Secondary metabolite biosynthetic gene diversity – Bacillales non-ribosomal peptides and polyketides in plant-microbe interactions

Dissertation
Submitted to obtain the doctoral degree at the University of Natural Resources and Life Sciences, Vienna in association with AIT Austrian Institute of Technology GmbH, Bioresources Unit, Tulln

Advisor: Priv.-Doz. DI Dr. Angela Sessitsch
AIT Austrian Institute of Technology GmbH, Bioresources Unit

Submitted by
Gajender Aleti

Vienna, August 2016

Acknowledgements

I express my gratitude to Dr. Angela Sessitsch for providing me the opportunity to pursue PhD at Austrian Institute of Technology. You have been an amazing mentor and I have always appreciated your prompt response and timely help with the proof reading of manuscripts. I would like to cherish your continuous support and encouragement.

I am very grateful to my supervisor, Dr. Günter Brader for his guidance, patience, motivation and sharing knowledge. Your friendly approach has always made it easy to openly discuss any issue, which I always admire. You have encouraged me during my hardship and always made yourself available and guided me to accomplish desired results. I am glad to have gained research experience under your supervision. You have been a great support throughout.

I would also like to thank Dr. Friederike Trognitz for her valuable suggestions and support. You have helped boosting my confidence with your positive energy.

Many thanks to Dr. Birgit Mitter, Dr. Evelyn Hackl and Dr. Tanja Kostic for their valuable suggestions in team meetings.

My heartfelt gratitude to Prof. Helen White-Cooper and Prof. Jim Murray from Cardiff University, who encouraged me to pursue a PhD and believed in my ability to grow towards research. You gave me a keen insight into academic research settings and I believe that it helped me immensely during my PhD tenure.

Special thanks to Dr. Janet Horrocks, Dr. Yusuf Deeni and Dr. Simon Plummer my mentors during the Masters study from Abertay University for supporting my career choice.

I wish to thank Dr. Markus Bacher, Dr. Antonielli Livio, Maja Plesko, Erika Corretto, Dr. Stefan Pfeiffer, Branislav Nikolić and Dr. Stéphane Compant for their contribution in the research projects. You all have made it possible to submit most of the work by estimated time.

Special regards to all my colleagues at AIT. You all have made my tenure at AIT a memorable one. Without you all the fun part at AIT would not have been possible.

I extend my deepest gratitude to my father Narasimha Aleti and mother Andalu Aleti and my extended family for their love and support. Despite the struggles and challenges, they have always encouraged me to dream big. I would also like to thank my wife Subashini Aleti for believing in me and for her unconditional love and support.

This work was supported by the life science project LS11-014 of the NFB, Niederösterreichische Forschungs- und Bildungs-G.m.b.H.

Table of Contents

List of Tables

List of Figures

Abstract

This study aimed to investigate the biosynthetic diversity of non-ribosomal peptides and polyketides in Bacillales due to their high capacity to produce secondary metabolites and their use in the biological control of plant diseases. Genome-mining in the Bacillales suggested that a substantial fraction of the predicted non-ribosomal peptides and polyketides are uncharacterized in plant-associated *Bacillus* and *Paenibacillus* strains. Surprisingly, many genera of Bacillales from other environments produce few of such compounds indicating the importance of these metabolites in plant-associated niches. The genome of *Paenibacillus polymyxa* strain CCI-25 encompasses genes encoding fusaricidin C, iturin-like, tridecaptin and polymyxin variants with altered monomer composition, a lantibiotic similar to paenicidin A, and a polyketide synthase type 1. Given the fact that 6.6% of the total genome is devoted to secondary metabolite biosynthesis, CCI-25 has high potential to be exploited for medical or agricultural applications. *Bacillus atrophaeus* strain 176s protected plants from *Rhizoctonia solani* infection and co-produced three lipopeptide families, which may play a role in biocontrol of plant pathogens. We isolated surfactin C from *B. atrophaeus* with subtle structural differences compared to surfactin A from *B. subtilis* and *B. amyloliquefaciens*. The dissimilarity is encoded in an adenylation domain of the surfactin synthetase and importantly, surfactin variants are distributed in a species-specific manner in all *Bacillus*. Further, the surfactin variants were associated with species-specific biofilm induction and root colonization.

The results of this thesis show that there is a huge yet untapped potential of secondary metabolites and their genetic diversity in the genera *Bacillus* and *Paenibacillus,* which may play a key role in plant-microbe interactions.

Zusammenfassung

Diese Studie untersucht die biosynthetische Vielfalt von nicht-ribosomalen Peptiden und Polyketiden in Bacillales auf Grund ihrer weit verbreiteten Kapazität Sekundärmetabolite zu produzieren und ihrem Einsatz in der biologischen Schädlingsbekämpfung. Genome-Mining in Bacillales weist darauf hin, dass ein wesentlicher Teil der vorhergesagten nicht-ribosomalen Peptide und Polyketide in Pflanzen-assoziierten *Bacillus* und *Paenibacillus* Stämmen uncharakterisiert ist. Überraschenderweise produzieren viele Bacillales Gattungen aus anderen Habitaten nur wenige solcher Verbindungen, was auf die Wichtigkeit dieser Metabolite in Pflanzen-assoziierten Nischen hinweist. Das Genom von *Paenibacillus polymyxa* CC1-25 enthält Gene, die für Fusaricidin C, Iturin-ähnliche, Tridecaptin und Polymyxin Varianten mit veränderter Monomerzusammensetzung kodieren. Ebenso sind Gene für die Produktion eines Paenicidin A - ähnlichem Lantibiotikum und eine Polysynthetase 1 enthalten. Angesichts der Tatsache, dass 6,6 % des gesamten Genoms der Synthese von Sekundärmethaboliten gewidmet ist, bietet CCI-25 ein hohes Potenzial für medizinische und landwirtschaftliche Anwendungen. *Bacillus atrophaeus* 176s schützt Pflanzen von *Rhizoctonia solani* Infektionen. Dieser Stamm produziert drei Lipopeptidfamilien, die in der Biokontrolle von pflanzlichen Pathogenen eine Rolle spielen könnten. Wir isolierten Surfactin C von *B. atrophaeus* welches subtile strukturelle Unterschiede in der Adenylierungsdomäne der Surfactinsynthetase im Vergleich zu Surfactin A von *B. subtilis* und *B. amyloliquefaciens* aufweist. Abweichungen in der Surfactinsynthetase sind über alle *Bacillus* Arten verteilt. Ferner waren die Surfactin Varianten mit Art-spezifischer Biofilmbildung und Wurzelbesiedelung assoziiert. Die Ergebnisse dieser Doktorarbeit zeigen das große, dennoch ungenutzte Potenzial von Sekundärmetaboliten und deren genetischer Vielfalt in den Gattungen *Bacillus* und *Paenibacillus*, die eine Schlüsselfunktion in Pflanzen-Mikroben-Interaktion darstellen könnten.

General Introduction

1.1 *Bacillus* and *Paenibacillus* as source of bioactive molecules

Infectious diseases of plants have long been a major threat to the food safety. Given the fact that the current world population is growing at an annual rate of 77 million people per year, the world's population is estimated to reach 9.1 billion by 2050, which is 34 percent higher than today (FAO 2009). Over the next years, crop production will have to increase rapidly to meet the needs of a rising world population and also ensuring a sustainable supply of food for such large population will be a major challenge.

Despite the fact that the agro-chemicals usage over the past half century has drastically reduced fungal infections and contributed to substantial increase in yields for most field crops (Den Herder et al., 2010), this created ecological imbalance leading to the emergence of highly resistant pathogens (Raposo et al., 2000). Irrational selection and extensive usage of certain chemical pesticides and fertilizers may have toxic effects on groundwater, on rhizosphere and soil (beneficial) microflora sharing the ecosystem and eventually toxic chemical residues may enter the food chain, and subsequent accumulation in the human body may pose obvious health risks (Bartlett et al., 2002). Every year more and more chemicals are lost into the soil, as a result the efficacy of pesticides and fertilizers may be reduced leading to poor soil fertility. Therefore, alternative strategies for ecologically compatible plant disease management practices are rapidly gaining interest worldwide.

Often plant beneficial microbes or microbial derived products have been shown to display diverse modes of action for controlling plant pathogens, referred as biological control agents, and are considered environmentally friendly and have emerged as promising alternatives to chemical inputs (pesticides and fertilizers) (Mizumoto et al., 2006). Application of beneficial microorganisms could be relatively the most efficient strategy for a healthy and secured agricultural practices where chemical solutions are ineffective and where conventional pesticides cannot be used owing to residue concerns, or in the rapidly growing sector of organic farming. The overall contribution of biopesticides for plant health management at present is relatively small, which represents ~2.5% of total agricultural sales (Ongena and Jacques, 2008). However, global market for biopesticides is estimated to reach billion dollars in near future.

Many studies have extensively investigated the potential of rhizosphere-associated bacteria for plant growth promotion and biocontrol activities against plant diseases caused by soil-borne and post-harvest pathogens (Romero et al., 2007; Ongena and Jacques, 2008; Chen et al., 2009; Kim et al., 2010). Representatives of the Gram-positive *Bacillus* and *Paenibacillus* genera encompass important plant beneficial strains that are noted for excellent plant colonization ability, plant growth promoting activities and for the production of array of bioactive compounds for effective biocontrol of plant pathogens (Ongena and Jacques, 2008; Raijmakers 2009; Borriss, 2011). For instance, one of the most commonly used and well-studied rhizobacterium, *B. subtilis*, devotes 4–5% of its genome for antibiotic biosynthesis and has the potential to produce non-ribosomally more than two dozen structurally diverse antimicrobial compounds including lipopeptides and polyketides (Stein, 2005). Bacilli dominate the biopesticide market to this day because of their excellent spore forming ability that allows the preparation of stable bioformulations with a long shelf-life (Borriss, 2015). Commercially available *Bacillus* based products for agriculture are listed in Table1.

The plant growth promoting bacteria (PGPB) *Paenibacillus polymyxa* (formerly *Bacillus polymyxa*) induce plant growth promotion largely by producing phytohormones, including indole acetic acid, cytokinins, gibberellins, ethylene, and volatile compounds, and like *Bacillus* have potential to produce many bioactive secondary metabolites that have antibacterial and antifungal activities.

Table 1. Examples of some commercially available *Bacillus* derived products for plant protection (adapted from Borris, 2015)

Trade name	*Bacillus* strain	Known properties	Company
Kodiak™	*B. subtilis* GB03	EPA-registered (71065–2) biological and seed treatment fungicide	Bayer Crop Science
Companion	*B. subtilis* GB03	EPA-registered (71065–2) biofungicide, Produces iturin and triggers induced systemic resistance (ISR) to control plant diseases	Growth Products Ltd
Yield Shield	*B. pumilus* GB34 (=INR7)	EPA-registered biofungicide (264–985), Suppression of root diseases caused by *Rhizoctonia* and *Fusarium*	Bayer Crop Science
BioYield™	*B. amyloliquefaciens* GB99 + *Bacillus subtilis* GB122	Combination of strong ISR activity (GB99) with phytostimulaton (GB122)	Bayer Crop Science
Subtilex®, INTEGRAL ®	*B. subtilis* MBI600	EPA-registered (71840–8.) biofungicide, protective against *Rhizoctonia solani*, *Pythium* spp. and *Fusarium* spp.	BASF
VAULT®	*B subtilis* MBI601	Produced by "BioStacked®" technology, enhancing growth of soybeans and peanuts	Becker Underwood
	B. pumilus BU F-33	EPA-registered (71840-RG,-RE,2013) plant growth stimulator, induced systemic resistance	Becker Underwood
SERENADE Max	*B. subtilis* QST713	EPA-registered (69592–11) biofungicide, Annex1 listing of the EU agrochemical registration directive (91/414)	Bayer Crop Science
SERENADE SOIL(R)	*B. subtilis* QST713	EPA-registered (69592-EI, 2012) biofungicide for food crop	Bayer Crop Science
SERENADE Optimum®	*B. subtilis* QST713	EPA-registered (2013) Active against fungal (*Botrytis*, *Sclerotinia*), and bacterial pathogens (*Xanthomonas* and *Erwinia*)	Bayer Crop Science
CEASE(R)	*B. subtilis* QST713	Aqueous suspension biofungicide, Recommended for leafy and fruiting vegetables, herbs and spices, and ornamentals	Bio Works
SONATA®	*B. pumilus* QST2808	EPA-registered (69592–13) biofungicide, powdery mildew control	Bayer Crop Science
Rhizo Vital®	*B. amyloliquefaciens* FZB42	Biofertilizer, plant growth promoting activity, provides protection against various soil borne diseases, stimulation of ISR	ABiTEP GmbH
RhizoPlus®	*B. subtilis*	Plant growth-promoting rhizobacterium and biocontrol agent. It can be used for potatoes, corn, vegetables, fruits and also turf	ABiTEP GmbH
Taegro®	*B. subtilis* FZB24	EPA-registered biofungicide	Syngenta
POMEX	*B. subtilis* CMB26	Microbial fungicide, control and inhibition germination effect on powdery mildew, *Cladosporium fulvum* and *Botrytis cinerea*	NIN Co.Ltd
	B. subtilis CX9060	EPA-registered 71840-RG, -RE (2012) fungicide, bactericide for food crops, turf and ornamentals	Certis Columbia

Easy Start® TE-Max	*B. subtilis* E4-CDX	Rhizosphere bacterium protects against soilborne fungi	COMPO Expert GmbH
Double Nickel 55™	*B. amyloliquefaciens* D747	EPA-registered (70051-RNI, 2011) biofungicide against powdery mildew, *Sclerotinia*, *Botrytis*, *Alternaria*, bacterial leaf spot, bacterial spot and speck, fire blight, *Xanthomonas*, *Monilia*	Certis Columbia
Amylo-X®	*B. amyloliquefaciens* D747	Annex 1 listing of the EU agrochemical registration directive. control of *Botrytis* and other fungal diseases of grapes, strawberries and vegetables, and bacterial diseases such as fire blight in pome fruit and PSA in kiwi fruit	Certis Columbia
BmJ WG	*B. mycoides* BmJ	It works entirely as a microbial SAR activator with no direct effect on the plant pathogen itself. Under development	Certis Columbia
	B. pumilus GHA181	EPA-registered fungicide (2012), food crops, seeds, ground cover, and ornamentals	Premier Horticulture
BioNem	*B. firmus* GB-126	EPA-registered (2008), suppressing plant pathogenic nematodes	Agro Green

1.2 Secondary metabolite biosynthesis

Bacteria produce metabolites either by classical pathway i.e. ribosomally or by a ribosome independent pathway involving mainly non-ribosomal peptide synthetase (NRPS) or polyketide synthases (PKS) or their hybrids (NRPS/PKS). The non-ribosomally produced products clearly differ from the ribosomally synthesized structures by encompassing non-proteogenic amino acids, D-isomers and macrocyclic (branched) structures. Unlike primary metabolites, they are not directly involved in the survival of an organism but they exhibit a wide-range of biological activity (Finking and Marahiel, 2004; Walsh, 2004).

1.3 Non-ribosomal peptide biosynthetic pathway

Non-ribosomal peptide synthetase (NRPS) encompass large multi-enzymatic proteins which are typically organized into modules. These modular proteins synthesize several hundred bioactive secondary metabolites so called non-ribosomal peptides (Walsh, 2004). NRPS catalyze repeated condensation of amino acids in a succession manner for the synthesis of non-ribosomal peptides (Fischbach and Walsh, 2006; Finking and Marahiel, 2004). Remarkably, NRPS can recruit natural amino acids from a pool of 500 potential building units of non-ribosomal peptides (nrps), of which 20 (a minor fraction of 4%) constitute proteinogenic amino acids and the remaining 96% constitute non-proteinogenic amino acids. The presence of epimerization domain in the NRPS modifying the building blocks from L-monomer into its D-isomer are important for the structural heterogeneity and diverse biological activity displayed by the nrps. In addition, fatty acids and α-hydroxy acids attached to the amino acid core chain further contributes to further structural diversity (Caboche et al., 2008).

1.3.1 The core catalytic domains

Often NRPS are organized into several iterative functional units known as modules (which usually comprise ~1000 amino acid residues) and each module is crucially involved in the incorporation of the activated monomer unit into the growing peptide chain. Each module further encompasses core and accessory catalytic domains that are responsible for specific enzyme activities. The four core catalytic domains crucially involved in the activity of each module of NRPS enzymes include adenylation (A), thiolation (T) and condensation (C) domains (Finking and Marahiel, 2004). In addition to these core domains, NRPS enzymes may also encompass auxiliary or accessory domains, which either act in *cis* or *trans* for post-synthetic modifications such as substrate epimerization (E), hydroxylation, methylation and heterocyclic ring formation leading to further structural heterogeneity. *Bacillus* lipopeptides are synthesized by non-ribosomal peptide synthetases (NRPSs) or hybrid polyketide synthases. A thioesterase (TE) domain is typically present in the last module to facilitate the cleavage of the thioester bond between the peptide and the last PCP domain. In many cases, thioesterase is also responsible for the cyclization of the peptide (Finking and Marahiel, 2004).

1.3.1.1 The adenylation domain and the non-ribosomal code

The A-domain is mainly responsible for the selection of the amino acid substrates, and the activation of cognate amino acids into amino acyl adenylates at the expense of ATP. The amino acyl adenylate will be incorporated into the non-ribosomal peptide chain (Dieckmann et al., 1995). The size of the A domain is around 500 to 600 amino acids, of which 10 amino acid residues are crucial for conferring the substrate specificity. This so called non-ribosomal code (or Stachelhaus code) encompassed within the 10 amino acid residue responsible for selective activation of the monomer units has been utilized by *in silico* secondary metabolite prediction tools to accurately predict the A-domain selectivity and thus, the end products can be predicted based on primary sequences (Stachelhaus et al., 1998).

The A domain activated amino acid is then transferred to the immediate peptidyl-carrier protein (PCP) domain, which is also termed thiolation (T) domain as it acts as a transporter unit (Stachelhaus et al., 1998). The activated amino acid residue is covalently tethered to its 4'phosphopantheteinic (4'PP) cofactor as thioester. Next, the cofactor is post-translationally transferred to a serine within the PCP domain, which is carried out via phosphopantetheinyl transferase encoded by the *sfp* gene in *Bacillus subtilis*, *sfp* is crucially involved in the transformation of the inactive apoform of PCP domain to the active holoform (Mofid et al., 2004). This step is known as priming and initiates biosynthesis. The 4'PP cofactor functions as a flexible arm and enables the transfer of the bound amino acyl and peptidyl substrates between different catalytic domains (Stachelhaus et al., 1998).

1.3.1.2 Thiolation, condensation and thioesterase domains

In the next step of non-robosomal peptide biosynthesis, the C-domain catalyzes the condensation reaction resulting in a peptide bond formation between the two amino acyl substrates linked to PCPs of adjacent modules (Stachelhaus et., 1998). The C-domain is generally absent within the initiation module.

The termination of non-ribosomal peptide biosynthesis can be achieved by the terminal TE domain of the last module, which facilitates release of the peptide. This process usually generates a macrocyclic product (lactones and lactams) or the oligomerisation of peptide units (Kopp and Marahiel 2007). Alternatively, linear peptides from NRPS can be achieved by the reduction of the peptidyl-S-PCP which results in the release of a linear aldehyde or alcohol.

The last NRPS module in the row is an additional thioesterase (T) or reductase (Re) domain that releases the peptide chain. In each module there might be further modification domains acting as cyclisation, methylation and epimerization domains that may further enhance the complexity of peptides (Finking and Marahiel, 2004).

1.3.2 Epimerization domain

The composition of D-Amino acids is one of the key feature influencing the structure of non-ribosomal peptides. The epimerization (E) domain is involved in the racemization of either the PCP-bound amino acid or the C-terminal amino acid of the polypeptide chain and brings further structural variations in NRPS (Stachelhaus and Walsh 2000). These reactions are carried out by two different types of E-domains including aminoacyl epimerases and peptidyl epimerases, which are part of the initiation and the elongation module, respectively.

The C-starter domain present at the beginning of a NRPS is involved in the condensation of a lipid moiety onto the first activated monomer of the peptide chain (Gao et al., 2012). The majority of NRPS contain an E-domain immediate downstream of the T-domain resulting in a particular module architecture (C-A-T-E). All cyclic lipopeptide synthetases from *Bacillus* and *Paenibacillus* harbor E-domains in the modules encoding for D-monomers.

1.4 Sequence based prediction for natural product biosynthetic pathways

1.4.1 Co-linearity rule of assembly lines

Sequence based analysis of biosynthetic pathways is a promising approach for the discovery of secondary metabolites, however, such approaches are limited by the lack of computational platforms that will enable to predict secondary metabolite clusters on a large scale. Intriguingly, genes involved in secondary metabolite biosynthetic pathway are often organized on the chromosome in clusters, and these large collection of genes is referred as a 'secondary metabolite biosynthesis gene cluster', and such genetic architecture enables straightforward detection of secondary metabolite biosynthetic pathways (Blin et al., 2013).

Often NRPS and PKS type I systems follow co-linearity rule in which the incorporation of amino acids (for NRPS) and carboxylic acids (for PKS) for the biosynthesis and final assembly of the structure is usually the same as the order of catalytic modules in the genome (Staunton and Weissman, 2001; Rawlings, 2001). This structural feature and insight into the organization of the catalytic modules and domains within natural-product assembly lines often enables prediction of the final product based on the genomic sequence (Bachmann and Ravel, 2009). However, deviations from the conventional biosynthetic logic such as module iteration and skipping has been noticed for several multimodular assembly lines (Wenzel and Müller, 2005).

1.4.2 Genome sequencing

In conventional microbial natural product discoveries, the active compounds are isolated by cultivation of mainly fungi and bacteria under different growth conditions to induce the production of secondary metabolites, followed by testing their pharmacological or biological activities (Bentley et al., 2002; Li and Vederas, 2009). The structures of the active compounds are further elucidated by employing sophisticated analytical chemistry. However, a substantial

fraction of all microorganisms are difficult or impossible to cultivate with the present techniques. In addition, genome sequencing of various natural product-producing organisms have shown that even in well-characterized organisms the biosynthesis pathways of the vast majority of natural products are still unknown, largely because they are silent until specific conditions trigger their expression or remain undetected by conventional screening approaches (Koehn and Carter, 2005). This seems to be one of the major bottlenecks in the discovery of natural products.

The advent of high-throughput next-generation sequencing techniques has significantly reduced the costs and technical difficulty of sequencing and, thus, rapidly increased sequence repository from both isolated strains and microbial communities from metagenomes. Also the advent of omics technologies such as genomics, transcriptomics, proteomics and metabolomics has created a paradigm shift in natural product discovery. By finding secondary metabolite gene clusters within a genome, it is possible to estimate the biosynthetic potential of producer strains, which may provide access to previously uncharacterized metabolites (Cimermancic et al., 2014, Blin et al., 2013).

Moreover, concomitant development of computational tools for mining biosynthetic gene clusters opened up new avenues to identify and characterize secondary metabolites with novel chemical structures (Caboche, 2014).

1.5 Genome mining tools for natural product discovery

Although natural products have vast chemical scaffolds, the molecular principles underlying for most secondary metabolite biosynthesis is often conserved. A set of non-ribosomal peptide enzyme families are responsible for the synthesis of diverse classes of secondary metabolites and the sequence information from these known gene families can be exploited to mine genomes to identify diverse secondary metabolite biosynthetic pathways. Increasing and re-emerging interest in the discovery of microbial derived natural products has fueled recent development of several databases comprising information on biosynthetic gene clusters and the corresponding secondary metabolites.

Different strategies that are largely implemented by genome mining tools to identify secondary metabolite biosynthetic gene clusters have been recently highlighted in several reviews (Bachmann et al., 2014; Weber et al., 2014; Medema et al., 2015; Boddy et al., 2014). The recent Secondary Metabolite Bioinformatics Portal (SMBP) provides links to the commonly used computational tools and databases (Weber and Kim, 2016). A comprehensive list of freely accessible databases and tools that are currently available for natural product research (mining secondary metabolite biosynthetic pathways) are summarized in Table 2.

Before automated prediction tools became available, many common secondary metabolite biosynthetic pathways encoding type I or II polyketides, ribosomally and post-translationally modified peptides (RiPPs) and non-ribosomal peptides can be identified with high accuracy by BLAST or PSI-BLAST (Altschul et al., 1997) and HMMer (Eddy, 2011), and more improved search using tools like MultiGeneBlast (Medema et al., 2013). However, this manual annotation is cumbersome and time consuming, moreover lead to incomplete annotations. Automatic annotation of secondary metabolite gene clusters may enable complete annotation with improved accuracy.

The two strategies, rule-based and rule-independent approaches are mostly exploited in automated genome mining. Modular PKS or NRPS derived conserved protein domains within the genome can be easily screened by NaPDoS (Ziemert et al., 2012), NP.searcher (Li et al., 2009), GNP/PRISM (Skinnider et al., 2015), and SMURF (Khaldi et al., 2010). These tools, which implement predefined rules, can assign gene sequences to known families and classes with high accuracy, however, cannot detect gene sequences encoding novel pathways. Alternatively, rule-independent algorithms have been developed to overcome these problems as implemented in ClusterFinder (Cimermancic et al., 2014) and EvoMining (Cruz-Morales et al., 2015) (Table 2).

Bioinformatic tools such as antibiotics and Secondary Metabolite Analysis SHell (antiSMASH) implements both rule-based and rule-independent algorithms to unambiguously identify gene sequences (Weber et al., 2015; Blin et al., 2013; Medema et al., 2011). The prediction process integrated into antiSMASH pipeline has been described by Medema et al. (2011). The first step involves prediction and annotation of genes by gene prediction algorithms (Glimmer3 (Delcher et al., 2007) for bacteria) on the uploaded DNA sequence file (FASTA file, GenBank or EMBL). In order to identity potential biosynthetic gene clusters, the translated amino acid sequences from protein encoding genes are then compared against a curated profile Hidden Markov Models (pHMMs) of the 44 different classes of biosynthetic genes from both bacteria and fungi, using the HMMer3 software (Eddy, 2011). Additionally, less typical and new biosynthetic pathways are also identified using the rule independent ClusterFinder algorithm (Cimermancic, 2014). Following predicted protein domains in the detected gene cluster are assigned to secondary metabolite-specific clusters of orthologous groups (smCOGs). After cluster detection, specific analysis modules analyze pathway in more detail and predict the product encoded by a given cluster. Using pHMM-based approach, identified protein domains are assigned to conserved motifs within multimodular PKS and NRPS natural products. Further substrate specificities for acyltransferase (AT) of PKS and adenylation (A) domains of NRPS are assigned by analyzing active site residues using methods of Minowa et al. (2007) and NRPSpredictor2 (Röttig, 2011) and thus, a core chemical structure of the putative biosynthetic product is predicted. Moreover, an integrated MultiGeneBlast (Medema et al., 2013) and ClusterBlast is performed to identify similar experimentally characterized gene clusters.

Table 2. Overview of the most applicable tools and methods specialized for prediction of NRPS and PKS biosynthetic pathways (adapted from Weber and Kim, 2016)

Software program or database	Functions	Reference
Tools for mining of SMBGcs		
2metDB^{R}	Standalone (Mac) tool to mine PKS/NRPS gene clusters	Bachmann and Ravel, 2009
antiSMASH$^{R/N}$	Web application and standalone tool (LINUX, MacOS and MS Windows) to mine and analyze BGCs	Weber et al., 2015; Blin et al., 2013; Medema et al., 2011
BAGEL^{R}	Web application to mine and analyze RiPPs	van Heel et al., 2013; de Jong et al., 2010; de

		Jong et al., 2006
CLUSEAN[R]	Standalone (LINUX and MacOS) tool to mine and analyze BGCs, mainly PKS/NRPS	Weber et al., 2009
ClusterFinder[N]	Standalone tool (LINUX and MacOS) to identify BGCs with an non-rule based approach	Cimermancic et al., 2014
eSNaPD[R]	Web application to mine metagenomic datasets for BGCs	Charlop-Powers et al., 2015
EvoMining[N]	Web application for phylogenomic approach of cluster identification	Cruz-Morales et al., 2015
GNP/GenomeSearch[R]	Web application to mine and analyze BGCs, mainly PKS/NRPS	Johnston et al., 2015
GNP/PRISM[R]	Web application to mine and analyze BGCs, mainly PKS/NRPS, including glycosylations and structure prediction	Skinnider et al., 2015
MIDDAS-M[N]	Web application to use transcriptome data to identify BGC coordinates in fungal genomes	Umemura, et al., 2013
MIPS-CG[N]	Web application to identify BGC coordinates in fungal genomes without transcriptome data	Takeda et al., 2014; Umemura et al., 2015
NaPDoS[R]	Web application offering phylogenomic analysis of PKS-KS and NRPS-C domains	Ziemert et al., 2012
SMURF[R]	Web application to mine PKS/NRPS/terpenoid gene clusters in fungal genome	Skinnider et al., 2015
Software for the analysis of typeI PKS and NRPS pathways		
ClustScan Professional	Java-based standalone tool to mine for PKS/NRPS BGCs	Starcevic et al., 2008
NP.searcher	Web application/standalone tool (LINUX) to mine for PKS/ NRPS BGCs	Li et al., 2009
NRPS-PKS/SBSPKS	Web application to mine for PKS BGCs	Anand et al., 2010; Ansari et al., 2004
SEARCHPKS	Web application to mine for PKS BGCs	Yadav et al., 2003
Software for predicting substrate specificities		
LSI-based A-domain function predictor	Web application to predict A-domain specificities	Baranasic et al., 2014

NRPS/PKS substrate predictor	Web application to predict A-domain/AT-domain specificities	Khayatt et al., 2013
NRPS predictor/NRPS predictor2	Web application/standalone tool (LINUX, MS Windows, MacOS) to predict A-domain specificities	Röttig et al., 2011; Rausch et al., 2005
NRPSsp	Web application to predict A-domain specificities	Prieto et al., 2012
PKS/NRPS Web Server/Predictive Blast Server	Web application to determine domain organization and A-domain specificities	Bachmann and Ravel, 2009
SEARCHGTr	Web application to predict glycosyltransferase specificities	Kamra et al., 2005
SEQL-NRPS	Web application to predict A-domain specificities	Knudsen et al., 2015
Databases focusing on gene clusters		
Bactibase	Web accessible database for bacteriocins	Hammami et al., 2007; Hammami et al., 2010
ClusterMine360	Web accessible database of BGCs	Conway and Boddy 2013
ClustScan Database	Web accessible database of PKS/NRPS BGCs	Diminic et al., 2013
DoBISCUIT	Web accessible database of PKS/NRPS BGCs	Ichikawa et al., 2013
IMG-ABC	Web accessible database of BGCs, tightly integrated into JGI's IMG platform	Hadjithomas et al., 2015
MIBiG	Web accessible repository of BGCs	Medema et al., 2015
Recombinant ClustScan Database	Database of in silico recombined BGCs	Starcevic et al., 2012
Databases focusing on bioactive compounds		
Antibioticome	Web accessible database on compounds, compound families and modes of action	Unpublished
ChEBI	Web accessible database and ontology on compounds focused on small molecules	Hastings et al., 2016
ChEMBL	Web accessible database on bioactive compounds with druglike properties	Davies et al., 2015
ChemSpider	Web accessible database on structures and properties of over 35 million structures	Kelly and Kidd, 2015
KNApSAcK database	Web accessible database on compounds; standalone version of KNApSAcK metabolite database available	Nakamura et al., 2014; Afendi et al.,

		2012
NORINE	Web accessible database on NRPs	Caboche et al., 2008; Flissi et al., 2015
Novel Antibiotics Database	Web accessible database on compounds	Unpublished
PubChem	Web accessible database on compounds and bioactivities; source data available for download	Bolton et al., 2008
StreptomeDB	Web accessible database on compounds produced by streptomycetes; download of compounds and metadata in SD format.	Lucas et al., 2013; Klementz et al., 2016

[R]:rule-based,[N]:non-rule based algorithms used to detect the BGCs

1.6 Secondary metabolite biosynthetic potential of bacteria

In a recent genome mining study, Wang and collaborators by examining a total of 3,339 gene clusters from 2,699 genomes, from 991 organisms have demonstrated that modular non-ribosomal peptide, polyketide and hybrid NRPS-PKS biosynthetic pathways are widely distributed across bacteria, fungi, archaea, animals, plants, and protists. Most importantly, among these a significant high number of secondary metabolite gene clusters (2,976, 89%) can be found in bacteria. In particular, Proteobacteria, Actinobacteria and Firmicutes phyla dominated these counts, therefore a great diversity of non-ribosomal peptide and polyketide structures can be anticipated from these groups, and with such untapped wealth, these groups can be of more interest for natural product discovery (Wang et al., 2014).

1.7 Secondary metabolites of *Bacillus* and *Paenibacillus*

Bacillus and *Paenibacillus* spp. synthesize many functionally diverse low molecular weight secondary metabolites including aminoglycosides, polyketides, and several small proteinaceous and peptidal structures including bacteriocins, lantibiotics, aminoglycosides, catecholate siderophores (mycobactin, enterobactin etc.) oligopeptides and lipopeptides with cyclic, linear or branched structures (Walsh et al., 2008). They exhibit broad range of biological activities such as bactericidal, immune suppression and tumor suppression properties, and therefore they are of immense importance in medicine and agriculture. Lipopeptides and polyketides represent the predominant class and from biocontrol perspective, they act as antagonists by inhibiting phytopathogen growth, as immuno-stimulators reinforce host plant resistance potential, and most importantly mediate beneficial plant-microbe interactions. Nevertheless, also enhance rhizospheric competence and ecological fitness of the producing strain (Kinsinger et al., 2003; Ongena and Jacques, 2008). A more detailed natural functions of lipopeptides and polyketides has been described in Chapter 1.10.

1.7.1 Lipopeptides of *Bacillus* and *Paenibacillus*

Many of the well-known lipopeptides of *Bacillus* and *Paenibacillus* spp. have been the subject of several recent reviews (Stein, 2005; Ongena and Jacques, 2008; Raaijmakers et al., 2010; Roongsawang et al., 2010; Jacques, 2011; Cochrane and Vederas 2014; Aleti et al., 2015).

The majority of lipopeptide structures synthesized by *Bacillus* and *Paenibacillus* spp. encompass short peptides, typically in a range of 6 and 13 amino acid residues, which can be cyclic or linear. Cyclic lipopeptide structures can be either cationic or non-cationic based on the net positive charge of the molecule. A substantial number of lipopeptides from *Paenibacillus* comprise cyclic cationic lipopeptides, for instance, polymyxins, polypeptins, octapeptins and paenibacterins contain 2,4-diaminobutyric acid (dab), a non-proteogenic amino acid, which is responsible for overall positive charge of these lipopeptides. Non-cationic structures in *Bacillus* and *Paenibacillus* spp. include surfactin, iturin, fengycin, and fusaricidin. While saltavalin, jolipeptin, tridecaptins from paenibacilli and cerexins from bacilli with strong bactericidal activities belong to the linear cationic class.

Many of the lipopeptides can cyclize at the C-terminus of the peptide by forming an ester or amide bond. Furthermore, a lipid tail either linear or branched fatty acids can be added to the peptide by acylation of the N-terminal amino acid (Ongena and Jacques, 2008; Cochrane and Vederas 2014), which is considered as the hallmark of the lipopeptide biosynthesis in *Bacillus* and *Paenibacillus*.

In the following sections well noted lipopeptides and the modular architecture of their biosynthetic gene clusters from *Bacillus* and *Paenibacillus* have been described in more detail.

1.7.1.1 Lipopeptides of *Bacillus* spp.

Bacillus species produce a large number of lipopeptides, but especially lipopeptides from *B. subtilis* and *B. amyloliquefaciens* represent a substantial better investigated fraction. Among these, the three families of cyclic lipopeptides iturin, surfactin, and fengycin have potential role in biotechnology and bio-pharmaceutical applications (Romero et al., 2007; Chen et al., 2009). Structurally, *Bacillus* lipopeptides are cyclic structures interlinked to a β-amino or β-hydroxy fatty acid chain (Fig.1).

Lipopeptides of *Bacillus* and their biosyntheis have been described in depth in several recent reviews (Ongena and Jacques, 2008; Roongsawang et al., 2010; Jacques, 2011).

Surfactin and fengycins synthesis is mediated by non-ribosomal peptide synthetases (NRPS) (Finking, and Marahiel, 2004; Stein, 2005) and iturin is produced by a hybrid of NRPS-PKS (polyketide synthases) (Fig.2) (Duitman et al., 1999; Koumoutsi et al., 2004).

1.7.1.1.1 The surfactin family

Chemically, members of the surfactin family have a cyclic heptapeptide structure in common, connected to the β-hydroxy fatty acid chain (typically C12 to C16) form a lactone ring structure (Fig.1) (Peypoux et al., 1999). The surfactin family contains structural variants with the same peptide length but differ in individual peptide moieties (Ongena and Jacques, 2008; Jacques, 2011). Some of the well-known surfactin variants lichenysin, pumilacidin and surfactin largely represent the surfactin family and are remarkably confined to specific taxonomic groups (Ongena and Jacques, 2008). Surfactin variants that mainly differ at

position 7 have been found in *B. subtilis*. They include surfactin A, B and C comprising Leu, Val and Ile, respectively (Stein, 2005). Lichenysin and its biosynthetic operon have been reported in *B. licheniformis* (Horowitz et al., 1990; Konz et al., 1999). Lichenysin differs from surfactin as Gln is found instead of Glu at position 1 (Yakimov et al., 1999) (Fig.2). The variant pumilacidin containing Leu at position 4 has been described in *B. pumilus* M937-B1 but do not correlates with the structure of synthetases described for other surfactin producers (Naruse et al., 1990). Furthermore, each of these variants contain isoforms that can vary in length and branching of the fatty acid chain that can be linear, iso and anteiso (Ongena and Jacques, 2008; Jacques, 2011).

Surfactins are reported as powerful biosurfactants with exceptional emulsification and foaming properties, and their mode of activity is largely by membrane permeabilization and disruption. Due to the amphiphilic nature, surfactins can tightly associate and penetrate lipid layers, and thereby interfere with biological membrane integrity in a dose-dependent manner (Heerklotz et al., 2007). Findings of Chen et al. (2009) indicate that surfactin production by *B. amyloliquefaciens* FZB42 protects it from other bacteria, as well as enhances motility and helps it to establish robust biofilms, thus equipping the bacterium with powerful antagonistic advantage during root colonization. Surfactins as powerful biosurfactants display a wide range of hemolytic, antimicrobial and antiviral activities, although antifungal activities have not been observed (Vollenbroich et al., 1997; Heerklotz et al., 2007).

A typical surfactin gene encoding for a surfactin synthetase comprises three large open reading frames (ORFs) named *srfA*, *srfB* comprising three modules each, and *srfC* comprising one module (Peypoux, 1999). Overall seven modules are organized in a linear fashion for surfactin production i.e., one module per residue. Each module contains core catalytic domains such as condensation, adenylation and thiolation domain (Fig.2). The adenylation domain responsible for the selection and recruitment of amino acids on the peptide chain is largely responsible for biosynthesis of structurally diverse cyclic lipopeptides that differ in peptide moiety. The selective activity of an adenylation domain is encompassed in ten amino acids of the primary sequence, of which eight amino acids are variable. This non-ribosomal code allows prediction of A-domain specificity based on the primary sequence (Stachelhaus et al., 1999; Jacques, 2011). The fatty acid chain is added to the amino acid activated by the first module, which is the hallmark of lipopeptide biosynthesis. The first thioesterase connected to the C-terminal end of the last activation PCP domain encoded by *srfC* operon catalyzes release of the peptide from the NRPS synthetase (Finking, and Marahiel, 2004; Stein, 2005). A second thioesterase domain encoded by a fourth gene, *srfD* is involved in initiation of surfactin biosynthesis (Steller et al., 2004).

The mechanism involved in regulation of surfactin biosynthesis is known to be associated with the competence development pathway, described in depth in later section below (1.10.6).

1.7.1.1.2 The iturin family

Similar to surfactins, all members of the iturin family encompass a cyclic heptapeptide backbone but are interlinked with a β-amino group of fatty acid chain (C14 to C17 carbons) (Fig.1). Bacillomycins, mycosubtilins and iturins represent main structural variants of this group and encompass a common part of the peptide backbone L-Asx – D-Tyr – D-Asn while the next two residues (Gln, Pro of *mycB* and *ituB*; Pro, Glu of *bmyB*) and the two last residues (Ser, Asn of *mycC*; Asn, Ser of *ituC* and Ser, Thr of b*myC*) are different (Fig.2) (Ongena and Jacques, 2008; Cochrane and Vederas, 2014; Jacques, 2011).

The members of iturin family are encoded by a hybrid PKS-NRPS synthetases, for instance, the mycosubtilin synthetases encoded by the three *myc* operons (*mycA*, *mycB* and *mycC*) contain seven modules responsible for the incorporation of one, four and two amino acid residues, respectively (Duitman et al., 1999) (Fig.2).The first module *fen*F codes for a malonyl-CoA transacylase (MCT-domain), and the following *myc*A also encompasses polyketide synthases (AL, ACP, KS and AMT) involved in the biosynthesis of the fatty acid chain, which is transferred to the activated first amino acid prior to its incorporation in the peptide moiety (Ongena and Jacques, 2008).

