Complete Genome Sequence and Methylome Analysis of *Beggiatoa leptomitoformis* strains D-401 and D-402.

Alexey Fomenkov[a*], Zhiyi Sun[a] Tamas Vincze[a], Maria Orlova [b], Brian P. Anton[a], Margarita Yu. Grabovich[b], Richard J. Roberts [a].

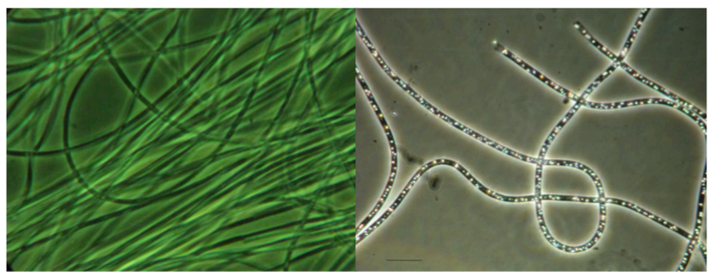
[a]New England Biolabs, Ipswich, MA, USA
[b]Department of Biochemistry and Cell Physiology, Voronezh State University, Voronezh, Russia.

*Correspondence: fomenkov@neb.com

The taxonomy of *Beggiatoa* genus is still a work in progress. Despite many morphotypes of the *Beggiatoa* genus having been described in the literature, only one species, *B. alba* has been validated until now. In 2016, we described a second species, *B. leptomitoformis*. Two strains of *B. leptomitoformis* D-401 and D-402 have been isolated from different regions of Russia. They differ in their morphology and physiology and especially their ability to grow lithotrophically in the presence of thiosulfate. While *B. leptomitoformis* D-402 is able to accumulate elemental sulfur, strain D-401 cannot. We performed genomic sequencing of these two strains using the PacBio SMRT platform and assembled the reads into two complete circular genomes: *B. leptomitoformis* D-401 with 4,266,286 bp and D-402 with 4,265,296 bp. Both genome sequences have been deposited in GenBank with accession numbers CP018889 and CP012373 respectively (1). Surprisingly these two genomes showed almost 99% identity. The preliminary analysis of metabolic operons of thiosulfate oxidation (soxAXBZY) and autotrophic assimilation of CO_2 (the genes encoded for the enzymes of the Calvin-Bassham cycle) in both strains did not reveal any noticeable differences.

One advantage of the PacBio sequencing platform is its ability to detect the epigenetic state of the sequenced DNA, which allows for the identification of modified nucleotides and the corresponding motifs in which they occur. Thirteen DNA methyltransferase recognition motifs were found. They include one m4C and nine m6A modifications that were detected by direct SMRT sequencing and an additional three m5C motifs were detected in Tet2 treated DNA. The motifs were then matched with methyltransferase genes in the genome, and the results have been deposited in REBASE (2).

Beggiatoa leptomitoformis

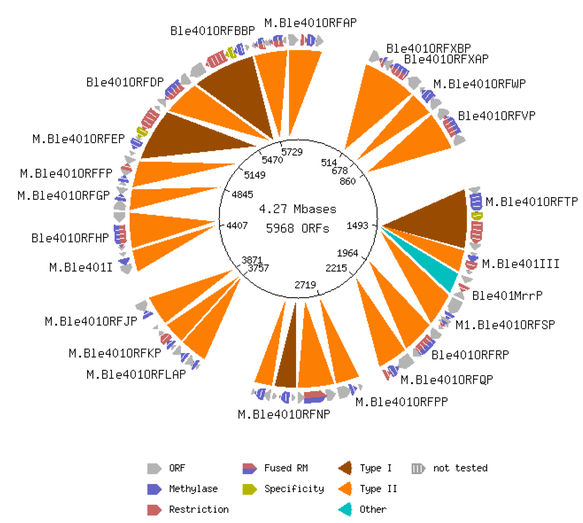


D-401 D-402

The figure demonstrates microphotograph of the bacterial filaments of *Beggiatoa leptomitoformis* strains D-401without and D-402 with granules of elemental sulfur. The size of this bacterium can be as long as 100 micron and even longer. Bar on a right figure, 10um.

