



**HAL**  
open science

## Differential evolution in 3'UTRs leads to specific gene expression in *Staphylococcus*

Pilar Menendez-Gil, Carlos Caballero, Arancha Catalan-Moreno, Naiara Irurzun, Inigo Barrio-Hernandez, Isabelle Caldelari, Alejandro Toledo-Arana

► **To cite this version:**

Pilar Menendez-Gil, Carlos Caballero, Arancha Catalan-Moreno, Naiara Irurzun, Inigo Barrio-Hernandez, et al.. Differential evolution in 3'UTRs leads to specific gene expression in *Staphylococcus*. *Nucleic Acids Research*, 2020, 48 (5), pp.2544-2563. 10.1093/nar/gkaa047 . hal-02966057v1

**HAL Id: hal-02966057**

**<https://hal.science/hal-02966057v1>**

Submitted on 27 Oct 2020 (v1), last revised 5 Apr 2023 (v2)

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

1 **Differential evolution in 3'UTRs leads to specific gene**  
2 **expression in *Staphylococcus***

3

4 Pilar Menendez-Gil<sup>1</sup>, Carlos J. Caballero<sup>1</sup>, Arancha Catalan-Moreno<sup>1</sup>, Naiara Irurzun<sup>1</sup>,  
5 Inigo Barrio-Hernandez<sup>2</sup>, Isabelle Caldelari<sup>3</sup> and Alejandro Toledo-Arana<sup>1\*</sup>

6

7 <sup>1</sup>Instituto de Agrobiotecnología (IdAB), CSIC-UPNA-Gobierno de Navarra, 31192-  
8 Mutilva, Navarra, Spain

9 <sup>2</sup>European Molecular Biology Laboratory, European Bioinformatics Institute (EMBL-  
10 EBI), Wellcome Genome Campus, Hinxton, Cambridgeshire, CB10 1SD, United  
11 Kingdom

12 <sup>3</sup>Université de Strasbourg, CNRS, Architecture et Réactivité de l'ARN, UPR9002, F-  
13 67000-Strasbourg, France

14

15

16 \*Corresponding author: Alejandro Toledo-Arana  
17 E-mail: [a.toledo.arana@csic.es](mailto:a.toledo.arana@csic.es)  
18 Phone: +34 948 16 9752  
19 Laboratory of Bacterial Gene Regulation  
20 Instituto de Agrobiotecnología  
21 CSIC-UPNA-Gobierno de Navarra  
22 Avda. de Pamplona 123  
23 31192-Mutilva, Navarra  
24 Spain  
25 Phone: +34 948 16 9752

26

27 **ABSTRACT**

28 The evolution of gene expression regulation has contributed to species differentiation.  
29 The 3' untranslated regions (3'UTRs) of mRNA include regulatory elements that  
30 modulate gene expression; however, our knowledge of their implications in the  
31 divergence of bacterial species is currently limited. In this study, we performed  
32 genome-wide comparative analyses of mRNAs encoding orthologous proteins from the  
33 genus *Staphylococcus* and found that mRNA conservation was lost mostly  
34 downstream of the coding sequence (CDS), indicating the presence of high sequence  
35 diversity in the 3'UTRs of orthologous genes. Transcriptomic mapping of different  
36 staphylococcal species confirmed that 3'UTRs were also variable in length. We  
37 constructed chimeric mRNAs carrying the 3'UTR of orthologous genes and  
38 demonstrated that 3'UTR sequence variations affect protein production. This  
39 suggested that species-specific functional 3'UTRs might be specifically selected during  
40 evolution. 3'UTR variations may occur through different processes, including gene  
41 rearrangements, local nucleotide changes, and the transposition of insertion  
42 sequences. By extending the conservation analyses to specific 3'UTRs, as well as the  
43 entire set of *Escherichia coli* and *Bacillus subtilis* mRNAs, we showed that 3'UTR  
44 variability is widespread in bacteria. In summary, our work unveils an evolutionary bias  
45 within 3'UTRs that results in species-specific non-coding sequences that may  
46 contribute to bacterial diversity.

47

48 **Keywords:** 3'UTRs, post-transcriptional regulation, mRNA evolution, *Staphylococcus*,  
49 species diversity, bacterial evolution.

50

## 51 INTRODUCTION

52 A prototypical bacterial mRNA comprises the protein coding sequence (CDS) and the  
53 5' and 3' untranslated regions (5'UTRs and 3'UTRs), which flank the CDS upstream  
54 and downstream, respectively. In addition to the Shine-Dalgarno (SD) sequence, large  
55 bacterial 5'UTRs often contain riboswitches, thermosensors, and RNA structures that  
56 are targeted by sRNAs and RNA-binding proteins. These regulatory elements control  
57 premature transcription termination, mRNA translation, and/or mRNA processing (1-3).  
58 Recent discoveries have revealed that 3'UTRs contain regulatory elements that  
59 modulate the central metabolism, virulence, and biofilm formation through different  
60 mechanisms. Several of these regulatory elements control the expression of the  
61 protein encoded in the same mRNA (*cis*-acting mechanism), while others behave as  
62 regulatory RNAs that base pair with other mRNAs (*trans*-acting mechanism) (4-19).

63 About *cis*-acting 3'UTRs, we previously showed that the long 3'UTR of the *icaR* mRNA,  
64 which encodes the main repressor of the PIA-PNAG exopolysaccharide synthesis in  
65 *Staphylococcus aureus*, interacts with the 5'UTR of the same mRNA to inhibit IcaR  
66 translation. This occurs through base-pairing of the anti-SD motif (UCCC sequence  
67 located in the 3'UTR) at the ribosome binding site (RBS), impairing ribosome  
68 accessibility by creating a double stranded RNA substrate that is cleaved by  
69 ribonuclease III (RNase III) (5). Similar 5'-3'UTR interactions were recently described in  
70 the *hbs* mRNA of *Bacillus subtilis*. The cleavage of the *hbs* 5'UTR by RNase Y is  
71 countered by a 14-nt perfect base-pairing interaction between the 5' and the 3'UTR,  
72 which partially overlaps with the cleavage site (8). Other bacterial 3'UTRs are often  
73 used as direct entry points for ribonucleases to initiate mRNA degradation. This is the  
74 case for the *hmsT* mRNA, which encodes a protein that modulates c-di-GMP synthesis,  
75 required for allosteric activation of polysaccharide production in *Yersinia pestis* biofilm  
76 formation. Polynucleotide phosphorylase (PNPase)-dependent cleavage of the *hmsT*  
77 3'UTR affects the turnover of the mRNA and, ultimately, HsmT expression (6, 9).  
78 Additionally, several AU-rich 3'UTRs of *Y. pestis* are targeted by PNPase when the

79 transcriptional terminator (TT) is Rho-dependent (9). In *Salmonella*, HliD, one of the  
80 main transcriptional activator factors of the pathogenicity island 1 (SPI-1) (20), is  
81 processed through its 3'UTR by RNase E and PNPase (21). Similarly, the *aceA* mRNA  
82 in *Corynebacterium glutamicum*, which encodes the isocitrate lyase protein that is part  
83 of the glyoxylate cycle (22), is processed by endoribonucleases E/G (RNase E/G) at  
84 the 3'UTR (4). On the contrary, there are examples of 3'UTRs that prevent mRNA  
85 degradation. When *Escherichia coli* grows under iron-limiting conditions, the apo-AcnB  
86 protein interacts with a stem-loop structure located at the 3'UTR of its own mRNA, and  
87 prevents RyhB-induced degradation. This occurs because the apo-AcnB binds close to  
88 the RNase E cleavage site, which is essential for *acnB* mRNA degradation (7).

89 Regarding *trans*-acting regulatory RNAs, *S. aureus* RNAIII was for decades the sole  
90 example of a prokaryotic mRNA with a regulatory 3'UTR. RNAIII consists of a 5'UTR of  
91 84-nt, a short CDS encoding the  $\delta$ -toxin (*hld*), and a long structured 3'UTR (352-nt) that  
92 controls the temporal expression of several virulence factors and the transcriptional  
93 repression of Rot exotoxins by targeting their mRNAs (23, 24). More recently, several  
94 small RNAs (sRNAs) derived from 3'UTRs have been described, some of which are  
95 expressed from alternative mRNA-internal promoters (Type I 3'UTR-derived sRNAs),  
96 while others are generated from specific mRNA processing events (Type II 3'UTR-  
97 derived sRNAs) (18).

98 Previous studies have shown the sequence conservation of some 3'UTRs to be  
99 variable depending on the bacterial species. For example, the 3'UTR of *dapB* mRNA,  
100 which produces the type I *dapZ* 3'UTR-derived sRNA, is not conserved among  
101 enterobacterial species (except for the R1 seed and the Rho-independent terminator)  
102 (11). Similarly, the *sdhX* sRNA, which is generated by processing the 3' end of the *sdh*-  
103 *suc* operon by RNase E in *E. coli* and *Salmonella*, also shows nucleotide variations  
104 between different enterobacterial species (16, 17). Interestingly, sequence differences  
105 in the *sdhX* sRNA could establish an mRNA target divergence between *Salmonella*  
106 and *E. coli* (16).

107 Although, these and other pioneering studies have indicated the importance of 3'UTRs  
108 as a new class of post-transcriptional regulatory elements controlling relevant  
109 physiological processes in bacteria, several questions remain unanswered. Are 3'UTR  
110 sequences preserved within and between bacterial species? How often do conserved  
111 genes from closely related bacteria contain different 3' UTRs and how is 3'UTR  
112 variability originated? Do differences in 3'UTRs between orthologous genes have  
113 consequences in their expression at the protein level?

114 In this study, we aimed to evaluate the evolutionary relationship between the CDSs  
115 and their corresponding 3'UTR sequences in bacteria. To this end, we performed  
116 genome-wide comparative analyses of orthologous mRNAs among closely related  
117 species of the genus *Staphylococcus*. As a result, we discovered that most of the  
118 3'UTRs from these mRNAs were rather variable at the sequence level. In most cases,  
119 the inter-species nucleotide conservation of mRNAs was lost downstream of the coding  
120 sequence, suggesting a large variability within the 3'UTRs. *In vivo* mapping using RNA-  
121 seq of the transcript boundaries in three species of the genus *Staphylococcus*  
122 confirmed that 3'UTRs vary in length and sequence when compared with *S. aureus*  
123 transcriptomic data. We demonstrated that such differences may have an impact on  
124 the regulation of orthologous genes and, thus, the functionality of the 3'UTR could  
125 differ depending on the species being analysed. Finally, by extending the 3'UTR  
126 conservation analysis to other bacteria genera, we found similar sequence variations.

127 In summary, this study proposes that 3'UTRs from orthologous genes may be  
128 selectively targeted by evolution to create specific sequence differences that contribute  
129 to bacterial diversity.

## 130 MATERIALS AND METHODS

131

### 132 Strains, plasmids, oligonucleotides and growth conditions

133 The bacterial strains, plasmids and oligonucleotides used in this study are listed in  
134 Table S1, Table S2 and Table S3, respectively. *Staphylococcus* sp. strains were grown  
135 in Tryptic Soy Broth (Pronadisa) supplemented with 0.25% glucose (TSBg) or, when  
136 indicated, in Brain Heart Infusion (BHI). *Escherichia coli* was grown in Luria-Bertani  
137 (LB) broth (Pronadisa). The B2 (casein hydrolysate, 10 g l<sup>-1</sup>; yeast extract, 25 g l<sup>-1</sup>;  
138 NaCl, 25 g l<sup>-1</sup>; K<sub>2</sub>HPO<sub>4</sub>, 1 g l<sup>-1</sup>; glucose, 5 g l<sup>-1</sup>; pH 7.5) and SuperBroth (tryptone, 30 g  
139 l<sup>-1</sup>; yeast extract, 20 g l<sup>-1</sup>; MOPS, 10 g l<sup>-1</sup>; pH 7) media were used to prepare *S. aureus*  
140 and *E. coli* competent cells, respectively. For selective growth, media were  
141 supplemented with the appropriated antibiotics at the following concentrations:  
142 Erythromycin (Erm), 10 µg ml<sup>-1</sup>; Ampicillin (Amp), 100 µg ml<sup>-1</sup>.

143

### 144 Nucleotide conservation analysis

145 The conservation of UTRs among orthologous monocistronic mRNAs from  
146 phylogenetically-related bacterial species was determined by performing blastn  
147 comparisons using the Microbial Nucleotide BLAST tool  
148 (<https://blast.ncbi.nlm.nih.gov/>). Monocistronic mRNA boundaries were manually  
149 annotated by visualizing in the Jbrowse application (25) the *S. aureus* transcriptomic  
150 maps, which were previously determined (5, 26, 27). For the batch comparison of *S.*  
151 *aureus* 3'UTRs, a file including the query sequences in FASTA format was used to  
152 compare one representative genome at a time. Each query sequence included the  
153 annotated 3'UTR and the last 200 nt of the CDS (Fig. S1). This restriction was applied  
154 to normalize the starting point of each UTR in the obtained blast result data, facilitating  
155 further analysis and plotting. We considered that 200 nt were enough to properly  
156 identify the corresponding orthologous genes when comparing them to the genomes.  
157 Similarly, for batch comparison of *S. aureus* 5'UTRs, a FASTA file including the

158 annotated 5'UTRs and the first 200 nt of the CDSs were used. Again, in order to  
159 normalize the starting point of the CDS, we used reverse-complement sequences for  
160 5'UTR blastn comparisons (Fig. S1). The UTR conservation length was registered as  
161 the last nucleotide showing a positive alignment of the query sequence with the region  
162 of the genome under comparison. For this purpose, the information contained in the  
163 blastn result TXT files was transformed into tables using a custom R script. Nucleotide  
164 alignments that started at values higher than 100, finished at values lower than 100, or  
165 were shorter than 80 nucleotides were not considered. The UTR conservation lengths  
166 of each *Staphylococcus* species were plotted against the reference UTR lengths of *S.*  
167 *aureus* (Fig. S1). The square correlation coefficient ( $R^2$ ) was calculated in R from the  
168 resulting fitted linear model. This analysis was also applied to compare the mapped  
169 3'UTRs of *B. subtilis* (28) and *E. coli* (29) with phylogenetically-related species of the  
170 *Bacillus* genus and *Enterobacteriaceae* family, respectively. Note that in the case of *B.*  
171 *subtilis*, 3'UTRs longer than 150 nt were not annotated (28).

172

### 173 **RNA sequencing and data analysis**

174 In order to perform transcriptomic maps of the *S. simiae* CCM 7213T, *S. capitis* SK14  
175 and *S. epidermidis* RP62A strains, bacteria were grown in TSBg at 37°C and 200 rpm  
176 until exponential phase was reached and total RNAs were extracted, as previously  
177 described (30). RNA sequencing and preliminary data analysis was carried out by the  
178 Stab Vida company. RNA-seq reads from the *S. simiae* CCM 7213T, *S. capitis* SK14  
179 and *S. epidermidis* RP62A samples were aligned with the Rockhopper program (31)  
180 using the complete genome sequences from *S. simiae* NCTC13838 (NZ\_LT906460.1),  
181 *S. capitis* AYP1020 (NZ\_CP007601.1) and *S. epidermidis* RP62A (NC\_002976.3) as  
182 references, respectively. Next, the obtained wiggle read coverage (.wig) files from the  
183 plus and minus strands were converted to BigWig (.bw) files with the wig2BigWig  
184 program. The generated .bw files and the gene annotation files, including the positions  
185 of transcriptional terminators (TTs) predicted by TransTerm HP program (32), were

186 loaded into a public web server based on Jbrowse (25) (<http://rnamaps.unavarra.es/>).  
187 The 3'UTR lengths of the orthologous monocistronic mRNAs were manually annotated  
188 and plotted against the lengths of previously described *S. aureus* 3'UTRs (5).

189

### 190 **Simultaneous mapping of 5' and 3' ends of RNA molecules**

191 The mapping of the 5' and 3' ends of mRNAs was performed using a modified version  
192 of the Rapid Amplification of cDNA Ends (mRACE) method. (33). Briefly, total RNA  
193 samples were treated with the Cap-Clip Acid Pyrophosphatase (Tebu-Bio) following the  
194 manufacturer's recommendations. After 1 hour of incubation, RNAs were extracted by  
195 phenol-chloroform and precipitated with AcNa and ethanol. Different dilutions of Cap-  
196 Clip and non-Cap-Clip treated RNA were ligated using the T4 RNA Ligase I (New  
197 England Biolabs) at 16°C ON. An RT-PCR for all the ligated RNA dilutions was carried  
198 out using the SuperScript<sup>TM</sup> III One-Step RT-PCR System with the Platinum Taq DNA  
199 polymerase (Invitrogen) and the outward primers A and B (Table S3). For the mapping  
200 of the 3'UTR of *rpiRc* from *S. epidermidis* and *S. capitis* primers RACE-3XFLAG-A and  
201 RACE-RpiRc-sau-B were used. Bands of the expected size were purified and ligated  
202 into the pGEM-T Easy vector (Promega). The resulting plasmids were cloned in *E. coli*  
203 XL1-Blue (Stratagene) and ten clones analyzed by Sanger sequencing. Transcript  
204 boundaries were determined by blastn and nucleotide frequencies representing the 5'  
205 and 3' ends, respectively, were registered as percentages (%).

206

### 207 **Chromosomal mutagenesis**

208 The mutants generated in this study (Table S1) were obtained using the pMAD plasmid  
209 system, which consists in a two-step homologous recombination that exchanges a  
210 specific chromosomal region by the mutant allele present in the pMAD plasmid (34).  
211 These marker-less mutants were verified by PCR using oligonucleotides E and F  
212 (Table S3) and Sanger sequencing.

213

214 **Plasmid construction**

215 In general, the plasmids used in this study were engineered as previously described  
216 (35). Briefly, PCR fragments were amplified from chromosomal or plasmidic DNA with  
217 the DreamTaq DNA polymerase or Phusion High-Fidelity DNA Polymerase (Thermo  
218 Scientific), using the oligonucleotides listed in Table S3. The resulting PCR products  
219 were purified from agarose gels using the NucleoSpin® Gel and PCR Clean-up  
220 Macherey-Nagel kit, ligated into the pJET 1.2 vector (Thermo Scientific) and cloned in  
221 *E. coli* XL1-Blue. Plasmids were purified from overnight (ON) cultures with the  
222 NucleoSpin® Plasmid Macherey-Nagel kit and verified by Sanger sequencing. When  
223 required, DNA fragments were excised using FastDigest restriction enzymes (Thermo  
224 Scientific) and ligated into the appropriate vector with the Rapid DNA ligation kit  
225 (Thermo Scientific). The final plasmids were introduced into *S. aureus* strains by  
226 electroporation (36).

