Genome sequence and description of the heavy metal tolerant bacterium *Lysinibacillus sphaericus* strain OT4b.31

Tito David Peña-Montenegro, and Jenny Dussán*

Centro de Investigaciones Microbiológicas – CIMIC, Universidad de los Andes, Bogotá, Colombia

*Correspondence: jdussan@uniandes.edu.co

Keywords: *Lysinibacillus sphaericus* OT4b.31, DNA homology, *de novo* assembly, heavy metal tolerance, Sip1A coleopteran toxin

Lysinibacillus sphaericus strain OT4b.31 is a native Colombian strain having no larvicidal activity against *Culex quinquefasciatus* and is widely applied in the bioremediation of heavy-metal polluted environments. Strain OT4b.31 was placed between DNA homology groups III and IV. By gap-filling and alignment steps, we propose a 4,096,672 bp chromosomal scaffold. The whole genome (consisting of 4,856,302 bp long, 94 contigs and 4,846 predicted protein-coding sequences) revealed differences in comparison to the *L. sphaericus* C3-41 genome, such as syntenial relationships, prophages and putative mosquitocidal toxins. Sphaericolysin B354, the coleopteran toxin Sip1A and heavy metal resistance clusters from *nik, ars, czc, cop, chr, czr* and *cad* operons were identified. *Lysinibacillus sphaericus* OT4b.31 has applications not only in bioremediation efforts, but also in the biological control of agricultural pests.

Introduction

Biological control of vector-borne diseases, such as dengue and malaria, and agricultural pests have been an issue of special concern in the recent years. Since Kellen et al. [1] first described *Lysinibacillus sphaericus* as an insect pathogen, studies have shown mosquitoes to be the major target of this bacterium [2-4], but toxic activity against other species has also been reported [5,6]. *L. sphaericus* larvicidal toxicity has been reported due to vegetative mosquitocidal toxins (Mtx) [7], the binary toxin (BinA/BinB) [4], Cry48/Cry49 toxin [8] and recently the S-layer protein [9]. To date, no larvicidal activity has been identified in *Lysinibacillus sphaericus* OT4b.31 against *Culex quinquefasciatus* [10].

On the other hand, *Lysinibacillus* species are potential candidates for heavy metal bioremediation. Some *Bacillaceae* strains have been successfully isolated from nickel contaminated soil [11], industrial landfills [12], naturally metalliferous soils [13] and a uranium-mining waste pile [14]. In addition, native Colombian *Lysinibacillus* strains have been reported as potential metal bioremediators. Strain CBAM5 is resistant to arsenic, up to 200 mM, and contains the arsenate reductase gene [15]. *L. sphaericus* OT4b.31 showed heavy metal biosorption in living and dead biomass. The S-layer protein was also shown to be present [16]. We observed 19 mosquito-pathogenic *L. sphaericus* strains and 6 non-pathogenic strains (including OT4b.31) that were able to grow in arsenate, hexavalent chromium and/or lead [17]. The moderate heavy metal tolerance in a *Lysinibacillus* strain isolated from a non-polluted environment generates interest in characterizing the genomic properties of *L. sphaericus* OT4b.31, in addition to its biotechnological potential in biological control.

Here we present a summary classification and a set of features for *Lysinibacillus sphaericus* OT4b.31 including previously unreported aspects of its phenotype, together with the description of the complete genomic sequencing and annotation.

Classification and features

Formerly known as *Bacillus sphaericus*, the species was defined as having a spherical terminal spore and by its inability to ferment sugars [18]. According to physiological and phylogenetic analysis, it was reassigned to the genus *Lysinibacillus* [19]. Strains of *L. sphaericus* can be divided into five DNA homology groups (I–V). Some mosquito pathogenic strains are allocated in subgroup II-A, while *Lysinibacillus fusiformis* species is in subgroup II-B [20]. Later, according to 16S rDNA and lipid profile comparisons, *Lysinibacillus sphaericus sensu lato* was classified into seven similarity subgroups, of which only four retained the previous description by Krych et al. [21]. Recently, by using 16S rDNA phylogenetic analysis some mosquito pathogenic native strains were found in group II with heterogeneous heavy metal tolerance levels. [17].

Partial 16S rRNA gene sequences (1,421 bp) were aligned to establish the phylogenetic neighborhood of *Lysinibacillus sphaericus* OT4b.31 (Figure 1). The phylogenetic tree was constructed by neighbor-joining [23] using the SEAVIEW [24] and TreeGraph2 [25] packages. Genetic distances were estimated by using the Jukes-Cantor model [23]. The stability of relationships was assessed by bootstrap analysis based on 1,000 resamplings for the tree topology. Interestingly, *L. sphaericus* OT4b.31 did not fall into any existing DNA similarity group; it was found between DNA similarity groups III and IV [21]. Consistent with Lozano & Dussán [17], *L. sphaericus* OT4b.31 did not fall into DNA similarity groups I, II or III.

Dussán et al. [10] evaluated physiological diversity and genetic potential in native *Bacillaceae* isolates from highlands of the Colombian Andes, where *Lysinibacillus sphaericus* OT4b.31 was first described (Table 1). *L. sphaericus* OT4b.31 is an aerobic free-living bacterium isolated from coleopteran (beetle) larvae collected in the highlands of the Colombian Andes [10]. Vegetative cells stain Gram positive, but in sporulating stages, cell stain Gram variable (Figure 2). By using a JEOL JSM-5800LV (Japan) scanning electron microscope, *L. sphaericus* OT4b.31 is estimated to measure 0.61 to 0.65 µm in width and 1.9 to 2.3 µm long (Figure 3). *L. sphaericus* HOT4b.31 showed slow sporulation rates (undetectable up to 40 hours of growth) and positive evidence of binary toxin which does not exhibit larvicidal activity against *Culex quinquefasciatus* [10]. Cultures grow at 10 to 40°C over a pH range of 6.0 to 9.0. Antibiotic resistance was evaluated separately by adding filter sterilized antibiotic solutions in Luria-Bertani broths and checking turbidity after 15 hours of growth. *L. sphaericus* OT4b.31 is sensitive to kanamycin (12.5 µg/mL), chloramphenicol (25 µg/mL), erythromycin (5 µg/mL), and gentamicin (25 µg/mL), while it showed resistance to trimethoprim/sulfamethoxazol up to 30 µg/mL/150 µg/mL.

Genome sequencing information

Genome project history

The genome sequencing of *Lysinibacillus sphaericus* OT4b.31 was supported by the CIMIC (Centro de Investigaciones Microbiológicas) laboratory at the University of Los Andes within the Grant (1204-452-21129) of the Instituto Colombiano para el fomento de la Investigación Francisco José de Caldas. Whole genomic DNA extraction and bioinformatics analysis was performed at CIMIC laboratory, whereas libraries construction and whole shotgun sequencing at the Beijing Genome Institute (BGI) Americas Laboratory (Tai Po, Hong Kong). The applied pipeline included quality check of reads, de novo assembly, a gap-filling step and mapping against a reference genome. This whole genome shotgun project has been deposited at DDBJ/EMBL/GenBank under the accession AQPX00000000. The version described in this paper is the first version, AQPX01000000. A summary of the project information is shown in Table 2.

Growth conditions and DNA isolation

Lysinibacillus sphaericus strain OT4b.31 was grown in nutrient broth for 16 hours at 30ºC and 150 rev/min. High molecular weight DNA was isolated using the EasyDNA® Kit (Carlsbad, CA, USA. Invitrogen) as indicated by the manufacturer. DNA purity and concentration were determined in a NanoDrop spectrophotometer (Wilmington, DE, HUSA. Thermo Scientific).

