Stalking the fourth domain in metagenomic data: searching for, discovering, and interpreting novel, deep branches in phylogenetic trees of phylogenetic marker genes

Dongying Wu, Martin Wu, Aaron Halpern, Doug Rusch, Shibu Yooseph, Marvin Frazier, J. Craig Venter, Jonathan A. Eisen

---

Supplementary Data

(1) recA data: recA.tgz

recA.tgz contains the following files:
recA_GOS.pep ------------------ Amino acid sequences for GOS RecAs
recA_ref.pep ------------------ Amino acid sequences for RecAs from NRAA and geno
recA_cluster.txt -------------- Lek clusters of RecA sequences (Table 1)
recA.ali ---------------------- Original alignment for the RecA tree (Figure 1)
recA.trim.ali ----------------- Trimmed RecA alignment that the RecA tree is buil
recA.tre ---------------------- RecA tree in Newick format (Figure 1)
recA_GOS_pepID_assemblyID.map - The assembly IDs of the *recA* encoding GOS assembli
recA_linked.pep --------------- The Amino Acid sequences of the genes that share

(2) rpoB data: rpoB.tgz

rpoB.tgz contains the following files:
rpoB_GOS.pep ------------------- Amino acid sequences for GOS RpoBs
rpoB_ref.pep ------------------- Amino acid sequences for RpoBs from NRAA and gen
rpoB_cluster.txt --------------- Lek clusters of RpoB sequences (Table 3)
rpoB.tre.ali ------------------- Original alignment for the RpoB tree (Figure 3)
rpoB.tre.trim ------------------ Trimmed RpoB alignment that the RpoB tree is bui
rpoB.tre ----------------------- RpoB tree in Newick format (Figure 3)

(3) ss-rRNA data: ssu.tgz

ssu.tgz contains the following files:
SSU_GOSreads.fa ---------------- GOS ss-rRNA sequences
SSU_GOSreads_deepbrach.fa ------ Potential GOS deep-branching ss-rRNA

(4) Lek Clustering Program: lek.tgz

lek.tgz contains scripts for the Lek clustering protocol.
Instructions can be found in the included README file.