

Research article

Open Access

Recombination and selectional forces in cyanopeptolin NRPS operons from highly similar, but geographically remote *Planktothrix* strains

Trine B Rounge¹, Thomas Rohrlack^{2,4}, Tom Kristensen^{3,4} and Kjetill S Jakobsen*¹

Address: ¹University of Oslo, Department of Biology, Centre for Ecological and Evolutionary Synthesis (CEES), 0316 Oslo, Norway, ²NIVA, Norwegian Institute for Water Research, 0411 Oslo, Norway, ³University of Oslo, Department of Molecular Biosciences, 0316 Oslo, Norway and ⁴University of Oslo, Microbial Evolution Research Group (MERG), 0316 Oslo, Norway

Email: Trine B Rounge - t.b.rounge@bio.uio.no; Thomas Rohrlack - thomas.rohrlack@niva.no; Tom Kristensen - tom.kristensen@imbv.uio.no; Kjetill S Jakobsen* - k.s.jakobsen@bio.uio.no

* Corresponding author

Published: 26 August 2008

Received: 23 December 2007

BMC Microbiology 2008, 8:141 doi:10.1186/1471-2180-8-141

Accepted: 26 August 2008

This article is available from: <http://www.biomedcentral.com/1471-2180/8/141>

© 2008 Rounge et al; licensee BioMed Central Ltd.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/2.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Abstract

Background: Cyanopeptolins are nonribosomally produced heptapeptides showing a highly variable composition. The cyanopeptolin synthetase operon has previously been investigated in three strains from the genera *Microcystis*, *Planktothrix* and *Anabaena*. Cyanopeptolins are displaying protease inhibitor activity, but the biological function(s) is (are) unknown. Cyanopeptolin gene cluster variability and biological functions of the peptide variants are likely to be interconnected.

Results: We have investigated two cyanopeptolin gene clusters from highly similar, but geographically remote strains of the same genus. Sequencing of a nonribosomal peptide synthetase (NRPS) cyanopeptolin gene cluster from the Japanese strain *Planktothrix* NIES 205 (205-oci), showed the 30 kb gene cluster to be highly similar to the oci gene cluster previously described in *Planktothrix* NIVA CYA 116, isolated in Norway. Both operons contained seven NRPS modules, a sulfotransferase (S) and a glyceric acid loading (GA)-domain. Sequence analyses showed a high degree of conservation, except for the presence of an epimerase domain in NIES 205 and the regions around the epimerase, showing high substitution rates and Ka/Ks values above 1. The two strains produce almost identical cyanopeptolins, cyanopeptolin-1138 and oscillapeptin E respectively, but with slight differences regarding the production of minor cyanopeptolin variants. These variants may be the result of relaxed adenylation (A)-domain specificity in the nonribosomal enzyme complex. Other genetic markers (16S rRNA, *ntcA* and the phycocyanin *cpcBA* spacer) were identical, supporting that these geographically separated *Planktothrix* strains are closely related.

Conclusion: A horizontal gene transfer event resulting in exchange of a whole module-encoding region was observed. Nucleotide statistics indicate that both purifying selection and positive selection forces are operating on the gene cluster. The positive selection forces are acting within and around the epimerase insertion while purifying selection conserves the remaining (major) part of the gene cluster. The presence of an epimerase in the gene cluster is in line with the D-configuration of Htyr, determined experimentally in oscillapeptin E in a previous study.

Background

Cyanopeptolins are nonribosomally produced peptides with highly variable composition. The general structure of the cyanopeptolin peptide family encompasses 7 amino acids, including the residue 3-amino-6-hydroxy-2-piperidone (Ahp), where the six C-terminal amino acids form a ring [1,2] and the N-terminal amino acid frequently is N-modified. The N-terminal amino acid and all positions in the ring except position 2 (threonine) and position 4 (Ahp) can be occupied by variable amino acids, giving rise to a large number of cyanopeptolin variants [3].

The succession of the modules [4,5] and specificity of A-domain binding pockets in nonribosomal peptide synthetases (NRPSs) [6,7] can give a good prediction of peptide composition and structure. NRPSs do not always perform stringent substrate selection and incorporation [7], thus, relaxed substrate specificity is common in NRPS [6,8,9]. In addition to the common module domains including the adenylation (A)-, condensation (C)- and thiolation (T)-domains, several tailoring domains have been found associated with cyanopeptolin synthetases. Methyltransferases are present in three cyanopeptolin gene clusters from *Anabaena*, *Microcystis* and *Planktothrix* (*apd*, *mcn* and *oci*). Halogenases are found in *apd* and *mcn*, while the tailoring domains responsible for side chain modification of the N-terminal amino acid are unique for each strain (i.e.; formyl transferase in *apd*, sulfotransferase and glyceric acid (GA) transferase in *oci*, absent in *mcn*).

So far, only cyanopeptolin gene clusters derived from the genera *Anabaena* [10], *Microcystis* [11] and *Planktothrix* [12] have been characterized. They share the same basic domain structure but possess unique tailoring genes and A-domain substrate binding pockets, indicating independent evolution of cyanopeptolin genes within each lineage. Sequence identity is high (approximately 80% in the NRPS module coding regions) between *Microcystis* (*mcn*) and *Planktothrix* (*oci*) cyanopeptolin gene clusters. The more thoroughly investigated microcystin gene clusters show higher sequence identity within a genus than between genera. The same is likely to be the case also for the cyanopeptolin genes.

Sequence variation in microcystin synthetase clusters has been investigated within strains of the genera *Microcystis* [13,14] and *Planktothrix* [15]. Modifications and reorganizations due to several recombination events have been reported [14-16], and together with differences in substrate specificity between equivalent A-domains [17-19] are the reason for the different peptide variants.

Planktothrix NIVA CYA 116 (NIVA CYA 116), isolated from a Norwegian lake, produces cyanopeptolin-1138 [12] for which the amino acids configurations are

unknown. This peptide was found to be highly similar to oscillapeptin E produced by *Planktothrix* NIES 205 (NIES 205), isolated in Japan [20]. Both peptides have the same molecular mass, but slightly different polarities [12]. A different content of L-/D-amino acids in the peptides was suggested as a possible reason for the observed difference [12]. To investigate the genetic basis of the differences between the peptides, we have cloned and sequenced the NIES 205 cyanopeptolin gene cluster and compared it to the previously characterized NIVA CYA 116 gene cluster. This has allowed us to explore NRPS evolution and genetic variations in closely related strains and to investigate to what extent selectional forces operate on these gene clusters.

