AEM
Journals.ASM.org

Mining for Nonribosomal Peptide Synthetase and Polyketide Synthase Genes Revealed a High Level of 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] 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, Coventry, United Kingdom[c]

***Sphagnum* bog ecosystems are among the oldest vegetation forms harboring a specific microbial community and are known to produce an exceptionally wide variety of bioactive substances. Although the *Sphagnum* metagenome shows a rich secondary metabolism, the genes have not yet been explored. To analyze nonribosomal peptide synthetases (NRPSs) and polyketide synthases (PKSs), the diversity of NRPS and PKS genes in *Sphagnum*-associated metagenomes was investigated by *in silico* data mining and sequence-based screening (PCR amplification of 9,500 fosmid clones). The *in silico* Illumina-based metagenomic approach resulted in the identification of 279 NRPSs and 346 PKSs, as well as 40 PKS-NRPS hybrid gene sequences. The occurrence of NRPS sequences was strongly dominated by the members of the *Protebacteria* phylum, especially by species of the *Burkholderia* genus, while PKS sequences were mainly affiliated with *Actinobacteria*. Thirteen novel NRPS-related sequences were identified by PCR amplification screening, displaying amino acid identities of 48% to 91% to annotated sequences of members of the phyla *Proteobacteria*, *Actinobacteria*, and *Cyanobacteria*. Some of the identified metagenomic clones showed the closest similarity to peptide synthases from *Burkholderia* or *Lysobacter*, which are emerging bacterial sources of as-yet-undescribed bioactive metabolites. This report 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.**

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 are of 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 plant-microbe interactions as well as the ecology of plant-associated bacteria has been reported previously (4). *Sphagnum* mosses in particular are characterized by a specific but diverse microbial community (4, 5–7) whose members 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 abundances of functional systems that are responsible for oxidative and drought stress responses 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 colonization (8). Bryophytes have been traditionally used in China and India and among Native Americans for their antifungal properties, and *Sphagnum* moss was employed as a natural disinfectant for diapers 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 species revealed that *Sphagnum* moss harbors an extraordinary high proportion of antifungal strains, as well as a lower proportion of antibacterial strains, which can partly explain the medicinal use (8). However, the major part (97%) of microbial communities associated with *Sphagnum* mosses belongs to noncultivable forms (7). Therefore, the antimicrobial potential of the moss microbiome remains mostly unexplored.

Prominent classes of active compounds of microbial and plant origin (antibiotics, antifungals, and antitumor agents) are synthesized by large multimodular enzymes, the nonribosomal peptide synthetases (NRPSs) and polyketide synthases (PKSs) or hybrids thereof (13). The widespread occurrence of the NRPS and PKS genetic machinery across all three domains of life (bacteria, ar-

Received 24 February 2015 **Accepted** 12 May 2015

Accepted manuscript posted online 22 May 2015

Citation Müller CA, Oberauner-Wappis L, Peyman A, Amos GCA, Wellington EMH, Berg G. 2015. Mining for nonribosomal peptide synthetase and polyketide synthase genes revealed a high level of diversity in the *Sphagnum* bog metagenome. Appl Environ Microbiol 81:5064–5072. doi:10.1128/AEM.00631-15.

Editor: V. Müller

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

C.A.M. and L.O.-W. contributed equally to this article.

Supplemental material for this article may be found at http://dx.doi.org/10.1128/AEM.00631-15.

doi:10.1128/AEM.00631-15

chaea, and eukarya) has been reported previously (14). Bacteria host the majority of the described NRPS/PKS gene clusters, which are especially common in the phyla *Proteobacteria*, *Actinobacteria*, *Firmicutes*, and *Cyanobacteria* (14). Both NRPSs and PKSs are regarded as multienzymatic megasynthases (200 to 2,000 kDa), organized in a modular assembly-line fashion, that contain catalytic modules for single rounds of chain elongation and variable modifications of the intermediate product (13). In nonribosomal peptides (NRPs), defined monomers, amino acids or nonproteinogenic monomers, are incorporated by specific modules consisting of three essential catalytic domains. The adenylation (A) domain catalyzes the activation of the amino acid, which is then transferred to the peptidyl carrier protein (PCP), followed by condensation of the bound amino acid (condensation [C] domain) (15). In a similar way, PKS megaenzymes consist of an acyltransferase (AT) domain for selection of the monomer substrate, usually malonyl- or methyl-malonyl-coenzyme A (CoA), priming it to the acyl carrier protein (ACP), followed by chain elongation and condensation (C-C-bond formation) by a ketosynthase (KS) domain (16). In addition to the core domains, a variable set of domains for further modifications of the peptide chain (epimerization, heterocyclization) (15) or the polyketide chain (ketoreduction, dehydration) are available (16). Termination of the chain is catalyzed in both NRPSs and PKSs by a thioester (TE) domain. Because of structural and functional similarities between the elements of each class, NRPSs and PKSs 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 permits 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 and PKS genes of bacterial communities in soil (20) and marine environments (21) but not in plant-associated microhabitats or the 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 bacterial species, 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*. The 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.

