Microbiology Papers in Press. Published October 17, 2014 as doi:10.1099/mic.0.083519-0

Microbiology

Section: Physiology & Biochemistry

Running title:

Gene function in C30 carotenoid synthesis of *Bacillus* species

Title:

Annotation and functional assignment of the genes for the C30 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], Gerhard Sandmann[1*]

[1]Biosynthesis Group, Molecular Biosciences, Goethe University Frankfurt, Frankfurt, Germany

[2]School of Biological Sciences, Royal Holloway, University of London, Egham, Surrey, UK

*Corresponding author at: Institute of Molecular Biosciences, Goethe University Frankfurt/M, Max von Laue Str. 9, D-60438 Frankfurt, Germany. Tel.: +49 69 798 29611; fax: +49 69 798 29600.

E-mail address: sandmann@bio.uni-frankfurt.de

Genome sequences are available from http://www1.uni-frankfurt.de/fb/fb15/english/institute/inst-3-mol-biowiss/AK-Sandmann/bacillus/index.html

Key words:

C30 carotenoid pathway; carotenogenic gene cluster; 1-deoxy-D-xylulose 5-phosphate synthase pathway; genome sequencing;

Size

Abstract: 241 words

Text: 4720 words

Abstract

Bacillus indicus and *Bacillus firmus* synthesize C30 carotenoids via farnesyl pyrophosphate (FPP) 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-oate and 4,4'-diapolycopen-4,4'-dioic acid with putative glycosyl esters. The genomes of both bacteria have been 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 the genes of the mevalonate pathway were absent. The gene for specific carotenoid synthesis and conversion were found on gene clusters which are differently organised in the two species. The genes involved in the formation of the carotenoid cores have been 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. By this approach, the genes for a 4,4'-diapophytoene synthase, *crtM*, for a -diapophytoene desaturase, *crtNa*, for a 4,4'-diapolycopene ketolase, *crtNb*, and for a 4,4'-diapolycopene aldehyde oxidase, *crtNc*, were identified. The three *crtN* genes are closely related and belong to the *crtI* gene family with a similar reaction mechanism of their enzyme products. Additional genes encoding glycosyl transferases and acyl transferases for the modification of the carotenoid skeleton of the diapolycopenoic acids were identified by comparison to the corresponding genes from other bacteria.

INTRODUCTION

Carotenoids are colored terpenoids. Their chromophore consists of a conjugated double bond system. Typically, carotenoids originate from the condensation of two molecules of geranylgeranyl pyrophosphate (GGPP) yielding a C40 carbon skeleton. However, among some bacteria carotenoid biosynthesis may utilize two molecules of C15 farnesyl pyrophosphate (FPP). This reaction yielding 4,4'-diapophytoene (7,8,11,12,7',8',11',12'-octyhydro-4,4'-diapo-ψ,ψ-carotene) is the starting point to a variety of acyclic C30 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 C30 carotenoid

derivatives were found in unrelated species like *Methylobacterium rhodium* (formerly *Pseudomonas rhodos*) (Kleinig *et al.,* 1979), *Methylomonas* (Tao *et al.*, 2005), *Staphylococcus aureus* (Pelz *et al.,* 2005) and *Rubritalea squalenifaciens* (Shindo *et al.*, 2007). In addition, C30 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 C30 carotenoid pathway are *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). For two *Bacillus* species, a C30 carotenoid biosynthesis pathway has been established. In *Bacillus indicus*, the major carotenoid is methyl 4'-[6-O-acyl-glycosyl)oxy]-4,4'-diapolycopen-4-oate (Perez-Fons *et al.,* 2011) whereas in *B. 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′-diapolycopene-4-oic acid (4,4'-diapo-ψ,ψ-carotene-4-oic acid) as precursor. In some pigmented bacilli, formation of end products of the carotenoid pathway is associated with spore formation (Duc *et al*., 2005; 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 C30 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 pathway (Wilding *et al.,* 2000). However, among some gram-positive bacteria the mevalonate pathway exists as an alternative route. In our genome sequences, we searched for genes of either the DXS or the mevalonate pathway and identified a carotenogenic gene cluster in both bacilli genomes. These genes encoding biosynthetic enzymes for C30 carotenoid formation were annotated and their function identified by genetic complementation in *Escherichia coli*.

METHODS

Cultivation, DNA isolation, plasmid construction and genetic complementation.

Escherichia 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 before (Khaneja *et al.*, 2010). They 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 (Hilden, Germany) genomic DNA isolation kit. Expression plasmids were constructed by PCR amplification of crt genes from genomic DNA. The genes *crtNc* from both *Bacillus* species were codon optimized for *E. coli* synthesized by MWG Operon Eurofins (Ebersberg, Germany) 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 Supplemental Table S1. 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* (Tao *et al.*, 2005). This gene was amplified by PCR with the primers *Bam*HI-forward (5'-CAGGATCCAATGAACTCAAATGACAACCA-3') and *Sac*I-reverse (5′-GTGAGCTCTTATTGCAAATCCGCCAC-3'). These restriction sites were used to clone *crtNb* into and pBBR1-MCS2 (Kovach *et al.*, 1995).

