Genome mining for antimicrobial compounds in wild marine animals-associated enterococci

Janira Prichula¹, Muriel Primon-Barros¹, Romeu C. Z. Luz¹, Ícaro M. S. Castro¹, Thiago G. S. Paim¹, Mauricio Tavares², Rodrigo Ligabue-Braun³, Pedro A. d’Azevedo¹, Jeverson Frazzon¹, Ana P. G. Frazzon¹, Adriana Seixas³ and Michael S. Gilmore⁶,⁷.*

¹ Gram-positive Cocci Laboratory, Federal University of Health Sciences of Porto Alegre (UFCSPA), Porto Alegre, Rio Grande do Sul (RS), Brazil.
² Centro de Estudos Costeiros, Limnológicos e Marinhos (CECLIMAR), Campus Litoral Norte, Federal University of Rio Grande do Sul (UFRGS), Imbé, RS, Brazil.
³ Department of Pharmacosciences, UFCSPA, Porto Alegre, RS, Brazil.
⁴ Food Science Institute, UFRGS. Porto Alegre, RS, Brazil.
⁵ Department of Microbiology, Immunology and Parasitology, UFRGS. Porto Alegre, RS, Brazil.
⁶ Department of Ophthalmology, Harvard Medical School, Massachusetts Eye and Ear Infirmary, Boston, Massachusetts, USA.
⁷ Department of Microbiology, Harvard Medical School, Boston, Massachusetts, USA.

*Correspondence: michael_gilmore@meei.harvard.edu.

Abstract: New ecosystems are being actively mined for new bioactive compounds. Because of the large amount of unexplored biodiversity, bacteria from marine environments are especially promising. Further, host-associated microbes are of special interest because of their low toxicity and compatibility with host health. Here we identified and characterized biosynthetic gene clusters encoding antimicrobial compounds in host-associated enterococcal recovered from fecal samples of wild marine animals remote from human-affected ecosystems. Putative biosynthetic gene clusters in the genomes of 22 Enterococcus strains of marine origin were predicted using antiSMASH5 and Bagel4 bioinformatic software. At least one gene cluster encoding a putative bioactive compound precursor was identified in each genome. Collectively, 73 putative antimicrobial compounds were identified, including 61 bacteriocins (83.56 %), 10 terpenes (13.70 %), and two (2.74 %) related to putative nonribosomal peptides (NRPs). Two of the species studied, Enterococcus avium and Enterococcus mundtii, are rare causes of human disease and were found to lack any known pathogenic determinants but yet possessed bacteriocin biosynthetic genes, suggesting possible additional utility as probiotics. Wild marine animal-associated enterococci from human-remote ecosystems provide a potentially rich source for new antimicrobial compounds of therapeutic and industrial value, and potential probiotic application.

Keywords: Enterococci; Genomic Screening; Bacteriocins; Probiotics, Wild Marine Species

Introduction

Drug-resistant bacteria kill an estimated 700,000 people worldwide each year, and the discovery of new antimicrobial drugs is urgently needed¹–³. This is motivating the search of new ecologies for novel natural products of potential therapeutic value. Human-proximal terrestrial life has been screened for diverse natural products to a much greater extent than larger but less accessible marine ecosystems. Blue biotechnology (or marine biotechnology) is an emerging field that investigates the rich diversity of bioactive molecules produced by marine organisms with potential industrial and
therapeutic applications. Some already have been derived from sponges, cyanobacteria, fungi, algae, and bacteria.

Recently, host-associated microbes also have drawn attention as a potential source for low toxicity agents compatible with host health, but active against pathogenic microbes. It was therefore of interest to us, to explore marine animals from remote habitats for host-associated microbes that encode novel natural product biosynthetic pathways. Further, we focused on host-associated enterococci, a genus of gut microbes associated with all classes of land animals studied, and with animals that have returned to the marine environment. Although most enterococci exist as harmless commensals, some lineages of the species Enterococcus faecalis and Enterococcus faecium have emerged as leading causes of multidrug resistant hospital infections.

Enterococci are known to produce bacteriocins with narrow to broad antimicrobial activity. Bacteriocins have found use as natural antimicrobial agents so far mainly in the food industry but could complement traditional antibiotics in controlling important human and animal pathogens.

Different classification schemes have been proposed for bacteriocins produced by Lactic Acid Bacteria (LAB), although still a subject of debate. Class I bacteriocins are posttranslationally modified peptides, with less than 10 kDa, that require enzymatic modification during biosynthesis, thereby the molecules have uncommon amino acids and structures that impact on their properties. Class II bacteriocins are also less than 10 kDa, although they are heat stable and unmodified peptides with the exception of disulfide bridging, circularization, and methionine formylation. This class has been subclassified: IIa) pediocin-like bacteriocins; IIb) two-peptide bacteriocins; II) circular bacteriocins; II) leaderless; and other bacteriocins do not fall into any of the recognized subclasses.

On the other hand, Class III bacteriocins are large-molecular-weight (more than 10 kDa) and heat-labile antimicrobial proteins usually composed of different domains. Divergently, some authors have been classified circular bacteriocins as class IV or as Class Ib, since these head-to-tail cyclized peptides whose N- and C-termini are linked by a peptide bond, thereby rendering a circular molecule.

The bacteriocins synthetized by enterococci, enterocins, are generally small molecular weight (20-60 amino acids), often post-translationally modified peptides with cationic, hydrophobic, and heat-stable properties. They vary in their mode of action, activity spectrum (restricted or broad), molecular mass, biochemical properties, and genetic origin. Most known enterocins are produced by E. faecium and E. faecalis, but a few peptides have also been isolated from Enterococcus mundtii, Enterococcus avium, Enterococcus hirae, and Enterococcus durans. Most characterized
enterocins derive from enterococci associated with food, waste, feces, and gastrointestinal tract of humans and other animals. Few have been described from enterococci from wild ecologies. Traditionally, new bioactive compounds have been identified by screening microorganism extracts for biological activity, or by amplification of new genes using polymerase chain reaction (PCR) These screening strategies are limited by time consuming and laborious test methods. Advances in molecular biology, bioinformatics, and genomics have been providing important new tools for exploration and development. Genome screening has identified a large pool of potential compounds encoded by biosynthetic gene clusters (BGCs) in DNA databases. The identification of new BGCs may be performed by applying algorithms based on indicators (e.g. evolutionary hallmarks, signature protein domains, and distant paralogs of primary metabolic enzymes), and using bioinformatic tools such as antiSMASH and BAGEL. High throughput computational technologies are being used for screening, presumptive chemical elucidation, and understanding of activities and biological aspects of new compounds.

Therefore, genome mining may represent a fertile strategy for identifying new biomolecules for future therapeutic and industrial applications. In this sense, the aim of the present study was to examine 22 genomes of Enterococcus species isolated from fecal samples of 17 wild marine animals from remote ecologies for potential antimicrobial compounds.

**Results**

**Diversity of wild marine animals associated- enterococci**

The genomes of 22 Enterococcus spp. isolated from wild sea turtles, seabirds and marine mammals were sequenced (Table 1). A summary of the sequencing statistics is presented in Supplementary Table 1. The genomes sizes were between 2.6 – 4.5 Mb, with GC contents ranging from 36.3 to 42.4%. All genomes share average nucleotide identities (ANI) above 95% with known species, confirming designation as Enterococcus avium (4.54 %); Enterococcus casseliflavus (13.64 %); Enterococcus faecalis (45.45 %); Enterococcus faecium (4.54 %), Enterococcus hirae (27.27 %), Enterococcus mundtii (4.54 %) species (Table 1; Supplementary Table 2).