227 The pMAD plasmids used for chromosomal mutations were constructed by amplifying  
228 the flanking regions (AB and CD) of the region to be mutated using primers A and B  
229 and C and D (Table S3). After cloning them into pJET they were digested and ligated  
230 into pMAD in a two-insert ligation process using BamHI, NheI and EcoRI (Table S2 and  
231 Table S3).

232 Plasmids expressing chimeric *icaR* mRNAs were constructed taking advantage of  
233 conserved natural restriction sites located at the beginning of the *icaR* 3'UTRs  
234 Specifically, *icaR* mRNA regions were amplified by PCR using chromosomal DNA from  
235 the corresponding staphylococcal species and specific oligonucleotides IcaR+1 and  
236 IcaR-Term (Table S3). PCR products were cloned into pJET and digested with SpeI  
237 and EcoRI, with the exception of the PCR product from *S. simiae*, which was digested  
238 with HincII and EcoRI. The digested fragments containing different 3'UTRs were  
239 ligated into the p<sup>3XF</sup>IcaRm or plcaRm plasmid (5), producing plasmids  
240 p<sup>3XF</sup>IcaR+3'UTR<sup>Sarg</sup>, p<sup>3XF</sup>IcaR+3'UTR<sup>Ssim</sup>, p<sup>3XF</sup>IcaR+3'UTR<sup>Sepi</sup>, p<sup>XF</sup>IcaR+3'UTR<sup>Scap</sup>,  
241 plcaR+3'UTR<sup>Sarg</sup>, plcaR+3'UTR<sup>Ssim</sup>, plcaR+3'UTR<sup>Sepi</sup> and plcaR+3'UTR<sup>Scap</sup> (Table S2).

242 The *ftnA* and *rpiRc* genes were labelled with the 3xFLAG tag sequence by overlapping  
243 PCRs. Specifically, the tagged *ftnA* gene was generated by linking two partially  
244 overlapping PCR fragments that were amplified using oligonucleotides Ftn-A, 3XFLAG-  
245 ftn-izqda, 3XFLAG-ftn-drcha and Term-ftn (Table S3). A second PCR with  
246 oligonucleotides +1-ftn and Term-ftn (Table S3) was performed to amplify the <sup>3x</sup>FtnA  
247 mRNA region. The PCR fragment was ligated into pJET, digested with restriction  
248 enzymes BamHI and KpnI and cloned into pEW, generating the p<sup>3x</sup>FtnA+3'UTR<sup>Sau</sup>  
249 plasmid (Table S2). The pEW plasmid was constructed by subcloning the  
250 transcriptional terminator region from pCN47 into the pCN40 plasmid (37). Similarly,  
251 the *rpiRc* gene was tagged by introducing the 3xFLAG sequence through overlapping  
252 PCRs using oligonucleotides +1-RpiRc, 3XFLAG-RpiRc-izqda, 3XFLAG-RpiRc-drcha  
253 and Term-RpiRc (Table S3). This PCR fragment was digested with BamHI and EcoRI  
254 and subcloned into the p<sup>3x</sup>RpiRc+3'UTR<sup>Sau</sup> plasmid (Table S2).

255 Plasmids p<sup>3x</sup>FtnAΔ3'UTR and p<sup>3x</sup>RpiRcΔ3'UTR, which express the *ftnA* and *rpiRc*  
256 mRNAs lacking the 3'UTRs, were constructed by amplifying PCR fragments from the  
257 p<sup>3x</sup>FtnA+3'UTR<sup>Sau</sup> and p<sup>3x</sup>RpiRc+3'UTR<sup>Sau</sup> plasmids with oligonucleotide pairs +1-ftn /  
258 Δ3UTR-ftn-term and +1-RpiRc / Δ3UTR-RpiRc-term (Table S3), respectively, and  
259 subcloned into pEW (Table S2).

260 To generate the plasmids expressing the chimeric tagged *ftnA* and *rpiRc* mRNAs  
261 (Table S2), the corresponding 3'UTRs from different staphylococcal species were  
262 fused to *S. aureus* FtnA or RpiRc CDSs by overlapping PCRs. First, PCR fragments  
263 including the 5'UTR and the 3xFLAG CDS of *ftnA* or *rpiRc* genes were generated by  
264 using the p<sup>3x</sup>FtnA+3'UTR<sup>Sau</sup> and p<sup>3x</sup>RpiRc+3'UTR<sup>Sau</sup> plasmids as templates and  
265 oligonucleotide pairs +1-ftn / CDS-stop-ftn and +1-RpiRc / CDS-stop-RpiRc as primers,  
266 respectively. Second, the 3'UTR regions from different staphylococcal species were  
267 amplified using their corresponding fw and rvs oligonucleotides (Table S3). Finally, the  
268 PCR fragments including the 3xFLAG tagged CDS and the 3'UTR were fused by PCR  
269 using oligonucleotides +1 and rvs (Table S3). Next, they were ligated into pJET and

270 digested with BamHI/ KpnI and BamHI/EcoRI in the case of <sup>3x</sup>FtnA and <sup>3x</sup>F RpiRc,  
271 respectively, and subcloned into pEW.

272 The p<sup>3x</sup>F RpiRc+3'UTR<sup>IS256</sup> plasmid was generated by PCR simulating the 3'UTR of  
273 *rpiRc* from the *S. aureus* strain 2010-60-6511-5, which carries an IS256 insertion. The  
274 IS256 was amplified from the *S. aureus* strain 15981, using primers IS256-fw and  
275 IS256-3'UTR-RpiRc-rvs. The 3'UTR fragment was amplified from p<sup>3x</sup>F RpiRc, using  
276 RpiRc-CDS-SpeI-fw and 3UTR-RpiRc-42-IS256-rvs (Table S3). An overlapping PCR  
277 with the primer pair RpiRc-CDS-SpeI-fw / IS256-3'UTR-RpiRc-rvs was used to  
278 generate the 3'UTR carrying the IS256 insertion, which was then cloned into pJET.  
279 Natural restriction sites were used to digest p<sup>3x</sup>F RpiRc with SpeI/EcoRI, pJET<sup>3x</sup>F RpiRc  
280 with NspI/EcoRI and pJET-3'UTR+IS1181 with SpeI/NspI, followed by a double-  
281 fragment ligation into the pEW plasmid to reconstitute an *rpiRc* gene with the IS256  
282 insertion in the 3'UTR.

283 The plasmid carrying <sup>3x</sup>F RpiRc with the IS1181 insertion in the 3'UTR,  
284 p<sup>3x</sup>F RpiRc+3'UTR<sup>IS1181</sup>, was generated simulating the 3'UTR of *rpiRc* from the *S.*  
285 *aureus* strain DAR1183 by PCR. This was done by amplifying the IS1181 from the *S.*  
286 *aureus* strain N315, using primers 3'UTR-RpiRc-IS1181-fw and IS1181-rvs and the  
287 3'UTR from p<sup>3x</sup>F RpiRc, using IS1181-3UTR-RpiRc-fw and pCN-univ-AT (Table S3). An  
288 overlapping PCR with the primer pair 3'UTR-RpiRc-IS1181-fw/pCN-univ-rv-AT was  
289 used to generate the 3'UTR carrying IS1181, which was then cloned into pJET. Natural  
290 restriction sites were used to digest p<sup>3x</sup>F RpiRc with SpeI/KasI or SpeI/HindIII and pJET-  
291 3'UTR+IS1181 with HindIII/KasI, followed by a double-fragment ligation into the pEW  
292 plasmid.

293 The GFP reporter plasmids were constructed using the *Listeria monocytogenes* pAD-  
294 cGFP plasmid as template (38). The *hly* 5'UTR and GFP sequences were amplified  
295 with primers Sal-GFP-fw and BcuI-TT-BamHI-GFP-rvs (Table S3), and the PCR  
296 fragment was cloned into the pEW plasmid using Sall and BamHI. The 3'UTR of *rpiRc*  
297 was amplified using primers BamHI-EcoRI-3UTR-RpiRc-fw and SmaI-3UTR-RpiRc-rvs

298 (Table S3) and cloned downstream of the *gfp* gene using the restriction sites BamHI  
299 and SmaI. The 3'UTRs carrying the IS were amplified from p<sup>3x<sub>F</sub></sup>RpiRc+3'UTR<sup>IS256</sup> and  
300 p<sup>3x<sub>F</sub></sup>RpiRc+3'UTR<sup>IS1181</sup> using BamHI-EcoRI-3UTR-RpiRc-fw and KpnI-term-RpiRc and  
301 introduced into the pEW-GFP using BamHI and KpnI. Lastly, the pGFP-Δ3'UTR-*rpiRc*  
302 was constructed using the pGFP-3'UTR-*rpiRc* as a template and primers Sall-GFP-fw  
303 and KpnI-term-RpiRc. The amplification product was then introduced into the pEW-  
304 GFP using Sall and KpnI.

305

### 306 **Total protein extraction and Western Blotting**

307 Preinocula were grown in 5 mL of TSBg supplemented with Erm (TSBg+Erm) and  
308 incubated ON at 37°C and 200 rpm. Bacterial strains were inoculated into Erlenmeyer  
309 flasks containing TSBg+Erm until a starting OD<sub>600</sub> of 0.02 was reached. Cultures were  
310 grown in TSBg+Erm at 37°C and 200 rpm until OD<sub>600</sub> = 0.5 (exponential phase) for  
311 strains carrying *IcaR* and *FtnA* constructs and OD<sub>600</sub> 5-6 (stationary phase) for strains  
312 carrying RpiRc plasmids. Then, culture samples were harvested by centrifugation (10  
313 min at 4400 g and 4°C) and bacterial pellets stored at -20°C until needed. Pellets were  
314 thawed, washed and resuspended in 1 mL of PBS 1X. Next, bacterial suspensions  
315 were transferred to Fast Prep tubes containing acid-washed 100 μm glass beads  
316 (Sigma) and bacteria were lysed using a FastPrep-24 instrument (MP Biomedicals) at  
317 speed 6 for 45 s twice. Total cell extracts were obtained by centrifuging the Fast Prep  
318 tubes for 10 min at 21000 g and 4°C. Total protein concentration was quantified using  
319 the Bio-Rad protein assay kit. Samples were prepared at the desired concentration in  
320 Laemmli buffer and stored at -20°C until needed.

321 Western blotting was performed as previously described (35). The 3xFLAG tagged  
322 *IcaR*, RpiRc and *FtnA* samples were incubated with anti-FLAG antibodies diluted  
323 1:1000 (Sigma) whereas the GFP samples were incubated with monoclonal anti-GFP  
324 antibodies 1:5000 (Living Colors, Clontech) and peroxidase-conjugated goat anti-  
325 mouse immunoglobulin G and M antibodies 1:2500 (Thermo Scientific). Proteins were

326 quantified by densitometry of western blot images using ImageJ  
327 (<http://rsbweb.nih.gov/ij/>). Each of the protein levels was relativized to the levels of *S.*  
328 *aureus* (Sau) or WT sample.

329

330

### 331 **RNA extraction and Northern Blotting**

332 Bacteria were grown as described for protein extraction, centrifuged for 3 min and 4400  
333 g at 4°C and bacterial pellets were stored at -80°C. RNA extraction was performed as  
334 described by Toledo-Arana et al. (30), while Northern blotting was performed as  
335 previously described by (35) with the following variations. Radiolabeled-RNA probes  
336 were synthesized from PCR products carrying the T7 promoter (Table S3) using the  
337 MAXIscript T7 transcription kit (Ambion) and <sup>32</sup>P- $\alpha$ -UTP, following the manufacturer's  
338 recommendations. Probes were then purified with Illustra MicroSpin G-50 columns (GE  
339 Healthcare) and membranes were hybridized with the RNA probes at 68°C ON. mRNA  
340 levels were quantified by densitometry of northern blot autoradiographies using ImageJ  
341 (<http://rsbweb.nih.gov/ij/>). Each of the mRNA levels was relativized to the levels of *S.*  
342 *aureus* (Sau) or WT sample.

343

344

### 345 **RNA stability assay**

346 Preinocula were grown in 5 mL of TSBg supplemented with Erm (TSBg+Erm) and  
347 incubated ON at 37°C and 200 rpm. Bacterial strains were inoculated into Erlenmeyer  
348 flasks containing TSBg+Erm until a starting OD<sub>600</sub> of 0.02 was reached. Cultures were  
349 grown in TSBg+Erm at 37°C and 200 rpm until OD<sub>600</sub> = 0.5 was reached. 20 mL of the  
350 culture were transferred to 50 mL Falcon tubes containing 300  $\mu$ g ml<sup>-1</sup> of Rifampicin  
351 and were incubated at 37 °C for 0, 2, 4, 8, 15 and 30 min. After the incubation time, 5  
352 mL of the Stop solution (95% Ethanol, 5% Phenol) were added to the samples and  
353 these were centrifuged for 2 min at 4400 g. Pellets were frozen in liquid nitrogen and

354 stored at -80°C. RNA extraction and Northern blot analysis were conducted as  
355 described above.

356

### 357 **Identification of elements disrupting the *rpiRc* mRNA sequence**

358 Updated *S. aureus* genome sequences were retrieved from  
359 <ftp://ftp.ncbi.nlm.nih.gov/genomes/> through the wget command-line utility. The Fasta  
360 files were loaded into a local nucleotide database using the Geneious software  
361 package. Next, a Blastn algorithm was locally run using *rpiRc* mRNA as a query.  
362 Genomes presenting pairwise alignment disruption located at the *rpiRc* 3'UTR were  
363 registered. When possible, sequences surrounding the position of alignment  
364 disruptions were used to look for insertion sequences at the ISfinder web application  
365 (<https://isfinder.biotoul.fr/>) (39). Note that in some cases, due to the assembly protocol,  
366 repeated sequences are eliminated from contig ends avoiding the identification of the  
367 IS responsible for the alignment disruption.

368

### 369 **Haemolysis assay**

370 Bacterial cultures were grown in 15 mL of BHI ON at 37°C and 200 rpm. Cultures were  
371 centrifuged for 10 min and 4400 g at 4°C and supernatants were equalized using the  
372 wet weight of the pellets. A 0.2 µm filtration process was applied to remove any  
373 remaining cells in the supernatant samples. These were then introduced in a  
374 SpeedVac Concentrator to make a 10 times volume reduction. 5-mm holes were made  
375 to Columbia Sheep blood (5%) agar plates (Biomérieux) and 30 µl of the concentrated  
376 supernatants were transferred into them. Plates were grown at 37°C ON and then kept  
377 at 4°C for several days. Haemolytic halos were then photographed.

378

### 379 ***In vitro* transcription**

380 PCR fragments containing the T7 promoter were used as templates for *in vitro*  
381 transcription. For this purpose, the regions of interest were amplified by PCR, with the

382 corresponding Fw oligonucleotides carrying the T7-RNA Polymerase binding site  
383 (Table S3). *In vitro* transcription was carried out using the T7 polymerase at 37°C ON  
384 followed by the removal of the PCR templates with a DNase I treatment. Next,  
385 transcripts were run on a denaturing 6% polyacrylamide gel and visualized with UV  
386 shadowing. Bands were cut and incubated with elution buffer (500 mM AcNH<sub>4</sub>, 1 mM  
387 EDTA, 0,1% SDS) and acid phenol pH 4.5 at 4°C ON. Subsequently, the transcribed  
388 RNA was purified using a phenol-chloroform extraction followed by precipitation with  
389 absolute ethanol. Pellets were then washed with 70% ethanol and resuspended in  
390 water. RNA integrity was checked through polyacrylamide gel electrophoresis and  
391 concentration was measured with a NanoDrop Instrument (Agilent Technologies).

392

### 393 **5'-end labelling of synthetic RNAs**

394 Before labelling, the RNA was dephosphorylated with FastAP (Thermo Scientific) at  
395 37°C for 15 min followed by a phenol-chloroform extraction. Next, 7 µg of  
396 dephosphorylated RNA were incubated with <sup>32</sup>P- γ-ATP and PNK (Thermo Scientific) at  
397 37°C for 1h. Labelled nucleic acids were then purified on a denaturing 6%  
398 polyacrylamide gel as described for the *in vitro* transcripts. The efficiency of the  
399 labelling was measured using a scintillation counter.

400

### 401 **Electrophoretic mobility shift assays**

402 Labelled and non-labelled RNAs were separately prepared at the desired  
403 concentrations by adding sterile milliQ water and 5X renaturing buffer (100 mM Tris-  
404 HCl pH 7.5, 300 mM KCl, 200 mM NH<sub>4</sub>Cl, 15 mM DTT). Samples were then denatured  
405 at 90°C for 1 min and chilled on ice for 1 min. Next, 10 mM of MgCl<sub>2</sub> were added to the  
406 samples followed by an incubation period of 10 min at 22°C. Subsequently, the labelled  
407 RNA was mixed with increasing concentrations of non-labeled samples in 1X reaction  
408 buffer (20 mM Tris-HCl pH 7.5, 60 mM KCl, 40 mM NH<sub>4</sub>Cl, 3 mM DTT, 10 mM MgCl<sub>2</sub>)  
409 in a total volume of 20 µl. Samples were then incubated at 37°C for 15 min before

410 being loaded in a non-denaturing 6% polyacrylamide gel and run in 1X TB containing 1  
411 mM MgCl<sub>2</sub> at 300V and 4°C. The gel was developed by autoradiography.

412

413 **PIA-PNAG quantification.**

414 Cell surface PIA/PNAG exopolysaccharide levels were extracted and quantified as  
415 previously described (5, 40). Cell surface extracts (5 μl), which were obtained from  
416 bacteria grown ON in TSBg, were spotted onto a nitrocellulose membrane using a Bio-  
417 Dot microfiltration apparatus (Bio-Rad). The membranes were blocked ON with 5%  
418 skimmed milk in phosphate-buffered saline with 0.1% Tween 20 (PBS-Tw20). After  
419 washing several times with PBS-Tw20, the membranes were initially incubated for 2 h  
420 with specific anti-PNAG antibodies diluted 1:10,000 (41) and then for an additional hour  
421 with peroxidase-conjugated goat anti-rabbit immunoglobulin G antibodies (Jackson  
422 ImmunoResearch Laboratories, Inc., Westgrove, PA) diluted 1:10,000. Finally,  
423 membranes were washed with PBS-Tw20 and developed using the SuperSignal West  
424 Pico Chemiluminescent Substrate (Thermo Scientific) and Amersham Hyperfilm ECL  
425 films (GE Healthcare Life Sciences).

