BMC Genomics

RESEARCH

Open Access

From the depths of the Java Trench: genomic analysis of *Priestia flexa* JT4 reveals bioprospecting and lycopene production potential

Ocky K. Radjasa[1,13*], Ray Steven[2], Yosua Natanael[2], Husna Nugrahapraja[2,13], Septhy K. Radjasa[2], Tati Kristianti[3], Maelita R. Moeis[4], Joko P. Trinugroho[5], Haekal B. Suharya[2], Alfito O. Rachmatsyah[2], Ari Dwijayanti[6], Mutiara R. Putri[7], Charlie E. de Fretes[1], Zen L. Siallagan[1,2], Muhammad Fadli[1], Rafidha D. A. Opier[1], Jandinta D. Farahyah[2], Viana Rahmawati[2], Meirifa Rizanti[2], Zalfa Humaira[8], Ary S. Prihatmanto[9], Nugroho D. Hananto[10], R. Dwi Susanto[11,12], Agus Chahyadi[13], Elfahmi[13,14], Neil Priharto[2], Kamarisima[2] and Fenny M. Dwivany[2,13*]

Abstract

Background The marine environment boasts distinctive physical, chemical, and biological characteristics. While numerous studies have delved into the microbial ecology and biological potential of the marine environment, exploration of genetically encoded, deep-sea sourced secondary metabolites remains scarce. This study endeavors to investigate marine bioproducts derived from deep-sea water samples at a depth of 1,000 m in the Java Trench, Indonesia, utilizing both culture-dependent and whole-genome sequencing methods.

Results Our efforts led to the successful isolation and cultivation of a bacterium *Priestia flexa* JT4 from the water samples, followed by comprehensive genome sequencing. The resultant high-quality draft genome, approximately 4 Mb, harbored 5185 coding sequences (CDSs). Notably, 61.97% of these CDSs were inadequately characterized, presenting potential novel CDSs. This study is the first to identify the "open-type" ($\alpha < 1$) pangenome within the genus *Priestia*. Moreover, our analysis uncovered eight biosynthetic gene clusters (BGCs) using the common genome mining pipeline, antiSMASH. Two non-ribosomal peptide synthetase (NRPS) BGCs within these clusters exhibited the potential to generate novel biological compounds. Noteworthy is the confirmation that the terpene BGC in *P. flexa* JT4 can produce lycopene, a compound in substantial industrial demand. The presence of lycopene in the *P. flexa* JT4 cells was verified using Ultra-performance liquid chromatography-mass spectrometry (UPLC-MS/MS) in multiple reaction modes.

Conclusions This study highlights the bioprospecting opportunity to explore novel bioproducts and lycopene compounds from *P. flexa* JT4. It marks the pioneering exploration of deep-sea bacterium bioprospecting in Indonesia, seeking to unveil novel bioproducts and lycopene compounds through a genome mining approach.

*Correspondence:
Ocky K. Radjasa
ocky001@brin.go.id
Fenny M. Dwivany
fennym@itb.ac.id
Full list of author information is available at the end of the article

Keywords Deep-sea waters, Java Trench, Whole genome sequencing, *Priestia flexa*, Biosynthetic gene cluster, Bioprospecting

Background

The deep sea, a marine zone situated 200 m below sea level, is deemed an extreme living environment for bacteria due to the absence of sunlight, high pressure, low temperature, and lack of oxygen [1, 2]. Bacteria adapt to these environmental conditions by producing secondary metabolites, enhancing their survival in such extreme environments [2, 3]. These secondary metabolites exhibit various biological activities, including antioxidant, antibacterial, antiviral, antifungal, and anticancer properties, making them valuable for human applications. Consequently, secondary metabolites find extensive use in industrial products across pharmaceutical, food, cosmetics, agriculture, and aquaculture sectors [2].

Indonesia, the world's largest archipelagic country and one of the most biodiverse, possesses an ocean area that constitutes 77% of its total area, with depths exceeding 200 m covering 68% of the total ocean area, particularly in Eastern Indonesia (Flores Sea, Banda Sea, and Makassar Strait) and the open ocean (the Indian and Pacific Oceans) [4]. This vast expanse offers an enormous potential for advantageous biological compounds. Despite this, a mere 128 original research papers on Indonesia's deep sea have been published from 1872 to 2017, with only 16% focusing on the biological activities of the marine deep sea [5]. Genetically encoded deep-sea sourced secondary metabolites have been underexplored due to extreme conditions (high pressure, high temperature, and the absence of light), limited access, and the costs associated with sampling at greater depths.

Current conventional biochemical methods used to screen and isolate secondary metabolites from bacterial species suffer from limitations, including labor-intensive processes and varying cultivation parameters (e.g., light, pH, aeration, temperature, and nutrients) [6]. Additionally, several biosynthetic gene clusters (BGCs) remain silent or cryptic under standard laboratory growth conditions, hindering their exploration [7]. To overcome these challenges, genome mining methods, utilizing genome sequencing, computational analysis, and compound characterization, have proven effective in identifying novel natural bioproducts in bacteria [8, 9].

Our earlier research employed a Biosynthetic Gene Cluster Profiling analysis to unveil potential hidden metabolites within the bacterium *Virgibacillus salarius* 19.PP.Sc1.6 from the North Java Sea [10]. Given the underexplored nature of marine bioresources, further investigations employing a genomic-driven exploration strategy are imperative to discover unknown and potentially novel bioproducts [11–13].

Industries currently rely on various secondary metabolites (SM) with diverse bioactivities for their products. Lycopene, one of the most widely used SMs, finds applications in cosmetics, food coloring, health and diet, and pharmaceuticals. The increasing demand for lycopene in the market underscores its positive trend, with biotechnological methods utilizing microbes for lycopene biosynthesis still prevalent [14–16].

In contemporary times, lycopene stands as a key ingredient in many dietary supplements due to its health benefits on the eyes, prostate, skin, and cardiovascular system [17]. Its antioxidant activity is harnessed in the cosmetic industry, particularly in anti-UV light creams, contributing to skin cell regeneration and revitalization. Furthermore, lycopene exhibits rejuvenation properties that decelerate the aging process of skin cells by regulating the formation of procollagen (collagen precursor) [18]. The global lycopene market is estimated to have reached 107.2 million dollars in 2020 and is predicted to grow to 187.3 million dollars by 2030, with a compound annual growth rate (CAGR) of 5.2% from 2021 to 2030 [19].

While a prior study successfully mapped the biosynthetic potential of microbial genomes in the global ocean [12], no microbial genomes from the Indonesian deep seas have been reported. Unraveling the hidden potential of marine bacterial genomics, especially in countries with rich marine biodiversity like Indonesia, is essential. This study explores the genome of *Priestia flexa* isolated and cultured from deep-sea water samples at a depth of 1,000 m in the Java Trench, Indonesia, using whole-genome sequencing. *P. flexa*, a bacterium within the *Priestia* genus was identified by Gupta et al., 2020, adding a new genus classification to the *Bacillus* genus. *Priestia* genus exhibits various characteristics, including being aerobic bacteria, mainly Gram-positive (though some species, like *P. koreensis* and *P. flexa*, exhibit different Gram characters), rod-shaped, motile, and capable of forming endospores. The genus is found in diverse habitats such as soil, sea sediment, feces, rhizosphere of plant roots, and the upper atmosphere. It thrives in temperatures ranging from 5 °C to 48 °C, with optimal growth occurring between 28 °C and 37 °C [20].

In this study, genome mining was conducted to identify potential novel bioproducts, with a focus on

lycopene produced by the bacterium *P. flexa* JT4. Our findings lay the foundation for further research on novel bioproducts derived from *P. flexa* isolated from Indonesian deep-sea waters.

Methods

Sample collection

The study was conducted in the southeastern tropical Indian Ocean (SETIO) around the Java Trench area (102° 15.077' E and 7° 22.551' S) during the TRIUMPH Expedition on November 21, 2019. The sampling location is situated approximately 204 nautical miles southwest of the Sunda Strait, with a depth of 4,552 m (Fig. 1). Seawater samples were collected using an SBE-32 carousel water sampler equipped with an SBE 911+conductivity-temperature-depth (CTD) sensor to measure the vertical profile of seawater, including pressure, temperature, salinity, fluorescence, and dissolved oxygen concentration. Vertical profiles of physical parameters were presented as variables versus depth graphics, and water-mass characteristics were analyzed using the temperature–salinity diagram (TS diagram). Based on the TS diagram of the CTD4 Station (Fig. S1, Table S1), we identified five types of water masses, namely Indonesian Upper Water (IUW), Indian Central Water (ICW), Antarctic Intermediate Water (AAIW), Indian Deep Water (IDW), and AABW (Antarctic Bottom Water).

Nitrate and phosphate concentrations in deep-sea water samples were assessed using a method akin to the one previously outlined [21]. Pre-filtered water (350 mL stored in black plastic bottles) underwent filtration through a membrane (pore size: 0.45 µm; diameter: 47 mm) using a vacuum pump with a maximum pressure of 0.3 bar, followed by storage at −20 °C. The concentrations of nitrate (wavelength: 543 nm) and phosphate (wavelength: 885 nm) were determined utilizing a 1700 UV–Vis Spectrophotometer (Shimadzu, Japan). Microorganisms were gathered from the filtered seawater sample using Whatman cellulose nitrate membranes (pore size: 0.22 µm; diameter: 47 mm), and these membranes were subsequently stored in Falcon tubes at −20 °C.

Bacterial isolation

The collected membrane was cultured on half-strength Zobell 2216E agar medium and incubated at 4 °C for two weeks. Bacteria from the agar plates were isolated and purified utilizing a streaking method on half-strength Zobell 2216E agar medium. Subsequently, the pure cultures were prepared for a psychrophilic test. For long-term storage, bacterial cells were first grown in liquid Zobell 2216E medium at 20 °C and shaken at 200 rpm for 2—3 days. Then, the cells were harvested by centrifugation at 12,000 rpm for 2 min. The pellets were resuspended in liquid 1/5 Zobell 2216E medium containing 50% (v/v) glycerol, flash frozen in liquid nitrogen and stored at −80 °C.

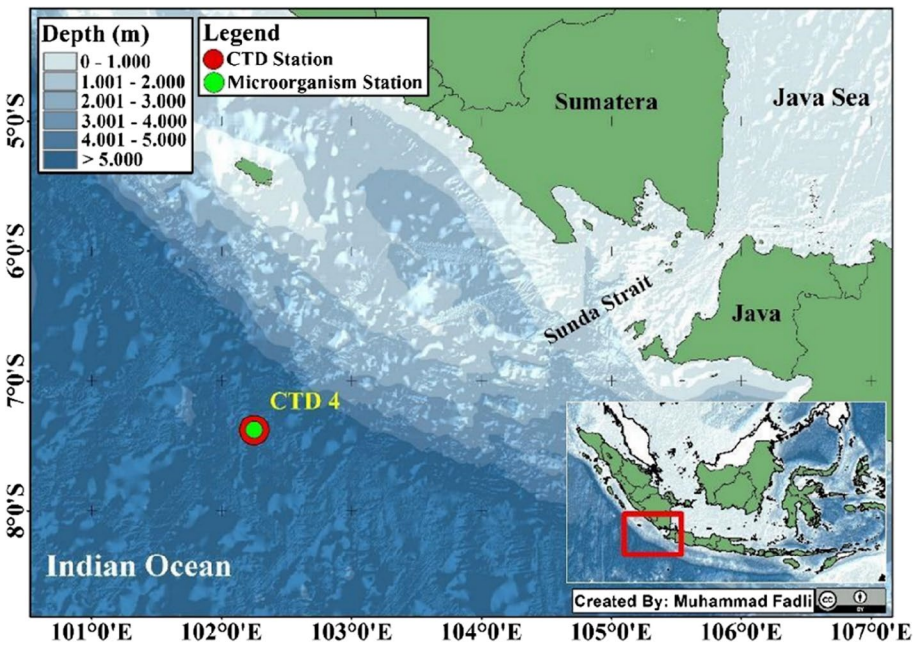

Fig. 1 The location from which seawater samples were collected at the southeastern tropical Indian Ocean (SETIO) around the Java Trench area (Maps Data: Indonesian Geospatial Information Agency)

Psychrophilic test

To ascertain whether the bacterial isolates exhibited psychrotrophic or psychrophilic characteristics, a psychrophilic test was conducted. Single colonies of the bacterium were cultivated at two distinct temperatures simultaneously, namely 4 °C and 20 °C, for a period of 14 days, after which the growth was carefully observed. Psychrotrophic bacteria were identified as strains capable of growth at both incubation temperatures, whereas psychrophiles were strains that demonstrated growth at 4 °C but were unable to grow at 20 °C [22, 23].

