

Genes for cooperation are not more likely to be carried by plasmids

Supplementary Material 1

Abstract

This document contains full details of all our analyses. It is also available to download as both an R markdown file and an R script, to enable full access to all our code. Data and phylogeny files are also provided.

Contents

1	All models & results	2
1.1	Genes for cooperation are not more likely to be carried on plasmids	2
1.1.1	MCMCglmm with supertree phylogeny	2
1.1.2	MCMCglmm with GTDB phylogeny	2
1.1.3	MCMCglmm with GTDB phylogeny (individual genome level)	2
1.2	Chromosomes carry the vast majority of genes for cooperation	3
1.2.1	Genes coding for cooperative & non-cooperative traits on plasmids & chromosomes . .	3
1.2.2	Proportion of genes carried on plasmid(s) in an average genome	4
1.2.3	Proportion of genes coding for cooperative traits in an average genome	4
1.2.4	Proportion of genes for cooperative traits carried on plasmid(s)	5
1.3	Influence of including other kinds of genes for cooperative traits	5
1.3.1	Genes coding for extracellular proteins	5
1.3.2	Genes with a cooperative functional annotation	5
1.3.3	Genes which are part of a cooperative secondary metabolite cluster	5
2	Supplementary figures	8
2.1	Additional data figure	8
2.2	Phylogenies	9
2.2.1	Supertree	9
2.2.2	GTDDB tree	10
3	Species table	11

1 All models & results

1.1 Genes for cooperation are not more likely to be carried on plasmids

1.1.1 MCMCglmm with supertree phylogeny

First, we tested whether plasmid(s) had a proportion of genes coding for cooperative traits compared to chromosome(s). We calculated the difference in plasmid and chromosome proportion for each genome, and then calculated the mean for each species. We then used a MCMCglmm analysis to examine this mean difference across species, controlling for phylogenetic relationships between species.

We found that across species, chromosome(s) had a higher proportion of genes coding for cooperative traits compared to plasmid(s).

Table S1: Results from the above MCMCglmm with species' mean difference in plasmid and chromosome proportion of genes coding for cooperative traits as the response variable and phylogeny as a random effect.

	post.mean	l-95% CI	u-95% CI	eff.samp	pMCMC	signif.
(Intercept)	-0.06638	-0.1179	-0.01383	9700	0.0167	*
			R-squared value			
Random effect			0.4834			

1.1.2 MCMCglmm with GTDB phylogeny

We also did the same analysis for species as defined by the GTDB phylogeny. There were 4364 genomes for which we could identify a representative genome within the GTDB tree, comprising 284 species clusters. Of these, 132 had at least 10 corresponding genomes in our dataset. To make the analysis as analogous to our main model as possible, we first took the mean difference for those 132 species clusters and used a MCMCglmm to compared this across species, while controlling for phylogeny using the GTDB tree.

We found very similar results to our main analysis: chromosomes carried a higher proportion of genes coding for cooperative traits compared to plasmids.

Table S2: Results from the above MCMCglmm with species' mean difference in plasmid and chromosome proportion of genes coding for cooperative traits as the response variable and phylogeny (GTDB tree) as a random effect.

	post.mean	l-95% CI	u-95% CI	eff.samp	pMCMC	signif.
(Intercept)	-0.06138	-0.115	-0.008094	9700	0.02825	*
			R-squared value			
Random effect			0.462			

1.1.3 MCMCglmm with GTDB phylogeny (individual genome level)

As an additional robustness check, we also analysed the difference in plasmid and chromosome proportion of genes coding for cooperative traits across all 4364 genomes with representatives in the GTDB tree. Because the data is now individual genomes, we controlled for both phylogenetic relationship between species clusters and also sample size within species clusters.

We again found that chromosomes carried a higher proportion of genes coding for cooperative traits compared to plasmids.