Localization of Restriction and Modification genes on assembeled chromosome map from *B.leptomitoformis* strains.

D-401



D-402

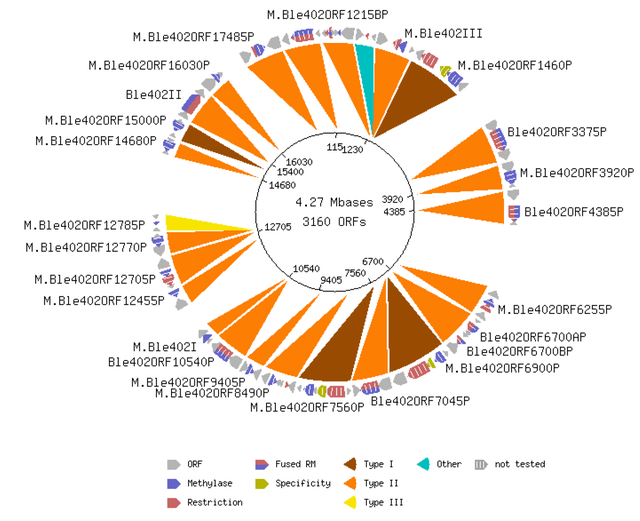


Table1 Genomic differences detected between *B. leptomitoformis* strains D-401 and D-402