Although members of this group display strong hemolytic activity, their biological activity is largely different from surfactin group. Their microbial activities toward bacteria and viruses are limited (Aranda et al., 2005) but they display strong antagonistic activity against a wide range of fungi (Duitman et al., 1999; Moyne et al, 2001; Yu et al., 2002). The toxicity of iturins towards fungi is mainly through membrane permeabilization and formation of ion-conducting pores leading to osmotic perturbation (Aranda et al., 2005). Members of the iturin group are considered as potential alternative to synthetic antifungal agents. For instance, Yu et al. (2002) have reported that *B. amyloliquefaciens* strain B94 suppresses *Rhizoctonia solani* infection mainly by production of iturin.

Bacillomycin has been shown to exhibit broad spectrum of antifungal activity. For example, the antifungal activity exhibited by well-known plant growth promoting rhizobacterium *B. amyloliquefacienes* FZB42 against phytopathogen *Fusarium oxysporum* has been associated with bacillomycin D biosynthesis (Koumoutsi et al., 2004). Increased biosynthesis of mycosubtilin by *B. subtilis* mutant showed enhanced antagonistic activities toward *Botrytis cinerea, Fusarium oxysporum* and *Pythium aphanidermatum* (Leclere et al., 2005).

In addition to biocontrol activities, members of iturin family are also known to influence swarming and motility activities of *Bacillus*, thus contributing to the ecological fitness of the producer strain (Leclere et al. 2005).

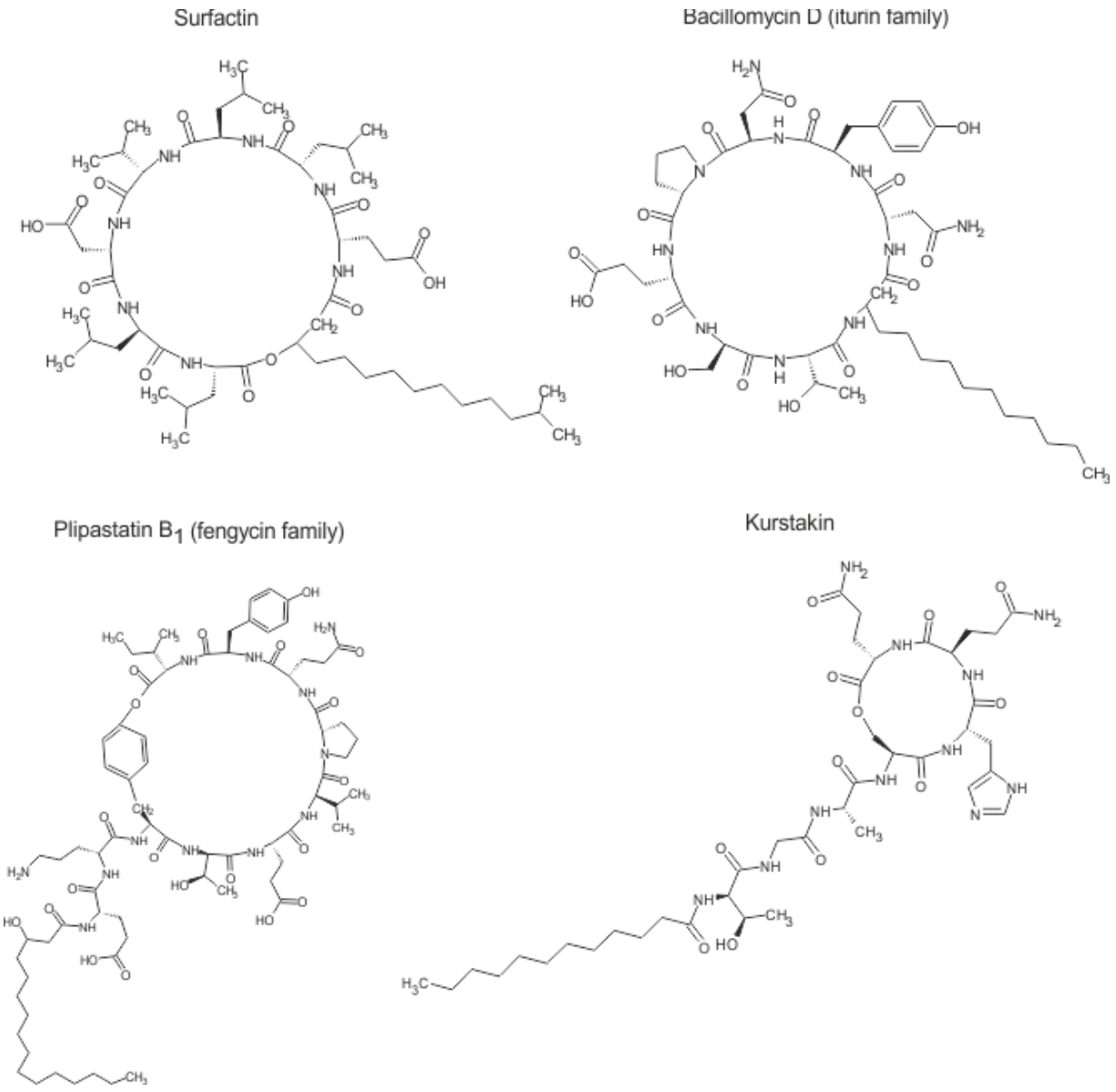


Fig.1. Chemical structures of representative molecules from the four families of lipopeptides from *Bacillus*. Some important lipopeptide structures produced by B. amyloliquefaciens FZB42 (surfactin; bacillomycin, an iturin member; plipastatin, a fengycin member) and kurstakin from B. thuringiensis kurstaki HD-1 (Aleti et al., 2015).

1.7.1.1.3 The fengycin family

The third family of lipopeptides of *Bacillus* comprise fengycins and the closely related plipastatins. These molecules contain decapeptide as a peptide backbone with an internal lactone ring between the C-terminus and tyrosine at position three, and the N terminus of glutamic acid is connected to different β-hydroxy fatty acid chains by an amide bond (C_{14} to C_{18}) (Fig.1). Fengycin and plipastatin mainly differ by the L and D forms of tyrosine, which are in position 3 and 9, respectively, for plipastatins and 9 and 3 for fengycins. However, the fengycin production cannot be correlated with the structure of the synthetases described (Koumoutsi et al. 2004).

Plipastatins are synthesized by peptide synthetases encoded by a large operon with five open reading frames ppsA–E (Fig.2). The first three enzymes are involved in incorporation of two

residues each, the fourth enzyme is involved in selection of three residues and the last enzyme is responsible for incorporation of one residue in the peptide moiety of plipastatin (Steller et al., 1999).

Fengycins are strongly antifungal in nature, especially towards several filamentous fungi but are less hemolytic than iturins and surfactins. Mechanistically, the action of fengycins is less understood but similar to other lipopeptides as they also readily associate and penetrate lipid layers to interfere with cell membrane integrity and permeability in a dose-dependent way.

The three lipopeptides surfactin, iturin and fengycin are known to act synergistically enhancing their individual activities (Romero et al., 2007). Co-production of fengycins and iturins in *Bacillus* spp. has been shown to be involved in suppression of several pathogens both in vitro and in vivo in plants (Kim et al., 2010; Ongena et al., 2005; Romero et al., 2007). Romero et al. (2007) demonstrated that production of both iturins and fengycins by *B. subtilis* can confer protective activity against *Podosphaera fusca* infections on melon plants. Site-directed mutagenesis of biosynthetic genes responsible for the production of different lipopeptides showed no biocontrol activity, which further confirms the above findings. The role of fengycins as direct antifungal agents against *Botrytis cinerea* in the plant system has been studied using *B. subtilis* S499 (Ongena et al., 2005). The above findings indicate that fengycins can be *Bacillus* powerful antibiotic arsenal during root colonization.

1.7.1.1.4 Kurstakins

Another family of lipopeptides known as kurstakins were isolated from *B. thuringiensis* subsp *kurstaki*, and the presence of this lipophilic marker has been considered as a phylogenetic marker of this species (Hathout et al., 2000). Kurstakins comprise a heptapeptide backbone with a peculiar lactone bond between serine at position 4 and the C-terminus of glutamine at position 7. The first threonine residue is attached by an amide bond to β-hydroxy fatty acid chain (C_{11} to C_{14}) (Fig.1). Kurstakins display limited antifungal activity (Béchet et al., 2012; Hathout et al., 2000).

The predicted kurtsakin cluster encompasses three genes including *krsA*, *krsB* and *krsC* comprising one, two and four modules each, respectively (Fig.2) (Béchet et al., 2012). The first condensation of KrsA catalyses the initiation of a threonine residue with a fatty acid chain attached, and subsequently an epimerization domain transforms L-Thr to D-Thr. The second synthetase KrsB catalyzes the incorporation of Gly and Ala amino acid residues, and the third synthetase KrsC is responsible for incorporation of amino acid residues: Ser, His, Gln and Gln. Similar to other NRPS two thoiesterases are present one in KrsC and KrsD (Béchet et al., 2012). The sixth gene *krsE* which is present upstream of the *krsA-C* genes, is presumed to encode efflux protein that maybe involved in the secretion of the peptide (Béchet et al., 2012).

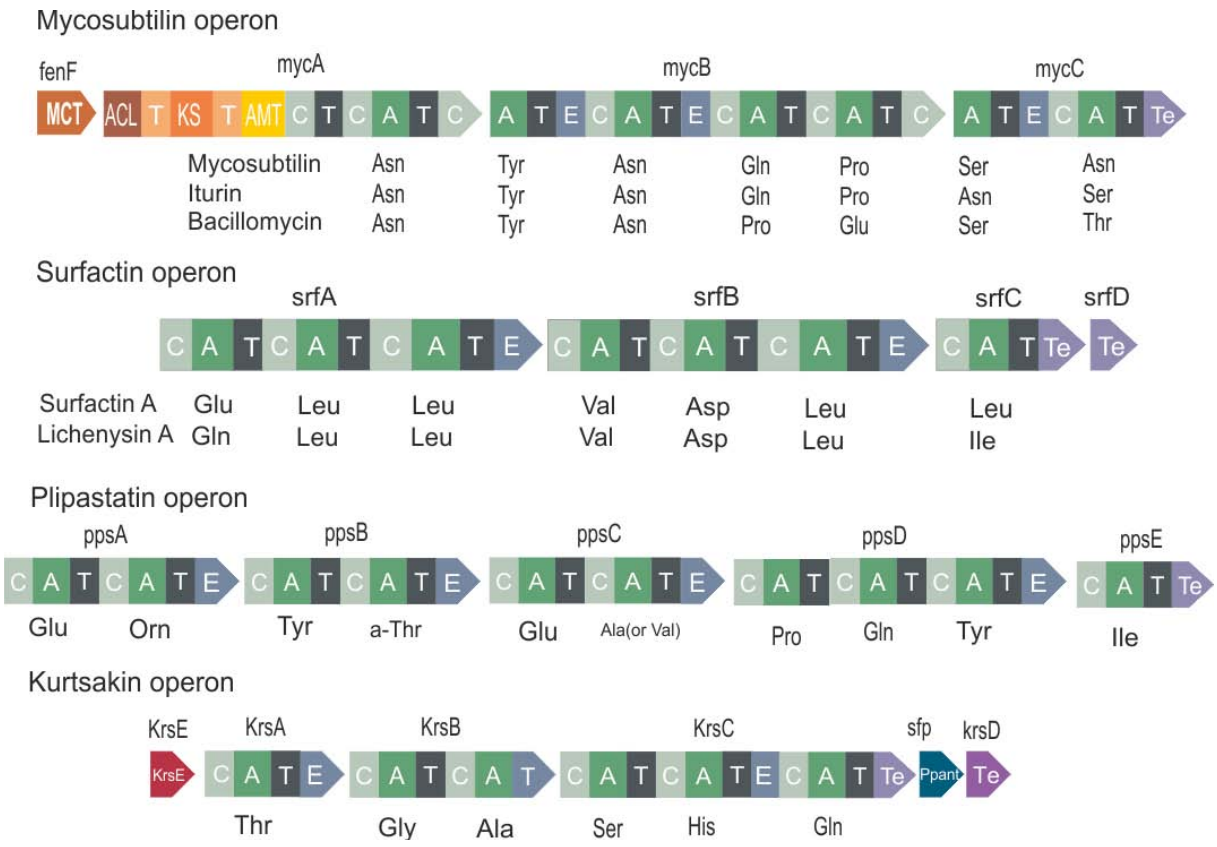


Fig.2. Schematic representation of the mycosubtilin (iturin family), surfactin, plipastatin (fengycin family) and kurstakin encoding operons in *Bacillus* species. Iterative domains: MCT, malonyl-CoA transacylase; ACL, acyl-coA ligase; A, adenylation; T, thiolation; E, epimerization; AMT, aminotransferase; KS, keto synthetase; TE, thioesterase. Gene names are indicated above and amino acids incorporated by adenylation domain within each module are shown below (adapted from Ongena and Jacques, 2008; Aleti et al., 2015).

1.7.1.1.5 Cerexins

Cerexins are linear structures isolated from *B. cereus* and contain a decapeptide backbone with a β-hydroxyl lipid tail. They show strong bactericidal effect on gram-positive bacteria, including *Staphylococcus aureus* and *Streptococcus pneumoniae*. Cerexin analogues mostly differ in monomer composition at positions 3, 6 and 8, and N-terminal lipid tail (Cochrane and Vederas 2014).

1.7.1.2 *Paenibacillus* lipopeptides

1.7.1.2.1 Polymyxins

Polymyxins are considered as the oldest class of lipopeptides mainly isolated from *Paenibacillus polymyxa*. All analogues of polymyxin encompass five or six Dab residues within the decapeptide backbone that impart net positive charge to polymyxins (Fig.3) (Velkov et al., 2010). The amine side chain of Dab at the 4th position forms an amide linkage with the C-terminus, resulting in cyclization of the peptide. Furthermore, the N-terminus is attached to a fatty acid chain by acylation. Polymyxins are known for their strong antibacterial activity, specifically against Gram-negative bacteria. Polymyxins primarily target the lipid A component of lipopolysaccharide within the outer cell membrane, resulting in disruption of the membrane, followed by further permeabilization of cell leading to

disruption of the inner cell membrane (Velkov et al., 2010; Soon et al., 2011). Recent in vitro studies by McCoy and collaborators have demonstrated that polymyxins can also impair eukaryotic translation by binding to ribosomal RNA (McCoy et al., 2013). In addition to the bactericidal activity, polymyxins displayed biocontrol effects against phytopathogenic *Erwinia* spp. (Niu et al., 2013).

P. polymyxa E681, a plant growth promoting rhizobacterium, has been shown to encompass a gene cluster encoding polymyxin synthetase (Fig.4). The gene cluster includes *pmxA*, *pmxB* and *pmxE*, which encode the polymyxin synthetase, while *pmxD* and *pmxC* are responsible for polymyxin transportation (Choi et al., 2009). Polymyxin variants A-F, M, S and T are known to contain altered monomer compositions at positions 3, 6, 7 and 10 (Cochrane and Vederas 2014).

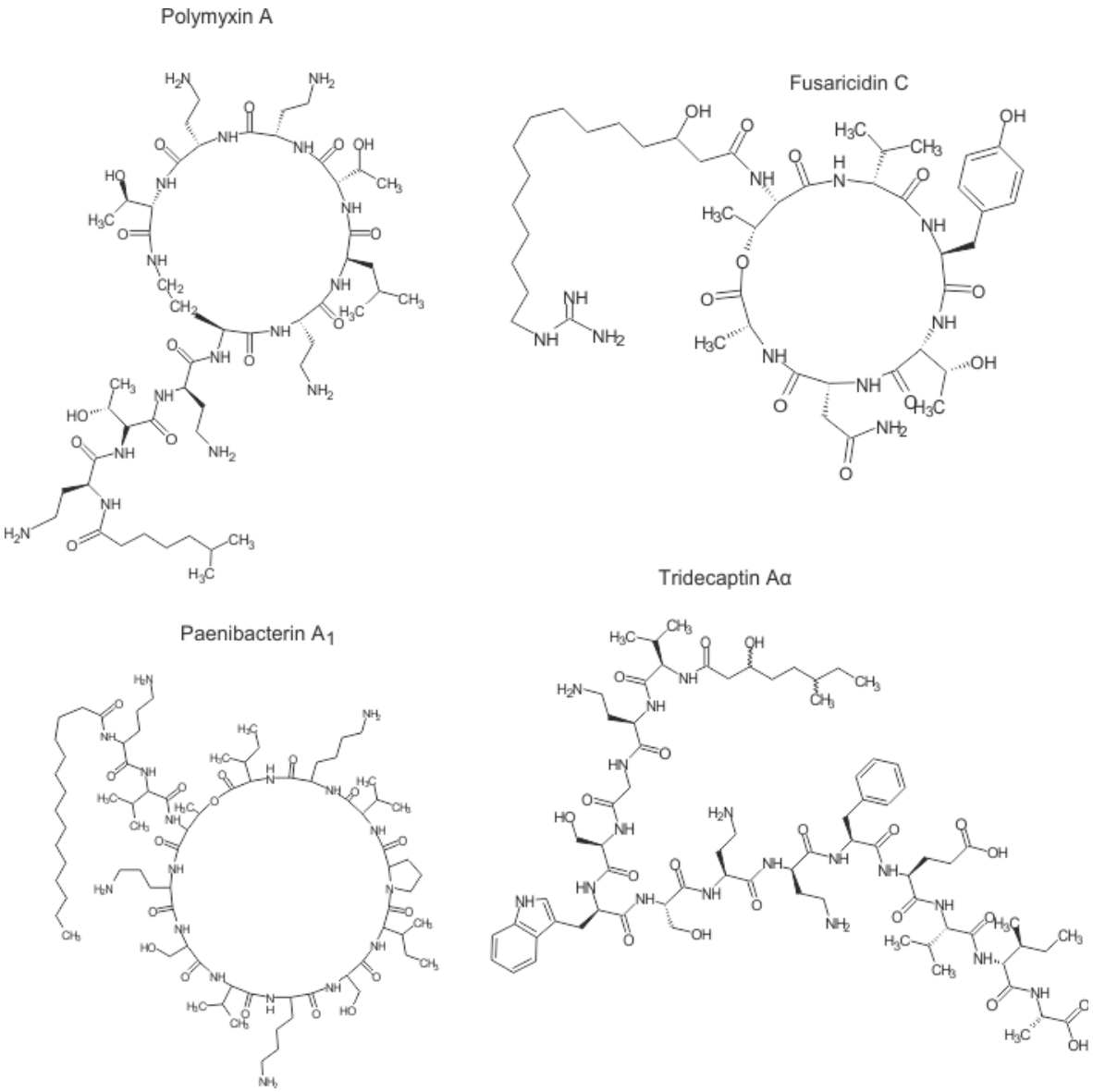


Fig.3. Chemical structures of representative lipopeptides from *Paenibacillus*. Polymyxin A and fusaricidin C from *P. polymyxa* E681, paenibacterin from *Paenibacillus* sp. OSY-SE, tridecaptin from *P. terrae* NRRL B-30644 (Aleti et al., 2015).

1.7.1.2.2 Octapeptins

Octapeptin structures belong to the class of cyclic cationic lipopeptides and contain eight monomers as peptide backbone, but are truncated when compared to polymyxins. Both polymyxins and octapeptins contain cyclic heptapeptide structure in common with an amide linkage among the monomer at position 1 and the C-terminus but differ at 7th position with L-Thr in polymyxins, and L-Leu in octapeptins. Unlike polymyxins, octapeptins display broad bactericidal activity by targeting cell membranes of both gram-positive and gram-negative bacteria. Also they have been shown to suppress filamentous fungi, protozoa and yeasts (Cochrane and Vederas 2014).

Octapeptin variants differing in composition of acyl lipid tail and altered monomer composition at positions 1, 4 and 5 have been described but the biosynthetic gene cluster has not been reported yet (Qian et al., 2012; Cochrane and Vederas 2014). Although octapeptins were reported previously from bacilli, in a more recent study, Qian et al. (2012) isolated battacin (a member of octapeptins) from paenibacilli. This clearly shows that the octapeptins belong to the genus *Paenibacillus* and the previously identified octapeptins may also belong to this genus as they were once classified as *Bacillus*.

1.7.1.2.3 Polypeptins

Polypeptins are cyclic cationic lipopeptides with nine monomers as peptide backbone, and comprise polypeptins and pelgipeptins. The first monomer at the N-terminus is acylated by a β-hydroxy fatty acid chain, whose hydroxyl moiety is connected by an ester linkage to the C-terminus of the last peptide, leading to cyclization. Polypeptins from *P. ehimensis* B7 and pelgipeptins from *P. elgii* B69 exhibit bactericidal activity against both Gram-positive and Gram-negative bacteria, as well as suppress several soil-borne pathogens, for instance, *Fusarium graminearum* and *Rhizoctonia solani* (Ding, 2011). The complete gene cluster coding for pelgipeptin has been recently identified and characterized in *P.elgii* B69 (Qian et al., 2012). Pelgipeptin gene cluster contains three ORFs, *plpD* and plpF encode one monomer each, and *plpE* encodes seven monomers.

Polypeptin analogues with variation in fatty acid chain and altered monomer composition at positions 2 and 9 have been summarized by Cochrane and Vederas (2014).

1.7.1.2.4 Other cyclic cationic lipopeptides of *Paenibacillus*

Besides cyclic cationic lipopeptide structures described above, *Paenibacillus* spp. also synthesize few structurally different cyclic cationic lipopeptides such as gavaserin and paenibacterins. The structure of gavaserin synthesized by *P. polymyxa* is anticipated to contain cyclic octapeptide structure with glutamic acid, alanine, serine, diaminobutyric acid, and octanoic acid as part of the peptide backbone (Pichard et al., 1995). However, thus far, structural data for gavaserin are not known.

Paenibacterins have been isolated from *Paenibacillus* sp. OSY-SE, whose draft genome sequence is also publicly available (Guo et al., 2012; Huang et al., 2012). Paenibacterins contain thirteen amino acid peptide backbone cyclized by an ester bond between the C-terminus and Thr at 3rd position, and acylated with a lipid tail at the N-terminus, however, no exact stereochemistry data and the biosynthetic gene cluster is available till date (Fig.3). Paenibacterin is active against gram positive and Gram-negative bacteria (Guo et al., 2012).

1.7.1.2.5 Fusaricidins

Fusaricidins are cyclic noncationic hexapeptide structures with cyclic portion comprising an ester bond between Thr at position 1 and Ala at the C terminus, and the N-terminus of Thr is attached to a distinctive fatty acid chain, to 15-guanidino-3-hydroxypentadecanoic acid (Fig.3) (Cochrane and Vederas 2014).

Fusaricidins have been shown to exhibit antimicrobial activities toward gram-positive bacteria, as well as *Fusarium* and *Aspergillus*. But unlike, cationic lipopeptide structures, fusaricidins display weak activities toward gram-negative bacteria (Kajimura et al., 1997). Most importantly, fusaricidins display strong activity against fungal infections on crop. They have been shown to protect watermelon from *Fusarium* wilt (Raza et al., 2009), canola from blackleg disease caused by *Leptosphaeria maculans* (Beatty et al., 2002), and red pepper plants from *Phytophthora* blight by inducing systemic resistance in the plant (Lee et al., 2013). Several fusaricidins with altered monomer composition at positions 2, 3 and 5 have been reported (Fig.4) (Cochrane and Vederas 2014).

1.7.1.2.6 Linear cationic structures

In addition to cyclic lipopeptides, *Paenibacillus* is capable of synthesizing linear cationic lipopeptides with different numbers of amino acid residues. Tridecaptins, isolated from *P. polymyxa* contain a tridecapeptide backbone and exhibit strong bactericidal activities toward gram-negative bacteria, but show little activity towards gram-positive bacterium (Fig.3). The gene cluster responsible for biosynthesis of tridecaptinAα has been recently described from *P. terrae* NRRL B-30644 (Fig.4) (Lohans et al., 2014). Recent study suggests that the mode of action of tridecaptin A_1 is by a membrane disruption (Cochrane and Vederas 2014). However, their specific activity against gram-negative bacteria is yet unclear. Tridecaptin variants with substituted amino acid residues have been described for positions 1, 5, 9, 12 and 13 (Cochrane and Vederas 2014).

Other classes such as saltavalin (Pichard et al., 1995) and jolipeptin (Ito et al., 1972) might also exist but needs confirmation and more data.

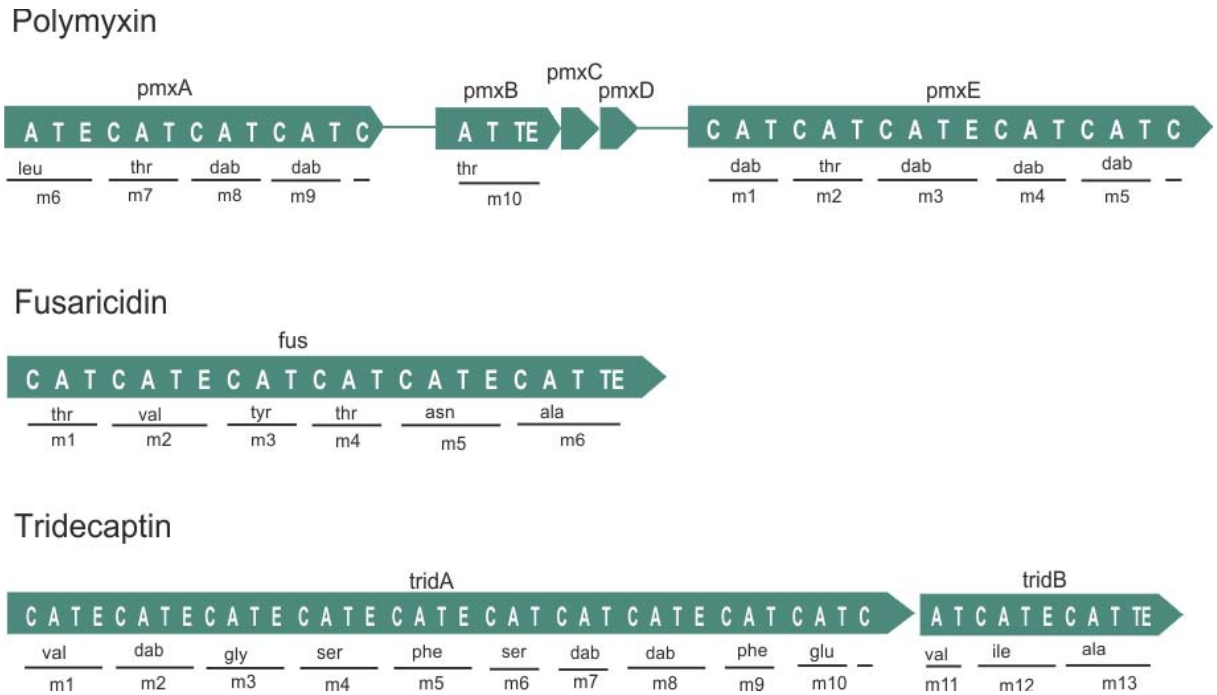


Fig.4. Schematic organization of polymyxin, fusaricidin and tridecaptin encoding operons in *Paenibacillus* species. Iterative domains: A, adenylation; T, thiolation; E, epimerization; TE, thioesterase; dab, 2,4-diaminobutyric acid. Gene names are indicated above and amino acids incorporated by adenylation domain within each module are shown below (modified from Aleti et al., 2015).

1.8 Polyketide biosynthetic pathway

Bacterial polyketide synthases are classified in three types based on their primary structures and catalytic activities, first, PKS type I are organized into modules and each module encompasses a series of non-iteratively acting domains responsible for loading the precursor units and elongation of the polyketide chain (Staunton and Weissman, 2001). Second, PKSs type II are large multienzyme complex which comprise a single set of iteratively acting domains. This class are involved in the biosynthesis of bacterial aromatic polyketides (Shen, 2000). Third, PKS type III also called chalcone synthase like PKS, contain iteratively acting acyl CoA condensing enzymes without the involvement of acyl carrier protein (Moore and Hopke, 2001).

Majority of *Bacillus* and *Peanibacillus* polyketides are of PKS type I and share similar architecture with NRPS. In PKS type I, each module is used only ones during biosynthetic processing of related polyketide. The precursors of polyketides are small carboxylic acids, such as acetate, propionate and malonate that are activated as their coenzyme A (CoA) thioesters. Each module of PKS-I comprises an acyltransferase (AT), an acyl carrier protein (ACP) and a ketosynthase (KS) domain. The AT domain is believed to determine which extender is incorporated at each step of polyketide chain growth while the KS domain catalyses condensation reaction via decarboxylation. The ACP domains tether the growing polyketide chain to the PKS between condensations and also accept extender units from the AT domains in preparation for a condensation. Thioesterase (TE) domain facilitates the release of the final product from the enzyme. In addition, several modifying domains might also be present in PKS-I moduls as β-keto reductase (KR), dehydratase (DH), enoyl reductase (ER), methyltransferase (MT) (Fischbach and Walsh, 2006).

1.9 Polyketides of *Bacillus* and *Paenibacillus*

In *B. amyloliquefaciens* FZB42, as much as 8.5% of the total genome is dedicated to the production of NRPS and PKS. Three important types of polyketides including bacillane, difficidin and macrolactin represent the diversity of *Bacillus* polyketides and these three large polyketide synthetases were assigned unambiguously by MALDI-TOF MS and LC-ESI MS techniques (Chen et al., 2006; Stein, 2005). Two polyketides paenilamicin and paenimacrolidin have been identified so far in the genus *Paenibacillus*. The underlying genetic cluster involved in paenimacrolidin remains to be established (Wu et al., 2011), while the biosynthetic gene clusters for recently identified paenilamicins from *P. larvae* have been described (Müller et al., 2014). In the following sections known polyketides and their biosynthetic gene clusters from *Bacillus* and *Paenibacillus* have been described in more detail.

1.9.1 Polyketides of *Bacillus*

1.9.1.1 Bacillaene

Bacillaene was first isolated from the culture supernatant of *B. subtilis* strains 3610, and 55422 (Patel et al., 1995). Butcher et al. (2007) and their structure was elucidated by combining differential NMR (nuclear magnetic resonance) spectroscopy techniques with genetically engineered strains of *B. subtilis* reported that bacillaene is a linear structure with two amide bonds. The first amide bond is formed between an α-hydroxy carboxylic acid and an ω-amino of a carboxylic acid containing a conjugated hexaene, while the second amide bond is formed between the carboxylic acid of this hexaene with and an ω-amino of a carboxylic acid with a conjugated triene (Fig.5).

The polyketide synthase gene cluster *bae* encodes a hybrid type I PKS-NRPS gene cluster responsible for the synthesis of bacillaene has been characterized from *B. amyloliquefaciens* FZB42 (Fig.6). Intriguingly, the bae gene cluster shares striking architectural and sequence similarity with the *pksX* gene cluster of *B. subtilis* 168 indicating that the *bae* gene cluster is orthologous to *pksX*. The bae gene cluster encompasses five giant open reading frames (ORFs) including *baeJ, baeL, baeM, baeN* and *baeR* (Chen et al., 2006). The first and second adenylation domains encoded by *baeJ* incorporate α-hydroxy-isocaproic acid and glycine, respectively, while the third adenylation domain encoded by *baeN* is responsible for incorporation of alanine (Calderone et al., 2006). Modules 4, 8 and 14 are splitted between adjacent genes (Fig.). The *bae* operon contains another unusual feature, the PKS system lacks the acyltransferase (AT) domains that are normally present in each biosynthetic module. Instead, three short ORFs including *baeC, baeD, baeE* upstream of *baeJ* encode three discrete AT domains, which load malonyl-CoA (Calderone et al., 2006). Bacillaene and dihydrobacillaene represent structural variants of this group (Chen et al., 2009; Butcher et al., 2007) (Fig.6).

Cell survival assays indicate that bacillaene is active against a wide range of bacteria. Bacillaene exhibits selective suppression of protein biosynthesis in prokaryotes, but not in eukaryotes (Patel et al., 1995).

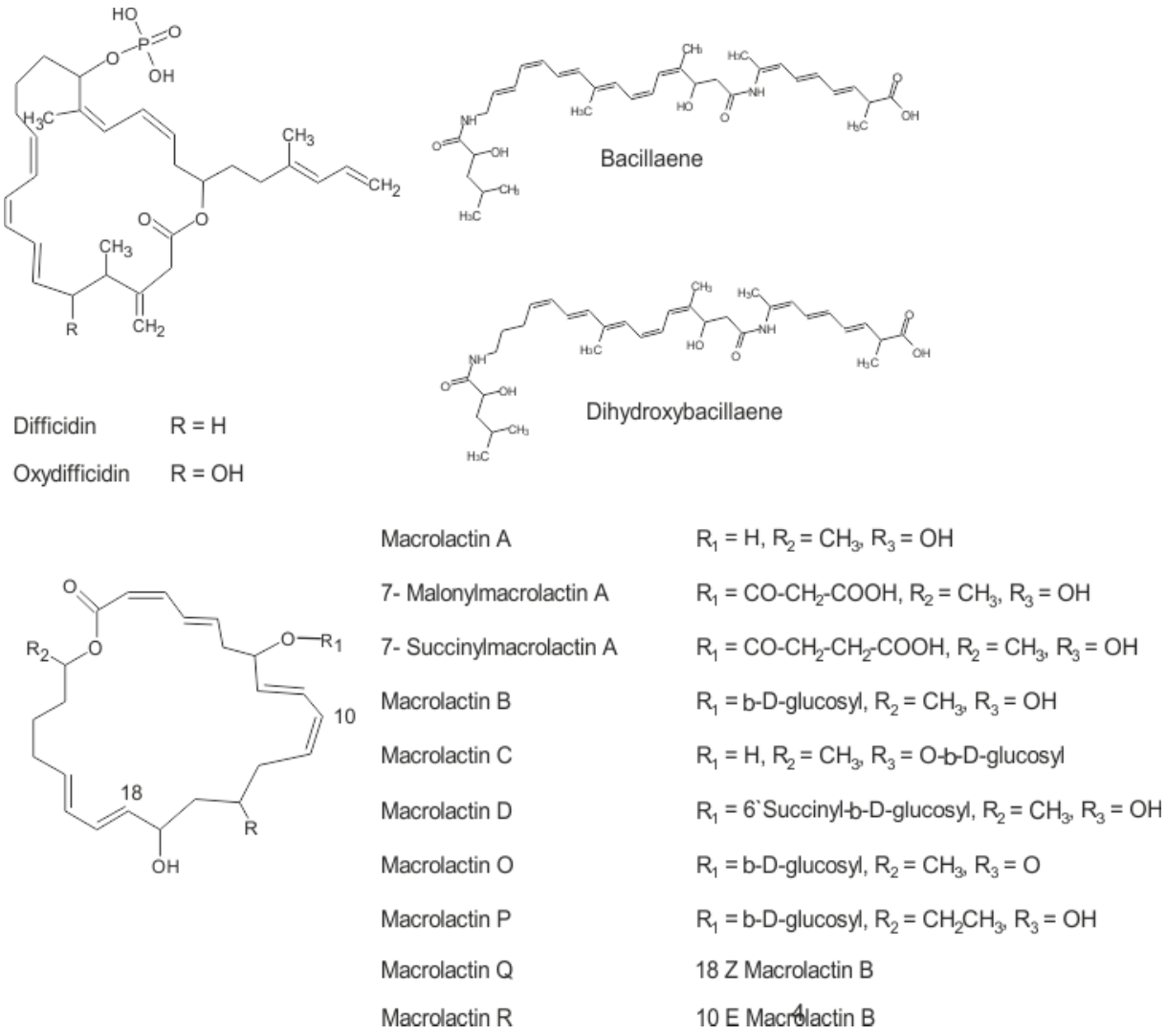


Fig.5. The structural diversity of polyketide products including bacillaene and difficidin from *B. amyloliquefaciens* FZB42, and macrolactins from *Bacillus* sp. AH159-1 (Aleti et al., 2015).

1.9.1.2 Difficidin

Difficidin was isolated from *B. subtilis* strains ATCC 39320 and ATCC 39374 (Wilson et al., 1987). The polyene antibiotic difficidin is a highly unsaturated macrolide comprising a 22-member carbon skeleton similar to bacillaene but contains a phosphate group, rarely found in secondary metabolites. Oxydifficidin is a structural variant of difficidin, which bears an additional hydroxyl group at position 5 (Wilson et al., 1987) (Fig.5).

Polyketide synthase gene cluster encoding difficidin has been characterized from *B. amyloliquefaciens* FZB42 (Fig.6). Difficidin is synthesized on a large gene cluster designated as *dif,* which contains 14 open reading frames from *difA* to *difO* (Chen et al., 2006). The *dif* system has an unusual architecture that makes it diverse from other polyketide enzymes. Biosynthesis of difficidin and its variant oxydifficidin clearly deviates from the colinearity rule because a KR domain within module 3, two DH domains within modules 4 and 9, and two ER domains within modules 2 and 8 are lacking within the polyketide synthase gene cluster that can be presumed from the polyketide structure. Nevertheless, these domain activities are accomplished by discrete enzymes found in trans. Other deviations from non-ribosomal colinearity code include sharing the biosynthesis of module 5 by both *difG* and

difH and skipping the biosynthesis of module 11 in the final product encoded by *difJ* and *difK* (Chen et al., 2006) (Fig.6).

Difficidin and oxydifficidin are shown to exhibit antagonistic activities toward a broad spectrum of bacteria including *Erwinia amylovora*, the phytopathogen responsible for fire blight (Chen et al., 2009). Difficidin and oxydifficidin also inhibit protein biosynthesis in *Escherichia coli* (Zweerink and Edison, 1987).

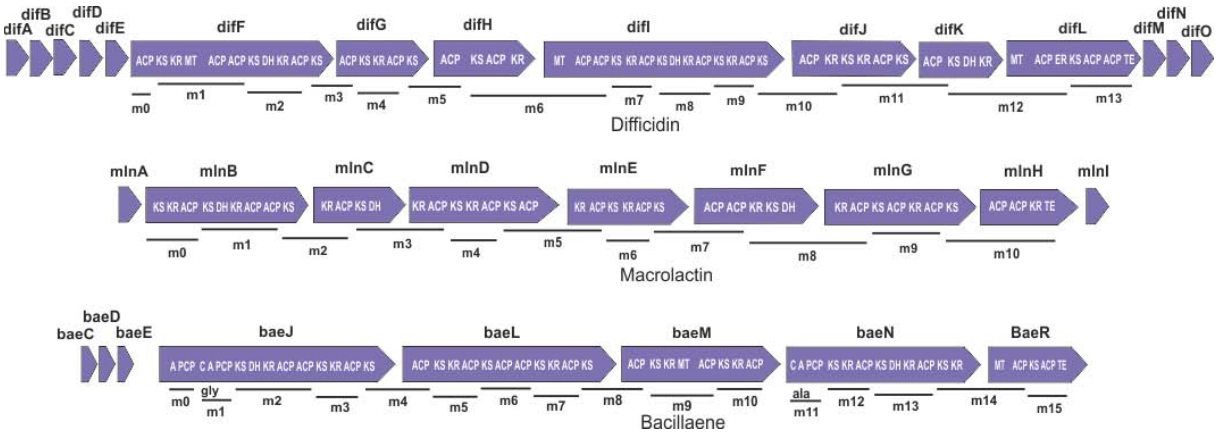


Fig.6. Gene structure of the three types of characterized polyketide synthetases from *B. amyloliquefaciens* FZB42: difficidin, macrolactin and bacillaene (adapted from Aleti et al., 2015). Iterative domains: ACP, acyl carrier protein; PCP, peptidyl carrier protein; A, adenylation; KS, ketosynthase; DH, dehydratase; MT, methyl transferase; KR, ketoreductase; TE, thioesterase.

1.9.1.3 Macrolactins

Production of macrolactins have been noticed in several bacilli including *B. amyloliquefaciens* FZB42, soil bacterium *Bacillus* sp. AH159-1 and marine *B. actinomadura*, as well as in several uncharacterized species (Schneider et al., 2007). Macrolactins contain a 24-membered lactone ring structure with three diene moieties in the carbon backbone and this ring can be attached to a glucose beta-pyranoside. (Fig.5), however, they might be also found in linear forms.

The polyketide gene cluster called mln has been assigned to the biosynthesis of cyclic macrolactins in *B. amyloliquefaciens* FZB42. The large *mln* cluster encompasses nine ORFs designated from *mlnA* to *mlnI* encoding 11 KS domains in total and utilizes malonate and acetate as building units for the biosynthesis of macrolactin backbone (Chen et al., 2006; Schneider et al., 2007) (Fig.6). Unlike the bacillaene gene cluster, the macrolactin gene cluster contains only one trans AT domain upstream to the *mlnA* operon, but organization of *mln* shows an unusual splitting of the modules similar to bacillaene gene cluster. Both *mlnB* and *mlnC* share to encode the module 2, and a similar split module organization has been noticed for modules 5, 7, 8 and 10. Comparison of macrolactin structure and the assembly order of the catalytic domains in polyketide synthase have revealed that module 2 lacks the ER domain while modules 7 and 10 lack two DH domains. Similar to difficidin assembly, the activities of missing ER and DH are accomplished by domains found in trans (Schneider et al., 2007) (Fig.6).

In *in vitro* assays, macrolactins selectively suppressed the growth of B16-F10 murine melanoma cancer cells as well as proliferation of mammalian Herpes simplex viruses, and arrested HIV proliferation in lymphoblast cells (Schneider et al., 2007). Therefore, macrolactins are of particular interest for their medical properties.

