Genomics 95 (2010) 120–128

Contents lists available at ScienceDirect

Genomics

journal homepage: www.elsevier.com/locate/ygeno

Genome mining of mycosporine-like amino acid (MAA) synthesizing and non-synthesizing cyanobacteria: A bioinformatics study

Shailendra P. Singh [a], Manfred Klisch [a], Rajeshwar P. Sinha [b], Donat-P. Häder [a,*]

[a] Department for Biology, Chair of Plant Ecophysiology, Friedrich-Alexander University Erlangen-Nuremberg, Staudtstrasse 5, D-91058 Erlangen, Germany
[b] Laboratory of Photobiology and Molecular Microbiology, Centre of Advanced Study in Botany, Banaras Hindu University, Varanasi-221005, India

ARTICLE INFO

Article history:
Received 20 April 2009
Accepted 15 October 2009
Available online 29 October 2009

Keywords:
Cyanobacteria
Genome mining
MAA biosynthesis
Phylogenetic analysis
Ultraviolet radiation

ABSTRACT

Mycosporine-like amino acids (MAAs) are a family of more than 20 compounds having absorption maxima between 310 and 362 nm. These compounds are well known for their UV-absorbing/screening role in various organisms and seem to have evolutionary significance. In the present investigation we tested four cyanobacteria, e.g., *Anabaena variabilis* PCC 7937, *Anabaena* sp. PCC 7120, *Synechocystis* sp. PCC 6803 and *Synechococcus* sp. PCC 6301, for their ability to synthesize MAA and conducted genomic and phylogenetic analysis to identify the possible set of genes that might be involved in the biosynthesis of these compounds. Out of the four investigated species, only *A. variabilis* PCC 7937 was able to synthesize MAA. Genome mining identified a combination of genes, YP_324358 (predicted DHQ synthase) and YP_324357 (O-methyltransferase), which were present only in *A. variabilis* PCC 7937 and missing in the other studied cyanobacteria. Phylogenetic analysis revealed that these two genes are transferred from a cyanobacterial donor to dinoflagellates and finally to metazoa by a lateral gene transfer event. All other cyanobacteria, which have these two genes, also had another copy of the DHQ synthase gene. The predicted protein structure for YP_324358 also suggested that this product is different from the chemically characterized DHQ synthase of *Aspergillus nidulans* contrary to the YP_324879, which was predicted to be similar to the DHQ synthase. The present study provides a first insight into the genes of cyanobacteria involved in MAA biosynthesis and thus widens the field of research for molecular, bioinformatics and phylogenetic analysis of these evolutionary and industrially important compounds. Based on the results we propose that YP_324358 and YP_324357 gene products are involved in the biosynthesis of the common core (deoxygadusol) of all MAAs.

© 2009 Elsevier Inc. All rights reserved.

Introduction

Rapid industrialization in the past few decades has resulted in an increase in anthropogenically released atmospheric pollutants such as chlorofluorocarbons, chlorocarbons and organobromides that are responsible for the depletion of the UV-screening ozone layer in the stratosphere [6,22]. Due to depletion of the ozone layer there is an increase in the intensity of harmful ultraviolet radiation (UVR; 280–400 nm) reaching on the Earth's surface. The high-energetic ultraviolet-B (UV-B; 280–315 nm) has the greatest potential for cell damage caused by both direct effects on DNA and proteins and indirect effects *via* the production of reactive oxygen species [17,45]. Ultraviolet-A (UV-A; 315–400 nm) which is not absorbed by the native DNA molecule, can still induce DNA damage either by inducing secondary photoreactions of existing DNA photoproducts or *via* indirect photosensitizing reactions [15]. Morphology, cell differentiation, survival, growth, pigmentation, motility and orientation, N_2 metabolism, phycobiliprotein composition, protein profile, DNA and $^{14}CO_2$ uptake have been reported to be affected by UVR [12,21,39].

Several organisms, including cyanobacteria that are simultaneously exposed to photosynthetic active radiation (PAR; 400–700 nm) and UVR, have evolved a number of mitigation strategies to reduce the damaging effects of UVR. These strategies include repair of UV-induced damage of DNA by several mechanisms such as photoreactivation, nucleotide excision repair, base excision repair, mismatch repair, recombinational repair [18,19], spore germination and reproduction [14], accumulation of carotenoids and detoxifying enzymes or radical quenchers and antioxidants [23,25] as well as synthesis of UV-absorbing/screening compounds such as scytonemin and mycosporine-like amino acids (MAAs) [34,37]. In recent years, among UV-absorbing compounds MAAs have obtained much consideration for their photoprotective role as these compounds are not only present in cyanobacteria but have been also reported in macroalgae, phytoplankton and various animals such as arthropods, rotifers, molluscs, fishes, cnidarians, tunicates, poriferans, nemertineans, echinodermates, platythelminthes, polychaetes, bryozoans and protozoans [40]. This also indicates the evolutionary significance of these compounds during the course of evolution by conserving them, contrary to scytonemin, the occurrence of which is limited to cyanobacteria.

MAAs, originally identified in fungi as having a role in UV-induced sporulation [20], are small (<400 Da), colorless, water-soluble

* Corresponding author. Fax: +49 9131 8528215.
E-mail address: dphaeder@biologie.uni-erlangen.de (D.-P. Häder).

0888-7543/$ – see front matter © 2009 Elsevier Inc. All rights reserved.
doi:10.1016/j.ygeno.2009.10.002

compounds composed of a cyclohexenone or cyclohexenimine chromophore conjugated with the nitrogen substituent of an amino acids or its imino alcohol [34]. Generally, the ring system contains a glycine subunit at the third carbon atom. Some MAAs also contain sulfate esters or glycosidic linkages through the imine substituent [3,47]. Several characteristics of MAAs, such as strong UV absorption maxima between 310 and 362 nm and high molar extinction coefficients (ε = 28,100–50,000 M^{-1} cm^{-1}), suggest their UV-screening role. Differences in the absorption spectra of MAAs are due to variations in the attached side groups and nitrogen substituent. The biosynthesis of MAAs and mycosporines has been suggested to occur *via* the first part of the shikimate pathway in both cyanobacteria and fungi based on feeding experiments with radiolabelled compounds in these organisms [9,29]. The results from these experiments also suggested that 3-dehydroquinate (DHQ), which is formed in the middle of the shikimate pathway, acts as a precursor for the synthesis of fungal mycosporines and MAAs *via* deoxygadusol. This view was also supported by the inhibition of MAA synthesis in *Stylophora pistillata* by the application of glyphosate which is a specific shikimate pathway inhibitor [32]. The synthesis of MAAs has been reported to occur in bacteria, cyanobacteria, phytoplankton and macroalgae but not in animals, where these compounds are supposed to be accumulated either *via* the food chain or synthesized by their symbiotic algal partner due to the lack of the shikimate pathway [31]. In addition to their photoprotective role in various organisms, MAAs are also of immense importance for humans as these compounds have been found to effectively block thymine dimer formation by UVR in vitro, have antioxidant property and were recently found to provide UV protection as well as growth stimulation activity in human cells [8,24,27].

Although there is biochemical evidence regarding the biosynthetic route of these ecologically, evolutionary and industrially important compounds, no study has been conducted so far to identify the genes involved in the biosynthesis of these compounds. In the present investigation we conducted a bioinformatics study to identify the possible genes involved in MAA biosynthesis by analyzing the genome of MAA-synthesizing and non-synthesizing cyanobacteria.

Experimental methods

Experimental organisms and growth conditions

Four fully sequenced cyanobacteria, *Anabaena variabilis* PCC 7937, *Anabaena* sp. PCC 7120, *Synechocystis* sp. PCC 6803 and *Synechococcus* sp. PCC 6301, were selected for this study. These cyanobacteria were selected since their genome has been fully sequenced and molecular techniques, e.g., transformation, mutant development and expression of new genes have also been well developed. These organisms have served as model for studying the hydrogen and nitrogen metabolism as well as photosynthesis at molecular level. The other cyanobacteria that are known to synthesize MAAs [40] have neither been sequenced nor any molecular technique has been developed to elucidate the genes. All cyanobacteria were obtained from the Pasteur culture collection (Institute Pasteur, France) and are being maintained under axenic conditions in our laboratory. Cultures were routinely grown in an autoclaved BG11 medium [43] at a temperature of 20 ± 2 °C and continuous fluorescence lamp illumination of 12 ± 2 W m^{-2}. For MAA induction exponentially growing cultures were used.

