



Genetic Evidences of Biosurfactant Production in Two *Bacillus subtilis* Strains MB415 and MB418 Isolated From Oil Contaminated Soil

Azra Yasmin^{1†}, Fozia Aslam^{1†} and Anila Fariq^{1,2*†}

¹Microbiology and Biotechnology Research Lab, Department of Biotechnology, Fatima Jinnah Women University, Rawalpindi, Pakistan, ²Department of Biotechnology, University of Kotli Azad Jammu and Kashmir, Kotli, Pakistan

OPEN ACCESS

Edited by:

Murat Ozdal,
Atatürk University, Turkey

Reviewed by:

Shaohua Chen,
South China Agricultural University,
China
Vivek Sharma,
Chandigarh University, India
Furkan Orhan,
Ağrı İbrahim Çeçen University, Turkey

*Correspondence:

Anila Fariq
neelaahmad@gmail.com

[†]These authors have contributed
equally to this work and share first
authorship

Specialty section:

This article was submitted to
Industrial Biotechnology,
a section of the journal
Frontiers in Bioengineering and
Biotechnology

Received: 15 January 2022

Accepted: 25 March 2022

Published: 26 April 2022

Citation:

Yasmin A, Aslam F and Fariq A (2022)
Genetic Evidences of Biosurfactant
Production in Two *Bacillus subtilis*
Strains MB415 and MB418 Isolated
From Oil Contaminated Soil.
Front. Bioeng. Biotechnol. 10:855762.
doi: 10.3389/fbioe.2022.855762

Biosurfactants are a diverse group of amphiphilic compounds obtained from microbes. In the present study, the genomic analysis of biosurfactant-producing *Bacillus subtilis* MB415 and MB418 obtained from oil-contaminated soil was performed. Initially, the strains were screened for biosurfactant production by hemolytic assay, emulsification index, and oil displacement. Further FTIR analysis of extracted biosurfactants revealed the presence of lipopeptides. The sequenced genomes of MB415 and MB418 were of 4.2 Mbps with 43% GC content. Among more than 4,500 protein-coding genes, many were involved in virulence, metal/multidrug resistances, flagella assembly, chemotactic response, and aromatic ring hydroxylating dioxygenases. An annotation analysis revealed that both genomes possessed non-ribosomal synthetase gene clusters for the lipopeptide synthetases *srf* and *fen* responsible for surfactin and fengycin production. Comparative studies of both genomes highlighted variability in gene operons mainly for surfactin biosynthesis.

Keywords: biosurfactant production, secondary metabolites biosynthesis, oil remediation, lipopeptides, fengycin, surfactin

INTRODUCTION

Natural products comprise a variety of bioactive compounds like bacteriocins, terpenoids, antibiotics, and biosurfactants (Raaijmakers et al., 2010; Anuradha, 2010). Among these, biosurfactants are surface metabolites produced by microbes during the stationary phase of growth. The broad spectrum of structurally different biosurfactants, generally glycolipids, phospholipids, lipopeptides, neutral lipids, and fatty acids, consist of hydrophobic moiety (fatty acids) of varying lengths linked to hydrophilic peptide chains of 7–10 amino acids (Geys et al., 2014). Lipopeptides are one of the major classes of lower molecular weight biosurfactants produced by microorganisms. Around 90 compounds from 26 different lipopeptide families have been identified in the last two decades (Liu et al., 2015). Microbial biosurfactants have gained special interest nowadays because of their surface and interfacial properties, eco-friendly characters, and diverse applications in industrial, health, and environmental sectors. For example, microbial glycoconjugates enhance degradation of toxic organic pollutants by increasing their bioavailability by lowering surface tension and producing a solvent interface (Bhatt et al., 2021).

Biosurfactants are encoded by several genes and synthesized by a variety of microorganisms such as *Bacillus* sp., *Pseudomonas* sp., *Acinetobacter* sp., *Candida* sp., *Cryptococcus* sp., *Penicillium* sp.,

Aspergillus sp., *Kurtzmanomyces* sp., *Rhodococcus* sp., *Sphingomonas* sp., *Arthrobacter* sp., *Lactococcus* sp., and *Pseudozyma* sp. (Borah et al., 2021). Among these, the genus *Bacillus* is the most prevalent in biosurfactant production. Carolin et al. (2021) reported on the lipopeptide biosurfactant production by *Bacillus* sp., and its significant role in the degradation of aromatic amine 4-chloroaniline compound. Similarly, biosurfactants also exhibit a promising potential to efficiently eliminate toxic heavy metal pollutants by using multiple biosurfactant-metal binding stratagems, that is, emulsification, mobilization, complexation, and solubilization (Mishra et al., 2021).

Lipopeptides are biosurfactants mainly comprise a hydrophilic peptide ring and hydrophobic fatty acid moieties. Based on their structure, they are further characterized into cyclic lipopeptides (CLPs) and linear lipopeptides. The cyclic lipopeptides, commonly produced by *Bacillus subtilis*, include fengycin, surfactin, and iturin. They are composed of a peptide ring of 7 or 10 amino acids with a long hydrophobic fatty acid chain. The fatty acid chain lengths vary in each type, that is, surfactin has a chain length of C₁₃–C₁₆, while iturin's chain is C₁₄–C₁₇, and that of fengycin is C₁₄–C₁₈ (Hu et al., 2019).

Surfactin has many congeners because of the variation in the length of the fatty acid chain and the types of amino acids. The structure of surfactin offers diverse applications in several global issues such as medicine and environmental protection (Gudiña et al., 2013). It was discovered as a potent fibrin clot-inhibitor and as an antibacterial and antiviral agent. It also showed hypocholesterolemic and anti-tumor activities. Owing to its remarkable surface-, interface-, and membrane-active action, it has the ability to cross plasma membrane barriers and viral envelopes (Santos et al., 2018). It can lower the surface tension of water from 72 to 27 mN/m and is highly thermally stable and salt tolerant. Because of these properties, it possesses huge potential in the microbial enhancement of oil recovery (MEOR) and considered a good candidate for the bioremediation of contaminated soils and sub-surface environments (Hu et al., 2019).

Fengycin is another lipopeptide consisting of a decapeptide linked to a saturated or unsaturated fatty acid (C₁₄–C₁₈). The ring structure of its cyclic peptide is formed by linking the residue at position 3 to the C-terminal–COOH group of the amino acid at position 10 (Fanaei et al., 2021). Fengycins are predominantly produced by *Bacillus* spp. with potential bioactivities. They are active against phytopathogenic fungi and bind with their plasma membrane, thereby causing cell lysis and leakage (Gimenez et al., 2021).

The biosynthesis of these lipopeptides is governed by a complex set of proteins called non-ribosomal peptide synthetases (Wang et al., 2014). Gene clusters for the synthesis of such compounds have been found in the *Bacillus* genera (Donio et al., 2013). Several structurally and functionally diverse molecules need to be screened and characterized to create a better understanding of such biomolecules (Lu et al., 2014; Aleti et al., 2015; Singh and Tiwary, 2016).

Next-generation sequencing has revolutionized the discovery of natural products, chemicals, and biosynthetic enzymes greatly

used in the fields of biotechnology and biomedicine (Luo et al., 2014). Biosurfactant production in various *Bacillus* species has been widely reported in many studies. Moreover, many *Bacillus* species are known for their metabolic capability, environmental versatility, and their ability to remove xenobiotic compounds and heavy metals via biosurfactant production (Arora, 2020). However, only limited literature is available on the genetic basis of biosurfactant production in bacteria. To this end, our research is focused on the whole genome sequencing and analysis of biosurfactant-producing and hydrocarbon-degrading indigenous *Bacillus subtilis* strains MB415 and MB418 isolated from hydrocarbon-contaminated soil. The genome analysis of both strains also indicated the gene clusters for surfactin and fengycin lipopeptides.

