Identification and classification of known and putative antimicrobial compounds produced by a wide variety of Bacillales species

Xin Zhao 1,2 and Oscar P. Kuipers 1*

Abstract

Background: Gram-positive bacteria of the Bacillales are important producers of antimicrobial compounds that might be utilized for medical, food or agricultural applications. Thanks to the wide availability of whole genome sequence data and the development of specific genome mining tools, novel antimicrobial compounds, either ribosomally- or non-ribosomally produced, of various Bacillales species can be predicted and classified. Here, we provide a classification scheme of known and putative antimicrobial compounds in the specific context of Bacillales species.

Results: We identify and describe known and putative bacteriocins, non-ribosomally synthesized peptides (NRPs), polyketides (PKs) and other antimicrobials from 328 whole-genome sequenced strains of 57 species of Bacillales by using web based genome-mining prediction tools. We provide a classification scheme for these bacteriocins, update the findings of NRPs and PKs and investigate their characteristics and suitability for biocontrol by describing per class their genetic organization and structure. Moreover, we highlight the potential of several known and novel antimicrobials from various species of Bacillales.

Conclusions: Our extended classification of antimicrobial compounds demonstrates that Bacillales provide a rich source of novel antimicrobials that can now readily be tapped experimentally, since many new gene clusters are identified.

Keywords: Antimicrobials, Bacillales, Bacillus, Genome-mining, Lanthipeptides, Sactipeptides, Thiopeptides, NRPs, PKs

Background

Most of the species of the genus Bacillus and related Firmicutes are Gram-positive, aerobic endospore-forming and rod-shaped bacteria, which are found in diverse environments such as soil and clay, rocks, dust, aquatic environments, on vegetation, in food and in the gastrointestinal tracts of various insects and animals [1]. Antimicrobial compounds have been used for a variety of purposes, such as delaying spoilage by plant pathogens in agriculture and extending product shelf life in the food industry [2, 3]. In particular, Bacillus strains are known to produce a wide variety of biocontrol metabolites, including the ribosomally synthesized antimicrobial peptides (bacteriocins) [4], as well as non-ribosomally synthesized peptides (NRPs) and polyketides (PKs) [5].

The discovery of biosynthetic gene clusters of antimicrobial compounds by genome mining is a rewarding task, because this methodology can lead to the identification and subsequent isolation of novel molecules of pharmacological and biotechnological interest [6]. Various powerful tools with broad databases have been created for the automated screening of bacteriocin gene clusters. BAGEL3 (http://bagel.molgenrug.nl/) is a versatile fast genome-mining tool valid not only for modified- and non-modified bacteriocins, but also for non-bacteriocidal ribosomally produced and posttranslationally modified peptides (RiPPs) [7]. The corresponding database [7] contains all the records belonging to one of the three classes of proteins being core to BAGEL3: Class I contains RiPPs of less than 10 kDa, which currently is divided into more than 12 supported subclasses; Class II...
contains unmodified peptides not fitting the criteria of the first database; Class III contains antimicrobial proteins larger than 10 kDa. BAGEL3 uses DNA nucleotide sequences in FASTA format as input; multiple sequence entries per file are allowed. The input DNA sequences are analyzed in parallel via two different approaches; one is the context of bacteriocin- or RiPP gene-based mining, the other is precursor (structural gene)-based mining directly by Glimmer, which increases the success rate and lowers the need for manual evaluation of results [7]. The output is visualized in an html page, by a table of putative bacteriocins or modified peptides classified into the detailed bacteriocin class found in the mining sequence; graphics of gene clusters; annotation of each ORF in the context; as well as detailed information of putative bacteriocins, such as BLAST hits in the bacteriocin database, or the pI (Isoelectric point) value. A detailed prediction of the gene clusters of NRPs, PKs and other antimicrobials is provided by antiSMASH (http://antismash.secondarymetabolites.org), a web server and stand-alone tool for the automatic genomic identification and analysis of biosynthetic gene clusters [8–10]. A database of classes specific for many types [10] of biosynthesis signature genes is constructed by Hidden Markov Models (pHMMs) covering a wide range of known or putative secondary metabolite compounds. The antiSMASH web server allows uploading of sequence files of not only a variety of types (FASTA, GBK, or EMBL files), but also GenBank/RefSeq accession numbers. Gene clusters are first predicted and identified by Glimmer and pHMMs, respectively. Subsequently, several downstream analyses can be performed by different modules: NRPS/PKS domain analysis and annotation; prediction of the core chemical structure of PKSs and NRPSs; ClusterBlast gene cluster comparative analysis; active enzyme site analysis; and secondary metabolism Clusters of Orthologous Groups (smCOG) analysis. Moreover, the ClusterFinder algorithm is used to detect putative gene clusters of unknown types. Finally, an html output is generated and putative gene clusters are listed in a Table [8–10]. Further details including gene cluster description, annotation, percentage of gene homology with known gene clusters or published genome sequences; genomic loci for this biosynthetic pathway are shown by clicking on the related words. Biochemical properties of the putative compounds are also predicted, especially chemical structures of NRPs and PKs. Results, stored in an EMBL/XLS/GenBank/BiosynML file, can be downloaded for additional analysis.

Although a description of Bacillus subtilis antimicrobials has been made before (excellent review of Stein [11]), we aim to give an updated overview and classification of bacteriocins covering various species of Bacillales, as well as NRPs and PKs, by genome mining of 57 different species within 328 whole-genome sequenced strains of Bacillales reported before March 2016 (Table 1, Additional file 1: Table S1 and Fig. 1). We also highlight examples of each class by describing the genetics and structure, with a keen eye on biocontrol properties and applications. Within the genus Bacillus, B. subtilis, B. amyloliquefaciens, B. licheniformis, B. cereus and B. thuringiensis are the best studied species for antimicrobials production [12]. Genome mining and subsequent analyses and classification of antimicrobials of other less explored Bacillales, including Paenibacillus, Brevibacillus, Alicyclobacillus, Anoxybacillus, Lysinibacillus and Geobacillus will be also included in this analysis, revealing interesting new features and distributions.

**Results**

**Classification of antimicrobial peptides encountered in Bacillales**

The main classification scheme for ribosomally synthesized antimicrobial peptides currently available is that of the lactic acid bacteria (LAB) bacteriocins [13], which was recently reviewed and revised by Alvarez-Sieiro et al. [14]. The main classification scheme for RiPPs (Class I) was provided by the paper of Arnison et al. [15]. Although some bacteriocins produced by Bacillus are similar to those of LABs, the Bacillus antimicrobial compound classification system now is lagging behind that of LAB classifications. Conveniently, BAGEL3 can be used for mining bacteriocin gene clusters, some of which were not identified before. Moreover, some cryptic gene clusters of bacteriocins were identified that have not been isolated yet from wild type microorganisms. In this study, we identified 583 putative bacteriocin gene clusters from 328 strains of 57 species of Bacillales (Table 1), and these gene clusters were further classified into three classes harboring 46 types of bacteriocins covering 50 species of Bacillales (Additional file 2: Table S2) according to their gene organization and the homologies of their structural and biosynthetic genes. In addition to the published bacteriocins, many novel putative bacteriocin gene clusters were discovered. Combining this with the genome mining results of antiSMASH, we also address the non-ribosomally synthesized and polyketide synthesized antimicrobial compounds. In total 1231 putative non-ribosomal antimicrobial gene clusters were detected and subgrouped into 23 types of NRPs, five types of PKs and three types of NRPS/PKS hybrid synthesized compounds distributed over 49 species of Bacillales (Additional file 3: Table S3). In the following sections, we will describe the various classes of ribosomally synthesized peptides, NRPs, PKs and other antimicrobials present in Bacillales and indicate their presence in the various genomes.