MAA induction

The cultures of *A. variabilis* PCC 7937, *Anabaena* sp. PCC 7120, *Synechocystis* sp. PCC 6803 and *Synechococcus* sp. PCC 6301 were continuously exposed for 72 h in three replicates to artificial radiation of UVR and PAR in open glass Petri dishes (75 mm in diameter). As induction of MAAs has been reported to be UV-B dependent [41] 295 nm cut-off filters (Ultraphan, UV Opak Digefra, Munich, Germany) were placed over the Petri dishes to achieve the PAR+UV-A+UV-B radiation regime. Visible (PAR) light was supplied by Osram L 36 W/32 Lumilux de luxe warm white and radium NL 36 W/26 Universal white tubes (Munich, Germany). Ultraviolet-B TL 40 W/12 fluorescence tubes (Philips, The Netherlands) were used as a source of artificial UV-B irradiation. Since UV-B tubes used in this study were also emitting UV-A radiation no additional source for UV-A radiation was used. Fig. 1 shows the spectral irradiances of the light sources as transmitted by 295 nm cut-off filter and transmission spectra of corresponding cut-off filter measured with a double-monochromator spectroradiometer (OL 754, Optronic Laboratories, Orlando, FL) and a UV-VIS spectrophotometer (UV-2550, Shimadzu Scientific Instruments, Inc., Riverwood Drive, Columbia, MD), respectively. The irradiances effectively received by the samples were 0.89 W m^{-2} for UV-B, 1.06 W m^{-2} for UV-A and 10 W m^{-2} for PAR based on the transmission characteristics of the 295 nm cut-off filter.

All experiment cultures were shaken several times during the radiation exposure to avoid self-shading. From each sample 2 ml aliquots were withdrawn at 0, 24, 48 or 72 h intervals in triplicates and tested for the induction of MAAs.

MAA extraction

MAA extraction was performed as described earlier [33]. Briefly, cells were harvested by centrifugation (GP centrifuge, Beckman, Palo Alto, CA) and MAAs extracted in 2 ml of 100% HPLC grade methanol overnight at 4 °C. Thereafter, aliquots were centrifuged at 5000 × *g* for 10 min and supernatants were transferred to new Eppendorf tubes. Afterwards, supernatants were evaporated to dryness at 45 °C in a vacuum evaporator and the extracts were redissolved in 500 µl double distilled water. To this 200 µl of chloroform was added and mixed by vigorous shaking. Thereafter, 300 µl from the upper water phase was transferred into new Eppendorf tubes following centrifugation to remove contaminant photosynthetic pigments. Finally, the samples were filtered through 0.2 µm pore-sized microcentrifuge filters (Mikro-Spin Zentrifugenfilter, Roth, Karlsruhe, Germany) and further subjected to high performance liquid chromatographic (HPLC) analysis.

HPLC analysis

The analysis of partially purified MAAs was done using an HPLC system (Waters 996 photodiode array detector, Waters, Milford, MA; pump L-7100, Hitachi, Darmstadt, Germany) equipped with a Licrospher RP 18 column and guard (5 µm packing; 250 × 4 mm I.D.). The samples (50 µl) were injected into the HPLC column through a Waters 717 plus autosampler (Waters). The wavelength used for MAA detection was 334 nm to get the chromatogram of MAA in *A. variabilis*

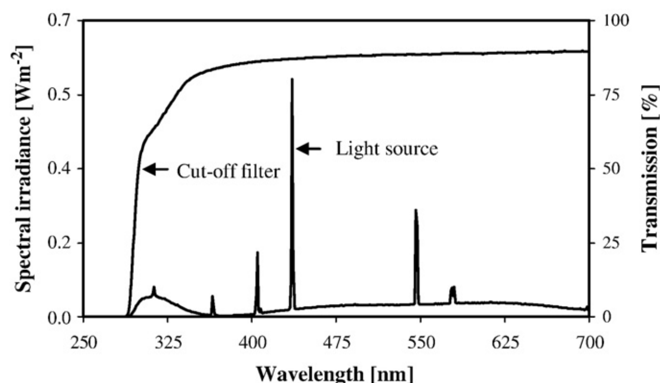

Fig. 1. Spectral irradiances of the light source used in the MAA induction experiments based on the transmission spectra of the 295 nm cut-off filter.

since only this cyanobacterium synthesizes MAA (shinorine) having an absorption maximum at 334 nm. The mobile phase was 0.02% acetic acid (v/v) in double distilled water, run isocratically at a flow rate of 1.0 ml/min.

Genomic analysis, sequence alignment and phylogenetic analysis

The genomes of all cyanobacteria used in this study are fully sequenced and available at Comprehensive Microbial Resource (CMR; http://cmr.jcvi.org/tigr-scripts/CMR/CmrHomePage.cgi). The comparison of genomic regions was performed using the service provided by CMR under gene-specific tools (http://cmr.jcvi.org/cgi-bin/CMR/CmrManual.cgi). The nucleotide sequences of promising genes were translated into amino acid sequences using the facility of the Open reading frame (ORF) finder at NCBI (http://www.ncbi.nlm.nih.gov/gorf/gorf.html) and similarity searches for the corresponding proteins were performed using the BLAST service at NCBI (http://blast.ncbi.nlm.nih.gov/Blast.cgi). Nucleotide sequences from similarity search were collected and a phylogenetic tree was constructed using the CLC Main workbench software (CLC bio A/S, Denmark). All alignments were constructed using a progressive alignment algorithm [11]. Phylogenetic trees were inferred for aligned sequences by using a distance based method [10]. The neighbor-joining algorithm [30] was used to build a tree where the evolutionary rates are free to differ in different lineages and reliability of the inferred tree was evaluated by bootstrap analysis. It is important to mention here that the used software always draws trees with roots for practical reasons, but with the neighbor-joining method, no particular biological hypothesis can be postulated by the placement of the root.

Protein structure and binding site prediction

For the prediction of protein structure and binding sites, amino acid sequences were submitted at I-TASSER internet service for protein structure and function predictions (http://zhang.bioinformatics.ku.edu/I-TASSER/). Models were built based on multiple-threading alignments by LOMETS and iterative TASSER simulations [48,51,52].

Results and discussion

Chromatograms obtained from HPLC analysis of all partially purified aqueous solutions revealed that out of the four cyanobacteria studied only *A. variabilis* PCC 7937 was able to synthesize MAA. *Anabaena* sp. PCC 7120, *Synechocystis* sp. PCC 6803 and *Synechococcus* sp. PCC 6301 were not able to synthesize this evolutionary important photoprotective compound even at trace levels during 72 h of radiation exposure (Fig. 2). Numerous studies have shown that biosynthesis of MAAs is regulated mainly by UV-B radiation in different organisms including cyanobacteria as evidenced from the action spectrum while UV-A and PAR had very little effect on their biosynthesis [36,41]. A UV-B-specific photoreceptor has also been proposed for the induction of MAAs in cyanobacteria [28]. It is important here to mention that the ratio of PAR:UV-A:UV-B in our study was quite different from that of solar radiation but has been shown to be sufficient enough to induce different MAAs in cyanobacteria [35]. Therefore, it is unlikely that we did not get MAAs in non-synthesizing cyanobacteria due to low UV-A. The biosynthesis of the MAA shinorine by *A. variabilis* PCC 7937 has been recently reported, and furthermore, it was found that biosynthesis of MAA in this cyanobacterium was regulated by a number of abiotic stressors such as UV radiation, salt and ammonium in a single as well as in a synergistic way [33]. The inability of *Anabaena* sp. PCC 7120 to synthesize MAAs in our experiment is also supported by the study of Gao et al. [10], in which they showed that this strain lacks the ability to synthesize MAAs. Similarly, there are also reports for the *Anabaena*

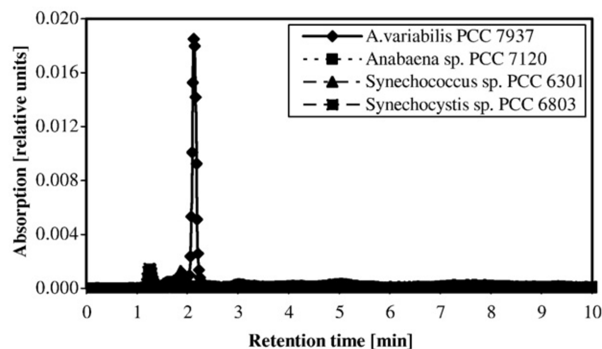

Fig. 2. HPLC chromatograms of *A. variabilis* PCC 7937, *Anabaena* sp. PCC 7120, *Synechocystis* sp. PCC 6803 and *Synechococcus* sp. PCC 6301 after 72 h of exposure showing a peak for MAAs only in *A. variabilis* PCC 7937.