Results

NIVA CYA 116 and NIES 205 have similar but not identical peptide profiles

The major peptides in the two strains consist of HO₃SO-CH₂-CH(OMe)-CO-HTyr-Thr-HTyr-Ahp-Ile-Phe(Me)-Ile ([Table 1, Additional file 1 figure 1] and Rounge *et al* [12]). However, spiking experiments (data not shown) revealed a slight difference in polarity between cyanopeptolin-1138 from NIVA CYA 116 [12] and oscillapeptin E from NIES 205 [20]. In contrast to NIVA CYA 116 producing only cyanopeptolins, screening of NIES 205 shows production of additional peptide from other peptide-classes (data not shown).

Several cyanopeptolin variants were also detected in both strains. LC-MS-MS data identified minute amounts of seven cyanopeptolins in NIVA CYA 116, with variation in the first, third, fifth and/or seventh positions compared to cyanopeptolin-1138/oscillapeptin E [Additional file 1 figure 2]. An earlier study has shown that NIES 205 produce oscillapeptin C, D and E, based on spectroscopic analyses including 2D NMR [20]. Our LC-MS-MS analysis of NIES 205 confirmed the production of oscillapeptin D and E, but also identified a cyanopeptolin with the mass 1074, which is found in NIVA CYA 116 as well [Table 1 and Additional file 1 figure 3].

NIVA CYA 116 and NIES 205 produced similar – but not identical – cyanopeptolin variants. The identified NIVA CYA 116 cyanopeptolins were mainly combinations of Hty/Ile/Leu in positions AA1 and AA3 and Ile/Leu/Val in positions AA5 and AA7. Other unidentified apolar amino acid-like residues were detected in position AA3. In contrast, the only variations observed in the NIES 205 peptides were Hty, Ile/Leu and HcAla in position AA3 (Table 1).

Comparison of the 205-oci and 116-oci gene clusters

Anticipating that two strains producing almost identical cyanopeptolins also should contain similar gene clusters,

Table 1: Oci A-domains binding pockets and peptide profiles

Binding pockets	OciA-A1	OciA-A2	OciB-A3	OciB-A4	OciB-A5	OciB-A6	OciC-A7
NIVA CYA 116	DLGFTGAVCK	DFWNIGMVHK	DAQSMGAIK	DVENAGVVTK	DAFFLGVTFK	DAWTIAGVCK	DAFFLGVTFK
NIES 205	DLGFTGAVCK	DFWNIGMVHK	DAEGMGAIK	DVENAGVVTK	DAFFLGVTFK	DAWTIAGVCK	DAFFLGVTFK

NIVA CYA 116								
Mass Da	Side chain	AA 1	AA2	AA3	AA4	AA5	AA6	AA7
1138	HO ₃ -SO-CH ₂ -CH(OMe)-COH	HTyr	Thr	HTyr	Ahp	Ile	Phe(Me)	Ile
1124	HO ₃ -SO-CH ₂ -CH(OMe)-COH	HTyr	Thr	HTyr	Ahp	Val	Phe(Me)	Ile
1124	HO ₃ -SO-CH ₂ -CH(OMe)-COH	HTyr	Thr	HTyr	Ahp	Ile	Phe(Me)	Val
1074	HO ₃ -SO-CH ₂ -CH(OMe)-COH	HTyr	Thr	Ile/Leu	Ahp	Ile	Phe(Me)	Ile
1010	HO ₃ -SO-CH ₂ -CH(OMe)-COH	Ile/Leu	Thr	Ile/Leu	Ahp	Ile	Phe(Me)	Ile
1088	HO ₃ -SO-CH ₂ -CH(OMe)-COH	HTyr	Thr	X	Ahp	Ile	Phe(Me)	Ile
1122	HO ₃ -SO-CH ₂ -CH(OMe)-COH	HTyr	Thr	Y	Ahp	Ile	Phe(Me)	Ile

NIES 205								
Mass Da	Side chain	AA 1	AA2	AA3	AA4	AA5	AA6	AA7
1138*	HO ₃ -SO-CH ₂ -CH(OMe)-COH	Htyr	Thr	Htyr	Ahp	Ile	Phe(Me)	Ile
1074	HO ₃ -SO-CH ₂ -CH(OMe)-COH	Htyr	Thr	Ile/Leu	Ahp	Ile	Phe(Me)	Ile
1128**	HO ₃ -SO-CH ₂ -CH(OMe)-COH	Htyr	Thr	HcAla	Ahp	Ile	Phe(Me)	Ile

The binding pocket residues of the NIVA CYA 116 and NIES 205 A-domains were identified by comparison to the *GrsA-Phe* A-domain (Residue 235, 236, 239, 278, 299, 301, 322, 330, 331, 517). The Oci-A3 binding pockets are different (in grey) between the two strains, and the divergent amino acids are shown in bold. The composition of cyanopeptolins produced by NIVA CYA 116 and NIES 205 and their molecular weights (M+H⁺) are shown with amino acids correlated with the putative binding pockets. HcAla is 3-(4'-hydroxy-2'-cyclohexenyl) alanine, and X and Y is unidentified amino acid derivatives. Mutual peptides in NIVA CYA 116 and NIES 205 are highlighted in dark grey and light grey. See [Additional file 1 figure 1, 2 and 3 for the peptide structure and more details on MS data.* peptide named oscillapeptin E and **peptide named oscillapeptin D

we sequenced a cyanopeptolin gene cluster in NIES 205 (205-*oci*) using primers designed for the cyanopeptolin (*oci*) gene cluster in NIVA CYA 116 (116-*oci*) [12]. The two gene clusters, including the ABC transporter genes and the intergenic spacers, were highly similar (93% identity between the nucleotide sequences), and the domain structures of the encoded synthetases were almost identical; except that 205-*oci* contained an epimerase encoding (E)-domain between T2 and C2 (Figure 1). The position of the E-domain corresponds to the Htyr in D-configuration in oscillapeptin E determined by Itou *et al* [20]. Both gene clusters included a GA-domain and a sulfotransferase domain. Comparison with cyanopeptolin gene clusters characterized in *Microcystis* (*mcn*) [11] and *Anabaena* (*apd*) [10] (Figure 1) showed a higher degree of similarity within the *Planktothrix* genus than between genera (70% identity between *OciB* and *AdpB* with the additional methyltransferase excluded and 77% identity between *OciC* and *AdpD*). A-domains and A-domain binding pockets signatures were identified from the gene clusters and aligned. The binding pocket signatures in 116-*Oci* and the corresponding 205-*oci* signatures were identical, except for 116-*OciB*-A3 (DAQSMGAIK) and 205-*OciB*-A3 (DAEGMGAIK) (Table 1). Corresponding pairs of