Sequencing, assembly and annotation. Initiated by the authors in a joint EU project, genome sequencing of both strains by 454 sequencing and assembly provided the sequences of the *Bacillus 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 re-sequenced and corrected *crt* genes (see also Supplemental 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 (NCBI) (Altschul *et al.*, 1997) searches against *E. coli* DXS pathway and *Halobacillus 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 were performed with the program Clustal X (Thomson *et al.*, 1997) and the alignments visualized with TreeView. Completion of nucleotide sequences, corrections and filling of gaps was carried out by re-sequencing 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 petrol, the upper phase with the carotenoids was collected and evaporated to dryness. Carotenoids were re-suspended in acetone or methanol/dichloromethane (50:50) with acetic acid (2 mM final concentration) directly before HPLC analysis on a 15 cm (column 1) or a 25 cm (column 2) Nucleosil C18, 3µ column at 10°C with a flow of 0.8 ml/min. Depending on the polarity of the carotenoids elution was either with acetonitrile / methanol / 2-propanol (85:10:5, v/v/v) (mobile phase 1) or with the same mobile phase plus 3% water and acetic acid to a concentration of 2 mM (mobile phase 2), the latter used especially for the samples containing carotenoid acids. The added acetic acid prevents 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 (Kontron Instruments, Neufahrn, Germany). Standards were generated in *E. coli* by combination of different *crt* genes as described recently (Steiger *et al.*, 2012b).

RESULTS

In non-carotenogenic *Bacillus subtilis*, terpenoid biosynthesis starts with the DXS pathway (Wagner *et al.*, 2000). Since *B. indicus* and *B. firmus* synthesize C30 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 IPP and 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 around 50% to the corresponding *E. coli* genes. In comparison to the corresponding genes from *H. halophilus*, identity values are higher reaching up to 87% identities. Their positions in the genome sequences are indicated in Fig. 1. It should be mentioned that the *dxr* gene which is not completely present in our *B. firmus* genome sequence. Nevertheless, the fragment lacking about 300 amino acids was sufficient for its definite identification. We also looked for genes from the mevalonate pathway which in other gram positive bacteria use to provide the terpenoid precursors (Wilding *et al.*, 2000) using the genes from *S. aureus* another bacterium synthesizing C30 carotenoids for comparison. However, none of the mevalonate pathway genes could be found either in the *B. indicus* or in the *B. firmus* genome.

The *crt* genes were further analysed in all three reading frames by simultaneous comparison with different homologous bacterial sequences. When frame shifts were observed, corrections

were made by re-sequencing. This revealed missing nucleotides or missing short inserts in some of the sequences. The corrected sequences of the *crt* genes from *B. indicus* and *B. firmus* are listed in Supplemental Tables S2 and S3. The organisation of the genes related to the carotenoid pathway in both bacilli is shown in Fig. 2. The positions of start and stop of the putative carotenogenic genes are indicated. All necessary genes to cover the whole C30 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 esters (Perez-Fons *et al.*, 2011) requires the catalytic activity of a glycosyl transferase (GT) with the terminal hydroxyl group acting as the sugar acceptor (Lerson *et al.*, 2008). A corresponding *AT* gene is present next to *crtNb*. However, a *GT* gene encoding a glycosyl transferase which can link a sugar molecule to a terminal hydroxy group was found isolated on another contig. In the gene cluster of *B. firmus*, 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 proposed 4,4'-diapolycopen-4,4'-dioic acid (4,4'-diapo-ψ,ψ-carotene-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 instead of hydrolysis function.

The organisation of the genes from the carotenoid pathway is completely different between *B. indicus* and *B. firmus* and compared to *Methylomonas* and *S. aureus*. However, there is a good match between *B. indicus* and *H. halophilus* (Köcher *et al.*, 2009) concerning gene arrangement and direction of transcription. Phylogenetic analysis comparing diapophytoene synthase genes *crtM* with bacterial phytoene synthase genes *crtB* shows the evolutionary relationship of both synthase genes which group at distinct clusters (Fig. 3, left). It has already been shown that the 4,4'-diapophytoene desaturase gene *crtNa* belongs to the phytoene desaturase (*crtI*) gene family (Raisig and 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 right. The genes *crtNb* formed a distinct group. Out of the *crtNc* genes, three of them 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*.

By alignment with genes from the C30 carotenoid biosynthesis pathways of other bacteria, the carotenogenic genes from *B. indicus* and *B. firmus* were putatively assigned. 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 gene 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. Trace A and B show that the *crtM* genes from *B. indicus* and *B. firmus* both synthesize 4,4'-diapophytoene isomers (peak 1 and 1') from the precursors already existing in the non-transformed *E. coli*. The 4,4'-diapophytoene peaks co-chromatograph with the standard in trace C and exhibit the characteristic absorbance spectrum at 275, 285 and 297 nm. The *crtNa* genes in a 4,4'-diapophytoene background (generated by plasmid pACCRT-M) are responsible for the formation of two isomers of 4,4'-diapolycopene (traces D and E, peak 2, maxima at 443, 469, 500 nm). They co-chromatograph with standard 4,4'-diapolycopene isomers (trace F) and show 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-carotene-4-al, peak 3, 445, 468, 495 nm) together with traces of 4,4'-diapolycopen-4-al (4,4'-diapo-ψ,ψ-carotene-4-al, peak 4 455, 475, 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-carotene-4,4'-dial, peak 5, 472, 505, 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 re-synthesized with an optimized *E. coli* codon usage, 4,4'-diapolycopen-4,4'-dial was metabolised 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, 506 nm) together with small amount of 4,4'-diapolycopen-4,4'-dioic acid (peak 7, maxima at 470, 492, 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