strains that one *Anabaena* sp., isolated from a rice field, synthesizes the MAA shinorine [38] while another *Anabaena* sp. isolated from the same habitat lacks the ability to synthesize this compound [16]. It is important to mention that we were not able to detect any trace of MAAs from *Synechocystis* sp. PCC 6803, while Zhang et al. [50] have reported the presence of some unusual and new MAAs from this cyanobacterium, different from the normally reported MAAs (asterina-330, euhalothece-362, mycosporine-glycine, palythene, palythinol, Porphyra-334 and shinorine) in cyanobacteria [40]. According to them [50], the UV-absorbing compounds they found in *Synechocystis* sp. PCC 6803 were MAAs. However, the conclusions they draw from their measurements on the identity of the compounds are questionable. Their HPLC method for the analysis of MAAs has neither been used previously nor has it been tested with authentic MAA standards. The UV absorption spectra they show are cropped outside the region of the main absorption band. Moreover, in Fig. 1 [48] it can be seen that the absorbance of compound A increases below 290 nm, which is not at all an indicative of a pure MAA (for examples of MAAs spectra, see [5] Fig. 4). Their compound A has been tentatively identified as mycosporine-taurine based on a detected ion at m/z 318 that was interpreted as a Na^+ adduct, even though the protonated molecule was not found in the spectrum. As mycosporine-taurine is a relatively rare MAA and has yet not been found in any cyanobacterium, this line of evidence seems to be rather thin. Similarly, mass spectral measurements of other compounds do not correspond to known MAAs. The designation of compound C as dehydrolxyusurijene is also not correct, because removing of a hydroxyl group from usurijene results in a molecular weight of 268, and not, as the authors incorrectly assume, of 267 (as the standard valency of carbon has to be saturated by an additional hydrogen, resulting in a net loss of oxygen only). Thus the MS peak at m/z 268 cannot be assigned to the protonated molecule of dehydroxylusurijene.

Once it became evident from the MAA induction experiment and HPLC analysis that only *A. variabilis* PCC 7937 has the capability to synthesize MAAs, we decided on genomic analysis of these four fully sequenced cyanobacteria to determine some possible set of genes that might be involved in MAA biosynthesis and is only present in *A. variabilis* PCC 7937. Genome mining of these four cyanobacteria revealed that *A. variabilis* PCC 7937 and *Anabaena* sp. PCC 7120 possess two sets of the 3-dehydroquinate synthase (DHQS) gene while *Synechocystis* sp. PCC 6803 and *Synechococcus* sp. PCC 6301 had only one set of this gene in their genome. The accession numbers of these genes are YP_324358 (locus name: Ava_3858) and YP_324879 (locus name: Ava_4386) in *A. variabilis* PCC 7937, NP_485964 (locus name: Alr1924) and NC_003272 (locus name: All0417) in *Anabaena* sp. PCC 7120, NP_441388 (locus name: slr2130) in *Synechocystis* sp. PCC 6803 and YP_171706 (locus name: syc0996_d) in *Synechococcus* sp. PCC 6301. Hereafter, in the text the accession number will be used to refer to a particular gene. Genomic region analysis for these genes

in the four cyanobacteria revealed that in *A. variabilis* PCC 7937 gene YP_324358 has a O-methyltransferase (accession: YP_324357, locus: Ava_3857) gene downstream to it while in other cyanobacteria this combination was absent (Fig. 3). Further, we compared the genomic region of *A. variabilis* PCC 7937 with *Anabaena* sp. PCC 7120, *Synechocystis* sp. PCC 6803 and *Synechococcus* sp. PCC 6301 for YP_324879 and YP_324358 genes. The results from the genomic region comparison revealed that gene YP_324879 has more similarity than YP_324358 to NP_485964, NP_441388 and YP_171706 of *Anabaena* sp. PCC 7120, *Synechocystis* sp. PCC 6803 and *Synechococcus* sp. PCC 6301, respectively, based on the *P* values (Figs. 3 and 4). The second copy of the DHQS (NC_003272) gene of *Anabaena* sp. PCC 7120 was not close to either YP_324358 or YP_324879 in *A. variabilis* PCC 7937 (Figs. 3 and 4) and is found to be closest to the DHQS gene of *Nostoc punctiforme* ATCC 29133 from a region which has been recently found to be dedicated to the biosynthesis of another photoprotective compound, scytonemin [42]. However, in addition to DHQS gene which has been found to be a part of the 18-gene cluster associated with scytonemin biosynthesis, *N. punctiforme* was also found to have another DHQS (YP_001868848) which is 73% identical to the identified MAA-related gene and this gene also has O-methyltransferase (YP_001868847) downstream which is also needed for the formation of the core of MAAs. The involvement of these identified genes in MAA biosynthesis is also strengthened by the fact that *N. punctiforme* ATCC 29133 also has the ability to synthesize the MAA shinorine (data not shown). The biosynthesis of scytonemin is regulated by UV-A radiation while MAAs induction is regulated by UV-B radiation [7,13,36,41] so it is most unlikely that MAAs and scytonemin biosynthesis are under the same UV-based control mechanism. This is also supported by the different loci of identified genes for MAA biosynthesis and genes for scytonemin biosynthesis in *N. punctiforme* genome. The genomic region of *A. variabilis* PCC 7937 was also compared for O-methyltransferase (YP_324357) with other cyanobacteria (Fig. 5). The results from all genomic region comparisons and their *P* values indicated that in *A. variabilis* PCC 7937 DHQS (YP_324358) and O-methyltransferase (YP_324357) are unique and present in one reading frame while this combination was absent in the other three cyanobacteria (Figs. 3 and 5). Therefore our further analysis concentrated on these genes (YP_324358, YP_324879 and YP_324357) in *A. variabilis* PCC 7937. The O-methyltransferase gene was also included in this study as the common core of MAAs; 6-deoxygadusol has a methyl group at its second carbon atom and thus methylation is also an important step in formation of deoxygadusols.

The nucleotide sequence of these three genes was downloaded and used for further analysis. The nucleotide sequence of YP_324358 was translated into the amino acid sequence and the predicted amino acid sequence (410 amino acids) was used for a BLAST search. The result revealed the presence of putative conserved domains for the DHQS superfamily and identified number of sequences for predicted DHQS from cyanobacteria, DHQS and 2-epi-5-epi-valiolone synthase from fungi, chloroplast fused protein of DHQS and O-methyltransferase from dinoflagellates, predicted DHQS protein in the metazoan *Nematostella vectensis* and DHQS in bacteria (Supplementary material 1). Contrary to this, when the amino acid sequence (363 amino acids) of YP_324879 was used for a BLAST search, the closest fit was with the DHQS from cyanobacteria and bacteria and sequences were not obtained for fungi, dinoflagellates and metazoan which were detected by BLAST search with YP_324358 (Supplementary material 2). The results from BLAST analysis for both predicted DHQS (YP_324358 and YP_324879) of *A. variabilis* PCC 7937 also clearly indicated that both gene products are significantly different from each other as none of the BLAST analysis recognized each other and thus we hypothesized that these two proteins might be catalyzing different reactions in *A. variabilis* PCC 7937. Our hypothesis is also supported by the study of Wu et al. [49] in which they reported the comparison of highly conserved 13 binding pocket residues (active sites) of an enzymatically characterized DHQS of *Aspergillus nidulans* with the active site of DHQS, amino DHQ synthases, 2-deoxy-scyllo-inosose synthases and 2-epi-5-epi-valiolone synthases of different organisms. From this study it was clear that the product of the YP_324879 gene from *A. variabilis* PCC 7937 has similar 13 binding pocket residues (active site) as in the DHQS of *A. nidulans* and thus most likely this product might be exclusively involved in the biosynthesis of shikimate pathway-dependent aromatic amino acids by catalyzing the formation of dehydroquinate. In contrast, the YP_324358 product was found to have different amino acid residues from the conserved K250, N268, H275 and K356 of *A. nidulans* DHQS within the active site. In YP_324358 these residues are similar to the residues in 2-epi-5-epi-valiolone synthases having proline except at position 268 where YP_324358 has an alanine residue instead of aspartic acid in the 2-epi-5-epi-valiolone synthases. 2-Epi-5-epi-valiolone synthases are required for the biosynthesis of clinically important C7-cyclitol-containing compounds such as acarbose, validamycin and cetoniacytones from sedoheptulose 7-phosphate and share homology with the reaction catalyzed by the DHQS [49].