MATERIALS AND METHODS

***In silico* analysis of the *Sphagnum* moss data set.** The recently published metagenomic data set 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 a quality score of less than Q30 on average (error probability of 0.001) and poly(N) nucleotide tails] using PRINSEQ software according to the manual (http://prinseq.sourceforge.net). The normalized data set consisted of single-stranded DNA (ssDNA) sequences after treatment and separation by hydroxyapatite chromatography (4). Untreated and normalized data sets were pooled, and the generated mixed data set was used for *de novo* assembly with SOAPdenovo2 software (http://soap.genomics.org.cn/soapdenovo.html) using default parameters for metagenomic data sets (23). Briefly, the SOAPec correction tool was used first to filter short reads (kmer size of ≤17, quality value of 33, thread value 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 insertion size of 200 bp, a read length cutoff value of 100 bp, a paired-end cutoff value of 3 bp, and a minimum alignment length of 32 bp. The resulting fasta file from *de novo* assembly, including the assembled scaffolds and contigs, was employed as a 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 abundances of the genes encoding NRPS and PKS with the abundances of other commonly found microbial genes, a search was performed for monooxygenases and *rpoD* RNA polymerases.

Sampling and total community DNA isolation. Sampling of *Sphagnum magellanicum* gametophytes and preparation of the microbiome for total community DNA isolation (enrichment of the microbial fraction and removal of plant debris) were performed as reported by Bragina et al. (4). To construct the fosmid library for PCR-based screening, total community DNA was extracted using a 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* moss microbiome was constructed using a CopyControl fosmid library production kit (Epicentre) as described in 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 insertion DNA). The ligated DNA was packaged with MaxPlax Lambda packaging extracts. The packaged phage particles were employed to infect *Escherichia 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, 9,500 clones were randomly transferred to 96-well microtiter plates (MTPs) containing 150 μl of LB medium with chloramphenicol (12.5 μg ml^{-1}) using sterile tooth picks, with each plate consisting of 95 different clones and 1 negative control (medium only). MTP cultures were grown at 37°C overnight with shaking at 225 rpm and were finally stored at −70°C after addition of glycerol to each well to achieve a final concentration of 25% (vol/vol). To estimate the average insertion size in the fosmid clones, restriction digestion was performed with BamHI. For the PCR screening, 10 clones were pooled in LB medium (12.5 μg ml^{-1} chloramphenicol) for a total of 10 MTP pools. The pooled MTPs were cultivated under the conditions described above, upon addition of 1× Fosmid Autoinduction Solution (Epicentre) to induce high copy numbers. Denaturation (15 min at 99°C) of MTP cultures (diluted 1:2 in double-distilled water [ddH_2O]) and centrifugation (4,000 rpm, 5 min) were performed in order to make the fosmid DNA accessible to PCR screening with the two previously reported degenerated NRPS1 and NRPS2 primer pairs (24) which are given in Table S1 in the supplemental material. A standard PCR mixture (25 μl) contained 1× *Taq* 2× master mix (New England BioLabs, Ipswich, United Kingdom) (12.5 μl), 0.4 μM (each) primer (1 μl degenerated primer) (Sigma-Aldrich, Vienna, Austria) (see Table S1 in the supplemental material), ddH_2O (4.25 μl), 5% (vol/vol) dimethyl sulfoxide (DMSO) (1.25 μl), and 5 μl of pooled tem-

plate DNA. The following PCR program was used: 95°C for 5 min; 35 cycles of 95°C for 1 min, 57°C for 1 min, and 68°C for 1 min; and elongation at 68°C for 10 min. PCR products were analyzed by 2% agarose–Tris-acetate-EDTA (TAE) gel electrophoresis. Localization of positive clones was achieved by repetition of the PCR as described above, employing the 10 single clones from the previously identified positive MTP pool in this case.

Phylogenetic analysis of identified fosmid clones. PCR products (diluted 1:1,000) from single fosmid clones identified as positive hits during rescreening of the library were amplified with shorter nondegenerated primer pairs NRPS1ndeg and NRPS2ndeg (nested PCR), employing the above-mentioned PCR program. These primers resemble those used for library screening but lack the degenerated nucleotides in the 3′ region (see Table S1 in the supplemental material). 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, and 6-H4) were partially sequenced by primer walking using a sequence-specific primer for each (see Table S1). This allowed retrieval of longer DNA sequences contiguous to the previously identified NRPS gene region (up to 1,100 bp) directly from the fosmid clones. Analysis of the obtained sequences was performed using BLASTx (25) against the nonredundant (nr) protein sequence database at NCBI (http://www.ncbi.nlm.nih.gov/protein) or the KEGG database (http://www.genome.jp/kegg). The retrieved amino acid sequences of fosmid clones with positive test results were employed for phylogenetic analysis, together with the most similar gene sequences from the blastx search. Alignment of amino acid 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 average linkages (UPMGA), Kimura Protein as the distance measure, and a bootstrap value of 1,000 replicates.