Among gram-positive bacteria with low GC content, species with the DXS or the MVA pathway feeding into terpenoid biosynthesis can be found (Wilding *et al.,* 2000). In *B. indicus* and *B. firmus* which both synthesize C30 carotenoids all genes of the DXS 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 C30 carotenoid pathway. The identification of the *ipp* and *fdps* in *B. indicus* and *B. firmus* indicate that IPP and DMAPP produced in the DXS pathway are in equilibrium catalysed by an IPP isomerase and are converted by FPP synthase to form the direct substrate for C30 carotenoid synthesis.

The initial reactions of the carotenoid pathways in *B. indicus* and *B. firmus* 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 case of *B. firmus* whereas in *B. indicus* only one end is carboxylated (Fig. 5). All 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* a 4,4’-diapophytoene desaturase catalysing a 4-step desaturation reaction to diapolycopene (Fig. 4). In this respect, CrtNa from both bacilli resemble the desaturase found in *Methylomonas* and are distinct to the 3-step desaturase from *S. aureus* (Raisig and Sandmann, 1999). We could further demonstrate that *crtNb* functions as an aldehyde synthase utilizing either 4,4’-diaponeurosporene or4,4’-diapolycopene as substrates 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 one terminal aldehyde and carboxylic group 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’-diapolycopene-4-oic acid mono or 4,4’-diapolycopene-4,4’-dioic acid in both bacilli,

respectively (Fig. 5). When considering the C30 pathway via the synthesis of 4,4'-diapolycopen-4,4'dioic acid which is found in many bacteria as the most advanced, a modification of CrtNa in *S. aureus* from a 4-step to a 3-step desaturase (Raisig and Sandmann, 1999) determines the alternative synthesis to staphyloxanthin (Pelz *et al.*, 2005; Kim & Lee, 2012) and the modification of CrtNb from a diketolase to a monoketolase the diversion to 4-glycosyl-4'-methyl-4,4'-diapolycopene-4'-oate fatty acid esters (Shindo *et al.*, 2008; Ogawa *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*. Since the predicted product in *B. firmus* is a 4,4'-diapolycopene-4,4'-dioic acid glycosyl ester (Steiger *et al.*, 2012b), one of the genes *CAE* and/or *GT* 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 water addition 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 C40 carotenoid pathway (Fig. 3 left). 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 is re-named 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'-diapolycopene-4,4'-dial was observed in *B. firmus* cells indicating that the oxidation of the 4,4'-diapolycopene-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 C4 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 stabilisation of the resulting carbo cation differes by reaction with a hydroxyl anion. The same reaction mechanism may be involved in *crtNb* and *crtNc* catalysis.

CONCLUSION

This investigation on C30 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 pathways 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 mechanism. In addition, the organisation 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 sequences strains (Duc *et al*., 2005; Perez-Fons *et al*., 2011) is differing with the final pathway steps dominating in the spores. The sequences of the carotenoid gene clusters should be helpful to find controlling factors for transcriptional up-regulation of these crt genes.

ACKNOWLEDGEMENTS

The authors are grateful for funding through the EU FP7 Colorspore project no. 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. Nucl 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). XcmI-Containing vector for direct cloning of PCR products. *BioTechn* **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.

Duc, L.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.

Britton, G., Liaaen-Jensen, S., Pfander, H. (2004). Carotenoids Handbook, compiled by Mercadante, A.Z. and Egeland, E.S., Birkhäuser Verlag, Basel.

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

Hanna, Z., Fregeau, C., Prefontaine, 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., Fraser, P. D. & Cutting, S. M. (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.

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.

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

Lairson, L. L., Henrissat, B., Davies, G. J. & Withers, S. G. (2008). Glycosyltransferases: Structures, functions, and mechanisms. *Ann 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 Microbiology* **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., Alber, 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). 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 Laboratory 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. (2008). Methyl glucosyl-3,4-dehydro-apo-8′-lycopenoate, a novel antioxidative glyco-C30-carotenoic acid produced by a marine bacterium *Planococcus maritimus*. *J Antibiot* **61**, 729–735.