MATERIALS AND METHODS

Isolation and Characterization of Biosurfactant-Producing Bacteria

Soil samples were collected from the oil-contaminated site of Missa Kiswal oil field, Gujjar Khan, Pakistan. Bacterial strains were isolated on Bushnell Hass mineral salt medium (Bushnell and Haas, 1941) by the spread plate method. Isolated bacterial strains were screened for biosurfactant production by a blood hemolysis assay (Mulligan et al., 1984). For this purpose, 5% defibrinated goat blood was added to the blood agar base. The emulsification activity (Cooper and Goldenberg, 1987) was determined by taking 2 ml of kerosene with an equal volume of the culture supernatant and vortexed vigorously for a few minutes. The emulsion formed was allowed to be stable for 24 h and the emulsification index was calculated as the percentage of the height of the emulsion layer (mm) divided by the total height of the liquid (mm). Furthermore, the oil-spreading technique and surface tension of culture supernatants were measured by using a torsion balance (Lin and Timasheff, 1996; Morikawa et al., 2000). Extraction and purification of biosurfactants from selected isolates were done by following the protocol of Qiao and Shao (2010). The extracted biosurfactants were analyzed through FTIR spectroscopy.

Whole Genome Sequencing

For molecular characterization, isolates were sequenced for 16S rRNAs that are submitted to NCBI. In addition, the whole genome sequence of both strains was performed to get an insight into the genetic composition. Cells were grown in Luria-Bertani (LB) broth at room temperature. The genomic DNA was extracted in the late exponential phase. The cells were collected by centrifugation at 4°C and 10,000 g for 10 min. The supernatant was discarded and the cells were suspended in 200 µL of PBS. The samples were then treated with lysozyme (1.33 mg/ml) and incubated for 5 min. Proteinase K (20 µL) was added to the cells that were incubated for 1 hour at 37°C. After this, 200 µL of buffer AL was added to the samples, mixed thoroughly, and kept at 56°C for 10 min. Furthermore, 96–100% ethanol (200 µL) was used for washing nucleic acid pellets. The mixture collected was centrifuged at 8,000 rpm for

TABLE 1 | Results of biosurfactant screening assays.

Bacterial isolates	Hemolysis	Oil spreading	Emulsification index (% α)	Surface tension (dynes/cm)
<i>Bacillus subtilis</i> MB415	A	+ (10s)	20 \pm 2	20 \pm 0.5
<i>Bacillus subtilis</i> MB418	B	+(2min)	15 \pm 2.5	35 \pm 5

α = incomplete hemolysis β = complete hemolysis.

1 min and the flow-through was discarded. Buffer AW2 (500 μ L) was added to the spin column, centrifuged at 14,000 rpm, and the flow-through was discarded. Finally, the DNA was eluted with an elution buffer AE (200 μ L) that was incubated for 1 min at room temperature and centrifuged at 6,000 rpm for 1 min to collect the DNA in a sterile Eppendorf tube (Aslam et al., 2018). The concentration and purity of DNA were determined by using a nanodrop spectrophotometer ND-1000 (Nanodrop Technology, Wilmington, DE). The qualitative assessment was done by visualizing genomic DNA on 1% agarose gel stained with ethidium bromide. The genomic DNA of both strains MB415 and MB418 were sequenced using the Illumina MiSeq (2 \times 300 bp) platform.

Sequence Assembly and Annotation

The obtained sequence reads were assembled through SPAdes 3.1 including error correction, deBruijn graph assembly for arrangement into contigs and their scaffolding (Bankevich et al., 2012). The quality of the assembled sequence data was assessed using Quast (Gurevich et al., 2013). Finally, the genomic data were annotated through different annotation servers.

Gene prediction and functional annotation analysis was performed using three different pipelines: Integrated Microbial Genomes expert review (<http://img.jgi.doe.gov>), Rapid Annotation Subsystem Technology (Aziz et al., 2008), and Genome Annotation Pipeline (PGAP) by NCBI (http://www.ncbi.nlm.nih.gov/genome/annotation_prok). The genes were identified by Prodigal (Hyatt et al., 2010). The predicted CDSs were translated and searched against NCBI, non-redundant databases, UniProt, TIGRFam, Pfam, PRIAM, COG, KEGG, and InterPro databases performed by the IMG ER platform that is developed by the Joint Genome Institute, Walnut Creek, CA, United States (Markowitz et al., 2009). In addition, transfer RNAs and rRNAs were identified using tRNAscan-SE (Lowe and Eddy, 1997) and RNAmmer (Lagesen et al., 2007).

RESULTS

Screening of Biosurfactant-Producing Bacteria

Fifteen hydrocarbon-degrading bacteria were isolated on Bushnell Haas Mineral salt medium from oil-polluted soil samples. Isolates with distinct morphological characteristics and competent to utilize different hydrocarbon substrates as a sole carbon source were selected, purified, and screened for biosurfactant production. Multiple screening assays were performed for the identification of potential biosurfactant producers, that is, oil spreading assay, emulsification index,

and surface tension and hemolytic assays. Two isolates, that is, MB415 and MB418 showed a remarkable reduction in surface tension values, that is, 20 and 35 dyn/cm, respectively. Both strains showed distinct oil displacement and low emulsification index. Strain MB418 exhibited complete, while MB415 displayed partial hemolysis of erythrocyte cells (Table 1).

In the present study, the maximum emulsification activity of diesel was recorded by culture supernatants of bacterial strains. The emulsification ability is attributed to the augmented biodegradation potential of petroleum hydrocarbons. The FTIR spectrum strongly reflects that the extracted compound was a potential biosurfactant with peptide and aliphatic hydrocarbon moieties. Significant absorbance peaks of the biosurfactant isolated from *B. subtilis* MB415 and MB418 observed C-O, C-H, and C=O in the regions of 1,000–1,320/cm, 2,850–3,000/cm, and 1,665–1760/cm stretching mode, respectively, verifying the presence of glycolipids (Figure 1).

Genome Sequencing

The genomes of *B. subtilis* strains MB415 and MB418 were sequenced to identify their hydrocarbon degradation potential and applications in bioremediation. The genome project was deposited in the Integrated Microbial Genome (IMG ER) online database and the Genome Online Database (GOLD) (Pagani et al., 2012). The whole-genome shotgun (WGS) project of both isolates was deposited at DDBJ/EMBL/GenBank under accession numbers LYDW00000000 and MQR000000000 for *B. subtilis* MB418 and MB415, accordingly.

Genome Properties

The genome sequences of *B. subtilis* MB415 and MB418 were generated with the help of Illumina Miseq sequenced data. For both the genomes, the paired end reads 4, 120, 685 bp and 4, 266, 089 bp respectively, were assembled into contigs and scaffolds. Genome statistics according to the NCBI Prokaryotic Genome Annotation Pipeline are summarized in Table 2.

Annotation Analysis and Subsystem Features

The draft genome of *B. subtilis* MB418 contains 4,543 genes containing 4,055 genes with coding sequences and 362 pseudogenes. The genome was found to have 126 RNA genes including 85 tRNAs and 35 rRNAs comprising 11 genes for 5S, 12 each for 16S and 23S rRNAs, while six were non-coding RNAs identified through RNAmmer and tRNA Scan S.E. The *B. subtilis* MB415 genome of 4, 267, 672 bps contains 4,812 genes with 4, 416 protein-coding genes and 121 rRNAs including 86 genes for

