

## problem 1A

```
yg336@wind:~/inf503/homework3
The number of arguments passed: 2
The first argument is: main
The second argument is: problem1A
The third argument is: /common/contrib/classroom/inf503/hw3_dataset.fa
The fourth argument is: /common/contrib/classroom/inf503/test_genome.fasta
successfully read data from file: /common/contrib/classroom/inf503/hw3_dataset.fa
init success!
record success!
The size of the direct access table is: 4294967296
There are 1584688 collisions in the direct access table
There are 4414557 unique sequences in the direct access table.
The alpha_T of the direct access table is: 0.00102784
number of sequences is: 5999245
```

## problem 1B

```
yg336@wind:~/inf503/homework3
The number of arguments passed: 4
The first argument is: main
The second argument is: problem1B
The third argument is: /common/contrib/classroom/inf503/hw3_dataset.fa
The fourth argument is: /common/contrib/classroom/inf503/test_genome.fasta
successfully read data from file: /common/contrib/classroom/inf503/hw3_dataset.fa
init success!
record success!
Total used to complete the entire search process is: 3 s
2046995 genome 16-mer fragments were found in the read set.
```

## problem 2A

```
yg336@wind:~/inf503/homework3
The number of arguments passed: 4
The first argument is: main
The second argument is: problem2A
The third argument is: /common/contrib/classroom/inf503/hw3_dataset.fa
The fourth argument is: /common/contrib/classroom/inf503/test_genome.fasta
It takes 8 s to read the sequences when the table size is 10,000
5989245 collisions found when the table size is 10,000
It takes 8 s to read the sequences when the table size is 100,000
5899245 collisions found when the table size is 100,000
It takes 8 s to read the sequences when the table size is 1,000,000
5051407 collisions found when the table size is 1,000,000
It takes 9 s to read the sequences when the table size is 10,000,000
2615972 collisions found when the table size is 10,000,000
```

The results make sense. It is more likely to have more collisions when the table size is small, and vice versa. The total search time costs are similar because of the structure of the hash table. Finding out the location of an element in a hash table is  $O(C)$ , but because of the collisions and the chain structure, the total search time depends on the length of the chain.

## problem 2B

```
yg336@wind:~/inf503/homework3
The number of arguments passed: 4
The first argument is: main
The second argument is: problem2B
The third argument is: /common/contrib/classroom/inf503/hw3_dataset.fa
The fourth argument is: /common/contrib/classroom/inf503/test_genome.fasta
Total used to complete the entire search process is: 6 s
2046995 genome 16-mer fragments were found in the read set.
```

The hash table is slower than the direct access table because of the collisions and chains.