

Table S5: Recapture fractions of other datasets by the GPC (at 90% similarity).

taxon	EMP	SILVA	UBA	IMG/M	RDP	GTDB
Bacteria	0.98	0.997	0.98	0.98	0.97	0.99
Archaea	0.97	0.977	0.95	NA	0.97	0.92