

High-quality draft genome sequence of *Pantanelinema* sp. GBBB05, a cyanobacterium from Cerrado biome

Supplementary Material

Table S1

Statistics obtained through CheckM of the bins recovered from the photosynthetic consortium metagenome and NCBI (genome) accession numbers

Accession number	Organisms	Completeness (%)	Contamination (%)	N50 (bp)	L50	Size (bp)	Coverage	GC (%)	bin
VRZD000000000	<i>Pantanelinema</i> sp. GBBB05	54.94	2.844	5621	131	2143083	7	65.79	bin_16
JACVCF000000000	<i>Blastomonas</i> sp. GBBB05	99.67	0.0	647436	2	2604855	66	67.15	bin_2
JACVCG000000000	<i>Brevundimonas</i> sp. GBBB05	98.77	0.9	225365	7	4299830	17	65.54	bin_6
JACVCH000000000	Rhizobiaceabacterium GBBB05	51.72	1.724	5728	231	3937098	7	63.33	bin_7
JACVCI000000000	<i>Pseudomonas</i> sp. GBBB05	99.82	1.075	97281	14	3725532	16	52.27	bin_8
JACVCJ000000000	Bacteroidiabacterium GBBB05	66.39	2.346	12962	63	2586752	10	70.35	bin_9
JACVCK000000000	Phycisphaeralesbacterium GBBB05	99.05	0.4	142797	17	7181771	18	48.43%	bin_11

TableS2

Bins classification using two different classifiers: GTDB-TK and Classify_bins

Bins	GTDDB-TK	Classify_bins
bin_2	<i>Blastomonas</i> (g)	<i>Blastomonas</i> (g)
bin_6	<i>Brevundimonas</i> (g)	Caulobacteraceae (f)
bin_7	Rhizobiaceae (f)	Phyllobacteriaceae (f)
bin_8	<i>Pseudomonas</i> (g)	<i>Pseudomonas mendocina</i> (s)
bin_9	Bacteroidia (c)	Bacteroidota (p)
bin_11	Phycisphaerales (o)	Bacteria (d)
bin_16	Leptolyngbyaceae (f)	Cyanobacteria (p)

Taxonomic level: d: domain; p: phylum; o: order; c: class; f: family; g: genus; s: specie.

Table S3CRISPR/CAS sequences identified in *Pantanalinema* GBBB05 genome through bioinformatic analysis

Element	CRISPR Id / CasType	Start	End	Spacer / Gene	Repeat consensus / cas genes	Direction	Evidence Level
CRISPR	NODE_6_length_425267_cov_7_059467_1	23208 4	23221 4	1	TCGGCTCCCTCCTTATTAAGGAGGGTTGGGGA GGATC	ND	1
CRISPR	NODE_12_length_325133_cov_7_187926_1	52707	52821	1	ATGTGCTCTATGACAACACTAGGGCGATCGC	ND	1
CRISPR	NODE_20_length_239621_cov_7_852330_1	15708 2	15716 8	1	ATTCAGTGTTTCCGCAAGCAAAC	ND	1
CRISPR	NODE_32_length_193142_cov_7_329306_1	16513	16599	1	GTTACTCGATGTTGGGACGATCGT	ND	1
CRISPR	NODE_33_length_192458_cov_6_760205_1	58493	58609	1	CCGAAGGATATGATTTGACTTCTGAACTGAA A	ND	1
CRISPR	NODE_41_length_154863_cov_7_417485_1	10604 7	10613 6	1	TTTCTTATGAATGCTCGATCGTTTGTT	ND	1
CRISPR	NODE_49_length_129209_cov_7_774880_1	84384	84487	1	GGTGGGAATCGAACCCACAAAATCT	ND	1

TableS3(Cont.)

