

Research article

Open Access

A genome-wide analysis of nonribosomal peptide synthetase gene clusters and their peptides in a *Planktothrix rubescens* strain

Trine B Rounge¹, Thomas Rohrlack^{2,4}, Alexander J Nederbragt¹,
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, Department of Biology, Microbial Evolution Research Group (MERG), 0316 Oslo, Norway

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

* Corresponding author

Published: 25 August 2009

Received: 13 November 2008

BMC Genomics 2009, 10:396 doi:10.1186/1471-2164-10-396

Accepted: 25 August 2009

This article is available from: <http://www.biomedcentral.com/1471-2164/10/396>

© 2009 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: Cyanobacteria often produce several different oligopeptides, with unknown biological functions, by nonribosomal peptide synthetases (NRPS). Although some cyanobacterial NRPS gene cluster types are well described, the entire NRPS genomic content within a single cyanobacterial strain has never been investigated. Here we have combined a genome-wide analysis using massive parallel pyrosequencing ("454") and mass spectrometry screening of oligopeptides produced in the strain *Planktothrix rubescens* NIVA CYA 98 in order to identify all putative gene clusters for oligopeptides.

Results: Thirteen types of oligopeptides were uncovered by mass spectrometry (MS) analyses. Microcystin, cyanopeptolin and aeruginosin synthetases, highly similar to already characterized NRPS, were present in the genome. Two novel NRPS gene clusters were associated with production of anabaenopeptins and microginins, respectively. Sequence-depth of the genome and real-time PCR data revealed three copies of the microginin gene cluster. Since NRPS gene cluster candidates for microviridin and oscillatorin synthesis could not be found, putative (gene encoded) precursor peptide sequences to microviridin and oscillatorin were found in the genes *mdnA* and *oscA*, respectively. The genes flanking the microviridin and oscillatorin precursor genes encode putative modifying enzymes of the precursor oligopeptides. We therefore propose ribosomal pathways involving modifications and cyclisation for microviridin and oscillatorin. The microviridin, anabaenopeptin and cyanopeptolin gene clusters are situated in close proximity to each other, constituting an oligopeptide island.

Conclusion: Altogether seven nonribosomal peptide synthetase (NRPS) gene clusters and two gene clusters putatively encoding ribosomal oligopeptide biosynthetic pathways were revealed. Our results demonstrate that whole genome shotgun sequencing combined with MS-directed determination of oligopeptides successfully can identify NRPS gene clusters and the corresponding oligopeptides. The analyses suggest independent evolution of all NRPS gene clusters as functional units. Our data indicate that the *Planktothrix* genome displays evolution of dual pathways (NRPS and ribosomal) for production of oligopeptides in order to maximize the diversity of oligopeptides with similar but functional discrete bioactivities.

Background

Cyanobacteria produce a high number of chemically diverse oligopeptides exhibiting various types of bioactivities ranging from mild enzyme inhibition to initiation of acute toxic effects in pro- or eukaryotes [1]. More than 600 individual compounds have been described and probably many peptides have remained undiscovered. Most oligopeptides can be assigned to chemical classes, of which aeruginosins, anabaenopeptins, cyanopeptolins, microcystins, microginins, and microviridins are among the most recognized ones [2]. Some classes have well characterized biosynthetic pathways [3-6], while for others, hardly anything is known. The biological functions of cyanobacterial oligopeptides are unknown, despite knockout studies for microcystins and cyanopeptolins [3,4,7]. However, gene knockouts in the relevant cyanobacterial strains have turned out to be difficult. Thus, the link between a given oligopeptide and the gene cluster has generally been difficult to establish. Since the precise functions of the peptides are unknown, the reasons for the vast structural peptide diversity (both within- and between classes of oligopeptides) remain obscure.

Several of the oligopeptides classes are produced by non-ribosomal peptide synthetases (NRPSs). NRPS pathways have been verified for microcystins [3,6,8], cyanopeptolins [4,9,10] and aeruginosins [5]. NRPSs have a modular structure with distinct activation domains (A-domains), thiolation domains (T-domains), and condensation domains (C domains) that are easily identified due to signature sequences [11]. Conserved sequence motifs also allow identification of additional modules and domains that often are present in NRPS enzyme complexes, including methyltransferases (M-domains), epimerases (E-domains), thioesterase (TE domains), halogenases, and ABC transporters. The oligopeptide products of NRPS gene clusters may be predicted *in silico* based on binding pocket analyses of A-domains [12,13], phylogeny and the co-linearity rule, i.e. the order of A-domains is co-linear with the amino acid sequence of the finished peptide [14]. Similar predictions of secondary metabolites from NRPS gene clusters have been performed successfully previously in *Streptomyces* ([15]). Some oligopeptides are, however, synthesized ribosomally [16-18], and it is at present unknown if this is the case also for other classes. Due to insufficient genomic information this question has been hard to address.

Evolutionary and phylogenetic studies of individual NRPS gene clusters have revealed frequent horizontal gene transfer (HGT) between related strains [19-21]. This is likely to generate new or recurrent variants of the enzymatic modules, leading to a change in oligopeptide profiles. In addition to HGT (intergenomic recombination),

recombination between sequences within the same genome (intragenomic) may occur even between different classes of NRPS clusters due to a general high genetic similarity (see Majewski and Cohan [22] and Papke *et al.* [23]) among the building blocks (i.e. the modules) [9,10,24] of NRPS gene clusters. Recombination events and point mutations may be reinforced by positive selection for the new variant, indicating that the new oligopeptide variant may have biological significance [25].

Welker *et al.*, 2006 [2] indicated that production of oligopeptides is concentrated within certain genera of cyanobacteria. Why do most strains belonging to in the NRPS producing genera produce several classes of oligopeptides and multiple variants of the same oligopeptide class? Further, what is the biological significance of the many recombination events within the NRPS gene clusters? So far, most studies have investigated single NRPS operons (and their oligopeptides) or compared variants of the same class of operons in different strains. It is likely that all the different oligopeptides within a strain contribute to its survival and fitness, therefore the most relevant approach to take is to examine the entire genome of a strain for NRPS genes. Such an approach should establish relationship between each oligopeptide and operon, and consequently reduce the need for knock-out mutants. Furthermore, the possibility of ribosomal synthesis of some oligopeptide classes can be addressed. In addition, it will be possible to investigate whether exchange of modules between different classes of NRPS gene clusters within the same genome occurs.