Table 1. Classification and general features of *Lysinibacillus sphaericus* OT4b.31 according to the MIGS recommendations [26]

MIGS ID	Property	Term	Evidence code[a]
		Domain *Bacteria*	TAS [27]
		Phylum *Firmicutes*	TAS [28-30]
		Class *Bacilli*	TAS [31,32]
	Current classification	Order *Bacillales*	TAS [33,34]
		Family *Bacillaceae*	TAS [33,35]
		Genus *Lysinibacillus*	TAS [19,36]
		Species *Lysinibacillus sphaericus*	TAS [19,37]
		Type strain OT4b.31	TAS [10]
	Gram stain	Positive in vegetative cells, variable in sporulating stages	IDA
	Cell shape	Straight rods	IDA
	Motility	Non-motile	IDA
	Sporulation	Sporulating	IDA
	Temperature range	Mesophile, grows > 14°, < 37°C	TAS [10]
	Optimum temperature	30°C	TAS [10]
	Carbon source	Complex carbohydrates	TAS [10]
	Energy metabolism	Heterotroph	TAS [10]
MIGS-6	Habitat	Coleopteran (beetle) larvae	TAS [10]
MIGS-6.3	Salinity	Growth in Luria-Bertani broth (5% NaCl)	IDA
MIGS-22	Oxygen requirement	Aerobic	TAS [10]
MIGS-15	Biotic relationship	Free living	TAS [10]
MIGS-14	Pathogenicity	Unknown	TAS [10]
MIGS-4	Geographic location	Tenjo, Cundinamarca, Colombia	TAS [10]
MIGS-5	Sample collection time	1995	TAS [10]
MIGS-4.1	Latitude	4.88727	TAS [10]
MIGS-4.2	Longitude	-74.132831	TAS [10]
MIGS-4.3	Depth	Surface	TAS [10]
MIGS-4.4	Altitude	2,685 m above sea level	TAS [10]

a) Evidence codes - IDA: Inferred from Direct Assay; TAS: Traceable Author Statement (i.e., a direct report exists in the literature); NAS: Non-traceable Author Statement (i.e., not directly observed for the living, isolated sample, but based on a generally accepted property for the species, or anecdotal evidence). These evidence codes are from the Gene Ontology project [38].

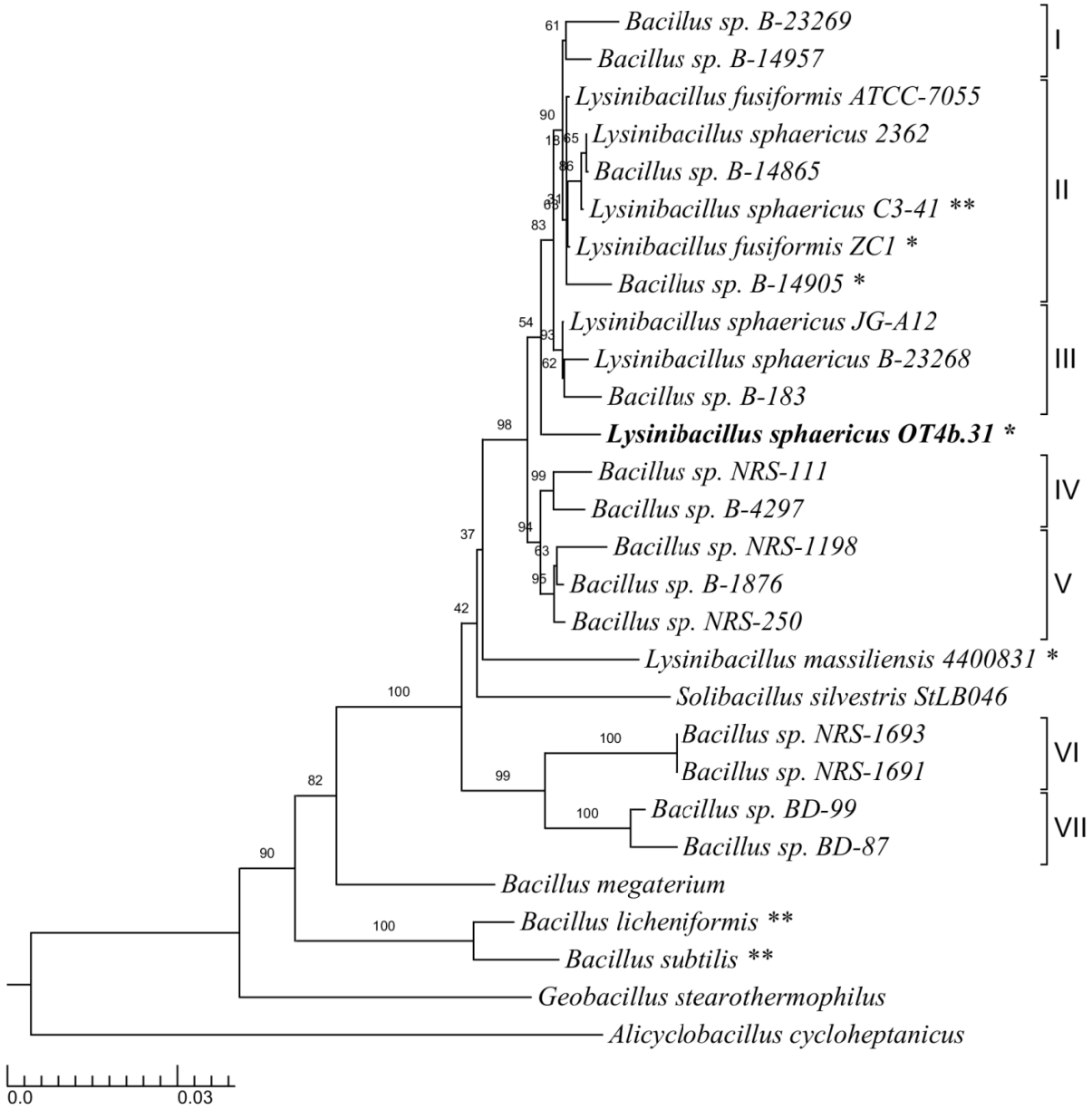

Figure 1. Phylogenetic tree highlighting the position of *Lysinibacillus sphaericus* OT4b.31 relative to the available type strains and other non-assigned species within the families *Alicyclobacillaceae* and *Bacillaceae*. *Alicyclobacillus cycloheptanicus* was designated as the outgroup species for the analyses. Right brackets encompass each homology group (I–VII) according to Nakamura's benchmarks [21]. Nucleotide sequences obtained from GenBank and used in the phylogenetic analyses were as follows: *Alicyclobacillus cycloheptanicus* 1457 (X51928), *Geobacillus stearothermophilus* 10 (X57309), *Bacillus subtilis* 168^{T} (X60646), *Bacillus licheniformis* DSM 13^{T} (X68416), *Bacillus megaterium* IAM 13418^{T} (D16273), *Bacillus sp.* BD-87 (AF169520), *Bacillus sp.* BD-99 (AF169525), *Bacillus sp.* NRS-1691 (AF169531), *Bacillus sp.* NRS-1693 (AF169533), *Solibacillus silvestris* StLB046 (NR_074954), *Lysinibacillus massiliensis* 4400831 (NR_043092), *Bacillus sp.* NRS-250 (AF169536), *Bacillus sp.* B-1876 (AF169494), *Bacillus sp.* NRS-1198 (AF169528), *Bacillus sp.* B-4297 (AF169507), *Bacillus sp.* NRS-111 (AF169526), *Lysinibacillus sphaericus* OT4b.31 (AQPX00000042.1:91-1546), *Bacillus sp.* B-183 (AF169493), *Lysinibacillus sphaericus* B-23268^{T} (AF169495), *Lysinibacillus sphaericus* JG-A12 (AM292655), *Bacillus sp.* B-14905 (AF169491), *Lysinibacillus sphaericus* ZC1 (NZ_ADJR01000054.1:1-1487), *Lysinibacillus sphaericus* C3-41 (NC_010382.1:16887-18287), *Bacillus sp.* B-14865 (AF169490), *Lysinibacillus sphaericus* 2362 (L14011), *Lysinibacillus fusiformis* ATCC-7055 (AJ310083), *Bacillus sp.* B-14957 (AF169492) and *Bacillus sp.* B-23269 (AF169496). The branches are scaled in terms of the expected number of substitutions per site. Numbers adjacent to the branches represent percentage bootstrap values based on 1,000 replicates. Lineages with type strain genome sequencing projects registered in GOLD [22] are labeled with one asterisk, those also listed as 'Complete and Published' with two asterisks.

