Mining for NRPS and PKS genes revealed a high diversity in the *Sphagnum* bog metagenome

Christina A. Müller[a,b]*, Lisa Oberauner-Wappis[a,b,]*, Armin Peyman[a,b], Gregory C. A. Amos[c], Elizabeth M. H. Wellington[c] and Gabriele Berg[b#]

Austrian Centre of Industrial Biotechnology (ACIB GmbH), Graz, Austria[a]; Institute of Environmental Biotechnology, Graz University of Technology, Graz, Austria[b]; School of Life Sciences, University of Warwick, United Kingdom[c]

Running title: NRPS and PKS screening in the *Sphagnum* microbiome

C.A.M and L.OW. contributed equally to this work.

[#] Address correspondence to Gabriele Berg, gabriele.berg@tugraz.at.

Abstract

Sphagnum bog ecosystems are one of the oldest vegetation forms harbouring a specific microbial community, which is known to produce an exceptionally wide variety of bioactive substances. Although the *Sphagnum* metagenome indicate a rich secondary metabolism, the genes are not yet explored. To analyse non-ribosomal peptide synthetases (NRPS) and polyketide synthases (PKS) the diversity of NRPS and PKS genes in *Sphagnum*-associated metagenome was investigated by *in silico* data mining and sequence-based screening (PCR-amplification of 9500 fosmid clones). The *in silico* Illumina-based metagenomic approach resulted in the identification of 279 NRPS, 346 PKS, as well as 40 PKS-NRPS hybrid gene sequences. Occurrence of NRPS sequences was strongly dominated by the phyla *Protebacteria*, especially by the genus *Burkholderia*, while PKS sequences were mainly affiliated to *Actinobacteria*. Thirteen novel NRPS-related sequences were identified by PCR-amplification screening, displaying amino acid sequence identities of 48 to 91% to annotated sequences of the phyla *Proteobacteria*, *Actinobacteria* and *Cyanobacteria*. Some of the identified metagenomic clones showed closest similarity to peptide synthases from *Burkholderia* or *Lysobacter*, which are emerging bacterial sources of yet undescribed bioactive metabolites. This study highlights the role of the extreme natural ecosystems as a promising source for detection of secondary compounds and enzymes, serving as a source for biotechnological applications.

Keywords: *Sphagnum* moss, NRPS/PKS, metagenome, fosmid library, *in silico* data mining

Introduction

The plant microbiome has established itself in recent years as an important player in the field of plant health and agricultural productivity (1). Mosses, especially *Sphagnum* species, are a phylogenetically old group of land plants in bog ecosystems, which are unique extreme habitats displaying high acidity, low temperature and water saturation, together with extremely low concentrations of mineral nutrients (2). *Sphagnum* bogs in particular reflect an enormous importance because of their approved role in the global carbon cycle and have therefore been used globally as an indicator of climate change (3). The role of *Sphagnum* mosses as an important model for examining the plant-microbe interactions as well as the ecology of plant-associated bacteria has been reported (4). *Sphagnum* mosses are in particular characterized by a specific but diverse microbial community (5–7, 4), which fulfil important functions in cooperation with the host, promoting plant growth by enhancing nutrient supply and showing antagonistic activity against plant pathogens (4, 8). In fact, high abundance of functional systems that are responsible for oxidative and drought stress, repair, resistance and genetic exchange were detected recently by metagenomic analysis of the *Sphagnum* microbiome (4). The biological activity of bryophytes and their traditional use in medicine and agriculture are well known (9). It has been shown, that *Sphagnum* species produce bioactive secondary metabolites, which influence their microbial colonisation (8). Bryophytes have been traditionally used in China, India and among Native Americans for their antifungal properties, and *Sphagnum* moss was employed as natural disinfectant for natural nappies or wound dressings in Europe (9). There are more than 300 natural compounds that have been isolated from bryophytes, mainly from liverworts (*Marchantiophyta*) but also from mosses (*Bryophyta*) (10). Some of the reported natural products in mosses are highly unsaturated fatty acids, alkanones, triterpenoids and flavonoids (10). Biological effects observed for extracts of mosses include antimicrobial, antifungal, cytotoxic, and antitumor activities (11, 12). The analysis of endo- and ectophytic bacterial strains revealed that *Sphagnum* moss harbours an

63 extraordinary high proportion of antifungal isolates, as well as a lower proportion of
64 antibacterial isolates, which can partly explain the medicinal use (8). However, the major part
65 (97%) of microbial communities associated with *Sphagnum* mosses belong to non-cultivable
66 forms (7). Therefore, the antimicrobial potential of the moss microbiome remains mostly
67 unexplored.

68 Prominent classes of active compounds from microbial and plant origin (antibiotics,
69 antifungals or antitumor agents) are synthesized by large multi-modular enzymes, the non-
70 ribosomal peptide synthetases (NRPSs) and polyketide synthases (PKSs) or hybrids thereof
71 (13). The widespread occurrence of the NRPS and PKS genetic machinery across all three
72 domains of life (bacteria, archaea and eukarya) has been reported (14). Bacteria host the
73 majority of the described NRPS/PKS gene clusters, which are especially common in the phyla
74 *Proteobacteria*, *Actinobacteria*, *Firmicutes* and *Cyanobacteria* (14). Both NRPSs and PKSs
75 are regarded as multi-enzymatic mega-synthases (200-2000 kDa) organized in a modular
76 assembly line fashion, that contains catalytic modules for single rounds of chain elongation
77 and variable modifications of the intermediate product (13). In non-ribosomal peptides
78 (NRPs) defined monomers, amino acids or non-proteinogenic monomers, are incorporated by
79 specific modules consisting of three essential catalytic domains. The adenylation (A) domain
80 catalyse the activation of the amino acid, which is then transferred to the peptidyl carrier
81 protein (PCP), followed by condensation of the bound amino acid (condensation (C) domain)
82 (15). In a similar way, PKS mega-enzymes consist of an acyltransferase (AT) domain for
83 selection of the monomer substrate, usually malonyl- or methyl-malonyl-CoA, priming it to
84 the acyl carrier protein (ACP), followed by chain elongation and condensation (C-C-bond
85 formation) by a ketosynthase (KS) domain (16). In addition to the core domains, a variable
86 set of domains for further modifications of the peptide chain (epimerization,
87 heterocyclization) (15) or the polyketide chain (ketoreduction, dehydration) are available
88 (16). Termination of the chain is catalysed in both NRPS and PKS by a thioester (TE)

domain. Because of structural and functional similarities between elements of each class, NRPS and PKS can form mixed assembly lines (hybrid gene clusters) (14). Rational design of combinatorial PKS and NRPS modules is an emerging strategy to design tailor-made antibiotics or therapeutic compounds (17).

Rapid development of new metagenomic approaches permit the assessment and exploitation of the taxonomic as well as the functional diversity of microbial communities (18, 19). The discovery of novel biocatalysts for production of natural active compounds can be accomplished through screening of metagenomic libraries, for example by PCR-based screening techniques. Metagenomic applications were recently used for detection of NRPS/PKS genes of bacterial communities in soil (20) and marine environments (21), but not for plant-associated microhabitats or extreme bog ecosystems known to be rich in antimicrobial activity (22).

