Progress Log – Genome Analysis Tool

Week 9 – Project Planning

Tasks Completed

- Identified core modules:
  - SequenceStore
  - RollingHash
  - SuffixTrie
  - MotifFinder
  - QueryEngine
- Created initial GitHub repository

Challenges:

- Understanding how modules would interact

Resolution:

- Designed clear module responsibilities

Week 10 – Core Data Structures

Tasks Completed:

- Implemented SequenceStore class
- Added DNA cleaning and validation
- Implemented RollingHash (computeHash, rollHash)
- Implemented Suffix Trie for exact search
- Inserted all suffixes of a sequence
- Verified pattern search correctness

Challenges:

- Handling invalid DNA characters.
- Managing large number of suffixes

Resolution:

- Added cleaning logic before storage.
- Used efficient Trie node structure

Week 11 – Motif Finding & Query Engine

Tasks Completed:

- Implemented MotifFinder using Rolling Hash
- Counted k-mers using hash maps
- Calculated motif frequency

- Sorted results by frequency
- Implemented Exact Search using Suffix Trie
- Implemented Approximate Search (edit distance $\leq$ k)
- Integrated QueryEngine with SequenceStore

Challenges:
- Debugging hash collisions and counts.
- Matching approximate patterns efficiently
Resolution:
- Verified rolling hash logic and window handling.
- Used sliding window with edit-distance checking

Week 12 – Integration & Testing
Tasks Completed:
- Integrated all modules in main()
- Performed integration testing
- Fixed crashes and logic errors
- Added debug output for verification

Challenges:
- Runtime crashes due to pointer misuse
Resolution:
- Proper initialization and validation

Week 13 – Finalization
Tasks Completed:
- Cleaned code and removed executables
- Final testing with realistic DNA sequences
-handled edge cases and fixed the code.
- Tagged final release on GitHub

Challenges:
- Ensuring output consistency
Resolution:
- Verified outputs against expected results.

Fatima: Implemented Suffix Tries, Tries and Query Engine modules.
Yusra: Implemented Motif Finder, Rolling Hash and Sequence Store modules.
Testing in main and integration done by both members.