Structure and biosynthesis of carotenoids produced by a novel *Planococcus* sp. isolated from South Africa

Anesu Conrad Moyo[1,2] , Laurent Dufossé[3], Daniele Giuffrida[4], Leonardo Joaquim van Zyl[1] and Marla Trindade[1*]

Abstract

Background: The genus *Planococcus* is comprised of halophilic bacteria generally reported for the production of carotenoid pigments and biosurfactants. In previous work, we showed that the culturing of the orange-pigmented *Planococcus* sp. CP5-4 isolate increased the evaporation rate of industrial wastewater brine effluent, which we attributed to the orange pigment. This demonstrated the potential application of this bacterium for industrial brine effluent management in evaporation ponds for inland desalination plants. Here we identified a C_{30}-carotenoid biosynthetic gene cluster responsible for pigment biosynthesis in *Planococcus* sp. CP5-4 through isolation of mutants and genome sequencing. We further compare the core genes of the carotenoid biosynthetic gene clusters identified from different *Planococcus* species' genomes which grouped into gene cluster families containing BGCs linked to different carotenoid product chemotypes. Lastly, LC–MS analysis of saponified and unsaponified pigment extracts obtained from cultures of *Planococcus* sp. CP5-4, revealed the structure of the main (predominant) glucosylated C_{30}-carotenoid fatty acid ester produced by *Planococcus* sp. CP5-4.

Results: Genome sequence comparisons of isolated mutant strains of *Planococcus* sp. CP5-4 showed deletions of 146 Kb and 3 Kb for the non-pigmented and "yellow" mutants respectively. Eight candidate genes, likely responsible for C_{30}-carotenoid biosynthesis, were identified on the wild-type genome region corresponding to the deleted segment in the non-pigmented mutant. Six of the eight candidate genes formed a biosynthetic gene cluster. A truncation of *crtP* was responsible for the "yellow" mutant phenotype. Genome annotation revealed that the genes encoded 4,4′-diapolycopene oxygenase (CrtNb), 4,4′- diapolycopen-4-al dehydrogenase (CrtNc), 4,4′-diapophytoene desaturase (CrtN), 4,4′- diaponeurosporene oxygenase (CrtP), glycerol acyltransferase (Agpat), family 2 glucosyl transferase 2 (Gtf2), phytoene/squalene synthase (CrtM), and cytochrome P450 hydroxylase enzymes. Carotenoid analysis showed that a glucosylated C_{30}-carotenoid fatty acid ester, methyl 5-(6-$C_{17:3}$)-glucosyl-5, 6′-dihydro-apo-4, 4′-lycopenoate was the main carotenoid compound produced by *Planococcus* sp. CP5-4.

Conclusion: We identified and characterized the carotenoid biosynthetic gene cluster and the C_{30}-carotenoid compound produced by *Planococcus* sp. CP5-4. Mass-spectrometry guided analysis of the saponified and unsaponified pigment extracts showed that methyl 5-glucosyl-5, 6-dihydro-apo-4, 4′-lycopenoate esterified to heptadecatrienoic acid ($C_{17:3}$). Furthermore, through phylogenetic analysis of the core carotenoid BGCs of *Planococcus* species we show

*Correspondence: ituffin@uwc.ac.za
[1] Institute for Microbial Biotechnology and Metagenomics (IMBM),
Department of Biotechnology, University of the Western Cape, Bellville,
Cape Town 7535, South Africa
Full list of author information is available at the end of the article

that various C_{30}-carotenoid product chemotypes, apart from methyl 5-glucosyl-5, 6-dihydro-apo-4, 4′-lycopenoate and 5-glucosyl-4, 4-diaponeurosporen-4′-ol-4-oic acid, may be produced that could offer opportunities for a variety of applications.

Keywords: Halophilic bacteria, Pigments, Carotenoid biosynthesis, Whole genome sequencing

Background

Of the pigments produced by bacteria, carotenoids have diverse biological functions that include, coloration, photoprotection, light harvesting, and regulating the fluidity of the bacterial phospholipid bilayer membrane [1–6]. Biotechnologically, carotenoids have been utilized as food colorants, antioxidants, animal feed supplements, nutraceuticals, cosmetics, and pharmaceuticals [7–10].

Numerous bacterial species, including some from the *Planococcus* genus, are known to produce carotenoids [11–18]. The genus *Planococcus* comprises of; halophilic, aerobic, Gram-positive, motile cocci from various environments including saltern ponds [19–21]. The growth of *Planococcus* bacteria and other pigmented halophilic microorganisms in saltern ponds contributes to the red–orange coloration and enhanced evaporation rates of the brine during salt production [22–24]. Recently, we showed that culturing the orange pigmented *Planococcus* sp. CP5-4 isolate in industrial wastewater reverse osmosis brine effluent, at different salt compositions to those of saltern pond brines, resulted in a 20% increase in the evaporation rate of the brine. The resulting increase in the brine evaporation rate was attributed to the production of a carotenoid pigment by *Planococcus* sp. CP5-4, thus, showcasing the potential of pigmented halophilic bacteria as an environmentally friendly and sustainable alternative option to the use of chemical dyes for this purpose [25–29].

Although several strains belonging to the *Planococcus* genus have been reported as carotenoid producers, information on the identification and characterization of the genes and/or biosynthetic gene clusters (BGCs) responsible for carotenoid production in this genus is not extensively reported in the literature. As of the time of writing, the carotenoid BGCs of two strains, *Planococcus maritimus* iso-3 [30] and *Planococcus faecalis* AJ003[T] [16, 31], have been analyzed, while only mentions of the existence of carotenoid biosynthetic genes in the sequenced genomes of *Planococcus maritimus* MKU009 [12], *Planococcus* sp. ANT_H30 [1], *Planococcus donghaensis* JH1T [32], *Planococcus halotolerans* SCU63[T] [33], and *Planococcus rifietoensis* M8[T] [34] has been made.

Furthermore, glucosylated carotenoid acid esters, in which a glucose molecule is esterified to both a triterpenoid carotenoid carboxylic acid and fatty acid (FA) [35], have been reported to be the end products of carotenoid biosynthesis in heterotrophic bacteria that include, but are not limited to, *Staphylococcus aureus* [36–38], *Planococcus maritimus* DSM 17275 [39, 40], *Planococcus faecalis* AJ003[T] [16, 41], *Methylomonas* sp. [42], *Halobacillus halophilus* [43], *Bacillus indicus* [44, 45], and *Bacillus firmus* [45, 46]. However, the occurrence of glucosylated carotenoid esters in bacteria is often overlooked because of the routine use of saponification to hydrolyze the esters and remove the fatty acids before subsequent characterization of the carotenoid compounds. This practice results in the generation of data for glucosylated carotenoid compounds devoid of the esterified fatty acid chain [47–49] and the reporting of incomplete carotenoid structures in the literature.

Given the potential of carotenoids in industry and our demonstration that these may assist in improved evaporation rates of brines, we investigated the nature of the *Planococcus* sp. CP5-4 carotenoid biosynthetic pathway and the compounds produced. Here we present the genome sequence of *Planococcus* sp. CP5-4 and demonstrate that the proposed carotenoid biosynthetic pathway is indeed responsible for production of a glucosylated carotenoid ester.

Results and discussion

Isolation of carotenoid mutant strains, genome sequencing and classification

Yellow and unpigmented mutant phenotypes designated *Planococcus* sp. CP5-4_YE and CP5-4_UN, respectively, were generated following MMS mediated mutagenesis (Fig. 1). Reports in the literature on the elucidation of pigment biosynthesis pathways in halophilic bacteria provide support for the production of white (or unpigmented) and yellow mutant strains following chemical mutagenesis using DNA alkylating agents [41, 48].

Genome sequencing of the *Planococcus* sp. CP5-4 strains and the analysis of the genomes were conducted to obtain an accurate classification of the wild type *Planococcus* sp. CP5-4 isolate that was previously tentatively identified as a novel species and to identify the pigment biosynthetic gene(s) or gene cluster(s) responsible for the expression of the orange pigment produced by this strain. The draft genome assembly of *Planococcus* sp. CP5-4 consists of 34 contigs that generated a genome size of 3 488 448 bp with a 47.5% G + C content (Additional file 1: Table S1 and S2). Of the 3612 predicted genes, 3 469 are
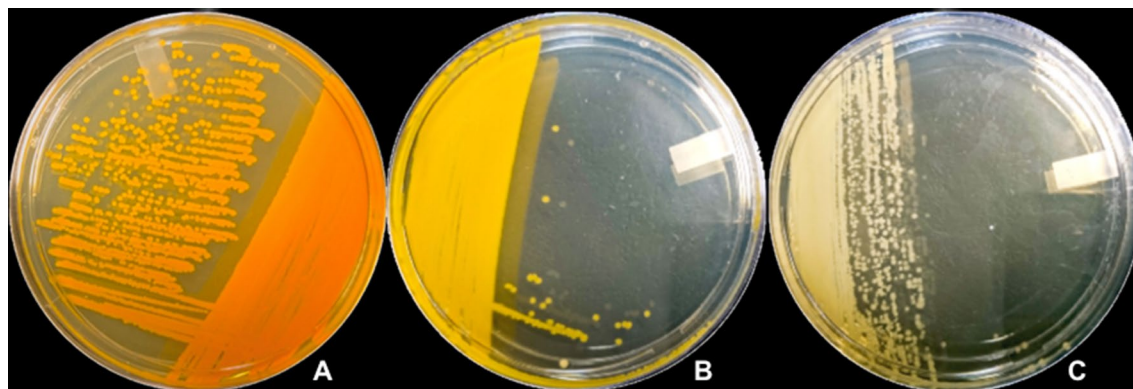
Moyo et al. Microbial Cell Factories   (2022) 21:43

Page 3 of 13

Fig. 1 Wild type *Planococcus* sp. CP5-4 producing orange pigment (**A**), yellow (**B**), and unpigmented (**C**) mutant strains

protein-coding genes and 76 are RNAs (17 rRNA and 65 tRNA genes) (Additional file 1: Table S2). The wild-type *Planococcus* sp. CP5-4 genome grouped together with *Planococcus plakortidis, Planococcus maitriensis, Planococcus* sp. 002833405, *Planococcus rifietoensis, Planococcus maritimus*_B and *Planococcus maritimus* following classification (Additional file 1: Fig. S1). The placement of the *Planococcus* sp. CP5-4 strain in the tree suggests that *Planococcus* sp. CP5-4 is a new species in the *Planococcus* genus. However, when compared to genomes of related *Planococcus* species, the draft genome assembly for *Planococcus* sp. CP5-4 is unremarkable with respect to size, number of genes encoded, number of tRNA's and G+C content (Additional file 1: Table S5).