Element	CRISPR Id / CasType	Start	End	Spacer / Gene	Repeat consensus / cas genes	Direction	Evidence Level
CRISPR	NODE_49_length_129209_cov_7_77488_0_2	84594	84781	2	GGTGGGAATCGAACCCACAAAATCT	ND	1
CRISPR	NODE_52_length_125719_cov_9_29386_4_1	3634	3782	2	CCCATTCTCATTCCGCCAATCAT	ND	1
CRISPR	NODE_66_length_102395_cov_7_12220_8_1	75737	80489	65	GCACCAAAACCATTGAATCCCTTG CAGGGATT GAAAC	ND	4
CRISPR	NODE_68_length_101587_cov_7_39772_3_1	33629	33706	1	GAATCTCGCCTGCCTGATCTTTAA	ND	1
CRISPR	NODE_87_length_83373_cov_7_663407_1	26195	26282	1	AATTGAGTGCGATCGTGGCACAGTT	ND	1
CRISPR	NODE_106_length_61554_cov_8_71997_7_1	7539	7646	2	TCTTGAGTAAGATCGATCTTACTCA	ND	1
CRISPR	NODE_143_length_37220_cov_8_29744_2_1	35114	35220	1	CTCCTGTAGCCCAATTGGAAGAGG	ND	1
CRISPR	NODE_158_length_32324_cov_7_42097_1_1	26934	27049	1	CGACTGCTCAAATCGACACCTTGCCCCAAAAAA	ND	1

TableS3(Cont.)

Element	CRISPR Id / CasType	Start	End	Spacer / Gene	Repeat consensus / cas genes	Direction	Evidence Level
CRISPR	NODE_169_length_28977_cov_7_44187_2_1	5946	6040	1	CCTGTATTGTAGATACCCCCGCC	ND	1
CRISPR	NODE_268_length_15810_cov_7_60466_7_1	10883	11959	14	GTTTCCGTCCTCGCGGGGAAAGGGTAATGA AAG	ND	4
CAS cluster	General-Class1	9689	10840	1	Cas6_0_I-III		

TableS4

Prophage identified in *Pantanelinema* GBBB05 genome using PHAST server

Region	Region length	Completeness	Score	CDS	Region Position	Possible Phage	GC percent
1	9.9 Kb	Incomplete	40	16	4040422-4050410	<i>Synechococcus</i> phage (NC_031906)	50.89%

TableS5

The *in silico* DDH values for *Pantalaninema* GBBB05 and selected reference genomes of Leptolyngbyaceae family.

Reference Genome (Accession Number)	Difference in % G+C	DDH (%)	Distance	Probability that DDH > 70% (i.e., same species)
<i>Leptolyngbya</i> (NZ_AP017308.1)	1.78	20.00	0.2196	0
<i>Leptolyngbyaboryana</i> dg5 (NZ_AP014642.1)	1.43	22.80	0.1917	0
<i>Leptolyngbya frigida</i> (NZ_PVWK00000000.1)	2.59	20.00	0.2201	0
<i>Leptolyngbyaohadii</i> IS1 (NZ_NKFP00000000.1)	3.66	21.80	0.2012	0
<i>Leptolyngbya</i> sp. 'hensonii' (NZ_MQTZ00000000.1)	3.89	22.30	0.1966	0
<i>Leptolyngbyavalderiana</i> BDU 20041 (NZ_LSYZ00000000.1)	11.34	19.10%	0.2304	0
<i>Neosynechococcusphagnicola</i> sy1 (NZ_JJML00000000.1)	3.17	20.50	0.2142	0
<i>Phormidesmispriestleyi</i> BC1401 (NZ_MPPI00000000.1)	0.19	18.40	0.2388	0
<i>Geitlerinema</i> sp. PCC 7407 (NC_019703.1)	10.04	20.7	0.2127	0

Table S5(Cont.)