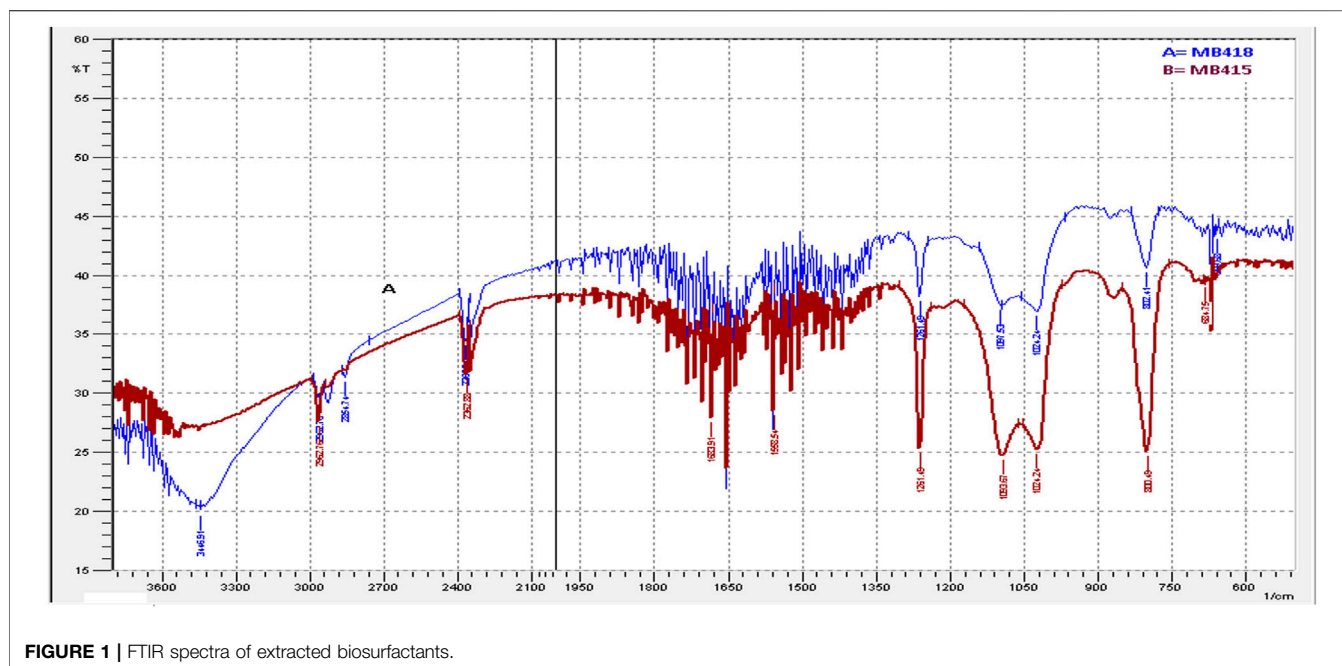


FIGURE 1 | FTIR spectra of extracted biosurfactants.

TABLE 2 | Genome characteristics of *Bacillus subtilis* MB415 and *Bacillus subtilis* MB418.

Strain	Accession number	Size Mbps	GC content	Number			
				Contigs	Subsystems	Cds	RNA
MB415	MQSR00000000	4267093	43.1	246	479	4,610	116
MB418	LYDW00000000	4118930	43.6	289	479	4,455	121

tRNAs and 11, 12, 7 genes carrying codes for 5S, 16S, and 23S respectively.

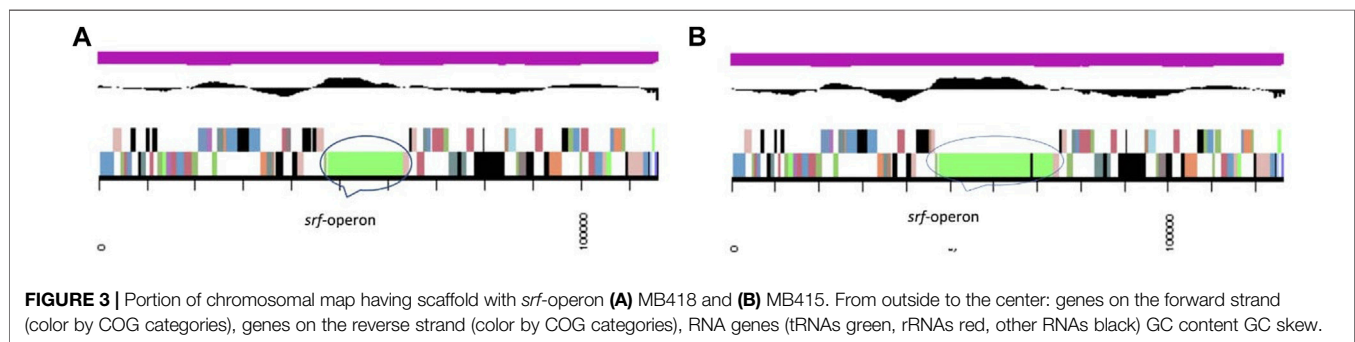
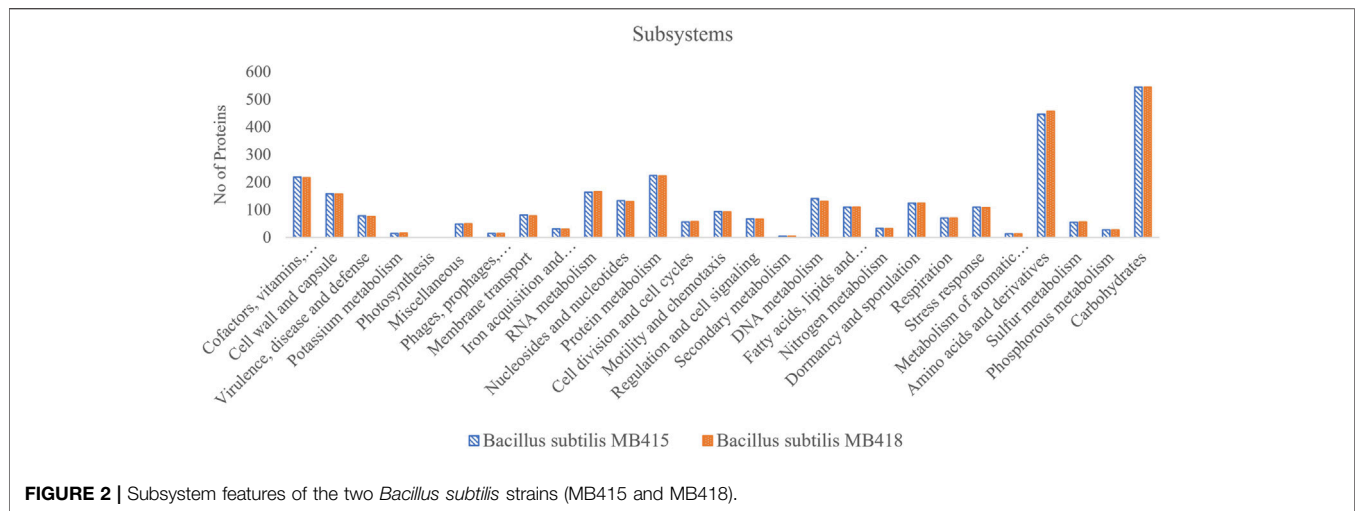
The annotation of the genome *B. subtilis* MB418 using the RAST server (Aziz et al., 2008) identified genes involved in the tolerance against metals, that is, arsenic (6 genes), copper (4 genes), and Co/Zn/Cd (4 genes). In addition, the genome has a multidrug resistance operon mdtRP mostly present in the genus *Bacillus*, and beta-lactamase, streptothricin, and vancomycin which are mostly found in gram-positive bacteria. Among 479 subsystems, 545 genes were specified for carbohydrates and 458 for amino acids and their derivatives. The genome also possesses a significant number of genes allocated for chemotaxis (15 genes) and flagellar motility (78 genes). There were around 117 genes which confer proteins for dormancy, sporulation-like genes designated for biofilm matrix protein component TsA, sporulation gene operons, spore germination, hydrolysis, and sporulation cluster IIIA and cluster II, while in *B. subtilis* MB415, many genes involved in metal tolerance including arsenic (6), copper (5), and Co/Zn/Cd resistance (4) were identified. However, few genes were found to be involved in antibiotic resistance among multidrug resistance operon efflux pumps (8) and beta-lactamase (3), along with genes for vancomycin, streptothricin, and fosfomycin resistance. Both genomes possess genes for secondary metabolite biosynthesis that includes surfactin, fengycin, and siderophore. In addition, genes

encoding enzymes involved in different aromatic compound metabolisms, that is, peripheral anaerobic and central aromatic intermediates formed during degradation were also predicted in both genomes (Figure 2).