Nucleotide sequence data. Query sequences from the *S. magellanicum* metagenome that showed homology to NRPS, PKS, or hybrid NRPS-PKS genes (from *in silico* data mining) can be found at http://dx.doi.org/10.5061/dryad.hf56v.

RESULTS

Data mining in the moss metagenomic data set. The metagenomic data set of *Sphagnum magellanicum* moss (Illumina HiSeq 2-by-100 paired-end sequencing) consists of 17,323 Mbp of raw (86,617,475 bp) and 14,141 Mbp of normalized (70,705,608 bp) metagenomic DNAs (4). *De novo* assembly of the pooled metagenome (raw and normalized reads; $n = 50$ of 199) yielded 1,062,181 contig sequences (168,393 scaffolds and 893,788 contigs), featuring a total size of 188.2 Mbp with an average length of 183 bp (see Table S2 in the supplemental material).

BLASTX analysis of the metagenomic data set revealed that NRPS, PKS, and NRPS-PKS hybrid gene sequences are present in the moss microbiome. Without cutoff settings, the blastx data set consisted of 279 NRPS, 346 PKS, and 40 hybrid or mixed-gene sequences (Table 1). This translates to a rate of 0.063% contigs containing NRPS or PKS gene sequences in the assembled *Sphagnum* metagenome (665 of 1,062,181 contigs). In comparison, sequences of other common microbial genes such as those coding for monooxygenases (3,244 contigs) and 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 subject to change 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 the closest NRPS matches belonged to species of 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, the closest matches were 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*, *Anabaena*, *Pseudanabaena*, *Microcystis*, *Fischerella*). Hybrid gene matches were mainly affiliated with the phylum *Proteobacteria* (*Pseudomonas* spp., $n = 7$, 18%; *Lysobacter* spp., $n = 6$, 15%; *Myxococcus* spp., $n = 6$, 15%).

TABLE 1 Abundance and occurrence (by phylum) of NRPS, PKS, and NRPS-PKS hybrid sequences from *in silico* data mining in the *Sphagnum* moss metagenomic database[a]

Phylum/class	Genus	No. of occurrences		
		NRPS	PKS	Hybrid
Actinobacteria	*Mycobacterium*	5	**92**	0
Actinobacteria	*Streptomyces*	**25**	**46**	2
Actinobacteria	*Rhodococcus*	**10**	1	0
Actinobacteria	*Brachybacterium*	0	0	3
Alphaproteobacteria	*Bradyrhizobium*	**10**	4	3
Betaproteobacteria	*Burkholderia*	**50**	5	0
Betaproteobacteria	*Rubrivivax*	0	6	0
Betaproteobacteria	*Bordetella*	0	6	0
Betaproteobacteria	*Ralstonia*	2	0	4
Gammaproteobacteria	*Pseudomonas*	**40**	2	7
Gammaproteobacteria	*Lysobacter*	0	0	6
Gammaproteobacteria	*Pectobacterium*	**12**	0	0
Gammaproteobacteria	*Xanthomonas*	9	0	0
Gammaproteobacteria	*Xenorhabdus*	6	1	0
Deltaproteobacteria	*Myxococcus*	**27**	5	6
Deltaproteobacteria	*Candidatus*	4	2	2
Firmicutes	*Paenibacillus*	1	6	1
Uncultured bacterium		**12**	**18**	1
Remaining strains		**66**	**152**	5
Total		**279**	**346**	**40**

[a] Counts of >10 are highlighted in bold, and the highest values are underlined.

Employing an E value cutoff of 10^{-20} (bit score > 88), it was possible to select the best matching sequences in the database. This resulted in a confined selection of 34 NRPS and 28 PKS genes, as well as three NRPS-PKS hybrids (see Table S3 in the supplemental material). These sequences displayed diverse identities to their closest neighbor from the blastx analysis ranging from 35% to 98%. Many of the selected PKS sequences with higher similarity (>60% identity) to annotated genes in the nr database are mainly related to the genus *Mycobacterium*. The remaining sequences (35% to 60% identity) show similarity to sequences of PKSs from diverse genera, such as *Streptomyces* or *Rubrivivax*. In the case of NRPS gene sequences, the most abundant genus from the closest hits was *Burkholderia*, followed by *Bradyrhizobium*, *Pseudomonas*, *Mycobacterium*, and *Pectobacterium*. One hybrid gene sequence is related to the genus *Lysobacter*, and the remaining two resemble the well-studied yersiniabactin synthase (26) from *Pseudomonas syringae*. Additionally, five of the NRPS and PKS gene sequences display the highest similarity to annotated genes from as-yet-uncultured bacteria.