Identification of the carotenoid biosynthetic gene cluster
Eight ORFs were predicted to be involved in carotenoid biosynthesis in the wild type *Planococcus* sp. CP5-4 genome. Six of the eight predicted ORFs were located on contig 1 as a 7876 bp biosynthetic cluster, while the remainder of the ORFs were located on contig 5 (Fig. 2).

A query of the protein sequence of the putative ORFs against the UniProtKB/Swiss-Prot database revealed that the sequences were related to those responsible for the biosynthesis of C_{30}-carotenoids (Additional file 1: Table S3). Although functional similarity is challenging to infer when global sequence similarity is low, very different sequences can have largely similar activities based on conserved active-sites or functionally relevant regions [52]. As such, many proteins of barely detectable sequence similarity have the same function. It is generally considered that sequences with greater than 30–40% identity are functionally similar [53, 54]. Therefore, we predicted that the pigment produced by *Planococcus* sp. CP5-4 would be a C_{30}-carotenoid. The CrtNb protein, a putative oxygenase, shared the lowest identity at 35.8%, suggesting that it might display different
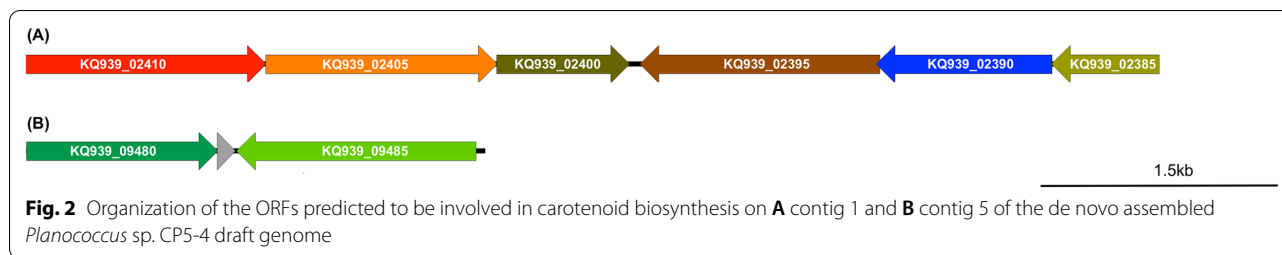
substrate specificity or kinetics compared to those that have been characterized [38].

Mapping the sequence reads from the *Planococcus* sp. CP5-4_YE and CP5-4_UN mutant strains to the wild type de novo assembled contigs revealed that copy number variants, specifically deletions measuring greater than 1 Kb in size from the wild type *Planococcus* sp. CP5-4 genome, were responsible for the altered phenotypes. For *Planococcus* sp. CP5-4_YE a 3 Kb sequence deletion resulted in the truncation of the *crtP* and three additional ORFs (Additional file 1: Figure S2). For *Planococcus* sp. CP5-4_UN a 146.691 Kb sequence deletion that included the six ORFs making up the carotenoid biosynthetic cluster resulted in the production of the unpigmented mutant strain following MMS mutagenesis (Additional file 1: Figure S3).

Carotenoid BGC core gene phylogenetic analysis
According to Takemura et al. [30], the genomes of several *Planococcus* species, such as *P. maritimus* iso-3, *P. faecalis* AJ003^T, *P. plakortidis* and *P. halocryophilus* contain a similar carotenoid biosynthesis gene cluster that includes an additional biosynthesis gene to the six predicted on contig 1 of the wild type *Planococcus* sp. CP5-4 genome.

Guided by this information and the results from the functional annotation and identification of the carotenoid biosynthetic gene cluster in *Planococcus* sp. CP5-4, an investigation of the carotenoid BGCs from different *Planococcus* species was conducted to resolve their evolutionary relationship and group them into gene cluster families (GCFs) containing BGCs linked to a highly similar carotenoid product chemotype. The genetic diversity of BGCs within GCFs is often directly related to structural differences between their molecular products [55, 56]. Hence, the evolutionary relationships between the

Moyo et al. Microbial Cell Factories   (2022) 21:43

Page 4 of 13

Fig. 2 Organization of the ORFs predicted to be involved in carotenoid biosynthesis on **A** contig 1 and **B** contig 5 of the de novo assembled *Planococcus* sp. CP5-4 draft genome

core genes of the carotenoid BGCs of the *Planococcus* species were inferred using CORASON.

Using the *crtM* gene sequence as a query, 36 of 39 carotenoid BGCs detected by antiSMASH 5.0 in the queried *Planococcus* bacterial genomes were placed in the BGC phylogeny shown in Fig. 3. The 36 carotenoid BGCs were primarily grouped into two distinct clades. The first clade was named the 'methyl 5-glucosyl-5, 6-dihydro-apo-4, 4′-lycopenoate clade' because it contained *Planococcus maritimus* DSM 17275 which is a known methyl 5-glucosyl-5, 6-dihydro-apo-4, 4′-lycopenoate producer [13, 40]. The second clade of interest, termed the '5-glucosyl-4, 4-diaponeurosporen-4′-ol-4-oic acid clade,' comprised BGCs with genes related to 5-glucosyl-4, 4-diaponeurosporen-4′-ol-4-oic acid biosynthesis which incorporates the recently characterized *Planococcus faecalis* AJ003 carotenoid BGC. The *Planococcus* sp. CP5-4 query BGC was placed into the methyl 5-glucosyl-5, 6-dihydro-apo-4, 4′-lycopenoate clade, indicating that the *Planococcus* sp. CP5-4 BGC should mediate synthesis of a methyl 5-glucosyl-5, 6-dihydro-apo-4, 4′-lycopenoate-like compound. Although several *Planococcus* carotenoid pathways group within these two clades, their genomic distance to characterized BGCs indicate that these species may produce novel chemotypes and should be a focus for future investigation.
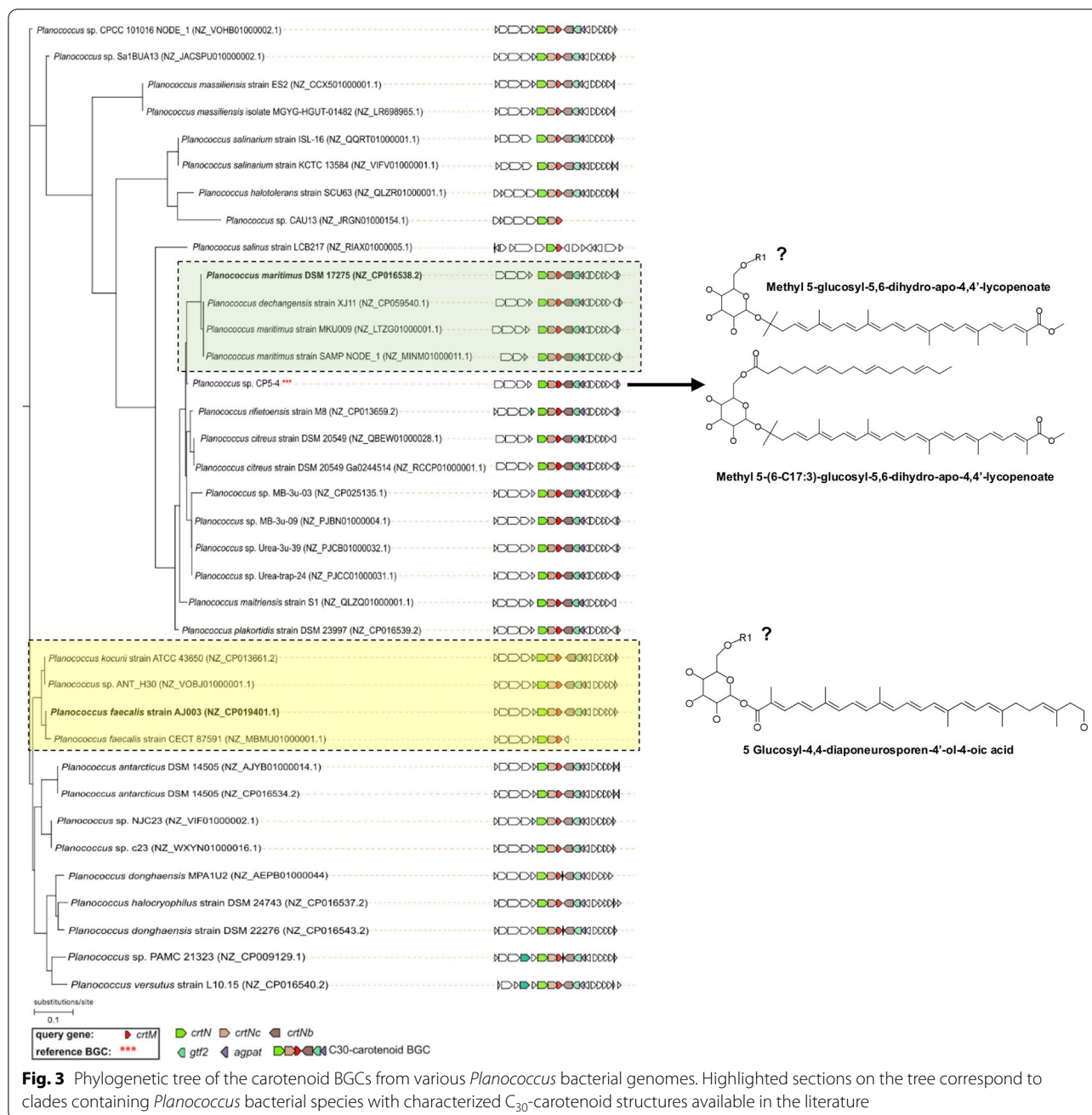
Pigment characterization

Analysis of saponified and unsaponified pigment extracts from *Planococcus* sp. CP5-4 were conducted in an attempt to resolve the complete structure of the carotenoid compound produced since the process of saponification results in the removal of fatty acid moieties that form part of the carotenoid structure in addition to removing lipids that may interfere with chromatographic separation [57]. The main compounds identified from the saponified pigment extract are presented in Table 1.