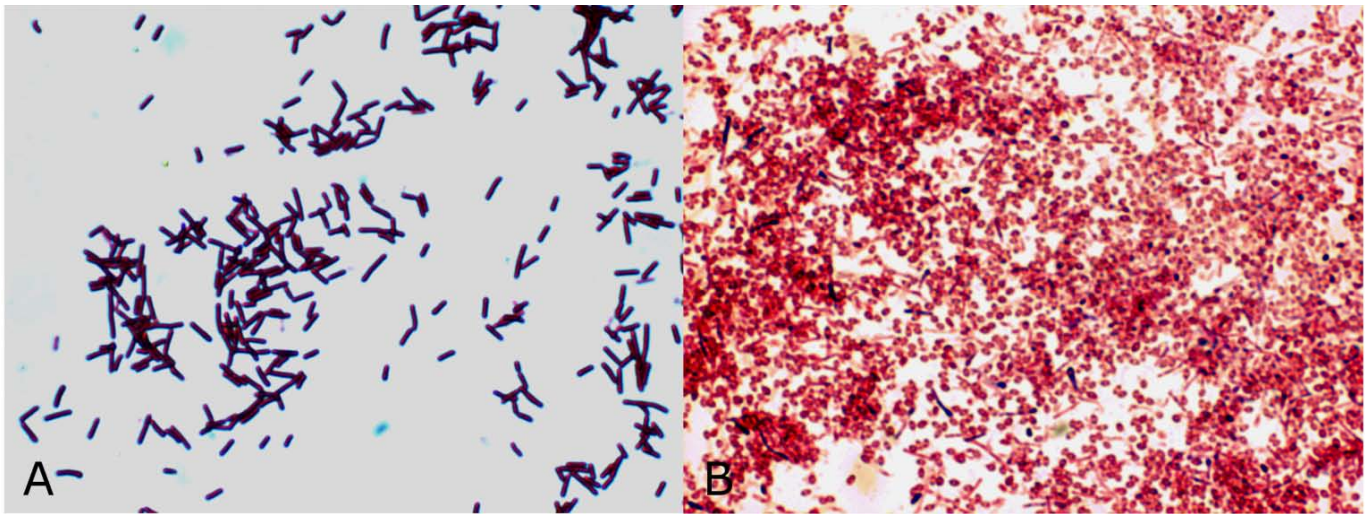

Figure 2. Gram staining of (A) vegetative cells and (B) spores of *Lysinibacillus sphaericus* OT4b.31.

Genome sequencing and assembly

After DNA extraction, samples were sent to the Beijing Genome Institute (BGI) Americas Laboratory (Tai Po, Hong Kong). Purified DNA sample was first sheared into smaller fragments with a desired size by a Covaris E210 ultrasonicator. Then the overhangs resulting from fragmentation were converted into blunt ends by using T4 DNA polymerase, Klenow Fragment and T4 polynucleotide kinase. After adding an "A" base to the 3' end of the blunt phosphorylated DNA fragments, adapters were ligated to the ends of the DNA fragments. The desired fragments were purified though gel-electrophoresis, then selectively enriched and amplified by PCR. The index tag was introduced into the adapter at the PCR stage as appropriate, and a library quality test was performed. Lastly, qualified, short, paired-ends of 90:90 bp length with 500 bp insert libraries were used to cluster preparation and to conduct whole-shotgun sequencing in Illumina Hi-Seq 2000 sequencers.

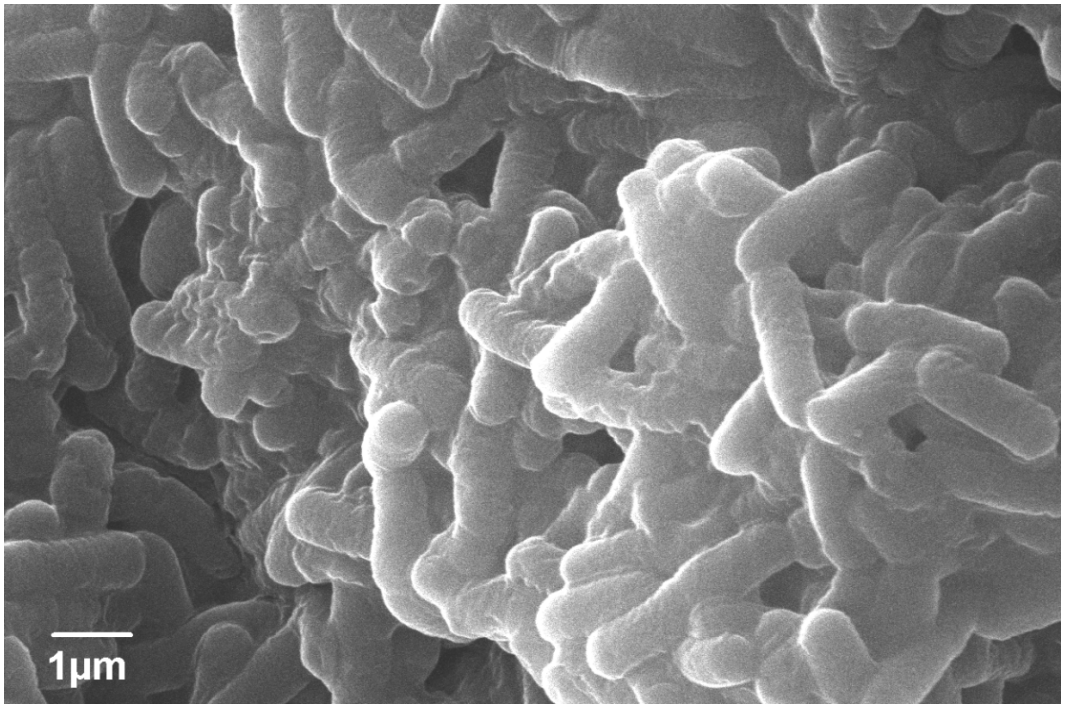

Figure 3. Scanning electron micrograph of *Lysinibacillus sphaericus* OT4b.31 at an operating voltage of 20 kV.

Using the FASTX-Toolkit version 0.6.1 [39] and clean_reads version 0.2.3 from the ngs_backbone pipeline [40] reads were trimmed and quality filtered. Then, with the CLC Assembly Cell version 4.0.10 [41], assembly and scaffolding steps were conducted via a de novo assembly pipeline. The assembly included automatic scaffolding and k-mer/overlapping optimization steps. Some gaps were successfully filled by using GapFiller [42] within 30 iterations. No more gaps reached convergence by running more iterations. To obtain structural insight of a chromosomal scaffold, we used CONTIGuator.2 [43], using the *Lysinibacillus sphaericus* strain C3-41 chromosome (accession number: CP000817.1) as reference. Gap-filling steps and mapping to reference sequences were performed again to confirm convergence. Quality assessment of the assembly was performed with iCORN [44]. The error rate of the final assembly is less than 1 in 1,000,000. Lastly, by using PROmer from the MUMmer [45] and Mauve [46] packages, we compared the chromosomal assembly and the chromosome of *L. sphaericus* C3-41.

Genome annotation

The Glimmer 3 gene finder was used to identify and extract sequences for potential coding regions. To achieve the functional annotation steps, the RAST server [47] and Blast2GO pipelines [48] were used. Blast2GO performed the blasting, GO-mapping and annotation steps; which included a description according to the ProDom, FingerPRINTScan, PIR-PSD, Pfam, TIRGfam, PROSITE, ProDom, SMART, SuperFamily, Pattern, Gene3D, PANTHER, SignalIP and TM-HMM databases. The results were summarized with InterPro [49]. Additionally, a GO-EnzymeCode mapping step was used to retrieve KEGG pathway-maps. tRNA genes were identified by using tRNAscan-SE [50] and rRNA genes by using RNAmmer [51]. The possible orthologs of the genome were identified based on the COG database and classified accordingly [52]. Prophage region prediction was also conducted by using the PHAST tool [53].