Gram staining

Gram staining was conducted to determine the Gram type of the bacterium. Colonies from the agar plates were sampled using an inoculating loop and transferred onto microscope slides. These colonies were spread with a small amount of water and heat-fixed over a flame. The colony area was then stained with crystal violet dye for a duration ranging from 10 to 60 s, followed by rinsing with distilled water. Iodine solution was applied for a period of 10 to 60 s, and after pouring off the excess iodine, the smear was washed using distilled water. Subsequently, the microscope slides were rinsed with 95% alcohol for destaining and washed with distilled water for 5 s. The smear was counterstained with safranin solution for 40 to 60 s. Following this, the microscope slides were rinsed with distilled water and air-dried. The dried slides were observed under a microscope with 1,000 times magnification [24].

Bacterial genome isolation and identification

The pure bacterial culture was inoculated in liquid Zobell 2216E medium at 20 °C and shaken at 200 rpm for 2—3 days. Bacterial cells were then harvested from the liquid Zobell 2216E medium through centrifugation at 12,000 rpm for 2 min. Subsequently, total genomic DNA was extracted from the collected biomass using 0.22-mm filters of the Genomic ZymoBIOMICS DNA Miniprep Kit (Zymo Research, US), following the manufacturer's instructions, and visualized using a 1% agarose gel. The purity and concentration of genomic DNA were assessed using a NanoDrop UV–Vis spectrophotometer, and the genomic DNA was subsequently stored at −20 °C.

The bacterial genomic DNA was amplified using the 16S rRNA primer pair 27F-YM (5'-AGAGTT TGATYMTGGCTCAG-3') [25] and 1492R (5'-GGT TACCTTGTTACGACTT-3') [26], which yielded PCR products of approximately 1.5 kb. The total reaction mixture was 50 µl, consisting of 100 ng of bacterial DNA in 25 µl Green Master Mix PCR Promega, and 0.5 µM of each primer. Reaction mixtures were incubated for 5 min at 95 °C, followed by cycles of denaturation for 30 s at 95 °C, annealing for 30 s at 55 °C, and extension for 2 min at 72 °C for a total of 35 cycles. Each PCR product was sequenced at Macrogen, Inc. (Singapore) using the primer 27F-YM and 1492R.

Genome characterization

The genomic DNA was sequenced using GridION (Oxford Nanopore, UK) at Genetika Science (Indonesia) and subsequently analyzed with MinKNOW software ver. 20.06.9. Base calling was executed using the high accuracy mode of Guppy ver. 4.0.11. Raw reads underwent filtration using Filtlong ver. 0.2.1, with a filtering criterion of the best 800,000,000 bp, and read quality was visualized using NanoPlot [27].

High-quality reads were assembled using the de novo approach with Flye software ver. 2.8.1 to generate contigs [28]. The contigs were further refined using Medaka ver. 1.2.0 (https://github.com/nanoporetech/medaka) and aligned to the reference genome utilizing Mauve ver. 2.4.0 [29]. Genomic attributes, including coding sequences (CDS), transfer RNA (tRNA), and ribosomal RNA (rRNA), were annotated using DFAST ver. 1.2.0 [30]. Annotation assessment was validated using both Busco ver. 5.3.0 [31] and CheckM ver. 1.2.2 [32].

Functional annotations were assigned to coding sequences (CDS) based on the KEGG pathway (KO) and the cluster of orthologous genes (COG) databases. KO annotation was conducted using TRAPID ver. 2.0 [33], while COG annotation was performed using eggnog ver. 5.0 [34].

Identification and taxonomic analysis

GenBank files resulting from whole-genome sequencing were parsed using a Python script (https://github.com/raysteven/gbk_parser) that utilized the package Biopython ver. 1.79 to convert GenBank file format into Table data structure to ease downstream analysis such as the estimation of genomic characteristics, pangenome analysis, and genomic island analysis. Identification of sequenced bacteria was conducted using the MinHash algorithm provided by PATRIC 3.16.12 (https://patricbrc.org/app/GenomeDistance) [35]. Bacterial genera that hit with p-value = 0 are *Priestia* and *Bacillus*, thus both genera were searched in the NCBI Assembly Database. We only picked type strain genomes with assembly level "scaffold" and "complete chromosome". The resulting genomes were curated manually to obtain the final data set for estimating the phylogenetic proximity of the sequenced bacteria.

Overall genome relatedness indices (OGRI), such as average nucleotide identity (ANI), were calculated using FastANI ver. 0.1.3 [36], and digital DNA–DNA hybridization (dDDH) were estimated using Genome-to-Genome

Distance Calculator (GGDC) ver. 3.0 [37]. Thereafter, a phylogenomic tree was constructed using the constructed data set and OrthoFinder ver. 2.5.4 [38] and visualized using iTOL ver. 6.5.2 [39].

The final dataset containing genomes for phylogenomic analysis, pangenomic analysis, and genomic island analysis are available in Table S2.

Estimation of genomic characteristics

General characteristics of the genome were determined using a Python script (https://github.com/raysteven/gbk_parser) with the sequenced genome. We determined genome size (bp), the length of the coding region (bp), G + C content (%), and the number of genes (tRNA, rRNA, tmRNA, and coding sequences). The parsed results were visualized as a circular genome map using the Proksee web server with a CGView-based draw engine [40].

Pangenome Analysis and prediction of Genomic islands and secondary metabolites

The genomes involved in this analysis were the *Priestia* bacteria from the Supplementary Table S2. The software GET_HOMOLOGUES [41] was used for pan genome analysis to determine characteristics such as genome size, estimated core-genome size, and occupancy statistics. The pangenome type was evaluated by calculating the α value from the pangenome size-estimation curve using the equation derived from a previous study [42]. Genomic islands and BGCs of secondary metabolites were predicted using the web servers IslandViewer ver. 4 [43] and antiSMASH ver. 6.0.1 [44], respectively. The reaction pathway of each gene based on the KO annotation was visualized using the KEGG Mapper [45].

In silico nonribosomal peptides structure similarity search and its antibacterial activity prediction

The putative non-ribosomal peptides (NRP) chemical structures were obtained from the antiSMASH result in SMILES (simplified molecular-input line-entry system) format. Then NRP's SMILES were used as a search query in the NRP database using NORINE (https://bioinfo.lifl.fr/norine/) to find the putative compound similarity against currently known NRP structures in the database [46]. The NRP's SMILES were also used to predict the antibacterial effect in silico with QSAR (quantitative-structure activity relationship) methods using the anti-Bac-Pred web app (http://www.way2drug.com/antibac/) from Way2Drug [47].

Bgc comparison

Analysis was also carried out using lycopene-producing gene clusters taken from several databases. The nine bacterial genomes are from the NCBI database with accession numbers, namely *Bacillus subtilis* PS832 (GenBank: CP010053.1), *Staphylococcus aureus* NCTC 8325 (GenBank: CP000253.1), *Priestia megaterium* MARUCO02 (GenBank: CP107543.1), *Priestia filamentosa* H146 (GenBank: CP136435.1), *Priestia aryabhattai* UASWS1812 (GenBank: NZ_JAKKWP010000001.1), *Priestia abyssalis* DSM 25875 (GenBank: NZ_KV917369.1), *Pantoea ananatis* PA13 (GenBank: CP003085.1), *Pantoea agglomerans* FDAARGOS 1447 (GenBank: CP077366.1, and *Brevundimonas* sp. (GenBank: CP059260.1). The lycopene gene cluster nucleotide sequences were predicted for structure and arrangement to determine the lycopene biosynthetic gene cluster in bacterial isolates, using the antiSMASH web server. Visualization of the lycopene biosynthetic gene cluster from Priestia flexa bacterial samples and other lycopene-producing bacteria was carried out using the Illustrator for Biological Sequences (IBS) software version 1.0 [48].

Lycopene production and identification

In order to confirm the existence of lycopene in *P. flexa* JT4 colonies, a testing procedure was carried out that involved inducing culture to generate lycopene, followed by a thorough characterization process. Culture activation was carried out by inoculating single colony *P. flexa* JT4 into a new Zobell 2216E medium with 15% salinity as stress [49]. The activation process was carried out until the culture became cloudy. After activation, 10% (v/v) of the culture was taken and inoculated into a new medium in Erlenmeyer 250 mL with a working volume of 150 mL. *Priestia flexa* was cultivated for 132 h using Zobell 2216E broth medium at 15% salinity with an initial pH of 7.6 ± 0.2, a temperature of 37 °C and agitation of 132 rpm. Sampling was then carried out every 12 h with sampled volume of 11.5 mL. The volume intended for pH checking (2 mL), sugar concentration measurement (1.5 mL), and dry weight of cells and lycopene product (8 mL). Lycopene product analysis is carried out using UPLC-MS/MS. The cell pellet was taken, dried at 40 °C and extracted using methanol, while the supernatant was extracted using ethyl acetate three times. The supernatant was extracted with an equal volume of ethyl acetate. The ethyl acetate layer was collected. The methanol and ethyl acetate fractions were evaporated under reduced pressure using a rotary evaporator [10, 50].

Liquid chromatography-mass spectrometry

The authentic standard of lycopene (Markherb, Bandung, Indonesia) was freshly prepared in chloroform and sonicated for 2 min. The stock solution (1 mg/ml) was then filtered through a 0.22-μm PTFE filter. Analysis was performed on an Acquity UPLC H-Class system equipped

with binary solvent manager, sample manager, and column heater (Waters, Milford, Michigan, USA) coupled triple quadrupole. The Masslynx analysis software was used to operate the instrument and execute the data analyses.

The separation was performed using ACQUITY UPLC BEH Shield RP18 100×2.1 mm, 1.7-µm column. The mobile phase consisted of solvent A: Water + 0.1% formic acid and solvent B: acetonitrile + 0.1% formic acid. The elution was performed in gradient mode using 10% B (0–0.5 min); 10–20% B (0.5–3.0 min); 20–30.3% B (3.0–4.0 min); 30.3–33.3% B (4.0–5.0 min); 33.3–50% B (5.0–5.5 min); 50–70% B (5.5–6.5 min); 70–80% B (6.5–7.0 min); 80–100% B (7.0–8.5 min); and hold 100% B (8.5–20 min) with a flow rate 0.3 mL/min. The column was operated at 40 °C and the sample/standard injection volume was 3 µL.

The detection was performed by multiple reaction monitoring (MRM). The ESI was operated in positive mode (ESI+) with the parameters in the source as follows: spray voltage, source gas flow at 900 L/h, and source temperature at 450 °C. The MRM mode was set up as follows: collision energy at 30 V for the quantifier and 11 V for the qualifier, capillary voltage at 2.5 kV, and cone voltage at 40 V. Under these conditions, the product ions (m/z) of all-*E*-lycopene used as the authentic standard were observed as follows: $[M]^+$ 536, $[M-69]^+$ 467, $[M-92]^+$ 444, $[M-161]^+$ 375, $[M-267]^+$ 269, $[M-353]^+$ 183 and $[M-467]^+$ 69 (Figure S2). The m/z 536 (parent ion), 444 (qualifier) and 69 (quantifier) were chosen for the MRM analysis of lycopene.