The pigment extracts were monitored at 280 nm and 450 nm to detect both the shorter polyene chain colourless carotenoids, which absorb maximally in the 280–320 nm region [58], and the long polyene chain coloured carotenoids that absorb in the visible region of the spectrum between 400 and 500 nm [59] Fig. 4).

The molecular formulae deduced for the compounds present in the saponified pigment extract from *Planococcus* sp. CP5-4 were mostly for C_{30} compounds (Table 1). Among these, the dominant peak (peak 3) in the chromatogram obtained at 450 nm in Fig. 4 B was taken to represent the most abundant C_{30}-carotenoid compound in the saponified extract. The $[M+H]^+$ peak of this compound had a m/z of 625.3732 (calcd: 625.3737; error:—1.3 ppm), and a predicted molecular formula of $C_{37}H_{53}O_{8}$. The compound was therefore deduced to be methyl 5-glucosyl-5, 6-dihydro-apo-4, 4′-lycopenoate with a molecular formula of $C_{37}H_{52}O_8$ informed by comparison with the MS data of the C_{30} carotenoids previously reported [32, 33] (Fig. 5A). The presence of a hexose sugar ($\Delta m/z$ of 180 Fig. 5B) on the structure of methyl 5-glucosyl-5, 6-dihydro-apo-4, 4′-lycopenoate was revealed by the fragmentation peak with a measured m/z of 445.31 $[M+H]^+$ (error: -0.2 ppm). The 445.31 mass can be accounted for as the difference of 180 (loss of glucose) in the mass of methyl 5-glucosyl-5, 6-dihydro-apo-4, 4′-lycopenoate detected following CID (Figs. 5B).

When analysing the unsaponified carotenoid extract, peak 3 eluting at 8 min 37 s was dominant with a UV–Vis spectral absorption maximum of 466 nm (Fig. 6A) and an absorption spectrum typical of a methyl-glucosyl-dihydro-apo-4, 4′-lycopenoate fatty acid ester [44, 60]. A base peak with m/z 871.5715 $[M+H]^+$ was obtained from the fragmentation of parent ion for peak 3 (Fig. 6C). The presence of an additional fatty acid moiety is revealed by subtracting the m/z of 445.3098 from the base peak of the unsaponified carotenoid (m/z 871.5715 $[M+H]^+$) to give a $\Delta m/z$ of 426. Removing the mass of the glucosyl moiety (162) yields a remainder of m/z 264.1993, likely representing a $C_{17:3}$ fatty acid moiety (Fig. 6D). Considering that a H_2O molecule was eliminated during the esterification process the mass of the fatty acid moiety becomes 264.1993, which represents that of a $C_{17:3}$—heptadecatrienoic acid. Thus, peak 3 in Fig. 6A was designated as methyl 5-(6-$C_{17:3}$)-glucosyl-5, 6-dihydro-apo-4, 4′-lycopenoate derived from the parent mass of 871.5715

Moyo *et al. Microbial Cell Factories*   (2022) 21:43

Page 5 of 13

Fig. 3 Phylogenetic tree of the carotenoid BGCs from various *Planococcus* bacterial genomes. Highlighted sections on the tree correspond to clades containing *Planococcus* bacterial species with characterized C$_{30}$-carotenoid structures available in the literature

$[M+H]^+$ comprised of a hexose sugar ($\Delta m/z$ 162) esterified to a heptadecatrienoic acid ($\Delta m/z$ 264) and attached to 445.31 $[M+H]^+$ dihydro-methyl-apo-4, 4'-lycopenoate (Fig. 6C and D).

Although several polyunsaturated fatty acids have been reported to be produced by bacteria [61–64], prior to this study there has been no report of the production of an odd chain polyunsaturated fatty acid by species in the *Planococcus* genera.

The production of the heptadecatrienoic acid esterified to methyl 5-glucosyl-5, 6-dihydro-apo-4, 4'-lycopenoate may be attributed to the possible conversion of succinyl-CoA to propionyl-CoA by *Planococcus* sp. CP5-4. Three ORFs coding for the large and small subunits of a methylmalonyl-CoA mutase, and YgfD; a protein forming a complex with the methylmalonyl-CoA mutase subunits in the pathway for conversion of succinyl-CoA to propionyl-CoA were identified (Additional file 1: Table S5).

Moyo et al. Microbial Cell Factories   (2022) 21:43

Page 6 of 13

Table 1 UPLC-DA-MS detected compounds in the saponified pigment extract from *Planococcus* sp. CP5-4

*Peak	t_R (mins)	Formula	λ_{max} (nm)	Measured Mass (m/z)	Calculated Mass (m/z)	Error (ppm)
1	1.74	ND	271.1	ND	ND	ND
2	2.10	ND	236.1		ND	
3	3.04	$C_{37}H_{53}O_8$	447, 467.12, 492.12	625.3732	625.3737	− 1.3
4	3.46	$C_{26}H_{37}O_6$	268.12, 277.12, 283.12	445.2589	445.2590	− 0.2
5	6.06	ND	230	ND	ND	ND
6	7.62	$C_{30}H_{43}O_2$	451.12	435.3277	435.3263	1.4

ND not determined

*Peak numbers refer to Fig. 4

The condensation of both propionyl-CoA and malonyl-CoA results in the formation of 3-oxovaleryl-ACP, which is the launching point for odd-chain FA synthesis. This five-carbon compound goes through elongation, where two carbons are added in each cycle leading to synthesis of odd-chain FAs [65, 66]. Moreover, the introduction of the three double bonds to the heptadecanoic FA chain to produce the heptadecatrienoic acid may have been mediated by a cell membrane bound fatty acid desaturase, a putative protein that could fulfill this role was also identified (Additional file 1: Table S5).

Regarding the truncation of *crtP*, UPLC analysis of the saponified extract from the yellow mutant strain revealed peak 6, previously observed eluting as a minor peak in the chromatogram of the wild type strains' saponified extract (Fig. 4A), as the predominant peak eluting at 7 min 59 s (Fig. 7A).

