

RESEARCH ARTICLE

Open Access

In silico analysis highlights the frequency and diversity of type 1 lantibiotic gene clusters in genome sequenced bacteria

Alan J Marsh^{1,2}, Orla O'Sullivan¹, R Paul Ross^{1,3*}, Paul D Cotter^{1,3*}, Colin Hill^{2,3}

Abstract

Background: Lantibiotics are lanthionine-containing, post-translationally modified antimicrobial peptides. These peptides have significant, but largely untapped, potential as preservatives and chemotherapeutic agents. Type 1 lantibiotics are those in which lanthionine residues are introduced into the structural peptide (LanA) through the activity of separate lanthionine dehydratase (LanB) and lanthionine synthetase (LanC) enzymes. Here we take advantage of the conserved nature of LanC enzymes to devise an *in silico* approach to identify potential lantibiotic-encoding gene clusters in genome sequenced bacteria.

Results: In total 49 novel type 1 lantibiotic clusters were identified which unexpectedly were associated with species, genera and even phyla of bacteria which have not previously been associated with lantibiotic production.

Conclusions: Multiple type 1 lantibiotic gene clusters were identified at a frequency that suggests that these antimicrobials are much more widespread than previously thought. These clusters represent a rich repository which can yield a large number of valuable novel antimicrobials and biosynthetic enzymes.

Background

Bacteriocins are bacterially produced peptide antibiotics. Two major classes of gram-positive bacteriocins have been recognised, Class I undergo significant post-translationally modifications while the Class II are unmodified [1,2]. The majority of the class I bacteriocins are lantibiotics; small peptides containing internal bridges resulting from the formation of (β -methyl)lanthionine residues. The structural gene encodes a ribosomally synthesised precursor prepeptide which is generically named LanA. This prepeptide contains a leader sequence at the N-terminus, which is ultimately cleaved, and a propeptide at the C-terminus. Many or all of the serine and threonine residues within the propeptide are dehydrated to form dehydroalanine (Dha) and dehydrobutyrine (Dhb), respectively. When these modified residues interact with an intrapeptide cysteine, a thioether bond is formed resulting in the formation of lanthionine (Lan, from Dha) or β -methyl lanthionine (meLan, from Dhb).

The lantibiotics and lantipeptides (lanthionine containing peptides which lack antimicrobial activity) can be divided into four groups according to the nature of the enzymes which catalyse (me)Lan formation [3]. In the case of type 1 lantibiotics two enzymes are involved; LanB, the lanthionine dehydratase which catalyses the dehydration of the amino acids, and LanC, the lanthionine synthetase which catalyses thioether formation. Type 2 lantibiotics contain a single LanM enzyme which performs both functions. Type 3 and 4 are lantipeptides which are also catalysed by distinct enzymes such as the RamC-like and LanL enzymes [4,5]. The type 1 and 2 lantibiotics can also be further subdivided on the basis of the amino acid sequence of the unmodified propeptide. In the case of the type 1 lantibiotics, five such subgroups have been identified, each of which is named after the corresponding prototypical lantibiotic; the nisin-like, epidermin-like, Pep5-like, streptin-like and planosporocin-like lantibiotic subgroups [2,6]. The nisin-like group is named for nisin A, which is the most extensively studied bacteriocin and is currently sold in more than 50 countries as a food preservative [7]. In addition to LanA, B and C, other proteins involved in

* Correspondence: paul.ross@teagasc.ie; paul.cotter@teagasc.ie

¹Teagasc, Moorepark Food Research Centre, Fermoy, Cork, Ireland
Full list of author information is available at the end of the article

the the production of nisin A and other type 1 lantibiotics include LanP, a serine protease that cleaves the leader from the propeptide; LanT, an ABC transporter responsible for the transport of the lantibiotic precursor across the cell membrane; LanIEFG encode proteins involved in immunity and; LanK, a histidine kinase and LanR, a response regulator, that together operate as a two-component regulatory system. LanD enzymes, such as that responsible for the oxidative decarboxylation of C-terminal cysteines in epidermin [8], are less common.

Given the broad antimicrobial spectrum of many lantibiotics, the possibility of applying lantibiotics in a medicinal capacity has become the subject of much attention. This is supported by an enhanced understanding of their mechanisms of action [9] and the dearth of novel antibiotics. Of the type 1 lantibiotics, nisin, mutacin and planosporicin have been shown to be active against multi-drug resistant gram-positive pathogens [10-12], Pep5 and epidermin inhibit *Staphylococcus epidermidis* adhesion to catheters [13] and epidermin and gallidermin are active against *Propionibacterium acnes* [14]. Other lantibiotics, or their producer strains, have been used as food preservatives and as oral and gastrointestinal antimicrobials/probiotics [15-17]. As a consequence of this increased interest in lantibiotics, a concerted effort has taken place to identify new and improved forms of these peptides. Culture based screening strategies have in the past been responsible for the identification of lantibiotics produced by bacteria isolated from diverse microbial niches including the oral cavity, intestine, soil, kefir grains and milk [12,18-21]. Recently, an alternative means of identifying novel lantibiotics has emerged as a consequence of the increasing generation and availability of genomic and metagenomic sequence data. The availability of such information has recently led to the identification of the type 1 epidermin-like lantibiotic, Bsa [22] as well as type 2 lantibiotics such as haloduracin [23,24], lichenicidin [25,26], as well as a range of cyanobacteria-associated lantipeptides [27]. This has prompted the development of on-line tools and repositories such as BAGEL and BACTIBASE to facilitate such screening strategies [28-31]. Notably, although an *in-silico* screen for *lanM* genes has recently resulted in the identification of 61 novel type 2 lantibiotic-like gene clusters [25], a corresponding screen for type 1 lantibiotics has not yet been described. Here we address this issue by screening for clusters containing genes homologous to the nisin A biosynthetic genes *nisB* (representing *lanB*) and *nisC* (representing *lanC*). In each case, the regions flanking the newly identified *lanB/lanC* genes were subjected to further *in silico* analysis to determine if they are potential lantibiotic/lantipeptide-associated gene clusters. This included a search of nearby open reading frames (orfs) which might

encode a corresponding LanA, defined as being of relatively short length (approx 60 amino acids) and containing an uneven distribution of cysteine, threonine and serine amino acids within the propeptide region. Using this approach, 27 novel type 1 lantibiotic/lantipeptide-encoding clusters were identified. Subsequent screening using the newly identified LanA, B and C homologs as driver sequences revealed a further 22 gene clusters, resulting in a total of 49 putative novel type 1 lantibiotic clusters. Significantly, many of these clusters are present in species, genera and phyla not previously associated with lantibiotic/lantipeptide production and are predicted to encode peptides which represent completely new type 1 subclasses.

Results and Discussion

In silico screen for *lanC* genes

An *in silico* screen for LanC homologues, using the NisC sequence as a driver, resulted in the identification of 56 homologues. Of these 7 have previously been associated with lantibiotic production, 11 were orphan homologs (in that no other lantibiotic-associated genes were identified in close association) (Table 1), 9 were encoded within a cluster in which no *lanA* could be detected (Table 2) and one cluster contained two LanCs (but no structural peptide). The remaining 27 potential lantibiotic/lantipeptide-encoding gene clusters all contained putative *lanA*, *B* and *C* genes (Table 3). The genes flanking the 27 novel *lanC*-like genes were subjected to further bioinformatic analysis to determine the presence of other orfs that share homology with genes linked to lantibiotic production or immunity. While these 27 clusters are the primary focus of this *in silico* analysis, the sequences of the newly identified LanA, B and C proteins associated with these clusters were in turn used for further *in silico* screens. This approach uncovered an additional 22 clusters (Table 4) that were also predicted to be novel lantibiotic/lantipeptide-encoding clusters, thereby yielding a total of 49 novel type 1 clusters.

All except one of the 27 gene clusters revealed following the initial screen were located within the genomes of *Firmicutes* and *Actinobacteria*. The exception was *Chitinophaga pinensis* DSM 2588 of the phylum *Bacteroidetes*. Of the other 26, the genera most commonly associated with lantibiotic production were *Bacillus*, *Geobacillus*, *Clostridium*, *Enterococcus*, *Streptococcus*, *Frankia* and *Streptomyces*. In many cases the novel clusters associated with a specific genus, such as those found on the *Streptomyces* and *Frankia* genomes, showed at least some similarity to each other. It was also noted that several of the genomes in which a cluster was located also contained an additional cluster(s) (Table 3), or other genes predicted to encode additional

Table 1 A selection of bacterial genomes in which isolated genes encoding LanA, LanB or LanC homologs were identified

Clusterless Homologs	Accession No.	LanA only	LanB only	LanC only
<i>Anoxybacillus flavithermus</i> WK1	NC_011567		Aflv_2440	
<i>Bacillus</i> sp. B14905	NZ_AAAXV000000000		BB14905_21668	
<i>Paenibacillus larvae</i> subsp. <i>larvae</i> BRL-230010	NZ_AARF000000000			Plarl_010100024193
<i>Lactobacillus crispatus</i> MV-1A-US	NZ_ACOG000000000			HMPREF0507_00422
<i>Lactobacillus crispatus</i> JV-V01	NZ_ACKR000000000			HMPREF0506_0642
<i>Haliangium ochraceum</i> DSM 14365	NC_013440			Hoch_3102
<i>Haliangium ochraceum</i> DSM 14365	NC_013440			Hoch_4144
<i>Haliangium ochraceum</i> DSM 14365	NC_013440		Hoch_0066	
<i>Pedobacter</i> sp. BAL39	NZ_ABCM000000000		PBAL39_02527	
<i>Peptoniphilus lacrimalis</i> 315-B	NZ_ADDO000000000		HMPREF0628_0526	
<i>Peptoniphilus lacrimalis</i> 315-B	NZ_ADDO000000000		HMPREF0628_0527	
<i>Lactococcus lactis</i> subsp. <i>lactis</i> KF147	NC_013656			
<i>Frankia alni</i> ACN14a	NC_008278		FRAAL2701	
<i>Frankia</i> sp. Cc13	NC_007777			Francci3_0205
<i>Frankia</i> sp. Cc13	NC_007777			Francci3_3997
<i>Peptoniphilus lacrimalis</i> 315-B	NZ_ADDO000000000		HMPREF0628_0527	
<i>Bifidobacterium longum</i> subsp. <i>infantis</i> ATCC 15697	NC_011593			
<i>Streptomyces</i> sp. AA4	NZ_ACEV000000000			
<i>Streptococcus pneumoniae</i> CGSP14	NC_010582			
<i>Saccharopolyspora erythraea</i> NRRL 2338	NZ_ABFV000000000			SACE_4959
<i>Kordia algicida</i> OT-1	NZ_ABIB000000000			KAOT1_07113
<i>Streptococcus dysgalactiae</i> GGS_124	NC_012891			SDEG_0295
<i>Microscilla marina</i> ATCC 23134	NZ_AAWS000000000		M23134_07394	
<i>Microscilla marina</i> ATCC 23134	NZ_AAWS000000000		M23134_05752	
<i>Microscilla marina</i> ATCC 23134	NZ_AAWS000000000		M23134_07275	
<i>Microscilla marina</i> ATCC 23134	NZ_AAWS000000000		M23134_01545	
<i>Microscilla marina</i> ATCC 23134	NZ_AAWS000000000			M23134_07404
<i>Streptomyces</i> sp. Mg1	NZ_ABJF000000000		SSAG_05771	
<i>Frankia</i> sp. Eul1c	NZ_ADDL000000000	FraEul1cDRAFT_6351		
<i>Geobacillus</i> sp. Y412MC52	NZ_ACNM000000000	GYMC52DRAFT_3129		
<i>Geobacillus</i> sp. Y412MC61	NC_013411	GYMC61_1158		
<i>Streptococcus pyogenes</i> M1 GAS	NC_002737	SPy_1083		
<i>Streptomyces griseus</i> subsp. <i>griseus</i> NBRC 13350	NC_010572			SGR_6574
<i>Clostridium kluveri</i> DSM 555	NC_009706			CKL_3505
<i>Spirosoma linguale</i> DSM 74	NC_013730		Slin_0903	
<i>Spirosoma linguale</i> DSM 74	NC_013730		Slin_2131	

LanA, B or C proteins (Table 2), elsewhere in the genome. The 27 clusters are described below and are grouped according to the phylum and genus of the associated strain.

Type 1 lantibiotic gene clusters in Actinobacteria
Identification of novel Frankia-associated lantibiotic gene clusters

The *Frankia* are nitrogen-fixing, root nodule-forming filamentous *Actinobacteria* that live in symbiosis with actinorhizal plants. All species of *Frankia* are closely related [32]. To date, four *Frankia* genomes have been sequenced, i.e. *Frankia alni* ACN14a, *Frankia* sp.

EAN1pec, *Frankia* sp. Cc13 and *Frankia* sp. EUI1c, and although no *Frankia*-associated bacteriocins have previously been reported, a number of predicted lantibiotic clusters can be found in each case (Figure 1) in addition to a number of apparently LanB- and LanC-encoding genes which do not have an accompanying *lanA* (Table 2). This latter phenomenon could be a result of the frequent rearrangements which occur in *Frankia* strains [32]. Of the clusters identified, many resemble clusters associated with another genus of *Actinobacteria*, the *Streptomyces*.

