

**This is a self-archived version of an original article. This version may differ from the original in pagination and typographic details.**

**Author(s):** Mareš, Jan; Hájek, Jan; Urajová, Petra; Kust, Andreja; Jokela, Jouni; Saurav, Kumar; Galica, Tomáš; Čapková, Kateřina; Mattila, Antti; Haapaniemi, Esa; Permi, Perttu; Mysterud, Ivar; Skulberg, Olav M.; Karlsen, Jan; Fewer, David P.; Sivonen, Kaarina; Tønnesen, Hanne Hjorth; Hrouzek, Pavel

**Title:** Alternative biosynthetic starter units enhance the structural diversity of cyanobacterial lipopeptides

**Year:** 2019

**Version:** Accepted version (Final draft)

**Copyright:** © 2019 American Society for Microbiology

**Rights:** In Copyright

**Rights url:** <http://rightsstatements.org/page/InC/1.0/?language=en>

**Please cite the original version:**

Mareš, J., Hájek, J., Urajová, P., Kust, A., Jokela, J., Saurav, K., Galica, T., Čapková, K., Mattila, A., Haapaniemi, E., Permi, P., Mysterud, I., Skulberg, O. M., Karlsen, J., Fewer, D. P., Sivonen, K., Tønnesen, H. H., & Hrouzek, P. (2019). Alternative biosynthetic starter units enhance the structural diversity of cyanobacterial lipopeptides. *Applied and Environmental Microbiology*, 85(4), Article e02675-18. <https://doi.org/10.1128/AEM.02675-18>

1 **Alternative biosynthetic starter units enhance the structural diversity of cyanobacterial**  
2 **lipopeptides**

3

4 Jan Mareš,<sup>1,2,3</sup> Jan Hájek,<sup>2,3</sup> Petra Urajová,<sup>2</sup> Andreja Kust,<sup>1,2,3</sup> Jouni Jokela,<sup>4</sup> Kumar Saurav,<sup>2</sup>

5 Tomáš Galica,<sup>2,3</sup> Kateřina Čapková,<sup>1</sup> Antti Mattila,<sup>4</sup> Esa Haapaniemi,<sup>5,6</sup> Perttu Permi,<sup>5</sup> Ivar

6 Mysterud,<sup>7</sup> Olav M. Skulberg,<sup>8</sup> Jan Karlsen,<sup>9</sup> David P. Fewer,<sup>4</sup> Kaarina Sivonen,<sup>4</sup> Hanne

7 Hjorth Tønnesen,<sup>9</sup> Pavel Hrouzek<sup>2,3#</sup>

8

9 <sup>1</sup>The Czech Academy of Sciences, Biology Centre, Institute of Hydrobiology, České  
10 Budějovice, Czech Republic

11 <sup>2</sup>The Czech Academy of Sciences, Institute of Microbiology, Center Algatech, Třeboň, Czech  
12 Republic

13 <sup>3</sup>University of South Bohemia, Faculty of Science, České Budějovice, Czech Republic

14 <sup>4</sup>Department of Microbiology, Biocenter 1, Helsinki, Finland

15 <sup>5</sup>Department of Chemistry, University of Jyväskylä, Finland

16 <sup>6</sup>Department of Biological and Environmental Science, Nanoscience Center, University of  
17 Jyväskylä, Finland

18 <sup>7</sup>Department of Biosciences, University of Oslo, Oslo, Norway

19 <sup>8</sup>Norwegian Institute for Water Research (NIVA), Oslo, Norway

20 <sup>9</sup>School of Pharmacy, University of Oslo, Oslo, Norway

21

22 Running title: Cyanobacterial lipopeptide biosynthesis

23

24 #Address correspondence to Pavel Hrouzek, hrouzekp@gmail.com

25

26 **Abstract:** Puwainaphycins (PUWs) and minutissamides (MINs) are structurally analogous  
27 cyclic lipopeptides possessing cytotoxic activity. Both types of compound exhibit high  
28 structural variability, particularly in the fatty acid (FA) moiety. Although biosynthetic gene  
29 clusters for several PUW variants has been proposed in a cyanobacterial strain, the genetic  
30 background for MINs remains unexplored. Herein, we report PUW/MIN biosynthetic gene  
31 clusters and structural variants from six cyanobacterial strains. Comparison of biosynthetic  
32 gene clusters indicates a common origin of the PUW/MIN hybrid nonribosomal peptide  
33 synthetase and polyketide synthase. Surprisingly, the gene clusters encode two alternative  
34 biosynthetic starter modules, and analysis of structural variants suggests that initiation by each  
35 of the starter modules results in lipopeptides of differing length and FA substitution. Among  
36 additional modifications of the FA chain, chlorination of minutissamide D was explained by  
37 the presence of a putative halogenase gene in the PUW/MIN gene cluster of *Anabaena*  
38 *minutissima* UTEX B 1613. We detected PUW variants bearing an acetyl substitution in  
39 *Symplocastrum muelleri* NIVA-CYA 644, consistent with an *O*-acetyltransferase gene in its  
40 biosynthetic gene cluster. The major lipopeptide variants did not exhibit any significant  
41 antibacterial activity, and only the PUW F variant was moderately active against yeast,  
42 consistent with previously published data suggesting that PUW/MIN interact preferentially  
43 with eukaryotic plasma membranes.

44  
45 **Importance:** Herein, we aimed to decipher the most important biosynthetic traits of a  
46 prominent group of bioactive lipopeptides. We reveal evidence for initiation of biosynthesis  
47 by two alternative starter units hardwired directly in the same gene cluster, eventually  
48 resulting in the production of a remarkable range of lipopeptide variants. We identified  
49 several unusual tailoring genes potentially involved in modifying the fatty acid chain. Careful  
50 characterization of these biosynthetic gene clusters and their diverse products could provide

51 important insight into lipopeptide biosynthesis in prokaryotes. Some of the identified variants  
52 exhibit cytotoxic and antifungal properties, and some are associated with a toxigenic biofilm-  
53 forming strain. The findings may prove valuable to researchers in the fields of natural product  
54 discovery and toxicology.

55

## 56 **Introduction**

57 Bacterial lipopeptides are a prominent group of secondary metabolites with  
58 pharmaceutical potential as antibacterial, antifungal, anticancer, and antiviral agents (1).  
59 Compounds such as fengycin, the iturin family antibiotics, octapeptins, and daptomycin are  
60 important pharmaceutical leads, the latter of which is already in clinical use (1–3). Their  
61 biological activity is the result of an amphipathic molecular structure that allows micellar  
62 interaction within the cell membranes of target organisms (4).

63 Lipopeptides are widespread in cyanobacteria and possess cytotoxic and antifungal  
64 activities (5-8). Puwainaphycins (PUWs) and minutissamides (MINs) are lipopeptides  
65 featuring a  $\beta$ -amino fatty acid and a 10-membered peptide ring (5, 9–11). Both classes exhibit  
66 considerable structural variability in terms of length and functionalization of the fatty acyl  
67 (FA) side chain attached to the stable peptide core (10–14). Only minor discrepancies in  
68 length and substitution of the FA chain separate these two types of lipopeptides. A wide array  
69 of bioactivities has been reported for these compounds. PUW C is a cardioactive compound  
70 (15) as demonstrated by positive inotropic activity in mouse atria, while PUW F/G exhibit  
71 cytotoxicity against human cells *in vitro* through cell membrane permeabilization (5). MINs  
72 A–L exhibited antiproliferative effects when tested against human cancer cell lines over a  
73 concentration range similar to PUWs (10, 11). The overall structural similarity suggests that  
74 PUWs and MINs share a similar biosynthetic origin. However, the biosynthetic mechanisms  
75 generating the conspicuous chemical variability remain unknown.

76 PUWs are synthesized by a hybrid polyketide/non-ribosomal peptide synthetase  
77 (PKS/NRPS) accompanied by tailoring enzymes (12). A characteristic feature of the PUW  
78 synthetase is the fatty acyl-AMP ligase (FAAL) starter unit (12). This enzyme specifically  
79 binds and adenylates FAs, and passes the activated acyl-adenylate to a downstream  
80 phosphopantetheine arm of the PKS acyl carrier protein (ACP) for further processing (12).  
81 The whole process bears resemblance to the biosynthesis of iturin-family lipopeptides (16–19)  
82 as well as small lipopeptide-like cyanobacterial metabolites such as hectochlorin (20),  
83 hapalosin (21), and jamaicamide (22), as discussed previously (23). Bacterial FAAL enzymes  
84 originate from basal cell metabolism, and likely evolved from fatty acyl-CoA ligases (FACLs)  
85 following a specific insertion that hampered subsequent ligation to CoASH (24) or altered the  
86 catalytic conformation (25). FAAL enzymes play an important role in the assembly of other  
87 metabolites including olefins (26) and unusual lipids (27) in addition to lipopeptide synthesis.  
88 The exact substrate-binding mechanism employed by FAALs was demonstrated  
89 experimentally in *Mycobacterium tuberculosis* using several homologous FAAL enzymes and  
90 FA substrates as models (28). The substrate specificity of these enzymes corresponds to the  
91 structure of the substrate-binding pocket (25, 28), although it overlaps among homologs.

92 Herein, we combined recently developed bioinformatics and high performance liquid  
93 chromatography combined with high resolution tandem mass spectrometry (HPLC-  
94 HRMS/MS) approaches (13, 23) to identify biosynthesis gene clusters for PUWs/MINs in  
95 five new cyanobacterial strains, and characterized the chemical variability of their products.  
96 We discuss the specific structural properties of the identified lipopeptide variants, and  
97 compare the predicted functions of synthetase enzymes.

98

## 99 **Results and Discussion**

### 100 *Structural variability vs. common biosynthetic origin of PUWs and MINs*

101 In the present study, we collected all known PUW/MIN producers (except for  
102 *Anabaena* sp. UIC10035). The strains were originally isolated from various soil habitats  
103 (Table 1). HPLC-HRMS/MS analysis detected multiple PUW and MIN variants in each of the  
104 strains studied (Fig. 1), ranging from 13 to 26 in strains 3 and 1, respectively (Table S1).

105 The MS/MS data acquired for crude extracts were used to create a molecular network  
106 (Fig. 2), analysis of which demonstrated that *Cylindrospermum* strains 1–3 and *Anabaena*  
107 strains 4 and 5 formed a single group with MIN A as the only variant common to all the  
108 strains (Fig. 2a). All major structural variants of these strains shared the common peptide  
109 sequence FA<sup>1</sup>-Val<sup>2</sup>-Dhb<sup>3</sup>-Asn<sup>4</sup>-Dhb<sup>5</sup>-Asn<sup>6</sup>-Ala<sup>7</sup>-Thr<sup>8</sup>-NMeAsn<sup>9</sup>-Pro<sup>10</sup> (Fig. 3), described  
110 previously for PUW F and MIN A (5, 10). The pattern of variant production was almost  
111 identical in *Cylindrospermum* strains 2 and 3, which in addition to MIN A contained PUW F  
112 (Fig. 1, Table S1). By contrast, *Anabaena* strains 4 and 5 produced MIN C and D in addition  
113 to the major variant MIN A (Fig. 1). The peptide core of the molecule was different in  
114 *Symplocastrum muelleri* strain 6 (Fig. 3), forming a separate group in the molecular network  
115 (Fig. 2b), with the general peptide sequence FA<sup>1</sup>-Val<sup>2</sup>-Dhb<sup>3</sup>-Thr<sup>4</sup>-Thr/Val<sup>5</sup>-Gln<sup>6</sup>-Ala<sup>7</sup>-OMe-  
116 Thr<sup>8</sup>-NMeAsn<sup>9</sup>-Pro<sup>10</sup> (Fig. 3), identical to PUW A–D and MIN I, K, L isolated previously  
117 from *Anabaena* sp. (9, 11).

118 The peptide core of the variants included in the network differed to some degree, but  
119 most variation was detected in the FA moiety (Fig. 4) when crude extracts were analyzed for  
120 the presence of characteristic FA immonium fragments (13).

121 Accordingly, bioinformatic analysis identified putative PUW and MIN gene clusters in  
122 each of the five newly sequenced strains (Fig. 5, Table 2). Based on BLASTp, CDD, and  
123 AntiSMASH searches, these gene clusters exhibited synteny and functional homology with  
124 the previously characterized *puw* biosynthesis gene cluster in strain 1 (12) (Fig. 5). Therefore,

125 our results strongly indicate a common biosynthetic origin of PUWs and MINs in  
126 cyanobacteria.

