

**SUPPLEMENTAL TABLE 1. 60 Genome Comparative Set**

Phylum	Class	Order	Species			
Actinobacteria	Actinobacteridae	Actinomycetales	<i>Mycobacterium tuberculosis</i> CDC1551			
			<i>Corynebacterium glutamicum</i> ATCC 13032			
			<i>Streptomyces coelicolor</i> A3(2)			
Aquificae	Aquificales	Aquificaceae	<i>Aquifex aeolicus</i> VF5			
Bacteroidetes	Bacteroides (class)	Bacteroidales	<i>Porphyromonas gingivalis</i> W83			
Chlamydiae	Chlamydiales	Chlamydiaceae	<i>Chlamydia pneumoniae</i> AR39			
		Parachlamydiaceae	<i>Parachlamydia</i> sp UWE25			
Chlorobi	Chlorobia	Chlorobiales	<i>Chlorobium tepidum</i> TLS			
Chloroflexi	Dehalococcoides	Dehalococcoides ethenogenes	<i>Dehalococcoides ethenogenes</i> 195			
Cyanobacteria	Chroococcales	Synechocystis	<i>Synechocystis</i> sp CC6803			
	Prochlorophytes	Prochlorococcaceae	<i>Prochlorococcus marinus</i> MIT9313			
Deinococcus- Thermus	Deinococci	Deinococcales	<i>Deinococcus radiodurans</i> R1			
Firmicutes	Bacillales	Bacillaceae	<i>Bacillus subtilis</i> 168			
			<i>Bacillus anthracis</i> Ames			
			<i>Bacillus cereus</i> ATCC14579			
			Listeriaceae	<i>Listeria monocytogenes</i> 4b F2365		
			Staphylococcus	<i>Staphylococcus aureus</i> MW2		
		Clostridia	Clostridiales	<i>Clostridium perfringens</i> 13		
				<i>Clostridium tetani</i> E88		
		Lactobacillales		Enterococcaceae	<i>Enterococcus faecalis</i> V583	
				Lactobacillaceae	<i>Lactobacillus johnsonii</i> NCC 533	
				Streptococcaceae	<i>Streptococcus pneumoniae</i> TIGR4	
					<i>Streptococcus pyogenes</i> SF370 serotype M1	
					<i>Phytoplasma asteris</i> Onion Yellows strain	
		Mollicutes		Acholeplasmatales	<i>Mycoplasma genitalium</i> G-37	
				Mycoplasmataceae	<i>Mycoplasma mycoides</i> SC G1	
		Fusobacteria	Fusobacteriales	Fusobacteriaceae	<i>Fusobacterium nucleatum</i> ATCC 25586	
		Proteobacteria	Alphaproteobacteria	Caulobacterales	<i>Caulobacter crescentus</i> CB15	
Rhizobiales	<i>Brucella suis</i> 1330					
	<i>Agrobacterium tumefaciens</i> C58 UWash					
				Rickettsiales	<i>Rickettsia prowazekii</i> Madrid E	
					<i>Wolbachia pipientis</i> wMel	
				Betaproteobacteria	Burkholderiales	<i>Bordetella pertussis</i> Tohama I
					Neisseriales	<i>Neisseria meningitidis</i> MC58
				Nitrosomonadales	<i>Nitrosomonas europaea</i> ATCC 19718	
				Gammaproteobacteria	Alteromonadaceae	<i>Shewanella oneidensis</i> MR-1
				Enterobacteriales	<i>Buchnera aphidicola</i> ( <i>Baizongia pistaciae</i> )	
					<i>Escherichia coli</i> O157:H7 EDL933	
					<i>Escherichia coli</i> O157:H7 VT2-Sakai	
					<i>Escherichia coli</i> CFT073	
					<i>Salmonella typhimurium</i> LT2 SGSC1412	
					<i>Salmonella enterica</i> serovar Typhi Ty2	
					<i>Shigella flexneri</i> 2a 2457T	
					<i>Yersinia pestis</i> KIM	
			Legionellaceae group		<i>Coxiella burnetii</i> RSA 493	
			Pasteurellaceae		<i>Pasteurella multocida</i> M70	
	Pasteurellales	<i>Haemophilus influenzae</i> KW20 Rd				
	Pseudomonadales	<i>Pseudomonas aeruginosa</i> AO1				

			<i>Pseudomonas putida</i> KT2440
		Vibrionales	<i>Vibrio cholerae</i> El Tor N16961
		Xanthomonadales	<i>Xylella fastidiosa</i> 9a5c
		Xanthomonas group	<i>Xanthomonas campestris</i> pv <i>campestris</i> ATCC33913
		Magnetotactic cocci	<i>Magnetococcus</i> sp MC-1
	Deltaproteobacteria	Desulfovibrionaceae	<i>Desulfovibrio vulgaris</i> Hildenborough
	Epsilonproteobacteria	Campylobacterales	<i>Campylobacter jejuni</i> NCTC 11168
			<i>Helicobacter pylori</i> 26695
Spirochaetes	Spirochaetales	Leptospiraceae	<i>Leptospira interrogans</i> serovar lai str 56601
		Spirochaetaceae	<i>Treponema pallidum</i> Nichols
			<i>Borrelia burgdorferi</i> B31
Thermotogae	Thermotogales	Thermotogaceae	<i>Thermotoga maritima</i> MSB8

