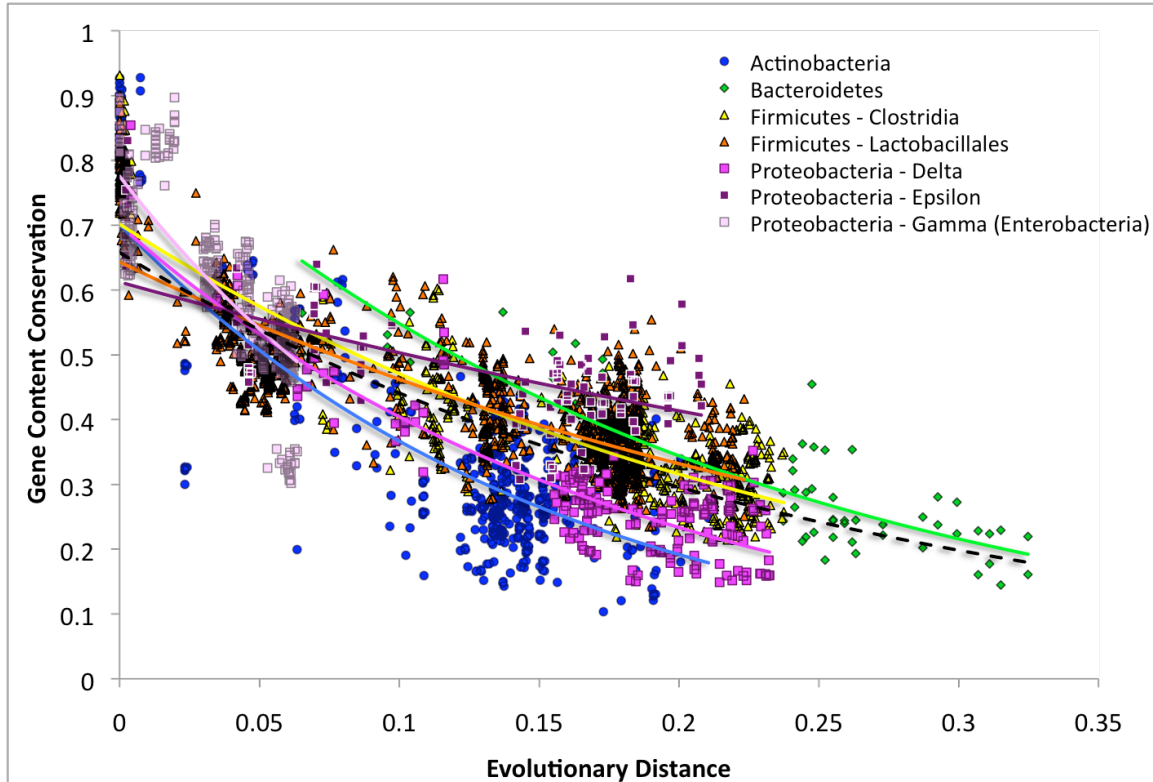


613 **Supplemental Figures**

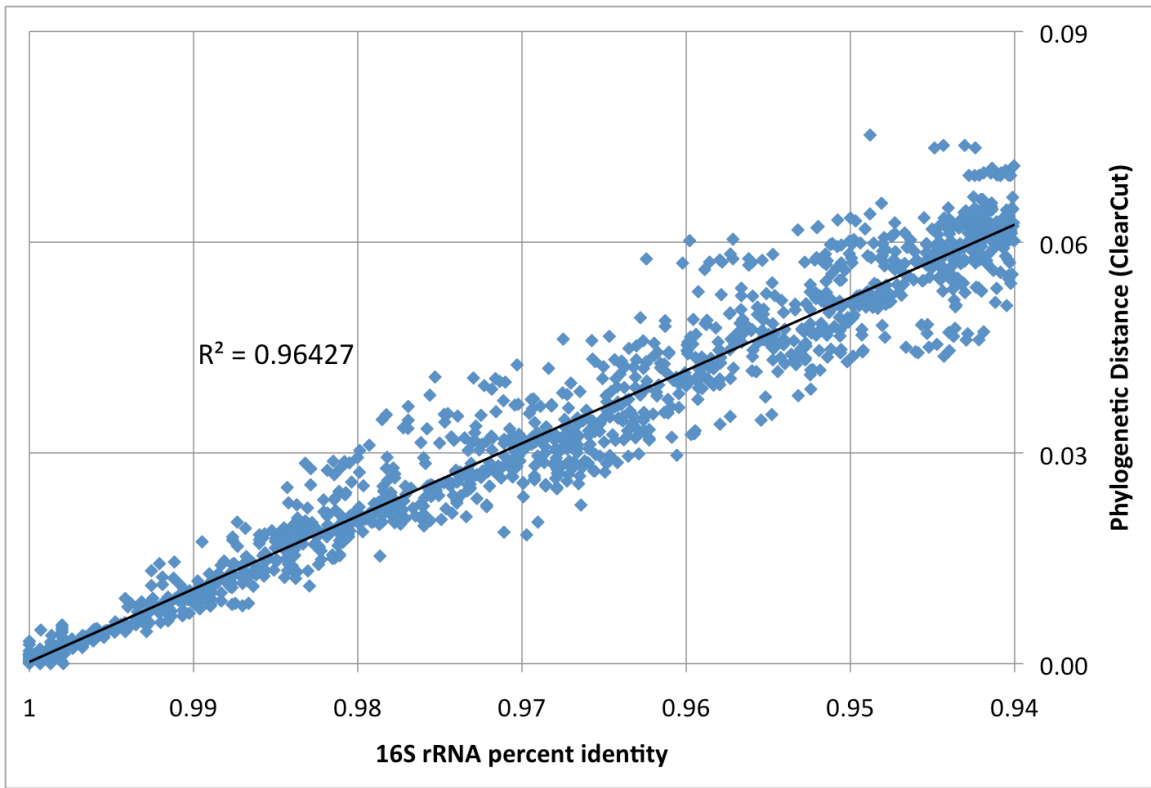


614

615 **Supplemental Figure 1. Gene conservation by evolutionary distance.** Gene
 616 sequence conservation at the peptide level. Each point represents a BLAST comparison
 617 between two genomes at a 10^{-10} threshold. The x-axis represents the 16S distance
 618 between the two genomes, while the y-axis represents the proportion of genes in the
 619 query genome present in the subject genome. Genome - genome comparisons were
 620 subdivided into series by taxonomic group, with members of the same division
 621 represented by the same shape and shades of the same color. The dashed black line
 622 represents exponential regression of the full data set. r^2 values for exponential
 623 regression of each series were: All data, $r^2 = 0.60$; Actinobacteria, $r^2 = 0.53$;
 624 Bacteroidetes, $r^2 = 0.75$; Clostridia, $r^2 = 0.57$; Lactobacillales, $r^2 = 0.70$; δ -
 625 Proteobacteria, $r^2 = 0.63$; ϵ -Proteobacteria $r^2 = 0.37$; γ -Proteobacteria $r^2 = 0.60$

626

627



628

629 **Supplemental Figure 2. Correlation between phylogenetic distance and 16S rRNA**

630 **percent identity.**

631

631

632 **Supplemental Tables**

633

634

635

Taxon	N	p (Mantel test)	P (Mantel test Log(y))
Actinobacteria	41	0.0000	0.0000
Bacteroidetes	9	0.0001	0.0000
Firmicutes Clostridia	22	0.0000	0.0000
Firmicutes - Bacilli (Lactobacillales)	43	0.0000	0.0000
Proteobacteria - Delta	16	0.0000	0.0000
Proteobacteria - Gamma	11	0.0004	0.0005
Proteobacteria - Gamma (Enterobacteria)	28	0.0000	0.0000

636 **Supplementary Table 1. Significance of correlation between evolutionary distance**

637 **and gene content conservation within diverse taxonomic groups.** Significance was

638 assessed within each taxonomic group using a Mantel test with 10,000 permutations.

639 Values were calculated for either the unmodified matrices or for the log (gene content

640 conservation). In all cases, the correlation between evolutionary distance and gene

641 content conservation was statistically significant

642

643

643

