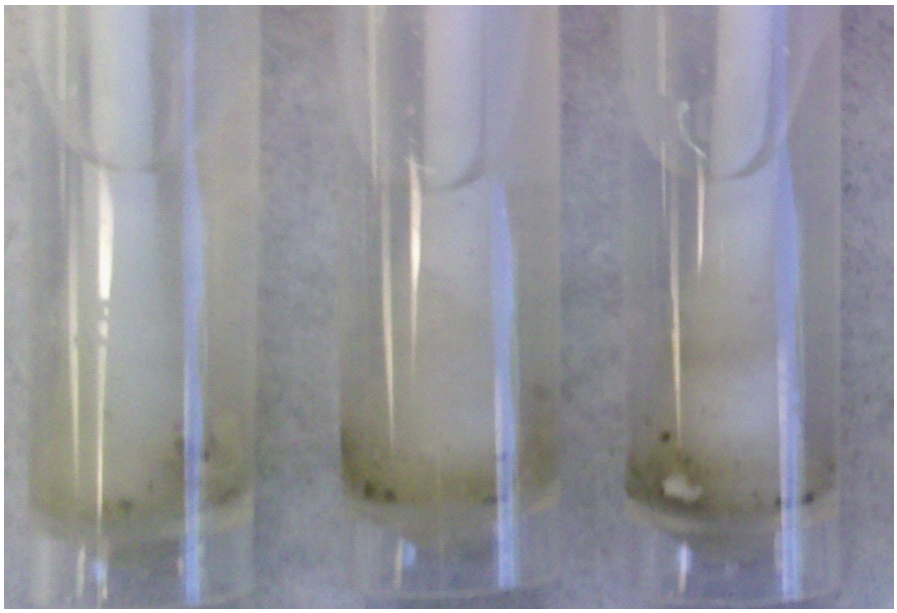
Supporting Information

McLean et al. "**Candidate Phylum TM6 Genome Recovered from a Hospital Sink Biofilm Provides the First Genomic Insights into this Uncultivated Phylum "**

SI Appendix

a



b

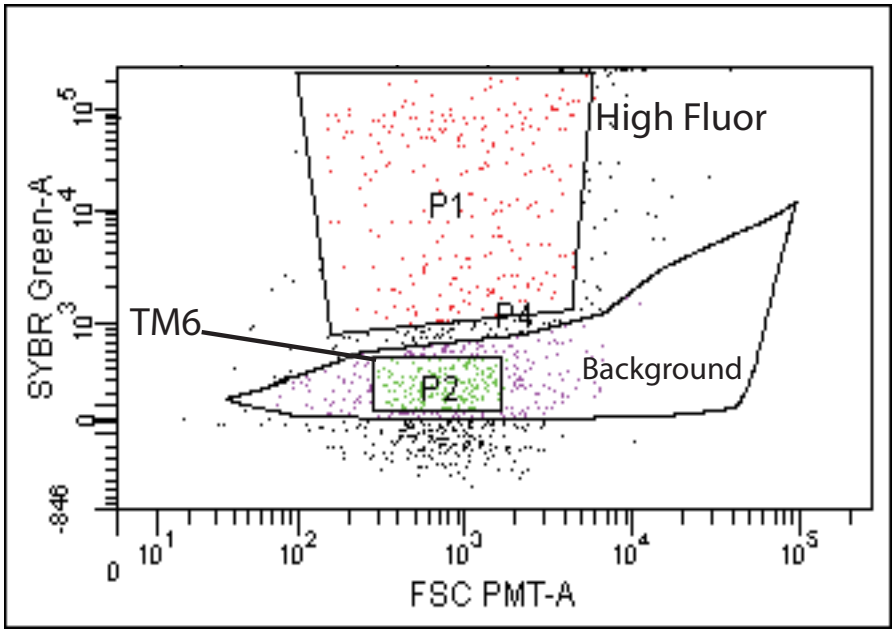


Fig. S1. Biofilm sample biomass and Fluorescence Activated Cell Sorting (FACS) plot of a biofilm sample. a) Sample biomass collected directly into buffer solution and from a biofilm in a sink drain within a restroom adjacent to an emergency waiting room. b) Sorting gates set to sort events after staining with SYBR Green DNA stain. The P1 gate includes high fluorescent SYBR Green stained particles, and the background gate indicates that region in which unstained sample events were located. The low fluorescent P2 region was chosen as a sort gate to target a total of 100 events in each of 32 wells of a 384 well plate.