Frankia alni ACN14a: The genome sequence of *F. alni* ACN14a [32] contains one complete cluster, *F. alni*

Table 2 Gene clusters encoding LanB and LanC, but not LanA, homologs

Species (Cluster No.)	Accession No.	LanB	LanC
<i>Frankia alni</i> ACN14a I	NC_008278	FRAAL2701	FRAAL2700
<i>Frankia</i> sp. Cc13 IV	NC_007777	Francci3_2033	Francci3_2032
<i>Frankia</i> EAN1pec II	NC_009921	Franean1_2799	Franean1_2800
<i>Frankia</i> sp. Eul1c II	NZ_ADDL000000000	FraEul1cDRAFT_6786	FraEul1cDRAFT_6785
<i>Bacillus clausii</i> KSM-K16	NC_006582	ABC3559	ABC3558
<i>Clostridium cellulovorans</i> 743B	NZ_ACPD000000000	ClocelDRAFT_0447	ClocelDRAFT_0445/_0452
<i>Bacillus cereus</i> AH1273	NZ_ACMT000000000	bcere0030_58380	bcere0030_58400
<i>Bacillus thuringiensis</i> serovar berliner ATCC 10792	NZ_ACNF000000000	bthur0008_53920	bthur0008_53930
<i>Bacillus thuringiensis</i> IBL 200	NZ_ACNK000000000	bthur0013_59170	bthur0013_59180
<i>Streptococcus pyogenes</i> MGAS9429	NC_008021	MGAS9429_Spy0926	MGAS9429_Spy0924
<i>Catenulispora acidiphila</i> DSM 44928	NC_013131	Mentioned; Caci_4205	Caci_4204
<i>Frankia</i> sp Cc13 V	NC_007777	Francci3_3530	Francci3_3531
<i>Microscilla marina</i> ATCC 23134	NZ_AAWS000000000	M23134_05752	M23134_05756
<i>Staphylococcus capitis</i> SK14	NZ_ACFR000000000	STACA0001_2327	STACA0001_2326
<i>Streptomyces</i> sp. Mg1 I	NZ_ABJF000000000	SSAG_03540	SSAG_03541

Table 3 Bacterial genomes in which 27 uncharacterised type 1 lantibiotic clusters were identified following a NisC-led *in silico* screen

Species (Cluster No.)	Accession No.
<i>Frankia alni</i> ACN14a (II)	NC_008278
<i>Frankia</i> sp Cc13 (I)	NC_007777
<i>Frankia</i> sp Cc13 (II)	NC_007777
<i>Frankia</i> sp Cc13 (III)	NC_007777
<i>Frankia</i> EAN1pec (I)	NC_009921
<i>Frankia</i> EAN1pec (III)	NC_009921
<i>Frankia</i> sp. Eul1c (I)	NZ_ADDL000000000
<i>Salinispora arenicola</i> CNS-205*	NC_009953
<i>Stackebrandtia nassauensis</i> DSM 44728 (I)	NC_013947
<i>Stackebrandtia nassauensis</i> DSM 44728 (II)	NC_013947
<i>Streptomyces clavuligerus</i> ATCC 27064 (I)	NZ_ADGD000000000
<i>Streptomyces clavuligerus</i> ATCC 27064 (II)	NZ_ADGD000000000
<i>Streptomyces coelicolor</i> A3(2) (I)	NC_003888
<i>Streptomyces coelicolor</i> A3(2) (II)	NC_003888
<i>Streptomyces</i> sp. Mg1 (II)	NZ_ABJF000000000
<i>Streptomyces griseoflavus</i> Tu4000 (I)	NZ_ACFA000000000
<i>Streptomyces griseoflavus</i> Tu4000 (II)	NZ_ACFA000000000
<i>Streptomyces griseoflavus</i> Tu4000 (III)	NZ_ACFA000000000
<i>Bacillus cereus</i> F65185	NZ_ACMO000000000
<i>Bacillus mycoides</i> DSM 2048	NZ_ACMU000000000
<i>Clostridium perfringens</i> CPE str. F4969	NZ_ABDX000000000
<i>Enterococcus faecalis</i> Fly1	NZ_ACAR000000000
<i>Geobacillus kaustophilus</i> HTA426	NC_006510
<i>Geobacillus thermodenitrificans</i> NG80-2	NC_009328
<i>Geobacillus</i> sp. G11MC16	NZ_ABVH000000000
<i>Streptococcus thermophilus</i> LMG 18311*	NC_006448
<i>Chitinophaga pinensis</i> DSM 2588 I	NC_013132

*The existence of a lantibiotic gene cluster within these strains has been referred to briefly, [36] and [79,80] respectively, but these clusters have not been the focus of a detailed bioinformatic analysis.

ACN14a II which includes the predicted LanA prepropeptide, FRAAL6345, FRAAL6344 and FRAAL6343 (encoding a putative LanB and LanC, respectively). The LanA peptide does not resemble any of the previously characterised type 1 lantibiotic propeptides but is 79%

Table 4 Bacterial genomes in which 22 additional type 1 lantibiotic gene clusters were identified following an *in silico* screen using the LanA, B, and C homologs, corresponding to the clusters referred to in Table 3, as leader sequences

Species (Cluster No.)	Accession No.
<i>Thermomonospora curvata</i> DSM 43183	NC_013510
<i>Frankia</i> EAN1pec (IV)	NC_009921
<i>Streptomyces viridochromogenes</i> DSM 40736	NZ_ACEZ000000000
<i>Streptomyces</i> sp. SPB74	NZ_ABJG000000000
<i>Streptomyces lividans</i> TK24	NZ_ACEY000000000
<i>Catenulispora acidiphila</i> DSM 44928	NC_013131
<i>Streptomyces</i> sp. Mg1 (III)	NZ_ABJF000000000
<i>Nocardioopsis dassonvillei</i> subsp. <i>dassonvillei</i> DSM 43111	NZ_ABUI000000000
<i>Micromonospora aurantiaca</i> ATCC 27029 (I)	NZ_ADBZ000000000
<i>Micromonospora aurantiaca</i> ATCC 27029 (II)	NZ_ADBZ000000000
<i>Bacillus cereus</i> AH1272	NZ_ACMS000000000
<i>Staphylococcus aureus</i> subsp. <i>aureus</i> D139	NZ_ACSR000000000
<i>Staphylococcus aureus</i> subsp. <i>aureus</i> H19	NZ_ACSS000000000
<i>Actinomyces</i> sp. oral taxon 848	NZ_ACUY000000000
<i>Parachlamydia acanthamoebae</i> str. Hall's coccus	NZ_ACZE000000000
<i>Corynebacterium lipophiloflavum</i> DSM 44291	NZ_ACHJ000000000
<i>Staphylococcus aureus</i> A9765	NZ_ACSN000000000
<i>Chitinophaga pinensis</i> DSM 2588 (II)	NC_013132
<i>Spirosoma linguale</i> DSM 74	NC_013730
<i>Pedobacter heparinus</i> DSM 2366	NC_013061
<i>Kordia algicida</i> OT-1	NZ_ABIB000000000
<i>Microscilla marina</i> ATCC 23134	NZ_AAWS000000000

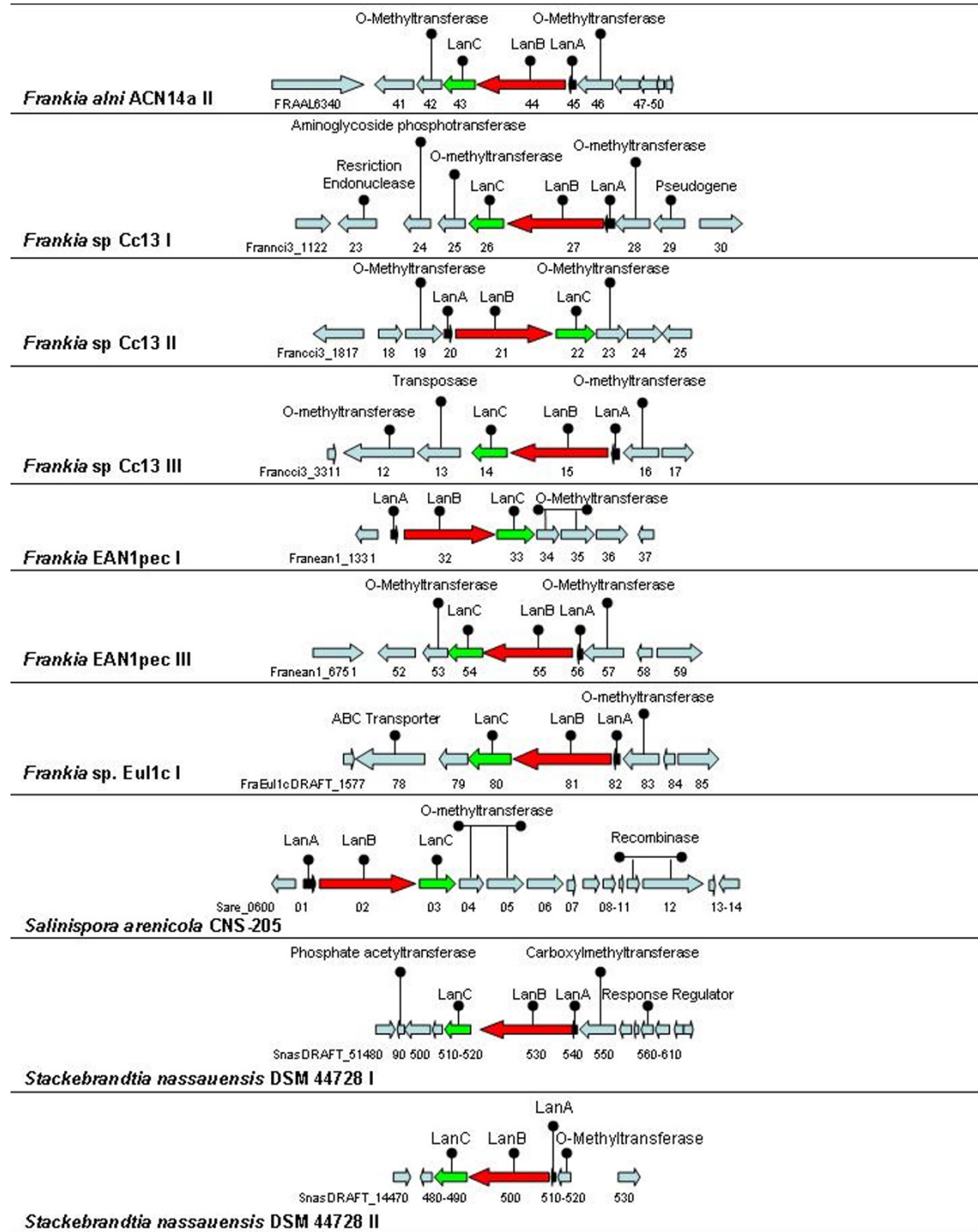


Figure 1 Diagrammatic representation of the non-streptomyces *Actinobacteria* type 1 lantibiotic operons, found in the original NisC screen, which contain genes predicted to encode a structural peptide LanA, and the modification enzymes LanB and LanC.

identical to Franean1_0057 of *Frankia* sp. EAN1pec (see below). The LanB and LanC proteins resemble those associated with other *Frankia*, as well as *Streptomyces*, clusters. The LanC protein is also 62% identical to Tcur_4648 of *Thermomonospora curvata* DSM 43183

(NC_013510), which itself appears to be encoded by an orf within a novel lantibiotic gene cluster. The ACN14a II cluster is also predicted to encode two proteins which share homology with O-methyltransferases (FRAAL6342 and FRAAL6346). O-methyltransferases contribute to the

production of a number of non-ribosomal antibiotics [33,34] and catalyse the methylation of hydroxyl group(s) on deoxysugar rings to protect the reactive hydroxyl group from undesired modifications and can alter the solubility and pharmacokinetic properties of the resulting molecule [35]. Although O-methyltransferases have not previously been associated with lantibiotic production, this study reveals that many *Actinobacteria*-associated type 1 clusters possess genes predicted to encode these enzymes.

Frankia sp. Cc13: The *Frankia sp. Cc13* genome [32] contains three gene clusters of interest (*Frankia sp. Cc13 I*, *Frankia sp. Cc13 II* and *Frankia sp. Cc13 III*). Two O-methyltransferase-encoding genes were identified in each case and it was noted that the associated LanB and LanC proteins are similar to one another and to numerous others predicted to be produced by *Frankia* and *Streptomyces* species. In contrast, there is a lack of homology between the three *lanA* genes. The *lanA* gene from cluster I was not previously annotated and was only identified following closer inspection of the DNA sequence. The cluster II-associated LanA, Francci3_1820, most closely resembles *Frankia sp. EuI1c* FraEuI1cDRAFT_6351 (69% identity) while the third, and also previously unannotated LanA appears to be one of an extended group of *Frankia*- and *Streptomyces*-associated LanAs that includes *Frankia sp. EuI1c* FraEuI1cDRAFT_1582 (56% identity).

Frankia EAN1pec: The *Frankia EAN1pec* genome [32] contains 4 putative LanB-encoding genes three of which correspond to potential lantibiotic/lantipeptide-associated gene clusters (*Frankia EAN1pec I*, *Frankia EAN1pec III* and *Frankia EAN1pec IV*) which again resemble those of *Streptomyces* and other *Frankia* species, and contain O-methyltransferase-encoding genes. Within the first cluster, a putative LanA prepropeptide, encoded by a previously unannotated orf located between Franean1_1331 and the LanB determinant, is homologous to a number of other LanAs, including Sare_0601 of *Salinispora arenicola* CNS-205 (55% identity). The cluster is also noteworthy by virtue of the presence of two LanC-encoding genes, Franean1_1333 and Franean1_1336. Within *Frankia EAN1pec III*, the LanA peptide, encoded by Franean1_6756 is 42% identical to FraEuI1cDRAFT_6351 of *Frankia sp. EuI1c* while finally, a screen using *F. alni* ACN14a FRAAL6345 as a driver led to the identification of yet another cluster (consisting of at least Franean1_0057-0055) which closely resembles cluster II of *F. alni* ACN14a II.

Frankia sp. EuI1c: *Frankia sp. EuI1c* contains a single putative lantibiotic/lantipeptide gene cluster (*Frankia sp. EuI1c I*) which again contains LanB, C and O-methyltransferase genes typical of *Frankia* and *Streptomyces* clusters. The associated LanA homolog (FraEuI1cDRAFT_1582) is

notable by virtue of being 46% identical to SSCG_03316, a known LanA of *Streptomyces clavuligerus* ATCC 27064 while a gene encoding an ABC transporter related protein (FraEuI1cDRAFT_1578) is also present.