1.9.2 Polyketides of *Paenibacillus*

1.9.2.1 Paenimacrolidin

Paenimacrolidin has been isolated from *Paenibacillus* sp. F6-B70. Paenimacrolidin is a highly unstable 22-membered macrocyclic lactone ring comprising a triene in the carbon backbone (Fig.7) (Wu et al., 2011). Three of the KS domains of the paenimacrolidin synthase are found to share high similarity to that of difficidin or oxydifficidin synthase of *B. amyloliquefaciens*, also the structure of paenimacrolidin is similar to difficidin or oxydifficidin, which suggests some potential similarities in the underlying genetic architecture and the biosynthesis. Nevertheless, the complete biosynthetic gene cluster involved in the synthesis of panimacrolidin remains to be characterized.

Paenimacrolidin has shown to inhibit multidrug resistant *Staphylococcus* spp. with a promising potential in clinical applications (Wu et al., 2011).

1.9.2.2 Paenilamicins

Paenilamicins have been described in *P. larvae* DSM25430, the causative agent of a fatal infectious disease (American Foulbrood) in honey bees, but the production of paenilamicins have not been directly associated with host killing, but is rather believed to be produced against ecological niche competitors (Müller et al., 2014; Garcia-Gonzalez et al., 2014). Based on comparative genome analysis the *pam* gene encoding a large NRPS/PKS hybrid gene cluster has been found to be associated with the production of paenilamicins. The chemical structure (Fig. 7) of paenilamicins were characterized by combining techniques HPLC–ESI-MS, GC–MS, and NMR spectroscopy (Müller et al., 2014). Variants of paenilamicins with altered amino acid composition in the positions first (lysine or arginine) and fourth (lysine or ornithine) of the peptide backbone have been described. The other amino acid residues of the peptide backbone comprise alanine, N-methyl-diaminopropionic acid (mDap), serine, mDap and glycine at positions 2, 3, 5, 6 and 7, respectively. The polyketide portion of the *pam* gene cluster is involved in the generation of a 2,3,5-trihydroxy pentanoic acid and its subsequent attachment to alanine at first position. Although *pam* gene cluster lacks thioesterase for termination, the assembled paenilamicins are released via nucleophilic cleavage mediated by spermidine (Müller et al., 2014).

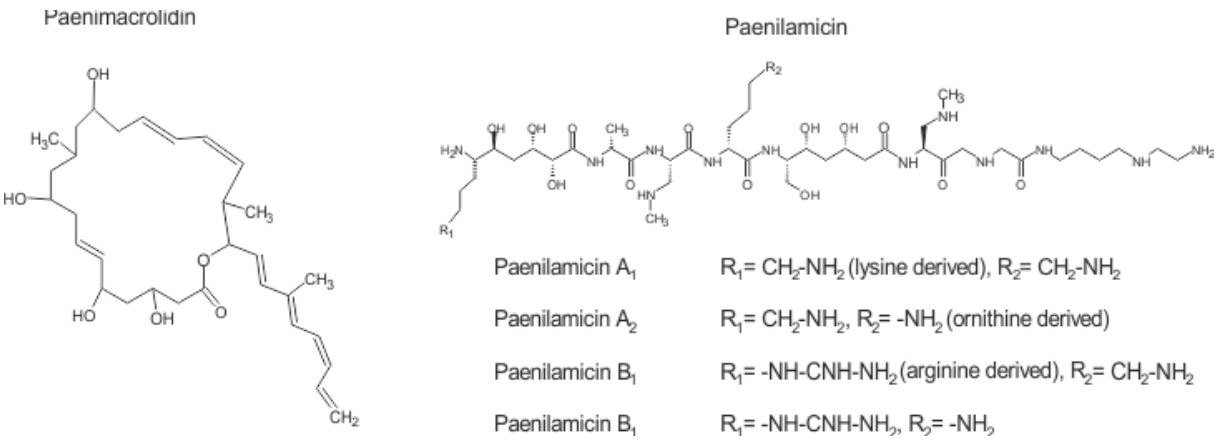


Fig.7. Chemical structures of polyketide products including paenimacrolidin from *Paenibacillus* sp. F6-B70 and paenilamicin from *P. larvae* DSM25430 (Aleti et al., 2015).

1.10 Natural roles of lipopeptides and polyketides from *Bacillus* and *Paenibacillus*

Lipopeptides of *Bacillus* and *Paenibacillus* are notably structurally heterogeneous, this suggests that these metabolites may have different ecological roles in natural habitats and may be species or group specific. The important functions of lipopeptides are their role in biocontrol activity, surface mobility and signaling for growth and differentiation and the following sections discuss their natural functions in detail (Ongena and Jacques, 2008; Raaijmakers et al., 2006; Lopez et al., 2009a, b).

1.10.1 Role in biocontrol of phytopathogens

Bacilli and Paenibacilli associated with phytosphere mount their antibiotic arsenal against pathogens invading plants (Table 3, Fig.8). Lipopeptides from these genera have been shown to exhibit lytic and growth-inhibitory activities toward viruses, mycoplasmas, bacteria, fungi and oomycetes (Fig.8). Surfactin has been shown to disintegrate the lipid envelope of various virus particles but this activity is not clear in the plant system. Surfactin also directly targets membranes of wide range human and animal pathogenic mycoplasmas (Vollenbroich et al., 1997; Ongena and Jacques, 2008).

Production of surfactin by *Bacillus* at root level is firstly important for establishing a stable relationship with the plant and secondly this can be exploited for biocontrol activity, thus outcompeting other competitors inhabiting the same niche (Bais et al., 2004). Surfactin protection is due to its direct biocontrol activity or an indirect effect caused as a result of inhibition of biofilm formation of pathogens on root surface (Ongena and Jacques, 2008).

Biocontrol activities against pathogens have been reported for lipopeptides and polyketides of *Bacillus* and *Paenibacillus* (Table 3). Previous studies have shown that biocontrol activity of iturins and fengycins against various pathogens in different plant species. For instance, iturin A from *B. subtilis* RB14 has led to reduction of damping-off in tomato, a soil-borne disease caused by *Rhizoctonia solani* (Ongena and Jacques, 2008). Also overexpression of mycosubtilin (iturin family) production in *B. subtilis* ATCC 6633 displayed significant reduction of seedling infection caused by *Pythium aphanidermatum* (Leclere, 2005). Regarding biocontrol of phyllosphere diseases, both iturins and fengycins were reported as the main antibiotic products responsible for the antagonistic activity of *B. subtilis* against *Podosphaera fusca* infecting melon leaves (Romero et al., 2007).

Table 3. Functions of secondary metabolites produced by selected *Bacillus* and *Paenibacillus* spp. in management of plant pathogens (Borris, 2015)

Metabolite	Occurrence	Gene cluster	Size (kb)	Inhibiting effect	Reference
Sfp-dependent non-ribosomal synthesis of lipopeptides					
Surfactin	BAP, BAA, BSU	*srfABCD*	32	Virus	Stein 2005
Iturin	BAP, BAA, BSU	*bmyCBAD*	39.7	Fungi	Chen et al. 2007
Fengycin	BAP, BSU	*fenABCDE*	38.2	Fungi	Chen et al. 2007
Polymyxin	PPO	*pmxABCDE*	40.7	Bacteria	Niu et al. 2013
Fusaricidin	PPO	*fus GFEDCBA*	32.4	Fungi	Li and Jensen 2008
Bacillibactin	BAP, BAA, BSU	*dhbABCDEF*	12.8	Bacterial competitors	Chen et al. 2007
Sfp-dependent non-ribosomal synthesis of polyketides					
Macrolactin	BAP	*mlnABCDEFGHI*	53.9	Bacteria	Chen et al. 2007
Bacillaene	BAP, BAA, BSU	*baeBCDE, acpK, baeGHIJLMNRS*	74.3	Bacteria	Chen et al. 2007
Difficidin	BAP	*dfnAYXBCDEFGHIJK LM*	71.1	Bacteria	Chen et al. 2007
Sfp-independent non-ribosomal synthesis					
Bacilysin	BAP, BSU	*bacABCDE, ywfG*	6.9	Bacteria, cyanobacteria	Chen et al. 2007
Ribosomal synthesis of processed and modified peptides (bacteriocins)					
Plantazolicin	BAP FZB42	*pznFKGHIAJC DBEL*	9.96	*B. anthrax*, nematodes	Scholz et al. 2011
Amylocyclicin	BAP FZB42	*acnBACDEF*	4.49	Closely related bacteria	Scholz et al. 2014
Mersacidin	BAPY2	*mrsK2R2FGEAR1DM T*	12	Gram-positive bacteria	Stein 2005
Amylolysin	BAP GA1	*amlAMTKRIFE*	9.36	Gram-positive bacteria	Arguelles Arias et al. 2014
Subtilin	BSUATCC 6633	*spaBTCAIFGRK*	12	Closely related bacteria	Stein 2005
Ericin	BAPA1/3	*eriBTCASIFEGRK*	12.5	Closely related bacteria	Stein 2005
Sublancin	BSU	*sunAT bdbA yolJ bdbB*	4.5	Closely related bacteria	Stein 2005
SubtilosinA	BSU	*sboA albABCDEFG*	7	Closely related bacteria	Stein 2005

BAP *B. amyloliquefaciens plantarum*, BAA *B. amyloliquefaciens amyloliquefaciens*, BSU *B. subtilis subtilis*, PPO *Paenibacillus polymyxa*

1.10.2 Role in motility

The rhizosphere is a nutrient rich environment, comprises a wide range of low molecular weight and macromolecular root exudates such as sugars, organic acids or amino acids (Bais, 2006). Some of the root exudates can act as signal molecules to stimulate a chemotactic response in motile bacteria like *Bacillus* and facilitate mobility on the root surface (Somers et al., 2004).

As phytopathogenic fungi and other soil associated microbes are highly motile in nature, motility is considered as an indispensable factor for rhizospheric competence and ecological fitness of the PGPB. Thus, rapid colonization by PGPR on plant roots can restrict other antagonists competing for space and nutrients in the same ecological niche (Kinsinger et al., 2003) (Fig.8). Kinsinger et al. (2003) suggested that *Bacillus* swarm to new nutrient rich niches and their secreted lipopeptides are indispensable for the motility.

Koumoutsi et al. (2004) and Raaijmakers et al. (2010) confirmed the crucial role of surfactin in swarming phenotypes of *Bacillus* amyloliquefaciens strains S499 and FZB42 by examining the swarming phenotypes of the lipopeptide mutants.

1.10.3 Lipopeptides as elicitors of induced systemic resistance in plants

Several reports suggest that *Bacillus* species including *B. amyloliquefaciens*, *B. subtilis*, *B. pasteurii*, *B. cereus*, *B. pumilus*, *B. mycoides* and *B. sphaericus* are potential elicitors of ISR in the plant (Ongena and Jacques, 2008). The role of lipopeptides in the plant environment is summarized in Fig.8.

Both fengycins and surfactins have been shown to interact with plant cells and trigger plant immune response in bean and tomato plants and the reduction in disease was largely due to the metabolic changes associated with plant defense responses (Ongena et al., 2007). Lipopeptide overproduction by *Bacillus* in tomato plants stimulated the key enzyme activities of the oxylipin pathway. In another study, application of purified fengycins on potato tuber cells led to the accumulation of plant phenolic compounds via phenylpropanoid metabolism, which is believed to be activated during the plant defense (Ongena et al., 2005). Treatment of surfactins and fengycins also triggered defense-associated early events and brought changes in the phenolic pattern of tobacco cells (Jourdan et al., 2009). Each lipopeptide family retains a specific ability to stimulate different plant cells, it is however not known, how exactly the lipopeptides are perceived by plant cells.

1.10.4 Role in chelation of metal ions

Biosurfactant role of lipopeptides in chelation of metal ions has been reported. The chelation potential of lipopeptides is largely influenced by minor changes in amino acid composition of the peptide backbone. In this regard, lichenysin with glutamine at the first position showed increased chelation of Ca2+ compared to surfactin with glutamic acid (Grangemard et al., 2001). When leucine substituted for isoleucine at position 2, the surfactin affinity for a Ca2+ ions increased threefold because of increased affinity of acidic side chains and carboxylate groups that constitute calcium-binding site (Grangemard et al., 2001). Iturin chelate metal cations such as Na+, K+ and Rb+ (Rautenbach et al., 2000). Biosurfactant properties of lipopeptides can be exploited for bioremediation of heavy metals by detachment of the metal

ion from soil particles and subsequent incorporation in the biosurfactant micelles (Mulligan, 2005).

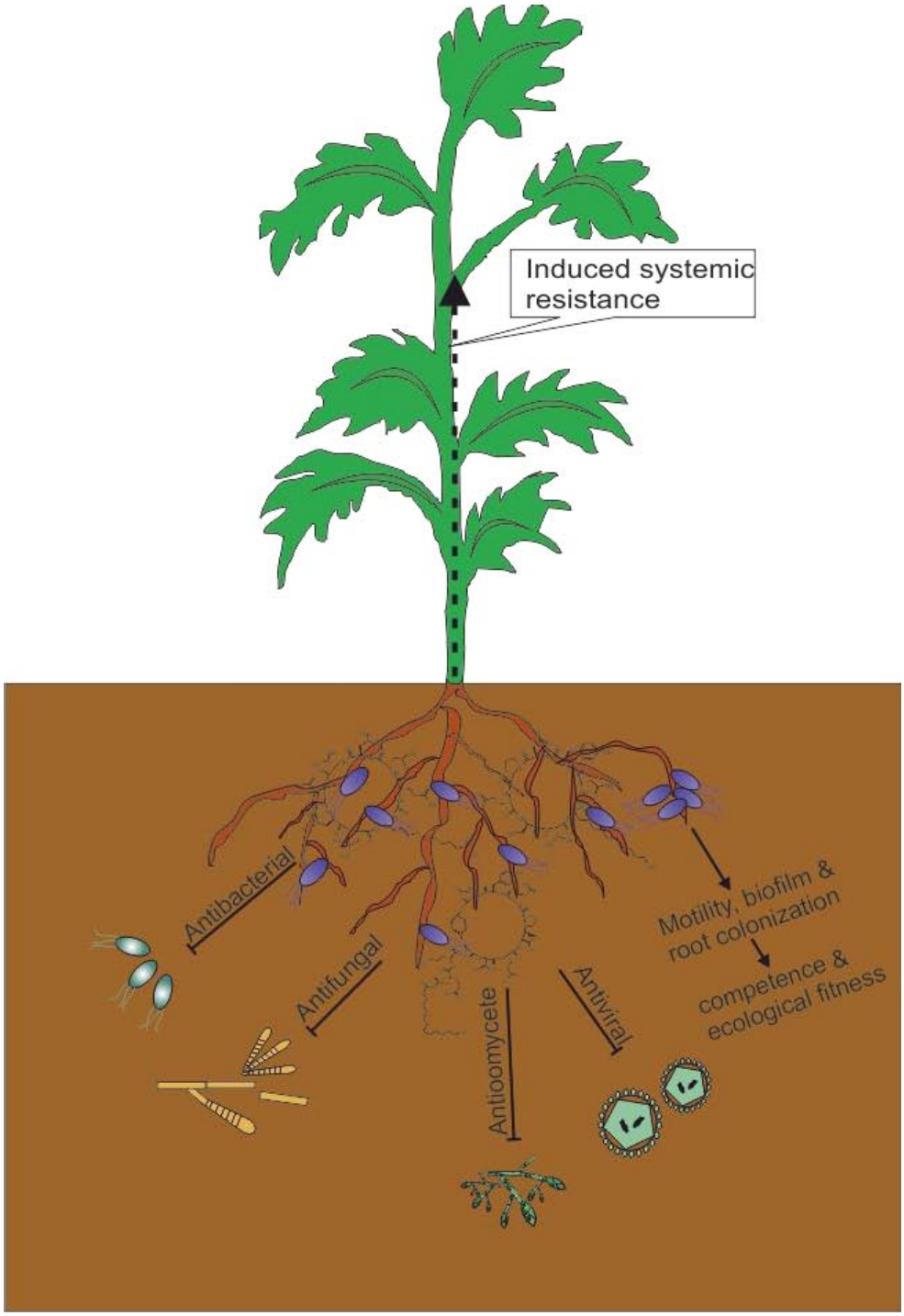


Fig.8. Overview of distinct natural functions of lipopeptides derived from *Bacillus* and *Paenibacillus* in the context of biocontrol of phytopathogens. Lipopeptides have signaling role in biofilm formation and root colonization in *Bacillus*. Moreover, lipopeptides facilitate mobility on the root surface and are largely responsible for the antimicrobial activities displayed by *Bacillus* and *Paenibacillus* species, and for induced systemic resistance in the plant.

1.10.5 Role in plant tissue colonization

Bacterial root colonization can be complex and may involve lipopeptides at different levels. Rhizobacteria attach and aggregate in clusters of cells to form biofilms and spread over the root surfaces, which is the basis for plant root colonization (Lugtenberg et al., 2001; Ramey et al., 2004) (Fig.8). Thus, *Bacillus* strains are believed to behave as highly structured biofilm communities on root and soil particle surfaces. In this regard, several studies have reported that surfactins but not fengycins or iturins are crucially involved in pellicle formation (a robust biofilm formation under laboratory conditions) at the air–liquid interface (Branda et al., 2001; Kinsinger et al., 2003; Hofemeister et al., 2004).

Spatial and temporal organization of cellular differentiation during the biofilm development has been extensively studied in *B. subtilis* (Branda et al., 2001). Bais et al. (2004) demonstrated the role of surfactins in biofilm formation. Wild type *B. subtilis* capable of surfactin production was able to develop robust biofilms on *Arabidopsis* roots while its surfactin gene mutant lacking the production of surfactin led to weak biofilms, reduced colonization and biocontrol activity. Also, the well-studied PGPR *B. amyloliquefaciens* FZB42 is known to form robust pellicles. Surfactin biosynthesis is believed to be a prerequisite for biofilm formation and swarming motility as surfactin biosynthetic mutants failed to form robust biofilms (Chen et al. 2007).

1.10.6 *Bacillus* biofilm formation

In natural environments microbes often thrive in complex sessile communities known as biofilms. In a biofilm development typically individual cells aggregate by an extracellular polymeric matrix often composed of exopolysaccharides, proteins and nucleic acids (Branda et al., 2006). Bacteria live in close proximity and interact to compete or cooperate for space and resources in an ecological niche (Nadell et al., 2009). These bacterial communications may require cell–cell interactions mediated by quorum sensing, quorum quenching molecules or by production of antimicrobial secondary metabolites (Dong and Zhang, 2005; Lopez et al., 2009 c,d).

Many strains of *B. subtilis* derived from the laboratory cultured *B. subtilis* strain 168 display significant reduction in the biofilm development due to the accumulation of mutations during laboratory propagation, a process known as domestication (McLoon et al., 2011). On the contrary, natural *B. subtilis* strains display sophisticated colony architectures and robust biofilms, which is an attribute for a wide range of undomesticated strains isolated from natural environments (Fig.9) (Mielich-Süss and Lopez, 2015). Biofilm formation can confer several advantages to bacilli, for instance, individual cells can be protected from environmental conditions when cells reach new nutrient rich niches and because of these properties biofilms have potential to be utilized in agricultural applications (Bais et al., 2004).

B. subtilis, which is capable of forming robust biofilms has been widely used for the studies on biofilm formation (Branda et al., 2001; Vlamakis et al., 2013). These biofilms are characterized by highly structured floating pellicles formed on the surface of stable liquid cultures and on agar plates (Branda et al., 2001; Vlamakis et al., 2013). In the following section, molecular process involved in the formation of biofilm in *B. subtilis* is discussed in depth (Cairns et al., 2013; Mielich-Süss and Lopez, 2015) (Fig.9). Regulatory networks controlling biofilm formation in *B. subtilis* is presented in more detail in the sections below.

1.10.6.1 Biofilm development

During the biofilm development, motile bacterial cells transform into a sessile state by down-regulation of the expression of flagellar genes and a concomitant up-regulation of genes responsible for the biosynthesis of the extracellular matrix (Cairns et al., 2013; Kolodkin-Gal et al., 2013). Individual sessile cells form cluster of chains, following encase in a self-produced exogenous matrix to form robust biofilm structures Further expansion of biofilm is enabled by the activity of motile cells and the production of surfactant molecules (Branda et al., 2006; Angelini et al., 2009) (Fig.9).

Biosynthesis of exopolysaccharide is mediated by the expression of the *epsA-O* operon (eps) (Branda et al., 2001; 2006; Kearns et al., 2005). Two additional extracellular proteins TasA and TapA encoded by the three-gene operon *tapA-sipW-tasA* provide structural integrity to the matrix while SipW facilitates secretion of TasA (an amyloid protein) into the extracellular space (Branda et al., 2006). TasA then assembles into fibres, which are anchored to the cell wall in association with TapA (Romero et al., 2011). Another protein BslA, a hydrophobin protein is secreted during the biofilm maturation, which develops into a hydrophobic layer over the biofilm and acts as a water-repellent barrier for protection of *B. subtilis* community (Hobley et al., 2013). Taken together, cells are essentially encased in an extracellular matrix comprising exopolysaccharide and protein polymers TasA and BslA (Fig.9).

1.10.6.2 Heterogeneous cell types within the biofilm

Regulation of cell differentiation in *Bacillus* is a complex network. Three main master regulators DegU, ComA and Spo0A trigger specific gene expression cascades, consequently leading to the differentiation of genetically identical cells into subpopulations of distinct phenotypes (Fig.9) (Lopez and Kolter, 2010a).

In a free-living population all constituent cells are motile because the three master regulators are in an unphosphorylated state which triggers the expression of genes involved in the motility (Guttenplan et al., 2010) (Fig.9). When cells become sessile, these master regulators are activated depending on the environmental cues. This favors differentiation of cell types with concomitant reduction of the population of motile cells, and thus the existence of subpopulation of motile cells is noticed only when the master regulators are inactive and can be considered mutually exclusive for other specialized cell types (Vlamakis et al., 2008). The remaining cell types usually share differentiation programmes and possibly overlap unlike the subpopulation of motile cells.

The expression of Spo0A~P varies with the levels of activation, high level activation of Spo0A (Spo0A~P) induces sporulation in a subpopulation of *B. subtilis* cells, while low level of activation triggers cannibalism and matrix production (Lopez et al., 2009b). The Spo0A-active cells essentially evolve as matrix producers before initiating sporulation, hence it is evident that both the differentiation programmes overlap during the transition of the cells. Spo0A-ON cells activate a molecular clutch (EpsE protein) encoded by the eps operon, which inhibits the flagellar motility (Guttenplan et al., 2010) (Fig.9). These SpoA-ON cells are cannibals since they destroy the other sensitive Spo0A-OFF cells by discharging two peptide toxins Skf and Sdp, and feed on them to compensate the nutritional limitation and, thus, postpone sporulation. However, cannibal cells are unaffected by these self-secreted toxins.

The activation of ComA (ComA~P) induces competence in cells and develops the ability to combine with external DNA to further increase genetic variability within bacterial

community. ComA activates the expression of the paracrine signal molecule surfactin triggering the matrix producers (Lopez et al., 2009c). ComA-ON cells differentiate into two different specialized cell types, surfactin producers and competent cells. Activation of DegU (DegU~P) generates miners which synthesize exoprotease enzymes that degrade exogenous proteins into tiny molecules, which are absorbed by the community (Veening et al., 2008). These distinct subpopulations are generated from either motile or matrix cells located near the air surface (Marlow et al., 2014) (Fig.9). Also, DegU-ON cells are capable of producing BslA protein that contributes to liquid repellency on the biofilm surface to maintain biofilm integrity (Hobley et al., 2013).

Motile cells located predominantly at the bottom and edges of the biofilm contribute to the expansion of the community, while matrix producing cells are present throughout the biofilm, encase extracellular matrix and maintain the rigidity (Vlamakis et al., 2008) (Fig.9). Sporulating cells are primarily found at the top of the biofilm that facilitates their dispersion (Vlamakis et al., 2008). DegU-ON cells preferentially found on the agar surface, since it cascades water repelling BslA protein layer that surrounds the biofilm (Marlow et al., 2014) (Fig.9). The spatiotemporal distribution of competent cells and surfactin producers is still unknown due to sparse distribution of this subset in the overall population.

1.10.6.3 Signal input to the cell differentiation programmes in biofilm

The activity of the master regulator protein Spo0A is regulated by phosphorylative action of five different kinases (KinA–E) (Fig.9). These kinases are involved in transfer of a phosphoryl group to Spo0A by a phosphorelay system. Spo0A~P induces the expression of a repressor SinI that in turn binds and represses the repressor SinR, as a result, represses the genes involved in the matrix biosynthesis (Kearns et al., 2005). In addition to SinI, a repressor SlrR can also inhibit the activity of SinR and thus, SinR-SlrR formation is controlled by double negative feedback loop, in which SinR-SlrR protein complex titrates SinR and prevents SinR from repressing the slrR (Fig.9). Spo0A~P inhibits AbrB, an alternative repressor regulatory protein that is involved in expression of the matrix related genes (Lopez et al., 2009d).

It has been proposed that activity of the KinA-E is governed by specific signals. For instance, activity of surfactin can generate pores in the membranes of *Bacillus*, which results in leakage of potassium ions, thus the membrane perturbations sensing by KinC results in induction of low level expression of Spo0A~P (Lopez et al., 2009d). It has been reported that plant root exudates such as glycerol, manganese and L-malic acid can trigger the activity of KinD (Beauregard et al., 2013). KinD can also act as a checkpoint protein that senses extracellular polysaccharides to trigger sporulation when critical thresholds of matrix is sensed (Aguilar et al., 2010) (Fig.9).

KinA and KinB are not only involved in sporulation but they are also responsible for biofilm formation. KinB acts in association with the respiratory apparatus, especially triggered when electron transport is impaired due to environmental stresses (e.g., low oxygen or high iron). Also KinA senses NAD+/NADH levels in the cytoplasm by direct binding to NAD+ (Kolodkin-Gal et al., 2013) (Fig.9). Based on the phosphorylation status, DegU-P can function as as an inhibitor or as an activator (Murray et al., 2009). DegU~P is believed to be activated during changes in osmolarity, in cells attached to a surface as they cease flagellum rotation (Cairns et al., 2013) and ClpC-mediated proteolysis (Ogura and Tsukahara, 2010).

For activation of the ComA-P, pheromone ComX binds and activates the ComP, which activates ComA and that in turn triggers ComA-P. This signal cascade generates natural competence and activates surfactin biosynthetic pathway (Fig.9). ComX is a quorum-sensing signal that exhibits strain specificity. Competence and sporulation factor (CSF) molecule mediates communication between different strains (Pottathil et al., 2008). Majority of the above mentioned signal cascades are inhibited by repressor family Rap phosphatase, which repress the activity of the three master regulators by dephosphorylation (Pottathil and Lazazzera, 2003). RapGH, RapABEJ and RapCFGHK possibly dephosphorylate DegU∼P, Spo0A∼P and ComA∼P, respectively (Veening et al., 2005).

The membrane of *B. subtilis* comprises microdomains named flotillins (FloT and FloA), which facilitate the activation of KinC and, thereby differentiation of SpoA-ON cells (Lopez and Kolter, 2010b). The microdomains also comprise protease FtsH, which degrade Rap phosphatases (Mielich-Süss et al., 2013). *Bacillus* biofilm development and an overview of the regulatory networks controlling the cell differentiation programs is illustrated schematically in Fig.9.

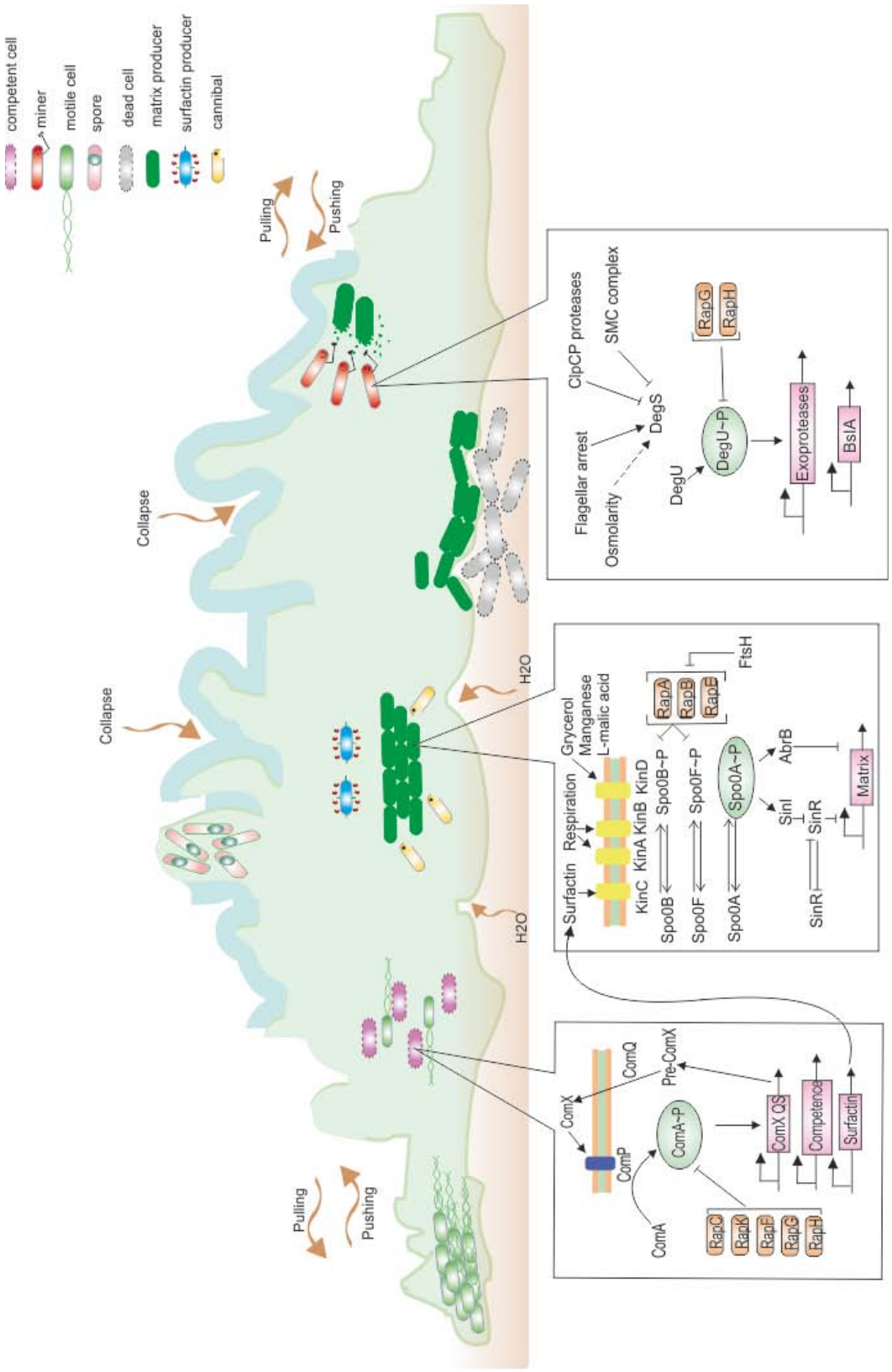


Fig.9. Regulatory networks controlling biofilm formation in *B. subtilis*. Schematic representation of spatio-temporal localization of distinct subpopulations of *B. subtilis* in a mature biofilm. The different Cell differentiation programs and the genes involved in each differentiation process are shown within the specific frame. Arrows represent activation while dashed arrows represent indirect activation, and T-bars represent repression (adapted from Cairns et al., 2014; Mielich-Süss and Lopez, 2015).

References

1. Aguilar C, Vlamakis H, Guzman A, Losick R, Kolter R. 2010. KinD is a checkpoint protein linking spore formation to extracellular-matrix production in *Bacillus subtilis* biofilms. MBio. doi: 10.1128/mBio.00035-10.
2. Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, et al. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 25: 3389–402.
3. Aleti G, Sessitsch A, Brader G. 2015. Genome mining: Prediction of lipopeptides and polyketides from *Bacillus* and related *Firmicutes*. doi: 10.1016/j.csbj..03.003.
4. Angelini TE, Roper M, Kolter R, Weitz DA, Brenner MP. 2009. *Bacillus subtilis* spreads by surfing on waves of surfactant. Proc Natl Acad Sci U S A. 106: 18109–18113
5. Aranda FJ, Teruel JA, Ortiz A. 2005. Further aspects on the hemolytic activity of the antibiotic lipopeptide iturin A. Biochim Biophys Acta 1713: 51–56.
6. Bachmann BO, Ravel J. 2009. Methods for in silico prediction of microbial polyketide and nonribosomal peptide biosynthetic pathways from DNA sequence data Methods. Enzymol 458: 181–217.
7. Bachmann BO, Van Lanen SG, Baltz RH. 2014. Microbial genome mining for accelerated natural products discovery: is a renaissance in the making. J Ind Microbiol Biotechnol 41: 175–84.
8. Bartlett DW, Clough JM, Godwin JR, Hall AA, Hamer M, Parr-Dobrzanski B. 2002. The strobilurin fungicides. Pest Manag Sci 58: 649–662.
9. Bais HP, Park SW, Weir TL, Callaway RM, Vivanco JM. 2004. How plants communicate using the underground information superhighway. Trend Plant Sci 9: 26–32.
10. Bentley SD, Chater KF, Cerdeno-Tarraga AM, Challis GL, Thomson NR, James KD et al. 2002. Complete genome sequence of the model actinomycete *Streptomyces coelicolor* A3(2). Nature 417: 141–147.
11. Beatty PH, Jensen SE. 2002. *Paenibacillus polymyxa* produces fusaricidin-type antifungal antibiotics active against *Leptosphaeria maculans*, the causative agent of blackleg disease of canola. Can J Microbiol 48(2): 159–169.
12. Balibar CJ, Vaillancourt FH, Walsh CT. 2005. Generation of D aminoacid residues in assembly of arthrofactin by dual condensation/epimerization domains. Chem Biol 12: 1189–1200.
13. Béchet M, Caradec T, Hussein W, Abderrahmani A, Chollet M, et al. 2012. Structure, biosynthesis and properties of kurstakins, nonribosomal lipopeptides from *Bacillus* spp. Appl Microbiol Biotechnol 95: 593–600.
14. Beauregard PB, Chai Y, Vlamakis H, Losick R, Kolter R. 2013. *Bacillus subtilis* biofilm induction by plant polysaccharides. Proc Natl Acad Sci U S A 110: E1621–1630.
15. Blin K, MH. Medema D. Kazempour MA, Fischbach R, Breitling, et al. 2013. antiSMASH 2.0- a versatile platform for genome mining of secondary metabolite producers. Nucleic Acids Res: W204–W212.
16. Borriss R, Chen XH, Rueckert C, Blom J, Becker A, Baumgarth B, et al. 2011. Relationship of *Bacillus amyloliquefaciens* clades associated with strains DSM 7T and *Bacillus amyloliquefaciens* subsp. plantarum subsp. nov. based on their discriminating complete genome sequences. Int J Syst Evol Microbiol 61: 1786–1801.
17. Branda SS, Gonzalez-Pastor JE, Ben-Yehuda S, Losick R, Kolter R. 2001. Fruiting body formation by *Bacillus subtilis*. Proc Natl Acad Sci U S A 98: 11621–11626.

18. Branda SS, Chu F, Kearns DB, Losick R, Kolter R. 2006. A major protein component of the *Bacillus subtilis* biofilm matrix. Mol Microbiol 59: 1229–1238.
19. Borris R. 2015. Principles of Plant-Microbe Interactions. Springer International Publishing, Switzerland 379 (ed.), Lugtenberg B., DOI 10.1007/978-3-319-08575-3_40.
20. Boddy CN. 2014. Bioinformatics tools for genome mining of polyketide and nonribosomal peptides. J Ind Microbiol Biotechnol 41: 443–50.
21. Butcher RA, Schroeder FC, Fischbach MA, Straight PD, Kolter R, et al. 2007. The identification of bacillaene, the product of the PksX megacomplex in *Bacillus subtilis* Proc Natl Acad Sci U S A 104: 1506–1509.
22. Caboche S, Pupin M, Leclere V, Fontaine A, Jacques P, Kucherov G. 2008. NORINE: a database of nonribosomal peptides. Nucleic Acids Res 36: D326–331.
23. Cimermancic P, Medema MH, Claesen J, Kurita K, Wieland Brown LC, Mavrommatis K, et al. 2014. Insights into secondary metabolism from a global analysis of prokaryotic biosynthetic gene clusters. Cell158: 412–21.
24. Cruz-Morales P, Martínez-Guerrero CE, Morales-Escalante MA, Yáñez-Guerra LA, Kopp JF, Feldmann J, et al. 2015. Recapitulation of the evolution of biosynthetic gene clusters reveals hidden chemical diversity on bacterial genomes. doi:10.1101/020503.
25. Choi SK, Park SY, Kim R, Kim SB, Lee CH, et al. 2009. Identification of a polymyxin synthetase gene cluster of *Paenibacillus polymyxa* and heterologous expression of the gene in *Bacillus subtilis.* J Bacteriol 191: 3350–3358.
26. Chen XH, Vater J, Piel J, Franke P, Scholz R, et al. 2006. Structural and functional characterization of three polyketide synthase gene clusters in *Bacillus amyloliquefaciens* FZB42. J Bacteriol 188: 4024–4036.
27. Cairns LS, Marlow VL, Bissett E, Ostrowski A, Stanley-Wall NR. 2013. A mechanical signal transmitted by the flagellum controls signalling in *Bacillus subtilis*. Mol Microbiol90: 6–21.
28. Walsh CT. 2008. The chemical versatility of natural-product assembly lines. Acc Chem Res 41: 4–10.
29. Chen XH, Koumoutsi A, Scholz R, Borriss R. 2009a. More than anticipated—Production of antibiotics and other secondary metabolites by *Bacillus amyloliquefaciens* FZB42. J Mol Microbiol Biotechnol 16: 14–24.
30. Caboche S, Pupin M, Leclere V, Fontaine A, Jacques P, Kucherov G. 2008. NORINE: a database of nonribosomal peptides. Nucleic Acids Res 36: D326–31.
31. Calderone CT, Kowtoniuk WE. Kelleher NL, Walsh CT, Dorrestein PC. 2006. Convergence of isoprene and polyketide biosynthetic machinery: isoprenyl-S-carrier proteins in the pksX pathway of *Bacillus subtilis*. Proc Natl Acad Sci U S A 103: 8977–8982.
32. Cochrane SA, Vederas JC. 2014. Lipopeptides from *Bacillus* and *Paenibacillus* spp.: a gold mine of antibiotic candidates. Med Res Rev. http://dx.doi.org/10.1002/med.21321
33. Ding R, Wu XC, Qian CD, Teng Y, Li O, et al. 2011. Isolation and identification of lipopeptide antibiotics from *Paenibacillus elgii* B69 with inhibitory activity against methicillinresistant *Staphylococcus aureus.* J Microbiol 49: 942–949.
34. Duitman EH, Hamoen LW, Rembold M, Venema G, Seitz H, et al. 1999. The mycosubtilin synthetase of *Bacillus subtilis* ATCC6633: a multifunctional hybrid between a peptide synthetase, an amino transferase, and a fatty acid synthase. Proc Natl Acad Sci U S A 96: 13294–13299
35. Den Herder G, Van Isterdael G, Beeckman T, De Smet I. 2010. The roots of a new green revolution. Trends Plant Sci 15: 600–607.

36. Dong YH, Zhang LH. 2005. Quorum sensing and quorum-quenching enzymes. J Microbiol 43: 101–109.
37. Dieckmann R, Lee YO, Liempt H, Döhren H, Kleinkauf H. 1995. Expression of an active adenylate-forming domain of peptide synthetases corresponding to acyl-CoA-synthetases. FEBS Lett 357: 212–216.
38. Delcher AL, Bratke KA, Powers EC, Salzberg SL. 2007. Identifying bacterial genes and endosymbiont DNA with glimmer. Bioinformatics 23: 673–679.
39. Eddy SR. 2011. Accelerated profile HMM searches. PLoS Comput Biol 7: e1002195.
40. Finking R, Marahiel MA. 2004. Biosynthesis of nonribosomal peptides 1. Annu Rev Microbiol 58: 453–488.
41. Fischbach MA, Walsh CT. 2006. Assembly-line enzymology for polyketide and nonribosomal peptide antibiotics: logic, machinery, and mechanisms. Chem Rev 106: 3468–3496.
42. Guttenplan SB, Blair KM, Kearns DB. 2010. The EpsE flagellar clutch is bifunctional and synergizes with EPS biosynthesis to promote *Bacillus subtilis* biofilm formation. PLoS Genet 6: e1001243.
43. Gao X, Haynes SW, D'Ames B, Wang P, Vien LP et al. 2012. Cyclization of fungal nonribosomal peptides by a terminal condensation-like domain. Nature Chem Biol 8: 823–830.
44. Grangemard I, Wallach J, Maget-Dana R, Peypoux F. 2001. Lichenysin: a more efficient cation chelator than surfactin. Appl Biochem Biotechnol 90: 199–210.
45. Guo Y, Huang E, Yuan C, Zhang L, Yousef AE. 2012. Isolation of a *Paenibacillus* sp. strain and structural elucidation of its broad-spectrum lipopeptide antibiotic. Appl Environ Microbiol 78: 3156–3165.
46. Hobley L, Ostrowski A, Rao FV, Bromley KM, Porter M, Prescott AR, et al. 2013. BslA is a self-assembling bacterial hydrophobin that coats the *Bacillus subtilis* biofilm. Proc Natl Acad Sci U S A 110: 13600–13605.
47. Hofemeister J, Conrad B, Adler B, Hofemeister B, Feesche J, Kucheryava NG, et al. 2004. Mol. Genet. Genomics 272: 363–378.
48. Heerklotz H, Seelig J. 2007. Leakage and lysis of lipid membranes induced by the lipopeptide surfactin. Eur Biophys J 36: 305–314.
49. Hathout Y, Ho YP, Ryzhov V, Demirev P, Fenselau C. 2000. Kurstakins: a new class of lipopeptides isolated from *Bacillus thuringiensis*. J Nat Prod 63: 1492–1496.
50. Ito M, Koyama Y. 1972. Jolipeptin, a new peptide antibiotic I Isolation, physico-chemical and biological characteristics. J Antibiot 25: 304–308.
51. Jacques P. Surfactin and Other Lipopeptides from *Bacillus* spp, in: A. Steinbüchel, G. Soberón-Chávez (Eds.), Biosurfactants, Microbiology Monographs, 20 (2011), pp. 57–91.
52. Kim PII, Ryu J, Kim YH, ChI YT. 2010. Production of biosurfactant lipopeptides Iturin A, fengycin and surfactin A from *Bacillus subtilis* CMB32 for control of *Colletotrichum gloeosporioides*. J Microbiol Biotechnol 20: 138–145.
53. Kinsinger RF, Shirk MC, Fall R. 2003. Rapid surface motility in *Bacillus subtilis* is dependent on extracellular surfactin and potassium ion. J Bacteriol. 185: 5627–5631.
54. Khaldi N, Seifuddin FT, Turner G, Haft D, Nierman WC,Wolfe KH, et al. 2010. SMURF: genomic mapping of fungal secondary metabolite clusters. Fungal Genet Biol 47: 736–41.
55. Kolodkin-Gal I, Elsholz AK, Muth C, Girguis PR, Kolter R, Losick R. 2013. Respiration control of multicellularity in *Bacillus subtilis* by a complex of the cytochrome chain with a membrane-embedded histidine kinase. Genes Dev 27: 887–899.

