Journal of Oral Microbiology

ISSN: 2000-2297 (Online) Journal homepage: www.tandfonline.com/journals/zjom20

# *Streptococcus dentisani* 7746 encodes a cocktail of 14 bacteriocins associated with Com and Blp-like quorum sensing regulatory systems

**Ainhoa Revilla-Guarinos, Anny Camelo Castillo, Rubén Cebrián, María D. Ferrer, Arantxa López-López, Ana Adrados-Planell, Sandra Lahoz Oliva, Laura Ledesma, Pascal Hols & Álex Mira**

**To cite this article:** Ainhoa Revilla-Guarinos, Anny Camelo Castillo, Rubén Cebrián, María D. Ferrer, Arantxa López-López, Ana Adrados-Planell, Sandra Lahoz Oliva, Laura Ledesma, Pascal Hols & Álex Mira (2026) ***Streptococcus dentisani* 7746 encodes a cocktail of 14 bacteriocins associated with Com and Blp-like quorum sensing regulatory systems**, Journal of Oral Microbiology, 18:1, 2633915, DOI: 10.1080/20002297.2026.2633915

**To link to this article:** https://doi.org/10.1080/20002297.2026.2633915

JOURNAL OF ORAL MICROBIOLOGY
2026, VOL. 18, NO. 1, 2633915
https://doi.org/10.1080/20002297.2026.2633915

Taylor & Francis
Taylor & Francis Group

RESEARCH ARTICLE

OPEN ACCESS Check for updates

# *Streptococcus dentisani* 7746 encodes a cocktail of 14 bacteriocins associated with Com and Blp-like quorum sensing regulatory systems

Ainhoa Revilla-Guarinos[a], Anny Camelo Castillo[a,1], Rubén Cebrián[b,c], María D. Ferrer[a], Arantxa López-López[a,2], Ana Adrados-Planell[a], Sandra Lahoz Oliva[a,3], Laura Ledesma[d], Pascal Hols[d] and Álex Mira[a]

[a]Genomics and Health Department, FISABIO-Public Health, Valencia, Spain; [b]Department of Clinical Microbiology, Instituto de Investigación Biosanitaria de Granada (ibs.Granada), University Hospital San Cecilio, Granada, Spain; [c]Centro de Investigación Biomédica en Red de Enfermedades Infecciosas (CIBERINFEC), Madrid, Spain; [d]Biochemistry and Genetics of Microorganisms (BGM), Louvain Institute of Biomolecular Science and Technology (LIBST), UCLouvain, Ottignies-Louvain-la-Neuve, Belgium

**ABSTRACT**

**Aim:** We explored *in silico* and *in vitro* the complete bacteriocin profile of the oral probiotic *Streptococcus oralis* subsp. *dentisani* strain 7746 with the primary objective of providing a descriptive analysis of bacteriocin genomic organization, regulatory context, and transcriptional expression.

**Methods:** The recently closed genome of 7746 was subjected to genome mining searches for bacteriocin biosynthetic gene clusters with BAGEL4 and antiSMASH. Orthology conservation analyses were performed to distinguish between bacteriocin-like peptides (Blp) and competence (Com) related peptides. We assessed bacteriocins' transcription by non-quantitative cross-gene RT-PCR.

**Results:** Three new bacteriocin-coding genes were identified, which increased to 14 the number of bacteriocins encoded by *S. dentisani* 7746. We proved that all 14 identified bacteriocins are transcriptionally expressed. We have assigned names to bacteriocins with unnamed orthologs in other species, proposing the name Denticins (from Denticin A to Denticin H). Our analysis led us to propose a model for competence and bacteriocin regulation in this strain, ruled by complete sets of Com and Blp-like quorum sensing systems.

**Conclusion:** Our results suggest that *S. dentisani* 7746 is the bacterial isolate with the largest repertoire of bacteriocin genes known to date and that part of its *blp*-like region might have been acquired by horizontal gene transfer from pneumococci.

**ARTICLE HISTORY**
Received 27 October 2025
Revised 16 January 2026
Accepted 13 February 2026

**KEYWORDS**
Antimicrobial peptide (AMP); bacteriocins; oral microbiota; oral health; probiotic bacteria; oral pathogens

## Introduction

Antimicrobials produced by human commensal bacteria and probiotic strains are gaining renewed attention for applications in agriculture, the food industry, human medicine and human oral healthcare [1–4]. In this regard, bacteriocins, which are ribosomally synthesised antimicrobial peptides (AMPs) produced by bacteria [3,5], are of particular interest because low rates of bacterial resistance development against them have been described [6–9]. Bacteriocins may undergo post-translational modifications or remain unmodified, and they are secreted into the extracellular environment [10,11]. Bacteriocins can display either a broad or narrow spectrum of activity, and the producer strain typically encodes immunity proteins to protect itself from their action [11]. Gene mining has been revealed as a powerful approach harnessing genomic information and has emerged as a cornerstone in the discovery of antimicrobial peptides produced by bacteria by a systematic exploration of microbial genomes, revealing cryptic

**CONTACT** Ainhoa Revilla-Guarinos ainhoa.revilla@fisabio.es Genomics and Health Department, FISABIO-Public Health, 46020, Valencia, Spain
[1]Digestive System Unit, Torrecárdenas University Hospital, Almería, Spain and Foundation for Biosanitary Research in Eastern Andalusia - Alejandro Otero (FIBAO).
[2]Servei de Microbiologia i Parasitologia, Hospital can Misses, Ibiza, Spain.
[3]Department of Biomedical Sciences, Faculty of Health Sciences, Universidad CEU Cardenal Herrera, CEU Universities, Valencia, Spain.
Supplemental data for this article can be accessed online at https://doi.org/10.1080/20002297.2026.2633915.

**Key messages**

- *S. dentisani* 7746 has complete sets of Com- and Blp-like quorum sensing systems. The *blp*-like region might have been acquired by horizontal gene transfer from pneumococci.
- *S. dentisani* 7746 encodes 14 bacteriocins, all of which are transcriptionally expressed during the cell cycle.
- We propose the name Denticins A to H for the bacteriocins with unnamed orthologs.

biosynthetic gene clusters responsible for AMP production. The increasing use of genome mining tools such as BAGEL4 or antiSMASH has demonstrated the widespread occurrence of gene clusters encoding bacteriocins [12,13]. In these lines, a previous study by Conrads et al. [14] reported the presence of at least 11 bacteriocin peptides in a bacteriocin-coding region of the probiotic bacterium *Streptococcus oralis* subsp. *dentisani* CECT 7746 (henceforth abbreviated *S. dentisani* 7746 [15]) genome. The species *S. dentisani* is found in the oral microbiota of individuals globally, and its relative abundance has been reported to be significantly higher in dental plaque from caries-free children compared to caries-active ones [16].

*S. dentisani* 7746 exhibits notable efficacy in inhibiting the growth of various oral pathogens, including Gram-positive and Gram-negative species such as *Streptococcus sobrinus*, *Streptococcus mutans*, *Fusobacterium nucleatum,* or *Prevotella intermedia* [17]. This antimicrobial activity is attributed to the extensive repertoire of antimicrobial peptides encoded in its genome. Indeed, the observed antimicrobial mechanism of action aligns with conventional modes associated with antimicrobial peptides, involving membrane pore formation and cellular lysis when exposed to *S. dentisani* 7746 supernatants. Importantly, the proteinogenic nature of this antimicrobial activity was substantiated by its susceptibility to suppression upon the addition of proteases [17]. However, assays to address the expression of the full repertoire of peptidic antimicrobial molecules under laboratory conditions have not yet been performed.

Among the most interesting characteristics of the bacteriocin-coding region described in *S. dentisani* 7746, is that it is linked to quorum sensing and competence-related genes [14]. Quorum sensing dependent cellular functions, like the production of bacteriocins and the induction of genetic competence, are coordinated by peptide-regulated Two-Component Systems (TCS) [18]. In *Streptococcus pneumoniae* strains, the regulation of competence depends on the ComC (pheromone)-ComDE (peptide-regulated TCS) system [19,20]. The *comC* gene codes for ComC, the large precursor peptide of competence stimulating peptide (CSP) with a Gly-Gly motif leader peptide sequence. ComC is exported and processed by the ABC-type transporter ComA and its associated transmembrane transport accessory protein ComB. The leaderless mature form CSP is sensed by the ComD histidine kinase (HK); ComD phosphorylates the response regulator (RR) ComE which in turn, regulates the expression of the competence-involved genes such as *comX*/*sigX* encoding $\sigma^{x}$. This sigma factor rules the entry into the X-state by the activation of the late competence genes, which include the *cib* genes coding for the competence-induced bacteriocins involved in fratricide [21]. The regulation of bacteriocin production by the BlpAB-BlpC-BlpHR system functions in a similar way [22]. The *blpC* gene codes for BlpC, a pheromone precursor with a Gly-Gly motif leader peptide. BlpC is exported and processed by the BlpA transporter and its transport accessory protein BlpB. The mature form of BlpC is the bacteriocin-inducing peptide (BIP). BIP is sensed by the quorum sensing HK BlpH which activates the RR BlpR. The activated BlpR regulates the expression of the bacteriocin and self-immunity genes. Interestingly, in the study by Conrads et al. 2019 [14] with *S. dentisani* 7746, a processing and export system located next to the bacteriocins was named ComAB. A TCS far away in the genome was postulated to be the regulatory system of this bacteriocin region and it was accordingly named ComCDE [14]. Synchronically, Kilian and Tettelin [23] reported the presence of the same bacteriocin cassette with a full regulatory set (composed of a TCS and its cognate pheromone) located upstream of the processing and export system, which they described as a Blp-like bacteriocin cassette [23]. These discrepancies suggested the existence of more than one quorum sensing system associated with bacteriocin coding regions in *S. dentisani* 7746.

The objective of the present study is dual. Firstly, a deep *in silico* characterisation of the bacteriocin biosynthetic gene clusters (BGCs) in the complete genome of *S. dentisani* 7746. For that purpose, we performed genome mining searches for antimicrobial peptides with the web-based bacteriocin genome

mining tools BAGEL4 and antiSMASH. We used the conservation of cassettes, genomic arrangements and context to distinguish between the Blp and Com systems, as previously described [24]. Secondly, we assessed the expression of the bacteriocin genes by RT-PCR analysis. Our analyses increase to 14 the number of bacteriocins encoded by *S. dentisani* 7746 in two different clusters, prove their expression, and lead us to propose a model for competence and bacteriocin regulation in this probiotic strain.

## Materials and methods

### Accession number

The complete genome of *S. oralis* subsp. *dentisani* 7746 is deposited in the European Nucleotide Archive (ENA) database under the assembly title: SDENT7746.1 assembly for *S. oralis* subsp. *dentisani*; accession number: GCA_963669195; Sequence Accession: OY769914; last updated on 14$^{th}$ December 2023.

### Genome mining searching for bacteriocin biosynthetic gene clusters

The online tools BAGEL4 [12] and antiSMASH 6.0 [25] were used to screen the genome of *S. dentisani* 7746 for bacteriocin coding genes. The genomic sequences identified for putative precursor peptides, TCS, putative immunity proteins and post-translation modification enzymes were subsequently manually curated and annotated using the online tools NCBI Blast [26] and Kyoto Encyclopaedia of Genes and Genomes (KEGG) (Kanehisa Laboratories). Orthologs with the highest amino acid identity are indicated. The software Clone Manager and CLC Main Wokbench were used for all subsequent *in silico* molecular studies. Comparative studies of genomic context were performed using MicrobesOnline [27]. Putative stem-loops that could act as terminators were identified by searching for Dyad Symmetries with the Clone Manager programme. Only stem-loops with free energy ranging from −9.3 to −25.9 kcal/mol, for the Blp-like region, and from −9.2 to −11.5 kcal/mol, for the Com-like region, are depicted in Figures 1 and 2.

### Protein sequence-structure analysis

Predictions of the proteins structures and domains were made with SMART [28] and InterPro [29].

### Bacterial strains and growth conditions

*S. dentisani* 7746 was routinely incubated aerobically at 37 °C without agitation. Overnight cultures were prepared in Brain Hearth Infusion Broth (BHI broth 02-599-500, Scharlau Microbiology) and incubated until they reached an optical density at 600 nm ($OD_{600}$) of around 1.8. BHI adjusted to pH 7 with HCl was used for sample collection for gene expression studies. The day cultures (50 ml) were inoculated at $OD_{600}$ = 0.02 with the overnight cultures and they were distributed into two 24 well plates (1.6 ml per well). Plate one was inoculated with two technical replicas of a biological triplicate (6 wells in total), plus the corresponding media blanks, and it was used to monitor the growth rate. For that, plate one was incubated at 37 °C in the Infinite® M Plex TECAN microplate reader (multimode microplate reader, monochromator optics, Product *n*. 3019008, TECAN Trading AG, Switzerland) where $OD_{600}$ was measured every 30 min. Plate two was inoculated with three technical replica of each biological triplicate and it was used for sample collection. For that, it was incubated statically at 37 °C in a bench incubator. Sampling was performed at mid-exponential phase and at stationary phase; one replica was left as backup. At each time point, 1.5 ml from 3 wells of one replica (4.5 ml total volume per sample) were collected in a 15 ml falcon tube, kept on ice, and immediately centrifuged for 5 minutes at 4000 rpm at 4 °C in an Eppendorf 5810 R Centrifuge (Eppendorf AG, Hamburg, Germany). The pellets where pipet-resuspended in RNA later™ solution (RNA stabilisation solution, AM7024, Invitrogen by Thermo Fisher Scientific, TFS Baltics UAB Vilnius, Lithuania) and they were immediately frozen at −20 °C until processing.