Genome properties

The genome summary and statistics are provided in Tables 3 and 4 and Figure 4. The genome consists of 96 scaffolds in 4,856,302 bp total size with a GC content of 37.5%. A total of 23 scaffolds were successfully aligned to a reference sequence, comprising 4,096,672 bp of sequence and are represented by the red and blue bars within the outer ring of Figure 4. Of the 4,938 genes predicted, 4,846 were protein-coding genes, 46 RNAs, and 1,623 pseudogenes were identified. Genes assigned a putative function comprised 67.13% of the protein-coding genes while the remaining ones were annotated as hypothetical proteins. The distribution of genes into COGs functional categories is presented in Table 5.

Table 2. Genome sequencing project information

MIGS ID	Property	Term
MIGS-31	Finishing quality	Improved high-quality draft
MIGS-28	Libraries used	One paired end tags 90:90 bp with 500 bp insert
MIGS-29	Sequencing platforms	Illumina Hi-Seq 2000
MIGS-31.2	Fold coverage	100×
MIGS-30	Assemblers	CLC Assembly Cell version 4.0.10
MIGS-32	Gene calling method	Glimmer3, tRNAscan-SE
	Genbank ID	AQPX00000000
	Genbank Date of Release	May 10, 2013
	GOLD ID	Gi39289
	Project relevance	Biotechnology, metabolic pathway

Table 3. Summary of genome

Label	Size (Mb)	Topology	INSDC identifier
Chromosomal scaffold	4,096,672	Circular	KB933398.1
Extrachromosomal elements	759,630	Linear	KB933399.1-KB933469.1

Table 4. Nucleotide content and gene count levels of the genome

Attribute	Value	% of total[a]
Genome size (bp)	4,856,302	100.00
DNA GC content (bp)	1,821,262	37.50
DNA coding region (bp)	3,924,297	80.81
Number of replicons	1	
Extrachromosomal	0	
Total genes	4,938	100
RNA genes	46	0.93
rRNA operons	7	
tRNA genes	38	0.77
Pseudogenes	1,623	32.87
Protein-coding genes	4,846	98.14
Genes in paralog clusters	658	13.33
Genes assigned to COGs	2,946	59.66
1 or more conserved domains	2,946	59.66
2 or more conserved domains	529	10.71
3 or more conserved domains	98	1.98
Genes with function prediction	3,315	67.13
Genes assigned Pfam domains	2,799	56.68
Genes with signal peptides	1,206	24.42
Genes with transmembrane helices	1,206	24.42
CRISPR repeats	0	0.00

a) The total is based on either the size of the genome in base pairs or the total number of protein coding genes in the annotated genome.

Table 5. Number of genes associated with the 25 general COG functional categories

Code	Value	%age[a]	Description
J	180	3.80	Translation
A	118	2.49	RNA processing and modification
K	354	7.48	Transcription
L	167	3.53	Replication, recombination and repair
B	1	0.02	Chromatin structure and dynamics
D	37	0.78	Cell cycle control, mitosis and meiosis
Y	0	0	Nuclear structure
V	75	1.58	Defense mechanisms
T	293	6.19	Signal transduction mechanisms
M	159	3.36	Cell wall/membrane biogenesis
N	95	2.01	Cell motility
Z	31	0.66	Cytoskeleton
W	28	0.59	Extracellular structures
U	48	1.01	Intracellular trafficking and secretion
O	96	2.03	Posttranslational modification, protein turnover, chaperones
C	169	3.57	Energy production and conversion
G	146	3.09	Carbohydrate transport and metabolism
E	351	7.42	Amino acid transport and metabolism
F	85	1.80	Nucleotide transport and metabolism
H	142	3.00	Coenzyme transport and metabolism
I	133	2.81	Lipid transport and metabolism
P	273	5.77	Inorganic ion transport and metabolism
Q	98	2.07	Secondary metabolites biosynthesis, transport and catabolism
R	450	9.51	General function prediction only
S	234	4.95	Function unknown
-	1,694	37.74	Not in COGs

a) The total is based on the total number of protein coding genes in the annotated genome.

Insights into the genome

To complete the assembly process, a resequencing pipeline was applied that set whole genome sequences as references such as *Lysinibacillus sphaericus* C3-41, *Bacillus sp.* strain B-14905, *Bacillus sp.* NRRL B-14911, *Bacillus megaterium* QM B1551, *Bacillus anthracis* Ames, *Lysinibacillus boronitolerans* F1182 and *Lysinibacillus fusiformis* ZC1. Mapping coverage was lower than 30% in any case (data not shown). In addition, GC content, and depth–GC correlation analysis demonstrated neither a biased distribution nor heterogeneity in the GC content of raw data. Thus, a *de novo* assembly was conducted in the CLC Assembly Cell version 4.0.10, as discussed above, resulting in a 123-scaffold assembly with a N50=96,816 bp. After the gap-filling step, all intrascaffold gaps and 29 interscaffold gaps were closed, leaving 94 scaffolds with a N50=205,086 bp. Finally, a mapping step was conducted using the sequences mentioned above as references. This yielded 26 supercontigs that mapped to *L. sphaericus* strain C3-41 chromosome corresponding to 88.9% of the reference chromosome. This alignment was proposed as a chromosomal scaffold. Other reference sequences lead to no significant coverage levels and extrachromosomal scaffolds did not align to previously sequenced plasmids of related species (data not shown). Chromosomal comparison from the PROmer analysis between *L. sphaericus* strains OT4b.31 and C3-41 showed that most of the two chromosomes mapped onto each other, revealing large segments of high similarity (Figure 5). However, a region comprising around 2 to 3.25 Mbp in the C3-41 chromosome and the contigs 15 to 19 in

the chromosomal scaffold were remarkably scattered in the dot-plot, revealing low coverage levels and different syntenial relationships to the reference sequence.

The origin of replication of the chromosome of *L. sphaericus* OT4b.31 was estimated by similarities to several features of the corresponding regions in *L. sphaericus* C3-41, *Bacillus sp.* B-14905 and other close related bacteria, including colocalization of the genes: *dnaX*, *recR*, *holB, dnaA*, *recG* and *recA*; and GC nucleotide skew [(G–C)/(G+C)] analysis. In the first 40 Kbp of contig 1, we found *dnaX, recR,* and *holB,* while *dnaA, recG* and *recA* were found at the end (after 290 Kbp) of contig 13. This may suggest that contig 13 should be allocated immediately before contig 1. Besides, there was no evidence of multiple *dnaA* boxes around the potential origin.

The replication termination site of the chromosomal scaffold is believed to be localized near 2.5 Mbp in the contig 18, according to GC skew analysis, and the coding bias for the two strands of the chromosome is for the majority of CDSs to be on the outer strand from 0 to ~2.5 Mbp and on the inner strand from ~2.5 Mbp to the end of the chromosomal scaffold (contig 26, Figure 4). This was also confirmed by the presence of *parC* (H131_12178) and *parE* (H131_12183), which encode the subunits of the chromosome-partitioning enzyme topoisomerase IV [54]. Similar to the *L. sphaericus* C3-41 genome [55], we did not find the homolog of *rtp* (replication terminator protein-encoding gene) in the chromosomal assembly of OT4b.31.