Fosmid library screening. The *Sphagnum* moss fosmid library was generated employing 1 μg of metagenomic DNA, obtained from the enriched microbial fraction (1 g) contained in 200 g moss. Based on the number of clones obtained (96,025) and an average insertion size of 27 kb, the library size was estimated to be 2.6 Gb. In total, 9,500 randomly selected clones (0.26 Gb) were employed for the screening of NRPS genes by PCR amplification in MTPs using two different degenerated oligonucleotide primer pairs (NRPS1 and NRPS2). The primers were previously designed to target the adenylation domain of NRPS gene clusters in diverse soil samples (24). In total, 25 NRPS1 and 33 NRPS2 wells, each containing a pool of 10 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-hit clones were subjected to optimization of PCR conditions with shorter nondegenerated 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 sequence database (NCBI) and the KEGG database, 14 sequences showed similarity to sequences of genes encoding peptide synthases (NRPSs; Table 2) and could be therefore assigned to these protein families. Six and eight clones were detected with primer pairs NRPS1 and 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 ranged from 48% to 99% (Table 2). Most of the identified closest neighbor sequences belonged to the phylum of *Proteobacteria* (12 hits; 86%) and, in particular, to the genus of *Pseudomonas* (4 hits; Table 2). The remaining two hits included the phyla *Actinobacteria* (*Kutzneria albida*) and *Cyanobacteria* (*Rubidibacter lacunae*). Despite clone 3-F3 showing a very high amino acid sequence homology of 99.4% to a peptide synthase from *Pseudomonas* sp. strain Ag1, all 13 of the other clones harbored novel, not-yet-annotated amino acid sequences with maximal identities of up to 91%.

The obtained amino acid sequences aligned partially to conserved domains of annotated NRPS sequences, i.e., 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 Fig. S1 in the supplemental material).

To gain a better overview of the putative NRPS amino acid sequences found in the fosmid library, a phylogenetic tree was generated (Fig. 1). Sequences with a hit length of >100 amino acids (aa) and a gap value of <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 *Alphaproteobacteria*, containing NRPS sequences 7-F1 and 6-H4. These sequences are most closely related to sequences of peptide synthetases from the family of *Caulobacteraceae* and *Agrobacterium* spp. The products of these peptide synthase gene clusters are, however, still unknown. The second and more diverse group includes NRPS sequences 4-B4, 7-D4, and 6-B1, clustering in closer proximity to the sequences of peptide synthetases from *Alphaproteobacteria*, *Betaproteobacteria*, and *Gammaproteobacteria*. Sequences from clones 7-D4 and 6-B1 are close related to sequences from 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, moderate (50%) similarity 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 syringopeptin synthetase b from *Photorhabdus asymbiotica* (65% amino acid sequence similarity; Table 2). The third group, which includes NRPS sequences 3-F3, 7-B9, and 2-F4, comprises species from *Pseudomonas* and *Lysobacter* (*Gammaproteobacteria*). The sequences 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 the NRPS sequences found by the two screening methodologies, the fosmid clones sequences were aligned (using blastn) against the NRPS contig data set from the *in silico* screening. Overlaps with high (up to 100%) sequence similarity were obtained only for DNA fragments that were very short (<15 bp). The best alignment score (63 bit score; E value of 7.00E-12) was found for fosmid clone 7-B9 and scaffold 30678, with sequence similarity of 91% (alignment length of 47 bp).

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 valuable insight into the abundance and origin of NRPS and PKS genes present in the *Sphagnum* moss microbiome. Our hypothesis based on ecological knowledge (4, 8) that the *Sphagnum* microbiome is a promising source for novel NRPS and PKS genes was fulfilled. The biological activity of bryophytes is well known (9), and several natural compounds, including antibiotics, antifungals, and cytotoxic compounds, have been elucidated (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, a 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 siderophores (4). Siderophore production and antibiotic or antifungal activity have been detected in many bacteria isolated from *Sphagnum* spp. (5, 6, 28, 29), although no microbial bioactive compounds have been isolated so far. Toxins, siderophores, and antibiotics are commonly synthetized by NRPS and PKS systems (30). Such compounds are often involved in characteristic reactions involving microbial antagonisms, where microbes inhibit each other (antibiotics and 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 such as iron (32).

