

Supplementary material for
Loss of genetic redundancy in reductive genome evolution

André Mendonça, Renato J. Alves and José B. Pereira-Leal

1. Supplementary information

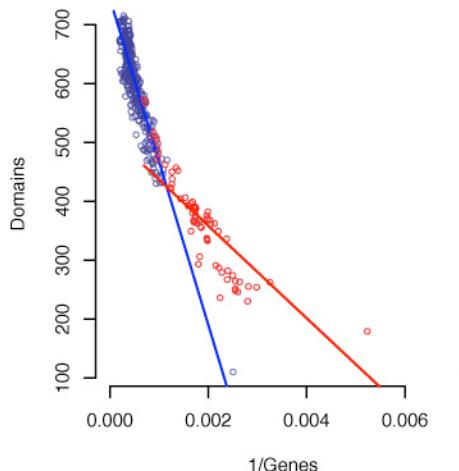


Figure S1 - Reduced genomes have a higher number of domains per gene than FL. Plotted are the domains vs. number of genes for FL (Blue) and Reduced genomes (red). Solid lines represent a fit by an inverse function. The slopes and intersects are statistically different ($p < 1 \times 10^{-66}$, $p < 3.72 \times 10^{-65}$, respectively, t test; $R^2 = 0.915$).

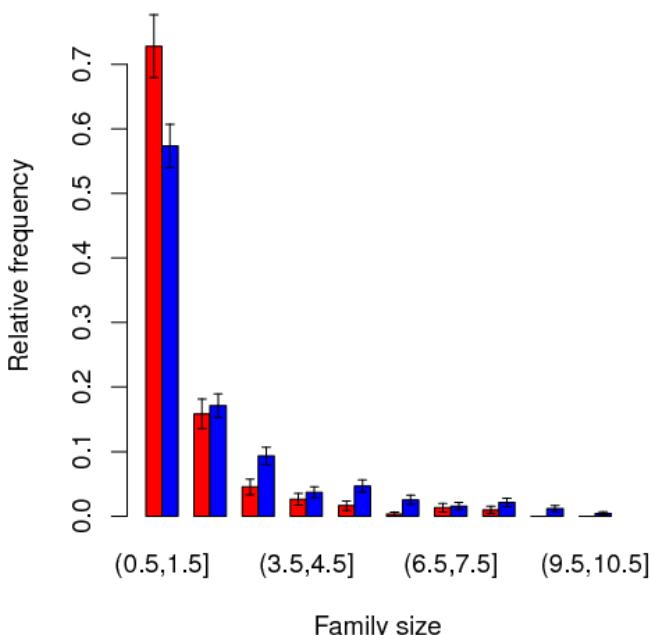


Figure S2 - Reduced Actinomycetales genomes have a higher number of singletons and smaller number of protein families with more than one element ($N_{FL}=27$; $N_R=2$; $p=1.064e-07$). This corresponds to the analysis shown on figure 1C, performed on a subset of organisms (see tables M1 and M2)

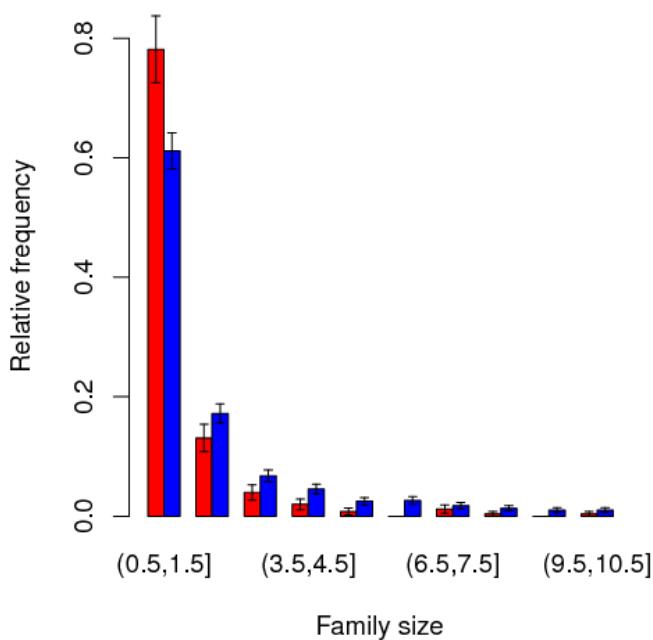


Figure S3 - Reduced Enterobacteriaceae genomes have a higher number of singletons and smaller number of protein families with more than one element ($N_{FL}=10$; $N_R=8$; $p= 3.815e^{-06}$). This corresponds to the analysis shown on figure 1C, performed on a subset of organisms (see tables M1 and M2)

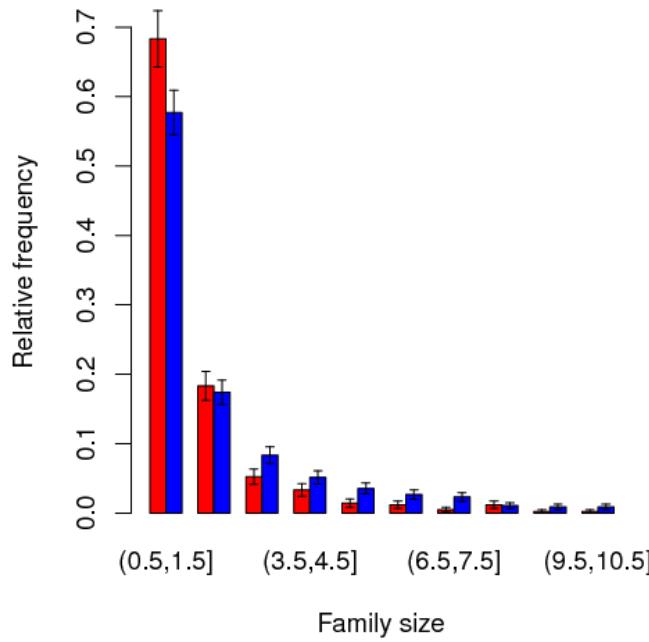


Figure S4 - Reduced Rhizobiales genomes have a higher number of singletons and smaller number of protein families with more than one element ($N_{FL}=19$; $N_R=3$; $p= 1.695e^{-07}$). This corresponds to the analysis shown on figure 1C, performed on a subset of organisms (see tables M1 and M2)