Reference Genome (Accession Number)	Difference in % G+C	DDH (%)	Distance	Probability that DDH > 70% (i.e., same species)
<i>Leptolyngbyaboryana</i> PCC 6306 (NZ_ALVM00000000.1)	1.41	22.3	0.1969	0
[<i>Leptolyngbya</i>] cyanobacterium JSC 12 (NZ_CM001633.1)	0.94	18.70	0.2348	0
[<i>Leptolyngbya</i>] sp. JSC-1(NZ_JMKF00000000.1)	2.30	19.70	0.2229	0
<i>Alkalinema</i> sp. RU_4_3 (JAAUSA)	5.61	12.50	0.9991	0
<i>Leptolyngbya</i> sp. NIES-3755 (NZ_AP017308.1)	1.78	12.60	0.9957	0
<i>Leptolyngbyaboryana</i> dg5 (NZ_AP014642.1)	1.43	12.60	0.9911	0
<i>Stenomitosfrigidus</i> ULC18 (NZ_PVWK00000000.1)	2.59	12.60	0.9916	0
<i>Leptolyngbyaohadii</i> IS1 (NZ_NKFP00000000.1)	3.66	12.60	0.9937	0
<i>Leptolyngbya</i> sp. 'hensonii'(NZ_MQTZ00000000.1)	3.89	12.60	0.9934	0
<i>Leptolyngbyavalderiana</i> BDU 20041 (NZ_LSYZ00000000.1)	11.34	12.50	0.9995	0
<i>Neosynechococcusphagnicola</i> sy1 (NZ_JJML00000000.1)	3.17	12.60	0.9955	0
<i>Phormidesmispriestleyi</i> ULC007 (NZ_MPPI00000000.1)	0.19	12.60	0.9951	0
<i>Geitlerinema</i> sp. PCC7407(NC_019703.1)	10.04	12.60	0.9960	0
<i>Oscillatorialescyanobacterium</i> C42_A2020_001 (NZ_CM001633.1)	0.94	12.70	0.9910	0
<i>Myxacorysalmedinensis</i> A (NZ_WVIE00000000.1)	1.28	12.60	0.9941	0

Table S6

Number of protein coding genes (CDS) assigned by COG for representatives of core, accessory and unique orthologous clusters present in all four genomes analyzed.

Organism	Number of core genes	Number of accessory genes	Number of unique genes	Total	GenBank Accession Number
<i>Geitlerinema</i> sp. PCC 7407	1,624	676	1,360	3,660	(NC_019703.1)
<i>Leptolyngbyaboryana</i> PCC 6306	1,624	1,206	3,229	6,059	(NZ_ALVM000000000.1)
<i>Leptolyngbyacyanobacterium</i> JSC 12	1,624	1,206	1,622	4,452	(NZ_CM001633.1)
<i>Pantanelinema</i> sp. GBBB05	1,624	1,575	2,434	5,633	(VRZD000000000)

Table S7

Overview of secondary metabolites identified by antiSMASH

Region	Type	From	To	Most similar known cluster	Similarity
Region 1.1	terpene	1	15713		
Region 1.2	bacteriocin	202358	212591		
Region 2.1	betalactone	221361	245891		
Region 3.1	resorcinol	1	30652		
Region 3.2	terpene	197847	218851		
Region 5.1	NRPS,T1PKS	198374	239621	Nostophycin (NRPS)	18%
Region 8.1	NRPS	15748	62374		
Region 8.2	bacteriocin	138633	149037		
Region 10.1	NRPS	1	52895	Nostopeptolide (NRPS)	37%
Region 10.2	terpene	81537	102469		
Region 15.1	bacteriocin	74363	84572		
Region 18.1	terpene	36106	58082		

Table S7(Cont.)

Region	Type	From	To	Most similar known cluster	Similarity
Region 21.1	NRPS-like,NRPS	51	108460	Anabaenopeptin (NRPS)	100%
Region 28.1	NRPS	1	23542		
Region 49.1	terpene	11455	32687		
Region 52.1	bacteriocin	11014	23757		
Region 62.1	NRPS	1	28438	Cyanopeptin (NRPS)	75%

Table S8

Biosynthetic gene clusters predicted in the *Pantanalinema* GBBB05 genome by NaPDos server for KS domain

ID	Locus	Database Match Id	Identity (%)	Alignment Length	e-value pathway	Product domain	Class	BLAST nr	Identity (%)
KS1	NODE_201_length_22649_cov_7.4 28559_2_0_12_386	FabF_Bacillus_FAS	34	406	2,00E-49	<u>fattyacids</u> <u>ynthesis</u>	FAS	No results	-
KS2	NODE_180_length_26464_cov_7.2 18894_1_0_40_290	FcsE_AAQ82567_3KS B	58	36	2,00E-06	<u>candididin</u>	Modular	acetyl-CoA C-acyltransferase [Cyanobacteriabacterium QH_8_48_120]	88,89
KS3	NODE_32_length_193142_cov_7.3 29306_2_10_43_450	FabF_Bacillus_FAS	60	408	2,00E-132	<u>fattyacids</u> <u>ynthesis</u>	FAS	beta-ketoacyl-ACP synthase II [Leptolyngbya sp. 'hensonii']	88,70
KS4	NODE_20_length_239621_cov_7.8 52330_1_26_39_464	EpoC_Q9L8C8_H	59	428	3,00E-138	<u>epothilone</u>	hybridKS	type I polyketide synthase [Pseudanabaena sp. 'RoaringCreek']	80

