PNAS RESEARCH ARTICLE | MICROBIOLOGY OPEN ACCESS

The squalene route to C30 carotenoid biosynthesis and the origins of carotenoid biosynthetic pathways

Carlos Santana-Molina[a,1,2], Valentina Henriques[a], Damaso Hornero-Méndez[b], Damien P. Devos[a,1], and Elena Rivas-Marin[a,1]

Edited by Eugene Koonin, NIH, Bethesda, MD; received June 20, 2022; accepted November 4, 2022

Carotenoids are isoprenoid lipids found across the tree of life with important implications in oxidative stress adaptations, photosynthetic metabolisms, as well as in membrane dynamics. The canonical view is that C40 carotenoids are synthesized from phytoene and C30 carotenoids from diapophytoene. Squalene is mostly associated with the biosynthesis of polycyclic triterpenes, although there have been suggestions that it could also be involved in the biosynthesis of C30 carotenoids. However, demonstration of the existence of this pathway in nature is lacking. Here, we demonstrate that C30 carotenoids are synthesized from squalene in the Planctomycetes bacteria and that this squalene route to C30 carotenoids is the most widespread in prokaryotes. Using the evolutionary history of carotenoid and squalene amino oxidases, we propose an evolutionary scenario to explain the origin and diversification of the different carotenoid and squalene-related pathways. We show that carotenoid biosynthetic pathways have been constantly transferred and neofunctionalized during prokaryotic evolution. One possible origin of the squalene pathway connects it with the one of C40 carotenoid synthesis of Cyanobacteria. The widespread occurrence of the squalene route to C30 carotenoids in *Bacteria* increases the functional repertoire of squalene, establishing it as a general hub of carotenoids and polycyclic triterpenes synthesis.

squalene | C30 carotenoids | hopanoids | Planctomycetes | terpenoid evolution

Carotenoids are isoprenoid lipids found in all photosynthetic and some non-photosynthetic organisms. The most abundant carotenoids are produced by photosynthetic organisms and have C40 backbones, although some chemoorganotrophic bacteria are capable of producing *de novo* C30, C45, or C50 carotenoids (1, 2). The production of C30 carotenoids is widespread in *Bacteria*, but how the precursor is synthesized is unclear in most cases. Thus far, bacterial C30 carotenoid production has only been characterized in *Firmicutes* where it starts with the enzyme CrtM, a 4,4′-diapophytoene synthase. Carotenoid biosynthesis is evolutionarily related to the biosynthesis of polycyclic triterpenes, such as hopanoids and sterols, as some of the initial enzymes involved are homologues (Fig. 1) (3, 4). Squalene ($C_{30}H_{50}$) is the precursor of polycyclic triterpenes, which can be synthesized *via* two routes: the HpnCDE pathway, found mostly in *Bacteria*; and squalene synthase (Sqs), a single enzyme found in the three domains of life. HpnC, HpnD, and Sqs belong to the trans-isoprenyl diphosphate synthases head-to-head (Trans IPPS HH) family, to which the enzymes initiating C30 and C40 carotenoid biosynthesis, CrtB and CrtM, also belong. These Trans IPPS HHs generate the initial backbones of polycyclic triterpenes or carotenoids, which then become the substrate for specific amino oxidases, also known as phytoene desaturases: CrtN/P and HpnE act on C30 backbones, while CrtI/D or CrtP-Qc/CrtH act on C40 backbones ("c" in CrtP-Qc indicates the cyanobacterial-chlorobi pathway).

Some bacteria, such as Planctomycetes, have C30-specific amino oxidases and subsequent carotenoid-modifying enzymes, but do not have the Trans IPPS HHs usually associated with carotenoid synthesis, CrtM. Instead, planctomycetal genomes encode the HpnCDE enzymes responsible for squalene production, suggesting that squalene could be the substrate for these C30-specific amino oxidases (4). This possibility is supported by three observations. First, it has been artificially demonstrated that the C30-specific amino oxidases, CrtN, can accept squalene as substrate (5). Second, a sterol synthesis-deficient mutant of the planctomycete *Gemmata obscuriglobus* that accumulates squalene shows a brighter red pigmentation (6). And third, interrupting the *hpnE* genes in the planctomycete *Planctopirus limnophila* and in the α-proteobacterium *Methylobacterium extorquens* results in non-pigmented colonies due to the lack of carotenoid production (4, 7). These evidences suggest the existence of a squalene route to carotenoids in nature but this possibility requires further confirmation.

Significance

Carotenoids are isoprenoids lipids found in many prokaryotes and eukaryotes with important physiological implications, including antioxidative functions and modulating membrane properties. Here, we unveil a biosynthetic pathway of C30 carotenoids that uses squalene as a precursor, simultaneously showing that this route is widespread in bacteria. This new pathway expands the functional repertoire of the squalene which has been exclusively associated to the synthesis of polycyclic triterpenes, such as sterols. Reconstructing the carotenoid amino oxidases evolution, we infer two main origins of these carotenoid pathways, showing that these routes have been spread by lateral gene transfer between prokaryotes and readapted once acquired. Together, we provide a comprehensive understanding for the biosynthesis and evolution of carotenoid pathways in an extensive diversity of prokaryotes.

The authors declare no competing interest.

This article is a PNAS Direct Submission.

[1]To whom correspondence may be addressed. Email: csantmol@gmail.com, damienpdevos@gmail.com, or erivmar@upo.es.

[2]Present address: Department of Marine Microbiology and Biogeochemistry, Royal Netherlands Institute for Sea Research, Utrecht University, 1790, Den Burg, The Netherlands

This article contains supporting information online at https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2210081119/-/DCSupplemental.

Published December 19, 2022.

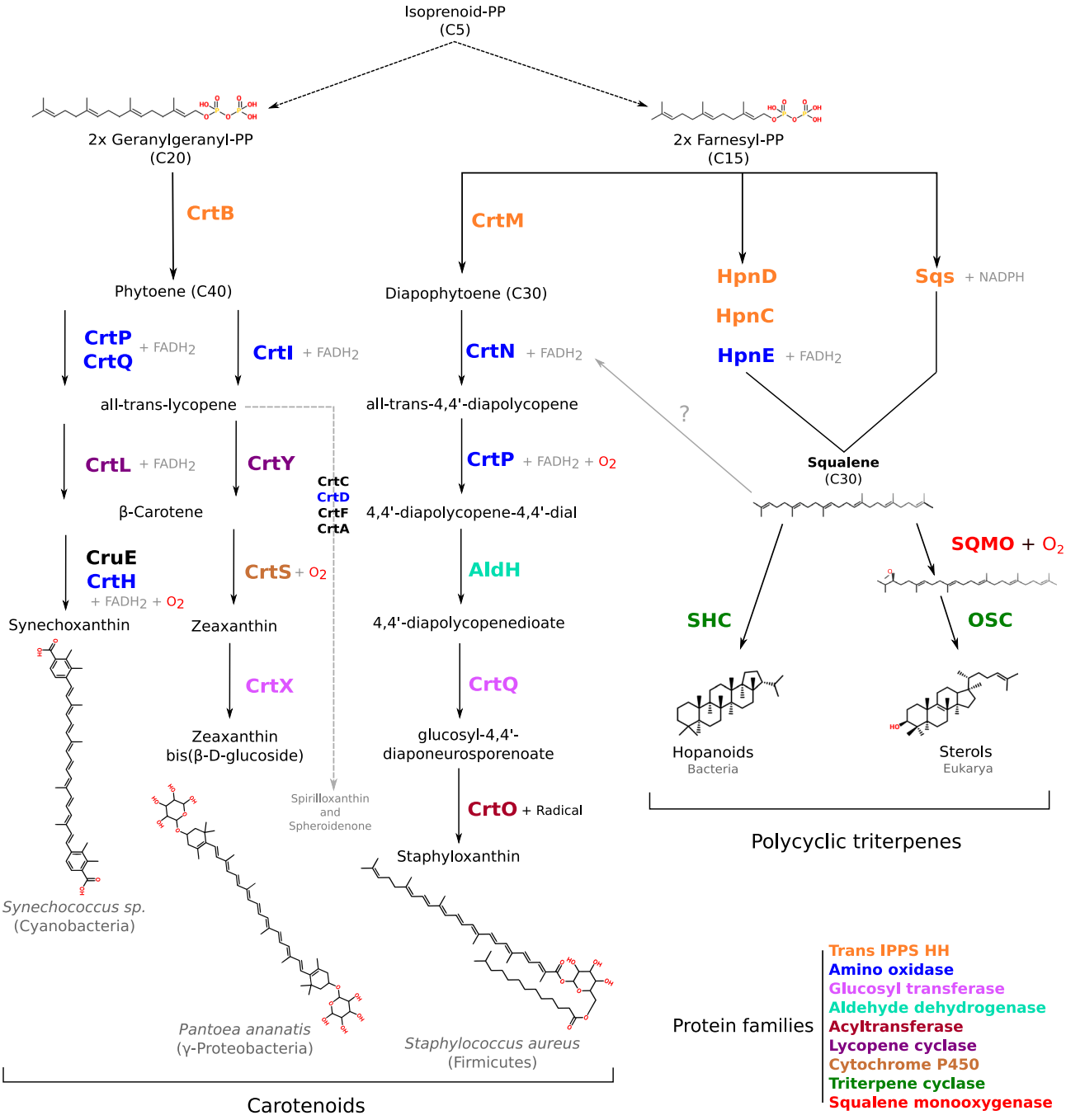

Fig. 1. Schematic of the canonical polycyclic and linear terpenoid biosynthesis pathways. Biosynthetic pathways of polycyclic triterpenes and carotenoids (C30 and C40), starting from farnesyl-PP or geranylgeranyl-PP. Homologous enzymes are shown in the same color. Note that the pathways shown are representative; alternatives are possible.

Due to the physiological relevance of carotenoids as antioxidant, membrane modulators, and in photosynthesis, these molecules have played an important role in the evolution of the tree of life (8–10). Understanding the early evolution of carotenoids biosynthesis pathways is important to trace back the origin of adaptative strategies to oxidative environments such as the one that occurred 2.5 giga years ago during the Great Oxidation Event (GOE) (11). However, the evolutionary relationships between the carotenoid production pathways (C30, C40, and C50), as well as the biosynthetic and evolutionary relationship of carotenoids with squalene remain unclear. In prokaryotes, the evolution and distribution of the carotenoid pathways is currently explained mostly by lateral gene transfer (LGT) events followed by neofunctionalizations. This is presented as a "bramble" model whereby downstream biosynthetic steps exhibit greater evolutionary plasticity and diversification as compared to those upstream, closer to the initial steps of the biosynthetic path (12). The phylogeny of the Trans IPPS HH family has limited resolution when it comes to solve these questions (4). On the other hand, carotenoid amino oxidases still remain as pathway-specific with possibly more phylogenetic signal than Trans IPPS HH and, thus, represent good candidates to solve the origin and evolution of the carotenoid pathways. In this framework, demonstrating the synthesis of C30 carotenoids *via* squalene would fill an important knowledge gap connecting the physiology and evolution of carotenoids with the one of polycyclic triterpenes.

