BIODIVERSITAS
Volume 22, Number 11, November 2021
Pages: 5211-5222

ISSN: 1412-033X
E-ISSN: 2085-4722
DOI: 10.13057/biodiv/d221160

Genome analysis of *Lysinibacillus sphaericus* isolate 6.2 pathogenic to *Culex quinquefasciatus* Say, 1823 (Diptera: Culicidae)

AFIANNISA VIERSANOVA, HARI PURWANTO♥
Faculty of Biology, Universitas Gadjah Mada. Jl. Teknika Selatan, Sekip Utara, Bulaksumur, Sleman 55281,Yogyakarta, Indonesia.
Tel./fax.: +62-274-580839, ♥email: hari.purwanto@ugm.ac.id

Manuscript received: 28 June 2021. Revision accepted: 31 October 2021.

Abstract. *Viersanova A, Purwanto H. 2021. Genome analysis of* Lysinibacillus sphaericus *isolate 6.2 pathogenic to* Culex quinquefasciatus *Say, 1823 (Diptera: Culicidae). Biodiversitas 22: 5211-5222. Lysinibacillus sphaericus* is an entomopathogenic bacteria that is specific to vector mosquitoes, especially *Culex* spp., and *Anopheles* spp., so it has been widely used as a bioinsecticide. *L. sphaericus* has a wide variation of toxicity efficiencies, which have led to continuous exploration of new isolates with higher toxicity and a new toxin to deal with resistance problems. This study aimed to identify the genomic characteristics and toxin characteristics of isolate 6.2 based on whole genome analysis and analyze the identification of isolate 6.2. Isolate 6.2 was previously obtained from rhizosphere in Yogyakarta. To analyze the genome and toxins, the *NGS* technique was used and then the analysis was carried out using a couple of freely available bioinformatics tools. Molecular identification was carried out with the 16SrRNA gene and the relationship was analyzed by reconstructing the phylogenetic tree using Neighbours-Joining. The genomic analysis of isolate 6.2 showed good results with G+C content and genome size that matched the reference genome of *L. sphaericus*. The result of the 16SrRNA gene blasting showed that the closest related gene of isolate 6.2 is *L. fusiformis* (NR_042072.1). However, the reconstructed phylogenetic tree did not show the formation of clusters according to the species. Toxin analysis indicates that isolate 6.2 has Mtx, s-layer protein, hemolysin, and chitin-binding protein genes. All of which are known to be associated with the toxicity of *L. sphaericus* to binary toxin resistant population of *Culex quinquefasciatus.*

Keywords: *Culex quinquefasciatus*, entomopathogen, genome sequencing, *Lysinibacillus sphaericus*

Abbreviations: CDS: coding sequence, NGS: next-generation sequencing, CBP: chitin-binding protein, LPMO: lytic polysaccharide monooxygenase, Mbp: megabase pairs

INTRODUCTION

Lysinibacillus sphaericus Neide is one of bacterium that is widely used as bioinsecticide and a part of vector control programs for infectious tropical diseases, such as malaria, filariasis, yellow fever, dengue fever, and West Nile virus (Poopathi and Abidha 2010; Berry 2012). Even so, not all *L. sphaericus* strains show toxicity against mosquitoes. *L. sphaericus* has a very wide genetic variation (Berry 2012). This fact creates many opportunities for the identification and exploration of new *L. sphaericus* strains with more toxic proteins or at least different protein structures.

Based on the degree of toxicity, *L. sphaericus* strains with mosquitocidal activity were divided into two groups, namely high and low levels of toxicity (Charles et al. 1996). Strains with high levels of toxicity produce binary toxins encoded by the BinA and BinB genes in the sporulation phase. While both strains with high and low levels of toxicity synthesize Mosquitocidal toxin (Mtx) during their vegetative cell growth (Clark and Baumann 1991). Decades later, it was found that the larvicidal toxicity of *L. sphaericus* also be caused by the expression of the Cry48/Cry49 toxin genes and S-layer protein (Jones et al. 2007; Lozano et al. 2011).

The most common toxins found in strains with high toxicity levels are binary toxins (BinA and BinB). This binary toxin gene is known to have a low variation. This means that the binary toxins in all strains with a high level of toxicity have genetic similarities. The gene coding for the binary toxin is highly conserved among strains, and until now, only five variations have been reported. This variation does not exist at the active site and the difference is no more than six amino acids in each protein between the two variants (Hire et al. 2009; Berry 2012). This low genetic variation in the binary toxin means that only a few toxin options can be used. Meanwhile, the uncontrolled use of *L. sphaericus* isolate with certain toxin genes can cause a resistance reaction against the target (Silva-Filha et al. 2014). Therefore, research continues to be developed to find a new collection of isolates that produce toxins that are more effective and solve the resistance problem. The discovery of new toxins can potentially overcome resistance cases against binary toxin genes began to appear. To detect the presence of a new toxin protein gene in an isolate, a bacterial genome sequence analysis using Next Generation Sequencing is promising method because it is a more easy, fast, and inexpensive one (van Dijk et al. 2014).

Previous study conducted in our lab was succeeded in isolating *L. sphaericus* from root-soil samples (rhizosphere). The pathogenicity test of this isolate, designated as isolate 6.2. showed higher pathogenicity compared to the strain 1593 that was used as positive

control (Indayati and Purwanto 2021). However, the molecular characteristics and types of toxins produced by this isolate have not yet been elucidated.

In order to develop *L. sphaericus* isolate 6.2 as a biolarvicide, it is necessary to ascertain the identity of the isolate, which until now was identified as *L. sphaericus* using morphological characters only. This study aimed to identify the genomic characteristics and toxin characteristics of isolate 6.2 based on whole genome analysis and analyze the identification of isolate 6.2. The results of this study provide basic information needed for the development of bacterial isolates with pathogenic toxins against vector mosquitoes.

MATERIALS AND METHODS

Research materials

Isolate 6.2, which was identified as *L. sphaericus* based on its morphological appearance, was isolated from the collection of the Entomology Laboratory, Faculty of Biology, Universitas Gadjah Mada, Yogyakarta, Indonesia. This isolate was previously taken from soil around plant root (rhizosphere) in Yogyakarta.

Procedures

Isolate 6.2 was prepared for shipment to MicrobesNG (Birmingham, UK) as a genome analysis service provider according to provided protocol. Sample preparation was carried out by mixing the pure isolate from agar slope with Ringer solution and streaked on the agar plate media (1/3 of the plate was filled with bacteria, while the remaining 2/3 was streaked out) and incubated until abundant growth, and then checked on phase-contrast microscope for culture purity. The bacterial culture then taken from the agar plate and mixed with preservative solution into the barcoded bead tube provided by MicrobesNG. The tube was homogenized by inverting up and down 10 times, and then sent to MicrobesNG at room temperature. The samples then went through the DNA isolation and genome sequencing analysis processes. The sample preparation, DNA isolation, and genome sequencing analysis processes were carried out based on the protocol determined by MicrobesNG (MicrobesNG 2018).

Data analysis

Contigs assembly

The sequenced bacterial genome then went through the contigs assembly stage as the first data analysis process. This process includes reads trimming, identification of reading quality, contigs assembly (de novo assembled), and identification of quality of assembled contigs. This process is performed by the sequencing service provider, MicrobesNG using software that conforms to their protocol, Trimmomatic for reads trimming, SPAdes for contigs assembly, and QUAST for assessing the quality of contigs (MicrobesNG 2021).

Genome annotation

Genome annotation aims to identify functional elements along the genome sequence. It was performed using Bowtie 2 (Langmead and Salzberg 2012), RAST (Overbeek et al. 2014), and Prokka 1.11 (Seemann 2014) to annotate with reference genomes. Genome annotation using Prokka is carried out by the sequencing service provider. Meanwhile, genome annotation with RAST and alignment of sequenced reads to long reference sequences with Bowtie 2 was carried out via the KBase facility (https://www.kbase.us/) (Arkin et al. 2018). The gene encoding binary toxin (BinA and BinB), Mtx toxin, crystalline toxin (cry48/ cry49), S-layer protein, hemolysin, and chitin-binding protein genes were detected by BLASTP and COBALT Multiple Alignment on the NCBI website (Krauthammer et al. 2000) and Mauve tool (Darling et al. 2004).

Molecular Identification and biological relationship analysis

Molecular identification was carried out using the 16S rRNA gene obtained by cutting the 16S rRNA sequence from the genomic annotation results with Artemis (Rutherford et al. 2000). The 16S rRNA gene sequence was then analyzed using Nucleotide BLAST on the NCBI website (https://blast.ncbi.nlm.nih.gov/Blast.cg). The results of the Nucleotide BLAST analysis are in the form of percent identity and query coverage that shows the similarity of the sample to the species in the GenBank database. The results obtained are then used to predict the sample species.

The relationship among taxa can be determined by conducting phylogenetic analysis. Phylogenetic relationships are visualized in the form of a phylogenetic tree which is a diagram showing the relationships among organisms based on genetic characters and evolutionary relationships. The sequences were first aligned using Mega X, which is aligned by ClustalW. Furthermore, the phylogenetic tree was reconstructed using the Neighbor-Joining method and the Kimura 2-Parameter (K2P) model (Kimura 1980) with a 1000 bootstrap. Phylogenetic tree reconstruction was carried out with Mega X (Kumar et al. 2018) between *L. sphaericus* isolate 6.2 juxtaposed with reference isolates identified as strains with high toxicity to mosquito larvae, C3-41 (Hu et al. 2008), 2362 (Hernández-Santana et al. 2016), and OT4b.25 (Rey et al. 2016); non-toxic strains KCTC 3346T or DSM 28 (Jeong et al. 2013), OT4b.31 (Dussán et al. 2002), and B1-CDA (strains with the ability to accumulate arsenic) (Rahman et al. 2016)), and *L. fusiformis*, *L. macroides*, and *Bacillus subtilis* as outgroups (Table 1).

