

## Additional File 1 – Andersson *et al.*

### The hybrid-cluster protein dataset.

gi	Short	Species	Domain	Group	Comments
32470344	Aci.fer	<i>Acidithiobacillus ferrooxidans</i>	B	$\gamma$ -Proteobacteria	
32033384	Act.ple	<i>Actinobacillus pleuropneumoniae</i> serovar 1 str. 4074	B	$\gamma$ -Proteobacteria	
53765312	Ana.var	<i>Anabaena variabilis</i> ATCC 29413	B	Cyanobacteria	
30263400	Bac.ant	<i>Bacillus anthracis</i> str. Ames	B	Low G+C Gram positive	
30021542	Bac.cer	<i>Bacillus cereus</i> ATCC 14579	B	Low G+C Gram positive	97,1% identity with Bac.ant
53713436	Bac.fra	<i>Bacteroides fragilis</i> YCH46	B	Bacteroidetes/Chlorobi group	
29346097	Bac.the	<i>Bacteroides thetaiotaomicron</i> VPI-5482	B	Bacteroidetes/Chlorobi group	
46312750	Bur.cep	<i>Burkholderia cepacia</i> R18194	B	$\beta$ -Proteobacteria	
JGI	Chl.re1	<i>Chlamydomonas reinhardtii</i>	E	Green algae	
JGI	Chl.re2	<i>Chlamydomonas reinhardtii</i>	E	Green algae	
21674516	Chl.tep	<i>Chlorobium tepidum</i> TLS	B	Bacteroidetes/Chlorobi group	
15896007	Clo.ac1	<i>Clostridium acetobutylicum</i>	B	Low G+C Gram positive	
15896669	Clo.ac2	<i>Clostridium acetobutylicum</i>	B	Low G+C Gram positive	
17976847	Clo.pe1	<i>Clostridium perfringens</i>	B	Low G+C Gram positive	
18311566	Clo.pe2	<i>Clostridium perfringens</i>	B	Low G+C Gram positive	97,6% identity with Clo.pe1
28210163	Clo.tet	<i>Clostridium tetani</i> E88	B	Low G+C Gram positive	
23021132	Clo.the	<i>Clostridium thermocellum</i> ATCC 27405	B	Low G+C Gram positive	
23111958	Des.haf	<i>Desulfitobacterium hafniense</i>	B	Low G+C Gram positive	
50875631	Des.ps1	<i>Desulfotalea psychrophila</i> LSv54	B	$\delta$ -Proteobacteria	
50876115	Des.ps2	<i>Desulfotalea psychrophila</i> LSv54	B	$\delta$ -Proteobacteria	
50876114	Des.ps3	<i>Desulfotalea psychrophila</i> LSv54	B	$\delta$ -Proteobacteria	
20663611	Des.de1	<i>Desulfovibrio desulfuricans</i>	B	$\delta$ -Proteobacteria	
20663615	Des.de2	<i>Desulfovibrio desulfuricans</i>	B	$\delta$ -Proteobacteria	100,0% identity with Des.de1
400847	Des.de3	<i>Desulfovibrio desulfuricans</i>	B	$\delta$ -Proteobacteria	100,0% identity with Des.de1
23474024	Des.de4	<i>Desulfovibrio desulfuricans</i> G20	B	$\delta$ -Proteobacteria	
46580947	Des.vH1	<i>Desulfovibrio vulgaris</i> subsp. <i>vulgaris</i> str. Hildenborough	B	$\delta$ -Proteobacteria	
46580421	Des.vH2	<i>Desulfovibrio vulgaris</i> subsp. <i>vulgaris</i> str. Hildenborough	B	$\delta$ -Proteobacteria	
18252516	Ent.his	<i>Entamoeba histolytica</i>	E	Entamoebidae	
this study	Ent.in1	<i>Entamoeba invadens</i>	E	Entamoebidae	frameshift
this study	Ent.in2	<i>Entamoeba invadens</i>	E	Entamoebidae	
this study	Ent.mos	<i>Entamoeba moshkovskii</i>	E	Entamoebidae	
this study	Ent.ter	<i>Entamoeba terrapinae</i>	E	Entamoebidae	fragment missing 272 of 418 aligned positions - 80,1% identity with Ent.his
49612112	Erw.car	<i>Erwinia carotovora</i> subsp. <i>atroseptica</i> SCRI1043	B	$\gamma$ -Proteobacteria	
1787099	Esc.co1	<i>Escherichia coli</i> K12	B	$\gamma$ -Proteobacteria	
33347490	Esc.co2	<i>Escherichia coli</i> K12	B	$\gamma$ -Proteobacteria	100,0% identity with Esc.co1
4062452	Esc.co3	<i>Escherichia coli</i> K12	B	$\gamma$ -Proteobacteria	100,0% identity with Esc.co1
19704019	Fus.nuc	<i>Fusobacterium nucleatum</i> subsp. <i>nucleatum</i> ATCC 25586	B	Fusobacteria	
48844072	Geo.met	<i>Geobacter metallireducens</i> GS-15	B	$\delta$ -Proteobacteria	
39995780	Geo.sul	<i>Geobacter sulfurreducens</i> PCA	B	$\delta$ -Proteobacteria	96,2% identity with Geo.met
18252518	Gia.lam	<i>Giardia intestinalis</i>	E	Diplomonad	
48832063	Mag.sp.	<i>Magnetococcus</i> sp. MC-1	B	Proteobacteria (unclassified)	
52306398	Man.suc	<i>Mannheimia succiniciproducens</i> MBEL55E	B	$\gamma$ -Proteobacteria	95,7% identity with Act.ple
this study	Mas.bal	<i>Mastigamoeba balamuthi</i>	E	Pelobiont	
15668946	Met.ja1	<i>Methanocaldococcus jannaschii</i>	A	Euryarchaeota	
7397414	Met.ja2	<i>Methanocaldococcus jannaschii</i>	A	Euryarchaeota	100,0% identity with Met.ja1
46142247	Met.bu1	<i>Methanococcoides burtonii</i> DSM 6242	A	Euryarchaeota	
53731535	Met.bu2	<i>Methanococcoides burtonii</i> DSM 6242	A	Euryarchaeota	
45358283	Met.mar	<i>Methanococcus maripaludis</i> S2	A	Euryarchaeota	
21229315	Met.maz	<i>Methanosarcina mazei</i> Goe1	A	Euryarchaeota	
7404412	Met.the	<i>Methanothermobacter thermautotrophicus</i> str. Delta H	A	Euryarchaeota	
49235681	Mor.the	<i>Moorella thermoacetica</i>	B	Low G+C Gram positive	
this study	Nae.gru	<i>Naegleria gruberi</i>	E	Heterolobosea	
26522739	Pho.pho	<i>Photobacterium phosphoreum</i>	B	$\gamma$ -Proteobacteria	
34540658	Por.gin	<i>Porphyromonas gingivalis</i> W83	B	Bacteroidetes/Chlorobi group	
14521200	Pyr.abv	<i>Pyrococcus abyssi</i>	A	Euryarchaeota	
18977064	Pyr.fur	<i>Pyrococcus furiosus</i> DSM 3638	A	Euryarchaeota	
33415324	Rho.cap	<i>Rhodobacter capsulatus</i>	B	$\alpha$ -Proteobacteria	
22966189	Rho.rub	<i>Rhodospirillum rubrum</i>	B	$\alpha$ -Proteobacteria	
16759810	Sal.ent	<i>Salmonella enterica</i> subsp. <i>enterica</i>	B	$\gamma$ -Proteobacteria	96,4% identity with Esc.co1

