTAAAGAGTGGTTACGAAAAAATGTTTAATAAAGCCGAACACCCCCTCAGCCGTCATACGCACCTGGAGGCACGAAGACCTACTGCGAAAGCAGCTTTAATTGTACCTGAACCCACGACAGCTACGACA
5	TTTATATTTTATCTTTGGAGCATGAGCTGGAATAGTAGGAACAGCATTAAGTTTATTAATCCGAGCAGAACTGGGGTCTCCTGGTAGATTGATTGGAAATGATCAAATTTATAATGTAGTTGTCACAGCACATGCTTTCGTAATAATTTTCTTTATAGTAATACCTATTCTTATTGGAGGATTTGGAAATTGATTAGTACCTTTAATACTAGGGGCACCTGATATG
6	CCTATATCTCGTATTTGGTGCCTGAGCCGGAATAGTCGGTACTGCACTGAGCCTTCTAATCCGTGCCGAATTAAGTCAACCAGGCGCCCTTCTTGGAGATGACCAAATTTACAATGTCATCGTCACAGCGCATGCCTTTGTAATGATTTTCTTTATAGTAATGCCAGTAATAATCGGAGGATTTGGCAACTGACTTGTGCCATTAATAATCGGCGCTCCAGACATA
7	CACCGCGGTTATACGAGAGGCCCAAGTTGAAAGACCCCGGCGTAAGGCGTGGTTAAGTTAAAATTTGTACTAAAGCCGAACATCTTCACGGCTGTTATACGCACCCGAAGATAAGAAGTTCAACCACGAAGGTAGCTTTATTTAATCTGAACCCACGAAAGCTACGGCA
8	ATAACAGATATAAAGGTTGTTTTGATTAGCGGTCTCAACTCATCAAATAGTGTGTATACAAATTATAGAACTTTCTTTTGAACTTTTATGTCATCAGATAGACTTGTGAGTGACTTTGTCATTTATTTGCAAAAGTTGTTATCTAACCACAATAGTCACTAATATTACA
9	CCTCTACCTAATCTTTGGTGCCTGAGCAGGTATAGTCGGAACTGGCCTAAGTCTTTTAATTCGAGCAGAGTTGAGCCAGCCCGGATCACTTCTAGGTGATGATCAGATTTATAATGTCCTTGTTACAGCCCATGCCTTAGTAATAATCTTTTTTATGGTTATACCAATTATAATTGGAGGGTTTGGCAATTGACTCGTCCCTTTAATGATTGGCTCTCCAGACATA
10	CCTTTACTTCATTTTCGGAGCATGAGGAGGCCTTCTTGGCACCTCCATAAGTCTCCTTATTCGAGCTGAGCTTGGACAACCTGGATCCCTTCTAGGAAGGGACCAGCTCTATAACACTATTGTTACCGCTCACGCCTTT

	CTAATAATTTTCTTTCTTGTTATACCAGTATTTATTGGAGGCTTTGGAAACTGACTCATCCCCCTAATACTAGGGGCCCCAGATATG
11	ACTAGCTTCTGGAATTTCGCACAGAGGACCCTCTGTTGATATAGCAATTTTTTCTCTTCATCTTGCCGGTGCTTCCTCTATCTTAGGAGCAGTAAACTTTATTTCTACAATTATTAATATACGTCCCCCAGGAATAACCTTCGATAAAATACCTTTGTTTGTTTGATCAATTTTTATTACAGCCATTTTATTACTCCTATCTTTACCAGTTTTAGCAGGAGCTATTACTATACTTCTCACTGATCGTAACTTAAATACAACTTTTTTTGACCCTGCAGGAGGAGGAGATCCTATTTTATACCAACATCTTTTT
12	TTAAGGAGTAACATTTCTCATGCAGGAGCGTCTGTAGACTACGCTATCTTTTCTTTACATTTAGCCGGAGTTTCTTCTATCTTAGGTGCAGTAAACTTTATTAGAACTCTCGGAAATATACGAACGTTCGGTATATTTCTTGACCGCATACCCTTATTCGCTTGAGCAGTTTTAATTACGGCTATCTTACTTTTATTATCTCTGCCTGTCTTAGCAGGTGCTATTACAATACTATTGACTGACCGCAACCTGAATACTTCTTTCTATGACCCCAGGGGCGGGGGAGACCCAATCTTATACCAGCACCTATTT