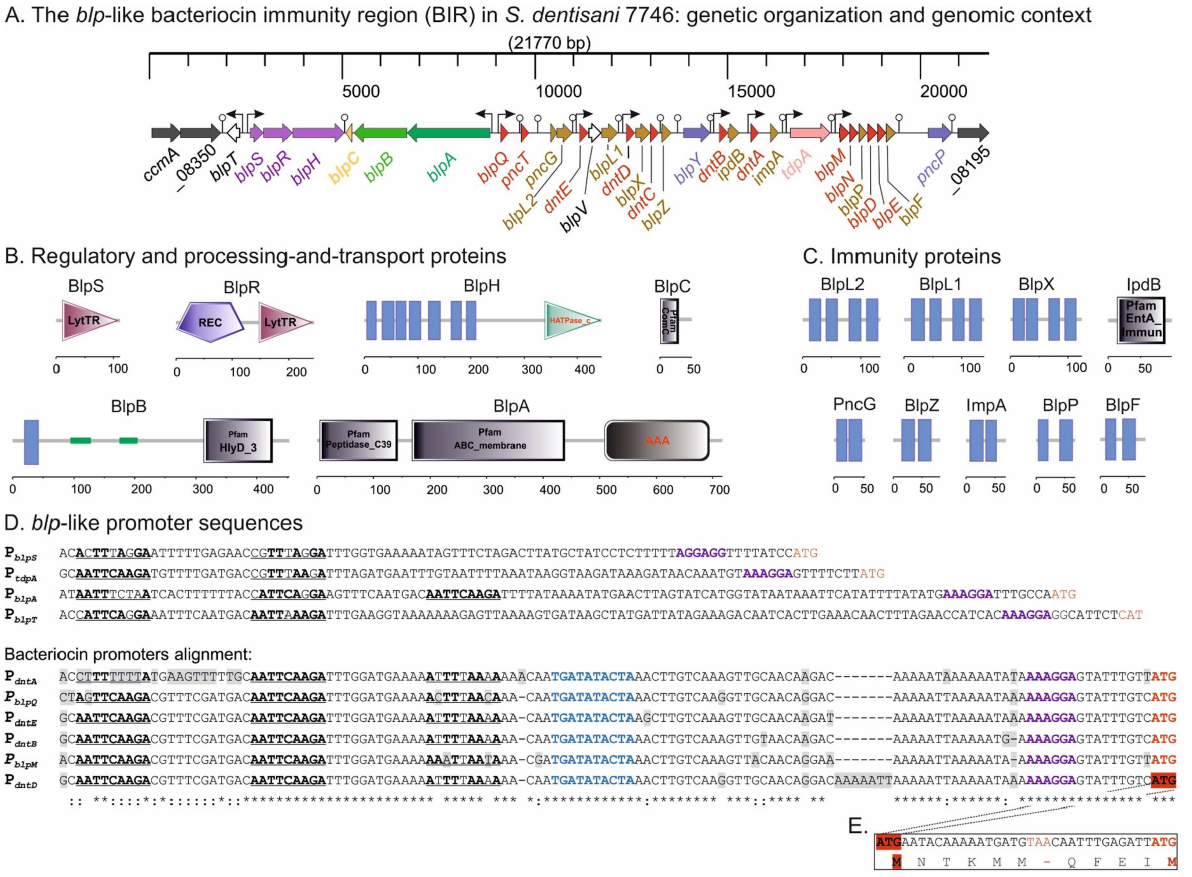


**Figure 1.** The *blp*-like bacteriocin immunity regions (BIR) of *S. dentisani* 7746. In this figure, '*blp* like' refers to a bacteriocin–immunity genomic region homologous to the *blp* locus described in pneumococci. A. Genetic organisation of the *blp*-like region is illustrated with genes drawn to scale. *In silico* identified promoters are denoted by black bent arrows, while stem-loops, potentially serving as terminators with a free energy ranging from –9.3 to –25.9 kcal/mol, are represented by lollipops. Black filled arrows indicate genes within the genomic context of the bacteriocin region, white arrows denote hypothetical proteins, purple arrows signify regulatory proteins (RR and HK), red arrows represent bacteriocins with GG leader peptides, the golden arrow designated the gene *blpC* coding for the quorum sensing system pheromone BlpC, light green arrow indicates transport accessory protein with biotinyl-lipoyl domain, dark green arrow denotes the Sun-T like transporter (featuring peptidase C-39, permease and ATPase predicted domains), brown arrows signify putative bacteriocin immunity proteins, blue arrows represent CAAX-proteases and the pink arrow designates thioredoxin domain-containing protein. B and C. Domain architectures of the regulatory, processing-and-transport proteins (B), and putative immunity proteins (C) as predicted by SMART [28,31]. The colour code for protein domains involved in signalling is as follows: pink triangle, LytTR DNA-binding domain found in a variety of bacterial transcriptional regulators; purple pentagon (REC), cheY-homologous receiver domain which contains a phosphoracceptor site that is phosphorylated by histidine kinase homologues; green triangle (HATPase_c), histidine kinase-like ATPase. Transmembrane regions are depicted as blue vertical rectangles, and coiled coil regions as horizontal green rectangles. AAA stands for ATPase. Description of Pfam domains: ComC, streptococcal competence stimulating peptide precursors; EntA_immun, enterocin A-like immunity protein; HlyD_3, biotinyl-lipoyl domain of unknown function; Peptidase_C39, maturation protease responsible for cleavage of the leader peptide; ABC_membrane, ABC transporter transmembrane region. D. Sequences of the ten putative *blp*-like promoters identified through *in silico* analyses are presented: the less conserved promoter sequences are shown in the upper part, while the alignment of the highly conserved bacteriocin-driving promoters is displayed below. In all promoters, putative ribosome binding sites are indicated in purple bold text and the ATG start codon of the downstream genes is indicated in red. Perfect and imperfect direct repeats of the sequence AATTCAAGA are underlined, with conserved residues indicated in bold letters. Within the alignment: nucleotides conserved in all six promoters are marked with an asterisk (*), nucleotides conserved in five out of six promoters are marked with two dots (:), absent residues are represented by a dash (-), and dissimilar nucleotides are grey shadowed. The extended -10 box [22,32] is highlighted in bold blue text. E. Stop codon mutation in the sequence encoding for the leader peptide of the Denticin D bacteriocin. The original ATG codon is indicated in black letters over a red background, and the actual ATG start codon is indicated in bold red text. Putative coded amino acids are denoted beneath the corresponding nucleotides.

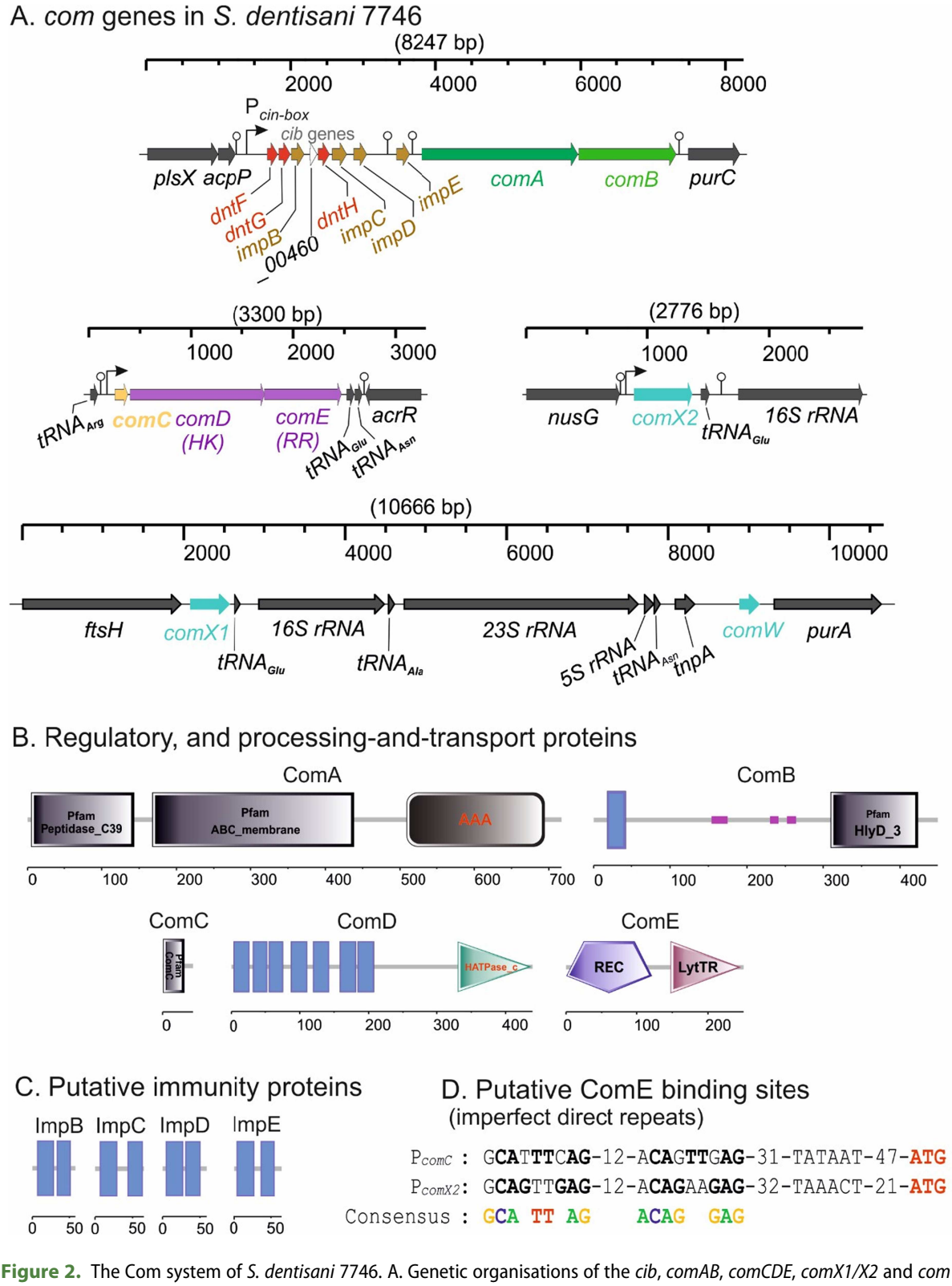


**Figure 2.** The Com system of *S. dentisani* 7746. A. Genetic organisations of the *cib*, *comAB*, *comCDE*, *comX1/X2* and *comW* are depicted with genes drawn to scale. *In silico* identified promoters are denoted by black bent arrows (see also D) and stem-loops potentially serving as putative terminators are represented by lollipops. The colour code is as follows: black filled arrows signify genes in the genomic context, white arrows represent hypothetical proteins, purple arrows denote regulatory proteins (RR and HK), red arrows indicate putative bacteriocins with GG leader peptides, the golden arrow designates the gene *comC* coding for ComC, the precursor peptide of the quorum sensing system pheromone CSP, the light green arrow denotes the transport accessory protein with biotinyl-lipoyl domain, the dark green arrow represents the Sun-T like transporter with peptidase C-39, permease and ATPase predicted domains, brown arrows signify putative

(*Caption on next page*)

bacteriocin immunity proteins, turquoise arrows denote genes for the alternative sigma factors $\sigma^X$, and the gene for the ComW protein (involved in the posttranscriptional control of ComX). B. and C. Domain architectures of the regulatory and processing-and-transport proteins (B), and putative immunity proteins (C) as predicted by SMART [31]. Protein domains are depicted as in Figures 1B and C. D. Sequences of the putative ComE binding sites in the $P_{comC}$ and $P_{comX2}$ promoters, identified by *in silico* analyses based on [33]. The nucleotides constituting the imperfect direct repeats within each promoter sequence are highlighted in bold letters, and the consensus sequence of the imperfect direct repeats is displayed in colour below the promoters. The putative -10 box and the ATG start codon (red) of the downstream genes are indicated.

### Reverse transcription PCR (RT-PCR) to detect polycistronic transcripts

Total nucleic acid samples were isolated using MagNA Pure LC 2.0 Instrument (Roche Molecular Systems, Barcelona, Spain), with the MagNA Pure LC DNA/RNA Isolation Kit III for Bacteria and Fungi (Roche Molecular Systems, GmbH, Mannheim, Germany) following the manufacturer's instructions but an additional enzymatic lysis step with lysozyme 1 mg/ml, lysostaphin 52 units/ml and mutanolysin 26 units/ml (1 hour incubation at 37 °C and 400 rpm) was performed prior to Proteinase K addition. Total nucleic acids were eluted in 100 µl of elution buffer and frozen at −20 °C until further analysis. The obtained DNAs/RNAs were quantified using a Qubit 3 Fluorometer (Invitrogen by Thermo Fisher Scientific) together with the Qubit™ RNA HS assay kit (Invitrogen by Thermo Fisher Scientific, Oregon, USA), and the Qubit™ 1x dsDNA HS assay kit (Invitrogen by Thermo Fisher Scientific, Oregon, USA), following manufacturer's instructions. Each sample with total nucleic acids (DNA/RNA) was split in two aliquots, one to be used as positive control DNA in the PCR, and the other one to be treated with DNAse to obtain clean RNA samples for the RT- PCR.

The RNA samples were diluted with nuclease free water (Invitrogen) when required, and they were subsequently treated to remove genomic DNA using the Turbo DNA-*free*™ kit (Invitrogen by Thermo Fisher Scientific, Baltics UAB Vilnius, Lithuania) following manufacturer's instructions for the rigorous DNAse treatment. Complete removal of genomic DNA from the DNAse treated-RNA samples was verified by absence of amplification when performing a control PCR with the Fusion hot start II high fidelity DNA polymerase (Thermo Fisher Scientific). A second rigorous DNAse treatment was performed to the RNA samples if necessary for complete removal of residual DNA. The quality of the RNA samples was visually assessed in a 1.4% agarose gel stained with the fluorescent nucleic acid dye GelRed® Nucleic Acid Gel Stain, 10.000X, (Biotium, San Francisco, USA).

