

Genome-guided development of a bacterial two-strain system for low-temperature soil biocementation

Karol Ciuchcinski¹, Grzegorz Czerwonka², Przemyslaw Decewicz¹, Zofia Godlewska¹, Katarzyna Misiolek³, Katarzyna Zegadlo², Michal Styczynski¹, Lukasz Dziewit^{1,*}

¹ Department of Environmental Microbiology and Biotechnology, Institute of Microbiology, Faculty of Biology, University of Warsaw, Warsaw, Poland.

² Department of Microbiology, Institute of Biology, Faculty of Natural Sciences, Jan Kochanowski University, Kielce, Poland

³ Department of Hydro–Engineering and Hydraulics, Faculty of Building Services, Hydro and Environmental Engineering, Warsaw University of Technology, Warsaw, Poland

* Corresponding author:

Lukasz Dziewit; phone: +48 225542004; e-mail: l.dziewit@uw.edu.pl

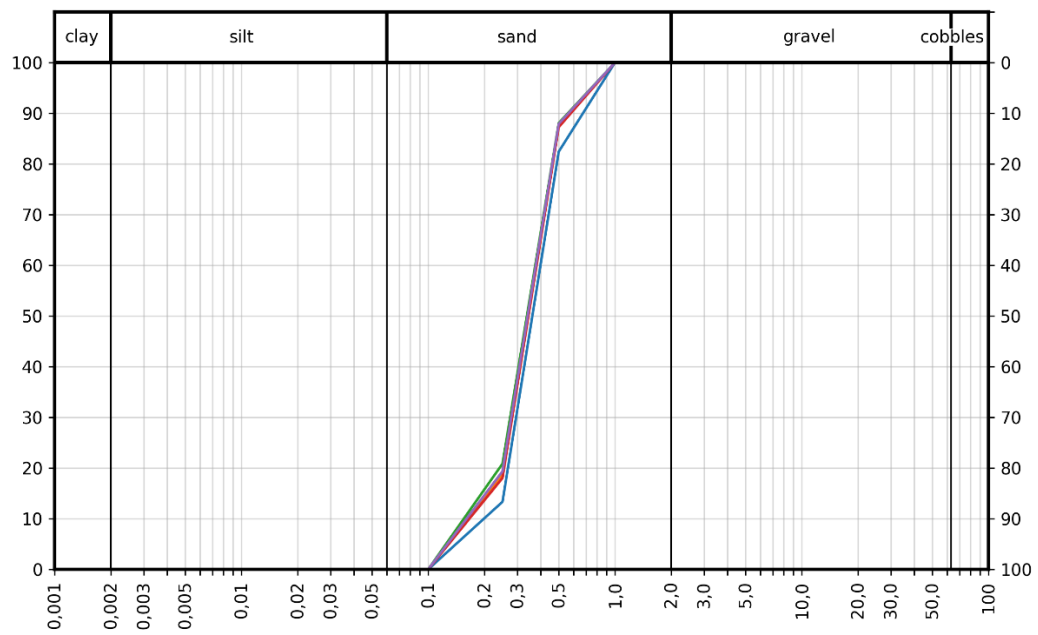


Figure S1. Sands grain-size distribution curves.



Figure S2. The grains of sand (marine origin) - light microscope photograph.

Table S1. General *Sporosarcina* sp. ANT_H38 genome characterisation.

Feature	Calculation
Number of contigs	5
Genome size (bp)	4,631,474 bp (chromosome); 6,243 bp (pA38H1); 10,106 bp (pA38H2); 10,532 bp (pA38H3); 27,152 bp (pA38H4)
GC content for chromosome (%)	39.9
Number of genes	4,512
Number of proteins with functional assignments	3,785
Number of proteins with EC number assignments	441
Number of tRNA genes	83
Number of regulatory RNA genes	104

Table S2. Genes involved in cold resistance identified within *Sporosarcina* sp. ANT_H38 genome.