Identification of novel *Salinispora*-associated lantibiotic gene clusters

Salinispora are marine *Actinobacteria*. There are two recognised species, *S. tropicalis* and *S. arenicola*. Representatives have been sequenced in each case and genes predicted to encode non-lantibiotic bacteriocins have been identified in both cases [36]. The existence of a putative lantibiotic/lantipeptide cluster, between Sare_0602 and Sare_0623, in the genome of *S. arenicola* CNS-205 was noted previously [36]. However, this cluster has not been the subject of a detailed bioinformatic characterisation. Our analysis reveals that Sare_0601 apparently encodes a LanA peptide which is 88% identical to that encoded by MicauDRAFT_5818 of *Micromonospora aurantiaca* ATCC 27029. The proteins encoded by Sare_0602 (LanB) and Sare_0603 (LanC) also resemble other ATCC 27029-associated proteins (encoded by MicauDRAFT_5819 (71% identity) and MicauDRAFT_5820 (75% identity)), thereby revealing an additional novel cluster in *Micromonospora*, a genus better known for its production of non-ribosomal antibiotics such as gentamycin and netamycin [37] (Table 4).

Identification of novel *Stackebrandtia*-associated lantibiotic gene clusters

Stackebrandtia are aerobic, non-motile *Actinobacteria* which have been isolated from soil [38]. There are only 2 associated species i.e. *S. albiflava* and *S. nassauensis* and *in silico* analysis of *S. nassauensis* DSM 44728 (NC_013947) reveals the presence of two similar clusters (*S. nassauensis* DSM 44728 I and *S. nassauensis* DSM 44728 II) (Figure 1). The hypothetical LanA, encoded by Snas_5416, of the first cluster showed a singular homology of 78% identity to Snas_3601 of the second cluster. The corresponding LanBs (Snas_5417 and Snas_3602) are 62% identical while the LanCs (Snas_5418 and Snas_3603) are 68% identical.

Identification of novel *Streptomyces*-associated lantibiotic gene clusters

Bacteria from the genus *Streptomyces*, comprising over 500 species, are filamentous, high G-C bacteria found frequently in soil and rotting vegetation. They are the most numerous and ubiquitous soil bacteria [39]. *Streptomyces* are also responsible for the production of over two-thirds of the clinically useful antibiotics of natural origin (e.g., neomycin, chloramphenicol) [40]. Although a number of *Streptomyces*-associated bacteriocins, such as ancovenin [41] and cinnamycin [42], have been identified, this number is relatively small considering the size of the genus. As was apparent above, our *in silico* analysis has revealed that many *Streptomyces* possess

potentially lantibiotic-encoding gene clusters which resemble those found in *Frankia*. Once again, the majority of these clusters contain O-methyltransferases (Figure 2).

***Streptomyces clavuligerus* ATCC 27064** *S. clavuligerus* is an aerobic, mesophilic *Streptomyces* sp. While there have been no previous reports of bacteriocin production by this species, two lantibiotic clusters were found to be present on the genome of *S. clavuligerus* ATCC 27064. In the first of these clusters, the associated hypothetical LanA, B and C proteins (SSCG_01498-01496) are 63%, 42% and 50% identical to the corresponding proteins of *Frankia* sp. CcI3 II. BLAST analysis of these proteins also revealed another novel cluster in *Streptomyces viridochromogenes* DSM 40736 corresponding to SvirD4_23440 (LanA; 50% identity), SvirD4_23449 (LanB; 36% identity) and SvirD4_23454 (LanB; 45% identity) (Table 4). The second *S. clavuligerus* cluster, which contains SSCG_03316 (LanA), SSCG_03317 (LanB) and SSCG_03318 (LanC), resembled clusters present in a number of other strains such as that of *Streptomyces griseus* subsp. *griseus* NBRC 13350 [43] (73%, 56% and 64% identity, respectively). BLAST analysis of these sequences also led to the identification of yet another novel cluster in *Streptomyces* sp. SPB74 (SSBG_01041 [LanA] 69% identity and SSBG_01042 [LanB] 58% identity).

***Streptomyces coelicolor* A3(2)** *Streptomyces coelicolor* A3(2) (NC_003888) is the best characterised representative of its genus [44] and was the first *Streptomyces* strain to have its genome sequenced [45]. Although bacteriocins/bacteriocin-like peptides are known to be produced by this species (e.g. the class III morphogenic peptide SapB [4]), such peptides have not previously been associated with this strain. Here BLAST analysis revealed the presence of two lantibiotic/lantipeptide clusters (*S. coelicolor* A3(2) I and *S. coelicolor* A3(2) II). The first of these clusters, containing SCO0268 (LanA), SCO0269 (LanB) and SCO0270 (LanC), very closely resembles *Streptomyces griseoflavus* Tu4000 cluster II (see below). Subsequent BLAST searches with the A3(2) cluster I-associated proteins led in turn to the discovery of an almost identical cluster in *Streptomyces lividans* TK24 which contains SSPG_07329 (LanA; 100% identity), SSPG_07328 (LanB; 99% identity) and SSPG_07327 (LanC; 100% identity) (Table 4). The second cluster, *Streptomyces coelicolor* A3(2) II, is predicted to encode two LanA peptides, (SCO6932 [43aa] and SCO6931 [59aa]) which are 97% identical to each other, as well as LanB (SCO6930) and LanC (SCO6929) proteins with homology (39-46% identity) with corresponding proteins associated with *Frankia* sp. CcI3. Such analysis also revealed another cluster of interest in the actinomycete, *Catenulispora acidiphila* DSM 44928 (NC_013131; [Table 4]).

***Streptomyces griseoflavus* Tu4000** Three lantibiotic/lantipeptide clusters were identified on the genome of this anaerobic, terrestrial *Streptomyces*. Although the LanA encoded within the first cluster (SgriT_010100000925) does not significantly resemble any other protein, the associated LanB (SgriT_010100000920) and LanC (SgriT_010100000915) proteins are homologous to the corresponding proteins of *S. coelicolor* A3(2) cluster I and Tu4000 cluster II. BLAST searches using the cluster I proteins as drivers also resulted in the identification of several additional clusters in *Nocardioopsis dassonvillei* DSM 43111 (NdasDRAFT_3161 [LanB] 30% identity), *Streptomyces* sp. Mg1 (SSAG_05771 [LanB] 37% identity) and two clusters on the genome of *M. aurantiaca* ATCC 27029 (NZ_ADBZ000000000; MicauDRAFT_5820 and MicauDRAFT_3008 [both LanBs] 35% identity). In addition to the components of the second cluster referred to above, an associated LanA (SgriT_010100036309) was also noted. In addition to the Tu4000 I and A3(2) I clusters, this cluster is also highly identical to that of *S. lividans* TK24 (SSPG_07329 [LanA] 97% identity; SSPG_07328 [LanB] 87% identity and SSPG_07327 [LanC] 89% identity). The LanA associated with the final cluster (SgriT_010100029766) again bears no homology with any other known peptides whereas the LanB (SgriT_010100029771) and LanC (SgriT_010100029776) corresponded to those of *Frankia* sp. EAN1pec II (39% and 44% identity, respectively).

Type 1 lantibiotic gene clusters in Firmicutes

Identification of novel *Bacillus*-associated lantibiotic gene clusters

Bacillus is a large and diverse genus of rod-shaped, sporulating, obligate aerobes which contains both free living and pathogenic species. A number of type 1 lantibiotics have previously been characterized in this genus (e.g. subtilin [46] and ericin [47]). The NisC-driven screen highlighted the presence of a type 1 lantibiotic cluster in the genomes of two *Bacillus* strains i.e. *Bacillus cereus* F65185 and *Bacillus mycoides* DSM 2048 (Figure 3). Bioinformatic analysis of these clusters revealed two further clusters in *B. cereus* ATCC 14579 and *B. cereus* AH1272.

***Bacillus cereus* F65185** *B. cereus* F65185 is a mesophilic bacterium sourced from a human wound containing one lantibiotic/lantipeptide cluster which is unusual in that 3 orfs separate the putative LanB and C genes and the two have a divergent orientation. The predicted LanA (bcere0025_48310) does not resemble any other known lantibiotic prepropeptides. The LanB homolog (bcere0025_48320) resembles a putative LanB associated with *Clostridium cellulovorans* 743B (ClonelDRAFT_0452, 30% identity) while the predicted LanC (bcere0025_48280) most closely resembles two further

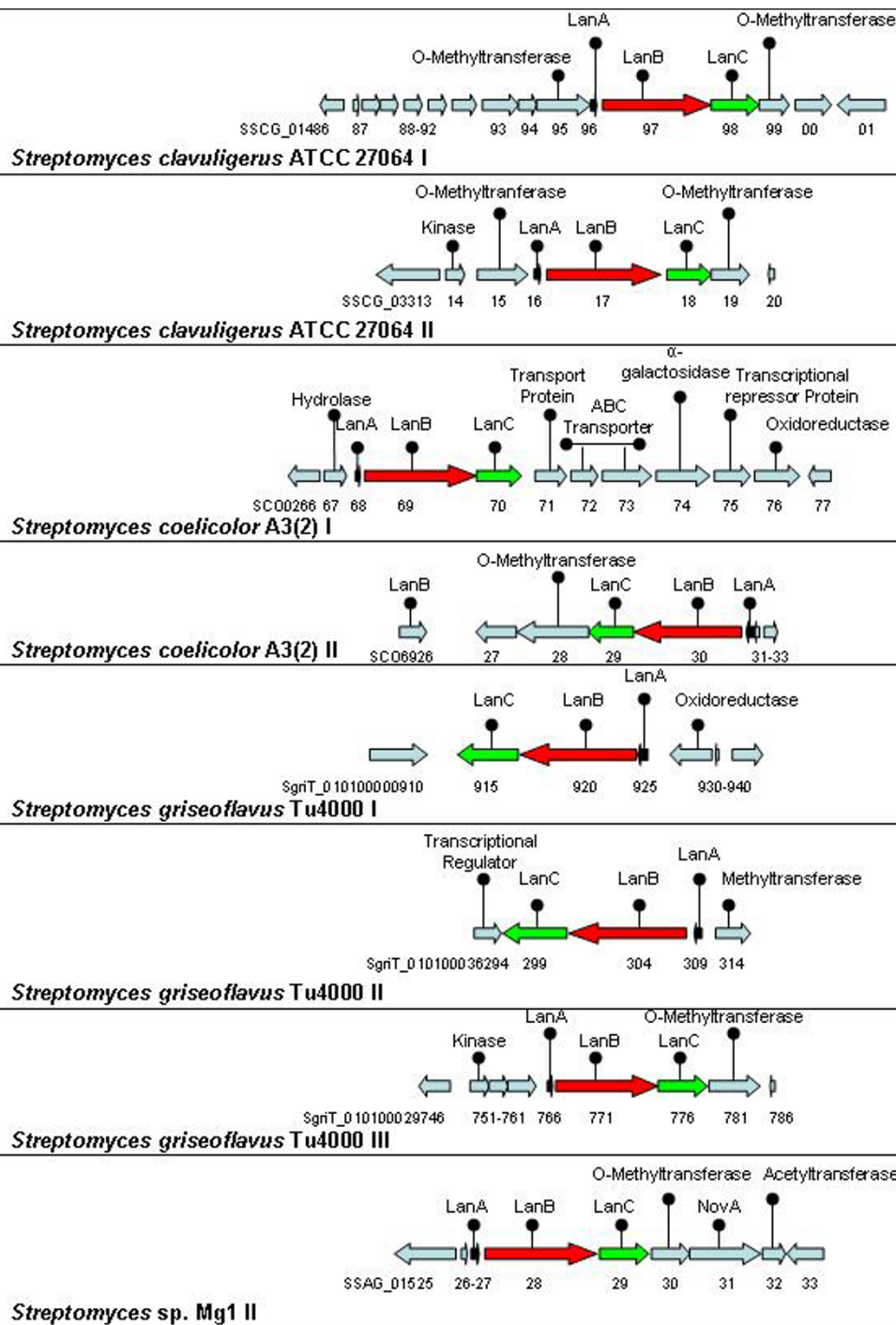


Figure 2 Diagrammatic representation of the *Streptomyces* type 1 lantibiotic operons, found in the original NisC screen, which contain genes predicted to encode a structural peptide LanA, and the modification enzymes LanB and LanC.

