

Title	In silico identification of bacteriocin gene clusters in the gastrointestinal tract, based on the Human Microbiome Project's reference genome database
Authors	Walsh, Calum J.;Guinane, Caitriona M.;Hill, Colin;Ross, R. Paul;O'Toole, Paul W.;Cotter, Paul D.
Publication date	2015-09-16
Original Citation	WALSH, C. J., GUINANE, C. M., HILL, C., ROSS, R. P., O'TOOLE, P. W. & COTTER, P. D. 2015. In silico identification of bacteriocin gene clusters in the gastrointestinal tract, based on the Human Microbiome Project's reference genome database. BMC Microbiology, 15:183, 1-11. http://dx.doi.org/10.1186/s12866-015-0515-4
Type of publication	Article (peer-reviewed)
Link to publisher's version	10.1186/s12866-015-0515-4
Rights	© Walsh et al. 2015. ; licensee BioMed Central Ltd. 2015. This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The Creative Commons Public Domain Dedication waiver (http://creativecommons.org/publicdomain/zero/1.0/) applies to the data made available in this article, unless otherwise stated. - http://creativecommons.org/licenses/by/4.0/
Download date	2025-07-06 16:46:16
Item downloaded from	https://hdl.handle.net/10468/2182



University College Cork, Ireland
Coláiste na hOllscoile Corcaigh

RESEARCH ARTICLE

Open Access



In silico identification of bacteriocin gene clusters in the gastrointestinal tract, based on the Human Microbiome Project's reference genome database

Calum J. Walsh^{1,3}, Caitriona M. Guinane¹, Colin Hill^{2,3}, R. Paul Ross^{2,3}, Paul W. O'Toole^{2,3} and Paul D. Cotter^{1,2*}

Abstract

Background: The human gut microbiota comprises approximately 100 trillion microbial cells which significantly impact many aspects of human physiology - including metabolism, nutrient absorption and immune function. Disturbances in this population have been implicated in many conditions and diseases, including obesity, type-2 diabetes and inflammatory bowel disease. This suggests that targeted manipulation or shaping of the gut microbiota, by bacteriocins and other antimicrobials, has potential as a therapeutic tool for the prevention or treatment of these conditions. With this in mind, several studies have used traditional culture-dependent approaches to successfully identify bacteriocin-producers from the mammalian gut. *In silico*-based approaches to identify novel gene clusters are now also being utilised to take advantage of the vast amount of data currently being generated by next generation sequencing technologies. In this study, we employed an *in silico* screening approach to mine potential bacteriocin clusters in genome-sequenced isolates from the gastrointestinal tract (GIT). More specifically, the bacteriocin genome-mining tool BAGEL3 was used to identify potential bacteriocin producers in the genomes of the GIT subset of the Human Microbiome Project's reference genome database. Each of the identified gene clusters were manually annotated and potential bacteriocin-associated genes were evaluated.

Results: We identified 74 clusters of note from 59 unique members of the Firmicutes, Bacteroidetes, Actinobacteria, Fusobacteria and Synergistetes. The most commonly identified class of bacteriocin was the >10 kDa class, formerly known as bacteriolysins, followed by lantibiotics and sactipeptides.

Conclusions: Multiple bacteriocin gene clusters were identified in a dataset representative of the human gut microbiota. Interestingly, many of these were associated with species and genera which are not typically associated with bacteriocin production.

Background

Bacteriocins are ribosomally synthesized antimicrobial peptides produced by bacteria that are active against other bacteria, either within the same species (narrow spectrum) or across genera (broad spectrum), and to which the producing organism is immune by a specific immunity protein(s) [1]. Some bacteriocins, most notably nisin, have a long history of use as preservatives in the food industry [2] and these antimicrobials are also

receiving increased attention as potential alternatives to antibiotics [3].

The intestinal microbiota comprises a dynamic community with 100–1000 phylotypes [4, 5] playing an integral role in gastrointestinal (GI) health and disease [6, 7]. As a consequence of advances in DNA sequencing technologies, there is now a clearer understanding of the composition of the GI microbiota and of associations between specific taxa with health and disease [6, 8]. This knowledge can potentially be utilised through the modulation of the gut microbiota to address certain GI disorders [9, 10]. Bacteriocins are ideal candidates with respect to the targeting of undesirable populations due

* Correspondence: paul.cotter@teagasc.ie

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

²APC Microbiome Institute, University College Cork, Cork, Ireland

Full list of author information is available at the end of the article

to their generally low toxicity, high potency and, particularly in the case of gut-associated isolates, the possibility of *in situ* production [11]. There have been some notable proof of concept studies, such as the use of a representative of the sactibiotic group of bacteriocins, thuricin CD, to specifically inhibit *Clostridium difficile* in a distal colon model, without significantly impacting on other members of the microbiota [12]. Similarly, bacteriocin production by the probiotic *Lactobacillus salivarius* UCC118 was shown to be directly responsible for significantly protecting mice against *Listeria monocytogenes* infection [13]. Bacteriocin production has also been investigated to assess the extent to which it can control weight gain as a consequence of changing the composition of the gut microbiota [14, 15].

There are a variety of strategies by which novel bacteriocin producers can be identified [16]. These can be broadly divided into traditional, culture-based approaches and newer, *in silico*-based, strategies. The latter take advantage of the vast amount of data generated by genome and metagenome sequencing projects and the fact that many features of bacteriocin gene clusters, and especially bacteriocin modification genes, are highly conserved. These modification genes encode enzymes responsible for the post-translational modification of Class 1 bacteriocins into their active forms. Other features common to bacteriocin gene clusters include specific immunity genes, ABC transporters for bacteriocin export, and leader cleavage peptidases for removing the leader sequence from the structural prepeptide (for a review see Arnison *et al.* [17]). To date, *in silico* bacteriocin screening strategies have led to the identification of many novel lantibiotic [16, 18–21], microcin [22] and sactibiotic [23] gene clusters of interest. While in a number of instances standard BLAST-based approaches have been employed to identify such clusters, the BAGEL web-based bacteriocin mining tool (<http://bagel.molgenrug.nl/>) has been a particularly valuable resource [24]. BAGEL combines direct mining for the structural gene with indirect mining for bacteriocin-associated genes. The latter is particularly useful for identifying peptides which undergo significant post-translational modification such as those observed in lantibiotics. The most recent iteration of this tool, BAGEL3 [24], was recently used to evaluate the density and diversity of bacteriocins in the human microbiome [25]. A previous version of this software was, for example, used in the identification of the novel, two-peptide lantibiotic lichenicidin [18] and 24 putative novel lantibiotics from genomic data [20]. BAGEL3 classifies clusters in a manner consistent with the generally accepted approach of dividing bacteriocins on the basis of whether they are modified (class I) or unmodified/minimally modified (class II) [1, 11]. The former can be sub-divided into a number of subclasses including the lantibiotics, sactibiotics,

some microcins, bottromycins, and linear azol(in)e-containing peptides (LAPs) [11, 17]. In addition, it also identifies antimicrobial proteins larger than 10 kDa in size (i.e. bacteriolysins, previously referred to as Class III).

Among the large databases of microbiota data that can be screened using *in silico* approaches are those generated by the Human Microbiome Project (HMP). The HMP was established with the goals of characterising the human microbiome, elucidating its role in health and disease, and developing new tools and databases to aid researchers. Among the data generated by the HMP is a reference genome database, which is a collection of genome-sequences from species/strains isolated from a variety of human body sites (<http://www.hmpdacc.org/>). The gastrointestinal tract (GIT) subset of this reference genome database was chosen as the focus of this study, which aimed to find bacteriocin-producers with the potential to alter the composition of the gut microbiota *in situ*. Indeed, previous culture-based approaches have shown the human gut is a rich reservoir of bacteriocin-producers [26–28]. Here we employ the bacteriocin genome-mining tool BAGEL3 to screen the GIT subset of the HMP reference genome database and identify 74 putative bacteriocin-encoding gene clusters (PBGCs) from 59 unique producers.