Here, we have selected a *Planktothrix rubescens* strain, NIVA CYA 98, that produces all major classes of oligopeptides according to Welker *et al.* [2] and shotgun sequenced the genome by massive parallel pyrosequencing (454) to high depth (18.5×). We have characterized the oligopeptides of this strain in detail by LC-MS-MS. This has enabled a correlation of NRPS gene clusters found within the genome with the identified oligopeptides. A putative gene cluster could be identified for all structurally characterized oligopeptides. However, for two NRPS gene clusters no oligopeptide were detected. Overall, the data show that the combined LC-MS-MS and 454 genome sequencing is a very powerful approach for finding putative secondary metabolite genes.

Results

Oligopeptides produced by *Planktothrix* NIVA CYA 98

Prior to 454-genome sequencing we performed a LC-MS-MS mass spectroscopic fragmentation analysis of *Planktothrix* NIVA CYA 98. This analysis revealed thirteen oligopeptides representing all major oligopeptide classes (Additional file 1: Table S1, Additional file 2: Figure S1).

Aeruginosins, anabaenopeptins, microcystins, and microviridins were found to occur with at least two variants each, typically differing by one amino acid or, as for microviridins, by the presence/absence of Cl in the molecule (Additional file 1: Table S1). An aeruginosin with a previously unreported molecular mass of 593.5 Da [M+H]⁺ was detected, but not fully elucidated. The oligopeptide nature of a second unknown compound with a mass of 1971.8 Da [M+H]⁺ could not be verified due to the production of immonium ions in MS fragmentation experiments. The molecular mass suggested that the oligopeptide might be a microviridin, which is the only known cyanobacterial oligopeptide class within this size range. A full structural elucidation was not attempted.

The genome of *Planktothrix NIVA CYA 98* and detection of NRPS gene clusters

The 454 pyrosequencing-generated genomic sequence dataset consisted of 570912 reads, with the average read length at 262 bp, totaling 150 Mb (Short read archive SRA008127). An assembly produced 692 contigs larger than 500 bp. The average contig size was 8240 bp and the largest contig was 129206 bp. The genome size was estimated at 5.5 Mb with an average GC content of 40.0%. An average sequencing depth of 18.5× was obtained, and the gaps between contigs were no larger than 100 bp within the NRPS gene clusters regions (positions shown in Figure 1). In this work, we were able to detect all NRPS gene clusters by performing BLAST [26] comparisons of the entire

