Genomics Data 9 (2016) 78–86

Contents lists available at ScienceDirect

Genomics Data

journal homepage: www.elsevier.com/locate/gdata

Data in Brief

Complete genome sequencing and comparative genomic analysis of functionally diverse *Lysinibacillus sphaericus* III(3)7

Andrés Rey, Laura Silva-Quintero, Jenny Dussán *

Centro de Investigaciones Microbiologicas (CIMIC), Universidad de los Andes, Bogotá D.C., Cundinamarca, Colombia

ARTICLE INFO

Article history:
Received 31 May 2016
Accepted 16 June 2016
Available online 26 June 2016

Keywords:
Lysinibacillus sphaericus
Complete genome sequencing
Putative extrachromosomal elements
Comparative genomics

ABSTRACT

Lysinibacillus sphaericus III(3)7 is a native Colombian strain, the first one isolated from soil samples. This strain has shown high levels of pathogenic activity against *Culex quinquefaciatus* larvae in laboratory assays compared to other members of the same species. Using Pacific Biosciences sequencing technology we sequenced, annotated (*de novo*) and described the genome of strain III(3)7, achieving a complete genome sequence status. We then performed a comparative analysis between the newly sequenced genome and the ones previously reported for Colombian isolates *L. sphaericus* OT4b.31, CBAM5 and OT4b.25, with the inclusion of *L. sphaericus* C3-41 that has been used as a reference genome for most of previous genome sequencing projects. We concluded that *L. sphaericus* III(3)7 is highly similar with strain OT4b.25 and shares high levels of synteny with isolates CBAM5 and C3-41.

1. Introduction

Lysinibacillus sphaericus is an aerobic, gram positive, spore-forming bacterium, widely used in biological control of vector-borne diseases like Malaria and Dengue, due to its highly lethal larvicidal action [1,2]. However, *L. sphaericus* is a versatile microorganism, which also has been described as either tolerant or resistant to several toxic metals such as arsenic, hexavalent chromium and lead. These toxic metals have been largely associated with oily sludge, the latter being a contamination problem of water sources and soils in developing countries like Colombia, and in general in countries where oil exploitation has a huge environmental impact [3].

Some of the strains have been reported to be highly toxic against some mosquito species like *Culex sp*., *Anopheles* sp. and *Aedes* sp. [4], the larvicidal activity of *L. sphaericus* focuses mainly on second and third instar larvae. Also there are some other insect species targeted by this action, including nematodes, grass shrimps, cockroaches, cutworms and hemipterans [2], nevertheless it has been reported that there are some species that are not affected by *L. sphaericus*. The first examples of insects resistant to *L. sphaericus* are honey bees, in which adult bees longevity and reproduction are not affected by its insecticidal effects [5]. Resistance to *L. sphaericus* is also found in beneficial species from sewage treatment plants [6], and toxic or pathogenic effects have been reported negative in eukaryotes like shrimps, fishes, birds and mammals [7,8]. The fact that *L. sphaericus* pathogenic effects are limited against insects such as *Culex* sp. and *Aedes* sp. is of major interest in biological control because it implies both ecological, environmental and public health safety in the widespread usage of *L. sphaericus* as an effective controller of vector borne diseases, specially in tropical countries Like Colombia where endemic diseases such as Yellow fever, Dengue, Chikungunya and Zika represent a considerable public health issue [9].

There have been reports on multiple mechanisms that allow the larvicidal action in *L. sphaericus,* including several mosquitocidal and specifically larvicidal toxins expressed in vegetative or sporulation phases, at vegetative growth phase proteins like toxins from the Mtx1 and Mtx2 family, comprising the toxins Mtx2, Mtx3 and Mtx4 [2,10], also binary toxins BinA and BinB [2]. In addition the larvicidal activity of *L. sphaericus* has been reported when vegetative cells, spores and S-layer proteins are administered to larvae [11,12].

During sporulation, highly toxic strains produce a binary toxin composed of proteins BinA and BinB. First BinB binds to a receptor in epithelial midgut cells that allows BinA to enter the cell in order to cause cellular lysis [1]. On the other hand in vegetative cells, both high and low-toxicity strains produce the Mtx1, Mtx2 and Mtx3 toxins, however Mtx1 and Mtx2 proteins are degraded by proteases during the stationary growth phase, hence these proteins are not detectable when cultures undergo sporulation [13].

Bacillus sphaericus was reassigned to the genus *Lysinibacillus* due to both phylogenetic analyses and physiological differences [14]. *L. sphaericus* is a functionally heterogeneous species, being divided into five DNA homology groups. Pathogenic (mosquitocidal) strains are found in subgroup IIA, nevertheless this homology group also contains non-pathogenic isolates. Subgroup IIB has been allocated to

* Corresponding author.
E-mail address: jdussan@uniandes.edu.co (J. Dussán).

http://dx.doi.org/10.1016/j.gdata.2016.06.002

Lysinibacillus fusiformis [15]. Nakamura classified *L. sphaericus sensu lato* into seven similarity groups using their 16S rRNA sequence. These similarity groups are in accordance with whole-cell fatty acid profiles, four of the phylogenetic groups correspond to the DNA hybridization groups.

In this study we present the complete genome analysis of *L. sphaericus* III(3)7, sequenced using exclusively Pacific Biosciences sequencing technology (PacBio RS II). We then performed a comparative genomic analysis of the sequenced strain with the 3 previously reported genomes for Colombian *L. sphaericus* isolates [16,17,18] and with their respective reference genome *L. sphaericus* C3-41 [19].

2. Materials and methods

2.1. Bacterial strains and culture conditions

The *L. sphaericus* strain III(3)7 used in this study was previously isolated from soil samples in an oak forest near Bogotá D.C., Colombia, and belonged to the CIMIC Culture Collection, (Table 1)[28]. For this isolate we started from previously cultured nutritive agar plates, then it was incubated in nutrient broth at 30 °C, 150 rpm, until absorbance at 600 nm reached 0.9, which is equivalent to 1×10^9 UFC/mL (data not shown). This strain was chosen due to its considerably high levels of pathogenic activity in Culicidae larvae and its potential in toxic metal bioremediation processes [12,20].

2.2. DNA sample preparation

Genomic DNA was extracted and purified using the GeneJET Genomic DNA Purifiaction Kit (Thermo Scientific, K0721), using the standard protocol for Gram-Positive Bacteria Genomic DNA Purification with modifications in the lysis procedure extending incubation time with lysis buffer to 1 h and doubling the recommended lysozyme concentration. Identity of the DNA samples was confirmed by amplification of the 16S rRNA gene, then sequenced and compared to Ribosomal Database Project RDP [33] and NCBI databases. DNA samples were quantified using Qubit 2.0 fluorometer (Thermo Scientific) and Nanodrop 2000 spectrophotometer (Thermo Scientific) in order to fulfill sample quality requirements (quantity 10 μg, concentration: >50 ng/μg, <200 ng/μg).

