2. To perform semi-global, I based my algorithm off of notes I found here <http://www.cs.cmu.edu/~durand/03-711/2015/Lectures/PW_sequence_alignment_2015.pdf>. To allow gaps at the beginning and end of the second sequence, I initialized the first column to 0’s and the traced back from the highest score in the last column. This leads to the program not printing the rest of the first string after the alignment is reached because the path may not lead back to (0,0), but I believe still performs the alignment correctly.

For example, on the strings from the assignment, I got:

CAGCACTTGG

CAG—CGTGG

3.

a. (black is homo sapien, blue is m D. melanogast)

The alignment with a score of 679 is:

HSGVNQLGGVFVNGRPLPDSTRQKIVELAHSGARPCDISRILQVSNGCVSKILGRYYETGSIRPRAIGGSKPRVATPEVVS

HSGVNQLGGVFVGGRPLPDSTRQKIVELAHSGARPCDISRILQVSNGCVSKILGRYYETGSIRPRAIGGSKPRVATAEVVS

KIAQYKRECPSIFAWEIRDRLLSEGVCTNDNIPSVSSINRVLRNLASEKQQMGADGMYDKLRMLNGQTGSWGTRPGW

KISQYKRECPSIFAWEIRDRLLQENVCTNDNIPSVSSINRVLRNLAAQKEQQ-STG---S-G--SSST-

YPGTSVPGQPTQDGCQQQEGGGENTNSISSNGEDSDEAQMRLQLKRKLQRNRTSFTQEQIEAL-EK-EFERT-H-Y-

SAGNSISAKVSVSIGGN-VSN-V---ASGSRGT--LSSS-TDLMQTATPLNSSESGGASNSGEGSEQ-

PDVFARERLAAKI-DLPEARIQVWFSNRRAKWRREEKLRNQRRQASNTPSHIPISSSFSTSVYQ-PIPQ-PT-

EAIYEKLRLLNTQHAAGPGPLEPARAAPLVGQSPN-HLGTRSSHPQLVHGNHQALQ-QHQQQSWPPRHYS-

TPVSSFTSGSMLGRTDTALTNTYSAL

GSWYPTSLSEIPISSAPNIASVTAYASGPSLAHS-LSPPNDIESL

b. The scores from the 100 randomly generated sequences are : [132.0, 155.0, 132.0, 122.0, 123.0, 116.0, 131.0, 143.0, 125.0, 180.0, 132.0, 144.0, 129.0, 151.0, 163.0, 167.0, 130.0, 112.0, 124.0, 120.0, 147.0, 144.0, 145.0, 149.0, 140.0, 129.0, 155.0, 132.0, 151.0, 133.0, 124.0, 176.0, 113.0, 149.0, 192.0, 153.0, 154.0, 123.0, 156.0, 108.0, 150.0, 176.0, 141.0, 133.0, 123.0, 121.0, 120.0, 119.0, 121.0, 124.0, 120.0, 114.0, 115.0, 158.0, 133.0, 169.0, 156.0, 130.0, 158.0, 182.0, 151.0, 156.0, 149.0, 126.0, 170.0, 132.0, 138.0, 171.0, 177.0, 145.0, 165.0, 139.0, 132.0, 142.0, 148.0, 130.0, 137.0, 144.0, 141.0, 145.0, 134.0, 107.0, 123.0, 147.0, 143.0, 147.0, 135.0, 142.0, 143.0, 140.0, 141.0, 154.0, 147.0, 133.0, 150.0, 131.0, 144.0, 146.0, 132.0, 140.0]. They are all around 100-200, which would seem to imply that the sequences above indicate a statistically significant similarity.