

Etymology of candidate phyla

Candidatus Dormibacteraeota (AD3) was named based on Bin13 (Ga0137693). Dor.mi.bac.ter.ae.o'ta: N.L. masc. n. *Dormibacter* a bacterial genus; suff. -aeota ending denoting a phylum; N.L. neut. pl. n. *Dormibacteraeota* the *Dormibacter* phylum. *Dormibacter* (Dor.mi.bac'ter): L. v. *dormio* to sleep, N. L. masc. n. *Dormibacter* a dormant rod. *Candidatus* Eremiobacteraeota (WPS-2) was named based on Bin22 (Ga0117865). E.re.mi.o.bac.ter.ae.o'ta: N.L. masc. n. *Eremiobacter* a bacterial genus; suff. -aeota ending denoting a phylum; N.L. pl. neut. n. *Eremiobacteraeota* the *Eremiobacter* phylum. *Eremiobacter* (E.re.mi.o.bac'ter): Gr. fem. n. *eremia* desert, wilderness; N.L. masc. n. *bacter* a rod; N.L. masc. n. *Eremiobacter* a rod from a desert.

Supplementary Table 1. Geographical, physical and chemical parameters of the Robinson Ridge and Adams Flat soil samples.

Soil Number	Latitude (Deg N)	Longitude (Deg E)	Elevation (m)	Aspect (Deg)	TC (% w/w)	TKN [†] (mg/kg)	TKP [†] (mg/kg)	Dry Matter Fraction	pH
RR 1	-66.3695	110.5856	45.94	29.46	0.13	160	1300	0.95997	5.4
RR 2	-66.3695	110.5856	45.85	27.67	0.14	120	1200	0.94414	5.9
RR 3	-66.3695	110.5856	46.05	28.22	0.24	230	1000	0.96303	5.1
AF 1	-68.5502	78.02106	1.76	310.10	0.093	88	960	0.99550	7.7
AF 2	-68.5502	78.02106	1.78	310.10	0.097	76	760	0.99768	8.2
AF 3	-68.5502	78.02106	1.69	310.10	0.081	84	730	0.99409	6.6
	Conductivity (uS/cm) at 25°C	NO ₂ ^{-†} (mg/kg)	Br ^{-†} (mg/kg)	NO ₃ ^{-†} (mg/kg)	PO ₄ ^{3-†} (mg/kg)	SO ₄ ^{2-*} (mg/kg)	NH ₄ ^{+†} (mg/kg)	Cl ⁻ (ppm)	Cl ⁻ (mg/kg)
RR 1	9.1	<0.15	<0.15	<0.76	5.0	2.1	0.55	75	3.7
RR 2	9.2	<0.15	<0.15	<0.76	5.6	1.9	<0.55	87	4.9
RR 3	12.4	<0.15	<0.15	<0.76	3.9	2.5	0.73	93	5.8
AF 1	1590	<0.15	1.40	12.47	18.22	547.2	<0.55	3869	<0.15
AF 2	944	<0.15	0.98	6.52	1.76	204.5	<0.55	950	<0.15
AF 3	230	<0.15	<0.15	0.76	<0.15	137.5	<0.55	608	214.3
	% Mud 63 µm	% Sand 63-2000 µm	% Gravel > 2 mm	Minimum (µm)	Maximum (µm)	Mean (µm)	Quartile Deviation (PQD)	Sorting Coefficient	Kurtosis
RR 1	5.1	77.2	17.7	30.0	1061.4	338.1	1.1	0.48	0.87
RR 2	11.6	68.8	19.6	44.6	1057.3	327.8	1.2	0.43	0.85
RR 3	4.8	76.2	19.0	31.3	1255.1	418.9	1.2	0.44	0.83
AF 1	5.8	85.6	8.7	13.8	1148.4	294.9	1.1	0.47	1.19
AF 2	3.4	89.3	7.4	23.6	1432.8	441.3	1.2	0.45	1.04
AF 3	3.3	87.8	8.9	15.4	1343.6	391.1	1.4	0.39	0.98
	SiO ₂ ^{**} (%)	TiO ₂ ^{**} (%)	Al ₂ O ₃ ^{**} (%)	Fe ₂ O ₃ ^{**} (%)	MnO ^{**} (%)	MgO ^{**} (%)	CaO ^{**} (%)	Na ₂ O ^{**} (%)	SO ₃ ^{**} (%)
RR 1	69.5	0.7	13.3	4.6	0.10	1.3	3.2	2.9	0.01
RR 2	67.0	1.0	13.6	6.0	0.13	1.6	3.9	2.9	0.01
RR 3	67.6	0.9	13.3	5.8	0.12	1.6	3.5	2.9	0.01
AF 1	64.1	0.7	14.4	14.4	0.10	3.7	5.0	3.7	0.10
AF 2	64.9	0.7	14.1	14.1	0.11	3.8	5.3	3.5	0.03
AF 3	64.9	0.7	14.3	14.3	0.10	3.5	5.0	3.5	0.02

