580.488 FCBB - Problem 2
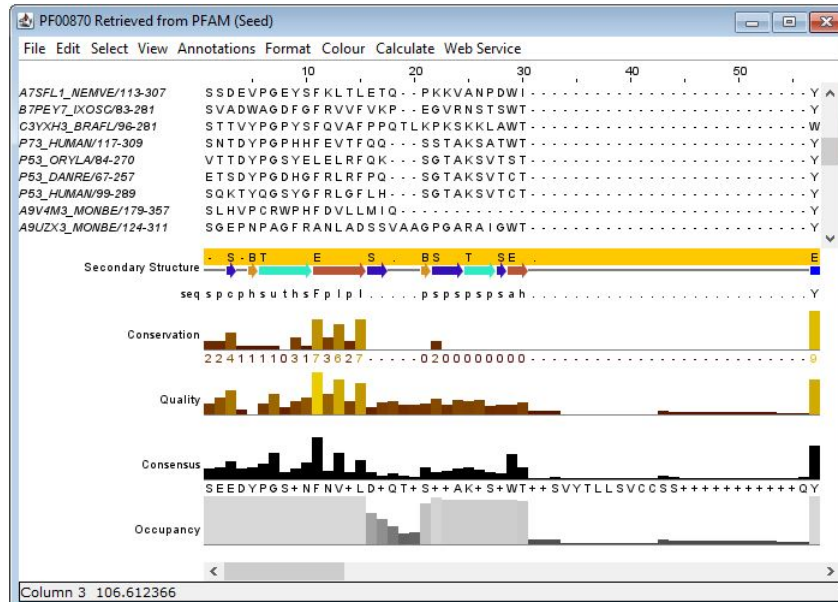
Sofía García del Barrio Cervera
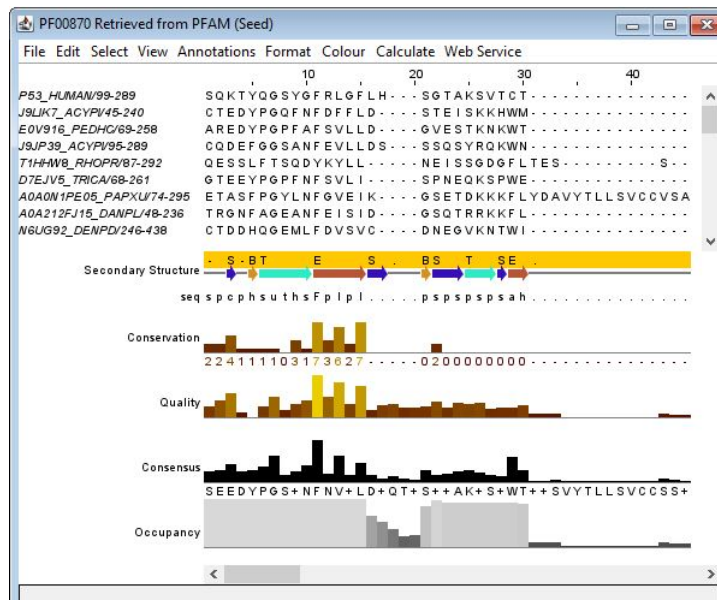
02/11/2022

