Insight into the genome and brackish water adaptation strategies of toxic and bloom-forming Baltic sea Dolichospermum sp. UHCC 0315

Jonna E. Teikari1, Rafael V. Popin1, Shengwei Hou2, Matti Wahlsten1, Wolfgang R. Hess2 & Kaarina Sivonen1

The Baltic Sea is a shallow basin of brackish water in which the spatial salinity gradient is one of the most important factors contributing to species distribution. The Baltic Sea is infamous for its annual cyanobacterial blooms comprised of Nodularia spumigena, Aphanizomenon spp., and Dolichospermum spp. that cause harm, especially for recreational users. To broaden our knowledge of the cyanobacterial adaptation strategies for brackish water environments, we sequenced the entire genome of Dolichospermum sp. UHCC 0315, a species occurring not only in freshwater environments but also in brackish water. Comparative genomics analyses revealed a close association with Dolichospermum sp. UHCC 0090 isolated from a lake in Finland. The genome closure of Dolichospermum sp. UHCC 0315 unraveled a mixture of two subtypes in the original culture, and subtypes exhibited distinct buoyancy phenotypes. Salinity less than 3 g L\(^{-1}\) NaCl enabled proper growth of Dolichospermum sp. UHCC 0315, whereas growth was arrested at moderate salinity (6 g L\(^{-1}\) NaCl). The concentrations of toxins, microcystins, increased at moderate salinity, whereas RNA sequencing data implied that Dolichospermum remodeled its primary metabolism in unfavorable high salinity. Based on our results, the predicted salinity decrease in the Baltic Sea may favor toxic blooms of Dolichospermum sp.
salinity may thus promote the distribution and abundance of Dolichospermum and further favor the formation of toxic Dolichospermum blooms.\(^{6,19}\). Due to the genetic and morphological heterogeneity of the genus Anaebaena, Dolichospermum was recently distinguished from a benthic Anaebaena-type cluster, resulting in the gas vacuole-forming and pelagic morphotype Dolichospermum sp.\(^{21}\). However, Dolichospermum sp. still harbors high genetic and morphological diversity, which results in ambiguity in nomenclatural and phylogenetic clustering. Even mixing with Aphanizomenon spp. and Nostoc sp. has been described.\(^{21-22}\). We have previously defined seven Nostocales subgroups (I a–d; II a–c), based on the comparison of average-nucleotide and amino-acid identities of 30 sequenced Nostocales strains.\(^{21}\). In this classification, Dolichospermum/Anaebaena species were found in four subgroups, implementing a high demand for further taxonomic clarification. However, a latter study suggested even more complex taxonomy due to the close relationship of these species to a third cyanobacterial genus, Aphanizomenon.\(^{24}\)

Here, we sequenced the entire genome of Baltic Sea Dolichospermum sp. UHCC 0315 to reinforce the genomic knowledge of filamentous and toxic Baltic Sea cyanobacteria and to gain insight into the niche adaptation strategies of the examined strain. The isolate was obtained from a coastal cyanobacterial bloom in the Helsinki area.\(^5\). Therefore, we addressed the question of how a slight increase in salinity relevant to the Gulf of Finland can affect the growth and construction of the transcriptomes. Furthermore, comparative genomic analysis was applied to elucidate the ambiguity within the sequenced Dolichospermum/Anaebaena species.

**Results**

**Insight into the genome of Dolichospermum sp. UHCC 0315.** The whole genome sequence of Dolichospermum sp. UHCC 0315 was obtained using Pacific Biosciences (PacBio) chemistry, and sequencing errors were corrected with Illumina Hiseq2500 reads by two rounds of corrections (Freebayes 217 errors and Pilon 8 errors). The assembly yielded one chromosome of 5.17 Mbp and three plasmids (0.85, 0.53, and 0.94 Mbp). Automatic annotation of the Dolichospermum sp. UHCC 0315 genome revealed a total of 5,362 proteins of which 3,570 (66%) were predicted to be part of 367 metabolic subsystems. Among a wide variety of functions, several response mechanisms to osmotic, oxidative and heat stress, as well as to antibiotic and toxic heavy metal defense were identified. Furthermore, one possible CRISPR array with 12 spacers and the direct repeat consensus sequence 5'-CTTGCAATTAACCTAATTACCAAAGCTATTTCACC-3' but without any known Cas proteins was detected (Supplementary Fig. S1). Interestingly, spacer 7 matches prophage antirepressor sequences present in the genomes of other Nostocales, Nostoc 'Peltigera membranacea cyanobiont' N6 and Nostoc flagelliforme CCNUN1 (Supplementary Fig. S1), supporting the further function of this genetic element as a likely CRISPR repeat spacer array further.

Interestingly, the assembly of the Dolichospermum sp. UHCC 0315 genome unraveled two chromosomal subtypes, A and B, of which subtype A carried an additional five adjacent hypothetical genes (BMF77_1250–1254). Based on the Protein Basic Local Alignment Search Tool (BlastP) search, these five genes encode a putative methyltransferase (BMF77_01250), glycosyltransferase (BMF77_01251; BMF77_01253; BMF77_01254), and a hypothetical protein (BMF77_01252). PacBio coverages of the two identified subtypes were 250 (subtype A) and 70 (subtype B). To ensure that both subtypes were properly distinguished by sequence assembly, we isolated 70 filaments under a microscope and cultivated them in liquid medium. Surprisingly 39 of the cultures were buoyant (subtype A) and 31 grew on the bottom of the flask (subtype B) (Fig. 1a,b), despite both subtypes carried full gvp gene cluster for gas vesicle formation. Differences in cellular morphology under the light microscope were additionally observed and indicate the presence of aerotopes in substrate A (Fig. 1c,d). Presence or absence of the 5-gene region was confirmed by PCR amplification of a segment from gene BMF77_1254 using specific primers (Supplementary Table S1). The amplification product was found only in the buoyant subtype A (Fig. 1e, Supplementary Fig. S2), which further showed that the absence of the BMF77_1250–1254 genes in subtype B was real. Furthermore, the distribution of all three plasmids was tested, using primers amplifying certain regions of the plasmids, and the PCR results indicated that all three plasmids were present in both substraits (Fig. 1f–h, Supplementary Fig. S3).