127

### 128 ***Variability in the peptide core***

129 A common set of NRPS genes (*puwA*, *puwE-H*; Fig. 5) encoding a sequence of nine  
130 amino acid-incorporating modules (Fig. 6) was detected in all analyzed strains. Individual  
131 NRPS modules displayed variability in amino acid adenylation and tailoring domains that was  
132 generally congruent with the PUW/MIN peptide cores inferred using HPLC-HRMS/MS (Fig.  
133 3). The two major observed types of peptide cores (represented by PUW A and PUW F,  
134 respectively) differed in the amino acids at positions 4 (Thr→Asn), 5 (Thr→Dhb), 6  
135 (Gln→Asn), 7 (Ala→Gly) and 8 (Thr→*OMe*-Thr), as shown in Fig. 3 and Table S1. This was  
136 reflected in the predicted substrates of the corresponding A-domains, and by the presence of  
137 an *O*-methyltransferase domain in *PuwH* of *S. muelleri* strain 6, which is responsible for the  
138 methoxylation of Thr<sup>8</sup> (Fig. 6, Table S2). In contrast to the variability observed at the  
139 previously noted amino acid positions, the two positions adjacent to both sides to the  
140 modified fatty acid (*NMeAsn*<sup>9</sup>-*Pro*<sup>10</sup>-(*FA*<sup>1</sup>)-*Val*<sup>2</sup>-*Dhb*<sup>3</sup>) are conserved in all known PUW/MIN  
141 variants described here and previously (5, 9, 13–15) (Fig. 3, Table S1). Accordingly, no  
142 functional variation in A-domains corresponding to these positions was observed within the  
143 deduced *PuwA*, *PuwE*, and *PuwF* proteins (Table S2). This is interesting because these four  
144 hydrophobic amino acids surround the FA moiety, which is likely responsible for the  
145 membrane disruption effect suggested previously (5). Thus, we hypothesize that such an  
146 arrangement could further support hydrophobic interactions with the lipid layer of the plasma  
147 membrane.

148 For some of the other positions, minor variants were observed involving substitution  
149 of amino acids similar in structure and hydrophobicity, including Asn-Gln at position 4, Thr-

150 Val at position 5, Ala-Gly at position 7, and Thr-Ser at position 8 (Fig. 3, Table S1),  
151 indicative of probable substrate promiscuity in their respective adenylation domains (29). The  
152 A6-domains in strains 4 and 5 activated Ala as a major substrate, and Gly to a lesser extent,  
153 even though *in silico* analysis predicted Gly as their main substrate (Table S2). In strain 6,  
154 Gly was incorporated, in agreement with the predicted substrate specificity. An epimerase  
155 domain was present in each of the sixth NRPS modules of the pathways (Fig. 6), indicating  
156 probable formation of a D-amino acid enantiomer at position 7 of the peptide core. Indeed, the  
157 presence of D-Ala was previously confirmed in PUW F (5) and MIN A-H (10, 11), and D-Gly  
158 was identified in MIN I-L (10, 11). In two cases, the adenylation domains A3 (PuwF) and A6  
159 (PuwG) are capable of incorporating significantly different amino acids such as Asn<sup>4</sup>-Thr<sup>4</sup>  
160 and Ala<sup>7</sup>-Ser<sup>7</sup>, respectively (Fig. 6). This degree of substrate promiscuity is relatively  
161 uncommon. Activation of two divergent amino acids (Arg/Tyr) by a single adenylation  
162 domain, based on point mutations in just three codons, was previously demonstrated in the  
163 anabaenopeptin synthetase from the cyanobacterium *Planktothrix agardhii* (30). The substrate  
164 exchange of Ala vs. Ser was previously reported from fungal class IV adenylation-forming  
165 reductases that contain A-domains homologous to NRPS enzymes (31).

166 The last synthetase enzyme in the pathway (PuwA) is equipped with a terminal  
167 thioesterase domain (Fig. 6), which presumably catalyzes cleavage of the final product and  
168 formation of the cycle via a peptide bond between the terminal prolyl and the  $\beta$ -amino group  
169 of the FA chain, as previously suggested (12).

#### 170 ***Two hypothetical starter units and their substrate range***

171 The biosynthesis of bacterial lipopeptides is typically commenced by FA-activating  
172 enzymes (16, 18). Initiation of the biosynthesis of PUW/MIN is performed by a FAAL enzyme  
173 (12) and allows a much broader array of activated substrates than the relatively conserved  
174 oligopeptide core (13) (Fig. 4). We identified three alternative arrangements of the putative



175 FAAL starter units (Fig. 5 and 6), each corresponding to a different array of FA side chains  
176 detected by HPLC-HRMS/MS, which presumably reflects the range of FA substrates  
177 activated during their biosynthesis (Fig. 4). *Cylindrospermum* sp. strains 1–3 possess the Type  
178 I putative starter unit consisting of a standalone FAAL enzyme PuwC and a separate ACP  
179 PuwD (Fig. 5, Table 2). By contrast, the biosynthetic gene cluster of *S. muelleri* strain 6  
180 contains the Type II putative starter unit (PuwI) consisting of a FAAL fused to an ACP (Fig.  
181 5, Table 2). *Anabaena* spp. strains 4 and 5 combine both Type I and Type II putative starter  
182 units in their biosynthetic gene clusters (Fig. 5, Table 2). Although the functions and substrate  
183 ranges of these hypothetical starter units requires further confirmation by gene manipulation  
184 experiments, they are supported by the patterns of lipopeptide variants detected by HPLC-  
185 HRMS/MS (Fig. 4, Table S1). In *Cylindrospermum* strains 1–3 that exclusively contain the  
186 Type I starter unit, the PUW/MIN products exhibited an almost continuous FA distribution  
187 between C<sub>10</sub>–C<sub>15</sub> (up to C<sub>17</sub> in negligible trace amounts; Fig. 4). In *S. muelleri* strain 6, the  
188 presence of the Type II loading module resulted in production of PUW/MIN variants with  
189 discrete FA lengths of C<sub>16</sub> and C<sub>18</sub>. Strains containing both Type I and Type II starter units  
190 (*Anabaena* strains 4 and 5) produced two sets of PUW/MIN products with no overlap  
191 (C<sub>12</sub>–C<sub>14,15</sub> for the Type I pathway, and C<sub>16</sub> for the Type II pathway), but exhibited a slightly  
192 shifted length distribution (Fig. 4). Based on these results, it seems plausible that PuwC/D and  
193 PuwI represent two alternative FAAL starter modules capable of initiating PUW/MIN  
194 biosynthesis (Figs. 5, 6). An analogous situation was previously described for the alternative  
195 NRPS starter modules in the anabaenopeptin synthetase (32).

196 In the FA residue of the lipopeptide, proximal carbons in the linear aliphatic chain are  
197 incorporated into the nascent product by PKS enzymes (12). The PKS domains of PuwB and  
198 PuwE (Fig. 6, Table 2) catalyze two elongation steps. Therefore, the fatty acid is expected to  
199 be extended by four carbons.. The substrate length specificity of the FAAL enzymes in

200 *Mycobacterium tuberculosis* was recently shown to be determined by the size and position of  
201 specific amino acid residues protruding into the FA-binding pocket (28). Experimental  
202 replacement of Gly or Ile by a larger Trp residue in the upper and middle parts of the pocket  
203 blocked the binding of the original C<sub>12</sub> substrate, but shorter chains (C<sub>2</sub> or C<sub>10</sub>, respectively)  
204 were still activated (28). Experimental data on FAAL substrate specificity in cyanobacteria  
205 are currently lacking. Alignment of amino acid residues from all putative PuwC and PuwI  
206 proteins demonstrates overall homology (Fig. S2a), including the positions corresponding to  
207 the FA-binding pocket, as previously shown in *Mycobacterium* (28) (Fig. S2b). Experimental  
208 evidence such as *in vitro* activity assays and crystallization of protein-ligand complexes is  
209 required to explain the variable substrate specificity of PuwC vs. PuwI. Also, we cannot  
210 exclude the possibility that the FA substrate length range is partially determined by the pool  
211 of free FAs available to the FAAL enzyme. Indeed, this possibility is supported by  
212 observations of *Cylindrospermum* strains 1–3, which share highly conserved PuwC proteins  
213 (Fig. S2a) with identical residues in the predicted FA-binding pockets (Fig. S2b), but display  
214 slightly different ranges and ratios of incorporated FAs in the PUW/MIN variants produced  
215 (Fig. 1, 4).

#### 216 ***FA tailoring reactions: oxidation, halogenation, and acetylation***

217 Intriguingly, all products originating from biosynthesis initiated by the Type II starter  
218 unit (variants with a C<sub>16</sub> and C<sub>18</sub> FA tail in *Anabaena* strains 4–5 and *S. muelleri* strain 6)  
219 include substitution of a hydroxy- or oxo-moiety (Fig. 6). For minutissamides C and D, this  
220 substitution takes place on the third carbon from the FA terminus (C<sub>14</sub>), as described  
221 previously (10), and this position was confirmed by NMR in variants produced by *Anabaena*  
222 *sp.* strain 4 in our study (Table S3, Figs. S3–6). In agreement with this hydroxy- and oxo-  
223 substitution, the respective gene clusters each encode PuwJ, a putative cytochrome P450-like  
224 oxidase (Table 2), immediately downstream of the gene encoding the Type II starter module.

225 We therefore hypothesize that the PuwJ enzyme is responsible for hydroxylation of FA  
226 residues activated by PuwI (Fig. 6). However, the formation of the keto variant remains  
227 unexplained by our data.

228 Another gene, the putative halogenase *puwK*, was associated with the Type II starter  
229 module in *Anabaena* sp. strain 5 (Table 2). Although no conserved enzymatic domain was  
230 detected in the deduced protein, it shares similarity with proteins postulated to be involved in  
231 halogenation of cyanobacterial chlorinated acyl amides known as columbamides (33), and *N*-  
232 oxygenases similar to *p*-aminobenzoate *N*-oxygenase AurF (34–36). The possible functional  
233 designation of this enzyme as a halogenase is further supported by the fact that the  $\omega$ -  
234 chlorinated product MIN B, originally described in strain *Anabaena* sp. strain 5 (10), was also  
235 detected in this study (Table S1) as one of the major variants, while no MIN B or any other  
236 chlorinated PUW/MIN products were detected in *Anabaena* sp. strain 4 (Fig. 1). *Anabaena* sp.  
237 strains 4 and 5 share identical organization across the entire gene cluster, and lack of the  
238 putative halogenase gene *puwK* is the only difference between these two clusters in terms of  
239 presence of genes (Fig. 5).

240 In *Cylindrospermum* sp. strains 1–3 that exclusively possess the Type I starter unit, the  
241 presence of minor amounts of hydroxylated and chlorinated variants (Fig. 4) suggests the  
242 involvement of another biosynthetic mechanism unexplained by the current data. This  
243 ambiguity warrants experimental research such as gene knock-out experiments to confirm the  
244 proposed functions of *puwJ* and *puwK*.

245 Finally, the gene cluster identified in *S. muelleri* strain 6 was the only one containing  
246 gene *puwL*. The deduced product of this gene shares 53.4% similarity with the *O*-  
247 acetyltransferase McyL (Table 2) involved in acetylation of the aliphatic chain of microcystin  
248 in cyanobacteria (37). Additionally, this gene is similar to chloramphenicol and streptogramin  
249 A *O*-acetyltransferases that serve as antibiotic resistance agents in various bacteria (38). The

250 functional annotation of PuwL as a putative *O*-acetyltransferase is consistent with the  
251 detection of *O*-acetylated lipopeptide variants in *S. muelleri* strain 6 (Table 3, Fig. 7). Five  
252 PUW variants ( $m/z$  1265.7338, 1279.7496, 1277.7695, 1291.7870 and 1293.7654) yielding  
253 high-energy fragments, proving the presence of an acetyl group bonded to the FA moiety,  
254 have been detected. In the  $m/z$  1279.8 and 1293.8 peaks, the high-energy fragment ion at  $m/z$   
255 312 corresponds to the FA immonium ion bearing an acetyl group, and fragment ion at  $m/z$   
256 439 corresponds to the prolyl-FA-acetyl fragment. The subsequent loss of an acetyl group  
257 resulted in the presence of ions at  $m/z$  252 and 379, respectively (Table 3, Fig. 7). Similarly,  
258 analysis of the  $m/z$  1265.7 peak revealed analogous fragments at  $m/z$  284/411 and 351/224  
259 (Table 3).

#### 260 ***Antimicrobial activity***

261 Both PUWs and MINs possess cytotoxic activity against human cells *in vitro* (5, 10,  
262 11). In the current study, we demonstrated that the major PUW/MIN variants (PUW F and  
263 MINs A, C, and D) did not exert antibacterial effects against either Gram-positive or Gram-  
264 negative bacteria using a panel of 13 selected strains (Table 4). PUW F was the only tested  
265 variant manifesting antagonistic activity against two yeast strains utilized in our experiment,  
266 namely *Candida albicans* HAMBI 261 and *Saccharomyces cerevisiae* HAMBI 1164, with  
267 inhibition zones of 14 and 18 mm, respectively, and minimum inhibitory concentration (MIC)  
268 values of  $6.3 \mu\text{g mL}^{-1}$  ( $5.5 \mu\text{M}$ ; Fig. 8). No antifungal activity was recorded for the MIN C  
269 and D variants, and only weak inhibition of the two yeast strains was recorded for MIN A  
270 (Fig. S7). PUW F differs only slightly from MIN A by a  $-\text{CH}_2-\text{CH}_3$  extension of the FA  
271 moiety, indicating that the FA length affects bioactivity. Furthermore, the lack of bioactivity  
272 for MIN C and MIN D suggests that hydroxy- and oxo- substitution also compromises  
273 antifungal efficacy. As previously demonstrated, cytotoxicity is due to membrane  
274 permeabilization activity accompanied by calcium flux into the cytoplasm (5), consistent with

275 the membrane effects documented for other bacterial lipopeptides (4). However, as apparent  
276 from our data (Fig. 8), PUW/MIN products appear to be effective solely against eukaryotes  
277 (thus far tested only on human and yeast cells). This finding is in contrast to the typical  
278 antibacterial activity frequently described for many lipopeptides produced by Gram-positive  
279 bacteria (4). Analogously, the cyanobacterial lipopeptides anabanenolysin A and hassalidins  
280 preferentially interact with cholesterol-containing membranes, hence their predisposition for  
281 activity against eukaryotic cells (6, 8).