We investigated the synthesis of carotenoids in the Planctomycete *P. limnophila* demonstrating the existence of the squalene route to carotenoid synthesis in nature, and further report its widespread occurrence in *Bacteria*. Together with our evolutionary analyses, these insights put into context the ancestral diversification of terpenoid metabolism comprising polycyclic triterpenes and carotenoids. The widespread occurrence of the squalene route in *Bacteria* decouples squalene from the exclusiveness of the polycyclic triterpenes biosynthesis, to which it has traditionally been associated, and increases its functional repertoire.

Results

Planctomycetes Produce C30 Carotenoids Using Squalene as the Precursor. To decipher carotenoid production in Planctomycetes, we performed random mutagenesis by Tn*5* transposition to select *P. limnophila* colonies with altered pigmentation. The transposition events in the selected clones mapped to the genes *hpnD, hpnE, crtN, crtP, aldH, crtQ*, and *crtO*, comprising all genes computationally identified with the exception of *hpnC* (4). We deleted *hpnC* by directed mutagenesis. The *hpnC, hpnD, hpnE*, and *crtN* mutants resulted in colorless colonies, while the other mutants showed colonies with altered pigmentation, with colors ranging from light yellow to bright red (Fig. 2*A*).

We also constructed a hopanoid-deficient mutant by deleting the *shc* (squalene hopene cyclase) gene. The resulting colonies grew under standard conditions, thus discarding the essentiality of hopanoids in *P. limnophila,* in contrast to the essentiality of sterols in its close relative *G. obscuriglobus* (6). The Δ*shc* mutant colonies displayed an intense red color (Fig. 2*A*), indicating that the accumulated squalene (or its precursors) is re-directed toward carotenoid production, as observed in *M. extorquens* (7, 13). Indeed, we confirmed squalene production *via* HpnCDE using the Δ*crtN*-Δ*shc* strain, since squalene is the precursor of hopanoids (Fig. 2*B*). The production and accumulation of squalene in this mutant was confirmed by gas chromatography with flame-ionization detection (GC-FID). Together with the lack of color in the *hpnCDE* mutants (Fig. 2*A* and *SI Appendix*, Fig. S1), these results support a role for squalene as an intermediate in carotenoid biosynthesis.

To characterize the carotenoid pigments produced by wild-type *P. limnophila* and the selected mutants, the corresponding cell extracts were analyzed by high-performance liquid chromatography (HPLC), and UV-visible spectra were obtained for each peak. Wild-type *P. limnophila* showed a complex HPLC profile with many peaks having UV-visible spectra in agreement with chromophore structures that have 11 to 13 conjugated double bonds (Fig. 2*C*). The fact that most peaks presented almost identical UV-visible spectra but different chromatographic mobility suggested the possibility that these peaks corresponded to an array of different esterified forms. This was analyzed upon alkaline hydrolysis of wild-type extract, which produced a simpler chromatogram (Fig. 2*C*), confirming the esterified nature of the pigments. Moreover, acidification of the hydrolyzed pigments solution was necessary to transfer them to organic solvent (diethyl ether), suggesting that the major pigments in the wild type were acidic carotenoids in which the native form was glycosyl esters, in agreement with previous studies (14). In fact, the spectroscopic (UV-visible and MS) and chromatographic properties of the major peaks in the saponified extract from wild type were in agreement with that of carotenoid acids derived from 4,4′-diapolycopene, a C30 carotenoid observed in other bacterial species such as *Methylobacterium rhodium* (formerly *Pseudomonas rhodos*) (15, 16), *Rubritella squalenifaciens* (17, 18), *Planococcus maritimus* (19), and *Bacillus firmus* (20). In addition, we did not find C40 carotenoids—such as lycopene, neurosporene, and β-carotene—in wild type or mutants. Altogether, these data identify the *P. limnophila* pigments as C30 carotenoids, with the most predominant ones being carotenoid acids derived from 4,4′-diapolycopene (Fig. 2*C*). The additional acylation of the sugar derivatives with different fatty acids by the action of CrtO (acyl transferase) produced the complex pigment profile observed in the wild-type strain (Fig. 2*C*). Additional experimental work is needed to identify both the sugar moieties and the fatty acids involved in the formation of the carotenoid acid glycosyl derivatives responsible for the native carotenoid profile and color of *P. limnophila.* Carotenoids have also been detected in the Planctomycetes *Rhodopirellula rubra* LF2^T and *Rubinisphaera brasiliensis* Gr7 (21). However, the nature and biosynthetic pathways of these carotenoids are still unclear.

To further explore the action of the detected enzymes, pigment extracts from different mutants were analyzed by HPLC-DAD (Fig. 2*D* and *SI Appendix*, Fig. S2). Additionally, the mass spectra for the most predominant compounds were obtained by HPLC-MS (APCI) (*SI Appendix*, Fig. S3). The *hpnC*, *hpnD*, and *hpnE* mutants lacked carotenoids (Fig. 2*A* and *SI Appendix*, Fig. S1). Similarly, carotenoids were absent in the *crtN* mutant (Fig. 2*D*). Additional analyses of the other mutants confirmed the nature of the intermediary products of each genes (*SI Appendix*, Text S1).

To corroborate the synthesis of carotenoids from squalene in *P. limnophila*, we assembled a heterologous expression system in *E. coli* BL21, a strain that lacks both carotenoids and hopanoids. *E. coli* was simultaneously transformed with three plasmids (*SI Appendix*, Table S1). The first plasmid ensured isopentenyl diphosphate (IPP) production (22), to enhance the biosynthesis of isoprenoid derivatives. The second plasmid carried the squalene synthesis genes (*hpnCDE*). Different versions of the third plasmid contained the downstream carotenoid modification genes in an additive fashion (*crtN, crtP, aldH, crtO*, or *crtQ*). In addition to these strains, alternative versions of plasmid 2 were also used as controls. One version contained only the *hpnC* and *hpnD* genes. Colonies containing this plasmid were colorless, confirming that carotenoids were produced *via* squalene and not *via* intermediates of the HpnCDE pathway (*SI Appendix*, Fig. S4). Another version contained the cyanobacterial *sqs* gene for squalene production. An *E. coli* strain containing this plasmid yielded the same carotenoids as the strain with the plasmid bearing the *hpnCDE* genes, although with different proportions, which confirmed that both pathways are equivalent for carotenoid synthesis (*SI Appendix*, Fig. S4). HPLC analysis verified that *E. coli* expressing the whole pathway produced a mixture of carotenoids, including 4,4′-diapolycopene as the most abundant, 4,4′-diapolycopen-4-al, 4,4′-diapolycopen-4′-al-4-oic acid, 4,4′-diapolycopen-4,4′-dioic acid, and 4,4′-diapolycopenoic acid. The glycosyl- and ester-modified carotenoids did not appear in the *E. coli* extracts (*SI Appendix*, Fig. S4), in agreement with the lack of the specific enzymes in the reconstructed system.

Compiling the results from genetic, bioinformatic, and chemical analyses, we propose a pathway for the synthesis of C30 carotenoids in *P. limnophila via* squalene (*SI Appendix*, Fig. S5).

Widespread Distribution of the Squalene Route to C30 Carotenoid in Prokaryotes. As CrtN is, with CrtP, the enzyme specifying C30 carotenoid production, we used this enzyme to detect C30 carotenoid-producing organisms. CrtN was found in *Firmicutes-Bacilli* (including *Clostridia*), few Proteobacterial classes (mostly α- and γ-), *Verrucomicrobia*, *Planctomycetes*, *Fusobacteria*, *Acidobacteria*, *Armatimonadetes*, and in the euryarchaeal class *Candidatus* Poseidoniia (also found in others underrepresented and uncultured phyla; Fig. 3*A* and *SI Appendix*, Fig. S6 and Dataset S1,

