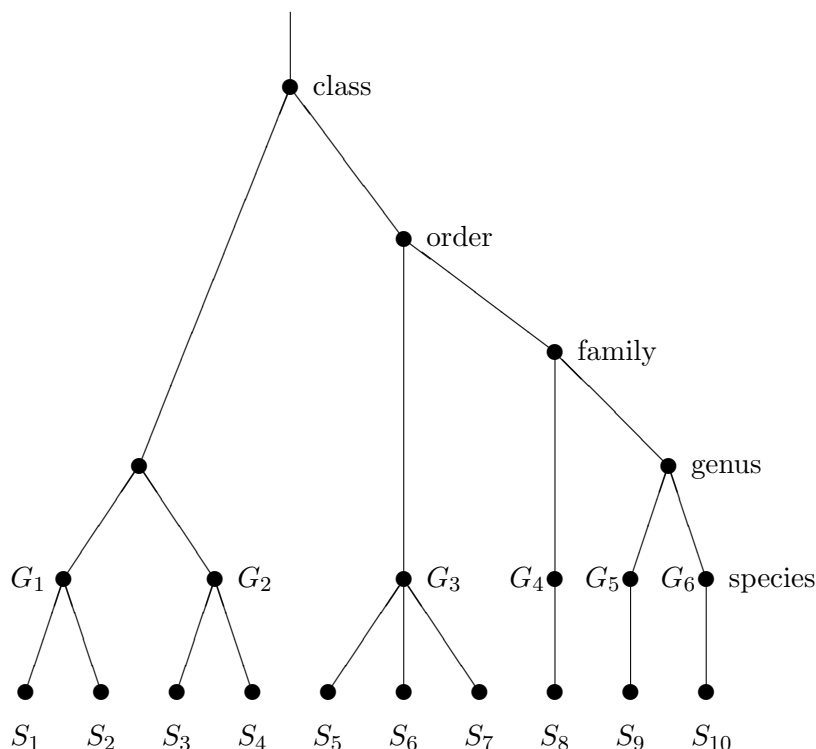


Algorithms in Genome Research
Winter 2009/2010

Exercises

Number 13, Discussion: 2010 February 05

Assume we have given 5 metagenomic reads R_1, \dots, R_5 and a database containing 10 sequences S_1, \dots, S_{10} from 6 different genomes G_1, \dots, G_6 . The following taxonomy is given for the genomes and the contained sequences:



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

$R_1 : (S_8, 50)(S_{10}, 47)(S_6, 43)(S_9, 42)(S_3, 42)(S_5, 34)(S_1, 33)(S_2, 27)(S_7, 24)(S_4, 23)$
 $R_2 : (S_7, 74)(S_5, 69)(S_9, 64)(S_6, 63)(S_{10}, 61)(S_4, 61)(S_1, 59)(S_2, 54)(S_8, 34)(S_3, 33)$
 $R_3 : (S_3, 60)(S_4, 60)(S_1, 58)(S_2, 55)(S_7, 53)(S_8, 51)(S_6, 50)(S_9, 47)(S_5, 45)(S_{10}, 38)$
 $R_4 : (S_5, 38)(S_4, 35)(S_6, 34)(S_7, 34)(S_{10}, 33)(S_1, 28)(S_8, 27)(S_2, 25)(S_9, 23)(S_3, 18)$
 $R_5 : (S_2, 87)(S_4, 41)(S_1, 40)(S_3, 39)(S_6, 39)(S_8, 37)(S_{10}, 33)(S_9, 32)(S_7, 30)(S_5, 29)$

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.