### 282 ***Distribution of PUWs and MINs in cyanobacteria***

283 PUWs and MINs form one of the most frequently reported groups of lipopeptides in  
284 cyanobacteria, and have been isolated from heterocytous cyanobacteria, particularly members  
285 of the genera *Anabaena* and *Cylindrospermum* that inhabit soil (5, 9–11). Only a single study  
286 has mentioned the probable occurrence of puwainaphycins in a planktonic cyanobacterium  
287 (*Sphaerospermopsis*) (39). Our current comprehensive analysis of these lipopeptides and their  
288 biosynthetic genes further supports the hypothesis that lipopeptides occur predominantly in  
289 non-planktic biofilm-forming cyanobacteria (23). In this context, it is worth mentioning that  
290 *S. muelleri* strain 6 was isolated from a wetland bog in alpine mountains in coastal Norway  
291 (40). This strain is a toxigenic member of a biofilm microbiome, and suspected to play a role  
292 in the development of severe hemolytic Alveld disease among outfield grazing sheep (41, 42).  
293 Biomass harvested from pure cultures of this strain exhibited strong cytotoxic activity toward  
294 primary rat hepatocytes (43, 44), which indicates the production of secondary metabolites  
295 with cytotoxic properties. Thus, the possible toxic potential of cyanobacterial lipopeptides  
296 such as PUWs and MINs in the environment warrants further attention.

### 297 ***Conclusions***

298 Our study highlights and explores the extensive structural versatility of cyanobacterial  
299 lipopeptides from the PUW/MIN family by introducing previously unknown variants and

300 newly sequenced biosynthetic gene clusters. Intriguingly, all variants are synthesized by a  
301 relatively conserved PKS/NRPS machinery with a common genetic origin. We hypothesize  
302 that chemical diversity is generated largely by the presence of two alternative fatty acyl-AMP  
303 ligase starter units, one of which exhibits an unusually broad specificity for FA substrates of  
304 variable length. Additionally, putative halogenase and *O*-acetyltransferase genes were present  
305 in some gene clusters. This knowledge provides novel insight into the genetic background  
306 underpinning the biosynthesis of bacterial lipopeptides. The proposed biosynthetic  
307 mechanisms allow the studied microbes to generate a large pool of products that can be  
308 readily expanded by introducing relatively small genetic changes. This is consistent with the  
309 so-called ‘Screening Hypothesis’ (45, 46), which predicts an evolutionary benefit for  
310 organisms producing a large chemical diversity of secondary metabolites at minimum cost.  
311 Accessory antimicrobial tests on bacteria and yeasts, together with previously published  
312 results, suggest a specific toxic effect of PUWs against eukaryotic cells. Thus, their toxic  
313 potential for humans and other animals clearly warrants further investigation, and their  
314 possible use as antifungal agents is ripe for exploration.

315

## 316 **Materials and Methods**

### 317 *Cultivation of cyanobacterial strains*

318 Six cyanobacterial strains were included in the present study: *Cylindrospermum moravicum*  
319 CCALA 993 (strain 1), *Cylindrospermum alatosporum* CCALA 994 (strain 2),  
320 *Cylindrospermum alatosporum* CCALA 988 (strain 3), *Anabaena sp.* UHCC-0399  
321 (previously *Anabaena sp.* SMIX 1; strain 4), *Anabaena minutissima* UTEX B 1613 (strain 5),  
322 and *Symplocastrum muelleri* NIVA-CYA 644 (strain 6). The origins of the strains are listed in  
323 Table 1. For chemical analysis, strains 1–5 were cultivated in BG-11 media (47) in glass  
324 columns (300 mL) bubbled with air enriched in 1.5% CO<sub>2</sub> at a temperature of 28°C and

325 constant illumination of  $100 \mu\text{mol photons m}^{-2} \text{s}^{-1}$ . Strain 6 was maintained in culture using a  
326 custom liquid medium obtained by mixing 200 mL of Z8 medium (48), 800 mL distilled  
327 water, 30 mL soil extract, and common vitamin pre-mix (according to SAG – Sammlung von  
328 Algenkulturen der Universität Göttingen, but without biotin). Cultivation was performed in  
329 100–200 mL Erlenmeyer flasks at  $20^\circ\text{C}$  with a 16:8 light:dark photoperiod under static  
330 conditions. Cultures were kept at low irradiance ( $4 \mu\text{mol m}^{-2} \text{s}^{-1}$  PHAR generated using RGB  
331 LED strips). Cells were harvested by centrifugation ( $3125 \times g$ ), stored at  $-80^\circ\text{C}$ , and  
332 subsequently lyophilized. Strain 4 was cultivated at a larger scale for purification of major  
333 lipopeptide variants in a 10 L tubular photobioreactor under the above-mentioned conditions  
334 in BG-11 medium.

### 335 ***Molecular and bioinformatic analyses***

336 Single filaments of strains 2, 3, 5, and 6 were isolated for whole-genome amplification  
337 (WGA) and subsequent preparation of a whole-genome sequencing (WGS) library, as  
338 described previously (12). Briefly, the glass capillary technique was used to isolate filaments  
339 excluding minor bacterial contaminants. A set of 20 filaments from each strain was then used  
340 as a template for WGA. Multiple displacement amplification (MDA) using a Repli-g Mini Kit  
341 (Qiagen, Hilden, Germany) was followed by PCR and sequencing to monitor the  
342 cyanobacterial 16S rRNA gene using primers 16S387F and 16S1494R (49). Positive samples  
343 (7–10 MDA products yielding clear 16S rRNA gene sequences of the respective genera) were  
344 then pooled to create a template for WGS. DNA samples were sent for commercial *de novo*  
345 genome sequencing (EMBL Genomics Core Facility, Heidelberg, Germany) using the  
346 Illumina MiSeq platform (Illumina, San Diego, CA, USA) with a ~350 bp average insert  
347 length Pair-End library and 250 bp reads (~1.4 Gbp data yield per strain). Raw data from *de*  
348 *novo* WGS were assembled using CLC Bio Genomics Workbench v. 7.5 (CLC Bio, Aarhus,

349 Denmark). Genomic DNA was isolated from strain 4 as previously described (37) and the  
350 quality was assessed using a NanoDrop 1000 spectrophotometer (Thermo Fisher Scientific,  
351 Waltham, MA, USA) and an Agilent TapeStation (Agilent Technologies, Santa Clara, CA,  
352 USA). High-molecular-weight DNA was used to construct an Illumina TruSeq PCR Free 350  
353 bp library and sequenced using an Illumina HiSeq 2500 platform with a paired-end 100 cycles  
354 run. Genome data (1Gb for each strain) were first checked using SPAdes version 3.7.1 (51)  
355 for read correction and removal of erroneous reads, and then assembled using Newbler  
356 version 3.0 (454 Life Sciences, Branford, CT, USA). Genomic scaffolds were loaded into  
357 Geneious Pro R10 (Biomatters; available from <http://www.geneious.com>) and investigated for  
358 FAAL and NRPS genes using BLASTp searches to identify putative lipopeptide synthetase  
359 gene clusters (23). FAAL and NRPS adenylation domains (A-domains) from the single  
360 known PUW gene cluster (strain 1; KM078884) were used as queries since homologous gene  
361 clusters were expected. Contigs yielding high similarity hits (E-value  $<10^{-20}$ ) were then  
362 analyzed using the Glimmer 3 (50) algorithm to discover putative open-reading frames  
363 (ORFs). Functional annotation of ORFs was conducted by applying a combination of  
364 BLASTp/CDD searches against the NCBI database, and using the antiSMASH 4.0 secondary  
365 metabolite gene cluster annotation pipeline (52, 53). Pairwise sequence identities and the  
366 presence of conserved residues in homologous putative proteins encoded in the gene clusters  
367 were assessed using Geneious Pro software based on amino acid alignment (MAFFT plugin,  
368 default parameters). Minor assembly gaps were identified in the genomic scaffolds of all  
369 investigated strains, either directly after pair-end read assembly, or based on mapping to the  
370 reference gene cluster from *C. alatosporum* CCALA 988. Gaps in PUW/MIN gene clusters  
371 were closed by PCR, and subsequent Sanger sequencing of PCR products was performed  
372 using custom primer annealing to regions adjacent to the assembly gaps.

373 ***Extraction and analysis of PUWs/MINs***



374 To obtain comparable results, each strain was extracted using an identical ratio of  
375 lyophilized biomass (200 mg) to extraction solvent (10 mL of 70% MeOH, v/v). Extracts  
376 were evaporated using a rotary vacuum evaporator at 35°C and concentrated to 1 mL of 70%  
377 MeOH. The methanolic extracts were analyzed using a Thermo Scientific Dionex UltiMate  
378 3000 UHPLC+ instrument equipped with a diode-array detector connected to a Bruker Impact  
379 HD (Bruker, Billerica, MA, USA) high-resolution mass spectrometer with electrospray  
380 ionization. Separation of extracts was performed on a reversed-phase Phenomenex Kinetex  
381 C18 column (150 × 4.6 mm, 2.6 μm) using H<sub>2</sub>O (A)/acetonitrile (B) containing 0.1%  
382 HCOOH as a mobile phase, at a flow rate of 0.6 mL min<sup>-1</sup>. The gradient was as follows: A/B  
383 85/15 (0 min), 85/15 (over 1 min), 0/100 (over 20 min), 0/100 (over 25 min), and 85/15 (over  
384 30 min). For better resolution of minor PUW variants, another analytical method with a  
385 longer gradient (67 min) adopted from our previous study (13) was applied. The peptide  
386 sequence was reconstructed from the *b* ion series obtained after opening of the ring between  
387 the proline and *N*-methylassparagine residues, followed by the sequential loss of water and all  
388 the amino acids with exception of the last residue (Pro). The number of carbons in the FA  
389 moiety in PUW/MIN variants containing nonsubstituted and hydroxy-/chloro-substituted FA  
390 was determined using a method described previously by our team (13). Characteristic FA  
391 immonium fragments in oxo-substituted PUW/MIN variants were identified by employing  
392 this method to crude extracts of *Anabaena* strain 5 containing the oxo-substituted MIN D  
393 variant (10). Since a stable, prominent, and characteristic FA immonium fragment with the  
394 sum formula C<sub>15</sub>H<sub>30</sub>NO<sup>+</sup> was obtained for MIN D (Fig. S1), analogous fragments with general  
395 formula C<sub>x</sub>H<sub>2x</sub>NO<sup>+</sup> were used to identify oxo-substituted components in unknown PUW/MIN  
396 variants from other investigated strains.

397 ***Molecular networking***

398 A molecular network was created using the Global Natural Products Social Molecular  
399 Networking (GNPS) online workflow (54). Data were filtered by removing all MS/MS peaks  
400 within +/- 17 Da of the precursor  $m/z$ . MS/MS spectra were window-filtered by choosing only  
401 the top six peaks in the +/- 50 Da window throughout the spectrum. Data were then clustered  
402 with MS-Cluster with a parent mass tolerance of 0.1 Da and a MS/MS fragment ion tolerance  
403 of 0.025 Da to create consensus spectra. Additionally, consensus spectra comprised of fewer  
404 than two spectra were discarded. A network was then created in which edges were filtered  
405 using a cosine score above 0.75 and more than three matched peaks. Additional edges  
406 between pairs of nodes were retained in the network only when both nodes were included in  
407 each other's respective top 10 most similar nodes. Spectra in the network were then searched  
408 against the GNPS spectral libraries, and library spectra were filtered in the same manner as  
409 the input data. All matches obtained between network spectra and library spectra were  
410 retained only when the score was above 0.7 and at least four peaks matched. Analog searching  
411 was performed against the library with a maximum mass shift of 200 Da.

