Homework 2: Sequence Alignment and Homology Search *(reformatted using chatGPT and updated from the 2023 instance)*

Due Date: September 5, 2025
Reading:

- Section 3.8 (Pairwise sequence alignment) – week of 8/25
- Chapter 4 and supplemental material on the Viterbi algorithm – week of 9/10
- Chapter 2 and the Glimmer paper – for 9/15

You are encouraged to **reuse** code across Tasks 2–4. Modularizing this code is optional.

💼 Assignment Tasks

1. Sequence Retrieval *(1 point)*

Download the following sequences from GenBank:

- `X79493` (Drosophila gene)
- `AY707088` (Human gene)
  Include both in your submission as FASTA files.

2. Global Alignment Implementation *(10 points)*

Write a program called `globalign` to compute the global alignment **score and alignment** of the two sequences from Task 1 using:

- +2 for a match
- -1 for a mismatch
- -2 for a gap

Perform full traceback and format the alignment clearly. Include:

- Source code and compilation instructions
- Alignment output in `report.txt`

💡 *Optional:* Highlight mismatches and gaps in the formatted alignment for better clarity.

3. Local Alignment Implementation *(10 points)*

Write a program called `localign` to compute the **local alignment** (Smith-Waterman) between the same sequences.

- Use the same scoring scheme as above
- Include traceback and user-friendly alignment output

In `report.txt`, compare this alignment to your global alignment from Task 2. Consider:

- Biological meaning
- Score difference
- Overlap region length

---

4. End-Gap Free Alignment (Score Only) *(5 points)*

Write a third program `egfalign` that computes the score (no traceback required) for an **end-gap free alignment** between:

- `NC_012920` and `AF254446` (from HW #1)
  Use the same scoring scheme (+2 match, -1 mismatch, -2 gap). Report score in `report.txt`.

---

5. BLAST Comparison *(3 points)*

Use NCBI's **bl2seq** utility to BLAST the sequences from Task 1. In `report.txt`, include:

- Length of alignment
- Identity percentage
- E-value
  Discuss how this result compares to your local alignment (Task 3).

---

6. BLAST of Mystery Sequence *(2 points)*

Download the mystery DNA sequence from the course website.

- Use `blastn` and `blastx` (via NCBI web interface)
- Report the top hits from each search (accession, species, function if available)

---

7. (COSC 594 only) Linear Space End-Gap Free Alignment *(6 points)*

Implement a linear-space version of the **end-gap free alignment** algorithm to align `NC_012920` and `AF254446`.

- Report score only (same scheme: +2, -1, -2)
- Discuss space complexity trade-offs vs. previous methods

Grading Rubric Summary:

Task	Description	Points
1	Sequence retrieval	1
2	Global alignment code + result	10
3	Local alignment code + comparison	10
4	End-gap free score	5
5	BLAST summary	3
6	Mystery sequence BLAST	2
7 (594 only)	Linear-space alignment	6
Total (UG)		**31**
Total (Grad)		**37**