Biosurfactant-Producing Genes

Emerging next-generation sequencing technology coupled with other computational efforts of annotation has enabled us to look into several potential genes simultaneously. This has improved our understanding of essential metabolic pathways and the adaptive variation of organisms in stressed environments. Similarly, bacterial diversity inhabiting soil with a high concentration of hydrocarbons (such as soil in and around oil fields) possessed certain genes for biosurfactant (surfactin, fengycin, and iturins) production.

Annotation results revealed that both genomes MB415 and MB418 carry operons responsible for the biosynthesis of secondary metabolites such as surfactin and fengycin. In *B. subtilis* MB418, surfactin-producing genes including *urfAA* codes for the surfactin family lipopeptide synthetase-A. This protein belonged to the COG1020 category of clusters of orthologous groups (COG) for the conserved protein domain family *EntF* called non-ribosomal peptide synthetase component F. In addition, it showed a close relationship with the phosphopantetheine attachment site protein (IPR006162) and

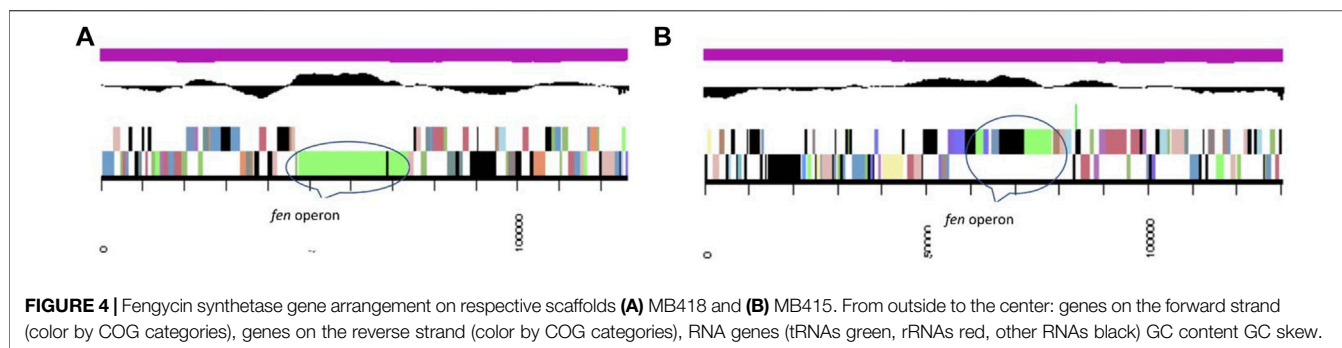


acyl carrier protein of super family SSF47336 (IPR009081). It also shares similarities with the peptide chains of IPR020806 polyketide synthase, phosphopantetheine-binding domain, and others including IPR020845, SSF56801, and TIGR01720 of the non-ribosomal peptide synthase domain. It was relevant to many other protein domains including PP-binding, condensation, AMP binding_C, and pfam13745-HXXPF_rpt. Next to this is the gene *srfAC* encode for the protein surfactin family lipopeptide synthetase C. This gene of 3,828 bp is translated into a protein of 1,275 amino acids, a non-ribosomal peptide synthetase component F. This protein belongs to the phosphopantetheine attachment site (IPR006162) and other protein classes of databases InterPro, pfam, and TIGRfam like *srfAA*.

The genome also possesses gene *srfAD* (729 bp) encodes for the thioesterase domain-containing protein comprising 242 amino acids. According to clusters of orthologous groups, it belongs to the category COG3208 surfactin synthase thioesterase subunit exhibiting 91.32% similarity with alpha/beta-hydrolases of pfam00975 and SSF53474. Lastly, the gene for *ycxA* similar to COG2271 called sugar-phosphate permease comprises 1,227 bp and encodes for 408 amino acid sequences. The genome encodes a permease of the major facilitator superfamily (KOG2533) related to the transporter class TC: 2. A. 1 of IPR016196, IPR020846 with 85.78% alignment with pfam07690.

This gene arrangement was different in *B. subtilis* MB415. The operon starts with the gene *srfAA* of 10,764 bp encode for the surfactin family lipopeptide synthetase A (COG1020). The acyl-CoA carrier protein, also called surfactin family lipopeptide synthetase B gene *srfAB* (3329bp), was identified in the genome of *B. subtilis* MB415. This protein of 3,329 amino acids has a maximum similarity with IPR009081 (acyl carrier protein-like), IPR020845 (AMP-binding, conserved site), and SSF52777 (CoA-dependent acyltransferases). Next to this is the gene for the condensation protein SrfAB of pfam00668. In addition, it also possesses the gene *srfAC* of 3828bp, *srfAD* thioesterase domain-containing protein, and *ycxA*-sugar phosphate permease. The arrangement of top COG hits on the scaffold with *srf*-operon in the genome of *B. subtilis* MB418 and MB415 (Figure 3).

Fengycin produced by various *Bacillus* strains is expected to form a lactone between the hydroxyl group of L-Tyr3 and C-terminal carboxyl group L-II-e and fengycin synthase (FenC, FenA, and FenB). In the MB418 genome, the gene *fenC* encoding fengycin family lipopeptide synthetase A (KO: K15664 ppsA) was identified (Figure 4). However, in case of the *Bacillus subtilis* MB415 genome, the operon of fengycin starts with the gene *fenB*, comprising 243 bps and belonging to fengycin lipopeptide synthetase E (KO: K15668). It contains the non-ribosomal synthetase component F of COG1020. This protein of



2,042 amino acids consists of fengycin family lipopeptide synthetase A (KO: K15664). In between these genes, a gene for the non-ribosomal peptide synthase and condensation domain-containing protein, amino-acid adenylation domain, AMP-binding enzyme C-terminal domain, and an uncharacterized lipoprotein *yddw* were identified.

DISCUSSION

In the present study, we investigated the biosurfactant production from the hydrocarbon degraders of the Missa Kiswal oilfield. Two *B. subtilis* strains were able to produce lipopeptide biosurfactants. Similarly, a study reported the strain *B. amyloliquefaciens* An6 as a potent biosurfactant producer with diesel oil degradation efficiency (Ayed et al., 2015). Hydrocarbon-degrading bacteria possess an innate potential to produce biosurfactants which aid them in the bioavailability of hydrocarbon substrates (Antoniou et al., 2015).

The genomes of the strains isolated in the present study also decipher *Srf* operons responsible for the biosynthesis of surfactin and fengycins. In the case of *B. subtilis*, mostly fengycin and surfactin have been reported in many studies (Pereira et al., 2013). Jadeja et al. (2019) reported the *sfp* and NPRS gene in *Bacillus* sp. AKBS9 and emulsan biosynthetic gene cluster in *Acinetobacter* sp. AKBS16 for biosurfactant production through the whole genome sequence analysis of the aforementioned strains.

B. subtilis is one of the predominant lipopeptide cell factories (Steller et al., 1999). The genetic makeup of biosurfactant-producing organisms is one of the major factors governing the biosynthesis of biosurfactants. Different molecular studies demonstrated the metabolic pathways, operons, and enzymes for the extracellular production of many biosurfactants. Wang et al. (2022) identified entire core synthetic genes for bacillibactin, fengycin, and surfactin production in a novel facultative-halophilic, long-chain hydrocarbon degrader, that is, *Bacillus* sp. XT-2. Surfactin is a cyclic lipopeptide produced non-ribosomally by surfactin synthetase, a multi-enzyme peptide synthetase complex. The biosynthesis of other lipopeptides is also mediated by similar enzyme complexes. Non-ribosomal peptide synthetases (NRPSs) involved in lipopeptide biosynthesis exhibit a high degree of similarity in structural conformation even in highly distant microbial species (Dhasayan et al., 2015). Fengycin is another cyclic

lipodecapeptide containing four D-amino acids and ornithine in the peptide chain. Fengycin looks like a mixture of isoforms that show differences in length and branching of the β -hydroxy fatty acid, and the amino-acid composition of the peptide ring (Hu et al., 2007).