**Marine enterococcal genomes harbor diverse biosynthetic gene clusters (BGCs) coding for antimicrobial compounds**

All 22 genomes were analyzed using antiSMASH and Bagel bioinformatic software (Supplementary Table 3). This analysis identified one or more gene clusters encoding a bioactive compound precursor in each genome. In total, seventy-three antimicrobial compound BGCs were predicted, including 61 (83.56 %) bacteriocins, 10 (13.70 %) related to terpene synthesis, and two (2.74...
related to putative nonribosomal peptides (NRPs). The NRPs biosynthetic gene clusters were found only in *E. faecium* genome (MP10-1), whereas terpene BGCs were found among *E. casseliflavus* (HT1-1, J2, J4), *E. hirae* (C7, DMW1-1, MP1-1, MP1-2, MP1-4, MP1-5), and *E. mundtii* (MP7-18) species (Supplementary Table 3). NRPS and terpene BGCs were predicted only by antiSMASH5, whereas bacteriocins were identified by both tools.

**Diversity of bacteriocins genes among wild marine animals-associated enterococci**

A total of 30 unique bacteriocin species were identified, including eight belonging to class I, 19 to class II, and three to class III (Fig. 1). Although class II bacteriocins showed the greatest diversity, class III bacteriocins were most common and widely distributed. Interestingly, eight new putative bacteriocins with no significant identity to known peptides were found amongst marine enterococci genomes, including two new putative lanthipeptides (I and II) identified as class I, five unknown bacteriocins (I, II, III, IV, and V) identified as class II, and one unknown class III bacteriocin (VI) (Fig. 1; Supplementary Table 4).

The most frequent class I bacteriocins were putative sactipeptides (n = 9), followed by unknown lanthipeptide 1 (n = 5), lasso peptides (n = 4), and thiopeptides (n = 4). Enterocin SE-K4 (n = 5) and enterocin P (n = 3) were the most frequent class II bacteriocins. In turn, the class III bacteriocin enterolysin A (n = 17) was the most frequent bacteriocin found in the 22 sequenced genomes (Fig. 1).

Eight enterococcal genomes belonging to *E. hirae* (C7, DMW1-1, MP1-5), *E. avium* (L8), and *E. faecalis* (GT3-2, GT6-1, MP8-1, and ST1-20) species showed four or more bacteriocin biosynthetic genes (Fig. 1). Four of these genomes (C7, DMW1-1, MP1-5, and MP8-1) encode bacteriocins belonging to three different classes (I, II, and III). Moreover, four enterococci genomes (C7, L8, ST1-20, and MP1-2) exhibited evidence of bacteriocin gene duplication (Fig. 1; Supplementary Table 3). Because of their potentially new bacteriocins and/or amenability large-scale synthesis, putative class II and III bacteriocins were of special interest for further analysis.

**Phylogenetic relationship among class II and III bacteriocins predicted from wild marine animal-associated enterococcal genomes**

To gain insights into the phylogeny of the 30 class II and 19 class III bacteriocins genes identified, phylogenetic analysis was performed (Fig. 2), to determine their relationship (Supplementary Table 5) to 16 reference sequences in Bagel4 and Uniprot databases (Supplementary Table 6). This identified two groups with significant branch support (Fig. 2). Group 1 included bacteriocins of both classes II and III. Class II bacteriocin gene clusters in Group 1 could be divided into subclasses a, b, and others. Included within each are: *IIa* mundticin AT06, enterocin P, bacteriocin T8, bacteriocin 3L, and enterocin SE-K4; *Ilb* enterocin X chain alpha, enterocin X chain beta; *II leaderless* enterocin EJ97; *II*...
circular bacteriocin) carnocyclin A; II other subclasses) sakacin Q, enterocin 96, uviB, and enterocin NKR-5-3D; and unknown bacteriocins I, II, III, IV, and V. Class III bacteriocins in Group 1 included: enterolysin A, propionicin SM1, and unknown bacteriocin VI. In contrast, phylogenetic Group 2 included only the class II bacteriocin, lactococcin 972.

Interestingly, the 17 Class III enterolysin A-related sequences occurring in Group 1 could be grouped into three subbranches. The first and second branches included sequences derived from *E. hirae* strains C7, MP1-1, MP1-2, MP1-4, MP1-5, DMW1-1, while the third branch included enterolysins A from *E. faecalis* strains GT3-2, ST1-20, MP8-1, RD1-1, MP5-1, MP9-10, and B9. The alignment of enterolysin A sequences within each branch (Supplementary Figs. 1, 2, and 3) shows high similarity among them, although they have few conserved amino acids compared to the enterolysin A reference sequences (Supplementary Fig. 4).

Alignment of the other bacteriocin sequences with reference sequences was performed (Supplementary Figs. 5, 6, 7, 8, 9, and 10). Among identities found were conserved motifs such as YGN and cysteine residues (all class IIa bacteriocins – Supplementary Fig. 6), and GxxxG or AxxxA motifs among class IIb and circular bacteriocin members (Supplementary Figs. 7 and 8).

New putative bacteriocins I, II, and VI showed greater similarity to carnocyclin A, while the unknown bacteriocins III, IV and V were more closely related to enterocin X chain alpha (Xα) (Fig. 2). Alignment of unknown bacteriocins with carnocyclin A and Enterocin Xα reference sequences allowed detection of conserved amino acid residues and motifs such as GxxxG or AxxxA (Fig. 3).

Putative novel bacteriocins I, II, VI and carnocyclin A showed only 1.3% overall amino acid sequence identity (Fig. 3A), whereas bacteriocins I and II share 55.22% identity between them (Fig. 3B). Putative bacteriocin III, IV, and V, which were closely related to enterocin Xα, have 9.2% overall amino acid sequence identity (Fig. 3C); and bacteriocins III and V share 43.4% identity between them (Fig. 3D). Structural modeling of these putative class II and III bacteriocins using the I-TASSER package to build models using a combination of fragment and ab initio model building is shown in Figure 4. Insights into structural features are important for the biosynthesis, mode of action, and potency of bacteriocins. The molecular models are in agreement with the expected protein folds (mostly alpha-helices with coil regions). Likewise, the most divergent model (Bacteriocin VI) is also isolated in its group in the phylogenetic reconstruction, supporting its uniqueness among other unknown bacteriocins.

*Detection of genes associated with enhanced enterococcal virulence*

Among the 22 genomes evaluated, *E. avium* (L8) and *E. mundtii* (MP7-18) were found to be devoid of determinants that have mainly been identified in *E. faecalis* and *E. faecium* strains associated with enhanced virulence (Fig. 5a and Fig. 5b). All other enterococci strains possessed at least one
potential virulence-associated traits (Fig. 5b). As expected, these were most common in *E. faecium* and *E. faecalis* where they have been most thoroughly studied. Some of these traits are encoded within the core genomes\(^\text{16,17}\). The unique *E. faecium* harbored *efaAfm* and *acm* genes, while all *E. faecalis* contained several genes associated with adhesion (*ace, efaAfs*), biofilm production (*ebpA, ebpB* and *ebpC*), proteases (*gelE* and *srtA*), protection against oxidative stress (*tpx*), and quorum sensing and sex pheromone (*cad, camE, cCF10, cOB1, and *fsrB*). *E. faecalis* genomes varied in the presence of hyaluronidase genes (*hyaA* and *hyaB*) and adhesion-associated gene (*ElrA*).

Resistome analysis (Fig. 5b) revealed that all *E. casseliflavus* genomes (*n* = 3) possessed genes related to low level vancomycin resistance (*vanRC* and *vanXCY*), as expected since these are part of the core genome for that species\(^\text{52}\). All *E. faecalis* genomes (*n* = 10) contained genes within the core genomes\(^\text{16}\) conferring resistance to trimethoprim (*dfrE*); to macrolide, fluoroquinolone, and rifamycin (*efrA* and *efrB*); to pleuromutilin, lincosamide, and streptogramin (*lsaA*); and have a multidrug and toxic compound extrusion transporter (*emeA*). On the other hand, the unique *E. faecium* genome possessed genes related to resistance to aminoglycosides (*aac(6′)-Ii*); to macrolide, lincosamide, streptogramin, tetracycline, oxazolidinone, phenicol, pleuromutilin (*eatAv*); and to macrolide, lincosamide, streptogramin (*msrC*). In addition, *E. hirae* genomes harbored genes related to aminoglycoside (*acc(6′)-Iid*; *n* = 6), and tetracycline [*tet(W/N/M), n* = 2; *tet(L); n* = 1] resistance.