In the further study we constructed a phylogenetic tree using YP_324358 of *A. variabilis* PCC 7937 with the closest BLAST hit for cyanobacteria, dinoflagellates and metazoan (Fig. 6). The YP_324358 sequence appeared to have close similarity to those of cyanobacteria, dinoflagellates and metazoan (see E values in Fig. 6). Furthermore, all cyanobacteria, which have sequences closest to the YP_324358 of *A. variabilis* PCC 7937, interestingly also, had a second copy of a DHQS

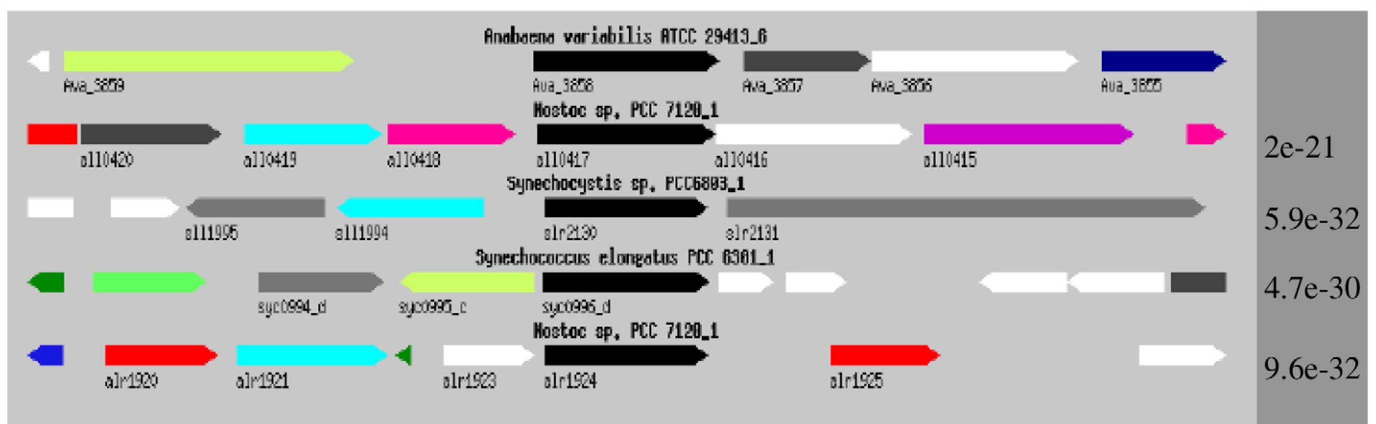

Fig. 3. Genomic region comparison of *A. variabilis* PCC 7937 (ATCC 29413), *Anabaena* (*Nostoc*) sp. PCC 7120, *Synechocystis* sp. PCC 6803 and *Synechococcus* sp. PCC 6301 for the YP_324358 (Ava_3858) gene (black arrow) of *A. variabilis* PCC 7937. The numerical value on the right side represents the *P* value for corresponding genes with respect to YP_324358.

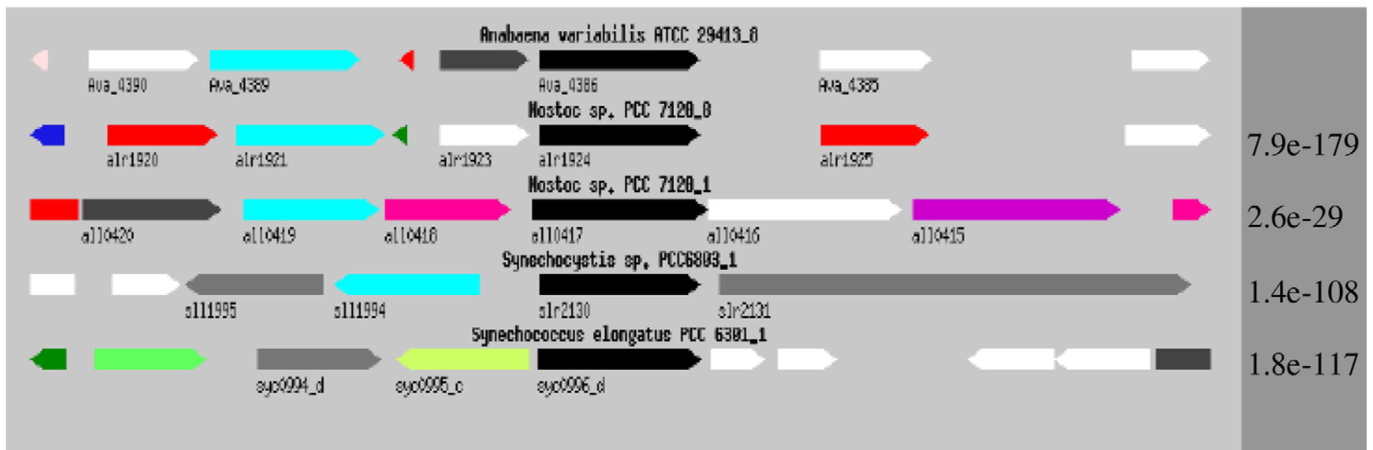

Fig. 4. Genomic region comparison of *A. variabilis* PCC 7937 (ATCC 29413), *Anabaena* (*Nostoc*) sp. PCC 7120, *Synechocystis* sp. PCC 6803 and *Synechococcus* sp. PCC 6301 for the YP_324879 (Ava_4386) gene (black arrow) of *A. variabilis* PCC 7937. The numerical value on the right side represents the *P* value for corresponding genes with respect to YP_324879.

gene like *A. variabilis* PCC 7937 and these were found to be close to the YP_324879 and NP_485964 of *A. variabilis* PCC 7937 and *Anabaena* sp. PCC 7120, respectively (see *E* values in Fig. 6). In the dinoflagellates *Heterocapsa triquetra* and *Oxyrrhis marina*, this product has been reported to be present in the chloroplast and to be fused to O-methyltransferase [46]. However, in *Karlodinium micrum* both DHQS and O-methyltransferase are not fused proteins but the O-methyltransferase is present downstream of the DHQS in the same reading frame (Figs. 6 and 7). In other cyanobacteria O-methyltransferase was also found to be present downstream to the predicted DHQS gene in the same reading frame similar to YP_324358 in *A. variabilis* PCC 7937 (Figs. 6 and 7). The metazoa *Nematostella vectensis* was also found to have the similar sequence of YP_324358, and the O-methyltransferase gene was also present downstream to it in this organism (Figs. 6 and 7). In the dinoflagellates *H. triquetra*, *O. marina* and *K. micrum* both genes have been reported to be transferred from cyanobacteria *via* a prokaryote-to-eukaryote lateral/horizontal gene transfer event during evolution (Figs. 6 and 7) [46]. Similarly, recently molecular evidence has been disclosed for the horizontal transfer of these two genes in *Nematostella vectensis* from dinoflagellate donors [44]. The presence of DHQS and O-methyltransferase genes together in *A. variabilis* PCC 7937 (YP_324358 and YP_324357) and in other cyanobacteria as well as their transfer from cyanobacteria to dinoflagellates and finally to an animal (*N. vectensis*) (Figs. 6 and 7) also support their possible involvement in the biosynthesis of MAAs as these compounds are not only present in cyanobacteria but have also been reported in macroalgae, phytoplankton and various animals [40]. However, in animals these compounds are supposed to be accumulated *via* the food chain as they can not synthesize these compounds *de novo* due to the lack of the shikimate pathway. Contrary to this, there are also some exceptions to this view as there are several studies which show that endosymbionts (algal partner) or algal diet are not the source of MAAs in the host animal [1,2] and propose some unique origin of MAAs instead of endosymbiotic or dietary origin in some animals [31].

