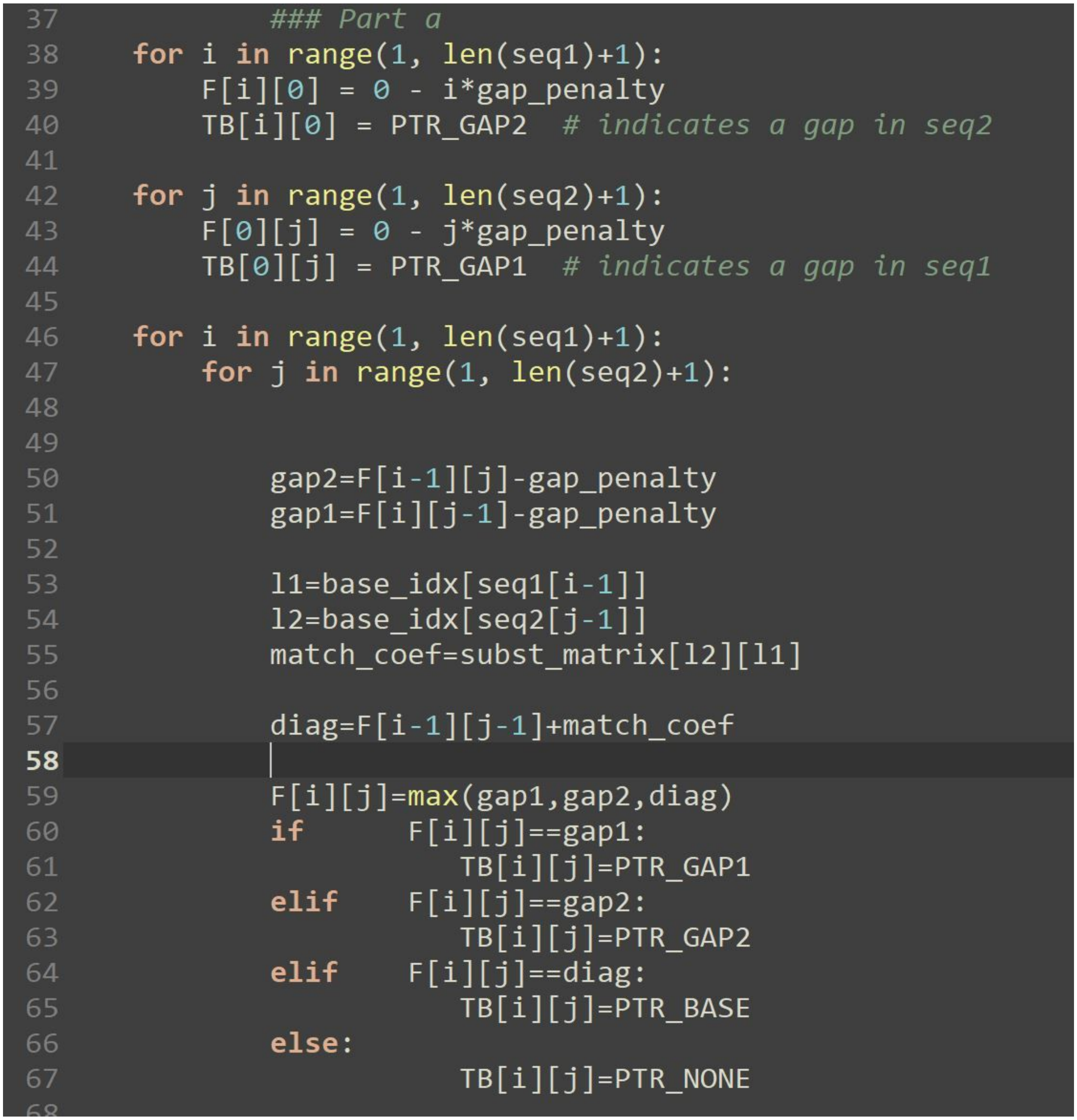
Kevork Hamparian
BE562
HW1

8
A.
1)



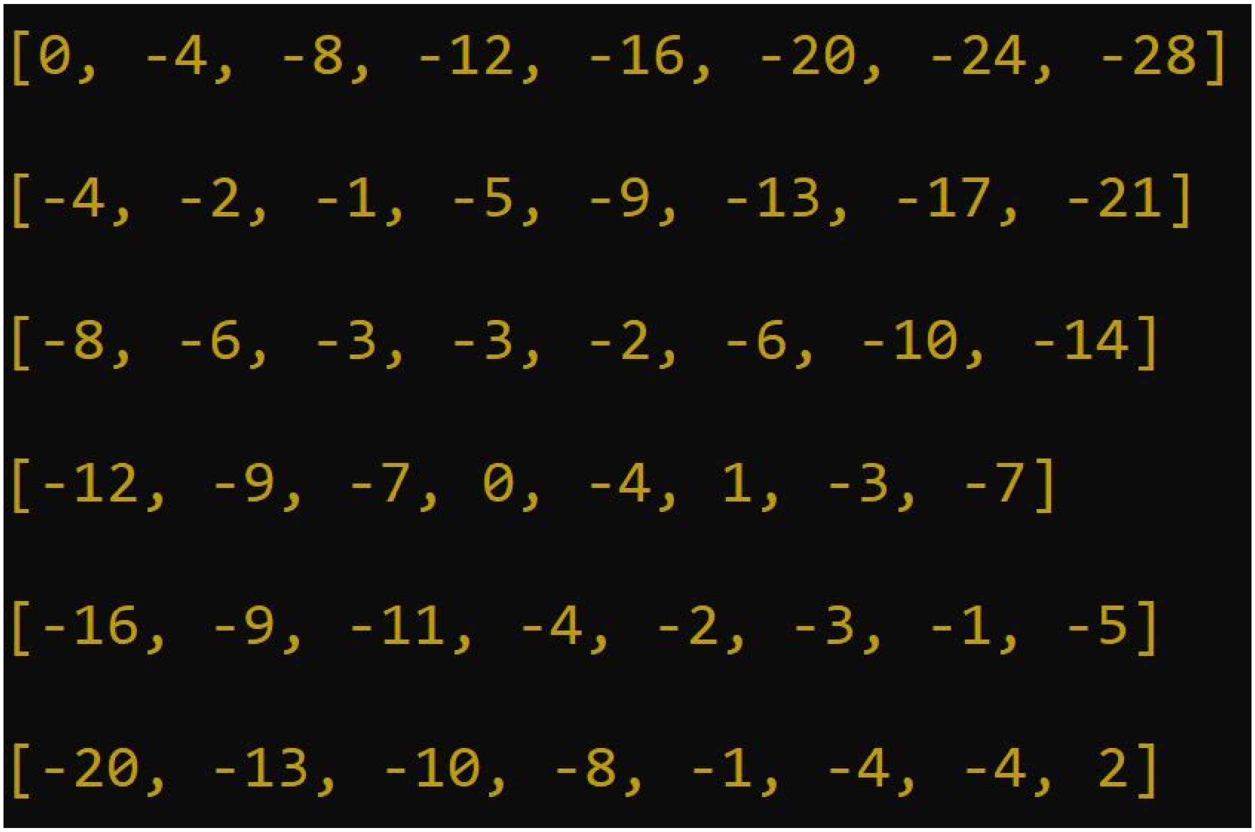
```
            ### Part a
    for i in range(1, len(seq1)+1):
        F[i][0] = 0 - i*gap_penalty
        TB[i][0] = PTR_GAP2  # indicates a gap in seq2

    for j in range(1, len(seq2)+1):
        F[0][j] = 0 - j*gap_penalty
        TB[0][j] = PTR_GAP1  # indicates a gap in seq1

    for i in range(1, len(seq1)+1):
        for j in range(1, len(seq2)+1):


            gap2=F[i-1][j]-gap_penalty
            gap1=F[i][j-1]-gap_penalty

            l1=base_idx[seq1[i-1]]
            l2=base_idx[seq2[j-1]]
            match_coef=subst_matrix[l2][l1]

            diag=F[i-1][j-1]+match_coef

            F[i][j]=max(gap1,gap2,diag)
            if      F[i][j]==gap1:
                       TB[i][j]=PTR_GAP1
            elif    F[i][j]==gap2:
                       TB[i][j]=PTR_GAP2
            elif    F[i][j]==diag:
                       TB[i][j]=PTR_BASE
            else:
                       TB[i][j]=PTR_NONE
```

2)

Optimal Alignment: - A - G C T G

T A C G C A G

F matrix:



```
[0, -4, -8, -12, -16, -20, -24, -28]

[-4, -2, -1, -5, -9, -13, -17, -21]

[-8, -6, -3, -3, -2, -6, -10, -14]

[-12, -9, -7, 0, -4, 1, -3, -7]

[-16, -9, -11, -4, -2, -3, -1, -5]

[-20, -13, -10, -8, -1, -4, -4, 2]
```

3)

Score: 2971

B.