56. Koumoutsi A, Chen XH, Henne A, Liesegang H, Hitzeroth G, Franke P, et al. 2004. Structural and functional characterization of gene clusters directing nonribosomal synthesis of bioactive cyclic lipopeptides in *Bacillus amyloliquefaciens* strain FZB42. J Bacteriol 186: 1084–1096.
57. Kearns DB, Chu F, Branda SS, Kolter R, Losick R. 2005. A master regulator for biofilm formation by *Bacillus subtilis*. Mol Microbiol 55: 739–749.
58. Koehn FE, Carter GT. 2005. The evolving role of natural products in drug discovery. Nat Rev Drug Discov 4:206–20.
59. Kajimura Y, Kaneda M. 1997. Fusaricidins B, C and D, new depsipeptide antibiotics produced by *Bacillus polymyxa* KT-8: Isolation, structure elucidation and biological activity. J Antibiot 50: 220–228.
60. Kopp F, Marahiel MA. 2007. Where chemistry meets biology: the chemoenzymatic synthesis of nonribosomal peptides and polyketides. Curr Opin Biotechnol 18: 513–520.
61. Kim PII, Ryu J, Kim YH, ChI YT. 2010. Production of biosurfactant lipopeptides Iturin A, fengycin and surfactin A from *Bacillus subtilis* CMB32 for control of Colletotrichum gloeosporioides. J Microbiol Biotechnol 20: 138–145.
62. Li MH, Ung PM, Zajkowski J, Garneau-Tsodikova S, Sherman DH. 2009. Automated genome mining for natural products. BMC Bioinformatics. doi: 10.1186/1471-2105-10-185.
63. Lopez D, Kolter R. 2010a. Extracellular signals that define distinct and coexisting cell fates in *Bacillus subtilis*. FEMS Microbiol Rev 34: 134–149.
64. Lopez D, Vlamakis H, Kolter R. 2009a. Generation of multiple cell types in *Bacillus subtilis*. FEMS Microbiol Rev 33: 152–163.
65. Lopez D, Vlamakis H, Losick R, Kolter R. 2009b. Cannibalism enhances biofilm development in *Bacillus subtilis*. Mol Microbiol 74: 609–618.
66. Lopez D, Vlamakis H, Losick R, Kolter R. 2009c. Paracrine signaling in a bacterium. Genes Dev 23: 1631–1638.
67. Lopez D, Fischbach MA, Chu F, Losick R, Kolter R. 2009d. Structurally diverse natural products that cause potassium leakage trigger multicellularity in *Bacillus subtilis*. Proc Natl Acad Sci U S A106: 280–285.
68. Li JW, Vederas JC. 2009. Drug discovery and natural products: end of an era or an endless frontier. Science 10: 161–165.
69. Lee SH, Cho YE, Park S-H, Balaraju K, Park JW, Lee SW, Park K. 2013. An antibiotic fusaricidin: A cyclic depsipeptide from *Paenibacillus polymyxa* E681 induces systemic resistance against phytophthora blight of red-pepper. Phytoparasitica 41: 49–58.
70. Lohans CT, Van Belkum MJ, Cochrane SA, Huang Z, Sit CS, McMullen LM, Vederas JC. 2014. Biochemical, structural, and genetic characterization of tridecaptin A1, an antagonist of *Campylobacter jejuni*. Chembiochem 15: 243–249.
71. Leclere V, Bechet M, Adam A, Guez JS, Wathelet B, Ongena M. et al. 2005. Mycosubtilin overproduction by *Bacillus subtilis* BBG100 enhances the organism's antagonistic and biocontrol activities. Appl. Environ. Microbiol 71: 4577–4584.
72. Leclere V, Bechet M, Adam A, Guez JS, Wathelet B, Ongena M. et al. 2005. Mycosubtilin overproduction by *Bacillus subtilis* BBG100 enhances the organism's antagonistic and biocontrol activities. Appl. Environ. Microbiol 71: 4577–4584.
73. McLoon AL, Guttenplan SB, Kearns DB, Kolter R, Losick R. 2011. Tracing the domestication of a biofilm-forming bacterium. J Bacteriol 193: 2027–34.
74. Mielich-Süss B, Lopez D. 2015. Molecular mechanisms involved in *Bacillus subtilis* biofilm formation. Environ Microbiol 17: 555–565.

75. Minowa Y, Araki M, Kanehisa M. 2007. Comprehensive analysis of distinctive polyketide and nonribosomal peptide structural motifs encoded in microbial genomes. J Mol Biol 368: 1500–1517.
76. Mielich-Suss B, Schneider J, Lopez D. 2005. Overproduction of flotillin influences cell differentiation and shape in *Bacillus subtilis*. MBio 4: e00719–00713.
77. Mulligan CN. 2005. Environmental applications for biosurfactants. Environ Poll 133: 183–198.
78. Medema MH, Takano E, Breitling R. 2013. Detecting sequence homology at the gene cluster level with MultiGeneBlast. Mol Biol Evol. doi:10.1093/molbev/ mst025.
79. Mizumoto S, Hirai M, Shoda M. 2006. Production of lipopeptide antibiotic iturin A using soybean curd residue cultivated with *Bacillus subtilis* in solid-state fermentation. Appl Microbiol Biotechnol 72: 869–875.
80. McCoy LS, Roberts KD, Nation RL, Thompson PE, Velkov T, Li J, Tor Y. 2013. Polymyxins and Analogues bind to ribosomal RNA and interfere with eukaryotic translation in vitro. Chembiochem 14: 2083–2086.
81. Medema MH, Fischbach MA. 2015. Computational approaches to natural product discovery. Nat Chem Biol 11: 639–48.
82. Murray EJ, Kiley TB, Stanley-Wall NR. 2009. A pivotal role for the response regulator DegU in controlling multicellular behaviour. Microbiology 155: 1–8.
83. Marlow VL, Cianfanelli FR, Porter M, Cairns LS, Dale JK, Stanley-Wall NR. 2014. The prevalence and origin of exoprotease-producing cells in the Bacillus subtilis biofilm. Microbiology 160: 56–66.
84. Moore BS, Hopke JN. 2001. Discovery of a new bacterial polyketide biosynthetic pathway. Chem Bio Chem 2: 35–38.
85. Mofid MR, Finking R, Essen L, Marahiel MA. 2004. Structure-basedmutational analysis of the 4′-phosphopantetheinyl transferases Sfp from Bacillus subtilis: carrier protein recognition and reactionmechanism. Biochemistry 14: 4128–4136.
86. Medema MH, Blin K, Cimermancic P, de Jager V, Zakrzewski P, Fischbach MA, et al. 2011. antiSMASH: rapid identification, annotation and analysis of secondary metabolite biosynthesis gene clusters in bacterial and fungal genome sequences. Nucleic Acids Res 39: W339–46.
87. Müller S, Garcia-Gonzalez E, Mainz A, Hertlein G, Heid NC. et al. 2014. Paenilamicin: structure and biosynthesis of a hybrid nonribosomal peptide/polyketide antibiotic from the bee pathogen Paenibacillus larvae. Angew Chem Int Ed Engl 53: 10821–10825.
88. Naruse N, Tenmyo O, Kobaru S, Kamei H, Miyaki T, Konishi M, Oki T. 1990. Pumilacidin, a complex of new antiviral antibiotics. Production, isolation, chemical properties, structure and biological activity. Journal of Antibiotics 43: 267–280.
89. Niu B, Vater J, Rueckert C, Blom J, Lehmann M, Ru JJ. et al. 2013. Polymyxin P is the active principle in suppressing *phytopathogenic Erwinia* spp. by the biocontrol rhizobacterium *Paenibacillus polymyxa* M-1 BMC Microbiol 13: 137.
90. Nadell CD, Xavier JB, Foster KR. 2009. The sociobiology of biofilms. FEMS Microbiol Rev 33: 206–224.
91. Ogura M, Tsukahara K. 2010. Autoregulation of the *Bacillus subtilis* response regulator gene degU is coupled with the proteolysis of DegU-P by ClpCP. Mol Microbiol 75: 1244–1259.
92. Ongena M, Jacques P. 2008. Bacillus lipopeptides: versatile weapons for plant disease biocontrol. Trends Microbiol 16: 115–125.
93. Ongena M, Jourdan E, Adam A, Paquot M, Brans A, et al. 2007. Surfactin and fengycin lipopeptides of *Bacillus subtilis* as elicitors of induced systemic resistance in plants. Environ Microbiol 9: 1084–1090.

94. Ongena M, Jacques P, Toure Y, Destain J, Jabrane A, Thonart P. 2005. Involvement of fengycin-type lipopeptides in the multifaceted biocontrol potential of Bacillus subtilis. Appl Microbiol Biotechnol 69: 29–38.
95. Peypoux F, Bonmatin JM, Wallach J. 1999. Recent trends in the biochemistry of surfactin. Appl Microbiol Biotechnol 51: 553–563.
96. Pichard B, Larue JP, Thouvenot D. 1995. Gavaserin and saltavalin, new peptide antibiotics produced by *Bacillus polymyxa*. FEMS Microbiol Lett133: 215–218.
97. Patel PS, Huang S, Fisher S, Pirnik D, Aklonis C, et al. 1995. Bacillaene, a novel inhibitor of procaryotic protein synthesis produced by *Bacillus subtilis*: production, taxonomy, isolation, physico-chemical characterization and biological activity J Antibiot 48: 997–1003.
98. Pottathil M, Jung A, Lazazzera BA. 2008. CSF, a species-specific extracellular signaling peptide for communication among strains of *Bacillus subtilis* and *Bacillus mojavensis*. J Bacteriol 190: 4095–4099.
99. Pottathil M, Lazazzera BA. 2003. The extracellular Phr peptide-Rap phosphatase signaling circuit of *Bacillus subtilis*. Front Biosci 8: d32–45.
100. Pichard B, Larue JP, Thouvenot D. 1995. Gavaserin and saltavalin, new peptide antibiotics produced by *Bacillus polymyxa* FEMS. Microbiol Lett 133: 215–218.
101. Qian CD, Wu XC, Teng Y, Zhao WP, Li O. et al. 2012. Battacin (octapeptin B5), a new cyclic lipopeptide antibiotic from *Paenibacillus tianmuens* is active against multidrugresistant gram-negative bacteria. Antimicrob Agents Chemother 56: 1458–1465.
102. Romero D, de Vicente A, Rakotoaly RH, Dufour SE, Veening JW, Arrebola E. et al. 2007. The iturin and fengycin families of lipopeptides are key factors in antagonism of *Bacillus subtilis* toward *Podosphaera fusca*. Mol Plant Microbe Interact 20: 430–440.
103. Röttig M, Medema MH, Blin K, Weber T, Rausch C, Kohlbacher O. 2011. NRPSpredictor2—a web server for predicting NRPS adenylation domain specificity. Nucleic Acids Res 39: W362–W367.
104. Raaijmakers JM, Paulitz TC, Steinberg C, Alabouvette C, Moenne Loccoz Y. 2009. The rhizosphere: a playground and battlefield for soilborne pathogens and beneficial microorganisms. Plant Soil 321: 341–361.
105. Romero D, Vlamakis H, Losick R, Kolter R. 2011. An accessory protein required for anchoring and assembly of amyloid fibres in *B. subtilis* biofilms. Mol Microbiol 80: 1155–1168.
106. Raposo R, Gomez V, Urrutia T, Melgarejo P. 2000. Fitness of *Botrytis cinerea* associated with dicarboximide resistance. Phytopathology 90: 1246–1249.
107. Raaijmakers J, De Bruin I, Nybroe O, Ongena M. 2010. Natural functions of cyclic lipopeptides from *Bacillus* and *Pseudomonas*: more than surfactants and antibiotics. FEMS Microbiol Rev 34: 1037–1062.
108. Raza W, Yang X, Wu H, Wang Y, Xu Y, Shen Q. 2009. Isolation and characterisation of fusaricidin-type compound-producing strain of *Paenibacillus polymyxa* SQE-21 active against *Fusarium oxysporum* F. Sp. nevium. Eur J Plant Pathol 125: 471–483.
109. Rausch C, Hoof I, Weber T, Wohlleben W, Huson DH. 2007. Phylogenic analysis of condensation domains in NRPS sheds light on their functional evolution. BMC Evol Biol 7: 78.
110. Rawlings BJ. 2001. Type I polyketide biosynthesis in bacteria (Part A-erythromycin biosynthesis). Nat Prod Rep 18: 190–227.

111. Roongsawang N, Washio K, Morikawa M. 2010. Diversity of nonribosomal peptide synthetases involved in the biosynthesis of lipopeptide biosurfactants. Int J Mol Sci. 12: 141-172.
112. Stein T. 2005. *Bacillus subtilis* antibiotics: structures, syntheses and specific functions. Mol Microbiol 56: 845–857.
113. Stachelhaus T, Mootz HD, Bergendahl V, Marahiel MA. 1998. Peptide-bond formation in non-ribosomal peptide biosynthesis: catalytic role of the condensation domain. J Biol Chem. 273: 22773–22781.
114. Stachelhaus T, Walsh CT. 2000. Mutational analysis of the epimerization domain in the initiation module PheATE of gramicidin S synthetase. Biochemistry 39: 5775–5783.
115. Skinnider MA, Dejong CA, Rees PN, Johnston CW, Li H, Webster AL, et al. 2015. Genomes to natural products PRediction Informatics for Secondary Metabolomes (PRISM). Nucleic Acids Res. doi:10.1093/nar/gkv1012.
116. Steller S, Sokoll A, Wilde C, Bernhard F, Franke P, Vater J. 2004. Biochemistry. 43: 11331–11343.
117. Staunton J, Weissman KJ. 2001. Polyketide biosynthesis: a millennium review. Nat Prod Rep 18: 380–416.
118. Shen B. 2000. Biosynthesis of aromatic polyketides. Curr Top Chem 209: 1-51.
119. Somers E, Vanderleyden J, Srinivasan M. 2004. Rhizosphere bacterial signalling: a love parade beneath our feet. Crit Rev Microbiol 304: 205–240.
120. Soon RL, Velkov T, Chiu F, Thompson PE, Kancharla R, et al. 2011. Design, synthesis, and evaluation of a new fluorescent probe for measuring polymyxin lipopolysaccharide binding interactions. Anal Biochem 409: 273–283.
121. Schneider K, Chen XH, Vater J, Franke P, Nicholson G. et al. 2007. Macrolactin is the polyketide biosynthesis product of the pks2 cluster of *Bacillus amyloliquefaciens* FZB42. J Nat Prod 70: 1417–1423.
122. Thakore Y. 2006. The biopesticide market for global agricultural use. Ind Biotechnol. 2: 194–208.
123. Vollenbroich D, Ozel M, Vater J, Kamp RM, Pauli G. 1997. Mechanism of inactivation of enveloped viruses by the biosurfactant surfactin from *Bacillus subtilis* Biologicals 25: 289–297.
124. Veening JW, Igoshin OA, Eijlander RT, Nijland R, Hamoen LW, Kuipers OP. 2008. Transient heterogeneity in extracellular protease production by *Bacillus subtilis*. Mol Syst Biol 4: 184.
125. Veening JW, Hamoen LW, Kuipers OP. 2005. Phosphatases modulate the bistable sporulation gene expression pattern in *Bacillus subtilis*. Mol Microbiol 56: 1481–1494.
126. Vlamakis H, Aguilar C, Losick R, Kolter R 2008. Control of cell fate by the formation of an architecturally complex bacterial community. Genes Dev 22: 945–953.
127. Vlamakis H, Chai Y, Beauregard P, Losick R, Kolter R. 2013. Sticking together: building a biofilm the *Bacillus subtilis* way. Nat Rev Microbiol 11: 157–168.
128. Velkov T, Thompson PE, Nation RL, Li J. 2010. Structure-activity relationships of polymyxin antibiotics. J Med Chem 53: 1898–1916.
129. Wenzel SC, Müller R. 2005. Formation of novel secondary metabolites by bacterial multimodular assembly lines: deviations from textbook biosynthetic logic. Curr Opin Chem Biol 9: 447–458.
130. Walsh CT. 2004. Polyketides and nonribosomal peptide antibiotics: modularity and versatility Science 303: 1805–1810.

131. Weber T. 2014. In silico tools for the analysis of antibiotic biosynthetic pathways. Int J Med Microbiol 304: 230–5.
132. Weber T and Kim HU. 2016. The secondary metabolite bioinformatics portal: Computational tools to facilitate synthetic biology of secondary metabolite production. Synthetic and Systems Biotechnology. doi: 10.1016/j.synbio.2015.12.002.
133. Weber T, Charusanti P, Musiol-Kroll EM, Jiang X, Tong Y, Kim HU, et al. 2015. Metabolic engineering of antibiotic factories: new tools for antibiotic production in actinomycetes. Trends Biotechnol 33: 15–26.
134. Wanga H, Fewera DP, Holmb L, Rouhiainena L, Sivonena 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
135. Wu XC, Qian CD, Fang HH, Wen YP, Zhou JY, et al. 2011. Paenimacrolidin, a novel macrolide antibiotic from *Paenibacillus* sp. F6-B70 active against methicillin-resistant *Staphylococcus aureus*. Microb Biotechnol 4: 491–502
136. Wilson KE, Flor JE, Schwartz RE, Joshua H, Smith JL, et al. 1987. Difficidin and oxydifficidin: novel broad spectrum antibacterial antibiotics produced by *Bacillus subtilis*, II Isolation and physicochemical characterization. J Antibiot 40: 1682–1691.
137. Yakimov MM, Abraham WR, Meyer H, Laura Giuliano, Golyshin PN. 1999. Structural characterization of lichenysin A component by fast atom bombardment tandem mass spectrometry. Biochim Biophys Acta. May 18;1438(2):273-80.
138. Yu GY, Sinclair JB, Hartman GL, Bertagnolli BL. 2002. Production of iturin A by *Bacillus amyloliquefaciens* suppressing *Rhizoctonia solani*. Soil Biol Biochem. 34: 955–963
139. Ziemert N, Podell S, Penn K, Badger JH, Allen E et al. 2012. The natural product domain seeker NaPDoS: a phylogeny based bioinformatic tool to classify secondary metabolite gene diversity. PLoS ONE 7: e34064. http://napdos.ucsd.edu/napdos_home.html.
140. Zweerink MM, Edison A. 1987. Difficidin and oxydifficidin: novel broad spectrum antibacterial antibiotics produced by *Bacillus subtilis*, III Mode of action of difficidin J Antibiot. 40: 1692–1697.

Objectives of the thesis

This Ph.D. thesis mainly focuses on unraveling the potential secondary metabolite capacity of representative members of *Bacillus* and *Paenibacillus* genera to produce, especially lipopeptides and type 1 polyketides with biocontrol potential against broad spectrum of phytopathogens, as well as to elucidate the underlying gene clusters encoding the production of secondary metabolites of these types, and in particular elucidate surfactin signaling role in species-specific interactions among *Bacillus*. More specific aims were as follows:

1) To review the structural diversity and the underlying gene clusters for known lipopeptides and type 1 polyketides synthesized by Bacillales and to evaluate by genome mining the yet uncharacterized gene clusters encoding unidentified secondary metabolites of these types in distinct taxonomic groups of the *Bacillus* and related genera within the Firmicutes based on the sequence information.

2) To sequence and predict the secondary metabolite capacity of *Paenibacillus polymyxa* strain CCI-25 based on the draft genome sequence.

3) To assess the metabolic potential of *B. atrophaeus* strain 176s to produce cyclic lipopeptides with antifungal activities against *Rhizoctonia solani* infection in plants and demonstrate that the surfactin variants differing by subtle structural differences synthesized in a species-specific manner are acting in their cognate form as signal molecules for biofilm formation and root colonization in *Bacillus*.

Outline of the thesis

The thesis is structured in 3 chapters (one paper per chapter) besides the general introduction, objectives, thesis outline, and final conclusions and future perspectives. The Introduction section describes in detail the background and objectives of the research, as well as provides up-to-date information on chemical structures and the underlying modular architecture of gene clusters of lipopeptides and polyketides from *Bacillus* and *Paenibacillus* and their roles in the context of biocontrol.

Chapter 1 systematically reviews current knowledge on structural and functional information, and the underlying gene clusters of well-noted structures of lipopeptides and polyketides produced by Bacillales. Moreover, it shows anticipated novel compounds by genome mining the published genome sequences of Bacillales. Using secondary metabolite prediction tools, we identified here several novel gene clusters of lipopeptides and polyketides from the genomes of Bacillales deposited in the public database. Our findings suggest that majority of the plant-associated genera, predominantly *Bacillus* and *Paenibacillus* harbor secondary metabolite biosynthetic gene clusters predicted to encode lipopeptides and type I polyketides and intriguingly many of these are uncharacterized, and their functions remain to be studied. While many genera from other environments sparsely encode such compounds indicating the role of these secondary metabolites in plant-associated niches.

Chapter 2 highlights the secondary metabolite capacity of *P. polymyxa* strain CCI-25 isolated from vermicompost. The draft genome sequence of *P. polymyxa* strain CCI-25 encompasses several non-ribosomal peptide synthetases predicted to encode varaints of tridecaptin and polymyxin, fusaricidin C, an iturin-like synthetase and a lantibiotic similar to paenicidin A, as well as a type 1 polyketide synthase. Given the fact that a gene fragment of about 370 kb, which corresponds to 6.6% of the total genome, is devoted to secondary metabolite production, CCI-25 has high potential to be exploited for medical or agricultural applications.

The biocontrol potential of *Bacillus atrophaeus* strain 176s and its capacity to produce cyclic lipopeptides with emphasis on surfactin variant production as signals for species-specific biofilm induction and root colonization is described in chapter 3. Our findings show that *B. atrophaeus* 176s synthesizing three families of cyclic lipopeptides including fengycins, iturins and surfactins have been shown to protect plants against *Rhizoctonia solani* infection. In *B. atrophaeus* we identified for the first time the production of variant surfactin C with subtle structural differences in amino acid composition when compared to surfactin A produced by *B. subtilis* and *B. amyloliquefaciens*. Examining the surfactin gene cluster revealed that the dissimilarity is encoded by the adenylation domain of *srfC* and further demonstrated that the biosynthesis of these surfactin variations is species-specific in bacilli and may have varying signal strengths on biofilm induction and root colonization dependent on the producing strain. This became evident as biofilm formation and root colonization in surfactin biosynthesis mutants was restored differentially in the presence of exogenously supplemented cognate and non-cognate surfactin variants. Furthermore, we show that surfactin variations are distributed in a species-specific manner in bacilli suggesting an intrinsic signaling role in biofilm formation. Our findings may prove to be important for biocontrol applications.

Publications

Chapter 1 Genome mining: Prediction of lipopeptides and polyketides from *Bacillus* and related Firmicutes

Published in
Computational and Structural Biotechnology Journal 2015, Mar 24;13:192-203
doi: 10.1016/j.csbj.2015.03.003. eCollection 2015.

Authors
Aleti G, Sessitsch A, Brader G

Author information
AIT Austrian Institute of Technology GmbH, Health & Environment Department, Bioresources Unit, Konrad Lorenz Straße 24, A-3430 Tulln, Austria.

Computational and Structural Biotechnology Journal 13 (2015) 192–203

COMPUTATIONAL AND STRUCTURAL BIOTECHNOLOGY JOURNAL

journal homepage: www.elsevier.com/locate/csbj

Mini Review

Genome mining: Prediction of lipopeptides and polyketides from *Bacillus* and related Firmicutes

Gajender Aleti, Angela Sessitsch, Günter Brader *

AIT Austrian Institute of Technology GmbH, AIT, Health & Environment Department, Bioresources Unit, Konrad Lorenz Strasse 24, A-3430 Tulln, Austria

ARTICLE INFO

Article history:
Received 1 November 2014
Received in revised form 15 March 2015
Accepted 16 March 2015
Available online 24 March 2015

Keywords:
Polyketides
Lipopeptides
Non-ribosomal protein synthetase
Genome mining
Structure prediction
Paenibacillus

ABSTRACT

Bacillus and related genera in the Bacillales within the Firmicutes harbor a variety of secondary metabolite gene clusters encoding polyketide synthases and non-ribosomal peptide synthetases responsible for remarkable diverse number of polyketides (PKs) and lipopeptides (LPs). These compounds may be utilized for medical and agricultural applications. Here, we summarize the knowledge on structural diversity and underlying gene clusters of LPs and PKs in the Bacillales. Moreover, we evaluate by using published prediction tools the potential metabolic capacity of these bacteria to produce type I PKs or LPs. The huge sequence repository of bacterial genomes and metagenomes provides the basis for such genome-mining to reveal the potential for novel structurally diverse secondary metabolites. The otherwise cumbersome task to isolate often unstable PKs and deduce their structure can be streamlined. Using web based prediction tools, we identified here several novel clusters of PKs and LPs from genomes deposited in the database. Our analysis suggests that a substantial fraction of predicted LPs and type I PKs are uncharacterized, and their functions remain to be studied. Known and predicted LPs and PKs occurred in the majority of the plant associated genera, predominantly in *Bacillus* and *Paenibacillus*. Surprisingly, many genera from other environments contain no or few of such compounds indicating the role of these secondary metabolites in plant-associated niches.

Contents

* Corresponding author.
E-mail address: guenter.brader@ait.ac.at (G. Brader).

http://dx.doi.org/10.1016/j.csbj.2015.03.003

1. Introduction

Bacteria are known to produce structurally diverse secondary metabolites including aminoglycosides, polyketides (PKs) and several small proteinaceous and peptidal structures such as bacteriocins, oligopeptides and lipopeptides (LPs) [1–3]. A substantial number of these metabolites have been described for their bactericidal, immune suppression and tumor suppression properties and represent potentially valuable agents in medical and veterinary medical applications, but especially PKs and LPs play also essential roles for applications in agriculture. They are vital for bacterial activities in suppressing disease pressure in plants by antimicrobial activities and activating plant defense and are important for biofilm formation and root colonization of crop plants [4–8]. LPs and PKs encompass a variety of cyclic, linear and branched structures and are generated by complex enzymes known as non-ribosomal peptide synthetases (NRPS) and polyketide synthases (PKS), respectively [9,10]. NRPS and type I PKS share to a large extent similar modular architecture and are largely organized into modules containing multiple domains, allowing the repetitive incorporation of building blocks into larger resulting compounds [11]. However, for the biosynthesis of smaller compounds (e.g. some siderophores), non-modular NRPS have been reported [12]. Often NRPS and type I PKS enzymes work using a co-linearity code, so that the recruitment of amino acids (for NRPS) and carboxylic acids (for PKS) for the biosynthesis and final structure assembly is the same as the order of catalytic domains in the genome [13,14]. This feature and insight into the architecture of modules and domains of NRPS and PKS often facilitate prediction of compound structures based on genomic sequences [15,16]. Nevertheless, variations from this conventional organization have been described and include for instance module iteration and skipping in several biosynthetic processes [17].

In this review, we will focus on Bacillales, an order belonging to the phylum Firmicutes, as genera within this order represent a rich source for diverse secondary metabolite gene clusters. Based on a recent whole genome mining study, 31% of the Firmicutes are estimated to harbor NRPS and PKS secondary metabolite gene clusters. 70% of these encode NRPS and 30% hybrid NRPS/PKS or PKS [18]. The total percentage of Firmicutes producing secondary metabolites is certainly higher, also because genes responsible for many common secondary metabolite classes (e.g. many oligosaccharides) are not detected by widely used prediction tools such as antiSMASH[19,20]. The distribution of NRPS and PKS gene clusters within different orders of the Firmicutes is not uniform and *Bacillus* and *Paenibacillus* from the order Bacillales dominate this secondary metabolite gene clusters count. These two genera in particular are well noted for their capability to produce structurally diverse LPs and PKs [4,7], but the genome information from most other Bacillales members remains largely untapped.

Despite the fact that next generation sequencing technology has contributed to the ample availability of the whole genome sequence data and a number of analysis tools for metabolite prediction exist [19–23], yet little is accomplished to explore the sequence wealth to identify novel LPs and PKs in these genomes and to predict uncharacterized secondary metabolites. We briefly review current knowledge on well characterized LPs and PKs from the Bacillales and show which novel compounds can be anticipated based on published Bacillales genome data using genome mining study and secondary metabolite prediction tools. The questions addressed here are to review the structural and functional information and the underlying gene clusters of known type I PKs and LPs produced by Bacillales and to elucidate by genome mining potential products of uncharacterized gene clusters and the potential of producing yet unidentified secondary metabolites of these types in distinct taxonomic groups of the Bacillales.

1.1. Bacillus and Paenibacillus polyketides

Polyketides are generated from simpler building units by repeated decarboxylation and condensation cycles on PKS enzymes [24]. The PKS machinery comprises three core domains: the acyl transferase (AT), the acyl carrier protein (ACP) and the ketosynthase (KS). The AT domain is responsible for activation and transfer of a simpler building unit (malonyl coenzyme A) to the ACP domain. The KS domain catalyzes decarboxylation and condensation reaction between the two ACP linked malonates [25]. Other domains include ketoreductases (KR) which catalyze hydroxy group formation, dehydratases (DH) which form double bonds after water elimination, enoyl reductases (ER) which catalyzes reduction reaction of the double bonds and methyl transferases (MT) which introduce methyl groups and branching in the carbon backbone. A phosphopantetheinyl transferase (PPT) encoded by a *sfp* gene is essential for the activation of the ACP domains [26,27]. The arrangement and the order of the catalytic domains within PKS influence PKs biosynthesis leading to a remarkable diversity in the PKs production. The PKS enzymes can be broadly categorized into three types, depending on the architecture of catalytic domains [28]. Type I PKS enzymes contain modules organized in multiple catalytic domains within a single protein that carry out decarboxylation and condensation steps to generate PKs from the starter unit malonyl-CoA [11]. In the type II and type III PKS enzymes, catalytic domains are found in separate proteins [28]. A large group of bacterial PKs are produced by modular PKS I enzymes with iterative KS, ACP and modification domains. These type I PKS mostly lack AT domains within the clusters, malonyl-CoA is transfered by acyl transferases acting in trans [29]. A large number of PKS is often found in association with NRPS as hybrid enzymes type I PKS-NRPS [30].

Metabolites produced by *Bacillus amyloliquefaciens* and *Bacillus subtilis* represent a substantial part of the diversity of LPs and PKs from the genus *Bacillus* [31,32]. The majority of the plant growth promoting and biocontrol agents commercially available are produced by these two species [4]. They produce three types of polyene PKs comprising bacillaene, difficidin and macrolactin [26,32]. *B. amyloliquefaciens* FZB42 contains a genome size of 3918 kb, of which nearly 200 kb are devoted to the production of polyketides. These three giant PKs gene clusters were assigned unambiguously by a mutagenesis study, utilizing MALDI-TOF MS and LC-ESI MS techniques [26]. In the genus *Paenibacillus* two PKs have been described so far. The underlying genetic cluster remains to be unambiguously identified in the case of paenimacrolidin [33], while for the recently described paenilamicins from *Paenibacillus larvae* also the responsible gene clusters have been reported [34]. In the following we describe the five known types of PKS from *Bacillus* and *Paenibacillus* in more detail.

1.1.1. Bacillaene

Bacillaene was first reported in the culture medium of *B. subtilis* strains 3610, and 55422 [35,36]. It has a linear structure comprising a conjugated hexaene (Fig. 2A) [35,36]. The biosynthesis of bacillaene has been described in *B. amyloliquefaciens* FZB42 and is encoded by a hybrid type I PKS-NRPS gene cluster called *bae* [26] (Fig. 1A). This cluster shares architectural characteristics with *pksX* of *B. subtilis* strain 168, presumably also encoding bacillaene [26]. The *bae* gene cluster contains five long open reading frames (ORFs) including *baeJ*, *baeL*, *baeM*, *baeN* and *baeR* [26]. The first and the second adenylation domains of *baeJ* are responsible for the incorporation of α-hydroxy-isocaproic acid and glycine, respectively. The third adenylation domain of *baeN* is involved in the incorporation of alanine [37]. Modules 4, 8 and 14 are splitted between adjacent genes (Fig. 1A). Three short ORFs found upstream of *baeJ* are *baeC, baeD, baeE*, encode for the three discrete AT domains that load malonyl-CoA [37]. Bacillaene and dihydrobacillaene are structural variants represented in this group of PKs [27,36] (Fig. 2A). Cell viable assays revealed that bacillaene selectively inhibits protein biosynthesis in prokaryotes, but not in eukaryotes,

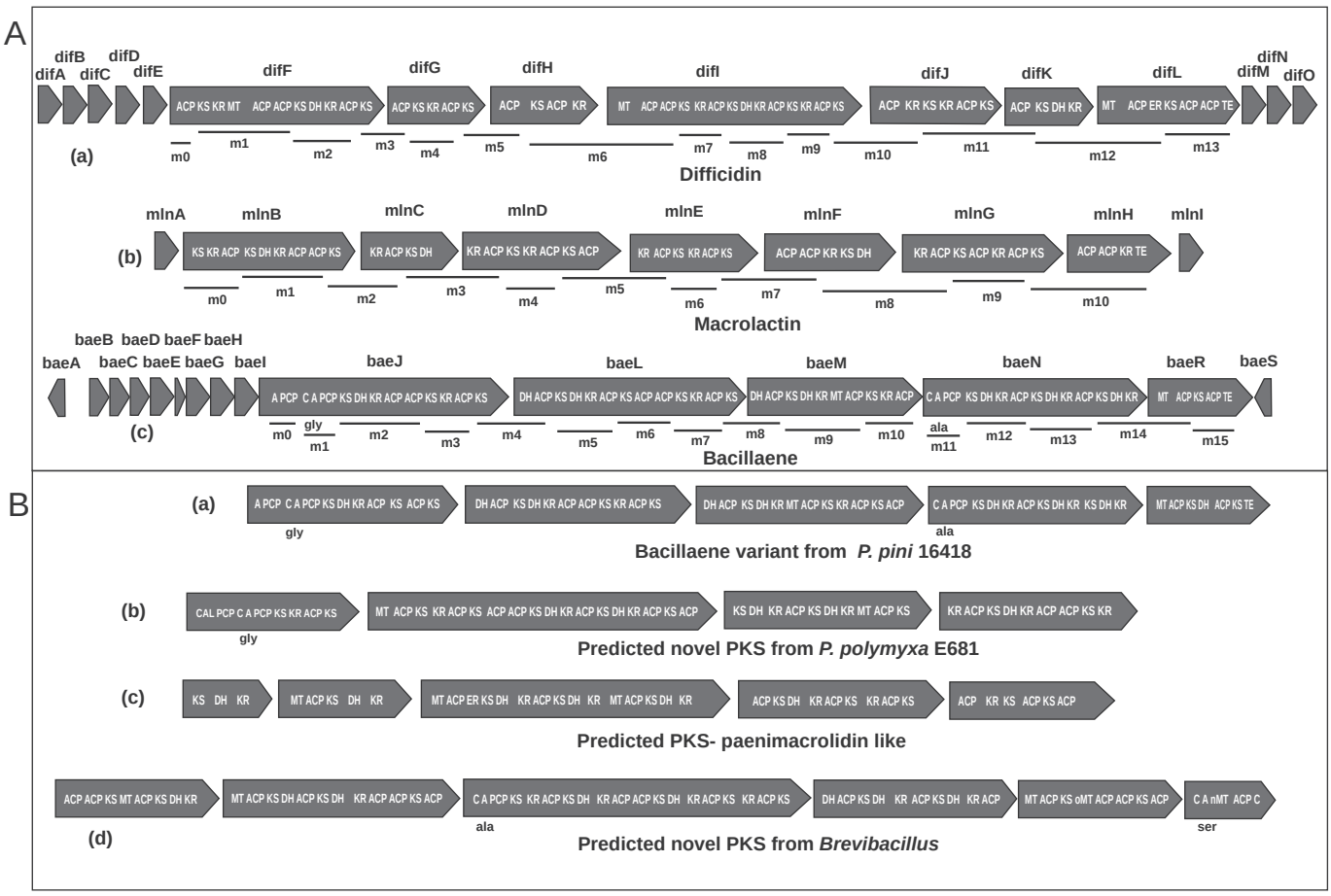


Fig. 1. Architectures of type I polyketide synthases (PKS) showing similarities and dissimilarities in known and predicted PKs. Iterative domains: ACP, acyl carrier protein; PCP, peptidyl carrier protein; A, adenylation; KS, ketosynthase; DH, dehydratase; MT, methyl transferase; KR, ketoreductase; TE, thioesterase. Further details of domains are described in Table 1. Modules and recruited amino acids indicated below, gene names indicated above each illustration. (A) Gene clusters of the three types of well-known PKS from *B. amyloliquefaciens* FZB42: (a) difficidin, (b) macrolactin, (c) bacillaene. Modular regions of predicted PKS: (a) bacillaene variant from *P. pini* 16418; number and order of the domains differ from *B. amyloliquefaciens* FZB42 bacillaene, (b) novel PKS from *P. polymyxa* E681; an adenylation domain specifies glycine, (c) paenimacrolidine like PKS from P. durus DSM 1735, (d) novel PKS form *Brevibacillus brevis NBRC 100599;* adenylation domains specify ala and ser, also contains the methylation domains- oMT and nMT. These predicted PKS machinery in *Paenibacillus* may work without thioesterase.

indicating a potential selective inhibition of other prokaryotes in their environment [35].

1.1.2. Difficidin

Difficidin is known to be produced by *B. amyloliquefaciens* strains ATCC 39320 and ATCC 39374 (originally classified as *B. subtilis* in the original paper [38]), and is a highly unsaturated macrocyclic polyene comprising a 22 member carbon skeleton with a phosphate group rarely found in secondary metabolites. Oxydifficidin, a structural variant of difficidin has an additional hydroxyl group incorporated at position 5 [38] (Fig. 2A). Difficidin is encoded by the gene cluster dif with 14 open reading frames from *difA* to *difN* and *difY* (Fig. 1A). Difficidin and oxydifficidin biosynthesis deviates from the colinearity rule as a number of KR, DH and ER domains are absent within the gene cluster. So module 3 lacks the KR domain, module 4 and 5 two DH domains and modules 2 and 8 two ER domains, but these domains are found acting in trans. The contribution of the genes difJ and difK are unclear and their potential activities are not seen in the final product [26]. Difficidin shows antagonistic activity against broad range of bacteria [39]. Difficidin has been shown to be active against the phytopathogen *Erwinia amylovora* causing fire blight [31]. In *Escherichia coli* it has been demonstrated that difficidin is responsible for inhibiting protein biosynthesis [40].

1.1.3. Macrolactin

Macrolactins have been isolated from *B. amyloliquefaciens* FZB42, the soil bacterium *Bacillus* sp. AH159-1 and from marine *Bacillus, Actinomadura* and uncharacterized species [41,42]. Most macrolactines are consisting of a 24 membered lactone ring with three diene moieties in the carbon backbone (Fig. 2A). The cyclic macrolactins are encoded in *B. amyloliquefaciens* FZB42 by the gene cluster *mln*, containing nine operons including *mlnA-I* [42] (Fig. 1A). The cluster contains 11 KS domains with malonate and acetate as the only used building units. Unlike in the bacillaene gene cluster, only one trans AT domain is found upstream of the *mlnA* gene. Similar to the *dif* gene cluster organization, *mln* shows an unusual splitting of the modules. Module 2 is splitted between *mlnB* and *mlnC* and a similar organization is seen for modules 5, 7, 8 and 10. A comparison of the order of the catalytic domains has shown that module 2 lacks the ER domain while modules 7 and 10 lack two DH domains. Like in *dif*, the activity of the missing domains can be accomplished by domains located in trans [42].