The aforementioned peak gave a visible spectral maximum at 451 nm and a $[M+H]^+$ ion peak at *m/z* 435. 3269 corresponding to a predicted formula of $C_{30}H_{43}O_2$ (error: 1.4 ppm; Fig. 7B). These data suggest that the yellow pigment may be a putative glucosylated diapolycopene, supported by the yellow glucosylated diapolycopene and glucosylated diapolycopene fatty acid esters absorbing maximally between 449 to 454 nm from saponified and unsaponified carotenoid extracts from yellow *Bacillus* spore formers [67]. The truncation of *crtP* is theorized to have resulted in the translation of a non-functional enzyme that could not mediate the addition of a terminal O-methyl ester group to glucosylated diapolycopene during the biosynthesis of the orange methyl 5-(6-C17:3)-glucosyl-5, 6-dihydro-apo-4, 4′-lycopenoate, thus resulting in the production of the yellow pigment. According to Sy et al. [68], the main structural difference between the glucosylated orange and yellow C_{30} pigments is the absence of the terminal O-methyl ester group in the yellow pigment.
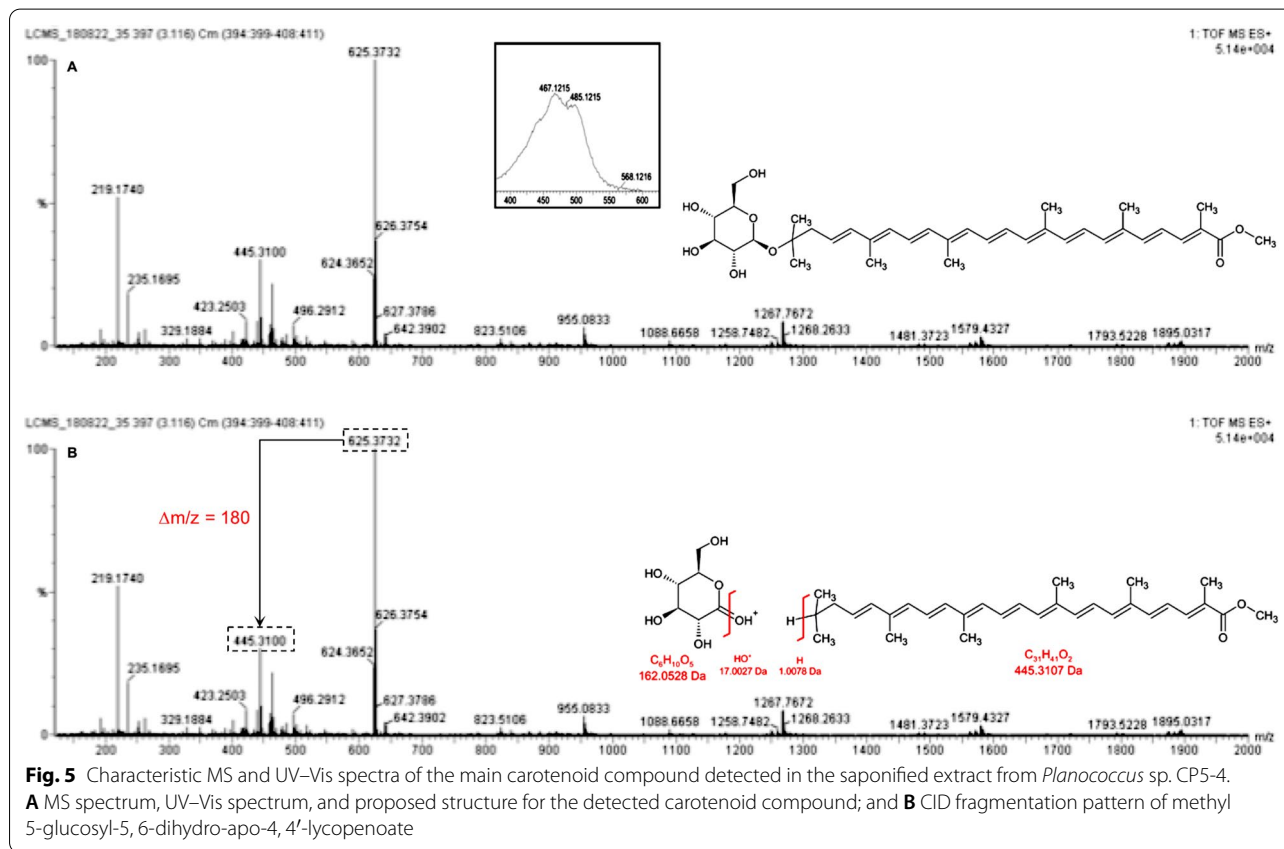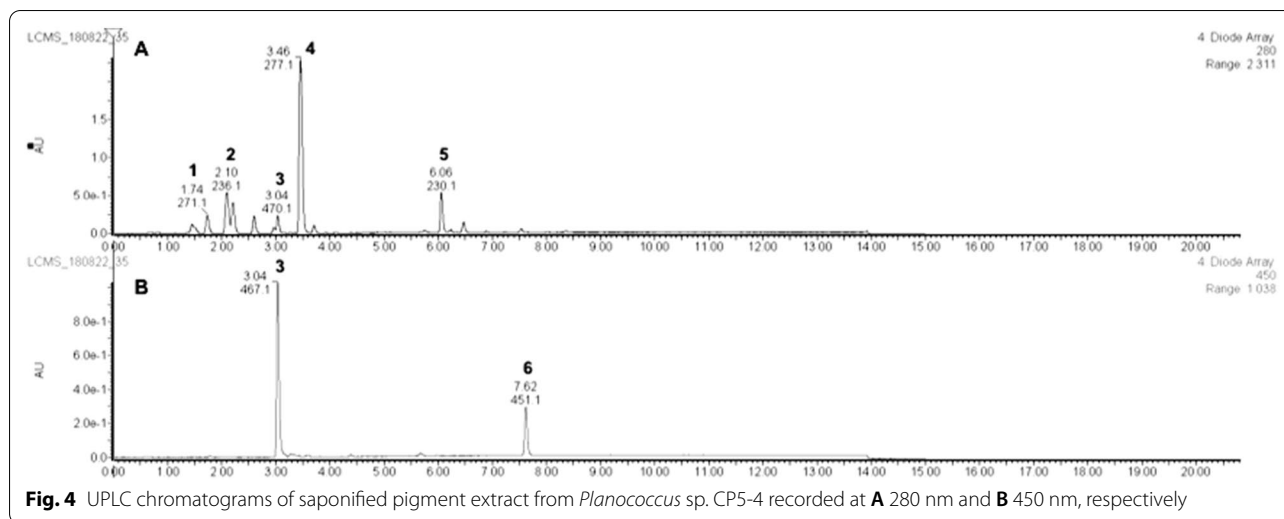
Conclusions

Through isolation of pigment mutants combined with genome sequencing and mass-spectrometry we identified the genes responsible for the biosynthesis of a C_{30}-carotenoid compound produced by *Planococcus* sp. CP5-4. The unsaponified pigment consists of methyl 5-glucosyl-5, 6-dihydro-apo-4, 4′-lycopenoate esterified to an odd chain unsaturated ($C_{17:3}$) heptadecatrienoic acid. It is possible that the production of the unsaturated FA esterified to glucosyl-5, 6-dihydro-apo-4, 4′-lycopenoate is a mechanism for *Planococcus* sp. CP5-4 to maintain membrane fluidity. Furthermore, through genomic core phylogenetic analysis of the carotenoid BGCs of species within the *Planococcus* genus we hypothesize that carotenoid chemotypes apart from methyl 5-glucosyl-5, 6-dihydro-apo-4, 4′-lycopenoate and 5-glucosyl-4, 4-diaponeurosporen-4′-ol-4-oic acid may be produced thus paving the way for future studies to conduct an informed determination of the potential applications of these molecules.
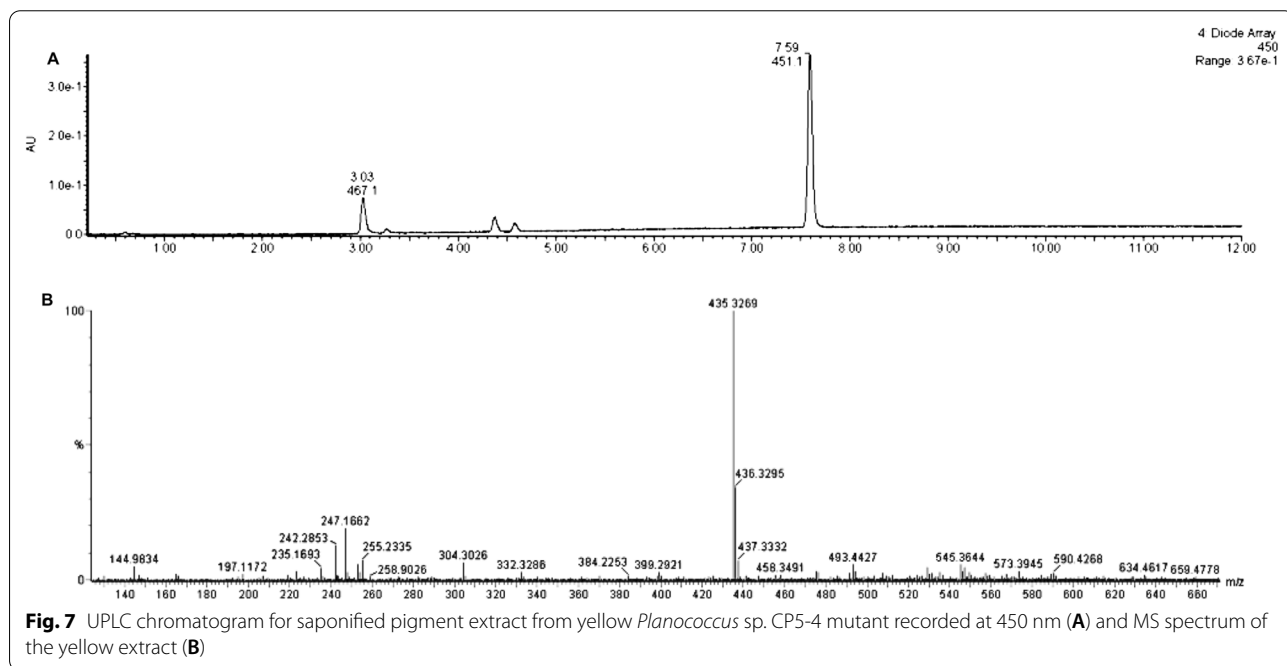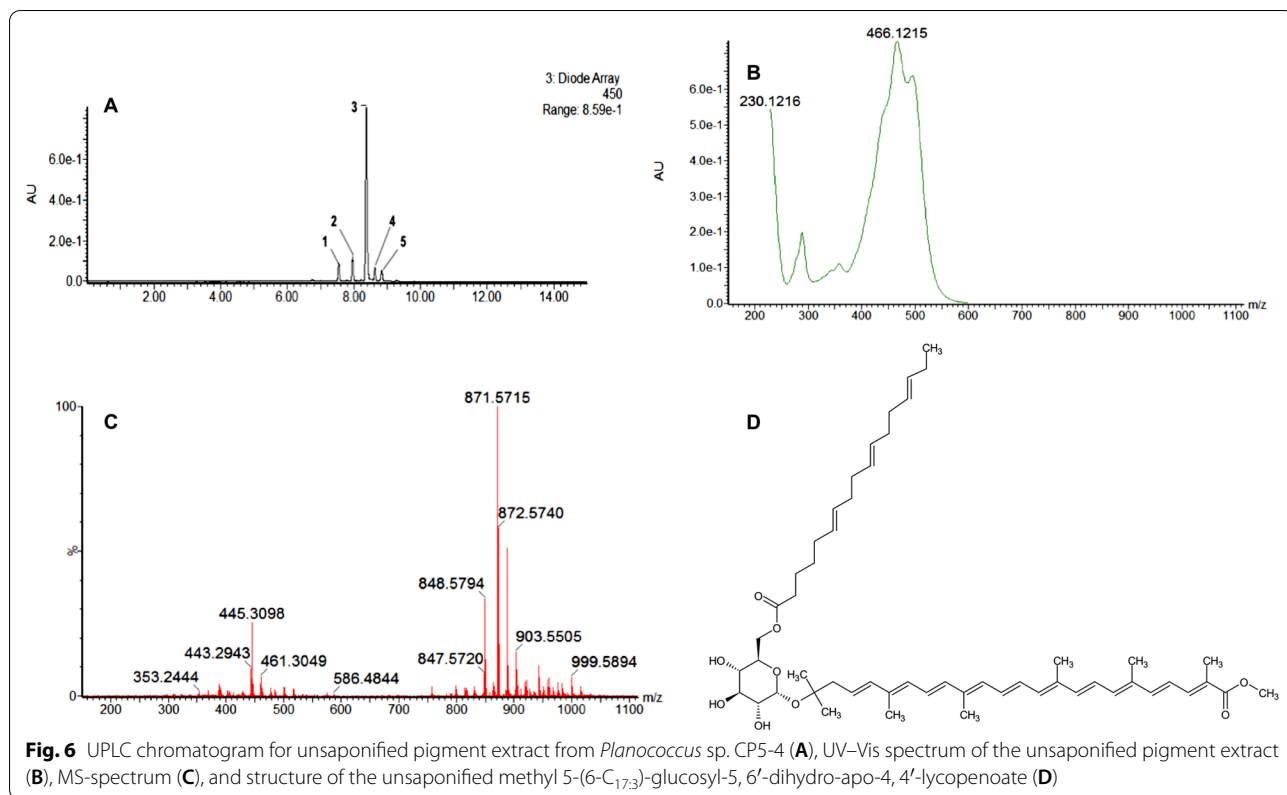
Methods