---

		serovar Typhi			
16764299	Sal.typ	<i>Salmonella typhimurium</i> LT2	B	$\gamma$ -Proteobacteria	96,6% identity with Esc.co1
24372941	She.one	<i>Shewanella oneidensis</i> MR-1	B	$\gamma$ -Proteobacteria	
24112244	Shi.fle	<i>Shigella flexneri</i> 2a str. 301	B	$\gamma$ -Proteobacteria	99,5% identity with Esc.co1
27983818	Spi.bar	<i>Spironucleus barkhanus</i>	E	Diplomonad	
51892166	Sym.the	<i>Symbiobacterium thermophilum</i> IAM 14863	B	High G+C Gram positive	
JGI	Tha.pse	<i>Thalassiosira pseudonana</i>	E	Diatom	
20809101	The.ten	<i>Thermoanaerobacter tengcongensis</i>	B	Low G+C Gram positive	
15643928	The.mar	<i>Thermotoga maritima</i>	B	Thermotogales	
42525995	Tre.den	<i>Treponema denticola</i> ATCC 35405	B	Spirochete	
18252520	Tri.va1	<i>Trichomonas vaginalis</i>	E	Parabasalid	
TIGR	Tri.va2	<i>Trichomonas vaginalis</i>	E	Parabasalid	
59714046	Vib.fis	<i>Vibrio fischeri</i> ES114	B	$\gamma$ -Proteobacteria	
28897963	Vib.par	<i>Vibrio parahaemolyticus</i> RIMD 2210633	B	$\gamma$ -Proteobacteria	
27366144	Vib.vul	<i>Vibrio vulnificus</i> CMCP6	B	$\gamma$ -Proteobacteria	96,6% identity with Vib.par
34557553	Wol.suc	<i>Wolinella succinogenes</i>	B	$\epsilon$ -Proteobacteria	
22126696	Yer.pes	<i>Yersinia pestis</i> KIM	B	$\gamma$ -Proteobacteria	
56542587	Zym.mob	<i>Zymomonas mobilis</i> subsp. mobilis ZM4	B	$\alpha$ -Proteobacteria	