First-strand cDNA was synthesised using the High-capacity cDNA Reverse Transcription kit from Applied Biosystems (Thermo Fisher Scientific) following manufacturer's instructions. Cross-gene PCRs were performed with the Q5 High Fidelity 2x Master Mix (New England Biolabs) to amplify genomic DNA (positive control), the purified RNA (negative control) and the cDNA (tested sample). The PCR reactions were adjusted so that the concentration of all the templates was 0.125 ng/µl. The PCR products were visualised in a 1.4% agarose gel.

### Primer design for polycistronic transcripts detection

Cross-gene primers were designed with the help of NCBI primer BLAST [30] and the online resource Thermo Fisher Scientific Multiple Primer analyser to rule out the formation of selfdimers and cross primer dimers. For the regions suspected to be co-transcribed, the primers were designed to bind at the 3' end of the upstream gene (forward primer) and the 5' end of the downstream gene (reverse primer), so that the full inter-genic region length was amplified. Primers were designed trying to reduce binding in the highly conserved coding regions for the bacteriocins' leader peptide sequences. The primer sequences are presented in Table 1.

## Results

### The genome of S. dentisani 7746 contains two bacteriocin cassettes with SunT-type transporters

The examination of *S. dentisani* 7746 complete genome utilising BAGEL4 revealed a noteworthy region containing antimicrobial peptides (*blp*-like locus), situated in the fourth quadrant of the genome (Sup.

**Table 1.** Primers used in this study.

| Primer | Target gene | Sequence (5′→3′) |
|---|---|---|
| AM033 | *blpQ* (SDENT7746_08310) | AGAGGGAGGAATATTTGGGGT |
| AM028 | *pncT* (SDENT7746_08305) | GGAGGTGCAGGATTCCATGA |
| AM025 | *dntE* (SDENT7746_08290) | AGTTGAAGGAGGTTTCGGAGG |
| AM066 | *blpL1* (SDENT7746__08280) | TAAGCAAGCATAGAACCGCGCC |
| RG243 | *dntD* (SDENT7746_08275) | GTGGAGATAAAGTAGGGGCTGG |
| RG244 | *dntC* (SDENT7746_08265) | CTGCATACAAAGCGAAATCCCC |
| RG245 | *dntB* (SDENT7746_08250) | CGGTAATGGTTTATATTGTGATAAC |
| RG246 | *ipdB* (SDENT7746_08245) | CAATGTTCATTCGACTAAGG |
| RG247 | *dntA* (SDENT7746_08240) | TGGAATTGGCGTAGGTGCTG |
| RG248 | *impA* (SDENT7746_08235) | TCTCCAGAGTTCCGTTGAAATC |
| RG249 | *blpM* (SDENT7746_08225) | GAGCATGTGGATACATAGGAGC |
| RG250 | *blpF* (SDENT7746_08205) | AGCCATAACGAAAAGGTACAAAGC |
| RG251 | *dntF* (SDENT7746_00450) | CTTATTGGCGGTATTGCTTGGG |
| RG252 | *impD* (SDENT7746_00475) | GTGGCGATGTAGATAATGAGCG |
| RG257 | *dntH* (SDENT7746_00465) | GCTATTGCTGTGGAAGTAGTAAC |
| RG258 | *impB* (SDENT7746_00456) | AATCACCTAGAACGCAGAGTGC |
| RG259 | *dntH* (SDENT7746_00465) | GTTACTACTTCCACAGCAATAG |
| RG260 | *impC* (SDENT7746_00470) | AAATAGCTCGAAGCAAGCAAG |

Fig. S1). This genomic region aligns with the previously reported bacteriocin cassette [14,23]. Subsequent validation through antiSMASH analysis corroborated this finding and identified a previously undocumented second potential region (*com*-like locus), located in the first quadrant of the genome, associated with bacteriocin production (see Sup. Fig. S1). These two loci were subjected to an in-depth investigation.

Bacteriocin biosynthesis loci frequently encompass ABC transporters responsible for processing and exporting newly synthesised peptides, as well as facilitating self-immunity in the producing organism [11]. Examination of the identified areas of interest within the genome of *S. dentisani* 7746 revealed the presence of genes encoding two transporters: locus SDENT7746_00485 in the first quadrant and locus SDENT7746_08315 in the fourth quadrant, both encoding for proteins of 717 amino acids [aa]. Structural predictions conducted through SMART [28,31] for the proteins encoded in these loci indicated the presence of an *N*-terminal peptidase C39 domain, a permease domain and a C-terminal ATPase domain (Figures 1 and 2; Tables 2 and 3). This domain architecture aligns with the SunT-type transporters previously characterised, known for their involvement in the processing and export of antimicrobial peptides or peptide pheromones, concurrently cleaving the leader peptides [11]. Typically, SunT-type transporters are associated with transport accessory proteins featuring one predicted TM-helix and a biotinyl-lipoyl domain of unknown function [11]. Accordingly, two genes (loci SDENT7746_00490 and SDENT7746_08320) encoding proteins embodying this domain architecture were found near the coding genes of the two transporters (Figures 1 and 2; Tables 2 and 3).

SunT-type transporters are usually under the regulation of two-component systems (TCS) involving a peptide quorum sensor histidine kinase (HK) and a LytTR family response regulator (RR) [11,18]. Genomic context analysis, coupled with subsequent structural predictions of protein products, unveiled the presence of a previously undescribed TCS exhibiting these characteristics, adjacent to the SunT-type transporter on the fourth quadrant: locus SDENT7746_08330, encoding for a sensor histidine kinase of 446 aa, and locus SDENT7746_08335, encoding a response regulator of 245 aa with a LytTR DNA binding domain. The domain-based sequence analysis of the HK predicts an *N*-terminal domain with 7 transmembrane helices, and a *C*-terminal histidine kinase-like ATPase domain similar to *Staphylococcus aureus* AgrC and *S. pneumoniae* ComD, both implicated in quorum sensing [29,31]. An additional LytTR binding domain was identified in locus SDENT7746_08340, and a locus encoding a quorum-sensing system pheromone (*blp*-inducing peptide, locus SDENT7746_08325) was positioned between the TCS and ABC transporter genes (Figure 1; Table 2). This genomic arrangement likely signifies a regulatory connection between the TCS, the inducing peptide and the transporter genes, mirroring previously described bacteriocin immunity regions (BIR) in *S. pneumoniae* strains [34]. Amino acid sequence comparisons with allelic variants of the regulatory proteins BlpS, BlpR, BlpH, and BlpC found in pneumococci [34] substantiate a robust conservation of these regulatory proteins in *S. dentisani* 7746 (Sup. File 1). Given these observed similarities, explored and discussed in greater detail subsequently, and for the sake of simplicity, we have renamed genes SDENT7746_08315/08320/08325 to *blpA/B/C*,

**Table 2.** Genes of the *blp* cluster from *S. dentisani* 7746 and their putative functions based on their homology with previously described genes (see also Sup. File 1).

| Locus in *S. dentisani* 7746 | Gene[a] | Predicted function of product |
|---|---|---|
| SDENT7746_08345 | *blpT* | Hypothetical protein associated with the *blp* cluster |
| SDENT7746_08340 | *blpS* | LytTR additional DNA binding domain associated with the TCS |
| SDENT7746_08335 | *blpR* | Response regulator of the TCS with LytTR DNA binding domain |
| SDENT7746_08330 | *blpH* | Sensor histidine kinase of the TCS |
| SDENT7746_08325 | *blpC* | Quorum-sensing system pheromone, *blp*-inducing peptide (BIP) |
| SDENT7746_08320 | *blpB*[b] | Transport accessory protein with a biotinyl-lipoyl domain |
| SDENT7746_08315 | *blpA*[b] | SunT-like ABC transporter with peptidase C-39-, transmembrane permease and ATPase domains |
| SDENT7746_08310 | *blpQ* | Bacteriocin precursor |
| SDENT7746_08305 | *pncT* | Bacteriocin precursor |
| SDENT7746_08300 | *pncG* | Putative membrane-associated immunity protein (group with four transmembrane helices) |
| SDENT7746_08295 | *blpL2* | Membrane-associated immunity protein |
| SDENT7746_08290 | *dntE*[c] | Bacteriocin precursor of Denticin E |
| SDENT7746_08285 | *blpV* | Hypothetical protein associated with the *blp* cluster |
| SDENT7746_08280 | *blpL1* | Putative membrane-associated immunity protein (two copies, group with two transmembrane helices) |
| SDENT7746_08275 | *dntD*[c] | Bacteriocin precursor of Denticin D |
| SDENT7746_08270 | *blpX* | Putative membrane-associated immunity protein (group with two transmembrane helices) |
| SDENT7746_08265 | *dntC*[c] | Bacteriocin precursor of Denticin C |
| SDENT7746_08260 | *blpZ* | Putative membrane-associated immunity protein (group with two transmembrane helices) |
| SDENT7746_08255 | *blpY* | Type II CAAX prenyl endopeptidase Rce1-like[d] |
| SDENT7746_08250 | *dntB*[c] | Bacteriocin precursor of Denticin B |
| SDENT7746_08245 | *ipdB*[c] | Putative immunity protein for Denticin B[e] |
| SDENT7746_08240 | *dntA*[c] | Bacteriocin precursor of Denticin A |
| SDENT7746_08235 | *impA*[b,c] | Membrane-associated immunity protein |
| SDENT7746_08230 | *tdpA*[b] | Thioredoxin domain-containing protein |
| SDENT7746_08225 | *blpM* | Bacteriocin precursor |
| SDENT7746_08220 | *blpN* | Bacteriocin precursor |
| SDENT7746_08216 | *blpP* | Putative membrane-associated immunity protein (group with two transmembrane helices) |
| SDENT7746_08215 | *blpD* | Bacteriocin precursor |
| SDENT7746_08210 | *blpE* | Bacteriocin precursor |
| SDENT7746_08205 | *blpF* | Putative membrane-associated immunity protein (group with two transmembrane helices) |
| SDENT7746_08200 | *pncP* | Type II CAAX prenyl endopeptidase Rce1-like[d] associated with self-immunity. |

[a]The alignments with the ortholog proteins are provided in Sup. File 1.
[b]Assignment based on protein domain conservation.
[c]Names assigned in this study. Example: *dntB* standing for Denticin B coding gene; *ipdB* standing for coding gene for the immunity protein for Denticin B; *impA*, standing for coding gene for a putative membrane immunity protein A.
[d]InterPro family IPR003675.
[e]InterPro family IPR015046, enterocin A immunity-like protein.

**Table 3.** Genes of the *com* cluster from *S. dentisani* 7746 putatively involved in competence development and fratricide, along with their putative functions based on homology with previously described genes [21,41].

| Locus in *S. dentisani* 7746 | Gene | Predicted function of product |
|---|---|---|
| Early competence genes, X-state regulation | | |
| SDENT7746_00485 | *comA*[a] | Competence-factor transporting and processing protein. |
| SDENT7746_00490 | *comB*[a] | Competence-factor transport accessory protein |
| SDENT7746_10175 | *comC* | Competence stimulating peptide (CSP) precursor |
| SDENT7746_10170 | *comD*[a] | Sensor histidine kinase for CSP |
| SDENT7746_10165 | *comE*[a] | Response regulator with a LytTR DNA binding domain |
| SDENT7746_00080 | *comX1*[b] | Alternative sigma factor $\sigma^X$/ComX/SigX involved in competence development |
| SDENT7746_01535 | *comX2*[b] | Alternative sigma factor $\sigma^X$/ComX/SigX involved in competence development |
| SDENT7746_00120 | *comW* | Posttranscriptional control of $\sigma^X$ |
| Bacteriocin and self-immunity genes | | |
| SDENT7746_00450 | *dntF*[c] | Bacteriocin precursor of Denticin F |
| SDENT7746_00455 | *dntG*[c] | Bacteriocin precursor of Denticin G |
| SDENT7746_00456 | *impB*[c] | Putative membrane-associated immunity protein B[d] |
| SDENT7746_00460 | - | Hypothetical protein |
| SDENT7746_00465 | *dntH*[c] | Bacteriocin precursor of Denticin H |
| SDENT7746_00470 | *impC*[c] | Putative membrane-associated immunity protein C[d] |
| SDENT7746_00475 | *impD*[c] | Putative membrane-associated immunity protein D[d] |
| SDENT7746_00480 | *impE*[c] | Putative membrane-associated immunity protein E[d] |

[a]Assignation based on protein domain conservation.
[b]100 % identical ComX paralogs.
[c]Names assigned in this study. Example: *dntF* standing for Denticin F coding gene; *impB*, standing for coding gene for a putative membrane immunity protein B. See main text for details.
[d]Putative function assigned based on structural predictions.

respectively, and the genes SDENT7746_08330/08335/08340 to *blpH/R/S*, respectively. We will employ this nomenclature henceforth (Figure 1; Table 2).

The second region of interest implicated in bacteriocin production was located upstream of the transporter and transport accessory protein in loci SDENT7746_00485/00490, flanked by the *acpP* and *purC* genes (Figure 2A). This genomic arrangement bears a striking resemblance to the *cib* (competence-induced bacteriocin) and *bib* (BIP-induced bacteriocin; BIP denoting bacteriocin-inducing peptide) loci found in *Streptococcus mitis* and *S. pneumoniae*, respectively. Notably, the *cib* and *bib* loci are situated upstream of *comAB*, which encodes ComA, the competence-stimulating-peptide export-and-processing transporter, and ComB, the transport accessory protein [21]. Given the parallels in genomic organisation and functional predictions of the gene products in *S. dentisani* 7746, we have renamed SDENT7746_00485/00490 as *comA/B*, respectively (Figure 2; Table 3). Further elucidation of this bacteriocin region will be discussed in more detail below.