* Dry matter basis, † Water extracted chemicals, ‡ Extracted with 2M KCl, **X-ray fluorescence used to obtain data

Supplementary Table 2. Sequence statistics of the Robinson Ridge metagenome.

	Total count	Proportion of reads
Sequence reads (paired)	336,45,878	
Reads after quality filtering (paired)	31,562,291	93.8%
Number of contigs	157,031	
Number of contigs (> = 1kb)	61,711	
Longest contig	462,793 bp	
Shortest contig	500 bp	
Total genome size	264,245,447 bp	
GC content	62.89%	
N50 contig size	157,031 bp	
Coverage (25913)	2.89	
Coverage (25914)	2.58	
Coverage (25919)	3.18	
Predicted genes (Prokka)	208,343	
Genes with predicted functions	112,716	54%
Genes assigned with KEGG IDs	63,244	30%
Genes assigned with COG IDs	71,413	34%

Supplementary Table 3. Alpha diversity metrics of Robinson Ridge and Adams Flat communities based on 16S rRNA gene sequencing.

Soil number	Number of OTUs	Chao1 richness	Shannon diversity
RR 1	511	966.9	4.9
RR 2	411	673.9	5.0
RR 3	424	751.2	4.9
AF 1	279	340.0	3.7
AF 2	384	517.2	3.9
AF 3	532	694.0	4.2

Supplementary Table 4. Sequence statistics of Robinson Ridge draft genomes.

Bin ID	Phylogeny (Phylum; Order)	Relative abundance* (%)	Completeness† (%)	Contamination† (%)	16S	Genome size (Mb)	Contigs (#)	N50 (bp)	Genes‡ (#)	GC content (%)
1	Actinobacteria; Acidimicrobiales	0.5 - 2.1	93.1	5.6	+	2.3	247	13870	2560	68.3
2	Actinobacteria	1.1 - 3.0	94.1	1.1	+	2.4	91	39959	2624	59.6
3	Actinobacteria; Pseudonocardiales	0.3 - 4.8	84.9	2.1	+	3.0	93	49695	3098	70.2
4	Actinobacteria; Pseudonocardiales	0.5 - 4.7	83.5	5.3		3.7	228	24463	3900	69.1
5	Actinobacteria; Pseudonocardiales	1.3 - 4.1	81.0	6.6		4.6	462	14183	5070	68.4
6	Actinobacteria; Pseudonocardiales	0.0 - 3.2	77.9	3.1		3.9	437	10907	4295	70.5
7	Actinobacteria; Pseudonocardiales	8.8 - 10.0	91.8	3.0		4.6	379	20402	5155	67.4
8	Actinobacteria; Pseudonocardiales	0.2 - 12.5	89.5	3.4		5.4	608	15960	5966	67.5
9	Actinobacteria; Pseudonocardiales	0.6 - 1.7	67.5	9.8		3.7	747	6168	4526	68.2
10	Actinobacteria; Solirubrobacterales	2.1 - 3.6	95.3	0.9	+	2.7	100	52627	2729	69.6
11	Actinobacteria; Solirubrobacterales	1.7 - 3.0	87.3	6.7		3.0	307	16457	3358	70.1
12	Dormibacteraeota (AD3)	0.3 - 3.0	95.3	2.3	+	3.0	302	23210	3231	69.2
13	Dormibacteraeota (AD3)	1.7 - 4.9	96.3	0.0	+	3.0	71	62854	3135	66.7
14	Dormibacteraeota (AD3)	1.9 - 4.0	92.4	4.6	+	5.3	358	25215	5615	68.3
15	Proteobacteria; Rhizobiales	1.7 - 2.8	90.0	1.4		3.6	429	10900	4096	59.8
16	Chloroflexi	1.3 - 5.2	55.8	0.0		2.7	523	5651	2592	60.5
17	Chloroflexi	0.3 - 2.0	84.8	5.0		3.7	440	12432	4203	61.6
18	Chloroflexi	1.0 - 2.6	91.7	0.5	+	2.2	48	98473	2218	69.6
19	Thaumarchaeota	1.6 - 2.6	98.1	3.9	+	4.0	278	22997	5405	38.9
20	Verrucomicrobia; Chthoniobacterales	0.7 - 1.8	92.9	5.1	+	3.6	173	31661	3513	58.4
21	Verrucomicrobia; Chthoniobacterales	0.0 - 1.7	66.0	2.3	+	2.5	422	6693	2825	58.7
22	Eremiobacteraeota (WPS-2)	1.9 - 2.6	99.1	0.9	+	2.6	19	214289	2529	62.6
23	Eremiobacteraeota (WPS-2)	0.3 - 2.7	94.9	0.0	+	1.9	43	74679	1988	60.1