Results and Discussion

In silico screen for putative bacteriocin-encoding gene clusters

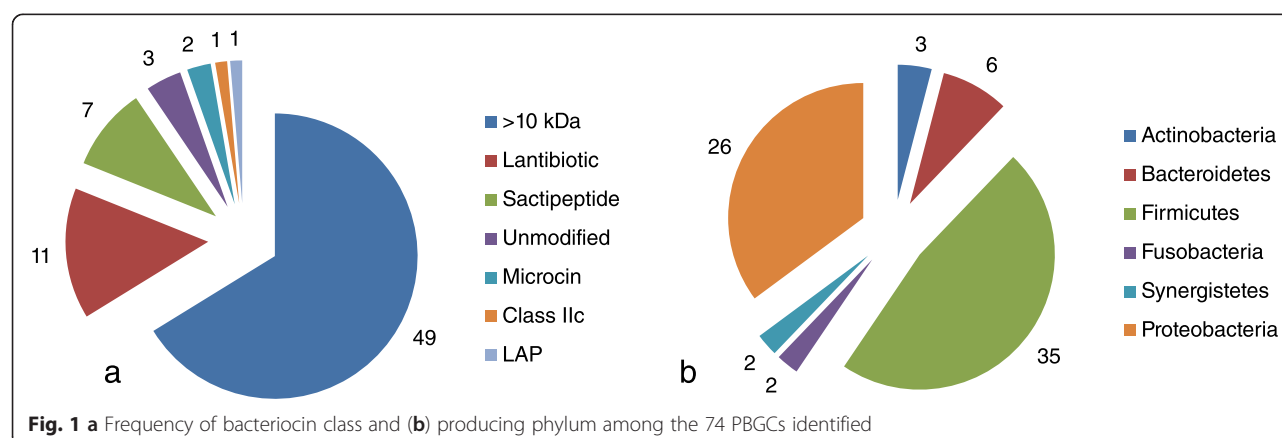
The GIT subset of the HMP reference genome database contained 382 fully sequenced genomes. The bacteriocin mining software tool BAGEL3 initially identified 217 areas of interest (AOIs) from 130 unique putative producers (Additional file 1: Table S1). Subsequent manual annotation and Blast analysis determined that 74 of these were PBGCs (Table 1). The remaining AOIs were eliminated following manual annotation due to the absence of key bacteriocin associated genes. However, we accept the possibility that these gene products may work in concert with other novel bacteriocin-related genes encoded elsewhere on the genome. Selection of the 74 PBGCs was achieved based on the presence of bacteriocin-associated genes, arrangement of those genes in the AOI, and by overall similarity to previously described gene clusters. An overall breakdown of the 74 PBGCs according to phylum and predicted bacteriocin type can be seen in Fig. 1a, b, respectively. The vast majority of PBGCs belonged to members of the Firmicutes and Proteobacteria phyla, and, in the latter case, *Escherichia coli* strains in particular. PBGCs were also identified in the Bacteroidetes, Actinobacteria, Fusobacteria and Synergistetes phyla. The most commonly identified clusters were > 10 kDa bacteriolysins followed by lantibiotics and sactipeptides (Fig. 1).

Table 1 Additional information on PBGCs and whether the initial identification of the AOI by BAGEL3 was based on the presence of bacteriocin-associated genes (context) or a specific bacteriocin structural gene

Potential Producer	Phylum	Class	BAGEL3 prediction
<i>Bifidobacterium longum</i> subsp. <i>infantis</i> JCM 1222	Actinobacteria	Lantibiotic	Context
<i>Bifidobacterium</i> sp. 12 1 47BFAA	Actinobacteria	Lantibiotic	BLD_1648
<i>Eggerthella</i> sp. HGA1	Actinobacteria	>10 kDa	Linocin M18
<i>Bacteroides dorei</i> DSM 17855	Bacteroidetes	Sactipeptide	Context
<i>Bacteroides fragilis</i> 3 1 12	Bacteroidetes	Sactipeptide	Context
<i>Bacteroides</i> sp. 2 1 16	Bacteroidetes	Lantibiotic	Manual
<i>Bacteroides</i> sp. 2 1 56FAA	Bacteroidetes	Unmodified	Manual
<i>Bacteroides</i> sp. 9 1 42FAA	Bacteroidetes	Sactipeptide	Context
<i>Bacteroides uniformis</i> ATCC 8492	Bacteroidetes	Sactipeptide	Context
<i>Anaerofustis stercorihominis</i> DSM 17244	Firmicutes	>10 kDa	Linocin M18
<i>Bacillus</i> sp. 7 6 55CFAA CT2	Firmicutes	>10 kDa	Colicin E9
<i>Bacillus</i> sp. 7 6 55CFAA CT2	Firmicutes	Lantibiotic	Haloduracin
<i>Dorea formicigenerans</i> 4 6 53AFAA	Firmicutes	Sactipeptide	Context
<i>Enterococcus faecalis</i> PC1.1	Firmicutes	>10 kDa	Enterolysin A
<i>Enterococcus faecalis</i> TX1302	Firmicutes	>10 kDa	Enterolysin A
<i>Enterococcus faecalis</i> TX1302	Firmicutes	Lantibiotic	Context
<i>Enterococcus faecalis</i> TX1341	Firmicutes	>10 kDa	Enterolysin A
<i>Enterococcus faecalis</i> TX1342	Firmicutes	>10 kDa	Enterolysin A
<i>Enterococcus faecalis</i> TX1342	Firmicutes	Lantibiotic	Context
<i>Enterococcus faecalis</i> TX1467	Firmicutes	>10 kDa	Enterolysin A
<i>Enterococcus faecalis</i> TX1467	Firmicutes	Lantibiotic	Context
<i>Enterococcus faecalis</i> TX2137	Firmicutes	>10 kDa	Enterolysin A
<i>Enterococcus faecalis</i> TX2137	Firmicutes	Lantibiotic	Context
<i>Enterococcus faecalis</i> TX4244	Firmicutes	>10 kDa	Enterolysin A
<i>Enterococcus faecalis</i> TX4244	Firmicutes	>10 kDa	Enterolysin A
<i>Holdemania filiformis</i> DSM 12042	Firmicutes	>10 kDa	Linocin M18
<i>Lactobacillus acidophilus</i> ATCC 4796	Firmicutes	>10 kDa	Enterolysin A
<i>Lactobacillus acidophilus</i> ATCC 4796	Firmicutes	>10 kDa	Helveticin J
<i>Lactobacillus antri</i> DSM 16041	Firmicutes	>10 kDa	Enterolysin A
<i>Lactobacillus brevis</i> subsp. <i>gravesensis</i> ATCC 27305	Firmicutes	Unmodified	Plantaricin NC8
<i>Lactobacillus delbrueckii</i> subsp. <i>lactis</i> DSM 20072	Firmicutes	>10 kDa	Enterolysin A
<i>Lactobacillus helveticus</i> DSM 20075	Firmicutes	>10 kDa	Helveticin J
<i>Lactobacillus helveticus</i> DSM 20075	Firmicutes	>10 kDa	Helveticin J
<i>Lactobacillus ultunensis</i> DSM 16047	Firmicutes	>10 kDa	Helveticin J
<i>Lactobacillus ultunensis</i> DSM 16047	Firmicutes	>10 kDa	Enterolysin A
<i>Lactobacillus ultunensis</i> DSM 16047	Firmicutes	>10 kDa	Helveticin J
<i>Listeria innocua</i> ATCC 33091	Firmicutes	LAP	Context
<i>Marvinbryantia formatexigens</i> DSM 14469	Firmicutes	Sactipeptide	Context
<i>Roseburia intestinalis</i> L1 82	Firmicutes	Sactipeptide	Context
<i>Ruminococcus obeum</i> A2 162	Firmicutes	Lantibiotic	Context
<i>Ruminococcus</i> sp. 5 1 39B FAA	Firmicutes	Lantibiotic	Context
<i>Streptococcus anginosus</i> 1 2 62CV	Firmicutes	Unmodified	Multiple
<i>Streptococcus infantarius</i> subsp. <i>infantarius</i> ATCC BAA 102	Firmicutes	Unmodified	Multiple

Table 1 Additional information on PBGCs and whether the initial identification of the AOI by BAGEL3 was based on the presence of bacteriocin-associated genes (context) or a specific bacteriocin structural gene (*Continued*)