Table S9

Biosynthetic gene clusters predicted in the *Pantanalinema* GBBB05 genome by NaPDos server for C domain

ID	Locus	Database Match Id	Identity (%)	Alignment Length	e-value pathway	Product domain	Class	BLAST nr	Identity (%)
C1	NODE_82_length_8627 6_cov_7.810480_6_6_85 _345	syryn1_C9_LCL	40	256	2,00E-46	<u>syngomycin</u>	LCL	amino acid adenylation domain- containing protein, partial [filamentous cyanobacterium CCP2]	53,70
C2	NODE_173_length_284 38_cov_9.152697_1_1_1 2_316	syryn1_C7_LCL	43	299	1,00E-58	<u>syngomycin</u>	LCL	non-ribosomal peptide synthetase [Phormidiumambiguum]	75,34
C3	NODE_173_length_284 38_cov_9.152697_3_0_9 6_398	syryn1_C6_LCL	47	300	2,00E-57	<u>syngomycin</u>	LCL	amino acid adenylation domain- containing protein, partial [filamentous cyanobacterium CCP2]	71,47
C4	NODE_33_length_1924 58_cov_6.760205_1_1_6 18_921	syryn1_C7_LCL	49	300	2,00E-79	<u>syngomycin</u>	LCL	non-ribosomal peptide synthetase, partial [Chloroflexi bacterium]	63,42
C5	NODE_33_length_1924 58_cov_6.760205_1_1_1 802_2104	syryn1_C5_LCL	44	300	8,00E-55	<u>syngomycin</u>	LCL	putative PAS/PAC sensor protein [Microcoleusvaginatus FGP-2]	58,72
C6	NODE_33_length_1924 58_cov_6.760205_1_2_4 0_343	syryn1_C9_LCL	47	295	8,00E-67	<u>syngomycin</u>	LCL	non-ribosomal peptide synthetase, partial [Chloroflexi bacterium]	60,07

Table S9(Cont.)

ID	Locus	Database Match Id	Identity (%)	Alignment Length	e-value pathway	Product domain	Class	BLAST nr	Identity (%)
C7	NODE_28_length_2089 57_cov_7.404319_3_3_9 00_1200	syrin1_C9_LCL	47	298	1,00E-70	<u>syringomycin</u>	LCL	non-ribosomal peptide synthetase [Nostoc sp. 'Peltigeramalaceacyanobiont' DB3992]	61,69
C8	NODE_55_length_1229 05_cov_7.418235_5_9_6 61_961	syrin1_C5_LCL	46	298	4,00E-62	<u>syringomycin</u>	LCL	non-ribosomal peptide synthetase, partial [Chloroflexi bacterium]	59,06
C9	NODE_20_length_2396 21_cov_7.852330_2_16_ 84_384	bleom9_C2_cyc	47	295	7,00E-67	<u>bleomycin</u>	cyc	non-ribosomalpeptidesynthetase [Pseudanabaena sp. SR411]	79,66
C10	NODE_20_length_2396 21_cov_7.852330_2_17_ 45_349	syrin1_C9_LCL	43	297	6,00E-58	<u>syringomycin</u>	LCL	non-ribosomalpeptidesynthetase [Pseudanabaenabiceps]	70,17
C11	NODE_20_length_2396 21_cov_7.852330_2_17_ 1125_1426	syrin1_C9_LCL	37	304	3,00E-45	<u>syringomycin</u>	LCL	non-ribosomal peptide synthetase [Pseudanabaena sp. 'Roaring Creek']	63,48

Table S10

Biosynthetic Gene Clusters (BGC) identified in *Pantanalinema* GBBB05 genome by PRISM server.

Cluster	Type	Number of ORF's
01	Polyketide/nonribosomalpeptide	4
02	Nonribosomalpeptide	1
03	Nonribosomalpeptide	2
04	Nonribosomalpeptide	1
05	Nonribosomalpeptide	2
06	Prochlorosin	6
07	Nonribosomalpeptide	3