Table S3: Results from the above MCMCglmm with genome-level difference in plasmid and chromosome proportion of genes coding for cooperative traits as the response variable, and phylogeny (GTDB tree) and species name as a random effect.

	post.mean	l-95% CI	u-95% CI	eff.samp	pMCMC	signif.
(Intercept)	-0.0634	-0.1132	-0.01264	9373	0.0167	*
				R-squared value		
Random effect: phylogeny				0.2670		
Random effect: species				0.2481		
Total model				0.5151		

1.2 Chromosomes carry the vast majority of genes for cooperation

1.2.1 Genes coding for cooperative & non-cooperative traits on plasmids & chromosomes

We first estimated the average number of genes coding for cooperative and non-cooperative traits. To do this, we first calculated the mean number of each kind of gene for each species, and then used an intercept-only MCMCglmm analysis to estimate the average number of genes while controlling for phylogenetic relationships across species. Below are the results for these MCMCglmm analyses.

Note: The variance across species in the number of genes on plasmids & chromosomes was very large. For this reason, the 95% Credible Intervals are inflated compared to if we were to calculate $1.96 \times \text{SE}$ of the data, which would be an approximate 95% Confidence Interval. These analyses are not designed as hypothesis tests (i.e. we are not examining whether a certain value is contained within the 95% Credible Interval), so we have instead reported Standard Error of the Means in the main text as a more meaningful value for the reader to assess variability in the data. Additionally, if you run the corresponding R code on your own computer, the exact values may vary slightly due to the stochasticity of MCMCglmm analyses.

Table S4: Results from the above MCMCglmm with species' mean number of genes coding for cooperative traits on plasmid(s) as the response variable and phylogeny as a random effect.

	post.mean	l-95% CI	u-95% CI	eff.samp	pMCMC	signif.
(Intercept)	5.599	-1.241	12.14	9807	0.09649	.
				R-squared value		
Random effect				0.2886		

Table S5: Results from the above MCMCglmm with species' mean number of genes coding for non-cooperative traits on plasmid(s) as the response variable and phylogeny as a random effect.

	post.mean	l-95% CI	u-95% CI	eff.samp	pMCMC	signif.
(Intercept)	201.9	-60.99	456.8	9700	0.1153	
				R-squared value		
Random effect				0.537		

Table S6: Results from the above MCMCglmm with species' mean number of genes coding for cooperative traits on chromosome(s) as the response variable and phylogeny as a random effect.

	post.mean	l-95% CI	u-95% CI	eff.samp	pMCMC	signif.
(Intercept)	86.86	-142	316	9326	0.4299	
			R-squared value			
Random effect			0.9945			

Table S7: Results from the above MCMCglmm with species' mean number of genes coding for non-cooperative traits on chromosome(s) as the response variable and phylogeny as a random effect.

	post.mean	l-95% CI	u-95% CI	eff.samp	pMCMC	signif.
(Intercept)	2748	-2251	8129	9700	0.2845	
			R-squared value			
Random effect			0.9992			

1.2.2 Proportion of genes carried on plasmid(s) in an average genome

We also estimated the average proportion of genes in a bacterial genome that are carried on plasmid(s). We did this first by calculating the proportion per genome, and then taking the mean proportion for each species. We then examined this value across species using a MCMCglmm analysis controlling for phylogeny.

Table S8: Results from the above MCMCglmm with species' mean proportion of genes carried on plasmid(s) as the response variable and phylogeny as a random effect. We arcsine square-root transformed the proportions before analysis, and we then backtransformed the output of the MCMCglmm to a proportion.

	post.mean	l-95% CI	u-95% CI	eff.samp	pMCMC	signif.
(Intercept)	0.0422	0.0125	0.0907	9700	0.001237	**
			R-squared value			
Random effect			0.6503			

1.2.3 Proportion of genes coding for cooperative traits in an average genome

We also estimated the average proportion of genes in a bacterial genome that coded for cooperative traits. We did this first by calculating the proportion per genome, and then taking the mean proportion for each species. We then examined this value across species using a MCMCglmm analysis controlling for phylogeny.

Table S9: Results from the above MCMCglmm with species' mean proportion of genes coding for cooperative traits as the response variable and phylogeny as a random effect. We arcsine square-root transformed the proportions before analysis, and we then backtransformed the output of the MCMCglmm to a proportion.

	post.mean	l-95% CI	u-95% CI	eff.samp	pMCMC	signif.
(Intercept)	0.028	0.0096	0.0556	9857	0.0004124	***
			R-squared value			
Random effect			0.9659			

1.2.4 Proportion of genes for cooperative traits carried on plasmid(s)

We also estimated the average proportion of genes for cooperative traits which are carried on plasmid(s). As above, we first calculated the proportion for each genome, and then calculated the mean proportion for each species. We examined this across species using a MCMCglmm controlling for phylogeny.