The blastx analysis revealed a significant number of NRPS and PKS sequences and also a clear difference between the NRPS and PKS sequences in bacterial diversity 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*. Species of these two genera of *Actinobacteria* are well-studied producers of bioactive com-

TABLE 2 BLASTX analysis of nonribosomal peptide synthetase putative sequences obtained through PCR-based screening of the *Sphagnum* moss metagenomic fosmid library[a]

Clone ID (query)	Primer pair (screening)	Close hit (accession no.)	Description [source]	E value	Score	Bit score	Alignment length (aa)	% identity	% positivity	% gaps
3-F3	NRPS1	WP_008434512	Peptide synthase [*Pseudomonas* sp. strain Ag1]	0	1,691	656	345	99.4	99.4	0
		WP_005745216	Pyoverdine side-chain 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	Nonribosomal peptide synthetase module(s), 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	Nonribosomal peptide synthetase [*Xenorhabdus doucetiae*]	9.97E−10	151	63	39	74.4	76.9	0
2-C8	NRPS1	WP_022609339	Nonribosomal peptide synthetase module, partial [*Rubidibacter lacunae*]	2.29E−08	138	58	29	86.2	89.7	0
7-C3	NRPS2	WP_032631609	Hypothetical protein, partial [*Pseudomonas syringae*]	0.05	88	39	20	90.0	90.0	0

[a] For fosmid clones 3-F3, 3-H3, 2-D4, 2-F4, 7-B9, and 6-H4, longer sequences (759 to 1,155 bp, 252 to 384 aa) were retrieved by primer walking. BLASTX searches were performed against the nonredundant protein database (http://www.ncbi.nlm.nih.gov/protein; 7 April 2015). ID, identifier; NRPS, nonribosomal peptide synthetase.

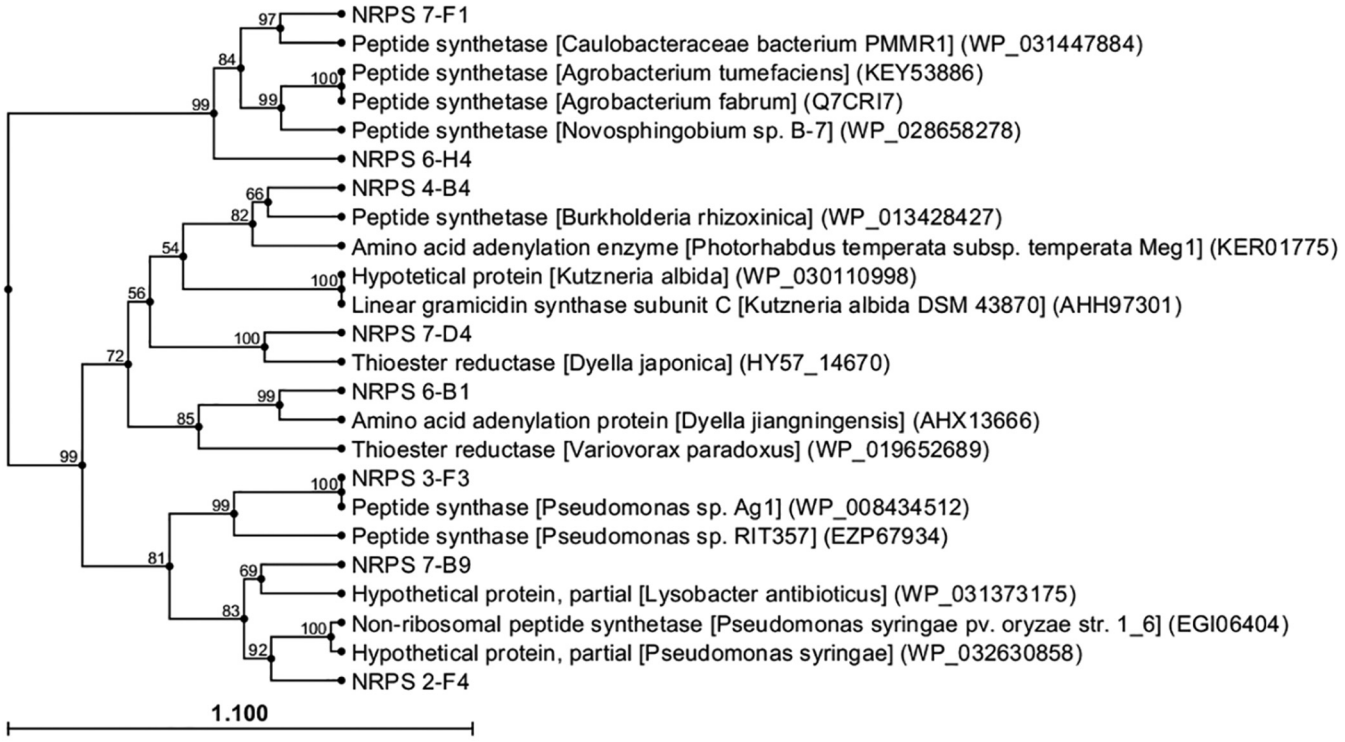

FIG 1 Phylogenetic tree of identified NRPS gene metagenomic sequences, obtained from the sequenced-based screening of a *Sphagnum* moss fosmid library. Putative NRPS sequences were aligned with reference sequences from the protein database (NCBI; accession numbers in parentheses). The tree was generated using CLC Main Workbench 6.9.1 software and the UPGMA algorithm, Kimura Protein, and a bootstrap of 1,000 replicates. Bootstrap values higher than 50% are indicated at branch points. The bar indicates 1.1 substitutions per amino acid position.