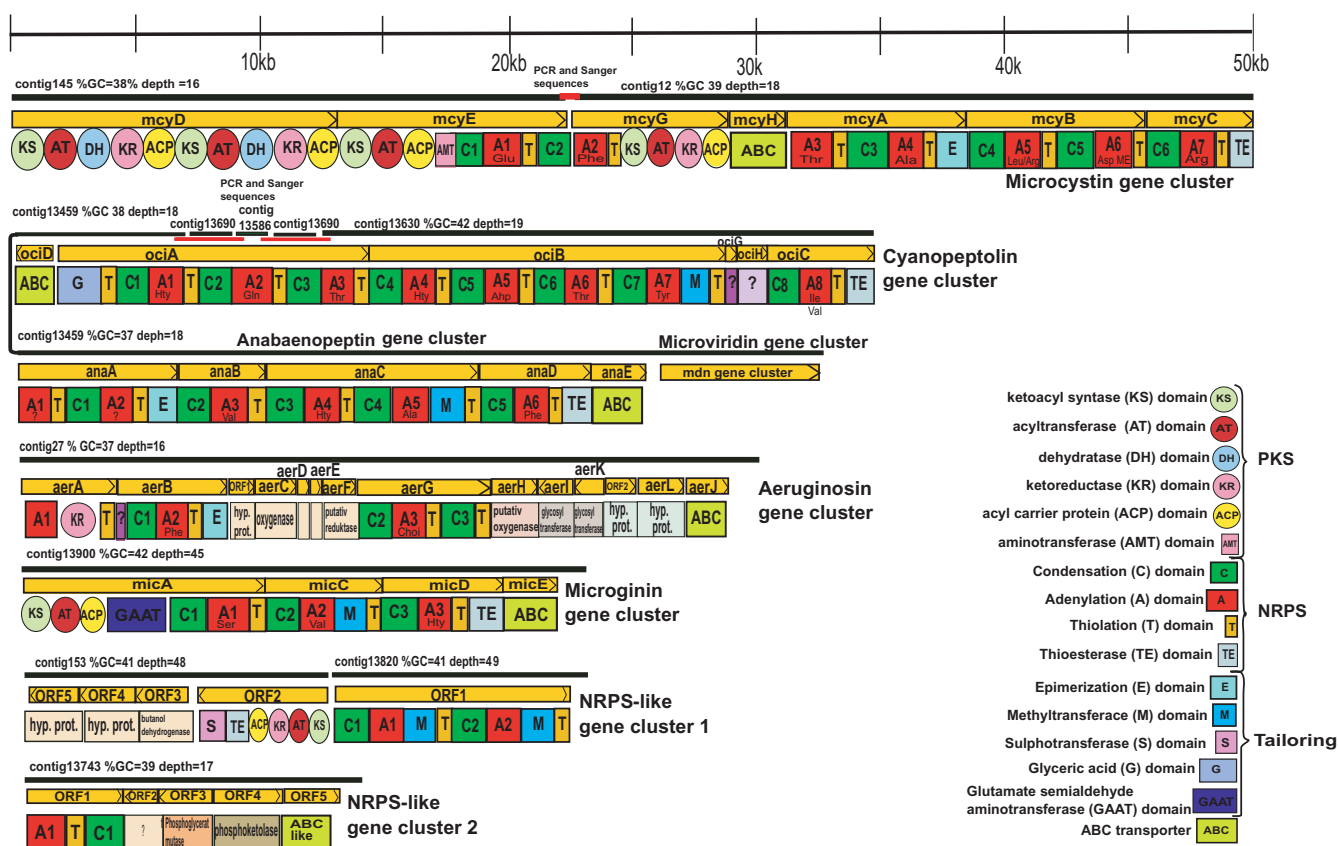


Figure 1
Organization of the NRPS gene clusters found in the *Planktothrix NIVA CYA 98* genome. Gene clusters shown are: microcystin [EMBL: AM990462], cyanopeptolin, anabaenopeptin and microviridin (putative ribosomal produced oligopeptide) on one contig [EMBL: AM990463], aeruginosin [EMBL: AM990465], microiridin [EMBL: AM990464] synthetase gene clusters and two gene clusters with unassigned oligopeptides [EMBL: AM990466 and AM990467]. Contigs from the assembly of 454 reads are illustrated with black lines and gaps closed with Sanger sequencing are illustrated with red lines. Gene names, transcription directions and approximate sizes are illustrated with yellow boxes with arrows. Adenylation- (red; showing *in silico* predicted activated amino acids), condensation- (green), thiolation- (yellow), methyltransferase- (blue), epimerase- (turquoise) glyceric acid- (lilac), halogenation- (purple) and termination domains (grey) are shown by their first letter abbreviations. Polyketide domains are shown as circles. The ABC transporter genes (light green) are connected to six of the seven gene clusters.

genome assembly with a relevant selection of NRPS domains. We have focused our assembly on the putative NRPS sequences. For these sequences we have performed closing of gaps (5 gaps, maximum gap length of 7 bp) and confirmation of correct assembly with PCR and Sanger sequencing (total 5700 bp).

Gene clusters for all structurally determined oligopeptides detected by LC-MS/MS were found among the contigs. These were identified utilizing BLAST searches with characterized A- and C-domains as query sequences. A BLASTn search identified 30 distinct regions with high similarity to previously characterized A-domains (E value < 10⁻¹⁰). Other regions were similar to known A-domains (E value range 10⁻¹⁰ to 10⁻¹), but lacked typical A-domain characteristics. Twenty-eight regions similar to known NRPS C-domains (E value < 10⁻¹⁰) were also identified using BLASTn search. In addition, a number of other domains and genes connected to NRPSs were found, including those encoding ABC-transporters.

Seven distinct NRPS gene clusters present in *Planktothrix NIVA CYA 98*

The annotation of the genome uncovered seven distinct NRPS clusters (Figure 1 and Additional file 3: Table S1) distributed on twelve contigs. Three NRPS gene clusters – *aer*, *oci*, and *mcy* – were highly similar in domain structure and sequence to clusters previously associated with the production of aeruginosins (96% identity to [GenBank: [AM071396](#)]) [5], cyanopeptolins (95% identity to [GenBank: [DQ837301](#)]) [4,9,10], and microcystins (99% identity to [GenBank: [AJ441056](#)]) [3,6,8], respectively. The oligopeptide structures predicted by gene identity and order were in agreement with the structural data obtained in MS fragmentation experiments (Additional file 1: Table S1). We are therefore confident that the NRPS encoded by clusters *mcy*, *oci*, and *aer* are responsible for the synthesis of two demethylated microcystins (*mcy*), the cyanopeptolin oscillapeptin G (*oci*), and two aeruginosins (*aer*) by NIVA CYA 98.

A large NRPS gene cluster was situated in close proximity to the cyanopeptolin gene cluster (*oci*), but transcribed in the opposite direction on the same contig (contig 13459). The novel gene cluster (*anaABCDE*) encodes six basic NRPS modules, including an epimerase and a methyltransferase domain, and an ABC transporter (Figure 1). The size, organization and amino acid specificity of A-domains predicted a link between this gene cluster and production of anabaenopeptins. A reliable prediction of amino acid specificity in the AnaA-A1 and AnaA-A2 domains was not possible. However, the occurrence of lysine in all anabaenopeptins suggested that the protein complex named AnaABCDE was responsible for production of all four anabaenopeptin variants – with AnaA-A2

domain activating Lys and relaxed amino acid specificity of the AnaA-A1 (Tyr/Arg) and AnaB-A3 (Val/Ile).

The *Planktothrix* CYA 98 genome also contains another new NRPS polyketide hybrid gene cluster (*micACDE*). The gene cluster encodes three basic NRPS modules, a polyketide module and an aminotransferase. *In silico* analyses of the novel MicACDE suggested that the gene cluster corresponded to the microginin oscillagin B produced by NIVA CYA 98. Oscillagin A is chlorinated and thus a halogenase is predicted to be part of the biosynthetic pathway. However, BLAST searches could not identify a complete halogenase in the gene cluster or in the entire genome.

The products of the two remaining NRPS gene clusters could not be assigned to oligopeptides detected by LC-MS-MS. NRPS-like gene cluster 1 did not contain an ABC transporter and the assembly of contig 153 (polyketide) and 13820 (peptide synthetase) into a single contig displayed low sequence depth in the contig breakpoint. Non-NRPS genes flanked NRPS-like gene cluster 2. The comparison of oligopeptides detected by mass spectroscopy and predicted by genetic analysis did not identify any NRPS gene clusters associated to the production of oscillatorin and the putative microviridin.

Multi copy NRPS gene clusters

To assess the quality of the contigs containing NRPS gene clusters, sequence depth per base were analyzed (Additional file 1: Table S1). Contig 13900 (length 32 kb) containing the 21 kb microginin gene cluster and the contig containing NRPS-like gene cluster 1 displayed higher sequence depth than the average sequence depth of the genome. Higher sequence depth per base is expected to be observed when reads from repeats and duplicated regions are collapsed into a single contig during the assembly process. The sequence depth per base of the microginin contig (13900) indicated three copies and the read-alignment also showed about 28 sites to be variable among the reads (Additional file 3: Figure S1), were each read variant was represented by approximately 1/3 of the reads covering that site. The 3X copy number of the microginin gene cluster was confirmed by real-time PCR in an independent study (see Nederbragt AJ, Rounge TB, Jakobsen KS: Identification and quantification of genomic repeats and sample contamination in assemblies of 454 pyrosequencing reads, submitted).