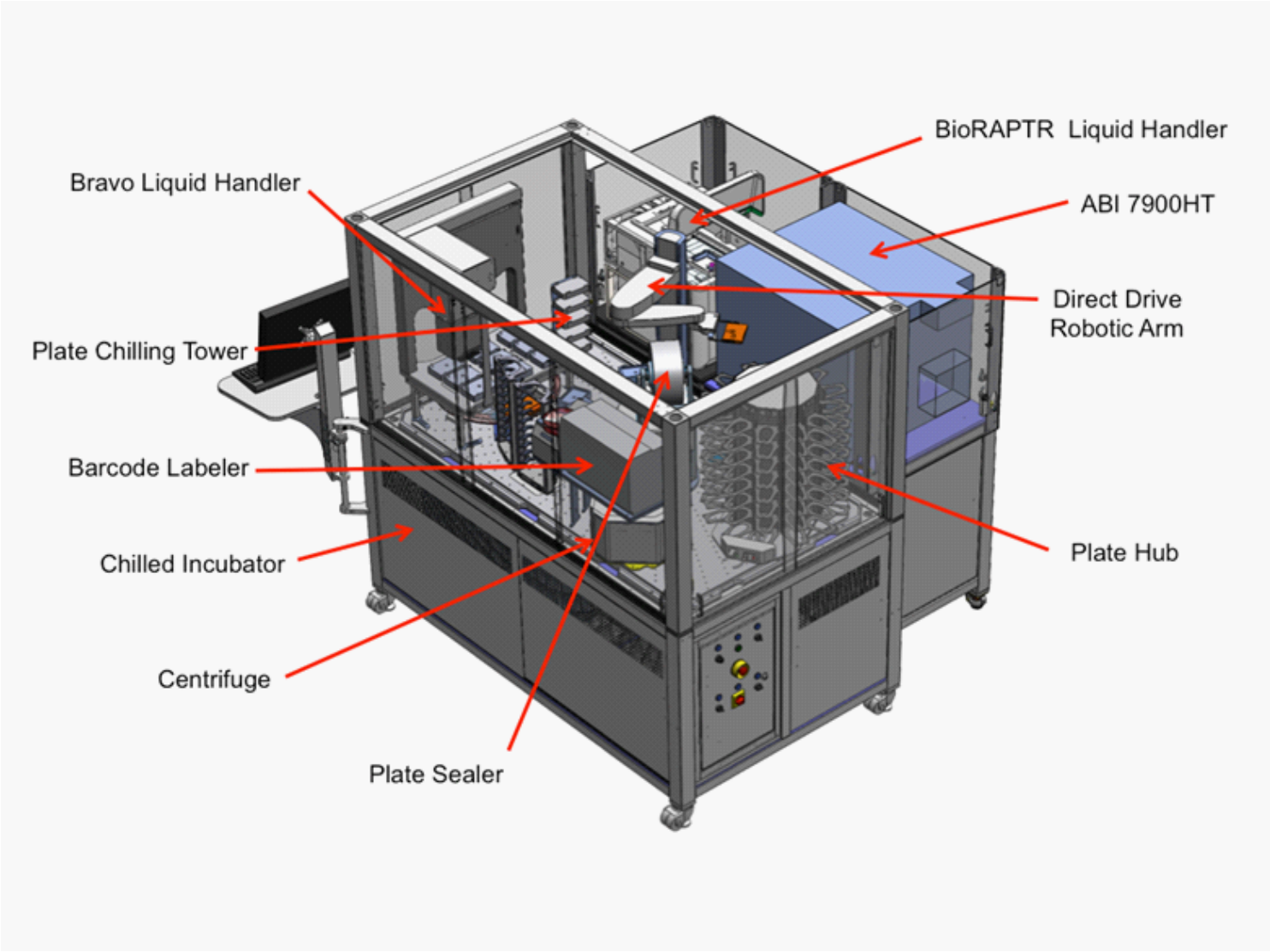


Fig. S2. Custom integrated Agilent Technologies BioCel 1200 liquid handling automated platform for high throughput single cell genomics. The BioCel platform allows processing of more than 5,000 single cells per week through a multi-stage protocol that includes multiple displacement amplification (MDA) of DNA, MDA dilution and 16S PCR, MDA and PCR hit picking, Picogreen (Life Technologies) DNA quantitation, 16S Syto 9 (Life Technologies) melt curve assay, 16S Taqman qPCR, and SAP/Exonuclease I (Affymetrix) PCR treatment. All liquid handling is performed on the BioCel with the BioRAPTR (Beckman Coulter) and Bravo (Agilent) performing non-contact dispensing and liquid transfer steps, respectively. The MDA isothermal reaction and PCR are performed offline on GeneAmp PCR system 9700 thermocyclers (Applied Biosystems), while TaqMan or melt curve analysis are performed in-line on the ABI 7900HT (Applied Biosystems). The platform includes barcode tracking of 384-well plates, and is integrated with a JCVI Laboratory Information Management System (LIMS).

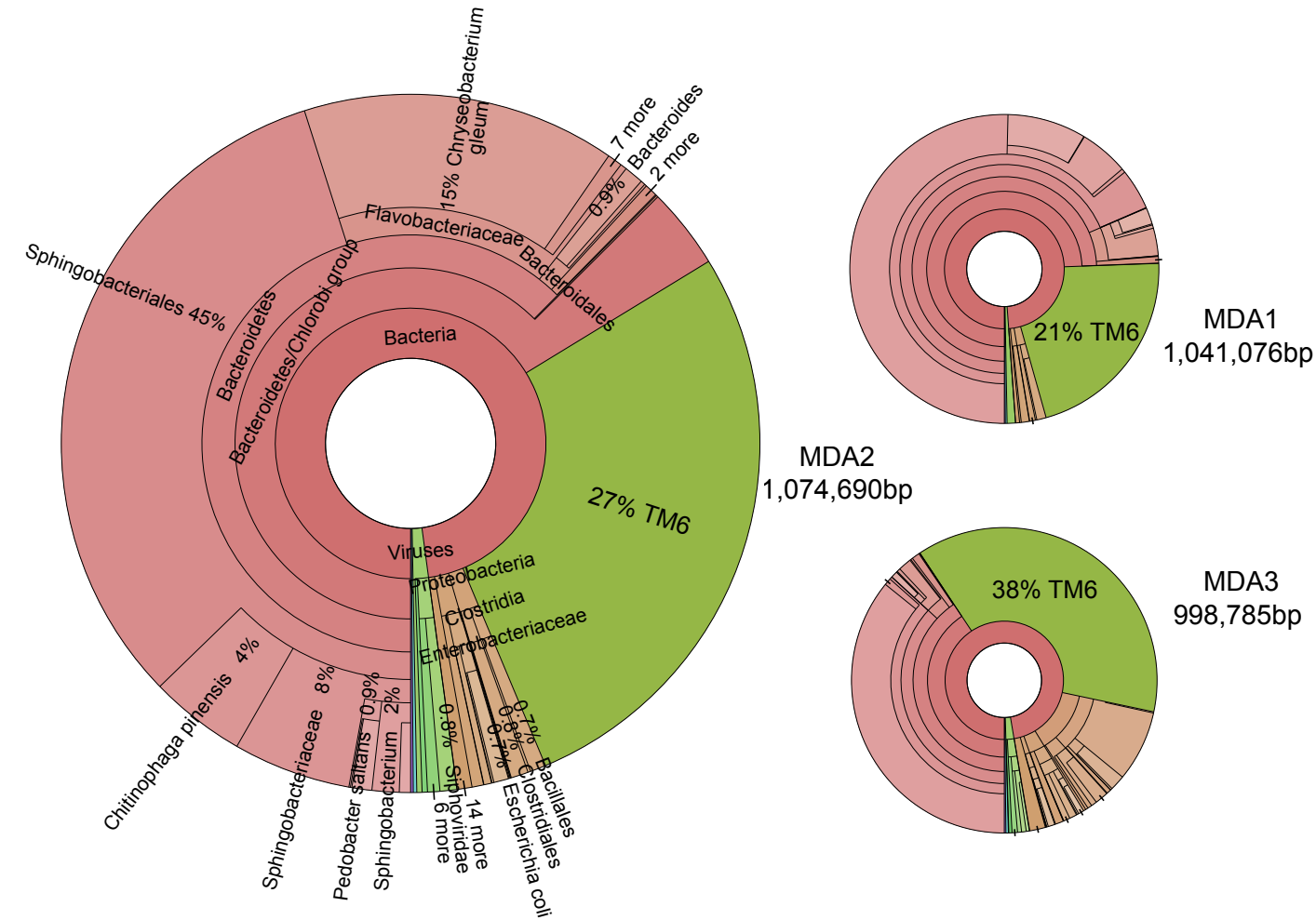


Fig. S3. **Identification of TM6 contigs in the assembled metagenome.** Taxonomic classification of assembled contigs for three independent wells representing the mini-metagenomes from 100 event sorts using MGTAXA software. A 273 kb contig containing the TM6 16S rRNA gene was used as a training sequence to generate a predictive model of nucleotide patterns for this genome. All assembled contigs were run through this pipeline and the percentage of contigs sharing similar profiles were identified and classified as belonging to TM6 (green). The total contig size representing TM6 are shown for each of the three independent amplified genomes.