**Ribosomally synthesized antimicrobial peptides**

The classification system used in this paper for Bacillus ribosomally synthesized antimicrobial peptides (Table 1) comprises the major Class I: small RiPPs (based on Arnison...
| Genera                          | Lanthipeptides type I | Lanthipeptides type II | Head to tail cyclized peptides | Sacti-peptides | Glyco-cins | Lasso peptides | LAPs | Thio-peptides | TOTAL |
|--------------------------------|-----------------------|------------------------|--------------------------------|----------------|------------|----------------|------|---------------|-------|
| Bacillus subtilis (39)         | 6                     | 53                     | 15                             | 1              | 1          | 8              | 84   | 168           | 66    |
| Bacillus thuringiensis (46)    | 4                     | 8                      | 6                              | 5              | 3          | 4              | 24   | 31            | 99    |
| Bacillus anthracis (39)        |                       |                        |                                | 25             |            |                |      | 25            | 117   |
| Bacillus cereus (55)           | 1                     | 16                     | 11                             | 4              | 1          | 9              | 34   | 22            | 4     |
| Bacillus amyloliquefaciens (13)|                      |                        |                                |                |            |                |      | 105           | 144   |
| Bacillus licheniformis (3)     | 1                     |                        |                                |                |            |                |      | 12            | 30    |
| Bacillus coagulans (5)         |                       |                        |                                |                |            |                |      | 4             | 6     |
| Bacillus megaterium (5)        | 1                     |                        |                                |                |            |                |      | 2             | 5     |
| Bacillus pumilus (8)           |                       |                        |                                |                |            |                |      | 14            | 17    |
| Bacillus atrophaeus (4)        |                       |                        |                                |                |            |                |      | 4             | 8     |
| Bacillus weihenstephanensis (2)|                       |                        |                                |                |            |                |      | 2             | 8     |
| Bacillus mycoides (5)          | 2                     | 4                      |                                |                |            |                |      | 5             | 4     |
| Bacillus cytotoxicus (1)       |                       |                        |                                |                |            |                |      | 20            | 14    |
| Bacillus claussenii (2)        | 1                     |                        |                                |                |            |                |      | 1             | 1     |
| Bacillus halodurans (1)        |                       |                        |                                |                |            |                |      | 3             | 2     |
| Bacillus celluloslyticus (1)   |                       |                        |                                |                |            |                |      | 1             | 3     |
| Bacillus infantis (1)          |                       |                        |                                |                |            |                |      |               |       |
| Bacillus selenitireducens (1)  |                       |                        |                                |                |            |                |      |               |       |
| Bacillus methylotrophicus (15)| 7                     |                        |                                |                |            |                |      | 15            | 26    |
| Bacillus paralicheniformis (3)|                       |                        |                                |                |            |                |      | 68            | 58    |
| Bacillus methanolicus (1)      |                       |                        |                                |                |            |                |      | 9             | 12    |
| Bacillus endophyticus (1)      | 2                     |                        |                                |                |            |                |      | 1             | 1     |
| Bacillus smithii (1)           |                       |                        |                                |                |            |                |      | 4             | 2     |
| Bacillus pseudomycoideas (1)   |                       |                        |                                |                |            |                |      | 1             | 1     |
| Bacillus pseudofirmus (1)      |                       |                        |                                |                |            |                |      | 1             | 1     |
| Bacillus bombysepticus (1)     |                       |                        |                                |                |            |                |      |               | 4     |
| Bacillus lehensis (1)          |                       |                        |                                |                |            |                |      | 1             | 4     |
| Bacillus toyonensis (1)        |                       |                        |                                |                |            |                |      | 1             | 1     |
| Bacillus gabiensis (1)         |                       |                        |                                |                |            |                |      | 3             | 3     |
| Bacillus sp. (13)              | 3                     | 1                      | 3                              | 5              | 1          | 1              | 2    | 6             | 24    |

Zhao and Kuipers BMC Genomics (2016) 17:882
Table 1 Number of putative antimicrobial gene clusters identified in 328 Bacillales genomes (reported in Genbank) (Continued)

| Genus                  | Number of Genomes | Number of Putative Antimicrobial Gene Clusters |
|------------------------|-------------------|-----------------------------------------------|
| **Kyrpidia tusciae**   | (1)               | 3                                             |
| **Alicyclobacillus acidocaldarius** | (2)   | 3                                             |
| **Anoxybacillus flavithermus** | (1)   | 2                                             |
| **Geobacillus stearothermophilus** | (2)   | 1                                             |
| **Geobacillus thermodenitrificans** | (1)   | 1                                             |
| **Geobacillus kaustophilus** | (1)   | 1                                             |
| **Geobacillus sp. (9)** |                   |                                               |
| **Lysinibacillus sphaericus** | (1)   | 1                                             |
| **Lysinibacillus fusiformis** | (1)   | 1                                             |
| **Brevibacillus laterosporus** | (1)   | 1                                             |
| **Brevibacillus brevis** | (1)   | 1                                             |
| **Paenibacillus polypeoxy** | (7)   | 11                                            |
| **Paenibacillus larvae** | (1)   | 1                                             |
| **Paenibacillus mucilaginosus** | (3)   | 2                                             |
| **Paenibacillus peoriae** | (1)   | 1                                             |
| **Paenibacillus odorifer** | (1)   | 1                                             |
| **Paenibacillus stellifer** | (1)   | 1                                             |
| **Paenibacillus borealis** | (1)   | 1                                             |
| **Paenibacillus bovis** | (1)   | 1                                             |
| **Paenibacillus naphthalenovorans** | (1)   | 1                                             |
| **Paenibacillus beijingensis** | (1)   | 1                                             |
| **Paenibacillus graminis** | (1)   | 1                                             |
| **Paenibacillus durus** | (2)   | 1                                             |
| **Paenibacillus terrae** | (1)   | 1                                             |
| **Paenibacillus riograndroides** | (1)   | 1                                             |
| **Paenibacillus sabinae** | (1)   | 1                                             |
| **Paenibacillus sp. (12)** |       |                                               |
| **Total**              |                   | 34                                            |

Numbers in parentheses () indicate the number of genomes analyzed per genus

*RPs is short for ribosomally synthesized peptides
Fig. 1 (See legend on next page.)
Zhao and Kuipers BMC Genomics (2016) 17:882 Page 6 of 18

et al. [15] Class II: unmodified bacteriocins; Class III: large antimicrobial proteins (see also Alvarez-Sieiro et al. [14]). Characteristics of the identified bacteriocins of Bacillales are listed in Additional file 2: Table S2, describing their precursor sequences, gene clusters and predicted producer species, respectively.

**Class I: Ribosomally produced and posttranslationally modified peptides (RiPPs)**

This class consists of antimicrobial peptides (less than 10 kDa) that are ribosomally synthesized, undergoing posttranslational modifications (PTMs), resulting in different structures and properties. In this study, we found 438 putative gene clusters of class I bacteriocins, widely distributed over 49 species of Bacillales (Table 1). According to the modification differences, this class can be subdivided into seven subclasses. Subclass 1 includes peptides with modifications typical for lantibiotics (e.g. lanthionine), while subclasses 2–7 include peptides with other unique modifications [15–17].

**Subclass 1: Lanthipeptides** Lanthipeptides are peptides containing unusual amino acids, such as dehydroalanine/dehydrobutyryne, lanthionine/methyl-lanthionine residues, introduced by different kinds of PTMs [15]. Lanthipeptides with antimicrobial activity form the so-called lantibiotics [17], which can be subdivided into four subclasses, following the classification scheme of lanthipeptides [18]. The main differences between classes I, II, III and IV lantipeptides are the PTM enzymes involved. Class I lantipeptides are modified by two distinct enzymes that carry out the PTM process: dehydratase LanB and cyclase LanC, while class II peptides are modified by a bifunctional lanthionine-introducing enzyme, called LanM. There are also two-component lantipeptides consisting of two peptides, which belong to class II lantipeptide, because they are processed by a single modifying enzyme, called LanM [19–23]. For other lantipeptides (class III and IV), the dehydration and cyclization reactions are catalyzed by multifunctional enzymes (RamC/LabKC or LanL) or they lack significant antibiotic activity, which are not further described here [24].

Subtilin is a well-investigated class I lantipeptide produced by *B. subtilis*, the encoding gene cluster of which is also found in the genome of *Bacillus* sp. YP1. The gene encoding subtilin encodes a 56-residue peptide precursor that is processed to yield the 32-residue mature peptide, which is structurally related to the lantibiotic nisin of *Lactococcus lactis* [25]. The subtilin gene cluster includes the structural gene *spaS*, encoding its prepeptide; PTM genes *spaB* and *spaC*, encoding a dehydratase and a cyclase for lanthionine formation, respectively; transporter gene *spaT* for modified precursor export and immunity genes *spaIFEG* (Additional file 2: Table S2) [26–28]. The presubtilin will be converted to mature subtilin by serine proteases secreted by *B. subtilis* [29]. Subtilin exhibits bactericidal activity against a broad spectrum of Gram-positive bacteria, based on pore formation in the cytoplasmic membrane, using cell wall precursors such as lipid II and undecaprenyl pyrophosphate, the hydrophobic carrier module for peptidoglycan monomers, as docking module and as a central constituent of the pore [30, 31]. The class II lanthipeptide mersacidin produced by several *B. amyloliquefaciens* strains [32–34], with a more globular structure comprising 20 amino acid residues, inhibits cell wall biosynthesis by binding to lipid II [35, 36]. The mersacidin gene cluster includes the structural gene *mrsA*, two modification genes (Additional file 2: Table S2), i.e. *mrsM* coding for both dehydration and cyclation and *mrsD* coding for a C-terminal S-[(Z)-2-aminovinyl]-3-methyl-D-cysteine formation enzyme, and the gene *mrsT* coding for a transporter with an associated protease domain, as well as three genes, *mrsEFG*, coding for immunity and three genes, *mrsR1, R2, K1*, coding for regulation [37–39].