<i>Streptococcus</i> sp. 2 1 36FAA	Firmicutes	Class IIc	Context
<i>Fusobacterium ulcerans</i> ATCC 49185	Fusobacteria	>10 kDa	Linocin M18
<i>Fusobacterium varium</i> ATCC 27725	Fusobacteria	>10 kDa	Linocin M18
<i>Arcobacter butzleri</i> JV22	Proteobacteria	>10 kDa	Colicin E9
<i>Escherichia coli</i> MS 110 3	Proteobacteria	>10 kDa	Colicin
<i>Escherichia coli</i> MS 110 3	Proteobacteria	>10 kDa	Colicin
<i>Escherichia coli</i> MS 119 7	Proteobacteria	>10 kDa	Colicin
<i>Escherichia coli</i> MS 124 1	Proteobacteria	>10 kDa	Colicin
<i>Escherichia coli</i> MS 146 1	Proteobacteria	>10 kDa	Linocin M18
<i>Escherichia coli</i> MS 153 1	Proteobacteria	>10 kDa	Colicin
<i>Escherichia coli</i> MS 16 3	Proteobacteria	>10 kDa	Colicin
<i>Escherichia coli</i> MS 16 3	Proteobacteria	>10 kDa	Colicin
<i>Escherichia coli</i> MS 16 3	Proteobacteria	>10 kDa	Colicin
<i>Escherichia coli</i> MS 185 1	Proteobacteria	>10 kDa	Colicin E9
<i>Escherichia coli</i> MS 196 1	Proteobacteria	>10 kDa	Colicin-10
<i>Escherichia coli</i> MS 200 1	Proteobacteria	>10 kDa	Colicin
<i>Escherichia coli</i> MS 21 1	Proteobacteria	>10 kDa	Colicin
<i>Escherichia coli</i> MS 45 1	Proteobacteria	Microcin	Microcin H47
<i>Escherichia coli</i> MS 45 1	Proteobacteria	>10 kDa	Colicin
<i>Escherichia coli</i> MS 57 2	Proteobacteria	>10 kDa	Colicin E9
<i>Escherichia coli</i> MS 78 1	Proteobacteria	>10 kDa	Colicin
<i>Escherichia coli</i> MS 85 1	Proteobacteria	>10 kDa	Colicin E9
<i>Escherichia coli</i> SE11	Proteobacteria	>10 kDa	Colicin
<i>Escherichia</i> sp. 3 2 53FAA	Proteobacteria	>10 kDa	Colicin
<i>Klebsiella</i> sp. MS 92 3	Proteobacteria	>10 kDa	Klebicin B
<i>Providencia rettgeri</i> DSM 1131	Proteobacteria	>10 kDa	Colicin A
<i>Yokenella regensburgei</i> ATCC 43003	Proteobacteria	Microcin	Context
<i>Anaerobaculum hydrogeniformans</i> ATCC BAA 1850	Synergistetes	>10 kDa	Linocin M18
<i>Synergistes</i> sp. 3 1 syn1	Synergistetes	>10 kDa	Linocin M18



Further analysis of PBGCs of particular interest

Sixty-three PBGCs are described in the Additional file 2: Supplementary Text and depicted in Additional file 3: Figures S1, Additional file 4: Figure S2 and Additional file 5: Figure S3. 11 PBGCs from 3 different phyla were deemed of particular interest and were selected for further *in silico* analysis based on the relative rarity with which bacteriocin production has been associated with the corresponding genus (*Bacteroides* and *Roseburia*), on the probiotic potential of strains from the genus (*Bifidobacterium*) or due to the importance/perceived importance of the genus in a gut environment (*Bacteroides*, *Roseburia*, *Ruminococcus*) (Fig. 2).

Identification of novel PBGCs in bifidobacteria

Bifidobacteria are an important group of human gut commensal bacteria, accounting for between 3 and 7 % of the gut microbiota in adults and up to 91 % in newborns [29]. Members of this genus have a long history of use as health-promoting/probiotic strains due to traits such as the regulation of intestinal microbial homeostasis, the inhibition of pathogens, the modulation of local and systemic immune responses, the maintenance of gastrointestinal barrier function, the production of vitamins and the bioconversion of a number of dietary compounds into bioactive molecules [30]. Bifidobacteria have the potential to

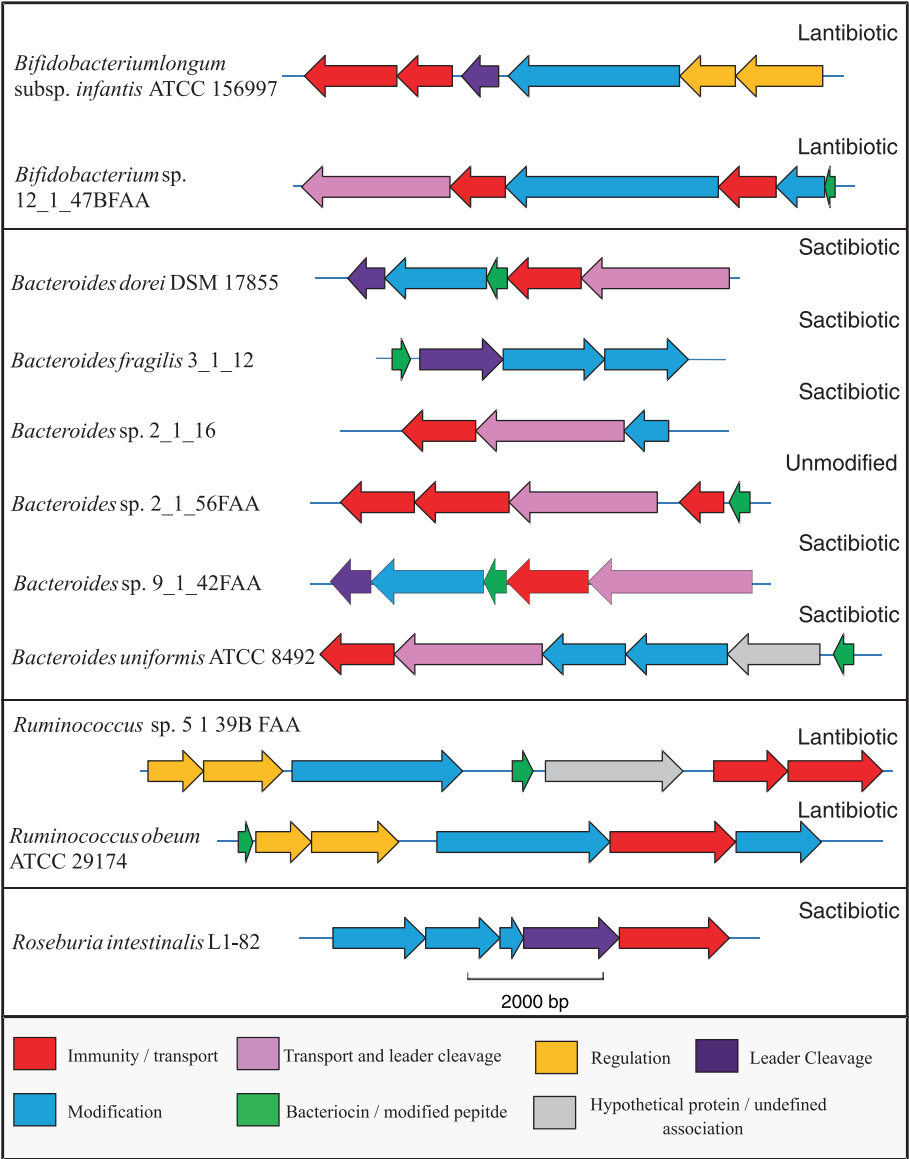


Fig. 2 Diagrammatic representation of PBGCs deemed of particular interest

suppress the growth of both Gram-negative and Gram-positive bacteria but, to date, this activity has been more often attributed to the inhibitory action of organic acids rather than bacteriocin production [31, 32]. For a review of the relatively rare examples of bacteriocin production by bifidobacteria see Martinez *et al.* [31]. Our *in silico* screen identified PBGCs of note in *Bifidobacterium longum* subsp. *infantis* ATCC 15697 and *Bifidobacterium* sp. 12_1_47BFA (Fig. 1).