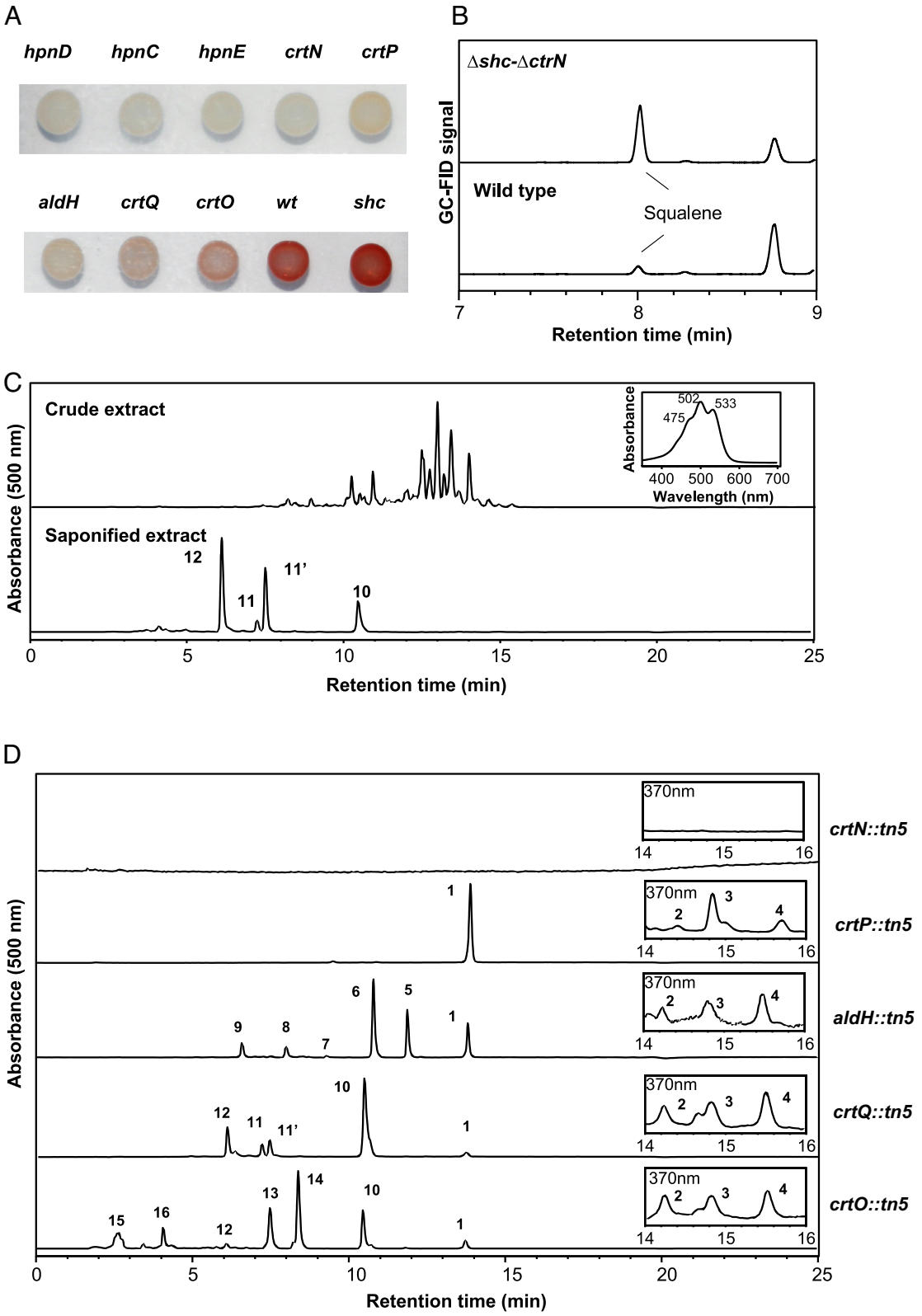

Fig. 2. *P. limnophila* carotenoids analysis. (*A*) Culture drops of wild type and various mutants lacking the indicated genes, (*B*) GC-FID analysis of squalene in wild type and Δ*crtN*-Δ*shc* mutant strains, (*C*) HPLC separation of carotenoid pigments present in crude and saponified (alkaline hydrolysis) extracts obtained from wild-type strain and total UV-visible spectrum for the crude extract. Peak numbers are in accordance with the chromatograms in *D* and the pathway scheme (*SI Appendix,* Fig. S5): 4,4′-diapolycopenoic acid (10); 4,4′-diapolycopen-4′-al-4-oic acid (11 & 11′); 4,4′-diapolycopen-4,4′-dioic acid (12). (*D*) HPLC chromatograms corresponding to the analyses of *P. limnophila* carotenoid mutants. Detection wavelengths at 370 and 500 nm. Peaks: 4,4′-diapolycopene (1); 4,4′-diaponeurosporene (2); 4,4′-diapo-ζ-carotene (3); 4,4′-diapophytofluene (4); 4,4′-diapolycopen-4-al (5); 4,4′-diapolycopene-4-ol (6); ***4,4′-diapolycopene dial (7); 4,4′-diapolycopene-4-ol-4′-al (8); 4,4′-diapolycopene-4,4′-diol (9); 4,4′-diapolycopenoic acid (10); 4,4′-diapolycopen-4′-al-4-oic acid (11 & 11′); 4,4′-diapolycopen-4,4′-dioic acid (12); glycosyl esters of 4,4′-diapolycopenoic acid (13 & 14); and glycosyl esters of 4,4′-diapolycopen-4,4′-dioic acid (15 & 16).

Tab A and B). We then investigated the potential substrate of these carotenoid amino oxidases. The C30-specific amino oxidases usually co-occur with squalene, or 4,4′-diapophytoene, biosynthesis enzymes; with the exception of *Clostridia* which only contains CrtN, *Firmicutes* contains CrtN, CrtP, and CrtM and produces carotenoid *via* this route (Fig. 3*A*). With minor exceptions, such as some *Fusobacteria* or *Fibrobacteria*, all other bacteria containing CrtN contain HpnCDE or Sqs. HpnCDE being the most common pathway of precursor production in *Planctomycetes*, *α-proteobacteria*, and *Acidobacteria*. *Methylococcales* and few members of *Firmicutes* have Sqs instead. In other cases, such as *Verrucomicrobia* and *Candidatus* Poseidoniia, both pathways are found, although *hpnC* and *hpnE* genes are usually lost and only *hpnD* and *sqs* are found, possibly indicating replacement of *hpnCDE* by *sqs* (4). Related, *R. squalenifaciens* is a strong squalene and carotenoid producer (23), but this bacteria only has the *hpnD* gene (manual verification; genome completeness 99.3%), suggesting that this enzyme is enough to produce squalene in this organism or that another pathway is present. This organism is thus an ideal candidate to investigate the unknown roles of HpnD in isolation.

The squalene route to C30 carotenoid production is further supported by the genomic context of *crtN* where *hpnD* is found in the archaeal class *Ca.* Poseidoniia or *hpnCDE* in some

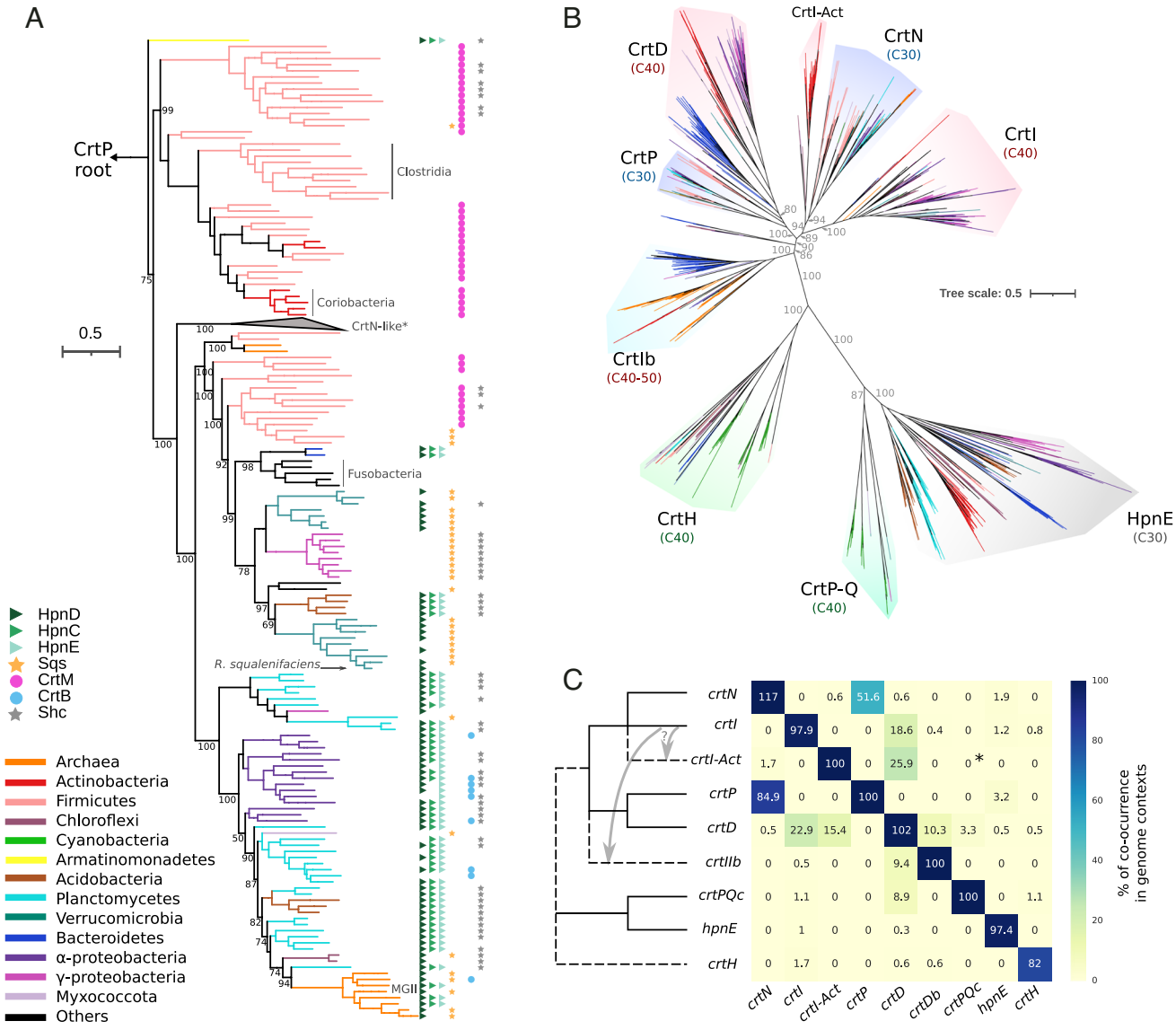

Fig. 3. Evolution of carotenoid pathways based on carotenoid amino oxidases. (*A*) Phylogeny of C30-specific amino oxidase CrtN rooted at CrtP subfamily, and phylogenetic profile of related genes. To ease visualization, the tree was pruned selecting the main taxonomic groups with highest genome completeness, and the Ufboot2 values of the nodes of interest. The respective multiple sequence alignment MSA contained 550 positions and the evolution model used was LG+I+G4. CrtN-like includes few sequences from *Cyanobacteria, Chloroflexi,* and *Myxococcota* likely acting in C40 pathways. Extended version of the original tree with taxonomic labels and support values is provided (*SI Appendix,* Fig. S8). (B) Unrooted phylogeny of carotenoid and squalene amino oxidases. UfBoot2 values for only those nodes of interest are shown, and branches are colored according to the taxonomy. The respective MSA contained 546 positions and the evolution model used was LG+F+I+G4. Note that the taxonomic composition of each carotenoid enzyme was subsampled by taxonomic redundancy (see *Methods*) and manually cleaned removing spurious sequences. (*C*) Heatmap for the co-occurrence of amino oxidases in their respective genomic contexts. Integers indicate the number of amino oxidases present in the genomic context divided by the total number of enzymes for each subfamily; for example, 117 *crtN* genes were found in the genomic context of 100 *crtN*. Amino oxidases were identified using Blastp against a home-made database with a strict threshold ($1e^{-150}$). Asterisk denotes a significant co-occurrence that is detected when the threshold for amino oxidase identification is relaxed ($1e^{-10}$, *SI Appendix,* Fig. S9). Subfamilies are sorted according to the phylogeny (*B*), in which dashed lines denote unclear relationships. Gray arrows suggest possible evolutionary derivation due to cooperativity with other amino oxidase. Raw data for phylogenetic profiles and genome context analyses provided in Dataset S1, Tab A and B.

α-proteobacterial orders (*SI Appendix*, Fig. S7). Thus, our results show that the squalene route to C30 carotenoid production is widespread in Bacteria, and, contrary to common belief, that squalene is the most common precursor to carotenoid synthesis. In addition, the biosynthesis of the precursor (4,4′-diapophytoene or squalene) has shifted between different groups, *i.e.*, *via* HpnCDE, Sqs, or CrtM (Fig. 3).