We also modelled protein structure and binding sites for enzymes encoded by YP_324879 and YP_324358 of *A. variabilis* PCC 7937. The predicted structure and binding site for these two sequences are shown in Fig. 8. The results from structure prediction revealed that the product of YP_324879 is similar to the DHQS [4]; however, YP_324358 product is structurally similar to the 2-deoxy-scyllo-inosose synthase (DOIS) [26]. DOIS is a key enzyme in the biosynthesis of aminoglycoside antibiotics which catalyzes the cyclization of D-glucose-6-phosphate into 6-membered carbocycle DOI [26].The reaction mechanism of DIOS is similar to that of DHQS which catalyzes the cyclization of 3-deoxy-D-arabino-heputulosonate-7-phosphate to dehydroquinate in the shikimate pathway [4]. The active site of DIOS is very similar to that of DHQS; however, the presence of Glu 243 and Glu 235 at the active site of DIOS makes them significantly different from DHQS [26]. Thus, DHQS and DIOS differ from each other

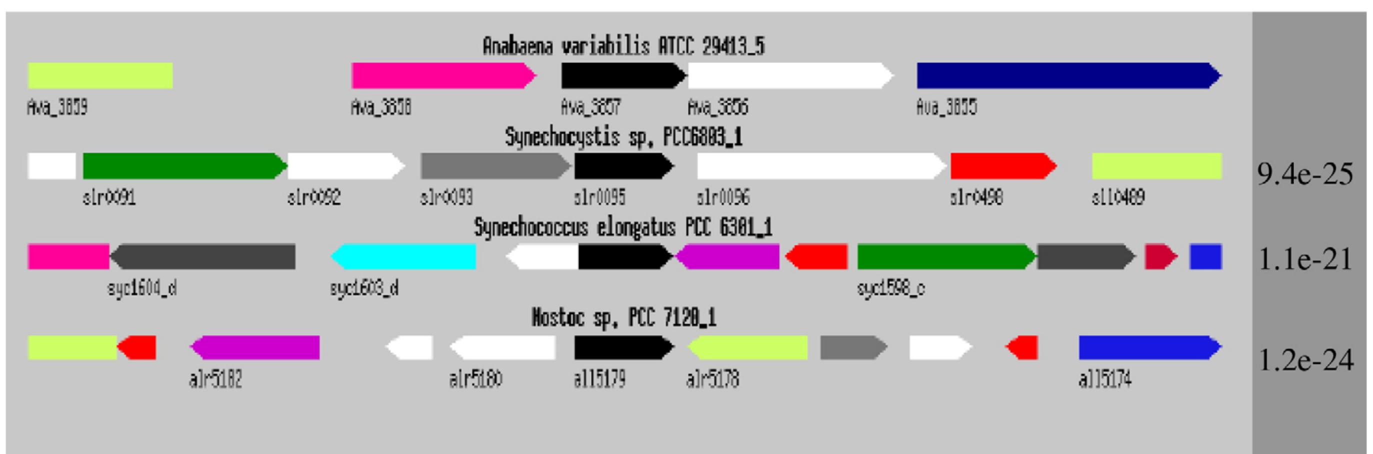

Fig. 5. Genomic region comparison of *A. variabilis* PCC 7937 (ATCC 29413), *Anabaena* (*Nostoc*) sp. PCC 7120, *Synechocystis* sp. PCC 6803 and *Synechococcus* sp. PCC 6301 for the YP_324357 (Ava_3857) gene (black arrow) of *A. variabilis* PCC 7937. The numerical value on the right side represents the *P* value for corresponding genes with respect to YP_324357.

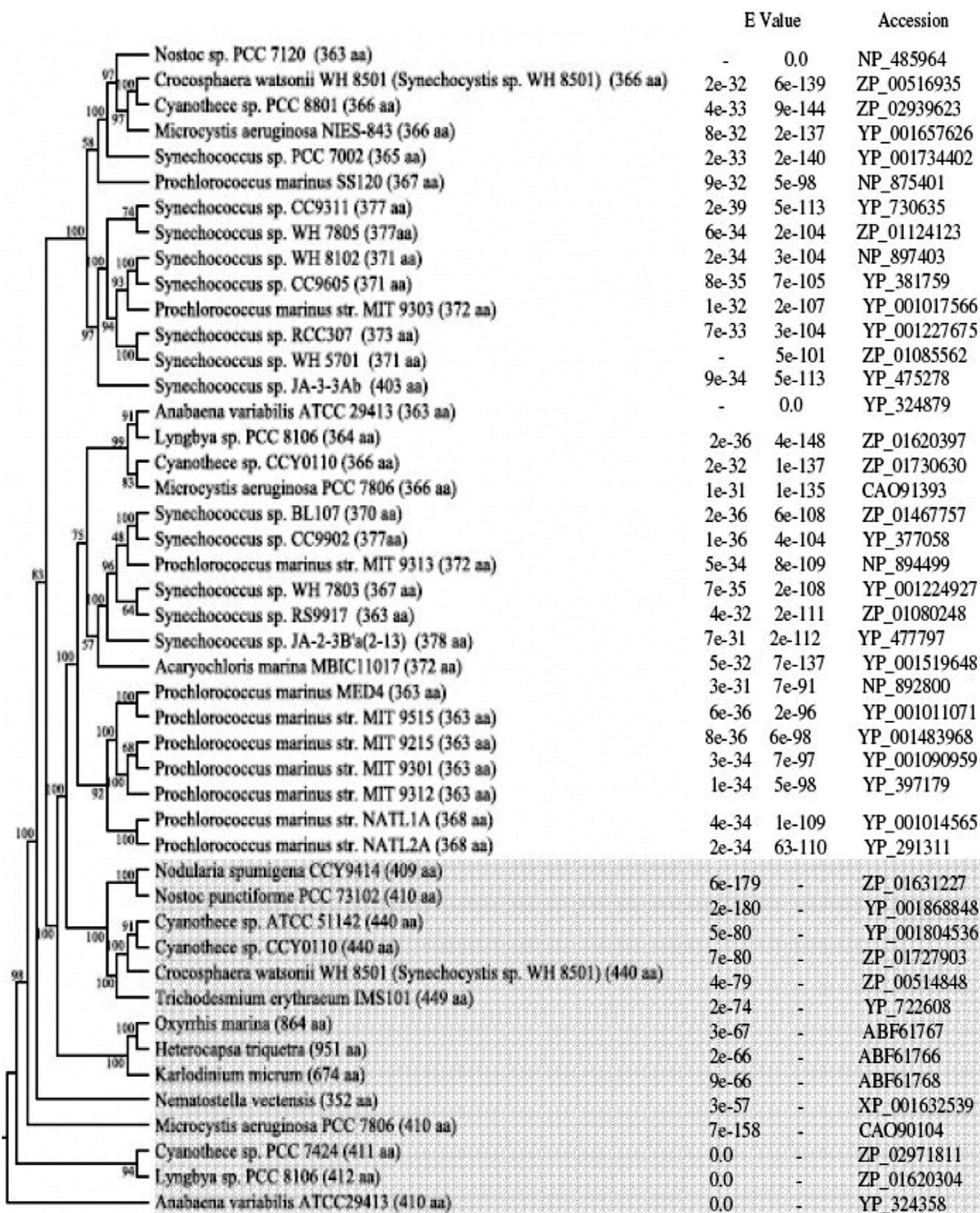

Fig. 6. Neighbor-joining phylogenetic tree including the homologues from cyanobacteria, dinoflagellate and metazoa found when amino acid sequences of YP_324358 or YP_324879 of *A. variabilis* PCC 7937 (ATCC 29413) were used for BLAST search. Numbers at nodes correspond to bootstrap values. The shaded part of the tree represents identified products that might be involved in MAA biosynthesis in cyanobacteria, dinoflagellates and metazoa. The first column next to the species name represents the *E* values when the YP_324358 product was used for BLAST and the second column represents the *E* value when the YP_324879 product was used for BLAST. '-' indicates that a product was not identified by BLAST. Accession numbers of all sequences used in the analysis are given next to the *E* value column.

only by two amino acid residues in their active site but the chemical reaction catalyzed by them differs greatly. Similarly, the predicted binding sites of both YP_324879 and YP_324358 are similar to each other; however, Thr 46, Val 108, Leu 134, Thr 174, Leu, Glu 179 and Lys 230 residues of YP_324879 are substituted by Val 59, Leu 120, Ile 146, Leu 183, Thr 186, Asn 194 and Met 250, respectively, in YP_324358 (Fig. 8E). In addition, Thr 144 residue of YP_324358 is missing in YP_324879 and Ala 182 residue of YP_324879 is missing in