Bifidobacterium longum subsp. *infantis* ATCC 15697 was isolated from human infant faeces and sequenced by the Joint Genome Institute (JGI) [33, 34]. A previous study has shown that this strain has the ability to reduce the levels of plasma endotoxins via modulation of the gut microbiota. However the authors concluded that the effect was mediated by increased levels of faecal organic acids [35]. The cluster of six genes identified are predicted to encode a LanL-type lantipeptide based on the presence of a LanL-type lanthionine synthetase gene. More specifically, the 8,139 bp cluster contains several lantibiotic-related genes including a putative lanthionine synthetase (conserved domain pfam05147 3.10e-10), a putative oligopeptidase (conserved domain pfam00326 5.24e-08) and a putative ABC transporter containing ATP-binding and permease subunits (conserved domains cd03255 and pfam02867 respectively). The cluster also contained a two-component regulatory system consisting of a putative histidine kinase (conserved domain CGO4585 6.70e-18) and a putative transcriptional response regulator (conserved domain COG2197 8.85e-57).

Bifidobacterium sp. 12_1_47BFA was recovered from inflamed biopsy tissue from a 25-year-old female patient with Crohn's disease and its genome was found to contain a 7,996 bp lantibiotic cluster comprising six genes (Fig. 1). A putative lantibiotic prepeptide LanA was found to be similar to BLD_1648 (BAGEL3 bacteriocin I database 4e-43), a feature that was further supported by manual annotation (conserved domain TIGR03893 6.47e-9). Also present in the area of interest was a putative LanM lantibiotic biosynthesis protein (conserved domain cd04792 0.0), a putative multidrug ABC transporter ATP-binding protein putatively involved in lantibiotic immunity (conserved domain cd03230 8.53e-42) and an ABC-type bacteriocin/lantibiotic exporter (conserved domain COG2274 7.59e-145) significantly similar (BlastP 4e-117) to the *crnT* protein responsible for transport and leader cleavage of the bacteriocin carnolysin [36]. The area of interest also contained a FMN-dependent reductase (conserved domain pfam03358 5.13e-09) similar to that located within the carnolysin-associated *crnJ* protein [36]. This family of proteins has been suggested to be an atypical lantibiotic post-translational modification protein [20, 37].

Identification of novel PBGCs in *Bacteroides* spp.

Bacteroides are Gram-negative, non-spore-forming, obligate anaerobes and near universal constituents of the human gut microbiota, especially prevalent in those individuals whose long-term diets are rich in protein and animal fat [38]. Translocation from the GIT can however result, in some cases, in bacteraemia and abscess formation [39]. Weight loss in obese humans subjected to dietary or surgical intervention has been associated with increased relative abundance in the phylum Bacteroidetes, with specific members including *Bacteroides* spp., *Bacteroides-Prevotella* spp. or the *Bacteroides fragilis* group bacteria having been associated with this phenomenon [40–43]. Despite their importance as a human gut commensal, there have been relatively few reports of bacteriocin production by members of the *Bacteroides* to date [44–47]. In this study, six PBGCs were identified in *Bacteroides* strains that possessed features typical of sactipeptide (4), lantibiotic (1) or unmodified bacteriocin (1) clusters.

Bacteroides dorei has been observed to be common in patients with active coeliac disease and it has also been proposed that the species be used as an indicator of water contamination by human faecal material [48, 49]. *B. dorei* DSM 17855 was isolated from a healthy, 23 year old, Japanese male [50] and its genome was found to contain a five gene, 5,711 bp sactipeptide-like gene cluster (Fig. 1). The cluster contained genes encoding a putative ABC-type transporter ATP-binding protein (BlastP 0.0, conserved domain COG2274 3.02e-34), a putative hemolysin secretion protein HlyD (BlastP 0.0), a structural gene belonging to pfam family pf10439 (Bacteriocin class II with double-glycine leader peptide), a radical SAM domain-containing protein hypothesised to be involved in peptide modification (conserved domain TIGR03962 1.46e-06) and a putative bacteriocin-associated C39 family peptidase (conserved domain pfam03412 1.13e-11). The latter may be involved in transport across the membrane in addition to leader cleavage, either alone or in conjunction with HlyD.

Bacteroides fragilis-produced metabolites are important in the activation and regulation of the T-cell-dependent immune response [39, 51] and its administration as a therapeutic has been proposed for gastrointestinal and behavioural symptoms associated with human neurodevelopmental disorders [52]. The genome of *B. fragilis* 3_1_12 found to contain a four gene, 4,267 bp sactipeptide-like cluster (Fig. 1). The putative structural gene belongs to pfam family PF14406 (Ribosomally synthesized peptide in Bacteroidetes) and BlastP identified it as a putative bacteriocin-type signal sequence containing a predicted leader sequence associated with peptide modification (conserved domain TIGR04149 1.34e-12). Immediately downstream is a putative lipoprotein belong to pfam family PF08139 followed by a pair of putative radical SAM

proteins, predicted to be involved in peptide modification. These radical SAM proteins, members of families TIGR04085 and TIGR04150, respectively, are known to occur in cassettes together with the bacteriocin signal sequence noted above [53].

Bacteroides sp. 2_1_16 was isolated from a healthy biopsy of the descending colon of a 58-year old female patient undergoing colonoscopy its genome was found to contain a 4,167 bp, three-gene cluster predicted to be sactipeptide-encoding based on the presence of a SacCD homolog (Fig. 1). However, manual annotation also revealed a cluster of several genes with homology with those typically associated with lantibiotic production. Specifically, the cluster contained a putative LanC-like lanthionine synthetase (conserved domain cd04793 6.02e-08), a putative ABC transporter predicted to be a bacteriocin/lantibiotic transporter based on conserved domains (COG2274 0.0) and a putative ABC transporter secretion protein closely related to hemolysin secretors (conserved domain TIGR01843 1.86e-22). However, a putative structural peptide-encoding gene could not be identified in this gene cluster.

The genome of *Bacteroides* sp. 2_1_56FAA was found to possess a 6,069 bp cluster containing five genes of note (Fig. 1). Manual annotation revealed a gene predicted to encode a ribosomally synthesised peptide (pfam PF14406 0.00024 [54]), located immediately upstream of a putative CAAX protease self-immunity family determinant (conserved domain pfam02517 8.17e-11). A gene encoding a putative ABC transporter containing a C39B peptidase domain (COG2274 7.75e-159), predicted to be responsible for transport and leader cleavage, was also present. Two additional possible transport genes were identified immediately downstream, both putative hemolysin secretion proteins (conserved domain pfam13437 5.74e-09 and conserved domain pfam13437 5.37e-11, respectively). The lack of any bacteriocin-modification genes suggests that this cluster encodes an unmodified bacteriocin.

Bacteroides sp. 9_1_42FAA was isolated from the duodenum of a 47 year old female patient and its genome contained a 5,714 bp area of interest comprised five genes. This cluster was identified as a potential sactipeptide based on the presence of a SacCD homolog (Fig. 1). The structural peptide putatively encoded within this cluster also possesses features associated with pfam family PF10439.4 i.e. unmodified subclass IIc bacteriocins. The area of interest also contains a putative ABC-type bacteriocin/lantibiotic exporter (contains conserved domain COG2274 0.0), a putative hemolysin secretion family protein (conserved domain TIGR01843 3.45e-06), a putative radical SAM peptide modification protein (conserved domain TIGR03962 1.47e-17), and a putative bacteriocin transporter containing an endopeptidase C39 domain (potentially involved in bacteriocin preprocessing;

conserved domain pfam03412 1.13e-11) [55]. This sequence exhibited very high (99 %) nucleotide identity to the aforementioned gene cluster in *B. dorei* DSM 17855. This similarity includes structural genes with 100 % amino acid sequence identity.