Table S10: Results from the above MCMCglmm with species' mean proportion of genes for cooperative traits carried on plasmid(s) as the response variable and phylogeny as a random effect. We arcsine square-root transformed the proportions before analysis, and we then backtransformed the output of the MCMCglmm to a proportion.

	post.mean	l-95% CI	u-95% CI	eff.samp	pMCMC	signif.
(Intercept)	0.0217	0.0031	0.0586	10009	0.00433	**
				R-squared value		
Random effect				0.5429		

1.3 Influence of including other kinds of genes for cooperative traits

We compared the plasmid and chromosome proportions of genes for cooperative traits found by each of these three modules of SOCfinder, to examine whether our results were consistent when we considered each separately.

1.3.1 Genes coding for extracellular proteins

Table S11: Results from the above MCMCglmm with species' mean difference in plasmid and chromosome proportion of genes coding for extracellular proteins as the response variable and phylogeny as a random effect.

	post.mean	l-95% CI	u-95% CI	eff.samp	pMCMC	signif.
(Intercept)	-0.03863	-0.09841	0.01739	9502	0.174	
				R-squared value		
Random effect				0.5527		

1.3.2 Genes with a cooperative functional annotation

Table S12: Results from the above MCMCglmm with species' mean difference in plasmid and chromosome proportion of genes with a cooperative functional annotation as the response variable and phylogeny as a random effect.

	post.mean	l-95% CI	u-95% CI	eff.samp	pMCMC	signif.
(Intercept)	-0.07326	-0.1113	-0.03896	9700	0.0006186	***
				R-squared value		
Random effect				0.5316		

1.3.3 Genes which are part of a cooperative secondary metabolite cluster

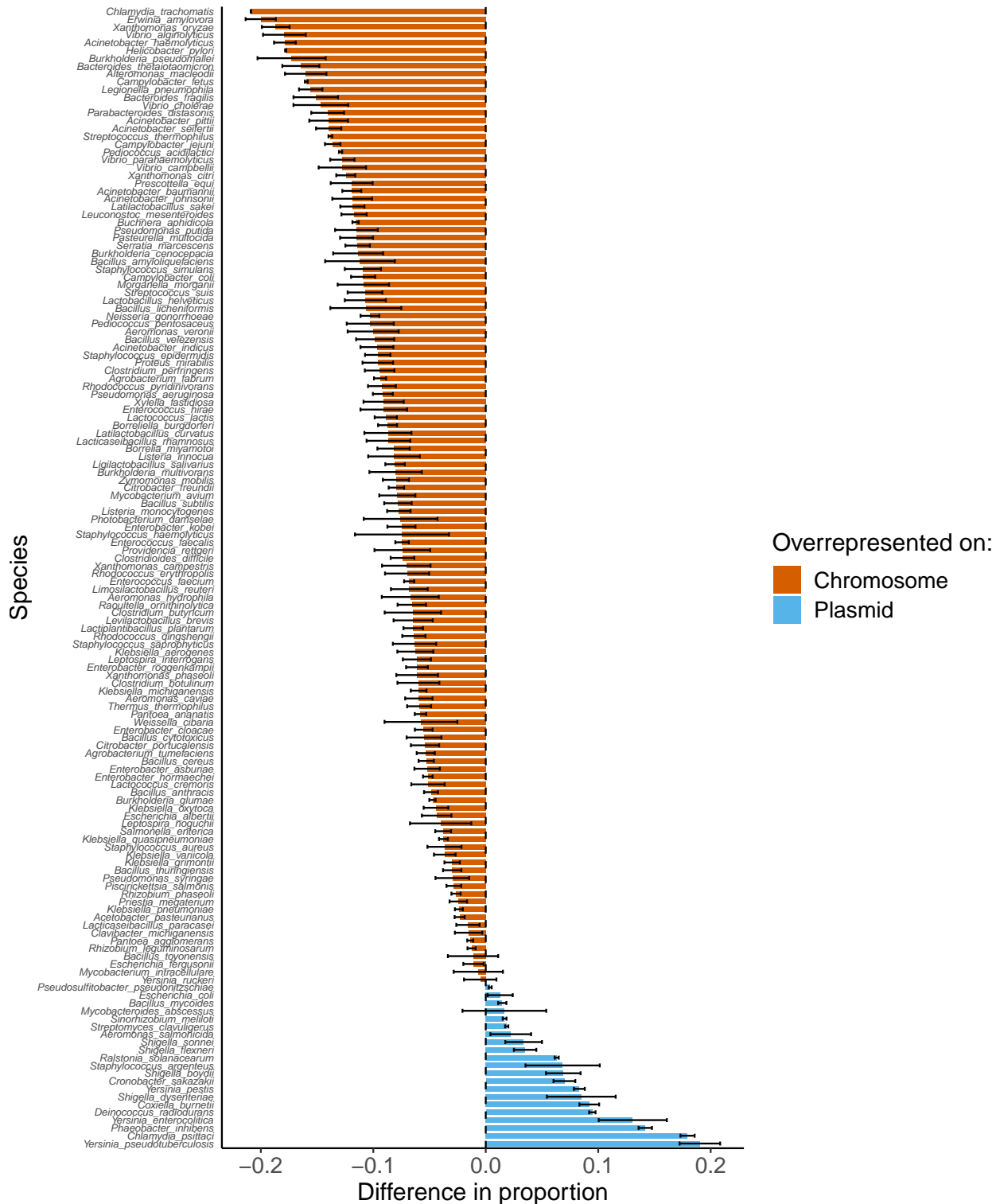
Note: N=97 species for this analysis, because 49 species had genomes which did not have any genes found by this SOCfinder module. Therefore we removed them from this analysis, since there is no mathematical definition for a proportion with 0 as both the numerator and denominator.