**Comparative genomics.** A phylogenetic tree, based on 31 conserved marker genes was constructed to show the placement of the newly sequenced Dolichospermum sp. UHCC 0315 within the cyanobacterial phylum (Fig. 2a). The strain branched together with 16 other Dolichospermum sp., Anaebaena sp. and Aphanizomenon flos-aquae with diverse biosynthetic gene clusters, forming a distinct Anaebaena/Dolichospermum/Aphanizomenon clade (ADA)\(^{24}\). Other 4 Anaebaena strains (PCC 7122, PCC 7108. ATCC 33047 and 4-3) formed different branches in the tree. Furthermore, Average Amino Acid and Nucleotide Identity analyses (Fig. 2b,c, respectively) highlighted the division of the genomes in 6 different subclades (Ia–b, Ia and III). Considering the genome of Dolichospermum sp. UHCC 0315, there are five genomes of the genus that are publicly available (Dolichospermum cirratile AWQC310F and AWQC131C and Dolichospermum sp. NIES-806 and UHCC 0090). However, while only the UHCC 0315 and UHCC 0090 strains grouped together in a subclade with other 3 Anaebaena strains (AL09, LE011-02 and MDT14b) in the genomic analyses, the other ones were scattered in other subclades. A 16S rRNA phylogenetic tree was constructed to further explore the ADA clade (Supplementary Fig. S4).

Orthologous gene clustering was employed to estimate the core and pan-genome of the 21 Anaebaena, Dolichospermum and Aphanizomenon genomes (Fig. 3). Whereas the core-genome asymptote is composed of approximately 1,500 genes, the current pan-genome reached more than 12,000 different genes belonging to the ADA clade.

The closest relative to Dolichospermum sp. UHCC 00315 was the freshwater strain Dolichospermum sp. UHCC 0090 (previously Anaebaena sp. 90). Synteny analysis showed a high number of conserved blocks between the two strains in the chromosomes and the pDOL01 and pUHCC0315a plasmids. Nonetheless, the other two plasmids of each strain seemed to be considerably less conserved (Supplementary Fig. S5). The two genomes encode a total
of 9,873 proteins of which 7,101 formed shared orthologous clusters (Supplementary Fig. S6). Around 93% of the proteins in those clusters were predicted to be involved in a wide diversity of functions of the cellular primary and secondary metabolisms. On the other hand, 1,816 and 956 proteins were found to be specific to Dolichospermum sp. UHCC 0315 and UHCC 0090, respectively (21.19% and 33.86% of the total number of proteins encoded in each genome). Functions for these two groups of proteins were less successfully predicted compared to the shared genes, pointing to many previously unknown genes among them.

Growth of Dolichospermum sp. UHCC 0315. Dolichospermum sp. UHCC 0315 growth rate was followed at four different salinities (0, 3, 6, or 9 g L\(^{-1}\) NaCl) by measuring the chlorophyll a concentration (Fig. 4a). The highest tested salinity (9 g L\(^{-1}\) NaCl) was deleterious for Dolichospermum sp. UHCC 0315, while 6 g L\(^{-1}\) NaCl inhibited the growth of the culture. Dolichospermum sp. UHCC 0315 was able to grow properly in 3 g L\(^{-1}\) NaCl, but the fastest growth was obtained under freshwater conditions (no added NaCl). Due to the appearance of a yellowish color in the culture, cell counting using the microscope technique was applied to follow the growth (Fig. 4b). The patterns of chlorophyll a and cell counting seemed to correlate, and thus the concentration of chlorophyll a was later used for normalization of the toxin concentration.

Reconstruction of transcriptomes under unfavorable salinities. To understand the cellular response of Dolichospermum sp. UHCC 0315 at moderate salinity, we applied RNA sequencing (RNA-Seq) analysis to compare the normal (0 g L\(^{-1}\) NaCl) and unfavorable conditions (6 g L\(^{-1}\) NaCl). We mapped a total of 4.1 \times 10^8 sequenced paired-end Illumina HiSeq reads for three biological replicates of the control and 4.2 \times 10^8 for three biological replicates of the treatment against the sequenced reference genome. Since the Dolichospermum sp. UHCC 0315 B lacked five genes and the RNA-seq data was generated from the mixture of the subcultures, Dolichospermum sp. UHCC 0315 A was used as a reference genome for mapping and calling of the differentially
Figure 2. Maximum likelihood phylogenomic tree based on the concatenated alignment of 31 universal marker genes from 75 cyanobacterial genomes. Those analyzed in this study are shown in bold and the presence of gene clusters of natural products are represented (a). Average Amino Acid Identity (AAI) (b), and Average Nucleotide Identity (ANI) (c) heatmaps of genomes in subgroups (Iα–δ, II and III). Subgroups were determined as described earlier.23-24

Figure 3. Core and pan-genome plots estimated using the ADA clade (Anabaena sp. WA102, WA93, AL09, LE011-02, MDT14b, WA113 and CRK333; Aphanizomenon flos-aquae NIES-81, 2012/KM1/D3, LD13, MDT14a and WA102; Dolichospermum sp. NIES-86, UHCC00315, UHCC 0090; and Dolichospermum circinale AWQC310F and AWQC131C plus Anabaena sp. PCC 7122, PCC7108, ATCC 33047 and 4-3. (a) Number of genes in the core-genome size by the number of genomes and (b) number of genes in the pan-genome by the number of genomes.
expressed genes. Using a Log2 fold change (Log2FC) of ±1 and a False-Discovery Rate (FDR) cutoff value <0.01 provided by DESeq software25, the number of differentially expressed genes for *Dolichospermum* sp. UHCC 0315 was 348, of which 197 genes were upregulated and 151 were downregulated (Supplementary Table S2). A total of 136 upregulated and 90 downregulated genes were classified as hypothetical proteins or proteins with unknown functions, showing an extensive gap in our knowledge of the metabolism and genetic reservoir of the examined cyanobacteria. *Dolichospermum* sp. UHCC 0315 was struggling in the unfavorable environment of moderate salinity, suppressing many metabolic processes and protecting cellular biomolecules (Supplementary Fig. S7). The high demand for metabolism remodeling was implemented by adjusting the gene expression of transcription and translation, including 10 upregulated and two downregulated genes (Table 1). Furthermore, the pool of chaperones was reorganized. Nitrogen fixation is energetically expensive for cyanobacteria and pathway seemed to be mainly repressed (BMF77_01942-3; BMF77_01966-67; BMF77_01994-96) despite one nitrogenase gene was upregulated (BMF77_00202). Moreover, the nitrogen requirement was most likely fulfilled by reorganizing existing nitrogen reservoirs e.g. as indicated by the induced expression of BMF77_01290, nitrate reduction (BMF77_3991; 3993), and fine-tuning amino-acid metabolism (BMF77_00644; 02305; 03528). High demand for the reorganization of nitrogen metabolism was also seen among 30 most up and downregulated genes (Fig. 5). A general stress response was implemented, hindering the expression of the genes associated with photosynthesis (BMF77_00624; 00701; 01361; 02844; 02798) and electron transfer (BMF77_00909; 01935; 01997; 02996-7; 03623). In addition, induction of genes involved in modifying cell-wall components or encoding multiple transporters illustrated a need for reorganization of the cell-wall structure. In freshwater cyanobacteria, trehalose and sucrose are the major compatible solutes. Sucrose synthase genes were found in *Dolichospermum* UHCC 0315 but, surprisingly, there was no differential gene expression detected for the sucrose synthase genes. Hence, the absence of certain genes for compatible solute synthesis and the lacking induction of certain genes are probably important factors limiting the acclimation of *Dolichospermum* UHCC 0315 to higher salinity.

