Rearranged Biosynthetic Gene Cluster and Synthesis of Hassallidin E in Planktothrix serta PCC 8927.

Claire Pancrace, Jouni Jokela, Nathalie Sassoon, Christelle Ganneau, Marie Desnos-Ollivier, Matti Wahlsten, Anu Humisto, Alexandra Calteau, Sylvie Bay, David P. Fewer, et al.

▶ **To cite this version:**

Claire Pancrace, Jouni Jokela, Nathalie Sassoon, Christelle Ganneau, Marie Desnos-Ollivier, et al.. Rearranged Biosynthetic Gene Cluster and Synthesis of Hassallidin E in Planktothrix serta PCC 8927.. ACS Chemical Biology, 2017, 12 (7), pp.1796-1804. 10.1021/acschembio.7b00093 . pasteur-01535773

HAL Id: pasteur-01535773

https://pasteur.hal.science/pasteur-01535773

Submitted on 22 Feb 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Distributed under a Creative Commons Attribution 4.0 International License

Rearranged biosynthetic gene cluster and synthesis of hassallidin E in *Planktothrix serta* PCC 8927

C. Pancrace[1,2], J. Jokela[3], N. Sassoon[1], C. Ganneau[4,5], M. Desnos-Ollivier[6], M. Wahlsten[3], A. Humisto[3], A. Calteau[7], S. Bay[4,5], D. P. Fewer[3], K. Sivonen[3], M. Gugger[1]

[1] Institut Pasteur, Collection of Cyanobacteria, Paris, France

[2] Sorbonne Universités, UPMC Univ Paris 06, UPEC, UDD, CNRS, INRA, IRD, IEES-Paris, Paris, France.

[3] Microbiology and Biotechnology Division, Department of Food and Environmental Science, University of Helsinki, Helsinki, Finland

[4] Institut Pasteur, Unit Chemistry of Biomolecules, Paris, France

[5] CNRS UMR 3523, Paris, France

[6] Institut Pasteur / CNRS URA3012, National Refence Center for Invasive Mycoses and Antifungals, Molecular Mycology Unit, Paris, France

[7] Commissariat à l'Energie Atomique et aux Energies Alternatives (CEA), Genoscope / CNRS, UMR 8030, Laboratoire d'Analyse Bioinformatique en Génomique et Métabolisme, Evry, France.

Corresponding author : Muriel Gugger (muriel.gugger@pasteur.fr)

ABSTRACT

Cyanobacteria produce a wide range of natural products with antifungal bioactivity. The cyclic glycosylated lipopeptides of the hassallidin family have potent antifungal activity and display a great degree of chemical diversity. Here we report the discovery of a hassallidin biosynthetic gene cluster from the filamentous cyanobacterium *Planktothrix serta* PCC 8927. The hassallidin gene cluster showed heavy rearrangement and marks of genomic plasticity. Nucleotide bias, differences in GC content and phylogenetic incongruence suggested the acquisition of the hassallidin biosynthetic gene cluster in *Planktothrix serta* PCC 8927 by horizontal gene transfer. Chemical analyses by liquid chromatography and mass spectrometry demonstrated that this strain produced hassallidin E, a new glycosylated hassallidin variant. Hassallidin E was the only structural variant produced by *Planktothrix serta* PCC 8927 in all tested conditions.

Further evaluated on human pathogenic fungi, hassallidin E showed an antifungal bioactivity. Hassallidin production levels correlated with nitrogen availability, in the only nitrogen-fixing *Planktothrix* described so far. Our results provide insights into the distribution and chemical diversity of cyanobacterial antifungal compounds as well as raising questions on their ecological relevance.

INTRODUCTION

Blooms formed by toxic cyanobacteria are a well-known phenomenon in fresh and brackish waters.[1-3] Many benthic cyanobacteria form blooms or mats on river banks or the river bed.[4-6] Benthic cyanobacterial blooms were reported mostly for dog neurotoxicosis occurrences from rivers[7] and for the stratification and nitrogen fixation observed in microbial mats dominated by cyanobacteria.[8]

Cyanobacteria produce a great diversity of natural products[9] with a wide range of bioactivities of interest for pharmaceutical and biotechnological applications.[10, 11] Many cyanobacterial natural products belonging to diverse families have antifungal properties[11],[12] including cyclic peptides (calophycin[13]), alkaloids (hapalindoles[14]), macrolides (scytophycin, tolytoxin, amantelide[15]), terpenes (scytoscalarol[16]) and glycosylated lipopeptides like hassallidins and anabaenolysin.[17-19] Surprisingly, the majority of cyanobacterial antifungal natural products are reported from Subsection IV.[20] Their restricted distribution among various groups of heterocyst-forming cyanobacteria raise questions about their cellular function, metabolic cost and ecological relevance.[21]

Natural products with antifungal activity could confer a selective advantage allowing cyanobacteria to dominate their natural habitat. Fungi are reported to play an important role in cyanobacterial bloom dynamics in the natural environment.[22, 23] A parasitic chytrid fungus belonging to the division *Chytridiomycota* was shown to specifically infect and regulate the development of the *Planktothrix* population in Norwegian lakes.[24] Strains of the genus *Planktothrix* produce different ribosomal and non-ribosomal oligopeptide cocktails as an antiparisitic defense system[25], these compounds taking part in regulation and coevolution of chytrid/*Planktothrix* equilibrium.[26]

Hassallidins are macrocyclic glycosylated lipopeptides with potent antifungal activity.[17, 18] These compounds are composed of an eight-residue cyclic peptide including

proteinogenic and non proteinogenic amino acids, linked with an amino acid, a fatty acid and carbohydrate residues.[17, 18] Other cyanobacterial antifungal compounds belong to the same chemical family as hassallidins, namely the balticidins from *Anabaena cylindrica*, linear and circular glycolipopeptides that share the same core-sequence.[27] The biosynthesis pathway for the production of hassallidins was reported from the bloom-forming heterocystous genus *Anabaena*.[21] The hassallidin gene cluster was subsequently found in several genera of heterocystous cyanobacteria, including *Nostoc, Cylindrospermopsis, Tolypothrix* and *Hassallia*[21] and only in these cyanobacterial taxa so far.

In this study, we report the serendipitous discovery of a hassallidin biosynthetic gene cluster from the nitrogen-fixing strain *Planktothrix serta* PCC 8927, a filamentous non heterocystous forming cyanobacterium. We report the discovery and characterization of a new chemical variant of hassallidin, hassallidin E from *Planktothrix serta* PCC 8927. Hassallidin E was the only variant produced by the strain and had antifungal bioactivity. The hassallidin biosynthetic gene cluster of *Planktothrix serta* PCC 8927 showed heavy rearrangement compared to the orthologous gene clusters from heterocystous cyanobacteria, nucleotide bias and phylogenetic incongruence suggesting a possible acquisition of the entire gene cluster by a horizontal gene transfer event.

RESULTS AND DISCUSSION

Four genes encoding the HasN, HasO, HasV and HasY NRPS proteins were identified on contig 13 of the *Planktothrix serta* PCC 8927 draft genome (accession number CZCU01000013). These genes were predicted to be part of a 48 kb-long biosynthetic gene cluster encoding 16 genes. The four NRPS genes encoded the enzymes required for the biosynthesis of the hassallidin cyclic core. Analysis of the NRPS adenylation domain substrate specificity predicted the incorporation of five threonines, two glutamines, one tyrosine, and one glycine (Table 1). The HasO, HasN, and HasY encoded also three epimerization and one N-methyltransferase domains (Table 1) to obtain the enantiomeric amino acids and the N-methylated amino acid in the final molecule. The genome assembly of the NRPS regions revealed several tailoring genes in their vicinity that coded for two hydroxylases, reductase and glycosyl transferase, HasB, HasM, HasL and HasT respectively, needed for the macrocyclization and the glycosylation of

hassallidin (Table 1). Thirteen of the 16 genes encoded in the hassallidin gene cluster of *Planktothrix serta* PCC 8927 shared between 64-89 % amino acid identity with homologs in other hassallidin gene clusters (Table 1).