Steiger, S., Mazet, A. & Sandmann, G. (2003). Heterologous expression, purification, and

411 enzymatic characterization of the acyclic carotenoid 1,2-hydratase from *Rubrivivax*
412 *gelatinosus*. *Arch Biochem Biophys* **414**, 51-58.
413 **Steiger, S., Perez-Fons, L., Fraser, P.D. & Sandmann, G. (2012a).** The biosynthetic
414 pathway to a novel derivative of 4,4'-diapolycopene-4,4'-oate in a red strain of
415 *Sporosarcina aquimarina*. *Arch Microbiol* **194**, 779-784.
416 **Steiger S., Perez-Fons L., Fraser P. D. & Sandmann G. (2012b).** Biosynthesis of a novel
417 C30 carotenoid in *Bacillus firmus* isolates. *J Appl Microbiol* **113**, 888-895.
418 **Sun, Z., Shen, S., Wang, C., Wang, H., Hu, Y., Jiao, J., Ma, T., Tian, B. & Hua, Y.**
419 **(2009).** A novel carotenoid 1,2-hydratase (CruF) from two species of the non-
420 photosynthetic bacterium *Deinococcus*. *Microbiol* **155**, 2775–2783.
421 **Takaichi, S., Oh-oka, H., Maoka, T., Jung, D. O. & Madigan, M. T. (2003).** Novel
422 carotenoid glucoside esters from alkaliphilic heliobacteria. *Arch Microbiol* **179**, 95–
423 100.
424 **Tao, L., Schenze, A., Odom, J. M., Cheng, Q. (2005).** Novel carotenoid oxidase involved in
425 biosynthesis of 4,4'diapolycopene dialdehyde. *Appl Environ Microbiol* **77,** 3294-3301.
426 **Taylor, R. F. (1984).** Bacterial triterpenoids. *Microbiol Rev* **48**, 181–198.
427 **Thompson, J. D., Higgins, D. G. & Gibson, T. J. (1994).** CLUSTALW: improving the
428 sensitivity of progressive multiple sequence alignment through sequence weighting,
429 positions-specific gap penalties and weight matrix choice. *Nucl Acids Res* **22**, 4673–
430 4680.
431 **Wagner, W. P., Helmig, D. & Fall, R. (2000).** Isoprene biosynthesis in *Bacillus subtilis* via
432 the methylerythritol phosphate pathway. *J Nat Prod* **63**, 37–70.
433 **Wilding, E. I., Brown, J. R., Bryant, A. P., Chalker, A. F., Holmes, D. J., Ingraham, K.**
434 **A., Iordanescu, S., So, C. Y., Rosenberg, M. & Gwynn, M. N. (2000).**
435 Identification, evolution and essentiality of the mevalonate pathway for isopentenyl
436 diphosphate biosynthesis in Gram-positive cocci. *J Bacteriol* **182**, 4319-4327.
437
438

figure legends

Figure 1. The 1-deoxy-D-xylulose 5-phosphate synthase pathway to isopentenyl pyrophosphate and dimethylallyl pyrophosphate and its continuation to farnesyl pyrophosphate in *Bacillus indicus* HU36 and *Bacillus firmus* GB1. Genes are positioned next to the corresponding reactions; the inserted 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-methyl-D-erythritol 2,4-cyclodiphosphate synthase, *gcpE* (=*ispG*) (E)-4hydroxy-3-methylbut-2-enyl-diphosphate synthase, *lytB* (=*ispH*) 4-hydroxy-3-methylbut-2-enyl-diphosphate reductase, *fdps* farnesyl diphosphate synthase, *ipp* isopentenylpyrophosphat isomerase *Gene *dxr* from *Bacillus firmus* is incomplete.

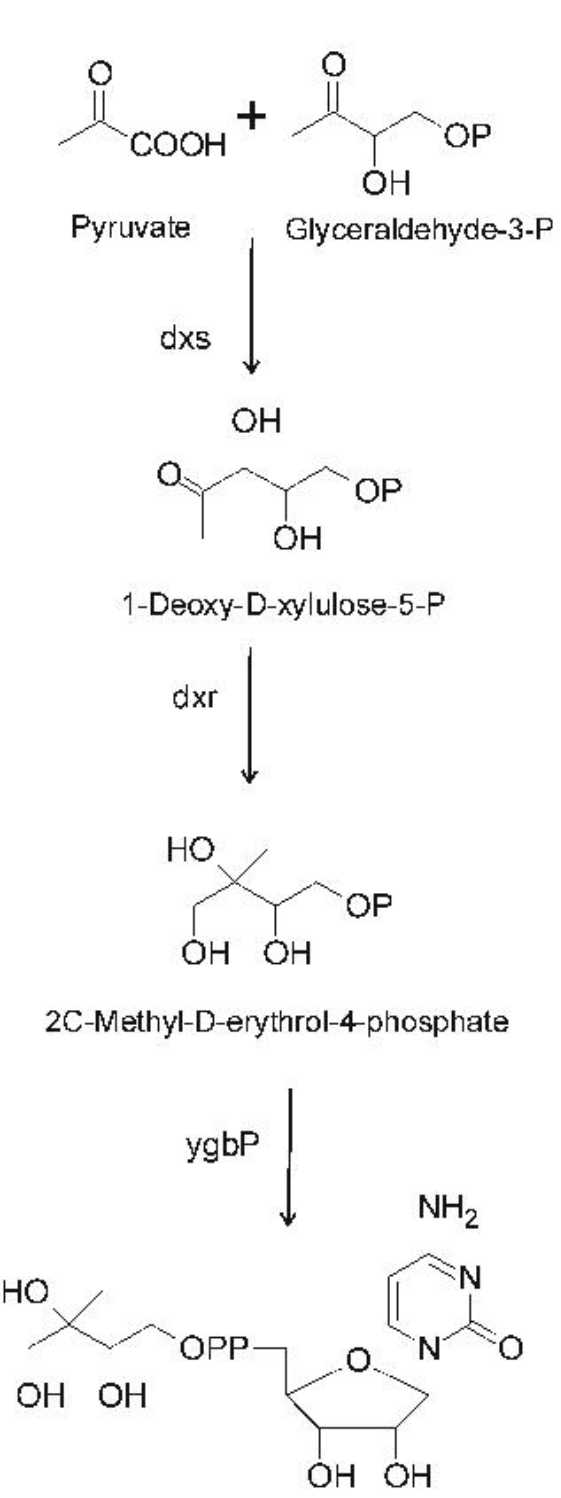
Figure 2. Organisation of the carotenogenic genes from *Bacillus indicus* HU36 and *Bacillus firmus* GB1 in comparison to the gene clusters of other C30 carotenoid synthesizing bacteria, *Halobacillus halophilus* (Köcher *et al.,* 2009), *Methylomonas* (Tao *et al.,* 2005) and *Staphylococcus aureus* (Kim and Lee, 2012). Numbers indicate the nucleotide position of the beginning and end of the *Bacillus* genes. Final assignment of genes from *Bacillus indicus* HU36 and *Bacillus firmus* was by functional pathway complementation in *Escherichia coli* except for those marked with an asterix.Genes: *crtM* 4,4'-diapophytoene synthase, *crtNa* 4,4'-diapophytoene desaturase, *crtNb*, 4,4'-diapolycopene ketolase, *crtNc* 4,4'-diapolycopene aldehyde oxidase, *AT* acyltransferase, *CAE* carboxyl esterase, *GT* glycosyl transferase.