**Biosynthetic potential.** Based on antiSMASH analysis, the genome of *Dolichospermum* sp. UHCC 0315 harbors six known and four unknown gene clusters for the synthesis of naturally bioactive metabolites (Fig. 1a, Supplementary Table S3). Due to the capability of the strain to produce microcystins, our next interest was to follow the toxin concentration at freshwater and moderate salinities. At the latter salinity, the microcystin gene cluster was slightly upregulated (Table 2). Interestingly, the normalized microcystin concentration increased at unfavorably high salinities, reaching the highest amount at day 8 (892.4 ng microcystin/µg chlorophyll, Fig. 6). After that time, the concentration decreased slightly and approached the toxin level of the control cultures during the later phase of the experiment, whereas total quota of the microcystin concentration in the system increased during the experiment (Supplementary Fig. S8a). Combined intra- and extracellular microcystin concentrations were measured every fourth day at 0 and 6 g L−1 NaCl (Supplementary Fig. S8b).

**Discussion**
Species diversity in brackish water ecosystems is usually relatively narrow, because most of the organisms are adapted to live in either a marine or freshwater environment2. A recent study showed that *Nodularia spumigena* has remarkable genomic plasticity in acclimation strategies to the brackish water Baltic Sea23. However, in brackish water regions, *N. spumigena* coexisted with toxic *Dolichospermum* spp. but the former managed to sense the higher salinities fundamentally better than in freshwater environments23. In the light of projected salinity decreases in the Baltic Sea area, the question of whether freshwater origin *Dolichospermum* spp. may predominate in future cyanobacterial blooms in the Baltic Sea has arisen. Cyanobacteria from this genus are filamentous, capable of fixing atmospheric nitrogen, usually produce toxins, conditionally buoyant, and form akinetes.
| Gene          | Product                                               | Log2 FC |
|--------------|-------------------------------------------------------|---------|
| **Photosynthesis and electron transport**                |                                                     |         |
| BMF77_02844  | Allophycocyanin alpha chain                          | −1,17   |
| BMF77_02798  | Photosystem I reaction center subunit PsaK            | −1,10   |
| BMF77_01361  | Photosystem II protein Y                              | −1,03   |
| BMF77_00624  | Photosystem II reaction center protein K              | −1,28   |
| BMF77_00701  | Phycobilisome rod-core linker polypeptide CpcG4       | −1,13   |
| BMF77_00909  | Cytochrome c oxidase subunit 2                        | −1,05   |
| BMF77_02997  | Ferredoxin                                           | −1,89   |
| BMF77_01997  | Ferredoxin                                           | −1,38   |
| BMF77_01935  | Ferredoxin, heterocyst                                | −1,57   |
| BMF77_02996  | Ferredoxin-3                                         | −1,79   |
| BMF77_03623  | Plastocyanin                                         | −1,25   |
| **Nitrogen cycle**                                      |                                                     |         |
| BMF77_04749  | 2-isopropylmalate synthase                           | −1,21   |
| BMF77_01994  | Nitrogenase iron protein                             | −2,00   |
| BMF77_01967  | Nitrogenase iron protein 1                           | −1,98   |
| BMF77_01966  | Nitrogenase molybdenum-iron protein alpha chain       | −2,21   |
| BMF77_01944  | Nitrogenase molybdenum-iron protein beta chain        | −1,95   |
| BMF77_01942  | Nitrogenase molybdenum-iron protein beta chain        | −1,50   |
| BMF77_01938  | Nitrogenase-stabilizing/protective protein NifW 2     | −1,41   |
| BMF77_03247  | Global nitrogen regulator                            | −1,04   |
| BMF77_03993  | Sulphite reductase [ferredoxin]                       | 2,86    |
| BMF77_03991  | Nitrate reductase                                    | 3,37    |
| BMF77_00202  | Nitrogenase iron protein 2                           | 1,32    |
| **Amino acid metabolism**                               |                                                     |         |
| BMF77_03528  | Glutamate racemase                                    | 1,12    |
| BMF77_02305  | LL-diaminopimelate aminotransferase                  | 1,06    |
| BMF77_00605  | High-affinity branched-chain amino acid transport system permease protein LivH | 1,10 |
| BMF77_01290  | Putative serine protease HtrA                         | 1,01    |
| BMF77_00644  | Aspartate aminotransfer                               | −1,20   |
| BMF77_01666  | Cystathionine beta-lyase MetC                         | −1,04   |
| BMF77_01996  | Cysteine desulfurase NiF                             | −1,22   |
| BMF77_01034  | ATP-dependent Clp protease ATP-binding subunit ClpC  | −1,08   |
| **Transcription and translation**                        |                                                     |         |
| BMF77_04949  | 50S ribosomal protein L16 arginine hydroxylase       | 1,18    |
| BMF77_01396  | ATP-dependent RNA helicase RhIE                       | 1,42    |
| BMF77_04171  | ECF RNA polymerase sigma factor SigE                 | 1,52    |
| BMF77_04319  | NAPDH-dependent 7-cyano-7-deazaguanine reductase      | 1,11    |
| BMF77_02346  | Putative dual-specificity RNA methyltransferase RlmN | 1,10    |
| BMF77_00639  | Ribosome biogenesis GTPase A                          | 1,21    |
| BMF77_04465  | RNA polymerase sigma factor SigA                     | 1,01    |
| BMF77_00231  | Serine–tRNA ligase                                   | 1,36    |
| BMF77_02360  | SarA-binding protein                                 | 1,04    |
| BMF77_03965  | tRNA thymylcarbamoyladenosine biosynthesis protein TsaB | 1,38 |
| BMF77_02233  | tRNA pseudouridine synthase A                         | −1,02   |
| BMF77_00676  | tRNA-His( gtg)                                       | −1,79   |
| **Chaperones**                                          |                                                     |         |
| BMF77_03414  | 10kDa chaperonin                                      | −1,17   |
| BMF77_02118  | Chaperone protein ClpB                               | −1,35   |
| BMF77_00359  | Chaperone protein DnaK2                              | −1,39   |
| BMF77_00729  | Chaperone protein DnaK                              | 1,01    |
| BMF77_04122  | Chaperone protein DnaK                              | 1,19    |
| **Cell wall and lipids**                                |                                                     |         |
| BMF77_02633  | Malonyl CoA acyl carrier protein transacylase        | 1,14    |
| BMF77_03151  | Cellulosome-anchoring protein                        | 1,44    |
| BMF77_02469  | TVP38/TMEM64 family inner membrane protein YdjZ       | 1,09    |
| BMF77_02020  | Murein DD-endopeptidase MeprS/Murein LD-carboxypeptidase | 1,10 |