426

## 427 RESULTS

428

### 429 **3'UTRs are highly variable among staphylococcal species.**

430 Next-generation RNA sequencing (NGS) technologies allow to accurately determine  
431 mRNA boundaries and, therefore, determine the lengths of the untranslated regions  
432 (UTRs). To evaluate the evolutionary relationship between coding sequences (CDSs)  
433 and their corresponding 3'UTRs, we performed genome-wide comparative analyses.  
434 Using high-resolution transcriptome maps of *S. aureus* as a reference, we analysed the  
435 3'UTR sequence conservation of close phylogenetic members of the genus  
436 *Staphylococcus* (Fig. S1). To simplify the study, we focused on monocistronic mRNAs.  
437 This ensured that each analysed CDS was flanked by a 5'UTR and a 3'UTR. First, we  
438 generated a query database including the 3'UTR sequences plus the last 200  
439 nucleotides (nt) of each monocistronic CDS (~590 sequences). This stretch of  
440 nucleotides was intended to serve as an indicator for the correct identification of  
441 orthologous genes among the staphylococcal species. Using blastn (42), we compared  
442 the selected sequences against 8 representative genomes of the most closely-related  
443 staphylococcal species according to the TimeTree knowledge-base (43) (Fig. S2). The  
444 last conserved nucleotide position of each of the orthologous mRNAs was registered  
445 (Table S4) and plotted against the length of their corresponding *S. aureus* mRNA (Fig.  
446 1A and S3). The comparison of *S. aureus* NCTC 8325 and MW2 strains revealed that  
447 97% of the analysed mRNAs fell on the diagonal axis of the plot, indicating no  
448 difference in 3'UTRs conservation. Only 17 (3%) out of 589 monocistronic mRNAs  
449 showed 3'UTR sequence variations, which could be explained by the presence of  
450 unique insertion sequences (ISs) or *Staphylococcus aureus* repeat (STAR) elements  
451 (44).

452 However, 3'UTR conservation among the analysed mRNAs decreased as the  
453 phylogenetic distance between *S. aureus* and the staphylococcal species increased.  
454 This phenomenon is reflected by the dots positioned along the horizontal line that

455 represents the stop codon position in Figure 1A and S3A. The lack of conservation was  
456 already noticeable in several 3'UTRs of *S. argenteus*, which is the closest species to *S.*  
457 *aureus*. This increased considerably in *S. simiae* and left the remaining species of the  
458 genus with only a few conserved 3'UTRs (Fig. 1A and S3A). As one would expect,  
459 among the conserved 3'UTRs (diagonal line) for all the staphylococcal species, we  
460 identified RNAlII and seven putative riboswitch-dependent 3'UTRs, meaning that the  
461 mRNA transcription ended at the TT of the downstream riboswitch when in the OFF  
462 configuration (5, 23, 24, 30). We also found two mRNAs, encoding hypothetical  
463 proteins SAOUHSC\_02781 and SAOUHSC\_02702, which produce the 3'UTR-derived  
464 sRNAs RsaOT (i.e. *srr43* or SAOUHSCs080) and Teg130 (i.e. SAOUHSCs100),  
465 respectively (45-47). The remaining mRNAs with conserved 3'UTRs coded for enolase  
466 (*eno*), ribosomal protein L32 (*rpmF*), and a small membrane protein  
467 (SAOUHSC\_01055a).

468 It is worth noting that, in most of the mRNAs, nucleotide conservation was lost  
469 downstream of the orthologous CDS (Fig. 1A and S3A). This is illustrated in Figure 1B  
470 and S3B, in which the distribution of the last conserved nucleotide position (quantified  
471 in windows of 10 nt) is shown. The top left panel of this figure shows the 3'UTR  
472 conserved length distribution in *S. aureus*, where the maximum of the peak is about 60  
473 nt downstream of the stop codon. Meanwhile, in most of the other staphylococcal  
474 species, the maximum is located within the region comprised between -10/+10 nt from  
475 the stop codon (Fig. 1B and S3B).

476 It is expected that non-coding sequences accumulate more changes throughout the  
477 course of evolution than their corresponding CDS, which cannot significantly change  
478 without dramatically affecting protein functionality. In agreement with this, 5'UTRs  
479 (excluding the ribosome binding region) should suffer a similar degree of nucleotide  
480 variation than their 3'UTR counterparts. We performed the same blastn comparison,  
481 using a query database that included the 5'UTR sequences plus the first 200 nt of the  
482 CDS (Table S5). As shown by the scatter plots in Figure S4, evident variations among

483 the 5'UTRs of orthologous genes existed. However, the number of conserved 5'UTRs  
484 was higher than their 3'UTR counterparts. This is significant when comparing the  
485 5'UTR and 3'UTR scatter plots of a specific staphylococcal species, as the number of  
486 dots closer to the diagonal line is always greater for the 5'UTRs (Fig. S4). Although this  
487 analysis may be biased by the presence of the ribosome binding site, it supports the  
488 idea that 3'UTRs are more prone to evolutionary changes than 5'UTRs.

489

490 **Most of the mRNAs encoding orthologous proteins contain 3'UTRs with different**  
491 **lengths and sequences in *Staphylococcaceae***

492 The fact that the conservation of several 3'UTRs was almost fully lost after the CDSs  
493 indicates that the mRNAs encoding orthologous proteins have different 3'UTRs. To  
494 verify this hypothesis, we performed RNA sequencing of whole transcripts to determine  
495 and compare the real length of 3'UTRs among closely-related staphylococcal species.  
496 We loaded the obtained transcriptomic maps of the *S. simiae*, *S. epidermidis*, and *S.*  
497 *capitis* strains into a Jbrowse-based server (25), which is publicly available at  
498 <http://rnamaps.unavarra.es>, and manually annotated the lengths of the 3'UTRs from  
499 each *S. aureus* orthologous monocistronic gene by combining the RNA-seq data and  
500 the Rho-independent transcriptional terminator prediction (using TransTermHP  
501 algorithm) (32). Subsequently, we plotted the real 3'UTR lengths from each  
502 staphylococcal species against the 3'UTR lengths of their corresponding *S. aureus*  
503 orthologous genes. Figure 2A shows that most of the 3'UTR lengths from *S. aureus* do  
504 not correlate with those of the three other analysed staphylococcal species (Fig. 2A).  
505 This lack of correspondence is exemplified when selecting some biologically relevant  
506 genes (Fig. 2B, 2C, and 2D). As reflected in Fig. 2B and 2D, in most cases, the length  
507 of the 3'UTR shows variability among all the analysed species, while a similar 3'UTR  
508 length for a given gene was only preserved in a few examples. Overall, these data  
509 confirmed that most of the mRNAs encoding orthologous genes in staphylococcal  
510 species have 3'UTRs with different lengths in addition to sequence variation.

511

512 **3'UTR sequence differences are originated by local nucleotide changes and gene**  
513 **rearrangements.**

514 In order to understand how 3'UTR differences originated among mRNAs encoding  
515 orthologous proteins, we compared the whole genomic sequence of *S. aureus* (Sau)  
516 against that of *S. simiae* (Ssim), *S. epidermidis* (Sepi), and *S. capitis* (Scap) using  
517 Mauve (48). Note that we only focused on the monocistronic mRNAs, as stated above.  
518 We expected 3'UTR sequence variations to occur due to different genomic  
519 rearrangements, as previously described for some sRNAs located at intergenic regions  
520 (IGRs) (49). Therefore, we checked the sequence conservation in genomic regions  
521 downstream of the orthologous CDSs, including IGRs and adjacent CDSs, for each  
522 genomic pair (Sau vs Ssim, Sau vs Sepi, and Sau vs Scap). We assigned a value of 1  
523 or 0 depending on whether the downstream IGR/CDS was conserved or not,  
524 respectively (Fig. 3A). Confirming previous blastn analysis (Fig. 1), we found that only  
525 around 6.8–10.6% of the *S. aureus* CDSs had a conserved downstream IGR/CDS  
526 when compared to the abovementioned species (Fig. 3B). On the other hand, 28–35%  
527 of the orthologous CDSs lacked a conserved downstream IGR/CDS, indicating that  
528 gene rearrangements in these loci led to 3'UTR variations. The rest of CDSs showed a  
529 non-conserved downstream IGR but a preserved the downstream CDS (between 58%  
530 to 62% depending on the species analysed) (Fig. 3B). When performing this analysis,  
531 we realized that such variations substantially altered the IGR lengths (Fig. 3C).  
532 Considering that the 3' end conservation was lost around the protein stop codon (Fig.  
533 1), IGR length variations may be attributed to shifts in the position of the stop codon,  
534 which may ultimately affect the CDS length. To address such a possibility, we  
535 compared the lengths of *S. aureus* CDSs to those of their corresponding orthologous  
536 genes of the three abovementioned staphylococcal species. As Figure 3D shows, most  
537 of the orthologous CDSs have similar lengths, indicating that variations in the lengths  
538 of the IGRs were not due to differences in the analysed CDS lengths (Fig. 3D).

539 Therefore, we concluded that the differences on the 3'UTR sequences occur partly due  
540 to gene rearrangements and mostly due to sequence variations in the IGR sequences.  
541 The causes of local variations are unknown.

542

543 **Species-specific 3'UTRs variations affect the expression of orthologous genes.**

544 3'UTR variations would be physiologically relevant if they led to differences in the  
545 expression of the orthologous genes of a particular bacterial species. To study this in  
546 detail, we selected the *icaR* (5), *ftnA*, and *rpiRc* genes as examples since we had  
547 previously found that their expression was affected by the deletion of their  
548 corresponding 3'UTR. In addition, according to the single nucleotide polymorphism  
549 (SNP) analysis carried out by Joseph and colleagues (50), the 3'UTRs sequences of  
550 these mRNAs are highly conserved among different *S. aureus* strains (Fig. S5). At the  
551 same time, these 3'UTRs present different levels of conservation when compared to  
552 their closest phylogenetically-related staphylococcal species (Fig. 4A). In addition,  
553 transcript mapping by RNA-seq revealed variations in the 3'UTR lengths (Fig. 2B and  
554 D). According to synteny analysis, the 3'UTR of *IcaR* shows significant variations in  
555 length and sequence due to gene rearrangements (Fig. S6A), while the differences in  
556 the *ftnA* and *rpiRc* 3'UTRs are most likely caused by local nucleotide variations in the  
557 IGRs, as no gene rearrangements were found by synteny analysis (Fig. S6B and C).  
558 Although a CDS insertion is found downstream of the *ftnA* in *S. argenteus*, it does not  
559 affect the mRNA sequence (Fig. S6B).

560 To confirm the transcript boundaries of the *icaR*, *ftnA*, and *rpiRc* mRNAs from *S.*  
561 *aureus*, *S. simiae*, *S. epidermidis*, and *S. capitis*, we performed a simultaneous  
562 mapping of their 5' and 3' mRNA ends using a modified version of the rapid  
563 amplification of cDNA ends technique (mRACE). Figures S7 to S9 show the mRACE  
564 results that confirmed the mRNA mappings from the RNA-seq data in combination with  
565 the transcriptional terminator predictions. Note that *rpiRc* in *S. epidermidis* and *S.*  
566 *capitis* is the last gene of a polycistronic transcript. Therefore, we were unable to obtain

567 mRACE mapping results from the native gene. We used Mfold to predict the putative  
568 transcriptional terminator structures of the transcripts (51). Interestingly, these stem-  
569 loops were different for the majority of the cases (Fig. S7B, S8B, and S9B).

570 Based on the information collected, we constructed chimeric mRNAs that combined the  
571 CDSs of the three selected *S. aureus* genes with the 3'UTR of their corresponding  
572 orthologous genes (from the same closely staphylococcal species analysed in Figure  
573 4A). We fused such 3'UTRs to their corresponding *S. aureus* CDS, which carried the  
574 3xFLAG sequence at the N-terminal to monitor their protein expression (Fig. 4B).

575 These plasmids were then electroporated into the mutant strains that lacked the  
576 original genes to facilitate the detection of the mRNAs that were expressed from the  
577 plasmids. To confirm the mapping of the 3' end of the chimeric *rpiRc* mRNA carrying  
578 the 3'UTRs from *S. epidermidis* and *S. capitis*, we performed mRACE. The results  
579 showed that the transcripts ended at the expected sites, which were downstream of the  
580 predicted stem-loop (Fig. S9).

581 Expression analysis of the 3xFLAG-tagged proteins by western blotting revealed that  
582 the deletion of the 3'UTRs produced an increase of IcaR, Ftn, and RpiRc proteins.

583 Interestingly, the chimeric *icaR* mRNAs harbouring the UCCCC motif (3'UTRs from *S.*  
584 *simaie* and *S. argenteus*) (Fig. 4A), which was required for modulation of IcaR  
585 translation in *S. aureus* (5), presented similar IcaR protein levels to those shown by *S.*  
586 *aureus* wild-type (WT) mRNA. In contrast, chimeric *icaR* mRNAs lacking the UCCCC  
587 motif (*S. epidermidis* and *S. capitis* *icaR* 3'UTRs) expressed a comparable amount of  
588 the IcaR protein to that of a *S. aureus* mRNA carrying the 3'UTR deletion (Fig. 4C).

589 Chimeric mRNAs expressing FtnA and RpiRc showed protein yields similar to those of  
590 the same mRNAs lacking the 3'UTR. The exception to such behaviour was the  
591 chimeric construct carrying the *ftnA* mRNA in combination with the 3'UTR of *S.*  
592 *argenteus*, whose sequence was conserved when compared to *S. aureus* (Fig. 4A).

593 The same samples were then subjected to northern blot analyses. The results showed  
594 that, in most cases, the steady-state mRNA levels correlated with the protein levels.

595 This suggested that the increase in protein levels, which occurred when the original  
596 3'UTR sequence was substituted by the one present in its orthologous mRNA, was  
597 most likely due to an increase in mRNA stability, as previously described for  
598 *icaR*Δ3'UTR mRNA (5). To confirm this, we measured the half-life of the chimeric  
599 *icaR*+3'UTR<sup>Sepi</sup> mRNA as an example. Figure S10 shows that this chimeric mRNA  
600 presented a higher half-life in comparison with the one from the *icaR* WT mRNA (Fig.  
601 S10).

602 Note that the chimeras of *icaR*-3'UTR<sup>Scap</sup> and *ftnA*-3'UTR<sup>Sepi</sup> did not show increased  
603 mRNA levels, implying that protein expression could be affected in other ways.  
604 Altogether, these results indicated that species-specific variations in the 3'UTR  
605 sequences could not reproduce the expression levels found in *S. aureus*, suggesting  
606 putative distinct functional roles of 3'UTR in the different species.

607

#### 608 **IS1181 transposition to *rpiRc* 3'UTR modifies RpiRc expression and haemolysis** 609 **production**

610 The *rpiRc* 3'UTR in the *S. simiae* strain is considerably longer than the 3'UTRs from  
611 other staphylococcal species (Fig. 4A). A deep sequence analysis revealed that the *S.*  
612 *simiae* *rpiRc* 3'UTR contains three copies of STAR elements (44), which are distributed  
613 among staphylococcal species (52). Since the number and genomic locations of the  
614 STAR element copies are variable among species and strains (52), we investigated  
615 whether insertions of STAR elements may also occur in the *rpiRc* 3'UTR of *S. aureus*  
616 and, therefore, alter protein expression as observed in the chimeric *rpiRc* construct  
617 carrying the 3'UTR of *S. simiae* (Fig. 4C). Since SNP analysis were not indicative of the  
618 presence of insertion/deletion sequences, we performed *rpiRc* mRNA sequence  
619 pairwise alignments to look for alignment disruptions in the *rpiRc* 3'UTR. We reasoned  
620 that alignment disruptions could help us identify repeated regions. Following this idea,  
621 we used the whole *rpiRc* mRNA sequence as the query to perform blastn against all  
622 draft and complete *S. aureus* genomes available on the NCBI database. The results

623 showed that 33 *S. aureus* genomes presented alignment disruptions in the *rpiRc*  
624 3'UTR (Table S6). Since many of these genomes were not fully closed, some  
625 alignment disruptions occurred due to the query sequence aligning to different contigs.  
626 By analysing the extremes of the aligned contigs, we identified that 16 out of the 33  
627 genomes contained IS insertions in the *rpiRc* 3'UTR sequence in different locations.  
628 Specifically, we were able to map nine IS1181 (53), five IS256 (54), and two ISs of the  
629 IS30 transposase family (Fig. 5A and Table S6). The remaining genomes presented  
630 sequence deletions or duplications. Interestingly, we identified duplications of 8-12 nt in  
631 some of the strains, which was indicative of the presence/excision of ISs. The fact that  
632 no IS sequences were found in the related contigs may be explained by either the use  
633 of assembly algorithms that eliminate repeated sequences from the extremes of the  
634 contigs or the ISs being excised. However, the latter would not explain why these  
635 contigs were not assembled together, suggesting that an unknown IS may be inserted  
636 in the *rpiRc* 3'UTR of these genomes.

637 In order to analyse the consequence of such insertions, we constructed two plasmids  
638 containing the IS256 and IS1181 sequences 42 and 108 nt downstream of the *RpiRc*  
639 stop codon, which mimicked the chromosomic configurations found in the *rpiRc* region  
640 of the *S. aureus* 2010-60-6511-5 and DAR1183 strains, respectively (Fig. 5B). In both  
641 cases, the insertion orientation of the IS placed the transposase gene in the DNA  
642 strand opposite to that of *rpiRc*. Northern blot analysis revealed different chimeric  
643 mRNA lengths (Fig. 5C). On the one hand, IS1181 carried a stem-loop downstream of  
644 the transposase gene and in the same DNA strand as the *rpiRc* gene that acted as a  
645 transcriptional terminator (Fig. 5B). Therefore, the IS1181 insertion produced a  
646 chimeric *rpiRc-IS1181* mRNA of ~1.4 Kb (3'UTR-IS1181 of 222 nt) that included a few  
647 nucleotides of the IS1181 sequence. This band migrated in a similar manner as the  
648 original *rpiRc* mRNA. The impact of this insertion was an increase in the mRNA levels  
649 (Fig. 5B and C), which correlated with an increase in the <sup>3x</sup>F<sub>RpiRc</sub> protein levels (Fig.  
650 5D). On the other hand, since the IS256 lacked stem-loops able to prevent

651 transcription from adjacent genes, a chimeric *rpiRc-IS256* mRNA of ~2.8 Kb (3'UTR-  
652 IS256 ~1.5 Kb) was produced. In contrast to the IS1181 insertion, the chimeric *rpiRc-*  
653 *IS256* mRNA included the whole IS256 sequence (Fig. 5B and C). In this case, there  
654 were no considerable changes neither in the mRNA nor <sup>3x</sup>F RpiRc protein levels (Fig.  
655 5C and D). Overall, these examples indicate that IS insertions may produce different  
656 RpiRc protein expression outcomes depending on the type of IS, the insertion site, and  
657 the orientation of the IS.