Table S13: Results from the above MCMCglmm with species' mean difference in plasmid and chromosome proportion of genes which are part of a cooperative secondary metabolite cluster as the response variable and phylogeny as a random effect.

	post.mean	l-95% CI	u-95% CI	eff.samp	pMCMC	signif.
(Intercept)	-0.04372	-0.08043	-0.006574	9305	0.02907	*
			R-squared value			
Random effect			0.3747			

2 Supplementary figures

2.1 Additional data figure



2.2.2 GTDB tree

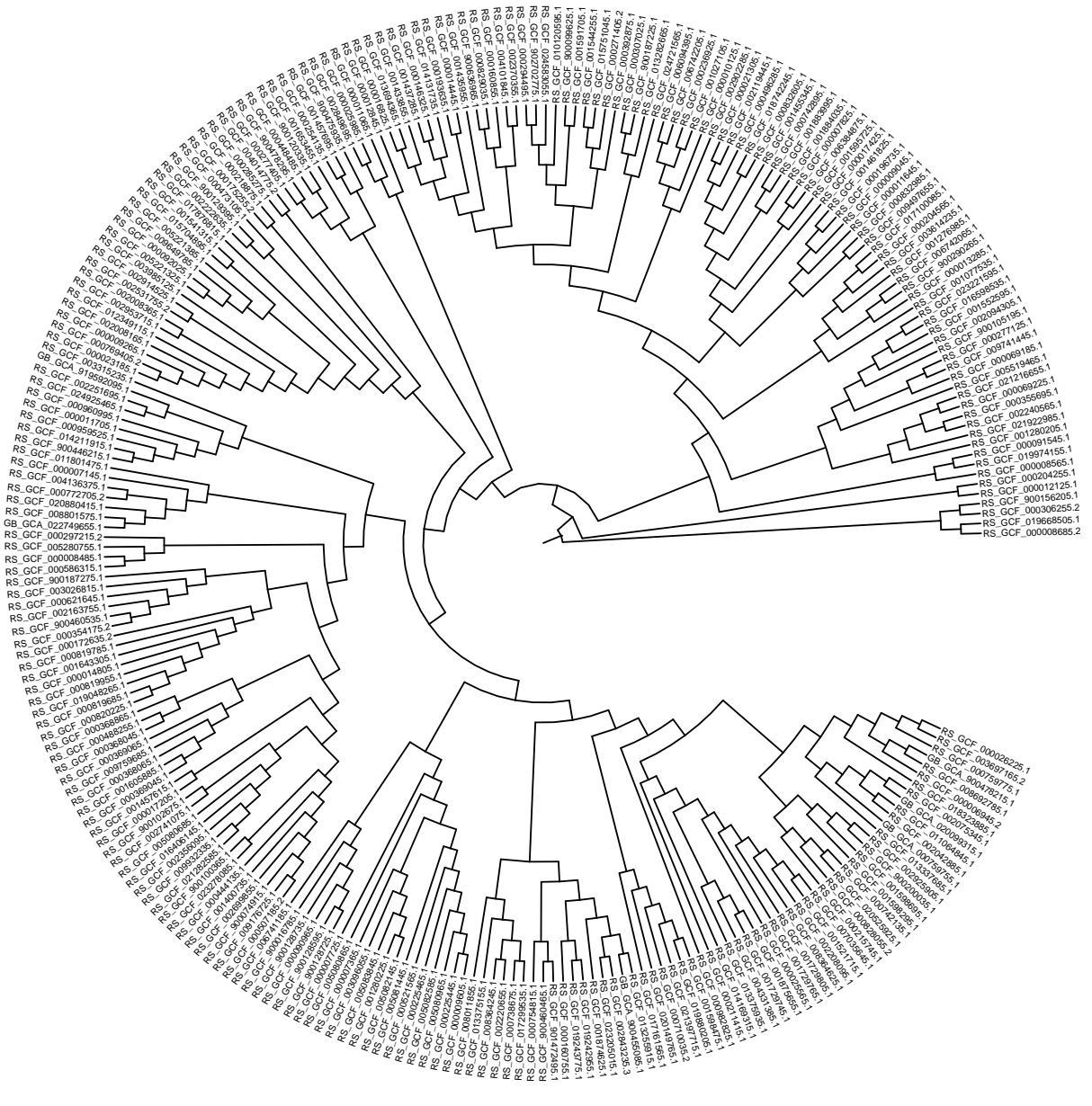


Figure S3: GTDB phylogeny for the 4364 of our genomes that we could identify a representative genome within the GTDB. See methods and code for more details. Raw tree file available to download.

3 Species table

Table S14: List of all all 146 species in our dataset, and their mean number of genes for cooperative and non-cooperative traits carried on plasmid(s) and chromosome(s) in a genome.

Species	Plas. & coop.	Chr. & coop.	Plas. & non-coop.	Chr & non-coop.
<i>Acetobacter pasteurianus</i>	8	74	341	2435
<i>Acinetobacter baumannii</i>	3	155	146	3467
<i>Acinetobacter haemolyticus</i>	0	111	87	3022
<i>Acinetobacter indicus</i>	1	78	108	2599
<i>Acinetobacter johnsonii</i>	1	90	166	3053
<i>Acinetobacter pittii</i>	1	162	147	3480
<i>Acinetobacter seifertii</i>	2	152	189	3540
<i>Aeromonas caviae</i>	3	146	127	3836
<i>Aeromonas hydrophila</i>	4	173	104	4089
<i>Aeromonas salmonicida</i>	8	174	158	3822
<i>Aeromonas veronii</i>	2	163	73	3874
<i>Agrobacterium fabrum</i>	5	144	587	4267
<i>Agrobacterium tumefaciens</i>	11	143	520	4477
<i>Alteromonas macleodii</i>	1	187	110	3621
<i>Bacillus amyloliquefaciens</i>	1	139	29	3664
<i>Bacillus anthracis</i>	4	172	185	4869