Table 1. List of the selected differentially expressed genes of *Dolichospermum* sp. UHCC 0315 in moderate salinity (6 g L−1 NaCl). Negative Log2 Fold Change values showed downregulated genes whereas positive values show up-regulated genes.
under unfavorable conditions\textsuperscript{24,26}. The abundance of Dolichospermum in the Baltic Sea is thus a consequence of ecological niche adaptation, for which the genomic inventory and regulatory apparatus provide the genetic underpinning.

Two substrains of Dolichospermum sp. UHCC 0315 (A and B) were distinguished at the genomic and physiological levels. Substrain B lacked a region of five genes, whereas substrain A harbored these genes and was capable of buoyancy. Otherwise, the genomic structure, including the same combination of plasmids, was similar albeit information on single-nucleotide polymorphisms (SNPs) was not gained by the methods used in this study. The association between the genes and the observed buoyancy phenotype is not directly clear, because both subtypes carry the full set of genes for gas vesicle formation (\textit{gvpA}, \textit{gvpC}, \textit{gvpN}, \textit{gvpK}, \textit{gvpF}, \textit{gvpG}, \textit{gvpV}, \textit{gvpW}, \textit{BMF77_4471–4480}), similar to \textit{Nostoc} sp. PCC7120 and Dolichospermum sp. UHCC 009027,\textsuperscript{28}. Moreover, SNPs may have a great impact on the gene function. Even though the buoyancy of the cyanobacteria can be controlled by mechanisms other than gas vesicles\textsuperscript{29}, the functions tentatively assigned to these five genes point towards the modification of surface properties, so a link, even if circumstantial, with buoyancy cannot be excluded. It is unknown when these changes occurred, during the years in the culture collection or before. However, stable laboratory conditions force cyanobacteria to diminish their energy demand by eliminating functionally redundant and nonessential parts of the genome that enhance fitness in the environment but are unnecessary under stable laboratory conditions\textsuperscript{28}. Moreover, water column vertical migration and co-occurrence of buoyant/non-buoyant populations of cyanobacteria have been documented as result of variations in light incidence and nutrient concentration\textsuperscript{30}. Identification of two substrains in the same culture underpins the importance of whole-genome analysis to elucidate explicit genetic structure and enhance the knowledge of novel genetic elements. The method coupling the short-read Illumina and long-read PacBio sequences used in this study serves as a platform for sequencing plastic genomes with high numbers of repetitive elements.

Figure 5. Heatmaps presenting Log2 Fold Changes of 30 most upregulated (a) and downregulated (b) genes in triplicates at two different salinities (0 g L\textsuperscript{-1} and 6 g L\textsuperscript{-1}, of added NaCl).

| Gene   | Gene name | Product                                      | Log2 FC | FDR       |
|--------|-----------|----------------------------------------------|---------|-----------|
| BMF77_03385 | mcyC      | non-ribosomal peptide synthetase              | 0.53    | 7.36E-09  |
| BMF77_03386 | mcyB      | non-ribosomal peptide synthetase              | 0.26    | 2.54E-02  |
| BMF77_03387 | mcyA      | McyA protein                                  | 0.27    | 1.81E-02  |
| BMF77_03388 | mcyG      | peptide synthetase polyketide synthase fusion protein McyG | 0.36    | 1.14E-03  |
| BMF77_03389 | mcyD      | polyketide synthase                           | 0.38    | 4.77E-05  |
| BMF77_03390 | mcyE      | methyltransfer                                | 0.30    | 2.33E-03  |
| BMF77_03391 | mcyF      | hybrid non-ribosomal peptide synthase/polyketide synthase | 0.41    | 8.97E-05  |
| BMF77_03392 | mcyG      | Asp/Glu racemase McyF                         | 0.56    | 1.43E-06  |
| BMF77_03393 | mcyF      | dehydrogenase McyF                            | 0.44    | 3.37E-07  |
| BMF77_03394 | mcyH      | ABC transporter ATP-binding protein            | 0.31    | 2.00E-02  |

Table 2. Expression of the genes in the microcystin synthetase (mcy) gene cluster. Positive Log2Fold Change values showed that the whole mcy gene cluster was slightly significantly upregulated in moderate salinity (6 g L\textsuperscript{-1} NaCl). Log2 FC = Log2 fold change. FDR = False discovery rate.

The genus Anabaena was recently separated into two genera: benthic Anabaena and gas-vesicle-producing planktonic Dolichospermum\textsuperscript{20}. However, the new nomenclature has not yet been fully adopted by the scientific community, and mixing between these two genera and even Aphanizomenon and Nostoc species still occurs\textsuperscript{9,17}. Nonetheless, a recent study indicates that Dolichospermum/Anabaena possess closely related genomic features.
Anabaena studies will possibly indicate whether the ADA clade is polyphyletic or should these four Dolichospermum reclassified in a different taxon. As expected, the addition of Cas12b (C2c1) and C2c344. The possibility that this array is part of a functional CRISPR system is supported fur-similarity to TnpB transposases of the IS605 family, a feature associated also with CRISPR effector proteins unanswered42.

Dolichospermum study, the expression of group 2 sigma factors was heavily repressed in the pan-genome of the clade is yet open and calculated to be currently similar to the one for Microcystis sp. UHCC 0315 branched with several other members of the Anabaena/Dolichospermum/Aphanizomenon clade (ADA), while N. spumigena was separated into a single tight cluster in accordance to previous results23. However, 4 Anabaena strains (PCC 7122, PCC 7108, ATCC 33047 and 4-3) genomes formed 2 divergent clades as previ-ously shown29 and might indicate a yet not fully understood taxonomic relationship. In the case of strains ATCC 33047 and 4-3, the taxonomy seems to be even more unclear since 16S rRNA allocated these two strains among Nostoc, Desmonostoc and Alalinostoc. The inclusion of new genomes of these cyanobacteria and further taxonomic studies will possibly indicate whether the ADA clade is polyphyletic or should these four Anabaena genomes be reclassified in a different taxon. As expected, the addition of Dolichospermum sp. UHCC 0315 to the ADA group core-genome analysis did not affect the previously estimated common gene pool of 1,500 genes36. However, the pan-genome of the clade is yet open and calculated to be currently similar to the one for Microcystis (12,000)31. Therefore, future inclusions of new genomes are likely to expand the full complement of genes of the clade.