Evolution of the C30 Carotenoid Pathway. Having shown the taxonomic spread of the squalene route, we reconstructed the evolutionary history of the main enzymes that define the C30 carotenoid biosynthetic pathways, the CrtN and CrtP amino oxidases. The inferred phylogenies of these amino oxidases provided globally congruent topologies, showing that these enzymes mostly have a common evolutionary history (*SI Appendix*, Fig. S6). This is in agreement with the fact that these genes form an operon together with other carotenoid-related genes, such as *aldH, crtQ*, or *crtO*, in distantly related genomes (*SI Appendix*, Fig. S7). CrtN has a broader distribution than CrtP, which could be related to the fact that it is the first amino oxidase enzyme in the pathway, responsible for making the first desaturation of the C30 backbone. The CrtN/P phylogeny shows a taxonomically mixed topology characterized by *Firmicutes-Bacilli* (and *Clostridia,* which contain only CrtN) branching paraphyletically and basally in the respective subfamilies (Fig. 3*A* and *SI Appendix*, Fig. S6). Embedded in this group are other bacterial orders such as *Verrucomicrobiales*, *Acidobacteriales*, and *Methylococcales*, among others. The other main basal branch in the CrtN/P phylogeny is composed by *Planctomycetes* phylum (classes *Planctomycetia* and *Phycisphaerae*), which contains other taxonomic groups embedded paraphyletically. The groups encompass bacterial orders with limited representatives, such as *Rhizobiales*, *Acetobacteriales*, *Rhodobacterales* (*α-proteobacteria*), *Acidobacteriales*, and the euryarchaeal class *Candidatus* Poseidoniia, among others. The latter contains only CrtN (including MGII archaea from *Candidatus* Thermoplasmatota). This taxonomic and phylogenetic distribution of the C30-specific amino oxidases suggests that there have been multiple LGT events between prokaryotes, mainly from *Firmicutes-Bacilli* and *Planctomycetes*. In addition to the main branches of *Firmicutes-Bacilli* and *Planctomycetes*, there is an intermediary branch in the CrtN subfamily, CrtN-like, including a few members of *Myxococcota*, *Cyanobacteria*, and *Chloroflexi*, possibly indicating LGT between these groups. The origin of this branch is unclear, but its position was stable in all reconstructions. The protein from *Myxococcota* has been described to be involved in C40 carotenoid synthesis (24).

The taxonomic distribution of the genes involved in C30 carotenoid synthesis is more limited than the ones of squalene or hopanoids (4). This limited distribution narrows the possible taxonomic origin of this pathways. As Firmicutes are the only bacteria-bearing CrtM, the biosynthesis of C30 carotenoids *via* 4,4′-diapophytoene originated in these organisms. However, this does not imply that *Firmicutes* is at the origin of C30 carotenoids. *Firmicutes* forms the most basal clades in the CrtN/P phylogeny (Fig. 3*A* and *SI Appendix*, Fig. S6), which could be a sign of ancestrality or of more intense divergence. The exclusive presence of CrtN in the anaerobic *Clostridia* might indicate an early origin of the CrtN subfamily. However, the function of CrtN in these organisms remains unclear since they do not have CrtM nor CrtP (*SI Appendix*, Fig. S7 and Dataset S1, Tab A and B). Their monophyly with Bacilli sequences could indicate origin of these enzymes in Firmicutes, or LGTs between these classes. By contrast, the biosynthesis of C30 carotenoids *via* squalene shows an ancestral evolution in Planctomycetes and represents an alternative to the origin of C30 carotenoids that is independent of CrtM. Alternatively, the deep and congruent branching of *Armatimonadetes*, containing the HpnCDE enzymes in the CrtN/P phylogeny, could suggest early origin or faster evolutionary rates. However, insufficient taxonomic sampling makes a clear inference difficult to draw.

Diversification of Carotenoid Biosynthesis Pathways. The carotenoid amino oxidases usually work in pairs, and their evolution provides a broad view of how the different carotenoid pathways evolved (Fig. 3*B*). HpnE and CrtN/P act on C30 backbones, while CrtI/D and CrtP-Qc/H act in the two main pathways of C40 carotenoid biosynthesis. The latter two branch separately suggesting two independent origins of C40 synthesis. One of these pathways, *via* CrtP-Qc and CrtH, is present exclusively in aerobic *Cyanobacteria* and green sulfur anaerobic bacteria (*Chlorobi*; *SI Appendix*, Fig. S8), suggesting LGT between these. While cyanobacterial CrtP-Qc is closely related to HpnE, CrtH is more distantly related to the other families (Fig. 3*B*). Thus, this C40 CrtP-Qc/H pathway most likely originated in *Cyanobacteria* or *Chlorobi* and closely related to HpnE. The other C40 pathway, *via* CrtI/D, most likely had an ancestral evolution in *Proteobacteria*, *Bacteroidetes*, *Actinobacteria*, and *Deinococcus*, among others, suggesting LGT events between the ancestors of these phyla (*SI Appendix*, Fig. S8). While CrtD is more widespread, CrtI is mainly found in *Proteobacteria* and *Actinobacteria* (CrtI of *Actinobacteria* here called CrtI-Act). *Bacteroidetes*, however, has a different carotenoid amino oxidase for C40 carotenoid biosynthesis, here called CrtIb, instead of the "classical" CrtI. This CrtIb has been laterally transferred between the ancestor of *Bacteroidetes*, *Actinobacteria*, *Thermoplasmatota*, *Halobacteria,* and a few members of *Thermoprotei*, among others (*SI Appendix*, Fig. S8). In some cases, such as *Actinobacteria* and *Halobacteria*, this enzyme is involved in C50 carotenoid biosynthesis (12). Halobacteria are expected to produce linear C50 carotenoids, that is independent of CrtI, and their CrtIb do not produce the first desaturation step (25), although it is important to remark that CrtIb is duplicated in *Haloarcula japonica* (Dataset S1, Tab A and B). Remarkably, a monophyletic group of *Archaea* comprising few sequences from *Lokiarchaeia*, *Nitrosphaeria*, *Methanobacteria,* and *Methanomicrobia*, most likely illustrates LGT events between close relatives (*SI Appendix*, Text S2 and Fig. S8). C50 carotenoids are derived from all-*trans*-lycopene (C40) (20), suggesting that the CrtIb subfamily emerged from the CrtN-P (C30) or CrtI-D (C40) subfamilies. In our controls, the phylogenetic location of CrtIb varies from basal to CrtI-N/D-P or associated to the CrtI-N subfamilies (*SI Appendix*, Fig. S9), showing the phylogenetic instability of this subfamily, but with a possible origin from the CrtI subfamily (see below). However, the topology of the CrtI-N and CrtD-P clusters was stable in controls [as in previous analyses (12)], except for some CrtP of *Firmicutes*, branching intermediary between CrtP and CrtD subfamilies (Fig. 3*B*).

We next analyzed the genomic context of the amino oxidases to detect possible mixed pathways. Enzymes with unclear or intermediary positions like CrtIb and CrtI-Act are genomically associated to CrtD (Fig. 3*C*), suggesting that they could have originated from CrtI, but diverged considerably. Relaxing the similarity threshold for the identification of carotenoid amino oxidases in the genomic context (from $1e^{-150}$ to $1e^{-10}$ e-value), we noticed that CrtI-Act is genomically associated to CrtP-Qc-related amino oxidases of the cyanobacterial/chlorobi pathway that were not included in our reconstructions. This result could illustrate an ancestral relationship between the CrtP-Qc/H and CrtD/I

pathways or, alternatively, suggests a shuffling of carotenoid pathways due to LGT events. The C30 genes *crtN* and *crtP* are usually found in close genomic proximity (51.6 and 84.9% in Fig. 3*C*), and *crtN* tends to undergo gene duplications (117%; Fig. 3*C*). Few *crtN*/*P* are associated with *hpnE*, but there is not specific genomic association of C30 amino oxidases with other carotenoids amino oxidases, suggesting that the C30 and C40 amino oxidases do not use the same substrate.

Enzymes of the C30 carotenoid pathway, CrtN-P, are closely related to those of the *Proteobacteria* (and few others) C40 carotenoid pathway, CrtI-D, but distantly related to those of *Cyanobacteria* (Fig. 3*B*). Two facts suggest that these pathways have a common origin, possibly due to neofunctionalization of an operon. First, the pairs CrtN-CrtI and CrtP-CrtD branch together. Second, one Trans IPPS HH enzyme (CrtM or CrtB), two amino oxidases (CrtN-CrtP or CrtD-CrtI), and even a glucosyl-transferase (CrtQ or CrtX) and an acyltransferase are found in the C30 and C40 carotenoid operons. In the phylogeny of carotenoid amino oxidases, CrtN and CrtP are branching closer to the deep nodes (Fig. 3*B*), which suggests that the C30 carotenoid pathway diverged earlier than the C40 pathway. In agreement with this assumption, the direct evolution of the function of CrtB from CrtM has been demonstrated by random mutagenesis, but not the opposite (26). This ancestral origin of C30 carotenoid biosynthesis is further supported by the fact that C30 biosynthesis enzymes belong to widespread protein families, while C40 carotenoids usually undergo cyclization at the extremities of the backbone, which add an additional and specific enzymatic steps involving the lycopene cyclases, the enzymes of the CrtY family (Fig. 1). Together, these observations suggest that the C30 carotenoid pathway is ancestral to C40 (and C50) carotenoid biosynthesis. In addition, *Firmicutes* or *Planctomycetes* are the most likely phyla of origin, depending on whether it originated from 4,4′-diapophytoene or from squalene.

Using the phylogenetic proximity between HpnE and CrtP-Qc and their taxonomic distribution, we attempted to resolve the order of evolution of both subfamilies. The HpnCDE enzymes show a more widespread and ancestral distribution in *Bacteria* [*i.e.,* ancestral in phyla such as *Actinobacteria*, *Planctomycetes,* and *Proteobacteria*, among others (4)] than the CrtP-Qc/H enzymes, which are present only in *Cyanobacteria* and *Chlorobi*, suggesting that squalene production (*via* HpnCDE) predates the cyanobacterial carotenoid pathway. Alternatively, it is also possible that both CrtP-Qc and HpnE subfamilies originated from a common ancestor rather than one from the other, and therefore, the origin of HpnCDE and CrtP-Qc is independent. Nevertheless, the independent origin of CrtP-Qc combined with the fact that Sqs is ancestrally associated to *Cyanobacteria* (4) could illustrate an ancestral co-evolution between the isoprenoid-derivative metabolic pathways assuming that these enzymes originated in *Cyanobacteria*.