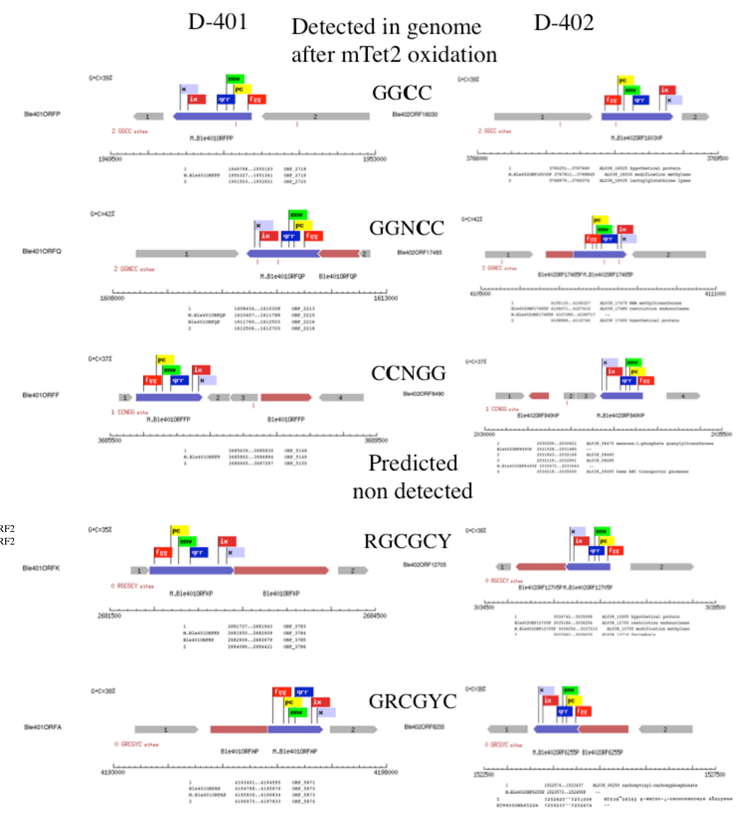
TYPE	LENGT1	INDEL COORDS	GENE AFFECTE1	GENE COORDS		
deletion	1	12064 .. 12064	lAL038_RS00995	11777 .. 12115	internal deletion	hypothetical protein
deletion	1	47013 .. 47013	lAL038_RS00290	46477 .. 47267	internal deletion	hypothetical protein
deletion	782	61181 .. 61962	lAL038_RS00335	61269 .. 62270		diguanylate cyclase response regulator
deletion	1	441052 .. 441052	lAL038_RS01885	439926 .. 441278	internal deletion	magnesium transporter
deletion	1	719628 .. 719628	lAL038_RS03120	719389 .. 719694	internal deletion	hypothetical protein
deletion	782	899337 .. 900118	lAL038_RS03840	898334 .. 900254	internal deletion	long-chain fatty acid--CoA ligase
deletion	1	926216 .. 926216	lAL038_RS03945	926161 .. 928076	internal deletion	1-deoxy-D-xylulose-5-phosphate synthase
deletion	1	1065356 .. 1065356	lAL038_RS04450	1063612 .. 1065471	internal deletion	hypothetical protein
deletion	1	1209976 .. 1209976	lAL038_RS05010	1209202 .. 1210313	internal deletion	microcin ABC transporter permease
deletion	853	1232192 .. 1233044	lAL038_RS05105	1231830 .. 1233056	internal deletion	hypothetical protein (цитуаллат циклаза)
deletion	782	1476518 .. 1477299	lAL038_RS06065	1476218 .. 1476664		hypothetical protein (Blr3520 protein homolog, hypothetical protein)
deletion	782	1476518 .. 1477299	lAL038_RS06070	1476916 .. 1477611		urea ABC transporter ATP-binding subunit UrtE
deletion	853	1803622 .. 1804474	lAL038_RS07520	1803331 .. 1803648		hypothetical protein
deletion	853	1803622 .. 1804474	lAL038_RS07525	1803901 .. 1804767		hypothetical protein (Helicase PriA essential for oriC/DnaA-independent DNA replication)
deletion	1	1932373 .. 1932373	lAL038_RS08070	1931320 .. 1932756	internal deletion	ammonia channel protein
deletion	1	2044685 .. 2044685	lAL038_RS08545	2044536 .. 2044994	internal deletion	ribonuclease HI
deletion	782	2225440 .. 2226221	lAL038_RS09260	2224368 .. 2226986	internal deletion	hypothetical protein (Alkaline phosphatase (EC 3.1.3.1))
deletion	853	2636742 .. 2637594	lAL038_RS11000	2636025 .. 2636819		peptidase
						hypothetical protein (Probable transmembrane protein)
deletion	853	2636742 .. 2637594	lAL038_RS11005	2636941 .. 2638794		
deletion	853	2808212 .. 2809064	lAL038_RS11790	2806941 .. 2809562	internal deletion	hypothetical protein (Serine phosphatase RsbU, regulator of sigma subunit)
deletion	1	3181477 .. 3181477	lAL038_RS13330	3181223 .. 3182407	internal deletion	hypothetical protein
deletion	1	3201097 .. 3201097	lAL038_RS13410	3200947 .. 3201966	internal deletion	hypothetical protein
deletion	1	3202285 .. 3202285	lAL038_RS13415	3201970 .. 3202605	internal deletion	threonylcarbamoyl-AMP synthase
deletion	1	3629468 .. 3629468	lAL038_RS15340	3627872 .. 3629689	internal deletion	hypothetical protein
deletion	1	3904261 .. 3904261	lAL038_RS16620	3903872 .. 3904276	internal deletion	transposase
deletion	1	19525 .. 19525	INTERGENIC			
deletion	1	1168050 .. 1168050	INTERGENIC			
deletion	1	1468221 .. 1468221	INTERGENIC			
deletion	1	1729712 .. 1729712	INTERGENIC			
deletion	1	2931338 .. 2931338	INTERGENIC			
insertion	1	67570 .. 67570	ORF_1884l	67407 .. 67826	internal insertion	Heat shock protein 60 family co-chaperone GroES
insertion	1	68032 .. 68032	ORF_1883l	67971 .. 68360	internal insertion	Alkaline phosphatase (EC 3.1.3.1)
insertion	1	68606 .. 68606	ORF_1880l	68321 .. 68749	internal insertion	hypothetical protein