In this part I changed two things. I changed the scoring matrix so that a match adds 0. Another thing I did was change the scoring matrix so that a mismatch that changes the score subtractively now changes it additively. For example, a mismatch that subtracted 2, now adds 2. The other change was that the ideal alignment now looks for the minimum of the same calculations. So a match is 0, a mismatch is either +2 or +1, and a gap is +4.

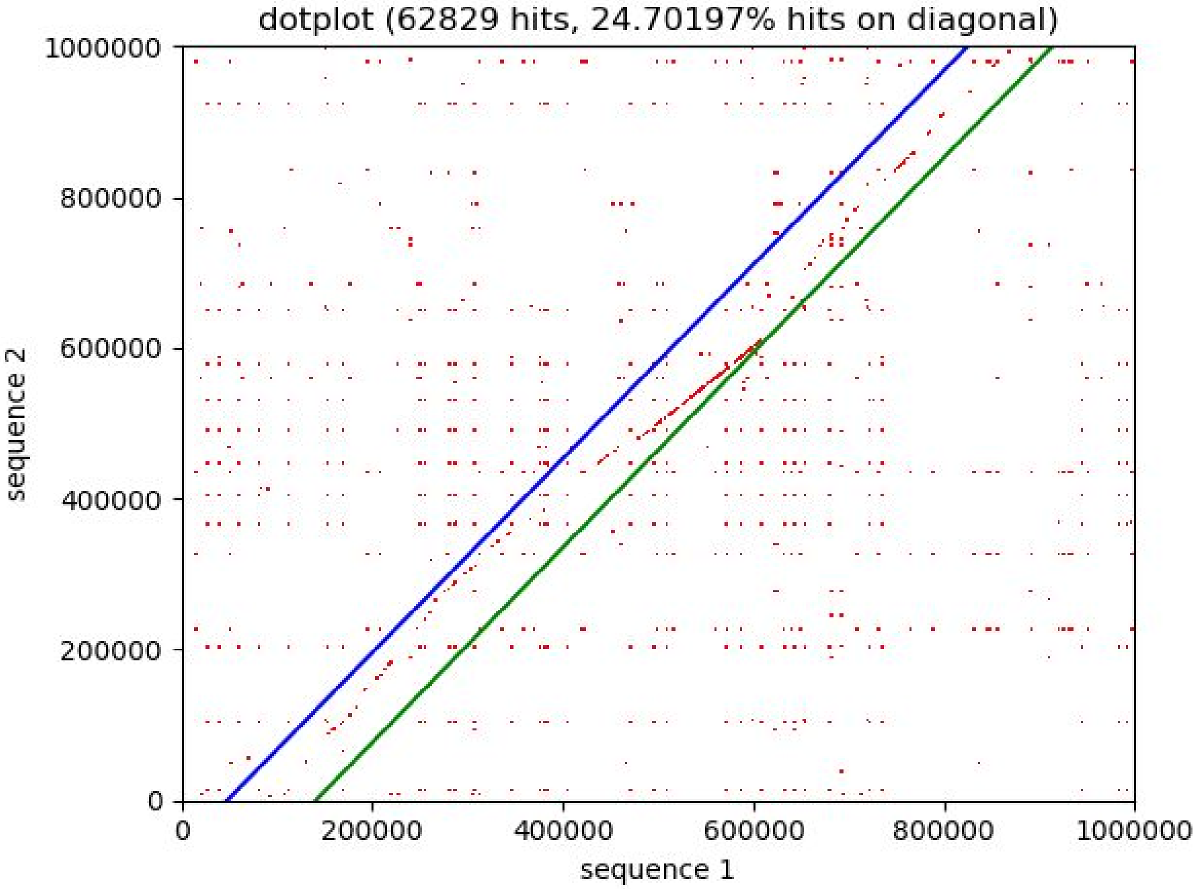
C.

The calculated distance between human HoxA13 and mouse HoxA13 genes are 197.

D.