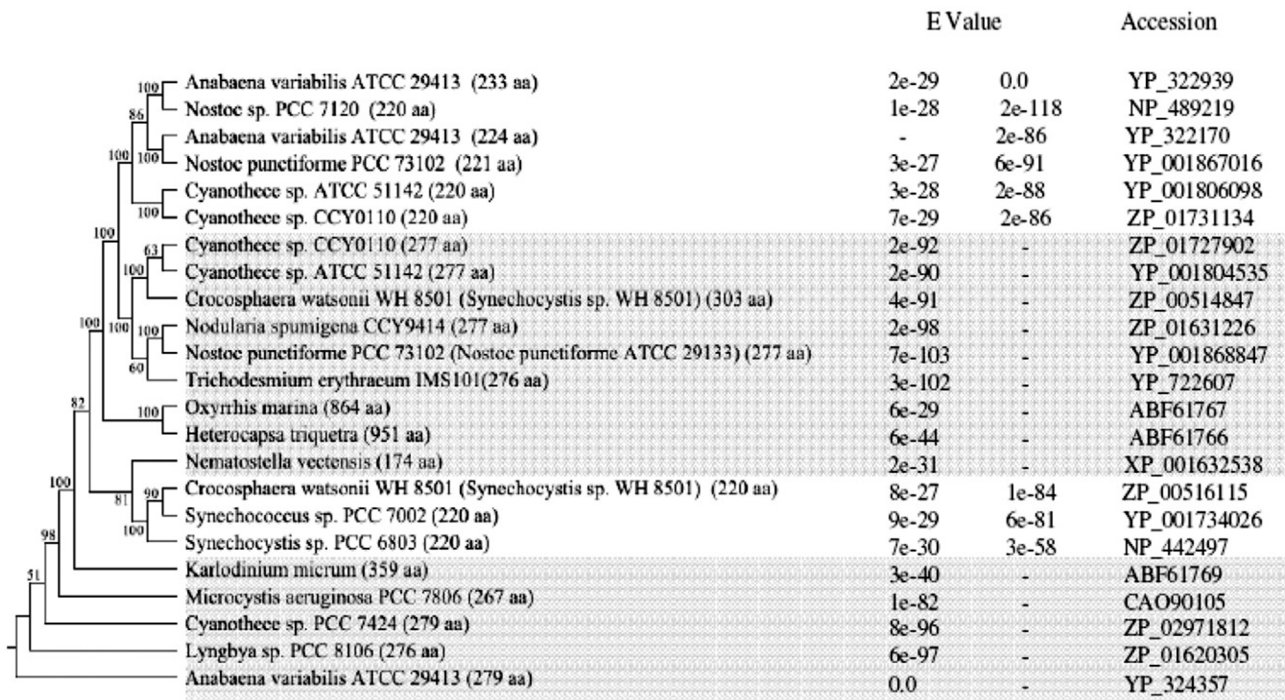

Fig. 7. Neighbor-joining phylogenetic tree including the homologues from cyanobacteria, dinoflagellate and metazoa found when amino acid sequences of YP_324357 or YP_322939 of *A. variabilis* PCC 7937 (ATCC 29413) were used for BLAST. Numbers at nodes correspond to bootstrap values. The shaded part of the tree represents identified products that might be involved in MAA biosynthesis in cyanobacteria, dinoflagellates and metazoa. The first column next to the species name represents *E* values when theYP_324357 product was used for BLAST and the second column represents *E* values when the YP_322939 product was used for BLAST. '-' indicates that a product was not identified during BLAST. Accession numbers of all sequences used in the analysis are given next to the *E* value column.

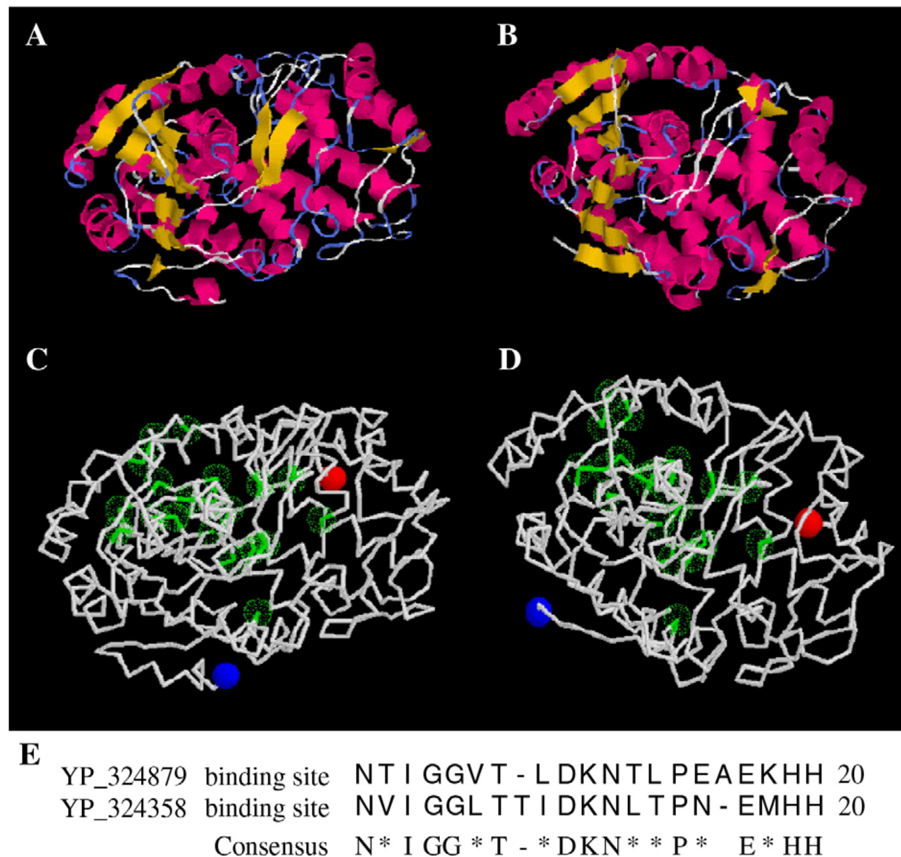

Fig. 8. Predicted I-TASSER structure and binding site of YP_324358 (A and C) and YP_324879 (B and D). Sequence alignment of predicted binding site residues from YP_324879 and YP_324358 (E). Green sphere in (C) and (D) represents binding sites residues while N and C termini in the model are marked by blue and red sphere, respectively.

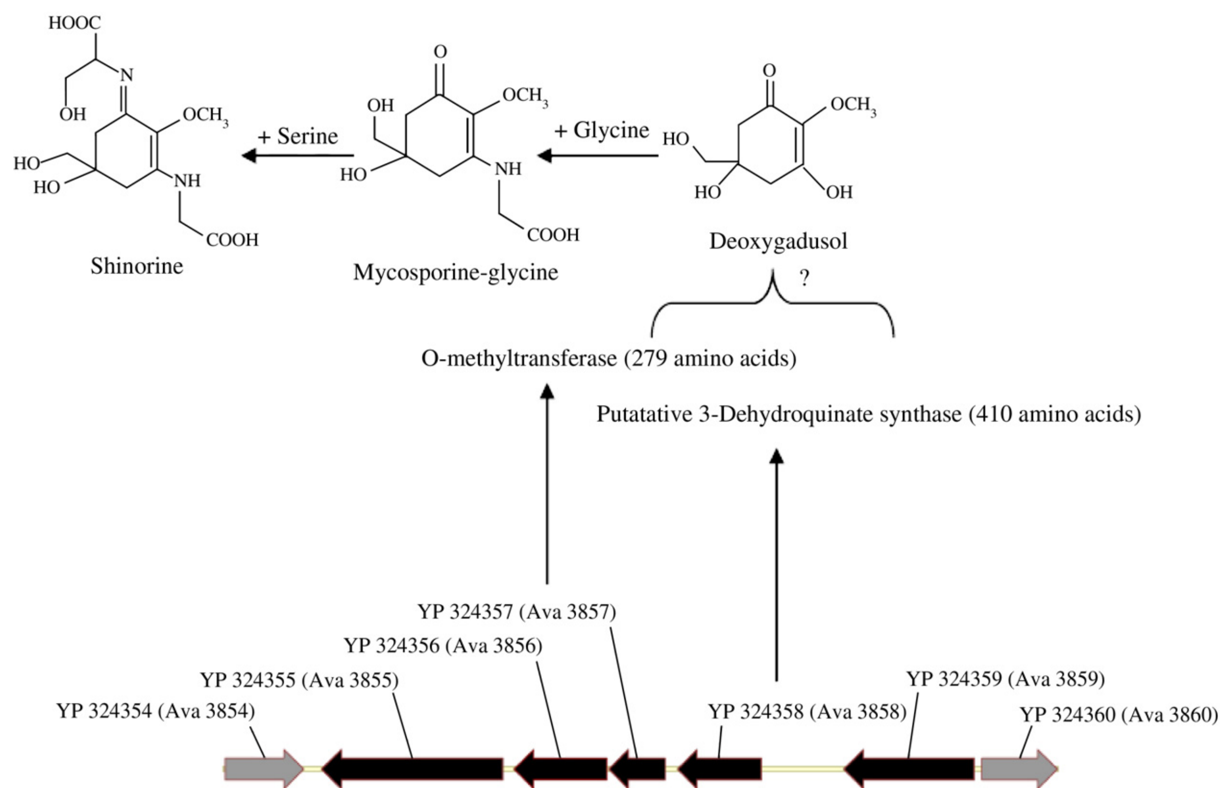

Fig. 9. Proposed role for the identified genes YP_324358 (Ava_3858) and YP_324357 (Ava_3857) from *A. variabilis* PCC 7937 in the biosynthesis of MAA shinorine in this cyanobacterium.