For each step, I will show the JalView results.
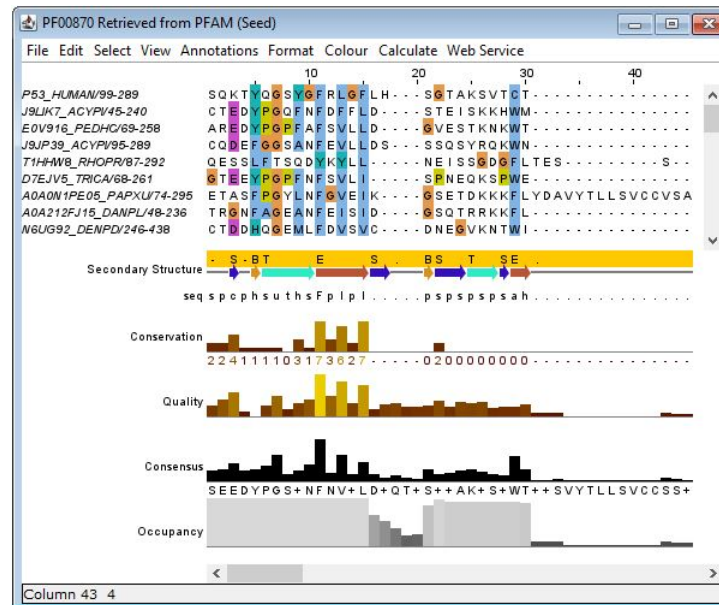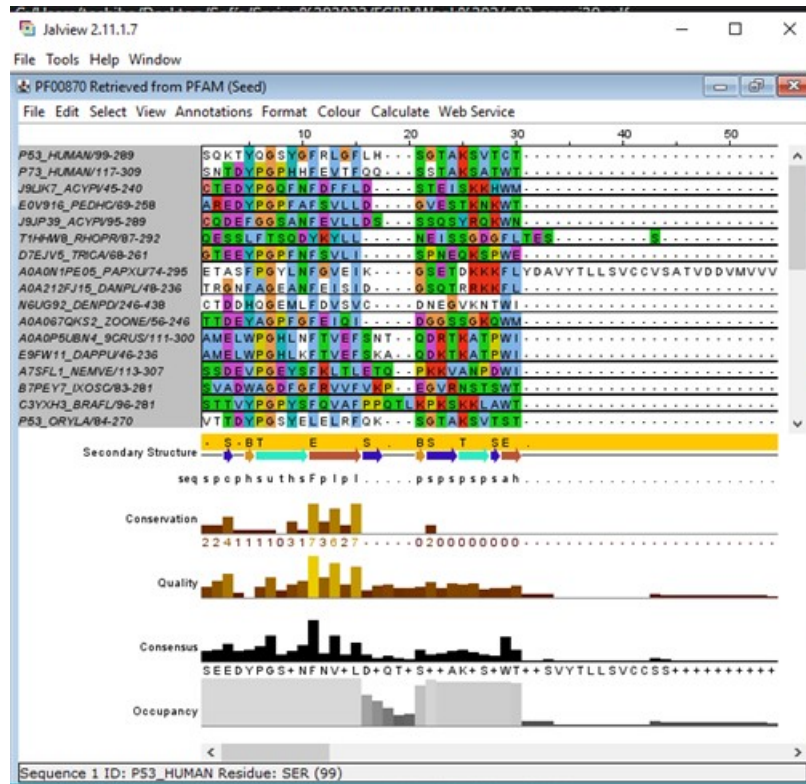
Question 1.

Question 2.

Question 3.

ClustalX colours aminoacids by some of their properties, in most of the cases its charge. The code is:

- Blue: Hydrophobic. Aminoacids A, I, L, M, F, W and V.

- Red: Positively charged. Aminoacids K and R.

- Magenta: Negative charge. Aminoacids E and D.

- Green: Polar. Aminoacids N, Q, S and T.

- Pink: Cysteine, C.

- Orange: Glycines, G.

- Yellow: Prolines, P.

- Cyan: Aromatic. Aminoacids H and Y.

Source: "Clustal Colour Scheme". 2022. Jalview.Org. https://www.jalview.org/help/html/colourSchemes/clustal.html.

Question 4.
The residue number of the first Serine is 99.

Question 5.

The type of aminoacids that are conserved in a column are:

- Blue, all hydrophobic aminoacids. Columns 64, 76, 96, 112, 149, 151, 249 and 258.

- Pink, all cysteines. Columns 115, 225 and 231.

- Red, all positively charged. Column 273.

The columns with a single aminoacid completely conserved are: 115, 225 and 231 with Cysteine (C), and 273 with Arginine (R).

Question 6.

The positions where the four binding sites occur, obtained from the UniProt page, are:

- Position 176: column 115. It is completely conserved, as all of them are Cysteines.

- Position 179: column 118. The aminoacid in the human sequence is Histidine (H). It is almost completely conserved, but there are three sequences in which have Aspartic Acid (D): Anopheles gambiae (African malaria mosquito), Anopheles darlingi (mosquito) and Anopheles sinensis (mosquito).

- Position 238: column 225. It is completely conserved, as all of them are Cysteines.

- Position 242: column 231. It is completely conserved, as all of them are Cysteines.

Question 7.

Five of the columns with missense mutations, obtained from the UniProt web, are:

- Position 132: column 63. The mutation is K− >E. No sequence has it, although some species have other aminoacids different from K.

- Position 144: column 75. The mutation is Q−>L. The sequence T1EE77HELRO from Helobdella robusta (Californian leech) has the mutation.

- Position 152: column 91. The mutation is Q−>L. The sequence J9LIK7ACYPI from Acyrthosiphon pisum (pea aphid) has the mutation.

- Position 181: column 120. There are two mutations, R−>C and R−−>H. There are two sequences with H: D7EJV5TRICA from Tribolium castaneum (Red flour beetle) and A0A0N1PE05PAPXU from Papilio xuthus (Asian swallowtail butterfly).