The genome size of newly sequenced Dolichospermum sp. UHCC 0315 was 5.4 Mbp, which is approximately the same size as its closest counterpart Dolichospermum sp. UHCC 0090 (5.31 Mbp)38. Moreover, the comparison of the two complete genomes indicated a relatively high level of conservation between them and no clear diver-gence of biosynthetic potential or coding sequences. Therefore, the close genetic association of UHCC 0315 with a freshwater cyanobacteria and the harmfulness of small increases in salinity indicate that Dolichospermum spp. have most probably been transported from freshwater ecosystems to the Baltic Sea.

The most rapid growth of Dolichospermum sp. UHCC 0315 was obtained under freshwater conditions (0 g L−1 NaCl), while higher salinities (>3 g L−1 NaCl) clearly hindered growth. However, slight increases in salinity (3 g L−1 NaCl) were not harmful for the strain, suggesting that the examined cyanobacterium sensed minor increases in salinity in studied environments, which may be caused by water fluctuation and saltwater pulses. Our findings agree well with field studies in the Baltic Sea in which salt concentrations play crucial roles in the abun-dance and intensity of N. spumigena and Dolichospermum spp. in cyanobacterial blooms19,23. Furthermore, the abundance of Dolichospermum is higher in the low-salinity coastal regions, whereas N. spumigena predominates in the more saline areas8,10,

In general, salt shock inhibits proper functioning of several proteins and metabolic processes such as photosynthesis, central metabolism, and cellular growth, clearly requiring the reconstruction of cellular metabo-lism32–40. For example, nitrogen fixation is an energy-demanding process for diazotrophic cyanobacteria38, and thus reducing expression of the nitrogen fixation (nif) gene cluster decreased the energy requirement of Dolichospermum sp. UHCC 0315 under stressful conditions. The transcript accumulation of chaperones, the molecules that play crucial roles in protein folding, protection, and repair under stressful conditions, were heavily modified. The role of chaperones in maintaining protein integrity under low and high salt stress was previously described35,41, and this study further showed the demand for different types of chaperones in salinity stress. In this study, the expression of group 2 sigma factors was heavily repressed in Dolichospermum UHCC 0315 growing at high salinity. The question of which sigma factors participate in gene expression in changing salinities remains unanswered32.

The genomes of cyanobacteria are highly dynamic entities. One factor affecting genome composition and especially the responses to phage attacks are CRISPR/Cas systems. Therefore, the finding of an unusual CRISPR repeat-spacer array lacking any known cas gene is of interest. Possible orphan repeat-spacer arrays exist in many filamentous cyanobacteria41. However, just next to this array is a gene encoding a 696 amino acid protein with similarity to TnpB transposases of the IS605 family, a feature associated also with CRISPR effector proteins Cas12b (C2c1) and C2c344. The possibility that this array is part of a functional CRISPR system is supported fur-ther by our finding that spacer 7 likely targets a phage antirepressor gene present in at least two other Nostocales.

The induction of either general or stress-specific genes and regulatory systems enables cyanobacteria to thrive, even in changing environments. The transcriptional regulatory system of model cyanobacteria for adapting to

Figure 6. Amount of microcystins normalized against chlorophyll a concentration produced by Dolichospermum sp. UHCC 0315 during the experiment of 24 days. The salinities refer to the added NaCl.
showed that the number of differentially expressed genes peaked after 30 min of salt shock, and the majority of these genes returned to the control level after acclimation for 24 h52. Relying on the number of differentially expressed genes found after 16 days of incubation in different salt concentrations, we assumed that reconstruction of the transcriptional pattern remained vigorous long enough to minimize the harmfulness of unfavorable salt conditions and fine-tune the metabolic patterns and structural components. Accumulation and production of compatible solutes together with ion exchange through the cell membrane are major salt-stress adaptation strategies for minimizing the harmfulness of increased intracellular ion concentrations56. Of the compatible solutes, low-halotolerance strains are able to produce trehalose and sucrose37. However, only genes for sucrose synthesis were found in the genome of Dolichospermum sp. UHCC 0315 but increased abundance of the transcripts was not identified.

Cyanobacterial genomes harbor a wide variety of gene clusters responsible for the production of biologically active natural products, of which toxins, e.g. microcystins, are the most studied molecules, due to their harmfulness to mammals. A search of the gene clusters responsible for the production of natural products in Dolichospermum sp. UHCC 0315 resulted in similar genomic construction to that found in the genome of Dolichospermum sp. UHCC 009027. However, while the anabaenopeptin gene cluster was absent in the genome of the later, the hassallidin was absent in the former. The concentration of microcystin/cell produced by Dolichospermum sp. UHCC 0315 increased at moderate salinity (6 g L\(^{-1}\) NaCl), which is well in line with previous studies in which high salinity triggered microcystin production in Dolichospermum52. Moreover, salt shock may cause oxidative stress in cyanobacteria which is harmful especially to lipids, and activation of lipid metabolism in the target strain cells was also found52,46. Despite the role of toxins in cyanobacteria being partially unclear, their function as protective elements against oxidative stress has been proposed47,48. Our findings thus further indicated that unfavorably high salinity induces oxidative stress and, together with other cellular stress responses, microcystins may play a role as protective molecules against oxidative stress.

**Conclusion**

The present study provides new insight into the genetic diversity of the Dolichospermum genus and explores the phylogenetic relationship of this toxic bloom-forming cyanobacterium with other members of the Anabaena and Aphanizomenon flos-aquae taxa. Despite the distinct morphological aspects of these cyanobacteria, the results support previous studies suggesting a demand for an explicit revision of these taxa. Beyond that, the study of Dolichospermum sp. UHCC 0315 genome revealed adaptive responses employed by the strain to explore different salinities. Future alteration in the Baltic Sea water due to climate change will possibly lead to an expansion of blooms dominated by Dolichospermum over Nodularia spumigena.