205-*Oci* and 116-*Oci* A-domains clustered together in phylogenetic analyses that included A-domains from cyanopeptolin [10-12], microcystin [17,21,22] nostocyclopeptide [23] and nostopeptolide [24] synthetases [Additional file 1 figure 4].

E-domains are common in cyanobacterial NRPS, found in microcystin, aeruginosin and nostocyclopeptide synthetases, notably, E-domains have until now not been found in cyanopeptolin synthetases. The E-domain produces the D-isomer of the amino acid activated by the upstream A-domain and is also involved in the stereospecific selection of the D-isomer for incorporation in the peptide product. Most E-domains are flanked by T (T_E)- and C-domains with special motives [25,26], and this was the case also in 205-*Oci* – as shown by the phylogenetic analyses (see Figure 2).

The NIES 205-E-domain is localized downstream of 205-A1 and T2. A phylogenetic analysis of E-domains (Figure 2), including E-domains from microcystin synthetase (*McyA-E*) and nostocyclopeptide synthetase (*NcpA-E*), showed a close relationship between *NcpA-E* and 205-

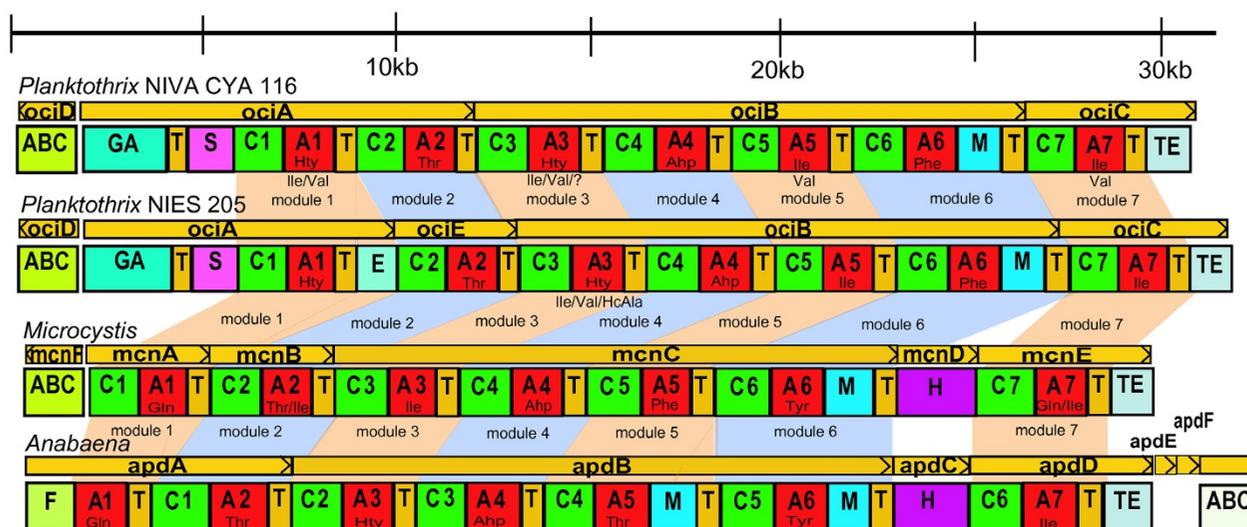


Figure 1
Comparison of the known cyanopeptolin operons. The overall structure of cyanopeptolin operons *oci* from *Planktothrix* NIES 205 [GenBank: EU109504] and NIVA CYA 116 [GenBank: DO837301], *mcn* from *Microcystis* [GenBank: DO075244] and *apd* from *Anabaena* [GenBank: AJ269505]. Gene names, transcription directions and approximate sizes are indicated above each gene cluster. Adenylation (red), condensation (green), thiolation (yellow), epimerization (turquoise), methyltransferase (blue) sulfotransferase (pink), halogenisation (purple) and termination domains (grey) are shown with their abbreviations. The putative activated amino acids are indicated for each A-domain. Amino acids detected in smaller amounts are beneath the major amino acid. Equivalent modules are depicted in light blue and light orange. The ABC transporter is transcribed in the opposite direction in the *oci* and *mcn* operons, and an ABC transporter is predicted downstream of the *apd* operon.

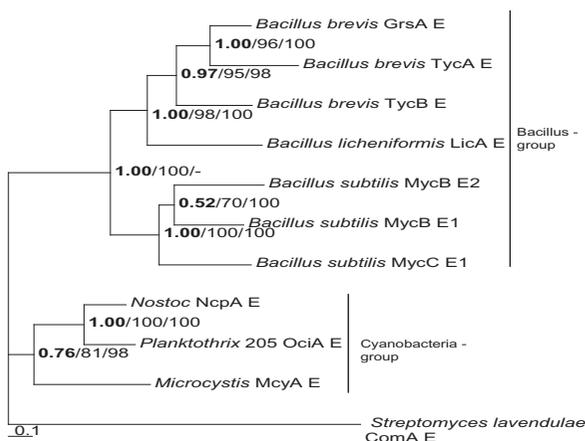


Figure 2
Phylogenetic analyses of E-domains. The E-domain phylogenetic tree was constructed utilizing MrBayes 3.1., WAg protein substitution model and gamma-shaped distribution. In addition, the bootstrap obtained for NJ (MEGA 3.1) at default settings and ML (RAXML) trees are indicated. Only posterior probability values and bootstrap replica values above 50% (out of 1000 (NJ) and 100 (ML) trees) are shown.