#### 412 ***Purification of MINs from Anabaena sp. strain 4 and its NMR analysis***

413 Freeze-dried biomass of strain 4 (10 g) was extracted with 70% MeOH (500 mL). The extract  
414 was evaporated using a rotary vacuum evaporator to reduce the MeOH content, and the  
415 sample was subsequently diluted with distilled water to reach a final MeOH concentration  
416 >5%. The diluted extract was pre-purified using a Supelco C18 SPE cartridge (10 g, 60 mL)  
417 pre-equilibrated with 60 mL of MeOH and 120 mL of H<sub>2</sub>O. After loading, retained  
418 components were eluted with 60 mL of pure MeOH, concentrated to dryness, and  
419 resuspended in 10 mL of pure MeOH. MINs A, C, and D were purified in two HPLC  
420 purification steps. The first step was performed on a preparative chromatographic system  
421 (Agilent 1260 Infinity series) equipped with a multi-wavelength detector and automatic  
422 fraction collector. A preparative Reprosil 100 C18 column (252 × 25 mm) was employed for

423 separation at a flow rate of 10 mL min<sup>-1</sup> using the following gradient of MeOH containing  
424 0.1% HCOOH (A) and 10% MeOH containing 0.1% HCOOH (B): 0 min (100% B), 6 min  
425 (100% B), 15 min (43% B), 43 min (12% B), 45 min (0% B), 58 min (0% B), 60 min (100%  
426 B), and 64 min (100% B). Fractions were collected using an automatic fraction collector at 1  
427 min intervals, and fractions were analyzed for MIN A, C, and D using the method described  
428 above. Fractions containing MIN A, C, and D were collected in separate vials and  
429 concentrated using a rotary evaporator. The second purification step was performed on a semi-  
430 preparative HPLC (Agilent 1100 Infinity series) using a Reprosil 100 Phenyl column (250 × 8  
431 mm) with (A) acetonitrile containing 0.1% HCOOH and (B) water containing 0.1% HCOOH  
432 using the following gradient: 0 min (60% B), 2 min (60% B), 6 min (50% B), 28 min (18%  
433 B), 30 min (0% B), 30 min (0% B), 32 min (0% B), 31 min (60% B), and 36 min (60% B).  
434 The flow rate was 1 mL min<sup>-1</sup> throughout, fractions were collected manually, and the purity  
435 was analyzed using the HPLC-HRMS method described above. NMR spectra of  
436 minutissamides were measured in dimethyl sulfoxide (DMSO)-*d*<sub>6</sub> at 30°C. All NMR spectra  
437 were collected using a Bruker Avance III 500 MHz NMR spectrometer, equipped with a 5  
438 mm Ø BBI probehead with actively shielded z-gradient.

439

#### 440 ***Antibacterial and antifungal assays***

441 The antimicrobial activity of four major variants (PUW F, and MINs A, C, and D) was  
442 tested against 13 bacterial and two yeast strains (Table 4) using disc diffusion assays (8) in  
443 three independent experiments with kanamycin/nystatin and MeOH as positive and negative  
444 controls, respectively. Antifungal activity of PUW F was further evaluated by determining the  
445 MIC against *Candida albicans* (HAMBI 261) and *Saccharomyces cerevisiae* (HAMBI 1164)  
446 as described previously (8). PUW F was isolated from *Cylindrospermum* strain 1 according to

447 a protocol described previously (5), and isolation of MIN A, C, and D was performed as  
448 described above. The variants produced by *S. muelleri* strain 6 were impossible to isolate due  
449 to the slow growth of the cyanobacterium, resulting in low biomass yields during the study  
450 period.

451 **Accession numbers** for the newly sequenced complete putative biosynthetic gene  
452 clusters uploaded to the NCBI GenBank database are MH325197-MH325201.

453

#### 454 **Acknowledgements**

455 This work was supported by Czech Science Foundation grant no. 16-09381S (Bioactive  
456 cyanobacterial lipopeptides: genome mining, detection, and structure-activity relationships),  
457 and by the Ministry of Education, Youth and Sports of the Czech Republic, National  
458 Programme of Sustainability I, ID: LO1416 project ALGAMIC (ID: CZ.1.05/2.1.00/19.0392)  
459 and MSCA IF II project (CZ.02.2.69/0.0/0.0/18\_070/0010493). Access to instruments and  
460 other facilities was supported by the Czech Research Infrastructure for Systems Biology  
461 (C4SYS; project no. LM2015055). This research was also supported by a grant from the  
462 NordForsk NCoE programme “NordAqua” (project no. 82845). The Norwegian participation  
463 was supported by grants from the Department of Agriculture and Forestry, the County  
464 Governor of Møre og Romsdal, the County Governor of Sogn og Fjordane, the University of  
465 Oslo, and the Norwegian Institute for Water Research. The authors declare no conflict of  
466 interest. The funders had no role in study design, data collection and interpretation, or the  
467 decision to submit the work for publication.

468

#### 469 **References**

- 470 1. Cochrane SA, Vederas JC. 2016. Lipopeptides from *Bacillus* and *Paenibacillus* spp.:  
471 A gold mine of antibiotic candidates. *Med Res Rev* 36:4–31.
- 472 2. Taylor SD, Palmer M. 2016. The action mechanism of daptomycin. *Bioorg Med Chem*  
473 24:6253–6268.
- 474 3. Velkov T, Roberts KD, Li J. 2017. Rediscovering the octapeptins. *Nat Prod Rep*  
475 34:295–309.
- 476 4. Ines M, Dhouha G. 2015. Lipopeptide surfactants: Production, recovery and pore  
477 forming capacity. *Peptides* 71:100–112.
- 478 5. Hrouzek P, Kuzma M, Černý J, Novák P, Fišer R, Šimek P, Lukešová A, Kopecný J.  
479 2012. The cyanobacterial cyclic lipopeptides puwainaphycins F/G are inducing  
480 necrosis via cell membrane permeabilization and subsequent unusual actin  
481 relocalization. *Chem Res Toxicol* 25:1203–1211.
- 482 6. Oftedal L, Myhren L, Jokela J, Gausdal G, Sivonen K, Doskeland SO, Herfindal L.  
483 2012. The lipopeptide toxins anabaenolysin A and B target biological membranes  
484 in a cholesterol-dependent manner. *Biochim Biophys Acta-Biomembr* 1818:3000–  
485 3009.
- 486 7. Tomek P, Hrouzek P, Kuzma M, Sýkora J, Fišer R, Černý J, Novák P, Bártová S,  
487 Šimek P, Hof M, Kavan D, Kopecný J. 2015. Cytotoxic lipopeptide muscotoxin A,  
488 isolated from soil cyanobacterium *Desmonostoc muscorum*, permeabilizes  
489 phospholipid membranes by reducing their fluidity. *Chem Res Toxicol* 28:216–  
490 224.
- 491 8. Vestola J, Shishido TK, Jokela J, Fewer DP, Aitio O, Permi P, Wahlsten M, Wang H,  
492 Rouhiainen L, Sivonen K. 2014. Hassallidins, antifungal glycolipopeptides, are  
493 widespread among cyanobacteria and are the end-product of a nonribosomal  
494 pathway. *P Natl Acad Sci USA* 111:E1909–E1917.

- 495 9. Gregson JM, Chen JL, Patterson GML, Moore RE. 1992. Structures of  
496 puwainaphycins A-E. *Tetrahedron* 48:3727–3734.
- 497 10. Kang HS, Kronic A, Shen Q, Swanson SM, Orjala J. 2011. Minutissamides A-D,  
498 antiproliferative cyclic decapeptides from the cultured cyanobacterium *Anabaena*  
499 *minutissima*. *J Nat Prod* 74:1597–1605.
- 500 11. Kang HS, Sturdy M, Kronic A, Kim H, Shen Q, Swanson SM, Orjala J. 2012.  
501 Minutissamides E-L, antiproliferative cyclic lipodecapeptides from the cultured  
502 freshwater cyanobacterium cf. *Anabaena* sp. *Bioorg Med Chem* 20:6134–6143.
- 503 12. Mareš J, Hájek J, Urajová P, Kopecký J, Hrouzek P. 2014. A hybrid non-ribosomal  
504 peptide/polyketide synthetase containing fatty-acyl ligase (FAAL) synthesizes the  
505 beta-amino fatty acid lipopeptides puwainaphycins in the cyanobacterium  
506 *Cylindrospermum alatosporum*. *Plos One* 9: e111904.
- 507 13. Urajová P, Hájek J, Wahlsten M, Jokela J, Galica T, Fewer DP, Kust A, Zapomělová-  
508 Kozlíková E, Delawská K, Sivonen K, Kopecký J, Hrouzek P. 2016. A liquid  
509 chromatography-mass spectrometric method for the detection of cyclic beta-amino  
510 fatty acid lipopeptides. *J Chromatogr A* 1438:76–83.
- 511 14. Cheel J, Urajová P, Hajek J, Hrouzek P, Kuzma M, Bouju E, Faure K, Kopecký J.  
512 2017. Separation of cyclic lipopeptide puwainaphycins from cyanobacteria by  
513 countercurrent chromatography combined with polymeric resins and HPLC. *Anal*  
514 *Bioanal Chem* 409:917–930.
- 515 15. Moore RE, Bornemann V, Niemczura WP, Gregson JM, Chen JL, Norton TR,  
516 Patterson GML, Helms GL. 1989. Puwainaphycin C, a cardioactive cyclic peptide  
517 from the blue-green alga *Anabaena* BQ-16-1. Use of two dimensional carbon-13-  
518 carbon-13 and carbon-13-nitrogen-15 correlation spectroscopy in sequencing the  
519 amino acid units. *J Am Chem Soc* 111:6128–6132.

- 520 16. Duitman EH, Hamoen LW, Rembold M, Venema G, Seitz H, Saenger W, Bernhard F,  
521 Reinhardt R, Schmidt M, Ullrich C, Stein T, Leenders F, Vater J. 1999. The  
522 mycosubtilin synthetase of *Bacillus subtilis* ATCC6633: A multifunctional hybrid  
523 between a peptide synthetase, an amino transferase, and a fatty acid synthase. P  
524 Nat Acad Sci USA 96:13294–13299.
- 525 17. Tsuge K, Akiyama T, Shoda M. 2001. Cloning, sequencing, and characterization of  
526 the iturin A operon. J Bacteriol 183:6265–6273.
- 527 18. Koumoutsi A, Chen XH, Henne A, Liesegang H, Hitzeroth G, Franke P, Vater J,  
528 Borriss R. 2004. Structural and functional characterization of gene clusters  
529 directing nonribosomal synthesis of bioactive cyclic lipopeptides in *Bacillus*  
530 *amyloliquefaciens* strain FZB42. J Bacteriol 186:1084–1096.
- 531 19. Sood S, Steinmetz H, Beims H, Mohr KI, Stadler M, Djukic M, von der Ohe W,  
532 Steinert M, Daniel R, Muller R. 2014. Paenilarvins: Iturin family lipopeptides from  
533 the honey bee pathogen *Paenibacillus larvae*. ChemBioChem 15:1947–1955.
- 534 20. Ramaswamy AV, Sorrels CM, Gerwick WH. 2007. Cloning and biochemical  
535 characterization of the hectochlorin biosynthetic gene cluster from the marine  
536 cyanobacterium *Lyngbya majuscula*. J Nat Prod 70:1977–1986.
- 537 21. Micallef ML, D'Agostino PM, Sharma D, Viswanathan R, Moffitt MC. 2015. Genome  
538 mining for natural product biosynthetic gene clusters in the Subsection V  
539 cyanobacteria. BMC Genomics 16:669.
- 540 22. Edwards DJ, Marquez BL, Nogle LM, McPhail K, Goeger DE, Roberts MA, Gerwick  
541 WH. 2004. Structure and biosynthesis of the jamaicamides, new mixed polyketide-  
542 peptide neurotoxins from the marine cyanobacterium *Lyngbya majuscula*. Chem  
543 Biol 11:817–833.

- 544 23. Galica T, Hrouzek P, Mareš J. 2017. Genome mining reveals high incidence of  
545 putative lipopeptide biosynthesis NRPS/PKS clusters containing fatty acyl-AMP  
546 ligase genes in biofilm-forming cyanobacteria. *J Phycol* 53:985–998.
- 547 24. Arora P, Goyal A, Natarajan VT, Rajakumara E, Verma P, Gupta R, Yousuf M,  
548 Trivedi OA, Mohanty D, Tyagi A, Sankaranarayanan R, Gokhale RS. 2009.  
549 Mechanistic and functional insights into fatty acid activation in *Mycobacterium*  
550 *tuberculosis*. *Nat Chem Biol* 5:166–173.
- 551 25. Liu Z, Ioerger TR, Wang F, Sacchettini JC. 2013. Structures of *Mycobacterium*  
552 *tuberculosis* FadD10 protein reveal a new type of adenylate-forming enzyme. *J*  
553 *Biol Chem* 288:18473–18483.
- 554 26. Coates RC, Podell S, Korobeynikov A, Lapidus A, Pevzner P, Sherman DH, Allen EE,  
555 Gerwick L, Gerwick WH. 2014. Characterization of cyanobacterial hydrocarbon  
556 composition and distribution of biosynthetic pathways. *Plos One* 9:e85140.
- 557 27. Mohanty D, Sankaranarayanan R, Gokhale RS. 2011. Fatty acyl-AMP ligases and  
558 polyketide synthases are unique enzymes of lipid biosynthetic machinery in  
559 *Mycobacterium tuberculosis*. *Tuberculosis* 91:448–455.
- 560 28. Goyal A, Verma P, Anandhakrishnan M, Gokhale RS, Sankaranarayanan R. 2012.  
561 Molecular basis of the functional divergence of fatty acyl-AMP ligase biosynthetic  
562 enzymes of *Mycobacterium tuberculosis*. *J Mol Biol* 416:221–238.
- 563 29. Villiers BRM, Hollfelder F. 2009. Mapping the limits of substrate specificity of the  
564 adenylation domain of TycA. *ChemBioChem* 10:671–682.
- 565 30. Christiansen G, Philmus B, Hemscheidt T, Kurmayer R. 2011. Genetic variation of  
566 adenylation domains of the anabaenopeptin synthesis operon and evolution of  
567 substrate promiscuity. *J Bacteriol* 193:3822–3831.



