Practical 9

Questions

08 April 2024

1. Identify the pair of sequences which are close to each other using Hamming and Euclidean distance methods.

(i)AMENLNMDLLYMAAAVMMGLAAIGAAIGIGILGGKFLEGAARQPDLIPLLRTQFFIVMGLVDAIPMIAVG
LGLYVMFAVA

(ii)AADVSAAVGATGQSGMTYRLGLSWDWDKSWWQTSTGRLTGYWDAGYTYWEGGDEGAGKHSLSFAP
VFVYEFAGDSIKPFIEAGIGVAAFSGTRVGDQNLGSSLNFEDRIGAGLKFANGQSVGVRAIHYSNAGLKQPN
DGIESYSLFYKIPI

(iii)MALLPAAPGAPARATPTRWPVGCFNRPWTKWSYDEALDGIKAAGYAWTGLLTASKPSLHHATATPEY
LAALKQKSRHAA

2. Get the non-redundant sequences of beta barrel membrane proteins with sequence identities of less than 40%, 50%, 75% and 90% using CD-HIT

3. Get the non-redundant sequences of the same type of proteins with sequence identities of less than 20%, 30%, 40% and 50% using PISCES (https://dunbrack.fccc.edu/pisces/)

4. Compare the results obtained with the cut-offs 40% and 50%.

5. Extract the data with the cut-off of 50% from Uniprot and compare with CD-HIT and PISCES.

Deadline: 14th April 2024