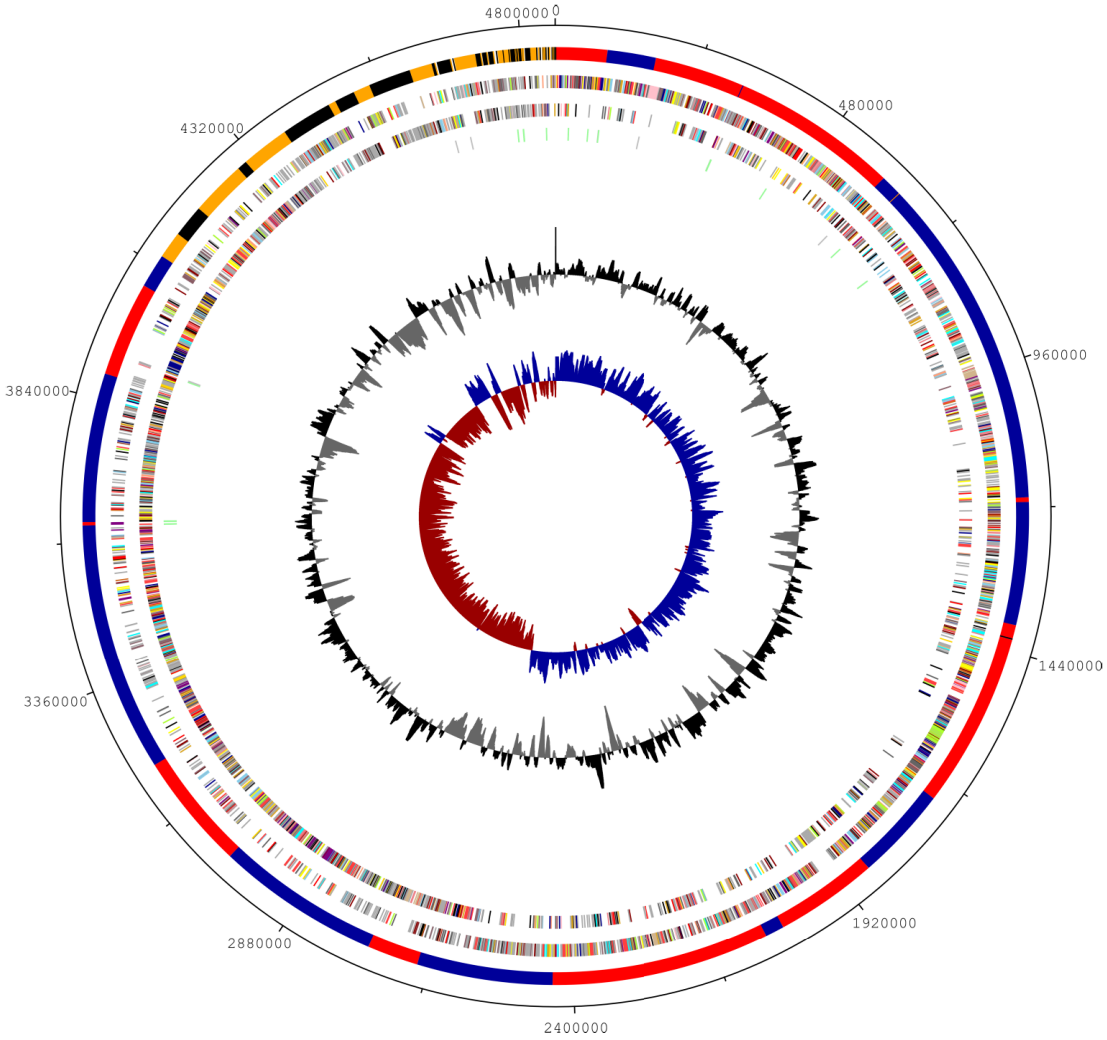

Figure 4. Graphical map of the genome. From outside to the center: Ordered and oriented scaffolds assigned to chromosome in blue and red, extrachromosomal scaffolds in orange and black, Genes on forward strand (color by COG categories), Genes on reverse strand (color by COG categories), RNA genes (tRNAs green, rRNAs gray), GC content and GC skew.

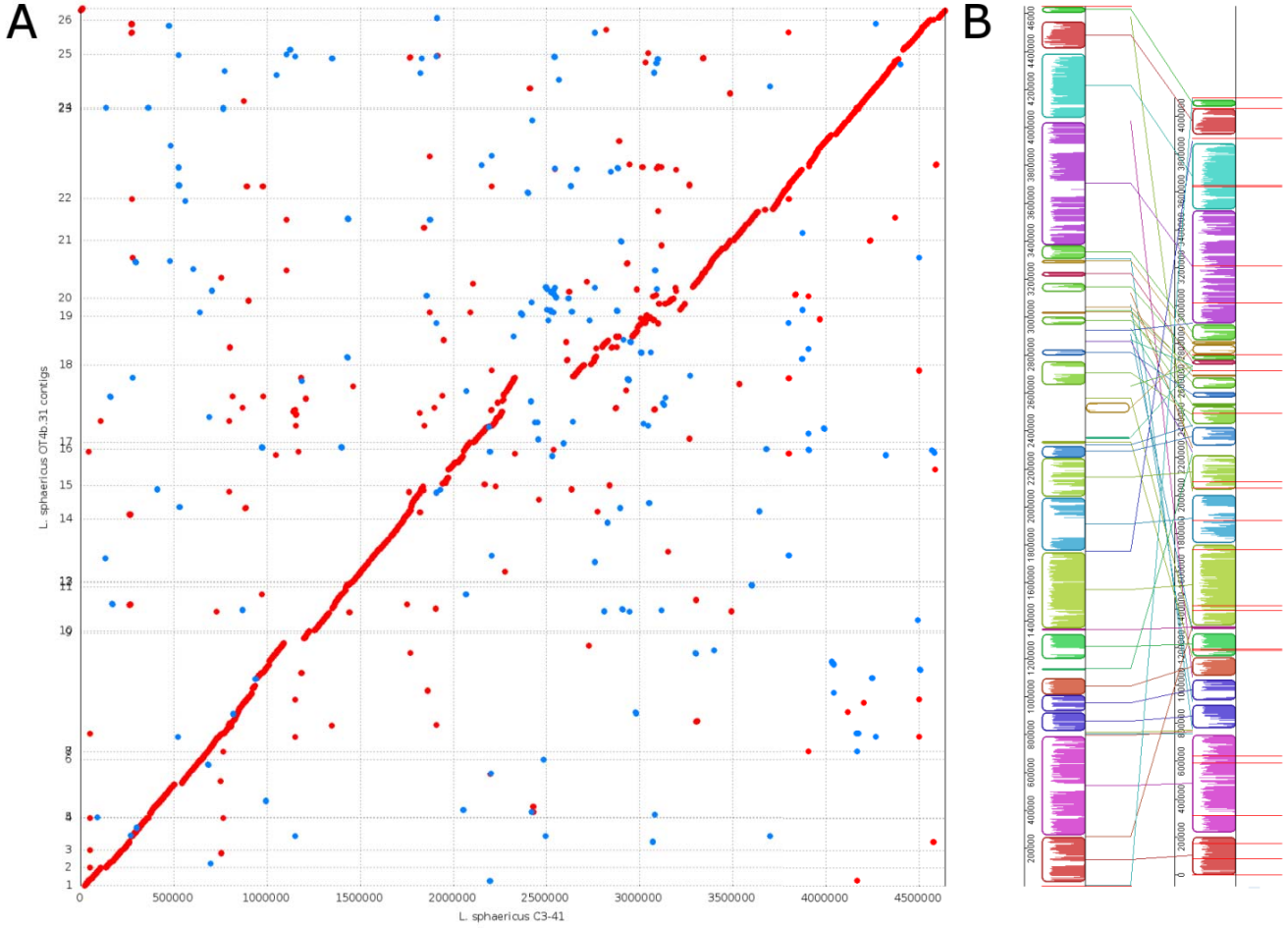

Figure 5. (A) Dot-plot of amino-acid-based alignment of a 4.09 Mbp chromosomal scaffold of *L. sphaericus* OT4b.31 (y-axis) to a 4.6 Mbp chromosome of *L. sphaericus* C3-41 (x-axis). Aligned segments are represented as dots or lines. Forward matches are plotted in red, reverse matches in blue. Figure generated by PROmer [45]. (B) Nucleotide-based alignment of a 4.09 Mbp chromosomal scaffold of *L. sphaericus* OT4b.31 (right) to a 4.6 Mbp chromosome of *L. sphaericus* C3-41 (left). A total of 27 homologous blocks are shown as identically colored regions and linked across the sequences. Regions that are inverted relative to *L. sphaericus* OT4b.31 are shifted to the right of center axis of the sequence. The origin of replication in each sequence is approximately at coordinate 1. Red bars show the limits of each contig in the chromosomal scaffold. Contigs 1 to 26 are numbered in ascending order start in coordinate 1. The figure was generated by Mauve [46].