PCR and Sanger sequencing showed that contig 13690 was present in two repeats within the cyanopeptolin gene cluster. This contig (13690) also showed two-fold higher sequence depth per base (Additional file 1: Table S1) and which is in agreement with the sequence depth per base/number of gene copy relationship as shown by Neder-

included a leader and core peptide region with conserved motives (Figure 2b). Based on *in silico* analyses of the MdnA core peptide and comparison to the microviridin biosynthetic pathway in *Microcystis*, a microviridin biosynthesis pathway in *Planktothrix* NIVA CYA 98 can be suggested. The pathway includes cleavage, ligation, acetylation and possible methylation and results in an *in silico* predicted microviridin with an unknown side chain. The genomic island, constituting the microviridin gene cluster and the two NRPS gene clusters *ana* and *oci*, we term an oligopeptide island. Interestingly, it includes both putative ribosomal and NRPS biosynthetic pathways.

tBLASTn search identified an 1853 bp open reading frame in the *Planktothrix* genome that encoded an eight amino acid sequence identical to Pro-Asn-Glu-Arg-Gly-Tyr-Gly-Leu sequence of oscillatorin. The protein, here named OsaA (accession [AM990468](http://www.ncbi.nlm.nih.gov/nuccore/AM990468)), has unknown function. The ninth amino acid flanking Leu is tryptophan suggested by Sano and Kaya [27] to be modified to oscillatoric acid (Osc), another part of oscillatorin. The probability of a coincidental occurrence of the nine amino acid sequence is 1.0×10^{-12} (see calculation in Methods) and we therefore suggest that the main part of oscillatorin is produced ribosomally as an integral part of this protein. Similar schemes have been shown for other oligopeptides [16-18]. Thus, we propose a biosynthetic pathway for oscillatorin involving ribosomally encoded amino acids (Pro-Asn-Glu-Arg-Gly-Tyr-Gly-Leu), plus a valine of unknown origin, and cyclization of the oligopeptide by a modifying enzyme possibly encoded by the putative formyltetrahydrofolate cycloligase found in the *osc* gene cluster (Figure 2c). A BLAST search showed similarity to a hypothetical protein L8106_04601 from *Lyngbya* sp. PCC 8106 (53% identity). Interproscan <http://www.ebi.ac.uk/InterProScan/>[28] identified motifs similar to pectate lyases (Pfam CL0268).

Intra-genomic evolutionary processes unveiled by phylogeny

Phylogenetic analyses of A-, C-, M-, E-domains (Additional file 3: Figures S2, S3, S4 and S5) and NRPS-ABC-transporter genes (Figure 3) including the *Planktothrix* CYA 98 NRPS sequences, supplemented with other NRPS sequences, showed clustering according to gene cluster type rather than strain or genera. A- and C- domains also showed functional clustering according to amino acid specificity and position in the gene cluster, respectively (Additional file 3: Figures S2 and S3). The phylogenetic analyses including only *Planktothrix* CYA98 domains indicated the same clustering according to gene cluster and position. However, due to the low number of sequences and few variable sites, the topology did not receive sufficient support (data not shown). Split decomposition analyses were used to identify conflicting phylogenies,

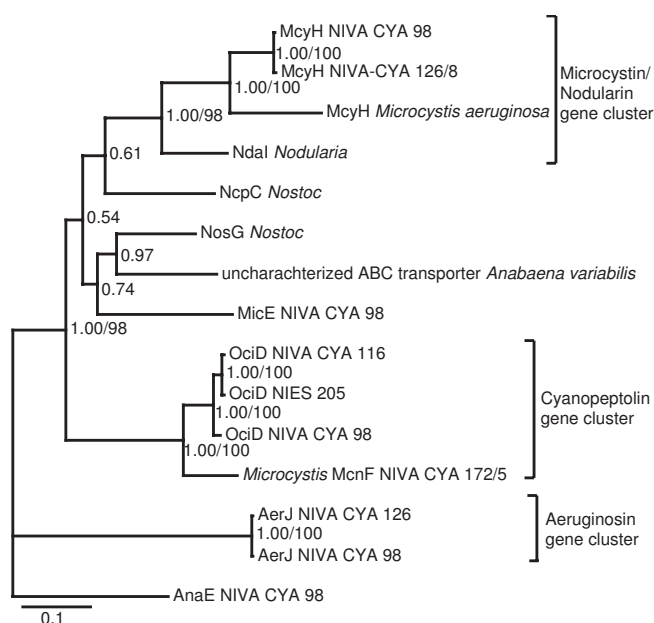


Figure 3
ABC transporter tree constructed using Bayesian inference. Support values for the nodes are Bayesian posterior probability and NJ bootstrap replicates above 50%. Genus origin is denoted for taxa other than *Planktothrix*. Accession numbers for the sequences are listed in Additional file 3: Table S3. The CpRev substitution model and gamma-shaped distribution were used in the Bayesian analyses. The MCMC chains were carried out for 4 million generations with sampling of trees every 100 generations, removing the first 3000 trees.

which represent recombination events (Additional file 4: Figure S1). Splits analyses did not indicate any conflicting phylogenies between domains of the NRPS gene clusters that were determined to be statistically significant by the Phi test analyses (Additional file 4: Figure S1), indicating no recombination events.

Discussion

Metabolomics and the genome-wide approach

Here, we show that combining information about all oligopeptides (from LC-MS-MS) with whole genome sequencing of a *Planktothrix* strain is a powerful method to deduce the relations between oligopeptides and the corresponding NRPS gene clusters. Identification of putative ribosomal pathways of oscillatorin and microviridin further demonstrates the power of the genome-wide approach. All sequence information needed to identify and characterize the oligopeptide gene clusters was derived from *de novo* shotgun sequencing of a previously unsequenced cyanobacterial genus, *Planktothrix*. The gene clusters were assembled to completion by Sanger-sequencing only a few short PCR fragments. The *Plank-*

tothrix strain was not axenic, but *in silico* analyses of the 454-sequences solved this potential problem (Nederbragt AJ, Rounge TB, Jakobsen KS: Identification and quantification of genomic repeats and sample contamination in assemblies of 454 pyrosequencing reads, submitted). The gene clusters were identified without making an effort to assemble the complete genome. This is the most complete oligopeptide investigation in a single cyanobacterial strain to date, and we demonstrate here that the genome-wide approach using high-throughput sequencing combined with structural determination of the metabolites has the potential to accelerate oligopeptide research and increase our understanding of secondary metabolite synthesis in general.