It has been previously documented that orally administering *Bacteroides uniformis* (strain CECT 7771) ameliorated high fat diet-induced metabolic and immune dysfunction associated with an altered gut microbiota in adult C57BL-6 mice [56]. Inspection of the genome of *B. uniformis* ATCC 8492 revealed a 7,976 bp, five-gene sactipeptide-like cluster (Fig. 1). Manual annotation identified a putative bacteriocin-type signal sequence containing a conserved TIGR04149 domain (7.43e-09). The area of interest also contained a pair of putative peptide-modifying radical SAM proteins (conserved domains TIGR04148 and TIGR04150 respectively) similar to those in *B. fragilis* 3_1_12 that were referred to above, a putative ABC-type bacteriocin exporter (conserved domain COG2274 0.0) and a putative hemolysin secretion protein (conserved domain pfam13437 1.02e-16).

Identification of novel PBGCs in *Ruminococcus* spp.

Ruminococci are Gram-positive anaerobes commonly found in the human gut, where they have been proposed to play a pivotal role in the fermentation of resistant starch [57]. There have been several previous reports of bacteriocin production by members of the ruminococci, including a class IIa lantibiotic, ruminococcin A, produced by *Ruminococcus gnavus* E1 and two distinct class III bacteriocins produced by *Ruminococcus albus* 7 [58–60]. We identified two apparently novel *Ruminococcus*-associated PBGCs, from among a total of 35 Firmicutes-associated clusters (Additional file 2: Supplementary Text).

The genome of *Ruminococcus* sp. 5_1_39_B_FAA contained a 13,553 bp lantibiotic-like cluster containing six genes (Fig. 1). The cluster contained a putative response regulator receiver protein (conserved domain COG3279 3.95e-24), a putative histidine kinase (conserved domain pfam14501 3.5e-20), a putative type 2 lantibiotic biosynthesis protein LanM (conserved domain TIGR03897 0.0), a putative UviB-like bacteriocin (BAGEL3 bacteriocin II database 3e-11), a putative ABC transporter ATP-binding protein (conserved domain COG1136 8.20e-111) and a putative efflux ABC transporter permease protein.

Strains of *Ruminococcus obeum* have been shown to restrict *Vibrio cholerae* infection via a quorum-sensing-mediated mechanism [61]. *Ruminococcus obeum* ATCC 29174 was isolated from human faeces and sequenced by the Washington University Genome Sequencing Centre. A 8,879 bp lantibiotic-like cluster comprising six genes was identified (Fig. 1). The putative structural gene was found to resemble geobacillin I (BAGEL3 bacteriocin I database 5e-12), a nisin homolog isolated from

Geobacillus thermodenitrificans [62]. Also present in the area of interest were genes that appear to encode a two-component regulatory system, consisting of a putative histidine kinase (conserved domain COG0642 1.84e-24) and a putative *NisR* homolog containing signal receiver and effector domains (cd00156 and cd00383 respectively). Furthermore, genes potentially encoding a lantibiotic dehydratase similar to the entianin (lantibiotic) modification protein *EtnB* (BlastP 0.0) [63], an ABC transport protein similar to SpaT (transportation of the lantibiotic subtilin; BlastP 0.0) and a lanthionine synthetase protein similar to SpaC (modification of subtilin; BlastP 6e-117) were identified.

Identification of a novel PBGC in *Roseburia* spp.

Roseburia is a genus of Gram positive, butyrate-producers found to be negatively associated with type 2 diabetes and ulcerative colitis [64, 65]. It has also been linked with ameliorating high-fat diet induced metabolic alterations in mice [66]. The only *Roseburia*-associated bacteriocin-producer to have been identified to date is *Roseburia faecis* M72/1 [67]. *Roseburia intestinalis* L1-82, the type strain, was found to contain a five gene, 6078 bp sactipeptide-like cluster (Fig. 1). The area of interest contained a putative bacteriocin-associated radical SAM protein (conserved domain TIGR04068 0.0), a putative peptide maturation system protein (conserved domain TIGR04066 8.58e-165), a putative peptide maturation system acyl carrier-related protein (conserved domain TIGR04069 1.15e-29), a subtilase family serine protease (conserved domain cd07492 7.11e-40) and a putative ABC transporter (conserved domain cd03228 5.95e-65). However, there were no immediately obvious bacteriocin structural or immunity genes in the area of interest and so it is particularly unclear if this cluster has the potential to produce an antimicrobial.

Conclusions

The large number of fully sequenced genomes available in public repositories means that genome-mining approaches are increasingly valuable with respect to the identification of novel genes and gene clusters [68–70]. As it has already been established that *in silico* approaches can be applied to the human microbiome for the purpose of identifying antimicrobial-producing microorganisms [25, 71], and that bacteriocins identified in this manner can be produced *in vitro* [18], it is apparent that there are considerable potential benefits in screening for and harnessing putative bacteriocin gene clusters from such databases.

It is commonly reported that between 30 and 99 % of bacteria have the potential to produce at least one bacteriocin [72, 73]. It is thus notable that this *in silico*-based study identified just 59 genomes encoding

probable PBGCs from 382 reference genomes, a frequency of just 15.4 %. It is unclear whether this low number is representative of bacteriocin-production in the human GIT or an underestimation due to biases in identification of gene clusters. In support of the former of these theories, a recent study on the human microbiome by Zheng *et al.* reported that the gut contained the lowest density of putative bacteriocin genes of all body sites investigated [25]. That study identified 123 putative lantibiotic, 56 putative class II bacteriocin and 148 putative class III bacteriocin gene clusters in the gut environment. Interestingly, only one sactipeptide of gut origin, a subtilisin A, was reported by Zheng *et al.* [25]. The discrepancy between the results reported by this study and those reported by Zheng *et al.* can be explained by differences in methodology. This method used BAGEL3 for the initial analysis while Zheng *et al.* performed a PSI-BLAST-based approach using the amino acid sequences from the BAGEL3 bacteriocin database as driver sequences. Furthermore, we manually annotated the potential clusters returned initially, resulting in a dramatic decrease in reported PBGCs. It is noteworthy that *in silico* screens are limited by their dependence on similarity to previously described bacteriocin-associated genes, meaning that it is possible to overlook completely novel bacteriocin clusters.

The vast majority of known/characterised lantibiotics are produced by members of the Firmicutes [74]. Similarly, of the 11 lantibiotic PBGCs identified in this study, seven were found in the genomes of Firmicutes, with two associated with bifidobacteria (Actinobacteria) and two with *Bacteroides* spp. (Bacteroidetes). While these clusters typically contained features that are common to lantibiotic-associated gene clusters, two putative lantibiotic clusters (in *Bifidobacterium* sp. 12_1_47BFAA and *Enterococcus faecalis* TX1342 (Additional file 2: Supplementary Results; Additional file 3: Figures S1 and Additional file 5: Figure S3 respectively)) contained predicted FMN reductase genes in addition to those more traditionally associated with lantibiotic modification.

It is apparent that the *in silico* screen identified gene clusters representative of some classes of bacteriocin more frequently than others. Clusters resembling those associated with the production of bacteriolysins (formerly referred to as class III bacteriocins) were most common. The large numbers of colicin-like and enterolysin A-like clusters was possibly due to the overrepresentation of *E. coli* in the reference genome database and the relative ease of detection. It appears that enterolysin A does not possess a specific immunity gene; instead, resistance results from the absence of specific binding receptors [75], making this single gene potentially easier to detect than a multi-gene operon. On the other hand, the relatively low frequency of class II bacteriocins (three unmodified and

one class IIc) cannot be explained in a similar manner. It is unclear whether this paucity is due to the methodology or an actual scarcity of class II bacteriocin producers in the gut microbiota. Comparatively, Zheng *et al.* identified 56 class II bacteriocin structural genes from gut-associated strains [25] suggesting that either this is an overestimation due to the lack of manual annotation or the approach used in this study is not ideal for the identification of Class II bacteriocins.

In several cases, complete gene clusters were identified that lacked an obvious bacteriocin structural gene. Compared to other classes, the number of described and characterised sactipeptides is relatively small so it may be possible that BAGEL3 and the nr database do not contain any homologs of the structural proteins encoded by *Bacteroides* sp. 2_1_16 and *Roseburia intestinalis* L1-82. This may also explain the relatively low incidence of sactipeptides reported by Zheng *et al.* [25]. The putative lantibiotic cluster identified in *Bifidobacterium longum* subsp. *infantis* ATCC 156997 was also missing an obvious structural gene but may be explained by the same hypothesis, as it is a potential LanL-type lantibiotic, a subclass which contains only one previously described member Venezuelin [76].

