JOURNAL OF PURE AND APPLIED MICROBIOLOGY, June 2017. Vol. **11**(2), p. 879-884

Genome Sequences of *Salisediminibacterium haloalkalitolerans* 10nlg, *Bacillus lonarensis* 25nlg, *Bacillus caseinilyticus* SP, *Pelagirhabdus alkalitolerans* S5, *Salibacterium halotolerans* S7 and *Salipaludibacillus aurantiacus* S9 Six Novel, Recently Described Compatible Solute Producing Bacteria

Vishnuvardhan Reddy Sultanpuram* and Thirumala Mothe

Microbial Ecology Laboratory, Department of Applied Biosciences, Mahatma Gandhi University,
Anneparthy, Yellareddygudem (PO), Nalgonda - 508 254, Telangana, India.

http://dx.doi.org/10.22207/JPAM.11.2.26

(Received: 03 February 2017; accepted: 30 April 2017)

Six novel Gram-stain-positive, rod shaped strains of bacteria designated as 10nlg^T, 25nlg^T, SP^T, S5^T, S7^T and S9^T which were isolated from marine ecosystems and soda lake from India, represents novel members of the family *Bacillaceae*. Apart from 10nlg^T, 25nlg^T and SP^T all the other three strains represented members of novel genus from the family. Here we report the draft genome of the strains 10nlg^T, 25nlg^T, SP^T, S5^T, S7^T and S9^T. Strains 10nlg^T, 25nlg^T and SP^T genomes comprised ~ 2.93 Mb, ~ 2.58 Mb and ~ 5.87 Mb with the G + C content of 47.13 %, 43.26 % 44.85 %, respectively. A total of 2772 protein-coding genes in strain 10nlg^T, 2581 protein-coding genes in strain 25nlg^T and 5199 protein-coding genes in strain SP^T were reported. Strain S5^T genome comprised ~ 2.52 Mb with the G + C content of 37.08 % and a total of 2407 protein-coding genes. Strains S7^T and S9^T genomes comprised ~ 3.61 Mb, ~ 4.43 Mb with the G + C content of 47.65 %, 42.42 % and a total of 3662, 4232 protein-coding genes, respectively.

Keywords: Compatible solute producing, Firmicutes, Whole genome sequencing.

The strain *Salisediminibacterium haloalkalitolerans* 10nlg was recently described as a novel species of the genus *Salisedimini bacterium*. The cells are Gram-stain-positive, rod shaped, non-motile, orange pigmented and non-endospore forming. The species showed catalase positive and oxidase negative activities (Sultanpuram *et al*., 2015a). The two novel *Bacillus* species described *Bacillus lonarensis* 25nlg^T and *Bacillus caseinilyticus* SP^T were Gram-stain-positive, terminal endospore forming, motile rods (Sultanpuram *et al*., 2015b; Sultanpuram *et al*., 2015c). Strains *Pelagirhabdus alkalitolerans* S5, *Salibacterium halotolerans* S7 and *Salipaludibacillus aurantiacus* S9 were three novel recently described members representing novel genus from the family *Bacillaceae.* These strains were Gram-stain-positive, rod shaped, non-motile and non-endospore forming, except *Salipaludibacillus aurantiacus* S9 (Sultanpuram *et al*., 2016; Sultanpuram *et al*., 2015d; Sultanpuram and Mothe 2016). Here we report the genome sequencing of these six strains.

MATERIAL AND METHODS

Library construction, genome sequencing and data analysis

Sequencing of the type strains was part of Genomic Encyclopedia of Type Strains, Phase

---

* To whom all correspondence should be addressed.
E-mail: svr.micro@gmail.com

III: the (KMG-III) project (Whitman *et al*., 2015). Project information is deposited in the Genomes on Line Database (GOLD) (Reddy *et al*., 2015) and the high quality draft genome sequence is deposited in GenBank and in the Integrated Microbial Genomes database (IMG) (Markowitz *et al*., 2014). Draft sequencing, initial gap closure and annotation were performed by the DOE Joint Genome Institute (JGI) using state-of-the-art sequencing technology (Mavromatis *et al*., 2012).