Link between NRPS gene clusters and oligopeptides

In silico analyses of the microcystin, cyanopeptolin and aeruginosin gene clusters correlate with the produced microcystin (desmethyl-RR), oscillapeptin G (cyanopeptolin) and aeruginosin A and a new aeruginosin variant detected by the MS analyses.

The link between the novel *ana* gene cluster, situated in close proximity to the cyanopeptolin gene cluster, and the anabaenopeptins is highly likely due to the correlation between the bioinformatic predictions and the LC-MS-MS deduction of the anabaenopeptins. It is likely that AnaA-BCDE produce all four anabaenopeptin variants due to relaxed specificity of the AnaA1 and AnaA2 domain. The epimerase domain suggests a D-amino acid in the A2 position, corresponding to the conserved Lysine in the oligopeptide, in agreement with structural analyses of anabaenopeptins [29]. However, the mass spectrometry analysis cannot differentiate between the D- and L-configurations of amino acids.

In silico analyses of MicACDE show that this gene cluster corresponds to the oscillagin B produced by the strain. The chlorinated oscillagin A requires an additional halogenase, but we were unable to detect a complete oligopeptide halogenase in the genome. The chlorine atom is attached to the polyketide part of the molecule, so a potential different halogenase outside the contig or in *trans* may be involved.

The microcystin, cyanopeptolin, and aeruginosin and the newly discovered anabaenopeptin and microginin gene clusters all share many features including ABC transporter genes. The function of NRPS ABC-transporters has not been determined, but it has been suggested that they are involved in the export of oligopeptides from the cells. Additionally, the ABC transporter *mcyl* is shown to be essential for microcystin production [30]. All five NRPS gene clusters are co-linear and consist of very similar

NRPS and PKS modules. Compared to previously characterized gene clusters, we observed small changes in aeruginosin gene cluster architecture [5], new tailoring genes/domains (*ociG* and *ociH* encoding unknown functions) and an additional NRPS module in the cyanopeptolin gene clusters (eight NRPS modules cyanopeptolin gene cluster, in contrast to previously described seven modules [10]) and a *mcyA* gene without a methyltransferase in the *mcy* gene cluster [31]. NRPS-like gene cluster 1, lacking an ABC transporter and therefore believed to be incomplete and NRPS-like gene cluster 2, containing only one NRPS encoding module, could not be linked to oligopeptides.

An almost three-fold higher sequencing depth per base and abundance in real-time PCR analysis suggested three copies of the microginin and the NRPS-like 1 gene clusters (Nederbragt AJ, Rounge TB, Jakobsen KS: Identification and quantification of genomic repeats and sample contamination in assemblies of 454 pyrosequencing reads, submitted). The variations (point mutations) in the reads detected in the microginin gene clusters point towards some sequence variation between the copies.

Independent NRPS gene cluster evolution and finding of an oligopeptide island support similar, but discrete, oligopeptide functions

The phylogenetic analyses showed individual functional domains within a NRPS class to be more closely related, compared to NRPS modules from other NRPS classes within the same strain. This indicates independent evolution of each gene cluster, in line with data from single NRPS studies [9,10,25]. Our results therefore indicate that NRPS domain evolution is influenced more by functional constraints acting on the domains, than by the evolutionary relationships between strains. ABC transporter genes were shown to evolve together with their respective gene clusters as sign of a functional unit. Consequently, they are useful for studying NRPS gene cluster evolution. While the general module architecture was conserved within a gene cluster type, frequent tailoring domain rearrangements were shown resulting in new features like the G-domain, and *OciH* of the cyanopeptolin and partial halogenases in *OciG* and *AerB*. A module rearrangement by intra-gene cluster recombination (identical *ociA* C2 and C3) seems likely in the cyanopeptolin gene cluster.

An oligopeptide island – a parallel to pathogenicity islands including the three closely located oligopeptide gene clusters *oci*, *ana* and *mdn* (4597 and 2139 bp distance between gene clusters respectively) was observed. Co-localization of cyanobacterial oligopeptide gene clusters has not been reported previously and may indicate a common target or class of targets for the peptides. There were no further signs of closely linked NRPS gene clusters.

However, a completely assembled genome is needed to determine the relative positions of the NRPS gene clusters to each other.

Dual origin of oligopeptides through NRPS and ribosomal pathways

If oscillatorin and microviridin were produced by NRPSs, the corresponding gene clusters would be large and thus easily detected in the genome by this method. These oligopeptides contain only L-amino acids suggesting a possible ribosomal pathway with subsequent modification and cyclization. It is therefore not unexpected that we have identified genes that are likely to ribosomally encode the precursor of oscillatorin and microviridin. The high similarity between the microviridin (*mdn*) gene cluster presented here and in two *Microcystis* strains [17] strongly suggest that the *mdn* gene cluster produce microviridin. However the length of the side chain, modifications of amino acids and occurrence of additional functional groups are unknown. Ribosomal pathways through the activity of a set of processing enzymes have been shown for microviridins [17] microcyclamide [18] and patellamides [16], demonstrating that at least two fundamentally different biosynthetic pathways are used by cyanobacteria to produce oligopeptides.

Conclusion

The present study suggests that NRPS gene clusters encoding the synthesis of aeruginosins, anabaenopeptins, cyanopeptolins, and microginins evolve independently of one another. Using phylogenetic methods we showed that there are low levels of recombination between similar modules in functionally different NRPS gene clusters within one and the same strain. The recombination levels have not been possible to assess without this genome-wide sequencing approach. In addition, the strain NIVA CYA 98 features a likely ribosomal production of oscillatorin and microviridin. All of these six oligopeptide classes are known as potent inhibitors of different types of proteolytic enzymes, including serine proteases, metalloproteases, and carboxypeptidases [27,32-37]. This strongly suggests dual pathway of evolution of protease inhibitors in *Planktothrix* NIVA CYA 98, allowing different rates of evolution for each NRPS and ribosomal oligopeptide gene clusters. Elucidation of evolutionary rates between the different oligopeptide gene clusters, in particular NRPS- and ribosomal derived, requires further studies. It is tempting, however, to attribute the dual evolution and high diversity of protease inhibitors in NIVA CYA 98 to an "arms-race" between unknown pathogens, competitors or grazers in cyanobacteria [38,39].