RESULTS AND DISCUSSION

Genome characteristic and sequencing quality assessment

The sequence results of *L. sphaericus* isolate 6.2 showed that the quality of reads and genome sequences met the requirements and has a good quality coverage. This is indicated by the high coverage value, which is 76×. The quality of de novo sequence data coverage depends on the size and continuity of the contig (the number of gaps in the data). For a whole genome sequencing study, recommended coverage that is stated to be of good quality is 30-50 × per genome (Illumina Inc. 2017).

Table 1. Sequence of 16S rRNA used to conduct the phylogenetic analysis

Species	Isolate	Accession number
Lysinibacillus sphaericus	6.2	
L. sphaericus	C3-41	DQ286309.1
L. sphaericus	OT4b.25	JQ744623.1
L. sphaericus	2362	JX535356.1
L. sphaericus	OT4b.31	JQ744623.1
L. sphaericus	NBRC 15095	NR 112627.1
L. sphaericus	DSM 28	NR 042073.1
L. sphaericus	B1-CDA	KF 961041.1
L. fusiformis	C5	HF 952729.1
L. fusiformis	DSM 2898	NR 042072.1
L. fusiformis	NBRC 15717	NR 112628.1
L. fusiformis	NBRC 15717	NR 112569.1
L. fusiformis		AB 662957.1
L. macroides	PWS-1	LC557047.1
L. macroides	LMG 18474	NR 114920.1
Bacillus subtilis	IAM 12118	NR 112116.2

Based on the assembly metrics that measure the quality of contigs, this sequence is classified as good to best. This classification is based on the value of the G + C content, the number of contigs, the total size of the genome, and the values of N50, N75, L50, and L75. The results of the assembly metrics show that the G+C content is 37.1%, the number of contigs is 62 with the largest contigs size of 1,535,091 bp, the total size of the genome sequence of 4,685,755 bp, and the value of N50, N75, L50, and L75 of 1,366,529, 211,746, 2, and 5 respectively (Table 2).

Genome annotation is carried out to identify the functional elements along with the genome sequences that have been arranged in the previous stage. Information on the functional elements identified in each genome sequence of the five isolates was then affixed to the sequence which was then stored in the GenBank format (.gbk) for later visualization with Artemis. The results of the genomic annotation with RASTtk for each isolate covering various functional categories are presented in Table 3.

Molecular identification

The bacterial genome obtained was identified for the presence of the 16S rRNA gene for molecular identification purposes, then searched for its equivalent using the online Nucleotide BLAST on the NCBI website (https://blast.ncbi.nlm.nih.gov/Blast.cgi). From the analysis, it is obtained the query cover value and the similarity value which is indicated by the identity value. The higher value of this parameter and the more similar the order of the sample based on the DNA database (Aprilyanto and Sembiring 2016). Isolate 6.2 had the highest similarity with *L. fusiformis* (NR_042072.1). The top five results show very small differences in similarity with each other (Table 4).

Table 2. Assembly metrics result

Statistics	Isolate 6.2
Contigs (>= 0 bp)	62
Contigs (>=1000 bp)	26
Contigs (>=5000 bp)	21
Contigs (>=10.000 bp)	20
Contigs (>=25.000 bp)	17
Contigs (>=50.000 bp)	15
Largest contigs	1.535.091
Total length	4.685.755
Total length (>= 0 bp)	4.694.566
Total length (>=1000 bp)	4.682.721
Total length (>=5000 bp)	4.673.724
Total length (>=10.000 bp)	4.667.857
Total length (>=25.000 bp)	4.621.417
Total length (>=50.000 bp)	4.566.204
N50	1.366.529
N75	211.746
L50	2
L75	5
GC (%)	37.1
Mismatches	
N's	0
N's per 100 kbp	0

Worst *Median* *Best*

Table 3. Genome annotation result of *L. sphaericus* isolate 6.2

Statistic	Isolate 6.2
Number of features	4792
Genetic code	11
Functional categories	
Carbohydrates	264
Respiration	37
Nucleosides and nucleotides	57
Stress response	46
Protein metabolism	105
Regulation and cell signaling	23
Cell wall and capsule	66
Miscellaneous	10
RNA metabolism	71
Metabolism of aromatic compounds	28
Clustering-based subsystems	59
Phosphorous metabolism	32
Secondary metabolism	7
Dormancy and sporulation	38
Amino acid and derivatives	279
Iron metabolism	42
Phages, prophages, transposable elements, plasmid	8
Cell division and cell cycle	34
Membrane transport	50
Nitrogen metabolism	29
Sulfur metabolism	13
DNA metabolism	72
Cofactors, vitamins, prosthetic groups, and pigments	134
Fatty acids, lipids, and isoprenoids	68
Virulence, disease, and defense	41
Potassium metabolism	5
Motility and chemotaxis	74

Table 4. BLAST analysis result of 16S rRNA gene

Isolate	Closest sequence in GenBank	Accession number	Query cover	Identity	Reference
6.2	*L. fusiformis*	NR_042072.1	98%	99.72%	Swiderski (2001)
	L. fusiformis	NR_112569.1	97%	99.72%	Matsuda (2005)
	L. fusiformis	NR_112628.1	97%	99.72%	Miyashita (2006)
	L. sphaericus	NR_042073.1	98%	99.3%	Swiderski (2001)
	L. sphaericus	NR_112627.1	97%	99.29%	Miyashita (2006)

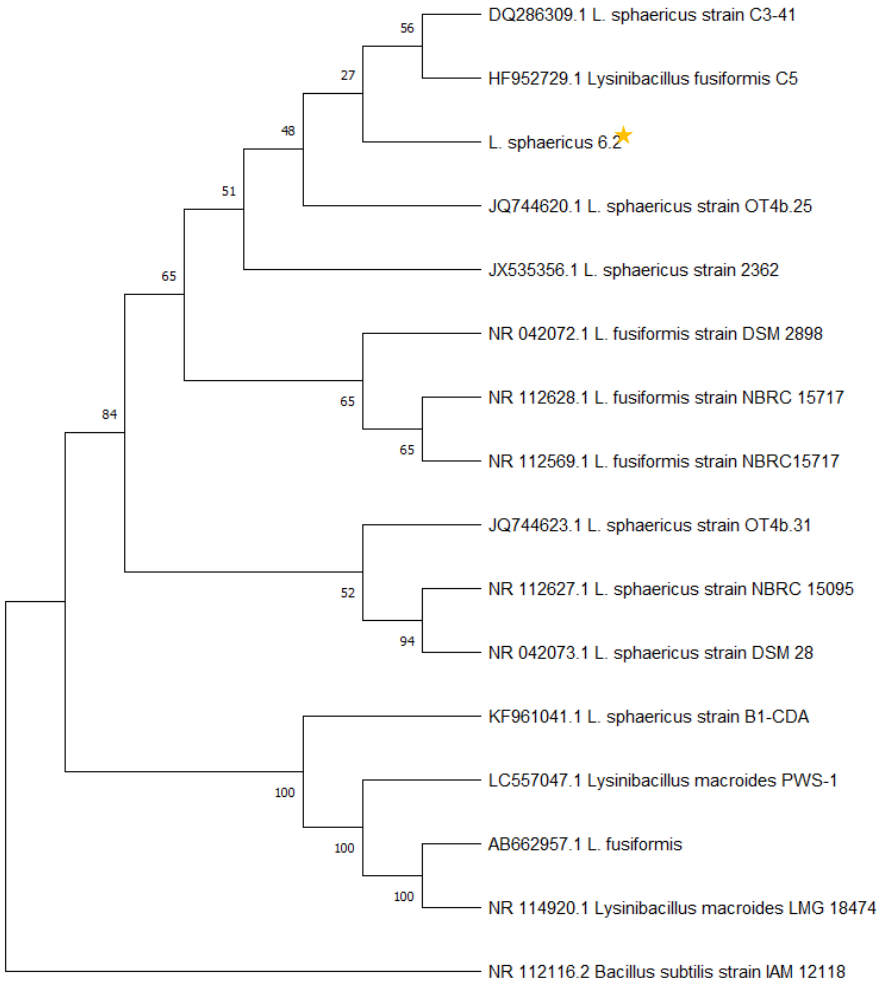

Figure 1. Phylogenetic tree of *L. sphaericus* isolate 6.2 with high-level toxicity and non-toxic *L. sphaericus* and reference strain of *L. fusiformis*

Phylogenetic tree reconstruction using 16S rRNA sequences on isolate 6.2 samples was conducted using MegaX software with ingroup and outgroup comparison sequence data downloaded from the NCBI website (Figure 1 The 16S rRNA sequences of strain with high-level toxicity, such as C3-41, 2362, and OT4b.25, 16S rRNA sequences of non-toxic strain, such as OT4b.31, DSM 28 or KCTC 3346T, and B1-CDA (strains with the ability to accumulate arsenic), and *L. macrolides*, *L. fusiformis*, and *Bacillus subtilis* as outgroups. The reconstructed phylogenetic tree showed that the isolates did not appear to be grouped into one clade according to their species. Phylogenetic tree reconstruction showing reference strains for *L. fusiformis* and *L. macroides* are intermingled with *L. sphaericus* strains.

Toxin genes analysis

Larvicidal activity of *L. sphaericus* against *Culex quinquefasciatus* is determined by the presence of toxin genes (Mtx, Bin, Cry, and Sphaericolysin) and protein-coding genes associated with the toxicity of *L. sphaericus* isolates (S-layer protein, hemolysin, and chitin-binding protein). To detect and identify these genes, Mauve software, as well as BLAST-P and COBALT Multiple alignments, were used (on the NCBI page). Toxin analysis indicates that isolate 6.2 did not have any Bin, Cry, or Sphaericolysin genes, but Mtx, s-layer protein, hemolysin, and chitin-binding protein genes were identified, these are known to be associated with the toxicity of *L. sphaericus* to mosquitoes.