**Methods**

**Strains, cultivation, and toxin analysis.** Microcystin-producing Dolichospermum sp. strain UHCC 0315 (former name Anabaena sp. strain 315) is maintained at the University of Helsinki, Faculty of Agriculture and Forestry, Department of Microbiology Culture Collection (HAMBI, UHCC) under continuous illumination of 3.2–3.7-μmol photons m\(^{-2}\) s\(^{-1}\). The strain was isolated in August 1997 from a cyanobacterial bloom in the coastal Helsinki area, made axenic, and since isolation cultivated in Z8 medium without nitrogen49. Three replicates were cultivated in four different saline solutions (0, 3, 6, and 9 g L\(^{-1}\) NaCl) at 20°C under continuous illumination of 3.2–3.7-μmol photons m\(^{-2}\) s\(^{-1}\) for 24 d. Chlorophyll a was measured at 4-d intervals by filtering 1 mL of culture through 21-mm GF/C glass microfiber filters (GE HealthCare, Chicago, IL, USA). Chlorophyll a was extracted using 1 mL of 90% acetone, and absorbance was determined at 664, 647, and 630 nm and calculated as described earlier49. Cell counting was applied accordingly to Coloma et al.31. The combined intra- and extracellular concentrations of microcystins were determined using LC-MS as described earlier21. A standard curve containing a dilution series of known microcystin concentrations was run alongside the toxin samples.

**Genome sequencing and annotation.** A NucleoBond® anion-exchange (AXG) kit (Macherey-Nagel GmbH & Co. KG, Düren, Germany) was used to extract high-molecular-weight genomic DNA from the middle-logarithmic growth phase of Dolichospermum sp. UHCC 0315. The DNA libraries, PacBio RS II sequencing with P6-C4 chemistry, and genome assembly using Hierarchical Genome Assembly Process 3 (HGAP3) protocol with Quiver polishing52 were conducted in the DNA Sequencing and Genomics Laboratory, Institute of Biotechnology, University of Helsinki. Paired-end Illumina Hiseq2500 reads (Macrogen Inc., Seoul, South Korea) were mapped twice against the PacBio assembly, using Pilon v. 1.2.0 software and resulting in correction of 225 sites53. The genome of Dolichospermum sp. UHCC 0315 was annotated, using Prokka v. 1.1254 and noncoding RNA (ncRNA) by querying the RNA family (rfam) database35, using cme-Scan from Infernal46. An additional automatic annotation and functional classification in subsystems were performed using the default parameters in the RAST server47 and SEED viewer58. InterProScan 5 was used to polish the protein annotation, followed by manual curation in the genome browser Artemis59. Blast KEGG Orthology And Links Annotation (BlastKOALA) v. 2.160 was used to annotate the KEGG Orthology (KO) orthologs and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways. Clusters of Orthologous Groups (COG) orthology was annotated by querying the National Center for Biotechnology Information (NCBI) Conserved Domain Database (CDD)61 with an e-value cutoff of 1e-5, using Reversed Position Specific BLAST (RPSBLAST) 2.2.31. BLAST2GO63 was used to annotate the Gene Ontology (GO) term. The putative secondary metabolites and biosynthetic gene clusters were predicted, using the antiSMASH 3.0 online server64. The genome and plasmid sequences were scanned against the Clustered Regularly
Interspaced Short Palindromic Repeat database (CRISPRdb), using CRISPRfinder and CRISPRCasFinder to annotate the CRISPR loci. The Insertion Sequence (IS) elements were annotated, using the IS Semiautomatic Genomic Annotation (ISSaga) webserver.

To ensure two genetic substrains, a total of 70 single filaments of Dolichospermum UHCC 0315 were extracted on Z8X agar plates and further cultivated in liquid Z8X medium. DNA was extracted from three surface-growing and three bottom-growing cultures, using the Macherey-Nagel NucleoSpin kit. The PCR primers were designed to amplify the 300-bp region of the deleted gene BMF77_01254, as well as three regions in plasmids 1–3 (Supplementary Table S1). The PCR mixture contained 1 × DyNAzyme PCR buffer, 0.2 mM deoxynucleoside triphosphates (dNTPs), 0.5 µM primers, and 0.5 U DyNAzyme II polymerase in a total volume of 20 µL. As a template, 50–70 ng of genomic DNA was used. DNA from the original Dolichospermum sp. UHCC 0315 culture was used as a positive control and water as a negative control. The PCR program was 94 °C for 3 min; 30 cycles of 94 °C for 1 min, 56 °C for 1 min, and 72 °C for 1 min; and 72 °C for 10 min. PCR amplification products were visualized on 1% agarose gel electrophoresis.

Phylogenetic inference and comparative genomics. A maximum-likelihood phylogenomic tree was constructed by the concatenation of 31 universal marker genes. In brief, the profiles of these marker genes were scanned against the open reading frames (ORFs) of the 76 reference genome sequences and aligned using anvio v5.1.1 workflow for phylogenomics. The protein substitution model LG + I + G was assigned as the best one for the 31 protein sequences by performing BIC calculation in ProtTest v3.2 (default parameters). A maximum-likelihood tree was estimated with Randomized Axelerated Maximum Likelihood (RAxML) v. 8.2.10, based on selected partitions and substitution models, with 1000 rapid bootstrap searches. Glocobacter violaceus PCC 7421 was selected as an outgroup to root the phylogenomic tree. Average Amino Acid and Nucleotide Identity heatmaps were estimated using the average identity matrices protocol in the program GET_HOMOLOGUES and generated using a seaborn v0.9 library heatmap script.

The 16S rRNA gene tree was constructed by RAxML v8.0.0 with 1000 bootstrap replicate using 75 cyanobacterial strains. The genes were aligned using MEGA v10.0.5 (default parameters) and the evolutionary model GTR + I + G selected as best fitting by jModelTest v2.1.178.

Pan and core-genomes were estimated using the OrthoMCL algorithm and Tettelin function implemented in the program GET_HOMOLOGUES. OrthoVenn with default parameters was employed to generate the Venn diagram of Dolichospermum sp. UHCC 0315 and UHCC 0090. Shared and specific proteins were extracted using the server tools and automatically annotated by the default parameters in the metagenomics analysis server (MG-RAST). Pairwise genome alignments and dotplots were performed, using MUMmer v. 3.1.81. Synteny blocks were predicted and visualized, using the jcvi v0.7.1 Python library and local synteny blocks were visualized, using the genoPlotR v. 0.8.4 package.

RNA sequencing. On day 16, cells from the cultures containing 0 and 6 g L⁻¹ NaCl were collected on 0.22 µm polycarbonate filters (GE Water and Process Technologies). The filters were frozen in liquid nitrogen and stored at −80 °C. Prior to harvesting, the liquid cultures were fixed by 5% phenol in 10% ethanol. RNA was extracted, using the RNeasy mini kit (Qiagen N.V., Venlo, The Netherlands). For the depletion of genomic DNA, the TURBO DNA-free™ kit (Life Technologies, now ThermoFisher Scientific, Carlsbad, CA, USA) was used, and the RNA was further cleaned and concentrated with the RNA Clean & Concentrator™ kit (Zymo Research Corp., Irvine, CA, USA). RNA integrity and purity were determined, using the Bioanalyzer RNA 6000 Nano Kit (Agilent Technologies Inc., Santa Clara, CA, USA). Ribosomal RNA (rRNA) removal (MICROBExpress™) was carried out at the Institute for Molecular Medicine Finland (FIMM).