* Based on lineage-specific marker sets following analysis with CheckM.

† The proportion of reads mapped to each genome was used to estimate relative abundance.

‡ Inferred from Prodigal

Supplementary Table 5. Key KEGG genes identified in the Robinson Ridge draft genomes.

KEGG ID	Function	AD3	Thaum archaeota	Actino bacteria	WPS -2	Chloro flexi	Proteo bacteria	Verruco microbia	Meta genome
Carbon fixation									
K03520	Hydrogenase large subunit	2	0	5	1	1	0	1	23
K06282	Hydrogenase small subunit	2	0	7	1	1	0	1	21
K03520	Carbon monoxide dehydrogenase large subunit	1	0	2	0	0	0	0	14
K01601	RuBisCO large subunit	2	0	5	1	0	0	0	19
K01602	RuBisCO small subunit	2	0	6	1	0	0	0	16
Energy metabolism									
Various	Type I NADH dehydrogenase	3	1	11	2	3	1	2	N/A
Various	Succinate dehydrogenase / fumarate reductase	3	1	10	2	3	1	2	N/A
Various	Cytochrome <i>aa</i> ₃ oxidase	3	1	11	2	3	1	2	N/A
Various	Cytochrome <i>cbb</i> ₃ oxidase	0	0	0	1	1	0	1	N/A
Various	Cytochrome <i>bd</i> oxidase	2	0	6	1	1	1	0	N/A
Various	F ₁ F _o -type ATPase	3	0	11	2	3	1	2	N/A
Nitrogen cycling									
K02588	<i>nifH</i> , nitrogen fixation	0	0	0	0	0	0	0	0
K00367	<i>narB</i> , assimilatory nitrate reductase	0	0	1	0	0	0	0	2
K00372	<i>nasA</i> , assimilatory nitrate reductase	0	0	1	0	0	0	0	14
K00366	<i>nirA</i> , assimilatory nitrite reductase	1	0	2	0	0	0	0	3
K00362	<i>nirB</i> , assimilatory nitrite reductase	0	0	7	0	0	1	0	19
K00370	<i>narG</i> , dissimilatory nitrate reductase	0	0	1	0	0	0	0	7
K00368	<i>nirK</i> , dissimilatory nitrite reductase	0	3	0	1	0	1	2	14
K00376	<i>nosZ</i> , nitrous oxide reductase	0	1	0	0	0	0	0	1
K10945	<i>amoC</i> , ammonia monooxygenase	0	1	0	0	0	0	0	1
Carbohydrate metabolism									
K01179	Endoglucanase	0	0	5	0	4	0	0	19
K05349	Beta-glucosidase	6	0	1	0	0	1	0	29
K05350	Beta-glucosidase	4	1	2	0	2	0	0	9
K00700	1,4-alpha-glucan branching enzyme	5	0	11	1	2	0	0	33
K16149	1,4-alpha-glucan branching enzyme	0	0	0	0	0	0	0	12
K00688	Starch phosphorylase	3	0	7	0	1	0	2	43
K01176	Alpha-amylase	0	0	2	0	0	0	0	5
K01178	Glucoamylase	1	0	2	0	0	1	3	17
K01196	Glycogen debranching enzyme	0	0	0	0	0	0	0	12
K00705	4-alpha-glucanotransferase	3	0	6	0	0	0	1	21
K12047	Maltase-glucoamylase	0	0	0	0	0	0	0	0
K01187	Alpha-glucosidase	3	0	11	0	2	0	0	36
K15922	Alpha-glucosidase	0	0	0	0	0	0	0	3
K01051	Pectinesterase	0	0	0	0	0	0	1	1
K01198	Xylan 1,4-beta-xylosidase	1	0	0	0	1	0	1	3
K01183	Chitinase	0	0	1	0	0	0	0	4
Psychrotolerance									
K00697	Trehalose 6-phosphate synthase	3	0	10	0	1	1	1	31
K16055	Trehalose 6-phosphate synthase/phosphatase	0	0	1	0	0	0	0	4
K01087	Trehalose 6-phosphate phosphatase	3	0	6	0	1	1	0	18
K05343	Maltose alpha-D-glucosyltransferase / alpha-amylase, <i>treS</i>	3	0	11	0	3	2	1	43
K02438	Glycogen operon protein, <i>treX</i>	3	0	16	0	1	0	2	40
K06044	(1->4)-alpha-D-glucan 1-alpha-D-glucosylmutase, <i>treY</i>	4	0	9	0	0	0	2	27
K01236	Maltotrioglucosyl trehalose trehalohydrolase, <i>treZ</i>	3	0	10	0	0	0	2	28
K05342	Alpha, alpha-trehalose phosphorylase, <i>treP</i>	0	0	0	0	0	0	0	0