Score of human HoxA13 and human HoxD13 is 1145.
Score of mouse HoxA13 and mouse HoxD13 is 1095.
1100/197 = ~5.5
5.5*70million=385 million years ago.
If a score of 197 is 70 million years of divergence, and we assume a <u>linear relationship</u> they diverged around 385 million years ago.

10
A.

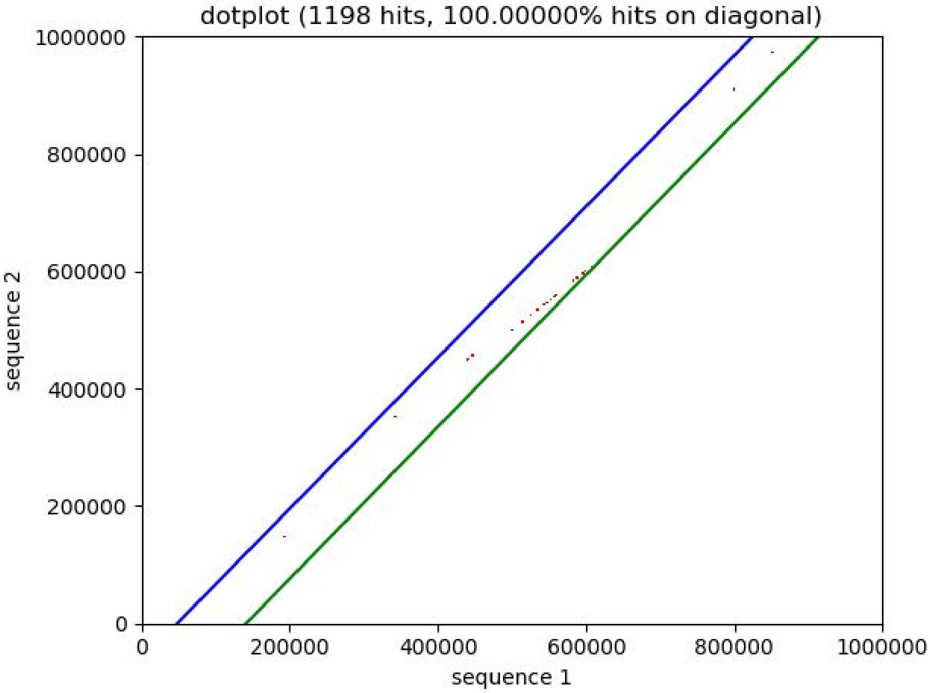


There seems to be a clear line of matches along the diagonal, but there are also quite a few random clusters throughout the matrix. There are 62,829 hits but only a quarter are along the diagonal. The clusters seem to show a checkered pattern. These are random sets of 30mers that match but not sequential matches. The matches along the diagonal show sequential matching 30mers that are in the same region of the genome.

B.

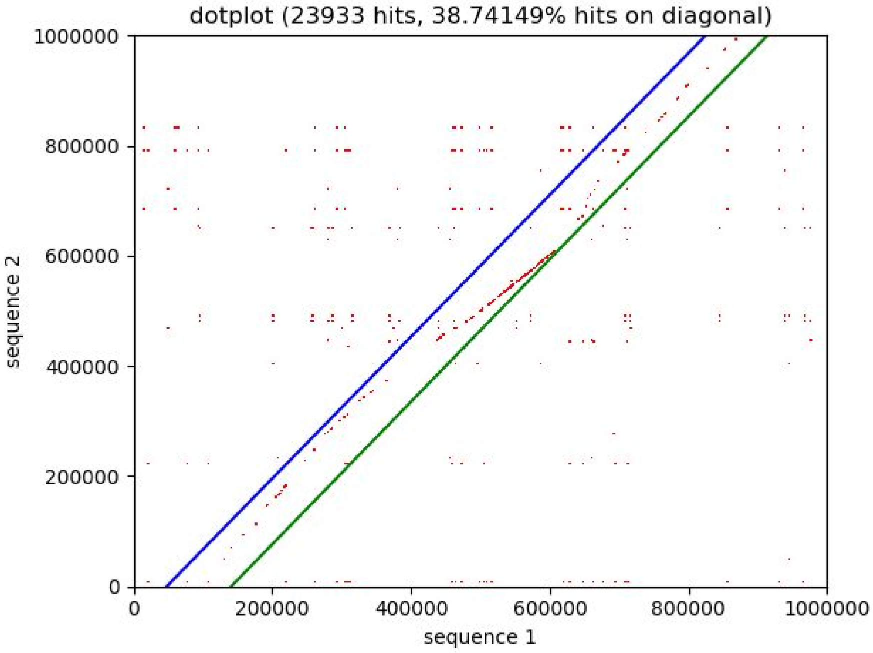
I.

