

RESEARCH ARTICLE

Open Access

# Insights into 6S RNA in lactic acid bacteria (LAB)



Pablo Gabriel Cataldo<sup>1</sup>, Paul Klemm<sup>2</sup>, Marietta Thüring<sup>2</sup>, Lucila Saavedra<sup>1</sup>, Elvira Maria Hebert<sup>1</sup>, Roland K. Hartmann<sup>2</sup> and Marcus Lechner<sup>2,3\*</sup>

## Abstract

**Background:** 6S RNA is a regulator of cellular transcription that tunes the metabolism of cells. This small non-coding RNA is found in nearly all bacteria and among the most abundant transcripts. Lactic acid bacteria (LAB) constitute a group of microorganisms with strong biotechnological relevance, often exploited as starter cultures for industrial products through fermentation. Some strains are used as probiotics while others represent potential pathogens. Occasional reports of 6S RNA within this group already indicate striking metabolic implications. A conceivable idea is that LAB with 6S RNA defects may metabolize nutrients faster, as inferred from studies of *Escherichia coli*. This may accelerate fermentation processes with the potential to reduce production costs. Similarly, elevated levels of secondary metabolites might be produced. Evidence for this possibility comes from preliminary findings regarding the production of surfactin in *Bacillus subtilis*, which has functions similar to those of bacteriocins. The prerequisite for its potential biotechnological utility is a general characterization of 6S RNA in LAB.

**Results:** We provide a genomic annotation of 6S RNA throughout the *Lactobacillales* order. It laid the foundation for a bioinformatic characterization of common 6S RNA features. This covers secondary structures, synteny, phylogeny, and product RNA start sites. The canonical 6S RNA structure is formed by a central bulge flanked by helical arms and a template site for product RNA synthesis. 6S RNA exhibits strong syntenic conservation. It is usually flanked by the replication-associated recombination protein A and the universal stress protein A. A catabolite responsive element was identified in over a third of all 6S RNA genes. It is known to modulate gene expression based on the available carbon sources. The presence of antisense transcripts could not be verified as a general trait of LAB 6S RNAs.

**Conclusions:** Despite a large number of species and the heterogeneity of LAB, the stress regulator 6S RNA is well-conserved both from a structural as well as a syntenic perspective. This is the first approach to describe 6S RNAs and short 6S RNA-derived transcripts beyond a single species, spanning a large taxonomic group covering multiple families. It yields universal insights into this regulator and complements the findings derived from other bacterial model organisms.

**Keywords:** 6S RNA, SsrS, ncRNA, CcpA, cre site, Lactic acid bacteria, LAB

\*Correspondence: [lechner@staff.uni-marburg.de](mailto:lechner@staff.uni-marburg.de)

<sup>2</sup>Philipps-Universität Marburg, Institut für Pharmazeutische Chemie, Marbacher Weg 6, 35032 Marburg, Germany

<sup>3</sup>Philipps-Universität Marburg, Center for Synthetic Microbiology (Synmikro), Hans-Meerwein-Straße 6, 35043 Marburg, Germany

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



© The Author(s). 2021 **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

## Background

### Lactic acid bacteria

Lactic acid bacteria (LAB) constitute a genotypically, phenotypically, and phylogenetically diverse group of Gram-positive bacteria that belongs to the taxonomic order of the *Lactobacillales*. Shared metabolic characteristics and evolutionary relationships have been used as common markers for the identification, classification, typing, and phylogenetic analysis of LAB species [1]. During the last few decades, the analysis of 16S rRNA gene similarity was combined with the study of the carbohydrate fermentation profile to classify new bacterial isolates. The ongoing exploration of the *Lactobacillus* genus has led to frequent taxonomic rearrangements [2]. One reason is the presence of odd similarities and ambiguities in 16S rRNA gene sequence comparisons, resulting in a biased annotation of strains, species, and even LAB genera at short and long phylogenetic distances [3]. Currently, LAB are grouped into six families: *Aerococcaceae*, *Carnobacteriaceae*, *Enterococcaceae*, *Lactobacillaceae*, *Leuconostocaceae*, and *Streptococcaceae*. These groups share the ability to catabolize sugars for the efficient production of lactic acid [4]. LAB constitute the most competitive and technologically relevant group of microorganisms Generally Recognized as Safe (GRAS). Their biotechnological relevance is a result of the many beneficial features that can be exploited, for instance, as starter cultures in the food industry, mediating the rapid acidification of raw material [4], or as probiotics, preventing the adherence, establishment, and replication of several enteric mucosal pathogens via exerting multiple antimicrobial activities [5]. Nevertheless, some LAB are opportunistic pathogens and can cause infections in individuals presenting some underlying disease or predisposing condition. The most prominent opportunistic pathogens are members of the genera *Streptococcus* (*S.*) and *Enterococcus* [6].

LAB are usually exposed to a wide range of harsh stresses, both in industrial environments and throughout the gastrointestinal tract. This includes acid, cold, drying, osmotic, and oxidative stresses [7]. Surviving these unfavorable conditions is a prerequisite to exert their expected activities [8]. While main stress-resistance systems have been documented in some LAB species, their regulation at the molecular level, including the role of non-coding RNAs (ncRNAs), is still far from being understood [9].

### 6S RNA

Over the last decades many small non-coding RNAs have been identified as key regulators in a variety of bacterial stress response pathways and in bacterial virulence [10–12]. A prominent example among these is 6S RNA encoded by a gene frequently termed *ssrS* according to the original gene designation in *Escherichia coli* [13, 14]. A 6S gene is found in nearly all bacterial genomes sequenced

so far [15, 16]. This includes species with highly condensed genomes such as the hyperthermophile *Aquifex aeolicus*, species that obtain energy through photosynthesis like *Rhodobacter sphaeroides*, as well as pathogens such as *Helicobacter pylori* [16–19]. The dissemination of 6S RNA and its usually growth phase-dependent and condition-specific expression profile are indicators of the RNA's regulatory impact. Its mechanistic features have been more intensely studied for the two model organisms *E. coli* and *Bacillus subtilis* [20, 21]. The latter belongs to the *Bacillales*, a sister-order of *Lactobacillales*. 6S RNA is about 160–200 nucleotides in length and adopts a rod-shaped structure with an enlarged internal loop or bulge flanked by large helical arms on both sides [22, 23]. 6S RNA can bind the DNA-dependent RNA polymerase (RNAP) in complex with the housekeeping sigma factor ( $\sigma^{70}$  in *E. coli* and  $\sigma^A$  in *B. subtilis*) in competition with regular DNA promoters. This sequestration of RNAP alters the housekeeping transcription at a global level that is seemingly advantageous when facing numerous types of stress [22, 24, 25]. When RNAP is bound, it can utilize 6S RNA as a template for the transcription of short product RNAs (pRNAs). Upon relief of stress, the transcribed pRNAs become increasingly long. When reaching a certain length (~14 nt in *B. subtilis*), pRNAs can persistently rearrange the structure of 6S RNA to induce RNAP release, thus restoring regular transcription [21, 26–30]. Studies in *E. coli* have provided evidence that nutrients are metabolized faster in 6S RNA knockout strains than in the parental wild type strain [29, 31]. Furthermore, knockout strains might have the so far unexplored potential to produce elevated levels of secondary metabolites such as surfactants.

### 6S RNA in lactic acid bacteria

The importance of 6S RNA in LAB is indicated by studies that report its abundant expression as well as metabolic changes upon its knockout. However, specific 6S RNA analyses in this important group of bacteria are scarce or the studied ncRNA was not recognized as 6S RNA. It is annotated only in about half of all LAB species analyzed in this study (539/1,092 genomes). Here, we identified it in about 91% of all known LAB species. An example is *L. delbrueckii*, an industrial starter for dairy products, where a highly abundant ncRNA was reported [32]. Though its function could not be specified further, the authors suspected it to act as an antisense RNA. In our study, we identified this 210 nt long ncRNA as 6S RNA. In another study, 6S RNA was identified along with two types of pRNAs via RNA sequencing of *S. pyogenes* [33].

For *Lactococcus lactis*, the expression of 6S RNA has been linked to the carbon catabolite repression protein CcpA that binds to DNA at *cis*-acting sequences. These sites are called catabolite responsive elements (*cre*) [34];

*cre* sites are degenerate pseudo-palindromes. In *Bacilli* a CcpA dimer was shown to bind to dsDNA upon association with the Ser46-phosphorylated form of histidine-containing phosphocarrier protein (HPr-Ser46-P) [35]. In *L. lactis*, 6S RNA levels were found to be increased during stationary and exponential phase in the presence of galactose or cellobiose, but not fructose, as the sole carbon source. CcpA repression is known to be relieved by galactose and cellobiose, but not by fructose. Moreover, 6S RNA was found to be about 3-fold upregulated in a CcpA-deficient mutant [34] and a *cre* element was identified upstream of the -35 region of its promoter. This indicates a potential interaction between CcpA and the 6S RNA gene that might be relevant for LAB in general. Notably, *B. subtilis* 6S-1 and 6S-2 RNA were not identified as a target for CcpA [36].

For *E. faecalis*, a major opportunistic human pathogen, an additional transcript antisense to 6S RNA was detected [37]. The authors proposed its participation in degradation or maturation of 6S RNA as both ncRNA products were present in a processed form. To our knowledge, an equivalent antisense product is not described for *E. coli* [37], *B. subtilis* or any other species to date (own observation). However, interdependent expression of genes around the 6S RNA locus was noticed for other bacteria, e.g. *R. sphaeroides* (Proteobacteria), where a salt stress-induced membrane protein gene on the opposite strand immediately downstream of the 6S RNA locus is expressed at elevated levels in a 6S RNA knockout strain [18].

Apart from these isolated findings, little is known about the sequence, structure, and physiological role of this regulatory ncRNA in the large and widely heterogeneous group of LAB. In this study, we have annotated and analyzed 6S RNAs systematically to lay a foundation for further investigations regarding its role in stress responses, metabolic processes and interactions with eukaryotic cells. Moreover, we investigated how wide-spread and universally relevant the species-specific observations stated above are for LAB (link to CcpA and the presence of an antisense transcript). This is also the first comparative study covering 6S RNAs in a set of taxonomic families, thus making it possible to draw more representative conclusions than in species-wise studies.