A total of 105 putative lantipeptide gene clusters were discovered in Bacillales in this study (Table 1). Among them, gene clusters of class I lantipeptides distribute over the genomes of *B. subtilis, B. thuringiensis, B. cereus, B. megaterium, B. mycoides, B. clausii, Bacillus sp., Geobacillus thermodenitrificans, Geobacillus kaustophilus, Paenibacillus polymyxa, Paenibacillus larvae, Paenibacillus peoriae and Paenibacillus durus*, while gene clusters of class II lantipeptides distribute over the genomes of *B. thuringiensis, B. cereus, B. amyloliquefaciens, B. licheniformis, B. mycoides, B. halodurans, B. methylotrophicus, B. paralicheniformis, B. endoplytoticus, B. pseudomycoides, Bacillus sp., G. thermodenitrificans, P. polymyxa, P. durus* and *Paenibacillus sp.* (Additional file 2: Table S2 and Fig. 2). Class I lantipeptides identified by BAGEL3 includes subtilin, clausin, subtilomycin and geobacillin I [22, 40–42]. Gene clusters of entianin, ericinA/S, paenibacillin, paenicidin A, B, thuricin 4A and its derivative thuricin 4D were not found by genome mining (Additional file 2: Table S2) [43–47]. Class II lantipeptides usually exhibit a globular structure, including mersacidin, amylolysin, pseudomycoicidin, cerecidin...
Fig. 2 Distribution of antimicrobials biosynthetic gene clusters among different Bacillales. a Class I ribosomally synthesized peptides; b Class II ribosomally synthesized peptides; c Class III ribosomally synthesized peptides; d Total ribosomally synthesized peptides, NRPs and PKs, respectively.
A1-A6 and geobacillin II; also two-component class II lanthipeptides including haloduracin and lichenicidin were identified [19, 22, 23, 48–52]. It is notable that gene clusters of two novel subtilin-like lantibiotics were found in several _P. polymyxa_ strains. By further analysis, both of their sequence of core peptides showed high similarity with the N-terminal part of subtilin but were quite different in the C-terminal part. Moreover, we report a novel gallidermin/N-terminal part of subtilin but were quite different in the C-terminal part. 

Looking into the sequence of its precursor peptide (see Additional file 2: Table S2), it has the conserved F(N/D)LID motif in its leader and theoretically could form the same rings as gallidermin/nisin according the position of serine and cysteine residues. All of the three putative lantibiotics have _lanBC_ genes in their gene clusters, which suggest they are involved in their production. A gene cluster of a two-peptide bacteriocin was found in the genome of _B. cereus_ Q1. Due to the existence of a _lanM_ gene, it was predicted to be a class II lanthipeptide. Interestingly, the C-terminal parts of both its core peptides are similar to lichenicidin and haloduracin, and the N-terminal part of one of the core peptides shows high similarity with one of cytolysins produced by _Enterococcus faecalis_ [53].

**Subclass 2: Head to tail cyclized peptides**  
Head to tail cyclic peptides are named by their unifying feature, which is the head to tail circularization of their peptide backbones by direct linkage of their N- and C-terminal amino acids, resulting in a well-defined three-dimensional structure, by folding in α-helical manner [54–57]. To our knowledge, these peptides contain no lanthionine, β-methyl-lanthionine, and dehydrated residues, making them clearly distinguishable from lanthipeptides [58].

Amylocyclcin was recently reported to be produced by _B. amyloquefaciens_ FZB42 and identified as a novel circular bacteriocin [59], which is derived from the 112 amino acid precursor AcnA (Additional file 2: Table S2) encoded by _acnA_, with a 48 amino acid derived leader cleaved by a protease that is still unknown, and then circularization occurring between Leu-1 and Trp-64 [59]. There are gene clusters present, regulating their maturation (e.g. circularization and cleavage), transportation and self-protection. The first gene of the putative operon, _acnB_, encodes a membrane-anchored protein comprising five transmembrane helices with unknown function. _acnD_ is likely to encode the transporter complex, whereas AcnC might act as circularization enzyme showing high similarity with the sequence of UclB, which brings uberlys in to maturation [60]. _AcnEF_ are proposed to be the putative immunity genes. Amylocyclcin has the ability to inhibit Gram-positive bacteria like _B. subtilis_, but not against Gram-negative bacteria.

There are 52 gene clusters of putative head to tail cyclized peptides identified in this genome-mining study, which distribute over the genomes of _B. thuringiensis_, _B. cereus_, _B. coagulans_, _B. pumilus_, _P. paralicheniformis_, _B. gobiensis_, _B. subtilis_ sp., _Kyrpidia tusciae_, _Geobacillus stearotherophilus_, _G. kaustophilus_, _Geobacillus sp._, _P. larvae_ and _Paenibacillus mucilaginosus_ (Table 1 and Fig. 2). An amylocyclcin-like circular bacteriocin gene cluster was found in the genomes of _B. coagulans_. The core peptide sequence is identical to that of amylocyclcin of _B. amyloquefaciens_ FZB42, but the leader peptide sequence is quite different (Additional file 2: Table S2). It is noteworthy that a gene cluster of an uberolysin-like peptide was detected in the genome of _Bacillus_ sp. INAL3E and gene clusters of circularin A/bacteriocin AS-48 like peptide were detected in several _Geobacillus_ sp., while uberolysin was produced by _Streptococcus uberis_, circularin A was produced by _Clostridium beijerincki_ and bacteriocin AS-48 was produced by _E. faecalis_ [54, 60–63].

From the core peptide sequences, their circularization is most likely being formed between leucine and tryptophan (Additional file 2: Table S2). There are also other putative gene clusters of head to tail cyclized peptides found in this study, but notably these show no similarity with reported peptides. Whether these are real circular bacteriocins or not, need to be further investigated experimentally.

**Subclass 3: Sactipeptides**  
Sactipeptides form a class of cyclic antimicrobial peptides with unusual sulfur to α-carbon cross-links, which are catalyzed by radical S-adenosylmethionine (SAM) enzymes in a leader peptide-dependent manner [64, 65]. Posttranslational linkage of a thiole to the α-carbon of an amino acid residue responsible for their antimicrobial bioactivities is rare in ribosomal synthesized peptides and they are classified as an independent group [66–68]. These unusual linkages differ from lanthionine bridges containing sulfur to β-carbon linkages.

Subtilosin A is a 35-residue peptide, formed by cleavage of a seven amino acid leader peptide, cyclization of the N- and C-terminal parts, and further modification of cysteine, threonine and phenylalanine residues. The maturation of subtilosin A begins with the transcription and translation of the _sbo-alb_ genes (Additional file 2: Table S2), resulting in the precursor peptide SboA [69, 70]. Subsequently, the radical SAM enzyme AlbA generates the thioether linkages between the sulfur atom of the cysteine residue and the α-carbon of the threonine residue [68]. Afterwards, either AlbE or AlbF (putative proteases) cleaves off the leader peptide. In the last step, the peptide backbone is circularized by one of the two proteases, resulting in subtilosin A, which is subsequently exported by the putative ABC transporter AlbC. The operon is induced under anaerobic conditions and is controlled by the transition state regulator protein AbrB [4]. It shows antibacterial activity against...
Bacillus spp., E. faecalis, Gardnerella vaginalis and Listeria monocytogenes by targeting their membranes and forming pores [71–73].

In this study, we found 87 putative gene clusters of sactipeptides in the genomes of Bacillales (Table 1), most of which belong to three reported types of sactipeptides (Additional file 2: Table S2): subtilosin A from B. subtilis, B. atrophaeus, B. simthii and Bacillus sp. strains; sporulation killing factor (SKF) from B. atrophaeus, B. pumilus and B. subtilis strains; and thuricins, such as thuricin H (17) and thuricin CD from B. thuringiensis and B. cereus [67, 74–77]. We found several other putative gene clusters of sactipeptides in the genomes of B. clausii, G. stearothermophilus, Brevibacillus laterosporus, P. larvae, Paenibacillus odorifer, Paenibacillus graminis, Paenibacillus riograndensis and Paenibacillus sp. (Additional file 2: Table S2 and Fig. 2), which showed very limited similarity with reported sactipeptides, and that need to be further experimentally confirmed.

Subclass 4: Linear azole-containing peptides (LAPs)
The linear azole-containing peptides (LAPs), form an important subgroup of RfPPs with a distinguishing heterocyclic ring of oxazoles and thiazoles derived from serine/threonine and cysteine by enzymatic cyclodehydration and dehydration [78–81]. Prominent natural products such as microcin B17 produced by Escherichia coli and streptolysin S produced by LAB, are model of represent LAPs peptides [82–86]. The LAP family has already been extended with plantazolicin A and B produced by B. amyloliquefaciens and B. methylotrophicus [80, 81].

Plantazolicin A (Additional file 2: Table S2) and its demethyl analogue plantazolicin B represent an unusual type of thioazole/oxazole-containing peptide antibiotic with a hiterto unknown mechanism of action, which show inhibition against Bacillus [80, 87]. The mature product plantazolicin is a linear 41 amino acid precursor peptide with the 14 amino acid core-peptide encoded by the structural gene pznA. The trimeric protein complex PznBCD (cyclodehydratase, dehydrogenase, and docking/scaffolding protein) likely catalyzes PTMs of ten cyclodehydrations followed by nine dehydrogenations. After the protease PznE cleaves off the leader peptide to yield desmethylplantazolicin plantazolicin B, a final N,N-bismethylation by methyltransferase PznL gives plantazolicin A [80, 81].