<i>Bacillus cereus</i>	9	188	287	4973
<i>Bacillus cytotoxigenus</i>	2	114	91	3689
<i>Bacillus licheniformis</i>	2	168	66	4085
<i>Bacillus mycoides</i>	20	202	414	4904
<i>Bacillus subtilis</i>	1	152	43	3906
<i>Bacillus thuringiensis</i>	20	201	564	5137
<i>Bacillus toyonensis</i>	18	202	415	4905
<i>Bacillus velezensis</i>	1	138	55	3631
<i>Bacteroides fragilis</i>	0	142	53	3970
<i>Bacteroides thetaiotaomicron</i>	0	166	50	4540
<i>Borrelia miyamotoi</i>	2	12	340	771
<i>Borrelia burgdorferi</i>	1	13	479	797
<i>Buchnera aphidicola</i>	0	7	7	506
<i>Burkholderia cenocepacia</i>	16	271	416	6227
<i>Burkholderia glumae</i>	9	230	338	5029
<i>Burkholderia multivorans</i>	3	219	112	5379
<i>Burkholderia pseudomallei</i>	1	294	57	5630
<i>Campylobacter coli</i>	1	39	70	1625
<i>Campylobacter fetus</i>	0	41	51	1597
<i>Campylobacter jejuni</i>	0	40	52	1574
<i>Chlamydia psittaci</i>	1	38	6	928
<i>Chlamydia trachomatis</i>	0	38	8	847
<i>Citrobacter freundii</i>	4	169	232	4445
<i>Citrobacter portucalensis</i>	6	159	238	4382
<i>Clavibacter michiganensis</i>	2	103	61	2740
<i>Clostridioides difficile</i>	1	66	87	3547
<i>Clostridium botulinum</i>	4	98	117	3300
<i>Clostridium butyricum</i>	12	110	454	3359
<i>Clostridium perfringens</i>	1	87	90	2625
<i>Coxiella burnetii</i>	2	27	36	1894
<i>Cronobacter sakazakii</i>	9	114	155	3808
<i>Deinococcus radiodurans</i>	8	42	158	2584
<i>Enterobacter asburiae</i>	7	184	234	4191
<i>Enterobacter cloacae</i>	8	195	301	4282
<i>Enterobacter hormaechei</i>	9	196	304	4191
<i>Enterobacter kobei</i>	7	202	290	4228
<i>Enterobacter roggenkampii</i>	5	184	201	4207
<i>Enterococcus faecalis</i>	1	61	116	2605
<i>Enterococcus faecium</i>	2	46	263	2477
<i>Enterococcus hirae</i>	2	54	85	2311