Figure 3. Phylogenetic trees of diapophytoene synthase genes *crtM* and phytoene synthase genes *crtB* (indicated as M and B, respectively) (left) and of the *crtN* type genes encoding diapophytoene desaturase, *crtNa*; diapolycopene ketolase, *crtNb*, and diapolycopene-al oxydase, *crtNc* (right). Abbreviations: B *crtB*, M *crtM*, Na *crtNa*, Nb *crtNb*, C *crtNc*. Species with genes and accession numbers: *Cronobacter helveticus* (*crtB* WP_029589748.1), Halob *Halobacillus 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.agg *Pantoea agglomerans* (*crtB* BAB79604.1), *Paracoccus marcusii* (*crtB* CAB56063) Rba.caps *Rhodobacter capsulatus* (*crtB* YP_003576852.1), Rba.pal *Rhodobacter palustris* (*crtB*

NP_946861.1), *Rhodospirillum centeneum* (*crtB* YP_002298289.1); Staph *Staphylococcus aureus* (*crtM* CAA52097.1, *crtNa* CAA52098.1, *crtNb* WP_023487165.1, *crtNc* WP_001084326.1).

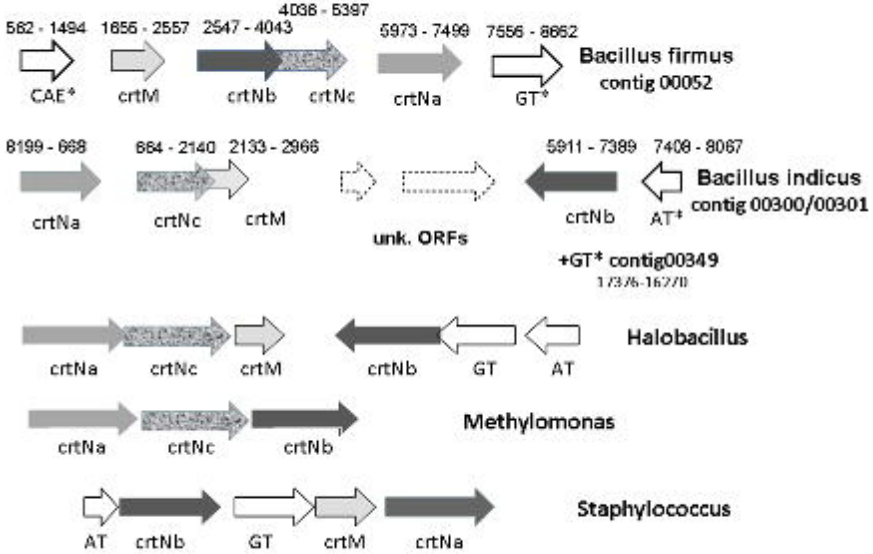
Figure 4. HPLC separation of carotenoids from different complementation experiments with the crt genes from *Bacillus indicus* HU36 and *Bacillus firmus* GB1. Traces C and F are chromatograms with standard carotenoids, trace F shows the carotenoid background of the complementation with *crtNc* from both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