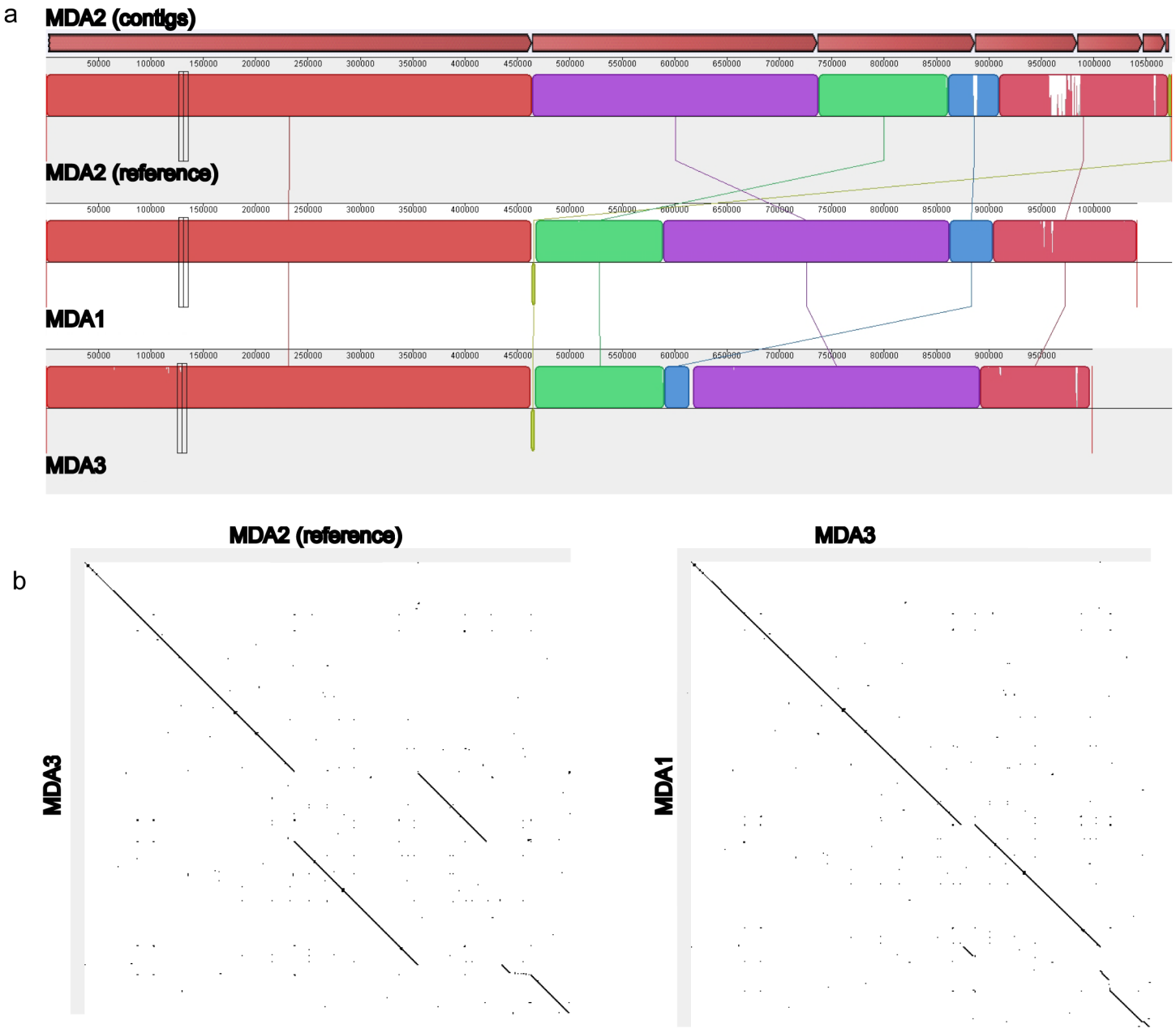


Fig. S4. Comparison of assembled TM6 genomes from MDA1 and MDA3 datasets with the concatenated MDA2 TM6 contigs (contigs were ordered by length and then concatenated). a) Contigs for each MDA were aligned with Progressive Mauve against the concatenated MDA2 contigs. b) Similarity dotplots between the concatenated MDA2 TM6 contigs and TM6 contigs from MDA1 and MDA3.

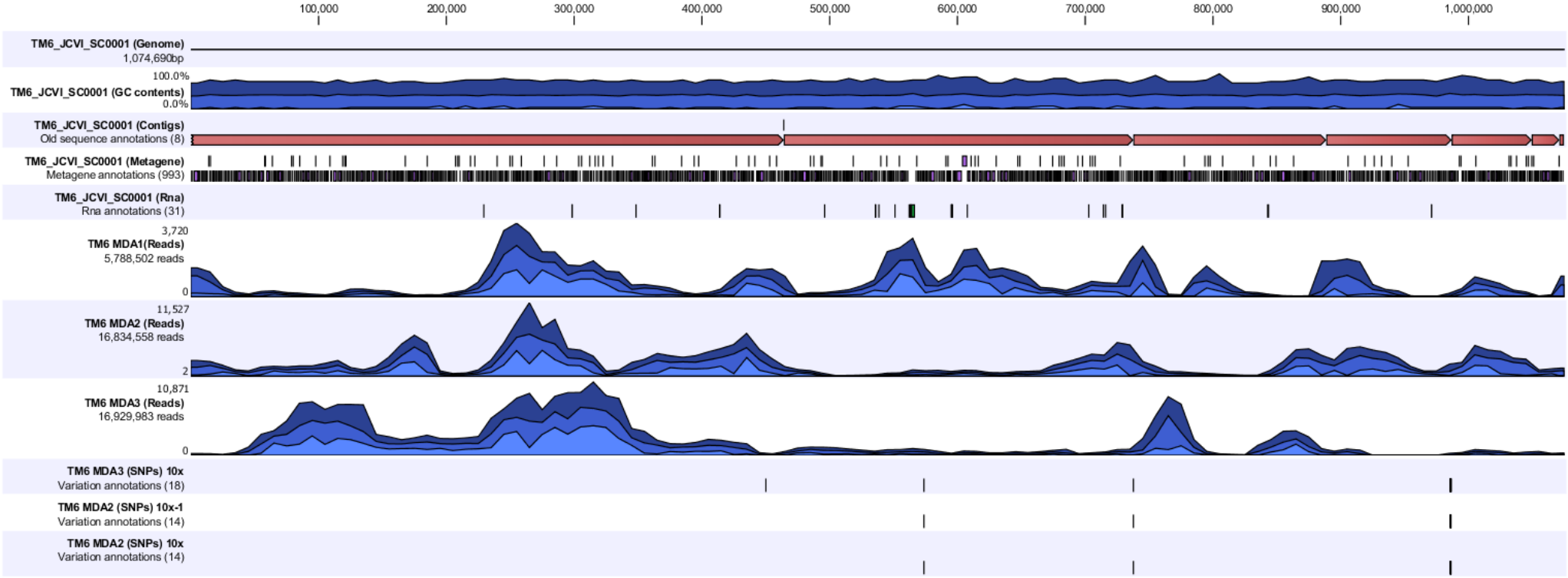


Fig. S5. Read coverage and single nucleotide polymorphisms across the concatenated MDA2 TM6 contigs (contigs were ordered by length and then concatenated). Row 1) Reference TM6 MDA2 contigs; Row 2) GC content; Row 3) MDA2 contigs; Row 4) CDS; Row 5) RNA genes; Row 6-8) depth of mapped Illumina reads from each amplified sample; Rows 9-11) SNPs at a cutoff of 10X coverage for each single cell amplification.