Table S14: List of all 146 species in our dataset, and their mean number of genes for cooperative and non-cooperative traits carried on plasmid(s) and chromosome(s) in a genome. (continued)

Species	Plas. & coop.	Chr. & coop.	Plas. & non-coop.	Chr & non-coop.
<i>Erwinia amylovora</i>	0	144	25	3070
<i>Escherichia albertii</i>	7	191	182	4108
<i>Escherichia coli</i>	10	206	196	4395
<i>Escherichia fergusonii</i>	8	168	206	4043
<i>Helicobacter pylori</i>	0	44	6	1372
<i>Klebsiella aerogenes</i>	4	189	185	4565
<i>Klebsiella grimontii</i>	8	199	315	5103
<i>Klebsiella michiganensis</i>	7	220	349	5183
<i>Klebsiella oxytoca</i>	7	216	259	5164
<i>Klebsiella pneumoniae</i>	9	171	319	4802
<i>Klebsiella quasipneumoniae</i>	7	168	306	4687
<i>Klebsiella variicola</i>	7	171	295	4912
<i>Lactocaseibacillus paracasei</i>	1	41	93	2727
<i>Lactocaseibacillus rhamnosus</i>	0	40	42	2617
<i>Lactiplantibacillus plantarum</i>	2	54	115	2804
<i>Lactobacillus helveticus</i>	0	28	23	1764
<i>Lactococcus cremoris</i>	2	37	141	2155
<i>Lactococcus lactis</i>	1	42	130	2236
<i>Latilactobacillus curvatus</i>	1	28	58	1691
<i>Latilactobacillus sakei</i>	0	34	52	1729
<i>Legionella pneumophila</i>	1	108	92	2809
<i>Leptospira interrogans</i>	3	118	191	3550
<i>Leptospira noguchii</i>	3	102	108	3412
<i>Leuconostoc mesenteroides</i>	0	32	84	1801
<i>Levilactobacillus brevis</i>	1	38	115	2130
<i>Ligilactobacillus salivarius</i>	1	31	235	1569
<i>Limosilactobacillus reuteri</i>	1	28	63	1887
<i>Listeria innocua</i>	1	49	106	2696
<i>Listeria monocytogenes</i>	3	56	151	2729
<i>Morganella morganii</i>	1	101	40	3524
<i>Mycobacterium avium</i>	1	87	121	4480
<i>Mycobacterium intracellulare</i>	1	89	105	5058
<i>Mycobacteroides abscessus</i>	1	71	43	4814
<i>Neisseria gonorrhoeae</i>	1	52	23	1922
<i>Pantoea agglomerans</i>	22	123	711	3474
<i>Pantoea ananatis</i>	6	140	317	3879
<i>Parabacteroides distasonis</i>	0	114	31	3913
<i>Pasteurella multocida</i>	1	42	50	2013
<i>Pediococcus acidilactici</i>	0	30	51	1783
<i>Pediococcus pentosaceus</i>	0	25	73	1606
<i>Phaeobacter inhibens</i>	30	58	403	3334
<i>Photobacterium damsela</i>	4	113	104	3411
<i>Piscirickettsia salmonis</i>	6	68	259	2498
<i>Prescottella equi</i>	0	120	71	4478
<i>Priestia megaterium</i>	7	118	413	4911
<i>Proteus mirabilis</i>	2	124	83	3416
<i>Providencia rettgeri</i>	2	108	94	3688
<i>Pseudomonas aeruginosa</i>	4	313	126	5673
<i>Pseudomonas putida</i>	2	184	157	5029
<i>Pseudomonas syringae</i>	6	280	124	4954
<i>Pseudosulfitobacter pseudonitzschiae</i>	13	46	874	3404
<i>Ralstonia solanacearum</i>	108	143	1375	3154
<i>Raoultella ornithinolytica</i>	6	166	372	4806
<i>Rhizobium leguminosarum</i>	50	122	2117	4431
<i>Rhizobium phaseoli</i>	37	106	1972	3987
<i>Rhodococcus erythropolis</i>	3	142	254	5489
<i>Rhodococcus pyridinivorans</i>	2	113	312	4263
<i>Rhodococcus qingshengii</i>	5	142	457	5485
<i>Salmonella enterica</i>	6	185	151	4208