## Results

### Dissemination & phylogeny

We searched 6S RNA sequences in 1,092 genomes covering strains from all 371 sequenced LAB species publicly available in the NCBI database at the time of this study [38]. While two 6S RNA copies were reported for some *Firmicutes* including *Bacillus subtilis*, *Bacillus halodurans*, *Clostridium acetobutylicum*, *Oceanobacillus iheyensis*, and *Thermoanaerobacter tengcongensis* [15], only one

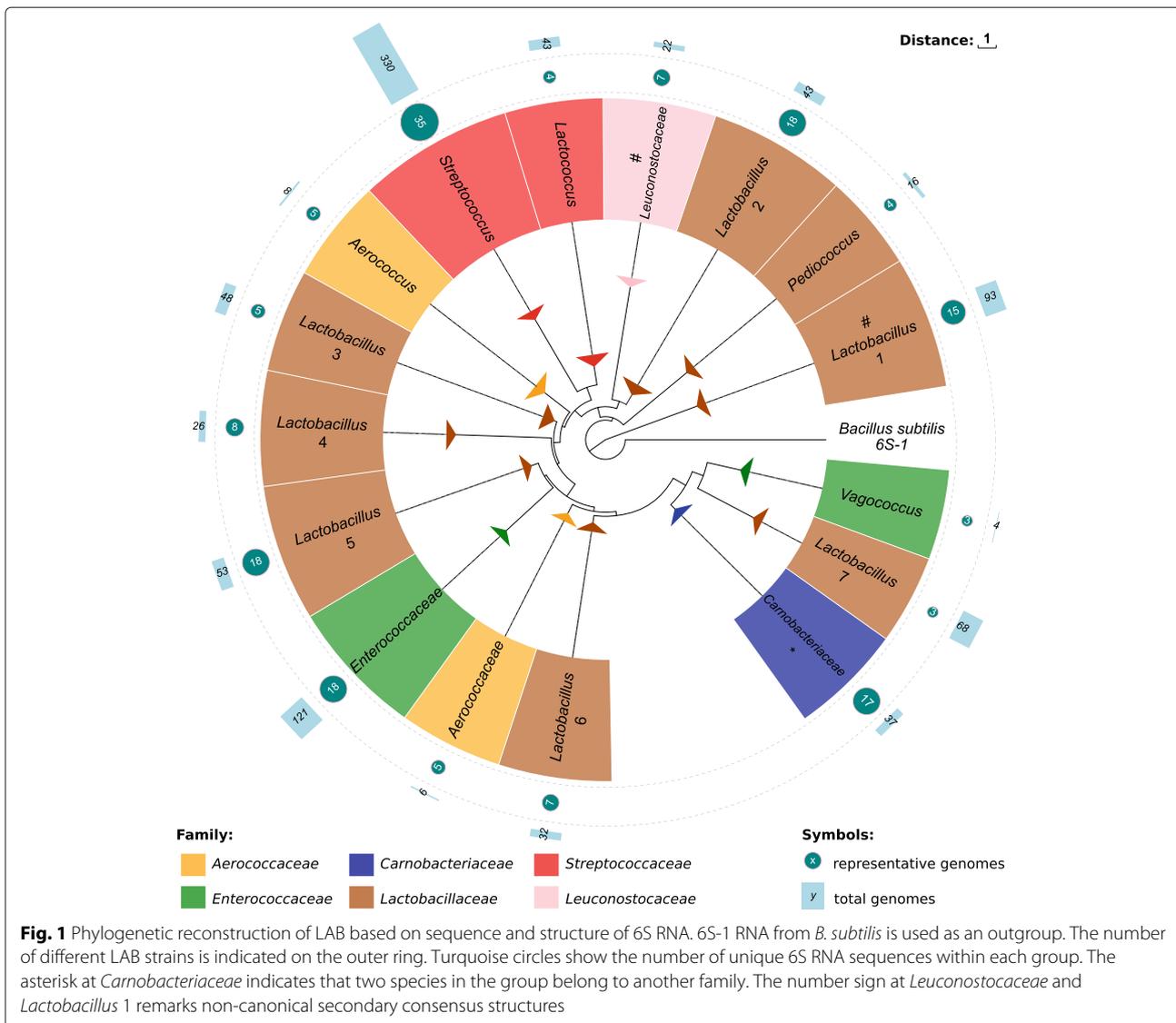
copy is present in LAB species. It shows more similarity to the major and well described *Bacillus subtilis* 6S-1 RNA than to its paralog 6S-2 RNA [39].

6S RNA was located in 1001 genomes (> 91%). Additional File 1 lists all loci. Genomes in which a 6S RNA gene could not be identified are predominantly partial genomes with a large number of contigs or scaffolds. When a 6S RNA gene was found in genomes of closely related species/strains, we assumed that the ncRNA is present but not part of the assembly yet. A peculiarity is the genus *Weissella* of the *Leuconostocaceae* family, represented with 13 species in our dataset. While only a weak 6S RNA locus was predicted in no more than four species of this genus, a significant amount of transcription could be shown for the syntenically conserved intergenic region downstream of *rarA* in publicly available RNA-Seq data for *W. confusa* and *W. koreensis* [40, 41]. Moreover, this locus is confined by a transcription terminator in most *Weissella* species. See Additional File 8 for details. This indicates that 6S RNAs in *Weissella* have a distinct singularity that was hardly picked up by our covariance-based search strategy. The typical rod-shaped structure with a central loop or bulge could not be confirmed for these non-canonical candidates.

Figure 1 shows the phylogeny of canonical 6S RNAs identified here based on their sequences and structural properties reconstructed using RNAclust [42] and mlocarna [43]. An alternative version with a resolution that reaches the species level is provided in Additional File 2. The phylogeny well resembles the taxonomic units at the level of genera. A minor exception is the *Carnobacteriaceae* group (blue) that includes *Abiotrophia defectiva* (*Aerococcaceae*) and *Bavariococcus seileri* (*Enterococcaceae*). At the level of taxonomic families, the genus *Vagococcus* is significantly different from other *Enterococcaceae* (green). Similarly, *Aerococcus* is different from other *Aerococcaceae*. *Lactobacillus* is known to be the most heterogeneous genus within LAB [1]. This is also reflected phylogenetically since the 6S RNAs of this genus are divided into eight well distinguishable groups (*Lactobacillus* 1-7, *Pediococcus*, brown).

### Relation to 16S rRNA phylogeny

The phylogenetic reconstruction of LAB species based on a sequence alignment of selected 16S rRNA sequences is shown analogous to the 6S RNA-based reconstruction in Additional File 3. As expected, the 16S rRNA-based approach better resembles the current taxonomic annotation [2, 44]. The majority of *Lactobacillaceae* species share a common subtree. Notably, a number of species from the *Lactobacillus* 6 group (6S RNA-based, see Fig. 1) is also located in a separate subtree in the 16S rRNA phylogeny. Similarly, the *Vagococcus* group is isolated from the remaining *Enterococcaceae* in both phylogenies and



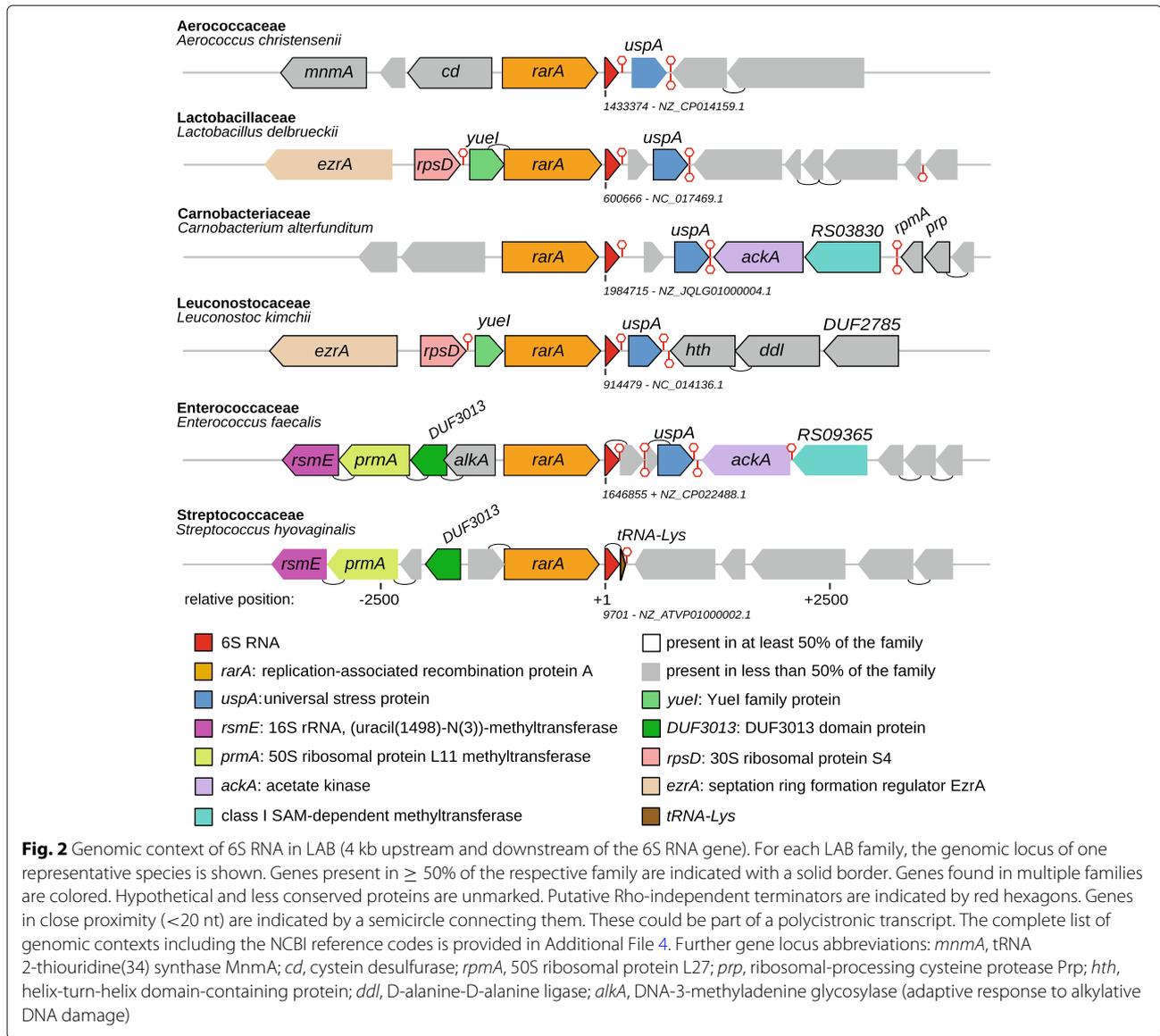
the same two family-foreign species are found within the *Carnobacteriaceae* subtree, namely *A. defectiva* (*Aerococcaceae*) and *B. seileri* (*Enterococcaceae*). In the 16S rRNA tree, the grouped *Aerococcaceae* are closely related to *Carnobacteriaceae*. The 6S RNA tree, in contrast, splits this group into two subgroups that are not closely related to *Carnobacteriaceae*.