A total of 42 hypothetical protein coding sequences were assigned as putative transposable elements, with the most frequent families being IS66, IS110, IS1272 and IS3. In addition, five prophage regions were identified, of which one region is intact and 4 regions are incomplete. *Lactobacillus* phage C5 (intact), *Bacillus* phage φ105, *Clostridium* phage c-st, *Bacillus* Phage SPP1 and *Bacillus* phage Wβ predicted regions were allocated at contigs 34, 8, 15, 18 and 37, respectively. Only lysis proteins were predicted in phages C5 and c-st regions. The only genes remaining in the phage φ105 region are those for coat proteins, integrase, and hypothetical and phage-like coding sequences. This is probably the remnant of phage invasion and genome deterioration during evolution. In addition, any previously reported phages in the genome of *L. sphaericus* C3-41 are in the genome of OT4b.31.

Two elements contain conserved domains from the *Listeria* pathogenicity island LIPI-1, functionally assigned as a thiol-activated cytolysin and a phosphatidylinositol phospholipase C. The first was confirmed to correspond to the *L. sphaericus* B354 sphaericolysin coding gene in contig 18

(H131_12483). Sphaericolysin B354 has been reported to be widespread across *L. sphaericus* DNA homology groups not only including IIA, IIB, IV and V [56] but also non-grouped species such as OT4b.31. Upstream, in the same contig, a *Bacillus* toxin from the family Mtx2 (PFam PF03318) was found and described as a hypothetical Sip1A toxin coding sequence (H131_12498). Purified from *Bacillus thuringiensis* strain EG2158, Sip1A is a secreted insecticidal protein of 38 KDa having activity against Colorado Potato beetle (*Leptinotarsa decemlineata*) [57]. Considering that *L. sphaericus* OT4b.31 was isolated from beetle larvae, we suggest potential coleopteran larvicidal activity. To our knowledge, strain OT4b.31 is the first report of a predicted Sip1A-like toxin in a native *Lysinibacillus sphaericus*. Unexpectedly, *mtx* or *bin* mosquito pathogenic genes were not found in the OT4b.31 genome, despite a previous report showing positive evidence of BinA/B toxins with no larvicidal activity [10].

A total of 32 CDSs were described as surface (S) layer proteins or S-layer homologs (SLH). The putative S-layer gene *sllB* (H131_05299) previously reported in *L. sphaericus* JG-A12 [58] was found in a 3,696 bp sequence allocated in contig 8. Three sequences with conserved domains similar to Slp5 and Slp6 were identified in contigs 8 (H131_05339, H131_05344) and 22 (H131_16838). *Bacillus sp.* B-14905 was the most similar sequence for the majority of S-layer protein domains. In addition, a putative glycoprotein (H131_22117), a bifunctional periplasmic precursor (H131_05993) and an S-layer fusion (H131_05409) coding sequence associated with S-layer proteins were recognized. On the other hand, a cluster of spore germination genes were determined near the termination of the replication site (including genes from the *ger* and *ype* operons) among other genes widespread in the genome. Three clusters of sporulation genes were allocated at contigs 1, 10 and 13 (including genes from *spo*II, *spo*V, *yaa* and *sig* operons).

Responses against toxic metal(oid)s in *L. sphaericus* OT4b.31 could be controlled by efflux pumps related genes in clusters found in contigs. Putative coding sequence order is as follows: *yozA*→*czcD*→*csoR*→*copZA* (contig 1, H131_00045: H131_00065); *nikABC*→*oppD*→*nikD* (contig 17, H131_11103:H131_11123); *cadC*-like→*cadA* (contig 24, H131_17086:H131_17081); *arsRBC* – putative extracellular secreted protein CDS – *arsR*-like→*arsR*-like→ putative excinuclease CDS (contig 18, H131_11998:H131_12028). The function of YozA is still unknown [59], but is similar to CzrA and CadC belonging to the ArsR transcriptional family regulators. YozA, CsoR (from the copper-sensitive operon), CadC-like and ArsR proteins seem to be the direct regulators of each cluster. At least one additional copy of ChrA, CzrB and CzcD CDSs were found. Upstream the *nik* cluster, we could not find transcriptional regulators. In summary, *L. sphaericus* OT4b.31 has protein encoding sequences probably involved in the resistance against Cd, Zn, Co, Cu, Ni, Cr, and As. In fact, prior reports of resistance to toxic metals [16,17] in *L. sphaericus* OT4b.31 may be explained due to participation of heavy-metal resistance proteins.

Strain OT4b.31 probably has a diverse defense repertoire according to the following responses and predicted genes: bacitracin stress responses, genes *bceBASR* and *yvcPQRS;* multidrug resistance, MATE (multidrug and toxin extrusion) family efflux pump genes *ydhE*/*norM* and *acrB*; antibiotics resistance, genes *vanRSW*, *tetP*-like group II, *fusA* (elongation factor G), *fosB*, *blaZ* and *ampC*-like. Based in the KEGG analysis, some predicted proteins might participate in peripheral pathways for the degradation of benzoate, aminobenzoate, quinate, toluene, naphthalene, geraniol, limonene, pinene, chloroalkane, chloroalkene, styrene, ethilbenzene, caprolactam and atrazine compounds, and biosynthesis of streptomycin, novobiocin, zeatin, ansamycins, penicillin and cephalosporins.

Conclusions

The native Colombian strain *Lysinibacillus sphaericus* OT4b.31, isolated from beetle larvae, is classified between DNA similarity groups III and IV. A comparison of the chromosomal sequences of strain OT4b.31 and its closest complete genome sequence, *L. sphaericus* C3-41, demonstrates the presence of only a few similar regions with syntenial rearrangements, and no prophage or putative mosquitocidal toxins are shared. Sphaericolysin B354 and the coleopteran toxin Sip1A were predicted in the strain OT4b.31, a finding which may be useful not only in bioremediation of polluted environments, but also for biological control of agricultural pests. Finally, Cd, Zn, Co, Cu, Ni, Cr and As resistances probably are supported by efflux pumps genes.

Acknowledgements

We would like to gratefully acknowledge the help of Dr. rer. nat. Diego Riaño-Pachón at Centro Nacional de Pesquisa em Energia e Materiais for his instructions in data analysis and the Group of Computational and Evolutionary Biology at the University of Los Andes for providing us access to the computing grid cluster. This work was performed under the auspices of the Grant (1204-452-21129) of the Instituto Colombiano para el fomento de la Investigación Francisco José de Caldas, Colciencias and by the Centro de Investigaciones Microbiológicas - CIMIC laboratory.