In this study, our aim was to explore the diversity of sequences assigned to NRPS and PKS genes in the *Sphagnum*-associated bacteria, allowing us greater insight into potentially novel synthetic pathways and biocatalysts. We combined two sequence-based screening methods to search for NRPS and PKS related sequences: *in silico* mining in the moss metagenomic database and PCR-amplification screening of a metagenomic fosmid library in *E. coli*. Origin and abundance of the identified metagenomic sequences were investigated. Our results demonstrate how sequence-based screenings can be used to detect NRPS and PKS genes involved in the biosynthesis of secondary metabolites within the *Sphagnum* moss microbiome.

Methods

In silico analysis of the Sphagnum moss dataset

The recently published metagenomic dataset of the *Sphagnum* microbiome (4) was employed for data mining of NRPS and PKS gene sequences. The generated paired-end reads from untreated and normalized sequences from a previous study by Bragina *et al*. (4) were quality filtered (trimming of read ends with an average <Q30 and polyN nucleotides) using the PRINSEQ software according to the manual (http://prinseq.sourceforge.net). The normalized dataset consisted of ssDNA sequences after treatment and separation by hydroxyapatite chromatography (4). Untreated and normalized datasets were pooled and the generated mixed dataset was used for *de novo* assembly with the SOAPdenovo2 software (http://soap.genomics.org.cn/soapdenovo.html) using default parameters for metagenomic datasets (23). Briefly, the SOAPec correction tool was used first to filter short reads (kmer size ≤17, quality value of 33, thread of 12), using a low frequency cutoff value of 5. For contig assembly (de Bruijn graph) a kmer size of 23 was employed, using an average insert size of 200 bp, a read length cut-off of 100 bp, a paired end cutoff value of 3 and a minimum alignment length of 32. The resulting fasta file from *de novo* assembly, including the assembled scaffolds and contigs, was employed as query for blastx analysis using a self-developed bioinformatics workflow (fasta-file splitting, blastx, and generation of the blastx database). Then, the resulting moss metagenomic blastx database was mined for NRPS and PKS gene sequences employing a self-developed script, which works on the basis of a search term. The terms “non ribosomal peptide synthetase” and “polyketide synthase” were used for the search. Additionally, to compare the abundance of NRPS and PKS with other commonly found microbial genes a search was performed for monooxygenases and rpoD RNA polymerases.

Sampling and total community DNA isolation

Sampling of *S. magellanicum* gametophytes and preparation of the microbiome for total community DNA isolation (enrichment of microbial fraction and removal of plant debris) was performed as reported in Bragina *et al*. (4). To construct the fosmid library for PCR-based screening, total community DNA was extracted using the Meta-G-Nome™ DNA Isolation Kit (Epicentre, Madison, WI, USA) according to the manufacturer's protocol. Metagenomic DNA was randomly sheared to fragment sizes of approximately 40 kb that were used for construction of the fosmid library.

Metagenomic fosmid library generation and PCR screening

A metagenomic fosmid library from the *Sphagnum* moos microbiome was constructed using the CopyControl Fosmid Library Production Kit (Epicentre) as described by the manufacturer's instructions. In short, the isolated metagenomic DNA (1 µg) of approximately 40 kb was directly used for blunt-end repair and was ligated into the CopyControl pCC2FOS™ vector (1.1 µg vector, 0.62 µg insert DNA). The ligated DNA was packaged with MaxPlax Lambda Packaging Extracts. The packaged phage particles were employed to infect *E. coli* EPI300-T1^{R} cells. The fosmid library was spread onto LB agar plates containing 12.5 µg ml^{-1} chloramphenicol and incubated at 37 °C overnight. In total, 9500 clones were randomly transferred to 96-well microtiter plates (MTP) containing 150 µl of LB medium with chloramphenicol (12.5 µg ml^{-1}) using sterile tooth picks, each plate consisting of ninety-five different clones and one negative control (only medium). MTP cultures were grown at 37 °C overnight by shaking at 225 rpm, and finally stored at -70 °C after addition of glycerol to a final concentration of 25% (v/v) to each well. To estimate the average insert size in the fosmid clones, restriction digestion was performed with BamHI. For the PCR screening, 10 clones were pooled together for a total of 10 MTP-pools in LB medium (12.5 µg ml^{-1} chloramphenicol). The pooled MTPs were cultivated under the conditions described above, upon addition of 1x Fosmid Autoinduction Solution (Epicentre) to induce high copy number.

Denaturation (15 min at 99°C) of diluted MTP cultures (1:2 in ddH_2O) and centrifugation (4000 rpm, 5 min) were performed in order to make the fosmid DNA accessible to PCR screening with the two previously reported degenerated primer pairs NRPS1 and NRPS2 (24) which are given in Supplementary Material, Table S1. A standard PCR reaction (25 µl) contained 1x Taq 2xMaster Mix (12.5 µl, New England Biolabs, Ipswich, UK), 0.4 µM of each primer (1 µl degenerated primer, Table 1; Sigma-Aldrich, Wien, Austria), ddH_2O (4.25 µl), 5% (v/v) DMSO (1.25 µl) and 5 µl of pooled template DNA. Following PCR program was used: 95 °C, 5 min; 35 cycles of 95 °C, 1 min; 57 °C, 1 min; 68 °C, 1 min; and elongation at 68 °C, 10 min. PCR products were analysed by 2% agarose/TAE gel electrophoresis. Localization of positive clones was achieved by repetition of the PCR as described above, employing in this case the 10 single clones from the previously identified positive MTP-pool.

Phylogenetic analysis of identified fosmid clones

Diluted PCR products (1:1000) from single fosmid clones identified as positive hits during rescreening of the library were amplified with shorter non-degenerated primer pairs NRPS1ndeg and NRPS2ndeg (nested PCR), employing the above mentioned PCR program. These primers resemble those used for library screening but lacking the degenerated nucleotides in the 3'-region (Supplementary Material, Table S1). The resulting PCR products were purified using a Wizard® SV 96 PCR Clean-Up System (Promega, Mannheim, Germany) and sent for Sanger sequencing at LGC Genomics (Berlin, Germany). Based on the first sequencing results, selected fosmids (3-F3, 3-H3, 2-D4, 2-F4, 7- B9, 6-H4) were partially sequenced by primer walking using for each a sequence-specific primer (Supplementary Material, Table S1). This allowed retrieval of longer DNA sequences contiguous to the previously identified NRPS gene region (up to 1100 bp) directly from the fosmid clones. Analysis of the obtained sequences was performed using BLASTx (25) against the non-

redundant (nr) protein sequences database at NCBI (http://www.ncbi.nlm.nih.gov/protein) or the KEGG database (http://www.genome.jp/kegg). The retrieved amino acid (aa) sequences of positive fosmid clones were employed for phylogenetic analysis, together with the most similar gene sequences from the blastx search. Alignment of aa sequences and construction of the phylogenetic tree were performed with CLC Main Workbench 6.9.1. The phylogenetic tree was generated using the Unweighted Pair Group Method using Arithmetic averages (UPMGA), Kimura Protein as distance measure and a bootstrap of 1000 replicates.