The genomes presented in our study decrypt the *B. subtilis* sequence variation and detailed comparisons of this vital model species. Comparisons of whole-genome sequences of different bacteria revealed significant genomic variability among phylogenetically interrelated bacterial species. Even sequences with 100% homology exhibit low conservation in the total gene content. *B. subtilis* is one of the most extensively studied spore forming, non-pathogenic, gram-positive bacteria ubiquitously found in soil (Earl et al., 2007; Chen et al., 2013). Another study demonstrated the complete genome sequence analysis of polystyrene-degrading deep sea *B. paralicheniformis* G1 strain comprising 4,281,959 bp with 45.88% GC content and encoded 4,213 protein-coding genes. Numerous genes encoding monooxygenase, dioxygenase, peroxidase, esterase, and hydrolase involved in the degradation of synthetic polymers were identified along with the genes associated with flagellum-dependent motility, chemotaxis, biofilm formation, and siderophores biosynthesis (Kumar et al., 2021).

The genetic regulation of lipopeptides is governed by the four open reading frames (ORFs) in the *srfA* operon directing the surfactin synthesis including *srfAA*, *srfAB*, *srfAC*, and *srfAD* (Zhu et al., 2021). Surfactin is synthesized by a multi-enzyme synthetase complex mainly comprising *SrfA*, *SrfB*, and *SrfD* subunits. *Srf* operon for biosurfactant pathways consists of the gene *srfAA*, *-AB*, *-AC*, and *-AD* in *Bacillus* spp. The annotation analysis of the genome *B. subtilis* MB415 and MB418 revealed that both strains also possess this operon (Figure 5). According to previous studies, in the *Bacillus subtilis* species, the synthesis of surfactin is based on three amino-acid activating components of surfactin synthetase (*SrfA*, *-B*, and *-C*) that are activated by *SrfD*. The gene *srfAD* also possesses thioesterase that is required for the last amino acid in the growing peptide chain that results in the cyclic structure of the biosurfactant and *B. subtilis* (Vedaraman and Venkatesh, 2011).

A comparative genome analysis showed gene arrangements of surfactin-producing operon in the draft genome of *B. subtilis* MB418 shared its neighborhood region showing the same top COG hits with *Bacillus subtilis inaquosorum* BGSC 3A28 and *B. subtilis* B4143. With the difference in the *SrfAA* protein, in

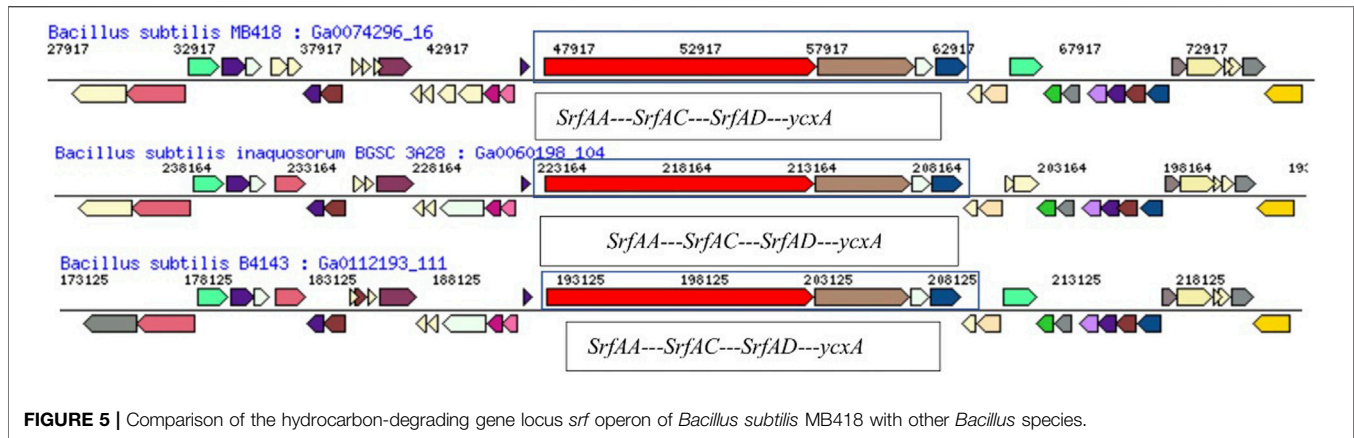


FIGURE 5 | Comparison of the hydrocarbon-degrading gene locus *srf* operon of *Bacillus subtilis* MB418 with other *Bacillus* species.

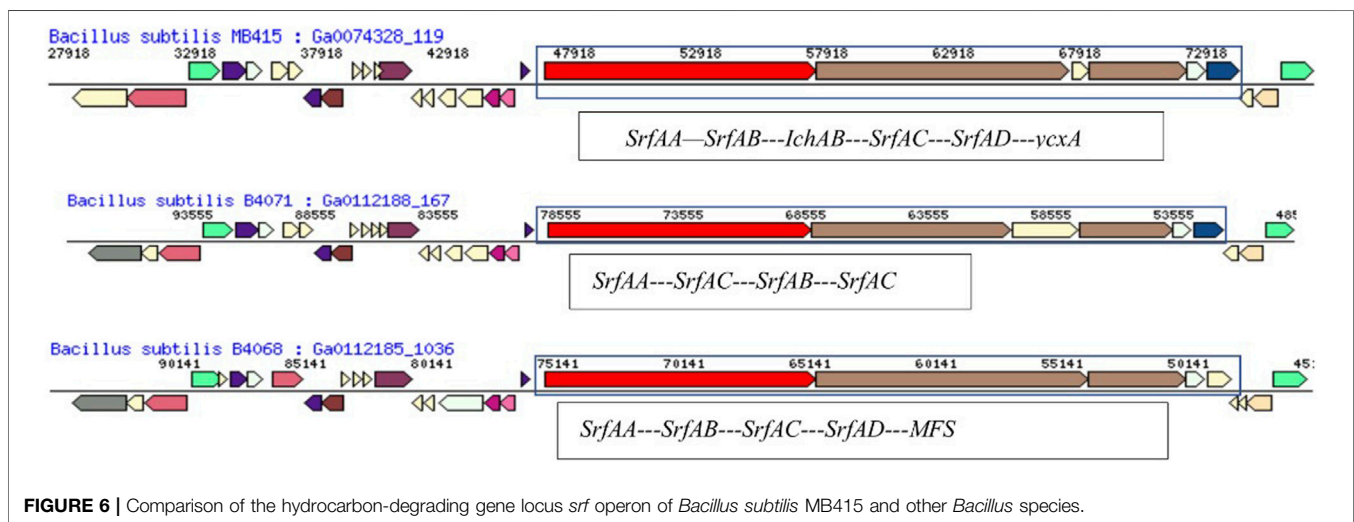


FIGURE 6 | Comparison of the hydrocarbon-degrading gene locus *srf* operon of *Bacillus subtilis* MB415 and other *Bacillus* species.