**Discussion**

Microbes associated with marine animals from remote ecologies may be important sources for new tools to manage human and/or microbial interactions. In this study, we explored *Enterococcus* strains from the microbiota of wild sea birds, sea turtles, and marine mammals that range from the Antarctic to the coast of Brazil to identify potentially novel BGCs. These BCGs were found in generalist species *E. faecalis* and *E. faecium*, as well as less common and less studied species, including *E. avium*, *E. casseliflavus*, *E. hirae*, and *E. mundtii*.

Bacteriocin genes were present in all enterococcal strains investigated, highlighting the competitive nature of the gut niche. Bacteriocin-encoding genes are known to be widely disseminated among enterococci species of different origins\(^\text{21,43-45}\). However, likely because of the novel environmental source of these strains, we found considerable diversity and novelty (Fig. 1), with eight genomes possessing four or more bacteriocin gene clusters. This may be driven by variation in wild marine animal diets along migratory routes, combined with selection pressure for factors to control population structure and niche control in the host gut.

Enterococcal bacteriocins are of interest because of their antimicrobial activities, with activity against different Gram-positive and Gram-negative bacteria including species of *Listeria, Clostridium, Staphylococcus, Streptococcus, Cutibacterium, Pseudomonas* and *Salmonella*\(^\text{6,23,24,53}\). Enterocins have also
been described as effective against antibiotic-resistant bacteria such as vancomycin resistant enterococci (VRE), and methicillin-resistant Staphylococcus aureus (MRSA)\textsuperscript{25,36}. Furthermore, antiviral activities have been reported against herpes simplex viruses (HSV-1 and HSV-2), polio virus (PV3), measles virus, and influenza virus\textsuperscript{31,34}. Immunomodulatory and anticancer properties of enterocins have not been widely explored but may also be of potential interest\textsuperscript{55–57}.

In this study, we identified known bacteriocins, natural variants, and potentially new bacteriocins distributed among different enterococcal species. The potency and spectrum of bacteriocins against important pathogens vary according to the peptide subclass\textsuperscript{24,25,54,58}. Class I bacteriocins were identified in our \textit{in-silico} screening, with sactipeptides, new lanthipeptides I, lasso peptides, and thiopeptides being found in high numbers (Fig. 1). Sactipeptides are produced mainly by Gram-positive organisms, and according to previous studies, the sactipeptides from Bacillus \textit{subtilis} (subtilisin A) and Bacillus \textit{thuringiensis} (Thuricin I) have broad and narrow antimicrobial activity spectra, respectively\textsuperscript{24,59}. A previous study also identified sactipeptide BGC in \textit{Enterococcus multii QU25}, similar to one found in this study. Lantibiotics and thiopeptides are most active against Gram-positive pathogens, including MRSA, VRE, and \textit{Clostridium difficile}\textsuperscript{13,24}. In contrast, most lasso peptide show activity against Gram-negative pathogens, e.g. bacteriocin MccJ25, which is active against some strains of \textit{Escherichia coli} and \textit{Salmonella} spp.\textsuperscript{34}.

The present study provides further evidence of significant biodiversity of BGCs for class II, 19 bacteriocin, including five new putative bacteriocins (Figs. 1 and 2; Supplementary Table 4). Class II bacteriocins are of special interest as potential therapeutic agents and has been proposed on a larger scale, whether in the food industry or in human health and veterinary medicine\textsuperscript{60–62}. Because they consist of unmodified peptides (they do not require enzymes for their maturation) and are small structures, less than 10 kDa\textsuperscript{26,61}, that may subject of low-cost production than other classes by chemical synthesis\textsuperscript{61}. Complementing the recombinant technologies, chemical synthesis of bacteriocins may allow further molecular engineering for enhanced potency, improved pharmacological properties, increased stability and modified spectra of activity\textsuperscript{61}. Class II bacteriocins and analogs thereof have been successfully prepared by chemical synthesis, such as aureocin A53 (AucA), durancin A5-11, enterocin CRL35, lactococcin MMFII, leucocin A, pediocin PA-1, curvacin A, lacticin Q (LnqQ), mesentericin Y105, and sakacin P\textsuperscript{60–62}.

In general, the class II bacteriocins are most active against Gram-positive pathogens, specially the class IIa bacteriocins, which are active against \textit{L. monocytogenes} and other Gram-positive pathogens\textsuperscript{23,24,60,63}. Enterocin SE-K4 and enterocin P were the most frequent class II bacteriocins in this study (Fig. 1). Enterocin SE-K4 have been showed antimicrobial activity against Gram-positive bacteria, \textit{B. subtilis}, \textit{Clostridium beijerinckii}, \textit{E. faecium}, \textit{E. faecalis}, and \textit{L. monocytogenes}\textsuperscript{30}. In contrast, enterocin P has a broad antimicrobial spectrum that includes activity against food-borne pathogens,
C. botulinum, C. perfringens, L. monocytogenes and S. aureus, as well as clinical strains, L. monocytogenes, Salmonella (S.) typhi, Salmonella paratyphi C, Shigella dysenteriae, vancomycin-resistant enterococci (VRE), and carbapenem-resistant Pseudomonas aeruginosa.

It is also important to highlight that class III bacteriocins were most common and widely distributed from wild marine animals, and also included the unknown bacteriocin VI (Fig. 1). Furthermore, three different enterolysin A sequences were verified among enterococci species, being two of them from E. hirae genomes that has been reported for the first time in this species. Enterolysin A is a cell wall-degrading bacteriocin first reported to be produced by E. faecalis isolated from fish in Iceland. Despite class III bacteriocins are large proteins (more than 10 kDa) currently complex by chemical approaches, enterolysin A have been reported as broad-spectrum activity against pathogenic and nonpathogenic bacteria since they are able to cleave the peptide bonds within the stem peptide as well as in the interpeptide bridge of Gram-positive bacterial cell walls.

In addition to bacteriocins, a wide variety of novel gene clusters encoding putative terpenes, NRPs, polyketides, and other active compounds have been uncovered by in silico analysis, creating new opportunities for drug development. In this study, we also found terpene biosynthesis-related clusters in E. casseliflavus, E. hirae, and E. mundtii species. Terpenes are secondary metabolites found in plants, bacteria, and fungi and have been shown to act as antibiotics, hormones, flavor or odor constituents, and pigments. Beukers and collaborators also identified putative genes or operons-involved in terpene synthesis in E. hirae, E. villorum, E. gallinarum, E. durans, and E. casseliflavus strains isolated from bovine feces. The role of terpenes in enterococcal biology remains unclear.

Previous studies have examined the probiotic potential of enterococci from the marine environment. Marine probiont strains have been used in finfish aquaculture due to their health beneficial effect and low potential to transfer antibiotic resistance genes to pathogens through horizontal gene transfer. In the present study, genome screening for bacteriocins highlighted potential probiotic enterococcal strains lacking known virulence or resistance traits (Fig. 5a and Fig. 5b). In particular, the E. avium (L8) genome contained gene clusters for bicereucin BsJ1 and BsJ2, enterocin NKR-5-3D, mundticin AT06, and unknown bacteriocin I; and the E. mundtii genome (MP7-18) encoded sacpeptide and mundticin AT06 variants. Members of the genus Enterococcus have not yet obtained the status of generally recognized as safe (GRAS), although some are already being used as probiotics and in the production of animal food additives to prevent diseases or to improve growth. New regulations for probiotics that distinguish between safe and potentially harmful strains are needed. The application of genomic approaches in probiotic research would improve the understanding of the molecular mechanisms that endow the genera with safe and favorable traits.
Host-associated microbes are a rich source of factors that regulate community structure in a manner compatible with host health. In this study, we identified new putative enterococcal bacteriocins as well as many variants of known ones by high throughput sequencing approaches. This data will inform broadens the limits of known structural variation, knowledge of how structure relates to activity, and synthetic biology, promoting future development of bacteriocin-based drugs for potential use in managing animal and human health, and as food preservatives.