A total of 117 putative gene clusters of LAPs occupy 20% of the total putative gene clusters of bacteriocins in this study and are widely distributed in more than 20 species of Bacillales (Table 1 and Fig. 2). However, only plantazolicin A and B produced by B. amyloliquefaciens and B. methylotrophicus have been reported before (Additional file 2: Table S2). This means that many novel LAPs can be found and need further experimental investigation.

Subclass 5: Thiopeptides Thiopeptides, or thiazolyl peptides are highly modified via either non-ribosomal or ribosomal assembly, with a six membered nitrogenous macrocycle being central of piperidine/pyridine/dehydropiperidine and including additional thiazoles and dehydrated amino acid residues [15, 88, 89]. Because of the trithiazolyl (tetrahydro) pyridine core, they display high affinity binding to either the 50S ribosomal subunit or elongation factor Tu.

In the thiocillins, found in the producer B. cereus ATCC 14579, at least 10 and up to 13 of the 14 C-terminal residues undergo PTM to generate a set of eight related antibiotics. The thiocillin gene cluster contains four identical copies of a gene encoding a 52-residue precursor peptide (tclE-H), which is thought to be posttranslationally modified to yield the mature antibiotic scaffold (Additional file 2: Table S2). Four of the eight thioicillins produced abundantly by B. cereus display similar efficacies against B. subtilis and two methicillin-resistant Staphylococcus aureus (MRSA) strains [90, 91]. Thiopeptide gene clusters involved in ribosomal synthesis are found in the genome sequences of several B. cereus, B. subtilis and Lysinibacillus sphaericus (Additional file 2: Table S2 and Fig. 2), which might go beyond the classification for LAB bacteriocins [14].

Subclass 6: Glycocins Glycocins are bacteriocins with glycosylated residues. There are various unique and diverse putative glycopeptide containing bacteriocins named glycocins in Firmicutes [15, 92].

There is one model glycopeptide bacteriocinc, sublancin 168 (Additional file 2: Table S2), produced by B. subtilis with a β-S-linked glucose moiety attached to cysteine22 and two disulfides [92–95]. The sublancin 168 biosynthetic gene cluster contains the precursor gene sunA coding a 56-residue polypeptide consisting of a 19-residue leader peptide and a 37-residue mature peptide and genes bdbA and bdbB encoding two thiol-disulfide oxidoreductases, i.e., BdBa and BdBb [95, 96]. In addition, it contains two open reading frames of unknown function, yolF and yolJ. YolF was recently suggested to be important for immunity of the producing strain and was renamed SunT; the function of YolJ has not yet been reported [97]. SunT is responsible for transport. The antimicrobial activity spectrum of sublancin 168 was like that of lantibiotics, inhibiting Gram-positive bacteria, but not Gram-negative bacteria; and acts also similar to the lantibiotics nisin and subtilin in its ability to inhibit both bacterial spore outgrowth and vegetative cell growth [17].

In addition to sublancin 168 found in B. subtilis, genomining study indicated that nine other putative gene clusters of glycocins were found in genomes of B. thuringiensis, B. cereus, B. weihenstephanensis, B. lehensis, Bacillus sp., Geobacillus sp. and Paenibacillus sp., which need further characterization (Table 1 and Fig. 2).
**Subclass 7: Lasso peptides** Lasso peptides, which form an emerging class of RiPPs from bacteria, were first described in 1991 [98]. Their defining structural feature is an N-terminal macro lactam ring that is threaded by the C-terminal tail resulting in a unique lasso structure—the so-called lariat knot. The ring is formed by an isopeptide bond between the N-terminal α-amino group of a glycine, alanine, serine, or cysteine and the carboxylic acid side chain of an aspartate or glutamate, which can be located at positions 7, 8, or 9 of the amino acid sequence [16, 99].

In general, lasso peptide production requires at least three genes encoding a precursor peptide A, a cysteine protease B, and an ATP-dependent lactam synthetase C. Gene clusters might contain additional genes, but so far no system was proven to be in need of an additional enzyme to produce mature lasso peptides [100–104]. Microcin J25 produced by *E. coli* AY25 has served as a model for studies of lasso peptides [105]. Known lasso peptides display antimicrobial activity by enzyme inhibition [106, 107].

Genome mining of *Bacillales* indicated 48 gene clusters of hypothetical peptides, which are likely lasso peptides in the genomes of 20 *Bacillales* species (Table 1 and Fig. 2), but these still need to be experimentally confirmed.

**Class II: unmodified bacteriocins**

Class II bacteriocins include small (less than 10 kDa), ribosomally synthesized, heat-stable, membrane-active linear peptides [4, 108, 109]. According to genome mining results, we found in total 121 putative gene clusters of class II bacteriocins distributed over 16 species of *Bacillales* (Table 1 and Fig. 2). This class can be subdivided into two subclasses: 1. Pediocin-like peptides; 2. Other unmodified peptides (Additional file 2: Table S2).

**Subclass 1: Pediocin-like peptides** The pediocin-like bacteriocins are antilisterial peptides that have a YGNGVXC consensus motif [110, 111]. Coagulin produced by *B. coagulans* I4 is a peptide of 44 residues has an amino acid sequence similar to that described for pediocins AcH and PA-1 [109, 112]. Coagulin and pediocin differ only by a single amino acid at their C-terminus (asparagine41threonine). Gene clusters of coagulin are located on a plasmid including the structural gene *coaA*, immunity gene *coaB*, and ABC transporter genes *coaC* and *coaD* [113].

**Subclass 2: Other unmodified peptides** Subclass 2 includes other unmodified peptides, such as lichenin produced by *B. licheniformis*, or cereins produced by *B. cereus*, which have already been described in a previous review although not yet detected in the reported complete genome sequences [4]. We found a lactocin A family protein [114] and a lactococcin A1 family protein [115] belonging to class II bacteriocins from *Anoxybacillus flavithermus* WK1. Here, we mainly added some new members of *Bacillus* class II bacteriocins detected by BAGEL3, in particular holins and holin-like peptide BhlA, antimicrobial peptide LCI, and leaderless bacteriocin aureocin A53 (Additional file 2: Table S2).

Analysis of all *Bacillales* genome sequences revealed the presence of a structural gene encoding a holin in *Geobacillus* sp. WCH70 and BhlA encoding genes in most of *B. subtilis*, *B. amyloliquefaciens*, *B. mycoides*, *B. pseudomycoide*, *B. licheniformis*, *B. pumilus* and *B. thuringiensis*, and further structural analysis of their sequence revealed features similar to holin (Additional file 2: Table S2) [116, 117]. Holins are phage-encoded proteins involved in the disruption of bacterial membrane to facilitate the release of progeny phage particles [118–121]. However, the functions of these specific ORFs have not yet been identified. The bacteriocin-related holin-like peptide BhlA from *Bacillus* showed antibacterial activity against several Gram-positive bacteria, including MRSA and *Micrococcus luteus* by destroying cell membranes [122]. BhlA consists of 70 amino acid residues with a single transmembrane domain at the N-terminus, a number of highly charged amino acid residues at the C-terminus. The presence of hydrophilic residues and the membrane topology of BhlA make it different from holins [122].

The *ici* gene encoding LCI was found in the genomes of *B. amyloliquefaciens* and *B. methylotrophicus* strains (Additional file 2: Table S2), sharing 98–100 % identity with the LCI sequence of *B. subtilis*. The antimicrobial peptide LCI was first identified and isolated by Liu et al. [123] from a *B. subtilis* strain named A014 that possesses very strong antagonistic activity against the Gram-negative pathogen *Xanthomonas campestris pv oryzae* causing rice leaf-blight disease, which is a serious threat to rice production and causes great losses in yields in most rice fields annually. LCI is a β-structure antimicrobial peptide containing 47 residues of 5460 Da with no disulfide bridge or circular structure. It also contains a hydrophobic core formed by valine5, tyrosine41 and tryptophan44 as well as 23 H-bonds which contribute to its considerable thermal stability [124, 125]. According to our BAGEL3 gene cluster mining results, there are two genes: a structural gene *ici* and an immunity/transporter-like gene which was still unknown. LCIs positively charged residues lead to a short-lived channel in the bacterial membrane of sensitive strains [126].

Another new member of *Bacillus* class II bacteriocins is leaderless aureocin A53, whose gene cluster was identified in the genome sequence of *B. pumilus* strains (Additional file 2: Table S2). It is active against *L. monocytogenes* by dissipating the membrane potential and simultaneously stopped biosynthesis of DNA, polysaccharides, and protein [127]. Aureocin A53 is a highly cationic 49-residue peptide containing six lysine and four tryptophan residues. Unlike most class II bacteriocins, aureocin A53 is synthesized without a leader peptide and retains a formylated N terminus. Notably, genes for biosynthetic enzymes, immunity functions, or
regulation of biosynthesis are not found in the vicinity of the aureocin A53 structural gene [128].