Results and discussion

Environmental condition and bacterial isolation

P. flexa JT4 was isolated from 1000 m deep at the SETIO around the Java Trench area. The depth of the third layer water mass (where *P. flexa* JT4 was isolated) ranging from 950 to 1300 m exhibited a sigma-theta (σ_Θ) value of 27.34–27.53 kg/m^3 and was filled with AAIW. The environmental characteristics where P. flexa JT4 was isolated can be seen on Table 1.

The effective separation of this bacterium from its native surroundings corresponds with its ecological traits, which are well within the documented parameters for its growth [20]. The water-mass type of the surrounding environment was AAIW, which was characterized by reduced salinity, low temperature, and reduced dissolved oxygen concentration. Because the native environment of the bacterium exhibited low temperatures, we performed a psychrotolerant/psychrophilic test on the bacteria isolated from the water samples collected from the sampling location.

Table 1 Descriptive information on the environmental characteristics of Java Trench (sampling location)

Parameter	Java Trench
Temperature	5.60 °C
Salinity	34.72 PSU
pH	8.0
Dissolved Oxygen concentration	2.20 mg/L
NO_3^- concentration	2.122 µmol/L
PO_4^{3-} concentration	4.067 µmol/L

Psychrophilic test and gram staining

The psychrophilic test showed that *P. flexa* JT4 could grow at 4 °C and 20 °C; therefore, the bacterial isolate met the definition of psychrotolerant [22]. Gram staining results showed that *P. flexa* JT4 was stained as a gram-negative bacterium (Fig. 2), but based on previous study, *Priestia flexa* was recorded as a gram-variable bacteria [20]. *Priestia flexa* have rod shaped cells and 0.9 µm in length, with opaque, cream, raised margin, and smooth colony characteristics.

Identification and taxonomic analysis

The 16S rRNA sequence of the bacterium (Supplementary Information 1) has been obtained from sequencing result from Macrogen, Inc., (Singapore). Then, BLASTn (NCBI) search with default parameter was conducted to determine the species of the bacterium. The first five hit were *Priestia flexa* with > 99% percent identity, 100% query cover, E-value = 0, with the highest total score is 33607 against *P. flexa* KLBMP 4941 (CP016790.1). We concluded that *Priestia flexa* is the species of our bacterial isolate, thus we named this bacterium strain as *Priestia flexa* JT4. The OGRI for species identification (Fig. 3) revealed that *P. flexa* JT4 exhibited an ANI value of 99.61 and a dDDH value of 96.8 compared to *P. flexa* DMP08 (ANI value: 98.82; dDDH value: 91.6) and *P. flexa* QDU. The ANI and dDDH threshold for the query sequence to belong to the same species as the reference sequence are 95 and 70 [46, 51], respectively. The ANI and dDDH values obtained for *P. flexa* exceeded the threshold values for the species, indicating that *P. flexa* JT4 belonged to the same species as *P. flexa* strain DMP08 and QDU.

After estimating the OGRI values, we performed phylogenomic tree construction for further identification. Figure 4 illustrates the phylogenomic relationship of *P. flexa* JT4 with the members of the genus *Priestia*. The phylogenomic tree was constructed using 22 *Priestia* and *Bacillus* genomes from 15 species, with *Sphingobium lactosutens* as the outgroup. Analysis using OrthoFinder revealed 318 single-copy gene orthogroups belonging to the entire query dataset, which were used to construct

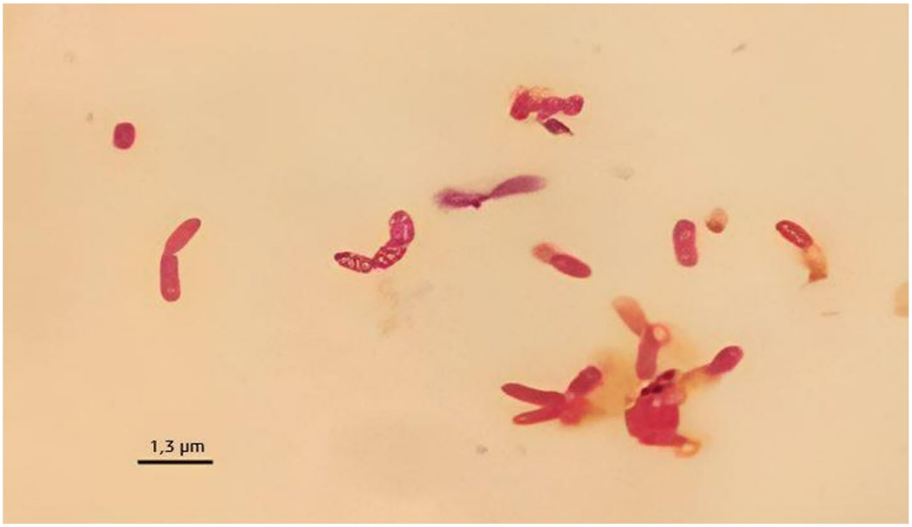

Fig. 2 Gram staining result of *Priestia flexa* JT4 indicating that the bacterium was gram-negative, because of the bright purple coloring of the bacterial cells

a phylogenomic tree using the STAG algorithm. We observed that *P. flexa* JT4 formed a clade with *P. flexa* DMP08. These results were confirmed by the ANI and dDDH results, indicating that the bacterial strain isolated from the Java Trench belonged to the species *P. flexa*.

Genomic characteristics

We obtained 800,000,258 bp high-quality reads with an estimated coverage of 198.7x, from genome sequencing. Then, the genome of *P. flexa* JT4 was successfully assembled, and its characteristics are presented in Table 2. The genome "completeness" and "contamination" values, referring to the marker for the Bacillales lineage, were 98.09% and 0.02%, respectively. Additionally, *P. flexa* JT4 genome contained 19 of the 24 COG categories; the following five COG categories were missing from the genome of *P. flexa* JT4: Y (nuclear structure), Z (cytoskeleton), A (RNA processing and modification), W (extracellular structure), and R (prediction of general function only. The data obtained from KEGG Mapper showed that the *P. flexa* JT4 genome contained genes involved in 271 pathways, with 74 completely annotated KEGG modules. Furthermore, we identified 5,185 CDS in the genome of *P. flexa* JT4, out of which 85.79% (4,448/5,185) were COG annotated (Fig. 5) and 63.63% (3,299/5,185) were KO annotated. We also successfully annotated 61.97% (3,213/5,185) of the known CDS using both COG and KO databases. However, the remaining 38.03% of the known CDS could not be annotated using both COG and KO databases (Table 2). This indicated that a few genes in the *P. flexa* JT4 genome were not well-characterized and could serve as candidates for further study.

Pan genome analysis

As shown in Fig. 6, the core genome-size estimation curve and the genome size of the genus *Priestia* revealed an open pan genome curve, as indicated in previous studies [52, 53]. The estimated pan genome size of the 26 *Priestia* genomes was 14,671 genes, with 1,351 genes belonging to the core genome, 2,234 genes to the soft-core genome, 3,490 genes to the shell genome, and 8,947 genes belonging to the cloud genome. The core genome accounted for only 9.20% (1,351/14,671) of the entire pangenome.

Among the other *Priestia* species, *P. flexa* JT4 has cloud genome (genes shared with ≤ 2 genomes combinatorically, e.g., *Priestia* sp. 1 with *Priestia* sp. 2 (*Priestia* species 1 with *Priestia* species 2), *Priestia* sp. 1 with *Priestia* sp. 3, *Priestia* sp. 2 with *Priestia* sp. 3, and so on) with the highest cloud gene number (2271 genes), almost two times the second highest cloud gene number from *P. filamentosa* Hbe603 (1177 genes). A high cloud gene number means that *P. flexa* JT4 shares a lot of specific genes with the other *Priestia* species in the pangenome data set and indicates that the Java Trench *P. flexa* has a high repertoire of gene diversity. Aside from high-shared gene numbers, cloud genes may also contain unique genes that exclusively belong to that species. Surprisingly, the cloud genome of *P. flexa* JT4 contains 95.64% (2172/2271) unique genes. It means only 99 genes that the *P. flexa* JT4 shared with one other *Priestia* species in the pangenome data set.

Strain	No	1	2	3	4	5	6	7	8	9	10	11	12	13	14
P.flexa_JavaSea	1		21.00	21.20	21.00	18.50	96.80	91.60	22.30	21.00	20.10	21.30	21.40	21.00	21.00
P.aryabhattai_K13	2	79.92		64.70	94.80	19.50	21.10	21.10	22.90	65.30	64.00	64.20	64.30	94.40	92.10
P.aryabhattai_LAD	3	80.16	95.28		64.80	20.50	21.20	21.40	22.50	74.00	68.00	70.90	72.10	64.50	64.60
P.aryabhattai_NM1A2	4	79.90	99.26	95.30		19.50	21.10	21.00	22.70	65.40	64.10	64.50	64.70	94.40	92.60
P.filamentosa_Hbe603	5	77.07	77.17	77.21	77.10		18.50	18.50	17.80	19.50	20.30	21.90	20.50	19.40	19.40
P.flexa_DMP08	6	99.61	80.13	80.37	80.08	77.16		93.30	22.30	21.00	20.00	21.30	21.10	21.00	21.10
P.flexa_QDU	7	98.82	80.25	80.48	80.23	77.12	98.84		22.20	21.00	20.20	21.50	21.50	21.30	21.00
P.koreensis_FS1	8	78.85	79.30	78.90	79.49	77.26	78.75	78.92		22.50	20.40	23.00	22.90	22.80	22.70
P.megaterium_5-3	9	80.20	95.56	96.85	95.57	77.10	80.19	80.26	78.67		72.60	96.70	97.00	65.00	65.30
P.megaterium_BHS1	10	79.58	95.31	95.70	95.33	77.42	79.59	79.62	78.12	95.56		70.20	70.30	63.90	63.80
P.megaterium_H2	11	80.02	95.19	96.14	95.26	77.44	80.10	80.20	78.56	99.59	96.13		92.70	64.00	64.20
P.megaterium_Mes11	12	80.39	95.28	96.40	95.32	77.53	80.22	80.45	78.93	99.60	96.11	98.81		63.90	64.50
P.megaterium_YC4R4	13	79.84	99.31	95.30	99.26	77.11	79.98	80.13	79.41	95.57	95.31	95.23	95.29		91.70
P.sp_J2	14	80.04	99.00	95.23	99.04	77.17	80.05	80.11	79.22	95.51	95.32	95.23	95.36	98.98	

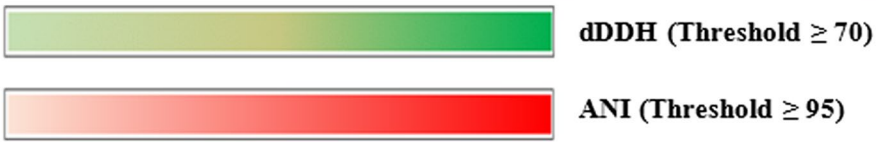

Fig. 3 Calculation results of ANI and dDDH of 14 *Priestia* genomes for sample identification

Prediction of Genomic islands

Genomic islands (GIs) can affect the evolution and growth conditions of bacteria. To analyze Genomic islands (GIs) from *P. flexa* JT4 genome, three different algorithms from IslandViewer 4.0 were used, including IslandPick, IslandPath-DIMOB, and SIGI-HMM. The IslandPick algorithm uses a comparative genomics approach to detect GIs on the genome. Meanwhile, IslandPath-DIMOB uses nucleotide bias and presence of mobility gene and SIGI-HMM uses codon usage bias with Hidden Markov Model approach [43].

A total of 21 genomic islands (413 genes) were identified in the genome of *P. flexa* JT4 (Fig. 7). The length of GIs varied from 3,528 bp to 47,628 bp, with the number of genes in a genomic island ranging from 4 to 45. Few of the GIs were shorter (3.5–9.1 kb) compared to the length of most GIs (10–200 kb) [46].