### Synteny

To characterize the genomic locus of 6S RNA in LAB, a synteny analysis was performed. Proteinortho [45] was used to group the protein-coding genes in the vicinity of the 6S RNA locus. An overview of the genomic context of 6S RNA in LAB is shown in Fig. 2 and in more detail in Additional File 4. The genomic neighborhood of 6S RNA is conserved at the family level. Typically, the same genes are encoded

up- and downstream of 6S RNA in the majority of genera from the same taxonomic family but not across LAB in general. Exceptions are the replication-associated recombination protein A gene (*rara*), that is found upstream of the 6S RNA locus in nearly all species, and the universal stress protein A gene (*uspA*), that is found downstream across almost all species except for *Streptococcaceae* and a few *Aerococcaceae* members.

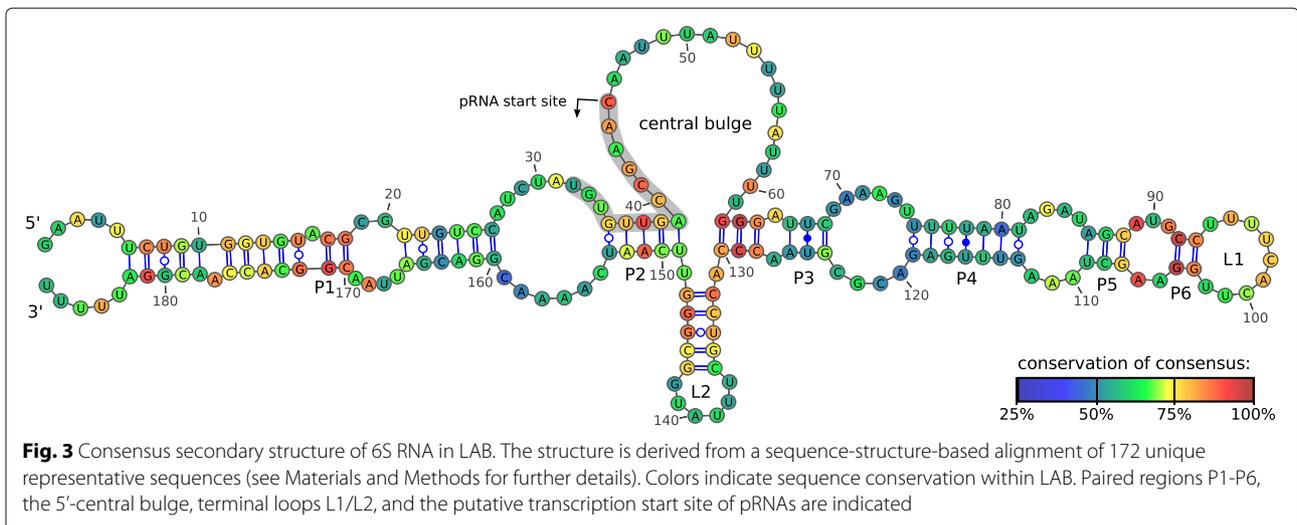
The upstream *rara* gene is part of a highly conserved family of ATPases found in prokaryotes as well as eukaryotes. Homologs are known as *mgsA* in *E. coli*, *mgs1* in yeast (maintenance of genome stability A/1), and *WRNIP1* (Werner interacting protein 1) in mammals. The encoded protein is involved in cellular responses to stalled or collapsed replication forks, likely by modulating replication restart [46–48].



The downstream *uspA* gene belongs to a superfamily that encompasses an ancient and highly conserved group of proteins that are widely distributed among bacteria, archaea, fungi, flies, and plants. It was found to be induced during metabolic, oxidative, and temperature stress in *Salmonella typhimurium* [49] and linked to cell sensitivity to ultraviolet light in *E. coli* [50]. *uspA* is known to be differentially expressed in response to a large number of different environmental stresses such as acid and salt stresses, starvation, exposure to heat, oxidants, metals, ethanol, antibiotics, and other stimulants - particularly within the genera *Lactobacillus*, *Streptococcus*, *Enterococcus* and *Lactococcus* [51–53].

### Structure and sequence conservation

The consensus structure and sequence conservation of 6S RNA in LAB based on a mLocARNA [43] alignment combined with RNAalifold [54] is illustrated in Fig. 3. Additional File 5 shows the consensus structures at the family level. The consensus of 6S RNA in LAB follows the well-known secondary structure of the canonical 6S RNA [15, 23], featuring an outer closing stem with smaller bulges and loops, a large 5'-central bulge and an apical stem with smaller internal loops capped by the terminal loop L1. Opposite to the 5'-central bulge a hairpin is predicted that was also shown to form in *B. subtilis* 6S-1 RNA [26]. The central bulge harbors the initiation site for



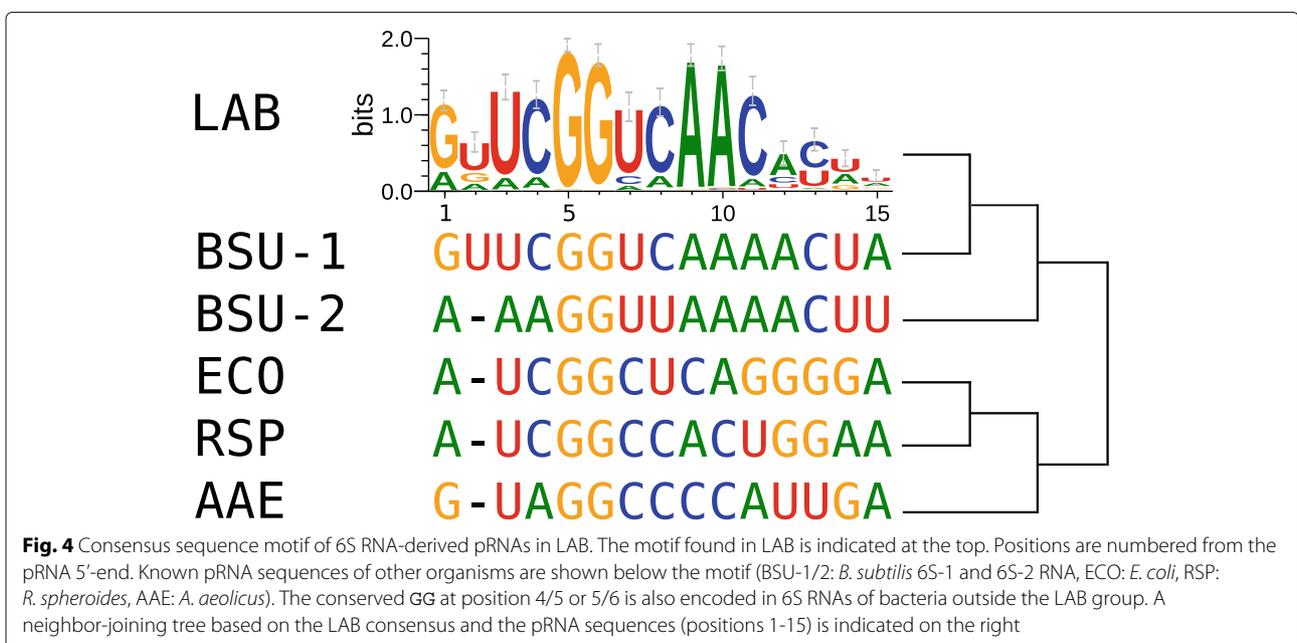
product RNA (pRNA) transcription. This consensus and canonical 6S secondary structure is evident in most of the 6S RNA groups: *Aerococcaceae*, *Aerococcus*, *Carnobacteriaceae*, *Vagococcus*, *Enterococcaceae*, *Pediococcus*, *Lactobacillus* 2, 3, 4, 6, 7, *Streptococcus*, and *Lactococcus*, see Additional File 5.

#### Product RNAs

Putative pRNA transcription start sites were inferred from a structural alignment (see Materials and Methods) of 172 representative 6S RNA sequences from LAB species and in relation to those of *E. coli*, *R. spheroides* and *B. subtilis* for which the start sites are experimentally proven. Fig. 4 shows the overall sequence motif.

The first eleven nucleotides of the pRNAs are well conserved. This conservation diminishes starting at position 12. GG at position 5/6 as well as AA at position 9/10 are the most conserved in this group. Two G residues are also conserved in experimentally verified pRNAs from more distantly related bacteria such as the Gram-negatives *E. coli*, *A. aeolicus* and *R. spheroides*, but in these cases at positions 4/5 (Fig. 4). Notably, a highly conserved adenine immediately upstream of the pRNA start sites was identified in the 6S RNAs of LAB species as well as in the reference 6S RNAs included in Fig. 4.

Based on the pRNA sequence (positions 1-15), LAB pRNAs are closely related to pRNAs synthesized from



*B. subtilis* 6S-1 RNA as template (Fig. 4). Although the 6S-1 pRNA sequence shows differences to the LAB pRNA consensus, major hallmarks (upstream adenine, GG dinucleotide, AA at position 9/10) are still present. Hence, despite the considerable phylogenetic distance, similarities to the pRNA sequence found in LAB are clearly recognizable.

We screened 115 publicly available RNA-Seq datasets for expression of 6S RNA and the presence of pRNAs. These small transcripts are usually depleted in sample preparation for RNA-Seq or neglected in data processing that typically focuses on longer RNAs such as tRNAs or mRNAs. Moreover, we found that pRNAs are underrepresented in adapter ligation libraries compared to poly(A)-tailing libraries [55]. It is thus not surprising that only small numbers of pRNA reads were identified in most RNA-Seq libraries. We yet found robust evidence for pRNAs in *Streptococcus pneumoniae* and *Streptococcus pyogenes* RNA-Seq data (Fig. 5), which also supports the predicted pRNA start site (Figs. 3 and 4) [56, 57]. Two pRNA transcripts were previously reported for *S. pyogenes*, but their sequences were not provided [33]. Here we confirm these findings. We find one alternative transcription start site (pRNA\*) located around position 136 that starts at the beginning of the L2 loop (see Fig. 3). The alternative pRNA transcript likely results from 6S RNA binding RNAP in inverse orientation. Similar observations have been made for *Helicobacter pylori* [19]. Notably, neither the pRNA nor the pRNA\* sequences have alternative matches in the respective genomes. It is thus unlikely that these transcripts derive from another locus. Additional File 6 illustrates further RNA-Seq results. While pRNAs were also found in libraries from *E. faecalis*, the number of reads is too low to draw safe conclusions.