**Class III: large antimicrobial proteins**

This group includes large proteins (larger than 10 kDa) with antimicrobial activity. Gene clusters of these proteins normally include an immunity gene and a structural gene [126]. We found 24 putative gene clusters of class III bacteriocins distributed over seven species of Bacillales (Table 1 and Fig. 2). In a previous review, megacins produced by *B. megaterium* ATCC 19213 were reported as class III bacteriocins [4, 129]. Here, we identified and introduced some class III bacteriocins by BAGEL3 represented by colicin, M23 peptidase and pyocin AP41 (Additional file 2: Table S2).

Gene clusters of colicins were identified in the genomes of *B. thuringiensis*, *B. cereus* and *Bacillus* sp. BH072 (Additional file 2: Table S2). Channel-forming colicins (colicins A, B, E1, Ia, Ib, and N) are transmembrane proteins that depolarize the cytoplasmic membrane, leading to dissipation of cellular energy. Their immunity gene is often produced constitutively, while the bacteriocin release protein is generally produced only as a read-through of the stop codon on the colicin structural gene. The colicin itself is repressed by the SOS response and may be regulated in other ways, as well [130]. Pyocin AP41 is also discovered as a large bacteriocin from *B. thuringiensis* (Additional file 2: Table S2), which was first isolated from *Pseudomonas aeruginosa* PAF41. According to literature, it showed a similar mode of action to that of colicin [131]. Interestingly, we found gene clusters of M23 peptidase in the genomes of *B. thuringiensis*, *B. coagulans* and *B. halodurans* (Additional file 2: Table S2), while M23 peptidase has not been reported to be secreted by *Bacillus* before, and so needs to be further experimentally confirmed. Over the past years, many members of the M23 metallopeptidase family have been identified and biochemically characterized. Structures have been determined for some of them, e.g. LytM, LasA and recently lysostaphin, a prototypic enzyme of the M23B group and the best studied bacteriocin of this group [132, 133].

**Non-ribosomal synthesized peptides (NRPs) and polyketides (PKs) of Bacillales**

NRPs and PKs encompass a variety of linear, cyclic and branched structures, which are generated by complex enzymes known as non-ribosomal peptide synthetases (NRPS), polyketide synthetases (PKS) and hybrid NRPS/PKS, respectively [134, 135]. Among them, NRPs produced by Bacillales include lipopeptides (LPs) and others, with significant antimicrobial activity [136]. Here we present an extended collection based on members described in a previous review by Aleti et al. [136]. By use of antiSMASH, we identified 31 types of putative NRPs, PKs and NRPS/PKS hybrid synthesized antimicrobials, which will be described in detail below. Characteristics of them are listed in Additional file 3: Table S3 by displaying their chemical structures, gene clusters and predicted producer species, respectively.

**Lipopeptides (LPs)**

Lipopeptides (LPs) are natural compounds of bacterial origin consisting of a hydrophobic long alkyl chain linked to a hydrophilic polypeptide to form a cyclic or linear structure [137]. According to our mining results, *B. amyloliquefaciens*, *B. methylotrophicus*, *B. atrophaeus*, *B. subtilis*, *B. licheniformis*, *B. paralicheniformis*, *B. pumilus*, *B. lehensi*, *B. laterosporus*, *Bacillus* sp., *P. polymyxa*, *P. larvae*, *P. mucilaginosus*, *P. peoriae*, *P. bovis*, *Paenibacillus terrae* and *Paenibacillus* sp. are likely to be the main producers of LPs, which are mainly known for their antifungal properties [138–140]. Based on a previous genome mining work (see review Aleti et al. [136]), we identified locilolymycins as novel members of LPs in species of Bacillales (Additional file 3: Table S3).

Traditional LPs (comprising the surfactins, iturins and fengycins) from *Bacillus* are homologues differing in length, branching, and saturation of their acyl chain. The surfactin family (exemplified by surfactin, lichenysins and pumilacins) contain a cyclic heptapeptide that forms a lactone bridge with ß hydroxy fatty acids [141]. The iturin group includes A, C, D and E isomers, bacillomycin D, F and L and mycosubtilin. All these compounds contain a cyclic heptapeptide acylated with ß amino fatty acids [142, 143]. The fengycin family comprises the decapeptide fengycin A and fengycin B, which differ in a single amino acid at the sixth position (D-alanine and D-valine, respectively) [144]. Kurstakins form another family of LPs composed of four partially cyclic heptalipeptides, which differ only in their fatty acid chains [145]. The gene clusters of the *Bacillus* LPs encoding the surfactin, fengycin, iturin and kurstakin families have been described and summarized in a number of recent reviews [6, 11, 136, 145]. Cerexins are linear LPs with strong antimicrobial activity against *S. aureus* and *Streptococcus pneumoniae* [146]. Kurstakins and cerexins are isolated and identified from *B. thuringiensis* and *B. cereus* strains before, respectively [146, 147]. Locilolymycins (locilolymycin A, B, and C derivatives), a novel family of cyclic lipopeptides active against bacteria and viruses produced by *B. subtilis* 916 [148, 149], include a unique nonapeptide sequence and macrocyclization. The locilolymycin biosynthetic gene cluster encodes four proteins (LocA, LocB, LocC, and LocD) that form a hexamodular NRPS to biosynthesize cyclic nonapeptides. *Paenibacillus* now are found to produce a large number of LPs [136]. Polymyxins are cyclic cationic LPs which contain the non-proteogenic amino acid 2, 4-diaminobutyric acid contributing to the overall positive charge of the cationic LPs, exhibiting antibacterial activity against both Gram-positive and Gram-negative bacteria by acting on their
membranes. The gene cluster consists of five genes, of which pmxA, B and E encode the polymyxin synthetase, whereas pmxC and D are involved in transport [136, 150]. Another cationic lipopeptide, paenibacterin is a new broad-spectrum antimicrobial agent consisting of a cyclic 13-residue peptide and an N-terminal C15 fatty acyl chain [151]. There are also cyclic noncationic LPs from Paenibacillus comprising fusaricidins containing a cyclic hexapeptide structure with antagonistic activity against Fusarium oxysporum, tridecapins with strong antimicrobial activity against Gram-negative bacteria. Polypeptins, octapeptins, pelgipeptins, gavaserin and saltavalin are LPs isolated from Paenibacillus sp. strains, reported before by other scientists, and should also be included in this collection [136, 152–155].

Other NRPs

By antiSMASH, we also found non-lipopeptide but NRPSs gene clusters putatively encoding NRPs with antimicrobial activity mainly in the species of Bacillus, Paenibacillus and Brevibacillus. We collected them as a group of other NRPs, which is exemplified by the following NRPs (Additional file 3: Table S3).

The non-ribosomal dodecapeptide bacitracin, released by some B. licheniformis and B. subtilis strains, proved to be an inhibitor of cell wall biosynthesis of Gram-positive bacteria [156, 157]. Small peptide bacilysin secreted by B. subtilis, B. amyloliquefaciens and B. pumilus contains an N-terminal alanine residue and L-anticapsin with antibacterial activity against S. aureus [158]. B. subtilis also produces rhizocins, phosphonate oligopeptide antibiotics containing the C-terminal non-proteinogenic amino acid (Z)-1-2-amino-5-phosphono-3-pentenoic acid displaying antifungal activity [159]. Petrobactin and bacillicin produced by several Bacillus strains under iron-limited conditions, are catecholate siderophores associated with two operons, asb (for petrobactin) and bac (for bacillibactin) [160].

Sevadacin is a tripeptide (D-phenylalanine-D-alanine-cysteine) produced by a NRPS encoded by a gene cluster found in the genome of P. larvae, which was shown to have antibacterial activity [161].

Both the cyclic peptides gramicidin S and tyrocidine, produced by Brevisbacillus, consist of 10 amino acid residues [162]. Gramicidin S consists of two identical pentapeptides, which are linked head to tail, and together form the stable amphiphilic cyclic decapetide. The first amino acid residue of the two pentapeptides is in the D-configuration [163]. The peptide exhibits strong antibacterial and antifungal activity [164]. Tyrocidine, actually a mixture of slightly different decapetides, is active against several Gram-positive bacteria and it has been suggested that this peptide plays a role in the regulation of sporulation of B. brevis [162]. The gramicidin S biosynthesis operon (grs) contains thee genes, which are grsA, encoding the gramicidin S synthetase 1; grsB, encoding the gramicidin S synthetase 2, and grsT, encoding a protein of unknown function. The sequence of the grsA gene product showed a high similarity with the tyrocidine synthetase 1 (TycA protein) [165, 166].