Taxon	Env	n	p (Mantel test)	p (Mantel test Log(y))
Actinobacteria	N	39	0.0000	0.0000
Bacteroidetes	G	4	0.0406	0.0414
	N	5	0.0164	0.0175
Firmicutes - Clostridia	G	4	0.0777	0.0844
	N	18	0.0000	0.0000
Firmicutes - Lactobacillales	GC	18	0.0000	0.0000
	N	24	0.0000	0.0000
Proteobacteria- Delta	N	15	0.0000	0.0000
	G	6	0.0207	0.0223
	GC	2	1.0000	1.0000
Proteobacteria- Epsilon	N	3	1.0000	1.0000
	GC	22	0.0000	0.0000
Proteobacteria- Gamma (Enterobacteria)	N	16	0.0000	0.0000

644

645 **Supplementary Table 2. Significance of correlation between evolutionary distance**

646 **and gene content conservation within gut specialist, cosmopolitan, or non-gut**

647 **members of diverse taxonomic groups.** Significance was assessed within each

648 taxonomic group using a Mantel test with 10,000 permutations. Values were calculated

649 for either the unmodified matrices or for log (gene content conservation). The

650 correlation between evolutionary distance and gene content conservation was statistically

651 significant, except in cases where very few ($n < 5$) genomes were present in a category.

652 G, gut-specialist to gut specialist comparisons; GC, gut – cosmopolitan to gut-

653 cosmopolitan comparisons; N, non-gut to non-gut comparisons.

654

654
655

Taxon	N	p (Mantel test)	p (Mantel test Log(y))
Actinobacteria	13	0.0000	0.0000
Bacteroidetes	3	0.1609	0.1724
Firmicutes - Clostridia	8	0.0000	0.0000
Firmicutes - Lactobacillales	7	0.0198	0.0386
Proteobacteria - Delta	6	0.0027	0.0029
Proteobacteria - Epsilon	-	-	-
Proteobacteria - Gamma (Enterobacteria)	23	0.0001	0.0000

656
657
658
659
660
661
662
663
664

Supplementary Table 3. Significance of correlation between evolutionary distance and gene content conservation in plasmids. Significance was assessed within each taxonomic group using a Mantel test with 10,000 permutations. Values were calculated for either the unmodified matrices or for evolutionary distance and log (gene content conservation). Surprisingly, the correlation between evolutionary distance and gene content conservation was statistically significant, except in Bacteroidetes, where few (n=3) plasmids were present in the data set.