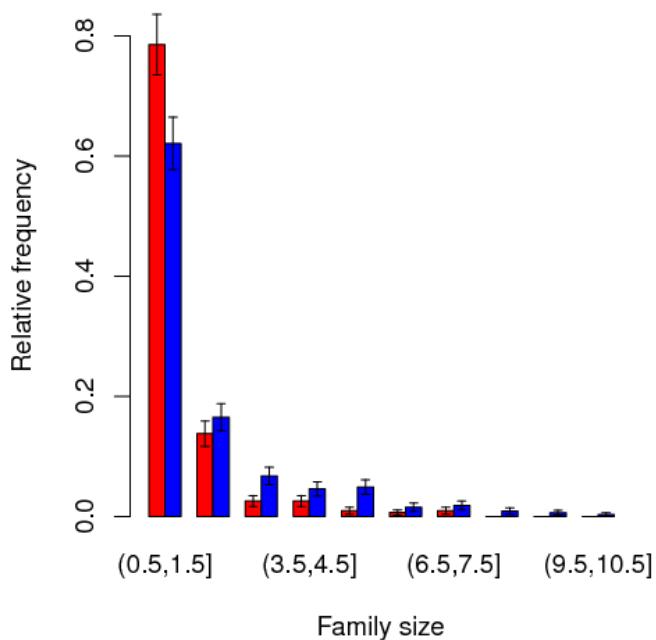


Figure S5 - Reduced Spirochaetales genomes have a higher number of singletons and smaller number of protein families with more than one element ($N_{FL}=4$; $N_R=3$; $p= 1.266e^{-05}$). This corresponds to the analysis shown on figure 1C, performed on a subset of organisms (see tables M1 and M2)

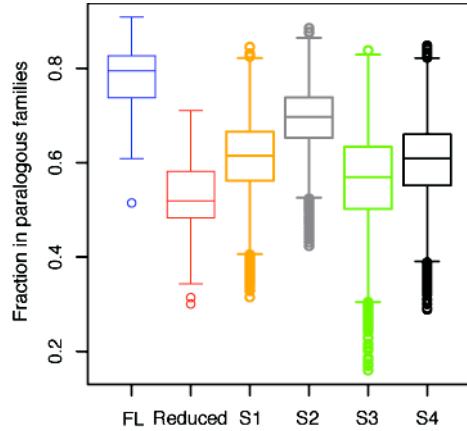


Figure S6 – Free-living organisms have significantly more genes in paralogous families, *i.e.* family size ≥ 2 , than reduced genomes (FL= 0.784 ± 0.063 ; Reduced= 0.546 ± 0.066 , $p<2.2\times 10^{-16}$; Mann-Whitney). Reduced genomes also have less genes in paralogous families than expected under any of the simulated scenarios. ($p_{S1}=1.5\times 10^{-5}$, $p_{S2}=2.2\times 10^{-6}$, $p_{S3}=0.0002$, $p_{S4}=3.0\times 10^{-13}$, Wilcoxon rank sum test)

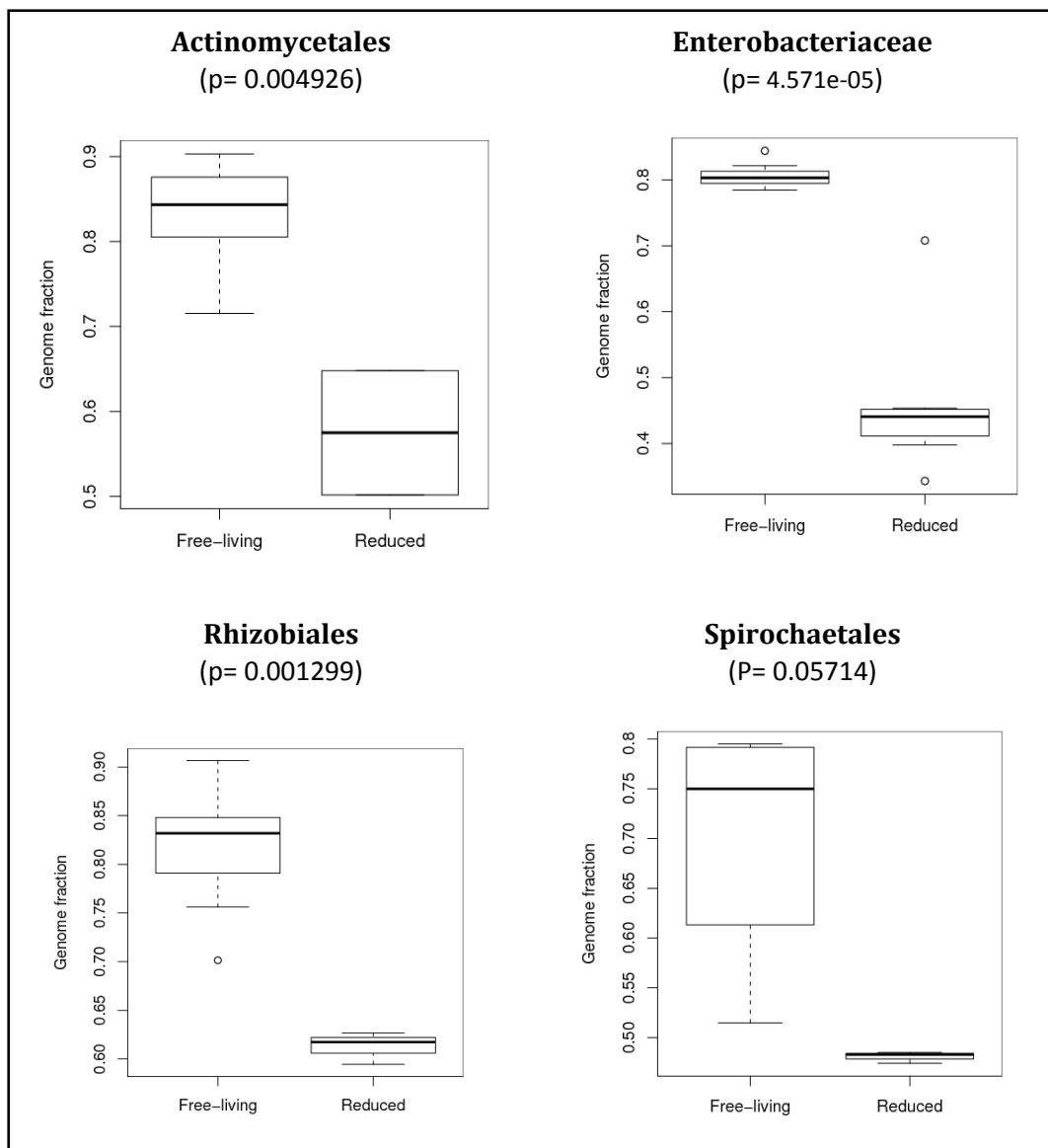


Figure S7 - Relative frequency of genes in paralogous families in FL, Reduced, by taxonomic group. P value calculated with the Wilcoxon rank sum test with continuity correction. This corresponds to the analysis shown in figure S6 for a subset of phylogenetically close organisms (see tables M1 and M2)