Results

Data mining in the moss metagenomic dataset

The metagenomic dataset of *Sphagnum magellanicum* moss (Illumina HiSeq 2x100 paired-end sequencing) consist of 17323 Mbp raw (pair-number: 86617475) and 14141 Mbp normalised metagenomic DNA (pair-number: 70705608) (4). *De novo* assembly of the pooled metagenome (raw and normalized reads; N50 of 199) yielded 1062181 contig sequences (168393 scaffolds and 893788 contigs), featuring a total size of 188.2 Mbp with an average length of 183 bp (Supplementary Material, Table S2).

Blastx analysis of the metagenomic dataset revealed that NRPS, PKS and NRPS-PKS hybrid gene sequences are present in the moss microbiome. Without cutoff settings, the blastx dataset consist of 279 NRPS, 346 PKS, and 40 hybrid or mixed gene sequences (Table 1). This translates into a rate of 0.063% contigs containing NRPS/PKS gene sequences in the assembled *Sphagnum* metagenome (665 out of 1062181 contigs). In comparison, other common microbial gene sequences such as those coding for monooxygenases (3244 contigs) or the rpoD RNA polymerase (sigma 70 factor; 160 contigs), contributed to the assembly with rates of 0.305% and 0.015%, respectively. It has to be considered that the estimated rates rely on the availability of annotated homolog genes in the employed databases, which can be subjected to changes along with the discovery of novel sequences. Therefore, our results may underrepresent the real frequency of these protein families in the microbial community. The highest abundance of NRPS closest matches belong to the phyla *Proteobacteria* (*Burkholderia* spp., n=50, 18%; *Pseudomonas* spp., n=40, 14%; *Myxococcus* spp., n=27, 10%) and *Actinobacteria* (*Streptomyces* spp., n=25, 9%; *Rhodococcus* spp., n=10, 4%). In the case of PKS, closest matches are mainly represented by *Actinobacteria* (*Mycobacterium* spp., n=92, 27%; *Streptomyces* spp., n=46, 13%) and uncultured bacteria (n=18, 5%). The remaining hits from the *in silico* search show a diverse distribution of underrepresented taxa from *Proteobacteria* and *Actinobacteria*, but also from *Cyanobacteria* (e.g. *Nostoc*,

223 *Anabaena*, *Pseudoanabaena*, *Microcystis*, *Fischerella*). Hybrid-gene matches are mainly
224 affiliated to the phylum *Proteobacteria* (*Pseudomonas* spp., n=7, 18%; *Lysobacter* spp., n=6,
225 15%, and *Myxococcus* spp., n=6, 15%).

226 Employing an E value cutoff of 10^{-20} (bitscore >88) it was possible to select the best
227 matching sequences in the database. This resulted in a confined selection of 34 NRPS and 28
228 PKS genes, as well as three NRPS-PKS hybrids (Supplementary Material, Table S3). These
229 sequences display diverse identities to their closest neighbour from the blastx analysis ranging
230 from 35 to 98%. Many of the selected PKS sequences with a higher similarity (>60%
231 identity) to annotated genes in the nr database are mainly related to the genus *Mycobacterium*.
232 The remaining sequences (35-60% identity) show similarity to PKSs from diverse genera,
233 such as *Streptomyces* or *Rubrivivax*. In the case of NRPS gene sequences, the most abundant
234 genera from the closest hits are *Burkholderia*, followed by *Bradyrhizobium, Pseudomonas*,
235 *Mycobacterium* and *Pectobacterium*. One hybrid gene sequence is related to the genus
236 *Lysobacter* and the remaining two resemble the well-studied Yersiniabactin synthase (26)
237 from *Pseudomonas syringae*. Additionally, five of the NRPS and PKS gene sequences display
238 the highest similarity to annotated genes from yet uncultured bacteria.
239
240 *Fosmid library screening*
241 The *Sphagnum* moss fosmid library was generated employing 1 µg of metagenomic DNA,
242 obtained from the enriched microbial fraction (1 g) contained in 200 g moss. Based on the
243 number of clones obtained (96,025) and an average insert size of 27 kb, the library size was
244 estimated as 2.6 Gb. In total, 9,500 randomly selected clones (0.26 Gb) were employed for the
245 screening of NRPS genes by PCR amplification in MTP using two different degenerated
246 oligonucleotide primer pairs (NRPS1 and NRPS2). The primers were previously designed to
247 target the adenylation domain of NRPS gene clusters in diverse soil samples (27). In total, 25
248 NRPS1 and 33 NRPS2 wells, each containing a pool of ten clones, gave a positive

amplification result. A second round of screening of the corresponding single clones resulted in 11 NRPS1 and 26 NRPS2 putative positive hits. Positive clones were subjected to optimization of PCR-conditions with shorter non-degenerated primer pairs in a nested PCR to avoid the amplification of unspecific products. Amplicons of 21 NRPS positive fosmids that were obtained as a pure DNA band were sent for sequencing.

Based on blastx analysis against the nr protein sequences database (NCBI) and the KEGG database, 14 sequences showed similarity to genes encoding for peptide synthases (NRPS, Table 2) and could be therefore assigned to these protein family. Seven clones were detected with each primer pair (NRPS1 or NRPS2) respectively, resulting in an average hit rate of one NRPS gene per 37 Mb of screened moss metagenomic DNA.

NRPS sequence identities to the closest hits from blastx analysis range from 48 to 99% (Table 2). Most of the identified closest neighbour sequences belong to the phylum of *Proteobacteria* (12 hits; 86%), and in particular to the genus of *Pseudomonas* (4 hits, Table 2). The remaining two hits include the phyla *Actinobacteria* (*Kutzneria albida*) and *Cyanobacteria* (*Rubidibacter lacunae*). Despite of the clone 3-F3, showing a very high aa sequence homology of 99.4% to a peptide synthase from *Pseudomonas* sp. Ag1, all other 13 clones harbour novel, not yet annotated aa sequences with maximal identities of up to 91%. The obtained aa sequences aligned partially to conserved domains of annotated NRPS sequences, either to the adenylation domain (A_NRPS motif; cd05930), the phosphopantetheine prosthetic group attachment site (pp-binding motif; pfam00550) or the condensation domain (pfam00668) (see multiple sequence alignment in Supplementary Material, Figure S1).