Conclusions

Our findings show that there is considerable novelty to be found through exploring the genomes of host-associated microbes from animals in remote ecologies for biosynthetic pathways with the potential to shape host-associated microbial population structures. The novel compounds and natural bacteriocin variants discovered provide first leads for the derivation of new approaches for managing human-microbe interactions in health and disease.

Materials and Methods

Bacterial strains

Twenty-two enterococci strains previously described were evaluated in the present study. Briefly, the collection includes Enterococcus species isolated from fecal samples (cloacal/anal swabs or intestinal content) collected from 17 wild marine animals. These animals, including sea turtles (n = 3), seabirds (n = 8), and marine mammals (n = 6), were found along the North Coast of Rio Grande do Sul, Southern Brazil, from Torres Beach (29°21′32.2′′S; 49°44′10.3′′W) to Dunas Altas Beach, Palmares do Sul (30°23′58.75′′S; 50°17′24.73′′W), between July 2012 and April 2014 (Table 1). The enterococci collection was stored frozen at -20 °C in skim milk supplemented with 20 % glycerol and cultures were routinely grown in brain heart infusion (BHI) at 37 °C for 18 h.

Genomic DNA preparation, high-throughput sequencing, assembly, and annotation

The Enterococcus spp. strains were grown in BHI at 37 °C for 18 h. Genomic DNA was extracted using a commercial kit (QIAGEN DNeasy Blood & Tissue Kit, USA). Manufacturer instructions were followed with minor modification, namely the addition of 50 µL of lysozyme (50 mg/mL) and 10 µL mutanolysin (2500 U/mL, Sigma-Aldrich) for 30 min at 37 °C before the addition of 20 µL proteinase K (20 mg/mL). Extracted DNA was quantified using the Qubit double-stranded DNA (dsDNA) high-sensitivity (HS) assay kit (Life Technologies). Libraries for genome sequencing were prepared using the Nextera XT DNA kit and index primers (Illumina), and reads were generated by HiSeq/MiSeq reagent kit version 2 with 250 cycles on an Illumina HiSeq/Miseq platforms. Reads were subjected to de novo assembly using CLC genomics workbench v8.0.3, and open reading frames (ORFs) were
predicted using NCBI Prokaryotic Annotation Pipeline – PGAP\textsuperscript{31}. The enterococci species assignment was confirmed by pairwise comparison of their average nucleotide identity (ANI) using JSpeciesWS\textsuperscript{32} and the following reference genomes available from GenBank (https://www.ncbi.nlm.nih.gov): Enterococcus avium ATCC 14025; Enterococcus casseliflavus ATCC 12755; Enterococcus faecalis ATCC 19433; Enterococcus faecium Aus0004 (Clade A1); Enterococcus faecium EnGen0007 (Clade A2); Enterococcus faecium Com12 (Clade B); Enterococcus hirae ATCC 9790; Enterococcus mundtii ATCC 882. The GenBank accession number of reference strains are presented in Supplementary Table 2. The novel genome sequences were deposited at DDBJ/ENA/GenBank as whole genome shotgun projects under the accession numbers VODK00000000, VODL00000000, VODM00000000, VODN00000000, VODO00000000, VODQ00000000, VODT00000000, SJAU00000000, SMGV00000000, SMHA00000000, SJAV00000000, SMGY00000000, VODX00000000, VODY00000000, VOEA00000000, VOEB00000000, VOEC00000000, VOED00000000, VOEE00000000, VOEF00000000, VOEG00000000, and VOEH00000000 (Table S1).

**Genome mining for antimicrobial compounds**

Putative biosynthetic gene clusters (BGCs) were predicted using antiSMASH (antibiotics & Secondary Metabolite Analysis Shell 5.0)\textsuperscript{47} and Bagel4 (bacteriocins and RiPP - Ribosomally synthesized and Post-translationally modified Peptides)\textsuperscript{48} using the default parameters. The bacteriocin classification is in accordance with previous proposals for enterococci\textsuperscript{23} and lactic acid bacteria\textsuperscript{26} that accommodate the novel subclasses that are appearing over the last years, based on the biosynthesis mechanism and biological activity.

**Phylogenetic analysis**

Amino acid sequences corresponding to bacteriocin genes (class II and class III) found in this work, along with reference sequences identified by AntiSMASH 5.0\textsuperscript{47} and Bagel4\textsuperscript{48}, and Uniprot databases were aligned using MAFFT\textsuperscript{83}. Guidance\textsuperscript{24} was used to filter unreliable positions and generate a mega alignment encompassing 5 alternative alignments for the sequences. The mega alignment was used to infer the evolutionary history of these proteins by using the Maximum Likelihood method, based on the VT model\textsuperscript{85}. A discrete Gamma distribution was used to model evolutionary rate differences among sites, and the rate variation model allowed for some sites to be evolutionarily invariable\textsuperscript{86}. Significance was assessed via aLRT\textsuperscript{87}. All evolutionary analyses were conducted in PhyML 3.0\textsuperscript{88}. Tree visualization and annotation were performed on Interactive Tree Of Life (iTOL) v4\textsuperscript{89}.

**Molecular modeling**
The structural modeling of unknown bacteriocins (I, II, III, IV, and VI) was performed using the I-TASSER package\textsuperscript{50,51}, since they were not suitable for traditional comparative modeling, requiring a combination of fragment and ab initio model building. UCSF Chimera\textsuperscript{50} was used to visualize and edit the new bacteriocin structural models. Physico-chemical parameters were calculated with ProtParam\textsuperscript{91}.

**Potential virulence markers**

Comprehensive antibiotic resistance database – CARD/RGI-2017\textsuperscript{92} and Resfinder 3.2\textsuperscript{93} were used to identify antimicrobial resistance genes with default parameters and identification threshold of 60 \% identity over a length of 60 \% coverage, respectively. Virulence genes were predicted using VirulenceFinder\textsuperscript{94}, with threshold of 85 \% identity over a length of 60 \%.

**Figures Design**

Figures were designed using D3 (or D3.js, a JavaScript library for visualizing data using web standards)\textsuperscript{95}, R software (R Development Core Team, 2019)\textsuperscript{96}, and Adobe Illustrator.

**Acknowledgments**

The authors would like to thank the Center for Coastal Studies, Limnology and Marine (CECLIMAR) for assistance in collecting samples and to members of the Enteromar Research Group for critical feedback on the manuscript.

**Author Contributions**

J.P., A.S., M.P.B., and M.S.G. designed the study. Samples were collected by M.T. The bacteria isolation, extraction, and genome sequencing were performed by J.P. and T.G.S.P. Bioinformatics approaches, analyze of data, and figures were designed by J.P., M.P.B., R.C.Z.L., I.M.S.C., and R.L.B. The original draft manuscript was writing by J.P., M.P.B., A.S., and M.S.G. This study was funding and/or supervised by A.S., J.F., P.A.A., A.P.G.F., and M.S.G. All authors reviewed, editing, and approved the manuscript.

**Funding**

This research was funded by NIH grant AI083214 – *Harvard-wide program on antibiotic resistance* –, Conselho Nacional de Desenvolvimento Científico e Tecnológico do Brasil and Coordenação e Aperfeiçoamento de Pessoal de Nível Superior (CAPES) of the Brazilian government (CNPq: 40788/2018-4).