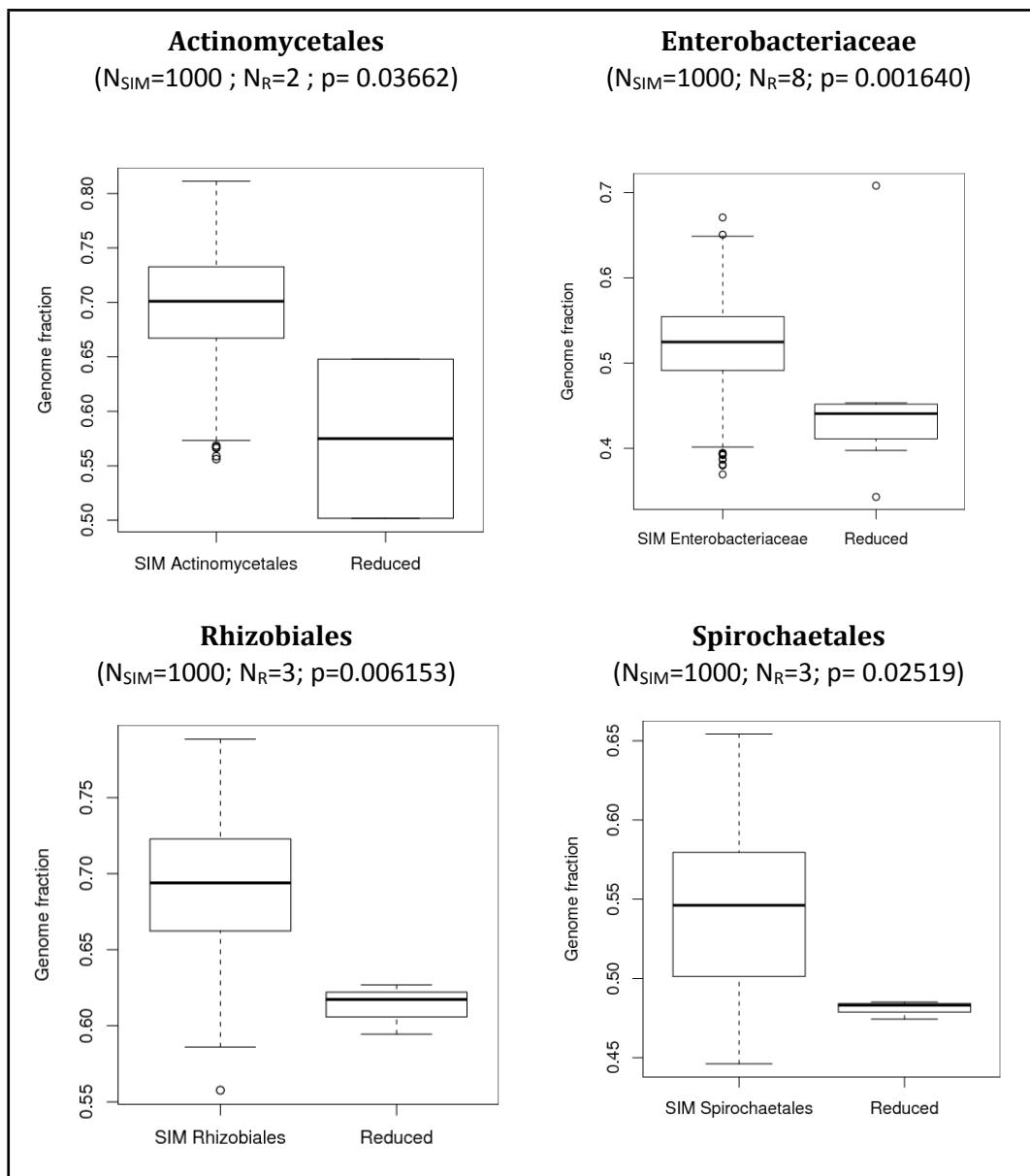


Figure S8 - Fraction of singletons in the genomes of simulated and Reduced genomes. P value calculated with the Wilcoxon rank sum test with continuity correction. N_{SIM} = number of simulated genomes; N_R =number of reduced genomes. This corresponds to the analysis shown in figure 2C for a subset of phylogenetically close organisms (see tables M1 and M2)

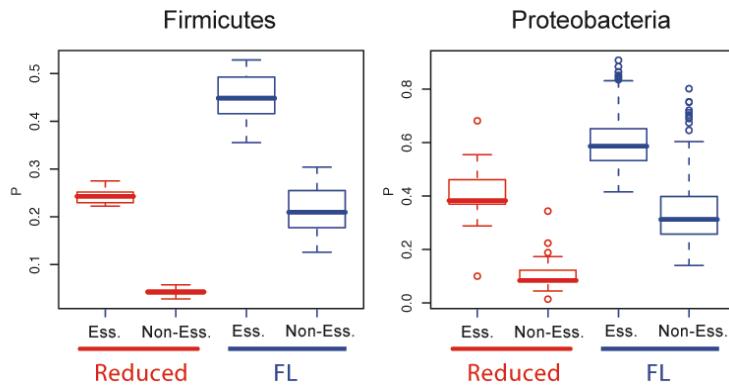


Figure S9 - Probability of conservation of genes according to their deletion phenotype in *E. coli* (Essential and Non-Essential) (Krylov et al., 2003), in free living (Blue) and reduced genome organisms (Red). Orthology was determined by BLAST bi-directional best hit at a cut off of 10^{-2} and is only show for the Phyla for which there were sufficient reduced and free living to allow statistical analysis. As expected, essential genes are more likely to be conserved in free-living organisms than non-essential genes . We observed that this is also the case in organisms that suffered extensive genome reduction, at, $p_{\text{Firmicutes}}=6.9 \times 10^{-7}$, $p_{\text{Proteobacteria}}=9.6 \times 10^{-14}$ (Mann-Whitney)

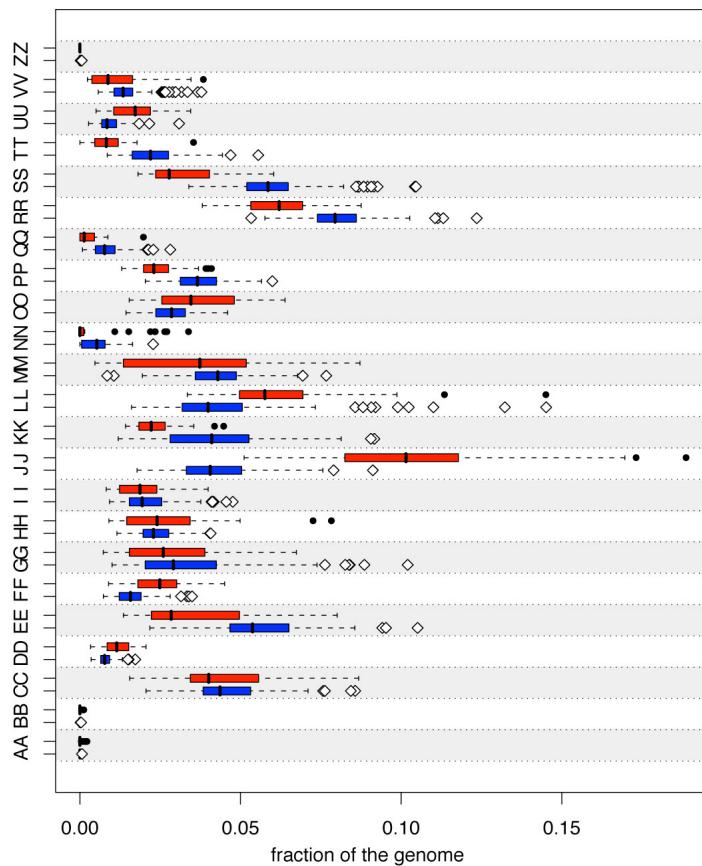


Figure S10 - COG functional class distribution in Free living (blue) and reduced (red) genomes.