To gain a better overview of putative NRPS amino acid sequences found in the fosmid library a phylogenetic tree was generated (Figure 1). Sequences with a hit length of >100 aa and a gap value <2% were placed on the tree. The phylogenetic analysis exhibits the distribution of the putative NRPS sequences, which clustered into three main groups. The first

group is composed of *α-Proteobacteria*, containing the sequences NRPS 7-F1 and 6-H4. These sequences are most related to peptide synthetases from the family of *Caulobacteraceae* and *Agrobacterium* spp. The product of these peptide synthase gene clusters are, however, still unknown. The second and more diverse group includes the NRPS sequences 4-B4, 7-D4, and 6-B1, clustering in closer proximity to peptide synthetases from α-, β-, and *γ-Proteobacteria*. Sequences from clones 7-D4 and 6-B1 are close related to thioester-reductases from *Dyella* and *Variovorax* species. This group is also in close proximity to the well-studied Gramicidin synthase from *Kutzneria albida* (phylum of *Actinobacteria*). 7-D4 displays furthermore a distant similarity (50%) to the Gramicidin synthetase (LgrC) from *Streptacidiphilus albus* (Table 2). The closest match of clone 4-B4 to a synthetase coding for a known product is the Syringopeptin synthetase b from *Photorhabdus asymbiotica* (65% aa sequence similarity; Table 2). The third group, which includes the NRPS sequences 3-F3, 7-B9, and 2-F4, comprises species from *Pseudomonas* and *Lysobacter* (*γ-Proteobacteria*). The sequence of clones 3-F3 and 2-F4 match partially to the biosynthetic pathway genes of the siderophore Pyoverdin from *Pseudomonas amygdali* and the toxin Syringomycin from *P. syringae*, respectively (61% and 57% similarity, Table 2).

Additionally, to investigate a possible overlap of NRPS sequences found by both screening methodologies the fosmid clones sequences were aligned (blastn) against the NRPS contig dataset from the *in silico* screening. Overlaps with high sequence similarity (up to 100%) were obtained only for very short DNA fragments (<15 bp). The best alignment score (63 bit, E value of 7.00E-12) was found for the fosmid clone 7-B9 and the scaffold30678 with a sequence similarity of 91% (alignment length of 47 bp).

Applied and Environmental Microbiology

Discussion

Our strategy to identify NRPS and PKS genes within moss-associated bacteria using two different approaches resulted in new findings. By *in silico* data mining we gained a valuable insight into the abundance and origin of NRPS and PKS genes present in the *Sphagnum* moss microbiome. Our hypothesis that the *Sphagnum* microbiome is a promising sources for novel NRPS and PKS genes based on ecological knowledge (4, 8) was fulfilled. The biological activity of bryophytes is well known (9) and several natural compounds have been elucidated including antibiotics, antifungals or cytotoxic compounds (10). We could show that the associated microbiota of *Sphagnum* has the biosynthetic potential to synthetize a significant amount of natural products by NRPS and PKS systems. In fact, the previous functional analysis of the *Sphagnum* metagenome revealed a high availability of subsystems that are responsible for the synthesis of bioactive compounds, such as quorum sensing molecules, toxins-antitoxins, adhesins, and especially sideropohores (4). Siderophore production and antibiotic/antifungal activity has been detected in many bacteria isolated from *Sphagnum* sp. (5, 6, 28, 29), although none microbial bioactive compounds has been isolated so far. Toxins, siderophores and antibiotics are commonly synthetized by NRPS/PKS systems (30). This compounds are often involved in characteristic reactions of microbial antagonisms, where microbes inhibit each other (antibiotics, toxins) or compete for space, nutrients and minerals (release of siderophores) in a shared microenvironment (31). A high incidence of biosynthetic systems for siderophore production was expected since *Sphagnum*-dominated peat bogs are nutrient deficient environments with low concentrations of bioavailable minerals like iron (32).

The blastx analysis revealed a significant number and also a clear difference in the bacterial diversity between NRPS and PKS sequences in the *Sphagnum* metagenome (Table 1). The composition of microbial communities derived from PKS-related sequences is strongly dominated by the genera *Mycobacterium* and *Streptomyces*. These two

Actinobacteria are well-studied producers of bioactive compounds including both NRPs and PKs (33). The synthesis of flavonoids in higher plants is nearly ubiquitous and involves the use of chalcone synthases (CHS), which belongs to the family of type III PKS (34). However, type III PKSs (CHS-like enzymes) have been also identified in bacteria (e.g. *Streptomyces griseus, S. coelicolor, Mycobacterium tuberculosis, Bacillus subtilis, Pseudomonas fluorescens*) and fungi (35). This strongly supports the possible microbial synthesis of flavonoid and aromatic polyketides (polyphenols such as stilbenes and chalcones) in mosses.

In contrast to PKS, NRPS-related sequences showed higher abundance of species belonging to *Proteobacteria* over those from *Actinobacteria*. A high abundance of protein coding sequences from *Proteobacteria* and *Actinobacteria* was expected, since our previous analysis on taxonomic structure and diversity based on 16S rRNA genes of the *Sphagnum* moss metagenome revealed a dominant role of these two phyla (65.8 % *Proteobacteria* and 5.6% *Actinobacteria*). A similar taxonomic hit distribution of *Proteobacteria* and *Actinobacteria* (62% and 8% respectively) was estimated on the basis of predicted protein coding regions and ribosomal RNA genes (4). At the class level the 16S rRNA analysis revealed high abundance of α-, β-, and γ-*Proteobacteria*, which correlates well to the high occurrence of NRPS-related sequences from this taxa (Table 1). A high portion of PKS sequences (44%, remaining strains in Table 1) is affiliated to diverse bacterial taxa, mainly *Actinobacteria*, *Cyanobacteria*, *Proteobacteria* or *Firmicutes*. Both *Cyanobacteria* and *Firmicutes* represent rather subdominant phyla in the *Sphagnum* metagenome (4).

Extensively studied bacterial sources for antibiotics are *Streptomyces*, myxobacteria, cyanobacteria, *Bacillus* and *Pseudomonas* (36). The distribution of PKS and NRPS clusters in bacterial genomes was also comprehensively reviewed by Donadio *et al*. (30). In the last years growing knowledge has been gained through sequencing of whole bacterial genomes, revealing the potential of many unexpected bacterial strains, which harbour the genetic machinery for production of secondary metabolites. One of the newly discovered groups of

350 secondary metabolites producers is the genus of *Burkholderia*. *Burkholderia* spp. synthetize
351 toxins (bongkretic acid), antifungal compounds (rhizoxin), and also antibacterial compounds
352 (enacycloxin) among others (37). Culture-dependent analysis of *Sphagnum magellanicum*-
353 associated antagonists demonstrated the dominant role of the genus *Burkholderia*, accounting
354 for 38% of the isolates (6). Besides, new moss-associated *Burkholderia* species (*B. bryophila*
355 and *B. megapolitana*) displaying antifungal activities and tested positively for the production
356 of siderophores were also described in this collection (28). Species from the plant-associated
357 *Burkholderia* cluster were identified as cosmopolitan core members of the *Sphagnum*
358 microbiome; they were present in the *Sphagnum* sporophyte as well as the gametophyte (38).
359 By *in silico* screening in the moss database we also detected a significantly high abundance of
360 NRPS sequences affiliated to *Burkholderia* (18%). Our results undermine the dominant role
361 of the genus *Burkholderia* in *Sphagnum* mosses, especially in regard to the production of
362 NRPS-synthetized secondary metabolites.