Biosynthetic gene clusters for the production of known cyanobacterial natural products have been discovered serendipitously from genomes.[28, 35] A phylum-wide examination revealed that the majority of these natural product gene clusters appears to be specific to individual strains or to a selection of cyanobacterial taxa.[28, 29] The hassallidin and balticidin biosynthetic gene clusters were previously described only from heterocystous cyanobacteria[12, 21], while *Planktothrix* is a well-known for its synthesis of microcystin, anabaenopeptin and microviridin among others.[30] Indeed, the strain *Planktothrix serta* PCC 8927 appeared to produce a hassallidin variant (this study) as well as a yet unknown ribosomal compound.[29] Several benthic *Planktothrix* strains were recently shown to exhibit a wider potential for natural products than the planktic strains of this genus. The synthetic pathways for tolytoxin and luminaolide B were detected in *Planktothrix* sp. PCC 11201 and *Planktothrix paucivesiculata* PCC 9631 respectively.[29, 31] Interestingly, the two compounds tolytoxin and hassallidin generally associated with heterocystous cyanobacteria are now found in biofilm-forming benthic *Planktothrix* strains.

Planktothrix serta PCC 8927 is one of the most distantly related benthic *Planktothrix* to the planktic *Planktothrix* based on concatenated housekeeping protein phylogeny,[28] and was not closely related to any other hassallidin producing strains, known to cluster among the heterocystous cyanobacteria (Figure 1). On the other hand, a phylogeny based on a concatenated alignment of the genes encoding HasO, Has V, HasN and HasY revealed that the NRPS genes of *Planktothrix serta* PCC 8927 are more closely related to the ones of *Tolypothrix* sp. PCC 9009 and *Hassalia byssoidea* VB512170 (Figure 2). Thus, the phylogenies on housekeeping proteins and on the *has* genes are incongruent. The presence of the hassallidin gene cluster in this *Planktothrix* strain could be explained by a horizontal gene transfer event, as the following features indicated the likeliness of such an evolutionary event. The *has* gene cluster was located in a 95 kb-long region of the draft genome of *Planktothrix serta* PCC 8927, which comprised the 16 *has* genes flanked by 48 genes. The flanking genes were absent in the 14 other *Planktothrix* genomes available.[29, 32] Moreover, this 95-kb region was characterized by genomic

plasticity (Figure 3). The *has* gene cluster and 32 of its closest flanking genes are characterized by a positive IVOM (interpolated variable order motifs) mark, indicating a compositional bias in nucleotide patterns from 2-mers to 8-mers in this region compared to the rest of the genome. Additionally, these flanking genes presented a GC% deviation compared to the whole genome. Biases in the nucleotide composition of the locus, both evidenced with the GC% and the IVOM patterns, were typical of an exogenous genetic element. The presence of fragment of mobility elements, transposase remnants and insertion sequence in this region was also consistent with an acquisition. The genome of *Planktothrix serta* PCC 8927, slightly enriched in predicted RGP compared to other *Planktothrix* genomes (data not shown), could have favoured this event. The *has* gene clusters present in the genomes of other hassallidin producers were not associated with features of genomic plasticity such as nucleotide composition bias or presence of mobility elements ([28], this study). In cyanobacteria, other natural product biosynthetic gene clusters exhibited features of genomic plasticity[28], and this phenomenon has been shown in other bacteria with mobile genetic islands dedicated to natural product synthesis.[33] Our results provide traces of the successful transfer of a natural product biosynthetic gene cluster.

The organization of *Planktothrix serta* PCC 8927 *has* gene cluster was further compared to counterparts available from five heterocystous cyanobacteria (four hassallidin producers and the non-producing *Anabaena* sp. 90). The content of the *has* gene cluster in the six strains was well conserved contrary to the architecture of the gene cluster that was heavily rearranged (Figure 4). The hassallidin biosynthetic gene cluster of *Planktothrix serta* PCC 8927 was relatively compact with only 13 *has* genes from the reference cluster of *Anabaena* sp. SYKE748A, and possessed three additional ORFs, including ORF2 encoding a putative O-acyl-transferase. Moreover, several of the 26 *has* genes of *Anabaena* sp. SYKE748A were absent of the homologous gene cluster in *Planktothrix serta* PCC 8927 notably among the tailoring enzymes (genes missing: *hasC, D, E, G, H, I, J, P, Q, R, S, W*, and *X*). The overall comparison of the *has* gene cluster revealed that only nine genes are conserved in the six strains that might be sufficient to generate the backbone of the hassallidins. High variability of tailoring enzymes of similar biosynthetic gene cluster in various cyanobacteria have been extensively described in the case of the aeruginosin gene cluster[34] or more specifically for some tailoring

enzymes such as halogenases.[35] A high level of recombination in NRPS/PKS gene cluster of comparable size to the one of hassallidin has also been reported for cylindrospermopsin, that exhibited signs of horizontal gene transfer[36] and transfers between distinct cyanobacterial genera.[37]

Based on the gene cluster rearrangement and content in tailoring enzymes, we hypothesized that *Planktothrix serta* PCC 8927 produced a novel hassallidin variant. We tested this hypothesis by chemical analysis using liquid chromatography coupled to mass spectrometry. When the methanol extract of *Planktothrix sera* PCC 8927 was analyzed with UPLC-QTOFMS and LC-ITMS, the total ion current chromatograms revealed only one prominent peak in each condition (R_t 4.07 and 27.8 min, respectively) (Figures S1 and S2). This peak turned out to be a new hassallidin variant named hereafter hassallidin E, which contained a cyclic nonadepsipeptide with a dehydrobutyric acid (dhb), a lipid part made of 2,3-dihydroxyhexadecanoic acid (dhh), and a hexose as carbohydrate component (Figure 5). This hassallidin E was distinguished from hassallidin A, by the length of the fatty acid chain (Figure 5). Indeed, the mass of hassallidin E (*m/z* 1410 $[M+H]^+$) was 28 Da larger than that of hassallidin A. Fragments from protonated hassallidin E indicated that this delta mass was due to the C_{16} fatty acid chain of hassallidin E instead of C_{14} in the hassallidin A (Figures S1 and S2; Table 2). Hassallidin E differed in one sugar moiety from hassallidins B, C and D. Product ion spectrum from sodiated open chain methyl ester (*m/z* 1464) yielded the sequence of amino acids and location of the hexose (Figure S1). In all the hassallidins A-D and balticidins A-D previously described in cyanobacteria, a hexose sugar mannose was attached to the methylated threonine.[17, 18, 21, 27] A local maximum of 222 nm in the UV spectrum of the peak eluting at 27.8 min is in agreement with the peptidic nature of the hassallidin E. Moreover, low intensity local maximum at 278 nm was consistent with the presence of tyrosine in hassallidin E (Figure S3). The presence of dehydrobutyrine, threonine, methyl threonine, glutamine and tyrosine was confirmed by the corresponding immonium ions observed in the QTOF product ion spectrum of protonated hassallidin E (Figure S1, Table 2). These first analyses of *Planktothrix serta* PCC 8927 extracts revealed only one congener of hassallidin. To test whether the growth conditions given to the organism could impact this production, the strain PCC 8927 was cultivated in the same medium in which the heterocystous cyanobacteria produce

numerous variants of hassallidins.[21] The temperatures and the day/night rhythm were also tested. *Planktothrix serta* PCC 8927 produced the unique variant hassallidin E regardless of the growing conditions. Cyanobacteria rarely produce a single variant of a natural product, but it occurred with a single congener of aeruginosin in *Microcystis aeruginosa* NIES-843.[34] On the contrary, several examples showed multiple variants produced by single strains, including 40 structural hassallidin variants produced by *Anabaena* SYKE 748A.[21] The growth experiments in different media indicated however that hassallidin production per mg of dry weight was higher at 25°C than at 18° C, but also that the yield of hassallidin was significantly higher in BG11 medium than in Z8 medium (Figure S4). Both media differ in $NaNO_3$ concentration, which was almost 20-fold higher in BG11 compared to Z8 and thus, could impact the production of peptides such as hassallidin. As *Planktothrix serta* PCC 8927 is capable of fixing nitrogen,[29] this suggested a potential link between hassallidin production and nitrogen metabolism in the strains known to perform both.