2.3. Genome sequencing and assembly

DNA samples that met the quality requirements were sent to Genome Quebec (Montreal, Canada). Genomic DNA samples were sequenced using an exclusively PacBio based sequencing strategy (Pacific Biosciences RS II) and as we can observe in Table 2 a Large insert library strategy was used, this strategy targets 20 kb fragments which affects detection of small plasmids, in this case reported plasmids in *L. sphaericus* are high molecular weight, hence the effect should not be that drastic as to completely avoid plasmid detection.

Table 1
Classification and general features of *Lysinibacillus sphaericus* III(3)7 according to the MIGS recommendations.

MIGS-ID	Property	Term	Evidence code[a]
	Current classification	Domain *Bacteria*	TAS[b]
		Phylum *Firmicutes*	TAS[c, d, e]
		Class *Bacilli*	TAS[f, g]
		Order *Bacillales*	TAS[h, i]
		Family *Bacillaceae*	TAS[h, j]
		Genus *Lysinibacillus*	TAS[k, l]
		Species *Lysinibacillus sphaericus*	TAS[k, m]
		Type strain III(3)7	TAS[b]
	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[n]
	Optimum temperature	30 °C	TAS[n]
	Carbon source	Complex carbohydrates	TAS[n]
	Energy metabolism	Heterotroph	TAS[n]
MIGS-6	Habitat	Coleopteran (beetle) larvae	TAS[n]
MIGS-6.3	Salinity	Growth in Luria-Bertani broth (5% NaCl)	IDA
MIGS-22	Oxygen requirement	Aerobic	TAS[n]
MIGS-15	Biotic relationship	Free living	TAS[n]
MIGS-14	Pathogenicity	Known, Coleopteran and Dipteran larvae	TAS[n]
MIGS-4	Geographic location	Chicaque Natural Reserve, Cundinamarca, Colombia	TAS[n]
MIGS-5	Sample collection time	1995	TAS[n]
MIGS-4.1	Latitude	4.607037	TAS[n]
MIGS-4.2	Longitude	−74.303202	TAS[n]
MIGS-4.3	Depth	20–40 cm	TAS[n]
MIGS-4.4	Altitude	2583 m above sea level	TAS[n]

[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 [21].
[b] Woese et al. [22].
[c] Gibbons and Murray [23].
[d] Garrity and Holt [24].
[e] Murray [25].
[f] Ludwig et al. [26].
[g] [27].
[h] Skerman et al. [28].
[i] Prévot et al. [29].
[j] A. Fischer [30].
[k] Ahmed et al. [14].
[l] Jung et al. [31].
[m] Claus and Berkeley [32].
[n] Lozano et al. [12].

Table 2
Genome sequencing project information.

MIGS ID	Property	Term
MIGS-31	Finishing quality	Completed genome
MIGS-28	Libraries used	Large insert
MIGS-29	Sequencing platforms	Pacific Biosciences (PacBio) RS II
MIGS-31.2	Fold coverage	242×
MIGS-30	Assemblers	Hierarchical Genome Assembly Process (HGAP)
MIGS-32	Gene calling method	RAST, Blast2Go, PGAAP, tRNAscan-SE
	Project relevance	Biological control of vector-borne diseases, metabolic pathway, enzymes, insect pathogen

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

Attribute	Value
Chromosomal size (bp)	4,663,526
DNA GC content (bp)	1,732,966 (37.16%)
Number of replicons	1
Extrachromosomal	1
Total genes	4485
RNA genes	149
tRNA genes	107
ncRNA genes	5
Pseudogenes	87
CRISPR arrays	1
Genes assigned to COGs	2468

Genomic assembly was done using Hierarchical Genome Assembly Protocol (HGAP) workflow [34], the outcome was a *de novo* assembly that was compared to genomes previously reported on databases using Mega BLAST (NCBI), which uses an algorithm capable of aligning sequences that differ slightly as a result of sequencing or other similar "errors" (Data not shown).

2.4. Genome annotation

The genome sequence was annotated using the automated prokaryotic annotation server: Rapid Annotations using Subsystem Technology (RAST) [35], then in order to obtain more information on the predicted coding regions we performed a Blast2Go [36] annotation, through the usage of this tool we obtained information on coding sequences that were not included in RAST subsystem calculations. We also used the NCBI Prokaryotic Genome Automatic Annotation Pipeline (PGAAP) [37]. The possible orthologs present in the chromosomal contig of both strains, were identified based on the COG database and classified accordingly [38].

2.5. Comparative genomic analysis

2.5.1. Multiple genome alignment

In order to compare the newly sequenced genome to the previously reported of *L. sphaericus* C3-41 and Colombian isolates OT4b.25, CBAM5 and OT4b.31 we used MAUVE [39], as a tool to check for synteny amongst large blocks of genomic sequences. We performed a multiple genome comparing strain III(3)7 against *L. sphaericus* OT4b.25, CBAM5, OT4b.31 and C3-41. We also executed the same analysis with pBsph of strain C3-41 and the putative extrachromosomal elements found in *L. sphaericus* III(3)7 and previously reported strain OT4b.25.

2.5.2. Whole genome alignment

We used MUMmer [40] to run the global nucleotide based alignments to check for synteny amongst the sequences, we aligned strain by strain to analyze specific synthenial rearrangements in a case by case scenario. We performed the same analysis on the plasmid sequences separately.

2.5.3. Whole genome comparative visualization

BLAST Ring Image Generator (BRIG) [41] was used to show a genome wide visualization of coding sequences identity between *L. sphaericus* III(3)7 and those genomes of the strains mentioned above (*L. sphaericus* C3-41, OT4b.25, CBAM5 and OT4b.31).

Table 3
Genome sequencing project summary.

Label	Size (pb)	Topology
Chromosomal contig	4,663,526	Circular
Extrachromosomal element	173,793	Circular

2.5.4. Multi-Fasta comparative analysis

Using the Multi-Fasta reference option within BRIG [41] we compared the genes associated with multiple functions shown in laboratory assays with bioprospection importance. First we compared a set of genes related with larvicidal activity of *L. sphaericus* against larvae of vector-borne diseases. This analysis included sequences of genes such as: binary toxin genes (*binA, binB*), S-Layer proteins, hemolysin-D, chitin-binding proteins and chitin deacetylases. Secondly we compared genes directly involved in the nitrogen cycle, such as nitroreductases, regulatory proteins, transporters and proteins involved in nitric oxide synthesis. Finally we made the same comparison with genes related bioremediation of toxic metals, such as nickel, cobalt, arsenic and zinc.

3. Results and discussion

3.1. Genome sequencing and assembly

Summary and statistics for the genome-sequencing project can be observed in Tables 2, 3 and 4, after assembly we obtained two contigs in both strains. Sequencing coverage averaged 207×. As a result of the HGAP assembly process *L. sphaericus* III(3)7 genome resulted in 2 contigs, both contigs were aligned *via* Megablast, to *L. sphaericus* strain C3-41 with a similarity percentage over 99%. Contig 1 of 4.66 Mpb aligned with the chromosomal sequences of strain C3-41 and contig 2 of 173 kpb aligned with its plasmid (pBpsh), suggesting that contig 2 might be a plasmid itself. GC content along the genome averaged 37.16%. Circular visualizations of the genome generated by DNAPlotter [42] can be observed in Fig. 1.