Pervasive LGT of Carotenoid Biosynthetic Pathways and Their Patterns of Distribution in Prokaryotes. Finally, we mapped the presence of the main enzymes studied here onto the species tree at two taxonomic levels. At the level of classes, we observed a scattered distribution of these pathways, with the remarkable pattern of the co-occurrence of pathways in close relatives [*e.g.*, *Actinobacteria*, *Chloroflexi Halobacterota*, and *Thermoplasmatota*, among others (Fig. 4 *A* and *B*)]. However, it is important to note that some of the absences are detected in under-represented taxonomic classes, pointing to the possibility of even higher widespread distribution of these pathways in prokaryotes. Nevertheless, the distribution of these enzymes (Fig. 4 *A* and *B*), combined with their phylogenies (Fig. 3 *A* and *B* and *SI Appendix*, Fig. S8), illustrates the prevalence of LGT events in the evolution of carotenoid pathways.

On the other hand, at the level of species (Fig. 4*C*), we observe that most of these co-occurrences have non-overlapping patterns of distribution, *e.g.*, in the *Halobacterota* phylum, *Halobacteria* and *Methanonatronarchaeia* classes have C40 and C30 pathways, respectively, or Methylococcales which switched squalene and carotenoid synthesis to Sqs and C30 pathways or in *Bacteroidetes* presenting both C40 pathways. Conversely, other cases have both complete C30 and C40 pathways, such as some α-proteobacterial orders, although in *M. extorquens* only the C30 pathway is functional in normal conditions (7). Other cases, such as some *Firmicutes* species, present a mix of one complete (C30) and other partial (CrtH) pathway, showing the coexistence of enzymes from different pathways in the same organism. Further analyses are required to clarify whether these mixing enzymes are using the same substrate, or not, although cases such as the *Myxobacteria* CrtN-like suggests that this is a possibility.

Discussion

In this work, we characterized the biosynthesis of C30 carotenoids *via* squalene in the Planctomycete *P. limnophila.* We identified the predicted genes through random mutagenesis and reconstructed the pathway in a heterologous system to corroborate the route. The carotenoid profile of *P. limnophila* is characterized by an array of C30 carotenoids derived from 4,4-diapophytoene, in which unusual red carotenoids with carboxylic acid moieties predominate. This structural feature raises the possibility of industrial and pharmaceutical applications for these molecules, since carotenoic acids have higher polarity and water solubility than common carotenoids.

Phylogenetic analyses of C30 carotenoid amino oxidases, CrtN and CrtP, combined with phylogenetic profiles of precursor enzymes, demonstrate that this squalene route to carotenoid biosynthesis is widespread in prokaryotes. Some organisms are expected to produce carotenoids from squalene *via* Sqs, others *via* HpnCDE, and other have both pathways, although in most of these later cases, only HpnD and Sqs remain. In some clades such as *Clostridria* and *Fusobacteria*, no Trans IPPS HH, such as HpnD, CrtM, nor Sqs, were found, suggesting that the carotenoid amino oxidases encoded in their genomes could have other functions. *Candidatus* Poseidoniia (archaeal MGII) is the only Archaea containing HpnCDE or Sqs, and CrtN, but no CrtP, CrtQ, nor AldH (manual inspection), suggesting that the products of CrtN, intermediaries between squalene and all-trans-4,4′-diapolycopene, could have alternative functions. As the accumulation of squalene is not toxic for *P. limnophila* (shown in the Δ*crtN*-Δ*shc* mutant) and squalene can influence membrane properties by itself [reported in *Halobacteria* (27), fungi (28) and mammals (29)], similar functions of the CrtN products could be expected.

Our evolutionary analyses decipher the biosynthetic and evolutionary relationships between the different carotenoid synthesis pathways and pinpoint their evolutionary origins (Fig. 5). C40 pathways have diversified more (CrtI-CrtD, CrtD-CrtIb, or CrtP/Qc-CrtH) than the unique C30 carotenoid pathway, which diversified at the level of the precursor (4,4′-diapophytoene or squalene). C30 and C40/C50 (*via* CrtI/D) pathways have a common origin, likely by operon neofunctionalization, while the C40 pathway of *Cyanobacteria* originated independently and closely related to the squalene pathway (HpnE). The C30 carotenoid pathway is likely to be ancestral to C40 (and C50) carotenoid biosynthesis, and its most likely phylum of origin is *Firmicutes* or *Planctomycetes*, depending on whether it originated from 4,4′-diapophytoene or from

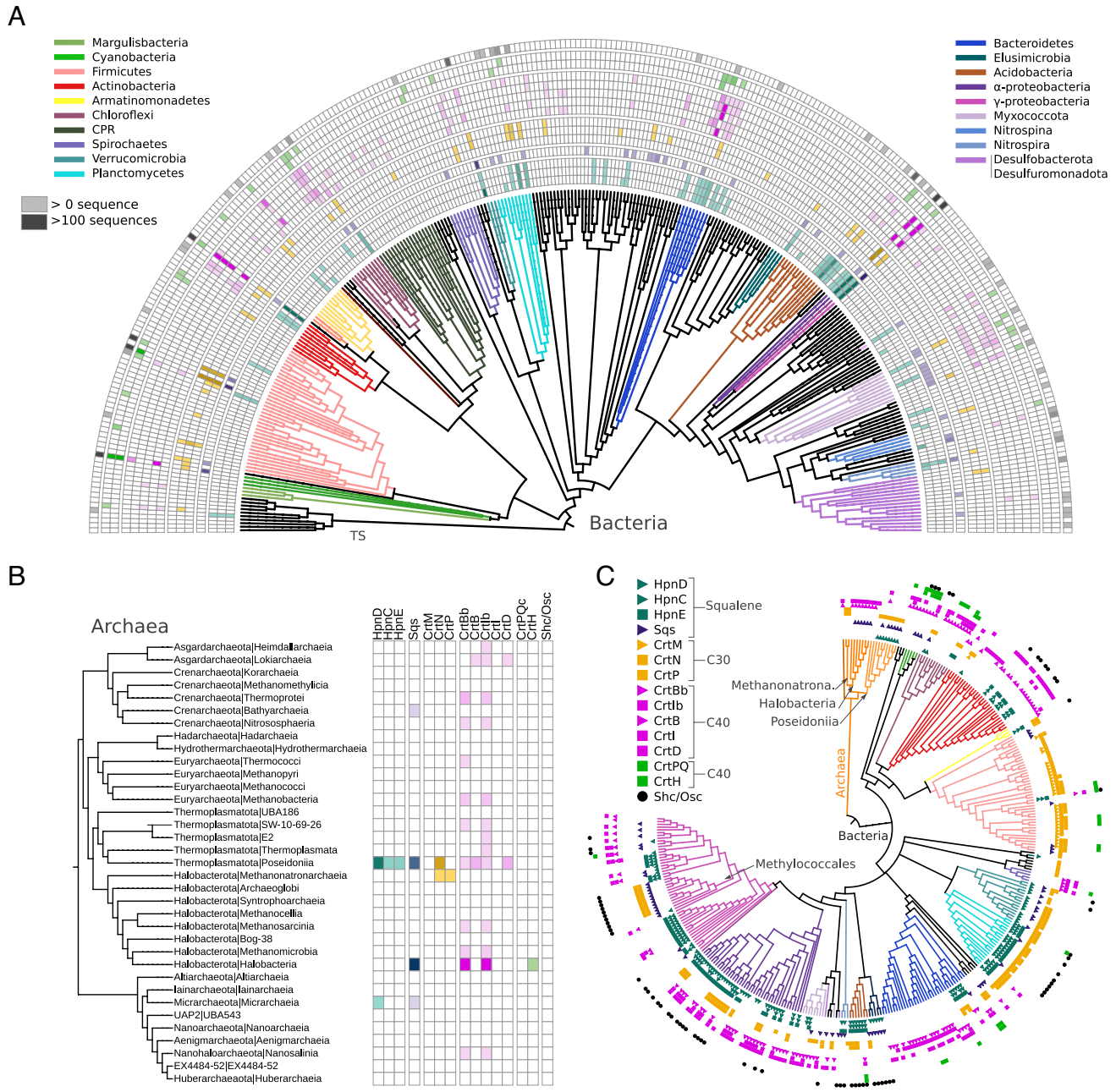

Fig. 4. Taxonomic distribution of the main squalene and carotenoid pathways mapped onto species trees. Phylogenetic profile at the taxonomic class level of (*A*) Bacteria and (*B*) Archaea. The presence of enzymes of different pathways are denoted by colored cells, in which light color indicates that at least one sequence was found and darkest color indicates that more than 100 sequences were found in the respective taxonomic class. Labels in (*A*) and (*B*) follow the same order. TS is comprised by deep bacterial taxa like Thermotogae and Synergistota among others. (*C*) Phylogenetic profile at the species level of selected organisms. Triangles indicate Trans IPPS HHs and squares amino oxidases. Methanonatrona. indicates Methanonatronarchaeia. Taxonomic trees were extracted from GTDB, pruning the leaves of interests. Extended trees with labels and taxonomic representation in database provided in *SI Appendix,* Fig. S10.

squalene. However, the high rates of LGT in these pathways, together with a limited taxonomic sampling, blur the evolutionary inferences about their origins. All these carotenoid pathways have been transferred multiple times between prokaryotes, as operons, and probably neofunctionalized or degenerated, or even metabolically re-adapted. In most squalene-producing organisms also bearing a triterpene cyclase, there could be a physiological synergy between C30 carotenoids and hopanoid production, as sharing the precursor squalene would be an efficient way to generate more complex and versatile membranes with both kinds of molecules.

The combination of the distribution (Fig. 4) and the phylogeny (Fig. 3 *A* and *B*) illustrates the high propensity of squalene and carotenoid pathways to be transferred by LGT. This is also illustrated by individual cases that have C30 and C40 enzymes, or the presence of different pathways within the same taxonomic clade (Fig. 4*C*).