Table S1 – Statistics of the COG functional class distributions in FL and Reduced. f is the average frequency of the functional class, S the standard deviation, FL and R stand for free living and reduced, respectively. Functional classes whose frequency is significantly different in FL and R are highlighted in the rejection column.

	\bar{f}^{FL}	s^{FL}	\bar{f}^R	s^R	t	p-value	Rejection	Sign
A	0.01%	0.02%	0.03%	0.07%	-2.28	0.026		
B	0.00%	0.00%	0.01%	0.04%	-2.54	0.014		
C	4.58%	1.18%	4.46%	1.70%	0.48	0.634		
D	0.81%	0.24%	1.16%	0.41%	-6.11	6.42E-08	*	+
E	5.56%	1.33%	3.67%	1.85%	6.98	1.30E-09	*	-
F	1.63%	0.53%	2.46%	0.86%	-6.81	3.65E-09	*	+
G	3.33%	1.73%	2.94%	1.66%	1.50	0.137		
H	2.44%	0.62%	2.69%	1.52%	-1.20	0.235		
I	2.12%	0.79%	1.98%	0.88%	1.01	0.315		
J	4.27%	1.32%	10.43%	3.15%	-14.02	2.06E-20	*	+
K	4.19%	1.66%	2.35%	0.68%	11.83	4.41E-25	*	-
L	4.48%	2.34%	6.30%	1.97%	-5.70	1.16E-07	*	+
M	4.28%	0.99%	3.61%	2.46%	1.96	0.055		
N	0.51%	0.42%	0.37%	0.85%	1.20	0.235		
O	2.81%	0.66%	3.67%	1.32%	-4.63	1.93E-05	*	+
P	3.69%	0.78%	2.42%	0.65%	11.88	4.49E-21	*	-
Q	0.86%	0.50%	0.24%	0.34%	10.38	1.18E-18	*	-
R	8.02%	1.05%	6.09%	1.13%	11.13	3.89E-18	*	-
S	5.96%	1.27%	3.21%	1.11%	15.30	8.05E-28	*	-
T	2.27%	0.83%	0.85%	0.58%	14.07	1.50E-27	*	-
U	0.92%	0.38%	1.71%	0.77%	-7.29	7.23E-10	*	+
V	1.44%	0.56%	1.15%	0.94%	2.17	0.034		
Z	0.00%	0.01%	0.00%	0.00%	1.62	0.106		
unk	35.83%	7.00%	38.18%	9.84%	-1.63	0.107		

Table S2 – COG functional classes

	Function
A	RNA processing and modification
B	Chromatin structure and dynamics
C	Energy production and conversion
D	Cell cycle control, mitosis and meiosis
E	Amino acid transport and metabolism
F	Nucleotide transport and metabolism
G	Carbohydrate transport and metabolism
H	Coenzyme transport and metabolism
I	Lipid transport and metabolism
J	Translation
K	Transcription
L	Replication, recombination and repair
M	Cell wall/membrane biogenesis
N	Cell motility
O	Posttranslational modification, protein turnover, chaperones
P	Inorganic ion transport and metabolism
Q	Secondary metabolites biosynthesis, transport and catabolism
R	General function prediction only
S	Function unknown
T	Signal transduction mechanisms
U	Intracellular trafficking and secretion
V	Defense mechanisms
W	Extracellular structures
Z	Cytoskeleton

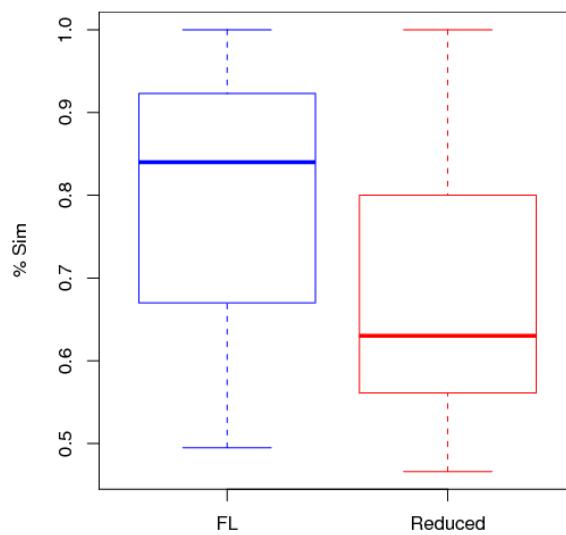


Figure S11 - The similarity between the closest paralogues in the similarity network in each protein family in families that have not reduced in size is not statistically different (p.value: 0.14, Wilcoxon)

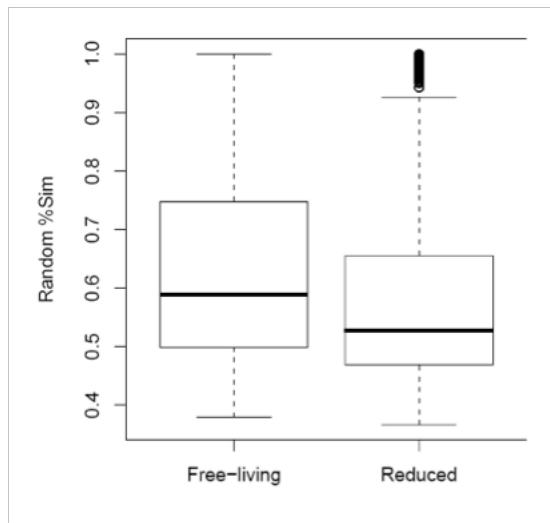


Figure S12 - Family size does not affect similarity difference. This analysis aims to address a the possibility that the difference observed in figure 5 may be due to Reduced organisms having smaller families. This is a well known problem of extreme statistics: the minimum distance between two members in a network is expected, by chance alone, to be smaller if the number of members is larger. To control for this we conducted a new experiment, which corresponds to artificially setting all family sizes to two elements only, thus controlling for family size. We randomly pick a distance from a family in Reduced, and then from the same family in Free-Living (blind to the minimum value) - we do this 10,000 times, with repetition, randomly sampling protein families. In this way we control for family size, as the numbers of elements are now two in each family group, and thus the same. However, one group (Reduced) suffered gene loss, whereas the other (FL), did not. If the observation reported in figure 5 is artifactual, i.e that we are not actively losing close paralogues, then we expect to see no difference in this scenario. However, what we observe is that Reduced families of 'size two' are significantly less similar ($p=0.018$, figure S13) than free living families of the same size. This argues that what we observe is not a statistical effect obtained by chance but in fact a loss of the closest family members in a network by Reduced organisms.