---

## The A-type flavoprotein dataset.

gi	Short	Species	Domain	Group	Comment
37913107	Aer.hyd	<i>Aeromonas hydrophila</i>	B	$\gamma$ -Proteobacteria	fragment missing 120 of 269 aligned positions – 91,5% identical with Erw.car
45510255	Ana.va1	<i>Anabaena variabilis</i> ATCC 29413	B	Cyanobacteria	99,3% identity with Nos.sp5
53764133	Ana.va2	<i>Anabaena variabilis</i> ATCC 29413	B	Cyanobacteria	96,3% identity with Nos.sp3
53764449	Ana.va3	<i>Anabaena variabilis</i> ATCC 29413	B	Cyanobacteria	97,8% identity with Nos.sp1
53764450	Ana.va4	<i>Anabaena variabilis</i> ATCC 29413	B	Cyanobacteria	100,0% identity with Nos.sp2
45510253	Ana.va5	<i>Anabaena variabilis</i> ATCC 29413	B	Cyanobacteria	97,0% identity with Nos.sp6
53764134	Ana.va6	<i>Anabaena variabilis</i> ATCC 29413	B	Cyanobacteria	96,3% identity with Nos.sp4
11497784	Arc.fu1	<i>Archaeoglobus fulgidus</i> DSM 4304	A	Euryarchaeota	
11499115	Arc.fu2	<i>Archaeoglobus fulgidus</i> DSM 4304	A	Euryarchaeota	
56475711	Azo.spE	<i>Azoarcus</i> sp. EbN1	B	$\beta$ -Proteobacteria	
60491887	Bac.fra	<i>Bacteroides fragilis</i> NCTC 9343	B	Bacteroidetes/Chlorobi group	
29349534	Bac.the	<i>Bacteroides thetaiotaomicron</i> VPI-5482	B	Bacteroidetes/Chlorobi group	
21675091	Chl.tep	<i>Chlorobium tepidum</i> TLS	B	Bacteroidetes/Chlorobi group	
15894314	Clo.ac1	<i>Clostridium acetobutylicum</i>	B	Low G+C Gram positives	
15895714	Clo.ac2	<i>Clostridium acetobutylicum</i>	B	Low G+C Gram positives	
18309270	Clo.pe1	<i>Clostridium perfringens</i>	B	Low G+C Gram positives	
18309763	Clo.pe2	<i>Clostridium perfringens</i>	B	Low G+C Gram positives	
18309996	Clo.pe3	<i>Clostridium perfringens</i>	B	Low G+C Gram positives	
28210710	Clo.te1	<i>Clostridium tetani</i> E88	B	Low G+C Gram positives	
28211064	Clo.te2	<i>Clostridium tetani</i> E88	B	Low G+C Gram positives	
28211481	Clo.te3	<i>Clostridium tetani</i> E88	B	Low G+C Gram positives	
28211510	Clo.te4	<i>Clostridium tetani</i> E88	B	Low G+C Gram positives	
23020486	Clo.the	<i>Clostridium thermocellum</i> ATCC 27405	B	Low G+C Gram positives	
45528365	Cro.wa1	<i>Crocospaera watsonii</i> WH 8501	B	Cyanobacteria	
45527451	Cro.wa2	<i>Crocospaera watsonii</i> WH 8501	B	Cyanobacteria	
41722689	Dec.aro	<i>Dechloromonas aromatica</i> RCB	B	$\beta$ -Proteobacteria	
50876536	Des.ps1	<i>Desulfotalea psychrophila</i> LSv54	B	$\delta$ -Proteobacteria	
50877008	Des.ps2	<i>Desulfotalea psychrophila</i> LSv54	B	$\delta$ -Proteobacteria	
50876535	Des.ps3	<i>Desulfotalea psychrophila</i> LSv54	B	$\delta$ -Proteobacteria	
23473638	Des.des	<i>Desulfovibrio desulfuricans</i> G20	B	$\delta$ -Proteobacteria	
14916694	Des.gig	<i>Desulfovibrio gigas</i>	B	$\delta$ -Proteobacteria	
46581587	Des.vu1	<i>Desulfovibrio vulgaris</i> subsp. vulgaris str. Hildenborough	B	$\delta$ -Proteobacteria	
46580422	Des.vu2	<i>Desulfovibrio vulgaris</i> subsp. vulgaris str. Hildenborough	B	$\delta$ -Proteobacteria	
56474182	Ent.hi1	<i>Entamoeba histolytica</i>	E	Entamoebidae	
56471042	Ent.hi2	<i>Entamoeba histolytica</i>	E	Entamoebidae	
56468393	Ent.hi3	<i>Entamoeba histolytica</i>	E	Entamoebidae	
56468595	Ent.hi4	<i>Entamoeba histolytica</i>	E	Entamoebidae	
this study	Ent.inv	<i>Entamoeba invadens</i>	E	Entamoebidae	
this study	Ent.mo1	<i>Entamoeba moshkovskii</i>	E	Entamoebidae	
this study	Ent.mo2	<i>Entamoeba moshkovskii</i>	E	Entamoebidae	
this study	Ent.ter	<i>Entamoeba terrapinae</i>	E	Entamoebidae	
49610370	Erw.car	<i>Erwinia carotovora</i> subsp. atroseptica SCRI1043	B	$\gamma$ -Proteobacteria	
16130617	Esc.col	<i>Escherichia coli</i> K12	B	$\gamma$ -Proteobacteria	
19703847	Fus.nu1	<i>Fusobacterium nucleatum</i> subsp. nucleatum ATCC 25586	B	Fusobacteria	
19704755	Fus.nu2	<i>Fusobacterium nucleatum</i> subsp. nucleatum ATCC 25586	B	Fusobacteria	
23055925	Geo.me1	<i>Geobacter metallireducens</i>	B	$\delta$ -Proteobacteria	
23056906	Geo.me2	<i>Geobacter metallireducens</i>	B	$\delta$ -Proteobacteria	100,0% identity with Geo.me1
39998383	Geo.sul	<i>Geobacter sulfurreducens</i> PCA	B	$\delta$ -Proteobacteria	
27981645	Gia.lam	<i>Giardia intestinalis</i>	E	Diplomonads	
37521344	Glo.vi1	<i>Gloeobacter violaceus</i>	B	Cyanobacteria	
37521345	Glo.vi2	<i>Gloeobacter violaceus</i>	B	Cyanobacteria	
this study	Mas.bal	<i>Mastigamoeba balamuthi</i>	E	Pelobionts	
46141957	Met.bur	<i>Methanococcoides burtonii</i> DSM 6242	A	Euryarchaeota	
15668714	Met.ja1	<i>Methanococcus jannaschii</i>	A	Euryarchaeota	
15668913	Met.ja2	<i>Methanococcus jannaschii</i>	A	Euryarchaeota	
15668929	Met.ja3	<i>Methanococcus jannaschii</i>	A	Euryarchaeota	
2127904	Met.ja4	<i>Methanococcus jannaschii</i>	A	Euryarchaeota	100,0% identity with Met.ja1
45047591	Met.ma1	<i>Methanococcus maripaludis</i> S2	A	Euryarchaeota	
45047564	Met.ma2	<i>Methanococcus maripaludis</i> S2	A	Euryarchaeota	
45047563	Met.ma3	<i>Methanococcus maripaludis</i> S2	A	Euryarchaeota	95,9% identity with Met.ma2
20092195	Met.ac1	<i>Methanosarcina acetivorans</i> str. C2A	A	Euryarchaeota	
20092541	Met.ac2	<i>Methanosarcina acetivorans</i> str. C2A	A	Euryarchaeota	