Together, our results demonstrate the existence and widespread occurrence of the squalene route to C30 carotenoid synthesis. This provides insights into the links between carotenoid and polycyclic triterpene metabolisms, and their diversification, demonstrating pervasive LGTs in carotenoids pathways as previously pointed (12). Likewise, our results reveal the evolutionary history of these molecules, which are commonly used as biomarkers, throughout geological time scales (30), but also as biomarkers for oxidative adaptations (31). The widespread occurrence of the squalene route

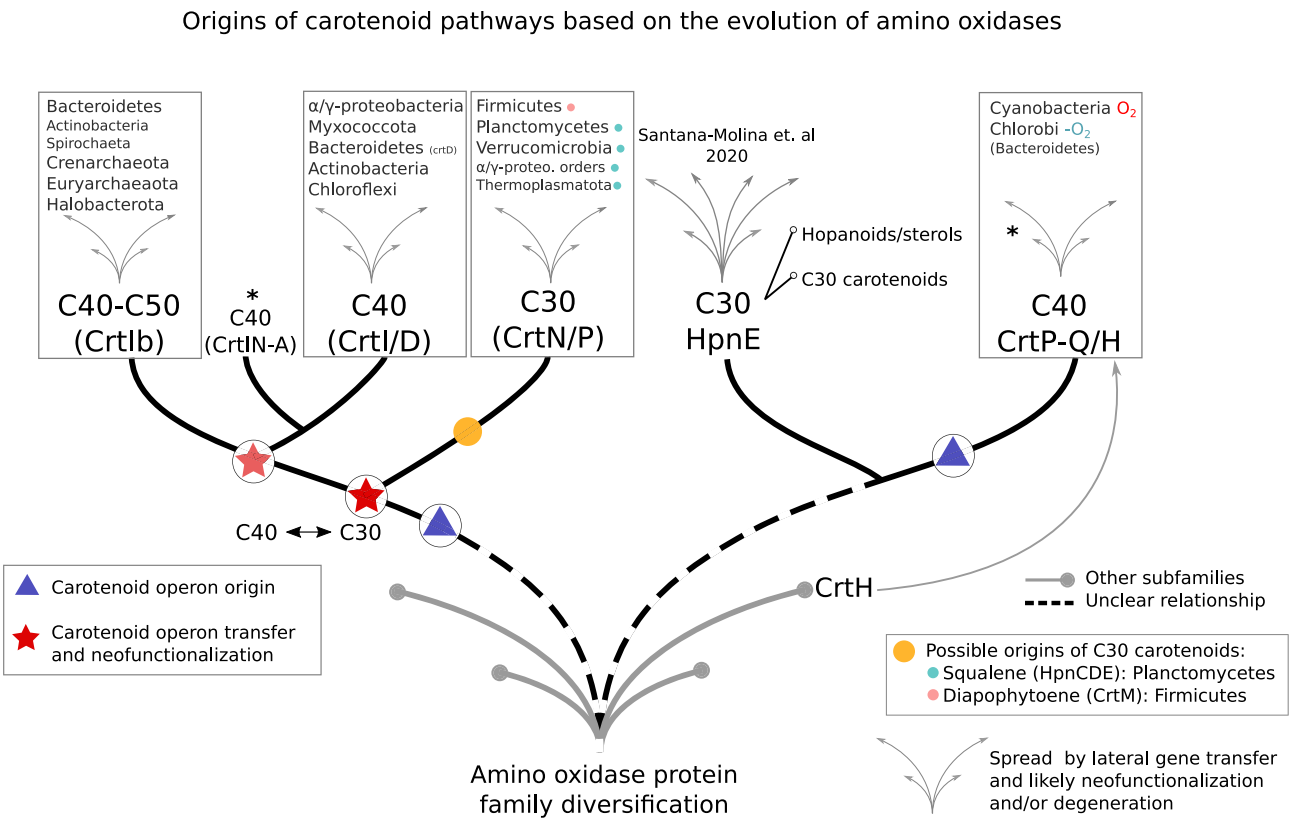

Fig. 5. Evolutionary model for the origin and diversification of carotenoids and squalene Hpn(CD)E pathways. Carotenoid pathways evolved in two main branches, one for C30 and C40-50 pathway of *Proteobacteria* and *Bacteroidetes*, and other for the C40 pathway of *Cyanobacteria* that have close relationship with squalene HpnCD(E) pathway. Only the main taxonomic groups are shown.

to carotenoid biosynthesis increases the functional repertoire of squalene and uncovers an important fraction of carotenoid producers in nature.

Methods

Bacterial Strains and Culture Conditions. The bacterial strains used in this work are listed in *SI Appendix*, Table S1. *Escherichia coli* DH5α, used for cloning purposes, and *E. coli* SoluBL21 (32), used for carotenoid analysis, were grown in lysogeny broth (LB) medium at 37 °C. *E. coli* SoluBL21 bearing the expression plasmids was grown in LB-based ZYM-5052 auto-induction medium (33). *P. limnophila* DSM3776^T was cultivated in M3-modified medium (34), pH 7.5, at 28 °C. We added 1.5% bacto-agar to solid media for *E. coli* and 1% for *P. limnophila*. When required, kanamycin (Km) or gentamycin (Gm) was used at 50 and 20 μg mL^{-1} for *P. limnophila*. For *E. coli*, antibiotics were added at the following concentrations (μg mL^{-1}): Km (25), Gm (10), chloramphenicol (Cl, 15), and ampicillin (Ap, 100). Methyl viologen dichloride hydrate (paraquat, 98% purity) and Isopropil-β-D-1-tiogalactopiranósido (IPTG) were purchased from Merck.

Plasmid Construction. The plasmids, strains, and oligonucleotides used in this work are summarized in *SI Appendix*, Tables S1–S3, respectively. All DNA manipulations were performed using standard protocols. Plasmids used for gene deletion in a double event of homologous recombination were derived from pEX18Tc vector (35), which is a suicide plasmid containing a tetracycline resistance gene. To construct knockout plasmids for the target genes, fragments containing 700 to 1,400 bp sequences of the flanking region of the target genes were amplified by PCR from genomic DNA of *P. limnophila* using the primers summarized in *SI Appendix*, Table S2. The upstream and downstream fragments were cloned into pEX18Tc by three-way ligation using the appropriate restriction enzymes listed in *SI Appendix*, Table S2. Finally, the Km or Gm resistance genes from the pUT-miniTn5km (36) or pBBR1MCS-5 (37) plasmid, respectively, were subsequently cloned as a *Bam*HI fragment between the flanking regions. Plasmids used for the reconstruction of carotenoid synthesis pathway in *E. coli* (*SI Appendix*, Table S3) were divided into three categories: category 1, a plasmid that contains the genes required for the synthesis of the precursor IPP (Addgene 35151(22)); category 2, the plasmids necessary for squalene biosynthesis through Sqs or HpnCDE; and category 3, the plasmids that bear the genes required for carotenoid biosynthesis. All the genes contained in category 2 and 3 plasmids were amplified from genomic DNA using the pair of oligonucleotides listed in *SI Appendix*, Table S2, cut with the appropriate restriction enzyme, and cloned into pBR322 (38) and pSEVA231 (39), respectively. The genes were expressed under the lacV5 promoter and a strong ribosome binding site. More details are presented in *SI Appendix*, Table S3.

Mutant Strain Construction. Genetic transformations of *P. limnophila* for the construction of deletion mutants were performed by electroporation, as previously described (34). In summary, fresh electrocompetent cells were prepared from 400 mL of a culture at OD_{600} 0.4 in modified M3. The cells were washed twice with ice-cold double-distilled sterile water (100 mL and then 50 mL) and once with 2 mL of ice-cold 10% glycerol. The pellet was then resuspended in 400 μL of ice-cold 10 % glycerol, and aliquots of 100 μL were dispensed into 0.1-mm-gapped electroporation cuvettes along with 1 μg of plasmid DNA and 1 μL of type-one restriction inhibitor (Epicentre). Electroporation was performed with a Bio-Rad Micropulser (Ec3 pulse, voltage [V] 3.0 kV). Electroporated cells were immediately recovered in 1 mL of cold fresh medium and incubated at 28 °C for 2 h with shaking. The cells were then plated onto agar plates supplemented with Km or Gm and were incubated at 28 °C until colony formation after approximately 7 d. Colonies were segregated onto fresh selection plates and genotyped by PCR and sequencing.

For random mutagenesis by transposition, 1 μL of EZ-Tn*5* solution and 1 mL of type-one restriction inhibitor were electroporated following the aforementioned protocol. The cells were then plated onto modified M3 supplemented with Km and incubated at 28 °C until colony formation. White colonies were segregated onto fresh selection plates. To verify Tn*5* insertions and their locations, DNA was isolated using the Wizard Genomic DNA Purification Kit (Promega) and analyzed by semirandom PCR (40). Genomic DNA was used as the template DNA in a 20 μL PCR mixture containing primer Map Tn5 A fwd and a mix of primers CEKG 2A,

CEKG 2B, and CEKG 2C; 1 μL of a 1:5 dilution of this reaction mixture was used as the template DNA for a second PCR performed with primers Map Tn5 B fwd and CEKG 4. For the first reaction, the thermocycler conditions were 94 °C for 2 min; followed by six cycles of 94 °C for 30 s, 42 °C for 30 s (with the temperature reduced 1 °C per cycle), and 72 °C for 3 min; and then 25 cycles of 94 °C for 30 s, 65 °C for 30 s, and 72 °C for 3 min. For the second reaction, the thermocycler conditions were 30 cycles of 94 °C for 30 s, 65 °C for 30 s, and 72 °C for 3 min. The DNA of purified PCR products (GFX PCR DNA and Gel Band Purification Kit GE Healthcare) was sequenced using primer Map Tn5 B fwd.

Carotenoid Pathway Reconstruction. To construct the different *E. coli* expression strains, the appropriate plasmids (*SI Appendix*, Table S3) were transformed into *E. coli* SoluBL21 by heat shock. SoluBL21 competent cells were prepared by TSS methods (41). Cells were plated in LB containing Cl, Ap, and Km. When liquid cultures were required, preinocula with the appropriate antibiotics were grown in LB at 37 °C until saturation. Once grown, the cultures were diluted to OD_{600} 0.1 in fresh media and then grown at 37 °C until OD_{600} 0.4, at which point they were induced with IPTG 0.5 mM (Merck) and incubated at 28 °C for 48 h. Cell were collected by centrifugation at 6,000 ×g and 4 °C, and the pellets were kept at –80 °C.

Carotenoid Production and Extraction. For pigment production, strain cultures (250 to 1,000 mL flasks) were performed according to respective *P. limnophila* and *E. coli* culture conditions. Cells were then harvested by centrifugation at 5,000 ×g and 4 °C, and the pellets were washed with phosphate buffer 1× (6.05 g L^{-1} of $Na_2HPO_4.12H_2O$ and 1.0 g L^{-1} of KH_2PO_4), frozen at –80 °C, and lyophilized (VirTis BenchTop 2 K Freeze Dryer, SP Industries Inc.). Approximately 0.15 g of lyophilized biomass was sequentially extracted with ethanol, methanol, and acetone (5 mL each) until no more color was extracted. Extraction was aided by vortex shaking for 1 min and sonication for 30 s. Extraction fractions were collected after centrifugation of samples (5,000×g at 4 °C), the solvent was evaporated to dryness under vacuum in a rotary evaporator (<30 °C), and the dry extract was dissolved in acetone-ethanol (1:1) for chromatographic analysis. We performed all the operations under dimmed light to avoid isomerization and photo-degradation of carotenoid pigments.