Methods

Cyanobacterial culture and growth conditions

Planktothrix rubescens NIVA CYA 98 was isolated in 1982 from Lake Steinsfjorden, Norway. The strain has been

maintained in continuous culture in the NIVA culture collection of Algae in Z8 medium and light at a photon flux density of 10 $\mu\text{mol m}^{-2}\text{s}^{-1}$, and a light-dark cycle of 12:12 hours.

Oligopeptide analysis

Oligopeptides were extracted from filters with cultured *Planktothrix* after lyophilisation using 50% MeOH as described previously [37]. For the detection and identification of oligopeptides, Liquid Chromatography Mass Spectroscopy (LC-MS-MS) was used. The MS-MS detector was run in total scanning mode for the mass range of 500 to 2000 Da during the entire Waters Acquity Ultra-performance Liquid Chromatography (UPLC) linear gradient (from 10% to 45% acetonitrile in water, both containing 0.1% formic acid, within 10 minutes at a flow rate of 0.25 mL min⁻¹). Compounds with a molecular mass within the range of 500–2000 Da were further analyzed in fragmentation experiments with the detector in daughter ion scanning mode.

The structural elucidation of cyanobacterial oligopeptides on the basis of MS fragmentation experiments were identical to previously described studies [9,10,40,41].

Sequencing and bioinformatics

DNA was extracted from cells frozen in isopropanol using a phenol-chloroform extraction. Massive parallel pyrosequencing was performed according to manufacturer's protocol on two PicoTiter Plates with a GS FLX instrument performed at Roche Penzberg, and at University of Oslo, Norway. PCR and Sanger-sequencing with specific primers (Additional file 3: Table S2) was used to close two gaps in the NRPS gene clusters and confirm the correct assembly of the cyanopeptolin gene cluster which was not possible to assemble with only 454 sequences. The gsAssembler 1.1.02 program (Roche-454, Basel, Switzerland) was used to assemble the 454-reads in contigs at default settings.

Axenic *Planktothrix* cultures are very difficult to obtain. A BLASTn search against the 16S rDNA database [42,43] confirms the contamination in *Planktothrix* NIVA CYA 98. 16S sequences similar to *Planktothrix*, *Pseudodevostia insularae*, *Mesorhizobium*, *Flavobacterium psychrophilum* and uncultured cyanobacteria were identified. However, due to large domination of *Planktothrix* in the culture only small fragments of other genomes were sequenced, and these sequences were not assembled in to large contigs due low number of reads. BLASTn of contigs showing atypical GC content and/or low sequence depth per base were shown to be contamination (Nederbragt AJ, Rounge TB, Jakobsen KS: Identification and quantification of genomic repeats and sample contamination in assemblies of 454 pyrosequencing reads, submitted).

Identification of contigs containing NRPS domains was performed with BLAST searches at the UiO Bioportal <http://www.bioportal.uio.no> with cyanopeptolin and microcystin A- and C-domains as query sequences, and confirmed using blast2go against the non-redundant protein sequences (nr) database. ORFfinder <http://www.ncbi.nlm.nih.gov/projects/gorf/> [44] and Vector NTI (Invitrogen, Carlsbad, USA) were used to find and translate open reading frames (ORF). Identification of domains, binding pocket signatures and A-domain substrate specificities were performed using the NRPS database [45]. A-domain specificities were also substantiated using phylogenetic analysis (Additional file 3: Figure S2). Functional predictions of other genes in connection with NRPS gene clusters were determined using BLAST, and conserved motifs searches using InterProScan <http://www.ebi.ac.uk/Tools/InterProScan/> [28]. A-, C-, M-, E-domains and ABC transporter gene Bayesian inference and neighbor joining (NJ) phylogeny was performed as described in Rounge *et al* [10] using MEGA 3.1 [46], MrBayes 3.1 [47], ProTest [48] and the UiO Bioportal. In addition to all the NRPS-domains from *Planktothrix* NIVA CYA 98, the phylogenetic analyses also include NRPS domains from other cyanobacteria (M-domain phylogenies also include other bacteria) (Accession numbers for the sequences used in the phylogenetic analyses are listed in Additional file 3: Table S3). The additional domains from other strains showed that the NIVA CYA 98 domains cluster in several different groups. Split decomposition analyses were performed using SplitsTree4 [49] with default settings (uncorrectedP method) and 1000 bootstrap replicas, and Phi test for recombination [50]. The probability of a coincidental encoded amino acid (P_{aa}) is the number of codons encoding the amino acid divided by the total number of possible codons (i.e. 64). The probability of a coincidental occurrence of an amino acid sequence is $P = P_{aa}^1 \times P_{aa}^2 \times \dots \times P_{aa}^n$. The coincidental occurrence of the amino acid sequence of oscillatorin is therefore $P = (4/64) \times (2/64) \times (2/64) \times (6/64) \times (4/64) \times (2/64) \times (4/64) \times (6/64) \times (1/64) = 1.0 \times 10^{-12}$. This calculation does not take into account codon usage bias.

%GC content and sequencing depth of each base, were determined from the alignment file utilizing a custom perl scripts.

Authors' contributions

TBR carried out the genetic experimentations and TR carried out all MS experiments and oligopeptide analyses. All authors have contributed to the experimental and analytical design. TBR and AJN performed the sequencing, bioinformatics and phylogenetic analysis TBR, KSJ, AJN, TK and TR wrote the manuscript. All authors have read and approved the final manuscript.

Additional material

Additional file 1

Overview of oligopeptides and gene clusters. Table S1: Overview of oligopeptides and gene clusters.

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1471-2164-10-396-S1.pdf>]

Additional file 2

Oligopeptide structures. Figure S1: Oligopeptide structures of Oscillaginins, Aeruginosins, Anabaenopeptin A, Microcystins-RR, Oscillapeptin G and Oscillatorin, The structure of the putative microviridin is not given since no structural elucidation has been conducted.

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1471-2164-10-396-S2.pdf>]

Additional file 3

NRPS protein sequence data, primers and accession numbers. Table S1: Proteins encoded by the NRPS gene clusters. Table S2: PCR and sequencing primers. Table S3: Accession numbers for the sequences used in the phylogenetic analyses. Figure S1: Variation in the microginin gene cluster. Figure S2: A-domain tree – confirms amino acid activation. Figure S3: Bayesian C-domain tree. Figure S4: E-domains phylogeny. Figure S5: M-domains phylogeny.