**Polyketides (PKs)**

Polyketides represent a group of secondary metabolites, exhibiting remarkable diversity both in terms of their structure and function. Polyketide natural products are known to possess a wealth of pharmacologically important activities, including antimicrobial, antifungal, antiparasitic, antitumor and agrochemical properties (http://www.nii.ac.in/~pksdb/polyketide.html). Novel gene clusters likely encoding similar PKSs were identified using antiSMASH. They were most prominent in B. subtilis, B. amyloliquefaciens, B. methylotrophicus, B. atrophaeus, B. laterosporus and Paenibacillus sp. (Additional file 3: Table S3 and Fig. 2). The genus Bacillus produces three types of PKs including bacilaene, difficidin and macrolactin; Paenibacillus produces paenimacrolin [6, 167]. B. laterosporus also produced the polyketide basiliskamide with antifungal activity [168], and it was added as novel member of PKs in species of Bacillales (Additional file 3: Table S3).

Bacilaene was first isolated from B. subtilis strains [169], are found to display a linear structure comprising a conjugated hexaene, while its gene clusters bae (bae, L, M, N and R) has now been discovered in several other Bacillus genomes, including B. amyloliquefaciens, B. atrophaeus and P. polymyxa. It is an inhibitor of prokaryotic protein synthesis, constituted by an open-chain enamine acid with an extended polypeptide chain and shows good antimicrobial activity against human pathogens such as Serratia marcescens, Klebsiella pneumoniae and S. aureus [37, 169, 170]. Difficidin is known to be produced by B. amyloliquefaciens strains, which is active against the phytopathogen Erwinia amylovora causing fire blight, and contains a highly unsaturated macrolactine polyene comprising a 22 membered carbon skeleton with a phosphate group rarely found in secondary metabolites [171]. Difficidin is encoded by the gene cluster dif with 14 open reading frames from difA to difN and difY. The contribution of the genes dif and difK are unclear and their potential activities are not seen in the final product [172]. Macrolactin has also been isolated from B. amyloliquefaciens strains [173]. Most macrolactins consist of a 24 membered lactone ring with three diene moieties in the carbon backbone, which is encoded by the gene cluster mln, containing nine operons including mlnA-I [174]. As the other Bacillus polyketides, macrolactins show antibacterial activity and might have the potential to be used in medical application. Moreover, they could inhibit the proliferation of murine melanoma cancer cells and the replication of mammalian Herpes simplex virus and HIV in lymphoblast cells [136, 173]. Paenimacrolin was isolated from Paenibacillus sp. F6-B70 with a 22 membered lactone ring showed high similarity with difficidin, which has
antimicrobial activity against *Staphylococcus* [167]. The polyketide antibiotics basilikamides A and B, which exhibit potentiactivity against *Candida albicans* and *Aspergillus fumigatus*, both comprise a 21 membered carbon skeleton, structurally identical in every respect, except for the position of the cinnamate ester: C9 in basilikamide A and C7 in basilikamide B [175, 176].

**NRPS/PKS hybrid synthesized compounds**

There are three NRPS/PKS hybrid synthesized NRPs or PKs of Bacillales identified in this study (Additional file 3: Table S3). Paenilarvin A is a hybrid polyketide-nonribosomal peptide produced by certain *B. cereus* group strains, inhibiting certain Gram-positive, Gram-negative, and eukaryotic microorganisms [181, 182]. Paenilaminic acid is another hybrid NRPS/PKS synthesized peptide with antibacterial and antifungal activity, whose encoded gene clusters (pam) were found the genomic sequence of the Gram-positive bacterium *P. larvae* [183].

In this study, 10 novel gene clusters encoding putative NRPs, PKs or NRPS/PKS hybrids were predicted from the genome of *B. brevis* NBRC 100599, *B. cereus* AH820, *B. cereus* G9842, *B. cereus* B4264, *B. cereus* E33L, *B. thuringiensis* HD771, *B. thuringiensis* HD789, *B. amyloliquefaciens* DSM7, *B. amyloliquefaciens* CC178, *B. methylotrophicus* NAU-B3, *B. anthracis* str. A0248, *B. anthracis* str. H9401 and *Bacillus* sp. BH072. The identified gene clusters (an characterized) show limited homology with gene clusters in the integrated databases. Related genes encoding the biosynthesis, predicted structures and antimicrobial activity of these compounds deserve to be experimentally validated.

**Discussion**

An extensive investigation of 328 published whole genome sequences of Bacillales for the presence of ribosomally synthesized antimicrobials, NRPs or PKs encoding genes, revealed that most species of the genus *Bacillus*, *Paenibacillus* and *Geobacillus* have good potential to produce a wide variety of antimicrobials and there is a high occurrence of putative biosynthetic gene clusters. The ability of *Bacillus* from different species to produce putative antimicrobial compounds relate to their phylogenetic relationship. According to the phylogenetic tree (Fig. 1), Bacillales are divided into several groups. Among them, the group of *B. subtilis* and *B. atrophaeus*, the group of *B. amyloliquefaciens*, *B. methylotrophicus*, *B. parlicheniformis*, *B. licheniformis*, *B. punillius* and *B. endophytics* are excellent producers of all the three kinds of antimicrobials. Additionally, the *B. cereus* group, *Paenibacillus* strains are rich sources of bacteriocins and NRPs, while *Geobacillus* strains mainly produce bacteriocins and PKs.

More than 89% strains covering 50 species have a predisposition towards producing ribosomally synthesized peptides (Fig. 1), some gene clusters of which show similarity with those of known bacteriocins, while some are uncharacterized or show limited homology. When it comes to the distribution of biosynthetic gene clusters of ribosomally synthesized antimicrobials among different Bacillales, lanthipeptides, head to tail cyclized peptides, sactipeptides, lasopeptides and LAPS of Class I are the most common types (Fig. 2), whilst glycosin and thiopetide genes are present predominantly in *B. subtilis* and *B. cereus* strains, respectively. Gene clusters of class II and III appear to be also regularly contained within genus *Bacillus* genomes.

Although the emphasis here is on ribosomally synthesized peptide classes, several new NRPs and PKs with potential antimicrobial activity were also identified. Bacillales are potential NRPs producers, and the gene clusters are widely spread in 40 species of the sequenced genomes in our analysis (Fig. 1). In contrast, only half of the genomes of these organisms appear to have PKs encoding genes. *Bacillus* and *Paenibacillus* genera in particular are well noted for their capability to produce structurally diverse NRPs and PKs. Approximately 35% strains of the Bacillales analyzed have the ability to produce all three types of antimicrobial compounds simultaneously. In this study, most of the genomes (255 of 328) were completely sequenced yielding one or only a few contigs, while there are some other level sequence data (shown in Additional file 1: Table S1) composed of relatively many single contigs. Some of these contigs are not in the correct order, which can result in higher mining counts of NRPs (caused by duplications or multiplications) than actual correct. In order to avoid this overestimation, the numbers of putative gene clusters of NRPs identified in Bacillales genomes, especially for *B. cereus* group strains, were adjusted by removing duplications or multiplications of NRPs manually. It is valuable to take this issue into account in further and future data mining and analyses.

The massive numbers of bacteria with whole genome sequence data and the development of various specific genome mining tools have made it possible to identify an informative set of putative antimicrobial gene clusters across the genomes that can be developed into new antimicrobials. Novel information found in this genome mining study includes three types that are novel: class I bacteriocins with either a new leader sequence or new core sequence; known antimicrobial compounds previously produced by other microorganisms; and completely novel gene clusters that need experimental confirmation. Another value of this study is that the post-genome mining analysis includes a number of potential species never considered to
be antimicrobial producers before and provide a reference for future Bacilli to be sequenced.

Conclusions
A multitude of antimicrobial compounds have been found to be produced by a variety of Bacillus strains. In the past, these compounds had to be identified by intensive screening for antimicrobial activity against appropriate targets and subsequently purified using fastidious methods prior to assess their potential utilization as antibacterial or antifungal compound. Nowadays, gene clusters encoding for ribosomally produced bacteriocins, NRPs and PKs can readily be identified in the genomic sequences by genome-mining tools that not only add missing ones, but also predict novel ones. Notably, genomic tools like BAGEL3 and antiSMASH combined with specific BLAST searches, makes the identification of new compounds much easier. Although several novel gene clusters of putative antimicrobials were found, they are as yet uncharacterized and their functions remain to be studied. Our extended classification of antimicrobial compounds demonstrates that Bacillales provides a rich source of novel antimicrobials that can now be readily tapped experimentally, since many new gene clusters were identified.

Methods
Genome sequences
Whole genome sequences of 328 strains of Bacillales (Additional file 1: Table S1) were obtained from NCBI Genome database (http://www.ncbi.nlm.nih.gov/genome). All proteins of all genomes were compared by bidirectional BLAST using Proteinortho and newick tree file was generated by p02tree [184]. The newick tree file was visualized using FigTree v1.4.3 (http://tree.bio.ed.ac.uk/software/figtree/).