Carotenoid Identification. Carotenoid identification was based on the chromatographic and UV-visible spectroscopic (UV-visible and mass spectrometry) data obtained by HPLC coupled with a diode array detector (HPLC-DAD) and HPLC coupled with a mass spectrometer (HPLC-MS(APCI)). Data were compared with those of literature values (14,20,42–47). HPLC-DAD analysis was carried out using a Waters e2695 Alliance chromatograph fitted with a Waters 2998 photodiode array detector and controlled with Empower2 software (Waters Cromatografía, SA). The separation was performed in a reverse-phase C18 (20 mm × 4.6 mm i.d., 3 μm, Mediterranea SEA18; Teknokroma) fitted with a guard column of the same material (10 mm × 4.6 mm). The chromatographic method used was previously described in Delgado-Pelayo et al. (39), although we added formic acid (0.1% final concentration) to the mobile phase. Briefly, carotenoid separation was carried out by a binary-gradient elution using an initial composition of 75% acetone and 25% deionized water (both containing 0.1% formic acid), which was increased linearly to 95% acetone in 10 min, held for 7 min, raised to 100% in 3 min, and held for 10 min. Initial conditions were reached in 5 min. The temperature of the column was kept at 25 °C, and the sample compartment was refrigerated at 15 °C. An injection volume of 10 μL and a flow rate of 1 mL min^{-1} were used. Detection was performed at 500 nm for major pigments and 370 nm for early precursors, and the online spectra were acquired in the 350 to 700 nm wavelength range. HPLC-MS(APCI) analysis was carried out with a Dionex Ultimate 3000RS U-HPLC (Thermo Fisher Scientific) coupled in series with a diode array detector (DAD) and a micrOTOF-QII high-resolution time-of-flight mass spectrometer (UHR-TOF) with qQ-TOF geometry (Bruker Daltonics) and fitted with an APCI (atmospheric pressure chemical ionization) source. The chromatographic conditions were identical to those described for HPLC-DAD analysis. A flow-split of the eluent from the DAD detector was set up in order to allow a 0.4 mL/min flow rate directly into the mass spectrometer (connected in series after the DAD detector). The instrument control was performed using Bruker Daltonics Hystar 3.2 software, and data evaluation was performed with the Bruker Daltonics DataAnalysis 4.0 software. The MS parameters were set as follows: positive mode; current corona, 4,000 nA; source (vaporizer) temperature, 350 °C; drying gas, N2; gas temperature, 250 °C; gas flow, 4 L/min; nebulizer pressure, 60 psi; scan range of m/z 50 to 1,200.

For alkaline hydrolysis of carotenoid extracts from *P. limnophila* wild type, 1 mL of crude extract was evaporated to dryness under a nitrogen stream, and the residue was dissolved in 3 mL of 0.25 N NaOH (aqueous) and left to react for 24 h at room temperature (<25 °C) in the dark. The mixture was acidified with formic acid and the pigments recovered with diethyl ether. The ether phase was collected, evaporated under nitrogen stream, and dissolved in acetone-ethanol (1:1) for chromatographic analysis.

Analysis of Squalene by Gas Chromatography. Lyophilized biomass pellet (0.1 g) was submitted to alkaline hydrolysis with 2 mL of 2 % (w/v) KOH-ethanol at 80 °C for 15 min. Squalane (20 μL; stock solution 10.8 mg mL^{-1}) was added as internal standard. After cooling to room temperature, the mixture was diluted with 3 mL of distilled water and extracted with 1 mL n-hexane. An aliquot of the upper hexane phase (0.5 mL) was transferred to a vial for GC-FID analysis. Gas chromatography analysis was performed on an Agilent Technologies 7890A gas chromatograph (Agilent Technologies España) fitted with a flame ionization detector, a split/splitless injector, and a 7683B series automatic liquid sampler. The chromatograph was fitted with a HP-5 capillary column (J&W Scientific; 30 m length; 0.32 mm i.d.; 0.25 μm thickness). Helium was used as carrier gas with a constant linear flow of 1.75 mL min^{-1}. The injector and detector temperatures were 300 °C and 325 °C, respectively. The oven temperature started at 250 °C and increased at a rate of 4 °C min^{-1} to 270 °C, where it was held for 3 min. The injection volume was 1 μL at a split ratio of 1:20.

Phylogeny. To infer the evolution of carotenoid amino oxidases, we searched for homologous sequences to amino oxidases such as HpnE, CrtN, and CrtI. We performed PHMMER searches (48) with an e-value threshold of $1e^{-5}$, against a local database containing NCBI proteomes of those organisms described in GTDB [version bac120 and ar122 (49)]. We combined the resulting target sequences (~14,000) into a single dataset, aligned them using MAFFT (50), and trimmed gap positions using trimAL (–gt 0.2) (51) and some other non-informative regions manually. With this alignment, we then performed a guide tree using Fasttree (default parameters) (52) and selected the subfamilies of interest according to the presence of characterized enzymes. For each subfamily, we removed redundant sequences for each taxonomic class using cd-hit (53), by applying different cut-offs depending on the number of sequences (from 20 sequences, 95% identity, up to 250 sequences, 55%). These individual reduced subfamilies were again combined to perform the final phylogenies of carotenoid/squalene amino oxidases and those for CrtN/P. We aligned these datasets using MAFFT-linsi, trimmed gap positions, and removed spurious sequences and unstable groups. For example, in order to clarify the relationship between the main subfamilies, we removed divergent CrtI-like sequences from *Deionococcus*, that can influence the topology of carotenoid amino oxidase phylogeny (*SI Appendix*, Fig. S6). These final alignments were used for phylogenetic reconstructions using IQ-TREE (54). We obtained branch supports with the ultrafast bootstrap (55), and the evolutionary models of each set of sequences were automatically selected using ModelFinder (56) and chosen according to BIC criterion. All trees were visualized and annotated using iTOL (57).

For the phylogenetic profiles mapped onto the CrtN/P phylogeny and those provided in Dataset S1, Tab A and B, we made use of a previously defined dataset for HpnCDE, CrtB/M, Sqs, and Shc subfamilies (4), in combination with the carotenoid amino oxidases identified in this study.

Genome Context. We defined the genome context as the arrangement of neighboring genes relative to the gene of interest. To analyze the genome contexts of genes containing the amino oxidase domain (PF01593), we extracted the genomic sequence of the genes 10 Kb upstream and downstream. We extracted the coding DNA sequence from these genomic fragments using PRODIGAL (58), annotated the coding proteins using the Pfam database (59) running HMMSCAN (48), and parsed the output to keep the longest coverage and best e-value in order to minimize the effect of overlapping domains. To identify co-occurrence of the amino oxidases in the genomic context (Dataset S1, Tab C), all the coding genes containing the amino oxidase Pfam domain were searched against a homemade database of the different carotenoid amino oxidase subfamilies identified in this study, using blastp with $1e^{-150}$ and $1e^{-10}$ as the e-value threshold (60). Note that some carotenoid amino oxidases were

not annotated with amino oxidase Pfam domain (annotated as *NAD_binding_8* instead), which were not included in the analyses and explain the percentages below 100 in the reciprocal hits in Fig. 3*C*.

Data, Materials, and Software Availability. All data needed to evaluate the conclusions in the paper are present in the paper and/or the *SI Appendix*.

ACKNOWLEDGMENTS. This work was supported by the Spanish Ministry of Economy and Competitiveness (Grant No. BFU2016-78326-Pand and MDM-2016-0687). This research is funded in part by the Gordon and Betty Moore foundation through Grant GBMF9733 to D.P.D. We also thank Berend Snel and Anja Spang for financial support of C.S.-M. through a grant funded by Utrecht University and the Netherlands Organisation for Scientific Research (NWO).

Author affiliations: [a]Centro Andaluz de Biología del Desarrollo, Consejo Superior de Investigaciones Científicas, Campus Universidad Pablo de Olavide, 41013, Seville, Spain; and [b]Instituto de la Grasa, Consejo Superior de Investigaciones Científicas, Campus Universidad Pablo de Olavide, 41013, Seville, Spain

Author contributions: C.S.-M., D.H.-M., D.P.D., and E.R.-M. designed research; C.S.-M., V.H., D.H.-M., and E.R.-M. performed research; C.S.-M., V.H., D.H.-M., D.P.D., and E.R.-M. analyzed data; and C.S.-M., D.H.-M., D.P.D., and E.R.-M. wrote the paper.