3.2. Genome annotation

We can observe in Table 4 that after annotation *L. sphaericus* III(3)7 has 4485 coding sequences, 149 RNA coding genes and 87 pseudogenes. Table 5 contains the COG functional annotation performed on the chromosomal contig.

The genome show a wide repertoire of potential protein encoding sequences in terms of mosquitocidal toxins and genes of crucial in the high levels of larvicidal activity that *L. sphaericus* III(3)7 has shown in laboratory assays. This activity has been previously reported in laboratory experiments determining the LC_{50} of this strain against *Culex* sp. [3,12] and *Aedes* sp. (Data not shown).

Protein encoding sequences for both *binA* and *binB,* previously reported as larvicidal toxins present in *L. sphaericus,* are also found in the chromosomal contig as contiguous open reading frames [17].

RAST Annotation revealed a set of subsystems with coding sequences for several metabolic processes of environmental importance, such as 21 subsystems dedicated to nitrogen cycling including denitrification, ammonia assimilation and nitric oxide synthesis. It also revealed 12 subsystems related with aromatic compound metabolism, including toluene, benzoate and catechol degradation pathways. Finally RAST

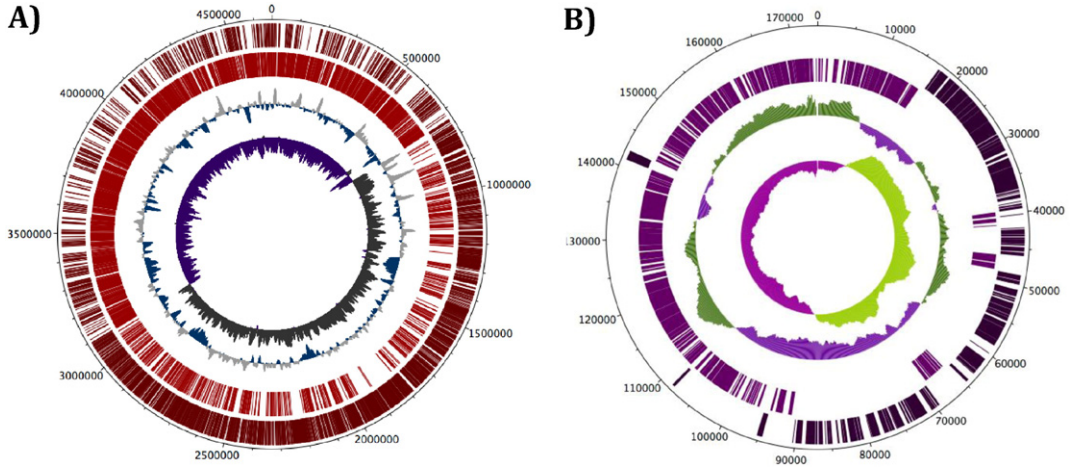


Fig. 1. Circular visualization of: A) *Lysinibacillus sphaericus* III(3)7 chromosomal contig, B) *L. sphaericus* III(3)7 putative extrachromosomal element. The inner circle represents the outer and second circles represent predicted coding regions on the forward (clockwise) and reverse (counterclockwise) DNA strands respectively. The third circle shows the GC content of the sequence, the final circle show the GC skew calculated as (G − C) / (G + C). The numbers on the outside of these circles indicate locations within the genomic contig. Image generated by DNAPlotter.

annotation showed a wide variety of efflux pump subsystems, resistance to toxic metals like Arsenic and Cobalt.

The genome *L. sphaericus* III(3)7 only shows a coding sequence for a Haemolysin-D which activity might have potential implications in *L. sphaericus* III(3)7 pathogenic activity in larvae. We can confirm the presence of a coding sequences for chitin deacetylases, these proteins are directly involved in degradation processes of chitin in a water environment into chitosan and acetone, a chitin deacetylase coding sequence has previously been reported in the genome of native Colombian strain of *L. sphaericus* CBAM5, having a putative domain of the protein NodB. Complementing the presence of the chitin deacetylase downstream along the genome there are two genes encoding chitin binding proteins with a 100% identity with the same protein reported on the genome of both the reference strain *L. sphaericus* C3-41 and Colombian isolate strain CBAM5. [17] These two genes compose an interesting metabolic pathway that may be involved in the process of inhibiting cuticle synthesis when larvae are undergoing instar switching.

Additionally the genome *L. sphaericus* III(3)7 shows 13 coding sequences for S-Layer and S-Layer like proteins, proteins that have shown a direct involvement in larvicidal activity. These sequences are coherent on what has been reported for all *L. sphaericus* strains both experimentally and through genome sequencing and annotation [12].

As a result of sequencing, assembly and annotation, we propose a potential extrachromosomal element, taking into account most of the proteins encoded by contig 2 correlates with the presence of a plasmid.

Annotation of contig 2 showed a coding sequence for protein TraG highly involved in conjugation processes in F and F-like plasmids. There is a 100% identity with the conjugal transfer protein TraG of *L. sphaericus* C3-41, and the plasmid replication protein involved in plasmid replication of this same strain with a 100% identity. Using Blast-domain this protein showed a domain similar to FtsZ that a protein considered the prokaryotic homologue to tubulin and is mainly involved in cell division. Also two protein-encoding sequences were found for site-specific recombinases like XerS that is also present on *L. sphaericus* C3-41 plasmid and DNA repair proteins like RadC. We also found sequences that belong to a Type I—C CRISPR array that includes three proteins Cas7/Csd2, Cas8c/Csd1 and Cas5, multiple DNA binding proteins, restriction endonucleases, helicases and reverse transcriptases.