References

1. Kellen WR, Clark TB, Lindegren JE, Ho BC, Rogoff MH, Singer S. *Bacillus sphaericus* Neide as a pathogen of mosquitoes. *J Invertebr Pathol* 1965; **7**:442-448. PubMed http://dx.doi.org/10.1016/0022-2011(65)90120-5
2. Porter AG, Davidson EW, Liu JW. Mosquitocidal toxins of bacilli and their genetic manipulation for effective biological control of mosquitoes. *Microbiol Rev* 1993; **57**:838-861. PubMed
3. Berry C. The bacterium, *Lysinibacillus sphaericus*, as an insect pathogen. *J Invertebr Pathol* 2012; **109**:1-10. PubMed http://dx.doi.org/10.1016/j.jip.2011.11.008
4. Baumann P, Clark MA, Baumann L, Broadwell AH. *Bacillus sphaericus* as a mosquito pathogen: properties of the organism and its toxins. *Microbiol Rev* 1991; **55**:425-436. PubMed
5. Bone LW, Tinelli R. *Trichostrongylus colubriformis*: Larvicidal activity of toxic extracts from *Bacillus sphaericus* (strain 1593) spores. *Exp Parasitol* 1987; **64**:514-516. PubMed http://dx.doi.org/10.1016/0014-4894(87)90066-X
6. Key PB, Scott GI. Acute toxicity of the mosquito larvicide, *Bacillus sphaericus*, to the grass shrimp,*Palaemonetes pugio*, and mummichog,*Fundulus heteroclitus*. *Bull Environ Contam Toxicol* 1992; **49**:425-430. PubMed http://dx.doi.org/10.1007/BF01239647
7. Thanabalu T, Porter AG. Efficient expression of a 100-kilodalton mosquitocidal toxin in protease-deficient recombinant *Bacillus sphaericus*. *Appl Environ Microbiol* 1995; **61**:4031-4036. PubMed
8. Jones GW, Nielsen-Leroux C, Yang Y, Yuan Z, Dumas VF, Monnerat RG, Berry C. A new Cry toxin with a unique two-component dependency from *Bacillus sphaericus*. *FASEB J* 2007; **21**:4112-4120. PubMed http://dx.doi.org/10.1096/fj.07-8913com
9. Lozano LC, Ayala JA, Dussán J. *Lysinibacillus sphaericus* S-layer protein toxicity against *Culex quinquefasciatus*. *Biotechnol Lett* 2011; **33**:2037-2041. PubMed http://dx.doi.org/10.1007/s10529-011-0666-9
10. Dussán J, Andrade D, Lozano L, Vanegas S. Caracterización fisiológica y genética de cepas nativas de *Bacillus sphaericus*. *Rev Colomb Biotecnologia* 2002; **4**:89-99.
11. Abou-Shanab RAI, van Berkum P, Angle JS. Heavy metal resistance and genotypic analysis of metal resistance genes in gram-positive and Gram-negative bacteria present in Ni-rich serpentine soil and in the rhizosphere of *Alyssum murale*. *Chemosphere* 2007; **68**:360-367. PubMed http://dx.doi.org/10.1016/j.chemosphere.2006.12.051
12. Desai C, Jain K, Madamwar D. Evaluation of *in vitro* Cr(VI) reduction potential in cytosolic extracts of three indigenous *Bacillus sp.* isolated from Cr(VI) polluted industrial landfill. *Bioresour Technol* 2008; **99**:6059-6069. PubMed http://dx.doi.org/10.1016/j.biortech.2007.12.046
13. Pal A, Paul AK. Aerobic chromate reduction by chromium-resistant bacteria isolated from serpentine soil. *Microbiol Res* 2004; **159**:347-354. PubMed http://dx.doi.org/10.1016/j.micres.2004.08.001
14. Selenska-Pobell S, Panak P, Miteva V, Boudakov I, Bernhard G, Nitsche H. Selective accumulation of heavy metals by three indigenous *Bacillus* strains, *B. cereus*, *B. megaterium* and *B. sphaericus*, from drain waters of a uranium waste pile. *FEMS Microbiol Ecol* 1999; **29**:59-67. http://dx.doi.org/10.1111/j.1574-6941.1999.tb00598.x
15. Villegas-Torres MF, Bedoya-Reina OC, Salazar C, Vives-Florez MJ, Dussan J. Horizontal *arsC* gene transfer among microorganisms isolated from arsenic polluted soil. *Int Biodeterior Biodegradation* 2011; **65**:147-152. http://dx.doi.org/10.1016/j.ibiod.2010.10.007
16. Velásquez L, Dussan J. Biosorption and bioaccumulation of heavy metals on dead and living biomass of *Bacillus sphaericus*. *J Hazard Mater*

2009; **167**:713-716. PubMed http://dx.doi.org/10.1016/j.jhazmat.2009.01.044

17. Lozano LC, Dussán J. Metal tolerance and larvicidal activity of Lysinibacillus sphaericus. World J Microbiol Biotechnol 2013; 29:1383-1389. PubMed http://dx.doi.org/10.1007/s11274-013-1301-9

18. White PJ, Lotay HK. Minimal nutritional requirements of *Bacillus sphaericus* NCTC9602 and 26 other strains of this species: the majority grow and sporulate with acetate as sole major source of carbon. *J Gen Microbiol* 1980; **118**:13-19.

19. Ahmed I, Yokota A, Yamazoe A, Fujiwara T. Proposal of *Lysinibacillus boronitolerans* gen. nov. sp. nov., and transfer of *Bacillus fusiformis* to *Lysinibacillus fusiformis* comb. nov. and *Bacillus sphaericus* to *Lysinibacillus sphaericus* comb. nov. *Int J Syst Evol Microbiol* 2007; **57**:1117-1125. PubMed http://dx.doi.org/10.1099/ijs.0.63867-0

20. Krych VK, Johnson JL, Yousten AA. Deoxyribonucleic acid homologies among strains of *Bacillus sphaericus*. *Int J Syst Bacteriol* 1980; **30**:476-484. http://dx.doi.org/10.1099/00207713-30-2-476

21. Nakamura LK. Phylogeny of *Bacillus sphaericus*-like organisms. *Int J Syst Evol Microbiol* 2000; **50**:1715-1722. PubMed

22. Liolios K, Chen IM, Mavromatis K, Tavernarakis N, Hugenholtz P, Markowitz VM, Kyrpides NC. The Genomes On Line Database (GOLD) in 2009: status of genomic and metagenomic projects and their associated metadata. *Nucleic Acids Res* 2010; **38**:D346-D354. PubMed http://dx.doi.org/10.1093/nar/gkp848

23. Baxevanis AD, Ouellette BFF. Bioinformatics: A Practical Guide to the Analysis of Genes and Proteins. John Wiley & Sons; 2004.

24. Galtier N, Gouy M, Gautier C. SEAVIEW and PHYLO_WIN: two graphic tools for sequence alignment and molecular phylogeny. *Comput. Appl. Biosci. Cabios* 1996; **12**:543-548. PubMed

25. Stöver BC, Müller KF. TreeGraph 2: Combining and visualizing evidence from different phylogenetic analyses. *BMC Bioinformatics* 2010; **11**:7. PubMed http://dx.doi.org/10.1186/1471-2105-11-7

26. Field D, Garrity G, Gray T, Morrison N, Selengut J, Sterk P, Tatusova T, Thomson N, Allen MJ, Angiuoli SV, *et al*. The minimum information about a genome sequence (MIGS) specification. *Nat Biotechnol* 2008; **26**:541-547. PubMed http://dx.doi.org/10.1038/nbt1360

27. Woese CR, Kandler O, Wheelis ML. Towards a natural system of organisms: proposal for the domains *Archaea, Bacteria*, and *Eucarya*. *Proc Natl Acad Sci USA* 1990; **87**:4576-4579. PubMed http://dx.doi.org/10.1073/pnas.87.12.4576

28. Gibbons NE, Murray RGE. Proposals Concerning the Higher Taxa of *Bacteria*. *Int J Syst Bacteriol* 1978; **28**:1-6. http://dx.doi.org/10.1099/00207713-28-1-1

29. Garrity GM, Holt J. The Road Map to the Manual. In: Garrity GM, Boone DR, Castenholz RW (eds), Bergey's Manual of Systematic Bacteriology.Vol 1. Second Edition. New York: Springer; 2001:119–169.

30. Murray RGE. The Higher Taxa, or, a Place for Everything...? In: Holt JG (ed), Bergey's Manual of Systematic Bacteriology.Vol 1. First Edition. Baltimore: The Williams and Wilkins Co.; 1984:31–34.

31. Ludwig W, Schleifer K, Whitman W. Class I. *Bacilli* class nov. In: De Vos P, Garrity GM, Jones D, N.R. Krieg W. Ludwig W, Rainey EA, Schleifer KH, Withman WB (eds) Bergey's Manual of Systematic Bacteriology. In: Vol 3 (The Firmicutes). Second Edition. Dordrecht, Heidelberg, London, New York: Springer; 2009:19–20.