Table S3 – Reduced genomes used in this study. Lifestyle codes are P – parasite, pathogen, E – endosymbionte, X - extracellular

Species	Phylum	Lifestyle
<i>Anaplasma marginale</i>	Proteobacteria	P
<i>Anaplasma phagocytophilum</i>	Proteobacteria	P
<i>Aster yellows witches-broom phytoplasma</i>	Firmicutes	P , X(?)
<i>Bartonella bacilliformis</i>	Proteobacteria	P
<i>Bartonella henselae</i>	Proteobacteria	P
<i>Bartonella quintana</i>	Proteobacteria	P
<i>Baumannia cicadellinicola</i>	Proteobacteria	E
<i>Borrelia afzelii</i>	Spirochaetes	P
<i>Borrelia burgdorferi</i>	Spirochaetes	P, X
<i>Borrelia garinii</i>	Spirochaetes	P, X
<i>Buchnera aphidicola APS</i>	Proteobacteria	E
<i>Buchnera aphidicola Bp</i>	Proteobacteria	E
<i>Buchnera aphidicola Cc</i>	Proteobacteria	E
<i>Buchnera aphidicola Sg</i>	Proteobacteria	E
<i>Candidatus Blochmannia floridanus</i>	Proteobacteria	E
<i>Candidatus Blochmannia pennsylvanicus</i>	Proteobacteria	E

<i>Candidatus Carsonella ruddii</i>	Proteobacteria	E
<i>Candidatus Protochlamydia amoebophila</i>	Chlamydiae	P
<i>Candidatus Ruthia magnifica</i>	Proteobacteria	E
<i>Candidatus Sulcia muelleri</i>	Bacteroidetes	E
<i>Candidatus Vesicomyosocius okutanii</i>	Proteobacteria	E, X(?)
<i>Chlamydia muridarum</i>	Chlamydiae	P
<i>Chlamydia trachomatis</i>	Chlamydiae	P
<i>Chlamydophila abortus</i>	Chlamydiae	P
<i>Chlamydophila caviae</i>	Chlamydiae	P
<i>Chlamydophila felis</i>	Chlamydiae	P
<i>Chlamydophila pneumoniae</i>	Chlamydiae	P
<i>Coxiella burnetii</i>	Proteobacteria	P
<i>Ehrlichia canis</i>	Proteobacteria	P, X(?)
<i>Ehrlichia chaffeensis</i>	Proteobacteria	P
<i>Ehrlichia ruminantium</i>	Proteobacteria	P
<i>Francisella tularensis</i> ssp. <i>holarctica</i>	Proteobacteria	P
<i>Francisella tularensis</i> ssp. <i>novicida</i>	Proteobacteria	P
<i>Francisella tularensis</i> ssp. <i>tularensis</i>	Proteobacteria	P
<i>Lawsonia intracellularis</i>	Proteobacteria	P
<i>Mesoplasma florum</i>	Firmicutes	P
<i>Mycobacterium leprae</i>	Actinobacteria	P, X
<i>Mycoplasma agalactiae</i>	Firmicutes	P, X
<i>Mycoplasma capricolum</i> ssp. <i>capricolum</i>	Firmicutes	P, X
<i>Mycoplasma gallisepticum</i>	Firmicutes	P, X
<i>Mycoplasma genitalium</i>	Firmicutes	P, X
<i>Mycoplasma hyopneumoniae</i> 232	Firmicutes	P, X
<i>Mycoplasma hyopneumoniae</i> 7448	Firmicutes	P, X
<i>Mycoplasma hyopneumoniae</i> J	Firmicutes	P, X
<i>Mycoplasma mobile</i>	Firmicutes	P, X
<i>Mycoplasma mycoides</i> ssp. <i>mycoides</i>	Firmicutes	P, X
<i>Mycoplasma penetrans</i>	Firmicutes	P, X
<i>Mycoplasma pneumoniae</i>	Firmicutes	P, X
<i>Mycoplasma pulmonis</i>	Firmicutes	P, X
<i>Mycoplasma synoviae</i>	Firmicutes	P, X
<i>Neorickettsia sennetsu</i>	Proteobacteria	P
<i>Onion yellows phytoplasma</i>	Firmicutes	P
<i>Orientia tsutsugamushi</i>	Proteobacteria	P
<i>Polynucleobacter</i>	Proteobacteria	P
<i>Rickettsia akari</i>	Proteobacteria	P
<i>Rickettsia bellii</i>	Proteobacteria	P
<i>Rickettsia canadensis</i>	Proteobacteria	P
<i>Rickettsia conorii</i>	Proteobacteria	P
<i>Rickettsia felis</i>	Proteobacteria	P
<i>Rickettsia massiliae</i>	Proteobacteria	P
<i>Rickettsia prowazekii</i>	Proteobacteria	P
<i>Rickettsia rickettsii</i>	Proteobacteria	P
<i>Rickettsia typhi</i>	Proteobacteria	P
<i>Sodalis glossinidius morsitans</i>	Proteobacteria	E
<i>Tropheryma whipplei</i>	Actinobacteria	P, X
<i>Ureaplasma parvum</i>	Firmicutes	P, X
<i>Wigglesworthia glossinidia</i>	Proteobacteria	E
<i>Wolbachia</i> endosymbiont of <i>Drosophila melanogaster</i>	Proteobacteria	E
<i>Wolbachia</i> endosymbiont TRS of <i>Brugia malayi</i>	Proteobacteria	E