An Illumina standard shotgun library was constructed and sequenced using the Illumina HiSeq 2500-1TB. Illumina paired-end reads of average 250-bp length and the clean data were assembled using Velvet (version 1.1.04)(Zerbino and Birney 2008), simulated paired end reads were created from Velvet contigs using wgsim and simulated read pairs were reassembled using Allpaths-LG (version r42328) (MacCallum *et al*., 2009). Protein-coding genes were identified using Prodigal (Hyatt *et al*., 2010), as part of the DOE-JGI genome annotation pipeline (Huntemann *et al*., 2015). Additional gene prediction analysis and manual functional annotation were performed within the Integrated Microbial Genomes (IMG) platform, which provides tools for analyzing and reviewing the structural and functional annotations of genomes in a comparative context (Markowitz *et al*., 2014). Genome annotation procedures are detailed in Markowitz *et al*. and references therein. Briefly, the predicted CDSs were translated and used to search the NCBI nonredundant database, UNIProt, TIGRFam, Pfam, KEGG, COG and InterPro databases. Transfer RNA genes were identified using the tRNAScan-SE tool and other non-coding RNAs were found using INFERNAL. Ribosomal RNA genes were predicted using hmmsearch against the custom models generated for each type of rRNA.

RESULTS AND DISCUSSIONS

Genome features

The Table 1 describes the classification, general features and genome sequencing projects information for the strains according to the MIGS recommendations (Field *et al*., 2008). The assembled draft genome of *Salisediminibacterium haloalkalitolerans* 10nlg contained 2,93,7914 bps with 47.13 % G + C content. Further, the genome analysis predicted 2,282 proteins coding sequences (CDS) of which 80.35 % were assigned to recognized functional genes. The genome analysis revealed 57 tRNA and 13 rRNA genes. The genome of *Bacillus lonarensis* 25nlg^T contained 2,58,2209 bps with 43.26 % G + C content. Further, the genome analysis predicted 2,581 proteins coding sequences (CDS) of which 96.34 % were assigned to recognized functional genes with 67 tRNA and 16 rRNA genes. The assembled genome of *Bacillus caseinilyticus* SP^T contained 5,87,2627 bps with 44.85 % GC content. Further, the genome analysis predicted 5,199 proteins coding sequences (CDS) of which 97.41 % were assigned to recognized functional genes. The genome analysis revealed 101 tRNA and 12 rRNA genes. The genome of *Pelagirhabdus alkalitolerans* S5 contained 2,52,7601 bps with 88.99 % DNA coding bases and 37.08 % GC content. The genome revealed 2,033 protein coding sequences which were about 81.61 % of the total genome. Around 15.01 % CDS present weren't assigned to any of the known functions. Further, it revealed 56 tRNA and 13 rRNA genes. Strain *Salibacterium halotolerans* S7 showed 3, 61, 4063 bps with 47.65% G + C mol%. Around 75.89% of the identified CDS were protein sequences with known functions and about 807 protein coding genes had some unknown functions. Sixty one tRNA and 13 rRNA genes were revealed in the analysis of strain *Salibacterium halotolerans* S7 genome. *Salipaludibacillus aurantiacus* S9 draft genome revealed 4,43,5185 bps genome size with 42.42% G + C mol%. The protein functions of 3349 CDS were known, whereas, 883 CDS protein functions were unknown. Further, it revealed 73 tRNA and 16 rRNA genes.

The genomes of all the above described six strains harboured a cluster of genes coding for compatible solute production, thus explaining their osmotolerance. These genes from strain *Salisediminibacterium haloalkalitolerans* 10nlg include, L- ectoine synthase (130 aa; locus tag SAMN05444126_10550; accession number FOGV01000005.1) and choline dehydrogenase (561 aa, locug tag SAMN05444126_10967; accession number FOGV01000009.1). The strain, *Bacillus lonarensis* 25nlg^T contains ectoine synthase (127 aa; locus tag SAMN05421737_10167; accession number FMYM01000001), glycerol-3-phosphate dehydrogenase (NAD(P)+ (339 aa;

Table 1. Classification, general features and genome sequencing projects information for the strains according to the MIGS recommendations (Field et al., 2008)