658 To further confirm that the *rpiRc* 3'UTR has a functional role that could be affected by  
659 IS insertions, we analysed the effect of the WT 3'UTR and the chimeric 3'UTR-ISs on  
660 the expression of a heterologous gene, such as the green fluorescent protein (*gfp*). To  
661 this end, we fused the *rpiRc* 3'UTR WT, the 3'UTR-IS1181, and 3'UTR-IS256  
662 sequences downstream of the *gfp* gene. As a control, we included the  $\Delta$ 3'UTR  
663 sequence that carried only the *rpiRc* TT fused downstream of *gfp* (Fig. 5E). Western  
664 blot analyses showed that the *rpiRc* 3'UTR lead to a lower GFP expression when  
665 compared to the  $\Delta$ 3'UTR version (Fig. 5E). This indicated that the *rpiRc* 3'UTR could  
666 reduce the expression of a heterologous gene and, therefore, act as a functional  
667 module. However, the presence of IS1181 in the *rpiRc* 3'UTR increased GFP  
668 expression while IS256 only slightly decreased it. This confirms that the transposition  
669 of ISs in the *rpiRc* 3'UTR could have different consequences on protein expression  
670 depending on the insertion events (Fig. 5E).

671 Next, we decided to analyse the relevance of the *rpiRc* 3'UTR on the *S. aureus*  
672 physiology and the implications of the IS insertions in its role. In previous studies, it  
673 was shown that RpiRc represses the expression of RNAIII. RNAIII is the master  
674 regulator of *S. aureus* virulence, which activates many virulence factors, including  
675 haemolysins. A mutant lacking the *rpiRc* gene showed increased haemolysis due to  
676 the lack of RNAIII repression (55-57). To test whether alterations in the *rpiRc* 3'UTR  
677 could modify the haemolytic activity of *S. aureus*, we constructed chromosomal  
678 mutants in the reference *S. aureus* MW2, a community-acquired *agr*-positive

679 methicillin-resistant (MRSA) strain. On the one hand, we performed a 115-bp  
680 chromosomal deletion that removed most of the 3'UTR of *rpiRc* while preserving its  
681 intrinsic TT, as we did for plasmid construction (MW2  $\Delta$ 3'UTR). On the other hand, we  
682 inserted the IS1181 as in the *rpiRc* region of the *S. aureus* DAR1183 strain (MW2  
683 3'UTR+IS1181). As a control, we deleted the whole *rpiRc* gene (MW2  $\Delta$ *rpiRc*). These  
684 strains were grown overnight and the production of haemolysins in the culture  
685 supernatants was visualized in sheep-blood agar plates. As expected, the  $\Delta$ *rpiRc*  
686 mutant showed an increase in haemolysis compared to the WT strain (Fig. 5F). In  
687 contrast, the deletion of the *rpiRc* 3'UTR and the IS1181 insertion in the 3'UTR, which  
688 produced higher protein levels of the RpiRc repressor (Fig. 5D), resulted in lower levels  
689 of haemolysis (Fig. 5F). Overall, these results highlighted the relevance of *rpiRc* 3'UTR  
690 in the physiology of *S. aureus* while demonstrating that the RpiRc function could be  
691 altered by the insertion of IS1181 in the *rpiRc* 3'UTR.

692

### 693 **Species-specific *icaR* 3'UTRs produce different PNAG levels in *S. aureus***

694 The expression analysis of chimeric *icaR* mRNAs indicated that the *icaR* 3'UTRs from  
695 *S. epidermidis* and *S. capitis* were unable to reproduce the 3'UTR-mediated regulation  
696 observed in *S. aureus* (Fig. 4C). We previously showed that the *icaR* 3'UTR interacted  
697 with the 5'UTR of the same mRNA molecule through a UCCCC motif to repress IcaR  
698 translation and thus modulate PIA-PNAG biosynthesis and biofilm formation in *S.*  
699 *aureus* (5). The absence of the UCCCC motif in the *S. epidermidis* and *S. capitis* *icaR*  
700 3'UTRs may account for the differences in IcaR expression, which may be attributed to  
701 the lack of interaction between the 5' and 3'UTRs in these chimeras. To test this  
702 hypothesis, we carried out electrophoretic mobility shift assays (EMSAs). The *S.*  
703 *aureus* *icaR* 5'UTR was synthesized *in vitro*, radioactively labelled, and then incubated  
704 with increasing concentrations of cold *icaR* 3'UTRs from the same five staphylococcal  
705 species analysed in Figure 4. Figure 6A shows that the *S. aureus* *icaR* 3'UTR

706 produced the expected gel-shifts when interacting with the 5'UTR, as previously  
707 described (5). Consistently, a similar behaviour was observed when using the *S.*  
708 *simiae* 3'UTR instead of the *S. aureus icaR* 3'UTR (Fig. 6B). In the case of the *S.*  
709 *argenteus icaR* 3'UTR, despite sharing a relatively high sequence homology with its *S.*  
710 *aureus* counterpart (82% of identity), a much lower binding affinity was found (Fig. 6B).  
711 This indicated that although *S. argenteus* preserved the UCCCC motif, the nucleotide  
712 differences might affect the binding affinity *in vitro*. *S. epidermidis* and *S. capitis*  
713 3'UTRs, as one would expect, did not produce gel shifts when incubated with the  
714 5'UTR due to the absence of the UCCCC motif in their sequences (Fig. 6C). Moreover,  
715 recent results have showed that the 5'UTR and 3'UTR from *S. epidermidis*, unlike what  
716 observed for *S. aureus*, do not interact between them, excluding the existence of an  
717 orthologous mechanism (58). In contrast, it was found that the *S. epidermidis*-specific  
718 sRNA IcaZ, downstream of *icaR* 3'UTR, targets *icaR* 5'UTR and hampers *icaR* mRNA  
719 translation (58). Altogether, these results indicate that sequence variations around the  
720 3'UTR determine the presence or absence of specific motifs and, ultimately, the  
721 generation of functional distinctions.

722 In order to demonstrate that the differences observed in terms of IcaR expression by  
723 chimeric mRNAs were enough to alter PIA-PNAG biosynthesis and, thus, be of  
724 biological relevance, we introduced the plasmids expressing these chimeric mRNAs  
725 into the *S. aureus* 15981 strain, which produces high levels of PIA-PNAG  
726 exopolysaccharide. The quantification of PIA-PNAG levels by dot-blot using specific  
727 antibodies against the exopolysaccharide showed that the strains expressing higher  
728 levels of IcaR (chimeric Scap/Sepi and  $\Delta$ 3'UTRs strains, which lacked the UCCCC  
729 motifs) presented lower levels of PIA-PNAG in comparison with the strains producing  
730 IcaR from mRNAs containing the UCCCC motif (WT and chimeric Sarg/Ssim strains)  
731 (Fig. 6D). These results confirmed that evolutionary variations in 3'UTRs, which occur  
732 among phylogenetically-related bacteria, are enough to create gene expression  
733 differences that affect important biological processes, such as exopolysaccharide

734 production. This suggested that the synthesis of orthologous proteins might be  
735 differentially affected by 3'UTR sequence variations.

736

737 **3'UTR variability among mRNAs encoding orthologous proteins is a common**  
738 **trait in bacteria**

739 Having proved that the inter-species 3'UTR variability has consequences on gene  
740 expression in *S. aureus*, we decided to extend the 3'UTR conservation analysis to  
741 other bacterial genera. First, we focused on several functional 3'UTRs previously  
742 described in the literature (4, 6-8). We compared the last 200 nt of the CDSs plus the  
743 entire 3'UTR sequences from the *E. coli acnB*, *Y. pestis hmsT*, *C. glutamicum aceA*,  
744 and *B. subtilis hbs* mRNAs with those of their corresponding phylogenetically-related  
745 species (Fig. S11A to S14A). Although the CDSs were highly conserved among all  
746 closely-related species, there were 3'UTR sequence variations in all four gene  
747 examples (Fig. 7). The *E. coli acnB* 3'UTR was conserved in *Citrobacter*, *Enterobacter*,  
748 and *Klebsiella* species with only a few nucleotide variations. In contrast, and according  
749 to synteny analysis by SyntTax (59), the *Salmonella*, *Cronobacter*, and *Serratia acnB*  
750 3'UTRs sequences varied due to gene rearrangements (Fig. S11B). For example, in *S.*  
751 *enterica* serovar Typhimurium SL1344, an insertion of the SL1344\_0160 gene (which  
752 encodes a putative HNH restriction endonuclease) deleted the sequence downstream  
753 of the *acnB* CDS and included a different TT. As a result, the *Salmonella acnB* 3'UTR  
754 is only 34 nt long and is missing the stem-loop recognized by the apo-AcnB protein (7).  
755 Sequence variations were also present in the *C. glutamicum aceA*, *B. subtilis hbs*, and  
756 *Yersinia hmsT* 3'UTRs. Although the *hbs* and *hmsT* CDSs are widely distributed across  
757 the *Bacillus* and *Yersinia* genera, their 3'UTRs were only conserved in the closest  
758 species: *B. amyloliquefaciens*, *B. licheniformis*, and *B. pumilus* in the former and *Y.*  
759 *pestis* biovars and *Y. pseudotuberculosis* in the latter (Fig. 7). Finally, none of the 10  
760 compared species of the genus *Corynebacterium* that carried the *aceA* CDS presented  
761 a similar sequence to that of the *C. glutamicum aceA* 3'UTR (Fig. 7). We then carried

762 out the prediction of Rho-independent transcriptional terminator structures in the non-  
763 conserved 3'UTRs to determine the length variability (Fig. 7 and Fig. S11 to S14). Most  
764 of these 3'UTRs variations are explained by the gene rearrangements occurring  
765 downstream of CDSs, with the exception of the *hbs* 3'UTR, in which variations may  
766 have occurred locally (Fig. S11 to S14).

767 Finally, to further investigate whether 3'UTR variations could also be distributed among  
768 other genes of *E. coli* and *B. subtilis*, as we observed in *S. aureus*, we performed  
769 similar genome-wide 3'UTR conservation analyses. Since the 3'-end transcript  
770 boundaries were recently released for *E. coli* and *B. subtilis* (28, 29), we generated a  
771 query database including the sequences of all mapped 3'UTRs plus the last 200 nt of  
772 their corresponding CDSs to detect orthologous genes among the different species.  
773 Using blastn, we compared the *E. coli* and *B. subtilis* databases against representative  
774 genomes of species from the *Enterobacteriaceae* family and the genus *Bacillus*,  
775 respectively. We found a similar outcome to the one observed for the genus  
776 *Staphylococcus*, meaning that 3'UTR sequences were also variable among mRNAs  
777 encoding orthologous proteins (Fig. 8, Fig. S15, Table S7 and S8). The bacterial  
778 species of the *Enterobacteriaceae* family showed a higher number of conserved  
779 3'UTRs in comparison with the *Bacillus* species. This was probably due to the  
780 phylogenetic distances between the species in the former being shorter than the ones  
781 in the latter. However, both bacterial groups presented a very high variation rate of  
782 their 3'UTR sequences, as illustrated in Fig. 8 and Fig. S15, where most of the dots fall  
783 onto the horizontal line, which represents the protein stop codon. This indicates that  
784 nucleotide conservation was also lost downstream of most CDSs (Fig. S16). On the  
785 one hand, these comparative analyses confirmed that nucleotide conservation analysis  
786 might only be useful for detecting a small percentage of inter-species conserved  
787 3'UTRs. On the other hand, these results indicate that the 3'UTRs from orthologous  
788 genes underwent different evolutionary events, a phenomenon that seems to be  
789 widespread in bacteria.

790 **DISCUSSION**

791 Bacterial IGRs, which have been traditionally defined as the genomic regions between  
792 two CDSs, are known for being poorly conserved among phylogenetically-related  
793 bacteria (60-63). For this reason, they have often been regarded as “junk” genetic  
794 material lacking relevant biological functions. However, recent RNA-seq technologies  
795 have revealed that IGRs may be more complex than initially anticipated,  
796 accommodating a plethora of regulatory elements that are either transcribed  
797 independently as sRNAs or fully functional from within the UTRs of mRNAs. In recent  
798 years, several studies have shown the biological relevance of 3'UTRs in both bacteria  
799 and eukaryotic organisms (for recent reviews see (18, 19, 64)). Since the 3'UTR-  
800 regulatory elements are physically- and functionally-related to the adjacent CDSs  
801 through mRNAs, an evolutionary connection between them might be expected in  
802 phylogenetically-related species. This indicates that the 3'UTR sequences would be  
803 preserved to the same degree as their corresponding CDSs. However, in this study, we  
804 observed this to be true in only a few examples, since most of the 3'UTRs from mRNAs  
805 encoding orthologous proteins, in closely-related bacterial species, display different  
806 lengths, and sequences. The idea of synthesizing hundreds of nucleotides at the 3'  
807 ends without a relevant function seems contradictory as bacteria tend to efficiently  
808 manage energy and clones with a lower fitness are rapidly eliminated from populations.  
809 Therefore, an explanation for this high 3'UTR variation rate between related species  
810 should be sought. First, it is important to highlight that 3'UTR sequences are more  
811 highly conserved in comparison with non-transcribed regions of strains from the same  
812 bacterial species (5). This implies that an evolutionary pressure to preserve 3'UTR  
813 sequences in a particular species exists. In addition, our data indicates that specific  
814 inter-species variations in 3'UTR sequences often occur. Therefore, these variations  
815 could differentially affect the expression of orthologous proteins from different species.  
816 This could be a strategy used by bacteria to create diversity without changing the CDS,  
817 an idea supported by previous studies that show how bacterial genes can shift rapidly

818 between multiple regulatory nodes. Oren *et al.* demonstrated the existence of  
819 promoter-regulatory regions that changed among orthologous genes, contributing to  
820 expression divergence and conferring distinct fitness advantages (65). Analogously,  
821 the finding that 3'UTR sequence variations can affect protein expression depending on  
822 the species may be an evolutionarily-selected trait to create species-specific post-  
823 transcriptional regulatory processes (16, 18, 66). Since the 5'UTRs are closely related  
824 with the initiation of translation, any significant sequence variations could impair protein  
825 synthesis and, thus, explain the bias in variability existing in 3'UTRs. This hypothesis is  
826 also in agreement with the role of 3'UTRs in eukaryotes, which have been specifically  
827 targeted throughout the course of evolution to contribute to the divergence of species  
828 by accumulating regulatory elements in their sequences (64, 67, 68). It is worth noting  
829 that longer 3'UTRs (including multiple miRNA target sites) are preferred in functionally  
830 critical eukaryotic genes that are spatially or temporally expressed. In contrast,  
831 housekeeping genes are selected to have shorter 3'UTRs (69). Strikingly, the 3'UTRs  
832 of *Salmonella hilD* and *S. aureus icaR* mRNAs (which encode the main transcriptional  
833 regulators of the SP-1 pathogenicity island and biofilm formation, respectively) are  
834 among the longest *bona fide* 3'UTRs found in bacteria. Resembling the eukaryotic  
835 process in which mRNAs are targeted by miRNAs, it has been recently shown that *hilD*  
836 and *icaR* 3'UTRs are targeted by sRNAs (70, 71). In the former, the Spot 42 sRNA  
837 (transcriptionally repressed by CRP-cAMP) targets the *hilD* 3'UTR, exerting a positive  
838 effect on HilD expression (70, 72). In the latter, the RsaI sRNA modulates PIA-PNAG  
839 synthesis by interacting with the 3'UTR of the *icaR* mRNA (71). The RsaI interacting  
840 region is located between the UCCCC motif, which is required for the *icaR* 3'-5'UTR  
841 interaction, and the transcriptional terminator of the *icaR* mRNA (71). Interestingly,  
842 while the UCCCC motif is conserved among the *S. aureus*, *S. argenteus*, and *S. simiae*  
843 *icaR* mRNAs, the RsaI pairing region is only conserved in the *S. aureus icaR* mRNA.  
844 Note that RsaI sRNA is conserved in most species of the genus *Staphylococcus*. Spot  
845 42 sRNA is present in several bacterial genera, while the *hilD* gene is exclusive to

846 *Salmonella* (73). Overall, this shows that the 3'UTRs can accumulate more than one  
847 regulatory motif, which may be differentially selected during evolution.

848 The acquisition of *icaR* 3'UTR variations may be explained by gene rearrangements  
849 that created mRNA chimeras among staphylococcal species. The *S. aureus* *icaR* CDS  
850 is only present in 5 out of 9 staphylococcal species that encode the *icaADBC* operon. It  
851 is interesting to note that in the four species lacking the *icaR* CDS, depending on the  
852 species, either DspB (hexosaminidase), a TetR-like regulator, or proteins of a two-  
853 component system are encoded in place of the *icaR* gene. These different genomic  
854 organizations indicate that the regulatory locus evolved independently from the  
855 *icaADBC* operon and may also explain the appearance of 3'UTR variations in the *icaR*  
856 mRNA when such a locus was acquired or mobilized. Similarly, gene rearrangements  
857 could also be responsible for the disparity of sRNA content in phylogenetically-related  
858 bacterial species and may also justify some of the 3'UTR variations (49).