Locus tag	Gene/protein name	Function	Reference database
GGGNBK_00030	<i>rplK</i>	50S ribosomal protein L11	CAPP database
GGGNBK_00170	unknown	hypothetical protein	CAPP database
GGGNBK_00175	<i>acoD</i>	Acetaldehyde dehydrogenase 2	CAPP database
GGGNBK_00790	<i>ureA</i>	Urease subunit gamma	CAPP database
GGGNBK_01250	<i>rpmH</i>	50S ribosomal protein L34	CAPP database
GGGNBK_01255	<i>dnaA</i>	DNA binding and replication initiator, global transcription regulator	Barria <i>et al.</i>
GGGNBK_02325	unknown	hypothetical protein	CAPP database
GGGNBK_04775	unknown	hypothetical protein	CAPP database
GGGNBK_05395	unknown	hypothetical protein	CAPP database
GGGNBK_05755	<i>icl</i>	Isocitrate lyase	CAPP database
GGGNBK_06320	<i>pseI</i>	Pseudaminic acid synthase	CAPP database
GGGNBK_07760	<i>sucD</i>	Succinate--CoA ligase [ADP-forming] subunit alpha	CAPP database
GGGNBK_08030	<i>nusA</i>	Transcription termination/antitermination/elongation L factor	Barria <i>et al.</i>
GGGNBK_08045	<i>infB</i>	Protein chain initiation factor IF2, translation initiation, fMet-tRNA binding, protein chaperone	Barria <i>et al.</i>
GGGNBK_08075	<i>pnp</i>	3'-5' exoribonuclease; component of RNA degradosome; cold shock protein required for growth at low temperatures	Barria <i>et al.</i>
GGGNBK_08180	<i>recA</i>	General recombination and DNA repair; induction of the SOS response	Barria <i>et al.</i> , CAPP database
GGGNBK_10195	<i>cspB</i>	Cold shock-inducible; function unknown	Barria <i>et al.</i> , custom cspABC database
GGGNBK_10335	<i>sodA</i>	Superoxide dismutase [Mn]	CAPP database
GGGNBK_11075	unknown	hypothetical protein	CAPP database
GGGNBK_11195	<i>katA</i>	Catalase	CAPP database
GGGNBK_11645	unknown	hypothetical protein	CAPP database
GGGNBK_11825	unknown	hypothetical protein	CAPP database
GGGNBK_11850	<i>cspA</i>	Cold-inducible RNA chaperone and anti-terminator; transcriptional enhancer	custom cspABC database

GGGNBK_14565	unknown	hypothetical protein	CAPP database
GGGNBK_15425	unknown	hypothetical protein	CAPP database
GGGNBK_15580	<i>infC</i>	Protein chain initiation factor IF3, translation initiation, stimulates mRNA translation	Barria <i>et al.</i>
GGGNBK_17365	unknown	hypothetical protein	CAPP database
GGGNBK_17800	<i>menB</i>	1,4-dihydroxy-2-naphthoyl-CoA synthase	CAPP database
GGGNBK_18125	unknown	hypothetical protein	CAPP database
GGGNBK_19845	<i>clpP</i>	ATP-dependent Clp protease proteolytic subunit	CAPP database
GGGNBK_20925	<i>purE</i>	N5-carboxyaminoimidazole ribonucleotide mutase	CAPP database
GGGNBK_22000	unknown	hypothetical protein	CAPP database
GGGNBK_22350	<i>cspA</i>	Cold-inducible RNA chaperone and anti-terminator; transcriptional enhancer	custom cspABC database
GGGNBK_22900	<i>rpsK</i>	30S ribosomal protein S11	CAPP database
GGGNBK_22915	<i>infA</i>	Protein chain initiation factor IF1, translation initiation	Barria <i>et al.</i> , CAPP database
GGGNBK_22940	<i>rpsE</i>	30S ribosomal protein S5	CAPP database
GGGNBK_22960	<i>rpsZ</i>	30S ribosomal protein S14 type Z	CAPP database
GGGNBK_22975	<i>rplN</i>	50S ribosomal protein L14	CAPP database
GGGNBK_22990	<i>rplP</i>	50S ribosomal protein L16	CAPP database
GGGNBK_23005	<i>rpsS</i>	30S ribosomal protein S19	CAPP database
GGGNBK_23030	<i>rpsJ</i>	30S ribosomal protein S10	CAPP database
GGGNBK_23040	<i>tufA</i>	Elongation factor Tu	CAPP database
GGGNBK_23055	<i>rpsL</i>	30S ribosomal protein S12	CAPP database

Table S3. Selected genes and their predicted functions (excluding replication and mobilisation for conjugal transfer) found within the ANT_H38 plasmids.