48837637	Met.bar	<i>Methanosarcina barkeri</i> str. fusaro	A	Euryarchaeota	
21229289	Met.maz	<i>Methanosarcina mazei</i> Goe1	A	Euryarchaeota	
21362551	Met.mar	<i>Methanothermobacter marburgensis</i> str. Marburg	A	Euryarchaeota	
15678185	Met.th1	<i>Methanothermobacter thermautotrophicus</i> str. Delta H	A	Euryarchaeota	
15678248	Met.th2	<i>Methanothermobacter thermautotrophicus</i> str. Delta H	A	Euryarchaeota	
2127680	Met.th3	<i>Methanothermobacter thermautotrophicus</i> str. Delta H	A	Euryarchaeota	
21362560	Moo.the	<i>Moorella thermoacetica</i>	B	Low G+C Gram positives	
23123589	Nos.pu1	<i>Nostoc punctiforme</i>	B	Cyanobacteria	
23123808	Nos.pu2	<i>Nostoc punctiforme</i>	B	Cyanobacteria	
23123810	Nos.pu3	<i>Nostoc punctiforme</i>	B	Cyanobacteria	
23123861	Nos.pu4	<i>Nostoc punctiforme</i>	B	Cyanobacteria	
23129575	Nos.pu5	<i>Nostoc punctiforme</i>	B	Cyanobacteria	
23129576	Nos.pu6	<i>Nostoc punctiforme</i>	B	Cyanobacteria	
17227673	Nos.sp1	<i>Nostoc</i> sp. PCC 7120	B	Cyanobacteria	
17227674	Nos.sp2	<i>Nostoc</i> sp. PCC 7120	B	Cyanobacteria	
17231383	Nos.sp3	<i>Nostoc</i> sp. PCC 7120	B	Cyanobacteria	
17231387	Nos.sp4	<i>Nostoc</i> sp. PCC 7120	B	Cyanobacteria	
17231936	Nos.sp5	<i>Nostoc</i> sp. PCC 7120	B	Cyanobacteria	
17231938	Nos.sp6	<i>Nostoc</i> sp. PCC 7120	B	Cyanobacteria	
54308337	Pho.pro	<i>Photobacterium profundum</i> SS9	B	$\gamma$ -Proteobacteria	
34540586	Por.gin	<i>Porphyromonas gingivalis</i> W83	B	Bacteroidetes/Chlorobi group	
33864428	Pro.mM1	<i>Prochlorococcus marinus</i> str. MIT 9313	B	Cyanobacteria	
33864429	Pro.mM2	<i>Prochlorococcus marinus</i> str. MIT 9313	B	Cyanobacteria	
33239496	Pro.mC1	<i>Prochlorococcus marinus</i> subsp. marinus str. CCMP1375	B	Cyanobacteria	
33239497	Pro.mC2	<i>Prochlorococcus marinus</i> subsp. marinus str. CCMP1375	B	Cyanobacteria	
33860602	Pro.mC3	<i>Prochlorococcus marinus</i> subsp. pastoris str. CCMP1378	B	Cyanobacteria	
33860603	Pro.mC4	<i>Prochlorococcus marinus</i> subsp. pastoris str. CCMP1378	B	Cyanobacteria	
14521096	Pyr.abv	<i>Pyrococcus abyssi</i>	A	Euryarchaeota	
18977066	Pyr.fu1	<i>Pyrococcus furiosus</i> DSM 3638	A	Euryarchaeota	
18977123	Pyr.fu2	<i>Pyrococcus furiosus</i> DSM 3638	A	Euryarchaeota	
14590919	Pyr.hor	<i>Pyrococcus horikoshii</i>	A	Euryarchaeota	
1075539	Rho.ca1	<i>Rhodobacter capsulatus</i>	B	$\alpha$ -Proteobacteria	97,0% identity with Rho.ca2
21431763	Rho.ca2	<i>Rhodobacter capsulatus</i>	B	$\alpha$ -Proteobacteria	
22958095	Rho.sph	<i>Rhodobacter sphaeroides</i>	B	$\alpha$ -Proteobacteria	
16761617	Sal.ent	<i>Salmonella enterica</i> subsp. enterica serovar Typhi	B	$\gamma$ -Proteobacteria	96,3% identity with Esc.col
16766146	Sal.typ	<i>Salmonella typhimurium</i> LT2	B	$\gamma$ -Proteobacteria	96,3% identity with Esc.col
24114007	Shi.fle	<i>Shigella flexneri</i> 2a str. 301	B	$\gamma$ -Proteobacteria	99,3% identity with Esc.col
27984029	Spi.bar	<i>Spironucleus barkhanus</i>	E	Diplomonads	
56752293	Syn.el1	<i>Synechococcus elongatus</i> PCC 6301	B	Cyanobacteria	
45513282	Syn.el2	<i>Synechococcus elongatus</i> PCC 6301	B	Cyanobacteria	100,0% identity with Syn.el1
56752292	Syn.el3	<i>Synechococcus elongatus</i> PCC 6301	B	Cyanobacteria	
45513283	Syn.el4	<i>Synechococcus elongatus</i> PCC 6301	B	Cyanobacteria	99,6% identity with Syn.el3
33866897	Syn.sW1	<i>Synechococcus</i> sp. WH 8102	B	Cyanobacteria	
33866898	Syn.sW2	<i>Synechococcus</i> sp. WH 8102	B	Cyanobacteria	
16329320	Syn.sP1	<i>Synechocystis</i> sp. PCC 6803	B	Cyanobacteria	
16329322	Syn.sP2	<i>Synechocystis</i> sp. PCC 6803	B	Cyanobacteria	
16331061	Syn.sP3	<i>Synechocystis</i> sp. PCC 6803	B	Cyanobacteria	
16331685	Syn.sP4	<i>Synechocystis</i> sp. PCC 6803	B	Cyanobacteria	
20807480	The.ten	<i>Thermoanaerobacter tengcongensis</i>	B	Low G+C Gram positives	
57159073	The.kod	<i>Thermococcus kodakaraensis</i> KOD1	A	Euryarchaeota	
22298632	The.el1	<i>Thermosynechococcus elongatus</i> BP-1	B	Cyanobacteria	
22298916	The.el2	<i>Thermosynechococcus elongatus</i> BP-1	B	Cyanobacteria	
15643518	The.mar	<i>Thermotoga maritima</i>	B	Thermotogales	
42526588	Tre.den	<i>Treponema denticola</i> ATCC 35405	B	Spirochetes	
23039893	Tri.er1	<i>Trichodesmium erythraeum</i> IMS101	B	Cyanobacteria	
23043405	Tri.er2	<i>Trichodesmium erythraeum</i> IMS101	B	Cyanobacteria	
this study	Tri.va1	<i>Trichomonas vaginalis</i>	E	Parabasalids	
TIGR	Tri.va2	<i>Trichomonas vaginalis</i>	E	Parabasalids	
TIGR	Tri.va3	<i>Trichomonas vaginalis</i>	E	Parabasalids	
37908704	Tri.va4	<i>Trichomonas vaginalis</i>	E	Parabasalids	
59712389	Vib.fis	<i>Vibrio fischeri</i> ES114	B	$\gamma$ -Proteobacteria	
27367713	Vib.vul	<i>Vibrio vulnificus</i> CMCP6	B	$\gamma$ -Proteobacteria	
34556611	Wol.suc	<i>Wolinella succinogenes</i>	B	$\epsilon$ -Proteobacteria	