363 Furthermore, single sequences of both NRPS and PKS were affiliated to
364 cyanobacteria. Cyanobacteria are a rich source of structurally diverse oligopeptides, mostly
365 synthesized by NRPSs and NRPS/PKS hybrid pathways (39). Common cyanobacteria genus
366 like *Nostoc*, *Microcystis*, and *Anabaena*, that were present in the moss metagenome, produce
367 bioactive peptides (39). Interestingly, mixed/hybrid NRPS-PKS gene clusters were also
368 present within the moss microbiome dataset. Most of the sequences were assigned to the
369 phylum *Proteobacteria*, being the most abundant genera *Pseudomonas* and especially
370 *Lysobacter*. The latter has emerged in the last years, not only as a promising source of new
371 bioactive natural products, such as antibiotics, ß-lactams, cyclic lactams and depsipeptides,
372 but also as a biocontrol agent for fungal plant infections (36, 40, 41). Furthermore, similar
373 sequences coding for the siderophore Yersiniabactin from *P. syringae* were detected.
374 Yersiniabactin acts as a virulence factor, facilitating iron uptake in the host, and is synthetized
375 by a hybrid PKS/NRPS system located on a transmissible high-pathogenicity island. This

pathogenicity island has been encountered in various strains as a result of horizontal transfer (e.g. in enterobacteria) (26). In the case of Yersiniabactin and similar compounds, mobility by horizontal transfer to other pathogenic strains (mainly *Proteobacteria*) could account for the high incidence in metagenomic dataset, as in the *Sphagnum* moss microbiome.

Of special interest are those NRPS and PKS aa sequences with a rather low sequence identity to their next blastx neighbours or displaying similarity to sequences originated from uncultured bacteria (12 NRPS and 18 PKS sequences; Table 1). Recently, the huge potential of uncultured bacteria for the screening of novel bioactive compounds was confirmed through discovery of the novel antibiotic teixobactin, which shows promising properties against multi-resistant pathogenic strains (42).

In addition to *in silico* data mining, PCR-amplification screening of a moss metagenomic fosmid library led to identification of 13 novel NRPS-related sequences out of 14 detected clones. Only a minimal overlap between the fosmid clone sequences and the NRPS contigs from the *in silico* analysis was detected. This can be explained by the major differences between both methodologies. On the one hand, low rates for discovery of positive hits in metagenomic clone libraries are very common, in our case 0.15% (14 from 9500 screened clones). On the other hand, assemblies of metagenomic DNA usually contain poor or no coverage of complete genomes or genome portions and are prone to formation of chimeras (43). The *Sphagnum* assembly employed in this study has a rather low average contig length of 183 bp, which limits the recovery of complete genes and leads to the low observed overlap between sequences. Despite of this, the methods complement each other for studying the biosynthetic capacity of the *Sphagnum* metagenome.

In terms of taxonomy, the *in vitro* library screening reflects the findings of the *in silico* screening. Phylogenetic analysis of the retrieved aa sequences from the *Sphagnum* moss fosmid library revealed closer proximity of NRPS-related genes to the phylum *Proteobacteria*. Selected sequences clustered into three main groups, with one representative

group being closely related to *Pseudomonas and Lysobacter*. The relevance of *Lysobacter* as an "emerging" producer of bioactive compounds was discussed above. In a similar manner, novel NRPS systems were recently discovered in *Pseudomonas* spp. by new PCR-screening methods (44). We encountered partial similarity of some of the sequences (3-F3 and 2-F4) to the reported peptide synthetase products Pyoverdine and Syringomycin, originated from opportunistic pathogenic *Pseudomonas* strains. Pyoverdine, a siderophore that facilitates iron-uptake, and Syringomycin, a cyclic lipodepsipeptide phytotoxin, are both regarded as important virulence factors secreted by the host cell (45, 46). Production and release of siderophore has been reported for *Sphagnum*-associated bacteria like *Pseudomonas* sp., *Serratia* sp. and *Burkholderia* sp. (5)

The second group, containing clones 7-F1 and 6-H4, shows similarity to annotated peptide synthases from the *Caulobacteraceae* family and *Agrobacterium* spp. So far, only ribosomally encoded peptides (so-called Lasso peptides), have been isolated or described in bacteria belonging to the *Caulobacteraceae* (47). For the *Agrobacterium tumefaciens* strain C58 only one biosynthetic gene cluster has been characterized, a hybrid NRPS-PKS system that catalyses the formation of a novel siderophore (48).

In the third group, one of the NRPS-sequence (clone 4-B4) is closely related to peptide synthetases from *Burkholderia rhizoxinica* and *Photorabdus temperate*. Interestingly, both of these bacteria are symbionts of pathogenic organisms, the fungal pathogen *Rhizopus microspores* and of entomopathogenic nematodes, respectively. Complete genome sequencing of *Burkholderia rhizoxinica* showed the occurrence of 14 NRPS gene clusters with a yet unknown function (49). *Photorabdus temperata* is also known to produce a large number of bioactive compounds, especially stilbenes, where a significant proportion of the genome (6%) is devoted to the production of secondary metabolites (50). Derivatives of stilbenes have also been detected in Bryophytes (10), which demonstrates convergence between the metabolic capacity of the associated bacterium and the plant host. A lower similarity of clone 4-B4 was

found for the gene cluster producing Syringopeptin in *Photorhabdus asymbiotica*. Syringopeptin is like Syringomycin a well-known phytotoxin that is secreted by the host organism and has been studied in more detail in *Pseudomonas syringae* (51). The last two sequences in this group, belonging to clones 7-D4 and 6-B1, display similarity to annotated sequences of thioester-reductases rather than peptide synthetases. The occurrence of reductase domains in NRPS systems has been reported, for example for the peptaibol synthetase from *Trichoderma virens* that does not comprise the commonly encountered thioester domain (TE) for termination of peptide synthesis, but rather a reductase domain (52). Similarly, in the fungus *Aspergillus flavus* NRPS-like proteins that are involved in the synthesis of metabolites contain a reductase domain instead of a condensation domain (53). These sequences, especially 7-D4, show distant homology to a gene cluster coding for Gramicidin, a linear polypeptide antibiotic (toxin) that forms an ion membrane channel and has been intensively studied in *Bacillus brevis* (54).

By combining two different screening approaches, we gained an excellent overview of the taxonomic and functional composition of NRPS and PKS gene clusters within the *Sphagnum* microbiome. The *in silico* data mining approach provided a general survey on the occurrence and abundance of the NRPS and PKS genetic machinery in *Sphagnum* moss-associated bacteria. Additionally, single clones containing novel NRPS sequences were identified by PCR-amplification screening. Analysis of the amplicon sequences suggested the presence of several novel gene clusters for production of microbial metabolites, such as siderophores, phytotoxins or antibiotics. These findings are in accordance to the previous metagenomic analysis and antimicrobial assays that suggested the availability of such biosynthetic systems in *Sphagnum*. A further characterization of the identified metagenomic clones will provide a promising basis for the discovery of novel biosynthetic pathways.