#### CcpA-binding catabolite responsive elements

A functional *cre* site upstream of the 6S RNA promoter was reported in *L. lactis*, suggesting that 6S RNA expression is regulated depending on the available carbon source

[34]. An equivalent *cre* site could be found in about one-third of all LAB species. Fig. 6 illustrates the location and sequence conservation of the two *cre* sites at the 6S RNA locus. Additional File 2 shows a detailed overview of all species with *cre* sites in the 6S RNA region. Additional File 7 lists the respective motif sequences along with their positions and p-values. *cre* sites are most frequently found in *Enterococcaceae* but also in several *Streptococcaceae* and the *Lactobacillus* groups 6 and 7 (see Fig. 1). Mainly in *Streptococcaceae* and *Lactobacillus* group 6, potential *cre* sites were also identified within the 6S RNA coding sequence. Notably, *L. coryniformis*, *L. rennini*, *L. vaginalis*, *S. canis*, *S. didelphis*, *S. equi*, *S. pantholopis* and *S. phocae* do not have a strong, detectable *cre* site at the 6S RNA promoter but only within the 6S RNA coding region; both sites were detected in *L. backii*, *L. bifementans*, *S. castoreus*, *S. gallolyticus*, *S. halotolerans*, *S. ictaluri*, *S. iniae*, *S. parauberis* and *S. uberis*.

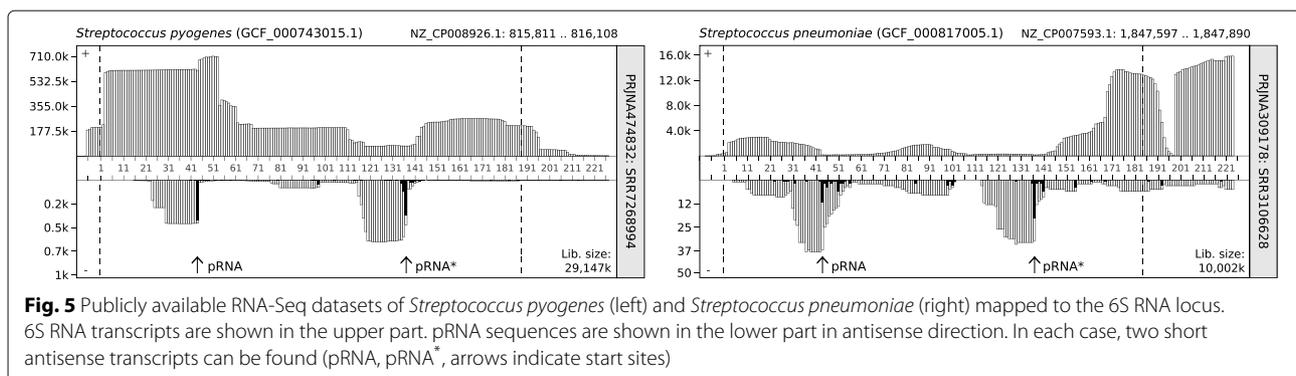
#### Expression and antisense transcripts

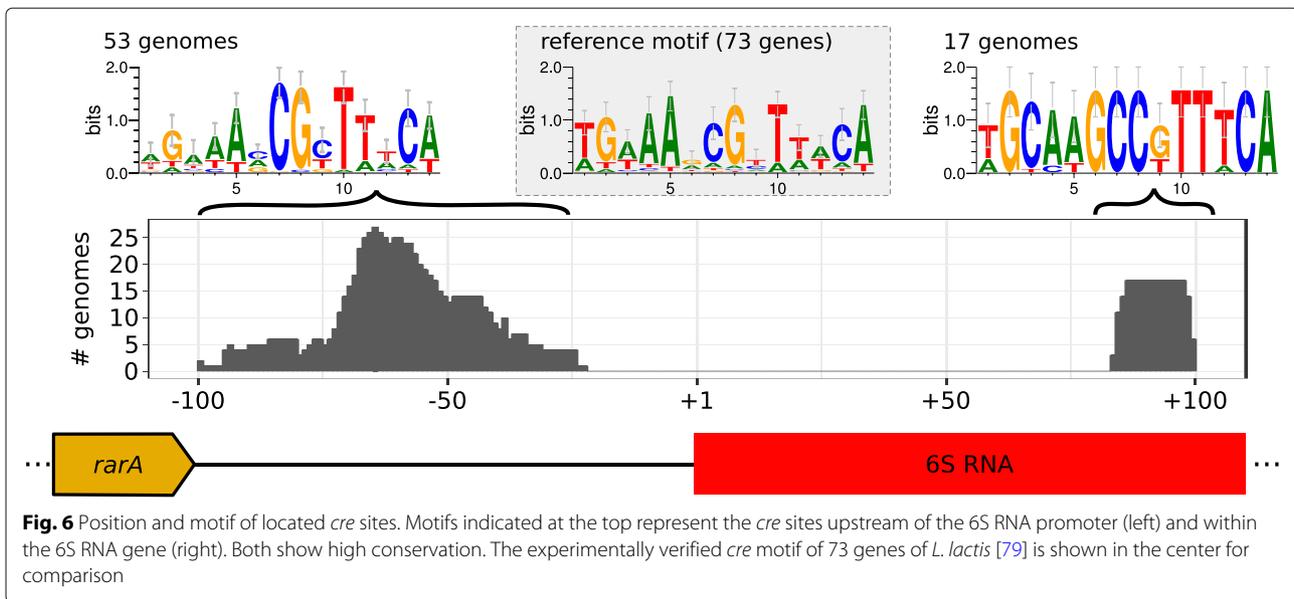
A total of 115 publicly available RNA-Seq libraries representing 24 different LAB genera were screened for the expression of 6S RNA, pRNAs and long antisense transcripts as described for the *Enterococcus faecalis* V583 strain [37]. Detailed results for each library are shown in Additional File 6.

6S RNA transcripts were highly abundant in general (usually 1-2% of all reads in the RNA-Seq libraries), indicating active transcription in LAB grown under a wide variety of culture conditions and stresses. In line with previous findings [37], however, we did not find evidence for long antisense transcripts of 6S RNA in any RNA-Seq library including those from other *Enterococcus faecalis* strains (OG1RF, 12030, and ATCC 29212), indicating that such transcripts are not a common trait among LAB.

#### Discussion

Here we identified the 6S RNA gene at a well-conserved genomic locus in LAB species that distinguishes this bacterial group from related bacterial clades. While the





consensus secondary structure is typically canonical as described for *B. subtilis* 6S-1 RNA, we could not verify this for candidates of the genus *Weissella*. Nevertheless, we identified evidence for significant transcription of the respective loci in publicly available RNA-Seq libraries for two strains, see Additional File 8. This confirms a weak 6S RNA candidate in *W. koreensis*. Although no relevant match was found for *W. confusa*, the intergenic region downstream of the syntenically conserved *rarA* showed transcription that matched a 6S RNA transcript even though its putative secondary structure did not match a canonical 6S RNA. A TATAAT sequence is present at the -10 region of all candidates reported for *Weissella*, indicating the presence of a promoter. Similarly, a rho-independent terminator was predicted at the RNA's proposed 3'-end. Thus, the presence of an actively transcribed 6S RNA-like transcript can be assumed. It will be interesting to investigate the functional consequences of this structural alteration.

Carbon catabolite control is a major regulatory mechanism for the modulation of metabolic activity of microorganisms to optimize carbon metabolism and energy use. It involves both carbon catabolite repression and activation. In most low-GC-content Gram-positive bacteria this regulation is mediated by the catabolite control protein A (CcpA) that binds to DNA at *cis*-acting sequences. These are called catabolite responsive elements (*cre*) and are located either in the promoter region or within the coding sequence of the regulated gene [36]. CcpA can function as an activator or may repress transcription depending on its location within a regulated gene or operon [58]. We found strong evidence for *cre* sites upstream of the 6S RNA promoter in about a third of all LAB species, mainly in

*Enterococcaceae* but also in *Streptococcaceae* and some *Lactobacillus* subgroups. For *Streptococcaceae* and *Enterococcaceae*, the presence and regulatory importance of these *cre* sites has been reported and studied previously [59, 60]. On the basis of previous reports, our findings suggest that 6S RNA expression is under the negative control of CcpA in many LAB species. This was shown e.g. for *L. lactis* where 6S RNA is 3-fold upregulated upon deletion of the *ccpA* gene [34].

For several 6S RNA genes, *cre* sites were also identified internally - in some cases in addition to the site at the 6S RNA promoter (see Additional File 2). The presence of two *cre* sites regulating the expression of *cid* and *lrg* genes in *Streptococcus mutans* has already been described, but in this case both sequences were upstream of the transcription start site of the above-mentioned genes [61]. In *B. subtilis*, *cre* sites upstream of promoters were found to be primarily activated by CcpA, while *cre* sites overlapping promoters had repressing effects [35]. As the *cre* sites in LAB overlap the -35 region of 6S RNA gene promoters (Fig. 6), CcpA-binding is likely inhibitory; *cre* sites located further downstream of the transcription start site may act as roadblocks or repress initiation of transcription through interaction with RNAP [62]. Future studies may address the interplay of the two *cre* sites at/within the 6S RNA gene. Although speculative at present, it is also a possibility that CcpA binds to 6S RNA at the internal *cre* site, taking into account that 6S RNAs mimic an open DNA promoter [22].

The identified *cre* sequences share a high degree of similarity to the consensus sequences previously described for other LAB such as *L. lactis*, (see Fig. 6) as well as to other Gram-positive bacteria such as *B. subtilis* [36, 63].

Recent studies on the promoter region of the *PTS-IIC* gene cluster of *L. lactis* demonstrated the importance of nucleotide identity at positions 7 and 12 of the 14-nt long *cre* site. Specific mutations within the -35 promoter element resulted in constitutive expression of the downstream gene in the presence of glucose, while other mutations enhanced promoter activity in the presence of cellobiose [63].

The prediction of transcription start sites for pRNAs was based on the structural alignment to other 6S RNAs and could be verified by RNA-Seq data in two cases. This study is the first that deduces pRNAs for a large taxonomic group covering multiple families. We found a highly conserved sequence up to around position 11. This may point to similar kinetics of pRNA synthesis and pRNA-induced 6S RNA refolding [26]. Strikingly, GG at positions 5/6 or 4/5 of the pRNAs appears to be a key feature conserved beyond LAB.