YP_324358. Thus, predicted models for both YP_324879 and YP_324358 also support that these products might be catalyzing different chemical reactions.

Conclusions

This study was conducted with an aim to identify the set of genes that are involved in the biosynthesis of MAAs using genome analysis approach in four fully sequenced cyanobacteria. This study will definitely provide foundation for further genetic studies (e.g., mutant development, promoter analysis) of these evolutionary important compounds which was so far completely clueless for workers interested in this field. The findings of this study will also help to understand the location of these compounds in eukaryotes as identified genes are located in chloroplasts in dinoflagellates. The cyanobacterium *A. variabilis* PCC 7937 was found to be the only one in a group of four to have the ability to synthesize MAAs while *Anabaena* sp. PCC 7120, *Synechocystis* sp. PCC 6803 and *Synechococcus* sp. PCC 6301 lack this ability. Genomic mining of these four cyanobacteria reveals the presence of a unique combination of two genes (DHQS; YP_324358 and O-methyltransferase; YP_324357) in *A. variabilis* PCC 7937 which are missing in other non-MAA-synthesizing cyanobacteria. Further phylogenetic analysis from the present and earlier studies revealed that these two genes were transferred from cyanobacteria to dinoflagellates and finally to metazoa by horizontal gene transfer. This additionally supports the possible involvement of these two gene products in MAA biosynthesis as these compounds are also present in other organisms and thought to be passed by cyanobacteria in the plastidic line [34]. The YP_324358 product was found to have similarity with the 2-epi-5-epi-valiolone synthase of actinobacteria (Supplementary material 1) and to have a different amino acid sequence from the sequences of the active site in the enzymatically characterized DHQS of *A. nidulans*. Contrary to this, the YP_324879 gene product has the conserved 13 binding pocket residues similar to the enzymatically characterized DHQS of *A. nidulans* [49]. The predicted model for YP_324879 also supported that this product is similar to DHQS of *A. nidulans*. In contrast, the YP_324358 product has different amino acid residues in the predicted binding site and structure was found similar to the DIOS. Based on these finding we propose that the YP_324879 gene product is exclusively involved in the shikimate pathway catalyzing the formation of dehydroquinate while the YP_324358 gene product together with the YP_324357 gene product (O-methyltransferase) catalyze the formation of deoxygadusol which is the core of all MAAs (Fig. 9). Once deoxygadusol is synthesized, free amino acids can be ligated to this core to synthesize different MAAs as reported earlier [29]. The genome mining for other cyanobacteria that fall under the category of having both the DHQS and O-methyltransferase gene was also performed and it was found that both genes in all cyanobacteria were present together and O-methyltransferase was always present downstream to DHQS in all identified cyanobacteria. These cyanobacteria are included in Figs. 6 and 7 and based on all these analyses a possible MAA biosynthesis clusters has been presented in Table 1. Table 1 presents information about accession number, locus name, product length and predicted function of other open reading frames

Table 1
Open reading frames from the identified genomic region of *A. variabilis* PCC 7937 which are in same transcriptional orientation with YP_324358 and YP_324357.

Accession number	Locus name	Product length (amino acids)	Predicted function
YP_324359	Ava_3859	641	Cd/Co/Hg/Pb/Zn-translocating P-type ATPase
YP_324358	Ava_3858	410	3-Dehydroquinate synthase
YP_324357	Ava_3857	279	O-methyltransferase family protein
YP_324356	Ava_3856	458	Hypothetical protein
YP_324355	Ava_3855	888	Amino acid adenylation

from the identified genomic region of *A. variabilis* PCC 7937. Thus, our study provides first insight into the genome of *A. variabilis* PCC 7937 and opens the field for molecular, bioinformatics and phylogenetic analysis of these evolutionary important photoprotective compounds in cyanobacteria and other organisms which was till now limited to biochemical analysis.

Acknowledgments

Authors are thankful to Martin Schuster for technical support and Ulrike Trenz for providing the literature cited in this study. We also sincerely thank Prof. Dr. Karl Forchhammer (University of Tuebingen, Germany) for providing the cyanobacterium *N. punctiforme* ATCC 29133.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.ygeno.2009.10.002.