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

In contrast to PKS-related sequences, NRPS-related sequences showed higher abundance of species belonging to *Proteobacteria* than of those belonging to *Actinobacteria*. A high abundance of protein-coding sequences from *Proteobacteria* and *Actinobacteria* species was expected, since our previous analysis of 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 rRNA genes (4). At the class level, the 16S rRNA analysis revealed high abundance of *Alphaproteobacteria*, *Betaproteobacteria*, and *Gammaproteobacteria*, which correlates well to the high number of occurrences of NRPS-related sequences from species of these taxa (Table 1). A high portion of PKS sequences (44%; "Remaining strains" in Table 1) are affiliated with species of 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 recent years, knowledge has been gained through sequencing of whole bacterial genomes, revealing the potential of many unexpected bacterial strains which harbor the genetic machinery for production of secondary metabolites. One of the newly discovered groups of secondary-metabolite producers is the genus of *Burkholderia*. *Burkholderia* spp. synthetize toxins (bongkrekic acid), antifungal compounds (rhizoxin), and also antibacterial compounds (enacyloxin), among other products (36, 37). Culture-dependent analysis of *Sphagnum magellanicum*-associated antagonists demonstrated the dominant role of the genus *Burkholderia*, accounting for 38% of the isolates (6). Besides, new moss-associated *Burkholderia* species (*B. bryophila* and *B. megapolitana*) displaying antifungal activities and giving positive test results for the production of siderophores in this collection were also described previously (28). Species from the plant-associated *Burkholderia* cluster were identified as cosmopolitan core members of the *Sphagnum* microbiome; they were present in the *Sphagnum* sporophyte as well as in the gametophyte (38). By *in silico* screening in the moss database, we also detected a significantly high (18%) abundance of NRPS sequences affiliated with *Burkholderia*. Our results undermine the idea of the dominant role of the genus

Burkholderia in *Sphagnum* mosses, especially in regard to the production of NRPS-synthetized secondary metabolites.

Furthermore, single sequences of both NRPSs and PKSs were affiliated with cyanobacteria. Cyanobacteria are a rich source of structurally diverse oligopeptides, mostly synthesized by NRPS and NRPS-PKS hybrid pathways (39). Common cyanobacteria genera such as *Nostoc*, *Microcystis*, and *Anabaena* that are present in the moss metagenome produce bioactive peptides (39). Interestingly, mixed/hybrid NRPS-PKS gene clusters were also present within the moss microbiome data set. Most of the sequences were assigned to the phylum *Proteobacteria*, with the most abundant genera being *Pseudomonas* and especially *Lysobacter*. In recent years, the latter has emerged not only as a promising source of new bioactive natural products, such as antibiotics, ß-lactams, cyclic lactams, and depsipeptides, but also as a biocontrol agent for fungal plant infections (36, 40, 41). Furthermore, similar sequences coding for the siderophore yersiniabactin from *P. syringae* were detected. Yersiniabactin acts as a virulence factor, facilitating iron uptake in the host, and is synthetized by a hybrid PKS/NRPS system located on a transmissible high-pathogenicity island. This pathogenicity island has been encountered in various strains (e.g., in enterobacteria) as a result of horizontal transfer (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 the metagenomic data set, as in the *Sphagnum* moss microbiome.

Of special interest are those NRPS and PKS amino acid sequences with rather low sequence identity to their next blastx neighbors or displaying similarity to sequences that 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 multiresistant 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 of 14 detected clones. Only minimal overlap of the fosmid clone sequences and the NRPS contigs from the *in silico* analysis was detected. This can be explained by the major differences between the two methodologies. On the one hand, low rates for discovery of positive hits in metagenomic clone libraries, in our case, 0.15% (14 from 9,500 screened clones), are very common. 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 of sequences. Despite 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 amino acid sequences retrieved 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 the members of one representative group being closely related to *Pseudomonas* and *Lysobacter*. The relevance of *Lysobacter* as an "emerging" producer of bioactive compounds is 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, which 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 siderophores has been reported for *Sphagnum*-associated bacteria such as *Pseudomonas* sp., *Serratia* sp., and *Burkholderia* sp. (5).