Table S4– free living genomes used in this study

<i>Species</i>	<i>Phylum</i>
<i>Acidobacteria bacterium</i>	Acidobacteria
<i>Acidothermus cellulolyticus</i>	Actinobacteria
<i>Acidovorax</i>	Proteobacteria
<i>Acidovorax avenae</i> ssp. <i>citrulli</i>	Proteobacteria
<i>Acinetobacter baumannii</i>	Proteobacteria
<i>Actinobacillus pleuropneumoniae</i>	Proteobacteria
<i>Actinobacillus succinogenes</i>	Proteobacteria
<i>Aeromonas hydrophila</i> ssp. <i>hydrophila</i>	Proteobacteria
<i>Aeromonas salmonicida</i> ssp. <i>salmonicida</i>	Proteobacteria
<i>Alcanivorax borkumensis</i>	Proteobacteria
<i>Alkalilimnicola ehrlichei</i>	Proteobacteria
<i>Alkaliphilus metallireducens</i>	Firmicutes
<i>Alkaliphilus oremlandii</i>	Firmicutes
<i>Anabaena variabilis</i>	Cyanobacteria
<i>Anaeromyxobacter</i>	Proteobacteria
<i>Anaeromyxobacter dehalogenans</i>	Proteobacteria
<i>Aquifex aeolicus</i>	Aquificae
<i>Arcobacter butzleri</i>	Proteobacteria
<i>Arthrobacter</i>	Actinobacteria
<i>Arthrobacter aurescens</i>	Actinobacteria
<i>Azoarcus</i>	Proteobacteria
<i>Azorhizobium caulinodans</i>	Proteobacteria
<i>Bacillus amyloliquefaciens</i>	Firmicutes
<i>Bacillus cereus</i>	Firmicutes
<i>Bacillus cereus</i> ssp. <i>cytotoxins</i>	Firmicutes
<i>Bacillus clausii</i>	Firmicutes
<i>Bacillus halodurans</i>	Firmicutes
<i>Bacillus licheniformis</i>	Firmicutes
<i>Bacillus pumilus</i>	Firmicutes
<i>Bacillus subtilis</i> ssp. <i>subtilis</i>	Firmicutes
<i>Bacillus thuringiensis</i>	Firmicutes
<i>Bacillus thuringiensis</i> ser. <i>konukian</i>	Firmicutes
<i>Bacillus weihenstephanensis</i>	Firmicutes
<i>Bacteroides fragilis</i>	Bacteroidetes
<i>Bacteroides vulgatus</i>	Bacteroidetes
<i>Bartonella tribocorum</i>	Proteobacteria
<i>Bdellovibrio bacteriovorus</i>	Proteobacteria
<i>Bifidobacterium adolescentis</i>	Actinobacteria
<i>Bordetella petrii</i>	Proteobacteria
<i>Bradyrhizobium</i>	Proteobacteria
<i>Bradyrhizobium japonicum</i>	Proteobacteria
<i>Brucella canis</i>	Proteobacteria
<i>Brucella melitensis</i>	Proteobacteria
<i>Brucella ovis</i>	Proteobacteria
<i>Brucella suis</i>	Proteobacteria
<i>Burkholderia</i>	Proteobacteria
<i>Burkholderia cenocepacia</i>	Proteobacteria
<i>Burkholderia cepacia</i>	Proteobacteria
<i>Burkholderia multivorans</i>	Proteobacteria
<i>Burkholderia thailandensis</i>	Proteobacteria
<i>Burkholderia vietnamiensis</i>	Proteobacteria
<i>Burkholderia xenovorans</i>	Proteobacteria
<i>Caldicellulosiruptor saccharolyticus</i>	Firmicutes
<i>Campylobacter concisus</i>	Proteobacteria

<i>Campylobacter curvus</i>	Proteobacteria
<i>Campylobacter fetus</i> ssp. <i>fetus</i>	Proteobacteria
<i>Campylobacter hominis</i>	Proteobacteria
<i>Campylobacter jejuni</i>	Proteobacteria
<i>Campylobacter jejuni</i> ssp. <i>doylei</i>	Proteobacteria
<i>Campylobacter jejuni</i> ssp. <i>jejuni</i>	Proteobacteria
<i>Candidatus Pelagibacter ubique</i>	Proteobacteria
<i>Carboxydothermus hydrogenoformans</i>	Firmicutes
<i>Caulobacter crescentus</i>	Proteobacteria
<i>Chlorobium chlorochromatii</i>	Chlorobi
<i>Chlorobium phaeobacteroides</i>	Chlorobi
<i>Chlorobium tepidum</i>	Chlorobi
<i>Chloroflexus aurantiacus</i>	Chloroflexi
<i>Chromobacterium violaceum</i>	Proteobacteria
<i>Chromohalobacter salexigens</i>	Proteobacteria
<i>Citrobacter koseri</i>	Proteobacteria
<i>Clavibacter michiganensis</i> ssp. <i>michiganensis</i>	Actinobacteria
<i>Clostridium acetobutylicum</i>	Firmicutes
<i>Clostridium beijerinckii</i>	Firmicutes
<i>Clostridium botulinum</i>	Firmicutes
<i>Clostridium difficile</i>	Firmicutes
<i>Clostridium kluyveri</i>	Firmicutes
<i>Clostridium novyi</i>	Firmicutes
<i>Clostridium perfringens</i>	Firmicutes
<i>Clostridium phytofermentans</i>	Firmicutes
<i>Clostridium tetani</i>	Firmicutes
<i>Clostridium thermocellum</i>	Firmicutes
<i>Colwellia psychrerythraea</i>	Proteobacteria
<i>Corynebacterium efficiens</i>	Actinobacteria
<i>Corynebacterium glutamicum</i>	Actinobacteria
<i>Corynebacterium jeikeium</i>	Actinobacteria
<i>Cytophaga hutchinsonii</i>	Bacteroidetes
<i>Dechloromonas aromatica</i>	Proteobacteria
<i>Dehalococcoides</i>	Chloroflexi
<i>Dehalococcoides ethenogenes</i>	Chloroflexi
<i>Deinococcus geothermalis</i>	Deinococcus-Thermus
<i>Deinococcus radiodurans</i>	Deinococcus-Thermus
<i>Delftia acidovorans</i>	Proteobacteria
<i>Desulfitobacterium hafniense</i>	Firmicutes
<i>Desulfococcus oleovorans</i>	Proteobacteria
<i>Desulfotalea psychrophila</i>	Proteobacteria
<i>Desulfotomaculum reducers</i>	Firmicutes
<i>Desulfovibrio desulfuricans</i>	Proteobacteria
<i>Desulfovibrio vulgaris</i> ssp. <i>vulgaris</i>	Proteobacteria
<i>Dichelobacter nodosus</i>	Proteobacteria
<i>Dinoroseobacter shibae</i>	Proteobacteria
<i>Enterobacter</i>	Proteobacteria
<i>Enterobacter sakazakii</i>	Proteobacteria
<i>Erwinia carotovora</i> ssp. <i>atroseptica</i>	Proteobacteria
<i>Erythrobacter litoralis</i>	Proteobacteria
<i>Fervidobacterium nodosum</i>	Thermotogae
<i>Flavobacterium johnsoniae</i>	Bacteroidetes
<i>Flavobacterium psychrophilum</i>	Bacteroidetes
<i>Frankia</i>	Actinobacteria
<i>Frankia alni</i>	Actinobacteria
<i>Fusobacterium nucleatum</i> ssp. <i>nucleatum</i>	Fusobacteria
<i>Geobacillus kaustophilus</i>	Firmicutes
<i>Geobacillus thermodenitrificans</i>	Firmicutes

