

A) leave-one-out validation

(complete genomes, in 1 kb pieces)

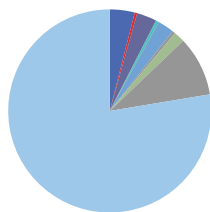
tested domain	nr. of genomes	correct position	distance 1 - 2	distance 3	distance 4 +	correct kingdom	correct phylum	correct class	correct order	correct family
Archaea (ML)	11	60.6%	15.7%	13.3%	10.3%	97.6%	86.7%	78.3%	72.2%	50.1%
Archaea (MP)	11	57.3%	19.9%	13.3%	9.5%	96.6%	85.1%	75.0%	69.4%	47.7%
Bacteria (ML)	64	53.6%	24.3%	6.0%	16.1%	99.8%	93.2%	83.8%	69.4%	53.3%
Bacteria (MP)	64	45.9%	26.7%	7.3%	20.1%	99.6%	90.7%	81.2%	64.5%	47.8%
Fungi (ML)	10	30.2%	23.4%	5.3%	41.1%	67.1%	54.7%	40.0%	36.6%	30.0%
Fungi (MP)	10	25.9%	25.1%	5.2%	43.8%	67.0%	51.4%	36.7%	33.9%	29.2%

assessing inferred tree position
(node-distance to actual target)

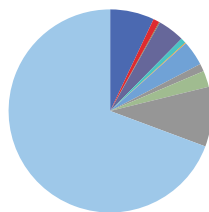
assessing inferred taxonomy
(vs. NCBI classification)

B) simulated metagenome

(*simMC*; Mavromatis et al.)



known input composition

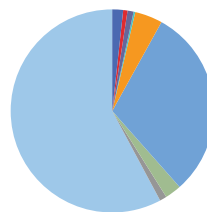


MLTreeMap classification

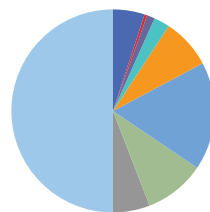


C) actual metagenome

(Singapore air sample; Tringe et al.)



16S rRNA data
(based on PCR)



MLTreeMap classification

