BIO310 Introduction to Bioinformatics

Computer Lab 4,

Homework 2

Spring 2019

Alihan Kerestecioğlu

25497

1 Crime Investigation

Protein 1

1) Alpha s1-casein
2) Bos taurus (cattle)
3) Important role in the capacity of milk to transport calcium phosphate. Casoxin D acts as opioid antagonist and has vasorelaxing activity mediated by bradykinin B1 receptors.
   Miscellaneous: In milk, the alpha s1- and beta-caseins precipitate in presence of calcium (so-called calcium-sensitive caseins). Kappa-casein prevents the precipitation of the other caseins by calcium through the formation of large stable colloidal particles termed micelles.
4) This protein is not a suspect, since it is involved in milk and there is a carton pf milk found in the crime scene.

Protein 2

1) Alpha-amylase 1
2) Homo sapiens (Human)
3) Endohydrolysis of (1->4)-alpha-D-glucosidic linkages in polysaccharides containing three or more (1->4)-alpha-linked D-glucose units.
4) It is not a suspect since it is and ordinary protein that can be found in the human saliva.

Protein 3

1) Snake venom metalloproteinase atrolysin-D
2) Crotalus atrox (Western diamondback rattlesnake)
3) Snake venom zinc metalloproteinase that causes hemorrhage by provoking the degradation of the sub-endothelial matrix proteins (fibronectin, laminin, type IV collagen, nidogen, and gelatins)
4) It can be responsible from the death of the tourist since it is found in dangerous, wild snakes and there is no reasonable explanation for the protein being found in the milk carton.

Protein 4

1) Beta-lactoglobulin
2) Bos taurus (Bovine)
3) Primary component of whey, it binds retinol and is probably involved in the transport of that molecule.
4) It is not a suspect since this protein is derived from cows and it is normal that is found inside the milk.

Conclusion: The victim got bitten by a Western diamondback rattlesnake, and died since the bites of these species are fatal.

3 Hidden Markov Models

1.

Path1 =

(1.0) X (0.25) X (0.9) X (0.25) X (0.9) X (0.25) X (0.1) X (0.95) X (1.0) X (0.4) X (0.1) = 0.00048094

Path2 = 0

Path3 =

(1.0) X (0.25) X (0.1) X (0.95) X (1.0) X (0.4) X (0.9) X (0.1) X. (0.9) X (0.4) X (0.1) = 0.00003078

	A G T G A							Probability
Path 1	Start	E	E	E	5	I	End	0.00048094
Path 2	Start	E	E	5	I	I	End	0
Path 3	Start	E	5	I	I	I	End	0,00003078

2.

Path 1 the most likely to annotate the sequence since it has the highest probability among the state paths.

3.

The probability of 5’ splice site to be in a position where C or T nucleotides are observed is 0. Also, G being observed has a higher probability than A in 5’ splice site. So, the splice site is more likely to be in the positions 2, 4 which G is observed.