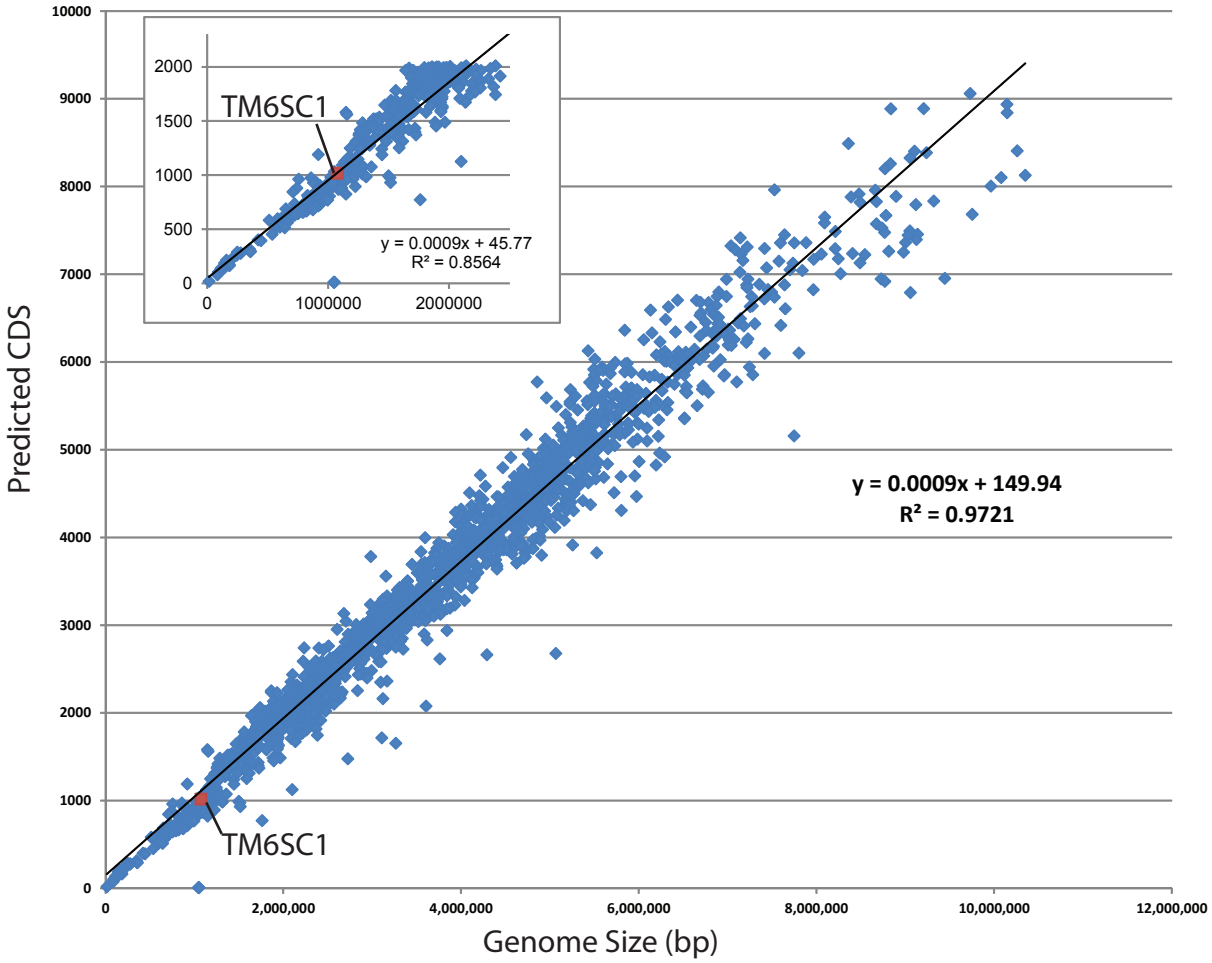
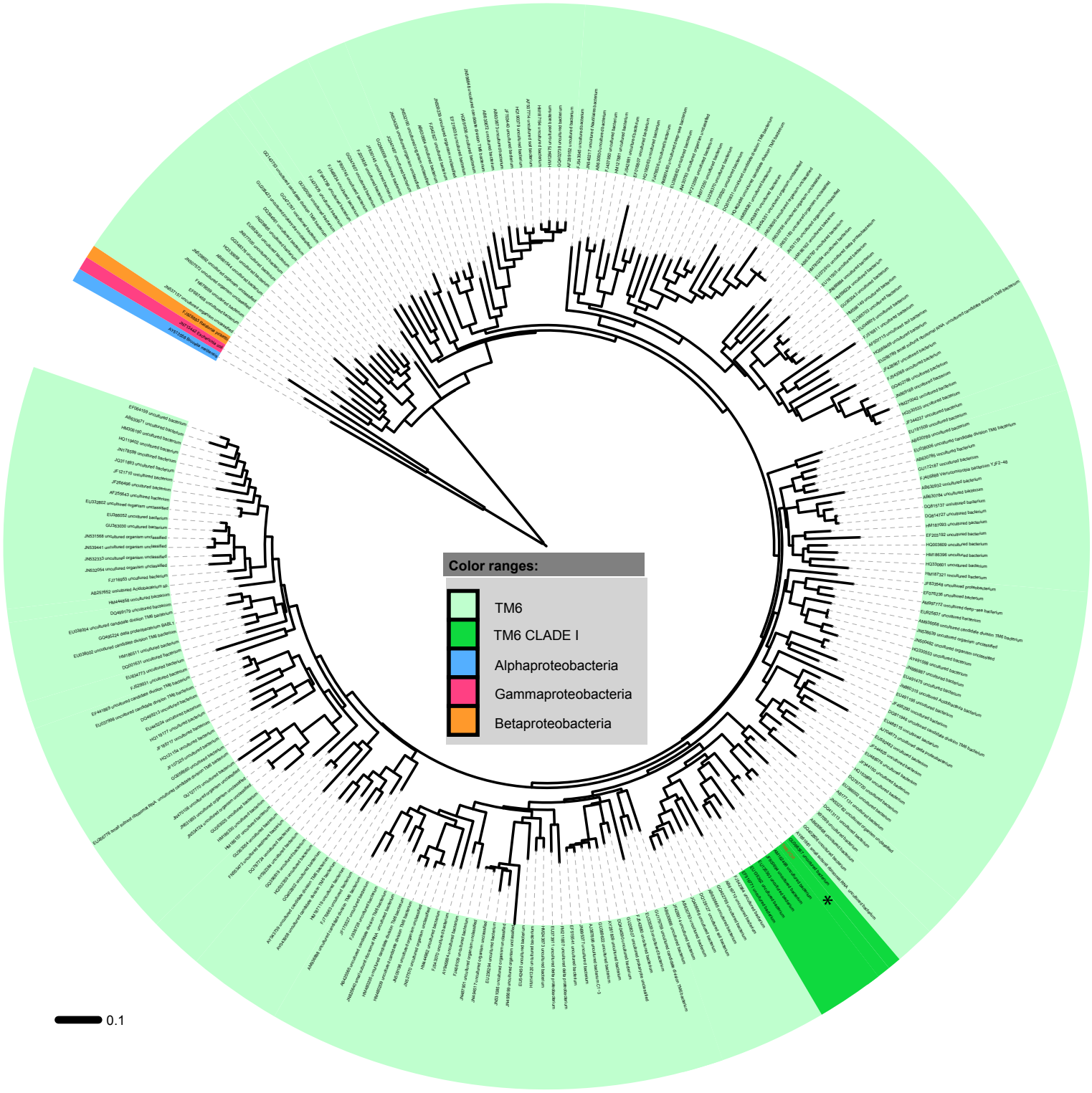