A general property of the 6S RNA locus in LAB is its location between the *rara* and *uspA* genes. Gene order conservation can be used not only to evaluate the orthology of genomic regions but might also hint at functional relationships between genes [64]. RarA is proposed to act at stalled DNA replication forks upon DNA damage and UspA alters the expression of a variety of genes that help to cope with stresses. As 6S RNA was shown to have a role in cellular stress responses to ensure long-time cell survival, all three gene products might be part of an overarching stress response network. The *rara* gene is in close vicinity to the 6S RNA locus across all families including the 6S-1 RNA locus of the non-LAB firmicute *B. subtilis* (see Additional File 4). In the latter, however, *rara* is encoded in the opposite direction and known to be monocistronic [65]. The RNA-Seq data presented in Additional Files 6 and 8 and the presence of a downstream terminator in most species indicates that the 6S RNA gene is monocistronic as well. However, several *Streptococcaceae* members encode a tRNA-Lys immediately downstream of 6S RNA, suggesting that both genes are part of the same operon. This assumption is supported by RNA-Seq data for *S. pneumoniae* (Additional File 6, p. 43) showing that both ncRNAs have the same transcript level [56]. Thus, both RNAs are likely processing products of the same primary transcript. Other notable syntenic bonds are not universally preserved for LAB but within and also across particular LAB families. Examples are the acetate kinase, class I SAM-dependent methyltransferase, 16S rRNA methyltransferase, and the 50S ribosomal protein L11 methyltransferase. While the function of the other frequently linked genes is unknown so far, this data suggests a cluster of growth-relevant and stress-related genes that 6S RNA is part of. Typically, these genes appear to be transcribed independently (with the exception of 6S RNA and tRNA-Lys in a number of *Streptococcaceae*).

Therefore, the possibility of a common functional context remains vague at present.

## Conclusions

Lactic acid bacteria include highly heterogeneous species and the study of the role of non-coding RNA molecules, particularly 6S RNA, in the regulation of the response of these bacteria to different stress conditions has many potential applications, both within industrial and health contexts. The global transcription regulator 6S RNA is present in nearly all species and well-conserved throughout this group. It generally resembles the canonical form that is well described for *B. subtilis* 6S-1 RNA. LAB 6S RNAs also share the syntenic proximity to *rara*, located upstream of 6S RNA in nearly all LAB genomes. Many species additionally encode the UspA protein downstream of 6S RNA, which makes its identification comparably easy. The experimental evidence that was processed and analyzed in this study also demonstrated that 6S RNA is expressed in a multitude of LAB species across all taxonomic families and under varying culture conditions. This also highlights the important regulatory role of this ncRNA in bacterial metabolism, further supported by the frequent presence of *cre* sites in its promoter and coding region. The conservation of 6S RNAs makes it plausible to generally apply our findings to any LAB species in order to explore its biotechnological potential.

## Methods

### Genomes

Several thousand genomes representing 576 species that cover 48 genera were listed as part of the *Lactobacillales* order according to the NCBI taxonomy classification (date of retrieval 10/09/2018) [38]. In order to work with a reasonably representative set, we focused on the genomes with the best respective assembly status for each species. The species *Enterococcus faecium* for example comprises 1109 genomes/subspecies. Fifty-one out of these are marked as “Complete Genome” and were thus considered in the present work. *Lactobacillus fuchuensis* is represented with three genomes out of which the most complete assembly is marked as “Chromosome” that was thus considered, and so on. Additionally, we added 13 strains that were characterized by our institute (CERELA-CONICET) even though they did not meet this criterion. Species with yet unclear specific names (sp.) were neglected. A total of 1,092 genomes were considered in this study. An overview of the genera analyzed here can be found in Table 1. A detailed list of the species and genome assembly levels is provided in Additional File 1. The respective genomes and genomic annotations were downloaded via <ftp.ncbi.nlm.nih.gov> from the NCBI database [38].

### 6S RNA prediction

Putative 6S RNAs encoded in LAB genomes were identified in multiple steps. A BLAST-based approach was performed using available 6S RNA annotations given in the NCBI RefSeq annotation, from Wehner *et al.*, and from the Rfam seed sequences for the 6S/SsrS RNA family (RF00013, Version 14) to cover the currently known 6S RNAs [16, 66, 67]. An e-value threshold of  $10^{-30}$  was

applied. Previously not annotated 6S RNAs were identified with a covariance-based search performed with INFERNAL (v1.1.1) [68] using the “6S/SsrS RNA” family model as query (see above). Initially, no thresholds were set. Based on the assumption that each genome should encode at least one 6S RNA gene, the highest-scoring hit for each genome was assumed as a true hit. Compared to this, the e-values of the second-best hits

**Table 1** Genomes overview

| Family              | Genus                      | Genomes used / Genomes available |
|---------------------|----------------------------|----------------------------------|
| Aerococcaceae       | <i>Abiotrophia</i>         | 1 / 2                            |
|                     | <i>Aerococcus</i>          | 8 / 61                           |
|                     | <i>Dolosicoccus</i>        | 2 / 3                            |
|                     | <i>Eremococcus</i>         | 1 / 2                            |
|                     | <i>Facklamia</i>           | 3 / 9                            |
|                     | <i>Globicatella</i>        | 1 / 4                            |
| Carnobacteriaceae   | <i>Agitococcus</i>         | 1 / 1                            |
|                     | <i>Alkalibacterium</i>     | 1 / 8                            |
|                     | <i>Allofustis</i>          | 1 / 1                            |
|                     | <i>Atopobacter</i>         | 1 / 1                            |
|                     | <i>Atopococcus</i>         | 1 / 1                            |
|                     | <i>Carnobacterium</i>      | 9 / 41                           |
|                     | <i>Dolosigranulum</i>      | 10 / 12                          |
|                     | <i>Granulicatella</i>      | 1 / 7                            |
|                     | <i>Jeotgalibaca</i>        | 1 / 4                            |
|                     | <i>Lactigenium</i>         | 1 / 1                            |
|                     | <i>Marinilactibacillus</i> | 1 / 5                            |
| <i>Trichococcus</i> | 7 / 15                     |                                  |
| Enterococcaceae     | <i>Bavariiococcus</i>      | 1 / 1                            |
|                     | <i>Enterococcus</i>        | 114 / 2105                       |
|                     | <i>Melissococcus</i>       | 2 / 14                           |
|                     | <i>Tetragenococcus</i>     | 5 / 19                           |
|                     | <i>Vagococcus</i>          | 4 / 6                            |
| Lactobacillaceae    | <i>Lactobacillus</i>       | 460 / 1680                       |
|                     | <i>Pediococcus</i>         | 25 / 61                          |
|                     | <i>Sharpea</i>             | 1 / 4                            |
| Leuconostocaceae    | <i>Convivina</i>           | 1 / 1                            |
|                     | <i>Fructobacillus</i>      | 5 / 9                            |
|                     | <i>Leuconostoc</i>         | 23 / 118                         |
|                     | <i>Oenococcus</i>          | 3 / 208                          |
|                     | <i>Weissella</i>           | 23 / 43                          |
| Streptococcaceae    | <i>Floricoccus</i>         | 2 / 2                            |
|                     | <i>Lactococcus</i>         | 44 / 168                         |
|                     | <i>Streptococcus</i>       | 328 / 12076                      |

Distribution and number of genomes that were retrieved and downloaded from the NCBI database according to the “most complete genome” criterion

were worse by orders of magnitude. A manual inspection on a sample basis confirmed that those were not likely to be valid 6S RNA candidates. Hence, an e-value threshold of  $10^{-8}$  was applied. In this case, a primary hit was found in most species while unexpected secondary hits were rare and could be judged manually in later stages. Overlapping hits were joined. Hits were found in 973 out of 1092 genomes. Redundant sequences were merged to a single representative sequence resulting in 330 unique sequences that were aligned using Clustal Omega (v1.2.1) [69]. Sequences with an edit distance of ten or less were merged to their consensus sequence to further reduce the amount of redundancy. 188 representative 6S RNA sequences remained. We checked for isolated sequences in the secondary structure clustering analysis (see below) and non-canonical secondary structures using RNAfold (v2.1.9) [54] as well as suspicious alignments to further remove non-canonical and doubtful hits. The following sixteen 6S RNA candidates were discarded manually in the first round: *Agitococcus lubricus*, *Lactococcus fujiensis*, *Facklamia hominis*, *Pediococcus damnosus*, *Lactobacillus babusae*, *Pediococcus cellicola*, *Lactobacillus cacaonum*, *Lactobacillus mucosae*, *Lactobacillus coleohominis*, *Lactobacillus gastricus*, *Lactobacillus equigenerosi*, *Lactobacillus malefermentans*, *Lactobacillus oryzae*, *Oenococcus oeni*, *Weissella kandleri*, and *Weissella koreensis*. In total 172 representative 6S RNA sequences covering 947 genomes remained. This set was used for further analyses.

For each genome without an annotated canonical 6S RNA (including those discarded manually in the first round), a second search iteration was performed with a LAB-specialized covariance model that was built based on all canonical 6S RNAs identified before. The e-value threshold was reduced to 0.1 and all search heuristics were turned off (`cmsearch -max`). In addition, the correct genomic locus was ensured by only allowing hits within 2000 nt from *upsA* and/or *rarA* homologs. Both are typically encoded in close vicinity to 6S RNA gene (see Results section “Synteny”). The homologs were annotated using BLAST (v2.8.1+) [66] with an e-value of  $10^{-40}$  based on the sequences found in the synteny analysis. In this way, additional syntenically supported 6S RNA candidate genes were identified in 54 genomes. These are marked as “2nd-iteration” in Additional File 1 that lists all 6S RNAs annotated for LAB.

#### Prediction of rho-independent terminators

Terminators were predicted using TransTermHP (v2.09) [70]. An adaptive threshold was used to ascertain significant predictions. Each genome was shuffled ten times while preserving its mono- and di-symbol composition. We then compared the number of hits above any given threshold between the shuffled genomes and original

genome. The threshold was chosen such that the average number of hits in the shuffled genomes was no more than 5% compared to the hits in the original genome. E.g. if we find 100 hits above a score of 90 in the genome, the average number of hits in the shuffled genomes above the same score cannot exceed 5, otherwise a higher threshold is chosen. In the absence of significance values provided by the prediction tool, this method roughly estimates a p-value threshold of 0.05 for terminator hits. Overlapping hits were merged. In addition, RNIE (v0.01) was used with default parameters for a genome-wide prediction [71]. For the relevant regions, the results were a subset of the former predictions.