Based on the genome annotation with RAST, 1 gene was found indicating the gene sequence encoding the mosquitocidal toxin, located in contigs 1 CDS_5929. Furthermore, based on the alignment between isolate 6.2 and the reference genome of *L. sphaericus* C3-41 using Mauve, gene sequences were found similar to mosquitocidal toxin in strain C3-41 which in the annotation results of isolate 6.2 with Prokka identified as hypothetical proteins. The four genes are located on CDS_1308314, CDS_1491487, CDS_1498341, and CDS_2977981 (Table 5).

Lysinibacillus sphaericus is known to have 4 types of Mtx which are characterized by the genes encoding the toxin, namely Mtx1, Mtx2, Mtx3, and Mtx4 (Table 5). To determine the type of Mtx, the five CDS genes were identified using the Constrain-based Multiple Alignment Tool (COBALT) on the NCBI website (http://www.st-va.ncbi.nlm.nih.gov). The alignment process was carried out by aligning the Mtx sequence of isolate 6.2 with the reference Mtx sequence from *L. sphaericus* C3-41, Mtx1 (BSPH_RS05330), Mtx2 (BSPH_RS05295), Mtx3 (BSPH_RS13250), and Mtx4 (BSPH_RS14895). The results displayed on COBALT showed that the amino acid sequence of Mtx isolate 6.2 was similar to the amino acid sequence of the reference Mtx, *L. sphaericus* C3-41 (Figures 2-6).

```
Mtx  _006.2        ------------------------------------------------------------------------------
Mtx1 _C3-41   1    ---MAIKKVLKIILAIIIISCQLPLNQKTVYASPNSPKDNTWIQA---ASLTWLMDMSSLLYQUSTRIPSFASPNGLH     72
Mtx2 _C3-41   1    -mKSTKLLFYVMIASFLFVNGSIYTAKATTIHENNHDIIKKQGVS-ieDIDKKIDNMIASIPPLFGFLPYSRFPYIFG     76
Mtx3 _C3-41   1    --MKNKAKVILMGATIGLSLLSSPIAMAANGDSNVKENQSIANFSpvkNSFPDAANGSRFLVNYYGRYLTSNGLGSIG     76
Mtx4 _C3-41   1    mrSKTLRNALIIGMVFITLGGTTTINPSQTYAESNSMQEKNIGIT---NVNQVLEKIGSYYYQRNLSLTWYEAPNSIG[8]  83

Mtx  _006.2   1    --------------     -------MGENTFINNSEQEQTFNITSFSESITKSTSTSIESGFKSSITTKGKVGIPFVAEG   55
Mtx1 _C3-41   73   --------------     -------MREQTIDSNTGQIQIONEHRLLRWDRRPPNDIFLNGFIP------RVTNONLSPV  121
Mtx2 _C3-41   77   ESVDVSGINIENTN     VTSWPLFIGSNTFENTTDRTMTENTVSFSKSITDSTTTOTLNGFKTAFEASGKVGIPLVAEG  152
Mtx3 _C3-41   77   KHPENIDFEVKNTY[ 9]ISQNPLWAGQSDLRNDTDRDQTLSSQEFRKSFSNTTTATTEHGFMFGTETSLATGIPFLAEG  161
Mtx4 _C3-41   84   KNPDSISSEVDLSI[10]DSTIPKFLGENIFENNTNQEQTYNTSKFSETYTESTSTSVSKGFKI--NVGRDFTIPLILNE  167

Mtx  _006.2   56   EVSAILEFNLSTT----ISTTSTITANSQAVKVP--PNKIYRHLSKQK-----------------   ------------   97
Mtx1 _C3-41   122  EDTHLLNYLRTNSPSIFVSTTRARYNNLGLEITPwtPHSANNNIIYRYEIFAPGGIDINASFSRN[4]PNEDEITFPRGN  202
Mtx2 _C3-41   153  QIKTTLEYNFSHTNSNTKSVITTYTVPPOPIPVP--PHTKTRTDVYLNQVSISGNVEIYADAITG   ------IKAESS  221
Mtx3 _C3-41   162  KITLKAEYNFSSSQANETSETVEYVAPSQSIVVP--PHTIARW-AVLEIKKIKGEMDIYAEVGLN[4]GYEELPISSMGG  239
Mtx4 _C3-41   168  GGKINLEYNSGSTNTNTLSKTYTLEAPSOPVKVP--PNKIYKAVVEYSQRTYKGTVKFYGRNTHN   PYPINTIKTTGS  242

Mtx  _006.2        ---   ------------    ------------------------------------------------------------
Mtx1 _C3-41   203  SS-   ------------    ------------------------------------------------------     204
Mtx2 _C3-41   222  GTV   ISIGDGLNLASN    TYGLIRSPODPDRVR-AIGSGKFNLINGADFTAITYDItSGEASARIIDVKEISF[1]  291
Mtx3 _C3-41   240  LKW   VSLGSIYEEAYN[12]IKIISRSVNNPDYFL-ASGKGRFESEYGSLFNVOVEYIsTKSNEVIKTENLMVSP[5]  325
Mtx4 _C3-41   243  YTG[9]FTFNDPLYKHYD[11]NDGWVIIPFLGATFVwVEGKGSFEGVYGAKLNVKTYDV-TDQNKVKLVDSRSIDL     331
```

Figure 2. Alignment results of Mtx toxin protein sequences from *L. sphaericus* isolate 6.2 annotated by RAST (CDS_5929) with Mtx sequences from *L. sphaericus* C3-41. Alignment was performed using COBALT on NCBI (http://www.st-va.ncbi.nlm.nih.gov). The color scheme taken from the program is interpreted as follows: red for conserved sequences, blue for gapless columns, gray for columns that have gaps. If less than 50% of the sequence contains gaps, it is shown in capital gray letters, 50% gaps will be shown in lower case gray.

```
CDS_1308314  1    --MKSKAKVILMGATIGLSLLSSPIAMAANGDSNVKENQSIANFSpVKNSFPDAANGSRFLVNYYGRYLTSNGLGSIG     76
Mtx1 _C3-41  1    ---MAIKKVLKIILAIIIISCQLPLNQKTVYASPNSPKDNTWIQA---ASLTWLMDMSSLLYQUSTRIPSFASPNGLH[4]  76
Mtx2 _C3-41  1    -mKSTKLLFYVMIASFLFVNGSIYTAKATTIHENNHDIIKKQGVS-IEDIDKKIDNMIASIPPLFGFLPYSRFPYIFG     76
Mtx3 _C3-41  1    --MKNKAKVILMGATIGLSLLSSPIAMAANGDSNVKENQSIANFSpVKNSFPDAANGSRFLVNYYGRYLTSNGLGSIG     76
Mtx4 _C3-41  1    mrSKTLRNALIIGMVFITLGGTTTINPSQTYAESNSMQEKNIGIT---NVNQVLEKIGSYYYQRNLSLTWYEAPNSIG[8]  83

CDS_1308314  77   KHPENIDFEVKNTYGKLSMEPQVI-SONPLWAGQSDLRNDTDRDQTLSSQEFRKSFSNTTTATTEHGFMFGTETSLATGI  155
Mtx1 _C3-41  77   --------TIDSNTGQIQIONEHR-----------LLRWDRRPPNDIFLNGFIPRVTNONLSPVEDTHLLNYLRTNSPSI  137
Mtx2 _C3-41  77   ESVDVSGINIENTN---------V-TSWPLFIGSNTFENTTDRTMTENTVSFSKSITDSTTTOTLNGFKTAFEASGKVGI  146
Mtx3 _C3-41  77   KHPENIDFEVKNTYGKLSMEPQVI-SQNPLWAGQSDLRNDTDRDQTLSSQEFRKSFSNTTTATTEHGFMFGTETSLATGI  155
Mtx4 _C3-41  84   KNPDSISSEVDLSISGEEIESLYYdSTIPKFLGENIFENNTNQEQTYNTSKFSETYTESTSTSVSKGFKIN--VGRDFTI  161

CDS_1308314  156  PFLAEGKITLKAEYNFSSSQANETSETVEYVAPSQSIVVPPHTIARVVAVLEIKKIKGEMDIYAEVGLNKEKFGYEELPI  235
Mtx1 _C3-41  138  FV-----STTRARYNNLGLEITPWT---------------PHSANNNIIYRYEIFAPGGIDINASFSRNHNPFPNEDEIT  197
Mtx2 _C3-41  147  PLVAEGQIKTTLEYNFSHTNSNTKSVITTYTVPPOPIPVPPHTKTRTDVYLNQVSISGNVEIYADAITG----------I  216
Mtx3 _C3-41  156  PFLAEGKITLKAEYNFSSSQANETSETVEYVAPSQSIVVPPHTIAR-WAVLEIKKIKGEMDIYAEVGLNKEKFGYEELPI  234
Mtx4 _C3-41  162  PLILNEGGKINLEYNSGSTNTNTLSKTYTLEAPSOPVKVPPNKIYKAVVEYSQRTYKGTVKFYGRNTHN----PYPINTI  237

CDS_1308314  236  SGVGGLK-WVSLGSIYEEAYNQAKLSGTHEFPD   IKIISRSVNNPDYFL-ASGKGRFESEYGSLFNVQVEYISTKSNE  310
Mtx1 _C3-41  198  FPRGNSS--------------------------   ---------------------------------------------  204
Mtx2 _C3-41  217  KAESSGT-VISIGDGLNLASN------------   TYGLIRSPODPDRVR-AIGSGKFNLINGADFTAITYDITSGEAS  279
Mtx3 _C3-41  235  SSMGGLK-WVSLGSIYEEAYNOAKLSGTHEFPD   IKIISRSVNNPDYFL-ASGKGRFESEYGSLFNVOVEYISTKSNE  309
Mtx4 _C3-41  238  KTTGSYTgWMGMQEIKOFTFNDPLYKHYDGLSD[7]NDGWVIIPFLGATFVwVEGKGSFEGVYGAKLNVKTYDV-TDQNK  320

CDS_1308314  311  VIKTENLMVSPT[4]  326
Mtx1 _C3-41       ------------
Mtx2 _C3-41  280  ARIIDVKEISFK     291
Mtx3 _C3-41  310  VIKTENLMVSPT[4]  325
Mtx4 _C3-41  321  VKLVDSRSIDL-     331
```