859 In addition, there may be other phenomena beyond gene rearrangements behind  
860 3'UTR variations, since only a minor percentage of them (around 28-35%) fall under  
861 this category (Fig. 3). It is likely that mobile insertion sequences and/or repeated  
862 sequences, such as those found in the IGRs of *S. aureus*, change the functionality of  
863 the 3'UTRs of orthologous mRNAs (52, 74, 75). Interestingly, we showed that the  
864 presence of STAR elements and ISs interrupting 3'UTR sequences affects *S. aureus*  
865 RpiRc expression (Fig. 5). This suggests that alternative IS transposition/excision  
866 processes may generate local 3'UTR variations that can be evolutionarily fixed if they  
867 produce an advantage for a specific clone in comparison to the rest of the bacterial  
868 population. Alternatively, ISs might be also responsible for phase variation  
869 mechanisms (76, 77). Here, we showed that the presence of the IS1181 inserted in the  
870 *rpiRc* 3'UTR of several *S. aureus* strains reduced haemolysin production by increasing  
871 the levels of the RNAIII-repressor RpiRc (Fig. 5). Further studies would be required to  
872 determine whether the transposition of ISs into the *rpiRc* 3'UTR provide an additional

873 way to regulate *S. aureus* virulence through the *agr* locus, as seen in the cases of *agr*-  
874 phase-variation (78).

875 Some of the genomic repeats include stem-loops, which serve as transcriptional  
876 terminators (Fig. 5) or provide specific targets for RNase III (79). Moreover, changes in  
877 the 3'UTR sequence could also affect mRNA turnover by modifying the 3'UTR  
878 accessibility to exoribonucleases. This indicates that the accumulation/elimination of  
879 ribonuclease target sites could be an alternative way to specifically modulate the  
880 expression of a protein at the post-transcriptional level in a particular species. In  
881 agreement with this, the deletion of 3'UTRs (e.g. *hilD*, *aceA*, *hmsT*, and *icaR*), which  
882 eliminates the putative ribonuclease target sites, increases protein expression (4-6, 9,  
883 21). Similarly, the presence of specific target sites for RNA-binding proteins in 3'UTRs  
884 could also contribute to regulatory differences among bacterial species (7, 11).

885 In addition to the differential effect on orthologous genes, it should be also considered  
886 that variations in 3'UTR sequences may affect the expression of downstream genes.  
887 For example, if variations in the 3'UTR sequence alter the strength of the  
888 corresponding TT or even eliminate it (e.g. Figure 4 shows that transcriptional read-  
889 through occurs in *ftnA* of some staphylococcal species), transcription would continue to  
890 increase the expression of downstream genes or create antisense RNAs if the  
891 downstream gene is encoded in the opposite direction (80). Moreover, if a downstream  
892 gene is close enough, variations in the 3'UTR sequence could also alter promoter  
893 regulatory elements of the next transcription unit, thereby affecting their expression.

894 Further studies are needed to improve our knowledge regarding how sequence  
895 variations in 3'UTRs may specifically affect gene expression in different bacterial  
896 species. The development of RNA sequencing techniques, mapping the precise 3'-  
897 ends of transcripts, as well as ribonuclease cleavage regions and RNA-protein binding  
898 sites will considerably contribute to this (28, 81-83). Combining such information with  
899 studies like the one presented here will help improve our understanding of the

900 evolutionary features responsible for the generation of species-specific sequences that  
901 may, ultimately, lead to bacterial species differentiation.

902

903

#### 904 **DATA AVAILABILITY**

905 Transcriptomic maps of *S. simiae* CCM 7213T, *S. capitis* SK14, and *S. epidermidis*  
906 RP62A strains are available at <http://rnamaps.unavarra.es/>. Other data or material that  
907 support the findings of this study can be made available by the corresponding author  
908 upon request.

909

#### 910 **SUPPLEMENTARY MATERIAL**

911 An appendix PDF file including Supplementary Figures S1 to S16, Supplementary  
912 Tables S1 to S8, and Supplementary References is available at NAR Online.

913

#### 914 **ACKNOWLEDGMENTS**

915 We thank Dr. Jaione Valle, Prof. Cristina Solano, Prof. Iñigo Lasa, Prof. Josep  
916 Casadesús, and Prof. Pascale Romby for their excellent advice and their critical  
917 reading of the manuscript; Prof. Tomas Maira-Litrán for the anti-PIA-PNAG antibodies;  
918 B. Garcia for her helpful assistance with RNA extraction. We also thank the following  
919 organizations for providing strains and plasmids: Network on Antimicrobial Resistance  
920 in *Staphylococcus aureus* (NARSA) for providing pCN40 (NR-46131) and pCN47 (Ref.  
921 NR-46141) plasmids constructed by Prof. Emmanuelle Charpentier and  
922 *Staphylococcus aureus* strains N315 (NR-45898) and *Staphylococcus epidermidis*  
923 strain RP62A (Ref. ATCC 35984) through BEI Resources, NIAID, NIH and the Human  
924 Microbiome Project for providing *Staphylococcus capitis* strain SK14 (Ref HM-117) also  
925 through BEI Resources, NIAID, NIH.

926

927

## 928 FUNDING

929 This work was supported by the European Research Council (ERC) under the  
930 European Union's Horizon 2020 research and innovation program (Grant Agreement  
931 No. 646869), and the Spanish Ministry of Economy and Competitiveness (BFU2014-  
932 56698-P). C.J.C. was supported by a predoctoral contract from the Public University of  
933 Navarre. I.C. was supported by recurrent funding from Centre National de la  
934 Recherche Scientifique (CNRS) and by a PICS program ("Projet International de  
935 Coopération Scientifique", No. PICS07507, CNRS) between France and Spain.

936

937

## 938 REFERENCES

- 939 1. Waters,L.S. and Storz,G. (2009) Regulatory RNAs in Bacteria. *Cell*, **136**, 615–628.
- 940 2. Gripenland,J., Netterling,S., Loh,E., Tiensuu,T., Toledo-Arana,A. and Johansson,J.  
941 (2010) RNAs: regulators of bacterial virulence. *Nat Rev Micro*, **8**, 857–866.
- 942 3. Bouloc,P. and Repoila,F. (2016) Fresh layers of RNA-mediated regulation in Gram-  
943 positive bacteria. *Current Opinion in Microbiology*, **30**, 30–35.
- 944 4. Maeda,T. and Wachi,M. (2012) 3' untranslated region-dependent degradation of the  
945 *aceA* mRNA, encoding the glyoxylate cycle enzyme isocitrate lyase, by RNase E/G  
946 in *Corynebacterium glutamicum*. *Appl Environ Microbiol*, **78**, 8753–8761.
- 947 5. Ruiz de Los Mozos,I., Vergara-Irigaray,M., Segura,V., Villanueva,M., Bitarte,N.,  
948 Saramago,M., Domingues,S., Arraiano,C.M., Fechter,P., Romby,P., *et al.* (2013)  
949 Base pairing interaction between 5'- and 3'-UTRs controls *icaR* mRNA translation  
950 in *Staphylococcus aureus*. *PLoS Genet*, **9**, e1004001.
- 951 6. Zhu,H., Mao,X.-J., Guo,X.-P. and Sun,Y.-C. (2015) The *hmsT* 3' untranslated region  
952 mediates c-di-GMP metabolism and biofilm formation in *Yersinia pestis*. *Mol*  
953 *Microbiol*, 10.1111/mmi.13301.
- 954 7. Benjamin,J.-A.M. and Massé,E. (2014) The iron-sensing aconitase B binds its own  
955 mRNA to prevent sRNA-induced mRNA cleavage. *Nucleic Acids Res*, **42**, 10023–  
956 10036.
- 957 8. Braun,F., Durand,S. and Condon,C. (2017) Initiating ribosomes and a 5'/3'-UTR  
958 interaction control ribonuclease action to tightly couple *B. subtilis hbs* mRNA  
959 stability with translation. *Nucleic Acids Res*, **45**, 11386–11400.
- 960 9. Zhao,J.-P., Zhu,H., Guo,X.-P. and Sun,Y.-C. (2018) AU-Rich Long 3' Untranslated  
961 Region Regulates Gene Expression in Bacteria. *Front Microbiol*, **9**, 633–10.

- 962 10. Kawano,M., Reynolds,A.A., Miranda-Rios,J. and Storz,G. (2005) Detection of 5'-  
963 and 3'-UTR-derived small RNAs and *cis*-encoded antisense RNAs in *Escherichia*  
964 *coli*. *Nucleic Acids Res*, **33**, 1040–1050.
- 965 11. Chao,Y., Papenfort,K., Reinhardt,R., Sharma,C.M. and Vogel,J. (2012) An atlas of  
966 Hfq-bound transcripts reveals 3' UTRs as a genomic reservoir of regulatory small  
967 RNAs. *EMBO J*, **31**, 4005–4019.
- 968 12. Peng,T., Berghoff,B.A., Oh,J.-I., Weber,L., Schirmer,J., Schwarz,J., Glaeser,J. and  
969 Klug,G. (2016) Regulation of a polyamine transporter by the conserved 3' UTR-  
970 derived sRNA SorX confers resistance to singlet oxygen and organic  
971 hydroperoxides in *Rhodobacter sphaeroides*. *RNA Biol*, **13**, 988–999.
- 972 13. Chao,Y. and Vogel,J. (2016) A 3' UTR-derived small RNA provides the regulatory  
973 noncoding arm of the inner membrane stress response. *Mol Cell*, **61**, 352–363.
- 974 14. Guo,M.S., Updegrove,T.B., Gogol,E.B., Shabalina,S.A., Gross,C.A. and Storz,G.  
975 (2014) MicL, a new  $\sigma$ E-dependent sRNA, combats envelope stress by repressing  
976 synthesis of Lpp, the major outer membrane lipoprotein. *Genes & Development*,  
977 **28**, 1620–1634.
- 978 15. Updegrove,T.B., Kouse,A.B., Bandyra,K.J. and Storz,G. (2018) Stem-loops direct  
979 precise processing of 3' UTR-derived small RNA MicL. *Nucleic Acids Res*,  
980 10.1093/nar/gky1175.
- 981 16. Miyakoshi,M., Matera,G., Maki,K., Sone,Y. and Vogel,J. (2018) Functional  
982 expansion of a TCA cycle operon mRNA by a 3' end-derived small RNA. *Nucleic*  
983 *Acids Res*, **250**, 1727–14.
- 984 17. De Mets,F., van Melderden,L. and Gottesman,S. (2019) Regulation of acetate  
985 metabolism and coordination with the TCA cycle via a processed small RNA.  
986 *Proceedings of the National Academy of Sciences*, **116**, 1043–1052.
- 987 18. Miyakoshi,M., Chao,Y. and Vogel,J. (2015) Regulatory small RNAs from the 3'  
988 regions of bacterial mRNAs. *Current Opinion in Microbiology*, **24**, 132–139.
- 989 19. Ren,G.-X., Guo,X.-P. and Sun,Y.-C. (2017) Regulatory 3' untranslated regions of  
990 bacterial mRNAs. *Front Microbiol*, **8**, 633–6.
- 991 20. Ellermeier,J.R. and Slauch,J.M. (2007) Adaptation to the host environment:  
992 regulation of the SPI1 type III secretion system in *Salmonella enterica* serovar  
993 Typhimurium. *Current Opinion in Microbiology*, **10**, 24–29.
- 994 21. López-Garrido,J., Puerta-Fernández,E. and Casadesús,J. (2014) A eukaryotic-like  
995 3' untranslated region in *Salmonella enterica* *hilD* mRNA. *Nucleic Acids Res*, **42**,  
996 5894–5906.
- 997 22. Gerstmeir,R., Wendisch,V.F., Schnicke,S., Ruan,H., Farwick,M., Reinscheid,D. and  
998 Eikmanns,B.J. (2003) Acetate metabolism and its regulation in *Corynebacterium*  
999 *glutamicum*. *Journal of Biotechnology*, **104**, 99–122.
- 1000 23. Novick,R.P., Ross,H.F., Projan,S.J., Kornblum,J., Kreiswirth,B. and Moghazeh,S.  
1001 (1993) Synthesis of staphylococcal virulence factors is controlled by a regulatory  
1002 RNA molecule. *EMBO J*, **12**, 3967–3975.
- 1003 24. Bronesky,D., Wu,Z., Marzi,S., Walter,P., Geissmann,T., Moreau,K.,

- 1004 Vandenesch,F., Caldelari,I. and Romby,P. (2016) *Staphylococcus aureus* RNAIII  
1005 and Its Regulon Link Quorum Sensing, Stress Responses, Metabolic Adaptation,  
1006 and Regulation of Virulence Gene Expression. **70**, 299–316.
- 1007 25. Skinner,M.E., Uzilov,A.V., Stein,L.D., Mungall,C.J. and Holmes,I.H. (2009)  
1008 JBrowse: a next-generation genome browser. *Genome Res*, **19**, 1630–1638.
- 1009 26. Lasa,I., Toledo-Arana,A., Dobin,A., Villanueva,M., de los Mozos,I.R., Vergara-  
1010 Irigaray,M., Segura,V., Fagegaltier,D., Penadés,J.R., Valle,J., *et al.* (2011)  
1011 Genome-wide antisense transcription drives mRNA processing in bacteria. *Proc*  
1012 *Natl Acad Sci USA*, **108**, 20172–20177.
- 1013 27. Koch,G., Yepes,A., Förstner,K.U., Wermser,C., Stengel,S.T., Modamio,J.,  
1014 Ohlsen,K., Foster,K.R. and Lopez,D. (2014) Evolution of resistance to a last-resort  
1015 antibiotic in *Staphylococcus aureus* via bacterial competition. *Cell*, **158**, 1060–  
1016 1071.
- 1017 28. Dar,D., Shamir,M., Mellin,J.R., Koutero,M., Stern-Ginossar,N., Cossart,P. and  
1018 Sorek,R. (2016) Term-seq reveals abundant ribo-regulation of antibiotics  
1019 resistance in bacteria. *Science*, **352**, aad9822–aad9822.
- 1020 29. Dar,D. and Sorek,R. (2018) High-resolution RNA 3'-ends mapping of bacterial Rho-  
1021 dependent transcripts. *Nucleic Acids Res*, **46**, 6797–6805.
- 1022 30. Toledo-Arana,A., Dussurget,O., Nikitas,G., Sesto,N., Guet-Revillet,H.,  
1023 Balestrino,D., Loh,E., Gripenland,J., Tiensuu,T., Vaitkevicius,K., *et al.* (2009) The  
1024 *Listeria* transcriptional landscape from saprophytism to virulence. *Nature*, **459**,  
1025 950–956.
- 1026 31. Tjaden,B. (2015) De novo assembly of bacterial transcriptomes from RNA-seq  
1027 data. *Genome Biol*, **16**, 1.
- 1028 32. Kingsford,C.L., Ayanbule,K. and Salzberg,S.L. (2007) Rapid, accurate,  
1029 computational discovery of Rho-independent transcription terminators illuminates  
1030 their relationship to DNA uptake. *Genome Biol*, **8**, R22.
- 1031 33. Britton,R.A., Wen,T., Schaefer,L., Pellegrini,O., Uicker,W.C., Mathy,N., Tobin,C.,  
1032 Daou,R., Szyk,J. and Condon,C. (2007) Maturation of the 5' end of *Bacillus subtilis*  
1033 16S rRNA by the essential ribonuclease YkqC/RNase J1. *Mol Microbiol*, **63**, 127–  
1034 138.
- 1035 34. Arnaud,M., Chastanet,A. and Debarbouille,M. (2004) New vector for efficient allelic  
1036 replacement in naturally nontransformable, low-GC-Content, Gram-positive  
1037 bacteria. *Appl Environ Microbiol*, **70**, 6887–6891.
- 1038 35. Caballero,C.J., Menendez-Gil,P., Catalan-Moreno,A., Vergara-Irigaray,M.,  
1039 García,B., Segura,V., Irurzun,N., Villanueva,M., Ruiz de Los Mozos,I., Solano,C.,  
1040 *et al.* (2018) The regulon of the RNA chaperone CspA and its auto-regulation in  
1041 *Staphylococcus aureus*. *Nucleic Acids Res*, **46**, 1345–1361.
- 1042 36. Lee,J.C. (1995) Electrotransformation of Staphylococci. *Methods Mol Biol*, **47**, 209–  
1043 216.
- 1044 37. Charpentier,E., Anton,A.I., Barry,P., Alfonso,B., Fang,Y. and Novick,R.P. (2004)  
1045 Novel cassette-based shuttle vector system for Gram-positive bacteria. *Appl*  
1046 *Environ Microbiol*, **70**, 6076–6085.