* The presence of the carbon monoxide dehydrogenase large subunit, RuBisCO large subunit, group 1h [NiFe] hydrogenase large subunit, and ammonia monooxygenase subunit C were confirmed by phylogenetic placement, while the rest were inferred from KEGG annotation. The presence of NADH dehydrogenase (k00330, K00331, k00332, k00333, k00334, k00335, k00336, k00337, k00338, k00339, k00340, k00341, k00342, k00343), succinate dehydrogenase/fumarate reductase (k00241, k00242, k00239, k00240, k00244, k00245, k00246, k00247), cytochrome *aa*₃ oxidase (k02274, k02275, k02276, k02277), cytochrome *cbb*₃ oxidase (k00404, k00405), cytochrome *bd* oxidase (k00425, k00426) and F₁F_o-ATPase (k02111, k02112, k02115, k02113, k02114, k02108, k02109, k02110) were determined by the presence of at least one subunit, the rest shown were the number of KEGG annotations identified.

Supplementary Table 6. Number of genes predicted to be associated with carbohydrate metabolism.

Bin ID	Phylum	Glycoside hydrolases	Glycosyl-transferases	Polysaccharide lyases	Carbohydrate esterases	Auxiliary activities	Carbohydrate-binding modules
1	Actinobacteria	10	35	0	4	0	7
2	Actinobacteria	7	17	1	5	1	2
3	Actinobacteria	16	29	2	10	2	13
4	Actinobacteria	26	29	1	7	1	10
5	Actinobacteria	26	33	0	17	0	12
6	Actinobacteria	21	20	1	14	4	13
7	Actinobacteria	19	30	0	14	5	7
8	Actinobacteria	25	31	0	14	5	8
9	Actinobacteria	13	22	0	10	2	6
10	Actinobacteria	11	34	0	14	0	4
11	Actinobacteria	14	48	1	9	0	3
12	Dormibacteraeota (AD3)	20	36	0	10	0	5
13	Dormibacteraeota (AD3)	16	32	0	7	1	12
14	Dormibacteraeota (AD3)	54	65	0	15	4	12
15	Proteobacteria	13	26	1	6	0	1
16	Chloroflexi	14	36	0	5	0	11
17	Chloroflexi	28	54	0	8	0	5
18	Chloroflexi	22	24	0	10	0	4
19	Thaumarchaeota	9	15	0	14	1	7
20	Verrucomicrobia	18	53	0	16	0	10
21	Verrucomicrobia	16	42	0	11	0	4
22	Eremiobacteraeota (WPS-2)	18	40	0	11	1	7
23	Eremiobacteraeota (WPS-2)	13	26	0	7	0	4

Gene numbers based on dbCAN search results with sequence coverage greater than 80%.

Supplementary Table 7. Protein coding sequences from the 3-hydroxypropionate carbon fixation pathway within the Thaumarchaeota draft genome (Bin19).