MIGS ID	Property Current Classification	Domain *Bacteria*	Domain *Bacteria*	Domain *Bacteria*	Domain *Bacteria*	Domain *Bacteria*	Domain *Bacteria*
		Phylum *Firmicutes*	Phylum *Firmicutes*	Phylum *Firmicutes*	Phylum *Firmicutes*	Phylum *Firmicutes*	Phylum *Firmicutes*
		Class *Bacilli*	Class *Bacilli*	Class *Bacilli*	Class *Bacilli*	Class *Bacilli*	Class *Bacilli*
		Order *Bacillales*	Order *Bacillales*	Order *Bacillales*	Order *Bacillales*	Order *Bacillales*	Order *Bacillales*
		Family *Bacillaceae*	Family *Bacillaceae*	Family *Bacillaceae*	Family *Bacillaceae*	Family *Bacillaceae*	Family *Bacillaceae*
		Genus *Salisediminibacterium*	Genus *Bacillus*	Genus *Bacillus*	Genus *Pelagirhabdus*	Genus *Salibacterium*	Genus *Salipaludibacillus*
		Species *haloalkalitolerans*	Species *lonarensis*	Species *caseinilyticus*	Species *alkalitolerans*	Species *halotolerans*	Species *aurantiacus*
		Strain 10nlg	Strain 25nlg	Strain SP	Strain S5	Strain S7	Strain S9
	Gram stain	Positive	Positive	Positive	Positive	Positive	Positive
	Cell shape	Rod	Rod	Rod	Rod	Rod	Rod
	Motility	Non-motile	Motile	Motile	Non-motile	Non-Motile	Non-motile
	Sporulation	Non-endospore former	Terminal endospore former	Terminal endospore former	Non-Endospore-former	Non-Endospore-former	Endospore-former
	Temperature range	20-50°C	20-50°C	15-60°C	20-55°C	25-45°C	20-45°C
	Optimum Temp.	37°C	35- 37°C	37°C	37°C	37°C	37°C
	Carbon source	Varied	Varied	Varied	Varied	Varied	Varied
	Energy source	Heterotrophic	Heterotrophic	Heterotrophic	Heterotrophic	Heterotrophic	Heterotrophic
MIGS-6	Habitat	Soda lake	Soda lake	Soda lake	Marine	Marine salt pan	Marine salt marsh
MIGS-6.3	Salinity	NaCl	NaCl	NaCl	NaCl	NaCl	NaCl
MIGS-22	Oxygen	Aerobic/Facultative anaerobe	Aerobic/ Facultative anaerobe	Aerobic/ Facultative anaerobe	Aerobic/ Facultative anaerobe	Aerobic/ Facultative anaerobe	Aerobic/ Facultative anaerobe
MIGS-15	Biotic relationship	Free living	Free living	Free living	Free living	Free living	Free living
MIGS-14	Pathogenicity	None	None	None	None	None	None
MIGS-4	Geographic location	Lonar soda lake, India	Lonar soda lake, India	Lonar soda lake, India	Pingaleshwar beach, India	Khavda, India	Narayan sarovar, India
MIGS-5	Sample collection time	05/08/2013	05/08/2013	05/08/2013	12/06/2015	18/09/2014	21/09/2014
MIGS-4.3	Depth	1 m	1 m	1 m	1 m	1 m	1 m
MIGS-4.4	Altitude	Not recorded	Not recorded	Not recorded	Not recorded	Not recorded	Not recorded
MIGS-31	Finishing quality	Permanent draft	Permanent draft	Permanent draft	Permanent draft	Permanent draft	Permanent draft
MIGS-28	Libraries used	Shotgun	Shotgun	Shotgun	Shotgun	Shotgun	Shotgun

MIGS-29	Sequencing platforms	Illumina HiSeq 2500-1TB	Illumina HiSeq 2500-1TB	Illumina HiSeq 2500-1TB	Illumina HiSeq 2500-1TB	Illumina HiSeq 2500-1TB	Illumina HiSeq 2500-1TB
MIGS-30	Assemblers	Newbler v.2.8	Newbler v.2.8	Newbler v.2.8	Newbler v.2.8	Newbler v.2.8	Newbler v.2.8
MIGS-32	Gene calling method	Glimmer 3.02	Glimmer 3.02	Glimmer 3.02	Glimmer 3.02	Glimmer 3.02	Glimmer 3.02
	NCBI project ID	335132	329819	329820	343906	335130	335131
	NCBI Bio-Project Accession	PRJNA335132	PRJNA329819	PRJNA329820	PRJNA330706	PRJNA335130	PRJNA335131
	Project relevance	Environmental and biotechnological	Environmental and biotechnological	Environmental and biotechnological	Environmental and biotechnological	Environmental and biotechnological	Environmental and biotechnological