Figure 3. Alignment results of Mtx toxin protein sequences annotated with Prokka and Mauve (CDS_1308314) with Mtx sequences from *L. sphaericus* C3-41. The coloring scheme taken from the program is interpreted as follows: red for conserved sequences, blue for columns without gaps, gray for columns that have gaps. If less than 50% of the sequence contains gaps, it is displayed in gray capital letters, while more than 50% gaps will be shown in lowercase gray.

```
CDS_1491487  1    --MAIKKVLKIILAIIIIISCQLPLNQKTVYASPNSPKDNTWIQA--------------------ASLTWLMDMSSLL     56
Mtx1 _C3-41  1    --MAIKKVLKIILAIIII-SCQLPLNQKTVYASPNSPKDNTWIQA--------------------ASLTWLMDMSSLL     55
Mtx2 _C3-41  1    -mKSTKLLFYVMIASFLFVNGSIYTAKATTIHENNHDIIKKQGVS-ieDIDKKIDNMIASIPPLFGFLPYSRFPYIFG     76
Mtx3 _C3-41  1    --MKNKAKVILMGATIGLSLLSSPIAMAANGDSNVKENQSIANFSpvkNSFPDAANGSRFLVNYYGRYLTSNGLGSIG     76
Mtx4 _C3-41  1    mrSKTLRNALIIGMVFITLGGTTTINPSQTYAESNSMQEKNIGIT---NVNQVLEKIGSYYYQRNLSLTWYEAPNSIG[8]  83

CDS_1491487  57   -------------YQLISTRIPSFASPNGLHMREQTIDSNTGQIQIONEHRLLRWDRRPPNDIFLNGFI-PRVTN--ONL  120
Mtx1 _C3-41  56   -------------YQU-STRIPSFASPNGLHMREQTIDSNTGQIQIONEHRLLRWDRRPPNDIFLNGFI-PRVTN--ONL  118
Mtx2 _C3-41  77   ESVDVSGINIE-NTN---------VTSWPLFIGSNTFENTTDRTMTENTVSFSKSITDSTTTOTLNGFKTAFEASgkVGI  146
Mtx3 _C3-41  77   KHPENIDFEVK-NTYGKLSMEPQVISQNPLWAGQSDLRNDTDRDQTLSSQEFRKSFSNTTTATTEHGFMFGTETSlaTGI  155
Mtx4 _C3-41  84   KNPDSISSEVDlSISGEEIESLYYDSTIPKFLGENIFENNTNQEQTYNTSKFSETYTESTSTSVSKGFKINVGRD--FTI  161

CDS_1491487  121  SPVEDTHLLNYLRTNSPSIFVSTTRARYN-NLGLEITPWTPHSANNNIIYRYEIFAPGGIDINASFSRNHNPFPNED-EI  198
Mtx1 _C3-41  119  SPVEDTHLLNYLRTNSPSIFVSTTRARYN-NLGLEITPWTPHSANNNIIYRYEIFAPGGIDINASFSRNHNPFPNED-EI  196
Mtx2 _C3-41  147  PLVAEGQIKTTLEYNFSHTNSNTKSVITTYTVPPOPIPVPPHTKTRTDVYLNQVSISGNVEIYADAITG----------I  216
Mtx3 _C3-41  156  PFLAEGKITLKAEYNFSSSQANETSETVEYVAPSQSIVVPPHTIARW-AVLEIKKIKGEMDIYAEVGLNKEKFGYEElPI  234
Mtx4 _C3-41  162  PLILNEGGKINLEYNSGSTNTNTLSKTYTLEAPSOPVKVPPNKIYKAVVEYSQRTYKGTVKFYGRNTHNPYPI---N-TI  237

CDS_1491487  199  TFPGGIR[6]TyEYHNGEIVRIWIN[4]NPST[12]VFWHENHSEGNNMDSKGFI-------LDLDYNQDFDMFAPNGEI  283
Mtx1 _C3-41  197  TFPRG--[2]S--------------   ----    ---------------------------------------------  204
Mtx2 _C3-41  217  KAESSGT   --VISIGDGLNLASN   ----    ----TYGLIRSPODPDRVR-AIGSGKFNLINGADFTAITYDITS  275
Mtx3 _C3-41  235  SSMGGLK   --WVSLGSIYEEAYN[4]SGTH    EFPDIKIISRSVNNPDYFL-ASGKGRFESEYGSLFNVOVEYIST  305
Mtx4 _C3-41  238  KTTGSY-   TgWMGMQEIKOFTFN   DPLY[11]KEVENDGWVIIPFLGATFVwVEGKGSFEGVYGAKLNVKTYDVT-  316

CDS_1491487  284  PNNNLLNNNSLNVIQN[1]  300
Mtx1 _C3-41       ----------------
Mtx2 _C3-41  276  GEASARIIDVKEISFK     291
Mtx3 _C3-41  306  KSNEVIKTENLMVSPT[4]  325
Mtx4 _C3-41  317  DQNKVKLVDSRSIDL-     331
```

Figure 4. Alignment results of Mtx toxin protein sequences annotated with Prokka and Mauve (CDS_1491487) with Mtx sequences from *L. sphaericus* C3-41. The coloring scheme taken from the program is interpreted as follows: red for conserved sequences, blue for columns without gaps, gray for columns that have gaps. If less than 50% of the sequence contains gaps, it is displayed in gray capital letters, while more than 50% gaps will be shown in lowercase gray.

```
CDS_1498341  1    -MKRTKLLFYIMIASFLFVNGSIYTAKATTIHENNHDSIKQQGVS-IEDIDKKIDNMIASIPPLFGFLSYSRFPYIFG     76
Mtx1 _C3-41  1    ---MAIKKVLKIILAIIIISCQLPLNQKTVYASPNSPKDNTWIQA---ASLTWLMDMSSLLYQUSTRIPSFASPNGLH[2]  74
Mtx2 _C3-41  1    -MKSTKLLFYVMIASFLFVNGSIYTAKATTIHENNHDIIKKQGVS-IEDIDKKIDNMIASIPPLFGFLPYSRFPYIFG     76
Mtx3 _C3-41  1    --MKNKAKVILMGATIGLSLLSSPIAMAANGDSNVKENQSIANFSpVKNSFPDAANGSRFLVNYYGRYLTSNGLGSIG     76
Mtx4 _C3-41  1    mRSKTLRNALIIGMVFITLGGTTTINPSQTYAESNSMQEKNIGIT---NVNQVLEKIGSYYYQRNLSLTWYEAPNSIG[8]  83

CDS_1498341  77   ESVDVSGINIENTN   VTSWVPLFIGSNTFENTTORTMTENTVSFSKSITDSTTTOTLNGFKTAFEASGKVGIPLVAE-  152
Mtx1 _C3-41  75   --------------   ----------EQTIDSNTGQIQIONEHRLLRWDRRPPNDIFLNGF-----------IPRVTNo  116
Mtx2 _C3-41  77   ESVDVSGINIENTN   VTSW-PLFIGSNTFENTTDRTMTENTVSFSKSITDSTTTOTLNGFKTAFEASGKVGIPLVAE-  151
Mtx3 _C3-41  77   KHPENIDFEVKNTY[8]VISQNPLWAGQSDLRNDTDRDQTLSSQEFRKSFSNTTTATTEHGFMFGTETSLATGIPFLAE-  160
Mtx4 _C3-41  84   KNPDSISSEVDLSI[9]YDSTIPKFLGENIFENNTNQEQTYNTSKFSETYTESTSTSVSKGFK--INVGRDFTIPLILN-  166

CDS_1498341  153  --GQIKTTLEYNFSHTNSNTKSVITTY   TVPPOPIPVPPHTKTRTDVYLNQVSISGNVEIVADAITG----   --IK  218
Mtx1 _C3-41  117  nlSPVEDTHLLNYLRTNSPSIFVSTTR[4]NLGLEITPWTPHSANNNIIYRYEIFAPGGIDINASFSRNHNPF[3]DEIT  197
Mtx2 _C3-41  152  --GQIKTTLEYNFSHTNSNTKSVITTY   TVPPOPIPVPPHTKTRTDVYLNQVSISGNVEIYADAITG----   --IK  217
Mtx3 _C3-41  161  --GKITLKAEYNFSSSQANETSETVEY   VAPSQSIVVPPHTIARW-AVLEIKKIKGEMDIYAEVGLNKEKF[4]LPIS  235
Mtx4 _C3-41  167  --EGGKINLEYNSGSTNTNTLSKTYTL   EAPSOPVKVPPNKIYKAVVEYSQRTYKGTVKFYGRNTHNPYPI   NTIK  238

CDS_1498341  219  AESSGKV   ISIGDGLNLASN    TYGLIRSPQDPDRVR-AIGSGKFNLIHGADFTAITYDITSGETSARIIDVKEIS  290
Mtx1 _C3-41  198  FPRGNSS   ------------    ----------------------------------------------------  204
Mtx2 _C3-41  218  AESSGTV   ISIGDGLNLASN    TYGLIRSPODPDRVR-AIGSGKFNLINGADFTAITYDITSGEASARIIDVKEIS  289
Mtx3 _C3-41  236  SMGGLKW   VSLGSIYEEAYN[12]IKIISRSVNNPDYFL-ASGKGRFESEYGSLFNVOVEYISTKSNEVIKTENLMVS  319
Mtx4 _C3-41  239  TTGSYTG[9]FTFNDPLYKHYD[11]NDGWVIIPFLGATFVwVEGKGSFEGVYGAKLNVKTYDVT-DQNKVKLVDSRSID  330

CDS_1498341  291  FK     292
Mtx1 _C3-41       --
Mtx2 _C3-41  290  FK     291
Mtx3 _C3-41  320  PT[4]  325
Mtx4 _C3-41  331  L-     331
```

Figure 5. Alignment results of Mtx toxin protein sequences annotated with Prokka and Mauve (CDS_1498341) with Mtx sequences from *L. sphaericus* C3-41. The coloring scheme taken from the program is interpreted as follows: red for conserved sequences, blue for columns without gaps, gray for columns that have gaps. If less than 50% of the sequence contains gaps, it is displayed in gray capital letters, while more than 50% gaps will be shown in lowercase gray.