Planktothrix serta PCC 8927 possessed the shortest *has* gene cluster among all hassallidin producer strains with available genomic sequence. Chemical analyses confirmed the antifungal activity and further revealed the new variant produced by strain PCC 8927. To better understand the link between the genetic variability of the cluster and variant structure and number, we investigated the *has* gene cluster composition and sequence. The NRPS catalytic domain organization was identical in all producers but the gene order differed. The substrate specificity of the NRPS enzymes encoded in the strain PCC 8927 (this study) and *Anabaena* sp. SYKE 748A[21] differed. However, this difference in substrate specificity did not seem to affect the peptide core of the hassalladin. The *has* gene cluster of PCC 8927 encoded a subset of the genes found in *Anabaena* SYKE 748A. In addition to the 13 *has* genes, the PCC 8927 hassallidin gene cluster encoded an extra acetyltransferase that was not found in the other versions of this gene cluster. However, the hassallidin E chemical structure did not contain acetyl group and no obvious role can be attributed to this enzyme in the synthesis of hassallidin E. This gene could be either non functional, or the encoded enzyme was involved in an intermediate step of the synthesis. Another inconsistency was the presence of two glycosyltransferases in the cluster while only one sugar was present in the hassallidin E structure. Similarly, there were more glycosyltransferase in the

Anabaena sp. SYKE 748A gene cluster than sugars on its diverse hassallidins. Overall, the gene cluster showed potential for more structural variant synthesis but only one variant, hassallidin E, was detected in all tested conditions. The capacity to produce multiple structural variants is considered an adaptive advantage for producer cells[38], while the selection of a single variant could correspond to a more specialized strategy.

Cyanobacterial hassallidins, including balticidins from *Anabaena cylindrica*[27], belong to a large family of chemical compounds produced from disparate microbes such as Sch 20561 and 20562 from *Aeromonas* sp.[39, 40] and herbicolin from *Herwinia herbicola*.[41] These molecules have been independently shown to exhibit antifungal properties. The hassallidin E was further tested to assess the antifungal activity of this new variant. A first disk diffusion assay with the crude extract of *Planktothrix serta* PCC 8927 showed an antifungal activity against *Candida albicans* HAMBI 485 (Figure S5), although slightly reduced when compared with a crude extract of the hassallidin D producer *Anabaena* sp. 258. As other metabolites could act in a synergistic way to alter hassallidin bioactivity[19], we further investigated the antifungal properties of purified hassallidin E.[21] A serial dilution assay of the hassallidin E from *Planktothrix serta* PCC 8927 indicated a MIC value of 32 mg/L (23 μM) against both *Candida* and *Cryptococcus* species tested. The purified hassallidin E used in this assay was confirmed as the cyclic form of the molecule by mass spectrometry (Figure S6). Difference between crude extract activity and purified compound had already been noted for hassallidin D and its producer *Anabaena* sp. SYKE748A.[21] The overall evaluation of *Planktothrix serta* PCC 8927 confirmed an antifungal activity of the native cyclic form of hassallidin variant E when tested against human pathogenic fungi (this study), although this activity is lower than that of hassallidin A-D.[17, 18, 21] The structural differences between variants could explain this reduced activity, as it could be affected by the absence of sugar acetylation or by the greater length of the fatty acid chain of the hassallidin E.

Antifungal properties of hassallidin might be ecologically relevant for *Planktothrix* as the need for a defense system against parasitic fungi could be the selective pressure that contributed to maintain the hassallidin gene cluster acquired by horizontal gene transfer. Planktic strains of *Planktothrix* have been previously shown to be susceptible to parasitic fungi chytrids[24], as well as other planktic cyanobacteria, including *Anabaena*[42], more commonly known as a hassallidin producer. No benthic cyanobacteria

have been shown to suffer from fungal parasitism yet, mostly due to the lack of studies. The niche *Planktothrix serta* PCC 8927, isolated from a sewage plant, is uncommon for this genus and poorly characterized in terms of evolutionary pressure sustained by the cyanobacterial strains. Such an environment and the need for adaptation might promote the selection of new or acquired compounds. The antifungal activity of hassallidin remains to be investigated against natural fungal parasites of cyanobacteria or ecosystem competitors, compared with other cyanobacterial compounds that might deter parasitic chytrids, and further completed using knock-out mutants of hassallidin producing strains. This will lead to better assess hassallidin functional role for *Planktothrix*, and more generally for cyanobacterial cells.

In summary, we characterized hassallidin E, a new variant of the hassallidin family from the filamentous cyanobacterium *Planktothrix serta* PCC 8927. This finding opened the way for a better understanding of the mechanisms leading to variant diversity through their dedicated gene clusters and biosynthetic pathways. This study revealed the first gene cluster coding for an antifungal glycosylated lipopeptide in a non-heterocystous cyanobacterial strain, and lead to expect other hassallidin producers. The localization of the cluster in a region exhibiting marks of a past horizontal gene transfer event contributed to explain its unusual presence in this benthic cyanobacterium. This further supported the role for horizontal gene transfer in natural product distribution across the cyanobacterial phylum, and illustrated both the potential and threat of finding bioactive compounds in unexpected cyanobacterial genera and their environments.

METHODS

***Planktothrix* strain.** The benthic strain *Planktothrix serta* PCC 8927 was originally isolated from a sewage plant of Berre-le-Clos, France in 1989, subsequently made axenic and maintained in the Pasteur Culture Collection of Cyanobacteria (PCC). The *Planktothrix serta* PCC 8927 genome sequence is available at GenBank under accession number CZCU00000000.[29]

Genomic comparison. Gene cluster and genomic context analysis were performed on the MicroScope platform[43] including *in silico* prediction of natural product biosynthetic

pathways using antiSMASH 3.0.[44] Protein BLAST was performed using NCBI BlastP[45] to manually annotate the genes of the hassallidin locus. The substrate specificity of the NRPS adenylation domains was assessed using NRPSpredictor2.[46] Regions of genomic plasticity (RGP) features were analyzed using RGPfinder including SIGI-HMM[47] and AlienHunter (IVOM)[48] as implemented on the MicroScope platform.

Phylogeny. The species tree was generated by a concatenation of twenty-nine conserved proteins using the Maximum Likelihood method as previously described.[28] Gene cluster phylogenetic tree was reconstructed on the filtered alignment of the four concatenated NRPS genes (*hasO*, *hasN*, *hasV* and *hasY*). Briefly, genes were independently aligned using MUSCLE[49], then the four alignments were filtered with Gblocks[50] and further concatenated. The tree was reconstructed on the concatenated alignment using PhyML 3.0[51] with a GTR (generalize time-reversible) substitution model and bootstrap analysis (1000 replicates). The tree was midpoint rooted. Visualization of gene cluster rearrangement was performed using genoPlotR package.[52]

Cyanobacterial culture. *Planktothrix serta* PCC 8927 was grown in 1.2 L BG11 medium[53] with 10 mM $NaHCO_3$ and bubbling 1% CO_2 under continuous light at 20 µmol photon.m^{-2}.s^{-1} to obtain enough biomass for genomic and chemical analyses. To test the production of hassallidin in different conditions, 40 mL cultures of the strain were inoculated in biological triplicates. They were maintained in media Z8[54] or BG11 at 18°C (Dark/Light 11/13h), 22°C (Dark/Light 11/13h) or 25°C (continuous light). All cultures were inoculated with cells originating from the same single culture grown in BG11. The cyanobacterial cultures were centrifuged after a 5-week incubation, washed in distilled water, frozen in liquid nitrogen and lyophilized prior further investigation.