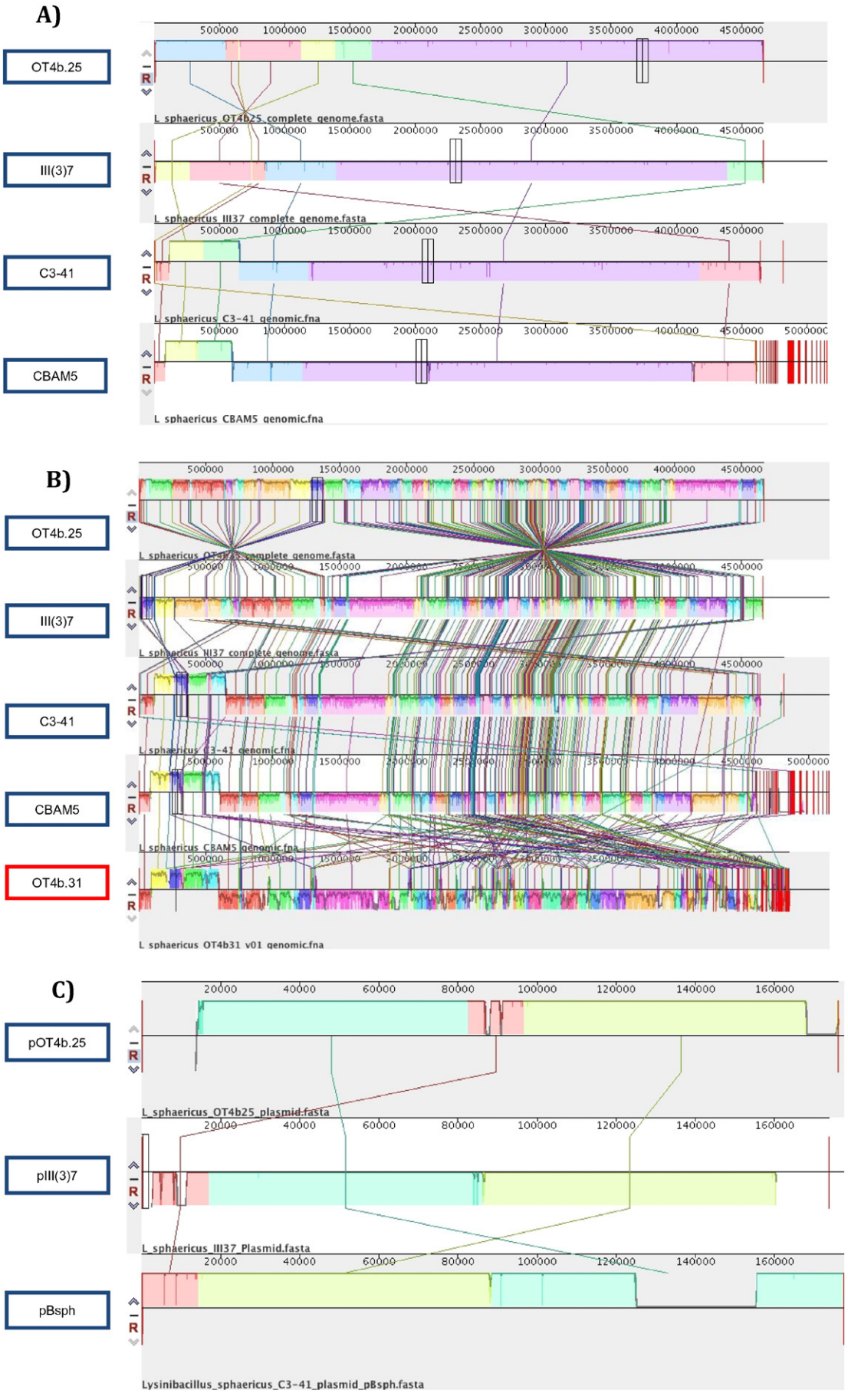
All the protein coding sequences previously mentioned, can be deemed evidence of the potential of the presence of a plasmid in *L. sphaericus* III(3)7, furthermore during the annotation of this contig we came across the presence of multiple hypothetical proteins that are related with high levels of identity to those reported on plasmid pBsph that belongs to *L. sphaericus* C3-41, pBsph is the only high molecular weight plasmid reported in *L. sphaericus* [19].

The presence of an extrachromosomal element in *L. sphaericus* III(3)7 is yet to be demonstrated by *in vitro* assays, but this evidence can be an initial step to describing a plasmid similar to the one found in *L. sphaericus* C3-41, perhaps due to low sequence representation in the sequenced samples we were not able to describe more proteins related with the presence of a plasmid this strain, nevertheless NCBI classifies contig 2 of both strains as plasmids (Accession numbers:

Table 5
Number of genes associated with the 25 general COG functional categories in the chromosomal contig of *L. sphaericus* III(3)7.

Code	Value	%age[a]	Description
J	112	4.53	Translation
K	171	6.91	Transcription
L	77	3.13	Replication, recombination and repair
B	1	0.04	Chromatin structure and dynamics
D	39	1.59	Cell cycle control, mitosis and meiosis
V	79	3.21	Defense mechanisms
T	224	9.08	Signal transduction mechanisms
M	106	4.27	Cell wall/membrane biogenesis
N	28	1.14	Cell motility
W	3	0.12	Extracellular structures
U	23	0.92	Intracellular trafficking and secretion
O	121	4.89	Posttranslational modification, protein turnover, chaperones
X	11	0.43	Phage derived proteins, transposases, mobilome components
C	101	4.09	Energy production and conversion
G	101	4.09	Carbohydrate transport and metabolism
E	274	11.10	Amino acid transport and metabolism
F	35	1.41	Nucleotide transport and metabolism
H	96	3.88	Coenzyme transport and metabolism
I	47	1.91	Lipid transport and metabolism
P	339	13.71	Inorganic ion transport and metabolism
Q	73	2.93	Secondary metabolites biosynthesis, transport and catabolism
R	343	13.90	General function prediction only
S	64	2.59	Function unknown
–	2168	46.76	Not in COGs

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


A)
OT4b.25
III(3)7
C3-41
CBAM5
L_sphaericus_OT4b25_complete_genome.fasta
L_sphaericus_III37_complete_genome.fasta
L_sphaericus_C3-41_genomic.fna
L_sphaericus_CBAM5_genomic.fna
B)
OT4b.25
III(3)7
C3-41
CBAM5
OT4b.31
L_sphaericus_OT4b25_complete_genome.fasta
L_sphaericus_III37_complete_genome.fasta
L_sphaericus_C3-41_genomic.fna
L_sphaericus_CBAM5_genomic.fna
L_sphaericus_OT4b31_v01_genomic.fna
C)
pOT4b.25
pIII(3)7
pBsph
L_sphaericus_OT4b25_plasmid.fasta
L_sphaericus_III37_Plasmid.fasta
Lysinibacillus_sphaericus_C3-41_plasmid_pBsph.fasta

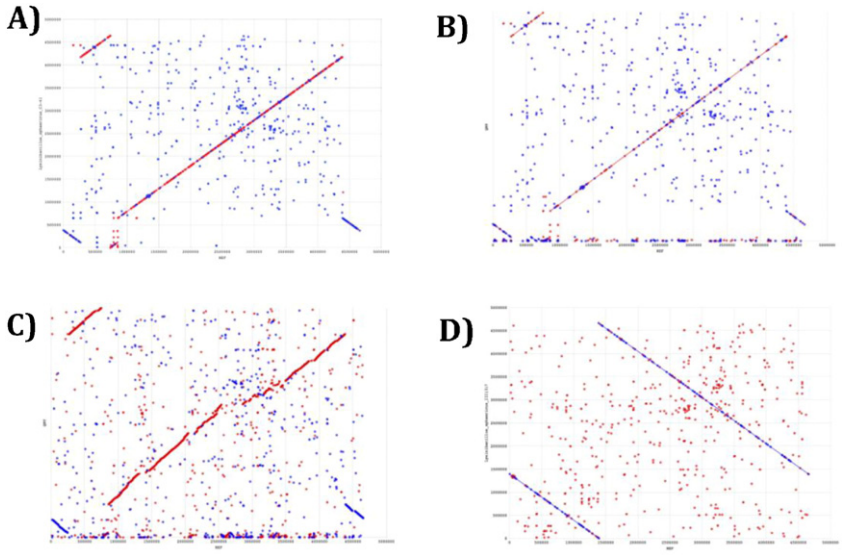


Fig. 3. Dot plot of a nucleotide-based alignment of *L. sphaericus* III(3)7 chromosomal contig with A) *L. sphaericus* C3-41 B) *L. sphaericus* CBAM5, C) *L. sphaericus* OT4b.31 and D) *L. sphaericus* OT4b.25. Aligned segments are represented as dots or lines. Forward matches are plotted in red, reverse matches in blue, figure generated by MUMmer.