OciA-E (72% identity on the DNA and 67% similarity on the protein level).

Phylogenetic analyses of the C-domains (Figure 3), including domains from cyanopeptolin [10-12], microcystin [17,21,22] nostocyclopeptide [23] and nostopeptolide [24] synthetases, clustered according to presence or absence of an upstream E-domain. The 205-Oci-C2-domain grouped with D-amino acid-specific C-domains, while the other 205-Oci-C domains formed a clade with the corresponding 116-Oci-C-domains.

The specialized T_E -domains associated with E-domains, show major differences within the core T motif compared to standard T-domains [25]. Comparisons of regular T-domains and T_E -domains, including 205-Oci- T_E2 , McnA-T1 and NcpA- T_E1 -domain, showed an H/D and L/I difference in addition to a gap in the T_E -domain motif [Additional file 1 figure 5]. N-terminal T-domains, including both 116-T1 and 205-T1, could also be distinguished from T_E -domains and regular T-domains [Additional file 1 figure 5].

Other genomic regions confirm a close relationship between the *Planktothrix* strains

Several markers were sequenced to further study the relationship between *Planktothrix* NIVA CYA 116 and NIES 205. The DNA sequences (16S rDNA (1357 bp), a part of *ntcA* (384 bp), a global transcriptional regulator of nitro-

gen assimilation in cyanobacteria, and the phycocyanin spacer *cpcBA*) displayed 100% identity between the two strains.

Variation in substitution rates throughout the cyanopeptolin gene clusters

Investigation of the substitution rates within the 30 kb 116- and 205-*oci* gene cluster alignment can identify both putative recombination events and regions under specific selection pressure. The region containing the epimerization domain (T2, E, C2) was excluded due to too large overall differences to produce a reliable alignment. Figure 4 shows segregating sites (black lines) and nonsynonymous vs. synonymous substitution rates (red lines) in a sliding window analysis of the alignment. Only a few scattered substitutions can be seen in the first part, containing the ABC transporter, GA, T1, S and C1 domains, and in the last part, containing A6, M, T7, C7, A7, T8 and TE domains. However, the C3 and A3 domains contained several substitutions and the rate of mutations in nonsynonymous sites compared with synonymous sites (K_a/K_s) exceeded 1 – a putative sign of positive selection. A high substitution rate was also observed in a small region in C6 and the last part of A1, but the K_a/K_s ratios did not exceed 1.

Discussion

Correlation between cyanopeptolin gene clusters and peptides

The presence of two highly similar NRPS gene clusters (*oci*) in NIVA CYA 116 and NIES 205, and the production of nearly identical peptides by the two strains corroborate the association between the *oci* gene cluster and cyanopeptolin-1138 proposed by Rounge *et al* [12]. This association is further substantiated by high degree of similarity to the cyanopeptolin gene cluster in *Anabaena* (*apd*), where the functional relationship between genes and peptides has been confirmed by a gene knock-out study [10] – as well as similarity to the *Microcystis* cyanopeptolin gene cluster (*mcn*) [11].

Global dispersal and distribution of cyanopeptolin genes

Based on the genomic regions studied here, two *Planktothrix* strains, NIVA CYA 116 and NIES 205, appear to be closely related despite the geographical separation. This is in accordance with the sequence comparison of 16S rDNA [27] identifying identical 16S rDNA sequences in Japan, China, The Netherlands, UK, Finland, Sweden and Norway, and thus may indicate a global distribution of closely related *Planktothrix* strains. Since Lake Årungen in Norway host international rowing competitions, a co-transport of this *Planktothrix* genotype with rowing equipment may be feasible. The data presented here do not allow any conclusions about global distribution without a more thorough analysis. The highly specific differences observed in the *oci*