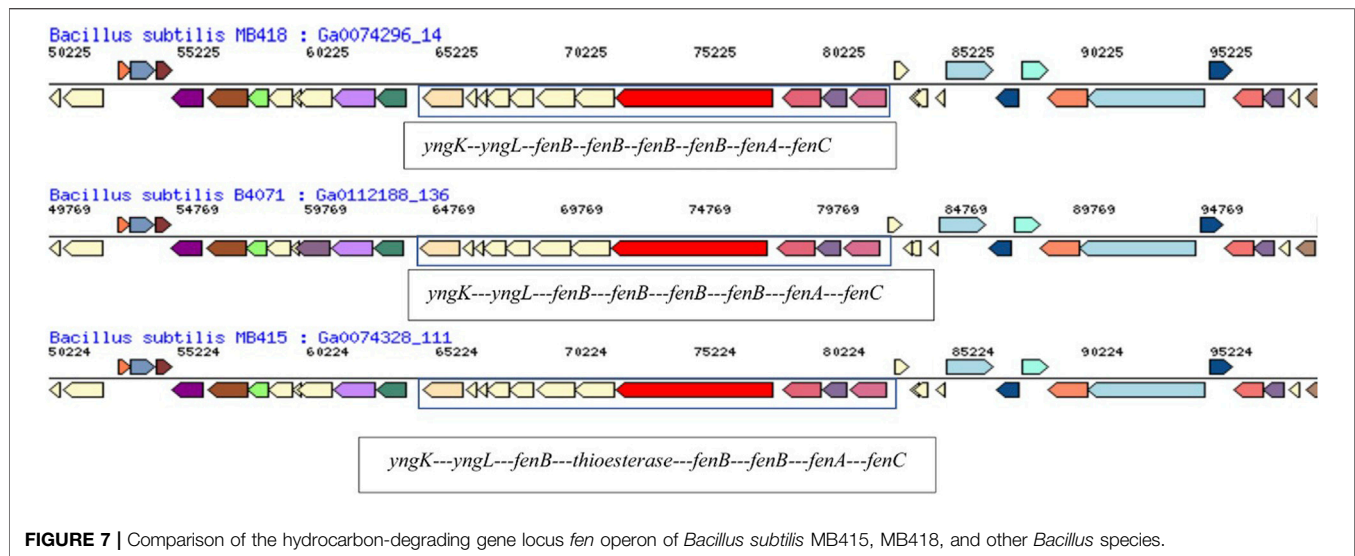
Bacillus subtilis MB418 there are 3,586 amino-acid sequences, while it is 3,587 amino acids in the rest of the two closely related genomes. However, in the case of MB415, the operon is entirely different with genes encoding for more proteins and a different arrangement as shown in **Figure 6**.

The gene *srfAB* encodes for surfactin family lipopeptide synthetase AB and consists of 3,329 amino acids. A comparison of the gene neighborhood showed maximum similarity with *B. subtilis* B4071 (Berendsen et al., 2016) having the gene for the protein SrfAB (2695aa) (**Figure 6**). With the exception, *B. subtilis* MB415 had an additional gene for non-ribosomal peptide synthase not present in the genome of B4071. The gene operon in the genome of MB415 ends at gene (*ycxA*) for sugar-phosphate permease, while in the *B. subtilis* MB4068 (Berendsen et al., 2016) operon, this gene is replaced with another gene for the major facilitator superfamily (**Figure 6**).

Furthermore, genomes which carry genes for the same proteins vary in sequence length than the ones present in the draft genome of MB415. For example, the *srfAA* gene of *B. subtilis* MB415 is of 10,764 bp codes for 3,587 amino acids. However, this gene sequence is relatively small in the previously reported

surfactant-producing *B. safensis* CCMA-560 and *B. amyloliqueficiens* comprising 10, 713bps (3570aa) and 10, 701bps (3566aa), respectively. The gene *srfAC* (3828bp) sequence is the same in the genomes of *B. subtilis* MB418 and MB415 and in the rest of the *Bacillus* species. The gene encoding thioesterase in the genomes of both *B. subtilis* MB415 and MB418 comprises 3,834 bps while in *B. pumilus*, it is encoded by the *srfAD* gene (732bp) (Jiang et al., 2016). Similarly, surfactin produced by *B. subtilis* strains isolated from soil consists of seven amino acids and 13–15 lipid chains forming a cyclic lipopeptide as described previously by researchers (Sekhon et al., 2011; Sachdev and Cameotra, 2013).

The genome analysis of both isolates unveiled potential gene operons for the biosynthesis of a secondary metabolite, fengycin. Fengycin, also called plipastatin, is an anti-fungal antibiotic that inhibits filamentous fungi but is ineffective against yeast and bacteria (Volpon et al., 2000). It also has the potential to inhibit phospholipase and biofilm formation in many bacteria (Thaniyavarn et al., 2003). Fengycin is a type of lipodecapeptide with moderate surfactant activity normally produced by several strains of *Bacillus* spp. *Fen* operon



comprises an N-acyl domain at the N-terminus of *FenC* and five non-ribosomal peptide synthetase subunits assembled in a co-linear chain in the following order: *FenC*, *FenD*, *FenE*, *FenA*, and *FenB* (Guo and Yu, 2014).

The fengycin synthetase gene operon in the draft genome of *B. subtilis* MB418 exhibited maximum similarity with *B. subtilis* B4071. However, in the genome of *B. subtilis* MB415, an esterase-producing gene is identified instead of *fenB* for fengycin lipopeptide synthetase E as indicated in other closely related *Bacillus* species (Figure 7). A recent study also reported three genes related to the gene cluster responsible for fengycin biosynthesis (*fenBCD*) in *B. velezensis* and *B. siamensis* (Zeng et al., 2021). Similarly, in another study, five ORFs, namely, *FenA*, *FenB*, *FenC*, *FenD*, and *FenE*, were identified for fengycin production in the *Bacillus velezensis* strain P45 (da Rosa et al., 2022).

CONCLUSION

Indigenous bacterial strains inhabiting a polluted environment acquire tremendous potential to utilize petroleum hydrocarbons than non-native microbes. The present study demonstrated the production of lipopeptide biosurfactants from *B. subtilis* MB415 and MB418. The genome analysis of both *B. subtilis* isolates

MB415 and MB418 contain a gene locus that enhanced their proficiency in degrading oil compounds. The presence of genes for flagella and secondary metabolite biosynthesis supported the experimental outcomes of emulsification and oil utilization capacities of both strains.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repository and accession number(s) can be found at: <https://www.ncbi.nlm.nih.gov/genbank/>, LYDW00000000, <https://www.ncbi.nlm.nih.gov/genbank/>, MQSR00000000.

AUTHOR CONTRIBUTIONS

The authors have contributed equally to this work. AY conceived the research idea and administered the research work. AF performed the laboratory experiments. FA performed the genome sequencing. All authors contributed to the writing of this manuscript. All authors have read and agreed to the publication of the manuscript.

REFERENCES

- Aleti, G., Sessitsch, A., and Brader, G. (2015). Genome Mining: Prediction of Lipopeptides and Polyketides from *Bacillus* and Related Firmicutes. *Comput. Struct. Biotechnol. J.* 13, 192–203. doi:10.1016/j.csbj.2015.03.003
- Antonou, E., Fodelianakis, S., Korkakaki, E., and Kalogerakis, N. (2015). Biosurfactant Production from marine Hydrocarbon-Degrading Consortia and Pure Bacterial Strains Using Crude Oil as Carbon Source. *Front. Microbiol.* 06, 274. doi:10.3389/fmicb.2015.00274
- Anuradha, S. (2010). Structural and Molecular Characteristics of Lichenysin and its Relationship With Surface Activity. *Biosurfactants* 672, 304–315. doi:10.1007/978-1-4419-5979-9_23
- Arora, P. K. (2020). Bacilli-mediated Degradation of Xenobiotic Compounds and Heavy Metals. *Front. Bioeng. Biotechnol.* 8, 570307. doi:10.3389/fbioe.2020.570307
- Aslam, F., Yasmin, A., and Thomas, T. (2018). Essential Gene Clusters Identified in *Stenotrophomonas* MB339 for Multiple Metal/antibiotic Resistance and Xenobiotic Degradation. *Curr. Microbiol.* 75 (11), 1484–1492. doi:10.1007/s00284-018-1549-2
- Aziz, R. K., Bartels, D., Best, A. A., DeJongh, M., Disz, T., Edwards, R. A., et al. (2008). The RAST Server: Rapid Annotations Using Subsystems Technology. *BMC Genomics* 9 (1), 75. doi:10.1186/1471-2164-9-75
- Bankevich, A., Nurk, S., Antipov, D., Gurevich, A. A., Dvorkin, M., Kulikov, A. S., et al. (2012). SPAdes: A New Genome Assembly Algorithm and its Applications to Single-Cell Sequencing. *J. Comput. Biol.* 19 (5), 455–477. doi:10.1089/cmb.2012.0021