CP014644.1 for strain OT4b.25 and CP014857.1 for strain III(3)7). We also have to take into account that there have been reports for cryptic plasmids in *L. sphaericus* LP1-G [43] and that evidence found in this study requires further characterization.

3.3. Comparative genomics

3.3.1. Multiple genome alignments using MAUVE

It was of our interest to compare the genome sequenced in this study to those previously sequenced of Colombian strains and the reference genome used for *L. sphaericus* genome sequencing projects to this date. We used MAUVE for multiple genome alignments, the results of these analyzes can be observed in Fig. 2.

As it can be observed in Fig. 2A there is a high level of synteny amongst strains OT4b.25, C3-41, CBAM5 and the strain sequenced in this study, as a result from the multiple genome alignment there are 5 homologous genomic blocks that are present in all strains in some cases with inversions and different positioning within each chromosome.

It is important to take into account that *L. sphaericus* C3-41 was the first genome to be sequenced of the species and has been used as a reference genome for assembly on most subsequent sequencing projects, including Colombian isolate CBAM5.

In Fig. 2B we can see the results of the same multiple alignments including *L. sphaericus* OT4b.31 sequenced at CIMIC by Peña-Montenegro and Dussán [16]. After the inclusion of this genomic sequence we can observe that the multiple alignment changes considerably. Instead of showing high levels of synteny amongst strains we see a divergence that can be reflected in over 30 homologous blocks scattered all over the genomic sequences. We can infer an important divergence between the genome of *L. sphaericus* OT4b.31 and the other strains included in the analysis, this coincides with the fact that out of the 5 genomes strain OT4b.31 is the only one with a *de novo* assembly approach and sequenced by using Illumina sequencing technology.

Fig. 2C shows the same analysis for pBsph and the putative extrachromosomal elements found in *L. sphaericus* III(3)7 and previously reported OT4b.25. Again we can observe a high level of synteny amongst the analyzed sequences which further supports the claim that strain III(3)7 possesses an extrachromosomal element.

3.3.2. Whole genome alignments using MUMmer

MUMmer was used to perform whole genome alignments. In Fig. 3 we can observe MUMmer dot-plots resulting from the alignment of the chromosomal sequence of *L. sphaericus* III(3)7 with strains C3-41, CBAM5, OT4b.31 and OT4b.25, Fig. 4 shows whole sequence alignments between pBsph and plasmid sequence found in this study and strain OT4b.25.

As it can be observed in Fig. 3A, B and D, the same level of synteny shown between *L. sphaericus* III(3)7, OT4b.25, and CBAM5 is maintained and the same inversions against strains CBAM5 and C3-41 shown in the multiple alignment in Fig. 2 can be seen in these dot-plots, furthermore there seems to be a higher similarity between the strains sequenced in this study than when compared to genome sequences of the other isolates. Again strain OT4b.31 seems to be the most divergent of the five showing the same basic outline of the dot-plot but not being able to achieve whole segment alignments.

When we performed the same analysis for the extrachromosomal elements present in strains C3-41, OT4b.25 and III(3)7 we could observe the same levels of synteny and similarity shown in the multiple sequence alignment. In this case we can see that the sequences of the putative extrachromosomal elements of strain C3-41 has lower similarity when compared with pIII(3)7 and pOT4b.25, than the latter when compared amongst themselves (Fig. 4).

3.4. BLAST ring image generator (BRIG)

3.4.1. Whole genome comparison

We compared the genomes of *L. sphaericus* III(3)7, OT4b.25, CBAM5, OT4b.31 and C3-41, as it can be observed in Fig. 5, using as reference genome the strain sequenced in this study. We can see that the similarity showed by strains CBAM5, C3-41, OT4b.25 and III(3)7 is maintained even when compared in an analysis like the one performed with BRIG, in which every open reading frame that is present in the reference genome, but absent in the genomes compared, is represented as a blank

Fig. 2. Multiple genome alignments of: A) *L. sphaericus* OT4b.25, III(3)7, CBAM5 and C3-41. B) Multiple genome alignment of: *L. sphaericus* OT4b.25, III(3)7, CBAM5, C3-41 and OT4b.31. C) Multiple global alignment of putative extra chromosomal elements of *L. sphaericus* OT4b.25 and III(3)7 with pBsph of *L. sphaericus* C3-41. Homologous blocks are shown as identically colored regions and linked across the sequences. Regions inverted relative to the reference genome are shifted downwards from the axis. Image generated by MAUVE.

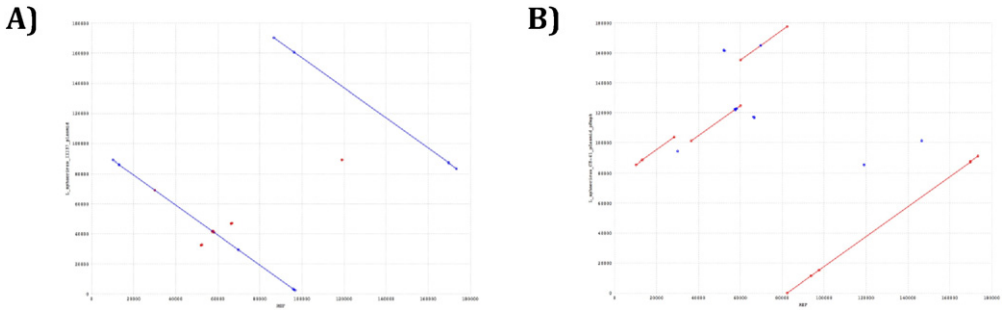


Fig. 4. Dot plot of a nucleotide-based alignment of the plasmid sequences reported in *L. sphaericus* C3-41 and found in strain III(3)7 A) pOT4b.25 and pIII(3)7 B) pIII(3)7 and pBsph. Forward matches are plotted in red, reverse matches in blue, figure generated by MUMmer.

space. Even though small or punctual differences between the most similar strains are not apparent in this kind of analysis, we can clearly see that strain OT4b.31 is again the most different amongst the strains analyzed.

3.4.2. Multi-FASTA reference gene analysis

Once we compared the whole genomes, we compared specific set of genes that are related with phenotypic characteristics in which *L. sphaericus* strains have excelled at and have been proven valuable for bioprospection purposes.

In the case of larvicidal activity (Fig. 6A and B) we can observe that when we compared the genes present in strain III(3)7 there is almost a perfect match (100% identity) with strains OT4b.25, CBAM5 and C3-41. However when compared with the genes present in strain OT4b.31 there is no match against fractions of the *binA* and *binB* genes, and some of the copies of S-layer protein are missing, as seems to be the case with both chitin deacetylase copies. These results go in accordance with the fact that *L. sphaericus* OT4b.31 is non-pathogenic and that in the annotation of its genome absence of larvicidal activity genes was recorded [16].