From the results of the COBALT analysis, it is known that the five coding sequence genes indicate the identified Mtx species as shown in Table 5. It was known that isolate 6.2 had all four types of Mtx. This naming is based on the old nomenclature system, while according to the new protein toxin naming system Mtx1, Mtx2, Mtx3, and Mtx4 are grouped into the ETX/MTX2 family of pore-forming toxins.

Based on the BLASTP, it is known that all CDS genes suspected of being Mtx belong to the ETX/MTX2 family of pore-forming toxin which has high similarity with ETX/MTX2 in *L. sphaericus* species. This is indicated by the top position in the BLASTP results which shows a high percent identity and query cover for in *L. sphaericus*'s ETX/MTX2 family (Table 6).

```
CDS_2977981  1    mKKRNYLVQKTLTIAAIASIGTTFALASPSLASANHISTSQ[17]QANAVGIKDVQSELNKIGKYYYTNEIAGTKYYSKG  93
Mtx1 _C3-41  1    ---MAIKKVLKIILAIIIISCQLPLNQKTVYASPNSPKDNT[ 9]-------------LMDMSSLLYQUSTRIPSFASPN  69
Mtx2 _C3-41  1    -MKSTKLLFYVMIASFLFVNGSIYTAKATTIHENNHDIIK-    KQGVS-IEDIDKKIDNMIASIP--PLFGFLPYSRF  71
Mtx3 _C3-41  1    --MKNKAKVILMGATIGLSLLSSPIAMAANGDSNVKENQS-    IANFSPVKNSFPDAANGSRFLV--NYYGRYLTSNG  71
Mtx4 _C3-41  1    mRSKTLRNALIIGMVFITLGGTTTINPSQTYAESNSMQEKN    ----IGITNVNQVLEKIGSYYYQRNLSLTWYEAPN  72

CDS_2977981  94   LPLI-fN   PAPDSVETELNFSITSNTL   NNLNYDSTTINYVTVGQLDNTkTPLEGILTTISYSESVQETTSTATON  166
Mtx1 _C3-41  70   GLHMR-E   QTIDSNTGQIQIONEHRLL[7]NDIFLNGFIPRVTNONLSPVEdTHLLNYLRTNSPSIFVSTTRARYNNL  149
Mtx2 _C3-41  72   PYIFG--   ESVDVSGINIENTN-----   -----VTSWPLFIGSNTFENT-TDRTMTENTVSFSKSITDSTTTOTLN  132
Mtx3 _C3-41  72   LGSIG--   KHPENIDFEVKNTYGKLSM   EP-QVISQNPLWAGQSDLRND-TDRDQTLSSQEFRKSFSNTTTATTEH  141
Mtx4 _C3-41  73   SIGVNhD[4]KNPDSISSEVDLSISGEEI   ESLYYDSTIPKFLGENIFENN-TNQEQTYNTSKFSETYTESTSTSVSK  149

CDS_2977981  167  GFKV   GGSGDLFFKVPLLV-EGLKLNAEFNSSTTNTTTKSVTRTLTAPSQNIRVPSGKKYKAVAVLKQLNFWGDVSFT  242
Mtx1 _C3-41  150  GLEI[6]SANNNIIYRYEIFAPGGIDINASFSRNHNPFPNEDEITFPRGNSS----------------------------  204
Mtx2 _C3-41  133  GFKT   AFEASGKVGIPLVAEGQIKTTLEYNFSHTNSNTKSVITTYTVPPOPIPVPPHTKTRTDVYLNQVSISGNVEIY  209
Mtx3 _C3-41  142  GFMF   GTETSLATGIPFLAEGKITLKAEYNFSSSQANETSETVEYVAPSQSIVVPPHTIARW-AVLEIKKIKGEMDIY  217
Mtx4 _C3-41  150  GFKI   --NVGRDFTIPLILNEGGKINLEYNSGSTNTNTLSKTYTLEAPSOPVKVPPNKIYKAVVEYSQRTYKGTVKFY  224

CDS_2977981  243  GEGVN-   PMTTIKGTAVYR-APNG[8]SKYTAQFWYELTNAQKNDLNGIEFNYFPSTGGVIVKAQGTGKFEGVMGSTL  322
Mtx1 _C3-41       ------   ----------------   --------------------------------------------------
Mtx2 _C3-41  210  ADAITG   ----IKAESSGT-VISI   GDGLNLASN------------TYGLIRSPODPDRVRAIGSGKFNLINGADF  266
Mtx3 _C3-41  218  AEVGLN[6]EELPISSMGGLK-WVSL   GSIYEEAYNOAKLSGTHEFPDIKIISRSVNNPDYFLASGKGRFESEYGSLF  296
Mtx4 _C3-41  225  GRNTHN[2]PINTIKTTGSYTgWMGM[8]NDPLYKHYDGLSDSQKKEVENDGWVIIPFLGATFVWVEGKGSFEGVYGAKL  308

CDS_2977981  323  EVDILDV-TNPANPILVESRPF--     343
Mtx1 _C3-41       ------------------------
Mtx2 _C3-41  267  TAITYDItSGEASARIIDVKEISF[1]  291
Mtx3 _C3-41  297  NVOVEYIsTKSNEVIKTENLMVSP[5]  325
Mtx4 _C3-41  309  NVKTYDV-TDQNKVKLVDSRSIDL     331
```

Figure 6. Alignment results of Mtx toxin protein sequences annotated with Prokka and Mauve (CDS_2977981) with Mtx sequences from *L. sphaericus* C3-41. The coloring scheme taken from the program is interpreted as follows: red for conserved sequences, blue for columns without gaps, gray for columns that have gaps. If less than 50% of the sequence contains gaps, it is displayed in gray capital letters, while more than 50% gaps will be shown in lowercase gray.

Table 5. The type of Mtx isolate 6.2 according to COBALT multiple alignment result

Annotation tools	Location	Mtx type
RAST	CDS_5929	Mtx4
Mauve and Prokka	CDS_1308314	Mtx3
Mauve and Prokka	CDS_1491487	Mtx1
Mauve and Prokka	CDS_1498341	Mtx2
Mauve and Prokka	CDS_2977981	Mtx4

Table 6. BLASTP result (top 5) isolate 6.2 amino acid sequence from RAST and Prokka

Description	Scientific name	Query cover	Percent identity	Accession number
CDS_5929				
ETX/MTX2 family pore-forming toxin	*L. sphaericus*	100%	100%	WP_099805144.1
MULTISPECIES: ETX/MTX2 family pore-forming toxin	Bacillaceae	100%	100%	WP_080695198.1
ETX/MTX2 family pore-forming toxin	*L. fusiformis*	100%	98.97%	WP_198695615.1
Epsilon-toxin family protein	*L. fusiformis*	93%	87.37%	WP_096364505.1
MULTISPECIES: epsilon-toxin family protein	*Lysinibacillus*	93%	84.21%	WP_107921992.1
CDS_2977981				
ETX/MTX2 family pore-forming toxin	*L. sphaericus*	100%	99.13%	WP_200990889.1
MULTISPECIES: ETX/MTX2 family pore-forming toxin	Bacillaceae	100%	99.42%	WP_051563179.1
ETX/MTX2 family pore-forming toxin	*L. sphaericus*	100%	99.13%	WP_051891081.1
ETX/MTX2 family pore-forming toxin	*L. sphaericus*	100%	99.13%	WP_051800214.1
ETX/MTX2 family pore-forming toxin	*L. sphaericus*	100%	98.25%	WP_051889585.1
CDS_1308314				
ETX/MTX2 family pore-forming toxin	*L. sphaericus*	100%	98.47%	WP_197223954.1
MULTISPECIES: epsilon toxin family protein	Bacillaceae	100%	98.77%	WP_012294440.1
MULTISPECIES: epsilon toxin family protein	*Lysinibacillus*	100%	98.47%	WP_099805143.1
ETX/MTX2 family pore-forming toxin	*L. fusiformis*	100%	96.93%	WP_171364833.1
Epsilon-toxin family protein	*L. fusiformis*	100%	96.63%	WP_096364507.1
CDS_1498341				
ETX/MTX2 family pore-forming toxin	*L. sphaericus*	100%	97.60%	WP_196244392.1
ETX/MTX2 family pore-forming toxin	*L. sphaericus*	100%	95.55%	WP_197224520.1
MULTISPECIES: epsilon-toxin family	Bacillaceae	100%	95.21%	WP_139015313.1
ETX/MTX2 family pore-forming toxin	*L. sphaericus*	100%	94.86%	WP_207889643.1
ETX/MTX2 family pore-forming toxin	*L. sphaericus*	100%	94.52%	WP_200990863.1

Based on the genome annotation with RAST, it was found that 8 copies of the s-layer protein-coding gene were found in 6.2. Genome annotation with Prokka shows 1 gene identified as S-layer protein. While result of alignment with Mauve shows some amount of sequence similarity to the s-layer protein gene in the reference genome C3-41, which was identified in the genome sequence as a hypothetical protein when annotated with Prokka. So, when the alignment results are calculated, the total gene s-layer protein is 21 copies for 6.2 (Table 7). The results of the analysis with different software showed different numbers. It happens because each software uses different features and methods in analyzing the sample genome, either by reference or alignment.