SUPPLEMENTAL TABLE 2. 49 Genome (Proteobacteria) Comparative Set

Class	Order	Species	
Alphaproteobacteria	Caulobacterales	<i>Caulobacter crescentus</i> CB15	
	Rhizobiales	<i>Rhodopseudomonas palustris</i> CGA009	
		<i>Brucella suis</i> 1330	
		<i>Sinorhizobium meliloti</i> 1021	
		<i>Agrobacterium tumefaciens</i> C58 UWash	
		Rickettsiales	<i>Rickettsia prowazekii</i> Madrid E
		<i>Rickettsia conorii</i> Malish 7	
Betaproteobacteria	Burkholderiales	<i>Wolbachia pipientis</i> wMel	
		<i>Bordetella bronchiseptica</i> RB50	
		<i>Bordetella pertussis</i> Tohama I	
		<i>Bordetella parapertussis</i> 12822	
		<i>Burkholderia mallei</i> ATCC:23344	
	Neisseriales	<i>Neisseria meningitidis</i> MC58	
		<i>Chromobacterium violaceum</i> ATCC12472	
		Nitrosomonadales	<i>Nitrosomonas europaea</i> ATCC 19718
		Ralstoniaceae	<i>Ralstonia solanacearum</i> GMI1000
		Gammaproteobacteria	Alteromonadaceae
Enterobacteriales	<i>Buchnera aphidicola</i> ( <i>Baizongia pistaciae</i> )		
	<i>Escherichia coli</i> O157:H7 EDL933		
	<i>Escherichia coli</i> O157:H7 VT2-Sakai		
	<i>Escherichia coli</i> CFT073		
	<i>Photorhabdus luminescens</i> TTO1		
	<i>Salmonella typhimurium</i> LT2 SGSC1412		
	<i>Salmonella enterica</i> serovar Typhi Ty2		
	<i>Shigella flexneri</i> 2a 2457T		
	<i>Yersinia pestis</i> KIM		
	<i>Blochmannia floridanus</i>		
	Legionellaceae		<i>Coxiella burnetii</i> RSA 493
	Methylococcales		<i>Methylococcus capsulatus</i> Bath
	Pasteurellaceae		<i>Pasteurella multocida</i> M70
	Pasteurellales		<i>Haemophilus ducreyi</i> 35000H
			<i>Haemophilus influenzae</i> KW20 Rd
	Pseudomonadales		<i>Pseudomonas aeruginosa</i> AO1
			<i>Pseudomonas putida</i> KT2440
			<i>Pseudomonas syringae</i> DC3000
	Vibrionales		<i>Vibrio cholerae</i> El Tor N16961
			<i>Vibrio parahaemolyticus</i> RIMD 2210633
			<i>Vibrio vulnificus</i> YJ016
	Xanthomonadales		<i>Xylella fastidiosa</i> 9a5c
	Xanthomonas group		<i>Xanthomonas campestris</i> pv <i>campestris</i> ATCC33913
			<i>Xanthomonas axonopodis</i> pv <i>citri</i> 306
	Magnetotactic cocci		<i>Magnetococcus</i> sp MC-1
	Deltaproteobacteria		Bdellovibrionales
Desulfovibrionaceae		<i>Desulfovibrio vulgaris</i> Hildenborough	
Desulfuromonas group		<i>Geobacter sulfurreducens</i> CA	
Epsilonproteobacteria	Campylobacterales	<i>Campylobacter jejuni</i> NCTC 11168	
		<i>Helicobacter hepaticus</i> ATCC 51449	
		<i>Helicobacter pylori</i> 26695	
		<i>Wolinella succinogenes</i> DSMZ 1740	

**SUPPLEMENTAL TABLE 3.** Area under ROC Curve Measurements for Predictions shown in Supplemental Figure 1.

Panel	Prediction set	Area under ROC curve
A	35 Genomes, Combined information	0.894
	60 Genomes, Combined information	0.812
	49 Proteobacteria, Combined information	0.821
B	35 Genomes, Combined information	0.894
	22 Proteobacteria, Combined information	0.915
	13 Genomes, Combined information	0.915
C	22 Proteobacteria, Combined information, Genomes with <1500 total orthologs removed (19 total genomes used)	0.909
	22 Proteobacteria, Combined information, Genomes with <2000 total orthologs removed (16 total genomes used)	0.910
	22 Proteobacteria, Combined information, Genomes with <2500 total orthologs removed (12 total genomes used)	0.912
D	22 Proteobacteria, Combined information, Genomes limited to Pearson's correlation of <0.9 (21 total genomes used)	0.913
	22 Proteobacteria, Combined information, Genomes limited to Pearson's correlation of <0.8 (20 total genomes used)	0.907
	22 Proteobacteria, Combined information, Genomes limited to Pearson's correlation of <0.7 (17 total genomes used)	0.912
E	35 Genomes, Combined information, BLASTP cutoff 1e-04	0.894
	35 Genomes, Combined information, BLASTP cutoff 1e-08	0.895