- 568 31. Brandenburger E, Braga D, Kombrink A, Lackner G, Gressler J, Künzler M,  
569 Hoffmeister D. 2018. Multi-genome analysis identifies functional and  
570 phylogenetic diversity of basidiomycete adenylylase-forming reductases. *Fungal*  
571 *Genet Biol* 112:55–63.
- 572 32. Rouhiainen L, Jokela J, Fewer DP, Urmann M, Sivonen K. 2010 Two alternative  
573 starter modules for the non-ribosomal biosynthesis of specific anabaenopeptin  
574 variants in *Anabaena* (Cyanobacteria). *Chem Biol* 17:265–273.
- 575 33. Kleigrewe K, Almaliti J, Tian IY, Kinnel RB, Korobeynikov A, Monroe EA, Duggan  
576 BM, Di Marzo V, Sherman DH, Dorrestein PC, Gerwick L, Gerwick WH. 2015.  
577 Combining mass spectrometric metabolic profiling with genomic analysis: A  
578 powerful approach for discovering natural products from cyanobacteria. *J Nat Prod*  
579 78:1671–1682.
- 580 34. Voráčová K, Hájek J, Mareš J, Urajová P, Kuzma M, Cheel J, Villunger A, Kapuscik  
581 A, Bally M, Novák P, Kabeláč M, Krumschnabel G, Lukeš M, Voloshko L,  
582 Kopecký J, Hrouzek P. 2017. The cyanobacterial metabolite nocuolin A is a  
583 natural oxadiazine that triggers apoptosis in human cancer cells. *Plos One*  
584 12:e0172850.
- 585 35. He J, Hertweck C. 2004. Biosynthetic origin of the rare nitroaryl moiety of the  
586 polyketide antibiotic aureothin: Involvement of an unprecedented N-oxygenase. *J*  
587 *Am Chem Soc* 126:3694–3695.
- 588 36. Choi YS, Zhang HJ, Brunzelle JS, Nair SK, Zhao HM. 2008. In vitro reconstitution  
589 and crystal structure of p-aminobenzoate N-oxygenase (AurF) involved in  
590 aureothin biosynthesis. *P Natl Acad Sci USA* 105:6858–6863.

- 591 37. Fewer DP, Wahlsten M, Osterholm J, Jokela J, Rouhiainen L, Kaasalainen U,  
592 Rikkinen J, Sivonen K. 2013. The genetic basis for O-Acetylation of the  
593 microcystin toxin in cyanobacteria. *Chem Biol* 20:861–869.
- 594 38. Murray IA, Shaw WV. 1997. O-acetyltransferases for chloramphenicol and other  
595 natural products. *Antimicrob Agents Chemother* 41:1–6.
- 596 39. Zapomělová E, Jezberová J, Hrouzek P, Hisem D, Řeháková K, Komárková J. 2009.  
597 Polyphasic characterization of three strains of *Anabaena reniformis* and  
598 *Aphanizomenon aphanizomenoides* (cyanobacteria) and their reclassification to  
599 *Sphaerospermum* gen. nov. (incl. *Anabaena kisseleviana*). *J Phycol* 45:1363–1373.
- 600 40. Skulberg OM, Mysterud I, Karlsen J, Tønnesen HH, Laane CMM, Schumacher T.  
601 2012. Alveld research per annum 2012: Searchlight on cyanobacteria we have  
602 minor knowledge of. (In Norwegian). *Biolog* 30: 32–41.
- 603 41. Tønnesen HH, Mysterud I, Karlsen J, Skulberg OM, Laane CMM., Schumacher T.  
604 2013. Identification of singlet oxygen photosensitizers in lambs drinking water in  
605 an alveld risk area in West Norway. *J Photochem Photobiol B* 119: 37–45.
- 606 42. Hegge AB, Mysterud I, Karlsen J, Skulberg OM, Laane CMM, Schumacher T,  
607 Tønnesen HH. 2013. Impaired secondary oxidant deactivation capacity and  
608 enhanced oxidative stress in serum from alveld affected lambs. *J Photochem*  
609 *Photobiol B* 126: 126–134.
- 610 43. Heinze R. 1996. A biotest for hepatotoxins using primary rat hepatocytes. *Phycologia*  
611 35 (Supplement), 89–93.
- 612 44. Skulberg, OM. 1996. Toxins produced by cyanophytes in Norwegian inland waters –  
613 health and environment. – In *Chemical data as a basis of geomedical*  
614 *investigations.* (J. Låg, J. Ed), pp. 197–216, The Norwegian Academy of Science  
615 and Letters.

- 616 45. Jones CG, Firm RD. 1991. On the evolution of plant secondary chemical diversity. *Phil*  
617 *Trans R Soc Lond B* 333:273–280.
- 618 46. Firm RD, Jones CG. The evolution of secondary metabolism - a unifying model. *Mol*  
619 *Microbiol* 37: 989–994.
- 620 47. Rippka R, Deruelles J, Waterbury JB, Herdman M, Stanier RY. 1979. Generic  
621 assignments, strain histories and properties of pure cultures of cyanobacteria. *J*  
622 *Gen Microbiol* 111: 1–61.
- 623 48. Skulberg R, Skulberg OM. 1990. Research with algal cultures – NIVA’s Culture  
624 *Collection of Algae. Norwegian Institute for Water Research*, 1–32. ISBN: 82–  
625 577–1743–6.
- 626 49. Taton A, Grubisic S, Brambilla E, De Wit R, Wilmotte A. 2003. Cyanobacterial  
627 diversity in natural and artificial microbial mats of Lake Fryxell (McMurdo dry  
628 valleys, Antarctica): A morphological and molecular approach. *Appl Environ*  
629 *Microbiol* 69:5157–5169.
- 630 50. Delcher AL, Bratke KA, Powers EC, Salzberg SL. 2007. Identifying bacterial genes  
631 and endosymbiont DNA with Glimmer. *Bioinformatics* 23:673–679.
- 632 51. Bankevich A, Nurk S, Antipov D, Gurevich A, Dvorkin M, Kulikov AS, Lesin V,  
633 Nikolenko S, Pham S, Prjibelski A, Pyshkin A, Sirotkin A, Vyahhi N, Tesler G,  
634 Alekseyev MA, Pevzner PA. 2012. SPAdes: A new genome assembly algorithm  
635 and its applications to single-cell sequencing. *J Comput Biol* 19:455–497.
- 636 52. Weber T, Blin K, Duddela S, Krug D, Kim HU, Bruccoleri R, Lee SY, Fischbach MA,  
637 Muller R, Wohlleben W, Breitling R, Takano E, Medema MH. 2015. antiSMASH  
638 3.0-a comprehensive resource for the genome mining of biosynthetic gene clusters.  
639 *Nucleic Acids Res* 43:W237–W243.

- 640 53. Blin K, Wolf T, Chevrette MG, Lu XW, Schwalen CJ, Kautsar SA, Duran HGS,  
641 Santos E, Kim HU, Nave M, Dickschat JS, Mitchell DA, Shelest E, Breitling R,  
642 Takano E, Lee SY, Weber T, Medema MH. 2017. antiSMASH 4.0-improvements  
643 in chemistry prediction and gene cluster boundary identification. *Nucleic Acids*  
644 *Res* 45:W36–W41.
- 645 54. Wang MX, Carver JJ, Phelan VV, Sanchez LM, Garg N, Peng Y, Nguyen DD,  
646 Watrous J, Kapono CA, Luzzatto-Knaan T, Porto C, Bouslimani A, Melnik AV,  
647 Meehan MJ, Liu WT, Criisemann M, Boudreau PD, Esquenazi E, Sandoval-  
648 Calderon M, Kersten RD, Pace LA, Quinn RA, Duncan KR, Hsu CC, Floros DJ,  
649 Gavilan RG, Kleigrew K, Northen T, Dutton RJ, Parrot D, Carlson EE, Aigle B,  
650 Michelsen CF, Jelsbak L, Sohlenkamp C, Pevzner P, Edlund A, McLean J, Piel J,  
651 Murphy BT, Gerwick L, Liaw CC, Yang YL, Humpf HU, Maansson M, Keyzers  
652 RA, Sims AC, Johnson AR, Sidebottom AM, Sedio BE, *et al.* 2016. Sharing and  
653 community curation of mass spectrometry data with Global Natural Products  
654 Social Molecular Networking. *Nat Biotechnol* 34:828–837.
- 655 55. Johansen JR, Bohunická M, Lukešová A, Hřčková K, Vaccarino MA, Chesarino NM.  
656 2014. Morphological and molecular characterization within 26 strains of the genus  
657 *Cylindrospermum* (Nostocaceae, Cyanobacteria) with descriptions of three new  
658 species. *J. Phycol.* 50:187-202.
- 659 56. Tamrakar A. 2016. Isolation of benthic cyanobacteria and screening of bioactivities  
660 and natural products from culture collection strains. Master Thesis. University of  
661 Helsinki, Faculty of Agriculture and Forestry, Department of Food and  
662 Environmental Sciences.

- 663 57. Kantz T., Bold HC. 1969. Phycological studies IX. Morphological and taxonomic  
664 investigations of *Nostoc* and *Anabaena* in culture. Publication no. 6924. University  
665 of Texas, Austin.
- 666 58. Skulberg OM, Mysterud I, Karlsen J, Tønnesen HH, Laane CMM, Schumacher T.  
667 2012. Alveld research per annum 2012: Searchlight on cyanobacteria we have  
668 minor knowledge of. (In Norwegian). *Biolog* 30: 32-41.
- 669

670 **Table 1. Strains analyzed for PUW/MIN production**

Strain No.	Strain	Isolated by	Date	Locality	Reference
1	<i>Cylindrospermum alatosporum</i> CCALA 988	A. Lukešová	1989	Canada, Manitoba, Riding Mountain National Park, soil	Johansen <i>et al.</i> 2014 <sup>55</sup>
2	<i>Cylindrospermum moravicum</i> CCALA 993	A. Lukešová	2008	Czech Republic, South Moravia, Moravian Karst, Amaterska Cave, cave sediment	Johansen <i>et al.</i> 2014 <sup>55</sup>
3	<i>Cylindrospermum alatosporum</i> CCALA 994	A. Lukešová	2011	Czech Republic, Moravian Karst, earthworms collected from soil above Amaterska Cave, earthworm casings	Johansen <i>et al.</i> 2014 <sup>55</sup>
4	<i>Anabaena</i> sp. UHCC-0399	M. Wahlsten	N/A	Finland, Jurmo, Southwestern Archipelago National Park, copepods	Tamrakar 2016 <sup>56</sup>
5	<i>Anabaena minutissima</i> UTEX B1613	T. Kantz	1967	South Texas, USA, soil	Kantz & Bold 1969 <sup>57</sup>
6	<i>Symplocastrum muelleri</i> NIVA-CYA 644	O.M. Skulberg	2009	Norway; Møre og Romsdal county; Halså municipality, western slope of Slettfjellet mountain in semiterrestrial alpine habitat, biofilm on turf in ombrotrophic blanket bog	Skulberg <i>et al.</i> 2012 <sup>58</sup>

671

672

673

674

675

676

677 **Table 2. Deduced proteins encoded by the *puw* gene cluster in six cyanobacterial strains,**  
 678 **including length and functional annotation. ACP, acyl carrier protein; FAAL, fatty acyl-**  
 679 **AMP ligase; PKS, polyketide synthase; NRPS, non-ribosomal peptide synthetase.**

Protein	Strain No.						Predicted Function
	1	2	3	4	5	6	
ORF1	659	664	664	643	643	647	ABC transporter
PuwA	2870	2870	2870	2854	2854	2866	NRPS
ORF2	1116	1499	1875	643	670	376	patatin-like phospholipase
ORF3	-	-	-	696	696	-	dynamain family protein
PuwI	-	-	-	709	702	711	FAAL, ACP
PuwJ	-	-	-	427	427	529	cytochrome-like protein
PuwB	2534	2592	2592	2549	2537	2555	hybrid PKS/NRPS, aminotransferase, oxygenase
PuwC	597	590	590	597	589	-	FAAL
PuwD	101	104	96	93	92	-	ACP
PuwK	-	-	-	-	465	-	halogenase
PuwE	3077	3121	3121	3099	3112	3113	NRPS
PuwF	2370			5877		3310	NRPS
PuwG	3492	5851	5851		5871		NRPS
PuwH	1102	1081	1102	1121	1121	1408	NRPS
PuwL	-	-	-	-	-	217	<i>O</i> -acetyltransferase

**Table 3. Fragmentation of PUW variants from *Symplocastrum muelleri* strain 6 bearing acetyl substitutions on the FA moiety revealed by high energy (100 eV) fragmentation, and amino acid composition deduced by fragmentation at 60 eV.**