**Competing interests**

The authors declare no competing interests.

**Additional information**
Supplementary Information: Table S1: Sequencing statistics, genome sizes, fold coverage, G+C content, of the Enterococcus spp. sequenced. Table S2: Reference genomes used to confirm the enterococci species. Table S3: Putative antimicrobial compounds biosynthesis gene clusters (BGCs) data predicted with antiSMASH5 and Bagel4 software. Table S4: Class I, class II, and class III unknown bacteriocins BGCs data that were not previously identified in antiSMASH5 and Bagel4 databases. Table S5: Class II and class III bacteriocin sequences predicted with antiSMASH5 and Bagel4 software. Table S6: Reference sequences from Bagel4 and Uniprot databases. Figure S1: Alignment of putative enterolysin A (class III) sequences (first branch) from E. hirae genomes using Clustal Omega software. Figure S2: Alignment of putative enterolysin A (class III) sequences (second branch) from E. hirae genomes using Clustal Omega software. Figure S3: Alignment of putative enterolysin A (class III) sequences (third branch) from E. faecalis genomes using Clustal Omega software. Figure S4: Alignment of four different enterolysin A (class III) and three different references (Bagel 62.3: E. faecalis LMG 2333; Bagel 63.3: E. faecalis; and Bagel: 64.3 - Lactobacillus acidophilus) using Clustal Omega software. Figure S5: Alignment of putative propionicin SM1 (class III) and reference sequence using Clustal Omega software. Figure S6: Alignment of putative Class Ia bacteriocins and reference sequences using Clustal Omega software. Figure S7: Alignment of putative class Ib bacteriocins and reference sequences using Clustal Omega software. Figure S8: Alignment of putative class II circular bacteriocin carnocyclin A and reference sequence using Clustal Omega software. Figure S9: Alignment of putative class II leaderless bacteriocin enterocin EJ97 and reference sequence using Clustal Omega software. Figure S10: Alignment of putative class II other bacteriocins and reference sequences using Clustal Omega software.

References
1. Wohlleben, W., Mast, Y., Stegmann, E. & Ziemert, N. Antibiotic drug discovery. Microb. Biotechnol. 9, 541–548 (2016).
2. Willyard C. The drug-resistant bacteria that pose the greatest health threats. Nature, 543, 15 (2017).
3. Tacconelli, E. et al. Discovery, research, and development of new antibiotics: the WHO priority list of antibiotic-resistant bacteria and tuberculosis. Lancet Infect. Dis. 18, 318–327 (2018).
4. Buonocore, F. Marine biotechnology: developments and perspectives. J. Aquac. Res. Dev. 4, 9546 (2013).
5. Reen, F.J., Gutiérrez-Barranquero, J.A., Dobson, A.D.W., Adams, C., O’Gara, F. Emerging concepts promising new horizons for marine biodiscovery and synthetic biology. Mar. Drugs 13, 2924–2954 (2015).
6. Calle, F. Marine microbiome as source of natural products. Microb. Biotechnol. 10, 1293–1296 (2017).
7. Smith, D. et al. Discovery pipelines for marine resources: an ocean of opportunity for biotechnology? World J Microb. Biotechnol 35, 1–8 (2019).
8. Desriac, F. et al. Bacteriocin as weapons in the marine animal-associated bacteria warfare: inventory and potential applications as an aquaculture probiotic. Mar. Drugs 8, 1153–1177 (2010).
9. Debbab, A., Aly, A. H., Lin, W. H. & Proksch, P. Bioactive compounds from marine bacteria and fungi. Microb. Biotech. 3, 544–563 (2010).
10. Barone, R. et al. Marine metagenomics, a valuable tool for enzymes and bioactive compounds discovery. Front. Mar. Sci. 1, 1–6 (2014).

11. Romano, G. et al. Marine microorganisms as a promising and sustainable source of bioactive molecules. Mar. Environ. Res. 128, 58–69 (2017).

12. Rahman, M. M. et al. Whole-genome sequence of Bacillus subtilis WS1A, a promising fish probiotic strain isolated from marine sponge of the Bay of Bengal. Microbiol Resour Announc 9, e00641-20 (2020).

13. Donia, M. S. et al. A Systematic analysis of biosynthetic gene clusters in the human microbiome reveals a common family of antibiotics. Cell 158, 1402–1414 (2014).

14. Wang, L., Ravichandran, V., Yin, Y., Yin, J. & Zhang, Y. Natural Products from Mammalian Gut Microbiota. Trends Biotechnol 37, 492–504 (2019).

15. Lebreton, F. et al. Tracing the Enterococci from Paleozoic Origins to the Hospital. Cell 169, 849-861.e13 (2017).

16. Prichula, J. et al. Enterococci from wild Magellanic penguins (Spheniscus magellanicus) as an indicator of marine ecosystem health and human impact. Appl Environ Microbiol 86, e01662-20 (2020).

17. Lebreton, F. et al. Emergence of epidemic multidrug-resistant Enterococcus faecium from animal and commensal strains. mBio 4, e00534-13 (2013).

18. Raven, K. E. et al. A decade of genomic history for healthcare-associated Enterococcus faecium in the United Kingdom and Ireland. Genome Res. 26, 1388–1396 (2016).

19. Raven, K. E. et al. Genome-based characterization of hospital-adapted Enterococcus faecalis lineages. Nat Microbiol 1, 1–7 (2016).

20. Fiore, E., Van Tyne, D. & Gilmore, M. S. Pathogenicity of Enterococci. Microbiol Spectr 7, 1–23 (2019).

21. Brock, T. D., Peacher, B. & Pierson, D. Survey of the bacteriocines of enterococci. J. Bacteriol. 86, 702–707 (1963).

22. Nes, I. F., Diep, D. B. & Holo, H. Bacteriocin diversity in Streptococcus and Enterococcus. J. Bacteriol. 189, 1189–1198 (2007).

23. Ness, I. F., Diep, D. B. & Ike, Y. Enterococcal bacteriocins and antimicrobial proteins that contribute to niche control. In Enterococci: From commensals to leading causes of drug resistant infection (eds. Gilmore M. S. et al.) 1-33 (Massachusetts Eye and Ear Infirmary, Boston, 2014).

24. Cotter, P. D., Ross, R. P. & Hill, C. Bacteriocins — a viable alternative to antibiotics? Nat Rev Microbiol 11, 95–105 (2013).

25. Hanchi, H., Mottawea, W., Sebei, K. & Hammami, R. The genus Enterococcus: between probiotic potential and safety concerns—An Update. Front. Microbiol. 9, 1–16 (2018).

26. Alvarez-Sieiro, P., Montalbán-López, M., Mu, D. & Kuipers, O. P. Bacteriocins of lactic acid bacteria: extending the family. Appl Microbiol Biotechnol 100, 2939–2951 (2016).

27. Zimina, M. et al. Overview of global trends in classification, methods of preparation and application of bacteriocins. Antibiotics 9, 1–21, (2020).
28. Maqueda, M. et al. Genetic features of circular bacteriocins produced by Gram-positive bacteria. *FEMS Microbiol Rev* 32, 2–22 (2008).

29. Chen, H. & Hoover, D. G. Bacteriocins and their food applications. *Compr. Rev. Food Sci. Food Saf.* 2, 82–100 (2003).

30. Franz, C. M. A. P., Van Belkum, M. J., Holzapfel, W. H., Abriouel, H. & Gálvez, A. Diversity of enterococcal bacteriocins and their grouping in a new classification scheme. *FEMS Microbiol Rev* 31, 293–310 (2007).

31. Todorov, S. D., Wachman, M. B., Knoetze, H., Meincken, M. & Dicks, L. M. T. An antibacterial and antiviral peptide produced by *Enterococcus mundtii* ST4V isolated from soya beans. *Int. J. Antimicrob. Agents* 25, 508–513 (2005).