Prediction of secondary metabolite BGCS

The results of antiSMASH revealed that *P. flexa* genome exhibited eight gene clusters, namely cyclic-lactone-autoinducer/RiPP_like (1), nonribosomal peptide synthetase (NRPS) (2), siderophore (1), lasso peptide (1), terpene (2), and type three polyketide synthase (T3PKS) (1) (Table 3).

The NRPS/NRPS-like, lasso peptide, terpene, and NRPS gene clusters have been predicted by antiSMASH

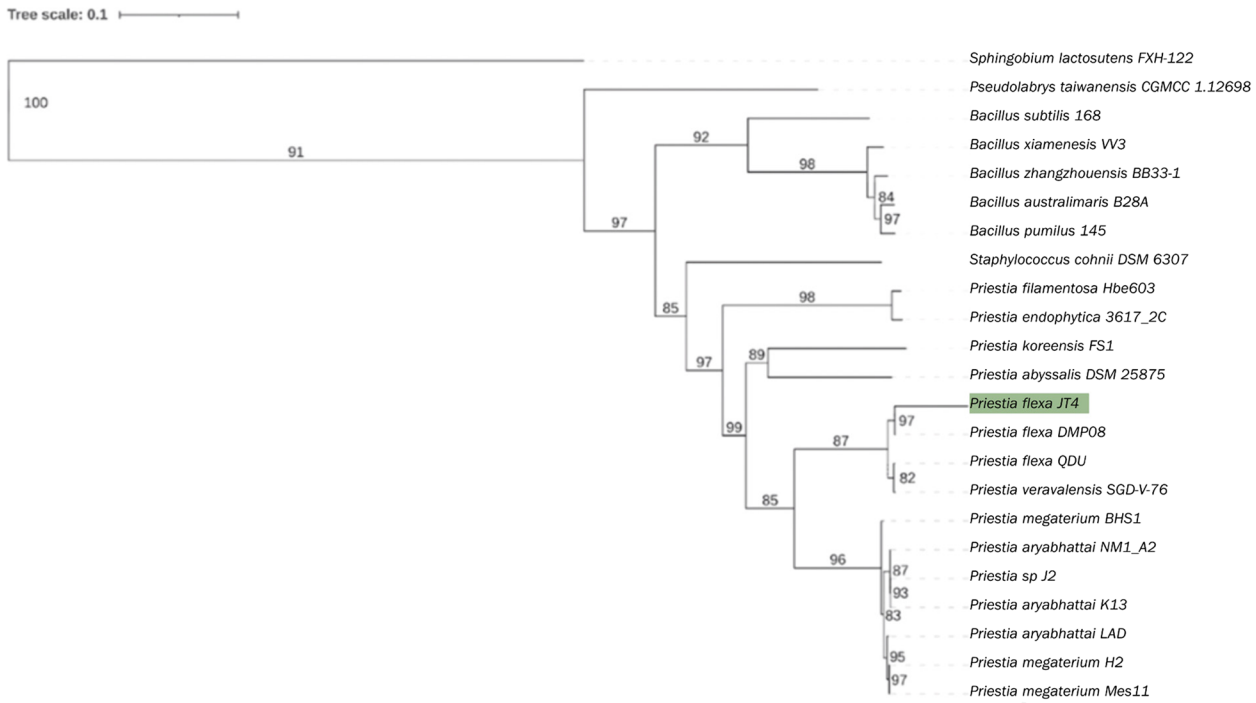

Fig. 4 Phylogenomic analysis of the *P. flexa* JT4 genome and the members of genera *Priestia* and *Bacillus*. The phylogenomic tree was constructed using 22 *Priestia* and *Bacillus* genomes from 15 species, with *Sphingobium lactosutens* as the outgroup, with OrthoFinder software and visualized using iTOL Software

Table 2 General characteristics of the *Priestia flexa* genome from Whole Genome Sequencing (WGS) Result

Attribute	Value	Percentage[a]
Contigs	1	-
Genome size (bp)	4,026,611	100
DNA coding Region (bp)	3,345,237	83.08
DNA G + C content (bp)	1,526,086	37.9
Total genes	5,315	100
rRNA	35	0.66
tRNA	94	1.77
tmRNA	1	0.02
CDS	5,185	100
CDS assigned to COGs	4,448	85.79
CDS assigned to KO	3,299	63.63
CDS assigned to both COGs and KO	3,213	61.97

CDS Coding sequence, *COGs* Clusters of orthologous groups, *KO* KEGG orthology

[a] Percentage is based on genome size (bp) or the total number of protein-coding genes in the annotated genome

to produce bacitracin, paeninodin, carotenoids, and bacillibactin, respectively, with low similarity scores. However, the remaining clusters did not return any hits with KCBLAST. Thus, further studies are needed to characterize these gene clusters.

The NRPSs module and its predicted product are shown in Figs. 8 and 9. The NRP1 cluster includes 55 genes, but only nine genes have predicted functions in the synthesis of NRP products. This cluster exhibits 22% similarity with the bacitracin producing NRP cluster. NORINE similarity search indicated that NRP1 resembled cepaciachelin, with a similarity value of 0.728, whereas Way2Drug antimicrobial activity of NRP1 revealed that NRP1 exhibited potential antimicrobial activity against *Bacillus subtilis*, with a confidence level of 0.4373.

The NRPS 2 cluster contained 51 genes, but only eight genes exhibited predicted functions in the synthesis of NRP products. The NRPS2 cluster has been reported to exhibit 46% similarity with the bacillibactin-producing NRPS cluster. A search performed on the NORINE database using SMILES structure as a query revealed that NRP2 resembled N-coronafacoyl-L-threonine, with a similarity value of 0.918. Furthermore, the antimicrobial activity of NRP2 predicted using the Way2Drug web server indicated that NRP2 exhibited antimicrobial activity against *Prevotella oralis*, with a confidence level of 0.5357.

The second terpene gene cluster (Fig. 10) consisted of a gene that synthesized a product like carotenoids (50% similarity), or lycopene. Lycopene was synthesized from the primary biosynthetic gene *crtM* at LOCUS_51590,

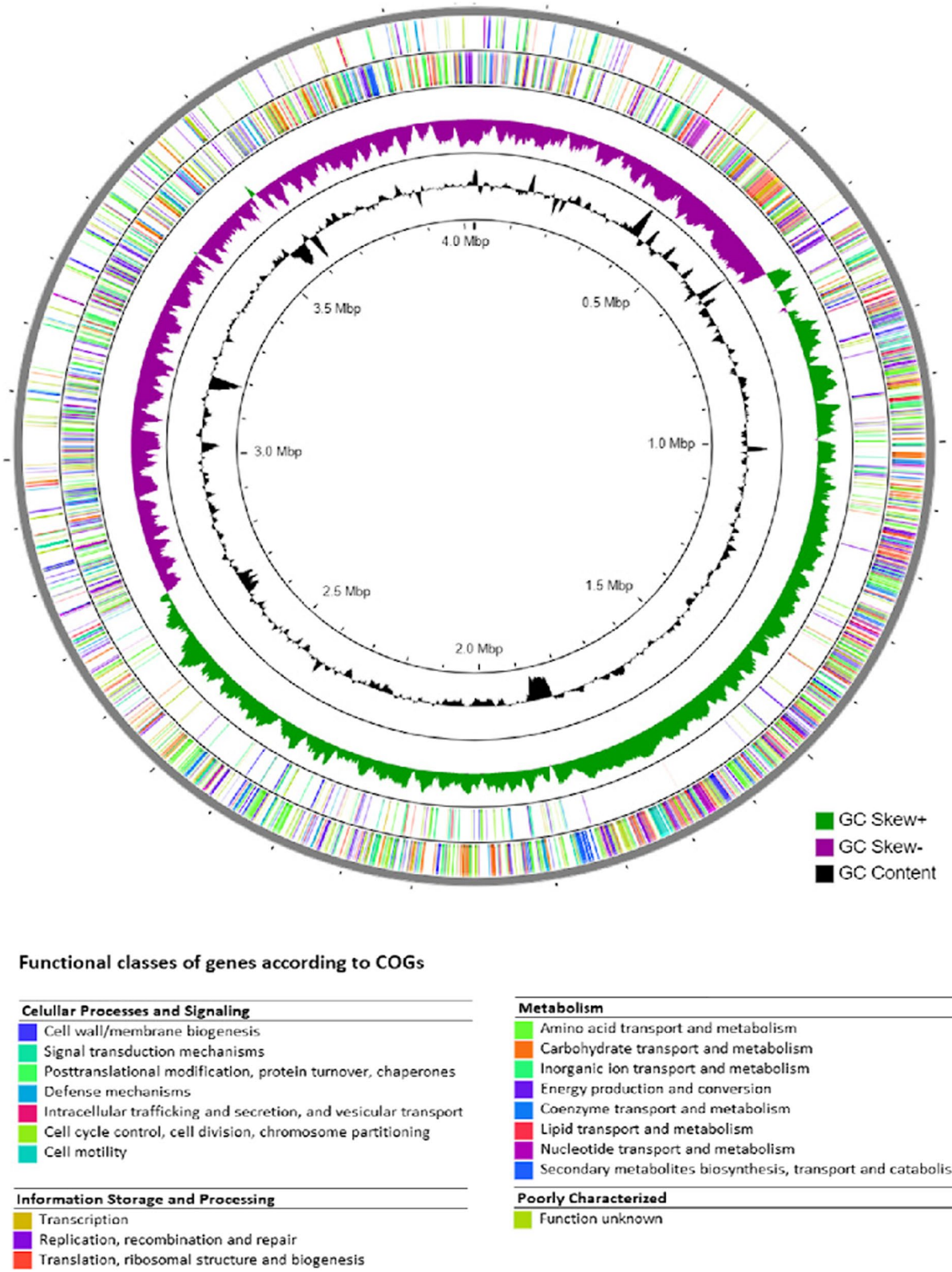

Fig. 5 Circular analysis and visualization of *Priestia flexa* genome. From the outside to the inside, the first and second circle represent genes with COG annotation. Circle 3 (green and purple) and 4 (black) show GC skew and GC content as the deviation from the average for the complete genome

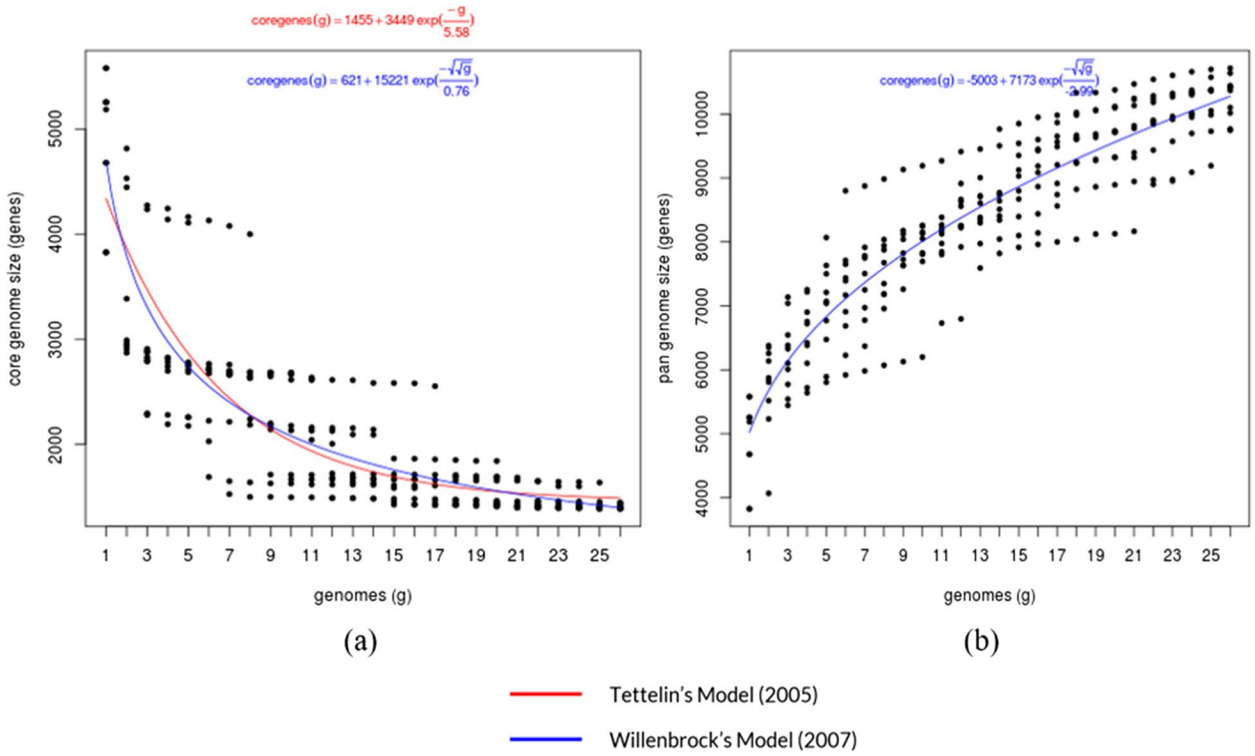

Fig. 6 Pan- and core- genome estimation curves. **a** The size of the pan-genome increases for every included genome indicating an open pan-genome. **b** Core-genome size decreases with more genomes included in analysis