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1471-2164-10-396-S3.pdf>]

Additional file 4

Splits decomposition analyses. Figure S1: a) Splits decomposition analysis of A-domains from *Planktothrix* CYA 98. b) Splits decomposition analysis of C-domains from *Planktothrix* CYA 98. c) Split decomposition analysis of M-domains from *Planktothrix* CYA 98. d) Splits decomposition analysis of E-domains. e) Splits decomposition analysis of ABC transporters.

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1471-2164-10-396-S4.pdf>]

Acknowledgements

We are grateful to Randi Skulberg at NIVA for providing the strain NIVA CYA 98 and Tom Inge Sønju, Bård Mathiesen and Ave Tooming-Klunderud for excellent technical assistance. We are appreciative for the excellent sequences and GS-FLX training from Roche-454-Life sciences. The work was supported by grants (157338/140 and 183732) to KSJ from the Norwegian Research Council.

References

1. Namikoshi M, Rinehart KL: **Bioactive compounds produced by cyanobacteria.** *J Ind Microbiol* 1996, **17**:373-384.
2. Welker M, von Döhren H: **Cyanobacterial peptides – nature's own combinatorial biosynthesis.** *FEMS Microbiol Rev* 2006, **30**(4):530-563.
3. 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**(10):753-764.

4. 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(1)**:156-167.
5. Ishida K, Christiansen G, Yoshida WY, Kurmayer R, Welker M, Valls N, Bonjoch J, Hertweck C, Börner T, Hemscheidt T, Dittmann E: **Biosynthesis and structure of aeruginoside 126A and 126B, cyanobacterial peptide glycosides bearing a 2-carboxy-6-hydroxyoctahydroindole moiety.** *Chem Biol* 2007, **14(5)**:565-576.
6. Moffitt MC, Neilan BA: **On the presence of peptide synthetase and polyketide synthase genes in the cyanobacterial genus *Nodularia*.** *FEMS Microbiol Lett* 2001, **196(2)**:207-214.
7. Christiansen G, Fastner J, Erhard M, Börner T, Dittmann E: **Microcystin biosynthesis in *Planktothrix*: Genes, evolution, and manipulation.** *J Bacteriol* 2003, **185(2)**:564-572.
8. 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(2)**:686-692.
9. 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(Pt 5)**:1382-1393.
10. 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(22)**:7322-7330.
11. Sieber SA, Marahiel MA: **Molecular mechanisms underlying nonribosomal peptide synthesis: approaches to new antibiotics.** *Chem Rev* 2005, **105(2)**:715-738.
12. Stachelhaus T, Mootz HD, Marahiel MA: **The specificity-conferring code of adenylation domains in nonribosomal peptide synthetases.** *Chem Biol* 1999, **6(8)**:493-505.
13. Challis GL, Ravel J, Townsend CA: **Predictive, structure-based model of amino acid recognition by nonribosomal peptide synthetase adenylation domains.** *Chem Biol* 2000, **7(3)**:211-224.
14. Marahiel MA, Stachelhaus T, Mootz HD: **Modular Peptide Synthetases Involved in Nonribosomal Peptide Synthesis.** *Chem Rev* 1997, **97(7)**:2651-2674.
15. Omura S, Ikeda H, Ishikawa J, Hanamoto A, Takahashi C, Shinose M, Takahashi Y, Horikawa H, Nakazawa H, Osonoe T, Kikuchi H, Shiba T, Sakaki Y, Hattori M: **Genome sequence of an industrial microorganism *Streptomyces avermitilis*: deducing the ability of producing secondary metabolites.** *Proc Natl Acad Sci USA* 2001, **98(21)**:12215-12220.
16. Schmidt EW, Nelson JT, Rasko DA, Sudek S, Eisen JA, Haygood MG, Ravel J: **Patellamide A and C biosynthesis by a microcin-like pathway in *Prochloron didemni*, the cyanobacterial symbiont of *Lissoclinum patella*.** *Proc Natl Acad Sci USA* 2005, **102(20)**:7315-7320.
17. Ziemert N, Ishida K, Liaimer A, Hertweck C, Dittmann E: **Ribosomal Synthesis of Tricyclic Depsipeptides in Bloom-Forming Cyanobacteria.** *Angew Chem Int Ed* 2008, **47(40)**:7756-7759.
18. Ziemert N, Ishida K, Quillardet P, Bouchier C, Hertweck C, de Marsac NT, Dittmann E: **Microcylamide biosynthesis in two strains of *Microcystis aeruginosa*: from structure to genes and vice versa.** *Appl Environ Microbiol* 2008, **74(6)**:1791-1797.
19. Mikalsen B, Boison G, Skulberg OM, Fastner J, Davies W, Gabrielsen TM, Rudi K, Jakobsen KS: **Natural variation in the microcystin synthetase operon *mcy ABC* and impact on microcystin production in *Microcystis* strains.** *J Bacteriol* 2003, **185(9)**:2774-2785.
20. Kurmayer R, Gumpenberger M: **Diversity of microcystin genotypes among populations of the filamentous cyanobacteria *Planktothrix rubescens* and *Planktothrix agardhii*.** *Mol Ecol* 2006, **15(12)**:3849-3861.
21. Tanabe Y, Kaya K, Watanabe MM: **Evidence for recombination in the microcystin synthetase (*mcy*) genes of toxic cyanobacteria *Microcystis* spp.** *J Mol Evol* 2004, **58(6)**:633-641.
22. Majewski J, Cohan FM: **DNA sequence similarity requirements for interspecific recombination in *Bacillus*.** *Genetics* 1999, **153(4)**:1525-1533.
23. Papke RT, Zhaxybayeva O, Feil EJ, Sommerfeld K, Muise D, Doolittle WF: **Searching for species in haloarchaea.** *Pro Nat Acad Sci USA* 2007, **104(35)**:14092-14097.
24. 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(1)**:143-151.
25. Rounge TB, Rohrlack T, Kristensen T, Jakobsen KS: **Recombination and selection forces in cyanopeptolin NRPS operons from highly similar, but geographically remote *Planktothrix* strains.** *BMC Microbiol* 2008, **8**:141.
26. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ: **Basic local alignment search tool.** *J Mol Biol* 1990, **215(3)**:403-410.
27. Sano T, Kaya K: **Oscillatorin, A chymotrypsin inhibitor from toxic *Oscillatoria agardhii*.** *Tetrahedron Lett* 1996, **37(38)**:6873-6876.
28. Hunter S, Apweiler R, Attwood TK, Bairoch A, Bateman A, Binns D, Bork P, Das U, Daugherty L, Duquenne L, Finn RD, Gough J, Haft D, Hulo N, Kahn D, Kelly E, Laugraud A, Letunic I, Lonsdale D, Lopez R, Madera M, Maslen J, McAnulla C, McDowal J, Mistry J, Mitchell A, Mulder N, Natale D, Orengo C, Quinn AF, Selengut JD, Sigrist CJA, Thimma M, Thomas PD, Valentin F, Wilson D, Wu CH, Yeats C: **InterPro: the integrative protein signature database.** *Nucleic Acids Res* 2009:D211-215.
29. Harada K, Fujii K, Shimada T, Suzuki M, Sano H, Adachi K, Carmichael WW: **2 Cyclic-Peptides, Anabaenopeptins, a 3rd Group of Bioactive Compounds from the Cyanobacterium *Anabaena-Flos-Aquae* Nrc-525-17.** *Tetrahedron Lett* 1995, **36(9)**:1511-1514.
30. Pearson LA, Hisbergues M, Börner T, Dittmann E, Neilan BA: **Inactivation of an ABC transporter gene, *mcyH*, results in loss of microcystin production in the cyanobacterium *Microcystis aeruginosa* PCC 7806.** *Appl Environ Microbiol* 2004, **70(11)**:6370-6378.
31. Kurmayer R, Christiansen G, Gumpenberger M, Fastner J: **Genetic identification of microcystin ecotypes in toxic cyanobacteria of the genus *Planktothrix*.** *Microbiology* 2005, **151(Pt 5)**:1525-1533.
32. 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(22)**:6871-6882.
33. Itou Y, Suzuki S, Ishida K, Murakami M: **Anabaenopeptins G and H, potent carboxypeptidase A inhibitors from the cyanobacterium *Oscillatoria agardhii* (NIES-595).** *Bioorg Med Chem Lett* 1999, **9(9)**:1243-1246.
34. Sano T, Kaya K: **Oscillamide-Y, a Chymotrypsin Inhibitor from Toxic *Oscillatoria-Agardhii*.** *Tetrahedron Lett* 1995, **36(33)**:5933-5936.
35. Shin HJ, Matsuda H, Murakami M, Yamaguchi K: **Aeruginosins 205A and -B, serine protease inhibitory glycopeptides from the cyanobacterium *Oscillatoria agardhii* (NIES-205).** *J Org Chem* 1997, **62(6)**:1810-1813.
36. Ishida K, Kato T, Murakami M, Watanabe M, Watanabe MF: **Microginins, zinc metalloproteases inhibitors from the cyanobacterium *Microcystis aeruginosa*.** *Tetrahedron* 2000, **56(4)**:8643-8656.
37. Rohrlack T, Christoffersen K, Hansen PE, Zhang W, Czarnecki O, Henning M, Fastner J, Erhard M, Neilan BA, Kaebernick M: **Isolation, characterization, and quantitative analysis of microviridin J, a new *Microcystis* metabolite toxic to *Daphnia*.** *J Chem Ecol* 2003, **29(8)**:1757-1770.
38. Rohrlack T, Christoffersen K, Kaebernick M, Neilan BA: **Cyanobacterial protease inhibitor microviridin J causes a lethal molting disruption in *Daphnia pulex*.** *Appl Environ Microbiol* 2004, **70(8)**:5047-5050.
39. Zainuddin EN, Mentel R, Wray V, Jansen R, Nimtz M, Lalk M, Mundt S: **Cyclic depsipeptides, ichthyopeptins a and b, from *Microcystis ichthyoblabe*.** *J Nat Prod* 2007, **70(7)**:1084-1088.
40. 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(11)**:5069-5076.
41. Welker M, Brunke M, Preussel K, Lippert I, von Döhren H: **Diversity and distribution of *Microcystis* (Cyanobacteria) oligopeptide chemotypes from natural communities studied by single-col-**

