Annotation and functional assignment of the genes for the C\textsubscript{30} carotenoid pathways from the genomes of two bacteria: \textit{Bacillus indicus} and \textit{Bacillus firmus}

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

\textbf{INTRODUCTION}

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

A series of differently pigmented bacilli have been isolated and their pigments tentatively assigned as carotenoids (Khaneja \textit{et al.}, 2010). A C\textsubscript{30} carotenoid biosynthesis
pathway has been established for two *Bacillus* species. In *Bacillus indicus*, the major carotenoid is methyl 4'-[6-O-acyl-glycosyl]oxy]-4,4'-diapocopen-4-oic acid (Perez-Fons et al., 2011), whereas in *Bacillus firmus*, 4,4'-diapocopen-4,4'-dioic acid with putative glycosyl esters accumulate (Steiger et al., 2012b; Osaka et al., 2013), all sharing 4,4'-diapocopen-4-oic acid (4,4'-dialo-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'-diapocarotenopeine-derived biosynthesis of staphyloxyanthin were identified and functionally assigned (Pelz et al., 2005; Kim & Lee, 2012), the genes for the oxidative C\(_{30}\) pathway via 4,4'-diapocopen 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 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 DH5\(\alpha\) 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 (Khanuja 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 *crt*NC genes from both *Bacillus* species were codon optimized for *E. coli* synthesized by MWG Operon Eurofins and provided already inserted into the SalI/HindIII sites of plasmid pEX-K. Out of this plasmid, they were both cloned into the SalI/HindIII 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 *Methylobacterium* sp. 16a (Tao et al., 2005). This gene was amplified by PCR with the primers BamHI-forward (5'-CAGGGATCCATGAAGATGACACCA-3') and SacI-reverse (5'-GTGAGGCTTTATGAAACTCGAC-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/lbf/bib15/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 *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 cm (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 (DAMPA) 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*. 

http://mic.sgmjournals.org
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 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-4,4'-methyl-4,4'-diaplycopene-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
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 phytene 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 phytene desaturase (crtI) gene family (Raisig & Sandmann, 2001). In addition to crtNa genes, two related clusters of the 4,4’-diapolycopeone ketolase genes (crtNb) and the 4,4’-diapolycopeone 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 C30 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’-diapolycopeone (traces D...
and E, peak 2, maxima at 443, 469 and 500 nm). They co-chromatographed with standard 4,4'-diapolycepone isomers (trace F) and showed the same absorbance spectra.

Plasmid pACCRT-MN was used to generate a 4,4'-diapolycepone 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'-diapolo-\( \psi,\psi \)-caroten-4-al, peak 3, maxima at 445, 468 and 495 nm) together with traces of 4,4'-diapolycepone-4-al (4,4'-diapolo-\( \psi,\psi \)-caroten-4-al, peak 4, maxima at 455, 475 and 506 nm). The latter compound was the major product when \( \text{crtNc} \) from \( B. \) \( firmus \) was expressed (trace H) together with 4,4'-diaponeurosporen-4-al and diapolycepone-dial (4,4'-diapolo-\( \psi,\psi \)-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 \( \text{crtNc} \) genes from both bacilli did not result in conversion of 4,4'-diapolycepone-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 \( \text{crtNc} \) genes were resynthesized with an optimized \( E. \) \( coli \) codon usage, 4,4'-diapolycepone-4,4'-dial was metabolized to 4,4'-diapolycepone-4,4'-dioic acid. With the \( \text{crtNc} \) gene from \( B. \) \( indicus \), formation of 4,4'-diapolycepone-4-oic acid (peak 6, maxima at 455, 476 and 506 nm) together with a small amount of 4,4'-diapolycepone-4,4'-dioic acid (peak 7, maxima at 470, 492 and 522 nm) was observed (trace J), whereas \( \text{crtNc} \) from \( B. \) \( firmus \) (trace K) mediated conversion to 4,4'-diapolycepone-4,4'-dioic acid as the main product. Trace L is a chromatogram of 4,4'-diapolycepone-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
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* and *B. firmus* (Balibar et al., 2009) – another carotenogenic species with a C₃₀ 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 C₃₀ 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 diapolycope, are the same. Both pathways differ by the symmetrical two-step carboxylation at both ends of the diapolycope 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 diapolycope (Fig. 4). In this respect, *CrtNa* from both bacilli resemble the desaturase found in *Methylophilus* 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′-diapolycope as substrate and that *crtNc* is an aldehyde dehydrogenase which oxidizes a 4-aldehyde group of 4,4′-diapolycope 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′-diapolycope-4-oic acid or 4,4′-diapolycope-4,4′-dioic acid in *B. indicus* or *B. firmus*, respectively (Fig. 5). When considering the C₃₀ pathway via the synthesis of 4,4′-diapolycope-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...
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′-diapolyplcopen-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...

**Fig. 5.** Pathway and functionality of the carotenogenic genes. Biosynthesis via 4,4′-diapo carotenoid acids to methyl 4′-[6-O-acyl-glycosyloxy]-4,4′-diapolyplcopen-4-oic acid in *B. indicus* HU36 (left) and to 4,4′-(diglycosyl)-4,4′-diapolyplcopen-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.
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′-diapopolycopen-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 diapopolycopenoic 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. 3a). The same holds for the 4,4′-diapophytoene desaturase CrtNa, which belongs to the Crtl (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 ctna was discovered (Tao et al., 2005). Recently, diphenylamine-dependent accumulation of 4,4′-diapolycopogen-4,4′-dial was observed in B. firmus cells, indicating that the oxidation of 4,4′-diapolycopogen-4,4′-dial to the corresponding acid is catalysed by a Crtl/CrtN-related enzyme (Steiger et al., 2012b). In the phylogenetic tree (Fig. 3), the ctna genes from B. firmus and H. halophilus cluster within the ctna 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 ctna. The catalytic function of the ctna gene products can be explained by a common mechanism. Desaturation by ctnl and ctna proceeds via hydride transfer to NAD and proton abstraction (Sandmann, 2009). It has recently been demonstrated that the ctna 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 ctnb and ctna 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 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 ctn genes.

**ACKNOWLEDGEMENTS**

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

**REFERENCES**


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