Table S14: List of all 146 species in our dataset, and their mean number of genes for cooperative and non-cooperative traits carried on plasmid(s) and chromosome(s) in a genome. (*continued*)

Species	Plas. & coop.	Chr. & coop.	Plas. & non-coop.	Chr & non-coop.
<i>Serratia marcescens</i>	2	183	89	4547
<i>Shigella boydii</i>	10	144	123	4004
<i>Shigella dysenteriae</i>	12	146	124	3863
<i>Shigella flexneri</i>	11	148	174	4039
<i>Shigella sonnei</i>	9	169	135	4211
<i>Sinorhizobium meliloti</i>	71	74	2660	3417
<i>Staphylococcus argenteus</i>	2	101	27	2388
<i>Staphylococcus aureus</i>	1	108	25	2544
<i>Staphylococcus epidermidis</i>	1	68	34	2232
<i>Staphylococcus haemolyticus</i>	1	54	37	2192
<i>Staphylococcus saprophyticus</i>	1	51	59	2295
<i>Staphylococcus simulans</i>	1	54	61	2310
<i>Streptococcus suis</i>	0	36	7	1988
<i>Streptococcus thermophilus</i>	0	30	3	1571
<i>Streptomyces clavuligerus</i>	36	223	696	5095
<i>Thermus thermophilus</i>	2	33	274	1949
<i>Vibrio alginolyticus</i>	1	212	87	4283
<i>Vibrio campbellii</i>	8	255	258	4369
<i>Vibrio cholerae</i>	2	187	108	3404
<i>Vibrio parahaemolyticus</i>	2	217	126	4311
<i>Weissella cibaria</i>	1	39	39	1886
<i>Xanthomonas campestris</i>	2	186	85	3845
<i>Xanthomonas citri</i>	1	203	101	3893
<i>Xanthomonas oryzae</i>	0	150	50	3465
<i>Xanthomonas phaseoli</i>	3	188	105	3885
<i>Xylella fastidiosa</i>	1	96	40	2023
<i>Yersinia enterocolitica</i>	10	151	77	3727
<i>Yersinia pestis</i>	17	209	157	3619
<i>Yersinia pseudotuberculosis</i>	12	192	64	3742
<i>Yersinia ruckeri</i>	4	134	75	2986
<i>Zymomonas mobilis</i>	1	53	102	1651