The quality of the raw reads was checked using FastQC v. 0.11.4 (Babrahan Bioinformatics, Babrahan Institute, Cambridge, UK) and the reads were trimmed with Trimmomatic85. The clean reads were aligned to the reference genomes, using Burrows-Wheeler Aligner—Maximum Exact Match (BWA-MEM) v. 0.7.7) in paired-end mode with default parameters. Sequence Alignment/Map tools (SAMtools) v. 1.277 was applied to convert the resulting SAM format to the Binary Alignment Map (BAM) format and filtered with the BAM filter (Galaxy Version 0.5.7.1) to remove the reads that were unmapped, less than 20 nucleotides long, flanked as secondary alignments, marked as PCR duplicates, or of low quality. The filtered BAM files were sorted with chromosomal coordinates using SAMtools v. 1.1997 and converted to Wiggle format, using the script bam2wig.py from the RNA Sequence Quality Control (RseQC) v. 2.4 package in paired-end mode, and normalized to 1000000000 wigsum.

The Wiggie format was then converted to Artemis v5.9-compatible genome coverage graphs, using custom scripts available at https://github.com/housw/GRPutils. FeatureCounts v. 1.4.6.986 was used to quantify unambiguously aligned fragments. Only paired-end reads that constituted a fragment size within 50–600 nucleotides long were considered, and chimeric fragments were excluded. The Galaxy instance of Freiburg University was used for RNA-Seq analysis.

To exclude weakly expressed features, the counts were converted to counts per million (CPM); only features having a higher CPM (sum of CPM across all samples should be more than 3 and at least three samples had a CPM more than 1) were taken for differential expression analysis. Features belonging to tRNA and transfer RNA (tRNA) were excluded in this study. The differentially expressed genes between control and treatment were called, using both edgeR v3.14.092 and DESeq v1.9.1225 following the simple design protocol. A GO enrichment test was performed for the differentially expressed genes, using GOstats v. 2.40.094 with all the expressed genes as background. Semantic similarities of the enriched GO terms were calculated, using the REVIGO online server.
The sequences were deposited in the NCBI database under the BioProject accession number PRJNA377208.

