

L-data set

Wei et al. [2] collected the genomic islands in 11 genomes identified using a genome-wide comparative approach from published literature and constructed the L-data set to evaluate the prediction methods. These distinct data are from seven orders, namely *Burkholderiales* (*Burkholderia cenocepacia* J2315: NC_011000, NC_011001, NC_011002; *Bordetella petrii* DSM 12804: NC_010170), *Corynebacteriales* (*Corynebacterium diphtheriae* NCTC 13129: NC_002935), *Enterobacteriales* (*Cronobacter sakazakii* ATCC BAA-894: NC_009778; *Escherichia coli* CFT073: NC_004431; *Proteus mirabilis* HI4320: NC_010554; *Salmonella typhi* CT18: NC_003198), *Micrococcales* (*Clavibacter michiganensis* NCPPB 382: NC_009480), *Rhizobiales* (*Bartonella tribocorum* CIP 105476: NC_010161), *Lactobacillales* (*Streptococcus equi* 4047: NC_012471) and *Vibrionales* (*Vibrio cholerae* N16961: NC_002505, NC_002506).