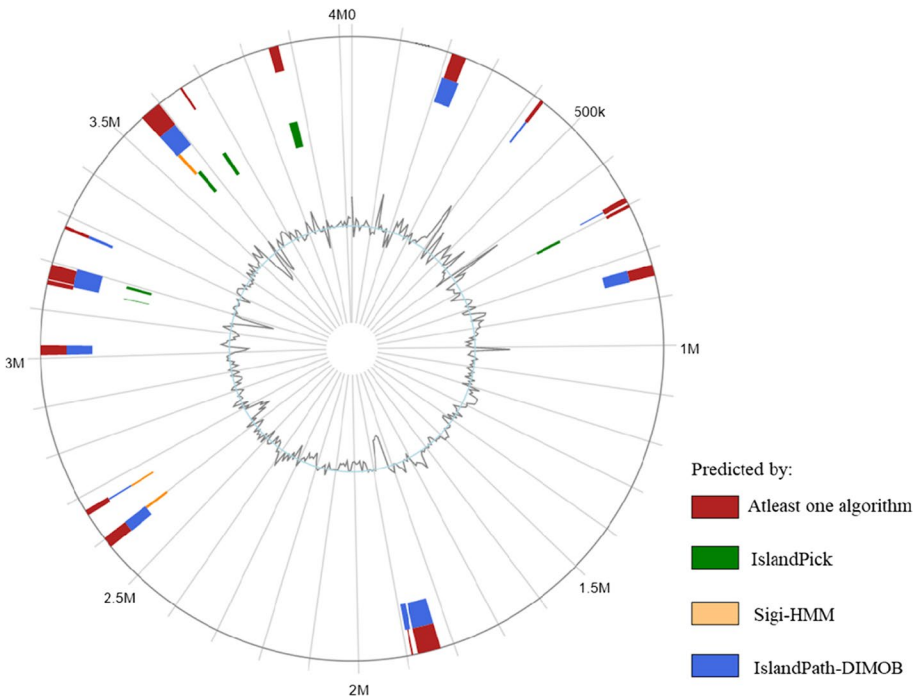

Fig. 7 Visualization of Genomic Island detected in the *Priestia flexa* JT4 North Java Sea Genome using IslandViewer 4.0

Table 3 Secondary metabolite biosynthetic BGCs identified in the genome of *P. flexa* from AntiSMASH Analysis Result

Cluster	Start	Stop	Length (bp)	KCBLAST[a] result	Similarity (%)	Number of genes
RiPP-like	183,212	203,959	20,747	-	-	35
NRPS	1,812,954	1,875,648	62,694	Bacitracin	22	55
Siderophore	1,925,003	1,941,570	16,567	-	-	18
Lasso peptide	2,920,293	2,944,228	23,935	Paeninodin	80	31
Terpene	2,950,161	2,970,658	20,497	-	-	27
T3PKS	3,242,453	3,283,508	41,055	-	-	47
NRPS	3,838,023	3,882,063	44,040	Bacillibactin	60	51
Terpene	3,991,378	4,012,214	20,836	Carotenoid	50	29

[a] KnownClusterBLAST or "Most Similar Known Clusters" identifies clusters from the MiBIG database that are similar to the current region [44]

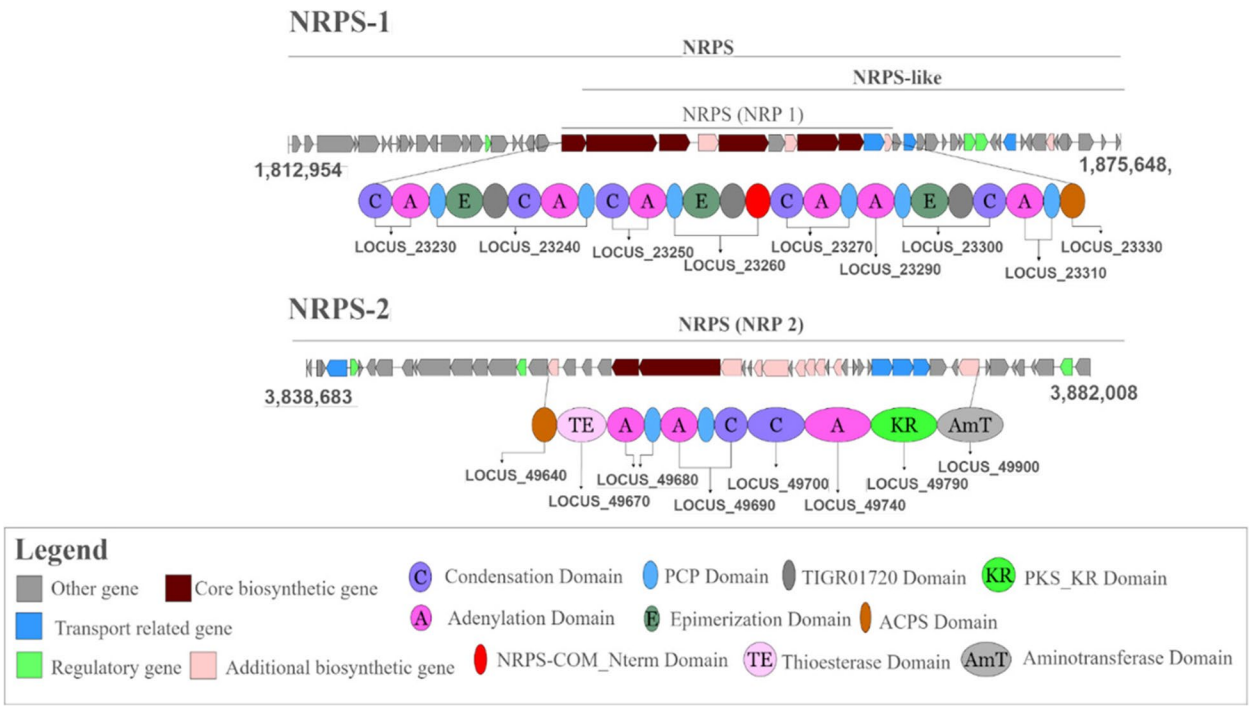

Fig. 8 Predicted BGC structures of NRP1 and NRP2 identified in *P. flexa* JT4

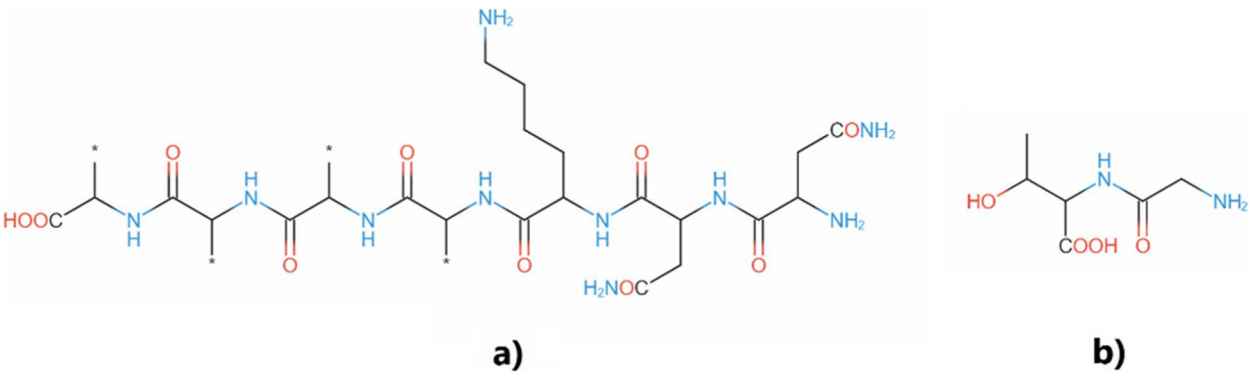

Fig. 9 Predicted chemical structures of (**a**) NRP1 and (**b**) NRP2 of the NRP-like and NRP BGCs identified in *P. flexa* JT4

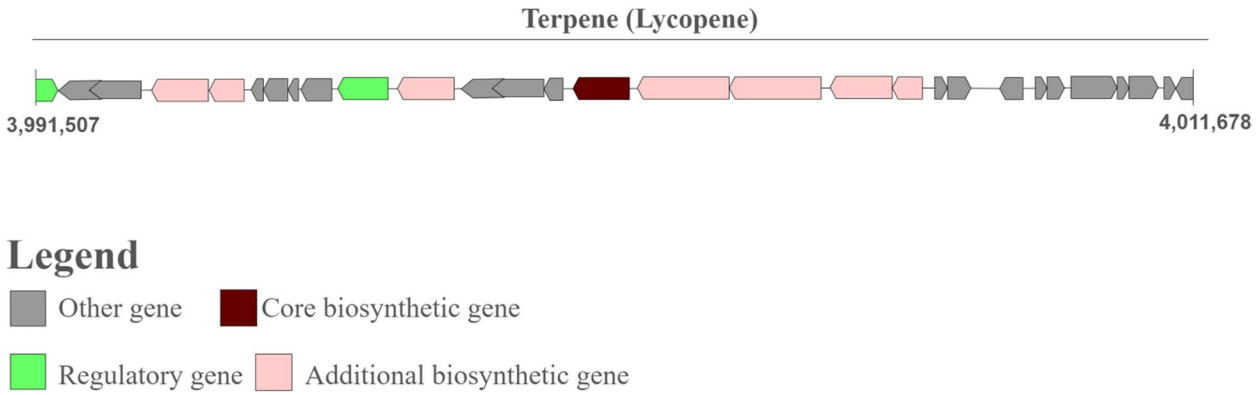

Fig. 10 Predicted BGC structure of lycopene (terpene) identified in *P. flexa* JT4

which exhibited two KO annotations, namely KO2291 and K10208. K02291 indicates 15-cis-phytoene synthase, which results from the expression of *crtB*, whereas K10208 indicates *crtM,* which produces the enzyme 4,4-diapo phytoene synthase. In the structure of the gene cluster, other *crt* genes, such as *crtN* and *crtI*, were included as additional biosynthetic genes. In this study, we further analyzed lycopene to its production capability in *Priestia flexa* using UPLC-MS/MS method. Lycopene is chosen mainly due to its potential for monetization, as it holds significant market value and can be generated by bacteria. This presents an opportunity for industry to tap into lycopene-producing bacteria and utilize them for commercial purposes.