- 1047 38. Balestrino,D., Hamon,M.A., Dortet,L., Nahori,M.-A., Pizarro-Cerda,J., Alignani,D.,  
1048 Dussurget,O., Cossart,P. and Toledo-Arana,A. (2010) Single-cell techniques using  
1049 chromosomally tagged fluorescent bacteria to study *Listeria monocytogenes*  
1050 infection processes. *Appl Environ Microbiol*, **76**, 3625–3636.
- 1051 39. Siguier,P., Perochon,J., Lestrade,L., Mahillon,J. and Chandler,M. (2006) ISfinder:  
1052 the reference centre for bacterial insertion sequences. *Nucleic Acids Res*, **34**,  
1053 D32–6.
- 1054 40. Cramton,S.E., Gerke,C., Schnell,N.F., Nichols,W.W. and Götz,F. (1999) The  
1055 intercellular adhesion (*ica*) locus is present in *Staphylococcus aureus* and is  
1056 required for biofilm formation. *Infect Immun*, **67**, 5427–5433.
- 1057 41. Maira-Litrán,T., Kropec,A., Goldmann,D.A. and Pier,G.B. (2005) Comparative  
1058 opsonic and protective activities of *Staphylococcus aureus* conjugate vaccines  
1059 containing native or deacetylated staphylococcal poly-N-acetyl- $\beta$ -(1-6)-  
1060 glucosamine. *Infect Immun*, **73**, 6752–6762.
- 1061 42. Johnson,M., Zaretskaya,I., Raytselis,Y., Merezhuk,Y., McGinnis,S. and  
1062 Madden,T.L. (2008) NCBI BLAST: a better web interface. *Nucleic Acids Res*, **36**,  
1063 W5–9.
- 1064 43. Hedges,S.B., Marin,J., Suleski,M., Paymer,M. and Kumar,S. (2015) Tree of life  
1065 reveals clock-like speciation and diversification. *Mol. Biol. Evol.*, **32**, 835–845.
- 1066 44. Cramton,S.E., Schnell,N.F., Götz,F. and Brückner,R. (2000) Identification of a new  
1067 repetitive element in *Staphylococcus aureus*. *Infect Immun*, **68**, 2344–2348.
- 1068 45. Bohn,C., Rigoulay,C., Chabelskaya,S., Sharma,C.M., Marchais,A., Skorski,P.,  
1069 Borezée-Durant,E., Barbet,R., Jacquet,E., Jacq,A., *et al.* (2010) Experimental  
1070 discovery of small RNAs in *Staphylococcus aureus* reveals a riboregulator of  
1071 central metabolism. *Nucleic Acids Res*, **38**, 6620–6636.
- 1072 46. Beaume,M., Hernandez,D., Farinelli,L., Deluen,C., Linder,P., Gaspin,C., Romby,P.,  
1073 Schrenzel,J. and Francois,P. (2010) Cartography of Methicillin-resistant *S. aureus*  
1074 transcripts: detection, orientation and temporal expression during growth phase  
1075 and stress conditions. *PLoS ONE*, **5**, e10725.
- 1076 47. Carroll,R.K., Weiss,A., Broach,W.H., Wiemels,R.E., Mogen,A.B., Rice,K.C. and  
1077 Shaw,L.N. (2016) Genome-wide annotation, identification, and global  
1078 transcriptomic analysis of regulatory or small RNA gene expression in  
1079 *Staphylococcus aureus*. *mBio*, **7**, e01990–15–13.
- 1080 48. Darling,A.C.E., Mau,B., Blattner,F.R. and Perna,N.T. (2004) Mauve: multiple  
1081 alignment of conserved genomic sequence with rearrangements. *Genome Res*, **14**,  
1082 1394–1403.
- 1083 49. Raghavan,R., Kacharia,F.R., Millar,J.A., Sislak,C.D. and Ochman,H. (2015)  
1084 Genome rearrangements can make and break small RNA genes. *Genome Biology  
1085 and Evolution*, **7**, 557–566.
- 1086 50. Joseph,S.J., Li,B., Petit,R.A., Qin,Z.S., Darrow,L. and Read,T.D. (2016) The single-  
1087 species metagenome: Subtyping *Staphylococcus aureus* core genome sequences  
1088 from shotgun metagenomic data. *PeerJ*, **2016**, e2571.
- 1089 51. Zuker,M. (2003) Mfold web server for nucleic acid folding and hybridization

- 1090 prediction. *Nucleic Acids Res*, **31**, 3406–3415.
- 1091 52. Purves,J., Blades,M., Arafat,Y., Malik,S.A., Bayliss,C.D. and Morrissey,J.A. (2012)  
1092 Variation in the genomic locations and sequence conservation of STAR elements  
1093 among staphylococcal species provides insight into DNA repeat evolution. *BMC*  
1094 *Genomics*, **13**, 515.
- 1095 53. Chesneau,O., Lailier,R., Derbise,A. and Solh,EI,N. (1999) Transposition of IS1181  
1096 in the genomes of *Staphylococcus* and *Listeria*. *FEMS Microbiology Letters*, **177**,  
1097 93–100.
- 1098 54. Lyon,B.R., Gillespie,M.T. and Skurray,R.A. (1987) Detection and characterization  
1099 of IS256, and insertion sequence in *Staphylococcus aureus*. *J. Gen. Microbiol.*,  
1100 **133**, 3031–3038.
- 1101 55. Zhu,Y., Nandakumar,R., Sadykov,M.R., Madayiputhiya,N., Luong,T.T., Gaupp,R.,  
1102 Lee,C.Y. and Somerville,G.A. (2011) RpiR homologues may link *Staphylococcus*  
1103 *aureus* RNAIII synthesis and pentose phosphate pathway regulation. *J Bacteriol*,  
1104 **193**, 6187–6196.
- 1105 56. Gaupp,R., Wirf,J., Wonnenberg,B., Biegel,T., Eisenbeis,J., Graham,J.,  
1106 Herrmann,M., Lee,C.Y., Beisswenger,C., Wolz,C., *et al.* (2016) RpiRc Is a  
1107 pleiotropic effector of virulence determinant synthesis and attenuates pathogenicity  
1108 in *Staphylococcus aureus*. *Infect Immun*, **84**, 2031–2041.
- 1109 57. Balasubramanian,D., Ohneck,E.A., Chapman,J., Weiss,A., Kim,M.K., Reyes-  
1110 Robles,T., Zhong,J., Shaw,L.N., Lun,D.S., Ueberheide,B., *et al.* (2016)  
1111 *Staphylococcus aureus* coordinates Leukocidin expression and pathogenesis by  
1112 sensing metabolic fluxes via RpiRc. *mBio*, **7**, e00818–16–13.
- 1113 58. Lerch,M.F., Schoenfelder,S.M.K., Marincola,G., Wencker,F.D.R., Eckart,M.,  
1114 Förstner,K.U., Sharma,C.M., Thormann,K.M., Kucklick,M., Engelmann,S., *et al.*  
1115 (2019) A non-coding RNA from the intercellular adhesion (*ica*) locus of  
1116 *Staphylococcus epidermidis* controls polysaccharide intercellular adhesion (PIA)-  
1117 mediated biofilm formation. *Mol Microbiol*, 10.1111/mmi.14238.
- 1118 59. Oberto,J. (2013) SyntTax: a web server linking synteny to prokaryotic taxonomy.  
1119 *BMC Bioinformatics*, **14**, 4.
- 1120 60. Molina,N. and van Nimwegen,E. (2008) Universal patterns of purifying selection at  
1121 noncoding positions in bacteria. *Genome Res*, **18**, 148–160.
- 1122 61. Luo,H., Tang,J., Friedman,R. and Hughes,A.L. (2011) Ongoing purifying selection  
1123 on intergenic spacers in group A streptococcus. *Infect. Genet. Evol.*, **11**, 343–348.
- 1124 62. Sridhar,J., Sabarinathan,R., Balan,S.S., Rafi,Z.A., Gunasekaran,P. and Sekar,K.  
1125 (2011) Junker: An intergenic explorer for bacterial genomes. *Genomics*,  
1126 *Proteomics & Bioinformatics*, **9**, 179–182.
- 1127 63. Thorpe,H.A., Bayliss,S.C., Hurst,L.D. and Feil,E.J. (2017) Comparative analyses of  
1128 selection operating on nontranslated intergenic regions of diverse bacterial  
1129 species. *Genetics*, **206**, 363–376.
- 1130 64. Mayr,C. (2017) Regulation by 3'-Untranslated Regions. *Annu Rev Genet*, **51**, 171–  
1131 194.

- 1132 65. Oren, Y., Smith, M.B., Johns, N.I., Kaplan Zeevi, M., Biran, D., Ron, E.Z., Corander, J.,  
1133 Wang, H.H., Alm, E.J. and Pupko, T. (2014) Transfer of noncoding DNA drives  
1134 regulatory rewiring in bacteria. *Proc Natl Acad Sci USA*, **111**, 16112–16117.
- 1135 66. Updegrove, T.B., Shabalina, S.A. and Storz, G. (2015) How do base-pairing small  
1136 RNAs evolve? *FEMS Microbiology Reviews*, **39**, 379–391.
- 1137 67. Mazumder, B., Seshadri, V. and Fox, P.L. (2003) Translational control by the 3'-UTR:  
1138 the ends specify the means. *Trends Biochem Sci*, **28**, 91–98.
- 1139 68. Mayr, C. and Bartel, D.P. (2009) Widespread Shortening of 3'UTRs by Alternative  
1140 Cleavage and Polyadenylation Activates Oncogenes in Cancer Cells. *Cell*, **138**,  
1141 673–684.
- 1142 69. Cheng, C., Bhardwaj, N. and Gerstein, M. (2009) The relationship between the  
1143 evolution of microRNA targets and the length of their UTRs. *BMC Genomics*, **10**,  
1144 431.
- 1145 70. Mouali, E.I., Gaviria-Cantin, T., Sánchez-Romero, M.A., Gibert, M.,  
1146 Westermann, A.J., Vogel, J. and Balsalobre, C. (2018) CRP-cAMP mediates  
1147 silencing of *Salmonella* virulence at the post-transcriptional level. *PLoS Genet*, **14**,  
1148 e1007401.
- 1149 71. Bronesky, D., Desgranges, E., Corvaglia, A., Francois, P., Caballero, C.J., Prado, L.,  
1150 Toledo-Arana, A., Lasa, I., Moreau, K., Vandenesch, F., *et al.* (2019) A multifaceted  
1151 small RNA modulates gene expression upon glucose limitation in *Staphylococcus*  
1152 *aureus*. *EMBO J*, **38**, e99363.
- 1153 72. Mouali, E.I. and Balsalobre, C. (2019) 3' untranslated regions: regulation at the end  
1154 of the road. *Curr. Genet.*, **65**, 127–131.
- 1155 73. Bækkel, C. and Haugen, P. (2015) The Spot 42 RNA: A regulatory small RNA  
1156 with roles in the central metabolism. *RNA Biol*, **12**, 1071–1077.
- 1157 74. Hardy, K.J., Ussery, D.W., Oppenheim, B.A. and Hawkey, P.M. (2004) Distribution  
1158 and characterization of staphylococcal interspersed repeat units (SIRUs) and  
1159 potential use for strain differentiation. *Microbiology*, **150**, 4045–4052.
- 1160 75. Geissmann, T., Chevalier, C., Cros, M.-J., Boisset, S., Fechter, P., Noirot, C.,  
1161 Schrenzel, J., Francois, P., Vandenesch, F., Gaspin, C., *et al.* (2009) A search for  
1162 small noncoding RNAs in *Staphylococcus aureus* reveals a conserved sequence  
1163 motif for regulation. *Nucleic Acids Res*, **37**, 7239–7257.
- 1164 76. Ziebuhr, W., Krimmer, V., Rachid, S., Lößner, I., Götz, F. and Hacker, J. (1999) A novel  
1165 mechanism of phase variation of virulence in *Staphylococcus epidermidis*:  
1166 Evidence for control of the polysaccharide intercellular adhesin synthesis by  
1167 alternating insertion and excision of the insertion sequence element IS256. *Mol*  
1168 *Microbiol*, **32**, 345–356.
- 1169 77. Vandecraen, J., Chandler, M., Aertsen, A. and Van Houdt, R. (2017) The impact of  
1170 insertion sequences on bacterial genome plasticity and adaptability. *Crit. Rev.*  
1171 *Microbiol.*, **43**, 709–730.
- 1172 78. Gor, V., Takemura, A.J., Nishitani, M., Higashide, M., Medrano Romero, V.,  
1173 Ohniwa, R.L. and Morikawa, K. (2019) Finding of Agr phase variants in  
1174 *Staphylococcus aureus*. *mBio*, **10**.

- 1175 79. De Gregorio,E., Abrescia,C., Carlomagno,M.S. and Di Nocera,P.P. (2003)  
1176 Ribonuclease III-mediated processing of specific *Neisseria meningitidis* mRNAs.  
1177 *Biochem J*, **374**, 799–805.
- 1178 80. Lasa,I., Toledo-Arana,A. and Gingeras,T.R. (2012) An effort to make sense of  
1179 antisense transcription in bacteria. *RNA Biol*, **9**, 1039–1044.
- 1180 81. Chao,Y., Li,L., Girodat,D., Förstner,K.U., Said,N., Corcoran,C., Šmiga,M.,  
1181 Papenfort,K., Reinhardt,R., Wieden,H.-J., *et al.* (2017) In Vivo Cleavage Map  
1182 Illuminates the Central Role of RNase E in Coding and Non-coding RNA Pathways.  
1183 *Mol Cell*, **65**, 39–51.
- 1184 82. Altuvia,Y., Bar,A., Reiss,N., Karavani,E., Argaman,L. and Margalit,H. (2018) In vivo  
1185 cleavage rules and target repertoire of RNase III in *Escherichia coli*. *Nucleic Acids*  
1186 *Res*, **17**, 138–15.
- 1187 83. Holmqvist,E. and Vogel,J.X.R. (2018) RNA-binding proteins in bacteria. *Nat Rev*  
1188 *Micro*, 10.1038/s41579-018-0049-5.
- 1189
- 1190

1191 **Figures legends**

1192

1193 **Figure 1. High-throughput conservation analysis of 3'UTRs from mRNAs**  
1194 **encoding orthologous proteins among phylogenetically-related staphylococcal**

1195 **species. (A)** Scatter plots representing the conservation of 3' end regions of *S. aureus*

1196 mRNAs compared to different closely-related staphylococcal species. *S. aureus* NCTC

1197 8325 3'UTR sequence query database was compared by blastn to *S. aureus* MW2, *S.*

1198 *simiae* NCTC13838, *S. capitis* AYP1020, and *S. epidermidis* RP62A genome

1199 sequences. Each dot represents the last conserved nucleotide (y axis) of a specific

1200 species in function of the *S. aureus* 3'UTR length (x axis). Based on the transcriptomic

1201 maps previously described (5), each query sequence included the last 200 nucleotides

1202 of the corresponding CDS plus the whole 3'UTR of monocistronic mRNAs. The plot

1203 was coloured by applying the Kernel density estimation, which indicates the proximity

1204 of the dots: blue, isolated dots; red, overlapping dots. The number (n) of plotted

1205 mRNAs encoding orthologous proteins and the square correlation coefficient ( $R^2$ ) are

1206 indicated. **(B)** Histogram plots showing the distribution of the last conserved nucleotide

1207 position (blue bars) among the *S. aureus* monocistronic mRNAs compared to the

1208 indicated staphylococcal species. Each blue bar represents the number of conserved

1209 3'UTRs at a given position in windows of 10 nt (width of the bar). Grey bars represent

1210 the *S. aureus* distribution, which is included as a reference. The dashed red line

1211 indicates the position of the stop codon.

1212

1213 **Figure 2. The lengths of the 3'UTRs do not correlate among staphylococcal**

1214 **species. (A)** Scatter plots representing the real 3'UTR lengths of each staphylococcal

1215 species in function of the 3'UTR lengths of the corresponding orthologous *S. aureus*

1216 mRNA. The length of the 3'UTRs from *S. aureus* orthologous monocistronic mRNAs

1217 were annotated by combining the transcriptomic data and the prediction of Rho-

1218 independent transcriptional terminators by TransTermHP (32). Only 3'UTRs shorter

1219 than 500 nt are represented. The plot was coloured by applying the Kernel density  
1220 estimation, which indicates the proximity of the dots: blue, isolated dots; red,  
1221 overlapping dots. **(B)** Plot representing the 3'UTR length of relevant orthologous genes  
1222 in the indicated staphylococcal species. **(C)** Plot representing the 3'UTR conservation  
1223 length of the orthologous genes analysed in **B**, which was determined by the blastn  
1224 algorithm. RNAIII is included as an example of an mRNA with a highly conserved  
1225 3'UTR. **(D)** Browser images showing the RNA transcribed from the *icaR*, *ftnA*, and  
1226 *rpiRc* chromosomal regions of the *S. aureus*, *S. simiae*, *S. epidermidis*, and *S. capitis*  
1227 strains. The RNA-seq data was mapped to the corresponding genomes, and the  
1228 transcriptomic maps were loaded onto a server based on Jbrowse (25). The complete  
1229 transcriptomic maps are available at <http://rnamaps.unavarra.es/>.

1230

1231 **Figure 3. Nucleotide sequence variation occurring downstream of orthologous**  
1232 **CDSs may explain 3'UTR diversity. (A)** Schematic representation of the conservation  
1233 analysis performed on IGRs and CDSs located downstream of an orthologous CDS.  
1234 Whole-genome comparisons between *S. aureus* and its phylogenetically-related  
1235 species were performed using Mauve (48). Values 1 and 0 were assigned to  
1236 conserved and non-conserved IGRs/CDSs, respectively, and a different colour was  
1237 attributed depending on the conservation configuration. Blue: the IGR and CDS  
1238 downstream of the orthologous CDS are conserved; orange: the downstream CDS is  
1239 conserved but not the IGR; green: both downstream regions are not conserved, and  
1240 grey: no orthologous CDS was found in the analysed species. **(B)** Pie chart quantifying  
1241 the different categories represented in **A**. **(C)** Plot showing the length of the IGR  
1242 downstream of each orthologous CDS in the different species compared to that of *S.*  
1243 *aureus*. **(D)** Plot showing the length of the orthologous CDS in the different species  
1244 compared to that of *S. aureus*. Note that only the IGRs and CDSs that fall under the  
1245 orange category are plotted in **C** and **D**.

1246

1247 **Figure 4. Changes in 3'UTRs sequences affect protein expression. (A)** Schematic  
1248 representation of the conservation analysis of the *icaR*, *ftnA*, and *rpiRc* mRNAs among  
1249 phylogenetically-related species. The colour code indicates the blastn bit score. The  
1250 3'UTR lengths in the corresponding species are indicated and represented as grey  
1251 lines. Black triangles indicate the presence of the UCCCC motif. **(B)** Schematic  
1252 representation of the constructed chimeric mRNAs, which comprises the *S. aureus*  
1253 CDS and a 3'UTR from a different staphylococcal species (for strains and plasmids  
1254 details see supplementary data). Red flags indicate the insertion of the 3xFLAG tag in  
1255 the N-terminus. **(C)** Western blot showing the levels of an orthologous protein when it  
1256 is expressed from different chimeric mRNAs. The western blot was developed using  
1257 peroxidase conjugated anti-FLAG antibodies. A Coomassie stained gel portion is  
1258 shown as a loading control. **(D)** Northern blot showing the mRNA levels expressed  
1259 from the constructions shown in **B**. These were detected using CDS-specific antisense  
1260 radiolabelled RNA probes. Ribosomal RNAs stained with Midori Green are shown as  
1261 loading controls. Western and northern blots images show the representative results  
1262 from at least three independent replicates. Protein and mRNA levels were quantified by  
1263 densitometry of western blot images and northern blot autoradiographies using ImageJ  
1264 (<http://rsbweb.nih.gov/ij/>). Each of the protein or mRNA levels was relativized to the  
1265 levels of *S. aureus* (Sau).