Low fragmentation energy (60eV)	X=Ala, Y=Thr, FA=C <sub>16</sub>			X=Gly, Y=Thr, FA=C <sub>18</sub>			X=Ala, Y=Thr, FA=C <sub>18</sub>			X=Gly, Y=Val, FA=C <sub>18</sub>			X=Ala, Y=Val, FA=C <sub>18</sub>		
	<i>m/z</i>	$\Delta$ (ppm)	Sum formula	<i>m/z</i>	$\Delta$ (ppm)	Sum formula	<i>m/z</i>	$\Delta$ (ppm)	Sum formula	<i>m/z</i>	$\Delta$ (ppm)	Sum formula	<i>m/z</i>	$\Delta$ (ppm)	Sum formula
[M] <sup>+</sup>	1265.7338	+0.7	C <sub>59</sub> H <sub>101</sub> N <sub>12</sub> O <sub>18</sub>	1279.7496	+0.9	C <sub>60</sub> H <sub>103</sub> N <sub>12</sub> O <sub>18</sub>	1293.7654	+0.8	C <sub>61</sub> H <sub>105</sub> N <sub>12</sub> O <sub>18</sub>	1277.7695	+1.6	C <sub>60</sub> H <sub>105</sub> N <sub>12</sub> O <sub>17</sub>	1291.7870	+0.1	C <sub>62</sub> H <sub>107</sub> N <sub>12</sub> O <sub>17</sub>
[M-CH <sub>2</sub> OH] <sup>+</sup>	1233.7170	-6.6	C <sub>58</sub> H <sub>99</sub> N <sub>12</sub> O <sub>17</sub>	1247.7194	+4.1	C <sub>59</sub> H <sub>99</sub> N <sub>12</sub> O <sub>17</sub>	1261.7494	-7.3	C <sub>60</sub> H <sub>101</sub> N <sub>12</sub> O <sub>17</sub>	low int.		C <sub>60</sub> H <sub>101</sub> N <sub>12</sub> O <sub>16</sub>	low int.		C <sub>61</sub> H <sub>103</sub> N <sub>12</sub> O <sub>16</sub>
[M-CH <sub>2</sub> OH-NMeAsn] <sup>+</sup>	1105.6558	-4.9	C <sub>53</sub> H <sub>89</sub> N <sub>10</sub> O <sub>15</sub>	1119.6619	+3.7	C <sub>54</sub> H <sub>91</sub> N <sub>10</sub> O <sub>15</sub>	1133.681	+0.6	C <sub>55</sub> H <sub>93</sub> N <sub>10</sub> O <sub>15</sub>	1117.6924	-5.0	C <sub>55</sub> H <sub>93</sub> N <sub>10</sub> O <sub>14</sub>	1131.7307	-25.0	C <sub>56</sub> H <sub>95</sub> N <sub>10</sub> O <sub>14</sub>
[M-CH <sub>2</sub> OH-NMeAsn-dhb] <sup>+</sup>	1022.6180	-4.7	C <sub>49</sub> H <sub>84</sub> N <sub>9</sub> O <sub>14</sub>	1036.6365	-7.3	C <sub>50</sub> H <sub>86</sub> N <sub>9</sub> O <sub>14</sub>	1050.6478	-3.1	C <sub>51</sub> H <sub>88</sub> N <sub>9</sub> O <sub>14</sub>	1134.6603	-10.3	C <sub>51</sub> H <sub>88</sub> N <sub>9</sub> O <sub>13</sub>	1048.6671	-1.7	C <sub>52</sub> H <sub>90</sub> N <sub>9</sub> O <sub>13</sub>
[M-CH <sub>2</sub> OH-NMeAsn-dhb-X] <sup>+</sup>	951.5785	-2.5	C <sub>48</sub> H <sub>83</sub> N <sub>9</sub> O <sub>13</sub>	979.589	+18.8	C <sub>48</sub> H <sub>83</sub> N <sub>9</sub> O <sub>13</sub>	979.6041	+3.4	C <sub>48</sub> H <sub>82</sub> N <sub>9</sub> O <sub>13</sub>	977.6481	-20.5	C <sub>48</sub> H <sub>82</sub> N <sub>9</sub> O <sub>12</sub>	977.6518	-24.1	C <sub>49</sub> H <sub>83</sub> N <sub>9</sub> O <sub>12</sub>
[M-CH <sub>2</sub> OH-NMeAsn-dhb-X-Gln] <sup>+</sup>	823.5253	-9.4	C <sub>41</sub> H <sub>71</sub> N <sub>8</sub> O <sub>11</sub>	851.5473	+1.8	C <sub>41</sub> H <sub>73</sub> N <sub>8</sub> O <sub>11</sub>	851.5478	+1.2	C <sub>43</sub> H <sub>75</sub> N <sub>8</sub> O <sub>11</sub>	849.5838	-16.7	C <sub>44</sub> H <sub>76</sub> N <sub>8</sub> O <sub>10</sub>	849.5589	+12.5	C <sub>44</sub> H <sub>77</sub> N <sub>8</sub> O <sub>10</sub>
[M-CH <sub>2</sub> OH-NMeAsn-dhb-X-Gln-Y] <sup>+</sup>	722.4729	-4.2	C <sub>37</sub> H <sub>64</sub> N <sub>8</sub> O <sub>9</sub>	750.5005	+0.9	C <sub>38</sub> H <sub>66</sub> N <sub>8</sub> O <sub>9</sub>	750.5147	-18.1	C <sub>39</sub> H <sub>68</sub> N <sub>8</sub> O <sub>9</sub>	low int.		C <sub>40</sub> H <sub>72</sub> N <sub>8</sub> O <sub>8</sub>	low int.		C <sub>40</sub> H <sub>72</sub> N <sub>8</sub> O <sub>8</sub>
[M-CH <sub>2</sub> OH-NMeAsn-dhb-X-Gln-Y-Thr] <sup>+</sup>	621.4223	-0.2	C <sub>33</sub> H <sub>53</sub> N <sub>8</sub> O <sub>7</sub>	649.4526	+1.4	C <sub>33</sub> H <sub>53</sub> N <sub>8</sub> O <sub>7</sub>	649.4539	-0.6	C <sub>33</sub> H <sub>53</sub> N <sub>8</sub> O <sub>7</sub>	649.465	-17.8	C <sub>33</sub> H <sub>53</sub> N <sub>8</sub> O <sub>7</sub>	649.4483	+8.0	C <sub>33</sub> H <sub>53</sub> N <sub>8</sub> O <sub>7</sub>
High fragmentation energy (100eV)															
Fragment 1	411.3208	+2.2	C <sub>23</sub> H <sub>43</sub> N <sub>5</sub> O <sub>4</sub>	439.3559	-6.5	C <sub>24</sub> H <sub>47</sub> N <sub>5</sub> O <sub>4</sub>	439.3556	-5.8	C <sub>25</sub> H <sub>47</sub> N <sub>5</sub> O <sub>4</sub>	439.3556	-5.8	C <sub>25</sub> H <sub>47</sub> N <sub>5</sub> O <sub>4</sub>	439.3508	+5.1	C <sub>25</sub> H <sub>47</sub> N <sub>5</sub> O <sub>4</sub>
Fragment 1 - C <sub>2</sub> H <sub>4</sub> O <sub>2</sub>	351.3006	+0.0	C <sub>21</sub> H <sub>39</sub> N <sub>5</sub> O <sub>2</sub>	379.3334	-4.0	C <sub>23</sub> H <sub>43</sub> N <sub>5</sub> O <sub>2</sub>	379.3329	-2.6	C <sub>23</sub> H <sub>43</sub> N <sub>5</sub> O <sub>2</sub>	379.3328	-2.4	C <sub>23</sub> H <sub>43</sub> N <sub>5</sub> O <sub>2</sub>	379.3360	-10.8	C <sub>23</sub> H <sub>43</sub> N <sub>5</sub> O <sub>2</sub>
Fragment 2	284.2583	+0.4	C <sub>17</sub> H <sub>33</sub> NO <sub>2</sub>	312.2919	-6.9	C <sub>18</sub> H <sub>35</sub> NO <sub>2</sub>	312.2892	+1.6	C <sub>19</sub> H <sub>38</sub> NO <sub>2</sub>	low int.		C <sub>19</sub> H <sub>38</sub> NO <sub>2</sub>	low int.		C <sub>19</sub> H <sub>38</sub> NO <sub>2</sub>
Fragment 2 - C <sub>2</sub> H <sub>4</sub> O <sub>2</sub>	224.2367	+2.6	C <sub>15</sub> H <sub>30</sub> N	252.2686	0.0	C <sub>17</sub> H <sub>34</sub> N	252.2687	-0.5	C <sub>17</sub> H <sub>34</sub> N	252.2684	+0.7	C <sub>17</sub> H <sub>34</sub> N	252.2677	+3.5	C <sub>17</sub> H <sub>34</sub> N

682

683



684 **Table 4. Bacterial and yeast strains used for antimicrobial testing of PUW F and MIN A,**  
 685 **C, and D. HAMBI, culture collection of University of Helsinki, Faculty of Agriculture**  
 686 **and Forestry, Department of Microbiology.**

Test organisms (HAMBI nr.)	Media <sup>a</sup>	Incubation temp. (°C)	Incubation time (h)	Gram strain reaction (+/-)
<i>Pseudomonas</i> sp. (2796)	TGY	28	24	-
<i>Micrococcus luteus</i> (2688)	TGY	28	24	+
<i>Bacillus subtilis</i> (251)	TGY	28	24	+
<i>Pseudomonas aeruginosa</i> (25)	TGY	37	24	-
<i>Escherichia coli</i> (396)	TGY	37	24	-
<i>Bacillus cereus</i> (1881)	TSA	28	24	+
<i>Burkholderia cepacia</i> (2487)	TSA	37	24	-
<i>Staphylococcus aureus</i> (11)	TSA	37	24	+
<i>Xanthomonas campestris</i> (104)	NA	28	24	-
<i>Burkholderia pseudomallei</i> (33)	NA	37	24	-
<i>Salmonella typhi</i> (1306)	NA	37	24	-
<i>Arthrobacter globiformis</i> (1863)	NA	28	24	-
<i>Kocuria varians</i> (40)	NA	28	24	+
<i>Candida albicans</i> (261)	YM agar	37	24	yeast
<i>Cryptococcus albidus</i> (264)	YM agar	28	24	yeast
<i>Saccharomyces cerevisiae</i> (1164)	YM agar	28	24	yeast

687

688 <sup>a</sup>The composition of all media was obtained from the American Type Culture Collection (ATCC).

689 TGY, tryptone glucose yeast; TSA, tryptic soy agar; NA, Nutrient agar; YM agar, yeast malt agar.

690

691 **Figure legends**

692

693 **Fig. 1.** HPLC-HRMS/MS analysis of crude extracts from the investigated strains. Major  
694 puwainaphycin (PUW) and minutissamide (MIN) variants are highlighted. Numbers in  
695 brackets following the peak designation refer to the corresponding Supplementary Figure  
696 containing full MS/MS data. For variants without complete structural information, only  $m/z$   
697 values are shown.

698 **Fig. 2.** Molecular network created using the Global Natural Products Social Molecular  
699 Networking (GNPS) web platform. Two separate networks were obtained during GNPS  
700 analysis; (a) a group containing *Cylindrospermum* strains 1–3 and *Anabaena* strains 4–5, and  
701 (b) a group containing only variants detected in *Symplocastrum muelleri* strain 6. The separate  
702 groups differ mainly in the peptide core of the molecule. Numbers in brackets following the  
703 peak designation refer to the corresponding Supplementary Figure containing full MS/MS  
704 data. For variants without complete structural information, only  $m/z$  values are shown. (\*)  
705 refers to compounds present in trace amounts and (#) refers to compounds for which MS/MS  
706 data failed to resolve the structural information.

707 **Fig. 3.** Structural variability of the peptide core of PUW/MIN variants. Examples of  
708 structural variants PUW F (a) and PUW A (b) with designated aminoacid positions  
709 representing the two major peptide cores. (c) Table summarizing all types of the PUW/MIN  
710 peptide core found in known compounds reported in literature and compounds (Comp.)  
711 detected in studied strains. Columns shaded in grey highlight the conserved aminoacid  
712 positions.

