

INF-503 – Homework 1

I could not figure out how to run the program on monsoon, so I've only run the program on my laptop using the sample dataset. Results for problems C and E are as follows:

C)

```
number of unique sequences: 1000
number of reads in dataset 1: 126
number of reads in dataset 2: 106
number of reads in dataset 3: 124
number of reads in dataset 4: 109
number of reads in dataset 5: 116
number of reads in dataset 6: 107
number of reads in dataset 7: 95
number of reads in dataset 8: 110
number of reads in dataset 9: 122
number of reads in dataset 10: 84
number of reads in dataset 11: 110
number of reads in dataset 12: 100
number of reads in dataset 13: 102
number of reads in dataset 14: 119
number of A's: 14072
number of C's: 10988
number of G's: 11316
number of T's: 13544
number of other charaters (like N's): 80
```

E)

```

sequence nr: 0
sequence is AAAAAACAACGGCTCAATAAACGACATACATTTCCCGTATGATGACGAGC
sequence nr: 1
sequence is AAAAAGCAGCAACACAGGCTATTTCAGGATTTATCAGTTAAANAGGGTGAA
sequence nr: 2
sequence is AAAACATCTTCTTCATGAACGAAATGTTTTGAGCAACGTTTAAAGCTTG
sequence nr: 3
sequence is AAAACGCATTATTTCAAGCTACATATACATTTCAATATAACGGAGTATCA
sequence nr: 4
sequence is AAAACTTTAATAACGGAATGTAAGAGATATAAGGGGATGGGTTAGTTTGG
sequence nr: 5
sequence is AAAAGCGAAATCCATTCATATCGAAACAGAGGAAGAAAAAGAAGTAGATA
sequence nr: 6
sequence is AAAAGGATATTTGTCACCATCTACAGAATACATGTTACTATCAAGATAAC
sequence nr: 7
sequence is AAAAGTGAAGAACTAATTGCTATTATACTGGCCGGAGCATTAATTTGGG
sequence nr: 8
sequence is AAAATATCAAAATATCCAGACAGTAATAACAAAAATATACATAAACTAG
sequence nr: 9
sequence is AAACACATGGACTATATACAGTATCATCAAGTCAATCGGCANTTTTGACT
sequence nr: 10
sequence is AAACATGTTGATACAACCTGAGTACATTAACTAGACTATAGCTATTCATA
sequence nr: 11
sequence is AAACCAGACACAGAAGCTCCAAGCATTCCAAATAGAATTCAGGCACAAAA
sequence nr: 12
sequence is AAACCCTTAGCTCAACAAAGTGGTTTGCTACTGATGAACCCGCCATATCC
sequence nr: 13
sequence is AAACGAACATTTTTTGGAGTGGTCGTATCAGAGTGTAAACGATTGAGAGGAG
sequence nr: 14
sequence is AAACGGAATCAAATGTCCGATATGCATACTATCACCAGTCGGATCAACTC
sequence nr: 15
sequence is AAATATGCGGTCCTTTTAGCTGTTGGAGTTGGCTTATCGGTAATAATTC
sequence nr: 16
sequence is AAAGAGCCTTCATCGCAAGAAGAGATGTTAATCAAACAGTTAGAGGAAAT
sequence nr: 17
sequence is AAAGATACGAACGGAGATGGATTCCGTGAAGATCCGAATGGTAAAGAATT
sequence nr: 18
sequence is AAAGATGTAACAACAGCGATATTTCAATTTATATCCCGTTTTTCATTAA
sequence nr: 19
sequence is AAAGCAAAAATGTTACAAAATTTATCCAAATGAATAACTTCTAACCTGAA

```

I've implemented a bubble sort algorithm to sort the sequences in alphabetic order, and the big O notation of the bubble sort algorithm is $O(n^2)$.