<i>Geobacter metallireducens</i>	Proteobacteria
<i>Geobacter sulfurreducens</i>	Proteobacteria
<i>Geobacter uraniumreducens</i>	Proteobacteria
<i>Gloeobacter violaceus</i>	Cyanobacteria
<i>Gluconacetobacter diazotrophicus</i>	Proteobacteria
<i>Gluconobacter oxydans</i>	Proteobacteria
<i>Gramella forsetii</i>	Bacteroidetes
<i>Granulibacter bethesdensis</i>	Proteobacteria
<i>Haemophilus influenzae</i>	Proteobacteria
<i>Haemophilus somnus</i>	Proteobacteria
<i>Hahella chejuensis</i>	Proteobacteria
<i>Halorhodospira halophila</i>	Proteobacteria
<i>Herminiimonas arsenicoxydans</i>	Proteobacteria
<i>Herpetosiphon aurantiacus</i>	Chloroflexi
<i>Hyphomonas neptunium</i>	Proteobacteria
<i>Idiomarina loihiensis</i>	Proteobacteria
<i>Jannaschia</i>	Proteobacteria
<i>Janthinobacterium</i>	Proteobacteria
<i>Kineococcus radiotolerans</i>	Actinobacteria
<i>Klebsiella pneumoniae</i> ssp. <i>< i>pneumoniae</i>	Proteobacteria
<i>Lactobacillus acidophilus</i>	Firmicutes
<i>Lactobacillus brevis</i>	Firmicutes
<i>Lactobacillus casei</i>	Firmicutes
<i>Lactobacillus delbrueckii</i> ssp. <i>< i>bulgaricus</i>	Firmicutes
<i>Lactobacillus gasseri</i>	Firmicutes
<i>Lactobacillus helveticus</i>	Firmicutes
<i>Lactobacillus plantarum</i>	Firmicutes
<i>Lactobacillus reuteri</i>	Firmicutes
<i>Lactobacillus sakei</i> ssp. <i>< i>sakei</i>	Firmicutes
<i>Lactobacillus salivarius</i> ssp. <i>< i>salivarius</i>	Firmicutes
<i>Lactococcus lactis</i> ssp. <i>< i>cremoris</i>	Firmicutes
<i>Lactococcus lactis</i> ssp. <i>< i>lactis</i>	Firmicutes
<i>Legionella pneumophila</i>	Proteobacteria
<i>Legionella pneumophila</i> ssp. <i>< i>pneumophila</i>	Proteobacteria
<i>Leifsonia xyli</i> ssp. <i>< i>xyli</i>	Actinobacteria
<i>Leptospira borgpetersenii</i>	Spirochaetes
<i>Leptospira interrogans</i>	Spirochaetes
<i>Leuconostoc mesenteroides</i> ssp. <i>< i>mesenteroides</i>	Firmicutes
<i>Listeria innocua</i>	Firmicutes
<i>Listeria monocytogenes</i>	Firmicutes
<i>Listeria welshimeri</i>	Firmicutes
<i>Magnetospirillum magneticum</i>	Proteobacteria
<i>Maricaulis maris</i>	Proteobacteria
<i>Marinobacter aquaeolei</i>	Proteobacteria
<i>Marinomonas</i>	Proteobacteria
<i>Mesorhizobium</i>	Proteobacteria
<i>Mesorhizobium loti</i>	Proteobacteria
<i>Methylibium petroleiphilum</i>	Proteobacteria
<i>Methylobacillus flagellatus</i>	Proteobacteria
<i>Methylobacterium extorquens</i>	Proteobacteria
<i>Methylococcus capsulatus</i>	Proteobacteria
<i>Moorella thermoacetica</i>	Firmicutes
<i>Mycobacterium avium</i>	Actinobacteria
<i>Mycobacterium gilvum</i>	Actinobacteria
<i>Mycobacterium smegmatis</i>	Actinobacteria
<i>Mycobacterium ulcerans</i>	Actinobacteria
<i>Mycobacterium vanbaalenii</i>	Actinobacteria

<i>Myxococcus xanthus</i>	Proteobacteria
<i>Neisseria gonorrhoeae</i>	Proteobacteria
<i>Nitratiruptor</i>	Proteobacteria
<i>Nitrobacter hamburgensis</i>	Proteobacteria
<i>Nitrobacter winogradskyi</i>	Proteobacteria
<i>Nitrosoccus oceanii</i>	Proteobacteria
<i>Nitrosomonas europaea</i>	Proteobacteria
<i>Nitrosomonas eutropha</i>	Proteobacteria
<i>Nitrosospira multiformis</i>	Proteobacteria
<i>Nocardia farcinica</i>	Actinobacteria
<i>Nocardiooides</i>	Actinobacteria
<i>Nostoc</i>	Cyanobacteria
<i>Novosphingobium aromaticivorans</i>	Proteobacteria
<i>Oceanobacillus iheyensis</i>	Firmicutes
<i>Ochrobactrum anthropi</i>	Proteobacteria
<i>Oenococcus oeni</i>	Firmicutes
<i>Parabacteroides distasonis</i>	Bacteroidetes
<i>Paracoccus denitrificans</i>	Proteobacteria
<i>Parvibaculum lavamentivorans</i>	Proteobacteria
<i>Pasteurella multocida</i> ssp. <i>< i>multocida</i>	Proteobacteria
<i>Pediococcus pentosaceus</i>	Firmicutes
<i>Pelobacter carbinolicus</i>	Proteobacteria
<i>Pelobacter propionicus</i>	Proteobacteria
<i>Pelodictyon luteolum</i>	Chlorobi
<i>Pelotomaculum thermopropionicum</i>	Firmicutes
<i>Petrotoga mobilis</i>	Thermotogae
<i>Photobacterium profundum</i>	Proteobacteria
<i>Photorhabdus luminescens</i> ssp. <i>< i>laumontii</i>	Proteobacteria
<i>Polaromonas</i>	Proteobacteria
<i>Polaromonas naphthalenivorans</i>	Proteobacteria
<i>Prochlorococcus marinus</i>	Cyanobacteria
<i>Prochlorococcus marinus</i> ssp. <i>< i>marinus</i>	Cyanobacteria
<i>Prochlorococcus marinus</i> ssp. <i>< i>pastoris</i>	Cyanobacteria
<i>Prosthecochloris vibrioformis</i>	Chlorobi
<i>Pseudoalteromonas atlantica</i>	Proteobacteria
<i>Pseudoalteromonas haloplanktis</i>	Proteobacteria
<i>Pseudomonas aeruginosa</i>	Proteobacteria
<i>Pseudomonas entomophila</i>	Proteobacteria
<i>Pseudomonas fluorescens</i>	Proteobacteria
<i>Pseudomonas mendocina</i>	Proteobacteria
<i>Pseudomonas putida</i>	Proteobacteria
<i>Pseudomonas stutzeri</i>	Proteobacteria
<i>Pseudomonas syringae</i> pv. <i>< i>phaseolicola</i>	Proteobacteria
<i>Pseudomonas syringae</i> pv. <i>< i>syringae</i>	Proteobacteria
<i>Pseudomonas syringae</i> pv. <i>< i>tomato</i>	Proteobacteria
<i>Psychrobacter</i>	Proteobacteria
<i>Psychrobacter arcticus</i>	Proteobacteria
<i>Psychrobacter cryohalolentis</i>	Proteobacteria
<i>Psychromonas ingrahamii</i>	Proteobacteria
<i>Ralstonia eutropha</i>	Proteobacteria
<i>Ralstonia metallidurans</i>	Proteobacteria
<i>Renibacterium salmoninarum</i>	Actinobacteria
<i>Rhizobium etli</i>	Proteobacteria
<i>Rhizobium leguminosarum</i> bv. <i>< i>viciae</i>	Proteobacteria
<i>Rhodobacter sphaeroides</i>	Proteobacteria
<i>Rhodococcus</i>	Actinobacteria
<i>Rhodoferax ferrireducens</i>	Proteobacteria
<i>Rhodopirellula baltica</i>	Planctomycetes