- Ben Ayed, H., Jemil, N., Maalej, H., Bayouhd, A., Hmidet, N., and Nasri, M. (2015). Enhancement of Solubilization and Biodegradation of Diesel Oil by Biosurfactant from *Bacillus Amylolyquefaciens* An6. *Int. Biodeterioration Biodegradation* 99, 8–14. doi:10.1016/j.ibiod.2014.12.009
- Berendsen, E. M., Wells-Bennik, M. H., Krawczyk, A. O., de Jong, A., van Heel, A., Eijlander, R. T., et al. (2016). Draft Genome Sequences of 10 *Bacillus Subtilis* Strains that Form Spores with High or Low Heat Resistance. *Genome Announc* 4 (2), e00124–16. doi:10.1128/genomeA.00124-16
- Bhatt, P., Verma, A., Gangola, S., Bhandari, G., and Chen, S. (2021). Microbial Glycoconjugates in Organic Pollutant Bioremediation: Recent Advances and Applications. *Microb. Cel Factories*. 20 (1), 1–18. doi:10.1186/s12934-021-01556-9
- Borah, D., Chaubey, A., Sonowal, A., Gogoi, B., and Kumar, R. (2021). “Microbial Biosurfactants and Their Potential Applications: An Overview,” in *Microbial Biosurfactants. Environmental and Microbial Biotechnology*. Editors Ahamed, M. I. Inamuddin and R. Prasad (Singapore: Springer). doi:10.1007/978-981-15-6607-3_5
- Bushnell, L. D., and Haas, H. F. (1941). The Utilization of Certain Hydrocarbons by Microorganisms. *J. Bacteriol.* 41 (5), 653–673. doi:10.1128/jb.41.5.653-673.1941
- Chen, Y., Yan, F., Chai, Y., Liu, H., Kolter, R., Losick, R., et al. (2013). Biocontrol of Tomato Wilt Disease by *Bacillus Subtilis* Isolates from Natural Environments Depends on Conserved Genes Mediating Biofilm Formation. *Environ. Microbiol.* 15 (3), 848–864. doi:10.1111/j.1462-2920.2012.02860.x
- Cooper, D. G., and Goldenberg, B. G. (1987). Surface-active Agents from Two *Bacillus* Species. *Appl. Environ. Microbiol.* 53 (2), 224–229. doi:10.1128/aem.53.2.224-229.1987
- Dhasayan, A., Selvin, J., and Kiran, S. (2015). Biosurfactant Production from marine Bacteria Associated with Sponge *Callyspongia Diffusa*. 3 *Biotech.* 5 (4), 443–454. doi:10.1007/s13205-014-0242-9
- Donio, M. B. S., Ronica, F. A., Vijji, V. T., Velmurugan, S., Jenifer, J. S. C. A., Michaelbabu, M., et al. (2013). Halomonas Sp. BS4, A Biosurfactant Producing Halophilic Bacterium Isolated from Solar Salt Works in India and Their Biomedical Importance. *SpringerPlus* 2 (1), 149. doi:10.1186/2193-1801-2-149
- Earl, A. M., Losick, R., and Kolter, R. (2007). *Bacillus Subtilis* Genome Diversity. *J. Bacteriol.* 189, 1163–1170. doi:10.1128/jb.01343-06
- Fanaei, M., Jurcic, K., and Entiazi, G. (2021). Detection of Simultaneous Production of Kurstakin, Fengycin and Surfactin Lipopeptides in *Bacillus Mojavensis* Using a Novel Gel- Based Method and MALDI-TOF Spectrometry. *World J. Microbiol. Biotechnol.* 37 (6), 111. doi:10.1007/s11274-021-03064-9
- Femina Carolin, C., Senthil Kumar, P., Chitra, B., Fetcia Jackulin, C., and Ramamurthy, R. (2021). Stimulation of *Bacillus Sp.* By Lipopeptide Biosurfactant for the Degradation of Aromatic Amine 4-Chloroaniline. *J. Hazard. Mater.* 415, 125716. doi:10.1016/j.jhazmat.2021.125716
- Ganesh Kumar, A., Hinduja, M., Sujitha, K., Nivedha Rajan, N., and Dharani, G. (2021). Biodegradation of Polystyrene by Deep-Sea *Bacillus Paralicheniformis* G1 and Genome Analysis. *Sci. Total Environ.* 774, 145002. doi:10.1016/j.scitotenv.2021.145002
- Geys, R., Soetaert, W., and Van Bogaert, I. (2014). Biotechnological Opportunities in Biosurfactant Production. *Curr. Opin. Biotechnol.* 30, 66–72. doi:10.1016/j.copbio.2014.06.002
- Gimenez, D., Phelan, A., Murphy, C., and Cobb, S. L. (2021). Fengycin A Analogues with Enhanced Chemical Stability and Antifungal Properties. *Org. Lett.* 23, 4672–4676. doi:10.1021/acs.orglett.1c01387
- Gudiña, E. J., Rangarajan, V., Sen, R., and Rodrigues, L. R. (2013). Potential Therapeutic Applications of Biosurfactants. *Trends Pharmacol. Sci.* 34, 667–675. doi:10.1016/j.tips.2013.10.002
- Guo, J., and Yu, J. (2014). Sorption Characteristics and Mechanisms of Pb(II) from Aqueous Solution by Using Bioflocculant MBFR10543. *Appl. Microbiol. Biotechnol.* 98 (14), 6431–6441. doi:10.1007/s00253-014-5681-z
- Gurevich, A., Saveliev, V., Vyahhi, N., and Tesler, G. (2013). QUAST: Quality Assessment Tool for Genome Assemblies. *Bioinfo* 29 (8), 1072–1075. doi:10.1093/bioinformatics/btt086
- Hu, F., Liu, Y., and Li, S. (2019). Rational Strain Improvement for Surfactin Production: Enhancing the Yield and Generating Novel Structures. *Microb. Cel fact.* 18 (1), 42–13. doi:10.1186/s12934-019-1089-x
- Hu, L. B., Shi, Z. Q., Zhang, T., and Yang, Z. M. (2007). Fengycin Antibiotics Isolated from B-F501 Culture Inhibit the Growth of *Fusarium moniliforme* Sheldon ATCC 38932. *FEMS Microbiol. Lett.* 272 (1), 91–98. doi:10.1111/j.1574-6968.2007.00743.x
- Hyatt, D., Chen, G.-L., LoCascio, P. F., Land, M. L., Larimer, F. W., and Hauser, L. J. (2010). Prodigal: Prokaryotic Gene Recognition and Translation Initiation Site Identification. *BMC Bioinformatics* 11 (1), 119. doi:10.1186/1471-2105-11-119
- Jadeja, N. B., Moharir, P., and Kapley, A. (2019). Genome Sequencing and Analysis of Strains *Bacillus Sp.* AKBS9 and *Acinetobacter Sp.* AKBS16 for Biosurfactant Production and Bioremediation. *Appl. Biochem. Biotechnol.* 187 (2), 518–530. doi:10.1007/s12010-018-2828-x
- Jiang, J., Gao, L., Bie, X., Lu, Z., Liu, H., Zhang, C., et al. (2016). Identification of Novel Surfactin Derivatives from NRPS Modification of *Bacillus Subtilis* and its Antifungal Activity against *Fusarium Moniliforme*. *BMC Microbiol.* 16 (1), 31. doi:10.1186/s12866-016-0645-3
- Lagesen, K., Hallin, P., Rodland, E. A., Stærfeldt, H.-H., Rognes, T., and Ussery, D. W. (2007). RNAMmer: Consistent and Rapid Annotation of Ribosomal RNA Genes. *Nucleic Acids Res.* 35 (9), 3100–3108. doi:10.1093/nar/gkm160
- Lin, T.-Y., and Timasheff, S. N. (1996). On the Role of Surface Tension in the Stabilization of Globular Proteins. *Protein Sci.* 5 (2), 372–381. doi:10.1002/pro.5560050222
- Liu, Q., Lin, J., Wang, W., Huang, H., and Li, S. (2015). Production of Surfactin Isoforms by *Bacillus Subtilis* BS-37 and its Applicability to Enhanced Oil Recovery under Laboratory Conditions. *Biochem. Eng. J.* 93, 31–37. doi:10.1016/j.bej.2014.08.023
- Lowe, T. M., and Eddy, S. R. (1997). tRNAscan-SE: a Program for Improved Detection of Transfer RNA Genes in Genomic Sequence. *Nucleic Acids Res.* 25 (5), 955–964. doi:10.1093/nar/25.5.955
- Lu, J., Guo, C., Zhang, M., Lu, G., and Dang, Z. (2014). Biodegradation of Single Pyrene and Mixtures of Pyrene by a Fusant Bacterial Strain F14. *Int. Biodeterioration Biodegradation* 87, 75–80. doi:10.1016/j.ibiod.2013.11.004
- Luo, C., Liu, X., Zhou, H., Wang, X., and Chen, Z. (2014). Identification of Four NRPS Gene Clusters in *Bacillus Subtilis* 916 for Four Families of Lipopeptides Biosynthesis and Evaluation of Their Intricate Functions to the Typical Phenotypic Features. *Appl. Environ. Microbiol.* 81, 422–431. doi:10.1128/AEM.02921-14
- Markowitz, V. M., Mavromatis, K., Ivanova, N. N., Chen, I.-M. A., Chu, K., and Kyrpides, N. C. (2009). IMG ER: a System for Microbial Genome Annotation Expert Review and Curation. *Bioinformatics* 25 (17), 2271–2278. doi:10.1093/bioinformatics/btp393
- Mishra, S., Lin, Z., Pang, S., Zhang, Y., Bhatt, P., and Chen, S. (2021). Biosurfactant Is a Powerful Tool for the Bioremediation of Heavy Metals from Contaminated Soils. *J. Hazard. Mater.* 418, 126253. doi:10.1016/j.jhazmat.2021.126253
- Morikawa, M., Hirata, Y., and Imanaka, T. (2000). A Study on the Structure-Function Relationship of Lipopeptide Biosurfactants. *Biochim. Biophys. Acta (Bba) - Mol. Cel Biol. Lipids* 1488 (3), 211–218. doi:10.1016/s1388-1981(00)00124-4
- Mulligan, C., Cooper, D., and Neufeld, R. (1984). Selection of Microbes Producing Biosurfactants in media without Hydrocarbons. *J. Ferment. Technol.* 62, 311–314.
- Pagani, I., Liolios, K., Jansson, J., Chen, I.-M. A., Smirnova, T., Nosrat, B., et al. (2012). The Genomes OnLine Database (GOLD) v.4: Status of Genomic and Metagenomic Projects and Their Associated Metadata. *Nucleic Acids Res.* 40 (D1), D571–D579. doi:10.1093/nar/gkr1100
- Pereira, J. F. B., Gudiña, E. J., Costa, R., Vitorino, R., Teixeira, J. A., Coutinho, J. A. P., et al. (2013). Optimization and Characterization of Biosurfactant Production by *Bacillus Subtilis* Isolates towards Microbial Enhanced Oil Recovery Applications. *Fuel* 111, 259–268. doi:10.1016/j.fuel.2013.04.040
- Qiao, N., and Shao, Z. (2010). Isolation and Characterization of a Novel Biosurfactant Produced by Hydrocarbon-Degrading bacterium *Alcanivorax dieselolei* B-5. *J. Appl. Microbiol.* 108, 1207–1216. doi:10.1111/j.1365-2672.2009.04513.x
- Raaijmakers, J. M., De Bruijn, I., Nybroe, O., and Ongena, M. (2010). Natural Functions of Lipopeptides from *Bacillus* and *Pseudomonas*: More Than Surfactants and Antibiotics. *FEMS Microbiol. Rev.* 34 (6), 1037–1062. doi:10.1111/j.1574-6976.2010.00221.x
- Rosa, C. E., Pinilla, C. M. B., Stincone, P., Pereira, J. Q., Varela, A. P. M., Mayer, F. Q., et al. (2022). Genomic Characterization and Production of Antimicrobial Lipopeptides by *Bacillus Velezensis* P45 Growing on Feather By-products. *J. Appl. Microbiol.* 132 (3), 2067–2079. doi:10.1111/jam.15363