References
1. Zillén, L., Conley, D. J., Andrén, T., Andrén, E. & Björck, S. Past occurrences of hypoxia in the Baltic Sea and the role of climate variability, environmental change and human impact. Earth-Science Rev. 91, 77–92 (2008).
2. Kjellström, E. & Ruosteenoja, K. Present-day and future precipitation in the Baltic Sea region as simulated in a suite of regional climate models. Clim. Change 81, 281–291 (2007).
3. von Storch, H., Omstedt, A., Pawlak, J. & Reckermann, M. Introduction and summary. In Second assessment of climate change for the Baltic Sea basin (ed. The BACC II Author Team) 1–22 (Springer International Publishing AG 2015).
4. Graham, L. P. Climate change effects on river flow to the Baltic Sea. Ambio 38, 235–241 (2016).
5. Johannesson, K. & André, C. Life on the margin: genetic isolation and diversity loss in a peripheral marine ecosystem, the Baltic Sea. Mol. Ecol. 15, 2013–2029 (2006).
6. Herlemann, D. P. et al. Transitions in bacterial communities along the 2000 km salinity gradient of the Baltic Sea. ISME J. 5, 1571–1579 (2011).
7. Telefsen, I., Schubert, H. & Skarlato, S. Life in the salinity gradient: discovering mechanisms behind a new biodiversity pattern. Estuar. Coast. Shelf Sci. 135, 317–327 (2013).
8. Celepli, N., Sundh, J., Ekman, M., Dupont, C. L. & Youssef, S. Meta-omic analyses of Baltic Sea cyanobacteria: diversity, community structure and salt acclimation. Environ. Microbiol. 19, 1–30 (2017).
9. Halinen, K., Jokela, J., Fewer, D. P., Wåhlin, M. & Sivonen, K. Direct evidence for production of microcystins by Anabaena strains from the Baltic Sea. Appl. Environ. Microbiol. 73, 6543–6550 (2007).
10. Sivonen, K. et al. Molecular diversity and function in the Baltic Sea with an emphasis on cyanobacteria. Ambio 36, 180–187 (2007).
11. Fewer, D. P. et al. Culture-independent evidence for the persistent presence and genetic diversity of microcystin-producing Anabaena (Cyanobacteria) in the Gulf of Finland. Environ. Microbiol. 11, 855–866 (2009).
12. Sivonen, K. et al. Occurrence of the hepatotoxyc cyanobacterium Nodularia spumigena in the Baltic Sea and structure of the toxin. Appl. Environ. Microbiol. 55, 1990–1995 (1989).
13. Chen, Y., Shen, D. & Fang, D. Nodularins in poisoning. Clin. Chim. Acta 425, 18–29 (2013).
14. Laamanen, M. J., Muriel, F. G., Lehtimäki, J. M., Haukkia, K. & Sivonen, K. Diversity of toxic and nontoxic Nodularia isolates (Cyanobacteria) and filaments from the Baltic Sea. Appl. Environ. Microbiol. 67, 4638–4647 (2011).
15. Lyra, C., Laamanen, M., Lehtimäki, J. M., Surakka, A. & Sivonen, K. Benthic cyanobacteria of the genus Nodularia are non-toxic, without gas vacuoles, able to glide and genetically more diverse than planktonic Nodularia. Int. J. Syst. Bacteriol. 55, 555–568 (2005).
16. Ripppka, R., Iteman, I. & Herdman, M. Form – Anabaena in Bergey’s manual of systematics of archaea and bacteria (ed. Whitman, W. B. et al.), https://doi.org/10.1007/9781118960608.qbm00452 (John Wiley & Sons 2015).
17. Halinen, K. et al. Genetic diversity in strains of the genus Anabaena isolated from planktonic and benthic habitats of the Gulf of Finland (Baltic Sea). FEMS Microbiol. Ecol. 64, 199–208 (2008).
18. Vuorinen, L. et al. Scenario simulations of future salinity and ecological consequences in the Baltic Sea and adjacent North Sea areas–implications for environmental monitoring. Ecol. Indic. 50, 196–205 (2015).
19. Brutemark, A., Vandellanoote, A., Engström-Öst, J. & Suikkanen, S. A less saline Baltic Sea promotes cyanobacterial growth, hampers intracellular microcystin production, and leads to strain-specific differences in algalpathy. PLoS One 10, e0128904, https://doi.org/10.1371/journal.pone.0128904 (2015).
20. Wacklin, P., Hoffmann, L. & Komárek, J. Nomenclatural validation of the genetically revised cyanobacterial genus Dolichospermum (RALFS ex BORNET et FLAHAULT) comb. nova. Fottea 9, 59–64 (2009).
21. Lachance, M.-A. Genetic relatedness of heterocystous cyanobacteria by deoxyribonucleic acid-deoxyribonucleic acid reassociation. Int. J. Syst. Bacteriol. 31, 139–147 (1981).
22. Li, X., Dreher, T. W. & Li, R. An overview of diversity, occurrence and toxin production of bloom-forming Microcystis aeruginosa. Microb. Ecol. 64, 358, https://doi.org/10.1007/s00248-018-10035 (2018).
23. Driscoll, C. B. et al. A closely-related clade of globally distributed bloom-forming cyanobacteria within the Nostocales. Harmful Algae 77, 93–107 (2018).
24. Anders, S. & Huber, W. Differential expression analysis for sequence count data. Genome Biol. 11, R106, https://doi.org/10.1186/gb-2010-11-10-r106 (2010).
25. Castenholz, R. W. General characteristics of the cyanobacteria in Bergey’s manual of systematics of archaea and bacteria. (ed. Whitman, W. B. et al.), https://doi.org/10.1007/9781118960608.qbm00019 (John Wiley & Sons 2015).
26. Kaneko, T. et al. Complete genomic sequence of the filamentous nitrogen-fixing cyanobacterium Anabaena sp. strain PCC 7120. DNA Res. 31, 227–253 (2001).
27. Wang, H. et al. Genome-derived insights into the biology of the hepatotoxic bloom-forming cyanobacterium Anabaena sp. strain 90. BMC Genomics 13, 613, https://doi.org/10.1186/1471-2164-13-613 (2012).
28. Chu, Z., Jin, X., Yang, B. & Zeng, Q. Buoyancy regulation of Microcystis flos-aquae during phosphorus-limited and nitrogen-limited growth. J. Plankton Res. 29, 739–745 (2007).
29. Brookes, J. D. & Ganf, G. G. Variations in the buoyancy response of Microcystis aeruginosa to nitrogen, phosphorus and light. J. Plankton Res. 23, 1399–1411 (2001).
30. Humblet, J.-F. et al. A tribute to disorder in the genome of the bloom-forming freshwater cyanobacterium Microcystis aeruginosa. PLoS ONE 8, e70747, https://doi.org/10.1371/journal.pone.0070747 (2013).
31. Quas, J. et al. Integrated proteomic and transcriptomic analysis reveals novel genes and regulatory mechanisms involved in salt stress responses in Synechococcus sp. PCC 6803. Appl. Microbiol. Biotechnol. 97, 8253–8264 (2013).
32. Marin, K. et al. Gene expression profiling reflects physiological processes in salt acclimation of Synechocystis sp. strain PCC 6803. Plant. Physiol. 136, 300–329 (2004).
33. Hagemann, M. Molecular biology of cyanobacterial salt acclimation. FEMS Microbiol. Rev. 35, 87–123 (2011).
34. Khan, S. & Hagemann, M. Compatible solute biosynthesis in cyanobacteria. Environ. Microbiol. 13, 531–532 (2011).
35. Allahkhverdiev, S. I. & Murata, N. Salt stress inhibits photosystems II and I in cyanobacteria. Photosynth. Res. 98, 529–539 (2008).
36. Al-Hosani, S., Oudah, M. M., Henschel, A. & Yousef, L. F. Global transcriptome analysis of salt acclimated Prochlorococcus AS9601. Microbiol. Res. 176, 21–28 (2015).
37. Quas, J. et al. Integrated transcriptomic and proteomic analysis reveals novel genes and regulatory mechanisms involved in salt stress responses in Synechocystis sp. PCC 6803. Appl. Microbiol. Biotechnol. 97, 8253–8264 (2013).
85. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014).
86. Li, H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. Preprint at https://arxiv.org/abs/1303.3997 (2013).
87. Li, H. et al. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
88. Wang, L., Wang, S. & Li, W. RsqQC: quality control of RNA-seq experiments. *Bioinformatics* **28**, 2184–2185 (2012).
89. Liao, Y., Smyth, G. K. & Shi, W. featureCounts: an efficient general-purpose program for assigning sequence reads to genomic features. *Bioinformatics* **30**, 2214–2220 (2014).
90. Afgan, E. et al. The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2016 update. *Nucleic Acids Res.* **44**, W3–10, https://doi.org/10.1093/nar/gkw343 (2016).
91. Cock, P. J. A., Grünig, B. A., Passiakiewicz, K. & Pritchard, L. Galaxy tools and workflows for sequence analysis with applications in molecular plant pathology. *PeerJ* **1**, e167, https://doi.org/10.7717/peerj.167 (2013).
92. Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139–140 (2010).
93. Anders, S. et al. Count-based differential expression analysis of RNA sequencing data using R and Bioconductor. *Nat. Protoc.* **8**, 1765–1786 (2013).
94. Falcon, S. G. R. Using GOstats to test gene lists for GO term association. *Bioinformatics* **23**, 257–258 (2007).
95. Supek, F., Bošnjak, M., Škunca, N. & Šmuc, T. REVIGO summarizes and visualizes long lists of Gene Ontology terms. *PLoS One* **6**, e21800, https://doi.org/10.1371/journal.pone.0021800 (2011).

**Acknowledgements**
This work, resulting from the BONUS Blueprint project, was supported by BONUS (Art 185) and funded jointly by the E.U. and Academy of Finland to K.S. The bioinformatics analyses were supported by the German Federal Ministry for Education and Research (BMBF) program de.NBI-Partner (Grant 031 L0106B) to WRH and by a China Scholarship Council grant to S.W. R.V.P. was supported by an EDUFI fellowship of the Centre for International Mobility (CIMO). Moreover, the authors acknowledge the support of the Freiburg Galaxy Team, funded by BMBF grant 031 A538 A RBC (de.NBI). We thank L. Saari for maintenance of the cyanobacterial cultures and the personnel in the sequencing centers at the Institute of Biotechnology, University of Helsinki, FIMM, and Macrogen Ltd.

**Author Contributions**
J.E.T. and K.S. designed and executed the experiment. M.W. performed LC-MS analysis. R.V.P., J.E.T., S.H. and W.R.H. were responsible for the analysis of the data. All authors participated in the writing and in the discussion of the results.

**Additional Information**
Supplementary information accompanies this paper at https://doi.org/10.1038/s41598-019-40883-1.

**Competing Interests:** The authors declare no competing interests.

Publisher’s note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/.

© The Author(s) 2019