The members of the second group, containing clones 7-F1 and 6-H4, show 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 *Agrobacterium tumefaciens* strain C58, only one biosynthetic gene cluster has been characterized, a hybrid NRPS-PKS system that catalyzes the formation of a novel siderophore (48).

In the third group, one of the NRPS sequences (clone 4-B4) is closely related to peptide synthetases from *Burkholderia rhizoxinica* and *Photorhabdus temperate*. Interestingly, both of these bacterial species are symbionts of pathogenic organisms, the fungal pathogen *Rhizopus microspores* and entomopathogenic nematodes, respectively. Complete genome sequencing of *Burkholderia rhizoxinica* showed the occurrence of 14 NRPS gene clusters with as-yet-unknown functions (49). *Photorhabdus temperata* is also known to produce a large number of bioactive compounds, especially stilbenes, where a significant proportion (6%) of the genome 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 that of the plant host. A lower similarity to clone 4-B4 was found for the gene cluster producing syringopeptin in *Photorhabdus asymbiotica*. Syringopeptin, like Syringomycin, is 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 (TE) domain 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 that of 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, and antibiotics. These findings are in accordance with 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.

ACKNOWLEDGMENTS

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 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.

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. http://dx.doi.org/10.3390/agronomy3040648.
2. **Daniels R, Eddy A.** 1985. Handbook of European *Sphagna*. Institute of Terrestrial Ecology, Natural Environment Research Council, Cambrian News, Aberystwyth, United Kingdom.
3. **Whinam J, Copson G.** 2006. *Sphagnum* moss: an indicator of climate change in the sub-Antarctic. Polar Rec **42:**43–49. http://dx.doi.org/10.1017/S0032247405004900.
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. http://dx.doi.org/10.1111/mec.12885.
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. http://dx.doi.org/10.1128/AEM.70.11.6569-6579.2004.
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. http://dx.doi.org/10.1111/j.1462-2920.2007.01391.x.
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. http://dx.doi.org/10.1038/ismej.2011.151.
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. http://dx.doi.org/10.1111/j.1574-6941.2007.00323.x.
9. **Frahm J-P.** 2004. Recent developments of commercial products from bryophytes. Bryologist **107:**277–283. http://dx.doi.org/10.1639/0007-2745(2004)107[0277:RDOCPF]2.0.CO;2.
10. **Asakawa Y, Ludwiczuk A, Nagashima F.** 2013. Phytochemical and biological studies of bryophytes. Phytochemistry **91:**52–80. http://dx.doi.org/10.1016/j.phytochem.2012.04.012.
11. **Banerjee RD, Sen SP.** 1979. Antibiotic activity of bryophytes. Bryologist **82:**141–153. http://dx.doi.org/10.2307/3242073.
12. **Basile A, Giordano S, López-Sáez JA, Cobianchi RC.** 1999. Antibacterial activity of pure flavonoids isolated from mosses. Phytochemistry **52:**1479–1482. http://dx.doi.org/10.1016/S0031-9422(99)00286-1.
13. **Walsh CT.** 2008. The chemical versatility of natural-product assembly lines. Acc Chem Res **41:**4–10. http://dx.doi.org/10.1021/ar7000414.
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 U S A **111:**9259–9264. http://dx.doi.org/10.1073/pnas.1401734111.
15. **Strieker M, Tanović A, Marahiel MA.** 2010. Nonribosomal peptide synthetases: structures and dynamics. Curr Opin Struct Biol **20:**234–240. http://dx.doi.org/10.1016/j.sbi.2010.01.009.
16. **Staunton J, Weissman KJ.** 2001. Polyketide biosynthesis: a millennium review. Nat Prod Rep **18:**380–416. http://dx.doi.org/10.1039/a909079g.
17. **Walsh CT.** 2002. Combinatorial biosynthesis of antibiotics: challenges and opportunities. Chembiochem **3:**125–134. http://dx.doi.org/10.1002/1439-7633(20020301)3:2/3<124::AID-CBIC124>3.0.CO;2-J.
18. **Simon C, Daniel R.** 2011. Metagenomic analyses: past and future trends. Appl Environ Microbiol **77:**1153–1161. http://dx.doi.org/10.1128/AEM.02345-10.
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. http://dx.doi.org/10.1111/1574-6976.12028.
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. http://dx.doi.org/10.1128/AEM.03080-09.
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. http://dx.doi.org/10.1007/s10126-011-9410-7.
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. http://dx.doi.org/10.1038/ismej.2007.58.
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. SOAPdenovo2: an empirically improved memory-efficient short-read *de novo* assembler. GigaScience **1:**18. http://dx.doi.org/10.1186/2047-217X-1-18.
24. **Laskaris P.** 2009. Evolution of the streptomycin and viomycin biosynthetic clusters and resistance genes. Ph.D. thesis. University of Warwick, Coventry, England.
25. **Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ.** 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res **25:**3389–3402. http://dx.doi.org/10.1093/nar/25.17.3389.
26. **Bultreys A, Gheysen I, de Hoffmann E.** 2006. Yersiniabactin production by *Pseudomonas syringae* and *Escherichia coli*, and description of a second yersiniabactin locus evolutionary group. Appl Environ Microbiol **72:**3814–3825. http://dx.doi.org/10.1128/AEM.00119-06.
27. Reference deleted.
28. **Vandamme P, Opelt K, Knöchel N, Berg C, Schönmann S, Brandt ED, Eberl L, Falsen E, Berg G.** 2007. *Burkholderia bryophila* sp. nov. and *Burkholderia megapolitana* sp. nov., moss-associated species with antifungal and plant-growth-promoting properties. Int J Syst Evol Microbiol **57**(Pt 10)**:**2228–2235. http://dx.doi.org/10.1099/ijs.0.65142-0.
29. **Shcherbakov AV, Bragina AV, Kuzmina EY, Berg C, Muntyan AN, Makarova NM, Malfanova NV, Cardinale M, Berg G, Chebotar VK, Tikhonovich IA.** 2013. Endophytic bacteria of *Sphagnum* mosses as promising objects of agricultural microbiology. Microbiology **82:**306–315. http://dx.doi.org/10.1134/S0026261713030107.
30. **Donadio S, Monciardini P, Sosio M.** 2007. Polyketide synthases and nonribosomal peptide synthetases: the emerging view from bacterial genomics. Nat Prod Rep **24:**1073. http://dx.doi.org/10.1039/b514050c.
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. http://dx.doi.org/10.1007/s00253-009-2092-7.
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. http://dx.doi.org/10.1093/aob/mcq071.
33. **Chen Y, Ntai I, Ju K-S, Unger M, Zamdborg L, Robinson SJ, Doroghazi JR, Labeda DP, Metcalf WW, Kelleher NL.** 2012. A proteomic survey of nonribosomal peptide and polyketide biosynthesis in Actinobacteria. J Proteome Res **11:**85–94. http://dx.doi.org/10.1021/pr2009115.
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, United Kingdom.
35. **Moore BS, Hopke JN.** 2001. Discovery of a new bacterial polyketide biosynthetic pathway. Chembiochem **2:**35–38. http://dx.doi.org/10.1002/1439-7633(20010105)2:1<35::AID-CBIC35>3.0.CO;2-1.