#### Consensus secondary structure

All representative 6S RNA candidates were aligned using mLocARNA (v2.0.0RC8), a local structural alignment algorithm for RNA secondary structures [43]. To locate the putative start sites for pRNAs in LAB, three well-studied 6S RNA instances were added as references from which the start sites were then projected to the LAB 6S RNAs. Namely *Escherichia coli* K12 (GCF\_000005845.2) and *Bacillus subtilis* 168 (GCF\_000009045.1), which codes for two paralogs, 6S-1 and 6S-2 RNA (also known as BsrA and BsrB) [39, 72]. The consensus secondary structure was then calculated with RNAalifold (v2.4.13) [54] and visualized using VARNA (v3.93) [73], excluding the folding references.

#### Prediction of pRNAs

The transcription start of 6S RNA-derived pRNAs was determined based on the structural alignment mentioned above. Based on previously characterized transcription start sites in other bacteria [26, 55, 74], we assumed the equivalent positions within LAB 6S RNAs. The putative pRNA sequences of 16 nt length were aligned with Clustal Omega (v1.2.1) [69]. We found a strong consensus sequence motif (see Results) that we used to further adjust the pRNA start site by shifting it for up to three nucleotides in case of suboptimal matches. The motif composition was calculated using WebLogo (v2.8.2) [75].

#### Phylogeny with secondary structure clusters

The sequences of the 6S RNA candidates identified in the first round were clustered hierarchically based on their structured RNA motifs using RNAclust [42]. This approach combines the base pair probability matrix of the secondary structure distributions (via RNAfold (v2.1.9) [54]) and a sequence-structure alignment based on LocARNA [43]. *Bacillus subtilis* 168 (GCF\_000009045.1) 6S-1 RNA (BsrB) was added as an outgroup [39]. The resulting tree can be found in Additional File 2, while a condensed version is shown in Fig. 1, visualized using Evolview (v3) [76].

### 16S rRNA phylogeny

16S rRNA sequences were identified using BLAST (v2.8.1+) [66] with an e-value of  $10^{-20}$  based on the 16S rRNA reference sequences provided by the NCBI database [38]. Redundant sequences were merged. Sequences were aligned using *muscle* (v3.8.1551) [77]. The 5'- and 3'-end of the 16S rRNA alignment were trimmed such that < 25% of all sequences had remaining gaps in these regions. The phylogenetic reconstruction was performed with *RAxML* (v8.1.20) [78] using the General Time Reversible model (GTR) with optimization of substitution rates and the GAMMA model of rate heterogeneity and 1000 bootstrap iterations. The phylogenetic reconstruction was visualized using *Evolview* (v3) [76].

### Synteny

The amino acid sequences of ten protein-coding genes 5000 nt up- and downstream of the predicted 6S RNA locus were fetched from the NCBI database. Orthologous groups were predicted with *Proteinortho* (v6.13) [45]. To avoid an overrepresentation bias, equivalent and similar 6S RNA sequences were represented by a single reference strain rather than all strains of the respective species (see “Detection of 6S RNAs”). Genes found in fewer than 50% of each family were omitted from the analysis. For each LAB family, one species that best represented the genomic context of all family members was chosen.

### CcpA-binding catabolite responsive elements

The sequence motif for *cre* sites was derived from experimental *B. subtilis* data [36] that also fits previously derived *L. lactis* data [79] as shown in Fig. 6. However, we preferred the former as it yields a higher number of underlying sequences, which strengthens the derived p-values for motif matches and thus avoids false positive predictions. The 6S RNA sequences along with their 100 nt upstream regions were used to find sequences matching the *cre* motif using *MAST* [80]. Typically, this position overlapped with the 3'-end of the *rarA* gene. Hence, we did not expect binding sites further upstream to be relevant to 6S RNA. We used the dinucleotide distribution of the respective genomes as background for each e-value calculation. The default e-value threshold of 10 and p-value threshold of  $10^{-5}$  was applied. The resulting motifs were separated in two groups: Upstream of the 6S RNA promoter and within the 6S RNA coding region as shown in Fig. 6.

### Expression

Available RNA-Seq datasets for LAB were located in the NCBI SRA archive and downloaded on 12-11-2018 [38]. In total 115 RNA-Seq libraries were analyzed

covering 24 different LAB species. Read sequences were extracted using the NCBI-provided *fastq-dump* (v2.8.2). Adapter removal and read trimming was performed using *cutadapt* (v1.12) [81] followed by a quality control with *fastqc* (v0.11.5) [82]. Processed reads were mapped to the respective genomes with *segemehl* (v0.2.0) [83]. An e-value threshold of 0.0001 was applied. The mapped data was visualized for each 6S RNA locus using custom scripts. Additional File 6 shows all results and data sources in detail.

### Abbreviations

GRAS: Generally Recognized as Safe; LAB: Lactic acid bacteria; RNA: ribonucleic acid; RNAP: DNA-dependent RNA polymerase complex; *cre* site: *ccpA*-binding catabolite responsive element

### Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12863-021-00983-2>.

**Additional file 1:** List of genomes and 6S RNAs (xls). List of LAB genomes used in this study including tax annotation, assembly status, location of the predicted 6S RNA.

**Additional file 2:** Full 6S RNA phylogeny (pdf). Sequence- and structure-based reconstruction of 6S RNA phylogeny in LAB including the annotation of species with located *cre* sites. Full taxonomic resolution of Fig. 1.

**Additional file 3:** 16S rRNA phylogeny (pdf). Phylogenetic reconstruction of LAB 16S rRNA.

**Additional file 4:** Full genomic context of 6S RNA in LAB (pdf). Full genomic context of 6S RNA in LAB. Full taxonomic resolution of Fig. 2.

**Additional file 5:** 6S RNA grouped consensus alignment (pdf). Folded consensus structure of the 6S RNA groups analogous to Fig. 3.

**Additional file 6:** RNA-Seq results (pdf). Visualization of RNA-Seq libraries mapped to the respective 6S RNA loci.

**Additional file 7:** Predicted *cre* site motifs (xls). Predicted *cre* site sequences and positions relative to the 6S RNA start site.

**Additional file 8:** 6S RNA evidence in *Weissella* (pdf). RNA-Seq data, genomic context and sequences of putative 6S RNA loci in *Weissella*.

### Acknowledgments

We thank Florian Taube for implementing the RNA-Seq visualizations.

### Author's contributions

ML conceived the study. PGC, PK, MT and ML carried out the bioinformatic analyses. PGC, PK, RKH, LS, EMH and ML wrote the manuscript. All authors read and approved the final manuscript.

### Funding

This work was supported by Deutscher Akademischer Austauschdienst (Short-Term Research Grant) and Deutsche Forschungsgemeinschaft (RTG 2355). Open Access funding enabled and organized by Projekt DEAL.

### Availability of data and materials

Genomes of lactic acid bacteria were downloaded from <ftp://ftp.ncbi.nlm.nih.gov/genomes/> (2019-08-07). Species names, chromosome and tax ids, fasta paths and annotated 6S RNAs are provided in Additional File 1. RNA-Seq data was retrieved from NCBI SRA at <https://www.ncbi.nlm.nih.gov/sra>. Bioproject and SRA ids are listed in Additional Files 6 and 8. The 6S/SsrS RNA family seed sequences RF00013 provided by RFAM at <http://rfam.xfam.org/family/RF00013> (Version 14) and the sequences by Wehner et al. provided at <https://www.rna.uni-jena.de/supplements/6SRNA/> (2019-08-07) were used for initial 6S RNA prediction.

## Declarations

### Ethics approval and consent to participate

Not applicable.

### Consent for publication

Not applicable.

### Competing interests

The authors declare that they have no competing interests.

### Author details

<sup>1</sup>Centro de Referencia para Lactobacilos (CERELA-CONICET), Chacabuco 145, 4000 San Miguel de Tucumán, Argentina. <sup>2</sup>Philipps-Universität Marburg, Institut für Pharmazeutische Chemie, Marbacher Weg 6, 35032 Marburg, Germany. <sup>3</sup>Philipps-Universität Marburg, Center for Synthetic Microbiology (Synmikro), Hans-Meerwein-Straße 6, 35043 Marburg, Germany.