Comparative lycopene BGC analysis

To identify the difference in lycopene BGC between *P. flexa* and other lycopene-producing bacteria, we conducted structural comparison of lycopene BGC among these bacteria. Figure 11 revealed that the lycopene BGC, annotated by antiSMASH in *P. flexa*, *P. megaterium*, *P. abyssalis*, and *P. aryabhattai*, contains three important genes related to producing lycopene: *crtI*, *crtN*, and *crtM*. *B. subtilis* only has the annotations for *crtN* and *crtM*. On the other hand, in the non-Bacillaceae family group, namely *Brevundimonas* sp., *Pantoea agglomerans*, and *Pantoea ananatis*, a more complex BGC was shown. All three species have genes in the carotenoid biosynthesis pathway, such as *crtE*, *idi*, *crtZ*, *crtW*, *crtY*, *crtB*, and *crtX* [54]. Based on Fig. 11, it is known that several bacteria have lycopene biosynthesis gene clusters. However, bacteria can specifically produce different types of lycopene, with C30, C40, or both backbones. In compiling the C30 carotenoid compound (lycopene), the *crtM* and crtN genes are present. In another pathway, in C40, there are the *crtE*, *crtB*, and *crtI* genes [54, 55]. Furthermore, in *Brevundimonas* sp., *Pantoea ananatis*, and *Pantoea agglomerans*, there are genes related to post-modification of lycopene biosynthesis. Based on the validation of the analysis with KEGG, it can be traced that the lycopene biosynthesis pathway in *Priestia flexa* has the main compound product being 4,4-diapolycopene.

Identification of lycopene compound in P. flexa JT4

The presence of lycopene in the *P. flexa* JT4 cells was confirmed using UPLC-MS/MS operated in multiple reaction modes as shown in the chromatograms (Fig. 12). The multiple reaction monitoring is a highly sensitive method for the selective detection and quantification of the analyte using the targeted mass spectrometry. This method has been used to identify and quantify a number of lycopene isomers [56–59]. The suspected peak of lycopene from the sample in Fig. 11a has a slight difference in the retention time with the authentic standard (Fig. 11c) which might be affected by the matrix effect or the concentration. Although we were not able provide the MS or MS/MS spectra of the suspected peak of the lycopene from the sample, we are very confident that lycopene is present in the *P. flexa* JT4. This proved that the presence of biosynthetic genes of lycopene in *P. flexa* JT4 and the culture condition enabled the cells to produce lycopene. Lycopene is synthesized by the *crt* genes family which is found in the *P. flexa* JT4 genome. Based on the KEGG pathway (Fig. 13), two main *crt* genes functioning in synthesizing lycopene are *crtI* and *crtB*. These two genes were found in the results of gene cluster analysis using antiSMASH. In the lycopene synthesis pathway, geranyldiphosphate (GPP) acts as a precursor for the pathway. GPP resulted from the conversion of IPP (isopentenyl pyrophosphate) and DMAPP (dimethylallyl pyrophosphate) by the enzyme farnesyl pyrophosphate. IPP and DMAPP were the products from mevalonate pathway (MVA) [51]. The *crtB* gene produces a phytoene synthase enzyme, which converts 2 GPP molecules into phytoene.

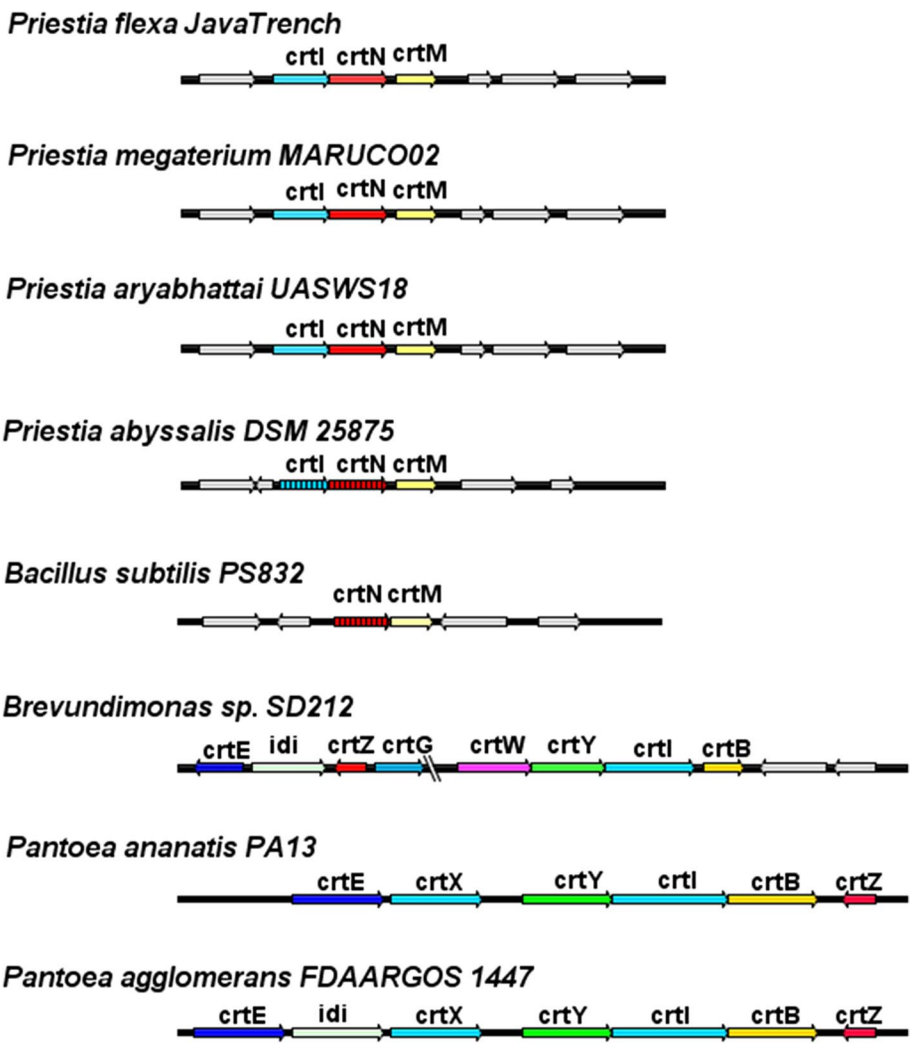

Fig. 11 BGC structure comparison of several lycopene-producing bacteria

The Enzyme lycopene synthase (CrtI) will consecutively carry out 4 desaturation reactions on phytoene to transform it into lycopene [60].

Lycopene is classified as a carotenoid compound which has a tetraterpene structure with 8 isoprene units and 11 double linear bonds (conjugations). Lycopene has various biological activities with the main activity as an antioxidant. Lycopene compounds also have anticancer, anti-inflammatory, antihypertensive, anti-aggregative, and other biological properties. They also operate as a starting point for the synthesis of additional carotenoid compounds. Lycopene has a potential effect to treat cardiovascular and neurological illnesses, due to its antioxidant action, which protects cells from reactive oxygen species (ROS) [18, 60, 61].

In comparison to another producing lycopene bacteria, *Blakeslea trispora* bacterium was used in the industrial world to naturally produce lycopene compounds. Fermentation accelerators such as imidazole, nicotinic acid, and 4-methylmorpholine are used to inhibit the transformation of lycopene to beta-carotene, which is catalyzed by the enzyme lycopene cyclase in the lycopene synthesis pathway in *B. trispora* bacteria. *B. trispora* yielded 4.78 103 mg/g DCW when imidazole was used, which was 73.5 times higher than the control output of 6.5×10–5 mg/g DCW [62]. After 48 h of culture, the addition of 42 mg/L geraniol increased lycopene production in *B. trispora*, reaching 578 mg/L compared to 317 mg/L in the control group [63].

Conclusion

A psychrotolerant bacterium from deep-sea water at Java Trench, Indonesia, was identified as *Priestia flexa* JT4 based on the results of phenotypic and genotypic analyses. Genome characteristic analysis showed pan genome types and number of secondary metabolite

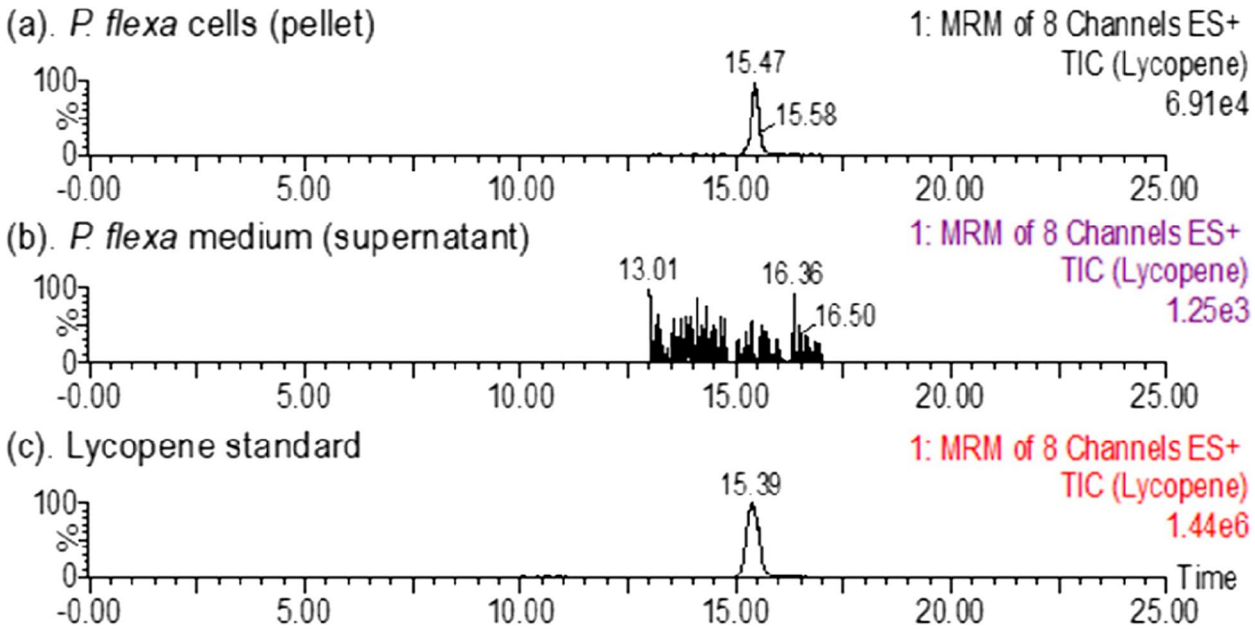

Fig. 12 UPLC-MRM-MS/MS of lycopene extracted from (**a**) *P. flexa* cells and (**b**) medium culture in comparison with (**c**) the authentic standard of lycopene

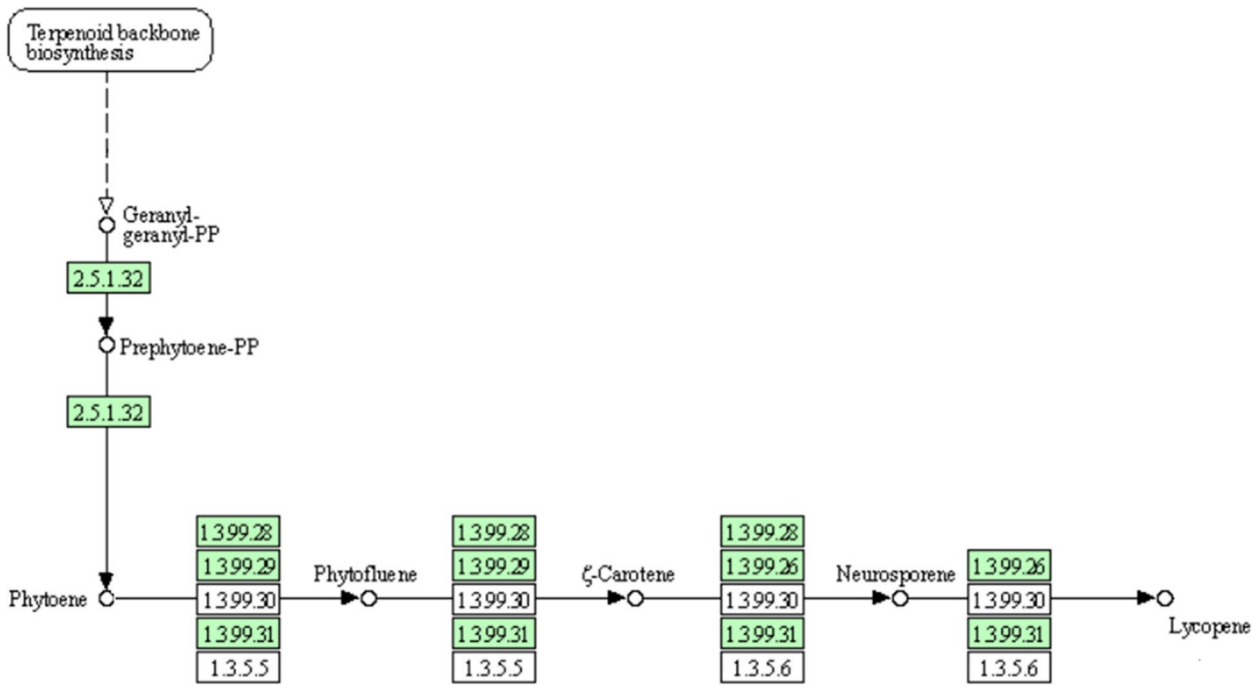

Fig. 13 KEGG Pathway for Lycopene Production based on BGCs Prediction using antiSMASH. The KEGG pathway was composed with two main genes, such as, *crtI* and *crtB*

gene clusters in *P. flexa* JT4 genome, with bioproduct potentials, including lycopene. Analysis of the terpene gene cluster indicated its lycopene production ability, which was identified and confirmed with UPLC-MS/MS. Based on the results of the experiments, more research into optimizing lycopene production from *P. flexa* JT4 is needed so that lycopene can be produced more efficiently.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12864-024-11115-2.