Genome mining for gene clusters of putative antimicrobials by BAGEL 3 and antiSMASH
Genomes were analyzed for gene clusters of putative bacteriocins, NRPs, PKs or other antimicrobials by using web-based genome mining tools BAGEL3 (http://bagel/molgenrug.nl/) [7] and antiSMASH (http://antismash.secondarymetabolites.org) [8–10]. Genome mining data were collected and putative gene clusters were classified manually. By BLAST, known and novel antimicrobials were predicted and identified.

Additional files

Additional file 1: Table S1. Accession numbers of whole genome sequences (reported in Genbank) of Bacillales analyzed in this study. (XLSX 23 kb) (XLSX 23 kb)

Additional file 2: Table S2. Characteristics of ribosomally synthesized antimicrobial peptides of Bacillales. (DOCX 201 kb)

Abbreviations
LAB: Lactic acid bacteria; LAP(φ): Linear azole-containing peptide(s); LPs: Lipopeptides; MRSA: Methicillin-resistant Staphylococcus aureus; NRPs: Non-ribosomally synthesized peptides; NRPS: Non-ribosomal peptide synthetases; PKs: Polyketides; PKS: Polyketide synthetases; PTM(s): Posttranslational modification(s); RiPPs: Ribosomally produced and posttranslationally modified peptides; SAM: S-adenosylmethionine; SKF: Sporulation killing factor

Acknowledgements
We thank Dr. Anne de Jong for his help with phylogenetic tree construction. Dr. Chunxu Song is acknowledged for critical reading of the manuscript. Xin Zhao was supported by the China Scholarship Council (no. 201406250069).

Funding
Xin Zhao holds a China Scholarship Council Studentship with the University of Groningen.

Availability of data and materials
Genome mining work was done base on the whole genome sequences of 328 strains of Bacillales listed in Additional file 1: Table S1, which can be obtained from NCBI Genome database by corresponding accession numbers (http://www.ncbi.nlm.nih.gov/genome). All the other data supporting the findings is contained within the manuscript.

Authors’ contributions
XZ collected and analyzed the genome mining data. XZ and Prof. OPK wrote the manuscript together, and Prof. OPK revised the manuscript. All authors read and approved the final manuscript.

Competing interests
The authors declare that they have no competing interests.

Consent for publication
Not applicable.

Ethics approval and consent to participate
Not applicable.

Author details
1 Department of Molecular Genetics, University of Groningen, Nijenborgh 7, Groningen 9747AG, The Netherlands. 2 School of Chemical Engineering and Technology, Tianjin University, Tianjin 300072, People’s Republic of China.

Received: 2 July 2016 Accepted: 27 October 2016
Published online: 07 November 2016

References
1. Nicholson WL. Roles of Bacillus endophores in the environment. Cell Mol Life Sci. 2002;59:410–6.
2. Lucera A, Costa C, Conte A, Del Nobile MA. Food applications of natural antimicrobial compounds. Front Microbiol. 2012;3:287.
3. Sumi CD, Yang BW, Yeo I-C, Hahm YT. Antimicrobial peptides of the genus Bacillus: a new era for antibiotics. Can J Microbiol. 2015;61(2):93–103.
4. Abriouel H, Franz CM, Ben Omar N, Galvez A. Diversity and applications of Bacillus bacteriocins. FEMS Microbiol Rev. 2011;35(1):201–32.
5. Finken R, Marahiel MA. Biosynthesis of nonribosomal peptides. Annu Rev Microbiol. 2004;58:453–88.
6.ICKER P, Antibiotic Compounds from Bacillus: Why are they so Amazing? Am J Biotech Biotechnol. 2012;8(1):38–43.
7. van Heel AJ, de Jong A, Montalban-Lopez M, Kok J, Kuipers OP, BAGEL3: Automated identification of genes encoding bacteriocins and nonribosomal posttranslationally modified peptides. Nucleic Acids Res. 2013;41:1448–53.
8. Medema MH, Blin K, Cimermancic P, de Jager V, Zakrzewski P, Fischbach MA, et al. antiSMASH: rapid identification, annotation and analysis of secondary metabolite biosynthesis gene clusters in bacterial and fungal genome sequences. Nucleic Acids Res. 2011;39:339–46.
57. Van Belkum MJ, Martin-Visscher LA, Vederas JC. Structure and genetics of circular bacteriocins. Trends Microbiol. 2011;19(8):411–8.
58. Gonzalez C, Langdon GM, Brux M, Galvez A, Valdivia E, Maqueda M, et al. Bacteriocin AS-48, a microbial cyclic polypeptide structurally and functionally related to mammalian NKG2y. Proc Natl Acad Sci U S A. 2000;97(2):1121–6.
59. Scholz R, Vater J, Budhiraja A, Wang Z, He Y, Dietel K, et al. Amylocyclin, a novel circular bacteriocin produced by Bacillus amyloliquifaciens. FEMS Bacteriol. 2014;196(10):1842–5.
60. Wirawan RE, Swanson KM, Kleffmann T, Jack RW, Tagg JR. Uberolysin, a novel cyclic bacteriocin produced by Streptococcus uberis. Microbiology. 2007;153(S5):1619–24.
61. Grande Burgos MJ, Pulido RP, Del Carmen Lopez Aguayo M, Galvez A, Lucas R. The cyclic antibacterial peptide enterocin A548: isolation, mode of action, and possible food applications. Int J Mol Sci. 2014;15(2):2270–27.
62. Kawai Y, Kempermann R, Kok J, Saito T. The circular bacteriocins gassericin A and circularin A. Curr Pept Pept Sci. 2004;5(5):393–8.
63. Borromeo J, Brede DA, Skauge M, Dep CB, Herranz C, Nes IF, et al. Characterization of garvicin ML, a novel circular bacteriocin produced by Lactococcus garvieae DC-C4S, isolated from mallard ducks (Anas platyrhynchos). Appil Environ Microbiol. 2011;77(11):3697–370.
64. Azevedo AC, Bento CB, Ruiz JC, Queiroz MV, Mantovani HC. Isolation of the novel precursor-centric genome-mining approach for lasso peptide discovery. Proc Natl Acad Sci U S A. 2010;107(20):9352–7.
65. Yang X, van der Donk WA. Ribosomally synthesized and post-translationally modified peptide natural products: new insights into the role of leader and core peptides during biosynthesis. Chemistry. 2013;19(24):7662–77.
66. Flühe L, Marahel MA. Radical S-adenosylmethionine enzyme catalyzed thioether bond formation in sactipeptide biosynthesis. Curr Opin Chem Biol. 2013;17(4):605–12.
67. Kawulka K, Sprules T, McKay RT, Mercier P, Diaper CM, Zuber P, et al. Thiopeptide antibiotics: retrospective and recent advances. Mar Drugs. 2014;12(3):1377–51.
68. Bosiers A, Walsh CT, Acker MG. Genetic interception and structural characterization of thioriboflavin, the precursor of the antibiotic thiorflavin. Proc Natl Acad Sci U S A. 2003;100(9):5879–84.
69. Liu Z, Budhiraja A, Wang P, Shi H, Fang J, Borris R, et al. The highly modified microcin-like bacteriocin microcin J25 is associated with nematicidal activity of Bacillus amyloliquefaciens. J Bacteriol. 2011;193(3):215–24.
70. Heddle JS, Blance SJ, Zamble DB, Holfelder F, Miller DA, Wetzell LM, et al. The antibiotic microcin B17 is a DNA gyrase poison: characterization of the mode of inhibition. J Mol Biol. 2001;307(5):1223–34.
71. Cow CL, Dorphozhri JR, Mitchell DA. The genomic landscape of ribosomal peptides containing thiazole and oxazole heterocycles. BMC Genomics. 2015;16(1):1–16.
72. Lee SW, Mitchell DA, Markley AL, Hersler ME, Gonzalez D, Wohlrab A, et al. Discovery of a widely distributed toxin biosynthetic gene cluster. Proc Natl Acad Sci U S A. 2008;105(15):5879–84.
73. Nizet V, Beall B, Bant DJ, Datta V, Kilburn L, Low DE, et al. Genetic locus for streptolysin S production by group A Streptococcus. Infect Immun. 2000;68(7):4245–54.
74. Borrero J, Brede DA, Skauge M, Dep CB, Herranz C, Nes IF, et al. Characterization of garvicin ML, a novel circular bacteriocin produced by Lactococcus garvieae DC-C4S, isolated from mallard ducks (Anas platyrhynchos). Appil Environ Microbiol. 2011;77(11):3697–370.
75. Bolhuis A, Venema G, Clardy J, Marahel MA. Thirteen post-translational modifications convert a 14-residue peptide into the antibiotic thiocillin. Proc Natl Acad Sci U S A. 2009;106(8):2549–53.
76. Shoji H, Hinoo H, Wakisaka Y, Kozumi K, Mayama M. Isolation of three new antibacterial peptides, thiocillins I, II and III, related to micrococcin P. Studies on antibiotics from the genus Bacillus. VIII. J Antibiot (Tokyo). 1976;29(4):366–74.
77. Serizawa M, Kodama K, Yamamoto H, Kobayashi K, Ogasawara N, Sekiguchi M, et al. Functional analysis of thiopeptide cyclization precursors from Bacillus subtilis. Angew Chem Int Ed. 2013;52(36):9518–21.
78. Liu Z, Budhiraja A, Wang P, Shi H, Fang J, Borris R, et al. The highly modified microcin-like bacteriocin microcin J25 is associated with nematicidal activity of Bacillus amyloliquefaciens. J Bacteriol. 2011;193(3):215–24.
79. Heddle JS, Blance SJ, Zamble DB, Holfelder F, Miller DA, Wetzell LM, et al. The antibiotic microcin B17 is a DNA gyrase poison: characterization of the mode of inhibition. J Mol Biol. 2001;307(5):1223–34.
80. Cox CL, Dorphozhri JR, Mitchell DA. The genomic landscape of ribosomal peptides containing thiazole and oxazole heterocycles. BMC Genomics. 2015;16(1):1–16.
81. Lee SW, Mitchell DA, Markley AL, Hersler ME, Gonzalez D, Wohlrab A, et al. Discovery of a widely distributed toxin biosynthetic gene cluster. Proc Natl Acad Sci U S A. 2008;105(15):5879–84.
82. Nizet V, Beall B, Bant DJ, Datta V, Kilburn L, Low DE, et al. Genetic locus for streptolysin S production by group A Streptococcus. Infect Immun. 2000;68(7):4245–54.
83. Stepper J, Sashiri S, Loo TS, Preston JC, Novak P, Man P, et al. Cysteine S-glycosylation, a new post-translational modification found in glycopeptidic bacteriocins. FEBS Lett. 2011;585(46):645–50.
84. Hsieh YS, Wilkinson BL, O’Connell MR, Mackay JP, Matthews JM, Payne RJ. Synthesis of the bacteriocin glycopeptide sublancin 168 and S-glycosylated variants. Org Lett. 2012;14(7):1910–3.
85. Oman TJ, Boetther JM, Wang H, Okalbe XN, van der Donk WA. Sublancin is not a lantibiotic but an S-linked glycopeptide. Nat Chem Biol. 2011;7(7):228–30.
86. Paik SH, Chakicherla A, Hansen JN. Identification and characterization of the structural and transporter genes governing the chemical and biological properties of, sublancin 168, a novel lantibiotic produced by Bacillus subtilis 168. J Biol Chem. 1998;273(36):23134–42.
87. Bolhuis A, Venema G, Quax WJ, Bro H, van Dijl JM. Functional analysis of paralogous thiol-disulfide oxidoreductases in Bacillus subtilis. J Biol Chem. 1999;274(35):24351–8.
88. Serizawa M, Kodama K, Yamamoto H, Kobayashi K, Ogasawara N, Sekiguchi M, et al. Functional analysis of the YvrGFr two-component system of Bacillus subtilis: identification of the regulated genes by DNA microarray and northern blot analyses. Biosci Biotechnol Biochem. 2005;69(1):2155–69.
89. Weber W, Fischl W, Hochuli E, Kupfer E, Weibel HK. Anti-infection activity, fermentation, isolation and biological activity. J Antibiot (Tokyo). 1991;44(2):164–71.
90. Hegemann JD, Zimmermann M, Kie X, Marshall MA, Lasso peptides: an intriguing class of bacterial natural products. Acc Chem Res. 2015;48(7):1909–19.
91. Makimov MO, Pletser I, Link AJ. Precursor-centric genome-mining approach for lasso peptide discovery. Proc Natl Acad Sci U S A. 2012;109(38):15223–8.
92. Makimov MO, Pan SJ, James Link A. Lasso peptides: structure, function, biosynthesis, and engineering. Nat Prod Rep. 2012;29(9):996–1006.
93. Makimov MO, Link AJ. Discovery and characterization of an isopeptidase that linearizes lasso peptides. J Am Chem Soc. 2013;135(32):12038–47.
94. Solbiati JO, Ciaccio M, Farias RN, González-Pastor JE, Moreno F, Salomón RA. Sequence analysis of the four plasmid genes required to produce the circular peptide antibiotic microcin J25. J Bacteriol. 1999;181(8):2659–62.
154. Huang Z, Hu Y, Shou L, Song M, Isolation and partial characterization of cyclic lipopeptide antibiotics produced by Paenibacillus brevis. B. BMC Microbiol. 2013;13:87.