The Blastp suite-2sequences comparisons, assessing amino acid conservation in the transport and accessory proteins of both SunT-type transporters from *S. dentisani* 7746, revealed that BlpA and ComA share a notable identity of 65% (465/713 aa). In contrast, BlpB and ComB exhibit a lower identity, with only 34% similarity (154/451 aa). These identity rates are consistent with those reported for the corresponding proteins in *S. pneumoniae* KNR.7/87, where BlpA-ComA share 65% identity, and BlpB-ComB demonstrate 31% identity [22].

### *The blp-like bacteriocin cassette from S. dentisani 7746*

The bacteriocin cassette located on the fourth quadrant encompasses four regulatory genes: *blpS*, *blpR*, *blpH* and *blpC* (Sup. Fig. S1). The TCS genes, *blpR* (response regulator) and *blpH* (histidine kinase), along with an additional DNA binding domain (*blpS*), are positioned adjacent to the genes coding the SunT-like transporter, *blpA*, and the accessory protein, *blpB*. Furthermore, the *blpC* gene encoding for BlpC, the precursor peptide for a BIP, is situated between the TCS and the transport genes (Figure 1). This genomic organisation closely resembles the BIR responsible for coding bacteriocin-like peptides (*blp*) in *S. pneumoniae* species. Comparative gene content and sequence similarity studies of *blp* prototype cassettes in pneumococci have revealed a broad allelic diversity among the predicted products of the *blp* genes [22,34].

The BIR region in strain 7746 retains at the beginning and ending common genes for *blp* cassettes found in *Pneumococcus* strains, specifically *blpT* and *pncP* [34]. Amino acid sequence alignments with previously described allelic variants confirm the presence of the gene encoding the hypothetical protein BlpT alongside the regulatory, processing-and-transport proteins [34]. At the opposite end of the *S. dentisani* 7746 cassette, a less conserved gene encoding the putative type II CAAX prenyl endopeptidase PncP is located (Figure 1; Table 2; Sup. File 1). Type II CAAX prenyl endopeptidases play a role in bacteriocin self-immunity [35,36].

A total of 23 additional genes/ORFs were identified between *blpA* and *pncP* (Figure 1). To elucidate the protein products of these genes, we conducted protein structure/domain prediction studies using InterPro [29] and performed amino acid sequence comparisons with allelic variants of previously described Blp-like proteins [34] (Sup. File 1). Notably, we identified a second type II CAAX prenyl endopeptidase corresponding to an allelic variant of BlpY (locus SDENT7746_08255) and named it accordingly. Additionally, gene SDENT7746_08285 was found to encode the hypothetical protein BlpV, with the protein in 7746 being 23 aa longer at the N-terminus compared to the alleles described in pneumococci (Sup. File 1). Locus SDENT7746_08230 codes for a conserved thioredoxin domain-containing protein; hence, based on protein domain conservation and genomic context, the gene was named *tdpA* (Table 2). Among the remaining 20 genes, nine encoded putative immunity proteins, and eleven encoded putative *blp*-like bacteriocin structural precursor peptides. These genes exhibited an organisation in putative transcriptional units, with one or two bacteriocin precursor peptides followed by one or two immunity proteins (Figure 1). Further details regarding the *blp*-like bacteriocin structural precursor peptides will be discussed later.

The putative immunity proteins could be classified in three groups according to their predicted secondary structure [28,31] (Figure 1): three proteins with four putative transmembrane helices, five proteins with two

putative transmembrane helices, and one protein containing an enterocin A-like immunity protein domain (Pfam: EntA_Immun, IPR015046 [29,37]). Amino acid sequence comparisons revealed that seven of these proteins were allelic variants of putative membrane-associated immunity proteins already described in *blp* cassettes from pneumococci. Accordingly, these genes were named as follows: *pncG* (SDENT7746_08300), *blpL2* (SDENT7746_08295), *blpL1* (SDENT7746_08280), *blpX* (SDENT7746_08270), *blpZ* (SDENT7746_08260), *blpP* (SDENT7746_08216) and *blpF* (SDENT7746_08205) (Table 2 and Sup. File 1). Notably, the *S. dentisani* 7746 BIR region contains two copies of *blpL,* encoding two proteins with a 90% identity (120/134 aa), as was previously described for three BIR categories found in *S. pneumoniae* strains [34].

### Regulatory elements within the blp-like region

The upstream sequences of putative operons in the BIR region of *S. dentisani* 7746 were analysed for potential promoter sequences. Using conserved BlpR binding sites and extended - 10 sequences from *blp* promoters in the BIR regions of *S. pneumoniae* strains as references [22,32], ten putative *blp*-like promoter sequences were identified: the less conserved $P_{blpS}$, $P_{tdpA}$, $P_{blpA}$ and $P_{blpT}$ promoter regions, and the very well conserved bacteriocin-operon promoters, namely $P_{dntA}$, $P_{blpQ}$, $P_{dntE}$ $P_{dntB}$, $P_{blpM}$ and $P_{dntD}$. The later six promoters showed a high degree of sequence conservation over a sequence length of ±120 bp upstream of the bacteriocins' ATG start codons (Figure 1A and D). Putative ribosome binding sites located −9 to −7 from the ATG translation start codons were identified in all ten cases. Two direct repeats—or less conserved variations— of the nine-base pairs sequence AATTCAAGA spaced by 12 bp, which could serve as putative BlpR binding site [22,32], were identified in all the *blp*-like promoter sequences. A third less conserved reminiscent direct repeat was located 12 bp downstream of the two conserved direct repeats of the bacteriocin promoter regions and upstream of the direct repeat from $P_{blpA}$. A conserved extended -10 region was identified in the promoter regions upstream of the six bacteriocin operons but not in $P_{blpS}$, $P_{tdpA}$, $P_{blpA}$ and $P_{blpT}$ promoters. The genomic sequence of *S. dentisani* 7746 was searched with CLC Main Workbench for additional AATTCAAGA conserved binding sites, but it failed to identify any additional *blp*-like promoters upstream of putative bacteriocin coding genes. Several stem-loops with free energy ranging from −9.3 to −25.9 kcal/mol were identified in this region, which could act as putative terminators (Figure 1A).

### Com-like systems from S. dentisani 7746

The *comAB*-like bacteriocin cassette in the first quadrant of the genome comprises ten ORFs (Sup. Fig. S1 and Figure 2A). The *comA* and *comB* genes are accompanied by eight additional ORFs: four putative bacteriocin immunity proteins (which we named *impB*, *impC*, *impD* and *impE*, loci SDENT7746_00456, SDENT7746_00470, SDENT7746_00475 and SDENT7746_00480, respectively), three structural precursor peptides (loci SDENT7746_00450, SDENT7746_00455, SDENT7746_00465 respectively) and one ORF coding for a hypothetical protein (locus SDENT7746_00460) (Table 3). Four stem-loops with a free energy ranging from −11.5 to −9.2 Kcal/mol were identified downstream of *acpP*, *impD*, *impE*, and downstream of *comB*, which could act as transcriptional terminators (Figure 2A).

The four putative bacteriocin immunity proteins exhibit a predicted secondary structure with two transmembrane helices (Figure 2C). The hypothetical protein encoded by SDENT7746_00460 has 32 amino acids, and it is conserved in *S. mitis* BCC33 (100% identities; NCBI Protein sequence ID RSJ04135.1) and in *Streptococcus infantis* UC6950A (91% aa identities; NCBI Protein sequence ID KJU95414.1). Although SMART [28] predicts one transmembrane helix, InterProScan did not assign it to any protein family [29].

The *comAB* genes, responsible for the ComAB processing and export system, are typically induced during the competence state of streptococci by the ComE response regulator [21]. These genes are part of the ComE response regulator's regulon, along with *comCDE*, *comX1/X2*, and *comW*, which encoded the ComCDE pheromone and cognate sensory system, the competence master regulator $\sigma^X$/ComX/SigX alternative sigma factor, and the competence co-regulator ComW protein, which is involved in the posttranscriptional control of ComX [33,38,39]. A manual inspection of the genomic surroundings of the *comAB*-like bacteriocin area from *S. dentisani* 7746 did not reveal genes coding for regulatory systems.

However, in the fourth quadrant of the genome, an additional AgrC-ComD-like quorum sensing HK of 439 aa (locus SDENT7746_10170), together with a RR of 250 aa with a LytTR DNA-binding domain (locus SDENT7746_10165), and a *comC* gene (locus SDENT7746_10175) coding for the ComC pheromone precursor (InterPro entry IPR004288 [29]) were identified (Sup. Fig. S1 and Figure 2A,B). This regulatory cluster, previously identified by Conrads et al. [14], was found to be 82% identical (358/439 aa) to the HK of the competence regulon ComD from *S. oralis* DD16 (NCBI protein sequence ID: KXT87852.1). The RR showed 96% identity (239/250 aa) to the RR of the competence regulon ComE from *S. mitis* DD22 (NCBI protein sequence ID: KXU11222.1). In the genome of *S. dentisani* 7746, these three genes are located as an operonic structure flanked by the genes encoding transfer RNAs for arginine and glutamate ($tRNA^{arg}$ and $tRNA^{glu}$, respectively) (Figure 2A), consistent with the arrangement observed in *comCDE* operons of other *Streptococcus* strains [20,40]. Considering all this, we named SDENT7746_10175/10170/10165 as *comC/D/E*, respectively, in agreement with the previous nomenclature assigned to these genes by Conrads et al. [14] (Table 3).

We conducted a manual search for *comX* and *comW* early competence genes, taking into account information about the conservation of their genomic contexts in other *Streptococcus* genomes.

The number of *comX* orthologs copies varies between species, with one or two copies being a common occurrence in most strains (reviewed in [41]). In *S. dentisani* 7746, two *comX* paralogs were identified (loci SDENT7746_00080 and SDENT7746_01535). These paralogs code for proteins of 157 amino acids, showing 100% identity to the annotated sigma-70 family RNA polymerase sigma factor from *S. oralis* (strains SOD and FDAARGOS_885; NCBI Protein ID WP_197906202.1). The first paralog, SDENT7746_00080, is located between the genes *ftsH* and the gene for a transfer RNA for glutamate ($tRNA^{glu}$) (Figure 2A). This genomic context resembles the previously described for *comX1* in *S. pneumoniae* [42], and as such, we named locus SDENT7746_00080 as *comX1*. The second paralog, SDENT7746_01535, is positioned between the genes *nusG* and the gene for a transfer RNA for glutamate ($tRNA^{glu}$), flanked by putative stem loops terminators with a free energy of –8.6 and –9.5 Kcal/mol (downstream of *nusG* and downstream of *tRNA-glu*, respectively) (Figure 2A). This genomic context resembles the previously described for *comX2* in *S. pneumoniae* [42], thus we named locus SDENT7746_01535 as *comX2*.

In *S. pneumoniae*, *comW* is typically situated upstream of *purA*, an adenylosuccinate synthetase gene, and it encodes for a small 80 aa protein [43,44]. An orthologous *comW* gene was also identified in *S. dentisani* 7746, specifically at locus SDENT7746_00120. This gene encodes a 78 amino acids protein that shares 100% identity with the sigma(X)-activator ComW from *S. mitis* BCC33 (NCBI Protein ID WP_038804352.1). Therefore, we have designated locus SDENT7746_00120 as *comW* (Table 3).

### Regulatory sequences upstream of the com genes

The upstream sequences of the early *com* genes *comAB*, *comCDE*, *comX1*, *comX2* and *comW* were aligned and searched for potential regulatory elements suggestive of promoter sequences containing ComE-*P* binding sites. Conserved ComE binding sites and extended -10 sequences from streptococci were used as references [33,44,45]. In *S. dentisani* 7746, ComE binding sites were identified only in front of the *comC* and *comX2* genes. These ComE binding sites consist of imperfect direct repeats with a consensus GCAnTTnAG-12n-ACAGnnGAG (Figure 2D).

Regarding the expression of the late-competence-related genes —which include the *cib* genes coding for the competence-induced bacteriocins involved in fratricide— it is induced when $\sigma^X$ associates with the RNA polymerase to target the $\sigma^X$-box DNA-binding motif (also named Cin-box) in promoter regions [21]. A manual search using the Cin-box consensus sequence TACGAATA from *S. pneumoniae* as a reference [46] allowed us to identify a putative Cin-box with the sequence TTCGAATA in the upstream sequence of the *cib* genes from 7746 (Figure 2A). This supports that the expression of these genes is driven by a $\sigma^X$ target promoter, and hence that *dntF*, *dntG* and *dntH* encode late-competence-induced bacteriocins involved in fratricide.