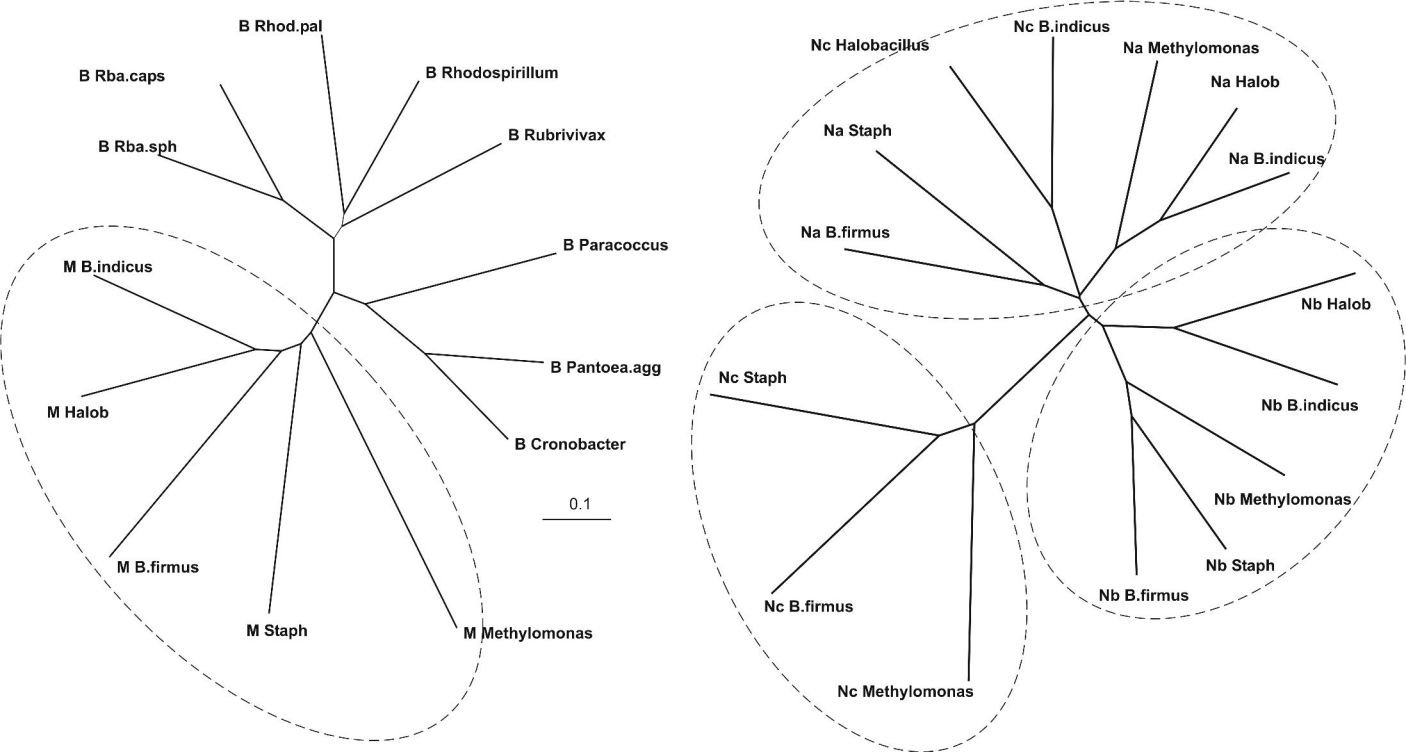
Figure 5. Pathway and functionality of the carotenogenic genes. Biosynthesis via 4,4'-diapo carotenoid acids to methyl 4'-[6-O-acyl-glycosyl)oxy]-4,4'-diapolycopen-4-oate in *Bacillus indicus* HU36 (left part) and to a 4,4'-(diglycosyl)-4,4'-diapolycopene-4,4'-dioate in *Bacillus firmus* GB1 (right part). R indicates a putative glycosyl moity. Genes are placed next to the corresponding reactions and gene functions related to the formation of structural elements in both bacilli.