Supplementary Material 1.

Acknowledgements
We gratefully acknowledge the C 2019 TRIUMPH expedition team; the First Institute of Oceanography (FIO), China; the University of Maryland (UMD), USA; and all research teams for all their support.

Authors' contributions
Conceptualization, O.K.R., F.M.D.; methodology, T.K., S.K.R., M.R.P., M.F., R.D.A.O., C.E.F., Z.L.S., M.R., R.S., Y.N., H.B.S., A.O.R.; software, H.N., A.S.P.; validation, H.N.; formal analysis, H.N., J.D.F., M.F., R.D.A.O., R.S., Y.N., A.C., N.P., K., H.B.S., A.O.R.; investigation, O.K.R., F.M.D., C.E.F., M.R.P.; resources, M.R.P., A.S.P. N.D.H., R.D.S., E.; writing—original draft preparation, O.K.R., F.M.D., R.S., Y.N., V.R.; writing—review and editing, J.P.T., A.D., O.K.R. F.M.D.; visualization, A.D., M.F., R.S., Y.N.; supervision, O.K.R., F.M.D., H.N., M.R.M.; project administration, T.K.; funding acquisition, M.R, Z.H. All authors have read and agreed to the published version of the manuscript.

Funding
This research was funded by a grant from the Institut Teknologi Bandung Indonesia International Research Program 2021 (No. 247/IT1.B07.1/SPP-LPPM/I/2021) to F.M.D. O.K.R., H.N., A.S.P., M.R.P., M.R.M., and A.D. R.D.S. was supported by NASA through the University of Maryland grant (#80NSSC18K0777).

Data availability
The dataset(s) supporting the findings of this study have been deposited in National Center for Biotechnology Information (NCBI) under the accession number CP110484.1.

Declarations

Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Competing interests
The authors declare no competing interests.

Author details
[1]Research Center for Deep Sea, The Earth Sciences and Maritime Research Organization, National Research and Innovation Agency, Jakarta, Indonesia. [2]Institut Teknologi Bandung, School of Life Sciences and Technology, Bandung, West Java, Indonesia. [3]Institut Pendidikan Indonesia, Garut, West Java, Indonesia. [4]Department of Biotechnology, Faculty of Science and Technology, Universitas Muhammadiyah Bandung, Bandung, West Java, Indonesia. [5]Research Center for Genetic Engineering, Research Organization for Life Science and Environment, National Research and Innovation Agency, Cibinong, Bogor, Indonesia. [6]Indonesia Biogeography and Biodiversity Research Institute, Bandung 40115, Indonesia. [7]Department of Oceanography, Faculty of Earth Science and Technology, Institut Teknologi Bandung, Bandung, West Java, Indonesia. [8]Korean Collection for Type Cultures (KCTC), Biological Resource Center, Korea Research, Institute of Bioscience & Biotechnology (KRIBB), Jeongeup, Jeollabuk-Do 56212, Republic of Korea. [9]Institut Teknologi Bandung Research Center On Information and Communication Technology, Bandung, West Java, Indonesia. [10]Directorate of Research Vessel Management, National Research, and Innovation Agency, Jakarta, Indonesia. [11]Department of Atmospheric and Oceanic Science, University of Maryland, College Park, Maryland, USA. [12]Faculty of Fisheries and Marine Science, Universitas Diponegoro, Semarang, Indonesia. [13]Bioscience and Biotechnology Research Center, Bandung Institute of Technology, Bandung, West Java, Indonesia. [14]School of Pharmacy, Bandung Institute of Technology, Bandung, West Java, Indonesia.

Received: 29 April 2024 Accepted: 2 December 2024
Published online: 30 December 2024

References

1. Siallagan ZL, Fadli M, de Fretes CE, Opier RDA, Susanto RD, Wei Z, Suhardi VSH, Nugrahapraja H, Radjasa OK, Dwivany FM. Metagenomic Analysis of Deep-Sea Bacterial Communities in the Makassar and Lombok Straits. Sci Rep. 2024;14:25472. https://doi.org/10.1038/s41598-024-74118-9.
2. Srinivasan R, Kannappan A, Shi C, Lin X. Marine Bacterial Secondary Metabolites: A Treasure House for Structurally Unique and Effective Antimicrobial Compounds. Mar Drugs. 2021;19:530. https://doi.org/10.3390/md19100530.
3. Wang Y-N, Meng L-H, Wang B-G. Progress in Research on Bioactive Secondary Metabolites from Deep-Sea Derived Microorganisms. Mar Drugs. 2020;18:614. https://doi.org/10.3390/md18120614.
4. Handayani I, Saad H, Ratnakomala S, Lisdiyanti P, Kusharyoto W, Krause J, Kulik A, Wohlleben W, Aziz S, Gross H, et al. Mining Indonesian Microbial Biodiversity for Novel Natural Compounds by a Combined Genome Mining and Molecular Networking Approach. Mar Drugs. 2021;19:316. https://doi.org/10.3390/md19060316.
5. Suyadi; Satrioajie, W.N.; Syahailatua, A.; Arifin, Z. Banda Deep-Sea Research: History, Mission and Strategic Plan. IOP Conf Ser Earth Environ Sci 2018, 184, 012001, https://doi.org/10.1088/1755-1315/184/1/012001.
6. Webster G, Jones C, Mullins AJ, Mahenthiralingam E. A Rapid Screening Method for the Detection of Specialised Metabolites from Bacteria: Induction and Suppression of Metabolites from Burkholderia Species. J Microbiol Methods. 2020;178: 106057. https://doi.org/10.1016/j.mimet.2020.106057.
7. Liu, Z.; Zhao, Y.; Huang, C.; Luo, Y. Recent Advances in Silent Gene Cluster Activation in Streptomyces. Front Bioeng Biotechnol 2021, 9, https://doi.org/10.3389/fbioe.2021.632230.
8. Albarano L, Esposito R, Ruocco N, Costantini M. Genome Mining as New Challenge in Natural Products Discovery. Mar Drugs. 2020;18:199. https://doi.org/10.3390/md18040199.
9. Belknap KC, Park CJ, Barth BM, Andam CP. Genome Mining of Biosynthetic and Chemotherapeutic Gene Clusters in Streptomyces Bacteria. Sci Rep. 2003;2020:10. https://doi.org/10.1038/s41598-020-58904-9.
10. Radjasa OK, Steven R, Humaira Z, Dwivany FM, Nugrahapraja H, Trinugroho JP, Kristianti T, Chahyadi A, Natanael Y, Priharto N, et al. Biosynthetic Gene Cluster Profiling from North Java Sea Virgibacillus Salarius Reveals Hidden Potential Metabolites. Sci Rep. 2023;13:19273. https://doi.org/10.1038/s41598-023-44603-8.
11. Ren, H.; Shi, C.; Zhao, H. Computational Tools for Discovering and Engineering Natural Product Biosynthetic Pathways. iScience 2020, 23, 100795, https://doi.org/10.1016/j.isci.2019.100795.
12. Paoli L, Ruscheweyh H-J, Forneris CC, Hubrich F, Kautsar S, Bhushan A, Lotti A, Clayssen Q, Salazar G, Milanese A, et al. Biosynthetic Potential of the Global Ocean Microbiome. Nature. 2022;607:111–8. https://doi.org/10.1038/s41586-022-04862-3.
13. Steven, R.; Humaira, Z.; Natanael, Y.; Dwivany, F.M.; Trinugroho, J.P.; Dwijayanti, A.; Kristianti, T.; Tallei, T.E.; Emran, T. Bin; Jeon, H.; et al. Marine Microbial-Derived Resource Exploration: Uncovering the Hidden Potential of Marine Carotenoids. Mar Drugs 2022, 20, 352, https://doi.org/10.3390/md20060352.
14. Bogacz-Radomska L, Harasym J. β-Carotene—Properties and Production Methods. Food Quality and Safety. 2018;2:69–74. https://doi.org/10.1093/fqsafe/fyy004.
15. Ligia, A. da C.C.; Karen, Y.F.K.; Susan, G.K. Microbial Production of Carotenoids A Review. Afr J Biotechnol 2017, 16, 139–146, https://doi.org/10.5897/AJB2016.15763.
16. Wang G-S, Grammel H, Abou-Aisha K, Sägesser R, Ghosh R. High-Level Production of the Industrial Product Lycopene by the Photosynthetic Bacterium Rhodospirillum Rubrum. Appl Environ Microbiol. 2012;78:7205–15. https://doi.org/10.1128/AEM.00545-12.
17. Shao A, Hathcock JN. Risk Assessment for the Carotenoids Lutein and Lycopene. Regul Toxicol Pharmacol. 2006;45:289–98. https://doi.org/10.1016/j.yrtph.2006.05.007.