This comprehensive *in silico* study led to the identification of PBGCs in species not previously associated with bacteriocin production, for example *Bacteroides uniformis* and *Roseburia intestinalis*. We also identified potential bacteriocin gene clusters in two *Bifidobacterium* species, a genus which has long been thought of as beneficial to the human host. It is not possible, by *in silico* methods alone, to state conclusively if these bacteriocins are produced *in vitro*. However, if even a portion of these gene clusters are responsible for bacteriocin production in the corresponding strain, it could greatly expand the arsenal of bacteriocins available for use in food and healthcare. Such investigations will be the focus of our future studies.

Methods

Initial screening of reference genomes for bacteriocin gene clusters

The GIT subset (382 available sequences as of 20/11/2014) of the HMP's reference genome database (<http://www.hmpdacc.org/HMRGD/>) was downloaded in multi-FASTA format and both complete and draft genomes were screened for putative bacteriocin gene clusters using the web-version of BAGEL3 (<http://bagel2.molgenrug.nl/index.php/bagel3>).

Further investigation of individual gene clusters

Approximately 20 kb of sequence data containing the gene/genes identified as being of potential interest by BAGEL3 were extracted and the sequences were manually annotated using the software ARTEMIS [77].

Predicted coding regions were analysed using the BlastP web server on NCBI (<http://www.ncbi.nlm.nih.gov/BLAST>) and the nr database. The coding regions were also analysed for the presence of conserved domains and, where applicable, compared to previously described gene clusters using the Artemis Comparison Tool (ACT) [78].

Additional files

Additional file 1: Table S1. List of the 130 unique putative producers identified by BAGEL3 (DOCX 15 kb)

Additional file 2: Supplementary Text. Description of remaining 63 PBGCs not covered in main text (DOCX 39 kb)

Additional file 3: Figure S1. Diagrammatic representation of remaining PBGCs identified in the Actinobacteria, Fusobacteria and Synergistetes (DOCX 25 kb)

Additional file 4: Figure S2. Diagrammatic representation of remaining PBGCs identified in the Firmicutes (DOCX 62 kb)

Additional file 5: Figure S3. Diagrammatic representation of remaining PBGCs identified in the Proteobacteria (DOCX 52 kb)

Abbreviations

ACT: Artemis Comparison Tool; AOI: Area of Interest; GI: Gastrointestinal; GIT: Gastrointestinal Tract; HMP: Human Microbiome Project; JGI: Joint Genome Institute; LAP: Linear Azol(in)e-containing Peptide; PBGC: Putative Bacteriocin Gene Cluster.

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

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

Acknowledgements

CJW, CMG and PDC are supported by a SFI PI award to PDC "Obesibiotics" (11/PI/1137).

Author details

¹Teagasc Food Research Centre, MooreparkFermoy, Cork, Ireland. ²APC Microbiome Institute, University College Cork, Cork, Ireland. ³School of Microbiology, University College Cork, Cork, Ireland.

Received: 7 May 2015 Accepted: 8 September 2015

Published online: 16 September 2015

References

- Cotter PD, Hill C, Ross RP. Bacteriocins: developing innate immunity for food. *Nat Rev Microbiol.* 2005;3(10):777–88. doi:10.1038/nrmicro1273.
- Deegan LH, Cotter PD, Hill C, Ross P. Bacteriocins: Biological tools for bio-preservation and shelf-life extension. *Int Dairy J.* 2006;16(9):1058–71. <http://dx.doi.org/10.1016/j.idairyj.2005.10.026>.
- Piper C, Cotter PD, Ross RP, Hill C. Discovery of medically significant lantibiotics. *Curr Drug Discov Technol.* 2009;6(1):1–18.
- Qin J, Li R, Raes J, Arumugam M, Burgdorf KS, Manichanh C, et al. A human gut microbial gene catalogue established by metagenomic sequencing. *Nature.* 2010;464(7285):59–65. doi:10.1038/nature08821.
- Faith JJ, Guruge JL, Charbonneau M, Subramanian S, Seedorf H, Goodman AL, et al. The Long-Term Stability of the Human Gut Microbiota. *Science.* 2013;341:6141. doi:10.1126/science.1237439.