### Orthology analysis of the bacteriocins from S. dentisani 7746

All the bacteriocins from *S. dentisani* 7746 exhibit an *N*-terminal double-glycine type leader peptide (Figure 3A). Leader sequences are typically conserved as they are recognised by processing and export

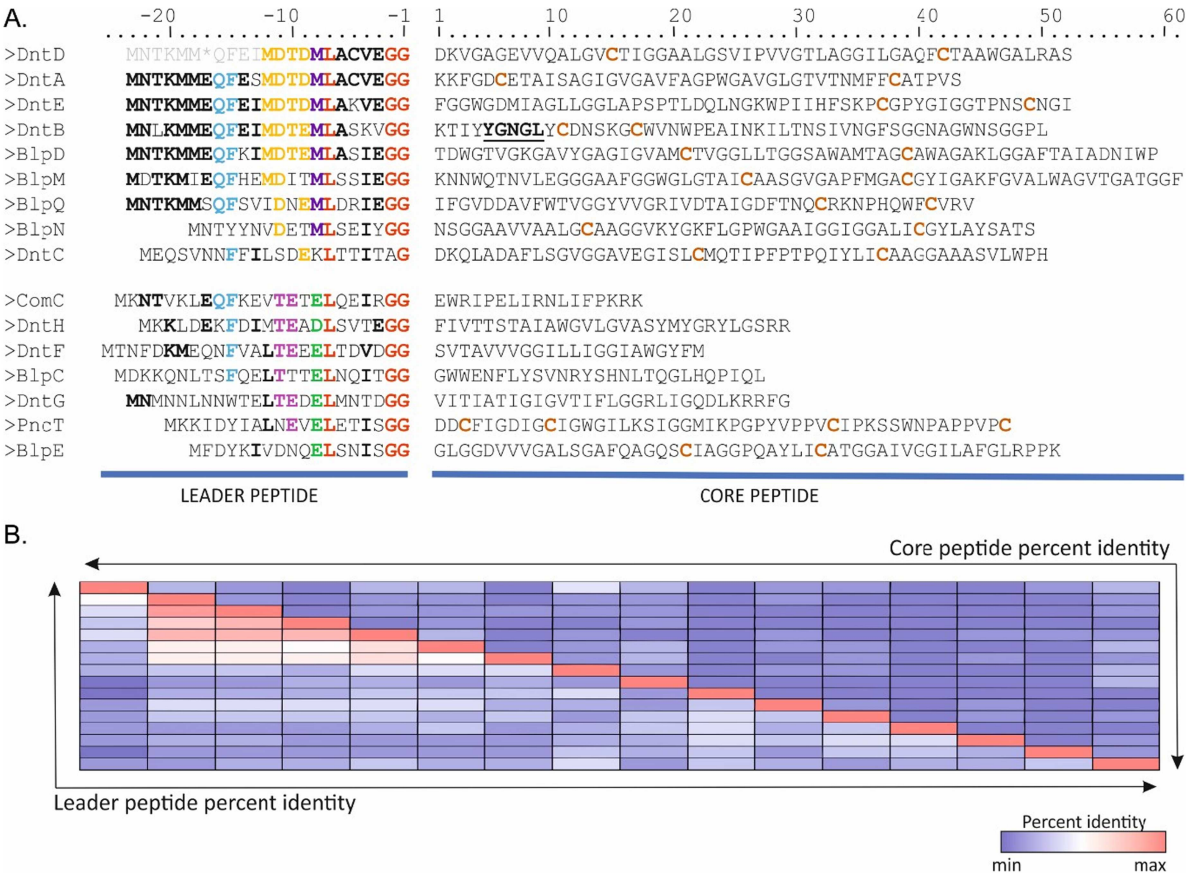


**Figure 3.** Primary sequence analysis of the 14 double-glycine bacteriocins and the two autoinducer molecules (ComC and BlpC) encoded in the *S. dentisani* 7746 genome. A. Primary sequence alignment. The bacteriocins and autoinducers are grouped according to the leader peptide sequence conservation. Conserved amino acids are highlighted in coloured bolded text. In the core peptide sequences: cysteines are indicated in bold brown; the conserved pediocin box of Denticin B is indicated in bold underlined text. The truncated *N*-terminal region of Denticin D is indicated in pale grey text (see Figure 1E). B. Pairwise conservation indicated as percent identity of the leader peptides (bottom left) and the core peptides (top right). The sequences were sorted as in (A). The pairwise alignment was performed with CLC Main Workbench 23.0.4. See Supplementary File. 2 for the complete set of numerical values.

systems [10]. The conservation of full-length precursor peptides, particularly the mature parts of the peptides (defined here as core peptides following nomenclature from [10]), was used to identify orthologs in other species (Table 4).

Six out of the eleven bacteriocins located in the *blp*-like region of *S. dentisani* demonstrate precursor peptide conservation ranging between 87—97% (amino acid (aa) identities), and core peptide conservation between 84—98% (aa identities) compared to multi-strain *S. pneumoniae* Blp bacteriocins (Table 4 and Sup. File 1). These orthologs from *S. pneumoniae* were previously described by Conrads et al. [14]. The high conservation of core peptides suggests that they are allelic variants of pneumococcal bacteriocin-like peptides, potentially acquired through horizontal gene transfer events. These genes were named according to these sequence similarities as follows: *blpQ* (SDENT7746_08310), *pncT* (SDENT7746_08305), *blpM* (SDENT7746_08225), *blpN* (SDENT7746_08220), *blpD* (SDENT7746_08215) and *blpE* (SDENT7746_08210). The predicted bacteriocin products of the three loci SDENT7746_08290, SDENT7746_08275 and SDENT7746_08240 share precursor peptide conservation between 85—96% (aa identities), and core peptide conservation between 93—100% (aa identities) with unnamed orthologs present in *S. mitis* BCA11, in the case of SDENT7746_08240 and SDENT7746_08275, and in different *S. mitis* strains in the case of SDENT7746_08290 (Table 4). In the absence of a gene name, we assigned the following names to these loci: *dntA* (SDENT7746_08240), *dntD* (SDENT7746_08275), and *dntE* (SDENT7746_08290), coding for Denticin A, Denticin D and Denticin E bacteriocins from *S. dentisani*, respectively. The predicted bacteriocin product

**Table 4.** Characteristics of the bacteriocins with double-glycine type leader peptide in *S. oralis* subsp. *dentisani* 7746 and ortholog proteins in other *Streptococcus* strains according to amino acid conservation of the precursor and core peptides.

| Bacteriocin in *S. oralis* subsp. *dentisani* 7746 | [a]aa leader + aa core = aa precursor peptide | Class[b] | Family type[c] | Ortholog (species)[d] | Conservation[e] | |
|---|---|---|---|---|---|---|
| | | | | | precursor peptide (aa identities) | core peptide (aa identities) |
| *Blp like bacteriocin- immunity region* | | | | | | |
| BlpQ | 23 + 44 = 67 | n.p.[a] | n.p. | BlpQ (*S. pneumoniae*) | 88 % (59/67 aa) | 84 % (37/44 aa) |
| PncT | 20 + 47 = 67 | n.p. | n.p. | PncT (*S. pneumoniae*) | 94 % (63/67 aa) | 96 % (45/47 aa) |
| Denticin E | 23 + 52 = 75 | IIb | Lactacin-related | WP_084953064.1 (*S. mitis*)[f] | 96 % (72/75 aa) | 98 % (51/52 aa) |
| Denticin D | 12 + 52 = 64 | IIb | Lactacin-related | WP_125450131.1 (*S. mitis*)[g] | 85 % (64/75 aa)[h] | 100 % (52/52 aa) |
| Denticin C | 22 + 50 = 72 | IIb [i] | Lactacin-related[i] | Mutacin IV *(S. mutans)* | 44 % (32/72 aa) | 48 % (24/50 aa) |
| | | | | WP_061817056.1 (*S. pneumoniae*) | 97 % (70/72 aa) | 100 % (50/50 aa) |
| Denticin B | 23 + 50 = 73 | IIa | Pediocin-like | Ubericin A (*S. uberis*)[j] | 54 % (38/70 aa) | 65 % (32/49 aa) |
| Denticin A | 23 + 43 = 66 | IIb | Lactacin-related | WP_125450135.1 (*S. mitis*)[k] | 95 % (63/66 aa) | 93 % (40/43 aa) |
| BlpM | 23 + 61 = 84 | IIb | Lactacin-related | BlpM (*S. pneumoniae*) | 96 % (81/84 aa) | 98 % (60/61 aa) |
| BlpN | 18 + 49 = 67 | IIb | Lactacin-related | BlpN (*S. pneumoniae*) | 87 % (58/67 aa) | 90 % (44/49 aa) |
| BlpD | 23 + 59 = 82 | IIb | Lactacin-related | BlpD (*S. pneumoniae*) | 90 % (74/82 aa) | 93 % (55/59 aa) |
| BlpE | 18 + 51 = 69 | IIb | Lactacin-related | BlpE (*S. pneumoniae*) | 97 % (67/69 aa) | 98 % (50/51 aa) |
| *Com-like bacteriocin immunity region* | | | | | | |
| Denticin F | 25 + 22 = 47 | IIb | lactobin A/cerein 7B family | WP_038804302.1 (multispecies: *S. mitis* and *S. oralis*) | 100 % (47/47 aa) | 100 % (22/22 aa) |
| Denticin G | 23 + 29 = 52 | IIb | lactobin A/cerein 7B family | WP_001056955.1 (multispecies: *S. mitis* and *S. oralis*) | 96 % (50/52 aa) | 97 % (28/29 aa) |
| Denticin H | 22 + 29 = 51 | n.p. | n.p. | WP_000732133.1 (multispecies: *S. mitis* and *S. oralis*) | 100 % (51/51 aa) | 100 % (29/29 aa) |

[a]Abbreviations: aa, amino acids; n.p., non predicted, generally assigned to Class IId —miscellaneous—[6].

[b]Bacteriocin classes according to InterPro [29]. Class IIb (two-peptide bacteriocins). Class IIa (pediocin-like bacteriocins): characterised by the N-terminal region containing a conserved 'pediocin box' motif (YGNGV/L) and two conserved cysteine residues joined by a disulphide bridge.

[c]Family types according to InterPro [29]: lactacin-related (IPR019493); pediocin-like (IPR002633); lactobin A/cerein 7B family (IPR023991).

[d]Orthologs with the highest amino acid identity are indicated. The orthologs BlpQ, PncT, BlpM, BlpN, BlpD and BlpE are in accordance with Conrads et al. [14]. NCBI Protein Reference Sequence IDs are indicated for unnamed orthologs.

[e]Alignments of the Blp-like bacteriocins with orthologs are included in Sup. File 1.

[f]Denticin E corresponds to ORF50 in Conrads et al. [14]. BlpJ from *S. pneumoniae* is indicated there as an ortholog. Alignments with Denticin E and BlpJ from Bogaardt et al. [34] show a query cover of 26 %, with a conservation of 75 % [18/24 aa identities], mainly covering the leader peptide sequence from both proteins (Sup. File 1). Given the conservation of leader peptide sequences of such bacteriocins, the similarity of BlpJ is not considered indicative of allelic variants.

[g]Denticin D corresponds to ORF45 in Conrads et al. [14]. BlpN/BlpK from *S. pneumoniae* are indicated as orthologs. Aligments with Denticin D and BlpN from Bogaardt et al. [34] indicate a query cover of 77 %, with a conservation of 43 % [23/54 aa identities] for the precursor peptide and a query cover of 77 % with a conservation of 44 % [17/39 aa identities] for the core peptides. In the case of BlpK, our alignments with Denticin D and BlpK from Bogaardt et al. [34] indicate a query cover of 15 % covering the leader peptide sequence from both proteins with a conservation of 92 % [11/12 aa identities]. These similarities are not considered indicative of allelic variants.

[h]Denticin D lacks 11 aa at the N-terminal due to a stop codon mutation accounting for the discrepancy in the length of the precursor peptide.

[i]InterPro did not predict any protein family membership for Denticin C. The classification of mutacin IV according to InterPro is indicated.

[j]Denticin B corresponds to ORF37 in Conrads et al. [14]. BlpJ from *S. pneumoniae* is indicated there as an ortholog. Alignments with the full-length precursor peptides BlpJ (from [34]) and Denticin B indicate a query cover of 25 %, with a conservation of 65 % [15/23 aa identities], mainly covering the leader peptide sequence. Considering the degree of conservation of the leader peptide sequences of this type of bacteriocins, we do not consider the similarity of BlpJ as indicative of allelic variants.

[k]Denticin A corresponds to ORF34 in Conrads et al. [14]. BlpN/BlpK from *S. pneumoniae* are indicated there as orthologs. Manual alignments with the full-length precursor peptides show a conservation of 24 % [16/66 aa identities] for BlpN-Denticin A. In the case of BlpK, alignments indicate a query cover of 38 %, with a conservation of 69 % [20/29 aa identities], mainly covering the leader peptide sequence from both proteins. Comparison of core peptides did not yield significant similarities. We do not consider these similarities as indicative of allelic variants.

from locus SDENT7746_08265 exhibits a 44% conservation with Mutacin IV from *S. mutans* and a 97% conservation with an unnamed hypothetical protein from *S. pneumoniae* (strains B10622 and C14376X). The bacteriocin product from locus SDENT7746_08250 has a 54% and a 65% conservation (aa identities) of the precursor peptide and the core peptide, respectively, with Ubericin A from *Streptococcus uberis* (Table 4). Considering the high conservation (≥84%) of the bacteriocins with allelic variants in pneumococci—namely BlpQ, PncT, BlpM, BlpN, BlpD and BlpE—, we did not consider a similarity lower than 85% as indicative of

allelic variants. Again, in the absence of gene names, we designated these loci *dntB* (SDENT7746_08250) and *dntC* (SDENT7746_08265), coding for Denticin B and Denticin C, bacteriocins from *S. dentisani*, respectively.

The bacteriocins located in the *com*-like region of *S. dentisani* are highly conserved with unnamed bacteriocins in *S. oralis* and *S. mitis* species (Table 4). We named these loci *dntF* (locus SDENT7746_00450), *dntG* (SDENT7746_00455) and *dntH* (SDENT7746_00465) coding for the putative bacteriocin precursor peptides of Denticin F (DntF), Denticin G (DntG) and Denticin H (DntH) from *S. dentisani*, respectively (Tables 3 and 4).

### Pairwise comparisons of 7746's bacteriocins and autoinducers

The primary sequence alignment and pairwise similarity comparisons of all the bacteriocins, along with the two putative autoinducer molecules (ComC and BlpC) are illustrated in Figure 3. With the exception of Denticin C, all 13 bacteriocins and the two autoinducer molecules exhibit a conserved double-glycine cleavage site.