32. Poeta, P. et al. Detection of antimicrobial activities and bacteriocin structural genes in faecal enterococci of wild animals. *Microbiol. Res.* 162, 257–263 (2007).

33. Comerlato, C.B., Buboltz, J.R., Santos, E. & Frazon, A.P.G. Antimicrobial compounds produced by *Enterococcus* spp. isolates from faecal samples of wild South American fur seals. *J. Microbiol. Antimicrob.* 8, 14–21 (2016).

34. Ghomrassi, H., Hani, K., Chobert, J.-M. & Ghrairi, T. Evaluation of marine bacteriocinogenic enterococci strains with inhibitory activity against fish-pathogenic Gram-negative bacteria. *Dis. Aquat. Organ.* 118, 31–43 (2016).

35. Ogaki, M. B., Rocha, K. R., Terra, M. R., Furlaneto, M. C. & Furlaneto-Maia, L. Screening of the enterocin-encoding genes and antimicrobial activity in *Enterococcus* species. *J. Microbiol. Biotechnol.* 26, 9 (2016).

36. Phumisantiphong, U., Siripanichgon, K., Reamtong, O. & Diraphat, P. A novel bacteriocin from *Enterococcus faecalis* 478 exhibits a potent activity against vancomycin-resistant enterococci. *PLoS ONE* 12, e0186415 (2017).

37. Rahmeh, R. et al. Characterization of semipurified enterocins produced by *Enterococcus faecium* strains isolated from raw camel milk. *J. Dairy Sci.* 101, 4944–4952 (2018).

38. Vandera, E., Tsirkaa, G., Kakouria, A., Koukkoub, A. & Samelisa, J. Approaches for enhancing in situ detection of enterocin genes in thermized milk, and selective isolation of enterocin-producing *Enterococcus faecium* from Baird-Parker agar. *Int. J. Food Microbiol.* 281, 23–31 (2018).

39. Zhao, X.-Q. Genome-based studies of marine microorganisms to maximize the diversity of natural products discovery for medical treatments. *Evid. Based Complement. Alternat. Med.* 2011, 1–12 (2011).

40. Zhang, M.M., Qiao, Y., Ang, E.L. & Zhao, H. Using natural products for drug discovery: the impact of the genomics era. *Expert. Opin. Drug Discov.* 12, 475–487 (2017).

41. Gregory, K., Salvador, L. A., Akbar, S., Adaikpoh, B. I. & Stevens, D. C. Survey of biosynthetic gene clusters from sequenced myxobacteria reveals unexplored biosynthetic potential. *Microorganisms* 7, 1–10 (2019).

42. Sekurova, O. N., Schneider, O. & Zotchev, S. B. Novel bioactive natural products from bacteria via bioprospecting, genome mining and metabolic engineering. *Microb. Biotechnol.* 12, 828–844 (2019).

43. Xin, B. et al. *In Silico* analysis highlights the diversity and novelty of circular bacteriocins in sequenced microbial genomes. *mSystems* 5, e00047-20 (2020).
479 44. Bonacina, J. et al. A genomic view of food-related and probiotic Enterococcus strains. DNA Res. 24, 11–24 (2016).
480 45. Suárez, N.; Bonacina, J.; Hebert, E.M. & Saavedra L. Genome mining and transcriptional analysis of bacteriocin genes in Enterococcus faecium CRL1879. J. Data Mining Genomics Proteomics. 06, 1–8 (2015).
481 46. Vezina, B., Rehm, B. H. A. & Smith, A. T. Bioinformatic prospecting and phylogenetic analysis reveals 94 undescribed circular bacteriocins and key motifs. BMC Microbiol 20, 1–16 (2020).
482 47. Blin, K. et al. antiSMASH 5.0: updates to the secondary metabolite genome mining pipeline. Nucleic Acids Res. 47, W81–W87 (2019).
483 48. van Heel, A. J. et al. BAGEL4: a user-friendly web server to thoroughly mine RiPPs and bacteriocins. Nucleic Acids Res. 46, W278–W281 (2018).
484 49. Richter, M. & Rosselló-Móra, R. Shifting the genomic gold standard for the prokaryotic species definition. PNAS 106, 19126–19131 (2009).
485 50. Yang, J. et al. The I-TASSER Suite: protein structure and function prediction. Nat. Methods 12, 7–8 (2015).
486 51. Zhang, Y. I-TASSER server for protein 3D structure prediction. BMC Bioinf. 9, 1–8 (2008).
487 52. Palmer, K. L. et al. Comparative genomics of enterococci: variation in Enterococcus faecalis, clade structure in E. faecium, and defining characteristics of E. gallinarum and E. casseliflavus. mBio 3, e00318-11 (2012).
488 53. Jennes, W., Dicks, L. M. T. & Verwoerd, D. J. Enterocin 012, a bacteriocin produced by Enterococcus gallinarum isolated from the intestinal tract of ostrich. J. Appl. Microbiol. Biochem. 88, 349–357 (2000).
489 54. Ermolenko, E. I. et al. Anti-influenza activity of enterocin B in vitro and protective effect of bacteriocinogenic enterococcal probiotic strain on influenza infection in mouse model. Probiotics Antimicrob. Proteins 11, 705–712 (2019).
490 55. Al-Fakharany, O. M., A. Abdel Aziz, A., El-Banna, T. E.-S. & Ibrahim Sonbol, F. Immunomodulatory and anticancer activities of enterocin Oe-342 produced by Enterococcus faecalis isolated from stool. J. Clin. Cell Immunol. 09, 1000558 (2018).
491 56. Ankaiah, D. et al. Cloning, overexpression, purification of bacteriocin enterocin-B and structural analysis, interaction determination of enterocin-A, B against pathogenic bacteria and human cancer cells. Int. J. Biol. Macromol. 116, 502–512(2018).
492 57. Pirkhezranian, Z., Tanhaeian, A., Mirzaei, M. & Sekhavati, M. H. Expression of Enterocin-P in HEK platform: evaluation of its cytotoxic effects on cancer cell lines and its potency to interact with cell-surface glycosaminoglycan by molecular modeling. Int J Pept Res Ther 26, 1503–1512 (2020).
493 58. Caly, D. L. et al. The safe enterocin DD14 is a leaderless two-peptide bacteriocin with anti-Clostridium perfringens activity. J. Antimicrob. Agents 49, 282–289 (2017).
494 59. Mathur, H., Rea, M., Cotter, P., Hill, C. & Ross, R. The sactibiotic subclass of bacteriocins: an update. CPPS 16, 549–558 (2015).
495 60. Bédard, F. et al. Synthesis, antimicrobial activity and conformational analysis of the class IIa bacteriocin pediocin PA-1 and analogs thereof. Sci Rep 8, 1–13 (2018).
61. Bédard, F. & Biron, E. Recent progress in the chemical synthesis of class II and s-glycosylated acteriocins. *Front. Microbiol.* 9, 1–14 (2018).

62. Lander, A. J., Li, X., Jin, Y. & Luk, L. Y. Total chemical synthesis of aureocin A53, lacticin Q and their enantiomeric counterparts. Preprint at https://chemrxiv.org/articles/Total_Chemical_Synthesis_of_Aureocin_A53_Lacticin_Q_and_Their_Enantiomers/12444554/1 (2020).

63. Tanhaeian, A., Damavandi, M. S., Mansury, D. & Ghaznini, K. Expression in eukaryotic cells and purification of synthetic gene encoding enterocin P: a bacteriocin with broad antimicrobial spectrum. *AMB Expr* 9, 1–9 (2019).

64. Cintas, L. M., Casaus, P., Håvarstein, L. S., Hernández, P. E. & Nes, I. F. Biochemical and genetic characterization of enterocin P, a novel sec-dependent bacteriocin from *Enterococcus faecium* P13 with a broad antimicrobial spectrum. *Appl Environ Microbiol* 63, 4321–4330 (1997).