As the other *Bacillus* polyketides, macrolactins show antibacterial activity and might have the potential to be used in medical application [42]. In *in vitro* assays, they have also been shown to inhibit the proliferation of murine melanoma cancer cells and the replication of mammalian *Herpes simplex* virus and HIV in lymphoblast cells [43].

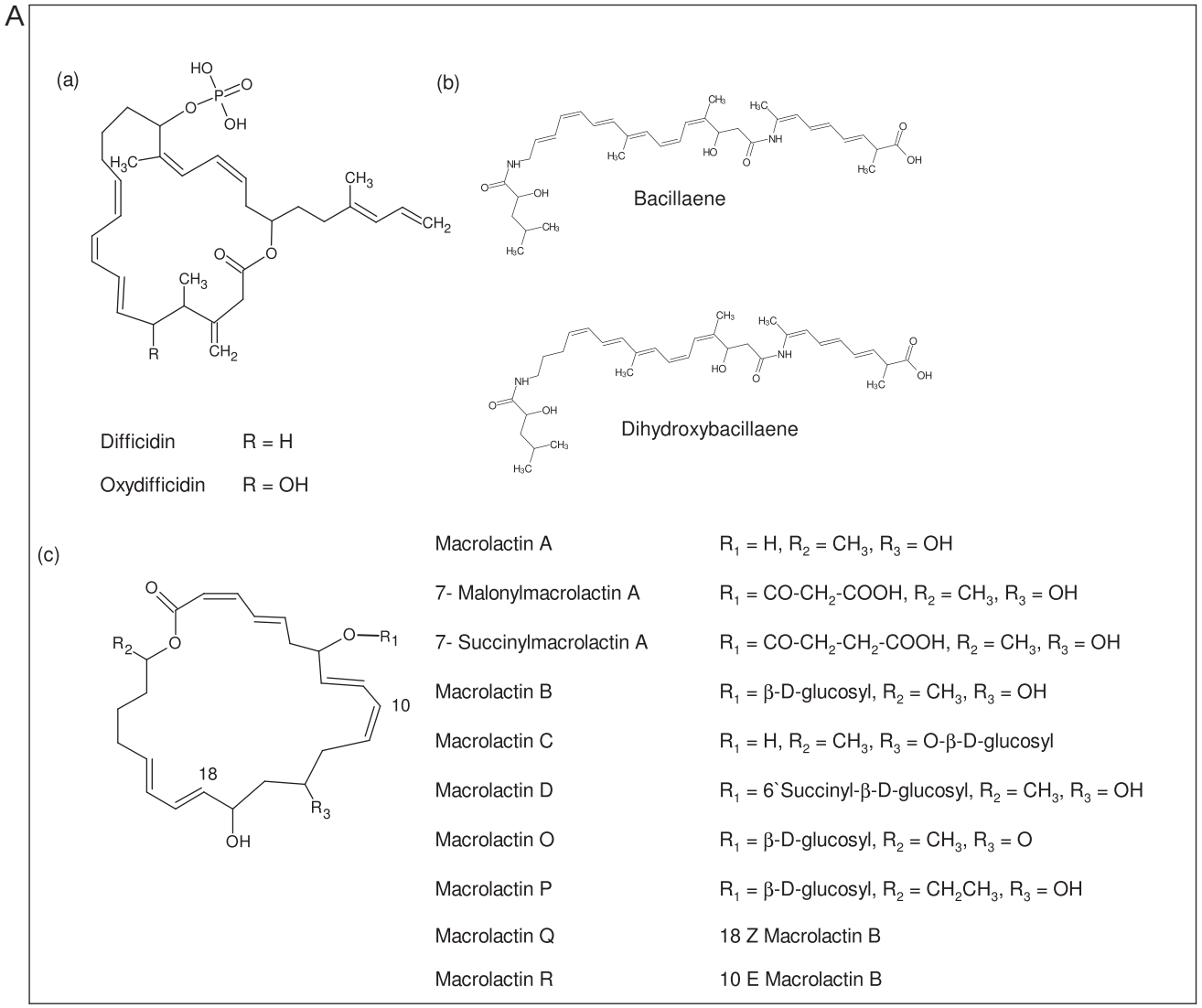


Fig. 2. Chemical structures of polyketides of *Bacillus* and *Paenibacillus*. (A) Polyketides from *B. amyloliquefaciens* FZB42 (a, b, c) and *Bacillus* sp. AH159-1 (c): (a) difficidins, (b) bacillaenes and (c) macrolactins. Stereochemistry not shown. (B) Polyketides from *Paenibacillus*: (a) Paenimacrolidin from *Paenibacillus* sp. F6-B70. Stereochemistry unknown. (b) Paenilamicin from *P. larvae* DSM25430.

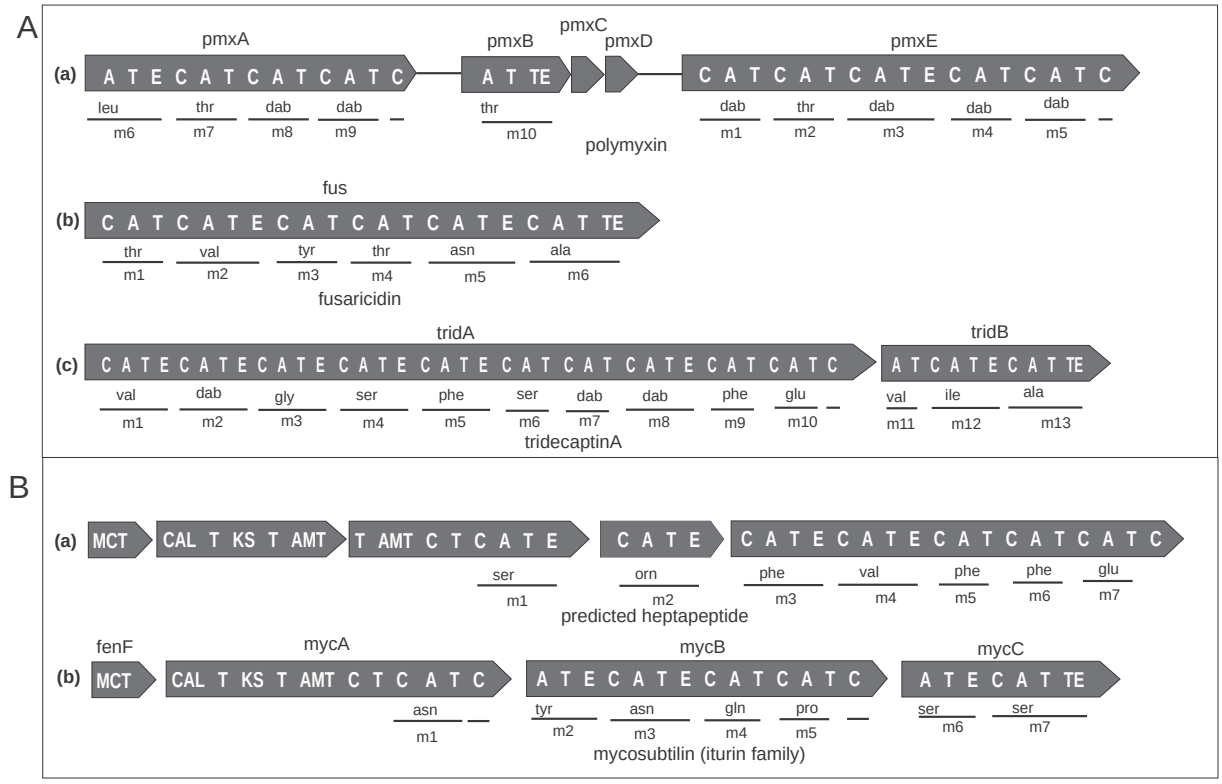


Fig. 3. Organization of the non-ribosomal peptide synthetases (NRPS) encoding lipopeptides in *Paenibacillus* and *Bacillus*. Iterative domains: A, adenylation; T, thiolation; E, epimerization; MCT, malonyl-CoA transacylase; ACL, acyl-coA ligase; AMT, aminotransferase; dab, 2,4-diaminobutyric acid; orn, ornithine; KS, keto synthetase; TE, thioesterase. Further details of domains are described in Table 1. Modules and recruited amino acids indicated below, gene names indicated above each illustration. (A) Organization of the known NRPS (a) polymyxin A in *P. polymyxa* E681, (b) fusaricidin in *P. polymyxa* E681 and (c) tridecaptin A in *P. terrae* NRRL B-30644. (B) Organization of the predicted novel NRPS encoding (a) a heptapeptide in *P. polymyxa* E681; modular architecture is similar to the known Iturin but predicted amino acid composition is completely different and (b) organization of the known mycosubtilin operon [69], an iturin member from *B. subtilis* for comparison.

1.1.4. Paenimacrolidin

Paenimacrolidin is a highly unstable macrocyclic lactone isolated from *Paenibacillus* sp. F6-B70 and comprises a 22 membered lactone ring with a triene in the carbon backbone [33] (Fig. 2B). Three out of four partial genes of the paenimacrolidin synthase showed high similarity to difficidin synthase of *B. amyloliquefaciens* and the structure of paenimacrolidin has similarities with difficidin, implying potential similarities in the biosynthesis and underlying genetic structures (Fig. 2A). Paenimacrolidin also exhibits antimicrobial activity against *Staphylococcus* with potential in clinical applications [3].

1.1.5. Paenilamicin

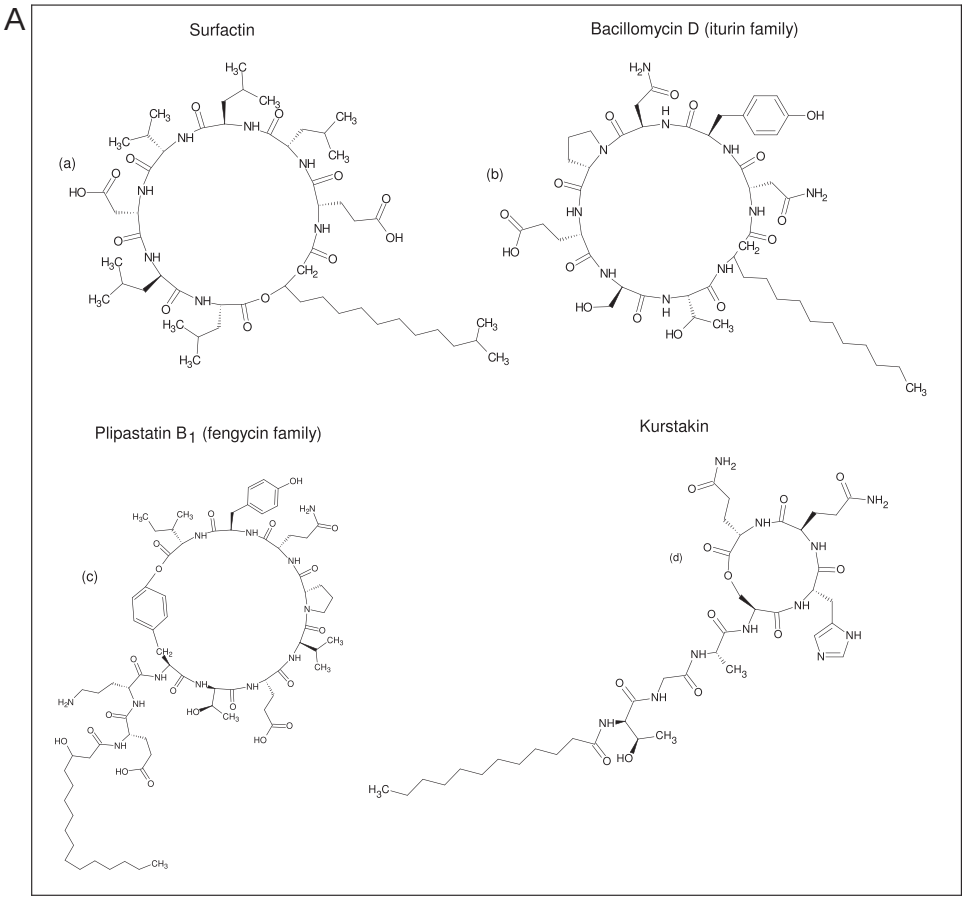
Paenilamicins with antibacterial and antifungal activity have been isolated from *P. larvae* DSM25430, a honey bee pathogen [44]. Despite their activities these compounds do not seem to be involved in host killing, but rather in niche competition [34]. Based on gene activation studies the biosynthesis of paenilamicins has been assigned to the *pam* gene cluster (a complex NRPS/PKS hybrid gene cluster), and the structure (Fig. 2B) was elucidated using HPLC–ESI-MS, GC–MS, and NMR spectroscopy [34]. Different variants of paenilamicins are found due to variation in the first (lysine or arginine) and fourth (lysine or ornithine) recruited amino acid, but synthesis is performed by the very same enzyme complex encoded by *pam*. The non-ribosomal peptide synthetases 2, 3, 5, 6 and 7 encode alanine, N-methyl-diaminopropionic acid (mDap), serine, mDap and glycine, respectivly. Both PKS 1 and 2 mediate the formation of 2,3,5-trihydroxy pentanoic acid, which is then condensed to alanine. Finally, termination is achieved by nucleophilic cleavage by spermidine without involving thioesterase [34].

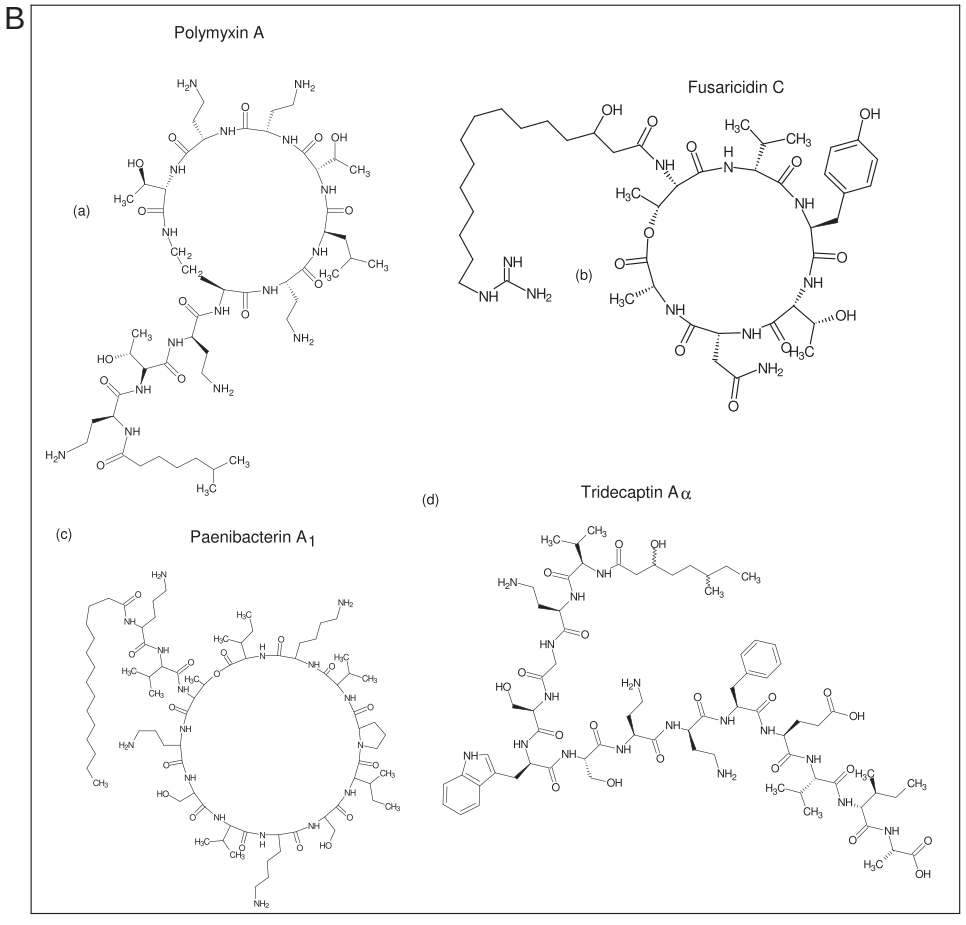
1.2. Bacillus and Paenibacillus lipopeptides

Lipopeptides from *Bacillus* and *Paenibacillus* have been described in a number of recent reviews [4,6,7,32,45,46]. These LPs are synthesized by non-ribosomal peptide synthetases (NRPS) [47]. NRPS comprise organized modules, each module containing catalytic domains: the adenylation (A) domain responsible for selection and monomer activation, the thiolation (T) domain for transfer of the adenylated monomer to a NRPS bound PPT, the condensation domain (C) for peptide bond formation and the thioesterase (TE) domain for release of the peptide monomer from NRPS. Also modification domains such as epimerization (E) domain catalyzing the isomerization of L- into D-amino acid monomers and methyl transferase (MT) are found. The starter condensation domain within the first module catalyzes the attachment of a fatty acid chain to the amino acid activated by the first adenylation domain [47] (Fig. 3). The gene clusters of the *Bacillus* LPs encoding the surfactin, fengycin, iturin and kurstakin families have been described and summarized in detail in a number of reviews [4,45,46].

Structurally, LPs consist of short oligopeptides (6–13 AA) with attached linear or branched fatty acids. For *Bacillus* and *Paenibacillus* linear and cyclic structures have been described (Fig. 4 shows examples of the variation) [7]. A large fraction of the *Paenibacillus* LPs are cyclic cationic LPs which contain the non-proteogenic amino acid 2,4-diaminobutyric

Fig. 4. Chemical structures of lipopeptides from *Bacillus* and *Paenibacillus*. (A) Lipopeptides from *B. amyloliquefaciens* FZB42 (a,b,c): (a) surfactin, (b) bacillomycin (an iturin member), (c) plipastatin (a fengycin member) and (d) kurstakin from *B. thuringiensis kurstaki* HD-1. (B) Lipopeptides from *Paenibacillus*: (a) polymyxin A from *P. polymyxa* E681, (b) fusaricidin C from *P. polymyxa* E681, (c) paenibacterin from *Paenibacillus* sp. OSY-SE (d) tridecaptin from *P. terrae* NRRL B-30644.


A
Surfactin
(a)
Bacillomycin D (iturin family)
(b)
Plipastatin B1 (fengycin family)
(c)
Kurstakin
(d)


B
Polymyxin A
(a)
Fusaricidin C
(b)
(c)
Paenibacterin A1
(d)
Tridecaptin Aα

acid (dab) contributing to the overall positive charge of the cationic lipopeptides. The polymyxins, octapeptins and polypeptins belong to this group enriched in dab (for review see [7]). The cationic lipopeptides have been reported as strong antibacterial agents against gram-negative bacteria and their mode of action is through permeabilization and disruption of the cell membrane [48,49]. Besides their clinical use as bactericidal agents, they have been shown to be active against plant pathogenic *Erwinia amylovora* and *Pectobacterium carotovorum*. [50]. The gene cluster responsible for synthesizing polymyxin synthetase has been described in plant growth promoting rhizobacteria such as *P. polymyxa* E681. The cluster encompasses five genes, of which *pmxA, pmxB* and *pmxE* encode the polymyxin synthetase, whereas *pmxD* and *pmxC* are involved in polymyxin transport [51] (Fig. 3A). Based on the amino acid substitutions at the positions 3, 6, 7 and 10, polymyxins are known to have variants (Fig. 4B). Octapeptins contain eight monomers and appear to be truncated polymyxins with cyclic heptapeptide structures in common. Like polymyxins they exhibit antibacterial activity against both gram-positive and gram-negative bacteria by acting on the membranes and are found in *Paenibacillus* spp. [52].

Polypeptins and pelgipeptins are cyclic nonapeptides isolated from *P. ehimensis* B7 and *P. elgii* B69, respectively. They are active against gram-positive and gram-negative bacteria, but also show antifungal activity against *Fusarium graminearum* and *Rhizoctonia solani* [53,54]. The gene cluster encoding pelgipeptin has been recently characterized in *P.elgii* B69 [55]. Other cyclic cationic LPs include gavaserin and paenibacterins. Gavaserin is isolated from *P. polymyxa* and speculated to contain a cyclic octapeptide structure [56]. Nevertheless, no structural data are available. Paenibacterins are known from *Paenibacillus* sp. OSY-SE and contain a tridecapeptide backbone (Fig. 4B). As the other cationic polypeptides they are active against gram-positive and gram-negative bacteria [57].

Cyclic noncationic lipopeptides from *Paenibacillus* comprise fusaricidins containing cyclic hexapeptide structure (Fig. 4B). They have been reported to exhibit strong antagonistic activity against *Fusarium oxysporum* [58]and induction of systemic resistance in red pepper plants against *Phytophthora* [59]. In addition, also a group of linear cationic LPs with different numbers of amino acids produced by *Paenibacillus* has been described. They include tridecaptins, with strong antimicrobial activity against gram-negative bacteria [60] (Fig. 4B). The gene cluster coding for tridecaptinA_α has been recently characterized from *P. terrae* NRRL B-30644 [61] (Fig. 3A). Cerexins are linear decapeptides, isolated from *B. cereus*, which display strong antimicrobial activity against gram-positive bacteria [62].

Most prominently, *B. amyloliquefaciens* and *B. subtilis* encompass gene clusters coding for cyclic LPs including surfactin, iturin, fengycin and kurstakin (46,63) (Fig. 4A). Several variants that differ in few amino acids have been reported within each family except for kurstakin. The LPs contain regularly variation in the fatty acid chain length and have linear, iso or aniso structural variations.

All surfactins contain cyclic heptapeptide structure, but differ in amino acid composition [64]. Known variants such as pumilacidin, lychenisin and surfactin represent this group and are remarkably confined to specific taxonomic groups [4]. Surfactins are vital for biofim formation and root colonization, but also exhibit a wide range of hemolytic, antimicrobial and antiviral activities, while fungicide activity has not been reported [65–68]. Surfactins are amphiphilic compounds, whose mode of activity seems mainly through membrane permeabilization and disruption [66].

All members of the iturin family have a cyclic heptapeptide structure, but differ from surfactins with distinct amino acid composition and cyclic closure of the lipopeptide structure by a beta-amino group of the fatty acid. Variants named bacillomycins, mycosubtilins, iturins and marihysins are noted [4,7,46]. They are mainly known for strong antifungal activity against several fungi [69–71]. Unlike surfactins their antibacterial activity is limited [72].

Fengycins and plipastatins are decapeptides which form a lactone ring structure between the C-terminus and a tyrosine at position three. They show remarkable antagonistic activity against filamentous fungi. The three LPs surfactin, iturin and fengycin may also act synergistically, enhancing their activities [73,74].

Kurstakins are another family of LPs isolated from *B. thuringiensis* strains and have been identified as phylogenetic markers for the species [75]. Kurstakins contain a lactone bond between Ser4 and the C-terminus of Gln7 and consequently form a cyclic tetrapeptide with a tetrapeptide side chain. They exhibit limited antifungal activity [63,75].

1.3. Genome mining tools for novel NRPS and PKS prediction

In order to discover novel secondary metabolites, several bioinformatics tools are available to perform genome mining. Some of the web based tools such as antiSMASH [20,21], NP.searcher [76] and NaPDoS [22] use hidden Markov models to identify NRPS and PKS in bacterial genomes. A more detailed prediction of the clusters is also possible through antiSMASH, which allows BLAST search on the predicted cluster to identify closest homologue in the database. antiSMASH allows the analysis of fragmented genomes and metagenomes making it a powerful prediction tool. Predicted peptides can be queried on NORINE database [77] containing more than 1000 non-ribosomal peptides to find similar structures [78]. Another useful prediction tool is the NRPS/PKS substrate predictor [23], which mainly focuses on the specificity of A domains (from NRPS) and AT domains (from PKS), which is useful to narrow the ambiguity of A domains specificity that occur in other prediction tools.

1.4. Prediction of lipopeptides and polyketides in published genome sets

In the following we evaluate the potential of type I PKs and LPs production based on genome mining and analysis, and show a clear potential for the discovery of several undiscovered variants and different structures. The next generation sequencing revolution of the last years have resulted and will result in a fast growing number of sequenced bacterial genomes and metagenomes. To evaluate the potential chemical space encoded in these genomes, the genome mining tools described above can facilitate the prediction of secondary metabolites, especially type I PKs and LPs. The cumbersome task, especially of various unstable PKs, to isolate and elucidate structures by NMR methods requiring milligram amounts can be pipelined by predicting the potential of novelty, also assisted by developments in mass spectrometry [79]. A limitation in prediction of PKs is that the colinearity rule common for LPs does not always apply. However, based on the predicted modular architecture and the number of core domains, it is still possible to predict the types of PKs and their variants as we show for Bacillales in the following (see Table 1 and Supplemental Table for an overview). A total of 160 published genomes the Bacillales were analyzed, of which 91 genomes contained metabolic clusters encoding LPs, type I PKs or both (57%). Intriguingly, a clear higher percentage, 85% of the 40 isolates, from rhizosphere and endophytes contained at least one of these metabolic clusters (Supplemental Table). However, the origin of almost a third of the isolates is unclear, making it difficult to foresee, if the higher incidence of these secondary metabolites in plant associated environments will also be seen when more genomes will be sequenced. A trend can be also seen phylogenetically with certain *Bacillus* spp. and *Paenibacillus* spp. as the taxa with the highest numbers of both type I PKs and LPs (Supplemental Fig.). How far also this observation just reflects a higher density of available genomes in these taxa than e.g. in *Salinibacillus* spp. remains to be seen.

Genome mining revealed the potential for known and novel LPs and PKs. Based on the prediction of the general architecture, undescribed, novel clusters can be identified (Supplemental Table, Table 1). Prediction of recruited substrates allows also the prediction of novel variants

Table 1
Predicted lipopeptides and type I polyketides from selected members of Bacillales.

GenBank ID	Organism	Lipopeptide*	Type I polyketide*
CP000154.1	*Paenibacillus polymyxa* E681	Polymyxin A, structure and biosynthetic gene cluster confirmed [SKChoi 2009, Catch JR 1949] L-dab-L-thr-D-dab-L-dab-L-dab-D-leu-L-thr-L-dab-L-dab-L-thr Fusaricidin C, structure and biosynthetic gene cluster confirmed [Soo-Keun Choi 2008] L-thr-D-val-L-tyr-D-thr-D-asn-D-ala Predicted tridecaptin variant D-val-D-dab-D-gly-D-ser-D-phe-L-ser-L-dab-D-dab-L-phe-L-glu-L-val-D-ile-L-val Predicted unknown heptapeptide (mal) + (pk) + D-ser-D-orn-D-phe-D-val-L-phe-L-phe-L-glu 47% identity to bacillomycin of B. *amyloliquefaciens* FZB42	Novel polyketide gly (DH = 5, KS = 12, KR = 9, cMT = 2, ACP = 14); 43% identity to known bacillaene of *B. amyloliquefaciens* FZB42
ARIL00000000.1	*Paenibacillus polymyxa* SQR-21	Polymyxin A variant L-dab-L-thr-D-dab-L-dab-L-dab-D-leu-L-leu-L-dab-L-dab-L-thr Fusaricidin C (peptide sequence is similar to E681); 93%identity to fusaricidin of P. polymyxa E681 Predicted tridecaptin variant, peptide sequence is similar to *P. polymyxa* E681 Predicted unknown heptapeptide, peptide sequence similar to *P. polymyxa* E681 Predicted decapeptide (maybe a truncated tridecaptin) D-gly-D-dab-D-gly-D-ser-D-phe-L-ser-L-dab-D-dab-L-ile-L-glu	Novel polyketide (same as above) - modular architecture is similar to *P. polymyxa* E681. 43% identity to bacillaene of B. *amyloliquefaciens* FZB42
ARIL00000000.1	*Paenibacillus massiliensis DSM 16942*	Novel fusaricidin variant L-thr-D-val-L-ile-D-ser-D-asn-L-ala; 49% identity to fusaricidin of *P.polymyxa* E681.	No clusters found
CP006941.1	*Paenibacillus polymyxa* CR1	Predicted heptapeptide variant (pk-nrp) + (thr-ser-ala) + (phe-gln-glu) 48% identity to bacillomycin of *B. amyloliquefaciens* FZB42	Incomplete PKS predicted
CP003235.1	*Paenibacillus mucilaginosus* 3016	Predicted heptapeptide variant phe + (orn-val-ile-phe-nrp-phe) 44% identity to bacillomycin of B. *amyloliquefaciens* FZB42	Incomplete PKS predicted
CP009288.1	*Paenibacillus durus* DSM 1735	Incomplete NRPS predicted	Paenimacrolidine (KS = 9, DH = 6, cMT = 2, KR = 6, ER = 1, ACP = 14) 40% identity to known difficidin of *B. amyloliquefaciens* FZB42
BAVZ00000000.1	*Paenibacillus pini* JCM 16418	Incomplete NRPS predicted	Bacillaene variant, gly; ala (KS = 14, DH = 8; KR = 8, cMT = 2, ACP = 16); also the order of domains differ; share 56% identity to bacillaene of *B. amyloliquifaciens* FZB42
ANAT00000000.1	*Paenibacillus lentimorbus* NRRL B-30488	Bacillomycin D, surfactin, plipastatin; similar to *B. amyloliquefaciens* FZB42	Bacillaene, macrolactin, difficidin; similar to *B. amyloliquefaceins* FZB42
AULE00000000.1	*Paenibacillus taiwanensis* DSM 18679	Paenibacterin variant (orn-val-thr-orn) + (tyr-orn-ser-ile-pro) + (pro) + (ile-ile); 69% identity with known paenibacterin of *Paenibacillus* sp. OSY-SE	Incomplete PKS predicted
ARMT00000000.1	*Paenibacillus fonticola* DSM 21315	Unknown heptapeptide-architecture similar to Iturin family (mal) + (pk-gly) + (orn-glu) + (lys-tyr) + (ile-val); 36% identity with known Bacillomycin of *B. amyloliquefaciens* FZB42	Incomplete PKS predicted
CP003355.1	*Paenibacillus larvae* DSM 25430	IturinA	Paenilamicins: A1, B1, A2, B2, -a complex NRPS/PKS hybrid lys/arg, ala, mdap, lys/orn, ser, mdap, gly (KS = 4, KR = 4, nMT = 2, ACP = 4)
CP003763.1	*Bacillus thuringiensis* HD-789	Kurstakin, structure confirmed [Hathout et al. 2000] D-thr-L-gly-L-ala-L-ser-L-his-D-gln-L-gln	No clusters found
CP004069.1	*Bacillus thuringiensis serovar kurstaki* HD73	Kurstakin variant D-thr-L-ser-L-ala-L-ser-L-leu-D-nrp-L-gln 99% identity to known kurstakin of *Bacillus thuringiensis serovar kurstaki* HD-1	No clusters found
CP000560.1	*Bacillus amyloliquefaciens* FZB42	SurfactinA [Peypoux F 1994, Koumoutsi A 2004] L-glu-L-leu-D-leu-L-val-L-asp-D-leu-L-leu Plipastatin B [Nishikiori 1986, Koumoutsi A 2004] L-glu-D-orn-L-tyr-D-thr-L-glu-D-val-L-pro-L-gln-D-tyr-L-ile Bacillomycin D [Peypoux F 1984, Koumoutsi A 2004] L-asn-D-tyr-D-asn-L-pro-L-glu-D-ser-L-thr	Bacillaene gly; ala (KS = 14, DH = 8, KR = 9, cMT = 2, ACP = 14) Difficidin (KS = 14; DH = 9, KR = 10, cMT = 3, ER = 1, ACP = 19) Macrolactin (KS = 11, DH = 5, KR = 11, ACP = 15) [Stein, 2005; Chen et al., 2006]
JOKF00000000.1	*Bacillus amyloliquefaciens plantarum* W2	SurfactinA-similar to FZB42, Plipastatin B (similar to FZB42 but Glu instead of Gln)	Macrolactin variant (KS = 11, DH = 3, KR = 11, ACP = 15); 97% identity with known macrolactin of *B. amyloliquefaciens* FZB42 Difficidin variant (KS = 14, DH = 9, KR = 10, CMT = 3, ER = 0, ACP = 19); 98% identity with know difficidin of

(continued on next page)

Table 1 (*continued*)

GenBank ID	Organism	Lipopeptide*	Type I polyketide*
			B. amyloliquefaciens FZB42 Bacillaene-similar to FZB42; 98% identity to bacillaene of *B. amyloliquefaciens* FZB42
NC_014639.1	*Bacillus atrophaeus* 1942	SurfactinC L-glu-L-leu-D-leu-L-val-L-asp-D-leu-L-ile; 78% identity to *B. amyloliquefaciens* FZB42 Plipastatin B; mycosubtilin; similar to FZB42	Bacillaene variant, similar to FZB42 in terms of specificity of A domains but (KS = 16, DH = 7, KR = 9, cMT = 2, ACP = 16); 64% identity to *B. amyloliquefaciens* FZB42
CM000488.1	*Bacillus subtilis NCIB 3610*	SurfactinA; plipastatin B; similar to FZB42	bacillaene similar to FZB42 A domains specificity gly, nrp (KS = 15, DH = 8, KR = 9, cMT = 2, ACP = 17), 64% identity to known bacillaene *of B. amyloliquefaciens* FZB42
AP008955.1	*Brevibacillus brevis* NBRC 100599	Incomplete NRPS predicted	Novel polyketide (KS = 14,cMT = 3, oMT = 1, nMT = 1, KR = 8, ACP = 20), A domain specificity ala, ser; 38% identity to difficidin of *B. amyloliquefaciens FZB42*
AEWH00000000.1	*Ornithinibacillus scapharcae* TW25	Incomplete NRPS predicted	Macrolactin like polyketide 44% identity *to B. amyloliquefaciens FZB42* (KS = 13, DH = 4, KR = 8, ACP = 16) Bacillaene, similar to *B. amyloliquefaciens* FZB42
APIS00000000.1	*Salinibacillus aidingensis* MSP4	Surfactin, plipastatin B; similar to *B. amyloliquefaciens* FZB42	Macrolactin like polyketide (KS = 12, DH = 5, KR = 6, ACP = 14) 45% idenity to bacillaene of *B. amyloliquefaciens* FZB42

* Sequence prediction using antiSMASH, NaPDos and NRPS/PKS substrate predictor tools, peptides in bold are predicted novel peptides, monomers in both bold and underline differ from described metabolites in that position (in case of polyketides they differ in number and maybe in the order of domains); monomers in underline are known variants, previously described. *B. subtilis* 3610 and *B. amyloliquefaciens* FZB42 are reported to produce similar bacillaene [Rebecca A. Butcher 2006, Chen 2009]. However, they differ in number of domains predicted. Abbreviations: mal, malonyl-CoA; pk, polyketide; dab, 2,4-diaminobutyric acid; KS, ketosynthase; DH, dehydratase; MT, methyl transferase; KR, ketoreductase; orn, ornithine, nrp, unassigned non ribosomal peptide, mdap, N-methyl-diaminopropionic acid, NRPS, non-ribosomal peptide synthetase, PKS, type 1 polyketide synthase.

with same cluster architecture. Of course, even the same architecture and substrate prediction cannot exclude additional secondary modifications. These clusters were not considered as "novel" in the current analysis, but indicated as similar to described clusters in Table 1 and in the Supplemental Table. Especially in several *Paenibacillus* strains, we found a high potential for novel undescribed PKs and LPs variants of heptapeptides, nonapeptides, tridecaptins and decapeptides (truncated tridecaptins). Besides this, many *Paenibacillus* strains encompass known LPs such as polymyxins and fusaricidins and variants that differ in monomer composition (Table 1). We found also a novel fusaricidin variant in *P. massiliensis* DSM 16942 differing at the 4th position substituted by serine, which is believed to be highly specific for allo-threonine.

Predicted heptapeptides from *Paenibacillus* strains have a modular architecture similar to iturin (Fig. 3B). Monomers of the peptide backbone in these heptapeptides are however completely different from the known iturin members. The genes in the heptapeptide operon of *P. polymyxa* E681 show up to 46% identity to bacillomycin D, an iturin member of *B. amyloliquefaciens* FZB42. Therefore, we hypothesize that these may belong to a novel class of iturins. Also, such heptapeptide variants with different peptide composition were found in other *Paenibacillus* strains such as *P. polymyxa* CR1, SC2, and *Paenibacillus* sp. HGH0039, *P. mucilaginosus* 3016 and *P. fonticola* DSM 21315. Moreover, we found an undescribed nonapeptide and its variants in *P. mucilaginosus* 3016, *P. elgii* B69 and *P. terrae* HPL-003. We discovered tridecaptin variants in *P. polymyxa* strains including E681, SQR21 and ATCC 842 (Table 1). In addition, we predicted decapeptides containing ten monomers, but with similar composition to tridecaptins. These seem to be truncated tridecaptins and therefore undescribed potential LPs of the *P. polymyxa* strains SQR21, M1 and SC2. We also identified a novel paenibacterin variant in *P. taiwanensis* DSM 18679 and *P. alvei* DSM 29 with four different amino acids to described metabolites of *Paenibacillus* sp. OSY-SE (Fig. 3B).

The majority of the *Bacillus* species that harbor lipopeptide gene clusters from the three families comprising surfactin, iturin and fengycin are *B. amyloliquefaciens*, *B. atrophaeus* and *B. subtilis*. Moreover, LPs (surfactins and fengycins) are predicted for *B. licheniformis*, *B. mojavensis* and *B. pumilus* with known metabolic potential but also for strains so far not characterized for their potential and less well investigated species such as *Salinibacillus aidingensis* (Table 1, Supplemental Table). The fourth family kurstakin is confined to *B. thuringiensis* strains. A kurstakin variant is found in *B. thuringiensis serovar kurstaki* HD73 with altered amino acid composition in position 2 and 5. The D and L forms of the monomers in a lipopeptide can also be predicted depending on presence and absence of the epimerization domains [80]. For instance, many *B. subtilis* encode plipastatin B, a member of fengycin family. Although plipastatin B and fengycin B are fengycin members and share identical monomers in the backbone, they differ in L-Tyr and D-Tyr, respectively, as also the chirality in monomers can be predicted with prediction tools. Altogether, it can be noted that the so far collected genome information confirms well known LPs for a number of *Bacillus* and *Paenibacillus* strains, but also shows a clear potential to produce a number of novel lipopeptides, especially in the genus *Paenibacillus*. A large number of strains from other genera of the Bacillales seem to lack the potential to produce LPs and PKs type 1 (Supplemental Table). However, it cannot be excluded that draft genomes may hinder the prediction of LPs and PKs (discussed below) if larger gaps within the published genomes exist.

For the defined structure of the polyketide paenimacrolidin from *Paenibacillus* sp. F6-B70, the biosynthetic gene cluster is not characterized. Based on partial 16S rRNA gene analysis of *Paenibacillus* sp. F6-B70 it has been shown to be closely related to *P. elgii* and *P. ehimensis* [33]. We predicted a novel polyketide gene cluster that is similar in *P. durus* DSM1735, *P. elgii* and *P. ehimensis* (Fig. 1B). The partial paenimacrolidin synthase genes from *Paenibacillus* sp. F6-B70, have high similarity

with part of *P. durus* genome. Furthermore, by examining the structure of paenimacrolidin using prediction tools, we speculate that a gene cluster with similarity to the difficidin cluster of *B. amyloliquefaciens* FZB42 may be responsible for the production of paenimacrolidin or a related PKS in these species (Table 1).

A number of very likely novel PKs with gene cluster architecture similar to bacillaene (Fig. 1B) are found in the *P. polymyxa* strains E681, SQR21, in *P. pini* JCM 16418 and in *Brevibacillus brevis* NBRC 100599 (Table 1). Intriguingly, in *P. polymyxa* strains, only one adenylation domain specifying glycine was found, instead of glycine and alanine as described in the bacillaene producer *B. amyloliquefaciens* (Table 1). PKS modules from *P. polymyxa* E681 shared up to 43% nucleotide sequence identity with *baeN* of *B. amyloliquefaciens*. Also for this polyketide, we identified variants that differ in number of the catalytic domains KS, DH, cMT and KR. In other *P. polymyxa* strains such as ATCC 842, M1 and SC2 a similar PKS cluster can be found with one DH domain less (Supplemental Table). In *P. pini*, the first adenylation domain specifies glycine like in bacillaene, while the second adenylation domain specifies serine instead of alanine. In *B. brevis*, the first adenylation domain specifies alanine and the second adenylation domain specifies serine. Besides it contains special methylation domains such as oMT and nMT that are not found in other polyketide clusters, clearly pointing to an uncharacterized PKs encoded in this genome (Fig. 1B).

Regarding the PKs anticipated from *Bacillus*, several strains contained well described clusters for bacillaene, macrolactin and difficidin synthesis. Surprisingly, we also found variants of those, which have not been anticipated to date, even in strains of *B. amyloliquefaciens* and *B. subtilis* (Table 1 and Supplemental Table). However, prediction has to be careful here as it has been shown that small variation in the domain structure does not result in the production of different bacillaenes [31,36]. Generally, and not surprisingly *B. amyloliquefaciens* and *B. subtilis* are noted as prolific producers of PKs. Other *Bacillus* spp. encompassing PKS are *B. atrophaeus, B. mojavensis* and *Brevibacillus brevis* with clearly different PKs potential. In more detail, macrolactin variants are found in *B. amyloliquefaciens* strains such as IT-45, DC-12, UASWS BA1 and B1895 and B. *amyloliquefaciens plantarum* such as UCMB 5036, W2 and AH159-1. Bacillaene variants are found in *B. atrophaeus, B. subtilis* strains and *B. mojavensis* RRC 101. In *B. atrophaeus* and *B. mojavensis* RRC 101 variants have similar amino acids like in *B. amyloliquefaciens* FZB42 but differ in number of catalytic domains. In *B. subtilis* strains, we found variation to bacillaene as the second adenylation domain specifies glutamine, but the number of catalytic domains is identical to *B. amyloliquefaciens* FZB42. It has also to be stated that not all metabolite clusters of these species are expressed or even be functional as seen in *B. subtilis* 168 [81]. This lab strain obviously does not require its secondary metabolites anymore, very likely unlike its relatives living in plant association in nature.

