Supplementary material

Example showing how assignment algorithms operate

Consider the following hypothetical example of a DNA sequencing read that has three equally significant BLASTX matches to different sequences in the NCBI-NR database:

DNA sequencing read:

```
>002246_0060_0880 length=102
TTTCCAGTAACTGGTGGCTTTTGCCAGCACCATACCAATACTGCCTTCTGATAGCCTATA
CGTAACGCTTCAGGCGAGGATTCTGCAGCAACCGCAAAGTAG
```

BLASTX hits to NR:

```
>gi|56383327|gb|AAN42564.2| orf, conserved hypothetical protein
     [Shigella flexneri 2a str. 301] Length = 333
Score = 38.1 bits (87), Expect = 0.095
Identities = 18/18 (100%), Positives = 18/18 (100%) Frame = -2
Query: 98 FAVAAESSPEALRIGYQK 45
          FAVAAESSPEALRIGYQK
Sbjct: 32 FAVAAESSPEALRIGYQK 49

>gi|49176061|ref|NP_415456.3| alkanesulfonate transport protein
    (ABC superfamily, peri_bind) [Escherichia coli K12] Length = 333
Score = 38.1 bits (87), Expect = 0.095
Identities = 18/18 (100%), Positives = 18/18 (100%) Frame = -2
Query: 98 FAVAAESSPEALRIGYQK 45
          FAVAAESSPEALRIGYQK
Sbjct: 32 FAVAAESSPEALRIGYQK 49

>gi|75512092|ref|ZP_00734677.1| COG0715:
    ABC-type nitrate/sulfonate/bicarbonate transport systems,
    periplasmic components [Escherichia coli 53638] Length = 319
Score = 38.1 bits (87), Expect = 0.095
Identities = 18/18 (100%), Positives = 18/18 (100%) Frame = -2
Query: 98 FAVAAESSPEALRIGYQK 45
          FAVAAESSPEALRIGYQK
Sbjct: 18 FAVAAESSPEALRIGYQK 35
```

The LCA algorithm assigns this read to the taxonomic family of *Enterobacteriaceae*, because it has significant hits to both *Shigella flexneri* and to two different strains of *Escherichia coli*. The first match does not have an NBCI RefSeq accession number and is thus not used by MEGAN to perform a functional classification. The second match has RefSeq accession number `NP_415456` (shown as "`|ref|NP_415456`"). In the SEED classification, this maps to the functional role of *alkanesulfonates-binding protein*, which can be found in the *alkanesulfonate assimilation* subdivision of *sulfur metabolism*. This functional role also appears in the *putative sulfate assimilation* cluster of *clustering-based* subsystems. In the KEGG classification, this maps to the KEGG orthology group K02051, which is listed as a *sulfonate*-related *ABC transporter*.