Organism cultivation

The *Planococcus* sp. CP5-4 strain used in this study originated from a collection of bacterial cultures previously isolated from brine samples taken from the Cerebos crystallizer salt ponds in Velddrif, Western Cape, South Africa (S 32°47′10,632, E 18°10′9,499) [69]. *Planococcus* sp. CP5-4 was grown in complex medium (Tryptic Soy Broth; Sigma-Aldrich, Darmstadt, Germany, supplemented with 5% (w/v) NaCl and 0.5 M sorbitol). Where necessary 15 g/L of agar were added to the broth prior to sterilizing. The inoculated broth was incubated at room temperature with shaking at 150 rpm, while the agar plates were incubated at 28 °C.

Fig. 4 UPLC chromatograms of saponified pigment extract from *Planococcus* sp. CP5-4 recorded at **A** 280 nm and **B** 450 nm, respectively

Fig. 5 Characteristic MS and UV–Vis spectra of the main carotenoid compound detected in the saponified extract from *Planococcus* sp. CP5-4. **A** MS spectrum, UV–Vis spectrum, and proposed structure for the detected carotenoid compound; and **B** CID fragmentation pattern of methyl 5-glucosyl-5, 6-dihydro-apo-4, 4'-lycopenoate

Random chemical mutagenesis

Random chemical mutagenesis was employed to generate pigment mutants to be used in the characterization of the pigment biosynthetic pathway. Mutagenesis was carried out on cells at the mid exponential phase of growth (OD660nm 0.7–0.75) as follows: The cells were harvested by centrifugation at $3214 \times g$ for 10 min, washed with, and re-suspended in 50 mL of the growth medium. The cell suspension was then concentrated 50-fold in quarter strength Ringer's solution (Sigma-Aldrich, Steinheim, Germany) and methyl methanesulfonate (Sigma-Aldrich, Steinheim, Germany) added to a final concentration

Fig. 6 UPLC chromatogram for unsaponified pigment extract from *Planococcus* sp. CP5-4 (**A**), UV–Vis spectrum of the unsaponified pigment extract (**B**), MS-spectrum (**C**), and structure of the unsaponified methyl 5-(6-$C_{17:3}$)-glucosyl-5, 6′-dihydro-apo-4, 4′-lycopenoate (**D**)

Fig. 7 UPLC chromatogram for saponified pigment extract from yellow *Planococcus* sp. CP5-4 mutant recorded at 450 nm (**A**) and MS spectrum of the yellow extract (**B**)

of 0.15 M. This mixture was inverted several times and incubated for 5 min at room temperature in a fume hood. At the end of the incubation period, the dosed wild type cells were washed thrice with 30 mL Ringer's solution through repeated suspension and centrifugation for 10 min at $3\,889 \times g$ in a refrigerated centrifuge

Moyo et al. Microbial Cell Factories   (2022) 21:43

Page 9 of 13

set to 20 °C. After the final washing step, the pelleted cells were re-suspended in 1 mL TSB containing 15%(v/v) glycerol and 5% (w/v) NaCl to make the mutant library, which was stored at − 80° C until needed. Screening of the library was conducted by spread plating a one in one thousand dilution of the library in Ringer's solution on TSB-salt-sorbitol agar plates. Colonies were then visually inspected for pigmentation differing from that of the wild type after incubation at 28° C for 48 h.

Genome sequencing and analysis
Sequencing, assembly and genome-based classification of Planococcus sp. CP5-4

Genomic DNA was extracted from cells pelleted from 5 mL cultures of the *Planococcus* sp. CP5-4 strains using a modified version of the method reported by Wang et al. [70]. The following modifications were made: lysozyme concentration was increased to 25 mg/mL, 0.2 mg/mL (final concentration) of proteinase K was added to the lysis buffer and cells were incubated overnight at 37 °C in the lysis buffer. The quality and quantity of the extracted DNA were verified through agarose gel electrophoresis and fluorometric measurement using a Qubit 2.0 fluorometer prior to library preparation using an Illumina Nextera XT kit following the manufacturer's recommendations. The paired end libraries were sequenced on an in-house Illumina MiSeq instrument using the MiSeq v2 chemistry (Illumina, San Diego, CA, USA) yielding fastq files containing 2×250 bp reads. Following trimming of the reads, de novo assembly was performed using Geneious ver. 11.1.0 [71, 72] with default settings. Contigs with a minimum length less than 1000 bp were removed from the final assembly. The final numbers of contigs were 34, 43 and 45 for the wild type, yellow and unpigmented *Planococcus* CP5-4 strains, respectively. This Whole Genome Shotgun project was deposited in the DDBJ/ENA/GenBank database under Bio Project PRJNA738312 and Accession Numbers JAHREQ000000000.1, JAHPZO000000000.1 and JAHRBB000000000.1.

To obtain an accurate taxonomic classification of the *Planococcus* sp. CP5-4 strain, the contigs were queried using the Genome Taxonomy Database Toolkit (GTDB-Tk) [73] in the Kbase web portal [74, 75]. The GTDB-Tk was used in preference to the NCBI [76] database because the GTDB provides a comprehensive genome-based taxonomy with bacterial and archaeal taxa circumscribed on the basis of monophyly and relative evolutionary divergence [77] compared to the NCBI database, reported to show inconsistent taxonomies with many polyphyletic groupings for bacteria [73, 77, 78].

Annotation, and identification of carotenoid biosynthetic gene cluster

Functional annotation of the de novo assembled *Planococcus* sp. CP5-4 contigs was conducted in Blast2Go (B2G) ver. 5.2.2 [79, 80]. The annotations for the open reading frames (ORFs) predicted to be involved in carotenoid biosynthesis by B2G were validated by querying the translated gene sequences in the UniProtKB/Swiss-Prot database [81] for homologous sequences. The UniProtKB/Swiss-Prot database was used because it contains manually annotated records with information extracted from literature and curator-evaluated computational analysis that brings together experimental results, computed features and scientific conclusions [82, 83], as opposed to the computationally analyzed records awaiting full manual annotation obtained by the NCBI QBlast algorithm in the non-redundant nucleotide (nr/nt) collection database used by B2G. To confirm that the predicted ORFs were responsible for carotenogenesis, the paired-end sequence reads of the yellow and unpigmented mutant strains were mapped to the de novo assembled contig(s) of the *Planococcus* sp. CP5-4 wild type strain, which contained the predicted carotenogenesis ORFs using Geneious as the mapper. The sensitivity of the operation was set on 'Medium-Sensitivity/Fast' and the 'Fine Tuning' option set on 'iterate' up to five times to improve the results by aligning reads to each other in addition to the reference sequence. Following mapping, variant calling on the mapped data was conducted to identify the mutations responsible for the altered phenotypes of the mutant strains. Variants were detected in the mapped data using default settings in Geneious with the option to 'Analyze the Effect of Polymorphisms on Translation' chosen, and the default genetic code changed to 'Bacterial'.

Carotenoid BGC genomic core phylogenetic analysis

The genomic core of carotenoid BGCs identified in 38 *Planococcus* genome sequences downloaded from the National Center for Biotechnology Information, U.S. National Library of Medicine database (https://www.ncbi.nlm.nih.gov/genome/browse#!/prokaryotes/planococcus) and that of *Planococcus* sp. CP5-4, were analyzed using CORASON software version 1.0 [56]. To conduct the analysis, the *Planococcus* genomes were first analyzed using antiSMASH 5.0 [84] at default settings and the predicted BGCs associated with the queried *Planococcus* genome sequences downloaded as.gbk files.

A database of all the identified carotenoid BGCs in the query sequences was created. The carotenoid BGC predicted to be present in the *Planococcus* sp. CP5-4

Moyo *et al. Microbial Cell Factories*   (2022) 21:43

Page 10 of 13

genome was chosen as the reference BGC for the database, while the *crtM* gene from the biosynthetic core of the reference BGC was chosen as the query protein. The choice for the reference BGC and query protein for the database was made after verification that: the reference BGC was one of the longest BGCs in the database, the query protein came from a biosynthetic core gene or an additional biosynthetic gene close to the core, and that the query protein was present in at least half of the BGCs in the database as suggested by Chanson, Moreau, and Duplais [85]. Once the reference BGC and protein query were selected for the database, the CORASON software (https://github.com/nselem/corason) was used to determine the genomic core of carotenoid BGCs, and to infer the evolutionary relationship of the BGCs.