Received: 21 June 2021 Accepted: 12 August 2021

Published online: 03 September 2021

## References

- Duar RM, Lin XB, Zheng J, Martino ME, Grenier T, Pérez-Muñoz ME, Leulier F, Gänzle M, Walter J. Lifestyles in transition: evolution and natural history of the genus *Lactobacillus*. *FEMS Microbiol Rev.* 2017;41(Supp\_1): 27–48. <https://doi.org/10.1093/femsre/fux030>.
- Zheng J, Wittouck S, Salvetti E, Franz CMAP, Harris HMB, Mattarelli P, O'Toole PW, Pot B, Vandamme P, Walter J, Watanabe K, Wuys S, Felis GE, Gänzle MG, Lebeer S. A taxonomic note on the genus *Lactobacillus*: Description of 23 novel genera, emended description of the genus *Lactobacillus* Beijerinck 1901, and union of *Lactobacillaceae* and *Leuconostocaceae*. *Int J Syst Evol Microbiol.* 2020;70(4):2782–858. <https://doi.org/10.1099/ijsem.0.004107>.
- Salveti E, Harris HMB, Felis GE, O extquoterightToole PW. Comparative genomics of the genus *Lactobacillus* reveals robust phylogroups that provide the basis for reclassification. *Appl Environ Microbiol.* 2018;84(17): <https://doi.org/10.1128/AEM.00993-18>. <https://aem.asm.org/content/84/17/e00993-18.full.pdf>.
- Leroy F, De Vuyst L. Lactic acid bacteria as functional starter cultures for the food fermentation industry. *Trends Food Sci Technol.* 2004;15(2): 67–78. <https://doi.org/10.1016/j.tifs.2003.09.004>.
- Hill C, Guarner F, Reid G, Gibson GR, Merenstein DJ, Pot B, Morelli L, Canani RB, Flint HJ, Salminen S, Calder PC, Sanders ME. The International Scientific Association for Probiotics and Prebiotics consensus statement on the scope and appropriate use of the term probiotic. *Nat Rev Gastroenterol Hepatol.* 2014;11(8):506–514. <https://doi.org/10.1038/nrgastro.2014.66>.
- Mattila-Sandholm T, Mättö J, Saarela M. Lactic acid bacteria with health claims—interactions and interference with gastrointestinal flora. *Int Dairy J.* 1999;9(1):25–35. [https://doi.org/10.1016/S0958-6946\(99\)00041-2](https://doi.org/10.1016/S0958-6946(99)00041-2).
- Smid EJ, Hugenholtz J. Functional genomics for food fermentation processes. *Ann Rev Food Sci Technol.* 2010;1:497–519. <https://doi.org/10.1146/annurev.food.102308.124143>.
- Zhang Y, Li Y. Engineering the antioxidative properties of lactic acid bacteria for improving its robustness. *Curr Opin Biotechnol.* 2013;24(2): 142–7. <https://doi.org/10.1016/j.copbio.2012.08.013>.
- Papadimitriou K, Alegría Á, Bron PA, de Angelis M, Gobetti M, Kleerebezem M, Lemos JA, Linares DM, Ross P, Stanton C, Turroni F, van Sinderen D, Varmanen P, Ventura M, Zúñiga M, Tsakalidou E, Kok J. Stress physiology of lactic acid bacteria. *Microbiol Mol Biol Rev.* 2016;80(3):837–90. <https://doi.org/10.1128/MMBR.00076-15>. <https://mmb.asm.org/content/80/3/837.full.pdf>.
- Gottesman S. Micros for microbes: non-coding regulatory RNAs in bacteria. *Trends Genet.* 2005;21(7):399–404. <https://doi.org/10.1016/j.tig.2005.05.008>.
- Holmqvist E, Wagner EGH. Impact of bacterial sRNAs in stress responses. *Biochem Soc Trans.* 2017;45(6):1203–12. <https://doi.org/10.1042/BST20160363>.
- Kok J, van Gijtenbeek LA, de Jong A, van der Meulen SB, Solopova A, Kuipers OP. The evolution of gene regulation research in *Lactococcus lactis*. *FEMS Microbiol Rev.* 2017;41(Supp\_1):220–43. <https://doi.org/10.1093/femsre/fux028>.
- Wassarman KM, Storz G. 6S RNA regulates E. coli RNA polymerase activity. *Cell.* 2000;101(6):613–23.
- Hsu L, Zagorski J, Wang Z, Fournier M. *Escherichia coli* 6S RNA gene is part of a dual-function transcription unit. *J Bacteriol.* 1985;161(3):1162–70.
- Barrick JE, Sudarsan N, Weinberg Z, Ruzzo WL, Breaker RR. 6S RNA is a widespread regulator of eubacterial RNA polymerase that resembles an open promoter. *RNA (New York, N.Y.)* 2005;11(5):774–84. <https://doi.org/10.1261/rna.7286705>.
- Wehner S, Damm K, Hartmann RK, Marz M. Dissemination of 6S RNA among bacteria. *RNA Biol.* 2014;11(11):1467–78. <https://doi.org/10.4161/rna.29894>.
- Lechner M, Nickel AI, Wehner S, Riege K, Wieseke N, Beckmann BM, Hartmann RK, Marz M. Genomewide comparison and novel ncRNAs of aquificales. *BMC Genom.* 2014;15:522. <https://doi.org/10.1186/1471-2164-15-522>.
- Elkina D, Weber L, Lechner M, Burenina O, Weisert A, Kubareva E, Hartmann RK, Klug G. 6S RNA in *Rhodobacter sphaeroides*: 6S RNA and pRNA transcript levels peak in late exponential phase and gene deletion causes a high salt stress phenotype. *RNA Biol.* 2017;14(11):1627–37. <https://doi.org/10.1080/15476286.2017.1342933>.
- Sharma CM, Hoffmann S, Darfeuille F, Reignier J, Findeiss S, Sittka A, Chabas S, Reiche K, Hackermüller J, Reinhardt R, Stadler PF, Vogel J. The primary transcriptome of the major human pathogen *Helicobacter pylori*. *Nature.* 2010;464(7286):250–5. <https://doi.org/10.1038/nature08756>.
- Wassarman KM. 6S RNA: a small RNA regulator of transcription. *Curr Opin Microbiol.* 2007;10(2):164–8. <https://doi.org/10.1016/j.mib.2007.03.008>. *Cell regulation (RNA special issue)*.
- Steuten B, Hoch PG, Damm K, Schneider S, Köhler K, Wagner R, Hartmann RK. Regulation of transcription by 6S RNAs. *RNA Biol.* 2014;11(5):508–21. <https://doi.org/10.4161/rna.28827>.
- Chen J, Wassarman KM, Feng S, Leon K, Feklistov A, Winkelman JT, Li Z, Walz T, Campbell EA, Darst SA. 6S RNA mimics b-form dna to regulate *Escherichia coli* RNA polymerase. *Mol Cell.* 2017;68(2):388–397. <https://doi.org/10.1016/j.molcel.2017.09.006>.
- Wassarman KM. 6S RNA, a global regulator of transcription. *Microbiol Spectr.* 2018;6(3):. <https://doi.org/10.1128/microbiolspec.RWR-0019-2018>.
- Cavanagh AT, Klocko AD, Liu X, Wassarman KM. Promoter specificity for 6S RNA regulation of transcription is determined by core promoter sequences and competition for region 4.2 of sigma70. *Mol Microbiol.* 2008;67(6):1242–56. <https://doi.org/10.1111/j.1365-2958.2008.06117.x>.
- Steuten B, Setny P, Zacharias M, Wagner R. Mapping the spatial neighborhood of the regulatory 6S RNA bound to *Escherichia coli* RNA polymerase holoenzyme. *J Mol Biol.* 2013;425(19):3649–61. <https://doi.org/10.1016/j.jmb.2013.07.008>.
- Beckmann BM, Hoch PG, Marz M, Willkomm DK, Salas M, Hartmann RK. A pRNA-induced structural rearrangement triggers 6S-1 RNA release from RNA polymerase in *Bacillus subtilis*. *EMBO J.* 2012;31(7):1727–38. <https://doi.org/10.1038/emboj.2012.23>.
- Panchapakesan SSS, Unrau P.J. *E. coli* 6S RNA release from RNA polymerase requires  $\sigma 70$  ejection by scrunching and is orchestrated by a conserved RNA hairpin. *RNA (New York, N.Y.)* 2012;18(12):2251–9. <https://doi.org/10.1261/rna.034785.112>.
- Willkomm DK, Hartmann RK. 6S RNA - an ancient regulator of bacterial RNA polymerase rediscovered. *Biol Chem.* 2005;386(12):1273–77. <https://doi.org/10.1515/BC.2005.144>.
- Cavanagh AT, Sperger JM, Wassarman KM. Regulation of 6S RNA by pRNA synthesis is required for efficient recovery from stationary phase in *E. coli* and *B. subtilis*. *Nucleic Acids Res.* 2012;40(5):2234–46.
- Beckmann BM, Burenina OY, Hoch PG, Kubareva EA, Sharma CM, Hartmann RK. In vivo and in vitro analysis of 6S RNA-templated short transcripts in *Bacillus subtilis*. *RNA Biol.* 2011;8(5):839–49.
- Cavanagh AT, Wassarman KM. 6S-1 RNA function leads to a delay in sporulation in *Bacillus subtilis*. *J Bacteriol.* 2013;195(9):2079–86.
- Zheng H, Liu E, Shi T, Ye L, Konno T, Oda M, Ji Z-S. Strand-specific RNA-seq analysis of the *Lactobacillus delbrueckii subsp. bulgaricus* transcriptome. *Mol bioSyst.* 2016;12(2):508–19. <https://doi.org/10.1039/c5mb00547g>.
- Le Rhun A, Beer YY, Reimegård J, Chylinski K, Charpentier E. RNA sequencing uncovers antisense RNAs and novel small RNAs in *Streptococcus pyogenes*. *RNA Biol.* 2016;13(2):177–95.