713 **Fig. 4.** Structural variability of the FA moiety of PUW/MIN variants. The relative proportion  
714 of variants with differences in FA length and substitution (y-axis) is depicted using a color  
715 scale (z-axis). For comparison, the peak area of a given variant was normalized against the

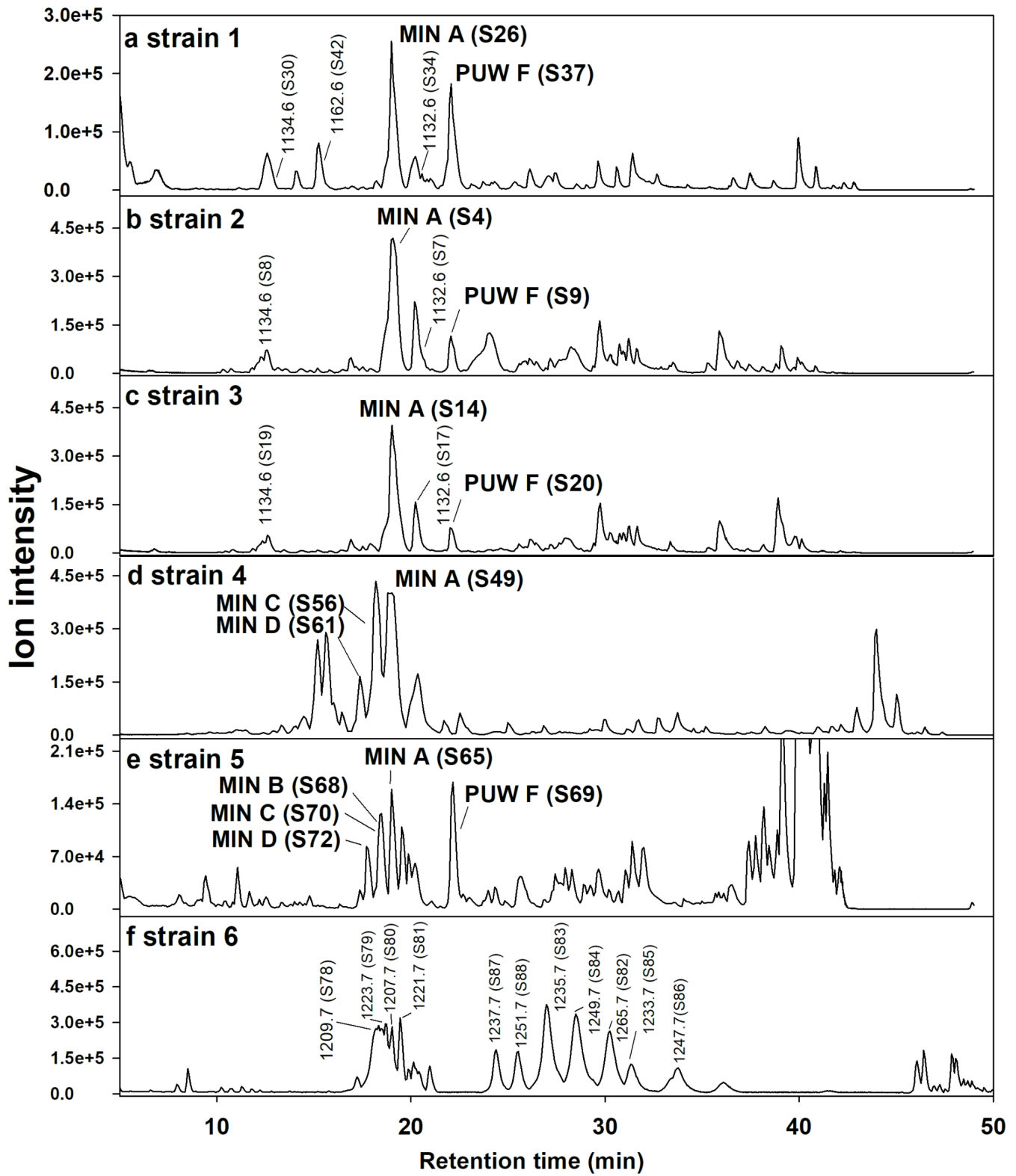
716 peak area of the major variant present in the strain (MIN A for strains 1–5, and  $m/z$  1235.7 for  
717 strain 6).

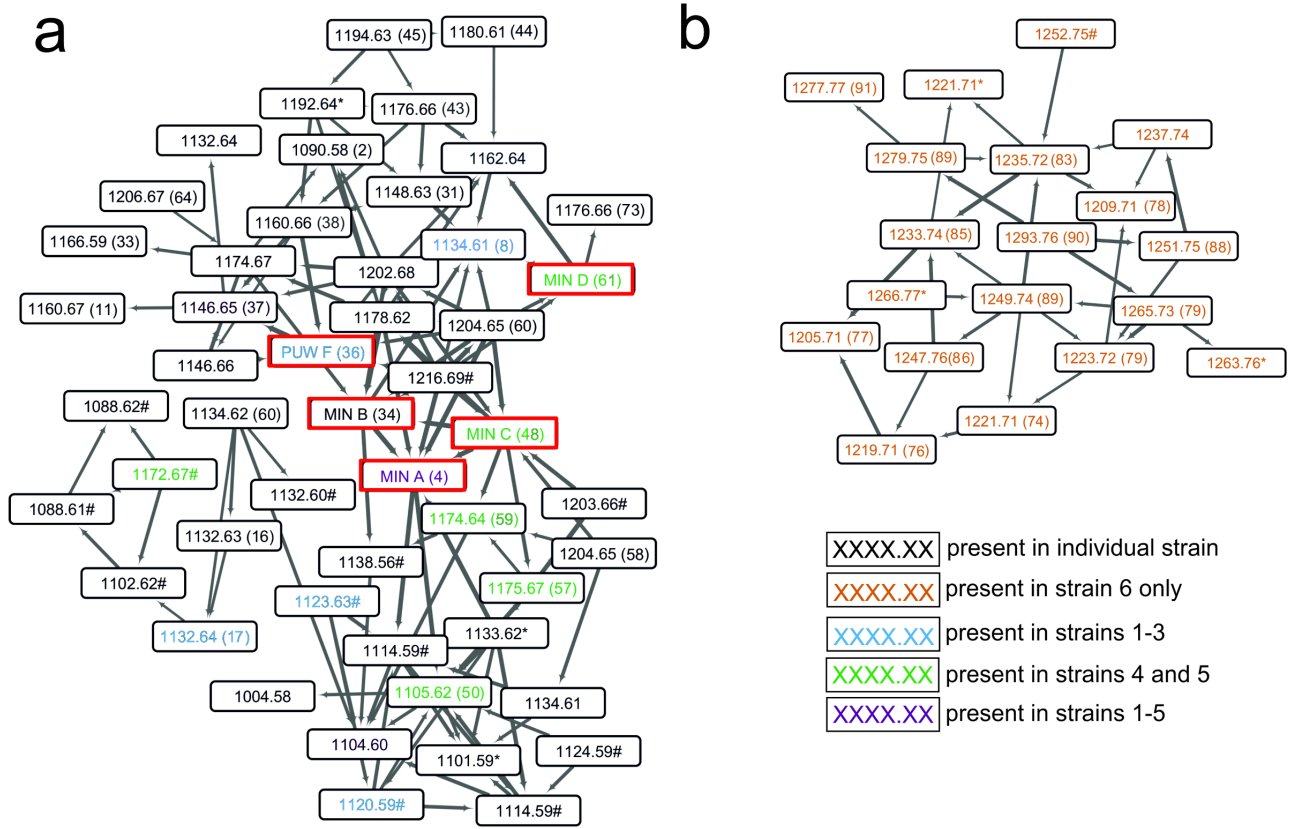
718 **Fig. 5.** Structure of the *puw* gene cluster in the six investigated cyanobacterial strains. Gene  
719 arrangement and functional annotation of *puwA–L* genes and selected PKS/NRPS tailoring  
720 domains is indicated by colored arrows. The distribution of the two observed types of putative  
721 starter modules (shaded boxes) is indicated by bars.

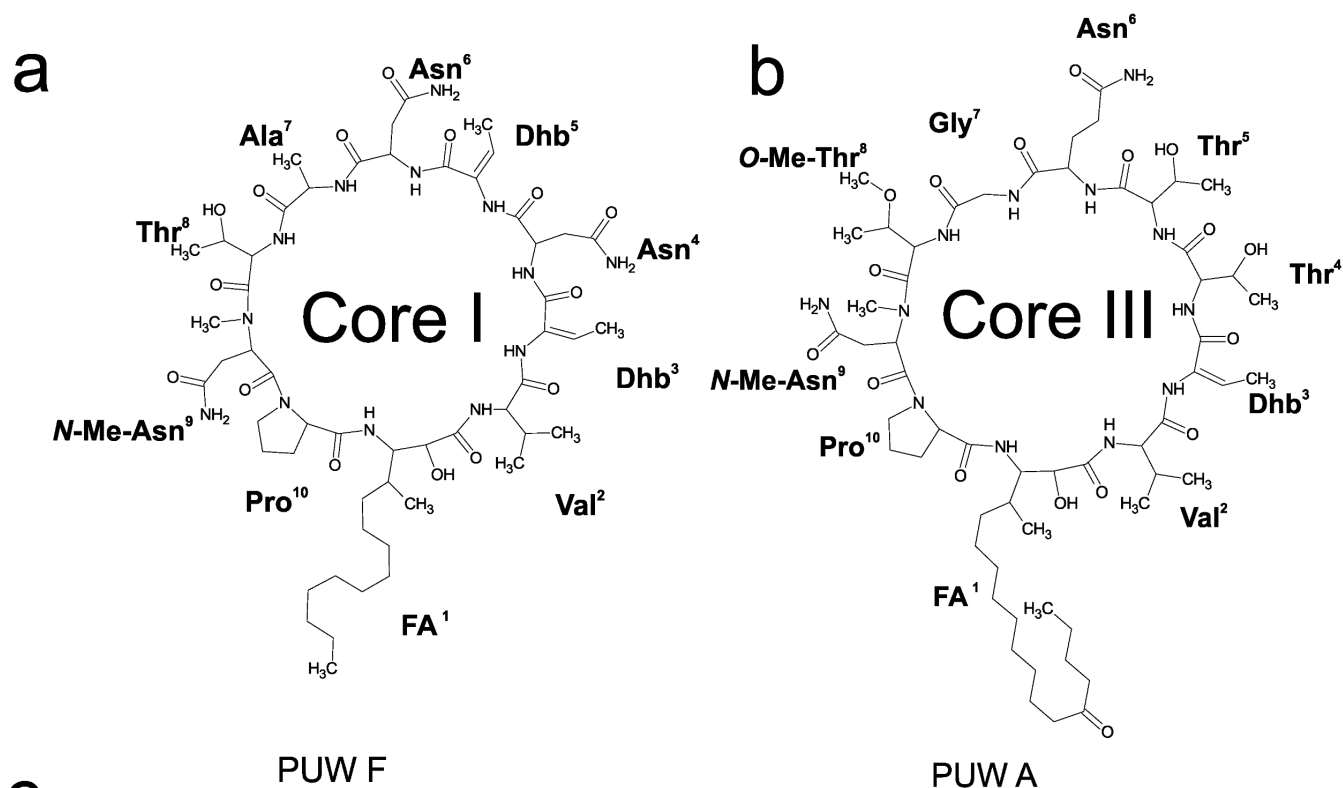
722 **Fig. 6.** Schematic view of the proposed biosynthesis assembly line of puwainaphycins and  
723 minutissamides. Variable amino acid positions and the ranges of fatty acyl lengths  
724 incorporated by the two putative alternative starter units are listed for individual strains. A,  
725 adenylation domain; ACP, acyl carrier protein; AmT, aminotransferase; AT, acyltransferase;  
726 C, condensation domain; DH, dehydratase; E, epimerase; ER, enoylreductase; FAAL, fatty  
727 acyl-AMP ligase; MT, methyltransferase; NRPS, non-ribosomal peptide synthetase; KR,  
728 ketoreductase; KS, ketosynthetase; Ox, monooxygenase; PCP, peptidyl carrier protein; PKS,  
729 polyketide synthetase; TE, thioesterase.

730 **Fig 7.** MS/MS fragmentation of MIN A (a, c, e) and the PUW variant at  $m/z$  1279 bearing an  
731 acetyl substitution of the fatty acid chain (b, d, f). (a–b) base peak chromatograms; (c–d)  
732 fragmentation of the protonated molecule at low fragmentation energy, yielding *b* series of  
733 ions corresponding to the losses of particular amino acid residues, (e–f) fragmentation of the  
734 protonated molecule at high energy (100 eV) yielding fragments characteristic for the  $\beta$ -  
735 amino fatty acid.