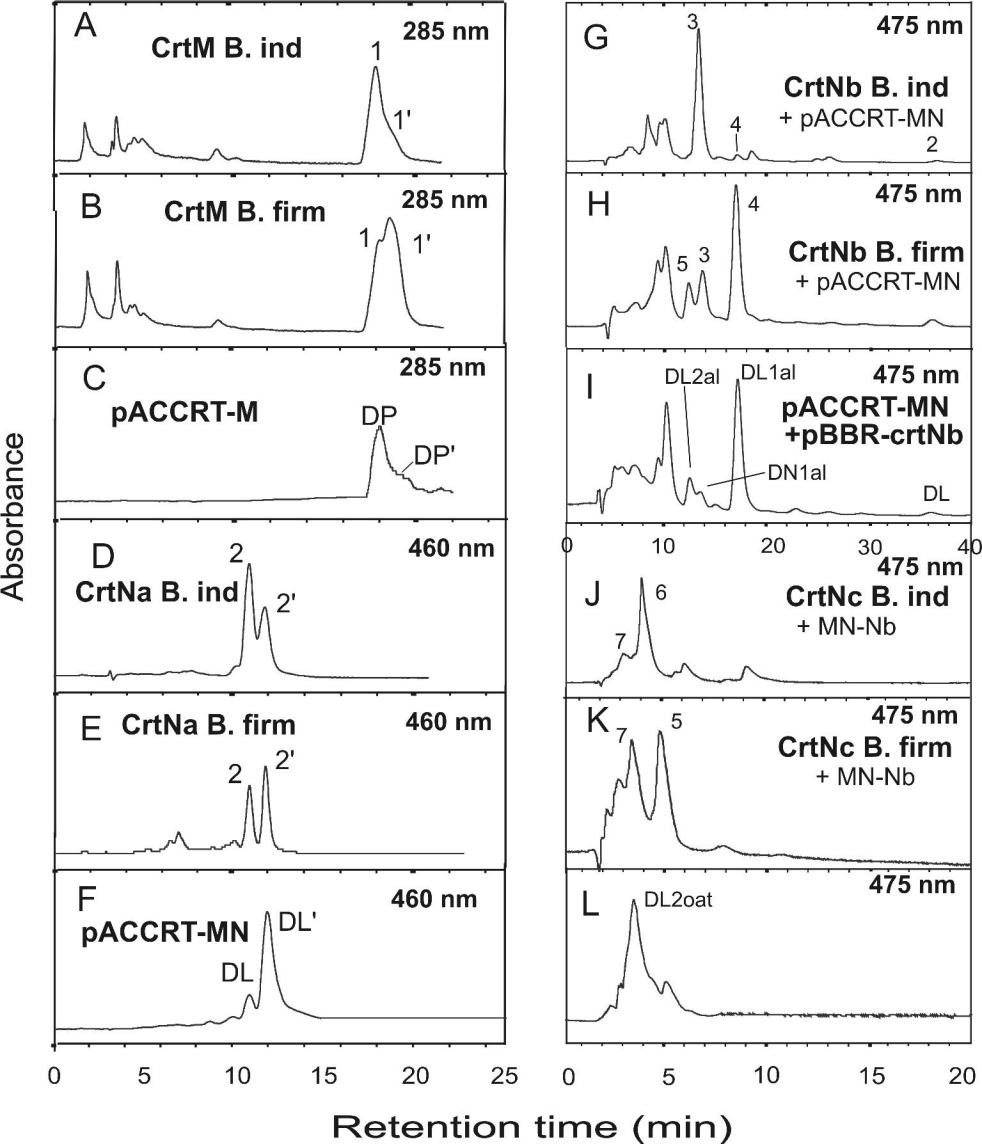


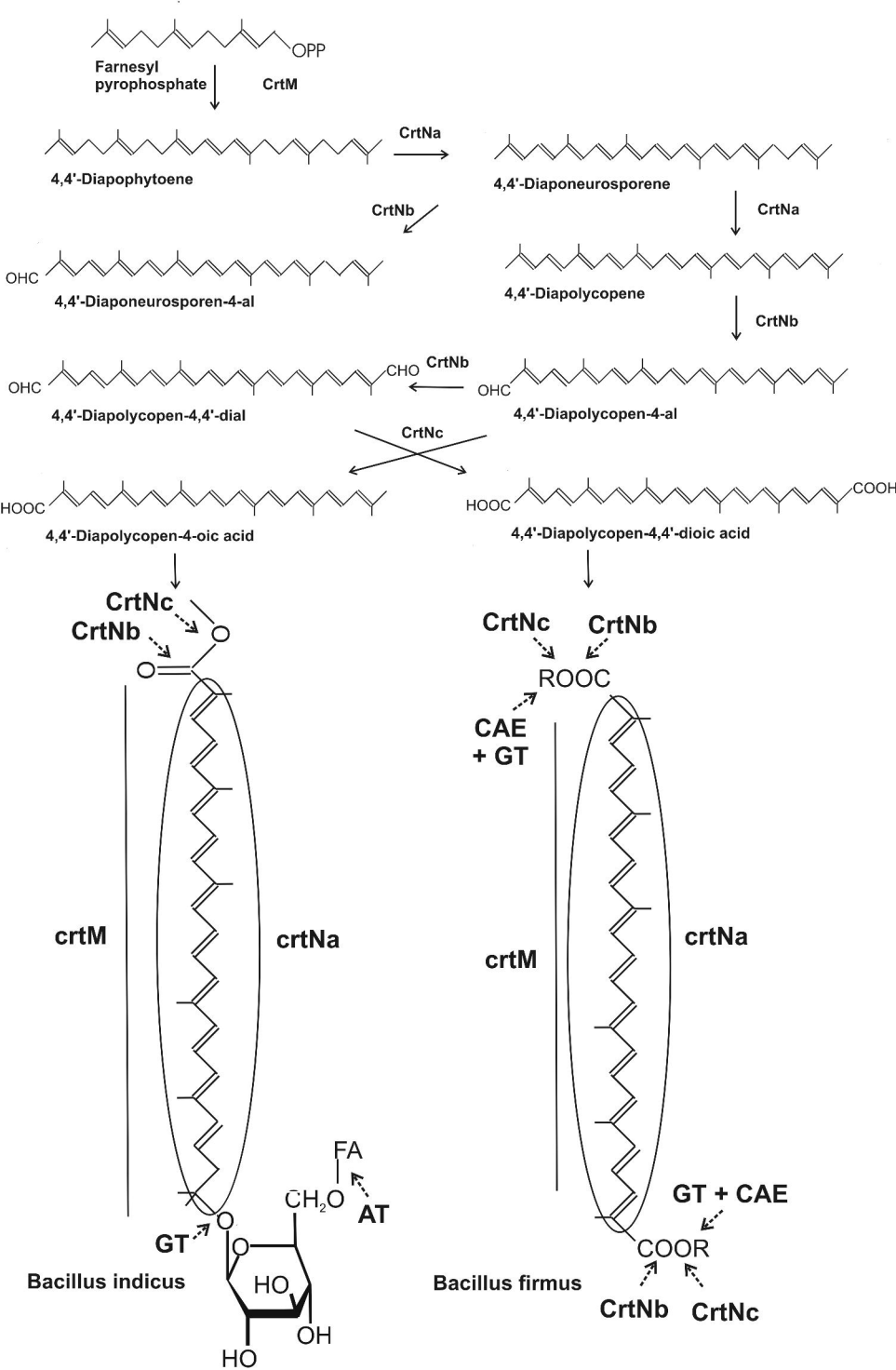
Gene	Position
Bacillus indicus	
dxs	Contig 00445, bp 8903-7047
dxr	Contig 00376, bp 12629-13750
ygbP	Contig 0006, bp 49460-48795
ychB	Contig 00718, bp 851-75,
ygbB	Contig 00006, bp 48764-48300,
gcpE	Contig 00030, bp 6823-7935,
lytB	Contig 00030, bp 15961-16908
ipp	Contig 00430, bp 14061-13045
fdps	Contig 00445, bp 10140-9267