Preparation of the extracts and LC-MS. Freeze dried *Plankthotrix serta* PCC 8927 biomass from 40 mL cultivation (between 5 – 20 mg of dry weight) was extracted with 1 mL of methanol as previously described,[55] and further detailed in Supporting Information.

Purification and analysis of hassallidin E. 1.5 g of freeze dried biomass of PCC 8927 was disrupted with glass beads in a mortar. Extraction was performed twice with

acetonitrile/DMSO (1:1) followed by centrifugation (9,000 g, 5 min at 20°C). The supernatant was collected and the acetonitrile was evaporated with a rotavapor at 40°C. The remaining solution was diluted in 20 volumes of water and loaded on a STRATA C18-E column (55 µm, 70 Å, 5 g/20 mL, Phenomenex). The column was rinsed with 60 mL of water/acetonitrile (95:5) and eluted with 60 mL of water/acetonitrile (10:90). After evaporation of this fraction and freeze-drying, the residue (10 mg) was resuspended in water/acetonitrile (80:20) and submitted in two batches to a second purification step on a STRATA C18-U cartridge (55 µm, 70 Å, 100 mg - 1 mL, Phenomenex) using a step gradient with water/acetonitrile (80:20, 70:30, 60:40, 50:50, 40:60, 25:75 and then 0:100).

The fractions were analyzed by LC-MS on an Alliance 2695 system coupled to a UV detector 2487 (220 nm) and to a Q-TofmicroTM spectrometer (MICROMASS) with an electrospray ionization (positive mode) source (Waters, France). The linear gradient was performed with acetonitrile + 0.025% formic acid (A) / water + 0.04% trifluoroacetic acid + 0.05% formic acid (B) over 20 min at 45°C. The column was a XBRIDGE BEH300 C18 (3.5 µm, 2.1x100 mm) (Waters) (gradient 5-100 % A). The source temperature was maintained at 120°C and the desolvation temperature at 400°C. The cone voltage was 40 V. The fractions containing hassallidin (from 50 to 75% acetonitrile, retention time 12.6 min) were combined and freeze-dried to yield 1.4 mg of compound. A final purification by HPLC was performed using an Agilent 1200 pump system with a UV detector at 220 and 280 nm. A Kinetex C18 column (5 µm, 100 Å, 250x4.6 mm, Phenomenex) was used, and a gradient of acetonitrile / water + 0.1 % TFA was applied over 20 min at 25°C, from 10 to 100 %. The peak eluted at 14.2 min was freeze-dried to yield 0.5 mg of white amorphous powder. The quantity evaluation was based on the UV absorbance at 226 nm measured by NanoDrop and on the molar extinction coefficient reported for hassallidin A.[17] The MS analysis of the purified hassallidin was recorded in the positive mode by direct infusion in the same spectrometer with a source temperature and a desolvatation temperature maintained at 80°C and 250°C, respectively. The samples were dissolved in water/acetonitrile (1:1) containing 0.1 % formic acid. MaxEnt 3 Software (Waters, France) was used for the deconvolution of mass spectra.

Antifungal assays. Crude extracts were prepared by methanol extraction of 100 mg of freeze-dried material as previously described.[21] A disk diffusion assay was performed by applying 300 µL of the methanol extract from freeze dried culture of *Planktothrix serta* PCC 8927 to a paper disc (Abteck Biologicals Ltd., Liverpool, United Kingdom), subsequently dried at room temperature. A second disk containing only 300 µL of methanol was prepared as a negative control and assessment of methanol toxicity. Positive controls were prepared with either a 300 µL of methanol extract of *Anabaena* sp. 258 (Hassallidin D producer[21]) or with 5 µL of Nystatin (5 mg/mL). *Candida albicans* HAMBI 485 was grown in yeast and mold agar medium. A 0.5 McFarland suspension of the fungal strain was prepared into 0.145 M NaCl solution and spread on agar plate. The disks were placed onto surface of agar plate containing the yeast and plates were incubated 24 hours at 37°C.

Antifungal activity of pure hassallidin E was determined *in vitro* by using broth microdilution method similar to the EUCAST method.[56] Antifungal activity was determined for *Candida albicans* CBS562, *Cryptococcus neoformans* H99, *Candida parapsilosis* ATCC22019 and *Candida krusei* ATCC6258. Inoculum of yeast waere prepared in sterile water, by using a Sabouraud agar culture of 24 hours. An inoculum of each strain was prepared in order to obtain a working preparation with a final concentration of 1-2.5 10^5 CFU/mL. RPMI1640 medium was prepared by two-fold dilution to obtain final working solutions of 1 % glucose and 32 mg/L to 0.067 mg/L of hassallidin E in duplicate, as well as a blank control with no inoculum and a growth control with an inoculum in a hassallidin-free medium. Plates were incubated at 35°C for 24 hours. Optical densities were determined for each well at 492 nm. Growth was considered as sufficient if OD was ≥0.2 after subtraction of blank value to growth control well. Finally, minimum inhibitory concentration (MIC) was determined as the lowest concentration (in mg/L) of hassallidin inhibiting the growth of yeast.

ASSOCIATED CONTENT

* Supporting information

Methods of preparation of the extracts for LC-MS are detailed. Six supplementary figures related to chromatograms and mass spectra from methanol extract of *Planktothrix serta* PCC 8927 obtained by UPLC-QTOFMS, obtained by LC-ITMS, effects of growth medium and temperature on hassallidin E production, disk diffusion assays and mass spectrum

of the purified hassallidin E used for antifungal assays, as indicated in the text. The supporting information is available free of charge on the ACS Publications Website (PDF).

AUTHOR INFORMATION Corresponding Author

*Email: muriel.gugger@pasteur.fr

NOTE

The authors declare no competing financial interest.

AKNOWLEDGEMENTS

This work was partially supported by the French National Research Agency under Contract N° ANR-15-CE34-0002-02, by incentive credits from the Conny Maeva Charitable Foundation. The Pasteur Culture Collection of Cyanobacteria was supported by the Institut Pasteur. CP was supported by the Ile-de-France ARDoC Grant for PhD. The work at the University of Helsinki was supported by the Academy of Finland grants (258827 and 273798) to KS and by the Academy of Finland grant (259505) to DPF. AH is student at the Doctoral Programme in Microbiology and Biotechnology, University of Helsinki. The authors thank Frédéric Bonhomme for the mass spectrometry analysis of the purified hassallidin E.

REFERENCES

[1] Francis, G. (1878) Poisonous Australian lakes, *Nature 18*, 11-12.

[2] Oliver, R. L., and Ganf, G. G. (2000) Freshwater blooms, In *The Ecology of Cyanobacteria - Their diversity in time and space* (Whitton, B. A., and Potts, M., Eds.), pp 149-194, Kluwer Academic Publishers, Dordrecht.

[3] Hudnell, H. K., and Dortch, Q. (2008) A synopsis of research needs identified at the Interagency, International Symposium on Cyanobacterial Harmful Algal Blooms (ISOC-HAB), In *Cyanobacterial Harmful Algal Blooms - State of the Science and Research Needs* (Hudnell, H. K., Ed.), pp 17-43, Springer, New York.

