LSDS Assignment 2

Command for running Problem A : srun homework_2 1 human.txt human_reads.fa

Command for running Problem B : srun homework_2 2 human.txt human_reads.fa

A) Unsorted Linear search :

a. How long did it take you to search for the first 10K, 100K, and 1M 32-character long fragments of the subject dataset within the query dataset?

Time taken for searching 10k fragments of the subject dataset is : 12190.066 seconds
Time taken for searching 100k fragments of the subject dataset is : 117431.09 seconds
Time taken for searching 1M fragments of the subject dataset is : 1210354.93 seconds

b. How long would it take to search for every possible 32-character long fragment of the subject dataset within the query dataset?

Time taken for searching 3B fragments of the subject dataset is : 3614457925.939 seconds

c. Print the first 10 fragments of the subject dataset that you found within the Query AR object (if any)?

```
File size of genome data is 3057186663
File size of query dataset is 27772294
=====================================================
First 10 matching fragments in 10000 fragments are

Matching index in genome data : 0
Matching index in genome data : 1
Matching index in genome data : 2
Matching index in genome data : 3
Matching index in genome data : 4
Matching index in genome data : 5
Matching index in genome data : 6
Matching index in genome data : 7
Matching index in genome data : 8
Matching index in genome data : 9
=====================================================
=====================================================
First 10 matching fragments in 100000 fragments are

Matching index in genome data : 0
Matching index in genome data : 1
Matching index in genome data : 2
Matching index in genome data : 3
Matching index in genome data : 4
Matching index in genome data : 5
Matching index in genome data : 6
Matching index in genome data : 7
Matching index in genome data : 8
Matching index in genome data : 9
=====================================================
=====================================================
```

```
==========================================================
First 10 matching fragments in 1000000 fragments are

Matching index in genome data : 0
Matching index in genome data : 1
Matching index in genome data : 2
Matching index in genome data : 3
Matching index in genome data : 4
Matching index in genome data : 5
Matching index in genome data : 6
Matching index in genome data : 7
Matching index in genome data : 8
Matching index in genome data : 9
==========================================================
```

B) Sorted Binary search:

a. How long did it take you to search for the first 10K, 100K, and 1M 32-character long fragments of the subject dataset within the query dataset?

Time taken for searching 10k fragments of the subject dataset is : 0.04 seconds
Time taken for searching 100k fragments of the subject dataset is : 0.51 seconds
Time taken for searching 1M fragments of the subject dataset is : 4.87 seconds

b. How long would it take to search for every possible 32-character long fragment of the subject dataset within the query dataset?

Time taken for searching 3B fragments of the subject dataset is : 14533.358 seconds

c. Print the first 10 fragments of the subject dataset that you found within the Query AR object (if any)?

```
File size of genome data is 3057186663
File size of query dataset is 27772294
Sorting and applying binary search

=====================================================
First 10 matching fragments in 10000 fragments are

Matching index in genome data : 0
Matching index in genome data : 1
Matching index in genome data : 2
Matching index in genome data : 3
Matching index in genome data : 4
Matching index in genome data : 5
Matching index in genome data : 6
Matching index in genome data : 7
Matching index in genome data : 8
Matching index in genome data : 9
Total time taken for program execution after sorting for 10000 fragments are 0.048922 seconds
=====================================================
Sorting and applying binary search

=====================================================
First 10 matching fragments in 100000 fragments are

Matching index in genome data : 0
Matching index in genome data : 1
Matching index in genome data : 2
Matching index in genome data : 3
Matching index in genome data : 4
Matching index in genome data : 5
Matching index in genome data : 6
Matching index in genome data : 7
Matching index in genome data : 8
Matching index in genome data : 9
Total time taken for program execution after sorting for 100000 fragments are 0.513192 seconds
```

```
Sorting and applying binary search

=====================================================
First 10 matching fragments in 1000000 fragments are

Matching index in genome data : 0
Matching index in genome data : 1
Matching index in genome data : 2
Matching index in genome data : 3
Matching index in genome data : 4
Matching index in genome data : 5
Matching index in genome data : 6
Matching index in genome data : 7
Matching index in genome data : 8
Matching index in genome data : 9
Total time taken for program execution after sorting for 1000000 fragments are 4.87 seconds
=====================================================
```