Applied and Environmental Microbiology

Acknowledgements

We thank Gerhard G. Thallinger (Graz), Christin Zachow (Graz), Henry Müller (Graz), Cornelia Rainer (Graz) and Barbara Fetz (Graz) for helpful discussions and valuable technical support. We also would like to acknowledge Timothy G. Mark for English revision of the manuscript. This work has been supported by a grant to G.B. from the Federal Ministry of Economy, Family and Youth (BMWFJ), the Federal Ministry of Traffic, Innovation and Technology (bmvit), the Styrian Business Promotion Agency SFG, the Standortagentur Tirol and ZIT - Technology Agency of the City of Vienna through the COMET-Funding Program managed by the Austrian Research Promotion Agency FFG.

Data accessibility

Query sequences from the *S. magellanicum* metagenome that showed homology to NRPS, PKS or hybrid NRPS-PKS genes (from *in silico* data mining) were deposited in the DRYAD repository (http://datadryad.org/pages/repository) under the accession number xxxx.

References

1. **Berg G, Zachow C, Müller H, Philipps J, Tilcher R**. 2013. Next-generation bio-products sowing the seeds of success for sustainable agriculture. Agronomy **3**:648–656.
2. **Daniels R, Eddy A**. 1985. Handbook of Europaen *Sphagna*. Institute of Terrestrial Ecology, Natural Environment Research Council, Cambrian News, Aberystwyth, UK.
3. **Whinam J, Copson G**. 2006. *Sphagnum* moss: an indicator of climate change in the sub-Antarctic. Polar Rec **42**:43–49.
4. **Bragina A, Oberauner-Wappis L, Zachow C, Halwachs B, Thallinger GG, Müller H, Berg G**. 2014. The *Sphagnum* microbiome supports bog ecosystem functioning under extreme conditions. Mol Ecol **23**:4498–4510.

5. **Opelt K, Berg G**. 2004. Diversity and antagonistic potential of bacteria associated with Bryophytes from nutrient-poor habitats of the Baltic Sea Coast. Appl Environ Microbiol **70**:6569–6579.
6. **Opelt K, Chobot V, Hadacek F, Schönmann S, Eberl L, Berg G**. 2007. Investigations of the structure and function of bacterial communities associated with *Sphagnum* mosses. Environ Microbiol **9**:2795–2809.
7. **Bragina A, Berg C, Cardinale M, Shcherbakov A, Chebotar V, Berg G**. 2012. *Sphagnum* mosses harbour highly specific bacterial diversity during their whole lifecycle. ISME J **6**:802–813.
8. **Opelt K, Berg C, Berg G**. 2007. The bryophyte genus *Sphagnum* is a reservoir for powerful and extraordinary antagonists and potentially facultative human pathogens. FEMS Microbiol Ecol **61**:38–53.
9. **Frahm J-P**. 2004. Recent developments of commercial products from Bryophytes. The Bryologist **107**:277–283.
10. **Asakawa Y, Ludwiczuk A, Nagashima F**. 2013. Phytochemical and biological studies of bryophytes. Phytochemistry **91**:52–80.
11. **Banerjee RD, Sen SP**. 1979. Antibiotic Activity of Bryophytes. The Bryologist **82**:141–153.
12. **Basile A, Giordano S, López-Sáez JA, Cobianchi RC**. 1999. Antibacterial activity of pure flavonoids isolated from mosses. Phytochemistry **52**:1479–1482.
13. **Walsh CT**. 2007. The chemical versatility of natural-product assembly lines. Acc Chem Res **41**:4–10.
14. **Wang H, Fewer DP, Holm L, Rouhiainen L, Sivonen K**. 2014. Atlas of nonribosomal peptide and polyketide biosynthetic pathways reveals common occurrence of nonmodular enzymes. Proc Natl Acad Sci **111**:9259–9264.

15. **Strieker M, Tanović A, Marahiel MA**. 2010. Nonribosomal peptide synthetases: structures and dynamics. Curr Opin Struct Biol **20**:234–240.
16. **Staunton J, Weissman KJ**. 2001. Polyketide biosynthesis: a millennium review. Nat Prod Rep **18**:380–416.
17. **Walsh CT**. 2002. Combinatorial biosynthesis of antibiotics: challenges and opportunities. ChemBioChem **3**:124–134.
18. **Simon C, Daniel R**. 2011. Metagenomic analyses: past and future trends. Appl Environ Microbiol **77**:1153–1161.
19. **Mendes R, Garbeva P, Raaijmakers JM**. 2013. The rhizosphere microbiome: significance of plant beneficial, plant pathogenic, and human pathogenic microorganisms. FEMS Microbiol Rev **37**:634–663.
20. **Parsley LC, Consuegra EJ, Kakirde KS, Land AM, Harper WF, Liles MR**. 2010. Identification of diverse antimicrobial resistance determinants carried on bacterial, plasmid, or viral metagenomes from an activated sludge microbial assemblage. Appl Environ Microbiol **76**:3753–3757.
21. **Hodges TW, Slattery M, Olson JB**. 2012. Unique Actinomycetes from marine caves and coral reef sediments provide novel PKS and NRPS biosynthetic gene clusters. Mar Biotechnol **14**:270–280.
22. **Opelt K, Berg C, Schönmann S, Eberl L, Berg G**. 2007. High specificity but contrasting biodiversity of *Sphagnum*-associated bacterial and plant communities in bog ecosystems independent of the geographical region. ISME J **1**:502–516.
23. **Luo R, Liu B, Xie Y, Li Z, Huang W, Yuan J, He G, Chen Y, Pan Q, Liu Y, Tang J, Wu G, Zhang H, Shi Y, Liu Y, Yu C, Wang B, Lu Y, Han C, Cheung DW, Yiu S-M, Peng S, Xiaoqian Z, Liu G, Liao X, Li Y, Yang H, Wang J, Lam T-W, Wang J**. 2012.