[4] Gugger, M., Lenoir, S., Berger, C., Ledreux, A., Druart, J. C., Humbert, J. F., Guette, C., and Bernard, C. (2005) First report in a river in France of the benthic cyanobacterium *Phormidium favosum* producing anatoxin-a associated with dog neurotoxicosis, *Toxicon 45*, 919-928.

[5] Cadel-Six, S., Peyraud-Thomas, C., Brient, L., de Marsac, N. T., Rippka, R., and Méjean, A. (2007) Different genotypes of anatoxin-producing cyanobacteria coexist in the Tarn River, France, *Appl Environ Microbiol 73*, 7605-7614.

[6] Wood, S. A., Selwood, A. I., Rueckert, A., Holland, P. T., Milne, J. R., Smith, K. F., Smits, B., Watts, L. F., and Cary, C. S. (2007) First report of homoanatoxin-a and associated dog neurotoxicosis in New Zealand, *Toxicon 50*, 292-301.
[7] Quiblier, C., Wood, S., Echenique-Subiabre, I., Heath, M. W., Villeneuve, A., and Humbert, J. F. (2013) A review of current knowledge on toxic benthic freshwater cyanobacteria- Ecology, toxin production and risk management, *Water Research 47*, 5464-5479.
[8] Severin, I., Acinas, S. G., and Stal, L. J. (2010) Diversity of nitrogen-fixing bacteria in cyanobacterial mats, *FEMS Microbiol Ecol 73*, 514-525.
[9] Dittmann, E., Gugger, M., Sivonen, K., and Fewer, D. P. (2015) Natural product biosynthetic diversity and comparative genomics of the Cyanobacteria, *Trends Microbiol 23*, 642-652.
[10] Tan, L. T. (2007) Bioactive natural products from marine cyanobacteria for drug discovery, *Phytochemistry 68*, 954-979.
[11] Niedermeyer, T. H. (2015) Anti-infective natural products from Cyanobacteria, *Planta Med 81*, 1309-1325.
[12] Shishido, T. K., Humisto, A., Jokela, J., Liu, L., Wahlsten, M., Tamrakar, A., Fewer, D. P., Permi, P., Andreote, A. P., Fiore, M. F., and Sivonen, K. (2015) Antifungal compounds from cyanobacteria, *Mar Drugs 13*, 2124-2140.
[13] Moon, S. S., Chen, J. L., Moore, R. E., and Patterson, G. M. L. (1992) Calophycin, a Fungicidal Cyclic Decapeptide from the Terrestrial Blue-Green-Alga Calothrix-Fusca, *J Org Chem 57*, 1097-1103.
[14] Kim, H., Lantvit, D., Hwang, C. H., Kroll, D. J., Swanson, S. M., Franzblau, S. G., and Orjala, J. (2012) Indole alkaloids from two cultured cyanobacteria, Westiellopsis sp and Fischerella muscicola, *Bioorgan Med Chem 20*, 5290-5295.
[15] Salvador-Reyes, L. A., Engene, N., Paul, V. J., and Luesch, H. (2015) Targeted natural products discovery from marine cyanobacteria using combined phylogenetic and mass spectrometric evaluation, *J Nat Prod 78*, 486-492.
[16] Mo, S. Y., Krunic, A., Pegan, S. D., Franzblau, S. G., and Orjala, J. (2009) An antimicrobial guanidine-bearing sesterterpene from the cultured cyanobacterium *Scytonema* sp., *J Nat Prod 72*, 2043-2045.
[17] Neuhof, T., Schmieder, P., Preussel, K., Dieckmann, R., Pham, H., Bartl, F., and von Dohren, H. (2005) Hassallidin A, a glycosylated lipopeptide with antifungal activity from the cyanobacterium *Hassallia* sp., *J Nat Prod 68*, 695-700.
[18] Neuhof, T., Schmieder, P., Seibold, M., Preussel, K., and von Dohren, H. (2006) Hassallidin B--second antifungal member of the Hassallidin family, *Bioorg Med Chem Lett 16*, 4220-4222.
[19] Shishido, T. K., Jokela, J., Kolehmainen, C. T., Fewer, D. P., Wahlsten, M., Wang, H., Rouhiainen, L., Rizzi, E., De Bellis, G., Permi, P., and Sivonen, K. (2015) Antifungal activity improved by coproduction of cyclodextrins and anabaenolysins in Cyanobacteria, *Proc Natl Acad Sci U S A 112*, 13669-13674.
[20] Castenholz, R. W. (2001) Phylum BX, In *Bergey's Manual of Systematic Bacteriology - 1: The Archaea and the Deeply Branching and Phototrophic Bacteria* (Garrity, G. M., Boone, D. R., and Castenholz, R. W., Eds.) 2 ed., pp 473-899, Springer-Verlag, New York.
[21] Vestola, J., Shishido, T. K., Jokela, J., Fewer, D. P., Aitio, O., Permi, P., Wahlsten, M., Wang, H., Rouhiainen, L., and Sivonen, K. (2014) Hassallidins, antifungal glycolipopeptides, are widespread among cyanobacteria and are the end-product of a nonribosomal pathway, *Proc Natl Acad Sci U S A 111*, E1909-1917.

[22] Gerphagnon, M., Macarthur, D. J., Latour, D., Gachon, C. M., Van Ogtrop, F., Gleason, F. H., and Sime-Ngando, T. (2015) Microbial players involved in the decline of filamentous and colonial cyanobacterial blooms with a focus on fungal parasitism, *Environ Microbiol 17*, 2573-2587.
[23] Rasconi, S., Niquil, N., and Sime-Ngando, T. (2012) Phytoplankton chytridiomycosis: community structure and infectivity of fungal parasites in aquatic ecosystems, *Environ Microbiol 14*, 2151-2170.
[24] Sonstebo, J. H., and Rohrlack, T. (2011) Possible implications of chytrid parasitism for population subdivision in freshwater cyanobacteria of the genus *Planktothrix*, *Appl Environ Microbiol 77*, 1344-1351.
[25] Rohrlack, T., Christiansen, G., and Kurmayer, R. (2013) Putative antiparasite defensive system involving ribosomal and nonribosomal oligopeptides in cyanobacteria of the genus *Planktothrix*, *Appl Environ Microbiol 79*, 2642-2647.
[26] Kyle, M., Haande, S., Ostermaier, V., and Rohrlack, T. (2015) The Red Queen race between parasitic chytrids and their host, *Planktothrix*: a test using a time series reconstructed from sediment DNA, *PLoS One 10*, e0118738.
[27] Bui, T. H., Wray, V., Nimtz, M., Fossen, T., Preisitsch, M., Schroder, G., Wende, K., Heiden, S. E., and Mundt, S. (2014) Balticidins A-D, antifungal hassallidin-like lipopeptides from the Baltic Sea cyanobacterium *Anabaena cylindrica* Bio33, *J Nat Prod 77*, 1287-1296.
[28] Calteau, A., Fewer, D. P., Latifi, A., Coursin, T., Laurent, T., Jokela, J., Kerfeld, C. A., Sivonen, K., Piel, J., and Gugger, M. (2014) Phylum-wide comparative genomics unravel the diversity of secondary metabolism in Cyanobacteria, *BMC genomics 15*, 977.
[29] Pancrace, C., Barny, M., Ueoka, R., Calteau, A., Scalvenzi, T., Pedron, J., Barbe, V., Piel, J., and Humbert, J. (2017) Insights into the *Planktothrix* genus: Genomic and metabolic comparison of benthic and planktic strains, *Sci Rep 7*, 41181.
[30] Kurmayer, R., Deng, L., and Entfellner, E. (2016) Role of toxic and bioactive secondary metabolites in colonization and bloom formation by filamentous cyanobacteria *Planktothrix*, *Harmful Algae 54*, 69-86.
[31] Ueoka, R., Uria, A. R., Reiter, S., Mori, T., Karbaum, P., Peters, E. E., Helfrich, E. J., Morinaka, B. I., Gugger, M., Takeyama, H., Matsunaga, S., and Piel, J. (2015) Metabolic and evolutionary origin of actin-binding polyketides from diverse organisms, *Nat Chem Biol 11*, 705-712.
[32] Tooming-Klunderud, A., Sogge, H., Rounge, T. B., Nederbragt, A. J., Lagesen, K., Glockner, G., Hayes, P. K., Rohrlack, T., and Jakobsen, K. S. (2013) From green to red: horizontal gene transfer of the phycoerythrin gene cluster between *Planktothrix* strains, *Appl Environ Microbiol 79*, 6803-6812.
[33] Penn, K., Jenkins, C., Nett, M., Udwary, D. W., Gontang, E. A., McGlinchey, R. P., Foster, B., Lapidus, A., Podell, S., Allen, E. E., Moore, B. S., and Jensen, P. R. (2009) Genomic islands link secondary metabolism to functional adaptation in marine Actinobacteria, *Isme J 3*, 1193-1203.
[34] Ishida, K., Welker, M., Christiansen, G., Cadel-Six, S., Bouchier, C., Dittmann, E., Hertweck, C., and Tandeau de Marsac, N. (2009) Plasticity and evolution of aeruginosin biosynthesis in cyanobacteria, *Appl Environ Microbiol 75*, 2017-2026.
[35] Cadel-Six, S., Dauga, C., Castets, A. M., Rippka, R., Bouchier, C., Tandeau de Marsac, N., and Welker, M. (2008) Halogenase genes in nonribosomal peptide synthetase gene clusters of *Microcystis* (cyanobacteria): sporadic distribution and evolution, *Mol Biol Evol 25*, 2031-2041.