- Sachdev, D. P., and Cameotra, S. S. (2013). Biosurfactants in Agriculture. *Appl. Microbiol. Biotechnol.* 97 (3), 1005–1016. doi:10.1007/s00253-012-4641-8
- Santos, V. S. V., Silveira, E., and Pereira, B. B. (2018). Toxicity and Applications of Surfactin for Health and Environmental Biotechnology. *J. Toxicol. Environ. Health B Crit. Rev.* 21 (6-8), 382–399. doi:10.1080/10937404.2018.1564712
- Sekhon, K., Khanna, S., and Cameotra, S. (2011). Enhanced Biosurfactant Production through Cloning of Three Genes and Role of Esterase in Biosurfactant Release. *Microb. Cell Fact.* 10 (1), 49. doi:10.1186/1475-2859-10-49
- Singh, P., and Tiwary, B. N. (2016). Isolation and Characterization of Glycolipid Biosurfactant Produced by a Pseudomonas Otitidis Strain Isolated from Chirimiri Coal Mines, India. *Bioresour. Bioproc.* 3 (1), 42. doi:10.1186/s40643-016-0119-3
- Steller, S., Vollenbroich, D., Leenders, F., Stein, T., Conrad, B., Hofemeister, J., et al. (1999). Structural and Functional Organization of the Fengycin Synthetase Multienzyme System from *Bacillus Subtilis* B213 and A1/3. *Chem. Biol.* 6 (1), 31–41. doi:10.1016/s1074-5521(99)80018-0
- Thaniyavarn, J., Roongsawang, N., Kameyama, T., Haruki, M., Imanaka, T., Morikawa, M., et al. (2003). Production and Characterization of Biosurfactants from *Bacillus licheniformis* F2.2. *Biosci. Biotechnol. Biochem.* 67 (6), 1239–1244. doi:10.1271/bbb.67.1239
- Vedaraman, N., and Venkatesh, N. (2011). Production of Surfactin by *Bacillus Subtilis* MTCC 2423 from Waste Frying Oils. *Braz. J. Chem. Eng.* 28 (2), 175–180. doi:10.1590/s0104-66322011000200001
- Volpon, L., Besson, F., and Lancelin, J.-M. (2000). NMR Structure of Antibiotics Plipastatins A and B from *Bacillus Subtilis* Inhibitors of Phospholipase A2. *FEBS Lett.* 485 (1), 76–80. doi:10.1016/s0014-5793(00)02182-7
- Wang, H., Fewer, D. P., Holm, L., Rouhiainen, L., and Sivonen, K. (2014). Atlas of Nonribosomal Peptide and Polyketide Biosynthetic Pathways Reveals Common Occurrence of Nonmodular Enzymes. *Proc. Natl. Acad. Sci. U.S.A.* 111 (25), 9259–9264. doi:10.1073/pnas.1401734111
- Wang, X.-T., Liu, B., Li, X.-Z., Lin, W., Li, D.-A., Dong, H., et al. (2022). Biosurfactants Produced by Novel Facultative-Halophilic *Bacillus Sp.* XT-2 with Biodegradation of Long Chain N-Alkane and the Application for Enhancing Waxy Oil Recovery. *Energy* 240, 122802. doi:10.1016/j.energy.2021.122802
- Zeng, Q., Xie, J., Li, Y., Chen, X., Gu, X., Yang, P., et al. (2021). Organization, Evolution and Function of Fengycin Biosynthesis Gene Clusters in the *Bacillus Amyloliquefaciens* Group. *Phytopathol. Res.* 3 (1), 1–12. doi:10.1186/s42483-021-00103-z
- Zhu, Z., Zhang, B., Cai, Q., Cao, Y., Ling, J., Lee, K., et al. (2021). A Critical Review on the Environmental Application of Lipopeptide Micelles. *Bioresour. Tech.* 339, 125602. doi:10.1016/j.biortech.2021.125602

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Yasmin, Aslam and Fariq. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.