529 SOAPdenovo2: an empirically improved memory-efficient short-read *de novo*
530 assembler. GigaScience **1**:18.
531 24. **Laskaris P**. 2009. Evolution of the streptomycin and viomycin biosynthetic clusters
532 and resistance genes. Ph.D. thesis, University of Warwick.
533 25. **Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ**.
534 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search
535 programs. Nucleic Acids Res **25**:3389–3402.
536 26. **Bultreys A, Gheysen I, Hoffmann E de**. 2006. Yersiniabactin production by
537 *Pseudomonas syringae* and *Escherichia coli*, and description of a second
538 Yersiniabactin locus evolutionary group. Appl Environ Microbiol **72**:3814–3825.
539 27. **Amos GCA, Borsetto C, Laskaris P, Krsek M, Berry AE, Newsham KK, Calvo-**
540 **Bado L, Pearce DA, Wellington EMH**. 2015. Diversity among natural product
541 synthetic genes in Antarctic and European soils. Plos One **in revision**.
542 28. **Vandamme P, Opelt K, Knöchel N, Berg C, Schönmann S, Brandt ED, Eberl L,**
543 **Falsen E, Berg G**. 2007. *Burkholderia bryophila* sp. nov. and *Burkholderia*
544 *megapolitana* sp. nov., moss-associated species with antifungal and plant-growth-
545 promoting properties. Int J Syst Evol Microbiol **57**:2228–2235.
546 29. **Shcherbakov AV, Bragina AV, Kuzmina EY, Berg C, Muntyan AN, Makarova NM,**
547 **Malfanova NV, Cardinale M, Berg G, Chebotar VK, Tikhonovich IA**. 2013.
548 Endophytic bacteria of *Sphagnum* mosses as promising objects of agricultural.
549 Microbiology **82**:306–315.
550 30. **Donadio S, Monciardini P, Sosio M**. 2007. Polyketide synthases and nonribosomal
551 peptide synthetases: the emerging view from bacterial genomics. Nat Prod Rep
552 **24**:1073.

31. **Berg G**. 2009. Plant–microbe interactions promoting plant growth and health: perspectives for controlled use of microorganisms in agriculture. Appl Microbiol Biotechnol **84**:11–18.

32. **Jin CW, Li GX, Yu XH, Zheng SJ**. 2010. Plant Fe status affects the composition of siderophore-secreting microbes in the rhizosphere. Ann Bot **105**:835–841.

33. **Chen Y, Ntai I, Ju K-S, Unger M, Zamdborg L, Robinson SJ, Doroghazi JR, Labeda DP, Metcalf WW, Kelleher NL**. 2011. A proteomic survey of nonribosomal peptide and polyketide biosynthesis in Actinobacteria. J Proteome Res **11**:85–94.

34. **Schröder J**. 1999. The chalcone/stilbene synthase-type family of condensing enzymes, p. 749–771. *In* Meth-Cohn, SDBN (ed.), Comprehensive Natural Products Chemistry. Pergamon, Oxford.

35. **Moore BS, Hopke JN**. 2001. Discovery of a new bacterial polyketide biosynthetic pathway. ChemBioChem **2**:35–38.

36. **Pidot SJ, Coyne S, Kloss F, Hertweck C**. 2014. Antibiotics from neglected bacterial sources. Int J Med Microbiol **304**:14–22.

37. **Partida-Martinez LP, Hertweck C**. 2007. A gene cluster encoding rhizoxin biosynthesis in "*Burkholderia rhizoxina*", the bacterial endosymbiont of the fungus *Rhizopus microsporus*. ChemBioChem **8**:41–45.

38. **Bragina A, Cardinale M, Berg C, Berg G**. 2013. Vertical transmission explains the specific *Burkholderia* pattern in *Sphagnum* mosses at multi-geographic scale. Front Microbiol **4**.

39. **Welker M, Von Döhren H**. 2006. Cyanobacterial peptides – Nature's own combinatorial biosynthesis. FEMS Microbiol Rev **30**:530–563.

40. **Xie Y, Wright S, Shen Y, Du L**. 2012. Bioactive natural products from *Lysobacter*. Nat Prod Rep **29**:1277.

41. **Islam MT, Hashidoko Y, Deora A, Ito T, Tahara S**. 2005. Suppression of damping-off disease in host plants by the rhizoplane bacterium *Lysobacter* sp. strain SB-K88 is linked to plant colonization and antibiosis against soilborne Peronosporomycetes. Appl Environ Microbiol **71**:3786–3796.

42. **Ling LL, Schneider T, Peoples AJ, Spoering AL, Engels I, Conlon BP, Mueller A, Schäberle TF, Hughes DE, Epstein S, Jones M, Lazarides L, Steadman VA, Cohen DR, Felix CR, Fetterman KA, Millett WP, Nitti AG, Zullo AM, Chen C, Lewis K**. 2015. A new antibiotic kills pathogens without detectable resistance. Nature **517**:455–459.

43. **Scholz MB, Lo C-C, Chain PS**. 2012. Next generation sequencing and bioinformatic bottlenecks: the current state of metagenomic data analysis. Curr Opin Biotechnol **23**:9–15.

44. **Rokni-Zadeh H, Mangas-Losada A, Mot RD**. 2011. PCR detection of novel non-ribosomal peptide synthetase genes in lipopeptide-producing *Pseudomonas*. Microb Ecol **62**:941–947.

45. **Meyer JM, Neely A, Stintzi A, Georges C, Holder IA**. 1996. Pyoverdin is essential for virulence of *Pseudomonas aeruginosa*. Infect Immun **64**:518–523.

46. **Guenzi E, Galli G, Grgurina I, Gross DC, Grandi G**. 1998. Characterization of the Syringomycin synthetase gene cluster. A link between prokaryotic and eukaryotic peptide synthetases. J Biol Chem **273**:32857–32863.

47. **Hegemann JD, Zimmermann M, Xie X, Marahiel MA**. 2013. Caulosegnins I–III: a highly diverse group of Lasso peptides derived from a single biosynthetic gene cluster. J Am Chem Soc **135**:210–222.

48. **Rondon MR, Ballering KS, Thomas MG**. 2004. Identification and analysis of a siderophore biosynthetic gene cluster from *Agrobacterium tumefaciens* C58. Microbiology **150**:3857–3866.

49. **Lackner G, Moebius N, Partida-Martinez L, Hertweck C**. 2011. Complete genome sequence of *Burkholderia rhizoxinica*, an endosymbiont of *Rhizopus microsporus*. J Bacteriol **193**:783–784.

50. **Clarke DJ**. 2008. *Photorhabdus*: a model for the analysis of pathogenicity and mutualism. Cell Microbiol **10**:2159–2167.

51. **Bender CL, Alarcón-Chaidez F, Gross DC**. 1999. *Pseudomonas syringae* phytotoxins: mode of action, regulation, and biosynthesis by peptide and polyketide synthetases. Microbiol Mol Biol Rev **63**:266–292.

52. **Manavalan B, Murugapiran SK, Lee G, Choi S**. 2010. Molecular modeling of the reductase domain to elucidate the reaction mechanism of reduction of peptidyl thioester into its corresponding alcohol in non-ribosomal peptide synthetases. BMC Struct Biol **10**:1.

53. **Forseth RR, Amaike S, Schwenk D, Affeldt KJ, Hoffmeister D, Schroeder FC, Keller NP**. 2013. Homologous NRPS-like gene clusters mediate redundant small-molecule biosynthesis in *Aspergillus flavus*. Angew Chem Int Ed **52**:1590–1594.

54. **Wallace BA**. 1990. Gramicidin channels and pores. Annu Rev Biophys Biophys Chem **19**:127–157.