- ony mass spectrometry. *Microbiology* 2004, **150**(Pt 6):1785-1796.
42. Maidak BL, Cole JR, Lilburn TG, Parker CT Jr, Saxman PR, Farris RJ, Garrity GM, Olsen GJ, Schmidt TM, Tiedje JM: **The RDP-II (Ribosomal Database Project)**. *Nucleic Acids Res* 2001, **29**(1):173-174.
 43. Cole JR, Chai B, Farris RJ, Wang Q, Kulam-Syed-Mohideen AS, McGarrell DM, Bandela AM, Cardenas E, Garrity GM, Tiedje JM: **The ribosomal database project (RDP-II): introducing myRDP space and quality controlled public data**. *Nucleic Acids Res* 2007:D169-172.
 44. **ORFfinder** [<http://www.ncbi.nlm.nih.gov/projects/gorf/>]
 45. **Database of nonribosomal peptide synthetases** [<http://www.nii.res.in/nrps-pks.html>]
 46. Kumar S, Tamura K, Nei M: **MEGA3: Integrated Software for Molecular Evolutionary Genetics Analysis and Sequence**. *Brief Bioinformatics* 2004, **5**:150-163.
 47. Ronquist F, Huelsenbeck JP: **MRBAYES 3: Bayesian phylogenetic inference under mixed models**. *Bioinformatics* 2003, **19**:1572-1574.
 48. Abascal F, Zardoya R, Posada D: **ProtTest: Selection of best-fit models of protein evolution**. *Bioinformatics* 2005, **21**(9):2104-2105.
 49. Huson DH: **SplitsTree: analyzing and visualizing evolutionary data**. *Bioinformatics* 1998, **14**(1):68-73.
 50. Bruen TC, Philippe H, Bryant D: **A simple and robust statistical test for detecting the presence of recombination**. *Genetics* 2006, **172**(4):2665-2681.

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