In *L. sphaericus* isolate 6.2, the genes encoding hemolysin and CBP were identified (Tables 8 and 9). Haemolysin was identified in 3 locations, namely CBP_699468, CBP_1391601, and CBP_2288207. While 1 CBP was identified in CDS_1018774. The BLASTP result showed that the hemolysins identified were *HlyA* and *HlyIII*. Meanwhile, CBP in BLASTP results was identified as Lytic Polysaccharide Monooxygenase (LPMO).

Based on BLASTP results for the CDS gene suspected of Haemolysin in isolate 6.2, it is known that all three are Haemolysin. In detail, HaemolysinA for CDS_699468 and CDS_2288207, and HaemolysinIII for CDS_1391601 (Table 8). HaemolysinA is also known as rRNA methyltransferase TlyA which is commonly abbreviated as HlyA or TlyA. Hemolysins are exotoxins that attack blood cell membranes and cause cell rupture. The mechanism of action is not well defined. Haemolysin A is induced by sodium ribonuclease, and is produced by pathogenic bacterial strains (ebi.ac.uk, interPro-Classification of Protein Family, Haemolysin A /rRNA methyltransferase TlyA (IPR004538) - InterPro entry - InterPro (ebi.ac.uk)).

Based on analysis with Mauve tools when aligned with the reference genome *L. sphaericus* C3-41, CDS_699468 has similarities with the gene sequences encoding Chitin-Binding Protein (CBP). While the results of BLASTP show that CDS_699468 has a high similarity with Lytic Polysaccharides Monooxygenase (LPMO) which is indicated by the top 5 results being the LPMO gene (Table 9). Different from the alignment result with Mauve tools and BLASTP result that shows CDS_699468 as CBP and LPMO, annotations with Prokka show that CDS_699468 is a GlcNAc binding protein A (GbpA).

Table 7. Number and location of s-layer protein genes analyzed by Prokka, RAST, and Mauve

Prokka		RAST		Mauve	
No.	**Loc.**	**No.**	**Loc.**	**No.**	**Loc.**
1	CDS 4266678	8	CDS.2221	21	CDS4102465
			CDS.2916		CDS4086312
			CDS.2919		CDS4338757
			CDS.312		CDS3682811
			CDS.313		CDS3987795
			CDS.3529		CDS3970794
			CDS.4223		CDS2996347
			CDS.4332		CDS2961831
					CDS2953962
					CDS2948564
					CDS2236945
					CDS2221600
					CDS2176440
					CDS2012074
					CDS1087379
					CDS1055934
					CDS849937
					CDS705080
					CDS509318
					CDS326609
					CDS14609

Note: Loc.: location

Table 8. BLASTP result (top 5) isolate 6.2 haemolysin amino acid sequence

Description	**Scientific name**	**Query cover**	**Percent identity**	**Accession number**
CDS_699468				
MULTISPECIES: TlyA Family RNA methyltransferase	Bacillaceae	100%	98.16%	WP_012295056.1
TlyA Family RNA methyltransferase	*L. sphaericus*	100%	97.79%	WP_197225283.1
TlyA Family RNA methyltransferase	*L. fusiformis*	100%	96.69%	WP_193832527.1
TlyA Family RNA methyltransferase	*L. fusiformis*	100%	96.32%	WP_069481921.1
TlyA Family RNA methyltransferase	*L. sphaericus*	100%	97.06%	WP_036216803.1
CDS_1391601				
MULTISPECIES: Haemolysin III family protein	Bacillaceae	100%	99.05%	WP_012294351.1
Haemolysin III family protein	*L. sphaericus*	100%	98.58%	WP_197223975.1
Haemolysin III family protein	*L. sphaericus*	100%	98.10%	WP_036168196.1
MULTISPECIES: Haemolysin III family protein	*Lysinibacillus*	100%	97.63%	WP_031418145.1
Haemolysin III family protein	*L. sphaericus*	100%	96.68%	WP_036216075.1
CDS_2288207				
Haemolysin XhlA Family Protein	*l. fusiformis*	100%	100%	WP_208704526.1
Haemolysin XhlA Family Protein	*L. sphaericus*	100%	100%	WP_051889624.1
MULTISPECIES: Haemolysin XhlA Family Protein	*Lysinibacillus*	100%	100%	WP_012293483.1
MULTISPECIES: Haemolysin XhlA Family Protein	Bacillaceae	100%	100%	WP_051563194.1
Haemolysin XhlA Family Protein	*L. fusiformis*	100%	98.77%	WP_150907955.1

Table 9. BLASTP result (top 5) isolate 6.2 CBP amino acid sequence

Description	Scientific name	Query cover	Percent ident.	Accession number
CDS_699468				
MULTISPECIES: Lytic Polysaccharide Monooxygenase	*Lysinibacillus*	99%	98.67%	WP_036152780.1
Lytic Polysaccharide Monooxygenase	*L. sphaericus*	100%	98.00%	WP_197223645.1
Lytic Polysaccharide Monooxygenase	*L. fusiformis*	99%	93.33%	WP_144789930.1
Lytic Polysaccharide Monooxygenase	*L. fusiformis*	99%	93.56%	WP_043990317.1
Lytic Polysaccharide Monooxygenase	*L. fusiformis*	99%	93.33%	WP_193833698.1

Discussion

The G+C content of *L. sphaericus* isolate 6.2 was 37.1%. This value is almost the same as the reference strain of *L. sphaericus* which usually has a G+C content in the range of 37-38%, such as isolate C3-41 with a G+C content of 37.29% (Hu et al. 2008). *L. sphaericus* entomopathogenic reference strain by WHO 2362 with a G+C content of 37.3% (Hernández-Santana et al. 2016), strain OT4b.25 with a G+C content of 37.15% (Rey et al. 2016), low level of toxicity KCTC 3346T or DSM 28 with a G+C content of 37.1% (Jeong et al. 2013), and strain LMG22257 which contains G+C 38.99% (Yan et al. 2017).

The total length of the genome sequence for *L. sphaericus* isolate 6.2 was 4.6 Mbp. Based on the reference of whole-genome sequencing studies on reference strains of *L. sphaericus*, these bacteria have a genome size ranging from 4-5 Mbp, such as C3-41 measuring 4.64 Mbp (Hu et al. 2008), WHO reference isolate 2362 measuring 4.67 Mbp (Hernández-Santana et al. 2016), CBAM5 metal tolerance isolate measuring 5.14 Mbp (Peña-Montenegro et al. 2015), entomopathogenic strain OT4b.25 measuring 4.66 Mbp (Rey et al. 2016), as well as strains with a low level of toxicity to mosquito larvae, KCT 3346T or DSM28, measuring 4.56 Mbp (Jeong et al. 2013).

The result of the genome annotation as presented in Table 3 indicate that the genome annotation process produces an output in the form of functional element data. It can be used to identify the presence of genes related to research objectives, such as protein toxin genes and other genes related to larvicidal activity against mosquitoes.

The top 5 sequences shown based on the results of analysis with BLASTn for isolate samples are shown in Table 4. The ordering is based on the largest to smallest percent identity followed by the largest to the smallest cover query. Based on the results of this analysis, 6.2 was identified as *L. fusiformis*. The BLASTn results show that the percent identity difference is classified as very small between one sequence and another, which ranges from 0.00-0.43%. Percent identity refers to a quantitative measurement of the similarity between two sequences, either DNA, amino acids, or something else. Species that are closely related have a higher percent identity for a particular sequence than species that are more closely related. Thus, the percent identity at a certain level indicates closeness or connection (Quick and Sikela 2021). Likewise with query coverage. Query coverage shows the percentage of the length of the sequence query that is included in the alignment (Newell et al. 2013).

Phylogenetics is an analytical study using molecular data, nucleotide, or amino acid sequences to reconstruct the kinship relationships among organisms. In the end, this analysis unites all organisms in the same ancestor or common ancestor. The positions of the organisms are visualized in the form of a phylogenetic tree (Sogandi 2018). The value that appears at the branch node is the bootstrap value. Bootstrap is a value that states the probability of changing the arrangement of clades and sister clades in the phylogenetic tree. If the bootstrap value is 70-100, the chance of changing the clade arrangement is low, so when analysis is carried out, the branches and trees that are formed reach consistency and will not change. Conversely, if the bootstrap value is less than 70, it is said to be unstable (Osawa et al. 2004), the chance of changing the clad arrangement is high, so that when the analysis is carried out, the branches and the trees formed can still change. Based on BLASTn, isolate 6.2 was identified as *L. fusiformis*.

However, the reconstruction of the phylogenetic tree showed that the isolates did not appear to be grouped into one clade according to their species. Phylogenetic tree reconstruction showing reference strains for *L. fusiformis* and *L. macroides* are intermingled with *L. sphaericus* strains. The concept of a bacterial species phylo-phenetic suggests that a bacterial species is a group of individual organisms that are monophyletic and genomically coherent, show a high level of similarity, and can be diagnosed by their phenotypic characteristics (Rosselló-Mora and Amann 2001). These results indicate that the concept of species in the genus *Lysnibacillus* needs to be reexamined. Sequencing using the 16S rRNA gene is by far one of the most common methods targeting housekeeping genes to study bacterial phylogeny as well as genus and species level classification (Woese 1987; Wang et al. 2014). However, based on the results of the relationship analysis visualized with the phylogenetic tree in this study, the 16S rRNA gene cannot be used for species identification. The reconstructed phylogenetic tree shows that *L. sphaericus* strains which are known to have high levels of toxicity, C3-41, 2362, and OT4b.25 were seen to spread with the low-level toxicity strains DSM 28, OT4b.31, and B1-CDA, and the strains of *L. macroides* and *L. fusiformis*. It is known that bacterial toxicity, which is indicated by the presence of a protein toxin gene, does not originate in an evolutionary manner from its ancestor, but is an acquired or adaptive gene.

