

Search strategy:

Search motif exhaustively, according to the order of A, C, G, T.

Compare the hamming distance and finally calculate the minimum hamming distance all sequences have.

For example, the length of the motif is 5, then we assume the motif is AAAAA, AAAAC, AAAAG,..., TTTTT. Every motif calculates the hamming distance and outputs the minimal.

Object function:

find_motif_width():

Processing the first line of the input file, return the length of the motif.

convert():

To control the loop iteration, we need to run the motif from A to T, so we make use of 4 digits number. The for loop runs 4^n times, converting the times of iteration to ACGT.

hamdist():

The function calculates the hamming distance between the motif and the sequences. Compare them, if there are different bases, then the distance will be incremented.

same_mutate_number():

Compare the hamming distance of all sequences. From 0 to n(the length of the motif), if all sequences have the same hamming distance, then we get the mutate number. We guarantee that is the minimum number since we compare the hamming distance from 0.

main function:

Recording the minimum number and the sequence. Finally, we can output the key, the motif of every sequence, and the location.

Snap shot of input and output:

```
(base) jason@jason-B660M-GAMING-X-AX-DDR4:~/BDM$ time ./motif.out Q1.txt
key = CGCGCGCGCG(10,1)
Sequence1: location: 279~288 CGCGCGCGAG
Sequence2: location: 201~210 CACGCGCGCG
Sequence3: location: 435~444 CGAGCGCGCG
Sequence4: location: 212~221 CGCGAGCGCG
Sequence5: location: 229~238 AGCGCGCGCG
Sequence6: location: 206~215 CGCGCGCGCA

real    0m15.445s
user    3m5.001s
sys     0m0.164s
```

```
(base) jason@jason-B660M-GAMING-X-AX-DDR4:~/BDM$ time ./motif.out Q2.txt
key = AATCCCCGGCCTTT(14,1)
Sequence1: location: 217~230 ATTCCCCGGCCTTT
Sequence2: location: 605~618 AATCCCCGGCCTTT
Sequence3: location: 139~152 AATCCCCGGCCTTT
Sequence4: location: 502~515 AACCCCCGGCCTTT
Sequence5: location: 490~503 AATCCCCGGCCTTT
Sequence6: location: 476~489 ACTCCCCGGCCTTT

real    106m50.595s
user    1253m10.156s
sys     1m3.217s
```

```
(base) jason@jason-B660M-GAMING-X-AX-DDR4:~/BDM$ time ./motif.out Q3.txt
key = GGGCCGGGATTAGGG(15,2)
Sequence1: location: 401~415 GGGCCGGGGTTGGGG
Sequence2: location: 340~354 GAGCCGGGATTAAGG
Sequence3: location: 594~608 GGCCCGGGATTAGCG
Sequence4: location: 372~386 GAGCCGGGATTTGGG
Sequence5: location: 407~421 GGGCAAGGATTAGGG
Sequence6: location: 299~313 GGGCCGGGCATAGGG
Sequence7: location: 11~25 GGGCCGGGACCAGGG
Sequence8: location: 201~215 GATCCGGGATTAGGG

real    436m57.315s
user    5219m4.306s
sys     2m55.187s
```

Input the command "make" and then get the executable file called "motif.out". Input "./motif.out {file}" and then we can get the results.