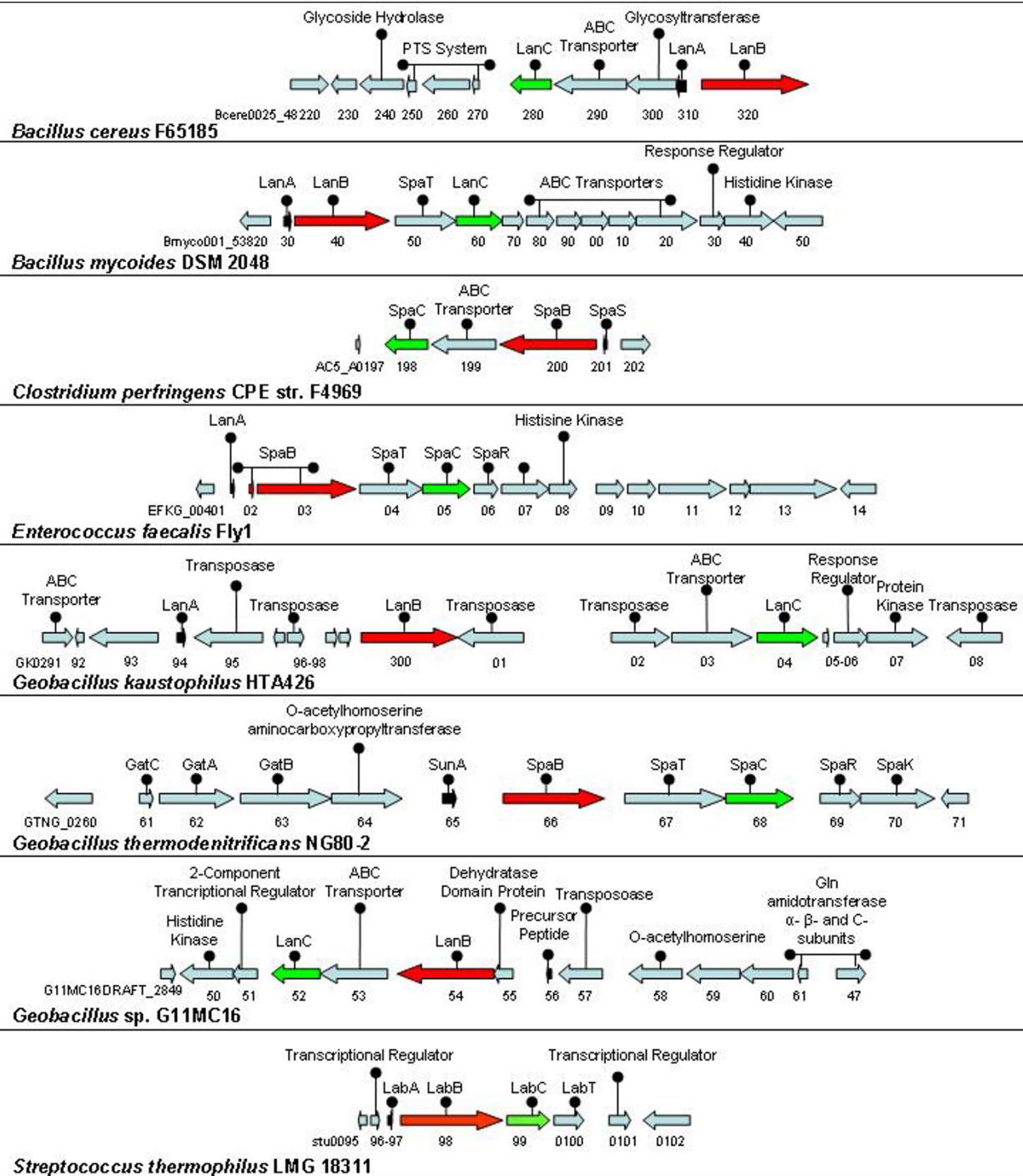


Figure 3 Diagrammatic representation of the *Firmicutes* type 1 lantibiotic operons, found in the original NisC screen, which contain genes predicted to encode a structural peptide LanA, and the modification enzymes LanB and LanC.

743B proteins (CloceLDRAFT_0452, 34% identity and CloceLDRAFT_0446, 28% identity). However the 743B strain lacks an associated LanA. Further BLAST analysis with the F65185-associated LanB highlighted the presence of a related protein within thiocillin-encoding gene cluster in *B. cereus* ATCC 14579 [48].

***Bacillus mycoides* DSM 2048** *B. mycoides* is a non-motile, non-pathogenic, saprophytic *Bacillus*, strains of which have been investigated with a view to their application as biological pesticides. Although representatives of this species have been associated with bacteriocin production [49], there are no published reports of lantibiotic-producing *B. mycoides*. The DSM 2048 genome contains a lantibiotic/lantipeptide cluster that contains putative *lanA*, *B*, *C* and other lantibiotic-associated genes which is very similar to other novel clusters in *B. cereus* AH1272 and *B. cereus* AH1273. Bmyco0001_53830 is the predicted prepropeptide and is 100% identical to the products of the *B. cereus* AH1272-associated bcere0029_28240 and bcere0029_28250. It is also 58% identical to prepropeptides associated with several *Staphylococcus*-associated Bsa lantibiotics, such as BsaA2_{RF122} of *S. aureus* RF122 [22]. It is thus apparent that the LanA is a member of the epidermin-like peptides. In addition to homologues in *B. cereus* AH1273 and AH1272, use of the DSM 2048-associated LanB and LanC sequences as drivers also surprisingly highlighted a gene cluster present in *Actinomyces* sp. oral taxon 848 (Table 4). In addition to homology with respect to LanB and LanC proteins (HMPREF0972_00932 and HMPREF0972_00933, respectively), the proposed LanA (HMPREF0972_00931; 86aa) is 43% identical to bmyco0001_53830 and bcere0029_28240 and bcere0029_28250 of *Bacillus mycoides* DSM 2048 and *Bacillus cereus* AH1272, respectively.

Identification of novel *Clostridium*-associated lantibiotic gene clusters

The *Clostridia* are gram positive anaerobic, endospore-forming *Firmicutes* of which there are approximately 100 species. These include important pathogens such as *Clostridium difficile*, *Clostridium perfringens* and *Clostridium tetani*. Several *Clostridium*-associated bacteriocins have been characterised from this genus [50-52], but no type 1 lantibiotic producers have been identified to date. Here, bioinformatic analysis revealed one *Clostridium*-associated lantibiotic/lantipeptide cluster, located on the genome of *C. perfringens* CPE str. F4969 (Figure 3).

***C. perfringens* CPE str. F4969** *C. perfringens* is one of the leading causes of food-borne illness in the developed world, usually as a result of the improper sterilization of canned foods in which endospores have germinated. It can also be responsible for wound and surgical infections [53,54]. The predicted LanA (AC5_A0201) of strain F4969 is most closely related to the LanAs of

B. mycoides DSM 2048 and *B. cereus* AH1272 (51% identity) while the proteins encoded by AC5_A0200 (LanB) and AC5_A0198 (LanC) most closely resemble proteins associated with *Geobacillus* sp. G11MC16 (G11MC16DRAFT_2954 and G11MC16DRAFT_2952; 34% and 35% identity respectively). Surprisingly, BLAST analysis also highlighted the presence of a related LanB homolog encoded within a lantibiotic/lantipeptide-like gene cluster in the genome of *Parachlamydia acanthamoebae* (phylum *Chlamydiae*) str. Hall's coccus (pah_c028o031; 25% identity). Adjacent genes of note within the *P. acanthamoebae* cluster include pah_c028o029 (LanA) and pah_c028o030 (LanC).

Identification of novel *Enterococcus*-associated lantibiotic gene clusters

The enterococci are gram positive lactic acid bacteria which are common commensal organisms in the intestines of humans but can also be pathogens. Many *Enterococcus*-associated bacteriocins (enterocins) have been identified [55]. Only one *Enterococcus*-associated lantibiotic, the type 2 peptide cytolysin, has been identified to date [56]. Here we describe genes which potentially encode the first type 1 *Enterococcus*-associated lantibiotic (Figure 3).

***Enterococcus faecalis* Fly1: *E. faecalis* Fly1** (NZ-ACAR00000000) is a non-motile, facultative anaerobe. Within its genome we identified a previously unannotated LanA determinant, through analysis of raw sequence data. The corresponding peptide is homologous to *C. perfringens* CPE str. F4969 (AC5_A0201; 68% identity), as well as a number of epidermin-like LanAs in other bacilli. The putative LanB protein is split across two orfs, EFKG_00402 (80 amino acids) and EFKG_0403 (942 amino acids), with both components most closely resembling the N-terminus of the dehydratase of *Streptococcus pyogenes* MGAS10270, MGAS10270_Spy0922. It is unclear whether the apparent frameshift in the Fly1 *lanB* is genuine or the result of a sequencing error. The LanC-like EFKG_00405, was most closely related to the corresponding protein in *G. thermodenitrificans* NG80-2 (SpaC GTNG_0268; 35% identity).

Identification of novel *Geobacillus*-associated lantibiotic gene clusters

Geobacilli are thermophilic (45-70°C), aerobic, spore-forming *Firmicutes*. They have been isolated from various terrestrial and marine environments, in geothermal, temperate and permanently cold habitats. Reclassified in 2001 [57], these bacteria are of industrial interest as sources of thermostable enzymes. Bacteriocins have been identified in *Geobacillus stearothermophilus* [58] and *Geobacillus thermoleovorans* [59], and while screening for LanM-producing gene clusters has highlighted the potential existence of a number of type 2 lantibiotics

[25], associated type 1 lantibiotics have not previously been described. Here, three putative type 1 lantibiotic/lantipeptide-encoding clusters within the genomes of *Geobacillus kaustophilus* HTA426, *Geobacillus thermodenitrificans* NG80-2 and *Geobacillus sp.* G11MC16 (Figure 3) are described.

***Geobacillus kaustophilus* HTA426** *G. kaustophilus* grows optimally in aquatic environments at 60°C with an upper temperature limit of 74°C. From a lantibiotic perspective, genome xsequencing of HTA426 revealed a hypothetical protein annotated as a 'lantibiotic precursor' GK0294. Our analysis revealed that this putative LanA is 91% identical to another prepeptide encoded by the closely located GK0286 gene. It is also 100% identical to orphan 'lantibiotic precursor' homologs (GYMC52DRAFT_3129 and GYMC61_1158) in *Geobacillus sp.* Y412MC52 and *Geobacillus sp.* Y412MC61, respectively. More distantly related LanAs (79% identity) are also associated with the genomes of *Geobacillus thermodenitrificans* NG80-2 (GTNG_0265) and *Geobacillus sp.* G11MC16 (G11MC16DRAFT_2956). The homology between the *Geobacillus* LanAs is highest within the leader regions, but, as is the case with nisin-, epidermin- and streptin-like lantibiotics, a conserved serine and CTPGC motif in the N-terminus of the prepeptide is present, which is believed to be involved in the binding of these lantibiotics to lipid II in the cell wall in gram positive bacteria [60]. BLAST analysis of the GK0286-encoded LanA highlighted the presence of another potential lantibiotic/lantipeptide cluster in *Corynebacterium lipophiloflavum* DSM 44291 (57% identity with HMPREF0298_1795). Within the HTA426 cluster, the proteins predicted to be encoded by GK0300/301 (an apparently frameshifted *lanB*) and GK0304 are homologous to those associated with many other geobacilli. It was also noted that this cluster is less condensed than typical lantibiotic gene clusters in that there are insertions of 7, 5 and 3 genes (predicted to encode many transposases and small, hypothetical proteins) between the lantibiotic associated genes.

***Geobacillus thermodenitrificans* NG80-2 and *Geobacillus sp.* G11MC16:** *G. thermodenitrificans* are facultative soil bacteria with denitrification qualities. Representatives of this species grow between 45°C and 73°C (optimum 65°C). NG80-2 was isolated from a deep-subsurface oil reservoir in Dagang oilfield, Northern China [61] and on the basis of *in-silico* analysis is potentially the producer of both a type 1 (see below) and type 2 lantibiotic [25]. Our analysis reveals that the type I lantibiotic/lantipeptide operons in *G. thermodenitrificans* NG80-2 and *Geobacillus sp.* G11MC16 are very highly conserved. The two LanAs are 100% identical and the homology between these, and indeed the associated B and Cs, and the corresponding *G. kaustophilus*

HTA426 proteins is discussed above. It was noted that the *lanB* of *Geobacillus sp.* G11MC16 is apparently frameshifted (G11MC16DRAFT_2955 (176aa) and G11MC16DRAFT_2954 (848aa)) but that this is not the case in *G. thermodenitrificans* NG80-2 (GTNG_0266).

Identification of novel *Staphylococcus*-associated lantibiotic gene clusters

The staphylococci are non-sporeforming, non-motile *Firmicutes*. The genus *Staphylococcus* contains 33 species, most of which are harmless and reside normally on the skin and mucous membranes of humans and other organisms. However, staphylococci can also cause a wide variety of diseases either through toxin production or penetration and are a common cause of food poisoning and nosocomial infections. Several strains of *Staphylococcus epidermidis* have been shown to be producers of type 1 lantibiotics, including epidermin [62], Pep5 [63], epicidin 280 [64] and epilancin K7 [65], gallidermin was isolated from *S. gallinarum* [14] while Staphylococin Au26 [66] and Bsa [22] were isolated from *S. aureus*. BLAST analysis has revealed that several other *S. aureus* strains possess gene clusters similar to those associated with Bsa and Bsa_{RF122} [22]. These clusters were identified in *S. aureus* A9765, D139 and H19. In A9765, SAPG_01762 and SAPG_01760 correspond to the BsaA1 and BsaA2 peptides of *S. aureus* MW2 (97% and 100% identity, respectively). The precursor peptides of the D139 (SATG_00575 and SATG_00574; 76% identical to each other) and H19 (SAUG_01228 and SAUG_01229; 76% identical to each other) strains are 100% identical. The peptides encoded by SATG_00575 and SAUG_01229 are 93% identical to BsaA1_{RF122} of *S. aureus* RF122 (93% identity) while those corresponding to SATG_00574 and SAUG_01228 are 100% identical to BsaA2_{RF122}.

Identification of novel *Streptococcus*-associated lantibiotic gene clusters

These facultative anaerobes of the phylum *Firmicutes* are spherical in shape and grow in long chains. Many species are part of the normal commensal flora of the mouth, skin, intestine and upper respiratory tract of humans but the genus also includes numerous human pathogens such as *Streptococcus pneumoniae*, *pyogenes* and *agalactiae*. The streptococci are known to producers of type 1 lantibiotics [67,68], such as streptin [69], some mutacins [21,70-72], nisin U and nisin U2 [73], as well as several non-lantibiotic bacteriocins. Here we discuss two clusters, identified in strains of *S. pyogenes* and *S. thermophilus* LMG 18311.

***Streptococcus pyogenes* MGAS10270** *S. pyogenes* (or Group A *Streptococcus*, GAS) is the cause of many important human diseases ranging from mild superficial skin infections to life-threatening systemic diseases. Bacteriocin production by these strains may give them a

competitive advantage against the natural skin microbiota. It has previously been established that many *S. pyogenes* strains, as well as strains of *Streptococcus salivarius*, produce the type 2 lantibiotic salivaricin A or closely related variants [16]. The type 1 streptins (1 and 2) and type 2 streptococcin A-FF22 are also *S. pyogenes* associated [69,74]. Here our analysis focuses on a type 1 cluster within the genome of *S. pyogenes* MGAS10270 [75]. This includes MGAS10270_Spy0919, which is 100% identical to the propeptide sequence of streptin. While this lantibiotic is thus not novel, subsequent BLAST searches were revealing in that they highlighted the presence of a LanA with 97% identity in *S. pyogenes* MGAS10750 (MGAS10750_Spy0955) which is contained within a cluster which also encodes a LanB (MGAS10750_Spy0958) and LanC (MGAS10750_Spy0957).