Figures and Tables

Table 1 Abundance and occurrence of phylas of NRPS, PKS and NRPS-PKS hybrid sequences from *in silico* data mining in the *Sphagnum* moss metagenomic database (counts >10 are highlighted in bold, the highest values are underlined).

Taxon		No. of occurrences		
Phylum/Class	**Genus**	**NRPS**	**PKS**	**Hybrids**
Actinobacteria	*Mycobacterium*	5	**92**	0
Actinobacteria	*Streptomyces*	**25**	**46**	2
Actinobacteria	*Rhodococcus*	**10**	1	0
Actinobacteria	*Brachybacterium*	0	0	3
α-Proteobacteria	*Bradyrhizobium*	**10**	4	3
β-Proteobacteria	*Burkholderia*	**50**	5	0
β-Proteobacteria	*Rubrivivax*	0	6	0
β-Proteobacteria	*Bordetella*	0	6	0
β-Proteobacteria	*Ralstonia*	2	0	4
γ-Proteobacteria	*Pseudomonas*	**40**	2	7
γ-Proteobacteria	*Lysobacter*	0	0	6
γ-Proteobacteria	*Pectobacterium*	**12**	0	0
γ-Proteobacteria	*Xanthomonas*	9	0	0
γ-Proteobacteria	*Xenorhabdus*	6	1	0
δ-Proteobacteria	*Myxococcus*	**27**	5	6
δ-Proteobacteria	*Candidatus*	4	2	2
Firmicutes	*Paenibacillus*	1	6	1
uncultured bacterium		**12**	**18**	1
Remaining strains		**66**	**152**	5
Total count		**279**	**346**	**40**

626 **Table 2** Blastx analysis of non-ribosomal peptide synthetase (NRPS) putative sequences obtained through PCR-based screening of the *Sphagnum* moss metagenomic fosmid
627 library. For fosmid clones 3-F3, 3-H3, 2-D4, 2-F4, 7-B9 and 6-H4 longer sequences (759-1155 bp, 252-384 aa) were retrieved by primer walking. Blastx was performed against
628 the non-redundant protein database (http://www.ncbi.nlm.nih.gov/protein; 07.04.2015).

Clone ID (Query)	Primer (Screening)	Close hits (Accession no.)	Description [source]	E-value	Score	Bit score	Alignment length [aa]	Identity [%]	Positive [%]	Gaps [%]
3-F3	NRPS1	WP_008434512	Peptide synthase [*Pseudomonas* sp. Ag1]	0	1,691	656	345	99.4	99.4	0
		WP_005745216	Pyoverdine sidechain peptide synthetase III, partial [*Pseudomonas amygdali*]	4.76E-131	1,019	397	342	60.6	73.9	0.87
7-B9	NRPS2	WP_031373175	Hypothetical protein, partial [*Lysobacter antibioticus*]	6.26E-116	969	378	320	62.9	73.8	1.25
		WP_032634710	Non-ribosomal peptide synthetase modules, partial [*Pseudomonas syringae*]	8.94E-105	845	330	322	57.2	68.0	3.08
6-H4	NRPS2	WP_031447884	Peptide synthetase [*Caulobacteraceae* bacterium PMMR1]	3.68E-114	948	370	301	64.1	76.1	0.66
2-F4	NRPS1	WP_032630858	Hypothetical protein, partial [*Pseudomonas syringae*]	4.78E-88	723	283	254	64.2	76.0	1.18
		WP_004417722	Syringomycin synthetase E, partial [*Pseudomonas syringae*]	2.55E-73	651	255	255	57.3	69.4	1.18
7-F1	NRPS2	WP_031447884	Peptide synthetase [*Caulobacteraceae* bacterium PMMR1]	1.47E-54	498	196	147	73.5	81.0	0
4-B4	NRPS2	WP_013428427	Peptide synthetase [*Burkholderia rhizoxinica*]	5.34E-48	451	178	143	70.1	78.5	0.69
		WP_015833634	Syringopeptin synthetase b [*Photorhabdus asymbiotica*]	2.32E-41	400	158	143	64.6	74.3	0.69
7-D4	NRPS2	AHX13665	Thioester reductase [*Dyella jiangningensis*]	1.10E-43	418	166	126	62.7	77.8	0
		WP_042439855	Gramicidin synthetase LgrC, partial [*Streptacidiphilus albus*]	2.08E-29	304	122	122	50	66.4	0.82
2-D4	NRPS1	WP_030110998	Hypothetical protein, partial [*Kutzneria albida*]	2.30E-41	415	165	209	47.9	58.1	3.72
8-C8	NRPS2	WP_011473791	Amino acid adenylation [*Rhodopseudomonas palustris*]	6.93E-34	346	138	156	50.3	61.6	5.03
6-B1	NRPS2	AHX13666	Amino acid adenylation protein [*Dyella jiangningensis*]	5.51E-31	323	129	103	72.8	80.6	0
3-H3	NRPS1	WP_010564295	Peptide synthetase [*Pseudomonas extremaustralis*]	1.56E-18	218	89	45	91.1	95.6	0
3-G9	NRPS1	CDG17982	Non-ribosomal peptide synthetase [*Xenorhabdus doucetiae*]	9.97E-10	151	63	39	74.4	76.9	0
2-C8	NRPS1	WP_022609339	Non-ribosomal peptide synthetase module, partial [*Rubidibacter lacunae*]	2.29E-08	138	58	29	86.2	89.7	0
7-C3	NRPS1	WP_032631609	Hypothetical protein, partial [*Pseudomonas syringae*]	0.05	88	39	20	90.0	90.0	0

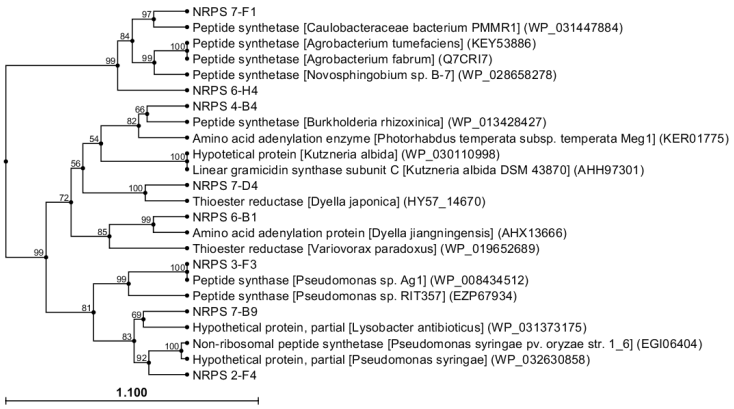

Figure 1 Phylogenetic tree of identified NRPS gene metagenomic sequences, obtained from the sequenced-based screening of a *Sphagnum* moss fosmid library. Putative NRPS sequences were aligned with reference sequences from the protein database (NCBI; accession no. in parenthesis). The tree was generated using CLC Main Workbench 6.9.1 software using the UPGMA algorithm, Kimura Protein, and a bootstrap of 1000 replicates. Bootstrap values higher than 50% are indicated at branch points. The bar indicates 1.1 substitutions per amino acid position.