[36] Stuken, A., and Jakobsen, K. S. (2010) The cylindrospermopsin gene cluster of *Aphanizomenon* sp. strain 10E6: organization and recombination, *Microbiol-Sgm 156*, 2438-2451.
[37] Kellmann, R., Mills, T., and Neilan, B. A. (2006) Functional modeling and phylogenetic distribution of putative cylindrospermopsin biosynthesis enzymes, *J Mol Evol 62*, 267-280.
[38] Firn, R. D., and Jones, C. G. (2000) The evolution of secondary metabolism - a unifying model, *Mol Microbiol 37*, 989-994.
[39] Afonso, A., Hon, F., and Brambilla, R. (1999) Structure elucidation of Sch 20562, a glucosidic cyclic dehydropeptide lactone - the major component of W-10 antifungal antibiotic, *J Antibiot 52*, 383-397.
[40] Afonso, A., Hon, F., Brambilla, R., and Puar, M. S. (1999) Structure elucidation of Sch 20561, a cyclic dehydropeptide lactone - a major component of W-10 antifungal antibiotic, *J Antibiot 52*, 398-406.
[41] Aydin, M., Lucht, N., Konig, W. A., Lupp, R., Jung, G., and Winkelmann, G. (1985) Structure elucidation of the peptide antibiotics Herbicolin-A and Herbicolin-B, *Liebigs Ann Chem*, 2285-2300.
[42] Gerphagnon, M., Latour, D., Colombet, J., and Sime-Ngando, T. (2013) Fungal Parasitism: Life Cycle, Dynamics and Impact on Cyanobacterial Blooms, *PLoS One 8*, e60894.
[43] Vallenet, D., Calteau, A., Cruveiller, S., Gachet, M., Lajus, A., Josso, A., Mercier, J., Renaux, A., Rollin, J., Rouy, Z., Roche, D., Scarpelli, C., and Medigue, C. (2017) MicroScope in 2017: an expanding and evolving integrated resource for community expertise of microbial genomes, *Nucleic Acids Res 45*, D517-D528.
[44] Weber, T., Blin, K., Duddela, S., Krug, D., Kim, H. U., Bruccoleri, R., Lee, S. Y., Fischbach, M. A., Muller, R., Wohlleben, W., Breitling, R., Takano, E., and Medema, M. H. (2015) antiSMASH 3.0-a comprehensive resource for the genome mining of biosynthetic gene clusters, *Nucleic Acids Res 43*, W237-243.
[45] Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J. H., Zhang, Z., Miller, W., and Lipman, D. J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs, *Nucleic Acids Res 25*, 3389-3402.
[46] Rottig, M., Medema, M. H., Blin, K., Weber, T., Rausch, C., and Kohlbacher, O. (2011) NRPSpredictor2-a web server for predicting NRPS adenylation domain specificity, *Nucleic Acids Res 39*, W362-W367.
[47] Waack, S., Keller, O., Asper, R., Brodag, T., Damm, C., Fricke, W. F., Surovcik, K., Meinicke, P., and Merkl, R. (2006) Score-based prediction of genomic islands in prokaryotic genomes using hidden Markov models, *BMC bioinformatics 7*, 142.
[48] Vernikos, G. S., and Parkhill, J. (2006) Interpolated variable order motifs for identification of horizontally acquired DNA: revisiting the Salmonella pathogenicity islands, *Bioinformatics 22*, 2196-2203.
[49] Edgar, R. (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput, *Nucleic Acids Res 32*, 1792-1797.
[50] Talavera, G., and Castresana, J. (2007) Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments, *Syst Biol 56*, 564-577.
[51] Guindon, S., Dufayard, J. F., Lefort, V., Anisimova, M., Hordijk, W., and Gascuel, O. (2010) New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0, *Syst Biol 59*, 307-321.

[52] Guy, L., Kultima, J. R., and Andersson, S. G. (2010) genoPlotR: comparative gene and genome visualization in R, *Bioinformatics 26*, 2334-2335.
[53] Rippka, R. (1988) Isolation and purification of cyanobacteria, *Methods in enzymology 167*, 3-27.
[54] Kotai, J. (1972) Instructions for the preparation of modified nutrient solution Z8 for algae, (Norwegian Institute for Water Research, B., Oslo, Ed.), p 5 pp.
[55] Leikoski, N., Fewer, D. P., Jokela, J., Alakoski, P., Wahlsten, M., and Sivonen, K. (2012) Analysis of an inactive cyanobactin biosynthetic gene cluster leads to discovery of new natural products from strains of the genus *Microcystis*, *PLoS One 7*, e43002.
[56] Arendrup, M. C., Guinea, J., Cuenca-Estrella, M., Meletiadis, J., Mouton, J. W., Lagrou, K., Howard, S. J., AFST and EUCAST (2015). Method for the determination of broth dilution minimum Inhibitory concentrations of antifungal agents for yeasts, pp 1-21, EUCAST DEFINITIVE DOCUMENT E.DEF 7.3.

Table 1. The proposed function of proteins encoded in the hassallidin gene cluster and flanking ORFs in *Planktothrix serta* PCC 8927. The percentage of amino-acid identity (AAI) is indicated for each gene with the corresponding Best BLASTp hit. NRPS domains: C for condensation, A for adenylation, PP for peptidyl carrier protein, E for epimerase and Te for thioesterase.