Pairwise comparisons of the full precursor peptide sequences reveal a low degree of conservation, with similarities going up to 38.5% for DntE and DntA, and 37.8% identity for BlpD and DntA (Sup. File 2). When comparing only the leader peptides or the core peptides individually, the former is notably more conserved than the latter (Figure 3B; Sup. File 2). The most similar core peptides are observed in the pairings of BlpN and DntD, exhibiting 39.6% identity, followed by BlpE and BlpM with 26.2% identity. Concerning leader peptides, two distinct groups emerge based on the sequence conservation (Figure 3). Particularly, the leader peptide sequence within the group comprising DntA, DntE, DntB, BlpD, BlpM, and BlpQ is highly conserved. For instance, the leader peptide conservation reaches up to 91.3% identity for DntE and DntA (differing only by two amino acids out of 23, Figure 3A), and 82.6% identity for BlpD and DntE, or BlpD and DntB. The hypervariability of the core peptides, coupled with the highest conservation of the leader sequences is in agreement with the role of the latter in the recognition by the processing enzymes and export systems [10]. It is noteworthy that bacteriocins encoded in the Blp region and those encoded in the Com region exhibit two clear distinguishing features consistent with their different biological contexts. Blp-region bacteriocins are longer (44–61 amino acids) and contain an even number of cysteine residues, compatible with the formation of disulphide bonds. Conversely, bacteriocins from the Com region (namely DntF, DntG and DntH) and the autoinducers (ComC and BlpC) are considerably shorter (17-29 amino acids) and lack any cysteine residues, which readily differentiates them from the Blp-derived peptides.

Interestingly, the alignment of the putative *blp*-like bacteriocin-operons' promoters has revealed a point mutation that could impact the expression of Denticin D. This mutation involves a substitution in a GAA codon for glutamic acid, changing to a TAA stop codon. Consequently, this mutation is anticipated to result in the premature termination of the precursor peptide sequence of Denticin D (Figure 1E). Prediction based on the expected ATG start codon, as per promoter sequences, indicate that Denticin D, lacks the first 11 amino acids of the otherwise conserved leader peptide (Figure 3A). The $P_{dntD}$ promoter may regulate an operon encompassing five genes: *dntD-blpX-dntC-blpZ-blpY*. This stop mutation could lead to the inadvertent translation of a six amino acids small peptide and the absence of another RBS downstream of the stop mutation might hinder the translation of the remaining sequence of Denticin D. However, the other four genes within this putative operon might be translated from their own RBSs.

### Bacteriocins' class predictions

Bacteriocin class predictions were conducted using InterPro [29]. BlpM, BlpN, BlpD, BlpE, Denticin A, Denticin D, and Denticin E were predicted as lactacin-related, class IIb (two-peptide) bacteriocins (InterPro entry IPR019493). Denticin B contains the conserved 'pediocin box' motif, YGNGL, and two cysteine residues at the *N*-terminal region of the core peptide (Figure 3A). Accordingly, it is predicted as a pediocin-like, class IIa bacteriocin (InterPro entry IPR002633) (Table 4). Denticin B is located adjacent to the only enterocin A-like immunity protein predicted coding gene of the *blp*-like region (Figure 1A and C). This genetic proximity suggests a functional bacteriocin/immunity protein relationship, leading us to

rename locus SDENT7746_08245 as the gene *ipdB*, representing the coding gene of the (putative) immunity protein for Denticin B (Table 2). Finally, within the *blp*-like region, no bacteriocin class was predicted for the *S. dentisani* 7746 peptide's sequences PncT, BlpQ, and Denticin C. Since Denticin C shares a 48% conservation of the precursor peptide with Mutacin IV, we speculate that it could also be a lactacin-related, class IIb bacteriocin (Table 4).

In the *com*-like cluster, Denticin H could not be assigned to any family type, whereas Denticin F and Denticin G belong to the class IIb lactobin A/cerein 7B family (InterPro entry IPR023991). This family is characterised by low-complexity pore-forming bacteriocins whose processing does not include post-translational modifications beyond the proteolytic cleavage at the double-glycine motif. Accordingly, Denticin F and Denticin G do not contain cysteine residues in their core peptides (Figure 3).

### *Analysis of the transcriptional expression of the bacteriocins*

BIP- and CSP-induced bacteriocins have different biological roles. As so, they are expressed differently through the cell cycle [21]. Based on our *in silico* observations, we next aimed at corroborating the expression *in vivo* of the 14 identified bacteriocins. Towards this end, we performed RT-PCR with cross-gene primers designed to amplify putative transcriptional units within the *blp*- and *com*-like regions with samples from mid-exponential, late-exponential and stationary phase cultures of *S. dentisani* 7746 (Sup. Fig. S2).

Within the *blp*-like region, cross-gene primers were designed to amplify the transcriptional units predicted according to the six identified promoter sequences (Figure 1D). In that way, six RT-PCR reactions were sufficient to amplify all the bacteriocin encoding genes from this region (Figure 4). A summary figure with transcript samples from the *blp*-like region at stationary phase is presented in Figure 4B. All the RT-PCR reactions yielded amplification products of the expected size, which confirms the transcription of the eleven bacteriocins from the *blp*-like region. Of note, despite the longer intergenic region between *blpQ* and *pncT*, where a transcriptional terminator with a free energy of –11.1 Kcal/mol is predicted, both genes are co-transcribed, suggesting that —if present— this terminator is overridden.

In the case of the *cib* genes, we performed several RT-PCR reactions with cross-gene primers designed to amplify different length fragments from *dntF* to *impD* (Figure 4C). At mid-exponential phase, all the RT-PCR reactions yielded amplification products of the predicted size (Figure 4D). This is expected since competence induction via ComX occurs in a short time window during the exponential phase. These results support *i)* the expression of the three bacteriocins from this region, and *ii)* the co-transcription of all the genes from *dntF* to *impD* in a single polycistronic mRNA, whose expression must be driven by the $\sigma^X$ target promoter identified upstream of *dntF*.

Altogether, these results demonstrate the expression of the 14 bacteriocin genes encoded by *S. dentisani* 7746. Future studies will address the detailed quantification of bacteriocin expression dynamics through the cell cycle.

## Discussion

As streptococci are prompt to natural transformation, horizontal gene transfer events might result in extensive genomic rearrangements and acquisition of new systems. In that way, some streptococcal strains possess multiple ComDE and BlpRH systems distributed along the genome and the high similarity between both systems has sometimes resulted in genomic miss-annotations [24]. In this regard, a previous study reported the presence of up to 11 bacteriocin peptides in a bacteriocin coding region in *S. dentisani* 7746 genome, and a processing and export system located next to the bacteriocins was identified as ComAB; a TCS far away in the genome was postulated to be the regulatory system of this bacteriocin region and it was accordingly named ComCDE [14]. Synchronically, a different study reported the presence of a similar bacteriocin cassette with a full regulatory set (composed by TCS and cognate pheromone) located upstream of the processing and export system. In this study the system was reported as a Blp-like bacteriocin cassette [23]. The discrepancies between these two previous works suggested the existence of more than one bacteriocin coding region in *S. dentisani* 7746. Our investigations have shed light on this matter.

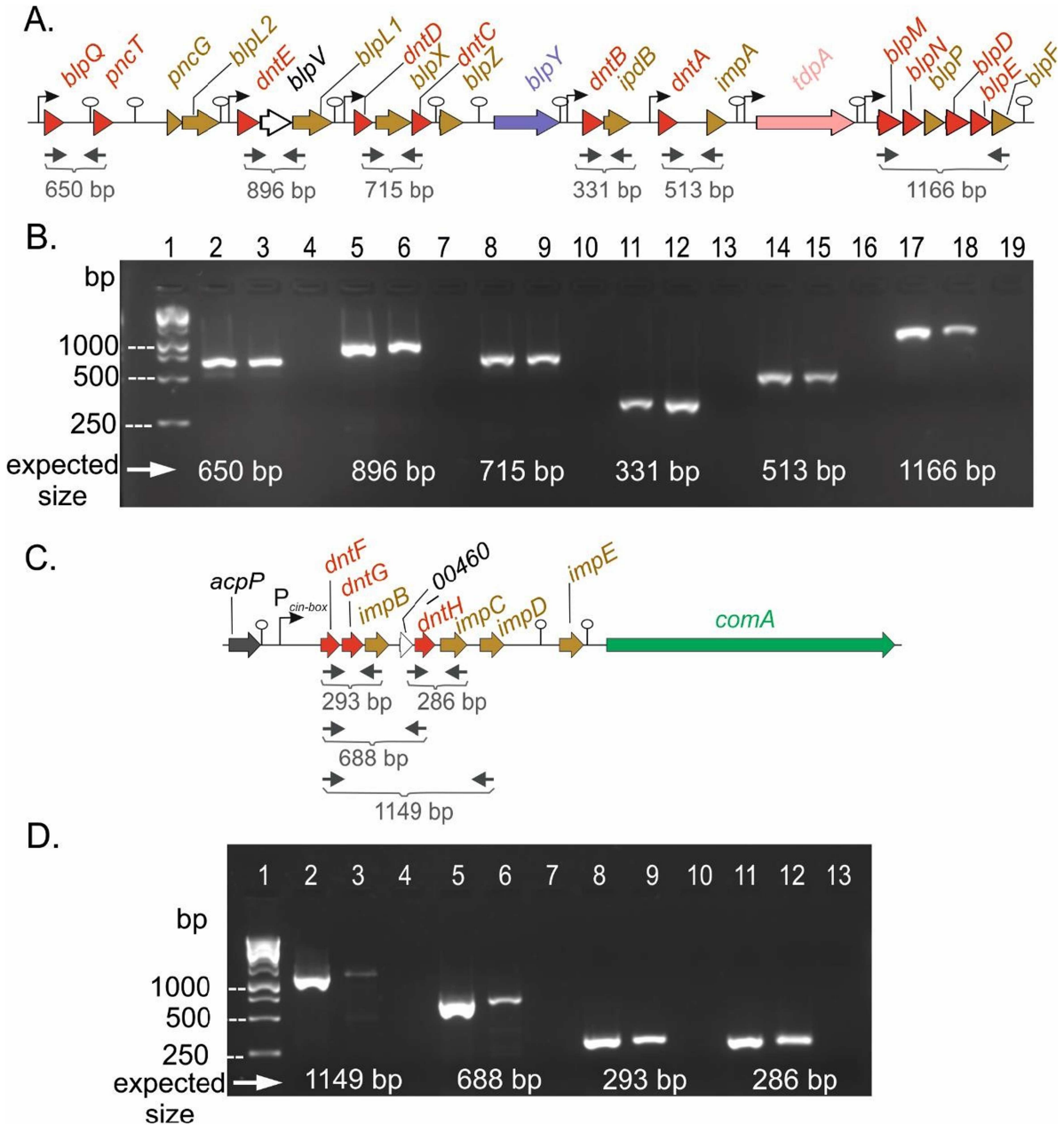


**Figure 4.** Analysis of the expression of the 14 bacteriocins from *S. dentisani* 7746 through the cell cycle. A and C: Schematics of the binding sites of the cross-gene PCR primers for the bacteriocin and immunity proteins from the *blp*-like region (A) and the *com*-like region (C). Genes are represented as in Figures 1 and 2, respectively; the thin grey arrows below the genes indicate the position and the product size of the cross-gene primers used for the RT-PCR presented in (B) and (D); the positions of putative terminators are indicated by lollipops; the *blp*-like promoters and the *cin-box* promoter are indicated by solid black bent arrows. B. RT-PCR analysis of transcripts from the *blp*-like region at stationary phase. The PCR amplification products correspond to the following cross-gene primers: AM033– AM028 (lanes 2-4, amplification of intergenic region *blpQ-pncT*, product size [650 bp]); AM025—AM066 (lanes 5-7, intergenic region *dntE-blpL1* [896 bp]); RG243—RG244 (lanes 8-10; intergenic region *dntD-dntC* [715 bp]); RG245—RG246 (lanes 11-13; intergenic region *dntB-ipdB* [331 bp]); RG247—RG248 (lanes 14-16; intergenic region *dntA-impA* [513 bp]); and RG249—RG250 (lanes 17-19; intergenic region *blpM-blpF* [1166 bp]). Line 1 contains the molecular weight ladder. Negative controls using RNA as a template are presented in lanes 4, 7, 10, 13, 16, and 19 to verify the absence of residual genomic DNA in the purified RNA samples. Positive controls amplified from genomic DNA are presented in lanes 2, 5, 8, 11, 14, and 17. Lanes 3, 6, 9, 12, 15, and 18 contain the RT-PCR products amplified from the cDNA. D. RT-PCR analysis of transcripts from the *com*-like region at mid-exponential phase. The PCR amplification products correspond to the following cross-gene primers: RG251—RG252 (lanes 1-4; intergenic region *dntF-impD* [1149 bp]); RG251—RG257 (lanes 5-7; intergenic region *dntF-dntH* [688 bp]); RG251—RG258 (lanes 8-10; intergenic region *dntF-impB* [293 bp]); and RG259—RG260 (lanes 11-13; intergenic region *dntH-impC* [286 bp]). Line 1 contains the molecular weight ladder. Negative controls using RNA as a template are

(*Caption on next page*)

presented in lanes 4, 7, 10 and 13. Positive controls amplified from genomic DNA are presented in lanes 2, 5, 8 and 11. Lanes 3, 6, 9 and 12 contain the RT-PCR products amplified from the cDNA. Cropped gels are depicted for clarity in B and D; the original gel pictures can be found in Supplementary Fig. S3 and S4.