We also performed genome mining on Bacillales genera growing in other environments. Intriguingly, the majority of these non-plant associated bacteria do not harbor LPS and PKS. On the contrary, a large fraction of the plant-associated bacteria contained LPS and PKS (Supplemental Table, Supplemental Fig.) with both *Bacillus* and *Paenibacillus* dominating the distribution. However, bacteria such as *Ornithinibacillus* and *Salinibacillus* occuring in soil environments seem also to have the capacity to produce macrolactin-like polyketides with higher dissimilarity to the macrolactin of *B. amyloliquefaciens* FZB42.

1.5. Conclusions and future perspectives

Bacillus and some related genera can be phylogenetically separated into ten distinct groups based on 16S rRNA gene sequence information [82,83]. It is intriguing that the LPS and PKS gene clusters seem to be constrained to particular species or groups (Supplemental Fig.), potentially indicating the ecological role for such gene clusters.

BLAST results can be often misleading in the prediction of metabolic capacity as part of the target gene cluster can share similarity within and between different gene clusters. Therefore, it is crucial to examine the whole architecture of a particular gene cluster to obtain precise results. With an increasing availability of genome information due to advanced and better affordable next generation sequencing, we anticipate that there will be enormous increase in the deposition of sequences in public databases derived from uncultured and less studied bacteria. Such sequence wealth can be a rich source for novel secondary metabolite production and can be explored to find novel gene clusters encoding secondary metabolites. Our results suggest that a substantial fraction of predicted LPs and PKs from the metabolomes of Bacillales are uncharacterized and their functions with regards to plant association still remains to be established and other so far neglected Bacillales with no published genomic data still remain unexplored.

2. Materials and methods

2.1. Genome sequences

NCBI accession numbers for the whole genome sequences of both characterized and uncharacterized group of isolates from selected members of the Bacillales were extracted. (Table 1, Supplemental Table). Contigs of draft genomes were extracted and saved as a fasta file.

2.2. Secondary metabolite gene cluster prediction and analysis tools

Three web based tools, antiSMASH, NaPDos, and NRPS/PKS substrate predictor tools were used for secondary metabolite gene cluster prediction and analysis. The architecture of the gene clusters were predicted using the antiSMASH program [20,21]. The catalytic domains of the predicted gene cluster are deduced using NaPDoS [22]. To analyze adenylation domains of NRPS and AT domains of PKS, NRPS/PKS substrate predictor [23] was used.

Firstly, Genbank accession numbers were given as input for antiSMASH. For draft genomes, the extracted files were uploaded to antiSMASH. The predicted secondary metabolite gene clusters from antiSMASH consisted of NRPS, PKS, hybrid PKS/NRPS, siderophore, bacteriocin and lantibiotics. The clusters responsible for biosynthesis of LPs and PKs were analyzed. Further predicted monomers were confirmed using NaPDos and NRPS/PKS substrate predictor. For accuracy, predictions from the three tools were analyzed. Regarding polyketides, the number of core catalytic domains KS, DH, KR, ACP and ER were noted. Finally, both lipopeptide and polyketide encoding gene clusters were subjected to BLAST to find the closest homologue available in the database.

2.3. Phylogenetic analysis of predicted LPs and type I PKs

The 16S rRNA gene sequences were downloaded from RDP [84]. These sequences were clustered at 97% identity using clustalW, and a tree was plotted using neighbor joining algorithm within MEGA6 [85]. The phylogenetic distribution of predicted LPs and PKs from genome mining is combined with the tree and visualized in iTOL2 [86].

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.csbj.2015.03.003.

Acknowledgements

This work was supported by the life science project LS11-014 of the NFB, Niederösterreichische Forschung & Bildung.

References

[1] Finking R, Marahiel MA. Biosynthesis of nonribosomal peptides1. Annu Rev Microbiol 2004;58:453–88.

[2] Walsh CT. The chemical versatility of natural-product assembly lines. Acc Chem Res 2008;41(1):4–10.
[3] Weissman KJ, Leadlay PF. Combinatorial biosynthesis of reduced polyketides. Nat Rev Microbiol 2005;3(12):925–36.
[4] Ongena M, Jacques P. *Bacillus* lipopeptides: versatile weapons for plant disease biocontrol. Trends Microbiol 2008;16(3):115–25.
[5] Ongena M, Jourdan E, Adam A, Paquot M, Brans A, et al. Surfactin and fengycin lipopeptides of *Bacillus subtilis* as elicitors of induced systemic resistance in plants. Environ Microbiol 2007;9(4):1084–90.
[6] Raaijmakers JM, De Bruijn I, Nybroe O, Ongena M. Natural functions of lipopeptides from *Bacillus* and *Pseudomonas*: more than surfactants and antibiotics. FEMS Microbiol Rev 2010;34(6):1037–62.
[7] Cochrane SA, Vederas JC. Lipopeptides from *Bacillus* and *Paenibacillus* spp.: a gold mine of antibiotic candidates. Med Res Rev 2014. http://dx.doi.org/10.1002/med.21321.
[8] Cane DE, Walsh CT, Khosla C. Harnessing the biosynthetic code: Combinations, permutations, and mutations. Science 1998;282(5386):63–8.
[9] Donadio S, Monciardini P, Sosio M. Polyketide synthases and nonribosomal peptide synthetases: the emerging view from bacterial genomics. Nat Prod Rep 2007;24:1073–109.
[10] Kopp F, Marahiel MA. Where chemistry meets biology: the chemoenzymatic synthesis of nonribosomal peptides and polyketides. Curr Opin Biotechnol 2007;18(6):513–20.
[11] Fischbach MA, Walsh CT. Assembly-line enzymology for polyketide and nonribosomal peptide antibiotics: logic, machinery, and mechanisms. Chem Rev 2006;106(8):3468–96.
[12] Crosa JH, Walsh CT. Genetics and assembly line enzymology of siderophore biosynthesis in bacteria. Microbiol Mol Biol Rev 2002;66(2):223–49.
[13] Staunton J, Weissman KJ. Polyketide biosynthesis: a millenium review. Nat Prod Rep 2001;18(4):380–416.
[14] Rawlings BJ. Type I polyketide biosynthesis in bacteria (Part A-erythromycin biosynthesis). Nat Prod Rep 2001;18:190–227.
[15] Marahiel MA, Stachelhaus T, Mootz HD. Modular peptide synthetases involved in nonribosomal peptide synthesis. Chem Rev 1997;97(7):2651–74.
[16] Bachmann BO, Ravel J. Methods for in silico prediction of microbial polyketide and nonribosomal peptide biosynthetic pathways from DNA sequence data. Methods Enzymol 2009;458:181–217.
[17] Wenzel SC, Müller R. Formation of novel secondary metabolites by bacterial multimodular assembly lines: deviations from textbook biosynthetic logic. Curr Opin Chem Biol 2005;9(5):447–58.
[18] Wang H, Fewer DP, Holm L, Rouhiainen L, Sivonen K. Atlas of nonribosomal peptide and polyketide biosynthetic pathways reveals common occurrence of nonmodular enzymes. Proc Natl Acad Sci U S A 2014;111(25):9259–64.
[19] Cimermancic P, Medema MH, Claesen J, Kurita K, Wieland Brown LC, et al. Insights into secondary metabolism from a global analysis of prokaryotic biosynthetic gene clusters. Cell 2014;158(2):412–21.
[20] Medema MH, Blin K, Cimermancic P, de Jager V, Zakrzewski P, et al. antiSMASH: rapid identification, annotation and analysis of secondary metabolite biosynthesis gene clusters in bacterial and fungal genome sequences. Nucleic Acids Res 2011;39:W339–46.
[21] Blin K, Medema MH, Kazempour D, Fischbach MA, Breitling R, et al. antiSMASH 2.0- a versatile platform for genome mining of secondary metabolite producers. Nucleic Acids Res 2013:W204–12.
[22] Ziemert N, Podell S, Penn K, Badger JH, Allen E, et al. The natural product domain seeker NaPDoS: a phylogeny based bioinformatic tool to classify secondary metabolite gene diversity. PLoS ONE 2012;7(3):e34064.
[23] Khayatt BI, Overmars L, Siezen RJ, Francke C. Classification of the adenylation and acyl-transferase activity of NRPS and PKS systems using ensembles of substrate specific hidden Markov models. PLoS ONE 2013;8(4):e62136.
[24] Cane DE. A special thematic issue on polyketide and nonribosomal polypeptide biosynthesis. Chem Rev 1997;97:2463–706.
[25] Walsh CT. Polyketides and nonribosomal peptide antibiotics: modularity and versatility. Science 2004;303(5665):1805–10.
[26] Chen X-H, Vater J, Piel J, Franke P, Scholz R, et al. Structural and functional characterization of three polyketide synthase gene clusters in *Bacillus amyloliquefaciens* FZB42. J Bacteriol 2006;188(11):4024–36.
[27] Chen XH, Koumoutsi A, Scholz R, Borriss R. More than anticipated – production of antibiotics and other secondary metabolites by *Bacillus amyloliquefaciens* FZB42. J Mol Microbiol Biotechnol 2008;2009(16):14–24.
[28] Shen B. Polyketide biosynthesis beyond the type I, II and III polyketide synthase paradigms. Curr Opin Chem Biol 2003;7(2):285–95.
[29] Cheng YQ, Tang GL, Shen B. Type I polyketide synthase requiring a discrete acyltransferase for polyketide biosynthesis. Proc Natl Acad Sci U S A 2003;100(6):3149–54.
[30] Hertweck C. The biosynthetic logic of polyketide diversity. Angew Chem Int Ed Engl 2009;48(26):4688–716.
[31] Chen XH, Scholz R, Borriss M, Junge H, Moegel G, et al. Difficidin and bacilysin produced by plant-associated *Bacillus amyloliquefaciens* are efficient in controlling fire blight disease. J Biotechnol 2009;140(1–2):38–44.
[32] Stein T. *Bacillus* subtilis antibiotics: structures, syntheses and specific functions. Mol Microbiol 2005;56(4):845–57.
[33] Wu XC, Qian CD, Fang HH, Wen YP, Zhou JY, et al. Paenimacrolidin, a novel macrolide antibiotic from *Paenibacillus* sp. F6-B70 active against methicillin-resistant Staphylococcus aureus. Microb Biotechnol 2011;4(4):491–502.
[34] Müller S, Garcia-Gonzalez E, Mainz A, Hertlein G, Heid NC, et al. Paenilamicin: structure and biosynthesis of a hybrid nonribosomal peptide/polyketide antibiotic from the bee pathogen *Paenibacillus larvae*. Angew Chem Int Ed Engl 2014;53(40):10821–5.
[35] Patel PS, Huang S, Fisher S, Pirnik D, Aklonis C, et al. Bacillaene, a novel inhibitor of procaryotic protein synthesis produced by *Bacillus subtilis*: production, taxonomy, isolation, physico-chemical characterization and biological activity. J Antibiot 1995;48(9):997–1003.
[36] Butcher RA, Schroeder FC, Fischbach MA, Straight PD, Kolter R, et al. The identification of bacillaene, the product of the PksX megacomplex in *Bacillus subtilis*. Proc Natl Acad Sci U S A 2007;104(5):1506–9.
[37] Calderone CT, Kowtoniuk WE, Kelleher NL, Walsh CT, Dorrestein PC. Convergence of isoprene and polyketide biosynthetic machinery:isoprenyl-S-carrier proteins in the pksX pathway of *Bacillus subtilis*. Proc Natl Acad Sci U S A 2006;103(24):8977–82.
[38] Wilson KE, Flor JE, Schwartz RE, Joshua H, Smith JL, et al. Difficidin and oxydifficidin: novel broad spectrum antibacterial antibiotics produced by *Bacillus subtilis*, II Isolation and physicochemical characterization. J Antibiot 1987;40(12):1682–91.
[39] Zimmerman SB, Schwartz CD, Monaghan RL, Pelak BA, Weissberger B, et al. Difficidin and oxydifficidin: novel broad spectrum antibacterial antibiotics produced by *Bacillus subtilis*. 1. production, taxonomy and antibacterial activity. J Antibiot 1987;40(12):1677–81.
[40] Zweerink MM, Edison A. Difficidin and oxydifficidin: novel broad spectrum antibacterial antibiotics produced by *Bacillus subtilis*, III Mode of action of difficidin. J Antibiot 1987;40(12):1692–7.
[41] Zheng CJ, Lee S, Lee CH, Kim WG. Macrolactins O–R, glycosylated 24-membered lactones from *Bacillus* sp. AH159-1. J Nat Prod 2007;70(10):1632–5.
[42] Schneider K, Chen XH, Vater J, Franke P, Nicholson G, et al. Macrolactin is the polyketide biosynthesis product of the pks2 cluster of *Bacillus amyloliquefaciens* FZB42. J Nat Prod 2007;70(9):1417–23.
[43] Gustafson K, Roman M, Fenical W. The macrolactins, a novel class of antiviral and cytotoxic macrolides from a deep-sea marine bacterium. J Am Chem Soc 1989;111(19):7519–24.
[44] Garcia-Gonzalez E, Müller S, Hertlein G, Heid N, Süssmuth RD, et al. Biological effects of paenilamicin, a secondary metabolite antibiotic produced by the honey bee pathogenic bacterium *Paenibacillus larvae*. Microbiologyopen 2014;3(5):642–56.
[45] Roongsawang N, Washio K, Morikawa M. Diversity of nonribosomal peptide synthetases involved in the biosynthesis of lipopeptide biosurfactants. Int J Mol Sci 2010;12(1):141–72.
[46] Jacques P. Surfactin and Other Lipopeptides from *Bacillus* spp. In: Steinbüchel A, Soberón-Chávez G, editors. Biosurfactants, 20. Microbiology Monographs; 2011. p. 57–91.
[47] Finking R, Marahiel MA. Biosynthesis of nonribosomal peptides 1. Annu Rev Microbiol 2004;58:453–88.
[48] Velkov T, Thompson PE, Nation RL, Li J. Structure-activity relationships of polymyxin antibiotics. J Med Chem 2010;53(5):1898–916.
[49] Soon RL, Velkov T, Chiu F, Thompson PE, Kancharla R, et al. Design, synthesis, and evaluation of a new fluorescent probe for measuring polymyxin lipopolysaccharide binding interactions. Anal Biochem 2011;409(2):273–83.
[50] Niu B, Vater J, Rueckert C, Blom J, Lehmann M, Ru JJ, et al. Polymyxin P is the active principle in suppressing phytopathogenic Erwinia spp. by the biocontrol rhizobacterium *Paenibacillus* polymyxa M-1. BMC Microbiol 2013;13:137.
[51] Choi SK, Park SY, Kim R, Kim SB, Lee CH, et al. Identification of a polymyxin synthetase gene cluster of *Paenibacillus polymyxa* and heterologous expression of the gene in *Bacillus subtilis*. J Bacteriol 2009;191(10):3350–8.
[52] Qian CD, Wu X-C, Teng Y, Zhao W-P, Li O, et al. Battacin (octapeptin B5), a new cyclic lipopeptide antibiotic from *Paenibacillus tianmuensis* active against multidrugresistant gram-negative bacteria. Antimicrob Agents Chemother 2012;56(3):1458–65.
[53] Sogn JA. Structure of the peptide antibiotic polypeptin. J Med Chem 1976;19(10):1228–31.
[54] Ding R, Wu XC, Qian CD, Teng Y, Li O, et al. Isolation and identification of lipopeptide antibiotics from *Paenibacillus elgii* B69 with inhibitory activity against methicillinresistant Staphylococcus aureus. J Microbiol 2011;49(6):942–9.
[55] Qian CD, Liu TZ, Zhou SL, Ding R, ZhaoWP Li O, Wu XC. Identification and functional analysis of gene cluster involvement in biosynthesis of the cyclic lipopeptide antibiotic pelgipeptin produced by *Paenibacillus elgii*. BMC Microbiol 2012;12:197–203.
[56] Pichard B, Larue JP, Thouvenot D. Gavaserin and saltavalin, new peptide antibiotics produced by *Bacillus polymyxa*. FEMS Microbiol Lett 1995;133(3):215–8.
[57] Guo Y, Huang E, Yuan C, Zhang L, Yousef AE. Isolation of a *Paenibacillus* sp. strain and structural elucidation of its broad-spectrum lipopeptide antibiotic. Appl Environ Microbiol 2012;78(9):3156–65.
[58] Raza W, Yang X, Wu H, Wang Y, Xu Y, et al. Isolation and characterisation of fusaricidin-type compound-producing strain of *Paenibacillus polymyxa* SQE-21 active against Fusarium oxysporum f.sp. nevium. Eur J Plant Pathol 2009;125(3):471–83.
[59] Lee SH, ChoYE Park S-H, Balaraju K, Park JW, et al. An antibiotic fusaricidin: a cyclic depsipeptide from *Paenibacillus* polymyxa E681 induces systemic resistance against phytophthora blight of red-pepper. Phytoparasitica 2013;41(1):49–58.
[60] Kato T, Sakazaki R, Hinoo H, Shoji J. The structures of tridecaptins B and C (studies on antibiotics from the genus *Bacillus*. XXV). J Antibiot 1979;32(4):305–12.
[61] Lohans CT, van Belkum MJ, Cochrane SA, Huang Z, Sit CS, et al. Biochemical, structural, and genetic characterization of tridecaptin A_1, an antagonist of Campylobacter jejuni. Chembiochem 2014;15(2):243–9.
[62] Shoji J, Kato T, Sakazaki R. The total structure of cerexin A (studies on antibiotics from the genus *Bacillus*. XVI). J Antibiot 1976;29(12):1268–74.
[63] Béchet M, Caradec T, Hussein W, Abderrahmani A, Chollet M, et al. Structure, biosynthesis and properties of kurstakins, nonribosomal lipopeptides from *Bacillus* spp. Appl Microbiol Biotechnol 2012;95(3):593–600.
[64] Peypoux F, Bonmatin JM, Wallach J. Recent trends in the biochemistry of surfactin. Appl Microbiol Biotechnol 1999;51:553–63.

[65] Dehghan Noudeh G, Housaindokht M, Bazzaz BSF. Isolation, characterization, and investigation of surface and hemolytic activities of a lipopeptide biosurfactant produced by *Bacillus subtilis* ATCC 6633. J Microbiol 2005;43(3):272–6.
[66] Heerklotz H, Seelig J. Leakage and lysis of lipid membranes induced by the lipopeptide surfactin. Eur Biophys J 2007;36:305–14.
[67] Vollenbroich D, Ozel M, Vater J, Kamp RM, Pauli G. Mechanism of inactivation of enveloped viruses by the biosurfactant surfactin from *Bacillus subtilis*. Biologicals 1997;25(3):289–97.
[68] Kracht M, Rokos H, Ozel M, Kowall M, Pauli G, Vater J. Antiviral and hemolytic activities of surfactin isoforms and their methyl ester derivatives. J Antibiot (Tokyo) 1999;52:613–9.
[69] Duitman EH, Hamoen LW, Rembold M, Venema G, Seitz H, et al. The mycosubtilin synthetase of *Bacillus subtilis* ATCC6633: a multifunctional hybrid between a peptide synthetase, an amino transferase, and a fatty acid synthase. Proc Natl Acad Sci U S A 1999;96(23):13294–9.
[70] Moyne AL, Shelby R, Cleveland TE, Tuzun S. Bacillomycin D: an iturin with antifungal activity against Aspergillus flavus. J Appl Microbiol 2001;90(4):622–9.
[71] Yu GY, Sinclair JB, Hartman GL, Bertagnolli BL. Production of iturin A by *Bacillus amyloliquefaciens* suppressing *Rhizoctonia solani*. Soil Biol Biochem 2002;34(7):955–63.
[72] Aranda FJ, Teruel JA, Ortiz A. Further aspects on the hemolytic activity of the antibiotic lipopeptide iturin A. Biochim Biophys Acta 2005;1713(1):51–6.
[73] Maget-Dana R, Thimon L, Peypoux F, Ptak M. Surfactin/iturin A interactions may explain the synergistic effect of surfactin on the biological properties of iturin A. Biochimie 1992;74(12):1047–51.
[74] Romero D, de Vicente A, Rakotoaly RH, Dufour SE, Veening JW, et al. The iturin and fengycin families of lipopeptides are key factors in antagonism of *Bacillus subtilis* toward Podosphaera fusca. Mol Plant Microbe Interact 2007;20(4):430–40.
[75] Hathout Y, Ho YP, Ryzhov V, Demirev P, Fenselau C. Kurstakins: a new class of lipopeptides isolated from *Bacillus thuringiensis*. J Nat Prod 2000;63(11):1492–6.
[76] Li MH, Ung PM, Zajkowski J, Garneau-Tsodikova S, Sherman DH. Automated genome mining for natural products. BMC Bioinforma 2009;10:185.
[77] Caboche S, Pupin M, Leclère V, Fontaine A, Jacques P, et al. NORINE: a database of nonribosomal peptides. Nucleic Acids Res 2008;36:D326–31.
[78] Boddy CN. Bioinformatics tools for genome mining of polyketide and non-ribosomal peptides. J Ind Microbiol Biotechnol 2014;41(2):443–50.
[79] Dorrestein PC. Mass spectrometry of small molecules and natural products. Nat Prod Rep 2014;31(6):704–5.
[80] Caradec T, Pupin M, Vanvlassenbroeck A, Devignes MD, Smaïl-Tabbone M, et al. Prediction of monomer isomery in Florine: a workflow dedicated to nonribosomal peptide discovery. PLoS ONE 2014;9(1):e85667.
[81] Mootz HD, Schwarzer D, Marahiel MA. Ways of assembling complex natural products on modular nonribosomal peptide synthetases. ChemBioChem 2002;3(6):490–504.
[82] Xu D, Côté JC. Phylogenetic relationships between *Bacillus* species and related genera inferred from comparison of 3' end 16S rDNA and 5' end 16S-23S ITS nucleotide sequences. Int J Syst Evol Microbiol 2003;53(3):695–704.
[83] Bhandari V, Ahmod NZ, Shah HN, Gupta RS. Molecular signatures for *Bacillus* species: demarcation of the *Bacillus subtilis* and *Bacillus cereus* clades in molecular terms and proposal to limit the placement of new species into the genus *Bacillus*. Int J Syst Evol Microbiol 2013;63(7):2712–26.
[84] Cole JR, Wang Q, Fish JA, Chai B, McGarrell DM, et al. Ribosomal Database Project: data and tools for high throughput rRNA analysis. Nucleic Acids Res 2014;42:D633–42.
[85] Tamura K, Stecher G, Peterson D, Filipski A, Kumar S. MEGA6: molecular evolutionary genetics analysis version 6.0. Mol Biol Evol 2013;30(12):2725–9.
[86] Letunic I, Bork P. Interactive Tree Of Life v2: online annotation and display of phylogenetic trees made easy. Nucleic Acids Res 2011;39:W475–8.

Chapter 2 The draft genome sequence of Paenibacillus polymyxa strain CCI-25 encompasses high potential for secondary metabolite production

Published in
Genome Announcements 2016, May 19;4(3). pii: e00366-16.
doi: 10.1128/genomeA.00366-16

Authors
Aleti G, Antonielli L, Corretto E, Nikolić B, Sessitsch A, Brader G.

Author information
AIT Austrian Institute of Technology GmbH, Health & Environment Department, Bioresources Unit, Konrad Lorenz Straße 24, A-3430 Tulln, Austria.

AMERICAN SOCIETY FOR MICROBIOLOGY | genomeAnnouncements

The Draft Genome Sequence of *Paenibacillus polymyxa* Strain CCI-25 Encompasses High Potential for Secondary Metabolite Production

Gajender Aleti, Livio Antonielli, Erika Corretto, Branislav Nikolić, Angela Sessitsch, Günter Brader

Bioresources Unit, Health and Environment Department, Austrian Institute of Technology (AIT), Tulln, Austria

We report here the draft genome sequence of *Paenibacillus polymyxa* strain CCI-25, which displays strong antifungal and antibacterial activities *in vitro*. The genome encompasses nonribosomal peptide synthetases predicted to encode a tridecaptin, polymyxin, fusaricidin, an iturin-like synthetase, a lantibiotic similar to paenicidin A, as well as a type 1 polyketide synthase.

Received 24 March 2016 **Accepted** 29 March 2016 **Published** 19 May 2016

Citation Aleti G, Antonielli L, Corretto E, Nikolić B, Sessitsch A, Brader G. 2016. The draft genome sequence of *Paenibacillus polymyxa* strain CCI-25 encompasses high potential for secondary metabolite production. Genome Announc 4(3):e00366-16. doi:10.1128/genomeA.00366-16.

Address correspondence to Günter Brader, guenter.brader@ait.ac.at.

Plant-associated *Paenibacillus polymyxa* strains are well noted for their production of a wide range of secondary metabolites (1–3), predominantly lipopeptides and polyketides involved in plant growth promotion and biocontrol of fungi (4–6). Here, we highlight the secondary metabolite capacity of *P. polymyxa* strain CCI-25 isolated from vermicompost. Both colonies and lipopeptide and polyketide crude extracts (7, 8) exhibited strong antimicrobial activity against *Escherichia coli* and fungi, including *Fusarium oxysporum* ACC01, *Botrytis cinerea* ofi 501-E (Austrian Institute of Technology [AIT] collection), and *Rhizoctonia solani* CBS101769, on plate assays.

To evaluate the molecular basis for secondary metabolite production, genomic DNA was isolated by phenol-chloroform extraction, and a library was prepared, according to the manufacturer's protocol, using the Nextera XT kit (Illumina, San Diego, CA). Library sequencing was performed using an Illumina MiSeq platform (MiSeq reagent kit version 3). Sequencing generated 2,213,773 paired-end reads with 124 ± 53-fold coverage after PhiX sequence removal by Bowtie2 (9). Adapter and quality trimming were performed using Trimmomatic-0.32 (10). Overlapping reads were merged with FLASH (11), and paired-end reads were assembled by SPAdes 3.1.0 (12). Quality control of mapping data was carried out by Qualimap 2.2 (13), and assembly quality was estimated by QUAST 3.2 (14). Assembly resulted in 117 contigs >1,000 bp, with an N_{50} size of 95,765 bp. The draft genome size is 5.61 Mb, with a G+C content of 44.95%. The identification of 40 highly conserved single-copy marker genes in the assembly by PhyloSift version 1.0.1 (15) indicated completeness of the genome and excluded contaminant sequences. Genomic BLAST showed similarities to *P. polymyxa* CR1. The NCBI Prokaryotic Genome Annotation Pipeline (PGAP) identified 5,146 genes, 4,953 coding sequences (CDSs), 15 complete 5S rRNAs, 30 partial 16S rRNAs, 37 partial 23S rRNAs (for a total of 15 putative rRNA operons), 107 tRNAs, 4 noncoding RNAs (ncRNAs), and 241 pseudogenes. The rRNAs were further confirmed by RNAmmer 1.2 (16). Prediction of secondary metabolite-encoding sequences was performed by antiSMASH (17).

The CCI-25 draft genome encompasses nonribosomal peptide synthetases with sequence similarities to published genes (1, 18–22), and the prediction includes the encoding of a tridecaptin with valine instead of isoleucine at the 13th position compared to *Paenibacillus terrae* NRRL B-30644 and fusaricidin C, and a polymyxin with leucine instead of phenylalanine at the 6th position compared to *P. polymyxa* M1. In addition, an iturin- and paenilarvin-like compound with altered monomer composition (D-Gly-D-Orn-D-Glu-D-nrp-L-nrp-L-Ile-L-Val) compared to the published metabolites from *Bacillus amyloliquefaciens* FZB42 (<58% identity) and *Paenibacillus larvae* DSM 25430 (<40% identity) (1, 23) has been predicted. CCI-25 contains a lantibiotic gene similar to paenicidin A and a predicted polyketide synthase with a different number of acyl carrier domains with 61% identity to bacillaene synthase from *B. amyloliquefaciens* FZB42 and 87% identity to *P. polymyxa* M-1 polyketide synthase (23). Given the fact that about 370 kb (6.6% of the total genome) is dedicated to secondary metabolite biosynthesis, CCI-25 has high potential to be exploited for medical or agricultural applications.

Nucleotide sequence accession number. The nucleotide sequences have been deposited at the DDBJ/EMBL/GenBank under the accession no. LTYJ00000000. The version described in this paper is the first version.

ACKNOWLEDGMENT

This work was supported by the life science project LS11-014 of the Niederösterreichische Forschungs-und Bildungsges.m.b.H (NFB).

FUNDING INFORMATION

This work, including the efforts of Günter Brader and Branislav Nikolić, was funded by NFB under grant number LS11-04.

REFERENCES

1. **Niu B, Rueckert C, Blom J, Wang Q, Borriss R.** 2011. The genome of the plant growth-promoting rhizobacterium *Paenibacillus polymyxa* M-1 contains nine sites dedicated to nonribosomal synthesis of lipopeptides and polyketides. J Bacteriol **193**:5862–5863. http://dx.doi.org/10.1128/JB.05806-11.
2. **Catch JR, Jones TSG, Wilkinson S.** 1949. The chemistry of polymyxin A. Ann N Y Acad Sci **51**:917–923. http://dx.doi.org/10.1111/j.1749-6632.1949.tb27318.x.

3. **He Z, Kisla D, Zhang L, Yuan C, Green-Church KB, Yousef AE.** 2007. Isolation and identification of a *Paenibacillus polymyxa* strain that coproduces a novel lantibiotic and polymyxin. Appl Environ Microbiol **73:** 168–178. http://dx.doi.org/10.1128/AEM.02023-06.
4. **Beatty PH, Jensen SE.** 2002. *Paenibacillus polymyxa* produces fusaricidin-type antifungal antibiotics active against *Leptosphaeria maculans*, the causative agent of blackleg disease of canola. Can J Microbiol **48:**159–169. http://dx.doi.org/10.1139/w02-002.
5. **Raza W, Yang X, Wu H, Wang Y, Xu Y, Shen Q.** 2009. Isolation and characterisation of fusaricidin-type compound-producing strain of *Paenibacillus polymyxa* SQR-21 active against *Fusarium oxysporum* f. sp. *nevium*. Eur J Plant Pathol **125:**471–483. http://dx.doi.org/10.1007/s10658-009-9496-1.
6. **Dijksterhuis J, Sanders M, Gorris LG, Smid EJ.** 1999. Antibiosis plays a role in the context of direct interaction during antagonism of *Paenibacillus polymyxa* towards *Fusarium oxysporum*. J Appl Microbiol **86:**13–21. http://dx.doi.org/10.1046/j.1365-2672.1999.t01-1-00600.x.
7. **Aktuganov G, Jokela J, Kivelä H, Khalikova E, Melentjev A, Galimzianova N, Kuzmina L, Kouvonen P, Himanen JP, Susi P, Korpela T.** 2014. Isolation and identification of cyclic lipopeptides from *Paenibacillus ehimensis*, strain IB-X-b. J Chromatogr B **973:**9–16. http://dx.doi.org/10.1016/j.jchromb.2014.09.042.
8. **Smyth TJP, Perfumo A, McClean S, Marchant R, Banat IM.** 2010. Isolation and analysis of lipopeptides and high molecular weight biosurfactants, p 3687–3704. In Timmis KN (ed), Handbook of hydrocarbon and lipid microbiology. Springer, Heidelberg, Germany.
9. **Langmead B, Salzberg SL.** 2012. Fast gapped-read alignment with Bowtie 2. Nat Methods **9:**357–359. http://dx.doi.org/10.1038/nmeth.1923.
10. **Bolger AM, Lohse M, Usadel B.** 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics **30:**2114–2120.
11. **Magoč T, Salzberg SL.** 2011. FLASH: fast length adjustment of short reads to improve genome assemblies. Bioinformatics **27:**2957–2963. http://dx.doi.org/10.1093/bioinformatics/btr507.
12. **Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Prjibelski AD, Pyshkin AV, Sirotkin AV, Vyahhi N, Tesler G, Alekseyev MA, Pevzner PA.** 2012. SPAdes: a new genome assembly algorithm and its applications to single cell sequencing. J Comput Biol **19:**455–477. http://dx.doi.org/10.1089/cmb.2012.0021.
13. **García-Alcalde F, Okonechnikov K, Carbonell J, Cruz LM, M, Götz S, Tarazona S, Dopazo J, Meyer TF, Conesa A.** 2012. Qualimap: evaluating next-generation sequencing alignment data. Bioinformatics **28:** 2678–2679. http://dx.doi.org/10.1093/bioinformatics/bts503.
14. **Gurevich A, Saveliev V, Vyahhi N, Tesler G.** 2013. QUAST: quality assessment tool for genome assemblies. Bioinformatics **29:**1072–1075. http://dx.doi.org/10.1093/bioinformatics/btt086.
15. **Darling AE, Jospin G, Lowe E, Matsen FA IV, Bik HM, Eisen JA.** 2014. PhyloSift: phylogenetic analysis of genomes and metagenomes. PeerJ **2:** http://dx.doi.org/10.7717/peerj.243.
16. **Lagesen K, Hallin P, Rødland EA, Staerfeldt HH, Rognes T, Ussery DW.** 2007. RNAmmer: consistent and rapid annotation of ribosomal RNA genes. Nucleic Acids Res **35:**3100–3108. http://dx.doi.org/10.1093/nar/gkm160.
17. **Blin K, Medema MH, Kazempour D, Fischbach MA, Breitling R, Takano E, Weber T.** 2013. antiSMASH 2.0 a versatile platform for genome mining of secondary metabolite producers. Nucleic Acids Res **41:** W204–W212 http://dx.doi.org/10.1093/nar/gkt449.
18. **Choi SK, Park SY, Kim R, Kim SB, Lee CH, Kim JF, Park SH.** 2009. Identification of a polymyxin synthetase gene cluster of *Paenibacillus polymyxa* and heterologous expression of the gene in *Bacillus subtilis*. J Bacteriol **191:**3350–3358. http://dx.doi.org/10.1128/JB.01728-08.
19. **Choi SK, Park SY, Kim R, Lee CH, Kim JF, Park SH.** 2008. Identification and functional analysis of the fusaricidin biosynthetic gene of *Paenibacillus polymyxa* E681. Biochem Biophys Res Commun **365:**89–95. http://dx.doi.org/10.1016/j.bbrc.2007.10.147.
20. **Lohans CT, Huang Z, van Belkum MJ, Giroud M, Sit CS, Steels EM, Zheng J, Whittal RM, McMullen LM, Vederas JC.** 2012. Structural characterization of the highly cyclized lantibiotic paenicidin A via a partial desulfurization/reduction strategy. J Am Chem Soc **134:**19540–19543. http://dx.doi.org/10.1021/ja3089229.
21. **Aleti G, Sessitsch A, Brader G.** 2015. Genome mining: prediction of lipopeptides and polyketides from *Bacillus* and related *Firmicutes*. Comput Struct Biotechnol J **13:**192–203.
22. **Lohans CT, Van Belkum MJ, Cochrane SA, Huang Z, Sit CS, McMullen LM, Vederas JC.** 2014. Biochemical, structural, and genetic characterization of tridecaptin A_1, an antagonist of *Campylobacter jejuni*. Chembiochem **15:**243–249. http://dx.doi.org/10.1002/cbic.201300595.
23. **Sood S, Steinmetz H, Beims H, Mohr KI, Stadler M, Djukic M, von der Ohe W, Steinert M, Daniel R, Müller R.** 2014. Paenilarvins: iturin family lipopeptides from the honey bee pathogen *Paenibacillus larvae*. Chembiochem **15:**1947–1955. http://dx.doi.org/10.1002/cbic.201402139.

Chapter 3 Surfactin variants mediate species-specific biofilm formation and root colonization in *Bacillus*

Published in
Environmental Microbiology 2016, Jun 16
doi: 10.1111/1462-2920.13405.

Authors
Aleti G[1], Lehner S[2], Bacher M[3], Compant S[1], Nikolic B[1], Plesko M[1], Schuhmacher R[2], Sessitsch A[1], Brader G[1].

Author information
[1]Health & Environment Department, Bioresources Unit, AIT Austrian Institute of Technology GmbH, AIT, Konrad Lorenz Straße 24, Tulln, A-3430, Austria.
[2]Department of Agrobiotechnology (IFA-Tulln), Center for Analytical Chemistry, University of Natural Resources and Life Sciences (BOKU), Konrad-Lorenz-Str. 20, 3430 Tulln, Austria.
[3]Department of Chemistry, University of Natural Resources and Life Sciences (BOKU), UFT Research Center Tulln, Konrad-Lorenz-Str. 24, Tulln, A-3430, Austria.

environmental microbiology

Environmental Microbiology (2016) **00**(00), 00–00 doi:10.1111/1462-2920.13405

Surfactin variants mediate species-specific biofilm formation and root colonization in *Bacillus*

Gajender Aleti,[1] Sylvia Lehner,[2] Markus Bacher,[3] Stéphane Compant,[1] Branislav Nikolic,[1] Maja Plesko,[1] Rainer Schuhmacher,[2] Angela Sessitsch[1] and Günter Brader[1]*

[1]*Health & Environment Department, Bioresources Unit, AIT Austrian Institute of Technology GmbH, AIT, Konrad Lorenz Strasse 24, Tulln A-3430, Austria.*

[2]*Department of Agrobiotechnology (IFA-Tulln), Center for Analytical Chemistry, University of Natural Resources and Life Sciences (BOKU), Konrad-Lorenz-Str. 20, 3430 Tulln Austria.*

[3]*Department of Chemistry, University of Natural Resources and Life Sciences (BOKU), UFT Research Center Tulln, Konrad-Lorenz-Str. 24, Tulln A-3430, Austria.*

Summary

Cyclic lipopeptides (cLP) and especially surfactins produced by *Bacillus* spp. trigger biofilm formation and root colonization and are crucial for biocontrol activity and systemic resistance in plants. *Bacillus atrophaeus* 176s isolated from the moss *Tortella tortuosa* produces the cLP fengycins, iturins and surfactins, possesses antifungal activities and can protect tomato, lettuce and sugar beet against *Rhizoctonia solani* infection. In *B. atrophaeus* we identified for the first time the variant surfactin C, which differs from surfactin A produced by *B. subtilis* and *B. amyloliquefaciens* by an isoleucine instead of a leucine at position 7 of the lipopeptide backbone. The analysis of the complete surfactin gene clusters revealed that the dissimilarity is encoded in the adenylation domain of *srfC* and show that surfactin variations are distributed in a species-specific manner in bacilli. We demonstrate that the surfactin A and C with subtle structural differences have varying signal strengths on biofilm formation and root colonization and act specifically on the respective producing strain. This became evident as biofilm formation and root colonization but not swarming motility in surfactin biosynthesis mutants was restored differentially in the presence of exogenously supplemented cognate and non-cognate surfactin variants.

Received 05 April 2016; accepted 31 May 2016. *For correspondence. E-mail guenter.brader@ait.ac.at.

Introduction

The bacterial genus *Bacillus* comprises important plant-associated strains utilized for control of plant diseases and for plant growth promotion (Ongena and Jacques, 2008). Three families of cyclic lipopeptides (cLP), surfactins, iturins and fengycins are considered as crucial components in these activities as they act as antifungal and antibacterial metabolites and have been shown to stimulate plant defense by inducing systemic resistance. Moreover, cLP have been demonstrated to play a vital role in biofilm formation and root colonization in *Bacillus subtilis* and *B. amyloliquefaciens* (Ongena *et al.*, 2007; Romero *et al.*, 2007; Ongena and Jacques, 2008).

The amphiphilic surfactins act hereby as powerful biosurfactants reducing surface tension alongside roots and facilitating bacterial swarming to nutrient rich niches (Kinsinger *et al.*, 2003; Bais *et al.*, 2004). Surfactin production is often associated with biofilm formation and subsequent root colonization and a strong correlation between defense-inducing activity and the amount of surfactin has been described (Coway *et al.*, 2015). Moreover, surfactins seem to be also involved in establishing intra- and interspecific communications between organisms thriving in the same ecological niche and are sensed by membrane disturbance recognized by the histidine kinase KinC (López *et al.*, 2009; Shah *et al.*, 2009; Oslizlo *et al.*, 2014).

Surfactins consist of a cyclic heptapeptide backbone connected via an amide and lactone bond to the carboxy- and β-hydroxy group to a fatty acid chain (variable length from C_{12} to C_{16}) respectively (Peypoux *et al.*, 1999), and their synthesis is mediated by mega-enzymes called non-ribosomal peptide synthetases (NRPS) (Finking and Marahiel, 2004; Stein, 2005). The adenylation domain, responsible for the selection and recruitment of amino acids on the peptide chain is responsible for biosynthesis of structurally diverse cLP that differ in peptide moiety (Ongena and Jacques, 2008). Variants of surfactin with the same peptide length but differences in amino acid

composition especially at position 1 and 7 have been described in *B. amyloliquefaciens, B. licheniformis, B. pumilus* and *B. subtilis*. Each of these variants contains isoforms that can vary both in length and branching point of the fatty acid chain (Ongena and Jacques, 2008; Jacques, 2011). Until now, little is known about specific roles and functions of variants and isoforms within a group of a given lipopeptide family.