Locus tag / gene name	Plasmid	Predicted protein	Putative function
GGGNBK_23135	pA38H2	Restriction endonuclease	Restriction/modification system
GGGNBK_23150	pA38H2	N-acetylmuramoyl-L-alanine amidase	Biofilm formation modulation/phage defense
GGGNBK_23155	pA38H2	MerR family transcriptional regulator	Metal-dependent metabolism regulation
GGGNBK_23210	pA38H3	Carbonate dehydratase	pH regulation
GGGNBK_23180	pA38H3	Ktr system potassium transporter B	Potassium uptake/osmotic stress resistance
GGGNBK_23195	pA38H3	MerR family transcriptional regulator	Metal-dependent metabolism regulation
GGGNBK_23260	pA38H4	MerR family transcriptional regulator	Metal-dependent metabolism regulation
GGGNBK_23275	pA38H4	Putative triple helix repeat-containing collagen	Spore formation
GGGNBK_23295	pA38H4	ABC transporter	ATP-binding protein
GGGNBK_23300	pA38H4	UBA/THIF-type NAD/FAD binding protein	Thiamine biosynthesis
GGGNBK_23350	pA38H4	MerR family transcriptional regulator	Metal-dependent metabolism regulation

Table S4. Insertion sequences identified within the *Sporosarcina* sp. ANT_H38 genome.

Locus tag(s)	Coordinates	IR	DR	IS family (according to ISfinder database)
GGGNBK_01305	273848 - 275021	IRL:AAATGCAACAAC IRR:GTTGTTGCATTT	AAT	IS30
GGGNBK_02100, GGGNBK_02105	438906 - 440642	IRL:ACGAGGACACTGAAAAA IRR: TTTTTCAGTGTCTCTCGT	-	IS1182
GGGNBK_02205	457357 - 457662	Unknown	Unknown	IS150
GGGNBK_04295	903872 - 905443	Unknown	Unknown	No significant similarity to other transposases from ISFinder DB
GGGNBK_05415	1114155 - 1114442 (complement)	Unknown	Unknown	No significant similarity to other transposases from ISFinder DB
GGGNBK_08435	1754374 - 1755276 (complement)	Not found	Not found	ISL3
GGGNBK_09275, GGGNBK_09280	1947278 - 19438346	IRL:AAAATAAAAAAAGTC, IRR:GACTTTTTTTATTTT	Not found	IS150
GGGNBK_09340, GGGNBK_09345	1958701 - 1960420 (complement)	IRL:GGACGCTTTCCGCGGGCACGGC IRR:GCCGTGCCCGCGGAAAGCGTCC	GAT/GCT	ISL3
GGGNBK_12810	2618952 - 2619736 (complement)	Not found	Not found	IS1182
GGGNBK_14925	3094101 - 3094502 (complement)	Not found	Not found	IS21
GGGNBK_17765	3705396 - 3705704 (complement)	Unknown	Unknown	No significant similarity to other transposases from ISFinder DB

Table S5. Genes responsible for carbohydrate metabolism identified uniquely in *Sporosarcina* sp. ANT_H38 genome.

Gene name	Function	Loci name
<i>araA</i>	L-arabinose isomerase	GGGNBK_13170
<i>araB</i>	Ribulokinase	GGGNBK_13175
<i>araC</i>	Arabinose operon regulatory protein	GGGNBK_00325, GGGNBK_08640
<i>araP</i>	L-arabinose transport system permease protein	GGGNBK_13150
<i>araQ</i>	L-arabinose transport system permease protein	GGGNBK_13125, GGGNBK_19645
<i>bga</i>	Beta-galactosidase	GGGNBK_13295
<i>bglK</i>	Beta-glucoside kinase	GGGNBK_16080
<i>bglY</i>	Beta-galactosidase	GGGNBK_08660
<i>crr</i>	PTS system glucose-specific EIIA component	GGGNBK_00395, GGGNBK_22850
<i>iolE</i>	Inosose dehydratase	GGGNBK_13195, GGGNBK_13410
<i>iolG</i>	Inositol 2-dehydrogenase/D-chiro-inositol 3-dehydrogenase	GGGNBK_13270
<i>larA</i>	Lactate racemase	GGGNBK_02645
<i>licH</i>	putative 6-phospho-beta-glucosidase	GGGNBK_16075
<i>malG</i>	Maltose/maltodextrin transport system permease protein	GGGNBK_08655
<i>malP</i>	PTS system maltose-specific EIICB component	GGGNBK_00385
<i>malX</i>	PTS system maltose-specific EIICB component	GGGNBK_08590
<i>melA</i>	Alpha-galactosidase	GGGNBK_13070,GGGNBK_13075,GGGNBK_13080
<i>melD</i>	Melibiose/raffinose/stachyose import permease protein	GGGNBK_13885, GGGNBK_22850
<i>mngR</i>	Mannosyl-D-glycerate transport/metabolism system repressor	GGGNBK_16130
<i>yteP</i>	putative multiple-sugar transport system permease	GGGNBK_13475