When comparing genes related with nitrogen cycling there seem to be a higher level of identity amongst all strains. However there in two cases there is a lower level of identity, a nitroreductase that only shows 50% identity and a NAD(P)H nitroreductase that has 70% identity with the reference strain.

Finally when comparing toxic metal remediation genes, strain OT4b.31 shows it is missing 3 genes important for arsenic resistance, including a transcriptional regulator ArsR from which it has another copy that presents 50% identity in its overall sequence. It is also missing an "arsenic resistance protein". In this case we found a difference with *L. sphaericus* C3-41 in a nickel transporter that shows 70% identity. As reported by Peña-Montenegro, et al. in 2015 *L. sphaericus* CBAM5 showed presence of resistance genes for both arsenic and cobalt.

Overall this Multi-FASTA analysis shows really low levels of genetic diversity within *L. sphaericus* strains.

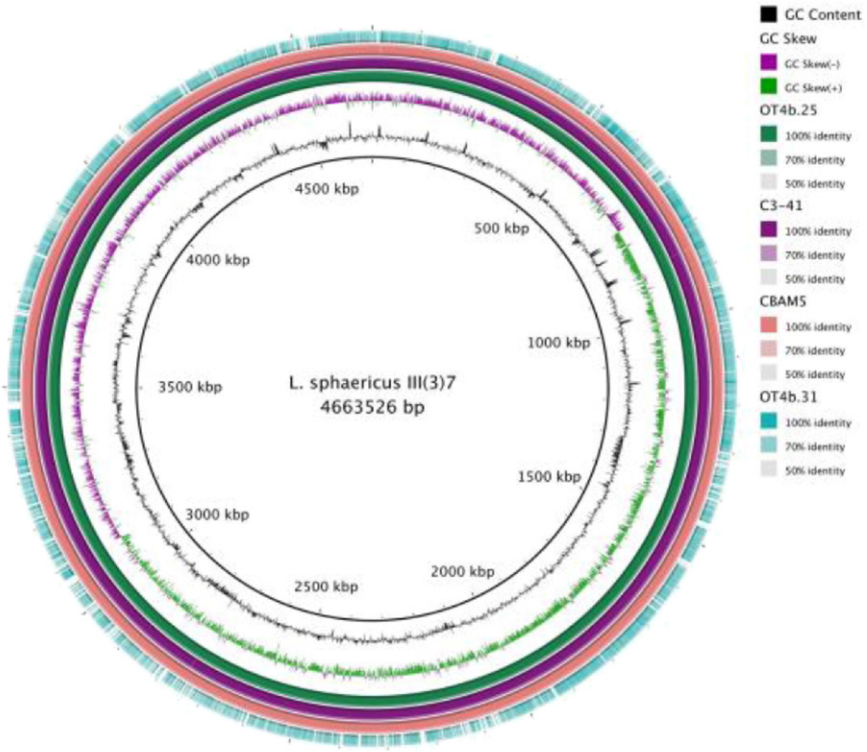


Fig. 5. Comparative circular genome (BLAST) visualization of *L. sphaericus* OT4b.25, CBAM5, C3-41 and OT4b.31, using as a reference genome *L. sphaericus* III(3)7. From inside to outside: Ring 1: GC content, Ring 2: GC Skew, Ring 3: BLAST comparison with strain OT4b.25, Ring 4: BLAST comparison with strain C3-41, Ring 5: BLAST comparison with strain CBAM5, Ring 6: BLAST comparison with strain OT4b.31. Image generated by BRIG.

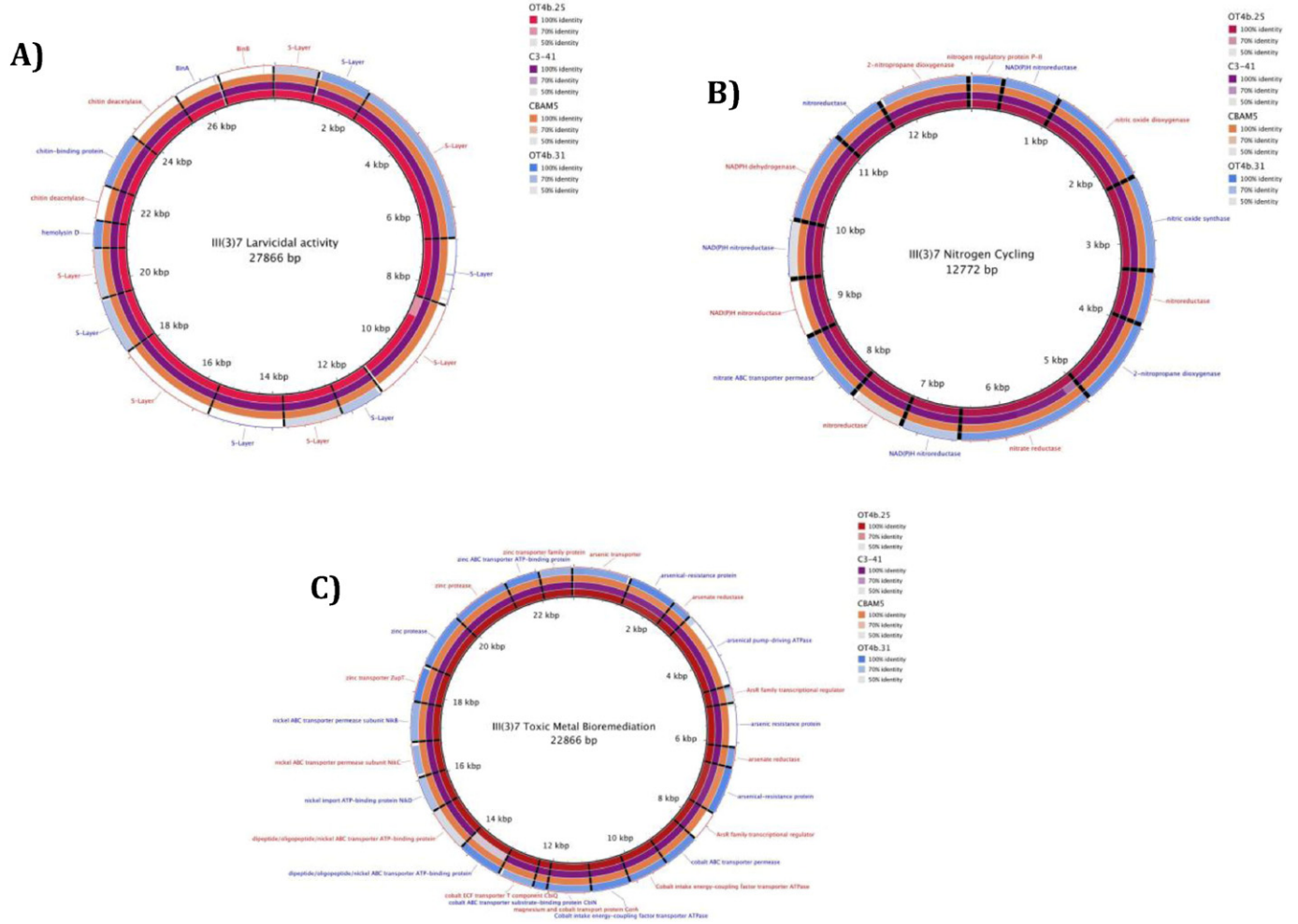


Fig. 6. Multi-FASTA reference comparison of specific set of genes amongst strains III(3)7, OT4b.25, CBAM5, OT4b.31 and C3-41. The comparisons were made of the following sets of genes. A) Genes associated with larvicidal activity, B) Genes associated with nitrogen cycle, C) Genes associated with toxic metal bioremediation. Image generated by BRIG.