Bacillus atrophaeus is a Gram-positive, non-pathogenic, sporulating bacterium distinguished from other *Bacillus* spp. by its dark pigmentation. It has been used as a surrogate for pathogenic *Bacillus anthracis* to study the spreading of its spores for safety and military aspects (Gibbons *et al.*, 2011) and recently, fengycin, a novel antifungal protein and several volatile compounds have been found in the plant-associated *B. atrophaeus* CAB-1 and were shown to contribute to the suppression of cucumber powdery mildew (Zhang *et al.*, 2013). However, production of other cLP in *B. atrophaeus* has not been reported so far.

In this study, we assessed the potential of *Bacillus atrophaeus* strain 176s for its capacity to produce cLP with antifungal activities and to control *Rhizoctonia solani* infection in plants and show that the surfactin variants produced in a species-specific manner are acting in their cognate form as signal for biofilm formation and root colonization.

Results

In vitro *antifungal and hemolytic activities of* B. Atrophaeus *176s*

The complete 16S rRNA gene sequence of strain 176s isolated from the moss *Tortella tortuosa* is identical to those of *Bacillus atrophaeus* strains 1942, BSS, NRS 1221A and UCMB-5137. Further *gyrA* gene sequence analysis revealed more than 99% identity to the *gyrA* genes of the *B. atrophaeus* strains. Phenotype characteristics of dark pigmentation observed on TSA plates confirmed the identity of 176s as *Bacillus atrophaeus*.

Bacillus atrophaeus 176s and its supernatants showed antifungal activity against the phytopathogens *Botrytis cinerea* ofi 501-E, *Fusarium oxysporum* ACC01, *Sclerotinia sclerotiorum* MA5092, *Rhizoctonia solani* CBS 101769 and FT1510 comparable to *B. amyloliquefaciens* FZB42 well known for its production of cLPs with antifungal activity (Supporting Information Fig. S1A). Crude cLPs extracts of *B. atrophaeus* 176s showed in liquid a minimum inhibitory concentration (MIC) against *R. solani* of 40 μg/mL comparable to a MIC of 100 μg/mL obtained with *B. amyloliquefaciens* FZB42 crude cLP extracts (data not shown), indicating a potential application in biocontrol.

The hemolytic activity of the crude cLP extracts of *B. atrophaeus* 176s was similar to commercial surfactin (Supporting Information Fig. S2A). Hemolytic activities of wild type (WT) and mutant strains of *B. atrophaeus* and *B. subtilis* published to have reduced hemolytic activities are shown in Supporting Information Fig. S2B. All tested WT strains showed hemolysis, but the *B. atrophaeus* mutant ATCC 9372-1 and the *B. subtilis* mutant OKB 105 showed no hemolysis on plate. The hemolysis mutants also displayed reduced antifungal activities. WT *B. atrophaeus* strain 1942 and WT *B. subtilis* strain OKB 105 exhibited antifungal activity, but less pronounced than *B. atrophaeus* 176s. Their natural mutants ATCC 9372-1and OKB 120 did not reveal antifungal activity (Supporting Information Fig. S1B).

Identification of the cLP based on LC-HRMS(/MS) analysis

To characterize cLP production potentially responsible for protective, antifungal and hemolytic activities, culture filtrates of *B. atrophaeus* 176s grown for 24 h, 48 h and 72 h on Landy, LB and TSB medium at 21 and 28°C were tested for their hemolytic and antifungal properties. Filtrates with the highest activity were grown on Landy at 28°C for 72 h and were subjected to LC-HRMS(/MS) analysis. As reference, *B. amyloliquefaciens* FZB42 known for its capacity to produce surfactin A, bacillomycin D and fengycin (plipastin) A and B, the fengycin and bacillomycin deficient mutant of FZB42 (AK3: ΔbmyA::EmR and Δfen::CmR) as well as commercial standards for surfactin A and iturin were used.

The analysis revealed a series of four peaks with identical mass to surfactins with a β-hydroxy fatty acid chain length from C_{13} to C_{16} in *B. atrophaeus* 176s. However, the retention times of the peaks in the extracted ion chromatograms (XICs, ± 5 ppm), were slightly shifted compared to the corresponding derivatives of different chain length of the authentic surfactin standard (Table 1). For each of the four compounds, the LC-HRMS(/MS) spectra with same precursor *m/z* exhibited the same fragmentation behavior as observed for the corresponding reference peptide (see Fig. 1 for the most intense signal at *m/z* 1022.67), which indicates an identical amino acid sequence and linkage of the β-hydroxy fatty acid as in the surfactin A standard. It should be noted however, the low collision energy fragmentation (Fig. 1) does not allow discrimination between isoleucine and leucine, D- and L-isomers or of different fatty acid isomers, which all pose potential reasons for the observed retention time shift.

LC-HRMS(/MS) analysis of *B. atrophaeus* 176s culture filtrates also showed intense peaks with identical masses (*m/z* 1449.8; 1463.8; 1477.8; 1491.8; 1505.9) to fengycin A and B and with fatty acid of a chain length from C_{15} to C_{17}. This series of peaks with the same retention time and the same fragmentation pattern indicative of fengycin A and B was also observed in *B. amyloliquefaciens* FZB42. The peaks were all missing in the mutant AK3 unable to

Table 1. Surfactins with different fatty acid chain length of culture filtrates of *B atrophaeus* 176s grown for 72 h on Landy broth and detected by LC-HRMS(/MS).

Fatty acid chain length	*m/z*	% Relative signal intensity of combined surfactins	Retention time (minutes) *B. atrophaeus* 176s	*B. amyloliquefaciens* FZB42
C_{13}	1008.66	21.9 ± 3.9%	18.9	18.2
C_{14}	1022.67	45.2 ± 1.9%	19.9	19.2
C_{15}	1036.69	31.6 ± 3.6%	20.7	19.9
C_{16}	1050.71	1.3 ± 0.3%	21.9	21.1

produce fengycins indicating that *B. atrophaeus* 176s produces the same series of fengycins as *B. amyloliquefaciens* FZB42. In addition, *B. atrophaeus* 176s culture filtrates contained low abundant of peaks with identical retention time and accurate mass (*m/z*: 1057.57; 1071.58; 1085.60) to the commercial iturin standard. In contrast, the bulk of bacillomycin D peaks observed in *B. amyloliquefaciens* FZB42 were missing in *B. atrophaeus* 176s. Together, the LC-HRMS(/MS) data indicate the production of iturins in *B. atrophaeus* 176s.

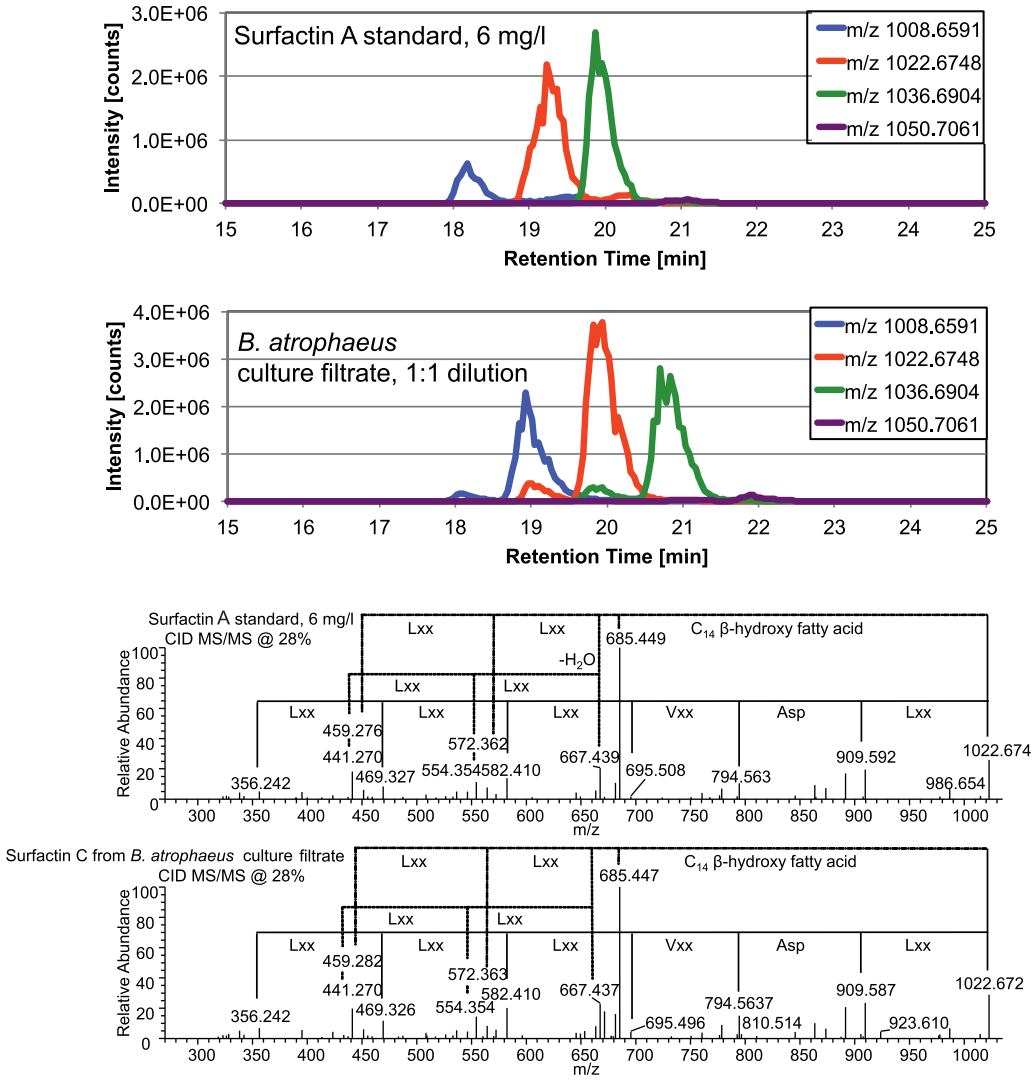


Fig. 1. Extracted ion chromatograms (XIC, ± 5ppm) and CID MS/MS spectra of the protonated molecule of the surfactin containing a C_{14} β-hydroxy fatty acid with *m/z* 1022.67 for the surfactin standard (6.67 mg/l) and a 1:1 dilution of *Bacillus atrophaeus* 176s culture filtrate.

Structure of surfactins and organization of cLP gene clusters in B. Atrophaeus *176s*

In order to determine the exact chemical structure of surfactins from *B. atrophaeus* 176s the purified surfactin (see SI Material and Methods for hemolytic and antifungal activity guided isolation, Supporting Information Fig. S3) was analyzed by NMR, which showed that the isolated compound contains a β-hydroxy fatty acid, 3 leucines and one glutamic acid, valine, aspartic acid and one isoleucine residue, clearly identified by its spin system in the TOCSY spectrum with resonances at δ 4.23 (H_α), 1.87 (H_β), 1.37 ($H_{\gamma 1}$-a), 1.15 ($H_{\gamma 1}$-b), 0.89 ($H_{\gamma 2}$) and 0.87 (H_δ) and the corresponding carbon signals at δ 57.86 (C_α), 38.47 (C_β), 25.78 ($C_{\gamma 1}$), 15.92 ($C_{\gamma 2}$), and 11.84 (C_δ), respectively. Leucine and Isoleucine can clearly be distinguished by the characteristic chemical shifts of their methyl groups (Supporting Information Table S1; Supporting Information Fig. S4). Furthermore the amino acid sequence of the lipopeptide was determined by long-range crosspeaks of the amide protons of the single amino acids to the carbonyl carbons of the neighboring amino acid in the band selective hmbc spectra. By selective excitation of the carbonyl region all carbonyl resonances could be unequivocally assigned. The amino acid sequence of the compound was established as that of surfactin C, previously isolated from different *Bacillus* strains and studied by NMR spectroscopy (Baumgart *et al.*, 1991; Itokawa *et al.*, 1994; Tang *et al.*, 2007), differing from surfactin A by a leucine isoleucine substitution. The crosspeak from H-3 of the fatty acid to Ile-CO clearly locates this amino acid at position 7 in the lipopeptide sequence. Mass spectral data suggest a mixture of different chain lengths of the fatty acids (C_{13} to C_{16}), but no information on the type of the isomers (n, anteiso, or iso) could be derived from the fragmentation pattern. A more detailed analysis of the ^{13}C and hsqc spectra revealed a mixture of iso- and ante-iso isomers (Lin *et al.*, 1994; Huszcza *et al.*, 2006), whereas non-alkyl could be detected. In addition by performing a quantitative hsqc (Heikkinen *et al.*, 2003) the ratio of anteiso: iso was determined as being approx. 60: 45.

In order to understand the genetic basis for production of the lipopeptides in *B. atrophaeus* 176s, whole genome 454 pyrosequencing was performed.190,069 reads with an average length of 431 bp, reads were generated and assembled to 51 contigs encompassing size of 4.3 Mbp. A total of 9 putative gene clusters containing genes encoding NRPS, PKS, NRPS/PKS hybrids, siderophores and terpenes were predicted using antiSMASH (Weber *et al.*, 2015) and subsequently clusters were assigned to the contigs. The predicted NRPS clusters involved in the synthesis of surfactin, iturin and fengycin were subjected to BLAST and in all cases high similarities to the corresponding clusters of *B. atrophaeus* 1942 were observed. Therefore, surfactin, iturin and fengycin gene sequences of *B. atrophaeus* 1942 were used as reference sequences to assemble the corresponding *B. atrophaeus* 176s contigs to scaffolds. AntiSMASH prediction for the fengycin cluster of *B. atrophaeus* 176s and 1942 is identical to the prediction for *B. amyloliquefaciens* FZB42 and expects a sequence of the lipopeptide backbone of L-Glu-D-Orn-L-Tyr-D-Thr-L-Glu-D-Val-L-Pro-L-Gln-D-Tyr-L-Ile, which is consistent with LC-HRMS(/MS) data and plipastin (fengycin) structures both in *B. atrophaeus* 176s and in *B. amyloliquefaciens* FZB42. The predictions for the iturin clusters of *B. atrophaeus* 176s and 1942 are different to that for *B. amyloliquefaciens* FZB42 and predict a mycosubtilin structure with the oligopeptide sequence L-Asn-D-Tyr-D-Asn-L-Gln-L-Pro-D-Ser-L-Asn in agreement with the LC-HRMS(/MS) data indicating the presence of iturins.

Due to the modular structure of the NRPS similar regions regarding *srfA* and *srfB* genes the surfactin cluster could not be unambiguously assigned with NGS data alone. A complete surfactin gene cluster was obtained by gap filling with sequences obtained from PCR with primers as indicated in Supporting Information Table S2 followed by Sanger sequencing. Overall, the complete surfactin gene cluster of *B. atrophaeus* 176s showed 97% identity to *B. atrophaeus* 1942 and 79% to *B. amyloliquefaciens* FZB42 on amino acid level. AntiSMASH analysis of the surfactin cluster of *B. atrophaeus* 176s and 1942 predicted a lipopeptide sequence of L-Glu-L-Leu-D-Leu-L-Val-L-Asp-D-Leu-L-Ile consistent with the LC-HRMS(/MS) and NMR data and the surfactin C structure (Fig. 2, Supporting Information Table S3).

Antifungal and 176s detection assays in greenhouse

To further evaluate the biocontrol activity of *B. atrophaeus* 176s greenhouse assays with lettuce, sugar beet and tomato infected with *R. solani* were performed under greenhouse conditions. *B. amyloliquefaciens* FZB42 was used as reference strain, as it has been shown to form robust biofilms, colonize roots, support plant growth and exhibit strong biocontrol activity against *R. solani* in lettuce (Chowdhury *et al.*, 2013). The assays revealed that untreated plants of lettuce, sugar beet and tomato were susceptible to fungal infections and majority of the plants died due to damping off within seven days after *R. solani* infection (Supporting Information Fig. S5). Plants inoculated with *B. amyloliquefaciens* FZB42 and *B. atrophaeus* 176s were protected in a comparable manner. Application of both strains resulted in a higher resistance to fungal infection with reduced symptoms (Supporting Information Fig. S5A) and increased plant recovery (Supporting Information Fig. S5B).

Similarly, lettuce plants were treated with WT and their natural surfactin-deficient mutant strains of *Bacillus* and

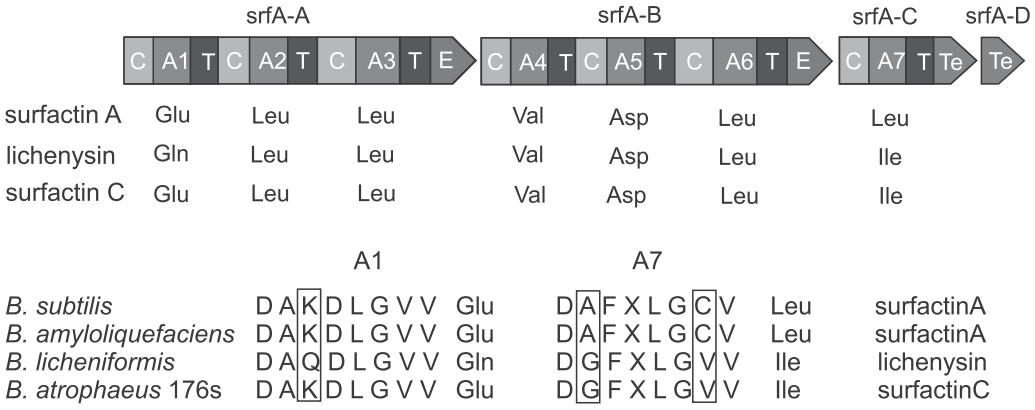


Fig. 2. The architecture of a typical surfactin synthetase gene coding for different surfactin variants.
Condensation domains (C), adenylation domains (A), thiolation domains (T) and thioesterase domains (TE) for the genes of the surfactin operons are shown. The amino acids recruited for the lipopeptide by each adenylation domain are indicated for the different surfactin variants. Dissimilarities between surfactin variants and species can be found in the active site of the adenylation domains A1 and A7 and responsible amino acids in the primary sequence are indicated in one letter code.

challenged with *R. solani* under greenhouse conditions in order to assess the effect of functional surfactin production on biocontrol activity. Two weeks post-inoculation of fungi a clear protective effect was shown by WT strains *B. atrophaeus* 1942 and *B. subtilis* OKB105 compared to the surfactin-deficient mutants *B. atrophaeus* ATCC9372 and *B. subtilis* OKB120, implying surfactin production is crucial for biocontrol activity (Supporting Information Fig. S6A and Supporting Information Fig. S6B).

Surfactin cluster sequences of *B. atrophaeus* 176s show enough differences in the primary nucleotide sequence to *B. atrophaeus* strains in the database to allow design of specific marker primers on the 3`end of the *sfrA* gene (Supporting Information Table S2). These primers allowed the specific detection of this strain in *B. atrophaeus* 176s pre-treated roots of tomato, sugar beet and lettuce, while control and *B. amyloliquefaciens* FZB42 treated plants and controls gave no specific PCR product. The amplicon was 700 bp and identity as 176s *srfA* was confirmed by Sanger sequencing.

Restoration of swarming motility and pellicle formation

To evaluate a potential specific role of the surfactin variants A and C, we examined differences in swarming motility and the pellicle formation in WT and mutant strains of *B. subtilis* and *B. atrophaeus*. The surfactin-deficient mutant ATCC9372 of *B. atrophaeus* showed slightly reduced swarming motility compared to the corresponding WT 1942 (Supporting Information Fig. S7). Also the surfactin-deficient *B. subtilis* mutant OKB120 had less swarming activity compared to its WT OKB 105, albeit less pronounced. After exogenous addition of surfactin A and surfactin C to the growth medium, both surfactins could slightly elicit swarming in the *B. atrophaeus* mutant ATCC9372 and in the *B. subtilis* mutant OKB120, comparable to levels of the corresponding WT strains (Supporting Information Fig. S7; lower panel).

In opposite to the effects on cell swarming, pellicle formation was restored in surfactin biosynthesis mutants to varying amounts depending on the exogenously supplemented surfactin. Importantly, *B. atrophaeus* mutant ATCC9372 with reduced surfactin production responded to exogenous, cognate surfactin C by forming robust pellicles, while only sparse pellicle was observed in the presence of *B. subtilis*–derived surfactin A. Similarly, exogenous, cognate surfactin A triggered robust pellicles in the mutant OKB120 of *B. subtilis*, while *B. atrophaeus*-derived surfactin C induced only sparse pellicles (Fig. 3). This indicates a compound-specific response.

Restoration of bacillus *biofilm development and root colonization by surfactin under gnotobiotic environment*

Given that the pellicle development in surfactin deficient *Bacillus* strains was restored by addition of its cognate surfactin exogenously, we further assessed whether these surfactin mutants might also restore biofilm and root colonization in lettuce plants upon addition of surfactin A or C. One week after inoculation of *Bacillus* strains, we performed live staining and examined the root system under confocal laser scanning microscope to visualize *Bacillus* biofilm and root colonization of lettuce plants grown under gnotobiotic conditions. The WT strains including *B. atrophaeus* 176s and *B. amyloliquefaciens* FZB42 were able to form clusters of cells along cell-walls and root hairs forming biofilm like structures, colonizing the roots (Fig. 4). The WT *B. subtilis* OKB105 showed reduced biofilms compared with the other WT strains, whereas the WT *B. atrophaeus* 1942 heavily colonized the root elongation

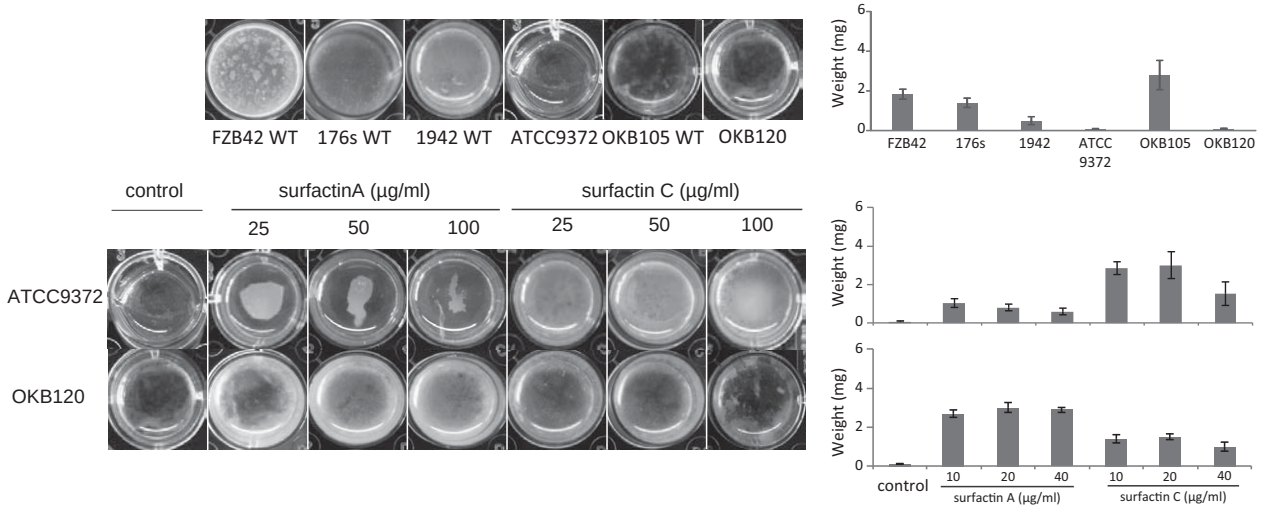


Fig. 3. Influence of surfactin variants on floating pellicle formation.
WT and mutant strains affected in surfactin biosynthesis of *B. amyloliquefaciens, B. subtilis* and *B. atrophaeus* as indicated were grown on LB medium (top and control lane) or LB supplemented with surfactin A from *B. subtilis* (bought from Sigma) and purified surfactin C from *B. atrophaeus* at different concentrations as indicated. Pellicle formation was restored to varying degrees in presence of exogenous surfactins. Bars on the right show pellicle weights determinate in triplicates with standard deviations.

zone but did not display visible biofilm like structures. In contrast, the corresponding surfactin biosynthesis mutant strains *B. atrophaeus* ATCC9372 and *B. subtilis* OKB120 showed only sporadic root hair colonization and showed no sign of biofilm development. Interestingly, addition of exogenous surfactin C but not of surfactin A to the rhizosphere soil has induced robust colonization of *B. atrophaeus* ATCC9372, whereas surfactin A supplementation resulted in increased colonization density and biofilm formation in *B. subtilis* OKB120 suggesting species-specific perception of surfactin variants in induction of biofilm development and root colonization (Fig. 4). The results of biofilm development and root colonization in lettuce plants are consistent with in vitro pellicle development assays.

Discussion

cLP play important roles in plant colonialization and biological control properties of bacilli. Here, we show that the *B. atrophaeus* 176s produces three classes of cLP including antifungal fengycins and surfactins, the latter a compound class which is crucial for biofilm formation and root colonization of plants (Bais et al., 2004; Ongena and Jacques, 2008). Interestingly, the surfactin variant from *B. atrophaeus* 176s differs from surfactins from *B. amyloliquefaciens* and *B. subtilis*. The difference is encoded in *srfC* of the surfactin biosynthesis gene cluster and species-specific surfactin variants are produced by *B. amyloliquefaciens, B. atrophaeus*, *B. licheniformis* and *B. subtilis* indicating that these surfactins may act as species-specific signaling compounds. Supplementation of *B. atrophaeus* and *B. subtilis* mutants with cognate and non-cognate surfactins suggests a specific signaling role in biofilm formation and root colonization.

We identified both gene clusters and showed production of surfactin, iturin and fengycin cLP families in *B. atrophaeus* 176s, compounds likely playingcrucial roles for the protective activities, similarly as described for *B. amyloliquefaciens* (Ongena and Jacques, 2008; Cawoy *et al*., 2015) and for fengycin fractions of *B. atrophaeus* CAB-1 (Zhang *et al*., 2013). While surfactins are likely responsible for hemolytic activity, the strong antifungal *in vitro* activity of *B. atrophaeus* 176s can be explained by the production of fengycins. Also co-production of the various cLP may synergistically enhance their individual activities and various lipopeptides can induce resistance in plants (Razafindralambo *et al*., 1997; Ongena *et al*., 2007; Romero *et al*., 2007). Up to now, this is the first report to show the presence and co-production of three lipopeptide families and specifically of surfactin C in *B. atrophaeus* and taken together, these results imply its potential application as biocontrol agent.

Intriguingly, the surfactins from *B. atrophaeus* differs from surfactin A of *B. subtilis* and *B. amyloliquefaciens* by a Leu instead of Ile at position 7 (Koumoutsi *et al*., 2004; Chen *et al*., 2009a; Chen *et al*., 2009b). The assembled surfactin NRPS cluster of *B. atrophaeus* 176s revealed 97% identity to *B. atrophaeus* 1942 and 79% to *B. amyloliquefaciens* FZB42 on amino acid level. The non-ribosomal code encoded by eight amino acids within adenylation domain allows the prediction of A-domain specificity for certain amino acids based on the primary sequence (Stachelhaus *et al*., 1999; Jacques, 2011). Here, we confirmed that the

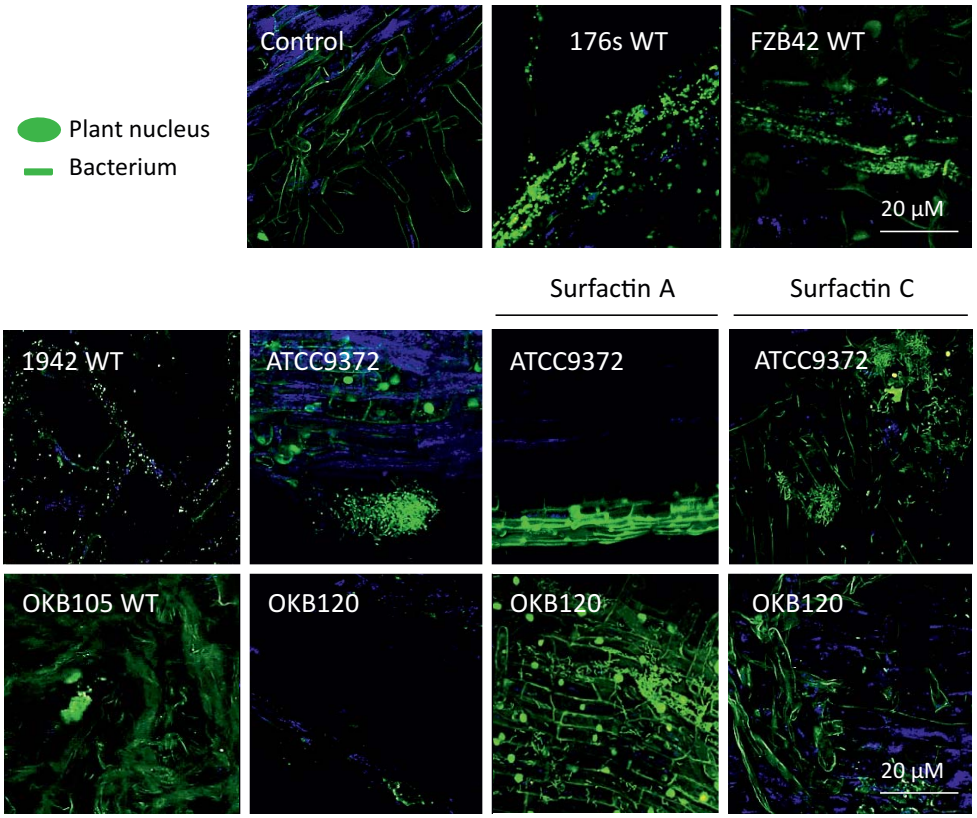


Fig. 4. Biofilm formation and root colonization of WT and surfactin biosynthesis mutant strains of *B. amyloliquefaciens*, *B. atrophaeus* and *B. subtilis*.
Bacillus cells on lettuce roots were visualized by live staining as described in methods. While WT strains were able to colonize the root and except strain 1942 form robust biofilms, the corresponding mutant strains deficient in surfactin production were able to colonize only few regions of the root with severely reduced cell numbers. Exogenous supplementation of surfactin C at a concentration of 40 μM has improved the colonization density of *B. atrophaeus* ATCC9372 compared to control and surfactin A, and vice versa in the case of *B. subtilis* OKB120 were biofilm formation was restored specifically. Pictures are representative of roots of three plants. Scale bars: 20 μm.

non-ribosomal code showed identical formation of amino acids in *B. atrophaeus* 176s and *B. amyloliquefaciens* FZB42 in the first six amino acids of the surfactin (Koumoutsi *et al.*, 2004), but striking dissimilarities can be found in of the A domain of *srfC* that encompasses the module 7 (Fig. 2). We observed that substitution of an Ala by a Gly and a Cys by a Val in the primary amino acid composition of A domain seems responsible for recruiting an Ile instead of a Leu in the formation of surfactin. These results are similar to the previous findings observed in the module 7 of lichenysin operon (Jacques, 2011). In our previous study (Aleti *et al.*, 2015), utilizing web based prediction tools we confirmed that the majority of the surfactin variants encoded in the genomes of bacilli and related species differed in peptide moiety particularly at position 1 and 7. It is also evident that the well-known *B. subtilis* and *B. amyloliquefaciens* species code for Leu at position 7, while *B. atrophaeus* code for Ile and *B. licheniformis* strains encode Gln and Ile at position 1 and 7, respectively (Aleti *et al.*, 2015) (Fig. 2, Supporting Information Table S3). Earlier work on cLP production especially in *Bacillus subtilis* has shown that growth medium supplemented with amino acids substitute L amino acid residues (Peypoux *et al.*, 1999). The variant production has been associated with the flexibility of adenylation domains, which can activate and recruit amino acids with similar structures (Shu *et al.*, 2002). The results here imply that under growth conditions without specific artificial amino acid the surfactin synthetase precisely produces either surfactin A or C and this information is genetically fixed in NRPS in the form of non-ribosomal code.

Surfactin variants are likely species-specific and genetically encoded suggesting that these compounds are involved in species-specific signaling potentially resulting in different ecological behavior. Intriguingly, surfactin has been implied as a signaling molecule in microbial communication (López *et al.*, 2009; Shah *et al.*, 2009) and is

believed to act as a quorum sensing molecule involved in the activation of a membrane-associated sensory receptor histidine kinase (KinC). KinC activates a pathway for biofilm formation and consequently root colonization (Ongena and Jacques, 2008; López *et al*., 2009). In addition, strong biosurfactant nature of surfactin plays an intricate role on bacterial swarming to nutrient rich environments and to the biofilm expansion (Kinsinger *et al*., 2003; Bais *et al*., 2004; Angelini *et al*., 2009). In accordance with previous research (Ghelardi *et al*., 2012) our results also indicated that the surfactins influenced swarming motility and were able to enhance the phenotype of swarming motility in surfactin-deficient *Bacillus* spp mutants, particularly in *B. atrophaeus*. We could not observe any surfactin A or C specific effect, pointing to an effect due to the amphiphilic nature of the surfactins or variant unspecific signaling. Biofilm formation and root colonialization, however, was strikingly restored to varying degrees by different surfactin variants. Surfactin mutants of *B. atrophaeus* and *B. subtilis*, which are deficient in pellicle formation androot colonialization, effectively restored phenotypes only in presence of their native surfactins. The differences in the ability of *Bacillus* spp. to respond to exogenous surfactins and to form biofilms suggest the variations in the signal perception of surfactin A and C depending on the genetic background and suggest that subtle structural differences may play a role in species-specific communication. In this respect it is also interesting to note that surfactin A and C show also differences in their confirmation in solution (Itokawa *et al*., 1994) indicating that even the small difference between a Ile and a Leu moiety can have effects on the three dimensional structure of a molecule and consequently on its ability to act as a signal component.

Small signaling molecules like homoserine lactones are very well described for their quorum sensing signaling in Gram-negative bacteria and small chain length differences in the structure of these molecules can lead to substantial differences in cell signaling (Sjöblom *et al*., 2006). However, the role of native surfactins in induction of robust biofilms in *Bacillus* spp. is not well understood. In *B. subtilis* five distinct sensor kinases (Kin A-E) are known to activate the master regulator Spo0A by phosphorylation, which in turn triggers the cellular events sporulation and biofilm formation in response to several environmental and physiological cues (LeDeaux *et al*., 1995; Jiang *et al*., 2000; Vlamakis *et al*., 2013). Surfactins act as a signaling molecule selectively induce potassium leakage and known to stimulate biofilm formation upon membrane disturbance and KinC perception (López *et al*., 2008). Recently, it has also been described that the membrane-associated chaperone protein, flotillin (FloT) embedded in the membrane micro-domains of *Bacillus*, interact with KinC to promote effective binding of specific-signaling proteins (Schneider *et al*., 2015). Also host related factors might modulate biofilm formation (Beauregard *et al*., 2013) and related histidine kinases, particularly KinD, seem to be involved in sensing products released by plant roots (Chen *et al*., 2012). It remains to be seen Kin histidine kinases signaling or so far uncharacterized components are involved in variant-specific recognition of surfactins.

Experimental procedures

Bacillus and fungal strains, cultivation conditions, and in vitro *assays*

B. atrophaeus strain 176s has been isolated from surface-sterilized *Tortella tortuosa* (Pottiaceae, Bryophyta) grown in an Austrian pine forest on limestone. Detailed description on *in vitro* assays and greenhouse assays for protection against *R. solani* are presented in Supporting Information Material and Methods. The identity of the *Bacillus* strain was evaluated by Sanger sequencing of 16S rRNA (8F and 1520R primers) (Weisburg *et al*., 1991) and *gyrA* genes (gyrA-F and gyrA-R primers) (Chun *et al*., 2000), further sequenced with the primers (1520R and gyrA-R, respectively) (Supporting Information Table S1). The gene sequences of 16S rRNA and *gyrA* have been deposited under GenBank accession number KT777650 and KT777651. The reference *Bacillus* strains *B. amyloliquefaciens* FZB42, its mutant AK3 ΔbmyA::EmR, Δfen::CmR deficient in bacillomycin and fengycin biosynthesis, *B. atrophaeus* 1942 and its natural mutant ATCC9372 affected in surfactin biosynthesis as well as *B. subtilis* OKB105 and its mutant OKB120 unable to produce surfactin (Nakano *et al*., 1988; Vollenbroich *et al*., 1994) were obtained from the *Bacillus* Genetic Stock Center (BGSC) (Table 2). All bacilli were cultivated in tryptic soy broth for maintenance and DNA isolation.

Fungal strains are summarized in Table 2. For biocontrol assays, fungal mycelial plugs were placed on YMA plates and incubated at 28°C. After one day, *Bacillus* strains or their cell free supernatants were placed 2cm away from the plug and plates were evaluated after one week. MIC assays were set up in 96 microtitre plates and dilution series of extracted cLP were incubated against *R. solani* at 28°C for 3 days according to the protocol Troskie *et al*. (2012). For hemolytic assays, 1 μL overnight cultures of *Bacillus* strains, cell free supernatants and lipopeptide fractions were tested on 10% sheep blood agar plates and incubated at 28°C for 48 h. Surfactin A from *B. subtilis* (from Sigma) were included as positive control.

Analysis and purification of cLP

For cLP production, a single colony from TSA plates was used as inoculant in three different media (Landy, TSB and LB medium) and grown at different temperatures (21 and 28°C). After 24, 48 and 72 h, the growth medium was centrifuged (5 min, 14,200*g*), the supernatant was filter-sterilized and 500 μL cell free culture filtrate was then mixed with 500 μL acetonitrile (Merck, pA) for LC-HRMS(/MS) analysis on a Accela HPLC system (Thermo Fisher Scientific, San Jose, CA, USA) coupled to an LTQ Orbitrap XL (Thermo Fisher Scientific) equipped with an electrospray ionization (ESI)

Table 2. Bacillus WT and mutant strains used in this study.

Microbial strain	Description	Source/reference
B. atrophaeus 176s	Wild type	this study; AIT
B. atrophaeus 1942	Wild type	Gibbons et al., 2011; BGSC
B. atrophaeus ATCC9372-1	Natural surfactin biosynthesis mutant	Gibbons et al., 2011; BGSC
B. subtilis OKB105	Wild type	Nakano et al., 1988; BGSC
B. subtilis OKB120	Surfactin biosynthesis mutant, srfa::Tn917 ΩOK120	Nakano et al., 1988; BGSC
B. amyloliquefaciens FZB42	Wild type	Koumoutsi et al., 2004; BGSC
B. amyloliquefaciens AK3	Bacillomycin & fengycin biosyntheis mutant, δbmya::emr, Δfen::cmr	Koumoutsi et al., 2004; BGSC
Botrytis cinerea ofi 501-E	Wild type fungus	AIT
Fusarium oxysporum ACC01	Wild type fungus	AIT
Sclerotinia sclerotiorum MA5092	Wild type fungus	ACBR
Rhizoctonia solani CBS101769	Wild type fungus from soy bean	CBS
Rhizoctonia solani FT1510	Wild type fungus from sugar beet	ACBR

ACBR: Vienna University of Natural Resources and Life Sciences strain collection; AIT: Austrian Institute of Technology strain collection; BGSC: *Bacillus* Genetic Stock Center; CBS: Centraalbureau voor Schimmelcultures.

interface which was operated in positive ionization mode (Supporting Information Methods).

For preparative isolation, crude cLP were extracted as previously described (Vater *et al.*, 2002; Smyth *et al.*, 2010). The cLP were further purified by an adapted SPE (solid phase extraction) technique (Kinsella *et al.*, 2009; Pertot *et al.*, 2013) as described in Supporting Information Methods.

For structural analysis, the purified surfactin was analyzed by NMR on a Bruker Avance II 400 (resonance frequencies 400.13 MHz for ^{1}H and 100.61 MHz for ^{13}C) equipped with a' 5 mm observe broadband probe head (BBFO) with z–gradients at room temperature with standard Bruker pulse programs (Supporting Information Methods). The samples were dissolved in 0.6 mL of CD3CN (99.8% D, euriso-top). Chemical shifts are given in ppm, referenced to residual solvent signals (1.94 ppm for ^{1}H, 118.26 ppm for ^{13}C).

Sequence analysis

Overnight cultures were harvested by centrifugation and DNA was isolated with phenol/chloroform and isopropanol precipitation. Five μg DNA of *B atrophaeus* 176s was subjected to pyrosequencing with a 454 SeqAnal Roche at GATC (Konstanz, Germany). Reads were assembled to contigs with the assembly package at GATC. Contigs were analyzed and annotated with RAST (Brettin *et al.*, 2015) and antiSMASH (Weber *et al.*, 2015) for the presence of contigs encoding for secondary metabolites. Contigs were assembled to scaffolds with *B atrophaeus* 1942 as reference. Contigs 4, 11, 18 and 33 contain surfactin encoding regions, which were further assembled by Sanger sequencing of the joint regions using the primers in the regions *srfA, srfB* and *srfC* (Supporting Information Table S2). The complete gene sequences of surfactin, iturin and fengycin biosynthesis clusters have been deposited in GenBank under accession numbers KP943734 to KP943745.

Swarming motility and floating pellicle formation assays

Swarming motility and biofilm formation assays were carried out in 24-well microtiter plates as previously reported (Luo *et al.*, 2015). 1 μL of cells from the overnight cultures at 28°C were inoculated on LB plates containing 20 μg/mL congo red and 10 μg/mL coomassie brilliant blue solidified by adding 0.7% agar. Following plates were incubated at 28°C for 72 h to evaluate swarming motility by measuring the diameter. For exogenous surfactin supplementation studies, appropriate dilutions of surfactin from Sigma and purified surfactin from *B. atrophaeus* were spread and air dried before the assay.