insertion	1	68606 .. 68606	ORF_1879l	68477 .. 68830	internal insertion	hypothetical protein
insertion	1	69038 .. 69038	ORF_1877l	68788 .. 69315	internal insertion	unnamed protein product
insertion	1	69174 .. 69174	ORF_1877l	68788 .. 69315	internal insertion	unnamed protein product
insertion	1	69038 .. 69038	ORF_1878l	68881 .. 69267	internal insertion	unnamed protein product
insertion	1	69174 .. 69174	ORF_1878l	68881 .. 69267	internal insertion	unnamed protein product
insertion	1	102043 .. 102043	ORF_1822l	101586 .. 102968	internal insertion	ATP-dependent RNA helicase RhlB
insertion	1	158205 .. 158205	ORF_1744l	157827 .. 158999	internal insertion	DNA repair protein RadA
insertion	1	202935 .. 202935	ORF_1693l	202819 .. 203187	internal insertion	Ribonuclease P protein component (EC 3.1.26.5)
insertion	1	223992 .. 223992	ORF_1673l	221116 .. 224784	internal insertion	Two-component hybrid sensor and regulator
insertion	1	227613 .. 227613	ORF_1666l	227158 .. 228060	internal insertion	Methylene tetrahydromethanopterin dehydrogenase (EC 1.5.99.9)
insertion	1	262243 .. 262243	ORF_1622l	257782 .. 262647	internal insertion	Adenylate cyclase (EC 4.6.1.1)
insertion	1	265884 .. 265884	ORF_1616l	265125 .. 265886	internal insertion	hypothetical protein
insertion	1	275848 .. 275848	ORF_1595l	275600 .. 276619	internal insertion	ATP-dependent DNA helicase UvrD/PcrA
insertion	853	343046 .. 343898	ORF_1475l	342817 .. 344478	internal insertion	Membrane protein
insertion	853	343046 .. 343898	ORF_1477l	343781 .. 344221		Membrane protein
insertion	11	344555 .. 344565	ORF_1474l	344459 .. 344875	internal insertion	hypothetical protein
insertion	853	403645 .. 404497	ORF_1404l	401686 .. 403647		sensory box histidine kinase/response regulator
insertion	853	403645 .. 404497	ORF_1401l	403814 .. 404590		Ferric siderophore transport system, periplasmic binding protein TonB
insertion	853	403645 .. 404497	ORF_1403l	403961 .. 404284		Ferric siderophore transport system, periplasmic binding protein TonB
insertion	853	403645 .. 404497	ORF_1402l	403975 .. 404325		Ferric siderophore transport system, periplasmic binding protein TonB
insertion	1	408039 .. 408039	ORF_1388l	407818 .. 408687	internal insertion	Signal peptidase I (EC 3.4.21.89)
insertion	1	520254 .. 520254	ORF_1229l	519698 .. 521881	internal insertion	Fe-S OXIDOREDUCTASE (1.8.-.-)
insertion	1	720562 .. 720562	ORF_939l	720492 .. 722714	internal insertion	hypothetical protein
insertion	1	724653 .. 724653	ORF_937l	723961 .. 727632	internal insertion	Glucokinase (EC 2.7.1.2)
insertion	1	729867 .. 729867	ORF_933l	729306 .. 730739	internal insertion	Arylsulfatase regulator (Fe-S oxidoreductase)
insertion	1	744204 .. 744204	ORF_915l	743694 .. 745556	internal insertion	Adenylate cyclase (EC 4.6.1.1)
insertion	1	813170 .. 813170	ORF_828l	812507 .. 813349	internal insertion	hypothetical protein
insertion	1	898865 .. 898865	ORF_702l	897384 .. 899303	internal insertion	Long-chain-fatty-acid--CoA ligase (EC 6.2.1.3)
insertion	1	1004515 .. 1004515	ORF_572l	1004299 .. 1005012	internal insertion	Glycosyl transferase, group 1
insertion	1	1004515 .. 1004515	ORF_573l	1004406 .. 1004591	internal insertion	Glycosyl transferase, group 1
insertion	1	1118889 .. 1118889	ORF_432l	1118721 .. 1120220	internal insertion	DNA polymerase III alpha subunit (EC 2.7.7.7)
insertion	1	1131646 .. 1131646	ORF_416l	1131484 .. 1132146	internal insertion	Protein DVU_0532 (HMC operon ORF 5)