<i>Rhodopseudomonas palustris</i>	Proteobacteria
<i>Rhodospirillum rubrum</i>	Proteobacteria
<i>Roseiflexus</i>	Chloroflexi
<i>Roseiflexus castenholzii</i>	Chloroflexi
<i>Roseobacter denitrificans</i>	Proteobacteria
<i>Rubrobacter xylanophilus</i>	Actinobacteria
<i>Saccharophagus degradans</i>	Proteobacteria
<i>Saccharopolyspora erythraea</i>	Actinobacteria
<i>Salinibacter ruber</i>	Bacteroidetes
<i>Salinispora arenicola</i>	Actinobacteria
<i>Salinispora tropica</i>	Actinobacteria
<i>Salmonella enterica</i> ssp. <i>arizona</i>	Proteobacteria
<i>Salmonella enterica</i> ssp. <i>enterica</i>	Proteobacteria
<i>Serratia proteamaculans</i>	Proteobacteria
<i>Shewanella</i>	Proteobacteria
<i>Shewanella amazonensis</i>	Proteobacteria
<i>Shewanella baltica</i>	Proteobacteria
<i>Shewanella denitrificans</i>	Proteobacteria
<i>Shewanella frigidimarina</i>	Proteobacteria
<i>Shewanella loihica</i>	Proteobacteria
<i>Shewanella oneidensis</i>	Proteobacteria
<i>Shewanella pealeana</i>	Proteobacteria
<i>Shewanella putrefaciens</i>	Proteobacteria
<i>Shewanella sediminis</i>	Proteobacteria
<i>Shigella boydii</i>	Proteobacteria
<i>Shigella dysenteriae</i>	Proteobacteria
<i>Shigella sonnei</i>	Proteobacteria
<i>Silicibacter</i>	Proteobacteria
<i>Silicibacter pomeroyi</i>	Proteobacteria
<i>Sinorhizobium medicae</i>	Proteobacteria
<i>Sinorhizobium meliloti</i>	Proteobacteria
<i>Solibacter usitatus</i>	Acidobacteria
<i>Sorangium cellulosum</i> So	Proteobacteria
<i>Sphingomonas wittichii</i>	Proteobacteria
<i>Sphingopyxis alaskensis</i>	Proteobacteria
<i>Staphylococcus aureus</i>	Firmicutes
<i>Staphylococcus aureus</i> ssp. <i>aureus</i>	Firmicutes
<i>Staphylococcus haemolyticus</i>	Firmicutes
<i>Staphylococcus saprophyticus</i> ssp. <i>saprophyticus</i>	Firmicutes
<i>Streptococcus gordoni</i>	Firmicutes
<i>Streptococcus sanguinis</i>	Firmicutes
<i>Streptococcus suis</i>	Firmicutes
<i>Streptococcus thermophilus</i>	Firmicutes
<i>Streptomyces avermitilis</i>	Actinobacteria
<i>Streptomyces coelicolor</i>	Actinobacteria
<i>Sulfurovum</i>	Proteobacteria
<i>Symbiobacterium thermophilum</i>	Actinobacteria
<i>Synechococcus</i>	Cyanobacteria
<i>Synechococcus elongatus</i>	Cyanobacteria
<i>Synechocystis</i>	Cyanobacteria
<i>Syntrophobacter fumaroxidans</i>	Proteobacteria
<i>Syntrophomonas wolfei</i> ssp. <i>wolfei</i>	Firmicutes
<i>Syntrophus aciditrophicus</i>	Proteobacteria
<i>Thermoanaerobacter tengcongensis</i>	Firmicutes
<i>Thermobifida fusca</i>	Actinobacteria
<i>Thermosiphon melanesiensis</i>	Thermotogae
<i>Thermosynechococcus elongatus</i>	Cyanobacteria
<i>Thermotoga lettingae</i>	Thermotogae

<i>Thermotoga maritima</i>	Thermotogae
<i>Thermotoga petrophila</i>	Thermotogae
<i>Thermus thermophilus</i>	Deinococcus-Thermus
<i>Thiobacillus denitrificans</i>	Proteobacteria
<i>Thiomicrospira crunogena</i>	Proteobacteria
<i>Thiomicrospira denitrificans</i>	Proteobacteria
<i>Treponema denticola</i>	Spirochaetes
<i>Treponema pallidum</i> ssp. <i>pallidum</i>	Spirochaetes
<i>Trichodesmium erythraeum</i>	Cyanobacteria
<i>Verminephrobacter eiseniae</i>	Proteobacteria
<i>Vibrio cholerae</i>	Proteobacteria
<i>Vibrio cholerae</i> O1 biovar. <i>eltor</i>	Proteobacteria
<i>Vibrio fischeri</i>	Proteobacteria
<i>Vibrio harveyi</i>	Proteobacteria
<i>Xanthobacter autotrophicus</i>	Proteobacteria
<i>Xanthomonas axonopodis</i> pv. <i>citri</i>	Proteobacteria
<i>Xanthomonas campestris</i> pv. <i>campestris</i>	Proteobacteria
<i>Xanthomonas campestris</i> pv. <i>vesicatoria</i>	Proteobacteria
<i>Xanthomonas oryzae</i> pv. <i>oryzae</i>	Proteobacteria
<i>Yersinia enterocolitica</i> ssp. <i>enterocolitica</i>	Proteobacteria
<i>Yersinia pseudotuberculosis</i>	Proteobacteria
<i>Zymomonas mobilis</i> ssp. <i>mobilis</i>	Proteobacteria

2. References

Krylov, D. M., Wolf, Y. I., Rogozin, I. B., and Koonin, E. V. (2003). Gene loss, protein sequence divergence, gene dispensability, expression level, and interactivity are correlated in eukaryotic evolution. *Genome research* 13, 2229-2235.