1. M. Rodriguez-Concepcion *et al.*, A global perspective on carotenoids: Metabolism, biotechnology, and benefits for nutrition and health. *Prog. Lipid Res.* **70**, 62–93 (2018).
2. M. Avalos *et al.*, Biosynthesis, evolution and ecology of microbial terpenoids. *Nat. Prod. Rep.* **23**, 249–272 (2021), 10.1039/D1NP00047K.
3. G. Ourisson, Y. Nakatani, The terpenoid theory of the origin of cellular life: The evolution of terpenoids to cholesterol. *Chem. Biol.* **1**, 11–23 (1994).
4. C. Santana-Molina, E. Rivas-Marin, A. M. Rojas, D. P. Devos, Origin and evolution of polycyclic triterpene synthesis. *Mol. Biol. Evol.* **2020**, **37** 1925–1941 (2020), 10.1093/molbev/msaa054.
5. M. Furubayashi, L. Li, A. Katabami, K. Saito, D. Umeno, Construction of carotenoid biosynthetic pathways using squalene synthase. *FEBS Lett.* **588**, 436–442 (2014).
6. E. Rivas-Marin *et al.*, Essentiality of sterol synthesis genes in the planctomycete bacterium Gemmata obscuriglobus. *Nat. Commun.* **10**, 2916–2916 (2019).
7. S. Rizk *et al.*, Functional diversity of isoprenoid lipids in Methylobacterium extorquens PA1. *Mol. Microbiol.* **116**, 1064–1078 (2021).
8. W. K. Subczynski, E. Markowska, W. I. Gruszecki, J. Sielewiesiuk, Effects of polar carotenoids on dimyristoylphosphatidylcholine membranes: A spin-label study. *Biochim. Biophys. Acta* **1105**, 97–108 (1992).
9. B. Mostofian, Q. R. Johnson, J. C. Smith, X. Cheng, Carotenoids promote lateral packing and condensation of lipid membranes. *Phys. Chem. Chem. Phys.* **22**, 12281–12293 (2020).
10. A. Sedoud *et al.*, The cyanobacterial photoactive orange carotenoid protein is an excellent singlet oxygen quencher. *Plant Cell* **26**, 1781–1791 (2014).
11. A. Bekker *et al.*, Dating the rise of atmospheric oxygen. *Nature* **427**, 117–120 (2004).
12. J. L. Klassen, Phylogenetic and evolutionary patterns in microbial carotenoid biosynthesis are revealed by comparative genomics. *PLoS One* **5**, e11257 (2010).
13. A. S. Bradley *et al.*, Hopanoid-free Methylobacterium extorquens DM4 overproduces carotenoids and has widespread growth impairment. *PLOS One* **12**, e0173323 (2017).
14. S. H. Kim, P. C. Lee, Functional expression and extension of staphylococcal staphyloxanthin biosynthetic pathway in Escherichia coli. *J. Biol. Chem.* **287**, 21575–21583 (2012).
15. H. Kleinig, R. Schmitt, W. Meister, G. Englert, H. Thommen, New C30- carotenoic acid glucosyl esters from Pseudomonas rhodos. *Z. Naturforsch. C* **34**, 181–185 (1979).
16. H. Kleinig, R. Schmitt, On the biosynthesis of C_{30} carotenoic acid glucosyl esters in Pseudomonas rhodos. Analysis of car-Mutants. *Z. Naturforsch. C* **37**, 758–760 (1982).
17. K. Shindo *et al.*, Rare carotenoids, (3R)-saproxanthin and (3R,2′S)-myxol, isolated from novel marine bacteria (Flavobacteriaceae) and their antioxidative activities. *Appl. Microbiol. Biotechnol.* **74**, 1350–1357 (2007).
18. K. Shindo *et al.*, Diapolycopenedioic Acid Xylosyl Esters A, B, and C, Novel Antioxidative Glyco-C30-carotenoic Acids Produced by a New Marine Bacterium Rubritalea squalenifaciens. *J. Antibiotics* **61**, 185–191 (2008).
19. K. Shindo *et al.*, Methyl glucosyl-3,4-dehydro-apo-8′-lycopenoate, a novel antioxidative Glyco-C30-carotenoic acid produced by a marine bacterium planococcus maritimus. *J. Antibiotics* **61**, 729–735 (2008).
20. S. Steiger, L. Perez-Fons, P. D. Fraser, G. Sandmann, Biosynthesis of a novel C30 carotenoid in Bacillus firmus isolates. *J. Appl. Microbiol.* **113**, 888–895 (2012).
21. N. Kallscheuer *et al.*, Pink- and orange-pigmented Planctomycetes produce saproxanthin-type carotenoids including a rare C45 carotenoid. *Environ. Microbiol. Rep.* **11**, 741–748 (2019).
22. P. P. Peralta-Yahya *et al.*, Identification and microbial production of a terpene-based advanced biofuel. *Nat. Commun.* **2**, 483–483 (2011).
23. H. Kasai *et al.*, Rubritalea squalenifaciens sp. nov., a squalene-producing marine bacterium belonging to subdivision 1 of the phylum 'Verrucomicrobia'. *Int. J. Syst. Evol. Microbiol.* **57**, 1630–1634 (2007).
24. A. A. Iniesta, M. Cervantes, F. J. Murillo, Cooperation of two carotene desaturases in the production of lycopene in Myxococcus xanthus. *FEBS J.* **274**, 4306–4314 (2007).
25. Y. Yang *et al.*, Complete biosynthetic pathway of the C50 carotenoid bacterioruberin from lycopene in the extremely halophilic archaeon Haloarcula japonica. *J. Bacteriol.* **197**, 1614–1623 (2015).
26. U. Daisuke, A. V. Tobias, F. H. Arnold, Evolution of the C30 carotenoid synthase CrtM for function in a C40 pathway. *J. Bacteriol.* **184**, 6690–6699 (2002).
27. S. F. Gilmore *et al.*, Role of squalene in the organization of monolayers derived from lipid extracts of Halobacterium salinarum. *Langmuir* **29**, 7922–7930 (2013).
28. E. Gnamusch, C. Kalaus, C. Hrastnik, F. Paltauf, G. Daum, Transport of phospholipids between subcellular membranes of wild-type yeast cells and of the phosphatidylinositol transfer protein-deficient strain Saccharomyces cerevisiae sec 14. *Biochim. Biophys. Acta* **1111**, 120–126 (1992).
29. J. A. Nathan, Squalene and cholesterol in the balance at the ER membrane. *Proc. Natl. Acad. Sci. U.S.A.* **117**, 8228–8230 (2020).
30. R. E. Summons, P. V. Welander, D. A. Gold, Lipid biomarkers: Molecular tools for illuminating the history of microbial life. *Nat. Rev. Microbiol.* **20**, 174–185 (2021), 10.1038/s41579-021-00636-2.
31. M. Khademian, J. A. Imlay, How microbes evolved to tolerate oxygen. *Trends Microbiol.* **29**, 428–440 (2021).
32. M. R. Dyson, S. P. Shadbolt, K. J. Vincent, R. L. Perera, J. McCafferty, Production of soluble mammalian proteins in Escherichia coli: Identification of protein features that correlate with successful expression. *BMC Biotechnol.* **4**, 32 (2004).
33. F. W. Studier, Protein production by auto-induction in high-density shaking cultures. *Protein Expr. Purif.* **41**, 207–234 (2005).
34. E. Rivas-Marín, I. Canosa, E. Santero, D. P. Devos, Development of genetic tools for the manipulation of the planctomycetes. *Front. Microbiol.* **7**, 914 (2016).
35. T. T. Hoang, R. R. Karkhoff-Schweizer, A. J. Kutchma, H. P. Schweizer, A broad-host-range Flp-FRT recombination system for site-specific excision of chromosomally-located DNA sequences: Application for isolation of unmarked Pseudomonas aeruginosa mutants. *Gene* **212**, 77–86 (1998).
36. M. Herrero, V. de Lorenzo, K. N. Timmis, Transposon vectors containing non-antibiotic selection markers for cloning and stable chromosomal insertion of foreign DNA in Gram-negative bacteria. *J. Bacteriol.* **172**, 6557–6567 (1990).
37. M. E. Kovach *et al.*, Four new derivatives of the broad-host-range cloning vector pBBR1MCS, carrying different antibiotic-resistance cassettes. *Gene* **166**, 175–176 (1995).
38. F. Bolivar *et al.*, Construction and characterization of new cloning vehicle. II. A multipurpose cloning system. *Gene* **2**, 95–113 (1977).
39. R. Silva-Rocha *et al.*, The Standard European Vector Architecture (SEVA): A coherent platform for the analysis and deployment of complex prokaryotic phenotypes. *Nucleic Acids Res.* **41**, D666–D675 (2013).
40. L. A. Gallagher, C. Manoil, *Pseudomonas aeruginosa* PAO1 Kills *Caenorhabditis elegans* by cyanide poisoning. *J. Bacteriol.* **183**, 6207–6214 (2001).
41. C. T. Chung, S. L. Niemela, R. H. Miller, One-step preparation of competent Escherichia coli: Transformation and storage of bacterial cells in the same solution. *Proc. Natl. Acad. Sci. U.S.A.* **86**, 2172–2175 (1989).
42. G. Britton, "General carotenoid methods" in *Methods in Enzymology*, (Academic Press, 1985), pp. 113–149.
43. P. C. Lee, A. Z. R. Momen, B. N. Mijts, C. Schmidt-Dannert, Biosynthesis of structurally novel carotenoids in Escherichia coli. *Chem. Biol.* **10**, 453–462 (2003).
44. L. Tao, A. Schenzle, J. M. Odom, Q. Cheng, Novel carotenoid oxidase involved in biosynthesis of 4,4′-diapolycopene dialdehyde. *Appl. Environ. Microbiol.* **71**, 3294–3301 (2005).
45. G. Britton, S. Liaaen-Jensen, H. Pfender, "Special Molecules, Special properties" in *Carotenoids: Natural Functions* (2008). Birkhäuser, Basel.
46. J. Garrido-Fernández, A. Maldonado-Barragán, B. Caballero-Guerrero, D. Hornero-Méndez, J. L. Ruiz-Barba, Carotenoid production in Lactobacillus plantarum. *Int. J. Food Microbiol.* **140**, 34–39 (2010).
47. A. Osawa, K. Iki, G. Sandmann, K. Shindo, 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 (2013).
48. S. C. Potter *et al.*, HMMER web server: 2018 update. *Nucleic Acids Res.* **46**, W200–W204 (2018).
49. D. H. Parks *et al.*, A standardized bacterial taxonomy based on genome phylogeny substantially revises the tree of life. *Nat. Biotechnol.* **36**, 996–1004 (2018).
50. K. Katoh, K. Misawa, K. Kuma, T. Miyata, MAFFT: A novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res.* **30**, 3059–3066 (2002).
51. S. Capella-Gutiérrez, J. M. Silla-Martínez, T. Gabaldón, trimAl: A tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* **25**, 1972–1973 (2009).
52. M. N. Price, P. S. Dehal, A. P. Arkin, FastTree: Computing large minimum evolution trees with profiles instead of a distance matrix. *Mol. Biol. Evol.* **26**, 1641–1650 (2009).
53. W. Li, A. Godzik, Cd-hit: A fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics* **22**, 1658–1659 (2006).
54. B. Q. Minh *et al.*, IQ-TREE 2: New models and efficient methods for phylogenetic inference in the genomic era. *Mol. Biol. Evol.* **37**, 1530–1534 (2020).
55. D. T. Hoang, O. Chernomor, A. von Haeseler, B. Q. Minh, L. S. Vinh, UFBoot2: Improving the ultrafast bootstrap approximation. *Mol. Biol. Evol.* **35**, 518–522 (2018).
56. S. Kalyaanamoorthy, B. Q. Minh, T. K. F. Wong, A. von Haeseler, L. S. Jermiin, ModelFinder: Fast model selection for accurate phylogenetic estimates. *Nat. Methods* **14**, 587–589 (2017).
57. I. Letunic, P. Bork, Interactive tree of life (iTOL) v4: Recent updates and new developments. *Nucleic Acids Res.* **47**, W256–W259 (2019).
58. D. Hyatt *et al.*, Prodigal: Prokaryotic gene recognition and translation initiation site identification. *BMC Bioinform.* **11**, 119 (2010).
59. R. D. Finn, J. Clements, S. R. Eddy, HMMER web server: Interactive sequence similarity searching. *Nucleic Acids Res.* **39**, W29–W37 (2011).
60. C. Camacho *et al.*, BLAST+: Architecture and application. *BMC Bioinform.* **10**, 421 (2009).