4. Conclusions

We sequenced, annotated and described the genome of native Colombian strain *L. sphaericus* III(3)7. When compared with its closest genome sequences also Colombian isolate *L. sphaericus* CBAM5, OT4b.25 and *L. sphaericus* C3-41, it shows similar regions with few synthenial arrangements, nevertheless when compared with Colombian strain OT4b.31 the assembled and annotated genome shows few similar regions and many synthenial rearrangements.

We found evidence that suggest that *L. sphaericus* III(3)7 have a plasmid similar to the one reported in *L. sphaericus* C3-41, however this fact still needs to be supported by *in vitro* evidence, the same case as in previously reported *L. sphaericus* OT4b.25.

After whole genome BLAST comparison and Multi-FASTA reference comparative analysis, we conclude that the genetic diversity amongst compared *L. sphaericus* strains is low, with the exception of *L. sphaericus* OT4b.31.

Acknowledgements

We acknowledge and appreciate the contribution of Frederick Robidoux, Genevieve Geneau and Alfredo Staffa at McGill University and Genome Québec Innovation Center who largely contributed to this study. This study was performed under the auspices of the Centro de Investigaciones Microbiológicas (CIMIC) and the Sciences faculty at the Universidad de los Andes. Special thanks to Tito Peña-Montenegro for his comments and suggestions.

References

[1] P. Baumann, M.A. Clark, L. Baumann, A.H. Broadwell, *Bacillus sphaericus* as a mosquito pathogen: properties of the organism and its toxins. Microbiol. Rev. 55 (3) (1991) 425–436.
[2] C. Berry, The bacterium, *Lysinibacillus sphaericus*, as an insect pathogen. J. Invertebr. Pathol. 109 (1) (2012) 1–10, http://dx.doi.org/10.1016/j.jip.2011.11.008.
[3] L.C. Lozano, J. Dussan, Metal tolerance and larvicidal activity of *Lysinibacillus sphaericus*. World J. Microbiol. Biotechnol. 29 (8) (2013) 1383–1389, http://dx.doi.org/10.1007/s11274-013-1301-9.
[4] Y. Zhang, E. Liu, C. Cai, Z. Chen, Isolation of two highly toxic *Bacillus sphaericus* strains. Insecticidal Microorg. 1 (1987) 98–99.
[5] E.W. Davidson, H.L. Morton, J.O. Moffett, S. Singer, Effect of *Bacillus sphaericus* strain SSII-1 on honey bees, *Apis mellifera*. J. Invertebr. Pathol. 29 (3) (1977) 344–346, http://dx.doi.org/10.1016/S0022-2011(77)80041–4.
[6] N.S. Tietze, M.A. Olson, P.G. Hester, J.J. Moore, Tolerance of sewage treatment plant microorganisms to mosquitocides. J. Am. Mosq. Control Assoc. 9 (4) (1993) 477–479.
[7] M.D. Brown, T.M. Watson, J. Carter, D.M. Purdie, B.H. Kay, T.M. Watson, ... B.H. Kay, Toxicity of VectoLex (*Bacillus sphaericus*) products to selected Australian mosquito and nontarget species. J. Econ. Entomol. 97 (1) (2004) 51–58, http://dx.doi.org/10.1603/0022-0493-97.1.51.
[8] L.A. Lacey, *Bacillus thuringiensis* serovariety israelensis and *Bacillus sphaericus* for mosquito control. J. Am. Mosq. Control Assoc. 23 (2 Suppl) (2007) 133–163, http://dx.doi.org/10.2987/8756-971x(2007)23[133:btsiab]2.0.co;2.
[9] M. Fischer, Dengue, chikungunya, and other Aedes mosquito-borne diseases. http://www.cdc.gov/cdcgrandrounds/pdf/archives/2015/may2015.pdf:CDC2015.
[10] A.G. Porter, E.W. Davidson, J.W. Liu, Mosquitocidal toxins of *Bacilli* and their genetic manipulation for effective biological control of mosquitoes. Microbiol. Rev. 57 (4) (1993) 838–861.
[11] M.C. Allievi, M.M. Palomino, M. Prado Acosta, L. Lanati, S.M. Ruzal, C. Sánchez-Rivas, ... C. Sánchez-Rivas, Contribution of S-layer proteins to the mosquitocidal activity of *Lysinibacillus sphaericus*. PLoS One 9 (10) (2014), e111114 http://dx.doi.org/10.1371/journal.pone.0111114.
[12] L.C. Lozano, J.A. Ayala, J. Dussan, *Lysinibacillus sphaericus* S-layer protein toxicity against *Culex quinquefasciatus*. Biotechnol. Lett. 33 (10) (2011) 2037–2041, http://dx.doi.org/10.1007/s10529-011-0666-9.