locus tag SAMN05421737_105223; accession number FMYM01000005.1), 1-acyl-sn-glycerol-3-phosphate acyltransferase (192 aa; locus tag SAMN05421737_105218; accession number FMYM01000005.1), choline dehyrogenase 402 aa; locus tag SAMN05421737_10199; accession number FMYM01000001.1).The strain, *Bacillus caseinilyticus* SP^{T} contains L-ectoine synthase (128 aa; locus tag SAMN05421736_104185; accession number FNPI01000004.1), glycerol-3-phosphate dehydrogenase (NAD(P)+ (338 aa; locus tag SAMN05421736_10427; accession number FNPI01000004.1), glycerol-1-phosphate dehydrogenase (351 aa; locus tag SAMN05421736_10511; accession number FNPI01000005.1), glycerol-1-phosphate dehydrogenase (561 aa; locus tag SAMN05421736_10785; accession number FNPI01000007.1), 1-acyl-sn-glycerol-3-phosphate acyltransferase (196 aa; locus tag SAMN05421736_10420; accession number FNPI01000004.1), choline dehyrogenase (569 aa; locus tag SAMN05421736_101228; accession number FNPI01000001.1). *Pelagirhabdus alkalitolerans* S5 contains ectoine synthase (129 aa; locus tag SAMN05421734_10588; accession number FMYI01000005.1), glycerol-3-phosphate dehydrogenase (NAD(P)+ (347 aa; locus tag SAMN05421734_102431; accession number FMYI01000002.1), 1-acyl-sn-glycerol-3-phosphate acyltransferase (243 aa; locus tag SAMN05421734_102368; accession number FMYI01000002.1/ 193 aa; locus tag SAMN05518683_10435; accession number FMYI01000002.1/ 237 aa; locus tag SAMN05518683_10446; accession number FMYI01000004.1). Whereas, strain *Salibacterium halotolerans* S7 codes for L- ectoine synthase (128 aa; locus tag SAMN05518683_10726; accession number FOXD01000007.1), ectoine hydroxylase (314 aa; locus tag SAMN05518683_10644; accession number FOXD01000006.1) and choline dehydrogenase (569 aa; locus tag SAMN05518683_10452; accession number FOXD01000004.1 / 265 aa; locus tag SAMN05518683_10464; accession number FOXD01000004.1 / 402 aa; locus tag SAMN05518683_12711; accession number FOXD01000027.1). *Salipaludibacillus aurantiacus* S9 codes for L- ectoine synthase (131 aa; locus tag SAMN05518684_11244; accession number FOGT01000012.1), ectoine hydroxylase (310 aa; locus tag SAMN05518684_10667; accession number FOGT01000006.1), choline dehydrogenase (563 aa; locus tag SAMN05518684_106165; accession number FOGT01000006.1) and betaine aldehydyde dehydrogenase (490 aa; locus tag SAMN05518684_106162; accession number FOGT01000006.1). These draft genome sequences will further help in understanding the genetic potential of the above described six strains belonging to *Firmicutes* for osmotolerance, especially in relation to ectoine and hydroxyectoine production which are very commercially viable products.

Nucleotide sequence accession numbers

The draft sequence of *Salisedimini bacterium haloalkalitolerans* 10nlg, *Bacillus lonarensis* 25nlg^T, *Bacillus caseinilyticus* SP^T, *Pelagirhabdus alkalitolerans* S5, *Salibacterium halotolerans* S7 and *Salipaludibacillus aurantiacus* S9 obtained in this Whole Genome Shotgun project has been deposited at GenBank under the accession numbers FOGV00000000, FMYM00000000, FNPI00000000, FOGV00000000, FOXD00000000 and FOGT00000000, respectively. The Genomes on Line Database (GOLD) IDs of the strains *Salisediminibacterium haloalkalitolerans* 10nlg, *Bacillus lonarensis* 25nlg^T, *Bacillus caseinilyticus* SP^T, *Pelagirhabdus alkalitolerans* S5, *Salibacterium halotolerans* S7 and *Salipaludibacillus aurantiacus* S9 are PRJEB16955, PRJEB15470, PRJEB16536, PRJEB15473, PRJEB17326 and PRJEB16964, respectively.

ACKNOWLEDGMENTS

We are thankful to Prof. William Whitman for proposing the genomic sequencing of these organisms under KMG-III. These organisms were isolated and described in Department of Science & Technology, Government of India, FTSYS project grants (SB/FT/LS-115/2012-13 and SB/FT/LS-320/2012-13, respectively).