The present work was designed as a descriptive genomic and transcriptional study aimed at characterising the bacteriocin repertoire encoded by *S. dentisani* 7746, rather than addressing the specific antimicrobial function or ecological role of each peptide. For that, we mined the recently closed, complete genome of *S. dentisani* 7746 using the web informatic tools BAGEL4 and antiSMASH and we identified a second bacteriocin coding region in the 7746 genome sequence. This region contains three additional bacteriocin coding genes, thus increasing to 14 the total number of bacteriocins encoded by 7746. The two bacteriocin containing regions were analysed in detail for regulatory systems, and besides the bacteriocin genes, full-length genes for the proteins potentially involved in regulation (including inducer peptides), synthesis, export-and-processing of the bacteriocins, as well as the putative self-immunity factors were identified and classified. Our results show that *S. dentisani* 7746 has two complete sets of Com- and Blp-like quorum sensing systems and indicate that in the work by Conrads et al. the *blpAB* genes were misidentified as *comAB* [14]. Therefore, the authors proposed a model with a direct regulatory link between ComCDE and BlpAB. Our results expand and complement this previous model since we demonstrate the presence of genes for more regulatory elements, named the *blpSRH-blpC* (Figure 1), the genes for the actual *comA* transporter and *comB* accessory proteins, as well as other regulatory elements such as *comX1* and *comX2* (Figure 2). Consequently, more complex regulatory interactions differing from the previously described ones might be inferred for *S. dentisani* 7746. Although it remains to be experimentally proven if strain 7746 is endowed for natural transformation, we are tempted to postulate a model for 7746 similar to what happens in *S. pneumoniae* strains: competence induction would be regulated by the ComAB-ComC-ComDE system, and the expression of the bacteriocins and self-immunity proteins from the *blp*-like BIR region would be regulated by BlpAB-BlpC-BlpSRH system (Figure 5). Importantly, in strain CECT 7746, both *blpA* and *comA* coded for proteins of 717 aa, and *blpB* and *comB* for transport accessory proteins of 453 and 449 aa, respectively. This indicates that all the transport proteins identified in *S. dentisani* 7746 were full-length alleles, which suggests that both systems involved in export and processing of CSP, BIP and the bacteriocins are more likely functional in this strain. Although the functionality of both systems in 7746 remains to be experimentally validated—for example, by generating mutant strains –, this is worth noticing since a study with more than four thousand *S. pneumoniae* genomes showed that only 23.5% of the genomes contained full-length *blpA* and *blpB* sequences, which encoded functional bacteriocin transport systems [32]. Knowledge of the regulatory mechanisms and conditions triggering bacteriocin production in *S. dentisani* 7746 could be truly relevant for its development as an industrial bacteriocin producer (e.g. through the addition of autoinducer peptides to increase yield) [47]. Future experimental efforts will be addressed to corroborate the functionality of the Com and Blp systems, as well as the putative interdependencies between them [48]. In this regard, we identified conserved promoter sequences within the BIR region. The high conservation of the $P_{dntA}$, $P_{blpQ}$, $P_{dntE}$, $P_{dntB}$, $P_{blpM}$ and $P_{dntD}$ promoters, and the partial conservation of the $P_{blpS}$, $P_{tdpA}$, $P_{blpA}$ and $P_{blpT}$, suggest their regulation by the BlpR response regulator (Figure 5). However, a comprehensive assessment of the regulatory mechanism is beyond the scope of this paper, and more studies, such as RR-DNA binding assays, will be necessary to confirm these hypotheses in future research.

We also analysed all the bacteriocin precursor peptides identified regarding orthology conservation and bacteriocins class predictions. The precursor peptide of Gly-Gly leader peptide bacteriocins contains the leader peptide and the core peptide. The leader peptide is highly conserved for the recognition by the export and processing enzymes [10]. The hypervariable core peptide will constitute the active mature peptide upon proteolytic removal of the leader peptide. Our results considering the precursor peptide and core peptide conservation support the previously described orthologs in *S. pneumoniae* for the bacteriocins BlpQ, PncT, BlpM, BlpN, BlpD, BlpE [14]. Conversely, we have assigned names for the bacteriocins with unnamed orthologs in other species, proposing the name Denticins (from Denticin A to Denticin H) to these eight novel peptides. The mechanism by which such a high number of diverse antimicrobial peptides has evolved should be studied in the future. The origin of the the *blp*-like region might be

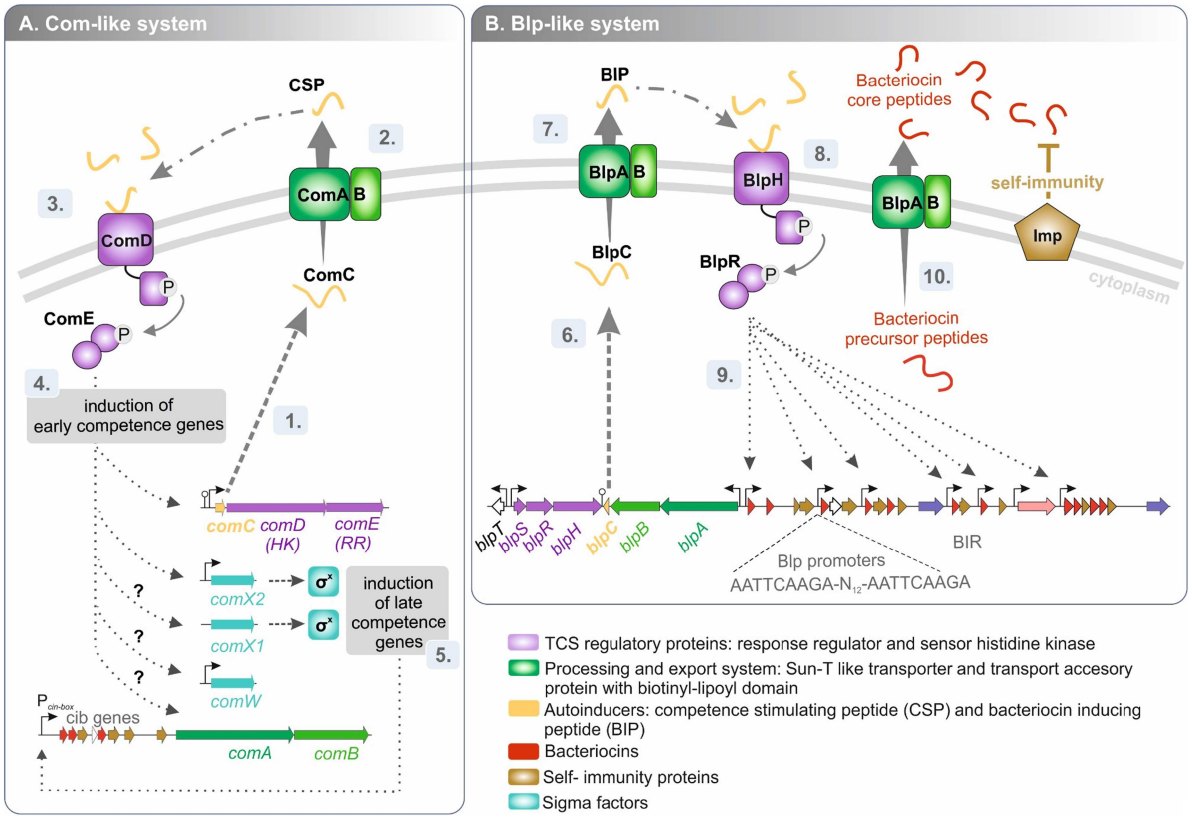


**Figure 5.** Predictive model for the regulation of quorum sensing dependent cellular functions in *S. dentisani* 7746 by Com- and Blp-like peptide-regulated systems, based on *in silico* predictions. A. The induction of genetic competence would depend on the ComAB-ComCDE system. 1) The *comC* gene encodes ComC, the precursor peptide for CSP (competence stimulating peptide) with a Gly-Gly motif leader peptide sequence. 2) ComC is post-translationally modified to CSP during its export by the ABC- transporter ComA and the transmembrane transport accessory protein ComB. 3) The processed CSP is detected by the HK ComD which activates the RR ComE. 4) Activated ComE might upregulate the expression of genes involved in the early stages of competence. 5) The induction of the *cib* (competence-induced bacteriocin) genes within the Com region, might be mediated by the ComX1/2 sigma factors at late stages of competence. B. The expression of bacteriocins and autoimmunity proteins from the *blp*-like region would be regulated by the BlpABC-BlpSRH system and would function likewise. 6) The *blpC* gene encodes for BlpC, a pheromone with a leader peptide with a Gly-Gly motif. 7) BlpC is exported and processed by the transporter BlpA and the transport accessory protein BlpB. 8) The mature form of BlpC, the bacteriocin-inducing peptide, BIP, is detected by HK BlpH, which activates the RR BlpR. 9) Activated BlpR might regulate the expression of bacteriocin and autoimmunity genes within the bacteriocin immunity region (BIR). 10) Bacteriocin precursor peptides are exported and processed by BlpAB. Immunity against the active peptides is mediated by the self-immunity proteins. Colour code for depicted proteins is indicated in the figure. Colour code for the genes is like in Figures 1 and 2. Transcription from *comC* and *blpC*, and translation of ComC and BlpC is indicated by densely dashed upward-facing arrows. Processing and transport of the autoinducer precursors ComC and BlpC, as well as of the precursor bacteriocin peptides, by the Sun-T like transporters is indicated by thick upward-facing arrows. Recognition of the CSP and BIP autoinducers by the HKs is indicated by dash-dotted arrows. Putative induction of transcription by the RRs is indicated by downward- facing dotted arrows. Phospho-transfer between the HKs and the RRs are indicated by grey thin solid arrows. Self-immunity is indicated with a barred, upward-facing brown line. Question marks indicate genes without clear binding sites in their promoter regions. This figure presents a conceptual model derived solely from *in silico* genomic analyses, and experimental validation remains necessary to confirm the regulatory interactions and functional relationships proposed.

horizontal gene transfer, but the possibility of internal duplication events followed by sequence divergence cannot be ruled out. In this scenario, new bacteriocin variants generated by mutation accumulation could confer antagonistic properties against different oral competitors. This could of course be used with applied purposes. Given that *S. dentisani* 7746 supernatants (containing an unknown blend of bacteriocins) have been found to be active against Gram-positive and Gram-negative bacteria, as well as *Candida albicans* yeasts [49], the identification of denticins with inhibitory activity against specific pathogens could allow the

design of bacteriocin cocktails to treat or prevent oral conditions such as tooth decay, gum disease or halitosis. To this end, our current experimental efforts are focused on identifying the bacterial targets of these bacteriocins. Since purifying the peptides from culture supernatants is technically challenging due to the complex composition of the culture media, we are instead pursuing chemical synthesis, followed by sensitivity assays against the pathogens of interest (Sup. File 5). However, this is an expensive and not always straightforward approach (Sup. Fig. S5). Additional strategies, such as heterologous expression of the peptides [50,51], may also be explored in the future.

In agreement with our bioinformatics analysis, our experimental results prove that all 14 bacteriocins are transcriptionally expressed by *S. dentisani* 7746. As for whether the bacteriocins are produced and exported out of the cell, our results suggest that apart from DntD, the remaining 13 are expected to be found in culture supernatants; however, post-transcriptional regulatory processes, as well as the antimicrobial activity, target specificity, and physiological relevance of individual denticins under *in vivo* conditions remain to be investigated. These aspects will require dedicated functional, biochemical, and ecological studies. With regard to the number of bacteriocins encoded by strain 7746, other works have reported up to seven bacteriocins per genome within *S. pneumoniae* strains [32,34], and up to ten bacteriocins per genome within *Streptococcus salivarius* [52]. As for now, *S. oralis* subsp. *dentisani* strain 7746 is one of the bacterial isolates with the largest repertoire of bacteriocin coding genes known to date. This exceptional bacteriocin arsenal may have been selected to defend this organism from the highly diverse and competitive oral biofilm it inhabits, where several hundred species may physically coexist and several other antimicrobial defence mechanisms have been identified [53,54].

## Author contributions

CRediT: **Ainhoa Revilla-Guarinos:** Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Supervision, Validation, Visualization, Writing – original draft, Writing – review & editing; **Anny Camelo Castillo:** Investigation; **Rubén Cebrián:** Investigation, Writing – review & editing; **María D. Ferrer:** Investigation; **Arantxa López-López:** Investigation; **Ana Adrados-Planell:** Methodology, Supervision; **Sandra Lahoz Oliva:** Methodology; **Laura Ledesma:** Data curation; **Pascal Hols:** Formal analysis, Writing – review & editing; **Álex Mira:** Conceptualization, Funding acquisition, Project administration, Resources, Supervision, Writing – review & editing.

## Disclosure statement

A.M. is the inventor of a patent protecting the use of *S. dentisani* as a probiotic. L. L. and P. H. declare that they are listed as inventors on patent(s) or patent application(s) related to bacteriocin production and uses. The other authors declare no conflicts of interest.

## Funding

A.R.-G.'s research was funded by the European Union's Horizon 2020 Research and Innovation Programme with the research grant SMILES (Marie Sklodowska-Curie grant agreement No 101026278); by the Fundación para el Fomento de la Investigación Sanitaria y Biomédica de la Comunitat Valenciana (FISABIO Foundation) with the research grant BACTIDENT (UGP-20-108); and by *Subvenciones a Grupos de Investigación Emergentes 2024* from the Generalitat Valenciana, Conselleria de Educación, Cultura, Universidades y Empleo through Project PepDent (CIGE/2023/211). R.C. is supported by the Instituto de Salud Carlos III (ISCIII, Miguel Servet programme, Ref: CP21/00113, Spain). P.H. is Research Director at FNRS. Funding sources were not involved in the study design, collection, analysis and interpretation of data, writing of the report and decision to submit the article for publication.