Putative genes involved in carbon fixation	NCBI accession	<i>Nitrososphaera evergladensis</i> SR1	<i>Nitrososphaera viennensis</i> EN7	<i>Nitrososphaera gargensis</i> Ga9.2	<i>Nitrosoarchaeum limnia</i> SFB1	<i>Nitrosopumilus koreensis</i> MY1
Biotin-(acetyl-CoA-carboxylase) ligase	AIF84040	43%	45%	45%	35%	38%
Acetyl-CoA carboxylase, carboxyltransferase subunit	AIF84813	80%	80%	80%	66%	65%
Acetyl-CoA carboxylase, biotin carboxylase subunit	AIF84814	72%	71%	71%	57%	57%
Acetyl-CoA carboxylase, biotin carboxyl carrier protein	AIF84815	53%	54%	53%	39%	39%
Methylmalonyl-CoA epimerase	AIF83180	75%	75%	79%	49%	52%
Methylmalonyl-CoA mutase N-terminal domain	AIF83181	74%	75%	77%	66%	65%
Methylmalonyl-CoA mutase C-terminal domain	AIF83184	78%	77%	78%	64%	64%
Fumarate hydratase (fumarase)	AIF82837	69%	70%	71%	61%	62%
Complex II superfamily, A subunit	AIF82619	74%	n.a.	34%	35%	35%
Complex II superfamily, A subunit	AIF82619	74%	n.a.	36%	37%	37%
Complex II superfamily, A subunit	AIF83657	78%	78%	82%	70%	70%
Complex II superfamily, B subunit	AIF83654	73%	69%	75%	54%	58%
Complex II superfamily, B subunit	AIF83654	69%	67%	71%	54%	55%
Complex II superfamily, C subunit	AIF83655	68%	70%	69%	52%	53%
Complex II superfamily, C subunit	AIF83655	71%	73%	80%	61%	60%
Complex II superfamily, C subunit	AIF83656	75%	75%	78%	51%	51%
Malonyl-CoA reductase	AIF85147	77%	76%	78%	58%	57%
3-hydroxyacyl-CoA dehydrogenase	AIF84511	71%	71%	74%	60%	59%
3-hydroxypropionyl-CoA synthetase	AIF84423	71%	71%	72%	61%	60%

Three ammonia oxidising *Nitrososphaera* and two ammonia oxidising *Nitrosopumilus* genomes were used as references. The complex II superfamily includes succinate dehydrogenase and fumarate reductase proteins, which cannot be distinguished by sequence alone.

Supplementary Table 8. Taxonomy of genes of the particulate methane monooxygenase and the soluble methane monooxygenase / ammonia monooxygenase superfamilies identified from the Robinson Ridge metagenome based on BLAST.

Gene	Predicted activity	Taxonomy	Closest protein sequence accession	Similarity
<i>amoC</i>	Soluble ammonia monooxygenase subunit	Thaumarchaeota (Bin19)	AFU59435.1	95%
<i>pmoB</i>	Particulate methane monooxygenase subunit	Alphaproteobacteria (Bin15)	WP_051953405	67%
<i>pmoB</i>	Particulate methane monooxygenase subunit	Unbinned	WP_051953405	61%
<i>pmoB</i>	Particulate methane monooxygenase subunit	Unbinned	CAJ01618	72%
<i>pmoA</i>	Particulate methane monooxygenase subunit	Alphaproteobacteria (Bin15)	CAJ01617	71%
<i>pmoA</i>	Particulate methane monooxygenase subunit	Unbinned	WP_036262482	74%
<i>pmoC</i>	Particulate methane monooxygenase subunit	Unbinned	CAJ01616	72%
<i>mmoX</i>	Soluble hydrocarbon monooxygenase subunit	Unbinned	WP_028063546	92%
<i>mmoX</i>	Soluble hydrocarbon monooxygenase subunit	Unbinned	WP_028063546	90%
<i>mmoY</i>	Soluble hydrocarbon monooxygenase subunit	Unbinned	WP_055494570	68%
<i>mmoY</i>	Soluble hydrocarbon monooxygenase subunit	Unbinned	WP_027935444	77%

Supplementary Table 9. Secondary metabolite biosynthesis gene clusters within the metagenome identified by Anti-smash.