## The glucosamine-6-phosphate isomerase dataset.

gi	Short	Species	Domain	Group	Comment
46143414	Act.ple	<i>Actinobacillus pleuropneumoniae</i> serovar 1 str. 4074	B	$\gamma$ -Proteobacteria	
45509250	Ana.var	<i>Anabaena variabilis</i> ATCC 29413	B	Cyanobacteria	97,8% identity with Nos.sp
31239371	Ano.gam	<i>Anopheles gambiae</i>	E	Metazoa	
66547108	Api.mel	<i>Apis mellifera</i>	E	Metazoa	
40745392	Asp.nid	<i>Aspergillus nidulans</i> FGSC A4	E	Fungi	
21402108	Bac.ant	<i>Bacillus anthracis</i> A2012	B	Low G+C Gram positives	
30022141	Bac.cer	<i>Bacillus cereus</i> ATCC 14579	B	Low G+C Gram positives	98,2% identity with Bac.ant
56963255	Bac.cla	<i>Bacillus clausii</i> KSM-K16	B	Low G+C Gram positives	
15612983	Bac.hal	<i>Bacillus halodurans</i>	B	Low G+C Gram positives	
52079159	Bac.li1	<i>Bacillus licheniformis</i> ATCC 14580	B	Low G+C Gram positives	
52082623	Bac.li2	<i>Bacillus licheniformis</i> ATCC 14580	B	Low G+C Gram positives	
31076788	Bac.sph	<i>Bacillus sphaericus</i>	B	Low G+C Gram positives	
16077305	Bac.su1	<i>Bacillus subtilis</i>	B	Low G+C Gram positives	
16080555	Bac.su2	<i>Bacillus subtilis</i>	B	Low G+C Gram positives	
60493857	Bac.fr1	<i>Bacteroides fragilis</i> NCTC 9343	B	Bacteroidetes/Chlorobi group	
60491888	Bac.fr2	<i>Bacteroides fragilis</i> NCTC 9343	B	Bacteroidetes/Chlorobi group	96,9% identity with Bac.th3
29345668	Bac.th1	<i>Bacteroides thetaiotaomicron</i> VPI-5482	B	Bacteroidetes/Chlorobi group	long
29348995	Bac.th2	<i>Bacteroides thetaiotaomicron</i> VPI-5482	B	Bacteroidetes/Chlorobi group	
29349535	Bac.th3	<i>Bacteroides thetaiotaomicron</i> VPI-5482	B	Bacteroidetes/Chlorobi group	
23335426	Bif.lon	<i>Bifidobacterium longum</i> DJO10A	B	High G+C Gram positives	
33519783	Blo.flo	<i>Blochmannia floridanus</i>	B	$\gamma$ -Proteobacteria	
15594497	Bor.bur	<i>Borrelia burgdorferi</i>	B	Spirochetes	
51572983	Bor.gar	<i>Borrelia garinii</i> PBi	B	Spirochetes	
61811995	Bos.tau	<i>Bos taurus</i>	E	Metazoa	98,3% identity with Hom.sa1
62422909	Bre.lin	<i>Brevibacterium linens</i> BL2	B	High G+C Gram positives	
39586581	Cae.bri	<i>Caenorhabditis briggsae</i>	E	Metazoa	
17554876	Cae.ele	<i>Caenorhabditis elegans</i>	E	Metazoa	
13702164	Can.al1	<i>Candida albicans</i>	E	Fungi	
417340	Can.al2	<i>Candida albicans</i>	E	Fungi	99,6% identity with Can.al1
57042718	Can.fam	<i>Canis familiaris</i>	E	Metazoa	96,1% identity with Hom.sa1
15893480	Clo.ace	<i>Clostridium acetobutylicum</i>	B	Low G+C Gram positives	
18311416	Clo.per	<i>Clostridium perfringens</i>	B	Low G+C Gram positives	
28212200	Clo.tet	<i>Clostridium tetani</i> E88	B	Low G+C Gram positives	
38233130	Cor.dip	<i>Corynebacterium diphtheriae</i> NCTC 13129	B	High G+C Gram positives	
25029074	Cor.eff	<i>Corynebacterium efficiens</i> YS-314	B	High G+C Gram positives	
19553842	Cor.glu	<i>Corynebacterium glutamicum</i> ATCC 13032	B	High G+C Gram positives	
45522698	Cro.wat	<i>Crocospaera watsonii</i> WH 8501	B	Cyanobacteria	
50254779	Cry.neo	<i>Cryptococcus neoformans</i> var. neoformans B-3501A	E	Fungi	
23135910	Cyt.hut	<i>Cytophaga hutchinsonii</i>	B	Bacteroidetes/Chlorobi group	long
62202396	Dan.rer	<i>Danio rerio</i>	E	Metazoa	
49657079	Deb.han	<i>Debaryomyces hansenii</i> CBS767	E	Fungi	
23112797	Des.haf	<i>Desulfitobacterium hafniense</i>	B	Low G+C Gram positives	
60466211	Dic.di1	<i>Dictyostelium discoideum</i>	E	Mycetozoa	long
60470057	Dic.di2	<i>Dictyostelium discoideum</i>	E	Mycetozoa	
19920764	Dro.mel	<i>Drosophila melanogaster</i>	E	Metazoa	
56470258	Ent.his	<i>Entamoeba histolytica</i>	E	Entamoebidae	long
this study	Ent.inv	<i>Entamoeba invadens</i>	E	Entamoebidae	long
this study	Ent.mos	<i>Entamoeba moshkovskii</i>	E	Entamoebidae	long
this study	Ent.ter	<i>Entamoeba terrapinae</i>	E	Entamoebidae	long
29375097	Ent.fa2	<i>Enterococcus faecalis</i> V583	B	Low G+C Gram positives	
22990925	Ent.fa1	<i>Enterococcus faecium</i>	B	Low G+C Gram positives	
49610791	Erw.car	<i>Erwinia carotovora</i> subsp. atroseptica SCRI1043	B	$\gamma$ -Proteobacteria	
20150540	Esc.co2	<i>Escherichia coli</i>	B	$\gamma$ -Proteobacteria	
3005594	Esc.co3	<i>Escherichia coli</i>	B	$\gamma$ -Proteobacteria	100,0% identity with Esc.co2
16131033	Esc.co1	<i>Escherichia coli</i> K12	B	$\gamma$ -Proteobacteria	
46113349	Exi.sp2	<i>Exiguobacterium</i> sp. 255-15	B	Low G+C Gram positives	
19704478	Fus.nuc	<i>Fusobacterium nucleatum</i> subsp. nucleatum ATCC 25586	B	Fusobacteria	
50747027	Gal.ga1	<i>Gallus gallus</i>	E	Metazoa	95,2% identity with Hom.sa2
57525040	Gal.ga2	<i>Gallus gallus</i>	E	Metazoa	
53130328	Gal.ga3	<i>Gallus gallus</i>	E	Metazoa	100,0% identity with Gal.ga2