36. **Pidot SJ, Coyne S, Kloss F, Hertweck C.** 2014. Antibiotics from neglected bacterial sources. Int J Med Microbiol **304:**14–22. http://dx.doi.org/10.1016/j.ijmm.2013.08.011.
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. http://dx.doi.org/10.1002/cbic.200600393.
38. **Bragina A, Cardinale M, Berg C, Berg G.** 18 December 2013, posting date. Vertical transmission explains the specific *Burkholderia* pattern in *Sphagnum* mosses at multi-geographic scale. Front Microbiol http://dx.doi.org/10.3389/fmicb.2013.00394.
39. **Welker M, Von Döhren H.** 2006. Cyanobacterial peptides—nature's own combinatorial biosynthesis. FEMS Microbiol Rev **30:**530–563. http://dx.doi.org/10.1111/j.1574-6976.2006.00022.x.
40. **Xie Y, Wright S, Shen Y, Du L.** 2012. Bioactive natural products from *Lysobacter*. Nat Prod Rep **29:**1277. http://dx.doi.org/10.1039/c2np20064c.
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. http://dx.doi.org/10.1128/AEM.71.7.3786-3796.2005.
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. http://dx.doi.org/10.1038/nature14098.
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. http://dx.doi.org/10.1016/j.copbio.2011.11.013.
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. http://dx.doi.org/10.1007/s00248-011-9885-9.
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. http://dx.doi.org/10.1074/jbc.273.49.32857.
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. http://dx.doi.org/10.1021/ja308173b.
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. http://dx.doi.org/10.1099/mic.0.27319-0.
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. http://dx.doi.org/10.1128/JB.01318-10.
50. **Clarke DJ.** 2008. *Photorhabdus*: a model for the analysis of pathogenicity and mutualism. Cell Microbiol **10:**2159–2167. http://dx.doi.org/10.1111/j.1462-5822.2008.01209.x.
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. http://dx.doi.org/10.1186/1472-6807-10-1.
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 Engl **52:**1590–1594. http://dx.doi.org/10.1002/anie.201207456.
54. **Wallace BA.** 1990. Gramicidin channels and pores. Annu Rev Biophys Biophys Chem **19:**127–157. http://dx.doi.org/10.1146/annurev.bb.19.060190.001015.