BT1051 计算生物学 2024-2025-2

Project 3. PurZ0 & PurZ 的 SSN、GNN、以及进化树的构建

BT1051 | Project 3 | ddl 2025.4.16

1. Merge the PurZ0s you got from ncbi nr and IMG/VR v4, remove duplicates, and use the nonredundant PurZ0s to make a SSN. Please display the SSN at E-value cutoffs 10^{-75} and 10^{-120}。Mapping PDB structures, function-known sequences and key residues for NTP substrate binding to the SNN with E-value cutoff 10^{-75}, respectively.

2. Based PurZ0's SNN (now you should display it at a more stringent E-value cutoff, to make sure each cluster is isofunctional), make a GNN. From your GNN, can you infer how many different pathways PurZ0s may engage in?

3. Please also make a SNN and a GNN for PurZs, similar to what you've done for PurZ0s.

4. Choose 5 PurZ0s and 5 PurZs from phages, and 5 PurAs from representative bacteria, archaea and eukaryotes to make a multiple sequence alignment (MSA). What alignment method do you choose? Why?

5. When making a phylogenetic tree based on the MSA you made in 4, choose the proper amino acid/nucleotide substitution model is important. You may use methods such as jmodeltest to choose the model you need, and you can use BIC or AIC as the criteria for model choosing. For this case, what model do you choose according to BIC? Use the NJ method to build a phylogenetic tree and evaluate the tree through bootstrapping (300 times) and label the bootstrapping results on your tree. From your phylogenetic tree, can you infer the evolutionary origin of PurZ?