56420811	Geo.kau	<i>Geobacillus kaustophilus</i> HTA426	B	Low G+C Gram positives	
6831555	Gia.la1	<i>Giardia lamblia</i>	E	Diplomonads	
6831556	Gia.la2	<i>Giardia lamblia</i>	E	Diplomonads	
6090573	Gia.la3	<i>Giardia lamblia</i>	E	Diplomonads	99,1% identity with Gia.la3
42548307	Gib.ze1	<i>Gibberella zeae</i> PH-1	E	Fungi	
42555043	Gib.ze2	<i>Gibberella zeae</i> PH-1	E	Fungi	
33152849	Hae.duc	<i>Haemophilus ducreyi</i> 35000HP	B	$\gamma$ -Proteobacteria	
16272108	Hae.inf	<i>Haemophilus influenzae</i> Rd	B	$\gamma$ -Proteobacteria	
32030023	Hae.som	<i>Haemophilus somnus</i> 2336	B	$\gamma$ -Proteobacteria	
13027378	Hom.sa1	<i>Homo sapiens</i>	E	Metazoa	
15930199	Hom.sa2	<i>Homo sapiens</i>	E	Metazoa	
16550419	Hom.sa3	<i>Homo sapiens</i>	E	Metazoa	98,7% identity with Hom.sa2
19923881	Hom.sa4	<i>Homo sapiens</i>	E	Metazoa	99,1% identity with Hom.sa2
58338189	Lac.aci	<i>Lactobacillus acidophilus</i> NCFM	B	Low G+C Gram positives	
62513852	Lac.cas	<i>Lactobacillus casei</i> ATCC 334	B	Low G+C Gram positives	
23003825	Lac.gas	<i>Lactobacillus gasserii</i>	B	Low G+C Gram positives	
42519872	Lac.joh	<i>Lactobacillus johnsonii</i> NCC 533	B	Low G+C Gram positives	99,1% identity with Lac.gas
28377158	Lac.pla	<i>Lactobacillus plantarum</i> WCFS1	B	Low G+C Gram positives	
15673555	Lac.lac	<i>Lactococcus lactis</i> subsp. lactis	B	Low G+C Gram positives	
50955485	Lei.xyl	<i>Leifsonia xylis</i> subsp. xylis str. CTCB07	B	High G+C Gram positives	
Sanger	Lei.maj	<i>Leishmania major</i>	E	Euglenozoa	
23024189	Leu.mes	<i>Leuconostoc mesenteroides</i> subsp. mesenteroides ATCC 8293	B	Low G+C Gram positives	
16799949	Lis.in1	<i>Listeria innocua</i>	B	Low G+C Gram positives	
16800025	Lis.in2	<i>Listeria innocua</i>	B	Low G+C Gram positives	98,7% identity with Lis.mo2
16801514	Lis.in3	<i>Listeria innocua</i>	B	Low G+C Gram positives	
16802918	Lis.mo1	<i>Listeria monocytogenes</i> EGD-e	B	Low G+C Gram positives	
16802997	Lis.mo2	<i>Listeria monocytogenes</i> EGD-e	B	Low G+C Gram positives	
16804396	Lis.mo3	<i>Listeria monocytogenes</i> EGD-e	B	Low G+C Gram positives	
38102084	Mag.gri	<i>Magnaporthe grisea</i> 70-15	E	Fungi	
23012890	Mag.mag	<i>Magnetospirillum magnetotacticum</i>	B	$\alpha$ -Proteobacteria	
52308312	Man.suc	<i>Mannheimia succiniciproducens</i> MBEL55E	B	$\gamma$ -Proteobacteria	
this study	Mas.bal	<i>Mastigamoeba balamuthi</i>	E	Pelobionts	fragment missing 128 of 229 aligned positions – 54,4% identity with Ent.inv; long version missing 438 of 562 aligned positions
3122434	Mes.aur	<i>Mesocricetus auratus</i>	E	Metazoa	96,9% identity with Hom.sa1
45914196	Mes.spB	<i>Mesorhizobium</i> sp. BNC1	B	$\alpha$ -Proteobacteria	
12644084	Mus.mu1	<i>Mus musculus</i>	E	Metazoa	96,9% identity with Hom.sa1
12853907	Mus.mu2	<i>Mus musculus</i>	E	Metazoa	97,4% identity with Hom.sa2
19882233	Mus.mu3	<i>Mus musculus</i>	E	Metazoa	97,4% identity with Hom.sa2
5353761	Mus.mu4	<i>Mus musculus</i>	E	Metazoa	97,4% identity with Hom.sa1
6754030	Mus.mu5	<i>Mus musculus</i>	E	Metazoa	96,1% identity with Hom.sa1
54020607	Myc.hyo	<i>Mycoplasma hyopneumoniae</i> 232	B	Low G+C Gram positives	
42561343	Myc.my1	<i>Mycoplasma mycoides</i> subsp. mycoides SC str. PG1	B	Low G+C Gram positives	
42560683	Myc.my2	<i>Mycoplasma mycoides</i> subsp. mycoides SC str. PG1	B	Low G+C Gram positives	
26553629	Myc.pen	<i>Mycoplasma penetrans</i>	B	Low G+C Gram positives	
15828833	Myc.pul	<i>Mycoplasma pulmonis</i>	B	Low G+C Gram positives	
32406955	Neu.cra	<i>Neurospora crassa</i>	E	Fungi	
23127811	Nos.pun	<i>Nostoc punctiforme</i>	B	Cyanobacteria	
17228222	Nos.sp	<i>Nostoc</i> sp. PCC 7120	B	Cyanobacteria	
23098066	Oce.ihe	<i>Oceanobacillus iheyensis</i> HTE831	B	Low G+C Gram positives	
23038068	Oen.oen	<i>Oenococcus oeni</i> MCW	B	Low G+C Gram positives	
Genoscope	Par.tet	<i>Paramecium tetraurelia</i>	E	Ciliates	long
15602740	Pas.mul	<i>Pasteurella multocida</i>	B	$\gamma$ -Proteobacteria	
48870313	Ped.pen	<i>Pediococcus pentosaceus</i> ATCC 25745	B	Low G+C Gram positives	
54302040	Pho.pr1	<i>Photobacterium profundum</i> SS9	B	$\gamma$ -Proteobacteria	
54308224	Pho.pr2	<i>Photobacterium profundum</i> SS9	B	$\gamma$ -Proteobacteria	
37525283	Pho.lum	<i>Photorhabdus luminescens</i> subsp. laumondii TTO1	B	$\gamma$ -Proteobacteria	
55730197	Pon.pyg	<i>Pongo pygmaeus</i>	E	Metazoa	99,6% identity with Hom.sa1
34540585	Por.gi1	<i>Porphyromonas gingivalis</i> W83	B	Bacteroidetes/Chlorobi group	
34540986	Por.gi2	<i>Porphyromonas gingivalis</i> W83	B	Bacteroidetes/Chlorobi group	long
33863691	Pro.mar	<i>Prochlorococcus marinus</i> str. MIT 9313	B	Cyanobacteria	
50841959	Pro.acn	<i>Propionibacterium acnes</i> KPA171202	B	High G+C Gram positives	
27695235	Rat.nor	<i>Rattus norvegicus</i>	E	Metazoa	97,8% identity with Hom.sa2