Protein	Proposed function (NRPS modules with substrate prediction)	Best BLASTp hit	Accession N°	AAI %
CDS1	Putative FAD dependent oxidoreductase	*Planktothrix* sp.	WP_026794628	94
CDS2	Pentapeptide repeat-containing protein of unknown function	*Planktothrix* spp.	WP_036832156	88
CDS3	Fragment of putative transposase	*Planktothrix agardhii*	WP_042158606	96
HasK	Major facilitator transporter	*Anabaena* sp. 90	WP_015080893	75
HasF	Dolichyl-phostphate-mannose—protein mannosyltransferase	*Tolypothrix* sp. PCC 9009	WP_029637428	64
Orf1	Protein of unknown function	None	-	-
Orf2	Putative O-acyl-transferase	*Oscillatoria* sp. PCC 10802	WP_017715552	53
Orf3	Fragment of transposase	*Planktothrix agardhii*	WP_026787096	65
HasO	NRPS (C-A_{Thr}-PP-C-A_{Gln}-PP-E-C-A_{Gly}-PP)	*Hassallia byssoidea* VB512170	WP_039754184	68
HasN	NRPS (C-A_{Thr}-PP-E-C-A_{Tyr}-PP-E)	*Tolypothrix* sp. PCC 9009	WP_029631929	71
HasM	Aspartyl/asparginyl beta-hydroxylase family protein	*Tolypothrix* sp. PCC 9009	WP_029631927	81
HasL	3-oxoacyl-acyl-carrier-protein reductase	*Tolypothrix* sp. PCC 9009	WP_029631926	89
HasT	Glycosyltransferase, DXD motif-containing protein	*Tolypothrix* sp. PCC 9009	WP_029631937	75
HasA	ABC-type transporter, permease	*Anabaena* sp. SYKE748A	AHZ20760	73
HasZ	Major facilitator transporter	*Tolypothrix* sp. PCC 9009	WP_038296765	71
HasU	Conserved protein of unknown function, MbtH-like	*Anabaena* sp. 90	WP_015080903	86
HasV	NRPS (C-A_{Thr}-PP-C-A_{Thr}-PP)	*Tolypothrix* sp. PCC 9009	WP_029631940	75
HasB	Glycosyl hydrolase	*Tolypothrix* sp. PCC 9009	WP_029631931	78
HasY	NRPS (C-A_{Thr}-nMT-PP-C-A_{Gln}-PP-Te)	*Hassallia byssoidea* VB512170	WP_039753382	70
CDS4	Fragment of putative	*Gloeocapsa* sp. PCC	ELR99152	54

	transposase	73106		
CDS5	Fragment of relaxase/mobilization nuclease family protein	*Planktothrix agardhii*	WP_042158465	97
CDS6	Fragment of relaxase/mobilization nuclease family protein	*Planktothrix agardhii*	WP_042158465	85

Table 2. Ions structures, calculated and measured ion masses and error values for the hassallidin E analyzed with UPLC-QTOF. M-CH_4O = Linear hassallidin methyl ester, Hex = hexose residue, Dhh = 2,3-dihydroxyhexadecanoic acid.

No	Ion structures	$[M+H/Na]^+$ (*m/z*) Calculated	Measured	Error (ppm)
1	$[M\text{-}CH_4O + Na]^+$	1,464.7332	1,464.7281	-3.50
2	$[M\text{-}CH_4O + H]^+$	1,442.7512	1,442.7491	-1.51
3	$[M + Na]^+$	1,432.7070	1,432.7047	-1.61
4	$[M + H]^+$	1,410.7250	1,410.7223	-1.96
5	$[M - Q + H]^+$	1,282.6664	1,282.6689	1.88
6	$[M\text{-}CH_4O - Hex + H]^+$	1,280.6984	1,280.6954	-2.39
7	$[M - Hex + H]^+$	1,248.6722	1,248.6705	-1.36
8	$[M - (Hex+H_2O) + H]^+$	1,230.6616	1,230.6601	-1.28
9	$[M - (Hex+H_2O+NH_3) + H]^+$	1,213.6351	1,213.6327	-2.00
10	$[M - (Hex+Q) + H]^+$	1,120.6136	1,120.6125	-1.04
11	$[M - (Hex+Q+H_2O) + H]^+$	1,102.6031	1,102.6013	-1.64
12	$[M - (mT(Hex)\text{-}Q) + H]^+$	1,005.5503	1,005.5456	-4.71
13	$[M - (mT(Hex)\text{-}Q + H_2O) + H]^+$	987.5397	987.5398	0.03
14	[Dhh-T-T-T-Y-Dhb-Q + $H]^+$	948.5288	948.5267	-2.29
15	[Dhh-T-T-T-Y-Dhb-Q − H_2O + $H]^+$	930.5183	930.5153	-3.24
16	[T-T-Y-Dhb-Q-G-mT-Q − H_2O + $H]^+$	859.3945	859.3928	-1.99
17	[Dhh-T-T-T-Y − H_2O + $H]^+$	719.4226	719.4241	2.06
18	[Dhh-T-T-T − H_2O + $H]^+$	556.3592	556.3573	-3.58

Figure 1. The distribution of the cyanobacterial strains encoding the hassallidin biosynthetic gene clusters. The phylogenetic tree was generated by a concatenation of 29 conserved proteins using Maximum Likelihood. Black dots indicate a support with a bootstrap of ≥ 70%. The highlighted strains encode the hassallidin biosynthetic gene cluster, and they produce or have produced hassallidins. Hassallidin producers belong either to the clade of nostocalean or to the oscillatorian clade in the case of *Planktothrix serta* PCC 8927.

Figure 2. Phylogenetic tree of hassallidin NRPS genes. The tree was reconstructed on the filtered and concatenated alignments of *hasN*, *hasO*, *hasV* and *hasY* genes using Maximum Likelihood and rooted by midpoint rooting. Bootstrap values are indicated.