The BLASTP result on the NCBI website shows that the toxin identified in isolate 6.2 is ETX/Mtx2, toxin family. It is known to be a group of the pore-forming toxin

family. The analysis with multiple alignment tools on the NCBI page using COBALT showed that the four types of Mtx, both Mtx1, Mtx2, Mtx3, and Mtx4 were identified in isolate 6.2. Pore-forming toxins are proteins capable of inserting pores in the membranes of the target cells which may lead to the lysis of the cell and release of nutrients (Iacovache et al. 2012). Overall results presented on BLASTP showed that the protein sequence in isolate 6.2 was identical to that of Mtx2 in *L. sphaericus.*

The pathogenic role of the S-layer protein has been shown in *L. sphaericus* against mosquito larvae *Cx.quinquefasciatus* and *Aedes aegypti* (Lozano et al. 2011; Allievi et al. 2014). In the sporulation phase, *L. sphaericus* spores maintain the presence of S-layer protein and associate with protein Bin. The presence of BinA-BinB toxin and S-Layer protein in the spores of an *L. sphaericus* strain contributes to the pathogenicity of *L. sphaericus*. Apart from being present in the sporulation phase of *L. sphaericus* and its association with other toxin proteins, S-layer proteins are also present in the vegetative phase of bacteria when spores are not present, and are mosquitocidal by themselves even though they are not associated with other toxin proteins (Allievi et al. 2014). S-Layer proteins had been observed to have a hemolytic activity. This activity causes damage to the target cell membrane due to the interaction of proteins with lipids. Hemolytic activity will help the pathogenic effect of the S-Layer protein on mosquito larvae and is thought to contribute to reducing the number of resistant mosquitoes (Allievi et al. 2014).

Based on the genome annotation, the gene coding for hemolysin and chitin-binding protein (CBP) was identified. Haemolysin, chitinase, and CBP are important virulence factors of entomopathogenic bacteria that increase the effectiveness of bacteria used in insect control (Andreeva et al. 2007; Manjeet et al. 2013). The presence of hemolysin is indicated by the *hlyD*, *hlyA*, and *hly*-III genes. Based on annotation and BLASTP results, *hlyA* and *hly*-III were found in isolate 6.2. Meanwhile, *hlyD* was not identified in the genome of isolate 6.2. The *hlyD* gene is a hemolysin gene that is associated with hemolysin protein secretion (Burgos and Beutin 2010). HaemolysinA (as known as rRNA methyltransferase TlyA and commonly abbreviated as HlyA or TlyA) is exotoxin that attacks blood cell membranes and causes cell rupture induced by sodium ribonucleate, and is produced by pathogenic bacterial strains. Another gene, *hly*-III, encodes hemolysin III which acts as pore-forming hemolysin (Baida and Kuzmin 1996).

A study reported the presence of a hemolytic domain in S-layer proteins and the larvicidal activity of these proteins was associated with the presence of these domains (Allievi et al. 2014). The hemolytic domain associated with the s-layer protein can be a toxic factor that contributes to the entomopathogenic activity of *L. sphaericus* (Rojas-Pinzón and Dussán 2017). As for the effect of hemolysin, after *L. sphaericus* is digested by the larvae, the bacteria recognize specific receptors and release hemolysin which lyses the midgut epithelial cells of the mosquito larvae on *Cx. quinquefasciatus* with Binary toxin (Lekakarn et al. 2015).

In *L. sphaericus* isolate 6.2, CBP was identified as LPMO and GbpA. The different results are shown from different bioinformatics tools used in this study, but basically, both CBP, LPMO, and GbpA are all related to chitinase which plays a role in degrading chitin. The entomopathogenic activity of some bacteria is associated with the production of chitinase which increases the effectiveness of the bacteria used in insect control. Chitinase is an enzyme that degrades chitin. Although larval degradation is mainly due to chitinase, CBP also contributes to this process (Manjeet et al. 2013; Suginta et al. 2016) by its ability to bind to chitin, facilitate chitinase accessibility, and act synergistically with chitinase (Frederiksen et al. 2013). LPMO is known to have the ability to degrade chitin, such as the chitinase enzyme which plays a role in entomopathogenic activity in bacteria. LPMO are abundant in nature and best known for their role in the oxidative degradation of various biopolymers such as cellulose and chitin (Eijsink et al. 2019; Labourel et al. 2020). LPMO are currently classified as Carbohydrate-Binding Module family 33 (CBM33) and Glycoside Hydrolase family 61 (GH61). CBM33 which is one of the classifications of LPMO is known to have a chitin active site called CBP21. So it is known that CBP is part of LPMO which is known to have some domains that have the ability to hydrolyze chitin (Aachmann et al. 2012). GlcNAc binding protein A (GbpA) that is shown on annotation result with Prokka is also related to chitin-binding activity. GbpA is known to bind to N-acetylglucosamine (GlcNAc) containing carbohydrates, such as chitin (Wong et al. 2012).

In conclusion, the results of genomic analysis of isolate 6.2 showed good results. This isolate was known to have G+C content and genome size that matched the reference genome for *L. sphaericus*, as well as coverage values that met the recommendations. The 16S rRNA gene blasting showed that the closest related gene of isolate 6.2 16S rRNA is *L. fusiformis* (Accession number: NR_042072.1). However, the reconstructed phylogenetic tree did not show the formation of clusters according to the species. From these results, it is known that the concept of species in the genus *Lysinibacillus* needs to be reviewed and 16S rRNA in this study cannot be used in phylogenetic identification. Toxin gene analysis carried out in this study showed that isolate 6.2 has no Bin, Cry, or sphaericolysin toxins. Meanwhile, Mtx was identified in isolate 6.2, namely Mtx1, Mtx2, Mtx3, and Mtx4. In *L. sphaericus* isolate 6.2, s-layer protein, hemolysin, and CBP genes were identified. All three are known to contribute to the entomopathogenic activity of the *Cx. quinquefasciatus* population which is resistant to binary toxin and *A. aegypti*.

ACKNOWLEDGEMENTS

This research was facilitated by Universitas Gadjah Mada (UGM), Yogyakarta, Indonesia through the 2020 Final Project Recognition Program (RTA) granted to Hari Purwanto. We thank our colleagues Ika Indayati and Larasati Kirana Putri, and staff of the Entomology Laboratory of the Faculty of Biology, UGM, and the Laboratory of Microbiology, Inter-University Center, UGM for their kind helps during this research.

REFERENCES

Aachmann FL, Sørlie M, Skjåk-Bræk G, Eijsink VGH, Vaaje-Kolstad G. 2012. NMR structure of a lytic polysaccharide monooxygenase provides insight into copper binding, protein dynamics, and substrate interactions. Proc Natl Acad Sci USA 109 (46): 18779-18784. DOI: 10.1073/pnas.1208822109

Allievi MC, Palomino MM, Acosta MP, Lanati L, Ruzal SM, Sánchez-Rivas C. 2014. Contribution of S-layer proteins to the mosquitocidal activity of *Lysinibacillus sphaericus*. PLoS One 9 (10). DOI: 10.1371/journal.pone.0111114.

Andreeva ZI, Nesterenko VF, Fomkina MG, Ternovsky VI, Suzina NE, Bakulina AY, Solonin AS, Sineva EV. 2007. The properties of *Bacillus cereus* haemolysin II pores depend on environmental conditions. Biochim Biophys Acta Biomembr BBA-Biomembranes 1768 (2): 253-263. DOI: 10.1016/j.bbamem.2006.11.004.

Arkin AP, Cottingham RW, Henry CS, et al. 2018. KBase: The United States department of energy systems biology knowledgebase. Nat Biotechnol 36 (7): 566-569. DOI: 10.1038/nbt.4163

Baida GE, Kuzmin NP. 1996. Mechanism of action of haemolysin III from *Bacillus cereus*. Biochim Biophys Acta Biomembr BBA-BIOMEMBRANES 1284 (2): 122-124. DOI: 10.1016/S0005-2736 (96)00168-X

Berry C. 2012. The bacterium, *Lysinibacillus sphaericus*, as an insect pathogen. J Invertebr Pathol 109 (1): 1-10. DOI: 10.1016/j.jip.2011.11.008

Burgos Y, Beutin L. 2010. Common origin of plasmid-encoded alpha-haemolysin genes in *Escherichia coli*. BMC Microbiol 10. DOI: 10.1186/1471-2180-10-193

Charles JF, Nielsen-LeRoux C, Delécluse A. 1996. *Bacillus sphaericus* toxins: Molecular biology and mode of action. Annu Rev Entomol 41 (1): 451-472. DOI: 10.1146/annurev.en.41.010196.002315

Clark MA, Baumann P. 1991. Modification of the *Bacillus sphaericus* 51- and 42-kilodalton mosquitocidal proteins: Effects of internal deletions, duplications, and formation of hybrid proteins. AEM 57 (1): 267-271. DOI: 10.1128/aem.57.1.267-271.1991

Darling ACE, Mau B, Blattner FR, Perna NT. 2004. Mauve: Multiple alignments of conserved genomic sequence with rearrangements. Genome Res 14 (7): 1394-1403. DOI: 10.1101/gr.2289704

Dussán J, Andrade D, Lozano L, Vanegas S. 2002. Caracterización fisiológica y genética de cepas nativas de *Bacillus sphaericus*. Rev Colomb Biotecnol 4 (1): 89-99.

Eijsink VGH, Petrovic D, Forsberg Z, Mekasha S, Røhr ÅK, Várnai A, Bissaro B, Vaaje-Kolstad G. 2019. On the functional characterization of lytic polysaccharide monooxygenases (LPMOs). Biotechnol Biofuels 12 (1): 1-16. DOI: 10.1186/s13068-019-1392-0.

Frederiksen RF, Paspaliari DK, Larsen T, Storgaard BG, Larsen MH, Ingmer H, Palcic MM, Leisner JJ. 2013. Bacterial chitinases and chitin-binding proteins as virulence factors. Microbiology 159 (5): 833-847. DOI: 10.1099/mic.0.051839-0.