34. van der Meulen SB, de Jong A, Kok J. Transcriptome landscape of *Lactococcus lactis* reveals many novel RNAs including a small regulatory RNA involved in carbon uptake and metabolism. *RNA Biol.* 2016;13(3):353–66. <https://doi.org/10.1080/15476286.2016.1146855>.
35. Schumacher MA, Sprehe M, Bartholomae M, Hillen W, Brennan RG. Structures of carbon catabolite protein a–(hpr-ser46-p) bound to diverse catabolite response element sites reveal the basis for high-affinity binding to degenerate dna operators. *Nucleic Acids Res.* 2011;39(7):2931–42.
36. Marciniak BC, Pabjaniak M, de Jong A, Duhring R, Seidel G, Hillen W, Kuipers OP. High- and low-affinity cre boxes for ccpa binding in *Bacillus subtilis* revealed by genome-wide analysis. *BMC Genom.* 2012;13(1):401. <https://doi.org/10.1186/1471-2164-13-401>.
37. Fouquier d'Hérouel A, Wessner F, Halpern D, Ly-Vu J, Kennedy SP, Serror P, Aurell E, Repoila F. A simple and efficient method to search for selected primary transcripts: non-coding and antisense RNAs in the human pathogen *Enterococcus faecalis*. *Nucleic Acids Res.* 2011;39(7):46. <https://doi.org/10.1093/nar/gkr012>.
38. Sayers EW, Barrett T, Benson DA, Bryant SH, Canese K, Chetvernin V, Church DM, DiCuccio M, Edgar R, Federhen S, Feolo M, Geer LY, Helmberg W, Kapustin Y, Landsman D, Lipman DJ, Madden TL, Maglott DR, Miller V, Mizrachi I, Ostell J, Pruitt KD, Schuler GD, Sequeira E, Sherry ST, Shumway M, Sirotkin K, Souvorov A, Starchenko G, Tatusova TA, Wagner L, Yaschenko E, Ye J. Database resources of the national center for biotechnology information. *Nucleic Acids Res.* 2009;37(Database issue):5–15. <https://doi.org/10.1093/nar/gkn741>.
39. Burenina OY, Hoch PG, Damm K, Salas M, Zatsepin TS, Lechner M, Oretskaya TS, Kubareva EA, Hartmann RK. Mechanistic comparison of *Bacillus subtilis* 6S-1 and 6S-2 RNAs—commonalities and differences. *RNA (New York, N.Y.)* 2014;20(3):348–59. <https://doi.org/10.1261/ma.042077.113>.
40. Qi J, Zhang D, Wang S, Huang L, Xia L, Dong W, Zheng Q, Liu Q, Xiao J, Xu Z. Transcriptome analysis of xylo-oligosaccharides utilization systems in *Weissella confusa* xu1. *AMS.* 2020;60(5):912–23.
41. Jeong SE, Chun BH, Kim KH, Park D, Roh SW, Lee SH, Jeon CO. Genomic and metatranscriptomic analyses of *Weissella koreensis* reveal its metabolic and fermentative features during kimchi fermentation. *Food Microbiol.* 2018;76:1–10.
42. Engelhardt J, Heyne S, Will S, Reiche R. RNAclust: A Tool for Clustering of RNAs Based on Their Secondary Structures Using LocARNA. <http://www.bioinf.uni-leipzig.de>. Accessed 03 Aug 2018.
43. Will S, Reiche K, Hofacker IL, Stadler PF, Backofen R. Inferring noncoding RNA families and classes by means of genome-scale structure-based clustering. *PLoS Comput Biol.* 2007;3(4):65. <https://doi.org/10.1371/journal.pcbi.0030065>.
44. Holzapfel WH, Haberer P, Geisen R, Björkroth J, Schillinger U. Taxonomy and important features of probiotic microorganisms in food and nutrition. *Am J Clin Nutr.* 2001;73(2):365–73. <https://doi.org/10.1093/ajcn/73.2.365>.
45. Lechner M, Findeiss S, Steiner L, Marz M, Stadler PF, Prohaska SJ. Proteinortho: detection of (co-)orthologs in large-scale analysis. *BMC Bioinformatics.* 2011;12:124.
46. Barre F-X, Søballe B, Michel B, Aroyo M, Robertson M, Sherratt D. Circles: the replication-recombination-chromosome segregation connection. *Proc Natl Acad Sci.* 2001;98(15):8189–95.
47. Stanage TH, Page AN, Cox MM. Dna flap creation by the RarA/MgsA protein of *Escherichia coli*. *Nucleic Acids Res.* 2017;45(5):2724–35.
48. Carrasco B, Seco EM, López-Sanz M, Alonso JC, Ayora S. *Bacillus subtilis* RarA modulates replication restart. *Nucleic Acids Res.* 2018;46(14):7206–20.
49. Liu W-T, Karavolos MH, Bulmer DM, Allaoui A, Hormaeche RDCE, Lee JJ, Khan CA. Role of the universal stress protein UspA of *Salmonella* in growth arrest, stress and virulence. *Microb Pathog.* 2007;42(1):2–10.
50. Gustavsson N, Diez A, Nyström T. The universal stress protein paralogues of *Escherichia coli* are co-ordinately regulated and co-operate in the defence against DNA damage. *Mol Microbiol.* 2002;43(1):107–17.
51. Kvint K, Nachin L, Diez A, Nyström T. The bacterial universal stress protein: Function and regulation. *Curr Opin Microbiol.* 2003;6:140–5. [https://doi.org/10.1016/S1369-5274\(03\)00025-0](https://doi.org/10.1016/S1369-5274(03)00025-0).
52. Huang G, Li C, Cao Y. Proteomic analysis of differentially expressed proteins in *Lactobacillus brevis* ncl912 under acid stress. *FEMS Microbiol Lett.* 2011;318(2):177–82. <https://doi.org/10.1111/j.1574-6968.2011.02257.x>.
53. Kaur G, Ali SA, Kumar S, Mohanty AK, Behare P. Label-free quantitative proteomic analysis of *Lactobacillus fermentum* ncdc 400 during bile salt exposure. *J Proteomics.* 2017;167:36–45. <https://doi.org/10.1016/j.jprot.2017.08.008>.
54. Lorenz R, Bernhart SH, Höner Zu Siederdisen C, Tafer H, Flamm C, Stadler PF, Hofacker IL. ViennaRNA Package 2.0. *Algorithms Mol Biol AMB.* 2011;6:26. <https://doi.org/10.1186/1748-7188-6-26>.
55. Hoch PG, Schlereth J, Lechner M, Hartmann RK. *Bacillus subtilis* 6S-2 RNA serves as a template for short transcripts *in vivo*. *RNA (New York, N.Y.)* 2016;22(4):614–22. <https://doi.org/10.1261/ma.055616.115>.
56. Donner J, Reck M, Bergmann S, Kirschning A, Müller R, Wagner-Döbler I. The biofilm inhibitor Carolacton inhibits planktonic growth of virulent pneumococci via a conserved target. *Sci Rep.* 2016;6(1):1–15.
57. Lécivain A-L, Le Rhun A, Renaut TT, Ahmed-Begrich R, Hahnke K, Charpentier E. *In vivo* 3'-to-5' exoribonuclease targetomes of *Streptococcus pyogenes*. *Proc Natl Acad Sci.* 2018;115(46):11814–9.
58. Muscariello L, Marasco R, De Felice M, Sacco M. The functional ccpa gene is required for carbon catabolite repression in *Lactobacillus plantarum*. *Appl Environ Microbiol.* 2001;67(7):2903–7. <https://doi.org/10.1128/AEM.67.7.2903-2907.2001>.
59. Giaretta S, Treu L, Vendramin V, da Silva Duarte V, Tarrah A, Campanaro S, Corich V, Giacomini A. Comparative transcriptomic analysis of *Streptococcus thermophilus* th1436 and th1477 showing different capability in the use of galactose. *Front Microbiol.* 2018;9:1765. <https://doi.org/10.3389/fmicb.2018.01765>.
60. Grand M, Aubourg M, Pikis A, Thompson J, Deutscher J, Hartke A, Sauvageot N. Characterization of the gen locus involved in b-1,6-oligosaccharide utilization by *Enterococcus faecalis*. *Mol Microbiol.* 2019;112(6):1744–56. <https://doi.org/10.1111/mmi.14390>.
61. Kim H-M, Waters A, Turner ME, Rice KC, Ahn S-J. Regulation of cid and Irg expression by ccpa in *Streptococcus mutans*. *Microbiology (Reading, England).* 2019;165(1):113–23. <https://doi.org/10.1099/mic.0.000744>.
62. Kim J-H, Yang Y-K, Chambliss GH. Evidence that *Bacillus* catabolite control protein CcpA interacts with RNA polymerase to inhibit transcription. *Mol Microbiol.* 2005;56(1):155–62.
63. Ogaugwu CE, Cheng Q, Fieck A, Hurwitz I, Durvasula R. Characterization of a *Lactococcus lactis* promoter for heterologous protein production. *Biotechnol Rep.* 2018;17:86–92. <https://doi.org/10.1016/j.btre.2017.11.010>.
64. Oberto J. Synttax: a web server linking synteny to prokaryotic taxonomy. *BMC Bioinformatics.* 2013;14:4. <https://doi.org/10.1186/1471-2105-14-4>.
65. Hernández-Tamayo R, Graumann PL. *Bacillus subtilis* RarA forms damage-inducible foci that scan the entire cell. *BMC Res Notes.* 2019;12(1):1–3.
66. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL. Blast+: architecture and applications. *BMC Bioinformatics.* 2009;10(1):421.
67. Kalvari I, Argasinska J, Quinones-Olvera N, Nawrocki EP, Rivas E, Eddy SR, Bateman A, Finn RD, Petrov AI. Rfam 13.0: shifting to a genome-centric resource for non-coding RNA families. *Nucleic Acids Res.* 2018;46(D1):335–42. <https://doi.org/10.1093/nar/gkx1038>.
68. Nawrocki EP, Eddy SR. Infernal 1.1: 100-fold faster RNA homology searches. *Bioinformatics (Oxford, England).* 2013;29(22):2933–5. <https://doi.org/10.1093/bioinformatics/btt509>.
69. Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, Li W, Lopez R, McWilliam H, Remmert M, Söding J, Thompson JD, Higgins DG. Fast, scalable generation of high-quality protein multiple sequence alignments using clustal omega. *Mol Syst Biol.* 2011;7:539. <https://doi.org/10.1038/msb.2011.75>.
70. Kingsford CL, Ayanbule K, Salzberg SL. Rapid, accurate, computational discovery of Rho-independent transcription terminators illuminates their relationship to DNA uptake. *Genome Biol.* 2007;8(2):1–12.
71. Gardner PP, Barquist L, Bateman A, Nawrocki EP, Weinberg Z. RNIE: genome-wide prediction of bacterial intrinsic terminators. *Nucleic Acids Res.* 2011;39(14):5845–52.
72. Brownlee GG. Sequence of 6S RNA of *E. coli*. *Nat New Biol.* 1971;229(5):147–9.
73. Darty K, Denise A, Ponty Y. Varna: Interactive drawing and editing of the RNA secondary structure. *Bioinformatics (Oxford, England).* 2009;25(15):1974–5. <https://doi.org/10.1093/bioinformatics/btp250>.
74. Wurm R, Neusser T, Wagner R. 6S RNA-dependent inhibition of RNA polymerase is released by RNA-dependent synthesis of small *de novo*

- products. *Biol Chem.* 2010;391(2-3):187–96. <https://doi.org/10.1515/BC.2010.018>.
75. Crooks GE, Hon G, Chandonia J-M, Brenner SE. Weblogo: a sequence logo generator. *Genome Res.* 2004;14(6):1188–90. <https://doi.org/10.1101/gr.849004>.
  76. Subramanian B, Gao S, Lercher MJ, Hu S, Chen W-H. Evolvview v3: a webserver for visualization, annotation, and management of phylogenetic trees. *Nucleic Acids Res.* 2019;47(W1):270–5. <https://doi.org/10.1093/nar/gkz357>.
  77. Edgar RC. Muscle: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* 2004;32(5):1792–7.
  78. Stamatakis A. Raxml version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* (Oxford, England). 2014;30(9):1312–3. <https://doi.org/10.1093/bioinformatics/btu033>.
  79. Zomer AL, Buist G, Larsen R, Kok J, Kuipers OP. Time-resolved determination of the ccpa regulon of *Lactococcus lactis subsp. cremoris* mg1363. *J Bacteriol.* 2007;189(4):1366–81. <https://doi.org/10.1128/JB.01013-06>.
  80. Bailey TL, Gribskov M. Combining evidence using p-values: application to sequence homology searches. *Bioinformatics* (Oxford, England). 1998;14(1):48–54.
  81. Martin M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet J.* 2011;17(1):10–12. <https://doi.org/10.14806/ej.17.1.200>.
  82. Andrews S. FastQC A Quality Control Tool for High Throughput Sequence Data. <http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>. Accessed 12 Dec 2017.
  83. Hoffmann S, Otto C, Kurtz S, Sharma CM, Khaitovich P, Vogel J, Stadler PF, Hackermüller J. Fast mapping of short sequences with mismatches, insertions and deletions using index structures. *PLoS Comput Biol.* 2009;5(9):1000502. <https://doi.org/10.1371/journal.pcbi.1000502>.

## Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more [biomedcentral.com/submissions](https://biomedcentral.com/submissions)