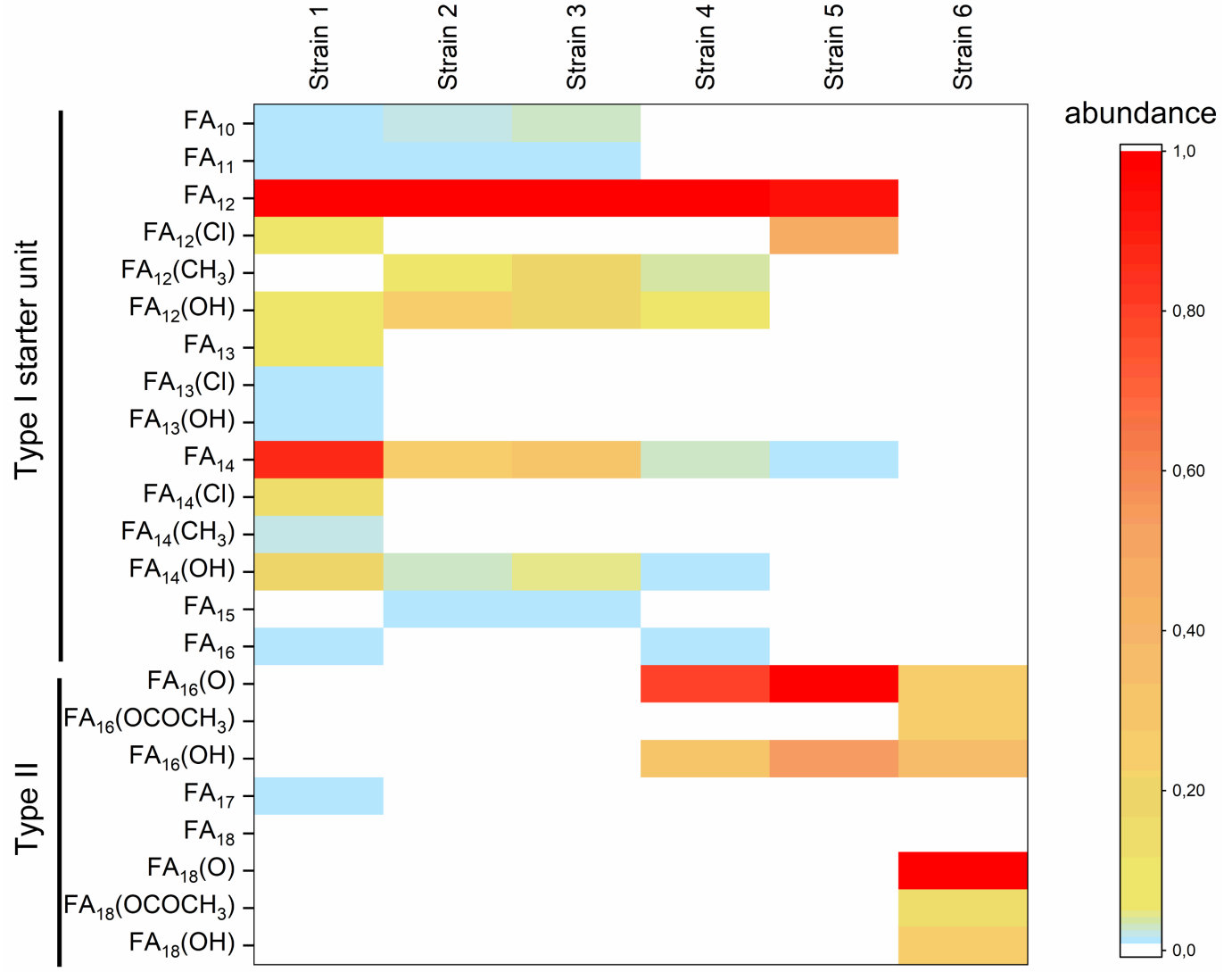
736 **Fig. 8.** Antifungal activity of PUW F against yeast strains (a) *Saccharomyces cerevisiae*  
737 HAMBI 1164 and (b) *Candida albicans* HAMBI 261. Discs were treated with a concentration  
738 range from  $25.2 \mu\text{g mL}^{-1}$  to  $0.0394 \mu\text{g/mL}$  to determine the minimum inhibitory concentration  
739 (MIC). Numbers represent concentrations: (1) =  $25.2 \mu\text{g mL}^{-1}$ ; (2) =  $12.6 \mu\text{g mL}^{-1}$ ; (3) =  $6.3$   
740  $\mu\text{g mL}^{-1}$ ; (+) = positive control (10  $\mu\text{g}$  of nystatin). (-) = negative control (10  $\mu\text{L}$  of methanol).

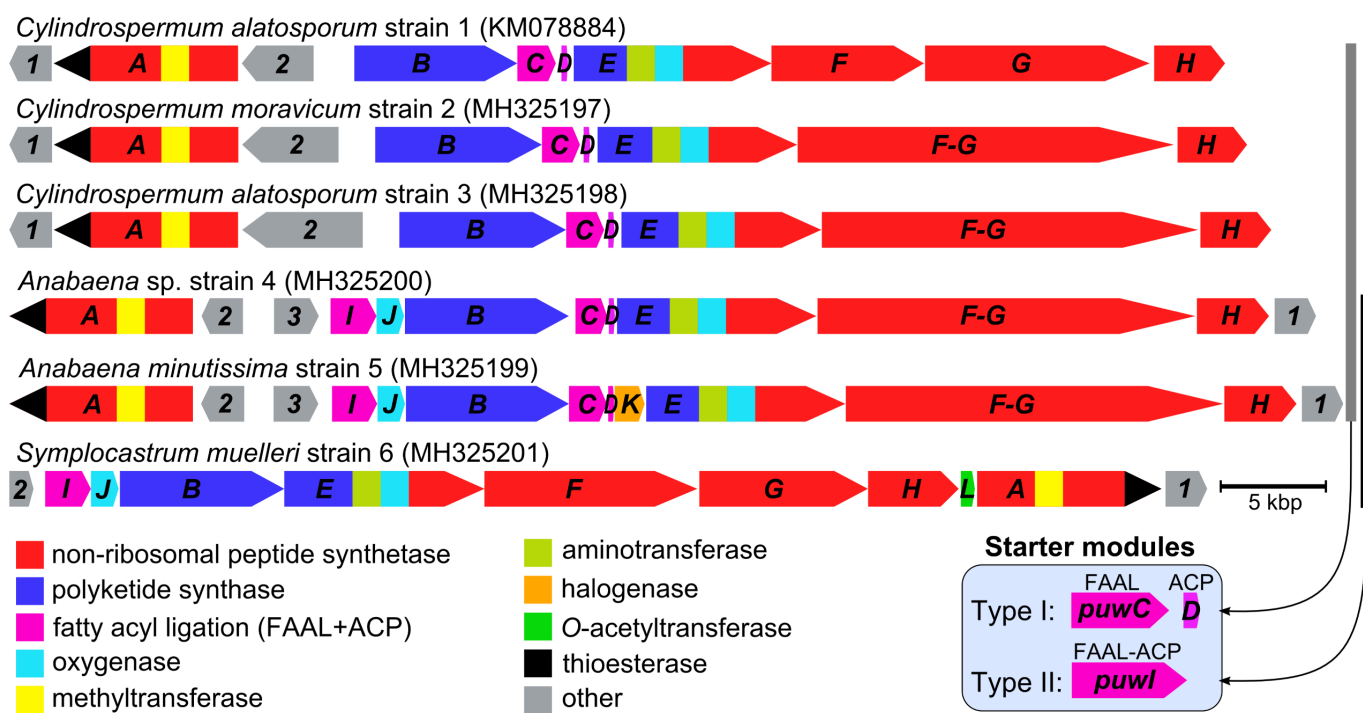




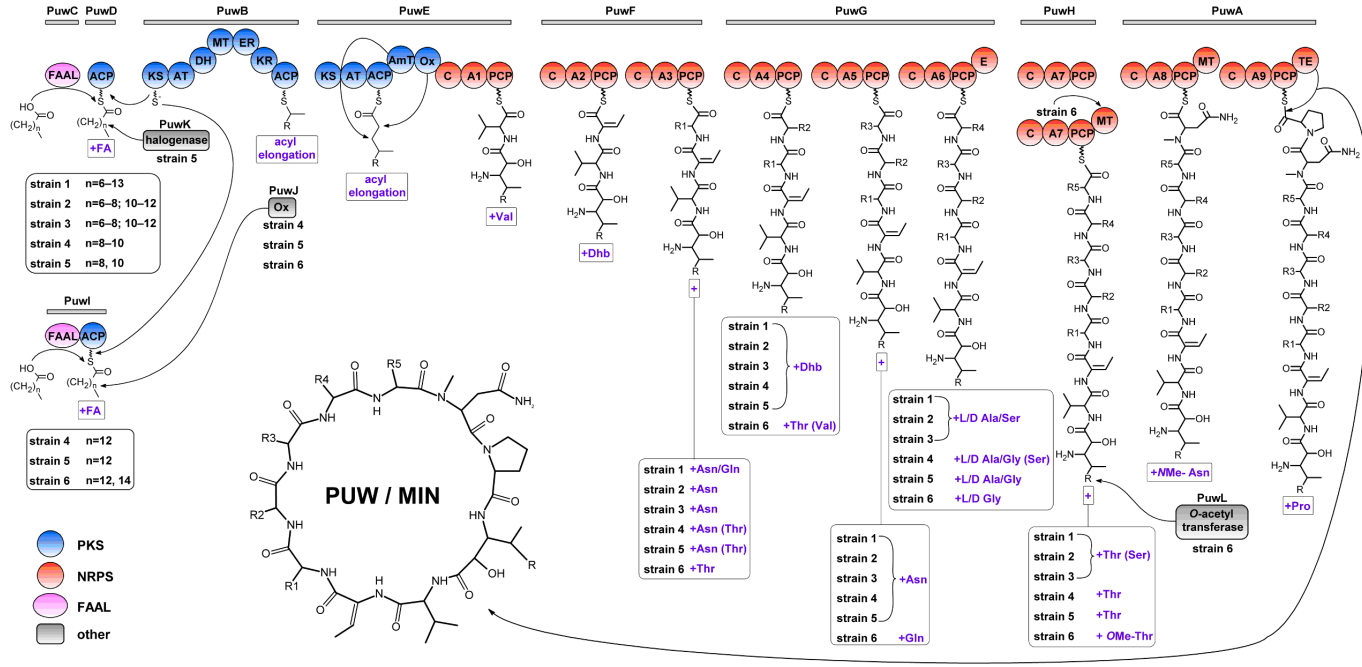


	AA <sub>1</sub>	AA <sub>2</sub>	AA <sub>3</sub>	AA <sub>4</sub>	AA <sub>5</sub>	AA <sub>6</sub>	AA <sub>7</sub>	AA <sub>8</sub>	AA <sub>9</sub>	AA <sub>10</sub>	known compounds	variants found in strains 1-6
Core I	FA	Val	Dhb	Asn	Dhb	Asn	Ala	Thr	<i>N</i> -Me-Asn	Pro	MIN A, MIN B, MIN C, MIN D, PUW F	MIN A, MIN B, MIN C, MIN D, PUW F, Comp. 1,3,10,11,12,14,16,19,20,25,26,28,30,31,37,45,51
Core II	FA	Val	Dhb	Gln	Dhb	Asn	Ala	Thr	<i>N</i> -Me-Asn	Pro	PUW G	PUW G, Comp. 2,6,13,15,17,27,29,32
Core III	FA	Val	Dhb	Thr	Thr	Gln	Gly	<i>O</i> -Me-Thr	<i>N</i> -Me-Asn	Pro	PUW A, PUW C, PUW E, MIN I, MIN K	Comp. 38,46,50,52,53,56,58
Core IV	FA	Val	Dhb	Thr	Thr	Gln	Ala	<i>O</i> -Me-Thr	<i>N</i> -Me-Asn	Pro	MIN E, MIN G	Comp. 39,47,53,57,59
Core V	FA	Val	Dhb	Thr	Val	Gln	Gly	<i>O</i> -Me-Thr	<i>N</i> -Me-Asn	Pro	PUW B, PUW D, MIN L	Comp. 40,48,54,60
Core VI	FA	Val	Dhb	Thr	Val	Gln	Ala	<i>O</i> -Me-Thr	<i>N</i> -Me-Asn	Pro	MIN H	Comp. 41,49,55,61
Core VII	FA	Val	Dhb	Asn	Dhb	Asn	Gly	Thr	<i>N</i> -Me-Asn	Pro		Comp. 7,36,44
Core VIII	FA	Val	Dhb	Asn	Dhb	Asn	Ser	Thr	<i>N</i> -Me-Asn	Pro		Comp. 8,23,35
Core IX	FA	Val	Dhb	Asn	Dhb	Asn	Ala	Ser	<i>N</i> -Me-Asn	Pro		Comp. 9,18,24
Core X	FA	Val	Dhb	Thr	Dhb	Asn	Ala	Thr	<i>N</i> -Me-Asn	Pro		Comp. 5,34,43

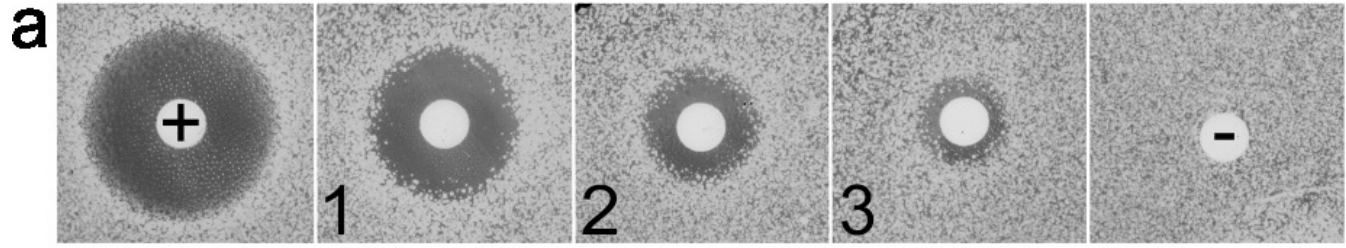










***Saccharomyces cerevisiae******Candida albicans***