insertion	1	1131646 .. 1131646	ORF_417l	1131645 .. 1131833	internal insertion	Protein DVU_0532 (HMC operon ORF 5)
insertion	1	1205050 .. 1205050	ORF_324l	1204718 .. 1205326	internal insertion	ABC transporter, ATP-binding protein
insertion	1	1205050 .. 1205050	ORF_325l	1204969 .. 1205271	internal insertion	ABC transporter, ATP-binding protein
insertion	853	1233216 .. 1234068	ORF_286l	1233072 .. 1233740		diguanylate cyclase/phosphodiesterase (GGDEF & EAL domains) with PAS/PAC sensor(s)
insertion	853	1233216 .. 1234068	ORF_287l	1233230 .. 1233631		diguanylate cyclase/phosphodiesterase (GGDEF & EAL domains) with PAS/PAC sensor(s)
insertion	853	1233216 .. 1234068	ORF_285l	1233641 .. 1235194		diguanylate cyclase/phosphodiesterase (GGDEF & EAL domains) with PAS/PAC sensor(s)
insertion	1	1240727 .. 1240727	ORF_276l	1240545 .. 1241426	internal insertion	Adenylate cyclase (EC 4.6.1.1)
insertion	1	1337119 .. 1337119	ORF_148l	1336927 .. 1338054	internal insertion	hypothetical protein
insertion	1	1412842 .. 1412842	ORF_54l	1412447 .. 1412851	internal insertion	Periplasmic septal ring factor with murein hydrolase activity EnvC/YibP
insertion	1	1412842 .. 1412842	ORF_53l	1412658 .. 1413515	internal insertion	Periplasmic septal ring factor with murein hydrolase activity EnvC/YibP
insertion	1	1777115 .. 1777115	ORF_553l	1775029 .. 1777428	internal insertion	Multimodular transpeptidase-transglycosylase (EC 2.4.1.129) (EC 3.4.-.-)
insertion	1	1838741 .. 1838741	ORF_543l	1838342 .. 1839412	internal insertion	Nitrogen regulation protein NtrB (EC 2.7.13.3)
insertion	1	1846221 .. 1846221	ORF_542l	1842310 .. 1846827	internal insertion	putative exported protein
insertion	1	1906363 .. 1906363	ORF_532l	1906302 .. 1907150	internal insertion	Hypothetical ATP-binding protein UPF0042, contains P-loop
insertion	1	2096122 .. 2096122	ORF_506l	2095859 .. 2096308	internal insertion	Periplasmic thiol:disulfide oxidoreductase DsbB, required for DsbA reoxidation
insertion	1	2096122 .. 2096122	ORF_506l	2095879 .. 2096265	internal insertion	Periplasmic thiol:disulfide oxidoreductase DsbB, required for DsbA reoxidation
insertion	1	2120882 .. 2120882	ORF_503l	2120616 .. 2120960	internal insertion	hypothetical protein
insertion	1	2286088 .. 2286088	ORF_480l	2284152 .. 2288318	internal insertion	conserved repeat domain
insertion	1	2287495 .. 2287495	ORF_480l	2284152 .. 2288318	internal insertion	conserved repeat domain
insertion	853	2359960 .. 2360812	ORF_471l	2359021 .. 2359977		hypothetical protein
insertion	853	2359960 .. 2360812	ORF_470l	2360353 .. 2361792		Sulfatase modifying factor 1 precursor (C-alpha-formylglycine- generating enzyme 1)
insertion	1	2477426 .. 2477426	ORF_454l	2476865 .. 2477506	internal insertion	Queuosine Biosynthesis QueE Radical SAM
insertion	1	2490818 .. 2490818	ORF_452l	2489602 .. 2491497	internal insertion	hypothetical protein
insertion	1	2607228 .. 2607228	ORF_435l	2606789 .. 2607781	internal insertion	hypothetical protein
insertion	1	2633711 .. 2633711	ORF_430l	2633620 .. 2634906	internal insertion	UDP-glucose dehydrogenase (EC 1.1.1.22)
insertion	853	2676217 .. 2677069	ORF_425l	2676289 .. 2677860		Glucans biosynthesis protein G precursor
insertion	1	2742466 .. 2742466	ORF_418l	2742086 .. 2742478	internal insertion	hypothetical protein
insertion	1	2742466 .. 2742466	ORF_418l	2742217 .. 2742528	internal insertion	hypothetical protein
