Christopher Elliott

Comp 363

Chandra Sekharan

Algorithmic Description

1. First, the code asks user for file and sends parameter to the reader function.
2. The reader function uses a stringbuilder to append the readlines from the text file
3. The stringbuilder that has the entire genome sequence in it is sent as a parameter to both encoding functions

Encoding Method 1

1. Reads in the stringbuilder as the variable genome.
2. Runs a for loop for the length of the genome sequence.
3. Uses if statements to check if the char at “i” is a A, C, G, or T.
4. Uses a counter variable to increment bitset index
5. Uses ones variable to increment every time the value “true” is set
6. If the if statement is executed, sets value for bitset for two bits to true or false accordingly.

The amount of the number '1' in the sequence is: 70002635

The amount of bits in use for the sequence is: 140000000

Encoding Method 2

1. Reads in the stringbuilder as the variable genome
2. Runs a for loop for the length of the genome sequence
3. Uses if statements to check if the char at “I” is a A, C, G, or T
4. Uses a counter variable to increment bitset index
5. Uses a ones variable to increment every time the value “true is set
6. If the if statement is executed, sets value for bitset for two bits to true or false accordingly.
7. Uses while statement that increments a loop variable if the char at the next index is a duplicate.
8. If it is a duplicate letter, while loop will keep incrementing loop variable up to 3 times because the max of mirror occurences is 3.
9. Depending on loop variable, executes an if statement that sets two bits according to how many duplicate occurences there were.

The amount of the number '1' in the sequence is: 118100072

The amount of bits in use for the sequence is: 239302584