1266

1267 **Figure 5. Disruption of *rpiRc* 3'UTR sequence by IS transposition events affects**  
1268 **RpiRc expression. (A)** Schematic representation showing the main insertion sites of  
1269 ISs in the *rpiRc* 3'UTR. **(B)** Schematic representation showing the putative chimeric  
1270 mRNAs that are generated when IS1181 and IS256 are inserted in the *rpiRc* 3'UTR of  
1271 the *S. aureus* DAR1183 and 2010-60-6511-5 strains, respectively. Since IS1181  
1272 harbours a stem loop that serves as a transcriptional terminator (TT) in the same  
1273 strand as the *rpiRc* gene, which is opposite to the transposase gene, the IS1181  
1274 insertion generates a chimeric 3'UTR that is 222 nt-long. In contrast, the IS256 lacks

1275 TTs, generating a chimeric mRNA that includes the whole IS256 sequence. **(C)**  
1276 Northern blots showing the mRNA levels expressed from the constructs harbouring the  
1277 WT *rpiRc* gene and the IS insertions mimicking the configurations found in the *S.*  
1278 *aureus* DAR1183 and 2010-60-6511-5 strains. The plasmid expressing the *rpiRc*  
1279 mRNA  $\Delta$ 3'UTR was included as a control. The mRNA levels of *rpiRc* were detected  
1280 with CDS-specific antisense radiolabelled RNA probe. Ribosomal RNAs stained with  
1281 Midori Green are shown as loading controls. **(D)** Western blot showing the  $^{3x}F$ RpiRc  
1282 protein levels expressed from the construction used in **C**. The western blot was  
1283 developed using peroxidase conjugated anti-FLAG antibodies. A Coomassie stained  
1284 gel portion is shown as a loading control. The western and northern blot images show  
1285 the representative results from at least three independent experiments. The protein  
1286 and mRNA levels were quantified by densitometry of western blot and northern blot  
1287 autoradiographies using ImageJ (<http://rsbweb.nih.gov/ij/>). Each of the protein or  
1288 mRNA levels was relativized to the levels of WT *rpiRc* mRNA. **(E)** Western blot  
1289 showing the GFP protein levels of the constructs harbouring the *gfp* fused to the 3'UTR  
1290 of *rpiRc*, the 3'UTR+IS256, 3'UTR+IS1181 and the  $\Delta$ 3'UTR as a control. Western blots  
1291 were developed and quantified as described in **D**. **(F)** Haemolytic halos produced by  
1292 the haemolysins contained in the supernatant of cultures of *S. aureus* MW2 wild-type  
1293 (WT) and their isogenic chromosomal mutants carrying the deletion of the *rpiRc* 3'UTR  
1294 ( $\Delta$ 3'UTR) and *rpiRc* gene ( $\Delta$ *rpiRc*) and the insertion of the IS1181 in the *rpiRc* 3'UTR  
1295 (as it occurred in the DAR1183 strain (3'UTR+IS1181)). The supernatants were  
1296 concentrated 10 times and loaded into 5-mm holes made in Columbia Sheep blood  
1297 (5%) agar plates.

1298

1299 **Figure 6. Species-specific variations on *icaR* 3'UTRs resulted in different PIA-**  
1300 **PNAG levels.** Electrophoretic mobility shift assays (EMSAs) between the synthetic *S.*  
1301 *aureus* *icaR* 5'UTR and the synthetic *icaR* 3'UTR from: **(A)** *S. aureus* as a positive  
1302 control; **(B)** *S. argenteus* and *S. simiae*, which carry the UCCCC motif necessary for

1303 5'UTR interaction, as previously described (5); **(C)** *S. epidermidis* and *S. capitis*, which  
1304 lack the UCCCC motif. The autoradiographies of the band-shifts result from incubating  
1305 a <sup>32</sup>P-labeled synthetic *S. aureus* 5'UTR RNA fragment (40,000 cpm) with increasing  
1306 amounts of the different synthetic 3'UTR RNA (100 to 500 nM). The UTR complexes  
1307 are indicated on the right side of the autoradiography. **(D)** Quantification of PIA-PNAG  
1308 exopolysaccharide production in the *S. aureus* 15981 strains expressing the different  
1309 chimeric *icaR* mRNAs. Serial dilutions of the samples were spotted onto nitrocellulose  
1310 membranes and PIA-PNAG was developed with specific anti-PIA-PNAG antibodies. ∅  
1311 indicates the presence of an empty pCN40 plasmid. Sau, *S. aureus*; Sarg, *S.*  
1312 *argenteus*; Ssim, *S. simiae*; Sepi, *S. epidermidis*; Scap, *S. capitis*. Images show  
1313 representative results from at least two independent experiments.

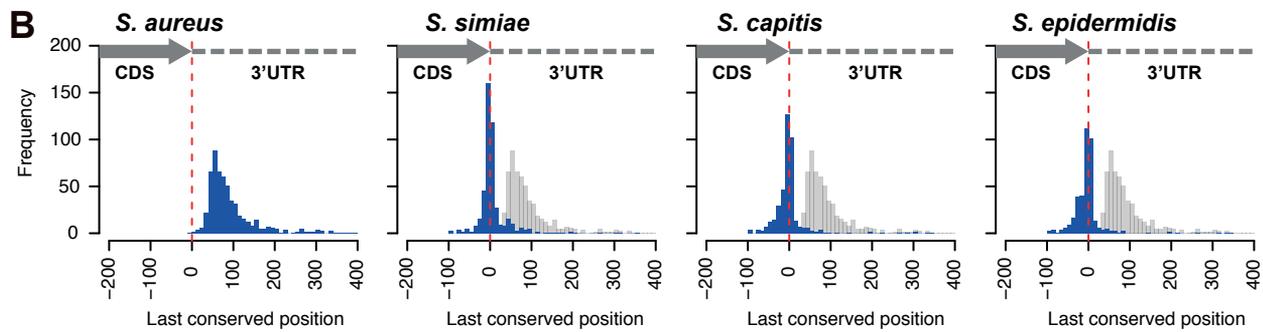
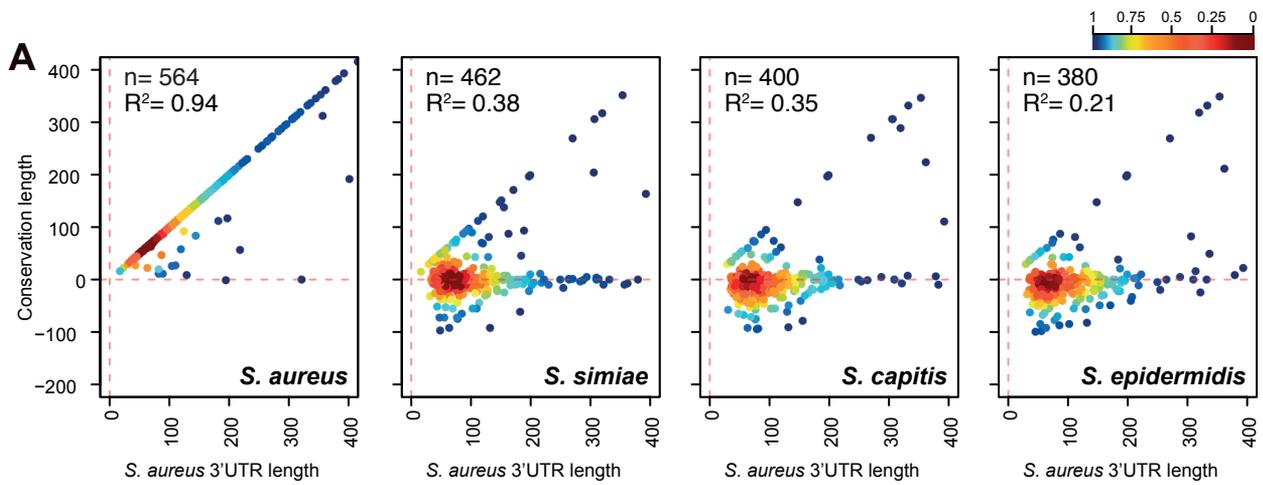
1314

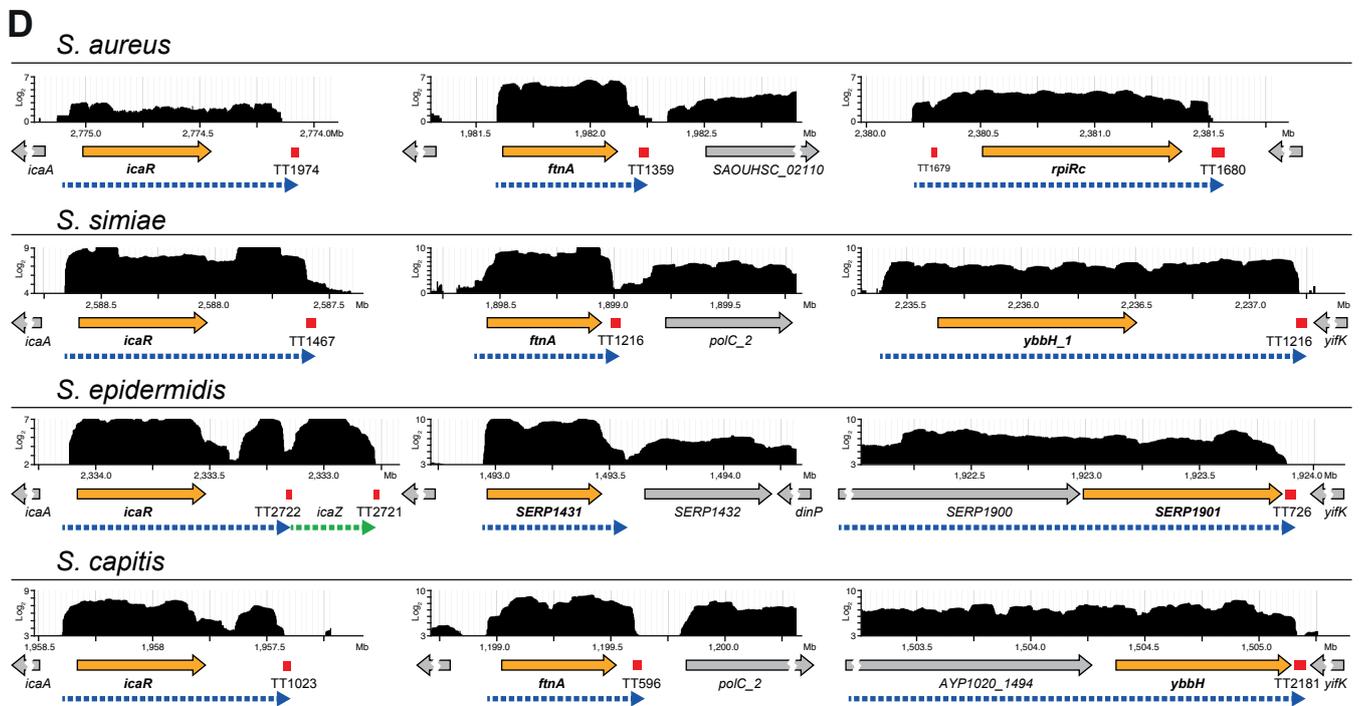
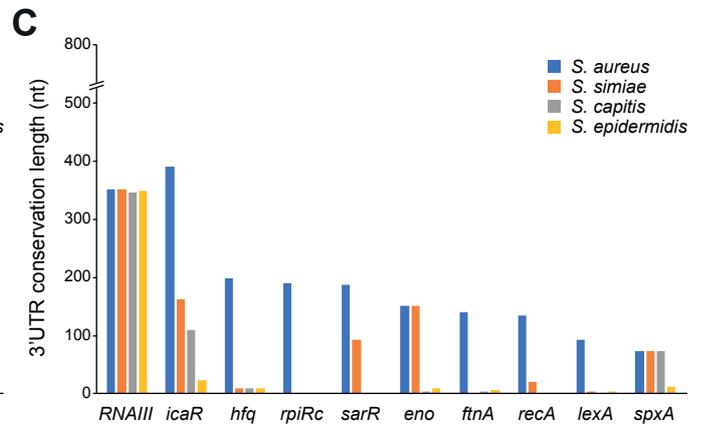
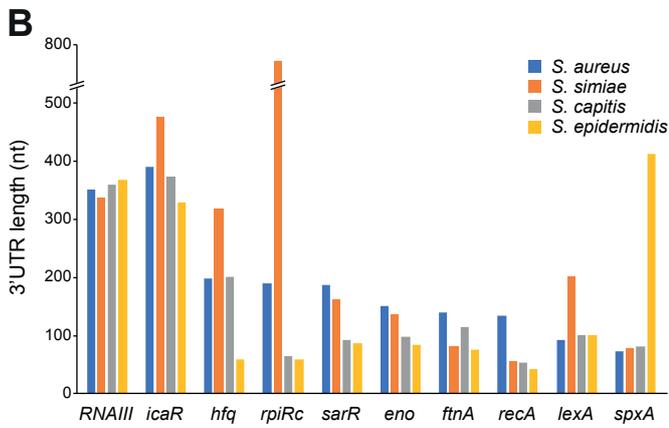
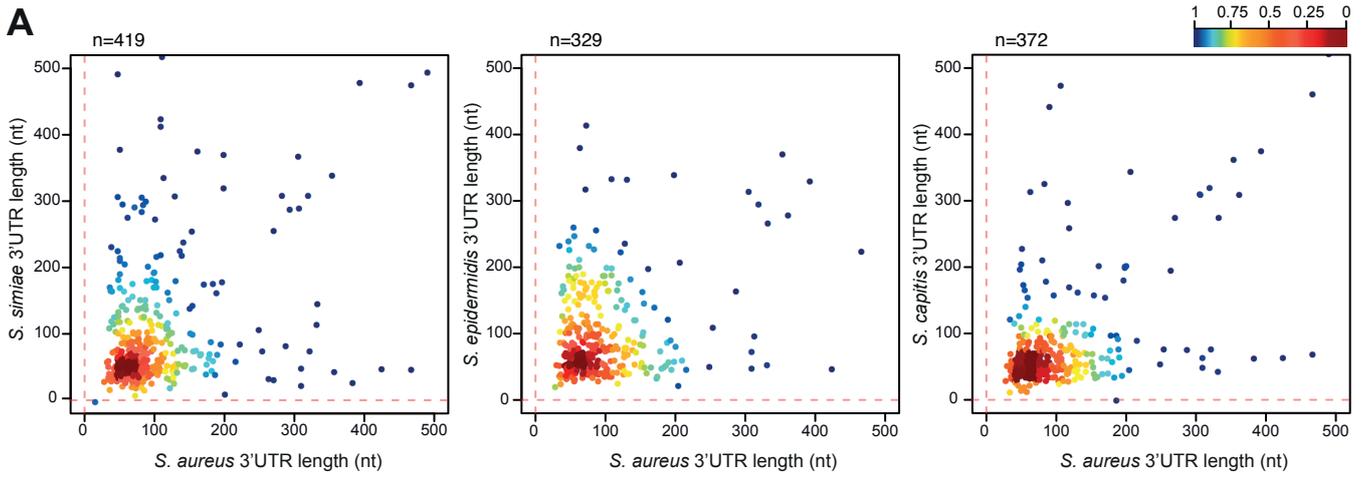
1315 **Figure 7. Schematic representation of the conservation analysis of regulatory**  
1316 **3'UTRs among phylogenetically-related species.** Blastn analyses were performed  
1317 using the *E. coli acnB* **(A)**, *Y. pestis hmsT* **(B)**, *C. glutamicum aceA* **(C)**, and *B. subtilis*  
1318 *hbs* **(D)** mRNAs as queries, which carried 3'UTRs with proven regulatory capacities (4,  
1319 6-8). The colour code indicates the blastn bit score. The estimated 3'UTR lengths are  
1320 indicated and represented by grey lines.

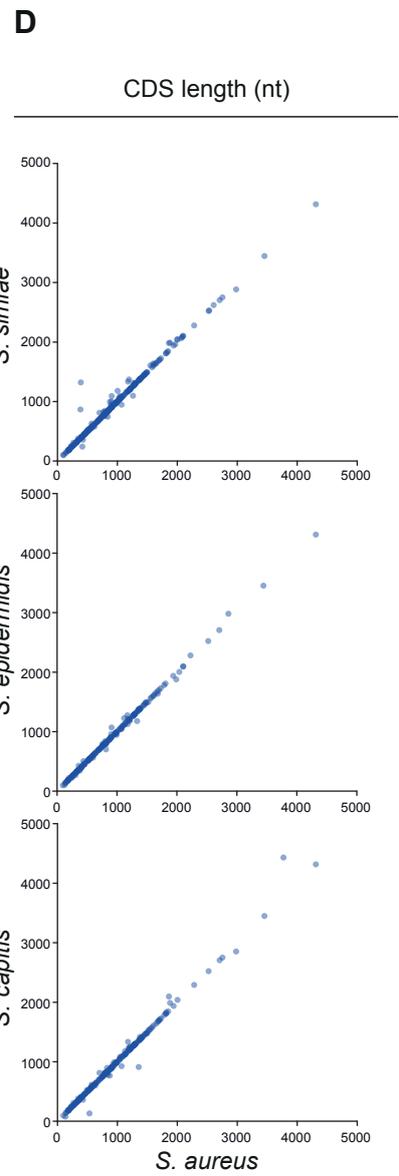
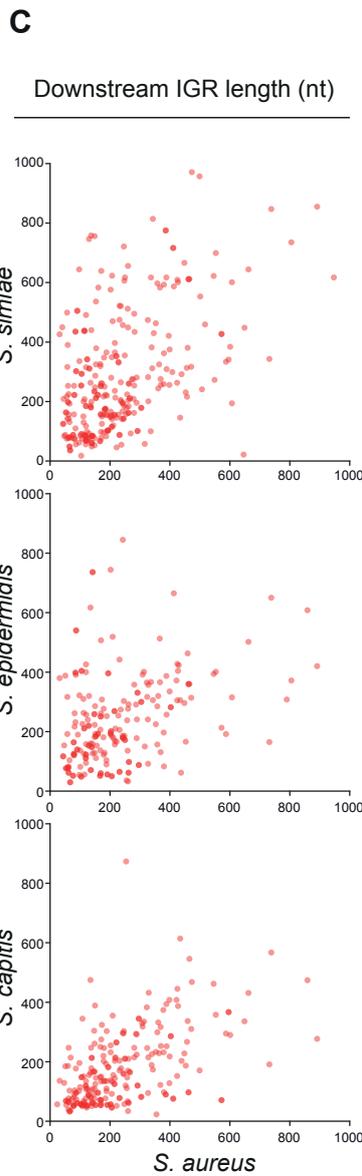
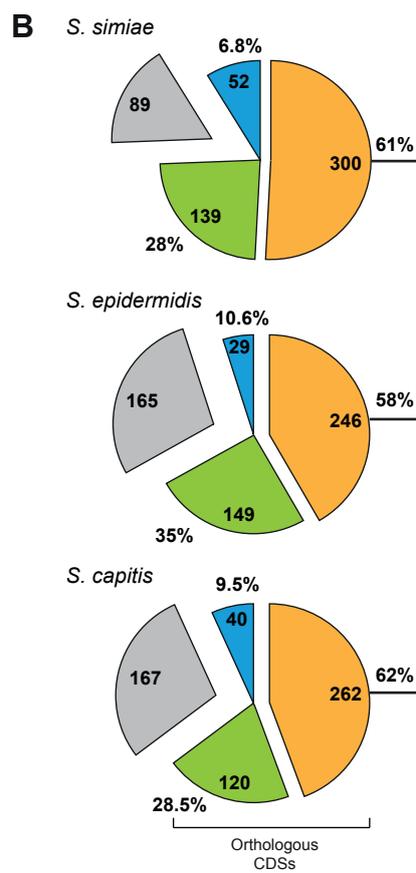
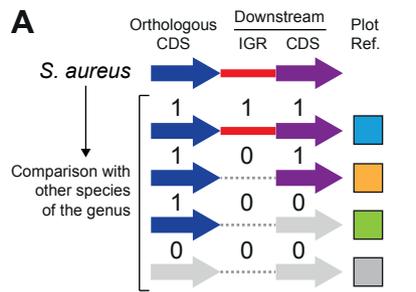
1321