155. Qian CD, Wu XC, Teng Y, Zhao WP, Li O, Fang SG, et al. Battacin (Octapeptin BS), a new cyclic lipopeptide antibiotic from Paenibacillus polymyxa active against multidrug-resistant Gram-negative bacteria. Antimicrob Agents Chemother. 2012;56(3):1458–65.

156. Azevedo EC, Rios EM, Fukushima K, Campos-Takai GM. Bacitracin production by a new strain of Bacillus subtilis. Extraction, purification, and characterization. Appl Biochem Biotechnol. 1993;42(1):1–7.

157. Ducluzeau R, Dubos F, Reilbaud P, Abrams GD. Inhibition of Cotidistrum perfringens by an antibiotic substance produced by Bacillus licheniformis in the digestive tract of gnotobiotic mice: effect on other bacteria from the digestive tract. Antimicrob Agents Chemother. 1976(1):20–2.

158. Ozcengiz G, Ogulur I. Biochemistry, genetics and regulation of bacilysin biosynthesis and its significance more than an antibiotic. New Biotechnol. 2015;32(6):612–9.

159. Borisova SA, Circello BT, Zhang JK, van der Donk WA, Metcalf WW. Presentation of the AT family. Angew Chem Int Ed. 2007;46(43):8195–8.

160. Barsby T, Kelly MT, Andersen RJ. Tupuseleiamides and basiliskamides, new macrocyclic lactones from Paenibacillus sp. FZB42. J Nat Prod. 2007;70(9):1417–23.

161. Gustafson K, Roman M, Fenical W. The macrolactins, a novel class of antiviral and antibiotic compounds from the deep-sea marine bacterium. J Am Chem Soc. 1999;121(19):9479–89.

162. Huang Z, Hu Y, Shou L, Song M. Isolation and partial characterization of Bacillus subtilis ATCC6633. Chem Biol. 2010;17(1):28–37.

163. Lee JY, Passalacqua KD, Hanna PC, Sherman DH. Regulation of petribactin and bacillibactin biosynthesis in Bacillus anthracis under iron and oxygen variation. Plos One. 2011;6(6):e20777.

164. Liao J, Zeng J, Wang H, Lin Z, Gao X, et al. Structural and biological activity. J Antibiot (Tokyo). 1995;48(9):997–1003.

165. Knüttel A, Krause M, Marahiel MA. Gramicidin S biosynthesis operon – complete nucleotide sequence and biochemical characterization of functional internal adenylation domains. J Bacteriol. 1997;179(21):6843–50.

166. Wu XC, Qian CD, Fang HH, Wen YP, Zhou JY, Zhan ZJ, et al. Battacin – a cyclic lipopeptide antibiotic. Microb Biotechnol. 2011;4(4):491–502.

167. Barsby T, Kelly MT, Andersen RJ. Tupuseliamides and basiliskamides, new acyldepsipeptides and antifungal polyketides produced in culture by a Bacillus laterosorus isolate obtained from a tropical marine habitat. J Nat Prod. 2002;65(10):1447–51.

168. Patel PS, Huang S, Fisher S, Pirnik D, Aklonis CJ, Dean L, et al. Bacillamide, a novel inhibitor of proacetylaminobutyric acid production produced by Bacillus subtilis: production, taxonomy, isolation, physico-chemical characterization and biological activity. J Antibiot (Tokyo). 1995;48(9):997–1003.

169. Molendhauer J, Chen XH, Borriss R, Piel J. Biosynthesis of the antibiotic bacilamycin, the product of a giant polyketide synthase complex of the trans-AT family. Angew Chem Int Ed. 2007;46(43):8195–7.

170. Wu L, Wu H, Chen L, Yu X, Borriss R, Gao X. Difffacin and bacilysin from Bacillus amyloliquefaciens FZB42 have antibacterial activity against Xanthomonas oryzae rice pathogen. Sci Rep. 2015;5:12975.

171. Chen XH, Vater J, Piel J, Franke P, Scholz R, Schneider K, et al. Structural and functional characterization of three polyketide synthase gene clusters in Bacillus amyloliquefaciens FZB 42. J Bacteriol. 2006;188(11):4024–36.

172. Gustafson K, Roman M, Fenical W. The macrolactins, a novel class of antiviral and antibiotic compounds from the deep-sea marine bacterium. J Am Chem Soc. 1989;111(19):7519–24.

173. Schneider K, Chen XH, Vater J, Franke P, Nicholson G, Borriss R, et al. Macrolactin is the Polyketide Biosynthesis Product of the pk2 Cluster of Bacillus amyloliquefaciens FZB42. J Nat Prod. 2007;70(9):1417–23.