insertion	1	2773084 .. 2773084	ORF_412l	2772610 .. 2773461	internal insertion	Protein-N(5)-glutamine methyltransferase PrmC, methylates polypeptide chain release factors RF1 and RF2
insertion	1	2773084 .. 2773084	ORF_412l	2772905 .. 2773237	internal insertion	Protein-N(5)-glutamine methyltransferase PrmC, methylates polypeptide chain release factors RF1 and RF2
insertion	1	2937617 .. 2937617	ORF_392l	2937078 .. 2938022	internal insertion	hypothetical protein
insertion	378	2979930 .. 2980307	ORF_386l	2978242 .. 2980254		Excinuclease ABC subunit B
insertion	378	2979930 .. 2980307	ORF_386l	2980246 .. 2980434		hypothetical protein
insertion	1	2983687 .. 2983687	ORF_385l	2981387 .. 2985472	internal insertion	High-affinity carbon uptake protein Hat/HatR
insertion	1	2994493 .. 2994493	ORF_384l	2994431 .. 2995222	internal insertion	Branched-chain amino acid transport ATP-binding protein LivG (TC 3.A.1.4.1)
insertion	1	3131012 .. 3131012	ORF_364l	3130760 .. 3131932	internal insertion	COGs COG3146
insertion	1	3161406 .. 3161406	ORF_360l	3160106 .. 3161809	internal insertion	Sensory histidine kinase BaeS
insertion	1	3161406 .. 3161406	ORF_360l	3161164 .. 3161553	internal insertion	Sensory histidine kinase BaeS
insertion	1	3178260 .. 3178260	ORF_357l	3177741 .. 3178868	internal insertion	diguanylate cyclase/phosphodiesterase (GGDEF & EAL domains) with PAS/PAC sensor(s)
insertion	1	3230792 .. 3230792	ORF_350l	3230140 .. 3231036	internal insertion	Putative hemolysin
insertion	1	3301634 .. 3301634	ORF_338l	3301448 .. 3301660	internal insertion	SSU ribosomal protein S3p (S3e)
insertion	1	3301634 .. 3301634	ORF_338l	3301497 .. 3302162	internal insertion	SSU ribosomal protein S3p (S3e)
insertion	1	3301634 .. 3301634	ORF_338l	3301618 .. 3301959	internal insertion	SSU ribosomal protein S3p (S3e)
insertion	1	3467358 .. 3467358	ORF_316l	3467170 .. 3467367	internal insertion	dTDP-4-dehydrorhamnose 3,5-epimerase (EC 5.1.3.13)
insertion	1	3479368 .. 3479368	ORF_314l	3478545 .. 3479423	internal insertion	methyltransferase, FkbM family domain protein
insertion	1	3506640 .. 3506640	ORF_311l	3506290 .. 3507393	internal insertion	ATPase, AFG1 family
insertion	1	3574545 .. 3574545	ORF_302l	3573833 .. 3574666	internal insertion	Aminodeoxychorismate lyase (EC 4.1.3.38)
insertion	1	3598781 .. 3598781	ORF_297l	3598733 .. 3599497	internal insertion	Methionine aminopeptidase (EC 3.4.11.18)
insertion	1	3735496 .. 3735496	ORF_278l	3733603 .. 3735840	internal insertion	Adenylate cyclase (EC 4.6.1.1)
insertion	1	3764653 .. 3764653	ORF_272l	3763972 .. 3765105	internal insertion	Ni,Fe-hydrogenase I cytochrome b subunit
insertion	1	3892676 .. 3892676	ORF_252l	3889873 .. 3893034	internal insertion	FB22 precursor
insertion	1	4009918 .. 4009918	ORF_234l	4008746 .. 4010092	internal insertion	tRNA-i(6)A37 methylthiotransferase
insertion	1	4058665 .. 4058665	ORF_227l	4056168 .. 4058795	internal insertion	Membrane alanine aminopeptidase N (EC 3.4.11.2)
insertion	853	4144221 .. 4145073	ORF_217l	4143477 .. 4144634		COG1565: Uncharacterized conserved protein
insertion	853	4144221 .. 4145073	ORF_217l	4144670 .. 4145419		FolM Alternative dihydrofolate reductase I
insertion	1	4172038 .. 4172038	ORF_213l	4171564 .. 4172226	internal insertion	Uncharacterized protein conserved in bacteria
insertion	1	3546252 .. 3546252	INTERGENIC			
SNP		3345637 (REF), 33474	AL038_RS14105 /	3344588 .. 3346021	(1 silent	