65. Nilsen, T., Nes, I. F. & Holo, H. Enterolysin A, a cell wall-degrading bacteriocin from *Enterococcus faecalis* LMG 2333. *AEM* 69, 2975–2984 (2003).

66. Khan, H., Flint, S. H. & Yu, P.L. Determination of the mode of action of enterolysin A, produced by *Enterococcus faecalis* B9510. *J Appl Microbiol* 115, 484–494 (2013).

67. Tran, P. N., Yen, M.-R., Chiang, C.Y., Lin, H.C. & Chen, P.Y. Detecting and prioritizing biosynthetic gene clusters for bioactive compounds in bacteria and fungi. *Appl Microbiol Biotechnol* 103, 3277–3287 (2019).

68. Brahmkshatriya, P. P. & Brahmkshatriya, P. S. Terpenes: chemistry, biological role, and therapeutic applications. In *Natural Products* (eds. Ramawat, K. G. & Mérillon, J.M.) 2665–2691 (Springer Berlin Heidelberg, 2013).

69. Chen, R., Wong, H. & Burns, B. New approaches to detect biosynthetic gene clusters in the environment. *Medicines* 6, 1-11 (2019).

70. Ziko, L., Adel, M., Malash, M. N. & Siam, R. Insights into red sea brine pool specialized metabolism gene clusters encoding potential metabolites for biotechnological applications and extremophile survival. *Mar. Drugs* 17, 1–25 (2019).

71. Beukers, A. G. *et al.* Comparative genomics of *Enterococcus* spp. isolated from bovine feces. *BMC Microbiol* 17, 1–18 (2017).

72. Román, L. *et al.* The effect of probiotic *Enterococcus gallinarum* L-1 on the innate immune parameters of outstanding species to marine aquaculture. *Journal of Appl.Animal Res.* 43, 177–183 (2015).

73. Safari, R., Adel, M., Lazado, C. C., Caipang, C. M. A. & Dadar, M. Host-derived probiotics *Enterococcus casseli flavus* improves resistance against *Streptococcus iniae* infection in rainbow trout (*Oncorhynchus mykiss*) via immunomodulation. *Fish Shellfish Immunol.* 52, 198–205 (2016).

74. Ringø, E. *et al.* Lactic Acid bacteria in finfish - an Update. *Front. Microbiol.* 9, 1818 (2018).

75. Franz, C.M.A.P.; Hutch, M.; Abriouel, H.; Holzapfel, W. & Gálvez, A. Enterococci as probiotics and their implications in food safety. *Int. J. Food Microbiol.* 151, 125–140 (2011).

76. Huys, G. *et al.* Microbial characterization of probiotics—advisory report of the working group “8651 Probiotics” of the Belgian Superior Health Council (SHC). *Mol. Nutr. Food Res.* 57, 1479–1504 (2013).
77. Joseph, J.P. Probiotic genomes: sequencing and annotation in the past decade. *Int. J. Pharm. Sci. Res.* 9, 1351-1362 (2018).

78. Donia, M. S. & Fischbach, M. A. Small molecules from the human microbiota. *Science* 349, 1254766 (2016).

79. Hashempour-Baltork, F. *et al.* Drug resistance and the prevention strategies in food borne bacteria: an update review. *Adv. Pharm. Bull.* 9, 335-347 (2019).

80. Prichula, J. *et al.* Resistance to antimicrobial agents among enterococci isolated from fecal samples of wild marine species in the southern coast of Brazil. *Mar. Pollut. Bull.* 105, 51–57 (2016).

81. Haft, D. H. *et al.* RefSeq: an update on prokaryotic genome annotation and curation. *Nucleic Acids Res.* 46, D851–D860 (2018).

82. Richter, M., Rosselló-Móra, R., Oliver Glöckner, F. & Peplies, J. JSpeciesWS: A web server for prokaryotic species circumscription based on pairwise genome comparison. *Bioinformatics* 32, 1254766 (2016).

83. Haft, D. H. *et al.* RefSeq: an update on prokaryotic genome annotation and curation. *Nucleic Acids Res.* 46, D851–D860 (2018).

84. Richter, M., Rosselló-Móra, R., Oliver Glöckner, F. & Peplies, J. JSpeciesWS: A web server for prokaryotic species circumscription based on pairwise genome comparison. *Bioinformatics* 32, 929–931 (2016).

85. Müller, T. & Vingron, M. Modeling Amino Acid Replacement. *J. Comput. Biol.* 7, 761–776 (2000).

86. Lefort, V., Longueville, J.E. & Gascuel, O. SMS: smart model selection in PhyML. *Mol. Biol. Evol.* 34, 2422–2424 (2017).

87. Anisimova, M. & Gascuel, O. Approximate likelihood-ratio test for branches: a fast, accurate, and powerful alternative. *Syst. Biol.* 55, 539–552 (2006).

88. Hordijk, W. & Gascuel, O. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst. Biol.* 59, 307–321 (2010).

89. Letunic, I. & Bork, P. Interactive Tree Of Life (iTOL) v4: recent updates and new developments. *Nucleic Acids Res.* 47, W256–W259 (2019).

90. Pettersen, E. F. *et al.* UCSF Chimera-A visualization system for exploratory research and analysis. *J. Comput. Chem.* 25, 1605–1612 (2004).

91. Gasteiger, E. *et al.* Protein identification and analysis tools on the ExPASy server. In *The Proteomics Protocols Handbook* 571–607 (Humana Press, Totowa, 2005).

92. Jia, B. *et al.* CARD 2017: expansion and model-centric curation of the comprehensive antibiotic resistance database. *Nucleic Acids Res.* 45, D566–D573 (2017).

93. Zankari, E. *et al.* Identification of acquired antimicrobial resistance genes. *J. Antimicrob. Chemother.* 67, 2640–2644 (2012).

94. Joensen, K. G. *et al.* Real-time whole-genome sequencing for routine typing, surveillance, and outbreak detection of verotoxigenic *Escherichia coli*. *J. Clin. Microbiol.* 52, 1501–1510 (2014).