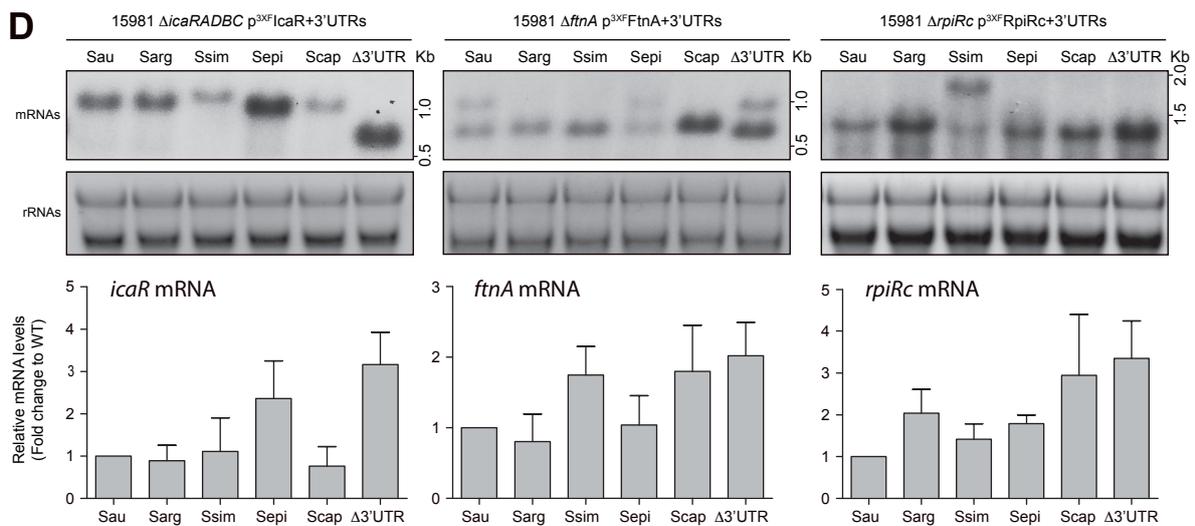
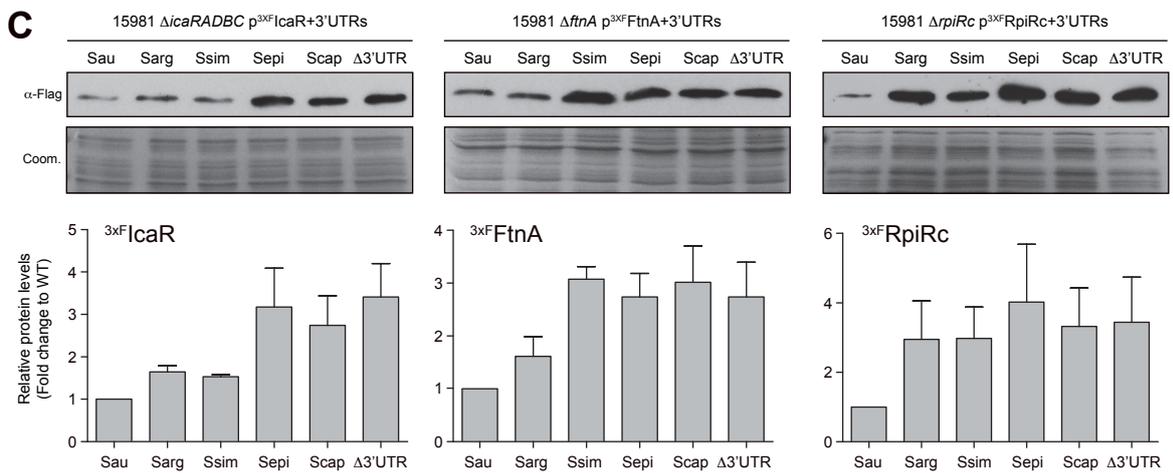
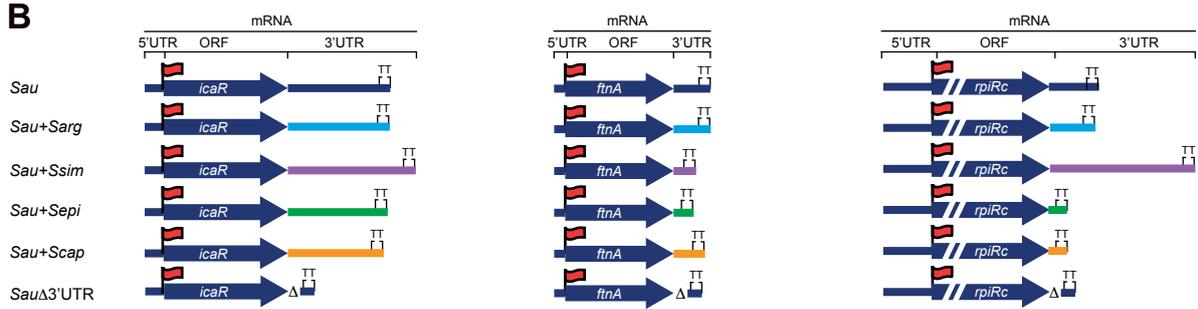
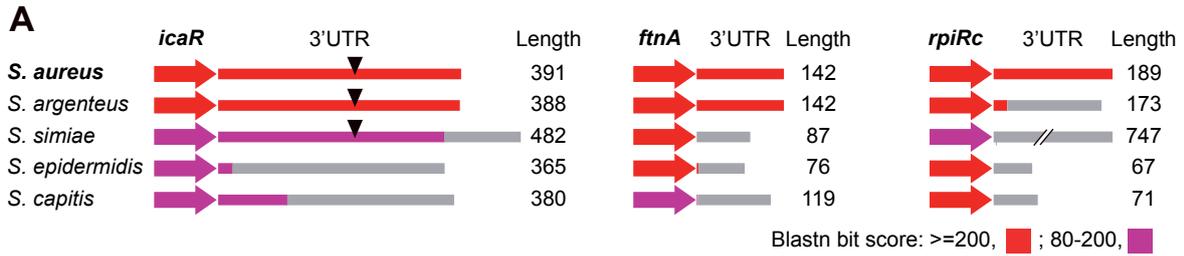
1322 **Figure 8. 3'UTR sequence variations are widely distributed in bacteria.** Scatter  
1323 plots representing the conservation of the 3' end regions of *E. coli* **(A)** and *B. subtilis*  
1324 **(B)** mRNAs compared to their corresponding 3' end regions in phylogenetically-related  
1325 species. The *E. coli* BW25113 3'UTR sequence query database was compared to *E.*  
1326 *coli* O157-H7 str EC4115, *Citrobacter koseri* ATCC BAA-895, *Salmonella* Typhimurium  
1327 SL1344, and *Enterobacter aerogenes* KCTC 2190. The *B. subtilis* 168 3'UTR  
1328 sequence query database was compared to *B. subtilis* OH 131.1, *B. amyloliquefaciens*  
1329 DSM7, *B. licheniformis* ATCC 14580, and *B. pumilus* SH-B9. Each dot represents the  
1330 last conserved nucleotide, according to blastn, of an indicated species (y axis) in

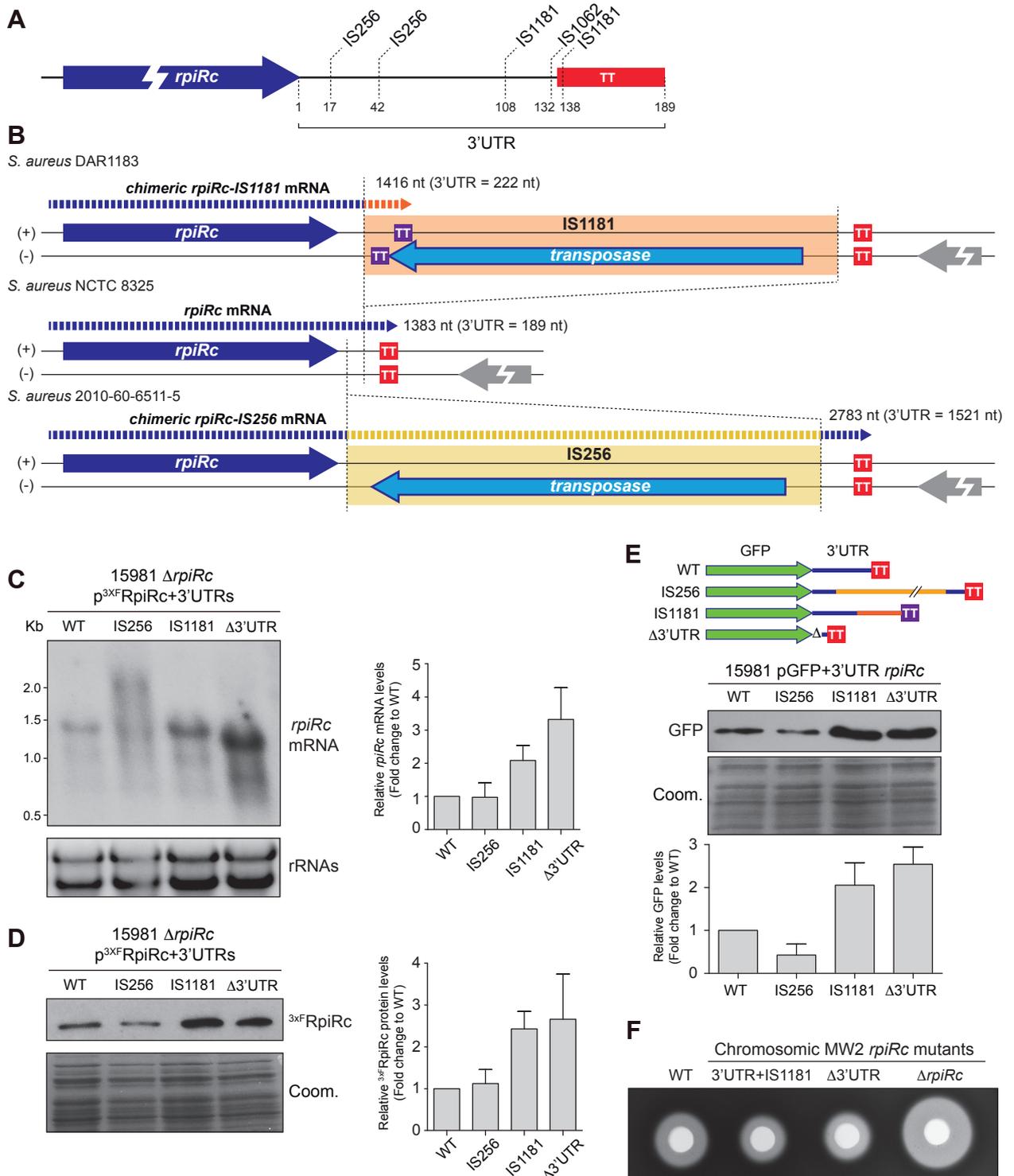
1331 function of the *E. coli* and *B. subtilis* 3'UTR lengths (x axis), respectively. The plot was  
1332 coloured by applying the Kernel density estimation, which indicates the proximity of the  
1333 dots. Blue: isolated dots; red: overlapping dots. The number (n) of plotted mRNAs  
1334 encoding orthologous proteins and the square correlation coefficient ( $R^2$ ) are indicated.

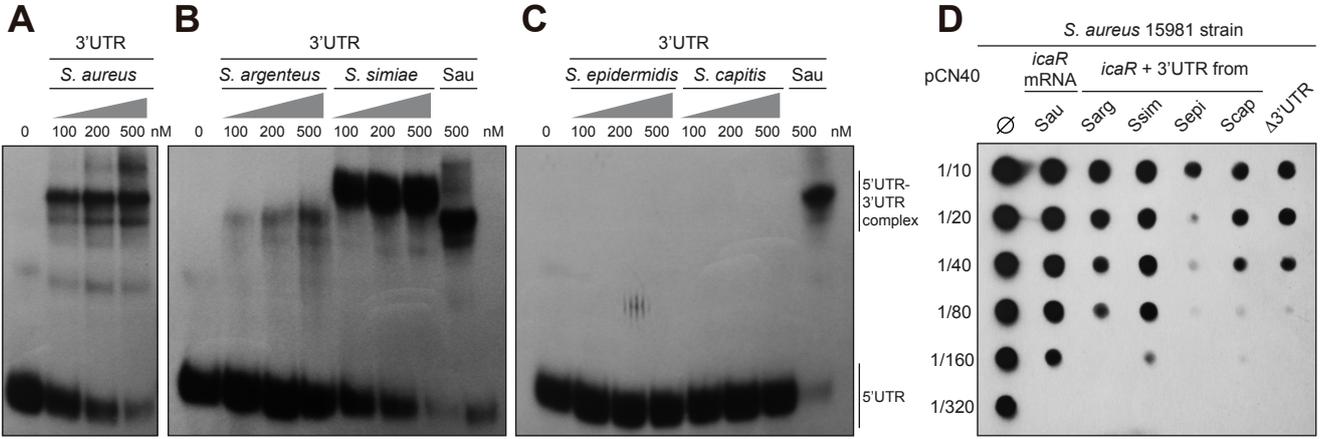


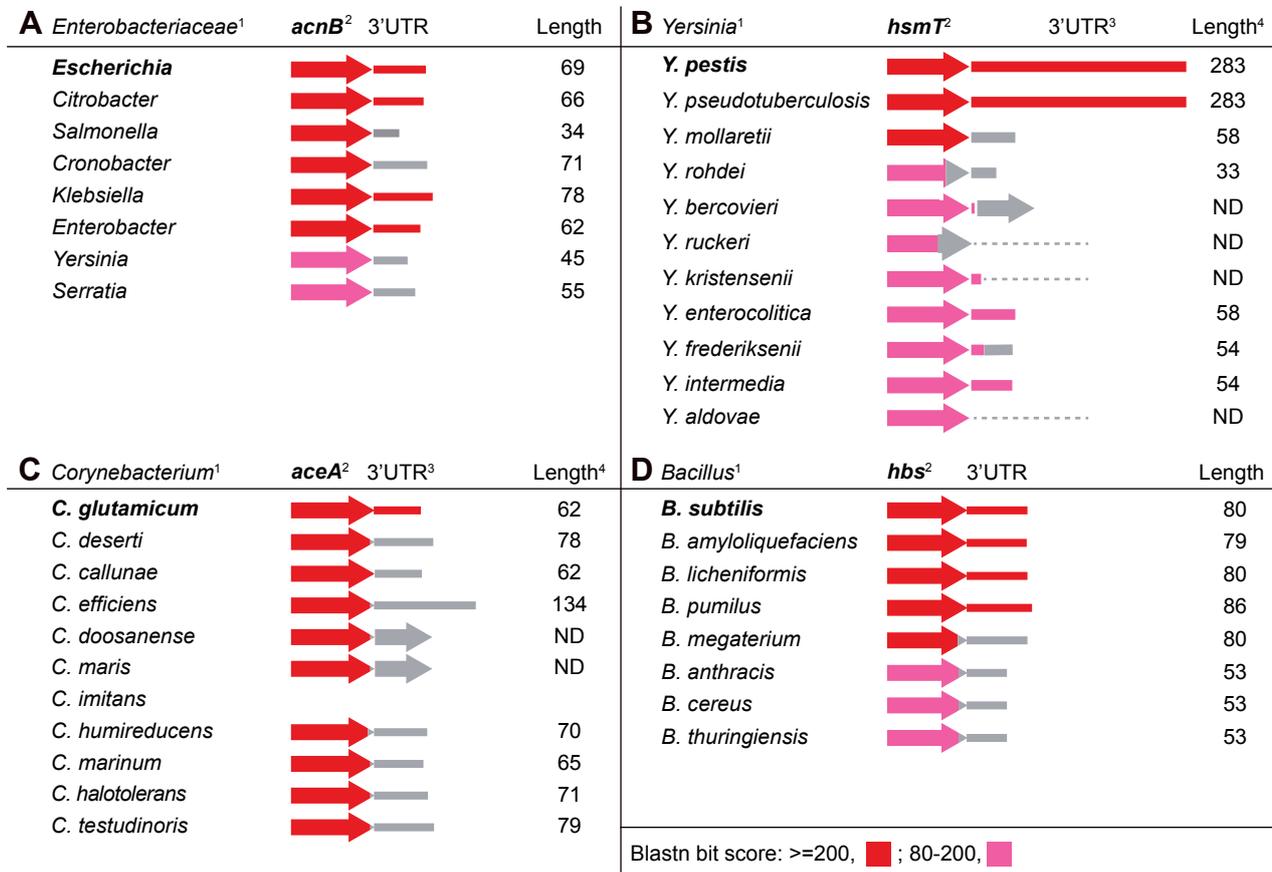












<sup>1</sup> Representative bacterial species according to the phylogenetic tree built at TimeTree knowledge-base ([www.timetree.org](http://www.timetree.org))

<sup>2</sup> Color code indicates blastn bit score. Gray lines represent the 3'UTR length in the corresponding species, which is non-conserved.

<sup>3</sup> Dashed line indicates absence of a putative TT, gray arrow indicates the presence of a CDS instead of a 3'UTR.

<sup>4</sup> ND, 3'UTR length not determined.