REFERENCES

1. Field D, Garrity G, Gray T, Morrison N, Selengut J, Strek P, Tatusova T, Thomson N, *et al.*, The minimum information about a genomic sequence (MIGS) specification. *Nat. Biotechnol.,* 2008; **26:** 541-547.
2. Huntemann M, Ivanova NN, Mavromatis K, Tripp HJ, Paez-Espino D, Palaniappan K, *et al.*, The standard operating procedure of the DOE-JGI Microbial Genome Annotation Pipeline (MGAP v.4). *Stand. Genomic Sci.,* 2015; **10:**86.
3. Hyatt D, Chen GL, Locascio PF, Land ML, Larimer FW, Hauser LJ., Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinfo*., 2010; **11:**119.
4. MacCallum I, Przybylski D, Gnerre S, Burton J, Shlyakhter I, Gnirke A, *et al.*, ALLPATHS 2: small genomes assembled accurately and with high continuity from short paired reads. *Genome Biol.,* 2009; **10:**R103.
5. Markowitz VM, Chen IM, Palaniappan K, Chu K, Szeto E, Pillay M, *et al.*, IMG 4 version of the integrated microbial genomes comparative analysis system. *Nucleic Acids Res.,* 2014; **42(Database issue):**D560–7.
6. Mavromatis K, Land M, Brettin TS, Quest DJ, Copeland A, Clum A, *et al.*, The fast changing landscape of sequencing technologies and their impact on microbial assemblies and annotation. *PLoS One.,* 2012; **7:**e48837.
7. Reddy TBK, Thomas A, Stamatis D, Bertsch J, Isbandi M, Jansson J, *et al.*, The Genomes OnLine Database (GOLD) v.5: a metadata management system based on a four level (meta) genome project classification. *Nucleic Acids Res.,* 2015; **43:**D1099.
8. Sultanpuram, VR, Mothe, T, Mohammed, F., *Salisediminibacterium haloalkalitolerans* sp. nov., isolated from a soda lake, emended description of *Bacillus locisalis* as *Salisediminibacterium locisalis* comb. nov and emened description of genus *Salisediminibacterium*. *Arch. Microbiol.,* 2015a; **197:** 553-560.
9. Sultanpuram, VR, Mothe, T, Mohammed, F., Chintalapati, S, Chintalapati, VR., *Bacillus lonarensis* sp. nov., a haloalkalophilic bacterium isolated from a soda lake. *Arch. Microbiol.,* 2015b; **197:** 27-34.
10. Sultanpuram, VR, Mothe, T, Mohammed, F., *Bacillus casienilyticus* sp. nov., a bacterium isolated from a soda lake. *Int. J. Sys.Evolu. Microbiol.,* 2015c; **65 (8):** 2441-2446.
11. Sultanpuram, VR, Mothe, T, Chintalapati, S, Chintalapati, VR., *Salibacterium halotolerans* gen. nov., sp. nov., a novel bacterium isolated from a salt pan, reclassification of *Bacillus qingdaonensis* as *Salibacterum qingdaonense* comb. nov. and *Bacillus halochares* as *Salibacterium halochares* comb. nov. *Int. J. Syst. Evolu. Microbiol.,* 2015d; **65:** 4270-4275.
12. Sultanpuram, VR, Mothe, T., *Salipaludibacillus*

aurantiacus gen. nov., sp. nov., a novel alkali tolerant bacterium, reclassification of *Bacillus agaradharens* as *Salipaludibacillus agaradharens* comb. nov. and *Bacillus neizhouensis* as *Salipaludibacillus neizhouensis* comb. nov. *Int. J. Syst. Evolu. Microbiol.,* 2016; **66:** 2747-2753.

13. Sultanpuram, VR, Mothe, T, Chintalapati, S, Chintalapati, VR., *Pelagirhabdus alkalitolerans* gen. nov., sp. nov., a novel alkali and thermo tolerant bacterium isolated from Pingaleshwar beach, India and reclassification of *Amphibacillus fermentum* as *Pelagirhabdus fermentum* com. nov. *Int. J. Syst. Evolu. Microbiol.,* 2016; **66:** 84-90.
14. Whitman WB, Woyke T, Klenk HP, Zhou Y, Lilburn TG, Beck BJ, et al., Genomic encyclopedia of bacterial and archaeal type strains, phase III: the genomes of soil and plant associated and newly described type strains. *Stand. Genomic Sci.,* 2015; **10:** 26.
15. Zerbino DR, Birney E, Velvet: algorithms for de novo short read assembly using de bruijn graphs. *Genome Res.,* 2008; **18:**821–829.