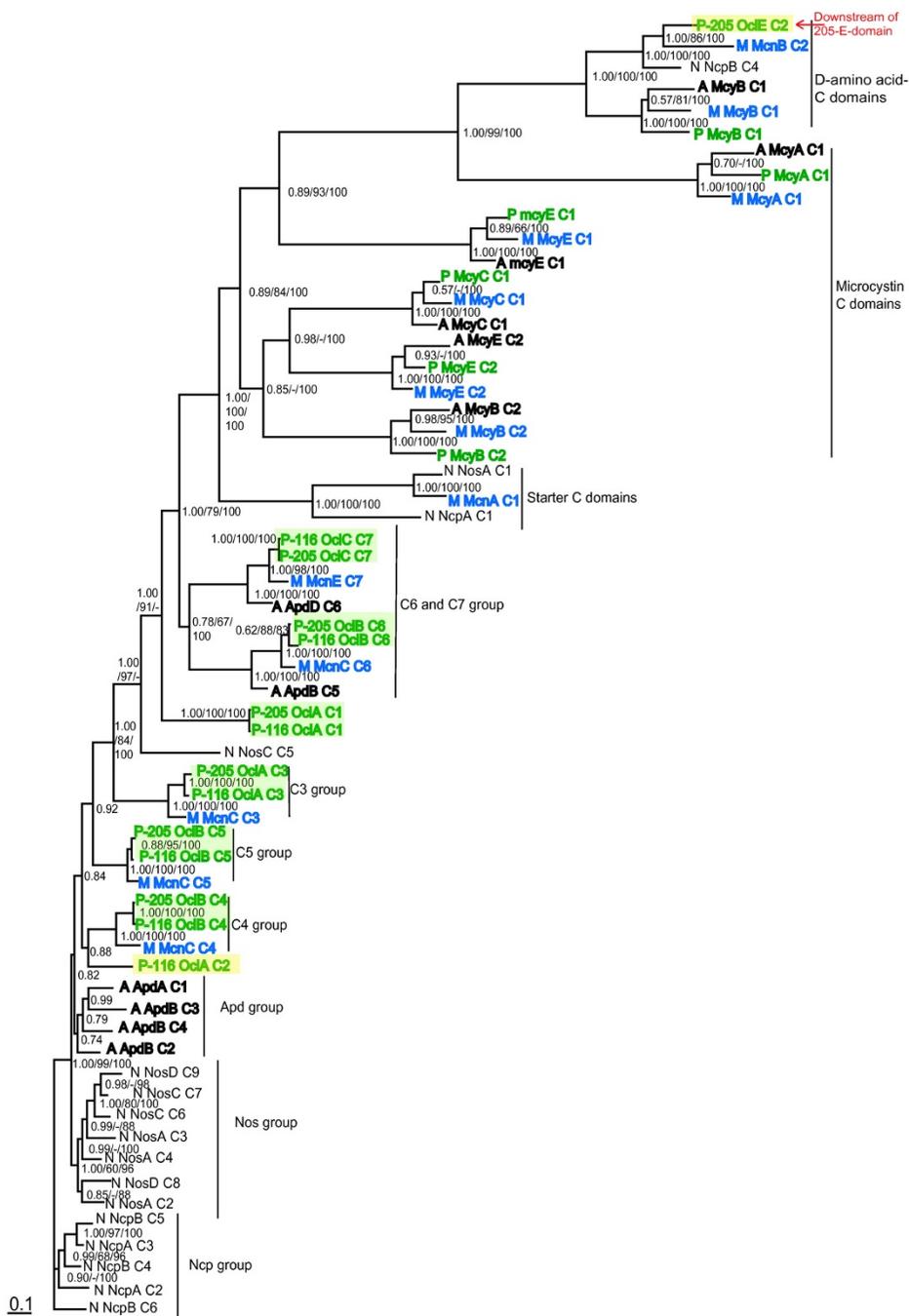


Figure 3
Phylogenetic analyses of C-domains showing groups according to gene cluster and position/function. The C-domain phylogeny was constructed using Bayesian inference with gamma distribution, 4 mill generations tree sampling every 100 generations and removal of the first 3000. The topologies generated using NJ (MEGA 3.1) and ML (RAxML) analyses show near identical branching patterns-only minor differences are seen within the Apd group. Bayesian posterior probability, NJ (1000 bootstrap values) and ML (100 trees) above 50% are shown. CpRev protein substitution model was used in the Bayesian and ML analyses. Genus origin is shown with first letter abbreviations (P = *Planktothrix*, M = *Microcystis*, A = *Anabaena* and N = *Nostoc*), and the C-domains are labeled in numerical order according to direction of transcription (i.e. seven *oci*, seven *mcn* and six *apd* C-domains). Corresponding *Oci* C domains group together, except for C2 situated downstream of the 205-E-domain. C1–C4 *apd*, *nos* and *ncp* C-domains do not group according to function

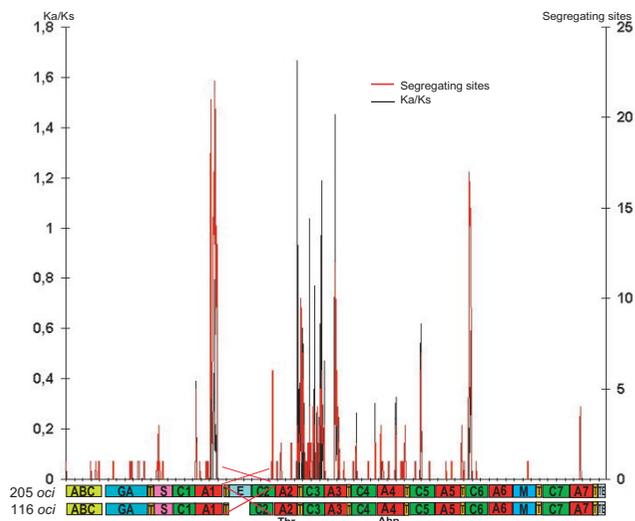


Figure 4
Distribution of segregating sites and Ka/Ks ratios in the *oci* gene cluster. The ratios are displayed using the program DnaSP and sliding windows analysis on the alignment of 205-*oci* and 116-*oci*. Window length was 50 bp and step size 10 bp. The distribution of segregation sites (red) and Ka/Ks (black) ratios are shown in correlation with the domain alignment. Module 2 (T2-(E)-C2) has been excluded from the analyses.

gene clusters are, independently of geographic distributions, intriguing. Our analyses indicate that the differences to some extent are due to positive selection at specific amino acid positions.

Variation in peptide content due to lack of specificity in the A-domains?

Previous studies have shown that lack of specificity in A-domains leads to activation of several amino acids with similar properties, thus giving rise to the synthesis of a series of related peptides from a single NRPS system [28]. Ile/Leu/Val activating A-domains have been reported in lichenysin biosynthesis [8], and fengycin synthetase [29] among others. It is likely that the 116-Oci-A5- and A7-domains can activate Leu, Ile and Val and that the 116-Oci-A1- and A3-domains, that mainly activates Htyr, also can activate Ile and Leu. Consequently, 116-Oci is responsible for production of all seven cyanopeptolin detected in NIVA CYA 116 in this study. Likewise, 205-Oci probably is responsible for all oscillapeptin variants. The biological significance of a single NRPS complex giving rise to several peptide variants is yet to be determined.

Six of the seven binding pockets signatures of corresponding A-domains in NIES 205 and NIVA CYA 116 are identical (Table 1). If the different peptide profiles observed in the two strains are due to genetic differences in the NRPS genes, they are likely to be due to differences not involving the amino acids constituting the binding pocket signatures. LC-MS-MS-analyses were performed on strains cultivated on the same media, but we cannot completely exclude substrate availability as a contributory cause of variable peptide amount and peptide profile in the strain.

Module exchange and amino acid configuration

Over a stretch of total of 30 kb including the ABC transporter, the 116-*oci* and 205-*oci* gene clusters are remarkable similar, except for the modules encoding the T2-(E)-C2 domains. Too low sequence similarity is found between the whole T2-(E)-C2 modules in NIVA CYA 116 and NIES 205 to make a reliable alignment, suggesting that in one of these strains an entire module may have been exchanged through recombination. The E-domain trees (Figure 2) show a close relationship between cyanobacterial E-domains.

Sequence similarity to other E-domains and the distinctive flanking C (Figure 3) and T [Additional file 1 figure 5] domains observed by phylogenetic analysis indicate that the Oci-E-domain is an active epimerase, and are responsible for epimerization of Htyr to D-configuration. The configuration of the amino acids in cyanopeptolin-1138 were not determined however, a D-Htyr in oscillapeptin E and a putative L-Htyr in cyanopeptolin-1138 might explain the small difference between the oligopeptides with regard to polarity observed by HPLC analysis, as reported by Rounge *et al.* [12].