For floating biofilm formation (pellicle) studies, 1 μL of cells from the overnight cultures at 28°C were subsequently inoculated in 2 mL of LB medium containing 20 μg/mL congo red and 10 μg/mL coomassie brilliant blue in 24-well microtiter plates, next incubated at 28°C without shaking and dry weight of floating pellicle was recorded. *B. subtilis* and its mutants were grown for 48 h and other strains for 24 h. For pellicle restoration assays, surfactin from Sigma and purified surfactin from *B. atrophaeus* 176s (see below) were supplemented in LB medium at different concentrations. These experiments were repeated at least twice with triplicates.

Microscopy and bacillus *root colonization under gnotobiotic conditions*

Wild-type strains and surfactin biosynthesis mutants of *Bacillus* were tested for biofilm formation and root colonization of lettuce roots with or without exogenously added surfactin A or C. Lettuce seeds were surface sterilized in 70% ethanol for 1 min followed by a wash step with sterile distilled water and then treated in 3.5% sodium hypochlorite for 15 min prior rinsing thoroughly with sterile distilled water thrice. Seeds were then placed in plastic magenta boxes filled with 50 g of soil (1:1 v/v potting soil/sand, autoclaved twice) that had been watered with 17 mL of sterile tap water and were grown in greenhouse with photoperiod of 16 h of light and 8 h of dark. Boxes were opened under sterile laminar air flow and 2 mL of bacteria (10^{7} cells/mL in 0.85% NaCl) with or without 40μM surfactin A or C were added two weeks after germination. One week post-inoculation, plants were delicately harvested, and roots were immersed for at least 15 minutes in Eppendorf tubes containing a Syto9 solution (3 μm in PBS pH 7.2) resulting in green fluorescence of bacteria. Samples were then observed under a confocal microscope (Olympus Fluoview

FV1000 with multiline laser FV5-LAMAR-2 HeNe(G)laser FV10-LAHEG230-2). Observations were carried out with 10X, 20X or 40X objectives. X, Y, Z pictures were taken at 405, 488, 549 nm and with 40X objective and then merged (RGB) using Imaris software. Pictures were cropped due to the convolution process in the microscope. Whole pictures were sharpened and the light balance was improved to observe the image details better as seen when samples were observed in the dark under the microscope as described in Glassner et al. (2015). All experiments were repeated on 3 plants. Images presented in this publication represent the average of colonization.

ACKNOWLEDGEMENTS

The authors thank the *Bacillus* Genetic Stock Center (BGSC) for providing bacterial strains. This work was supported by the life science project LS11-014 of the NFB, Niederösterreichische Forschungs- und Bildungsges.m.b.H.

REFERENCES

Aleti, G., Sessitsch, A., and Brader, G. (2015) Genome mining: prediction of lipopeptides and polyketides from *Bacillus* and related Firmicutes. *Comput Struct Biotechnol J* **13:** 192–203.

Angelini, T.E., Roper, M., Kolter, R., Weitz, D.A., and Brenner, M.P. (2009) *Bacillus subtilis* spreads by surfing on waves of surfactant. *Proc Natl Acad Sci USA* **106:** 18109–18113.

Baumgart, F., Kluge, B., Ullrich, C., Vater, J., and Ziessow, D. (1991) Identification of amino acid substitutions in the lipopeptide surfactin using 2D NMR spectroscopy. *Biochem Biophys Res Commun* **177:** 998–1005.

Bais, H.P., Fall, R., and Vivanco, J.M. (2004) Biocontrol of *Bacillus subtilis* against infection of arabidopsis roots by *pseudomonas syringae* is facilitated by biofilm formation and surfactin production. *Plant Physiol* **134:** 307–319.

Beauregard, P.B., Chai, Y., Vlamakis, H., Losick, R., and Kolter, R. (2013) *Bacillus subtilis* biofilm induction by plant polysaccharides. *Proc Natl Acad Sci USA* **110:** 1621–1630.

Brettin, T., Davis, J.J., Disz, T., Edwards, R.A., Gerdes, S., Olsen, G.J., *et al.* (2015) RASTtk: a modular and extensible implementation of the RAST algorithm for building custom annotation pipelines and annotating batches of genomes. *Sci Rep* 8365.

Campisano, A., Ometto, L., Compant, S., Pancher, M., Antonielli, L., Yousaf, S, *et al.* (2014) Interkingdom transfer of the acne causing agent, *Propionibacterium acnes*, from human to grapevine. *Mol Biol Evol* **31:** 1059–1065.

Cawoy, H., Mariutto, M., Henry, G., Fisher, C., Vasilyeva, N., Thonart, P, *et al.* (2015) Plant defense stimulation by natural isolates of Bacillus depends on efficient surfactin Production. *Mol Plant Microbe Interact* **27:** 87–100.

Chen, X.H., Koumoutsi, A., Scholz, R., and Borriss, R. (2009a) More than anticipated—Production of antibiotics and other secondary metabolites by *Bacillus amyloliquefaciens* FZB42. *J Mol Microbiol Biotechnol* **16:** 14–24.

Chen, X.H., Koumoutsi, A., Scholz, R., Schneider, K., Vater, J., Süssmuth, R., *et al.* (2009b) Genome analysis of *Bacillus amyloliquefaciens* FZB42 reveals its potential for biocontrol of plant pathogens. *J Biotechnol* **140:** 27–37.

Chen, Y., Cao, S., Chai, Y., Clardy, J., Kolter, R., Guo, J.H., and Losick, R. (2012) A *Bacillus subtilis* sensor kinase involved in triggering biofilm formation on the roots of tomato plants. *Mol Microbiol* **85:** 418–430.

Chowdhury, S.P., Dietel, K., Rändler, M., Schmid, M., Junge, H., Borriss, R., *et al.* (2013) Effects of *Bacillus amyloliquefaciens* FZB42 on lettuce growth and health under pathogen pressure and its impact on the rhizosphere bacterial community. *PLoS One* **8:**e68818.

Chun, J. and Bae, K.S. (2000) Phylogenetic analysis of *bacillus subtilis* and related taxa based on partial gyrA gene sequences. *Antonie Leeuwenhoek* **78:** 123–127.

Finking, R. and Marahiel, M.A. (2004) Biosynthesis of nonribosomal peptides. *Annu Rev Microbiol* **58:** 453–488.

Ghelardi, E., Salvetti, S., Ceragioli, M., Gueye, S.A., Celandroni, F., and Senesi, S. (2012) Contribution of surfactin and *swrA* to flagellin expression, swimming, and surface motility in *Bacillus subtilis*. *Appl Environ Microbiol* **78:** 6540–6544.

Gibbons, H.S., Broomall, S.M., McNew, L.A., Daligault, H., Chapman, C., and Bruce, D. (2011) Genomic signatures of strain selection and enhancement in *Bacillus atrophaeus* var. Globigii, a historical biowarfare simulant. *PLoS One* **6:** e17836.

Glassner, H., Zchori-Fein, E., Compant, S., Sessitsch, A., Katzir, N., Portnoy, V., and Yaron, S. (2015) Characterization of endophytic bacteria from cucurbit fruits with potential benefits to agriculture in melons (*Cucumis melo* L.). *FEMS Microbiol Ecol* **91:**pii: fiv074.

Heikkinen, S., Toikka, M.M., Karhunen, P.T., and Kilpeläinen, I.A. (2003) Quantitative 2D HSQC (Q-HSQC) via suppression of J-dependence of polarizatrion transfer in NMR spectroscopy: Application to wood lignin. *J Am Chem Soc* **125:** 4362–4367.

Huszcza, E. and Burczyk, B. (2006) Surfactin isoforms from *Bacillus coagulans*. *Z Naturforsch C* **61:** 727–733.

Itokawa, H., Miyashita, T., Morita, H., Takeya, K., Hirano, T., Homma, M., and Oka, K. (1994) Structural and conformational studies of [Ile^7] and [Leu^7] surfactins from *Bacillus subtilis* natto. *Chem Pharm Bull* **42:** 604–607.

Jacques, P. (2011) Surfactin and other lipopeptides from Bacillus spp. Steinbüchel, A. and Soberón-Chávez, G. (eds). Biosurfactants. Microbiology Monographs, 20. Berlin Heidelberg: Springer, pp. 57–91.

Jiang, M., Shao, W., Perego, M., and Hoch, J.A. (2000) Multiple histidine kinases regulate entry into stationary phase and sporulation in *Bacillus subtilis*. *Mol Microbiol* **38:** 535–542.

Kinsinger, R.F., Shirk, M.C., and Fall, R. (2003) Rapid surface motility in *Bacillus subtilis* is dependent on extracellular surfactin and potassium ion. *J Bacteriol* **185:** 5627–5631.

Kinsella, K., Schulthess, C.P., Morris, T.F., and Stuart, J.D. (2009) Rapid quantification of *Bacillus subtilis* antibiotics in the rhizosphere. *Soil Bio Biochem* **41:** 374–379.

Koumoutsi, A., Chen, X.H., Henne, A., Liesegang, H., Hitzeroth, G., Franke, P., *et al.* (2004) Structural and functional characterization of gene clusters directing nonribosomal synthesis of bioactive cyclic lipopeptides in *Bacillus amyloliquefaciens* strain FZB42. *J Bacteriol* **186:** 1084–1096.

LeDeaux, J.R., Yu, N., and Grossman, A.D. (1995) Different roles for KinA, KinB, and KinC in the initiation of sporulation in *Bacillus subtilis*. *J Bacteriol* **177:** 861–863.

Lin, S.C., Minton, M.A., Sharma, M.M., and Georgiu, G. (1994) Structural and immunological characterization of a biosurfactant produced by *Bacillus licheniformis* JF-2. *Appl Environ Microbiol* **60:** 31–38.

López, D., Fischbach, M.A., Chu, F., Losick, R., and Kolter, R. (2008) Structurally diverse natural products that cause potassium leakage trigger multicellularity in *Bacillus subtilis*. *Proc Natl Acad Sci USA* **106:** 280–285.

López, D., Vlamakis, H., Losick, R., and Kolter, R. (2009) Paracrine signaling in a bacterium. *Genes Dev* **23:** 1631–1638.

Luo, C., Zhou, H., Zou, J., Wang, X., Zhang, R., Xiang, Y., and Chen, Z. (2015) Bacillomycin L and surfactin contribute synergistically to the phenotypic features of *Bacillus subtilis* 916 and the biocontrol of rice sheath blight induced by *Rhizoctonia solani*. *Appl Microbiol Biotechnol* **99:** 1897–1910.

Nakano, M.M., Marahiel, M.A., and Zuber, P. (1988) Identification of a genetic locus required for biosynthesis of the lipopeptide antibiotic surfactin in *Bacillus subtilis*. *J Bacteriol* **170:** 5662–5668.

Ongena, M. and Jacques, P. (2008) *Bacillus* lipopeptides: versatile weapons for plant disease biocontrol. *Trends Microbiol* **16:** 115–125.

Ongena, M., Jourdan, E., Adam, A., Paquot, M., Brans, A., Joris, B., *et al*. (2007) Surfactin and fengycin lipopeptides of *Bacillus subtilis* as elicitors of induced systemic resistance in plants. *Environ Microbiol* **9:** 1084–1090.

Oslizlo, A., Stefanic, P., Dogsa, I., and Mandic-Mulec, I. (2014) Private link between signal and response in *Bacillus subtilis* quorum sensing. *Proc Natl Acad Sci USA* **111:** 1586–1591.

Pertot, I., Puopolo, G., Hosni, T., Pedrotti, L., Jourdan, E., and Ongena, M. (2013) Limited impact of abiotic stress on surfactin production in planta and on disease resistance induced by *Bacillus amyloliquefaciens* S499 in tomato and bean. *FEMS Microbiol Ecol* **86:** 505–519.

Peypoux, F., Bonmatin, J.M., and Wallach, J. (1999) Recent trends in the biochemistry of surfactin. *Appl Microbiol Biotechnol* **51:** 553–563.

Razafindralambo, H., Popineau, Y., Deleu, M., Hbid, C., Jacques, P., Thonart, P., and Paquot, M. (1997) Surface-active properties of surfactin/iturin a mixtures produced by *Bacillus subtilis*. *Langmuir* **13:** 6026–6031.

Romero, D., de, Vicente, A., Rakotoaly, R.H., Dufour, S.E., Veening, J.W., Arrebola, E., *et al*. (2007) The iturin and fengycin families of lipopeptides are key factors in antagonism of *Bacillus subtilis* toward *Podosphaera fusca*. *Mol Plant Microbe Interact* **20:** 430–440.

Schneider, J., Mielich-Süss, B., Böhme, R., and Lopez, D. (2015) In vivo characterization of the scaffold activity of flotillin on the membrane kinase KinC of *Bacillus subtilis*. *Microbiology* **161:** 1871–1887.

Shah, I.M. and Dworkin, J. (2009) Microbial interactions: bacteria talk to (some of) their neighbors. *Curr Biol* **19:** 689–691.

Shu, H.Y., Lin, G.H., Wu, Y.C., Tschen, J.S., and Liu, S.T. (2002) Amino acids activated by fengycin synthetase *Fen*E. *Biochem Biophys Res Commun* **292:** 789–793.

Sjöblom, S., Brader, G., Koch, G., and Palva, E.T. (2006) Cooperation of two distinct ExpR regulators controls quorum sensing specificity and virulence in the plant pathogen Erwinia carotovora. *Mol Microbiol* **60:** 1474–1489.

Smyth, T.J.P., Perfumo, A., McClean, S., Marchant, R., and Banat, I.M. (2010) *Isolation and Analysis of Lipopeptides and High Molecular Weight Biosurfactants. Handbook of Hydrocarbon and Lipid Microbiology*. Timmis, K.N. (ed). Berlin Heidelberg: Springer, pp. 3687–3704.

Stachelhaus, T., Mootz, H.D., and Marahiel, M.A. (1999) The specificity-conferring code of adenylation domains in nonribosomal peptide synthetases. *Chem Biol* **6:** 493–505.

Stein, T. (2005) *Bacillus subtilis* antibiotics: structures, syntheses and specific functions. *Mol Microbiol* **56:** 845–857.

Tang, J.S., Gao, H., Hong, K., Yu, Y., Jiang, M.M., Lin, H.P, *et al*. (2007) Complete assignments of ^{1}H and ^{13}C NMR spectral data of nine surfactin isomers. *Magn Reson Chem* **45:** 792–796.

Troskie, A.M., Vlok, N.M., and Rautenbach, M. (2012) A novel 96-well gel-based assay for determining antifungal activity against filamentous fungi. *J Microbiol Methods* **91:** 551–558.

Vater, J., Kablitz, B., Wilde, C., Franke, P., Mehta, N., and Cameotra, S.S. (2002) Matrix-assisted laser desorption ionization-time of flight mass spectrometry of lipopeptide biosurfactants in whole cells and culture filtrates of *Bacillus subtilis* C-1 isolated from petroleum sludge. *Appl Environ Microbiol* **68:** 6210–6219.

Vlamakis, H., Chai, Y., Beauregard, P., Losick, R., and Kolter, R. (2013) Sticking together: Building a biofilm the *Bacillus subtilis* way. *Nat Rev Microbiol* **11:** 157–168.

Vollenbroich, D., Mehta, N., Zuber, P., Vater, J., and Kamp, R.M. (1994) Analysis of surfactin synthetase subunits in *srfA* mutants of *Bacillus subtilis* 0KB105. *J Bacteriol* **176:** 395–400.

Weisburg, W.G., Barns, S.M., Pelletier, D.A., and Lane, D.J. (1991) 16S ribosomal DNA amplification for phylogenetic study. *J Bacteriol* **173:** 697–703.

Weber, T., Blin, K., Duddela, S., Krug, D., Kim, H.U., Bruccoleri, R, *et al*. (2015) antiSMASH 3.0-a comprehensive resource for the genome mining of biosynthetic gene clusters. *Nucleic Acids Res* **43:** 237–243.

Zhang, X., Li, B., Wang, Y., Guo, Q., Lu, X., Li, S., and Ma, P. (2013) Lipopeptides, a novel protein, and volatile compounds contribute to the antifungal activity of the biocontrol agent *Bacillus atrophaeus* CAB-1. *Appl Microbiol Biotechnol* **97:** 9525–9534.

Supporting information

Additional Supporting Information may be found in the online version of this article at the publisher's web-site:

Table S1. ^{1}H (400 MHz) and ^{13}C NMR spectral data for surfactin C with C-14 fatty acid in CD_3CN

Table S2. Primers used in this study

Table S3. An overview of surfactin variants. Differing Surfactin monomers at positions 1 and 7 are underlined.

Fig. S1. In vitro biocontrol assays. (A) Evaluation of antifungal activity of *B. atrophaeus* 176s against (a) *Fusarium oxysporum* ACC01, (b) *Botrytis cinerea* 501-E, (c) *Rhizoctonia solani* FT1510, (d) *Sclerotinia sclerotiorum* MA5092. (B) Antifungal activity of *B. atrophaeus* and *B. subtilis* strains against *F. oxysporum*. The WT strains of *B. atrophaeus* 176s and 1942, and the mutant ATCC9372 affected in surfactin production are shown on the left. *B. subtilis* OKB 105 and its

surfactin biosynthesis mutant *B. subtilis* OKB 120 on the right.

Fig. S2. Hemolytic activities on 10% sheep blood agar plate recorded 48 h after application. A) Hemolytic activities of WT strains *B. atrophaeus* 176s and 1942, *B. amyloliquefaciens* FZB42 and *B. subtilis* OKB 105 and lacking activity of the corresponding mutants *B. atrophaeus* 9372 and *B. subtilis* OKB 120. B) Crude cLP (500 μg) extracted from *B. atrophaeus* 176s with pronounced hemolytic activity. Surfactin from Sigma was used as a positive control.

Fig. S3. Evaluation of SPE purified cLP fractions through hemolytic and activity against *Fusarium oxysporum*. Fraction C displayed strong hemolytic activity but no antifungal activity, representing surfactins, while fraction A displayed strong antifungal activity but no hemolytic activity, indicating fengycins.

Fig. S4. Surfactin C structure shown with the dominant C-14 fatty acid chain as determined by NMR analysis. Numbering of amino acids in accordance with Table S2.

Fig. S5. Suppression of *R. solani* infection in plants pretreated with *B. atrophaeus* 176s and *B. amyloliquefaciens* FZB42. (A) Representative pictures one week after inoculation with *R. solani* strains and plants as indicated in the figure. (B) Protective effect of *B. atrophaeus* 176s and *B. amyloliquefaciens* FZB42 compared to untreated control one week after fungal inoculation. The health status of plants is indicated as H (healthy), S (symptomatic), and D (dead) 7 days after infection. The data shown here represent two independent experiments carried out in the greenhouse.

Fig. S6. Biocontrol activity in lettuce plants pretreated with WT and their surfactin biosynthetic mutants of *B. subtilis* and *B. atrophaeus* in response to *R. solani* infection under greenhouse conditions. (A) Representative pictures two week after fungal inoculation with *R. solani* as shown in the figure. (B) Protective effect of WT strains *B. atrophaeus* 1942 and *B. subtilis* OKB105 compared to their mutants impaired in surfactin biosynthesis two week after fungal inoculation. The health status of plants is indicated as H (healthy), S (symptomatic), and D (dead) 14 days post infection. The data shown here represent two independent experiments carried out in the greenhouse.

Fig. S7. Influence of surfactin variants on swarming motility. WT and mutant strains affected in surfactin biosynthesis of *B. amyloliquefaciens*, *B. subtilis* and *B. atrophaeus* were grown on LB agar plates supplemented with surfactin A from *B. subtilis* (bought from Sigma) and purified surfactin C from *B. atrophaeus* at different concentrations for 72 h. Strains are indicated on the left and concentrations of exogenous surfactins are indicated on the top. In the control lane and WT (on top) no surfactin has been added. Bars show diameter of colonies after 72 h indicative for swarming capacity. Bars signify standard deviations of three replicates.

Final conclusions

Bacterial and fungal plant diseases have been a severe concern for the food production. Although chemical fertilizers and pesticides undoubtedly suppressed the infections and boosted yields for most field crops, toxic chemical residues are a major concern for human health. Therefore, alternative approaches for ecologically compatible safe crop management practices are rapidly gaining importance. Plant beneficial microbes and biocontrol agents have emerged as promising alternatives to the chemical inputs. Representatives of the genera *Bacillus* and *Paenibacillus* encompass important plant beneficial strains noted for excellent colonization ability and plant growth promoting activities, as well as produce diverse bioactive compounds, for instance, lipopeptides and polyketides for the establishment of plant-microbe interaction and biocontrol of fungal plant pathogens. Cyclic lipopeptides especially surfactins are crucial for biocontrol activity and trigger signaling cascade for biofilm formation and consequently root colonization. Although surfactin variants differing in peptide moiety have been discovered in several plant-associated bacilli, how these subtle structural differences are encoded in a group specific way and how they might impact the bacterial interactions has not been explored so far.

Often NRPS and PKS type I assembly lines follow co-linearity rule in which the incorporation of amino acids (for NRPS) and carboxylic acids (for PKS) for the biosynthesis and final assembly of the structure is usually the same as the order of catalytic modules in the genome. This structural feature and insight into the organization of the catalytic modules and domains within natural-product assembly lines often enables prediction of the final product based on the genomic sequence and the otherwise cumbersome task to isolate often unstable polyketides and deduce their structure can be streamlined. Although next-generation sequencing technology has largely contributed to the ample availability of the whole genome sequences and bioinformatics methods and tools have revolutionized the pipeline for the prediction of secondary metabolite biosynthetic gene clusters, little is accomplished to explore the sequence wealth to identify potential novel lipopeptides and polyketides in these genomes and to predict uncharacterized secondary metabolites.

In this context, Chapter 1 and chapter 2 of the PhD thesis evaluate the secondary metabolite potential by genome mining the published sequence repository of Bacillales and a draft genome sequence of *Paenibacillus polymyxa*. In Chapter 3 we demonstrate the species-specific signal role for surfactin variants in biofilm induction and root colonization.

Genome mining: Prediction of lipopeptides and polyketides from *Bacillus* and related Firmicutes

The first phase of the PhD study was aimed to review the current structural information and the underlying biosynthetic gene clusters of known type I polyketides and lipopeptides synthesized by well-known *B. amyloliquefaciens*, *B. subtilis* and *P. polymyxa* from Bacillales and apply this knowledge to mine the published genomes using secondary metabolite gene cluster prediction tools to evaluate novel chemical space encoded in the distinct taxonomic groups of Bacillales within the phylum Firmicutes. Such a genome mining study may provide access to the previously uncharacterized secondary metabolite biosynthetic pathways. The huge sequence repository encompassing both bacterial genomes and metagenomes enable such genome-mining study to reveal the potential for novel structures of secondary metabolites. The major bottlenecks in the discovery of natural products to isolate milligram amounts of lipopeptides and in particular various unstable polyketides, and characterize their structure by NMR techniques can be streamlined by genome mining in combination with

mass spectrometry methods. Moreover, a recent genome mining showed that 31% of the Firmicutes are estimated to contain NRPS and PKS, of which 70% were NRPS and 30% were hybrid NRPS/PKS or PKS. *Bacillus* and *Paenibacillus* genera from the order Bacillales within Firmicutes have been reported to encompass higher number of these biosynthetic gene clusters, and predicted to encode structurally diverse lipopeptides and polyketides, however, the secondary metabolite potential encoded in the genomes of many other members of Bacillales are still largely unexplored.

Three polyene polyketides including bacillane, difficidin and macrolactin represent the most diversity of polyketides synthesized by well-known plant growth promoting and biocontrol agents from *Bacillus*, while paenilamicin and paenimacrolidin represent the diversity from the genus *Paenibacillus*. Majority of the lipopeptides encoded by *Paenibacillus* include cyclic cationic lipopeptides comprising polymyxins, polypeptins, octapeptins and paenibacterins, all-encompassing a non-proteogenic amino acid 2,4-diaminobutyric acid (dab) in the peptide backbone constituting to the overall positive charge of these lipopeptides. Non-cationic structures encoded by *Bacillus* and *Paenibacillus* spp. mostly include surfactin, iturin, fengycin, and fusaricidin. While linear cationic class from paenibacilli include saltavalin, jolipeptin, tridecaptins and bacilli include cerexins.

A systematic procedure was adapted for the secondary metabolite gene cluster prediction and analysis to avoid ambiguity in the discovery of the predicted monomers. The whole genome sequences derived from both characterized and uncharacterized strains from selected members of the Bacillales were fed to bioinformatic tools including antiSMASH, NaPDos, and NRPS/PKS substrate predictor. A more detailed analysis of the predicted gene clusters was performed by BLAST search enabled by antiSMASH pipeline to identify potential homologs found in other bacterial genomes for the predicted gene cluster. The predicted catalytic domains within the gene cluster were further evaluated by NaPDoS, and to narrow the ambiguity of predicted monomers especially substrate specificity for A domain of NRPS and AT domain of PKS were further assessed by NRPS/PKS substrate predictor.

Genome mining identified both known and several novel secondary metabolite biosynthetic gene clusters predicted to encode lipopeptides and polyketides. *Paenibacillus* strains contained a high potential for uncharacterized lipopeptide of heptapeptides, nonapeptides, decapeptides and tridecaptins, and their variants with altered monomer composition in the peptide backbone. Many *Paenibacillus* strains encompass known lipopeptide gene clusters predicted to encode polymyxins and fusaricidins and variants as well. Especially the modular organization of the predicted heptapeptides from strains of *P. polymyxa*, and *P. mucilaginosus* and *P. fonticola* were similar to iturin and shared up to 46% identity with iturin from *B. amyloliquefaciens*, however the predicted monomers of the peptide backbone were completely different when compared to known iturin class suggesting a novel class of iturins. Moreover, we predicted previously neglected nonapeptide and its variants in the genomes of *P. mucilaginosus, P. elgii* and *P. terrae*. A novel predicted paenibacterin variant is encoded by *P. taiwanensis* and *P. alvei*. Many novel PKS with modular organization similar to bacillaene were predicted in the genomes of *P. polymyxa*, *P. pini* and *Brevibacillus*. Paenimacrolidin synthase like genes were also found in *P. durus, P. elgii* and *P. ehimensis*.

The well described lipopeptide families, surfactin, iturin and fengycin were not only encoded in well-known *Bacillus* species such as *B. amyloliquefaciens, B. subtilis, B. licheniformis, B. atrophaeus*, *B. mojavensis* and *B. pumilus* but surprisingly also encoded in other neglected strains and less studied species including *Salinibacillus aidingensis*. Intriguingly, the lipopeptide kurstakin and its variants are encoded exclusively in the genomes of *B.*

thuringiensis strains. Regarding the PKs anticipated from *Bacillus*, several strains were predicted to encode bacillaene, macrolactin and difficidin and surprisingly, we also found variants that have not been described so far even in well-studied strains of *B. amyloliquefaciens* and *B. subtilis*. Other *Bacillus* species, in particular *B. atrophaeus*, *B. mojavensis*, *Brevibacillus brevis*, *Ornithinibacillus* and *Salinibacillus* are predicted to encode clearly a distinct polyketide potential.

In summary, a total of 160 published genomes of Bacillales analyzed, 57% (91 isolates) genomes harbor gene clusters predicted to encode either lipopeptides, type I polyketides or both. Intriguingly, a higher percentage of 85% of the 40 isolates derived from rhizosphere and endophytes predicted to contain at least one of these biosynthetic gene clusters. However, given the fact that the origin of up to a third of the isolates is ambiguous, it is difficult to foresee, if more frequency of these secondary metabolite clusters in plant-associated strains will also be seen when genomic data will be expanded. Important findings of this study suggests that a large fraction of lipopeptides and type I polyketides predicted in the current study are previously unexplored and found in majority of plant-associated *Bacillus* and *Paenibacillus* genera. While many genera from other environments harbor few of such gene clusters indicating the role of these lipopeptides and polyketides in plant-associated niches.

The draft genome sequence of *Paenibacillus polymyxa* strain CCI-25 encompasses high potential for secondary metabolite production

The second part of the study investigated the secondary metabolite biosynthetic capacity and biocontrol potential of *P. polymyxa* strain CCI-25 isolated from vermicompost. Results show that both CCI-25 colonies and lipopeptide and polyketide crude isolations exhibited strong antibacterial activity against *Escherichia coli* and antifungal activities towards *Fusarium oxysporum*, *Botrytis cinerea*, and *Rhizoctonia solani*, *in vitro*. In order to understand the genetic basis for secondary metabolite biosynthesis, whole genomic DNA was subjected for next-generation sequencing.

The genome assembly resulted in 117 contigs larger than 1,000 bp, with an N_{50} contig length of 95,765 bp. The size of the draft genome is 5.61 Mb with a G+C content of 44.95%. We identified 40 highly conserved single-copy marker genes in the genome assembly suggesting completeness of the genome and excluded contaminant sequences. Genomic BLAST showed similarities to plant growth promoting *P. polymyxa* CR1. We identified 5,146 genes, 4,953 coding sequences (CDSs), 15 complete 5S rRNAs, 30 partial 16S rRNAs, 37 partial 23S rRNAs (for 15 putative rRNA operons), 107 tRNAs, 4 noncoding RNAs (ncRNAs), and 241 pseudogenes based on the NCBI Prokaryotic Genome Annotation Pipeline (PGAP).

Secondary metabolite biosynthetic gene sequences in the genome were predicted using antiSMASH pipeline. Intriguingly, the CCI-25 draft genome is predicted to encompass non-ribosomal peptide synthetase encoding gene clusters with sequence similarities to published genes. However, the prediction includes the encoding of a fusaricidin C, and tridecaptin variant and a polymyxin variant with amino acid substitutions in the peptide backbone. In addition, an iturin- and paenilarvin-like compound with altered monomer composition shared less than 58% and 40% sequence identity with *Bacillus amyloliquefaciens* FZB42 and *P. larvae* DSM 25430 respectively have been predicted. Also CCI-25 contains a lantibiotic gene with similarities to paenicidin A and a polyketide synthase predicted to encode different number of acyl carrier domains with sequence identities of up to 61% with bacillaene synthase from *B. amyloliquefaciens* FZB42 and 87% with polyketide synthase derived from *P. polymyxa* M-1. Given the fact that 6.6% of the total genome is devoted to secondary

metabolite production, CCI-25 has high potential to be exploited for medical or agricultural applications.

Surfactin variants mediate species-specific biofilm formation and root colonization in *Bacillus*

The biocontrol potential of *Bacillus atrophaeus* strain 176s and its capacity to produce cyclic lipopeptides, in particular, the underlying molecular mechanism for the synthesis of surfactin variant structures and their signaling role in biofilm formation and root colonization is addressed in the third phase of the study. In *B. atrophaeus* 176s, isolated from moss plant, we reported the biosynthesis of three cyclic lipopeptide families, surfactin, iturin and fengycin. These lipopeptide antibiotics likely play crucial roles in protective activities similar to the well-known plant growth promoting *B. amyloliquefaciens* FZB42. Amphiphilic surfactins may display strong hemolytic activity, while the strong antifungal *in vitro* activity shown by *B. atrophaeus* 176s may be due to the production of fengycins. Co-production of the three cyclic lipopeptides may synergistically enhance their individual activities and are known to elicit systemic resistance in the plant. To our knowledge, this is the first study demonstrating the co-production of three different lipopeptides families and especially biosynthesis of surfactin C in *B. atrophaeus*. Taken together, these findings imply *B. atrophaeus* potential in biocontrol application.

In *B. atrophaeus* we identified for the first time the variant surfactin C, which differs from surfactin A synthesized by *B. subtilis* and *B. amyloliquefaciens* by substitution of an isoleucine instead of leucine at position 7 in the heptapeptide backbone. The assembly of the complete surfactin gene cluster showed 97% sequence identity with *B. atrophaeus* 1942 and 79% with *B. amyloliquefaciens* FZB42 on amino acid level. The non-ribosomal code encoded by eight amino acids within the adenylation domain of the NRPS enables the prediction of A-domain specificity for certain amino acids based on the primary sequence. The non-ribosomal code encoding the first six amino acids of the surfactin synthetase was identical in both *B. atrophaeus* 176s and *B. amyloliquefaciens* FZB42, but striking dissimilarities are encompassed in the A domain of *srfC* within the module 7. Intriguingly, Ala to Gly and a Cys to Val substitution in the amino acid sequence composition of A domain is responsible for the selection and incorporation of an Ile instead of a Leu in surfactin synthesis. These results are in accordance with the previously reported module 7 of lichenysin operon from *B. licheniformis*. However, Lys to Glu substitution in the A domain of *srfA* within the first module is responsible for recruiting Gln instead of Glu in lychenysin synthesis. Our previous genome mining findings confirmed that the well-known members of *B. subtilis* and *B. amyloliquefaciens* encode Leu at position 7, while *B. atrophaeus* encode Ile and *B. licheniformis* strains encode Gln and Ile. at position 1 and 7, respectively. These findings imply that the surfactin synthetase precisely synthesizes either surfactin A or C as these subtle specificities are genetically fixed in NRPS in the form of non-ribosomal code.

Surfactin variants are distributed in a species-specific manner suggesting that these surfactin variants may play a role in species-specific signaling potentially resulting in different ecological behavior. Surfactin as a signaling molecule mediates microbial communication and acts as a quorum sensing molecule responsible for the activation of a membrane-associated sensory receptor histidine kinase (KinC), which in turn believed to trigger a pathway for biofilm formation and consequently root colonization, and the strong biosurfactant property of surfactin may play an intricate role in bacterial swarming to nutrient-rich environments and contribute to the expansion of the biofilm. Our findings suggested that the surfactins effected the motility and rescued the phenotype of swarming motility in *Bacillus* deficient in surfactin

synthesis, especially *B. atrophaeus*. These results indicated no surfactin A or C specific effect on swarming motility, this effect is perhaps associated with the amphiphilic nature of the surfactin molecules or a nonspecific signaling by the variants. Intriguingly, biofilm formation and consequent root colonialization of bacilli, however, was strikingly restored to varying degrees by different surfactin variants. Surfactin biosynthesis mutants of *B. atrophaeus* and *B. subtilis* with impaired pellicle formation and root colonialization, effectively restored phenotype of biofilm formation only when supplemented with their native surfactins. Substantial differences in *Bacillus* spp. potential to respond to exogenously supplemented surfactins in biofilm development clearly point to the variations in the signal perception of surfactin A and C depending on the genetic background of bacilli. Also suggests that even subtle structural differences within moiety of surfactin A and C may influence the three dimensional structure of surfactin and consequently on its ability to act as a signal component and their role in species-specific communication.

Future perspectives

The availability of fully sequenced genomes derived from uncultured and less studied bacteria in public databases is constantly increasing due to advanced and better affordable next-generation sequencing techniques. Such sequence wealth can be a rich source for the discovery of novel secondary metabolites and can be explored by genome mining to predict novel secondary metabolite biosynthetic pathways. Our genome mining findings suggest that a large number of predicted non-ribosomal and polyketide products from the metabolomes of Bacillales are uncharacterized and their functions with regards to plant association still remains unknown and other so far neglected Bacillales with no published genomic data still remain untapped.

In the current study using genome mining tools we identified a novel NRPS gene cluster from *Paenibacillus* spp. that has a modular architecture similar to the iturin, however, the corresponding product yet remains to be analyzed and characterized. Further investigation of transcriptional activity of the novel gene cluster and other associated NRPS and PKS genes, especially in co-cultures with bacteria, fungi and their metabolites to enable biosynthesis of antifungal and antibacterial compounds may elucidate not only the role of these yet uncharacterized gene clusters, but also their potential bioactive products in the context of biocontrol of plant pathogens.

Surfactins are believed to cause selective potassium leakage trigger and known to induce biofilm development upon membrane disturbance and perception of membrane-associated sensory receptor histidine kinase KinC. Recent studies suggest that the membrane-associated chaperone like protein flotillin (FloT) embedded within the membrane micro-domains of *Bacillus*, seem to associate with KinC to promote effective binding of specific-signaling peptides. Host related factors might also modulate biofilm formation and related histidine kinases, in particular KinD, seem to be involved in sensing plant root exudates as well. It remains to be seen whether Kin histidine kinases signaling or so far uncharacterized membrane components are involved in species-specific perception of surfactin variants in induction of biofilm development and root colonization.

Root colonization by antagonistic bacteria can be considered as a prerequisite for effective biological control, and the instability of bacterial colonization for different environmental impacts has emphasized the need to improve the colonization potential. *Bacillus* root colonization is mostly determined by chemotaxis, swarming motility and robust biofilm formation, and surfactins are believed to play a key role in this respect. Therefore, another outstanding question remains whether exogenous supplementation of cognate surfactins can boost the plant colonization potential of a specific strain in terms of biocontrol application.

Finally, detection of NRPS and PKS gene clusters in the genome alone will not be sufficient but identification of the well-characterized homologs by phylogenetic analysis and the bioinformatics prediction of the related products, and their functionality with regard to biocontrol of plant diseases remains particularly tricky.

Curriculum Vitae
Gajender Aleti

Personal data
Date of Birth: 23rd August 1982
Nationality: Indian
Marital status: married
✉ gajender.aleti@gmail.com
☎ +436509836883
Address: Linkewienzeile 158/12, 1060, Vienna, Austria

Profile

- Expertise in the field of molecular microbial ecology including soil microbiology, biodiversity, biocontrol, metagenomics and secondary metabolite biosynthesis with an emphasis on plant growth promoting bacteria
- Possess a strong background in molecular biology and bioinformatics with competencies on next-generation sequencing and analysis
- Presentation and communication skills, and teaching experience

Education and Qualifications

Dec 2011-present	**PhD Microbial Ecology,** AIT Austrian Institute of Technology in association with the University of Natural Resources and Life Sciences, Tulln, Austria PhD title: Secondary metabolite biosynthetic gene diversity – Bacillales non-ribosomal peptides and polyketides in plant-microbe interactions Supervisor: PD Dl Dr. Angela Sessitsch
2007-2008	**MSc Bioinformatics**, University of Abertay Dundee, Dundee, Scotland. UK 6 months industrial research project at CXR Biosciences, UK. Thesis title: Designing a Chromatin Immunoprecipitations on chip microarray to test effects of phthalate chemicals on fetal testis of rat
2003-2005	**MSc Biotechnology**, Government City College, Hyderabad, INDIA. Qualified with distinction 5 months research project at National Institute of Nutrition, India Thesis title: Identification of molecular markers for the diagnosis of diabetes and tuberculosis.
1998-2002	**BSc Biology** (Botany, Zoology and Chemistry) Post Graduate College, Secunderabad, INDIA

Employment History

Feb 2011-Dec 2011	**Research Assistant** Institute for Population Genetics, Veterinary Medicine University, 1210, Vienna, Austria Assisting in Drosophila cultures, next-generation sequencing, DNA-microinjection and molecular biology techniques.
July 2009- Jan 2011	**Research Technician** School of Biosciences, Cardiff University, Cardiff, UK Worked on a Wellcome Trust funded project to understand the regulation of gene expression in testes of the fruit fly *Drosophila*.
June 2006-Aug 2007	**Lecturer in Biotechnology** Villamarie Degree College, India Lecturing and conducting practical courses.

Experimental skills

Microarrays	ChIP-on-chip microarrays and data analysis
Molecular biology	DNA/RNA/cDNA extraction, PCRs (qPCR, RT-PCR, gradient PCR), Degenerate primer design, bacterial transformation, molecular cloning and phylogenetic marker gene analysis (16S rRNA, gyrB…)
Microbiology	Isolation, identification and characterization of interesting bacteria, microbial diversity studies, secondary metabolite production and isolation of compounds, construction of metagenomic fosmid libraries and functional screening
Analytical chemistry	Knowledge of LC-MS and NMR techniques
Microscopy	DIC/Nomarski optics, fluorescence and confocal microscopy

Bioinformatics and computer skills

- NGS sequencing and analysis: QIIME, UPARSE…
- Assembly, annotation and analysis of prokaryotic genomes
- Metagenome analysis
- Working in Linux shell and R environment
- Tools: masurca, velvet, SPAdes, Amphora, MAUVE, conspred, Amphora, Inparanoid, MEGAN, PhyMMBL, GeneMarkS, Glimmer, Prodigal, BWA, TopHat, NoISeq, …

Review

1. **Aleti G**, Sessitsch A, Brader G, Genome mining: prediction of lipopeptides and polyketides from plant associated *Bacillus* and related Firmicutes. Computational and Structural Biotechnology Journal 2015;13: 192–203.

Research articles

2. **Aleti G**, Antonielli L, Corretto E, Nikolić B, Sessitsch A, Brader G, Draft genome sequence of *Paenibacillus polymyxa* CCI-25 encompasses high potential for secondary metabolite production. Genome Announcements 2016.

3. **Aleti G**, Lehner S, Bacher M, PleskoM, Branislav N, Sessitsch A, Schuhmacher R, Brader G, Surfactins mediate species-specific signalling in *Bacillus*. Environmental Microbiology 2016.

4. **Aleti G**, Branislav N, Brader G, Sessitsch A, Secondary metabolite genes encoded by potato rhizosphere microbiomes in the Andean highlands are diverse and depend on site and vegetation stage (manuscript in preparation).

5. **Aleti G**, Sessitsch A, Brader G, Secondary metabolite biosynthetic gene potential in shoot endosphere and rhizosphere bacteria (manuscript in preparation).

6. Doggett K, Jiang J, **Aleti G**, White-Cooper H, Wake-up-call, a lin-52 paralogue, and always early, a lin-9 homologue physically interact, but have opposing functions in regulating testis- specific gene expression. Developmental Biology 2011; 355:381–393.