Changed the length of the hash key to 100

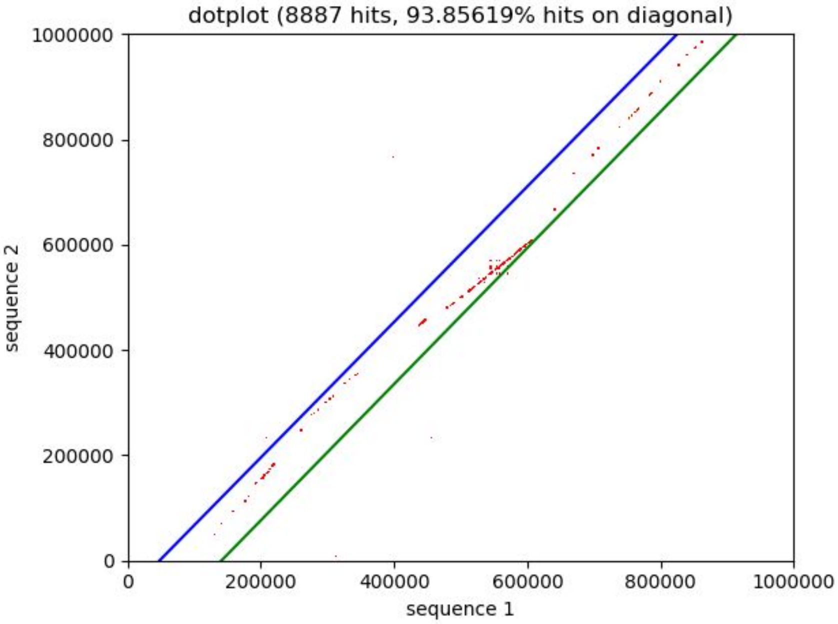


II.

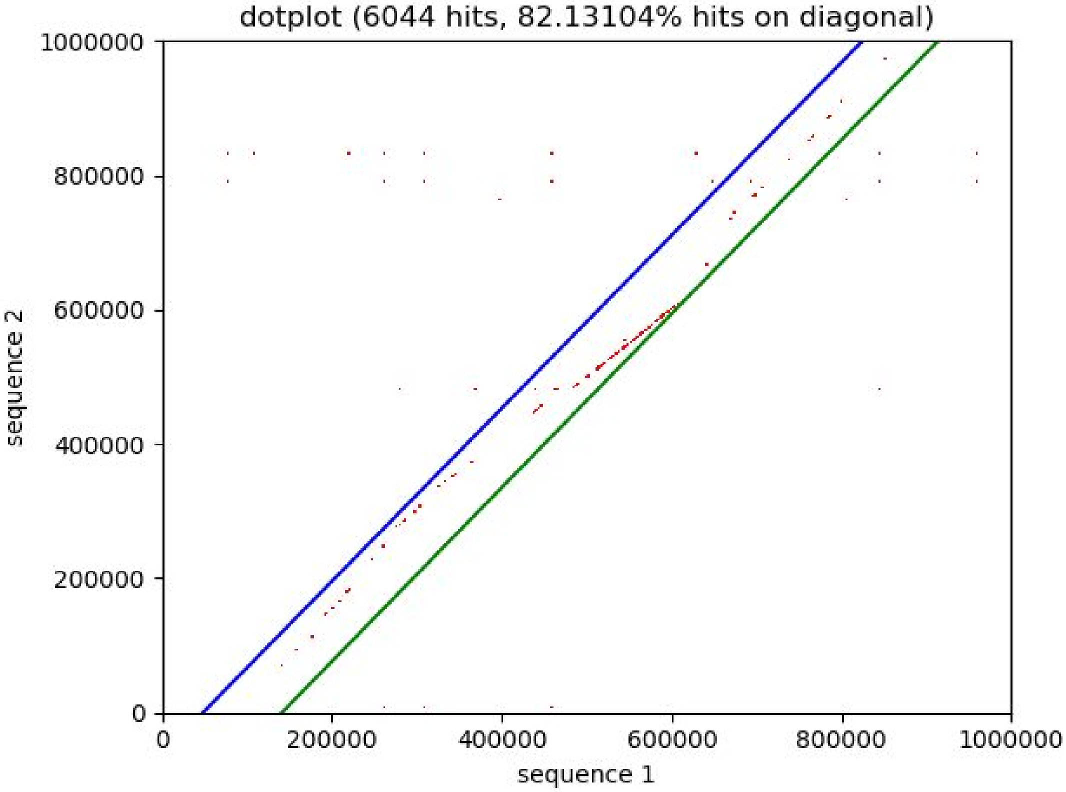
Lengthened the hash key then sliced it so it is only every other base