Pigment analysis
Carotenoid extraction
For carotenoid extraction, cells from a 5-day-old, 1 L culture were pelleted by centrifugation at 4 629 $\times g$ at 4 ℃ for 15 min, and washed three times with sterile water. Cell pellets were dried and a mixture of acetone:methanol (7:3 v/v) containing 0.1% butylhydroxytoluene (BHT) as antioxidant added. The resulting cell suspensions were subjected to freeze—thaw cycles in liquid N_2 to facilitate extraction of the pigment. The extracts were then centrifuged at 12,857 $\times g$ for 10 min at 4 ℃, and the coloured supernatant pipetted into foil-covered 50 mL Falcon tubes for protection from photo oxidation. Successive extractions were carried out on the cell pellets until both the solvent and the cells were colourless. The solvent phases were pooled together and evaporated to dryness under a N_2 gas stream. Duplicate extractions were conducted for the wild type *Planococcus* sp. CP5-4 culture and one set of the extracts saponified concurrently with the yellow *Planococcus* sp. CP5-4 mutant extract as described by Cardinault et al. [86]. Pigment from the yellow *Planococcus* sp. CP5-4 mutant strain was extracted to identify the effects of the mutation(s) induced by MMS on the structure of the pigment.

Following saponification, the mixture was centrifuged (10 min, 4629 $\times g$, 4 ℃) and the upper orange phase collected, washed with sterile water and recollected before being evaporated under N_2 gas to produce an 'oily' red extract. The lower phase containing the saponified products of carotenoid esters was discarded, and the 'oily' red residue stored together with the dried unsaponified extract under nitrogen at − 80 ℃ before further processing. Extraction and saponification procedures were conducted in the dark.

Ultra-performance liquid chromatography (UPLC)
Prior to the determination of the composition of the pigment extracts, the unsaponified and saponified extracts were re-dissolved in 500 μL of acetone containing 0.1%

BHT, loaded onto preparative thin layer chromatography (PLC) silica gel plates measuring 4 cm × 14 cm (60 F_{254} 2 mm; Merck, Darmstadt, Germany), which were developed in the dark in a chamber containing hexane: acetone (70:30) as mobile phase. The separated pigment phases were collected, dried under N_2 stream, and the carotenoid compositions analyzed using a Waters Synapt G2 High-Definition (HD) quadrupole time-of-flight (qTOF) mass spectrometer (MS) connected to a Waters Acquity ultra-performance liquid chromatograph (UPLC) (Waters, Milford, MA, USA). For the analysis, 5 μL of the samples were injected into a Waters BEH C_{18}, 2 × 100 mm, 1.7 μm column thermostated at 45 ℃.

The column was initially eluted for 5 min using a gradient of 30% of solvent A (0.1% formic acid) and 62% acetonitrile containing 0.1% formic acid (solvent B), then for a further 25 min using 100% (solvent B). After which, the column was returned to the initial elution conditions (30% solvent A and 62% solvent B) and equilibrated over 30 min. In all instances the flow rate was maintained at 0.4 mL/min. The separation of components was monitored using a Waters eλ photodiode array detector set to measure in the 200–600 nm range and a mass spectrometer at the 150–1500 m/z range in MSe positive mode. Data was acquired using two separation functions; F1: at a low collision energy (4 V) and F2: using a collision energy ramp (25 − 60 V) to simultaneously acquire both unfragmented and fragmented data. Leucine enkephalin was used as lock mass (reference mass) for accurate mass determination and the instrument was calibrated weekly with sodium formate. The following MS settings were used: cone voltage of 15 V, desolvation temperature of 275 ℃, desolvation gas at 650 L/h, and the rest of the MS settings optimized for best resolution and sensitivity. The identification of components was performed by the analysis of the absorbance spectra in connection with mass spectra using the Targertlynx module of the MassLynx ver. 4.1 software (Waters, Milford, USA).

Supplementary Information

Additional file 1: Table S1. De novo assembly report for Planococcus sp. CP5-4 strains. **Table S2.** Features for the de novo assembled CP5-4 genome. **Table S3.** ORFs predicted to be associated with carotenoid biosynthesis on contigs 1 and 5 of the de novo assembled Planococcus sp. CP5-4 genome. **Table S4.** Average nucleotide identity comparison between CP5-4 and closely related GTDB Planococcus species. **Table S5.** Genome property comparison between CP5-4 and related Planococcus species. **Table S6.** Locus tags in CP5-4 involved in unsaturated fatty acid biosynthesis. **Figure S1.** Placement of the Planococcus CP5-4 strain in the GTDB bac 120 tree relative to other Planococcus species. The genome identifiers of the organisms are shown on the tree branches, where appropriate, both the GTDB (release.95) and unfiltered NCBI taxonomy names are given in parenthesis. The tree is based on the topology of the

---

Moyo *et al. Microbial Cell Factories*   (2022) 21:43

Page 11 of 13

genomes and not on bootstrap resampling as this is computationally prohibitive and consequently may over-classify genomes relative to manual curation based on unsupported affiliations of user genomes to reference taxa. Part of the GTDB bac120 genome tree is shown. **Figure S2.** Deleted 3 Kb sequence region from the wild type CP5-4 strain's genome to produce a truncation of crtP and expression of the yellow mutant phenotype. The black coloured ORF represents crtP while the red ORF represents the flavodoxin reductase gene. The respective locus tags for the ORFs are also shown. Figure not drawn to scale. **Figure S3.** Deleted 146.691 Kb sequence region from the wild type CP5-4 strain's genome to produce the unpigmented mutant phenotype. The red ORFs represent the predicted carotenoid biosynthetic gene cluster while the green ORFs represent the additional genes that were also deleted as a result of DNA repair following MMS mutagenesis. The position of two insertion sequences ISBsp5 and ISBce1 (Table S4) that may have mediated the deletion are shown. Figure not drawn to scale.

Authors' contributions
ACM, LJZ and MT conceptualized and designed this study. ACM conducted the experimental work. ACM, LD, LJZ and DG analyzed and interpreted the LC-MS data. ACM and LJZ wrote the manuscript. MT Supervision, project administration and funding acquisition. All authors edited and approved the final manuscript.

Declarations

Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Competing interests
The authors declare that they have no competing interests.

Author details
[1]Institute for Microbial Biotechnology and Metagenomics (IMBM), Department of Biotechnology, University of the Western Cape, Bellville, Cape Town 7535, South Africa. [2]BioCiTi Laboratory, 4th Floor Block B, Bandwidth Barn, Woodstock Exchange Building, 66-68 Albert Road, Woodstock, Cape Town 7925, South Africa. [3]Chemistry and Biotechnology of Natural Products, CHEMBIOPRO, ESIROI Agroalimentaire, Université de La Réunion, 15 Avenue René Cassin, CS 92003, CEDEX 9, F-97744 Saint-Denis, France. [4]Università Degli Studi Di Messina, Dip. B.I.O.M.O.R.F, Polo Annunziata, 98168 Messina, ME, Italy.

References
1. Styczynski M, Rogowska A, Gieczewska K, Garstka M, Szakiel A, Dziewit L. Genome-based insights into the production of carotenoids by antarctic bacteria, *Planococcus* sp. ANT_H30 and *Rhodococcus* sp. ANT_H53B. Molecules. 2020. https://doi.org/10.3390/molecules25194357.
2. Nagarajan J, Ramanan RN, Raghunandan ME, Galanakis CM, Krishnamurthy NP. Carotenoids. In: Nagarajan J, editor. Nutraceutical and functional food components Effects of innovative processing techniques. Amsterdam: Elsevier Inc.; 2017. p. 256–96.
3. Seel W, Baust D, Sons D, Albers M, Etzbach L, Fuss J, et al. Carotenoids are used as regulators for membrane fluidity by *Staphylococcus xylosus*. Sci Rep. 2020. https://doi.org/10.1038/s41598-019-57006-5.
4. Domonkos I, Kis M, Gombos Z, Ughy B. Carotenoids, versatile components of oxygenic photosynthesis. Prog Lipid Res. 2013;52(4):539–61. https://www.sciencedirect.com/science/article/pii/S0163782713000453. Accessed 2 Feb 2021.
5. Liu GY, Essex A, Buchanan JT, Datta V, Hoffman HM, Bastian JF, et al. *Staphylococcus aureus* golden pigment impairs neutrophil killing and promotes virulence through its antioxidant activity. J Exp Med. 2005;202(2):209–15. https://doi.org/10.1084/jem.20050846.
6. Saviola B. Pigments of pathogenic bacteria. J Microbiol Exp. 2018;6(2):114–5.
7. Dufossé L, Galaup P, Yaron A, Arad SM, Blanc P, Murthy KNC, et al. Microorganisms and microalgae as sources of pigments for food use: A scientific oddity or an industrial reality? Trends Food Sci Technol. 2005. https://doi.org/10.1016/j.tifs.2005.02.006.
8. Higuera-Ciapara I, Félix-Valenzuela L, Goycoolea FM. Astaxanthin: a review of its chemistry and applications. Crit Rev Food Sci Nutr. 2006;46(2):185–96. https://doi.org/10.1080/10408690590957188.
9. Lee JH, Kim JW, Lee PC. Genome mining reveals two missing CrtP and AldH enzymes in the C_{30} carotenoid biosynthesis pathway in *Planococcus faecalis* AJ003(T). Molecules. 2020. 25(24):5892. https://pubmed.ncbi.nlm.nih.gov/33322786. Accessed 24 Feb 2021.
10 Aruldass CA, Dufossé L, Ahmad WA. Current perspective of yellowish-orange pigments from microorganisms- a review. J Clean Prod. 2018. https://doi.org/10.1016/j.jclepro.2018.01.093.
11. Vila E, Hornero-Méndez D, Azziz G, Lareo C, Saravia V. Carotenoids from heterotrophic bacteria isolated from Fildes Peninsula, King George Island, Antarctica. Biotechnol Rep. 2019. 21: e00306. https://www.sciencedirect.com/science/article/pii/S2215017X18302807. Accessed 16 Apr 2019.
12. Ganapathy A, Jayavel S, Natesan S. Draft genome sequence of carotenoid producing yellow pigmented *Planococcus maritimus* MKU009. J Genomics. 2016. https://doi.org/10.7150/jgen.15533.
13. Shindo K, Misawa N. New and rare carotenoids isolated from marine bacteria and their antioxidant activities. Mar Drugs. 2014;12(3):1690–8.
14. Kasai H, Misawa N, Miyake Y, Endo M, Bramley PM, Shindo K, et al. Methyl glucosyl-3,4-dehydro-apo-8′-lycopenoate (I), a novel antioxidative glyco-C_{30}-carotenoic acid produced by a marine bacterium *Planococcus maritimus*. ChemInform. 2009;40(29):729–35.
15 Ram S, Mitra M, Shah F, Tirkey SR, Mishra S. Bacteria as an alternate biofactory for carotenoid production: A review of its applications, opportunities and challenges. J Funct Foods. 2020. https://doi.org/10.1016/j.jff.2020.103867.
16 Kim JW, Choi BH, Kim JH, Kang HJ, Ryu H, Lee PC. Complete genome sequence of *Planococcus faecalis* AJ003T, the type species of the genus *Planococcus* and a microbial C_{30} carotenoid producer. J Biotechnol. 2018. https://doi.org/10.1016/j.jbiotec.2017.12.005.
17. Zhang B, Yang R, Zhang G, Liu Y, Zhang D, Zhang W, et al. Characteristics of *Planococcus antioxidans* sp. nov., an antioxidant-producing strain isolated from the desert soil in the Qinghai-Tibetan Plateau. Microbiol Open. 2020;9(6):1183–96.
18 Atam SI, Singh L, Dube S, Reddy GSN, Shivaji S. Psychrophilic *Planococcus maitriensis* sp. nov. from Antarctica. Syst Appl Microbiol. 2003. https://doi.org/10.1078/072320203770865792.
19 Shivaji S, Srinivas TNR, Reddy GSN. The family *Planococcaceae*. In: Rosenberg E, DeLong EF, Lory S, Stackebrandt E, Thompson F, editors. the prokaryotes: firmicutes and tenericutes. Berlin: Springer; 2014.
20. Shivaji S. Planococcus. In: Whitman WB, Rainey F, Kämpfer P, Trujillo M, Chun J, DeVos P, Hedlund B, Dedysh S, editors. Bergey's manual of systematics of Archaea and Bacteria. Hoboken: Wiley; 2015. p. 1–11. https://doi.org/10.1002/9781118960608.gbm00561.
21 Pearson MD, Noller HF. The draft genome of *Planococcus donghaensis* MPA1U2 reveals non-sporulation pathways controlled by a conserved Spo0A regulon. J Bacteriol. 2011;193:6106–6106.
22. Khanafari A, Khavarinejad D, Mashinchian A. Solar Salt Lake as natural environmental source for extraction halophilic pigments. Iran J Microbiol. 2010;2(2):103–9.
23 Oren A, Stambler N, Dubinsky Z. On the red coloration of saltern crystallizer ponds. Int J Salt Lake Res. 1992. https://doi.org/10.1007/BF02904363.
24 Oren A, Rodriguez-Valera F. The contribution of halophilic bacteria to the red coloration of saltern crystallizer ponds. FEMS Microbiol Ecol. 2001. https://doi.org/10.1016/S0168-6496(01)00124-6.

Moyo *et al. Microbial Cell Factories*   (2022) 21:43

Page 12 of 13

25. Keyes CG, Gunaji N. Effect of dye on solar evaporation of brine. New Mexico State University; 1966. http://hydrologie.org/redbooks/a078/iahs_078_0338.pdf. Accessed 13 June 2017.

26. Bloch MR, Farkas L, Spiegler KS. Solar evaporation of salt brines. Ind Eng Chem. 1951;43(7):1544–53. https://doi.org/10.1021/ie50499a025.

27. Winans DC. The relative stability of six dyes in a saline brine of constant salinity. 1967.

28. Solar evaporation of alkaline trona-process waste pond liquors with acid black 2 and acid green 1. 1976 Nov 18.

29. Potash Evaporation Ponds—Moab, Utah - Atlas Obscura. https://www.atlasobscura.com/places/potash-evaporation-ponds. Accessed 10 Oct 2021

30. Takemura M, Takagi C, Aikawa M, Araki K, Choi S-K, Itaya M, et al. Heterologous production of novel and rare C$_{30}$-carotenoids using *Planococcus* carotenoid biosynthesis genes. Microb Cell Fact. 2021;20(1):194. https://doi.org/10.1186/s12934-021-01683-3.

31. Lee JH, Kim JW, Lee PC. Genome mining reveals two Missing CrtP and AldH enzymes in the C$_{30}$ carotenoid biosynthesis pathway in *Planococcus faecalis* AJ003T. Molecules. 2020;25(24):5892.

32. See-Too WS, Chua KO, Lim YL, Chen JW, Convey P, Mohd Mohidin TB, et al. Complete genome sequence of *Planococcus donghaensis* JH1T, a pectin-degrading bacterium. J Biotechnol. 2017;20(252):11–4.

33. Gan L, Li X, Tian Y, Peng B. Genomic insights into the salt tolerance and cold adaptation of *Planococcus halotolerans* SCU63T. Arch Microbiol. 2020;202(10):2841–7. https://doi.org/10.1007/s00203-020-01979-9.

34. See-Too WS, Convey P, Pearce DA, Lim YL, Ee R, Yin WF, et al. Complete genome of *Planococcus rifietoensis* M8T, a halotolerant and potentially plant growth promoting bacterium. J Biotechnol. 2016;10(221):114–5.

35. Pelz A, Wieland KP, Putzbach K, Hentschel P, Albert K, Götz F. Structure and biosynthesis of staphyloxanthin from *Staphylococcus aureus*. J Biol Chem. 2005;280(37):32493–8. http://www.jbc.org/cgi/content/short/280/37/32493. Accessed 30 Jan 2020.

36. Marshal JH, Wilmoth GJ. Pigments of *Staphylococcus aureus*, a series of triterpenoid carotenoids. J Bacteriol. 1981;147(3):900–13.

37. Marshall JH, Wilmoth GJ. Proposed pathway of triterpenoid carotenoid biosynthesis in *Staphylococcus aureus* - evidence from a study of mutants. J Bacteriol. 1981;147:914–9.

38. Kim SH, Lee PC. Functional expression and extension of *Staphylococcal staphyloxanthin* biosynthetic pathway in *Escherichia coli*. J Biol Chem. 2012. https://doi.org/10.1074/jbc.M112.343020.

39. Shindo K, Endo M, Miyake Y, Wakasugi K, Morritt D, Bramley PM, et al. Methyl glucosyl-3,4-dehydro-apo-8′-lycopenoate, a novel antioxidative glyco-C$_{30}$-carotenoic acid produced by a marine bacterium *Planococcus maritimus*. J Antibiot (Tokyo). 2008;61(12):729–35.

40. Shindo K, Endo M, Miyake Y, Wakasugi K, Morritt D, Bramley PM, et al. Erratum: Methyl 5-glucosyl-5,6-dihydro-apo-4,4′-lycopenoate, a novel antioxidative glyco-C$_{30}$-carotenoic acid produced by a marine bacterium *Planococcus maritimus*. J Antibiotics. 2008; 61:729–735. https://doi.org/10.1038/ja.2008.86. Vol. 67, J Antibiotics. 2014. 67;731–2. www.nature.com/ja. Accessed 27 Mar 2019.

41. Kim JH, Kang HJ, Yu BJ, Kim SC, Lee PC. *Planococcus faecalis* sp. nov., a carotenoid-producing species isolated from stools of Antarctic penguins. Int J Syst Evol Microbiol. 2015;65(10):3373–8. https://doi.org/10.1099/ijsem.0.000423.

42. Tao L, Schenzle A, Odom JM, Cheng Q. Novel carotenoid oxidase involved in biosynthesis of 4,4′- diapolycopene dialdehyde. Appl Environ Microbiol. 2005;71(6):3294–301.

43. Köcher S, Breitenbach J, Müller V, Sandmann G. Structure, function and biosynthesis of carotenoids in the moderately halophilic bacterium *Halobacillus halophilus*. Arch Microbiol. 2009;191(2):95–104.

44. Perez-Fons L, Steiger S, Khaneja R, Bramley PM, Cutting SM, Sandmann G, et al. Identification and the developmental formation of carotenoid pigments in the yellow/orange Bacillus spore-formers. Biochim Biophys Acta - Mol Cell Biol Lipids. 2011;1811(3):177–85. https://doi.org/10.1016/j.bbalip.2010.12.009.

45. Steiger S, Perez-Fons L, Cutting SM, Fraser PD, Sandmann G. Annotation and functional assignment of the genes for the C$_{30}$ carotenoid pathways from the genomes of two bacteria: *Bacillus indicus* and *Bacillus firmus*. Microbiology (United Kingdom). 2015;161(1):194–202.

46. Steiger S, Perez-Fons L, Fraser PD, Sandmann G. Biosynthesis of a novel C*30* carotenoid in *Bacillus firmus* isolates. J Appl Microbiol. 2012;113(4):888–95.

47. Amorim-Carrilho KT, Cepeda A, Fente C, Regal P. Review of methods for analysis of carotenoids. TrAC Trend Anal Chem. 2014. https://doi.org/10.1016/j.trac.2013.12.011.

48. Mariutti LRB, Mercadante AZ. Carotenoid esters analysis and occurrence: What do we know so far? Arch Biochem Biophys. 2018;648:36–43. https://doi.org/10.1016/j.abb.2018.04.005.