***Streptococcus thermophilus* LMG 18311** *S. thermophilus* is a thermophilic, non-pathogenic *Streptococcus*. It is of major importance to the fermented dairy food industry. A number of non-lantibiotic bacteriocins (thermophilins) from this species have been characterized, including thermophilin 347 [76], thermophilin A [77] and thermophilin ST-1 [78]. Strain LMG 18311 was sequenced in 2004 and at the time it was noted that bacteriocin production was one of the characteristics that distinguishes it from strain CNRZ1066 [79]. While the existence of a putative lantibiotic/lantipeptide gene cluster in LMG 18311 has been reported [79,80], this cluster (Figure 3) has not been the focus of a detailed *in silico* analysis. The associated LanA, encoded by stu0097, is homologous with that predicted to be encoded by SPCG_0144 of *S. pneumoniae* CGSP14 (88% identical) which, on the basis of previous *in silico* analysis, is also within a lantibiotic gene cluster [81]. The LanB protein (Stu0098) is 73% identical to SPCG_0145 of *S. pneumoniae* CGSP14 and 97% identical to a truncated LanB associated with *S. thermophilus* CNRZ1066 [79].

Type 1 lantibiotic gene clusters in *Bacteroidetes*

The *Bacteroidetes* are a highly diverse phylum found in soil, seawater and the skin and intestines of animals. The *Bacteroidales* class, which includes the genus *Bacteroides*, are the best-studied of the phylum. *Bacteroides* comprises the most substantial portion of the human gastrointestinal tract [82] some of which are opportunistic pathogens [83].

Identification of novel *Chitinophaga*-associated lantibiotic gene clusters

Chitinophaga are rod-shaped mesophiles of the phylum *Bacteroidetes* which are noted for their ability to degrade chitin [84]. There have been no reports to date of bacteriocin production by any of the 10 *Chitinophaga* species. *Chitinophaga pinensis* DSM 2588 (NC_013132) is unusual in that it appears to be a *Bacteroidetes* possessing genes encoding a type 1 lantibiotic (Figure 4).

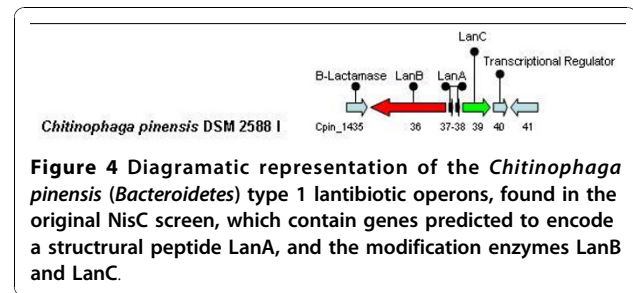


Figure 4 Diagrammatic representation of the *Chitinophaga pinensis* (*Bacteroidetes*) type 1 lantibiotic operons, found in the original NisC screen, which contain genes predicted to encode a structural peptide LanA, and the modification enzymes LanB and LanC.

There are two predicted LanA peptides, corresponding to Cpin_1438 and Cpin_1437, which are 50% identical as a consequence of similar N-termini. Adjacent orfs of note include Cpin_1435 and Cpin_1440, predicted to encode a β -lactamase and a 2-component transcriptional regulator of the LuxR family, respectively. BLAST analysis of the associated LanB and LanC proteins (Cpin_1436 and Cpin_1439 respectively) revealed another putative LanB (Cpin_3392; 36% identity) and LanC (Cpin_3397; 23% identity) encoded within the same genome. Within this second *C. pinensis*-associated cluster, Cpin_3393 possess a number of features which suggest that it may be a LanA-encoding gene. Interestingly, BLAST analysis of the Cpin_3397-encoded LanC also led to the identification of a number of additional homologs apparently encoded within the genomes of strains not previously associated with lantibiotic production. The genome of another *Bacteroidetes*, *Spirosoma linguale* DSM 74, is notable in that it contains 4 putative LanB-encoding genes and 5 putative LanC-encoding genes. Of these only one LanB protein (Slin_4704; 31% identity) and one LanC protein (Slin_4705; 26% identity), are encoded within what appears to be a novel lantibiotic-associated gene cluster. This cluster contains 3 potentially LanA-encoding genes, Slin_4706-4708. Slin_4706 and Slin_4707, which are identical and share 58% identity with Slin_4708. The genome of another *Bacteroidetes* species, *Pedobacter heparinus* DSM 2366 (NC_013061), contains a cluster encoding two LanBs (Phep_0556 and Phep_0557; 37% and 36% identity, respectively), a LanC (Phep_0555; 33% identity) and a potential LanA (Phep_0553; no significant BLAST hits). A cluster within *Kordia algicida* OT-1 contains LanB (KAOT1_15523; 36% identity), LanC (KAOT1_15518; 30% identity) and LanA determinants (KAOT1_15533; no significant BLAST hits) while *Microscilla marina* ATCC 23134 potentially has five associated LanB homologs, but of these, the LanB corresponding to M23134_03921 (28% identity) was the only one to be located in close proximity to one of multiple LanC proteins (M23134_03925; 22% identity). The putative LanA, M23134_03926, does not resemble any other known proteins.

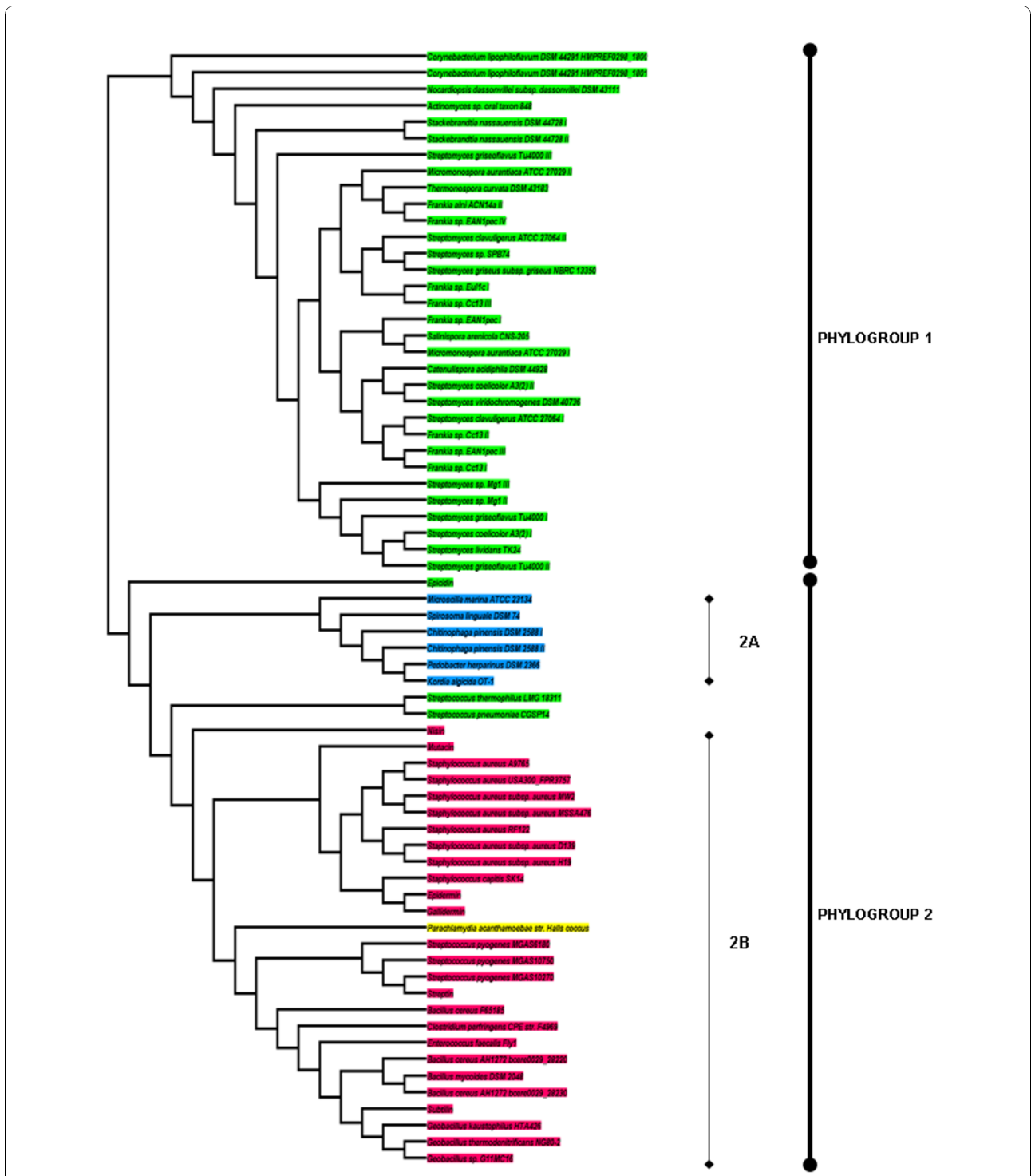


Figure 5 Cladogram of all the LanB enzymes from clusters encountered during the screen. Also included are dehydratases from some well-known lantibiotics. Green = Actinobacteria; Pink = Firmicutes; Blue = Bacteroidetes; Yellow = Chlamydiae

Phylogenetics of LanABC

The conserved nature of LanB and LanC proteins facilitated a phylogenetic analysis of their relatedness. The resultant cladogram of LanB enzymes (all those identified in both screens, as well as a number of LanBs from previously analysed clusters) highlights the existence of two distinct phylogroups (Figure 5). The first phylogroup contains *Actinobacteria*-associated LanBs, all of which are from strains not previously known to be producers of lantibiotics/lantipeptides. The second contains a variety of lanthionine synthetases associated with known lantibiotics (nisin, subtilin, epidermin etc), some uncovered by previous *in silico* analysis (e.g. *S. pneumoniae* CGSP14 [81]), novel clusters from genera with which lantibiotic production has previously been attributed as well as genera not previously associated with lantibiotic production. Within this second phylogroup one finds two subgroups; one consisting of *Bacteroidetes*-associated LanBs (2A) and a second consisting of *Firmicutes*-associated LanB's (2B) as well as that from *P. acanthamoeba*. Among the *Firmicutes*-associated LanBs further subclustering is evident. One common branch contains three offshoots; (i) *Bacillus/Geobacillus/Enterococcus/Clostridium*, (ii) *S. pyogenes* and (iii) *P. acanthamoebae* str. Hall's coccus LanBs. The *Staphylococcus* LanBs and that of mutacin 1140 (those associated with epidermin-like peptides) also form a distinct subgroup as do those encoded with the genomes of *S. thermophilus* LMG 18311 and *S. pneumoniae* CGSP14. Curiously the epicidin (*S. epidermidis*) LanB does not group with any other LanB.

The cladogram of the corresponding lanthionine synthetases (LanCs) is quite similar to that of the dehydratases (Figure 6). All can be positioned into one of two phylogroups (phylogroups 1 and 2). Phylogroup 1 contains six *Bacteroidetes*-associated LanC's. In contrast phylogroup 2 is large and can be further divided into group 2A, which are *Actinobacteria*-associated, and the *Firmicutes*-associated group 2B. Further subgrouping follows the patterns identified from analysis of the LanB cladogram. 8 exceptional LanCs that avoid subgrouping include those associated with *C. lipophiloflavum* DSM 44291, *S. thermophilus* LMG 18311/*S. pneumoniae* CGSP14, epicidin, *B. cereus* F65185, *N. dassonvillei* subsp. *dassonvillei* DSM 43111, nisin-producing lactococci and *Actinomyces* sp. oral taxon 848.

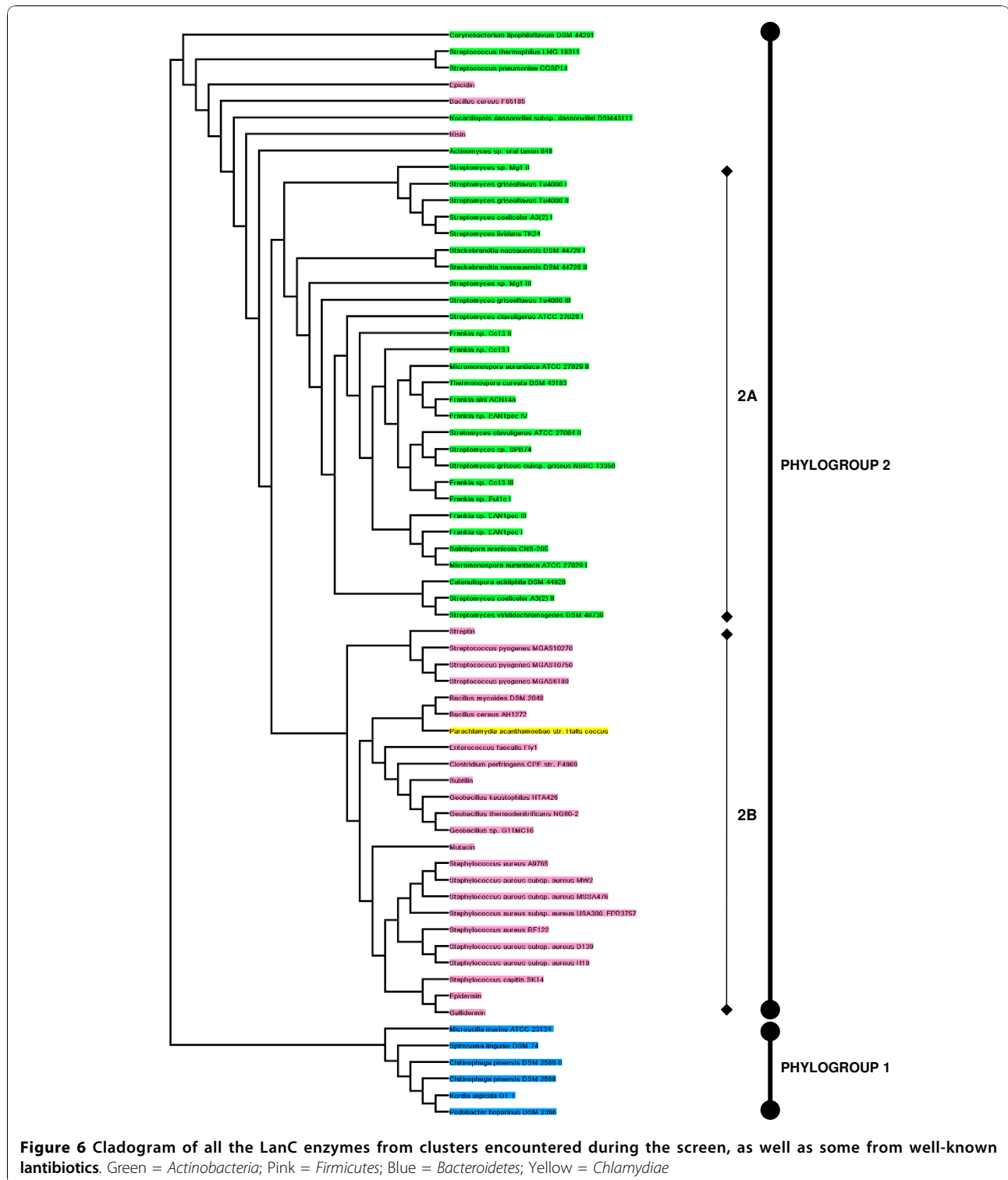
A cladogram of the less highly conserved LanAs revealed 13 major branches, several of which contain only one corresponding LanA (Figure 7). Notably the various phylogroups do not group in a phylum specific manner to the same extent as was evident in LanB and LanC cladograms. The largest phylogroups, i.e. phylogroups 11 and 13, are those containing the nisin-like and epidermin-like peptides, respectively. While