Fig. S6. Relationships between genome size of finished bacterial genomes and the number of predicted coding DNA sequences CDS. (Inset) Small bacterial genomes that have less than 2000 predicted CDS. The TM6SC1 genome is marked in red



Supplemental Figure 7. Evolutionary relationships of Candidate Division TM6. a) Phylogenetic relationship of sequences in RDP designated as members of TM6 reveal the global distribution and sequence diversity within this group. *indicates TM6 sequence from this study. Scale at bottom of figure indicates the branch length associated with 0.1 substitutions per position

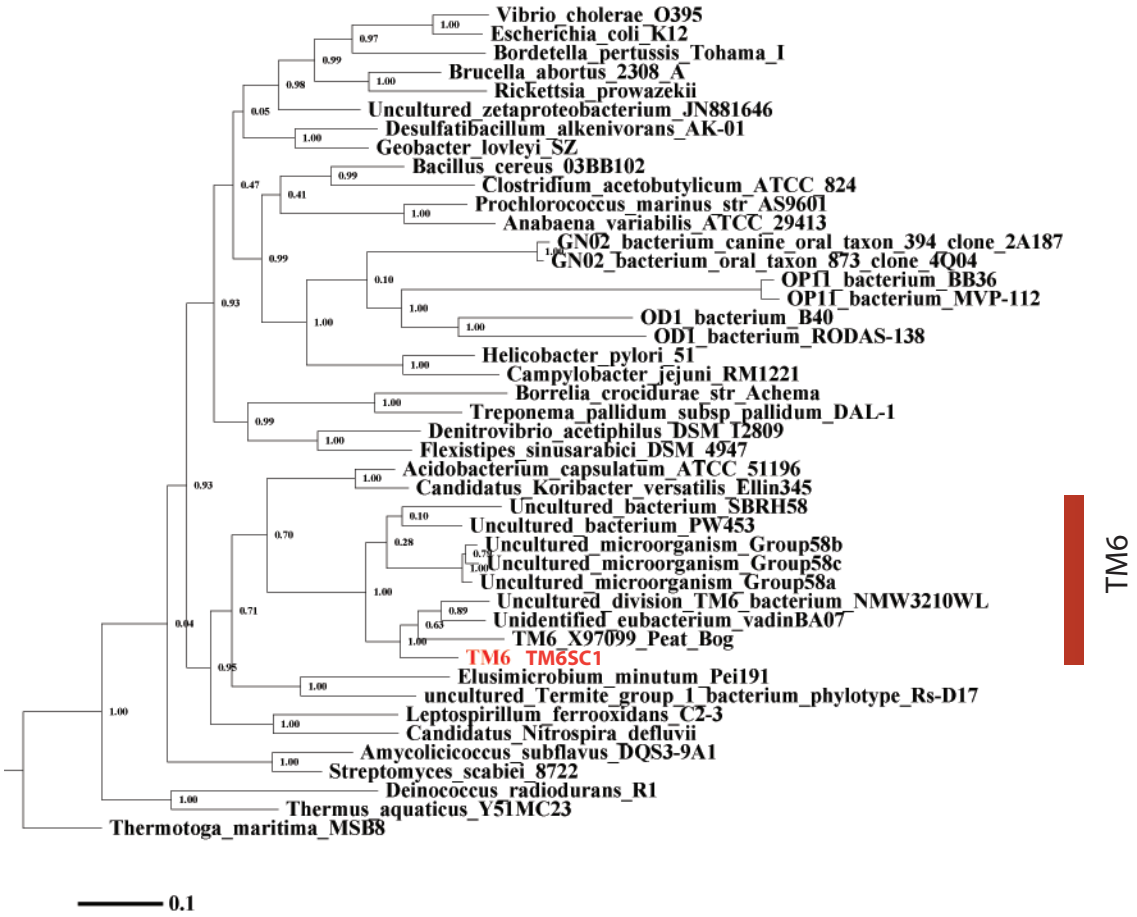


Fig. S8. **Phlyogenetic tree of 16S rRNA genes of major bacterial phyla in relation to TM6SC1**. a) Maximum likelihood tree of 16S rRNA genes from representative groups of phylogenetically distinct bacteria phyla. Support values shown are phyml's alRT values, which range between 0 and 1. Scale bar indicates the branch length associated with 0.1 substitutions per position.