## ORCID

Ainhoa Revilla-Guarinos 0000-0002-9894-3538
Anny Camelo Castillo 0000-0002-4080-6346
Rubén Cebrián 0000-0003-1575-1846
María D. Ferrer 0000-0002-4521-1632
Arantxa López-López 0000-0002-6081-7608
Ana Adrados-Planell 0000-0003-3944-2503
Sandra Lahoz Oliva 0000-0002-5599-5549

Laura Ledesma 0000-0003-3276-2192
Pascal Hols 0000-0001-8749-3650
Álex Mira 0000-0002-9127-3877

## Data availability statement

The data supporting the conclusions of this article can be found in the Supplementary Information files.

## References

[1] Hegarty JW, Guinane CM, Ross RP, et al. Bacteriocin production: a relatively unharnessed probiotic trait? F1000Res. 2016;5:2587. doi: 10.12688/f1000research.9615.1
[2] Hols P, Ledesma-García L, Gabant P, et al. Mobilization of microbiota commensals and their bacteriocins for therapeutics. Trends Microbiol. 2019;27(8):690–702. doi: 10.1016/j.tim.2019.03.007
[3] Sugrue I, Ross RP, Hill C. Bacteriocin diversity, function, discovery and application as antimicrobials. Nat Rev Microbiol. 2024;22(9):556–571. doi: 10.1038/s41579-024-01045-x
[4] Van Holm W, Carvalho R, Delanghe L, et al. Antimicrobial potential of known and novel probiotics on in vitro periodontitis biofilms. NPJ Biofilms Microbiomes. 2023;9(1):3. doi: 10.1038/s41522-023-00370y
[5] Cotter PD, Hill C, Ross RP. Bacteriocins: developing innate immunity for food. Nat Rev Microbiol 2005;3(10):777–788. doi: 10.1038/nrmicro1273
[6] Draper LA, Cotter PD, Hill C, et al. Lantibiotic resistance. Microbiol Mol Biol Rev. 2015;79(2):171–191. doi: 10.1128/MMBR.00051-14
[7] Jaumaux F, Gómez de Cadiñanos L, Gabant P. In the age of Synthetic Biology, will antimicrobial peptides be the next generation of antibiotics? Antibiotics (Basel). 2020;9(8):484. doi: 10.3390/antibiotics9080484
[8] Lewies A, Du Plessis LH, Wentzel JF. Antimicrobial peptides: the achilles' heel of antibiotic resistance? Probiotics Antimicrob Proteins. 2019;11(2):370–381. doi: 10.1007/s12602-018-9465-0
[9] Perron GG, Zasloff M, Bell G. Experimental evolution of resistance to an antimicrobial peptide. Proc Biol Sci. 2006;273(1583):251–256. doi: 10.1098/rspb.2005.3301
[10] Arnison PG, Bibb MJ, Bierbaum 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):108–160. doi: 10.1039/c2np20085f
[11] Gebhard S. ABC transporters of antimicrobial peptides in Firmicutes bacteria - phylogeny, function and regulation. Mol Microbiol. 2012;86(6):1295–1317. doi: 10.1111/mmi.12078
[12] van Heel AJ, de Jong A, Song C, et al. BAGEL4: a user-friendly web server to thoroughly mine RiPPs and bacteriocins. Nucleic Acids Res. 2018;46(W1):W278–W281. doi: 10.1093/nar/gky383
[13] Weber T, Blin K, Duddela S, et al. antiSMASH 3.0-a comprehensive resource for the genome mining of biosynthetic gene clusters. Nucleic Acids Res. 2015;43(W1):W237–W243. doi: 10.1093/nar/gkv437
[14] Conrads G, Westenberger J, Lürkens M, et al. Isolation and bacteriocin-related typing of streptococcus dentisani. Front Cell Infect Microbiol. 2019;9:110. doi: 10.3389/fcimb.2019.00110
[15] Camelo-Castillo A, Benítez-Páez A, Belda-Ferre P, et al. Streptococcus dentisani sp. nov., a novel member of the mitis group. Int J Syst Evol Microbiol. 2014;64(Pt 1):60–65. doi: 10.1099/ijs.0.054098-0
[16] Lopez-Santacruz HD, López-Santacruz HD, López-López A, et al. Streptococcus dentisani is a common inhabitant of the oral microbiota worldwide and is found at higher levels in caries-free individuals. Int Microbiol. 2021;24(4):619–629. doi: 10.1007/s10123-021-00222-9
[17] López-López A, Camelo-Castillo A, Ferrer MD, et al. Health-associated niche inhabitants as oral probiotics: the case of streptococcus dentisani. Front Microbiol. 2017;8:379–379. doi: 10.3389/fmicb.2017.00379
[18] Mascher T, Helmann JD, Unden G. Stimulus perception in bacterial signal-transducing histidine kinases. Microbiol Mol Biol Rev. 2006;70(4):910–938. doi: 10.1128/mmbr.00020-06
[19] Håvarstein LS, Coomaraswamy G, Morrison DA. An unmodified heptadecapeptide pheromone induces competence for genetic transformation in Streptococcus pneumoniae. Proc Natl Acad Sci U S A. 1995;92(24):11140–11144. doi: 10.1073/pnas.92.24.11140
[20] Pestova EV, Håvarstein LS, Morrison DA. Regulation of competence for genetic transformation in Streptococcus pneumoniae by an auto-induced peptide pheromone and a two-component regulatory system. Mol Microbiol. 1996;21(4):853–862. doi: 10.1046/j.1365-2958.1996.501417.x
[21] Claverys JP, Martin B, Håvarstein LS. Competence-induced fratricide in streptococci. Mol Microbiol. 2007;64(6):1423–1433. doi: 10.1111/j.1365-2958.2007.05757.x
[22] de Saizieu A, Gardès C, Flint N, et al. Microarray-based identification of a novel Streptococcus pneumoniae regulon controlled by an autoinduced peptide. J Bacteriol. 2000;182(17):4696–4703. doi: 10.1128/JB.182.17.4696-4703.2000
[23] Kilian M, Tettelin H. Identification of virulence-associated properties by comparative genome analysis of *Streptococcus pneumoniae*, *S. pseudopneumoniae*, *S. mitis*, Three S. oralis subspecies, and S. infantis. mBio. 2019;10(5):1–32. doi: 10.1128/mBio.01985-19

[24] Dhaked HPS, Biswas I. Distribution of two-component signal transduction systems BlpRH and ComDE across streptococcal species. Front Microbiol. 2022;13:960994. doi: 10.3389/fmicb.2022.960994
[25] Blin K, Shaw S, Kloosterman AM, et al. antiSMASH 6.0: improving cluster detection and comparison capabilities. Nucleic Acids Res. 2021;49(W1):W29–W35. doi: 10.1093/nar/gkab335
[26] Altschul SF, Gish W, Miller W, et al. Basic local alignment search tool. J Mol Biol. 1990;215(3):403–410. doi: 10.1016/S0022-2836(05)80360-2
[27] Dehal PS, Joachimiak MP, Price MN, et al. MicrobesOnline: an integrated portal for comparative and functional genomics. Nucleic Acids Res. 2010;38(Database issue):D396–D400. doi: 10.1093/nar/gkp919
[28] Letunic I, Khedkar S, Bork P. SMART: recent updates, new developments and status in 2020. Nucleic Acids Res. 2021;49(D1):D458–D460. doi: 10.1093/nar/gkaa937
[29] Blum M, Chang H, Chuguransky S, et al. The InterPro protein families and domains database: 20 years on. Nucleic Acids Res. 2021;49(D1):D344–D354. doi: 10.1093/nar/gkaa977
[30] Ye J, Coulouris G, Zaretskaya I, et al. Primer-BLAST: a tool to design target-specific primers for polymerase chain reaction. BMC Bioinform. 2012;13:134. doi: 10.1186/1471-2105-13-134
[31] Letunic I, Bork P. 20 years of the SMART protein domain annotation resource. Nucleic Acids Res. 2018;46(D1):D493–D496. doi: 10.1093/nar/gkx922
[32] Kjos M, Miller E, Slager J, et al. Expression of streptococcus pneumoniae bacteriocins is induced by antibiotics via regulatory interplay with the competence system. PLoS Pathog. 2016;12(2):e1005422. doi: 10.1371/journal.ppat.1005422
[33] Martin B, Soulet A, Mirouze N, et al. ComE/ComE~P interplay dictates activation or extinction status of pneumococcal X-state (competence). Mol Microbiol. 2013;87(2):394–411. doi: 10.1111/mmi.12104
[34] Bogaardt C, van Tonder AJ, Brueggemann AB. Genomic analyses of pneumococci reveal a wide diversity of bacteriocins - including pneumocyclicin, a novel circular bacteriocin. BMC Genomics. 2015;16:554. doi: 10.1186/s12864-015-1729-4
[35] Kjos M, Snipen L, Salehian Z, et al. The abi proteins and their involvement in bacteriocin self-immunity. J Bacteriol. 2010;192(8):2068–2076. doi: 10.1128/JB.01553-09
[36] Pei J, Grishin NV. Type II CAAX prenyl endopeptidases belong to a novel superfamily of putative membrane-bound metalloproteases. Trends Biochem Sci. 2001;26(5):275–277. doi: 10.1016/s0968-0004(01)01813-8
[37] Johnsen L, Dalhus B, Leiros I, et al. 1.6-Angstroms crystal structure of EntA-im. A bacterial immunity protein conferring immunity to the antimicrobial activity of the pediocin-like bacteriocin enterocin A. J Biol Chem. 2005;280(19):19045–19050. doi: 10.1074/jbc.M501386200
[38] Inniss NL, Prehna G, Morrison DA. The pneumococcal sigma(X) activator, ComW, is a DNA-binding protein critical for natural transformation. J Biol Chem. 2019;294(29):11101–11118. doi: 10.1074/jbc.RA119.007571
[39] Sung CK, Morrison DA. Two distinct functions of ComW in stabilization and activation of the alternative sigma factor ComX in *Streptococcus pneumoniae*. J Bacteriol. 2005;187(9):30523061. doi: 10.1128/JB.187.9.3052-3061.2005
[40] Havarstein LS, Håvarstein LS, Hakenbeck R, et al. Natural competence in the genus Streptococcus: evidence that streptococci can change pherotype by interspecies recombinational exchanges. J Bacteriol. 1997;179(21):6589–6594. doi: 10.1128/jb.179.21.6589-6594.1997
[41] Fontaine L, Wahl A, Fléchard M, et al. Regulation of competence for natural transformation in streptococci. Infect Genet Evol. 2015;33:343–360. doi: 10.1016/j.meegid.2014.09.010
[42] Lee MS, Morrison DA. Identification of a new regulator in *Streptococcus pneumoniae* linking quorum sensing to competence for genetic transformation. J Bacteriol. 1999;181(16):5004–5016. doi: 10.1128/jb.181.16.5004-5016.1999
[43] Luo P, Li H, Morrison DA. Identification of ComW as a new component in the regulation of genetic transformation in Streptococcus pneumoniae. Mol Microbiol. 2004;54(1):172–183. doi: 10.1111/j.1365-2958.2004.04254.x
[44] Peterson SN, Sung CK, Cline R, et al. Identification of competence pheromone responsive genes in Streptococcus pneumoniae by use of DNA microarrays. Mol Microbiol. 2004;51(4):1051–1070. doi: 10.1046/j.1365-2958.2003.03907.x
[45] Ween O, Gaustad P, Håvarstein LS. Identification of DNA binding sites for ComE, a key regulator of natural competence in Streptococcus pneumoniae. Mol Microbiol. 1999;33(4):817–827. doi: 10.1046/j.1365-2958.1999.01528.x
[46] Claverys JP, Havarstein LS. Extracellular-peptide control of competence for genetic transformation in *Streptococcus pneumoniae*. Front Biosci. 2002;7:d1798–d1814. doi: 10.2741/claverys
[47] Muñoz M, Munoz M, Jaramillo D, et al. Native and heterologous production of bacteriocins from gram-positive microorganisms. Recent Pat Biotechnol. 2011;5(3):199–211. doi: 10.2174/187220811797579088
[48] Wholey WY, Kochan TJ, Storck DN, et al. Coordinated bacteriocin expression and competence in streptococcus pneumoniae contributes to genetic adaptation through neighbor predation. PLoS Pathog. 2016;12(2):e1005413. doi: 10.1371/journal.ppat.1005413
[49] Llena C, Almarche A, Mira A, et al. Antimicrobial efficacy of the supernatant of Streptococcus dentisani against microorganisms implicated in root canal infections. J Oral Sci. 2019;61(1):184–194. doi: 10.2334/josnusd.18-0239
[50] Mejia-Pitta A, Mejía-Pitta A, Broset E, et al. Probiotic engineering strategies for the heterologous production of antimicrobial peptides. Adv Drug Deliv Rev. 2021;176:113863. doi: 10.1016/j.addr.2021.113863

[51] Vermeulen RR, Van Staden ADP, Dicks L. Heterologous expression of the class IIa Bacteriocins, Plantaricin 423 and Mundticin ST4SA, in escherichia coli using green fluorescent protein as a fusion partner. Front Microbiol. 2020;11:1634. doi: 10.3389/fmicb.2020.01634

[52] Damoczi J, Knoops A, Martou M, et al. Uncovering the arsenal of class II bacteriocins in salivarius streptococci. Commun Biol. 2024;7(1):1511. doi: 10.1038/s42003-024-07217-y

[53] Kreth J, Zhang Y, Herzberg MC. Streptococcal antagonism in oral biofilms: Streptococcus sanguinis and Streptococcus gordonii interference with Streptococcus mutans. J Bacteriol. 2008;190(13):46324640. doi: 10.1128/JB.00276-08

[54] Marsh PD, Zaura E. Dental biofilm: ecological interactions in health and disease. J Clin Periodontol. 2017;44(Suppl 18):S12–S22. doi: 10.1111/jcpe.12679