Gene	Position
Bacillus firmus	
dxs	Contig 00976, bp 8267-6771
dxr	Contig 00262, bp 4-165*
ygbP	Contig 00536, bp 1230-1886
ychB	Contig 00392, bp 2262-1831
ygbB	Contig 00536, bp 1914-2372
gcpE	Contig 00992, bp 213-1301,
lytB	Contig 00995, bp 450-1409
ipp	Contig 00956, bp 10428-10095 plus Contig 00957, bp 618-208
fdps	Contig 00976, bp 9904-9022,


562 - 1494
1656 - 2557
2547 - 4043
4036 - 5397
5973 - 7499
7556 - 8662
CAE*
crtM
crtNb
crtNc
crtNa
GT*
Bacillus firmus
contig 00052
8199 - 668
664 - 2140
2133 - 2966
5911 - 7389
7408 - 8067
crtNa
crtNc
crtM
unk. ORFs
crtNb
AT*
Bacillus indicus
contig 00300/00301
+GT* contig00349
crtNa
crtNc
crtM
crtNb
GT
AT
Halobacillus
crtNa
crtNc
crtNb
Methylomonas
AT
crtNb
GT
crtM
crtNa
Staphylococcus


B Rhod.pal
B Rba.caps
B Rhodospirillum
B Rba.sph
B Rubrivivax
B Paracoccus
M B.indicus
B Pantoea.agg
M Halob
B Cronobacter
0.1
M B.firmus
M Staph
M Methylomonas
Nc Halobacillus
Nc B.indicus
Na Methylomonas
Na Halob
Na Staph
Na B.indicus
Na B.firmus
Nb Halob
Nc Staph
Nb B.indicus
Nb Methylomonas
Nb Staph
Nb B.firmus
Nc B.firmus
Nc Methylomonas


A
CrtM B. ind
285 nm
1
1'
B
CrtM B. firm
285 nm
1
1'
C
pACCRT-M
285 nm
DP
DP'
D
CrtNa B. ind
460 nm
2
2'
E
CrtNa B. firm
460 nm
2
2'
F
pACCRT-MN
460 nm
DL
DL'
G
CrtNb B. ind
+ pACCRT-MN
475 nm
3
4
2
H
CrtNb B. firm
+ pACCRT-MN
475 nm
4
5
3
I
pACCRT-MN
+pBBR-crtNb
475 nm
DL2al
DL1al
DN1al
DL
J
CrtNc B. ind
+ MN-Nb
475 nm
6
7
K
CrtNc B. firm
+ MN-Nb
475 nm
7
5
L
475 nm
DL2oat
Absorbance
0
5
10
15
20
25
0
10
20
30
40
0
5
10
15
20
Retention time (min)


Farnesyl pyrophosphate
OPP
CrtM
4,4'-Diapophytoene
CrtNa
4,4'-Diaponeurosporene
CrtNb
CrtNa
OHC
4,4'-Diaponeurosporen-4-al
4,4'-Diapolycopene
CrtNb
CrtNb
OHC
CHO
4,4'-Diapolycopen-4,4'-dial
OHC
4,4'-Diapolycopen-4-al
CrtNc
HOOC
4,4'-Diapolycopen-4-oic acid
HOOC
COOH
4,4'-Diapolycopen-4,4'-dioic acid
CrtNc
CrtNb
crtM
crtNa
FA
CH2O
AT
GT
O
HO
OH
HO
Bacillus indicus
CrtNc
CrtNb
ROOC
CAE
+ GT
crtM
crtNa
GT + CAE
COOR
CrtNb
CrtNc
Bacillus firmus