CS170 Fall 2014 HW9

Name: Judy Wang
SID: 24208051
Login: cs170-bz
Collaborators: None

1. Writeup

(a) <u>Main Idea:</u> In my algorithm, we take the text file and save it as a list of strings. Then, we find the pair of fragments with the maximum overlap and merge them together. This will create potential increases in pair-wise overlap between other strings. We will loop through this until there is only one string left – therefore we will loop n times. This will return the optimal DNA sequence

(b) <u>Pseudocode:</u>

```
Algorithm ShotgunSequencing(List S):
1. while S has more than one element:
2.    int max = 0;
3.    string s1; string s2;
4.    for strings i in S:
5.       for strings j in S:
6.          if overlap(i,j) > max:
7.             max = overlap(i,j)
8.             s1 = i; s2 = j;
9.    merge(s1,s2)
10.return the string in S
```

(c) <u>Running Time:</u> $O(k^2n^3)$

The time it takes to find the overlap of two strings is worst case $O(k^2)$, since it must iterate through both strings of length k to find the overlap. To find the maximum overlap, the worst case running time is $O(k^2n^2)$, since it needs to find the overlap of all n^2 combinations of two strings in order to choose the pair with the most overlap. Since there are n strings, we will have to do this n times, giving us a final running time of $O(k^2n^3)$.