Sequence origin	Metabolite type	Most similar known cluster	Similarity
Proteobacteria (Bin15)	Bacteriocin	Oxazolomycin biosynthetic gene cluster	6%
Proteobacteria (Bin15)	Bacteriocin	N/A	
Actinobacteria (Bin10)	Bacteriocin	N/A	
AD3 (Bin13)	Terpene	N/A	
Actinobacteria (Bin9)	Terpene	N/A	
Actinobacteria (Bin3)	Terpene	Hopene biosynthetic gene cluster	46%
Actinobacteria (Bin3)	Terpene	Isorenieratene biosynthetic gene cluster	28%
Proteobacteria (Bin15)	Terpene	N/A	
Unbinned	Butyrolactone	N/A	
Unbinned	Lantipeptide	Catenulipeptin biosynthetic gene cluster	60%
Unbinned	Lantipeptide	N/A	
Unbinned	Nonribosomal peptide	N/A	
Unbinned	Nonribosomal peptide	Mannoheptimycin biosynthetic gene cluster	7%
Unbinned	Nonribosomal peptide	N/A	
Unbinned	Polyketide	Eicosapentaenoic acid biosynthetic gene cluster	22%
Unbinned	Polyketide	N/A	
Unbinned	Polyketide	N/A	
Unbinned	Siderophore	Asukamycin biosynthetic gene cluster	4%
Unbinned	Siderophore	N/A	
Unbinned	Polyketide	BE-7585A biosynthetic gene cluster	22%
Unbinned	Polyketide	Alkyl-O-dihydrogeranyl-methoxyhydroquinones biosynthetic gene cluster	17%
Unbinned	Terpene	N/A	
Unbinned	Terpene	Geosmin biosynthetic gene cluster	100%

Supplementary Table 10. Rate, threshold, and theoretical bacterial population sustained by atmospheric H₂ and CO oxidation in Robinson Ridge and Adams Flat soil samples.

Trace gas	Rate (nmol h ⁻¹ g ⁻¹ _{dw})	Threshold (ppmv)	Theoretical population (CFU g ⁻¹ _{dw})
Robinson Ridge			
Atmospheric H ₂ oxidation	3.49	190	4.8×10^7
Atmospheric CO oxidation	0.42	20	7.2×10^6
Total	3.91	N/A	5.5×10^7
Adams Flat			
Atmospheric H ₂ oxidation	5.54	100	7.6×10^7
Atmospheric CO oxidation	0.25	20	4.2×10^6
Total	5.79	N/A	8.0×10^7

Supplementary Figure- Uncropped scans with size marker indications for Extended Data Figure 8.

Presence and expression of trace gas scavenging genes in Robinson Ridge and Adams Flat soil samples. Raw agarose electrophoresis gels for two separate experiments are shown below. The presence and expression of the genes encoding the group 1h [NiFe]-hydrogenase large subunit (*hhyL*) and type IE RuBisCO large subunit (*rbcL1E*) genes was confirmed by agarose gel electrophoresis. Lane M shows the DNA ladder (Bioline 100 bp Molecular Weight Marker). Lanes 1 and 5 show amplifications from community DNA extracted from the soil samples. Lanes 2 and 6 show amplifications of cDNA derived from reverse transcription of community RNA extracted from the soil samples. No template negative controls for both the PCR (lanes 3, 7) and RT-PCR (lanes 4, 8) are also shown.

Figure 8a

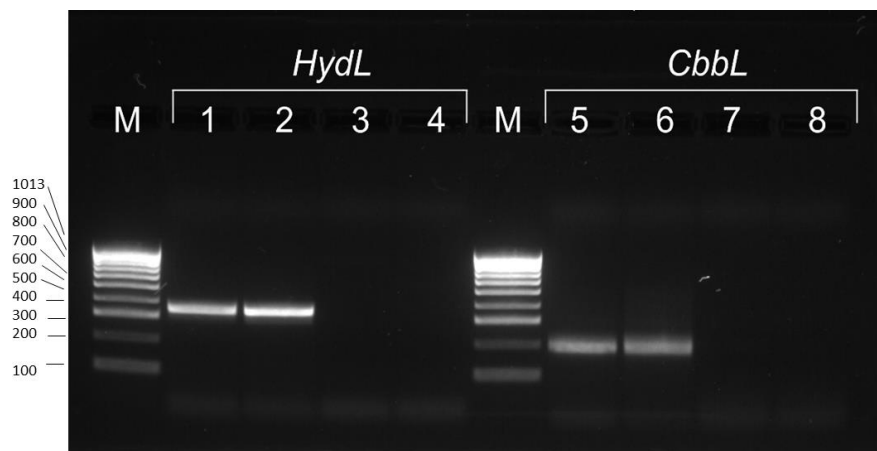


Figure 8b

