Samuel Young

August, 29, 2018

Introduction of problem

Is when this scientist went looking for a phylogenetic tree of cyanophages are virus that infect blue-green bacteria also know as cyanobacteria, they didn't fine one. There investiagation then came across several individual articles that leaded to being typed with several different phages that also affect athropods. That wasn't what they were looking for instead they were looking wheither these two distinctive of cyanophages types could have a similar gene in order for them to develop a phylogenetic tree.

Decription of the experiment

First, the scientist wrote a code the that would pick ramdomly pick the genome "Synechococcus phage ACG-2014a". In the same script they also include within it was one to randomly pick out a gene number "68". Then they created at two seperate database with the use of BLAST where there was the single gene and they other database with all the genomes and genes. The sceinetist was thus looking for a otrhology .The first BLASTp that they then ran as a database gene compared to all the other genomes . The second BLASTp that the ran against all the genomes as the database compared a majority of the other the genes. They went through both BLASTed files and found all the lowest e-values, in which they then compared the similar sizes of BLAST from both side. Considering that none of the base data pulled exactly the same. It was then decided to use the distance between the first and the second to determine how similar the were. Well in one file he had only most similar with that gene and in the other they had all the genes. Then from there they were aligned with all the genes that they had found with they uses of custal omega with the use of input method.

Result of the experiment

A majority of the organism that were looked at are primarly phages of synecoccus phages. With a possiblility of having some that actaully infect both of these two phages. Theme is that there is one that can infect only one speices and not the other in which was what the experiment was looking for.

1st protien id	amino acids
AIX14253.1.68	MDPKTRVERQDERVWCLEQLIRLEGMLDPRMYECAD YAASAGLVKDKKDLYKLWKEWKEDNPTDNPQIRNRL
YP.004323691.1.80	MDAKTRVERQDTRVWALEQLIRLEAFLDPRMYECAD YYTSSYASQVVEDLYTLWVEWKEDNPTDNPQVINRM
AD097265.1.89	MDSKTRIERQETRVWAIEHLIRHEGMLDPRMYECAD YYASSYASQVTEDLYTLWVEWKTNNPTNNPQVRNRL
AOO10255.1.71	MDPKTRLERQETRVWAIEQLIRYESFLDPRMYECAD YYASAYATQDTNYLYTLWVEWKIDNPTDNPQVVNRM
AD099474.1.78	MDSETRIERQETRVWAIEQLLRREGFLDPRMYECAD YYASGYASQNTNDLYTLWVEWKEDNPTNNPQIVNRL
YP.009133636.1.76	MDAKTRIERQETRVWALEQLIRLESFLDPRMYECAD YYTSSYASQVVEDLYTLWTEWKEDNPTSNPQVINRM
AON99309.1.76	MDPKTRVERQETRVWALEQLIRYEGFLDPRMYECAD YYTSAYASQVSEYLYTLWVEWKEEHPSDNPQVINRM

The inital aminoacid sequence match	gentical identitcal-matches
72	26
53	1
50	10
51	2
51	2
47	10

conclusion

Well instead of not knowing how many genomes I had I could have of just went with the amount of genes that are within this file. As I had determined later there were more Syneccous phage about 255 of them where that speices are about 73 of phrococcus phages. I could have of randomized that to determine what a single one that I was looking for.It was just that the data it came up with between the two files was around 198 phages were similar within that range probably about four of them are phroccuocus phages. Maybe next time I could try and find a better gene to match them all. Another thing that could have of been done would to be compare one speices form both genomes just to see the result of what one was closer for all the test. Another thing would be for me to determine which protein belong to what organism before I ran through the files because this became a tough task to keep track of.