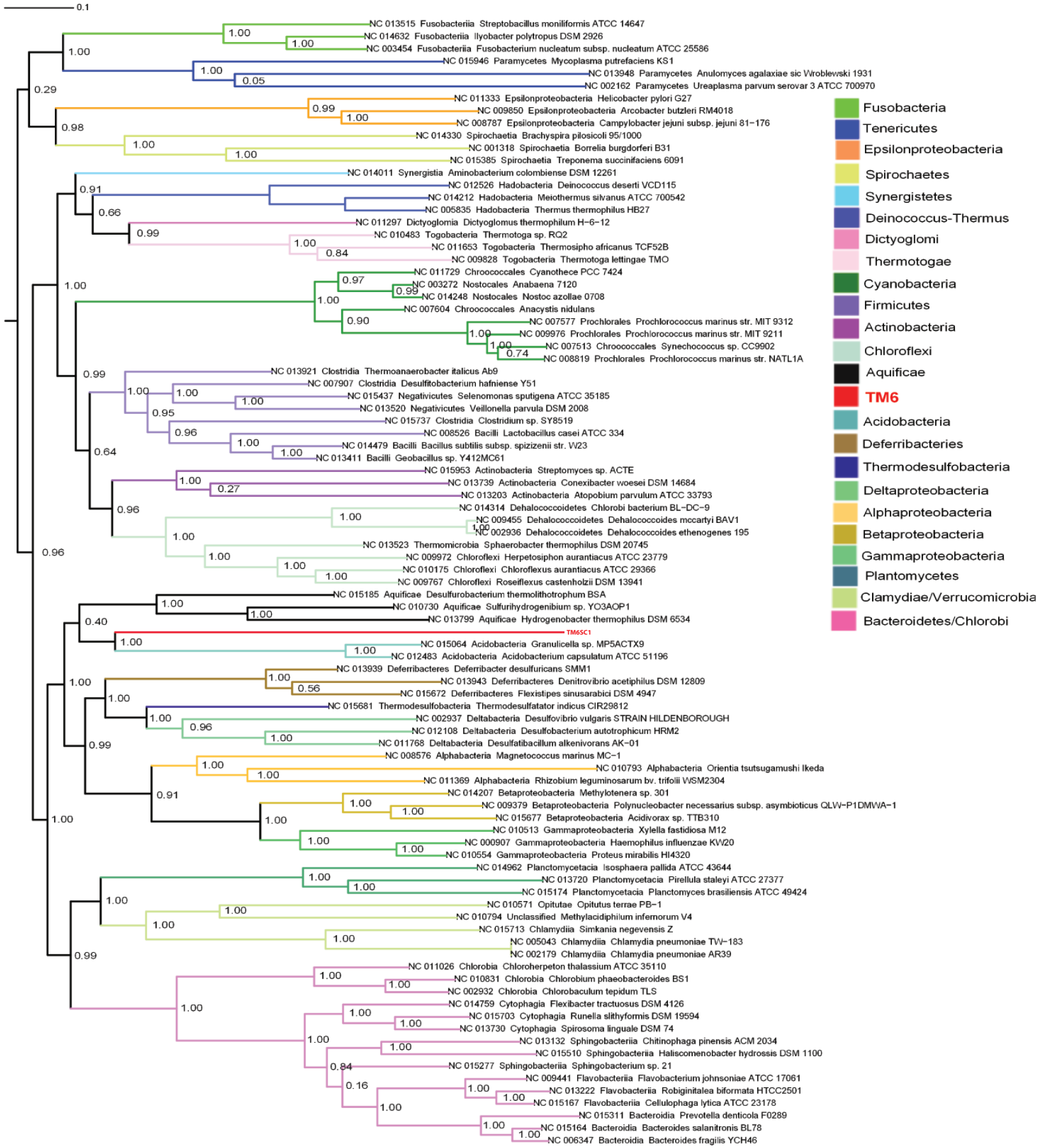


Fig. S9. Phyml tree of major phyla. This evolutionary distance dendrogram was constructed by aligning 29 conserved single copy genes from the TM6 genome to the AMPHORA2 seed alignments through HMM. All major phyla are separated into their monophyletic groups and highlighted by color. Support values shown are phyml's alRT values, which range between 0 and 1.

Table S1. Summary table of accession and publications for a number of TM6 related sequences.

Accession	Source	Title	Reference
GU368367	Copper pipe biofilm	Culture dependent and independent analyses of bacterial communities involved in copper plumbing corrosion	[1]
X97099	Peat bog	A molecular approach to search for diversity among bacteria in the Environment (original name derives from this study)	[2]
EU635154	Showerhead biofilm	Opportunistic pathogens enriched in showerhead biofilms	[3]
Q917125	Drinking water	Analysis of structure and composition of bacterial core communities in mature drinking water biofilms and bulk water of a citywide network in Germany	[4]
FJ203939	Stream biofilm	Biofilm bacterial community structure in streams affected by acid mine drainage	[5]
EU038006	Acidic cave-wall biofilm	Episodic subaerial speleogenesis controlled by mineralogy	*
DQ137999	Wetland	Microbial characteristics of the constructed wetland system receiving acid sulfate water	*
FJ265258	Paddy soil	Phylogenetic Analysis of 16S rDNA Sequences of Paddy Soil Under Long-term Fertilization	[6]
FJ625572	Forest soil	Influence of lead contamination on bacterial community in boreal pine forest soil	[7]
GU369202	Zebra mussel	Molecular characterization of bacterial communities within the zebra mussel (*Dreissena polymorpha*) in the Laurentian great lakes basin (USA)	*
GU518233	Wastewater biofilm	Bacterial community composition and diversity of a full-scale integrated fixed-film activated sludge system as investigated by pyrosequencing	[8]
AM162488	Subsurface	Vertical stratification of subsurface microbial community composition across geological formations at the Hanford Site	[9]
AY661981	Groundwater	Impacts on microbial communities and cultivable isolates from groundwater contaminated with high levels of nitric acid-bearing uranium waste at the NABIR-FRC	[10]
AM162488	Peat bog	Phylogenetic Analysis and In Situ Identification of Bacteria Community Composition in an Acidic Sphagnum Peat Bog	[11]
JF000694	BETEX aquifer	Evidence of monitored natural attenuation at BTEX contaminated aquifers: analysis of catechol 2, 3-dioxygenase and toluene/biphenyl dioxygenase genes	*
EU135362 EU135363	Prairie soil	Novelty and uniquene[i]ss patterns of rare members of the soil biosphere	
EF516771	Grassland soil	Despite strong seasonal responses, soil microbial consortia are more resilient to long-term changes in rainfall than overlying grassland	
GU179759	Oil well	Microbial diversity profiles of fluids from low-temperature petroleum reservoirs with and without exogenous water perturbation	*
EU335393	Soil	Changes in bacterial and archaeal community structure and functional diversity along a geochemically variable soil profiles	
DQ413113	SBR reactor	Comparative analysis of microbial communities from culture-dependent and –independent approaches in an anaerobic/aerobic SBR reactor	*
JN532616 JN532616	Hypersaline mat	Phylogenetic stratigraphy in the Guerrere Negro hypersaline microbial mat	*
AB630668	Aquatic moss pillars	MicroXorae of aquatic moss pillars in a freshwater lake, East Antarctica, based on fatty acid and 16S rRNA gene analyses	
GQ402806	Peat swamp forest soil	Insights into the phylogeny and metabolic potential of a primary tropical peat swamp forest microbial community by metagenomic analysis	
JF301303	Forest soil	Fungal and bacterial communities in relation to gradients of pH and N supply in boreal forest soils	*

. Unpublished Genbank record

REFERENCES

1. Pavissich, J., Vargas, I., González, B., Pastén, P. & Pizarro, G. Culture dependent and independent analyses of bacterial communities involved in copper plumbing corrosion. *Journal of applied microbiology* **109**, 771-853 (2010).
2. Rheims, H. & Rainey…, F. A molecular approach to search for diversity among bacteria in the environment. *Journal of Industrial Microbiology …* (1996).
3. Feazel, L.M. et al. Opportunistic pathogens enriched in showerhead biofilms. *Proceedings of the National Academy of Sciences* **106** (2009).
4. Henne, K., Kahlisch, L., Brettar, I. & Höfle, M. Analysis of structure and composition of bacterial core communities in mature drinking water biofilms and bulk water of a citywide network in Germany. *Applied and environmental microbiology* **78**, 3530-3538 (2012).
5. Lear, G., Niyogi, D., Harding, J., Dong, Y. & Lewis, G. Biofilm bacterial community structure in streams affected by acid mine drainage. *Applied and environmental microbiology* **75**, 3455-3515 (2009).
6. Wu, M., Qin, H., Chen, Z., Wu, J. & Wei, W. Effect of long-term fertilization on bacterial composition in rice paddy soil. *Biology and Fertility of Soils* **47**, 397-802 (2011).
7. Hui, N. et al. Influence of lead on organisms within the detritus food web of a contaminated pine forest soil. *Boreal Environ. Res* **14**, 70-155 (2009).
8. Kwon, S., Kim, T.-S., Yu, G., Jung, J.-H. & Park, H.-D. Bacterial community composition and diversity of a full-scale integrated fixed-film activated sludge system as investigated by pyrosequencing. *Journal of microbiology and biotechnology* **20**, 1717-1740 (2010).
9. Lin, X., Kennedy, D., Fredrickson, J., Bjornstad, B. & Konopka, A. Vertical stratification of subsurface microbial community composition across geological formations at the Hanford Site. *Environmental microbiology* **14**, 414-439 (2012).
10. Fields, M. et al. Impacts on microbial communities and cultivable isolates from groundwater contaminated with high levels of nitric acid-uranium waste. *FEMS microbiology ecology* **53**, 417-445 (2005).
11. Dedysh, S., Pankratov, T., Belova, S., Kulichevskaya, I. & Liesack, W. Phylogenetic analysis and in situ identification of bacteria community composition in an acidic Sphagnum peat bog. *Applied and environmental microbiology* **72**, 2110-2117 (2006).