Interestingly, in the corresponding region of the Mcn cyanopeptolin synthetase in *Microcystis* the McnA-T1 and McnB-C2 include motifs suggesting association with an E-domain [11]. In this case, however, no E-domain is present.

Sequence conservation and selection within cyanopeptolin modules

The two cyanopeptolin gene clusters (205-*oci* and 116-*oci*) are highly similar also at the third codon position. The first part (ABC-transporter, the spacer, GA-, T1-, S-, and C1-domains) and last part (C4-, A4-, T5-, C5-, A5-, T6-, C6-, M-, T7-, C7-, A7-, T8- and TE domains) of the *Planktothrix* cyanopeptolin gene cluster are nearly identical, despite the geographical distance separating the strains. Mechanisms for such sequence conservation may be frequent homology-driven genetic exchange within a genotype, leading to homogenization – in line with the general models suggested by Rudi *et al.* [30], Gogarten *et al.* [31] and Papke *et al.* [32]. Or alternatively sequence conserva-

tion may be due to low evolutionary rates caused by purifying selection or very short time of independent evolution.

Analysis of segregating sites and rates of nonsynonymous and synonymous nucleotide substitutions (Ka/Ks) indicate that module 3 (T3-, C3- and A3-domains) is different from the remaining domains by displaying higher substitution rates and signs of positive selection at several sites (Ka/Ks higher than 1). This is the module responsible for incorporation of the amino acid at position AA3 in the peptide.

According to data from Itou *et al* [20], a single amino acid replacement in the AA3 position of oscillapeptin E and F alters the protease inhibitory profile, indicating that this position could be pivotal for the inhibitory activity of cyanopeptolins. Positive selection in the third module could thus be expected to increase the adaptability of the inhibitory- or other putative functions of cyanopeptolin.

Conclusion

The *Planktothrix* strains of Japan and Norway harbor almost identical cyanopeptolin gene clusters and display very similar (but not identical) cyanopeptolin profiles. The notable gene cluster difference is the presence of an epimerase in NIES 205 corresponding to a D-Htyr in oscillapeptin E. Within a single gene cluster we have demonstrated both positive selection and purifying selection, the first promoting new gene cluster variants following recombination, the latter maintaining a high degree of conservation of the major parts of the gene cluster.

Methods

Bacterial cultures

Planktothrix agardhii NIVA CYA 116 was isolated in 1983 from Lake Årungen, Norway, and maintained in the NIVA culture collection of Algae. *Planktothrix agardhii* NIES 205 was isolated from Lake Kasumigaura/Ibaraki, Japan in 1982, and maintained in the NIES culture collection [20]. Both strains were cultured in Z8 [33] media at ~20°C with 12 hour illumination at about 15 μmol m⁻² s⁻¹ in Sanyo versatile environmental test chamber (FG-4P 36–40).

PCR and sequencing

DNA from NIES 205 was isolated utilizing Dynabeads (Invitrogen, Carlsbad, USA) [34]. Combinations of PCR primers designed for the cyanopeptolin (*oci*) gene cluster in NIVA CYA 116 [12] were used to amplify regions of a cyanopeptolin gene cluster in NIES 205. These PCR products were sequenced using primer walking. Additional PCR primers were designed to amplify regions between already obtained PCR products. BD Advantage 2 (BD Biosciences, Mountain View, USA) was utilized as polymerase in all PCR amplifications. The PCR products were

sequenced using an ABI 3730 sequencer and v3.1 Big Dye solution.

Sequence analysis and phylogeny

Open reading frames were identified and translated using Vector NTI (Invitrogen, Carlsbad, USA). Domains and their boundaries were identified using the NRPS database <http://www.nii.res.in/nrps-pks.html> [35], A-domain binding pocket residues identified by aligning the sequences with the GrsA-Phe A-domain [6] and substrate specificity predicted utilizing the NRPS database <http://www.nii.res.in/nrps-pks.html> and phylogenetic analysis. A-, C-, T- and E- domain protein sequences were aligned using MEGA 3.1 and Neighbor-Joining (NJ) trees were constructed using MEGA 3.1 at default settings (Poisson correction as the amino acid substitution model) [36]. Optimal protein evolution model was found by ProtTest [37]. Trees were constructed utilizing MrBayes [38] 3.0 and 3.1 [39] on the Uio Biportal <http://www.biportal.uio.no> [40] with an optimal protein substitution model. Variable substitution rates across sites were accounted for by gamma distribution. The MCMC chains were carried out for 4 million generations and trees were sampled every 100 generations, removing 3000 trees before the MCMC chain reached convergence. In addition, maximum likelihood inferences with RAxML [41] were performed on the E- and C- domain alignments. Similarity calculations were done in Vector NTI. DnaSP [42] was used to calculate Ka/Ks ratio and segregating sites with a sliding window with window length of 50 bp and step size 10 bp.

Mass spectrometry

Freeze-dried material of NIVA-CYA116 and NIES 205 was extracted with 50% MeOH (MeOH:water, v/v) and the extracts were subjected to a screening for cyanopeptolins by LC-MS. The instrument included a Waters Acquity UPLC system equipped with an Atlantis column (C18 2.1 × 150 mm, 5 μm particle size) and set to run a linear gradient starting with 80% solvent A (10 mM ammonium acetate, 0.1% acetic acid) and ending with 60% solvent A after 15 min. Solvent B was MeOH with 0.1% acetic acid. The flow rate was 0.2 ml min⁻¹. The LC system was connected to a Waters Quattro Premier XE tandem quadrupole mass spectrometer equipped with an electrospray probe. The detector was run in the positive ion mode at a cone voltage of 50 V. A total ion scan from 600 to 1400 Da was performed during the entire length of the LC gradient.