- Position 213: column 177. The mutation is R−>Q No sequence has it, although some species have other aminoacids different from R.

Question 8.

1. Neighbor Joining and the BLOSUM62 scoring matrix.

   (a) The closest protein to human P53 is P53-DANRE, which is in Danio rerio (Zebrafish).

   (b) P53-Human and P53-DANRE have the same latest ancestor. For the case of P73-Human, the closest common ancestor is 3 nodes behind.

   (c) When the red line is moved, the coloring of the tree changes. The clades created by the red line at a specific position have the same color.
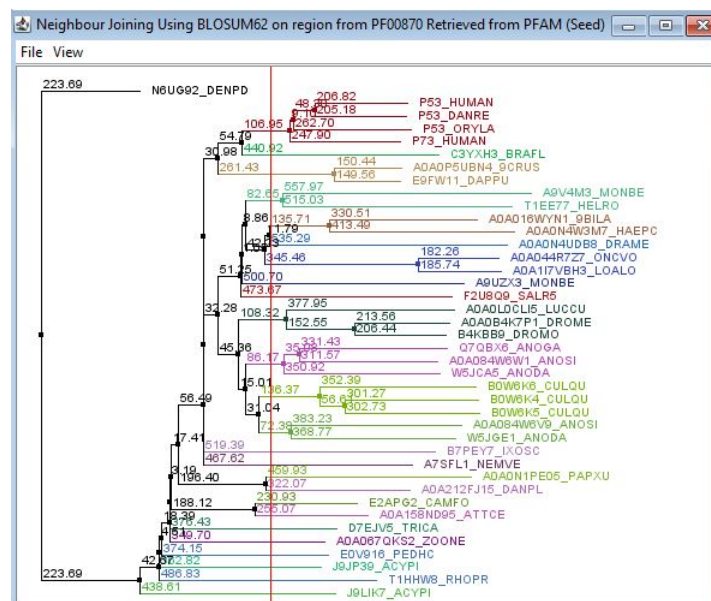
   (d) The most distant is N6GU92DENPD, which is in Dendroctonus ponderosae (Mountain pine beetle). They only share the first ancestor.

2. The plot that results from PCA and the BLOSUM62 scoring matrix is:

   (a) The three components are represented by a coordinate axis in three perpendicular directions. They are the yellow lines in the next picture.

   (b) Human P53 is selected in gray, next to the name sign.
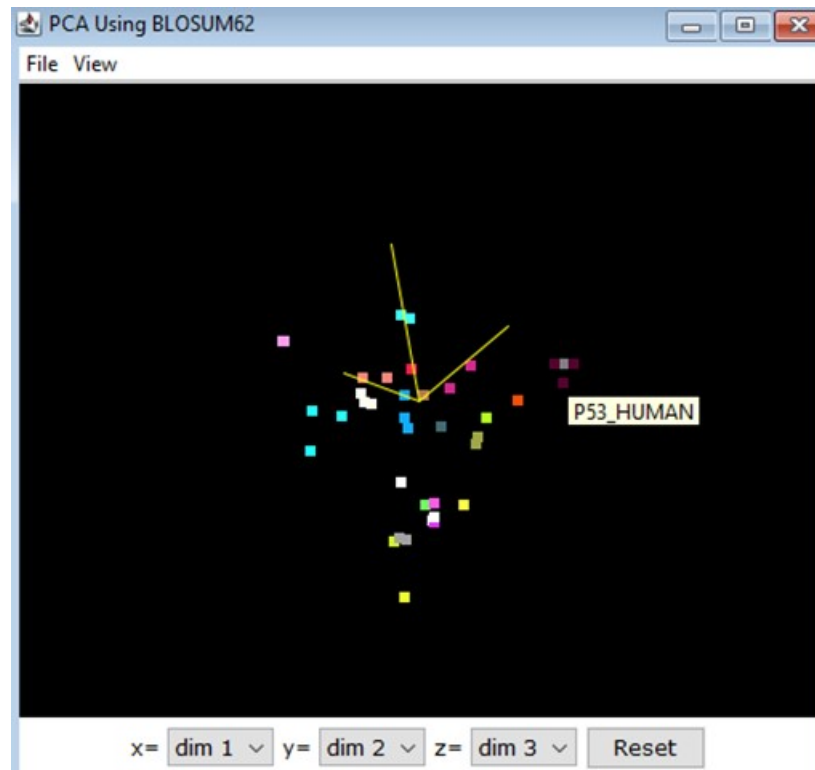
(c) The number of clades (points with the same color) seen in the PCA plot is determined by the position of the red line in the tree. For instance, in the previous PCA plot there are 19 clades represented. This is the tree plot: