Microbiology (2015), **161**, 194–202 **DOI** 10.1099/mic.0.083519-0

Annotation and functional assignment of the genes for the C_{30} carotenoid pathways from the genomes of two bacteria: *Bacillus indicus* and *Bacillus firmus*

Sabine Steiger,[1] Laura Perez-Fons,[2] Simon M. Cutting,[2] Paul D. Fraser[2] and Gerhard Sandmann[1]

Correspondence
Gerhard Sandmann
sandmann@bio.uni-frankfurt.de

[1]Biosynthesis Group, Molecular Biosciences, Goethe University Frankfurt, Frankfurt, Germany
[2]School of Biological Sciences, Royal Holloway, University of London, Egham, Surrey, UK

Bacillus indicus and *Bacillus firmus* synthesize C_{30} carotenoids via farnesyl pyrophosphate, forming apophytoene as the first committed step in the pathway. The products of the pathways were methyl 4′-[6-*O*-acyl-glycosyl)oxy]-4,4′-diapolycopen-4-oic acid and 4,4′-diapolycopen-4,4′-dioic acid with putative glycosyl esters. The genomes of both bacteria were sequenced, and the genes for their early terpenoid and specific carotenoid pathways annotated. All genes for a functional 1-deoxy-D-xylulose 5-phosphate synthase pathway were identified in both species, whereas genes of the mevalonate pathway were absent. The genes for specific carotenoid synthesis and conversion were found on gene clusters which were organized differently in the two species. The genes involved in the formation of the carotenoid cores were assigned by functional complementation in *Escherichia coli*. This bacterium was co-transformed with a plasmid mediating the formation of the putative substrate and a second plasmid with the gene of interest. Carotenoid products in the transformants were determined by HPLC. Using this approach, we identified the genes for a 4,4′-diapophytoene synthase (*crtM*), 4,4′-diapophytoene desaturase (*crtNa*), 4,4′-diapolycopene ketolase (*crtNb)* and 4,4′-diapolycopene aldehyde oxidase (*crtNc*). The three *crtN* genes were closely related and belonged to the *crtI* gene family with a similar reaction mechanism of their enzyme products. Additional genes encoding glycosyltransferases and acyltransferases for the modification of the carotenoid skeleton of the diapolycopenoic acids were identified by comparison with the corresponding genes from other bacteria.

Received 15 August 2014
Accepted 15 October 2014

INTRODUCTION

Carotenoids are coloured terpenoids. Their chromophore consists of a conjugated double-bond system. Typically, carotenoids originate from the condensation of two molecules of geranylgeranyl pyrophosphate to yield a C_{40} carbon skeleton. However, amongst some bacteria carotenoid biosynthesis may utilize two molecules of C_{15} farnesyl pyrophosphate (FPP). This reaction yields 4,4′-diapophytoene (7,8,11,12,7′,8′,11′,12′-octyhydro-4,4′-diapo-ψ,ψ-carotene) and is the starting point for a variety of acyclic C_{30} carotenoids (Taylor, 1984). After a series of desaturation steps, the resulting 4,4′-diaponeurosporene (7,8-dihydro-4,4′-diapo-ψ,ψ-carotene) and the fully desaturated 4,4′-diapolycopene (4,4′-diapo-ψ,ψ-carotene) are end-group modified. These C_{30} carotenoid derivatives were found in unrelated species, such as *Methylobacterium rhodium* (formerly *Pseudomonas rhodos*) (Kleinig *et al.*, 1979), *Methylomonas* sp. (Tao *et al.*, 2005), *Staphylococcus aureus* (Pelz *et al.*, 2005) and *Rubritalea squalenifaciens* (Shindo *et al.*, 2007). In addition, C_{30} carotenoids are group specific, as in the case of heliobacteria (Takaichi *et al.*, 2003) and pigmented *Bacillales*. Bacteria from the latter group with a well-established C_{30} carotenoid pathway include *Planococcus maritimus* (Shindo *et al.*, 2008), *Halobacillus halophilus* (Osawa *et al.*, 2010) and *Sporosarcina aquimarina* (Steiger *et al.*, 2012a).

A series of differently pigmented bacilli have been isolated and their pigments tentatively assigned as carotenoids (Khaneja *et al.*, 2010). A C_{30} carotenoid biosynthesis

Abbreviations: AT, acyltransferase; DMPP, dimethylallyl pyrophosphate; DXS, 1-deoxy-D-xylulose 5-phosphate synthase; FPP, farnesyl pyrophosphate; GT, glycosyltransferase; IPP, isopentenyl pyrophosphate; MVA, mevalonate.

Three supplementary tables are available with the online Supplementary Material.

pathway has been established for two *Bacillus* species. In *Bacillus indicus*, the major carotenoid is methyl 4′-[6-*O*-acyl-glycosyl)oxy]-4,4′-diapolycopen-4-oic acid (Perez-Fons *et al.*, 2011), whereas in *Bacillus firmus*, 4,4′-diapolycopen-4,4′-dioic acid with putative glycosyl esters accumulate (Steiger *et al.*, 2012b; Osawa *et al.*, 2013), all sharing 4,4′-diapolycopen-4-oic acid (4,4′-diapo-ψ,ψ-caroten-4-oic acid) as a precursor. In some pigmented bacilli, formation of the end products of the carotenoid pathway is associated with spore formation (Duc *et al.*, 2006; Perez-Fons *et al.*, 2011). In contrast to *S. aureus*, in which the genes for the 4,4′-diaponeurosporene-derived biosynthesis of staphyloxanthin were identified and functionally assigned (Pelz *et al.*, 2005; Kim & Lee, 2012), the genes for the oxidative C_{30} pathway via 4,4′-diapolycopene are unknown. Therefore, we cloned these genes from both *Bacillus* species starting from their genome sequences. In most bacteria, terpenoids are synthesized via the 1-deoxy-D-xylulose 5-phosphate synthase (DXS) pathway (Wilding *et al.*, 2000). However, the mevalonate (MVA) pathway exists as an alternative route amongst some Gram-positive bacteria. In our genome sequences, we searched for genes of either the DXS or MVA pathway and identified a carotenogenic gene cluster in the genomes of both bacilli. These genes encoding biosynthetic enzymes for C_{30} carotenoid formation were annotated and their function identified by genetic complementation in *Escherichia coli*.

METHODS

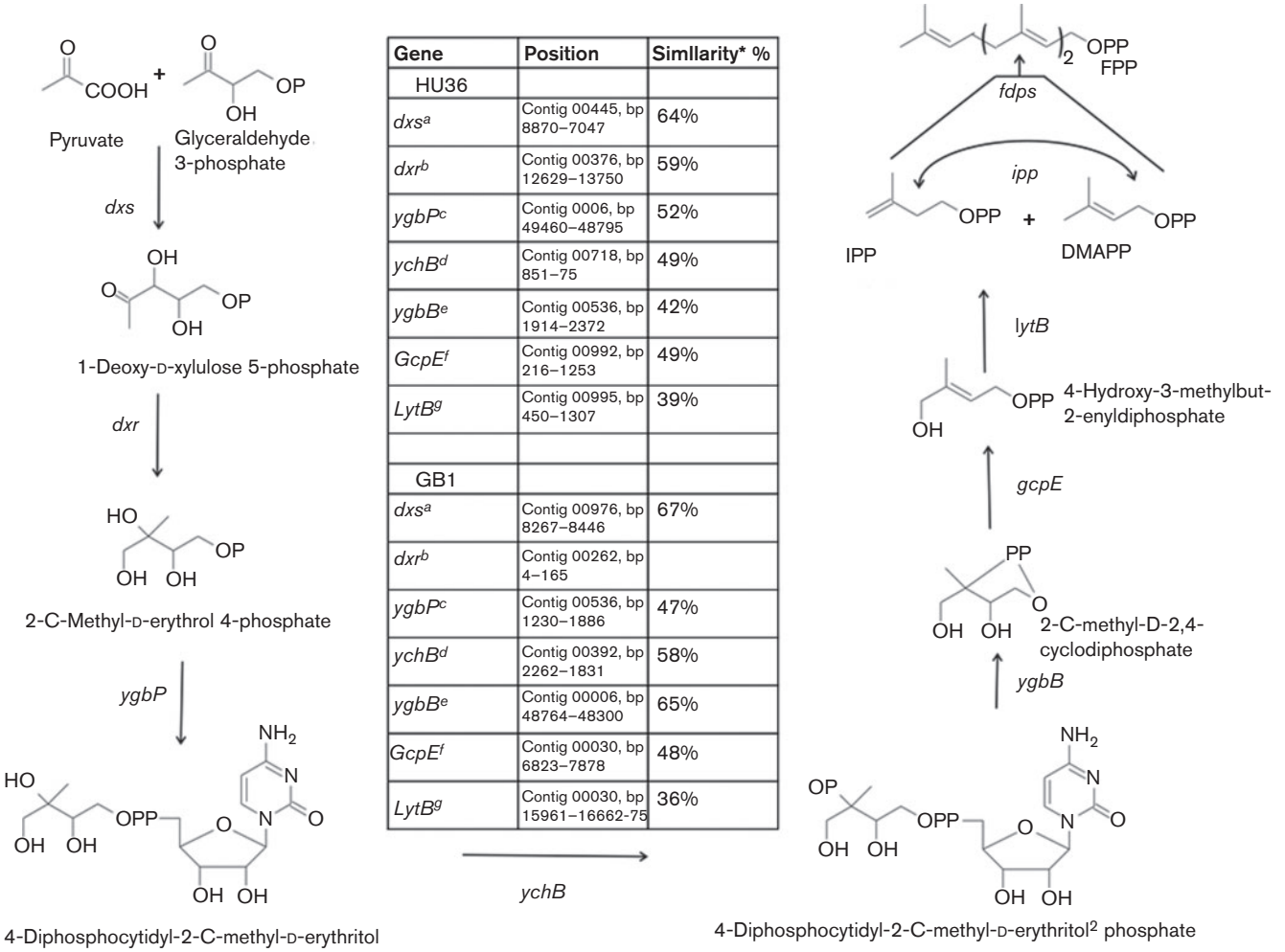
Cultivation, DNA isolation, plasmid construction and genetic complementation. *E. coli* strains DH5α and JM101 were used for cloning and genetic complementation of carotenogenic reactions. They were cultivated in LB medium with appropriate antibiotics according to Sambrook *et al.* (1989). *B. indicus* HU36 and *B. firmus* GB1 have been described previously (Khaneja *et al.*, 2010). These were grown in tryptone/yeast media in shake-culture for 48 h at 30 °C. DNA was isolated from both *Bacillus* strains using the Qiagen genomic DNA isolation kit. Expression plasmids were constructed by PCR amplification of *crt* genes from genomic DNA. The *crtNc* genes from both *Bacillus* species were codon optimized for *E. coli* synthesized by MWG Operon Eurofins and provided already inserted into the *Sal*I/*Hin*dIII sites of plasmid pEX-K. Out of this plasmid, they were both cloned into the *Sal*I/*Hin*dIII sites of pUC8-2 (Hanna *et al.*, 1984). The details are compiled in Table S1 (available in the online Supplementary Material) (Hanna *et al.*, 1984; Borovkov & Rivkin, 1997). Plasmids used for diapocarotenoid background formation in *E. coli* were pACCRT-M containing the gene for diapophytoene synthase, pACCRT-MN with the additional diapophytoene desaturase gene (Raisig & Sandmann, 1999), both from *S. aureus*, and pBBR-Nb with the *crtNb* gene from *Methylomonas* sp. 16a (Tao *et al.*, 2005). This gene was amplified by PCR with the primers *Bam*HI-forward (5′-CAGG-ATCCAATGAACTCAAATGACAACCA-3′) and *Sac*I-reverse (5′-GT-GAGCTCTTATTGCAAATCCGCCAC-3′). These restriction sites were used to clone *crtNb* into pBBR1-MCS2 (Kovach *et al.*, 1995).

Sequencing, assembly and annotation. Initiated by the authors in a joint European Union project, genome sequencing of both strains by 454 sequencing and assembly provided the sequences of the *B. indicus* HU36 and *B. firmus* GB1 genomes on 756 and 1175 contigs, respectively (Manzo *et al.*, 2011). Both genomes are available at http://www1.uni-frankfurt.de/fb/fb15/english/institute/inst-3-mol-biowiss/AK-Sandmann/bacillus/index.html together with the resequenced and corrected *crt* genes (see also Tables S2 and S3). Both genomes were analysed for the presence of genes involved in the terpenoid pathway finally leading to carotenoid biosynthesis using TBLAST_N (www.ncbi.nlm.nih.gov/blast/) (Altschul *et al.*, 1997) searches against the *E. coli* DXS pathway and *H. halophilus* and *Methylobacter* sp. carotenogenic genes, and alignment of the protein sequences using BioEdit Sequence Alignment 7.0.9.0 (Hall, 1999). The best gene models were selected and annotated. Reading frame analysis was carried out with the Clone Manager program. Phylogenetic analysis of amino acid sequences was performed with the program CLUSTAL_X (Thompson *et al.*, 1994) and the alignments visualized with TreeView. Completion of nucleotide sequences, corrections and filling of gaps was carried out by resequencing of the *crt* gene clusters.

Carotenoid extraction and analysis. Freeze-dried *E. coli* cells were extracted with methanol for 20 min at 65 °C and re-extracted with acetone. After partitioning of the combined extracts into 50 % diethyl ether in petroleum ether, the upper phase with the carotenoids was collected and evaporated to dryness. Carotenoids were resuspended in acetone or methanol/dichloromethane (50 : 50) with acetic acid (2 mM final concentration) directly before HPLC analysis on a 15 (column 1) or 25 cm (column 2) Nucleosil C18, 3 µm column at 10 °C with a flow rate of 0.8 ml min^{-1}. Depending on the polarity of the carotenoids, elution was with either acetonitrile/methanol/2-propanol (85 : 10 : 5, by vol.) (mobile phase 1) or the same mobile phase plus 3 % (v/v) water and acetic acid to a concentration of 2 mM (mobile phase 2), the latter being used especially for the samples containing carotenoid acids. The added acetic acid prevented the dissociation of the carboxyl group. Carotenoids were identified with individual standards and by their specific absorbance spectra (Britton *et al.*, 2004) recorded online with a Kontron DAD 440 diode array detector. Standards were generated in *E. coli* by the combination of different *crt* genes as described previously (Steiger *et al.*, 2012b).

RESULTS

In non-carotenogenic *B. subtilis*, terpenoid biosynthesis starts with the DXS pathway (Wagner *et al.*, 2000). As *B. indicus* and *B. firmus* synthesize C_{30} carotenoids derived from FPP, we annotated the DXS pathway genes of all reactions leading to FPP formation (Fig. 1). The genes of the complete DXS pathway to the formation of isopentenylpyrophosphate (IPP) and dimethylallyl pyrophosphate (DMAPP) were present in both genomes. In addition, we identified the genes *ipp* encoding an IPP isomerase and *fdps* encoding an FPP synthase. These genes exhibited identities of ~50 % with the corresponding *E. coli* genes. In comparison with the corresponding genes from *H. halophilus*, identity values were higher, reaching up to 87 % identity. Their positions in the genome sequences are indicated in Fig. 1. It should be noted that the *dxr* gene was incomplete in our *B. firmus* genome sequence; nevertheless, the fragment lacking ~300 aa was sufficient for its definite identification. We also looked for genes from the MVA pathway, which provide the terpenoid precursors in other Gram-positive bacteria (Wilding *et al.*, 2000), using the genes from another bacterium synthesizing C_{30} carotenoids, *S. aureus*, for comparison. However, no MVA pathway genes could be found in the genomes of *B. indicus* or *B. firmus*.



Gene	Position	Simllarity* %
HU36		
dxs[a]	Contig 00445, bp 8870–7047	64%
dxr[b]	Contig 00376, bp 12629–13750	59%
ygbP[c]	Contig 0006, bp 49460–48795	52%
ychB[d]	Contig 00718, bp 851–75	49%
ygbB[e]	Contig 00536, bp 1914–2372	42%
GcpE[f]	Contig 00992, bp 216–1253	49%
LytB[g]	Contig 00995, bp 450–1307	39%
GB1		
dxs[a]	Contig 00976, bp 8267–8446	67%
dxr[b]	Contig 00262, bp 4–165	
ygbP[c]	Contig 00536, bp 1230–1886	47%
ychB[d]	Contig 00392, bp 2262–1831	58%
ygbB[e]	Contig 00006, bp 48764–48300	65%
GcpE[f]	Contig 00030, bp 6823–7878	48%
LytB[g]	Contig 00030, bp 15961–16662-75	36%

Fig. 1. The DXS pathway to IPP and DMAPP and its continuation to FPP in *B. indicus* HU36 and *B. firmus* GB1. Genes are positioned next to the corresponding reactions; the tables indicate their position in the genome sequences. Genes: *dxs*, 1-deoxy-D-xylulose-5-phosphate synthase; *dxr*, 1-deoxy-D-xylulose-5-phosphate isomerase; *ygbP* (*ispD*), 2-C-methyl-D-erythritol-4-phosphate cytidyltransferase; *ychB* (*ispE*), 4-diphosphocytidyl-2-C-methyl-D-erythritol kinase; *ygbB* (*ispF*), 2-C-methyl-D-erythritol-2,4-cyclodiphosphate synthase; *gcpE* (*ispG*), 4-hydroxy-3-methylbut-2-enyldiphosphate synthase; *lytB* (*ispH*), 4-hydroxy-3-methylbut-2-enyldiphosphate reductase; *fdps*, farnesyldiphosphate synthase; *ipp*, isopentenylpyrophosphate isomerase. *The *dxr* gene from *B. firmus* is incomplete.

The *crt* genes were further analysed in all three reading frames by simultaneous comparison to different homologous bacterial sequences. When frameshifts were observed, corrections were made by resequencing. This revealed missing nucleotides or missing short inserts in some of the sequences. The corrected sequences of the *crt* genes from *B. firmus* and *B. indicus* are listed in Tables S2 and S3, respectively. The organization of the genes related to the carotenoid pathway in both bacilli is shown in Fig. 2. The start and stop positions of the putative carotenogenic genes are indicated. All genes necessary to cover the whole C_{30} carotenoid pathway of *B. firmus* are located on one contig in a range of 8100 bp. All their transcription is in the same direction. In *B. indicus*, the *crtNa*, *crtNc* and *crtM* genes are separated by two unknown reading frames from *crtNb* and *AT* (acyltransferase). Both sub-clusters are transcribed in opposite directions. In *B. indicus*, the carotenoid product 4-glycosyl-4′-methyl-4,4′-diapolycopene-4′-oate fatty acid ester (Perez-Fons *et al.*, 2011) requires the catalytic activity of a glycosyltransferase (GT) with the terminal hydroxyl group acting as the sugar acceptor (Lairson *et al.*, 2008). A corresponding *AT* gene is present next to *crtNb*. However, a gene encoding a GT that can link a sugar molecule to a terminal hydroxy group was found isolated on another contig. In the *B. firmus* gene cluster, a *GT* gene and a *CAE* gene encoding a carboxyl esterase, an enzyme which can hydrolyse fatty acid sugar alcohol esters (Hosokawa, 2008), are present. The formation of the

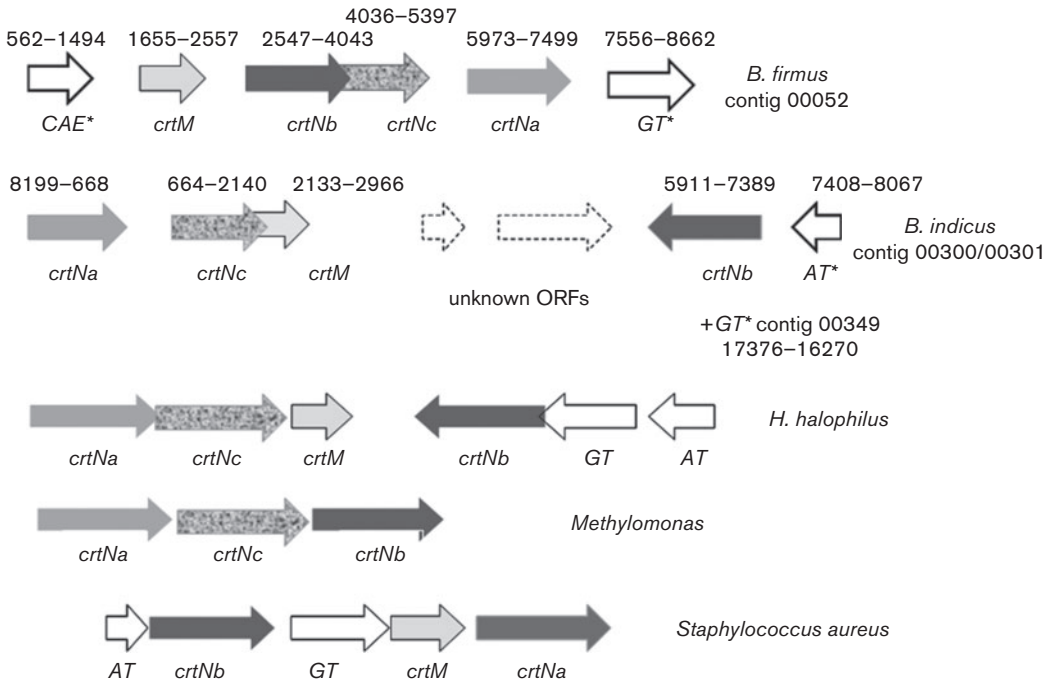


Fig. 2. Organization of the carotenogenic genes from *B. indicus* HU36 and *B. firmus* GB1 in comparison to the gene clusters of other C_{30} carotenoid-synthesizing bacteria: *H. halophilus* (Köcher *et al.*, 2009), *Methylomonas* sp. (Tao *et al.*, 2005) and *S. aureus* (Kim & Lee, 2012). Numbers indicate the nucleotide positions of the beginning and end of the *Bacillus* genes. Final assignment of genes from *B. indicus* HU36 and *B. firmus* was by functional pathway complementation in *E. coli*, except for those marked with an asterisk. Genes: *crtM*, 4,4′-diapophytoene synthase; *crtNa*, 4,4′-diapophytoene desaturase; *crtNb*, 4,4′-diapolycopene ketolase; *crtNc*, 4,4′-diapolycopenealdehyde oxidase; *AT*, acyltransferase; *CAE*, carboxyl esterase; *GT*, glycosyl-transferase.

proposed 4,4′-diapolycopen-4,4′-dioic acid (4,4′-diapo-ψ,ψ-caroten-4,4′-dioic acid) sugar esters of *B. firmus* (Steiger *et al.*, 2012b) may be catalysed either by a GT-type enzyme or a CAE with synthesis rather than hydrolysis function.

The organization of the genes from the carotenoid pathway is completely different between *B. indicus* and *B. firmus*, and compared with *Methylomonas* sp. and *S. aureus*. However, there is a good match between *B. indicus* and *H. halophilus* (Köcher *et al.*, 2009) with regard to gene arrangement and direction of transcription. Phylogenetic analysis comparing diapophytoene synthase genes (*crtM*) to bacterial phytoene synthase genes (*crtB*) showed the evolutionary relationship of both synthase genes, which group at distinct clusters (Fig. 3a). It has been shown previously that the 4,4′-diapophytoene desaturase gene (*crtNa*) belongs to the phytoene desaturase (*crtI*) gene family (Raisig & Sandmann, 2001). In addition to *crtNa* genes, two related clusters of the 4,4′-diapolycopene ketolase genes (*crtNb*) and the 4,4′-diapolycopene aldehyde oxidase genes (*crtNc*) were compared in the phylogenetic tree in Fig. 3(b). The *crtNb* genes formed a distinct group. Out of the *crtNc* genes, three of these also formed an individual clade, but *crtNc* from *B. indicus* and *H. halophilus* were positioned within the *crtNa* group. In all cases, the genes from *B. indicus* were closest to those of *H. halophilus*.

The carotenogenic genes from *B. indicus* and *B. firmus* were putatively assigned by alignment with genes from the C_{30} carotenoid biosynthesis pathways of other bacteria. For final functional identification, we analysed the catalytic activity of proteins related to *crtNa*, *crtNb* and *crtNc* gene candidates from both bacilli by genetic pathway complementation in *E. coli*. In each case, a substrate background was established with a complementation containing all genes for substrate synthesis and co-expression with a plasmid containing the gene of interest. The separation of the resulting carotenoids by HPLC analysis is shown in Fig. 4. Traces A and B showed that the *crtM* genes from *B. indicus* and *B. firmus*, respectively, both synthesized 4,4′-diapophytoene isomers (peaks 1 and 1′) from the precursors that already existed in the non-transformed *E. coli*. The 4,4′-diapophytoene peaks co-chromatographed with the standard in trace C, and exhibited the characteristic absorbance spectrum at 275, 285 and 297 nm. The *crtNa* genes in a 4,4′-diapophytoene background (generated by plasmid pACCRT-M) were responsible for the formation of two isomers of 4,4′-diapolycopene (traces D

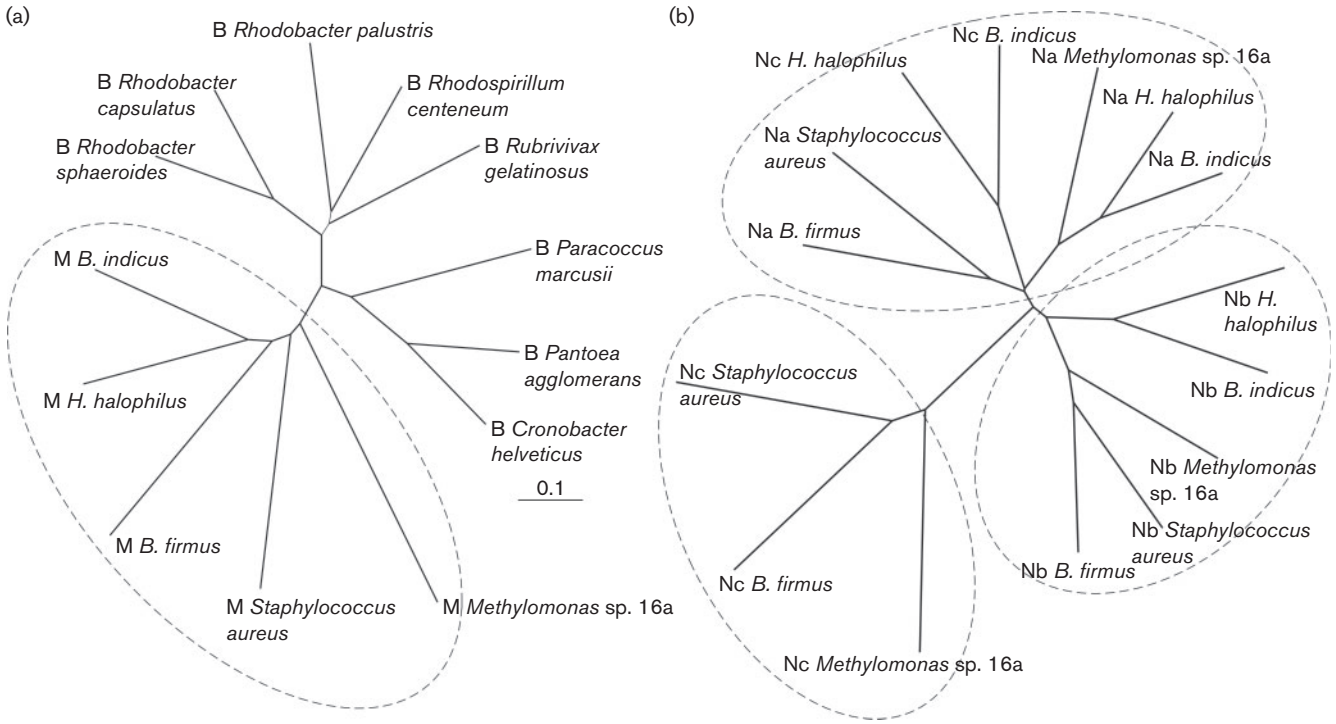


Fig. 3. Phylogenetic trees of (a) diapophytoene synthase genes (*crtM*) and phytoene synthase genes (*crtB*) (M and B, respectively), and (b) *crtN*-type genes encoding diapophytoene desaturase (*crtNa*), diapolycopene ketolase (*crtNb*) and diapolycopen-al oxidase (*crtNc*) (right). B, *crtB*; M, *crtM*; Na, *crtNa*; Nb, *crtNb*; C, *crtNc*. Species with genes and GenBank accession numbers: *Cronobacter helveticus* (*crtB*, WP_029589748.1), *H. halophilus* (*crtM*, WP_014643004.1; *crtNa*, YP_006180382.1; *crtNb*, ACM07427.1; *crtNc*, WP_014643003.1), *Methylomonas* sp. 16a (*crtM*, YP_004514339.1; *crtNa*, AAX46183.1; *crtNb*, AAX46185.1; *crtNc*, AAX46184.1), *Pantoea agglomerans* (*crtB*, BAB79604.1), *Paracoccus marcusii* (*crtB*, CAB56063), *Rhodobacter capsulatus* (*crtB*, YP_003576852.1), *Rhodobacter palustris* (*crtB*, NP_946861.1), *Rhodobacter sphaeroides* (*crtB*, AAB31139.1), *Rhodospirillum centeneum* (*crtB*, YP_002298289.1), *Rubrivivax gelatinosus* (*crtB*, AAB87738.2), *S. aureus* (*crtM*, CAA52097.1; *crtNa*, CAA52098.1; *crtNb*, WP_023487165.1; *crtNc*, WP_001084326.1).

and E, peak 2, maxima at 443, 469 and 500 nm). They co-chromatographed with standard 4,4′-diapolycopene isomers (trace F) and showed the same absorbance spectra.

Plasmid pACCRT-MN was used to generate a 4,4′-diapolycopene background. Co-expression of both *crtNb* genes resulted in a different combination of keto derivatives. With the gene from *B. indicus* (trace G), the major product was 4,4′-diaponeurosporen-4-al (7,8-dihydro-4,4′-diapo-ψ,ψ-caroten-4-al, peak 3, maxima at 445, 468 and 495 nm) together with traces of 4,4′-diapolycopen-4-al (4,4′-diapo-ψ,ψ-caroten-4-al, peak 4, maxima at 455, 475 and 506 nm). The latter compound was the major product when *crtNc* from *B. firmus* was expressed (trace H) together with 4,4′-diaponeurosporen-4-al and diapolycopen-dial (4,4′-diapo-ψ,ψ-caroten-4,4′-dial, peak 5, maxima at 472, 505 and 435 nm). Trace I shows the positions of the standard keto carotenoids in the HPLC diagram.

Expression of the *crtNc* genes from both bacilli did not result in conversion of 4,4′-diapolycopen-4,4′-dial (peak 5) when generated in the *E. coli* transformants by combination of plasmids pACCRT-MN and pBBR-crtNb (trace L). However when the *crtNc* genes were resynthesized with an optimized *E. coli* codon usage, 4,4′-diapolycopen-4,4′-dial was metabolized to 4,4′-diapolycopen-4,4′-dioic acid. With the *crtNc* gene from *B. indicus*, formation of 4,4′-diapolycopen-4-oic acid (peak 6, maxima at 455, 476 and 506 nm) together with a small amount of 4,4′-diapolycopen-4,4′-dioic acid (peak 7, maxima at 470, 492 and 522 nm) was observed (trace J), whereas *crtNc* from *B. firmus* (trace K) mediated conversion to 4,4′-diapolycopen-4,4′-dioic acid as the main product. Trace L is a chromatogram of 4,4′-diapolycopen-4,4′-dioic acid as reference.

DISCUSSION

Species with the DXS or MVA pathway feeding into terpenoid biosynthesis can be found amongst Gram-positive bacteria with a low DNA G+C content (Wilding *et al.*, 2000). In *B. indicus* and *B. firmus* which both synthesize C_{30} carotenoids, all of the genes of the DXS

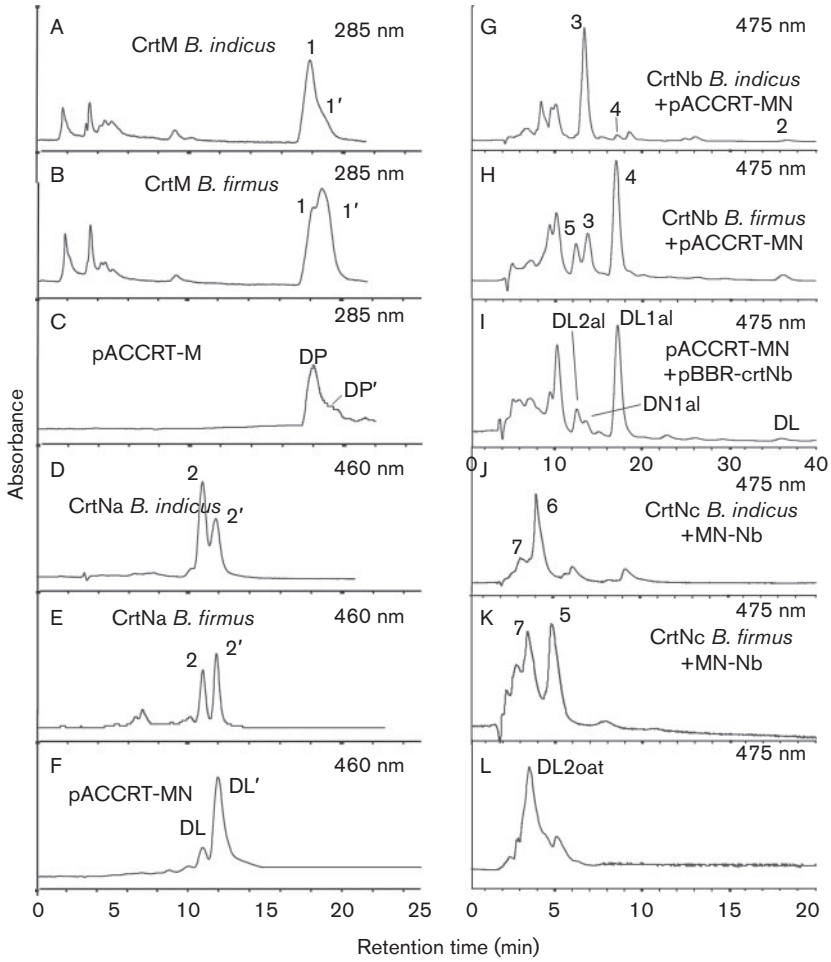


Fig. 4. HPLC separation of carotenoids from different complementation experiments with the *crt* genes from *B. indicus* HU36 and *B. firmus* GB1. Traces C and F are chromatograms with standard carotenoids; trace F shows the carotenoid background of the complementation with *crtNc* from both bacilli. Detection wavelengths varied as indicated. Mobile phase 1 was used for separations A–F and mobile phase 2 was used for separations G–L. Column 1 was used for separations A–I and column 2 was used for separations J–L. DP and DP′, 4,4′-diapophytoene isomers; DL and DL′, 4,4′-diapolycopene isomers; DL2al, 4,4′-diapolycopen-4,4′-dial; DL1al, 4,4′-diapolycopen-4-al; DN1al, 4,4′-diaponeurosporen-4-al; DL2oat, 4,4′-diapolycopen-4,4′-dioic acid. For source of standards, see Steiger *et al.* (2012b). MN-Nb, pACCRT-Mn + pBBR-crtNb plasmid combination.

pathway could be annotated (Fig. 1) corresponding to the DXS pathway in *B. subtilis* (Wagner *et al.*, 2000), and those of the MVA pathway were absent. This separates both species from other related Gram-positive, low-GC-containing bacteria using the MVA pathway, including *S. aureus* (Balibar *et al.*, 2009) – another carotenogenic species with a C_{30} carotenoid pathway. The identification of *ipp* and *fdps* in *B. indicus* and *B. firmus* indicates that IPP and DMAPP produced in the DXS pathway are in an equilibrium catalysed by an IPP isomerase and are converted by FPP synthase to form the direct substrate for C_{30} carotenoid synthesis.

The initial reactions of the carotenoid pathways in *B. indicus* and *B. firmus*, i.e. the synthesis of diapophytoene and the desaturation steps to diapolycopene, are the same. Both pathways differ by the symmetrical two-step carboxylation at both ends of the diapolycopene in the case of *B. firmus*, whereas only one end is carboxylated in *B. indicus* (Fig. 5). All of the genes for these reactions were located in gene clusters of both species. After annotation, functional expression of some of the genes demonstrated that *crtM* encodes a 4,4′-diapophytoene synthase and *crtNa* encodes a 4,4′-diapophytoene desaturase catalysing a four-step desaturation reaction to diapolycopene (Fig. 4). In this respect, CrtNa from both bacilli resemble the desaturase found in *Methylomonas* and are distinct from the three-step desaturase from *S. aureus* (Raisig & Sandmann, 1999). We could further demonstrate that *crtNb* functions as an aldehyde synthase utilizing either 4,4′-diaponeurosporene or 4,4′-diapolycopene as substrate and that *crtNc* is an aldehyde dehydrogenase which oxidizes a 4-aldehyde group of 4,4′-diapolycopene to a carboxy group. The reactions are indicated in Fig. 5. However, differences between the *crtNb* and *crtNc* gene products from *B. indicus* and *B. firmus* were found. CrtNb and CrtNc from *B. indicus* preferentially catalyse the formation of terminal aldehyde and carboxylic groups, whereas these gene products from *B. firmus* mediate the synthesis of carboxylic groups at both ends. These observed catalytic activities correspond well with the biosynthesis pathway via 4,4′-diapolycopen-4-oic acid or 4,4′-diapolycopen-4,4′-dioic acid in *B. indicus* or *B. firmus*, respectively (Fig. 5). When considering the C_{30} pathway via the synthesis of 4,4′-diapolycopen-4,4′-dioic acid which is found in many bacteria to be the most advanced, a modification of CrtNa in *S. aureus* from a four-step to a three-step desaturase (Raisig & Sandmann, 1999) determines the alternative synthesis to staphyloxanthin (Pelz

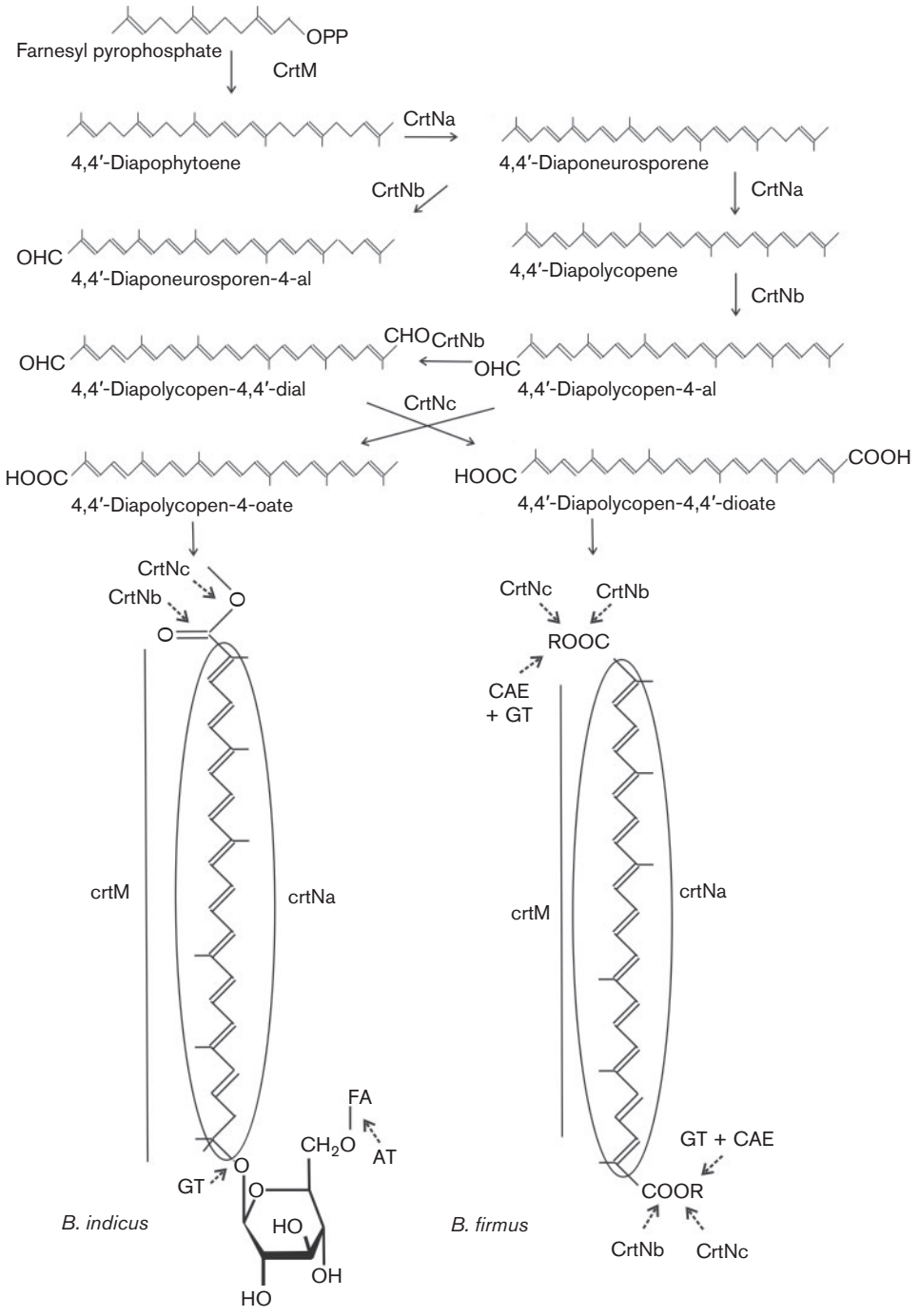


Fig. 5. Pathway and functionality of the carotenogenic genes. Biosynthesis via 4,4′-diapo carotenoid acids to methyl 4′-[6-*O*-acyl-glycosyl)oxy]-4,4′-diapolycopen-4-oic acid in *B. indicus* HU36 (left) and to 4,4′-(diglycosyl)-4,4′-diapolycopen-4,4′-dioic acid in *B. firmus* GB1 (right). R, putative glycosyl moiety. Gene products are indicated next to the corresponding reactions and gene functions related to the formation of structural elements in both bacilli.

et al., 2005; Kim & Lee, 2012), and the modification of CrtNb from a diketolase to a monoketolase determines the diversion to 4-glycosyl-4′-methyl-4,4′-diapolycopen-4′-oate fatty acid esters (Shindo *et al.*, 2008; Osawa *et al.*, 2010).

The catalytic gene functions are illustrated at structures of both carotenoid end products (Fig. 5). The other genes in the carotenogenic gene clusters shown in Fig. 2 indicated that these structures were identified only by gene comparison, but

their functions explain the synthesis of the different products of the carotenoid pathway in both bacilli (Perez-Fons *et al.*, 2011; Steiger *et al.*, 2012b). Genes *GT* and *AT* should be responsible for the glycosylation of a terminal hydroxyl group and the esterification of a sugar alcohol group with a fatty acid, respectively, in *B. indicus*. As the predicted product in *B. firmus* is a 4,4′-diapolycopen-4,4′-dioic acid glycosyl ester (Steiger *et al.*, 2012b), one of the *CAE* and/or *GT* genes found in the gene cluster may be responsible for the modification of the acid groups. One of the reactions in *B. indicus* to its carotenoid end product is the addition of water to the terminal double bond of diapolycopenoic acid. Two different carotenoid 1,2-hydratases, CrtC (Steiger *et al.*, 2003) and CruF (Sun *et al.*, 2009), are known. However, none of the corresponding genes could be identified in the genome.

The diapophytoene synthases CrtM are closely related to the phytoene synthases of the C_{40} carotenoid pathway (Fig. 3a). The same holds for the 4,4′-diapophytoene desaturase CrtNa, which belongs to the CrtI (phytoene desaturase) gene family and exhibits common sensitivity toward the inhibitor diphenylamine (Raisig & Sandmann, 2001). This enzyme was formerly named CrtN and was renamed CrtNa after a diapolycopene oxidase gene *crtNb* with sequence homology to *crtNa* was discovered (Tao *et al.*, 2005). Recently, diphenylamine-dependent accumulation of 4,4′-diapolycopen-4,4′-dial was observed in *B. firmus* cells, indicating that the oxidation of 4,4′-diapolycopen-4,4′-dial to the corresponding acid is catalysed by a CrtI/CrtN-related enzyme (Steiger *et al.*, 2012b). In the phylogenetic tree (Fig. 3), the *crtNc* genes from *B. firmus* and *H. halophilus* cluster within the *crtNa* group. Due to this close relationship and the inhibitory property of the enzyme, we assigned the gene responsible for the aldehyde to acid conversion as *crtNc*. The catalytic function of the *crtN* gene products can be explained by a common mechanism. Desaturation by *crtI* and *crtNa* proceeds via hydride transfer to NAD and proton abstraction (Sandmann, 2009). It has recently been demonstrated that the *crtO* gene product which mediates the formation of a keto group at C-4 of a carotenoid β-ionone ring acts by double hydroxylation at this position and water abstraction (Breitenbach *et al.*, 2013). Formation of the hydroxy group starts with hydride transfer similar to the desaturation reaction; however, the stabilization of the resulting carbocation differs by reaction with a hydroxyl anion. The same reaction mechanism may be involved in *crtNb* and *crtNc* catalysis.

CONCLUSION

This investigation on C_{30} carotenoid biosynthesis started from our previous analysis of the carotenoids in *B. indicus* and *B. firmus* (Perez-Fons *et al.*, 2011; Steiger *et al.*, 2012b). Although the bacilli are closely related, they form different pathway end products. We were able to reveal the details of their biosynthesis after cloning the carotenogenic genes and by reconstitution of the individual reactions *in vitro*. Some of the genes could be attributed to the same CrtI-type gene family encoding enzymes with similar initial reaction mechanisms. In addition, the organization of *Bacillus* carotenogenic genes is completely unrelated (Fig. 2) and so is the grouping of the genes in the phylogenetic tree (Fig. 3). With respect to these features, *B. indicus* resembles *H. halophilus*. The carotenoid pathway in vegetative cells and spores of some bacilli, including both sequenced strains (Duc *et al.*, 2006; Perez-Fons *et al.*, 2011) is different, with the final pathway steps dominating in the spores. The sequences of the carotenoid gene clusters should be helpful in finding control factors for transcriptional upregulation of these *crt* genes.

ACKNOWLEDGEMENTS

The authors are grateful for funding through the European Union FP7 Colorspore Project (207948).

REFERENCES

Altschul, S. F., Madden, T. L., Schäffer, A. A., Zhang, J., Zhang, Z., Miller, W. & Lipman, D. J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* **25**, 3389–3402.

Balibar, C. J., Shen, X. & Tao, J. (2009). The mevalonate pathway of *Staphylococcus aureus*. *J Bacteriol* **191**, 851–861.

Borovkov, A. Y. & Rivkin, M. I. (1997). *Xcm*I-containing vector for direct cloning of PCR products. *Biotechniques* **22**, 812–814.

Breitenbach, J., Gerjets, T. & Sandmann, G. (2013). Catalytic properties and reaction mechanism of the CrtO carotenoid ketolase from the cyanobacterium *Synechocystis* sp. PCC 6803. *Arch Biochem Biophys* **529**, 86–91.

Britton, G., Liaaen-Jensen, S. & Pfander, H. (editors) (2004). *Carotenoids Handbook*. Basel: Birkhäuser.

Duc, H., Fraser, P. D., Tam, N. K. M. & Cutting, S. M. (2006). Carotenoids present in halotolerant *Bacillus* spore formers. *FEMS Microbiol Lett* **255**, 215–224.

Hall, T. A. (1999). BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Sym Ser* **41**, 95–98.

Hanna, Z., Fregeau, C., Préfontaine, G. & Brousseau, R. (1984). Construction of a family of universal expression plasmid vectors. *Gene* **30**, 247–250.

Hosokawa, M. (2008). Structure and catalytic properties of carboxylesterase isozymes involved in metabolic activation of prodrugs. *Molecules* **13**, 412–431.

Khaneja, R., Perez-Fons, L., Fakhry, S., Baccigalupi, L., Steiger, S., To, E., Sandmann, G., Dong, T. C., Ricca, E. & other authors (2010). Carotenoids found in *Bacillus*. *J Appl Microbiol* **108**, 1889–1902.

Kim, S. H. & Lee, P. C. (2012). Functional expression and extension of staphylococcal staphyloxanthin biosynthetic pathway in *Escherichia coli*. *J Biol Chem* **287**, 21575–21583.

Kleinig, H., Schmitt, R., Meister, W., Englert, G. & Thommen, H. (1979). New C30-carotenoic acid glucosyl esters from *Pseudomonas rhodos*. *Z Naturforsch C* **34c**, 181–185.

Köcher, S., Breitenbach, J., Müller, V. & Sandmann, G. (2009). Structure, function and biosynthesis of carotenoids in the moderately

halophilic bacterium *Halobacillus halophilus*. *Arch Microbiol* **191**, 95–104.

Kovach, M. E., Elzer, P. H., Hill, D. S., Robertson, G. T., Farris, M. A., Roop, R. M., II & Peterson, K. M. (1995). Four new derivatives of the broad-host-range cloning vector pBBR1MCS, carrying different antibiotic-resistance cassettes. *Gene* **166**, 175–176.

Lairson, L. L., Henrissat, B., Davies, G. J. & Withers, S. G. (2008). Glycosyltransferases: structures, functions, and mechanisms. *Annu Rev Biochem* **77**, 521–555.

Manzo, N., D'Apuzzo, E., Coutinho, P. M., Cutting, S. M., Henrissat, B. & Ricca, E. (2011). Carbohydrate-active enzymes from pigmented Bacilli: a genomic approach to assess carbohydrate utilization and degradation. *BMC Microbiol* **11**, 198.

Osawa, A., Ishii, Y., Sasamura, N., Morita, M., Köcher, S., Müller, V., Sandmann, G. & Shindo, K. (2010). 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)* **63**, 291–295.

Osawa, A., Iki, K., Sandmann, G. & Shindo, K. (2013). Isolation and identification of 4, 4′-diapolycopene-4,4′-dioic acid produced by *Bacillus firmus* GB1 and its singlet oxygen quenching activity. *J Oleo Sci* **62**, 955–960.

Pelz, A., Wieland, K. P., Putzbach, K., Hentschel, P., Albert, K. & Götz, F. (2005). Structure and biosynthesis of staphyloxanthin from *Staphylococcus aureus*. *J Biol Chem* **280**, 32493–32498.

Perez-Fons, L., Steiger, S., Khaneja, R., Bramley, P. M., Cutting, S. M., Sandmann, G. & Fraser, P. D. (2011). Identification and the developmental formation of carotenoid pigments in the yellow/orange *Bacillus* spore-formers. *Biochim Biophys Acta* **1811**, 177–185.

Raisig, A. & Sandmann, G. (1999). 4,4′-Diapophytoene desaturase: catalytic properties of an enzyme from the C_{30} carotenoid pathway of *Staphylococcus aureus*. *J Bacteriol* **181**, 6184–6187.

Raisig, A. & Sandmann, G. (2001). Functional properties of diapophytoene and related desaturases of C_{30} and C_{40} carotenoid biosynthetic pathways. *Biochim Biophys Acta* **1533**, 164–170.

Sambrook, J., Fritsch, E. F. & Maniatis, T. (1989). *Molecular Cloning: A Laborator Manual*, 2nd edn. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory.

Sandmann, G. (2009). Evolution of carotene desaturation: the complication of a simple pathway. *Arch Biochem Biophys* **483**, 169–174.

Shindo, K., Mikami, K., Tamesada, E., Takaichi, S., Adachi, K., Misawa, N. & Maoka, T. (2007). Diapolycopenedioic acid xylosyl ester, a novel glycol-C30-carotenoic acid produced by a new marine bacterium *Rubritalea squalenifaciens*. *Tetrahedron Lett* **48**, 2725–2727.

Shindo, K., Endo, M., Miyake, Y., Wakasugi, K., Morritt, D., Bramley, P. M., Fraser, P. D., Kasai, H. & Misawa, N. (2008). 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)* **61**, 729–735.

Steiger, S., Mazet, A. & Sandmann, G. (2003). Heterologous expression, purification, and enzymatic characterization of the acyclic carotenoid 1,2-hydratase from *Rubrivivax gelatinosus*. *Arch Biochem Biophys* **414**, 51–58.

Steiger, S., Perez-Fons, L., Fraser, P. D. & Sandmann, G. (2012a). The biosynthetic pathway to a novel derivative of 4,4′-diapolycopene-4,4′-oate in a red strain of *Sporosarcina aquimarina*. *Arch Microbiol* **194**, 779–784.

Steiger, S., Perez-Fons, L., Fraser, P. D. & Sandmann, G. (2012b). Biosynthesis of a novel C30 carotenoid in *Bacillus firmus* isolates. *J Appl Microbiol* **113**, 888–895.

Sun, Z., Shen, S., Wang, C., Wang, H., Hu, Y., Jiao, J., Ma, T., Tian, B. & Hua, Y. (2009). A novel carotenoid 1,2-hydratase (CruF) from two species of the non-photosynthetic bacterium *Deinococcus*. *Microbiology* **155**, 2775–2783.

Takaichi, S., Oh-Oka, H., Maoka, T., Jung, D. O. & Madigan, M. T. (2003). Novel carotenoid glucoside esters from alkaliphilic heliobacteria. *Arch Microbiol* **179**, 95–100.

Tao, L., Schenzle, A., Odom, J. M. & Cheng, Q. (2005). Novel carotenoid oxidase involved in biosynthesis of 4,4′-diapolycopene dialdehyde. *Appl Environ Microbiol* **71**, 3294–3301.

Taylor, R. F. (1984). Bacterial triterpenoids. *Microbiol Rev* **48**, 181–198.

Thompson, J. D., Higgins, D. G. & Gibson, T. J. (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res* **22**, 4673–4680.

Wagner, W. P., Helmig, D. & Fall, R. (2000). Isoprene biosynthesis in *Bacillus subtilis* via the methylerythritol phosphate pathway. *J Nat Prod* **63**, 37–40.

Wilding, E. I., Brown, J. R., Bryant, A. P., Chalker, A. F., Holmes, D. J., Ingraham, K. A., Iordanescu, S., So, C. Y., Rosenberg, M. & Gwynn, M. N. (2000). Identification, evolution, and essentiality of the mevalonate pathway for isopentenyl diphosphate biosynthesis in Gram-positive cocci. *J Bacteriol* **182**, 4319–4327.

Edited by: G. Unden