

Table S2: Recapture fractions of other datasets by the GPC (at 97% similarity).

taxon	EMP	SILVA	UBA	IMG/M	RDP	GTDB
Bacteria	0.93	0.96	0.90 (393/436)	0.93	0.89	0.94
Archaea	0.81	0.88	0.92 (138/150)	NA	0.90	0.78
Cyanobacteria	0.90	0.95	1.00 (5/5)	0.96	0.88	0.97
Proteobacteria	0.94	0.98	0.98 (117/120)	0.97	0.91	0.98
Firmicutes	0.88	0.98	0.86 (25/29)	0.97	0.95	0.99