6. Clemente Jose C, Ursell Luke K, Parfrey Laura W, Knight R. The Impact of the Gut Microbiota on Human Health: An Integrative View. *Cell*. 2012;148(6):1258–70. <http://dx.doi.org/10.1016/j.cell.2012.01.035>.
7. Flint HJ, Scott KP, Louis P, Duncan SH. The role of the gut microbiota in nutrition and health. *Nat Rev Gastroenterol Hepatol*. 2012;9(10):577–89.
8. Karlsson F, Tremaroli V, Nielsen J, Backhed F. Assessing the human gut microbiota in metabolic diseases. *Diabetes*. 2013;62(10):3341–9. doi:10.2337/db13-0844.
9. Kadooka Y, Sato M, Imaizumi K, Ogawa A, Ikuyama K, Akai Y, et al. Regulation of abdominal adiposity by probiotics (*Lactobacillus gasseri* SBT2055) in adults with obese tendencies in a randomized controlled trial. *Eur J Clin Nutr*. 2010;64(6):636–43. doi:10.1038/ejcn.2010.19.
10. Xiao S, Fei N, Pang X, Shen J, Wang L, Zhang B, et al. A gut microbiota-targeted dietary intervention for amelioration of chronic inflammation underlying metabolic syndrome. *FEMS Microbiol Ecol*. 2014;87(2):357–67. doi:10.1111/1574-6941.12228.
11. Cotter PD, Ross RP, Hill C. Bacteriocins - a viable alternative to antibiotics? *Nat Rev Micro*. 2013;11(2):95–105.
12. Rea MC, Dobson A, O'Sullivan O, Crispie F, Fouhy F, Cotter PD, et al. Effect of broad- and narrow-spectrum antimicrobials on *Clostridium difficile* and microbial diversity in a model of the distal colon. *Proc Natl Acad Sci*. 2011;108(Supplement 1):4639–44. doi:10.1073/pnas.1001224107.
13. Corr SC, Li Y, Riedel CU, O'Toole PW, Hill C, Gahan CGM. Bacteriocin production as a mechanism for the anti-infective activity of *Lactobacillus salivarius* UCC118. *Proc Natl Acad Sci*. 2007;104(18):7617–21. doi:10.1073/pnas.0700440104.
14. Murphy EF, Cotter PD, Hogan A, O'Sullivan O, Joyce A, Fouhy F, et al. Divergent metabolic outcomes arising from targeted manipulation of the gut microbiota in diet-induced obesity. *Gut*. 2013;62(2):220–6. doi:10.1136/gutjnl-2011-300705.
15. Riboulet-Bisson E, Sturme MHJ, Jeffery IB, O'Donnell MM, Neville BA, Forde BM, et al. Effect of *Lactobacillus salivarius* Bacteriocin Abp118 on the Mouse and Pig Intestinal Microbiota. *PLoS One*. 2012;7(2):e31113. doi:10.1371/journal.pone.0031113.
16. Marsh AJ, O'Sullivan O, Ross RP, Cotter PD, Hill C. In silico analysis highlights the frequency and diversity of type 1 lantibiotic gene clusters in genome sequenced bacteria. *BMC Genomics*. 2010;11:679. doi:10.1186/1471-2164-11-679.
17. Arison PG, Bibb MJ, Bierbaum G, Bowers AA, Bugni TS, Bulaj G, et al. Ribosomally synthesized and post-translationally modified peptide natural products: overview and recommendations for a universal nomenclature. *Nat Prod Rep*. 2013;30(1):108–60. doi:10.1039/C2NP20085F.
18. 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):5451–60. doi:10.1128/aem.00730-09.
19. 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. doi:10.1111/j.1574-6968.2006.00539.x.
20. Singh M, Sareen D. Novel LanT associated lantibiotic clusters identified by genome database mining. *PLoS One*. 2014;9(3):e91352. doi:10.1371/journal.pone.0091352.
21. McClarren AL, Cooper LE, Quan C, Thomas PM, Kelleher NL, van der Donk WA. Discovery and in vitro biosynthesis of haloduracin, a two-component lantibiotic. *Proc Natl Acad Sci U S A*. 2006;103(46):17243–8. doi:10.1073/pnas.0606088103.
22. Scholz R, Molohon KJ, Nachtigall J, Vater J, Markley AL, Sussmuth RD, et al. Plantazolicin, a novel microcin B17/streptolysin S-like natural product from *Bacillus amyloliquefaciens* FZB42. *J Bacteriol*. 2011;193(1):215–24. doi:10.1128/jb.00784-10.
23. Murphy K, O'Sullivan O, Rea MC, Cotter PD, Ross RP, Hill C. Genome Mining for Radical SAM Protein Determinants Reveals Multiple Sactibiotic-Like Gene Clusters. *PLoS One*. 2011;6(7):e20852. doi:10.1371/journal.pone.0020852.
24. van Heel AJ, de Jong A, Montalban-Lopez M, Kok J, Kuipers OP. BAGEL3: Automated identification of genes encoding bacteriocins and (non-) bactericidal posttranslationally modified peptides. *Nucleic Acids Res*. 2013;41(Web Server issue):W448–53. doi:10.1093/nar/gkt391.
25. Zheng J, Gänzle MG, Lin XB, Ruan L, Sun M. Diversity and dynamics of bacteriocins from human microbiome. *Environmental microbiology*. 2014;n/a-n/a. doi:10.1111/1462-2920.12662.
26. Rea MC, Sit CS, Clayton E, O'Connor PM, Whittall RM, Zheng J, et al. Thuricin CD, a posttranslationally modified bacteriocin with a narrow spectrum of activity against *Clostridium difficile*. *Proc Natl Acad Sci*. 2010;107(20):9352–7.
27. Lakshminarayanan B, Guinane CM, O'Connor PM, Coakley M, Hill C, Stanton C, et al. Isolation and characterization of bacteriocin-producing bacteria from the intestinal microbiota of elderly Irish subjects. *J Appl Microbiol*. 2013;114(3):886–98. doi:10.1111/jam.12085.
28. O'Shea EF, Gardiner GE, O'Connor PM, Mills S, Ross RP, Hill C. Characterization of enterocin- and salivaricin-producing lactic acid bacteria from the mammalian gastrointestinal tract. *FEMS Microbiol Lett*. 2009;291(1):24–34. doi:10.1111/j.1574-6968.2008.01427.x.
29. Cheikhoussef A, Pogori N, Chen H, Tian F, Chen W, Tang J, et al. Antimicrobial activity and partial characterization of bacteriocin-like inhibitory substances (BLIS) produced by *Bifidobacterium infantis* BCRC 14602. *Food Control*. 2009;20(6):553–9. <http://dx.doi.org/10.1016/j.foodcont.2008.08.003>.
30. Mayo B, van Sinderen D. *Bifidobacteria: Genomics and Molecular Aspects*. Caister Academic Press, Norfolk, UK; 2010.
31. Martinez FA, Balciunas EM, Converti A, Cotter PD, de Souza Oliveira RP. Bacteriocin production by *Bifidobacterium* spp. A review. *Biotechnol Adv*. 2013;31(4):482–8. doi:10.1016/j.biotechadv.2013.01.010.
32. Fukuda S, Toh H, Hase K, Oshima K, Nakanishi Y, Yoshimura K, et al. *Bifidobacteria* can protect from enteropathogenic infection through production of acetate. *Nature*. 2011;469(7331):543–7. doi:10.1038/nature09646.
33. Sela DA, Chapman J, Adeyua A, Kim JH, Chen F, Whitehead TR, et al. The genome sequence of *Bifidobacterium longum* subsp. *infantis* reveals adaptations for milk utilization within the infant microbiome. *Proc Natl Acad Sci U S A*. 2008;48:18964–9. doi:10.1073/pnas.0809584105.
34. Reuter G. Designation of Type Strains for *Bifidobacterium* Species. *Int J Syst Bacteriol*. 1971;21(4):273–5. doi:10.1099/00207713-21-4-273.
35. Rodes L, Saha S, Tomaro-Duchesneau C. Microencapsulated *Bifidobacterium longum* subsp. *infantis* ATCC 15697 favorably modulates gut microbiota and reduces circulating endotoxins in F344 rats. 2014;2014:602832. doi:10.1155/2014/602832.
36. Tulini FL, Lohans CT, Bordon KCF, Zheng J, Arantes EC, Vederas JC, et al. Purification and characterization of antimicrobial peptides from fish isolate *Carnobacterium maltaromaticum* C2: Carnobacteriocin X and carnolysins A1 and A2. *Int J Food Microbiol*. 2014;173(0):81–8. <http://dx.doi.org/10.1016/j.jfoodmicro.2013.12.019>.
37. Cotter PD, O'Connor PM, Draper LA, Lawton EM, Deegan LH, Hill C, et al. Posttranslational conversion of l-serines to d-alanines is vital for optimal production and activity of the lantibiotic lactacin 3147. *Proc Natl Acad Sci U S A*. 2005;102(51):18584–9. doi:10.1073/pnas.0509371102.
38. Wu GD, Chen J, Hoffmann C, Bittinger K, Chen Y-Y, Keilbaugh SA, et al. Linking Long-Term Dietary Patterns with Gut Microbial Enterotypes. *Science*. 2011;334(6052):105–8. doi:10.1126/science.1208344.
39. Wexler HM. Bacteroides: the Good, the Bad, and the Nitty-Gritty. *Clin Microbiol Rev*. 2007;20(4):593–621. doi:10.1128/cmr.00008-07.
40. Furet JP, Kong LC, Tap J, Poitou C, Basdevant A, Bouillot JL, et al. Differential adaptation of human gut microbiota to bariatric surgery-induced weight loss: links with metabolic and low-grade inflammation markers. *Diabetes*. 2010;59(12):3049–57. doi:10.2337/db10-0253.
41. Nadal I, Santacruz A, Marcos A, Warnberg J, Garagorri JM, Moreno LA, et al. Shifts in clostridia, bacteroides and immunoglobulin-coating fecal bacteria associated with weight loss in obese adolescents. *Int J Obes (Lond)*. 2009;33(7):758–67. doi:10.1038/ijo.2008.260.
42. Santacruz A, Marcos A, Warnberg J, Marti A, Martin-Matillas M, Campoy C, et al. Interplay between weight loss and gut microbiota composition in overweight adolescents. *Obesity (Silver Spring)*. 2009;17(10):1906–15. doi:10.1038/oby.2009.112.
43. Santacruz A, Collado MC, Garcia-Valdes L, Segura MT, Martin-Lagos JA, Anjos T, et al. Gut microbiota composition is associated with body weight, weight gain and biochemical parameters in pregnant women. *Br J Nutr*. 2010;104(1):83–92. doi:10.1017/s0007114510000176.
44. Avelar K, Pinto L, Antunes L, Lobo L, Bastos M, Domingues R, et al. Production of bacteriocin by *Bacteroides fragilis* and partial characterization. *Lett Appl Microbiol*. 1999;29(4):264–8.
45. Booth S, Johnson J, Wilkins T. Bacteriocin production by strains of *Bacteroides* isolated from human feces and the role of these strains in the bacterial ecology of the colon. *Antimicrob Agents Chemother*. 1977;11(4):718–24.
46. Mossie K, Jones D, Robb F, Woods D. Characterization and mode of action of a bacteriocin produced by a *Bacteroides fragilis* strain. *Antimicrob Agents Chemother*. 1979;16(6):724–30.