32. List of new names and new combinations previously effectively, but not validly, published. *Int J Syst Evol Microbiol* 2010; **60**:1009-1010. PubMed http://dx.doi.org/10.1099/ijs.0.024562-0

33. Skerman VBD, McGowan V, Sneath PHA. Approved Lists of Bacterial Names. *Int J Syst Bacteriol* 1980; **30**:225-420. http://dx.doi.org/10.1099/00207713-30-1-225

34. Prévot A. Hauduroy P, Ehringer G, Guillot G, Magrou J, Prévot AR, Rosset, Urbain A (eds) Dictionnaire des Bactéries Pathogènes. 2nd ed. Paris: Masson; 1953, p. 1-692.

35. Fischer A. Untersuchungen über bakterien. *Jahrbücher Für Wiss. Bot.* 1895; **27**:1-163.

36. Jung MY, Kim JS, Paek WK, Styrak I, Park IS, Sin Y, Paek J, Park KA, Kim H, Kim HL, Chang YH. Description of *Lysinibacillus sinduriensis* sp. nov., and transfer of *Bacillus massiliensis* and *Bacillus odysseyi* to the genus *Lysinibacillus* as *Lysinibacillus massiliensis* comb. nov. and *Lysinibacillus odysseyi* comb. nov. with emended description of the genus *Lysinibacillus*. *Int J Syst Evol Microbiol* 2012; **62**:2347-2355. PubMed http://dx.doi.org/10.1099/ijs.0.033837-0

37. Claus D, Berkeley R. Genus *Bacillus* Cohn 1872, 174AL. In: Sneath PHA, Mair N, Sharpe M, Holt J, eds. Bergey's Manual of Systematic Bacteriology.Vol 2. Baltimore: The Williams and Wilkins Co.; 1986:1105–1139.

38. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, *et al*. Gene Ontology: tool for the unification of biology. *Nat Genet* 2000; **25**:25-29. PubMed http://dx.doi.org/10.1038/75556

39. FASTX-Toolkit. Available at: http://hannonlab.cshl.edu/fastx_toolkit/. Accessed November 20, 2012.

40. Blanca JM, Pascual L, Ziarsolo P, Nuez F, Cañizares J. ngs_backbone: a pipeline for read cleaning, mapping and SNP calling using Next Generation Sequencing. *BMC Genomics* 2011; **12**:285. PubMed http://dx.doi.org/10.1186/1471-2164-12-285

41. CLC Genomics Workbench - CLC bio. *Clc Bio*. Available at: http://www.clcbio.com/products/clc-genomics-workbench/. Accessed November 19, 2012.

42. Boetzer M, Pirovano W. Toward almost closed genomes with GapFiller. *Genome Biol* 2012; **13**:R56. PubMed http://dx.doi.org/10.1186/gb-2012-13-6-r56

43. Galardini M, Biondi EG, Bazzicalupo M, Mengoni A. CONTIGuator: a bacterial genomes finishing tool for structural insights on draft genomes. *Source Code Biol Med* 2011; **6**:11. PubMed http://dx.doi.org/10.1186/1751-0473-6-11

44. Otto TD, Sanders M, Berriman M, Newbold C. Iterative Correction of Reference Nucleotides (iCORN) using second generation sequencing technology. *Bioinformatics* 2010; **26**:1704-1707. PubMed http://dx.doi.org/10.1093/bioinformatics/btq269

45. Delcher AL, Phillippy A, Carlton J, Salzberg SL. Fast algorithms for large-scale genome alignment and comparison. *Nucleic Acids Res* 2002; **30**:2478-2483. PubMed http://dx.doi.org/10.1093/nar/30.11.2478

46. Darling AE, Mau B, Perna NT. progressiveMauve: multiple genome alignment with gene gain, loss and rearrangement. *PLoS ONE* 2010; **5**:e11147. PubMed http://dx.doi.org/10.1371/journal.pone.0011147

47. Aziz RK, Bartels D, Best AA, DeJongh M, Disz T, Edwards RA, Formsma K, Gerdes S, Glass EM, Kubal M, *et al*. The RAST Server: Rapid Annotations using Subsystems Technology. *BMC Genomics* 2008; **9**:75. PubMed http://dx.doi.org/10.1186/1471-2164-9-75

48. Götz S, García-Gómez JM, Terol J, Williams TD, Nagaraj SH, Nueda MJ, Robles M, Talón M, Dopazo J, Conesa A. High-throughput functional annotation and data mining with the Blast2GO suite. *Nucleic Acids Res* 2008; **36**:3420-3435. PubMed http://dx.doi.org/10.1093/nar/gkn176

49. Apweiler R, Attwood TK, Bairoch A, Bateman A, Birney E, Biswas M, Bucher P, Cerutti L, Corpet F, Croning MD, *et al*. The InterPro database, an integrated documentation resource for protein families, domains and functional sites. *Nucleic Acids Res* 2001; **29**:37-40. PubMed http://dx.doi.org/10.1093/nar/29.1.37

50. Lowe TM, Eddy SR. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res* 1997; **25**:955-964. PubMed

51. Lagesen K, Hallin P, Rødland EA, Staerfeldt HH, Rognes T, Ussery DW. RNAmmer: consistent and rapid annotation of ribosomal RNA genes. *Nucleic Acids Res* 2007; **35**:3100-3108. PubMed http://dx.doi.org/10.1093/nar/gkm160

52. Tatusov RL, Natale DA, Garkavtsev IV, Tatusova TA, Shankavaram UT, Rao BS, Kiryutin B, Galperin MY, Fedorova ND, Koonin EV. The COG database: new developments in phylogenetic classification of proteins from complete genomes. *Nucleic Acids Res* 2001; **29**:22-28. PubMed http://dx.doi.org/10.1093/nar/29.1.22

53. Zhou Y, Liang Y, Lynch KH, Dennis JJ, Wishart DS. PHAST: A Fast Phage Search Tool. *Nucleic Acids Res*. 2011. Available at: http://nar.oxfordjournals.org/content/early/2011/06/14/nar.gkr485. Accessed April 26, 2013.

54. Adams DE, Shekhtman EM, Zechiedrich EL, Schmid MB, Cozzarelli NR. The role of topoisomerase IV in partitioning bacterial replicons and the structure of catenated intermediates in DNA replication. *Cell* 1992; **71**:277-288. PubMed http://dx.doi.org/10.1016/0092-8674(92)90356-H

55. Hu X, Fan W, Han B, Liu H, Zheng D, Li Q, Dong W, Yan J, Gao M, Berry C, Yuan Z. Complete genome sequence of the mosquitocidal bacterium *Bacillus sphaericus* C3-41 and comparison with those of closely related *Bacillus* Species. *J Bacteriol* 2008; **190**:2892-2902. PubMed http://dx.doi.org/10.1128/JB.01652-07

56. From C, Granum PE, Hardy SP. Demonstration of a cholesterol-dependent cytolysin in a

noninsecticidal *Bacillus sphaericus* strain and evidence for widespread distribution of the toxin within the species. *FEMS Microbiol Lett* 2008; **286**:85-92. PubMed http://dx.doi.org/10.1111/j.1574-6968.2008.01256.x

57. Donovan WP, Engleman JT, Donovan JC, Baum JA, Bunkers GJ, Chi DJ, Clinton WP, English L, Heck GR, Ilagan OM, *et al*. Discovery and characterization of Sip1A: A novel secreted protein from *Bacillus thuringiensis* with activity against coleopteran larvae. *Appl Microbiol Biotechnol* 2006; **72**:713-719. PubMed http://dx.doi.org/10.1007/s00253-006-0332-7

58. Pollmann K, Raff J, Schnorpfeil M, Radeva G, Selenska-Pobell S. Novel surface layer protein genes in *Bacillus sphaericus* associated with unusual insertion elements. *Microbiology* 2005; **151**:2961-2973. PubMed http://dx.doi.org/10.1099/mic.0.28201-0

59. He M, Li X, Liu H, Miller SJ, Wang G, Rensing C. Characterization and genomic analysis of a highly chromate resistant and reducing bacterial strain *Lysinibacillus fusiformis* ZC1. *J Hazard Mater* 2011; **185**:682-688. PubMed http://dx.doi.org/10.1016/j.jhazmat.2010.09.072