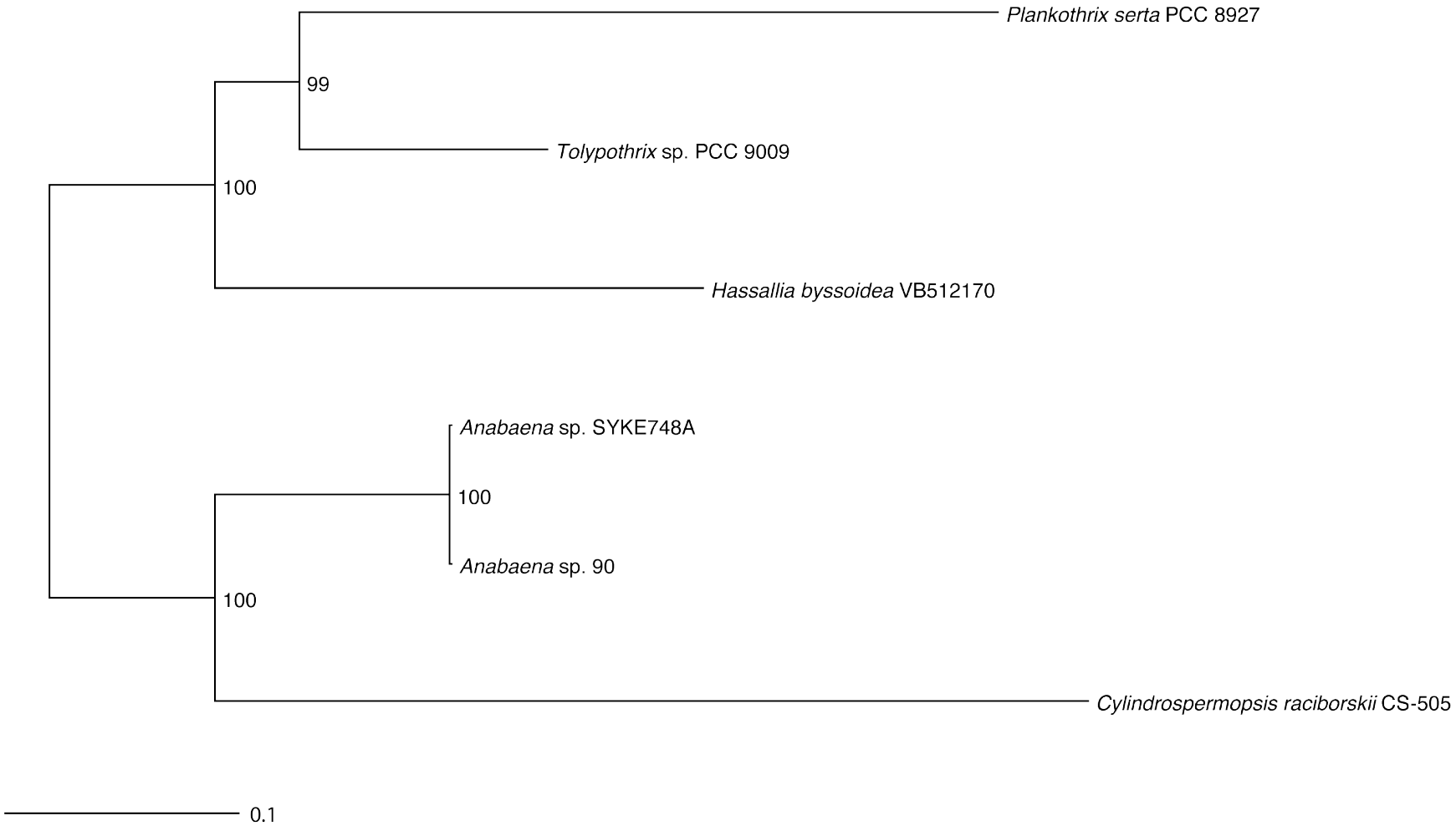

Figure 3. The region of genomic plasticity including the hassallidin locus in *Planktothrix serta* PCC 8927. This 95 kb region includes the hassallidin genes (blue) as well as transposase and integrase (yellow) and genes of other functions (white). The plotted graph represents the mean GC content of each gene (in %), the red line indicates the mean GC content of the genome and the asterixes indicate significant GC% deviation of the considered gene (+/- 1 or 2 Standard Deviation). The locus presenting an k-mers deviation (2-mers to 8-mers) deviation includes 48 genes presenting a biased use of nucleotides compared to the rest of the genome.

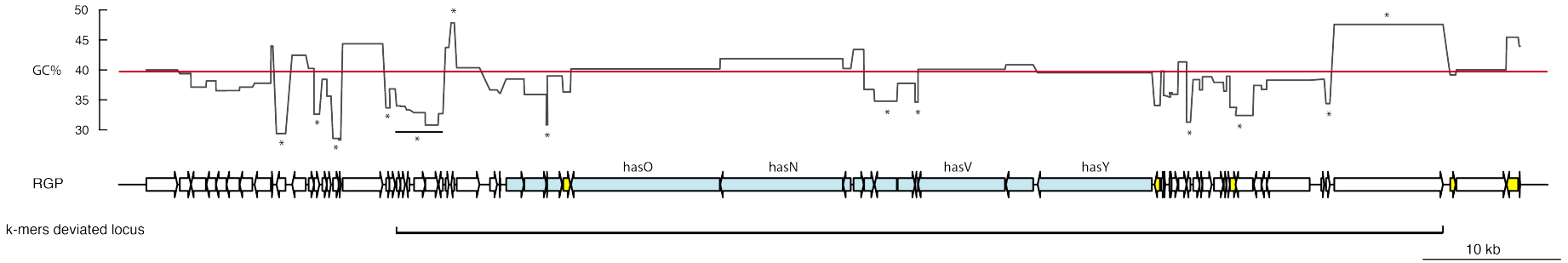

Figure 4. Hassallidin gene cluster rearrangement in six genomes. The hassallidin gene clusters include between 16 up to 26 genes encoding four NRPS (red), one ketoreductase (green), permease and transporter (brown), transposase (yellow) and other genes involved in hassallidin biosynthetic pathway (blue) and genes putatively involved (grey). The deletion in *hasV* gene of *Anabaena* sp. 90 is indicated by a triangle. The grey areas link conserved genes between the 6 pathways ordered in function of the phylogenetic relationship between the 4 NRPS genes.

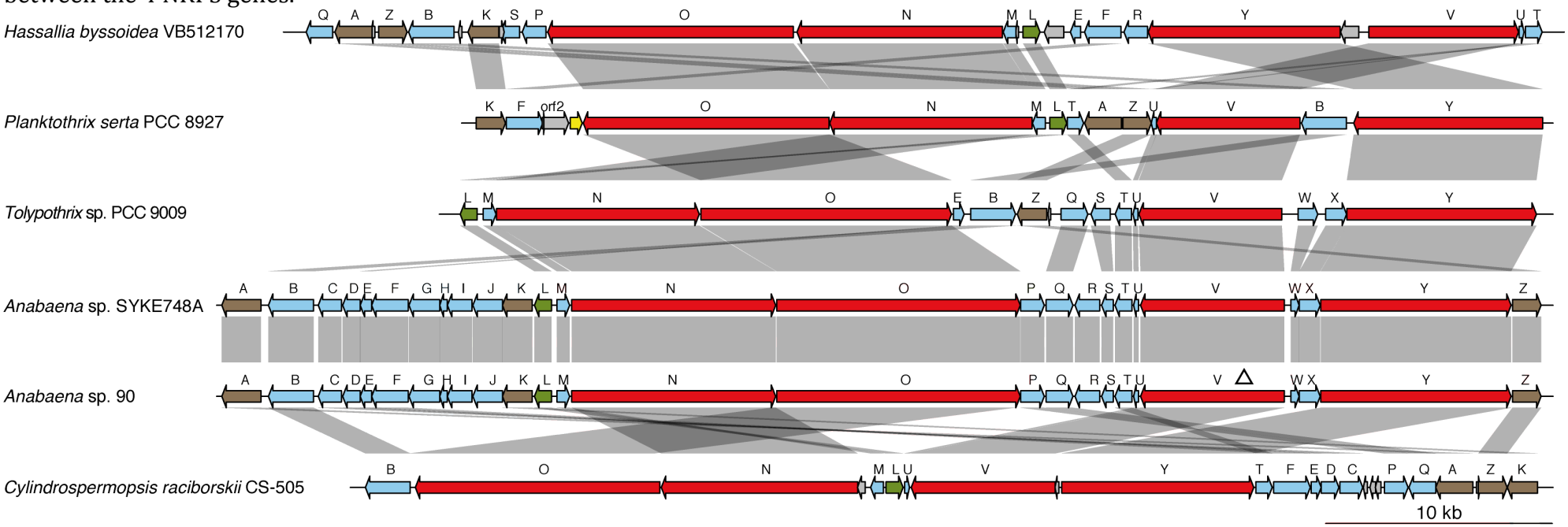
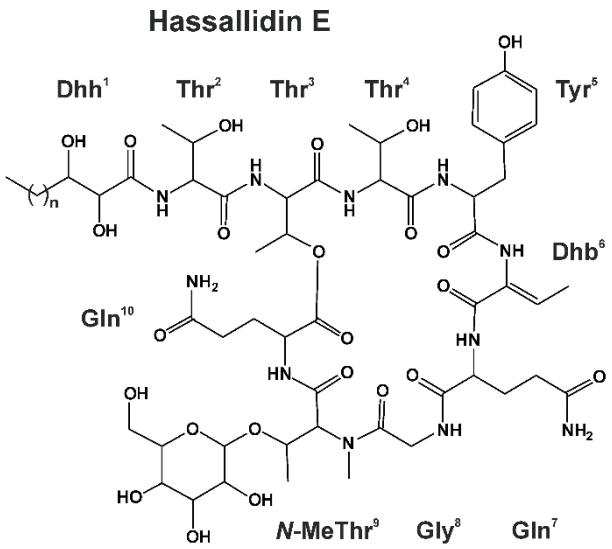
Figure 5. Structure of hassallidin E. Dhh = 2,3-dihydroxyhexadecanoic acid (n = 12), Dhb = dehydrobutyric acid. Hassallidin A has a closely related structure with n = 10.