Comparative Genomics 2018

Practical 3: Phylogenetic Reconstruction

Group number: 6

Group members: Kyle Kimler, Kajetan Juszczak

**Summary**

**Key Questions to Answer**

Ex.1 Finding homologs

1. -in input file name, we enter our fasta filename in here

-dbtype type of molecules in database in our case nucl for nucleotides

1. -
2. Blast results
   1. <Hit\_num>1</Hit\_num> - indicate best hit
   2. <Hsp\_hseq>sequence here </Hsp\_hseq>
      1. we can use max\_hsps flag to specify how many High-scoring pairs we want to get.
      2. We can specify what the output directory should be, I expect to see XLM file with best hits.

Ex.2 Parsing

B:

1. calling method .hsps with index 0 will give the best hit
2. Blast record correspond to BLAST output
3. that it contains all the hits we get while running BLAST with the flags we we specified. As default It goes from best to worst hit
4. output is a best hit

Ex.3

1. 217.00000000 gap open penalty

39.40000153 gap extension

292.60000610 terminal gap penalty

283.00000000 bonus

1. If the penalties are very high it means that the region is possibly extremally conserved.

Ex.4

1. belvu
   1. b -Scoredist distance correction – used as default

j - Jukes-Cantor distance correction

k -Kimura distance correction

s -Storm & Sonnhammer distance correction

r -uncorrected distances

* 1. We used UPGMA and neighbor-joining and for distance correction Kimura and Scoredist. There were no visible differences between different correction methods. In case of neighbor-joining and UPGMA there is a clear difference since neighbor-joining method creates unrooted tree while UPGMA tree is rooted.

1. RaxML
   1. [they are](http://etetoolkit.org/treeview/?treeid=9a49b40c1db891e02180b0ac2f80fb88&algid=ce443ed1e53858bf4e11d1e069c7a927)
   2. link to tree : <http://etetoolkit.org/treeview/treeid=9a49b40c1db891e02180b0ac2f80fb88&algid=ce443ed1e53858bf4e11d1e069c7a927>["](http://etetoolkit.org/treeview/?treeid=9a49b40c1db891e02180b0ac2f80fb88&algid=ce443ed1e53858bf4e11d1e069c7a927)
   3. -f a (f – select algorithm)it performs rapid Booststrap analysis and look for trees with highest score

-x 54321 set the number – random seed and start rapid bootstrapping

-N 100 number of alternative runs starting from different starting trees

-T 4 number of threads to run

-p 12345 specify a number of random seedfor the parsimony inferences

-m PROTCATBLOSUM62 uses substitution matrix called BLOSUM 62 for aminoacids as model.

-s input

-n output

* 1. =

Ex. 5

1. =