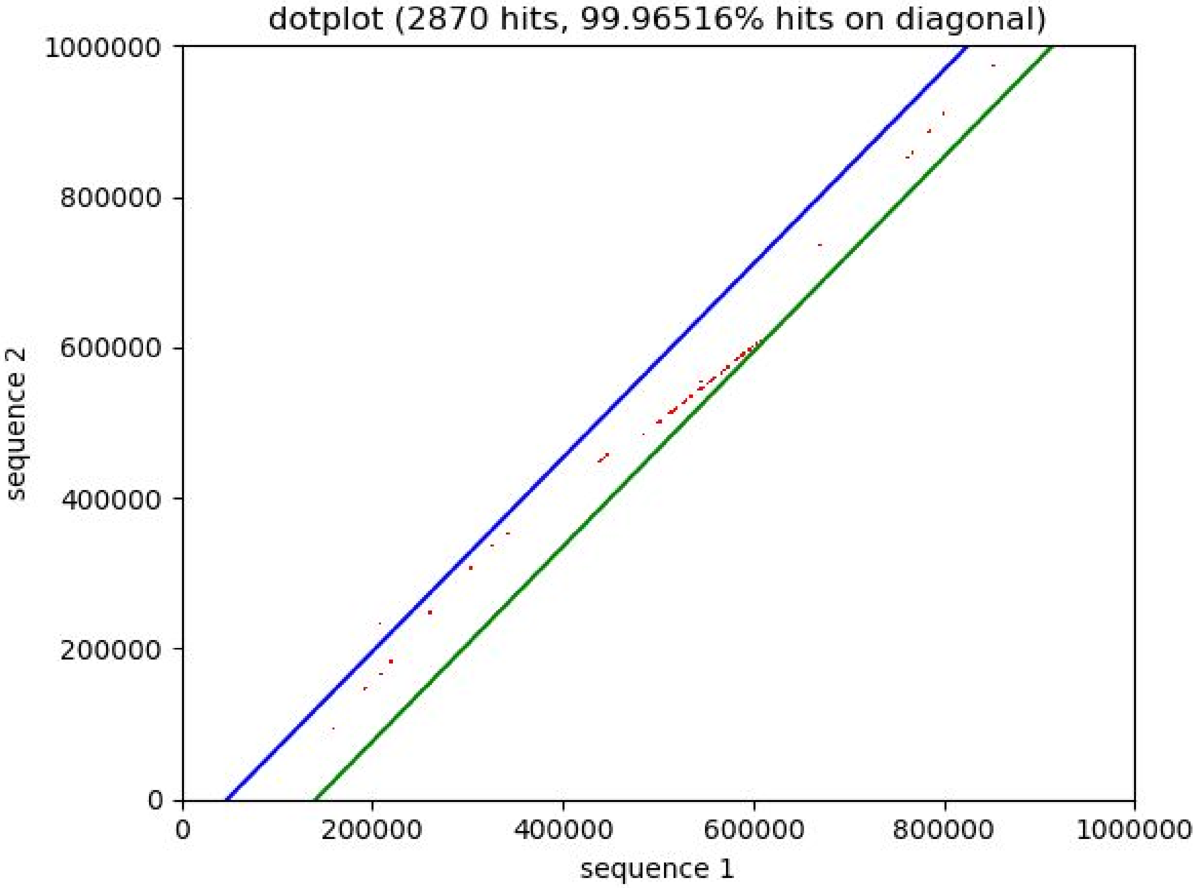
III.
Lengthened the hash key then sliced it until every third base



IV.
Lengthened the hash key to 120 but then cut away every fourth base



V.
Filled every third base of key with a filler value so that every third base could match regardless of original value


dotplot (2870 hits, 99.96516% hits on diagonal)
sequence 2
sequence 1