32471377	Rho.ba1	<i>Rhodopirellula baltica</i> SH 1	B	Planctomycetes	
32472019	Rho.ba2	<i>Rhodopirellula baltica</i> SH 1	B	Planctomycetes	
32472456	Rho.ba3	<i>Rhodopirellula baltica</i> SH 1	B	Planctomycetes	
32472581	Rho.ba4	<i>Rhodopirellula baltica</i> SH 1	B	Planctomycetes	long
46107074	Rub.xyl	<i>Rubrobacter xylanophilus</i> DSM 9941	B	High G+C Gram positives	
16759628	Sal.ent	<i>Salmonella enterica</i> subsp. enterica serovar Typhi	B	$\gamma$ -Proteobacteria	95,6% identity with Esc.co2
16764054	Sal.typ	<i>Salmonella typhimurium</i> LT2	B	$\gamma$ -Proteobacteria	95,6% identity with Esc.co2
56756765	Sch.jap	<i>Schistosoma japonicum</i>	E	Metazoa	
24112036	Shi.fl1	<i>Shigella flexneri</i> 2a str. 301	B	$\gamma$ -Proteobacteria	99,6% identity with Esc.co2
24114429	Shi.fl2	<i>Shigella flexneri</i> 2a str. 301	B	$\gamma$ -Proteobacteria	98,6% identity with Esc.co1
27982841	Spi.ba1	<i>Spironucleus barkhanus</i>	E	Diplomonads	fragment missing 107 of 229 aligned positions – 68,0% identity with Spi.ba2
27983038	Spi.ba2	<i>Spironucleus barkhanus</i>	E	Diplomonads	
15923559	Sta.aur	<i>Staphylococcus aureus</i> subsp. aureus Mu50	B	Low G+C Gram positives	
27467258	Sta.epi	<i>Staphylococcus epidermidis</i> ATCC 12228	B	Low G+C Gram positives	
22536963	Str.aga	<i>Streptococcus agalactiae</i> 2603V/R	B	Low G+C Gram positives	
24379109	Str.mut	<i>Streptococcus mutans</i> UA159	B	Low G+C Gram positives	
15901269	Str.pne	<i>Streptococcus pneumoniae</i> TIGR4	B	Low G+C Gram positives	
15675320	Str.pyo	<i>Streptococcus pyogenes</i>	B	Low G+C Gram positives	
50591240	Str.sui	<i>Streptococcus suis</i> 89/1591	B	Low G+C Gram positives	
62526838	Str.the	<i>Streptococcus thermophilus</i> LMD-9	B	Low G+C Gram positives	
29829560	Str.ave	<i>Streptomyces avermitilis</i> MA-4680	B	High G+C Gram positives	
21223604	Str.coe	<i>Streptomyces coelicolor</i> A3(2)	B	High G+C Gram positives	
51891744	Sym.the	<i>Symbiobacterium thermophilum</i> IAM 14863	B	High G+C Gram positives	
56750455	Syn.elo	<i>Synechococcus elongatus</i> PCC 6301	B	Cyanobacteria	
33865076	Syn.sp	<i>Synechococcus</i> sp. WH 8102	B	Cyanobacteria	
TIGR	Tet.the	<i>Tetrahymena thermophila</i>	E	Ciliates	long
47222915	Tet.nig	<i>Tetraodon nigroviridis</i>	E	Metazoa	
20807502	The.ten	<i>Thermoanaerobacter tengcongensis</i>	B	Low G+C Gram positives	
42525853	Tre.den	<i>Treponema denticola</i> ATCC 35405	B	Spirochetes	
23043712	Tri.ery	<i>Trichodesmium erythraeum</i> IMS101	B	Cyanobacteria	
this study	Tri.vag	<i>Trichomonas vaginalis</i>	E	Parabasalids	long
TIGR	Try.bru	<i>Trypanosoma brucei</i>	E	Euglenozoa	
TIGR	Try.cru	<i>Trypanosoma cruzi</i>	E	Euglenozoa	
46097301	Usi.may	<i>Ustilago maydis</i> 521	E	Fungi	
15601778	Vib.cho	<i>Vibrio cholerae</i> O1 biovar eltor str. N16961	B	$\gamma$ -Proteobacteria	
59712964	Vib.fis	<i>Vibrio fischeri</i> ES114	B	$\gamma$ -Proteobacteria	
28899893	Vib.par	<i>Vibrio parahaemolyticus</i> RIMD 2210633	B	$\gamma$ -Proteobacteria	
27367580	Vib.vul	<i>Vibrio vulnificus</i> CMCP6	B	$\gamma$ -Proteobacteria	
38014747	Xen.la1	<i>Xenopus laevis</i>	E	Metazoa	
38014417	Xen.la2	<i>Xenopus laevis</i>	E	Metazoa	
49647197	Yar.lip	<i>Yarrowia lipolytica</i> CLIB99	E	Fungi	
16122840	Yer.pes	<i>Yersinia pestis</i>	B	$\gamma$ -Proteobacteria	
51595466	Yer.pse	<i>Yersinia pseudotuberculosis</i> IP 32953	B	$\gamma$ -Proteobacteria	100,0% identity with Yer.pes