phylogroup 13 is, with the exception of the *Actinomyces* sp. oral taxon 848-associated LanA, composed of *Firmicutes*-associated LanAs, phylogroup 11 contains LanAs from both *Firmicutes* and *Actinobacteria*. This phylogroup contains three subgroups, with subgroup 11A containing LanAs from *E. faecalis* Fly1 and *C. perfringens* F4969, 11B contains *Actinobacteria*-associated LanAs and 11C contains both *Actinobacteria*- and *Firmicutes*-associated LanAs. Of the other phylogroups, phylogroups 6 and 7 are largest and contain *Chlamydiae/Bacteroidetes*- and *Actinobacteria*-associated LanAs, respectively.

Alignment of LanABC

The availability of a significant number of LanA, B and C protein sequences enabled further *in silico* analysis to identify conserved motifs and residues. Alignment of 66 LanB proteins highlighted a number conserved motifs which are summarised in Table 5. A YxxR motif (corresponding to residues 80-83 of NisB) is conserved in 89% of the LanB enzymes, while a GxG motif (363-365) is present in 92% of LanBs, with the LanB of *C. lipophiloflavum* being exceptional by virtue of lacking both glycine residues. A GRF motif (463-465) is fully conserved in 86% of LanBs with the *Streptomyces* sp. Mg1 III LanB being the only protein to lack this motif. An RxTPFG motif (87-94) is present in 77% of LanBs but is completely absent from the LanBs of *Geobacillus* sp. G11MC16, *Streptomyces* sp. Mg1 III, *M. aurantiaca* and *C. lipophiloflavum*. A FxxxYG motif (342-347) is present in 82% of LanBs and, although present in only 50% of LanBs, a PxxxRxxNV (501-509) motif is at least partially conserved in many such proteins i.e. 94% contain the proline, 71% contain the NV residues and the least conserved is the arginine with 71% conservation. Elsewhere, there is a RFL motif (585-587) conserved in 51% of LanBs, a RYG motif (826-828) conserved in 85% of LanBs and a HxxxNR motif (961-966) in 70% of the dehydratases. In addition to these, there are multiple highly conserved residues such as aspartates at residues 121, 299, 648 and 843, prolines at 612 and 639 and a leucine, tryptophan and phenylalanine at 97, 616 and 840.

Alignment of the LanC protein also revealed several conserved regions (summarised in Table 6). Of these, CHG and WCYG motifs were particularly notable. The CHG motif (corresponding to residues 330-333 of NisC) was found to be conserved in 98% of the LanCs. The cysteine³³⁰ and histidine³³¹ residues, which act as ligands to the zinc in the active site of NisC, have been shown to be necessary for enzyme activity [85]. The WCYG motif (283-286) was present in 95% of the aligned enzymes. Within the WCYG motif, tryptophan²⁸³ (W) and cysteine²⁸⁴ (C) have been shown to be



vital to subtilin and nisin biosynthesis (residue numbers refer to location in NisC) [85,86]. It has previously been shown that although alanine substitution of tyrosine²⁸⁵ (Y) results in enzyme inactivation, a phenylalanine change is tolerated indicating that the presence of an

aromatic ring at this position is of key importance [85]. In the same study, a preceding arginine residue (Arg²⁸⁰), present in 86% of these enzymes, was found not to be essential for enzyme activity. 92% of LanCs also contained a closely located Gly²⁸⁹ residue. The histidine²¹²



Figure 7 A cladogram of all the LanA prepeptides identified in this study, as well as a number of previously identified LanAs.

of another highly conserved motif, GxAHGxxG (209-216; conserved in 83% of LanCs), together with a conserved aspartic acid¹⁴¹ (91% of LanCs) are thought to be involved in the electrophilic activation of the carbonyl group of dehydroalanine/dehydrobutyrine or in the

protonation of the enolate (thiol substrate) [85]. The HG of this latter motif was conserved in 98% of the enzymes (the exception being *S. aureus* subsp. *aureus* D139). In addition to these, other motifs of note included LxxG (39-42; conserved in 83% of LanCs),

Table 5 Highly conserved residues shared by 66 cluster-associated (including those from the 49 novel clusters referred to in this paper) lanthionine dehydratases (LanB).

From Alignment	NisB Position	% Conservation
Motifs		
YxxR	80-83	89%
RxTPFG	87-94	77%
FxxxYG	342-347	82%
GxG	363-365	92%
GRF	463-465	86%
PxxxRxxNV	501-509	50%
RFL	585-587	51%
RYG	826-828	85%
HxxxNR	961-966	70%
Single Residues		
R	14	86%
D	121; 299; 648; 843	86%; 94%; 94%; 85%
N	145	86%
L	217	97%
P	612; 639	100%; 95%
E	975	89%
W	616	98%
F	840	95%
V	352	83%

Residues are numbered according to their position in NisB

GxxxGxxGxxLxL (377-389; 73%) and YDxxxGxxG (140-148; 67%). Highly conserved single residues include Gly⁹⁰ (94%) and Tryp²⁵⁸ and Tryp⁴⁰¹ (83% and 92% respectively).

Although LanA peptides are less conserved than their modification enzymes, some motifs were evident (Figure 8). A DLD motif present in the leader region of the almost all phylogroup 13 LanAs is also found in many other LanAs. Indeed, the leucine of this motif is conserved across 93% of

Table 6 Highly conserved residues shared by 66 cluster-associated (including those from 49 novel clusters) lanthionine cyclases. Residues are numbered according to their position in NisC

Conserved Residues	NisC Position	% Conservation
Motifs		
LxxG	39-42	83%
YDxxxGxxG	140-148	67%
GxAHGxxG	209-216	83%
WCYG	283-286	95%
CHG	330-332	98%
GxxxGxxGxxLxL	377-389	73%
Single Residues		
G	90	94%
W	258; 401	83%; 92%

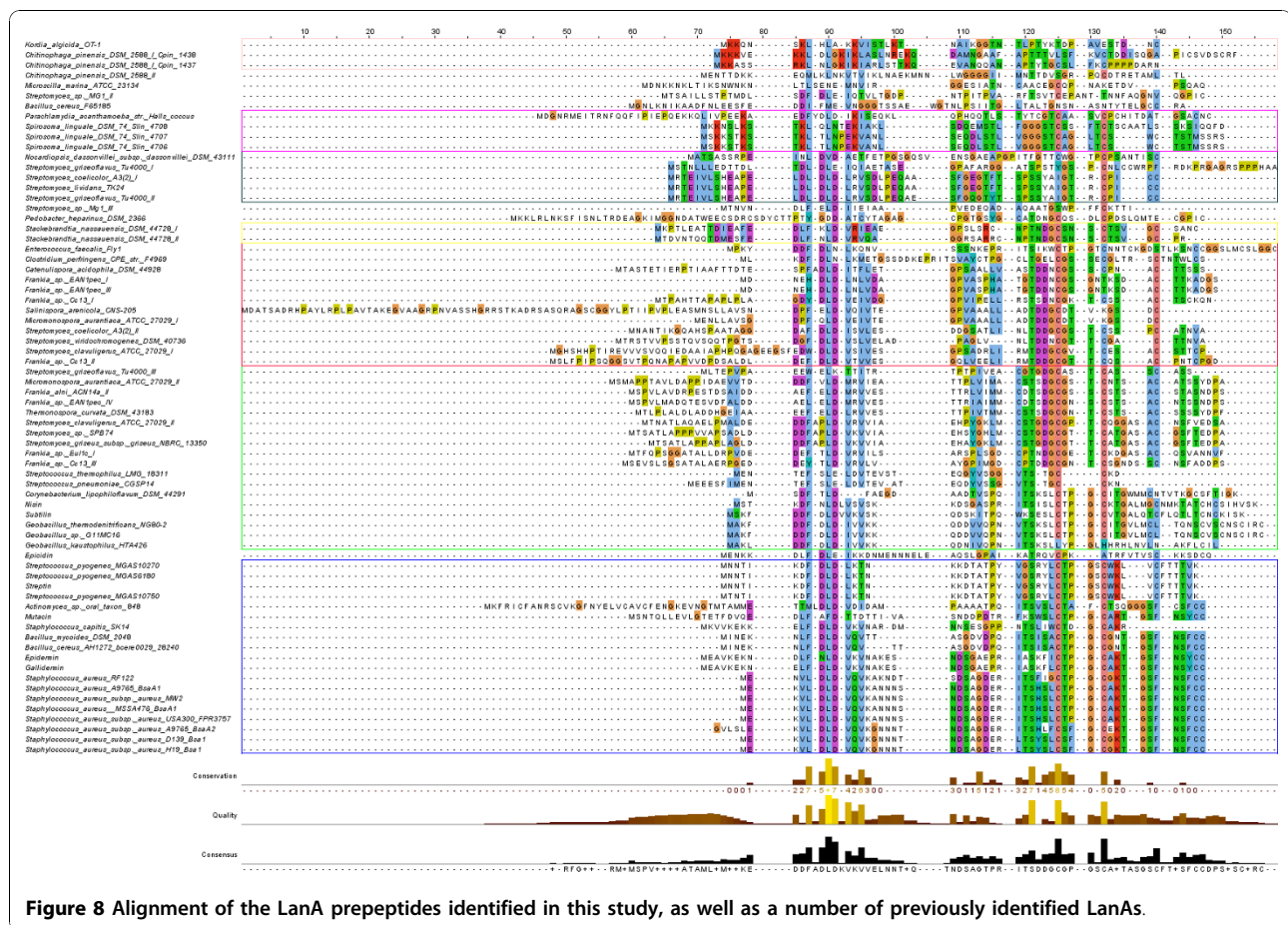
the 70 aligned peptides. The only peptides lacking this residue are those from *M. marina*, *B. cereus* F65185, *N. dassonvillei* subsp. *dassonvillei* and *P. heparinus* as well as mutacin 1140. The leader regions from many actinobacteria, and especially those from phylogroup 11, are also distinctive as a consequence of the frequent presence of proline residues.

Within the propeptide, the most highly conserved residues are cysteines corresponding to positions 30 and 34 of the Nisin A prepropeptide [60] which significantly are within the lipid II-binding region of the peptide. These cysteines are each found in 86% of the type 1 LanAs. The *Actinobacteria* in phylogroups 10 and 11 also share a conserved or partially conserved DGCG motif in the propeptide region. A less highly conserved AC motif which is closer to the C terminus is also evident. In addition to motifs which are conserved across motifs, a large number of motifs which are conserved within phylogroups are evident.

Conclusions

The *in silico* strategy adopted here resulted in the initial identification of 56 proteins which share 20-30% identity with NisC. Further investigation of novel LanC proteins in turn led to the identification of even more homologs, revealing novel lantibiotic/lantipeptide associated clusters and establishing the existence of subgroups of LanA, B and C proteins. Theoretically, additional homologs could be identified through a continuation of this method but the rate at which new homologs would be identified would begin to level off. The approach taken led to the identification of 49 novel clusters which, prior to this study, had not been the subject of a detailed bioinformatic analysis. While the *in silico* identification of gene clusters in a strain will not always be confirmed by detection of an associated lantibiotic/lantipeptide, past experience [22-25] suggests that there is likely to be a strong correlation. It is thus anticipated that the peptides produced by these gene clusters will represent a valuable resource, as will be the associated biosynthetic proteins.

This study reveals new details regarding type 1 lantibiotics and their associated clusters. Type 1 lantibiotics have been predominantly associated with the *Firmicutes*, with the *Actinobacteria*-produced planosporicin and microbisporicin being notable exceptions. It is thus interesting to find type 1 clusters distributed among the genomes of bacteria representing four different phyla, the *Actinobacteria*, *Firmicutes*, *Bacteroidetes* and *Chlamydiae*, which have been isolated from a diverse range of habitats including soil, skin, intestines and the deep-sea. Indeed, based on these investigations, it would appear that such clusters are as common among *Actinobacteria* as they are among *Firmicutes*, with *Streptomyces* and *Frankia* sp. being particularly rich sources.