47. Nakano V, Ignacio A, Fernandes MR, Fukugaiti MH, Avila-Campos MJ. Intestinal *Bacteroides* and *Parabacteroides* species producing antagonistic substances. *Microbiology*. 2006;1:61–4.
48. Shanks OC, Peed L, Sivaganesan M, Haugland RA, Chern EC. Human Fecal Source Identification with Real-Time Quantitative PCR. *Environmental Microbiology*. Springer; 2014. p. 85–99.
49. Sánchez E, Donat E, Ribes-Koninckx C, Calabuig M, Sanz Y. Intestinal *Bacteroides* species associated with coeliac disease. *J Clin Pathol*. 2010;63(12):1105–11.
50. Bakir MA, Sakamoto M, Kitahara M, Matsumoto M, Benno Y. *Bacteroides dorei* sp. nov., isolated from human faeces. *Int J Syst Evol Microbiol*. 2006;56(Pt 7):1639–43. doi:10.1099/ijs.0.64257-0.
51. Mazmanian SK, Round JL, Kasper DL. A microbial symbiosis factor prevents intestinal inflammatory disease. *Nature*. 2008;453(7195):620–5. doi:10.1038/nature07008.
52. Hsiao Elaine Y, McBride Sara W, Hsien S, Sharon G, Hyde Embriette R, McCue T, et al. Microbiota Modulate Behavioral and Physiological Abnormalities Associated with Neurodevelopmental Disorders. *Cell*. 2013;155(7):1451–63. http://dx.doi.org/10.1016/j.cell.2013.11.024.
53. Haft DH, Basu MK. Biological systems discovery in silico: radical S-adenosylmethionine protein families and their target peptides for posttranslational modification. *J Bacteriol*. 2011;193(11):2745–55. doi:10.1128/jb.00040-11.
54. Iyer LM, Abhiman S, Burroughs AM, Aravind L. Amidoligases with ATP-grasp, glutamine synthetase-like and acetyltransferase-like domains: synthesis of novel metabolites and peptide modifications of proteins. *Mol Biosyst*. 2009;5(12):1636–60. doi:10.1039/b917682a.
55. Havarstein LS, Diep DB, Nes IF. A family of bacteriocin ABC transporters carry out proteolytic processing of their substrates concomitant with export. *Mol Microbiol*. 1995;16(2):229–40.
56. Gauffin Cano P, Santacruz A, Moya Á, Sanz Y. *Bacteroides uniformis* CECT 7771 Ameliorates Metabolic and Immunological Dysfunction in Mice with High-Fat-Diet Induced Obesity. *PLoS One*. 2012;7(7):e41079. doi:10.1371/journal.pone.0041079.
57. Ze X, Duncan SH, Louis P, Flint HJ. *Ruminococcus bromii* is a keystone species for the degradation of resistant starch in the human colon. *ISME J*. 2012;6(8):1535–43. doi:10.1038/ismej.2012.4.
58. Chen J, Stevenson DM, Weimer PJ, Albusin B, a bacteriocin from the ruminal bacterium *Ruminococcus albus* 7 that inhibits growth of *Ruminococcus flavefaciens*. *Appl Environ Microbiol*. 2004;70(5):3167–70.
59. Dabard J, Bridonneau C, Philippe C, Anglade P, Molle D, Nardi M, et al. *Ruminococcin A*, a new lantibiotic produced by a *Ruminococcus gnavus* strain isolated from human feces. *Appl Environ Microbiol*. 2001;67(9):4111–8.
60. Wang HT, Chen IH, Hsu JT. Production and characterization of a bacteriocin from ruminal bacterium *Ruminococcus albus* 7. *Biosci Biotechnol Biochem*. 2012;76(1):34–41. doi:10.1271/bbb.110348.
61. Hsiao A, Ahmed AM, Subramanian S, Griffin NW, Drewry LL, Petri WA, et al. Members of the human gut microbiota involved in recovery from *Vibrio cholerae* infection. *Nature*. 2014. doi:10.1038/nature13738.
62. Garg N, Tang W, Goto Y, Nair SK, van der Donk WA. Lantibiotics from *Geobacillus thermodenitrificans*. *Proc Natl Acad Sci*. 2012;109(14):5241–6. doi:10.1073/pnas.1116815109.
63. Fuchs SW, Jaskolla TW, Bochmann S, Kötter P, Wichelhaus T, Karas M, et al. Entianin, a Novel Subtilin-Like Lantibiotic from *Bacillus subtilis* subsp. *spizizenii* DSM 15029(T) with High Antimicrobial Activity. *Appl Environ Microbiol*. 2011;77(5):1698–707. doi:10.1128/AEM.01962-10.
64. Machiels K, Joossens M, Sabino J, De Preter V, Arijis I, Eeckhaut V, et al. A decrease of the butyrate-producing species *Roseburia hominis* and *Faecalibacterium prausnitzii* defines dysbiosis in patients with ulcerative colitis. *Gut*. 2014;63(8):1275–83. doi:10.1136/gutjnl-2013-304833.
65. Qin J, Li Y, Cai Z, Li S, Zhu J, Zhang F, et al. A metagenome-wide association study of gut microbiota in type 2 diabetes. *Nature*. 2012;490(7418):55–60. doi:10.1038/nature11450.
66. Neyrinck AM, Possemiers S, Verstraete W, De Backer F, Cani PD, Delzenne NM. Dietary modulation of clostridial cluster XIVa gut bacteria (*Roseburia* spp.) by chitin-glucan fiber improves host metabolic alterations induced by high-fat diet in mice. *J Nutr Biochem*. 2012;23(1):51–9. doi:10.1016/j.jnutbio.2010.10.008.
67. Hatzioanou D, Mayer MJ, Duncan SH, Flint HJ, Nisbad A. A representative of the dominant human colonic Firmicutes, *Roseburia faecis* M72/1, forms a novel bacteriocin-like substance. *Anaerobe*. 2013;23:5–8. doi:10.1016/j.anaerobe.2013.07.006.
68. Eustaquio AS, Nam SJ, Penn K, Lechner A, Wilson MC, Fenical W, et al. The discovery of salinosporamide K from the marine bacterium “*Salinispora pacifica*” by genome mining gives insight into pathway evolution. *Chembiochem*. 2011;12(1):61–4. doi:10.1002/cbic.201000564.
69. Velázquez JE, van der Donk WA. Genome mining for ribosomally synthesized natural products. *Curr Opin Chem Biol*. 2011;15(1):11–21. http://dx.doi.org/10.1016/j.cbpa.2010.10.027.
70. Chen D, Feng J, Huang L, Zhang Q, Wu J, Zhu X, et al. Identification and Characterization of a New Erythromycin Biosynthetic Gene Cluster in *Actinopolyspora erythraea* YIM90600, a Novel Erythronolide-Producing Halophilic Actinomycete Isolated from Salt Field. *PLoS One*. 2014;9(9):e108129. doi:10.1371/journal.pone.0108129.
71. Donia Mohamed S, Cimermanic P, Schulze Christopher J, Wieland Brown Laura C, Martin J, Mitreva M et al. A Systematic Analysis of Biosynthetic Gene Clusters in the Human Microbiome Reveals a Common Family of Antibiotics. *Cell*. 158(6):1402–14. doi:10.1016/j.cell.2014.08.032.
72. Klaenhammer TR. Bacteriocins of lactic acid bacteria. *Biochimie*. 1988;70(3):337–49.
73. Riley MA. Molecular mechanisms of bacteriocin evolution. *Annu Rev Genet*. 1998;32(1):255–78.
74. Li X, O'Sullivan DJ. Contribution of the Actinobacteria to the growing diversity of lantibiotics. *Biotechnol Lett*. 2012;34(12):2133–45. doi:10.1007/s10529-012-1024-2.
75. Mendez-Vilas A, Antonio M. Science and Technology Against Microbial Pathogens: Research, Development and Evaluation, Proceedings of the International Conference on Antimicrobial Research (ICAR2010). Valladolid, Spain: World Scientific Publishing Company Pte Limited; 2011.
76. Goto Y, Li B, Claesen J, Shi Y, Bibb MJ, van der Donk WA. Discovery of Unique Lanthionine Synthetases Reveals New Mechanistic and Evolutionary Insights. *PLoS Biol*. 2010;8(3):e1000339. doi:10.1371/journal.pbio.1000339.
77. Rutherford K, Parkhill J, Crook J, Horsnell T, Rice P, Rajandream MA, et al. Artemis: sequence visualization and annotation. *Bioinformatics*. 2000;16(10):944–5.
78. Carver TJ, Rutherford KM, Berriman M, Rajandream M-A, Barrell BG, Parkhill J. ACT: the Artemis Comparison Tool. *Bioinformatics*. 2005;21(16):3422–3. doi:10.1093/bioinformatics/bti553.

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