[13] T. Thanabalu, A.G. Porter, Efficient expression of a 100-kilodalton mosquitocidal toxin in protease-deficient recombinant *Bacillus sphaericus*. Appl. Environ. Microbiol. 61 (11) (1995) 4031–4036.
[14] I. Ahmed, A. Yokota, A. Yamazoe, T. Fujiwara, 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. Micr. 57 (5) (2007) 1117–1125, http://dx.doi.org/10.1099/ijs.0.63867-0.
[15] L.K. Nakamura, Phylogeny of *Bacillus sphaericus*-like organisms. Int. J. Syst. Evol. Microbiol. 50 (5) (2000) 1715–1722, http://dx.doi.org/10.1099/00207713-50-5-1715.
[16] T.D. Peña-Montenegro, J. Dussán, Genome sequence and description of the heavy metal tolerant bacterium *Lysinibacillus sphaericus* strain OT4b.31. Stand. Genomic Sci. 9 (1) (2013) 42–56, http://dx.doi.org/10.4056/sigs.4227894.
[17] T.D. Peña-Montenegro, L. Lozano, J. Dussán, Genome sequence and description of the mosquitocidal and heavy metal tolerant strain *Lysinibacillus sphaericus* CBAM5. Stand. Genomic Sci. 10 (2015) 2, http://dx.doi.org/10.1186/1944-3277-10-2.
[18] A. Rey, L. Silva-Quintero, J. Dussán, Complete genome sequence of the larvicidal bacterium *Lysinibacillus sphaericus* strain OT4b.25. Genome Announc. 4 (3) (2016) http://dx.doi.org/10.1128/genomeA.00257–16.
[19] X. Hu, W. Fan, B. Han, H. Liu, D. Zheng, Q. Li, ... Z. Yuan, Complete genome sequence of the mosquitocidal bacterium *Bacillus sphaericus* C3-41 and comparison with those of closely related *Bacillus* species. J. Bacteriol. 190 (8) (2008) 2892–2902, http://dx.doi.org/10.1128/JB.01652-07.
[20] L. Velasquez, J. Dussan, Biosorption and bioaccumulation of heavy metals on dead and living biomass of *Bacillus sphaericus*. J. Hazard. Mater. 167 (1–3) (2009) 713–716, http://dx.doi.org/10.1016/j.jhazmat.2009.01.044.
[21] M. Ashburner, C.A. Ball, J.A. Blake, D. Botstein, H. Butler, J.M. Cherry, ... G. Sherlock, Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. Nat. Genet. 25 (1) (2000) 25–29, http://dx.doi.org/10.1038/75556.
[22] C.R. Woese, O. Kandler, M.L. Wheelis, Towards a natural system of organisms: proposal for the domains archaea, bacteria, and eucarya. Proc. Natl. Acad. Sci. U. S. A. 87 (12) (1990) 4576–4579.
[23] N.E. Gibbons, R.G.E. Murray, Proposals concerning the higher taxa of bacteria. Int. J. Syst. Evol. Microbiol. 28 (1) (1978) 1–6, http://dx.doi.org/10.1099/00207713-28-1-1.
[24] G.M. Garrity, J. Holt, The Road Map to the Manual. second ed., Bergey's Manual of Systematic Bacteriology, vol. 1, Springer, New York, 2001 119–169.
[25] R.G.E. Murray, The Higher Taxa, or, a Place for Everything…? Bergey's Manual of Systematic Bacteriology, vol. 1, The Williams and Wilkins Co., Baltimore, 1984 31–34.
[26] W. Ludwig, K. Schleifer, W. Whitman, Class I. Bacilli Class Nov. Bergey's Manual of Systematic Bacteriology, vol. 3 (The Firmicutes), Springer, Heidelberg, London, New York, 2009 19–20.
[27] List of new names and new combinations previously effectively, but not validly, published, Int. J. Syst. Evol. Microbiol. 60 (5) (2010) 1009–1010, http://dx.doi.org/10.1099/ijs.0.024562–0.
[28] V.B.D. Skerman, V. McGowan, P.H.A. Sneath, Approved lists of bacterial names. Int. J. Syst. Evol. Microbiol. 30 (1) (1980) 225–420, http://dx.doi.org/10.1099/00207713-30-1-225.
[29] A. Prévot, P. Hauduroy, G. Ehringer, G. Guillot, J. Magrou, P.A. R., P.A. R., Dictionnaire des Bactéries Pathogènes. second ed. Masson, París, 1953.
[30] A. Fischer, Jahrbuch Fur Wissenschaftliche Botanik. 1895.
[31] M.Y. Jung, J.-S. Kim, W.K. Paek, I. Styrak, I.-S. Park, Y. Sin, ... Y.-H. Chang, 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. 62 (10) (2012) 2347–2355, http://dx.doi.org/10.1099/ijs.0.033837–0.
[32] D. Claus, R. Berkeley, Genus Bacillus Cohn 1872, 174AL. Bergey's Manual of Systematic Bacteriology, vol. 2, The Williams and Wilkins Co., Baltimore, 1986 1105–1139.
[33] J.R. Cole, Q. Wang, J.A. Fish, B. Chai, D.M. McGarrell, Y. Sun, ... J.M. Tiedje, Ribosomal database project: data and tools for high throughput rRNA analysis. Nucleic Acids Res. 42 (Database issue) (2014) D633–D642, http://dx.doi.org/10.1093/nar/gkt1244.
[34] C.-S. Chin, D.H. Alexander, P. Marks, A.A. Klammer, J. Drake, C. Heiner, ... J. Korlach, Nonhybrid, finished microbial genome assemblies from long-read SMRT sequencing data. Nat. Methods 10 (6) (2013) 563–569, http://dx.doi.org/10.1038/nmeth.2474 (http://www.nature.com/nmeth/journal/v10/n6/abs/nmeth.2474.html - supplementary-information).
[35] R. Aziz, D. Bartels, A. Best, M. DeJongh, T. Disz, R. Edwards, ... O. Zagnitko, The RAST server: rapid annotations using subsystems technology. BMC Genomics 9 (1) (2008) 75.
[36] A. Conesa, S. Gotz, J.M. Garcia-Gomez, J. Terol, M. Talon, M. Robles, ... M. Robles, Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. Bioinformatics 21 (18) (2005) 3674–3676, http://dx.doi.org/10.1093/bioinformatics/bti610.
[37] S.V. Angiuoli, A. Gussman, W. Klimke, G. Cochrane, D. Field, G.M. Garrity, ... O. White, Toward an online repository of standard operating procedures (SOPs) for (meta) genomic annotation. OMICS 12 (2) (2008) 137–141, http://dx.doi.org/10.1089/omi.2008.0017.
[38] R.L. Tatusov, D.A. Natale, I.V. Garkavtsev, T.A. Tatusova, U.T. Shankavaram, B.S. Rao, ... E.V. Koonin, The COG database: new developments in phylogenetic classification of proteins from complete genomes. Nucleic Acids Res. 29 (1) (2001) 22–28, http://dx.doi.org/10.1093/nar/29.1.22.
[39] A.E. Darling, B. Mau, N.T. Perna, progressiveMauve: multiple genome alignment with gene gain, loss and rearrangement. PLoS ONE 5 (6) (2010), e11147 http://dx.doi.org/10.1371/journal.pone.0011147.
[40] A.L. Delcher, S. Kasif, R.D. Fleischmann, J. Peterson, O. White, S.L. Salzberg, Alignment of whole genomes. Nucleic Acids Res. 27 (11) (1999) 2369–2376, http://dx.doi.org/10.1093/nar/27.11.2369.
[41] N.-F. Alikhan, N.K. Petty, N.L. Ben Zakour, S.A. Beatson, BLAST ring image generator (BRIG): simple prokaryote genome comparisons. BMC Genomics 12 (1) (2011) 1–10, http://dx.doi.org/10.1186/1471-2164-12-402.
[42] T. Carver, N. Thomson, A. Bleasby, M. Berriman, J. Parkhill, DNAPlotter: circular and linear interactive genome visualization. Bioinformatics 25 (1) (2009) 119–120, http://dx.doi.org/10.1093/bioinformatics/btn578.
[43] E. Wu, L. Jun, Y. Yuan, J. Yan, C. Berry, Z. Yuan, Characterization of a cryptic plasmid from *Bacillus sphaericus* strain LP1-G. Plasmid 57 (3) (2007) 296–305, http://dx.doi.org/10.1016/j.plasmid.2006.11.003.