18. Baran MT, Miziak P, Bonio K. Characteristics of Carotenoids and Their Use in the Cosmetics Industry. Journal of Education, Health and Sport. 2020;10:192–6. https://doi.org/10.12775/JEHS.2020.10.07.020.
19. Himanshu, V.; Roshan, D. Lycopene Market by Form, (Powder and Others), Nature (Natural and Synthetic), and Application (Food and Beverage, Nutraceuticals and Pharmaceuticals, and Cosmetics and Personal Care): Global Opportunity Analysis and Industry Forecast, 2021–2030; 2021;
20. Gupta, R.S.; Patel, S.; Saini, N.; Chen, S. Robust Demarcation of 17 Distinct Bacillus Species Clades, Proposed as Novel Bacillaceae Genera, by Phylogenomics and Comparative Genomic Analyses: Description of Robertmurraya Kyonggiensis Sp. Nov. and Proposal for an Emended Genus Bacillus Limiting It Only to the Members of the Subtilis and Cereus Clades of Species. Int J Syst Evol Microbiol 2020, 70, 5753–5798, https://doi.org/10.1099/ijsem.0.004475.
21. Caspers, H. J. D. H. Strickland and T. R. Parsons: A Practical Handbook of Seawater Analysis. Ottawa: Fisheries Research Board of Canada, Bulletin 167, 1968. 293 Pp. $ 7.50. Internationale Revue der gesamten Hydrobiologie und Hydrographie 1970, 55, 167–167, https://doi.org/10.1002/iroh.19700550118.
22. Radjasa OK, Urakawa H, Kita-Tsukamoto K, Ohwada K. Characterization of Psychrotrophic Bacteria in the Surface and Deep-Sea Waters from the Northwestern Pacific Ocean Based on 16S Ribosomal DNA Analysis. Mar Biotechnol. 2001;3:454–62. https://doi.org/10.1007/s10126-001-0050-1.
23. Morita RY. Psychrophilic Bacteria. Bacteriol Rev. 1975;39:144–67. https://doi.org/10.1128/br.39.2.144-167.1975.
24. Tripathi N, Sapra A. Gram Staining. Treasure Island (FL): StatPearls Publishing; 2021.
25. Frank JA, Reich CI, Sharma S, Weisbaum JS, Wilson BA, Olsen GJ. Critical Evaluation of Two Primers Commonly Used for Amplification of Bacterial 16S RRNA Genes. Appl Environ Microbiol. 2008;74:2461–70. https://doi.org/10.1128/AEM.02272-07.
26. Lane DJ. 16S/23S RRNA Sequencing. In: Stackebrandt E, Goodfellow M, editors. Nucleic acid techniques in bacterial systematics. New York: John Wiley & Sons; 1991. p. 115–75.
27. De Coster W, D'Hert S, Schultz DT, Cruts M, Van Broeckhoven C. NanoPack: Visualizing and Processing Long-Read Sequencing Data. Bioinformatics. 2018;34:2666–9. https://doi.org/10.1093/bioinformatics/bty149.
28. Kolmogorov M, Yuan J, Lin Y, Pevzner PA. Assembly of Long, Error-Prone Reads Using Repeat Graphs. Nat Biotechnol. 2019;37:540–6. https://doi.org/10.1038/s41587-019-0072-8.
29. Darling ACE, Mau B, Blattner FR, Perna NT. Mauve: Multiple Alignment of Conserved Genomic Sequence With Rearrangements. Genome Res. 2004;14:1394–403. https://doi.org/10.1101/gr.2289704.
30. Tanizawa Y, Fujisawa T, Nakamura Y. DFAST: A Flexible Prokaryotic Genome Annotation Pipeline for Faster Genome Publication. Bioinformatics. 2018;34:1037–9. https://doi.org/10.1093/bioinformatics/btx713.
31. Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. BUSCO: Assessing Genome Assembly and Annotation Completeness with Single-Copy Orthologs. Bioinformatics. 2015;31:3210–2. https://doi.org/10.1093/bioinformatics/btv351.
32. Parks DH, Imelfort M, Skennerton CT, Hugenholtz P, Tyson GW. CheckM: Assessing the Quality of Microbial Genomes Recovered from Isolates, Single Cells, and Metagenomes. Genome Res. 2015;25:1043–55. https://doi.org/10.1101/gr.186072.114.
33. Bucchini, F.; Del Cortona, A.; Kreft, Ł.; Botzki, A.; Van Bel, M.; Vandepoele, K. TRAPID 2.0: A Web Application for Taxonomic and Functional Analysis of de Novo Transcriptomes. Nucleic Acids Res 2021, 49, e101–e101, https://doi.org/10.1093/nar/gkab565.
34. Huerta-Cepas, J.; Szklarczyk, D.; Heller, D.; Hernández-Plaza, A.; Forslund, S.K.; Cook, H.; Mende, D.R.; Letunic, I.; Rattei, T.; Jensen, L.J.; et al. EggNOG 5.0: A Hierarchical, Functionally and Phylogenetically Annotated Orthology Resource Based on 5090 Organisms and 2502 Viruses. Nucleic Acids Res 2019, 47, D309–D314, https://doi.org/10.1093/nar/gky1085.
35. Wattam AR, Davis JJ, Assaf R, Boisvert S, Brettin T, Bun C, Conrad N, Dietrich EM, Disz T, Gabbard JL, et al. Improvements to PATRIC, the All-Bacterial Bioinformatics Database and Analysis Resource Center. Nucleic Acids Res. 2017;45:D535–42. https://doi.org/10.1093/nar/gkw1017.
36. Jain C, Rodriguez-R LM, Phillippy AM, Konstantinidis KT, Aluru S. High Throughput ANI Analysis of 90K Prokaryotic Genomes Reveals Clear Species Boundaries. Nat Commun. 2018;9:5114. https://doi.org/10.1038/s41467-018-07641-9.
37. Meier-Kolthoff JP, Carbasse JS, Peinado-Olarte RL, Göker M. TYGS and LPSN: A Database Tandem for Fast and Reliable Genome-Based Classification and Nomenclature of Prokaryotes. Nucleic Acids Res. 2022;50:D801–7. https://doi.org/10.1093/nar/gkab902.
38. Emms DM, Kelly S. OrthoFinder: Phylogenetic Orthology Inference for Comparative Genomics. Genome Biol. 2019;20:238. https://doi.org/10.1186/s13059-019-1832-y.
39. Letunic I, Bork P. Interactive Tree Of Life (ITOL) v5: An Online Tool for Phylogenetic Tree Display and Annotation. Nucleic Acids Res. 2021;49:W293–6. https://doi.org/10.1093/nar/gkab301.
40. Stothard P, Wishart DS. Circular Genome Visualization and Exploration Using CGView. Bioinformatics. 2005;21:537–9. https://doi.org/10.1093/bioinformatics/bti054.
41. Contreras-Moreira B, Vinuesa P. GET_HOMOLOGUES, a Versatile Software Package for Scalable and Robust Microbial Pangenome Analysis. Appl Environ Microbiol. 2013;79:7696–701. https://doi.org/10.1128/AEM.02411-13.
42. Costa SS, Guimarães LC, Silva A, Soares SC, Baraúna RA. First Steps in the Analysis of Prokaryotic Pan-Genomes. Bioinform Biol Insights. 2020;14:117793222093806. https://doi.org/10.1177/1177932220938064.
43. Bertelli C, Laird MR, Williams KP, Lau BY, Hoad G, Winsor GL, Brinkman FS. IslandViewer 4: Expanded Prediction of Genomic Islands for Larger-Scale Datasets. Nucleic Acids Res. 2017;45:W30–5. https://doi.org/10.1093/nar/gkx343.
44. Blin, K.; Shaw, S.; Kloosterman, A.M.; Charlop-Powers, Z.; van Wezel, G.P.; Medema, M.H.; Weber, T. AntiSMASH 6.0: Improving Cluster Detection and Comparison Capabilities. Nucleic Acids Res 2021, 49, W29–W35, https://doi.org/10.1093/nar/gkab335.
45. Kanehisa M, Sato Y, Kawashima M. KEGG Mapping Tools for Uncovering Hidden Features in Biological Data. Protein Sci. 2022;31:47–53. https://doi.org/10.1002/pro.4172.
46. Kim M, Oh H-S, Park S-C, Chun J. Towards a Taxonomic Coherence between Average Nucleotide Identity and 16S RRNA Gene Sequence Similarity for Species Demarcation of Prokaryotes. Int J Syst Evol Microbiol. 2014;64:346–51. https://doi.org/10.1099/ijs.0.059774-0.
47. Pogodin PV, Lagunin AA, Rudik AV, Druzhilovskiy DS, Filimonov DA, Poroikov VV. AntiBac-Pred: A Web Application for Predicting Antibacterial Activity of Chemical Compounds. J Chem Inf Model. 2019;59:4513–8. https://doi.org/10.1021/acs.jcim.9b00436.
48. Liu W, Xie Y, Ma J, Luo X, Nie P, Zuo Z, Lahrmann U, Zhao Q, Zheng Y, Zhao Y, et al. IBS: An Illustrator for the Presentation and Visualization of Biological Sequences. Bioinformatics. 2015;31:3359–61. https://doi.org/10.1093/bioinformatics/btv362.
49. Zobell, C.E. Studies on Marine Bacteria. I. The Cultural Requirements of Heterotrophic Aerobes. J Mar Res 1941, 4.
50. Liu N, Liu B, Wang G, Soong Y-HV, Tao Y, Liu W, Xie D. Lycopene Production from Glucose, Fatty Acid and Waste Cooking Oil by Metabolically Engineered Escherichia Coli. Biochem Eng J. 2020;155: 107488. https://doi.org/10.1016/j.bej.2020.107488.
51. Meier-Kolthoff JP, Auch AF, Klenk H-P, Göker M. Genome Sequence-Based Species Delimitation with Confidence Intervals and Improved Distance Functions. BMC Bioinformatics. 2013;14:60. https://doi.org/10.1186/1471-2105-14-60.
52. Tettelin H, Masignani V, Cieslewicz MJ, Donati C, Medini D, Ward NL, Angiuoli SV, Crabtree J, Jones AL, Durkin AS, et al. Genome Analysis of Multiple Pathogenic Isolates of *Streptococcus Agalactiae* : Implications for the Microbial "Pan-Genome." Proc Natl Acad Sci. 2005;102:13950–5. https://doi.org/10.1073/pnas.0506758102.
53. Willenbrock H, Hallin PF, Wassenaar TM, Ussery DW. Characterization of Probiotic Escherichia Coliisolates with a Novel Pan-Genome Microarray. Genome Biol. 2007;8:R267. https://doi.org/10.1186/gb-2007-8-12-r267.
54. Misawa N. Carotenoids. In Comprehensive Natural Products II: Elsevier; 2010. p. 733–53.
55. Umeno D, Tobias AV, Arnold FH. Evolution of the C 30 Carotenoid Synthase CrtM for Function in a C 40 Pathway. J Bacteriol. 2002;184:6690–9. https://doi.org/10.1128/JB.184.23.6690-6699.2002.
56. Fang L, Pajkovic N, Wang Y, Gu C, van Breemen RB. Quantitative Analysis of Lycopene Isomers in Human Plasma Using High-Performance Liquid Chromatography−Tandem Mass Spectrometry. Anal Chem. 2003;75:812–7. https://doi.org/10.1021/ac026118a.

57. Richelle M, Lambelet P, Rytz A, Tavazzi I, Mermoud A-F, Juhel C, Borel P, Bortlik K. The Proportion of Lycopene Isomers in Human Plasma Is Modulated by Lycopene Isomer Profile in the Meal but Not by Lycopene Preparation. Br J Nutr. 2012;107:1482–8. https://doi.org/10.1017/S0007114511004569.
58. Amorim, A.G.N.; Souza, J.M.T.; Santos, R.C.; Gullón, B.; Oliveira, A.; Santos, L.F.A.; Virgino, A.L.E.; Mafud, A.C.; Petrilli, H.M.; Mascarenhas, Y.P.; et al. HPLC-DAD, ESI–MS/MS, and NMR of Lycopene Isolated From P. Guajava L. and Its Biotechnological Applications. European Journal of Lipid Science and Technology 2018, 120, https://doi.org/10.1002/ejlt.201700330.
59. Huang J, Hui B. The Determination of Lycopene *Z* -isomer Absorption Coefficient on C30-HPLC. Food Sci Nutr. 2020;8:5943–52. https://doi.org/10.1002/fsn3.1879.
60. Schwartz, C.; Frogue, K.; Misa, J.; Wheeldon, I. Host and Pathway Engineering for Enhanced Lycopene Biosynthesis in Yarrowia Lipolytica. Front Microbiol 2017, 8, https://doi.org/10.3389/fmicb.2017.02233.
61. Khan, U.M.; Sevindik, M.; Zarrabi, A.; Nami, M.; Ozdemir, B.; Kaplan, D.N.; Selamoglu, Z.; Hasan, M.; Kumar, M.; Alshehri, M.M.; et al. Lycopene: Food Sources, Biological Activities, and Human Health Benefits. Oxid Med Cell Longev 2021, 2021, https://doi.org/10.1155/2021/2713511.
62. Li L, Liu Z, Jiang H, Mao X. Biotechnological Production of Lycopene by Microorganisms. Appl Microbiol Biotechnol. 2020;104:10307–24. https://doi.org/10.1007/s00253-020-10967-4.
63. Shi Y, Xin X, Yuan Q. Improved Lycopene Production by Blakeslea Trispora with Isopentenyl Compounds and Metabolic Precursors. Biotechnol Lett. 2012;34:849–52. https://doi.org/10.1007/s10529-011-0839-6.

Publisher's Note