The structures of putative cyanopeptolins were analyzed by MS fragmentation studies. MS fragments hold valuable structural information and have been successfully used before to identify and structurally elucidate cyanobacterial oligopeptides including cyanopeptolins [43-45]. Frag-

mentation experiments were carried out with the hardware configuration described above. The mass spectrometer was run in daughter ion scanning mode and all settings were automatically optimized for fragmentation at 30 eV. Fragments were recorded during the entire length of the LC gradient. The identification of fragments was assisted by the HighChemMassFrontier software version 3.0. This software predicts MS fragmentation patterns on the basis of a putative structure. Comparing predicted and actual fragmentation patterns was used to assess the accuracy of a putative structure. Further hints to the structure were obtained from the occurrence of typical diagnostic ions such as immonium ions and from predictions on the amino acid occurrence made by the genetic analyses.

Authors' contributions

This work was performed as part of the PhD thesis for TBR. TBR and TR carried out all experimentation and all authors have contributed to the experimental and analytical design. TBR performed the bioinformatics and phylogenetic analysis under supervision of KSJ and TK. TR carried out the peptide analyses. TBR, KSJ (thesis advisor) TK and TR wrote the ms. All authors have read and approved the final manuscript.

Additional material

Additional file 1

Peptide structure and NRPS phylogeny. Figure 1: peptide structure of cyanopeptolin 1138. Figure 2: mass spectrometric fragmentation experiments data of NIVA CYA 116. Figure 3: mass spectrometric fragmentation experiments data of NIES 205. Figure 4: NRPS A-domain phylogeny. Figure 5: Sequence analyses of T-domains. Table 1: Accession numbers

Click here for file
[http://www.biomedcentral.com/content/supplementary/1471-2180-8-141-S1.doc]

Acknowledgements

We are grateful to Ann-Kristin Hansen and Beatriz Decenciere for excellent technical assistance, Ave Tooming-Klunderud for helpful discussions and Randi Skulberg at NIVA and The Microbial Culture Collection at the National Institute for Environmental Studies (NIES) for providing the *Planktothrix* NIES 205 strain. The work was supported by a grant to the project I57338/140 from the Norwegian Research Council.

References

- Weckesser J, Martin C, Jakobi C: **Cyanopeptolins, depsipeptides from cyanobacteria.** *Syst Appl Microbiol* 1996, **19**:133-138.
- Namikoshi M, Rinehart KL: **Bioactive compounds produced by cyanobacteria.** *J Ind Microbiol* 1996, **17**:373-384.
- Welker M, von Döhren H: **Cyanobacterial peptides – nature's own combinatorial biosynthesis.** *FEMS Microbiol Rev* 2006, **30**:530-563.
- Marahiel MA, Stachelhaus T, Mootz HD: **Modular Peptide Synthetases Involved in Nonribosomal Peptide Synthesis.** *Chem Rev* 1997, **97**:2651-2674.
- Stachelhaus T, Marahiel MA: **Modular Structure of Genes Encoding Multifunctional Peptide Synthetases Required for Nonribosomal Peptide-Synthesis.** *FEMS Microbiol Lett* 1995, **125**:3-14.
- Challis GL, Ravel J, Townsend CA: **Predictive, structure-based model of amino acid recognition by nonribosomal peptide synthetase adenylation domains.** *Chem Biol* 2000, **7**:211-224.
- Stachelhaus T, Mootz HD, Marahiel MA: **The specificity-conferring code of adenylation domains in nonribosomal peptide synthetases.** *Chem Biol* 1999, **6**:493-505.
- Konz D, Doekel S, Marahiel MA: **Molecular and biochemical characterization of the protein template controlling biosynthesis of the lipopeptide lichenysin.** *J Bacteriol* 1999, **181**:133-140.
- von Döhren H, Keller U, Vater J, Zocher R: **Multifunctional Peptide Synthetases.** *Chem Rev* 1997, **97**:2675-2706.
- Rouhiainen L, Paulin L, Suomalainen S, Hyytiäinen H, Buikema W, Haselkorn R, Sivonen K: **Genes encoding synthetases of cyclic depsipeptides, anabaenopeptilides, in *Anabaena* strain 90.** *Mol Microbiol* 2000, **37**:156-167.
- Tooming-Klunderud A, Rohrlack T, Shalchian-Tabrizi K, Kristensen T, Jakobsen KS: **Structural analysis of a non-ribosomal halogenated cyclic peptide and its putative operon from *Microcystis*: implications for evolution of cyanopeptolins.** *Microbiology* 2007, **153**:1382-1393.
- Rounge TB, Rohrlack T, Tooming-Klunderud A, Kristensen T, Jakobsen KS: **Comparison of cyanopeptolin genes in *Planktothrix*, *Microcystis*, and *Anabaena* strains: evidence for independent evolution within each genus.** *Appl Environ Microbiol* 2007, **73**:7322-7330.
- Kurmayer R, Dittmann E, Fastner J, Chorus I: **Diversity of microcystin genes within a population of the toxic cyanobacterium *Microcystis* spp. in Lake Wannsee (Berlin, Germany).** *Microb Ecol* 2002, **43**:107-118.
- Mikalsen B, Boison G, Skulberg OM, Fastner J, Davies W, Gabrielsen TM, Rudi K, Jakobsen KS: **Natural variation in the microcystin synthetase operon *mcvABC* and impact on microcystin production in *Microcystis* strains.** *J Bacteriol* 2003, **185**:2774-2785.
- Kurmayer R, Christiansen G, Gumpenberger M, Fastner J: **Genetic identification of microcystin ecotypes in toxic cyanobacteria of the genus *Planktothrix*.** *Microbiology* 2005, **151**:1525-1533.
- Tanabe Y, Kaya K, Watanabe MM: **Evidence for recombination in the microcystin synthetase (*mcv*) genes of toxic cyanobacteria *Microcystis* spp.** *J Molecular Evol* 2004, **58**:633-641.
- Christiansen G, Fastner J, Erhard M, Börner T, Dittmann E: **Microcystin biosynthesis in *Planktothrix*: Genes, evolution, and manipulation.** *J Bacteriol* 2003, **185**:564-572.
- Moffitt MC, Neilan BA: **Characterization of the nodularin synthetase gene cluster and proposed theory of the evolution of cyanobacterial hepatotoxins.** *Appl Environ Microbiol* 2004, **70**:6353-6362.
- Kurmayer R, Gumpenberger M: **Diversity of microcystin genotypes among populations of the filamentous cyanobacteria *Planktothrix rubescens* and *Planktothrix agardhii*.** *Mol Ecol* 2006, **15**:3849-3861.
- Itou Y, Ishida K, Shin SJ, Murakami M: **Oscillapeptins A to F, serine protease inhibitors from the three strains of *Oscillatoria agardhii*.** *Tetrahedron* 1999, **55**:6871-6882.
- Tillett D, Dittmann E, Erhard M, von Döhren H, Börner T, Neilan BA: **Structural organization of microcystin biosynthesis in *Microcystis aeruginosa* PCC7806: an integrated peptide-polyketide synthetase system.** *Chem Biol* 2000, **7**:753-764.
- Rouhiainen L, Vakkilainen T, Siemer BL, Buikema W, Haselkorn R, Sivonen K: **Genes coding for hepatotoxic heptapeptides (microcystins) in the cyanobacterium *Anabaena* strain 90.** *Appl Environ Microbiol* 2004, **70**:686-692.
- Becker JE, Moore RE, Moore BS: **Cloning, sequencing, and biochemical characterization of the nostocyclopeptide biosynthetic gene cluster: molecular basis for imine macrocyclization.** *Gene* 2004, **325**:35-42.
- Hoffmann D, Hevel JM, Moore RE, Moore BS: **Sequence analysis and biochemical characterization of the nostopeptolide A biosynthetic gene cluster from *Nostoc* sp. GSV224.** *Gene* 2003, **311**:171-180.
- Linne U, Doekel S, Marahiel MA: **Portability of epimerization domain and role of peptidyl carrier protein on epimerization**