Table S2. Conserved single copy gene list (111) and presence within TM6SC1 genome.

Conserved Genes (111)	HMM	TM6 genome
cgtA	TIGR02729	-
coaE	TIGR00152	-
fmt	TIGR00460	-
ligA	TIGR00575	-
pgk	PF00162	-
pheT	TIGR00472	-
proS	TIGR00409	-
rnc	TIGR02191	-
rpoC	TIGR02387	-
secE	TIGR00964	-
dnaK	TIGR02350	+
ffh	TIGR00959	+
glyS	TIGR00388	+
glyS	TIGR00389	+
gmk	TIGR03263	+
gyrB	TIGR01059	+
ksgA	TIGR00755	+
proS	TIGR00408	+
rpoC	TIGR02386	+
uvrB	TIGR00631	+
prfA	TIGR00019	+
alaS	TIGR00344	+
argS	PF00750	+
aspS	TIGR00459	+
cysS	TIGR00435	+
dnaA	TIGR00362	+
dnaG	TIGR01391	+
dnaN	TIGR00663	+
dnaX	TIGR02397	+
engA	TIGR03594	+
era	TIGR00436	+
frr	TIGR00496	+
ftsY	TIGR00064	+
grpE	PF01025	+
gyrA	TIGR01063	+
hisS	TIGR00442	+
ileS	TIGR00392	+
infB	TIGR00487	+
infC	TIGR00168	+
lepA	TIGR01393	+
leuS	TIGR00396	+
mnmA	TIGR00420	+
mraW	PF01795	+
nusA	TIGR01953	+
nusG	TIGR00922	+
pheS	TIGR00468	+
pheT	TIGR00472	+
rpsT	TIGR00029	+
secA	TIGR00963	+
secG	TIGR00810	+
secY	TIGR00967	+
serS	TIGR00414	+
smpB	TIGR00086	+
thrS	TIGR00418	+
tig	TIGR00115	+
tilS	TIGR02432	+
tsf	TIGR00116	+
tyrS	TIGR00234	+
valS	TIGR00422	+
ybeY	TIGR00043	+
ychF	TIGR00092	+

Conserved Genes (111)	HMM	TM6 genome
pyrG	TIGR00337	+
recA	TIGR02012	+
rfbA	TIGR00082	+
rplA	TIGR01169	+
rplB	TIGR01171	+
rplC	PF00297	+
rplD	PF00573	+
rplE	PF00281	+
rplF	PF00347	+
rplI	TIGR00158	+
rplJ	PF00466	+
rplK	TIGR01632	+
rplL	TIGR00855	+
rplM	TIGR01066	+
rplN	TIGR01067	+
rplO	TIGR01071	+
rplP	TIGR01164	+
rplQ	TIGR00059	+
rplR	TIGR00060	+
rplS	TIGR01024	+
rplT	TIGR01032	+
rplU	TIGR00061	+
rplV	TIGR01044	+
rplW	PF00276	+
rplX	TIGR01079	+
rpmA	TIGR00062	+
rpmB	TIGR00009	+
rpmC	TIGR00012	+
rpmF	TIGR01031	+
rpmH	TIGR01030	+
rpmI	TIGR00001	+
rpoA	TIGR02027	+
rpoB	TIGR02013	+
rpsB	TIGR01011	+
rpsC	TIGR01009	+
rpsD	TIGR01017	+
rpsE	TIGR01021	+
rpsF	TIGR00166	+
rpsG	TIGR01029	+
rpsH	PF00410	+
rpsI	PF00380	+
rpsJ	TIGR01049	+
rpsK	PF00411	+
rpsL	TIGR00981	+
rpsM	PF00416	+
rpsO	TIGR00952	+
rpsP	TIGR00002	+
rpsQ	PF00366	+
rpsR	TIGR00165	+
rpsS	TIGR01050	+

Table S3. Putative proteins with unique features in the TM6 genome.

Name, NCBI accession No.	Putative function
Sporulation related	
SpoIID/LytB domain-containing protein, ACM21796	Important for membrane migration and early steps of engulfment during endospore formation
spoVG, ADK84098	Control of sporulation initiation and cell division
Cell surface binding protein	
Ricin Lectin B[(1)], ABX04072	Binds to carbohydrates
Competence protein and related	
PilM, ACM21413	Type IV pilus assembly protein
ComEC/Rec2, ACL70042	Multi-membrane protein involved in DNA internalization
Secretion proteins	
Protein D, CAE79475	Type II and III secretion system proteins
Protein D precursor, EAT1437	Type II and III secretion system proteins
Virulence	
Streptomycin-6-phosphotransferase, BAG40005	Antibiotic inactivation
CarD family transcription regulator, AAS96050	Homologues to transcription regulator LtpA, exclusively expressed during *Borrelia burgdorferi* cultivation[(2)]
GroEL and GroES co-chaperonins, ACY19055	Bacteriophages T4 and RB49 proteins Gp31 and RB-49
Clostripan peptidase C11, ABB31446	Cystein protease domain performs proteolysis of host protein. A CPDmartx protein present in pathogenic Proteobacteria[(3)]
Phage spo1 DNA polymerase related protein, ADR36304	Induces viral DNA Polymerase activity during bacterial infection[(4)]
OMP18 outer membrane protein, SBK83275	Serologic detector for *Helicobacter* infection[(5)]
Putative CRISPR XR138021	Identical repeats (TTTCAAAAGTTATCCACC) encoding a putative disease resistance protein

Table S4. Proteins related to Archaea. BLASTP hits longer than 50 aa with greater than 50% sequence identity were included.

Closest matching Organism (MG-RAST)	Fragment size (bp)	KEGG Function	E-value	Sequence Identity
Pyrobaculum arsenaticum	(422 bp)	NUDIX hydrolase	1E-35	91/137 (66%)
Archaeaglobus profundus	(1577 bp)	AMP-dep. synthetase and ligase	4E-04	21/31 (68%)
Desulfurubacterium thermolithotrophum	(1844 bp)	Amino transferase	1E-152	356/562 (63%)
Pyrococcus furiosus	(1310 bp)	Deaminase	8E-07	101/126 (80%)
Methanococcus voltae A3	(1703 bp)	CTP synthase	7E-179	367/553 (66%)
Candidatus Nitrosoarchaeum	(962 bp)	Translation factor SUA5	1E-85	204/326 (63%)

Table S5. Taxonomic classification of CDS within TM6SC1 to known symbiotic organisms.