Reference strain is NZ_CP012373- Beggiatoa_leptomitiformis_D402
NEB strain is NZ_CP018889 -Beggiatoa_leptomitiformis_D401
Deletions are present in the reference but not in the NEB strain; **indel coordinates refer to the reference strain.**
Insertions are present in the NEB strain but not in the reference; **indel coordinates refer to the NEB strain.**
The same indel may be listed in more than 1 row if it affects more than 1 gene.

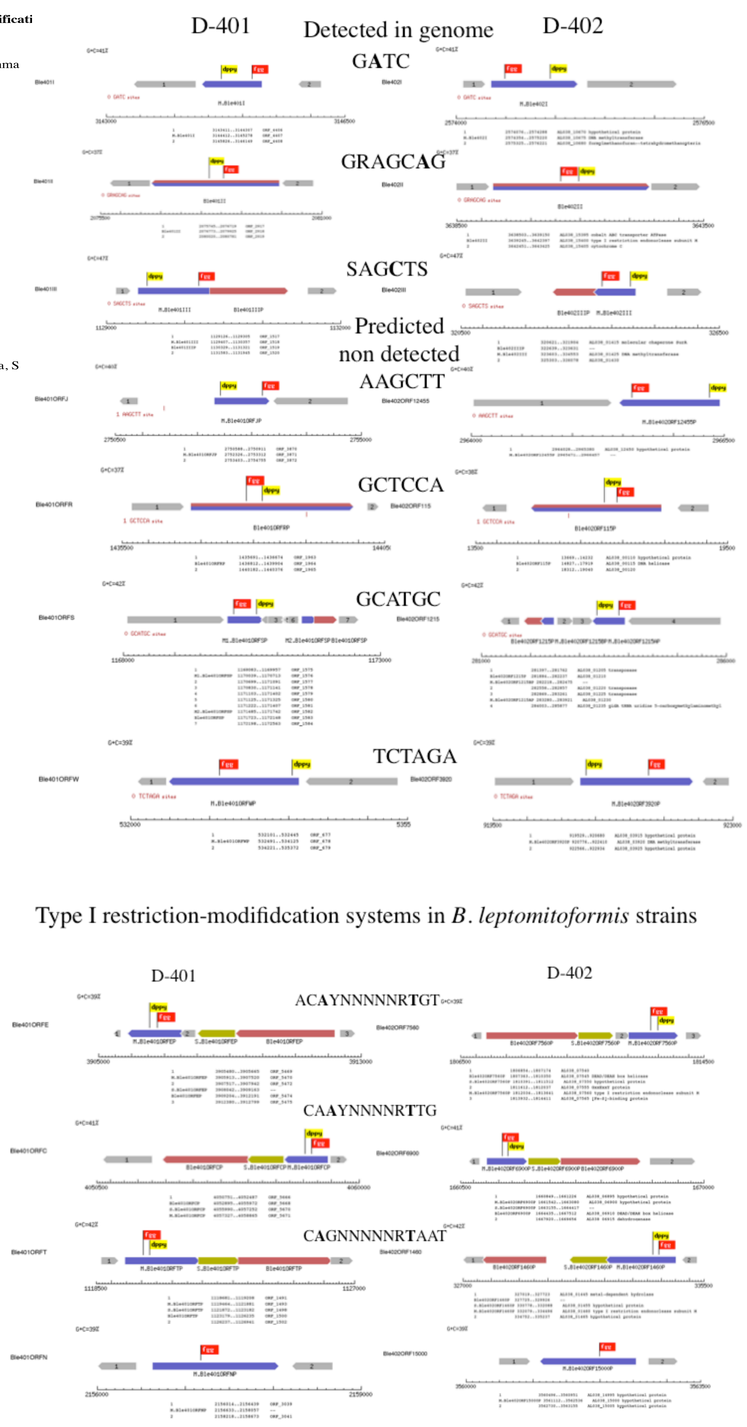
Table2 Methyltransferase genes and their cognate motifs identified in *B. leptomitoformis* strains D-401 and D-402