Hernández-Santana A, Gómez-Garzón C, Dussán J. 2016. Complete genome sequence of *Lysinibacillus sphaericus* WHO reference strain 2362. Genome Announc 4 (3): 9-10. DOI: 10.1128/genomeA.00545-16.

Hire RS, Hadapad AB, Dongre TK, Kumar V. 2009. Purification and characterization of mosquitocidal *Bacillus sphaericus* BinA protein. J Invertebr Pathol 101: 106- 111. DOI: 10.1016/j.jip.2009.03.005.

Hu X, Fan W, Han B, Liu H, Zheng D, Li Q, Dong W, Yan J, Gao M, Berry C, Yuan Z. 2008. Complete genome sequence of the mosquitocidal bacterium *Bacillus sphaericus* C3-41 and comparison with those of closely related *Bacillus* species. J Bacteriol Res 190 (8): 2892-2902. DOI: 10.1128/JB.01652-07

Iacovache I, Degiacomi MT, van der Goot FG. 2012. Pore-forming toxins. Comprehensive Biophysics, Elsevier: 164-188. DOI: 10.1016/B978-0-12-374920-8.00518-X.

Illumina Inc. 2017. Illumina sequencing introduction. Illumina Sequencing Introduction, October: 1-8. https://www.illumina.com/documents/products/illumina_sequencing_introduction.pdf

Indayati I, Purwanto H. 2021. Characterisation of *Lysinibacillus sphaericus* strains. Jurnal Biologi Tropis 21 (1), 298-304,

Jeong H, Jeong DE, Sim YM, Park SH, Choi SK. 2013. Genome sequence of *Lysinibacillus sphaericus* strain KCTC 3346T. Genome Announc 1 (4): 4-5. DOI: 10.1128/genomeA.00625-13.

Jones GW, Nielsen-Leroux C, Yang Y, Yuan Z, Fiúza Dumas V, Monnerat RG, Berry C. 2007. A new Cry toxin with a unique two-component dependency from *Bacillus sphaericus*. FASEB J 21 (14): 4112-4120. DOI: 10.1096/fj.07-8913com.

Kimura M. 1980. A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. J Mol Evol 16 (2): 111-120. DOI: 10.1007/BF01731581.

Krauthammer M, Rzhetsky A, Morozov P, Friedman C. 2000. Using BLAST for identifying gene and protein names in journal articles. Gene 259 (1-2): 245-252. DOI: 10.1016/S0378-1119 (00)00431-5.

Kumar S, Stecher G, Li M, Knyaz C, Tamura K. 2018. MEGA X: Molecular evolutionary genetics analysis across computing platforms. Mol Biol Evol 35 (6): 1547-1549. DOI: 10.1093/molbev/msy096.

Labourel A, Frandsen KEH, Zhang F, Brouilly N, Grisel S, Haon M, Ciano L, Ropartz D, Fanuel M, Martin F, Navarro D, Rosso MN, Tandrup T, Bissaro B, Johansen KS, Zerva A, Walton PH, Henrissat B, Leggio L, Berrin JG. 2020. A fungal family of lytic polysaccharide monooxygenase-like copper proteins. Nat Chem Biol 16 (3): 345-350. DOI: 10.1038/s41589-019-0438-8.

Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. Nat Methods 9 (4): 357-359. DOI: 10.1038/nmeth.1923.

Lekakarn H, Promdonkoy B, Boonserm P. 2015. Interaction of *Lysinibacillus sphaericus* binary toxin with mosquito larval gut cells: Binding and internalization. J Invertebr Pathol 132: 125-131. DOI: 10.1016/j.jip.2015.09.010.

Lozano LC, Ayala JA, Dussán J. 2011. *Lysinibacillus sphaericus* S-layer protein toxicity against *Culex quinquefasciatus*. Biotechnol Lett 33 (10): 2037-2041. DOI: 10.1007/s10529-011-0666-9.

Manjeet K, Purushotham P, Neeraja C, Podile AR. 2013. Bacterial chitin binding proteins show differential substrate binding and synergy with chitinases. Microbiol Res 168 (7): 461-468. DOI: 10.1016/j.micres.2013.01.006.

MicrobesNg. n.d.. MicrobesNg Methods Document. microbesng.com.

MicrobesNg. 2018. Preparing stock tubes for MicrobesNG. 6-7.

Newell PD, Roco CA, Fricker AD, Merkel SM, Chandrangsu P. 2013. A Small-Group Activity Introducing the Use and Interpretation of BLAST. J Microbiol Biol Educ 14 (2): 238-243. DOI: 10.1128/jmbe.v14i2.637.

Osawa S, Su ZH, Imura Y. 2002. Molecular Phylogeny and Evolution of Carabid Ground Beetles. Springer-Verlag, Tokyo. DOI: 10.1007/978-4-431-53965-0 Please check 2002 or 2004? see page 12.

Overbeek R, Olson R, Pusch GD, Olsen GJ, Davis JJ, Disz, Edwards RA, Gerdes S, Parrello B, Shukla M, Vonstein V, Wattam AR, Xia F, Stevens R. 2014. The SEED and the Rapid Annotation of microbial genomes using Subsystems Technology (RAST). Nucleic Acids Res Spec Publ 42 (D1): 206-214. DOI: 10.1093/nar/gkt1226.

Peña-Montenegro TD, Lozano L, Dussán J. 2015. Genome sequence and description of the mosquitocidal and heavy metal tolerant strain *Lysinibacillus sphaericus* CBAM5. Stand Genom Sci 10 (2). DOI: 10.1186/1944-3277-10-2.

Poopathi S, Abidha S. 2010. Mosquitocidal bacterial toxins (*Bacillus sphaericus* and *Bacillus thuringiensis* serovar israelensis): Mode of action, cytopathological effects and mechanism of resistance. J Physiol Pathophysiol 1 (3): 22-38.

Quick VS, Sikela J. n.d.. *P*ercent Identity of Genomic DNA and Amino Acid Sequences. Center for Academic Research and Training in Anthropogeny. Retrieved March 1, 2021, from https://carta.anthropogeny.org/moca/topics/percent-identity-genomic-dna-and-amino-acid-sequences#:~:text=Percent identity refers to a, to a degree reflects relatedness.

Rahman A, Nahar N, Jass J, Olsson B, Mandal A. 2016. Complete genome sequence of *Lysinibacillus sphaericus* B1-CDA, a bacterium that accumulates arsenic. Genome Announc 4 (1): 1-3. DOI: 10.1128/genomeA.00999-15.

Rey A, Silva-Quintero L, Dussán J. 2016. Complete genome sequence of the larvicidal bacterium *Lysinibacillus sphaericus* strain OT4b.25. Genome Announc 4 (3): 10-11. DOI: 10.1128/genomeA.00257-16.

Rojas-Pinzón PA, Dussán J. 2017. Contribution of *Lysinibacillus sphaericus* haemolysin and chitin-binding protein in entomopathogenic activity against insecticide-resistant *Aedes aegypti*. World J Microbiol Biotechnol 33 (10) DOI: 10.1007/s11274-017-2348-9.

Rosselló-Mora R, Amann R. 2001. The species concept for prokaryotes. FEMS Microbiol Rev 25 (1): 39-67. DOI: 10.1016/S0168-6445 (00)00040-1.

Rutherford K, Parkhill J, Crook J, Horsnell T, Rice P, Rajandream MA, Barrell B. 2000. *Artemis*: sequence visualization and annotation.

Bioinformatics 16 (10): 944-945. DOI: 10.1093/bioinformatics/16.10.944.

Seemann T. 2014. Prokka: Rapid prokaryotic genome annotation. J Bioinform 30 (14): 2068-2069. DOI: 10.1093/bioinformatics/btu153

Silva-Filha MHNL, Berry C, Regis L. 2014. *Lysinibacillus sphaericus*: Toxins and mode of action, applications for mosquito control and resistance management. In: Dhadialla TS, Gill SS. (eds.). Advances in Insect Physiology: Insect Midgut and Insecticidal Proteins. Elsevier, Oxford, U.K.

Sogandi S. 2018. Biologi Molekuler: Identifikasi Bakteri Secara Molekuler. Universitas 17 Agustus 1945, Semarang. [Indonesian]

Suginta W, Sirimontree P, Sritho N, Ohnuma T, Fukamizo T. 2016. The chitin-binding domain of a GH-18 chitinase from *Vibrio harveyi* is crucial for chitin-chitinase interactions. Intl J Biol Macromol 93: 1111-1117. DOI: 10.1016/j.ijbiomac.2016.09.066.

van Dijk EL, Auger H, Jaszczyszyn Y, Thermes C. 2014. Ten years of next-generation sequencing technology. Trends Genet 30 (9): 418-426. DOI: 10.1016/j.tig.2014.07.001.

Wang X, King Jordan I, Mayer LW. 2014. A Phylogenetic Perspective on Molecular Epidemiology. In: Molecular Medical Microbiology: 2nd ed (Vols. 1-3). Elsevier, New York. DOI: 10.1016/B978-0-12-397169-2.00029-9.

Woese CR. 1987. Bacterial evolution. Microbiol Rev 51 (2): 221-271. DOI: 0146-0749/87/020221-51$02.00/0.

Wong E, Vaaje-Kolstad G, Ghosh A, Hurtado-Guerrero R, Konarev PV, Ibrahim AFM, Svergun DI, Eijsink VGH, Chatterjee NS, van Aalten DMF. 2012. The *Vibrio cholerae* colonization factor GbpA possesses a modular structure that governs binding to different host surfaces. PLoS Pathog 8 (1): 1-13. DOI: 10.1371/journal.ppat.1002373.

Yan W, Xiao X, Zhang Y. 2017. Complete genome sequence of *Lysinibacillus sphaericus* LMG 22257, a strain with ureolytic activity inducing calcium carbonate precipitation. J Biotechnol 246*:* 33-35. DOI: 10.1016/j.jbiotec.2017.02.016.