Common_name	TAXON	EVIDENCE	gene_symbol	E.C.
hypothetical protein	*Wolbachia endosymbiont strain TRS of Brugia malayi*	UniRef100_Q5GS46		
conserved putative membrane protein	*Waddlia chondrophila WSU 86-1044*	UniRef100_D6YTP4		
hypothetical protein	*Waddlia chondrophila WSU 86-1044*	UniRef100_D6YVV7	cpsT	
methionyl aminopeptidase	*uncultured Termite group 1 bacterium phylotype Rs-D17*	UniRef100_B1GZA4		3.4.11.18
ribosomal protein L16	*uncultured Termite group 1 bacterium phylotype Rs-D17*	TIGR01164	rplP	
triose-phosphate isomerase	*Rickettsiella grylli*	PF00121	tpiA	5.3.1.1
valine--tRNA ligase	*Rickettsia prowazekii*	TIGR00422	valS	6.1.1.9
tolQ protein	*Rickettsia peacockii str. Rustic*	UniRef100_C4K1W3		
AAA+ superfamily protein	*Rickettsia endosymbiont of Ixodes scapularis*	UniRef100_C4YVA4		
phosphohydrolase	*Rickettsia endosymbiont of Ixodes scapularis*	UniRef100_C4YYI5		
AAA+ superfamily protein	*Rickettsia bellii RML369-C*	UniRef100_Q1RJW9		
trigger factor	*Rickettsia akari str. Hartford*	UniRef100_A8GQ54	tig	
hypothetical protein	*Planctomyces maris DSM 8797*	UniRef100_A6CGG7		
oxidoreductase, zinc-binding dehydrogenase family protein	*Planctomyces maris DSM 8797*	UniRef100_A6C733		
hypothetical protein	*Parachlamydia acanthamoebae str. Hall's coccus*	UniRef100_D1R7K7		
hypothetical protein	*Parachlamydia acanthamoebae str. Hall's coccus*	UniRef100_D1R7L0		
hypothetical protein	*Parachlamydia acanthamoebae str. Hall's coccus*	UniRef100_D1R6G7		
hypothetical protein	*Parachlamydia acanthamoebae str. Hall's coccus*	UniRef100_D1R7D6		
hypothetical protein	*Parachlamydia acanthamoebae str. Hall's coccus*	UniRef100_D1RAY7		
hypothetical protein	*Parachlamydia acanthamoebae str. Hall's coccus*	UniRef100_D1R681		
hypothetical protein	*Parachlamydia acanthamoebae str. Hall's coccus*	UniRef100_D1R6G6		
hypothetical protein	*Parachlamydia acanthamoebae str. Hall's coccus*	UniRef100_D1RA48		
hypothetical protein	*Parachlamydia acanthamoebae str. Hall's coccus*	UniRef100_D1R681		
hypothetical protein	*Parachlamydia acanthamoebae str. Hall's coccus*	UniRef100_D1R7K7		
hypothetical protein	*Parachlamydia acanthamoebae str. Hall's coccus*	UniRef100_D1RBI5		
hypothetical protein	*Parachlamydia acanthamoebae str. Hall's coccus*	UniRef100_D1RBK2		
UPF0235 protein pah_c178o054	*Parachlamydia acanthamoebae str. Hall's coccus*	UniRef100_D1R9K8		
glycosyltransferase, family 2	*Mycobacterium vanbaalenii PYR-1*	UniRef100_A1T4K0		
pyridine nucleotide-disulfide oxidoreductase	*Mycobacterium leprae*	PF07992	trxB	
16S rRNA methyltransferase gidB	*Listeria seeligeri serovar 1/2b str. SLCC3954*	TIGR00138	gidB	2.1.-.-
flavin containing amine oxidoreductase	*Legionella pneumophila str. Paris*	PF01593		
tipAS antibiotic-recognition domain	*Legionella pneumophila str. Paris*	PF07739		
hypothetical protein	*Legionella pneumophila str. Lens*	UniRef100_Q5WZY4		
hypothetical protein	*Legionella pneumophila str. Corby*	UniRef100_A5IFS2		
hypothetical protein	*Legionella longbeachae*	UniRef100_D3HJQ2		
similar to chloroperoxidase	*Legionella longbeachae*	UniRef100_D3HSL9	cpo	1.11.1.10
histone methylation protein DOT1	*Legionella drancourtii LLAP12*	PF08123		
hypothetical protein	*Legionella drancourtii LLAP12*	UniRef100_C6MZA6		
putative dioxygenase	*Coxiella burnetii RSA 334*	UniRef100_A9ZHF7		
hypothetical protein	*Coxiella burnetii*	UniRef100_A9KCX7		
ribonucleoside-diphosphate reductase subunit beta	*Chlamydia muridarum*	UniRef100_Q9PL92	nrdB	1.17.4.1
uncharacterized protein TC_0114	*Chlamydia muridarum*	UniRef100_Q9PLI5		
uncharacterized protein TC_0114	*Chlamydia muridarum*	UniRef100_Q9PLI5		
D-ala D-ala ligase N-terminal domain protein	*Candidatus Protochlamydia amoebophila UWE25*	PF01820	ddlA	
hypothetical protein	*Candidatus Protochlamydia amoebophila UWE25*	UniRef100_Q6MA51		
hypothetical protein	*Candidatus Protochlamydia amoebophila UWE25*	UniRef100_Q6MD30		
hypothetical protein	*Candidatus Protochlamydia amoebophila UWE25*	UniRef100_Q6MA19		
hypothetical protein	*Candidatus Protochlamydia amoebophila UWE25*	UniRef100_Q6MA19		
probable sodium/proline symporter	*Candidatus Protochlamydia amoebophila UWE25*	UniRef100_Q6MAG0	putP	
putative branched-chain amino acid transport system II carrier protein	*Candidatus Protochlamydia amoebophila UWE25*	UniRef100_Q6MD58	braB	
putative sodium/pantothenate symporter (pantothenate permease)	*Candidatus Protochlamydia amoebophila UWE25*	UniRef100_Q6MA53	panF	
tRNA-guanine transglycosylase	*Candidatus Protochlamydia amoebophila UWE25*	TIGR00430	tgt	2.4.2.29
holliday junction DNA helicase ruvB	*Candidatus Hamiltonella defensa 5AT (Acyrthosiphon pisum)*	TIGR00635	ruvB	3.6.1.-
tRNA nucleotidyltransferase	*Candidatus Hamiltonella defensa 5AT (Acyrthosiphon pisum)*	UniRef100_C4K3X0	rph	2.7.7.56
hypothetical protein	*Candidatus Amoebophilus asiaticus 5a2*	UniRef100_B3EU24		
hypothetical protein	*Candidatus Amoebophilus asiaticus 5a2*	UniRef100_B3ETM4		
hypothetical protein	*Candidatus Amoebophilus asiaticus 5a2*	UniRef100_B3ETM4		
MutS domain V protein	*Candidatus Amoebophilus asiaticus 5a2*	PF00488		
putative ABC transporter ATP-binding subunit	*Burkholderia pseudomallei*	UniRef100_Q63Q81		
amino acid permease-associated region	*Burkholderia multivorans ATCC 17616*	UniRef100_A9AKF8		
DNA polymerase ligD, polymerase domain	*Bradyrhizobium sp. BTAi1*	TIGR02778	ligD	
bll2851 protein	*Bradyrhizobium japonicum*	UniRef100_Q89RC1		
ribosomal protein L17	*Borrelia turicatae 91E135*	TIGR00059	rplQ	
tRNA (5-methylaminomethyl-2-thiouridylate)-methyltransferase	*Borrelia hermsii DAH*	TIGR00420	trmU	2.1.1.61
1-acyl-sn-glycerol-3-phosphate acyltransferase	*Borrelia duttonii Ly*	UniRef100_B5RKT9	plsC	