References

[1] A.T. Banaszak, T.C. LaJeunesse, R.K. Trench, The synthesis of mycosporine-like amino acids (MAAs) by cultured, symbiotic dinoflagellates, J. Exp. Mar. Biol. Ecol. 249 (2000) 219–233.
[2] A.T. Banaszak, R.K. Trench, Ultraviolet sunscreens in dinoflagellates, Protist 152 (2001) 93–101.
[3] G.A. Böhm, W. Pfleiderer, P. Böger, S. Scherer, Structure of a novel oligosaccharide-mycosporine-amino acid ultraviolet A/B sunscreen pigment from the terrestrial cyanobacterium *Nostoc commune*, J. Biol. Chem. 270 (1995) 8536–8539.
[4] E.P. Carpenter, A.R. Hawkins, J.W. Frost, K.A. Brown, Structure of dehydroquinate synthase reveals an active site capable of multi step catalysis, Nature 394 (1998) 299–302.
[5] J.I. Carreto, M.O. Carignan, G. Daleo, S.G. De Marco, Occurrence of mycosporine-like amino acids in the red-tide dinoflagellate *Alexandrium excavatum*: UV-protective compounds? J. Plankton Res. 12 (1990) 909–921.
[6] P.J. Crutzen, Ultraviolet on the increase, Nature 356 (1992) 104–105.
[7] J.G. Dillon, C.M. Tatsumi, P.G. Tandingan, R.W. Castenholz, Effect of environmental factors on the synthesis of scytonemin, a UV-screening pigment, in a cyanobacterium (*Chroococcidiopsis* sp.), Arch. Microbiol. 177 (2002) 322–331.
[8] W.C. Dunlap, Y. Yamamoto, Small-molecule antioxidants in marine organisms: antioxidant activity of mycosporine-glycine, Comp. Biochem. Physiol., B Biochem. Mol. Biol. 112 (1995) 105–114.
[9] J. Favre-Bonvin, J. Bernillon, N. Salin, N. Arpin, Biosynthesis of mycosporines: mycosporine glutaminol in *Trichothecium roseum*, Phytochemistry 26 (1987) 2509–2514.
[10] J. Felsenstein, Evolutionary trees from DNA sequences: a maximum likelihood approach, J. Mol. Evol. 17 (1981) 368–376.
[11] D.F. Feng, R.F. Doolittle, Progressive sequence alignment as a prerequisite to correct phylogenetic trees, J. Mol. Evol. 25 (1987) 351–360.
[12] K. Gao, H. Yu, M.T. Brown, Solar PAR and UV radiation affects the physiology and morphology of the cyanobacterium *Anabaena* sp. PCC 7120, J. Photochem. Photobiol., B Biol. 89 (2007) 117–124.
[13] F. Garcia-Pichel, R.W. Castenholz, Characterization and biological implications of scytonemin, a cyanobacterial sheath pigment, J. Phycol. 27 (1991) 395–409.
[14] T. Han, J.-A. Kong, Y.-S. Han, S.-H. Kang, D.-P. Häder, UV-A/blue light-induced reactivation of spore germination in UV-B irradiated *Ulva pertusa* (Chlorophyta), J. Phycol. 40 (2004) 315–322.
[15] A. Hargreaves, F.A. Taiwo, O. Duggan, S.H. Kirk, S.I. Ahmad, Near-ultraviolet photolysis of β-phenylpyruvic acid generates free radicals and results in DNA damage, J. Photochem. Photobiol., B Biol. 89 (2007) 110–116.
[16] Y.-Y. He, M. Klisch, D.-P. Häder, Adaptation of cyanobacteria to UV-B stress correlated with oxidative stress and oxidative damage, Photochem. Photobiol. 76 (2002) 188–196.
[17] D. Karentz, J.E. Cleaver, D.L. Mitchell, DNA damage in the Antarctic, Nature 350 (1991) 28.
[18] S.-T. Kim, A. Sancar, Photorepair of nonadjacent pyrimidine dimers by DNA photolyase, Photochem. Photobiol. 61 (1995) 171–174.
[19] S. Kumari, R.P. Rastogi, K.L. Singh, S.P. Singh, R.P. Sinha, DNA damage: detection strategies, EXCLI J. 7 (2008) 44–62.
[20] C.M. Leach, Ultraviolet-absorbing substances associated with light-induced sporulation in fungi, Can. J. Bot. 43 (1965) 185–200.
[21] M.P. Lesser, Effects of ultraviolet radiation on productivity and nitrogen fixation in the cyanobacterium, *Anabaena* sp. (Newton's strain), Hydrobiologia 598 (2008) 1–9.
[22] D. Lubin, E.H. Jensen, Effects of clouds and stratospheric ozone depletion on ultraviolet radiation trends, Nature 377 (1995) 710–713.
[23] E.M. Middleton, A.H. Teramura, The role of flavonol glycosides and carotenoids in protecting soybean from ultraviolet-B damage, Plant Physiol. 103 (1993) 741–752.
[24] T. Misonou, et al., UV-absorbing substance in the red alga *Porphyra yezoensis* (Bangiales, Rhodophyta) block thymine photodimer production, Mar. Biotechnol. 5 (2003) 194–200.
[25] R. Mittler, E. Tel-Or, Oxidative stress responses in the unicellular cyanobacterium *Synechococcus* PCC7942, Free Radic. Res. Commun. 12 (1991) 845–850.
[26] E. Nango, T. Kumasaka, T. Hirayama, N. Tanaka, T. Eguchi, Structure of 2-deoxy-scyllo-inosose synthase, a key enzyme in the biosynthesis of 2- deoxystreptamine-containing aminoglycoside antibiotics, in complex with a mechanism based inhibitor and NAD^+, Proteins 70 (2008) 517–527.
[27] C. Oyamada, M. Kaneniwa, K. Ebitani, M. Murata, K. Ishihara, Mycosporine-like amino acids extracted from scallop (*Patinopecten yessoensis*) ovaries: UV protection and growth stimulation activities on human cells, Mar. Biotechnol. 10 (2008) 141–150.
[28] A. Portwich, F. Garcia-Pichel, A novel prokaryotic UVB photoreceptor in the cyanobacterium *Chlorogloeopsis* PCC 6912, Photochem. Photobiol. 71 (2000) 493–498.
[29] A. Portwich, F. Garcia-Pichel, Biosynthetic pathway of mycosporines (mycosporine-like amino acids) in the cyanobacterium *Chlorogloeopsis* sp. strain PCC 6912, Phycologia 42 (2003) 384–392.
[30] N. Saitou, M. Nei, The neighbor-joining method: a new method for reconstructing phylogenetic trees, Mol. Biol. Evol. 4 (1987) 406–425.
[31] J.M. Shick, W.C. Dunlap, Mycosporine-like amino acids and related gadusols: biosynthesis, accumulation, and UV-protective functions in aquatic organisms, Annu. Rev. Physiol. 64 (2002) 223–262.
[32] J.M. Shick, S. Romaine-Lioud, C. Ferrier-Pagès, J.-P. Gattuso, Ultraviolet-B radiation stimulates shikimate pathway-dependent accumulation of mycosporine-like amino acids in the coral *Stylophora pistillata* despite decrease in its population of symbiotic dinoflagellates, Limnol. Oceanogr. 44 (1999) 1667–1682.
[33] S.P. Singh, M. Klisch, R.P. Sinha, D.-P. Häder, Effects of abiotic stressors on synthesis of the mycosporine-like amino acid shinorine in the cyanobacterium *Anabaena variabilis* PCC 7937, Photochem. Photobiol. 84 (2008) 1500–1505.
[34] S.P. Singh, S. Kumari, R.P. Rastogi, K.L. Singh, R.P. Sinha, Mycosporine-like amino acids (MAAs): chemical structure, biosynthesis and significance as UV-absorbing/screening compounds, Ind. J. Exp. Biol. 46 (2008) 7–17.
[35] S.P. Singh, R.P. Sinha, M. Klisch, D.-P. Häder, Mycosporine-like amino acids (MAAs) profile of a rice-field cyanobacterium *Anabaena doliolum* as influenced by PAR and UVR, Planta 229 (2008) 225–233.
[36] R.P. Sinha, N.K. Ambasht, J.P. Sinha, D.-P. Häder, Wavelength-dependent induction of a mycosporine-like amino acid in a rice-field cyanobacterium, *Nostoc commune*: role of inhibitors and salt stress, Photochem. Photobiol. Sci. 2 (2003) 171–176.
[37] R.P. Sinha, D.-P. Häder, UV-protectants in cyanobacteria, Plant Sci. 174 (2008) 278–289.
[38] R.P. Sinha, M. Klisch, D.-P. Häder, Induction of a mycosporine-like amino acid (MAA) in the rice-field cyanobacterium *Anabaena* sp. by UV irradiation, J. Photochem. Photobiol., B Biol. 52 (1999) 59–64.
[39] R.P. Sinha, S. Kumari, R.P. Rastogi, Impacts of ultraviolet-B radiation on cyanobacteria: photoprotection and repair, J. Sci. Res. 52 (2008) 125–142.
[40] R.P. Sinha, S.P. Singh, D.-P. Häder, Database on mycosporines and mycosporine-like amino acids (MAAs) in fungi, cyanobacteria, macroalgae, phytoplankton and animals, J. Photochem. Photobiol., B Biol. 89 (2007) 29–35.
[41] R.P. Sinha, J.P. Sinha, A. Gröniger, D.-P. Häder, Polychromatic action spectrum for the induction of a mycosporine-like amino acid in a rice-field cyanobacterium, *Anabaena* sp. J. Photochem. Photobiol., B Biol. 66 (2002) 47–53.
[42] T. Soule, V. Stout, W.D. Swingley, J.C. Meeks, F. Garcia-Pichel, Molecular genetics and genomic analysis of scytonemin biosynthesis in *Nostoc punctiforme* ATCC 29133, J. Bacteriol. 189 (2007) 4465–4472.
[43] R.Y. Stanier, R. Kunisawa, M. Mandel, G. Cohen-Bazire, Purification and properties of unicellular blue-green algae (order Choococcales), Bacteriol. Rev. 35 (1971) 171–205.
[44] A. Starcevic, et al., Enzymes of the shikimic acid pathway encoded in the genome of a basal metazoan, *Nematostella vectensis*, have microbial origins, Proc. Natl. Acad. Sci. U. S. A. 105 (2008) 2533–2537.
[45] W.F. Vincent, P.J. Neale, Mechanisms of UV damage to aquatic organisms, in: S.J. de Mora, S. Demers, M. Vernet (Eds.), The Effects of UV Radiation on Marine Ecosystems, Cambridge Univ. Press, Cambridge, 2000, pp. 149–176.
[46] R.F. Waller, C.H. Slamovits, P.J. Keeling, Lateral gene transfer of a multigene region from cyanobacteria to dinoflagellates resulting in a novel plastid-targeted fusion protein, Mol. Biol. Evol. 23 (2006) 1437–1443.
[47] J.J. Wu, WonB.E. , ChalkerJ.A. , Rideout, Two new UV-absorbing compounds from *Stylophora pistillata*: sulfate esters of mycosporine-like amino acids, Tetrahedron Lett. 38 (1997) 2525–2526.
[48] S. Wu, J. Skolnick, Y. Zhang, *Ab initio* modeling of small proteins by iterative TASSER simulations, BMC Biol. 5 (2007) 17.
[49] X. Wu, et al., A comparative analysis of the sugar phosphate cyclase superfamily involved in primary and secondary metabolism, ChemBioChem 8 (2007) 239–248.
[50] L. Zhang, L. Li, Q. Wu, Protective effects of mycosporine-like amino acids of *Synechocystis* sp. PCC 6803 and their partial characterization, J. Photochem. Photobiol., B Biol. 86 (2007) 240–245.
[51] Y. Zhang, Template-based modeling and free modeling by I-TASSER in CASP7, Proteins 8 (2007) 108–117.
[52] Y. Zhang, I-TASSER server for protein 3D structure prediction, BMC Bioinformatics 9 (2008) 40.