The *Actinobacteria* clusters are, in general, quite similar, typically encoding a LanA, B, C and a methyltransferase. The role of the methyltransferase is not clear but may serve to protect specific serine and threonine residues from LanB-mediated dehydration. The presence of five clusters within the genomes of five *Bacteroidetes*, a phylum in which bacteriocin production is purportedly quite rare, is particularly noteworthy. However, the sequencing of additional representatives of this species may well reveal this to be a common feature. The *P. acanthamoebae* cluster is unusual by virtue of its presence in a representative of the *Chlamydiae*. Phylogenetic analysis indicates that the LanB and LanC proteins from this strain are closely related to those of several *Firmicutes* and thus the cluster may originally have been acquired from such a source.

The availability of a much larger collection of LanA, B and C sequences for further *in silico* analysis is also extremely useful for a number of other reasons. In addition to providing greater certainty with respect to the proposed conservation of particular motifs, it also reveals the existence of a greater number of subgroups of sequences than was previously apparent. This is particularly important

with respect to LanAs as alignment of these peptides has previously been employed as a means of subgrouping type 1 lantibiotics [2,6]. Ultimately, the most significant outcome has been the number of new type 1 lantibiotic gene clusters. When one considers that less than 25 type 1 lantibiotics had been identified prior to this study, this represents a major expansion. While the genome sequenced strains themselves can be accessed with a view to purifying the associated peptides and/or utilising the biosynthetic machinery, the information gathered will also encourage researchers to include *Actinobacteria* and *Bacteroidetes* when carrying out wet lab-based screens for novel lantibiotic producers. A combination of this approach and analysis of newly generated bacterial genome sequence data will ensure that many more lantibiotics and lantipeptides will soon be discovered which are associated with unusual microorganisms and a wide variety of environments.

Methods

Screening of genomic databases

Using the nisin modification enzyme NisC (GenBank accession number CAA79470) as a driver sequence, all fully sequenced genomic sequences (approx. 1178 at

time of study; Dec 2009) were mined for homologs using Genomic-BLAST (http://www.ncbi.nlm.nih.gov/sutils/genom_table.cgi). BLASTs were carried out with default parameters; criteria for homolog detection were a threshold of $1e^{-7}$ and greater than 20% identity.

Bioinformatic analysis of *lanC*-containing gene clusters

In cases where novel *lanC*-like genes encoding enzymes were identified, the arrangement of adjacent genes was visualised using the genome viewer on NCBI, and individual orfs were subjected to BLAST analysis to identify those potentially involved in lantibiotic production or immunity. The predicted LanA, LanB and LanC proteins from these operons were each in turn used for further *in silico* screens to determine their similarities to corresponding proteins associated with known lantibiotics and to identify additional novel clusters. In instances where a LanC- and LanB-, but not a LanA-, encoding gene were annotated, intergenic regions were inspected following translation by the Seqbuilder program of the DNASTAR Lasergene 8 software package to investigate the presence of potentially unidentified *lanA* genes. The tblastn program was then used to search all sequenced DNA for related peptides.

Phylogenetic analysis

Protein alignments were generated by MUSCLE [87]. Sequence alignment were viewed and edited for publication with Jalview alignment editor [88]. These alignments were used to establish phylogenetic trees in Phylip [89] which were subsequently visualised using the Dendroscope package [90].

Acknowledgements

This work was supported by the Science Foundation of Ireland funded Centre for Science, Engineering and Technology (SFI-CSET): the Alimentary Pharmabiotic Centre (APC).

Author details

¹Teagasc, Moorepark Food Research Centre, Fermoy, Cork, Ireland.

²Microbiology Department, University College Cork, Cork, Ireland. ³Alimentary Pharmabiotic Centre, Cork, Ireland.

Authors' contributions

PDC conceived the study and designed the project. OOS participated in the design and coordination of the project. AJM performed the screening of the databases and the bioinformatic and phylogenetic analysis. AJM and PDC wrote the manuscript. OOS, PRR and CH contributed in the preparation of the manuscript. All authors have read and approved the final manuscript.

Received: 7 September 2010 Accepted: 30 November 2010

Published: 30 November 2010

References

1. Cotter PD, Hill C, Ross RP: **Bacteriocins: Developing innate immunity for food.** *Nat Rev Microbiol* 2005, **3**(10):777-788.
2. Rea M, Ross RP, Cotter PD, Hill C: **Classification of bacteriocins from gram positive bacteria.** In *Prokaryotic antimicrobial peptides* Springer Press; 2010.

3. Willey JM, van der Donk WA: **Lantibiotics: Peptides of diverse structure and function.** *Annual Review of Microbiology* 2007, **61**:477-501.
4. Kodani S, Hudson ME, Durrant MC, Buttner MJ, Nodwell JR, Willey JM: **The SapB morphogen is a lantibiotic-like peptide derived from the product of the developmental gene *ramS* in *Streptomyces coelicolor*.** *Proceedings of the National Academy of Sciences of the United States of America* 2004, **101**(31):11448-11453.
5. Goto Y, Li B, Claesen J, Shi Y, Bibb M, Van der Donk W: **Discovery of unique lanthionine synthetases reveals new mechanistic and evolutionary insights.** *PLoS Biology* 2010, **8**(3).
6. Piper C, Cotter P, Ross R, Hill C: **Discovery of medically significant lantibiotics.** *Current Drug Discovery Technologies* 2009, **6**(1):1-18.
7. Delves-Broughton J: **Nisin as a food preservative.** *Food Australia* 2005, **57**(12):525-527.
8. Kupke T, Stevanovic S, Sahl HG, Gotz F: **Purification and Characterization of Epid, a Flavoprotein Involved in the Biosynthesis of the Lantibiotic Epidermin.** *Journal of Bacteriology* 1992, **174**(16):5354-5361.
9. Bierbaum G, Sahl HG: **Lantibiotics: Mode of Action, Biosynthesis and Bioengineering.** *Current Pharmaceutical Biotechnology* 2009, **10**(1):2-18.
10. Severina E, Severin A, Tomasz A: **Antibacterial efficacy of nisin against multidrug-resistant Gram-positive pathogens.** *J Antimicrob Chemother* 1998, **41**(3):341-347.
11. Parrot M, Charest M, Lavoie MC: **Production of Mutacin-Like Substances by *Streptococcus mutans*.** *Canadian Journal of Microbiology* 1989, **35**(3):366-372.
12. Castiglione F, Cavaletti L, Losi D, Lazzarini A, Carrano L, Ferroggio M, Ciciliato I, Corti E, Candiani G, Marinelli F, et al: **A novel lantibiotic acting on bacterial cell wall synthesis produced by the uncommon actinomycete *Planomonospora* sp.** *Biochemistry* 2007, **46**(20):5884-5895.
13. Fontana MBC, de Bastos MDF, Brandelli A: **Bacteriocins Pep5 and epidermin inhibit *Staphylococcus epidermidis* adhesion to catheters.** *Current Microbiology* 2006, **52**(5):350-353.
14. Kellner R, Jung G, Horner T, Zahner H, Schnell N, Entian KD, Gotz F: **Gallidermin - a New Lanthionine-Containing Polypeptide Antibiotic.** *European Journal of Biochemistry* 1988, **177**(1):53-59.
15. Cleveland J, Montville TJ, Nes IF, Chikindas ML: **Bacteriocins: safe, natural antimicrobials for food preservation.** *International Journal of Food Microbiology* 2001, **71**(1):1-20.
16. Wescombe PA, Upton M, Dierksen KP, Ragland NL, Sivabalan S, Wirawan RE, Inglis MA, Moore CJ, Walker GV, Chilcott CN, et al: **Production of the lantibiotic salivaricin A and its variants by oral streptococci and use of a specific induction assay to detect their presence in human saliva.** *Appl Environ Microbiol* 2006, **72**(2):1459-1466.
17. Wescombe PA, Heng NCK, Burton JP, Chilcott CN, Tagg JR: **Streptococcal bacteriocins and the case for *Streptococcus salivarius* as model oral probiotics.** *Future Microbiology* 2009, **4**(7):819-835.
18. Dabard J, Bridonneau C, Philippe C, Anglade P, Molle D, Nardi M, Ladire M, Girardin H, Marcille F, Gomez A, et al: **Ruminococcin A, a new lantibiotic produced by a *Ruminococcus gnavus* strain isolated from human feces.** *Appl Environ Microbiol* 2001, **67**(9):4111-4118.
19. Ryan MP, Rea MC, Hill C, Ross RP: **An application in cheddar cheese manufacture for a strain of *Lactococcus lactis* producing a novel broad-spectrum bacteriocin, lacticin 3147.** *Appl Environ Microbiol* 1996, **62**(2):612-619.
20. Beasley SS, Saris PEJ: **Nisin-producing *Lactococcus lactis* strains isolated from human milk.** *Appl Environ Microbiol* 2004, **70**(8):5051-5053.
21. Hillman JD, Novak J, Sagura E, Gutierrez JA, Brooks TA, Crowley PJ, Hess M, Azizi A, Leung KP, Cvitkovitch D, et al: **Genetic and biochemical analysis of mutacin 1140, a lantibiotic from *Streptococcus mutans*.** *Infection and Immunity* 1998, **66**(6):2743-2749.
22. Daly KM, Upton M, Sandiford SK, Draper LA, Wescombe PA, Jack RW, O'Connor PM, Rossney A, Gotz F, Hill C, et al: **Production of the Bsa Lantibiotic by Community-Acquired *Staphylococcus aureus* Strains.** *Journal of Bacteriology* 2010, **192**(4):1131-1142.
23. McClerren AL, Cooper LE, Quan C, Thomas PM, Kelleher NL, van der Donk WA: **Discovery and *in vitro* biosynthesis of haloduracin, a two-component lantibiotic.** *Proceedings of the National Academy of Sciences of the United States of America* 2006, **103**(46):17243-17248.
24. Lawton EM, Cotter PD, Hill C, Ross RP: **Identification of a novel two-peptide lantibiotic, Haloduracin, produced by the alkaliphile *Bacillus halodurans* C-125.** *FEMS Microbiol Lett* 2007, **267**(1):64-71.