95. Bostock, M., Ogievetsky, V. & Heer, J. D* Data-Driven Documents. *IEEE Trans. Visual. Comput. Graphics* 17, 2301–2309 (2011).
96. R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/. (2019).
| Animal         | Common Name       | Scientific Name       | Age | Code | Collection Date | Location   | Enterococci Genomes | Species Identification (ANI) | Collection from |
|----------------|-------------------|-----------------------|-----|------|-----------------|------------|---------------------|-----------------------------|----------------|---------------------------|
| Sea turtles    | green turtle      | Chelonia mydas        | Y   | 2    | 29-May-13       | Cidreira   | GT3-2               | E. faecalis (98.38)           | Prichula et al. (2016) |
|                | green turtle      | Chelonia mydas        | Y   | 1    | 25-Apr-14       | Tramandai  | GT6-1               | E. faecalis (98.53)           | Prichula et al. (2016) |
|                | hawksbill turtle | Eretmochelys imbricata| Y   | 1    | 23-Dec-12       | Tramandai  | HT1-3               | E. casseliflavus (98.56)      | Prichula et al. (2020) |
| Seabirds       | Magellanic penguin| Spheniscus magellanicus| Y   | 1    | 2-Nov-12        | Cidreira   | MP1-1               | E. hirae (98.36)              | Prichula et al. (2020) |
|                | Magellanic penguin| Spheniscus magellanicus| Y   | 2    | 13-Nov-12       | Xangri-lá  | MP2-6               | E. faecalis (98.55)           | Prichula et al. (2020) |
|                | Magellanic penguin| Spheniscus magellanicus| Y   | 1    | 19-Sep-13       | Perfus     | MP5-1               | E. faecalis (98.54)           | Prichula et al. (2020) |
|                | Magellanic penguin| Spheniscus magellanicus| Y   | 1    | 14-Oct-13       | Cidreira   | MP7-18              | E. mundtii (97.04)            | Prichula et al. (2020) |
|                | Magellanic penguin| Spheniscus magellanicus| Y   | 1    | 16-Oct-13       | Cidreira   | MP8-1               | E. faecalis (98.52)           | Prichula et al. (2020) |
|                | Magellanic penguin| Spheniscus magellanicus| Y   | 1    | 23-Dec-13       | Torres     | MP10-1              | E. faecium (97.82)            | Prichula et al. (2020) |
|                | snowy-crowned tern| Spheniscus magellanicus| Y   | 1    | 4-Dec-13        | Xangri-lá  | MP2-1               | E. faecalis (98.55)           | Prichula et al. (2020) |
|                | dwarf minke whale | Spheniscus magellanicus| Y   | 1    | 21-Jan-13       | Xangri-lá  | MP4-1               | E. faecalis (98.54)           | Prichula et al. (2020) |
| Marine Mammals | Risso’s dolphin    | Arctocephalus australis| A   | 2    | 2-Aug-12       | Torres     | MP1-1               | E. faecalis (98.81)           | Santestevan et al. (2015) |
|                | South American fur seal | Arctocephalus australis | A   | 2    | 2-Aug-12       | Xangri-lá  | MP1-1               | E. hirae (98.67)              | Santestevan et al. (2015) |
|                | South American fur seal | Arctocephalus australis | A   | 2    | 12-Jul-12      | Palmares do Sul | MP1-1 | E. casseliflavus (98.56) | Santestevan et al. (2015) |
|                | South American fur seal | Arctocephalus australis | A   | 2    | 21-Jul-12      | Tramandai  | MP1-1               | E. avium (98.06)              | Santestevan et al. (2015) |

1 Age of the animals: A: adult; Y: young.
2 Code based on Geraci and Lounsbury (2005).
3 Strains were sequenced in this study. GT — green turtle; HT — hawksbill turtle; MP — Magellanic penguin; ST — snowy-crowned tern; DMW — dwarf minke whale; RD — risso’s dolphin, and B, C, J or L — South American fur seal.
4 Genomes sequenced in previously study (Prichula et al., 2020).
5 The enterococci species were confirmed by pairwise comparison of their average nucleotide identity (ANI) using as reference the following genomes: Enterococcus avium ATCC14025; Enterococcus casseliflavus ATCC12755; Enterococcus faecalis ATCC19433; Enterococcus faecium Aus0004 (Clade A1), Enterococcus faecium EnGen0007 (Clade A2); Enterococcus faecium Com12 (Clade B); Enterococcus hirae ATCC 9790; Enterococcus mundtii ATCC 882. The E. faecium MP10-1 genome showed average nucleotide identity (ANI) above 95% (97.82) with Enterococcus faecium Com12, confirming it belong to the Clade B.
Figure 1. Biosynthetic bacteriocins genes were found within 22 Enterococcus spp. genomes from wild marine animals. The Enterococcus genomes are represented in the external circle (grey). Diversity of bacteriocin genes within 22 Enterococcus spp. genomes are represented by color gradients: Class I (green gradient) and Class II (blue gradient), and Class III (purple). * Genomes showing duplicated bacteriocin genes (rectangles indicate the number of these genes).
Figure 2. Phylogenetic relationships among bacteriocins (Class II and III) predicted for wild marine animals-associated enterococci genomes. The different groups are represented by grey colors (light grey: Group 1 and dark grey: Group 2). Class II are represented in blue and class III in purple (bold purple are enterolysins A from *E. hirae* and regular purple are enterolysins A from *E. faecalis*). Unknown bacteriocins are highlighted in bold blue (I, II, III, IV, and V) and bold purple (VI).
| GENOME               | BACTERIOCIN               | AA SEQUENCE                                      |
|---------------------|---------------------------|--------------------------------------------------|
| Reference (Bagel: 148.1) | Carnocyclin A            |                                                  |
| E. faecalis GT6-1  | Unknown Bacteriocin VI   | MVPIRKKAGDFMELQSRSKKFCLAMALLIALGMFISAGTSVYAENVNDISEEDKVI |
| E. avium L8        | Unknown Bacteriocin I     | MIKKDV                                           |
| E. hirae DMW1-1    | Unknown Bacteriocin II    |                                                  |
| Consensus          |                           |                                                  |
| Reference (Bagel: 148.1) | Carnocyclin A            |                                                  |
| E. faecalis GT6-1  | Unknown Bacteriocin VI   | LDIVDVPNFSYSDANKGNFSSKAVSAMPINGKLALNIEAKDMFEGIEEYAVSVFIF |
| E. avium L8        | Unknown Bacteriocin I     | LKKVDLKKVIGGGSAGT                                |
| E. hirae DMW1-1    | Unknown Bacteriocin II    | LTAEMKQVGG--RV--                                 |
| Consensus          |                           |                                                  |
| Reference (Bagel: 148.1) | Carnocyclin A            |                                                  |
| E. faecalis GT6-1  | Unknown Bacteriocin VI   | FNSDNNFELGRFRDSKLTLAQGNLRQTRAG@AKCATEDAkNFGV@$AA $ |
| E. avium L8        | Unknown Bacteriocin I     | $NGQAGMLAGSNGCLGIGISIGGIAAGGC $ $               |
| E. hirae DMW1-1    | Unknown Bacteriocin II    | $NGQLGMLTGSVGGIGIIIGGGAIGAGGC$                 |
| Consensus          |                           |                                                  |
| Reference (Bagel: 148.1) | Carnocyclin A            |                                                  |
| E. faecalis GT6-1  | Unknown Bacteriocin VI   | $FIVKAAIAKQGIKAIQL$                             |
| E. avium L8        | Unknown Bacteriocin I     | $FESV--LNTYGAAKANLAKVASTGKAAAVLTLWTAMTCAPIEAE$ |
| E. hirae DMW1-1    | Unknown Bacteriocin II    | $FN------$                                       |
| Consensus          |                           |                                                  |
Figure 3. Alignment of putative unknown Class II bacteriocins and reference sequences using Clustal Omega software. (A) Alignment among I, II, VI, and carnocyclin A (reference) [Identity (*): 1.3 %; Strongly similar (:): 2.2 %; Weakly similar (.): 4.4 %]. (B) Alignment between I and II [Identity (*): 55.22 %; Strongly similar (:): 11.94 %; Weakly similar (.): 10.45 %]. (C) Alignment among III, VI, V, and enterocin Xα (reference) [Identity (*): 9.2 %; Strongly similar (:): 11.8 %; Weakly similar (.): 9.2 %]. (D) Alignment between I and II [Identity (*): 43.4 %; Strongly similar (:): 14.5 %; Weakly similar (.): 11.8 %].

Identical residues are shaded in grey and GxxxG or AxxxxA motives are represented in red color. (-) Gaps introduced to optimize alignments. (*) Positions with a single conserved residue. (:) Conservation between groups with strongly similar properties. (.) Conservation between groups with weakly similar properties.
Figure 4. The structural modeling of unknown Class II enterococcal bacteriocins from wild marine animal.
Figure 5. Wild marine animals-associated enterococci might represent a potentially valuable source of new compounds for biotechnological application and generation of new drug leads, and potential probiotic application. (A) Scheme showing the main marine enterococci applications suggested in this study. (B) Virulence markers analysis revealed potential probiotic enterococci from wild marine animals. Determinants of resistance (light yellow) and virulence (dark yellow) was associated to results of in silico screening by bacteriocins (green, blue, and purple colors). *Genomes showing duplicated bacteriocin genes (rectangles are representing the number of these genes). Blue dash representing the potential probiotic candidates strains (L8 and MP7-18).
