Nifh accessions.csv Compilation of geneIDs gene IDs.r Output list of gene IDs gene_id_parser.sh Retrieves *nifH* sequences nifh sequences1.fasta 100 *nifH* sequences in fasta format msa.r for multiple sequence alignment out.fasta MView.sh To create the alignment figure Figure 1.html Figure 2.html

pipeline.sh

BASH script to run

these scripts

automatically

Ratio test.py Filter out sequences with >60% gaps DMinput.fasta distance_matrix_and tree.r Calculate the distance matrix and construct a phylogenetic tree DistMat(Indel).pdf DistMat(JC69&K80&K81&TN9 3).pdf

DistMat(TS&TV).pdf

Phylogeny.pdf