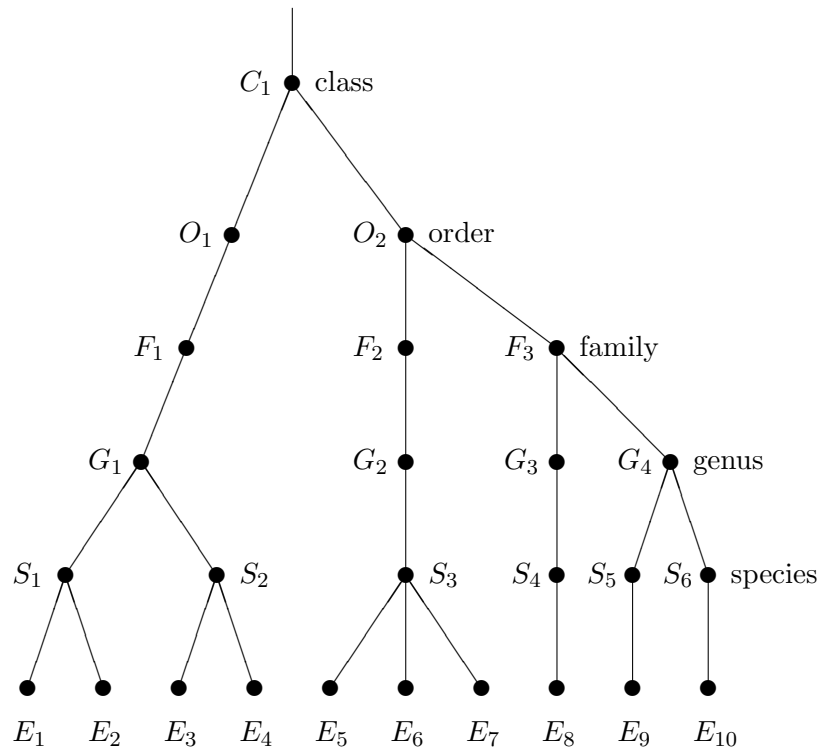


Algorithms in Genome Research  
Winter 2010/2011

Exercises

Number 07, Discussion: 2010 December 17

Assume we have given 5 metagenomic reads  $R_1, \dots, R_5$  and a database containing 10 entries  $E_1, \dots, E_{10}$  from 6 different species  $S_1, \dots, S_6$ . The following taxonomy is given for the genomes and the contained sequences:



Let the following be the BLAST output as pairs (entry, bit-score), when the reads are BLASTed against the database:

$R_1 : (E_8, 50)(E_{10}, 47)(E_6, 43)(E_9, 42)(E_3, 42)(E_5, 34)(E_1, 33)(E_2, 27)(E_7, 24)(E_4, 23)$   
 $R_2 : (E_7, 74)(E_5, 69)(E_9, 64)(E_6, 63)(E_{10}, 61)(E_4, 61)(E_1, 59)(E_2, 54)(E_8, 34)(E_3, 33)$   
 $R_3 : (E_3, 60)(E_4, 60)(E_1, 58)(E_2, 55)(E_7, 53)(E_8, 51)(E_6, 50)(E_9, 47)(E_5, 45)(E_{10}, 38)$   
 $R_4 : (E_5, 38)(E_4, 35)(E_6, 34)(E_7, 34)(E_{10}, 33)(E_1, 28)(E_8, 27)(E_2, 25)(E_9, 23)(E_3, 18)$   
 $R_5 : (E_2, 87)(E_4, 41)(E_1, 40)(E_3, 39)(E_6, 39)(E_8, 37)(E_{10}, 33)(E_9, 32)(E_7, 30)(E_5, 29)$

1. What would be the predictions by MG-RAST?
2. What would be the predictions by MEGAN with threshold 10%?
3. What would be the predictions by SOrt-ITEMS? (You will need more information to answer this question. Just make something up.)
4. Discuss additional criteria that may be used to classify metagenomic reads.