---

## The alcohol dehydrogenase E dataset.

gi	Short	Species	Domain	Group	Comment
46143771	Act.ple	<i>Actinobacillus pleuropneumoniae</i> serovar 1 str. 4074	B	$\gamma$ -Proteobacteria	
30264441	Bac.ant	<i>Bacillus anthracis</i> str. Ames	B	Low G+C Gram positives	97,0% identity with Bac.cer
30022446	Bac.cer	<i>Bacillus cereus</i> ATCC 14579	B	Low G+C Gram positives	
52005783	Bac.lic	<i>Bacillus licheniformis</i> ATCC 14580	B	Low G+C Gram positives	
49476825	Bac.thu	<i>Bacillus thuringiensis</i> serovar konkukian str. 97-27	B	Low G+C Gram positives	96,9% identity with Bac.thu
23336499	Bif.lon	<i>Bifidobacterium longum</i> DJO10A	B	High G+C Gram positives	
JGI	Chl.rei	<i>Chlamydomonas reinhardtii</i>	E	Green algae	
34496592	Chr.vio	<i>Chromobacterium violaceum</i> ATCC 12472	B	$\beta$ -Proteobacteria	
15004739	Clo.ac1	<i>Clostridium acetobutylicum</i>	B	Low G+C Gram positives	
15004865	Clo.ac2	<i>Clostridium acetobutylicum</i>	B	Low G+C Gram positives	
18311513	Clo.per	<i>Clostridium perfringens</i>	B	Low G+C Gram positives	
28211045	Clo.tet	<i>Clostridium tetani</i> E88	B	Low G+C Gram positives	
23020357	Clo.the	<i>Clostridium thermocellum</i> ATCC 27405	B	Low G+C Gram positives	
54659250	Cry.hom	<i>Cryptosporidium hominis</i>	E	Apicomplexa	100,0% identity with Cry.par
46228815	Cry.par	<i>Cryptosporidium parvum</i>	E	Apicomplexa	
56473007	Ent.hi1	<i>Entamoeba histolytica</i> HM-1:IMSS	E	Entamoebidae	
56467638	Ent.hi2	<i>Entamoeba histolytica</i> HM-1:IMSS	E	Entamoebidae	99,9% identity with Ent.hi1
56465626	Ent.hi3	<i>Entamoeba histolytica</i> HM-1:IMSS	E	Entamoebidae	99,9% identity with Ent.hi1
56472980	Ent.hi4	<i>Entamoeba histolytica</i> HM-1:IMSS	E	Entamoebidae	99,9% identity with Ent.hi1
this study	Ent.inv	<i>Entamoeba invadens</i>	E	Entamoebidae	
this study	Ent.mos	<i>Entamoeba moshkovskii</i>	E	Entamoebidae	
this study	Ent.ter	<i>Entamoeba terrapinae</i>	E	Entamoebidae	
29375484	Ent.fas	<i>Enterococcus faecalis</i> V583	B	Low G+C Gram positives	
22992835	Ent.fam	<i>Enterococcus faecium</i>	B	Low G+C Gram positives	
49611780	Erw.car	<i>Erwinia carotovora</i> subsp. atroseptica SCR11043	B	$\gamma$ -Proteobacteria	
15801467	Esc.col	<i>Escherichia coli</i> O157:H7 EDL933	B	$\gamma$ -Proteobacteria	
45531125	Exi.sp2	<i>Exiguobacterium</i> sp. 255-15	B	Low G+C Gram positives	
2052472	Gia.la1	<i>Giardia intestinalis</i>	E	Diplomonads	
29247272	Gia.la2	<i>Giardia lamblia</i> ATCC 50803	E	Diplomonads	97,3% identity with Gia.la1
58336794	Lac.aci	<i>Lactobacillus acidophilus</i> NCFM	B	Low G+C Gram positives	
23002679	Lac.gas	<i>Lactobacillus gasseri</i>	B	Low G+C Gram positives	97,2% identity with Lac.joh
42519642	Lac.joh	<i>Lactobacillus johnsonii</i> NCC 533	B	Low G+C Gram positives	
28379962	Lac.pla	<i>Lactobacillus plantarum</i> WCFS1	B	Low G+C Gram positives	
15674137	Lac.la1	<i>Lactococcus lactis</i> subsp. lactis	B	Low G+C Gram positives	
27527171	Leu.mes	<i>Leuconostoc mesenteroides</i> subsp. cremoris	B	Low G+C Gram positives	
16800743	Lis.inn	<i>Listeria innocua</i>	B	Low G+C Gram positives	99,4% identity with Lis.mon
45549371	Lis.iva	<i>Listeria ivanovii</i>	B	Low G+C Gram positives	98,6% identity with Lis.mon
16803674	Lis.mon	<i>Listeria monocytogenes</i> EGD-e	B	Low G+C Gram positives	
45549373	Lis.see	<i>Listeria seeligeri</i>	B	Low G+C Gram positives	99,2% identity with Lis.mon
45549375	Lis.wel	<i>Listeria welshimeri</i>	B	Low G+C Gram positives	98,2% identity with Lis.mon
52308297	Man.suc	<i>Mannheimia succiniciproducens</i> MBEL55E	B	$\gamma$ -Proteobacteria	
21435953	Mas.bal	<i>Mastigamoeba balamuthi</i>	E	Pelobionts	
23037201	Oen.oen	<i>Oenococcus oeni</i> MCW	B	Low G+C Gram positives	
15603318	Pas.mul	<i>Pasteurella multocida</i>	B	$\gamma$ -Proteobacteria	
54308296	Pho.pro	<i>Photobacterium profundum</i> SS9	B	$\gamma$ -Proteobacteria	
37526388	Pho.lum	<i>Photorhabdus luminescens</i> subsp. laumondii TTO1	B	$\gamma$ -Proteobacteria	
33578055	Pir.sp	<i>Piromyces</i> sp. E2	E	Fungi	
40644910	Pol.sp.	<i>Polytomella</i> sp. Pringsheim 198.80	E	Green algae	
16760134	Sal.ent	<i>Salmonella enterica</i> subsp. enterica serovar Typhi	B	$\gamma$ -Proteobacteria	98,0% identity with Esc.col
16765093	Sal.typ	<i>Salmonella typhimurium</i> LT2	B	$\gamma$ -Proteobacteria	98,1% identity with Esc.col
24373696	She.one	<i>Shewanella oneidensis</i> MR-1	B	$\gamma$ -Proteobacteria	
24112636	Shi.fle	<i>Shigella flexneri</i> 2a str. 301	B	$\gamma$ -Proteobacteria	100,0% identity with Esc.col
27983190	Spi.bar	<i>Spironucleus barkhanus</i>	E	Diplomonads	
15923138	Sta.aur	<i>Staphylococcus aureus</i> subsp. aureus Mu50	B	Low G+C Gram positives	
27467424	Sta.epi	<i>Staphylococcus epidermidis</i> ATCC 12228	B	Low G+C Gram positives	
22536238	Str.aga	<i>Streptococcus agalactiae</i> 2603V/R	B	Low G+C Gram positives	95,0% identity with Str.pyo
34787345	Str.bov	<i>Streptococcus bovis</i>	B	Low G+C Gram positives	
24378663	Str.mut	<i>Streptococcus mutans</i> UA159	B	Low G+C Gram positives	
15901847	Str.pne	<i>Streptococcus pneumoniae</i> TIGR4	B	Low G+C Gram positives	
21909572	Str.pyo	<i>Streptococcus pyogenes</i> MGAS315	B	Low G+C Gram positives	
50590075	Str.sui	<i>Streptococcus suis</i> 89/1591	B	Low G+C Gram positives	



22297771	The.elo	<i>Thermosynechococcus elongatus</i> BP-1	B	Cyanobacteria	
58578286	unc.bac	<i>uncultured bacterium</i>	B		
15642035	Vib.cho	<i>Vibrio cholerae</i> O1 biovar eltor str. N16961	B	$\gamma$ -Proteobacteria	
59711525	Vib.fis	<i>Vibrio fischeri</i> ES114	B	$\gamma$ -Proteobacteria	
28898895	Vib.par	<i>Vibrio parahaemolyticus</i> RIMD 2210633	B	$\gamma$ -Proteobacteria	95,1% identity with Vib.cho
37679359	Vib.vul	<i>Vibrio vulnificus</i> YJ016	B	$\gamma$ -Proteobacteria	96,9% identity with Vib.cho
16122410	Yer.pes	<i>Yersinia pestis</i>	B	$\gamma$ -Proteobacteria	

---