25. Begley M, Cotter PD, Hill C, Ross RP: **Identification of a Novel Two-Peptide Lantibiotic, Lichenicidin, following Rational Genome Mining for LanM Proteins.** *Appl Environ Microbiol* 2009, **75**(17):5451-5460.
26. Dischinger J, Josten M, Szeekat C, Sahl HG, Bierbaum G: **Production of the Novel Two-Peptide Lantibiotic Lichenicidin by *Bacillus licheniformis* DSM 13.** *Plos One* 2009, **4**(8).
27. Li B, Sher D, Kelly L, Shi Y, Huang K, Knerr P, Joewono I, Rusch D, Chisholm S, van der Donk W: **Catalytic promiscuity in the biosynthesis of cyclic peptide secondary metabolites in planktonic marine cyanobacteria.** *Proceedings of the National Academy of Sciences of the United States of America* 2010.
28. de Jong A, van Hijum SAFT, Bijlsma JJE, Kok J, Kuipers OP: **BAGEL: a web-based bacteriocin genome mining tool.** *Nucleic Acids Research* 2006, **34**: W273-W279.
29. de Jong A, van Heel A, Kok J, Kuipers O: **BAGEL2: mining for bacteriocins in genomic data.** *Nucleic Acids Research* 2010, **38**:W647-W651.
30. Hammami R, Zouhir A, Ben Hamida J, Fliss I: **BACTIBASE: a new web-accessible database for bacteriocin characterization.** *Bmc Microbiology* 2007, **7**.
31. Hammami R, Zouhir A, Le Lay C, Ben Hamida J, Fliss I: **BACTIBASE second release: a database and tool platform for bacteriocin characterization.** *Bmc Microbiology* 2010, **10**.
32. Normand P, Lapiere P, Tisa LS, Gogarten JP, Alloisio N, Bagnarol E, Bassi CA, Berry AM, Bickhart DM, Choise N, *et al*: **Genome characteristics of facultatively symbiotic *Frankia* sp strains reflect host range and host plant biogeography.** *Genome Research* 2007, **17**(1):7-15.
33. Li SY, Anzai Y, Kinoshita K, Kato F, Sherman DH: **Functional Analysis of MycE and MycF, Two O-Methyltransferases Involved in the Biosynthesis of Mycinamicin Macrolide Antibiotics.** *ChemBiochem* 2009, **10**(8):1297-1301.
34. Tsukada S, Anzai Y, Li SY, Kinoshita K, Sherman DH, Kato F: **Gene targeting for O-methyltransferase genes, mycE and mycF, on the chromosome of *Micromonospora griseorubida* producing mycinamicin with a disruption cassette containing the bacteriophage C31 attB attachment site.** *FEMS Microbiol Lett* 2010, **304**(2):148-156.
35. Zubieta C, He XZ, Dixon RA, Noel JP: **Structures of two natural product methyltransferases reveal the basis for substrate specificity in plant O-methyltransferases.** *Nature Structural Biology* 2001, **8**(3):271-279.
36. Penn K, Jenkins C, Nett M, Udway DW, Gontang EA, McGlinchey RP, Foster B, Lapidus A, Podell S, Allen EE, *et al*: **Genomic islands link secondary metabolism to functional adaptation in marine Actinobacteria.** *Isme Journal* 2009, **3**(10):1193-1203.
37. Berdy J: **Bioactive microbial metabolites - A personal view.** *Journal of Antibiotics* 2005, **58**(1):1-26.
38. Labeda DP, Kroppenstedt RM: ***Stackebrandtia nassauensis* gen. nov., sp nov and emended description of the family Glycomycetaceae.** *International Journal of Systematic and Evolutionary Microbiology* 2005, **55**:1687-1691.
39. Hodgson DA: **Primary metabolism and its control in streptomycetes: A most unusual group of bacteria.** *Advances in Microbial Physiology, Vol 42* 2000, **42**:47-238.
40. Watve MG, Tickoo R, Jog MM, Bhole BD: **How many antibiotics are produced by the genus *Streptomyces*?** *Archives of Microbiology* 2001, **176**(5):386-390.
41. Kido Y, Hamakado T, Yoshida T, Anno M, Motoki Y, Wakamiya T, Shiba T: **Isolation and Characterization of Ancovenin, a New Inhibitor of Angiotensin-I Converting Enzyme, Produced by Actinomycetes.** *Journal of Antibiotics* 1983, **36**(10):1295-1299.
42. Widdick DA, Dodd HM, Barraille P, White J, Stein TH, Chater KF, Gasson MJ, Bibb MJ: **Cloning and engineering of the cinnamycin biosynthetic gene cluster from *Streptomyces cinnamoneus* DSM 40005.** *Proceedings of the National Academy of Sciences of the United States of America* 2003, **100**(7):4316-4321.
43. Ohnishi Y, Ishikawa J, Hara H, Suzuki H, Ikenoya M, Ikeda H, Yamashita A, Hattori M, Horinouchi S: **Genome sequence of the streptomycin-producing microorganism *Streptomyces griseus* IFO 13350.** *Journal of Bacteriology* 2008, **190**(11):4050-4060.
44. Hopwood DA: **Forty years of genetics with *Streptomyces*: from in vivo through in vitro to in silico.** *Microbiology-Sgm* 1999, **145**:2183-2202.
45. Bentley SD, Chater KF, Cerdeno-Tarraga AM, Challis GL, Thomson NR, James KD, Harris DE, Quail MA, Kieser H, Harper D, *et al*: **Complete genome sequence of the model actinomycete *Streptomyces coelicolor* A3(2).** *Nature* 2002, **417**(6885):141-147.
46. Nishio C, Komura S, Kurahashi K: **Peptide Antibiotic Subtilin Is Synthesized Via Precursor Proteins.** *Biochemical and Biophysical Research Communications* 1983, **116**(2):751-758.
47. Stein T, Borchert S, Conrad B, Feesche J, Hofemeister B, Hofemeister J, Entian KD: **Two different lantibiotic-like peptides originate from the ericin gene cluster of *Bacillus subtilis* A1/3.** *Journal of Bacteriology* 2002, **184**(6):1703-1711.
48. Brown LCW, Acker MG, Clardy J, Walsh CT, Fischbach MA: **Thirteen posttranslational modifications convert a 14-residue peptide into the antibiotic thiocillin.** *Proceedings of the National Academy of Sciences of the United States of America* 2009, **106**(8):2549-2553.
49. Sharma N, Gautam N: **Antibacterial activity and characterization of bacteriocin of *Bacillus mycoides* isolated from whey.** *Indian Journal of Biotechnology* 2008, **7**(1):117-121.
50. Dineen SS, Bradshaw M, Johnson EA: **Cloning, nucleotide sequence, and expression of the gene encoding the bacteriocin boticin B from *Clostridium botulinum* strain 213B.** *Appl Environ Microbiol* 2000, **66**(12):5480-5483.
51. Li AW, Verpoorte JA, Lewis RG, Mahony DE: **Characterization of Bacteriocin 28 Produced by *Clostridium-Perfringens*.** *Canadian Journal of Microbiology* 1982, **28**(7):860-873.
52. Nieves BM, Castillo FJ: **Properties of a Bacteriocin Produced by *Clostridium-Bifermentans* 18137.** *Acta Cientifica Venezolana* 1982, **33**(1):38-43.
53. Narayan KG: **Food Borne Infection with *Clostridium-Perfringens* Type-A.** *International Journal of Zoonoses* 1982, **9**(1):12-32.
54. Pearson RD, Valenti WM, Steigbigel RT: ***Clostridium-Perfringens* Wound-Infection Associated with Elastic Bandages.** *Jama-Journal of the American Medical Association* 1980, **244**(10):1128-1130.
55. Franz CMAP, van Belkum MJ, Holzapfel WH, Abriouel H, Galvez A: **Diversity of enterococcal bacteriocins and their grouping in a new classification scheme.** *Fems Microbiology Reviews* 2007, **31**(3):293-310.
56. Booth MC, Bogie CP, Sahl HG, Siezen RJ, Hatter KL, Gilmore MS: **Structural analysis and proteolytic activation of *Enterococcus faecalis* cytolysin, a novel lantibiotic.** *Molecular Microbiology* 1996, **21**(6):1175-1184.
57. Nazina TN, Tourova TP, Poltarau AB, Novikova EV, Grigoryan AA, Ivanova AE, Lysenko AM, Petrunyaka VV, Osipov GA, Belyaev SS, *et al*: **Taxonomic study of aerobic thermophilic bacilli: descriptions of *Geobacillus subterraneus* gen. nov., sp nov and *Geobacillus uzonensis* sp nov from petroleum reservoirs and transfer of *Bacillus stearothermophilus* *Bacillus thermocatenulatus*, *Bacillus thermoleovorans*, *Bacillus kaustophilus*, *Bacillus thermoglucosidasius* and *Bacillus thermodenitrificans* to *Geobacillus* as the new combinations *G-stearothermophilus*, *G-thermocatenulatus*, *G-thermoleovorans*, *G-kaustophilus*, *G-thermoglucosidasius* and *G-thermodenitrificans*.** *International Journal of Systematic and Evolutionary Microbiology* 2001, **51**:433-446.
58. Pokusaeva K, Kuisienu N, Jasinskyte D, Rutiene K, Saleikiene J, Chitavichius D: **Novel bacteriocins produced by *Geobacillus stearothermophilus*.** *Central European Journal of Biology* 2009, **4**(2):196-203.
59. Novotny JF, Perry JJ: **Characterization of Bacteriocins from 2 Strains of *Bacillus-Thermoleovorans*, a Thermophilic Hydrocarbon-Utilizing Species.** *Appl Environ Microbiol* 1992, **58**(8):2393-2396.
60. Hsu STD, Breukink E, Tischenko E, Lutters MAG, de Kruijff B, Kaptein R, Bonvin AMJJ, van Nuland NAJ: **The nisin-lipid II complex reveals a pyrophosphate cage that provides a blueprint for novel antibiotics.** *Nature Structural & Molecular Biology* 2004, **11**(10):963-967.
61. Feng L, Wang W, Cheng JS, Ren Y, Zhao G, Gao CX, Tang Y, Liu XQ, Han WQ, Peng X, *et al*: **Genome and proteome of long-chain alkane degrading *Geobacillus thermodenitrificans* NG80-2 isolated from a deep-subsurface oil reservoir.** *Proceedings of the National Academy of Sciences of the United States of America* 2007, **104**(13):5602-5607.
62. Allgaier H, Jung G, Werner RG, Schneider U, Zahner H: **Elucidation of the Structure of Epidermin, a Ribosomally Synthesized, Tetracyclic Heterodetic Polypeptide Antibiotic.** *Angewandte Chemie-International Edition in English* 1985, **24**(12):1051-1053.
63. Kaletta C, Entian KD, Kellner R, Jung G, Reis M, Sahl HG: **Pep5, a New Lantibiotic - Structural Gene Isolation and Prepeptide Sequence.** *Archives of Microbiology* 1989, **152**(1):16-19.

64. Heidrich C, Pag U, Josten M, Metzger J, Jack RW, Bierbaum G, Jung G, Sahl HG: Isolation, characterization, and heterologous expression of the novel lantibiotic epicidin 280 and analysis of its biosynthetic gene cluster. *Appl Environ Microbiol* 1998, **64**(9):3140-3146.
65. Vandekamp M, Vandenhoooven HW, Konings RNH, Bierbaum G, Sahl HG, Kuipers OP, Siezen RJ, Devos WM, Hilbers CW, Vandeven FJM: Elucidation of the Primary Structure of the Lantibiotic Epilancin K7 from *Staphylococcus-Epidermidis* K7 - Cloning and Characterization of the Epilancin-K7-Encoding Gene and Nmr Analysis of Mature Epilancin K7. *European Journal of Biochemistry* 1995, **230**(2):587-600.
66. Scott JC, Sahl HG, Carne A, Tagg JR: Lantibiotic-Mediated Anti-Lactobacillus Activity of a Vaginal *Staphylococcus-Aureus* Isolate. *FEMS Microbiol Lett* 1992, **93**(1):97-102.
67. Geis A, Singh J, Teuber M: Potential of Lactic Streptococci to Produce Bacteriocin. *Appl Environ Microbiol* 1983, **45**(1):205-211.
68. Nes IF, Diep DB, Holo H: Bacteriocin diversity in *Streptococcus* and *Enterococcus*. *Journal of Bacteriology* 2007, **189**(4):1189-1198.
69. Wescombe PA, Tagg JR: Purification and characterization of streptin, a type A1 lantibiotic produced by *Streptococcus pyogenes*. *Appl Environ Microbiol* 2003, **69**(5):2737-2747.
70. MotaMeira M, Lacroix C, LaPointe G, Lavoie MC: Purification and structure of mutacin B-Ny266: A new lantibiotic produced by *Streptococcus mutans*. *Febs Letters* 1997, **410**(2-3):275-279.
71. Qi FX, Chen P, Caufield PW: Purification of mutacin III from group III *Streptococcus mutans* UA787 and genetic analyses of mutacin III biosynthesis genes. *Appl Environ Microbiol* 1999, **65**(9):3880-3887.
72. Qi FX, Chen P, Caufield PW: Purification and biochemical characterization of mutacin I from the group I strain of *Streptococcus mutans*, CH43, and genetic analysis of mutacin I biosynthesis genes. *Appl Environ Microbiol* 2000, **66**(8):3221-3229.
73. Wirawan RE, Kleese NA, Jack RW, Tagg JR: Molecular and genetic characterization of a novel nisin variant produced by *Streptococcus uberis*. *Appl Environ Microbiol* 2006, **72**(2):1148-1156.
74. Hynes WL, Ferretti JJ, Tagg JR: Cloning of the Gene Encoding *Streptococcus* a-Ff22, a Novel Lantibiotic Produced by *Streptococcus-Pyogenes*, and Determination of Its Nucleotide-Sequence. *Appl Environ Microbiol* 1993, **59**(6):1969-1971.
75. Beres SB, Richter EW, Nagiec MJ, Sumbly P, Porcella SF, Deleo FR, Musser JM: Molecular genetic anatomy of inter- and intraserotype variation in the human bacterial pathogen group A *Streptococcus*. *Proceedings of the National Academy of Sciences of the United States of America* 2006, **103**(18):7059-7064.
76. Villani F, Pepe O, Mauriello G, Salzano G, Moschetti G, Coppola S: Antilisterial Activity of Thermophilin-347, a Bacteriocin Produced by *Streptococcus-Thermophilus*. *International Journal of Food Microbiology* 1995, **25**(2):179-190.
77. Ward DJ, Somkuti GA: Characterization of a Bacteriocin Produced by *Streptococcus-Thermophilus* St134. *Applied Microbiology and Biotechnology* 1995, **43**(2):330-335.
78. Aktypis A, Kalantzopoulos G: Purification and characterization of thermophilin ST-1, a novel bacteriocin produced by *Streptococcus thermophilus* ACA-DC 0001. *Lait* 2003, **83**(5):365-378.
79. Bolotin A, Quinquis B, Renault P, Sorokin A, Ehrlich SD, Kulakauskas S, Lapidus A, Goltsman E, Mazur M, Pusch GD, et al: Complete sequence and comparative genome analysis of the dairy bacterium *Streptococcus thermophilus*. *Nature Biotechnology* 2004, **22**(12):1554-1558.
80. Liu MJ, Siezen RJ, Nauta A: In Silico Prediction of Horizontal Gene Transfer Events in *Lactobacillus bulgaricus* and *Streptococcus thermophilus* Reveals Protooperation in Yogurt Manufacturing. *Appl Environ Microbiol* 2009, **75**(12):4120-4129.
81. Ding F, Tang P, Hsu MH, Cui P, Hu S, Yu J, Chiu CH: Genome evolution driven by host adaptations results in a more virulent and antimicrobial-resistant *Streptococcus pneumoniae* serotype 14. *Bmc Genomics* 2009, **10**.
82. Qin JJ, Li RQ, Raes J, Arumugam M, Burgdorf KS, Manichanh C, Nielsen T, Pons N, Levenez F, Yamada T, et al: A human gut microbial gene catalogue established by metagenomic sequencing. *Nature* 2010, **464**(7285):59-U70.
83. Nobles ER: *Bacteroides* infections. *Annals of Surgery* 1973, **177**(5).
84. Sangkhobol V, Skerman VBD: Chitinophaga, A New Genus of Chitinolytic Myxobacteria. *International Journal of Systematic Bacteriology* 1981, **31**(3):285-293.
85. Li B, van der Donk WA: Identification of essential catalytic residues of the cyclase NisC involved in the biosynthesis of nisin. *Journal of Biological Chemistry* 2007, **282**(29):21169-21175.
86. Helfrich M, Entian KD, Stein T: Structure-function relationships of the lantionine cyclase SpaC involved in biosynthesis of the *Bacillus subtilis* peptide antibiotic subtilin. *Biochemistry* 2007, **46**(11):3224-3233.
87. Edgar RC: MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics* 2004, **5**:1-19.
88. Waterhouse AM, Procter JB, Martin DMA, Clamp M, Barton GJ: Jalview Version 2-a multiple sequence alignment editor and analysis workbench. *Bioinformatics* 2009, **25**(9):1189-1191.
89. Felsenstein J: PHYLIP - Phylogeny Inference Package (Version 3.2). *Cladistics* 1989, **5**:164-166.
90. Huson DH, Richter DC, Rausch C, Dezulian T, Franz M, Rupp R: Dendroscope: An interactive viewer for large phylogenetic trees. *BMC Bioinformatics* 2007, **8**.

doi:10.1186/1471-2164-11-679

Cite this article as: Marsh et al.: In silico analysis highlights the frequency and diversity of type 1 lantibiotic gene clusters in genome sequenced bacteria. *BMC Genomics* 2010 **11**:679.

Submit your next manuscript to BioMed Central and take full advantage of:

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at
www.biomedcentral.com/submit