49. Manzoni Maroneze M, Jacob-Lopes E, Queiroz Zepka L, Roca M, Pérez-Gálvez A. Esterified carotenoids as new food components in Cyanobacteria. Food Chem. 2019. https://doi.org/10.1016/j.foodchem.2019.02.102.

50. Osawa A, Ishii Y, Sasamura N, Morita M, Köcher S, Müller V, et al. Erratum: 5-Hydroxy-5,6-dihydro-apo-4,4′-lycopene and methyl 5-hydroxy-5,6-dihydro-apo-4,4′-lycopenoate, novel C$_{30}$-carotenoids produced by a mutant of marine bacterium *Halobacillus halophilus*. J Antibiot. 2010;63:291–5. https://doi.org/10.1038/ja.201.

51. Osawa A, Ishii Y, Sasamura N, Morita M, Köcher S, Müller V, et al. Hydroxy-3,4-dehydro-apo-8′-lycopene and methyl hydroxy-3,4-dehydro- apo-8′-lycopenoate, novel C$_{30}$ carotenoids produced by a mutant of marine bacterium *Halobacillus halophilus*. J Antibiot (Tokyo). 2010;63(6):291–5.

52. Shortridge MD, Powers R. Structural and functional similarity between the bacterial type III secretion system needle protein PrgI and the eukaryotic apoptosis Bcl-2 proteins. PLoS ONE. 2009. https://doi.org/10.1371/journal.pone.0007442.

53. Todd AE, Orengo CA, Thornton JM. Evolution of function in protein superfamilies, from a structural perspective. J Mol Biol. 2001;307(4):1113–43.

54. Addou S, Rentzsch R, Lee D, Orengo CA. Domain-based and family-specific sequence identity thresholds increase the levels of reliable protein function transfer. J Mol Biol. 2009;387(2):416–30. http://www.sciencedirect.com/science/article/pii/S0022283608015660. Accessed 23 Sept 2021.

55. Johnston CW, Skinnider MA, Wyatt MA, Li X, Ranieri MRM, Yang L, et al. An automated Genomes-to-Natural Products platform (GNP) for the discovery of modular natural products. Nat Commun. 2015;6(1):1–11.

56. Navarro-Muñoz JC, Selem-Mojica N, Mullowney MW, Kautsar SA, Tryon JH, Parkinson EI, et al. A computational framework to explore large-scale biosynthetic diversity. Nat Chem Biol. 2019. 16(1):60–8: https://pubmed.ncbi.nlm.nih.gov/31768033. Accessed 16 Jan 2020.

57. Armstrong GA. Genetics of eubacterial carotenoid biosynthesis: a colorful tale. Annu Rev Microbiol. 1997;51(1):629–59. https://doi.org/10.1146/annurev.micro.51.1.629.

58. Meléndez-Martínez AJ, Mapelli-Brahm P, Benítez-González A, Stinco CM. A comprehensive review on the colorless carotenoids phytoene and phytofluene. Arch Biochem Biophys. 2015;572:188–200.

59. Machmudah S, Goto M. Methods for extraction and analysis of carotenoids. In: Ramawat KG, Mérillon J-M, editors. Natural products: phytochemistry, botany and metabolism of alkaloids, phenolics and terpenes. Berlin: Springer; 2013. p. 3367–411.

60. Steiger S, Perez-Fons L, Fraser PD, Sandmann G. The biosynthetic pathway to a novel derivative of 4, 4′-diapolycopene-4, 4′-oate in a red strain of *Sporosarcina aquimarina*. Arch Microbiol. 2012;194(9):779–84.

61. Schujman GE, Grau R, Gramajo HC, Ornella L, De Mendoza D. *De novo* fatty acid synthesis is required for establishment of cell type-specific gene transcription during sporulation in *Bacillus subtilis*. Mol Microbiol. 1998. https://doi.org/10.1046/j.1365-2958.1998.01004.x.

62. Schujman GE, Guerin M, Buschiazzo A, Schaeffer F, Llarrull LI, Reh G, et al. Structural basis of lipid biosynthesis regulation in Gram-positive bacteria. EMBO J. 2006. https://doi.org/10.1038/sj.emboj.7601284.

63. de Mendoza D, Schujman GE, Aguilar PS. Biosynthesis and function of membrane lipids. In: Hoch JA, Losick R, editors. Bacillus subtilis and its closest relatives. Washington DC: ASM Press; 2014.

64. De Mendoza D, Schujman GE. Lipid biosynthesis. In: Encyclopedia of Microbiology. Elsevier; 2019. p. 780–90. https://doi.org/10.1016/B978-0-12-801238-3.02444-2.

65. Park Y, Ledesma-Amaro R, Nicaud J-M. *De novo* biosynthesis of odd-chain fatty acids in *Yarrowia lipolytica* enabled by modular pathway engineering. Front Bioeng Biotechnol. 2020;7:484. https://doi.org/10.3389/fbioe.2019.00484.

Moyo *et al. Microbial Cell Factories*   (2022) 21:43

Page 13 of 13

66. Froese DS, Dobson CM, White AP, Wu X, Padovani D, Banerjee R, et al. Sleeping beauty mutase (sbm) is expressed and interacts with ygfd in *Escherichia coli*. Microbiol Res. 2009;164(1):1–8. https://www.sciencedirect.com/science/article/pii/S0944501308000554. Accessed 23 Sept 2021.

67. Pérez-Fons L, Fraser PD. Analysis of diapocarotenoids found in pigmented *Bacillus* species. Methods Mol Biol. 2012;892:335–45.

68. Sy C, Gleize B, Chamot S, Dangles O, Carlin F, Caris C, et al. Glycosyl carotenoids from marine spore-forming *Bacillus* sp. strains are readily bioaccessible and bioavailable. Food Res Int. 2013;51(2):914–23. https://doi.org/10.1016/j.foodres.2013.02.003.

69. Silva-Castro GA, Moyo AC, Khumalo L, van Zyl LJ, Petrik LF, Trindade M. Factors influencing pigment production by halophilic bacteria and its effect on brine evaporation rates. Microb Biotechnol. 2019;12(2):334–45.

70. Wang Y, Zhang Z, Ruan J. A Proposal to transfer *Microbispora bispora* (Lechevalier 1965) to a new genus, *Thermobispora* gen. nov., as *Thermobispora bispora* comb. nov. Int J Syst Bacteriol. 1996;46(4):933–8.

71. Geneious 11.1.5. https://www.geneious.com. Accessed 14 Aug 2019.

72. Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, et al. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics.. 2012. 28(12):1647–9. https://www.ncbi.nlm.nih.gov/pubmed/22543367. Accessed 27 Apr 2012.

73 Chaumeil P-A, Mussig AJ, Hugenholtz P, Parks DH. GTDB-Tk: a toolkit to classify genomes with the Genome Taxonomy Database. Bioinformatics. 2019. https://doi.org/10.1093/bioinformatics/btz848.

74. http://kbase.us. Accessed 13 Oct 2020.

75. Arkin AP, Cottingham RW, Henry CS, Harris NL, Stevens RL, Maslov S, et al. KBase: The United States department of energy systems biology knowledgebase. Nat Biotechnol. 2018. https://doi.org/10.1038/nbt.4163.

76 Federhen S. The NCBI Taxonomy database. 2012. Nucleic Acids Res. https://doi.org/10.1093/nar/gkr1178.

77. Parks DH, Chuvochina M, Waite DW, Rinke C, Skarshewski A, Chaumeil PA, et al. A standardized bacterial taxonomy based on genome phylogeny substantially revises the tree of life. Nat Biotechnol. 2018. https://doi.org/10.1038/nbt.4229.

78 Parks DH, Chuvochina M, Chaumeil PA, Rinke C, Mussig AJ, Hugenholtz P. A complete domain-to-species taxonomy for Bacteria and Archaea. Nat Biotechnol. 2020. https://doi.org/10.1038/s41587-020-0501-8.

79. Conesa A, Götz S, García-Gómez JM, Terol J, Talón M, Robles M. Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. Bioinformatics. 2005;21(18):3674–6.

80. Götz S, García-Gómez JM, Terol J, Williams TD, Nagaraj SH, Nueda MJ, et al. High-throughput functional annotation and data mining with the Blast2GO suite. Nucleic Acids Res. 2008;36(10):3420–35.

81 https://www.uniprot.org/uniprot/. Accessed 24 May 2020.

82 Boutet E, Lieberherr D, Tognolli M, Schneider M, Bairoch A. Uniprotkb/swiss-prot. In: Edwards David, editor. Plant bioinformatics. Berlin: Springer; 2007. p. 89–112.

83 Boutet E, Lieberherr D, Tognolli M, Schneider M, Bansal P, Bridge AJ, et al. UniProtKB/Swiss-Prot, the manually annotated section of the UniProt KnowledgeBase: how to use the entry view. In: Edwards David, editor., et al., Plant bioinformatics. New York: Springer; 2016. p. 23–54.

84 Blin K, Shaw S, Steinke K, Villebro R, Ziemert N, Lee SY, et al. antiSMASH 5.0: updates to the secondary metabolite genome mining pipeline. Nucleic Acids Res. 2019;47(W1):W81-7. https://doi.org/10.1093/nar/gkz310.

85. Chanson A, Moreau CS, Duplais C. Assessing biosynthetic gene cluster diversity in a multipartite nutritional symbiosis between herbivorous turtle ants and conserved gut symbionts. bioRxiv. 2020. http://biorxiv.org/content/early/2020/05/07/2020.05.05.072835.abstract. Accessed 5 May 2020.

86. Doreau M, Lyan B, Cardinault N, Chauveau B, Rock E, Grolier P. Development of a method to determine carotenoid composition of fresh forages. Can J Plant Sci. 2011;88(6):1057–64.

Publisher's Note