Predicted Gene in Ble402	% Detected in Ble402	Predicted Gene in Ble401	% Direct detection in Ble401	Motif Assigned Motifs	Methylation Type	Restriction Modificati Type
M.Ble402I	99.8	M.Ble401I	100	GATC	m6A	II alpha
M.Ble402II	99.5	M.Ble401II	100	GRAGCAG	m6A	II G, S, gamma
M.Ble402III	20.1	M.Ble401III	99.4	SAGCTS	m4C	II beta
S.Ble402ORF7560P	96.4	S.Ble401ORFEP	non-detected*	ACAYNNNNNRTGT	m6A	I
S.Ble402ORF6900P	71.8	S.Ble401ORFCP	72.45	CAAYNNNNNRTTG	m6A	I
S.Ble402ORF1460P	96.9	S.Ble401ORFTP	98.8	CAGNNNNNRTAAT	m6A	I
				Unassigned motifs		
	100		100	CATCHAG	m6A	II G
	98.9		99.4	CGGAG	m6A	III
	98.2		99.2	CGGTCA	m6A	II G
	97.7		99.9	DCTGGATD	m6A	II
	99.6		99.9	GGCTGA	m6A	II G
	100		100	GTTGNAG	m6A	II G
	98.7		99.9	TCGA	m6A	II
				5mC detected with mTet2 oxidation		
M.Ble402ORF17485P	71.56	M.Ble401ORFQP	ND	GGHCC=GGNCC	5mC	II
M.Ble402ORF8490P	39.55	M.Ble401ORFFP	ND	CCDGG=CCNGG	5mC	II
M.Ble402ORF16030P	29.6	M.Ble401ORFPP	ND	GGCCNB=GGCC	5mC	II
				Predicted/ non detected		
M.Ble402ORF12705P		M.Ble401ORFKP		RGCGCY	5mC	II
M.Ble402ORF6255P		M.Ble401ORFAP		GRCGYC	5mC	II
Ble402ORF115P		Ble401ORFRP		GCTCCA	m6A	II G, gamma, S
M.Ble402ORF1215P		M1-2.Ble401ORFSP		GCATGC	ND	II alpha
M.Ble402ORF12455P		M.Ble401ORFJP		AAGCTT	ND	II beta
M.Ble402ORF3920P		M.Ble401ORFWP		TCTAGA	ND	II beta
M.Ble402ORFA1P				ATGCAT	ND	II

Type II m5C DNA methyltransferase genes, and their assigned specificities, from *B. leptomitoformis* strains D-401 and D-402



Type II m6A and m4C DNA methyltransferase genes, and their assigned specificities, from *B. leptomitoformis* strains D-401 and D-402



Type I restriction-modifidcation systems in *B. leptomitoformis* strains

References.

(1) Fomenkov,A., Vincze, T., Grabovich, M.Y., Dubinina, G., Orlova, M., Belousova, E., Roberts, R.J. (2015) Complete Genome Sequence of the Freshwater Colorless Sulfur Bacterium *Beggiatoa leptomitiformis* Neotype Strain D-402. Genome Announc., 6:e01436. DOI:10.1128/genomeA.01436-15

(2) Roberts,R.J., Vincze,T. Posfai, J. Macelis, D. REBASE a database for DNA restriction and modification enzymes,genes and genomes. Nuclic Acids Res. 43:D298-D299 (2015). DOI:10,1093/nar/gku1046