- activity in nonribosomal peptide synthetases. *Biochemistry* 2001, **40**:15824-15834.
26. Roongsawang N, Lim SP, Washio K, Takano K, Kanaya S, Morikawa M: **Phylogenetic analysis of condensation domains in the non-ribosomal peptide synthetases.** *FEMS Microbiol Lett* 2005, **252**:143-151.
 27. Suda S, Watanabe MM, Otsuka S, Mahakahant A, Yongmanitchai W, Nopartnaraporn N, Liu Y, Day JG: **Taxonomic revision of water-bloom-forming species of oscillatorioid cyanobacteria.** *Int J Syst Evol Microbiol* 2002, **52**:1577-1595.
 28. Sieber SA, Marahiel MA: **Molecular mechanisms underlying nonribosomal peptide synthesis: approaches to new antibiotics.** *Chem Rev* 2005, **105**:715-738.
 29. Lin TP, Chen CL, Chang LK, Tschen JSM, Liu ST: **Functional and transcriptional analyses of a fengycin synthetase gene, *fen C*, from *Bacillus subtilis*.** *J Bacteriol* 1999, **181**:5060-5067.
 30. Rudi K, Skulberg OM, Jakobsen KS: **Evolution of cyanobacteria by exchange of genetic material among phyletically related strains.** *J Bacteriol* 1998, **180**:3453-3461.
 31. Gogarten JP, Doolittle WF, Lawrence JG: **Prokaryotic evolution in light of gene transfer.** *Mol Biol Evol* 2002, **19**:2226-2238.
 32. Papke RT, Zhaxybayeva O, Feil EJ, Sommerfeld K, Muise D, Doolittle WF: **Searching for species in haloarchaea.** *Proc Natl Acad Sci USA* 2007, **104**:14092-14097.
 33. Kotai J: **Instructions for Preparation of Modified Nutrient Solution Z8 for Algae.** Publication B-11/69 Blindern, Oslo: Norwegian Institute for Water Research; 1972.
 34. Rudi K, Kroken M, Dahlberg OJ, Deggerdal A, Jakobsen KS, Larsen F: **Rapid, universal method to isolate PCR-ready DNA using magnetic beads.** *Biotechniques* 1997, **22**:506-511.
 35. **Database of nonribosomal peptide synthetases** [<http://www.nii.res.in/nrps-pks.html>]
 36. Kumar S, Tamura K, Nei M: **MEGA3: Integrated Software for Molecular Evolutionary Genetics Analysis and Sequence.** *Brief Bioinformatics* 2004, **5**:150-163.
 37. Abascal F, Zardoya R, Posada D: **ProtTest: Selection of best-fit models of protein evolution.** *Bioinformatics* 2005, **21**(9):2104-2105.
 38. Huelsenbeck JP, Ronquist F: **MRBAYES: Bayesian inference of phylogeny.** *Bioinformatics* 2001, **17**:754-755.
 39. Ronquist F, Huelsenbeck JP: **MRBAYES 3: Bayesian phylogenetic inference under mixed models.** *Bioinformatics* 2003, **19**:1572-1574.
 40. **The university of Oslo Bioportal** [<http://www.biportal.uio.no/>]
 41. Stamatakis A: **RAxML-VI-HPC: Maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models.** *Bioinformatics* 2006, **22**:2688-2690.
 42. Nei M, Gojobori T: **Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions.** *Mol Biol Evol* 1986, **3**:418-426.
 43. Erhard M, von Dohren H, Jungblut PR: **Rapid identification of the new anabaenopeptin G from *Planktothrix agardhii* HUB 011 using matrix-assisted laser desorption/ionization time-of-flight mass spectrometry.** *Rapid Commun Mass Sp* 1999, **13**:337-343.
 44. Fastner J, Erhard M, von Döhren H: **Determination of oligopeptide diversity within a natural population of *Microcystis* spp. (Cyanobacteria) by typing single colonies by matrix-assisted laser desorption ionization-time of flight mass spectrometry.** *Appl Environ Microbiol* 2001, **67**:5069-5076.
 45. Welker M, Marsalek B, Sejnohova L, von Dohren H: **Detection and identification of oligopeptides in *Microcystis* (cyanobacteria) colonies: Toward an understanding of metabolic diversity.** *Peptides* 2006, **27**(9):2090-2103.

Publish with **BioMed Central** and every scientist can read your work free of charge

"BioMed Central will be the most significant development for disseminating the results of biomedical research in our lifetime."

Sir Paul Nurse, Cancer Research UK

Your research papers will be:

- available free of charge to the entire biomedical community
- peer reviewed and published immediately upon acceptance
- cited in PubMed and archived on PubMed Central
- yours — you keep the copyright

Submit your manuscript here:
http://www.biomedcentral.com/info/publishing_adv.asp

